Still Acting Green: Continued Expression of Photosynthetic Genes in the Heterotrophic Dinoflagellate *Pfiesteria piscicida* (Peridiniales, Alveolata)

Gwang Hoon Kim¹*, Hae Jin Jeong²*, Yeong Du Yoo², Sunju Kim¹, Ji Hee Han¹, Jong Won Han¹, Giuseppe C. Zuccarello³

1 Department of Biology, Kongju National University, Kongju, Korea, 2 School of Earth and Environmental Science, Seoul National University, Seoul, Korea, 3 School of Biological Sciences, Victoria University of Wellington, Wellington, New Zealand

Abstract

The loss of photosynthetic function should lead to the cessation of expression and finally loss of photosynthetic genes in the new heterotroph. Dinoflagellates are known to have lost their photosynthetic ability several times. Dinoflagellates have also acquired photosynthesis from other organisms, either on a long-term basis or as "kleptoplastids" multiple times. The fate of photosynthetic gene expression in heterotrophs can be informative into evolution of gene expression patterns after functional loss, and the dinoflagellates ability to acquire new photosynthetic function through additional endosymbiosis. To explore this we analyzed a large-scale EST database consisting of 151,091 unique sequences (29,170 contigs, 120,921 singletons) obtained from 454 pyrosequencing of the heterotrophic dinoflagellate *Pfiesteria piscicida*. About 597 contigs from *P. piscicida* showed significant homology (E-value < e⁻³⁰) with proteins associated with plastid and photosynthetic function. Most of the genes involved in the Calvin-Benson cycle were found, genes of the light-dependent reaction were also identified. Also genes of associated pathways including the chorismate pathway and genes involved in starch metabolism were discovered. BLAST searches and phyllogenetic analysis suggest that these plastid-associated genes originated from several different photosynthetic ancestors. The Calvin-Benson cycle genes are mostly associated with genes derived from the secondary plastids of peridinin-containing dinoflagellates, while the light-harvesting genes are derived from diatoms, or diatoms that are tertiary plastids in other dinoflagellates. The continued expression of many genes involved in photosynthetic pathways indicates that the loss of transcriptional regulation may occur well after plastid loss and could explain the organism's ability to "capture" new plastids (i.e. different secondary endosymbioses or tertiary symbioses) to renew photosynthetic function.

Introduction

The genetic outcomes of plastid gain and loss have been actively investigated. Dinoflagellates together with apicomplexans [1–3] and ciliates [4,5] have drawn special attention in terms of plastid evolution. Evidence is accumulating that the ancestor of these lineages (e.g. apicomplexans) [1]. The genetic outcomes of plastid gain and loss have been actively investigated. Dinoflagellates together with apicomplexans [1–3] and ciliates [4,5] have drawn special attention in terms of plastid evolution. Evidence is accumulating that the ancestor of these lineages (e.g. apicomplexans) [1].

While the ciliates and most apicomplexans are non-photosynthetic, about 50% of dinoflagellates are photosynthetic. The majority of photosynthetic dinoflagellates contain peridinin as an accessory pigment, but the origin of the peridinin plastid is still enigmatic. Yoon et al. [11] suggested haptophyte origin of peridinin plastid but others suggested its red algal origin by secondary endosymbiotic event [12]. Some dinoflagellates replaced the peridinin-containing plastid with others either from a green alga, cryptophytes, haptophytes or diatoms via tertiary endosymbiosis events or a second secondary endosymbiotic event [13]. This unique evolutionary gain and loss and regain of plastids among dinoflagellates [14] and the transfer of genes to the nucleus have lead to plastid-derived genes, mostly genes involved in photosynthesis or other critical plastid functions, from several lineages. For example, in the fucoxanthin-containing dinoflagellate *Karlodinium veneficum* some photosynthetic genes have a peridinin-dinoflagellate origin and some are more closely related to haptophyte genes [15]. These metagenomic conclusions have recently been challenged due to undue care in phylogenetic interpretation [10,16].
In apicomplexans, photosynthetic ability has been lost, presumably over long evolutionary time scales, except for the early diverging apicomplexan *Chlororhiza* which is sister to the remaining parasitic apicomplexans [17], and many genes involved in photosynthetic function have been lost, or are no longer expressed, from both the plastid (apicoplast) genome and the host nucleus.

While photosynthesis is relatively easy to document the presence of cryptic plastids (retention of plastids that lack photosynthetic pigments and light-energy derived carbon fixation) is more difficult. Plastids in many lineages are degenerate and difficult to recognize and are considered as “cryptic” plastids [18]. The confirmation of a cryptic plastid (the apicoplast) in an extremely well studied group of parasite, the apicomplexans, is only about 20 years old [19], revealing that plastids are difficult to confirm in non-photosynthetic lineages.

The heterotrophic dinoflagellates obtain primary carbon by ingesting other organisms [20] and loss of photosynthesis has occurred multiple times in dinoflagellate evolution [14]. It appears that some of the lineages diverging early in dinoflagellate evolution are non-photosynthetic [14,21] again suggesting that they been without photosynthetic function for a long time. For example, recent studies on a heterotrophic early-diverging dinoflagellate *Oxyrrhis marina* indicate that it has a few genes for several biosynthetic pathways that are associated with plastids [22] but no genes associated with photosynthesis (e.g. the light harvesting or Calvin-Benson cycle), which may indicate an early ancestor with photosynthesis, but the loss of many unused plastid functions over a long period. The *Cryptophidium colbin* (Gonyaulacales), a non-photosynthetic but later diverging dinoflagellate, has some genes associated with photosynthesis (e.g. ribulose-1,5-bisphosphate carboxylase/oxygenase) [23]. This species and *Oxyrrhis marina* [22] clearly show that they originally derived from a photosynthetic ancestor, with the transfer of the photosynthonts to the host nucleus, and that these transferred genes may still serve a function. Clearly plastids have functions beyond light-energy capture.

Still very little is known about how and why photosynthesis is lost outright or what occurs when photosynthetic function is no longer needed [24,25]. The expression of genes may be a process that continues even after their functional utility has been lost. The selective advantage of reducing expression of these genes may be negligible and loss of function could be a stochastic process taking a long time. The continued expression of genes that have a plastid function, and plastid targeting, may aid in acquisition of alternate plastid, as has occurred often in dinoflagellates [5,26]. Heterotrophic dinoflagellates therefore may make good candidates to study these stages of gene regulation once functional constraints (e.g. photosynthesis) have been removed. With this in mind we studied the heterotrophic dinoflagellate *Pfiesteria piscicida* using a large-scale EST data set to estimate expression of plastid genes and try to account for genes in important plastid biosynthetic pathways.

*P. piscicida* is a member of the family Pfiesteriaceae in the order Peridinales [27], a group containing mainly non-photosynthetic dinoflagellates. *P. piscicida* has been studied over the last 20 years as it is involved in fish deaths. While some aspects of its biology (life cycle and toxicity) are still controversial [28,29], it is clear that *P. piscicida* is not photosynthetic and a TEM study could not find any membranous structures assignable as plastids [29]. *P. piscicida* has been implicated as being kleptoplastidic when it feeds upon cryptophyte algae as they have been shown to persist in vacuoles of starved *P. piscicida* for a week, apparently fixing carbon and accumulating starch [30].

In the present study, we identified many plastid-derived genes from both of the major photosynthetic pathways (light-dependent reaction and Calvin-Benson cycle) and other plastid-associated pathways (e.g. chorismate pathway).

**Materials and Methods**

**Culture conditions**

The strain of *P. piscicida* was originally isolated from Masan Bay (southern part of Korea) in July 2005 [31]. *P. piscicida* cells were added to 1-L polycarbonate (PC) bottles containing fresh medium. Bottles were capped, placed on a rotating wheel, incubated under an illumination of 20 μE/m²/s provided by cool-white fluorescent light on a 14:10 h light-dark cycle. Perch blood cells were collected from the live fish purchased at seafood market. The serum was removed by washing the fish blood three times using PBS buffer after mild centrifugation. The washed blood cells were provided to *P. piscicida*. As the concentration of *P. piscicida* increased, cells were transferred to new 1-L PC bottles every 2 days and 1 mL of washed Perch blood cells were provided together. The fish blood cells were checked with fluorescence microscope every time. Although bacterial mRNA data could be easily eliminated from the EST database of *P. piscicida* because they do not have a poly-A tail of eukaryotic organism, the cultures were maintained with no visible bacterial contamination. We obtained forty 1-L bottles of dense culture of *P. piscicida* for pyrosequencing in 6 months. Samples were taken from the culture whenever they were transferred to other bottles and observed with a light microscope as well as a fluorescent microscope to check for contamination of any photosynthetic algae. Whenever any auto-fluorescence of photosynthetic phytoplankton was detected, whole culture batches were discarded and a new culture started. Eventually, we conducted these experiments using cultures without any algal contamination. In addition, to exclude possible contamination by other *Pfiesteria*-like heterotrophic dinoflagellates (so called PLDs) such as *Stoeckera* spp., *Luciella* spp. and etc., PCR was performed using the DNA specific primers for detecting dinoflagellates before harvest [32].

**RNA isolation and pyrosequencing**

Total RNA from *P. piscicida* was isolated using Trizol (MRC Inc.) according to manufacturer’s protocol. Twenty liters of *P. piscicida* cells were taken after 3 days of starvation and harvested by centrifugation at 1200g. After confirming no contamination using fluorescence microscope and PCR, the pellets were immediately frozen with liquid nitrogen and stored at −80°C. Isolated RNA was quantified spectrophotometrically or using RNA gel electrophoresis. mRNA was purified using Oligotex (Qiagen) following the manufacturer’s instructions. Double-strand cDNA was synthesized using Just cDNA Double-stranded cDNA Synthesis Kit (Agilent Technologies, CA, USA) following the manufacturer’s instructions. The cDNA was then sent to GNCBio company (Daemyung, Korea) for 454 pyrosequencing. The library preparation, GS-FLX titanium sequencing, assembly and annotation of sequencing data were carried out by GNCBio. To analyze the sequence data a web-based pipeline program for EST data analysis was established (http://genebank.kongju.ac.kr).

**Identification of plastid-derived genes and bioinformatics**

Possible contamination of bacterial mRNA was removed easily from the ESTs database. Putative plastid-derived genes were identified by closest sequence similarity with an E-value <1e−30. Nucleotide and amino acid sequence homology searches and comparison were carried out using BLAST on the NCBI GenBank
database (http://blast.ncbi.nlm.nih.gov). Additional homology searches were carried out by comparing our translated EST database directly with the comprehensive chloroplast protein database of Arabidopsis thaliana [33] (Plastid protein database: http://www.plprot.ethz.ch, AT_Chloro database: http://www.grenoble.prabi.fr/at_chloro). The sequences generated in this study were deposited in GenBank under accession number SRR837773.

Phylogenetic analysis
Putative plastid-derived translated sequences of P. piscicida were aligned with the highest BLAST hit sequences plus other genes homologs from a selection of other lineages, especially stramenopiles, alveolates and plants were available. Prokaryote homologs were used as outgroup sequences. Sequences were aligned with MAFFT [34] in the Geneious software package [35]. Amino acids datasets were analyzed under the WAG + C + I model. The phylogeny of putative plastid-derived genes was inferred by maximum likelihood (ML) using RAxML 7.2.8 [36] and Bayesian analysis using MrBayes [37]. Likelihoods were estimated using the WAG protein substitution model [38]. For ML, bootstrap support was performed with 100 replicates. For Bayesian analysis, a total of 1,000,000 generations were run and sampled every 1,000 generations with burn-in of 100,000 generations. Stationarity was assessed using Tracer v1.5 [39] and a burn-in of 1000 generations was applied.

5'-RACE PCR
Total RNA was isolated as described above. First and second strand cDNA was synthesized using SMART cDNA Library construction kit (Clontech, CA, USA) according to manufacturer's protocol. Synthesized cDNA was used as a template. Specific primers were designed using the contigs sequence (Table S1). PCR was carried out in a 50 µL reaction mixture containing DNA template, 20 µmol Spliced Leader primer and 20 µmol specific primers (Table S1), 1X Taq buffer, 2.5 mM MgCl₂, and 1 unit of Taq DNA polymerase (Takara, Tokyo, Japan). PCR was performed for 35 cycles at 95°C for 20 sec, 50–55°C for 30 sec, and 72°C for 60 sec, followed by 72°C for 10 min. The PCR products were cloned into T-easy cloning vector (Promega, USA), and their sequences were determined using 3730xl DNA analyzer (Applied Biosystem).

Sub-cellular localization was predicted using CBS prediction program (http://www.cbs.dtu.dk/services).

Ethics statement
This research has been approved by Institutional Animal Care and Use Committees of Seoul National University.

Results
A total of 264 Mbp of ESTs were sequenced from P. piscicida and assembled to 151,091 unique clusters (29,170 contigs and 120,921 singletons) with an average length of 636 bp. When photosynthesis genes were searched in the databases, 8 major
Table 1. Photosynthesis genes involved in Calvin cycle and its associated pathways in *Pfiesteria piscicida*.

<table>
<thead>
<tr>
<th>Gene match</th>
<th>Contig (KNU ID)</th>
<th>No. of reads</th>
<th>Length (bp)</th>
<th>GenBank match (Acc. No.)</th>
<th>E-value (BlastX)</th>
<th>Origin</th>
<th>Phylum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calvin cycle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ribose-5-phosphate isomerase</td>
<td>320C004736</td>
<td>6</td>
<td>809</td>
<td>AAW79354</td>
<td>1.44E-96</td>
<td>Heterocapsa triquetra</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C018635</td>
<td>7</td>
<td>712</td>
<td>AAW79354</td>
<td>5.95E-108</td>
<td>Heterocapsa triquetra</td>
<td>Dinophyta</td>
</tr>
<tr>
<td>Ribulose-phosphate 3-epimerase</td>
<td>320C010448</td>
<td>10</td>
<td>829</td>
<td>XP_002907019</td>
<td>1.05E-91</td>
<td>Phytophthora infestans</td>
<td>Oomycota</td>
</tr>
<tr>
<td>Phosphoglycerate kinase</td>
<td>320C003700</td>
<td>12</td>
<td>719</td>
<td>AAW79324</td>
<td>1.88E-76</td>
<td>Heterocapsa triquetra</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C003327</td>
<td>11</td>
<td>897</td>
<td>BAE07167</td>
<td>1.55E-157</td>
<td>Karenia brevis</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C004509</td>
<td>9</td>
<td>301</td>
<td>AAW79324</td>
<td>1.56E-45</td>
<td>Heterocapsa triquetra</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C004555</td>
<td>40</td>
<td>2039</td>
<td>BAE07174</td>
<td>0</td>
<td>Heterocapsa triquetra</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C005438</td>
<td>44</td>
<td>1699</td>
<td>AAW79324</td>
<td>0</td>
<td>Heterocapsa triquetra</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C015232</td>
<td>15</td>
<td>538</td>
<td>EGZ09335</td>
<td>9.22E-58</td>
<td>Phytophthora sojae</td>
<td>Oomycota</td>
</tr>
<tr>
<td></td>
<td>320C019628</td>
<td>30</td>
<td>528</td>
<td>BAE07174</td>
<td>2.23E-77</td>
<td>Heterocapsa triquetra</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C019872</td>
<td>45</td>
<td>1233</td>
<td>XP_002771110</td>
<td>0</td>
<td>Perkins marinus</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C019994</td>
<td>48</td>
<td>1372</td>
<td>AAW79324</td>
<td>1.50E-116</td>
<td>Heterocapsa triquetra</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C029471</td>
<td>23</td>
<td>741</td>
<td>AAW79324</td>
<td>2.30E-127</td>
<td>Heterocapsa triquetra</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C029472</td>
<td>8</td>
<td>625</td>
<td>AAW79324</td>
<td>1.50E-116</td>
<td>Heterocapsa triquetra</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C029883</td>
<td>19</td>
<td>671</td>
<td>AAW79324</td>
<td>8.92E-127</td>
<td>Heterocapsa triquetra</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C030136</td>
<td>21</td>
<td>592</td>
<td>AAW79324</td>
<td>2.81E-105</td>
<td>Heterocapsa triquetra</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C019342</td>
<td>11</td>
<td>369</td>
<td>XP_002776078</td>
<td>1.06E-40</td>
<td>Perkins marinus</td>
<td>Dinophyta</td>
</tr>
<tr>
<td>Triose phosphate isomerase</td>
<td>320C005680</td>
<td>7</td>
<td>553</td>
<td>AAW79357</td>
<td>2.04E-88</td>
<td>Heterocapsa triquetra</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C008908</td>
<td>29</td>
<td>2188</td>
<td>ABP35605</td>
<td>0</td>
<td>Karlodinium micrum</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C023136</td>
<td>6</td>
<td>670</td>
<td>AAW79357</td>
<td>6.80E-68</td>
<td>Heterocapsa triquetra</td>
<td>Dinophyta</td>
</tr>
<tr>
<td>Transketolase</td>
<td>320C011913</td>
<td>13</td>
<td>1030</td>
<td>EGB05300</td>
<td>7.51E-114</td>
<td>Aureococcus anophagefferens</td>
<td>Heterokontophyta</td>
</tr>
<tr>
<td></td>
<td>320C012917</td>
<td>10</td>
<td>970</td>
<td>EGB05300</td>
<td>7.00E-107</td>
<td>Aureococcus anophagefferens</td>
<td>Heterokontophyta</td>
</tr>
<tr>
<td>Fructose-1,6-bisphosphatase</td>
<td>320C003750</td>
<td>48</td>
<td>1224</td>
<td>ACU44982</td>
<td>0</td>
<td>Pfiesteria piscicida</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C013991</td>
<td>14</td>
<td>1270</td>
<td>ZP_09081006</td>
<td>1.98E-73</td>
<td>Mycobacterium thermoautotrophic</td>
<td>Actinobacteria</td>
</tr>
<tr>
<td>Fructose-bisphosphate aldolase</td>
<td>320C018990</td>
<td>41</td>
<td>247</td>
<td>ACU44985</td>
<td>1.78E-44</td>
<td>Pfiesteria piscicida</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C022296</td>
<td>37</td>
<td>603</td>
<td>ZP_09685324</td>
<td>9.32E-80</td>
<td>Mycobacterium tuscii</td>
<td>Actinobacteria</td>
</tr>
<tr>
<td></td>
<td>320C025487</td>
<td>4</td>
<td>654</td>
<td>NP_001242086</td>
<td>1.57E-51</td>
<td>Glycine max</td>
<td>Steptophyta</td>
</tr>
<tr>
<td></td>
<td>320C026884</td>
<td>9</td>
<td>451</td>
<td>ACU44982</td>
<td>5.96E-86</td>
<td>Pfiesteria piscicida</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C029593</td>
<td>173</td>
<td>956</td>
<td>ACU44982</td>
<td>0</td>
<td>Pfiesteria piscicida</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C029818</td>
<td>9</td>
<td>428</td>
<td>ACU44982</td>
<td>8.91E-47</td>
<td>Pfiesteria piscicida</td>
<td>Dinophyta</td>
</tr>
<tr>
<td>Glycerinaldehyde-3-phosphate dehydrogenase</td>
<td>320C007095</td>
<td>23</td>
<td>442</td>
<td>ABI14256</td>
<td>1.42E-71</td>
<td>Pfiesteria piscicida</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C019945</td>
<td>31</td>
<td>1028</td>
<td>ABI14256</td>
<td>0</td>
<td>Pfiesteria piscicida</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C02033</td>
<td>34</td>
<td>839</td>
<td>ABI14256</td>
<td>4.72E-156</td>
<td>Pfiesteria piscicida</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C005475</td>
<td>54</td>
<td>1163</td>
<td>ABI14256</td>
<td>0</td>
<td>Pfiesteria piscicida</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C006386</td>
<td>9</td>
<td>379</td>
<td>ABI14256</td>
<td>4.22E-72</td>
<td>Pfiesteria piscicida</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C006750</td>
<td>17</td>
<td>1029</td>
<td>ABI14256</td>
<td>0</td>
<td>Pfiesteria piscicida</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C013130</td>
<td>5</td>
<td>558</td>
<td>AAM68968</td>
<td>7.23E-105</td>
<td>Pyrocystis fumula</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C018125</td>
<td>31</td>
<td>332</td>
<td>ABI14256</td>
<td>2.91E-70</td>
<td>Pfiesteria piscicida</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C019394</td>
<td>16</td>
<td>323</td>
<td>ABI14256</td>
<td>5.71E-66</td>
<td>Pfiesteria piscicida</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C019399</td>
<td>17</td>
<td>373</td>
<td>AAD01872</td>
<td>4.67E-61</td>
<td>Gonoeaulax polyedra</td>
<td>Dinophyta</td>
</tr>
<tr>
<td></td>
<td>320C019466</td>
<td>11</td>
<td>526</td>
<td>ABI14256</td>
<td>1.07E-99</td>
<td>Pfiesteria piscicida</td>
<td>Dinophyta</td>
</tr>
</tbody>
</table>
genes (57 contigs) involved in the Calvin-Benson cycle, out of 13 main genes usually placed in the cycle, were detected (Fig. 1). Four genes involved in chorismate pathway and starch metabolism were also detected. These Calvin-Benson cycle genes had several isoforms with BlastX hits not always to the same species (Table 1). For example, triose phosphate isomerase isoforms had hits to Dinophyta (Table 1). Fructose-1,6-bisphosphotase had most similarity to Heterokontophyta genes. This is partially supported in the phylogeny in which the isoforms form a clade with diatoms, to the exclusion of some, but not all Dinophyta (Fig. 2). Most Calvin-Benson cycle genes had either hits exclusively to Dinophyta genes, or contained a majority of hits to Dinophyta genes. The phylogeny of the two isoforms of ribose 5-phosphate isomerase was mostly poorly resolved, one isoform grouped (1.0 PP, 68% BP) with the peridinin-containing dinoflagellates (Heterocapsa and Prorocentrum) to the exclusion of other organisms (Fig. 3). While the alternate isoform showed no strong relationship but grouped weakly with a dinoflagellate containing a haptophyte plastid (Karlodinium). Transketolase grouped strongly with Dinophyta. An alternate isoform had unsupported relationships to any other member and appeared to be highly divergent.

Although *P. piscicida* does not have any visible chloroplast several genes annotated for light reaction center were also detected (Fig. 4). Most light-dependent reaction genes showed high similarity with those of photosynthetic dinoflagellates containing diatoms as endosymbionts (the so called dinotoms, e.g. *Durinskia*, Table 2) [40]. Our phylogenetic relationships clearly place the *psbA* isoform with this Heterokontophyta to the exclusion of Dinophyta (Fig. 3).

Phylogenetic analysis of other genes associated with photosynthetic organisms (hydroxymethylbilane synthase (HMBS) and ascorbate peroxidase) are either poorly resolved as far as placement of the *P. piscicida* (Fig. 2). An isoform of HMBS did not clearly group with any particular lineage (it weakly affiliated with the cryptophyte *Guillardia theta* but without any support) but was distinct from Archaeplastida sequences. In the ascorbate peroxidase phylogeny, *P. piscicida* formed a strongly supported clade (100%) with homologues from the non-photosynthetic

### Table 1. Cont.

<table>
<thead>
<tr>
<th>Gene match</th>
<th>Contig (KNU ID)</th>
<th>No. of reads</th>
<th>Length (bp)</th>
<th>GenBank match (Acc. No.)</th>
<th>E-value (BlastX)</th>
<th>Origin Phylum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chorismate pathway</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPSP synthase</td>
<td>320C004635</td>
<td>3</td>
<td>468</td>
<td>CBN78624</td>
<td>1.89E-26</td>
<td>Ectocarpus siliculosus Heterokontophyta</td>
</tr>
<tr>
<td>Chorismate synthase</td>
<td>320C003600</td>
<td>19</td>
<td>1113</td>
<td>XP_002773541</td>
<td>9.95E-138</td>
<td>Perkinsus marinus Dinophyta</td>
</tr>
<tr>
<td>Tryptophane synthase (alpha/beta chain)</td>
<td>33S011114</td>
<td>1</td>
<td>512</td>
<td>CBQ69006</td>
<td>2.55E-57</td>
<td>Sporisorium reilianum Basidiomycota</td>
</tr>
<tr>
<td>Gluconeogenesis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glucose-6-phosphate isomerase</td>
<td>320C002437</td>
<td>28</td>
<td>836</td>
<td>ABH11438</td>
<td>8.00E-167</td>
<td>Pyrocystis lumula Myzozoa</td>
</tr>
<tr>
<td>Starch and sucrose metabolism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soluble starch synthase</td>
<td>320C001989</td>
<td>14</td>
<td>1781</td>
<td>XP_004307998</td>
<td>3.14E-75</td>
<td>Fragaria vesca Steptophyta</td>
</tr>
<tr>
<td></td>
<td>320C00390</td>
<td>24</td>
<td>1506</td>
<td>EKX37680</td>
<td>1.47E-110</td>
<td>Guillardia theta Cryptophyta</td>
</tr>
<tr>
<td></td>
<td>320C018048</td>
<td>66</td>
<td>2042</td>
<td>EKX45880</td>
<td>4.44E-118</td>
<td></td>
</tr>
<tr>
<td>Total:</td>
<td>69</td>
<td></td>
<td>1736</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0068232.t001
dinoflagellate *Cryptothecodinium cohnii*, which has been revealed from molecular evidence to harbor a relict plastid [23] (Fig. 2).

When the ESTs data were compared with the comprehensive chloroplast protein database of *Arabidopsis thaliana* [33] about 544 contigs (1.86% of total contigs) from *P. piscicida* showed significant homology (E-value < e^-50) with the chloroplast proteins of *A. thaliana* (Table S2). About 23.5% (162 out of 690 proteins) of plastid-targeted proteins of *A. thaliana* were found in the *P. piscicida* EST dataset, with E-value < e^-50 suggesting that there are far more photosynthetic genes still remained and expressed in *P. piscicida*.

The presence of photosynthetic genes in *P. piscicida* genome was confirmed using 5′-RACE PCR for 15 selected genes. Sub-cellular localization of these genes were shown using CBS prediction program (Table 3). Among them five genes contained Spliced Leader, a signature sequence of *P. piscicida*, at their 5′ ends (Fig. 5).

**Discussion**

Our results show that *Pfiesteria piscicida* expresses numerous genes involved in metabolic pathways of plastids despite it not having any sub-cellular membranous structure assignable to plastids [29]. The heterogeneous origins of the plastid genes (especially the genes directly related to photosynthesis) suggest that *P. piscicida* had experienced multiple endosymbioses, both from a secondary plastid (grouping with peridinin-containing dinoflagellate lineages) and at least one tertiary endosymbiosis (grouping with diatoms that have formed endosymbioses with dinoflagellates). This mixed origin of photosynthetic genes has been reported previously for the photosynthetic dinoflagellate *Karlodinium veneficum* (as *K. micrum*) which contains genes both of secondary-endosymbiotic origin and tertiary-endosymbiotic origin, from a haptophyte [15]. Genes for photosynthesis have also been reported in heterotrophic dinoflagellates. The early branching dinoflagellate *Oxyrhizus marina* has several genes associated with plastids but no
genes directly involved in the light-reaction or the Calvin-Benson cycle [22]. This is the first report in which many (or a majority with reference to the Calvin-Benson cycle) of the genes involved in photosynthesis have been found in a heterotrophic dinoflagellate. Interestingly genes that are normally located in the plastid (e.g. \textit{psaA} and \textit{psbA}) are found in the transcriptome of \textit{P. piscicida}. Mass transfer of genes from the plastid genome to the nucleus is well documented in dinoflagellates that have a peridinin-containing plastid [41,42]. Our data would indicate that plastid gene transfer may even occur from a tertiary plastid. While all phylogenetic

Figure 3. Maximum likelihood trees (RaXML, WAG + I + I’ model) inferred from \textit{Pfiesteria piscicida} protein sequences and assorted plastid endosymbiont lineages. \textit{psbA} protein – alignment length 282 amino acids. Ribose-5-phosphate isomerase protein – alignment length 231 amino acids. Numbers above nodes indicate posterior probabilities and RAXML bootstrap percentages. * = 1.00 PP and 100% RAxML BP. Values <.50% are not shown.
doi:10.1371/journal.pone.0068232.g003

Figure 4. A diagram showing chloroplast membrane genes and light reaction center genes. The genes found in \textit{Pfiesteria piscicida} EST database are marked in red color.
doi:10.1371/journal.pone.0068232.g004
### Table 2. Light-dependent reaction genes expressed in *Pfiesteria piscicida*

<table>
<thead>
<tr>
<th>Gene match</th>
<th>Chloroplast/Plastid (pID)</th>
<th>Synthetic organism or plastid (i.e. kleptoplasty)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photosystem III protein F</td>
<td>353s051420 1 460</td>
<td>1.58E-61</td>
</tr>
<tr>
<td>Photosystem I protein F</td>
<td>353s051412 1 460</td>
<td>1.58E-61</td>
</tr>
<tr>
<td>Photosystem I core antenna</td>
<td>353s051420 1 417</td>
<td>1.58E-61</td>
</tr>
<tr>
<td>Photosystem II chlorophyll A core antenna</td>
<td>353s051420 1 417</td>
<td>1.58E-61</td>
</tr>
</tbody>
</table>

**Table:**

<table>
<thead>
<tr>
<th>Gene match</th>
<th>Chloroplast/Plastid (pID)</th>
<th>Synthetic organism or plastid (i.e. kleptoplasty)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photosystem III protein F</td>
<td>353s051420 1 460</td>
<td>1.58E-61</td>
</tr>
<tr>
<td>Photosystem I protein F</td>
<td>353s051412 1 460</td>
<td>1.58E-61</td>
</tr>
<tr>
<td>Photosystem I core antenna</td>
<td>353s051420 1 417</td>
<td>1.58E-61</td>
</tr>
<tr>
<td>Photosystem II chlorophyll A core antenna</td>
<td>353s051420 1 417</td>
<td>1.58E-61</td>
</tr>
</tbody>
</table>

The presence of a Spliced Leader, a signature sequence of plastids, indicates that *P. piscicida* may have cryptic plastid still not acquired from HGT from bacteria, e.g. *EF-Tu* [51]. The number becomes 1.86% of total ESTs. This would be a resource costly activity in *P. piscicida* if it had no utility. We suggest several possibilities: One that some photosynthetic genes may be used in some other cellular process. Secondly, the continued expression of plastid genes in *P. piscicida* is non-functional and selection has not removed expression of these genes and this is compensated for by aggressive feeding by *P. piscicida*. Thirdly, *P. piscicida* may get some benefit in being able to maintain these genes to prolong functioning of captured photosynthetic organism or plastid (i.e. kleptoplasty).

Many of the genes matching the *Arabidopsis* proteome database may function in other cellular compartments. Aromatic amino acids produced in the chorismate pathway are produced in the plastids of higher plants but their cellular location in other lineages may be elsewhere [49]. It is also suggested that photosynthetic genes may perform limited carbon fixation [50] while other genes clearly have homology to other non-plastid genes that may be acquired from HGT from bacteria, e.g. *EF-Tu* [51].

It is also possible that *P. piscicida* may have cryptic plastid still not found. Non-photosynthetic plastids are very difficult to identify. A good example is provided by stramenopiles (or heterokonts) belonging to the clade Dictyochophyceae. Based solely on ultrastructural data, it was postulated that they lost their secondary plastids [7]. However, later studies demonstrated the existence of non-photosynthetic plastids with four envelope membranes and an ER-like outermost membrane connected with the nuclear envelope in these stramenopiles [52]. Recently, Fernández-Rubledo et al. [53] reported problems with the identification of a non-photosynthetic plastid in the well-investigated parasite *Perkinus marinus*. Question still remains in either case; why are these photosynthetic genes still expressed if not for its own functional plastids?

**P. piscicida** is an aggressive predator that could even control other algal blooms [31] and predators on fish [54]. *P. piscicida* feeds with a peduncle (i.e. feeding tube) extracting cell contents from prey into food vacuoles and thus plastids of algal prey could be transferred into the predator's protoplasm without damage [31].
There is a possibility that these acquired plastids work as “kleptoplasts” inside the predator cell. Lewitus et al. [30] reported that plastid of ingested cryptophytes persisted in vacuoles of \textit{P. piscicida} for a week and were apparently fixing small amount of carbon and accumulating starch grains. Additionally, Feinstein et al. [55] showed that the growth rate of \textit{P. piscicida} fed on the cryptophyte \textit{Rhodomonas} sp. at saturating light levels was almost twice as in the darkness. Jeong et al. [31] reported that the gross growth efficiency of \textit{P. piscicida} fed on \textit{Rhodomonas salina} exceeded 100%, which suggests the possibility of kleptoplastidy. However, kleptoplastidic photosynthesis alone is not enough for the survival of \textit{P. piscicida} as the number of \textit{P. piscicida} cells decreased as soon as

Figure 5. Multiple alignments of 5’ end sequences of GAPDH and fructose-bisphosphate aldolase showing the location of Spliced Leader sequence and specific primer (black lines on the top). 5’ UTR and open reading frame (ORF) were shown with red arrows. doi:10.1371/journal.pone.0068232.g005

Expression of Photosynthetic Genes in \textit{Pfiesteria}
prey cells were removed and was even faster in the light than in the dark [5,31,55]. Recently, Johnson [5] suggested that the enhanced prey cells were removed and was even faster in the light than in the dinoflagellate P. piscicida than any other group of eukaryotic organisms [56]. Considering its wide spectrum of prey, it is hard to believe that the numerous plastid genes of P. piscicida are used only by its kleptoplastids. Actually, it was initially suggested that unusual dinoflagellate plastids (e.g. the fucoxanthin plastid of Karenia and Karlodinium) adapted the targeting machinery and hundreds of nucleus-residing plastid genes of the ancestral peridinin plastid. However, the genomic studies by Yoon et al. [11] and Patron et al. [15] questioned the hypothesis. They found that the fucoxanthin plastid uses mainly genes derived from its haptophyte ancestor.

So why does P. piscicida cell express so many photosynthetic genes? It is possible that there is reduced selection on the removal and reduction on their expression in these genes. A consequence of this is that heterotrophic dinoflagellates may more easily acquire and maintain symbiotic plastids. The serial replacement of one plastid and another has been seen in the dinotomes (diatom-containing dinoflagellates) [45] and is more prevalent in dinoflagellates than any other group of eukaryotic organisms [56].

Our comprehensive EST data set of the heterotrophic dinoflagellate P. piscicida indicated that this organism still expresses plastid throughout their history.

### Supporting Information

**Table S1** Used primer sets for 5’RACE PCR. (PDF)

**Table S2** ESTs database of *Pfiesteria piscicida* contigs associated with the plastid through the chloroplast protein database of *Arabidopsis thaliana* and their BLAST analysis. (PDF)

### Acknowledgments

We thank Dr. Tatyana A. Klochkova for the help during manuscript preparation.

### Author Contributions

Conceived and designed the experiments: GHK HJJ. Performed the experiments: YDY JWH JHH. Analyzed the data: GHK JWH GCZ SK. Contributed reagents/materials/analysis tools: GHK HJJ. Wrote the paper: GHK HJJ GCZ.

### References


### Table 3. List of contigs used for 5’-RACE and prediction of sub-cellular location.

<table>
<thead>
<tr>
<th>Gene match</th>
<th>Contig (KNU ID)</th>
<th>cTP</th>
<th>mTP</th>
<th>SP</th>
<th>Other</th>
<th>Loc</th>
<th>RC</th>
<th>TPlen</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chloroplast FtsH protease</td>
<td>320C008843</td>
<td>0.044</td>
<td>0.251</td>
<td>0.168</td>
<td>0.74</td>
<td>-</td>
<td>3</td>
<td>-</td>
<td>ESTs</td>
</tr>
<tr>
<td>Alpha-amylase</td>
<td>320C010144</td>
<td>0.071</td>
<td>0.146</td>
<td>0.109</td>
<td>0.825</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>ESTs</td>
</tr>
<tr>
<td>Chaperonin</td>
<td>320C019442</td>
<td>0.083</td>
<td>0.298</td>
<td>0.089</td>
<td>0.613</td>
<td>-</td>
<td>4</td>
<td>-</td>
<td>ESTs</td>
</tr>
<tr>
<td>Fructose-bisphosphate aldolase</td>
<td>320C018990</td>
<td>0.035</td>
<td>0.033</td>
<td>0.062</td>
<td>0.955</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>SL</td>
</tr>
<tr>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
<td>320C019568</td>
<td>0.031</td>
<td>0.616</td>
<td>0.102</td>
<td>0.293</td>
<td>M</td>
<td>4</td>
<td>55</td>
<td>SL</td>
</tr>
<tr>
<td></td>
<td>320C019569</td>
<td>0.031</td>
<td>0.672</td>
<td>0.05</td>
<td>0.38</td>
<td>M</td>
<td>4</td>
<td>55</td>
<td>SL</td>
</tr>
<tr>
<td></td>
<td>320C019675</td>
<td>0.022</td>
<td>0.62</td>
<td>0.093</td>
<td>0.325</td>
<td>M</td>
<td>4</td>
<td>55</td>
<td>SL</td>
</tr>
<tr>
<td></td>
<td>320C028250</td>
<td>0.02</td>
<td>0.761</td>
<td>0.08</td>
<td>0.22</td>
<td>M</td>
<td>3</td>
<td>55</td>
<td>SL</td>
</tr>
<tr>
<td></td>
<td>320C024424</td>
<td>0.014</td>
<td>0.637</td>
<td>0.133</td>
<td>0.297</td>
<td>M</td>
<td>4</td>
<td>55</td>
<td>ESTs</td>
</tr>
<tr>
<td>Peptidyl-prolyl isomerase</td>
<td>320C016005</td>
<td>0.03</td>
<td>0.263</td>
<td>0.813</td>
<td>0.015</td>
<td>S</td>
<td>3</td>
<td>18</td>
<td>ESTs</td>
</tr>
<tr>
<td></td>
<td>320C018648</td>
<td>0.113</td>
<td>0.205</td>
<td>0.04</td>
<td>0.813</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>ESTs</td>
</tr>
<tr>
<td>Phosphoglycerate kinase</td>
<td>320C005438</td>
<td>0.056</td>
<td>0.128</td>
<td>0.05</td>
<td>0.775</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>ESTs</td>
</tr>
<tr>
<td>Splicing factor Prp8</td>
<td>320C008393</td>
<td>0.047</td>
<td>0.125</td>
<td>0.142</td>
<td>0.904</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>ESTs</td>
</tr>
<tr>
<td>Pyruvate kinase</td>
<td>320C021827</td>
<td>0.098</td>
<td>0.101</td>
<td>0.149</td>
<td>0.795</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>ESTs</td>
</tr>
<tr>
<td>Triosephosphate isomerase</td>
<td>320C001323</td>
<td>0.334</td>
<td>0.207</td>
<td>0.018</td>
<td>0.4</td>
<td>-</td>
<td>5</td>
<td>-</td>
<td>ESTs</td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td></td>
<td>0.014</td>
<td>0.637</td>
<td>0.133</td>
<td>0.297</td>
<td>M</td>
<td>4</td>
<td>55</td>
<td>15</td>
</tr>
</tbody>
</table>

SL: PCR with Spliced Leader sequence. CBS prediction program was used for analyzing sub-cellular location of each gene. doi:10.1371/journal.pone.0068232.t003