Early Stage Adaptation of a Mesophilic Green Alga to Antarctica: Systematic Increases in Abundance of Enzymes and LEA Proteins

Yali Wang,[1,2] Xiaoxiang Liu,[1] Hong Gao,[1,2] Hong-Mei Zhang,[3] An-Yuan Guo,[3] Jian Xu,[4] and Xudong Xu*[,1,2]

[1]State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, Hubei, China

[2]Key Laboratory of Algal Biology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, Hubei, China

[3]College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei, China

[4]Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong, China

***Corresponding author:** E-mail: xux@ihb.ac.cn.

Associate editor: Irina Arkhipova

The *Chlorella vulgaris* NJ-7 Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number VATV00000000 (version VATV01000000). The *C. vulgaris* UTEX259 Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number VATW00000000 (version VATW010000000). Assembled transcripts ≥200 nt have been deposited in the NCBI Transcriptome Shotgun Assembly (TSA) sequence database with accession numbers GHLW00000000 for NJ-7 and GHLX00000000 for UTEX259. Raw sequence read data have been deposited in the NCBI Sequence Read Archive (SRA) with the following study identifiers: SRP198424 for NJ-7 transcriptome and RNA-seq; SRP198705 for UTEX259 transcriptome and RNA-seq. Organellar genome sequences have been deposited at the NCBI GenBank under accession numbers MK948100 (NJ-7 chloroplast genome), MK948101 (NJ-7 mitochondrial genome), MK948102 (UTEX259 chloroplast genome), and MK948103 (UTEX259 mitochondrial genome). The mass spectrometry proteomics raw data have been deposited to the ProteomeXchange Consortium via the MassIVE partner repository (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp; last accessed November 24, 2019) with the data set identifier PXD014018.

Abstract

It is known that adaptive evolution in permanently cold environments drives cold adaptation in enzymes. However, how the relatively high enzyme activities were achieved in cold environments prior to cold adaptation of enzymes is unclear. Here we report that an Antarctic strain of *Chlorella vulgaris*, called NJ-7, acquired the capability to grow at near 0 °C temperatures and greatly enhanced freezing tolerance after systematic increases in abundance of enzymes/proteins and positive selection of certain genes. Having diverged from the temperate strain UTEX259 of the same species 2.5 (1.1–4.1) to 2.6 (1.0–4.5) Ma, NJ-7 retained the basic mesophilic characteristics and genome structures. Nitrate reductases in the two strains are highly similar in amino acid sequence and optimal temperature, but the NJ-7 one showed significantly higher abundance and activity. Quantitative proteomic analyses indicated that several cryoprotective proteins (LEA), many enzymes involved in carbon metabolism and a large number of other enzymes/proteins, were more abundant in NJ-7 than in UTEX259. Like nitrate reductase, most of these enzymes were not upregulated in response to cold stress. Thus, compensation of low specific activities by increased enzyme abundance appears to be an important strategy for early stage cold adaptation to Antarctica, but such enzymes are mostly not involved in cold acclimation upon transfer from favorable temperatures to near 0 °C temperatures.

Key words: **Antarctica, cold adaptation, intraspecies divergence, omics, enzyme activity, *Chlorella vulgaris*.**

Article

Introduction

Antarctica is the coldest continent on the earth, with 99.8% of the area (Burton-Johnson et al. 2016) covered by a sheet of ice that averages about 2 km thick (Fretwell et al. 2013). The lowest air temperature in Antarctica may reach −89.2 °C (National Polar Research Institute (Japan) 1991), but the ground surface temperatures on ice-free sites are in a range of −35 to 5 °C in most time of a year (Guglielmin 2006). In Antarctica, animal and plant communities are naturally separated from those on other continents (Convey and Stevens 2007; Fraser et al. 2012) and largely limited to ice-free sites that receive meltwater in warm seasons (Convey et al. 2014). In contrast, microalgae, bacteria, and fungi not only grow on these ice-free sites but also thrive in sea ice (Thomas and Dieckmann 2002), snow fields (Davey et al. 2019), and ice-covered lakes (Karl et al. 1999; Possmayer et al. 2016). Microbes from other continents could be transported to Antarctica by atmospheric circulation (Mayol et al. 2017; Cáliz et al. 2018), even though their contribution to Antarctic microbial communities is limited (Archer et al. 2019).

In the geological history, Antarctica was formed from the breakup of the supercontinent Gondwana. About 34–33 Ma,

the atmospheric CO_2 level significantly dropped, and the ice cover expanded rapidly on the continent to near-modern dimensions or larger than present-day values; about 23 Ma, the Drake Passage opened between Antarctica and South America (Galeotti et al. 2016; Lear and Lunt 2016). Mesophilic microorganisms that experienced the quick cooling period might evolve into psychrophilic or psychrotrophic species; on the other hand, those microorganisms transported from other continents in later periods might also develop the cold-growth capability and freezing tolerance.

The main challenge faced by organisms in Antarctica is to keep cell activities at near 0 °C temperatures and maintain survivability under freezing conditions. The freezing tolerance depends on ice-binding proteins (Bar Dolev et al. 2016; Collins and Margesin 2019) and LEA (late embryogenesis abundant) proteins (Shih et al. 2008; Liu et al. 2011; Wang et al. 2011). These proteins prevent the growth/recrystallization of ice or freeze/thaw-induced inactivation of enzymes. Cell activities depend on various biochemical reactions. Our understanding of biochemical reactions in the cold is largely based on cold-adapted or psychrophilic enzymes (Morgan-Kiss et al. 2006). Compared with their mesophilic homologs, such enzymes feature significantly lowered optimal temperatures and adaptive mutations that destabilize the structures bearing the active site or the overall structure (Feller and Gerday 2003; Santiago et al. 2016). It has been shown that the optimal temperature or thermal stability of a mesophilic enzyme can be downshifted by site-directed mutagenesis (Saavedra et al. 2018; Liao et al. 2019) and that the activity at a low temperature can be increased by directed evolution (Zhao and Feng 2018). However, because adaptive mutations that cause coding differences are relatively rare (Jones et al. 2012) and coding differences toward cold adaptation must be even rarer, it remains to be answered whether there is an alternative strategy to maintain relatively high enzyme activities at near 0 °C temperatures before enzymes are significantly cold-adapted.

Without cold-growing microorganisms generated from mesophiles by experimental evolution, natural intraspecies divergence in the capability to grow at low temperatures would provide very valuable materials for studies. Here, we analyzed an Antarctic strain and a temperate strain of *Chlorella vulgaris* (green alga) at genomic, transcriptomic, and proteomic levels. Our results indicate that systematic elevation in abundance of enzymes could allow a mesophilic newcomer to develop the capability to grow at near 0 °C temperatures, representing an early stage adaptive mechanism in Antarctic environments. Remarkably, most enzymes with higher levels in the Antarctic strain are not upregulated upon transfer from 20 to 4 °C.

Results

Physiological Divergence between *C. vulgaris* NJ-7 and UTEX259

Chlorella vulgaris NJ-7 was isolated from rock samples collected near a transitory pond 5 km away from the Zhongshan Station (69°22′S–76°22′E) in Antarctica (Hu et al. 2008). Its 18S rRNA-ITS1-5.8S rRNA region is identical to that of *C. vulgaris* UTEX259, a strain initially collected and isolated from Delft (52°00′N, 04°21′E), the Netherlands (https://utex.org; last accessed November 24, 2019); its ITS2-28S rRNA region differs from UTEX259 at five bases (supplementary fig. S1, Supplementary Material online). They show similar growth at 20 °C (fig. 1*a*), but only NJ-7 is able to grow at 4 °C (fig. 1*b*); after being frozen at −20 °C, the survivability of NJ-7 is much higher than that of UTEX259, and the survivability can be enhanced by preconditioning (48-h exposure) at 4 °C (fig. 1). When cultured at 20 °C, both strains show maximal photosynthetic activities at 30 °C (supplementary fig. S2, Supplementary Material online); NJ-7 maintains higher photosynthetic activities than UTEX259 at high temperatures. These physiological characteristics indicate that NJ-7 is a mesophilic algal strain evolving toward a psychrotrophic one. The higher photosynthetic activities of NJ-7 at high temperatures could be either associated with or independent of the enhanced freezing tolerance.

Nuclear Genome Assembly and Synteny Analyses

To analyze the relationship between the two strains of the same species and the mechanism for NJ-7 to develop the cold-growth capability and freezing tolerance, we sequenced the whole genomes of both strains. Transcriptomes were also sequenced to facilitate genome assembly and gene predictions. The assembled NJ-7 nuclear genome consisted of 39.1 Mb in 753 scaffolds with an N50 size of 938 kb; the UTEX259 nuclear genome consisted of 39.1 Mb in 780 scaffolds with an N50 size of 498 kb. A total of 9,412 protein-coding genes were predicted in NJ-7 and 9,439 in UTEX259; 13,390 transcripts (isotigs) were identified in NJ-7, and 13,880 identified in UTEX259 (features are summarized in table 1, and detailed information for sequencing, assembly and annotation in supplementary tables S1–S6, Supplementary Material online).

The quality and gene coverage of genome assemblies were then assessed. First, completeness of genome assemblies was assessed by BUSCO (Simao et al. 2015). Of the 2,168 universal single-copy orthologs of the chlorophyta gene set, 90.0% were found in the NJ-7 genome assembly as complete genes, and 90.3% found in UTEX259; 4.6% and 4.3% were completely missing in NJ-7 and UTEX259 assemblies, respectively (supplementary table S7, Supplementary Material online). Numbers of predicted genes in *C. vulgaris* strains NJ-7 and UTEX259 are close to those of *C. variabilis* NC64A and *Coccomyxa subellipsoidea* C-169. Second, the genome assemblies were evaluated by aligning with assembled transcripts (Sanger-based fosmid-end sequences also used for assessing NJ-7). For NJ-7, about 97.3% of the fosmid-end sequences were mapped to the genome assembly, and 93.5% of the transcripts mapped over at least 90% of their length. Of 956 pairs of fosmid-end sequences, 790 pairs were mapped to the same scaffold, and the average distance was 34.7 kb, almost identical to the estimated average insert size of fosmid clones (35.0 kb). To the UTEX259 genome assembly, about 93.4% of the

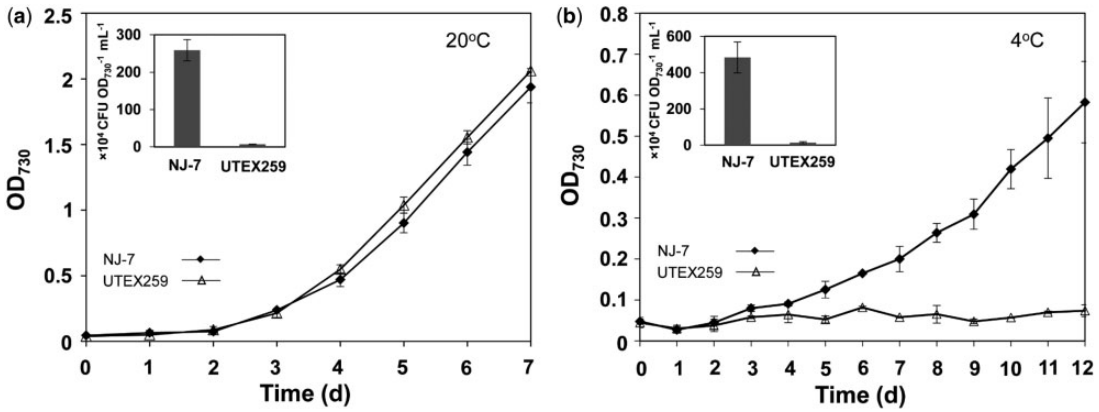

FIG. 1. Comparisons of the cold-growth capability and freezing tolerance between NJ-7 and UTEX259. (*a*) Growth of NJ-7 and UTEX259 at 20 °C and survivability of the two strains after being frozen at −20 °C (inset). (*b*) Growth of NJ-7 and UTEX259 at 4 °C and survivability of the two strains grown at 20 °C, pretreated at 4 °C (48 h) and frozen at −20 °C (inset). Data are means ± SD from three biological replicates.

Table 1. Genome and Transcriptome Statistics of NJ-7 and UTEX259.

Features	NJ-7	UTEX259
Nuclear genome size (Mb)	39.08	39.13
Genomic G+C content (%)	61.68	61.68
Number of scaffolds	753	780
Scaffold N50 (bp)	937,767	497,853
Repeated sequences (%)	6.2	5.9
Number of protein-coding genes	9,412	9,439
Genes with transcript support (%)	84.92	86.48
Genes with homology support (%)	87.78	86.16
Average protein length (aa)	507	515
Average exon length (bp)	178	180
Average intron length (bp)	216	218
Average number of exons per gene	8.54	8.57
Number of isotigs	13,390	13,880
Syntenic regions (%)	99.37	97.68

transcripts were mapped over at least 90% of their length. Single-base mismatch and insertion/deletion frequencies of the two genome assemblies were <1.0 base/10 kb (compared with sequences generated by Sanger sequencing). These results indicated the high coverage and high quality of the genome assemblies.

Similarities between nuclear genomes of the two strains were evaluated with synteny and collinearity analyses. Scaffolds longer than 10 kb, which covered over 97% of each genome, were used to generate the syntenic dot plot (supplementary fig. S3, Supplementary Material online). The result indicated that syntenic regions covered 99.37% of the NJ-7 scaffolds and 97.68% of the UTEX259 scaffolds. The longest 12 scaffolds (with a total length of 18.1 Mb) of NJ-7 and the corresponding scaffolds of UTEX259 were used to generate the dual synteny plot (supplementary fig. S4, Supplementary Material online). Only one large-scale genomic rearrangement was identified. The percentage of collinearity between NJ-7 and UTEX259 was 97.33%; amino acid sequence identity between their reciprocal best blast hits was 94.68%, significantly higher than those between *C. vulgaris* and other green algae (supplementary table S8, Supplementary Material online).

Comparison of Organelle Genomes and Estimation of Divergence Time

The relationship between the two strains was also shown with their organelle genomes and divergence time. Except for two genomic inversions and differences in *psbA* and *psbC* (with or without an intron that contains an endonuclease gene), the chloroplast genomes of NJ-7 and UTEX259 are identical in gene content and arrangement (supplementary fig. S5 and table S9, Supplementary Material online). The two chloroplast genomes closely resemble that of *C. vulgaris* C-27 [a strain isolated from Sendai (38°16′N, 141°02′E), Japan (NIES-2170, http://mcc.nies.go.jp/; last accessed November 24, 2019)] but differ from those of other *Chlorella* strains in gene arrangement (supplementary fig. S6, Supplementary Material online). The mitochondrial genomes of NJ-7 and UTEX259 are even more similar to each other (supplementary fig. S7 and table S9, Supplementary Material online). Organelle genomes of NJ-7 are smaller than that of UTEX259 due to the reduced noncoding regions.

Divergence times between NJ-7, UTEX259, other green algae, and higher plants were estimated based on sequences of 36 chloroplast genes, with three fossil calibrations as previously reported (Herron et al. 2009). Broad-scale analyses were performed with four combinations of species/strains (fig. 2*a*; supplementary fig. S8*a*–*c*, Supplementary Material online). It is deduced that NJ-7 diverged from UTEX259 about 2.5 (95% confidence interval 1.1–4.1) Ma to 2.6 (1.0–4.5) Ma, much later than the divergence between NJ-7/UTEX259 and NC64A [153.4 (69.2–247.1) Ma], also later than the opening of Drake Passage (~23 Ma). The structure of the chronogram in figure 2*a* is consistent with the dendrograms based on amino acid and nucleotide sequences of 1,080 nuclear genes (supplementary fig. S9*a* and *b*, Supplementary Material online). In addition to the broad-scale chronograms, we also constructed a fine-scale chronogram for *Chlorella* species, in which the divergence time between UTEX259 and NJ-7 (and C-27) was estimated to be 2.9 (0.6–6.4) Ma (fig. 2*b*). In light of the mesophilic characteristics of NJ-7 and the estimated time for its divergence from UTEX259, the ancestor of

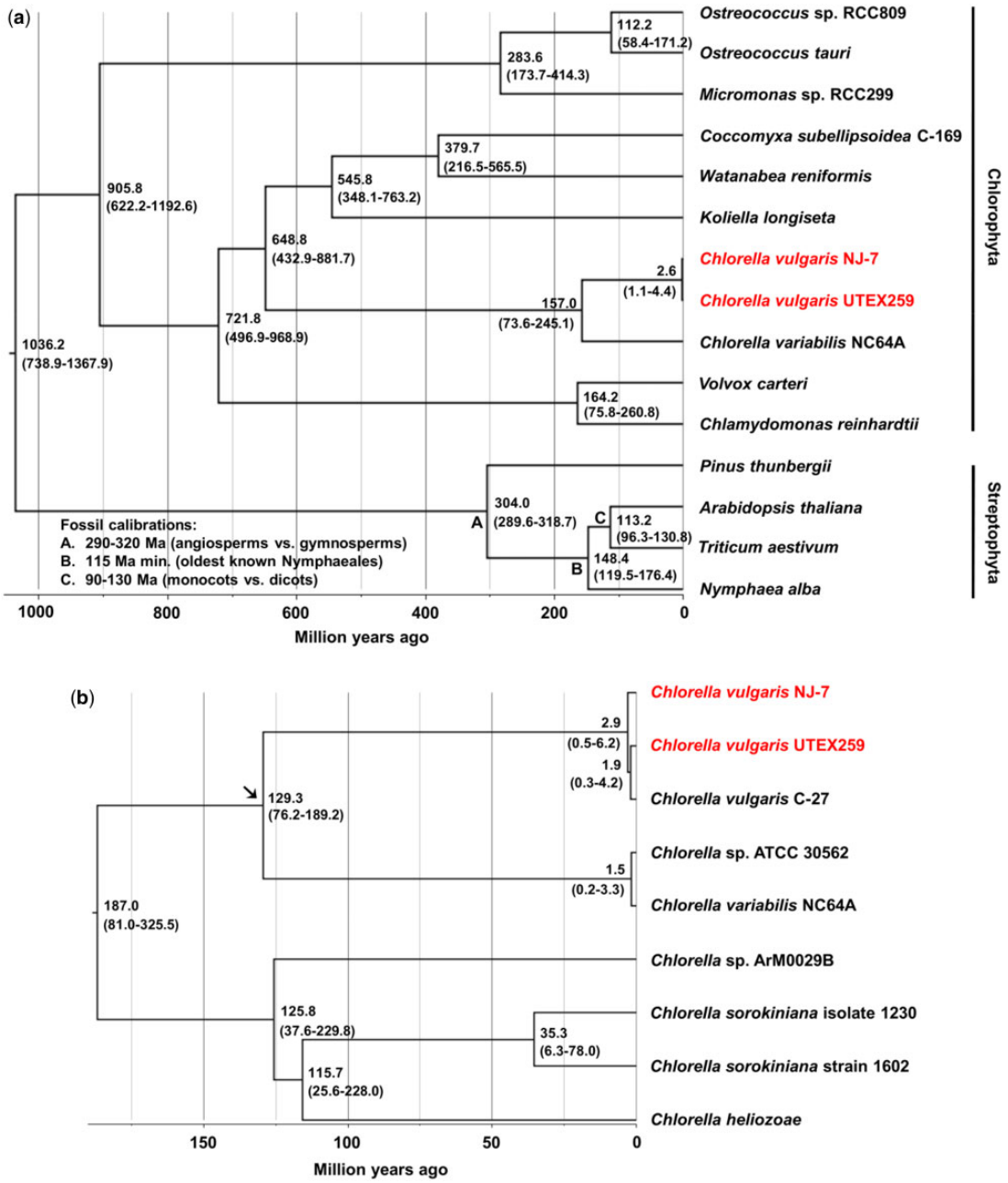

FIG. 2. Chronograms showing estimated divergence times for green algae and higher plants. Divergence times were estimated using BEAST based on 36 chloroplast genes. Branch lengths are proportional to the absolute ages of nodes (scale on *x*-axis in million years). Numbers to the right or left of nodes are the ages of nodes and 95% confidence intervals (in parentheses). (*a*) One of the broad-scale chronograms. A, B, and C indicate the nodes to which time constrains were applied based on fossil records. The other three chronograms (with different combinations of species/strains) are shown in supplementary figure S8, Supplementary Material online. (*b*) The fine-scale chronogram. The divergence time between *Chlorella variabilis* and *C. vulgaris* estimated from the broad-scale analysis, in particular 157.0 Ma (73.6–245.1 Ma), was used to calibrate the fine-scale analysis (indicated by an arrow) but slightly changed after the computations. Of the four divergence times between *C. variabilis* and *C. vulgaris*, 157.0 Ma (73.6–245.1 Ma) from (*a*) is close to the average.

NJ-7 was probably transported to the Antarctic continent from a temperate region.

Genomic Evolution toward Cold Adaptation of NJ-7

Cold adaptation is an evolutionary process toward the capabilities to grow at near 0 °C temperatures and survive under freezing conditions. This process depends on accumulation of adaptive mutations and leads to divergence between strains. With the genomic data, we first searched for genetic changes that may be associated with cold adaptation. NJ-7 shows much less nonsynonymous substitutions per site and gene duplications than UTEX259 (supplementary table S10, Supplementary Material online), but the two strains are comparable in numbers of positively selected genes (supplementary excel S1, Supplementary Material online) and alternative splicing events (supplementary excel S1, Supplementary

Material online). Positive selections involved regulation of RNA conformation, desaturation of membrane lipids, energy metabolism, anabolic and catabolic reactions, metabolite transport, protein processing and translocation, intracellular movement, transcription, DNA repair, etc. For an example, the sphingolipid delta(4)-desaturase (NJ-7.evm.TU.scaffold00053.17) was positively selected. Sphingolipids are enriched in the outer leaflet of plasma membrane in plants, and the long chain base of sphingolipids, called sphinganine, can be desaturated by sphingolipid delta(4)-desaturase and sphingolipid delta(8)-desaturase (Luttgeharm et al. 2016). The delta(8)-desaturase plays a major role in desaturation of long chain base and cold tolerance in higher plants (Zhou et al. 2016) but is not found in *Chlorella*, *Chlamydomonas*, and *Volvox*. The positive selection of sphingolipid delta(4)-desaturase in NJ-7 probably reflects the critical role of sphingolipid desaturation in cold adaptation. Of genes with alternative splicing, there are seven related to DNA repair in NJ-7 but only one in UTEX259 (supplementary table S11, Supplementary Material online). These alternative splicing events in NJ-7 might have contributed to adaptation to the strong UV radiation in Antarctica (Liao and Frederick 2005).

Unique genes (supplementary excel S1, Supplementary Material online), high-copy-number genes (supplementary fig. S10 and table S12, Supplementary Material online), and expanded gene families (supplementary tables S13–S15, Supplementary Material online) were also analyzed, but no apparent connection to cold adaptation was identified. Gene duplication has been shown to be a rapid mechanism for adaptation to stressful or novel environmental conditions (Kondrashov 2012). We identified genes with multiple copies in NJ-7 and UTEX259 and analyzed the relationship between gene expression and copy numbers. Only one gene (MFS transporter) with higher copy number in UTEX259 showed increased expression relative to its homolog in NJ-7 (supplementary fig. S10 and table S12, Supplementary Material online).

Cold Adaptation and Cold Acclimation of NJ-7 Based on Altered Gene Expression

Unlike cold adaptation, cold acclimation is a physiological process that relieves the cold stress in temperature-fluctuating environments and prepares for the freezing conditions (Morgan-Kiss et al. 2006). It depends on the regulation of gene expression and metabolic pathways. To which extent genes/proteins regulated in cold acclimation are involved in cold adaptation has not been addressed.

To understand how NJ-7 adapted to the Antarctica, we further analyzed mRNA and protein profiles in NJ-7 and UTEX259 cultured at 20 °C with or without treatment at 4 °C, and identified differential gene expression in the same strain at mRNA (RNA-seq) (supplementary table S16 and excel S2, Supplementary Material online) and protein (proteomic analysis) levels (supplementary excel S3, Supplementary Material online). Real-time quantitative polymerase chain reaction (RT-qPCR) analysis was conducted on eight genes in cells grown at 20 °C and exposed to 4 °C to evaluate the quality of RNA-seq data. Expression levels assessed with RT-qPCR correlated well ($R^2 = 0.903$) with those obtained from the RNA-seq analysis (supplementary fig. S11, Supplementary Material online). The RNA-seq and proteomic data were also used to analyze the differential expression between two strains at 20 or 4 °C (supplementary excels S2 and S3, Supplementary Material online), but the differences between protein abundance in two strains were analyzed by generating a new NJ-7/UTEX259 protein database, with which orthologous proteins were identified based on those identical trypsin-digested peptides. The RNA-seq and proteomic analyses were highly reproducible between biological replicates as shown with Pearson's correlation coefficients (supplementary fig. S12, Supplementary Material online). Numbers of differentially expressed genes in four data sets are shown in supplementary table S17, Supplementary Material online. Genes up- or downregulated in NJ-7 or UTEX259 after transfer from 20 to 4 °C are related to cold acclimation, whereas those with altered expression in NJ-7 relative to their homologs in UTEX259 are related to strain divergence.

Using upregulated genes as the example, we first analyzed overlaps between different data sets. The results are summarized in figure 3, from which four implications can be derived: 1) Strain divergence is less dependent on regulation at mRNA level than cold acclimation. During cold acclimation, over 70% of genes [215/(215 + 84)] upregulated in NJ-7 at the protein level are concurrently upregulated at the mRNA level (see fig. 3*a*-I); relative to their homologs in UTEX259, only 28.1% of genes [113/(113 + 299)] with higher protein levels in NJ-7 also show higher mRNA levels (at 4 °C, see fig. 3*b*-II); 2) most proteins [193/(193 + 106)] upregulated in NJ-7 during cold acclimation are also upregulated in UTEX259 (see fig. 3*c*-I); 3) most proteins [312/(312 + 100)] with higher abundance in NJ-7 than in UTEX259 at 4 °C show higher abundance at 20 °C (see fig. 3*c*-III); and 4) most proteins [374/(374 + 38)] with higher abundance in NJ-7 than in UTEX259 at 4 °C are not upregulated during cold acclimation (see fig. 3*c*-II).

For downregulated genes, overlaps between data sets of "NJ-7/UTEX259" exhibit a similar pattern (supplementary fig. S13*b*, *c*-II, and *c*-III, Supplementary Material online) to those for upregulated genes (fig. 3*b*, *c*-II, and *c*-III), but the number of proteins downregulated during cold acclimation (supplementary fig. S13*a*, Supplementary Material online) is much less than that of proteins upregulated (fig. 3*a*), and most proteins downregulated in NJ-7 and UTEX259 during cold acclimation do not overlap (supplementary fig. S13*c*-I, Supplementary Material online).

Divergence between the two strains was driven by adaptation to environments and genetic drifts. Adaptation of NJ-7 to the cold environment is one of the most important factors. Of the 38 upregulated proteins (fig. 3; supplementary table S18, Supplementary Material online) involved in both strain divergence and cold acclimation, sphingolipid delta(4)-desaturase was positively selected (supplementary excel S1, Supplementary Material online). Other proteins in the list are involved in freezing tolerance (cryoprotective protein), cold stress response and tolerance, cell division, carbon metabolism, etc. As we pointed out above, many more proteins with

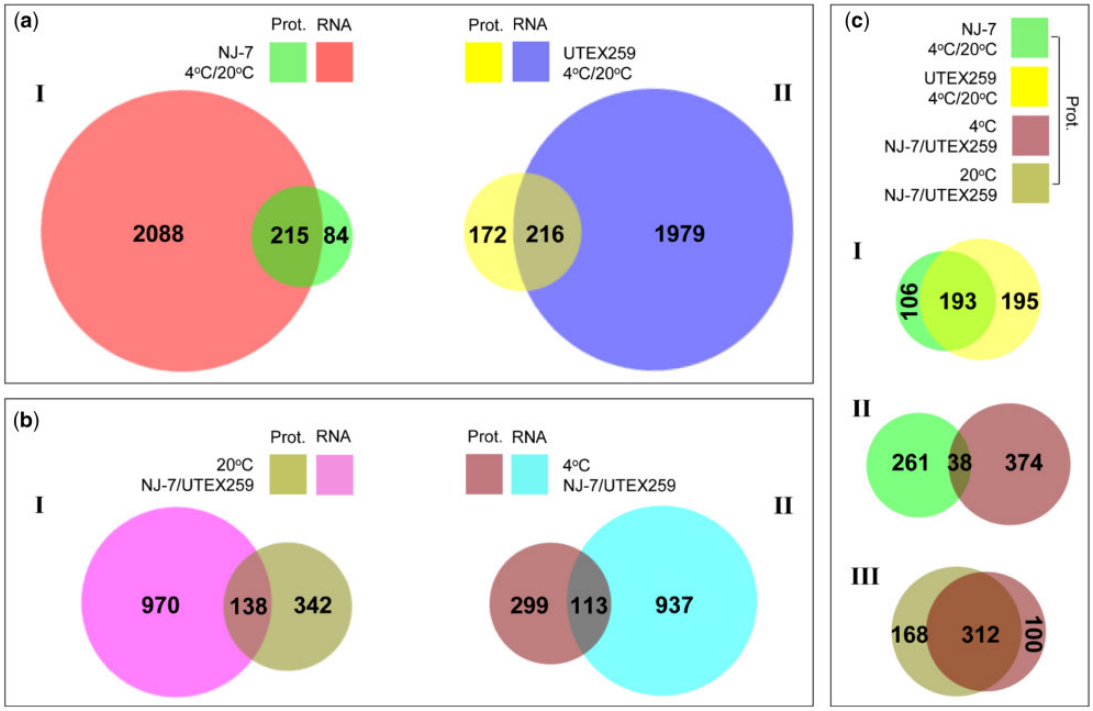

FIG. 3. Venn diagrams showing overlaps between genes upregulated in RNA-seq and proteomic analyses or between proteomic data sets. Upregulation at protein level is defined as a fold change in protein abundance $\geq$1.3 (P-value $<$ 0.05); upregulation at mRNA level is defined as a fold change $\geq$2.0 (P-value $<$ 0.05). NJ-7 4 °C/20 °C, UTEX259 4 °C/20 °C: genes upregulated at mRNA or protein level in NJ-7 or UTEX259 during cold acclimation; 4 °C NJ-7/UTEX259, 20 °C NJ-7/UTEX259: genes with higher expression in NJ-7 than in UTEX259 (strain divergence) at 4 °C or 20 °C. Prot., proteomic data; RNA, RNA-seq data. (*a*) The overlap between genes upregulated in RNA-seq and proteomic analyses for cold acclimation of NJ-7 (I) or UTEX259 (II). There are 215 genes upregulated at both mRNA and protein levels in NJ-7 at 4 °C relative to 20 °C, 216 genes at both levels in UTEX259. (*b*) The overlap between genes upregulated in RNA-seq and proteomic data for strain divergence at 20 °C (I) or 4 °C (II). At 20 °C, 138 genes show increased expression at both levels in NJ-7 relative to UTEX259; at 4 °C, 113 genes. (*c*) Overlaps between proteomic data sets showing 193 genes upregulated at the protein level in both NJ-7 and UTEX259 during cold acclimation (I), 312 genes upregulated in NJ-7 relative to UTEX259 at both 20 and 4 °C (III), but only 38 genes upregulated in NJ-7 in both cold acclimation and strain divergence (II).

higher abundance in NJ-7 are not upregulated during cold acclimation. Increases in abundance of these proteins may play an important role in the maintenance of cellular activities at near 0 °C temperatures and the greatly enhanced freezing tolerance.

To identify functions enhanced in NJ-7 relative to UTEX259, we performed gene set enrichment analysis (GSEA) of differentially expressed genes based on Gene Ontology (GO) categories (supplementary fig. S14, Supplementary Material online). NJ-7 (supplementary fig. S14*a* and *b*, Supplementary Material online) and UTEX259 (supplementary fig. S14*c* and *d*, Supplementary Material online) show positive enrichments for many GO terms in cold acclimation but only one to several GO terms in strain divergence (supplementary fig. S14*e*, Supplementary Material online). At 4 °C, only the GO category "chromatin" is enriched in NJ-7 compared with UTEX259. However, this analysis would not identify positive enrichments for functional categories not in the GO list or pathways with only one or two rate-limiting reactions enhanced. In addition, many NJ-7/UTEX259 genes involved in cold acclimation show the highest similarity to homologs of other green algae in public databases, but these algal homologs have not been assigned to the GO category of "cold acclimation" or "response to cold." We extracted predicted NJ-7/UTEX259 proteins based on their similarities to gene products of the two GO categories in AmiGO 2 (http://amigo.geneontology.org/amigo; last accessed November 24, 2019). GSEA showed positive enrichments for these two groups of proteins in cold acclimation (supplementary fig. S15, Supplementary Material online).

We also identified 152 transcription factors (TFs) in NJ-7 and 163 TFs in UTEX259 and analyzed their differential expression across temperatures and strains (supplementary excel S4, Supplementary Material online). The differential expression of TFs may lead to the differences in transcription of target genes. Several TFs with identified recognition motifs (for orthologs) in *C. variabilis* NC64A were used to test this possibility, but none of them showed a change in expression correlated with that of a potential target gene.

Elevated Abundance of LEA Proteins in NJ-7

LEA proteins were first identified in cotton and so named because they accumulate during the late maturation stages of seed development (Dure et al. 1981). Proteins in this family play important roles in stress tolerance in bacteria, fungi, plants, and animals (Shih et al. 2008). In particular, they can enhance freezing tolerance as cryoprotectants (Honjoh et al. 2001; Sasaki et al. 2014). In this study, we systematically

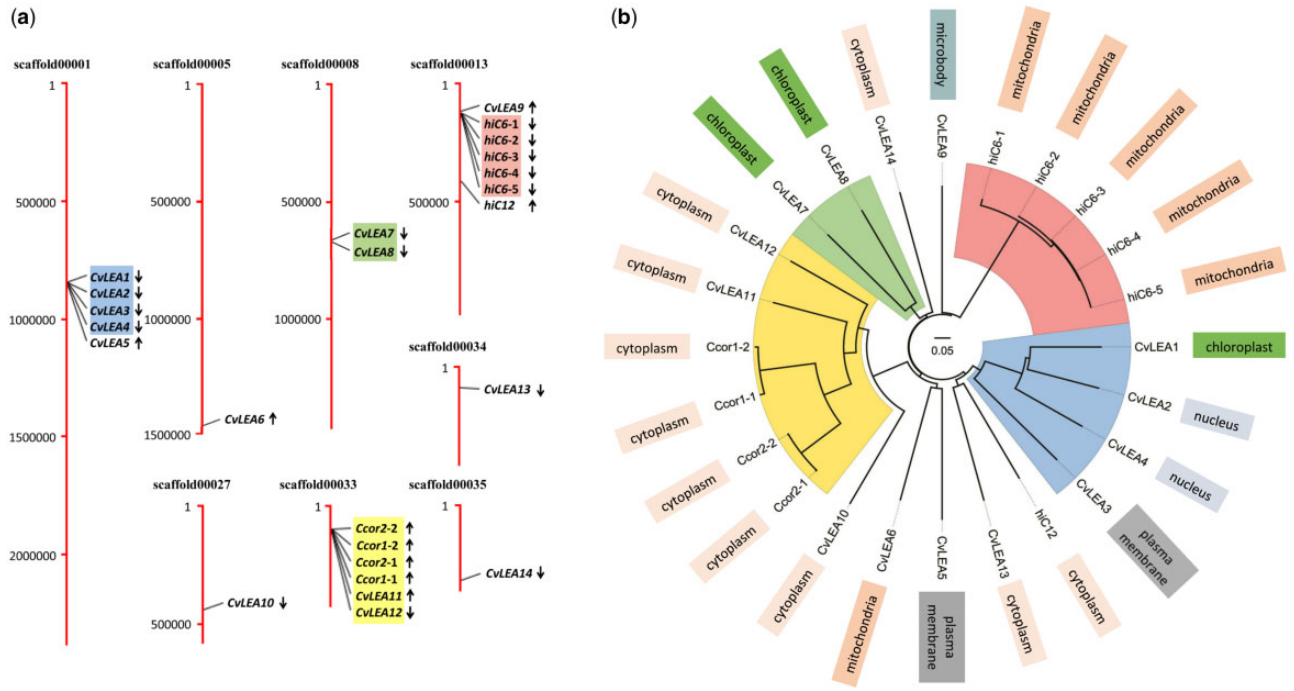

FIG. 4. Physical locations of LEA protein genes, predicted cellular locations of LEA proteins and their phylogenetic relationship. LEA proteins/genes are almost identical to each other between counterparts in NJ-7 and UTEX259, but *hiC6-5* is only found in NJ-7. (*a*) LEA protein genes in assembled scaffolds of NJ-7. Arrows indicate orientations of genes. (*b*) Unrooted phylogenetic tree of NJ-7 LEA proteins with the predicted cellular locations indicated. The scale bar shows expected substitutions per site. Genes clustered on the genome and the phylogenetic tree are highlighted with corresponding colors.

identified LEA protein genes in NJ-7 and UTEX259 by motif and secondary structure searches, Pfam domain searches, and similarity (to known LEA proteins) searches (supplementary table S19, Supplementary Material online). Compared with other green algae, *C. vulgaris* strains possess more LEA proteins (supplementary table S20, Supplementary Material online). Except *hiC6-5* that is not found in UTEX259, all LEA protein genes are very similarly arranged in the two genomes. In NJ-7, *CvLEA1*/*CvLEA2*/*CvLEA3*/*CvLEA4*, *CvLEA7*/*CvLEA8*, *CvLEA12*/*CvLEA11*/*Ccor1-1*/*Ccor2-1*/*Ccor1-2*/*Ccor2-2*, and *hiC6-1*/*hiC6-2*/*hiC6-3*/*hiC6-4*/*hiC6-5* are organized in four gene clusters (fig. 4*a*); genes from the same genomic cluster are also clustered in the phylogenetic tree (fig. 4*b*). In *C. variabilis* NC64A, there are genes homologous to *CvLEA1* (1 copy) and *hiC6* (1 copy), but no homologs to *CvLEA7*/*CvLEA8* or *Ccor1*/*Ccor2*. Homologs to *Chlorella* LEA genes are not found in other green algae. Apparently, these gene clusters in *C. vulgaris*, except for the occurrence of *hiC6-5*, had been formed by gene duplication before the divergence between NJ-7 and UTEX259. The encoded LEA proteins are predicted to be located in the cytoplasm, nucleus, or organelles (fig. 4*b*) to provide cryoprotection of enzymes/proteins in different cellular compartments, but the predicted locations need to be confirmed with experiments in the future. Most LEA proteins in the two strains differ from each other by <3% amino acid residues (supplementary table S21, Supplementary Material online).

Based on proteomic analyses, we identified seven LEA proteins (five isoforms of hiC6 and two versions of Ccor2 were respectively treated as one protein, plus CvLEA3, CvLEA5, CvLEA10, CvLEA11, and CvLEA14) with higher abundance in NJ-7 than in UTEX259 (fold change $\geq$ 1.3, *P*-value $<$ 0.05) at both 20 and 4 °C, three additional LEA proteins (CvLEA2, CvLEA13, and hiC12) with higher abundance in NJ-7 at 20 °C (fig. 5). All of their encoding genes, except *CvLEA2* and *CvLEA14*, also showed higher mRNA levels in NJ-7 (fold change $\geq$ 2.0, *P*-value $<$ 0.05) under the corresponding conditions. GSEA of LEA proteins differentially expressed between NJ-7 and UTEX259 showed significant positive enrichments in NJ-7 at both 20 and 4 °C (supplementary fig. S16*a*, Supplementary Material online). Because LEA proteins are not listed as a GO category, and only one from NJ-7 (NJ-7.evm.TU.scaffold00034.28) shows similarity to an LEA protein in the GO category "response to cold" (including cold acclimation), the positive enrichment would not be identified by the GSEA based on GO (supplementary fig. S14, Supplementary Material online).The increased abundance of so many LEA proteins would greatly enhance the freezing tolerance of NJ-7. On the other hand, because LEA proteins can reduce cellular peroxides and protect enzyme activities under stresses, they may promote tolerance against high temperature and other stresses (Zhang et al. 2014; Wang et al. 2017). In NJ-7, such effects of LEA proteins may indirectly promote photosynthetic activities at high temperatures (supplementary fig. S2, Supplementary Material online).

As shown in the fine-scale chronogram (fig. 2*b*), NJ-7 diverged from another temperate strain C-27 of *C. vulgaris* 1.9 (0.3–4.2) Ma. C-27 also accumulates LEA proteins upon

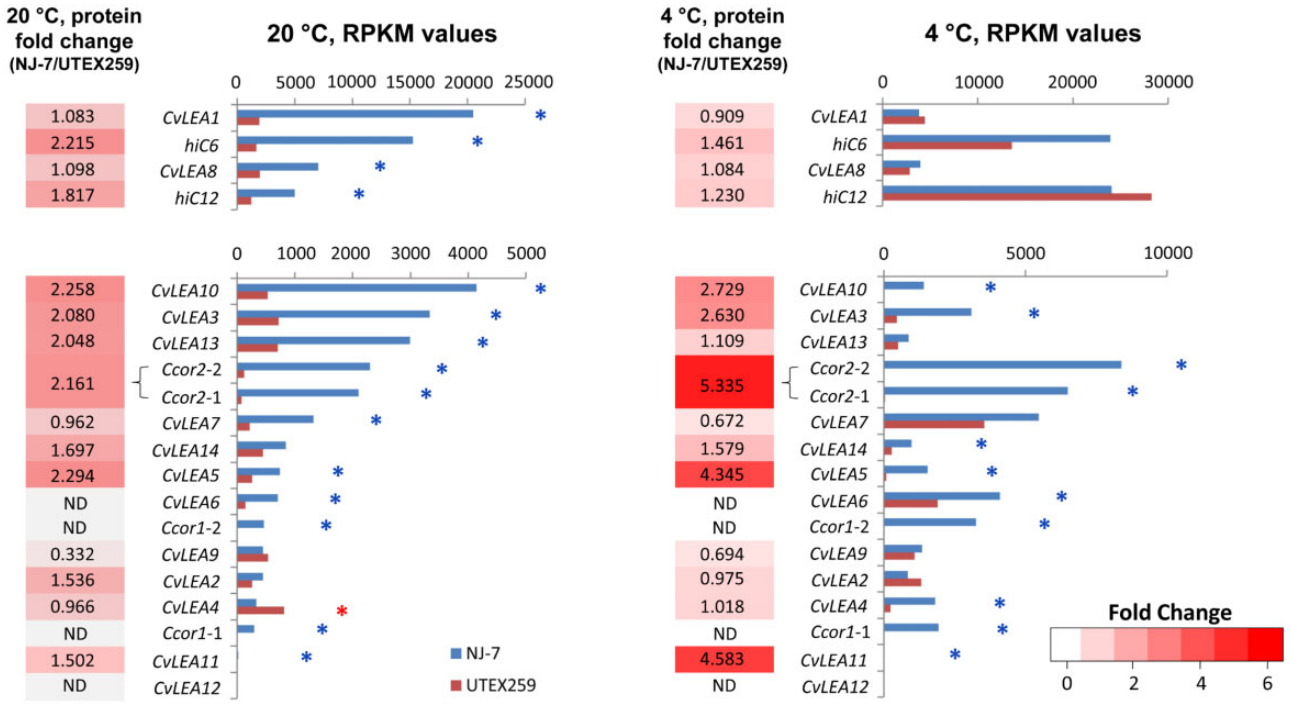

FIG. 5. Differential expression of LEA protein genes in NJ-7 and UTEX259 at 20 or 4 °C based on RNA-seq and proteomic analyses. ND, not detected. Different copies of *hiC6* could not be differentiated from each other at protein and mRNA levels; therefore they were treated as one gene. Different copies of *Ccor1* and *Ccor2* could be differentiated at mRNA level, but the two copies of *Ccor2* could not be differentiated at protein level. Data are means of three biological replicates, *P*-values are given in supplementary excels S2 and S3, Supplementary Material online. Blue and red asterisks indicate significantly higher mRNA levels (fold change $\geq$ 2.0, *P*-value $<$ 0.05) in NJ-7 and UTEX259, respectively.

exposure to cold treatment (3 °C) (Honjoh et al. 1995; Joh et al. 1995) and shows cold-induced freezing tolerance (Hatano et al. 1976). It is deduced that the ancestor of NJ-7 had possessed cold-inducible freezing tolerance before arrival at Antarctica, but the tolerance of NJ-7 was further developed to a higher level with much less dependence on cold induction (fig. 1). Even so, many LEA protein genes in NJ-7 remained to be cold-regulated. *hiC6*, *hiC12*, *CvLEA1*, *CvLEA2*, *CvLEA7*, and *CvLEA13* in NJ-7 showed weaker responses to cold than in UTEX259; *CvLEA14* showed similar responses in the two strains; the two *Ccor2* genes and *CvLEA3*, *CvLEA5*, *CvLEA9*, and *CvLEA11* even showed stronger responses in NJ-7 (supplementary fig. S17, Supplementary Material online).

Systematic Increases in Abundance of Metabolic Enzymes in NJ-7

In addition to the greatly enhanced freezing tolerance, NJ-7 acquired the capability to grow at 4 °C. How are metabolic activities maintained in NJ-7 at such a low temperature? Nitrate reductase (NR) is the most often used enzyme in studies of cold adaptation of green algae (Loppes et al. 1996; di Rigano et al. 2006). We first compared the amino acid sequences of NR (877-aa) in NJ-7 and UTEX259 but found only 19 substitutions (fig. 6*a*). Then, we assayed the NADH:NR activities in NJ-7 and UTEX259 at different temperatures. The two curves of NR activity versus temperature were similar to each other in shape, with the optimal temperature slightly shifted from 25 to 30 °C in UTEX259 to 25 °C in NJ-7 (fig. 6*b*). However, the NR activity in NJ-7 was significantly higher than that in UTEX259 at all the temperatures tested. The activity of NJ-7 NR at 4 °C was approximately equal to that of UTEX259 NR at 25–30 °C. This implied that NJ-7 probably has higher abundance of NR than UTEX259. Western blot analysis confirmed this idea (fig. 6*b*). The ratio of NR abundance in NJ-7 and UTEX259 (cultured at 20 °C) was 2.54 $\pm$ 0.62, very close to the ratio of NR activity, 2.70 $\pm$ 0.89. In the proteomic analysis data, the abundance of NR in NJ-7 was shown to be 1.36 $\pm$ 0.08-fold of that in UTEX259 at 20 °C and 1.37 $\pm$ 0.07-fold at 4 °C (fig. 6*c*). Compared with Western blot and enzyme activity analyses, proteomic analysis produced lower values for fold changes. This supported the use of fold change $\geq$1.3 (*P*-value $<$ 0.05) as the criterion for upregulation of protein abundance in the proteomic analysis. In addition to NR, proteomic analysis data also showed a higher abundance of nitrite reductase (NiR) in NJ-7 than in UTEX259 at 20 and 4 °C (fig. 6*c*). The higher abundance of NR and NiR in NJ-7 may not be necessarily associated with increased mRNA level (fig. 6*c*). In algal cells, NR converts NO_3^- into NO_2^-, then NiR converts NO_2^- into NH_4^+ for synthesis of glutamine. Increased abundance of NR and NiR in NJ-7 could significantly enhance the metabolism of nitrate. In other words, before NR and NiR are cold adapted in NJ-7, their concentrations are elevated to compensate for the low specific activities, so that nitrate can be utilized actively at low temperatures.

Many enzymes involved in other aspects of cell activities are also upregulated in NJ-7 relative to UTEX259

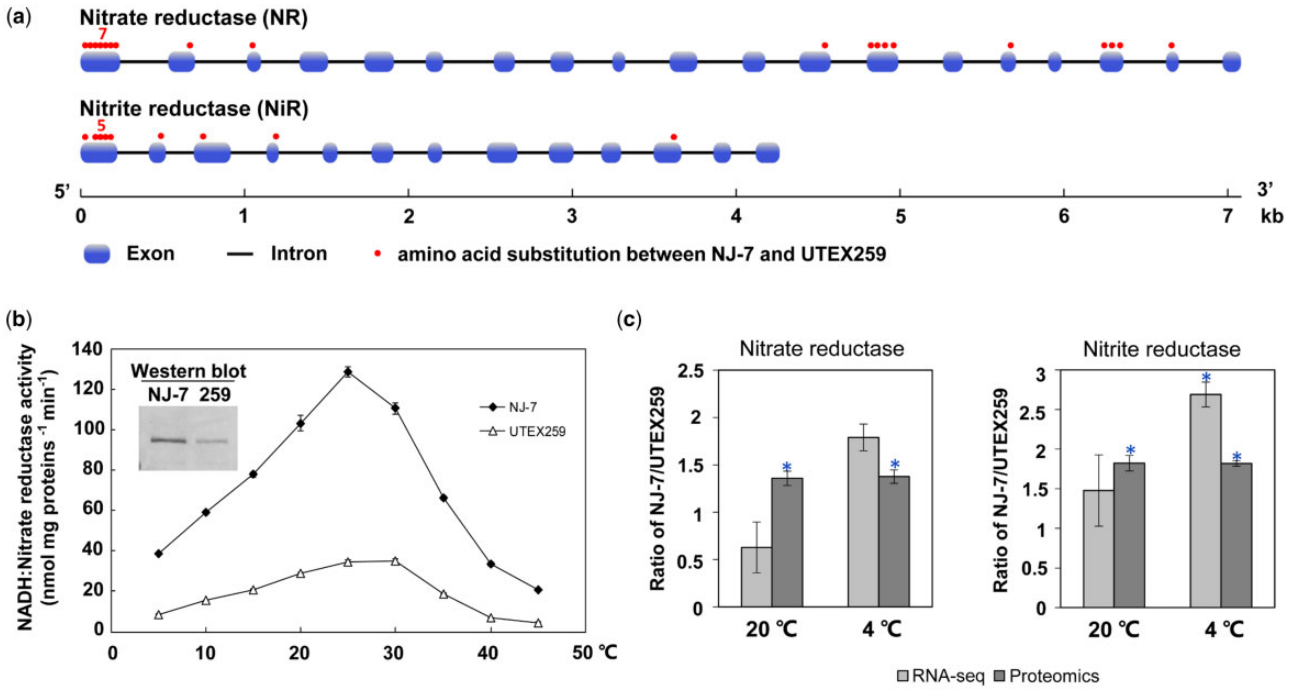

FIG. 6. Differential expression of NR and NiR genes in NJ-7 and UTEX259. (*a*) Structure of genes indicating amino acid substitutions. (*b*) Comparison of abundance and activity of NR in NJ-7 and UTEX259. Cells were cultured at 20 °C, and cell-free extracts were used in enzyme activity assays at different temperatures. The inset represents a typical result of Western blot analysis of NR in the two strains. (*c*) Differential expression of NR and NiR genes in the two strains as shown with RNA-seq and proteomic data. Asterisks indicate significantly higher expression in NJ-7 compared with UTEX259 (RNA-seq: fold change ≥ 2, *P*-value < 0.05; proteomic data: fold change ≥ 1.3, *P*-value < 0.05). Data are means $\pm$ SD of three biological replicates.

(supplementary excel S3, Supplementary Material online), especially some involved in carbon metabolism: for two critical steps in Calvin cycle, RuBisco activase/small subunit of RuBP carboxylase and sedoheptulose-1,7-bisphosphatase (supplementary fig. S18, Supplementary Material online); for glycolysis, the bifunctional enzyme phosphofructokinase (PFK-2)/fructose-2,6-bisphosphatase, phosphoglycerate kinase, phosphoglycerate mutase, and phosphate dikinase (fig. 7); linking glycolysis to TCA cycle or fatty acid synthesis, phosphoenolpyruvate carboxylase and pyruvate dehydrogenase E1 (α and β subunits) (fig. 7); for degradation of polysaccharides or oligosaccharides (including those covalently linked to proteins and lipids), α-mannosidase and seven other enzymes (fig. 7); for reutilization of monosaccharides in polysaccharide or oligosaccharide synthesis, UDP-*N*-acetylglucosamine pyrophosphorylase and three other enzymes (fig. 7). Differences in protein and mRNA levels of these genes between NJ-7 and UTEX259 are shown in supplementary figure S19, Supplementary Material online. No enzymes shown in figure 7 are downregulated in NJ-7. Taking genes for the enzymes in figure 7 as a gene set, we also performed GSEA of the differential expression between two strains at the protein level and found significant positive enrichments in NJ-7 at both temperatures (supplementary fig. S16*b*, Supplementary Material online). Systematic increases in abundance of critical enzymes would significantly accelerate carbon metabolisms at low temperatures. However, the early stage cold adaptation is still underway in NJ-7. For example, RuBisco activase and the small subunit of RuBP carboxylase increased in abundance, whereas the large subunit did not. Like LEA proteins, most of these enzymes from the two strains differ from each other by <3% amino acid residues (supplementary table S22, Supplementary Material online).

Discussion

Typical mesophilic microorganisms do not grow at near 0 °C temperatures, those transported by atmospheric circulation to Antarctica need to develop the cold-growth capability. To survive the deeper freezing conditions than encountered before, these species also need significantly higher levels of ice-binding proteins or LEA proteins. *Chlorella vulgaris* NJ-7 is such an example. Based on comparative omics data, we systematically analyzed the genetic divergence between NJ-7 and the temperate strain UTEX259 and deduced the underlying mechanism for the early stage cold adaptation of *C. vulgaris*.

The cold-growth capability may be developed in two steps: elevation of protein abundance and cold adaptation of enzymes. Adaptation of NJ-7 to the Antarctic cold environment is basically at the early stage—elevation of protein abundance. Positive selection of genes (leading to cold-adapted enzymes/proteins) should have contributed to the cold adaptation of NJ-7 (supplementary excel S1, Supplementary Material online) but is apparently not the dominant contributor. Instead, most homologous proteins in NJ-7 and UTEX259 are nearly identical to each other. Some of them, such as NR, NiR, and many enzymes involved in other metabolic pathways, showed higher abundance in NJ-7 than in UTEX259

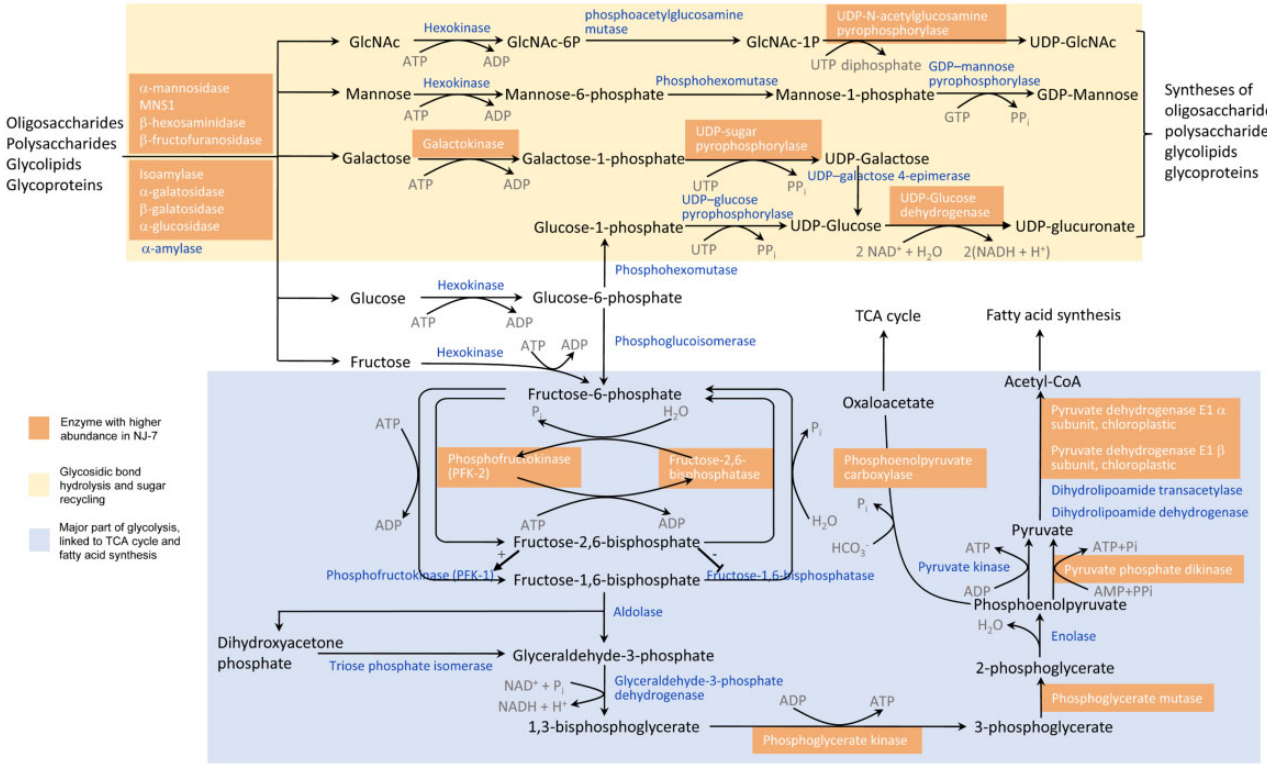

FIG. 7. Differential expression of enzymes involved in carbohydrate metabolism in NJ-7 and UTEX259. Biochemical reactions shown in this figure include hydrolysis of glycosidic linkages, reutilization of monosaccharides and glycolysis; two reactions (phosphoenolpyruvate to oxaloacetate, pyruvate to acetyl-CoA) that link glycolysis to TCA cycle and fatty acid synthesis are also included. Most of the high-lighted enzymes showed higher abundance in NJ-7 than in UTEX259 at both 20 and 4 °C (supplementary fig. S19, Supplementary Material online). MNS1, mannosyl-oligosaccharide 1,2-alpha-mannosidase; GlcNAc, *N*-acetyl-D-glucosamine; GlcNAc-6P, *N*-acetyl-D-glucosamine-6-phosphate; GlcNAc-1P, *N*-acetyl-D-glucosamine-1-phosphate.

irrespective of temperature (supplementary fig. S19 and excel S3, Supplementary Material online). The capability to grow at near 0 °C temperatures may depend on almost all aspects of cellular activities and involves a large number of enzymes/proteins. Relative to positive selection of enzymes for higher specific activities, increases in cellular concentrations would be a quick adaptation pathway for promoting biochemical reactions at low temperatures. Presumptively, the increase in cellular concentration is only necessary for enzymes whose activities are not sufficient to support cell proliferation and preferentially occurs at rate-limiting steps in metabolic pathways (such as fructose-2,6-bisphosphatase/phosphofructokinase in glycolysis shown in fig. 7 and sedoheptulose-1,7-bisphosphatase in Calvin cycle shown in supplementary fig. S18, Supplementary Material online).

There are different mechanisms for increasing the expression of enzymes/proteins. Gene duplication is one of the rapid mechanisms (Kondrashov 2012). In an Antarctic fish, augmented gene expression for cold adaptation is largely associated with gene duplication and family expansion (Chen et al. 2008). In microbes, experimentally evolved phenotypes (Wenger et al 2011) or development of drug resistance (Sandegren and Andersson 2009) may also be associated with gene amplifications. In NJ-7, however, the elevated abundance of enzymes is apparently dependent on the upregulation at transcription, posttranscription (RNA stability), translation, or posttranslation (protein stability) levels rather than increased gene copy numbers.

In coping with seasonally changing temperatures, mesophilic microorganisms developed the capability to relieve the cold stress through cold acclimation (Morgan-Kiss et al. 2006); based on "anticipation and associative learning" (Bleuven and Landry 2016), the cold acclimation also prepares for the freezing condition that may follow. In cold acclimation of *C. vulgaris*, sphingolipid delta(4)-desaturase, RNA helicases, and other enzymes are upregulated for resumption of cellular activities at 4 °C (predominantly for relief of cold stress), LEA proteins are upregulated to provide cryoprotection of cells at the "anticipated" subzero temperatures. Presumptively, the expression of some of these proteins/enzymes could have been further enhanced in NJ-7 after its arrival in Antarctica. A small proportion of cold-inducible proteins in NJ-7 indeed showed higher abundance than their orthologs in UTEX259 (fig. 3 and supplementary table S17, Supplementary Material online). However, most enzymes/proteins with higher abundance in NJ-7, such as those in figures 6 and 7, are not inducible in response to cold stress, because increases in their expression are not essential for cold stress relief but are required for cell proliferation at near 0 °C temperatures. The enhanced expression of "chromatin" proteins in NJ-7 (supplementary fig. S14*e*, Supplementary Material online) may also contribute to the cold-growth capability.

Relatively speaking, Antarctica is a permanently cold environment. In coastal ice-free regions, the ground surface temperature fluctuates at large amplitudes in December and January (Guglielmin 2006). These sites could allow newcomer mesophilic species to grow shortly every year. Some species may evolve to acquire the capability to grow at near 0 °C temperatures in a way like NJ-7. Because of the energy burden for keeping higher abundance of enzymes, selective pressure in permanently cold environments would continue to drive the cold adaptation in enzymes in subsequent evolutionary processes.

Materials and Methods

A short form of "Materials and methods" is presented as follows. More detailed descriptions are provided in supplementary text S1, Supplementary Material online.

Algal Strains, Culture Conditions, and Physiological Analyses

Chlorella vulgaris NJ-7 was isolated from rock samples collected (January 1999) near a transitory pond 5 km away from the Zhongshan Station (69°22′S–76°22′E) in Antarctica (Hu et al. 2008) and deposited at the Freshwater Algal Culture Collection of the Institute of Hydrobiology (FACHB2411). *Chlorella vulgaris* UTEX259 was purchased from the Culture Collection of Algae at The University of Texas at Austin (https://utex.org; last accessed November 24, 2019). Algal strains were purified by repeated streaking of single colonies on agar plates, and the axenicity was confirmed by microscopic examination and culture on solid Luria-Bertani medium and BG11 (Stanier et al. 1971) supplemented with glucose.

Chlorella strains were cultivated in BG11 under the light of 30 μE m^{-2} s^{-1} at 20 °C with aeration. Algal cells grown at 20 °C were rapidly cooled to 4 °C and exposed to the same temperature with aeration in a refrigerator with illumination of 30 μE m^{-2} s^{-1}. For comparison of their growth at 20 and 4 °C, cell density was monitored turbidimetrically at 730 nm. For evaluation of antifreeze capability, algal cells grown at 20 °C were exposed to 4 °C for 48 h or not. Cells collected by centrifugation were then rapidly cooled to −20 °C and frozen at −20 °C for 8 days. The frozen cells were diluted in liquid BG11 and allowed to grow on BG11 plates at 20 °C for 15 days. Colony-forming unit per OD_{730} milliliter was then calculated. For measurements of photosynthetic activities, algal cells cultured at 20 °C were used. Rates of photosynthetic oxygenic evolution were measured on a Clark-type oxygen electrode (Oxylab2, Hansatech, United Kingdom) with a saturating light of 2,000 μE m^{-2} s^{-1} at different temperatures.

For extraction of genomic DNA, cells cultured at 20 °C were used. For transcriptome analysis, RNA was extracted from a variety of culture conditions to increase the number of expressed genes. These conditions included different growth temperatures (20 and 28 °C), exposure to cold (4 °C for 24 h), salt (0.3 M NaCl for 24 h), or oxidation (0.2 mM H_2O_2 for 5 h) stress, exposure to N, P, C, Fe, Ca, Mg, K, or trace element starvation (48 h). For differential expression analyses, total RNA and proteins were extracted from algal cells cultured at 20 °C with or without exposure to 4 °C, harvested and frozen in liquid nitrogen. Considering the different rates for accumulation of mRNA and proteins of cold-induced genes (Liu et al. 2011), cells exposed to 4 °C were collected at 6 h for RNA extraction and at 24 h for protein extraction.

All physiological and differential expression (transcriptomic, proteomic, RT-qPCR) data were calculated from results of three biological replicates.

Genome Assembly and Annotation

High-molecular-weight genomic DNA was extracted using the cetyltrimethyl ammonium bromide method (Murray and Thompson 1980) with modifications. For 454 sequencing, shotgun and paired-end libraries were constructed. For Illumina sequencing, two short-insert paired-end genomic DNA libraries and a long-insert mate-pair library were constructed. Total RNA was extracted using Trizol reagent (Invitrogen) from cells cultured/treated under a variety of conditions to increase the number of expressed genes, and the pooled total RNA was used to construct the cDNA library for 454 sequencing.

The NJ-7 genome assembly was generated from 454 shotgun, 454 paired-end and Illumina GAIIx paired-end reads, using Newbler (GS De Novo Assembler, Roche) and Velvet (Zerbino and Birney 2008). The UTEX259 genome assembly was generated from 454 shotgun, 454 paired-end, Illumina GAIIx paired-end, Illumina MiSeq paired-end and mate pair reads, using Newbler, Velvet, ALLPATHS-LG assembler (Gnerre et al. 2011) and the scaffolder SSPACE (Boetzer et al. 2011).

De novo transcriptome assembly was performed using Newbler with parameters -cdna -urt -tr, generating isotigs that represent transcripts. Genome assemblies were further improved by L_RNA_scaffolder (Xue et al. 2013), which uses the transcripts to order and join genome sequences into larger scaffolds, and GapCloser tool in SOAPdenovo (Li et al. 2010), which makes use of the information of paired-end Illumina reads to fill gaps within scaffolds. Short scaffolds (<500 bp) and organelle sequences were excluded from the assemblies of nuclear genomes.

The quality of NJ-7 and UTEX259 genome assemblies was assessed in multiple ways, including BUSCO (Benchmarking Universal Single-Copy Orthologs) completeness assessment (Simao et al. 2015), mapping of fosmid-end sequences, and assembled transcripts onto the genome sequences.

Scaffolds and contigs corresponding to organelle genomes were extracted from the genome assemblies using BlastN searches against the *C. vulgaris* C-27 chloroplast genome (GenBank accession number AB001684) or *Prototheca wickerhamii* mitochondrial genome (GenBank accession number NC_001613). The gaps were filled by sequencing of PCR fragments, producing circular organelle genomes.

Prediction of protein-coding genes was performed using both ab initio gene predictions and transcript-based approaches, followed by integrating them using EVidenceModeler (EVM) (Haas et al. 2008). Protein motifs

and domains of predicted gene models were annotated using InterProScan search (Zdobnov and Apweiler 2001) against InterPro databases (Hunter et al. 2009). Gene functions were assigned based on BLASTP searches (*E*-value $\leq$ 1e-5). tRNA and rRNA genes were identified using tRNAscan-SE v1.3 (Lowe and Eddy 1997) and RNAmmer 1.2 (http://www.cbs.dtu.dk/services/RNAmmer/; last accessed November 24, 2019). Organelle genomes were annotated using DOGMA (http://dogma.ccbb.utexas.edu/; last accessed November 24, 2019), ORF-Finder (https://www.ncbi.nlm.nih.gov/orffinder/; last accessed November 24, 2019), and BLASTX searches based on sequence similarity to genes in other annotated organelle genomes.

Synteny and Comparative Analyses

A BLASTP alignment (*E*-value $\leq$ 1e-5) was performed between all protein sequences of the two strains. OrthoMCL (Li et al. 2003) was applied to identify orthologous protein pairs based on the BLASTP results, with the criteria *E*-value $\leq$1e-5, identity $\geq$80%, and length coverage $\geq$70%.

Synteny between genomes was identified by aligning genomic regions using orthologous gene pairs as anchors. Using MCScanX (Wang et al. 2012), pairwise syntenic segments were identified, grouped into blocks, and displayed by a dot plot graph. A dual synteny plot was generated with the longest 12 scaffolds of NJ-7 and the corresponding scaffolds of UTEX259. Nine chlorophyte chloroplast genomes were compared using the Progressive Mauve algorithm implemented in Mauve 2.4.0 (Darling et al. 2010).

Phylogenetic and Evolutionary Analyses

i. *Phylogenetic analysis*. Phylogenetic trees were constructed based on 1,080 single-copy genes shared by NJ-7, UTEX259, and other nine chlorophytes, with the land plant *Arabidopsis thaliana* as the outgroup. Using the MUSCLE program (Edgar 2004), 1,080 genes from these species/strains were aligned. Poorly aligned regions were removed from multiple-alignments using Gblocks implemented in TranslatorX (Abascal et al. 2010). The cleaned multiple alignments were concatenated and subjected to ML (maximum likelihood) tree reconstruction using PhyML (Guindon et al. 2010).

ii. *Divergence time estimates*. Divergence times between lineages were estimated using the Bayesian MCMC approach as implemented in the program BEAST v1.7.5 (Drummond et al. 2012) based on 36 chloroplast genes (*atpA*, *atpE*, *atpF*, *atpH*, *atpI*, *petA*, *petB*, *petG*, *psaA*, *psaB*, *psaC*, *psaJ*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbL*, *psbN*, *rbcL*, *rpl2*, *rpl14*, *rpl16*, *rpl20*, *rpl23*, *rps7*, *rps8*, *rps11*, *rps12*, *rps14*, *rps18*, and *rps19*). Two sets of analyses were performed: broad-scale analyses with 12–19 species/strains of green algae (in four combinations) and four species of higher plants; a fine-scale analysis with nine species/strains of *Chlorella*. BEAST used concatenated nucleotide alignments as input and estimated phylogeny and divergence times simultaneously. To calibrate nodes in the broad-scale tree, three reliable fossil dates were used: the divergence between angiosperms and gymnosperms (290–320 Ma) (Goremykin et al. 1997; Doyle 1998), the divergence between Nymphaeales and eudicots (115 Ma minimum) (Friis et al. 2001), and the divergence between monocots and dicots (90–130 Ma) (Crane et al. 1995). The maximum clade credibility tree was generated using TreeAnnotator v1.7.5 (Drummond et al. 2012) and visualized in FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/; last accessed November 24, 2019). The estimated divergence time between *C. variabilis* and *C. vulgaris* was further used to calibrate the fine-scale analysis.

iii. *Evolutionary rate estimates*. The number of nonsynonymous substitutions per nonsynonymous site (d_N) and the number of synonymous substitutions per synonymous site (d_S) were estimated based on the 1,080 nuclear genes, using codeml program in the PAML package v4.6 (Yang 2007).

Detection of Genes with Positive Selection

The codeml program in the PAML package was applied to all the 5,899 core genes shared by NJ-7, UTEX259, and NC64A. Codeml analysis with branch-site model was conducted on either NJ-7 or UTEX259 branch. The null hypothesis (assuming $\omega = 1$ or $\omega < 1$) and the alternative hypothesis (assuming $\omega > 1$) were compared to identify significantly higher likelihood values for the alternative hypothesis. Those genes with significance (χ^2 *P*-value < 0.01) were deemed to be under positive selection.

Analyses of Alternative Splicing Events, Repetitive Elements, Gene Duplication, and Gene Families

Alternative splicing events were detected by comparing the predicted genes on genome and the assembled isotigs and processing the clusters of aligned isotigs with at least one splice site.

Repetitive elements were identified with both homology-based methods and de novo repeat finding programs. The library of identified repeats was used to estimate repeat contents of each genome.

Duplicated genes were identified by performing self-versus-self BLASTP on protein sequences. The copy number of recent duplicated genes was deduced based on highly similar genes that were detected by performing self-versus-self BlastN, or roughly estimated by aligning trimmed Illumina reads onto gene sequences to calculate the relative depth of coverage.

Homologous protein families were constructed based on genome sequences of 11 species/strains of green algae, and all predicted protein sequences were compared using all-against-all BLASTP. Protein families were annotated according to Pfam domains or InterPro descriptions. The significance of family expansion or reduction was analyzed by chi-square test in R.

Identification of LEA Protein Genes

In addition to *Ccor1*/*Ccor2* identified by experimental analyses (Liu et al. 2011) and genes similar to *Ccor1*/*Ccor2*, LEA protein genes were identified by searching for LEA motifs and

secondary structures (Tunnacliffe and Wise 2007; Battaglia et al. 2008) using the fuzzpro program in EMBOSS package (Rice et al. 2000) and the PSIPRED Server (http://bioinf.cs.ucl.ac.uk/psipred; last accessed November 24, 2019), searching for Pfam domains using hmmsearch (Eddy 2011) against LEA HMM profiles, searching for homologs using BLASTP (*E*-value $\leq$ 1e-5) against LEA proteins in the Late Embryogenesis Abundant Proteins DataBase (http://forge.info.univ-angers.fr/~gh/Leadb/index.php; last accessed November 24, 2019). Subcellular locations were predicted using TargetP (http://www.cbs.dtu.dk/services/TargetP; last accessed November 24, 2019), MitoProt (https://ihg.gsf.de/ihg/mitoprot. html; last accessed November 24, 2019), and PSORT (http://psort1.hgc.jp/form.html; last accessed November 24, 2019).

Analyses of Differential Expression at mRNA Level

Differential expression at mRNA level was analyzed based on RNA-seq. Processed sequencing reads were aligned to exonic regions of predicted genes using Bowtie (Langmead et al. 2009) and Tophat (Trapnell et al. 2009) allowing one mismatch. Raw count data and RPKM (reads per kilobase of exon per million mapped reads) values were generated from the alignment files using bam2rpkm (http://bam2rpkm.sourceforge.net/; last accessed November 24, 2019). Raw count data were then used as input into DESeq (Anders and Huber 2010) for differential expression analyses (three biological replicates). Differentially expressed genes were identified with criteria: fold-change $\geq$ 2 (either up- or downregulated), FDR (false discovery rate) adjusted *P*-value $<$ 0.05, and RPKM $\geq$ 10 under at least one condition.

Real-Time Quantitative Polymerase Chain Reactions

DNA-free RNA was used to synthesize the first strand of cDNA using PrimeScript RT reagent Kit (Takara) and oligo (dT)15 primer (Promega). SYBR Green I (Takara) was added to the PCR reaction mixture according to manufacturers' protocol, and RT-qPCR was performed on Applied Biosystems ABI 7500. β-Actin was used as the internal standard. Primers were designed based on identical sequences of orthologous genes (supplementary table S23, Supplementary Material online).

Analyses of Differential Expression at Protein Level

Differential expression at the protein level was analyzed using the quantitative proteomics approach. Algal cells were sonicated and centrifuged. Proteins in the supernatant were precipitated and subjected to trypsin digestion. After labelling with Tandem Mass Tags/Isobaric Tag for Relative Absolute Quantitation (TMT Kit/iTRAQ), peptides were analyzed by liquid chromatography coupled with tandem mass spectrometry (MS/MS). The resulting MS/MS data were processed using Maxquant search engine (v.1.5.2.8) (Cox and Mann 2008). Peptide and protein ratios were obtained by direct comparison of signals of "light" and "heavy" isotope in the same liquid chromatography run. Cross-strain comparison of expression of orthologous proteins was performed using identical peptides instead of whole protein sequences. In cases where multiple copies of proteins could not be distinguished, the multiple copies were treated as one in the quantitative analyses. Differential expression with ratio $\geq$1.3 (*P*-value $<$ 0.05) was defined as upregulation, that with ratio $\leq$0.77 (*P*-value $<$ 0.05) as downregulation.

GO Enrichment Analysis

GSEA software version 3.0 (Subramanian et al. 2005) was used to identify significantly enriched GO gene sets. GSEA was run with 1,000 permutations, and functional gene sets with FDR-adjusted *q*-value $<$0.05 were considered significant.

Analyses of Regulation Networks

TFs were predicted by searching for Pfam domains using hmmsearch against the HMM profiles for TF families at PlantTFDB (Jin et al. 2014). DNA-binding motifs for orthologous TFs in *C. variabilis* NC64A were used to scan for recognition sites upstream of protein-coding genes in NJ-7 and UTEX259 by FIMO (Grant et al. 2011). The potential role of a TF in gene regulation was defined based on the correlation between its own expression and the expression of genes with its recognition site.

NR Assays and Western Blot Analysis

NR activity in crude extracts was assayed as described by di Rigano et al. (2006) using NADH as electron donor, with a few modifications. Reactions (three technical repeats) were carried out for 30 min at temperatures from 5 to 45 °C.

Western blot analysis of total soluble proteins was performed using the rabbit antiserum against the recombinant NR.

Data Availability

Annotations of NJ-7 and UTEX259 genomes are available in the Figshare repository (DOI: 10.6084/m9.figshare.c.4678916; https://figshare.com/s/66909bf96c4a06159c94).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

The authors express great gratitude to Guoxiang Liu for his advice on the taxonomic analysis of strains of *Chlorella vulgaris* based on the ribosomal RNA region. This research was supported by the Knowledge Innovation Project of Hubei Province (2017CFA021), the STS Project of Chinese Academy of Sciences (KFJ-SW-STS-163), the national special support program "Wan-Ren-Ji-Hua" of China and the State Key Laboratory of Freshwater Ecology and Biotechnology at IHB, CAS (2019FBZ06).

Author Contributions

X.X., J.X., and A.-Y.G. designed the project. Y.W. performed bioinformatic analyses. H.-M.Z. contributed to genome assembly/annotation, multicopy gene analyses, and identification of LEA protein genes. X.L. performed experimental analyses of the expression and activity of nitrate reductase;

Y.W. and H.G. analyzed the physiological characteristics of algal strains; Y.W. prepared DNA and RNA samples, performed PCR and RT-qPCR analyses. X.X., Y.W., X.L., H.G., A.-Y.G., and J.X. interpreted the results. X.X., J.X., and A.-Y.G. edited the manuscript. X.X. and Y.W. wrote the manuscript. All authors read and approved the content of the manuscript.

References

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res*. 38(2 Suppl):W7–W13.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol*. 11(10):R106.

Archer SDJ, Lee KC, Caruso T, Maki T, Lee CK, Cary SC, Cowan DA, Maestre FT, Pointing SB. 2019. Airborne microbial transport limitation to isolated Antarctic soil habitats. *Nat Microbiol*. 4(6):925–932.

Bar Dolev M, Braslavsky I, Davies PL. 2016. Ice-binding proteins and their function. *Annu Rev Biochem*. 85(1):515–542.

Battaglia M, Olvera-Carrillo Y, Garciarrubio A, Campos F, Covarrubias AA. 2008. The enigmatic LEA proteins and other hydrophilins. *Plant Physiol*. 148(1):6–24.

Bleuven C, Landry CR. 2016. Molecular and cellular bases of adaptation to a changing environment in microorganisms. *Proc R Soc B*. 283(1841):20161458.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27(4):578–579.

Burton-Johnson A, Black M, Fretwell PT, Kaluza-Gilbert J. 2016. An automated methodology for differentiating rock from snow, clouds and sea in Antarctica from Landsat 8 imagery: a new rock outcrop map and area estimation for the entire Antarctic continent. *Cryosphere* 10(4):1665–1677.

Cáliz J, Triadó-Margarit X, Camarero L, Casamayor EO. 2018. A long-term survey unveils strong seasonal patterns in the airborne microbiome coupled to general and regional atmospheric circulations. *Proc Natl Acad Sci U S A*. 115(48):12229–12234.

Chen Z, Cheng C-HC, Zhang J, Cao L, Chen L, Zhou L, Jin Y, Ye H, Deng C, Dai Z, et al. 2008. Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proc Natl Acad Sci U S A*. 105(35):12944–12949.

Collins T, Margesin R. 2019. Psychrophilic lifestyles: mechanisms of adaptation and biotechnological tools. *Appl Microbiol Biotechnol*. 103(7):2857–2871.

Convey P, Chown SL, Clarke A, Barnes DK, Bokhorst S, Cummings VJ, Ducklow H, Frati F, Green TGA, Gordon S, et al. 2014. The spatial structure of Antarctic biodiversity. *Ecol Monogr*. 84(2):203–244.

Convey P, Stevens MI. 2007. Antarctic biodiversity. *Science* 317(5846):1877–1878.

Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 26(12):1367–1372.

Crane PR, Friis EM, Pedersen KR. 1995. The origin and early diversification of angiosperms. *Nature* 374(6517):27–33.

Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5(6):e11147.

Davey MP, Norman L, Sterk P, Huete-Ortega M, Bunbury F, Loh BKW, Stockton S, Peck LS, Convey P, Newsham KK, et al. 2019. Snow algae communities in Antarctica: metabolic and taxonomic composition. *New Phytol*. 222(3):1242–1255.

di Rigano VM, Vona V, Lobosco O, Carillo P, Lunn JE, Carfagna S, Esposito S, Caiazzo M, Rigano C. 2006. Temperature dependence of nitrate reductase in the psychrophilic unicellular alga *Koliella antarctica* and the mesophilic alga *Chlorella sorokiniana*. *Plant Cell Environ*. 29(7):1400–1409.

Doyle JA. 1998. Molecules, morphology, fossils, and the relationship of angiosperms and Gnetales. *Mol Phylogenet Evol*. 9(3):448–462.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*. 29(8):1969–1973.

Dure L, Greenway SC, Galau GA. 1981. Developmental biochemistry of cotton seed embryogenesis and germination: changing messenger ribonucleic acid populations as shown by *in vitro* and *in vivo* protein synthesis. *Biochemistry* 20(14):4162–4168.

Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol*. 7(10):e1002195.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32(5):1792–1797.

Feller G, Gerday C. 2003. Psychrophilic enzymes: hot topics in cold adaptation. *Nat Rev Microbiol*. 1(3):200–208.

Fraser CI, Nikula R, Ruzzante DE, Waters JM. 2012. Poleward bound: biological impacts of southern hemisphere glaciation. *Trends Ecol Evol*. 27(8):462–471.

Fretwell P, Pritchard H, Vaughan D, Bamber JL, Barrand NE, Bell R, Bianchi C, Bingham RG, Blankenship DD, Casassa G, et al. 2013. Bedmap2: improved ice bed, surface and thickness datasets for Antarctica. *Cryoshere* 7(1):375–393.

Friis EM, Pedersen KR, Crane PR. 2001. Fossil evidence of water lilies (Nymphaeales) in the early cretaceous. *Nature* 410(6826):357–360.

Galeotti S, DeConto R, Naish T, Stocchi P, Florindo F, Pagani M, Barrett P, Bohaty SM, Lanci L, Pollard D, et al. 2016. Antarctic Ice Sheet variability across the Eocene-Oligocene boundary climate transition. *Science* 352(6281):76–80.

Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, et al. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A*. 108(4):1513–1518.

Goremykin VV, Hansmann S, Martin WF. 1997. Evolutionary analysis of 58 proteins encoded in six completely sequenced chloroplast genomes: revised molecular estimates of two seed plant divergence times. *Plant Syst Evol*. 206(1–4):337–351.

Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27(7):1017–1018.

Guglielmin M. 2006. Ground Surface Temperature (GST), Active layer and permafrost monitoring in continental Antarctica. *Permafr Periglac Process*. 17(2):133–143.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 59(3):307–321.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol*. 9(1):R7.

Hatano S, Sadakane H, Tutumi M, Watanabe T. 1976. Studies on frost hardiness in *Chlorella ellipsoidea* I. Development of frost hardiness of *Chlorella ellipsoidea* in synchronous culture. *Plant Cell Physiol*. 17:451–458.

Herron MD, Hackett JD, Aylward FO, Michod RE. 2009. Triassic origin and early radiation of multicellular volvocine algae. *Proc Natl Acad Sci U S A*. 106(9):3254–3258.

Honjoh K-I, Shimizu H, Nagaishi N, Matsumoto H, Suga K, Miyamoto T, Iio M, Hatano S. 2001. Improvement of freezing tolerance in transgenic tobacco leaves by expressing the *hiC6* gene. *Biosci Biotechnol Biochem*. 65(8):1796–1804.

Honjoh K-I, Yoshimoto M, Joh T, Kajiwara T, Miyamoto T, Hatano S. 1995. Isolation and characterization of hardening-induced proteins in *Chlorella vulgaris* C-27: identification of late embryogenesis abundant proteins. *Plant Cell Physiol*. 36:1421–1430.

Hu H, Li H, Xu X. 2008. Alternative cold response modes in *Chlorella* (Chlorophyta, Trebouxiophyceae) from Antactica. *Phycologia* 47(1):28–34.

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, et al. 2009. InterPro: the integrative protein signature database. *Nucleic Acids Res*. 37(Database):D211–D215.

Jin JP, Zhang H, Kong L, Gao G, Luo JC. 2014. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res*. 42(D1):D1182–D1187.

Joh T, Honjoh K-I, Yoshimoto M, Funabashi J, Miyamoto T, Hatano S. 1995. Molecular cloning and expression of hardening-induced genes in *Chlorella vulgaris* C-27: the most abundant clone encodes a late embryogenesis abundant protein. *Plant Cell Physiol*. 36:85–93.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484(7392):55–61.

Karl DM, Bird DF, Björkman K, Houlihan T, Shackelford R, Tupas L. 1999. Microorganisms in the accreted ice of Lake Vostok, Antarctica. *Science* 286(5447):2144–2147.

Kondrashov FA. 2012. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc R Soc B*. 279(1749):5048–5057.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 10(3):R25.

Lear CH, Lunt DJ. 2016. How Antarctica got its ice. *Science* 352(6281):34–35.

Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 13(9):2178–2189.

Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, et al. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res*. 20(2):265–272.

Liao Y, Frederick JE. 2005. The ultraviolet radiation environment of high southern latitudes: springtime behavior over a decadal timescale. *Photochem Photobiol*. 81(2):320–324.

Liao M-L, Somero GN, Dong Y-W. 2019. Comparing mutagenesis and simulations as tools for identifying functionally important sequence changes for protein thermal adaptation. *Proc Natl Acad Sci U S A*. 116(2):679–688.

Liu X, Wang Y, Gao H, Xu X. 2011. Identification and characterization of genes encoding two novel LEA proteins in Antarctic and temperate strains of *Chlorella vulgaris*. *Gene* 482(1–2):51–58.

Loppes R, Devos N, Willem S, Barthelemy P, Matagne R. 1996. Effect of temperature on two enzymes from a psychrophilic *Chloromonas* (Chlorophyta). *J Phycol*. 32(2):276–278.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 25(5):955–964.

Luttgeharm KD, Kimberlin AN, Cahoon EB. 2016. Plant sphingolipid metabolism and function. In: Nakamura Y, Li-Beisson Y, editors. Lipids in plant and algae development. Cham/Heidelberg/New York/Dordrecht/London: Springer. p. 249–286.

Mayol E, Arrieta JM, Jiménez MA, Martínez-Asensio A, Garcias-Bonet N, Dachs J, González-Gaya B, Royer S-J, Benítez-Barrios VM, Fraile-Nuez E, et al. 2017. Long-range transport of airborne microbes over the global tropical and subtropical ocean. *Nat Commun*. 8(1):201.

Morgan-Kiss RM, Priscu JC, Pocock T, Gudynaite-Savitch L, Huner NP. 2006. Adaptation and acclimation of photosynthetic microorganisms to permanently cold environments. *Microbiol Mol Biol Rev*. 70(1):222–252.

Murray MG, Thompson WF. 1980. Rapid isolation of high molecular weight plant DNA. *Nucl Acids Res*. 8:4321–4325.

National Polar Research Institute (Japan). 1991. Antarctic meteorology (in Chinese). Beijing (China): China Ocean Press.

Possmayer M, Gupta RK, Szyszka-Mroz B, Maxwell DP, Lachance M-A, Hüner NPA, Smith DR. 2016. Resolving the phylogenetic relationship between *Chlamydomonas* sp. UWO 241 and *Chlamydomonas raudensis* SAG 49.72 (Chlorophyceae) with nuclear and plastid DNA sequences. *J Phycol*. 52(2):305–310.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends Genet*. 16(6):276–277.

Saavedra HG, Wrabl JO, Anderson JA, Li J, Hilser VJ. 2018. Dynamic allostery can drive cold adaptation in enzymes. *Nature* 558(7709):324–329.

Sandegren L, Andersson DI. 2009. Bacterial gene amplification: implications for the evolution of antibiotic resistance. *Nat Rev Microbiol*. 7(8):578–588.

Santiago M, Ramírez-Sarmiento CA, Zamora RA, Parra LP. 2016. Discovery, molecular mechanisms, and industrial applications of cold-active enzymes. *Front Microbiol*. 7:1408.

Sasaki K, Christov NK, Tsuda S, Imai R. 2014. Identification of a novel LEA protein involved in freezing tolerance in wheat. *Plant Cell Physiol*. 55(1):136–147.

Shih M-D, Hoekstra FA, Hsing Y-I. 2008. Late embryogenesis abundant proteins. *Adv Bot Res*. 48:211–255.

Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.

Stanier RY, Kunisawa R, Mandel M, Cohen-Bazire G. 1971. Purification and properties of unicellular blue-green algae (order Chroococcales). *Bacteriol Rev*. 35(2):171–205.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 102(43):15545–15550.

Thomas DN, Dieckmann GS. 2002. Antarctic sea ice-a habitat for extremophiles. *Science* 295(5555):641–644.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105–1111.

Tunnacliffe A, Wise MJ. 2007. The continuing conundrum of the LEA proteins. *Naturwissenschaften*. 94(10):791–812.

Wang X, Zhang L, Zhang Y, Bai Z, Liu H, Zhang D. 2017. *Triticum aestivum* WRAB18 functions in plastids and confers abiotic stress tolerance when overexpressed in *Escherichia coli* and *Nicotiania benthamiana*. *PLoS One* 12(2):e0171340.

Wang Y, Liu X, Gao H, Xu X. 2011. Characterization of the tandem-arrayed *hiC6* genes in Antarctic and temperate strains of *Chlorella vulgaris*. *FEMS Microbiol Lett*. 325(2):130–139.

Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T-H, Jin H, Marler B, Guo H, et al. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 40(7):e49.

Wenger JW, Piotrowski J, Nagarajan S, Chiotti K, Sherlock G, Rosenzweig F. 2011. Hunger artists: yeast adapted to carbon limitation show trade-offs under carbon sufficiency. *PLoS Genet*. 7(8):e1002202.

Xue W, Li JT, Zhu YP, Hou GY, Kong XF, Kuang YY, Sun XW. 2013. L_RNA_scaffolder: scaffolding genomes with transcripts. *BMC Genomics* 14(1):604.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24(8):1586–1591.

Zdobnov EM, Apweiler R. 2001. InterProScan–an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17(9):847–848.

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 18(5):821–829.

Zhang X, Lu S, Jiang C, Wang Y, Lv B, Shen J, Ming F. 2014. RcLEA, a late embryogenesis abundant protein gene isolated from *Rosa chinensis*, confers tolerance to *Escherichia coli* and *Arabidopsis thaliana* and stabilizes enzyme activity under diverse stresses. *Plant Mol Biol*. 85(4–5):333–347.

Zhao HY, Feng H. 2018. Engineering *Bacillus pumilus* alkaline serine protease to increase its low-temperature proteolytic activity by directed evolution. *BMC Biotechnol*. 18(1):34.

Zhou Y, Zeng L, Fu X, Mei X, Cheng S, Liao Y, Deng R, Xu X, Jiang Y, Duan X, et al. 2016. The sphingolipid biosynthetic enzyme sphingolipid delta8 desaturase is important for chilling resistance of tomato. *Sci Rep*. 6(1):38742.