RESEARCH ARTICLE
doi: 10.2306/scienceasia1513-1874.2022.141

ScienceAsia 49 (2023): **85–93**

Identification, *in silico* characterization, and expression analysis of NBS-LRR class of R genes against stem and crown rot disease in *Trifolium alexandrinum* L.

Bukhtawer Nasir[a], **Siddra Ijaz**[a,*], **Imran Ul Haq**[b], **Nasir Ahmad Khan**[b], **Amer Habib**[b]

[a] Centre of Agricultural Biochemistry and Biotechnology (CABB), University of Agriculture, Faisalabad 38000 Pakistan

[b] Department of Plant Pathology, University of Agriculture, Faisalabad 38000 Pakistan

*Corresponding author, e-mail: siddraijazkhan@yahoo.com

Received 17 Dec 2021, Accepted 16 Aug 2022
Available online 10 Oct 2022

ABSTRACT: R genes with colossal contributions to disease resistance possess conserved domains with confounded roles. Based on the conserved domains, R genes encoded proteins are classified. The members of nucleotide-binding site, NBS-Leucine rich repeats, LRR class of R genes are significant candidates involved in plant disease-resistance pathways. In this study, the genetic elements contributing to stem and crown rot disease resistance were explored through transcriptome probing based on conserved NBS domain under disease stress in berseem clover. Expression analysis of identified gene sequences revealed their upregulation during stem and crown rot disease. The *in silico* characterization imparted strong support towards the predictive role of identified RGAs in disease resistance.

KEYWORDS: NBS-LRR domain, agricultural sustainability, Berseem clover, *R* genes

INTRODUCTION

Among diseases, fungal diseases are the most dangerous to plants' health and production. Chemical control measures are in practice to overcome these diseases in crop plants. However, the application of chemical control measurements is compromised by severe environmental effects and new strains of resistant pathogens. Therefore, scientists have been moving towards the understanding of innate defense mechanisms in crop plants. While pathogen attacks, the plant undergoes physiological and biochemical changes to activate defense mechanisms that ultimately trigger systemic acquired resistance (SAR) [1]. The pathogen infects the plant by breaking the first line of defense, i.e., bark and waxes; as the pathogen breaches the first line of defense, the plant starts to secrete secondary metabolites that are lethal for pathogens. Infectious particles enter the cell, replicating and moving towards other plant cells to invade the second defense line.

The R genes confer resistance against pathogens and have been identified in diverse plant species [2]. During the pathogen attack, these R gene encoding products are putative receptors responding to avirulence (avr) gene products. They have been cloned and characterized by monocotyledons and dicotyledon plants [3]. Sequence analysis of Resistance gene analogs (RGAs) or candidate R gene encoding proteins from different plants displayed substantial homology in specific regions that lately are known as conserved domains, i.e., NBS-LRR domain. As the putative role of these conserved domains had been demonstrated in plant disease resistance [4], so, the conserved motifs are being used to recover resistance gene analogs from within plant species or related plant species [5].

In real-time qPCR analysis, it is necessary to normalize the target gene expression using a suitable reference gene. The reference gene expression should be stable among samples taken from different tissues. The reference genes are housekeeping genes. They have constitutive expression levels in different plant cells, tissues, and environmental conditions. The housekeeping genes involve cellular processes such as protein folding, the glycolytic pathway, etc. [6]. The studies for selecting the stable reference genes have been done in various plant species [7].

Hence, in this research study, putative disease resistance gene sequences or RGAs based on the NBS-LRR domain were identified against stem and crown rot in *Trifolium alexandrinum* (berseem clover). For the identification of RGAs, we used the primers that gave cross species amplification in *T. alexandrinum* reported in our previous study [8]. Identified resistance gene sequences were characterized with bioinformatics. For expression analysis of identified putative disease resistance gene sequences using real-time qPCR, a comprehensive reference genes validation is needed in berseem. Therefore, in this study, four potential reference genes, namely *UBC2*, *ACT*, *GAPDH*, and *E1F4a*, were used for an expression stability analysis in *T. alexandrinum*. We monitored the expression pattern of these genes in samples under-investigated stem and crown rot disease conditions. The gene expression stability was validated using statistical algorithms to assign an appropriate reference gene suitable for accurate transcript normalization to study the expression in disease condition.

MATERIALS AND METHODS

Plant material collection and inoculation

Seeds of two cultivars of berseem clover, including Agaiti (susceptible to stem and crown rot disease), and Anmol (tolerant to stem and crown rot disease), were taken from Ayub Agriculture Research Institute (AARI) Faisalabad, Pakistan. Seed sowing of each cultivar was in trays filled with an equal ratio of clay and soil mixture and kept under a glasshouse. Inoculation of plants was performed using the sick soil method by mixing inoculum of *Sclerotinia trifoliorum* [9] (the fungal pathogen of stem and crown rot disease) in the soil.

RNA isolation and cDNA synthesis

Total RNA was extracted from the leave samples of plants after fifteen days post-inoculation (15 dpi). A total of 0.1 g of frozen plant material was ground in liquid nitrogen to fine powder following the procedural channel to get RNA using the GeneJET Plant RNA Purification kit (Thermo Scientific, USA), as per the protocol described by the manufacturer. The extracted RNA sample was treated with RapidOut DNA Removal kit to genomic DNA (gDNA) wipeout. UV visible NANODROP (8000 Spectrophotometer, Thermo Scientific) at an absorbance of 260/280 nm and 260/230 nm was used to determine RNA samples' quality and quantity. The gel electrophoresis (Wealtec USA, Model Elite 300) using 1.2% (w/v) agarose gel was performed to check RNA integrity. The cDNA synthesis was done using RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific), following the manufacturer's protocol. The synthesized cDNA was quantified (at 260/280nm absorbance) using UV visible NANODROP 8000 Spectrophotometer, diluted 20 times, and stored at −20 °C.

Identification of RGAs

Reverse transcription PCR (rt-PCR) analysis was performed using the primers [8], given in Table S1, to identify putative disease resistance gene sequences (RGAs) against stem and crown rot disease in berseem. PCR products were run through gel electrophoresis using 1% (w/v) agarose gel at 80 volts. The gel from the brink of required amplicons was excised following DNA elution using the FavorPrep gel purification kit (FAVORGEN, Biotech Corp., Taiwan). Eluted products were directly sequenced from Eurofins Genomics DNA sequencing services. Generated sequences were trimmed to get high-quality sequences using BioEdit software and *in silico* identified based on homology using the Basic local alignment search tool, Blastx (https://blast.ncbi.nlm.nih.gov). The deduced amino acid sequences of identified resistance gene sequences were generated from the ExPASy translate tool (https://web.expasy.org/translate). PSI-Blast (Position-Specific Iterated BLAST) based on Position-Specific Scoring Matrix (PSSM) was used to find a distant evolutionary relationship of deduced amino acid sequences of identified disease resistance gene sequences. The identified resistance gene sequences were submitted to the NCBI database to get the GenBank accession numbers.

Physicochemical characterization and subcellular localization

Physicochemical characterization of the identified putative disease resistance gene sequences was made by computing isoelectric point, molecular weight, aliphatic and instability index, and grand average of hydropathicity (GRAVY). An online tool, Protparam [10], analyzed the physicochemical properties. The deduced amino acid sequence of each identified putative disease resistance gene sequence against the stem and crown rot disease in berseem was generated using the ExPASy tool for Protparam analysis. However, their subcellular localization was predicted using the CELLO server and WoLF PSORT [11].

Multiple sequence alignment analysis

For multiple sequence alignment analysis, the identified disease resistance gene sequences were aligned with known R gene sequences of NBS-LRR class through MAFFT (Multiple Alignment using Fast Fourier Transform) programs [12] and analyzed in Jalview software [13].

In silico characterization

The web-based tools ScanProsite [14], ProtoNet [15], MOTIF (http://genome.jp/tools/motif), and CD-search [16] were employed to identify the functional sites, conserved regions, domains, and motifs. The SuperFamily database [17] was used to identify family level classifications, and functional family was predicted using the CATH database [18]. The COACH [19] server was used for protein-ligand binding site prediction. However, active site prediction was made using the I-TASSER suite [20], and gene ontology (GO) was determined through the COFACTOR server [21].

Candidate reference genes selection for RT-qPCR

Due to the unavailability of genomic data of berseem clover, expressed sequence tags (ESTs) sequences from *Trifolium pratense* were retrieved from National Centre for Biotechnology Information (NCBI) database, for reference genes identification to transcript normalization in qRT-PCR analysis. The selection of candidate reference genes was based on the homology sequence analysis of retrieved ESTs using Blastn and Blastx algorithms. The gene name and symbols were denoted based on gene ontology annotation by the consulting gene ontology database. We used it to determine the molecular function, involvement in biological processes, and ESTs' cellular location. Gene-

specific primers were designed from selected candidate reference gene sequences in *T. pratense* using the PrimerQuest tool (Integrated DNA Technologies, Inc., USA) for getting PCR amplification in berseem clover. The reaction mixture (10 µl) of PCR analysis was comprised of 5 µl Phusion High-Fidelity PCR Master Mix with HF Buffer (Thermo Scientific), 1 µl cDNA, 1 µM of each forward and reverse primer, and d_3H_2O added to the total volume. However, the thermal profile for PCR analysis was 95 °C for 3 min, followed by a loop of 30 cycles, comprising 95 °C for 30 s, 58 °C for 30 s, and 72 °C for 60 s, followed by a final extension at 72 °C for 10 min. PCR products were resolved on 1% agarose gel using gel electrophoresis. The gel was sliced at the place of required amplicons from its brink. Gel slices were treated using the FavorPrep gel purification kit (FAVORGEN, Biotech Corp) and directly sequenced from Eurofins Genomics DNA sequencing services. Generated sequences were trimmed through BioEdit software to get high-quality (HQ) sequences. These generated sequences were deposited into the NCBI database. They were subjected to the PrimerQuest tool to get primer pairs for RT-qPCR analysis to identify internal reference genes with a highly stable expression for transcript normalization of target genes in berseem clover. Four reference genes were selected as candidates (Table S2), and melt curve analysis was performed to check the specificity of primers for RT-qPCR analysis. The RT-qPCR reaction was performed in a non-skirted thin wall 96 × 0.2 ml plate (Bioplastics Netherlands) containing 1 µl of 20-times diluted cDNA, 2.5 µl Maxima Syber green/ROX-/qPCR master mix, 0.2 µM of each primer in a total reaction volume of 5 µl using CFX96 Touch Real-Time PCR detection system. In the RT-qPCR reaction, water was used as a negative control (no-template control). The reaction was conducted under a two-step cycling thermal profile comprising 50 °C for 2 min (UDG pretreatment), 95 °C for 3 min (initial denaturation), followed by a loop of 40 cycles at 95 °C for 30 s (denaturation), and 58 °C for 30 s (annealing/extension). The quantification cycle (Cq) value was automatically calculated for the reaction using the RT-PCR system with default parameters. After 40 cycles, amplification specificity was evaluated by heating from 60 °C to 95 °C with a ramp speed of 1.9 °C for 60 s, resulting in melting curves.

Expression stability analysis of candidate reference genes

Three different types of statistical algorithm-based programs, BestKeeper [22], NormFinder [23], and geNorm [24], were employed to rank and compare the expression stability of candidate reference genes in an investigated experimental condition. In the BestKeeper method, original Cq values of selected candidate genes were used. However, relative quantitation of original

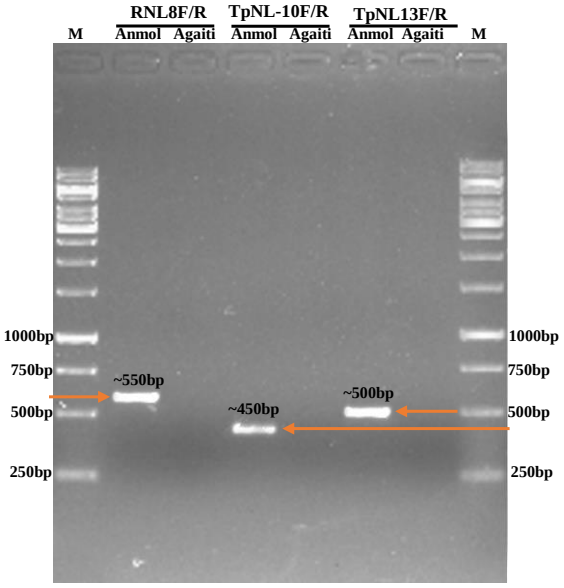

Fig. 1 RT-PCR analysis of *T. alexandrinum* that showed cross-species amplification.

Cq values were made using the formula; $2^{-\Delta Cq}$ where ΔCq = candidate reference gene Cq value − minimum Cq value, previously described by Wang et al [25] for geNorm and NormFinder methods. However, a comprehensive ranking of candidate reference genes was obtained using these three statistical algorithms' ranking results in RefFinder.

Expression analysis

Expression analysis of identified putative disease resistance gene sequences in berseem clover cultivar "Anmol" under stem and crown rot disease challenge was made using RT-qPCR. In this analysis, reference genes that showed stable expression for transcript normalization in expression stability analysis were used. This expression profiling was given differential expression data of identified putative disease resistance gene sequences and validated gene expression normalization accurately. The RT-qPCR primers were designed from the identified putative disease resistance gene sequences using PrimerQuest tool (Table S3).

RESULTS

Identification and *in silico* characterization of disease resistance gene sequences

In rt-PCR analysis, the primer pairs, RNL8-F/RNL8-R, TpNL10-F/TpNL10-R, and TpNL13-F/TpNL13-R gave amplification in berseem cultivar, Anmol (tolerant). However, these primers did not show any amplification in Agaiti (susceptible) under stem and crown rot disease stress (Fig. 1). The amplified products

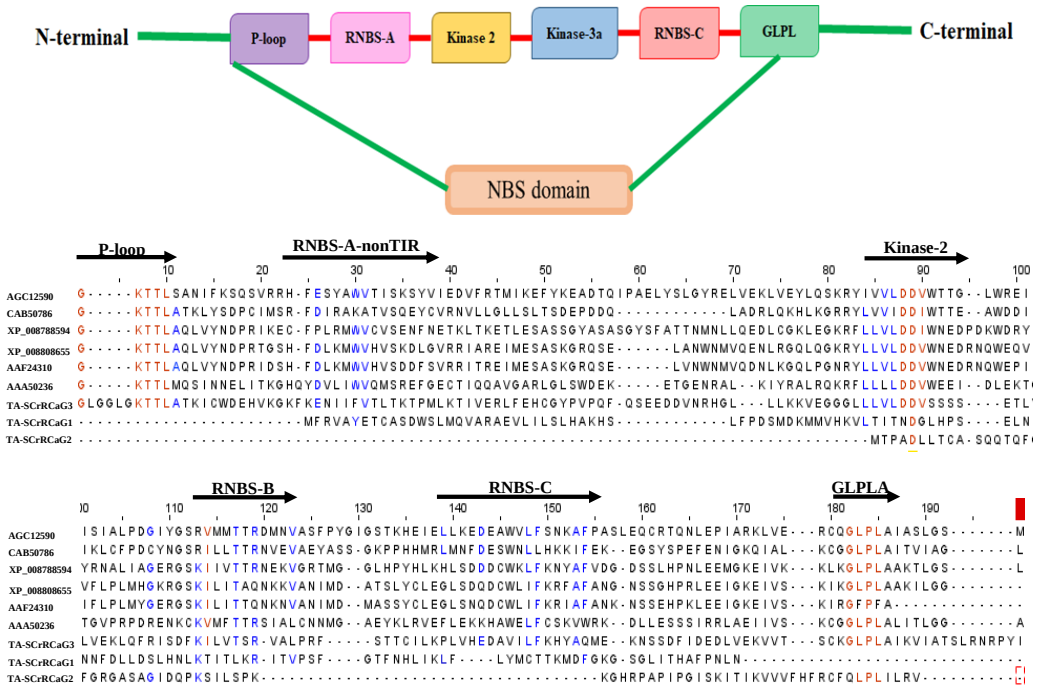

Fig. 2 Multiple Sequence alignment analysis to identify conserved motifs of NBS domain of TA-SCrRCaG1 (RGA1), TA-SCrRCaG2 (RGA2), and TA-SCrRCaG3 (RGA3).

Table 1 Physicochemical properties of identified putative disease resistance gene sequences.

Identified RGAs	Protein length	Mw (Da)	Molecular formula	Isoelectric point (pI)	Instability index (II)	Aliphatic index	GRAVY
TA-SCrRCaG1	120	13611.96	$C_{615}H_{968}N_{160}O_{170}S_9$	7.87	32.40	97.50	0.133
TA-SCrRCaG2	68	7382.81	$C_{335}H_{550}N_{94}O_{87}S_3$	10.56	31.64	98.97	0.144
TA-SCrRCaG3	181	20301.75	$C_{920}H_{1490}N_{238}O_{264}S_6$	7.83	35.12	108.12	0.028

were sequenced, and generated sequences were *in silico* characterized for homology search using BLAST tools (Blastx and PSI-Blast). They showed maximum homology with NBS-LRR protein, disease resistance proteins, and stress response proteins. Hence, the identified resistance gene sequences submitted to the NCBI database were assigned with GenBank accession numbers, MN862280 (TA-SCrRCaG1), MT946504 (TA-SCrRCaG2), and MW030521 (TA-SCrRCaG3). These identified sequences were *in silico* characterized using different web-based tools. The multiple sequence alignment analysis of identified berseem clover RGAs against the stem and crown rot disease with known RGAs sequences of NBS-LRR class in plants [5] displayed the conserved motifs of the NBS domain (P-loop, RNBSA-non TIR, kinase2, kinase 3a (RNBS-B), RNBS-C, and GLPL). The RGA1 (TA-SCrRCaG1) contained RNBSA-non TIR, kinase2, kinase 3a (RNBS-B), and RNBS-C motifs of the NBS domain. The RGA2 (TA-SCrRCaG2) showed kinase2, kinase 3a, and GLPL motifs. However, RGA3 (TA-SCrRCaG3) shows P-loop, RNBSA-non TIR, kinase2, kinase 3a (RNBS-B), RNBS-C, and GLPLA motifs of the NBS domain (Fig. 2). Among the identified putative disease resistance gene sequences, RGA3 (TA-SCrRCaG3) encoded translated protein was characterized as having the NBS-ARC domain. Besides a signature motif P-loop (GLGGLGKT) (ATP/GTP-binding site motif A), the characteristic molecular feature of disease resistance genes or R genes- encoding proteins was also found in the translated sequences of identified RGAs. The ProtoNet tools-based results also identified these putative disease resistance gene sequences as disease-resistant protein-like protein-encoding genes (Table S4). The functional family predicted through the CATH database was disease resistance protein ADR1 (3.80.10.10/FF/105273) for RGA1 and disease resistance protein RPP8 (3.40.50.300/FF/635446) for RGA3. However, it showed no result for RGA2. The SuperFamily database presented RGA3 under P-loop containing nucleoside triphosphate hydrolases superfamily. In the *in silico* characterization of a protein,

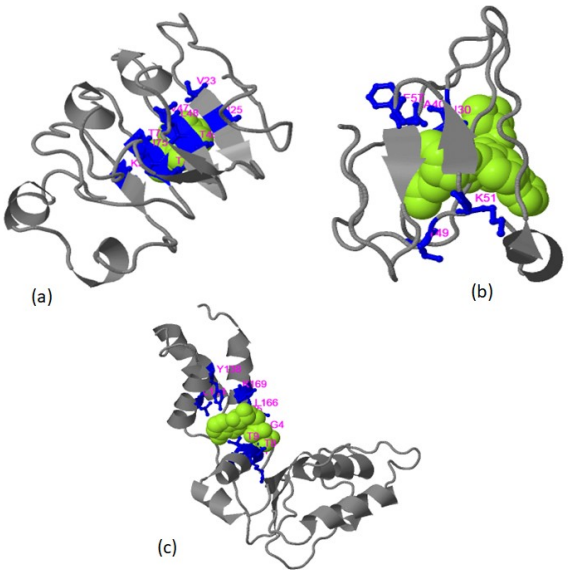

Fig. 3 Protein-ligand binding sites of (a) TA-SCrRCaG1, (b) TA-SCrRCaG2, and (c) TA-SCrRCaG3 predicted using COACH database

knowledge about its physicochemical properties is the core criterion. We used the Protparam web tool to compute the physicochemical properties of identified putative disease resistance gene sequences encoding deduced amino acid sequences. The stability index and aliphatic index scores represented their stable nature with high thermal stability. However, theoretical pI laid them under basic nature proteins. The computed GRAVY values showed their hydrophobic nature (Table 1). The subcellular localization analysis through the CELLO web server predicted the subcellular localization of TA-SCrRCaG1, TA-SCrRCaG2, and TA-SCrRCaG3 to extracellular, nuclear, and nuclear + cytoplasm, respectively, with significant hits. However, WoLF PSORT predicted subcellular protein localization of TA-SCrRCaG1, TA-SCrRCaG2, and TA-SCrRCaG3 in the cytoplasm with maximum hits.

The functional annotation of identified disease resistance gene sequences or RGAs-encoding translated proteins was made by COFACTOR, COACH, and I-TASSER. The COACH database predicted 23Val (V), 25Ile (I), 47Val (V), 48Leu (L), 49Thr (T), 74Thr (T), 75Ile (I), 76Thr (T), 93Lys (K), as protein-ligand binding sites in TA-SCrRCaG1 (RGA1) (Fig. 3a). For TA-SCrRCaG2 (RGA2), the ligand-binding sites were 29Ser (S), 30Ile (I), 40Ala (A), 49Thr (T), 51Lys (K), 57Phe (F) (Fig. 3b). However, the binding residues were, 4Gly (G), 5Leu (L), 6Gly (G), 7Lys (K), 8Thr (T), 9Thr (T), 134Leu (L), 138Tyr (Y), 165Pro (P), 166Leu (L), 169Lys (K) in TA-SCrRCaG3 (RGA3) (Fig. 3c). In RGA3, 7Lys (K), 8Thr (T), 9Thr (T) are ATP binding site residues of the P-loop domain. These protein-ligand binding sites of each identified RGA were based on a high C-score (confidence score), which relates to a more reliable prediction. The active site residues of TA-SCrRCaG1 predicted using the I-TASSER server were 54Gly (G), 60Leu (L), and 62Asn (N). The gene ontology (GO) explored through the COFACTOR server is given in Table S5, representing their predominant role in defense response to biotic stress and external stimuli.

Selection and stability analysis of candidate reference genes in the RT-qPCR analysis

Genomic information of berseem clover is unavailable, so 38104 ESTs of *T. pratense* were retrieved from the NCBI database and subjected to homology search using Blastn and Blastx algorithms. Four reference genes (Table S2) involved in different biological processes were selected to identify candidate reference genes in berseem clover. The generated HQ sequences got GenBank accession numbers of *TaACT* (MW533148), *TaUBC2* (MW560482), *TaEIF4a* (MW548610), and *TaGAPDH* (MW286365) lodging to the NCBI database. The primers for RT-qPCR analysis (Table S2) for assessing and identifying reliable reference genes with the stable expression for transcript normalization in berseem clover were designed from generated sequences of candidate reference genes.

Melt curve analysis yielded a single sharp peak with no primer dimer in all primer pairs. Besides, no amplification was observed in water control, proving the absence of primer dimers and reaction contamination. We used three statistical algorithms for reference gene stability analysis: BestKeeper, NormFinder, and geNorm. The BestKeeper is a Microsoft Excel-based program with utilization in stability analyses. It determines the stability ranking by comparing the candidate reference genes' expression stability based on the percentage of crossing point (% CP) to the BestKeeper Index and standard deviation (SD) from the geometric mean of the candidate reference genes' Cq values. The genes with the lowest CP and SD values are considered the best reference genes for accurate transcript normalization. In our experimental samples, BestKeeper analysis identified more stable reference genes: *TaEIF4a* (1.81 ± 0.28), *TaGAPDH* (6.77 ± 1.28), *TaACT* (15.63 ± 2.48), and *TaUBC2* (16.57 ± 3.80) respectively (Table 2). We also determined the stability analysis for candidate reference gene expression using the geNorm statistical algorithm (Table 2). This program does not use original Cq values; instead, it transforms them into relative quantities using the ΔCq method. It determines stability value (M) based on the geometric mean of all candidate reference genes and the average pair-wise variation of a gene from all other reference genes in a set of samples. A gene with a low stability value is identified as stable, and vice versa. The geNorm statistical tool identified

Table 2 Stability ranking of the reference genes in *T. alexandrinum* (berseem clover).

Rank	BestKeeper		geNorm		NormFinder		RefFinder	
	Genes	CP(%) ± SD	Genes	Normalization Value (M)	Genes	Stability Value (SV)	Genes	Geomean of ranking value
1	*TaEIF4a*	0.275	*TaUBC2*|*TaGAPDH*	0.018	*TaEIF4a*	0.023	*TaGAPDH*	1.414
2	*TaGAPDH*	1.28			*TaGAPDH*	0.075	*TaEIF4a*	1.732
3	*TaACT*	2.48	*TaEIF4a*	0.042	*TaUBC2*	0.136	*TaACT*	2.06
4	*TaUBC2*	3.795	*TaACT*	0.343	*TaACT*	0.644	*TaUBC2*	4

TaUBC2|*TaGAPDH* with the lowest stability value of 0.018, the best reference gene for normalization in RT-qPCR analysis, followed by *TaEIF4a* with "M" value, 0.042, and *TaACT* with 0.343 stability value, respectively. However, the reference genes with stability value (M) below the cut-off range, i.e., < 1.5 using geNorm analysis, are recommended as significantly best reference genes for high gene expression stability. The ranking of candidate reference genes for gene expression stability was also explored using the NormFinder algorithm. This statistical tool determines gene expression stability of candidate reference genes based on intra and inter-group variations. This program identified the *TaE1F4a* reference gene as highly significant expression stability in our experimental samples by scoring a stability value of 0.023. However, other reference genes, *TaGAPDH*, *TaUBC2*, and *TaACT*, ranked second, third, and fourth in expression stability with stability values of 0.075, 0.136, and 0.644, respectively (Table 2). The NormFinder results were similar compared with the BestKeeper results. Both statistical algorithms supported the *TaE1F4a* reference gene with highly stable expression in our experimental sample's condition, followed by the *TaGAPDH* gene in berseem clover for transcript normalization in RT-qPCR analyses. Comprehensive ranking based on geomean values using the RefFinder tool compared the results of BestKeeper, geNorm, and NormFinder programs to accurately validate the expression stability ranking of candidate reference genes for transcript normalization. The *TaGAPDH* reference gene scored 1.41 geomean ranking value in the overall comprehensive ranking, followed by *TaEIF4a*, *TaACT*, and *TaUBC2* reference genes with geomean ranking values of 1.73, 2.06, and 4.00, respectively (Table 2).

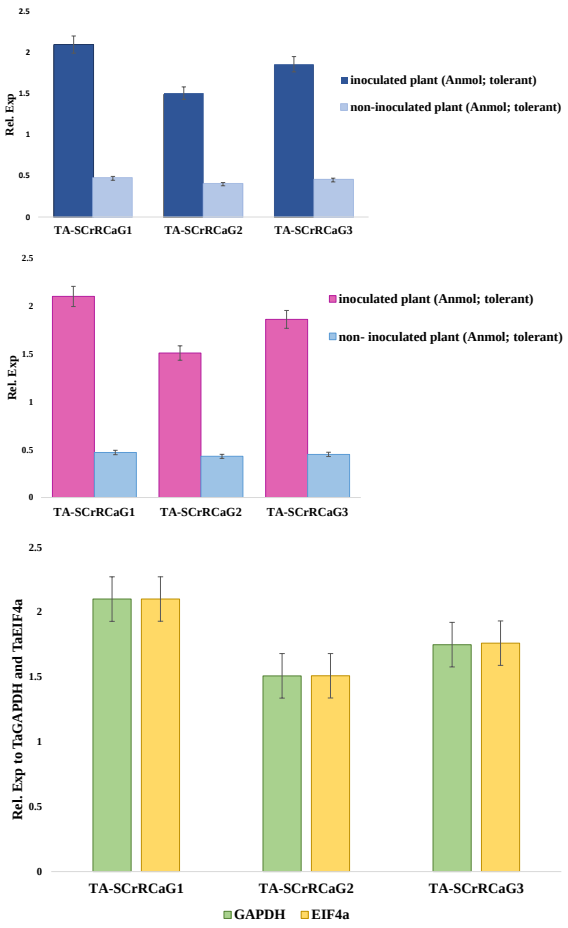

Fig. 4 Gene expression analysis of identified disease resistance gene sequences in berseem clover cultivar, "Anmol" under stem and crown rot disease challenge. (a) and (b) Differential gene expression pattern of identified gene sequences normalizing with *TaGAPDH* and *TaEIF4a* reference genes, respectively. (c) Validation of top-ranked reference genes, *TaGAPDH* and *TaEIF4a* for accurate transcript normalization by determining the relative transcript levels of TA-SCrRCaG1, TA-SCrRCaG2, and TA-SCrRCaG3 genes sequences.

Gene expression analysis and validation of selected reference genes

Three putative disease resistance gene sequences, TA-SCrRCaG1, TA-SCrRCaG2, and TA-SCrRCaG3 (identified from the transcriptome of the berseem clover cultivar Anmol tolerant to stem and crown rot disease) were investigated for determining their expression pattern using RT-qPCR analysis. This experiment was carried out using two reference genes after accurate normalization. The Reference genes, *TaGAPDH* and *TaEIF4a*, showed stable expression in our experimental sample's condition for reliable normalization by acquiring top ranking under different statistical algorithm programs. The expression levels of identi-

fied disease resistance gene sequences, TA-SCrRCaG1, TA-SCrRCaG2, and TA-SCrRCaG3, were determined. These gene sequences showed upregulation under stem and crown rot disease challenge in berseem clover variety Anmol (Fig. 4a,b). TA-SCrRCaG1 was expressed at a relatively high level, followed by TA-SCrRCaG3. However, TA-SCrRCaG2 showed an intermediate expression level. The expression analysis also validated the selected reference genes for accurate transcript normalization by determining the relative transcript levels of TA-SCrRCaG1, TA-SCrRCaG2, and TA-SCrRCaG3 gene sequences. Relative transcript levels of these gene sequences were calculated after normalizing with the top-ranked candidate reference genes as determined by BestKeeper, NormFinder, geNorm, and then as recommended by RefFinder in its comprehensive ranking (Fig. 4c). We observed more transcript abundance of TA-SCrRCaG1 gene sequence followed by TA-SCrRCaG3 gene sequence when normalized using reference genes, *TaEIF4a* revealed stable and constant relative expression pattern, as obtained when the *TaGAPDH* reference gene, determined and recommended by geNorm and RefFinder as top-ranked, was used for normalization.

DISCUSSION

Stresses, either biotic or abiotic, bring severe losses to plants every year. The bacteria, fungi, viruses, and nematodes infect plants and cause plant diseases, resulting in great loss in the agricultural economy by reducing crop plants' health and productivity. However, fungal diseases pose potential threats to crop production. While pathogens attack, plants activate their defense mechanisms [26]. The NBS-LRR domain-containing proteins directly block the pathogens by activating hypersensitive responses [5]. It has been documented that the R genes encoding proteins of the NBS-LRR superfamily are intracellular immune receptors that directly or indirectly recognize specific pathogen effectors and triggers plant defense responses.

Literature supports that the NBS-LRR class of R genes plays a crucial role in crop plants against diverse insects and pathogens. The members of this class possess common evolutionary origins and typical structure patterns [27]. Degenerate primers-based PCR strategies have been used to isolate and identify putative *R* genes from diverse plant species [28, 29]. A diverse array of R genes in plants had been cloned and displayed those encoded proteins having an NBS domain followed by LRR domain, and that is why they proved to be the members of the NBS-LRR superfamily [30]. In this study, putative disease-resistance gene sequences or RGAs based on the NBS-LRR domain were identified and recovered from berseem clover against stem and crown rot disease. The translated protein sequences of identified RGAs showed maximum homology with NBS-LRR protein, disease resistance proteins, and stress response proteins. Due to genetic code degeneracy, in RGAs, the comparison of sequences at the protein level has shown more homology with NBS-LRR region than at the nucleotide level [28].

Subcellular localization of R proteins significantly impacts the role against pathogen resistance. Several R proteins had been identified as having nuclear and cytoplasmic localization; even in most R protein sequences; no clear nuclear localization signals were identified. Indeed, in most cases, the R proteins get built up in the nucleus to respond to pathogen infection [31]. The CELLO and WoLF PSORT programs predict 12 subcellular localization classes in eukaryotes: cytoplasm, chloroplast, ER, extracellular space, golgi bodies, nucleus, lysosome vacuole, mitochondria, plasma membrane, peroxisomes, and the cytoskeleton [32]. Recent studies of NBS-LRR class receptors revealed that the R proteins are localized to and inside the nucleus to activate defense responses [33]. MLA10 R protein of *Hordeum vulgare* (Barley), which confers the resistance to *Blumeria graminis* f. sp. *hordei* (powdery mildew fungus), accumulates in the nucleus and is also found in the cytoplasm [33]. In *Arabidopsis*, β-*N*-acetylhexosaminidases, HEXO3 was soluble in extracellular space [34]. The HEXO proteins were involved in plant defense mechanisms and associated with fruit ripening by hydrolyzing *N*-acetylglucose residues from chitin. Similarly, our study predicted the subcellular localization of identified RGAs in extracellular, nuclear, and nuclear +cytoplasm with significant hits using CELLO; However, WoLF PSORT predicted their subcellular localization in the cytoplasm. The CELLO program predicts the subcellular localization of the protein in eukaryotes to 12 classes; however, WoLF PSORT predicts localization to 12 classes specifically in plants rather than in eukaryotes in general [35]. However, no nuclear localization signals were detected in deduced amino acid sequences of identified RGAs (TA-SCrRCaG1, TA-SCrRCaG2 and TA-SCrRCaG3) using NLStradamus program. Similarly, no signal peptide was found in translated amino acid sequences of identified RGAs, using TargetP bio-tools [8]. That's why we did not mention these tools in the manuscript.

Gene expression analysis reveals the contribution of regulatory networks of metabolic, genetic, and signaling pathways, which trigger plant-pathogen interaction. Real-time qPCR is used to analyze the gene expression level in tissues. The stable reference genes are commonly used for data normalization in gene expression experiments. The use of transcript abundance data from a species can identify reference genes in other species. Such an approach was used by Tashiro et al [36] to identify reference genes in *Vitis vinifera* L. (grapevine), based upon evidence from *Arabidopsis thaliana*, which might have broader application.

In berseem clover, transcript abundance stability was evaluated using four housekeeping genes, *ACT*, *GAPDH*, *EIF4a*, and *UBC2*, at two disease levels by BestKeeper and NormFinder geNorm, and RefFinder analysis methods. We found that the *TaGAPDH* reference gene scored a high geomean ranking value in the overall comprehensive ranking, followed by *TaEIF4a*, *TaACT*, and *TaUBC2* reference genes. *GAPDH* has been universally used as a candidate reference gene for RT-qPCR analysis [37].

Similarly, Tian et al [38] identified that the GAPDH reference gene was stably expressed in parsley and carrot leaves in response to hormone stimuli and more stable in *Arabidopsis pumila* under salt stress, heat, and drought stresses [39]. Yan and his colleagues conducted the research, which revealed that in perennial ryegrass under stress conditions, the *eIF4A* could be used as a stable reference gene expressed under herbicide stress in *Eleusine indica* [40]. TaACT was the least stable reference gene in our study, which was also ranked unsuitable for *A. pumila* under drought stress. However, low-level transcript abundance could be useful when studying gene expression at a low level [39].

Different studies suggested that using more than one reference gene could increase RT-qPCR's accuracy [24]. Generally, two reference genes are recommended for normalization. Hence, our study investigated the expression analysis of identified RGAs (TA-SCrRCaG1, TA-SCrRCaG2, and TA-SCrRCaG3) using the expression pattern of two reference genes, i.e., *TaGAPDH* and *TaEIF4a*.

CONCLUSION

Our study identified three putative disease-resistance gene sequences in berseem clover against stem and crown rot disease. These sequences were differentially expressed with upregulated expression in the Anmol tolerant cultivar. We further examined these identified genetic elements for R proteins features and expression profiling in response to stem and crown rot stress. The *in silico* characterization of their translated amino acid sequences showed that they possess motifs that are the attributes of disease resistance genes encoding proteins. Expression analysis of identified RGAs showed more contribution of putative disease-resistance gene sequence, TA-SCrRCaG1 towards stem and crown rot disease resistance in berseem clover.

Appendix A. Supplementary data

Supplementary data associated with this article can be found at http://dx.doi.org/10.2306/scienceasia1513-1874.2022.141.

Acknowledgements: This work was supported by Dr. Siddra Ijaz Laboratory, "Molecular Biology of Plant Disease Resistance Lab" of Centre of Agricultural Biochemistry and Biotechnology (CABB), University of Agriculture Faisalabad, and Dr. Imran Ul Haq Laboratory, "Fungal Molecular Biology Lab", Department of Plant Pathology, University of Agriculture Faisalabad.

REFERENCES

1. Haq IU, Ijaz S (2020) History and recent trends in plant disease control: An overview. In: Haq IU, Ijaz S (eds) *Plant Disease Management Strategies for Sustainable Agriculture through Traditional and Modern Approaches*, Springer, Cham, Switzerland, pp 1–13.
2. Cooley MB, Pathirana S, Wu HJ, Kachroo P, Klessig DF (2000) Members of the Arabidopsis HRT/RPP8 family of resistance genes confer resistance to both viral and oomycete pathogens. *Plant Cell* **12**, 663–676.
3. Haq IU, Ijaz S (2020) *Plant Disease Management Strategies for Sustainable Agriculture through Traditional and Modern Approaches*, 1st edn, Springer, Cham, Switzerland.
4. Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res* **8**, 1113–1130.
5. Rachana KE, Naganeeswaran SA, Fayas TP, Thomas RJ, Rajesh MK (2016) Cloning, characterization, and expression analysis of NBS-LRR-type resistance gene analogues (RGAs) in coconut. *Acta Bot Croat* **75**, 1–10.
6. Li MY, Wang F, Jiang Q, Wang GL, Tian C, Xiong AS (2016) Validation and comparison of reference genes for qPCR normalization of celery (*Apium graveolens*) at different development stages. *Front Plant Sci* **7**, 313.
7. Raman B, Priyanka M, Cassone BJ, Praveen M, Redinbaugh MG, Andy M (2015) Recommended reference genes for quantitative PCR analysis in soybean have variable stabilities during diverse biotic stresses. *PloS One* **10**, e0134890
8. Nasir B, Ijaz S, Awan FS, Haq IU (2021) Genome-wide probing of NBS-LRR encoding genes in red clover (*Trifolium pratense* L) for the identification of resistance gene analogs in *Trifolium alexandrinum* L. *ScienceAsia* **47**, 425–433.
9. Faraz A, Haq IU, Ijaz S (2021) First report of *Sclerotinia trifoliorum* stem and crown rot on *Trifolium alexandrinum* in Pakistan. *J Plant Pathol* **103**, 735–736.
10. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. In: Walker JM (ed) *The Proteomics Protocols Handbook*, Humana Press, New Jersey, United States, pp 571–607.
11. Yu CS, Chen YC, Lu CH, Hwang JK (2006) Prediction of protein subcellular localization. *Proteins* **64**, 643–651.
12. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey AR, et al (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res* **47**, W636–W641.
13. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2: a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191.
14. Sigrist CJ, De Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I (2012) New and continuing developments at PROSITE. *Nucleic Acids Res* **41**, D344–D347.
15. Rappoport N, Karsenty S, Stern A, Linial N, Linial M (2012) ProtoNet 6.0: organizing 10 million protein se-

quences in a compact hierarchical family tree. *Nucleic Acids Res* **40**, D313–D320.
16. Marchler-Bauer A, Bryant SH (2004) CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* **32**, W327–W331.
17. Pandurangan AP, Stahlhacke L, Oates ME, Smithers B, Gough J (2018) The SUPERFAMILY 2.0 database: a significant proteome update and a new webserver. *Nucleic Acids Res* **47**, D490–D494.
18. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM (1997) CATH: a hierarchic classification of protein domain structures. *Structure* **5**, 1093–1109.
19. Yang J, Roy A, Zhang Y (2013) Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics* **29**, 2588–2595.
20. Yang J, Zhang Y (2015) I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res* **43**, W174–W181.
21. Zhang C, Freddolino PL, Zhang Y (2017) COFACTOR: improved protein function prediction by combining structure, sequence, and protein-protein interaction information. *Nucleic Acids Res* **45**, W291–W299.
22. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper-Excel-based tool using pair-wise correlations. *Biotechnol Lett* **26**, 509–515.
23. Andersen CL, Jensen JL, Ørntoft TF (2004) Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res* **64**, 5245–5250.
24. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**, research0034.1.
25. Wang H, Wen H, Li Y, Zhang K, Liu Y (2018) Evaluation of potential reference genes for quantitative RT-PCR analysis in spotted sea bass (*Lateolabrax maculatus*) under normal and salinity stress conditions. *PeerJ* **6**, e5631.
26. Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, Yamaguchi-Shinozaki K, Shinozaki K (2006) Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. *Curr Opin Plant Biol* **9**, 436–442.
27. Baldi PAOLO, Patocchi A, Zini E, Toller C, Velasco R, Komjanc M (2004) Cloning and linkage mapping of resistance gene homologues in apple. *Theor Appl Genet* **109**, 231–239.
28. Totad AS, Fakrudin B, Kuruvinashetti MS (2005) Isolation and characterization of resistance gene analogs (RGAs) from sorghum (*Sorghum bicolor* L. Moench). *Euphytica* **143**, 179–188.
29. Babar M, Ijaz S, Khan MS, Haq IU (2021) Computational genomics-based probing of resistance gene analogs (RGAs) in mungbean under Cercospora leaf spot disease challenge. *Pak J Agri Sci* **58**, 1–14.
30. Hulbert SH, Webb CA, Smith SM, Sun Q (2001) Resistance gene complexes: evolution and utilization. *Annu Rev Phytopathol* **39**, 285–312.
31. Meier I, Somers DE (2011) Regulation of nucleocytoplasmic trafficking in plants. *Curr Opin Plant Biol* **14**, 538–546.
32. Yu CS, Chen YC, Lu CH, Hwang JK (2006) Prediction of protein subcellular localization. *Proteins* **64**, 643–651.
33. Shen QH, Saijo Y, Mauch S, Biskup C, Bieri S, Keller B, Seki H, Ülker B, et al (2007) Nuclear activity of MLA immune receptors links isolate-specific and basal disease-resistance responses. *Science* **315**, 1098–1103.
34. Liebminger E, Veit C, Pabst M, Batoux M, Zipfel C, Altmann F, Mach L, Strasser R (2011) β-*N*-acetylhexosaminidases HEXO1 and HEXO3 are responsible for the formation of paucimannosidic *N*-glycans in *Arabidopsis thaliana*. *J Biol Chem* **286**, 10793–10802.
35. Goldberg T, Hamp T, Rost B (2012) LocTree2 predicts localization for all domains of life. *Bioinformatics* **28**, i458–i465.
36. Tashiro RM, Philips JG, Winefield CS (2016) Identification of suitable grapevine reference genes for qRT-PCR derived from heterologous species. *Mol Genet Genom* **291**, 483–492.
37. Wang X, Wu Z, Bao W, Hu H, Chen M, Chai T, Wang H (2019) Identification and evaluation of reference genes for quantitative real-time PCR analysis in *Polygonum cuspidatum* based on transcriptome data. *BMC Plant biol* **19**, 498.
38. Tian C, Jiang Q, Wang F, Wang GL, Xu ZS, Xiong AS (2015) Selection of suitable reference genes for qPCR normalization under abiotic stresses and hormone stimuli in carrot leaves. *PLoS One* **10**, e0117569.
39. Jin Y, Liu F, Huang W, Sun Q, Huang X (2019) Identification of reliable reference genes for qRT-PCR in the ephemeral plant *Arabidopsis pumila* based on full-length transcriptome data. *Sci Rep* **9**, 8408.
40. Chen J, Huang Z, Huang H, Wei S, Liu Y, Jiang C, Zhang J, Zhang C (2017) Selection of relatively exact reference genes for gene expression studies in goosegrass (*Eleusine indica*) under herbicide stress. *Sci Rep* **7**, 46494.

Appendix A. Supplementary data

Table S1 Primers used to identify the putative disease resistance gene sequences against stem and crown rot disease in *T. alexandrinum*.

Sr #	Primer name	Primer sequence (5′ to 3′ direction)	Reference
1	RNL1-F RNL1-R	ACAACCCTTGCTACGACTTTAT CTATTGGGTGGGAACAAGATGA	[9]
2	RNL2-F RNL2-R	GGTGGAATTGGAAAGACAACAC CCCTTCCAATCTATCACCTCTTC	[9]
3	RNL4-F RNL4-R	AATCATCACGACCAGGGATAAG CTTAATTGGGATGGACCGAAATG	[9]
4	RNL6-F RNL6-R	GCAGAGACTACCATGACCTAAC TGATTGTGGTGAAGACCCTG	[9]
5	RNL8-F RNL8-R	CTGCCCGTATATTGTCGGTATC GTCGGTATAGCTTCCAAGTCAG	[9]
6	TpNL9-F TpNL9-R	TCCGAGTCCAGTCACCTAAT GGTTAAGCTACCCAAACCTCTC	[9]
7	TpNL10-F TpNL10-R	CATCTTGTTCCCACCTCACTT TGCTCGCTCTGCTCTTATTG	[9]
8	TpNL11-F TpNL11-R	GAAACGGTCTCCAAGGGAATAG CAACAGCAAGAGCCTGGTATAG	[9]
9	TpNL12-F TpNL12-R	GGGACAAAGGAAGAGGAAGAAG GACAAGGTTAGGATGGTGAAGTAG	[9]
10	TpNL13-F TpNL13-R	CTGCAAAGGGATTGGAGTATCT CAGCTATAACCGGACGCATATTA	[9]
11	TpNL14-F TpNL14-R	GATGAGATTTGGTGCTGGTATTG TATGACCGGAGGATTAGCTTTG	[9]

Table S2 Candidate reference genes used for the assessment of the expression stability in qRT-PCR analysis.

Gene name	Accession no.	Gene symbol	Primer sequences (5′ to 3′) for General PCR	Amplicon size (bp)	Primer sequences (5′ to 3′) for qRT-PCR	Amplicon size (bp)	GO annotation
Actin	AY372368	*ACT*	AGGCTGTCCTCTCC CTTTAT GGTGGAGCCACAAC CTTAAT	589	AGTGGTCGTACAACTG GTATTG GCAAGATCCAAACGAA GAATGG	110	Structural constituent of cytoskeleton
Glyceraldehyde-3-phosphate dehydrogenase	BB911051	*GAPDH*	CATCAAGCAAGGAC TGGAGAG CTGTCACCAACAAA GTCAGTAGA	298	CACCGTTGATGTTT CAGTTGTT CTTGAGCTTGCCCT CTGATT	106	Dehydrogenase, Oxidoreductase in glycolysis and gluconeogenesis
Eukaryotic Initiation factor 4a	BB919542	*EIF4a*	TGCTACAATGCCAC CAGAAG CCACTACGACCAAT ACGATGAAG	453	GCTCTTCACGAGTG CTCATAAC CAGGTTGGGTAGGC AAATCATAG	99	Hydrolase activity DNA binding ATP binding
Ubiquitin-conjugating enzyme E2A	BB916552	*UBC2*	GTCTGACTCTTCTT CACCTCTC GCTCAACAATCTCG CGTATTC	490	GTGTGATCCAAACC CAAACTC GCTCAACAATCTCG CGTATTC	101	DNA repair and recombination proteins

Table S3 qRT-PCR primer pairs used for expression analysis of identified disease resistance gene sequences in berseem clover against stem and crown rot disease.

Sr #	Primer name	qRT-PCR primer sequence (5′ to 3′ direction)	Length (bp)	Amplicon size
1	TA-SCrRCaG1-F	GGTGCACAAAGTTCTGACAATC	22	∼129
	TA-SCrRCaG1-R	CCGAAGGAAGGGACAGTAATTC	22	
2	TA-SCrRCaG2-F	CTGTTCTCGTTGAGGACTCTTATG	24	∼111
	TA-SCrRCaG2-R	GATCGACTTCGGTTGGTCAAT	21	
3	TA-SCrRCaG3-F	GTGGTTATCCTGTTCCTCAATTTC	24	∼118
	TA-SCrRCaG3-R	GAAGACGAAACATCATCAAGAACA	24	

Table S4 Conserved domains of identified RGAs with ScanProsite, CD search, Motif, and ProtoNet servers.

Identified RGAs	ScanProsite	CD-search	Motif	ProtoNet
TA-SCrRCaG1 (RGA1)	• N-myristoylation site • Protein kinase C phosphorylation site • cAMP and cGMP dependent protein kinase phosphorylation site • Casein kinase II phosphorylation site	–	• TCP-1 (CTT or eukaryotic type II) chaperonin family • chaperonin_type_I_II, chaperonin families, type I and type II • Cpn60_TCP1, TCP-1/cpn60 chaperonin family	• Disease resistance protein-like protein
TA-SCrRCaG2 (RGA2)	• Casein kinase II phosphorylation site • N-myristoylation site • Protein kinase C phosphorylation site	–	–	• Cytochrome P450 like_TBP (TATA box binding protein)
TA-SCrRCaG3 (RG3)	• ATP/GTP-binding site motif A (P-loop) • N-myristoylation site • Casein kinase II phosphorylation site • Tyrosine kinase phosphorylation site 1 • Protein kinase C phosphorylation site N-glycosylation site	NB-ARC superfamily	• NB-ARC • The AAA+ (ATPases Associated with a wide variety of cellular Activities) superfamily	• Disease resistance protein-like protein

Table S5 Functional characterization of identified disease-resistance gene sequences in berseem clover against stem and crown rot disease with the aid of bioinformatics.

Identified putative disease resistance genes or RGAs	Molecular function	Biological process	Cellular component
TA-SCrRCaG1	• catalytic activity • transferase activity • kinase activity • ATP binding • protein kinase activity • protein serine/threonine kinase activity • transmembrane receptor protein tyrosine kinase activity	• response to an external stimulus • defense response to bacterium • cellular process • cellular macromolecule metabolic process • protein metabolic process • peptidyl-tyrosine phosphorylation	• cell part • membrane • plasma membrane • an integral component of membrane
TA-SCrRCaG2	cis-trans isomerase activity	peptidyl-proline modification	• intracellular part • intracellular organelle part • cytoplasmic part • membrane • endoplasmic reticulum membrane • membrane-bounded organelle • intracellular membrane-bounded organelle • cytoplasm
TA-SCrRCaG3	• heterocyclic compound binding • organic cyclic compound binding • nucleotide-binding	• response to stress • defense response • response to biotic stimulus • response to an external stimulus • response to other organisms • defense response to other organisms • single-organism cellular process • cellular response to stress • programmed cell death • regulation of the cellular process	• cell part • intracellular part • membrane • protein complex • intracellular organelle • intracellular membrane-bounded organelle • plasma membrane • nucleus • cytoplasmic part