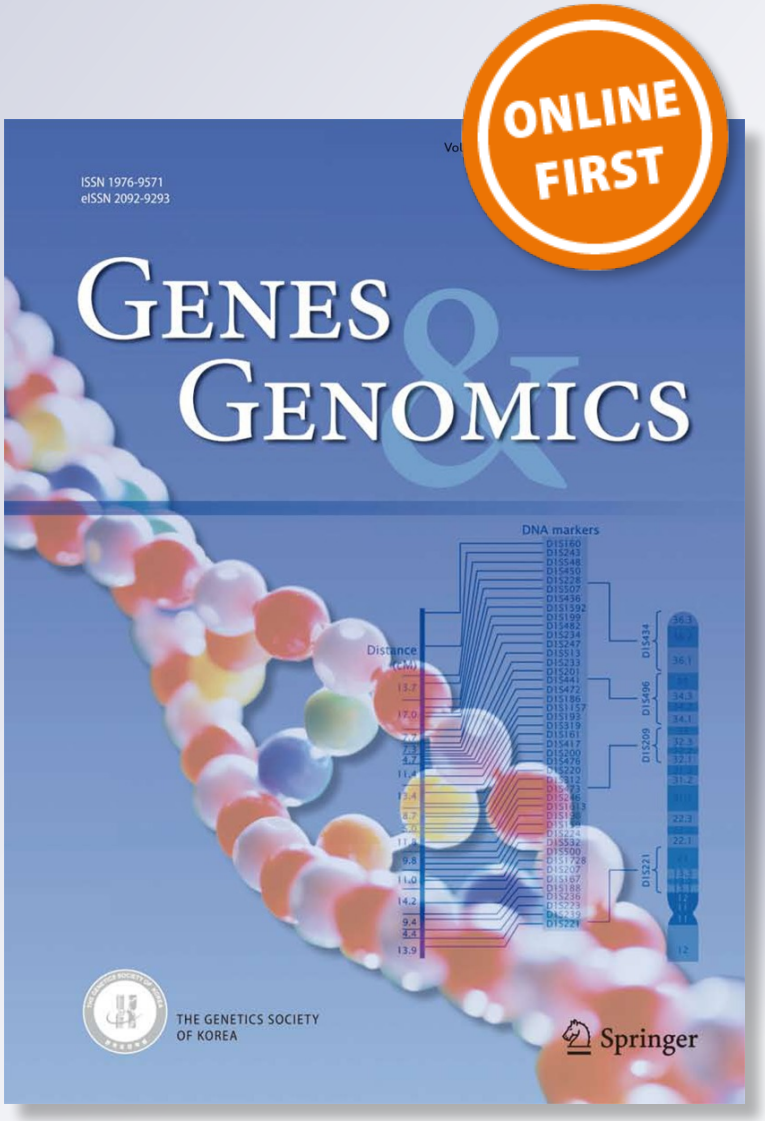
De novo transcriptome analysis of an Arctic microalga, Chlamydomonas *sp.*

Sanghee Kim, Min Jung Kim, Min Gui Jung, Sooyong Lee, Ye-Seul Baek, Sung-Ho Kang & Han-Gu Choi

Genes & Genomics

ISSN 1976-9571

Genes Genom
DOI 10.1007/s13258-013-0085-5



Springer

Genes Genom
DOI 10.1007/s13258-013-0085-5

RESEARCH ARTICLE

De novo transcriptome analysis of an Arctic microalga, *Chlamydomonas* sp.

Sanghee Kim · Min Jung Kim · Min Gui Jung · Sooyong Lee · Ye-Seul Baek · Sung-Ho Kang · Han-Gu Choi

Received: 28 August 2012 / Accepted: 24 September 2012

Abstract Microalgal *Chlamydomonas* species are abundant in polar regions and provide a novel model system for investigating the physiological characteristics and mechanisms related to multiple stresses. As part of an effort to understand the adaptation mechanism of polar organisms, we analyzed the transcriptome of an arctic microalgal strain, ArF0006, from the KOPRI Culture Collection of Polar Microorganisms (KCCPM). ArF0006 was isolated from ice-covered freshwater from the Arctic and, therefore, showed optimal growth at low temperatures. It was identified as a *Chlamydomonas* sp. based on morphology; this was confirmed by molecular phylogenetic analysis. A total of 64,712 reads were generated from pyrosequencing and assembled into 45,443 contigs and 19,424 singletons. Contigs ranged from 100–4,673 bp, with an average of 665 bp. Functional annotations of contigs with a high number of reads were related principally to cold response or photosynthesis. Also, several fatty acid and TGA biosynthesis-related genes and putative homologs of antifreeze protein were identified. A comparison of ArF0006 with *Chlamydomonas reinhardtii, Ostreococcus tauri*, and *Arabidopsis thaliana* was performed, and RT-PCR analysis verified the effect of growth temperature on gene expression pattern. Therefore, ArF0006 transcriptome analysis provided useful genetic information regarding cold tolerance that could be useful for bioindustrial applications.

Keywords Arctic *Chlamydomonas* sp. · ArF0006 · Pyrosequencing · Transcriptome · Cold adaptation · Fatty acid desaturase · Antifreeze protein

Introduction

Microorganisms are remarkably diverse and predominant in many Arctic and Antarctic habitats and represent the most abundant cold-adapted life-forms (Morgan-Kiss et al. 2006). As polar regions undergo dramatic territorial changes under the current warming climate, they provide valuable information for assessing the biological and biogeochemical impacts of environmental change. For example, much attention has been paid to polar-living phytoplankton because of their role in global warming as a carbon sink (Wohlers et al. 2009; Peck et al. 2009). Also, recent studies reported phytoplankton blooms underneath the Arctic sea ice caused by ice thinning (Arrigo et al., 2012). To understand how polar microorganisms thrive in this tough habitat by overcoming environmental pressures, it is necessary to investigate the physiological traits and genetic adaptations that render them tolerant to temperature extremes.

Cold generates various biochemical and molecular changes in living organisms. Examples include the evolution of cold-shock 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 (Kodama et al. 1995; Loppes et al. 1995; Suga et al. 2002; Morgan-Kiss et al., 2006; Hayward et al., 2007). Also, cold-adapted organisms have developed defense mechanisms, including prevention of ice crystallization, because these crystals can penetrate the cell membrane. Antifreeze proteins (AFPs) have been shown to bind directly to the ice surface, inhibit the growth of ice crystals,

Electronic supplementary material The online version of this article (doi:10.1007/s13258-013-0085-5) contains supplementary material, which is available to authorized users.

S. Kim · M. J. Kim · M. G. Jung · S. Lee · Y.-S. Baek · S.-H. Kang · H.-G. Choi (✉)
Division of Life Sciences, Korea Polar Research Institute, KORDI, Incheon 406-840, Korea
e-mail: hchoi82@kopri.re.kr

Published online: 29 January 2013

and reduce damage to frozen tissues; such proteins have been described only in species inhabiting cold environments (DeVries 1986; Storey and Storey 1996; Griffith et al. 1997; Tyshenko et al. 1997; Davies et al., 2002; Duman et al. 2004; Janech et al. 2006; Morgan-Kiss et al. 2006; Raymond et al. 2008, 2009; Young and Fletcher 2008; John et al. 2009; Gwak et al. 2010; Kiko 2010). A recent comparative study of extremophile and mesophilic *Chlamydomonas* suggested that phenotypic plasticity involves an acclimation process (Pocock et al. 2011). Most polar microorganisms are exposed not only to freezing temperatures but also to osmotic stress and desiccation (Morgan-Kiss et al. 2006). Therefore, they are good sources for the discovery of novel biochemical pathways and genes responding to environmental stress (van den Burg 2003; Laybourn-Parry and Pearce 2007).

To better understand molecular adaptations in polar species, we selected the Arctic freshwater microalgal strain ArF0006, *Chlamydomonas* sp., from the KOPRI Culture Collection of Polar Microorganisms (KCCPM). ArF0006 is advantageous for investigating the adaptation of microalgae to freezing environments: firstly, *Chlamydomonas reinhardtii* whole genome information is available in NCBI so the genome of ArF0006 could be compared to that of its mesophilic counterpart. Secondly, ArF0006 shows optimal growth at around 15 °C, whereas the mesophile *C. reinhardtii* exhibited maximum growth rates at 28 °C and is unable to grow below 10 °C (Vitova et al. 2011; our unpublished data).

In the present study, we report the transcriptome of ArF0006. This is to our knowledge the first report of genetic characterization of an arctic microalga of the genus *Chlamydomonas*. The data revealed the existence of stress-related and cold-adaptive genes, such as fatty acid desaturase, putative antifreeze genes, and a large number of ArF0006-specific genes.

Materials and methods

Algal collection, isolation, and culture conditions

ArF0006 was collected in a small pond near the Dasan Station located in NyÅlesund, Spitsbergen, Norway (78°55′N, 11°56′E) in 2005 using a 20-μm mesh and cells were deposited in the KOPRI Culture Collection of Polar Microorganisms (KCCPM). Single cells were manually isolated under a microscope and then cultured in Tris–acetate phosphate (TAP) buffer containing 17.5 mM acetate (Harris 1989) under a 16:8 h light: dark cycle at 4 °C. Cells were shaken at 250 rpm in glass flasks on an orbital shaker. Cells were cultured under various growth conditions and to different phases of the growth cycle. For the growth assay, 1 ml of cell suspension was filtered on Whatman glass microfiber filters GF/F (Whatman, Maidstone, Kent, UK). Each filter was incubated in 1 ml of 100 % dimethyl formamide (DMF) for 30 min at 4 °C and centrifuged at 13,000 rpm for 5 min. Chlorophyll levels in supernatants were measured using a Trilogy fluorometer (Turner Designs, Inc. Sunnyvale, CA, USA).

Molecular phylogeny

Genomic DNA was extracted using the DNeasy® Plant Mini Kit (Qiagen, Hilden, Germany). The nuclear SSU rDNA and chloroplastid *rbc*L were amplified from total genomic DNA using PCR and primer combinations described by Hoham et al. (2002) for SSU rRNA and new primer combinations (rbcLF1, 5′-GCWGGKTTYAAAGCWGGTGT-3′; rbcLF2, 5′-CGTGGTGGWYTWGACTTYAC-3′; rbcLR1, 5′-TCRTCTCCACCTTCACGAGC-3′; rbcLR2, 5′-CGGTCWCKCCAACGCATGAA-3′) for *rbc*L. PCR products were sequenced using the BigDye™ terminator cycle sequencing ready-reaction kit (PE Applied Biosytems [ABI], Foster City, CA, USA); sequence data were collected using an ABI PRISM 3730 DNA Analyzer.

Final alignments for the SSU and *rbc*L data consisted of 42 and 19 taxa, respectively. Ambiguously aligned regions were manually removed from the alignments (Hoham et al. 2002) and 1,738 (SSU) and 1,130 (*rbc*L) base pairs (bp) were used for phylogenetic inference. Distance, maximum-likelihood, and parsimony analyses were completed in PAUP* 4.0*b*10 (Swofford 2002). Modeltest 3.7 (Posada and Crandall, 1998; Posada and Buckley 2004) was used to determine whether the general time reversible (GTR) model with a gamma correction for among-site rate variation (Γ) and invariant sites (I) was an appropriate model for our data and was applied for maximum-likelihood analysis. A distance tree was obtained using the neighbor-joining (Saitou and Nei 1987) method and subjected to 2000 rounds of bootstrap resampling (Felsenstein 1985). Unweighted parsimony and maximum-likelihood analyses (gaps treated as missing data) were completed under a heuristic search (starting trees generated from 100 and 10 random sequence additions, respectively) with TBR branch swapping in effect. The robustness of internal nodes in the parsimony tree was estimated with 1000 replicates of bootstrap resampling (10 random addition replicates per bootstrap replicate).

SSU and *rbc*L data trees were also generated using MrBayes 3.0b4 (Huelsenbeck and Ronquist, 2001). The GTR+Γ+I model was used and 5,000,000 generations of four chains were run with sampling every 100 generations (burn-in subsequently identified at 29,000 and 22,600 generations for SSU and *rbc*L data, respectively).

RNA isolation, library preparation, and sequencing

Cells were harvested by centrifugation at 5,000× *g* for 10 min at 4 °C. The supernatant was discarded and cell

pellets were immediately frozen in liquid nitrogen and stored at −80 °C until further analysis. Total RNA from the frozen cells was purified using the RNeasy Plant Mini Kit (Qiagen, CA, USA) according to the manufacturer's instructions. cDNA was synthesized from poly(A)-enriched RNA samples and amplified by PCR to generate a cDNA library for Genome sequencer FLX Titanium (Roche/454 Life Science).

Assembly and Analysis

Sequencing and assembly were performed at the Macrogen Inc. Genomic Division (Macrogen, Republic of Korea) using the standard GS FLX protocol. The reads were assembled into contigs using the GS de novo assembler (NEWBLER, Roche/454 Life Sciences) and trimmed of low-quality and adapter sequences using the seqClean and Lucy software packages. A total of 64,710 reads were generated and assembled into 45,443 contigs and 19,424 singletons. Unigenes (contigs + singletons) were compared to the NCBI non-redundant (nr) database using the BLASTX algorithm with a cut-off e-value of $\leq 10^{-3}$. The GO terms were assigned to unique genes after annotation to the homologs in the UniProt database. The amino acids of putative AFPs were predicted from the DNA sequences and aligned using the Vector NTI program.

RT-PCR and RACE- PCR

cDNA was synthesized from 1 μg of RNA extracted from samples grown at 4 and 25 °C. Primers were designed according to the sequences of the contigs and singletons, and RT-PCR was performed using EX-Taq polymerase (Takara, Republic of Korea). To confirm the full lengths of the putative AFPs contig00093, 02042, and 01058, 3′- and 5′-end RACE PCR was carried out using gene-specific primers designed based on the partial sequences of each gene (Supplemental Table 1).

Results and Discussion

Phylogenetic relationships of ArF0006

The SSU rDNA and *rbc*L sequences of the Arctic green alga, *Chlamydomonas* sp. (strain ArF0006), have been deposited in GenBank (XXXXX and XXXXX). The lengths of the SSU rDNA and *rbc*L sequences of ArF0006 were 1,758 and 1,300 bp, respectively. They share 99.6 % (1,691/1,698) and 98.3 % (1,109/1,128) identities with those of *Chlamydomonas asymmetrica* (SSU, U70788; *rbc*L, EF113420). Final alignments for the SSU and *rbc*L data consisted of 42 and 19 taxa, respectively. Ambiguously aligned regions were manually removed from the alignments (Hoham et al., 2002) and 1,738 (SSU) and 1,130 (*rbc*L) base pairs (bp) were used for phylogenetic inference. ArF0006 was closely related to *C. asymmetrica* and relatively distant from *C. reinhardtii*. ArF0006 was grouped together within the clade including *C. asymmetrica*, *C. maxima*, and *Hormotila blennista* with strong-to-moderate support in the SSU tree (Fig. 1 and Supplemental Fig. 1). It was also related to the clade consisting of *C. asymmetrica*, *Chlamydomonas* sp. ICE-L, *Chlamydomonas* sp. ICE-W, and *Chloromonas* sp. ANT1 with moderate-to-weak support in the *rbc*L tree (Fig. 1).

Effect of temperature on growth rate

ArF0006 was able to grow over a broad temperature range (4–20 °C), and showed maximum growth at 15 °C (Fig. 2). ArF0006 was unable to grow at 25 °C, the temperature at which mesophilic *Chlamydomonas reinhardtii* showed maximum growth (Fig. 2). In contrast, Arf0006 was able to survive in frozen media for weeks (data not shown). Its physiological ability to survive at freezing temperatures reflects its successful adaptation to the Arctic habitat.

Highly expressed sequences and expression profiling

To determine the ArF0006 transcriptome that is responsible for its unique adaptation to the arctic environment, a whole cDNA library of ArF0006 grown at 4 °C was sequenced. A total of 64,712 reads were generated, which were assembled into 45,443 contigs and 19,424 singletons (Table 1). Contigs were 100–4,673 bp in length, with an average of 665 bp. A total of 1,922 (80 %) contigs were assembled with ≥20 reads, 2,387 contigs were assembled with ≥5 reads and three contigs (01809, 02062, and 00595) were assembled with ≥400 reads. The longest contig (00317) was 4,673 bp, aligned by 159 reads.

Contigs with a high number of reads (>100 reads) contained putative genes that were categorized to previously known stress-responsive genes based on a BLAST similarity search (Supplemental Table 2.). For example, the expression levels of homologous genes encoding the ABC transporter (contig00317), fatty acid desaturase (contig00093), a leucine-rich repeat containing protein (contig00076), and HSP70 (contig00063) have been reported to increase in various organisms when exposed to cold (Mock et al. 2006; Lee et al. 2008). Also, several ribosomal proteins and elongation factors known to be responsible for protein synthesis (e.g., contig00426, 00036, 00048, and contig 02062, etc.) were identified, supporting active metabolism at 4 °C. Ribosomal proteins have been reported to play roles as cellular sensors of temperature for the heat- and cold-shock response networks

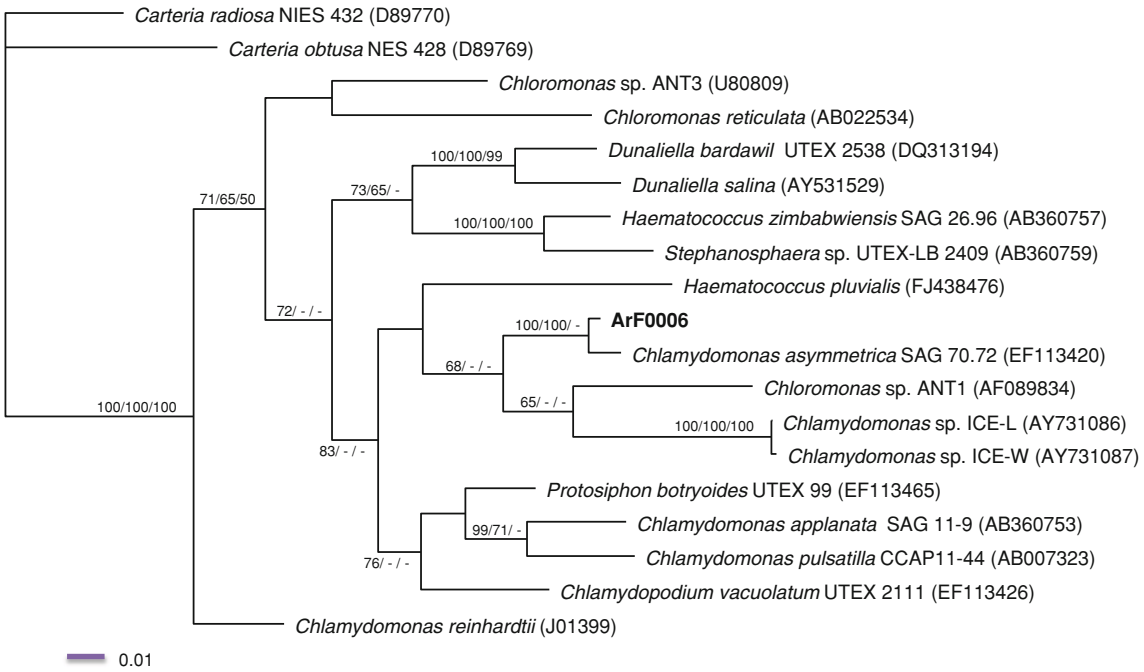


Fig. 1 *rbc*L alignment tree constructed by Bayesian inference. Values at branches represent Bayesian posterior probabilities (*left value*) and 2000 and 1000 bootstrap replicates for each neighbor-joining distance and parsimony (*middle and right values*, respectively). Branches lacking values received <50 % support

(VanBogelen and Neidhardt 1990). We also observed many typical stress-responsive genes, such as heat-shock proteins (two contigs and 22 singletons) and genes related to lipid biosynthesis, in the ArF0006 transcriptome. Transcriptome analysis of Antarctic krill also revealed that the most highly expressed genes were related to lipids, metabolism, or chaperones such as HSPs (Clark et al. 2011). RT-PCR analysis was performed to determine the effect of growth temperature on the expression of the above-mentioned genes (Supplemental Fig. 2).

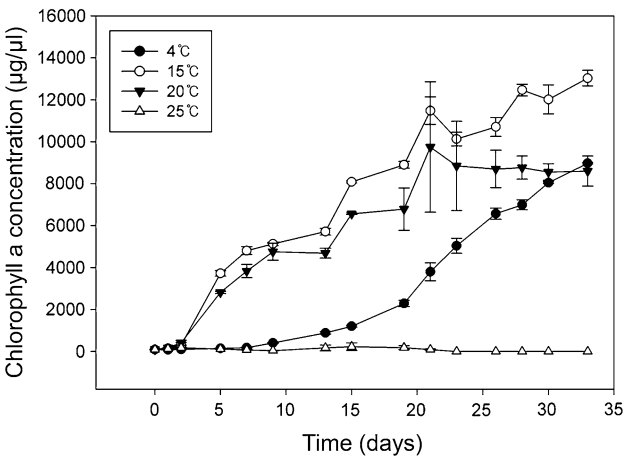


Fig. 2 Effect of temperature on the growth rate of ArF0006. ArF0006 was grown at 4, 15, 20, and 25 °C under a 18/6 h (light/dark) photoperiod. *X*-axis shows days and *Y*-axis shows chlorophyll-a contents. Error bars represent standard deviations of three biological replicates

Genome annotation and comparison

The 45,443 reads assembled as contigs were blasted against the NCBI protein database using BLASTX to identify similar sequences with an e-value cutoff $<10^{-3}$. A total of 6,835 reads (15 %) did not match any known sequence in GenBank, suggesting the existence of novel genes in ArF0006. Only 38,608 reads were assigned to gene ontology (GO) numbers. Figure 3 shows the percentage distributions of the three major GO categories. 'Cellular component' (41 %) was the predominant group, while 'biological process' and 'molecular function' constituted 22 % of the genes. Within the 'biological process' category, 29 % of the assignments corresponded to metabolic processes, followed by 'response to stimulus' (15 %) and 'cellular processes' (15 %). Within the 'cellular component' category, 'cell part' (65 %) was predominant, followed by 'organelle' (12 %) and 'unclassified' (11 %). Meanwhile, 'binding' (32 %), 'structural molecular activity' (25 %), and 'catalytic activity' (24 %) were represented relatively equally within the 'molecular function' category (Fig. 3).

To identify differences and overlap between ArF0006 genes and public databases, 21,801 unigenes newly obtained from ArF0006 were compared against three

Table 1 Sequences and assembly of 454 HQ reads and contigs

		Sequence (No.)	Bases (bp)
Sequencing (read)	High-quality (HQ) reads	64,712	22,867,847
	Average HQ read length	353 bp	
	Reads used in assembly	64,710	22,662,556
Contig	Reads assembled as contigs	45,443	
	Number of contigs	2,377	1,580,438
	Average length of contigs	665 bp	
	Range of contig length	100–4673 bp	
	Depth on contigs	19.12	
Singleton	Number of singletons	19,424	6,437,109
	Average length of singletons	331 bp	
	Range of singleton length	50–1,114 bp	
Unique sequences (contigs + singletons)		21,801	

reference data sets for selected species: *Chlamydomonas reinhardtii* (microalga), *Ostreococcus tauri* (ancestral microalga), and *Arabidopsis thaliana* (land plant) using BLASTX with a cut-off e-value of $\leq 10^{-3}$. Chlorophytes (*Chlamydomonas* and *Ostreococcus*) and Streptophytes (*Arabidopsis*) diverged from a common green plant ancestor; thus, they are expected to exhibit marked genomic information commonality. Of the ArF0006 proteins, 6,967 (35 %) were determined to be homologous to at least one of the genome references and 4,075 proteins (21 %) were common to all of these references (Fig. 4). Unexpectedly, only 6,282 (32 %) of ArF0006 unigenes were shared with *C. reinhardtii*, 5,391 (27 %) with *A. thaliana*, and 4,995 (26 %) with *O. tauri*, whereas in

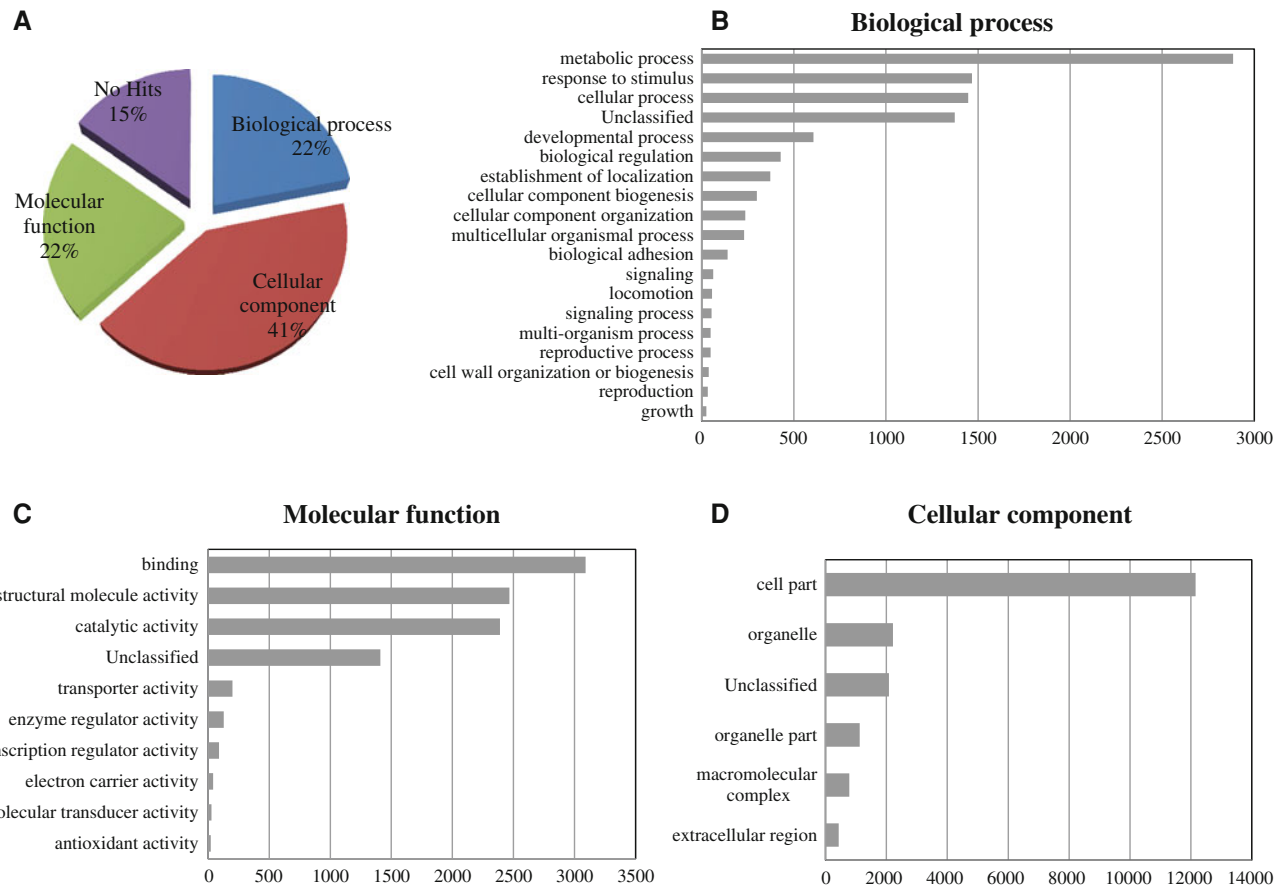


Fig. 3 Gene ontology assignments for the ArF0006 transcriptome. A *pie chart* shows proportional distribution of genes in each gene ontology category (**a**). The *graph* shows the percentage of genes observed in each category (**b–d**)

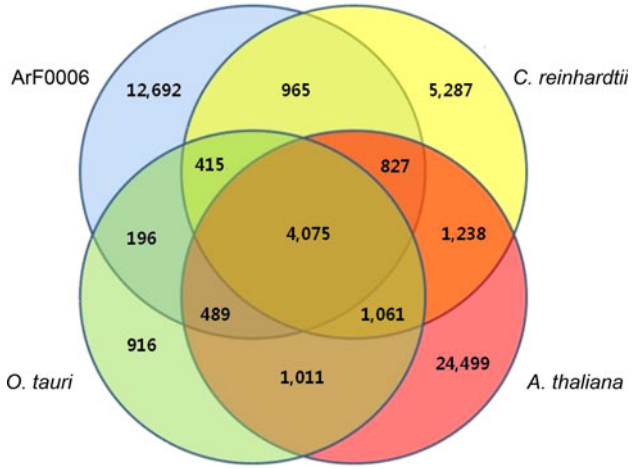


Fig. 4 Venn diagram showing genes unique and shared between and among four species (ArF0006, *Chlamydomonas reinhardtii*, *Ostreococcus tauri*, *Arabidopsis thaliana*)

C. reinhardtii, 7,201 (52 %) were shared with *A. thaliana* and 5,551 (40 %) with *O. tauri* (Fig. 4). Comparative analysis of the proteins and ESTs of *C. reinhardtii*, *O. tauri*, *A. thaliana* or in public databases indicated that a large number of ArF0006 genes have not yet been detected or did not show similarities to any other publicly available genes. Therefore, a large proportion of ArF0006-specific genes may be an evolutionary consequence, in concert with its physiological traits. Alternatively, they may represent a specialized set of genes that are expressed only at 4 °C. However, the latter explanation seems unlikely because the optimal growth temperature of ArF0006 was identical to that of *C. reinhardtii, A. thaliana*, and *O. tauri*. Thus the major of ESTs of each species represent actively expressed, rather than cold-dependent, genes.

Genes related to fatty acid and TGA biosynthesis

Due to their high lipid content, algae are considered a good source of biomass for a fossil fuel replacement. The potential of algae as a source of biomass to replace fossil fuels has been investigated, due to their high lipid content and the presence of unusual lipids (Guschina and Harwood, 2006; Rismani-Yazdi et al., 2011). Environmental factors, such as silicon deprivation and nutrient limitation, in the polar regions stimulate high rates of lipid synthesis in diatoms and other microalgae (Taguchi et al. 1987). Consistent with this, our data revealed that large numbers of genes were involved in the biosynthesis and catabolism of fatty acids, based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway assignment linked in the Blast2Go platform. For example, there were multiple copies of contig02131 (beta-ketoacyl-ACP synthase), 00116 (enoyl-ACP reductase), 00053 (acetyl-CoA acyltransferase), 02018 (acetyl CoA synthetase), and 00978 (enoyl CoA hydratase-like). With regard to singletons, C5SSA (stearoyl-ACP desaturase) and DHIP4 (acyl carrier protein thioesterase) were identified. In terms of genes involved in TGA biosynthesis, 00022 (glycerol kinase), 00745 (glycerol-3-phosphate O-acyltransferase), 00891 (triacylglycerol lipase), and several singletons were detected.

Table 2 Putative desaturases discovered in ArF0006

Desaturase	GO molecular function (species)
contig00039	Omega-3 fatty acid desaturase [*Chlorella vulgaris*]
contig00093	Plastid delta12 fatty acid desaturase [*Chlorella vulgaris*]
contig01298	Phytoene desaturase [*Dunaliella salina*]
contig01947	Phytoene desaturase [*Oncidium Gower Ramsey*]
GI1EB1N03C5SSA	Stearoyl-ACP desaturase [*Muriella zofingiensis*]
GI1EB1N03C6Y4Z	Zeta desaturase [*Tagetes erecta*]
GI1EB1N03C7TPK	Sterol 14 desaturase [*Chlamydomonas reinhardtii*]
GI1EB1N03C8Q9P	Delta-12 fatty acid desaturase [*Chlorella vulgaris*]
GI1EB1N03C98UZ	Endoplasmic reticulum 18:2 desaturase [*Populus tomentosa*]
GI1EB1N03DD0DQ	Phytoene desaturase [*Dunaliella salina*]
GI1EB1N03DGITX	Predicted protein [*Populus trichocarpa*]
GI1EB1N03DIC9 N	Delta-12 fatty acid desaturase [*Chlorella vulgaris*]
GI1EB1N03DMDW6	Zeta desaturase [*Tagetes erecta*]
GI1EB1N03DOCRH	Zeta-carotene desaturase [*Chlamydomonas reinhardtii*]
GI1EB1N03DQBKL	Microsomal delta12 fatty acid desaturase [*Chlorella vulgaris*]
GI1EB1N03DRMA6	Stearoyl-ACP desaturase [*Muriella zofingiensis*]
GI1EB1N03DRS9 J	Plastid stearoyl-acyl carrier protein desaturase [*Helicosporidium* sp.]
GI1EB1N03DTIBM	Fatty acid desaturase [*Chlamydomonas reinhardtii*]
GI1EB1N03DUKFH	Fatty acid desaturase [*Chlamydomonas reinhardtii*]

The genes encoding the candidate fatty acid desaturase are shown in Table 2. Of them, contig00093 (142 reads) was selected for further study because its transcript was detected at high levels at 4 °C. The full-length cDNA of contig00093 was confirmed by 5′ and 3′ end RACE analysis. Protein sequences showed high similarity (94 %) to the plastid delta 12 fatty acid desaturase of *Chlorella vulgaris* (ADB03432) in BLASTX matches. In contrast, the partial cDNA of contig00039 matched the omega-3 fatty acid desaturase of *Chlorella vulgaris* (BAB78717) with 79 % identity. It is interesting that putative desaturases of

Table 3 Antifreeze proteins discovered in ArF0006

Cluster ID	GO molecular function (species)
contig00914	Antifreeze glycopeptide AFGP polyprotein precursor [*Boreogadus saida*]
contig01058	Antifreeze protein [*Chlorella vulgaris*]
contig01488	Antifreeze protein [*Chlorella vulgaris*]
contig01832	Antifreeze protein [*Chlorella vulgaris*]
contig02402	Antifreeze protein [*Chlorella vulgaris*]
GI1EB1N03C2D3S	Antifreeze peptide 4 precursor, putative [*Brugia malayi*]
GI1EB1N03C863E	Antifreeze glycoprotein [*Boreogadus saida*]
GI1EB1N03DERF6	Antifreeze glycoprotein [*Boreogadus saida*]
GI1EB1N03DHQ4C	Antifreeze peptide 4 precursor, putative [*Brugia malayi*]
GI1EB1N03DR4YZ	Antifreeze glycoprotein [*Boreogadus saida*]

ArF0006 matched more closely with those of *Chlorella* than those of *Chlamydomonas*.

Putative antifreeze proteins

Unlike the sea, Arctic freshwater tends to be isolated both geologically and ecologically, with a low chance of invasion or disturbance. Thus, the habitat restriction of the arctic freshwater ecosystem may represent a selective pressure on microorganisms in terms of acquisition of unique stress-tolerance mechanisms over long periods of time. We found that at least five contigs (00914, 01058, 01488, 01832, and 02402) and five singletons (C2D3S, C863E, DERF6, DHQ4C, and DR4YZ) were categorized as AFP candidates based on similarity searches in other organisms (Table 3). Contig02402 and 01058 matched *Chlorella* ABR01234 with a 38 % identity, indicating a low sequence similarity, which matched only the Antarctic ecotype *Chlorella vulgaris* (Fig. 5). Meanwhile, 02402 and 01058 were 81 % similar when the extended amino acid sequences of contig01058 were excluded, suggesting that they are likely isoforms (Fig. 5). Also, the transcript levels of both genes were increased at low temperatures (Supplemental Fig. 2). Partial sequences of contigs01488 and 01832 matched that of *Chlorella* ABR01229 with 49 and 54 % identities, respectively. Since their similarities were lower and these AFPs were not found in the Antarctic *Chlamydomonas* CCMP681, further investigation of their role as antifreeze agents is necessary.

The origin of AFP remains controversial. Recent papers suggested that copepod AFP was generated by horizontal gene transfer (HGT) (Kiko 2010). Meanwhile, Deng et al. (2010) reported that the type III AFPs in Antarctic zoarcid fish evolved from ancestor genes that obtained a new function and, therefore, provide evidence of gene duplication under EAC (escape from adaptive conflict). It may also be possible to explain the evolutionary process of AFPs by means of multiple mechanisms that satisfy the conditions relevant to each organism. Discovery of other novel AFPs in non-model psychrophilic species, such as ArF0006, would facilitate resolution of this issue.

Conclusion

We report here the transcriptome annotation of a non-model Arctic freshwater microalga, ArF0006, using 454 pyrosequencing. ArF0006 was relatively phylogenetically close to *Chlamydomonas* spp. The ArF0006 transcriptome contained a large number of genes related to the cold response; these are also found in other psychrophilic species but are not present in temperate species. Furthermore, a large number of ArF0006 sequences did match those in other organisms,

Fig. 5 Multiple amino acid alignments of ArF0006 AFP candidates, contig02402, and 01058 with *Chlorella* AFP ABR01234. *Black boxes* indicate sequences identical in all; *grey boxes* indicate sequences conserved in at least two positions

indicating the presence of ArF0006-specific genes. Since global climate change has already had a tremendous impact on the polar ecosystem, the transcriptome of the arctic microalga, ArF0006, provides valuable genetic information for ecological and evolutionary studies.

Acknowledgments We thank Dr. Mark E. Siddall, Dr. Jungeun Lee, and Dr. Gi-Sik Min for helpful comments on the manuscript. This work was supported by the Basic Research Program of the Korea Polar Research Institute (PE12030).

References

Arrigo KR, Perovich DK, Pickart RS, Brown ZW, van Dijken GL, Lowry KE, Mills MM, Palmer MA, Balch WM, Bahr F et al (2012) Massive phytoplankton blooms under Arctic sea ice. Science 336:1408

Clark MS, Thorne MAS, Toullec J-Y, Meng Y, Guan LL, Peck LS, Moore S (2011) Antarctic krill 454 pyrosequencing reveals chaperone and stress transcriptome. PLoS One 6:e15919

Davies PL, Baardsnes J, Kuiper MJ, Walker VK (2002) Structure and function of antifreeze proteins. Philos Trans R Soc Lond B Biol Sci 357:927–935

Deng C, Cheng CHC, Ye H, He X, Chen L (2010) Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. Proc Natl Acad Sci USA 107:21593–21598

DeVries AL (1986) Antifreeze glycopeptides and peptides: interactions with ice and water. Methods Enzymol 127:293–303

Duman JG, Bennett V, Sformo T, Hochstrasser R, Barnes BM (2004) Antifreeze proteins in Alaskan insects and spiders. J Insect Physiol 50:259–266

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Griffith M, Antikainen M, Hon W-C, Pihakaski-Maunsbach K, Yu XM, Chun JU, Yang DSC (1997) Antifreeze proteins in winter rye. Physioloigia Plantarum 100:327–332

Guschina IA, Harwood JL (2006) Lipids and lipid metabolism in eukaryotic algae. Prog Lipid Res 45:160–186

Gwak IG, Jung WS, Kim HJ, Kang SH, Jin E (2010) Antifreeze protein in Antarctic marine diatom, *Chaetoceros neogracile*. Mar Biotechnol 12:630–639

Harris EH (1989) The *Chlamydomonas* sourcebook. Academic, San Diego

Hayward SA, Murray PA, Gracey AY, Cossins AR (2007) Beyond the lipid hypothesis: mechanisms underlying phenotypic plasticity in inducible cold tolerance. Adv Exp Med Biol 594:132–142

Hoham RW, Bonome TA, Martin CW, Leebens-Mack JH (2002) A combined 18S rDNA *rbc*L phylogenetic analysis of *Chloromonas* and *Chlamydomonas* (*Chlorophyceae, Volvocales*) emphasizing snow and other cold-temperature habitats. J Phycol 38:1051–1064

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755

Janech MG, Krell A, Mock T, Kang JS, Raymond JA (2006) Ice-binding protiens from sea ice diatoms (*Bacillariophyceae*). J Phycol 42:410–416

John UP, Polotnianka RM, Sivakumaran KA, Chew O, Mackin L, Kuiper MJ, Talbot JP, Nugent GD, Mautord J, Schrauf GE et al (2009) Ice recrystallization inhibition proteins (IRIPs) and freeze tolerance in the cryophilic Antarctic hair grass *Deschampsia antarctica* E. Desv Plant Cell Environ 32:336–348

Kiko R (2010) Acquisition of freeze protection in a sea-ice crustacean through horizontal gene transfer? Polar Biol 33:543–556

Kodama H, Horiguchi G, Nishiuchi T, Nishimura M, Iba K (1995) Fatty Acid Desaturation during Chilling Acclimation Is One of the Factors Involved in Conferring Low-Temperature Tolerance to Young Tobacco Leaves. Plant Physiol 107:1177–1185

Laybourn-Parry J, Pearce DA (2007) The biodiversity and ecology of Antarctic lakes: models for evolution. Philos Trans R Soc Lond B Biol Sci 362:2273–2289

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

Loppes R, Devos N, Willem S, Barthélemy P, Matagne RF (1995) Effect of temperature on two enzymes from a psychrophilic *Chloromonas* (Chlorophyta). J Phycol 32:276–278

Mock T, Krell A, Glöckner G, Kolukisaoglu Ü, Valentin K (2006) Analysis of expressed sequence tags (ESTs) from the polar diatom *Fragilariopsis cylindrus*. J Phycol 42:78–85

Morgan-Kiss R, 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

Peck LS, Barnes DKA, Cook AJ, Fleming A, Clarke A (2009) Negative feedback in the cold: ice retreat produces new carbon sinks in Antarctica. Glob Change Biol 9:2614–2623

Pocock T, Vetterli A, Falk S (2011) Evidence for phenotypic plasticity in the Antarctic extremophile *Chlamydomonas* raudensis Ettl. UWO 241. J Exp Bot 62:1169–1177

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of the AIC and Bayesian approaches over likelihood ratio tests. Syst Biol 53:793–808

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818

Raymond JA, Christner BC, Schuster SC (2008) A bacterial ice-binding protein from the Vostok ice core. Extremophiles 12:713–717

Raymond JA, Janech MG, Fritsen CH (2009) Novel ice-binding proteins from a psychrophilic antarctic alga (*Chlamydomonadaceae*, *Chlorophyceae*). J Phycol 45:130–136

Rismani-Yazdi H, Haznedaroglu BZ, Bibby K, Peccia J (2011) Transcriptome sequencing and annotation of the microalgae *Dunaliella tertiolecta*: pathway description and gene discovery for production of next-generation biofuels. BMC Genomics 12:148

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Storey KB, Storey JM (1996) Natural freezing survival in animals. Annu Rev Ecol Syst 27:365–386

Suga K, Honjoh K, Furuya N, Shimizu H, Nishi K, Shinohara F, Hirabaru Y, Maruyama I, Miyamoto T, Hatano S et al (2002) Two low-temperature-inducible *Chlorella* genes for delta12 and omega-3 fatty acid desaturase (FAD): isolation of delta12 and omega-3 fad cDNA clones, expression of delta12 fad in *Saccharomyces cerevisiae*, and expression of omega-3 fad in *Nicotiana tabacum*. Biosci Biotechnol Biochem 66:1314–1327

Swofford DL (2002) PAUP*: phylogenetic analysis using parsimony (*and other methods), Version 4.0b10. Sinauer, Sunderland

Taguchi S, Hirata JA, Laws EA (1987) Silicate deficiency and lipid synthesis of marine diatoms. J Phycol 23:260–267

Tyshenko MG, Doucet D, Davies PL, Walker VK (1997) The antifreeze potential of the spruce budworm thermal hysteresis protein. Nat Biotech 15:887–890

van den Burg B (2003) Extremophiles as a source for novel enzymes. Curr Opin Microbiol 6:213–218

VanBogelen RA, Neidhardt FC (1990) Ribosomes as sensors of heat and cold shock in *Escherichia coli*. Proc Natl Acad Sci USA 87:5589–5593

Vitova M, Bisova K, Hlavova M, Kawano S, Zachleder V, Cizkova M (2011) *Chlamydomonas reinhardtii*: duration of its cell cycle and phases at growth rates affected by temperature. Planta 234: 599–608

Wohlers J, Engel A, Zöllner E, Breithaupt P, Jürgens K, Hoppe H-G, Sommer U, Riebesell U (2009) Changes in biogenic carbon flow in response to sea surface warming. Proc Natl Acad Sci USA 106:7067–7072

Young HM, Fletcher GL (2008) Antifreeze protein gene expression in winter flounder pre-hatch embryos: implications for cryopreservation. Cryobiology 57:84–90