Auxenochlorella protothecoides and Prototheca wickerhamii plastid genome sequences give insight into the origins of non-photosynthetic algae

Dong Yan1,*,†, Yun Wang2,*, Tatsuya Murakami1, Yue Shen2, Jianhui Gong2, Huifeng Jiang3, David R. Smith4, Jean-Francois Pombert5, Junbiao Dai1 & Qingyu Wu1

The forfeiting of photosynthetic capabilities has occurred independently many times throughout eukaryotic evolution. But almost all non-photosynthetic plants and algae still retain a colorless plastid and an associated genome, which performs fundamental processes apart from photosynthesis. Unfortunately, little is known about the forces leading to photosynthetic loss; this is largely because there is a lack of data from transitional species. Here, we compare the plastid genomes of two “transitional” green algae: the photosynthetic, mixotrophic Auxenochlorella protothecoides and the non-photosynthetic, obligate heterotroph Prototheca wickerhamii.

Remarkably, the plastid genome of A. protothecoides is only slightly larger than that of P. wickerhamii, making it among the smallest plastid genomes yet observed from photosynthetic green algae. Even more surprising, both algae have almost identical plastid genomic architectures and gene compositions (with the exception of genes involved in photosynthesis), implying that they are closely related. This close relationship was further supported by phylogenetic and substitution rate analyses, which suggest that the lineages giving rise to A. protothecoides and P. wickerhamii diverged from one another around six million years ago.

There is a diversity of feeding strategies across the tree of life. Photoautotrophs, for instance, produce organic materials through photosynthesis and, thus, do not require exogenous organic matter. Heterotrophs, alternatively, survive on organic components from the environment. In eukaryotes, photosynthesis occurs in the chloroplast, which evolved circa 1.5 billion years ago through the endosymbiosis of a cyanobacterium by a unicellular, heterotrophic protist1,2. Despite the obvious benefits of photosynthesis, many eukaryotes have forfeited their photosynthetic capabilities, including various parasitic land plants and heterotrophic algae3–5. With few exceptions6, non-photosynthetic plants and algae still contain a colorless chloroplast (plastid) and a highly reduced plastid genome, both of which continue to carry out essential processes, apart from photosynthesis3,5,7,8–10. Among the leading models for understanding...
the evolutionary loss of photosynthesis are green algae, including the colorless genera Prototheca, Helicosporidium, Polytoma, and Polytonella.

Currently, the only complete plastid genome sequence available from non-photosynthetic green algae is that of the trebouxiophyte Helicosporidium sp. ATCC 50920, which is a parasite of various invertebrates. The Helicosporidium plastid DNA (ptDNA) is highly reduced (<40 kb), having lost various genes related to photosynthesis, and is similar in structure and content to the plastid genomes of apicomplexan parasites, such as Plasmodium falciparum. The closest known relatives of Helicosporidium are from the non-photosynthetic trebouxiophyte genus Prototheca, which is comprised of ubiquitous opportunistic animal pathogens, some of which can infect humans. Partial ptDNA sequence data from Prototheca wickerhamii suggest that its plastid genome is in a "transitional stage" between Helicosporidium and various photosynthetic trebouxiophytes. The closest known photosynthetic relative of Prototheca species is Auxenochlorella protothecoides, a free-living green alga that can use sugars as carbon sources for heterotrophic growth. When A. protothecoides cells are switched to heterotrophic conditions, their plastids can degenerate, resulting in the suppression and eventual elimination of photosynthesis. Remarkably, this process is reversible, depending on the conditions, and suggests that A. protothecoides could provide insights into the loss of photosynthesis.

Here, in the hopes of better understanding the shift from a photoautotrophic to heterotrophic lifestyle, we report and compare the plastid genome sequences of A. protothecoides and P. wickerhamii. Both genomes show a surprising amount of similarities, including severe ptDNA contraction and similar gene orders and gene contents, photosynthesis-related genes notwithstanding. Our phylogenetic inferences and other genomic analyses confirmed that A. protothecoides and P. wickerhamii are indeed closely related, with a recent divergence time of about six million years. Together, our results provide interesting clues about the loss of photosynthesis and the evolution of obligate heterotrophy within green algae.

Results
The A. protothecoides and P. wickerhamii plastid genomes are paragons of compactness. The A. protothecoides plastid genome is an 84.6 kb, AT-rich (69.2%), circular-mapping molecule (Fig. 1; Supplementary Table S1). It has a compact architecture (24.57% non-coding), with no inverted repeats or introns. The P. wickerhamii ptDNA architecture mirrors that of A. protothecoides: it is small (55.6 kb), circular, AT-rich (68.8%), and compact (28.8% non-coding) (Fig. 1; Supplementary Table S2). These two genomes are among the smallest and most reduced ptDNAs observed from the Trebouxiophyceae (Table 1) and green algae as a whole (Supplementary Table S3). The genomic compaction of the A. protothecoides and P. wickerhamii ptDNAs largely results from being no introns and relatively little intergenic DNA (Table 1). Moreover, genes essential for photosynthesis have been lost in P. wickerhamii (discussed below). Further contributing to the ptDNA streamlining in A. protothecoides and P. wickerhamii is the lack of plastid inverted repeat elements. The absence of these elements, however, is a reoccurring theme throughout the Trebouxiophyceae (Table 1).

The types of plastid genome reduction observed in A. protothecoides and P. wickerhamii are not uncommon for green algae, especially non-photosynthetic species. In fact Helicosporidium sp. has one of the smallest ptDNAs ever observed. Although only about 2/3 the size of that of A. protothecoides, the P. wickerhamii ptDNA is still larger and more expanded than that of Helicosporidium sp. (Reference and Table 1).

A. protothecoides and P. wickerhamii have similar plastid gene contents. The A. protothecoides ptDNA encodes 76 proteins, 3 rRNAs, and 30 tRNAs, which is among the lowest plastid gene contents currently found in green algae (Table 1; Supplementary Table S3). Not surprisingly, given its non-photosynthetic existence, the P. wickerhamii plastid genome encodes even fewer gene products than A. protothecoides—40 proteins, 3 rRNAs, and 27 tRNAs. All of the genes in the ptDNA of P. wickerhamii are also present in that of A. protothecoides (Supplementary Tables S1 and S2). In both A. protothecoides and P. wickerhamii, most of the ptDNA genes have the same transcriptional polarity, and, more importantly, the gene orders are highly conserved between these two algae (Fig. 1). Such a high degree of similarity in plastid gene arrangement is rarely observed between photosynthetic and non-photosynthetic species. A. protothecoides and P. wickerhamii show fewer regions of plastid gene collinearity with other trebouxiophytes, such as C. variabilis and Helicosporidium sp., as they do with each other (Fig. 2). Pairwise plastid-gene-order comparisons using a broader sampling of green algae (Supplementary Figure S1 and S2) further supports the hypothesis that the ptDNA synten between A. protothecoides and P. wickerhamii is among the highest yet observed within the Chlorophyta, when comparing species from distinct lineages.

The presence and absence of photosynthesis-related genes in the A. protothecoides and P. wickerhamii ptDNAs. A. protothecoides and P. wickerhamii have very different modes of energy production—the former is photosynthetic whereas the latter is an obligate heterotroph. Therefore, it is unexpected that various phylogenetic analyses showed that P. wickerhamii is more closely related to A. protothecoides than to Helicosporidium spp., implying that the loss of photosynthesis has occurred at least twice in the Chlorophytes; once in the lineage giving rise to Helicosporidium and once within that giving rise to Prototheca.
To gain more insight into the evolutionary relationships among *A. protothecoides*, *P. wickerhamii*, and other algae, we performed a Maximum-likelihood phylogenetic analysis, using peptide sequences from 12 single-copy plastid-encoded proteins from 26 species from throughout the Chlorophyta (Fig. 3A). The resulting tree placed *A. protothecoides*, *P. wickerhamii* and *Helicosporidium* sp. together within a clade adjacent to the one containing *Chlorella* sp. Arm0029B, *C. variabilis*, *C. vulgaris*, *P. kessleri*, *Chlorella sorokiniana*, consistent with the fact that all of these algae belong to Chlorellaceae. Moreover, *A. protothecoides* appears to be most closely (bootstrap support 100%) related to *P. wickerhamii*. The two algae are separated by relative short branch lengths, suggesting that they diverged from one another recently in evolutionary history. These results are consistent well with the phylogenies based on 18s rRNA (Supplementary Figure S3) and previous phylogenetic analyses using rRNA genes. In addition, a previous mitochondrial phylogenetic analysis placed *P. wickerhamii* and *Helicosporidium* sp. in the same clade. When we included *Helicosporidium* sp. in the plastid phylogeny we found that it is rather basal to the *A. protothecoides* and *P. wickerhamii* clade, suggesting that it is a close relative of *P. wickerhamii* and *A. protothecoides*.

The rate of synonymous substitution (Ks) can be used to estimate the divergent time among species. The average Ks of plastid genes between *A. protothecoides* and *P. wickerhamii* is 0.816 (Fig. 3B). If we assume that the nuclear mutation rate in unicellular green alga is $3.23 \times 10^{-10}$ substitution per generation, then the number of generation that occurred since their divergence between *A. protothecoides* and...
*P. wickerhamii* is about $2.52 \times 10^9$. When considering that single-celled algae typically have a short generation time and that they typically have similar mutation rates in their plastid and nuclear genomes 27–30, then the predicted time of divergent time for *A. protothecoides* and *P. wickerhamii* should be between six to twenty million years. Further supporting the hypothesis that *A. protothecoides* and *P. wickerhamii* are closely related is the fact that the levels of synonymous substitution in the ptDNA are not saturated ($<1$) and are in fact similar to those observed between other closely related algal strains or species 27–29.

Furthermore, we calculated the similarity of each gene, as well as the 5′ UTR (50 bp upstream of ATG) between *P. wickerhamii* ptDNA and its 17 relatives. We found that the overall similarities between the plastid genomes of *P. wickerhamii* and *A. protothecoides* are highest (Supplementary Figure S4) in all the comparisons. Among them, the tRNA genes are more conserved than other coding genes, while the upstream of tRNA genes is more diverged than the coding genes.

We also investigated the various plastid genomic changes that occurred in the *Prototheca* lineage following the loss of photosynthesis. Among 109 genes in the ptDNA of *A. protothecoides* ptDNA, 70 are also present in that of *P. wickerhamii*, meaning 39 genes were lost from the lineage of *P. wickerhamii* following its divergence from that of *A. protothecoides* (Supplementary Table S1 and S2). The majority

| Species                        | Size (bp) | %ncDNA (including introns) | Mean intergenic distance (bp) | Protein | rRNA | tRNA | G+C (%) | Inverted repeats | Accession |
|--------------------------------|-----------|-----------------------------|-------------------------------|---------|------|------|---------|------------------|-----------|
| Helicosporidium sp.            | 37,454    | 22.2                        | 319                           | 26      | 3    | 25   | 26.9    |                  | NC_008100 |
| *Prototheca wickerhamii*       | 55,636    | 28.8                        | 400                           | 40      | 3    | 27   | 31.2    |                  | KJ001761  |
| *Auxenochlorella protothecoides* | 84,580    | 24.6                        | 273                           | 76      | 3    | 30   | 30.8    |                  | KC843975  |
| *Chlorella sorokiniana*        | 109,811   | 43.0                        | 629                           | 75      | 3    | 31   | 34.0    |                  | NC_023835 |
| *Chlorella* sp. ArM0029B       | 119,989   | 46.2                        | 692                           | 80      | 3    | 32   | 33.9    |                  | KF554427  |
| *Parachlorella kessleri*       | 123,994   | 45.3                        | 668                           | 84      | 6    | 36   | 30.0    |                  | NC_012978 |
| *Chlorella variabilis*         | 124,579   | 49.5                        | 770                           | 80      | 3    | 32   | 33.9    |                  | NC_015359 |
| *Trebouxiophyceae* sp. MX-A201  | 149,707   | 52.2                        | 977                           | 80      | 3    | 32   | 57.7    |                  | NC_018569 |
| *Chlorella vulgaris*           | 150,613   | 43.3                        | 375                           | 174     | 3    | 33   | 31.6    |                  | NC_001865 |
| *Coccomyxa* sp. C-169          | 175,731   | 60.0                        | 1318                          | 80      | 3    | 32   | 50.7    |                  | NC_015084 |
| *Leptosira terrestris*         | 195,081   | 56.8                        | 1259                          | 88      | 3    | 28   | 27.3    |                  | NC_009681 |

Table 1. General features of known ptDNAs in Trebouxiophytes. aConserved genes, unique ORFs and intronic ORFs were counted as coding sequences. bncDNA; non-coding DNA. c,dGenes within inverted repeats were counted twice.

**Figure 2.** Gene order comparisons between trebouxiophyte plastid genomes. Genomes are drawn to scale. Genes are represented by filled boxes; photosynthetic, ATP synthase and rRNA-encoding genes are indicated by green, blue and red boxes, respectively. Identical genes between the genomes are connected by straight lines.
P. wickerhamii' cating a more complex metabolism in Helicosporidium in the plastid of sp.36.

ccsA and ptDNA, as are , which encode a plastid envelope membrane protein 34 and cytochrome strand bias—almost all genes are transcribed in one direction (Fig. 1).

A. protothecoides between the ptDNAs of these two algae is the loss of photosynthesis-related genes in the latter. Again, our non-photosynthetic relative P. wickerhamii photosynthetic algae, particularly those from the Trebouxiophyceae and Chlorophyceae. Moreover, the ptDNA mutation rates within the Chlorophyta.

Figure 3. Phylogenetic niche of A. protothecoides as inferred from plastid gene sequences and average plastid mutation rates within the Chlorophyta. (A) The best Maximum Likelihood phylogenetic tree computed with PHYML under the LG + G8 + I model of amino acid substitution is shown here, with prasinophytes used as outgroups. Bootstrap support for each clade is indicated on the corresponding node. (B) Average synonymous mutation rate (Ks) among A. protothecoides and nine related species. (C) Average non-synonymous mutation rate (Ka) among A. protothecoides and nine related species.

of the missing genes are related to photosynthesis. For instance, in the A. protothecoides plastid genome, 31 genes are involved in photosynthesis, and these genes have been lost from P. wickerhamii (and Helicosporidium sp.) (Supplementary Table S4 and S5). Three other genes (ycf3, ycf4 and ycf12), with ambiguous functions but likely connected to photosynthesis), are also absent from the P. wickerhamii ptDNA, as are ccoA and ccsA, which encode a plastid envelope membrane protein34 and cytochrome c-type biogenesis protein35, respectively. Finally, three tRNAs, (trnL(GAG)), trnS(GGA) and trnT(GGU) have also been eliminated from the P. wickerhamii plastid. Significant gene content differences were also observed between P. wickerhamii and Helicosporidium sp. (Fig. 2, Supplementary Table S3 and S4), indicating a more complex metabolism in P. wickerhamii's plastid compared with that predicted to be located in the plastid of Helicosporidium sp.36.

To investigate the molecular mechanisms of plastid gene loss from P. wickerhamii, we analyzed the junctions flanking deleted genes and gene clusters relative to A. protothecoides. In total, 17 regions containing missing genes were identified (labeled breakpoint BP 1 to BP17), ranging from <0.5 kb to >7.5 kb (Supplementary Table S6 and Supplementary Figure S5). Most of the junctions show no sequence similarity, but they do tend to be very AT rich (average >85%), and 13 of the 17 BPs are adjacent to a tRNA gene.

We compared in detail the ptDNAs of P. wickerhamii and Helicosporidium sp., and found that although both genomes are reduced, the overall architecture is quite different. In the Helicosporidium ptDNA, the rRNA operon is split, and the coding regions display a symmetric strand bias. In contrast, the P. wickerhamii plastid has a “typical” intact rRNA operon and the coding sequences have an asymmetric strand bias—almost all genes are transcribed in one direction (Fig. 1).

Discussion
The ptDNAs of non-photosynthetic species are generally <80 kb, making them much smaller than those of most photosynthetic plants and algae, which are about 100–200 kb, with some notable exceptions. In this study, we showed that the A. protothecoides ptDNA is among the smallest observed from photosynthetic algae, particularly those from the Trebouxiophyceae and Chlorophyceae. Moreover, the ptDNA architecture and sequence of A. protothecoides is similar in many ways to that of its close non-photosynthetic relative P. wickerhamii (Figs 1 and 2, Table 1). Indeed, the only significant difference between the ptDNAs of these two algae is the loss of photosynthesis-related genes in the latter. Again, our phylogenetic analyses are consistent with earlier studies showing that A. protothecoides is more closely
related to *P. wickerhamii* than it is to other species within the Chlorellales23, and ultimately support the independent loss of photosynthesis in the *P. wickerhamii* and *Helicosporidium* lineages.

*A. protothecoides* is a free-living mixotrophic alga and, thus, can survive heterotrophically, provided it has organic carbon sources—a feature that has been exploited to produce large amount of biomass in a short period of time20. *P. wickerhamii*, on the other hand, is a widely distributed, obligate heterotrophic alga, that can act as an opportunistic pathogens to infect humans39 and animals40. Although these two algae could not have more drastically different lifestyles41,42, we found that both share a surprisingly high level of ptDNA sequence identity, which explains that *A. protothecoides* and *P. wickerhamii* are more closely related to one another than they are to other members of the *Chlorella* and *Prototheca* genera, and validates previous results using phylogenetics43.

The major difference between the *A. protothecoides* and *P. wickerhamii* ptDNA is the presence of photosynthesis-related genes. Given the high similarity in gene content and genomic structure between *A. protothecoides* and *P. wickerhamii*, it is almost certain that they share a recent common photosynthetic ancestor, perhaps as recently as ~6 million years ago. At some point after the two lineages diverged, photosynthesis was lost in the *P. wickerhamii* lineage, resulting in the wholesale loss of photosynthetic genes, whereas in the *A. protothecoides* lineage photosynthesis has been maintained. Ultimately, the close relationship between the two algae suggests that they could represent an excellent duo for studying the evolutionary loss of photosynthesis.

**Methods**

**Strains and Cultivation Conditions.** The *A. protothecoides* strain, the medium and cultivation methods used in this study have been described previously43. The *P. wickerhamii* strain (SAG 263-11) was purchased from the Culture Collection of Algae at the University of Göttingen, Germany (SAG) and cultured in malt peptone medium containing 10 g/L malt extract (Sigma) and 2.5 g/L protease-peptone (Sigma).

**Organelle genome sequencing and annotation.** The complete *P. wickerhamii* plastid genome was obtained using Sanger chemistry and assembled with the SeqMan program from the DNASTAR Lasergene package. The sequencing primers were designed based on the partially sequenced genome (GenBank: AJ245645.1 and AJ236874.144). To close the gaps, a set of primers was designed based on the conserved genes found in both *A. protothecoides* and *Helicosporidium* sp. The *A. protothecoides* plastid genome was obtained as part of the whole genome-sequencing project18 (GenBank: APJ01001039.1 and APJ001001000.1). All primer sequences are available upon request.

Both the *A. protothecoides* and *P. wickerhamii* plastid genomes were annotated using the same methods. The gene sets were originally annotated by Dogma (Dual Organellar GenoMe Annotator)44 and then curated manually. Protein-coding genes and non-coding RNAs with percent identity lower than 25 and 80 respectively were cut off (E-value 1e-5). Protein-coding genes were examined by Blastx searches against the NCBI non-redundant protein (nr) database and their boundaries adjusted manually whereas tRNA-coding genes were identified by tRNAscan-SE 1.2345 using organelle parameters. The complete sequences of the *A. protothecoides* and *P. wickerhamii* plastid genomes have been deposited into GenBank under the access numbers KC843975 (*A. protothecoides*) and KJ001761 (*P. wickerhamii*).

**Pairwise alignment and comparison.** The pairwise alignments were performed by Blastn (E-value \(-1\) to 0.005) in tbl2asn 2.2.2346 with ‘-1’ as a penalty for a nucleotide mismatch. Each hit was used to estimate the average identity for all alignments or in 100bp-long windows. In gene order comparisons, protein-coding gene and rRNA gene were considered and identified by gene name. *C. variabilis* NC64A and *Helicosporidium* plastid genomes were adjusted for comparison correspondingly. *C. variabilis* NC64A was reversed and started with *atpI*, while *Helicosporidium* sp. started with *rpl12*.

For mutation rate analyses, we did pairwise alignments using GeneWise for each orthologous gene47. The software YN00 in the package PAML 4.8a was used to estimate the synonymous and non-synonymous mutation rate (KS&KA)48. Li’s model49 was used for estimating DNA divergence time (the generation times using for calculation were 24h = 0.00274y for heterotroph and 72h = 0.008219y for autotroph).

**Phylogenetic tree construction.** We used the TreeFam methodology50 to define orthologous genes among taxa as follows: the all-versus-all peptide sequence alignments of protein-coding genes were performed using blastp 2.2.23 (no SEG query sequence filtering, E-value threshold of 1e-7), the blast alignments were combined and filtered using SOLAR 0.9.6 (pairwise gene alignment rate of 0.24), and the clustering was performed with Hcluster�_sg 0.5.0 (minimum edge weight 10, minimum edge density 0.34).

A total of 12 single-copy TreeFam protein-encoding gene clusters (*tufA*, *rpoC1*, *rps7*, *rps8*, *rps11*, *rps12*, *rpl19*, *rpl20*, *rpl5*, *rpl14*, *rpl16*, *rpl20*) from 26 Chlorophyta taxa were defined. The amino acid sequences were aligned with MUSCLE 3.751 and the ambiguously aligned regions were filtered out with BMGE 1.152 using the default parameters. Maximum-Likelihood phylogenetic inferences were run with PhyML 3.053 under the LG + G + I model of amino acid substitution (8 gamma categories), selected with the Model Selection (ML) module from MEGA 6.0554. A total of 100 non-parametric bootstrap replicates were performed, as implemented in PHYML.
43. Yan, D., Lu, Y., Chen, Y. F. & Wu, Q. Waste molasses alone displaces glucose-based medium for microalgae fermentation towards cost-saving biodiesel production. *Bioresour Technol* **102**, 6487–6493 (2011).
44. Wyman, S. K., Jansen, R. K. & Boore, J. L. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **20**, 3252–3255 (2004).
45. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955–964 (1997).
46. Tatusova, T. A. & Madden, T. L. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbial Lett* **174**, 247–250 (1999).
47. Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res* **14**, 988–995 (2004).
48. Yang, Z. & Nielsen, R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol* **17**, 32–43 (2000).
49. Li, W. *Molecular Evolution* (MA: Sinauer Associates, 1997).
50. Li, H. *et al.* TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res* **34**, D572–580 (2006).
51. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).
52. Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* **10**, 210 (2010).
53. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307–321 (2010).
54. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**, 2725–2729 (2013).

**Acknowledgements**
This work was supported by MOST project 2011CB808804 and 2014AA02200, NSFC project 31370282 and 41030210 to Q.W. and by Tsinghua University Initiative Scientific Research Program 2011Z02296 and 2012Z08128 to J.D.

**Author Contributions**
Y.D. conducted experiments and wrote the paper. Y.W. assembled and annotated the genome, conducted analyses, and wrote part of the paper. TM prepared the *P. wickerhamii* genome samples. Y.S. and J.G. performed bioinformatics analyses. H.J. analyzed the mutation rates. D.R.S. and J.F.P. performed analyses and helped draft the manuscript. J.D. and Q.W. conceived the study, conducted analysis, and wrote the paper. All authors read and approved the final manuscript.

**Additional Information**
Supplementary information accompanies this paper at http://www.nature.com/srep

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Yan, D. *et al.* Auxenochlorella protothecoides and *Prototheca wickerhamii* plastid genome sequences give insight into the origins of non-photosynthetic algae. *Sci. Rep.* **5**, 14465; doi: 10.1038/srep14465 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/
Erratum: *Auxenochlorella protothecoides* and *Prototheca wickerhamii* plastid genome sequences give insight into the origins of non-photosynthetic algae

Dong Yan, Yun Wang, Tatsuya Murakami, Yue Shen, Jianhui Gong, Huifeng Jiang, David R. Smith, Jean-Francois Pombert, Junbiao Dai & Qingyu Wu

*Scientific Reports* 5:14465; doi: 10.1038/srep14465; published online 25 September 2015; updated on 14 December 2015

The Acknowledgements section in this Article is incomplete.

“This work was supported by MOST project 2011CB808804 and 2014AA02200, NSFC project 31370282 and 41030210 to Q.W. and by Tsinghua University Initiative Scientific Research Program 2011Z02296 and 2012Z08128 to J.D.”

should read:

“This work was supported by MOST project 2011CB808804 and 2014AA02200, NSFC project 31370282 and 41030210 to Q.W. and by Tsinghua University Initiative Scientific Research Program 2011Z02296 and 2012Z08128 to J.D., and partially by MOST project 2012AA02A701.

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/