



New Insights on Leucine-Rich Repeats Receptor-Like Kinase Orthologous Relationships in Angiosperms

Jean-François Dufayard¹, Mathilde Bettembourg¹, Iris Fischer², Gaetan Droc¹, Emmanuel Guiderdoni¹, Christophe Périn¹, Nathalie Chantret^{2*} and Anne Diévert^{1*}

¹ CIRAD, UMR AGAP, Montpellier, France, ² INRA, UMR AGAP, Montpellier, France

OPEN ACCESS

Edited by:

Eduard Akhunov,
Kansas State University, USA

Reviewed by:

Jianfei Zhao,
University of Pennsylvania, USA
Shichen Wang,
Texas A&M University, USA

*Correspondence:

Anne Diévert
anne.dievert@cirad.fr
Nathalie Chantret
nathalie.chantret@inra.fr

Specialty section:

This article was submitted to
Plant Genetics and Genomics,
a section of the journal
Frontiers in Plant Science

Received: 21 November 2016

Accepted: 06 March 2017

Published: 05 April 2017

Citation:

Dufayard J-F, Bettembourg M,
Fischer I, Droc G, Guiderdoni E,
Périn C, Chantret N and Diévert A
(2017) New Insights on Leucine-Rich
Repeats Receptor-Like Kinase
Orthologous Relationships
in Angiosperms.
Front. Plant Sci. 8:381.
doi: 10.3389/fpls.2017.00381

Leucine-Rich Repeats Receptor-Like Kinase (LRR-RLK) genes represent a large and complex gene family in plants, mainly involved in development and stress responses. These receptors are composed of an LRR-containing extracellular domain (ECD), a transmembrane domain (TM) and an intracellular kinase domain (KD). To provide new perspectives on functional analyses of these genes in model and non-model plant species, we performed a phylogenetic analysis on 8,360 LRR-RLK receptors in 31 angiosperm genomes (8 monocots and 23 dicots). We identified 101 orthologous groups (OGs) of genes being conserved among almost all monocot and dicot species analyzed. We observed that more than 10% of these OGs are absent in the Brassicaceae species studied. We show that the ECD structural features are not always conserved among orthologs, suggesting that functions may have diverged in some OG sets. Moreover, we looked at targets of positive selection footprints in 12 pairs of OGs and noticed that depending on the subgroups, positive selection occurred more frequently either in the ECDs or in the KDs.

Keywords: LRR, receptor, kinase, angiosperms, phylogeny, orthologs

INTRODUCTION

Receptor-like kinases constitute one of the largest gene families in the plant kingdom. They are typically composed of an amino-terminal ECD, a TM, and an intracellular domain (ICD) containing the KD. Several phylogenetic studies of the RLK family were conducted, initially focusing on *Arabidopsis* and later including other plant species (Shiu and Bleecker, 2001b, 2003; Shiu et al., 2004; Lehti-Shiu et al., 2009; Liu et al., 2009; Sakamoto et al., 2012; Zan et al., 2013). Using a phylogeny inferred from their KD alignment the *Arabidopsis* RLK genes were classified into 44 SGs or subfamilies (Shiu and Bleecker, 2001a). Fifteen SGs have been described containing common motifs in their ECD. The ECD of the largest SG possesses LRR and this SG has therefore been named LRR-RLK (Kobe and Deisenhofer, 1994; Kajava, 1998; Shiu and Bleecker, 2001a; Shiu et al., 2004; Lehti-Shiu et al., 2009). The first members of this large family were cloned in the 90s and their signaling pathways were extensively studied. Those members are ERECTA (ER), CLAVATA1 (CLV1),

Abbreviations: Cys-pair, cysteine-pair; ECD, extracellular domain; ICD, intracellular domain; KD, kinase domain; LRR, Leucine-Rich Repeat; LRR-RLK, Leucine-Rich Repeats Receptor-Like Kinase; MD, monocots dicots; MLD, malectin-like domain; OG, orthologous group; RLK, Receptor-Like Kinase; SG, subgroup; SP, signal peptide; TM, transmembrane domain.

BRASSINOSTEROID INSENSITIVE 1 (BRI1), SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE (SERK), HAESA-RLK5, and Xa21 (Horn and Walker, 1994; Song et al., 1995; Torii et al., 1996; Clark et al., 1997; Li and Chory, 1997; Schmidt et al., 1997). To date, functions have been assigned to ~35% of the ~230 LRR-RLK members in *A. thaliana* and – to a lesser extent – other species (Wu et al., 2016). They are important mediators of cell-cell communication to relay developmental cues and environmental stimuli or to activate defense/resistance against pathogens (Mu et al., 1994; Muschietti et al., 1998; Antolin-Llovera et al., 2014a; Belkhadir et al., 2014; Jaouannet et al., 2014).

Functional analyses conducted on *LRR-RLK* genes over the last twenty years revealed the role of the domains located in the ECD of these receptors. The LRR domains are highly versatile in number allowing a whole range of protein-protein interactions. These include homo- or hetero-dimerization of receptors, in addition to ligand binding. Furthermore, some LRR-RLK receptors possess island domains – devoid of LRRs – located between LRR motifs (Li and Chory, 1997). They were identified in the BRI1 receptor as the binding site for the brassinosteroid (BR) hormone (Kinoshita et al., 2005; Hothorn et al., 2011; She et al., 2011). Few studies have also described the functions of other ECD domains. For example, two Cys-pair have been reported. The first one is located in the N-terminal part of the LRR-RLKs, approximately 60 AA from the start codon between the SP and the first LRRs. The second one – if present – can be found between the last LRR and the TM domain (Dievart and Clark, 2003). Mutations in the Cys-pairs have been shown to affect the function of some LRR-RLKs, e.g., FLAGELLIN SENSING 2 (FLS2), a gene participating in the perception of the bacterial elicitor flagellin. However, there is also an example of a LRR receptor like protein (CLAVATA 2) for which mutations in Cys-pairs had no effect on the function of protein – at least in the meristem and roots (Noguchi et al., 1999; Song et al., 2010; Sun et al., 2012). In BRI1, a mutant harboring a mutation in Cys-pairs appears to be functional but seems to be retained in the endoplasmic reticulum and degraded. This suggests that this mutant protein does not pass the endoplasmic reticulum quality control (Hong et al., 2008). Although no general conclusions can be drawn so far on the importance of this motif, all the variations observed in Cys-pairs likely play a role in the folding, trafficking and/or the binding to other proteins. It was therefore suggested that this motif influences the signaling pathways activated downstream of the LRR-RLKs (Su et al., 2012). Another ECD, the MLD lying in between the SP and the LRRs, has also been described in one LRR-RLK SG (Hok et al., 2011). In legumes and actinorhizal plants, the SYMBIOSIS RECEPTOR LIKE KINASE (SYMRK, also known as NORK or DMI2) receptor, involved in phosphate-acquiring arbuscular mycorrhiza and in nitrogen-fixing root nodule symbiosis, possesses a lectin domain but the exact function of this receptor is still unclear (Antolin-Llovera et al., 2014a). It has been recently demonstrated that the SYMRK receptor is likely cleaved at the plasma membrane to release the N-glycosylated MLD (Antolin-Llovera et al., 2014b). Moreover, this cleavage would permit a physical interaction between the SYMRK and the LysM-type RLK NOD FACTOR RECEPTOR 5

and induces a rapid degradation of the SYMRK protein lacking its MLD. Thus, all the domains lying in the ECD with the LRRs play essential and complementary roles for specific LRR-RLK receptor functions.

Their central role in plant development and perception of environmental condition or stresses, their ubiquity in all angiosperms, and the complexity of their relationships make *LRR-RLK* genes an interesting candidate family to be studied in a phylogenetic framework (Shi et al., 2014). Such an analysis will be helpful to identify groups of orthologous genes and to compare functions between orthologs. However, inferring the phylogeny of such a large family raises several challenges. First, the vast number of sequences to be analyzed poses a problem of computational time and space. Second, the high rate of gene gains and losses during the evolution of the family, species-specific characteristics, and annotation errors result in complex orthologous relationships that are not always identified correctly by automatic gene annotation. For these reasons, large gene families – such as LRR-RLKs – are not well characterized on platforms like GreenPhylDB or Phytozome dedicated to automatic clustering (Conte et al., 2008; Rouard et al., 2011; Goodstein et al., 2012) and significant manual expertise is required to produce reliable results.

In the present article, we conducted a phylogenetic analysis of the *LRR-RLK* genes from 33 plant genomes with the objective to investigate the characteristics of genes belonging to the same OGs, expected to be conserved among most monocot and dicot species analyzed. To do so, we first looked for and identified 101 OGs of genes present in most genomes analyzed defined them as the LRR-RLK angiosperm “core” sets. We observed that ECD structural features were not always conserved in some OGs, suggesting that functions may have diverged among these orthologs. We also looked at selection footprints that led to the differentiations of pairs of OGs. This allowed us to investigate the putative role and function of uncharacterized genes in recently sequenced genomes from experimentally characterized *LRR-RLK* genes in model organisms.

MATERIALS AND METHODS

Plant Genomes Analyzed

The 33 species analyzed represent a broad spectrum of land plants (embryophyta), with one bryophyta genome, *Physcomitrella patens* (PHYPA, moss) (Rensing et al., 2008), the *Selaginella moellendorffii* genome (SELML, spikemoss) (Banks et al., 2011), representative of the lycopodiopsida, and 31 species of magnoliophyta (angiosperms), divided into eight monocot species (*Phoenix dactylifera* (PHODC, date palm) (Al-Dous et al., 2011), *Musa acuminata* (MUSAC, banana) (D’Hont et al., 2012), two subspecies of *Oryza sativa* (rice), *Oryza sativa* ssp. *japonica* (ORYSJ) (Goff et al., 2002) and *Oryza sativa* ssp. *indica* (ORYSI) (Yu et al., 2002; International Rice Genome Sequencing Project, 2005), *Brachypodium distachyon* (BRADI, purple false brome) (The International Brachypodium Initiative, 2010), *Zea mays* (MAIZE, corn) (Schnable et al., 2009), *Sorghum bicolor* (SORBI) (Paterson et al., 2009) and *Setaria italica*

(SETIT, Foxtail millet) (Bennetzen et al., 2012; Zhang et al., 2012), and 23 dicot species (two *Solanum* species, *Solanum tuberosum* (SOLTU, potato) (The Potato Genome Sequencing Consortium, 2011) and *Solanum lycopersicum* (SOLLIC, tomato) (Tomato Genome Consortium, 2012), *Vitis vinifera* (VITVI, Grape Vine) (Jaillon et al., 2007), *Lotus japonicus* (LOTJA) (Sato et al., 2008), *Medicago truncatula* (MEDTR, Barrel Medic) (Young et al., 2011), *Glycine max* (GLYMA, soybean) (Schmutz et al., 2010), *Cajanus cajan* (CAJCA, pigeon pea) (Varshney et al., 2012), *Prunus persica* (PRUPE, peach) (Ahmad et al., 2011), *Malus x domestica* (MALDO, apple) (Velasco et al., 2010; Jung et al., 2012), *Ricinus communis* (RICCO, castor oil plant) (Chan et al., 2010), *Jatropha curcas* (JATCU) (Sato et al., 2011), *Manihot esculenta* (MANES, Cassava) (Prochnik et al., 2012), *Populus trichocarpa* (POPTR, black cottonwood) (Tuskan et al., 2006), two *Cucumis* species, *Cucumis sativus* (CUCSA, cucumber) (Huang et al., 2009) and *Cucumis melo* (CUCME, Muskmelon) (Sebastian et al., 2010; Garcia-Mas et al., 2012), *Schrenkiella parvula* (SCHPA (formerly EUTPR)) (Dassanayake et al., 2011), *Eutrema salsugineum* (EUTSA (formerly THEHA)) (Oh et al., 2010), *Brassica rapa* (BRARA) (Wang et al., 2011), two *Arabidopsis* species, *Arabidopsis lyrata* (ARALY) and *Arabidopsis thaliana* (ARATH, thale cress) (The Arabidopsis Genome Initiative, 2000; Hu et al., 2011), *Carica papaya* (CARPA, papaya) (Ming et al., 2008), *Gossypium raimondii* (GOSRA, cotton) (Wang et al., 2012) and *Theobroma cacao* (THECC, cacao tree) (Argout et al., 2011). Note that the genus *Thellungiella* is now known as *Eutrema*, so the species formerly known as *Thellungiella halophila* (THEHA) is now known as *Eutrema salsugineum* (EUTSA). In addition, the species of *Eutrema* sequenced at the JGI has been determined to be *salsugineum* (a close relative of *halophila*). Therefore, this genome is actually classified as *Eutrema salsugineum*. Since we started our analysis before this change, all sequences related to the *Eutrema salsugineum* genome are annotated “THEHA” or “EUTSA” in our paper. Also, in the course of our study, the name *Eutrema parvula* (EUTPR) has been changed for *Schrenkiella parvula* (SCHPA). So, all *Schrenkiella parvula* (SCHPA) sequences are annotated “EUTPR” or “SCHPA”. See Supplementary Table S1 for bibliography and web links for download. Species tree representation (Supplementary Material) is based on several studies (Koch et al., 2000; Oh and Potter, 2005; Rensing et al., 2008; Forest and Chase, 2009; Magallcnd and Castillo, 2009; Wang et al., 2009; Oh et al., 2010; Sebastian et al., 2010; Arakaki et al., 2011; Reineke et al., 2011; Wang et al., 2011, 2012; Young et al., 2011; Tomato Genome Consortium, 2012; Varshney et al., 2012; Zhang et al., 2012).

LRR-RLKs Extraction, Phylogeny, and OGs

On each of the 33 plant proteomes, the hmmsearch program was run to extract peptide sequences containing both LRR(s) and a KD (Eddy, 2009). Sequences containing both LRRs and KD were classified into SGs using a global phylogenetic analysis (Fischer et al., 2016). First, the KD of all these sequences was aligned using MAFFT with a progressive strategy (Katoh et al., 2002). Then the alignment was cleaned with TrimAl configured to remove every sites with more than 20% of gaps

or with a similarity score lower than 0.001 (Capella-Gutierrez et al., 2009). A similarity matrix was computed using ProtDist with a JTT model, and then a global distance phylogeny was inferred using FastME configured with default settings and SPR movements to optimize the tree topology (Felsenstein, 1989; Desper and Gascuel, 2002). SGs were defined manually in the global phylogeny using the *Arabidopsis* genes as reference (Shiu et al., 2004; Lehti-Shiu et al., 2009; Fischer et al., 2016). To extend this dataset to receptor kinases devoid of LRRs in their ECD (sequences annotated “No_LRR”), the BLASTP algorithm (default parameters) has been run SG per SG, using each of the 7,767 KD sequences to search a database composed of the 33 proteomes as query (Altschul et al., 1997). Blast outputs were parsed to keep only homolog sequences sharing more than 90% identity with the query sequence. The new “No_LRR” sequences retrieved by blast were assigned to the same SG as the query sequence. Then, phylogenies were inferred for each of the 20 SGs. Each group of sequences was aligned using MAFFT with an iterative strategy (maximum of 100 iterations) (Katoh et al., 2002). Alignments were cleaned using TrimAl configured this time to remove sites with more than 80% of gaps (Capella-Gutierrez et al., 2009). Then maximum likelihood phylogenies were inferred using PhyML 3.0, configured with LG+gamma model, and the best of NNI and SPR topology optimization (Guindon and Gascuel, 2003). Statistical branch supports were computed using the aLRT/SH-like strategy (Guindon et al., 2010; Anisimova et al., 2011). Each of the 20 phylogenetic trees has been reconciled with the species tree using RAP-Green (Dufayard et al., 2005)¹. By comparing the gene tree with the species tree, this analysis allows us to root phylogenetic trees (Dufayard et al., 2005). We tested this approach of rooting (by minimizing the number of inferred duplications and losses) and compared it with rooting with outgroups (data not shown). The two methods provided very close root locations that did not change the overall conclusions.

To define the monocots dicots (MD) OGs, monocots/dicots bifurcations (branch support threshold >0.85) have been manually located in each of the 20 SG-specific trees. To be considered as MD OGs, the minimum number of monocots and dicots species represented was 3 and 4, respectively, to avoid keeping groups of misannotated sequences as MD OGs. Thus, 101 MD OGs have been defined. For each of them, the number of sequences in each of the 31 studied angiosperm species was recorded. If no sequence was discovered in one species, it was considered lost in this species.

Structural Features

Number of motifs and positions (LRRs (PF00560.24) and KDs (PF00069.16)) are outputs of the hmmsearch program (default parameters) (Eddy, 2009). Island domains were determined based on predicted LRR positions in sequences. For SP and TM domains, the TMHMM and TOPPRED softwares have been used (Claros and von Heijne, 1994; Krogh et al., 2001). For the malectin domains in SG_I and SG_VIII-2, sequences of the domains were extracted in SMART and aligned to build hmm

¹<https://github.com/SouthGreenPlatform/rap-green>

motifs with the hmmbuild program (Eddy, 2009; Letunic et al., 2009). For Cys-pairs, hmm motifs were built based on subsets of sequences known to possess these motifs (Eddy, 2009).

Test of Positive Selection on Ancestral Branches

Twelve sub-trees were considered: the OGs selected were those with a 'simple' organization, *i.e.* with a gene topology fitting approximately with the species tree. Three to four sequences among those that are the most closely related to the OG were selected from the whole SG tree as outgroup (Supplementary Figure S4). The sequences were re-aligned and the alignment was cleaned as described previously (Fischer et al., 2014, 2016). We ran codeml branch/site models implemented in the PAML4 software (Yang, 2007). For each OG, the following branch partition was defined: all branches but one were tagged as 'background' branches and the branch between the duplication node and the node corresponding to the split between monocot and dicot tagged as 'foreground' branch. Then two models were compared: the null model (A_0), in which sites on the foreground and background branches evolved under the same selective pressure (purifying or neutral), and a model including positive selection (model A) in which some sites on the foreground branch evolved under positive selection whereas sites on the background branches still evolved under purifying selection or neutrality. The most likely model was inferred by a likelihood ratio test (LRT). To take into account multiple testing, a Bonferroni correction was applied: the significance threshold of 0.05 was divided by the number of tested branches (24). Sites detected to be under positive selection at the codon level were manually curated for alignment quality and reliability. In branches identified to have evolved under positive selection, Bayes empirical Bayes was used to calculate the posterior probabilities at each codon and detect those under positive selection (*i.e.*, those with a posterior probability of $\omega > 1$ strictly above 95%).

RESULTS AND DISCUSSION

More than 200 LRR-RLK Genes on Average Per Angiosperm Species

We conducted a phylogenetic analysis of the *LRR-RLK* gene family in 33 fully sequenced plant genomes to classify them into SGs and to highlight and describe general characteristics of these LRR-RLK gene sets (Fischer et al., 2016). Briefly, besides 31 angiosperms genomes – represented by eight monocots (including six poaceae) and 23 dicots – one bryophyte genome of *Physcomitrella patens* (PHYPA, moss) and one lycopodiopsida genome of *Selaginella moellendorffii* (SELML, spikemoss) were included (Supplementary Table S1, see Section "Materials and Methods" for details and five-digit species code). As it has been done previously, we based our classification of *LRR-RLK* genes into SGs on the KD phylogeny (Shiu et al., 2004; Lehti-Shiu et al., 2009; Fischer et al., 2016). In our previous study (Fischer et al., 2016), the LRR-RLK dataset contained 7,767 sequences

possessing at least one LRR in their ECD. Since we focus on structural features of ECDs and presence/absence of the genes in LRR-RLK OGs in this present study, we included LRR-RLK homologues for which LRRs were completely lost or were degenerated in this new dataset. Thus, 593 sequences (prefixed "No_LRR") were added to the original set of 7,767 sequences which lead to a total of 8,360 sequences (Supplementary Table S2). Within each of the 20 SGs, KDs were aligned and SG-specific trees were obtained using a likelihood-based method (PHYML) (Figure 1A and Supplementary Material). Altogether, these *LRR-RLK* genes represent on average 0.71 and 0.66% of the monocot and dicot proteomes, respectively. Interestingly, in moss (PHYPA) and spikemoss (SELML), the proportions of LRR-RLKs per genome (0.41 and 0.36%, respectively) are approximately half the ratio observed in angiosperms. Likewise, the average number of LRR-RLK genes in angiosperms is 263.6, with 260.7 LRR-RLK proteins [± 20.2 (SE)] in dicots and 268.8 [± 18.2 (SE)] in monocots. In PHYPA and SELML, 134 and 81 *LRR-RLK* genes have been retrieved, respectively. There is no significant difference in the average number of *LRR-RLK* genes between monocots and dicots but the number almost doubled in most angiosperms compared to PHYPA and SELML. It has to be noted that in some genomes (*e.g.*, CARPA and LOTJA), the number of *LRR-RLK* genes is particularly low compared to other angiosperm genomes, suggesting that retention rates vary among genomes, and that many losses may have occurred in some genomes (Fischer et al., 2016). Nevertheless, our results highlight the fact that, after the first wave of expansion in early land plants (Embryophyta), a second large amplification occurred in angiosperm genomes which shaped the current LRR-RLK family size of more than 200 gene copies on average per genome.

Among the 20 SGs, SG_III, SG_XI, and SG_XIIa are the largest as they contain $\sim 50\%$ of the total number of *LRR-RLK* genes in the analyzed plant genomes (Figure 1A). The extensive expansions leading to their size do not follow the same amplification pattern (Figure 1B). This observation is highlighted in Figure 1B by the color code used for each species in the 20 SG-specific trees (with branches of monocots species in pink and red, and branches of dicots species in yellow, blue and green). Our results reveal that the high numbers of SG_XIIa genes is the consequence of many lineage-specific expansions (LSE) (See Fischer et al., 2016 for details). These LSEs are relatively recent as they can be observed in phyla as well as species-specific lineages. On the contrary, in SG_III and SG_XI, expansions occurred mainly before the early divergence of angiosperm lineages, even though LSEs can also be observed at different levels of resolution in the trees. Therefore, these numerous and diverse modes of expansions lead to complex paralogous and orthologous relationships.

101 OGs of Monocot and Dicot Genes Retained Along Angiosperm Evolution

With the aim of transferring functional annotation from well studied genes from model species to orthologous genes in other genomes, we first analyzed in depth orthologous relationships

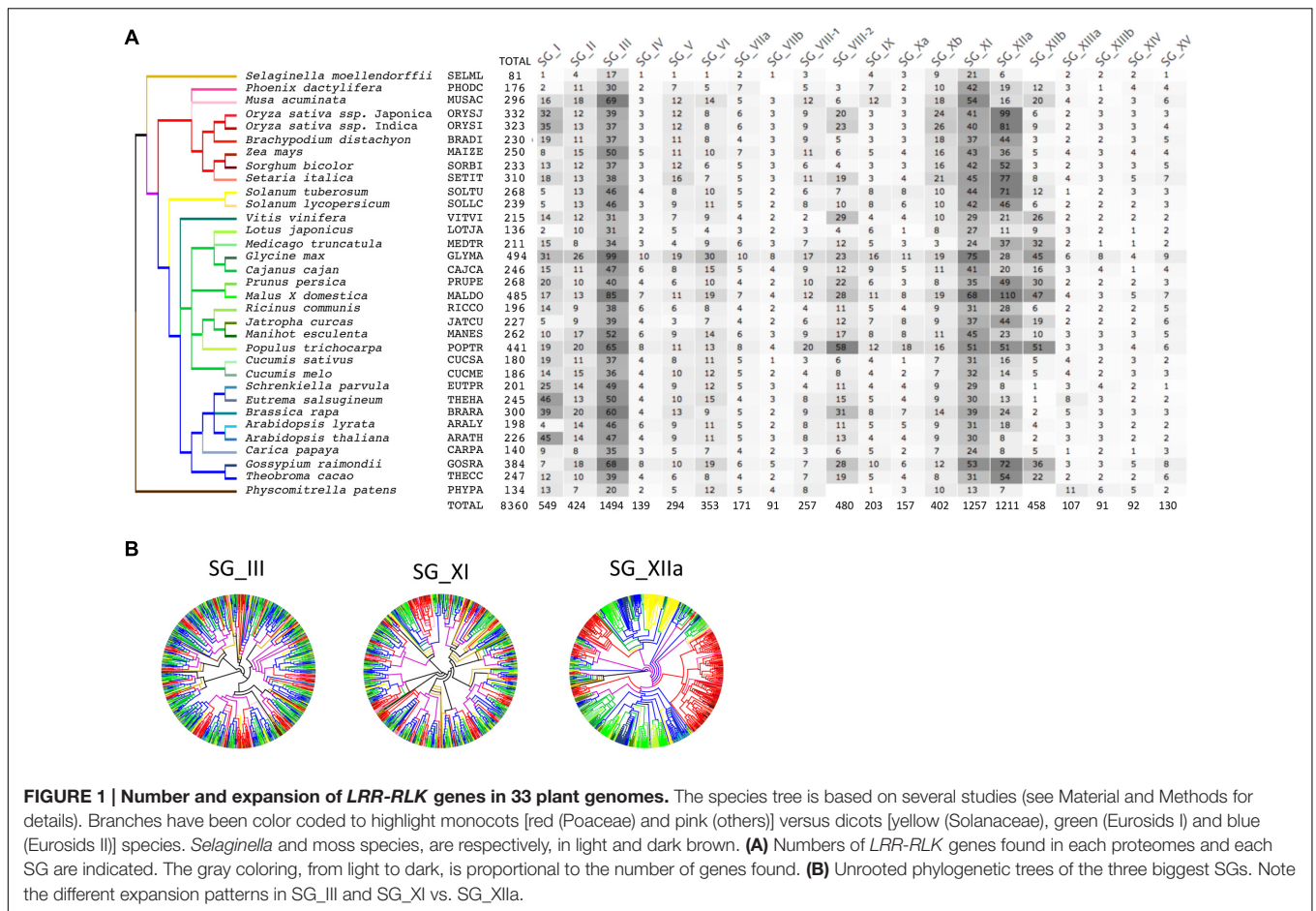
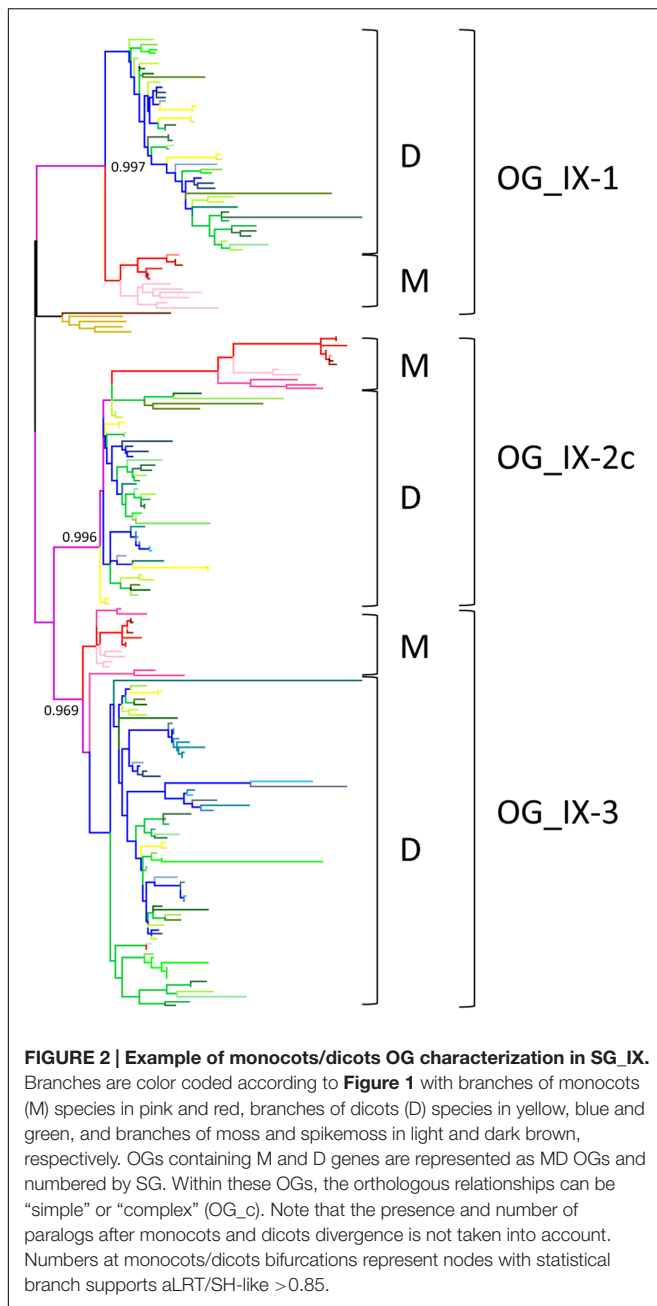


FIGURE 1 | Number and expansion of LRR-RLK genes in 33 plant genomes. The species tree is based on several studies (see Material and Methods for details). Branches have been color coded to highlight monocots [red (Poaceae) and pink (others)] versus dicots [yellow (Solanaceae), green (Eurosids I) and blue (Eurosids II)] species. *Selaginella* and moss species, are respectively, in light and dark brown. **(A)** Numbers of LRR-RLK genes found in each proteomes and each SG are indicated. The gray coloring, from light to dark, is proportional to the number of genes found. **(B)** Unrooted phylogenetic trees of the three biggest SGs. Note the different expansion patterns in SG_III and SG_XI vs. SG_XIIa.

between monocots and dicots LRR-RLK genes in each SG. This analysis led us to define what we named the “core set” of LRR-RLK genes in angiosperms: i.e., orthologous genes which have not been completely lost in either monocots or dicots throughout the angiosperms evolutionary history. To do so, the 20 SG-specific trees were scanned to locate monocots/dicots bifurcations (Figure 2). Based on this analysis, 101 OGs containing monocots and dicots sequences (named MD OGs) were characterized and defined as the “core” set of LRR-RLK genes in angiosperms.

The SG analysis revealed that these 101 MD OGs are present in 19 of the 20 SGs, with the majority of them in SG_III and SG_XI (Table 1). This highlights again the fact that most of the LRR-RLK genes which underwent expansions before the monocots/dicots split have been retained in these SGs. In order to go further in the description of orthologous relationships, we qualified OGs as either “simple” or “complex”. In “simple” OGs, the presence or absence of duplications within the monocot or dicot clades can clearly be inferred from the phylogenetic tree. On the other hand, if several duplications occurred disorderly with no obvious connection to the species tree we described these OGs as “complex” (OG_c). Interestingly, these OG_c are over-represented in SG_IV, VIII-1, VIII-2, and XI. A total of 6739 genes are contained in the 101 OGs, representing 82.7%

of the entire LRR-RLK gene family. However, while 2956 genes are included in the 24 OG_c (average of 123.2 genes per OG), 3783 genes belong to the 77 non-complex OGs (average of 49.2 genes per OG), highlighting differences in expansion/retention rates between these OGs. Moreover, looking at the percentage of genes contained in OGs per SG reveals that, except for SG_VIII-2 and XIb, more than 70% of the LRR-RLK genes belong to OGs, with genes mainly in complex OGs in SG_I, IV, VIII-1, VIII-2, XI, and XIIa (Figure 3A). In some SGs, some OGs contain a very large number of genes, such as 405 genes in one of SG_I OG (SG_I-3c), or 307 and 708 genes in two SG_XIIa OGs (Figure 3B and Supplementary Table S2 for details). In these large OGs, many species-specific duplications occurred, and among the 10 OGs containing more than 100 genes, 8 are complex. In SG_I-3c for example, several genes have been studied in *Arabidopsis*, such as IMPAIRED OOMYCETE SUSCEPTIBILITY 1 (IOS1), FLG22-INDUCED RECEPTOR-LIKE KINASE 1 (FRK1), and light-repressible receptor protein kinase (LRRPK) (Deeken and Kaldenhoff, 1997; Asai et al., 2002). All have been reported to be involved in abiotic and biotic responses in dicots but no gene from the same OG have been described so far in monocots. However, the fact that these genes are classified into the complex mode of expansion suggests that in monocots too, these genes could be involved in stress response.



More than 10% of the LRR-RLK Core Sets Are Absent in Brassicaceae

For each OG, we investigated whether some species were lacking members, focusing particularly on the Brassicales, for which six species are included in our analysis (Figure 4). Moreover, this clade contains the model plant *Arabidopsis* which is the reference for many studies on LRR-RLK functions (Wu et al., 2016). Interestingly, among the 101 OGs, 14 (13.8%) have been completely lost in the Brassicaceae, and 3 of them are even absent in all the Brassicales. This observation is an incentive to the extensive study of these receptors in other plants than *Arabidopsis*, adding an argument to the fact that functions or

TABLE 1 | Number of MD OGs per SG.

SG	Total number of MD OG	Number of OG_c
SG_I	3	1
SG_II	5	1
SG_III	24	3
SG_IV	1	1
SG_V	4	
SG_VI	5	
SG_VIIa	3	
SG_VIIb	2	
SG_VIII-1	3	3
SG_VIII-2	3	3
SG_IX	3	1
SG_Xa	2	
SG_Xb	7	
SG_XI	22	9
SG_XIIa	6	2
SG_XIIb	1	
SG_XIIIa	2	
SG_XIIIb	2	
SG_XIV	0	
SG_XV	3	
Total	101	24

interactions are sometimes phylum-specific. For example, the SG_I-2 OG contains the SYMBIOSIS RECEPTOR LIKE KINASE (SYMRK, also known as NORK or DMI2) receptor which is involved in actinorhizal plants and legumes, respectively, in phosphate-acquiring arbuscular mycorrhiza and nitrogen-fixing root nodule symbiosis (Antolin-Llovera et al., 2014a). For this gene, we noticed that besides being absent in the Brassicaceae, which do not form mycorrhizal associations and root nodule symbiosis with rhizobia, some other characteristics of these receptors have been observed in monocots (see below).

Inference of Functional Information from Experimentally Characterized LRR-RLK Genes to Uncharacterized Genes

The use of orthologous relationships to infer functional annotations relies on the fact that orthologs are expected to carry equivalent functions in different organisms. However, this can only be reliably inferred if, at least, structural characteristics and domain architecture are conserved between orthologs. In our analysis, the phylogeny of the LRR-RLK proteins was computed on the well conserved KDs. However, LRR-RLK sequences are composed of several domains and, notably, of LRRs in their ECD. One could wonder whether the domains belonging to genes of the same OG are conserved. First, we took a detailed look at the predicted number of LRRs of all these receptors. The number of LRR motifs per protein is an important feature for homo- and hetero-complex formation between LRR-RLKs (Macho and Zipfel, 2014). Second, we investigated the presence of island domains in between LRRs. These domains have been described to be the binding site for the BR hormone in some receptors (Kinoshita et al., 2005; Hothorn et al., 2011; She

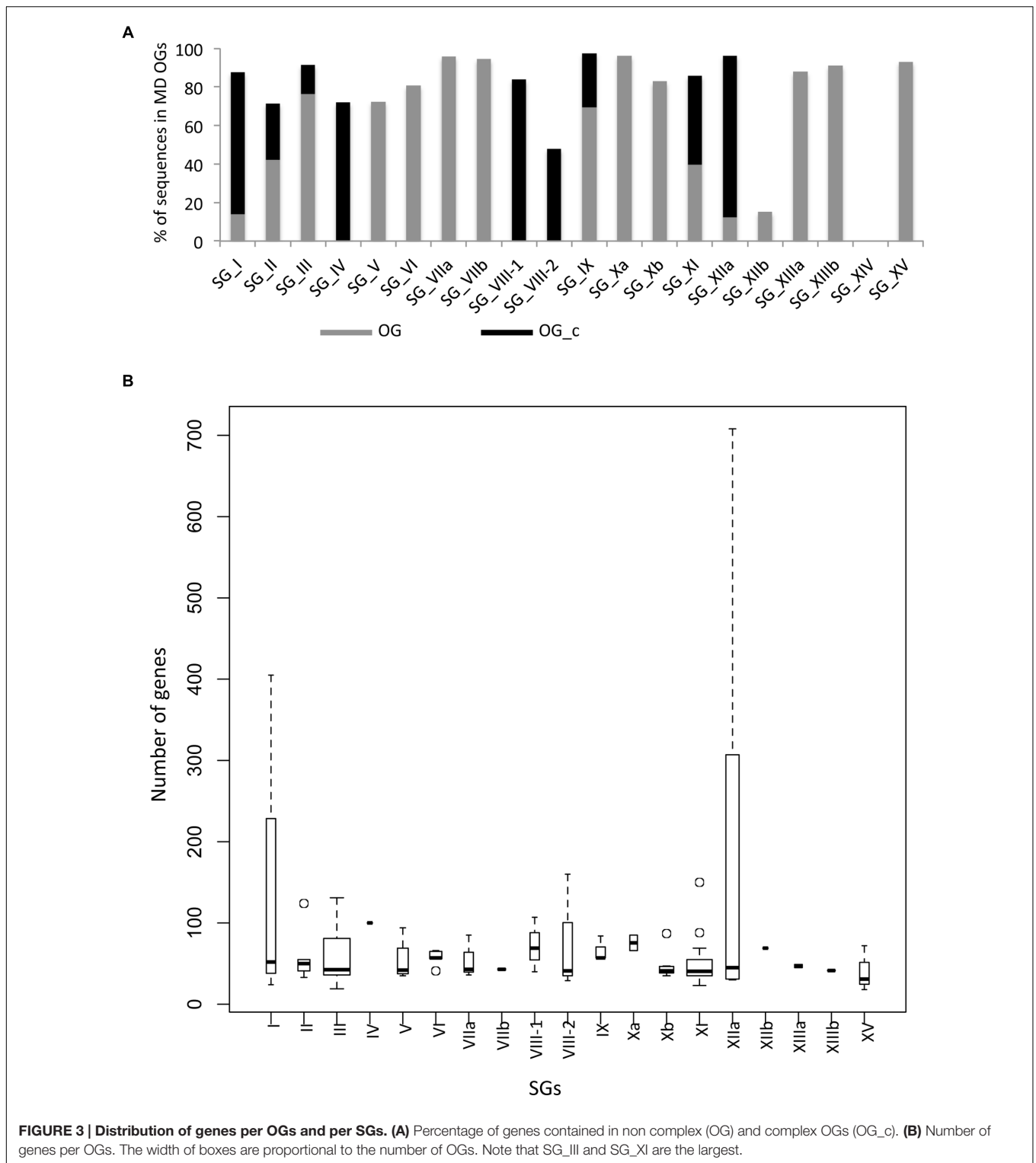


FIGURE 3 | Distribution of genes per OGs and per SGs. (A) Percentage of genes contained in non complex (OG) and complex OGs (OG_c). **(B)** Number of genes per OGs. The width of boxes are proportional to the number of OGs. Note that SG_III and SG_XI are the largest.

et al., 2011). Third, we analyzed the presence of the MLD, a carbohydrate-binding domain, and the GDPC, a protein cleavage motif, which were shown to be located before the LRRs in some SG_I receptors. Fourth, we investigated the presence of Cys-pairs surrounding the LRRs in some SGs. The presence

and organization of these domains is functionally important and has to be taken into account for transferring functional informations between orthologous genes. The description of structural features localized in the ECDs of these receptors allows a subclassification which, although reflecting the phylogeny of

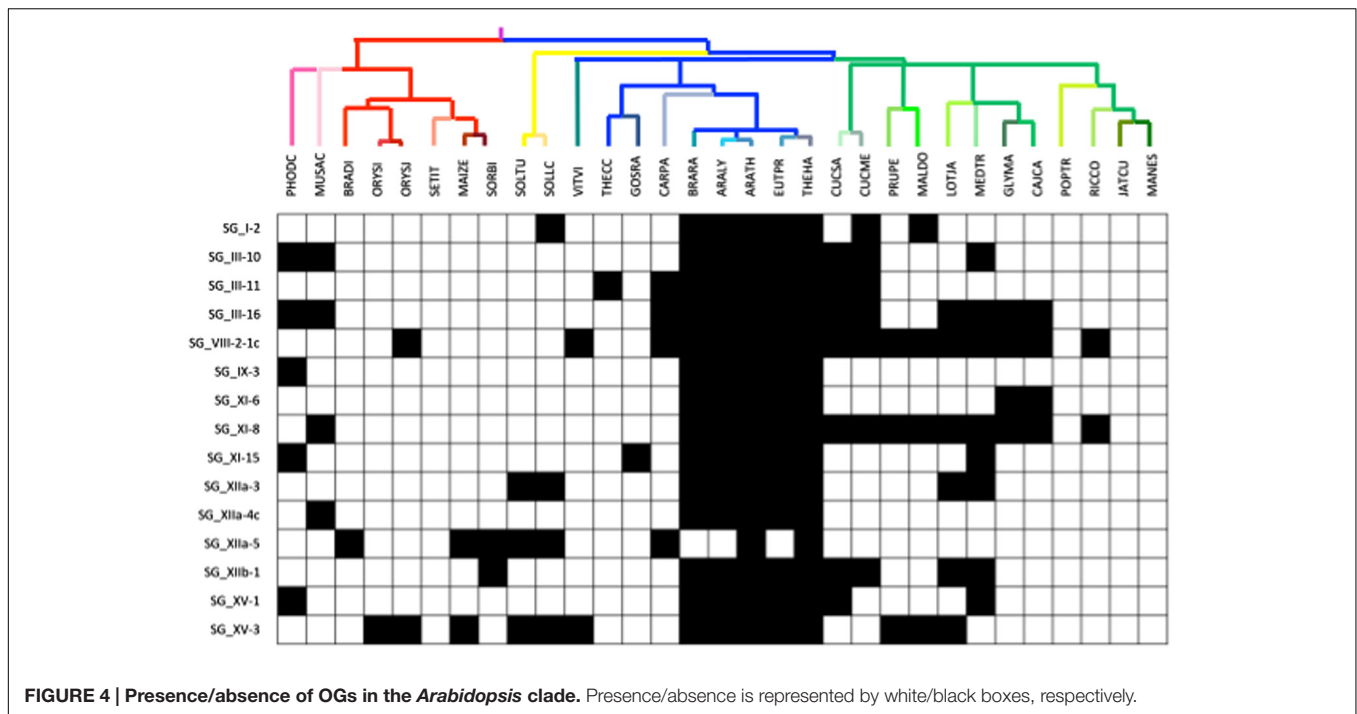


FIGURE 4 | Presence/absence of OGs in the *Arabidopsis* clade. Presence/absence is represented by white/black boxes, respectively.

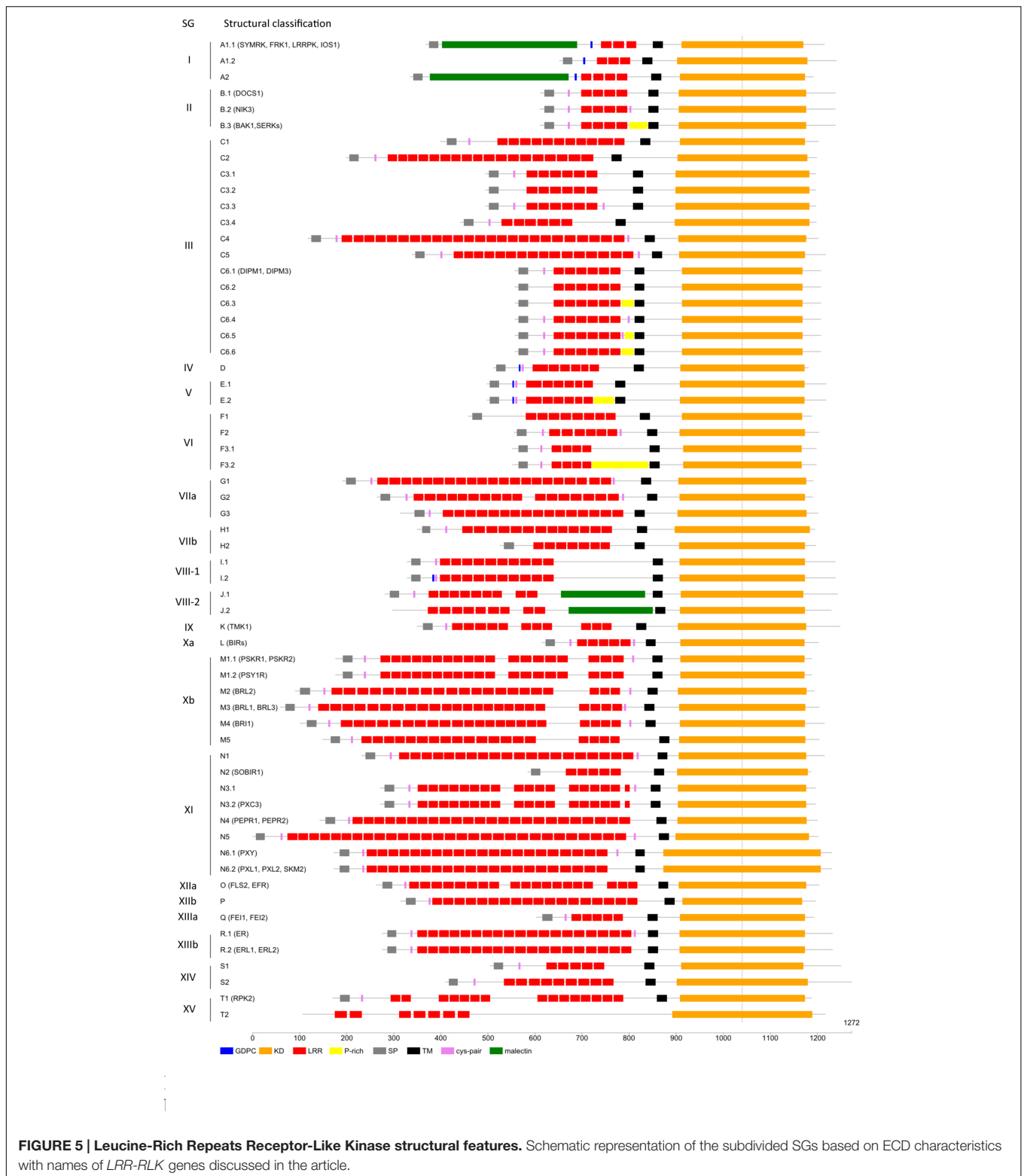
the KD of these receptors, also takes the structural differences of the ECD into account. Therefore, we subdivided the 20 SGs further according to these characteristics in their ECD (Figure 5 and Supplementary Table S3 for details).

Number of LRR Motifs

First, for the moss and/or spikemoss genes which are orthologous to the angiosperm core sets of genes, we investigated if some sets of receptors varied in the number of LRR motifs. All those genes have a common ancestor, predating the divergence between moss and/or spikemoss and angiosperms and their KDs have all evolved in concert for ~450 MYA. Despite speciation events, the close phylogenetic relationship of all these *LRR-RLK* genes with moss and spikemoss orthologs suggests that signaling pathways downstream of these receptors could be conserved. In the OG containing the FLS2 receptor (SG_XIIa), we noticed that the number of LRRs in the PHYPA orthologs was lower than in angiosperms (Figure 6A). This peculiar differences noted in the PHYPA ECDs of the FLS2 orthologs could affect ligand binding or even suggest that ligands are not conserved. This would be in agreement with publications stating that the moss *Physcomitrella patens* does not carry an FLS2 ortholog and also shows no response to flg22 (Boller and Felix, 2009; Tanigaki et al., 2014). All other core gene sets for which *Physcomitrella/Selaginella* ECDs are conserved compared to angiosperms, provide interesting cases for which it would be worth to verify if the functions described for monocots and/or dicots members are entirely conserved in bryophytes and lycopsids.

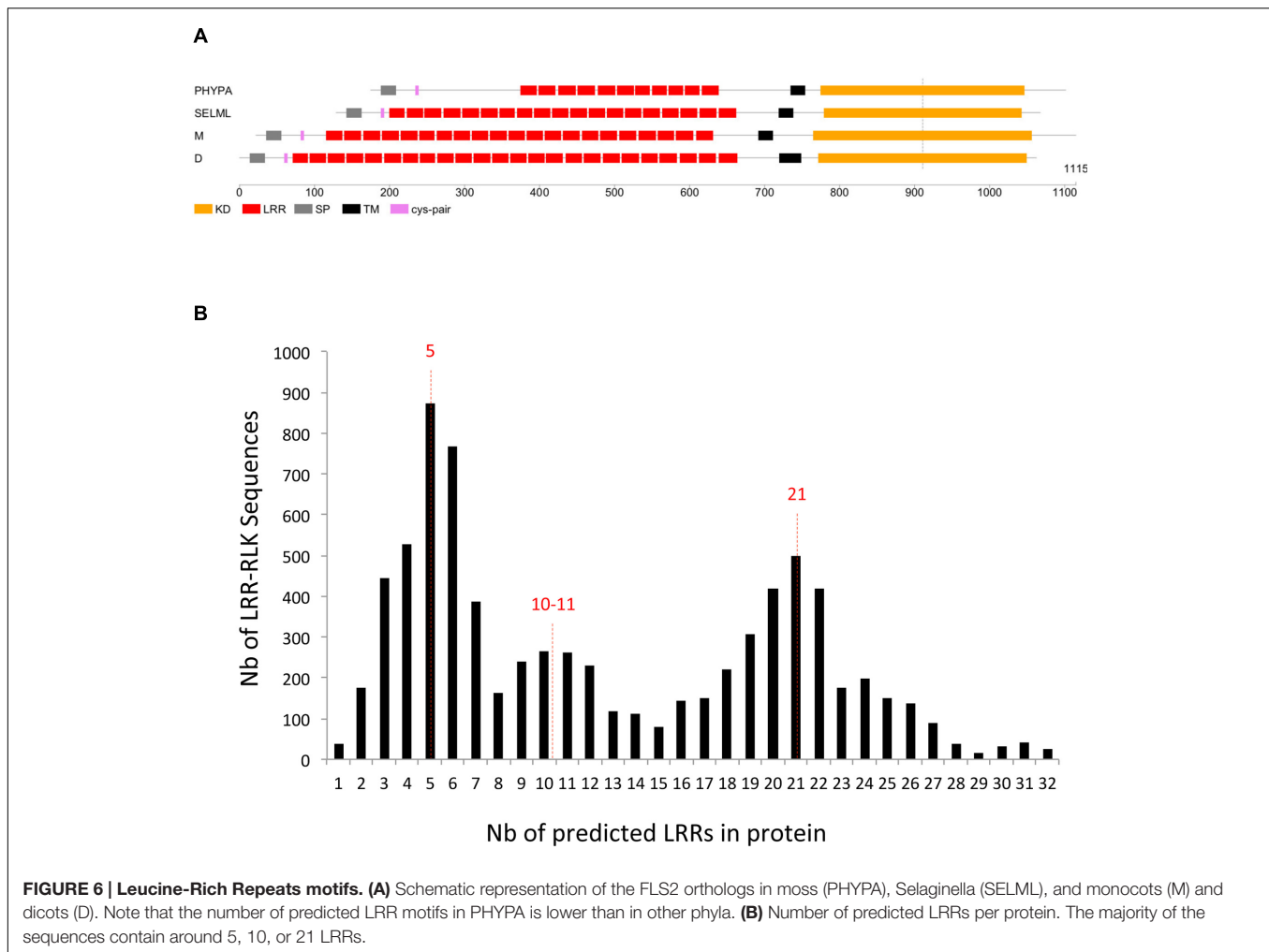
Second, we focused on the number of predicted LRRs in the 7,767 LRR-containing sequences. Even if the number of LRRs per sequence is very variable, the distribution of the number of LRR per sequences shows three peaks at 5, 20, or 21 (Figure 6B). This

observation suggests that these numbers of LRRs per sequences may be optimal for the 3D conformation of these receptors and their interactions in homo- or heterocomplexes. To our knowledge, this observation has not been explicitly made in animal LRR-containing proteins but could also be true (Ng et al., 2011). In plants, one hetero-oligomeric protein complex has been described between the BR receptor BRI1 (SG_Xb M4 in Figure 5) and BAK1/SERK3 or SERK1 receptors (SG_II B.3 in Figure 5) (Aker and de Vries, 2008; Chinchilla et al., 2009; Santiago et al., 2013; Sun et al., 2013). The complex crystal structure of SERK1 and BRI1 has revealed that the BRI1 C-terminal LRRs form a docking platform for the LRRs of the SERK1 co-receptor (Hothorn et al., 2011; Santiago et al., 2013). The SERK proteins have also been shown to serve various other BR-independent functions by forming heterocomplexes with SG_XIIa receptors (FLS2 and EFR, structure O in Figure 5) and the PEP1 RECEPTOR proteins (PEPR1, SG_XI N4 in Figure 5) (Chinchilla et al., 2007; Heese et al., 2007; Albrecht et al., 2008; Schulze et al., 2010; Roux et al., 2011; Koller and Bent, 2014). In rice, OsSERK2 (SG_II B.3 in Figure 5) forms a constitutive complex with the LRR-RLK Xa21 (SG_XIIa O in Figure 5) (Chen et al., 2014). Thus, these SERK co-receptors (4-5 LRRs) seem to play a central role in the regulation of multiple LRR-RLKs (>20 LRRs) by interacting directly with them (Aker and de Vries, 2008; Chinchilla et al., 2009; Kim et al., 2013; Santiago et al., 2013; Sun et al., 2013). Interestingly, SERK3/BAK1 has also been found in complex with the BAK1-INTERACTING RECEPTOR KINASE1 and 2 (BIR1 and BIR2) proteins, two receptors belonging to SG_Xa, another SG possessing five LRRs in its ECD (structure L in Figure 5) (Gao et al., 2009; Halter et al., 2014). The association between SERK3/BAK1 and BIR1/2 prevents the FLS2-BAK1 interaction before elicitation of the



immune response. It is still unknown if the co-receptor status is restricted to the SERK genes subfamily of SG_II B.3, or if other LRR-RLK SGs could also be part of various signaling pathways in interaction with receptors possessing 20-25 LRRs.

Indeed, other SGs such as SG_XI, SG_XIIIa, and SG_XIV corresponding to structures N2, Q and S1, respectively (Figure 5), also possess five LRRs in their ECD. In SG_XI-22 (structure N2 in Figure 5), the receptor SUPPRESSOR OF BIR1 1 (SOBIR1)



also named EVERSHED (EVR) has been shown to be involved in floral organ shedding and in the regulation of several resistance signaling pathways with the BIR1, BAK1, and FLS2 receptors (Gao et al., 2009; Leslie et al., 2010). Moreover, the SOBIR1 receptor is also described as a coreceptor/adaptor forming complexes with many LRR-Receptor like proteins (LRR receptors devoid of a KD), suggesting that the SERK-type receptors (5-LRRs) could be considered as general adaptors important for functionality in complex with their receptor partners (Liebrand et al., 2013; Gust and Felix, 2014). In SG_XIIIa (structure Q), the FEIs receptors (FEI1 and FEI2, named after the Chinese word for fat), whose single mutants were indistinguishable from the wild type in development, work against a co-receptor function (Xu et al., 2008). However, for other SGs possessing five LRRs in their ECD, the question about the putative co-receptor function will remain unanswered until further molecular characterization is performed. The receptors belonging to SG_II B.3, contrary to SG_II B.1 and B.2, contain also a Pro-rich motif in their ECD. The question on the functionality of the Pro-rich domain in the ECD of the SERK proteins also remains to be answered. This domain of unknown function could provide a flexible hinge to the ECD (Schmidt et al.,

1997; Kay et al., 2000; Baudino et al., 2001; Hecht et al., 2001; Chevalier et al., 2005). Interestingly, other SGs possess these kinds of motifs, e.g., SG_VI F3.2, for which no receptor has been studied yet.

Island Domains

We positioned all the predicted LRRs on the 7,767 proteins and searched for islands between them (Supplementary Table S4). These islands are of particular importance as they are the brassinolide hormone binding sites for the BRI1 and BRI1-like (BRL) receptors (structures M2–M4 in **Figure 5**) belonging to SG_Xb (Hothorn et al., 2011; She et al., 2011). We found that in SG_Xb, all sets of orthologs possess an island encompassing two (Structures M1 in **Figure 5**) or at least three (structures M2–M5) LRRs. One could ask if the island domain in genes of structure M5, for which no function has been described up to now and very similar to the M2–M4 structures, is also a binding site for the BR hormone. The remaining OGs in SG_Xb (structure M1) contain the three *Arabidopsis* genes: *PSKR1*, *PSKR2*, and *PLANT PEPTIDE CONTAINING SULFATED TYROSINE 1 RECEPTOR (PSY1R)*. These receptors have overlapping functions in promoting cellular proliferation,

longevity and expansion (Matsubayashi et al., 2006; Amano et al., 2007; Hartmann et al., 2013). The PSKR subfamily is also required for PSK peptide signaling in sexual reproduction in plants (Stuhrwoldt et al., 2015). Moreover, these receptors play a role in modifying responses to biotic pathogens and wounding (Loivamaki et al., 2010; Mosher et al., 2013). In the three *Arabidopsis* PSKRs, one island of ~60 AA was detected in addition to other smaller ones specific to each receptor. These islands could be important for hormone binding like in BRI1 and BRL receptors (Hothorn et al., 2011; She et al., 2011). Indeed, it has been shown that the BR hormone could play a role in the signaling pathways activated downstream of these receptors (Hartmann et al., 2013). In SG_IX (structure K) and SG_XV (structure T1), islands encompassing the size of at least two LRRs are also present. In SG_IX, the crystal structure of the *Arabidopsis* TRANSMEMBRANE KINASE 1 [TMK1, also known as BLK1 (BARK1-like Kinase 1)] suggests that the islands could be critical for structural integrity. In SG_XV, the crystal structure of RECEPTOR-LIKE PROTEIN KINASE 2 (RPK2, also known as TOAD2) suggests that the islands could be the site for ligand binding as in BRI1 (Liu et al., 2013; Song et al., 2014).

Additional Domains

As mentioned previously, a MLD lying in between the SP and the LRRs has been described in some SG_I receptors (Figure 5) (Hok et al., 2011; Antolin-Llovera et al., 2014a). One of them is the SYMRK receptor (structure A1.1) involved in mycorrhizal associations and rhizobium-legumes symbiosis, but its exact function is still unclear (Antolin-Llovera et al., 2014a). Recently, it has been demonstrated that the SYMRK receptor is cleaved at a GDPC motif placed at the end of the MLD to release the N-glycosylated ectodomain in the absence of symbiotic stimulation (Antolin-Llovera et al., 2014b). Moreover, protein cleavage on this motif would permit a physical interaction with the LysM-type RLK NOD FACTOR RECEPTOR 5 (NFR5) and induces a rapid degradation of the SYMRK protein lacking its MLD. In this form, SYMRK could act as a co-receptor to initiate symbiotic signaling with NFR5 and would mirror the role played by the receptors of the SERK family. After MLD release, the structure of SYMRK could indeed resemble that of BAK1/SERK3. Our analysis of structural features of ECDs reveals that the monocot orthologs of the SYMRK receptor were much shorter than the dicot ones and that the monocot receptors are devoid of the MLD present in dicots (structure A1.2). Thus, the SYMRK activation mechanism will have to be further investigated in monocots to evaluate if it can fit into the dicots model, or if other receptors possessing a MLD (SG_I or others, see below) are involved in this process.

We therefore looked for malectin domains in all LRR-RLK sequences and found that SG_VIII-2 receptors also contain one. However, in SG_VIII-2, this domain is not located just after the SP but between the LRRs and the TM domain (Structure J in Figure 5). We also found the GDPC cleavage motif in most of the SG_IV, V and SG_VIII-1 receptors. However, contrary to SG_I, the GDPC site is located just before the first Cys-pairs in

all other SGs. In these SGs, the Cys of the GDPC motif is the first site of the Cys-pair. It is still unknown if these receptors are also cleaved at this site and what the functional consequences would be. In SG_VIII-2, which contains a malectin domain C-terminal of the LRRs, no GDPC sites are present. This does not exclude the possibility that another cleavage site could be used to truncate the protein. Thus, the function of the malectin domains in SG_VIII-2 will have to be explored in the future to decipher their exact functional role and their potential involvement in protein stabilization.

Positive Selection in the Divergence between Ancestrally Duplicated OGs

Twelve pairs of MD OGs present in almost all monocot and dicot species, and harboring a gene topology fitting approximately the species tree, appear to be issued from ancestral duplications predating the monocot/dicot divergence (Table 2). As these OG pairs had a similar ECD structural organization and were kept in almost all species studied here, we searched for potential positive selection footprints in the divergence leading to their differentiation. Although these genes do not all have a known function, it is expected that they underwent amino-acid changes leading to their sub- and/or neo-functionalization, and one can wonder if and how positive selection could have driven these changes.

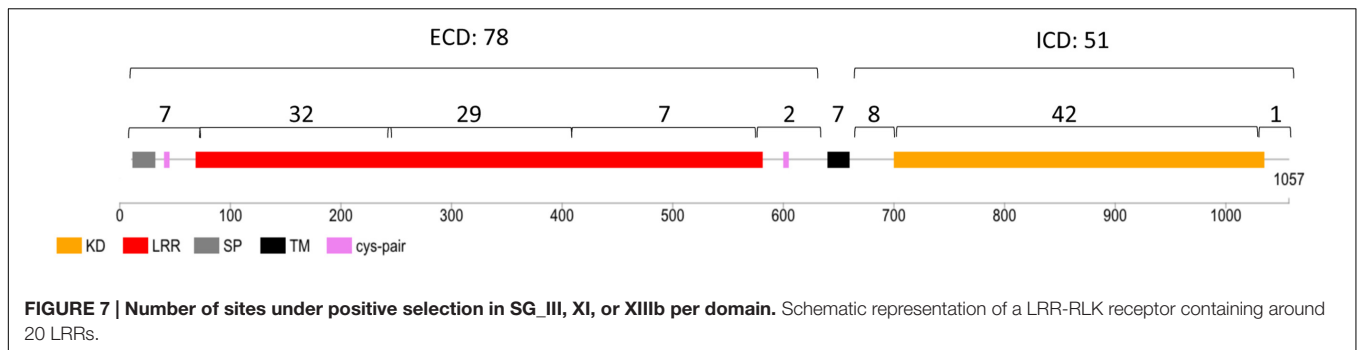
To answer this question, we tested whether some sites underwent positive selection on the two branches starting from the ancestral duplication and ending at the monocot/dicot divergence node of each OG. The detailed results of this analysis are presented in Table 2 and Supplementary Table S5. Two pairs showed no signal on either of the two branches (SG_Xa-1/2, SG_XIIa-2/3). For the pairs SG_II-3/4 and SG_XIIIa-1/2, a signal was detected for one branch only but the signal on SG_XIIIa-1/2 may be a false positive or the sign of a lack of power, since no sites appeared to be significant (see Materials and Methods for details). The eight other pairs showed a signal of positive selection on each branch. Although the model indicating positive selection performs significantly better than the null model, two pairs (SG_III-3/4 and SG_XI-19/20) have no significant sites for one of the two tested branches. This again indicates either a false positive or a lack of power. It is also possible that positive selection acted on a large number of sites which results in none of them exceeding the significance threshold. Finally, five pairs (SG_III-8/9, SG_XI-8/9, SG_XI-14/15, SG_XI-17/18, and SG_XIIIb-1/2) have a strong signal with up to 26 sites validated after manual curation. This result shows that in about half of the tested cases, several amino acid changes fixed in the divergence between these genes are compatible with a signal of positive selection.

A total of 141 sites were manually validated as having experienced an episode of positive selection during MD OG genes divergence. For the five pairs with a strong signal, the repartition of these sites across the different domains of the LRR-RLK protein showed that the LRRs and KDs are the most affected (Figure 7). More than half of the sites (78) fall in the ECD, among which 68 are in the LRR domain; 51 sites fall in the ICD, among which 42 are in the KD (Table 2). Considering that these LRR and KD

TABLE 2 | Positive selection on branches of pairs of OGs.

SG	Number of sequences	Name of branch	Structural SG	Known genes	Model	Number of parameters	InL	P-value		Number of sites	ECD	ICD
SG_II	101	SG_II-3	B.1	DOCS1	A ₀	203	-47756.49038	0.2960935	ns	—		
					A	204	-47755.94453					
SG_III	92	SG_III-4	B.2	NIK3	A ₀	203	-47761.1425	0.0012474	*	3	3	
					A	204	-47755.93314					
					A ₀	185	-87752.00789	0.0013461	*	1	1	
SG_III	92	SG_III-3	C2		A	186	-87746.86878					
					A ₀	185	-87749.51178	0.0000735	**	0		
					A	186	-87741.65272					
SG_III	126	SG_III-9	C6.1	DIPM1, DIPM3	A ₀	253	-58707.08864	0.000000	**	11	4	7
					A	254	-58683.61908					
SG_Xa	145	SG_Xa-1	L	BIR1	A ₀	291	-80168.41936	0.013983288	ns	—		
					A	292	-80165.39924					
					A ₀	291	-80166.7335	0.031737545	ns	—		
SGXb	162	SG_Xb-5	M1.1	PSKR2	A	292	-80164.42719					
					A ₀	325	-148131.1793	0.000381394	**	1	1	
					A	326	-148124.8687					
SG_XI	99	SG_Xb-6	M1.1	PSKR1	A ₀	325	-148131.1246	0.000004	**	1	1	
					A	326	-148120.4449					
					A ₀	199	-40840.3843	0.00030735	**	0		
SG_XI	99	SG_XI-20	N3.2	PXC3	A	200	-40833.87176					
					A ₀	199	-40838.95437	0.000838	*	1	1	
					A	200	-40833.37719					
SG_XI	116	SG_XI-9	N4	PEPR1, PEPR2	A ₀	233	-80676.72267	9.80354E-09	**	12	8	4
					A	234	-80660.28275					
SG_XI	115	SG_XI-8	N4		A ₀	233	-80679.22227	0.000000	**	9	3	6
					A	234	-80660.42399					
					A ₀	231	-126354.8399	3.48134E-19	**	20	12	6
SG_XI	115	SG_XI-14	N6.2	SKM2	A	232	-126314.768					
					A ₀	231	-126374.0677	0.000000	**	12	8	3
					A	232	-126335.0153					
SG_XI	132	SG_XI-15	N6.1		A ₀	265	-131270.9444	1.13374E-34	**	5	4	1
					A	266	-131195.5224					
SG_XI	132	SG_XI-18	N6.1	PXY	A ₀	265	-131270.9444	1.13374E-34	**	5	4	1
					A	266	-131195.5224					
					A ₀	265	-131261.1148	2.96009E-40	**	23	14	9
SG_XIIa	101	SG_XI-17	N6.2	PXL1, PXL2	A	266	-131172.9143					
					A ₀	203	-69490.32865	0.020062	ns	—		
					A	204	-69487.62542					
SG_XIIa	101	SG_XIIa-2	O		A ₀	203	-69490.53986	0.047369	ns	—		
					A	204	-69488.57373					
					A ₀	203	-69490.53986	0.047369	ns	—		
SG_XIIIa	143	SG_XIIa-3	O		A	204	-69488.57373					
					A ₀	287	-66690.11714	0.006790	ns	—		
					A	288	-66686.45331					
SG_XIIIa	143	SG_XIIIa-2	Q		A ₀	287	-66684.64365	0.000737	*	0		
					A	288	-66678.94678					
					A ₀	287	-66684.64365	0.000737	*	0		
SG_XIIIb	143	SG_XIIIa-1	Q	FEI1, FEI2	A	288	-66678.94678					
					A ₀	287	-66684.64365	0.000737	*	0		
SG_XIIIb	143	SG_XIIIb-1	R.2	ERL1, ERL2	A ₀	161	-79161.61197	3.21923E-09	**	26	14	8
					A	162	-79144.08871					
					A ₀	161	-79156.50926	0.000012	**	11	8	3
SG_XIIIb	143	SG_XIIIb-2	R.1	ER	A	162	-79146.91448					
					A ₀	161	-79156.50926	0.000012	**	11	8	3

InL, log-likelihood; P-values, ns, Likelihood ratio test not significant; *, significant at 5%-level; **, significant at 1%-level after the Bonferroni correction (see Materials and Methods).



are the largest domains, we normalized the number of positively selected sites by domain size. Kinase and LRR appeared then to be affected equally (Chi-square test, $p = 0.26$) by positive selection. This result is very different from positive selection signatures observed in the recent paralogs, for which LRR is the most strongly affected domain (Fischer et al., 2016). The number of sites laying in the LRR domain allowed us to look for any specific distribution across the 24 amino acids composing the motif. The repartition of the sites affected by positive selection within the LRR is not homogeneous and the majority of them (67) fall in the 13 non-canonical positions (Supplementary Table S5). However no notable pattern emerges from their distribution (data not shown). Again, this result contrasts with what is observed in lineage-specific expanded genes for which four positions are predominantly affected (Fischer et al., 2016). The remaining domains are affected by a number of sites varying from 1 to 9.

This approach revealed a prevalence of sites targeted by positive selection in the ECD for the three couples of genes belonging to SG_XI as well as SG_XIIIb. On the opposite, a tendency to target ICDs can be observed for SG_III. Two pairs of OGs for which positive selection footprints are detected in the ECD correspond to OGs whose duplication gave rise to the PXY and PXLs clades in SG_XI-17/18 (Fisher and Turner, 2007; Jung et al., 2015); and to the ERECTA (ER) and ERECTA-like (ERL) clades in SG_XIIIb-1/2 (Torii et al., 1996; Sanchez-Rodriguez et al., 2009). Other pairs of OGs concern differentiation of the PEPR1 and 2 clade (SG_XI-9), of the STERILITY-REGULATING KINASE MEMBER 2 (SKM2) gene (SG_XI-14) or of the DspA/E-interacting protein of *Malus x domestica* Borkh 1 and 3 (DIPM1 and 3) clade (SG_III-9), from their respective sister clades, SG_XI-8, SG_XI-15, and SG_III-8 (Meng et al., 2006; Krol et al., 2010; Kang and Hardtke, 2016). In these clades, no gene has been described yet. The strong signal of positive selection detected for these five groups of genes indicates that the divergence between ancestral copies may have procured a selective advantage: in the domain involved in ligands or partners binding for ECDs, or in the domain affecting downstream signaling pathways for ICDs. Indeed, during the early expansion of LRR-RLK that took place before angiosperm split, some duplicated LRR-RLK differentiated by fixation of a higher number of non-synonymous than synonymous mutations at some amino acid sites, indicating the emergence of probably new advantageous functions.

CONCLUSION

In this report, we provide a framework to aid in the classification and give new insights to new prospects for functional analysis of some plant LRR-RLKs. We have defined the “core set” of the large *LRR-RLK* gene family and classified these receptors based on their ECD features. These analyses reveal that even if the KDs of the LRR-RLKs are phylogenetically related, the ECDs may have been subjected to major (e.g., loss of LRRs revealed by the structural features characterization) or minor (e.g., point mutations revealed by the traces of positive selection analysis) modifications during the evolution of orthologs. These alterations could affect ligand recognition sites, dimerization with other receptors, and/or other processes involved in signal transduction. Indeed, the proper signal transduction via receptor kinases is not restricted to the binding of ligands to receptors located at the plasma membrane. Tightly regulated steps for proper folding of the proteins, trafficking from endomembranes to plasma membranes, and finally internalization and recycling of the receptors after ligand binding play essential roles in signal transduction (Shah et al., 2002; Robatzek et al., 2006; Salomon and Robatzek, 2006; Irani and Russinova, 2009; Beck et al., 2012; Di Rubbo et al., 2013; Offringa and Huang, 2013; Martins et al., 2015). Recently, an enthusiastic wave swept over the plant receptor kinases community concerning endoplasmic reticulum quality control since most of these steps take place in this cellular compartment (Saijo, 2010; Su et al., 2011; Huttner and Strasser, 2012; Tintor and Saijo, 2014). Newly synthesized membrane-resident proteins translocate first into the endoplasmic reticulum where they are subjected to folding and modifications like formation of disulfide bridges. It is also the place where nascent polypeptides are glycosylated – the most common post-translational modification which is a crucial event during protein folding and quality control processes (Bieberich, 2014). The LRR-RLKs are part of the large family of plant proteins which are N-glycosylated and many N-glycosylation acceptor sequences are present in all ECDs. In some pattern recognition receptors and receptors involved in developmental processes, proteins with mutations at residues which will create misfolded proteins have been shown to be part of endoplasmic reticulum protein complexes and directed to degradation (Hong et al., 2008, 2009, 2012; Li et al., 2009; Nekrasov et al., 2009; Lee et al., 2011; Su et al., 2011; Huttner and Strasser, 2012; Sun et al., 2012; Park et al., 2013;

Robatzek and Wirthmueller, 2013; Huttner et al., 2014). The significance of all the structural feature modifications which have been mentioned above are still mostly unknown but classic biochemical and cell biological studies (e.g., domain swapping among orthologs and/or targeted point mutations using CRISPR/Cas9) should help to explore their functions in details and will provide many novel insights into the molecular characterization of LRR-RLKs.

AUTHOR CONTRIBUTIONS

NC, CP, EG, and AD designed the study; GD and AD performed the LRR-RLK extraction; J-FD and AD performed the phylogenetic clustering; NC and IF performed the selection footprint analysis; IF, NC, J-FD, MB, and AD analyzed the data; J-FD, AD, NC, and IF wrote the article.

REFERENCES

- Ahmad, R., Parfitt, D. E., Fass, J., Oguniwin, E., Dhinra, A., Gradziel, T. M., et al. (2011). Whole genome sequencing of peach (*Prunus persica* L.) for SNP identification and selection. *BMC Genomics* 12:569. doi: 10.1186/1471-2164-12-569
- Aker, J., and de Vries, S. C. (2008). Plasma membrane receptor complexes. *Plant Physiol.* 147, 1560–1564. doi: 10.1104/pp.108.120501
- Albrecht, C., Russinova, E., Kemmerling, B., Kwaaitaal, M., and De Vries, S. C. (2008). Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASE proteins serve brassinosteroid-dependent and -independent signaling pathways. *Plant Physiol.* 148, 611–619. doi: 10.1104/pp.108.123216
- Al-Dous, E. K., George, B., Al-Mahmoud, M. E., Al-Jaber, M. Y., Wang, H., Salameh, Y. M., et al. (2011). De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nat. Biotechnol.* 29, 521–527. doi: 10.1038/nbt.1860
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Amano, Y., Tsubouchi, H., Shinohara, H., Ogawa, M., and Matsubayashi, Y. (2007). Tyrosine-sulfated glycopeptide involved in cellular proliferation and expansion in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18333–18338. doi: 10.1073/pnas.0706403104
- Anisimova, M., Gil, M., Dufayard, J. F., Dessimoz, C., and Gascuel, O. (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst. Biol.* 60, 685–699. doi: 10.1093/sysbio/syr041
- Antolin-Llovera, M., Petutsching, E. K., Ried, M. K., Lipka, V., Nurnberger, T., Robatzek, S., et al. (2014a). Knowing your friends and foes—plant receptor-like kinases as initiators of symbiosis or defence. *New Phytol.* 204, 791–802. doi: 10.1111/nph.13117
- Antolin-Llovera, M., Ried, M. K., and Parniske, M. (2014b). Cleavage of the SYMBIOSIS RECEPTOR-LIKE KINASE ectodomain promotes complex formation with Nod factor receptor 5. *Curr. Biol.* 24, 422–427. doi: 10.1016/j.cub.2013.12.053
- Arakaki, M., Christin, P. A., Nyffeler, R., Lendel, A., Egli, U., Ogburn, R. M., et al. (2011). Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8379–8384. doi: 10.1073/pnas.1100628108
- Argout, X., Salse, J., Aury, J.-M., Gaultinan, M. J., Droc, G., Gouzy, J., et al. (2011). The genome of *Theobroma cacao*. *Nat. Genet.* 43, 101–108. doi: 10.1038/ng.736
- Asai, T., Tena, G., Plotnikova, J., Willmann, M. R., Chiu, W. L., Gomez-Gomez, L., et al. (2002). MAP kinase signalling cascade in *Arabidopsis* innate immunity. *Nature* 415, 977–983. doi: 10.1038/415977a
- Banks, J. A., Nishiyama, T., Hasebe, M., Bowman, J. L., Gribskov, M., Depamphilis, C., et al. (2011). The selaginella genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332, 960–963. doi: 10.1126/science.1203810
- Baudino, S., Hansen, S., Brettschneider, R., Hecht, V. F., Dresselhaus, T., Lorz, H., et al. (2001). Molecular characterisation of two novel maize LRR receptor-like kinases, which belong to the SERK gene family. *Planta* 213, 1–10. doi: 10.1007/s004250000471
- Beck, M., Zhou, J., Faulkner, C., Maclean, D., and Robatzek, S. (2012). Spatio-temporal cellular dynamics of the *Arabidopsis* flagellin receptor reveal activation status-dependent endosomal sorting. *Plant Cell* 24, 4205–4219. doi: 10.1105/tpc.112.100263
- Belkhadir, Y., Yang, L., Hetzel, J., Dangl, J. L., and Chory, J. (2014). The growth-defense pivot: crisis management in plants mediated by LRR-RK surface receptors. *Trends Biochem. Sci.* 39, 447–456. doi: 10.1016/j.tibs.2014.06.006
- Bennetzen, J. L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A. C., et al. (2012). Reference genome sequence of the model plant *Setaria*. *Nat. Biotechnol.* 30, 555–561. doi: 10.1038/nbt.2196
- Bieberich, E. (2014). Synthesis, processing, and function of N-glycans in N-glycoproteins. *Adv. Neurobiol.* 9, 47–70. doi: 10.1007/978-1-4939-1154-7_3
- Boller, T., and Felix, G. (2009). A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu. Rev. Plant Biol.* 60, 379–406. doi: 10.1146/annurev.arplant.57.032905.105346
- Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Chan, A. P., Crabtree, J., Zhao, Q., Lorenzi, H., Orvis, J., Puiu, D., et al. (2010). Draft genome sequence of the oilseed species *Ricinus communis*. *Nat. Biotechnol.* 28, 951–956. doi: 10.1038/nbt.1674
- Chen, X., Zuo, S., Schwesinger, B., Chern, M., Canlas, P. E., Ruan, D., et al. (2014). An XA21-associated kinase (OsSERK2) regulates immunity mediated by the XA21 and XA3 immune receptors. *Mol. Plant* 7, 874–892. doi: 10.1093/mp/ssi003
- Chevalier, D., Batoux, M., Fulton, L., Pfister, K., Yadav, R. K., Schellenberg, M., et al. (2005). STRUBBELIG defines a receptor kinase-mediated signaling pathway regulating organ development in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9074–9079. doi: 10.1073/pnas.0503526102
- Chinchilla, D., Shan, L., He, P., De Vries, S., and Kemmerling, B. (2009). One for all: the receptor-associated kinase BAK1. *Trends Plant Sci.* 14, 535–541. doi: 10.1016/j.tplants.2009.08.002
- Chinchilla, D., Zipfel, C., Robatzek, S., Kemmerling, B., Nurnberger, T., Jones, J. D., et al. (2007). A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature* 448, 497–500. doi: 10.1038/nature05999
- Clark, S. E., Williams, R. W., and 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. doi: 10.1016/S0092-8674(00)80239-1

FUNDING

This work was supported by the German Research Foundation (DFG) grant number FI 1984/1-1 to IF; the Agropolis Resource Center for Crop Conservation, Adaptation and Diversity (ARCAD); the Centre de coopération Internationale de Recherche en Agronomie pour le Développement (CIRAD) Ph.D. fellowship to MB; and the Agence Nationale de la Recherche (ANR, France) ANR-08-GENM-021 to AD, CP, and EG.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://phylogeny.southgreen.fr/kinase2/>

- Claros, M. G., and von Heijne, G. (1994). TopPred II: an improved software for membrane protein structure predictions. *Comput. Appl. Biosci.* 10, 685–686. doi: 10.1093/bioinformatics/10.6.685
- Conte, M. G., Gaillard, S., Lanau, N., Rouard, M., and Perin, C. (2008). GreenPhylDB: a database for plant comparative genomics. *Nucleic Acids Res.* 36, D991–D998. doi: 10.1093/nar/gkm934
- Dassanayake, M., Oh, D.-H., Haas, J. S., Hernandez, A., Hong, H., Ali, S., et al. (2011). The genome of the extremophile crucifer *Thellungiella parvula*. *Nat. Genet.* 43, 913–918. doi: 10.1038/ng.889
- Deeken, R., and Kaldenhoff, R. (1997). Light-repressible receptor protein kinase: a novel photo-regulated gene from *Arabidopsis thaliana*. *Planta* 202, 479–486. doi: 10.1007/s004250050152
- Desper, R., and Gascuel, O. (2002). Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *J. Comput. Biol.* 9, 687–705. doi: 10.1089/106652702761034136
- D'Hont, A., Denoeud, F., Aury, J. M., Baurens, F. C., Carreel, F., Garsmeur, O., et al. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488, 213–217. doi: 10.1038/nature11241
- Di Rubbo, S., Irani, N. G., Kim, S. Y., Xu, Z. Y., Gadeyne, A., Dejonghe, W., et al. (2013). The clathrin adaptor complex AP-2 mediates endocytosis of brassinosteroid insensitive1 in *Arabidopsis*. *Plant Cell* 25, 2986–2997. doi: 10.1105/tpc.113.114058
- Dievart, A., and Clark, S. E. (2003). Using mutant alleles to determine the structure and function of leucine-rich repeat receptor-like kinases. *Curr. Opin. Plant Biol.* 6, 507–516. doi: 10.1016/S1369-5266(03)00089-X
- Dufayard, J.-F., Duret, L., Penel, S., Gouy, M., Rechenmann, F., and Perrière, G. (2005). Tree pattern matching in phylogenetic trees: automatic search for orthologs or paralogs in homologous gene sequence databases. *Bioinformatics* 21, 2596–2603. doi: 10.1093/bioinformatics/bti325
- Eddy, S. R. (2009). A new generation of homology search tools based on probabilistic inference. *Genome Inform.* 23, 205–211. doi: 10.1142/9781848165632_0019
- Felsenstein, J. (1989). PHYLIP – Phylogeny inference package. *Cladistics* 5, 164–165.
- Fischer, I., Dainat, J., Ranwez, V., Glemin, S., Dufayard, J. F., and Chantret, N. (2014). Impact of recurrent gene duplication on adaptation of plant genomes. *BMC Plant Biol.* 14:151. doi: 10.1186/1471-2229-14-151
- Fischer, I., Dievart, A., Droc, G., Dufayard, J. F., and Chantret, N. (2016). Evolutionary dynamics of the leucine-rich repeat receptor-like kinase (LRR-RLK) subfamily in angiosperms. *Plant Physiol.* 170, 1595–1610. doi: 10.1104/pp.15.01470
- Fisher, K., and Turner, S. (2007). PXY, a receptor-like kinase essential for maintaining polarity during plant vascular-tissue development. *Curr. Biol.* 17, 1061–1066. doi: 10.1016/j.cub.2007.05.049
- Forest, F., and Chase, M. W. (2009). “Eurosid I,” in *The Timetree of Life*, eds S. B. Hedges and S. Kumar (New York, NY: Oxford University Press), 188–196.
- Gao, M., Wang, X., Wang, D., Xu, F., Ding, X., Zhang, Z., et al. (2009). Regulation of cell death and innate immunity by two receptor-like kinases in *Arabidopsis*. *Cell Host Microbe* 6, 34–44. doi: 10.1016/j.chom.2009.05.019
- García-Mas, J., Benjak, A., Sanseverino, W., Bourgeois, M., Mir, G., González, V. M., et al. (2012). The genome of melon (*Cucumis melo* L.). *Proc. Natl. Acad. Sci. U.S.A.* 109, 11872–11877. doi: 10.1073/pnas.1205415109
- Goff, S. A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science* 296, 92–100. doi: 10.1126/science.1068275
- Goodstein, D., Shu, S., Howson, R., Neupane, R., Hayes, R., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520
- Gust, A. A., and Felix, G. (2014). Receptor like proteins associate with SOBIR1-type of adaptors to form bimolecular receptor kinases. *Curr. Opin. Plant Biol.* 21, 104–111. doi: 10.1016/j.pbi.2014.07.007
- Halter, T., Imkampe, J., Mazzotta, S., Wierzba, M., Postel, S., Bucherl, C., et al. (2014). The leucine-rich repeat receptor kinase BIR2 is a negative regulator of BAK1 in plant immunity. *Curr. Biol.* 24, 134–143. doi: 10.1016/j.cub.2013.11.047
- Hartmann, J., Stuhrowoldt, N., Dahlke, R. I., and Sauter, M. (2013). Phytosulfokine control of growth occurs in the epidermis, is likely to be non-cell autonomous and is dependent on brassinosteroids. *Plant J.* 73, 579–590. doi: 10.1111/tpj.12056
- Hecht, V., Vielle-Calzada, J. P., Hartog, M. V., Schmidt, E. D., Boutilier, K., Grossniklaus, U., et al. (2001). The Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1 gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. *Plant Physiol.* 127, 803–816. doi: 10.1104/pp.010324
- Heese, A., Hann, D. R., Gimenez-Ibanez, S., Jones, A. M. E., He, K., Li, J., et al. (2007). The receptor-like kinase SERK3/BAK1 is a central regulator of innate immunity in plants. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12217–12222. doi: 10.1073/pnas.0705306104
- Hok, S., Danchin, E. G., Allasia, V., Panabieres, F., Attard, A., and Keller, H. (2011). An *Arabidopsis* (malectin-like) leucine-rich repeat receptor-like kinase contributes to downy mildew disease. *Plant Cell Environ.* 34, 1944–1957. doi: 10.1111/j.1365-3040.2011.02390.x
- Hong, Z., Jin, H., Fitchette, A. C., Xia, Y., Monk, A. M., Faye, L., et al. (2009). Mutations of an alpha1,6 mannosyltransferase inhibit endoplasmic reticulum-associated degradation of defective brassinosteroid receptors in *Arabidopsis*. *Plant Cell* 21, 3792–3802. doi: 10.1105/tpc.109.070284
- Hong, Z., Jin, H., Tzfira, T., and Li, J. (2008). Multiple mechanism-mediated retention of a defective brassinosteroid receptor in the endoplasmic reticulum of *Arabidopsis*. *Plant Cell* 20, 3418–3429. doi: 10.1105/tpc.108.061879
- Hong, Z., Kajiura, H., Su, W., Jin, H., Kimura, A., Fujiyama, K., et al. (2012). Evolutionarily conserved glycan signal to degrade aberrant brassinosteroid receptors in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11437–11442. doi: 10.1073/pnas.1119173109
- Horn, M. A., and Walker, J. C. (1994). Biochemical properties of the autophosphorylation of RLK5, a receptor-like protein kinase from *Arabidopsis thaliana*. *Biochim. Biophys. Acta* 1208, 65–74. doi: 10.1016/0167-4838(94)90160-0
- Hothorn, M., Belkhadir, Y., Dreux, M., Dabi, T., Noel, J. P., Wilson, I. A., et al. (2011). Structural basis of steroid hormone perception by the receptor kinase BRI1. *Nature* 474, 467–471. doi: 10.1038/nature10153
- Hu, T. T., Pattyn, P., Bakker, E. G., Cao, J., Cheng, J.-F., Clark, R. M., et al. (2011). The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat. Genet.* 43, 476–481. doi: 10.1038/ng.807
- Huang, S., Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., et al. (2009). The genome of the cucumber, *Cucumis sativus* L. *Nat. Genet.* 41, 1275–1281. doi: 10.1038/ng.475
- Huttner, S., and Strasser, R. (2012). Endoplasmic reticulum-associated degradation of glycoproteins in plants. *Front. Plant Sci.* 3:67. doi: 10.3389/fpls.2012.00067
- Huttner, S., Veit, C., Vavra, U., Schoberer, J., Dicker, M., Maresch, D., et al. (2014). A context-independent N-glycan signal targets the misfolded extracellular domain of *Arabidopsis* STRUBBELIG to endoplasmic-reticulum-associated degradation. *Biochem. J.* 464, 401–411. doi: 10.1042/BJ20141057
- International Rice Genome Sequencing Project (2005). The map-based sequence of the rice genome. *Nature* 436, 793–800. doi: 10.1038/nature03895
- Irani, N. G., and Russinova, E. (2009). Receptor endocytosis and signaling in plants. *Curr. Opin. Plant Biol.* 12, 653–659. doi: 10.1016/j.pbi.2009.09.011
- Jaillon, O., Aury, J. M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449, 463–467. doi: 10.1038/nature06148
- Jaouannet, M., Rodriguez, P. A., Thorpe, P., Lenoir, C. J., Macleod, R., Escudero-Martinez, C., et al. (2014). Plant immunity in plant-aphid interactions. *Front. Plant Sci.* 5:663. doi: 10.3389/fpls.2014.00663
- Jung, C. G., Hwang, S. G., Park, Y. C., Park, H. M., Kim, D. S., Park, D. H., et al. (2015). Molecular characterization of the cold- and heat-induced *Arabidopsis* PXL1 gene and its potential role in transduction pathways under temperature fluctuations. *J. Plant Physiol.* 176, 138–146. doi: 10.1016/j.jplph.2015.01.001
- Jung, S., Cestaró, A., Troggio, M., Main, D., Zheng, P., Cho, I., et al. (2012). Whole genome comparisons of *Fragaria*, *Prunus* and *Malus* reveal different

- modes of evolution between Rosaceous subfamilies. *BMC Genomics* 13:129. doi: 10.1186/1471-2164-13-129
- Kajava, A. V. (1998). Structural diversity of leucine-rich repeat proteins. *J. Mol. Biol.* 277, 519–527. doi: 10.1006/jmbi.1998.1643
- Kang, Y. H., and Hardtke, C. S. (2016). *Arabidopsis* MAKR5 is a positive effector of BAM3-dependent CLE45 signaling. *EMBO Rep.* 17, 1145–1154. doi: 10.15252/embr.201642450
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Kay, B. K., Williamson, M. P., and Sudol, M. (2000). The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J.* 14, 231–241.
- Kim, B. H., Kim, S. Y., and Nam, K. H. (2013). Assessing the diverse functions of BAK1 and its homologs in arabidopsis, beyond BR signaling and PTI responses. *Mol. Cells* 35, 7–16. doi: 10.1007/s10059-013-2255-3
- Kinoshita, T., Cano-Delgado, A., Seto, H., Hiranuma, S., Fujioka, S., Yoshida, S., et al. (2005). Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. *Nature* 433, 167–171. doi: 10.1038/nature03227
- Kobe, B., and Deisenhofer, J. (1994). The leucine-rich repeat: a versatile binding motif. *Trends Biochem. Sci.* 19, 415–421. doi: 10.1016/0968-0004(94)90090-6
- Koch, M. A., Haubold, B., and Mitchell-Olds, T. (2000). Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol. Biol. Evol.* 17, 1483–1498. doi: 10.1093/oxfordjournals.molbev.a026248
- Koller, T., and Bent, A. F. (2014). FLS2-BAK1 extracellular domain interaction sites required for defense signaling activation. *PLoS ONE* 9:e111185. doi: 10.1371/journal.pone.0111185
- Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315
- Krol, E., Mentzel, T., Chinchilla, D., Boller, T., Felix, G., Kemmerling, B., et al. (2010). Perception of the arabidopsis danger signal peptide 1 involves the pattern recognition receptor AtPEPR1 and its close homologue AtPEPR2. *J. Biol. Chem.* 285, 13471–13479. doi: 10.1074/jbc.M109.097394
- Lee, H. Y., Bowen, C. H., Popescu, G. V., Kang, H. G., Kato, N., Ma, S., et al. (2011). *Arabidopsis* RTNLB1 and RTNLB2 Reticulon-like proteins regulate intracellular trafficking and activity of the FLS2 immune receptor. *Plant Cell* 23, 3374–3391. doi: 10.1105/tpc.111.089656
- Lehti-Shiu, M. D., Zou, C., Hanada, K., and Shiu, S. H. (2009). Evolutionary history and stress regulation of plant receptor-like kinase/pelle genes. *Plant Physiol.* 150, 12–26. doi: 10.1104/pp.108.134353
- Leslie, M. E., Lewis, M. W., Youn, J. Y., Daniels, M. J., and Liljegren, S. J. (2010). The EVERSLED receptor-like kinase modulates floral organ shedding in *Arabidopsis*. *Development* 137, 467–476. doi: 10.1242/dev.041335
- Letunic, I., Doerks, T., and Bork, P. (2009). SMART 6: recent updates and new developments. *Nucleic Acids Res.* 37, D229–D232. doi: 10.1093/nar/gkn808
- Li, J., and Chory, J. (1997). A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* 90, 929–938. doi: 10.1016/S0092-8674(00)80357-8
- Li, J., Zhao-Hui, C., Batoux, M., Nekrasov, V., Roux, M., Chinchilla, D., et al. (2009). Specific ER quality control components required for biogenesis of the plant innate immune receptor EFR. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15973–15978. doi: 10.1073/pnas.0905532106
- Liebrand, T. W., Van Den Berg, G. C., Zhang, Z., Smit, P., Cordewener, J. H., America, A. H., et al. (2013). Receptor-like kinase SOBIR1/EVR interacts with receptor-like proteins in plant immunity against fungal infection. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10010–10015. doi: 10.1073/pnas.1220015110
- Liu, P., Hu, Z., Zhou, B., Liu, S., and Chai, J. (2013). Crystal structure of an LRR protein with two solenoids. *Cell Res.* 23, 303–305. doi: 10.1038/cr.2012.159
- Liu, P., Wei, W., Ouyang, S., Zhang, J. S., Chen, S. Y., and Zhang, W. K. (2009). Analysis of expressed receptor-like kinases (RLKs) in soybean. *J. Genet. Genomics* 36, 611–619. doi: 10.1016/S1673-8527(08)60153-8
- Loivamaki, M., Stuhrowldt, N., Deeken, R., Steffens, B., Roitsch, T., Hedrich, R., et al. (2010). A role for PSK signaling in wounding and microbial interactions in *Arabidopsis*. *Physiol. Plant.* 139, 348–357. doi: 10.1111/j.1399-3054.2010.01371.x
- Macho, A. P., and Zipfel, C. (2014). Plant PRRs and the activation of innate immune signaling. *Mol. Cell.* 54, 263–272. doi: 10.1016/j.molcel.2014.03.028
- Magallcnd, S., and Castillo, A. (2009). Angiosperm diversification through time. *Am. J. Bot.* 96, 349–365. doi: 10.3732/ajb.0800060
- Martins, S., Dohmann, E. M., Cayrel, A., Johnson, A., Fischer, W., Pojer, F., et al. (2015). Internalization and vacuolar targeting of the brassinosteroid hormone receptor BRI1 are regulated by ubiquitination. *Nat. Commun.* 6:6151. doi: 10.1038/ncomms7151
- Matsubayashi, Y., Ogawa, M., Kihara, H., Niwa, M., and Sakagami, Y. (2006). Disruption and overexpression of *Arabidopsis* phyto-sulfokine receptor gene affects cellular longevity and potential for growth. *Plant Physiol.* 142, 45–53. doi: 10.1104/pp.106.081109
- Meng, X., Bonasera, J. M., Kim, J. F., Nissinen, R. M., and Beer, S. V. (2006). Apple proteins that interact with DspA/E, a pathogenicity effector of *erwinia amylovora*, the fire blight pathogen. *Mol. Plant Microbe Interact.* 19, 53–61. doi: 10.1094/MPMI-19-0053
- Ming, R., Hou, S., Feng, Y., Yu, Q., Dionne-Laporte, A., Saw, J. H., et al. (2008). The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452, 991–996. doi: 10.1038/nature06856
- Mosher, S., Seybold, H., Rodriguez, P., Stahl, M., Davies, K. A., Dayaratne, S., et al. (2013). The tyrosine-sulfated peptide receptors PSKR1 and PSY1R modify the immunity of *Arabidopsis* to biotrophic and necrotrophic pathogens in an antagonistic manner. *Plant J.* 73, 469–482. doi: 10.1111/tpj.12050
- Mu, J.-H., Lee, H.-S., and Kao, T.-H. (1994). Characterization of a pollen-expressed receptor-like kinase gene of *Petunia inflata* and the activity of its encoded kinase. *Plant Cell* 6, 709–721. doi: 10.1105/tpc.6.5.709
- Muschietti, J., Eyal, Y., and McCormick, S. (1998). Pollen tube localization implies a role in pollen-pistil interactions for the tomato receptor-like protein kinases LePRK1 and LePRK2. *Plant Cell* 10, 319–330. doi: 10.2307/3870591
- Nekrasov, V., Li, J., Batoux, M., Roux, M., Chu, Z. H., Lacombe, S., et al. (2009). Control of the pattern-recognition receptor EFR by an ER protein complex in plant immunity. *EMBO J.* 28, 3428–3438. doi: 10.1038/emboj.2009.262
- Ng, A. C., Eisenberg, J. M., Heath, R. J., Huett, A., Robinson, C. M., Nau, G. J., et al. (2011). Human leucine-rich repeat proteins: a genome-wide bioinformatic categorization and functional analysis in innate immunity. *Proc. Natl. Acad. Sci. U.S.A.* 108(Suppl. 1), 4631–4638. doi: 10.1073/pnas.1000093107
- Noguchi, T., Fujioka, S., Choe, S., Takatsuto, S., Yoshida, S., Yuan, H., et al. (1999). Brassinosteroid-insensitive dwarf mutants of *Arabidopsis* accumulate brassinosteroids. *Plant Physiol.* 121, 743–752. doi: 10.1104/pp.121.3.743
- Offringa, R., and Huang, F. (2013). Phosphorylation-dependent trafficking of plasma membrane proteins in animal and plant cells. *J. Integr. Plant Biol.* 55, 789–808. doi: 10.1111/jipb.12096
- Oh, D. H., Dassanayake, M., Haas, J. S., Kropornika, A., Wright, C., D'urzo, M. P., et al. (2010). Genome structures and halophyte-specific gene expression of the extremophile *Thellungiella parvula* in comparison with *Thellungiella salsuginea* (*Thellungiella halophila*) and *Arabidopsis*. *Plant Physiol.* 154, 1040–1052. doi: 10.1104/pp.110.163923
- Oh, S. H., and Potter, D. (2005). Molecular phylogenetic systematics and biogeography of tribe Neillieae (Rosaceae) using DNA sequences of cpDNA, rDNA, and LEAFY. *Am. J. Bot.* 92, 179–192. doi: 10.3732/ajb.92.1.179
- Park, C.-J., Sharma, R., Lefebvre, B., Canlas, P. E., and Ronald, P. C. (2013). The endoplasmic reticulum-quality control component SDF2 is essential for XA21-mediated immunity in rice. *Plant Sci.* 210, 53–60. doi: 10.1016/j.plantsci.2013.05.003
- Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., et al. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457, 551–556. doi: 10.1038/nature07723
- Prochnik, S., Marri, P., Desany, B., Rabinowicz, P., Kodira, C., Mohiuddin, M., et al. (2012). The cassava genome: current progress, future directions. *Trop. Plant Biol.* 5, 88–94. doi: 10.1007/s12042-011-9088-z
- Reineke, A. R., Bornberg-Bauer, E., and Gu, J. (2011). Evolutionary divergence and limits of conserved non-coding sequence detection in plant genomes. *Nucleic Acids Res.* 39, 6029–6043. doi: 10.1093/nar/gkr179
- Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., et al. (2008). The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319, 64–69. doi: 10.1126/science.1150646

- Robatzek, S., Chinchilla, D., and Boller, T. (2006). Ligand-induced endocytosis of the pattern recognition receptor FLS2 in *Arabidopsis*. *Genes Dev.* 20, 537–542. doi: 10.1101/gad.366506
- Robatzek, S., and Wirthmueller, L. (2013). Mapping FLS2 function to structure: LRRs, kinase and its working bits. *Protoplasma* 250, 671–681. doi: 10.1007/s00709-012-0459-6
- Rouard, M., Guignon, V., Aluome, C., Laporte, M. A., Droc, G., Walde, C., et al. (2011). GreenPhylDB v2.0: comparative and functional genomics in plants. *Nucleic Acids Res.* 39, D1095–D1102. doi: 10.1093/nar/gkq811
- Roux, M., Schwessinger, B., Albrecht, C., Chinchilla, D., Jones, A., Holton, N., et al. (2011). The *Arabidopsis* leucine-rich repeat receptor-like kinases BAK1/SERK3 and BKK1/SERK4 are required for innate immunity to hemibiotrophic and biotrophic pathogens. *Plant Cell* 23, 2440–2455. doi: 10.1105/tpc.111.084301
- Saijo, Y. (2010). ER quality control of immune receptors and regulators in plants. *Cell Microbiol.* 12, 716–724. doi: 10.1111/j.1462-5822.2010.01472.x
- Sakamoto, T., Deguchi, M., Brustolini, O., Santos, A., Silva, F., and Fontes, E. P. (2012). The tomato RLK superfamily: phylogeny and functional predictions about the role of the LRRII-RLK subfamily in antiviral defense. *BMC Plant Biol.* 12:229. doi: 10.1186/1471-2229-12-229
- Salomon, S., and Robatzek, S. (2006). Induced endocytosis of the receptor kinase FLS2. *Plant Signal. Behav.* 1, 293–295. doi: 10.4161/psb.1.6.3594
- Sanchez-Rodriguez, C., Estevez, J. M., Llorente, F., Hernandez-Blanco, C., Jorda, L., Pagan, I., et al. (2009). The ERECTA receptor-like kinase regulates cell wall-mediated resistance to pathogens in *Arabidopsis thaliana*. *Mol. Plant Microbe Interact.* 22, 953–963. doi: 10.1094/MPMI-22-8-0953
- Santiago, J., Henzler, C., and Hothorn, M. (2013). Molecular mechanism for plant steroid receptor activation by somatic embryogenesis co-receptor kinases. *Science* 341, 889–892. doi: 10.1126/science.1242468
- Sato, S., Hirakawa, H., Isobe, S., Fukai, E., Watanabe, A., Kato, M., et al. (2011). Sequence analysis of the genome of an oil-bearing tree, *Jatropha curcas* L. *DNA Res.* 18, 65–76. doi: 10.1093/dnares/dsq030
- Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., et al. (2008). Genome structure of the legume, *Lotus japonicus*. *DNA Res.* 15, 227–239. doi: 10.1093/dnares/dsn008
- Schmidt, E. D., Guzzo, F., Toonen, M. A., and De Vries, S. C. (1997). A leucine-rich repeat containing receptor-like kinase marks somatic plant cells competent to form embryos. *Development* 124, 2049–2062.
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670
- Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115. doi: 10.1126/science.1178534
- Schulze, B., Mentzel, T., Jehle, A. K., Mueller, K., Beeler, S., Boller, T., et al. (2010). Rapid heteromerization and phosphorylation of ligand-activated plant transmembrane receptors and their associated kinase BAK1. *J. Biol. Chem.* 285, 9444–9451. doi: 10.1074/jbc.M109.096842
- Sebastian, P., Schaefer, H., Telford, I. R., and Renner, S. S. (2010). Cucumber (*Cucumis sativus*) and melon (*C. melo*) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14269–14273. doi: 10.1073/pnas.1005338107
- Shah, K., Russinova, E., Gadella, T. W. Jr., Willemsse, J., and De Vries, S. C. (2002). The *Arabidopsis* kinase-associated protein phosphatase controls internalization of the somatic embryogenesis receptor kinase 1. *Genes Dev.* 16, 1707–1720. doi: 10.1101/gad.220402
- She, J., Han, Z., Kim, T. W., Wang, J., Cheng, W., Chang, J., et al. (2011). Structural insight into brassinosteroid perception by BRI1. *Nature* 474, 472–476. doi: 10.1038/nature10178
- Shi, T., Huang, H., Sanderson, M. J., and Tax, F. E. (2014). Evolutionary dynamics of leucine-rich repeat receptor-like kinases and related genes in plants: a phylogenomic approach. *J. Integr. Plant Biol.* 56, 648–662. doi: 10.1111/jipb.12188
- Shiu, S. H., and Bleecker, A. B. (2001a). Plant receptor-like kinase gene family: diversity, function, and signaling. *Sci. STKE* 2001:RE22. doi: 10.1126/stke.2001.113.re22
- Shiu, S. H., and 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. doi: 10.1073/pnas.181141598
- Shiu, S. H., and Bleecker, A. B. (2003). Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in *Arabidopsis*. *Plant Physiol.* 132, 530–543. doi: 10.1104/pp.103.021964
- Shiu, S. H., Karlowski, W. M., Pan, R., Tzeng, Y. H., Mayer, K. F., and Li, W. H. (2004). Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. *Plant Cell* 16, 1220–1234. doi: 10.1105/tpc.020834
- Song, W., Han, Z., Sun, Y., and Chai, J. (2014). Crystal structure of a plant leucine rich repeat protein with two island domains. *Sci. China Life Sci.* 57, 137–144. doi: 10.1007/s11427-013-4586-x
- Song, W. Y., Wang, G. L., Chen, L. L., Kim, H. S., Pi, L. Y., Holsten, T., et al. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* 270, 1804–1806. doi: 10.1126/science.270.5243.1804
- Song, X., Guo, P., Li, C., and Liu, C.-M. (2010). The cysteine pairs in CLV2 are not necessary for sensing the CLV3 peptide in shoot and root meristems. *J. Integr. Plant Biol.* 52, 774–781. doi: 10.1111/j.1744-7909.2010.00978.x
- Stuhrwoldt, N., Dahlke, R. I., Kutschmar, A., Peng, X., Sun, M. X., and Sauter, M. (2015). Phytosulfokine peptide signaling controls pollen tube growth and funicular pollen tube guidance in *Arabidopsis thaliana*. *Physiol. Plant.* 153, 643–653. doi: 10.1111/ppl.12270
- Su, W., Liu, Y., Xia, Y., Hong, Z., and Li, J. (2011). Conserved endoplasmic reticulum-associated degradation system to eliminate mutated receptor-like kinases in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 870–875. doi: 10.1073/pnas.1013251108
- Su, W., Liu, Y., Xia, Y., Hong, Z., and Li, J. (2012). The *Arabidopsis* homolog of the mammalian OS-9 protein plays a key role in the endoplasmic reticulum-associated degradation of misfolded receptor-like kinases. *Mol. Plant* 5, 929–940. doi: 10.1093/mp/sss042
- Sun, W., Cao, Y., Jansen Labby, K., Bittel, P., Boller, T., and Bent, A. F. (2012). Probing the *Arabidopsis* flagellin receptor: FLS2-FLS2 association and the contributions of specific domains to signaling function. *Plant Cell* 24, 1096–1113. doi: 10.1105/tpc.112.095919
- Sun, Y., Han, Z., Tang, J., Hu, Z., Chai, C., Zhou, B., et al. (2013). Structure reveals that BAK1 as a co-receptor recognizes the BRI1-bound brassinolide. *Cell Res.* 23, 1326–1329. doi: 10.1038/cr.2013.131
- Tanigaki, Y., Ito, K., Obuchi, Y., Kosaka, A., Yamato, K. T., Okanami, M., et al. (2014). *Physcomitrella patens* has kinase-LRR R gene homologs and interacting proteins. *PLoS ONE* 9:e95118. doi: 10.1371/journal.pone.0095118
- The *Arabidopsis* Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi: 10.1038/35048692
- The International Brachypodium Initiative (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463, 763–768. doi: 10.1038/nature08747
- The Potato Genome Sequencing Consortium (2011). Genome sequence and analysis of the tuber crop potato. *Nature* 475, 189–195. doi: 10.1038/nature10158
- Tintor, N., and Saijo, Y. (2014). ER-mediated control for abundance, quality, and signaling of transmembrane immune receptors in plants. *Front. Plant Sci.* 5:65. doi: 10.3389/fpls.2014.00065
- Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485, 635–641. doi: 10.1038/nature11119
- Torii, K. U., Mitsukawa, N., Oosumi, T., Matsuura, Y., Yokoyama, R., Whittier, R. F., et al. (1996). The *Arabidopsis* ERECTA gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. *Plant Cell* 8, 735–746. doi: 10.1105/tpc.8.4.735
- Tuskan, G., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (torr. & Gray). *Science* 313, 1596–1604. doi: 10.1126/science.1128691
- Varshney, R. K., Chen, W., Li, Y., Bharti, A. K., Saxena, R. K., Schlueter, J. A., et al. (2012). Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat. Biotechnol.* 30, 83–89. doi: 10.1038/nbt.2022
- Velasco, R., Zharkikh, A., Affourtit, J., Dhirgra, A., Cestaro, A., Kalyanaraman, A., et al. (2010). The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nat. Genet.* 42, 833–839. doi: 10.1038/ng.654

- Wang, H., Moore, M. J., Soltis, P. S., Bell, C. D., Brockington, S. F., Alexandre, R., et al. (2009). Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proc. Natl. Acad. Sci. U.S.A.* 106, 3853–3858. doi: 10.1073/pnas.0813376106
- Wang, K., Wang, Z., Li, F., Ye, W., Wang, J., Song, G., et al. (2012). The draft genome of a diploid cotton *Gossypium raimondii*. *Nat. Genet.* 44, 1098–1103. doi: 10.1038/ng.2371
- Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., et al. (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* 43, 1035–1039. doi: 10.1038/ng.919
- Wu, Y., Xun, Q., Guo, Y., Zhang, J., Cheng, K., Shi, T., et al. (2016). Genome-wide expression pattern analyses of the *Arabidopsis* leucine-rich repeat receptor-like kinases. *Mol. Plant* 9, 289–300. doi: 10.1016/j.molp.2015.12.011
- Xu, S. L., Rahman, A., Baskin, T. I., and Kieber, J. J. (2008). Two leucine-rich repeat receptor kinases mediate signaling, linking cell wall biosynthesis and ACC synthase in *Arabidopsis*. *Plant Cell* 20, 3065–3079. doi: 10.1105/tpc.108.063354
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Young, N. D., Debelle, F., Oldroyd, G. E. D., Geurts, R., Cannon, S. B., Udvardi, M. K., et al. (2011). The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480, 520–524. doi: 10.1038/nature10625
- Yu, J., Hu, S., Wang, J., Wong, G. K., Li, S., Liu, B., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 296, 79–92. doi: 10.1126/science.1068037
- Zan, Y., Ji, Y., Zhang, Y., Yang, S., Song, Y., and Wang, J. (2013). Genome-wide identification, characterization and expression analysis of populus leucine-rich repeat receptor-like protein kinase genes. *BMC Genomics* 14:318. doi: 10.1186/1471-2164-14-318
- Zhang, G., Liu, X., Quan, Z., Cheng, S., Xu, X., Pan, S., et al. (2012). Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nat. Biotechnol.* 30, 549–554. doi: 10.1038/nbt.2195

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Dufayard, Bettembourg, Fischer, Droc, Guiderdoni, Périn, Chantret and Diévert. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.