J. Microbiol. Biotechnol. (2012), **22**(7), 902–906
http://dx.doi.org/10.4014/jmb.1201.01002
First published online April 16, 2012
pISSN 1017-7825 eISSN 1738-8872

JOURNAL OF MICROBIOLOGY AND BIOTECHNOLOGY

Analysis of Expressed Sequence Tags from the Antarctic Psychrophilic Green Algae, *Pyramimonas gelidicola*

Jung, Woongsic[1,2†], Sung Gu Lee[1,3†], Se Won Kang[4], Yong Seok Lee[4], Jun Hyuck Lee[1,3], Sung-Ho Kang[3,5], Eon Seon Jin[2*], and Hak Jun Kim[1,3*]

[1]*Division of Polar Life Sciences, Korea Polar Research Institute, Incheon 406-840, Korea*
[2]*Department of Life Science, College of Natural Sciences, Hanyang University, Seoul 133-791, Korea*
[3]*Department of Polar Sciences, University of Science and Technology, Incheon 406-840, Korea*
[4]*Department of Parasitology, College of Medicine and Frontier Inje Research for Science and Technology, Inje University, Busan 614-735, Korea*
[5]*Division of Polar Climate Research, Korea Polar Research Institute, Incheon 406-840, Korea*

Received: January 3, 2012 / Revised: February 29, 2012 / Accepted: March 4, 2012

Expressed sequence tags (ESTs) from the Antarctic green algae *Pyramimonas gelidicola* were analyzed to obtain molecular information on cold acclimation of psychrophilic microorganisms. A total of 2,112 EST clones were sequenced, generating 222 contigs and 219 singletons, and 200 contigs and 391 singletons from control (4°C) and cold-shock conditions (–2°C), respectively. The complete EST sequences were deposited to the DDBJ EST database (http://www.ddbj.nig.ac.jp/index-e.html) and the nucleotide sequences reported in this study are available in the DDBJ/EMBL/GenBank. These EST databases of Antarctic green algae can be used in a wide range of studies on psychrophilic genes expressed by polar microorganisms.

Keywords: Expressed sequence tags (ESTs), *Pyramimonas gelidicola*, psychrophilic, Antarctic, database

Over 70% of the earth consists of low temperature ecosystems that have temperatures below or close to 0°C, the freezing point of water. Antarctica is a continent of extremely harsh environment on the planet, which has the lowest temperature of –89.6°C [20]. During seasonal changes, the Antarctic region undergoes principal environmental changes, which affect subsequent influence to living organisms. In transition from Fall to Winter, the temperature and the solar insolation decrease in this area [4]. The water temperature of the Antarctic coasts is around –1.8°C during Winter and increases to approximately 5°C in the Summer season [22]. Antarctic microorganisms, including photoautotrophs, have developed a number of ways to adapt themselves to the harsh environments over generations [7]. Cold-adapted microorganisms contain a high concentration of polyunsaturated fatty acids in the membrane to grow and photosynthesize at low temperatures [14]. In addition, Antarctic plants and marine microorganisms produce ice active substances such as antifreeze proteins (AFPs), which lower the freezing point; consequently, they are able to survive at below subfreezing temperatures [9, 19]. These survival strategies in cold environments were found in Prasinophytes [14]. The EST databases of the marine diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana* were integrated together [11] and the database currently contains 27,310 EST sequences (12,136 for *P. tricornutum* and 15,174 for *T. pseudonana*). The EST datasets of polar diatoms were constructed from *Fragilariopsis cylindrus* [13] and *Chaetoceros neogracile* [7]. The *F. cylindrus* and *C. neogracile* EST databases contain 1,376 EST sequences (996 tentative unique sequences) and 1,881 sequences, respectively. Most of the ESTs from *F. cylindrus* encoded the genes related to translation, ribosomal structure, and biosynthesis. It was revealed that over 28% of *C. neogracile* ESTs was related to cellular reactions for metabolism, genetic information processing, and photosynthesis.

Pyramimonas gelidicola is one of the dominant primary producers in the Antarctic area. This phototrophic nanoflagellate is mainly distributed in lacustrine and oligotrophic saline lakes in the Antarctic continent, such as Ace Lake [2, 3, 6, 10, 14, 21]. This flagellate grazed 0.4% to 16% *in situ* bacterial biomass (between 4% and over 100% *in situ* bacterial production) during the summers of 1997 and 1999 at Ace Lake in Vestfold Hills, eastern

*Corresponding author
Phone: +82-32-260-6253; Fax: +82-32-260-6256;
E-mail: hjkim@kopri.re.kr, esjin@hanyang.ac.kr

†These authors contributed equally to this manuscript.

Antarctica. In spite of their ecological and physiological importance in Antarctic ecosystems, there is little information of *P. gelidicola* at the level of molecular biology. In this report, we aimed to analyze ESTs of *P. gelidicola* expressed at cold temperature.

The Antarctic phototrophic nanoflagellate *Pyramimonas gelidicola* (KOPRI AnM0046) that was maintained at the Korea Polar Research Institute (KOPRI) was cultured in F/2 seawater media filtered with 0.2 μm pore size filters. The seed culture was grown at 2°C in a 10 L batch culture under 80 μmol photons/m^2s with continuous light. Carbon dioxide was supplied by air bubbling and the culture media were mixed continuously using a stirrer. Five liter samples of control microalgal were harvested when *in vitro* chlorophyll *a* concentration was 34.2 μg/l. The remaining 5 L of the culture was further maintained at −2°C (approximate freezing temperature of seawater) for a week under the same light condition as the control, and then harvested. The collected samples were kept immediately at −70°C. Total RNA was extracted from *P. gelidicola* with Trizol reagent (Invitrogen). The mRNA was isolated from approximately 100 μg of total RNA with the Absolutely mRNA Purification Kit (Stratagene, La Jolla, CA, USA). The cDNA library was synthesized with a ZAP Express cDNA Synthesis Kit and ZAP Express cDNA Gigapack III Gold Cloning Kit (Stratagene, La Jolla, CA, USA) according to the manufacturer's instructions. Total library titers of each library from control and cold-shock conditions were 1.3×10^5 and 3.0×10^5 recombinant plaque-forming units, respectively. A total of 2,112 clones (1,056 clones from each of two culture conditions) were randomly chosen and the plasmids were subjected to sequencing. The cDNA amplification was carried out with the T7 primer and ABI Prism BigDye terminator cycle sequencing reaction mix (Applied Biosystems, Foster City, CA, USA). The amplified cDNAs were purified by Quadra 3 (Tomtec, Hamden, CT, USA). Sequencing was performed with an ABI3730 XL DNA analyzer (Applied Biosystems, Foster City, CA, USA). More than 92% and 94% had over 500 nucleotide base pairs in control and treatment conditions, respectively. The 2,042 EST sequences (1,016 from control condition and 1,026 from cold-shock condition) with average Phred scores [5] greater than 20 were subjected to subsequent clustering and assembling by TGICL (TIGR Gene Indices clustering tools) [17]. All of the sequences longer than 100 nucleotides were clustered by the CAP3 (Contig Assembling Program). Sequences greater than 94% identity over a region longer than 30 base pairs were clustered. As a result, 441 unique sequences (222 contigs and 219 singletons) and 591 unique sequences (200 contigs and 391 singletons) were obtained from control and cold-shock conditions, respectively (Table 1).

The complete EST sequences were deposited to the DDBJ EST database (http://www.ddbj.nig.ac.jp/index-e.html). The nucleotide sequences reported in this study are available in the DDBJ/EMBL/GenBank database of nucleotide sequences, with the accession numbers FS592765 to FS593780 for control condition database and those from FS593781 to FS594806 for cold-shock database.

Table 1. Summary of ESTs from the Antarctic Prasinophyte *P. gelidicola.*

	Control condition	Cold-shock condition
Number of EST clones	1,056	1,056
Number of passed reads	1,016	1,026
Number of contigs	222	200
Number of singletons	219	391
Number of ESTs of matching NR database	668	775
Number of ESTs of matching KOG database	429	522

Passed reads had ≥ 100 nucleotides.

Individual unique sequences were compared with nonredundant and KOG (orthologous groups for eukaryotic complete genomes; http://www.ncbi.nlm.nih.gov/COG/grace/shokog.cgi) databases from the BLASTX algorithm. ESTs and genomic sequences for *Arabidopsis thaliana* [8], *Chlamydomonas reinhardtii* [12], *Micromonas pusilla* CCMP1545 [23], *Micromonas* sp. [23], *Picea sitchensis* [18], *Ostreococcus lucimarinus* [16], and *Physcomitrella patens* [15] were compared with EST sequences of *P. gelidicola* dominantly. Functional classification was accomplished from KOG categories.

A total of 2,042 EST sequences from the two different conditions were compared with 8 nonredundant and KOG (orthologous groups for eukaryotic complete genomes) databases to identify genetic function and categorize functional protein domains. The EST sequence comparison was carried out using BLASTX [1] with a cut-off expectancy value of 10^{-4}. *Micromonas* sp. RCC299 sequences showed the highest similarity and those of *Arabidopsis thaliana* revealed the lowest similarity in control condition (Fig. 1A). The most abundant EST sequences were matched with *Micromonas* sp. RCC299 in cold-shock condition (Fig. 1B). Most of both EST sequences from control and cold-shock conditions showed high similarities to eukaryotic organisms from algae (57.0% and 54.1%), animal (5.5% and 3.9%), bacteria (2.6% and 5.3%), fungi (0.9% and 1.7%), and higher plants (34.0% and 35.0%). More than 80% if the ESTs from both conditions were relevant to algae and higher plants, mainly genera *Chlamydomonas*, *Micromonas*, *Ostreococcus*, and *Physcomitrella*. All EST sequences were then compared with the KOG database to classify the protein domains (Fig. 2).

S-Adenosyl homocysteine hydrolase (13 ESTs), *S*-adenosylmethionine synthetase (11 ESTs), heat shock protein 70 (8 ESTs), and stress-related chlorophyll *a/b*

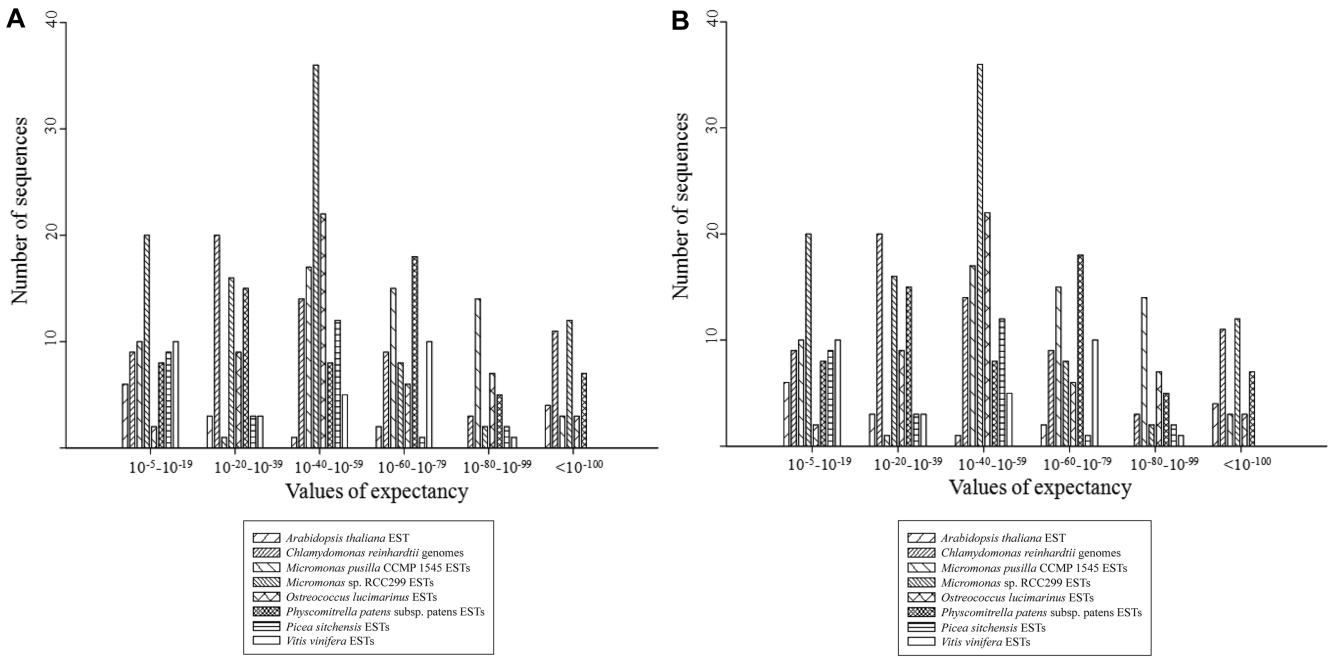

Fig. 1. Distribution of e-values from 8 nonredundant databases.
(**A**) Distribution of ESTs in control condition. (**B**) Distribution of ESTs in cold-shock condition.

binding protein 2 (7 ESTs) were dominantly counted in cold-shock condition (Table 2). The number of EST sequences encoding unknown (14 ESTs) was the highest in the cold-shock condition. Another unknown EST sequence (10 ESTs) and unnamed protein product (8 ESTs) were also counted dominantly. Further functional analysis of the unknown EST sequences in this report (Table 2) might provide new clues to understand the cold-adaptation mechanism of polar organisms. To date, Antarctic Prasinophyte green algae have been investigated in aspects of their importance

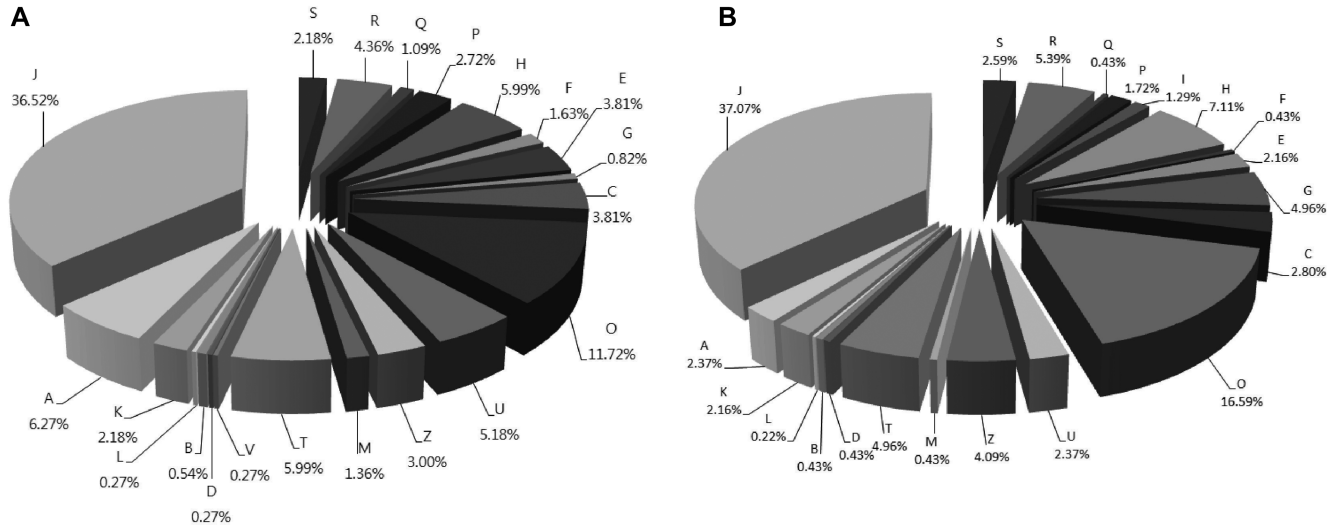

Fig. 2. KOG distributions of ESTs from control (**A**) and cold-shock (**B**) conditions.
Each number on the number of the categories indicates percentage against the total ESTs in the KOG database. The abbreviations for categories of genetic function are as follows: A, RNA processing and modification; B, Chromatin structure and dynamics; C, Energy production and conversion; D, Cell cycle control, cell division, and chromosome partitioning; E, Amino acid transport and metabolism; F, Nucleotide transport and metabolism; G, Carbohydrate transport and metabolism; H, Coenzyme transport and metabolism; I, Lipid transport and metabolism; J, Translation, ribosomal structure, and biogenesis; K, Transcription; L, Replication, recombination, and repair; M, Cell wall, membrane, and envelope biogenesis; O, Posttranslational modification, protein turnover, and chaperones; P, Inorganic ion transport and metabolism; Q, Secondary metabolites biosynthesis, transport, and metabolism; R, General function prediction only; S, Function unknown; T, Signal transduction mechanism; U, Intracellular trafficking, secretion, and vesicular transport; Z, Cytoskeleton.

Table 2. Most abundant ESTs of *P. gelidicola* in cold-shock condition compared with control condition.

Rank	Protein domains (Database entry number)	Number of EST reads in control condition	Number of EST reads in cold-shock condition
1	Unknown (ABK21901)	2	14
2	*S*-Adenosyl homocysteine hydrolase (XP_001693339)	1	13
3	*S*-Adenosylmethionine synthetase (XP_001696661)	6	11
4	Unknown (ABK92690)	1	10
5	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit (BAD42334)	11	9
	Tubulin alpha chain (Q9ZRJ4)	2	9
6	Unnamed protein product CAO61715)	0	8
	Heat shock protein 70, cytosolic (XP_001418384)	0	8
7	Chlorophyll *a/b* binding protein (AAB70556)	7	7
	Stress-related chlorophyll *a/b* binding protein 2 (XP_001696064)	0	7

as primary producers in polar ecosystems and ecological distributions, but studies related to cold-shock or cold acclimation mechanisms have been limited. Hence, analyses of *P. gelidicola* ESTs would provide the genetic information and preliminary expression profiles in low-temperature environments.

Acknowledgments

This work was supported by the National Agenda Project from The Korea Research Council of Fundamental Science & Technology (KRCF) and Korea Polar Research Institute (KOPRI) (Grant Nos. PG11010, PE11100, and PG12010).

References

1. Altschul, S. F., T. L. Madan, A. A. Schaffer, J. Zhang, J. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.
2. Bell, E. M. and J. Laybourn-Parry. 2003. Mixotrophy in the Antarctic phytoflagellate, *Pyramimonas gelidicola* (Chlorophyta: Prasinophyceae). *J. Phycol.* **39:** 644–649.
3. Burch, M. D. 1988. Annual cycle of phytoplankton in Ace Lake, an ice covered, saline meromictic lake. *Hydrobiologia* **165:** 59–75.
4. Cockell, C. S., M. D. Stokes, and K. E. Korsmeyer. 2000. Overwintering strategies of Antarctic organisms. *Environ. Rev.* **8:** 1–19.
5. Ewing, B., L. Hillier, M. C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8:** 175–185.
6. Jones, H. L. J. 1997. A classification of mixotrophic protists based on their behavior. *Freshwater Biol.* **37:** 35–43.
7. Jung, G. S., C. G. Lee, S. H. Kang, and E. S. Jin. 2007. Annotation and expression profile analysis of cDNAs from the Antarctic diatom *Chaetoceros neogracile*. *J. Microbiol. Biotechnol.* **17:** 1330–1337.
8. Kaneko, T., T. Katoh, S. Sato, A. Nakamura, E. Asamizu, and S. Tabata. 2000. Structural analysis of *Arabidopsis thaliana* chromosome 3. II. Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC and BAC clones. *DNA Res.* **7:** 217–221.
9. Kristiansen, E., S. A. Pedersen, and K. E. Zachariassen. 2008. Salt-induced enhancement of antifreeze protein activity: A salting-out effect. *Cryobiology* **57:** 122–129.
10. Madan, N. J., W. A. Marshall, and J. Laybourn-Parry. 2005. Virus and microbial loop dynamics over an annual cycle in three contrasting Antarctic lakes. *Freshwater Biol.* **50:** 1291–1300.
11. Maheswari, U., A. Montsant, J. Goll, S. Krishnasamy, K. R. Rajyashri, V. M. Patell, and C. Bowler. 2005. The diatom EST database. *Nucleic Acids Res.* **33:** 344–347.
12. Merchant, S. S., S. E. Prochnik, O. Vallon, E. H. Harris, S. J. Karpowicz, G. B. Witman, *et al.* 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* **318:** 245–250.
13. Mock, T., A. Krell, G. Glockner, U. Koluksaoglu, and K. Valentin. 2005. Analysis of expressed sequence tags (ESTs) from the polar diatom *Fragilariopsis cylindrus*. *J. Phycol.* **42:** 78–85.
14. Morgan-Kiss, R. M., J. C. Priscu, T. Pocock, L. Gudynaite-Savitch, and N. P. A. Huner. 2006. Adaptation and acclimation of photosynthetic microorganisms to permanently cold environments. *Microbiol. Mol. Biol. Rev.* **70:** 222–252.
15. Nishiyama, T., T. Fujita, T. Shin-I, M. Seki, H. Nishide, I. Uchiyama, *et al.* 2003. Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: Implication for land plant evolution. *Proc. Natl. Acad. Sci. USA* **100:** 8007–8012.
16. Palenik, B., J. Grimwood, A. Aerts, P. Rouze, A. Salamov, N. Putnam, *et al.* 2007. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc. Natl. Acad. Sci. USA* **104:** 7705–7710.
17. Pertea, G., X. Huang, F. Liang, V. Antonescu, R. Sultana, S. Karamycheva, *et al.* 2003. TIGR gene indices clustering tools (TGICL): A software system for fast clustering of large EST datasets. *Bioinformatics* **19:** 651–652.

18. Ralph, S. G., H. J. E. Chun, N. Kolosova, D. Cooper, C. Oddy, C. E. Ritland, *et al.* 2008. A conifer genomics resource of 200,000 spruce (*Picea* spp.) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs form Sitka spruce (*Picea sitchensis*). *BMC Genomics* **9:** 484.
19. Raymond, J. A. and C. A. Knight. 2003. Ice binding, recrystallization inhibition, and cryoprotective properties of ice-active substances associated with Antarctic sea ice diatoms. *Cryobiology* **46:** 174–181.
20. Teoh, M. L., W. L. Chu, H. Marchant, and S. M. Phang. 2004. Influence of culture temperature on the growth, biochemical composition and fatty acid profiles of six Antarctic microalgae. *J. Appl. Phycol.* **16:** 421–430.
21. Thomson, P. G., A. McMinn, I. Kiessling, M. Watson, and P. M. Goldsworthy. 2006. Composition and succession of dinoflagellates and chrysophytes in the upper fast ice of Davis Station, East Antarctica. *Polar Biol.* **29:** 337–345.
22. Wiencke, C. and I. Dieck. 1990. Temperature requirements for growth and survival of macroalgae from Antarctica and southern Chile. *Mar. Ecol. Progr. Ser.* **59:** 157–170.
23. Worden, A. Z., J. H. Lee, T. Mock, P. Rouze, M. P. Simmons, A. L. Aerts, *et al.* 2009. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* **324:** 268–272.