J. Microbiol. Biotechnol. (2005), **15**(5), 1060–1066

JOURNAL OF MICROBIOLOGY AND BIOTECHNOLOGY

© The Korean Society for Microbiology and Biotechnology

Gene Expression Profiling of Eukaryotic Microalga, *Haematococcus pluvialis*

EOM, HYUNSUK[1], SEUNGHYE PARK[2], CHOUL-GYUN LEE[1], AND EONSEON JIN[2]*

[1]*Department of Biotechnology, Department of Biological Engineering, Inha University, Incheon 402-751, Korea*
[2]*Department of Life Science, College of Natural Science, Hanyang University, Seoul 133-791, Korea*

Received: February 1, 2005
Accepted: March 24, 2005

Abstract Under environmental stress, such as strong irradiance or nitrogen deficiency, unicellular green algae of the genus *Haematococcus* accumulate secondary carotenoids, i.e. astaxanthin, in the cytosol. The induction and regulation of astaxanthin biosynthesis in microalgae has recently received considerable attention owing to the increasing use of secondary carotenoids as a source of pigmentation for fish aquacultures, and as a potential drug in cancer prevention as a free-radical quencher. Accordingly, this study generated expressed sequence tags (ESTs) from a library constructed from astaxanthin-induced *Haematococcus pluvialis*. Partial sequences were obtained from the 5' ends of 1,858 individual cDNAs, and then grouped into 1,025 non-overlapping sequences, among which 708 sequences were singletons, while the remainder fell into 317 clusters. Approximately 63% of the EST sequences showed similarity to previously described sequences in public databases. *H. pluvialis* was found to consist of a relatively high percentage of genes involved in genetic information processing (15%) and metabolism (11%), whereas a relatively low percentage of sequences was involved in the signal transduction (3%), structure (2%), and environmental information process (3%). In addition, a relatively large fraction of *H. pluvialis* sequences was classified as genes involved in photosynthesis (9%) and cellular process (9%). Based on this EST analysis, the full-length cDNA sequence for superoxide dismutase (SOD) of *H. pluvialis* was cloned, and the expression of this gene was investigated. The abundance of SOD changed substantially in response to different culture conditions, indicating the possible regulation of this gene in *H. pluvialis*.

Key words: *Haematococcus pluvialis*, astaxanthin, expressed sequence tags (ESTs), superoxide dismutase

The unicellular green alga *Haematococcus pluvialis* (Volvocales) is known for its massive accumulation of the ketocarotenoid astaxanthin (3,3'-dihydroxy-β,β-carotene-4,4'-dione) in an esterified form in response to, for example, nutritional or light stress conditions [3, 6]. Large amounts of the astaxanthin are also produced in the green alga *Haematococcus pluvialis* under unfavorable culture conditions, such as oxidative stress induced by the addition of reactive oxygen species (ROS), drought, high salt, and high temperatures [7, 9, 12, reviewed in 16]. During massive production of astaxanthin in *H. pluvialis* 50 mg/g dry cell [18], the vegetative cells of the alga form cysts and change their color from green to red. Carotenoids are synthesized by all photosynthetic organisms, and play an important role as light-harvesting pigments and protect the photosynthetic apparatus from photooxidative damage under excess light conditions (reviewed in [2, 14]). The increasing commercial interest in astaxanthin results from its antioxidative properties [15, 18], which are important for the pharmaceutical and cosmetic industries. The combination of nitrate deficiency stress together with increased illumination resulted in cyst cell formation, high astaxanthin accumulation, and a strong induction of carotenoid biosynthesis genes. After the transfer of *Haematococcus* cells from low light condition to high light condition, a higher astaxanthin production and cyst cell formation have been observed [24].

Although there have been lots of studies focused on the commercial points, as the induction conditions for astaxanthin and mass culture conditions in *H. pluvialis* [8, 17, 19, 23, 24], the molecular mechanisms that occur in response to inducible conditions, which may lead to cellular transformation and produce astaxanthin, have not been well studied. Accordingly, this study profiled the genes expressed by *H. pluvialis* as green vegetative cells and the cells under astaxanthin induction conditions using an expressed sequence tag (EST) analysis. The advent of high-throughput sequencing of partial cDNAs, or expressed sequence tags (ESTs), provides relatively fast and cost-effective access to the gene expression profile of an organism [5, 13, 20, 22]. ESTs add information on the expressed part of the genome, while also representing a valuable tool in the identification

*Corresponding author
Phone: 82-2-2220-2561; Fax: 82-2-2299-3495;
E-mail: esjin@hanyang.ac.kr

of new genes and the construction of linkage maps [10, 11, 26]. An interesting application of EST sequencing is the study of the gene expression pattern in response to a given environmental stimulus; the composition of the mRNA population offers an overall view of the transcribed genes, thereby providing a novel tool for gene discovery and understanding of the biochemical pathways involved in physiological responses. EST analyses are especially useful when studying organisms for which little sequence data exist, or when sequencing the genome is either not planned or not easily feasible because of the genome size.

To date, there are little genomic data available on *Chlorophytes* (green algae), a group far more diverse and evolutionarily divergent than all land plants combined. Currently, more than 20 million ESTs can be found in the EST databases maintained by US National Center of Biotechnology Information (dbEST release 101805). Nonetheless, despite all the potential applications, ESTs from only a few algal species can be found in the EST databases (about 167,621 for *Chlamydomonas reinhardtii*, 20,979 for *Porphyra yezoensis*, 900 for *Laminaria digitata*, 200 for *Gracilaria gracilaris*, and a few other algae). Therefore, this paper provides a preliminary report on the initial phases of the *H. pluvialis* EST sequencing project. A cDNA library was constructed from *H. pluvialis,* then the size, occurrence, and quality of the recombinant cDNA inserts were analyzed. Most of the sequenced cDNAs (ESTs) showed a significant similarity to genes from other organisms, thus revealing possible functions for these *H. pluvialis* sequences.

MATERIALS AND METHODS

Algal Growth and Induction of Carotenoid Biosynthesis

The unicellular green alga *Haematococcus pluvialis* was purchased from UTEX, the Culture Collection of Algae at the University of Texas at Austin. *H. pluvialis* was cultivated in a modified Bold's Basal Medium (MBBM) [20], where the composition per liter consisted of $NaNO_3$, 246.5 mg; $CaCl_2{\cdot}2H_2O$, 24.99 mg; $MgSO_4{\cdot}7H_2O$, 73.95 mg; $FeSO_4{\cdot}7H_2O$, 4.98 mg; K_2HPO_4, 74.9 mg; KH_2PO_4, 175.57 mg; NaCl, 25.13 mg; $C_{10}H_{16}N_2O_8$ (EDTA), 49.68 mg; $CuSO_4{\cdot}5H_2O$, 1.57 mg; $Na_2MoO_4{\cdot}2H_2O$, 1.19 mg; H_3BO_3, 11.13 mg; $MnCl_2{\cdot}4H_2O$, 1.44 mg; $ZnSO_4{\cdot}7H_2O$, 8.83 mg; $Co(NO_3)_2{\cdot}6H_2O$, 0.49 mg; MoO_3, 6.06 mg; KOH, 30.86 mg; and H_2SO_4, 0.98 mg, in distilled water. The seed culture was grown in 2-l scale bubble column photobioreactors in MBBM at 25°C under continuous light intensity of 40 $\mu E{\cdot}m^{-2}{\cdot}s^{-1}$ at the column surface. To prepare various cell types, immature and mature cells of *H. pluvialis* were harvested, where the immature *H. pluvialis* cells were harvested during the exponentially growing period. To induce astaxanthin biosynthesis and mature cell formation, the cultures were inoculated at a cell density of $5{\times}10^5$ $cell{\cdot}ml^{-1}$ with cells in the exponentially growing period and induced in $NaNo_3$-free MBBM at 25°C under a continuous light intensity of 1,200 $\mu E{\cdot}m^{-2}{\cdot}s^{-1}$ at the column surface for 1 day.

cDNA Library Construction

The total RNA was extracted from the immature and mature cells using the TRI reagent (Molecular Research Center, Inc., Cincinnati, OH, U.S.A.). $Poly(A)^+$ RNA was then prepared using a Poly(A) Track mRNA isolation system (Promega, Madison, WI, U.S.A.). A cDNA library was constructed using a ZAP-cDNA synthesis kit and the ZAP-cDNA Gigapack III Gold packaging extract (Stratagene, La Jolla, CA, U.S.A.) according to the manufacturer's instructions. The total primary titer of each library in recombinant plaque-forming units was $6.5{\times}10^6$. After amplification, samples from each cDNA library were used to subclone inserts by mass excision for the conversion of Lambda to a phagemid vector. The resulting phagemid libraries were plated at low density on Luria Bertani agar plates containing kanamycin (25 $mg{\cdot}l^{-1}$).

Nucleotide Sequencing

Nineteen-hundred colonies were randomly picked and cultivated for storage of the colonies and isolation of the plasmid DNA. The plasmid DNA was purified from the *E. coli* cultures by alkaline lysis, vacuum filtration, and anion-exchange chromatography using a high-throughput, 96-well format system (Millipore, MA, U.S.A.). The amplified cDNAs were purified with GeneClean (BIO 101, Vista, CA, U.S.A.), and the 5'-end from each cDNA sequenced with the T3 primer and ABI Prism BigDye Terminator Cycle Sequencing reaction mix (PE Applied Biosystems, Foster City, CA, U.S.A.). The reactions were purified using Centri-Sep spin columns (Princeton Separations, Adelphia, NJ, U.S.A.), and the eluted extension products resolved and analyzed on an ABI 3100 Genetic Analyzer (PE Applied Biosystems).

Analysis of cDNAs and Functional Annotation

The raw DNA sequence data were edited to remove the vector sequences and poor quality data using a computer program (Lasergene software, DNASTAR, Inc., Madison, WI, U.S.A.). The edited EST sequences were then arranged into nucleotide-matched clusters using an alignment program (SeqMan in Lasergene, DNASTAR, Inc.) to determine the frequency of sampling redundancy. To assign the functions of the ESTs, the *H. pluvialis* sequences were aligned with the GenBank nucleotide sequence database using the BLASTX algorithm with an E-value cut-off at 10^{-10}, identities cut off at 40% and positives 50%. Based on the BLASTX comparison results, the *H. pluvialis* sequences were classified according to their predicted function. Genes with predicted or known functions were classified into 11

putative cellular roles based on the functional catalogues established at KEGG sites and modified by functional catalogues of plant genes [4].

Northern Analysis

For Northern blot analysis, the total RNA (10 μg) was denaturated with 50% formamide and 6.3% formaldehyde, and separated on a denaturing agarose gel. The RNA was then transferred onto nylon filters (Hybond TM-H+, Amersham Pharmacia Biotechnology) and fixed by UV cross-linking. The cDNA fragments were labeled with biotin based on a system (KPL, Gaithersburg, MD, U.S.A.) and used as a probe for the Northern blot hybridization. The membranes were hybridized at 60°C for 16 h and washed twice with 0.1×SSC/1% SDS at 60°C for 15 min. Finally, the hybridized bands were detected using the CDP-Star™ system (Tropix, Bedford, MA, U.S.A.).

RESULTS AND DISCUSSION

cDNA Library and EST Analyses

A titer analysis of the primary library indicated that greater than 3.25×10^6 plaque-forming units (pfu) were produced by packaging the vector/cDNA ligation mixture. Phage amplification of the primary library yielded approximately 50 ml of phage-containing SM buffer with a titer of about 4.0×10^9 pfu·ml^{-1}. The amplified secondary library was plated out with *E. coli* host cells, and a total of 20 plaques randomly selected from the library for a PCR and sequence analysis. The results of these quality evaluations are summarized in Table 1. A PCR amplification using the T3 and T7 primers that flank the cloning vector's polylinker demonstrated that 18 out 20 (90%) of the selected phage clones were recombinant, containing cDNA inserts that ranged in size from 0.5 to 2.0 kb; the average size for the 18 amplified cDNAs was 1.25 kb. Thereafter, the 5'-end sequences of the 1,865 cDNA clones were obtained, and sequencing provided 1,667 high quality ESTs. The average read length of these ESTs, after vector trimming and the removal of low quality sequences, was around 500 bp. Such large-scale sequencing generally provides highly redundant ESTs that can be aligned and assembled for a set of unique genes. The ESTs were assembled into nonredundant sequence groups. The clustering analysis generated 1,025 nonredundant sequences, comprising 317 groups of sequences and 708 singletons.

Table 1. Quality assessment of *H. pluvialis* cDNA library and ESTs sequencing.

Primary cDNA library titer	3.25×10^6 total pfu[a]
Secondary (amplified) cDNA library titer	4×10^9 pfu·ml^{-1}
Total phage clones selected and PCR amplified	20
Clones with cDNA insert	18 (90%)
Average insert size	1.25 kb[b] (0.5–2.0 kb)
Average EST length	460 bp[c] (315–510 bp)

[a]Plaque-forming units.
[b]Kilobases.
[c]Base pairs.

Table 2. Results of the similarity search with a public database. The number of EST groups and clones exhibiting a similarity to genes with known functions and hypothetical genes with no known function are given.

	Group	Clone
Gene with known function	567	1,158
Hypothetical	81	97
No similarity	377	431
Total	1,025	1,686

Identification of Gene Function by Database Search

To identify the putative functions, the 1,025 nonredundant EST groups were subjected to a similarity search against the nonredundant database, nr, provided by NCBI using the 1,025 nonredundant EST groups were found to be similar to genes registered in the databases, whereas 81 EST groups (8%) showed similarity to protein and/or DNA sequences with unknown functions. The remaining 377 EST groups (36%) showed no significant similarity to the sequences of registered genes, and were thus classified as novel sequences (Table 2).

The ESTs exhibiting an identity or high similarity to known genes were classified into eleven categories (Table 3) on the basis of the putative general functions of the protein encoded, as described previously [4]. A condensed version of the detained scheme is presented in Table 3, including the number of ESTs in each category. *H. pluvialis* was found to consist of a relatively high percentage of genes

Table 3. Classification of the ESTs according to putative function.

Catalogue	Number
Respiration	20
Photosynthesis	94
Metabolism	116
Genetic information processing	151
Defense or stress resistance	27
Signal transduction	26
Cellular process	95
Hypothetical	81
Others	18
No similarity	377
Structure	20
Total group	1,025

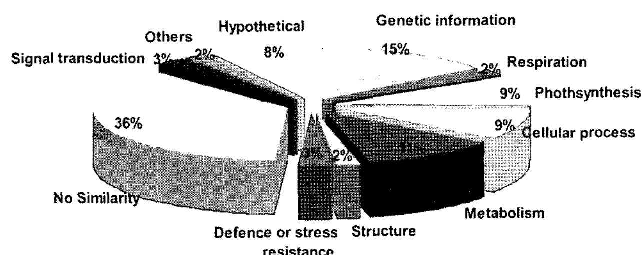

Fig. 1. Functional expression profile graph of *H. pluvialis*.

involved in genetic information processing (15%) and metabolism (11%), yet a relatively low percentage of sequences involved in the signal transduction (3%), defense or stress resistance (3%), and structure (2%). In addition, a relatively large fraction of *H. pluvialis* sequences was classified as genes involved in photosynthesis (9%) and cellular process (9%). These functional expression profiles are summarized in Fig. 1.

Comparison of *H. pluvialis* ESTs with Higher Plant and Unicellular Algal Genomes

The ESTs from *H. pluvialis* were also compared with other higher plants (*Arabidopsis*, *Oryzia sativa*, *Zea mays*, *Hordeum vulgare*, *Glycine max*) and unicellular algae (*Chlamydomonas reinhardtii*) at the protein level. The comparison of the *H. pluvialis* ESTs with the higher plant proteins revealed that 303 of the 1,025 ESTs that exhibited a significant similarity to proteins in the public databases also showed a significant similarity to the corresponding higher plant proteins (Table 4). When the ESTs from *H. pluvialis* were compared with ESTs from *C. reinhardtii* currently deposited in the GenBank to date, the BLASTX results revealed 193 matches out of the 1,025 *H. pluvialis* ESTs, indicating a significant similarity to the proteins in the public databases. Overall, the *H. pluvialis* EST sequences exhibited a higher percentage of similarity to the corresponding *C. reinhardtii* sequences than to the higher plant proteins. However, 31 ESTs showed a higher similarity to proteins from Cyanobacteria (Table 4).

Particular Genes of Interest in *H. pluvialis*

From the EST analysis, approximately 17% of the EST sequences revealed an open reading frame of cDNAs of putative functional genes (data not shown), among which particular genes of interest were cloned and characterized. Figure 2 shows the nucleotide and deduced amino acid sequences of the *H. pluvialis* SOD cDNA and α-tubuline. These sequences have been submitted to the GenBank under accession numbers AY878538 and AY894136, respectively. The nucleotide sequence of the SOD cDNA consisted of 972 bp with a 654 bp open reading frame (Fig. 2a), whereas the deduced amino acid sequence was 217 residues in length. Based on the similarity to other known SODs, the predicted size of the 217-amino acid final protein product was 23.7 kDa, calculated using the deduced amino acid content and specific masses of each amino acid (DNASTAR software). The *H. pluvialis* SOD amino acid sequence was aligned with the SOD sequences from other organisms, *Chlamydomonas reinhardtii* (GenBank accession number T08047), and *Arabidopsis thaliana* (AAM62550). The *H. pluvialis* SOD showed a high degree of sequence identity (72%) with the Mn-SOD from green alga, *C. reinhardtii*, and showed a 40% identity to the SOD from *A. thaliana*. Figure 2b also shows the nucleotide and deduced amino acid sequence for the *H. pluvialis* α-tubulin cDNA. The nucleotide sequence of the α-tubulin cDNA consisted of 1,593 bp with a 1,356 bp open reading frame, whereas the deduced amino acid sequence was 452 residues in length. Based on the similarity to other known α-tubulins, the predicted size of the 452-amino acid final protein product was 49.5 kDa, calculated using the deduced amino acid content and specific masses of each amino acid (DNASTAR software).

Under stress conditions, such as high light, the photosynthetic reduction of O_2 through the Mehler reaction results in the formation of free superoxide radicals. Under the present culture conditions, immature cells (40 $\mu E{\cdot}m^{-2}{\cdot}s^{-1}$) were irradiated with high light (1,200 $\mu E{\cdot}m^{-2}{\cdot}s^{-1}$) for 1 day to induce astaxanthin biosynthesis and mature cell formation. Although light is essential for photosynthetic organisms, the exposure of cells to excessive light leads to inactivation of the photosynthetic functions and the production of reactive oxygen species (ROS), such as hydrogen peroxide (H_2O_2), superoxide ($^{\bullet}O_2^-$) hydroxyl radicals, and singlet oxygen ($^1O_2$) [1, 21]. However, living organisms have evolved antioxidant defense mechanisms to remove excess ROS, where antioxidant enzymes, such as superoxide dismutase enzymes (SODs), remove $^{\bullet}O_2^-$ by accelerating its conversion to H_2O_2. In the green alga, *Dunaliella bardawil*, it has been

Table 4. Comparisons of *H. pluvialis* ESTs with genome of higher plants and ESTs of *C. reinhardtii*.

	H. pluvialis	Unicellular green algae	Higher plant	Cyano bacteria	Other eukaryote	Other prokaryote	Archaea
Gene with known function	567	189	261	26	52	35	3
Hypothetical	81	4	42	5	24	6	0
No. similarity				377			

(A) Mn-SOD

```
1     ACG AGG AAC GAG AGC AGC AGA TCT TAC ACG ATG GCA TTC ACC CTG CCC GCG CTG CCC TAC   60
                                              M   A   F   T   L   P   A   L   P   Y    10
61    GCC TTC GAC GCT CTA GAG CCT TTT GTG GAT GCC ACC ACG ATG CAG ATC CAC CAC GGC AAG   120
11     A   F   D   A   L   E   P   F   V   D   A   T   T   M   Q   I   H   H   G   K    30
121   CAT CAC CAG GCC TAC GTC ACA AAC CTG AAC AAC GCG CTG GAC AAG TTC CCG GAG CTG AAG   180
31     H   H   Q   A   Y   V   T   N   L   N   N   A   L   D   K   F   P   E   L   K    50
181   GAT TTG GGC ATC GTG GAC ATC AAC AAG AAG GTG GGC ACG GAC GCC ATT CCC AAG GAC ATC   240
51     D   L   G   I   V   D   I   N   K   K   V   G   T   D   A   I   P   K   D   I    70
241   GCG ACC GTG GTG CGC AAC AAT GGC GGG GGC CAC TGG AAC CAC AGC TTT TTC TGG AAG ATT   300
71     A   T   V   V   R   N   N   G   G   G   H   W   N   H   S   F   F   W   K   I    90
301   ATG ACC GCA CCC ACC AAC TCG AAC GGC CCG TCC GAC GAG CTC AAG GCC GCG GTT GAT GCG   360
91     M   T   A   P   T   N   S   N   G   P   S   D   E   L   K   A   A   V   D   A    110
361   TCG TTC GGC AGC ATG GAC GTG ATG AAG GAG AAG TTC AAC GCG GCA GCA GCC GGG CGA TTT   420
111    S   F   G   S   M   D   V   M   K   E   K   F   N   A   A   A   A   G   R   F    130
421   GGC TCT GGG TGG GCG TGG CTG GGT GTG AAG GCG GAC GGC AGC CTG GGC ATC ACC TCC ACC   480
131    G   S   G   W   A   W   L   G   V   K   A   D   G   S   L   G   I   T   S   T    150
481   CCC AAC CAG GAC AAC CCG CTG CAG GGC GTG GCC GAC GAG GCG CTG CTG CCT GTG CTG GGC   540
151    P   N   Q   D   N   P   L   Q   G   V   A   D   E   A   L   L   P   V   L   G    170
541   CTC GAC GTG TGG GAG CAT GCA TAC TAC CTG AAA TAC CAG AAC CGC CGC CCC GAG TAC ATC   600
171    L   D   V   W   E   H   A   Y   Y   L   K   Y   Q   N   R   R   P   E   Y   I    190
601   GCC GCG TTC TGG AAT ATT GTG AAC TGG GAG CAA GTG TCA GAG AAC TTC ACG GCG GCG AAA   660
191    A   A   F   W   N   I   V   N   W   E   Q   V   S   E   N   F   T   A   A   K    210
661   GCG GGC AGC CCG CCC GCG CTG TGA GAT TGC GTG AAG ATG AGG CCC ACC AGG CCG TGC TCT   720
211    A   G   S   P   P   A   L   *
721   AGG CTG CAG GCT TTG CTG CGA AAG GAT CAG GGA AGG TGT TTA GTT GAA CTG ATG GAT ATA   780
781   CAA AAG CTG GGC GTT GCT TTG CCT CTT CGC AGC AGT TGC GTT GGT GGT TAA ATT GTG GCG   840
841   GCT ACT TCA GCG TGC GGT GTG TGC ATT GCC ATT TCC TAC AGC GGT GTG CAG CTT CAA TAC   900
901   CAC ATT CTG TTT ATT GCA CCA CGT ACT GAG ATG CAT CTT TCG TTA CCT TTG GCC CAT AAA   960
961   AAA AAA AAA AAA   972
```

(B) α-Tubulin

```
1     CAC GAG GCC GGC AAG CAA CCC CAT TCC CCG TTC ATT CTC TCT CTT TAA ACA GCA CCC TTA   60
61    AAG ATG CGC GAG GCC ATT TCA ATC CAC TTG GGC CAA GCC GGT GTC CAG GTC GGC AAT GCC   120
           M   R   E   A   I   S   I   H   L   G   Q   A   G   V   Q   V   G   N   A    19
121   TGC TGG GAG CTT TAC TGC CTG GAG CAT GGC ATC CAG CCT GAT GGC CAG ATG CCC AGT GAC   180
20     C   W   E   L   Y   C   L   E   H   G   I   Q   P   D   G   Q   M   P   S   D    39
181   AAG ACC ATT GGC GGT GGC GAT GAT GCT TTC AAC ACC TTC TTC AGC GAG ACT GGT GCT GGC   240
40     K   T   I   G   G   G   D   D   A   F   N   T   F   F   S   E   T   G   A   G    59
241   AAG CAC GTG CCC CGT GCC ATC TTC CTG GAC CTG GAG CCC ACT GTG GTG GAT GAG GTC CGC   300
60     K   H   V   P   R   A   I   F   L   D   L   E   P   T   V   V   D   E   V   R    79
301   ACC GGT ACC TAC CGC CAG CTG TTC CAC CCT GAG CAG CTG ATC TCC GGC AAG GAG GAC GCT   360
80     T   G   T   Y   R   Q   L   F   H   P   E   Q   L   I   S   G   K   E   D   A    99
361   GCC AAC AAC TTT GCT CGT GGC CAC TAC ACC ATT GGC AAG GAG ATC GTG GAC CTG GCC CTG   420
100    A   N   N   F   A   R   G   H   Y   T   I   G   K   E   I   V   D   L   A   L    119
421   GAC CGC ATC CGC AAG CTG GCT GAC AAC TGC ACA GGC CTG CAG GGC TTC CTG GTG TTC AAT   480
120    D   R   I   R   K   L   A   D   N   C   T   G   L   Q   G   F   L   V   F   N    139
481   GCT GTG GGT GGT GGT ACT GGC TCT GGC CTT GGC TCC CTG CTG CTG GAG CGG CTG AGT GTG   540
140    A   V   G   G   G   T   G   S   G   L   G   S   L   L   L   E   R   L   S   V    159
541   GAC TAC GGC AAG AAG TCC AAG CTG GGC TTC ACC ATC TAC CCC TCC CCC CAG GTG TCC AAC   600
160    D   Y   G   K   K   S   K   L   G   F   T   I   Y   P   S   P   Q   V   S   N    179
601   GCC GTG GTG GAG CCC TAC AAC TCC GTG CTG TCC ACC CAC TCC CTG CTG GAG CAC ACT GAC   660
180    A   V   V   E   P   Y   N   S   V   L   S   T   H   S   L   L   E   H   T   D    199
661   GTG GCC ATC ATG CTG GAC AAT GAG GCC ATC TAC GAC ATC TGC CGC CGC TCC CTG GAC ATT   720
200    V   A   I   M   L   D   N   E   A   I   Y   D   I   C   R   R   S   L   D   I    219
721   GAG CGC CCC ACC TAC ACC AAC CTG AAC CGC CTG ATC GCC CAG GTC ATC TCC TCC TTG ACT   780
220    E   R   P   T   Y   T   N   L   N   R   L   I   A   Q   V   I   S   S   L   T    239
781   GCC TCC CTG CGC TTT GAT GGC GCC CTG AAC GTG GAC ATC ACT GAG TTC CAG ACC AAC CTG   840
240    A   S   L   R   F   D   G   A   L   N   V   D   I   T   E   F   Q   T   N   L    259
841   GTG CCT TAC CCC CGC ATC CAC TTT GTG CTG TCC TCC TAT GCC CCC ATC ATC AGC GCT GAG   900
260    V   P   Y   P   R   I   H   F   V   L   S   S   Y   A   P   I   I   S   A   E    279
901   AAG GCC TAC CAC GAG CAG CTG TCC GTG GCC GAG ATC ACC AAC GCA GTG TTT GAG CCT GCC   960
280    K   A   Y   H   E   Q   L   S   V   A   E   I   T   N   A   V   F   E   P   A    299
961   AGC ATG ATG GTC AAG TGC GAC CCC CGC CAT GGC AAG TAC ATG GCC TGC TGC ATG ATG TAC   1020
300    S   M   M   V   K   C   D   P   R   H   G   K   Y   M   A   C   C   M   M   Y    319
1021  CGT GGC GAC GTG GTG CCC AAG GAT GTG AAT GCT GCC GTT GCC ACC ATC AAG ACC AAG CGC   1080
320    R   G   D   V   V   P   K   D   V   N   A   A   V   A   T   I   K   T   K   R    339
1081  ACC ATC CAG TTT GTG GAC TGG TGC CCC ACT GGC TTC AAG TGT GGC ATC AAC TAC CAG CCC   1140
340    T   I   Q   F   V   D   W   C   P   T   G   F   K   C   G   I   N   Y   Q   P    359
1141  CCC ACC GTG GTG CCT GGT GGC GAC CTG GCC AAG GTG CAG CGC GCC GTG TGC ATG ATC TCC   1200
360    P   T   V   V   P   G   G   D   L   A   K   V   Q   R   A   V   C   M   I   S    379
1201  AAC TCT ACC GCC ATT GGC GAG GTG TTC TCC CGC CTG GAC CAC AAG TTC GAC CTG ATG TAT   1260
380    N   S   T   A   I   G   E   V   F   S   R   L   D   H   K   F   D   L   M   Y    399
1261  GCC AAG CGT GCC TTT GTG CAC TGG TAT GTG GGT GAG GGC ATG GAG GAG GGC GAG TTC TCT   1320
400    A   K   R   A   F   V   H   W   Y   V   G   E   G   M   E   E   G   E   F   S    419
1321  GAG GCC CGT GAG GAC CTG GCG GCC CTG GAG AAG GAC TTT GAG GAG GTC GGC GCC GAG TCT   1380
420    E   A   R   E   D   L   A   A   L   E   K   D   F   E   E   V   G   A   E   S    439
1381  GCT GAG GGC GCC GGA GAA GGA GAG GGC GAG GAG TAC TAA ACA TGA CGC TCA GGC CCA CAG   1440
440    A   E   G   A   G   E   G   E   G   E   E   Y   *
1441  CTG GGG GAG CTA GAG GCC AGC CGT GGC AAG CTG AAC ATG AAT CTG AAC GCA AGC ATC TGG   1500
1501  GTA TCC TAG CCT GCA GTG AAG CTG TCT AAT CAG TGT TCC GGT GGA GGA GGT GCT AGT TCT   1560
1561  GTG CTG GTG CTG AAT TGC TGA GAC ATT GGC   1590
```

Fig. 2. Nucleotide and deduced amino acid sequences of full-length cDNA encoding SOD (A) and α-tubulin (B) of *H. pluvialis*.
The nucleotides and amino acids are numbered in the margins. The cDNA sequences are shown in lower case. The translated regions of the sequence are underlined and indicated using a single-letter amino acid code.

suggested that photosynthetically produced active oxygen species, O_2^- (and its products) and $^1O_2$, are involved in triggering β-carotene biosynthesis, and the massive amount of carotenoid accumulated then protects the photosynthetic apparatus against photooxidative stress [25].

Figure 3 shows the results of the Northern blot analysis of particular genes derived from EST analysis. The expression of the putative Mn-SOD and α-tubulin genes were investigated under immature and mature culture conditions (see Materials and Methods). The transcript levels of the Mn-SOD and α-tubulin gene from *H. pluvilais* were significantly higher in the cells under the immature culture conditions. This higher transcript level of the SOD gene under immature conditions when the astaxanthin content was low is also consistent with the results previously reported by Kobayashi *et al.* [18], where an SOD assay using *H. pluvialis* whole-cell and cell-free extract samples detected SOD activity in the vegetative cells, yet only minimal activity in the mature cyst cells of *H. pluvialis*. Recently, Wang *et al.* [27] introduced a proteomic approach to show that the expression of Mn-SOD was constitutively high in vegetative cells and progressively downregulated after the onset of stress, decreasing to one-tenth of the basal level after 72 h of stress. Therefore, the downregulation of the Mn-SOD gene

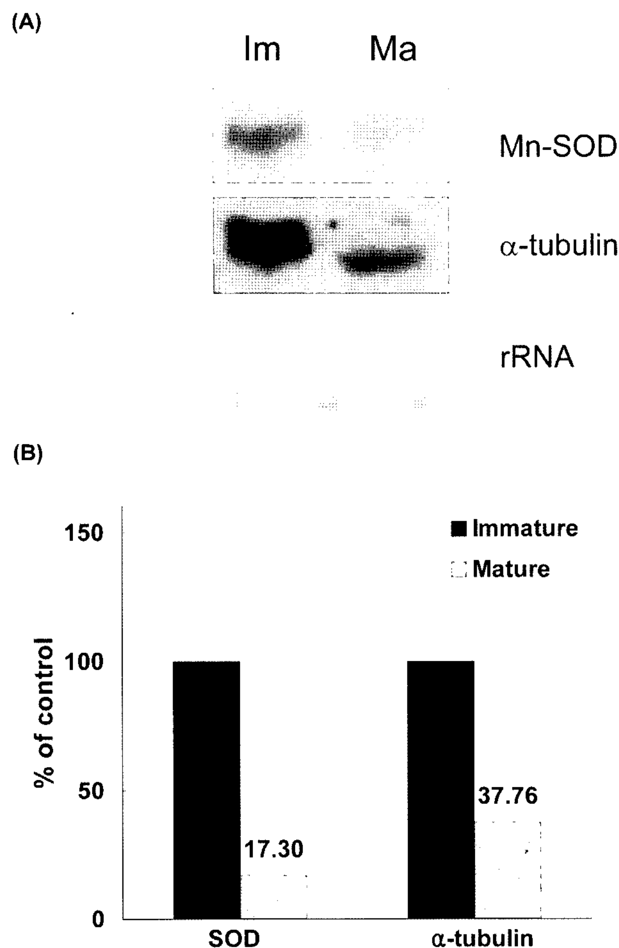

Fig. 3. Regulation of selected ESTs expression by high light stress.

(A) RNA blot analysis of the expression patterns of selected ESTs (homologous to Mn-SOD and α-tubulin gene sequences). The total RNA was isolated from immature (Im) condition (see Materials and Methods) grown and mature (Ma) condition grown *H. pluvialis*. The loading of an equal amount of total RNA in each lane was verified by ethidium bromide staining. (B) Quantification of RNA levels. The signal intensities were quantified by scanning and compared with the control value (Im), which was set as 100%.

in the mature cells in the present study was quite consistent with the results of previous studies [16, 18, 27], and strongly supports the hypothesis that *H. pluvialis* has two antioxidative mechanisms, antioxidative enzymes in vegetative cells and antioxidative astaxanthin in cyst cells, especially for the early events involved in the antioxidative defense process prior to the cellular accumulation of astaxanthin, and visible morphological and cellular structural changes.

Two of the current ESTs were identified as putative α-tubulin homologues. The tubulin protein is a very important eukaryotic cytoskeletal protein involved in maintaining the cell structure, coordinating intracellular movement, separating chromosomes during mitosis, and forming the backbone of the eukaryotic flagellum. The Northern blot analysis showed a remarkably reduced expression of α-tubulin when the cells entered the maturation stage (Fig. 3). Under carotenogenesis (maturation) conditions, the green flagellate cells, which could actively swim with two flagella, were gradually transform into spherical immotile cyst cells (mature cysts) with no flagella. Thus, since the immotile non-flagellate cysts of *H. pluvialis* did not need to form flagella, the expression of the α-tubulin gene was not apparently necessary in the red resting stage of the mature cells.

In summary, this study presented the first characterization of the ESTs from the commercially important green alga, *H. pluvialis*, which was found to consist of a relatively high percentage of genes involved with genetic information processing and metabolism, and a relatively low percentage of sequences involved in signal transduction, defense or stress resistance, and structure. In additon, a relatively large fraction of the *H. pluvialis* sequences was classified as genes involved in photosynthesis, structure and cellular process. A comparison of the *H. pluvialis* ESTs with the higher plant genome and *C. reinhardtii* ESTs revealed that most of the identified proteins were also present in both green plant and algae. However, a relatively large proportion of *H. pluvialis* (36%) was not identified, indicating that *H. pluvialis* also contains unique proteins. The putative Mn-SOD of *H. pluvialis* derived in the present study was cloned and an investigation of transcript levels, including the putative α-tubulin gene, revealed that these genes were significantly regulated under different culture conditions. Although the sequence data presented herein are limited, the availability of the ESTs in *H. pluvialis* identified in this study will still provide opportunities for gene mining and algal functional genomics.

Acknowledgment

This work was supported by grant No. KRF-2004-005-D00002 from Korea Research Fund (KRF).

REFERENCES

1. Apel, H. and H. Hirt. 2004 Reactive oxygen species: Metabolism, oxidative stress, and signal transduction. *Annu. Rev. Plant Biol.* **55:** 373–399.
2. Bartley, G. E. and P. A. Scolnik. 1995. Pigments for photoprotection, visual attraction, and human health. *Plant Cell* **7:** 1027–1038.
3. Ben Amotz, A. and M. Avron. 1983. On the factors which determine massive L-carotene accumulation in the halotolerant alga *Dunaliella bardawil*. *Plant Physiol.* **72:** 593–597.
4. Bevan, M. and The EU Arabidopsis Genome Project. 1998. Analysis of 1.9 Mb of contiguous sequence from

chromosome4 of *Arabidopsis thaliana*. *Nature* **391:** 485–488.

5. Boguski, M. S., T. M. Lowe, and C. M. Tolstoshev. 1993. dbEST-Database for "expressed sequence tags". *Nat. Genet.* **4:** 332–333.
6. Boussiba, S. 2000. Carotenogenesis in the green alga *Haematococcus pluvialis*: Cellular physiology and stress response. *Physiol. Plant* **108:** 111–117.
7. Boussiba, S. and A. Vonshak. 1991. Astaxanthin accumulation in the green alga *Haematococcus pluvialis*. *Plant Cell Physiol.* **32:** 1077–1082.
8. Boussiba, S., B. Wang, P. P. Yuan, A. Zarka, and F. Chen. 1999. Changes in pigments profile in the green alga *Haematococcus pluvialis* exposed to environmental stresses. *Biotechnol. Lett.* **21:** 601–604.
9. Boussiba, S., L. Fan, and A. Vonshak. 1992. Enhancement and determination of astaxanthin accumulation in green alga *Haematococcus pluvialis*. *Methods Enzymol.* **213:** 386–391.
10. Delseny, M., R. Cooke, M. Raynal, and F. Grellet. 1997. The *Arabidopsis thaliana* cDNA sequencing projects. *FEBS Lett.* **403:** 221–224.
11. Endo, M., T. Kokubun, Y. Takahata, A. Higashitani, S. Tabata, and M. Watanabe. 2000. Analysis of expressed sequence tags of flower buds in *Lotus japonicus*. *DNA Res.* **7:** 213–216.
12. Harker, M., A. J. Tsavalos, and A. J. Young. 1996. Autotrophic growth and carotenoid production of *Haematococcus pluvialis* in a 30 l air-lift photobioreactor. *J. Ferment. Bioeng.* **82:** 113–118.
13. Henry, I. M., M. D. Wilkinson, J. M. Hernandez, Z. Schwarz-Sommer, E. Grotewold, and D. F. Mandoli. 2004. Comparison of ESTs from juvenile and adult phases of the giant unicellular green alga *Acetabularia acetabulum*. *Plant Biol.* **4(3):** 1–15.
14. Jin, E. S., J. E. W. Polle, H. K. Lee, S. M. Hyun, and M. Chang. 2003. Xanthophylls in microalgae: From biosynthesis to biotechnological mass production and application. *J. Microbiol. Biotechnol.* **13:** 165–175.
15. Kim, N.-K., J.-K. Lee, and C.-G. Lee. 2004. Pigment reduction to improve photosynthetic productivity of *Rhodobacter sphaeroides*. *J. Microbiol. Biotechnol.* **14:** 442–450.
16. Kobayashi, M. 2003. Astaxanthin biosynthesis enhanced by reactive oxygen species in the green alga *Haematococcus pluvialis*. *Biotech. Bioproc. Engin.* **8:** 322–330.
17. Kobayashi, M., T. Kakizono, and S. Nagai. 1993. Enhanced carotenoid biosynthesis by oxidative stress in acetate-induced cyst cells of a green unicellular alga, *Haematococcus pluvialis*. *Appl. Environ. Microbiol.* **59:** 867–873.
18. Kobayashi, M., T. Kakizono, N. Nishio, S. Nagai, Y. Kurimura, and Y. Tsuji. 1997. Antioxidant role of astaxanthin in the green alga *Haematococcus pluvialis*. *Appl. Microbiol. Biotechnol.* **48:** 351–356.
19. Lorenz, R. T. and G. R. Cysewski. 2000. Commercial potential for *Haematococcus* microalgae as a natural source of astaxanthin. *Trends Biotechnol.* **18:** 160–167.
20. Nichols, H. W. and H. C. Bold. 1964. *Trichsarcina polymorpha* gen. et. nov. *J. Phycol.* **1:** 34–38.
21. Niyogi, K. K. 1999. Photoprotection revisited: Genetic and molecular approaches. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50:** 333–359.
22. Oh, K. S., O. S. Kwon, Y. W. Oh, M. J. Sohn, S. G. Jung, Y. K. Kim, M. G. Kim, S. K. Rhee, G. Gellissen, and H. A. Kang. 2004. Fabrication of a partial genome microarray of the methylotrophic yeast *Hansenula polymorpha*: Optimization and evaluation of transcript profiling. *J. Microbiol. Biotechnol.* **14:** 1239–1248.
23. Orosa, M., D. Franqueira, A. Cid, and J. Abalde. 2005. Analysis and enhancement of astaxanthin accumulation in *Haematococcus pluvialis*. *Bioresour. Technol.* **96:** 373–378.
24. Park, E. K. and C.-G. Lee. 2001. Astaxanthin production by *Haematococcus pluvialis* under various light intensities and wavelengths. *J. Microbiol. Biotechnol.* **11:** 1024–1030.
25. Shaish, A., M. Avron, U. Pick, and A. Ben-Amotz. 1993. Are active oxygen species involved in induction of β-carotene in *Dunaliella bardawil*. *Planta* **190:** 363–368.
26. Wan, K. L., S. P. Chong, S. T. Ng, M. W. Shirley, F. M. Tomley, and M. S. Jangi. 1999. A survey of genes in *Eimeria tenella* merozoites by EST sequencing. *Int. J. Parasitol.* **29:** 1885–1892.
27. Wang, S. B., F. Chen, M. Sommerfeld, and Q. Hu. 2004. Proteomic analysis of molecular response to oxidative stress by the green alga *Haematococcus pluvialis* (Chlorophyceae). *Planta* **220:** 17–29.