Ecotoxicology and Environmental Safety 167 (2019) 95–106

Contents lists available at ScienceDirect

Ecotoxicology and Environmental Safety

journal homepage: www.elsevier.com/locate/ecoenv

Identification and comprehensive analysis of the characteristics and roles of leucine-rich repeat receptor-like protein kinase (LRR-RLK) genes in *Sedum alfredii* Hance responding to cadmium stress

Xuelian He[a,b], Tongyu Feng[a,b], Dayi Zhang[c], Renying Zhuo[a,b], Mingying Liu[a,b,*]

[a] State Key Laboratory of Forest Genetics and Breeding, Xiangshan Road, Beijing 100091, P.R. China
[b] Key Laboratory of Tree Breeding of Zhejiang Province, The Research Institute of Subtropical of Forestry, Chinese Academy of Forestry, Hangzhou 311400, P.R. China
[c] School of Environment, Tsinghua University, Beijing 100084, P.R. China

ARTICLE INFO

Keywords:
Sedum alfredii Hance
Leucine-rich repeat receptor-like protein kinase (LRR-RLK)
Cadmium stress
Signal transduction

ABSTRACT

Sedum alfredii Hance is a Zn/Cd co-hyperaccumulator and its underlying molecular mechanism of Cd tolerance is worthy to be elucidated. Although numerous studies have reported the uptake, sequestration and detoxification of Cd in *S. alfredii* Hance, how it senses Cd-stress stimuli and transfers signals within tissues remains unclear. Leucine-rich repeat receptor-like protein kinases (LRR-RLKs) are vital for plant growth, development, immunity and signal transduction. Till now, there is lack of comprehensive studies addressing their functions in *S. alfredii* Hance responding to Cd stress. In the present study, we identified 60 LRR-RLK genes in *S. alfredii* Hance based on transcriptome analysis under Cd stress. They were categorized into 11 subfamilies and most of them had highly conserved protein structures and motif compositions. The inter-family diversity provided evidence for their functional divergence, supported by their expression level and profile in tissues under Cd stress. Co-expression network analysis revealed that the most highly connected hubs, *Sa0F.522*, *Sa0F.1036*, *Sa28F.115* and *Sa1F.472*, were closely related with other genes involved in metal transport, stimulus response and transcription regulations. Of the ten hub genes exhibiting differential expression dynamics under the short-term Cd stress (*Sa0F.522, Sa0F.1036* and *Sa28F.115*) were dramatically induced in the whole plant. Among them, *Sa0F.522* gene was heterologously expressed in a Cd-sensitive yeast cell line and its function in Cd signal perception was confirmed. For the first time, our findings performed a comprehensive analysis of LRR-RLKs in *S. alfredii* Hance, mapped their expression patterns under Cd stress, and identified the key roles of *Sa0F.522, Sa0F.1036* and *Sa28F.115* in Cd signal transduction.

1. Introduction

Cadmium (Cd) contamination is currently a worldwide problem affecting plant growth and threatening human health through plant-derived foods (Wang et al., 2017c; Yousaf et al., 2016). Increasing epidemiological evidences suggest the absence of safety threshold causing adverse health effects (Bernard, 2016; Callan et al., 2015; Satarug et al., 2017). Accordingly, eliminating Cd from polluted soils to reduce the risks of Cd exposure is of great urgency. Of all the strategies, phytoremediation is viewed as an effective and environmental-friendly approach to clean Cd-contaminated soils and hyperaccumulating plants are becoming the focus of many studies (Agnello et al., 2016; Tauqeer et al., 2016).

Sedum alfredii Hance (*Crassulaceae*) is a zinc/Cd (Zn/Cd) hyperaccumulator growing naturally in mining regions in China where Cd concentration was up to 400 mg/kg in soils (Tian et al., 2017). As it can hyperaccumulate up to 9000 μg/g Cd and 29,000 μg/g Zn in shoots without any toxicity symptoms (Yang et al., 2004), *S. alfredii* Hance has great potential in phytoremediation to decontaminate Cd-polluted soils (Wan et al., 2016; Zhu et al., 2012). Numerous studies have elucidated

Abbreviations: PP2C68, protein phosphatase 2C68; MPK5, mitogen protein kinase 5; LRR-RLK, leucine-rich repeat receptor-like kinase; NPR1, non-expressor pathogenesis-related 1; SERK, Somatic Embryogenesis Receptor Kinase; LRR_8, leucine-rich repeat 8; LRR_1, leucine-rich repeat 1; LRRNT_2, leucine-rich repeat N-terminal 2; MEME, Multiple Expectation Maximization for Motif Elicitation; PK, protein kinase; RLKs, receptor-like kinases; PCC, Pearson correlation coefficient; NCBI, National Center for Biotechnology Information; pI, isoelectric point; NJ, neighbor-joining

* Correspondence to: Institute of Subtropical Forestry, Chinese Academy of Forestry, Fuyang, Hangzhou 311400, P.R. China.
E-mail addresses: xuelhe@163.com (X. He), ichbindora@outlook.com (T. Feng), zhangdayi@tsinghua.edu.cn (D. Zhang), zhuory@gmail.com (R. Zhuo), lmyxin@163.com (M. Liu).

https://doi.org/10.1016/j.ecoenv.2018.09.122
Received 18 July 2018; Received in revised form 13 September 2018; Accepted 28 September 2018

the mechanisms of Cd sequestration and tolerance in *S. alfredii* Hance. Some genes related to Cd uptake (Chen et al., 2017; Liu et al., 2017a; Zhang et al., 2016b), chelation (Zhang et al., 2011b) and alleviation of reactive oxygen species (ROS) damage (Li et al., 2017a) have been characterized. Moreover, several genes are linked to novel Cd-tolerance traits in *S. alfredii* Hance (Liu et al., 2016a; Zhang et al., 2011a). In addition, the transfer and storage of Cd within *S. alfredii* Hance tissues are also explored. For instance, parenchyma cells in leaf mesophyll, stem pith and cortex tissues served as terminal storage sites for Cd sequestration (Tian et al., 2011, 2017). However, how *S. alfredii* Hance senses the Cd-stress and transfers signals within tissues remains unclear.

Leucine-rich repeat receptor-like kinase (LRR-RLK) is the largest family of receptor-like protein kinases (RLKs), actively involved in the regulation of plant growth, development, signal transduction, immunity and stress response (Halter et al., 2014; Park et al., 2014; Pitorre et al., 2010; Wang et al., 2017a). LRR-RLKs consist of three functional domains: an extracellular tandem arrayed leucine-rich repeat domain perceiving signals, a single-pass transmembrane (TM) domain anchoring the protein within the membrane, and an intracellular functional protein kinase (PK) domain participating the signal transduction *via* autophosphorylation and subsequent phosphorylation of specific substrates (Liu et al., 2017b). The huge diversity of LRR and the varying numbers of LRR repeats enable LRR-RLKs to sense a broader spectrum of ligands (ten Hove et al., 2011).

Many experimental studies have unraveled the functions of some representative LRR-RLKs. For example, the Somatic Embryogenesis Receptor Kinase (SERK) family of plant LRR-RLKs is found to participate in the regulation of immune response in plants and play a significant role in triggering the immune response through the interaction with pattern recognition receptors like FLS2 (Santiago et al., 2013; Sun et al., 2013). In Antarctic moss *Pohlia nutans*, 56 LRR-RLK genes were identified that PnLRR-RLK27 functioned as a signaling regulator conferring abiotic stress response and was associated with the regulation of the stress and abscisic-acid (ABA) mediated signaling network (Wang et al., 2017a). However, the large number and huge structural diversity of LRR-RLK family members challenge the complete understanding on the functions and mechanisms of individual genes within the signal transduction pathway, which is complex in plants. Additionally, complementary functions across proteins illustrate the necessity of systematic analyses to understand the precise roles of LRR-RLK family members using bioinformatics tools.

Recently, comprehensive analyses of LRR-RLK gene family have been conducted in *Solanum lycopersicum* (Wei et al., 2015), *Arabidopsis* (Wu et al., 2016), soybean (Zhou et al., 2016), two Citrus species (Magalhães et al., 2016), *Amborella trichopoda* (Liu et al., 2016b), five Rosaceae species (Sun et al., 2017) and other plant species (Liu et al., 2017b). Lin *et al.* performed a comprehensive global stress transcriptome analysis on rice seedlings to reveal the differences between early and delayed response genes, and identified four hub genes associated with the early stress response as protein phosphatase 2C68 (PP2C68), mitogen protein kinase 5 (MPK5), non-expressor pathogenesis-related 1 (NPR1) and LRR-RLK (Lin et al., 2017). Particularly, LRR-RLK genes might be responsible for modulating plant early stress response pathways as early alarm genes.

It is still unclear whether the LRR-RLKs superfamily members function as alarm genes in *S. alfredii* Hance. Exploring their responses to Cd-stress and characterizing their expression dynamics will contribute to the identification of their potential roles as early-transient, early-constant and delayed gene inductions. Therefore, a comprehensive investigation of LRR-RLK family in *S. alfredii* Hance is of great urgency to link the functions of LRR-RLK members to Cd stress response and hyperaccumulation. In the present study, we identified all the putative LRR-RLK genes at genome-scale using a previously published transcriptome database (Han et al., 2016) and analyzed their sequence phylogeny, protein organization, motif conservation, potential co-expression network and transcriptional expression profiles with or without Cd stress. Our work attempts to give an insight into the functions of LRR-RLK family in *S. alfredii* Hance and provide new clues for its Cd/Zn hyperaccumulating mechanisms at transcriptome level.

2. Materials and methods

2.1. Identification of LRR-RLK genes in S. alfredii Hance

To identify LRR-RLK members in *S. alfredii* Hance, we performed a batch BLAST search against *S. alfredii* Hance annotated transcriptome database (Han et al., 2016). *Arabidopsis*, a model eudicot plant, was used as the reference and the amino-acid sequences of all *Arabidopsis* LRR-RLK members were applied in a local BLAST against the protein database using the Blast + software supplied by National Center for Biotechnology Information (NCBI) (Altschul et al., 1997). Of all the proteins, 240 putative LRR-RLKs in *S. alfredii* Hance were selected for E-value less than 10^{-6}. They were further filtered by removing redundant sequences and annotated with functions using SMART (http://smart.embl-heidelberg.de) (Schultz et al., 1998) and PFAM (http://pfam.xfam.org/) (Finn et al., 2014). Conserved domains of LRR and PK for each gene were analyzed by scanning against the PFAM database version 28.0 using HMMER v3.1 (Finn et al., 2011). The sequences with coverage less than 50% were discarded, and overlapping domains were resolved by comparing their E-values and retaining the best-matched domain.

To ensure the presence of PK domains, the filtered LRR-RLK genes were further validated by searching hidden Markov models of the "typical" Pkinase clan [Pkinase (PF00069) and Pkinase_Tyr (PF07714)], obtained from the PFAM database version 28.0 (Finn et al., 2014), against the proteome in each studied species using HMMER v3.1 (Finn et al., 2011) with a gathering threshold. Then, the candidate genes were further analyzed to confirm the presence of signal peptides and TM domains with Signalp v4.0 (http://www.cbs.dtu.dk/services/SignalP/) (Petersen et al., 2011) and TMHMM v2.0 (http://www.cbs.dtu.dk/services/TMHMM/) (Krogh et al., 2001), respectively. Only proteins containing LRR, TM and KD domains were considered as putative LRR-RLK. The basic information of SaLRR-RLKs, including molecular weight (kDa), isoelectric point (pI) and the number of amino acids, was predicted by ExPASy ProtParam (http://web.expasy.org/protparam/) (Gasteiger, 2003).

2.2. Phylogenetic analysis and conserved motif analysis of SaLRR-RLKs

To study the classification and evolutionary relationships of LRR-RLK members in *S. alfredii* Hance, a phylogenetic tree was constructed using MEGA 7.0 (Kumar et al., 2016) *via* the neighbor-joining (NJ) method with bootstrap values from 1000 replicates indicated at each node. Full-length protein sequences of SaLRR-RLKs were subjected to MUSCLE software performing multiple sequence alignment with default settings (Gap opening penalty, −2.9; Gap extend, 0; Hydrophobicity Multiplier, 1.2; Clustering method, UPGMB) (Edgar, 2004).

As the LRR domains are ambiguous and the PK domains are relatively conserved with sizes ranging from 250 to 300 aa (Shiu and Bleecker, 2001), we further constructed a phylogenetic tree of the PK domain using the corresponding ones of *Arabidopsis* LRR-RLKs as reference. The amino acid sequences of PK domain in *S. alfredii* Hance were obtained from PFAM database, while those in *A. thaliana* were from a previous study (Wu et al., 2016). The phylogenetic trees were constructed by MEGA7.0 (Kumar et al., 2016) and visualized by iTOL tool (http://itol.embl.de) (Letunic and Bork, 2011).

2.3. Protein structure and conserved motif analysis of SaLRR-RLKs

Although the basic components of LRR-RLK are coherent as LRR, TM and PK domains, they vary significantly across LRRs. Protein features in each subfamily, *e.g.*, the number of LRR repeats and their