Polar Biol (2010) 33:617–626
DOI 10.1007/s00300-009-0737-8

ORIGINAL PAPER

Complementary DNA library construction and expressed sequence tag analysis of an Arctic moss, *Aulacomnium turgidum*

Shenghao Liu · Hyoungseok Lee · Pil-Sung Kang · Xiaohang Huang · Joung Han Yim · Hong Kum Lee · Il-Chan Kim

Received: 7 January 2009 / Revised: 24 September 2009 / Accepted: 25 September 2009 / Published online: 21 October 2009

Abstract Unique physiological and metabolic properties of Arctic mosses are responsible for their acclimation to the inclement polar environment. To perform transcriptome analysis of an Arctic moss species adapted to polar conditions, we constructed a complementary DNA (cDNA) library using total high-quality RNA extracted from the moss species *Aulacomnium turgidum*. The library consisted of 1.81×10^6 of independent clones with 97.41% of recombinants. A total of 509 cDNA clones were sequenced. After eliminating poor quality sequences, vector trimming and clustering, 360 unigenes consisting of 33 contigs and 327 singletons were identified. Basic Local Alignment Search Tool X searches generated 245 significant hits (*E* value $<10^{-5}$). For further Gene Ontology analysis, 158 unigenes were annotated and classified with terms for molecular function, biological process and cellular component. Among the expressed sequence tags, seven genes were selected based on their putative roles in stress response, and they showed enhanced transcripts level under various abiotic stresses such as low temperature, heat and high-salinity. Also, two rare-cold-inducible genes showed different expression patterns under low temperature and UV-B treatment, indicating their distinct roles in adaptation to Arctic environment. Although experiments have been conducted on a limited scale, this study provides useful information for better understanding the mechanism of stress acclimation of polar mosses and material basis for potential genomic modification for higher plants to increase stress tolerance.

Keywords Arctic moss · *Aulacomnium turgidum* · Expressed sequence tags (ESTs) · Gene Ontology (GO)

S. Liu · H. Lee · P.-S. Kang · J. H. Yim · H. K. Lee · I.-C. Kim (✉)
Polar BioCenter, Korea Polar Research Institute (KOPRI),
KORDI, Incheon 406-840, Republic of Korea
e-mail: ickim@kopri.re.kr

S. Liu · X. Huang
Key Laboratory of Marine Bioactive Substances,
The First Institute of Oceanography,
State Oceanic Administration, 266061 Qingdao, China

P.-S. Kang
Department of Biological Engineering, Inha University,
Incheon 402-751, Republic of Korea

Introduction

The Arctic constitutes a unique and important environment with a significant role in the dynamics and evolution of the earth system (Birks et al. 2004). In the Arctic, plants are exposed to extreme amplitudes of environmental factors: early snow and frost, drought, wind abrasion and extreme temperatures on open patches (Billings and Mooney 1968). They developed a variety of strategies that range from physiological adaptations to tolerance to stressful environmental factors. Examples of adaptive mechanisms to low temperature include the evolution of cold shock and antifreeze proteins, the modulation of the kinetics of key enzymes, and the development of more fluid biological membranes through the accumulation of polyunsaturated fatty acyl chains (Morgan-Kiss et al. 2006).

Mosses are one of the most diverse and widespread groups of plants and often form the dominant vegetation in many Arctic tundra plant communities (Longton 1997; Ayres et al. 2006; Gornall et al. 2007). In contrast to a wide range of taxa already analyzed in the Antarctic (Skotnicki et al. 2000, 2005), few studies so far focused on the genetic diversity of *Sphagnum* species in the Arctic (Flatberg and

Thingsgaard 2003; Greilhuber et al. 2003). Mosses represent one of the oldest clades of land plants, separated by approximately 450 million years of evolution from crop plants. Consequently, mosses contain metabolites and genes not known from these seed plants (Reski and Frank 2005). Therefore, Arctic mosses may offer exceptional opportunities for gaining novel insights into the mechanisms of plant survival under extreme conditions and plant genetic evolution.

Large-scale random sequencing of expressed sequence tags (ESTs) provides an effective means of identifying expressed genes in organisms across all kingdoms. In plants, for example, a number of genes from rice have been reported to be induced by drought, high-salinity, and low-temperature stresses, and their products are thought to function in stress tolerance and adversity response (Rabbani et al. 2003). To date, over 300,000 ESTs from the model species *Physcomitrella patens* have been deposited in GenBank (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). However, to our knowledge, there is no report on identifying the stress-related genes from a complementary DNA (cDNA) library of polar mosses. In the present study, we constructed a cDNA library and then conducted an EST analysis for an Arctic moss, *Aulacomnium turgidum*. In addition, we identified several stress-related genes using annotation searches from the dataset.

Materials and methods

Plant materials and stress treatments

Aulacomnium turgidum specimens growing under natural conditions were collected in the vicinity of the Korean Dasan Arctic Station (78°54′N; 11°57′W) near Ny-Ålesund, Svalbard, in August 2006. They were placed in plastic containers and transported to Korea, where they were then cultivated on BCD solid media (Ashton and Cove 1977) in a growth room at 25°C with continuous light. For the cold or the heat treatments, plants were transferred from 25 to 4 or 37°C and incubated for 24 h. For the salt treatment, plants were transferred to BCD solid media containing 300 mM NaCl and incubated at 25°C for 24 h. UV-B treatments were performed in a dark box at a constant 15°C. UV-B irradiation was supplied by two Philips UV-B lamps (TL 20 W/01 RS). Photon fluxes were adjusted to 9 W/m^2 by varying the distance of the leaves from the light source. Fluence was measured using an digital ultraviolet radiometer model SM 6.0 UVB (Solartech Inc., USA). Some green parts of shoots and protonemata were ground in liquid nitrogen with mortar and pestle and then the ground powder was transferred into 50 mL conical tubes and stored at −80°C.

Total and poly A$^+$ RNA preparation

Total RNA was extracted from 3 to 4 g of ground tissue using cetyltrimethylammonium bromide (CTAB) extraction buffer (Fu et al. 2004) composed of 2% CTAB, 1% polyvinylpyrrolidone (PVP) K-30 (soluble), 100 mM Tris HCl (pH 8.0), 20 mM EDTA, 1.4 M NaCl, 0.5 g/L spermidine (free acid), 2% β-mercaptoethanol (added just before use). Solutions were treated with diethylpyrocarbonate and autoclaved (Sambrook et al. 1989). Tris–HCl (pH 8.0), prepared with DEPC-treated water, was added to the appropriate solutions post-autoclaving.

After adding 0.5 mL CTAB buffer (2% β-ME) to 0.5 mL powder, the samples were incubated at 60°C for 10 min, mixed well by vortexing. The samples were further extracted twice with an equal volume of chloroform:isoamyl alcohol (24:1) centrifuged at 13,000 rpm for 10 min at 4°C and the aqueous phase was transferred to a new tube. To ensure obtaining a high-quality RNA, the interphase was carefully avoided being transferred. The gained supernatant was precipitated with 1/4 volume of 10.0 M LiCl to each tube and mixed well by inverting. The RNA was precipitated overnight at 4°C and harvested by centrifugation at 13,000 rpm for 30 min at 4°C. Pellet was washed with 75% ethanol twice and air-dried for 5 min. RNA was dissolved in 100 μL DEPC-treated water and RNA from the same samples were pooled together into new 1.5-mL RNAse-free microfuge tubes.

Poly A$^+$ RNA was isolated using an Oligotex Mini Kit (Qiagen, Los Angeles, CA). The purified total RNA and poly A$^+$ RNA were quantified with a ND-1000 spectrophotometer (NanoDrop Technologies Inc.) at wavelengths of 230, 260 and 280 nm. The integrity of total RNA and poly A$^+$ RNA was verified by running samples on 1.2% denaturing agarose gels.

Construction of cDNA library

First-strand cDNA was synthesized from approximately 500 ng of total RNA using the SMART cDNA Library Construction Kit (DB Clontech, USA) according to the manufacturer's protocol. Double stranded cDNA was synthesized by employing long-distance PCR (LD-PCR) from 2 μL of poly A$^+$ RNA. After size fractioning, cDNA was cloned into the bacteriophage expression vector λTriplEx2 and packaged into phage particles using ZAP-cDNA Gigapack III Gold Cloning Kit (Stratagene, USA) to construct the primary cDNA library. Phages were used to transfect XL1-Blue to amplify the primary cDNA library.

Positive clones were checked by PCR with λ TriplEx LD-insert screening primers. To increase the stability of cDNA library, primary library was amplified as described in the manufacturer's instructions.

Sequence processing

Clones for sequencing were selected randomly from the cDNA library. Each clone was cultured and the plasmid DNA was extracted with an AccuPrep Plasmid Extraction Kit (Bioneer, Korea). The sequencing reaction with pTriplEx2 5′ sequencing primer was performed on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, USA).

Raw sequences were first individually processed using Phred software (CodonCode Aligner) which linked base calling, eliminating sequences shorter than 100 bp, low-quality sequences and vector trimming. The overlapping sequences (from both 3′ and 5′ ends) were evaluated and aligned into full consensus sequence contigs using DNA analyzing software Lasergene 7.0 (DNAStar, Madison, WI, USA). The non-overlapping sequences were formatted and treated as singlet sequences.

Gene Ontology (GO) assignments

For functional classification according to GO assignments, all unigene sequences were queried against the National Center for Biotechnology Information (NCBI) non-redundant (nr) protein database using BLASTX algorithm with the E value threshold set to $\leq 10^{-5}$. Functional categories were defined using Gene Ontology Database (http://www.geneontology.org) and BLAST algorithm with default settings. According to BLAST results, the top 158 BLAST hits were selected for further analysis. Gene symbols associated with the best scoring BLAST hits were assigned back to the original query nucleotide sequence and later used in GO Slimmer search. All searched gene symbols were uploaded to the Slimmer tool of AmiGO software for mapping the granular annotations using plant GO slim (Suparna Mundodi, December 2002). Based on these GO term assignments, the complete unigene EST set for the cDNA library was organized around GO hierarchies and divided into three categories: biological processes, cellular components and molecular functions (Harris et al. 2004).

Quantitative, real-time RT-PCR

Quantitative real-time PCR (qRT-PCR) was performed to analyze the expression of selected genes from both samples using gene-specific PCR primers. Total RNA (2 μg) was used as template for reverse transcription with the MMLV Reverse Transcriptase (EnZynomics, Korea). qRT-PCR was conducted using the SYBR Green PCR Kit (TaKaRa) in a total volume of 10 μL on an Rotor-Gene 6000 PCR machine (Corbett, Australia). A melting curve was recorded at the end of every run to assess product specificity. For each target gene, PCR conditions (primer concentrations and cDNA quantity) were optimized and PCR efficiency was determined. Experimental samples were evaluated in triplicate and qRT-PCR reactions for each were run in duplicate. Target gene expression was normalized with actin and corrected according to the PCR efficiency value. The products of qRT-PCR were run on agarose gels and consistently showed accurate molecular weight.

Results

RNA isolation and cDNA library construction

High-quality total RNA was obtained using modified CTAB extraction buffer. The formaldehyde agarose gel electrophoresis demonstrated that the extraction protocol described here was efficient in yielding a high quality of total RNA from all moss tissues and control grass species (*Deschampsia antarctica*; Fig. 1a). The yields of total RNA (μg/g fresh weight) were as follows: 20–30 for brown tissue of mosses, 80–126 for green tissue of mosses and 364–415 for grass.

For all samples, the A260/230 ratio was higher than 2.0. This indicated that the RNA was of high purity and without polyphenol and polysaccharide contamination. The A260/280 ratios ranged from 1.91 to 2.02, indicating lack of protein contamination. The RNA integrity was assessed by the sharpness of ribosomal RNA bands visualized on a denaturing 1.2% agarose gel. For all RNA samples tested, two

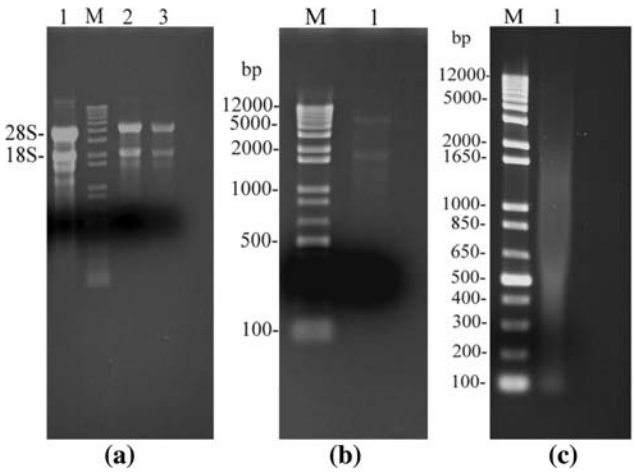

Fig. 1 **a** Total RNA isolated using the CTAB extraction protocol was separated on a 1.2% denaturing formaldehyde agarose gel and stained with ethidium bromide. *M* 1 kb DNA ladder, *1* control grass total RNA (~4 μg), *2* moss total RNA (~1 μg), *3* moss total RNA (~0.5 μg). **b** Agarose gel electrophoresis of Poly A^+ RNA extracted from total RNA of Arctic moss. *M* 1 kb DNA ladder, *1* poly A^+ RNA. **c** Agarose gel electrophoresis of double stranded cDNA amplified by long-distance PCR. *M* 1 kb DNA ladder, *1* double stranded cDNA

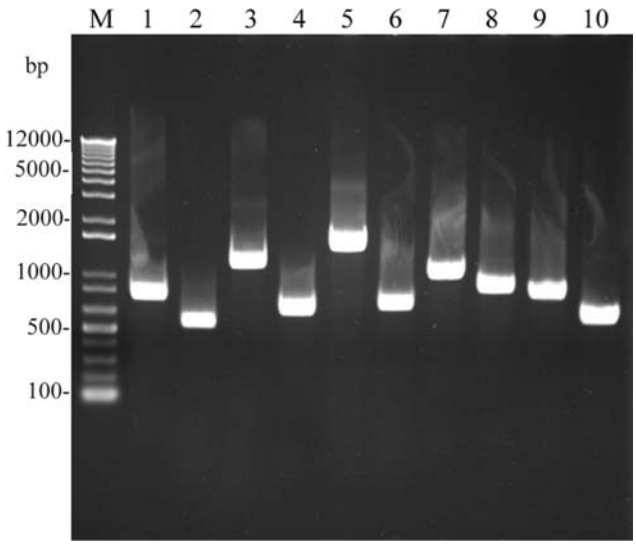

Fig. 2 cDNA inserts were verified by PCR and visualized on 1.2% agarose gel electrophoresis. *M* 1 kb DNA ladder, *1–10* PCR products

distinct bands (28S and 18S ribosomal RNA) at approximately 4.5 and 1.9 kb without degradation were observed, respectively (Fig. 1a).

Poly A^+ RNA was isolated using oligo d(T) cellulose column. Gel electrophoresis showed that moss poly A^+ RNA was enough for high-quality cDNA library construction (Fig. 1b). Double-strand cDNA was generated by LD-PCR and showed a large size distribution (0.1–10 kb) by electrophoresis gel analysis (Fig. 1c).

After construction of cDNA library, the titer was determined using phages to transfect XL1-Blue. The result showed that the titer was 3.62×10^6 pfu/mL and the library consisted of 1.81×10^6 recombinants, the recombinant rate was 97.41%. For better preservation, the cDNA library was amplified, resulting at least 9.83×10^9 pfu/mL. Ten monoclonal recombinant phages were picked up randomly and PCR was performed to verify the size of cDNA inserts in the phages. The PCR products were electrophoresed on 1.2% agarose gel and their sizes were ranged from 0.5 to 2.0 kb (Fig. 2).

Table 1 Representative genes in the *A. turgidum* cDNA library

ID	ESTs	Representative clone	*E* value	Length	Description
Contig02	8	MC28	8.00E−58	1,128	Cytochrome b559 subunit alpha [*Spinacia oleracea*]
Contig03	2	MD13	7.00E−06	1,334	Predicted protein [*Physcomitrella patens* subsp. *patens*]
Contig04	4	MC94	1.00E−24	1,048	Conserved hypothetical protein [*Magnetospirillum gryphiswaldense*]
Contig05	2	MB47	9.00E−40	1,025	Eukaryotic release factor 3 [*Ricinus communis*]
Contig06	2	MG85	6.00E−145	1,193	Polyubiquitin [*Pinus sylvestris*]
Contig07	3	MC78	1.00E−106	1,482	Predicted protein [*Physcomitrella patens*]
Contig08	2	MC39	2.00E−18	1,010	Choline kinase [*Pisum sativum*]
Contig09	2	MI31	9.00E−61	706	Heat-shock protein 60 [*Ageratina adenophora*]
Contig11	3	MG72	4.00E−59	950	Predicted protein [*Physcomitrella patens*]
Contig13	3	MA66	3.00E−22	711	Enhanced disease susceptibility 5 [*Arabidopsis thaliana*]
Contig14	3	MC60	2.00E−12	700	Low temperature-induced protein [*Hordeum vulgare* subsp. *vulgare*]
Contig15	2	MA84	3.00E−33	701	Probable thioredoxin H [*Picea mariana*]
Contig17	2	MF90	4.00E−59	975	Ribosomal protein S12 [*Physcomitrella patens*]
Contig19	2	ML45	6.00E−87	1,245	Predicted protein [*Physcomitrella patens*]
Contig20	2	MI20	9.00E−11	683	Predicted protein [*Physcomitrella patens*]
Contig21	2	MF71	7.00E−43	759	Binding protein [*Arabidopsis thaliana*]
Contig23	2	MD88	4.00E−27	929	Predicted protein [*Physcomitrella patens*]
Contig24	2	MA52	1.20E−86	863	19s proteasome subunit 9 [*Arabidopsis thaliana*]
Contig25	2	MN47	1.00E−57	680	Predicted protein [*Physcomitrella patens*]
Contig27	2	MC14	3.00E−55	1,249	Alcohol dehydrogenase [*Neosartorya fischeri* NRRL]
Contig28	2	MH90	9.00E−32	1,039	Predicted protein [*Physcomitrella patens*]
Contig29	2	ML95	9.00E−32	944	RNA polymerase [*Physcomitrella patens*]
Contig30	6	MJ89	2.00E−08	685	Low temperature and salt responsive protein [*Sonneratia alba*]
Contig31	2	MD82	7.00E−21	751	GSH-dependent glutathione dehydrogenase [*Arabidopsis thaliana*]
Contig32	2	MC97	2.00E−58	874	Phosphoglyceromutase [*Malus domestica*]
Contig33	2	MH46	7.40E−80	923	Cholinephosphate cytidylyltransferase [*Arabidopsis thaliana*]

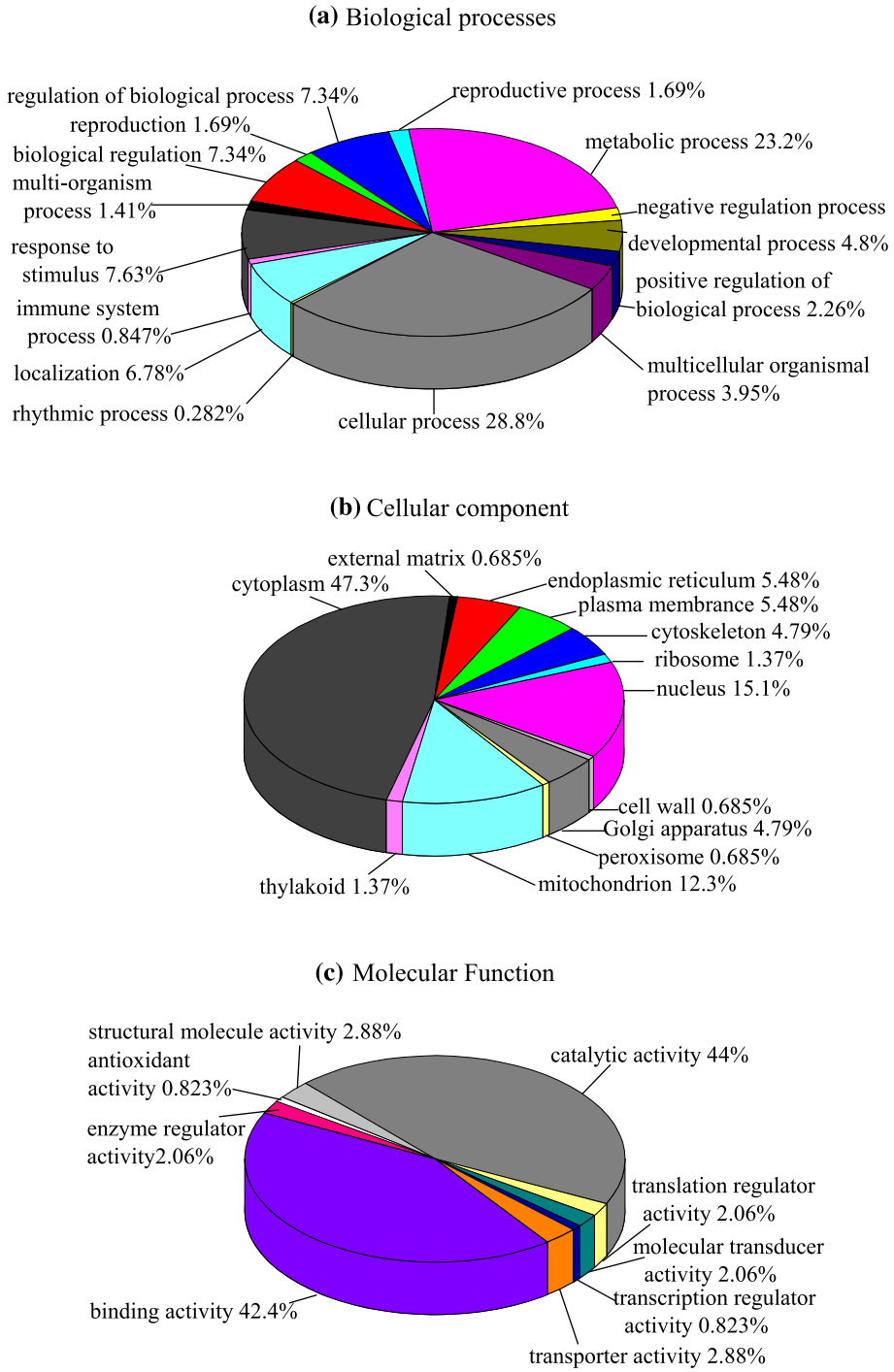

Fig. 3 Percentage representation of Gene Ontology mappings for ESTs of the Arctic moss *Aulacomnium turgidum*

General characteristics of *A. turgidum* ESTs

A total of 509 cDNA clones were randomly selected from the library and single-pass sequences were generated. The sequences of poor quality and of <100 bases were removed, and the final EST number was 437 with mean lengths of 617 bp (GenBank Accession No. from FL685677 to FL686115). Cluster analysis revealed 360 unigenes of 33 contigs and 327 singletons with a redundancy level of 9.2%. The sequences of 360 unigenes were translated into all possible reading frames, and compared with those in NCBI non-redundant protein database. Among these sequences, approximately 46.0 and 22.2%, respectively, had significant matches with sequences in the non-redundant protein database (E value $\leq 10^{-5}$), which encode for "known proteins" and "predicted/unknown proteins" based on the BLASTX results. In addition, 31.9% of the ESTs showed no similarity to any sequences in the database

Table 2 Stress-related genes isolated from the *A. turgidum* cDNA library

ID	*E* value	Putative identity	Related stress incidents
Contig08	3.00E−19	Choline kinase [*Oryza sativa*]	Salinity stress
MC32	1.40E−48	DEAD/DEAH box helicase, putative (RH17) [*Arabidopsis thaliana*]	Salinity stress
ML12	4.90E−72	NADP + isocitrate dehydrogenase [*Arabidopsis thaliana*]	Salinity stress
Contig32	2.40E−47	Phosphoglycerate mutase [*Arabidopsis thaliana*]	Drought-stress
MC48	4.50E−23	Oxidoreductase [*Arabidopsis thaliana*]	Drought-stress
Contig09	9.00E−61	Heat-shock protein 60 [*Ageratina adenophora*]	Thermal and oxidative stresses
MG69	1.00E−18	2,3-Biphosphoglycerate-independent phosphoglycerate mutase [*Arabidopsis thaliana*]	Thermal and oxidative stresses
Contig24	1.20E−86	19s proteasome subunit 9 [*Arabidopsis thaliana*]	Thermal and oxidative stresses
MG84	3.80E−67	20S proteasome alpha subunit B1 [*Arabidopsis thaliana*]	Thermal and oxidative stresses
MJ67	1.80E−85	20S proteasome beta subunit C1 [*Arabidopsis thaliana*]	Thermal and oxidative stresses
MG61	2.70E−31	26S proteasome regulatory subunit S2 1A [*Arabidopsis thaliana*]	Thermal and oxidative stresses
MC42	7.50E−32	Ubiquitin-protein ligase [*Arabidopsis thaliana*]	Thermal and oxidative stresses
Contig06	1.70E−135	Polyubiquitin10 [*Arabidopsis thaliana*]	Thermal and oxidative stresses
MB35	5.40E−41	Histone mono-ubiquitination 2 [*Arabidopsis thaliana*]	Thermal and oxidative stresses
MK56	3.70E−30	Ubiquitin-specific protease [*Arabidopsis thaliana*]	Thermal and oxidative stresses
MJ85	2.40E−24	Phospholipase A2 [*Arabidopsis thaliana*]	Thermal and oxidative stresses
Contig15	5.30E−29	Thioredoxin H-type 9 [*Arabidopsis thaliana*]	Oxidative stress
MG80	3.90E−19	Thioredoxin (Trx) [*Arabidopsis thaliana*]	Oxidative stress
MF95	5.70E−23	Aldo–keto reductase family 1, member A1a [*Arabidopsis thaliana*]	Oxidative stress
ML01	8.60E−14	UDP-glycosyltransferase [*Arabidopsis thaliana*]	Oxidative stress
MN78	1.50E−33	Peroxidase/catalase [*Arabidopsis thaliana*]	Oxidative stress
Contig02	8.00E−58	Cytochrome b559 subunit alpha [*Spinacia oleracea*]	Freezing and thawing stresses or solar radiation
Contig14	9.10E−16	Rare-cold-inducible 2a [*Arabidopsis thaliana*]	Cold and freezing stresses
Contig30	9.40E−10	Rare-cold-inducible 2b [*Arabidopsis thaliana*]	Cold-response or freezing tolerance
MD71	2.80E−22	Regulator of chromosome condensation [*Arabidopsis thaliana*]	Cold tolerance
MG53	1.60E−10	Alcohol dehydrogenase [*Arabidopsis thaliana*]	Low temperature
MA16	2.70E−09	Ankyrin repeat-containing protein [*Schizosaccharomyces pombe*]	Disease resistance and antioxidation metabolism
ME79	3.40E−61	1,3-Beta-glucan synthase [*Arabidopsis thaliana*]	Wounding, physiological stress, infection
MJ19	4.30E−32	Copper amine oxidase, putative [*Arabidopsis thaliana*]	Signal in abscisic acid (ABA)
MC67	1.00E−64	Protein phosphatase 2C family protein [*Arabidopsis thaliana*]	Hormone abscisic acid (ABA) signaling
Contig27	2.50E−45	Glutathione-dependent formaldehyde dehydrogenase [*Schizosaccharomyces pombe*]	Wounding, jasmonic acid and salicylic acid
Contig13	3.90E−21	Enhanced disease susceptibility 5 [*Arabidopsis thaliana*]	Detoxification
MJ02	4.20E−105	Pyruvate dehydrogenase kinase [*Arabidopsis thaliana*]	Abiotic stress
MA98	1.80E−38	Trehalose-6-phosphate synthase [*Arabidopsis thaliana*]	Stress tolerance/cell shape and plant architecture
MH85	2.10E−103	Long-chain fatty-acid–CoA ligase [*Arabidopsis thaliana*]	Tolerance to biotic and abiotic stresses

(*E* value $\geq 10^{-5}$) and were deemed "No significant similarity found". Furthermore, more than 74.5% of the "predicted/unknown protein" in our libraries matched *Physcomitrella patens* EST sequence, annotated as predicted. After sequence assembly, results showed that the highest number of sequences clustered into one contig was 8. The representative genes with *E* value $\leq 10^{-5}$ are shown in Table 1.

Functional analysis of ESTs

AmiGO software for BLAST searching the Gene Ontology Database was used under the default setting to determine similarity to the known genes and extract gene symbols for grouping sequences into functional classifications. In total, 158 annotations acquired with gene symbols were exported into a Microsoft Excel data sheet and sorted manually. In

Table 3 Primer pairs used in the qRT-PCR analysis

Target	Forward	Reverse
Sequence (5′ → 3′)		
MA52	TTGTCAGCGAGCTTCATCTCCAGT	GCCCACGTTCCTTCGTTCTTGAAA
MA66	TCCTTCAGCTCCAACTGCACTTCT	ATGGGTCAGCAGACTCCACAAT
MA98	CCAACAGCTTCAAAGACCGCAAGT	ACTTCACGAACGACACGACCTCAT
MC39	TAGTGCCTCTCCGACAATTGGGTT	AACTTCCCTGGAGATACCTTGCCA
MI31	GGCATGGGAGGCATGTACTAGAAA	TATGGCAGAGCATGCCAGATGACT
MJ89	AGTGTCGCAAGAAGAGAGCAGTGA	GCTCTTCCTGAAGGGAGCAATCTT
MK20	TTTGTGAAATGCTTGCTCCGGGAC	ACTTGCACACTTAGTTCGCCACAC
Actin	GACTGCTTGATTGATGAGCGCCAA	TGAACGGTACGGACCCTAATCACA

this way more than 93% of the matches came from the *Arabidopsis thaliana* genome, whereas only 11 matches were found in *Nicotiana sylvestris*, *Nostoc* sp. *PCC* 7120, *Schizosaccharomyces pombe* and *Dictyostelium discoideum*.

A more detailed functional annotation was performed by mapping assigned unique gene symbols to the Gene Ontology Consortium structure which provides a structured and controlled vocabulary to describe gene products according to three ontologies: cellular components, biological processes and molecular functions. The most represented biological processes were "cellular" and "metabolic", accounting for 28.8 and 23.2%, respectively, of the total 158 unique sequences assigned with at least one GO term (Fig. 3a). In terms of cellular components, 43.7% were related to "cytoplasm" (Fig. 3b). In terms of molecular functions, 44% were involved in "catalytic activity" and 42.4% in "binding activity" (Fig. 3c).

In addition, 35 unigenes showing similarity to proteins related to stress incident were presented in Table 2. Among them, 17, 5, 3 and 2 genes are candidates for oxidative stress, cold stress, salinity stress and abscisic acid (ABA)-inducible genes, respectively. To confirm the relationship of those genes and various stress responses, we performed quantitative RT-PCR analyses with gene-specific primer sets (Table 3). Seven genes from Table 2 were analyzed, and actin was used as a quantitative control. The results of the RT-PCR analysis agreed in most cases with expectations from GO analysis.

As two genes showing significant similarity with RCI2a (rare-cold-inducible 2a) and RCI2b from *Arabidopsis* showed four- and tenfold mRNA accumulation by low temperature (Fig. 4a), they were selected to investigate the relationship between specific gene expression and Arctic environmental conditions, cold and UV-B irradiation. AtuRCI2a and AtuRCI2b from *A. turgidum* are 54 and 59 amino acids, respectively. Their sequence alignment revealed 65% identity and 84% similarity between them (Fig. 5a). Their counterparts from other plant species showed overall similarity, which suggests that these proteins are conserved throughout the plant kingdom. Hydropathy analyses indicated that they are highly hydrophobic and contain two potential transmembrane domains (data not shown). Under cold treatment, both *AtuRCI2a* and *AtuRCI2b* mRNAs accumulated gradually, reaching the peak after 1 and 2 days, respectively (Fig. 5B). *AtuRCI2b* transcripts have shown more rapid and high expression level than *AtuRCI2a*. However, under UV-B treatment, *AtuRCI2a* mRNA accumulated upto 10 times the basal level and *AtuRCI2a* mRNA was not induced (Fig. 5c).

Discussion

Isolating high-quality nucleic acid is a prerequisite for molecular biological studies. However, isolating high-quality RNA from Arctic mosses is quite difficult because of their richness in polyphenols, polysaccharides and secondary metabolites. In the present study, we employed a CTAB extraction method with 1.4 M NaCl and 1% PVP, suitable for isolating high-quality RNA from moss tissues. LiCl is used to precipitate RNA rather than alcohol with a monovalent cation, because LiCl will not efficiently precipitate DNA, protein or carbohydrate (Birnboim 1992). Finally, the yield of total RNA was 80–126 μg/g for green tissue of mosses and the quality was confirmed to be superior by spectrophotometer and formaldehyde agarose gel (Fig. 1).

A cDNA library should represent all expressed genes in the tissue from which the library was constructed. In this research, SMART technique was used to construct a cDNA library of *A. turgidum* tissues, resulting in at least 1.81×10^6 recombinants in the primary library. ESTs with the mean lengths of 617 bp also showed that the sequence integrality of recombinant cDNA was good in quality. BLAST was used to search the entire NCBI GenBank with an *E* value threshold of 10^{-5}, which revealed 46% of the

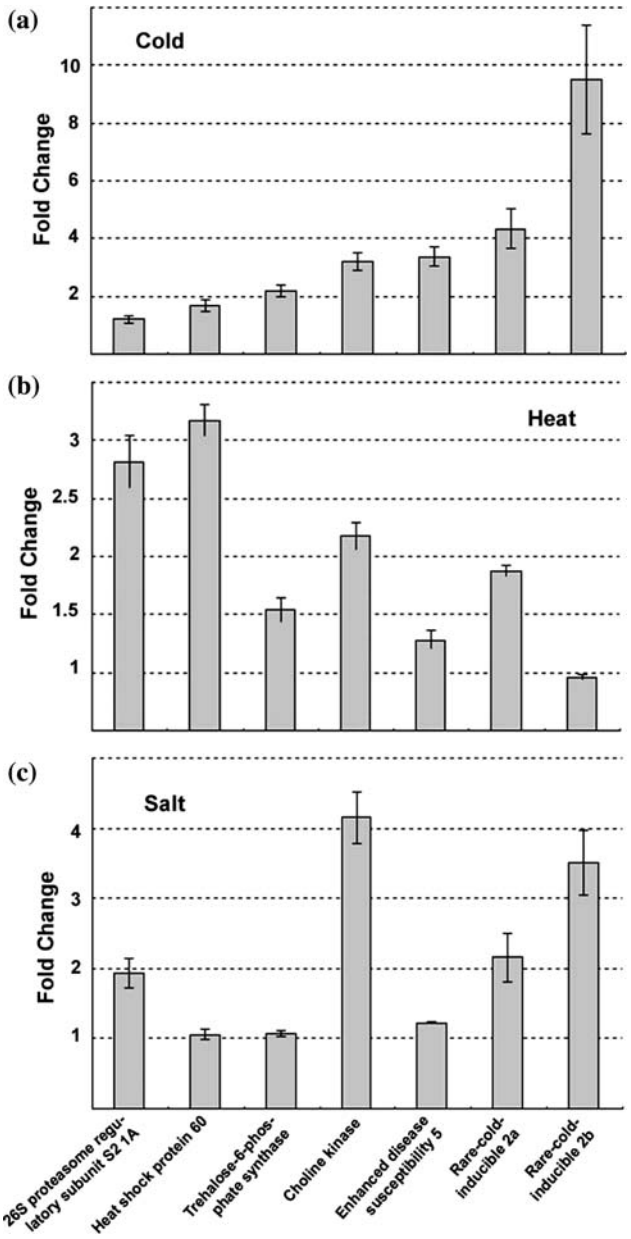

Fig. 4 Quantitative RT-PCR analysis of various stress-related genes from *A. turgidum* under abiotic stresses treatments. Fold changes were calculated by comparing gene expression in plants under stress conditions with control plants without stress treatment after normalization with the actin gene of *A. turgidum*. Cold, incubation at 4°C for 24 h (**a**); heat, incubation 37°C for 24 h (**b**); NaCl, incubation with 300 mM NaCl for 24 h (**c**)

cDNA clones with high homologies to genes with known functions in the database. The results showed that the most representative genes in our library are cytochrome b559 subunit alpha, low temperature and salt responsive protein, polyubiquitin and enhanced disease susceptibility 5 (Table 1). In a previous study of ESTs of *Deschampsia antarctica*, an Antarctic extremophile angiosperm plant species, photosynthesis-related genes were most abundant (Lee et al. 2008). In *A. turgidum*, the portion of stress-related genes is relatively high, which might be caused to some extent by the cDNA amplification step during cDNA library preparation.

Automatic functional annotation methods basically rely on the sequence, structure, phylogenetic or co-expression relationships between known and novel sequences (Frishman 2007). Most tools provide GO annotation of sequences data through homology searches. However, function transfer from homologous sequences is comparatively highly error prone (Jones et al. 2007). In our study, we explored BLAST Search of the GO website and assigned functional symbols manually. Manual curation guarantees a high level of annotation correctness. Finally, the official software AmiGO was used for the high-throughput functional annotation according to GO hierarchies. A relatively high number of stress-related genes were isolated from cDNA library and annotated in the dataset (Table 2).

A group of proteins, part of the plant antioxidant system, are rapidly activated in response to oxidative stress generated by heat, including superoxide dismutases, catalases and peroxidases. Modulation of the heat stress response is also dependent on cellular control of degradation and maintenance of quality of proteomes by the ubiquitin–proteasome system (Mathew and Morimoto 1998; Mathew et al. 1998). Therefore, we identified a total of 17 genes related to oxidative stress, which was more than the genes related to cold stress, salinity stress or ABA-induction (Table 2).

Heat-shock proteins (Hsps)/chaperones are responsible for protein folding, assembly, translocation and degradation in many normal cellular processes, stabilize proteins and membranes, and can assist in protein refolding under stress conditions (Wang et al. 2004). The HSP60 in chloroplasts is the Rubisco subunits binding protein also known as chloroplast chaperonin (Yordanov 1995). In present study, we identified a HSP60 gene which assembled from two ESTs of moss cDNA library and its expression was induced by the heat and also by the cold (Fig. 4). Therefore, the common action of HSPs in Arctic moss may involve in mitigation of temperature changing, protecting proteins against oxidation damage and folding of newly synthesized proteins.

To endure freezing and thawing stress in Arctic, mosses could also reduce the efficiency of photosystem II (PSII) which protect them from photoinhibitory damage in environments where freezing temperatures occur in conjunction with high levels of solar radiation (Lovelock et al. 1995). Using isolated reaction centers of PSII, Barber and De Las Rivas (1993) confirmed that the low potential form of cytochrome b559 can accept electrons directly from reduced pheophytin and protect the reaction center against photoinhibition. In this study, cytochrome b559 constitutes the most abundantly expressed gene in *A. turgidum*, indicating

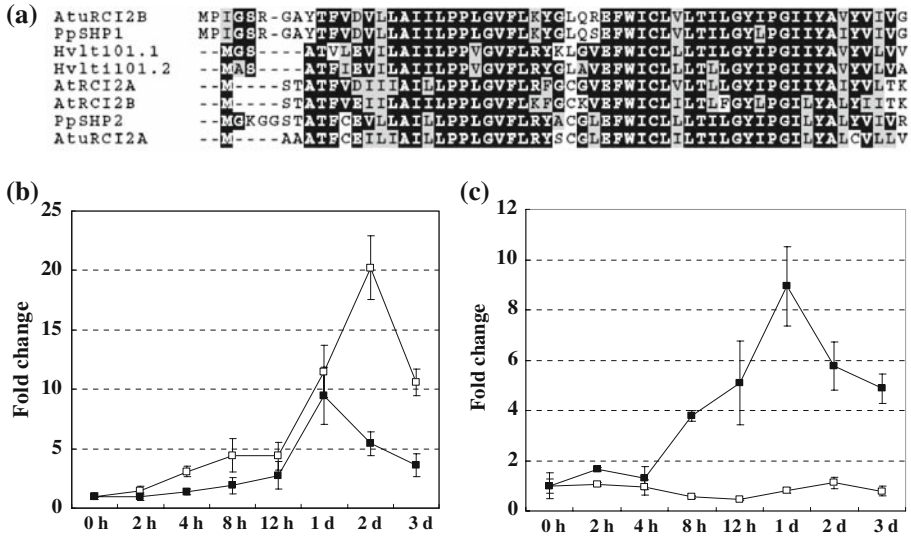

Fig. 5 Multiple sequence alignment plant RCI family proteins and expression analysis of *AtuRCI2a* and *AtuRCI2b*. **a** Sequence alignment between AtuRCI2a and AtuRCI2b of *A. turgidum*, related proteins Hvlti101.1 and Hvlti101.2 from barley (P68179 and Q9ARD5), AtRCI2A and AtRCI2B form *Arabidopsis thaliana* (NP_187239 and NP_187240), and PpSHP1 and PpSHP2 from *Physcomitrella patens* (AAR87655 and AAR87656). Quantitative RT-PCR analysis of *AtuRCI2a* and *AtuRCI2b* under 4°C (**b**) and UV-B (**c**) treatment during indicated periods. Fold changes were calculated by comparing gene expression in plants under stress conditions with control plants at starting points after normalization with the actin gene of *A. turgidum*. *AtuRCI2a* (*closed square*), *AtuRCI2b* (*open square*)

a photosynthetic acclimation to low levels of temperature and solar radiation.

Glutathione dehydrogenase, glutathione-dependent formaldehyde dehydrogenase and enhanced disease susceptibility 5 are related to plant antioxidation, detoxification and pathogen resistance, respectively (Achkor et al. 2003; Rogers and Ausubel 1997). Other genes identified from cDNA library and EST analysis may implement their possible roles in plant survival in harsh polar condition.

From *A. turgidum* cDNA library, two new rare-cold-inducible (RCI) genes were isolated. They encode very small (50–60 residues) and highly hydrophobic proteins with two potential transmembrane domains, constituting small gene families in various plant species (Goddard et al. 1993; Capel et al. 1997; Brown et al. 2001). They are considered to be localized in the plasma membrane, which is the primary target of stress recognition, suggesting their significance in a broad range of stress conditions such as low temperature, freezing, and high-salinity. In this study, two RCI genes were cloned from *A. turgidum*. While their expression patterns were similar in the cold treatment, these differed clearly more in the UV-B treatment, suggesting that they play divergent roles under different abiotic stress conditions in the Arctic (Fig. 5b, c). Further studies will focus on identifying the activation mechanisms of these genes under various abiotic stresses, as shown for *Arabidopsis* RCI proteins (Medina et al. 2001). As far as we know, this is the first report on RCI gene expression under UV-B irradiation.

Here, we have presented and discussed only an initial analysis of the EST dataset and further characterized selected examples with emphasis on genes with putative roles in stress response. The amount of redundancy present in the EST dataset is relatively low. However, identification of novel genes, determination of their expression patterns, and an improved understanding of their functions in stress adaptation gained through this study will provide us the basis of effective engineering strategies to improve stress tolerance.

Acknowledgments The research was financially supported by a Grant PE08050 from the Korea Polar Research Institute.

References

Achkor H, Díaz M, Fernández MR, Biosca JA, Parés X, Martínez MC (2003) Enhanced formaldehyde detoxification by overexpression of glutathione-dependent formaldehyde dehydrogenase from *Arabidopsis*. Plant Physiol 132:2248–2255. doi:10.1104/pp.103.022277

Ashton NW, Cove DJ (1977) The isolation and preliminary characterisation of auxotrophic and analogue resistant mutants of the moss, *Physcomitrella patens*. Mol Gen Genet 154:87–95. doi:10.1007/BF00265581

Ayres E, van der Wal R, Sommerkorn M, Bardgett RD (2006) Direct uptake of soil nitrogen by mosses. Biol Lett 2:286–288. doi:10.1098/rsbl.2006.0455

Barber J, De Las Rivas J (1993) A functional model for the role of cytochrome b559 in the protection against donor and acceptor side photoinhibition. Proc Natl Acad Sci USA 90:10942–10946. doi:10.1073/pnas.90.23.10942

Billings WD, Mooney HA (1968) The ecology of arctic and alpine plants. Biol Rev 43:481–529. doi:10.1111/j.1469-185X.1968.tb00968.x

Birks HJB, Jones VJ, Rose NL (2004) Recent environmental change and atmospheric contamination on Svalbard as recorded in lake sediments—an introduction. J Paleolimnol 31:403–410. doi:10.1023/B:JOPL.0000022542.87722.a6

Birnboim HC (1992) Extraction of high molecular weight RNA and DNA from cultured mammalian cells. Methods Enzymol 216:154–160

Brown AP, Dunn MA, Goddard NJ, Hughes MA (2001) Identification of a novel low-temperature-response element in the promoter of the barley (*Hordeum vulgare* L) gene blt101.1. Planta 213:770–780. doi:10.1007/s004250100549

Capel J, Jarillo JA, Salinas J, Martínez-Zapater JM (1997) Two homologous low-temperature-inducible genes from *Arabidopsis* encode highly hydrophobic proteins. Plant Physiol 115:569–576. doi:10.1104/pp.115.2.569

Flatberg KI, Thingsgaard K (2003) Taxonomy and geography of *Sphagnum tundrae* with a description of *S. mirum*, sp. nov. (Sphagnaceae, sect. Squarrosa). Bryol 106:501–515. doi:10.1639/0007-2745(2003)106[501:TAGOST]2.0.CO;2

Frishman D (2007) Protein annotation at genomic scale: the current status. Chem Rev 107:3448–3466. doi:10.1021/cr068303k

Fu XH, Geng SL, Su GH, Zeng QL, Shi SH (2004) Isolating high-quality RNA from mangroves without liquid nitrogen. Plant Mol Biol Report 22:197a–197e. doi:10.1007/BF02772728

Goddard NJ, Dunn MA, Zhang AJ, White PL, Jack PL, Hughes MA (1993) Molecular analysis and spatial expression pattern of a low-temperature-specific barley gene, *blt 101*. Plant Mol Biol 23:871–879

Gornall JL, Jónsdóttir IS, Woodin SJ, Van der Wal R (2007) Arctic mosses govern below-ground environment and ecosystem processes. Oecologia 153:931–941. doi:10.1007/s00442-007-0785-0

Greilhuber J, Såstad SM, Flatberg KI (2003) Ploidy determination in *Sphagnum* samples from Svalbard, Arctic Norway, by DNA image cytometry. J Bryol 25:235–239. doi:10.1179/037366803225013083

Harris MA, Clark J, Ireland A et al (2004) The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 32:D258–D261. doi:10.1093/nar/gkh036

Jones C, Brown A, Baumann U (2007) Estimating the annotation error rate of curated GO database sequence annotations. BMC Bioinformatics 8:170. doi:10.1186/1471-2105-8-170

Lee H, Cho HH, Kim IC, Yim JH, Lee HK, Lee YK (2008) Expressed sequence tag analysis of Antarctic hairgrass *Deschampsia antarctica* from King George Island, Antarctica. Mol Cells 25:258–264

Longton RE (1997) The role of bryophytes and lichens in polar ecosystems. In: Woodin SJ, Marquiss M (eds) Ecology of Arctic environments. Blackwell, Oxford, pp 69–96

Lovelock CE, Jackson AE, Melick DR, Seppelt RD (1995) Reversible photoinhibition in Antarctic moss during freezing and thawing. Plant Physiol 109:955–961

Mathew A, Morimoto RI (1998) Role of the heat-shock response in the life and death of proteins. Ann N Y Acad Sci 851:99–111. doi:10.1111/j.1749-6632.1998.tb08982.x

Mathew A, Mathur SK, Morimoto RI (1998) Heat shock response and protein degradation: regulation of HSF2 by the ubiquitin–proteasome pathway. Mol Cell Biol 18:5091–5098

Medina J, Catalá R, Salinas J (2001) Developmental and stress regulation of RCI2A and RCI2B, two cold-inducible genes of *Arabidopsis* encoding highly conserved hydrophobic proteins. Plant Physiol 125:1655–1666. doi:10.1104/pp.125.4.1655

Morgan-Kiss RM, Priscu JC, Pocock T, Gudynaite-Savitch L, Huner NPA (2006) Adaptation and acclimation of photosynthetic microorganisms to permanently cold environments. Microbiol Mol Biol Rev 70:222–252. doi:10.1128/MMBR.70.1.222-252.2006

Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K, Ito Y, Yoshiwara K, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. Plant Physiol 133:1755–1767. doi:10.1104/pp.103.025742

Reski R, Frank W (2005) Moss (*Physcomitrella patens*) functional genomics-gene discovery and tool development, with implications for crop plants and human health. Brief Funct Genomics Proteomics 4:48–57. doi:10.1093/bfgp/4.1.48

Rogers EE, Ausubel FM (1997) *Arabidopsis* enhanced disease susceptibility mutants exhibit enhanced susceptibility to severa1 bacterial pathogens and alterations in *PR-I* gene expression. Plant Cell 9:305–316. doi:10.1105/tpc.9.3.305

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Plainview

Skotnicki ML, Ninhami JA, Selkirk PM (2000) Genetic diversity, mutagenesis and dispersal of Antarctic mosses—a review of progress with molecular studies. Antarct Sci 12:363–373

Skotnicki ML, Mackenzie AM, Clements MA, Selkirk PM (2005) DNA sequencing and genetic diversity of the 18S–26S nuclear ribosomal internal transcribed spacers (ITS) in nine Antarctic moss species. Antarct Sci 17:377–384. doi:10.1017/S0954102005002816

Wang WX, Vinocur B, Shoseyov O, Altman A (2004) Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends Plant Sci 9:244–252. doi:10.1016/j.tplants.2004.03.006

Yordanov I (1995) Responses of photosynthesis to stress and plant growth regulators. Bulg J Plant Physiol 21:51–70