Genome scanning for identification and mapping of receptor-like kinase (RLK) gene superfamily in *Solanum tuberosum*

Sara Dezhsetan¹

Received: 31 May 2017 / Revised: 13 September 2017 / Accepted: 18 September 2017 / Published online: 3 October 2017
© Prof. H.S. Srivastava Foundation for Science and Society 2017

Abstract Receptor-like kinases (RLKs) are a key class of genes that contribute to diverse phenomena from plant development to defense responses. The availability of completed potato genome sequences provide an excellent opportunity to identify and characterize RLK gene superfamily in this lineage. We identified 747 non-redundant RLK genes in the potato genome that were classified into 52 subfamilies, of which 58% members organized into tandem repeats. Nine of potato RLK subfamilies organized into tandem repeats. Also, six subfamilies exhibited lineage-specific expansion compared to Arabidopsis. The majority of RLK genes were physically organized within heterogeneous and homogeneous clusters on chromosomes and were unevenly distributed on the genome. Chromosome 2, 3 and 7 contained the highest number of RLK genes and the most underrepresented chromosomes were chromosome 8, 10 and 11. Taken together, our results provide a framework for future efforts on comparative, evolutionary and functional studies of the members of RLK superfamily.

Keywords Bioinformatics · Comparative analyses · Genome analysis · RLK superfamily · Potato

Introduction

Plant cells continually respond to multiple signals that come from external and endogenous stimuli, which are converted to proper cellular and molecular responses that enable plant cells adaptation to diverse local environmental conditions (Morris and Walker 2003). Plant cells possess multiple systems of transmembrane receptor proteins that receive, perceive and process signals. Receptor-like kinase (RLK) superfamily is one of the largest classes of these receptors (Sakamoto et al. 2012). Plant RLKs have similar structure to animal's RTKs that possess a cytoplasmic tyrosine kinase domain (Hardie 1999). The RLK subfamilies are characterized by a ligand-binding extracellular domain that is very divergent in domain organization and sequences, followed by a single transmembrane spanning domain and a conserved cytoplasmic serine/threonine kinase domain. However, some RLK members, named as receptor-like cytoplasmic kinases (RLCKs) are localized in cytoplasm and do not possess extracellular region or a transmembrane segment (Shiu et al. 2004). It is revealed that the plant transmembrane RLKs perceive signals through their extracellular domains by binding to a specific signal molecule and relay signals via cytoplasmic kinase segments that trigger a signal transduction cascade(s) (Matsubayashi 2003; Shiu et al. 2004). Phylogenetic analysis of plant RLKs classify them into several distinct subfamilies based on structural comparison analyses of extracellular regions and domain organizations (Shiu et al. 2004). The diversity of RLKs implicates different potential functions of the RLK subfamilies.

RLKs have been identified in many plant species and their structure and function have been studied in the past investigations. Functional analysis of plant RLKs revealed that this superfamily could be categorized into two major

Electronic supplementary material The online version of this article (doi:10.1007/s12298-017-0471-6) contains supplementary material, which is available to authorized users.

✉ Sara Dezhsetan
sdezhsetan@uma.ac.ir

¹ University of Mohaghegh Ardabili, Ardabil, Iran

biological function classes (Shiu and Bleecker 2001a). The first major class contains RLKs that associate with plant growth and development (Becraft 2002). Examples of this class of RLKs include CLAVATA1 that controls meristem and flower development (Clark 1997), BRI1 that mediates brassinosteroid signal transduction (Li and Chory 1997) and ERECTA involved in regulating organ shape (Shpak et al. 2003). The second major class includes RLKs involved in plant-pathogen interactions and stress responses (Lehti-Shiu et al. 2009). Recent evidence suggest that some subfamilies like lectin RLK subfamily (Vaid et al. 2012; Yang et al. 2016), LRR-RLK subfamily (Lehti-Shiu et al. 2009; Wu et al. 2016) are induced by stress conditions and also play key roles in developmental stages.

RLKs perform wide variety of functions. Certain RLKs are involved in cell wall attachment, eg., extensin, proline-rich extension and lectin RLKs (Silva and Goring 2002); early stages of nodulation eg., LysM RLKs (Madsen et al. 2003); regulating rhizobial nod factor-induced infection (Limpens et al. 2003); plant-bacterial symbiotic interactions (Radutoiu et al. 2003) and self-incompatibility eg., S-locus glycoprotein-RLKs (Cui et al. 2000). Also, it is revealed that six CrRLK1L genes are highly related to fiber development in cotton (Niu et al. 2016). Two CrRLK1L genes in rice associate with water-use efficiency and one CrRLK1L gene in rice mediates a circadian regulation mechanism (Nguyen et al. 2015). Among all RLKs, LRR-RLKs include the highest frequencies in RLK superfamily. To date, genome-wide analysis has identified a large number of RLK members in several species, for example, 300 and 297 RLK proteins in *Citrus clementina* and *Citrus sinensis*, respectively (Magalhães et al. 2016) and 379 LRR-RLK genes in *Populus trichocarpa* (Zan et al. 2013). LRR-RLKs inclusive leucine-rich repeat (LRR) domains are composed of 15 subgroups (LRR I to LRR XV) in *Arabidopsis thaliana*, 14 subgroups in *P. trichocarpa* (Zan et al. 2013) that vary in number of LRR domains (from 1 to 25) and distribution pattern of LRRs in extracellular region (Shiu and Bleecker 2001b). However, supplementary phylogenetic analysis show that plant LRR-RLKs belong to 19 subfamilies, 18 of which originated from early land plants and one of them established in flowering plants (Liu et al. 2017).

The structural features of LRRs extracellular domain are not conserved among subfamilies, providing evidence for functional divergence among LRR-RLK subfamilies (Dufayard et al. 2017). Also, expression analyses of LRR-RLK genes provide further evidence for their functional diversification (Liu et al. 2016; Wu et al. 2016) and has confirmed that a large number of LRR-RLK genes play important molecular functions as receptors of external signals (Park et al. 2014).

Whole genome sequencing projects of model plants and field crops enable identification of various gene families.

Genome wide analyses indicate variation in the number of RLK members in different plants with 1070 in rice (Dardick et al. 2007), 605 in soybean (Liu et al. 2009), 647 in tomato (Sakamoto et al. 2012) and more than 600 members of RLKs in *Arabidopsis* (Shiu and Bleecker 2003). A comparative analysis of protein kinase of *Arabidopsis* and rice revealed common and unique protein kinases in these plants (Krupa and Anamika 2006). The RLKs involved in developmental function have approximately conserved size, while those associated with defense/disease resistance function increased their members by duplication events (Shiu et al. 2004).

Potato is the most important non-cereal world food crop, with more than 380 million tons field production in the world (<http://www.fao.org>). The genome sequence of *Solanum tuberosum* group Phureja, annotation genes and proteins and anchored sequence map were published in 2011 (Xu et al. 2011). Genomic analysis predicted that potato genome has near 740 mega bases of DNA and 39,031 genes and 86% of the genome anchored to linkage groups. The availability of complete potato genome sequence, annotation genes and proteins and anchored sequence map allow genomic studies on RLK superfamily. So far, any large-scale analysis hasn't been managed on RLK superfamily on potato genome.

Structural analysis of extracellular domains of RLKs *Arabidopsis* and tomato classified them into 50 (Shiu and Bleecker 2003) and 58 subfamilies (Sakamoto et al. 2012), respectively. The diversity of extracellular domains and association with potential functions is reasonable cause for investigation of RLK superfamily. In these analyses, we identified RLK superfamily in *S. tuberosum* group Phureja genome, classified these genes based on domains and anchored them to linkage groups of this genome. Collectively, genome-wide analysis of potato RLKs provide information that would facilitate in gaining comprehensive insight into potato RLKs.

Materials and methods

Data sets of potato genome sequence

Annotated genes and predicted proteins were downloaded from the potato genome sequencing consortium (PGSC) data release site at <http://potatogenomics.plantbiology.msu.edu/index.html> (Xu et al. 2011).

Identification of genes that encode RLKs domains

BLASTP analysis was used (E-value cutoff was 0.01) to retrieve the potato RLK candidate proteins using each subfamily of RLK superfamily of *A. thaliana* and *S.*

lycopersicum against a protein database of potato (PGSC v4.03 Pseudomolecules) available on http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml. Predicted proteins were screened using HMMER V.3 (Finn et al. 2011) by Hidden Markov Model (HMM) in the Pfam (<http://pfam.sanger.ac.uk/>) and E-value cutoff was 0.00001. The protein set of each subfamily with E-value < $1E - 20$ was used in the MEME Suite psp-gen script (version 4.8.1) (Bailey et al. 2010) to summarize information about anticipated discriminative motifs and identify the 10 most significant motifs. Then, based on these potato-specific motifs of each subfamily, a MAST search was performed on all of the predicted protein models (PGSC0003DMP.pep.v3.4) and ignored motifs with E-value upper than 0.00001. DMP sequences were considered as RLK candidates if their reported MAST E-value were lower than 0.00001. Re-annotating retrieved predicted RLKs and predicted domains were performed by SMART (<http://smart.embl-heidelberg.de/>) and Pfam databases. Also, classification and re-prediction of domains of forecasted RLK candidates was carried out using RLKs iTAK—Plant Transcription factor and Protein Kinase Identifier and Classifier (<http://bioinfo.bti.cornell.edu/cgi-bin/itak/index.cgi>). Proteins that did not consist of domains of interest in each subfamily were not considered for further analysis.

Multiple alignment and phylogenetic analysis

Multiple alignment and phylogenetic tree were built based on re-annotated RLK proteins. Full-length amino acid sequences of predicted members of RLK-subfamily were aligned using Clustal W with default parameters. Maximum likelihood method [JTT model, partial deletion, Nearest-Neighbor-Interchange (NNI)] with 1000 bootstrap replicates was used to construct a phylogenetic tree with MEGA6 (Tamura et al. 2013) software.

Anchoring annotated RLK genes to genome

All DM gene models (DMGs) corresponding to the identified RLK DMPs, were extracted from PGSC data sharing site (PGSC_DM_v3.4_gene.fasta). The location of genes on chromosomes, positive and negative DNA strand were extracted from the PGSC data sharing site. To anchoring RLK genes to *S. tuberosum* (group Phureja) genome, RLK genes were mapped to physical position map of chromosomes. Genomepixelizer (Kozik et al. 2002) was used to locate the RLK genes into twelve linkage groups of potato.

Identification of RLKs in tandem duplication

For identification of tandem duplication mechanism and determination of their frequency in each subfamily, we need to harbor annotated RLK genes to genome. Two genes were considered as tandem cluster if I) they belonged to the same subfamily or II) they separated from each other by less than 10 genes or 100 kb (Lehti-Shiu et al. 2009).

We used reconciliation methods in Notung software (2.8) (Stolzer et al. 2012) for inferring duplication/deletion gene events and determination of common ancestral genes after divergence of Arabidopsis and potato.

Results

Identification of RLK superfamily

In total, 1988 DMPs were identified by MEME. All retrieved DMPs were submitted to SMART and pfam databases (with E-value cutoff = 0.00001) for re-annotation of their domains. This resulted in choosing 1249 DMPs as RLK candidate proteins. This reduction was mainly related to chosen NBS-LRR family, due to common domains (LRRs) between NBS-LRR family and LRR-RLK subfamilies. Also, a portion of this reduction was as a result of retrieving proteins that were protein kinase but didn't belong to RLK family and identifying some proteins because of pseudo-similarity with characterized motifs by MEME and lacking a typical RLK structure, the latter had lower e-value and had no relationship to RLK superfamily. Also, RLK proteins with incomplete domains and short sequences were omitted manually. Finally, 1169 DMPs were distinguished as RLK candidate proteins. Also, 747 DMGs were identified, of which 31 didn't have specific chromosomal positions and were excluded manually from genomic distribution analysis (Table 1) (Supplementary material 2).

In spite of the fact that RLK superfamily profile of potato, tomato, Arabidopsis and rice (Table 1) have common characteristics, comparative analysis revealed some differences between these species based on presence of some subfamilies and member size of each subfamily. Interestingly, only one Lectin C-type was found in four species. The results showed that LRR-VII-3, Singleton and URK-2 subfamilies were present only in potato and LRR-I-3, LRR-IV-sis, LRR-VIII-2, LRR-Xc, LRR-XII-2, LRR-XV-sis, RLCK-VIII-sis, RLCK-XIII-sis, RLCK-XIV, SD-2a, SD-3, URK-I-sis and WAKL-OS subfamilies included any member in potato. Also, LRR-XV-sis, RLCK-OS, RLCK-XIII-sis, SD-2a, URK-I-sis and

Table 1 RLK-subfamilies organized based on structural comparisons of extracellular domains and phylogenetic analysis in four species. The information of tomato was inferred from Sakamoto et al. (2012) and Arabidopsis and rice were obtained from Shiu et al. (2004)

RLK subfamilies	<i>Solanum tuberosum</i>	<i>Arabidopsis thaliana</i>	<i>Solanum lycopersicum</i>	<i>Oryza sativa</i>
C-LEC	1	1	1	1
CR4L	9	8	7	7
CrRLK1L-1	29	15	23	18
DLSV	78	76	80	98
Extensin	5	5	6	8
L-LEC	30	45	23	103
LRK10L-1	10	0	0	0
LRK10L-2	47	13	14	38
LRR-I-1	5	43	2	36
LRR-I-2	2	2	2	2
LRR-I-3	0	3	4	4
LRR-II	13	14	13	13
LRR-III	44	47	44	41
LRR-IV, LRR-IV-sis	3, 0	3, 1	2, 1	3, 0
LRR-IX	8	4	8	3
LRR-V	8	8	9	12
LRR-VI	11	14	12	15
LRR-VII-1	5	7	8	6
LRR-VII-2	2	3	2	3
LRR-VII-3	3	0	0	0
LRR-VIII-1	6	8	8	10
LRR-VIII-2	0	14	19	32
LRR-Xa	9	4	6	3
LRR-Xb	11	10	11	25
LRR-Xc	0	2	4	1
LRR-XI	61	26	31	27
LRR-XII-1	88	8	48	117
LRR-XII-2	0	2	6	10
LRR-XIIIa	2	4	3	3
LRR-XIIIb	2	3	2	2
LRR-XIV	3	3	3	2
LRR-XV, LRR-XV-sis	2, 0	3, 0	8, 0	13, 1
LysM	15	4	13	12
PERK	17	19	19	18
RKF3	3	2	2	2
RLCK-II	2	4	2	1
RLCK-IV	3	3	2	5
RLCK-IXa	2	2	2	3
RLCK-IXb	13	19	15	28
RLCK-OS	0	0	0	40
RLCK-V	9	11	8	13
RLCK-VI	13	14	14	10
RLCK-VIIa	50	45	55	51
RLCK-VIIb	1	1	0	1
RLCK-VIII, RLCK-VIII-sis	7, 0	11, 1	7, 1	9, 1
RLCK-X	2	4	2	3
RLCK-XI	3	4	3	2

Table 1 continued

RLK subfamilies	<i>Solanum tuberosum</i>	<i>Arabidopsis thaliana</i>	<i>Solanum lycopersicum</i>	<i>Oryza sativa</i>
RLCK-XII	25	25	9	9
RLCK-XIII, RLCK-XIII-sis	3, 0	2, 0	2, 0	5, 1
RLCK-XIV	0	1	1	1
RLCK-XV	2	2	3	4
RLCK-XVI	1	1	1	2
SD-2a	0	0	0	30
SD-2b	38	7	27	81
SD-3	0	1	1	0
Singleton	1	0	0	0
URK-1, URK-I-sis	2, 0	2, 0	2, 0	2, 1
URK-2	4	0	0	0
URK-3	0	1	0	0
WAK	24	22	15	96
WAK_LRK10L-1	10	7	15	8
WAKL-OS	0	0	0	48
error_region1, error_region2	0, 0	0, 0	0, 0	64, 13
not_assigned	0	8	16	14
Total	747	606	647	1235

WAKL-OS subfamilies were rice specific RLK subfamilies. The rice RLK member size was almost twofold larger than that of Arabidopsis, potato and tomato. Most rice RLK subfamilies were nearly larger than those of Arabidopsis, potato and tomato; however it was not necessarily correct for each subfamily (Table 1).

Also, CrRLK1L-1, L-LEC, LRK10L-2, LRR-XI, LRR-XII-1, RLCK-XII, SD-2b and WAK subfamilies in potato displayed prominent expansion than those in tomato. Inversely, RLCK-VIIa and WAK_LRK10L-1 subfamilies in potato reduced in size in comparison to tomato.

In total, the existence and size of each subfamily didn't follow a specific pattern in these four species and exhibited approximately diverse way of evolution in this superfamily after divergence of these species from each other.

In order to infer the degree of expansion in RLK subfamilies, the occurrence of duplication/deletion events and determination of number of common ancestral genes in Arabidopsis and potato after divergence, we compared the RLK subfamilies trees with a species tree constructed at NCBI taxonomy browser using reconciliation method. Then, the frequency of duplication/deletion events was calculated in each RLK subfamily. As shown in Fig. 1, the expansions/reductions were identified in certain subfamilies. DLSV, L-LEC, LRK10L-2, RLCK-VIIa, RLCK-XII and WAK subfamilies have remarkably expanded in their member size in these two plant lineages. Whereas, in potato, lineage-specific expansions were identified in the CrRLK1L-1, LRR-Xa, LRR-XI, LRR-XII, LysM and SD-2b subfamilies. Also, LRR-I-1, LRR-III, PERK and

RLCK-IXb subfamilies showed considerable expansion in Arabidopsis. Also, RLCK-VIII subfamily disappeared from Arabidopsis. Others were approximately similar in member size between common ancestor and Arabidopsis and potato, showing that they have not much expanded/reduced after their speciation. In total, these findings suggested that the ancestral RLK gene family has dramatically expanded after Arabidopsis-potato split, from 414 member size in common ancestor to 606 in Arabidopsis and 747 in potato. Therefore, further expansions occurred after divergence.

Also, we compared the number of duplications/deletions that happened after Arabidopsis-potato split (Fig. 2). Most points are grouped around – 5 (deletion) to 5 (duplication) that exhibited limited rounds of duplication/deletion in each subfamily after divergence of Arabidopsis and potato. However, approximately few of subfamilies have noticeably expanded after these two lineage split. Also, different subfamilies expanded diversely, suggesting that these subfamilies are subjected to different selection pressure. On the other hand, most of lineage-specific expansions of RLK subfamilies display diverse evolutionary (duplication/deletion) rate in Arabidopsis and potato, suggesting these relate to differences in functions of RLKs and selection pressure on RLKs in these two lineage.

Tandem duplication mechanism

Tandem and large-scale duplication mechanisms represent a common role in expansion of RLK family (Shiu and Bleecker 2001b; Shiu et al. 2004; Sakamoto et al. 2012).

Fig. 1 The expansion profile of RLK subfamilies in Arabidopsis and potato before and after divergence

Fig. 2 The potato du/de and Arabidopsis du/de indicate the number of duplications/deletions that occurred in potato and Arabidopsis in each subfamily after divergence

We identified the RLKs clustered in tandem duplications and counted their frequency in each subfamily (Table 2) (Supplementary material 1). Among the 716 RLK members that mapped on genome, 418 members organized in tandem duplications (58%). Therefore, tandem duplication mechanism was one of the prominent expansion mechanisms in potato RLK family. There were 113 tandem duplication clusters with 2–18 genes in 12 linkage groups of potato. Subfamilies showed remarkable variation in the extent of tandem duplications. In potato, DLSV, L-LEC, LRR-XI-1, LRR-XII-1, RLCK-XII-2, SD-2b and WAK subfamilies indicated noticeable number of tandem duplications (more than 67% of their members located in tandem repeats). These subfamilies display critical roles in signal transduction and stress responses. Also, based on previous studies, LRR-XII subfamilies in tomato (Sakamoto et al. 2012) and DLSV, LRR-XII-2 and RLCK-XII-2 subfamilies in Arabidopsis (Shiu et al. 2004) contain a significant number of duplication clusters. In addition, it is documented that RLK members that are involved in resistance or defense signal transduction and stress responses significantly expanded by tandem duplication mechanism (Shiu et al. 2004; Lehti-Shiu et al. 2009; Sakamoto et al. 2012). In *P. trichocarpa* genome, approximately 82% (312 out of 379) of the LRR-RLK genes located in segmental duplication blocks, indicating the role of duplication process in the expansion of this gene family (Zan et al. 2013). Also, our finding indicated subfamilies that underwent noticeable lineage-specific expansion in potato tend to be located in tandem repeats.

Phylogenetic analysis

To assess relationship among retrieved RLKs genes, a phylogenetic tree was constructed from alignment of

Table 2 Chromosomal location, number of genes in tandem repeats, function of RLKs that located in tandem repeats

	RLK subfamilies	Chromosomal location (number of genes in tandem repeats)	Ratio*	Function	References
1	CR4L	11 (2)	0.22	Epidermal cell development L1 cell layer organization	Becraft et al. (1996) and Gifford et al. (2005)
2	CrRLK1L-1	2 (2), 3 (2), 5 (8, 2), 6 (2), 7 (2)	0.62	Regulation of fertilization and cell expansion mechanisms Up-regulated by most abiotic stresses	Galindo-Trigo et al. (2016) and Gao and Xue (2012)
3	DLSV	2 (2), 2 (10), 3 (5, 3, 2), 4 (2, 4), 5 (2), 7 (3, 3, 7, 12), 9 (2), 10 (3, 2), 12 (7)	0.88	Involved in biotic stresses	Lehti-Shiu et al. (2009)
4	L-LEC	2 (2), 3 (2, 2), 9 (3, 2, 3), 10 (3, 3)	0.67	Involved in biotic stresses Plant innate immunity	Lehti-Shiu et al. (2009) and Singh and Zimmerli (2013)
5	LRK10L-1	5 (9)	0.9	ABA-mediated signaling Drought resistance	Lim et al. (2015)
6	LRK10L-2	1 (8), 2 (12), 3 (2), 5 (12), 12 (2)	0.76	Resistance to fungal pathogen	Feuillet et al. (1997)
7	LRR-III	1 (2), 3 (2, 5, 2), 5 (3), 6 (2, 2), 8 (3), 9 (2, 2), 11 (2)	0.61	Meristem proliferation Involved in abiotic stress	Ogawa et al. (2008) and ten Hove et al. (2011)
8	LRR-VII-1	6 (2)	0.4	Involved in abiotic stress	ten Hove et al. (2011)
9	LRR-VIII-1	1 (2), 10 (3)	0.83	Involved in biotic stresses	Lehti-Shiu et al. (2009)
10	LRR-Xa	7 (3), 12 (2)	0.55	Involved in abiotic stress	ten Hove et al. (2011)
11	LRR-Xb-1	4 (2)	0.22	Involved in biotic stresses	Lehti-Shiu et al. (2009)
12	LRR-XI-1	1 (2), 2 (2), 3 (2, 2), 4 (2, 14), 8 (6), 9 (3, 2), 11 (2), 12 (2)	0.67	Expressed in drought stress	Gao and Xue (2012)
13	LRR-XII-1	2 (2, 12), 3 (4, 3, 2), 4 (18), 6 (6), 7 (2, 5), 8 (8), 10 (2), 11 (4)	0.77	Resistance to bacterial pathogen	Song et al. (1995)
14	LysM	2 (2), 9 (3)	0.33	Plant-bacterial symbiotic interactions Fungal and bacterial resistance	Radutoiu et al. (2003) and Gimenez-Ibanez et al. (2009)
15	RLCK-Ixb	3 (2), 5 (2)	0.30	Plant development Stress responses	Jurca et al. (2008)
16	RLCK-V	1 (2), 12 (2)	0.44		
17	RLCK-VI	1 (2), 1 (2)	0.30		
18	RLCK-VIIa-1	4 (2), 6 (2), 10 (2)	0.37	Resistance to bacterial pathogen	Swiderski and Innes (2001)
19	RLCK-VIIa-2	1 (4, 2), 2 (2), 5 (2), 6 (2), 7 (2), 9 (2), 12 (2)	0.52	Resistance to bacterial pathogen	Swiderski and Innes (2001)
20	RLCK-VIII	10 (2), 12 (2)	0.57	Regulate Brassinosteroids signaling	Kim et al. (2011)
21	RLCK-XII-2	6 (15)	0.83		
22	RLCK-XIII	12 (2)	0.66	Expressed in drought stress	Gao and Xue (2012)
23	SD-2b	1 (2), 3 (3), 7 (12), 9 (3, 2), 11 (7)	0.76	Involved in biotic stresses Expressed in salt stress	Lehti-Shiu et al. (2009) and Gao and Xue (2012)
24	URK-2	5 (2)	0.5	Defense response	He et al. (1998)
25	WAK	7 (2), 9 (4, 8), 10 (5)	0.79		
26	WAK_LRK10L-1	2 (2), 12 (2)	0.4	Involved in biotic stresses	Lehti-Shiu et al. (2009)

* Total number of tandem repeat genes in each subfamily divided to total number of genes in the same subfamily

DMPs. The phylogenetic analysis separated dramatically the RLK subfamilies. There were approximately clear separations between RLK subfamilies that reflected

differentiation of RLK encoding genes. However, a few exceptions were present in phylogenetic tree, as some potato RLKs were not clustered in its subfamily because,

Fig. 3 Maximum likelihood phylogenetic analysis of potato RLK superfamily. Phylogenetic tree was drawn by sequence alignment of RLK proteins (with 1000 bootstraps). RLK subfamilies were shown with different colors (color figure online)

some extracellular domain sequences are presented in several different of RLK subfamilies (Fig. 3).

Genomic organization of RLK genes

Genomic distribution for predicted RLK-encoding genes were established on 12 potato chromosomes (Fig. 4)

(Supplementary material 1). In total, 716 RLK-encoding genes were anchored on 12 potato chromosomes. There were mapped 58, 89, 82, 77, 66, 53, 82, 35, 63, 39, 35 and 37 RLK genes on chromosomes 1–12, respectively. It is obvious that genomic organization among the majority of linkage groups was not even. Chromosome 2, 3 and 7 contained the highest number of RLK genes and the most

Fig. 4 Genomic distribution of RLK genes on *Solanum tuberosum* group phureja genome. Gray bars indicated all 12 chromosomes of potato. Boxes across each bar showed the location of RLK genes. Genes encoded by positive DNA strand were shown below bar whereas those encoded by negative strand are designated above bar. Colors of boxes represented of RLK subfamilies (color figure online)

underrepresented chromosomes were chromosome 8, 10, 11 and 12.

RLK genes formed heterogeneous and homogenous clusters and were unevenly distributed on the genome. Interestingly, most subfamilies were greatly clustered on the genome. In addition, there were several cases where mixed subfamilies or one subfamily formed supercluster. LRR, LRK10L, RLCK, DLSV, CrRLK1L and SD-2b subfamilies formed superclusters. Heterogeneous and homogenous subfamilies in large clusters, which formed superclusters, majority located at the end or at the first part of chromosomes. The greatest number of LRK10L subfamily was found on chromosomes 2 and 5, and biggest RLCK cluster harbored on chromosome 6. Also, LRRs and RLCKs organized on all chromosomes in homogenous and heterogeneous clusters or located physically singleton. Interestingly, although some genes did not occur physically in clusters, most of them were located near clusters.

Discussion

The RLK superfamily is one of the largest gene families that expanded significantly in land plants (Shiu and Bleecker 2003) and structurally generated before algae-land plants split almost 10^9 years ago (Lehti-Shiu et al. 2009) because proteins with RLK configuration were identified in unicellular algae (Lehti-Shiu et al. 2009) and multicellular algae (Sasaki et al. 2007). In this attempt, we identified and characterized 747 potato RLK members. RLK superfamily has approximately similar size in tomato (647 RLKs) and Arabidopsis (> 600 RLKs), but the size of potato RLK superfamily was larger than tomato and Arabidopsis and lower than rice. The potato RLK superfamily was classified into 52 subfamilies (Table 1). The largest RLK subfamily belonged to LRR subfamily with 288 members. Also, the most abundant RLK subfamily is LRR subfamily in Arabidopsis and tomato with 236 and 257 members, respectively (Shiu et al. 2004; Sakamoto et al. 2012). The second biggest subfamily was RLCK with 136 out of 747 members of RLK gene family. RLCK include 128 members in tomato (Sakamoto et al. 2012) and 150 members in Arabidopsis (Shiu et al. 2004). Also, we estimated 414 ancestral RLK genes prior to divergence between Arabidopsis and potato. Also, Shiu et al. (2004) evaluated ~ 443 ancestral RLK genes before Arabidopsis-rice split. Therefore, RLK superfamily expanded in common ancestor of angiosperms that branched 15^7 – 20^7 years ago (Shiu et al. 2004). Evidence documented further expansions in potato RLK superfamily in comparison to Arabidopsis and their common ancestor. Also, the degree of expansion was widely diverse in subfamilies. In addition, specific and extensive expansions and reductions

imply lineage-specific responses to environmental conditions.

Previous studies showed some subfamilies, such as DLSV, L-LEC, LRR-I, LRR-VIII-2, LRR-Xb, RLCK-VIIa, SD-2b, WAK, and WAK_LRK10L-1, involved in biotic and abiotic stresses (Table 2). In addition, all of these subfamilies tend to be located in tandem repeats and specific expansions have mostly occurred for them. Taken together, our findings are consistent with Sakamoto et al. (2012).

It is evident that the distribution of these genes on the majority of chromosomes are not even. This unequal distribution of RLK genes is not novel in the plant genomes (Shiu and Bleecker 2001b). Also, we found RLK subfamilies have differential patterns of chromosomal distributions. Some genes are not grouped (singletons) but most members of this superfamily are located within clusters. As previously mentioned in R genes (Friedman and Baker 2007), the clustered nature of RLK genes is thought to facilitate the evolution and expansion of these genes through sequence exchange via unequal crossing over or recombinational mispairing events (Zhang 2003).

More than half of RLK genes organized in tandem clusters. Also, the percentage of tandem repeats and the degree of expansion for each subfamily have positive relationship. In addition, RLK genes located in tandem repeats are significantly more responsive to environmental stresses, especially biotic stresses, than those not located in tandem repeats (Hanada et al. 2008; Lehti-Shiu et al. 2009). On the other hand, it is documented that tandem repeats are associated with the evolution of other agronomic traits (Wang et al. 2015).

Comprehensive analysis of RLK genes to identify, characterize phylogenetic relationships and genomic organization may provide a framework to facilitate comparative, evolutionary and functional studies of the members of potato RLK superfamily.

Acknowledgement Funding was provided by University of Mohaghegh Ardabili (Grant No. 554).

References

- Bailey TL, Bodén M, Whittington T, Machanick P (2010) The value of position specific priors in motif discovery using MEME. *BMC Bioinform* 11:179. doi:[10.1186/1471-2105-11-179](https://doi.org/10.1186/1471-2105-11-179)
- Becraft PW (2002) Receptor kinase signaling in plant development. *Annu Rev Cell Dev Biol* 18:163–192. doi:[10.1146/annurev.cellbio.18.012502.083431](https://doi.org/10.1146/annurev.cellbio.18.012502.083431)
- Becraft PW, Stinard PS, McCarty DR (1996) CRINKLY4: a TNFR-like receptor kinase involved in maize epidermal differentiation. *Science* 273(80):1406
- Clark SE (1997) Organ formation at the vegetative shoot meristem. *Plant Cell* 9:1067–1076. doi:[10.1105/tpc.9.7.1067](https://doi.org/10.1105/tpc.9.7.1067)

- Cui Y, Bi YM, Brugiere N et al (2000) The S locus glycoprotein and the S receptor kinase are sufficient for self-pollen rejection in Brassica. *Proc Natl Acad Sci USA* 97:3713–3717. doi:[10.1073/pnas.050480297](https://doi.org/10.1073/pnas.050480297)
- Dardick C, Chen J, Richter T et al (2007) The rice kinase database. A phylogenomic database for the rice kinome. *Plant Physiol* 143:579–586. doi:[10.1104/pp.106.087270](https://doi.org/10.1104/pp.106.087270)
- Dufayard J-F, Bettembourg M, Fischer I et al (2017) New insights on leucine-rich repeats receptor-like kinase orthologous relationships in angiosperms. *Front Plant Sci* 8:18. doi:[10.3389/fpls.2017.00381](https://doi.org/10.3389/fpls.2017.00381)
- Feuillet C, Schachermayr G, Keller B (1997) Molecular cloning of a new receptor-like kinase gene encoded at the Lr10 disease resistance locus of wheat. *Plant J* 11:45–52
- Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39:W29–W37. doi:[10.1093/nar/gkr367](https://doi.org/10.1093/nar/gkr367)
- Friedman AR, Baker BJ (2007) The evolution of resistance genes in multiprotein plant resistance systems. *Curr Opin Genet Dev* 17:493–499. doi:[10.1016/j.gde.2007.08.014](https://doi.org/10.1016/j.gde.2007.08.014)
- Galindo-Trigo S, Gray JE, Smith LM (2016) Conserved roles of CrRLK1L receptor-like kinases in cell expansion and reproduction from algae to angiosperms. *Front Plant Sci* 7:1–10. doi:[10.3389/fpls.2016.01269](https://doi.org/10.3389/fpls.2016.01269)
- Gao L-L, Xue H (2012) Global analysis of expression profiles of rice receptor-like kinase genes. *Mol Plant* 5:143–153. doi:[10.1093/mp/ssr062](https://doi.org/10.1093/mp/ssr062)
- Gifford ML, Robertson FC, Soares DC, Ingram GC (2005) Arabidopsis CRINKLY4 function, internalization, and turnover are dependent on the extracellular crinkly repeat domain. *Plant Cell* 17:1154–1166. doi:[10.1105/tpc.104.029975](https://doi.org/10.1105/tpc.104.029975)
- Gimenez-Ibanez S, Ntoukakis V, Rathjen JP (2009) The LysM receptor kinase CERK1 mediates bacterial perception in Arabidopsis. *Plant Signal Behav* 4:539–541. doi:[10.1016/j.cub.2009.01.054](https://doi.org/10.1016/j.cub.2009.01.054)
- Hanada K, Zou C, Lehti-Shiu MD et al (2008) Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol* 148:993–1003. doi:[10.1104/pp.108.122457](https://doi.org/10.1104/pp.108.122457)
- Hardie DG (1999) Plant protein serine/threonine kinases: classification and functions. *Annu Rev Plant Physiol Plant Mol Biol* 50:97–131
- He Z-H, He D, Kohorn BD (1998) Requirement for the induced expression of a cell wall associated receptor kinase for survival during the pathogen response. *Plant J* 14:55–63
- Jurca ME, Bottka S, Fehér A (2008) Characterization of a family of Arabidopsis receptor-like cytoplasmic kinases (RLCK class VI). *Plant Cell Rep* 27:739–748. doi:[10.1007/s00299-007-0494-5](https://doi.org/10.1007/s00299-007-0494-5)
- Kim TW, Guan S, Burlingame AL, Wang ZY (2011) The CDG1 kinase mediates brassinosteroid signal transduction from BRI1 receptor kinase to BSU1 phosphatase and GSK3-like kinase BIN2. *Mol Cell* 43:561–571. doi:[10.1016/j.molcel.2011.05.037](https://doi.org/10.1016/j.molcel.2011.05.037)
- Kozik A, Kochetkova E, Michelmore R (2002) GenomePixelizer—a visualization program for comparative genomics within and between species. *Bioinformatics* 18:335–336
- Krupa A, Anamika N (2006) Genome-wide comparative analysis of domain organization of repertoires of protein kinase of *Arabidopsis thaliana* and *Oryza sativa*. *Gene* 380:1–13. doi:[10.1016/j.gene.2006.05.016](https://doi.org/10.1016/j.gene.2006.05.016)
- Lehti-Shiu MD, Zou C, Hanada K, Shiu S-H (2009) Evolutionary history and stress regulation of plant. *Plant Physiol* 150:12–26. doi:[10.1104/pp.108.134353](https://doi.org/10.1104/pp.108.134353)
- Li J, Chory J (1997) A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* 90:929–938
- Lim CW, Yang SH, Shin KH et al (2015) The AtLRK10L1.2, Arabidopsis ortholog of wheat LRK10, is involved in ABA-

- mediated signaling and drought resistance. *Plant Cell Rep* 34:447–455. doi:[10.1007/s00299-014-1724-2](https://doi.org/10.1007/s00299-014-1724-2)
- Limpens E, Franken C, Smit P et al (2003) LysM domain receptor kinases regulating rhizobial nod factor-induced infection. *Science* 302(80):630–633. doi:[10.1126/science.1090074](https://doi.org/10.1126/science.1090074)
- Liu P, Wei W, Ouyang S et al (2009) Analysis of expressed receptor-like kinases (RLks) in soybean. *J Gent Genomics* 36:611–619. doi:[10.1016/S1673-8527\(08\)60153-8](https://doi.org/10.1016/S1673-8527(08)60153-8)
- Liu P-L, Xie L-L, Li P-W et al (2016) Duplication and divergence of leucine-rich repeat receptor-like protein kinase (LRR-RLK) genes in basal angiosperm *amborella trichopoda*. *Front Plant Sci* 7:1–15. doi:[10.3389/fpls.2016.01952](https://doi.org/10.3389/fpls.2016.01952)
- Liu P-L, Du L, Huang Y et al (2017) Origin and diversification of leucine-rich repeat receptor-like protein kinase (LRR-RLK) genes in plants. *BMC Evol Biol* 17:16. doi:[10.1186/s12862-017-0891-5](https://doi.org/10.1186/s12862-017-0891-5)
- Madsen E, Madsen L, Radutoiu S et al (2003) A receptor kinase gene of the LysM type involved in legume perception of rhizobial signals. *Nature* 425:637–640. doi:[10.1038/nature02045](https://doi.org/10.1038/nature02045)
- Magalhães DM, Scholte LLS, Silva NV et al (2016) LRR-RLK family from two Citrus species: genome-wide identification and evolutionary aspects. *BMC Genom* 17:13. doi:[10.1186/s12864-016-2930-9](https://doi.org/10.1186/s12864-016-2930-9)
- Matsubayashi Y (2003) Ligand-receptor pairs in plant peptide signaling. *J Cell Sci* 116:3863–3870. doi:[10.1242/jcs.00733](https://doi.org/10.1242/jcs.00733)
- Morris ER, Walker JC (2003) Receptor-like protein kinases: the keys to response. *Curr Opin Plant Biol* 6:339–342
- Nguyen Q-N, Lee Y-S, Cho L-H et al (2015) Genome-wide identification and analysis of *Catharanthus roseus* RLK1-like kinases in rice. *Planta* 241:603–613. doi:[10.1007/s00425-014-2203-2](https://doi.org/10.1007/s00425-014-2203-2)
- Niu E, Cai C, Zheng Y et al (2016) Genome-wide analysis of CrRLK1L gene family in *Gossypium* and identification of candidate CrRLK1L genes related to fiber development. *Mol Genet Genomics* 291:1137–1154. doi:[10.1007/s00438-016-1169-0](https://doi.org/10.1007/s00438-016-1169-0)
- Ogawa M, Shinohara H, Sakagami Y, Matsubayashi Y (2008) Arabidopsis CLV3 peptide directly binds CLV1 ectodomain. *Science* 319(80):294. doi:[10.1126/science.1150083](https://doi.org/10.1126/science.1150083)
- Park S, Moon J-C, Park YC et al (2014) Molecular dissection of the response of a rice leucine-rich repeat receptor-like kinase (LRR-RLK) gene to abiotic stresses. *J Plant Physiol* 171:1645–1653. doi:[10.1016/j.jplph.2014.08.002](https://doi.org/10.1016/j.jplph.2014.08.002)
- Radutoiu S, Madsen Y, Madsen LH et al (2003) Plant recognition of symbiotic requires two LysM receptor-like kinases. *Nature* 425:585–592
- Sakamoto T, Deguchi M, Brustolini OJB et al (2012) The tomato RLK superfamily: phylogeny and functional predictions about the role of the LRR-II-RLK subfamily in antiviral defense. *BMC Plant Biol* 12:229. doi:[10.1186/1471-2229-12-229](https://doi.org/10.1186/1471-2229-12-229)
- Sasaki G, Katoh K, Hirose N et al (2007) Multiple receptor-like kinase cDNAs from liverwort *Marchantia polymorpha* and two charophycean green algae, *Closterium ehrenbergii* and *Nitella axillaris*: extensive gene duplications and gene shufflings in the early evolution of streptophytes. *Gene* 401:135–144. doi:[10.1016/j.gene.2007.07.009](https://doi.org/10.1016/j.gene.2007.07.009)
- Shiu S-H, Bleecker AB (2001a) Plant receptor-like kinase gene family: diversity, function, and signaling. *Sci STKE* 2001:13
- Shiu S-H, Bleecker AB (2001b) Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases. *Proc Natl Acad Sci USA* 98:10763–10768. doi:[10.1073/pnas.181141598](https://doi.org/10.1073/pnas.181141598)
- Shiu S-H, Bleecker AB (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](https://doi.org/10.1104/pp.103.021964)
- Shiu S-H, Karlowski WM, Pan R et al (2004) Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. *Plant Cell* 16:1220–1234. doi:[10.1105/tpc.020834](https://doi.org/10.1105/tpc.020834)
- Shpak ED, Lakeman KU, Torii MB (2003) Dominant-negative receptor uncovers redundancy in the Arabidopsis ERECTA leucine-rich repeat receptor-like kinase signaling pathway that regulates organ shape. *Plant Cell* 15:1095–1110
- Silva NF, Goring DR (2002) The proline-rich, extensin like receptor kinase-1 (PERK1) gene is rapidly induced by wounding. *Plant Mol Biol* 50:667–685
- Singh P, Zimmerli L (2013) Lectin receptor kinases in plant innate immunity. *Front Plant Sci* 4:4. doi:[10.3389/fpls.2013.00124](https://doi.org/10.3389/fpls.2013.00124)
- Song WY, Wang GL, Chen LL et al (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* 270(80):1804–1806
- Stolzer M, Lai H, Xu M et al (2012) Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* 28:i409–i415. doi:[10.1093/bioinformatics/bts386](https://doi.org/10.1093/bioinformatics/bts386)
- Swiderski MR, Innes RW (2001) The Arabidopsis PBS1 resistance gene encodes a member of a novel protein kinase subfamily. *Plant J* 26:101–112
- Tamura K, Stecher G, Peterson D et al (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729. doi:[10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197)
- ten Hove CA, Bochdanovits Z, Jansweijer VMA et al (2011) Probing the roles of LRR RLK genes in *Arabidopsis thaliana* roots using a custom T-DNA insertion set. *Plant Mol Biol* 76:69–83. doi:[10.1007/s11103-011-9769-x](https://doi.org/10.1007/s11103-011-9769-x)
- Vaid N, Pandey PK, Tuteja N (2012) Genome-wide analysis of lectin receptor-like kinase family from Arabidopsis and rice. *Plant Mol Biol* 80:365–388. doi:[10.1007/s11103-012-9952-8](https://doi.org/10.1007/s11103-012-9952-8)
- Wang Y, Xiong G, Hu J et al (2015) Copy number variation at the GL7 locus contributes to grain size diversity in rice. *Nat Genet* 47:944–948. doi:[10.1038/ng.3346](https://doi.org/10.1038/ng.3346)
- Wu Y, Xun Q, Guo Y 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](https://doi.org/10.1016/j.molp.2015.12.011)
- Xu X, Pan S, Cheng S et al (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475:189–195. doi:[10.1038/nature10158](https://doi.org/10.1038/nature10158)
- Yang Y, Labbé J, Muchero W et al (2016) Genome-wide analysis of lectin receptor-like kinases in *Populus*. *BMC Genom* 17:16. doi:[10.1186/s12864-016-3026-2](https://doi.org/10.1186/s12864-016-3026-2)
- Zan Y, Ji Y, Zhang Y et al (2013) Genome-wide identification, characterization and expression analysis of populus leucine-rich repeat receptor-like protein kinase genes. *BMC Genom* 14:318. doi:[10.1186/1471-2164-14-318](https://doi.org/10.1186/1471-2164-14-318)
- Zhang H (2003) Evolution by gene duplication: an update. *Trends Ecol Evol* 18:292–298. doi:[10.1016/S0169-5347\(03\)00033-8](https://doi.org/10.1016/S0169-5347(03)00033-8)