Article

In Silico Phylogenetic and Structural Analysis of Plant Damps

Athanasia Pavlopoulou[1],*, Ezgi Karaca[1],[2], Alma Balestrazzi[3], Alexandros G. Georgakilas[4],**

[1]Izmir Biomedicine and Genome Institute, Dokuz Eylül University, Izmir 35340, Turkey.
[2]Izmir Biomedicine and Genome Center (IBG), Izmir 35340, Turkey; ezgi.karaca@ibg.edu.tr
[3]Department of Biology and Biotechnology 'Lazzaro Spallanzani', University of Pavia, via Ferrata 1, 27100, Pavia, Italy; alma.balestrazzi@unipv.it
[4]DNA Damage Laboratory, Department of Physics, School of Applied Mathematical and Physical Sciences, National Technical University of Athens (NTUA), Athens, Greece
*Correspondence: athanasia.pavlopoulou@deu.edu.tr; Tel.: +90-232-412-6549
**Correspondence: alexg@mail.ntua.gr; Tel.: +30--210-7724453

Abstract: In plants and animals, endogenous biological molecules, termed damage-associated molecular patterns (DAMPs) or alarmins, are released by damaged, stressed or dying cells following abiotic stress such as radiation and drought stress. In turn, a cascade of downstream signaling events is initiated leading to the up-regulation of defense-related genes. In the present study, in an effort to investigate the conservation status of the molecular mechanisms implicated in the danger signaling, thorough *in silico* phylogenetic and structural analyses of the effector biomolecules were performed in taxonomically diverse plant species. On the basis of our results, the defense mechanisms appear to be largely conserved within the plant kingdom. Of note is our finding that the sequence and/or function of several components of these mechanisms were found to be conserved in animals, as well.

Keywords: DAMPs; alarmins; innate immune response; plants; phylogeny; *in silico*

1

1. Introduction

Death-, danger- or damage-associated molecular patterns (DAMPs) are of biological origin and are considered the major immunogenic mediators released passively by damaged, stressed or dying cells, including tumor cells targeted by radiation or chemotherapy [1]. DAMPs are host cell-derived, as opposed to foreign (non-self) molecules termed Microbe-Associated Molecular Patterns (MAMPs) or Pathogen-Associated Molecular Patterns (PAMPs) [2]. In a seminal study on this field by Matzinger (1994), it was suggested that the immune system does not actually distinguish between the self and non-self but rather detects "danger" through a series of positive and negative signals derived from damaged or stressed tissues mediated by DAMPs [3,4].

DAMPs are perceived by pattern recognition receptors (PRRs), such as plasma membrane-localized receptors, thereby leading to a cascade of events including cytoplasmic Ca^{2+} elevation, depolarization of the cell membrane, production of reactive oxygen species (ROS), the transient phosphorylation of mitogen activated protein kinases (MAPKs), and the transcriptional upregulation of defense-related genes.

Plants have evolved mechanisms of innate immunity against detrimental pathogenic microorganisms and herbivorous animals. As mentioned above, DAMPs or alarmins i.e. biomolecules released by stressed cells, in many ways share similarities between plants and animals. For example, the recently discovered *Arabidopsis* HMGB3, is the counterpart of the pivotal animal DAMP HMGB1 [2].

Plants' immune response against intruders is regulated mainly by two antagonistic defense signaling pathways: i) the salicylic acid (SA)-mediated signal transduction pathway elicited by biotrophic and hemiobiotrophic pathogens and ii) the octadecanoid signaling pathway with the key hormone jasmomic acid (JA) induced by heterotrophic pathogens and herbivores [5-7]. Of note, the oxidized lipids that participate in octadecanoid signaling are the plant functional equivalents of the mammalian oxidized phospholipids [8].

To date, there is still a knowledge gap concerning the evolutionary origin of the different DAMPs identified in the plant kingdom and how this valuable knowledge can be projected to more complex organisms, and especially animal cells. In this study, we have made an effort to assess the conservation of the defense mechanisms by conducting phylogenetic and structural analyses of the effector molecules implicated in danger signaling in 11 vascular plant species that represent diverse taxonomic divisions: *Arabidopsis thaliana* (thale-cress), *Zea mays*, *Oryza sativa* (rice), *Hordeum vulgare* (barley), *Medicago truncatula* (barrel medic), *Nicotiana tabacum* (tobacco), *Populus trichocarpa* (cottonwood), *Solanum lycopersicum* (tomato), *Prunus persica* (peach), *Vitis vinifera* (wine grape), *Pisum sativum* (pea) (Figure S1).

2. Methods

2.1. Sequence dataset & homology searching

The names and/or accession numbers of the characterized proteins reported in PubMed (https://www.ncbi.nlm.nih.gov/pubmed) articles were used to retrieve their corresponding sequences from the publicly accessible sequence databases UniProtKB [9] and NCBI's GenBank [10]. To identify more orthologous protein sequences, the known sequences were used as query in reciprocal BLASTp and tBLASTn [11] searches of the genomes of 11 vascular plant species, representing diverse taxonomic divisions (Figure S1). The canonical or longest transcripts were selected. Any partial or ambiguous sequences were not included in the subsequent steps of the study.

2.2. Phylogenetic analyses

The amino acid sequences of the corresponding proteins (Table S1) were aligned with MAFFT, version 3.7 [12]. Phylogenetic analyses were performed by employing a maximum likelihood (ML) method. To this end, the resulting multiple alignments were provided as input to PhyML v.3.0 [13], which optimizes a distance-based starting tree (BioNJ) [14] by employing a heuristic algorithm. The expected number of

amino acid substitutions per site was estimated with the JTT model [15]. Protein sequences distantly related to those under study were used as outgroups. The robustness of the reconstructed phylogenetic trees was evaluated by bootstrapping (200 bootstrap pseudo-replicates). Trees were illustrated using Dendroscope [16].

2.3. Pairwise distances estimation

Pairwise distances between protein sequences were computed using the software package MEGA, version 7.0 [17].

2.4. Protein Domain Organization

The protein sequences under study were searched against the protein signature databases SMART v.8.0 [18] and CDD v.3.16 [19] in order to determine the boundaries of their constituent domains.

2.5. Sequence Motifs

The amino acid sequences corresponding to the catalytic kinase domain of the orthologous proteins DORN1, MAPK3, MAPK6, PEPR1/2, SERK3/4/5, SR106 and WAKL (Table S1), were aligned using MAFFT v.3.7 [12] and edited with Utopia's CINEMA alignment editor [20]. Ungapped sequence motifs were extracted from the alignment and submitted to WebLogo 3.5.0 [21] with default parameters, to create consensus sequences.

2.6. Tertiary structure analysis

The plant MAPK3 and MAPK6 sequences under study were aligned using the PROMALS3D [22,23] multiple sequence alignment program, which incorporates evolutionary and tertiary structural information to improve alignment accuracy. The degree of conservation of amino acid residues of the homologous plant MAPK3 and MAPK6 proteins was estimated with the usage of the ConSurf program [24]. To this end, the resulting multiple sequence alignment of the MAPK3 and MAPK6 amino acid sequences was provided as input to ConSurf to map the conservation grades of the amino acid residues onto the resolved tertiary structure of the *Arabidopsis thaliana* MAPK6 (PDB ID: 5CI6; [25]). The protein structure was displayed using the PyMol molecular graphics program (http://www.pymol.org/).

2.7. Structural modeling of protein-protein interactions

The tertiary structure of the extracellular LRR domain of the *A. thaliana* PEPR2 was predicted by homology modeling. The X-ray crystal structures of the *Arabidopsis* PEPR1LRR (PDB ID: 5GR8, chain A; [26]) was used as template to model PEPR2LRR using the I-TASSER server [27]. The quality of the final modeled protein structures was evaluated with Procheck [28].

In order to assemble the trimeric PEPR2LRR-AtPEP1-BAK1 complex, the following steps were carried out: LRR domains (bold) present in the **PEPR1LRR**-AtPEP1 (PDB ID: 5GR8) and **FLS2LRR**-FLG22-BAK1 (PDB ID: 4MN8) complexes, and the **PEPR2LRR** homology model were structurally superimposed by using the alignment algorithm FATCAT [29]. Following this superimposition, the coordinates of FLS2LRR-FLG22-BAK1 and PEPR1LRR were deleted. As a result of this procedure, a crude model of the PEPR2LRR-AtPEP1-BAK1 complex was obtained. This crude model was refined with the water refinement step of the HADDOCK webserver [30]. During refinement, the interaction depicted in Figure 7B was used as a restraint, with the following restraint definition: i) assign (resi 438 and segid A) (resi 23 and segid C) 3.00 3.00 0.05 and ii) assign (resi 392 and segid A)(resi 23 and segid C) 3.00 3.00 0.05. All of the interface statistics were carried out through CoCoMaps tool (https://www.molnac.unisa.it/BioTools/cocomaps/).

3.   Results and Discussion

3.1. Classes of DAMPs

The diverse DAMPs that have been identified in plants are arbitrarily grouped, in this study, in the following classes.

3..1.1. Cell wall-derived DAMPs

Oligogalacturonide (OG) fragments are released from the degradation of cell wall constituent homogalacturonan (HGA) by pathogen-encoded polygalacturonases. The OG-mediated signal transduction pathway involves ROS production, increase in cytoplasmic Ca^{2+} concentration, activation of MAPK3 and MAPK6, upregulation of resistance genes, activation of components of the SA, JA, and ethylene pathways [31]. The wall-associated kinase 1 (WAK1) was suggested to perceive [32,33] OGs.

Likewise, in animals, glycosaminoglycan hyaluronan fragments are released during damage and are implicated in wound repair and regeneration. These fragments are recognized by the Toll-like receptors TLR2 and TLR4 and activate the expression of inflammatory genes [34].

Figure 1. Phylogenetic distribution of plant effector proteins.

The viridiplantae MAPK3 and MAPK6 proteins have orthologs in metazoa (Figure 1) which are also implicated in danger signaling [35]. The plant MAPK3 and MAPK6 have rather evolved independently from their animal orthologs as they form distinct monophyletic groups with bootstrap values of 100 and they are connected to the corresponding metazoan clade with a long branch which reflects long evolutionary distance (Figure 2 and 3). The monocotyledonous plant MAPK3 and MAPK6 sequences form well-separated clades in their respective phylogenetic trees, leading to the suggestion that these sequences have probably appeared after the monocot-eudicot divergence about 200 million years ago (MYA) [36]. The tomato and tobacco MAPK3 and MAPK6 sequences cluster together with high statistical confidence, indicating that these sequences emerged after the members of the order solanales branhed off from the other vascular plants (Figure 2 and 3).
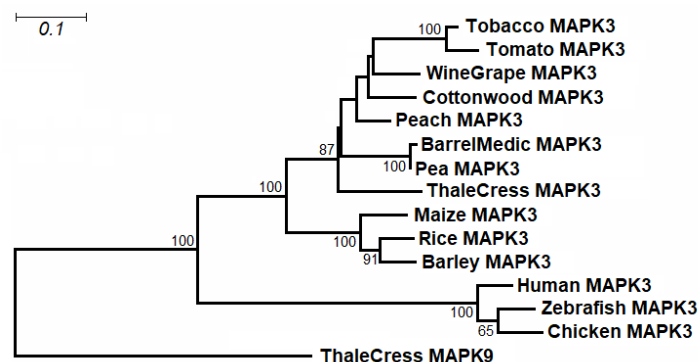
Figure 2. ML-based tree of MAPK3 protein sequences. The branch lengths are proportional to the evolutiotary distance. Bootstrap support values ≥ 50% are shown at the nodes. The scale bar at the upper right denotes the length of amino acid replacements per position. The sequence ThaleCress MAPK9 was used as outgroup.
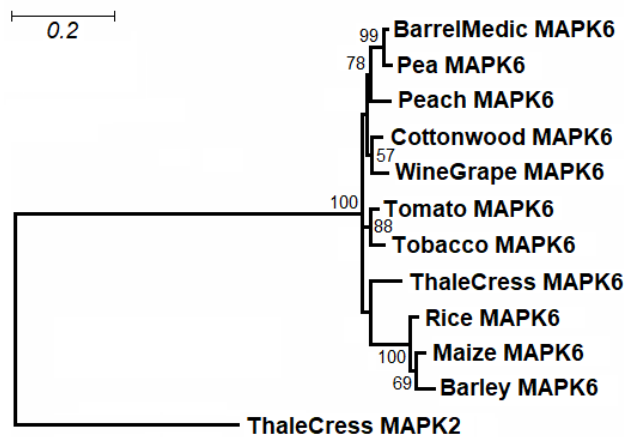

Figure 3. Tree of MAPK6 protein sequences. The sequence ThaleCress MAPK9 was used as outgroup. The conventions are the same as in Figure 2.

A cluster of five highly similar *WAK (WAK1-5)* genes are present in *A. thaliana* (thale cress) [37]. The *Arabidopsis* WAK1-5 protein sequences form a separate highly supported clade (Figure 4). Based on sequence database searches, a single copy of *WAK* was found in the other plant species (Figure 1). WAK homologs were not detected in pea; this is probably due to incomplete annotation of the *Pisum sativum* genome. The encoded protein was arbitrarily referred to as WAKL, where "L" stands for "Like", by virtue of sequence similarity to the thale cress WAK1-5 (Figure 4). A primordial *WAK* gene might have undergone a series of duplications in *A. thaliana*, thereby giving rise to five *WAK* paralogs. However, there are not available fully sequenced and annotated genomes of other members of the order brassicales, apart from the one of the model plant organism *A. thaliana*, which would allow us to examine whether the *WAK* gene expansion is species- or order-specific. Based on extensive bibliographic searches, no experimental evidence was found about *A. thaliana* WAK2-5 acting as PRRs. WAK1 has apparently acquired a more specialized function during the course of evolution. WAK proteins are likely restricted to vascular plants, since orthologous WAK sequences were not found in metazoa. The role of these WAKL proteins in the perception of OGs in their host plant organisms remains to be investigated experimentally.
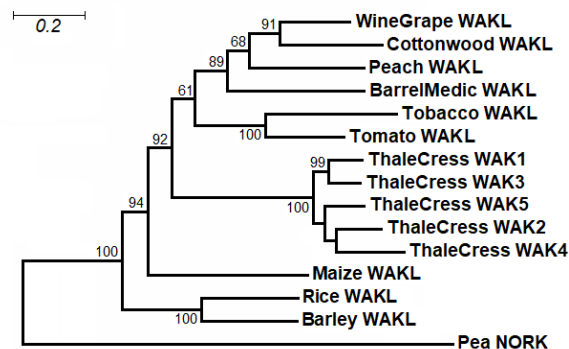

Figure 4. Tree of WAKL protein sequences. The sequence Pea NORK (nodulation receptor kinase) was used as outgroup. The conventions are the same as in Figure 2.

3.1.2. Extracellular ATP

5

Extracellular ATP (eATP), released from stressed or damaged cells into the extracellular milieu, constitutes a class of danger signals both in plants and animals. In plants, eATP plays an important signaling role in plant cells by participating in several processes including cell development, viability and stress responses [38,39]. It is recognized by the plant-specific PRR, DORN1 (DOes not Respond to Nucleotides 1), which is a membrane lectin receptor kinase [40]. DORN1 is required for the ATP-mediated intracellular influx of Ca^{2+}, formation of ROS, phosphorylation of MAPK3 and MAPK6 and stimulation of the expression of defense genes [40,41].

Putative DORN1 homologous sequences were detected in all plant species under investigation (Figure 1). The cereal plants as well as the solanales DORN1 sequences form their own monophyletic group with bootstrap values of 100, suggesting class- and order-specific evolution of *DORN1* (Figure 5). DORN homologous sequences were found exclusively in plants, wherein might be implicated in the recognition of eATP.

In animals, eATP is involved in inflammatory response by activating mast cells [42,43]. Two types of P2 purinoceptors, the ligand-gated cation channel P2X receptors and the metabotropic (G-protein-coupled) P2Y receptors [44], are activated by eATP and are involved in the eATP-induced increase of cytosolic Ca^{2+} levels [45].
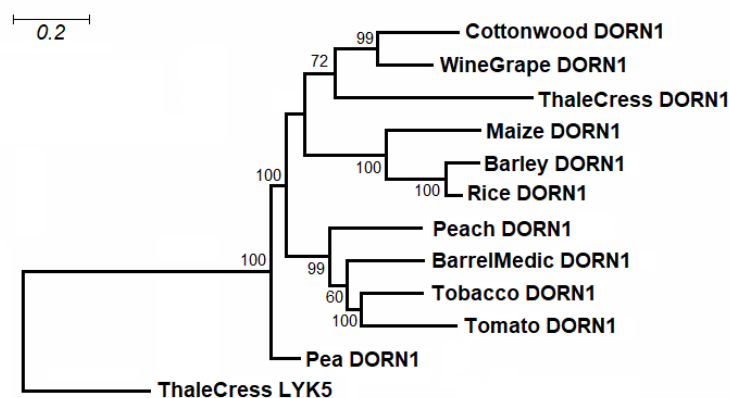

Figure 5. Tree of DORN1 protein sequences. The sequence ThaleCress LYK5 was used as outgroup. The conventions are the same as in Figure 2.

3.1.3. Polypeptide-based DAMPs

A wide range of polypeptide-based DAMPs are derived from larger precursors [46]. In *Arabidopsis thaliana* the 23-amino acid PEP1 (Plant Elicitor Peptide) is processed from a 92-amino acid precursor upon damage, JA, and ethylene [47]. Based on homology searches, only one ortholog of AtPEP1 was detected in the fellow plants, that is, ZmPEP1 in maize which exhibited DAMP-like behavior [48]. In *A. thaliana*, PEP1 is perceived by two PRRs, PEPR1 (PEPtide Receptor 1) and PEPR2, leading to ROS production and upregulation of defensin-like genes [26,47,49,50]. Of note, the only PEP orthologs were found in two species belonging to divergent taxonomic groups, eudicots (*A. thaliana*) and monocots (*Zea mays*), and not any taxonomically related species (such as fellow eudicots). This suggest that even distantly related organisms respond to similar stimuli with similar mechanisms.

PEPs' cognate receptors, however, were found in all plants under study (except pea) (Figure 1). Thale cress harbors a total of seven PEPs (PEP1-PEP7) and two *PEPR* paralogs, the encoded proteins of which share 67% identity. In the inferred phylogenetic tree, the AtPEPR1 and AtPEPR12 protein sequences form a highly monophyletic branch with bootstrap support 100 (Figure 6). However, a single copy of *PEPR* was detected in the other plant species, the corresponding protein of which is called PEPR1/2 due to its

6

amino acid sequence similarity to both AtPEPR1 and AtPEPR2. The cereal plants form a separate highly supported clade, leading to the suggestion that *PEPR1/2* genes evolved after the monocots-dicots split (Figure 6). *PEPRs* are also kingdom-restricted, since *PEPR* orthologs were detected exclusively in vascular plants.
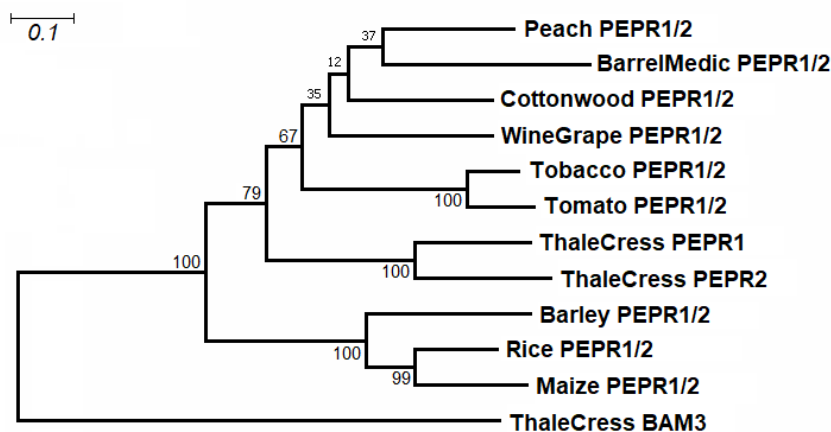

Figure 6. Tree of PEPR1/2 protein sequences. The sequence ThaleCress BAM3 was used as outgroup. The conventions are the same as in Figure 2.

AtPEPs exhibit functional similarity to the 18-amino acid peptide systemin, which is processed from the 200-amino acid hormone prosystemin (SYST) in tomato upon wounding [51]. Prosystemin was found only in tomato, and not in tobacco which is also member of the same solanales order (Figure 1). However, two hydroxyproline-rich systemin (HSY) peptides, designated HSYA and HSYB, cleaved from a larger preprotein, were identified in tobacco (Figure 1) which do not bear any sequence similarity to systemin [52]. Both SYST and HSY induce a JA-mediated signaling pathway that leads to the activation of defense-related genes [51,52].

Of particular note, as in the case of AtPEPRs, systemin cognate receptors (SR160) were detected in all plant species under investigation (Figure 1). It would be intriguing to suggest that *PEPR* and *SR160* genes were probably propagated through vertical gene transfer in plants, and the peptides that specifically bind to them evolved later in certain species in order to carry out species-specific functions.

Protein fragments, such as fibronectin fragments (FN-fs), are the equivalents of the plant polypeptide-based DAMPs in mammals, and they are perceived by Toll-like receptors (TLRs) [53]. PEPR1/2 and SR160 are transmembrane (TM) proteins comprised of LRR (Leucine Rich Repeats) motifs, one TM domain, and one catalytic kinase domain in the cytoplasmic region. PEPRs resemble the mammalian TLRs in terms that they share LRR motifs in their amino-termini, suggesting that the extracellular LRR domain is used for DAMP perception both in plants and mammals.

The interaction between PEPR2LRR-AtPEP1 (Figure 7) was also modeled by using the crystal structure of PEPR1LRR-AtPEP1 (PDB ID: 5GR8) as a template. According to 5GR8, there are 16 hydrogen bonds formed between PEPR1LRR and its AtPEP1 peptide [26]. In our PEPR2LRR-AtPEP1 model, this number increases to 17. Two hydrogen bonds, among all, were found to be strictly conserved, i.e. the ones between TYR395-GLN21 and ASP441-ASN23 across the PEPR1LRR-AtPEP1 interface; TYR346-GLN21 and ASP392-ASN23 across the PEPR2LRR-AtPEP1 interface. This reflects the importance of these stabilizing interactions within the PEPR family. The extent of the PEPR1LRR-AtPEP1 and PEPR2LRR-AtPEP1 interfaces appear to be conserved, where 66 amino acids interact over an area of 1033 A2 in the case of PEPR1LRR-AtPEP1, and 65 amino acids interact through a 1046 A2 sized interface.

7

PEPR2LRR-BAK1 and its template FLS2LRR-BAK1 (PDB ID: 4MN8) entail a similar number of amino acids across their interfaces (57 for PEPR2LRR-BAK1 and 56 for FLS2LRR-BAK1) (Figure 7). Although this is the case, PEPR2LRR-BAK1 makes more polar interactions (through 12 hydrogen bonds) compared to FLS2LRR-BAK1 (entails 5 hydrogen bonds only), indicating the polar character of the PEPR2LRR surface and a possibly higher affinity between PEPR2LRR-BAK1.
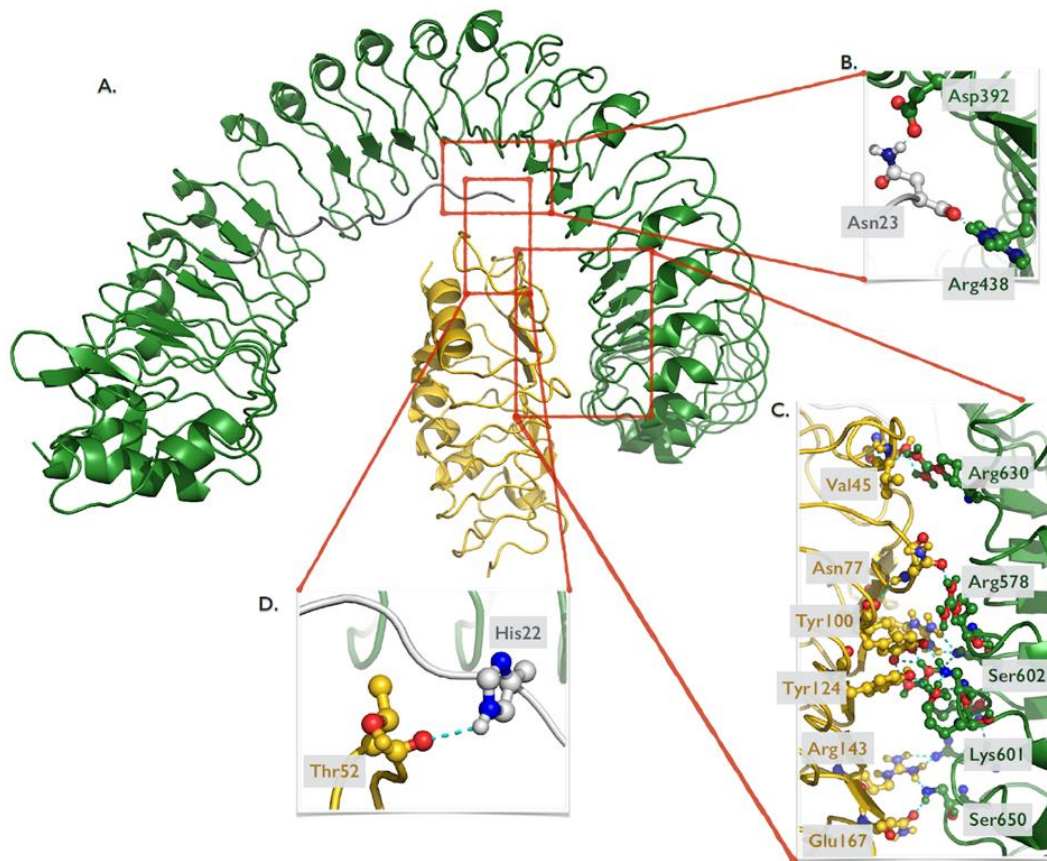

Figure 7. A. Structural model of the *Arabidopsis thaliana* PEPR2LRR(green)-AtPEP1(gray)-BAK1(gold) complex. Important polar interactions formed across the PEPR2LRR and AtPEP1, PEPR2LRR and BAK1, AtPEP1 and BAK1 interfaces are depicted in panels **B., C., D.**, respectively. Polar interactors are represented in ball-and-sticks, with oxygen atoms colored in red and nitrogens in blue.

3.1.4. HMGB3

A relatively novel class of plant DAMPs, the *Arabidopsis thaliana* HMBG3 (High Mobility Group Box 3), displayed DAMP-like responses, including activation of MAPK3/4/6 and defense-related genes, as well as increased resistance to necrotrophic pathogens. This response is mediated by the LRR-PRRs BAK1 (Brassinosteroid insensitive 1-Associated Kinase 1)/SERK3 (Somatic Embryogenesis Receptor-like Kinase 3) and BKK1 (BAK1-like 1)/SERK4 [54]. The mammalian counterpart of the plant HMBG3 is HMBG1, released upon tissue wounding and inflammation [55]. HMGB1 is a chromosomal scaffold protein involved in chromatin remodeling induced in response to DNA damage accumulation. HMGB1 promotes nucleotide excision repair (NER), base excision repair (BER) and mismatch repair (MMR) pathways, facilitating DNA protein kinase (DNA-PK) activity [56]; similarly to the catalytic subunit of DNA-PK, DNA-PKcs, HMGB1 binds CpG oligonucleotides, inducing the TLR9-mediated inflammatory response [57].

The human HMGB1 harbors two HMG boxes, BoxA and BoxB, whereas the *A. thaliana* HMGB proteins have only one HMG box. In the plant species under study, single HMBG homologs were detected that were found to be related to *Arabidopsis* HMBG1/2/3. This is because the *A. thaliana* proteins HMBG1, HMBG2 and HMBG3 share a high degree of sequence identity/similarity. HMBG1/2/3-related sequences were found in all plant species (Figure S1).

Genes encoding SERKs have been identified in the genomes of vascular plants, both monocots and dicots [58]. Furthermore, Sasaki et al. [59] reported the presence of SERK homologs in the less evolved plant liverwort *Marchantia polymorpha* and even in the unicellular green alga *Closterium ehrenbergii*. *SERKs* are ancient, essential genes, conserved during speciation. Multiple *SERK* genes are found in dicots, monocots, and non-vascular plants. This indicates that SERK genes are at least 450 million years old, and were present before the split of non-vascular and vascular plants [60]. Putative SERK-related proteins were detected in the species under investigation (except pea) (**Figure 1**).
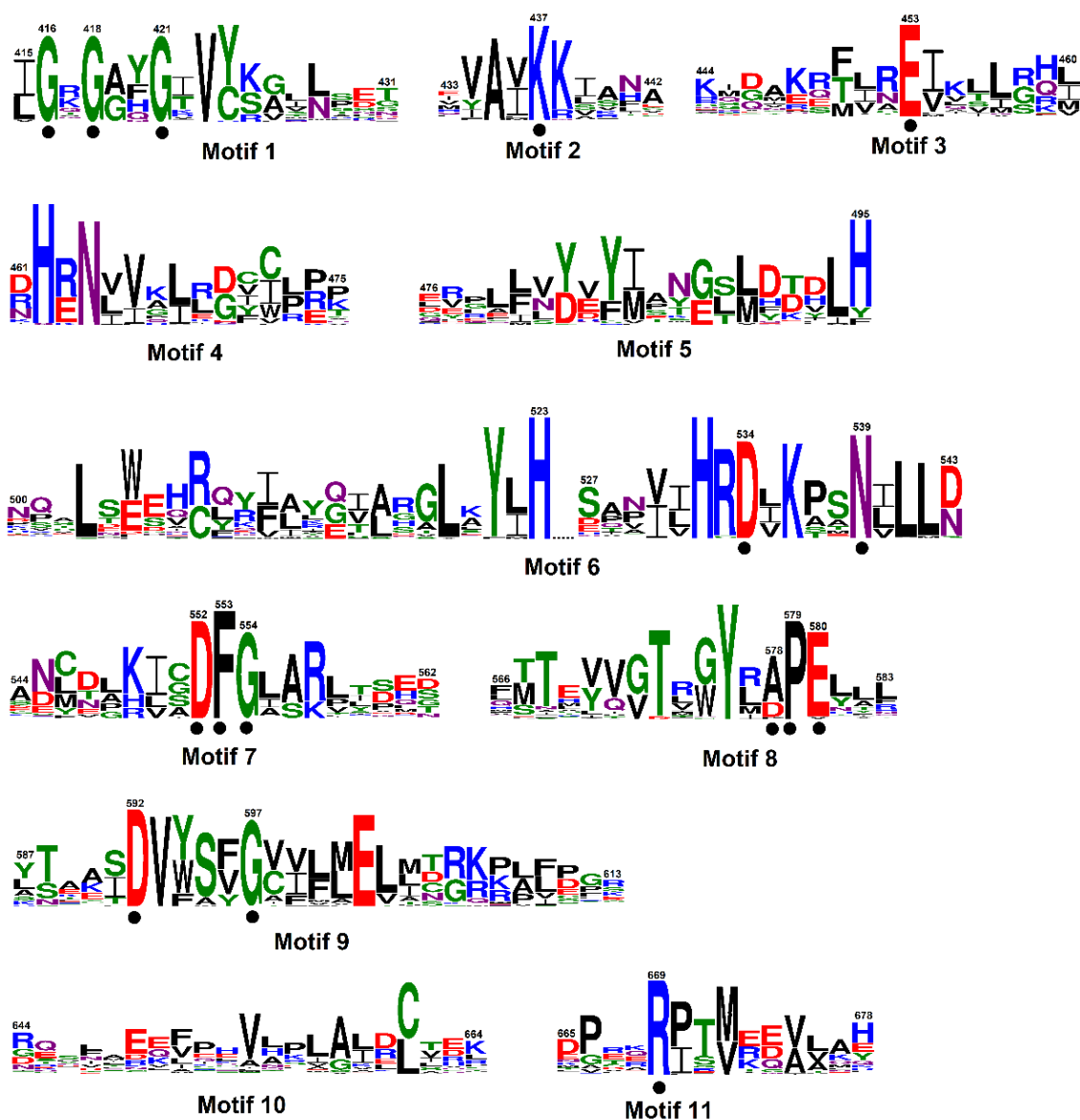
Figure 8. Sequence motifs identified in the kinase domain of plant proteins. Sequence logos of the conserved protein motifs identified in the catalytic domain of protein kinases. The first and last amino acid residues of each motif are numbered according to *A. thaliana* WAK1. The conserved/invariant amino acid residues critical for the protein kinase activity are denoted by dots. The height of each letter depicts the relative frequency of the corresponding amino at that position, and the letters are ordered in such a way that the most frequent one is on the top.

3.2. Conserved structural features of plant proteins implicated in danger signaling

In the present study, the 11 conserved catalytic sequence motifs reported by Hanks et al [61] were identified in the plant kinase domains under investigation (Figure 8). In motif 1, the consensus signature GxGxxG of phosphate binding, conserved across nucleotide binding proteins [62,63], was detected. The invariant lysine (K) in motif 2 is essential in protein kinase activity as it participates actively in the phosphate transfer reaction [64]. The invariant aspartate (D) and asparagine (N) amino acids in motif 6, as well as the stretch of Asp (D), Phe (F) and Gly (G) residues in motif 7 are actively involved in ATP binding [65]. In motif 8, the highly conserved Glu (E) and invariant Pro (P) were found to be flanked by a conserved alanine (A) residue in [61]. However, in the plant kinase domain, at the same position, Asp (D) was found to be the most frequent residue instead. Phosphorylation of these residues was associated with enhanced catalytic activity in a number of protein kinases [66,67]. Moreover, several highly conserved amino acids with an unknown role identified by Hanks et al [61] were also found to be conserved in plant kinases, including Asp (D) in motif 3, Asp (D) and Gly (G) in motif 9 and Arg (R) in motif 11.

The degree of conservation of the amino acid residues critical for kinase activity is depicted in the three-dimensional structure of *A. thaliana* MAPK6 [25] (Figure 9). A total 13 out of the 16 amino acids found to be conserved/invariant in the primary structures of the kinase domain of plant proteins (Figure 8) are also shown to be conserved in the tertiary structure of this domain (Figure 9).
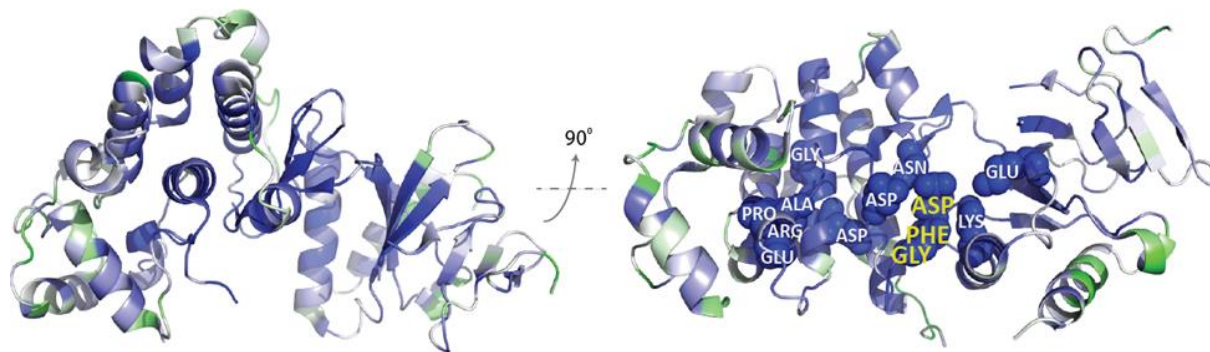

Figure 9. A. Tertiary structure … blue corresponds to the most highly conserved region, where green to the least one. The conserved DFG (207-209) motif is highlighted in yellow (left panel, at the center of the molecule). The other important conserved residues can be listed as (from left to right) GLY251, ALA 231, PRO232, GLU233, ARG336, ASP246, ASP189, ASN194, GLU110, LYS92. The numbering is given as in 5CI6.

4. Conclusion

It has been only about two decades has the significance of DAMPs for the survival and homeostasis of multicellular organisms has surfaced. Related to the more than twenty DAMPs discovered in animals to date, relatively few have been recognized in plants[68,69]. Based on our study, several of these plant DAMPs have counterparts in animals, including eATP, HMGBs and MAPKs. This finding supports the idea that all efforts to elucidate the pathway(s) through which innate immunity is triggered will likely

10

identify additional signaling components that are shared by all kingdoms, from planta to humans. The "ancestral" mechanisms of defense are conserved across kingdoms. Several key molecular "players" of those mechanisms appear to be largely conserved (e.g. eATP, calcium, MAPKs). Some plant effector proteins were found to have sequence homologs in animals (e.g., MAPKs), whereas most of them were detected exclusively in plants. Both in plants and animals, selective pressure has been likely exerted to maintain the functional integrity of the defense mechanisms by recruiting or co-opting into the existing molecular pathways novel players. These players may have no apparent sequence / structural similarity but, through convergent evolution, have rather acquired similar functions, as in the case of DORN1 and P2 purinoreceptors, in plants and animals, respectively. These findings could provide the foundation for expanding the current knowledge on DAMP molecules in plants and their role in adaptive defense mechanisms.

Supplementary Materials: Figure S1. A cladogram extracted from NCBI's Taxonomy database [70] depicting the evolutionary relationships among plant species. Table S1. Accession numbers of the protein sequences investigated in the present study.

Author Contributions: A.P. and A.G.G. designed and supervised the study. A.P. and E.K. performed data analyses. A.P., E.K., A.B. and A.G.G. wrote the manuscript. All authors reviewed and approved of the final manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Garg, A.D.; Galluzzi, L.; Apetoh, L.; Baert, T.; Birge, R.B.; Bravo-San Pedro, J.M.; Breckpot, K.; Brough, D.; Chaurio, R.; Cirone, M., *et al.* Molecular and translational classifications of damps in immunogenic cell death. *Frontiers in immunology* **2015**, *6*, 588.
2. Choi, H.W.; Klessig, D.F. Damps, mamps, and namps in plant innate immunity. *BMC plant biology* **2016**, *16*, 232.
3. Matzinger, P. Tolerance, danger, and the extended family. *Annual review of immunology* **1994**, *12*, 991-1045.
4. Seong, S.Y.; Matzinger, P. Hydrophobicity: An ancient damage-associated molecular pattern that initiates innate immune responses. *Nature reviews. Immunology* **2004**, *4*, 469-478.
5. Thaler, J.S.; Humphrey, P.T.; Whiteman, N.K. Evolution of jasmonate and salicylate signal crosstalk. *Trends in plant science* **2012**, *17*, 260-270.
6. Shah, J. The salicylic acid loop in plant defense. *Current opinion in plant biology* **2003**, *6*, 365-371.
7. Campos, M.L.; Kang, J.H.; Howe, G.A. Jasmonate-triggered plant immunity. *Journal of chemical ecology* **2014**, *40*, 657-675.
8. Pieterse, C.M.; Leon-Reyes, A.; Van der Ent, S.; Van Wees, S.C. Networking by small-molecule hormones in plant immunity. *Nature chemical biology* **2009**, *5*, 308-316.
9. The UniProt, C. Uniprot: The universal protein knowledgebase. *Nucleic acids research* **2017**, *45*, D158-D169.
10. Benson, D.A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Ostell, J.; Pruitt, K.D.; Sayers, E.W. Genbank. *Nucleic acids research* **2018**, *46*, D41-D47.
11. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *Journal of molecular biology* **1990**, *215*, 403-410.
12. Katoh, K.; Rozewicki, J.; Yamada, K.D. Mafft online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics* **2017**.
13. Guindon, S.; Lethiec, F.; Duroux, P.; Gascuel, O. Phyml online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic acids research* **2005**, *33*, W557-559.

14. Gascuel, O. Bionj: An improved version of the nj algorithm based on a simple model of sequence data. *Molecular biology and evolution* **1997**, *14*, 685-695.

15. Jones, D.T.; Taylor, W.R.; Thornton, J.M. The rapid generation of mutation data matrices from protein sequences. *Computer applications in the biosciences : CABIOS* **1992**, *8*, 275-282.

16. Huson, D.H.; Scornavacca, C. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Systematic biology* **2012**, *61*, 1061-1067.

17. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. Mega6: Molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution* **2013**, *30*, 2725-2729.

18. Letunic, I.; Bork, P. 20 years of the smart protein domain annotation resource. *Nucleic acids research* **2018**, *46*, D493-D496.

19. Marchler-Bauer, A.; Bryant, S.H. Cd-search: Protein domain annotations on the fly. *Nucleic acids research* **2004**, *32*, W327-331.

20. Parry-Smith, D.J.; Payne, A.W.; Michie, A.D.; Attwood, T.K. Cinema--a novel colour interactive editor for multiple alignments. *Gene* **1998**, *221*, GC57-63.

21. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. Weblogo: A sequence logo generator. *Genome research* **2004**, *14*, 1188-1190.

22. Pei, J.; Grishin, N.V. Promals3d: Multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods in molecular biology* **2014**, *1079*, 263-271.

23. Pei, J.; Kim, B.H.; Grishin, N.V. Promals3d: A tool for multiple protein sequence and structure alignments. *Nucleic acids research* **2008**, *36*, 2295-2300.

24. Ashkenazy, H.; Abadi, S.; Martz, E.; Chay, O.; Mayrose, I.; Pupko, T.; Ben-Tal, N. Consurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research* **2016**, *44*, W344-350.

25. Wang, B.; Qin, X.; Wu, J.; Deng, H.; Li, Y.; Yang, H.; Chen, Z.; Liu, G.; Ren, D. Analysis of crystal structure of arabidopsis mpk6 and generation of its mutants with higher activity. *Scientific reports* **2016**, *6*, 25646.

26. Tang, J.; Han, Z.; Sun, Y.; Zhang, H.; Gong, X.; Chai, J. Structural basis for recognition of an endogenous peptide by the plant receptor kinase pepr1. *Cell research* **2015**, *25*, 110-120.

27. Yang, J.; Zhang, Y. I-tasser server: New development for protein structure and function predictions. *Nucleic acids research* **2015**, *43*, W174-181.

28. Laskowski, R.A.; Rullmannn, J.A.; MacArthur, M.W.; Kaptein, R.; Thornton, J.M. Aqua and procheck-nmr: Programs for checking the quality of protein structures solved by nmr. *Journal of biomolecular NMR* **1996**, *8*, 477-486.

29. Ye, Y.; Godzik, A. Fatcat: A web server for flexible structure comparison and structure similarity searching. *Nucleic acids research* **2004**, *32*, W582-585.

30. Bonvin, A.; Karaca, E.; Kastritis, P.L.; Rodrigues, J. Defining distance restraints in haddock. *Nature protocols* **2018**, *13*, 1503.

31. Denoux, C.; Galletti, R.; Mammarella, N.; Gopalan, S.; Werck, D.; De Lorenzo, G.; Ferrari, S.; Ausubel, F.M.; Dewdney, J. Activation of defense response pathways by ogs and flg22 elicitors in arabidopsis seedlings. *Molecular plant* **2008**, *1*, 423-445.

32. Decreux, A.; Messiaen, J. Wall-associated kinase wak1 interacts with cell wall pectins in a calcium-induced conformation. *Plant & cell physiology* **2005**, *46*, 268-278.

33. Brutus, A.; Sicilia, F.; Macone, A.; Cervone, F.; De Lorenzo, G. A domain swap approach reveals a role of the plant wall-associated kinase 1 (wak1) as a receptor of oligogalacturonides. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 9452-9457.

34. Jiang, D.; Liang, J.; Noble, P.W. Hyaluronan in tissue injury and repair. *Annual review of cell and developmental biology* **2007**, *23*, 435-461.

35. Plotnikov, A.; Zehorai, E.; Procaccia, S.; Seger, R. The mapk cascades: Signaling components, nuclear roles and mechanisms of nuclear translocation. *Biochimica et biophysica acta* **2011**, *1813*, 1619-1633.

36. Wolfe, K.H.; Gouy, M.; Yang, Y.W.; Sharp, P.M.; Li, W.H. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proceedings of the National Academy of Sciences of the United States of America* **1989**, *86*, 6201-6205.

37. He, Z.H.; Cheeseman, I.; He, D.; Kohorn, B.D. A cluster of five cell wall-associated receptor kinase genes, wak1-5, are expressed in specific organs of arabidopsis. *Plant molecular biology* **1999**, *39*, 1189-1196.

38.   Cao, Y.; Tanaka, K.; Nguyen, C.T.; Stacey, G. Extracellular atp is a central signaling molecule in plant stress responses. *Current opinion in plant biology* **2014**, *20*, 82-87.

39.   Tanaka, K.; Choi, J.; Cao, Y.; Stacey, G. Extracellular atp acts as a damage-associated molecular pattern (damp) signal in plants. *Frontiers in plant science* **2014**, *5*, 446.

40.   Choi, J.; Tanaka, K.; Cao, Y.; Qi, Y.; Qiu, J.; Liang, Y.; Lee, S.Y.; Stacey, G. Identification of a plant receptor for extracellular atp. *Science* **2014**, *343*, 290-294.

41.   Choi, J.; Tanaka, K.; Liang, Y.; Cao, Y.; Lee, S.Y.; Stacey, G. Extracellular atp, a danger signal, is recognized by dorn1 in arabidopsis. *The Biochemical journal* **2014**, *463*, 429-437.

42.   Burnstock, G.; Boeynaems, J.M. Purinergic signalling and immune cells. *Purinergic signalling* **2014**, *10*, 529-564.

43.   Yoshida, K.; Ito, M.; Matsuoka, I. Divergent regulatory roles of extracellular atp in the degranulation response of mouse bone marrow-derived mast cells. *International immunopharmacology* **2017**, *43*, 99-107.

44.   Burnstock, G.; Kennedy, C. Is there a basis for distinguishing two types of p2-purinoceptor? *General pharmacology* **1985**, *16*, 433-440.

45.   Wareham, K.J.; Seward, E.P. P2x7 receptors induce degranulation in human mast cells. *Purinergic signalling* **2016**, *12*, 235-246.

46.   Albert, M. Peptides as triggers of plant defence. *Journal of experimental botany* **2013**, *64*, 5269-5279.

47.   Huffaker, A.; Pearce, G.; Ryan, C.A. An endogenous peptide signal in arabidopsis activates components of the innate immune response. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 10098-10103.

48.   Huffaker, A.; Dafoe, N.J.; Schmelz, E.A. Zmpep1, an ortholog of arabidopsis elicitor peptide 1, regulates maize innate immunity and enhances disease resistance. *Plant physiology* **2011**, *155*, 1325-1338.

49.   Krol, E.; Mentzel, T.; Chinchilla, D.; Boller, T.; Felix, G.; Kemmerling, B.; Postel, S.; Arents, M.; Jeworutzki, E.; Al-Rasheid, K.A., *et al.* Perception of the arabidopsis danger signal peptide 1 involves the pattern recognition receptor atpepr1 and its close homologue atpepr2. *The Journal of biological chemistry* **2010**, *285*, 13471-13479.

50.   Yamaguchi, Y.; Huffaker, A.; Bryan, A.C.; Tax, F.E.; Ryan, C.A. Pepr2 is a second receptor for the pep1 and pep2 peptides and contributes to defense responses in arabidopsis. *The Plant cell* **2010**, *22*, 508-522.

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

52.   Pearce, G.; Ryan, C.A. Systemic signaling in tomato plants for defense against herbivores. Isolation and characterization of three novel defense-signaling glycopeptide hormones coded in a single precursor gene. *The Journal of biological chemistry* **2003**, *278*, 30044-30050.

53.   Okamura, Y.; Watari, M.; Jerud, E.S.; Young, D.W.; Ishizaka, S.T.; Rose, J.; Chow, J.C.; Strauss, J.F., 3rd. The extra domain a of fibronectin activates toll-like receptor 4. *The Journal of biological chemistry* **2001**, *276*, 10229-10233.

54.   Choi, H.W.; Manohar, M.; Manosalva, P.; Tian, M.; Moreau, M.; Klessig, D.F. Activation of plant innate immunity by extracellular high mobility group box 3 and its inhibition by salicylic acid. *PLoS pathogens* **2016**, *12*, e1005518.

55.   Lotze, M.T.; Tracey, K.J. High-mobility group box 1 protein (hmgb1): Nuclear weapon in the immune arsenal. *Nature reviews. Immunology* **2005**, *5*, 331-342.

56.   Lange, S.S.; Vasquez, K.M. Hmgb1: The jack-of-all-trades protein is a master DNA repair mechanic. *Molecular carcinogenesis* **2009**, *48*, 571-580.

57.   Ivanov, S.; Dragoi, A.M.; Wang, X.; Dallacosta, C.; Louten, J.; Musco, G.; Sitia, G.; Yap, G.S.; Wan, Y.; Biron, C.A., *et al.* A novel role for hmgb1 in tlr9-mediated inflammatory responses to cpg-DNA. *Blood* **2007**, *110*, 1970-1981.

58.   Aan den Toorn, M.; Albrecht, C.; de Vries, S. On the origin of serks: Bioinformatics analysis of the somatic embryogenesis receptor kinases. *Molecular plant* **2015**, *8*, 762-782.

59.   Sasaki, G.; Katoh, K.; Hirose, N.; Suga, H.; Kuma, K.; Miyata, T.; Su, Z.H. Multiple receptor-like kinase cdnas from liverwort marchantia polymorpha and two charophycean green algae, closterium ehrenbergii and nitella axillaris: Extensive gene duplications and gene shufflings in the early evolution of streptophytes. *Gene* **2007**, *401*, 135-144.

60.   Bowman, J.L.; Floyd, S.K.; Sakakibara, K. Green genes-comparative genomics of the green branch of life. *Cell* **2007**, *129*, 229-234.

61.   Hanks, S.K.; Quinn, A.M.; Hunter, T. The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains. *Science* **1988**, *241*, 42-52.

62.   Bottoms, C.A.; Smith, P.E.; Tanner, J.J. A structurally conserved water molecule in rossmann dinucleotide-binding domains. *Protein science : a publication of the Protein Society* **2002**, *11*, 2125-2137.

63.   Zheng, Z.; Goncearenco, A.; Berezovsky, I.N. Nucleotide binding database nbdb--a collection of sequence motifs with specific protein-ligand interactions. *Nucleic acids research* **2016**, *44*, D301-307.

64.   Carrera, A.C.; Alexandrov, K.; Roberts, T.M. The conserved lysine of the catalytic domain of protein kinases is actively involved in the phosphotransfer reaction and not required for anchoring atp. *Proceedings of the National Academy of Sciences of the United States of America* **1993**, *90*, 442-446.

65.   Denhez, F.; Galibert, F.; Heimann, B.; Dozier, C.; Graf, T.; Moelling, K.; Stehelin, D.; Ghysdael, J. An invariant asparagine residue belonging to a highly conserved domain in all protein kinases is instrumental in the protein kinase activity of the v-mil gene product. *Oncogene research* **1988**, *3*, 199-205.

66.   Ellis, L.; Clauser, E.; Morgan, D.O.; Edery, M.; Roth, R.A.; Rutter, W.J. Replacement of insulin receptor tyrosine residues 1162 and 1163 compromises insulin-stimulated kinase activity and uptake of 2-deoxyglucose. *Cell* **1986**, *45*, 721-732.

67.   Weinmaster, G.; Zoller, M.J.; Smith, M.; Hinze, E.; Pawson, T. Mutagenesis of fujinami sarcoma virus: Evidence that tyrosine phosphorylation of p130gag-fps modulates its biological activity. *Cell* **1984**, *37*, 559-568.

68.   Land, W.G. The role of damage-associated molecular patterns (damps) in human diseases: Part ii: Damps as diagnostics, prognostics and therapeutics in clinical medicine. *Sultan Qaboos University medical journal* **2015**, *15*, e157-170.

69.   Liston, A.; Masters, S.L. Homeostasis-altering molecular processes as mechanisms of inflammasome activation. *Nature reviews. Immunology* **2017**, *17*, 208-214.

70.   Sayers, E.W.; Barrett, T.; Benson, D.A.; Bolton, E.; Bryant, S.H.; Canese, K.; Chetvernin, V.; Church, D.M.; Dicuccio, M.; Federhen, S., *et al.* Database resources of the national center for biotechnology information. *Nucleic acids research* **2012**, *40*, D13-25.