The Complete Plastid Genomes of the Two ‘Dinotoms’
Durinskia baltica and Kryptoperidinium foliaceum

Behzad Imanian*, Jean-François Pombert*, Patrick J. Keeling*
Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada

Abstract

Background: In one small group of dinoflagellates, photosynthesis is carried out by a tertiary endosymbiont derived from a diatom, giving rise to a complex cell that we collectively refer to as a ‘dinotom’. The endosymbiont is separated from its host by a single membrane and retains plastids, mitochondria, a large nucleus, and many other eukaryotic organelles and structures, a level of complexity suggesting an early stage of integration. Although the evolution of these endosymbionts has attracted considerable interest, the plastid genome has not been examined in detail, and indeed no tertiary plastid genome has yet been sequenced.

Methodology/Principal Findings: Here we describe the complete plastid genomes of the two closely related dinotoms, Durinskia baltica and Kryptoperidinium foliaceum. The D. baltica (116470 bp) and K. foliaceum (140426 bp) plastid genomes map as circular molecules featuring two large inverted repeats that separate distinct single copy regions. The organization and gene content of the D. baltica plastid closely resemble those of the pennate diatom Phaeodactylum tricornutum. The K. foliaceum plastid genome is much larger, has undergone more reorganization, and encodes a putative tyrosine recombinase (tyrC) also found in the plastid genome of the heterokont Heterosigma akashiwo, and two putative serine recombinases (serC1 and serC2) homologous to recombinases encoded by plasmids pCF1 and pCF2 in another pennate diatom, Cylindrotheca fusiformis. The K. foliaceum plastid genome also contains an additional copy of serC1, two degenerate copies of another plasmid-encoded ORF, and two non-coding regions whose sequences closely resemble portions of the pCF1 and pCF2 plasmids.

Conclusions/Significance: These results suggest that while the plastid genomes of two dinotoms share very similar gene content and genome organization with that of the free-living pennate diatom P. tricornutum, the K. foliaceum plastid genome has absorbed two exogenous plasmids. Whether this took place before or after the tertiary endosymbiosis is not clear.

Introduction

The path of plastid evolution has been neither simple nor linear, but rather full of twists and turns. After the divergence of glaucophytes, red and green algae following primary endosymbiosis, plastids spread by the secondary and tertiary uptake of these eukaryotic algae by new eukaryotic hosts [1,2,3,4]. Each of these endosymbiotic events involved a massive loss of genes from the symbiont as well as a large scale transfer of other genes to its new host. In the primary endosymbiosis this meant gene transfers from the ancient cyanobacterium, whereas in secondary and tertiary endosymbioses most gene transfer would be from the nucleus of the endosymbiont alga to the nucleus of its new host [5,6,7]. The products of many of these genes would be targeted to the plastid, which necessitated the development of a new protein targeting system to direct the protein products back to their correct location [8,9].

These processes have been most thoroughly studied in primary and secondary plastids, but tertiary endosymbioses add another layer of complexity to the process. In tertiary endosymbiosis an alga with a secondary plastid is taken up by another eukaryote, and to date the only lineage known to take up tertiary plastids is dinoflagellates, where tertiary plastids derived from three different lineages are known: Karenia and Karlodinium species with plastids derived from a haptophyte [7,10]; Dinophysis species with cryptophyte derived-plastids [11,12,13]; and a small but growing group of dinoflagellates harboring a diatom endosymbiont [14,15,16,17,18], which we refer to as dinotoms. By dinotoms, we refer to the whole biological system that includes both the dinoflagellate host and the diatom endosymbiont.

Dinotoms are widely distributed in both freshwater and marine environments and some, most notably Kryptoperidinium foliaceum and Peridinium quinquecorne, form blooms with occasional harmful effects.
[19,20,21]. The Dinoflagellate host component are currently divided into at least five distinct genera, *Kryptoperidinium, Peridinium, Gymnodinium,* and *Galeidiniium* [14,15,16,17,18,22], while the endosymbiont components have been shown to originate from three different diatom lineages, one pennate [23,24,25,26] and two centric [27,28].

In haptophyte and cryptophyte endosymbiont-containing dinoflagellates, the endosymbiont has reduced the point that only the plastid itself remains. In contrast, the diatom endosymbionts in dinoflagellates have preserved more of their genetic and cellular identity than any other secondary or tertiary plastid. The endosymbiont has lost some characters such as its cell wall, motility, and the ability to condense its chromosomes normally or divide mitotically [14,29,30], but it retains a large nucleus and the nuclear genome, mitochondria and the mitochondrial genome [24,31], as well as cytosolic ribosomes, endoplasmic reticulum (ER), and dictyosomes in an extensive cytoplasm that is separated by a double membrane [14,32]. Despite such unusual degree of character retention, the endosymbiont is permanently integrated within its host, and it is present at all different stages of the life cycle including cell division, sexual reproduction, and cyst formation [30,33,34].

The number of plastids in dinoflagellates varies from one or two (in gametes) to as many as 30 to 40 (in zygotes). Chlorophyll formation [30,33,34].

The D. baltica and K. foliaceum plastid genomes (Figure 1) map as circular molecules divided into large single-copy (LSC) and small single-copy (SSC) regions by the two inverted repeats (IRs), a quadripartite structure that is common to many other algal plastid genomes including the pennate and centric diatoms *P. tricornutum* and *Thalassiosira pseudonana,* respectively [41]. The general characteristics of all diatom and diatom-derived plastid genomes are juxtaposed in Table 1.

Like other related plastids, both dinotom plastid genomes use standard plastid/bacterial genetic code, with G TG as alternative start codon to ATG. This alternative start codon is found in the four plastid genes (chs, *sp23*, rps9, and *rpl5*) in all four diatom and diatom-derived plastid genomes.

The IRs in *D. baltica* are very similar to those of the free-living diatom *P. tricornutum* and feature almost the same gene composition (trnM, *ycf9*, *rs*, *trnD*, *trnA*, *rL*, and *rM*) and size. The slight difference in the size and composition of the IRs in these two plastids is due to the presence of *pshY* in the IRs of *P. tricornutum* instead of partial *cysA* in the IRs of *D. baltica.* The plastid genome size and gene content of *D. baltica* are remarkably similar to those of *P. tricornutum.* The *D. baltica* plastid genome is only about 900 bp shorter than that of *P. tricornutum,* and the two genomes share 159 genes in common. The *D. baltica* plastid genome encodes 127 protein-coding genes, three rRNAs, 27 tRNAs, a sufficient set for their plastid protein synthesis machinery, one transfer-messenger RNA (mRNA), *ssr,* and one plastid signal recognition particle RNA, *fsh.* Interestingly, like *P. tricornutum* it has retained *sflF,* encoding a tRNA synthetase, which is missing from the plastid genome of *T. pseudonana* but is present in red algal plastid genomes [41]. Only three genes present in the plastid genome of *P. tricornutum* are absent from the *D. baltica* genome: *tfs* (not found in other diatom plastid genomes), *acP,* and *ycf42.*

In contrast, the *K. foliaceum* plastid genome is considerably larger than the plastid genomes of *D. baltica* and *P. tricornutum,* by about 24 and 23 kb, respectively. The IRs in *K. foliaceum* are shorter than those of *D. baltica* and *P. tricornutum* by almost 1 kb because of the absence of *trnP* and *ycf9* in the *K. foliaceum* IRs, so its larger size is not due to the increased size of the IRs as seen in *T. pseudonana* [41]. Instead, both SSC and LSC in *K. foliaceum* are sizably larger than those observed in other diatoms, owing to the presence of more apparently non-coding DNA (see below) and protein-coding genes. In addition to the same 159 genes found in both *D. baltica* and *P. tricornutum,* the plastid genome of *K. foliaceum* encodes a putative tyrosine recombinase gene, *tyrC,* two putative serine recombinase genes, *serC1* and *serC2,* two smaller ORFs, ORF93 and ORF92 both related to *serC1,* and seven putative open reading frames (ORFs) larger than 150 amino acids (aa), or 15 ORFs if the threshold for annotation is lowered to 100 aa.

**Compactness of dinotom plastid genomes**

Like other chromist plastid genomes, the plastid genomes of the two dinotom possess some of the features of a compact genome. They lack introns, and the same four overlapping pairs of genes found in diatoms [41] are also found in both dinotom plastids with the identical length of overlap: *pshD-pshC,* *atpD-atpF,* *sufC-sufB,* and *rpl16-rpl23* with 53, 4, 1, and 8 nucleotides (nt) overlap, respectively. In addition, *dnaB* and *trnF* have no intergenic spacer in *D. baltica* and *P. tricornutum,* whereas this gene pair is separated by 1 nt in *K. foliaceum.* Similarly, *rpl14* and *rpl24* are separated by a single nt in *D. baltica,* *K. foliaceum,* and *P. tricornutum.*

The plastid genomes of *D. baltica,* *K. foliaceum,* and *P. tricornutum* demonstrate no considerable change in the length of their genes (Figure S1). Out of the common 159 genes, 108 are invariant in length and the sum of all differences between *P. tricornutum* genes

---

**Results**

**Genome structure, gene repertoire, and GC content of the D. baltica and K. foliaceum genomes**

The *D. baltica* CSIRO CS-38 plastid genome (GenBank: GU591327) assembly contained 18704 Titanium pyrosequencing 454 reads (363 bp average), amounting to 6.8 Mbp, or 38-fold coverage of the genome. The *K. foliaceum* CCMP 1326 plastid genome (GenBank: GU591328) assembly included 7274 reads (383 bp average) amounting to 2.8 Mbp, or 20-fold coverage. Over 20 kb of the *D. baltica* and 75 kb of *K. foliaceum*’s plastid genome sequences were also ascertained by PCR and Sanger sequencing (see Methods).
and those of *D. baltica* and *K. foliaceum* amount to a mere 199 and 142 bp, respectively (and only 57 bp between *K. foliaceum* and *D. baltica*; Figure S1).

Average intergenic spaces in *D. baltica* (94.3 bp) are only slightly longer than those of *P. tricornutum* (88.4 bp), but in *K. foliaceum* the spacing is more than twice as long (246.7 bp on average) (Table 1). Even when putative ORFs in *K. foliaceum* are brought into account, the average spacing is 180 bp, but more importantly when the average is calculated based only on the 159 shared genes, the average is only 94.1 bp, about equivalent to *D. baltica* and *P. tricornutum*.

### Conserved ordered gene blocks

To investigate the conservation of genome structure, MAUVE [42] was used to detect gene clusters. Overall, 23 conserved clusters were found in *T. pseudonana*, *P. tricornutum*, *D. baltica*, and *K. foliaceum*. If *T. pseudonana* (a more distantly related centric diatom) is removed from analysis, 14 larger blocks are found. In pairwise comparisons, nine large conserved blocks are shared between *P. tricornutum* and *D. baltica*, 13 between *P. tricornutum* and *K. foliaceum*, and nine between *D. baltica* and *K. foliaceum*. However, taking into account the presence or absence of a single gene between large blocks extends these blocks (to 16 conserved blocks among the three species amounting to more than 108 kb, 10 blocks between *P. tricornutum* and *D. baltica*, 14 blocks between *P. tricornutum* and *K. foliaceum*, and 13 blocks between *D. baltica* and *K. foliaceum*) (Figure 2). The largest block conserved among the three species spans more than 31 kb and includes 46 genes appearing in the same order, encoded on the same strands (*ycf33*, *trnI*, *trnS* … *rpoC1*, *rpoC2*, *rps2*). The largest conserved gene block between *P.*

---

**Table 1. General characteristics of plastid genomes in dinotoms compared to diatoms.**

|                        | *Durinskia baltica* | *Kryptoperidinium foliaceum* | *Phaeodactylum tricornutum* a | *Thalassiosira pseudonana* a |
|------------------------|---------------------|-------------------------------|-------------------------------|-------------------------------|
| **Size (bp)**          | 116470              | 140426                        | 117369                        | 128814                        |
| **IR**                 | 7067                | 6017                          | 6912                          | 18337                         |
| **SSC**                | 39813               | 56521                         | 39871                         | 26889                         |
| **LSC**                | 62523               | 71871                         | 63674                         | 65250                         |
| **GC content (%)**     | 32.5                | 32.4                          | 32.56                         | 30.66                         |
| **rRNA genes**         | 46.9                | 47.0                          | 47.2                          | 47.0                          |
| **tRNA genes**         | 53.5                | 53.7                          | 53.0                          | 52.6                          |
| **Other RNAs**         | 27.3                | 28.3                          | 26.0                          | 25.6                          |
| **Protein-coding genes** | 32.4               | 33.0                          | 32.9                          | 31.5                          |
| **Intergenic spacer b** | 22.1               | 26.5                          | 18.8                          | 16.3                          |
| **Coding sequence (%)** | 86.7               | 71.9                          | 87.5                          | 85.2                          |
| **Gene content d**     | 159                 | 160                           | 162                           | 159                           |
| **Protein-coding genes** | 127                | 128                           | 130                           | 127                           |
| **rRNA genes**         | 3                   | 3                             | 3                             | 3                             |
| **tRNA genes**         | 27                  | 27                            | 27                            | 27                            |
| **Other RNAs**         | 2                   | 2                             | 2                             | 2                             |
| **Introns**            | 0                   | 0                             | 0                             | 0                             |
| **Overlapping genes**  | 4                   | 4                             | 4                             | 4                             |
| **Average intergenic spacer (bp)** | 94.3              | 246.7                         | 88.4                          | 108.2                         |

*aData taken from Oudot-Le Secq et al [41].

*bDuplicated genes were taken into account (size/number of genes).

*cConserved genes (unique and duplicated) and ORFs were considered as coding sequences.

*dDuplicated genes and unique ORFs were not taken into account.*
Similarly, two small conserved blocks of genes (the transition of the three plastid genomes of have evolved, GRIMM [43] was used to identify 14 inversions in respect to the other two species. Inverted in *D. baltica* dnaK and *psbC* show two inversions (blocks 1, 4), five inversions/ translocations (blocks 2, 8, and 9) and two translocations (block 10 and *clpC*, *psbA*) appear in inverted orientation in *D. baltica* with respect to the other two species.

To see how the organization of blocks of shared genes might have evolved, GRIMM [43] was used to identify 14 inversions in the transition of the three plastid genomes of *P. tricornutum*, *D. baltica*, and *K. foliaceum*. If *T. pseudonana* is added, 23 inversions are required. In pairwise analyses, GRIMM also proposes 6 inversions for *P. tricornutum* and *D. baltica*, 9 for *P. tricornutum*, and *K. foliaceum*, and 8 for *D. baltica* and *K. foliaceum*.

Closer manual inspections reveal that compared to the plastid genome of *P. tricornutum* fewer rearrangements of the conserved gene blocks distinguish *D. baltica* from *K. foliaceum*: only three inversions (blocks 2, 8, and 9) and two translocations (block 10 and *clpC* gene) are detected in *D. baltica* versus two inversions (blocks 1, 4), six inversions/translocations (blocks 10, 11, 7, 6, 8, and 12) and three translocations (blocks 9, 13, and *clpC* gene) in *K. foliaceum* (Figure 2). Compared to the plastid genome of *D. baltica*, *K. foliaceum* shows two inversions (blocks 1, 4), five inversions/translocations (blocks 6, 10, 9, 8, and 7) and two translocations (blocks 12, and 11).

All the three missing genes from the plastid genomes of *D. baltica* and *K. foliaceum*, present in *P. tricornutum*, are located in its LSC region. Curiously, however, most of the rearrangements seem to have occurred in the SSC regions of the plastid genomes of *D. baltica* and *K. foliaceum* (Figure 2).

**Low gene density regions of the *Kryptoperidinium foliaceum* plastid genome**

There are nine distinct regions (labeled with Roman numerals in Figures 1 and 2) within the *K. foliaceum* plastid genome that have a low gene density and do not show any similarity to *D. baltica*, *P. tricornutum*, or *T. pseudonana*. Six of the nine regions are dispersed within the SSC (regions III-VIII, totaling to more than 17 kb) and three within the LSC (regions I, II, and IX, amounting to about 7.5 kb). All four junctions of the IRs with the SSC and LSC include such regions: II and IX at the boundary of IRa and LSC, and III and VIII at the junction of IRb and SSC. These nine distinct regions collectively amount to more than 24 kb ranging in size from 905 bp (region IX at the boundary of IRa and LSC) to 4852 bp (region III at the junction of IRb and SSC) with an overall GC content of 30.4%, which is 2% lower than the GC content of the genome as a whole (Table 1), and 2.4% lower than the rest of the genome.

Interestingly, regions I and III are each bounded by two imperfect palindromes. A 35 bp palindromic site is placed near the *rps2* gene and a 44 bp palindromic sequence is placed at its end, near *rbcS*. Region III is similarly bounded by two palindromic sequences: a 25 bp sequence near the *rps5* gene and a 42 bp sequence near *dnaK*. Another 32 bp palindromic site is located at one end of region V (near *psbA*).

**tyrC in *K. foliaceum* and Heterosigma akashiwo**

The tyrC gene located in region III shows strong similarity to a putative site-specific tyrosine recombinase protein (TyrC) encoded within the plastid genome of the raphidophyte heterokont *H. akashiwo* [44]. The conceptual translation of tyrC also shows similarity, albeit much weaker, to putative integrase/recombinases proteins encoded in the plastid genome of the chlorophycean alga *Oedogonium cardiaicum* and in the mitochondrial of the charophyte *Chaosphoridium globosum*. As revealed by NCBI Conserved Domain Database (CDD) searches [45], the *K. foliaceum* TyrC has conserved all the major catalytic, active, and DNA-binding sites required by this protein for integrase/recombinase activity, including His 250 and the four invariably conserved sites Arg 145, Arg 253, Lys 172, and Tyr 285 (Figure S2) [46,47,48]. RT-PCR was performed on tyrC and the amplicon sequenced (data not shown), confirming that this gene is transcribed and most likely expressed in the *K. foliaceum* plastid genome.
Similarity between the \textit{K. foliaceum} plastid genome and pCf1 and pCf2 plasmids in \textit{Cylindrotheca fusiformis}

A total of five ORFs (orf441, serC1 [orf205], serC2 [orf212], orf92, and orf92) in the \textit{K. foliaceum} plastid genome show strong similarity to ORFs found in the pCf1 and pCf2 plasmids of the pennate diatom \textit{C. fusiformis}. Each of these two plasmids includes several ORFs, two pairs of which share considerable similarity (ORF217 of pCf2 and ORF218 of pCf1 with almost 80\% aa identity and ORF484 of pCf2 and ORF482 of pCf1 with 54\%) [49]. \textit{K. foliaceum} ORF141 (region VI) shares 57\% and 47\% aa identity with ORF484 (aa 106 to aa 324) from pCf2 and ORF482 from pCf1 plasmid, respectively. The \textit{K. foliaceum} SerC1 shares 76\% and 66\% aa identity with ORF218 from pCf1 and ORF217 from pCf2, respectively, while SerC2 displays 60\% aa identity with \textit{C. fusiformis} ORF218 and 61\% with ORF217. Interestingly, SerC1 and SerC2 share less similarity to each other (57\% of aa identity) than they do with \textit{C. fusiformis} ORF218 and ORF217, and serC2 also shares a single codon insertion specifically with ORF217. \textit{K. foliaceum} orf92 and orf92 (region I) appear to be truncated versions of the \textit{C. fusiformis} ORF218, corresponding to amino acids 1 to 93 and 117 to 206, respectively. The two \textit{K. foliaceum} fragments are separated from each other by 69 bp, the conceptual translation of which shares 97\% identity with \textit{C. fusiformis} ORF218 amino acids 95 to 116, however this region contains two stop codons suggesting it is a pseudogene.

CDD searches [45] reveal that SerC1 and SerC2 in \textit{K. foliaceum} have retained almost all the catalytic, DNA-binding, presynaptic, and synaptic residues found in other site-specific serine recombinases (Figure S3). Once again, RT-PCR (data not shown) showed that both \textit{serC1} and \textit{serC2} are transcribed and most likely expressed in \textit{K. foliaceum}.

In addition to the five abovementioned ORFs, a number of dispersed non-coding stretches of DNA in several distinct regions of the \textit{K. foliaceum} plastid genome show strong similarity to the \textit{C. fusiformis} pCf1/pCf2 plasmid sequences. A 350-bp sequence in region III of \textit{K. foliaceum} plastid genome (Figure 1) shows 92\% nt identity to a portion of ORF484 (corresponding to aa 322 to 43) from pCf2. This 350-bp sequence does not include any ORF but contains two stop codons in the same frame that shows similarity to ORF484. Similarly, in region VI about 600 bp immediately downstream of ORF411 shows 71\% nt identity to \textit{C. fusiformis} ORF484 and region IV contains 240 bp, 110 bp, and 120 bp sequences with strong similarity to non-coding regions of pCf2 (72\%, 72\% and 74\% nt identity). There is also a 156 bp sequence in region VIII with strong similarity (66\% nt identity) to the non-coding region at the end of pCf1.

We also searched the non-plastid 454 sequence data (both assembled and singleton reads) for potential plasmids similar to those of \textit{C. fusiformis}, however, none were found. Since ORF218 and ORF482 are close to each other in pCf1 and similarly ORF217 and ORF484 are close in pCf2, we also designed outward primers for the two corresponding \textit{K. foliaceum} ORFs (ORF205 or \textit{serC1} and ORF414). All attempts to PCR amplify a small product were unsuccessful.

\section*{Discussion}

The divergent evolution of two tertiary plastid genomes of diatom origin

The plastid genomes of the tertiary endosymbionts of \textit{K. foliaceum} and \textit{D. baltica} share numerous common features with those of free-living diatoms, including gene content, ordered gene blocks, and overall genome structure. Especially striking is the similarity to the pennate diatom \textit{P. tricornutum}, with which they share more than 108 kb of syntenic gene clusters, reconfirming the pennate diatom ancestry for these endosymbionts also suggested by molecular phylogeny [23,24,25,26,27]. Recent phylogenetic analyses suggest a particularly close relationship with the genus \textit{Nitzschia} [24,26,27]. Unfortunately, at present the only pennate diatom plastid genome known is from the more distantly related \textit{P. tricornutum}. Considering the high degree of conservation between its plastid genome content, composition, and organization, and those of \textit{K. foliaceum} and \textit{D. baltica}, we suggest the plastid genome of closer free-living pennate diatom relatives will reveal even fewer structural changes have taken place since the tertiary endosymbiosis.

Notwithstanding the high degree of conservation between the tertiary plastid genomes and their free-living relative \textit{P. tricornutum}, the plastid genome of \textit{K. foliaceum} is different in one interesting respect. Its genome is more than 23 kb larger than those of its close relatives and the majority of the additional sequence falls into a handful of specific regions. Most of this sequence shows no strong similarity to known sequences, but a few regions share a strong similarity to the plasmids pCf1 and pCf2 in the pennate diatom \textit{C. fusiformis}. The genome also encodes two site-specific serine recombinase genes also shared with those plasmids, as well as a site-specific tyrosine recombinase gene present in the plastid genome of another heterokont, the raphidophyte \textit{H. akashiwo}.

Earlier hybridization experiments suggest that either pCf1/PCf2 or plasmids with considerable sequence similarity existed in three strains of \textit{C. fusiformis}, in one of three strains of \textit{C. closterium}, in \textit{Nitzschia angulata}, and in \textit{N. carolinensis} [49,50], but no sequence data were available to indicate the possible sites of hybridization or integration of such plasmids with the plastids. In \textit{K. foliaceum} we find no evidence for the presence of intact plasmids, only the putative \textit{serC1} and \textit{serC2} genes and some degenerated fragments integrated into the plastid genome. Overall we can conclude that both plasmids were present in an ancestor of \textit{K. foliaceum}, and fragments of both have persisted by integration into the plastid genome in \textit{K. foliaceum}, but not \textit{D. baltica}.

While the two serine recombinase genes in the \textit{K. foliaceum} plastid genome clearly originated from plasmids and probably functioned in spreading those plasmids, the origin of the tyrosine recombinase/integrase is less clear. TyrC in \textit{H. akashiwo} has been speculated to be involved in converting multimeric plastid molecules to monomeric forms [44], but no sequence data were available to indicate the possible sites of hybridization or integration of such plasmids with the plastids. In \textit{K. foliaceum} we find no evidence for the presence of intact plasmids, only the putative \textit{serC1} and \textit{serC2} genes and some degenerated fragments integrated into the plastid genome. Overall we can conclude that both plasmids were present in an ancestor of \textit{K. foliaceum}, and fragments of both have persisted by integration into the plastid genome in \textit{K. foliaceum}, but not \textit{D. baltica}.

The plastid genomes of the tertiary endosymbionts of \textit{K. foliaceum} and \textit{D. baltica} share numerous common features with those of free-living diatoms, including gene content, ordered gene blocks, and overall genome structure. Especially striking is the similarity to the pennate diatom \textit{P. tricornutum}, with which they share more than
The ancestral state of the tertiary endosymbiont genome

Despite the more recent common ancestry of the D. baltica and K. foliaceum plastids, the D. baltica and P. tricornutum plastid genomes share a much greater overall similarity in structure, in large part due to the presence of plasmid-associated sequences in K. foliaceum. Which of these two tertiary endosymbionts better represents the state of their common ancestor is not entirely clear. On one hand, the close similarity between the genomes of D. baltica and P. tricornutum might suggest this represents the ancestral state and that the genome of K. foliaceum subsequently acquired plasmids (which are selfish and frequently mobile elements) leading to its expansion and reorganization. On the other hand the plasmids are known to exist in some form in other pennate diatoms that are more closely related to the tertiary plastids, most notably some Nitzschia species [54]. If they were in the ancestor of the tertiary endosymbionts then D. baltica would have to have ridded itself of all evidence of both plasmids to revert to a highly similar form as P. tricornutum. Both explanations, the multiple movements of plasmids between close relatives or the complete loss of plasmids in certain lineages, are consistent with the seemingly sporadic interspecies and intraspecies distribution of these plasmids in diatoms: only 5 out of 18 examined diatom species and only 1 out of 3 strains of the pennate diatom C. clusterium are suggested to have similar plasmids [54]. Perhaps the most likely explanation is that the ancestor possessed unintegrated plasmids and a plastid genome with a structure highly similar to that of P. tricornutum and D. baltica. In K. foliaceum the plasmids would have integrated into the main plastid genome, degenerated, and promoted the reorganization of many gene blocks, whereas in D. baltica all traces of the plasmids were lost, which is not unlikely if they never integrated into the plastid genome.

Conclusions

Here we describe the first completely sequenced plastid genomes from tertiary endosymbionts, specifically the diatom-derived plastids of two dinoflagellates, D. baltica and K. foliaceum. Both genomes have retained many characteristics of the ancestral, free-living diatom, including elements of genome structure, gene content, and ordered gene clusters. The plastid genome of K. foliaceum is much larger than that of D. baltica, and contains a site-specific tyrosine recombinase gene also found in the heterokont H. akashio, and the incorporation, maintenance, and degradation of genetic material from two similar plasmids found in other pennate diatoms, which have resulted in the addition of two site-specific serine recombinases.

Materials and Methods

Strains and culture conditions

Cultures of Durvillaea baltica (Peridinium balticum) CSIRO CS-38 and Kryptoperidinium foliaceum CCMP 1326 were obtained respectively from the CSIRO Microalgal Supply Service (CSIRO Marine and Atmospheric Research Laboratories, Tasmania, Australia) and from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (West Boothbay Harbor, ME, USA). D. baltica cultures were maintained in GSe medium at 22 °C (12:12 light:dark cycle) whereas P. tricornutum reflective index was adjusted to 1.3995 and 1.4000 for D. baltica and K. foliaceum respectively and Hoechst 33258 (Invitrogen, Carlsbad, CA, USA) was added to the solution (100 μg/ml for D. baltica and 120 μg/ml for K. foliaceum). Ultracentrifugation was conducted in a Beckman L8 80 M ultracentrifuge, using a VTi 80 (Beckman) rotor at 55000 rpm and 20°C for 22 and 20 hours for D. baltica and K. foliaceum, respectively. The extracted A+T-rich satellite bands were washed 4 times with CsCl/TE buffer-saturated isopropanol to remove the Hoechst dye. The DNA was precipitated from CsCl as described previously [55] and eluted in Tris HCl pH 8.0. The purified DNA was amplified using the REPLI-g mini kit (Qiagen, Mississauga, ON, Canada). The total genomic DNA from both species used for the PCR reactions was obtained after 2 phenol, 1 Phenol:Chloroform:Isoamyl Alcohol (25:24:1), and 2 chloroform extractions and ethanol precipitation. Total RNA extraction and RT-PCR were carried out as described previously [31] using the following primers: tyrC_F, CCAGAACTGGCTGTTATATGGTATGTAATTCA. PCR was performed in a 25 μl reaction that included 1X E2 Buffer (20 mM Tris-HCl pH 9.0, 50 mM KCl, 1.5 mM MgCl₂, 0.2 mM dNTPs, 0.5 μM of each forward and reverse primer, 0.5 units of Taq DNA polymerase) and 200 μM each of forward and reverse primers. The PCR products were then inserted into pCR-TOPO vector (Invitrogen, Carlsbad, CA, USA) and sequenced (GATC Biotech, USA). DNA and RNA extractions, PCR, RT-PCR, and DNA fractionation and precipitation

Cells were collected and ground as described previously [31]. Ground cells were lysed in 50 mM Tris-HCl, 100 mM EDTA, 100 mM NaCl, pH 8.0 in the presence of β-mercaptoethanol (2%), SDS (2%) and protease K (300 μg/ml) at 50°C for 1 hour. In case of D. baltica, 6 phenol and 1 phenol/chloroform extractions were performed, whereas for K. foliaceum, 3 phenol, 1 phenol/chloroform, and 2 chloroform extractions were conducted. Organellar A+T-rich DNA was separated from nuclear DNA using CsCl gradient density centrifugation. The initial CsCl reflective index was adjusted to 1.3995 and 1.4000 for D. baltica and K. foliaceum respectively and Hoechst 33258 (Invitrogen, Carlsbad, CA, USA) was added to the solution (100 μg/ml for D. baltica and 120 μg/ml for K. foliaceum). Ultracentrifugation was conducted in a Beckman L8 80 M ultracentrifuge, using a VTi 80 (Beckman) rotor at 55000 rpm and 20°C for 22 and 20 hours for D. baltica and K. foliaceum, respectively. The extracted A+T-rich satellite bands were washed 4 times with CsCl/TE buffer-saturated isopropanol to remove the Hoechst dye. The DNA was precipitated from CsCl as described previously [55] and eluted in Tris HCl pH 8.0. The purified DNA was amplified using the REPLI-g mini kit (Qiagen, Mississauga, ON, Canada). The total genomic DNA from both species used for the PCR reactions was obtained after 2 phenol, 1 Phenol:Chloroform:Isoamyl Alcohol (25:24:1), and 2 chloroform extractions and ethanol precipitation. Total RNA extraction and RT-PCR were carried out as described previously [31] using the following primers: tyrC_F, CCA-TAACTGCGTAATATAGCCG, tyrC_R, TCTGAAAGAAT- TAAATCTAATCAAGG, serC1_F, CCAGTTAATCTTGTA- CTGTCGG, serC1_R TTGGCTCTGCTGCTAACG, serC2_F TGTGTCCTGAAAGTCACAAGG, and serC2_R AAC- TATCGGTTTATATGTGTTAATCT. PCR was performed using the EconoTaq PLUS GREEN kit (Lucigen, Middleton, WI, USA).

Genome sequencing

The D. baltica and K. foliaceum plastid genomes were sequenced using massively parallel GS-FLX DNA pyrosequencing (Roche 454 Life Sciences, Branford, CT, USA). The GS-FLX shotgun libraries and pyrosequencing using the GS-FLX Titanium reagents were carried out at the Genome Quebec Innovation Centre. The Newbler de novo assemblies edited and re-assembled with CONSED 19 [56]. Plastid sequences in assembled and unassembled sequence pools were identified by BLAST searches [57]. Ambiguous pyrosequencing homopolymer stretches in the assemblies were verified by PCR/Sanger sequencing, which invariably yielded sequence that preserved the open reading frame. The only exceptions were fragments of plasmid-derived genes in the K. foliaceum plastid genomes that are concluded to be pseudogenes.

Genome annotation and analysis

Genes were identified by DOGMA searches [58] and by BLAST homology searches [57] against the NCBI nonredundant database (http://www.ncbi.nlm.nih/BLAST), and annotated using Artemis 11 [59]. Protein-coding genes were identified using GETORF from EMBOSS 6.0.1 [60] and ORFFINDER at NCBI, with start codons ascertained by comparison with known homologues. Positions of tRNA-encoding genes were determined with tRNAscan-SE [61]. Ribosomal and miscellaneous RNA-encoding genes were annotated by comparison with P. tricornutum and T. pseudonana homologues. Repeated elements were searched for using PpMaker [62], REPuter [63], and FUZZYNUC from the EMBOSS package [60]. Physical maps were generated using GenomeVx [64] and further edited manually. Conserved gene clusters between the D. baltica, K. foliaceum and P. tricornutum plastid genomes were identified using MAUVe [42] and by visual inspection of the physical maps. Hypothetical gene inversions

Dinotom Plastid Genomes

PLOS ONE | www.plosone.org 7 May 2010 | Volume 5 | Issue 5 | e10711
between the 1.59 genes that are shared between the three genomes were examined using GRIMM [43]. Translocations were identified by manual inspections and defined as homologous portions of genomes (i.e. a gene or a conserved blocks of genes) appearing at different loci in the same orientation.

Supporting Information

Figure S1 Gene size differences. Pairwise comparisons of gene size differences (larger than 5 bp in at least 2 of the 3 species) in the plastid genomes of *Phaeodactylum tricornutum* (Pt), *Durinskia baltica* (Db), and *Kryptoperidinium foliaceum* (Kf).

Found at: doi:10.1371/journal.pone.0010711.s001 (0.92 MB EPS)

Figure S2 Conserved residues found in TyrC. The conserved residues found in the TyrC recombinase encoded in the plastid genomes of *Kryptoperidinium foliaceum* and other site-specific tyrosine recombinases. The sequences of *Heterosigma akashiwo's* TyrC were manually added to the Conserved Domain Database (CDD) alignment for *K. foliaceum* Tyrc. The conserved residues with specific functions are marked with a number sign (#) above the alignments. Shaded residues indicate invariable sites among all the recombinases in the alignment, and the long rectangular boxes highlight conserved sites among all the recombinases in the alignment except one.

Found at: doi:10.1371/journal.pone.0010711.s001 (1.74 MB EPS)

Acknowledgments

We would like to thank Dr. Beverley Green and Dr. Yunkun Dang for their advice and assistance with our CoCl gradient density experiments.

Author Contributions

Conceived and designed the experiments: BI JFP PJK.Performed the experiments: BI JFP. Analyzed the data: BI JFP PJK. Wrote the paper: BI JFP PJK.

References

1. McFadden GI (2001) Primary and secondary endosymbiosis and the origin of plastids. Journal of Phycology 37: 951–959.
2. Archibald JM, Keeling PJ (2002) Recycled plastids: A 'green movement' in eukaryotic evolution. Trends in Genetics 10: 577–594.
3. Palmer JD (2003) The symbiotic birth and spread of plastids: How many times and how often? Journal of Phycology 39: 4–11.
4. Bhattacharya D, Youn HS, Hackett JD (2004) Photosynthetic eukaryotes unite: Endosymbioses connect the dots. Bioessays 26: 56–60.
5. Archibald JM, Rogers MB, Toop M, Ishida K, Keeling PJ (2003) Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. Proceedings of the National Academy of Sciences of the United States of America 100: 7678–7683.
6. Dora JA, Fruhmoller M, Su V, Maier UG, Martin W, et al. (2006) Evidence for nucleomorph to host nucleus gene transfer: Light-harvesting complex proteins from cryptomonads and chlorarachniophytes. Protist 151: 239–252.
7. Patron NJ, Waller RF, Keeling PJ (2006) A tertiary plastid uses genes from two endosymbionts. Journal of Molecular Biology 357: 1373–1382.
8. Jarvis P, Self J (2002) Tox, and chloroplast protein import (vol 1541, pg 64, 2001). Biochimica Et Biophysica Acta-Molecular Cell Research 1590: 177–189.
9. Bruce BD (2001) The paradox of plastid transit peptides: Conservation of function despite divergence in primary structure. Biochimica et Biophysica Acta 1541: 2–21.
10. Teng S, Dahlberg OJ, Shalchian-Tabrizi K, Klavenson D, Rudi K, et al. (2003) Phylogenetic analyses indicate that the 19′hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. Molecular Biology & Evolution 17: 718–729.
11. Hewes CD, Mitchell BG, Moisan TA, Vernet M, Reid FMH (1998) The phycothyln signatures of chloroplasts from three dinoflagellate species: A microanalytical study of *Dinophysis canda*, *D. ficta* and *D. acuminata* (Dinophytales, Dinophyceae). Journal of Phycology 34: 945–951.
12. Schouten E, Elbrachter M (1988) Cryptophycean-like double membrane-bound chloroplasts in the dinoflagellate, *Dinophyces thor* evolution, phylogenetic and toxicological implications. Botanica Acta 101: 196–203.
13. Hackett JD, Maranda L, Youn HS, Bhattacharya D (2003) Phylogenetic evidence for the cryptophyte origin of the plastid of *Dinophyces* (Dinophytales, Dinophyceae). Journal of Phycology 39: 440–446.
14. Tomas RN, Cox ER (1973) Observations on symbiosis of *Peridinium Balticum* and its intracellular alga I. Ultrastructure. J Phycol 9: 304–323.
15. Dodge JD (1971) A dinoflagellate with both a mesokaryotic and a eukaryotic nucleus: Part 1 fine structure of the nuclei. Protoplasma 73: 145–157.
16. Hornguchi T, Pienaar RN (1994) Ultrastructure of a new marine sand-dwelling dinoflagellate, *Gymnodinium quadralvesulum* sp. nov. (Dinophyceae) with special reference to its endosymbiotic alga. European Journal of Phycology 29: 237–245.
Dinotom Plastid Genomes

34. Figueroa RI, Bravo I, Fraga S, Garces E, Llaveria G (2009) The life history and cell cycle of *Cryptoerythra longicollis*, a dinoflagellate with two eukaryotic nuclei. Protoplanet 160: 285–300.

35. Withers NW, Cox ER, Tomas R, Haxo FT (1977) Pigments of the dinoflagellate *Peridinium balticum* and its photosynthetic endosymbiont. Journal of Phycology 13: 354–358.

36. Jeffrey SW, Siedlicki M, Haxo FT (1975) Chloroplast pigment patterns in dinoflagellates. Journal of Phycology 11: 374–384.

37. Kite GC, Dodge JD (1965) Structural organization of plastid DNA in two anomalously pigmented dinoflagellates. Journal of Phycology 21: 59–66.

38. Mandelli EF (1968) Carotenoid pigments of dinoflagellate *Glenodinium foliacum* Stein. Journal of Phycology 4: 347–46.

39. Schnepf E, Ehlbrechter M (1999) Dinoflagelate chloroplasts and phylogeny - A review. Grana 38: 81–97.

40. Jeffrey SW, Vesk M (1976) Further evidence for a membrane bound endosymbiont within the dinoflagellate *Peridinium foliacum*. Journal of Phycology 12: 450–455.

41. Oudot-Le Secq MP, Grimwood J, Shapiro H, Armbrust EV, Bowler C, et al. (2009) Characterization of 2 circular plasmids from the marine diatom *Cylindrotheca fusiformis* - Plasmids hybridize to chloroplast and nuclear-DNA. Molecular & General Genetics 267: 179–186.

42. Barre FX, Sohalle B, Michel B, Aroyo M, Robertson M, et al. (2001) Circles: The replication-recombination-chromosome segregation connection. Proceedings of the National Academy of Sciences of the United States of America 98: 8189–8195.

43. Tesler G (2002) GRIMM: genome rearrangement web server. Bioinformatics 18: 492–493.

44. Darling ACE, Mau B, Blattner FR, Perna NT (2004) Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Research 14: 1394–1403.

45. Cattolico RA, Jacobs MA, Zhou Y, Chang J, Duplessis M, et al. (2008) Genome rearrangements and their cellular and evolutionary significance. Genome Research 18: 492–493.

46. Friesen H, Sadowski PD (1992) Mutagenesis of a conserved region of the gene encoding the Flp recombinase of *Saccharomyces cerevisiae* - a role for Arginine-191 in binding and ligation. Journal of Molecular Biology 225: 313–326.

47. Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organelar genomes with DOGMA. Bioinformatics 20: 3252–3255.

48. Hildebrand M, Corey DK, Ludwig JR, Kukel A, Feng TY, et al. (1993) Plasmids in diatom species. *Journal of Bacteriology* 173: 5924–5927.

49. Withers NW, Cox ER, Tomas R, Haxo FT (1977) Pigments of the dinoflagellate *Glenodinium foliacum* and its photosynthetic endosymbiont. *Journal of Molecular Biology* 13: 354–358.

50. Jacobs JD, Ludwig JR, Hildebrand M, Kukel A, Feng TY, et al. (1992) Characterization of 2 circular plasmids from the marine diatom *Cylindrotheca fusiformis* - Plasmids hybridize to chloroplast and nuclear-DNA. *Molecular & General Genetics* 267: 179–186.

51. Kurtz S, Xerc and Xerd with the recombination site dif. Nucleic Acids Research 22: 5613–5620.

52. Kite GC, Rothschild IJ, Dodge JD (1968) Nuclear and plastid DNAs from the binucleate dinoflagellates *Glenodinium (Peridinium) foliacum* and *Peridinium balticum*. Biosystems 21: 151–163.

53. Barre FX, Sohalle B, Michel B, Aroyo M, Robertson M, et al. (2001) Circles: The replication-recombination-chromosome segregation connection. Proceedings of the National Academy of Sciences of the United States of America 98: 8189–8195.

54. Blakely GW, Sherratt DJ (1994) Interactions of the site-specific recombinases XerC and XerD with the recombination site dif. Nucleic Acids Research 22: 5613–5620.

55. Lesterlin C, Barre FX, Cornet F (2004) Genetic recombination and the cell cycle: what we have learned from chromosome dimers. Molecular Microbiology 54: 1151–1160.

56. Mandelli EF (1968) Carotenoid pigments of dinoflagellate *Glenodinium foliacum* Stein. Journal of Phycology 4: 347–46.

57. Schnepf E, Ehlbrechter M (1999) Dinoflagelate chloroplasts and phylogeny - A review. Grana 38: 81–97.

58. Jeffrey SW, Vesk M (1976) Further evidence for a membrane bound endosymbiont within the dinoflagellate *Peridinium foliacum*. Journal of Phycology 12: 450–455.

59. Oudot-Le Secq MP, Grimwood J, Shapiro H, Armbrust EV, Bowler C, et al. (2009) Characterization of 2 circular plasmids from the marine diatom *Cylindrotheca fusiformis* - Plasmids hybridize to chloroplast and nuclear-DNA. *Molecular & General Genetics* 267: 179–186.

60. Rice P, Longden I, Bleasby A (2000) EMBOSS: The European molecular biology open software suite. Trends in Genetics 16: 276–277.

61. Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Research* 33: W686–W689.

62. Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, et al. (2000) PipMaker - A web server for aligning two genomic DNA sequences. *Genome Research* 10: 4677–4584.

63. Kurtz S, Xerc and Xerd with the recombination site dif. Nucleic Acids Research 22: 5613–5620.

64. Conant GC, Wolfe KH (2008) GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics* 24: 961–962.