Libertas Academica
FREEDOM TO RESEARCH

Open Access: Full open access to this and thousands of other papers at http://www.la-press.com.

Evolutionary Bioinformatics

A Multistep Screening Method to Identify Genes Using Evolutionary Transcriptome of Plants

Chang-Kug Kim[1], Hye-Min Lim[1], Jong-Kuk Na[2], Ji-Weon Choi[3], Seong-Han Sohn[1], Soo-Chul Park[4], Young-Hwan Kim[5], Yong-Kab Kim[6] and Dool-Yi Kim[2]

[1]Genomics Division, National Academy of Agricultural Science (NAAS), Rural Development Administration (RDA), Suwon, Korea. [2]Molecular Breeding Division, NAAS, RDA, Suwon, Korea. [3]Vegetable Science Division, National Institute of Horticultural and Herbal Science, Suwon, Korea. [4]The National Center for GM Crops, RDA, Suwon, Korea. [5]Policy Development Office, Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries, Anyang, Korea. [6]School of Electrical Information Communication Engineering, Wonkwang University, Iksan, Korea.

ABSTRACT: We introduced a multistep screening method to identify the genes in plants using microarrays and ribonucleic acid (RNA)-seq transcriptome data. Our method describes the process for identifying genes using the salt-tolerance response pathways of the potato (*Solanum tuberosum*) plant. Gene expression was analyzed using microarrays and RNA-seq experiments that examined three potato lines (high, intermediate, and low salt tolerance) under conditions of salt stress. We screened the orthologous genes and pathway genes involved in salinity-related biosynthetic pathways, and identified nine potato genes that were candidates for salinity-tolerance pathways. The nine genes were selected to characterize their phylogenetic reconstruction with homologous genes of *Arabidopsis thaliana*, and a Circos diagram was generated to understand the relationships among the selected genes. The involvement of the selected genes in salt-tolerance pathways was verified by reverse transcription polymerase chain reaction analysis. One candidate potato gene was selected for physiological validation by generating *dehydration-responsive element-binding 1* (*DREB1*)-overexpressing transgenic potato plants. The *DREB1* overexpression lines exhibited increased salt tolerance and plant growth when compared to that of the control. Although the nine genes identified by our multistep screening method require further characterization and validation, this study demonstrates the power of our screening strategy after the initial identification of genes using microarrays and RNA-seq experiments.

KEYWORDS: microarray, multistep method, RNA-seq, salt tolerance

CITATION: Kim et al. A Multistep Screening Method to Identify Genes Using Evolutionary Transcriptome of Plants. *Evolutionary Bioinformatics* 2014:10 69–78 doi: 10.4137/EBO.S14823.

RECEIVED: February 12, 2014. **RESUBMITTED:** March 25, 2014. **ACCEPTED FOR PUBLICATION:** March 28, 2014.

ACADEMIC EDITOR: Jike Cui, Associate Editor

TYPE: Original Research

FUNDING: This study was conducted with support from the Research Program for Agricultural Science and Technology Development (Project no PJ006651 and PJ00867701) of the National Academy of Agricultural Science, and the Next-Generation BioGreen 21 Program (SSAC, Grant No PJ009614 and PJ00951407), Rural Development Administration, Republic of Korea.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

COPYRIGHT: © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

CORRESPONDENCE: dykim22@korea.kr

Introduction

Plants are continuously exposed to abiotic stresses that negatively affect their growth. Plants can adapt to these stresses at the molecular, cellular, and physiological levels. In many agriculturally important plants, salt stress negatively affects germination, root development, and biomass production, which results in significant yield losses.[1] The salt-stress signaling can be divided into three functional categories: ionic and osmotic stress signaling for the reestablishment of cellular homeostasis; detoxification signaling to repair stress damages; and signaling to coordinate cell division.[2]

Potato (*Solanum tuberosum L.*) is moderately salt-sensitive to moderately salt-tolerant crops; growth and yield are considerably affected by salt stress.[3] In potato growth metabolism, the signaling pathway pointed to the importance of the plant hormone, abscisic acid; the calcium sensor, calcineurin B-like 1; potassium homeostasis; MYB; dehydration response element binding factor (*DREB*); members of the ethylene response factor/APETALA2 (AP2)

transcription factor family, and various genes in salt-stress responses that lead to environmental protection.[4,5] Under conditions of high salinity, these factors may affect plant productivity via three main negative effects, including water deficit, ionic (ie, Na^+, K^+, H^+, and Ca^{2+}) toxicity, and other nutrient imbalances.[6]

Hence, an excellent potato variety with resistance or tolerance to salt stress is required for the steady development of the potato industry. Current work on plant salt tolerance aims to identify the putative genes involved in the salt-tolerance response through a transcriptomic approach, followed by a targeted metabolite analysis. Gene expression profiling using microarrays of potatoes under several abiotic stresses, and transcriptomic approaches followed by targeted metabolite analysis, have been performed to investigate the pathways involved in salinity tolerance, as well as in carbohydrate and amino acid metabolism.[7,8] Molecular tools, such as genome-wide mapping, microarrays, transcriptome profiling, and proteomic techniques, have been used to generate salt-tolerant cultivars of different crops.[9] A recent study performed an analysis of the potato transcriptome under abiotic stress conditions using microarrays.[10] More recent work performed expression profiling of the potato genome under abiotic stress using Illumina ribonucleic acid (RNA)-seq transcriptome data.[11]

With respect to the relationship of RNA-seq and microarrays, Pearson's and Spearman's correlation coefficients of both datasets each exceeded 0.80, highlighting a 66%~68% overlap of genes;[12] RNA-seq provides better estimates of absolute transcript levels using standard microarray and RNA-seq protocols.[13] Using the combination of microarray and RNA-seq methods, disease-related genes were efficiently identified.[14]

Many studies have elucidated the mechanisms involved in salt stress using salt-stress-resistant mutants[15] and transformed potatoes.[16] These studies have provided specific information about salt tolerance. However, salt tolerance is a complex trait that involves numerous genes and products that function in salinity-resistant signaling pathways. Thus, it is necessary to apply high-throughput technologies to identify the genes involved in salt-responsive biosynthesis.[17,18]

In this study, we identified candidate genes for the regulation of salt-tolerance pathways in potatoes using a multistep screening method with the 56 K potato microarray and Illumina RNA-seq transcriptome data for potatoes. Using reverse transcription polymerase chain reaction (RT-PCR) and transgenic potatoes, the identified genes were examined for their effects on growth in a high-salinity environment.

Materials and Methods

We performed a three-step method to identify the potato genes involved in salt-tolerance pathways (Fig. 1). In step 1, we compared gene expression levels among three potato lines that displayed high, intermediate, and low salt tolerance by performing microarray and RNA-seq analyses on plants under conditions of salt stress. In step 2, we identified the candidate genes by combining the orthologous and biological pathways. A phylogenetic analysis and Circos diagram were calculated using the candidate genes. In step 3, we verified and validated the final selected genes by performing RT-PCR and generating transgenic *dehydration-responsive element-binding 1* (*DREB1*)-overexpressing potato lines.

Potato plants and salt treatments. We screened 50 transgenic potato (*S. tuberosum L.*) lines that were previously generated for drought resistance and salt tolerance. These potato lines were transformed in the Sumi potato cultivar using the AtbZIP17 transcription factor, which was activated in response to salt stress in *Arabidopsis*.[19] During the first step, we analyzed the salt tolerance phenotypes of the mutant lines. Plants were grown in a greenhouse with standard cultivation, and they were watered every 3 days with 150 mM of an NaCl solution. We selected two mutant lines designated as *SU-109* and *SU-14*, which showed higher and lower salt tolerance, respectively, compared to that of the control wild-type Sumi cultivar. These three potato lines shared a similar genetic background, but they exhibited different salt-tolerance phenotypes. Therefore, these lines were utilized for further studies to compare gene expression in response to salt-stress treatment.

Treatment and RNA extraction. The two mutant lines and the Sumi cultivar were grown for 2 months in individual pots in a greenhouse, and salt treatment was conducted using the protocol defined by Evers et al.[11] To extract RNA, we cut the youngest leaves on each plant at 0 hours, 6 hours, and 12 hours after salt treatment, froze the harvested leaf samples in liquid nitrogen, ground them to a powder with a mortar and pestle, and stored the samples in liquid nitrogen until further use. Total RNA was extracted from the leaf tissues using the RNeasy plant mini-kit (Qiagen, Inc., Valencia, CA, USA), and quantified using a Nanodrop ND-1000spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA).

Microarray experiments. We used 56 K potato microarray to perform a total of nine experiments (without replication) that compared gene expression among the two mutant lines and the control cultivar at three time points. The potato 56 K microarray was manufactured by Roche NimbleGen.[20] The microarray was designed from 43,553 coding sequences of PGSC version 2.1.10 pseudomolecules, which were based on version 3 of the genome assembly (Potato Genome Sequencing Consortium Public Data Release).[21] Genes reflecting alternative splice sites were designed to represent the exons. In total, the 56 K microarray was designed to include 56,647 predicted genes with 125,924 probes. We scanned the microarray for Cy3 signals with the Genepix 4000B Scanner (Molecular Instruments, LLC, Sunnyvale, CA, USA), and digitized the signals using Nimblescan (Roche NimbleGen, Inc., Madison, WI, USA).

RNA-seq transcriptome experiments. To compare the microarray experiments, we performed nine RNA-seq experiments using different potato leaf samples that were harvested

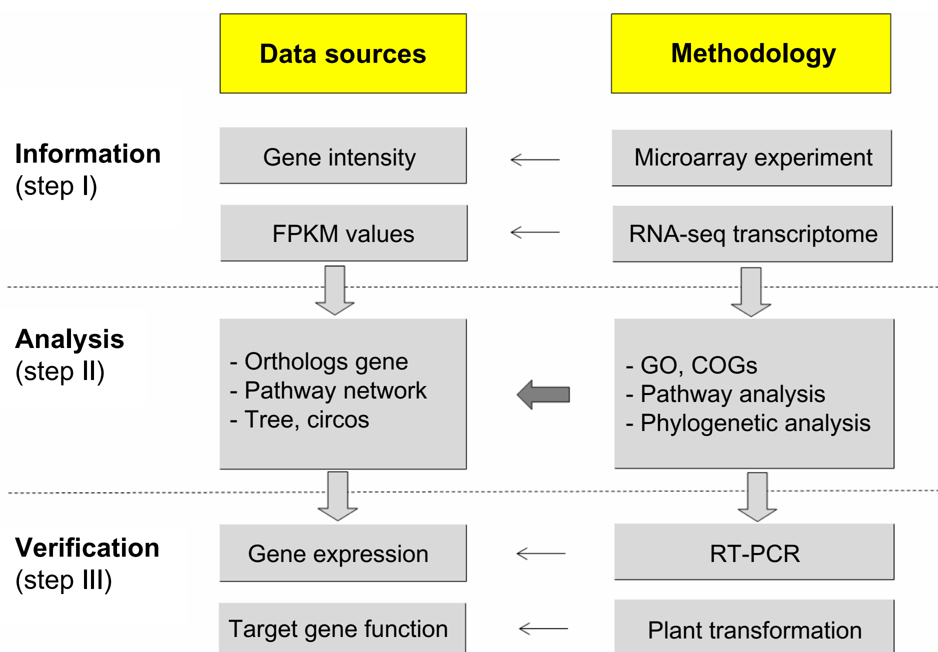

Figure 1. Flowchart of the multistep screening strategy used for this study.

under the same conditions as the microarray experiment. A comparison of the gene expression data generated by the microarray and RNA-seq experiments can increase the confidence in the experimental results. In the RNA-seq experiments, total RNA was isolated from the plant tissues using the GeneAll HybridR$^+$ kit (GeneAll Biotechnology Co., Ltd., Seoul, Korea), and messenger RNA was purified using the Illumina TruSeq RNA sample preparation kit (Illumina, Inc., San Diego, CA, USA) following the manufacturer's instructions. Library construction was performed using the general protocol,[22] and sequencing and assembly were performed using the Illumina Hi-Seq 1000 platform (Hayward, CA, USA). To perform quality-control checks on the raw sequence data, we used the FastQC program.[23] Useful transcripts were predicted using CLC Assembly Cell 3.2 (CLC Bio, Aarhus, Denmark) and the Trinity software package.[24] The parameters of the assembly process used default values, with the exception of essential options (ie, minimum length of contig, and length of paired fragment). Trinity is a novel method that is used for the efficient reconstruction of the transcriptome from RNA-seq data.[25] Transcript expression was tested using the Bowtie mapping program[26] with a false discovery rate of 5%, as described previously.[27]

Multiple-bioinformatics analysis. We performed a gene ontology enrichment analysis with GoMiner (National Cancer Institute, Bethesda, MD, USA).[28] This analysis first categorizes each gene according to its gene ontology (GO) function and mode of expression, and it then attempts to identify false discoveries with one-sided Fisher's exact tests ($P < 0.05$). To identify a network of enriched categories, we conducted a hyper-geometric test that included the calculation of the false discovery rate.[29] In the ortholog analysis, we classified the genes identified by comparing them to those in the eukaryotic National Center for Biotechnology Information (NCBI)/Clusters of Orthologous Groups (COGs) database.[30] We classified the differentially expressed potato genes into functional categories based on all-against-all Basic Local Alignment Search Tool (BLAST) searches of complete proteomes with an e-value cutoff of ≤1.0 e-20 using the NCBI/COGs database.

To identify the interaction network pathways for salt tolerance, we utilized the gene expression data that represented potential interactions between expressed genes and the salt-tolerance pathways. We used MedScan Reader™ 5.0 (Ariadne Genomics, Inc., Rockville, MD, USA) to extract information from the published literature about the known genes and regulatory interactions involved in salt-tolerance response pathways. We performed an enriched pathways analysis and Fisher's exact test to determine the most significant network responses using Pathway Studio software 9.0 (Ariadne Genomics, Inc.). To identify significant genes, we used a threshold with an absolute value of two, and calculated the *P*-values ($P \leq 0.05$ for statistical significance) with Fisher's exact tests using the Pathway Studio software.

For the phylogenetic analysis, amino acid sequences of the candidate genes involved in salt-tolerance pathways were aligned with the ClustalW method using the slow-accurate options in DNASTAR Lasergene® version 8.1 (DNASTAR, Inc., Madison, WI, USA).[31] The aligned sequences were trimmed at both ends to eliminate regions of poor alignment. The phylogenetic trees were constructed using the maximum likelihood algorithm in Molecular Evolutionary Genetics

Analysis version 5 (MEGA5).[32] The phylogeny of tree nodes was tested using the bootstrap method with 1,000 replicates. The Circos software was used to identify gene similarities from comparisons of genomes[33] and relationships among homologous genes in the genomes.[34,35] The Circos diagrams were constructed via a similarity index using Circos software version 0.64 (Circos, Vancouver, BC, Canada).[36] The similarity index was calculated using the mega-alignment function of the DNASTAR software, as follows:

$$\text{Similarity index} = 100[\text{Match}/(\text{Match} + \text{Mismatch} + \text{lenGaps} + \text{Gaps})]. \quad (1)$$

Reverse transcription-polymerase chain reaction analysis (RT-PCR). The RT-PCR reactions used 5 µg of RNA from each sample to synthesize complementary deoxyribonucleic acid (DNA) (cDNA) with the RNA to cDNA EcoDry Premix Oligo dT kit (Takara Bio, Inc., Otsu, Shiga, Japan). Then, we performed PCR using the synthesized cDNA and 10 pmol of each primer under the following conditions: 94 °C for 5 minutes; 25 cycles of 94 °C for 30 seconds, 56 °C for 1 minute; 72 °C for 1 minute; and a 10-minute extension at 72 °C. After the standard PCR of 35 cycles, the resulting PCR products (5 µL each) were analyzed using electrophoresis and ethidium bromide staining. To measure the relative expression level of each target gene, reference RNA (*Actin*) is used as a normalized template amount in each PCR reaction (Table S1).

Production of the transgenic potato. To perform validation of the selected *DREB*1 gene, first-strand cDNA for amplification was prepared using the Sprint™ RT Complete-Oligo(dT) kit according to the manufacturer's instructions (Clontech Laboratories, Inc.; Takara Bio, Inc.). The plasmid for the *DREB1* gene construct used the pCaMV35S promoter.[37] The pCaMV35S::*DREB1* binary vector structure is shown in Figure 2. We transformed a Sumi potato cultivar using the previously published protocol for the *Agrobacterium*-mediated transformation, selection, and regeneration of soybean.[38] The transgenic potato lines were tested for the expression of *DREB1* using RT-PCR. The growth phenotypes of the transgenic potato lines with high levels of *DREB1* expression were tested under two different environments, including treatment with 250 mM NaCl for 2 weeks using in vitro culture,[39] and watering with 150 mM of NaCl every 3 days for 3 weeks in a greenhouse.

Results and Discussion

Gene expression analysis using microarrays and RNA-seq experiments (step 1). We performed a total of nine microarray experiments that analyzed three potato lines at three time points of salt treatment (0 hours, 6 hours, and 12 hours). The potato lines were selected from a salt-tolerance screen of 50 mutant lines generated from the Sumi cultivar. The *SU-109* mutant displayed higher salt tolerance than that of the Sumi cultivar, whereas the *SU-14* mutant displayed lower salt tolerance than that of the Sumi cultivar. We examined the signal intensity values for a total of 56,647 unigenes from the potato microarrays, and compared the relative expression levels in the three cultivars. We screened 6,820 genes to identify those with at least a 2.0-fold difference in expression level among the 56,647 unigenes. In the RNA-seq experiments, we obtained a total of 320,078,806 reads with 48 Gbp of nucleotides using Illumina/Solexa pair-end sequencing (Table S2). In order to compare the expression profiling among samples without a point of reference, we assembled a dataset that merged all RNA-seq experiments sequence reads. In addition, we verified the expression difference, upon which each sample is mapped, using total assembled transcripts (contigs). We assembled 195,372 transcripts using a dataset that was merged with nine RNA-seq experiment sequence reads; following that, transcript expression was tested. The mapping percentage of the total reads is shown in Table S2.

We obtained 7,530 different genes ($P \leq 0.05$) that were identified using the fragments per kilobase of exon per million fragments mapped values in each treatment. Comparisons between the genes that were identified using RNA-seq and those identified using microarrays resulted in the identification of 5,416 common genes. The RNA-seq data validated approximately 72% of the regulated genes identified by the microarray gene expression data. Thus, our gene expression analysis screening procedure selected a total of 5,416 candidate genes for further analysis of potential involvement in salt-tolerance biosynthesis and/or metabolism in potato.

Gene identification using bioinformatics analyses (step 2). *Ortholog and ontology analyses.* To identify conserved orthologous genes, we performed GO and COG analyses using the 5,416 candidate genes. The GO enrichment analysis screened 2,962 genes with one-sided Fisher's exact test and the hypergeometric test. We screened the genes that were likely related to salt tolerance functions based on the presence of salinity-related categories; we then identified the poorly

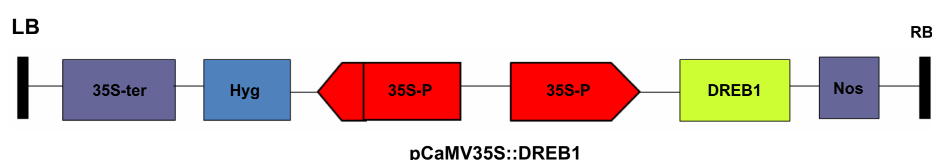

Figure 2. Schematic diagram of the pCaMV35S::DREB1 binary vector. The region shows the construct containing the pCaMV35S promoter and the *DREB1* coding region between the left border (LB) and right border (RB).

characterized group as possibly exhibiting abiotic stress-related functions, because these 5,416 candidate genes were assumed to include a variety of function categories.

For the COG analysis, we placed a total of 1,528 genes into functional categories, which were used to base all-against-all BLAST searches for each of the three experimental time points (0 hours, 6 hours, and 12 hours), and 133 COG genes (8.7%) were placed into multiple categories. We performed GO and COG analyses and screened a total of 1,273 orthologous genes to compare the ortholog similarities and to categorize the conserved genes.

Pathway analysis. To identify genes related to specific salt-tolerance pathways among the 5,416 differentially expressed genes, we performed an enriched-pathways analysis using the Pathway Studio software. First, we used published evidence in the literature to acquire a total of 653 genes that were candidates for involvement in salt-tolerance pathways. Second, we used Fisher's exact test to determine the most significant network responses, and found 42 interconnected genes that were linked to specific salt-tolerance pathways. Finally, we used these genes to construct an interconnected network. The constructed network of salt-tolerance pathways revealed the two main processes characteristically involved in salinity responses and salt metabolism, including water channels/water transport and cell death (Fig. 3).

The Na^+/H^+ antiporters are commonly present as integral membrane proteins in the plasma membrane. The Na^+ and K^+ transporters can direct movement out of the cell, or luminal movement into the vacuoles and intracellular organelles.[40] This activity contributes to control the pH value, while maintaining ion homeostasis in the cells.[5] The functional class network (Fig. 3, depicted with orange octagons) shared by NaCl pathways and salinity-response pathways showed that NaCl pathways contain the H^+-transporting adenosine triphosphatase and the Na^+/H^+ antiporter, whereas salinity-response pathways contain the Na^+ transporter.

Expression of the MYB-type transcription factors in salinity-related processes or salt-stress pathways were reported, in which R2R3-type *MYB* genes play a diverse role in plant development and response to salt stress in potato.[41] These results illustrate the functional diversity of proteins that may be activated during salt-tolerance responses.

Selection of candidate genes involved in salt tolerance. To screen the final candidate genes, we compared the 653 pathway–network genes and the 1,273 ortholog–ontology genes. We found nine potential candidate genes that were

Figure 3. Pathway diagram of the differentially regulated genes identified in the screening process for salt-tolerance responses in high, intermediate, and low salt tolerance potato lines. The pathway networks of genes involved in salinity responses reveal the biological interactions among different networks.

differentially expressed between the high, intermediate, and low salinity-tolerance potato lines at two time points (6 hours and 12 hours) with respect to the baseline levels observed at the 0-hour time point (before treatment). The nine selected genes encoded the following putative proteins: plasma membrane intrinsic protein 1; osmotin-like protein; aquaporin-like protein; dehydration-responsive family protein; thaumatin-like protein; *DREB1*; and three unknown proteins. The potato ST_23752 gene (the *Arabidopsis thaliana* homolog is AT4G11650) is well-known in the salt-stress pathway as an osmotin-like protein, and ST_11949 (the *A. thaliana* homolog is AT4G25480) is a *DREB1* transcription factor that responds to low-temperature and abiotic stress according to the *Arabidopsis* Information Resource.[42] The potato ST_03756 gene (the *A. thaliana* homolog is AT4G00430) is linked to water channel activity (GO:0015250) according to the European Bioinformatics Institute.[43] However, the other genes identified by our multistep screen were not well-known in salt-stress or salinity-response pathways. This may be because many of the previously identified genes are not expressed continuously during the treatment times that we examined (up to 12 hours after the start of salt stress), but they are strongly expressed at only one time point.

Phylogenetic analysis with the Circos diagram. A phylogenetic analysis was performed to identify relationships among the nine selected genes and the 42 known genes involved in the salinity-response pathways. To reveal the phylogenic relationships using the well-known genes, we detected 31 homologous genes in the *A. thaliana* genome using 42 sequences for genes involved in salinity-related pathways. Using the nine selected genes and the 31 homologous genes, phylogenetic analysis was performed based on the maximum likelihood method.

The phylogenetic tree indicates that the genes fall into one of four subgroups, which are designated as group 1, group 2, group 3, and group 4 (Fig. 4). A composite tree indicates that there are homologous relationships between the genes of the potato and *A. thaliana*, and all selected candidate genes were clustered into one of the four subgroups. Group 2 contains one candidate gene that was verified by our RT-PCR analysis among the nine selected genes, and group 3 contains three candidate genes.

The Circos diagram shows the chromosomal positions of the well-known *A. thaliana* and potato genes that were

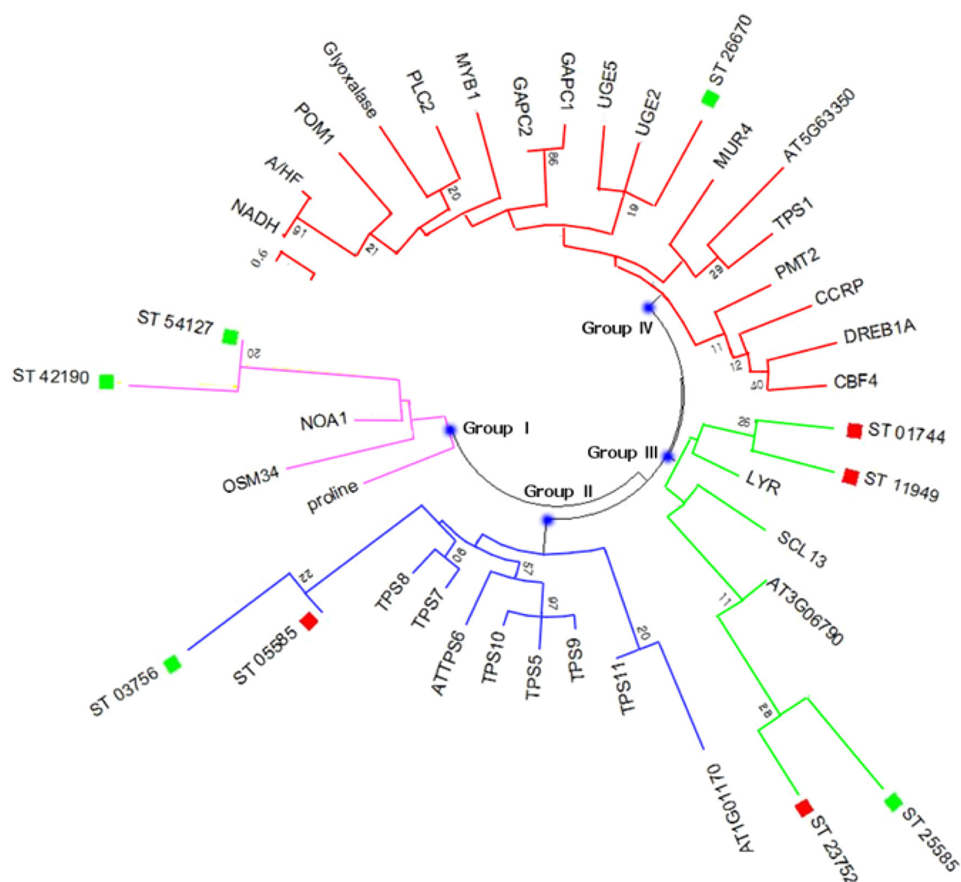

Figure 4. Phylogenetic relationships among nine selected genes identified in potato and homologous genes expressed in *Arabidopsis thaliana*. Four of the selected potato genes were verified by RT-PCR (marked with a red square), and the other five selected potato genes were generated using the maximum likelihood method of the MEGA5 program (marked with a green square). The phylogenetic tree was generated with 1,000 bootstrap repetitions; bootstrap values (greater than 50%) are given at the nodes.

identified by our analysis (Fig. 5). To facilitate the interpretation of the Circos diagram, we presented the pairwise chromosomal relationships among the 31 homologous genes and the nine potato genes on the chromosomes of *A. thaliana* (At1–At5) and the identified groups of potato. The potato groups contained the nine identified candidate genes including *DREB1*. The *DREB1* gene partly shows the shared ancestral origin of the *Arabidopsis* and potato genomes as CBF4/AP2/GAPC2. The *trehalose-phosphate synthase* 5 (*TPS5*) gene on At4 represents TPS7/8/10/11. Taken together, these results suggest that the nine selected genes do not share extensive similarities with the previously characterized and well-known homologous genes involved in the salinity-related pathways.

Gene verification and validation using RT-PCR and transgenic potato (step 3). To verify the nine candidate genes identified by our analyses, we performed RT-PCR on the RNA samples using different potato leaf samples that were harvested under the same conditions as the microarray and RNA-seq experiments at five different time points (0 hours, 3 hours, 6 hours, 12 hours, and 24 hours). Four genes were upregulated in the *SU-109* mutant (high salinity tolerance), whereas they were downregulated in the *SU-14* mutant (low salinity tolerance) when compared to the levels of expression in cultivar Sumi (intermediate salinity tolerance) during all of the treatment stages tested (Fig. 6). We assume that these four selected genes are specifically involved in salt-tolerance responses because the *SU*-109 mutant has the same genetic background and exhibits greater salt tolerance than the control cultivar Sumi.

We selected the verified *DREB1* gene that was isolated by our screening method for further validation by using it to generate transgenic potato lines. *DREB1* was previously reported to play an important role in increasing stress tolerance in *Arabidopsis*,[44] tobacco,[45] wheat and barley,[46] rice,[47] and potato.[48] To validate the function of *DREB1* identified by our screening method, we overexpressed *DREB1* driven by the CaMV35S promoter in the Sumi cultivar, and analyzed ten transgenic lines for the expression of *DREB1* and the response to salt stress. We selected five mutant lines (*M-18*, *M-19*, *M-21*, *M-28*, and *M-60*) that displayed high levels of *DREB1* expression based on RT-PCR (Fig. 7). Then, we monitored growth under salt-stress conditions using the five *DREB1* overexpression lines and the wild-type Sumi cultivar. Differences were observed in the growth

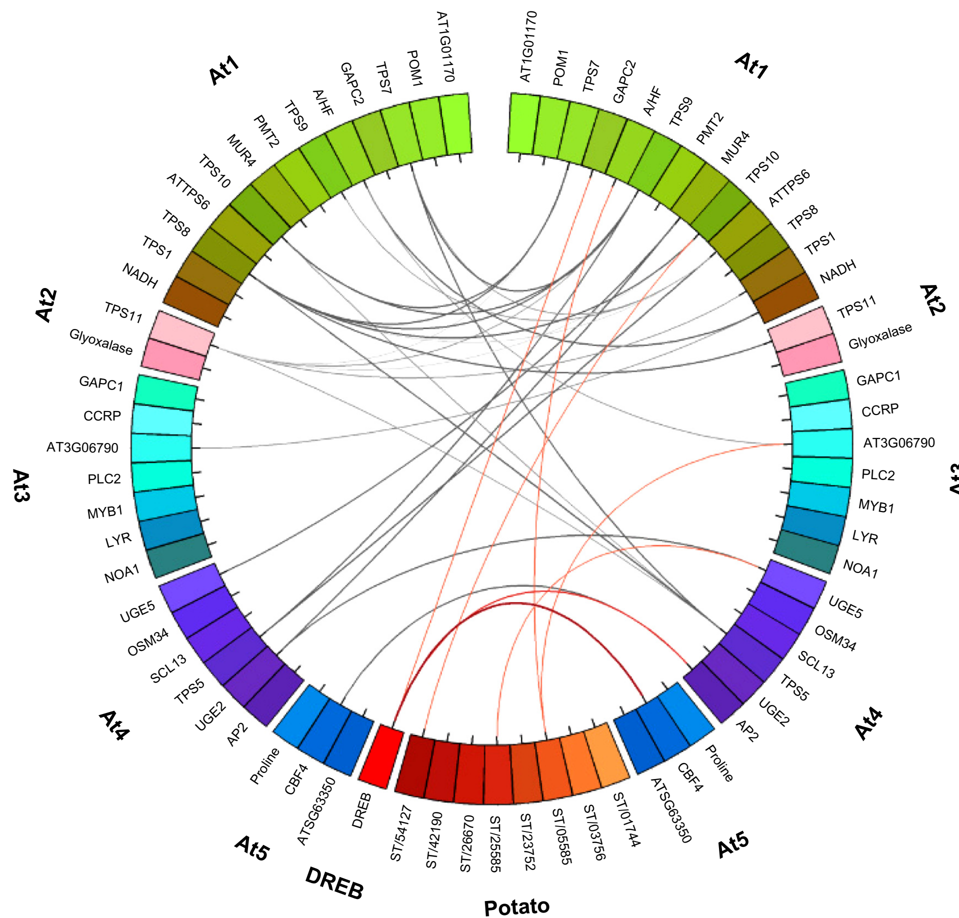

Figure 5. Circos diagram of gene expression involved in salinity-response pathways. The homologous genes in each group are plotted against their pairwise counterparts. Individual chromosomes of *Arabidopsis thaliana* (At1–At5) and potato groups (Potato/DREB) are shown as a similarity index to represent the shared ancestral origin of the genomes. Red arrow, selected gene of potato; black arrow, homologous gene of *Arabidopsis thaliana*.

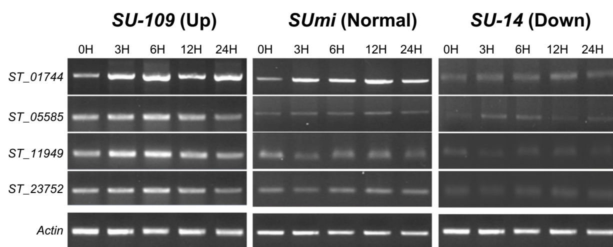

Figure 6. RT-PCR analysis of the genes involved in the salt-tolerance pathways identified by the multistep screening strategy. Expression of the five candidate genes was assessed in high-expression mutants (*SU-109*, high salt tolerance) and low-expression mutants (*SU-14*, low salt tolerance) during five salt-treatment time points, as well as in the control cultivar Sumi. The 0H, 3H, 6H, 12H, and 24H lanes represent hours after salt treatment.

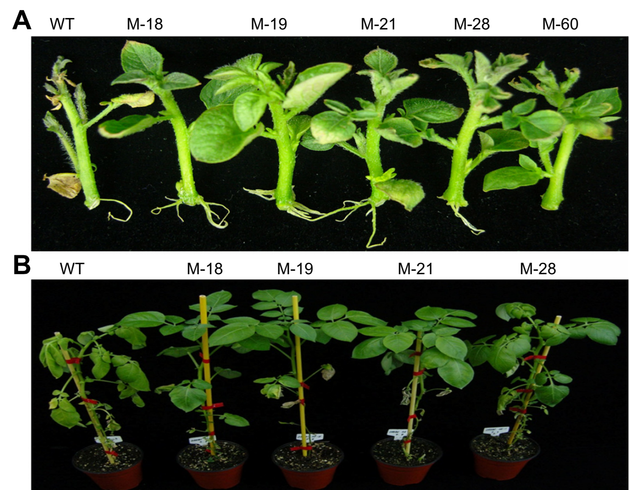

Figure 8. Growth analysis under salt stress of five transgenic potato lines (*M-18*, *M-19*, *M-21*, *M-28*, and *M-60*) and a wild-type control (Sumi cultivar). A. Plants were cultured in vitro in 250 mM of NaCl for 2 weeks. B. Plants were grown in a greenhouse and watered with 150 mM of NaCl every 3 days for 3 weeks.

between the transgenic overexpression mutants and the wild-type plant. Plants were exposed to 250 mM of NaCl for 2 weeks during in vitro culture (Fig. 8A), or 150 mM of NaCl for 3 weeks in a greenhouse (Fig. 8B). The results showed that all transgenic *DREB1* overexpression lines had greater salinity tolerance and increased growth compared to that of wild-type plants for both salt-stress treatment conditions. Thus, transgenic potato lines overexpressing *DREB1* were successfully generated using *Agrobacterium*-mediated transformation. This result validates our screening method that was used to identify the genes involved in salt-tolerance pathways in potato.

This study developed a novel, multistep screen that successfully identified candidate genes that may be involved in salinity-tolerance pathways in potatoes. We generated transgenic lines overexpressing *DREB1*, one of the verified genes isolated by our screen. The transgenic *DREB1*-overexpression potato lines exhibited increased salt tolerance compared to that of wild-type controls, and they could possibly lead to tolerance to multiple stresses in potato plants. Therefore, we registered the potato *DREB1* gene in NCBI/GenBank (HM641796.1).[49] We compared our isolated *DREB1* gene and *StDREB1* gene (NCBI/GenBank: JN125862.1), which increases tolerance to salt in potato plants, as reported by Bouaziz et al.[48] Two genes (*HM641796.1* and *JN125862.1*) have sequence lengths of 960 and 657, individually. In proteins, sequence alignments show the different degrees of similarity between the amino acids occupying a particular position; however, the domain analysis showed results where two genes have been assumed to be part of the same family of genes at the protein level. Therefore, we have collected the family genes in the rice and *Arabidopsis* genome database, and we performed the BLAST search with selected family genes. The result indicated the different structured genes between two genes, although the two genes have the similarity relationship at the protein level. (Table S3).

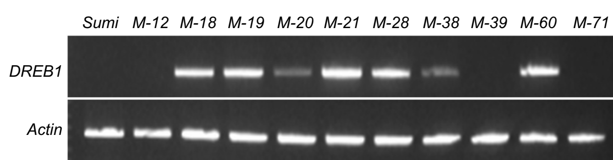

Figure 7. RT-PCR analysis of ten transgenic potato lines for overexpression studies of the potato *DREB1* gene. Potato *actin* was used as a control. WT, Sumi cultivar; *M-12, M-18*, *M-19*, *M-20*, *M-21*, *M-28*, *M-38*, *M-39*, *M-60*, *M-71*, transgenic *DREB1*-overexpression lines.

Four of the genes isolated by our screen were verified by RT-PCR and genetic transformation methods, and they were involved in specific salinity-responsive pathways. Five genes were not verified by RT-PCR. We assume that these genes either play regulatory roles in salinity resistance, or they are involved in salt-stress signaling during metabolite biosynthesis. Although there is no direct evidence linking the potato *DREB1* gene function to salt tolerance, our results with transgenic *DREB1* overexpression lines suggest that this gene plays a role in the regulation of salt tolerance. The genes identified in our screen require additional validation to further confirm our microarray and RNA-seq experimental strategy. Our study demonstrates the potential of multistep bioinformatics screening combined with microarray and RNA-seq transcriptome data to isolate genes involved in complex biosynthetic networks and pathways.

Author Contributions

CKK conceived and designed the experiments. JKN, JWC, SCP, YHK, and YKK analyzed the data. CKK wrote the first draft of the manuscript. DYK, SHS, and HML contributed to the writing of the manuscript and jointly developed the structure and arguments for the paper. The final manuscript was reviewed and approved by all authors.

DISCLOSURES AND ETHICS

As a requirement of publication the authors have provided signed confirmation of their compliance with ethical and legal obligations including but not limited to compliance with ICMJE authorship and competing interests guidelines, that the article is neither under consideration for publication nor published elsewhere, of their compliance with legal and ethical guidelines concerning human and animal research participants (if applicable), and that permission has been obtained for reproduction of any copyrighted material. This article was subject to blind, independent, expert peer review. The reviewers reported no competing interests.

Supplementary Materials

Supplementary Table 1. Primer sequences used for RT-PCR analysis to verify the four candidate potato genes.

Supplementary Table 2. Transcriptome assembly of each potato line in the potato genome.

Supplementary Table 3. Similarity comparison between the *HM641796* and *JN125862* genes.

REFERENCES

1. Kawasaki S, Borchert C, Deyholos M, et al. Gene expression profiles during the initial phase of salt stress in rice. *Plant Cell*. 2001;13(4):889–905.
2. Zhu JK. Salt and drought stress signal transduction in plants. *Annu Rev Plant Biol*. 2002;53:247–73.
3. Katerji N, van Hoorn JW, Hamdy A, Mastrorilli M. Salt tolerance classification of crops according to soil salinity and to water stress day index. *Agricultural Water Management*. 2000;43(1):99–109.
4. Xiong L, Schumaker KS, Zhu JK. Cell signaling during cold, drought, and salt stress. *Plant Cell*. 2002;14 Suppl:S165–83.
5. Yamaguchi T, Blumwald E. Developing salt-tolerant crop plants: challenges and opportunities. *Trends Plant Sci*. 2005;10(12):615–20.
6. Hasegawa PM, Bressan RA, Zhu JK, Bohnert HJ. Plant cellular and molecular responses to high salinity. *Annu Rev Plant Physiol Plant Mol Biol*. 2000;51:463–99.
7. Rensink WA, Iobst S, Hart A, Stegalkina S, Liu J, Buell CR. Gene expression profiling of potato responses to cold, heat, and salt stress. *Funct Integr Genomics*. 2005;5(4):201–7.
8. Legay S, Lamoureux D, Hausman JF, Hoffmann L, Evers D. Monitoring gene expression of potato under salinity using cDNA microarrays. *Plant Cell Rep*. 2009;28(12):1799–816.
9. Abbasi FM, Komatsu S. A proteomic approach to analyze salt-responsive proteins in rice leaf sheath. *Proteomics*. 2004;4(7):2072–81.
10. Kondrák M, Marincs F, Kalapos B, Juhász Z, Bánfalvi Z. Transcriptome analysis of potato leaves expressing the trehalose-6-phosphate synthase 1 gene of yeast. *PLoS One*. 2011;6(8):e23466.
11. Singh AK, Sharma V, Pal AK, Acharya V, Ahuja PS. Genome-wide organization and expression profiling of the NAC transcription factor family in potato (Solanum tuberosum L.). *DNA Res*. 2013;20(4):403–23.
12. Xu X, Zhang Y, Williams J, et al. Parallel comparison of Illumina RNA-Seq and Affymetrix microarray platforms on transcriptomic profiles generated from 5-aza-deoxy-cytidine treated HT-29 colon cancer cells and simulated datasets. *BMC Bioinformatics*. 2013;14 Suppl 9:S1.
13. Fu X, Fu N, Guo S, et al. Estimating accuracy of RNA-Seq and microarrays with proteomics. *BMC Genomics*. 2009;10:161.
14. Mooney M, Bond J, Monks N, et al. Comparative RNA-Seq and microarray analysis of gene expression changes in B-cell lymphomas of Canis familiaris. *PLoS One*. 2013;8(4):e61088.
15. Munns R, Tester M. Mechanisms of salinity tolerance. *Annu Rev Plant Biol*. 2008;59:651–81.
16. Fevziye CT, Behnam B, Serrano G, et al. Tolerance to salt stress of the transgenic tetrasomic tetraploid potato, Solanum tuberosum cv. Desiree appears to be induced by the DREB1 A gene and rd29A promoter of Arabidopsis thaliana. *Breed Sci*. 2005;55:311–20.
17. Seki M, Narusaka M, Ishida J, et al. Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J*. 2002;31(3):279–92.
18. Evers D, Legay S, Lamoureux D, Hausman JF, Hoffmann L, Renaut J. Towards a synthetic view of potato cold and salt stress response by transcriptomic and proteomic analyses. *Plant Mol Biol*. 2012;78(4–5):503–14.
19. Liu JX, Srivastava R, Howell SH. Stress-induced expression of an activated form of AtbZIP17 provides protection from salt stress in Arabidopsis. *Plant Cell Environ*. 2008;31(12):1735–43.
20. Roche NimbleGen, Inc. [homepage on the Internet]. Nimblegen. Madison, WI: NimbleGen, Inc.; 2014. Available from: http://www.nimblegen.com/. Accessed January 7, 2014.
21. MSU Potato Breeding and Genetics Program [webpage on the Internet]. Spud DB potato genomics resource. East Lansing, MI; 2013. Available from: http://solanaceae.plantbiology.msu.edu. Accessed December 5, 2013.
22. Mun JH, Yu HJ, Shin JY, Oh M, Hwang HJ, Chung H. Auxin response factor gene family in Brassica rapa: genomic organization, divergence, expression, and evolution. *Mol Genet Genomics*. 2012;287(10):765–84.
23. Babraham Bioinformatics [webpage on the Internet]. FastQC. Cambridge, UK; The Babraham Institute; 2012. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. Accessed November 12, 2013.
24. The Broad Institute; The Hebrew University of Jerusalem [webpage on the Internet]. RNA-seq de novo assembly using Trinity. Cambridge, MA: The Broad Institute; Jerusalem, Israel: The Hebrew University of Jerusalem; 2013. Available from: http://trinityrnaseq.sourceforge.net/. Accessed November 19, 2013.
25. Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29(7):644–52.
26. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25.
27. Bhargava A, Clabaugh I, To JP, et al. Identification of cytokinin-responsive genes using microarray meta-analysis and RNA-Seq in Arabidopsis. *Plant Physiol*. 2013;162(1):272–94.
28. Genomics and Bioinformatics Group [homepage on the Internet]. GoMiner. Bethesda, MD: Genomics and Bioinforamtics Group; 2011. Available from: http://discover.nci.nih.gov/gominer/. Accessed November 29, 2013.
29. Kim CK, Lee YH, Hong JK, et al. Identification and characterization of flowering repressor-related genes in Chinese cabbage. *BioChip Journal*. 2012;6(2):120–7.
30. National Center for Biotechnology Information [webpage on the Internet]. COGs: phylogenetic classification of proteins encoded in complete genomes. Bethesda, MD: 31. National Center for Biotechnology Information; 2003. Available from: http://www.ncbi.nlm.nih.gov/COG/. Accessed November 30, 2013.
31. DNASTAR, Inc. [webpage on the Internet]. DNASTAR. Madison, WI: DNASTAR; 2013. Available from: http://www.dnastar.com/. Accessed December 3, 2013.
32. Tamura K, Stecher G, Peterson D, Kumar S [webpage on the Internet]. MEGA: molecular evolutionary genetics analysis. 2014. Available from: http://www.megasoftware.net/. Accessed July 15, 2013.
33. Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19(9):1639–45.
34. Burkart-Waco D, Josefsson C, Dilkes B, et al. Hybrid incompatibility in Arabidopsis is determined by a multiple-locus genetic network. *Plant Physiol*. 2012;158(2):801–12.
35. Wang D, Xia Y, Li X, Hou L, Yu J. The Rice Genome Knowledgebase (RGKbase): an annotation database for rice comparative genomics and evolutionary biology. *Nucleic Acids Res*. 2013;41(Database issue):D1199–205.
36. Krzywinski M [homepage on the Internet]. Circos. Vancouver, BC: Canada's Michael Smith Genome Sciences Centre; 2014. Available from: http://circos.ca/. Accessed December 15, 2013.
37. Wachter A, Tunc-Ozdemir M, Grove BC, Green PJ, Shintani DK, Breaker RR. Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. *Plant Cell*. 2007;19(11):3437–50.
38. Polizel AM, Medri ME, Nakashima K, et al. Molecular, anatomical and physiological properties of a genetically modified soybean line transformed with rd29 A:AtDREB1 A for the improvement of drought tolerance. *Genet Mol Res*. 2011;10(4):3641–56.
39. Hussey G, Stacey NJ. Factors affecting the formation of in vitro tubers of potato (Solanum tuberosum L.). *Ann Bot*. 1984;53(4):565–78.
40. Bassil E, Tajima H, Liang YC, et al. The Arabidopsis Na+/H+ antiporters NHX1 and NHX2 control vacuolar pH and K+ homeostasis to regulate growth, flower development, and reproduction. *Plant Cell*. 2011;23(9):3482–97.
41. Yang A, Dai X, Zhang WH. A R2R3-type MYB gene, OsMYB2, is involved in salt, cold, and dehydration tolerance in rice. *J Exp Bot*. 2012;63(7):2541–56.
42. Tair [homepage on the Internet]. The Arabidopsis Information Resource. Stanford, CA: the Carnegie Institution for Science Department of Plant Biology; 2014. Available from: http://www.arabidopsis.org/. Accessed December 27, 2013.
43. European Molecular Biology Laboratory [homepage on the Internet]. Quick GO. Cambridgeshire, UK: EMBL-EBI, Wellcome Trust Genome Campus; 2014. Available from: http://www.ebi.ac.uk/QuickGO/. Accessed December 27, 2013.
44. Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K. DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. *Biochem Biophys Res Commun*. 2002;290(3):998–1009.

45. Akhtar M, Jaiswal A, Taj G, Jaiswal JP, Qureshi MI, Singh NK. DREB1/CBF transcription factors: their structure, function and role in abiotic stress tolerance in plants. *J Genet*. 2012;91(3):385–95.
46. Morran S, Eini O, Pyvovarenko T, et al. Improvement of stress tolerance of wheat and barley by modulation of expression of DREB/CBF factors. *Plant Biotechnol J*. 2011;9(2):230–49.
47. Ito Y, Katsura K, Maruyama K, et al. Functional analysis of rice DREB1/CBF-type transcription factors involved in cold-responsive gene expression in transgenic rice. *Plant Cell Physiol*. 2006;47(1):141–53.
48. Bouaziz D, Pirrello J, Charfeddine M, et al. Overexpression of StDREB1 transcription factor increases tolerance to salt in transgenic potato plants. *Mol Biotechnol*. 2013;54(3):803–17.
49. National Center for Biotechnology Information [webpage on the Internet]. Solanum tuberosum dehydration responsive element binding protein 1 (DREB1) mRNA, complete cds. Bethesda, MD: National Center for Biotechnology Information; 2010. Available from: http://www.ncbi.nlm.nih.gov/nuccore/HM641796. Accessed January 9, 2014.