



# Phylogeny of C<sub>4</sub>-Photosynthesis Enzymes Based on Algal Transcriptomic and Genomic Data Supports an Archaeal/Proteobacterial Origin and Multiple Duplication for Most C<sub>4</sub>-Related Genes

Shan Chi<sup>1</sup>✉, Shuangxiu Wu<sup>2,3</sup>✉, Jun Yu<sup>2,3</sup>, Xumin Wang<sup>2,3\*</sup>, Xuexi Tang<sup>1</sup>, Tao Liu<sup>1\*</sup>

**1** Ocean University of China, Qingdao, Shandong Province, People's Republic of China, **2** CAS Key Laboratory of Genome Sciences and Information, Beijing Key Laboratory of Genome and Precision Medicine Technologies, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, P. R. China, **3** Beijing Key Laboratory of Functional Genomics for Dao-di Herbs, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

## Abstract

Both Calvin-Benson-Bassham (C<sub>3</sub>) and Hatch-Slack (C<sub>4</sub>) cycles are most important autotrophic CO<sub>2</sub> fixation pathways on today's Earth. C<sub>3</sub> cycle is believed to be originated from cyanobacterial endosymbiosis. However, studies on evolution of different biochemical variants of C<sub>4</sub> photosynthesis are limited to tracheophytes and origins of C<sub>4</sub>-cycle genes are not clear till now. Our comprehensive analyses on bioinformatics and phylogenetics of novel transcriptomic sequencing data of 21 rhodophytes and 19 Phaeophyceae marine species and public genomic data of more algae, tracheophytes, cyanobacteria, proteobacteria and archaea revealed the origin and evolution of C<sub>4</sub> cycle-related genes. Almost all of C<sub>4</sub>-related genes were annotated in extensive algal lineages with proteobacterial or archaeal origins, except for phosphoenolpyruvate carboxykinase (PCK) and aspartate aminotransferase (AST) with both cyanobacterial and archaeal/proteobacterial origin. Notably, cyanobacteria may not possess complete C<sub>4</sub> pathway because of the flawed annotation of pyruvate orthophosphate dikinase (PPDK) genes in public data. Most C<sub>4</sub> cycle-related genes endured duplication and gave rise to functional differentiation and adaptation in different algal lineages. C<sub>4</sub>-related genes of NAD-ME (NAD-malic enzyme) and PCK subtypes exist in most algae and may be primitive ones, while NADP-ME (NADP-malic enzyme) subtype genes might evolve from NAD-ME subtype by gene duplication in chlorophytes and tracheophytes.

**Citation:** Chi S, Wu S, Yu J, Wang X, Tang X, et al. (2014) Phylogeny of C<sub>4</sub>-Photosynthesis Enzymes Based on Algal Transcriptomic and Genomic Data Supports an Archaeal/Proteobacterial Origin and Multiple Duplication for Most C<sub>4</sub>-Related Genes. PLoS ONE 9(10): e110154. doi:10.1371/journal.pone.0110154

**Editor:** Zhong-Jian Liu, The National Orchid Conservation Center of China; The Orchid Conservation & Research Center of Shenzhen, China

**Received:** April 2, 2014; **Accepted:** September 15, 2014; **Published:** October 14, 2014

**Copyright:** © 2014 Chi et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability:** The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are available from the GenBank database (accession numbers are listed in Supporting Information files).

**Funding:** This work was supported by the National Natural Science Foundation of China (NSFC No. 41206116 and No. 31271397, China Postdoctoral Science Foundation (No. 2011M501167), National Public Benefit Research Foundation of China (No. 201105021). The algal transcriptome sequencing was supported by 1KP Project (www.onekp.com). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* Email: wangxm@big.ac.cn (XW); liutao@ouc.edu.cn (TL)

✉ These authors contributed equally to this work.

## Introduction

The Calvin-Benson-Bassham (CBB) cycle is the most important autotrophic CO<sub>2</sub> fixation pathway on today's Earth, and widely distributes among all land plants, algae and cyanobacteria [1]. The characteristic enzyme involved in CBB cycle is ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO), which catalyzes the primary carboxylation of ribulose 1,5-bisphosphate (RuBP) and yields two molecules of 3-phosphoglycerate (PGA), a C<sub>3</sub> compound [2]. Therefore, the CBB cycle is also called C<sub>3</sub> cycle [2]. Numerous reviews have summarized the understandings of this important pathway [2,3].

Another important CO<sub>2</sub> fixation pathway is Hatch-Slack (H-S) cycle, found in sugarcane in 1965 [4]. It is also well known as C<sub>4</sub> cycle because, in counterpart with the C<sub>3</sub> cycle, the first enzyme involved in this cycle is phosphoenolpyruvate carboxylase (PEPC), which catalyzes the primary carboxylation of phosphoenolpyruvate (PEP) to form a 4-carbon acid compound oxaloacetate

(OAA) as the first photosynthetic product. As compared with C<sub>3</sub> plants, the C<sub>4</sub> cycle develops novel and efficient CO<sub>2</sub> concentration mechanisms, on anatomical and biochemical function, to enhance RubisCO performance even at limited ambient CO<sub>2</sub> levels and result in significant decreases in photorespiration, improvement of photosynthetic efficiency and water use efficiency during CO<sub>2</sub> fixation [5]. Thus, C<sub>4</sub> plants are capable of growing in habitats that may be too harsh for C<sub>3</sub> species, such as rock outcrops and hypersaline or arid soils of low latitude, and contribute about a quarter of the primary productivity on the planet [6].

For land plants, C<sub>4</sub> plants can be divided into three distinct biochemical variants based on different decarboxylation modes: NADP-malic enzyme (NADP-ME) type, NAD-malic enzyme (NAD-ME) type, and phosphoenolpyruvate carboxykinase (PCK) type [7]. Each C<sub>4</sub> subtype consists of two shared enzymes (PEPC and PPDK) and two or three other unique enzymes of their own,

which are all encoded by nuclear genome and transported to different position of cell to catalyze corresponding reaction.

Through the phylogenetic studies in species of *Flaveria* (Asteraceae) and *Brassica gravinae*, C<sub>4</sub> plants are believed to have evolved gradually from C<sub>3</sub> plants through several intermediate stages of C<sub>3</sub>–C<sub>4</sub> plants [8]. However, C<sub>4</sub>-cycle genes and pathway analysis are mostly limited to land plants till recently. It is widely accepted that land plants evolved from streptophyte algae (a diverse group of green, fresh water algae) [9]. Algae are the principal primary producers in oceanic and freshwater communities, and also are responsible for the net flux of about 2 gigatons of carbon per year from the atmosphere to the lithosphere [10]. Furthermore, some C<sub>4</sub>-cycle genes and intermediates were found in a few algal species [11–13], although the presence of the whole pathway was not confirmed in algae. Therefore the phylogenetic analysis of C<sub>4</sub> pathway-related (C<sub>4</sub>-related) genes using recently released algal genomic and transcriptomic data has important scientific values for understanding the origin and evolution of photosynthesis.

With the development of the new generation sequencing (NGS) technology, more algal genome and transcriptome sequencing data were released (Table S1, including their references), including several species of marine phytoplankton such as diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricorutum*, green algae *Ostreococcus tauri* and *Micromonas*, brown alga *Ectocarpus siliculosus*, and red algae *Pyropia haitanensis* and *Pyropia yezoensis*. In addition, as a part of the recent 1000 Plant (OneKP) Project (<http://www.onekp.com>), we provided 19 marine phaeophytes and 21 marine rhodophytes for transcriptome sequencing. These new released algal sequencing data provide us more valuable gene information and more extensive algal lineages to screen the key genes in algae. Therefore, in this study, by analyzing these sequencing data, we confirmed the existence of C<sub>4</sub>-related genes in extensive algal lineages. We further resolved to perform a comprehensive phylogenetic analysis using a much larger dataset (especially including much more diverse algal lineages, land plants, cyanobacteria,  $\gamma$ -proteobacteria and archaea) to elucidate the evolution of C<sub>4</sub>-related genes and possible pathway types. Our analyses, for the first time, support a non-cyanobacterial origin for almost all of the C<sub>4</sub>-related genes, with the exception of phosphoenolpyruvate carboxykinase gene (*PCK*) and aspartate aminotransferase (*AST*), which have both cyanobacterial endosymbiont and non-cyanobacterial origin. This result is quite different from the previous study on C<sub>3</sub> photosynthesis origin from a cyanobacterial primary endosymbiont.

## Results

### Sequencing yield and annotation of unigenes

The transcriptomic sequencing of 21 red and 19 brown algal species yielded a total of 503,310,608 raw reads, equal to the length of 89.2 Gb with an average sequence length of 180 bases. The reads were assembled into 2,161,986 scaffolds, with an average length of 717 bp and an N50 of 1751 bp. All sequences were aligned against the local nr protein database downloaded from NCBI using the BLASTx algorithm for gene annotation. When the E-value cutoff was set at 10<sup>-5</sup>, a total of 585,247 unigenes had significant BLAST matches.

### KEGG pathway analysis of CO<sub>2</sub> fixation genes in diverse algal lineages

These transcriptomic data were further performed on KEGG pathway analysis. We also included genomic data from 19 sequenced algae (Table S1) for a thorough identification of genes

encoding enzymes related to the carbon fixation pathways in detail and all enzymes in C<sub>3</sub> and C<sub>4</sub> pathways among diverse algal groups (Chlorophyta, Rhodophyta, Ochrophyta, Glaucophyta and Cryptophyta) (Fig. 1). The numbers of gene in C<sub>3</sub> and C<sub>4</sub> pathways in algal species and some representative tracheophytes are listed in Table 1. Our results provide an unequivocal molecular evidence that most of the C<sub>3</sub>-pathway and C<sub>4</sub>-pathway genes are actively transcribed in these algal groups and demonstrate the possibility of the extensive existence of different photosynthetic pathways in algae.

### Phylogenetic analysis of C<sub>4</sub>-related genes

In addition of diverse algal species and tracheophytes, wealth of candidate C<sub>4</sub> gene sequences (Table S2) were also detected among archaea, proteobacteria and cyanobacteria. Therefore, we built phylogenetic trees that display relationships of full amino acid sequences of C<sub>4</sub> related genes from archaea, proteobacteria, cyanobacteria, tracheophytes, and algae based on Bayesian method (only representative candidates are included to save space). The results show that the eukaryotic C<sub>4</sub> enzymes have an archaeal/proteobacterial core (Fig. 2, 3, and 4). The phylogenetic trees of *PEPC* (phosphoenolpyruvate carboxylase), *PPDK* (pyruvate, orthophosphate dikinase), *ALT* (alanine transaminase), *MDH* (malate dehydrogenase), *ME* (malic enzyme), and *PK* (pyruvate kinase) (see Figs. 2, 3, and 4) support their non-cyanobacterial origin in primary endosymbiotic algae. However, almost all eukaryotic *PCK*s have a cyanobacterial origin through endosymbiosis gene transfer (EGT). Some genes of ochrophytes and cryptophytes potentially have a red algal origin (e.g., *ME*) as expected under the secondary endosymbiosis hypothesis, and others have archaeal/proteobacterial origins inherited from their endosymbiosis host genomes or acquired from non-cyanobacterial archaea or proteobacteria via horizontal gene transfer (HGT) (e.g., *PEPC*).

### *PEPC* originated from archaea/proteobacteria and evolved into diverse types in different eukaryotic organisms

*PEPC* (EC 4.1.1.31) is an important enzyme for CO<sub>2</sub> fixation in the C<sub>4</sub>-cycle and shared by three subtypes of C<sub>4</sub>-pathway metabolisms. The enzyme is present in almost all plants, green algae, cyanobacteria, most archaea, and non-photosynthetic bacteria, but is absent from animals and fungi [14]. In this study, we, for the first time, find its homologous sequences widely exist in various algal species, including chlorophytes, ochrophytes, rhodophytes, glaucophytes and cryptophytes (Fig. 1).

There are two homologous of *PEPC* genes in chlorophytes and tracheophytes, termed as bacterial-type (BTPC) and plant-type (PTPC) according to their amino acid sequences and structures [15]. Especially, the deduced *PEPC* polypeptides are readily classified as a BTPC or PTPC based on their C-terminal tetrapeptide, which is either (R/K) NTG for BTPCs or QNTG for PTPCs [16]. We compared the full deduced sequence of *PEPC*s among all above five algal groups (chlorophytes, ochrophytes, rhodophytes, glaucophytes and cryptophytes), cyanobacteria, proteobacteria and archaea. The results show that rhodophytes only possess PTPCs, whereas ochrophytes, glaucophytes, proteobacteria, and cyanobacteria only contain BTPCs. *PEPC*s of archaea, cryptophytes and ciliates do not have the typical C-terminal tetrapeptide and can not be classified into any known types.

Phylogenetic analysis suggests that *PEPC* broadly distributes among algae and plants and is likely to have a non-cyanobacteria

origin. Within the strict consensus tree, archaeal PEPCs is the first group to diverge at the base of the tree, followed by ciliates clade, BTPCs clusters, cryptophytes clade, and a large group including PTPCs of chlorophytes and rhodophytes which form a single well-resolved clade in two clusters with strong support (Bayesian posterior probability, PP = 0.98) (Fig. 2A). This topology suggests that BTPCs maybe more primitive than PTPCs and the latter one presumably have arisen from the BTPC homologs. Furthermore, PTPC can be classified into C<sub>3</sub>-type, C<sub>3</sub>-like-type, C<sub>3</sub>-C<sub>4</sub> intermediate type, and C<sub>4</sub>-type according to the sequence context (the amino acid residue at 774 or those around it at the C-end of PEPCs) [17]. All C<sub>4</sub>-type PTPCs investigated to date harbor a serine (S) residue at the corresponding position but this very site is replaced by an alanine (A) residue in all non-photosynthetic PEPCs. Site-specific mutagenesis studies demonstrated that this amino acid residue plays a key role in enzyme kinetics [18]. In addition, we found that the differentiation between non-C<sub>4</sub>-type and C<sub>4</sub>-type is not limited in PTPCs but also happens in BTPCs (Fig. 5). According to the criteria mentioned above, BTPCs of chlorophytes and tracheophytes are identified as non-C<sub>4</sub>-type, whereas in ochrophytes, diatoms possess non-C<sub>4</sub>-type site and Phaeophyceae algae prefer to C<sub>4</sub>-type. Interestingly, fractions of cyanobacteria and archaea also have the C<sub>4</sub>-type site (Fig. 5).

### PPDK, ALT, and PK have a non-cyanobacteria archaea/proteobacteria origin

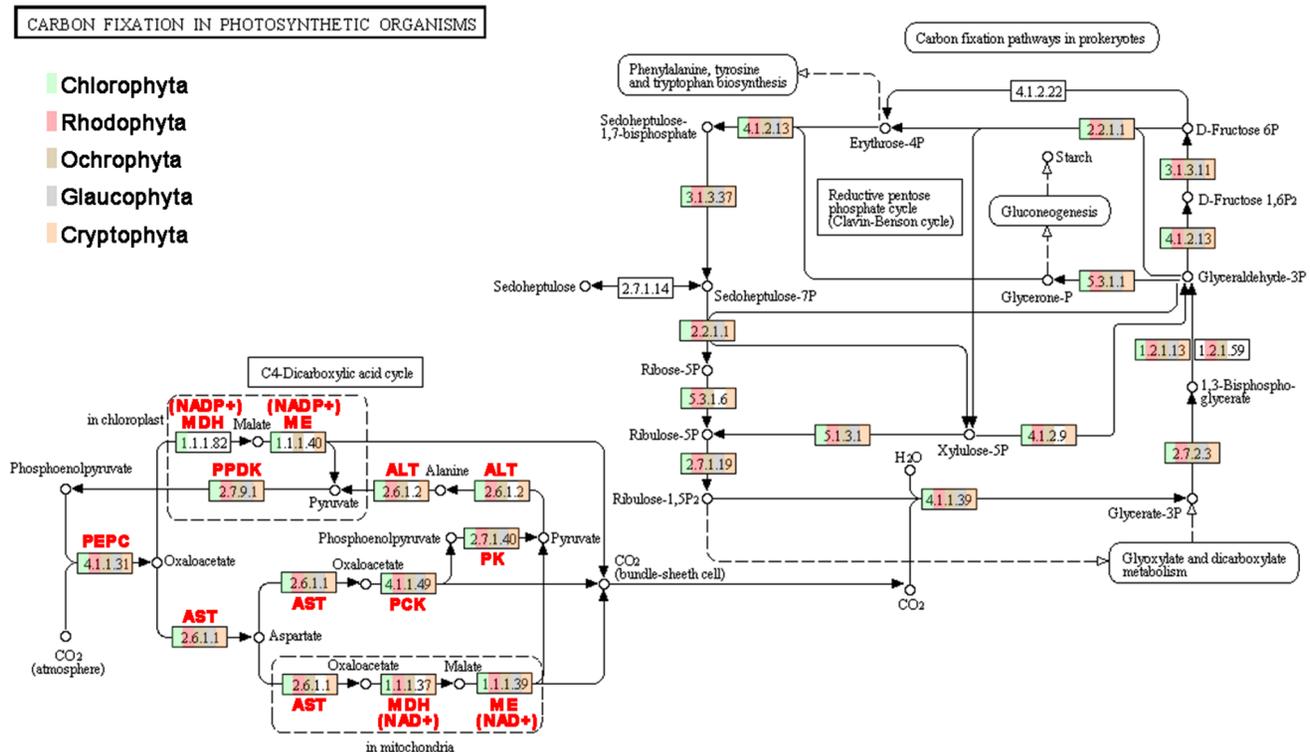
The enzyme PPDK (EC 2.7.9.1), another shared enzyme in the C<sub>4</sub> cycle (Fig. 1), catalyzes the reversible phosphorylation of pyruvate and inorganic phosphate yielding P-enolpyruvate and

inorganic pyrophosphate at the expense of a single molecule of ATP [19]. Using present gene data in NCBI, *PPDK* homologs have not been detected in the available cyanobacterial genomes. Actually we find that all published cyanobacterial *PPDK* are not real *PPDK*, for they lack the *PPDK* N-terminal nucleotide-binding domain. Instead, they should be classified as other PEP-utilizing genes, such as phosphoenolpyruvate synthetase (*PEPS*; pyruvate, water dikinase) (Fig. 6).

In the consensus tree of *PPDK* (Fig. 2B), archaeal sequences also cluster at the base like the PEPC tree, neighbored by clades of proteobacteria and protists, indicating its origin from archaea/proteobacteria. A separate clade is formed by rhodophytes with moderate bootstrap support, while Phaeophyceae algae (Ochrophyta) cluster independently and appear as the sister taxon of rhodophytes. Tracheophytes, chlorophytes, diatoms, and cryptophytes form another large clade. Therefore, primary endosymbiotic algal *PPDK* gene seems originated from archaea or proteobacteria, while Phaeophyceae algae possibly inherit the gene from red algal endosymbionts.

Similar to *PPDK*, neither dose *ALT* have any cyanobacterial homolog, indicating the eukaryotic *ALT* is also non-cyanobacterium-derived gene (Fig. 2C). Cryptophytes and diatoms formed a well supported clade with rhodophytes, which indicates their revolution relationship with red algae endosymbionts.

The analysis of *PK* genes is much more complicated due to its enormous duplications in eukaryotes. Our phylogenetic analysis also excluded the cyanobacterial origin of *PK* in algae and tracheophytes (Fig. 2D). Rhodophytes, glaucophytes, chlorophytes, and tracheophytes obtained this gene from archaea/



**Figure 1. Genes identified for coding enzymes in CO<sub>2</sub> fixation pathways in algal species generated by KEGG.** Different colors in the small boxes represent the genes identified from Rhodophyta, Ochrophyta, Chlorophyta, Glaucophyta and Cryptophyta, respectively. The numbers within the small boxes are the enzyme commission (EC) number. PEPC: phosphoenolpyruvate carboxylase; PPDK: pyruvate, orthophosphate dikinase; ALT: alanine transaminase; MDH: malate dehydrogenase; ME: malic enzyme; AST: aspartate aminotransferase; PK: pyruvate kinase; PCK: phosphoenolpyruvate carboxykinase.  
doi:10.1371/journal.pone.0110154.g001

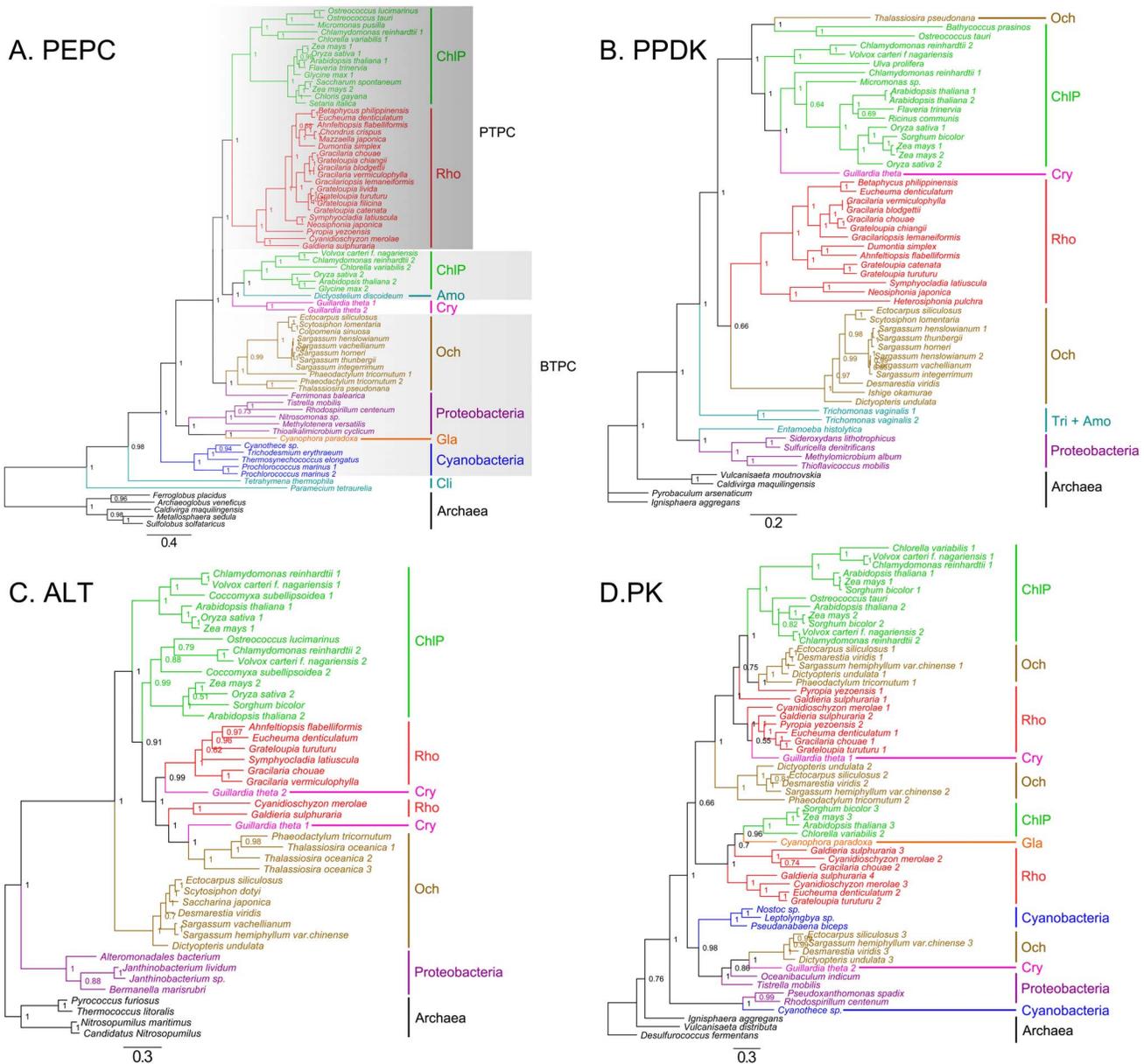
**Table 1.** The numbers of gene in C<sub>3</sub> and C<sub>4</sub> pathways in algae and tracheophytes species.

Taxonomy	Organism	ALT	AST	MDH	ME	PCK	PEPC	PK	PPDK
Chlorophyta	<i>Bathycoccus prasinos</i>								1
Chlorophyta	<i>Chlamydomonas reinhardtii</i>	2	5	4	2	1	2	2	2
Chlorophyta	<i>Chlorella variabilis</i>					2	2	2	
Chlorophyta	<i>Coccomyxa subellipsoidea</i>	2							
Chlorophyta	<i>Micromonas pusilla</i>						1		
Chlorophyta	<i>Micromonas sp.</i>	1							
Chlorophyta	<i>Ostreococcus lucimarinus</i>	1	2	2	3		1		
Chlorophyta	<i>Ostreococcus tauri</i>		2		2		1	1	1
Chlorophyta	<i>Volvox carterii f. nagariensis</i>	2	2	2	1	2	1	2	1
Cryptophyta	<i>Guillardia theta</i>	2	7	3	6	1	2	2	1
Glaucophyta	<i>Cyanophora paradoxa</i>		4	4		1	1	1	
Ochrophyta	<i>Aureococcus anophagefferens</i>	1						2	
Ochrophyta	<i>Nannochloropsis gaditana</i>		5	1	2	2	1	5	
Ochrophyta	<i>Nannochloropsis oceanica</i>								
Ochrophyta	<i>Colpomenia sinuosa</i>	1	2	2	1	1	1	3	
Ochrophyta	<i>Desmarestia viridis</i>	1	3	1	2	1		3	1
Ochrophyta	<i>Dictyopteris undulata</i>	1	2	2	2			3	1
Ochrophyta	<i>Ectocarpus siliculosus</i>	1	3	2	2	1	1	3	1
Ochrophyta	<i>Ishige okamurae</i>			1	1	1		3	1
Ochrophyta	<i>Petalonia fascia</i>	1	3	1	2	1		2	
Ochrophyta	<i>Phaeodactylum tricornutum</i>	1	4	1	2	1	2	2	
Ochrophyta	<i>Punctaria latifolia</i>	1	3	2	3	1		3	
Ochrophyta	<i>Saccharina japonica</i>	1	2	2	1	1		2	
Ochrophyta	<i>Saccharina sculpera</i>	1	3	2	2	1			
Ochrophyta	<i>Sargassum fusiforme</i>	1	2	1	1				
Ochrophyta	<i>Sargassum hemiphyllum var. chinense</i>	1	3	2		1		3	1
Ochrophyta	<i>Sargassum henslowianum</i>	1	2	1	1	1	1	2	1
Ochrophyta	<i>Sargassum horneri</i>	1	2	2	2	1	1	2	1
Ochrophyta	<i>Sargassum integerrimum</i>	1	3	1	1	1	1	2	1
Ochrophyta	<i>Sargassum muticum</i>	1	2	2	2	1			
Ochrophyta	<i>Sargassum thunbergii</i>	1	4	2	1	1	1	2	1
Ochrophyta	<i>Sargassum vachellianum</i>	1	3	2	2	1	1	3	1
Ochrophyta	<i>Scytosiphon dotyi</i>	1	2	2	2	1		3	
Ochrophyta	<i>Scytosiphon lomentaria</i>	1	3	2	1	1	1	3	1
Ochrophyta	<i>Thalassiosira oceanica</i>	3	1	1		1			

Table 1. Cont.

Taxonomy	Organism	ALT	AST	MDH	ME	PCK	PEPC	PK	PPDK
Ochrophyta	<i>Thalassiosira pseudonana</i>		3	2	1	1	1		1
Ochrophyta	<i>Undaria pinnatifida</i>	1	4	2	1	1		2	
Rhodophyta	<i>Ahnfeltiopsis flabelliformis</i>	1	1	1	1		1	2	1
Rhodophyta	<i>Betaphycus philippinensis</i>	1	1	1	1		1	2	1
Rhodophyta	<i>Ceramium kondoi</i>	1		2	1			2	
Rhodophyta	<i>Chondrus crispus</i>		2	1	2		1	2	
Rhodophyta	<i>Cyanidioschyzon merolae</i>	1	2	2	1		1	3	
Rhodophyta	<i>Dumontia simplex</i>	1	1	1	1		1	2	1
Rhodophyta	<i>Eucheuma denticulatum</i>	1	2	1	2		1	2	1
Rhodophyta	<i>Galdieria sulphuraria</i>	1	3	3	1	2	1	4	
Rhodophyta	<i>Gloiopeltis furcata</i>	1	1		2			2	
Rhodophyta	<i>Gracilaria blodgettii</i>	1	1	1	1		1	2	1
Rhodophyta	<i>Gracilaria chouae</i>	1		1	1		1	2	1
Rhodophyta	<i>Gracilaria vermiculophylla</i>	1	2	1	1		1	2	1
Rhodophyta	<i>Gracilariopsis lemaneiformis</i>	1	1	2	1		1	2	1
Rhodophyta	<i>Grateloupiia catenata</i>	1	1	1	1		1	2	1
Rhodophyta	<i>Grateloupiia chiangii</i>		1	1	1		1	2	1
Rhodophyta	<i>Grateloupiia filicina</i>		1				1		
Rhodophyta	<i>Grateloupiia livida</i>	1	1		1		1	1	
Rhodophyta	<i>Grateloupiia turuturu</i>	1	2		1		1	2	1
Rhodophyta	<i>Heterosiphonia pulchra</i>	1		1	1			1	1
Rhodophyta	<i>Kappaphycus alvarezii</i>		1		1			1	
Rhodophyta	<i>Mazzaella japonica</i>		1	3	2		1	2	
Rhodophyta	<i>Neosiphonia japonica</i>	1	1	1	2		1	1	1
Rhodophyta	<i>Pyropia yezoensis</i>		3	1		1	1	2	
Rhodophyta	<i>Symphocladia latiuscula</i>	1	1	2	1		1	2	1
Tracheophytes	<i>Arabidopsis thaliana</i>	2	6	11	6	4	4	13	1
Tracheophytes	<i>Oryza sativa</i>	2	6	10	7	4	6	10	4
Tracheophytes	<i>Sorghum bicolor</i>	1	3	7	11	3	4	11	3
Tracheophytes	<i>Zea mays</i>	9	5	2	1	1	3	11	1
Bryophytes	<i>Physcomitrella patens</i>	8	3	11	4	4	7	15	5

doi:10.1371/journal.pone.0110154.t001



**Figure 2. Bayesian phylogenetic trees of PEPC (A) PPDK (B), ALT (C) and PK (D) enzymes with bootstrap values (when >50%) indicated at the nodes.** The PEPCs are classified as PTPCs (in dark gray boxes) and BTPCs (in light gray boxes). ChIP, chlorophytes and plants. Gla, glaucophytes. Rho, rhodophytes. Och, ochrophytes. Cry, cryptophytes. Tri, trichomonad. Amo, amoeba. Cil, ciliates. doi:10.1371/journal.pone.0110154.g002

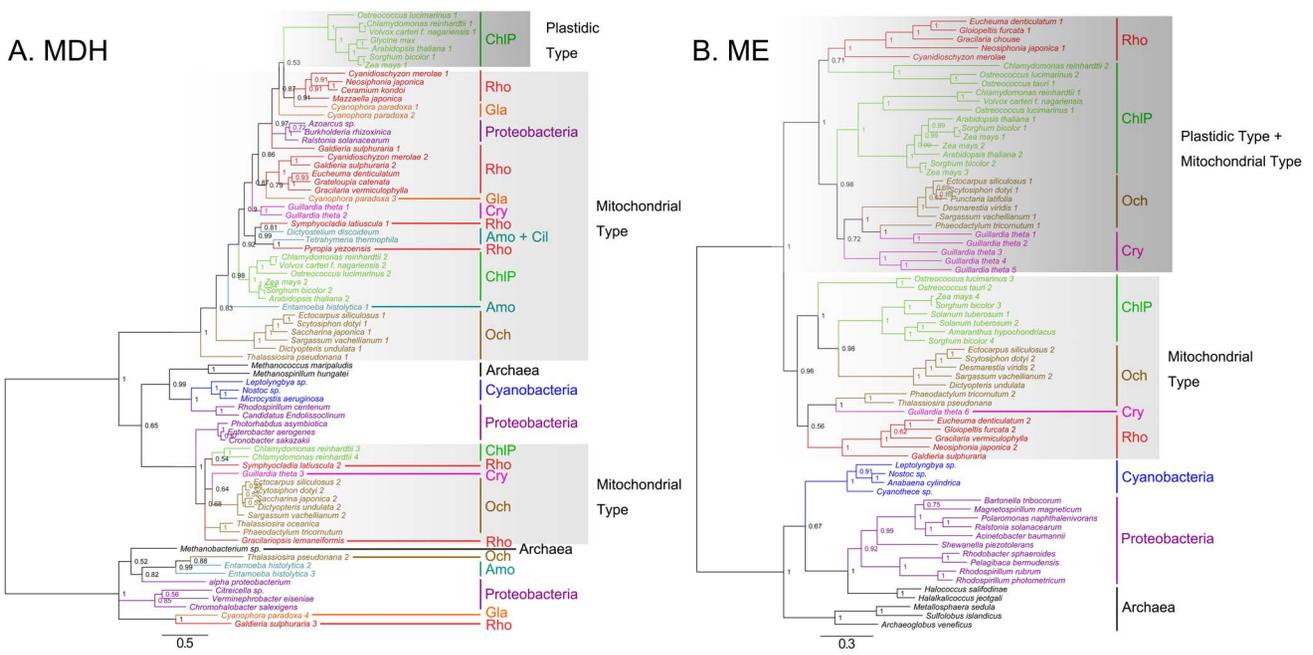
proteobacteria, and gene duplication occurred extensively at least twice in these species. Cryptophytes and Phaeophyceae algae have two types of *PK*, one is inherited from red algae through EGT (eg. *Guillardia theta* 1, EKX52520.1), another is transferred from some proteobacteria or archaea (eg. *Guillardia theta* 2, EKX43540.1). The latter type is not found in diatoms, and this absence suggests different evolutionary pattern, such as gene lost, between Phaeophyceae algae and diatoms.

Note that Sogin and Zillig *et al.* theorized that the eukaryotic nucleus arose from the cellular fusion between both a bacterium and an archaeon [20]. Lang *et al.* indicated several hundred eukaryotes genes are acquired from the mitochondrial ancestor derived from an endosymbiotic alpha-proteobacterium that was engulfed by a eukaryotic- or archaeobacteria-like cell more than

one billion years ago [21]. Therefore, we suggest that the archaea/proteobacteria-derived C<sub>4</sub>-related genes in eukaryotic organisms may inherit from the endosymbiotic host genome, which had acquired numerous genes from bacterium and archaeon or transferred from archaea/proteobacteria through HGT.

**MDH and ME originated from archaea/proteobacteria and evolved into different homologs after duplication**

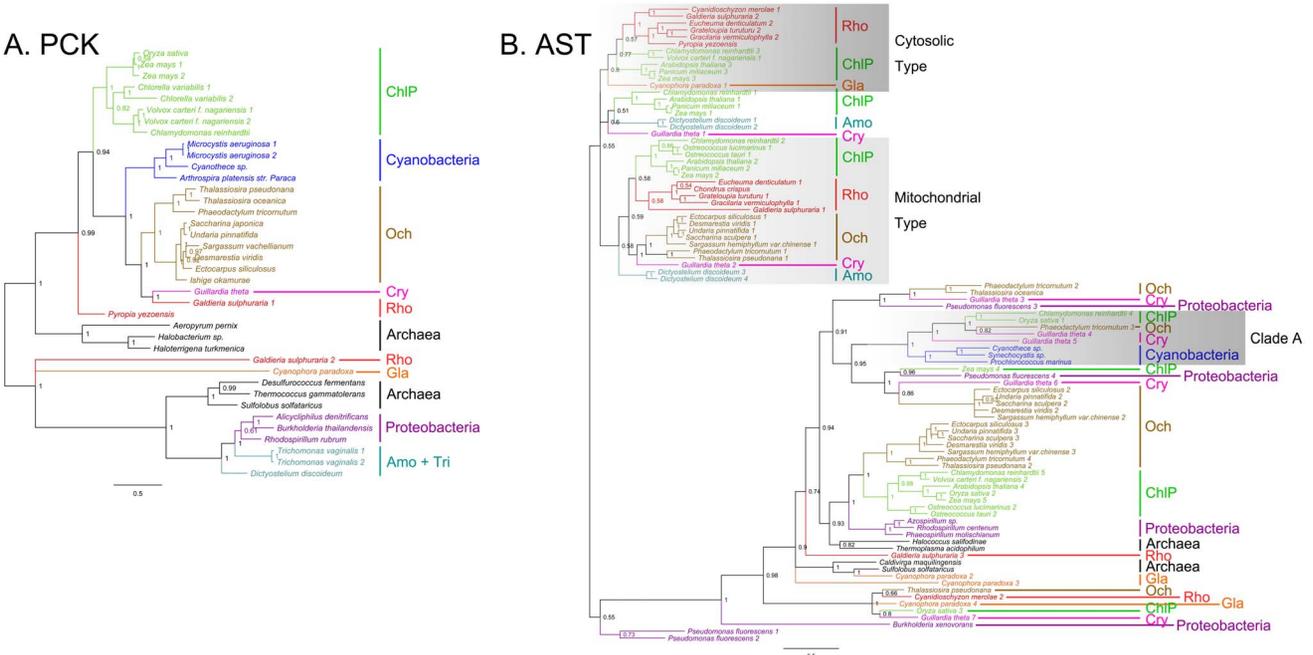
The phylogenic analysis of *MDH* and *ME* has encountered difficulties from sequence similarity between NADP- and NAD-dependent homologs. Therefore, we built phylogenies separately for *MDH* and *ME*, albeit they probably shared a common ancestry (Fig. 3). According to the phylogenetic tree, it is clear that eukaryotic *MDHs* are considered essentially archaeal/proteobac-



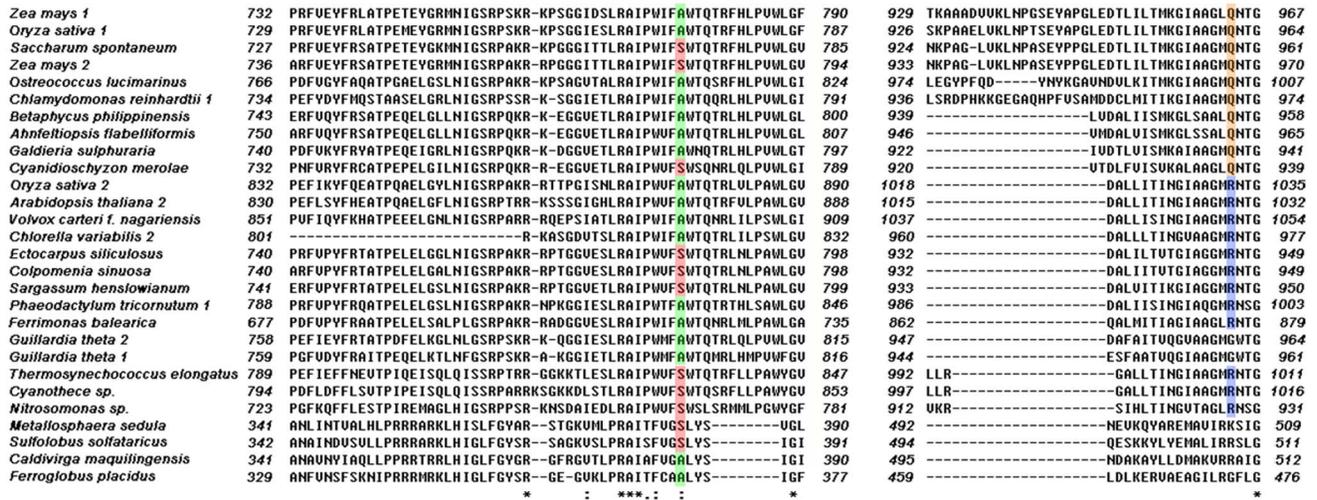
**Figure 3. Bayesian phylogenetic trees of MDH (A) and ME (B) enzymes with bootstrap values (when >50%) indicated at the nodes.** The plastic type MDHs are in dark gray boxes, and mitochondrial types in light gray boxes. Eukaryotic MEs cluster in two clades, one clade in dark gray boxes consists of both plastic and mitochondrial type, the other one only contains mitochondrial type. ChIP, chlorophytes and plants. Gla, glaucophytes. Rho, rhodophytes. Och, ochrophytes. Cry, cryptophytes. Tri, trichomonad. Amo, amoeba. Cil, ciliates. doi:10.1371/journal.pone.0110154.g003

terial in provenance (Fig. 3A). Targeting signal predictions for plastid and mitochondrion genes in algae and plants suggest that the gene encoding plastid-targeted MDH in tracheophytes and chlorophytes likely evolved from eukaryotic algal mitochondrial

homologs via gene duplication and acquisition of a plastid-import signal. Ochrophyte MDH has two separate origins: one from red algae endosymbionts (eg. *Saccharina japonica* 2, KM113483), the other from secondary endosymbiotic hosts or archaea/proteobac-



**Figure 4. Bayesian phylogenetic trees of PCK (A) and AST (B) enzymes with bootstrap values (when >50%) indicated at the nodes.** The cytosolic types of AST are in dark gray boxes and mitochondrial types are in light gray boxes. Besides this two clades, some chlorophyte, plant and cryptophyte ASTs are grouped with cyanobacteria (Clade A), which are also in light gray boxes. ChIP, chlorophytes and plants. Gla, glaucophytes. Rho, rhodophytes. Och, ochrophytes. Cry, cryptophytes. Tri, trichomonad. Amo, amoeba. Cil, ciliates. doi:10.1371/journal.pone.0110154.g004



**Figure 5. Alignments of partial amino acid sequences of PEPC isoenzymes between species of Rhodophyta, Ochrophyta, chlorophytes, Cryptophyta, bacteria and archaea.** The deduced PEPC sequences were aligned using ClustalX 1.83 software. Numbering indicates the position of the first and last residue in each aligned sequence. Semi-colons and asterisks indicate identical and conserved amino acids respectively. The C-terminal tetrapeptide (R/K)NTG for BTPCs is highlighted in blue color, whereas QNTG for PTPCs is highlighted in orange. The amino acid residue of No. 774 or around it is highlighted in green (non-photoautotrophic PEPCs) or red (C<sub>4</sub>-type PEPCs). doi:10.1371/journal.pone.0110154.g005

teria through HGT (eg. *Saccharina japonica* 1, KM113482). Cryptophytes only acquired *MDH* from red algal endosymbionts (Fig. 3A).

The strict *ME* consensus tree displays a relatively simple topology compared to the *MDH* tree. It shows that the *ME* genes are also obtained very early in algal evolution but start to duplicate before the emergence of primary endosymbiotic algae, might be in its last eukaryotic common ancestor. After duplication, some Archaeplastida (including red algae, green algae and plants) obtained two homologs of *ME* from archaea/proteobacteria, and these genes are likely to be transmitted into ochrophytes and cryptophytes by EGT from their red algal endosymbionts (Fig. 3B). Targeting signal predictions suggest two clades with different organellar locations. One clade is composed of plastid target genes and the other consists of both plastid and mitochondrial target genes.

**PCK and AST originated from both cyanobacteria and archaea/proteobacteria**

The remaining C<sub>4</sub> genes, *PCK* and *AST*, have a complex evolutionary history. Our phylogenetic analysis shows they may have arisen from multiple origins (Fig. 4A). For instance, two copies of *PCKs* from *Galdieria sulphuraria* cluster into separate clades: one is related to archaea and proteobacteria (eg. *Galdieria sulphuraria* 2, EME28834.1), the other groups with cyanobacteria (eg. *Galdieria sulphuraria* 1, EME27660.1). Therefore, rhodophytes may acquire *PCK* from two different ancestors. In contrast, glaucophytes possess *PCKs* of archaea/proteobacteria origin. Ochrophytes and cryptophytes may obtain *PCKs* from red algal endosymbionts.

The evolution of eukaryotic *AST* genes appears much more complicated. In our phylogenetic tree, the enzymes clustered into two separate clades. One clade consists of cytosolic and mitochondrial *ASTs*. In the constricted tree (Fig. 4B), mitochon-



**Figure 6. Partial sequence alignment of the N-terminal domains of PEPS and PPK enzymes.** Numbering indicates the position of the first and last residue in each aligned sequence. Proteins with biochemical enzyme activity are indicated by gray shading. Gray shading indicate the identified sequence signatures specific for PEPS and PPK respectively. doi:10.1371/journal.pone.0110154.g006

drial *AST* is present in tracheophytes, chlorophyte, rhodophyte, ochrophyte, and cryptophyte algae, and the cytoplasmic *AST* is present in archaeplastida (Rhodophyta, Viridiplantae and Glaucophyta) and forms a monophyletic group without any secondary endosymbiotic algae. These species all group with homologs of proteobacteria, suggesting their proteobacteria origin. However interestingly, in our phylogenetic tree, in addition to the mitochondrial and cytoplasmic types, *AST*s of some chlorophyte and cryptophyte algae are grouped into another clade (clade A) with cyanobacteria, indicating their cyanobacteria endosymbiont origin.

## Discussion

### C<sub>4</sub> genes distributed among diverse algae mostly with archaeal/proteobacterial origins

Our new data allow the identification of almost all the enzymes necessary for C<sub>4</sub> photosynthesis widely distributing among different eukaryotic algal lineages (Fig. 1). Furthermore, we find that most of eukaryotic C<sub>4</sub> genes have an archaeal/proteobacterial core but algal *PCK*s and *AST*s originated from both cyanobacteria and archaea/proteobacteria, based on our phylogenetic analyses by using diverse genes from archaea, proteobacteria, cyanobacteria, tracheophytes and various algae groups (Fig. 2, 3, 4).

Notably, we found that cyanobacteria do not possess complete C<sub>4</sub> pathway. Though *PEPC*s are proved existed in cyanobacteria (Fig. 2A), the cyanobacterial *PPDK* sequences published in GenBank are excluded from our analysis (Fig. 2B), because we for the first time found that the published cyanobacterial *PPDK* sequences should actually be annotated as *PEPS* or other PEP-utilizing genes because of the lacking of the *PPDK* N-terminal nucleotide-binding domain [22]. Neither are *ALT* and *PK* found their homologs in cyanobacteria (Fig. 2C, D), suggesting that the C<sub>4</sub> cycle does not arise from a cyanobacterial endosymbiosis. Our further phylogenetic analysis strongly supports this hypothesis that almost all of the C<sub>4</sub>-related genes have a non-cyanobacteria origin (Fig. 2B).

We also searched the cyanobacterium-like plastid (also termed as cyanelle) genome of Glaucophyta *Cyanophora paradoxa*, which is considered as a “living fossil” and a paradigm for the invasion of a eukaryotic cell by a cyanobacterium [23], but did not find any complete C<sub>4</sub> pathway, except for partial sequences of *PPDK* and *ALT* homologs (data not shown). As nucleomorphs are considered as the enslaved red or green algal nucleus residuals, we further searched the nucleomorph genomes of Cryptophyta *Guillardia theta* (AF165818.4, AJ010592.2, AF083031.2), *Hemiselmis andersenii* (CP000881.1, CP000882.1, CP000883.1), *Cryptomonas paramecium* (CP002172.1, CP002173.1, CP002174.1) and *Cercozoa Bigelowiella natans* (DQ158856.1, DQ158857.1, DQ158858.1), and did not find any C<sub>4</sub> genes, neither. The results are consistent with our hypothesis that the C<sub>4</sub> cycle is not evolved from cyanobacterial-endosymbiotic event.

### C<sub>4</sub>-related genes of diverse subtypes in different algal lineages

In land plants, C<sub>4</sub> photosynthesis are typically divided into three subtypes-NADP-ME, NAD-ME, and PCK-based on different decarboxylation models. Different C<sub>4</sub> species often use one of the three subtypes nearly exclusively [24]. Some experimental results suggest that the PCK subtype is maximal in biomass production and CO<sub>2</sub> fixation [25,26], and others show that the NAD-ME and PCK subtypes are more adaptive than the NADP-ME subtype for species in arid environments [24].

By surveying a selection of cyanobacterial and algal species to determine the gene composition (Table S3), we found all chlorophytes own all three varieties of C<sub>4</sub>-related genes, just like their derived lineage, tracheophytes. However, virtually all species of rhodophytes and ochrophytes, including diatoms which have been proved to be able to incorporate CO<sub>2</sub> into the C-4 carboxyl of C<sub>4</sub> acids [11], own NAD-ME and PCK subtypes genes, but lack the *MDH* (NADP) of NADP-ME subtype. This result suggests that diatoms, red algae and brown algae quite possibly have similar C<sub>4</sub>-related genes, and NADP-ME type C<sub>4</sub> genes may emerge in chlorophytes after they separated from other algal lineages.

In tracheophytes, the cytosolic and mitochondrial *AST* activities are connected by C<sub>4</sub> photosynthesis and they both participate in the NAD-ME subtype of C<sub>4</sub> pathway, while cytosolic *AST* itself can function as the PCK type [27]. In our analysis, Archaeplastida all evolved to have both cytosolic *AST* and mitochondrial *AST* except that glaucophytes only have cytosolic *AST* (Fig. 4B). In addition, no known C<sub>4</sub> cytosolic *AST* homolog has been annotated in ochrophytes and cryptophytes, but *AST*s of these secondary endosymbiotic algae are grouped into another clade with proteobacteria/archaea (Fig. 4B), indicating different evolution of C<sub>4</sub>-related genes between primary and secondary endosymbiotic algae.

### Gene duplication of C<sub>4</sub>-related enzymes

Our phylogenetic analyses also suggest that most C<sub>4</sub>-related genes undergo gene duplications subsequently at different evolutionary time scale (Fig. 2, 3, 4). Furthermore, there are diverse gene duplications among different eukaryotic lineages. For instance, there are two types of *PEPC* termed as PTPC and BTPC in chlorophytes and tracheophytes, which demonstrate that the *PEPC* gene duplication had already occurred in the ancestral chlorophytes, with BTPCs being the ancestral type.

Our study showed that the primary endosymbiotic algae evolved to have both cytosolic *AST* and mitochondrial *AST* except glaucophytes may only have cytosolic *AST* (Fig. 4B), suggesting *AST* duplicated and functional differentiated in a common algal ancestor before the primary endosymbiosis, and the absence of mitochondrial *AST* in glaucophytes may be due to gene lost. Another group of *AST* found in the secondary endosymbiotic algae, including diatoms, Phaeophyceae and cryptophytes (Fig. 4B), may also be the results of gene lost and functional differentiation after gene duplication which happened in their common ancestor before the secondary endosymbiosis.

Gene duplication and evolution in *MDH* and *ME* are even more fascinating. NADP-dependent and NAD-dependent *MDH*s of the two separate C<sub>4</sub> subtypes seem to share the same ancestor, and duplication took place in the last common ancestor of chlorophytes and tracheophytes (Fig. 3A). The same event happened during evolution of *ME*, yet duplication maybe occurred earlier in the last eukaryotic common ancestor (Fig. 3B). The ancestral mitochondrial enzymes endured duplication and gave rise to homologs which evolved plastid target signal peptide. These plastid *MDH* and *ME* genes finally took part in the construction of NADP-ME type C<sub>4</sub> pathway in tracheophytes.

Since gene duplication is a necessary contribution to genetic novelty and adaptation and requirement [28], the multiple copies of C<sub>4</sub>-related genes should have arisen accompanying the emergence of different C<sub>4</sub> cycle which made the regulation more delicate during the evolution of C<sub>4</sub> photosynthesis.

### Origin and evolution of C<sub>4</sub> pathway

Our analysis suggests that the C<sub>4</sub>-related genes (except for *PCK* and *AST*) in eukaryotic organisms originated from archaea/

proteobacteria. It is widely accepted that the eukaryotic nucleus may have arisen from the cellular fusion between either a bacterium or “protoeukaryote” and an archaeon [20]. Therefore, a substantial number of archaea/proteobacteria-derived genes are retained in the eukaryotic nuclear genome. Another mechanism for algae and plants acquiring archaea/proteobacteria genes is horizontal gene transfer (HGT) from archaea or proteobacteria that occurred throughout the history of eukaryotes [29]. Thus, the archaea/proteobacteria-derived C<sub>4</sub>-related genes may inherit from the endosymbiotic host genome or transferred from archaea/proteobacteria through HGT.

Previous study showed that photosynthetic eukaryotes (i.e., algae and plants) gained C<sub>3</sub> photosynthesis from a cyanobacterial primary endosymbiont [30]. The eukaryotic C<sub>3</sub> pathway consists of eleven enzymes, and molecular phylogenetic analyses show that recruitment of single enzymes from different pathways could be the driving force for C<sub>3</sub> pathway evolution in chlorophytes, tracheophytes and rhodophytes [31,32]. Therefore, we would like to propose a hypothesis on C<sub>4</sub> cycle formation.

First, when a cyanobacterium was engulfed and retained by a heterotrophic eukaryote, its *PCK* and *AST* genes transferred to the host nucleus and were inherited by the primary endosymbiotic algae to lay out the main framework of PCK subtype of C<sub>4</sub> pathway with some archaea/proteobacteria-derived C<sub>4</sub> genes founding in plants.

In addition to the PCK subtype, NAD-ME subtype was also constructed as the result of a patchwork assembly in tracheophytes, but it is difficult to predict which one is more primitive. However, the present work proved that NADP-ME subtype should be a derivant of NAD-ME subtype according to our phylogenetic analysis of *MDH* and *ME* genes. The plastid *MDHs* in chlorophytes and tracheophytes are originated from its mitochondrial homologs via duplication (Fig. 3A), and a clade contains both mitochondrial and plastid *ME* homologs are originated from mitochondrial ones (Fig. 3B).

The changing CO<sub>2</sub> concentration may also be a major environmental driving force for eukaryotes to develop C<sub>4</sub> metabolism, to suppress photorespiration. Throughout Earth's geological history, eukaryotes had been exposed to much higher CO<sub>2</sub> at the beginning of evolutionary history but then became starved by a steep decrease of CO<sub>2</sub> and increase of O<sub>2</sub> as the outcome of appearance of C<sub>3</sub> photosynthesis. Some plants can dramatically change the photosynthetic and anatomical traits to meet different environment pressure. For example, *Eleocharis vivipara*, an amphibious sedge, changes its photosynthetic pathway from C<sub>3</sub> to C<sub>4</sub> under conditions of CO<sub>2</sub> deficiency [33]. So did in aquatic environments, in order to reduce the impact of increasing oxygen pressures, aquatic photosynthetic organisms evolved special mechanisms to efficiently maintain carbon fixation. It has been demonstrated that many aquatic photosynthetic organisms can take up both CO<sub>2</sub> and HCO<sub>3</sub><sup>−</sup> from the surrounding media and this capacity is greatly strengthened under CO<sub>2</sub>-limiting conditions, which is generally known as the inorganic carbon-concentrating mechanism (CCM) [34].

For algae, it is believed that the C<sub>3</sub> cycle is predominant in the CO<sub>2</sub> fixation pathway [35], as a result of the endosymbiotic acquisition of a cyanobacterium that evolved into the chloroplasts [31]. However, the existence of a C<sub>4</sub> photosynthetic pathway in algae or marine phytoplankton is a long-standing debate [36]. Recent papers have reported the evidence for the operation of C<sub>4</sub> photosynthesis as an alternative CCM in the marine diatom *Thalassiosira weissflogii*, using <sup>14</sup>C-labeling experiments [11,12]. In addition, some intermediate products of the C<sub>4</sub> pathway had

been detected in brown alga *Ascophyllum nodosum*, euglenoid *Euglena gracilis*, and dinoflagellates [13], and C<sub>4</sub>-like photosynthetic characteristics had also been detected in green alga *Udotea flabellum* [37]. Until recently, few studies are focused on C<sub>4</sub>-cycle genes in limited algae species, and imperfect detection method in algae may bring about ambiguous results. Our current work provides molecular evidences for further biochemical and physiological experiment validation on more algal species to prove the existence of C<sub>4</sub>-related genes, or even C<sub>4</sub> cycle in some algal species or extensive algal lineages.

## Materials and Methods

### Ethics Statement

Marine brown and red algal samples were collected along the coast of China during October, 2010 to March, 2012, and were provided by Culture Collection of Seaweed at the Ocean University of China. The location is neither privately owned nor protected places. No specific permissions were required for these locations, and the study did not involve any endangered or protected species. The species information and the GPS coordinates of their specific locations are provided in Table S4.

### RNA extraction

Total RNA was extracted from algal tissues using an improved CTAB method for brown algal samples and an improved Trizol method for red algal samples, and RNA quantity and quality was examined as previous reports [38,39].

### Transcriptome sequencing and de novo assembly

cDNA library construction and sequencing were performed by the BGI (Shenzhen, China) on Illumina (San Diego, USA) HiSeq instruments in accordance with the manufacturer's instructions. Strict reads filtering was performed before *de novo* assembly using SOAP denovo-Trans (<http://soap.genomics.org.cn/SOAPdenovo-Trans.html>). Pair-end reads with primer or adaptor sequences were removed. Reads with more than 10% of the bases below Q20 quality or more than 5% of unknown nucleotides (Ns) were filtered from total reads. Gapcloser was further used for gap filling of the scaffolds.

### C<sub>4</sub>-cycle gene identification and KEGG pathway analysis

The assembled sequences were BLASTx against the nr protein database downloaded from NCBI with E-value < 10<sup>−5</sup>. Deduced algal C<sub>4</sub> protein coding sequences were further examined for their homology by using BLAST X and BLAST P available at NCBI. These sequences were uploaded to GenBank database, and their accession numbers were listed in Table S2. All other publicly available C<sub>4</sub> enzymes sequences were obtained from GenBank databases (Table S2). The accession numbers of gene in C3 and C4 pathways in tracheophytes species were listed in Table S5. To reconstruct the metabolic pathways, high-quality assembled algal sequences and 19 sequenced algal genome sequences downloaded from GenBank database (Table S1) were assigned to KEGG Automatic Annotation Server ([www.genome.jp/tools/kaas/](http://www.genome.jp/tools/kaas/)) to provide functional annotation of genes.

### Sequence alignments and phylogenetic analyses

Sequences were aligned with ClustalX 1.83 software before Bayesian analysis using MrBayes 3.1.2 software. Analyses were performed as two independent runs, each with four incrementally heated Metropolis-coupled Monte-Carlo Markov Chains running for 5,000,000 generations. Trees were sampled every 100 generations. A total of 50,001 trees in one files were read and

the first 25% of trees were discarded as the burn-in. The average standard deviation of split frequencies at the end of the run was below 0.01, indicating stationary conditions.

### Cellular target prediction

The probability of plastid or mitochondrion targeting was assessed according to the on-line prediction service at CBS (<http://www.cbs.dtu.dk/services/>) with the ChloroP V1.1 and TargetP V1.1 servers using the default values.

### Supporting Information

**Table S1** Update information of publicly available algal genome sequences. (XLSX)

**Table S2** Taxonomical list of taxa included in this study. The sequences used to reconstruct phylogenetic trees are written in bold letters, the other Rhodophyta and Ochrophyta full-length sequences in regular letters are also obtained in present study. Sequence data of *Cyanophora paradoxa* are acquired from Cyanophora Genome Database. (XLSX)

### References

- Raven JA (2009) Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments. *Aquat Microb Ecol* 56: 177–192.
- Tabita FR (2004) Research on carbon dioxide fixation in photosynthetic microorganisms (1971–present). *Photosynth Res* 80: 315–332.
- Falkowski PG, Fenchel T, DeLong EF (2008) The microbial engines that drive Earth's biogeochemical cycles. *Science* 320: 1034–1039.
- Hatch MD, Slack CR (1966) Photosynthesis by sugar-cane leaves. A new carboxylation reaction and the pathway of sugar formation. *Biochem J* 101(1): 103–111.
- Matsuoka M, Furbank RT, Fukuyama H, Miyao M (2001) Molecular engineering of C<sub>4</sub> photosynthesis. *Annu Rev Plant Physiol Plant Mol Biol* 52: 297–314.
- Brown RH (1999) Agronomic implications of C<sub>4</sub> photosynthesis. In: SageRF, MonsonRK, eds. *C<sub>4</sub> Plant biology*. USA: Academic Press 473–507.
- Hatch MD, Kagawa T, Craig S (1975) Subdivision of C<sub>4</sub>-pathway species based on differing C<sub>4</sub> acid decarboxylating systems and ultrastructural features. *Aust J Plant Physiol* 2(2): 111–128.
- Ueno O (2011) Structural and biochemical characterization of the C<sub>3</sub>–C<sub>4</sub> intermediate *Brassica graviniae* and relatives, with particular reference to cellular distribution of Rubisco. *J Exp Bot* 62(15): 5347–5355.
- Qiu YL and Palmer JD (1999) Phylogeny of early land plants: insights from genes and genomes. *Trends Plant Sci* 4: 26–30.
- Dorrell RG and Smith AG (2011) Do red and green make brown?: perspectives on plastid acquisitions within chromalveolates. *Eukaryot Cell* 10(7): 856–868.
- Reinfelder JR, Kraepiel AML, Morel FMM (2000) Unicellular C<sub>4</sub> photosynthesis in a marine diatom. *Nature* 407: 996–999.
- Reinfelder JR, Milligan AJ, Morel FMM (2004) The role of the C<sub>4</sub> pathway in carbon accumulation and fixation in a marine diatom. *Plant Physiol* 135: 2106–2111.
- Fan XL, Fang YJ, Hu SN, Wang GC (2007) Generation and analysis of 5318 expressed sequence tags from the filamentous sporophyte of *Porphyra haitanensis* (Rhodophyta). *J Phycol* 43: 1287–1294.
- Patel HM, Kraszewski JL, Mukhopadhyay B (2004) The phosphoenolpyruvate carboxylase from *Methanothermobacter thermoautotrophicus* has a novel structure. *J Bacteriol* 186: 5129–5137.
- O'Leary B, Park J, Plaxton WC (2011) The remarkable diversity of plant PEPCase (phosphoenolpyruvate carboxylase): recent insights into the physiological functions and post-translational controls of non-photosynthetic PEPCase. *Biochem J* 436(1): 15–34.
- Gennidakis S, Rao SK, Greenham K, Uhrig RG, O'Leary B, et al. (2007) Bacterial- and plant-type phosphoenolpyruvate carboxylase polypeptides interact in the hetero-oligomeric class-2 PEPCase complex of developing castor oil seeds. *Plant J* 52: 839–849.
- Westhoff P, Gowik U (2004) Evolution of C<sub>4</sub> phosphoenolpyruvate carboxylase. Genes and proteins: a case study with the genus *Flaveria*. *Ann Bot* 93(1): 13–23.
- Engelmann S, Blasing OE, Westhoff P, Svensson P (2002) Serine 774 and amino acids 296 to 437 comprise the major C<sub>4</sub> determinants of the C<sub>4</sub> phosphoenolpyruvate carboxylase of *Flaveria trinervia*. *FEBS Lett* 524(1–3): 11–14.
- Reeves RE, Menzies RA, Hsu DS (1968) The pyruvate-phosphate dikinase reaction. The fate of phosphate and the equilibrium. *J Biol Chem* 243(20): 5486–5491.
- Brown JR and Doolittle WF (1997) Archaea and the prokaryote-to-eukaryote transition. *Microbiol Mol Biol Rev* 61(4): 456–502.
- Lang BF, Gray MW, Burger G (1999) Mitochondrial genome evolution and the origin of eukaryotes. *Annu Rev Genet* 33(1999): 351–397.
- Tjaden B, Plagens A, Dörr C, Siebers B, Hensel R (2006) Phosphoenolpyruvate synthetase and pyruvate, phosphate dikinase of *Thermoproteus tenax*: key pieces in the puzzle of archaeal carbohydrate metabolism. *Mol Microbiol* 60(2): 287–298.
- Stürwalt VL, Michalowski CB, Löffelhardt W, Bohnert HJ, Bryant DA (1995) Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. *Plant Molecular Biology Reporter* 13: 327–332.
- Taub DR (2000) Climate and the U.S. distribution of C<sub>4</sub> grass subfamilies and decarboxylation variants of C<sub>4</sub> photosynthesis. *Am J Bot* 87(8): 1211–1215.
- Fravolini A, Williams DG, Thompson TL (2002) Carbon isotope discrimination and bundle sheath leakiness in three C(4) subtypes grown under variable nitrogen, water and atmospheric CO(2) supply. *J Exp Bot* 53(378): 2261–2269.
- Wang C, Guo L, Li Y, Wang Z (2012) Systematic comparison of c3 and c4 plants based on metabolic network analysis. *BMC Syst Biol* 2: S9.
- Taniguchi M, Kobe A, Kato M, Sugiyama T (1995) Aspartate aminotransferase isozymes in *Panicum miliaceum* L., an NAD-malic enzyme-type C<sub>4</sub> plant: comparison of enzymatic properties primary structures, and expression patterns. *Arch Biochem Biophys* 318(2): 295–306.
- Hittinger CT, Carroll SB (2007) Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449(7163): 677–681.
- Qiu H, Yoon HS, Bhattacharya D (2013) Algal endosymbionts as vectors of horizontal gene transfer in photosynthetic eukaryotes. *Front Plant Sci* 19(4): 366. eCollection 2013.
- Bhattacharya D, Yoon HS, Hackett JD (2004) Photosynthetic eukaryotes unite: endosymbiosis connects the dots. *BioEssays* 26: 50–60.
- Martin W, Schnarrenberger C (1997) The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. *Curr Genet* 32(1): 1–18.
- Reyes-Prieto A, Bhattacharya D (2007) Phylogeny of Calvin cycle enzymes supports Plantae monophyly. *Mol Phylogenet Evol* 45(1): 384–391.
- Ueno O (1996) Structural characterization of photosynthetic cells in an amphibious sedge, *Eleocharis vivipara*, in relation to C<sub>3</sub> and C<sub>4</sub> metabolism. *Planta* 199: 382–393.
- Tachibana M, Allen AE, Kikutani S, Endo Y, Bowler C (2011) Localization of putative carbonic anhydrases in two marine diatoms, *Phaeodactylum tricoratum* and *Thalassiosira pseudonana*. *Photosynth Res* 109: 205–221.
- Tsuji Y, Suzuki I, Shiraiwa Y (2009) Photosynthetic carbon assimilation in the coccolithophorid *Emiliania huxleyi* (Haptophyta): evidence for the predominant operation of the C<sub>3</sub> cycle and the contribution of b-carboxylases to the active anaerobic reaction. *Plant Cell Physiol* 50: 318–329.
- Falkowski PG, Raven JA (1997) *Aquatic Photosynthesis*. Blackwell Science, Malden, MA.

37. Reiskind JB, Bowes G (1991) The role of phosphoenolpyruvate carboxykinase in a marine macroalga with C<sub>4</sub>-like photosynthetic characteristics. *Proc Natl Acad Sci USA* 88: 2883–2887.
38. Li TY, Ren L, Zhou G, Liu C, Chi S, et al. (2012) A suitable method for extracting total RNA from red algae (in Chinese). *Transactions of Oceanology and Limnology* 4: 64–71.
39. Johnson MT, Carpenter EJ, Tian Z, Bruskewich R, Burris JN, et al. (2012) Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PLoS One* 7: e50226.