A hypervariable mitochondrial protein coding sequence associated with geographical origin in a cosmopolitan bloom-forming alga, *Heterosigma akashiwo*

Aiko Higashi, Satoshi Nagai, Sergio Seone and Shoko Ueki

1 Institute of Plant Science and Resources, Okayama University, 2-20-1 Chuo, Kurashiki, Okayama 710-0046, Japan
2 National Research Institute of Fisheries Science, 2-12-4 Fukuura, Kanazawa-ku, Yokohama, Kanagawa 236-8648, Japan
3 Department of Plant Biology and Ecology, University of the Basque Country, 48940 Leioa, Spain

Received: 23 December 2016
Accepted: 21 March 2017

**Subject Areas:**
ecology, molecular biology, evolution

**Keywords:**
*Heterosigma akashiwo*, mitochondrial DNA, hypervariable regions, phytoplankton

**Author for correspondence:**
Shoko Ueki
e-mail: shokoueki@okayama-u.ac.jp

Geographical distributions of phytoplankton species can be defined by events on both evolutionary time and shorter scales, e.g. recent climate changes. Additionally, modern industrial activity, including the transport of live fish and spat for aquaculture and aquatic microorganisms in ship ballast water, may aid the spread of phytoplankton. Obtaining a reliable marker is key to gaining insight into the phylogeographic history of a species. Here, we report a hypervariable mitochondrial gene in the cosmopolitan bloom-forming alga, *Heterosigma akashiwo*. We compared the entire mitochondrial genome sequences of seven *H. akashiwo* strains from Japanese and North American coastal waters and identified a hypervariable segment. The region codes for a hypothetical protein with no defined function, and its variations between Japanese and North American isolates were prominent, while the sequences were more conserved among Japanese strains and North American isolates. Comparison of the sequence in isolates obtained from different geographical points in the Northern Hemisphere revealed that the sequence variations largely correlated with latitude and longitude (i.e. Pacific/Atlantic oceans). Our results demonstrate the usefulness of the sequence in determining the phylogeographic history of *H. akashiwo*

1. **Introduction**

Phytoplankton species are primary producers, providing a foundation for aquatic food webs. On the other hand, some noxious algae form harmful algal blooms (HAB) that negatively affect ecosystems [1–3]. Therefore, phytoplankton distribution and dynamics have pivotal effects on the environment, and are thus of great interest to marine ecologists.

*Heterosigma akashiwo* is a photosynthetic, eukaryotic, HAB-causing alga that belongs to class *Raphidophyceae*, and is regarded as an ichthyotoxid species, having caused significant damage to mariculture [4–9]. While originally regarded as a temperate species, recent studies have indicated that it also inhabits arctic and tropical areas, including the Pacific Rim area, Oceania, and the North and the South Atlantic oceans [10]. Recent identification of the alga over a wider area may be merely the result of exhaustive surveillance. Alternatively, the species may have been recently introduced to these areas. Recent global climate changes, including temperature changes and ocean stream shifts, may
have resulted in the short- and long-distance dispersal of HAB-causing species [1,11,12]. Additionally, human-assisted dispersions, typically by ship ballast water [13–15] and the commercial transfer of spat and live fish [16], can further alter HAB species distribution. Several phylogeographic markers have been developed and successfully used to understand geographical dynamics of phytoplankton [17–19]. Here, we identified a hypervariable segment of *H. akashiwo* mitochondrial DNA (mtDNA), and demonstrated its associations with isolate origins. This information lays the foundation for elucidating the phylogeographic history of *H. akashiwo*.

2. Material and methods

(a) Algal strains and sequences

The sources of strains and culture conditions are provided in the electronic supplementary material, text. Cells were collected from 30 ml culture of each strain by centrifuging at 5000 *g* for 5 min, and total DNA was extracted from the cells using the DNeasy Blood & Tissue Kit (Qiagen). The TaKaRa LA Taq polymerase and primers listed in figure 1 and table 1 were used for DNA amplification. After amplification by PCR (cycle condition: 3 min at 95°C, 35 cycles of 15 s at 95°C, 30 s at 62°C, and 4 min at 72°C, and a final elongation of 10 min), the reaction mixtures were treated with Illustra ExoProStar (GE Healthcare), and were subjected to sequence analysis using BigDye Terminator ver3 (Applied Biosystems) according to the manufacturer’s instructions.

(b) Sequence analysis

Whole *H. akashiwo* strain mtDNA sequences available to date were downloaded from the NCBI database (table 2). The open reading frames (ORFs) were predicted by ORFfinder software (https://www.ncbi.nlm.nih.gov/orffinder/), and each ORF was compared with the genes in strain Y by BLASTN in BLAST+ v. 2.2.31 for identification. The ORFs and intergenic sequences were extracted from each genome, and sets of corresponding sequences were aligned using MUSCLE v. 3.8.31 [20,21]. Alignment and phylogenetic reconstructions were performed using the ‘build’ function in ETE 3 v. 3.0.0b32 [22] as implemented on GenomeNet (http://www.genome.jp/tools/ete/). Briefly, the MtORFvar sequences corresponding to 4442–5884 nt in H193616 were aligned using MAFFT v. 6.861b [23], and a distance-based tree was inferred by the most likelihood method using RAxML v. 8.1.20 run with model GTRGAMMA and default parameters [24]. The association of MtORFvar to geographical origin was evaluated by a nonparametric Mantel test [25]. The geographical distances were estimated in kilometres using GOOGLE EARTH 4.3, and averages of distances of the sampling points in Groups 1–4 less the origins of the exceptional strains (figure 2a) were adopted as distances between groups. The genetic distances were calculated based on the MAFFT alignment of all seventeen MtORFvar.

Figure 1. Circular representation of *H. akashiwo* mtDNA from different strains. The open reading frames (thick arcs) and tRNAs (short black bars) outside and inside are coded by positive and reverse strands, respectively. Genes with different predicted functions are colour-coded as follows: respiratory genes, orange; ribosomal subunit proteins, pink; secY-independent transporter (secY), red; ribosomal RNA subunits, blue; hypothetical proteins coded by strain Y, green; orphans, grey. The gene name abbreviations are as follows: ribosomal RNA large subunit, rRNA-L; ribosomal RNA small subunit, rRNA-S; cytochrome c oxidase subunit, *cox*; NADH dehydrogenase subunit, *nad*; ATP synthase F0 subunit, *atp*; ribosomal protein small subunit, *rbs*; cytochrome, *cycb*; ribosomal protein large subunit, *rbl*. Scales for nucleotide positions are represented as grids and are numbered clockwise. The approximate positions of primers for amplification (red triangles) and sequencing (blue triangles) are shown with their names (table 1).
3. Results and discussion

We conducted multiple sequence alignments of the mtDNA genomes to gain strain-specific sequence information. We recently completed the sequencing of the full-length mtDNA of four *H. akashiwo* strains, H93616, Ha00_17, HaTj01 and NEPCC522 [27] (figure 1 and table 2). For systematic comparison, we identified the ORFs on the mtDNA of the four newly sequenced strains and reannotated the previously sequenced strains, CCMP452 and NIES293 [28], based on the strain Y annotations [29]. The sequences and configurations of the ORFs with defined functions were well preserved in all of the strains, except for the absence of *rbs4* in CCMP452 (figure 1). Some ORFs with undefined functions were absent from some of the strains analysed (figure 1, grey). Despite differences in the coding capacities of these segments, their sequence variations were nominal, exhibiting a few single nucleotide polymorphisms. In contrast, we found major sequence variations in the ORFs coding a hypothetical protein preserved in all of the strains, except for strain Y (figure 1, MtORFvar): strain Y possesses two ORFs in the corresponding segment because of a single nucleotide deletion (electronic supplementary material, figure S1). Several single nucleotide substitutions and indels were observed in these segments (electronic supplementary material, figure S1). Importantly, differences between strains obtained from North American and Japanese coastal waters were particularly prominent. The sequence variations were rather conserved among the Japanese strains and between the North American isolates. We named the hypervariable ORF MtORFvar based on the feature. MtORFvar did not exhibit any homology to genes from other species (TBLASTN, E-value, $10^{-5}$), and did not contain any identifiable motif; thus, the function of the MtORFvar encoded protein is unclear.

To analyse the MtORFvar variations in *H. akashiwo* obtained from different geographical regions, we designed primers to amplify and sequence MtORFvar (figure 1 and table 1). The sequences preserved in all mtDNAs analysed that flanked MtORFvar were chosen for amplification primers, and reactions using the set were confirmed to yield no detectable off-target product (figure 1 and table 1). Additionally, sequences on both sides of and one inside MtORFvar were chosen to design the primers for sequencing (figure 1 and table 1). The PCR amplifications and sequencing of the product using the primer set designed were successfully conducted using DNA prepared from the strains from different geographical origins (table 2, strains 8–18). The regions corresponding to 4330–5800 nt on strain Y mtDNA were covered by the sequencing results using the primers.

### Table 1. Primers designed for mtDNA strain-specific markers.

<table>
<thead>
<tr>
<th>name for amplification</th>
<th>sequence</th>
<th>length</th>
<th>GC%</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mt_2043 F</td>
<td>GAGGCCTGACGGAAGGTTG</td>
<td>20</td>
<td>55</td>
<td>61.84</td>
</tr>
<tr>
<td>Mt_7473 R</td>
<td>GCTGACGAGAATCCGCAAC</td>
<td>20</td>
<td>55</td>
<td>66.98</td>
</tr>
<tr>
<td>name for sequencing</td>
<td>sequence</td>
<td>length</td>
<td>GC%</td>
<td>Tm (°C)</td>
</tr>
<tr>
<td>Mt_4283 F</td>
<td>GTCAACATCATTGAGGGTTG</td>
<td>21</td>
<td>43</td>
<td>63.95</td>
</tr>
<tr>
<td>Mt_5479 R</td>
<td>CGCTGATTGTCCCTTACCTTG</td>
<td>23</td>
<td>43</td>
<td>66.76</td>
</tr>
<tr>
<td>Mt_6057 R</td>
<td>AAAGCCTGATATAGGTTTGTAACTCTTG</td>
<td>26</td>
<td>31</td>
<td>60.46</td>
</tr>
</tbody>
</table>

Figure 2. (a) Phylogenetic analysis of MtORFvar collected from different regions in the Northern Hemisphere. Branch supports computed out of 100 bootstrapped trees larger than 70 are indicated at each node. The numbers on the branch labels correspond to table 2. The branches labelled with asterisks and ampersands were the exceptions of the Atlantic/Pacific and latitude categorizations, respectively. (b) Isolation by distance plot for the Group 1–4 populations based on the data deposited in Dryad Digital Repository [26].
Importantly, phylogenetic analysis of MtORFvar isolates from different origins revealed that the sequence variations correlated with their geographical origins (figure 2). According to the analysis, the sequences clustered to two groups, with a few exceptions, according to the latitude at which the strains originated. Moreover, strains that originated in high-latitude regions segregated into two groups, one with strains that originated in Atlantic regions, and the other with strains from Pacific regions. Association of MtORFvar sequence variations and their geographical origins was confirmed by nonparametric Mantel test analysis (figure 2b). There are three isolates, two from Florida and one from Rhode Island, United States, that were not categorized in the expected groups. The results may be due to incorrect prediction of their origin by the marker. Alternatively, categorization of the strains into unexpected groups may be because some of the strains established in the area might have been transferred from the geographical area of the group. This hypothesis is supported by the fact that the isolates from Florida and Rhode Island, United States, exhibited two different haplotypes, suggesting that *H. akashiwo*, at least in these areas, consists of a mixed population. The phylogeographic history of the isolates in these areas should be further tested by incorporating other haplotype markers for future study.

<table>
<thead>
<tr>
<th>no.</th>
<th>strains</th>
<th>location</th>
<th>accessions</th>
<th>full genome</th>
<th>MtORFvar</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>H93616</td>
<td>Uranouchi Bay, Kochi, Japan</td>
<td>KU561547</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Ha00_17</td>
<td>Fukuoka Bay, Fukuoka, Japan</td>
<td>KU561548</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>HaTj 01</td>
<td>Tajiri Bay, Hiroshima, Japan</td>
<td>KU561550</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Y</td>
<td>Bingo-nada, Hiroshima, Japan</td>
<td>NC_016738</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>NIES293</td>
<td>Onagwa Bay, Miyagi, Japan</td>
<td>GQ222227</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>NEPCC522</td>
<td>British Columbia, Canada</td>
<td>KU726247</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>CCMP452</td>
<td>Long Island, NY, USA</td>
<td>GQ222228.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>CMS_Haek9505-1</td>
<td>Tampa Bay, Florida, USA</td>
<td>LC202896</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>CMS_Haek9806-1</td>
<td>Tampa Bay, Florida, USA</td>
<td>LC202897</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>CCAP934/7</td>
<td>Puget Sound, Washington, USA</td>
<td>LC202894</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>CCAP934/9</td>
<td>Puget Sound, Washington, USA</td>
<td>LC202895</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>CCMP1596</td>
<td>Narragansett Bay, Rhode Island USA</td>
<td>LC202888</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>CCMP1870</td>
<td>Los Angeles River, California, USA</td>
<td>LC202889</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>CCMP2274</td>
<td>San Francisco Bay, California, USA</td>
<td>LC202890</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>CCMP3374</td>
<td>Greenwich Cove, Rhode Island, USA</td>
<td>LC202891</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>EHU_RP01</td>
<td>Bay of Biscay, Spain</td>
<td>LC202892</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>EHU_RP02</td>
<td>Bay of Biscay, Spain</td>
<td>LC202893</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These results also imply that MtORFvar may play important roles in modulating *H. akashiwo* mitochondrial functions to adapt to different conditions depending on latitude, presumably because of differences in water temperatures and day length. In this study, we found out that *H. akashiwo* MtORFvar was strongly correlated with strain origin. While the *H. akashiwo* microsatellite sequence information is useful as a molecular marker to identify strains, off-target amplification may hamper its usage [18]. In the case of mtDNA, the amplification step is straightforward because of the high mtDNA copy number. Therefore, mtDNA provides an important molecular tool to understand the phylogeographic history of *H. akashiwo*. The functional relevance of MtORFvar in *H. akashiwo* adaptation to different environments is also of interest and should be studied in more detail in the future.

**Data accessibility.** DNA sequences: GenBank accessions listed in table 2. The dataset for Mantel test for isolation by distance: Dryad: http://dx.doi.org/10.5061/dryad.42856 [29].

**Authors’ contributions.** S.U. conceived, designed, coordinated the study and drafted the manuscript. All the authors conducted molecular lab work and participated in writing the manuscript, and gave final approval for publication. All authors agreed to be held accountable for the content therein.

**Competing interests.** We have no competing interests.

**Funding.** S.U. was supported by KAKENHI (26291092) and Priority Areas ‘Comprehensive Genomics’ (no. 221S0002) provided by the Japan Society for the Promotion of Science.

**References**


Correction to ‘A hypervariable mitochondrial protein coding sequence associated with geographical origin in a cosmopolitan bloom-forming alga, Heterosigma akashiwo’

Aiko Higashi, Satoshi Nagai, Sergio Seoane and Shoko Ueki


The name of author ‘Sergio Seoane’ was misspelled in the original manuscript. We apologize for any inconvenience caused.