PERSPECTIVE ARTICLE

The Sequenced Genomes of Non-Flowering Land Plants Reveal the (R)Evolutionary History of Peptide Signaling

Chihiro Furumizu,[a1] Anders K. Krabberød,[b] Marta Hammerstad,[c] Renate M. Alling,[b] Mari Wildhagen,[b] Shinichiro Sawa,[a,2] and Reidunn B. Aalen[b,3]

[a] Graduate School of Science and Technology, Kumamoto University, Kumamoto, Japan.

[b] Section for Evolutionary Biology and Genetics, Department of Biosciences, University of Oslo, Norway

[c] Section for Biochemistry and Molecular Biology, Department of Biosciences, University of Oslo, Norway

[c] Contact information:

[1] furumizu@kumamoto-u.ac.jp

[2] sawa@kumamoto-u.ac.jp.

[3] Corresponding author: Reidunn B. Aalen (reidunn.aalen@ibv.uio.no)

Short title: (R)Evolutionary Peptide Signaling History

One sentence summary: Successful identification of orthologues of Arabidopsis signaling peptides and their receptors in non-flowering plants suggest their importance in cell-to-cell communication in all land plants .

The author(s) responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) are: Reidunn B. Aalen (reidunn.aalen@ibv.uio.no) and Chihiro Furumizu (furumizu@kumamoto-u.ac.jp).

1

Abstract

An understanding of land plant evolution is a prerequisite for in-depth knowledge of plant biology. Here we extract and explore information hidden in the increasing number of sequenced plant genomes, from bryophytes to angiosperms, to elucidate a specific biological question – how peptide signaling evolved. To conquer land and cope with changing environmental conditions, plants have gone through transformations that must have required a revolution in cell-to-cell communication. We discuss peptides mediating endogenous and exogenous changes by interaction with receptors activating intracellular molecular signaling. Signaling peptides were discovered in angiosperms and operate in tissues and organs like flowers, seeds, vasculature, and 3D meristems that are not universally conserved across land plants. Nevertheless, orthologues of angiosperm peptides and receptors have been identified in non-flowering plants. These discoveries provoke questions regarding the co-evolution of ligands and their receptors, and whether *de novo* interactions in peptide signaling pathways may have contributed to generate novel traits in land plants. The answers to such questions will have profound implications for the understanding of evolution of cell-to-cell communication and the wealth of diversified terrestrial plants. Under this perspective we have generated, analyzed and reviewed phylogenetic, genomic, structural, and functional data to elucidate the evolution of peptide signaling.

INTRODUCTION

Land plants have evolved from freshwater green algae that started to colonize land ~ 470 million years ago (Mya) (Delwiche and Cooper, 2015; Ishizaki, 2017). This successful evolution has required groundbreaking biological innovations, such as the transition from the ancestral algal life cycle of haploid multicellular plants with zygotic meiosis to that with multicellular alternating haploid and diploid generations. While the diploid sporophyte and the haploid gametophyte initially might have employed the same genes, gene duplications followed by neofunctionalization likely facilitated the development of new specialized organs that necessitated extensive changes in gene expression patterns (Ligrone et al., 2012; Bowman et al., 2017; Jill Harrison, 2017) (Figure 1). Lycophytes and ferns, which appeared ~430 Mya (Kenrick and Crane, 1997; Gensel, 2008; Steemans et al., 2009; Morris et al., 2018), are extant plants representing early diverging lineages with vasculature, which provided mechanical support and transport of nutrients (Ishizaki, 2017). Roots evolved independently in lycophytes and euphyllophytes (vascular plants excepting lycophytes); they provide anchorage for the plants and allow acquisition of water and nutrients from the soil (Raven and Edwards, 2001; Hetherington and Dolan, 2019). The evolution of novel cell types, tissue types, and organs with increasing multicellular complexity must have put a high demand on cell-to-cell communication in land plants (Grienenberger and Fletcher, 2015).

Recently, genomes of several hornwort *Anthoceros* species, the liverwort *Marchantia polymorpha,* and ferns (*Azolla filiculoides* and *Salvinia cucullata*) have been sequenced (Bowman et al., 2017; Li et al., 2018; Li et al., 2020; Yang et al., 2020; Zhang et al., 2020a). Together with the pioneering genome projects of the moss *Physcomitrium* (previously *Physcomitrella* (Rensing et al., 2020)) *patens*, the lycophyte *Selaginella moellendorffii*, the gymnosperm Norway spruce (*Picea abies*), and *Amborella trichopoda* representing the sister lineage to all other flowering plants, these resources have facilitated large-scale comparative genomic studies, which help us understand the evolution of land plants (Rensing et al., 2008; Banks et al., 2011; Amborella Genome, 2013; Nystedt et al., 2013). Advances in whole genome sequencing in streptophyte algae further provide opportunities to elucidate evolutionary processes of plant terresteralization (Nishiyama et al., 2018; Cheng et al., 2019; Chen et al., 2020; Jiao et al., 2020). For instance, the ancestry of several hormone and stress signaling pathways in land plants has been traced back to algae (Bowman et al., 2017; de Vries et al., 2018), underscoring the fundamental importance of these pathways for vegetative life.
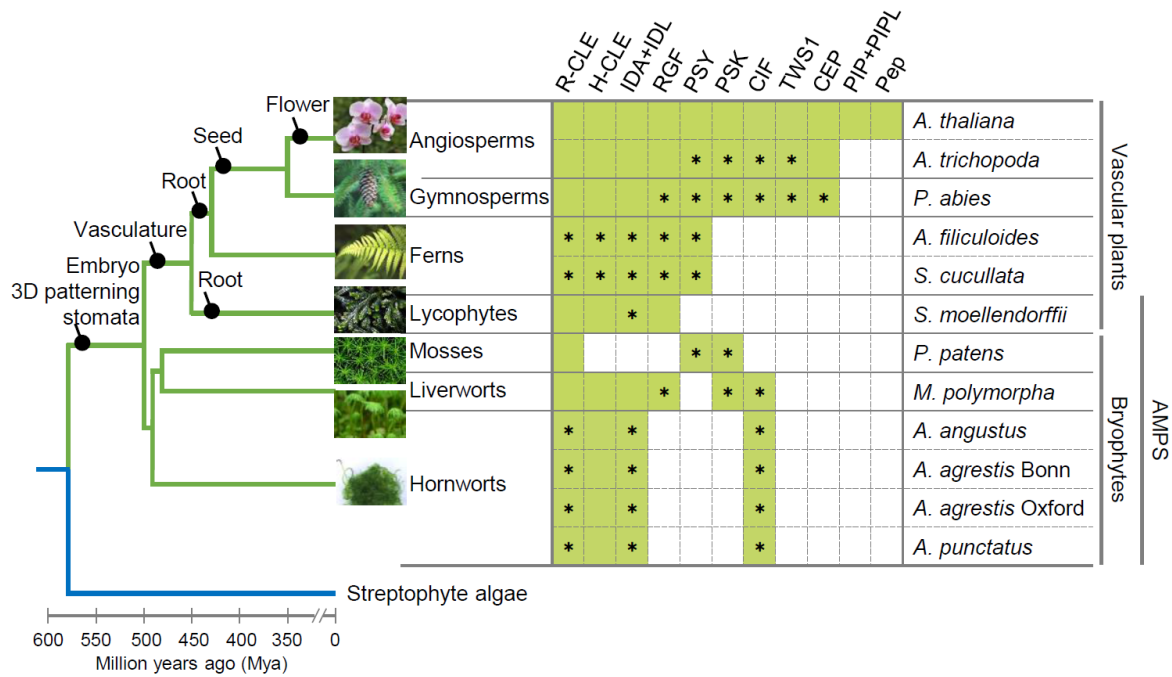
3

Figure 1. Phylogenetic distribution of small post-translationally modified peptides in land plants.

Left: Simplified phylogenetic tree of major extant land plant lineages (green lines) based on (Transcriptomes, 2019). The branch lengths are roughly proportional to the estimated divergence dates (Morris et al., 2018). Gains of key morphological innovations are shown as black circles mapped on the tree while some gain events are still under discussion.

Right: Green-colored boxes indicate the presence of homologues of the indicated small post-translationally modified signaling peptides for the model plants *Arabidopsis thaliana* (Lamesch et al., 2012), *Amborella trichopoda* (Amborella, 2013) (Amborella, 2013), *Picea abies* (Nystedt et al., 2013), *Azolla filiculoides* and *Salvinia cucullata* (Li et al., 2018), *Selaginella moellendorffii* (Banks et al., 2011), *Physcomitrium patens* (Lang et al., 2018), *Marchantia polymorpha* (Bowman et al., 2017), and *Anthoceros* hornworts (Li et al., 2020; Zhang et al., 2020a). Asterisks indicate putative peptides newly identified in our sequence analyses. Images represent each taxa and are from Keiko Sakakibara (*A. agrestis* Oxford) and Pixabay (https://pixabay.com), except for *M. polymorpha*.

Meanwhile, in flowering plants (angiosperms) the position of the classical hormones as the major mediators of signaling processes has over the last decades been challenged by small peptides interacting with leucine-rich repeat ectodomains of plasma-membrane-bound receptors with cytoplasmic serine/threonine kinase domains (KDs) (Matsubayashi, 2014; Olsson et al., 2019). Traditional hormones work over longer distances giving a so-called endocrine effect, while most peptides work over short distances and seem ideal for communication between neighboring cells with different functions (Wheeler and Irving, 2010). Changes in paracrine effects are likely to be instrumental in producing functional novelty or increasing complexity in an organism.

4

The first small signaling peptide was discovered in plants thirty years ago (Pearce et al., 1991), while more than 200 genes encoding leucine-rich repeat receptor-like kinases (LRR RLKs) were identified ten years later in the sequenced genome of *Arabidopsis thaliana* (Shiu and Bleecker, 2001b). They were assigned to fifteen subfamilies according to their unique structure, organization, and number of LRRs. Since then the increase in the number and quality of sequenced genomes, and advances in methods for *in silico* annotation of small genes, have together allowed the identification of hundreds of potential signaling peptides (Gong et al., 2002; Ghorbani et al., 2015; Tavormina et al., 2015; Goad et al., 2017). The peptides with known receptors represent two major groups, the cysteine-rich peptides (CRPs) (Supplemental Table 1), which attain their 3D structure by disulfide bridges formed between pairs of cysteines, and the post-translationally modified peptides (PTMPs, Table 1 and Supplemental Table 2), generated from prepropeptides and processed to mature peptides of 5-20 amino acids (Matsubayashi, 2011). To date, LRR-RLKs of the X, XI, and XIII subfamilies, with ligand-binding ectodomains built of >20 units of leucine-rich repeats, have been identified as main receptors for both groups of endogenous small signaling peptides. Here we will focus on the ancestry of the known PTMPs, which almost exclusively signal through the Arabidopsis receptor subfamily XI (hereafter LRXI).

LRR-RLKs have from their discovery been studied in a phylogenetic and evolutionary perspective (Shiu and Bleecker, 2001b; Lehti-Shiu et al., 2009; Liu et al., 2017; Chakraborty et al., 2019; Hosseini et al., 2020; Man et al., 2020). Signaling peptides have on the other hand been discovered due to their involvement in very "angiospermish" processes, like root growth, organ abscission, or seed development, and the phylogenetic investigations have focused on their distribution and function within the orders of flowering plants (Delay et al., 2013; Lori et al., 2015; Stø et al., 2015). Subsequent studies have matched LRR-RLKs with endogenous peptides as signaling pairs in angiosperms. Utilizing the increasingly available genome sequences of non-flowering plants, it is now feasible to elucidate peptide signaling in a phylogenetic and evolutionary perspective, taking into account the likely co-evolution of ligands and receptors.

LRXIs have been discovered in *M. polymorpha, P. patens* and *S. moellendorffii* (Bowman et al., 2017; Liu et al., 2017), but a detailed analyses of their relation to the well-studied Arabidopsis receptors has not been performed. Under the assumption that non-angiosperm LRXI receptors are activated by PTMP ligands, this discovery triggers questions regarding the

5

presence and the roles of signaling peptides in different lineages of land plants. Changes in peptide signaling systems might be linked to cell-to-cell communication pathways that facilitated the evolution from ancestral algae to land plants and revolutionized terrestrial ecosystems. To address this hypothesis, we started by going through literature and data on peptide signaling and LRXI. A number of papers have reviewed specific peptide or receptor families involved in peptide signaling (Fernandez et al., 2013; Muschietti and Wengier, 2018; Oh et al., 2018; Taleski et al., 2018; Kaufmann and Sauter, 2019; Segonzac and Monaghan, 2019; Shi et al., 2019) and (Supplemental Table 2). It has not been, however, comprehensively investigated how widely these receptors and ligands are conserved; here we identified the need to search for homologues of Arabidopsis *LRXI* genes in non-flowering plants as well as genes encoding specific peptides to match the presence of their presumptive receptors. Thus, our *Perspective* reviews published data but is also supplemented with novel findings. All genes and proteins used are listed in Supplemental Dataset 1. We offer cross-comparison of multiple peptide ligand-receptor pairs and argue that if we are to understand the evolutionary history and possibly revolutionary importance of signaling peptides, we need to study cell-to-cell communication through peptide signaling pathways in an integrative manner, using diverse *in silico* methods that can generate hypotheses to be addressed by experimental approaches.

PTMPs in land plants

PTMPs are generated from prepropeptides by post-translational processing and amino acid modifications (Matsubayashi and Sakagami, 2006; Olsson et al., 2019). Gene and genome duplications have led to evolutionary changes, but the amino acid sequences of C-terminal bioactive peptides have been conserved. Members of ten families of peptides with characteristic length and amino acid composition have been matched with LRR-RLKs (Table 1).

The CLAVATA3/EMBRYO SURROUNDING REGION-related (CLE) peptides have the broadest distribution, from bryophytes to angiosperms, and has largest family in Arabidopsis with 32 members (Jun et al., 2008; Goad et al., 2017). The 12-13 amino acids long CLE peptides are proline (Pro, P)-rich and characteristically have histidine-asparagine (His, H and Asn, N) or two His residues at the C-terminus (Table 1). CLE peptides are classified into two groups by the N-terminal residue: the R (arginine, Arg) and H groups. Many R-CLEs interact with CLAVATA1 (CLV1) or BARELY ANY MERISTEM (BAM) receptors involved in meristem maintenance in Arabidopsis (DeYoung et al., 2006; Hazak and Hardtke, 2016) while

6

the three H-CLE peptides interact with the PHLOEM INTERCALATED WITH XYLEM/TDIF RECEPTOR (PXY/TDR) involved in vascular differentiation in Arabidopsis (Hirakawa et al., 2010). The CLE peptides are the first PTMPs from bryophytes that have been assigned with biological functions. In *M. polymorpha* an R-type peptide (MpCLE2) functions in a paracrine manner as a haploid stem cell-promoting signal in dichotomous branching of the thallus and is genetically dependent on the MpCLV receptor (Hirakawa et al., 2020).

Table 1. Major PTMP families in land plants.

| Peptide family[1] | Major receptor(s) (LR subfamily)[1] | Amino acid sequence of representative mature peptide[2] | |
		A. thaliana	*M. polymorpha*
CEP	CEPR (XI)	CEP1: DFRQTNPGNSQGVGH	
CIF/TWS1	GSO1/SGN3, GSO2 (XI)	CIF1: DYGNNSQSQRLERPPFKLIPN TWS1: DYNFPVDPTPTTKASIKPGPIEH	Putative MpCIF: DYCDCYGFDPSPPLVHAEITF
CLE (CLV3/R-type)	CLV1, BAM (XI)	CLV3: RTVQSGoDPLHH	MpCLE2: KEVPNGPNPLHN
CLE (TDIF/H-type)	PXY/TDR (XI)	CLE41: HEVQSGQNPISN	MpCLE1: HKNPAGPNPIGN
IDA + IDL	HAE, HSL (XI)	IDA: PIPPSAQSKRHN	MpIDA1: EVPPQGPSPIHN
Pep	PEPR (XI)	Pep1: ATKVKAKQRGKEKVSSGRPGQHN	
PIP + PIPL	RLK7 (XI)	TOLS2/PIPL3: ASGQSRRGAGH	
PSK	PSKR (X)	PSK1: YIYTQ	Putative MpPSK: YIYTSQ
PSY	PSYR (X)	PSY1: DYGDPSANPKHDPGVoOS	
RGF/GLV/CLEL	RGFR/RGI (XI)	RGF1/GLV11/CLEL8: DYSNPGHHPORHN	Putative MpRGF: DYAEPDTHPPESN

[1]See Supplemental Data Set 1 for corresponding gene IDs.
[2]O, hydroxylated proline; o, arabinosylated hydroxyproline; and Y, sulfated tyrosine.

The H-type peptide MpCLE1, with an MpTDR receptor, is on the other hand confined to the proliferative activity of gametophytic meristem and affects the overall size of reproductive organs (gametangiophore) (Hirakawa et al., 2019). In *P. patens* the R-type PpCLE peptide with its PpCLV receptor is involved 3D growth by determination of the orientation of stem cell division planes (Whitewoods et al., 2018). The function of the H-type CLE peptide recently detected in *Anthoceros* remains to be revealed (Li et al., 2020; Zhang et al., 2020b).

INFLORESCENCE DEFICIENT IN ABSCISSION (IDA) and the IDA-LIKE peptides are, similar to CLEs, 12-14 amino acid long P-rich peptides with C-terminal HN residues (Butenko et al., 2003; Jun et al., 2008; Vie et al., 2015) (Table 1). In Arabidopsis IDA signals trough HAESA (HAE) and HAE-LIKE2 (HSL2) to trigger cell wall remodeling leading to cell

separation, in processes like floral organ abscission, lateral root emergence, and root cap sloughing (Cho et al., 2008; Stenvik et al., 2008; Butenko et al., 2014; Shi et al., 2018). Orthologues have been identified in all orders of flowering plants (Shi et al., 2019). Most recently, genes encoding putative IDA peptides and HSL receptors were identified in *M. polymorpha* (Bowman et al., 2017).

ROOT GROWTH FACTOR (RGF), also called GOLVEN (GLV) or CLE-LIKE (CLEL) peptides, are 12 amino acids long with Asp (Asp, D) and a sulfated tyrosine (Tyr, Y) at the N-terminus (Matsuzaki et al., 2010; Meng et al., 2012; Fernandez et al., 2013) (Table 1). Members of this family serve diverse roles in root development, like meristem homeostasis, gravitropism, and lateral root development (Fernandez et al., 2013). In addition to angiosperms, RGFs have been suggested to be present in the lycophyte *S. moellendorffii* (Ghorbani et al., 2015). RGFs can signal through RGF RECEPTOR (RGFR)/ROOT GROWTH FACTOR INSENSITIVE (RGI) receptors, which were discovered a few years ago (Shinohara et al., 2016; Song et al., 2016; Qiu et al., 2020).

When we initiated our study, the remaining peptide families of CASPAIAN STRIP INTERGRATING FACTORs (CIFs), CARBOXY-TERMINALLY ENCODED PEPTIDEs (CEPs), PAMP-INDUCED SECRETED PEPTIDEs (PIP/PIP-LIKE), PLANT ELICITOR PEPTIDES (Peps), Phytosulfokine (PSK), and PLANT PEPTIDEs CONTAINING SULFATED TYROSINE (PSY), Table 1) had not (yet) been identified in non-seed plants. However, the presence of putative *LRXI* (Bowman et al., 2017; Liu et al., 2017), raised the question of when signaling peptides evolved. This has not been well addressed in the literature; therefore we started to look for these peptide ligand homologues.

Methods for identification of peptide orthologues

A useful initial procedure for identification of homologues of Arabidopsis peptide-encoding genes is to search among more closely-related species, like monocots or *A. trichopoda*, using conventional Basic Local Alignment Search Tool for proteins (BLASTp) (Altschul et al., 1990) (https://blast.ncbi.nlm.nih.gov/Blast.cgi), confirm the homology of the hits with reciprocal BLASTp, and thereafter use identified peptides as queries in subsequent searches in gymnosperms and ferns, and later lycophytes and bryophytes. It is however, generally difficult to find distantly related peptide by BLASTp since prepropeptide sequences are highly variable except for the short peptide-encoding region. Small open reading frames encoding peptide

precursors are additionally often overlooked during the annotation of newly sequenced genomes. Consequential false-negative problems may possibly be circumvented by using exhaustive tBLASTn searches against whole genome sequences or RNA collections translated in all six frames (Altschul et al., 1997) .

BLASTp searches work more efficiently when limiting the query sequence to the region encompassing the mature peptide and using the BLAST-short option with the PAM30 matrix (Altschul et al., 1997). An alternative approach is to specify conserved motifs of a given peptide family in Pattern Hit Initiated (PHI) BLAST searches, and/or use queries with wildcards in positions where the amino acid residues are less conserved, e.g. we identify RGF homologues using the query DYxxxxxKPPIHN. PHI-BLAST with the DY motif facilitated also successful identification of *P. patens* PSK and PSKY1, and *M. polymorpha* CIF and PSK against the NCBI database (Zhang et al., 1998). The amino acids HN or HH, found at the C-terminus in several peptide families (Table 1), may also be used in PHI-BLAST. The likelihood of getting false-positive hits in BLAST searches increases with the size of the database searched, and it is therefore helpful to limit searches to coding sequences of less than 300 amino acids. This measure was employed to retrieve putative CIF and IDA peptides in *Anthoceros* species by the EMBOSS 6.5.7 tool *fuzzpro* (http://emboss.open-bio.org/rel/rel6/apps/fuzzpro.html). Queries with wildcards were generated based on consensus sequences from MUSCHLE alignments (Edgar, 2004) of mature peptides of Arabidopsis, *A. trichopoda, P. glauba, M. polymorpha* and *S. moellendorffii,* (DYxxxxPxPPLxxPxPF and PIPxSxPSKRHN, respectively) and used against protein libraries generated from gene-annotated *Anthoceros* genomes. Further options may be to test Machine Learning or Deep Learning methods that recently have been developed for peptide discovery (Plisson et al., 2020; Serrano, 2020; Zhang et al., 2020b).

The confidence of retrieved hits is usually low, and hits must therefore in all cases be analyzed for an appropriate prepropeptide length (<200 amino acids), the position of the assumed mature peptide sequence near or at the C-terminal end and the presence of a hydrophobic N-terminal secretion signal. When interpreting the sequences, it has to be kept in mind that sequences from extant species may have diversified from ancestral sequences.

Using these strategies above, we have identified putative peptide gene homologs in hornworts, lycophytes, ferns, and gymnosperms (Figure 1, indicated by asterisks, see Supplemental dataset 1), suggesting that orthologues of the respective interacting LRXI should be present.

LRR-RLK subfamily XI in land plants

Shiu and Bleecker pioneered the characterization of RLKs in Arabidopsis and also initiated a comparison to other species (Shiu and Bleecker, 2001b; a). The total number of *LRR-RLK* genes in land plant species varies approximately from 60 to 400 in the genomes analyzed to date (Liu et al., 2017). The ectodomain is characteristic for each of the fifteen LRRs subfamilies, but the KD, as well as the scaffold amino acids of the LRRs ectodomain, are highly conserved. Therefore BLASTp searches of LRR-RLKs return many high-scoring hits. The challenge is to identify the few top hits that are true orthologues of the query protein: This can be achieved by reciprocal BLASTp searches in combination with phylogenetic analyses where paralogues will branch off outside the clade with the original query sequence. Phylogenetic analyses of all RLK subfamilies have therefore been based on the KD, often with limited taxon sampling (Zan et al., 2013; Wei et al., 2015; Liu et al., 2016; Magalhães et al., 2016; Zhou et al., 2016; Bowman et al., 2017; Lin et al., 2018; Man et al., 2020). This gives ample information on the relation between different subfamilies.

Our focus, however, is different; it is to explore the evolutionary history of the group LRXI receptors known to interact with signaling peptides in angiosperms. One of the most plausible hypotheses of the land plant phylogeny proposes an early separation of both monophyletic bryophyte and vascular plant lineages (Figure 1) (2019; Li et al., 2020; Zhang et al., 2020b). We have therefore sampled putative LRXI receptors from *Anthoceros* hornworts, the liverwort *M. polymorpha*, the moss *P. patens*, and the lycophyte *S. moellendorffii* (hereafter collectively referred to as AMPS), which are extant species representing lineages that arose from ancient splits in land plant diversification before the emergence of angiosperms.

Methods for phylogenetic analyses

In order to explore the origins and evolution of the group LRXI peptide-interacting receptors, we used 27 full-length receptors encoded by *LRXI* genes of Arabidopsis as individual queries in BLASTp searches against protein sequences from 11 species in the Phytozome database database (v13), the NCBI Genome database (https://www.ncbi.nlm.nih.gov/genome/), and the Hornworts database (https://www.hornworts.uzh.ch/en.html). These species were selected to cover the major evolutionary clades of land plants. In addition to the AMPS including a recently sequenced hornwort (*Anthoceros agrestis*) (Frangedakis et al., 2020; Li et al., 2020; Zhang et al., 2020a), we sampled *LRXIs* from banana, rice and maize monocots (*Musa*

acuminata, Oryza sativa, and *Zea mays*), and three dicots, *Capsella rubella, Populus trichocarpa* and *Solanum lysopersicum* (pink Shepherd's-purse, poplar, and tomato) (Supplemental data set 1). *A. trichopoda* was included in the analyses as a landmark for identification of evolutionary changes that took place before and after the emergence of angiosperms. Finally, we identified LRR-RLKs with more than 20 LRRs from the recently sequenced algae *Penium margaritaceum* (the Penium Genome Database (http://bioinfo.bti.cornell.edu/cgi-bin/Penium/home.cgi) (Jiao et al., 2020), *Spirogloea muscicola,* and *Mesotaenium endlicherianum* (Cheng et al., 2019) to be used as outgroup.

For each of the proteins encoded by the 27 Arabidopsis *LRXI* genes, we retrieved the 20 top BLASTp hits from each of the genomes. Duplicates and isoforms were removed before the candidate LRR sequences (1521 in total) were aligned with the l-INS-I algorithm in MAFFT v7.453 (Katoh and Standley, 2013). Ambiguously aligned characters were masked with trimAl (Capella-Gutiérrez et al., 2009) using the *gappyout* algorithm, and a preliminary phylogenetic tree was constructed with FastTree2 (Bennett and Scheres, 2010). This tree was used to identify sequences that did not belong to XI-type LRR-RLKs. The final dataset consisted of 249 sequences which were unmasked, and after annotation of domains with Interproscan5 (Jones, 2014) were split on each side of the transmembrane region to generate an LRR and a KD dataset. (Supplemental Dataset 1) Since the ectodomains of the LRXI have at least 20 LRRs and can be aligned, we took the opportunity to compare the ligand-interacting domain of these receptors, not only the KD. After separate realignment with MAFFT l-INS-I, and trimming with trimAl (*gappyout*, the LRR-containing ectodomain alignment was 954 amino acids long, and the KD alignment was 439 amino acids long (see Supplemental data sets 1, 2 and 3 for sequences and alingments).

Phylogenetic trees were constructed separately for each domain with IqTree (Supplemental Data Sets 2 and 3), and a tanglegram (Scornavacca et al., 2011) was drawn that connects the corresponding taxa in the two trees to help explore the differences in phylogenetic topology between the two domains of the LRXI proteins (Figure 2, Supplemental Figure 1). The tree files are available as Supplemental Data Sets 4 and 5.

In addition, nucleotide sequences of *LRXI* genes were aligned using Pal2Nal (Suyam et al., 2006), and the pairwise ratios of nonsynonymous to synonymous substitutions (Ka/Ks, also known as ω or dN/dS ratio) were calculated with the maximum likelihood method using PAML (Yang, 1997) for the two domains separately. These data were examined to investigate whether

11

the selective pressure differs between the two domains of different functions: the ectodomain in perceiving peptide signals and the KD determining the subsequent intracellular response.

## Seven clades of LRXIs in land plants

The phylogenetic tanglegram consists of seven major monophyletic clades (Figure 2A-G, colored boxes)*,* all with high support (Shimodaira-Hasegawa approximate likelihood ratio test (SH-alrt) > 85 (Shimodaira and Hasegawa, 1999); see Supplemental Figure 1 for all support values). However, the branching pattern differed between the KD phylogeny and the LRR phylogeny with regards to the putative orthologues of PEP RECEPTOR (PEPR), which in Arabidopsis play a role in innate immunity, and GASSHO (GSO), involved in the regulation of stem cell identity, control of seedling root growth, and establishment of the Casparian strip, which works as a diffusion barrier in the root vasculature (Tsuwamoto et al., 2008; Doblas et al., 2017; Nakayama et al., 2017 {Racolta, 2014 #5882) (Figure 2A-B). In the KD phylogeny, a paraphyletic group of AMPS other than *M. polymorpha* branches in between GSO and PEPR, while in the LRR phylogeny these AMPS form a monophyletic clade with PEPR. Except for one *M. polymorpha* orthologue of GSO, the phylogenetic trees could not unambiguously resolve the relationship between AMPS and angiosperm homologues as the support values for the deep nodes were low in both phylogenies, which reflects a common problem of inferring ancient relationships of *LRR-RLK* genes (Liu et al., 2016).

The remaining five LRXI clades (Figure 2C-G) are all monophyletic in both the KD and LRR phylogenies, with high support (SH-aLRT support higher or equal to 99) and include sequences from both vascular plants and bryophytes, indicating that these clades existed in the most recent common ancestor of vascular plants and bryophytes. One of these clades (Figure 2C) encompass receptors encoded by three Arabidopsis genes (*At2g33170*, *At5g63930*, and *At1g17230*) of common ancestry with *S. moellendorffii* and *P. patens* orthologues, which all three are without known ligand(s) and function(s). Multivalent mutants may be needed to obtain lines with informative mutant phenotypes.

The neighboring RGI clade includes sequences from *M. polymorpha* and *S. moellendorffii*. Arabidopsis RGIs are in particular involved in developmental processes in roots in Arabidopsis (Fernandez et al., 2013; Shinohara et al., 2016). Roots have been suggested to have evolved independently in several lineages (Fujinami et al., 2020), and it would be interesting to explore whether RGFs and RGIs have laid a fundament for root development. Non-AMPS sequences

12

form three subgroups each of which consists of orthologues of Arabidopsis RGI5, RGI3 and 4, and RGI1 and 2 (Figure 2D, Table 2). In contrast, the monophyletic group encompassing the angiosperm HSL-related receptors (Figure 2E, Table 2) have orthologues from all AMPS.

Figure 2. Tanglegram of the kinase domains (KDs) and the LRR ectodomains encoded by *LRXI* genes.
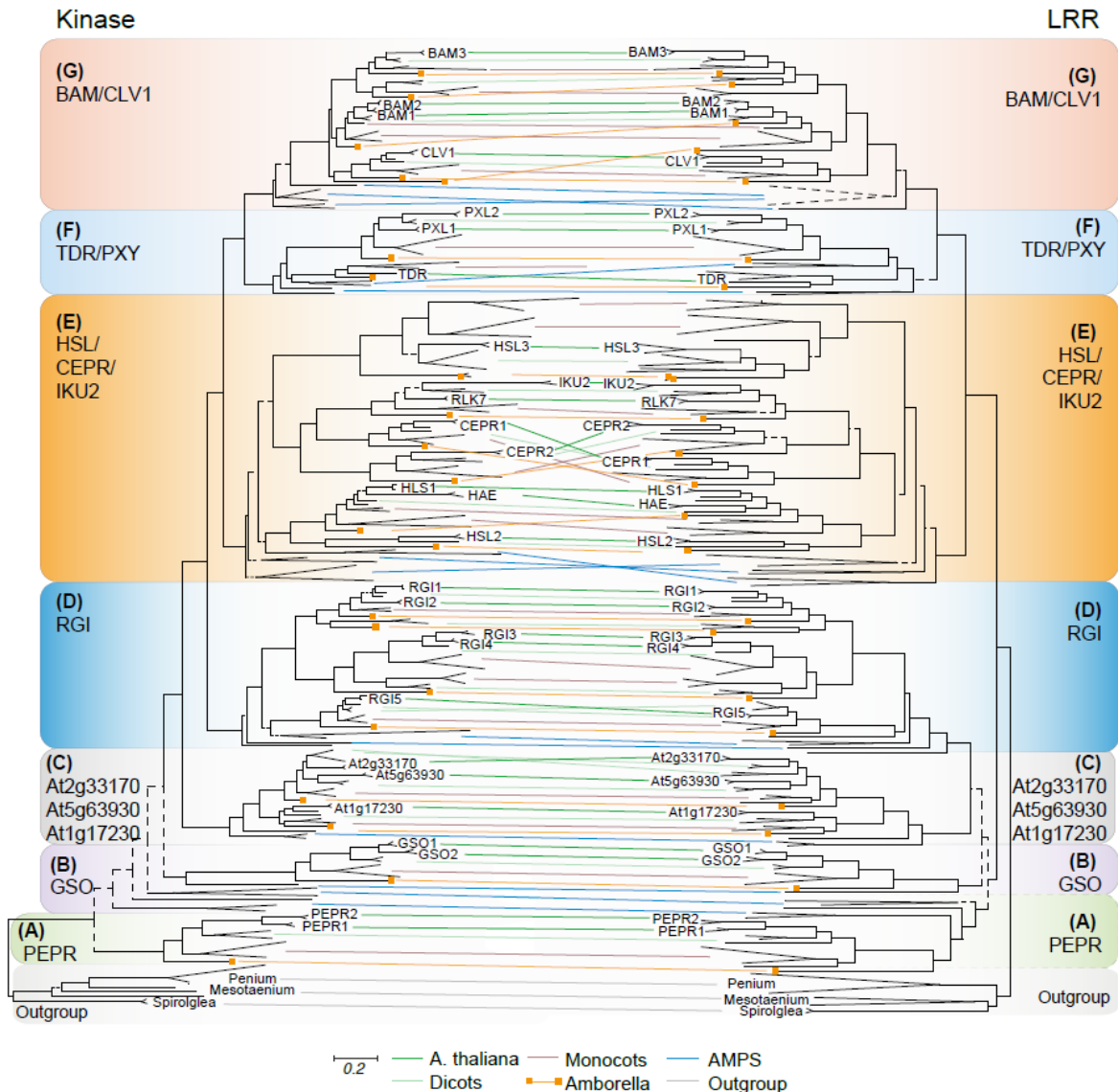
(A)-(G) The seven colored boxes represent monophyletic clades of the indicated angiosperm LRXIs with a common origin and a last common ancestor with one or more AMPS (see Table 2 and Supplemental Figure 1 for details). The phylogenetic trees were constructed separately for the LRRs and the KDs. The braches representing Arabidopsis KDs or LRRs are labeled at the tips. A tanglegram connects KD and LRR originating from the same LRXI. The connecting lines represents *A. thaliana* (dark green); three dicot species (light green); three monocot species (brown); the basal angiosperm *A. trichopoda* (orange with boxes); the hornwort *Anthoceros,* the liverwort *M. polymorpha*, the moss *P. patens*, and the lycophyte *S. moellendorffii* (blue, collectively referred to as AMPS), and three species of Zygnematophyceae algae used as the outgroup (grey). Connecting lines between monophyletic groups of the same kind (i.e. either monocots, dicots, *A. trichopoda*, AMPS, or the outgroup) have been collapsed for clarity. For data on each species see Supplemental Figure 1. Branches drawn with solid lines are well supported (SH-aLRT > 0.85), while dashed lines represent unsupported branches with SH-aLRT < 0.85. Support values for all branches can be found in Supplemental Figure 1. The scale bar for branch lengths represents the average number of changes in amino acids per site, and the two trees of the tanglegram are drawn to scale.

13

The HSL clade has two major branches. One includes HAESA (HAE) and HSL2, which control cell separation events in Arabidopsis with the IDA peptide ligand (Shi et al., 2019). Consistent with a phylogenetic investigation of HSL receptors in angiosperms (Stø et al., 2015) HAE appeared during the evolution of dicot flowering plants. The other branch includes the Arabidopsis At5g25930 (also called HSL3) which very recently was shown to be involved in regulation of stomata closure by modulation of H_2O_2 levels in guard cells (Liu et al., 2020). This receptor and its orthologues are worth an investigation, as an ancestral *HSL3* gene has proliferated in several species, especially in monocots, both before and after the emergence of flowering plants (Figure 2E, Supplemental Figure 1). Branching off from HSL3, *A. trichopoda* orthologues of C-TERMINALLY ENCODED PEPTIDE (CEP) RECEPTORs (CEPRs), and RECEPTOR-LIKE KINASE7 (RLK7) are found (Figure 2E, Supplementary Figure 1). And finally, HAIKU2 (IKU2) branched off from RLK7 during eudicot evolution. CEPR mediates nitrogen starvation signaling (Tabata et al., 2014). RLK7 is involved in innate immunity and lateral root formation (Hou et al., 2014; Toyokura et al., 2019). *IKU2* of the most recent origin controls seed size in Arabidopsis (Luo et al., 2005) and may have been recruited to regulate angiosperm- or eudicot-specific characters (Friedman and Williams, 2004). So far the ligand of HSL3 and IKU2 is not known.

Interestingly, the two CEPR receptors have swapped places in the two trees; RLK7 and IKU2 branches off from CEPR1 in the KD tree but from CEPR2 in the LRR tree (Figure 2E, Supplemental Figure 1). Close inspection of the alignment of these four receptors indicates that CEPR1 has lost a repeat compared to the other three (Supplemental Figure 2). Interestingly, in the related HAE receptor, this repeat is involved in the interaction with the SERK1 co-receptor (Santiago et al., 2016). In the KD, RLK7 and IKU2 are more similar to CEPR1, and the CEPR2 is deviating, for instance with a truncated kinase motif M3 (Supplemental Figure 2). Thus, RLK7 and IKU2 may be more like CEPR2 with regards to perceiving signals, but more similar to CEPR1 with regards to output from the KD.

TDR/PXY and BAM/CLV receptors form clades distinct from the HSL. As mentioned above, these receptors interact with CLE ligands and are mainly involved in the regulation of stem cell activity. Consistent with previous findings, no TDR/PXY orthologues were found in *P. patens* (Figure 2F, Table 2), but other species tested have at least one sequence in the TDR/PXY clade. In its sister group, BAM/CLV1, we identified sequences from all AMPS (Figure 2G, Table 2) (Bowman et al., 2017; Liu et al., 2017; Li et al., 2020; Yang et al., 2020;

Zhang et al., 2020a). The branch point of TDR/PXY and BAM/CLV1 clades is highly supported; these clades likely originated from gene duplication in a common ancestral gene before the divergence of vascular plants and bryophytes. Subsequently, TDR/PXY was lost in the lineage that includes *P. patens*. In the TDR/PXY branch, a gene duplication before the appearance of the angiosperms gave rise to a *PXY-LIKE* (*PXL*) gene, which was further duplicated into *PXL1* and *PXL2* in the eudicot lineage (Figure 2F). The BAM/CLV branch expanded through gene duplications, resulting in three *BAM* genes in addition to the *CLV1* gene in Arabidopsis (Figure 2G).

Our analyses indicate that the 27 *LRXI* Arabidopsis genes originate from six or seven genes (orthologues of *GSO* and/or *PEPR*, *At1g17230*, *RGI*, *HSL*, *TDR*, *BAM*) present in a common ancestor of vascular plants and bryophytes that have undergone duplications (Table 2).

Table 2. Presence of orthologues of Arabidopsis LRR-RLKs subfamily XI in *A. trichopoda* and AMPS.

Paralogues in Angiosperms[1]		Number of paralogues in AMPS			
		Lycophyte	Hornwort	Moss	Liverwort
Arabidopsis thaliana	*Amborella trichopoda*	*Selaginella moellendorffii*	*Anthoceros agrestis*	*Physcomitrium patens*	*Marchantia polymorpha*
PEPR1	*PEPR*	2[3]	2[3]	1[2]	0
PEPR2					
GSO1	*GSO*[2]				1
GSO2					
At1g17230	*At1g17230*	1	0	2	0
At2g33170	*At2g33170*[2]				
At5g63930					
RGI5	*RGI5*	3	0	0	2
RGI4	*RGI4*[2]				
RGI3					
RGI2	*RGI2*[2]				
RGI1					
HSL2	*HSL2*	3	3	2	1
HSL1	*HSL1*[2]				
HAE					
HSL3	*HSL3*				
CEPR1	*CEPR1*				
CEPR2	*CEPR2*				
RLK7	*RLK7*[2]				
IKU2					
TDR/PXY	*TDR/PXY*	2	1	0	1
PXL1	*PXL1*[2]				
PXL2					
BAM1	*BAM1*[2]	3	1	2	1
BAM2					
CLV1	*CLV1*				
BAM3	*BAM3*				
Total		14	7	7	6

[1]For gene IDs see Supplemental Data Set 1.
[2]Gene duplications in eudicots have generated additional paralogues in *A. thaliana*.
[3]Unsolved phylogeny.

S. moellendorffii has in total 14 genes with one to three paralogues for each founder gene. However, in the three bryophytes one or two of the founder genes have been lost, while in *A. trichopoda* orthologues of 18 of the Arabidopsis *LRXI* genes are present. Gene duplications followed by neofunctionalization with regards to ligand interaction may have refined cell-to-cell communication and facilitated the (r)evolution of flowering plants. Furthermore, nine of the present-day Arabidopsis *LRXIs* are likely results of genome duplication during the evolution of the eudicot species (Table 2) (Vanneste et al., 2014).

LRRs and KDs evolve at different pace

Mutational changes in *LRXI* receptor genes may affect the output from the KD or the perception of peptide ligands through the ectodomain. The two parts of these receptors may therefore experience different selective pressure. We noted that for most of the clades in the tanglegram, the branches were longer in the LRR ectodomain tree compared to the KD tree (Figure 2). The branch length is proportional to the number of amino acid changes per site; therefore, this difference suggested that the LRR evolved faster than the KD. To further investigate the rate of evolution in *LRXI* genes of the 21 angiosperm subclades as well as six AMPS subclades, we computed the pairwise ratio (ω) between non-synonymous substitutions at non-synonymous sites (K_a) and synonymous substitutions at synonymous sites (K_s) for the DNA sequences encoding the KDs and the ectodomains (Figure 3). Comparison of the ratio between non-synonymous and synonymous substitutions is an effective method for detection of positive or purifying selection. In general, a ω of 1 indicates neutral evolution without any selective pressure, while $\omega > 1$ indicates positive selection and $\omega < 1$ purifying selection (Yang and Bielawski, 2000).

A recent survey of the evolution of *LRR-RLK* genes in angiosperm did not find the *LRXI* genes to be under significant positive selective pressure compared to other *LRR-RLK* genes (Fischer et al., 2016). We instead wanted to investigate if KDs of *LRXI* genes are under different selective pressure relative to the LRR domains. The ω ratios were therefore calculated separately for the two domains, and the Wilcoxon rank-sum test, which calculate the difference between sets of pairs to establish if these differences are statistically significantly different from one another, was used to calculate p-values. While the $\omega < 1$ indicated purifying selection for both domains, the median ratio for all the angiosperm *LRXI* genes was significantly lower for *KD* (ω_{KD}=0.015) than for the *LRR* (ω_{LRR}=0.024, p-value $< 2.2*10^{-16}$, Figure 3), supporting the idea that the KD is under a stronger purifying selective pressure than the LRR ectodomain.
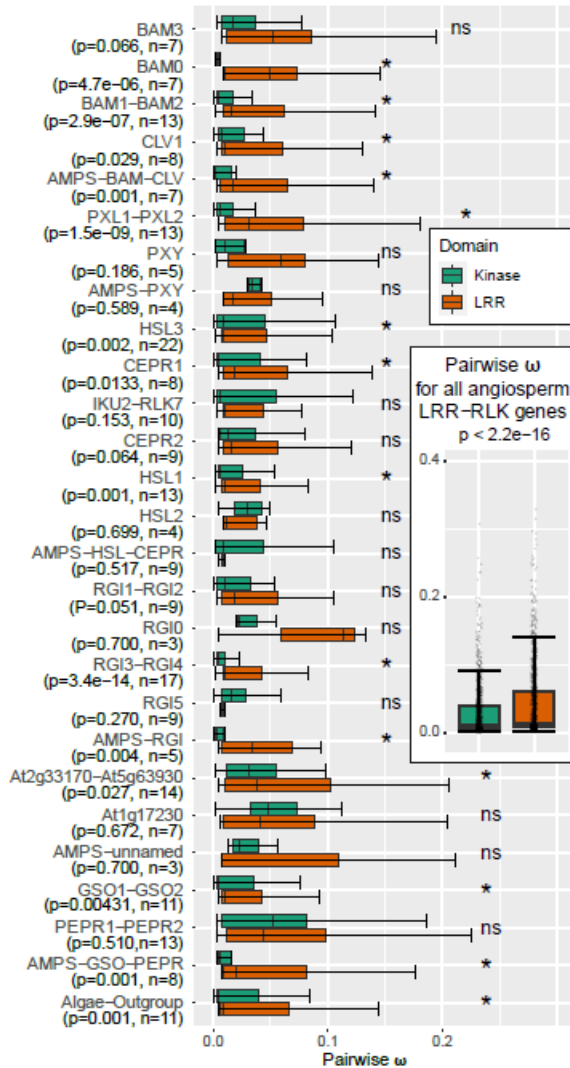
16

Figure 3. Ratio of nonsynonymous and synonymous substitutions in the angiosperm and AMPS LRXI gene sequences encoding the kinase domains (KDs) and LRR ectodomains.
The seven clades were divided into subclades defined from the tanglegram in Figure 2 and detailed in Supplemental Figure 2, and the pairwise ratio (ω) between non-synonymous substitutions at non-synonymous sites (Ka) and synonymous substitutions at synonymous sites (Ks) was calculated for each subclade. The boxplots delineate the mean and the interquartile range of ω. The error bars extend from a value 1.5 times smaller to a value 1.5 times larger than the interquartile range. p-values are based on a Wilcoxon test, n - number of sequences tested in the subclade. * - subclades with significant difference between ω_{LRR} and ω_{KD} (p < 0.05), "ns" – not significant (p> 0.05).
Insert: ω ratio between all pairs of sequences, both within and between subclades. Each pairwise comparison is marked with a point. Outliers have been removed from all boxplots for clarity.

This was a common trend for the angiosperm *LRXI* genes, as ω_{KD} of 15 of the 21 subclades was significantly lower than the ω_{LRR} according to the Wilcoxon test (significance level p<0.05), implying that the *LRRs* have evolved faster relative to the KDs.

Interestingly, the angiosperm subclades with the most significant differences between ω_{KD} and ω_{LRR} were enriched in duplicated monocot genes, and showed particularly low ω_{KD} values (Figure 3 and Table 3). Another interesting finding was the exceptionally low ω_{KD} for the BAM subclades (except for BAM3), where the receptors from bryophytes to Arabidopsis are known to be involved in control of meristem activity. The only subclade for which $\omega_{KD} > \omega_{LRR}$ (p-value 0.002), was HSL3, while for the rest of the clades the difference was insignificant. The HSL3 subclade contains the highest number of sequences with several duplications in both monocots and dicots (Figure 2, and Supplemental Figure 1), suggesting that with the expanded number of genes the selective pressure is relaxed on the KD relative to the LRR. Among the

17

AMPS subclades significant $\omega_{KD} < \omega_{LRR}$ was found for the sister subclades to *RGI* (p-value 0.004, AMPS-RGI in Figure 3), and to GSO-PEPR (p-value 0.001), while there were no significant differences for the remaining subclades.

Table 3. Median values of pairwise K_a/K_s (ω) for subclades of *LRXI* genes with p-values from Wilcoxcon rank-sum test.

Subclade[1]	ω_{KD}	ω_{LRR}	p-value[2]
AMPS-GSO-PEPR	0.0066	0.0203	**0.001**
PEPR1-PEPR2	0.0531	0.0436	0.510
GSO1-GSO2	0.0056	0.0112	**0.004**
AMPS-unnamed	0.0230	0.0081	0.700
At1g17230	0.0490	0.0409	0.672
At2g33170-At563930	0.0323	0.0386	**0.027**
AMPS-RGI	0.0033	0.0346	**0.004**
RGI5	0.0166	0.0096	0.270
RGI3-RGI4	0.0055	0.0111	**3.4×10^{-14}**
RGI0	0.0231	0.1142	0.700
RGI1-RGI2	0.0106	0.0192	0.051
AMPS-HSL-CEPR	0.0090	0.0093	0.517
HSL2	0.0298	0.0119	0.699
HSL1	0.0060	0.0109	**0.001**
CEPR2	0.0135	0.0171	0.064
IKU2-RKL7	0.0070	0.0114	0.153
CEPR1	0.0054	0.0187	**0.013**
HSL3	0.0100	0.0099	*0.002*
AMPS-PXY	0.0351	0.0173	0.589
PXY	0.0207	0.0245	0.186
PXL1-PXL2	0.0068	0.0313	**1.5×10^{-09}**
AMPS-BAM-CLV	0.0029	0.0174	**0.001**
CLV1	0.0083	0.0110	**0.029**
BAM1-BAM2	0.0053	0.0169	**2.9×10^{-07}**
BAM0	0.0047	0.0492	**4.7×10^{-05}**
BAM3	0.0176	0.0523	0.066

[1]See Supplemental Figure 1 for subclade classifications.

[2]Significantly lower ω_{KD} than ω_{LRR} in bold (p < 0.05), significantly higher ω_{KD} than ω_{LRR} in italics (p < 0.05).

The observed difference in selection pressure is likely due to the differential structural and functional requirements facing the respective domains. The stronger conservation of the amino acids in the KD may reflect a need for preservation of an overall configuration to secure the kinase activity by maintaining a compact structure of alpha-helices and beta-sheets based on 13 tightly positioned conserved motifs (Liu et al., 2017). The ectodomain on the other hand needs a structural scaffold guaranteed by the repeated leucines, in addition to amino acid residues involved in peptide recognition. The necessity of molecular interaction between signaling peptides and their receptors must have set constraints on mutational changes of amino

acids with properties and locations directly needed for ligand binding and receptor activation. However, as the peptide ligands vary in sequence and length, the LRRs must have sequence differences adapted to their specific ligand(s).

Additionally, the length of the whole ectodomain must fit the length of the peptide ligand, given that PTMPs interact along the inner surface of the ectodomain (Figure 4A-B). PEPRs and GSOs have for instance 26 and 31 LRRs, respectively, while most LRXI have 21-23 repeats. Correspondingly, the peptides interacting with PEPR and GSO are longer (17 and 20 amino acids, respectively) than the common 12-14 amino acids of the ligands of other subfamily XI receptors (Tang et al., 2015; Okuda et al., 2020). Thus, in searches for unidentified ligands, it is important to take into account the number of LRRs of the receptor of interest.

Methods for in silico homology mapping and modeling of peptide receptor interactions

Having identified presumptive orthologues of receptors and peptide ligands the challenge is to substantiate that candidates fit together as expected for a signaling module. *In silico* methods can be used to identify potential important residues in the LRRs. The simplest is to generate a visual representation of conserved of amino acids in alignments of putative receptor orthologues or members of peptide families, for instance using WebLogo (https://weblogo.berkeley.edu) (Crooks et al., 2004). Each position in the alignment has a stack of amino acid symbols and the height of the stack indicate the overall conservation at that position, while the relative height of the symbols at a given position indicates the relative frequency of specific residues (Crooks et al., 2004).

It may also be useful to generate a sequence logo of aligned repeats from single LRR ectodomain as that will demonstrate the scaffold amino acids as the most conserved residues in these repeats (see Figure 4C-D). The land plant LRR motif is 24 residues long and has specific amino acids, including six leucines, conserved in nearly fixed positions with the consensus sequence of PxxLxxLxxLxxLxLxxNxLxGxI (P – Pro; L – Leu; N – Asn: G – Gly; I – Ile; x – any amino acid). The LRR units that are most importantly involved in peptide binding are often deviating somewhat from the strict consensus of the scaffold (Figure 4 and Supplemental Figure 3). Conserved non-scaffold amino acids are on the other hand likely to

represent functional sites, including residues interacting with ligands or co-receptors (Helft et al., 2011; Orr and Aalen, 2017).
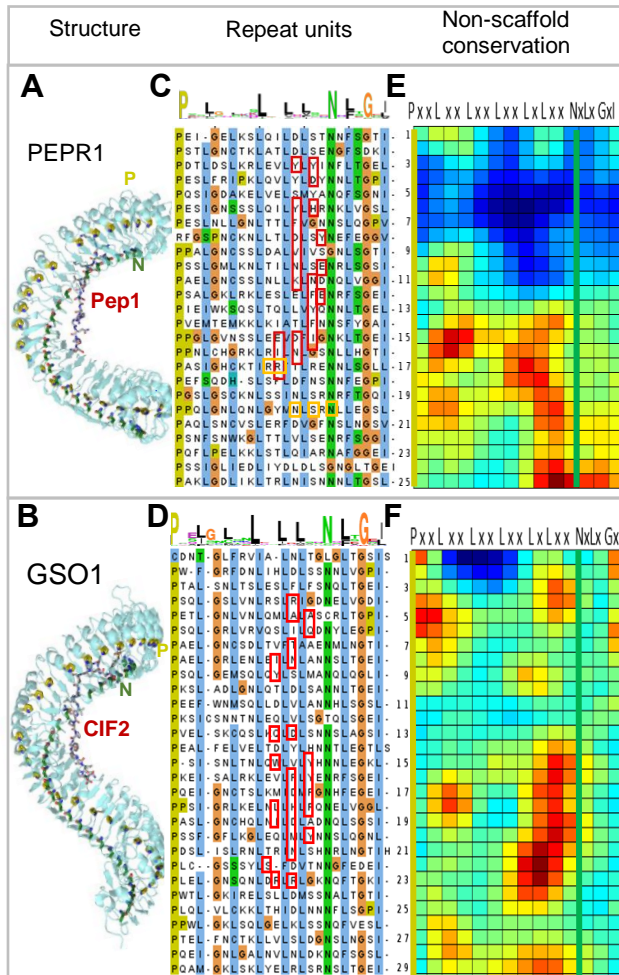

Figure 4. Conservation of LRRs scaffold residues and ligand-interacting amino acids.

(A) and **(B)** Crystal structure of PEPR1 (PDBid:5GR8) with the ligand Pep1 (Tang et al., 2015) and GSO1 (PDBid: 6S6Q) with the ligand CIF2 (Okuda et al., 2020), respectively, interacting along the inner side of the LRR structures. The conserved Pro (P) and Asn (N) residues of the scaffolds are highlighted in mustard and green colors.

(C) and **(D)** Alignment of the 24 amino acids long leucine-rich repeat units of PEPR1 and GSO1 with coloration based on amino acid properties. Above a WebLogo consensus sequence (http://weblogo.berkeley.edu/) visualizing the conservation of the scaffold residues. Residues of the LRR that according to the crystal structures shown in (A) and (D) interact with residues of the respective peptides, are marked with red rectangles, while orange squares represent residues interacting with the co-receptor BAK1/SERK3.

(E) and **(F)** Heat maps generated using Repeat Conservation Mapping (http://www.bentlab.russell.wisc.edu/main/main.php) reflecting the degree of amino acid identify and similarity in a given position (X-axis) and a given repeat (Y-axis) for the non-scaffold residues of PEPR and GSO orthologues, respectively. The most conserved residues are in red and least conserved in blue. The position of the conserved Ps and Ns of the scaffolds are indicated by mustard and green colored vertical lines, respectively. Note that the conserved amino acid residues of PEPR orthologues (LRR 13 and onwards) are not coinciding with the majority of the amino acids of PEPR1 interacting with the Pep1 ligand (LRRs 2-13), but rather with the co-receptor-interacting residues. In contrast, ligand-interacting residues overlap with a substantial number of conserved residues in GSO orthologues (LRRs 14-23).

Repeat Conservation Mapping (RCM) (http://www.bentlab.russell.wisc.edu/main/main.php) is a computational method that removes the scaffold residues from the alignment of orthologous LRRs and thereafter calculates the conservation of the remaining residues, presented as a heat map reflecting the degree of amino acid identify and similarity in a given position in a given repeat (Helft et al., 2011; Orr and Aalen, 2017). RCM can be used to identify conserved

residues in orthologous receptors, which opens for *in vitro* and *in planta* experimental testing of functional importance.

The 3-dimentional (3D) structure of ectodomains bound to ligands are crucial for an understanding of the specificity of ligand interaction. The crystal structure of a number of Arabidopsis receptors and their peptide ligands have been solved over the last few years (Chakraborty et al., 2019) and RCM predictions have been validated by comparison to such crystal structures (Koller and Bent, 2014; Shi et al., 2019).

Solved crystal structures can also be used as templates in comparative modelling. 3D protein models can be generated by extrapolating coordinates of solved crystal structures of Arabidopsis peptide-receptor pairs to evolutionarily related peptides and receptors, using SWISS-MODEL (Arnold et al., 2006). Provided with the amino acid sequence of a target this program can search for suitable templates in the Protein Data Bank (PDB), which is growing rapidly. Lately SWISS-modeling has been extended to protein complexes (Waterhouse et al., 2018) which may be useful, since LRR-RLKs not only interacts with ligands, but also interact with other receptors and with co-receptors (Smakowska-Luzan et al., 2018; Chakraborty et al., 2019). Small co-receptors as key partners in signaling complexes (Ma et al., 2016; Hohmann et al., 2018) is a topic of its own right, which awaits to be elucidated in an evolutionary perspective in land plants.

Conservation of amino acid residues in the ectodomain

When comparing LRR-RLKs one should keep in mind that high similarity observed between LRRs using BLASTp may result primarily from the conserved amino acids making up nearly half of each LRR. Thus, high BLASTp scores may not necessarily imply a conserved mode of ligand binding. One example is PEPR1 which in Arabidopsis is involved in amplification of biotic and abiotic stress responses and interact with endogenous stress-induced Pep peptides (Safaeizadeh and Boller, 2019). The phylogenetic distribution of the Pep peptides is, however, limited to angiosperms, and even within flowering plants Pep peptides show interfamily incompatibility. Peps of heterologous origin cannot be recognized as ligands due to rapid co-evolution of PEPR LRRs with PEPs (Lori et al., 2015). Consistent with this the ω_{LRR} of PEPR1-PEPR2 subclade has a higher value than most of the LRXI subclades (Table 3). The scaffold residues as well as putative co-receptor-interacting residues are conserved, but the majority of the Arabidopsis PEPR1 residues interacting with the peptide ligand Pep (according to the

solved crystal structure (Tang et al., 2015)), are not conserved among PEPR orthologues from the species used in our studies (Figure 4A, C and E).

This situation has likewise been found for an LRR-RLK of subfamily XII involved in defense, the FLAGELLIN-SENSITIVE2 (FLS2) which, like PEPR, is interacting with the co-receptor BAK1/SERK3 (Koller and Bent, 2014). This is in contrast to the conserved ligand-interacting residues of receptors involved in developmental processes, like GSO1 interacting with the confirmed ligand CIF2 (Figure 4B and F) (Okuda et al., 2020) or the *M. polymorpha* and *S. moellendorffii* orthologues of HSL2 interacting with putative IDA peptides of these species (Supplemental Figure 3).

Matching DY^{sulf} peptides with receptors

An additional twist when investigating potential peptide-receptor interactions is the presence of modified amino acid residues, in particular sulfated Tyr and hydroxylated Pro. We have confirmed that GSO and RGI homologues are present in non-flowering plants (Bowman et al., 2017; Liu et al., 2017; Man et al., 2020), and hypothesized that their respective ligands are conserved in the species where presumptive receptors are present. Using PHI-BLASTp, as described, we indeed retrieved ligand candidates from several species (Figure 1 and 5). The candidates show limited sequence identity (Figure 5A), however this is also the case for the four Arabidopsis CIF peptides and the recently identified TWISTED SEED1 (TWS1) peptide which show differential affinities to GSO1 and GSO2 (Barbosa et al., 2019) (Fiume et al., 2016; Doll et al., 2020; Okuda et al., 2020). TWS1 functions with the GSOs to form a functional cuticle around the developing embryo of Arabidopsis.(Tsuwamoto et al., 2008).

It may seem difficult to tell CIF and RGF peptides apart since they both start with DY (where Y is sulfated), they are enriched in Pro residues, and often end with a C-terminal Asn (N). They do, however, differ in length (Figure 5A-B, Table 1). The solved crystal structure of RGF1 and its receptor RGI3 (At4g26540) (Song et al., 2016); (note that the *RGFR/RGI* gene nomenclature differs in the literature) revealed that the DY^{sulf} residues are recognized by the amino acid motif RxGG of the receptor. Importantly, the RGI orthologues from different land plant lineages contain this motif, which distinguishes RGIs from other LRXI members (Supplemental Figure 4).
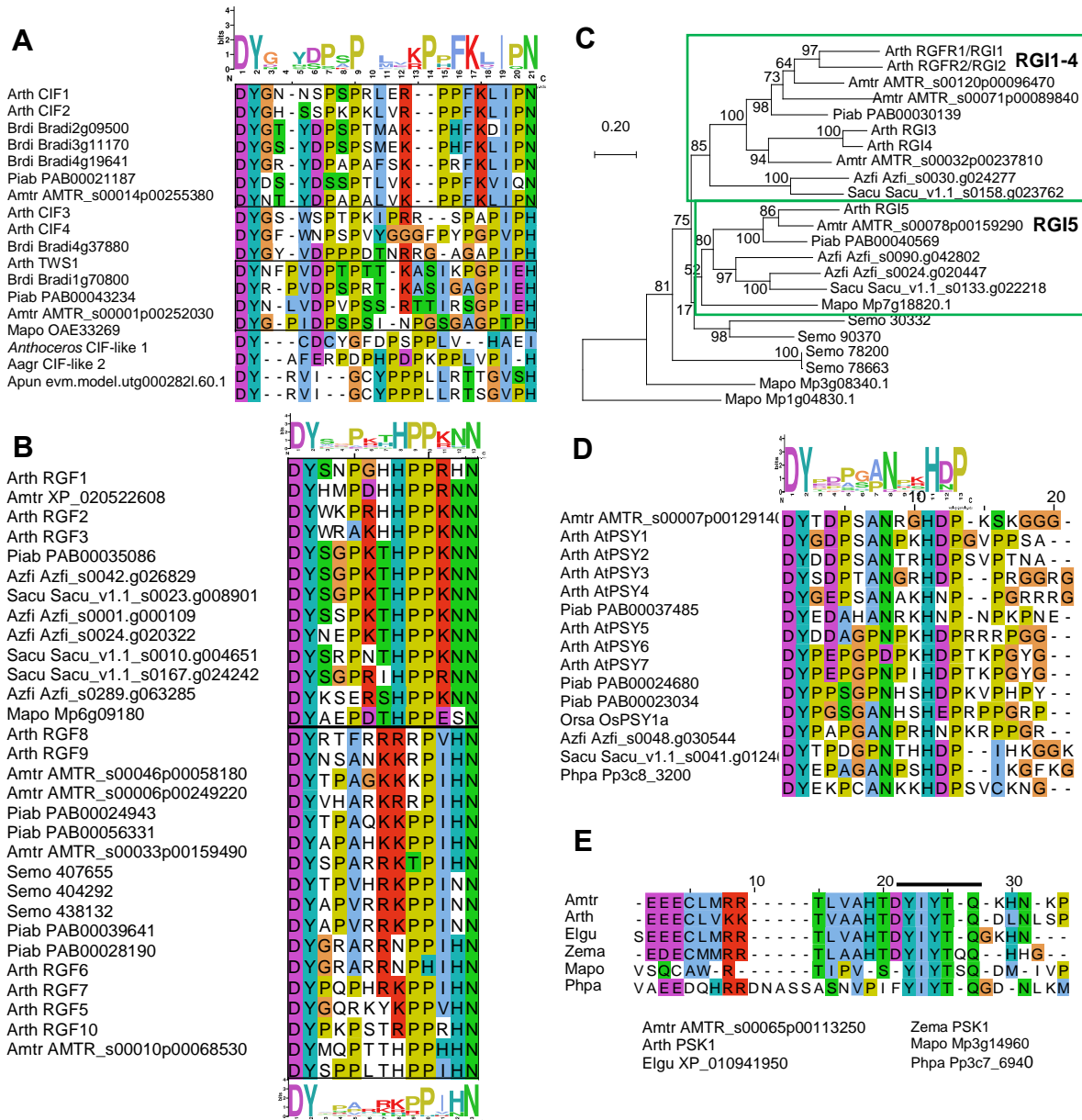
Figure 5. Conservation of sulfated peptides.

(A) Alignment of CIF- and TWS1-related peptides. Above the alignment shown is the sequence logo of the top seven sequences (boxed), which represent canonical CIFs. TWS1 and its putative orthologues are also boxed.

(B) Alignment of the presumptive RGF peptides, with amino acids colored according to chemical properties. RGFs are classified into two distinct groups (boxed), represented by the sequence logos above and below the alignment.

(C) Phylogenetic tree based on the ectodomain amino acid sequences of RGIs in Arabidopsis, and their orthologues in the *A. trichopoda,* the gymnosperm *P. abies*, the ferns *A. filiculoides* and *S. cucullata,* the lycophyte *S. moellendorffii* and the bryophyte *M. polymorpha* constructed by the maximum likelihood method. The GSO homologue of *M. polymorpha* (Mp1g04830.1) from the neighboring clade was used as outgroup. Bootstrap values were obtained from 500 bootstrap replicates and indicated at the nodes as percentages. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

(D) Alignment of the presumptive PSY peptides, with a sequence logo of the first 13 amino acids above the alignment.

(E) Alignment of the PSK1 mature peptide (YIYTQ) and adjacent conserved motif.

Sequence logos were generated using WebLogo 2.8.2 (https://weblogo.berkeley.edu).

23

To further clarify phylogenetic relationships among RGI homologues from different lineages, we bridged the evolutionary gap between Arabidopsis and *S. moellendorffii* by finding genes encoding highly similar receptors in ferns (duckweed, *Azolla filiculoide*, and buce plant, *Salvinia cucullata*) and gymnosperm (Norway spruce, *Picea abies*). In each of these species, two or three *RGI* genes were identified, one or two belonging to a clade that includes Arabidopsis *RGI5* and the other to a branch with three *A. trichopoda* and four Arabidopsis *RGIs* genes (Figure 5C), suggesting one gene duplication in the common ancestor ferns and gymnosperms generated the RGI1-4 and RGI5 clades. In the RGI1-4 clade, at least three gene duplications are inferred: one in the seed plant lineage, generating two subgroups, both of which experienced another duplication in the eudicot lineage. A previous study also indicates that gene copies increased through lineage-specific gene duplications (Man et al., 2020).

In Arabidopsis, Tyr sulfation regulates protein-protein interactions and affects receptor binding (Matsubayashi and Sakagami, 1996; Shinohara et al., 2016; Kaufmann and Sauter, 2019). The gene encoding the responsible enzyme, *TYROSYLPROTEIN SULFOTRANSFERASE* (*AtTPST*), was discovered in Arabidopsis (Komori et al., 2009). TPST-encoding genes are conserved across land plants and found in *S. moellendorffii*, *P. patens*, and *M. polymorpha*, as well in ferns and gymnosperms (Supplemental Figure 5) supporting the likely presence of sulfated signaling peptides and the plausible contribution of PTMs to the ligand diversity in early land plant evolution.

PSY1 and phytosulfokine (PSK) are two sulfated peptides of 18 and five amino acid, respectively. PSY1 has an N-terminal DY^{sulf}, and promotes cellular proliferation and expansion (Amano et al., 2007). The mature PSK peptide is only five amino acids long (YIYIQ) and promotes cell growth, acts in the root apical meristem, contributes to pollen tube guidance, and integrates growth and defense (Sauter, 2015; Kaufmann and Sauter, 2019). We identified putative orthologues both for PSY1 and PSK in *P. patens, M. polymorpha* as well as *P. abies* (Figure 5D-E). *PSK*-like genes and bioactivities have previously been detected in gymnosperms (Igasaki et al., 2003; Wu et al., 2019), but surprisingly, we did not find any orthologue of the Arabidopsis PSY1-receptror in the bryophytes. In contrast to all the subfamily XI receptors discussed here, the two PSK-RECEPTORs (PSKRs) of Arabidopsis belong to the subfamily X, of which binding to small ligands is facilitated by a so-called island domain (ID) in the LRR structure that doesn't fit in with the rest of the repeats. In the PSK-like sequences found in *P. patens* and *M. polymorpha* the Asp (D) residue preceding the YIYIQ

24

sequence is missing, but a conserved N-terminal region can still be recognized (Figure 5E). Such conserved sequences outside the assumed mature peptide might be involved in precursor processing and peptide maturation steps, which have been less characterized in plant peptidic signals.

Matching receptors and hydroxylated ligands

In silico searches for homologues of many known Arabidopsis peptide sequences have identified candidate ligand-receptor pairs across land plants, with the latest discovery of BAM/CLV1 receptor and an H-type CLE peptide in hornworts (Li et al., 2020; Zhang et al., 2020a). This was further complemented by our findings of hornwort TDR/PXY and R-type CLE homologs (Figures 1; Table 2). It is, therefore, feasible that both H-CLE and R-CLE signaling pathways are conserved in hornworts.

Putative orthologues of IDA and HSL2 were also identified in *M. polymorpha* (Bowman et al., 2017) and therefore we looked for IDA in other AMPS. We did not detect IDA orthologues in the moss, but identified potential hits in *S. moellendorffii* and *Anthoceros* hornworts (Figure1 and 6). The crystal structure of synthetic AtIDA peptide bound to its receptor AtHAE has previously been solved, and showed as expected that the ligand is positioned along the inner surface of the LRR ectodomain of the receptor (Figure 6A) (Santiago et al., 2016). Using this structure (PDBid:5ixq), we have recently generated a three-dimensional model of IDA-HSL2 orthologues from oil palm, thereby substantiating a likely interaction in this monocot species (Shi et al., 2019). Here we have used the same modeling strategy for putative peptide ligands and receptors of *M. polymorpha* and *S. moellendorffii*, which demonstrates that the potential overall hydrogen bonding interactions between the modeled MpHSL2 and SmHSL2 receptors and their corresponding IDA peptides (MpIDA1-2 and SmIDA1-3) support receptor-peptide binding, regardless of variations in IDA-peptide sequences (Figure 6B-C). A substantial fraction of the AtHAE amino acids generating hydrogen bonds with IDA residues are identical in MpHSL2 and SmHSL2 (Figure 6A-B and Supplemental Figure 3), and crucial residues, in particular the central Pro and the C-terminal His-Asn, are also conserved in the amino acid sequence of the MpIDA1-2 and SmIDA peptides (Figure 6C). Thus, the generated models are consistent with the possible function as ligand-receptor pairs in early-diverging land plant lineages.
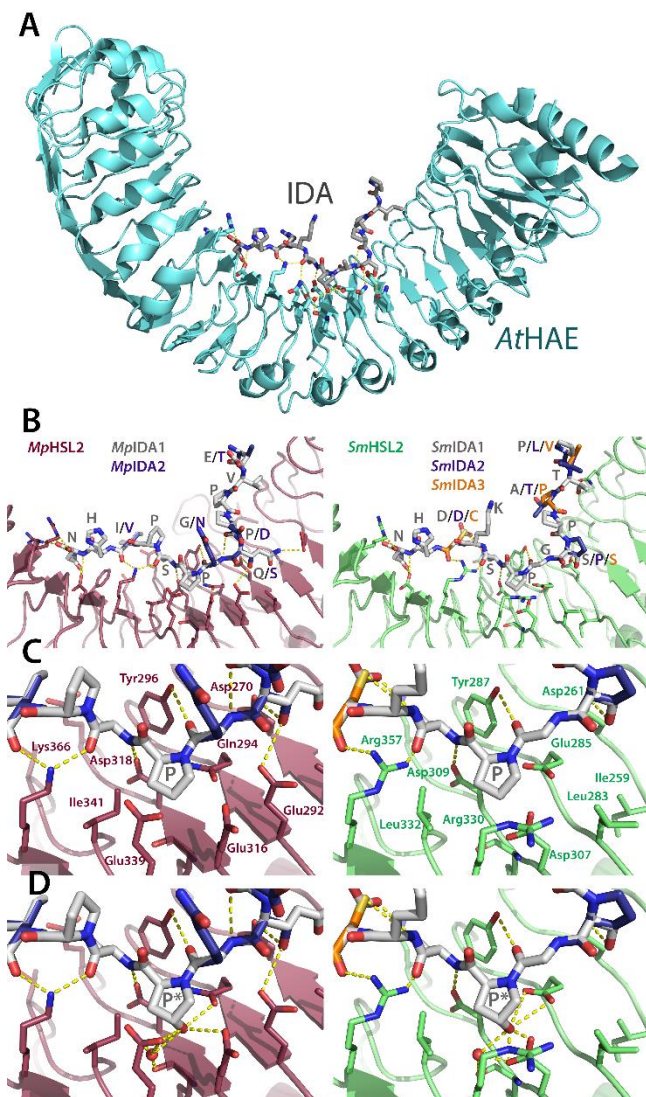
Figure 6. Modelling of the interaction between putative IDA ligands and HSL2 receptors of *M. polymorpha* and *S. moellendorffii.*

(A) The AtHAE-AtIDA crystal structure (PDBid:5ixq) (Santiago et al., 2016) the IDA peptide is lining up along the inner face of the LRR structure. **(B)** Overall models of the interaction between *M. polymorpha* and *S. moellendorffii* putative HSL2 receptors (MpHSL2 in magenta and SmHSL2 in green) with respectively two and three *M. polymorpha* and *S. moellendorffii* superimposed putative IDA peptides (grey backbones) built on the AtHAE-AtIDA structure using SWISS-MODEL (Arnold et al., 2006). Note in particular receptor interaction with the Asn (N) at the C-terminal end of the peptides.

(C) Close-up view of the central parts of the respective receptor models and the surrounding hydrogen bonding network, with a central Pro (P) in the ligands.

(D) Close-up view as in (C), however, with hydroxylation of the central Pro (P*), which facilitates formation of additional hydrogen bonds.

Central amino acids of the receptors, as well as the peptides, are shown as sticks and colored by atom type. Water molecules are shown as red spheres, and modelled based on coordinates from the AtHAE-AtIDA crystal structure. Hydrogen bonds are depicted as dotted lines (yellow). Residues involved in hydrogen bonding to the peptides are depicted with three-letter symbols in colors according to the respective structures. The peptide residues are shown in one-letter symbols. All structure figures were prepared using PyMOL (Schrödinger, LLC).

Synthetic peptides of the IDA/IDL family have been shown to bind and activate the HAE and HSL2 receptors more efficiently when the central Pro is hydroxylated (Table 1) (Butenko et al., 2014; Santiago et al., 2016), and this PTM has also been detected in planta on CLE peptides (Stührwohldt and Schaller, 2019). We therefore modeled MpHSL2 and SmHSL2 with hydroxylated MpIDA and SmIDA peptides. In these models, the hydroxylation postulated an increased number of hydrogen bonds, which would result in a stronger interaction (Figure 6C and D). Interestingly, genes encoding prolyl 4-hydroxylases (P4Hs) which mediate Pro hydroxylation, are evolutionarily conserved from algae to angiosperms (Myllyharju, 2003; Ogawa-Ohnishi et al., 2013). Thus, mature peptides produced from IDA and CLE orthologues may have been hydroxylated already in early land plants (Oelkers et al., 2008; Bowman et al., 2017). Hydroxyproline has also been found in members of other peptide families (Table 1).

26

In Arabidopsis, IDA and its receptors HAE and HSL2 are involved in cell separation processes, like floral organ abscission, lateral root emergence, and root cap sloughing (Aalen et al., 2013; Shi et al., 2018). In various angiosperm species that shed other organs, like fruits or leaves, the IDA-HAE/HSL2 signaling module has also been found expressed at the base of the organ to be shed (Shi et al., 2019), indicating that the molecular process underlying cell separation is highly conserved. This suggests that a change in timing or location of expression of peptide signaling components through changes in their cis-regulatory elements may have served as an important driving force for evolutionary changes in plant architecture and reproductive strategies.

Cell separation processes also play a role in bryophyte development, for example in air chamber formation of *M. polymorpha*. Ishizaki et al. (2015) revealed that E3 ubiquitin ligase regulates cell separation in *Marchantia* air chamber formation, and it is not known whether the IDL-HAE/HSL signaling module is involved or not in this process. It will be important to find out more about biological processes in bryophytes and lycophytes that are controlled by cell separation, where the putative IDA and HSL orthologues might have conserved roles.

Evolution of peptide ligands

In the 20 years since Shiu and Bleecker identified the LRR-RLKs of Arabidopsis, the knowledge about their function and their confirmed ligands have gradually increased. Principally it is expected that genes encoding a receptor and its ligand display the same mutant phenotype. The CLAVATA1 (CLV1) receptor and the CLV3 peptide ligand were discovered already at the end of the last century by their spectacular meristem phenotypes (Clark et al., 1995; Clark et al., 1997). A double mutant of the closely related receptors HAE and HSL2 was needed to disclose their involvement in floral organ abscission, like their peptide ligand IDA (Butenko et al., 2003; Cho et al., 2008; Stenvik et al., 2008). Furthermore, triple to quintuple mutants were necessary to disclose the involvement of the RGI receptors in root development, with RGFs peptides as ligands (Shinohara et al., 2016; Song et al., 2016; Qiu et al., 2020). Thus, functional redundancy on the receptor side as well as the peptide side has made it complicated to identify ligand-receptor pairs through genetics. However, in non-seed plants with often fewer duplicated genes and with the establishment of CRISPR-Cas mutagenesis, a genetic approach for identification of peptide ligand receptor-pair should be more feasible. In order to pursue this, we need to develop more non-seed, genetically tractable model plants (Rensing, 2017).

27

It has just started to be revealed that the CLE peptide signaling has evolutionarily conserved roles from *P. patens* and *M. polymorpha* to angiosperms. One of the known key roles of the CLE signaling in angiosperms is to regulate cell proliferation in the sporophytic ($2n$) meristems, which grow indeterminately. In the bryophytes, the sporophyte shows determinate growth while indeterminate meristems are present in the gametophyte body ($1n$). The CLE signaling in both *P. patens* and *M. polymorpha* is involved in cell proliferation in the gametophyte and determination of the orientation of cell division planes which thereby facilitating the transition from 2D to 3D growth (Whitewoods et al., 2018; Hirakawa et al., 2019). Similarly, Arabidopsis quadruple mutant for *CLV1* and the three *BAM* genes show abnormal cell division planes. These studies support the idea that peptide-receptor modules are conserved and employed in a parallel manner, although the evolutionary relationship between the bryophyte gametophytic meristems and the vascular plant sporophytic meristems is still debated (Bowman, 2013).

To our surprise, several Arabidopsis CLE peptides have been found to interact with LRRs of different clades. The AtCLE9/CLE10 peptides interact with both the HSL1 and BAM1 receptors, depending on tissue type (Qian et al., 2018). AtCLE14 peptide can signal through AtPEPR2 (Gutierrez-Alanis et al., 2017). While PEPs and CLEs differ in length, the C-terminal 12 amino acids of mature PEPs share sequence similarity to CLEs (Table 1), offering a possible explanation for the recognition mechanism. Another CLE peptide, AtCLE45, which differs from other CLE peptides by a number of arginine (Arg, R) residues near its N-terminus, is perceived by canonical CLE receptors as well as an RGF receptor homolog, AtRGI4/SKM2 (Endo et al., 2013; Hazak et al., 2017).

These "unconventional" ligand-receptor pairs raise the question of how the specificity of ligand-receptor pairs evolves. It is of note that genes encoding CLE peptides highly similar to AtCLE9/CLE10 are present in both *M. polymorpha* and *P. patens* (PpCLE5/CLE6) (Whitewoods et al., 2018) (Figure 7). Our phylogenetic investigations suggest that founder receptors, representing the seven clades, by the onset of the angiosperm era had increased about three times, and in each of the clades a new gene has evolved during dicot evolution. We therefore speculate that during early evolution of peptide signaling, there were fewer peptides, and receptors were less specific. Accordingly, large-scale clustering analyses of land plant CLE peptides based on their entire prepropeptide sequences found a smaller number of clusters in AMPS, indicative of the diversification of CLE peptide sequences, which could have resulted in changes in the specificity of ligand-receptor interactions during land plant evolution (Goad

et al., 2017). Hirakawa et al. (2017) found that a synthetic CLE peptide, KIN named after the K (2nd), I (10th), and N (12th) residues crucial for activity, exerts both R-type and H-type CLE activities and interact directly with both the CLV1 and the TDR receptors (Figure 7) (Hirakawa et al., 2017), illustrating the potential for a broader specificity in ligand-receptor interaction hidden in short mature peptide sequences (Figure 7).
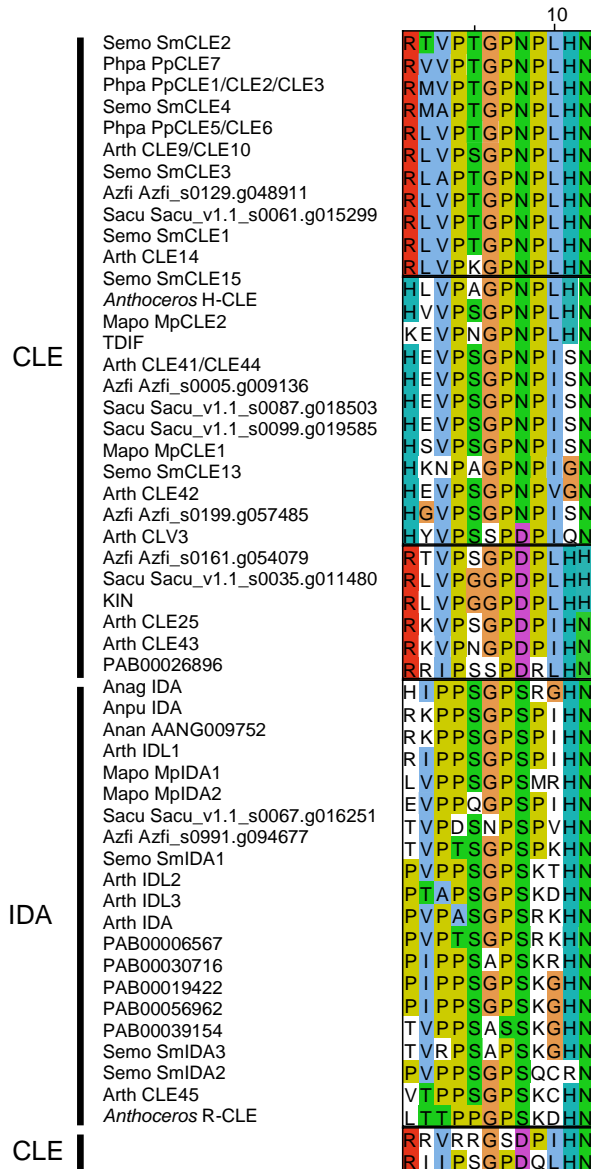

Figure 7. Conservation of CLE and IDA peptides. Alignments of mature CLE and IDA peptides in land plants and a synthetic peptide, KIN. Note shared core residues (PSGP) and C-terminal end (HN) between IDA and some of the CLE peptides, and the presence of almost identical R-CLE, H-CLE and IDA peptides in AMPS, ferns and Arabidopsis. The amino acids are colored according to chemical properties.

We may also learn from some common structural features in known endogenous peptide ligands (reviewed and discussed in (Zhang et al., 2016)). For instance, the IDA and CLE peptides are all 12-14 amino acids long, proline-rich, and need C-terminal HN or HH residues for function (Figure 7; Table 1). Similarities are also recognized among CEPs, PIP-PIPLs, and IDA peptides with respect to the size and amino acid composition. Most of the CEPs and PIP-PIPLs, however, lack the C-terminal HN or HH residues that have been shown to interact with

two closely positioned arginine (Arg, R) residues found in many receptors (Hou et al., 2014; Vie et al., 2015; Song et al., 2016) (Table 1). Instead, CEPs and PIP-PIPLs share the C-terminal GxGH motif (Vie et al., 2015) (Table 1). The CEPs and PIP-PIPLs have been demonstrated to signal through closely-related receptors of the HSL major clade, CEPRs and RLK7, respectively (Figure 2). This points to the possibility that a ligand recognition mechanism is conserved among recently diversified receptors, and in the case of the CEP and PIP-PIPL signaling, the GxGH motif could be crucial for peptide-receptor interactions.

Perspectives

With the increasing availability of genomic data, it is now possible to analyze the molecular inventory of peptide-signaling components in all major lineages of land plants. All known signaling peptides were originally discovered in angiosperms, however, the identification of LRXI in non-seed plants and our successful hunt for their likely peptide ligands, give a new perspective on the roles and evolution of cell-to-cell signaling. A fourth of the present-day XI receptors of Arabidopsis, with seemingly angiosperm-specific functions, have a history all the way back to the ancestor of the bryophytes and vascular plants. In addition to the previously identified orthologues of CLE peptide ligands and their receptors (BAM/CLV and TDR) (Whitewoods et al., 2018; Hirakawa et al., 2019; Hirakawa et al., 2020), our approach has disclosed early occurrence of orthologues of GSO and RGI receptors and their CIF and RGF ligands. Additionally, the direct interaction between HSL2 and IDA orthologues of *M. polymorpha* and *S. moellendorffii*, was substantiated by taking advantage of the solved 3D structure of the Arabidopsis peptide-receptor pair. We have also bridged the gap between AMPS and Arabidopsis by identifying likely orthologues of LRXI in ferns and gymnosperms. This may be a starting point for investigation of peptide signaling in these taxa. This would elucidate LRXI evolution prior to the emergence of angiosperms, and possibly identify common ancestors for receptors present both in *A. trichopoda* and Arabidopsis (Table 2). Nine of the Arabidopsis XI genes are results of gene duplication within the eudicots lineage (Table 2). Several of the receptor families have also proliferated in monocot lineages (Figure 2). This may reflect the increased need for cell-to-cell communication for development of more complex organisms with more complex organs. An approach similar to what we have shown here with the integration of phylogenetic, functional and structural information, should facilitate the disentanglement of distinct evolutionary histories of peptide ligands and their

receptors in mono- and dicots. One study along these lines have recently been published on the BAM/CLV1/TDR receptors and CLE peptides (Cammarata and Scanlon, 2020).

Interestingly, the GSO/RGI clades on one side and the HSL/TDR/BAM/CLV clades on the other, seem to have ligands representing two different groups – those with sulfated Tyr and the CLE/IDA-like peptides. Founder genes for both were possibly present in the most recent common ancestor of vascular plants and bryophytes. Alternatively, convergent evolution cannot be completely dismissed and might account for the highly variable central part of peptide precursor sequences. Signaling peptides can evolve not only from changes in the primary sequences, but through changes in proteolytic processing of prepropeptides, or by PTM enzymes conferring Tyr sulfation or Pro hydroxylation. The PTM processes tend to be irreversible and hence require precise regulation, pointing to the possible contribution of PTMs in diversification of peptide ligands and in defining the specificity of ligand-receptor interaction since the early stages of land plant evolution. Our structure-based modeling of IDA-HSL2 is consistent with an ancient origin of peptide hydroxylation. Structural information can be used to model whether suggested peptide-receptor pairs actually fit together, and structural analyses of bioactive forms of PTMPs in AMPS will be critical for confirming the prediction from the models.

It needs to be mentioned that a number of receptors, even some in Arabidopsis, are orphans, in the sense that their peptide ligands are unknown. Likewise, all PTMP families seem to have orphan members. These may have resulted from gene or genome duplications where some sister genes have accumulated mutations that weakened ligand-receptor interaction. An alternative (and likely) explanation is that more specific genetic, molecular and biochemical factors and conditions for peptide-mediated cell-to-cell communication remain to be discovered. In line with this suggestion, CLE9/10 of Arabidopsis is recruited to bind HSL1 very efficiently in the presence of the small multi-faceted co-receptors SERK1-2 or 3 (K_d = 1.5 μM), while in the absence of SERK the preferred receptor is BAM1 (K_d = 0.112 μM) (Qian et al., 2018). To dig deeper into the evolution of peptide signaling, we recommend the matching of additional phylogenetic, functional and structural data, especially on the interaction with other XI receptors (Smakowska-Luzan et al., 2018), and co-receptors like the SERKs as key partners in receptor complexes in diverse peptide signaling pathways (He et al., 2018; Liang and Zhou, 2018; Hohmann et al., 2017; Ma et al., 2016).

Land plant evolution must successfully have made use of a broadened spectrum of receptors as more complex organ evolved and interspecies interactions increased. However, except for the bryophyte CLE signaling pathways, biological roles have not been assigned for peptide LRXI signaling other than in angiosperms. Peptide signaling pathways have most often been studied in angiosperm-specific organs, of which homologous structures are not universally found outside the flowering plants. A question therefore arises as to whether these peptide signaling modules mainly have undergone neofunctionalization in more recently evolved lineages. This may well be the case from an organismal perspective, however, from a molecular and cellular perspective it may not. Where and when a signaling system triggers a given outcome (e.g. cell division, cuticle formation or cell separation) may actually be the crux of the matter. Thus, a very important next step in tracing the evolutionary origin and history of land plant peptide signaling pathways will be to investigate thoroughly the temporal and spatial expression patterns of LRXI receptors and their putative peptide ligands, and to study downstream molecular events.

Supplemental material

Supplemental Table 1. Major CRP families in land plants. Support for Table 1.

Supplemental Table 2. References for the signaling peptides. Support for Table 1.

Supplemental Figure 1. Phylogeny for subfamily XI LRR-RLKs based on the kinase domain and the LRR ectodomain. Support for Figure 2.

Supplemental Figure 2. Differential sequence similarities in the ectodomains and the kinase domains of CEPR1, CEPR2, RLK7 and IKU2 compared to HAE. Support for Figure 2.

Supplemental Figure 3. Conservation of amino acids residues involved in ligand binding. Support for Figure 4 and 6.

Supplemental Figure 4. Conservation of ligand-binding residues in RGI receptors. Support for Figure 5.

Supplemental Figure 5. Phylogenetic analysis of TYROSYLPROTEIN SULFOTRANSFERASE (TPST). Support for Figure 5.

Supplemental Data Set 1. Names and IDs for proteins and genes presented in Figures 1-6. Excel format.

Supplemental Data Set 2. Alignment of LRRs for phylogenetic analyses.

Supplemental Data Set 3. Alignment of kinase domains for phylogenetic analyses.

Supplemental Data Set 4. Phylogenetic tre file for LRRs.

Supplemental Data Set 5. Phylogenetic tre file for kinase domains.

Acknowledgements

Author Contributions

MW and SS initiated and drafted the work; CF, AKK, RMA and RBA identified and analyzed sequences; MH performed structure analyses and modelling; CF and RBA wrote the article. CF created Figure 1, 5 and 7, and Table 1 with supporting tables and figures with input from RMA and RBA; AKK generated Figures 2 and 3, and Table 2 and 3 with supporting figures with input from RMA and RBA; Figures 4 and 6 with supporting figures were created by MH and RBA.

Distribution of materials

The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) are Chihiro Furumizu (furumizu@kumamoto-u.ac.jp) and Reidunn B. Aalen (reidunn.aalen@ibv.uio.no).

34

References

Aalen, R.B., Wildhagen, M., Stø, I.M. Butenko, M.A. (2013). IDA: A peptide ligand regulating cell separation processes in Arabidopsis. *J. Exp. Bot.* **64:** 5253-5261.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215:** 403-410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucl. Acids Res.* **25**.

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

Amborella Genome, P. (2013). The *Amborella* genome and the evolution of flowering plants. *Science* **342:** 1241089.

Arnold, K., Bordoli, L., Kopp, J. Schwede, T. (2006). The SWISS-model workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* **22:** 195-201.

Banks, J.A., Nishiyama, T., Hasebe, M., et al. (2011). The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* **332:** 960-963.

Barbosa, I.C.R., Rojas-Murcia, N. Geldner, N. (2019). The Casparian strip—one ring to bring cell biology to lignification? *Curr. Opin. Biotech.* **56:** 121-129.

Bennett, T. Scheres, B. (2010). Root development--two meristems for the price of one? In: Marja C.P.T., ed. *Curr. Develop. Biol..* Academic Press, 67-102. (Volume 91.)

Bowman, J.L. (2013). Walkabout on the long branches of plant evolution. *Curr. Opin. Plant Biol.* **16:** 70-77.

Bowman, J.L., Kohchi, T., Yamato, K.T., et al. (2017). Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* **171:** 287-304.e215.

Butenko, M.A., Patterson, S.E., Grini, P.E., Stenvik, G.E., Amundsen, S.S., Mandal, A. Aalen, R.B. (2003). *INFLORESCENCE DEFICIENT IN ABSCISSION* controls floral organ abscission in *Arabidopsis* and identifies a novel family of putative ligands in plants. *Plant Cell* **15:** 2296-2307.

Butenko, M.A., Wildhagen, M., Albert, M., Jehle, A., Kalbacher, H., Aalen, R.B. Felix, G. (2014). Tools and strategies to match peptide-ligand receptor pairs. *Plant Cell* **26:** 1838-1847.

Cammarata, J. Scanlon, M.J. (2020). A functionally informed evolutionary framework for the study of LRR-RLKs during stem cell maintenance. *J. Plant Res.* **133:** 331-342.

Capella-Gutiérrez, S., Silla-Martínez, J.M. Gabaldón, T. (2009). TRIMAL:A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25:** 1972–1973.

Chakraborty, S., Nguyen, B., Wasti, S.D. Xu, G. (2019). Plant leucine-rich repeat receptor kinase (LRR-RK): Structure, ligand perception, and activation mechanism. *Molecules* **24**.

Chen, T., Wang, B., Wang, F., Niu, G., Zhang, S., Li, J. Hong, Z. (2020). The evolutionarily conserved serine residues in bri1 LRR motifs are critical for protein secretion. *Fron.t Plant Sci.* **11:** 32.

Cheng, S., Xian, W., Fu, Y., et al. (2019). Genomes of subaerial zygnematophyceae provide insights into land plant evolution. *Cell* **179:** 1057-1067.e1014.

Cho, S.K., Larue, C.T., Chevalier, D., Wang, H., Jinn, T.L., Zhang, S. Walker, J.C. (2008). Regulation of floral organ abscission in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. U S A* **105:** 15629-15634.

Clark, S.E., Running, M.P. Meyerowitz, E.M. (1995). *CLAVATA3* is a specific regulator of shoot and floral meristem development affecting the same processes as clavata1. *Development* **121:** 2057-2067.

Clark, S.E., Williams, R.W. Meyerowitz, E.M. (1997). The *CLAVATA1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in Arabidopsis. *Cell* **89:** 575-585.

Crooks, G.E., Hon, G., Chandonia, J.M. Brenner, S.E. (2004). WEBLOGO: A sequence logo generator. *Genome Res.,* **:** 1188-1190.

De Vries, J., Curtis, B.A., Gould, S.B. Archibald, J.M. (2018). Embryophyte stress signaling evolved in the algal progenitors of land plants. *Proc. Natl Acad. Sci.U S A* **115:** E3471-E3480.

Delay, C., Imin, N. Djordjevic, M.A. (2013). Cep genes regulate root and shoot development in response to environmental cues and are specific to seed plants. *J Exp Bot* **64:** 5383-5394.

Delwiche, C.F. Cooper, E.D. (2015). The evolutionary origin of a terrestrial flora. *Curr. Biol.* **25:** R899-910.

Deyoung, B.J., Bickle, K.L., Schrage, K.J., Muskett, P., Patel, K. Clark, S.E. (2006). The CLAVATA1-related BAM1, BAM2 and BAM3 receptor kinase-like proteins are required for meristem function in Arabidopsis. *Plant J.* **45:** 1-16.

Doblas, V.G., Smakowska-Luzan, E., Fujita, S., Alassimone, J., Barberon, M., Madalinski, M., Belkhadir, Y. Geldner, N. (2017). Root diffusion barrier control by a vasculature-derived peptide binding to the SGN3 receptor. *Science* **355:** 280-284.

Doll, N.M., Royek, S., Fujita, S., Okuda, S., Chamot, S., Stintzi, A., Widiez, T., Hothorn, M., Schaller, A., Geldner, N. Ingram, G. (2020). A two-way molecular dialogue between embryo and endosperm is required for seed development. *Science* **367:** 431-435.

Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucl. acids res.* **32:** 1792-1797.

Endo, S., Shinohara, H., Matsubayashi, Y. Fukuda, H. (2013). A novel pollen-pistil interaction conferring high-temperature tolerance during reproduction via CLE45 signaling. *Curr. Biol.* **23:** 1670-1676.

Fernandez, A., Drozdzecki, A., Hoogewijs, K., Nguyen, A., Beeckman, T., Madder, A. Hilson, P. (2013). Transcriptional and functional classification of the GOLVEN/ROOT GROWTH FACTOR/CLE-LIKE signaling peptides reveals their role in lateral root and hair formation. *Plant physiol.* **161:** 954-970.

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

Fiume, E., Guyon, V., Remoué, C., Magnani, E., Miquel, M., Grain, D. Lepiniec, L. (2016). TWS1, a novel small protein, regulates various aspects of seed and plant development. *Plant physiol.* **172:** 1732.

Frangedakis, E., Shimamura, M., Villarreal, J.C., Li, F.W., Tomaselli, M., Waller, M., Sakakibara, K., Renzaglia, K.S. Szövényi, P. (2020). The hornworts: Morphology, evolution and development. *New Phytol.* Aug 13. doi: 10.1111/nph.16874. Online ahead of print.

Friedman, W.E. Williams, J.H. (2004). Developmental evolution of the sexual process in ancient flowering plant lineages. *Plant Cell* **16 Suppl:** S119-132.

Fujinami, R., Yamada, T. Imaichi, R. (2020). Root apical meristem diversity and the origin of roots: Insights from extant lycophytes. *J. Plant Res.* **133:** 291-296.

Gensel, P.G. (2008). The earliest land plants. *Annu. Rev. Ecol. Evo. System.* **39:** 459-477.

Ghorbani, S., Lin, Y.C., Parizot, B., Fernandez, A., Njo, M.F., Van De Peer, Y., Beeckman, T. Hilson, P. (2015). Expanding the repertoire of secretory peptides controlling root development with comparative genome analysis and functional assays. *J. Exp. Bot.* **66:** 5257-5269.

Goad, D.M., Zhu, C. Kellogg, E.A. (2017). Comprehensive identification and clustering of *CLV3*/*ESR*-related (*CLE*) genes in plants finds groups with potentially shared function. *New phytol.* **216:** 605-616.

Gong, Z., Morales-Ruiz, T., Ariza, R.R., Roldan-Arjona, T., David, L. Zhu, J.K. (2002). ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. *Cell* **111:** 803-814.

Grienenberger, E. Fletcher, J.C. (2015). Polypeptide signaling molecules in plant development. *Curr. Opin. Plant Biol.* **23:** 8-14.

Gutierrez-Alanis, D., Yong-Villalobos, L., Jimenez-Sandoval, P., Alatorre-Cobos, F., Oropeza-Aburto, A., Mora-Macias, J., Sanchez-Rodriguez, F., Cruz-Ramirez, A. Herrera-Estrella, L. (2017). Phosphate starvation-dependent iron mobilization induces CLE14 expression to trigger root meristem differentiation through CLV2/PEPR2 signaling. *Dev. Cell* **41:** 555-570 e553.

Hazak, O., Brandt, B., Cattaneo, P., Santiago, J., Rodriguez-Villalon, A., Hothorn, M. Hardtke, C.S. (2017). Perception of root-active CLE peptides requires coryne function in the phloem vasculature. *EMBO Rep.* **18:** 1367-1381.

Hazak, O. Hardtke, C.S. (2016). CLAVATA 1-type receptors in plant development. *J. Exp. Bot.* **67:** 4827-4833.

Helft, L., Reddy, V., Chen, X., Koller, T., Federici, L., Fernandez-Recio, J., Gupta, R. Bent, A. (2011). LRR conservation mapping to predict functional sites within protein leucine-rich repeat domains. *PLoS One* **6:** e21614.

Hetherington, A.J. Dolan, L. (2019). Evolution: Diversification of angiosperm rooting systems in the early cretaceous. *Curr. Biol.* **29:** R1081-r1083.

Hirakawa, Y., Fujimoto, T., Ishida, S., Uchida, N., Sawa, S., Kiyosue, T., Ishizaki, K., Nishihama, R., Kohchi, T. Bowman, J.L. (2020). Induction of multichotomous branching by CLAVATA peptide in *Marchantia polymorpha*. *Curr. Biol*. **30**: 3833-3840.

Hirakawa, Y., Kondo, Y. Fukuda, H. (2010). Regulation of vascular development by cle peptide-receptor systems. *J. Integr. Plant Biol.* **52:** 8-16.

Hirakawa, Y., Torii, K.U. Uchida, N. (2017). Mechanisms and strategies shaping plant peptide hormones. *Plant Cell Physiol.* **58:** 1313-1318.

Hirakawa, Y., Uchida, N., Yamaguchi, Y.L., Tabata, R., Ishida, S., Ishizaki, K., Nishihama, R., Kohchi, T., Sawa, S. Bowman, J.L. (2019). Control of proliferation in the haploid meristem by CLE peptide signaling in *Marchantia polymorpha. PLoS Genet.* **15:** e1007997.

Hohmann, U., Santiago, J., Nicolet, J., Olsson, V., Spiga, F.M., Hothorn, L.A., Butenko, M.A. Hothorn, M. (2018). Mechanistic basis for the activation of plant membrane receptor kinases by SERK-family coreceptors. *Proc. Natl Acad. Sci. U S A* **115:** 3488-3493.

Hosseini, S., Schmidt, E.D.L. Bakker, F.T. (2020). Leucine-rich repeat receptor-like kinase II phylogenetics reveals five main clades throughout the plant kingdom. *Plant J.* **103:** 547-560.

Hou, S., Wang, X., Chen, D., Yang, X., Wang, M., Turra, D., Di Pietro, A. Zhang, W. (2014). The secreted peptide PIP1 amplifies immunity through receptor-like kinase 7. *PLoS pathogens* **10:** e1004331.

Igasaki, T., Akashi, N., Ujino-Ihara, T., Matsubayashi, Y., Sakagami, Y. Shinohara, K. (2003). Phytosulfokine stimulates somatic embryogenesis in *Cryptomeria japonica*. *Plant Cell Physiol.* **44:** 1412-1416.

Ishizaki, K. (2017). Evolution of land plants: Insights from molecular studies on basal lineages. *Biosci. Biotechnol. Biochem.* **81:** 73-80.

Jiao, C., Sørensen, I., Sun, X., et al. (2020). The *Penium margaritaceum* genome: Hallmarks of the origins of land plants. *Cell* **181:** 1097-1111.e1012.

Jill Harrison, C. (2017). Development and genetics in the evolution of land plant body plans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372**.

Jones, P. (2014). INTERPROSCAN 5: Genome-scale protein function classification *Bioinformatics* **30:** 1236-1240.

Jun, J., Fiume, E. Fletcher, J. (2008). The CLE family of plant polypeptide signaling molecules. *Cell. Mol. Life Sci. (CMLS)* **65:** 743-755.

Katoh, K. Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30:** 772-780.

Kaufmann, C. Sauter, M. (2019). Sulfated plant peptide hormones. *J. Exp. Bot.* **70:** 4267-4277.

Kenrick, P. Crane, P.R. (1997). The origin and early evolution of plants on land. *Nature* **389:** 33-39.

Koller, T. Bent, A.F. (2014). FLS2-BAK1 extracellular domain interaction sites required for defense signaling activation. *PloS One* **9:** e111185-e111185.

Komori, R., Amano, Y., Ogawa-Ohnishi, M. Matsubayashi, Y. (2009). Identification of TYROSYLPROTEIN SULFOTRANSFERASE in Arabidopsis. *Proc. Natl Acad. Sci. U S A* **106:** 15067-15072.

Lamesch, P., Berardini, T.Z., Li, D., et al. (2012). The Arabidopsis information resource (TAIR): Improved gene annotation and new tools. *Nucl. Acids Res.* **40:** D1202-1210.

Lang, D., Ullrich, K.K., Murat, F., et al. (2018). The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *Plant J*. **93:** 515-533.

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

Li, F.-W., Brouwer, P., Carretero-Paulet, L., et al. (2018). Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat. Plants* **4:** 460-472.

Li, F.W., Nishiyama, T., Waller, M., et al. (2020). *Anthoceros* genomes illuminate the origin of land plants and the unique biology of hornworts. *Nat. Plants* **6:** 259-272.

Ligrone, R., Duckett, J.G. Renzaglia, K.S. (2012). Major transitions in the evolution of early land plants: A bryological perspective. *Annals of botany* **109:** 851-871.

Lin, Y., Wang, K., Li, X., Sun, C., Yin, R., Wang, Y., Wang, Y. Zhang, M. (2018). Evolution, functional differentiation, and co-expression of the RLK gene family revealed in *Jilin ginseng*, *Panax ginseng* c.A. Meyer. *Mol. Genet. Genom.* **293:** 845-859.

Liu, P.L., Du, L., Huang, Y., Gao, S.M. Yu, M. (2017). Origin and diversification of leucine-rich repeat receptor-like protein kinase (LRR-RLK) genes in plants. *BMC Evol. Biol.* **17:** 47.

Liu, P.L., Xie, L.L., Li, P.W., Mao, J.F., Liu, H., Gao, S.M., Shi, P.H. Gong, J.Q. (2016). Duplication and divergence of leucine-rich repeat receptor-like protein kinase (LRR-RLK) genes in basal angiosperm *Amborella trichopoda*. *Front. Plant Sci.* **7:** 1952.

Liu, X.S., Liang, C.C., Hou, S.G., Wang, X., Chen, D.H., Shen, J.L., Zhang, W. Wang, M. (2020). The LRR-RLK protein HSL3 regulates stomatal closure and the drought stress response by modulating hydrogen peroxide homeostasis. *Front. Plant Sci.* **11:** 548034.

Lori, M., Van Verk, M.C., Hander, T., Schatowitz, H., Klauser, D., Flury, P., Gehring, C.A., Boller, T. Bartels, S. (2015). Evolutionary divergence of the PLANT ELICITOR PEPTIDES (Peps) and their receptors: Interfamily incompatibility of perception but compatibility of downstream signalling. *J.Exp. Bot.* **66:** 5315-5325.

Luo, M., Dennis, E.S., Berger, F., Peacock, W.J. Chaudhury, A. (2005). *Miniseed3* (*mini3*), a *WRKY* family gene, and *Haiku2* (iku2), a leucine-rich repeat (LRR) kinase gene, are regulators of seed size in Arabidopsis. *Proc. Natl Acad. Sci. U S A* **102:** 17531-17536.

Ma, X., Xu, G., He, P. Shan, L. (2016). SERKing coreceptors for receptors. *Trends Plant Sci.* **21:** 1017-1033.

Magalhães, D.M., Scholte, L.L., Silva, N.V., Oliveira, G.C., Zipfel, C., Takita, M.A. De Souza, A.A. (2016). LRR-RLK family from two citrus species: Genome-wide identification and evolutionary aspects. *BMC Genomics* **17:** 623.

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

Matsubayashi, Y. (2011). Post-translational modifications in secreted peptide hormones in plants. *Plant Cell Physiol.* **52:** 5-13.

Matsubayashi, Y. (2014). Posttranslationally modified small-peptide signals in plants. *Annu. Rev. Plant Biol.* **65:** 385-413.

Matsubayashi, Y. Sakagami, Y. (1996). Phytosulfokine, sulfated peptides that induce the proliferation of single mesophyll cells of *Asparagus officinalis l. Proc. Natl Acad. Sci. U S A* **93:** 7623-7627.

Matsubayashi, Y. Sakagami, Y. (2006). Peptide hormones in plants. *Annu. Rev.Plant Biol.* **57:** 649-674.

Matsuzaki, Y., Ogawa-Ohnishi, M., Mori, A. Matsubayashi, Y. (2010). Secreted peptide signals required for maintenance of root stem cell niche in Arabidopsis. *Science* **329:** 1065-1067.

Meng, L., Buchanan, B.B., Feldman, L.J. Luan, S. (2012). CLE-LIKE (CLEl) peptides control the pattern of root growth and lateral root development in Arabidopsis *Proc. Natl Acad. Sci. U S A* **109:** 1760-1765.

Morris, J.L., Puttick, M.N., Clark, J.W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Yang, Z., Schneider, H. Donoghue, P.C.J. (2018). The timescale of early land plant evolution. *Proc. Natl Acad. Sci. U S A* **115:** E2274-e2283.

Muschietti, J.P. Wengier, D.L. (2018). How many receptor-like kinases are required to operate a pollen tube. *Curr. Opin. Plant Biol.* **41:** 73-82.

Myllyharju, J. (2003). Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol.* **22:** 15-24.

Nakayama, T., Shinohara, H., Tanaka, M., Baba, K., Ogawa-Ohnishi, M. Matsubayashi, Y. (2017). A peptide hormone required for Casparian strip diffusion barrier formation in Arabidopsis roots. *Science* **355:** 284-286.

Nishiyama, T., Sakayama, H., De Vries, J., et al. (2018). The chara genome: Secondary complexity and implications for plant terrestrialization. *Cell* **174:** 448-464.e424.

Nystedt, B., Street, N.R., Wetterbom, A., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature* **497:** 579-584.

Oelkers, K., Goffard, N., Weiller, G., Gresshoff, P., Mathesius, U. Frickey, T. (2008). Bioinformatic analysis of the CLE signaling peptide family. *BMC Plant Biol.* **8:** 1.

Ogawa-Ohnishi, M., Matsushita, W. Matsubayashi, Y. (2013). Identification of three hydroxyproline o-arabinosyltransferases in *Arabidopsis thaliana*. *Nat. Chem. Biol.* **9:** 726-730.

Oh, E., Seo, P.J. Kim, J. (2018). Signaling peptides and receptors coordinating plant root development. *Trends Plant Sci.* **23:** 337-351.

Okuda, S., Fujita, S., Moretti, A., Hohmann, U., Doblas, V.G., Ma, Y., Pfister, A., Brandt, B., Geldner, N. Hothorn, M. (2020). Molecular mechanism for the recognition of sequence-divergent CIF peptides by the plant receptor kinases GSO1/SGN3 and GSO2. *Proc. Natl Acad. Sci. U S A* **117:** 2693-2703.

Olsson, V., Joos, L., Zhu, S., Gevaert, K., Butenko, M.A. De Smet, I. (2019). Look closely, the beautiful may be small: Precursor-derived peptides in plants. *Annu. Rev. Plant Biol.* **70:** 153-186.

Orr, R.J.S. Aalen, R.B. (2017). *In silico* prediction of ligand-binding sites of plant receptor kinases using conservation mapping. *Methods Mol. Biol.* **1621:** 93-105.

Pearce, G., Strydom, D., Johnson, S. Ryan, C.A. (1991). A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins. *Science* **253:** 895-897.

Plisson, F., Ramírez-Sánchez, O. Martínez-Hernández, C. (2020). Machine learning-guided discovery and design of non-hemolytic peptides. *Sci. Rep.* **10:** 16581.

Qian, P., Song, W., Yokoo, T., Minobe, A., Wang, G., Ishida, T., Sawa, S., Chai, J. Kakimoto, T. (2018). The CLE9/10 secretory peptide regulates stomatal and vascular development through distinct receptors. *Nat. Plants* **4:** 1071-1081.

Qiu, Z.L., Wen, Z., Yang, K., Tian, T., Qiao, G., Hong, Y. Wen, X.P. (2020). Comparative proteomics profiling illuminates the fruitlet abscission mechanism of sweet cherry as induced by embryo abortion. *Int. J. Mol. Sci.* **21**.

Raven, J.A. Edwards, D. (2001). Roots: Evolutionary origins and biogeochemical significance. *J. Exp. Bot.* **52:** 381-401.

Rensing, S.A. (2017). Why we need more non-seed plant models. *New Phytol* **216:** 355-360.

Rensing, S.A., Goffinet, B., Meyberg, R., Wu, S.Z. Bezanilla, M. (2020). The moss *Physcomitrium* (*Physcomitrella*) *patens*: A model organism for non-seed plants. *Plant Cell* **32:** 1361-1376.

Rensing, S.A., Lang, D., Zimmer, A.D., et al. (2008). The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* **319:** 64-69.

Safaeizadeh, M. Boller, T. (2019). Differential and tissue-specific activation pattern of the *ATPROPEP* and *ATPEPR* genes in response to biotic and abiotic stress in *Arabidopsis thaliana*. *Plant Signal. Behav.* **14:** e1590094.

Santiago, J., Brandt, B., Wildhagen, M., Hohmann, U., Hothorn, L.A., Butenko, M.A. Hothorn, M. (2016). Mechanistic insight into a peptide hormone signaling complex mediating floral organ abscission. *eLife* **5:** e15075.

Sauter, M. (2015). Phytosulfokine peptide signalling. *J. Exp. Bot.* **66:** 5161-5169.

Scornavacca, C., Zickmann, F. Huson, D.H. (2011). Tanglegrams for rooted phylogenetic trees and networks. *Bioinformatics* **27:** i248-256.

Segonzac, C. Monaghan, J. (2019). Modulation of plant innate immune signaling by small peptides. *Curr. Opin. Plant Biol.* **51:** 22-28.

Serrano, A. (2020). Deepmspeptide: Peptide detectability prediction using deep learning. *Bioinformatics* **36**.

Shi, C.-L., Von Wangenheim, D., Herrmann, U., et al. (2018). The dynamics of root cap sloughing in Arabidopsis is regulated by peptide signalling. *Nat. Plants* **4:** 596-604.

Shi, C.L., Alling, R.M., Hammerstad, M. Aalen, R.B. (2019). Control of organ abscission and other cell separation processes by evolutionary conserved peptide signaling. *Plants (Basel)* **8**.

Shimodaira, H. Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16:** 1114–1116.

Shinohara, H., Mori, A., Yasue, N., Sumida, K. Matsubayashi, Y. (2016). Identification of three LRR-RKs involved in perception of root meristem growth factor in Arabidopsis. *Proc. Natl Acad. Sci. U S A* **113:** 3897-3902.

Shiu, S.H. Bleecker, A.B. (2001a). Plant receptor-like kinase gene family: Diversity, function, and signaling. *Sci STKE* **2001:** re22.

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

Smakowska-Luzan, E., Mott, G.A., Parys, K., et al. (2018). An extracellular network of Arabidopsis leucine-rich repeat receptor kinases. *Nature* **553:** 342-346.

Song, W., Liu, L., Wang, J., et al. (2016). Signature motif-guided identification of receptors for peptide hormones essential for root meristem growth. *Cell Res* **26:** 674-685.

Steemans, P., Herisse, A.L., Melvin, J., Miller, M.A., Paris, F., Verniers, J. Wellman, C.H. (2009). Origin and radiation of the earliest vascular land plants. *Science* **324:** 353.

Stenvik, G.E., Tandstad, N.M., Guo, Y., Shi, C.L., Kristiansen, W., Holmgren, A., Clark, S.E., Aalen, R.B. Butenko, M.A. (2008). The EPIP peptide of INFLORESCENCE DEFICIENT IN ABSCISSION is sufficient to induce abscission in Arabidopsis through the RECEPTOR-LIKE KINASES HAESA and HAESA-LIKE2. *Plant Cell* **20:** 1805-1817.

Stührwohldt, N. Schaller, A. (2019). Regulation of plant peptide hormones and growth factors by post-translational modification. *Plant biol. (Stuttgart, Germany)* **21 Suppl 1:** 49-63.

Stø, I.M., Orr, R.J., Fooyontphanich, K., Jin, X., Knutsen, J.M., Fischer, U., Tranbarger, T.J., Nordal, I. Aalen, R.B. (2015). Conservation of the abscission signaling peptide IDA during angiosperm evolution: Withstanding genome duplications and gain and loss of the receptors HAE/HSL2. *Front. Plant Sci.* **6:** 931.

Suyam, M., Torrents, D. P., B. (2006). PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucl. Acids Res.***:** W609–W612.

Tabata, R., Sumida, K., Yoshii, T., Ohyama, K., Shinohara, H. Matsubayashi, Y. (2014). Perception of root-derived peptides by shoot LRR-RKs mediates systemic n-demand signaling. *Science* **346:** 343-346.

Taleski, M., Imin, N. Djordjevic, M.A. (2018). CEP peptide hormones: Key players in orchestrating nitrogen-demand signalling, root nodulation, and lateral root development. *J. Exp. Bot.* **69:** 1829-1836.

Tang, J., Han, Z., Sun, Y., Zhang, H., Gong, X. Chai, J. (2015). Structural basis for recognition of an endogenous peptide by the plant receptor kinase PEPR1. *Cell Research* **25:** 110-120.

Tavormina, P., De Coninck, B., Nikonorova, N., De Smet, I. Cammue, B.P.A. (2015). The plant peptidome: An expanding repertoire of structural features and biological functions. *Plant Cell* **27:** 2095-2118.

Toyokura, K., Goh, T., Shinohara, H., et al. (2019). Lateral inhibition by a peptide hormone-receptor cascade during Arabidopsis lateral root founder cell formation. *Dev. Cell.* **48:** 64-75.e65.

Transcriptomes, O.T.P. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574:** 679-685.

Tsuwamoto, R., Fukuoka, H. Takahata, Y. (2008). *GASSHO1* and *GASSHO2* encoding a putative leucine-rich repeat transmembrane-type receptor kinase are essential for the normal development of the epidermal surface in Arabidopsis embryos. *Plant J.* **54:** 30-42.

Vanneste, K., Baele, G., Maere, S. Van De Peer, Y. (2014). Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the cretaceous–paleogene boundary. *Genome Research* **24:** 1334-1347.

Vie, A.K., Najafi, J., Liu, B., Winge, P., Butenko, M.A., Hornslien, K.S., Kumpf, R., Aalen, R.B., Bones, A.M. Brembu, T. (2015). The *IDA/IDA-LIKE* and *PIP/PIP-LIKE* gene families in Arabidopsis: Phylogenetic relationship, expression patterns, and transcriptional effect of the PIPL3 peptide. *J.Exp. Bot.* **66:** 5351-5365.

Waterhouse, A., Bertoni, M., Bienert, S., et al. (2018). Swiss-model: Homology modelling of protein structures and complexes. *Nucl. Acids Res.* **46:** W296-W303.

Wei, Z., Wang, J., Yang, S. Song, Y. (2015). Identification and expression analysis of the LRR-RLK gene family in tomato (*Solanum lycopersicum*) Heinz 1706. *Genome* **58:** 121-134.

Wheeler, J.I. Irving, H.R. (2010). Evolutionary advantages of secreted peptide signalling molecules in plants. *Func .Plant Biol.* **37:** 382-394.

Whitewoods, C.D., Cammarata, J., Nemec Venza, Z., et al. (2018). *CLAVATA* was a genetic novelty for the morphological innovation of 3D growth in land plants. *Curr. Biol.* **28:** 2365-2376.e2365.

Wu, H., Zheng, R., Hao, Z., Meng, Y., Weng, Y., Zhou, X., Zhu, L., Hu, X., Wang, G., Shi, J. Chen, J. (2019). *Cunninghamia lanceolata* PSK peptide hormone genes promote primary root growth and adventitious root formation. *Plants (Basel)* **8**.

Yang, Y., Sun, P., Lv, L., et al. (2020). Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. *Nat. Plants* **6:** 215-222.

Yang, Z. (1997). Paml: A program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci*. **13:** 555-556.

Yang, Z. Bielawski, J.P. (2000). Statistical methods for detecting molecular adaptation.. *Trends Ecol. Evol.***15:** 1994-1997.

Zan, Y., Ji, Y., Zhang, Y., Yang, S., Song, Y. Wang, J. (2013). Genome-wide identification, characterization and expression analysis of *Populus LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE* genes. *BMC Genomics* **14:** 318.

Zhang, H., Han, Z., Song, W. Chai, J. (2016). Structural insight into recognition of plant peptide hormones by receptors. *Mol. Plant* **9:** 1454-1463.

Zhang, J., Fu, X.X., Li, R.Q., et al. (2020a). The hornwort genome and early land plant evolution. *Nat. Plants* **6:** 107-118.

Zhang, Z., Liu, L., Kucukoglu, M., Tian, D., Larkin, R.M., Shi, X. Zheng, B. (2020b). Predicting and clustering plant CLE genes with a new method developed specifically for short amino acid sequences. *BMC Genomics* **21:** 709.

Zhang, Z., Schäffer, A.A., Miller, M., Madden, T.L., Lipman, D.J., Koonin, E.V. Altschul, A.F. (1998). Protein sequence similarity searches using patterns as seeds. *Nucl. Acids Res.* **26:** 3986–3990.

Zhou, F., Guo, Y. Qiu, L.J. (2016). Genome-wide identification and evolutionary analysis of *LEUCINE-RICH REPEAT RECEPTOR-LIKE PROTEIN KINASE* genes in soybean. *BMC Plant Biol.* **16:**