

RESEARCH ARTICLE

Open Access

# The tomato RLK superfamily: phylogeny and functional predictions about the role of the LRRII-RLK subfamily in antiviral defense

Tetsu Sakamoto<sup>1</sup>, Michihito Deguchi<sup>1,2</sup>, Otávio JB Brustolini<sup>1,2</sup>, Anésia A Santos<sup>1,2</sup>, Fabyano F Silva<sup>3</sup> and Elizabeth PB Fontes<sup>1,2\*</sup>

## Abstract

**Background:** Receptor-like kinases (RLKs) play key roles during development and in responses to the environment. Despite the relevance of the RLK family and the completion of the tomato genome sequencing, the tomato RLK family has not yet been characterized, and a framework for functional predictions of the members of the family is lacking.

**Results:** To generate a complete list of all the members of the tomato RLK family, we performed a phylogenetic analysis using the Arabidopsis family as a template. A total of 647 RLKs were identified in the tomato genome, which were organized into the same subfamily clades as Arabidopsis RLKs. Only eight of 58 RLK subfamilies exhibited specific expansion/reduction compared to their Arabidopsis counterparts. We also characterized the LRRII-RLK family by phylogeny, genomic analysis, expression profile and interaction with the virulence factor from begomoviruses, the nuclear shuttle protein (NSP). The LRRII subfamily members from tomato and Arabidopsis were highly conserved in both sequence and structure. Nevertheless, the majority of the orthologous pairs did not display similar conservation in the gene expression profile, indicating that these orthologs may have diverged in function after speciation. Based on the fact that members of the Arabidopsis LRRII subfamily (AtNIK1, AtNIK2 and AtNIK3) interact with the begomovirus nuclear shuttle protein (NSP), we examined whether the tomato orthologs of *NIK*, *BAK1* and *NsAK* genes interact with NSP of *Tomato Yellow Spot Virus* (ToYSV). The tomato orthologs of NSP interactors, *SINIKs* and *SINsAK*, interacted specifically with NSP in yeast and displayed an expression pattern consistent with the pattern of geminivirus infection. In addition to suggesting a functional analogy between these phylogenetically classified orthologs, these results expand our previous observation that NSP-NIK interactions are neither virus-specific nor host-specific.

**Conclusions:** The tomato RLK superfamily is made-up of 647 proteins that form a monophyletic tree with the Arabidopsis RLKs and is divided into 58 subfamilies. Few subfamilies have undergone expansion/reduction, and only six proteins were lineage-specific. Therefore, the tomato RLK family shares functional and structural conservation with Arabidopsis. For the LRRII-RLK members *SINIK1* and *SINIK3*, we observed functions analogous to those of their Arabidopsis counterparts with respect to protein-protein interactions and similar expression profiles, which predominated in tissues that support high efficiency of begomovirus infection. Therefore, NIK-mediated antiviral signaling is also likely to operate in tomato, suggesting that tomato NIKs may be good targets for engineering resistance against tomato-infecting begomoviruses.

**Keywords:** Receptor-like kinase, NIK, NSP, BAK1, Virus infection, Protein-protein interaction, Functional divergence, Plant signaling, *Solanum lycopersicum*

\* Correspondence: [bbfontes@ufv.br](mailto:bbfontes@ufv.br)

<sup>1</sup>National Institute of Science and Technology in Plant-Pest Interactions, Universidade Federal de Viçosa, 36570-000, Viçosa, MG, Brazil

<sup>2</sup>Departamento de Bioquímica e Biologia Molecular/BIOAGRO, Universidade Federal de Viçosa, 36570-000, Viçosa, MG, Brazil

Full list of author information is available at the end of the article

## Background

Plant cells constantly react to multiple signals that come from the local environment, neighboring cells, or even from other organisms. Depending on the stimuli, plant cells may expand, divide, differentiate, synthesize compounds, prepare against pathogen infection, or induce necrosis [1]. To perceive and receive these signals, plant cells possess complex systems of transmembrane receptor proteins that facilitate communication between the intracellular environment and the outside world.

One of the largest groups of these receptors is the receptor-like kinase (RLK) superfamily, which contains over 600 members in *Arabidopsis* [2-4]. RLKs are structurally organized into an extracellular domain that can be highly divergent, followed by a transmembrane segment and a conserved intracellular serine/threonine kinase domain. Most RLKs are localized in the plasma membrane, although there are also RLK members that are found in the cytoplasm. In this case, RLKs do not possess either an extracellular region or a transmembrane domain and are called receptor-like cytoplasmic kinases (RLCKs). Analyses of *Arabidopsis* RLKs by structural comparison of their extracellular region and phylogenetic analysis of their kinase domain revealed that they can be divided into over 50 subfamilies [5].

Several distinct RLKs have been studied in the past decade, and a common theme that has emerged is that binding of a specific signal molecule to their extracellular domain is required to initiate a signal transduction cascade [6]. Generally, ligand-receptor interactions at the extracellular domain of RLKs initiate the propagation of the signal through the membrane by inducing a conformational change in the receptor kinase domain, which allows interactions with other RLKs resulting in homo- or heterodimers. Dimerized RLKs are then transphosphorylated by their cytoplasmic kinase domain, leading to both activation of the kinase and establishment of docking sites for phosphorylation of downstream phosphorylation targets [7,8]. This activation mechanism of plant RLKs is similar to that of signal transduction mediated by receptor tyrosine kinases in animal cells, which share a common origin with plant serine/threonine kinases [3].

Functional analysis of RLKs indicates that the majority of them are associated with plant development or defense response, but there are also RLKs involved in cell wall attachment (extensin, proline-rich extensin and lectin RLKs), plant-bacterial symbiotic interactions (LysM RLKs) and self-incompatibility (S-domain containing RLKs). Among all RLKs, those bearing a leucine-rich repeat (LRR) domains are overrepresented in the RLK superfamily, comprising over 38% of *Arabidopsis* RLKs, which are distributed into 15 subgroups (LRR I to LRR XV). The LRR domains in these receptors vary in number (from one to 25) and in the distribution pattern of the LRRs along

the extracellular region. Examples of well-known LRR-RLKs include CLAVATA1, which controls the size of stem cells in the apical meristem by forming a heterodimer with CLAVATA2 and then interacting with CLAVATA3 through the extracellular domain [9], and BRASSINOSTEROID INSENSITIVE-1 (BRI1) [10,11], which perceives brassinosteroids and interacts with its receptor partner, BAK1 (BRI1-ASSOCIATED KINASE-1) [12,13]. Other functions associated with LRR-RLKs include morphogenesis [14-20], embryogenesis [21-24], pollen self-incompatibility [25] and responses to environmental signals [26]. In addition, some LRR-RLKs are known to function as regulators of defense response to bacterial pathogen [27-29], necrotrophic fungus [30] and viral infection [31,32].

Most of the characterized RLKs are from model plants such as *Arabidopsis* and *Medicago truncatula*, but significant efforts have been made to expand these studies to relevant field crops. Large-scale comparative analyses of *Arabidopsis* RLKs with rice [5,33,34] and soybean [35] RLKs identified over 1000 kinase proteins in rice and 600 in soybean belonging to the RLK superfamily; almost all members were grouped into previously determined *Arabidopsis* RLK subfamilies. The RLK subfamilies with developmental function have conserved size, whereas those involved in defense response have expanded their members, mainly by tandem duplication [5].

Although tomato (*Solanum lycopersicum*) is one of the most consumed and cultivated field crops in the world, a large-scale phylogenetic analysis of tomato RLKs has not yet been performed, and few members of the tomato RLK/Pelle family (RLKs + RLCKs) have been studied and characterized. These members include Pto [36], Pti1 (Pto-INTERACTING 1) [37], and Bti9 (AvrPtoB-TOMATO INTERACTING PROTEIN 9) [38], which interact with *Pseudomonas syringae* elicitors; TARK1 (TOMATO ATYPICAL RECEPTOR KINASE-1) [39], which interacts with the *Xanthomonas campestris* elicitor; TPK1b (TOMATO PROTEIN KINASE 1) [40], whose expression is induced by mechanical wounding and oxidative stress; and SR160 (SYSTEMIN RECEPTOR) [41], which is the AtBRI1 ortholog and binds to systemin to respond to wounding or herbivore attack, although there is some debate about the function of this receptor [42]. Another well-studied RLK in tomato is NIK (NSP-INTERACTING KINASE), which interacts with nuclear shuttle protein (NSP) of geminivirus during infection [43]. Three homologs of NIK in *Arabidopsis* (AtNIK1, AtNIK2 and AtNIK3) have also been shown to interact with NSP through their kinase domain [31]. This interaction causes inhibition of the kinase activity of NIKs and hence prevents the activation of the signal transduction cascade that evokes a plant defense response [32]. These RLKs are members of the LRRII subfamily that also contains the SOMATIC EMBRYOGENESIS RECEPTOR KINASES (SERKs) [44].

With the completion of the tomato genome sequencing along with the annotation of the encoded proteins [45], it has become possible to study the RLK superfamily in this species using a large-scale phylogenetic approach. According to genomic analyses, the tomato genome was predicted to have approximately 900 megabases of DNA and encode 34,727 proteins. In this investigation, we identified and classified all putative tomato RLKs by comparison with previously described Arabidopsis RLKs [5]. We also showed that the tomato RLK members of LRRII subfamily, which comprises *NIK* and *SERK* genes, share similar biochemical activity (capacity to interact with the geminivirus NSP), genomic structure and partial overlapping expression profiles with the Arabidopsis orthologs. Our results provide a framework for understanding RLK function in tomato and reveal that some tomato and Arabidopsis LRRII-RLK orthologs may play similar roles in antiviral defense.

## Results

### The tomato RLK superfamily

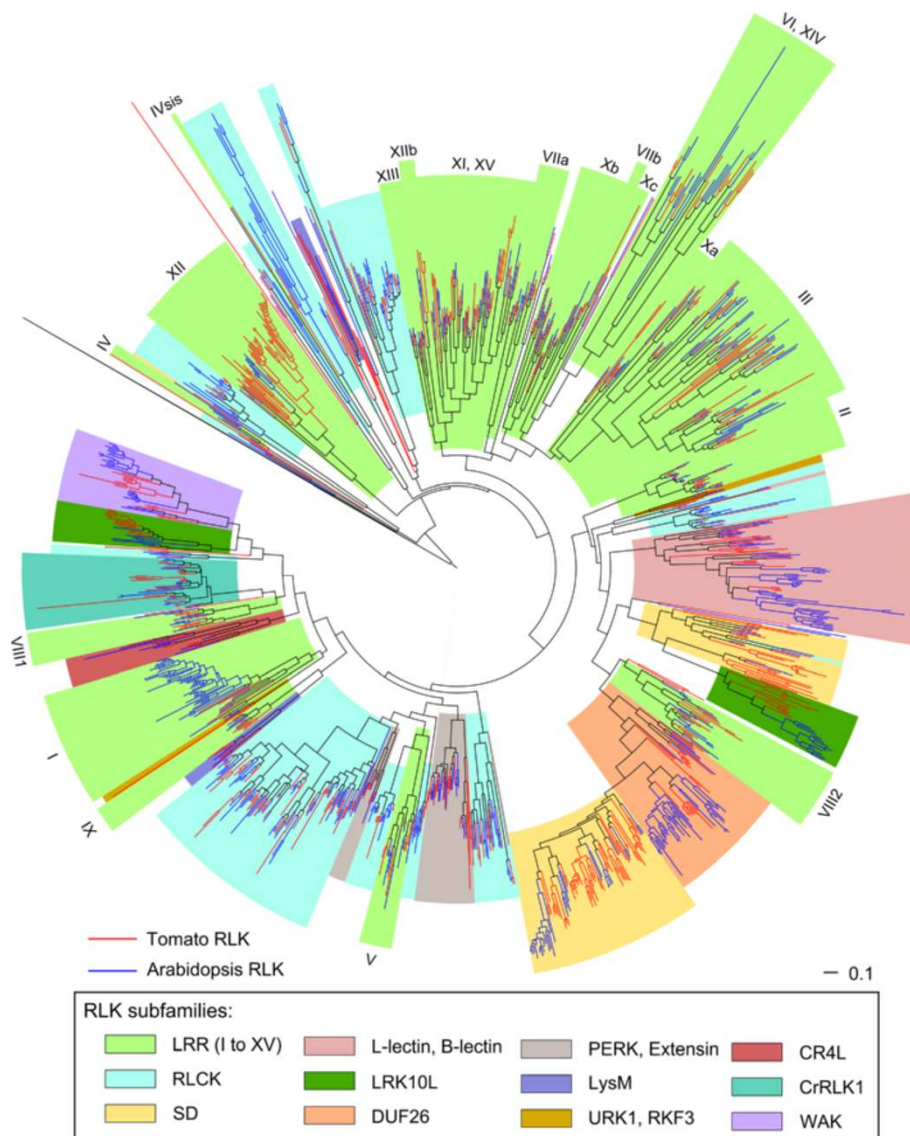
The identification of the RLK superfamily members in tomato was initially performed by a batch BLAST search against a tomato protein database (ITAG v2.3, available in solgenomics.net) using the kinase sequences of representative Arabidopsis RLKs as queries. This analysis retrieved 955 tomato proteins that seemed to be RLKs. All of these retrieved tomato proteins were submitted for annotation of their domain structure using SMART [46] (smart.embl-heidelberg.de) and Pfam [47] (pfam.sanger.ac.uk) databases. Four proteins that did not bear a kinase domain were not considered for further analysis. The remaining 951 proteins were used for phylogenetic analysis based on their kinase domain sequences. For this analysis, we included all Arabidopsis RLKs to compare with tomato RLKs and used representative proteins of other kinase families of Arabidopsis and human as outgroups (Additional file 1). All Arabidopsis RLKs were placed in a major cluster together with 647 tomato proteins that were identified as members of the RLK superfamily (Figure 1). The other 304 proteins were clustered with outgroups; consequently, they were not considered to be members of RLK superfamily (Additional file 2).

The size of the tomato RLK superfamily (647 RLKs) was similar to that of the Arabidopsis RLK superfamily (623 RLKs). Furthermore, almost all tomato RLKs (631 RLKs) were clustered with at least one Arabidopsis RLK. Therefore, the tomato RLK superfamily was divided into the same 58 subfamilies as described previously for Arabidopsis [5]. As in Arabidopsis, in which 236 out of all 623 RLKs belong to leucine-rich repeat (LRR) subfamilies, tomato LRR subfamilies were the most abundant and contained 257 proteins. Another large RLK subfamily was

RLCK, which included 128 members in tomato, almost the same number as in Arabidopsis (150). Among the 16 tomato RLKs that were not clustered in the same branches as Arabidopsis RLKs, ten proteins were quite small and lacked a typical RLK structure, but the other six proteins had a clear RLK structure and as such were considered to be tomato-specific RLKs. Among those six tomato-specific RLKs, Solyc03g080060 contained a legume lectin domain similar to members of the lectin subfamily, and Solyc02g083410 harbored an amino oxidase domain (flavin containing amine oxidoreductase activity), which is not found in any Arabidopsis RLKs. The remaining four proteins did not have any predicted protein domains in their extracellular region. RLK superfamily profiles in both species are summarized in Figure 2.

Although the RLK superfamilies of Arabidopsis and tomato share common features, a close inspection reveals some interesting differences between them. A comparison of membership size of each subfamily revealed some differences between the species. To infer the number of RLKs in their common ancestral and the occurrence of duplication/deletion events after the divergence of both species, we used a reconciliation method [48] to compare the RLK superfamily tree (Figure 1 and Additional file 2) with a species tree generated at NCBI taxonomy browser (Additional file 3). Then, the frequency of duplication or deletion events that occurred in each RLK subfamily was statistically analyzed. We identified twelve RLK subfamilies from these plant species that have significantly expanded/reduced in their size (Additional files 3 and 4, test of equal or given proportion,  $p < 0.05$ ). In tomato, the LRK10L2, SD1, SD2b and LRRXII subfamilies displayed significant number of duplication, while in Arabidopsis, expansion was observed in DUF26, L-LEC, LRK10L2, SD1, WAK, LRRIIa and RLCKXII/XIII subfamilies. Conversely, LysMII in Arabidopsis and LRRIIa and RLCKXII in tomato reduced in size. Another distinct aspect of RLKs in these plant species refers to the lack of a common domain structure in the extracellular region of RLK members of the LRK10L-2 subfamily. Whereas the extracellular region of Arabidopsis LRK10L-2 members harbors diverse structures, such as thaumatin, glycerophosphoryl diester phosphodiesterase family (GDPD) or malectin domains, the tomato RLKs of this same subfamily do not contain any predicted domain structure in their N-terminal region.

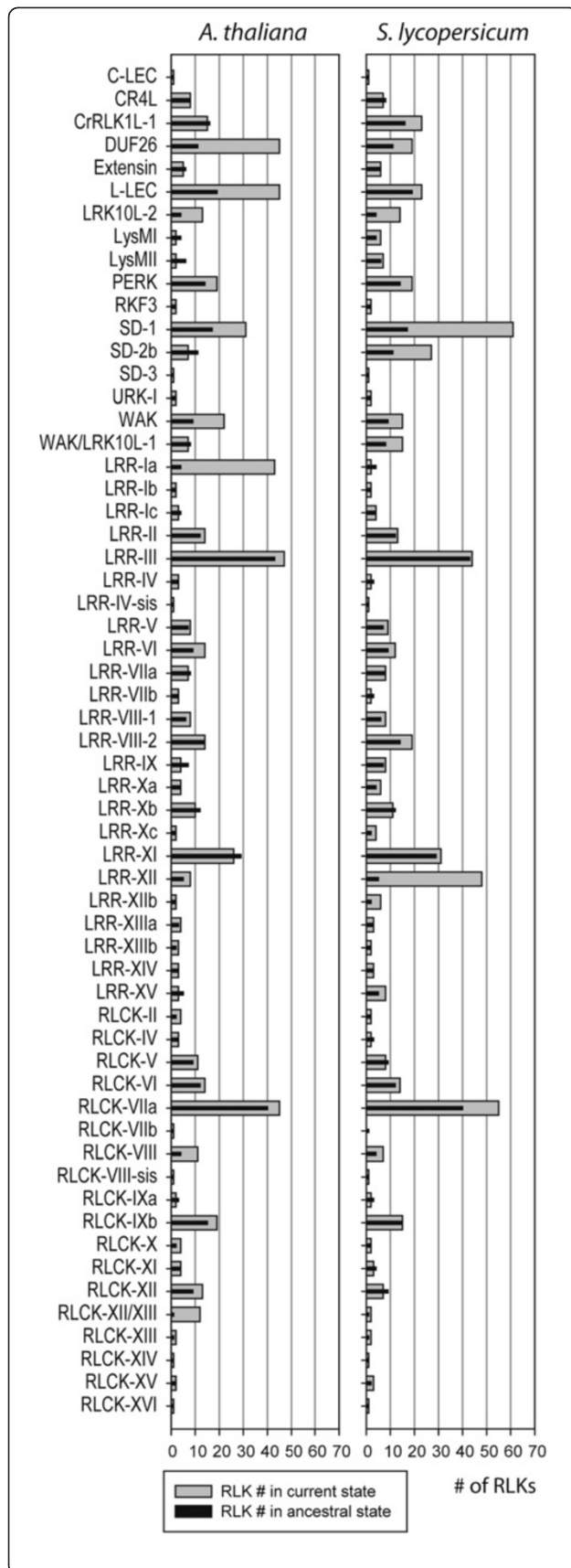
Further analyses were performed to predict function associated with expansion/reduction patterns. As RLKs are frequently associated to defense or developmental processes, we performed a search using the Gene Ontology (GO) terms [49] for functionally annotated RLKs in those categories (Additional file 3) and statistically compared the proportion of annotated genes in each RLK



**Figure 1 The tomato RLK superfamily is composed of 647 proteins.** Phylogenetic tree constructed by sequence alignment of kinase domain of Arabidopsis RLKs together with putative tomato RLKs. The alignment was carried out with CLUSTALW, and the phylogenetic tree reconstruction was made using FastTree. Almost all tomato RLKs (red branches) clustered with Arabidopsis RLKs (blue branches). Color ranges delimit the RLK subfamilies. LRR subfamilies (light green) are subdivided in 15 groups, and each group is identified in the tree with Roman numerals (I to XV).

subfamily with the proportion of annotated genes in the whole RLK superfamily (see Methods for more details). Compelling evidence in the literature has demonstrated that defense-related genes have high duplication rate and are organized in tandem repeats [34,50,51]. We also identified the RLKs organized in tandem repeats and determined their frequency in each RLK subfamilies (Additional file 3). These analyses demonstrated that most of the RLK subfamilies, which expanded after Arabidopsis and tomato species divergence, had their genes organized in tandem repeats. Functional annotation analysis of their genes revealed that the GO terms of 42 out

of 214 genes in tandem repeats were associated with defense response. RLKs annotated as developmental-related were overrepresented in CrRLK1-1, PERK, LRRVIIa, LRRXI and LRRXIII subfamilies and all those subfamilies did not expand or present genes in tandem arrays. The subfamilies that underwent reduction in their size were not associated with defense- or development-related functions, except for the LRRIIa subfamily, which may be related to defense response. Consistent with the involvement of members of the LRRII subfamily in defense and development, the LRRII subfamily of tomato and Arabidopsis had significantly high number of



**Figure 2** The number of members varies in some tomato and Arabidopsis RLK subfamilies. The distribution profile of tomato and Arabidopsis RLKs in subfamilies (gray bar) and the estimated number of RLK in their common ancestral (black bar) are presented. Almost all RLK subfamilies described in Arabidopsis have representatives in tomato.

annotated members in either defense or developmental categories.

### Motif prediction, genomic structure and phylogenetic analysis of the LRR-II subfamily

Compelling evidence in the literature has revealed a fundamental role for members of the Arabidopsis LRR-II-RLK subfamily as co-receptors for transducing developmental and defense signals [52-54]. The potential of the members of this subfamily as co-receptors involved in the activation of RLK-mediated signal transduction prompted us to perform a comprehensive analysis of the tomato LRR-II-RLK subfamily to uncover related functions in tomatoes. Based on the phylogenetic tree of all members of RLK superfamily, the tomato LRR-II-RLK subfamily encompassed 13 proteins. The members of this group from both plant species have over 600 amino acids on average. Phylogenetic analysis of this group using full-length protein sequences resulted in a tree with three well-resolved clusters; the tomato and Arabidopsis proteins were found in all clusters, although they had distinct sizes (Figure 3A). These clades were termed NIK, SERK and LRR-IIc based on annotation of the Arabidopsis members in each cluster. The NIK clade is formed by seven tomato members and six Arabidopsis members, including the three AtNIK genes. The SERK clade clustered the five well-characterized SERKs in Arabidopsis and three members of the tomato subfamily. The LRR-IIc clade consisted of three tomato proteins and three Arabidopsis proteins whose functions are unknown. Motif prediction analysis on these proteins revealed that tomato and Arabidopsis LRR-II-RLK members display similar protein domains organized in the same fashion (Figure 4). The consensus structural organization of the conserved domains between both species included an N-terminal signal peptide followed by a leucine zipper, five LRRs at the extracellular side and a transmembrane domain separating the N-terminal portion from the cytoplasmic C-terminal kinase domain. Among the SERK genes of both species, there was also a proline-rich domain (SPP) localized between the last LRR and the transmembrane domain (Figure 4). Sequence alignment of the LRR-II RLKs showed several conserved amino acid positions among members of both plant species. Exon/intron boundaries were also well conserved. Variation at the sequence level was observed within the SPP and signal peptide recognition domains. Among the

LRRII clades, the proteins comprising the SERK clade were more conserved with a larger number of conserved positions compared with the predicted proteins of the LRRIIc and NIK clades (Figure 4). Genomic structure analysis revealed that in general LRRII genes are organized into 11 exons (Figure 5A). Genes that varied in this number displayed fused exons, including *AtNIK1*, *At5g10290.1* and *SINIK3*, or had deleted exons, such as *At5g63710.1*, *Solyc02g072310.2.1* and *Solyc05g005140.2.1*. Intronic regions were larger in tomato members than in Arabidopsis members and in *SERK* genes compared with genes from other clades (Figure 5B).

#### Expression analysis of LRRII subfamily genes in different tissues

We examined the expression profiles of LRRII-RLK genes in different tomato tissues, including leaf, stem, root, flower, cotyledon and hypocotyl, by real-time PCR. The results are presented in Figure 6 and summarized in Figure 3B. Almost all analyzed genes exhibited expression in at least one organ, except for *Solyc05g005140.2.1* and *Solyc02g072310.2.1* in the NIK clade, which had very low expression in all tissues tested. The tomato genes in the SERK clade were expressed more highly in leaf and cotyledon tissue than in stem or flower tissue. The NIK clade genes were highly expressed in diverse organs, such as leaves, flowers and roots. The LRRIIc group encompassed genes with a higher level of expression in cotyledon, flower and leaf tissues and with lower expression in stem tissue.

The expression data for LRRII subfamily members from Arabidopsis (Figure 3C) were extracted from AtGenExpress [55] to examine whether there is some correspondence in the expression profiles between orthologous pairs of tomato and Arabidopsis genes. Statistical analyses of correlation between tomato and Arabidopsis expression data could not be performed because these data have been generated by different methods (qRT-PCR and microarray) and hence they have different units. Nevertheless, a subjective comparison of the expression analysis from both plants revealed that the majority of the orthologous genes displayed partially but not entirely overlapping expression profiles (Figure 3D). The orthologous groups that presented similar expression profile were *AtSERK1/AtSERK2* and *Solyc04g072570*, which had high expression levels in leaves, *AtNIK1* and *SINIK1*, which were lowly expressed in stem and cotyledon tissues, and *AtNIK3* and *SINIK3*, which were most highly expressed in the leaf.

#### Interactions between representatives of the LRRII subfamily and NSP of ToYSV

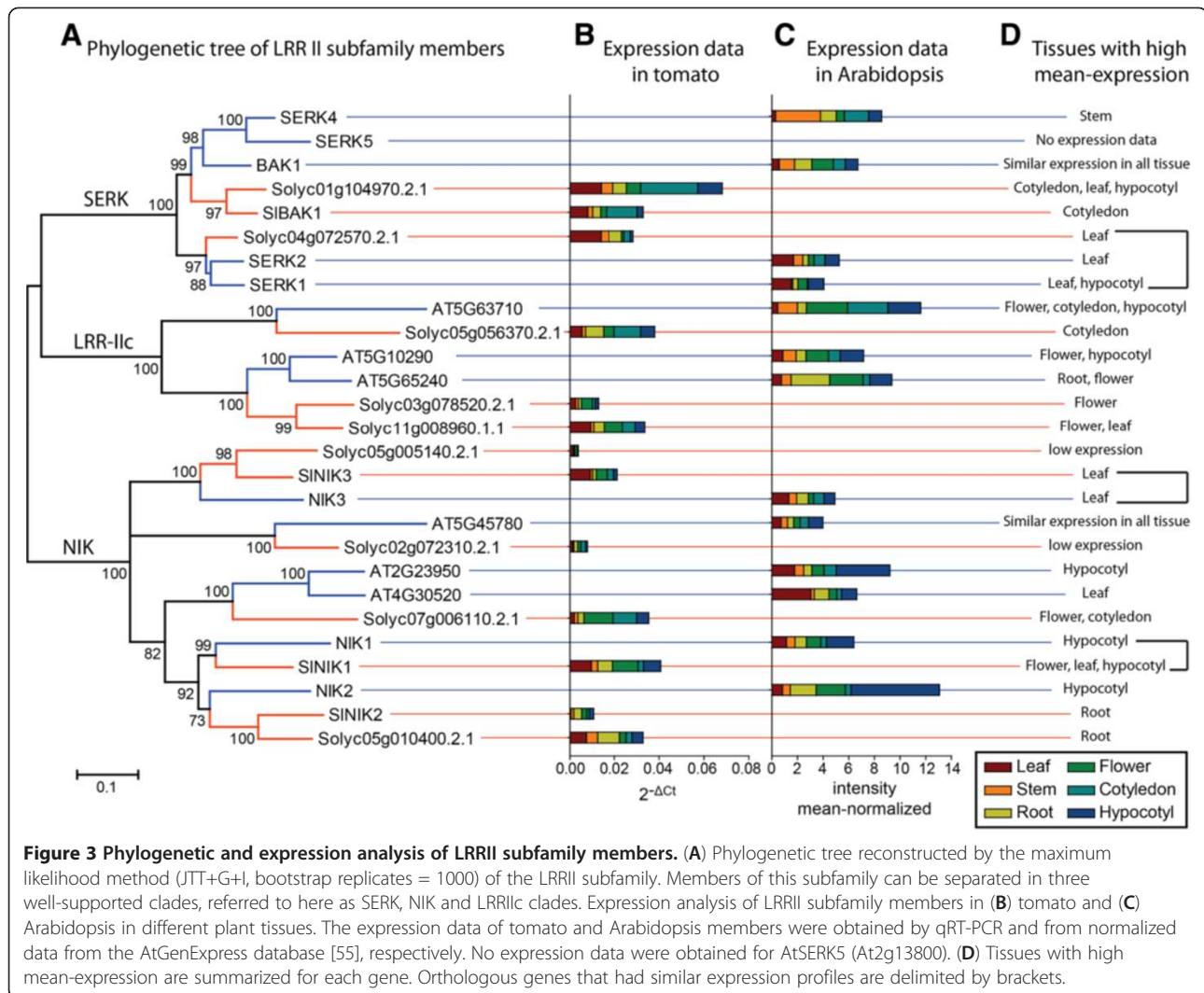
We have previously shown that NSP from begomovirus interacts with members of the LRRII-RLK subfamily, such as *AtNIK1*, *AtNIK2* and *AtNIK3*, to suppress host defense, and it interacts with a member of the PERK-like

RLK subfamily, NSP-ASSOCIATED KINASE (NsAK), to potentiate virus infection [31,56]. Either NIK from tomato and NsAK from Arabidopsis were isolated by their capacity to interact with NSP through two-hybrid screening [43,56]. The NSP interactions with the Arabidopsis *AtNIK1*, *AtNIK2* and *AtNIK3* and NsAK were further confirmed by yeast-two hybrid and in vitro pull down assays [31,56]. We have recently shown by bimolecular fluorescence complementation (BiFC) assay that NSP also interacts with NIK *in vivo*. Because begomovirus negatively impacts tomato cultivation worldwide, we selected representatives of tomato RLKs from the LRRII subfamily and examined their capability to interact with NSP of *Tomato Yellow Spot Virus* (ToYSV) similar to the interaction observed with Arabidopsis *NIK* genes [31]. Yeast two-hybrid experiments were performed using the ToYSV-NSP (accession number: YP\_459917.1) as prey and kinase domains of *SINIK1* (*Solyc02g089550*), *SINIK2* (*Solyc04g005910*), *SINIK3* (*Solyc04g039730*) and *SIBAK1* (*Solyc10g047140*) as bait. We also analyzed a PERK representative (*Solyc12g007110*, *SINsAK*) that is similar to the NSP-interactor PERK-like gene of Arabidopsis (*At5g24550*, *AtNsAK*). A tomato gene (*Solyc03g019980*) from the LRRXII subfamily, homolog of the Arabidopsis EF-Tu receptor (*AtEFR*), was used as a negative control. Interactions between the viral NSP and host proteins were detected after co-transforming the yeast cells with both bait and prey plasmids and monitoring for histidine prototrophy. NSP was found to interact with the kinase domains of *SINIK1*, *SINIK2*, *SINIK3* and *SIBAK1* or with the kinase domain of the PERK representative *SINsAK* (Figure 7, upper panel). The NSP interactions were specific to the tomato LRRII-RLK orthologs and to PERK-like *SINsAK* because the *HIS* marker gene was not activated in yeast cells co-transformed with TYNSP-p22 (pAD-NSP) and with either the empty vector or *SIEFR-p32* (EFR kinase domain). Furthermore, co-transformation of yeast with the NSP interactors fused to the GAL4-binding domain and the empty vector expressing the GAL4-activating domain alone also failed to activate the *HIS* marker gene (Figure 7, lower panel). These results expanded our previous observation that NSP-NIK complex formation was neither virus-specific nor host-specific [31,43]. They also suggest that *SINsAK* is a NSP target during begomovirus infection in tomato. Certainly, the *in vivo* demonstration of these interactions will further support these interpretations.

#### Discussion

##### The structure of the tomato RLK superfamily and the proposed evolution of RLK superfamily in plants

To date, the phylogenetic and structural characterization of the RLK superfamily has been limited to the following plant species: moss, rice, poplar, soybean and Arabidopsis [5,34,35]. The size of these families ranges from 300 to

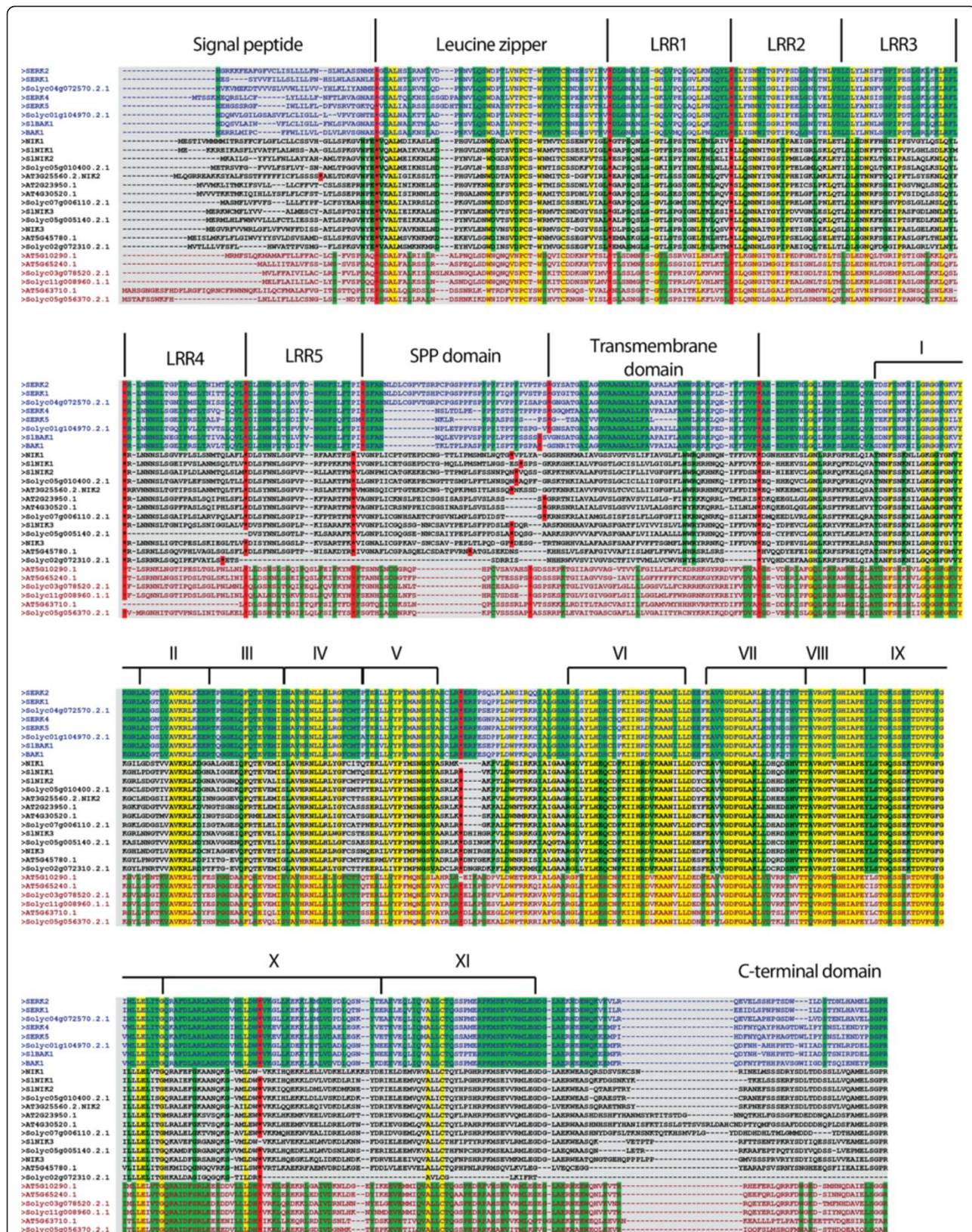


1200 proteins (Figure 8), and their extracellular regions bear a great variety of protein domain structures. In the present investigation, we characterized and generated a complete list of the tomato RLK superfamily members (Additional files 2 and 5), identifying 647 RLKs, which falls in the size range of the Arabidopsis (623 RLKs) and soybean (605 RLKs) superfamilies.

Evolutionary analyses of the RLK superfamily has suggested that the RLK structure was established prior to the divergence of land plants from algae because proteins with RLK configurations were discovered in the unicellular algae *Chlamydomonas reinhardtii* [34]. Comparative analysis of RLKs among moss, rice, poplar and Arabidopsis revealed that the RLK superfamily underwent expansion in the beginning of the land plant lineage, after the divergence of angiosperm and bryophyte and independently during diversification of each angiosperm lineage. The most dramatic expansion was observed in the rice and poplar lineages, which have almost twice as many as RLK

members as Arabidopsis [34]. This evolutionary scenario has not been changed by inclusion of data regarding the soybean [35] and tomato (this work) RLK superfamily expansion (Figure 8).

The phylogenetic tree of the members of the RLK superfamilies in tomato and Arabidopsis revealed that most of the RLK subfamilies have maintained approximately the same number of RLK members between these species. Exceptions were observed for the DUF26, L-LEC, LRK10L2, SD1, SD2b, WAK, LRRIIa, LRRXII, LRRXIIb, RLCKXII/XIII subfamilies, in which specific and extensive expansion was observed in one of the two plant species, as well as for the LysMII, LRRIIa and RLCKXII, in which specific reduction was observed (Additional file 3). Functional annotations of some Arabidopsis RLKs and the number of genes in tandem repeat that compose those subfamilies indicated a predominance of genes clustered in tandem array and defense-related RLKs (Additional file 3). Some of those subfamilies, such as DUF26, L-LEC, SD1, SD2b,



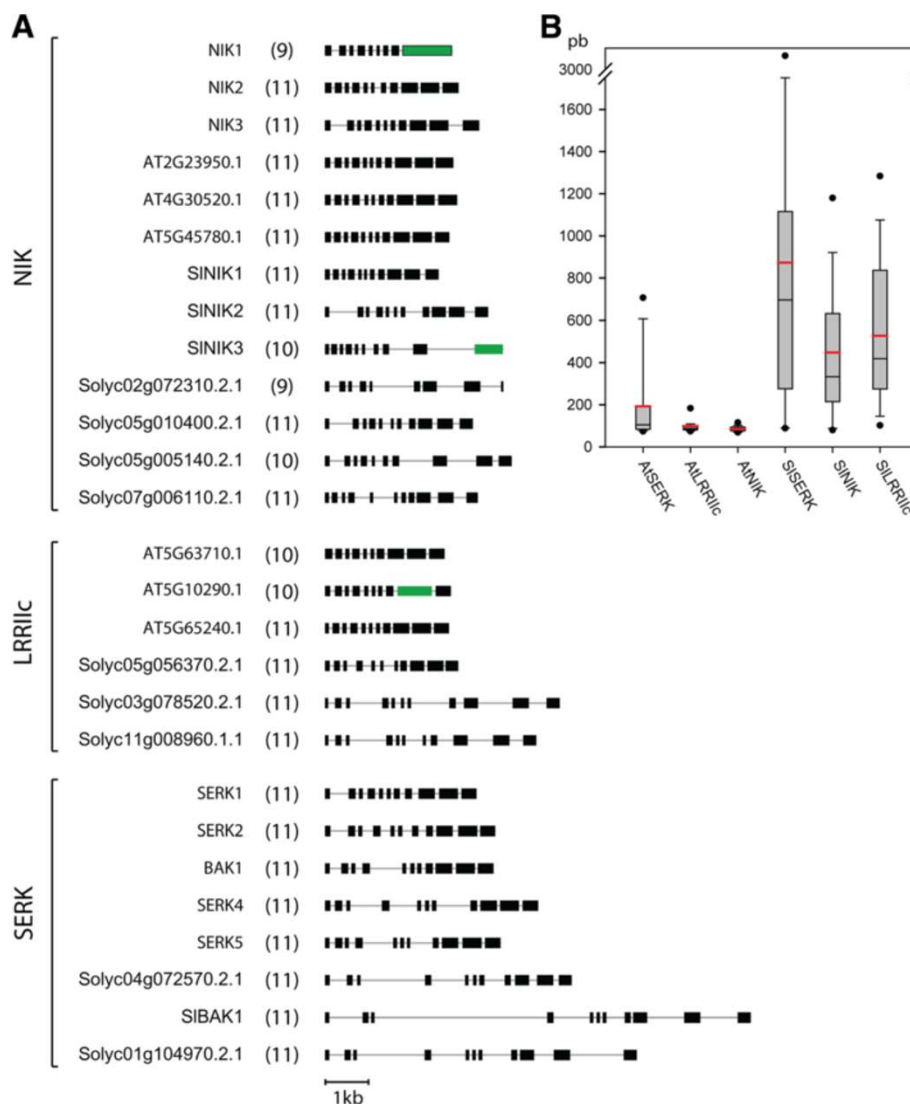


(See figure on previous page.)

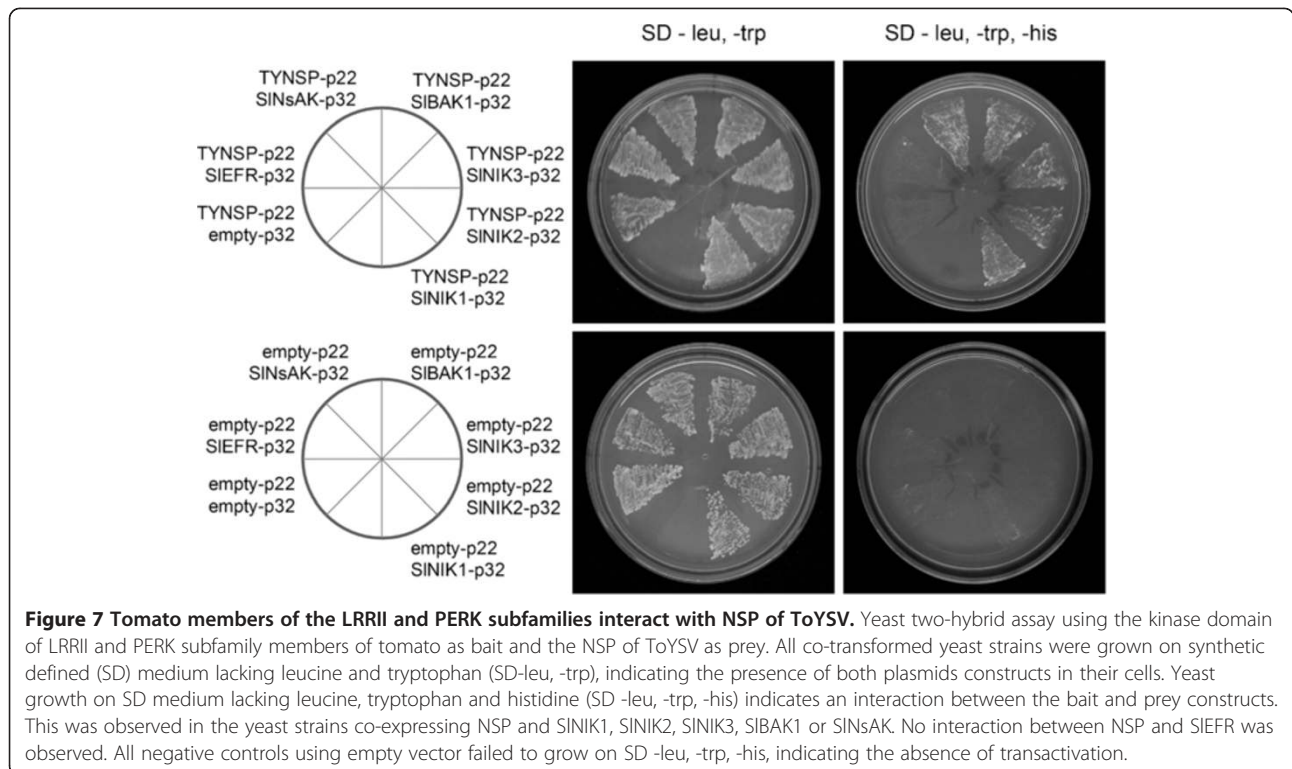
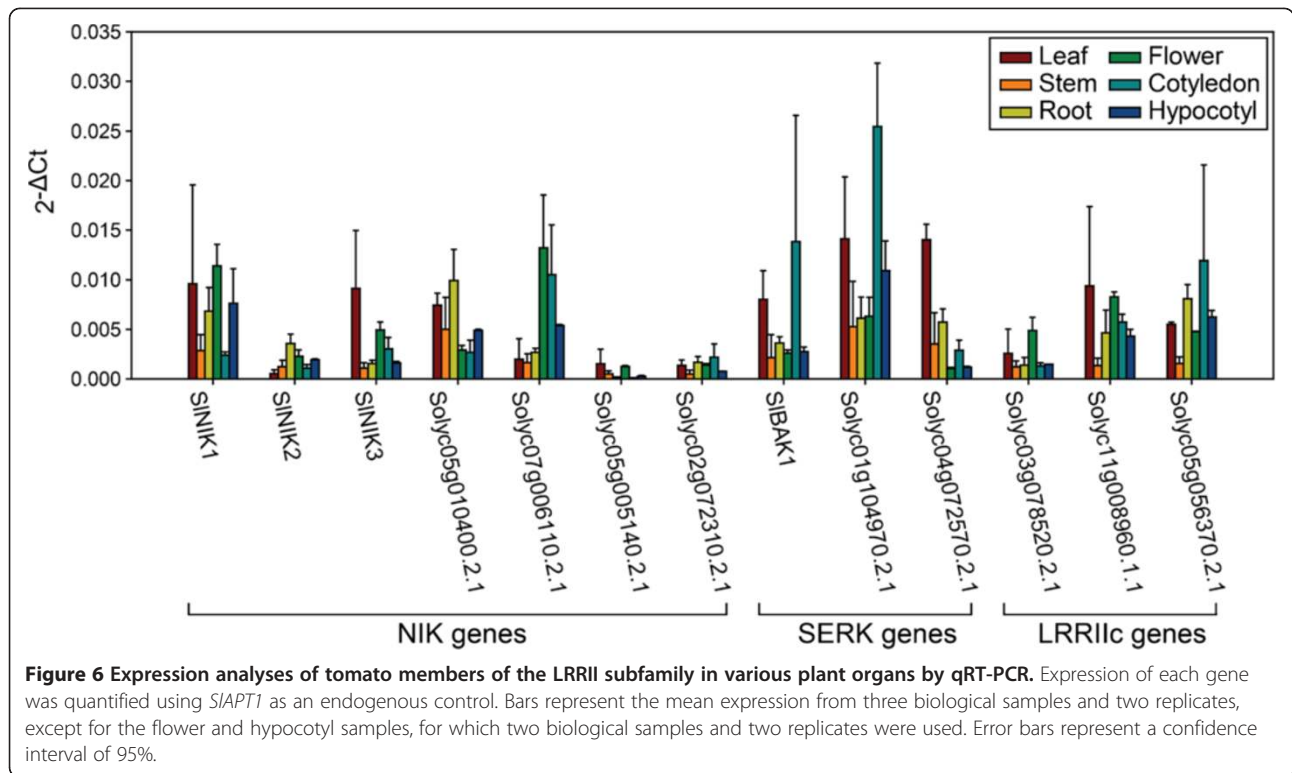
**Figure 4 Full-length sequence alignment of LRRII subfamily members of tomato and Arabidopsis demonstrate sequence and structure conservation.** Sequences of SERK, NIK and LRRIIc clades members are represented with blue, black and red letters, respectively. Yellow sites represent conserved sites in all sequences, and green sites represent conserved sites in each clade. Red sites represent the exon-exon junctions. Domain structures are indicated above the alignment. Roman numerals delimit the 11 subdomains of the kinase domain.

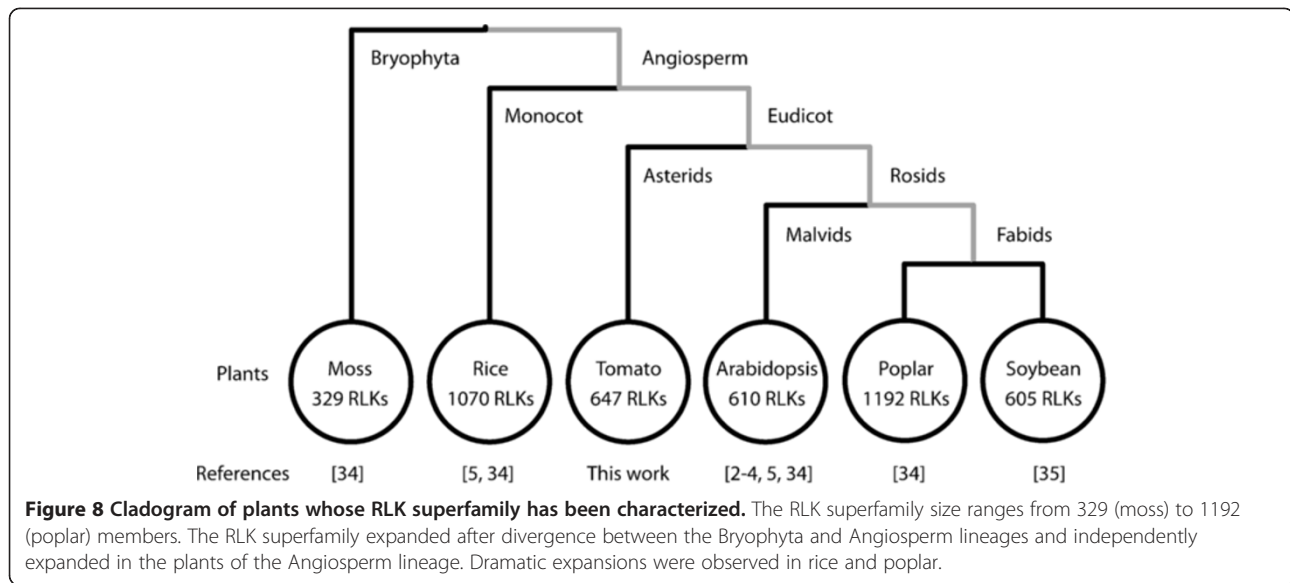
WAK and LRRIIa, had been shown previously to be overrepresented in microarray analysis of Arabidopsis under different stress conditions [34]. Taken together, these results are consistent with the previous assumption that specific expansions of RLK genes have occurred more frequently for those RLKs associated with defense response.

Specific expansion of tomato RLKs compared to Arabidopsis occurred in the LRRXII, SD1 and SD2b subfamilies. Interestingly, all of those RLK subfamilies, except for SD1, were also overrepresented in rice when compared with Arabidopsis [5]. The LRRXII subfamily comprises the EF-Tu RECEPTOR (EFR) and



**Figure 5 Genomic structure analysis of members of the LRRII subfamily of tomato and Arabidopsis. (A)** Exons are shown as dark boxes and introns as grey lines. Green boxes represent fused exons. The number between parentheses represents the number of exons in each gene. Almost all genes contain 11 exons. **(B)** Boxplot illustrating the distribution of intron length among LRRII clades of Arabidopsis and tomato. The red line marks the average intron length. Note the large length of intronic regions in SERK genes compared with genes from other clades, and in tomato sequences compared with Arabidopsis sequences.





FLAGELLIN-SENSITIVE 2 (FLS2) in Arabidopsis, and Xa21 in rice, all of which are associated with defense responses. The expansion of the LRRXII subfamily in cultivated plants such as tomato and rice has previously been suggested to be associated with the accumulation of resistance genes by intense breeding programs [5]. Likewise, the SD subfamily also has representatives involved in defense response, such as RK1 and RK3 [57], but this subfamily is also strongly associated with the self-incompatibility process [58]. Arabidopsis is known to be self-compatible, while rice, tomato and the close relative of *A. thaliana*, *A. lyrata*, are self-incompatible. Thus, the specific deletion of some representatives in *A. thaliana* could have contributed to the generation of the self-compatible mode in this species. In contrast, some RLK subfamilies were expanded specifically in Arabidopsis. Those include DUF26, L-LEC, LRK10L2, SD1, WAK, LRR1a and RLCKXII/XIII. Except for SD1 and LRK10L2 subfamilies, no significant expansion were observed in these RLK subfamilies in tomato, although all of them contain RLKs that are involved in defense response, such as FLS2-INDUCED RECEPTOR-KINASE 1 (FRK1, LRR1a) [59], LECTIN RECEPTOR KINASE 1.9 (LecRK-1.9, L-LEC) [60], PATHOGENESIS-RELATED 5-LIKE RECEPTOR KINASE (PR5K, LRK10L2) [61], RESISTANT TO FUSARIUM OXYSPORIUM 1 (RFO1, WAK) [62] and CYSTEINE-RICH RLK 5 (CRK5, DUF26) [63].

Another relevant distinction between tomato and Arabidopsis RLK subfamilies derives from the diversification of the extracellular domain patterns of the LRK10L2 subfamily representatives. Arabidopsis members of the LRK10L2 subfamily have unique domain structures, such as GDPD, thaumatin and malectin domains, while

the tomato members do not possess any characterized domain structure in their extracellular region. Likewise, these domain structures have not been found in rice, poplar or moss RLKs, indicating that within the RLK superfamily they are specific to Arabidopsis. We also identified a tomato-specific RLK that possesses an amino oxidase domain in its extracellular region. These RLKs may respond to molecular signals not perceived by other plants. Although gaining a novel protein structure could increase the repertoire of signals perceived by plants, the small number of lineage-specific RLKs in tomato, as was also reported in rice and poplar, further substantiates the hypothesis that the expansion of existing RLK kinase subfamilies is the major mechanism of evolution of these proteins.

The members of the RLK superfamily are involved in diverse biological processes at all steps of plant development. Thus, the gain or loss of a RLK gene could have serious repercussions on plant phenotype. The specific profiles of the RLK superfamily found in tomato and Arabidopsis are certainly responsible for several differences between these plants, such as morphology, reproduction and, importantly, responsiveness to different stress conditions. Among tomato and Arabidopsis, few RLK subfamilies have undergone specific expansion or reduction after their speciation. This scarcity may indicate that variation in RLK superfamily profiles in both plants appeared recently. The domestication process that tomatoes underwent could have been a significant factor contributing to this variation because some RLKs have been directly linked to traits targeted by artificial selection, such as disease resistance and growth. Nevertheless, most of the specific expansions of RLK subfamilies observed in Arabidopsis were also associated with

defense response and had occurred at similar rates (287 and 259 duplication events in Arabidopsis and tomato, respectively). These data suggest that, independently of artificial selection, Arabidopsis had also expanded and developed a specific machinery against abiotic or biotic stress response, which argues against the assumption that artificial selection leads to resistant genes accumulation. To further examine the influence of artificial selection on the repertoire of plant RLKs, new genetic resources for closely related wild plants are necessary.

### Functional expression analysis of the LRR II subfamily members in tomato and Arabidopsis

The LRRII subfamily contains RLKs with dual functions in development and defense response [52-54]. Characterized members of this subfamily include (i) *SERK* genes, which are associated with diverse processes, such as brassinosteroid signaling, flagellin, cell death, light and pathogen-associated molecular pattern (PAMP) responses [53], and (ii) *NIK* genes, which interact directly with geminivirus NSP during viral infection [31,32]. Phylogenetic and protein structure analyses on LRRII subfamily members of tomato and Arabidopsis demonstrated that this group is highly conserved between these species. In rice, in which the RLK superfamily has undergone a large expansion, the LRRII subfamily members are also conserved in number and sequence, indicating that biochemical pathways regulated by LRRII-RLKs have essential and conserved roles in angiosperm species.

Although LRRII members have well-conserved amino acid sequences among various species [64], expression analysis of the members of the tomato and Arabidopsis LRRII subfamilies demonstrated that only a few of the orthologous pairs resemble in their expression profiles. By analogy with some evidence in the literature from other plant species, one may envision that these orthologous genes could have functionally diverged after the speciation event separating tomato and Arabidopsis. Functional divergence in orthologous genes is not an uncommon event in both plants [65,66] and animals [67,68]. For example, the CRABS CLAW transcription factor in Arabidopsis is expressed in the carpel primordial abaxial region and in floral nectarines and regulates carpel morphology and nectar development, whereas its orthologous in rice, DROOPING LEAF (DL), is expressed in the whole carpel primordium and in central undifferentiated cells of leaves, where it regulates carpel identity and midrib development [66]. The expression of orthologous genes has also been shown to vary differently in response to a stress condition. In barley, which is tolerant to salinity, the expression of genes involved in root development, such as CONSTANS-LIKE 3 (COL3), is suppressed by high salinity, whereas the expression of the rice orthologous is unchanged under the

same stress condition [65]. Likewise, a large fraction of orthologous pairs of rice and Arabidopsis genes with receptor activity do not display conserved co-expression [69]. Therefore, different patterns of expression between the orthologous genes in the LRRII subfamilies from tomato and Arabidopsis may be a result of functional divergence that occurred between these genes. Functional divergence in receptor proteins with a developmental function may lead to a dramatic change in the plant phenotype because plant development is heavily guided by external signals. For example, a tissue that displays high expression of certain RLKs is likely to be more sensitive to perception of RLK-specific sensing signals, leading to a rapid and effective response. In contrast, reduced expression of an orthologous gene from a different species in the same tissue would decrease the effectiveness and delay the signal perception and response. This difference in the cell responsiveness to a specific signal could represent the differential timing of biochemical reactions that are regulated by this signal. In developmental process, small differences in reaction time may be sufficient to generate a distinct phenotype in the plant. In contrast, the orthologous pairs *SERK1/SERK2/Solyc04g072570.2.1*, which display similar expression profiles (Figure 3), also contain the most conserved extracellular and intracellular domains (approximately 80% and 93% of sequence identity, respectively). The other two orthologous pairs, *NIK1/SINIK1* and *NIK3/SINIK3*, which also displayed similar expression profiles, also had highly conserved extracellular and intracellular regions. In both orthologous pairs, sequence identity was approximately 65% in the extracellular regions and approximately 80% in the intracellular regions (Additional file 6). This finding may indicate a tight conservation of function between these members of the Arabidopsis and tomato LRRII RLK subfamilies.

### Conservation of geminivirus interactions with members of the RLK family in tomato

Although most of LRRII subfamily orthologous pairs exhibited functional divergence, we showed that the tomato orthologs of the LRRII-RLKs members *NIK1*, *NIK2* and *NIK3* retain the capacity to interact with geminivirus NSP in yeast (Figure 7) [31]. At least for the *NIK1* and *NIK3* ortholog pairs, the functional conservation associated with specific protein-protein interactions may be linked to the high conservation of their NSP-interacting kinase domain (approximately 80% sequence identity, Additional file 6) and similarity of expression profiles (Figure 3). The current model of *NIK*-mediated defense response posits that the immune receptor protects plant against geminiviruses by phosphorylating the ribosomal protein L10 (rpL10) [32,70]. Phosphorylation of rpL10 by *NIK* redirects the ribosomal protein to the

nucleus, where it may mount a defense mechanism to prevent viral proliferation. During geminivirus infection, NSP interacts with the kinase domain of NIKs to inhibit their kinase activity, preventing activation of the defense response. Despite the high similarity between *NIK* genes and *SERK* genes, AtBAK1/SERK3 and AtSERK1 do not functionally replace the AtNIK1 role in transducing an antiviral signaling response and do not interact with the viral NSP [31,32]. In contrast, we found that the *AtBAK1* ortholog from tomato interacts with NSP in yeast. Although the functional relevance of this interaction *in planta* remains to be determined, it is worth noting that the expression profiles of the *BAK1* orthologs are not similar, as would be expected for functionally divergent orthologs. Although both orthologs are ubiquitously expressed in the cognate plant species, they are expressed to different extents in distinct organs. Whereas AtBAK1 expression is quantitatively similar and relatively low in all organs analyzed, its ortholog from tomato displays a higher level of expression in the cotyledons, hypocotyls and leaves, where geminivirus infection largely takes place. Therefore, the expression profiles of the NSP interactors (SINIKs and SIBAK1) seemed to be related with the onset of geminivirus infection. Due to the high expression of the AtBAK1 tomato orthologous in leaves, one may envision the existence of evolving selective pressures to diverge the corresponding NSP-interacting domains of the BAK1 orthologs towards functional fitness with regard to geminivirus infection.

In contrast to NIK receptors, which are inhibited by NSP interaction, AtNsAK, a member of PERK subfamily, interacts with NSP and phosphorylates the viral protein *in vitro* [56]. Loss of *nsak* function enhances tolerance to geminivirus infection, indicating that AtNsAK is a positive contributor to geminivirus infection in Arabidopsis. Here, we showed that the NsAK tomato orthologous retains its capacity to interact with viral NSP. This demonstrates that specific members of the RLK family have conserved defense functions (such as NIKs) or compatibility functions (such as NsAK) in response to viral infection. Due to the emergence of new species of tomato-infecting begomoviruses that rapidly evolve through recombination or pseudo-recombination to produce divergent genome sequences that gives the virus an advantage over its host's recognition system, a survey of the interactions between NSPs from distinct tomato-infecting geminiviruses and SINIKs and SINsAK may add insights into the co-evolution of the viral protein and host defense/compatibility functions.

## Conclusions

The RLK superfamily is a large and diverse group of transmembrane receptors that enables plants to perceive a

diverse array of signals at the cell surface, creating an efficient mechanism for cell-environment communication. In this investigation, we generated a complete list of the members of the tomato RLK superfamily, which is made-up of 647 proteins. The tomato RLK sequences exhibited a typical receptor-like kinase configuration and almost all of them were phylogenetically clustered with at least one member of the Arabidopsis RLK superfamily. Therefore, the tomato RLK superfamily is similarly organized, with the same number and identity of subfamilies as previously defined for Arabidopsis RLKs. Among the 58 RLK subfamilies, twelve showed specific and extensive expansion or reduction in the number of their RLK members, which may be a reflection of lineage-specific responses to various biotic and abiotic stresses. The intense breeding programs tomatoes have been subjected to may also have contributed to the establishment of the current RLK superfamily profile in this species. This comprehensive analysis comparing the complete repertory of Arabidopsis and tomato RLKs may provide a framework to rationalize future functional studies of the members of this family.

Phylogenetic and structural analyses of LRRII subfamily members from both tomato and Arabidopsis reveal a well-conserved group both in terms of sequence and protein domain organization. As a consequence, the tomato LRRII-RLK subfamily is organized into the same three with phylogenetically supported clades, SERK, NIK and LRRIIc clusters. Nevertheless, a comparison of the expression between orthologous genes of this subfamily demonstrated that the majority of the orthologous pairs did not share a similar expression profile, indicating that these orthologous LRRII-RLKs may have undergone functional divergence. This finding is supported by the observation that, in contrast to the Arabidopsis AtBAK1, SIBAK1 interacts with the geminivirus NSP and is highly expressed in leaves and the cotyledon. This pattern of SIBAK1 expression is consistent with the pattern of infection by tomato-infecting begomoviruses, which infect leaf tissues and move through the phloem but do not invade roots. Additionally, as immune receptors, the orthologous pairs NIK1 and NIK3 displayed both the capacity to interact with the begomovirus virulence factor NSP and expression profiles that parallel the onset of begomovirus infection. Evidence for functional conservation between NIK1 orthologs has been previously provided with the demonstration that NIK1 from Arabidopsis is capable of protecting tomato plants against tomato-infecting begomovirus [70]. Collectively, our results indicate that NIK orthologs retain similar functions as defense receptors to protect plant cells against viral attack. Therefore, NIK-mediated antiviral signaling likely also operates in tomato, suggesting that the tomato NIKs may be good candidate targets for engineering resistance against tomato-infecting begomoviruses.

## Methods

### Identification and classification of tomato RLKs

Tomato RLK proteins were retrieved through a batch BLAST analysis (blastp, e-value cutoff = 0.01) [71] using an *A. thaliana* representative of each subfamily of the RLK superfamily against a protein database of tomato (iTAGv2.3) available on the Sol Genomics Network website (solgenomics.net) [72]. Through this procedure, 955 predicted proteins were retrieved and annotated using SMART (smart.embl-heidelberg.de) [46] and Pfam (pfam.sanger.ac.uk) [47] databases. Among these proteins, 951 contained a predicted kinase domain and hence were considered to be putative RLKs. The sequences of the kinase domains of Arabidopsis RLKs, previously described in [5], and tomato putative RLKs were submitted to sequence alignment and tree reconstruction using ClustalW (v. 2.0.12) [73] and FastTree (v. 2.1.4) [74], respectively (Figure 1 and Additional file 2) using default parameters. The kinase domain of other kinase protein families from *A. thaliana* and human were used as outgroups [3,75]. The accession numbers for all outgroup members are reported in Additional file 1. Those proteins that clustered with outgroup members were not considered to be RLKs and were discarded from further analysis. Additionally, short putative RLKs were deleted manually from the analysis. The identified RLK-related tomato sequences comprised a list of 647 members. Tomato RLKs that clustered with *A. thaliana* RLK subfamily members, as defined previously in [5], were classified as members of the same subfamily. Phylogenetic trees (Figure 1 and Additional file 2) and protein schemes (Additional file 2) were generated using iTOL tool (itol.embl.de) [76].

### Inference on duplication/deletion events, identification of RLKs in tandem repeats and functional categorization of RLKs subfamilies

Number of members in the common ancestral of Arabidopsis and tomato and the occurrence of gene duplication and deletion were inferred by reconciliation methods implemented in Notung (v.2.6) [48]. For this analysis, we used the RLK superfamily tree, showed in Figure 1 and in Additional file 2, as gene tree. The species tree was generated at NCBI taxonomy browser (www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi). For identification of RLKs in tandem repeats, we considered that two genes are clustered in tandem array when i) they are classified in the same subfamily, ii) they are distant from each other by less than 100kb and iii) they are separated by less than 10 genes from each other, as previously described in [34]. For identification of defense- or development-related genes, we used the GO terms associated to the Arabidopsis genes. Arabidopsis RLKs that had GO terms related to "response to stress" (GO:0006950) and/or "developmental process"

(GO:0032502) and their child terms were classified as defense- and/or developmental-related, respectively.

### Statistical test for expansion/reduction analysis and functional categorization of RLK subfamilies

To statistically verify if (i) RLK subfamilies have differentially expanded or reduced in their size, (ii) tandem duplications or (iii) a functional annotation (defense or development) are more often in a specific RLK subfamily, we used the test of equal or given proportions [77]. This statistical analysis tests if two different proportions ( $p_1$  and  $p_2$ ) are equal ( $H_0: p_1 = p_2$ ) or different ( $H_a: p_1 \neq p_2$ ). The two tested proportions were the occurrence of a given feature (number of duplication/deletion, tandem repeats or genes annotated as defense- or developmental-related) in a subpopulation (RLK subfamily,  $p_1$ ), and the proportion of the number of the same feature in the whole population (RLK superfamily,  $p_2$ ). As we analyzed whether those features were overrepresented in a given RLK subfamily, our alternative hypothesis was  $H_a: p_1 > p_2$ . Test calculations were performed in R environment. All p-values associated with tested values are summarized in Additional file 4.

### Motif prediction, genomic structure and phylogenetic analysis of the LRRII subfamily

Full-length amino acid sequences of members of the LRRII subfamily from tomato and Arabidopsis were aligned using ClustalW (v. 2.0.12) [73] using the default parameters. A phylogenetic tree was constructed using the maximum likelihood method (JTT model, bootstrap replicates = 1000) implemented in MEGA5 software [78]. Motif, signal peptide and transmembrane prediction were carried out using Pfam [47] and SMART [46] databases. The genomic structure of the LRRII subfamily members of tomato and Arabidopsis was determined by aligning the coding sequence (CDS) of each gene with genomic sequences of the respective organism. The alignment was carried out using the BLAST algorithm (blastn) [71] with high-stringency parameters. Amino acid, CDS and genomic sequences for tomato and Arabidopsis were retrieved from the Sol Genomics Network (solgenomics.net) [72] and TAIR (www.arabidopsis.org) [79] websites, respectively.

### Protein-protein interaction assays

The analysis of protein-protein interactions between viral NSP and the kinase domain of tomato RLKs was performed using the Proquest Yeast Two-Hybrid system with Gateway Technology (Invitrogen Inc.). The tomato RLKs that presented the highest identity with AtNIK1 (At5g16000), AtNIK2 (At3g25560), AtNIK3 (At1g60800), AtBAK1 (At4g33430) and AtNsAK (At5g24550) were selected for the assay. These tomato proteins are referred to as SINIK1 (Solyc02g089550), SINIK2 (Solyc04g005910),

SINIK3 (Solyc04g039730), SIBAK1 (Solyc10g047140) and SINsAK (Solyc12g007110). As a negative control, we used the kinase domain of the tomato RLK that displayed the highest identity with AtEFR (At5g20480), referred to as SIEFR (Solyc03g019980).

The NSP coding region was amplified from ToYSV (Tomato Yellow Spot Virus-Geminiviridae, *Begomovirus*) [80] using gene-specific primers with appropriate extensions for cloning via the Gateway system, as described in Additional file 7. The amplified fragment was cloned into pDONR201 to generate pUFV1780.1 and then transferred by recombination to pDEST22 yielding pUFV1781, also designated as TYNSP-p22.

For amplification of the C-terminal kinase domain of the tomato RLKs, we prepared cDNA from cotyledons of wild-type tomato plants (var. Santa Clara). Briefly, total RNA from tomato cotyledons was isolated using an RNeasy Kit (Qiagen Inc.). First-strand cDNA was synthesized from 1 µg of total RNA using the M-MLV Reverse Transcriptase (Invitrogen Inc.) according to the manufacturer's instructions. Primers used in the amplification step were designed with recombination sites for further cloning procedures using the Gateway System (Invitrogen Inc.). The primers used are listed in Additional file 7. PCR assays were performed using Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen Inc.) according to the manufacturer's instructions. The amplified fragments were cloned into the entry vector pDONR201 (Invitrogen Inc.) and sequenced. The resulting vectors were the following: pUFV1756.1, pUFV1596, pUFV1757.1, pUFV1734.2, pUFV1744.1 and pUFV1955.2, corresponding, respectively to the fragment encoding the kinase domain of SINIK1, SINIK2, SINIK3, SIBAK1, SINsAK and SIEFR. Then, the cloned fragment in pDONR201 was transferred to pDEST32, which contains the DNA-binding domain of the GAL4 promoter (Invitrogen Inc.). This procedure resulted in the following recombinant plasmids: pUFV1768.1, pUFV1760.1, pUFV1779.1, pUFV1769.1, pUFV1770.1 and pUFV1975.1, also designated as SINIK1-p32, SINIK2-p32, SINIK3-p32, SIBAK1-p32, SINsAK-p32 and SIEFR-p32, respectively.

Competent cells of yeast strain AH109 (Clontech Inc., genotype: *MATα*, *trp1-901*, *leu2-3,112*, *ura3-52*, *his3200*, *gal4Δ*, *gal80Δ*, *LYS2::GAL1<sub>UAS</sub>-GAL1<sub>TATA</sub>-HIS3*, *GAL2<sub>UAS</sub>-GAL2<sub>TATA</sub>-ADE2*, *URA3::MEL1<sub>UAS</sub>-MEL1<sub>TATA</sub>-lacZ*) were sequentially co-transformed with TYNSP-p22 and with one of the pDEST32 constructs. Co-transformed yeasts were plated onto synthetic dropout medium lacking leucine, tryptophan and histidine, and incubated at 28°C. Yeast growth was monitored for 5 days.

#### Expression analysis of the LRRII subfamily genes

The expression patterns of genes in the LRRII subfamily were assayed by quantitative Real-Time PCR (qRT-PCR)

in various tomato tissues. Wild-type tomato plants (var. Santa Clara) were cultivated in a greenhouse for 45 days after germination. Leaves, stems, roots and flowers from three plants were collected separately. We also cultivated plants in half-strength Murashige and Skoog medium (1/2 MS, Sigma-Aldrich Co.) for 10 days after germination under normal conditions to collect cotyledons and hypocotyls tissue. For these tissues, due to the small amount of material, each sample represented a pool of three young plants. Total RNA from each sample was extracted using TRIzol (Invitrogen Inc.), and the quality and integrity of extracted RNA were monitored by spectrophotometry and electrophoresis. For cDNA synthesis, 3 µg of total RNA from each sample was first treated with RNase-free DNase I (Promega Inc.) and then reverse-transcribed using M-MLV Reverse Transcriptase (Invitrogen Inc.) and oligo-dT primers. qRT-PCR assays were performed using an ABI7500 *Real Time PCR System* (Applied Biosystems Inc.) and *SYBR<sup>®</sup> Green PCR Master Mix* (Applied Biosystems Inc.). The amplification reactions were performed using default parameters for thermal cycling (50° for 10 min, 95° for 1 min, followed by 40 cycles of 95° for 15 sec and 60° for 1 min). Primers were designed using PerlPrimer [81], attempting to choose primer pairs in which at least one of them extended across an intron-exon boundary. Expression quantification of each gene was determined according to the Ct relative quantification method ( $2^{-\Delta C_t}$ ) [82] using *SIAPT1* (adenine phosphoribosyl transferase, Solyc04g077970.2.1) as an endogenous control for data normalization. Expression data from Arabidopsis were obtained from the AtGenExpress website ([jsp.weigelworld.org/expviz/expviz.jsp](http://jsp.weigelworld.org/expviz/expviz.jsp)) [55].

#### Additional files

**Additional file 1: List of outgroup proteins.** Summary of the names and accession numbers of proteins used as outgroups in the phylogenetic tree of Figure 1 and Additional file 2.

**Additional file 2: RLK Phylogenetic tree of tomato and Arabidopsis.** This is the same phylogenetic tree as presented in Figure 1, but displayed in more details. It contains additionally the accession numbers and schemes of the domain structures of each protein that composes the tree. Tomato proteins are represented by red branches and Arabidopsis proteins by blue branches. The local support values at the nodes were computed by resampling the site likelihoods 1,000 times and performing the Shimodaira-Hasegawa test.

**Additional file 3: Expansion/reduction in Arabidopsis and tomato RLK subfamilies and functional inference.** The membership size of RLK subfamilies in Arabidopsis (At) and tomato (Sl) is indicated. Values in bold and with asterisks indicate statistical significance by the test of equal or given proportions ( $\alpha=0.05$ ). Subfamilies with significantly large proportion of duplication (dup.) or deletion (del.) were considered to have specifically expanded or reduced respectively after the divergence of Arabidopsis and tomato species. Subfamilies that presented statistically large proportion of RLKs organized in tandem repeats (t.r.) and/or of RLKs functionally annotated in defense response (def.) category were considered to be defense-related (red arrows). Conversely, subfamilies

with significantly large proportion of members annotated in developmental process (dev.) category were classified as development-related (blue arrows). Green arrow indicates the LRRIL subfamily that presented large proportion in both functional categories. Legend: dup.: duplication events; del.: deletion events.

**Additional file 4: Analyses on expansion/reduction in Arabidopsis and tomato RLK subfamilies and on their functional inference.** The table contains information from Additional file 3 and presents the associated p-value from each test performed.

**Additional file 5: List of Arabidopsis and tomato RLKs and their respective RLK subfamilies.** Summary of all RLK IDs presented in the tree of Additional file 2.

**Additional file 6: Sequence identity between members of LRRIL-RLK subfamily of tomato and Arabidopsis.** (A) Full-length amino acid sequences, (B) intracellular and (C) extracellular regions of LRRIL subfamily members were aligned using CLUSTALW. Thick lines delimit the sequence comparison between members of the same clade (NIK, SERK, LRRILc). Blue cells indicate high sequence identity, whereas red cells denote low sequence identity.

**Additional file 7: List of primers used for yeast two-hybrid assay and for expression analysis by real-time PCR analysis.** Summary of all primers used for gene cloning and real-time PCR experiments.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

The project was coordinated by EPBF. OJBB and TS performed the bioinformatics and phylogenetic analyses. AAS, MD and TS were involved in the gene expression assays. MD and TS carried out the yeast two-hybrid assays. FFS designed and performed the statistical analysis. EPBF, MD and TS prepared the manuscript. All authors read and approved the final manuscript.

#### Acknowledgements

We acknowledge the FAPEMIG, FINEP and CNPq for financial support and the teams of the Sol Genomics Network, TAIR and AtGenExpress for publicly providing the genome, proteome and expression data for tomato and Arabidopsis.

#### Author details

<sup>1</sup>National Institute of Science and Technology in Plant-Pest Interactions, Universidade Federal de Viçosa, 36570-000, Viçosa, MG, Brazil. <sup>2</sup>Departamento de Bioquímica e Biologia Molecular/BIOAGRO, Universidade Federal de Viçosa, 36570-000, Viçosa, MG, Brazil. <sup>3</sup>Departamento de Estatística, Universidade Federal de Viçosa, 36570-000, Viçosa, MG, Brazil.

Received: 30 July 2012 Accepted: 18 November 2012

Published: 2 December 2012

#### References

- Morris ER, Walker JC: Receptor-like protein kinases: the keys to response. *Curr Opin Plant Biol* 2003, **6**:339–342.
- Shiu SH, Bleecker AB: Plant receptor-like kinase gene family: diversity, function, and signaling. *Sci STKE* 2001, **2001**:re22.
- Shiu S-H, Bleecker AB: Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *PNAS* 2001, **98**:10763–10768.
- Shiu SH, Bleecker AB: Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in Arabidopsis. *Plant Physiol* 2003, **132**:530–543.
- Shiu S-H, Karlowski WM, Pan R, Tzeng Y-H, Mayer KFX, Li W-H: Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. *Plant Cell* 2004, **16**:1220–1234.
- Matsubayashi Y: Ligand-receptor pairs in plant peptide signaling. *J Cell Sci* 2003, **116**:3863–3870.
- Wang X, Goshe MB, Soderblom EJ, Phinney BS, Kuchar JA, Li J, Asami T, Yoshida S, Huber SC, Clouse SD: Identification and functional analysis of in vivo phosphorylation sites of the Arabidopsis BRASSINOSTEROID-INSENSITIVE1 receptor kinase. *Plant Cell* 2005, **17**:1685–1703.
- Clouse SD: Brassinosteroid signal transduction: from receptor kinase activation to transcriptional networks regulating plant development REVIEW. *Plant Cell* 2011, **23**:1219–1230.
- Clark SE: Organ formation at the vegetative shoot meristem. *Plant Cell* 1997, **9**:1067–1076.
- Li J, Chory J: A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* 1997, **90**:929–938.
- Kinoshita T, Caño-Delgado A, Seto H, Hiranuma S, Fujioka S, Yoshida S, Chory J: Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. *Nature* 2005, **433**:167–171.
- Li J, Wen J, Lease KA, Doke JT, Tax FE, Walker JC: BAK1, an Arabidopsis LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. *Cell* 2002, **110**:213–222.
- Nam KH, Li J: BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. *Cell* 2002, **110**:203–212.
- Caño-Delgado A, Yin Y, Yu C, Vafeados D, Mora-García S, Cheng J-C, Nam KH, Li J, Chory J: BRL1 and BRL3 are novel brassinosteroid receptors that function in vascular differentiation in Arabidopsis. *Development* 2004, **131**:5341–5351.
- Shpak ED, McAbee JM, Pillitteri LJ, Torii KU: Stomatal patterning and differentiation by synergistic interactions of receptor kinases. *Science* 2005, **309**:290–293.
- Fisher K, Turner S: PXY, a receptor-like kinase essential for maintaining polarity during plant vascular-tissue development. *Curr Biol* 2007, **17**:1061–1066.
- Hara K, Yokoo T, Kajita R, Onishi T, Yahata S, Peterson KM, Torii KU, Kakimoto T: Epidermal cell density is autoregulated via a secretory peptide, EPIDERMAL PATTERNING FACTOR 2 in Arabidopsis leaves. *Plant Cell Physiol* 2009, **50**:1019–1031.
- Hord CLH, Chen C, DeYoung BJ, Clark SE, Ma H: The BAM1/BAM2 receptor-like kinases are important regulators of Arabidopsis early anther development. *Plant Cell* 2006, **18**:1667–1680.
- Jia G, Liu X, Owen HA, Zhao D: Signaling of cell fate determination by the TPD1 small protein and EMS1 receptor kinase. *PNAS* 2008, **105**:2220–2225.
- Kwak S-H, Schiefelbein J: A feedback mechanism controlling SCRAMBLED receptor accumulation and cell-type pattern in Arabidopsis. *Curr Biol* 2008, **18**:1949–1954.
- Tanaka H, Watanabe M, Sasabe M, Hiroe T, Tanaka T, Tsukaya H, Ikezaki M, Machida C, Machida Y: Novel receptor-like kinase ALE2 controls shoot development by specifying epidermis in Arabidopsis. *Development* 2007, **134**:1643–1652.
- Nodine MD, Yadegari R, Tax FE: RPK1 and TOAD2 are two receptor-like kinases redundantly required for Arabidopsis embryonic pattern formation. *Dev Cell* 2007, **12**:943–956.
- Lau S, Slane D, Herud O, Kong J, Jürgens G: Early embryogenesis in flowering plants: setting up the basic body pattern. *Annu Rev Plant Biol* 2012, **63**:483–506.
- Nodine MD, Bryan AC, Racolta A, Jerosky KV, Tax FE: A few standing for many: embryo receptor-like kinases. *Trends Plant Sci* 2011, **16**:211–217.
- Muschietti J, Eyal Y, McCormick S: Pollen tube localization implies a role in pollen-pistil interactions for the tomato receptor-like protein kinases LePRK1 and LePRK2. *Plant Cell* 1998, **10**:319–330.
- Deeken R, Kaldenhoff R: Light-repressible receptor protein kinase: a novel photo-regulated gene from Arabidopsis thaliana. *Planta* 1997, **202**:479–486.
- Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C, Ronald P: A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* 1995, **270**:1804–1806.
- Gómez-Gómez L, Boller T: FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. *Mol Cell* 2000, **5**:1003–1011.
- Godiard L, Sauviac L, Torii KU, Grenon O, Mangin B, Grimsley NH, Marco Y: ERECTA, an LRR receptor-like kinase protein controlling development pleiotropically affects resistance to bacterial wilt. *Plant J* 2003, **36**:353–365.
- Llorente F, Carlos A-B, Sánchez-Rodríguez C, Jorda L, Molina A: ERECTA receptor-like kinase and heterotrimeric G protein from Arabidopsis are



- required for resistance to the necrotrophic fungus *Plectosphaerella cucumerina*. *Plant J* 2005, **43**:165–180.
31. Fontes EPB, Santos AA, Luz DF, Wacławovsky AJ, Chory J: **The geminivirus nuclear shuttle protein is a virulence factor that suppresses transmembrane receptor kinase activity.** *Genes Dev* 2004, **18**:2545–2556.
  32. Santos AA, Lopes KVG, Apfata JAC, Fontes EPB: **NSP-interacting kinase, NIK: a transducer of plant defence signalling.** *J Exp Bot* 2010, **61**:3839–3845.
  33. Dardick C, Chen J, Richter T, Ouyang S, Ronald P: **The rice kinase database. a phylogenomic database for the rice kinome.** *Plant Physiol* 2007, **143**:579–586.
  34. Lehti-Shiu MD, Zou C, Hanada K, Shiu S-H: **Evolutionary history and stress regulation of plant receptor-like kinase/pelle genes.** *Plant Physiol* 2009, **150**:12–26.
  35. Liu P, Wei W, Ouyang S, Zhang J-S, Chen S-Y, Zhang W-K: **Analysis of expressed receptor-like kinases (RLKs) in soybean.** *J Genet Genomics* 2009, **36**:611–619.
  36. Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganai MW, Spivey R, Wu T, Earle ED, Tanksley SD: **Map-based cloning of a protein kinase gene conferring disease resistance in tomato.** *Science* 1993, **262**:1432–1436.
  37. Zhou J, Loh Y-T, Bressan RA, Martin GB: **The tomato gene *Pti1* encodes a serine/threonine kinase that is phosphorylated by Pto and is involved in the hypersensitive response.** *Cell* 1995, **83**:925–935.
  38. Zeng L, Velásquez AC, Munkvold KR, Zhang J, Martin GB: **A tomato LysM receptor-like kinase promotes immunity and its kinase activity is inhibited by *AvrPtoB*.** *Plant J* 2012, **69**:92–103.
  39. Kim J-G, Li X, Roden JA, Taylor KW, Aakre CD, Su B, Lalonde S, Kirik A, Chen Y, Baranage G, McLane H, Martin GB, Mudgett MB: **Xanthomonas T3S effector *XopN* suppresses PAMP-triggered immunity and interacts with a tomato atypical receptor-like kinase and TFT1.** *Plant Cell* 2009, **21**:1305–1323.
  40. AbuQamar S, Chai M-F, Luo H, Song F, Mengiste T: **Tomato protein kinase 1b mediates signaling of plant responses to necrotrophic fungi and insect herbivory.** *Plant Cell* 2008, **20**:1964–1983.
  41. Yin Y, Wu D, Chory J: **Plant receptor kinases: systemin receptor identified.** *Proc Natl Acad Sci U S A* 2002, **99**:9090–9092.
  42. Malinowski R, Higgins R, Luo Y, Piper L, Nazir A, Bajwa V, Clouse S, Thompson P, Stratmann J: **The tomato brassinosteroid receptor *BR11* increases binding of systemin to tobacco plasma membranes, but is not involved in systemin signaling.** *Plant Mol Biol* 2009, **70**:603–616.
  43. Mariano AC, Andrade MO, Santos AA, Carolino SMB, Oliveira ML, Baracat-Pereira MC, Brommonschenkel SH, Fontes EPB: **Identification of a novel receptor-like protein kinase that interacts with a geminivirus nuclear shuttle protein.** *Virology* 2004, **318**:24–31.
  44. Zhang X, Choi J, Heinz J, Chetty C: **Domain-specific positive selection contributes to the evolution of Arabidopsis leucine-rich repeat receptor-like kinase (LRR RLK) genes.** *J Mol Evol* 2006, **63**:612–621.
  45. Consortium TTG: **The tomato genome sequence provides insights into fleshy fruit evolution.** *Nature* 2012, **485**:635–641.
  46. Letunic I, Doerks T, Bork P: **SMART 7: recent updates to the protein domain annotation resource.** *Nucleic Acids Res* 2011, **40**:D302–D305.
  47. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD: **The Pfam protein families database.** *Nucleic Acids Res* 2011, **40**:D290–D301.
  48. Stolzer M, Lai H, Xu M, Sathaye D, Vernot B, Durand D: **Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees.** *Bioinformatics* 2012, **28**:i409–i415.
  49. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The gene ontology consortium.** *Nat Genet* 2000, **25**:25–29.
  50. Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu S-H: **Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli.** *Plant Physiol* 2008, **148**:993–1003.
  51. Kondrashov FA: **Gene duplication as a mechanism of genomic adaptation to a changing environment.** *Proc R Soc B* 2012, **279**: doi:10.1098/rspb.2012.1108.
  52. Vert G: **Plant signaling: brassinosteroids, immunity and effectors are BAK!** *Curr Biol* 2008, **18**:R963–R965.
  53. Chinchilla D, Shan L, He P, de Vries S, Kemmerling B: **One for all: the receptor-associated kinase *BAK1*.** *Trends Plant Sci* 2009, **14**:535–541.
  54. Postel S, Küfner I, Beuter C, Mazzotta S, Schwedt A, Borlotti A, Halter T, Kemmerling B, Nürnberger T: **The multifunctional leucine-rich repeat receptor kinase *BAK1* is implicated in Arabidopsis development and immunity.** *Eur J Cell Biol* 2010, **89**:169–174.
  55. Kilian J, Whitehead D, Horak J, Wanke D, Weini S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter K: **The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses.** *Plant J* 2007, **50**:347–363.
  56. Florentino LH, Santos AA, Fontenelle MR, Pinheiro GL, Zerbini FM, Baracat-Pereira MC, Fontes EPB: **A PERK-like receptor kinase interacts with the geminivirus nuclear shuttle protein and potentiates viral infection.** *J Virol* 2006, **80**:6648–6656.
  57. Pastuglia M, Swarup R, Rocher A, Saindrenan P, Roby D, Dumas C, Cock JM: **Comparison of the expression patterns of two small gene families of S gene family receptor kinase genes during the defence response in Brassica oleracea and Arabidopsis thaliana.** *Gene* 2002, **282**:215–225.
  58. Nasrallah JB: **Cell-cell signaling in the self-incompatibility response.** *Curr Opin Plant Biol* 2000, **3**:368–373.
  59. He P, Shan L, Lin N-C, Martin GB, Kemmerling B, Nürnberger T, Sheen J: **Specific bacterial suppressors of MAMP signaling upstream of MAPKKK in Arabidopsis innate immunity.** *Cell* 2006, **125**:563–575.
  60. Bouwmeester K, de Sain M, Weide R, Gouget A, Klamer S, Canut H, Govers F: **The lectin receptor kinase *LecRK-1.9* is a novel phytophthora resistance component and a potential host target for a RXLR effector.** *PLoS Pathog* 2011, **7**:e1001327.
  61. Wang X, Zafian P, Choudhary M, Lawton M: **The PR5K receptor protein kinase from Arabidopsis thaliana is structurally related to a family of plant defense proteins.** *Proc Natl Acad Sci U S A* 1996, **93**:2598–2602.
  62. Diener AC, Ausubel FM: **RESISTANCE TO FUSARIUM OXYSPORUM 1, a dominant Arabidopsis disease-resistance gene, is not race specific.** *Genetics* 2005, **171**:305–321.
  63. Chen K, Du L, Chen Z: **Sensitization of defense responses and activation of programmed cell death by a pathogen-induced receptor-like protein kinase in Arabidopsis.** *Plant Mol Biol* 2003, **53**:61–74.
  64. Nolan KE, Kurdyukov S, Rose RJ: **Characterisation of the legume SERK-NIK gene superfamily including splice variants: Implications for development and defence.** *BMC Plant Biol* 2011, **11**:44.
  65. Walia H, Wilson C, Ismail A, Close T, Cui X: **Comparing genomic expression patterns across plant species reveals highly diverged transcriptional dynamics in response to salt stress.** *BMC Genomics* 2009, **10**:398.
  66. Nakayama H, Yamaguchi T, Tsukaya H: **Expression patterns of *AaDL*, a CRABS CLAW ortholog in Asparagus asparagoides (Asparagaceae), demonstrate a stepwise evolution of CRC/DL subfamily of YABBY genes.** *Am J Bot* 2010, **97**:591–600.
  67. Huminiecki L, Wolfe KH: **Divergence of spatial gene expression profiles following species-specific gene duplications in human and mouse.** *Genome Res* 2004, **14**:1870–1879.
  68. Gharib WH, Robinson-Rechavi M: **When orthologs diverge between human and mouse.** *Brief Bioinform* 2011, **12**(5):436–441.
  69. Movahedi S, Van de Peer Y, Vandepoele K: **Comparative network analysis reveals that tissue specificity and gene function are important factors influencing the mode of expression evolution in Arabidopsis and Rice1 [W].** *Plant Physiol* 2011, **156**:1316–1330.
  70. Carvalho CM, Santos AA, Pires SR, Rocha CS, Saraiva DI, Machado JPB, Mattos EC, Fietto LG, Fontes EPB: **Regulated nuclear trafficking of *rPL10A* mediated by *NIK1* represents a defense strategy of plant cells against virus.** *PLoS Pathog* 2008, **4**:e1000247.
  71. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403–410.
  72. Bombarely A, Menda N, Tecle IY, Buels RM, Strickler S, Fischer-York T, Pujar A, Leto J, Gosselin J, Mueller LA: **The Sol genomics network (solgenomics.net): growing tomatoes using Perl.** *Nucleic Acids Res* 2011, **39**:D1149–D1155.
  73. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23**:2947–2948.
  74. Price MN, Dehal PS, Arkin AP: **FastTree 2-approximately maximum-likelihood trees for large alignments.** *PLoS One* 2010, **5**:e9490.

75. Hanks SK, Hunter T: Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J* 1995, **9**:576–596.
76. Letunic I, Bork P: Interactive tree of life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* 2011, **39**:W475–W478.
77. Newcombe RG: Interval estimation for the difference between independent proportions: comparison of eleven methods. *Stat Med* 1998, **17**:873–890.
78. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011, **28**(10):2731–2739.
79. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E: The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucl Acids Res* 2011, **40**(Database issue):D1202–D1210.
80. Calegario RF, de Ferreira SS, Andrade EC, Zerbini FM: Characterization of tomato yellow spot virus, a novel tomato-infecting begomovirus in Brazil. *Pesq Agropec Bras* 2007, **42**:1335–1343.
81. Marshall OJ: PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics* 2004, **20**:2471–2472.
82. Livak KJ, Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C(T)}$  Method. *Methods* 2001, **25**:402–408.

doi:10.1186/1471-2229-12-229

**Cite this article as:** Sakamoto *et al.*: The tomato RLK superfamily: phylogeny and functional predictions about the role of the LRRIL-RLK subfamily in antiviral defense. *BMC Plant Biology* 2012 **12**:229.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

