Acidophilic green algal genome provides insights into adaptation to an acidic environment

Shunsuke Hirooka[a,b,1], Yuu Hirose[c], Yu Kanesaki[b,d], Sumio Higuchi[e], Takayuki Fujiwara[a,b,f], Ryo Onuma[a], Atsuko Era[a,b], Ryudo Ohbayashi[a], Akihiro Uzuka[a,f], Hisayoshi Nozaki[g], Hirofumi Yoshikawa[b,h], and Shin-ya Miyagishima[a,b,f,1]

[a]Department of Cell Genetics, National Institute of Genetics, Shizuoka 411-8540, Japan; [b]Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Saitama 332-0012, Japan; [c]Department of Environmental and Life Sciences, Toyohashi University of Technology, Aichi 441-8580, Japan; [d]NODAI Genome Research Center, Tokyo University of Agriculture, Tokyo 156-8502, Japan; [e]Research Group for Aquatic Plants Restoration in Lake Nojiri, Nojiriko Museum, Nagano 389-1303, Japan; [f]Department of Genetics, Graduate University for Advanced Studies, Shizuoka 411-8540, Japan; [g]Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo 113-0033, Japan; and [h]Department of Bioscience, Tokyo University of Agriculture, Tokyo 156-8502, Japan

Edited by Krishna K. Niyogi, Howard Hughes Medical Institute, University of California, Berkeley, CA, and approved August 16, 2017 (received for review April 28, 2017)

Some microalgae are adapted to extremely acidic environments in which toxic metals are present at high levels. However, little is known about how acidophilic algae evolved from their respective neutrophilic ancestors by adapting to particular acidic environments. To gain insights into this issue, we determined the draft genome sequence of the acidophilic green alga *Chlamydomonas eustigma* and performed comparative genome and transcriptome analyses between *C. eustigma* and its neutrophilic relative *Chlamydomonas reinhardtii*. The results revealed the following features in *C. eustigma* that probably contributed to the adaptation to an acidic environment. Genes encoding heat-shock proteins and plasma membrane H^+-ATPase are highly expressed in *C. eustigma*. This species has also lost fermentation pathways that acidify the cytosol and has acquired an energy shuttle and buffering system and arsenic detoxification genes through horizontal gene transfer. Moreover, the arsenic detoxification genes have been multiplied in the genome. These features have also been found in other acidophilic green and red algae, suggesting the existence of common mechanisms in the adaptation to acidic environments.

environmental adaptation | acidic environment | acidophilic alga | comparative genomics | comparative transcriptomics

Several eukaryotic microalgae have been identified in acidic environments (pH <4.0) such as acid mine drainage (AMD) and geothermal hot springs (1). In this pH range, cyanobacteria are not present, and only acidophilic eukaryotic phototrophs are capable of photosynthesis (Fig. 1) (2, 3). The extremely low pH of these waters is due to the dissolution and oxidation of sulfur that is exposed to water and oxygen and produces sulfuric acid (4). The low pH facilitates metal solubility in water; therefore, acidic waters tend to have high concentrations of metals (5). Thus, acidophilic eukaryotic algae usually possess the ability to cope with toxic heavy metals in addition to low pH, both of which are lethal to most eukaryotes (2). Acidophilic algae are distributed throughout different branches of the eukaryotes, such as in red and green algae, stramenopiles, and euglenids. In most cases, neutrophilic relatives have been identified, suggesting that acidophilic algae evolved from their respective neutrophilic ancestors multiple times independently (6). However, it is largely unknown how several lineages of algae have successfully adapted to their acidic environments.

Thus, far, the genomes of three related thermo-acidophilic red algae, *Cyanidioschyzon merolae* (7), *Galdieria sulphuraria* (8), and *Galdieria phlegrea* (9), have been sequenced (all belong to the cyanidialean red algae, which inhabit sulfuric hot springs worldwide and grow optimally at 40–45 °C and pH 2–3). Genomic analyses showed that horizontal gene transfer (HGT) from environmental prokaryotes, the expansion of gene families, and the loss of genes have probably played important roles in the adaptation of Cyanidiales to acidic and high-temperature environments (8). Through HGT, cyanidialean red algae acquired arsenical-resistance efflux pumps that biotransform arsenic and archaeal ATPases, which probably contribute to the algal heat tolerance (8). In addition, the reduction in the number of genes encoding voltage-gated ion channels and the expansion of chloride channel and chloride carrier/channel families in the genome has probably contributed to the algal acid tolerance (8). Likewise, a study in the acidophilic green alga *Chlamydomonas acidophila* showed that phytochelatin synthase genes of bacterial HGT origin played an important role in the tolerance to cadmium (10).

However, the genomes of acidophilic algae other than cyanidialean red algae have not been sequenced. The green and red algae diverged relatively soon after the emergence of primitive eukaryotic algae (11). In addition, comparisons with neutrophilic relatives are feasible in the case of acidophilic green algae but are difficult in the case of cyanidialean red algae because their last common acidophilic ancestor diverged from other neutrophilic red algae 1.2–1.3 billion y ago (12). Thus, whole-genome comparisons between evolutionarily related neutrophilic and acidophilic green algae will give insights into how acidophiles evolved from their neutrophilic ancestors.

Significance

Extremely acidic environments are scattered worldwide, and their ecosystems are supported by acidophilic microalgae as primary producers. To understand how acidophilic algae evolved from their respective neutrophilic ancestors, we determined the draft genome sequence of the acidophilic green alga *Chlamydomonas eustigma* and performed comparative genome analyses between *C. eustigma* and its neutrophilic relative *Chlamydomonas reinhardtii*. The results suggest that higher expression of heat-shock proteins and H^+-ATPase, loss of some metabolic pathways that acidify cytosol, and acquisition of metal-detoxifying genes by horizontal gene transfer have played important roles in the adaptation to acidic environments. These features are also found in other acidophilic green and red algae, suggesting the existence of common mechanisms in the adaptation to acidic environments.

Author contributions: S. Hirooka and S.-y.M. designed research; S. Hirooka, Y.H., Y.K., S. Higuchi, T.F., R. Onuma, A.E., and S.-y.M. performed research; R. Ohbayashi, A.U., H.N., and H.Y. contributed new reagents/analytic tools; S. Hirooka, Y.H., Y.K., and S.-y.M. analyzed data; and S. Hirooka and S.-y.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in DNA Data Bank of Japan/European Molecular Biology Laboratory-European Bioinformatics Institute/GenBank under the accession codes PRJDB5468, PRJDB6154, and PRJDB6155.

[1]To whom correspondence may be addressed. Email: shirooka@nig.ac.jp or smiyagis@nig.ac.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1707072114/-/DCSupplemental.

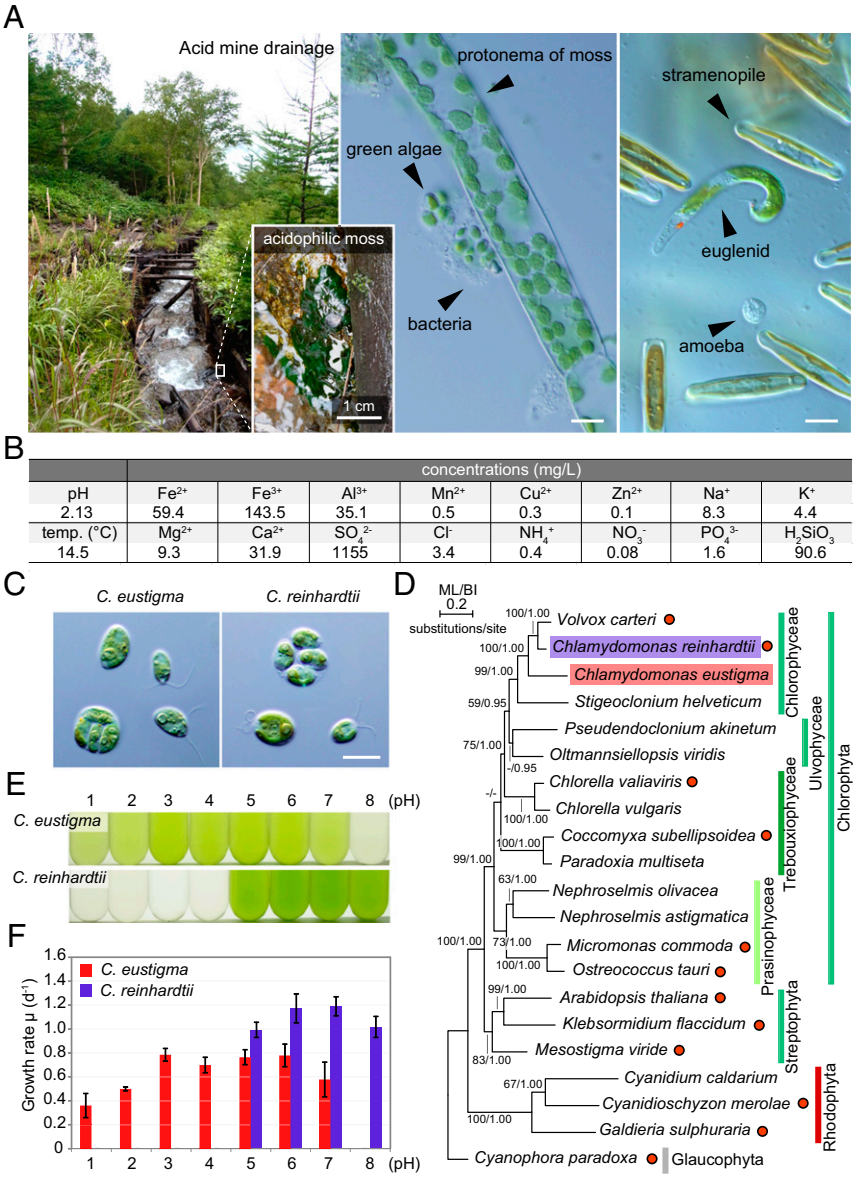

Fig. 1. Habitat, taxonomic position, and physiological features of the acidophilic green alga *C. eustigma*. (*A*) The algae inhabiting AMD in Yokote, Nagano Prefecture, Japan, and confirmation of the existence of *C. eustigma*. Algae were found predominantly in association with acidophilic mosses. (Scale bars: 10 μm.) (*B*) pH, temperature, and concentrations of some ions in the AMD. (*C*) Cells of *C. eustigma* NIES-2499 (*Left*) and *C. reinhardtii* 137c mt+ (*Right*). (Scale bar: 10 μm.) (*D*) A phylogenetic tree of green and red algae based on the concatenated datasets (21 taxa, 11,367 sites) of five chloroplast protein-coding genes (*atpB*, *psaA*, *psaB*, *psbC*, and *rbcL*) and chloroplast ribosomal DNA sequence (16S and 23S). The maximum likelihood (ML) (RaxML 8.0.0) and Bayesian (MrBayes 3.2.6) analyses were calculated under separate model conditions. Bootstrap values (BP) >50% obtained by ML and Bayesian posterior probabilities (BI) >0.95 obtained by Bayesian analysis (MrBayes 3.2.6) are shown above the branches. The branch lengths reflect the evolutionary distances indicated by the scale bar. Filled red circles on the right indicate organisms for which genomes have been sequenced thus far. (*E*) *C. eustigma* and *C. reinhardtii* were cultured for 1 d in the same photoautotrophic medium at a series of pH. (*F*) Growth rates of *C. eustigma* and *C. reinhardtii* based on the increase in the cell number at the indicated pH. The error bars represent the SD of three biological replicates.

Here, we determined the draft genome of the acidophilic green alga *Chlamydomonas eustigma* NIES-2499 isolated from sulfuric AMD and performed comparative genome and transcriptome analyses between *C. eustigma* and its neutrophilic relative *Chlamydomonas reinhardtii*, which was previously fully sequenced (13). The results suggest that up-regulation of genes encoding heat-shock proteins (HSPs) and plasma membrane H^+-ATPase (PMA), loss of fermentative genes that produce organic acids and thus reduce cytosolic pH, the acquisition of an energy shuttle and buffering system, and the acquisition and multiplication of genes involved in arsenic biotransformation and detoxification have contributed to the adaptation of *C. eustigma* to acidic conditions. The results also suggest that there are several commonalities in genomic evolution for adapting to acidic environments among red algae and green algae.

Results

Habitat, Taxonomic Position, and Physiological Features of the Acidophilic Green Alga *C. eustigma*. *C. eustigma* (a haploid vegetative cell) was originally isolated together with mosses (14) from sulfuric AMD in Nagano Prefecture, Japan, in August 1992. We confirmed that *C. eustigma* still thrived in that AMD (pH 2.13, 14.5 °C) in September 2013 (Fig. 1*A*). The water in the AMD contained high concentrations of iron, Al^{3+}, and SO_4^{2-} (Fig. 1*B*), as in the case of other AMDs (15). *C. eustigma* exhibits a cell size and morphology very similar to those of the neutrophilic *C. reinhardtii*. Both cells possess two flagella and a large cup-shaped chloroplast in which an eyespot and a large pyrenoid are formed and proliferate by forming autospores in the mother cell (Fig. 1*C*). As also shown by a previous phylogenetic study based on 18S rDNA sequences (16), phylogenetic analysis based on five chloroplast-encoded genes and two chloroplast ribosomal DNA sequences (Table S1) showed that *C. eustigma*, together with the fully sequenced neutrophile *C. reinhardtii* (13), belongs to the Chlorophyceae, which contains mainly neutrophilic algae (Fig. 1*D*). In the phylogenetic tree, all the Chlorophyta (green algae) except for *C. eustigma* are neutrophilic (Fig. 1*D*), suggesting that the acidophile *C. eustigma* evolved from a neutrophilic ancestor. In autotrophic synthetic medium at 20 °C, *C. eustigma* proliferated at pH 2.0–6.0 (at pH 1.0 cells grew for a few days, but after that they died), and pH 3.0–6.0 was optimal for its growth, whereas *C. reinhardtii* proliferated at pH 5.0–8.0, and pH 6.0–7.0 was optimal for its growth (Fig. 1 *E* and *F*).

Characteristics of the *C. eustigma* Nuclear Genome. To understand the genetic basis of the adaptation of *C. eustigma* to an acidic environment, we sequenced its nuclear genome (Tables S2 and S3). K-mer analysis of Illumina MiSeq reads yielded two peaks with similar frequency of coverage (19× and 38×, respectively), suggesting that the *C. eustigma* genome is a chimera of single and duplicated regions (Fig. 2*A*). Considering the length of the duplicated regions, the estimated genome size was ∼130 Mb (Table S3). Then we obtained Illumina HiSeq and Roche 454 GS FLX+ reads of the nuclear genome (Table S2). The sequenced DNA reads were assembled into 519 scaffolds (the N_{50} scaffold size was 465 kb, and the total length was 67 Mb) (Table S3). Consistent with the result of K-mer analysis, sequencing coverage ratios of some scaffolds were two times those of other scaffolds (Fig. 2*B* and Dataset S1).

Most of the scaffolds consist of only single or duplicated sequences, with a few exceptions (for example, the largest scaffold is a chimera of single and duplicated regions) (Fig. 2*B* and Dataset S1). This result suggests that *C. eustigma* has experienced genomic duplication at the chromosomal level but not considerable rearrangements of the duplicated regions. Considering the length of the duplicated regions, the assembled genome size of *C. eustigma* was ∼110 Mb (Table S3). The difference between the estimated (∼130 Mb) and assembled (∼110 Mb) genome sizes is probably due to the difficulty in resolving repeats, which is often encountered in genome sequencing studies (17). The *C. eustigma* genome exhibits relatively low GC content (45%) compared with the genomes of other green algae (64% in *C. reinhardtii*; 56% in *Volvox carteri*; 67% in *Chlorella variabilis*; 53% in *Coccomyxa subellipsoidea*) (Fig. 2*C*).

In the assembled *C. eustigma* draft genome, 14,105 protein-coding genes were identified by Augustus software with RNA-sequencing (RNA-seq) data (genes encoded in duplicated regions are counted as single genes) (Table S3). The BLASTP search against the Nationl Center for Biotechnology Information nonredundant (NCBI-nr) database (release 20160519) showed

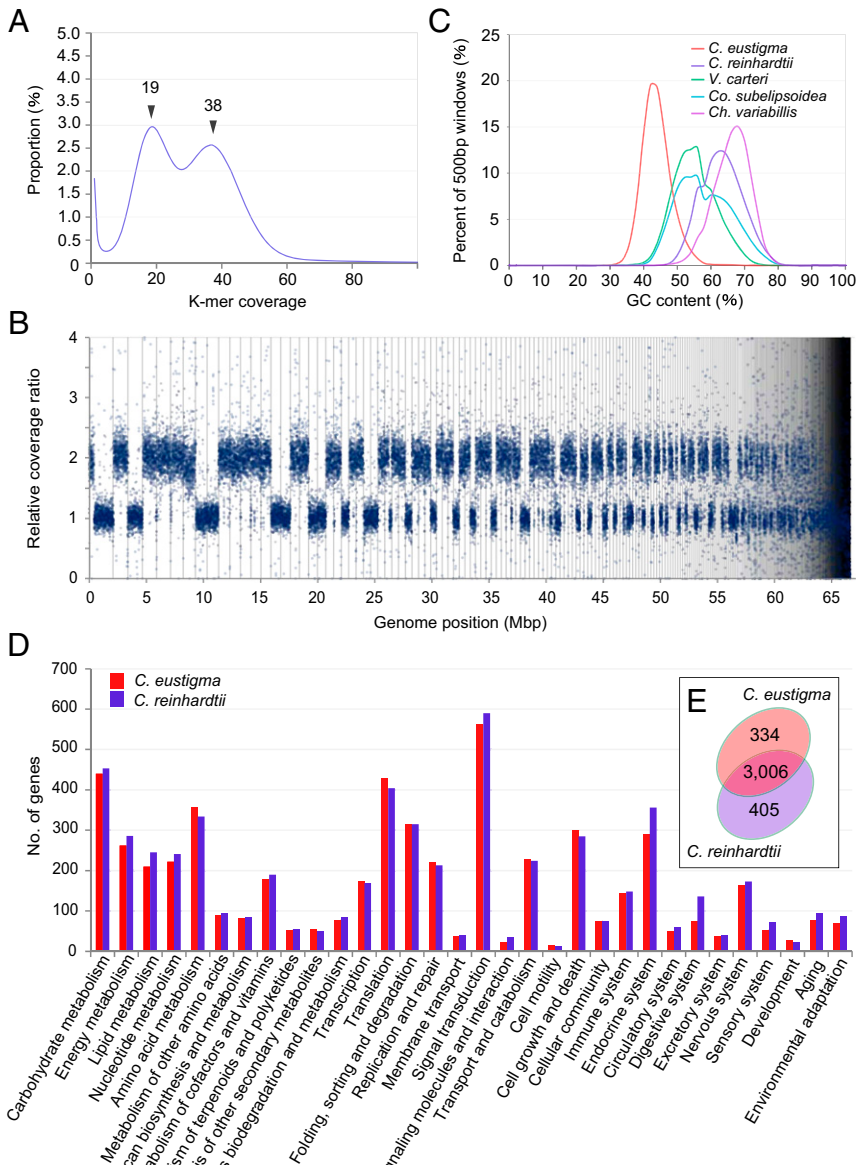

Fig. 2. The *C. eustigma* genome architecture and comparison of genome contents between *C. eustigma* and *C. reinhardtii*. (*A*) 31 K-mer depth distribution of whole-genome Illumina MiSeq reads. Two peaks at 19× and 38× were identified. (*B*) Distribution of the relative sequencing coverage ratio in the *C. eustigma* genome based on the coverage ratio of 2-kb windows. The scaffolds are ordered descendingly from the largest one on the *x* axis. Scaffolds are separated by black bars. (*C*) Comparison of the GC contents in *C. eustigma* and the evolutionarily related neutrophilic green algal species with sequence genomes. The *x* axis indicates the GC content, and the *y* axis indicates the proportion of the bin number divided by the total windows. We used 500-bp bins (with a 250-bp overlap) sliding along the genome. (*D*) Comparison of the number of genes in *C. eustigma* and *C. reinhardtii* whose functions were assigned to respective KEGG functional categories. Each bar indicates the number of genes that are assigned to the particular functional category. (*E*) Venn diagram of KEGG Orthology IDs to which one or more genes are assigned in *C. eustigma* and *C. reinhardtii*.

that 52.1% of *C. eustigma* proteins are most closely related to those of Volvocales (*C. reinhardtii*, *Gonium pectorale*, and *V. carteri*), whereas 19.7% showed no significant similarity to any known proteins (Fig. S1).

High Expression of *HSP* and *PMA* Genes in *C. eustigma*. To compare the genomic contents between acidophile *C. eustigma* and neutrophile *C. reinhardtii*, functional annotations were assigned to *C. eustigma* and *C. reinhardtii* gene models. Predicted genes were assigned to the Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology database through the KEGG Automatic Annotation Server (KAAS). The analysis assigned unique KEGG Orthology IDs to 4,470 *C. eustigma* and 4,741 *C. reinhardtii* protein-coding genes, respectively (Table S3). However, there were no marked differences in the number of genes classified into respective functional categories (Fig. 2*D*), and most of the KEGG Orthology IDs (3,006) were shared by the two species (Fig. 2*E*).

To examine the difference in the expression levels of the orthologous genes between *C. eustigma* and *C. reinhardtii*, we performed RNA-seq analyses of the two species under their individual optimal conditions (at 20 °C in the same autotrophic medium and at pH 3.0 for *C. eustigma* and pH 7.0 for *C. reinhardtii*). Before comparing the transcriptome, we identified 4,590 one-to-one orthologous genes in the three volvocalean species (*C. eustigma*, *C. reinhardtii*, and *V. carteri*) by OrthoMCL (Fig. 3 *A* and *B*). Of the 4,590 genes, 1,282 (~30%) showed a greater than fivefold difference in the mRNA levels between *C. eustigma* and *C. reinhardtii* (Fig. 3*C* and Dataset S2). Notably, in the group that was up-regulated in *C. eustigma*, *HSP* genes were enriched (Fig. 3 *C* and *D* and Fig. S2). Consistent with the result at the mRNA level, previous studies showed that the acidophile *C. acidophila* (CCAP 11/137 isolated from acidic fresh water in Germany), which is closely related to *C. eustigma* (16), had higher basal HSP levels (HSP70, HSP60, and HSP20) than *C. reinhardtii* (18). These observations suggest that *C. eustigma* is constantly exposed to higher stress despite being adapted to an acidic environment.

In addition, we found that *PMA* was highly expressed in *C. eustigma* [151th highest reads per kilobase of transcript per million reads mapped (RPKM) value among 14,105 protein-coding genes] compared with *C. reinhardtii* (1,553th highest RPKM value among 17,741 protein-coding genes) (Fig. 3 *C* and *D* and Fig. S2). Maintenance of a neutral pH in the cytosol despite being in an acidic environment of pH 3 indicates the presence of a 10^4-fold proton gradient across the plasma membrane. It has been suggested that this proton gradient in acidophiles is achieved by a combination of active transport and low permeability of protons (19). In the acidophile *Chlamydomonas* sp. (ATCC PRA-125 isolated from acidic fresh water in Spain), it was previously shown that average cytosolic pH is maintained at pH 6.6 in the culture medium at both pH 2 and pH 7 (20). In *C. reinhardtii* cultured at pH 7, the average cytosolic pH was 7.1 (20). In addition, it was shown that 7% more ATP was consumed to remove protons entering the cytosol across the membrane at pH 2 than at pH 7 (20). Thus, the high expression of *PMA* in *C. eustigma* probably contributes to maintaining the high proton-pumping activity against the acidic environment.

Selective Loss of Acid-Producing Fermentation Pathways from *C. eustigma*. The above comparison of genome contents showed that several hundred KEGG Orthology IDs (Fig. 2*E*) and gene families (Fig. 3*A*) are specific to either *C. eustigma* or *C. reinhardtii*, suggesting that the gene acquisitions and gene losses by *C. eustigma* after divergence from its neutrophilic ancestor also played roles in its adaptation to an acidic environment.

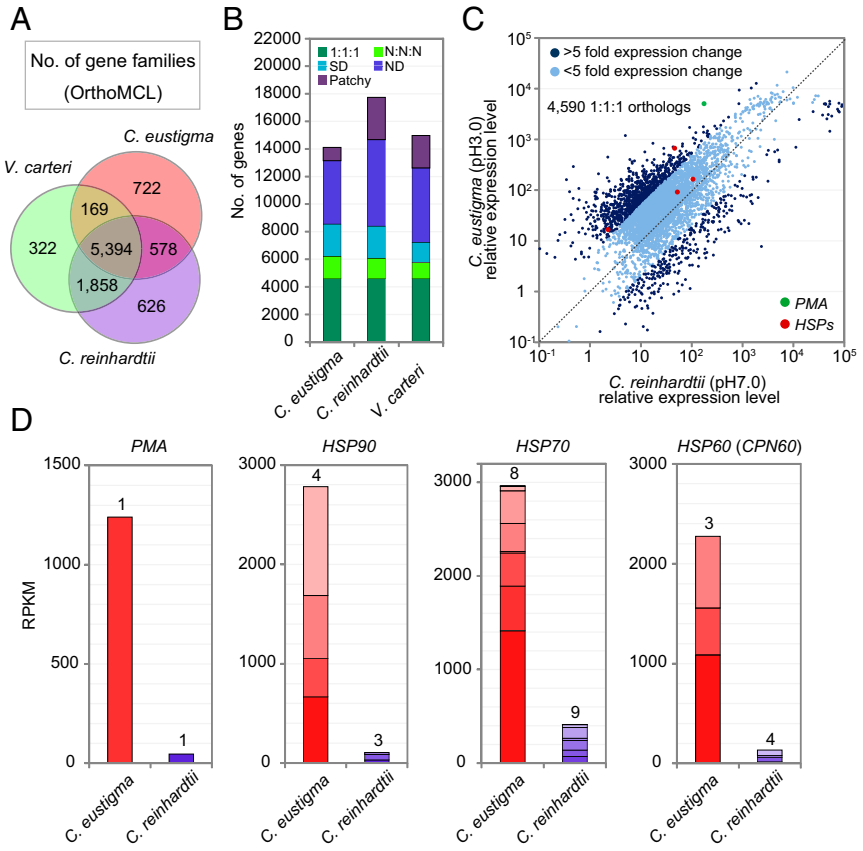

Fig. 3. Comparison of genome contents and transcriptome between the acidophile *C. eustigma* and neutrophile *C. reinhardtii*. (*A*) Venn diagram showing the number of protein families (by OrthoMCL) shared by *C. eustigma*, *C. reinhardtii*, and *V. carteri* genomes. (*B*) Gene orthologs of *C. eustigma*, *C. reinhardtii*, and *V. carteri* identified by OrthoMCL. "1:1:1" indicates an ortholog shared by three species as single copies; "N:N:N" indicates an ortholog shared by three species as multiple copies; "Patchy" indicates an ortholog shared by only two species. "ND" and "SD" indicate a species-specific gene in single or multiple copies, respectively. (*C*) Scatter plot of the mRNA levels of one-to-one orthologous genes between *C. eustigma* (pH 3.0, 20 °C) and *C. reinhardtii* (pH 7.0, 20 °C). The RPKM levels of 4,590 orthologous genes were plotted. (*D*) Comparison of RPKM levels of the *PMA* gene and *HSP* genes in *C. eustigma* (pH 3.0, 20 °C) and *C. reinhardtii* (pH 7.0, 20 °C). The number above the bar indicates the number of genes that exist in the respective nuclear genomes.

Regarding the gene losses by *C. eustigma*, we found that the genome had lost many genes involved in anaerobic fermentation pathways (Fig. 4 *A* and *B*). Several lineages of eukaryotic algae have evolved fermentation pathways that produce ATP when oxygenic respiration is compromised, for example, under anoxic/hypoxic conditions resulting from a low level of photosynthetic activity and local depletion of oxygen by microbial respiration (21). The alcohol fermentation pathway produces a diffusible, nonacidic, and relatively nontoxic end product, ethanol, while other pathways produce organic acids as end products that cause cytosolic acidification and damage (Fig. 4*A*) (22–24). Moreover, organic acids function as uncouplers of the respiratory chain at low pH by diffusion of the protonated form into the cell, followed by dissociation of a proton (19).

C. reinhardtii possesses fermentation pathways that produce both ethanol and organic acids, namely, lactate, formate, and acetate (Fig. 4 *A* and *B*) (24). Although the *C. eustigma* genome encodes pyruvate decarboxylase 3 (PDC3) and alcohol dehydrogenase (ADH) that produce ethanol, it lacks enzymes involved in organic acid fermentation pathways, such as lactate dehydrogenase (LDH) that produces lactate, pyruvate formate lyase (PFL) that produces formate, and both chloroplast and mitochondrial phosphate acetyltransferases (PAT2 and PAT1) and acetate kinases (ACK1 and ACK2) that produce acetate (Fig. 4 *A* and *B*). In addition to lacking these genes, *C. eustigma* lacks the genes encoding pyruvate:ferredoxin oxidoreductase (PFR) and hydrogenase (HYDA) and the proteins required for hydrogenase activation (HYDEF and HYDG) (25). All the above-mentioned genes absent in *C. eustigma* are present in other green algal genomes (*C. reinhardtii*, *V. carteri*, and *Ch. Variabilis*) (Fig. 4*B* and Table S4), suggesting that *C. eustigma* lost these genes during evolution after divergence from the common ancestor of *C. reinhardtii* and *V. carteri*. We found that some enzymes are also absent in *Co. subellipsoidea* C-169 (isolated from Antarctic dried algal peat) (26). However, based on the phylogenetic relationship between *C. eustigma* and *Co. subellipsoidea*, they probably lost these genes independently (Fig. 4*B*).

Consistent with the loss of genes involved in organic acid fermentation, HPLC analyses showed that *C. eustigma* produces little lactate, formate, and acetate (Fig. 4*C*). These three organic acids were detected in the supernatant fraction of *C. reinhardtii* culture but were scarcely detected in that of *C. eustigma* under aerobic conditions (Fig. 4*C*). When the cells were transferred to dark and anaerobic conditions, the formate and acetate levels increased in the supernatant fraction of *C. reinhardtii* culture 4 h after the transfer (Fig. 4*C*), as previously reported (27), but under these conditions these organic acids still were hardly detected in the supernatant fraction of *C. eustigma* (Fig. 4*C*). In contrast to the loss of organic acid fermentation by *C. eustigma*, a higher concentration of ethanol was detected in the supernatant fraction of *C. eustigma* culture than in that of *C. reinhardtii* by the gas chromatography analysis (Fig. 4*C*). The cellular ethanol level increased 4 h after the cells had been transferred from aerobic to dark and anaerobic conditions in both *C. eustigma* and *C. reinhardtii* (Fig. 4*C*). These results indicate that *C. eustigma* selectively lost organic acid-producing fermentative genes.

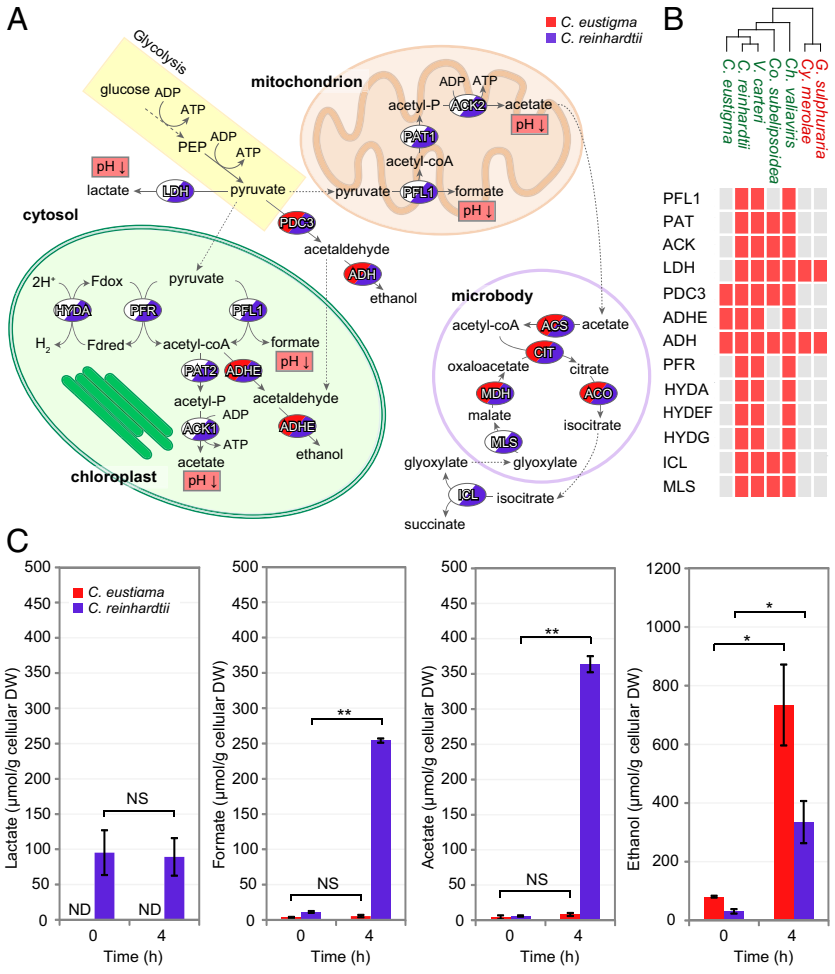

Fig. 4. Loss of organic acid-producing fermentation pathways by *C. eustigma*. (*A*) Overview of the anaerobic fermentation pathways in eukaryotic algae (24, 66). Glucose (stored as starch) synthesized by photosynthesis is oxidized to pyruvate via glycolysis. Under anaerobic conditions, the conversion of pyruvate to acetyl-coA is catalyzed by the pyruvate formate lyase (PFL1) pathway that generates formate or by the pyruvate-ferredoxin oxidoreductase (PFR) and hydrogenase (HYDA) pathway that generates hydrogen. In *C. reinhardtii*, HYDEF and HYDG are essential to activate HYDA (67). Acetyl-CoA enters the phosphate acetyltransferase (PAT) and acetate kinase (ACK) pathway that generates acetate or the aldehyde/alcohol dehydrogenase (ADHE) pathway that generates ethanol. Pyruvate can also be used as a substrate to generate ethanol or lactate via the PDC3 (pyruvate decarboxylase 3) and alcohol dehydrogenase (ADH) pathway or lactate dehydrogenase (LDH) pathway, respectively. Acetate is used for lipid biosynthesis or is converted into acetyl-CoA by acetyl-CoA synthetase (ACS), which is further processed in the glyoxylate cycle to regenerate malate and succinate. ACO, aconitase; CIT, citrate synthase; ICL, isocitrate lyase; MDH, malate dehydrogenase; MLS, malate synthase; PEP, phosphoenolpyruvate. (*B*) Presence or absence of fermentation genes in the genomes of five green algae (shown in green) and two thermo-acidophilic red algae (shown in red). The red boxes indicate the presence of the gene, and white boxes indicate the absence of the gene. (*C*) Concentrations of lactate, formate, acetate, and ethanol in the algal culture medium before (0 h) and 4 h after the dark anaerobic treatment. The error bars represent the SD of three biological replicates. DW, dry weight; ND, not detected; NS, not statistically significant; $*P < 0.02$, $**P < 0.01$ (*t* test).

In the rice bean *Vigna umbellata*, it was reported that exposure of plants to low pH leads to the accumulation of formate. In addition, overexpression of *V. umbellata* formate dehydrogenase in tobacco resulted in decreased sensitivity to low pH and aluminum stresses by reducing the accumulation of formate (28). Thus, the loss of organic acid-producing fermentation pathways by *C. eustigma* probably contributed to the adaptation to acidic environments with high concentrations of metals. This probably also accounts for the independent loss of fermentation genes from *Co. subellipsoidea* because the genus *Coccomyxa* contains several acidophilic members (29, 30).

The genome analyses also showed that *C. eustigma* had lost the key enzymes of the glyoxylate cycle, namely, malate synthase (MLS) and isocitrate lyase (ICL) (Fig. 4 *A* and *B*). The glyoxylate cycle converts acetyl-CoA to succinate for the synthesis of carbohydrates and plays an essential role in cell growth on acetate (31). The loss of these enzymes is probably consistent with the fact that *C. eustigma* produces little acetate (Fig. 4*C*) and/or prevents cytosolic acidification that is caused by succinate production.

Acquisition of the Energy Shuttle and Buffering System Based on Amidinotransferase and Phosphagen Kinase by *C. eustigma* Through HGT. Regarding the gene acquisition by *C. eustigma*, we found that the genome encodes two phosphagen kinases (PKs) and one amidinotransferase (AMGT), which were probably introduced through HGT (Fig. 5 *A–C* and Fig. S3). PK and AMGT exist in various animal, protozoan, and bacterial taxa and function as an energy shuttle and buffering system (32) (Fig. 5*C*). PK catalyzes the reversible transfer of a phosphate between ATP and guanidino compounds (e.g., arginine, creatine, glycocyamine, lombricine, and taurocyamine), which are produced from amino acids by AMGT (Fig. 5*C*) (32). However, *PK* or *AMGT* genes have not been identified in other Archaeplastida (land plants and eukaryotic algae whose chloroplasts are of cyanobacterial primary endosymbiotic origin) (11).

In the *C. eustigma* genome, PK1 and AMGT are encoded in the same scaffold close to each other, whereas PK2 is encoded in another scaffold (Fig. 5*A*). Phylogenetic analysis showed that *C. eustigma* PK1 and PK2 are most closely related to PK proteins of cryptophytes and stramenopiles, respectively (Fig. 5*B*). In addition, *C. eustigma* AMGT is most closely related to that of bacteria and cryptophytes (Fig. S3). PK possesses a guanidine specificity region, which probably defines the substrate specificity (33). To determine the substrate of *C. eustigma* PKs, the guanidine specificity region was compared with PKs of other organisms for which substrates have been determined. In the amino acid sequence alignment, the guanidine specificity regions of *C. eustigma* PK1 and PK2 were found to be most closely related to taurocyamine kinase (TK) of *Phytophthora infestans* (Fig. 5*D*). By HPLC/*o*-phthalaldehyde (OPA) fluorometry, taurocyamine was detected in *C. eustigma* cellular extract but not in that of *C. reinhardtii* (Fig. 5*E*). These results suggest that *C. eustigma* acquired TK and AMGT as an L-arginine:taurine amidinotransferase through HGT.

The maintenance of a neutral cytosolic pH by acidophiles consumes a large amount of ATP, as described above (20). It was previously shown that the "artificial HGT of PK," that is, the expression of exogenous arginine kinase in yeasts and *Escherichia coli*, which do not possess endogenous PKs, increased the resistance to transient pH reduction by building up an energy-storing phosphoarginine pool (34, 35). The RNA-seq results showed that *PK1*, *PK2*, and *AMGT* are relatively highly expressed in *C. eustigma*, exhibiting the 1,189th, 381th, and 1,050th highest RPKM values, respectively, among 14,105 protein-coding genes (Fig. S2 and Dataset S3). Thus, the acquisition of the PK–AMGT shuttle by *C. eustigma* has probably contributed to the supply of ATP needed to maintain cellular pH against an acidic environment (Fig. 5*C*).

Enhancement of Arsenic Biotransformation and Detoxification by *C. eustigma* Through HGT. In addition to gene loss and acquisition through HGT, the genomic analysis of *C. eustigma* suggests that gene amplifications within the genome have also contributed to the adaptation to an acidic environment (Fig. 6*A*). It is known that natural acidic drainage often contains a very high concentration of toxic metals such as arsenic (36). In addition to accelerating metal solubilization, acidic water protonates arsenic, which accelerates the penetration of arsenic into cells (2). Arsenate (AsO_4^{3-}), an analog of phosphate, is incorporated into cells along with phosphate, whereas arsenite (AsO_3^{3-}) is incorporated into cells through aquaglycoporins (Fig. 6*B*) (37). Arsenite oxidizes thiols of biomolecules and causes strong oxidative stress (38). Consistent with the higher toxicity of arsenic in acidic environments, we found that *C. eustigma* tolerates a >10 times higher concentration of arsenate than *C. reinhardtii* (Fig. 6*D*). Genomic analyses showed that genes involved in arsenic biotransformation and detoxification (37) have been multiplied in the *C. eustigma* genome (Fig. 6*A*). The genome possesses approximately 10 copies of genes encoding arsenate reductase (ArsC) and arsenite efflux transporter (ACR3), which are located side-by-side in the genome, and approximately seven copies of the gene encoding arsenite S-adenosylmethionine methyltransferase (ArsM) (Fig. 6*A*). In addition, genes encoding glutaredoxin (Grx) (~20 copies) and glutathione reductase (GR) (two copies), which are involved in the reduction of arsenate to arsenite (39), have also been multiplied in the genome (Fig. 6*A*). Consistent with the increase in the gene copy number, RNA-seq analysis showed that *ArsM*, *Grx*, and *GR* mRNA levels are higher in *C. eustigma* than in *C. reinhardtii*, even when both are cultured under their respective optimal growth conditions without arsenic (Fig. S2 and Datasets S3 and S4). In addition, *ArsC* and *ACR3* are also relatively highly expressed in *C. eustigma*, exhibiting the 1,001th and 177th highest RPKM, respectively, among 14,105 protein-coding genes (Fig. S2 and Dataset S3).

Several studies have already succeeded in enhancing the tolerance to arsenic by artificial HGT, for example, by overexpressing *E. coli* ArsC and γ-glutamylcysteine synthase (40) (to increase the thiol pool) or overexpression of the yeast ACR3 in *Arabidopsis thaliana* (41). Thus, the multiplication and high expression of arsenic biotransformation and detoxification genes in *C. eustigma* have probably contributed to the high algal resistance to arsenic.

The comparison of green algal genomes showed that, among the proteins related to arsenic biotransformation and detoxification, ArsC and ACR3 are not encoded in other green algal genomes except for those of *C. eustigma* and *Co. subellipsoidea* (Fig. 6*C* and Table S5). Based on the phylogenetic relationship between these two species, *C. eustigma* and *Co. subellipsoidea* probably acquired *ArsC* and *ACR3* genes independently (Fig. 6*C*). In the phylogenetic analyses, ArsC of *C. eustigma* and *C. subellipsoidea* formed a clade with those of acidobacteria, actinobacteria, and δ-proteobacteria (Fig. S4), suggesting the bacterial HGT origin of *C. eustigma* ArsC. On the other hand, *C. eustigma* and *Co. subellipsoidea* ACR3 formed a clade with proteins of charophycean algae and certain land plant species, and this clade is a sister group of fungal proteins (Fig. S5). Thus, the origin of *C. eustigma* ACR3 is not clear at this point; however, given that only a limited number of green algae and land plants possess *ACR3* (Fig. S5), it is likely that *ACR3* was acquired by these species multiple times independently through HGT. Thus, the multiplication of both genes derived from their eukaryotic ancestor (*ACR3*, *Grx*, and *GR*) and genes acquired through HGT (*ArsC* and *ACR3*) probably contributed to the adaptation of *C. eustigma*.

Discussion

The above analyses showed that the *C. eustigma* genome has experienced large-scale duplication throughout its genome (Fig. 2*B*)

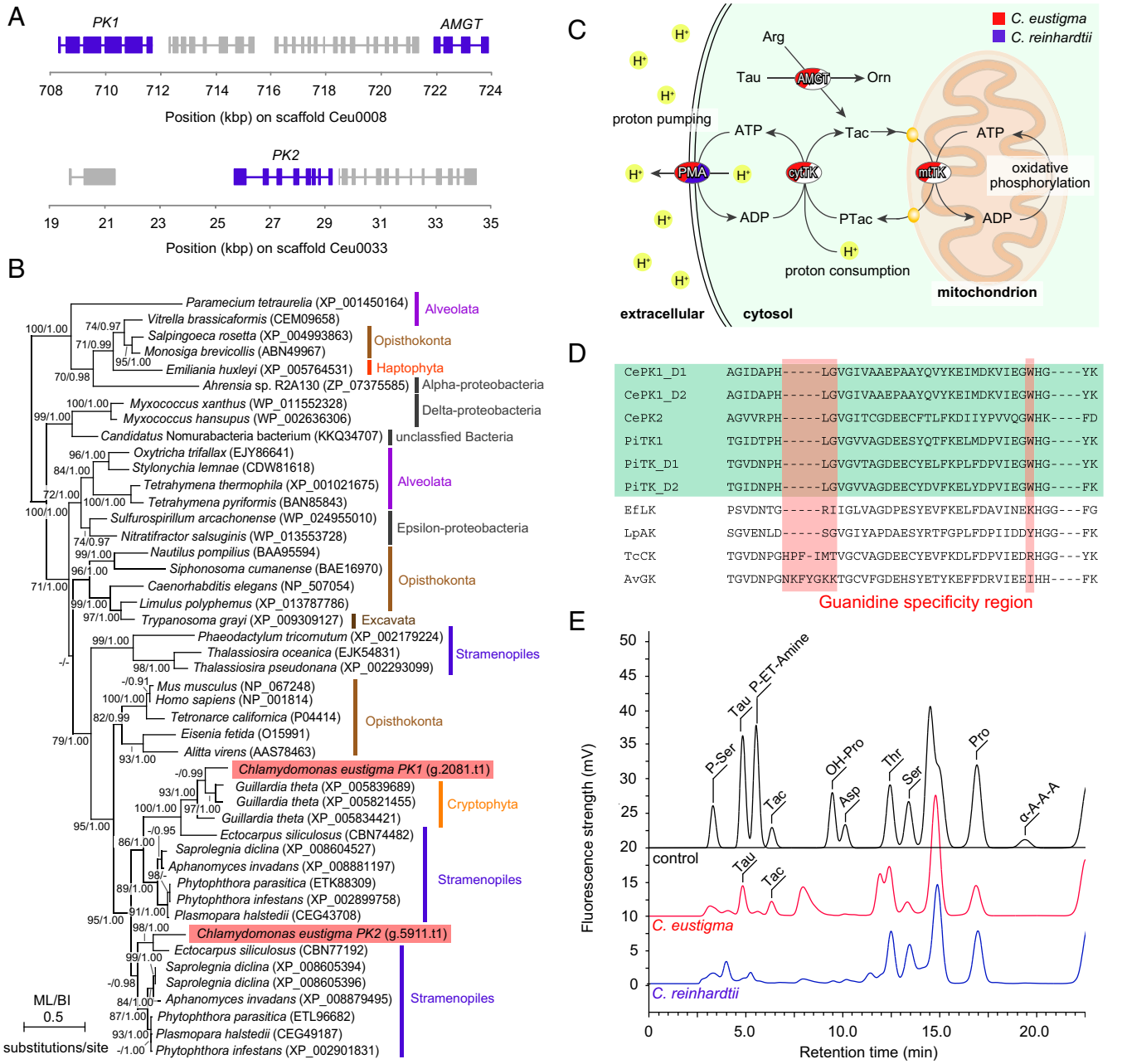

Fig. 5. Acquisition of the PK–AMGT energy shuttle and buffering system by *C. eustigma* through HGT. (*A*) Genomic location of the two *PK* genes and one *AMGT* gene. PK1 and AMGT are encoded in the scaffold Ceu0008, and PK2 is encoded in the scaffold Ceu0033 in the *C. eustigma* genome. (*B*) Phylogenetic relationship of PK proteins. The tree was constructed by the ML method (RaxML 8.0.0). ML BP >50% obtained by RaxML, and BI >0.95 obtained by Bayesian analysis (MrBayes 3.2.6) are shown above the branches. The accession numbers of the sequences are shown along with the names of the species. The branch lengths reflect the evolutionary distances indicated by the scale bar. (*C*) Overview of the phosphagen kinase energy buffering system (32). Taurocyamine (Tac) is produced by amidinotransferase (AMGT) from L-arginine (Arg) and taurine (Tau). Tac is phosphorylated to produce phosphotaurocyamine (PTac) by mitochondrial taurocyamine kinase (mtTK) by consuming ATP produced by oxidative phosphorylation. PTac is dephosphorylated by cytosolic taurocyamine kinase (cytTK) to produce ATP. ATP is, for example, consumed by plasma membrane H^+-ATPase (PMA) to pump protons from the cytosol to outside the cell. Orn, ornithine. (*D*) Amino acid sequence alignment of the guanidine specificity region in *C. eustigma* phosphagen kinase with those of other organisms. The guanidine specificity regions (33) are shaded red, and taurocyamine kinases are shaded green. D1 and D2 represent domain 1 and domain 2, respectively, of two-domain enzymes. AvGK, *Alitta virens* glycocyamine kinase; CePK, *C. eustigma* phosphagen kinase; EfLK, *Eisenia fetida* lombricine kinase; LpAK, *Limulus polyphemus* arginine kinase; PiTK, *Phytophthora infestans* taurocyamine kinase; TcCK, *Tetronarce californica* creatine kinase. (*E*) Fluorometric detection of amines in cellular extractions by HPLC/OPA. The control is a chromatogram of a standard mixture of amino acids and taurocyamine. Asp, aspartic acid; OH-Pro, hydroxyproline; P-ET-Amine, *o*-phosphoethanolamine; Pro, proline; P-Ser, *o*-phosphoserine; Ser, serine; Tac, taurocyamine; Tau, taurine; Thr, threonine; α-A-A-A, α-aminoadipic acid.

and has a lower GC content than evolutionarily related neutrophilic green algae sequenced thus far (Fig. 2*C*). However, it is currently unclear whether there are any relationships between these features in the genome structure and the adaptation to an acidic environment. Generally, genome or gene duplication is widely considered to facilitate environmental adaptation because the redundancy generated allows the evolution of new beneficial gene functions that are otherwise prohibited due to functional constraints (42). Genomic

GC content is predicted to affect genome functioning and species ecology significantly. However, the biological significance of GC content diversity remains elusive because of a lack of sufficiently robust genomic data (43).

Comparative genome and transcriptome analyses suggest that the following features of genomic evolution have contributed to the adaptation of *C. eustigma* to an acidic environment. (*i*) *HSP*s and *PMA* became expressed at high levels. (*ii*) The genome lost fermentative genes that produce organic acids in the cell. (*iii*) The genome acquired genes encoding the PK–AMGT energy shuttle and buffering system and genes that are involved in arsenic biotransformation and detoxification through HGT. (*iv*) The genes involved in arsenic biotransformation and detoxification, derived from a green algal ancestor or acquired through HGT, were multiplied in the genome.

In addition, based on this study and the results of previous studies in other acidophilic algae, it is suggested that these genomic changes are probably common trends in the adaptation to acidic or other extreme environments, as discussed below. Regarding (*i*), we also found that *HSP*s are highly expressed in the thermo-acidophilic red alga *Cy. merolae* under its optimal conditions (in an autotrophic medium at pH 2.5 and 42 °C), as in the case of *C. eustigma* compared with *C. reinhardtii* (Fig. S2 and Dataset S5). Regarding (*ii*), *Co. subellipsoidea* also lost organic acid-producing fermentation genes independently from *C. eustigma* (Fig. 4 *A* and *B*). The genus *Coccomyxa* contains several acidophiles (29, 30). However, it is currently unclear whether there is a correlation between the loss of organic acid-producing fermentation and adaptation to an acidic environment in acidophilic red algae, because they possess only the lactate fermentation pathway and alcohol dehydrogenases (Fig. 4 *A* and *B*). Regarding (*iii*), although the HGT of PK–AMGT into acidophiles has not been reported in other acidophiles, the acquisition of arsenic biotransformation and detoxification genes through HGT has been found in the green alga *Co. subellipsoidea* and thermo-acidophilic red algae (Fig. 6*C*) (8). The green alga *Co. subellipsoidea* acquired ACR3 (Fig. 6*C* and Fig. S5), and the red alga *G. sulphuraria* acquired ArsB through HGT (8) (Fig. 6*C*). Regarding (*iv*), in the *G. sulphuraria* genome, genes of the chloride channel and chloride carrier/channel families have been multiplied and are thought to be important to acid tolerance (8). In addition, an archaeal ATPase of HGT origin has been multiplied and probably contributes to heat tolerance (8).

This study and recent studies on the genomes of acidophiles have started to reveal commonalities in genomic evolution regarding adaptation to an acidic environment. Besides increasing our understanding of evolution, this information could also have important applications. Microalgae have been cultivated at a large scale to produce functional foods and pigments and are also considered to be an alternative source for biofuels because of their relatively rapid growth to a high concentration (44). Acidophilic microalgae have an advantage in that they can be cultivated outdoors without the risk of contamination by other undesirable organisms (45). In addition, trials using acidophiles

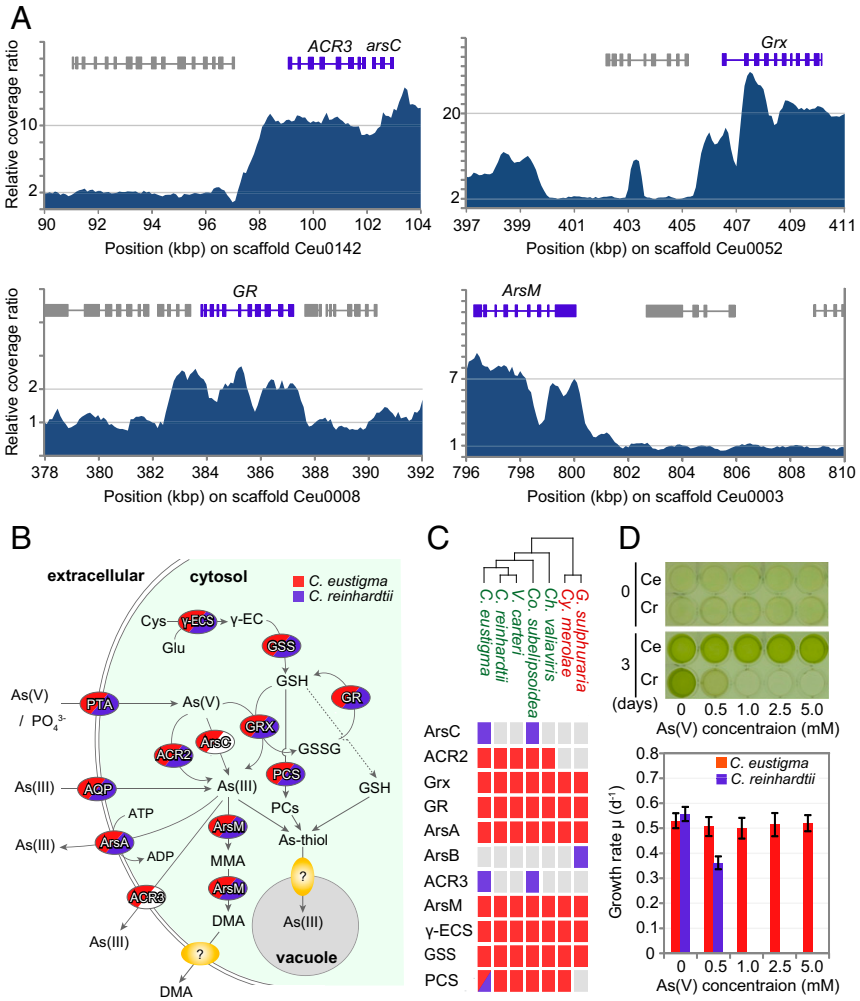

Fig. 6. Acquisition through HGT and expansion of genes in the arsenic biotransformation and detoxification system in *C. eustigma*. (*A*) Relative sequencing coverage ratio of Illumina MiSeq reads (200-bp window with 100-bp overlap) showing the copy numbers of arsenic biotransformation and detoxification genes. (*B*) Overview of the arsenic biotransformation and detoxification system (37). As(V) and As(III) are taken up through phosphate transporter (PTA) as a phosphate analog and through aquaglycoporins (AQP), respectively. As(V) is reduced to As(III) by glutaredoxin (Grx) by using glutathione (GSH) as a reductant or arsenate reductase ArsC or ACR2. As(III) is excreted by As(III)-specific transporter ArsA or ACR3. GSH is an antioxidant and is synthesized by γ-glutamylcysteine synthetase (γ-ECS) and glutathione synthetase (GSS). The methylation of As(III) to monomethylarsonic acid (MMA) and dimethylarsinic acid (DMA) is also thought to function as inorganic arsenic biotransformation and detoxification. As(III) is chelated by phytochelatins (PCs), which are synthesized by phytochelatin synthase (PCS) or GSH to form the As(III)-thiol complex, which is finally transported into the vacuole. (*C*) Presence or absence of arsenic biotransformation and detoxification genes in the genomes of five green algae (shown in green) and two thermo-acidophilic red algae (shown in red). The red and blue boxes indicate the presence of the gene; blue boxes indicate a gene that was horizontally transferred. The gray boxes indicate the absence of the gene. (*D*) Effects of the concentration of arsenate As(V) on the growth of *C. eustigma* and *C. reinhardtii*. Cells were cultured in modified original medium containing both 1 mM phosphate (originally 10 mM) and various concentrations of arsenate for 3 d. The error bars in the graph represent the SD of three biological replicates.

for bioremediation (46) and metal recovery have also been initiated (47). An understanding of the genetic basis of acid and/or metal tolerance is also necessary to confer such abilities to other organisms.

Methods

Algal Strains. *C. eustigma* (NIES-2499; Microbial Culture Collection at the National Institute for Environmental Studies) and *C. reinhardtii* 137c mt+ were used in the current study. *C. eustigma* and *C. reinhardtii* were maintained with gyration (100 rpm) on a rotary shaker (NR-2; Taitec) in photoautotrophic medium (9.35 mM NH_4Cl, 81.15 μM $MgSO_4{\cdot}7H_2O$, 68.04 μM $CaCl_2$, 10 mM KH_2PO_4/K_2HPO_4, 59.2 μM $FeCl_3$, 0.73 μM $MnCl_2{\cdot}4H_2O$, 0.31 μM $ZnSO_4{\cdot}7H_2O$, 0.0672 μM $CoCl_2{\cdot}6H_2O$, 0.0413 μM $Na_2MoO_4{\cdot}2H_2O$, 0.5046 μM $CuSO_4{\cdot}5H_2O$, 13.75 μM $Na_2EDTA{\cdot}2H_2O$) at pH 3.0 and pH 7.0, respectively, at 21 °C under a 12-h light/12-h dark photoperiod (30 $\mu E{\cdot}m^{-2}{\cdot}s^{-1}$).

Genomic DNA Sequencing. Genomic DNA of *C. eustigma* was extracted, and sequencing libraries were prepared according to ref. 48. The shotgun and paired-end libraries (8 kb) were sequenced by Roche 454 GS FLX+ Titanium (Roche Diagnostics). The paired-end (400-bp) library was sequenced by HiSeq 2500 with 100 base-paired end format with the TruSeq SBS kit v3 (Illumina, Inc.). The paired-end (800-bp) and mate-pair libraries (3, 5, and 8 kb) were sequenced by MiSeq (Illumina, Inc.) with the MiSeq reagent kit version 3 (600 cycles; Illumina). The MiSeq reads were filtered using ShortReadManager (49), based on a 17-mer frequency.

Estimation of the Genome Size by K-mer Analysis. The MiSeq paired-end reads were used for the K-mer 31 frequency distribution analysis using JELLYFISH (50). The total genome size was estimated by analyzing the occurrence and distribution of K-mers using the following formula: Estimated genome size in base pairs = K-mer number/depth. The 31 K-mer depth distribution of the MiSeq paired-end reads exhibited two peaks (Fig. 2*A*). The estimated genome size of *C. eustigma* was ∼130 Mb, when the ×19 was considered as the main peak.

Genome Assembly, Scaffolding, and Gap Closing. The Roche 454 shotgun and paired-end reads were assembled de novo by Newbler version 2.9 (Roche) with the following parameters: -mi 98 -mL 80 -scaffold -large -s 500. Subsequent scaffolding of the Newbler output contigs was performed by SSPACE (51) using the Illumina paired-end and mate-pair information (Table S2). GMcloser (52) was used for gap filling with preassembled contigs and Illumina paired-end reads. The genome sequence was improved with Illumina paired-end reads using iCORN2 (53). To remove mitochondrial and chloroplast DNA sequences, tblastn (54) searches were performed against scaffolds by using amino acid sequences of *C. reinhardtii* mitochondrion- and chloroplast-encoded proteins as queries. By the tblastn search, one mitochondrial DNA scaffold and two chloroplast DNA scaffolds were identified, and these scaffolds were removed from the assembly.

Estimation of Coverage Ratio in the *C. eustigma* Genome. The sequencing coverage ratio was assessed by calculating normalized coverage depth followed by manual inspection of depth variation along each scaffold. The MiSeq read data (6,660,746 reads) were mapped to the scaffolds using Bowtie2 versiom 2.1.0 (55) with default settings. Mapped reads in sequence alignment/map (SAM) format were converted to the binary version of the SAM file (BAM format) using the Samtools version 0.1.19 (56) <view>, <sort>, and <index> commands. The aligned reads in BAM format were filtered for duplicates using the Samtools version 0.1.19 <rmdup> command. After the removal of duplicates, BAM files were converted to browser extensible data (BED) format using the Bedtools version 2.17 (57) <bamToBed> command. Genome-wide windows were defined using the Bedtools <makewindows> command, and then coverage depth of each individual window was calculated using the Bedtools <coverage> command. The histogram of the coverage ratio exhibited two major peaks that probably correspond to single and duplicated regions. The relative coverage ratio (shown in Figs. 2*B* and 6*A*) was normalized by the averaged coverage depth of the probable single regions.

Prediction and Annotation of the Nuclear Genes. Nuclear genes were predicted by Augustus 3.0.3 (58). Assembled transcript sequences were mapped to the scaffolds by BLAT (59) to assess the likelihood that each sequence was indeed a transcript. The manually curated 1,900 gene models were used as Augustus training sets, and 14,105 genes were predicted by Augustus with transcript evidence. The KEGG Orthology ID assignment was performed for all predicted genes in the *C. eustigma* and *C. reinhardtii* genomes. The assignments were performed by the KAAS (60).

Comparison of mRNA Levels of Orthologous Genes in *C. eustigma* and *C. reinhardtii*. To identify one-to-one orthologous genes in the three Volvocales (*C. eustigma*, *C. reinhardtii*, and *V. carteri*), gene clustering analysis was performed by OrthoMCL (61) with the following parameters: inflation value = 1.5, percentMatchCutoff = 50, and evalueExponentCutoff = −10. A BLASTP search for predicted amino acid sequences with an E-value of $1e^{-10}$ in the three algal species was performed using NCBI BLAST+ version 2.2.30. In total, we found 4,590 one-to-one ortholog pairs. Gene-expression scores were obtained from RNA-seq data by mapping the clean reads to the genes by Bowtie2 version 2.1.0 (55). SAMtools (56), BEDtools (57), and R version 2.14.2 (62) were used to calculate the tag-count data that were mapped to the coding genes. Normalization of the orthologous gene-expression scores was performed by RPKM normalization. After obtaining the normalized expression scores of orthologous genes for each sample, the scores were log (base 10)-transformed and plotted to produce a scatter graph for comparison of the expression scores of the two algal species.

Comparison of mRNA Levels in Algal Species. The RNA-seq reads were mapped to the *C. eustigma*, *C. reinhardtii* (JGI, version 5.5), or *Cy. merolae* (*Cyanidioschyzon merolae* Genome Project) coding sequences by Bowtie2 (55) with the default parameters. The Bowtie2 outputs were processed to obtain tag counts. Since it has been shown that the GC content affects the read abundances in an RNA-seq dataset (63), counts were full-quantile normalized within a sample by GC content bias-correction methods implemented in the EDASeq R package (64). These normalized counts were used to calculate the mRNA level of each gene (in RPKM units) in the algal samples according to ref. 65.

Data Availability. The *C. eustigma* NIES-2499 whole-genome and gene models have been deposited in DNA Data Bank of Japan (DDBJ)/European Molecular Biology Laboratory (EMBL)/GenBank under the accession code PRJDB5468. The dataset includes sequences of the nuclear and mitochondrial genomes. Because the chloroplast genome was highly repetitive and the genome could not be assembled well, the chloroplast genome was omitted from the dataset. The RNA-seq data of *C. eustigma* and *C. reinhardtii* have been deposited in DDBJ/EMBL/GenBank (accession codes PRJDB6154 and PRJDB6155, respectively).

ACKNOWLEDGMENTS. We thank Dr. T. Kuroiwa, Dr. H. Kuroiwa, Dr. K. Tanaka, and Dr. Y. Kabeya for kind encouragement and advice on this work; Dr. K. Hori, Dr. N. V. Sasaki, Dr. M. Ishikawa, Dr. T. Fujisawa, and Mr. Y. Kazama for advice on bioinformatic analyses; Dr. U. Goodenough and Dr. Y. Nishimura for advice on *Chlamydomonas* research; Mr. T. Sone for advice on RNA extraction; and members of the S.-y.M. Laboratory for their technical advice and support. This work was supported by the Core Research for Evolutional Science and Technology Program of the Japan Science and Technology Agency (S.-y.M.), by Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science 25251039 (to S.-y.M.), and by the Ministry of Education, Culture, Sports, Science and Technology-supported Program for Strategic Research Foundation at Private Universities 2013–2017 S1311017. Computations were partially performed on the National Institute of Genetics (NIG) supercomputer at the Research Organization of Information and Systems of the NIG.

1. Oren A (2010) Acidophiles. *Encyclopedia of Life Sciences* (Wiley, Chichester, UK), pp 192–206.
2. Gross W (2000) Ecophysiology of algae living in highly acidic environments. *Hydrobiologia* 33:31–37.
3. Ferris MJ, et al. (2005) Algal species and light microenvironment in a low-pH, geothermal microbial mat community. *Appl Environ Microbiol* 71:7164–7171.
4. Novis P, Harding JS (2007) Extreme acidophiles: Freshwater algae associated with acid mine drainage. *Algae and Cyanobacteria in Extreme Environments*, ed Seckbach J (Springer, Heidelberg), pp 443–463.
5. Pedrozo F, et al. (2001) First results on the water chemistry, algae and trophic status of an Andean acidic lake system of volcanic origin in Patagonia (Lake Caviahue). *Hydrobiologia* 452:129–137.
6. Amaral Zettler LA, et al. (2002) Microbiology: Eukaryotic diversity in Spain's River of Fire. *Nature* 417:137.
7. Matsuzaki M, et al. (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 428:653–657.
8. Schönknecht G, et al. (2013) Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science* 339:1207–1210.
9. Qiu H, et al. (2013) Adaptation through horizontal gene transfer in the cryptoendolithic red alga Galdieria phlegrea. *Curr Biol* 23:R865–R866.
10. Olsson S, et al. (2017) Horizontal gene transfer of phytochelatin synthases from bacteria to extremophilic green algae. *Microb Ecol* 73:50–60.
11. Adl SM, et al. (2012) The revised classification of eukaryotes. *J Eukaryot Microbiol* 59: 429–493, and erratum (2013) 60:321.

12. Yoon HS, Hackett JD, Pinto G, Bhattacharya D (2002) The single, ancient origin of chromist plastids. *Proc Natl Acad Sci USA* 99:15507–15512.
13. Merchant SS, et al. (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* 318:245–250.
14. Higuchi S, et al. (2003) Morphology and phylogenetic position of a mat-forming green plant from acidic rivers in Japan. *J Plant Res* 116:461–467.
15. Nordstrom DK, Alpers CN (1999) Negative pH, efflorescent mineralogy, and consequences for environmental restoration at the Iron Mountain Superfund site, California. *Proc Natl Acad Sci USA* 96:3455–3462.
16. Yumoto K, Kasai F, Kawachi M (2013) Taxonomic re-examination of Chlamydomonas strains maintained in the NIES-collection. *Microbiol Cult Collect* 29:1–12.
17. Schatz MC, Delcher AL, Salzberg SL (2010) Assembly of large genomes using second-generation sequencing. *Genome Res* 20:1165–1173.
18. Gerloff-Elias A, Barua D, Mölich A, Spijkerman E (2006) Temperature- and pH-dependent accumulation of heat-shock proteins in the acidophilic green alga *Chlamydomonas acidophila*. *FEMS Microbiol Ecol* 56:345–354.
19. Baker-Austin C, Dopson M (2007) Life in acid: pH homeostasis in acidophiles. *Trends Microbiol* 15:165–171.
20. Messerli MA, et al. (2005) Life at acidic pH imposes an increased energetic cost for a eukaryotic acidophile. *J Exp Biol* 208:2569–2579.
21. Maier RM, Pepper IL, Gerba CP (2000) *Environmental Microbiology* (Gulf Professional Publishing, Houston).
22. van Dongen JT, Licausi F (2014) *Low-Oxygen Stress in Plants: Oxygen Sensing and Adaptive Responses to Hypoxia* (Springer, Vienna).
23. Gfeller RP, Gibbs M (1984) Fermentative metabolism of *Chlamydomonas reinhardtii*: I. Analysis of fermentative products from starch in dark and light. *Plant Physiol* 75: 212–218.
24. Catalanotti C, Yang W, Posewitz MC, Grossman AR (2013) Fermentation metabolism and its evolution in algae. *Front Plant Sci* 4:150.
25. Banti V, et al. (2013) Low oxygen response mechanisms in green organisms. *Int J Mol Sci* 14:4734–4761.
26. Blanc G, et al. (2012) The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biol* 13:R39.
27. Mus F, Dubini A, Seibert M, Posewitz MC, Grossman AR (2007) Anaerobic acclimation in *Chlamydomonas reinhardtii*: Anoxic gene expression, hydrogenase induction, and metabolic pathways. *J Biol Chem* 282:25475–25486.
28. Lou HQ, et al. (2016) A formate dehydrogenase confers tolerance to aluminum and low pH. *Plant Physiol* 171:294–305.
29. Fuentes JL, et al. (2016) Phylogenetic characterization and morphological and physiological aspects of a novel acidotolerant and halotolerant microalga *Coccomyxa onubensis* sp. nov. (Chlorophyta, Trebouxiophyceae). *J Appl Phycol* 28:3269–3279.
30. Koechler S, et al. (2016) Arsenite response in Coccomyxa sp. Carn explored by transcriptomic and non-targeted metabolomic approaches. *Environ Microbiol* 18:1289–1300.
31. Kunze M, Pracharoenwattana I, Smith SM, Hartig A (2006) A central role for the peroxisomal membrane in glyoxylate cycle function. *Biochim Biophys Acta* 1763: 1441–1452.
32. Ellington WR (2001) Evolution and physiological roles of phosphagen systems. *Annu Rev Physiol* 63:289–325.
33. Uda K, Hoshijima M, Suzuki T (2013) A novel taurocyamine kinase found in the protist *Phytophthora infestans*. *Comp Biochem Physiol B Biochem Mol Biol* 165:42–48.
34. Canonaco F, Schlattner U, Pruett PS, Wallimann T, Sauer U (2002) Functional expression of phosphagen kinase systems confers resistance to transient stresses in *Saccharomyces cerevisiae* by buffering the ATP pool. *J Biol Chem* 277:31303–31309.
35. Canonaco F, Schlattner U, Wallimann T, Sauer U (2003) Functional expression of arginine kinase improves recovery from pH stress of *Escherichia coli*. *Biotechnol Lett* 25: 1013–1017.
36. Cullen WR, Reimer KJ (1989) Arsenic speciation in the environment. *Chem Rev* 89: 713–764.
37. Tripathi RD, et al. (2007) Arsenic hazards: Strategies for tolerance and remediation by plants. *Trends Biotechnol* 25:158–165.
38. Birben E, Sahiner UM, Sackesen C, Erzurum S, Kalayci O (2012) Oxidative stress and antioxidant defense. *World Allergy Organ J* 5:9–19.
39. Mukhopadhyay R, Rosen BP (2002) Arsenate reductases in prokaryotes and eukaryotes. *Environ Health Perspect* 110:745–748.
40. Dhankher OP, et al. (2002) Engineering tolerance and hyperaccumulation of arsenic in plants by combining arsenate reductase and gamma-glutamylcysteine synthetase expression. *Nat Biotechnol* 20:1140–1145.
41. Ali W, et al. (2012) Heterologous expression of the yeast arsenite efflux system ACR3 improves *Arabidopsis thaliana* tolerance to arsenic stress. *New Phytol* 194: 716–723.
42. Kondrashov FA (2012) Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc Biol Sci* 279:5048–5057.
43. Šmarda P, et al. (2014) Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proc Natl Acad Sci USA* 111:E4096–E4102.
44. Milledge JJ (2012) Microalgae–Commercial potential for fuel, food and feed. *Food Sci Technol (Campinas)* 26:26–28.
45. Varshney P, Mikulic P, Vonshak A, Beardall J, Wangikar PP (2015) Extremophilic microalgae and their potential contribution in biotechnology. *Bioresour Technol* 184: 363–372.
46. Nishikawa K, Yamakoshi Y, Uemura I, Tominaga N (2003) Ultrastructural changes in *Chlamydomonas acidophila* (Chlorophyta) induced by heavy metals and polyphosphate metabolism. *FEMS Microbiol Ecol* 44:253–259.
47. Minoda A, et al. (2015) Recovery of rare earth elements from the sulfothermophilic red alga Galdieria sulphuraria using aqueous acid. *Appl Microbiol Biotechnol* 99: 1513–1519.
48. Hirose Y, et al. (2015) Complete genome sequence of cyanobacterium *Geminocystis* sp. strain NIES-3708, which performs type II complementary chromatic acclimation. *Genome Announc* 3:e00357–e15.
49. Ohtsubo Y, Maruyama F, Mitsui H, Nagata Y, Tsuda M (2012) Complete genome sequence of *Acidovorax* sp. strain KKS102, a polychlorinated-biphenyl degrader. *J Bacteriol* 194:6970–6971.
50. Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27:764–770.
51. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2011) Scaffolding preassembled contigs using SSPACE. *Bioinformatics* 27:578–579.
52. Kosugi S, Hirakawa H, Tabata S (2015) GMcloser: Closing gaps in assemblies accurately with a likelihood-based selection of contig or long-read alignments. *Bioinformatics* 31:3733–3741.
53. Otto TD, Sanders M, Berriman M, Newbold C (2010) Iterative correction of reference nucleotides (iCORN) using second generation sequencing technology. *Bioinformatics* 26:1704–1707.
54. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
55. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359.
56. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25:2078–2079.
57. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
58. Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24: 637–644.
59. Kent WJ (2002) BLAT–The BLAST-like alignment tool. *Genome Res* 12:656–664.
60. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35: W182–W185.
61. Li L, Stoeckert CJ, Jr, Roos DS (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189.
62. R Core Team (2013) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna), Version 2.14.2.
63. Zheng W, Chung LM, Zhao H (2011) Bias detection and correction in RNA-sequencing data. *BMC Bioinformatics* 12:290.
64. Risso D, Schwartz K, Sherlock G, Dudoit S (2011) GC-content normalization for RNA-seq data. *BMC Bioinformatics* 12:480.
65. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat Methods* 5:621–628.
66. Lauersen KJ, et al. (2016) Peroxisomal microbodies are at the crossroads of acetate assimilation in the green microalga *Chlamydomonas reinhardtii*. *Algal Res* 16: 266–274.
67. Posewitz MC, et al. (2004) Discovery of two novel radical S-adenosylmethionine proteins required for the assembly of an active [Fe] hydrogenase. *J Biol Chem* 279: 25711–25720.
68. Hirooka S, Higuchi S, Uzuka A, Nozaki H, Miyagishima SY (2014) Acidophilic green alga *Pseudochlorella* sp. YKT1 accumulates high amount of lipid droplets under a nitrogen-depleted condition at a low-pH. *PLoS One* 9:e107702.
69. Minoda A, Sakagami R, Yagisawa F, Kuroiwa T, Tanaka K (2004) Improvement of culture conditions and evidence for nuclear transformation by homologous recombination in a red alga, *Cyanidioschyzon merolae* 10D. *Plant Cell Physiol* 45: 667–671.
70. Nakazawa A, Nozaki H (2004) Phylogenetic analysis of the tetrasporalean genus Asterococcus (Chlorophyceae) based on 18S ribosomal RNA gene sequences. *Shokubutsu Kenkyu Zasshi* 79:255–261.
71. Nollet LML, De Gelder LSP (2014) *Handbook of Water Analysis* (Taylor & Francis Group, Boca Raton, FL), 3rd Ed.
72. Murota C, et al. (2012) Arsenic tolerance in a Chlamydomonas photosynthetic mutant is due to reduced arsenic uptake even in light conditions. *Planta* 236:1395–1403.
73. Yoon OK, Brem RB (2010) Noncanonical transcript forms in yeast and their regulation during environmental stress. *RNA* 16:1256–1267.
74. Fujiwara T, et al. (2015) A nitrogen source-dependent inducible and repressible gene expression system in the red alga *Cyanidioschyzon merolae*. *Front Plant Sci* 6:657.
75. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10–12.
76. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30:772–780.
77. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540–552.
78. Tanabe AS (2011) Kakusan4 and Aminosan: Two programs for comparing nonpartitioned, proportional and separate models for combined molecular phylogenetic analyses of multilocus sequence data. *Mol Ecol Resour* 11:914–921.
79. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
80. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
81. Durzan DJ (1969) Automated chromatographic analysis of free monosubstituted guanidines in physiological fluids. *Can J Biochem* 47:657–664.
82. Bennett MS, Guan Z, Laurberg M, Su XD (2001) *Bacillus subtilis* arsenate reductase is structurally and functionally similar to low molecular weight protein tyrosine phosphatases. *Proc Natl Acad Sci USA* 98:13577–13582.