De novo assembly and inferred functional annotation of the transcriptome of *Heterosigma akashiwo*

Masanao Sato¹, Masahide Seki², Yutaka Suzuki², and Shoko Ueki³

¹Hokkaido University

²The University of Tokyo Graduate School of Frontier Sciences

³Okayama University

January 4, 2022

Abstract

Heterosigma akashiwo is a eukaryotic, cosmopolitan, and unicellular alga (class: Raphidophyceae), and produces fish-killing blooms. There is a substantial scientific and practical interest in its ecophysiological characteristics that determine bloom dynamics and its adaptation to broad climate zones. A well-annotated genomic/genetic sequence information enables researchers to characterize organisms using modern molecular technology. The Chloroplast and the mitochondrial genome sequences and transcriptome sequence assembly (TSA) datasets with limited sizes for *H. akashiwo* are available in NCBI nucleotide database on December 2021: there is no doubt that more genetic information of the species will greatly enhance the progress of biological characterization of the species. Here, we conducted *H. akashiwo* RNA sequencing, a *de novo* transcriptome assembly (NCBI TSA ICRV01) of a large number of high-quality short-read sequences, and the functional annotation of predicted genes. Based on our transcriptome, we confirmed that the organism possesses genes that were predicted to function in phagocytosis, supporting the earlier observations of *H. akashiwo* bacterivory. Along with its capability for photosynthesis, the mixotrophy of *H. akashiwo* may partially explain its high adaptability to various environmental conditions. Our study here will provide an important toehold to decipher *H. akashiwo* ecophysiology at a molecular level.

Introduction

Heterosigma akashiwo is a eukaryotic, unicellular, bloom-forming alga that belongs to class *Raphidophyceae* (1, 2). The organism often forms harmful algal bloom (HAB) during the summer and exerts a negative impact on the local ecosystem and industries, particularly fisheries (1, 3-12). Because of their potential damage to the environment and the economy of an area, the behavior of *H. akashiwo* in the environment is of great interest and importance (13-15).

The species was once regarded as a temperate species. However, recent studies revealed that the species were also identified in the thalassic sea of both the northern and southern hemisphere, from the arctic to the tropic area, including the Pacific Rim area, Oceania, and the North and the South Atlantic oceans (1, 3-12, 16). The distribution of the organism over such a wide area is most probably due to its high adaptability: The adaptabilities to a different temperature (17-19), carbon dioxide (17), salinity (19, 20), cyst formation to survive conditions unfavorable for vegetative propagation (21-23), and resistance to turbulence (24) have been described for the organism. While the toxicity of *H. akashiwo* has been debated over years, the identity of the toxic substance has not been identified (18, 25-30). In addition, the possibility of bacterivory by the organism, that may enable the species to utilize both auto- and auxotrophic lifestyles has been reported (31-33). The characterization of the mechanisms underlying such biological features may lead to a deeper understanding of the biological flexibilities of *H. akashiwo*.

A collection of well-annotated genomic/genetic sequences enables researchers to characterize organisms

using modern molecular technology. However, as of October 2021, the complete mitochondrial (34) and chloroplast (35) genome sequences are the only two *H. akashiwo* nucleotide sequences available in the NCBI RefSeq database. A lack of nuclear genomic sequence information is a major impediment to investigating the molecular mechanisms underlying *H. akashiwo* physiological processes. Because long intergenic regions, introns, and repetitive sequences in eukaryotic genomes hinder the construction of a high-quality reference genome sequence, building reference gene models using the transcriptome is an efficient alternative. Five datasets of *H. akashiwo* transcriptome sequence assembly with limited data sizes without detailed annotations were deposited in a public database by February 2021 and were referenced for the past studies (36, 37). In this study, we conducted *de novo* transcriptome assembly (NCBI TSA ICRV01, <https://www.ncbi.nlm.nih.gov/Traces/wgs/ICRV01>), gene prediction, and extensive functional annotation for the genes. Further, utilizing our transcriptome, we explored the potential for phagotrophy in *H. akashiwo* using the gene-based trophic model from Burnset *al* (38). The organism possesses the set of genes required for phagocytosis activity, supporting earlier observation of bacterivory (31-33). The novel transcriptome will serve as an important information infrastructure to dissect *H. akashiwo* physiology at the molecular level.

Material and Methods

Heterosigma akashiwo culture conditions

Bacteria-free *H. akashiwo* strain H93616 [isolated from Uranouchi Bay, Kochi Prefecture, Japan and then rendered bacteria-free (39)] was maintained in filter-sterilized Daigo's IMK medium (Nihon Pharmaceutical Co., Ltd.) supplemented with penicillin (100 units/mL), streptomycin (100 µg/mL), ampicillin (100 µg/mL), and kanamycin (60 µg/mL). The culture was maintained in a controlled-environment chamber set at 21 °C with a 12-h light (100 µmol m⁻²s⁻¹)/12-h dark cycle. The absence of contaminating bacteria was confirmed by DAPI staining and a microscopic examination

For the viral infection experiments, actively growing *H. akashiwo* cells (5×10^4 cells/mL) in a 300-mL culture were inoculated with *Heterosigma akashiwo virus* strain 53 (*HaV* 53) at a multiplicity of infection of 5. The cells in 12.5-mL cultures were filter-trapped at 30, 90, 120, and 420 min post-inoculation on a polycarbonate membrane (0.8 µm pores) and then flash-frozen in liquid N₂. For the non-infected cells, the same number of non-inoculated *H. akashiwo* cells were filter-trapped and flash-frozen in liquid N₂.

Other growth and harvest conditions in this study were as follows: UV irradiation for 5 min and then an incubation under normal conditions for 7 h; low-temperature treatment at 16 °C for 7 h; exposure to high-intensity lights (300 µmol m⁻²s⁻¹ white fluorescent light) for 7 h; and harvesting at various times during the day (9:00, 12:00, 15:00, and 18:00). The cells (2×10^5 cells/mL) in a 5-mL growth medium subjected to the various treatments were collected and then flash frozen. All samples were stored at -80 degC until the RNA extraction.

RNA purification

Total RNA was isolated using the PureLink RNA Purification kit (Ambion) according to the manufacturer's protocol with the following modification: the cell extract was treated twice with chloroform-isoamyl alcohol before the RNA purification step using a column to remove impurities. The total RNA was eluted in RNase-free water and the concentration was determined using the QuantiFluor RNA system.

Library construction

To construct the multi-treatment sequencing library, equal amounts of RNA extracted from UV-irradiated cells, low-temperature-treated cells, high-light-irradiated cells, and the cells harvested at different times during the day were pooled. After verifying the RNA quality using the Agilent RNA 6000 Nano Kit, mRNA-seq libraries were constructed according to the Agilent SureSelect Strand Specific RNA Prep Kit protocol. To prepare total RNA libraries aiming to sequence 5' regions of mRNA and to target non-poly (A) RNA, the rRNA was depleted using the RiboZero Plant rRNA Removal kit (Illumina), after which the libraries were prepared according to the Agilent SureSelect Strand Specific RNA Prep Kit protocol without

the polyA selection. The libraries were quantified using the Agilent DNA 7500 kit and then sequenced on the Illumina HiSeq 2500 platform.

Assembly and annotation pipeline

The assembly and annotation pipeline is presented in Fig. 1. The paired-end reads in the fastq files for the sequenced libraries (Table 1) were individually deduplicated using the filter PCRdupl (<https://github.com/linneas/condetri/blob/master/filterPCRdupl.pl>) (40) by comparing the initial 100 nt of the reads. The deduplicated reads in the fastq files were trimmed and filtered using ConDeTri (version 2.3) (<https://github.com/linneas/condetri/blob/master/condetri.pl>) (40), with 30 and 28 set as the high- and low-quality thresholds, respectively, and with [?] 80% of the reads exceeding the high-quality threshold, using Illumina scoring table 33. The preprocessed reads in the fastq files were concatenated. For the *de novo* transcriptome assembly, Trinity (version 2.3.2) (41) was used, with a minimum count for K-mers set at 3 to balance the sensitivity for reconstructing sequences expressed at low levels and for potential sequencing errors. Contigs longer than 300 nt were subsequently analyzed using Transdecoder (version 3.0.1) to predict coding sequences based on a standard translation table, and the inferred CDSs longer than 150 bp were analyzed further.

Mapping reads from the assembled data and publicly available reads in the Short Read Archives (SRAs)

To validate the quality of the assembly, the reads in the fastq files used for the assembly were mapped to our contigs using the Burrows-Wheeler Aligner (version 0.1.17-r1188) (42). Similarly, the reads from various *H. akashiwo* transcriptome or meta-transcriptome analyses downloaded from Marine Microbial Eukaryote Transcriptome Sequencing Project SRA databases (Table 1) were preprocessed with Trimmomatic (43) and then mapped to our transcriptome assembly.

Annotation, GO analyses, and Trophic mode prediction

For functional annotations, gene ontology (GO) terms were assigned to the predicted peptides in a two-step process. First, the best-match homologs of the *H. akashiwo* peptides were identified following a BLASTP search (E-value < 1) of a custom database composed of RefSeq gene models of *Arabidopsis thaliana*, *Homo sapiens*, *Mus musculus*, and *Saccharomyces cerevisiae* (S288C). Second, the *H. akashiwo* peptides were annotated with the GO terms (<http://geneontology.org>) assigned to their best-match homologs. The Pfam database (<http://pfam.xfam.org>) and hmmer (44) were used to predict the domains in the *H. akashiwo* gene models. The phyletic distribution of *H. akashiwo* gene homologs was analyzed by a BLASTP search of the NCBI non-redundant (nr) database using CDSs and a custom script for counting the taxonomies of the source organisms of the best-hit sequences. To predict the trophic mode of *H. akashiwo*, the predicted protein sequences based on the transcriptome were tested for phagocytotic and prototrophic potentials using a gene-based Trophic Mode Prediction tool using default mode (38, 45). The gene models of a non-photosynthetic ciliate, *Paramecium tetraurelia* (NCBI BioProject PRJNA19409) as well as photosynthetic *Emiliania huxleyi* (NCBI AHAL000000000.1) and *Chlamydomonas reinhardtii* (<https://www.uniprot.org/uniprot/?query=proteome:UP000006906>), and conditionally symbiotic and photosynthetic *Breviolum minutum* (Previously termed *Symbiodinium minutum*, https://marinegenomics.oist.jp/symb/viewer/download?project_id=21, (46)), were also analyzed.

Results and Discussions

In this study, we attempted to obtain a comprehensive catalog of gene models that serves as an alternative for a reference genome sequence by compiling transcriptome of *H. akashiwo* under various experimental conditions. To generate reads unbiased for 3'-partial sequences, libraries prepared from both ribosomal RNA (rRNA)-depleted RNA and polyA-selected RNA were adopted for the study. More than 84 million high-quality reads (Table 2) included in 16 independent datasets (Table 1) were assembled through the pipeline presented in Fig. 1. To evaluate the quality of the assembly, the reads from our data files and those of 31 experiments involving *H. akashiwo* transcriptome or meta-transcriptome analyses in the NCBI Short Read Archives (SRA) database were mapped to the obtained contigs (Table 1). More than 90% of the reads

used for this study were mapped to the transcriptome, and among them, more than 89% of paired reads were properly mapped to the transcriptome. The mapping ratios of the read files obtained from the public SRA, that were derived from the library prepared from different strains of *H. akashiwo*, ranged between 67.5 to 97.5%. In these cases, too, more than 92.3% of the mapped reads were properly paired. These high mapping rates corroborate the quality of the assembly conducted in this study.

Next, the protein-coding sequences (CDSs) were predicted according to the standard translation table using Transdecoder. The number of the contigs and the CDSs obtained in this study (Table 2) were greater than those of the previously published *H. akashiwo* TSAs (Table 3), showing that the current result may be more exhaustive. The high mapping rates of the read files obtained from this study and public SRA to our *H. akashiwo* transcriptome (Table 1) demonstrated that our gene models were widely useful and suggest that the coverage of the *H. akashiwo* gene models obtained in this study to the complete gene set in the *H. akashiwo* genome is likely to be very high.

The homology-based analysis involving a BLASTP search of the RefSeq database revealed the phyletic distribution of *H. akashiwo* gene homologs (Fig. 2a). More than 86% of the *H. akashiwo* ORFs were most homologous to eukaryotic genes in the nr database (Fig. 2a). Additionally, 56% of the ORFs that matched a sequence in the nr database were most homologous to stramenopile genes (Fig. 2b). It is also noteworthy that, fourteen libraries used in this study were constructed from RNA purified from the *HaV53* -infected *H. akashiwo* (Table 1). Although only 0.18% of the total reads were mapped to the *HaV53* genome (Table 1), the resultant contigs containing 110 full-length and 69 partial open reading frames (ORFs) (out of 246 predicted *HaV53* ORFs) were assembled with 100% identity, also suggesting that the host genes were also assembled with a similar high quality. Approximately 13% of the predicted ORFs were homologous to prokaryotic genes (Fig. 2a and b). Because rRNA-depleted RNA was analyzed, bacterial mRNA derived from an undetectable amount of symbiotic or contaminating bacteria in the *H. akashiwo* culture may have been included in the analysis, which may have resulted in contigs. It is technically impossible to determine whether the sequences are *bona fide* *H. akashiwo* transcripts without a high-quality genome sequence: This point has to be clarified by the *H. akashiwo* genome assembly in the future.

To functionally characterize the *H. akashiwo* transcriptome, the gene models in the contigs were annotated according to GO and protein domains. Because there is no well-established model organism for stramenopiles, we initially attempted to annotate *H. akashiwo* using the gene models of the unicellular green alga model, *Chlamydomonas reinhardtii*. However, in the case of many *H. akashiwo* gene models, more the best-hit sequences were obtained from other organisms, such as human and mouse, rather than *C. reinhardtii*. Therefore, we assigned GO terms to *H. akashiwo* gene models based on human, mouse, *A. thaliana*, and yeast genes, for which there is a considerable abundance of information regarding their functions (Fig. 3, Table 3). A BLASTP search (E-value < 1) indicated that only 22% of the *H. akashiwo* gene models matched genes in the four organisms (Fig. 3a). When the *C. reinhardtii* gene models were similarly analyzed, 92% of the genes matched the ones in the reference organisms (Fig. 3b). These results imply that the *H. akashiwo* genes extensively diversified after the common ancestor or are unique. To gain more information regarding the gene functions, the *H. akashiwo* gene models were also annotated according to the information resulting from a protein domain search conducted using the Pfam database. Consequently, the GO and protein domain analyses functionally annotated 56% of the *H. akashiwo* gene models. The most frequently detected Pfam domains and assigned GO terms in the *H. akashiwo* transcriptome are listed in Tables 4 and 5. The transcriptome and the annotation information were deposited to NCBI database as TSA (ICRV01, ICRV01000001-ICRV01144777, <https://www.ncbi.nlm.nih.gov/Traces/wgs/ICRV01>). Because the gene models were predicted solely based on *in silico* approach without any experimental confirmation, all the gene models are deposited as hypothetical proteins: particularly, the information on BLAST-matched homologs and the GO terms for the gene models are not included in the NCBI TSA at this point because of the NCBI data registration policy. To share the information obtained from this study with the public, all the blast matches, pfam matches, and GO terms for the transcriptome was deposited in Dryad as <https://datadryad.org/stash/share/g1DLXFFEkJmo3Nel3SyBdUbJZAB-Lr-0yY91l0UrM>.

To gain more insight into *H. akashiwo* ecophysiology, we analyzed its potential for different trophic modes. Earlier studies revealed the possibility of bacterivory activity of *H. akashiwo* (31-33), while the ecophysiological feature of the species has been rarely discussed in depth since then. A recent study predicted the bacterivory potentials of several green algae based on their gene sets and confirmed the activity experimentally (45), suggesting that the usage of both photosynthesis and phagocytosis for survival may be more common than previously considered. Phagocytotic and photosynthetic capabilities of the organisms were computationally predicted using the predictTrophicMode tool using default mode (38, 45). We analyzed the gene models predicted based on the transcriptomes of various algal species presented in Table 3 and our *H. akashiwo* transcriptome using the package for their capabilities for being phagocytotic and photosynthetic (Fig. 4ab). *C. reinhardtii* was one of the training data used for the prediction tool, and the transcriptome of a non-photosynthetic protist *P. tetraurelia* that belongs to the ciliate phylum was also analyzed as a reference. Among the five organisms analyzed, *C. reinhardtii*, *E. huxley*, and *H. akashiwo* were predicted to be photosynthetic, while *P. tetraurelia* was predicted to be non-photosynthetic at > 95% probabilities. On the other hand, *P. tetraurelia*, *B. minutum*, and *H. akashiwo* were predicted to be phagocytotic at > 95% probabilities. Based on the results, *H. akashiwo* is presumably mixotrophic, supporting the previous observation of its bacterivory activity (31-33).

This study provides the foundation for future molecular-level investigations of the bloom-forming species and will contribute to the elucidation of the genetic features of unexploited stramenopiles. More thorough analyses, e.g., full-length transcriptome by long-read sequencing data using third-generation sequencers and the empirical validation of the gene models by phycologists, would refine and supplement the available information regarding the *H. akashiwo* transcriptome. As demonstrated potential of the our *H. akashiwo* gene models with inference of *H. akashiwo* mixotrophy in this study, the current *H. akashiwo* gene models and continued efforts to refine them will be essential to resolve some controversial biological features of the organism, such as production of toxin and its identity (18, 25-30).

Figure legends

Fig. 1 Analysis pipeline including read assembly, gene prediction, and annotation.

Fig. 2 Phyletic distribution of *H. akashiwo* gene homologs at the (a) empire and (b) subkingdom/phylum levels. The best-hit homologs in the NCBI nr database for each *H. akashiwo* gene determined by a BLASTP search (E-value < 1) were used to identify the source organisms. The ratio of the number of source organisms at each taxonomic level is indicated.

Fig. 3 The ratios of the *H. akashiwo* (a) and *C. reinhardtii* (b) gene models that best-hit to a protein of the four model organisms, human, mouse, *Arabidopsis thaliana*, and yeast at E-value [?] 1, and the ratios of genes that best-hit to either of four model organisms to the total gene models in *H. akashiwo* (c) and *C. reinhardtii* (d) are presented. Note that in *H. akashiwo*, a much smaller fraction of the total gene models were homologous to the genes of the four model organisms than in *C. reinhardtii*.

Fig. 4 Plots of the first two principal components (PC1 and PC2) of the GO category scores for each organism: The 37-GO-category, 243-protein photosynthesis model (a), and the 86-GO-category, 474-protein model defined by free-living phagocytes (b). The solid lines represent a 95% confidence ellipse around the photosynthetic (a) and phagocytotic (b) training sets and gray dashed lines represent a 95% confidence ellipse around the non-photosynthetic (a) and non-phagocytotic (b, d) training sets, respectively.

Data accessibilities

Read files: NCBI SRA SAMD00092771 - SAMD00092778, SAMD00092788, SAMD00092812, SAMD00092814 - SAMD00092818, Dryad

The transcriptome : NCBI TSA (ICRV01, ICRV01000001-ICRV01144777, <https://www.ncbi.nlm.nih.gov/Traces/wgs/ICRV01>)

The blast matches, pfam matches, and GO terms for the transcriptome:

<https://doi.org/10.5061/dryad.m0cfxpp56>

Competing interests

This study was supported by funding sources cited in the *Acknowledgments*. The authors have no conflicts of interest.

Author Contributions

Masanao Sato: Conceptualization (equal); data curation (lead); formal analysis (lead); funding acquisition (support); investigation (equal); methodology (equal); software (equal); supervision (support); validation (equal); visualization (equal); writing-original draft (support); writing-review & editing (equal).

Masahide Seki: data curation (equal); formal analysis (equal); investigation (support); methodology (equal); software (equal); validation (equal); visualization (equal); writing-review & editing (support).

Yutaka Suzuki: data curation (support); formal analysis (support); investigation (support); methodology (equal); ; project administration (support); software (equal); supervision (support); validation (equal); visualization (equal); writing-review & editing (support).

Shoko Ueki: Conceptualization (equal); data curation (support); formal analysis (support); funding acquisition (lead); investigation (equal); methodology (support); project administration (lead); resources (lead); software (support); supervision (lead); validation (equal); visualization (equal); writing-original draft (lead); writing-review & editing (equal).

Acknowledgments

We thank the Data Integration and Analysis Facility, National Institute for Basic Biology, for the access to high-performance computers. This work was supported by grants from KAKENHI (16H06449 and 221S0002) provided by the Japan Society for the Promotion of Science, the Science and Technology Research Partnership for Sustainable Development (No. 989459) provided by the Japan Science and Technology Agency and the Japan International Cooperation Agency, the NAGASE Science Technology Foundation, and the Casio Science Promotion Foundation to S.U., and the Joint Usage/Research Center, Institute of Plant Science and Resources, Okayama University to M.S. and S.U.

References

1. Honjo T. Overview on bloom dynamics and physiological ecology of *Heterosigma akashiwo*. In: Smayda TJ, Shimizu Y, editors. *Toxic Phytoplankton Blooms in the Sea*. Amsterdam: Elsevier; 1993. p. 33-41.
2. Smayda TJ. Harmful algal blooms: Their ecophysiology and general relevance to phytoplankton blooms in the sea. *Limnology and Oceanography*. 1997;42(5):1137-53.
3. Chang FH, Anderson C, Boustead NC. 1st Record of a *Heterosigma* (Raphidophyceae) Bloom with Associated Mortality of Cage-Reared Salmon in Big Glory Bay, New-Zealand. *New Zealand Journal of Marine and Freshwater Research*. 1990;24(4):461-9.
4. Mackenzie L. Toxic and Noxious Phytoplankton in Big Glory Bay, Stewart-Island, New-Zealand. *Journal of Applied Phycology*. 1991;3(1):19-34.
5. Black EA, Whyte JNC, Bagshaw JW, Ginther NG. The Effects of *Heterosigma-Akashiwo* on Juvenile *Oncorhynchus-Tshawytscha* and Its Implications for Fish Culture. *Journal of Applied Ichthyology-Zeitschrift Fur Angewandte Ichthyologie*. 1991;7(3):168-75.
6. Rensel J, Homer RA, Postel LE. Effects of phytoplankton blooms on salmon aquaculture in Puget Sound, Washington: initial research. *Northwest Environ J* 1989;5:53-9.

7. Taylor FJR. Current problems with harmful phytoplankton blooms in British Colombia waters. In: Smayda TJ, Shimizu Y, editors. Toxic Phytoplankton Blooms in the Sea. Amsterdam: Elsevier; 1993. p. 699-703.
8. O'Halloran C, Silver MW, Holman TR, Scholin CA. *Heterosigma akashiwo* in central California waters. *Harmful Algae*. 2006;5(2):124-32.
9. Lackey JB, Lackey E. Microscopic algae and protozoa in the waters near Plymouth in August 1962. . *J Mar Biol Ass UK*. 1963;43:797-805.
10. Throndsen J. Flagellates of Norwegian coastal waters. *Nytt Magazin for Botanikk*. 1969;16:161–216.
11. Rojas de Mendiola B. Red tide along the Peruvian coast. In: Taylor DL, Seliger HH, editors. Toxic Dinoflagellate Bloom. Amsterdam: Elsevier; 1979. p. 183-90.
12. Tseng CK, Zhou MJ, Zou JZ. Toxic phytoplankton studies in China. . In: Smayda TJ, Shimizu, Y. , editor. Toxic Phytoplankton Blooms in the Sea. Amsterdam: Elsevier; 1993. p. 347-52.
13. Fu FX, Tatters AO, Hutchins DA. Global change and the future of harmful algal blooms in the ocean. *Marine Ecology Progress Series*. 2012;470:207-33.
14. Hallegraeff G. A review of harmful algal blooms and their apparent global increase. *Phycologia*. 1993;32(2):79-99.
15. Maso M, Garcés E. Harmful microalgae blooms (HAB); problematic and conditions that induce them. *Mar Pollut Bull*. 2006;53(10-12):620-30.
16. Engesmo A, Eikrem W, Seoane S, Smith K, Edvardsen B, Hofgaard A, et al. New insights into the morphology and phylogeny of *Heterosigma akashiwo* (Raphidophyceae), with the description of *Heterosigma* minor sp. nov. *Phycologia*. 2016;55(3):279-94.
17. Fu FX, Zhang YH, Warner ME, Feng YY, Sun J, Hutchins DA. A comparison of future increased CO₂ and temperature effects on sympatric *Heterosigma akashiwo* and *Prorocentrum minimum*. *Harmful Algae*. 2008;7(1):76-90.
18. Ono K, Khan S, Onoue Y. Effects of temperature and light intensity on the growth and toxicity of *Heterosigma akashiwo* (Raphidophyceae). *Aquaculture Research*. 2000;31(5):427-33.
19. Martinez R, Orive E, Laza-Martinez A, Seoane S. Growth response of six strains of *Heterosigma akashiwo* to varying temperature, salinity and irradiance conditions. *Journal of Plankton Research*. 2010;32(4):529-38.
20. Gimpler H. Primary sodium plasma membrane ATPases in salt-tolerant algae: facts and fictions. *Journal of Experimental Botany*. 2000;51(348):1171-8.
21. Shikata T, Nagasoe S, Matsubara T, Yamasaki Y, Shimasaki Y, Oshima Y, et al. Effects of temperature and light on cyst germination and germinated cell survival of the noxious raphidophyte *Heterosigma akashiwo*. *Harmful Algae*. 2007;6(5):700-6.
22. Han MS, Kim YP, Cattolico RA. *Heterosigma akashiwo* (Raphidophyceae) resting cell formation in batch culture: Strain identity versus physiological response. *Journal of Phycology*. 2002;38(2):304-17.
23. Tobin ED, Grunbaum D, Cattolico RA. Cell motility, life stage transitions and cyst distribution of the harmful alga, *Heterosigma akashiwo*. *Integrative and Comparative Biology*. 2010;50:E175-E.
24. Sengupta A, Carrara F, Stocker R. Phytoplankton can actively diversify their migration strategy in response to turbulent cues. *Nature*. 2017;543(7646):555-8.
25. Mohamed ZA, Al-Shehri AM. The link between shrimp farm runoff and blooms of toxic *Heterosigma akashiwo* in Red Sea coastal waters. *Oceanologia*. 2012;54(2):287-309.

26. Pistocchi R, Guerrini F, Pezzolesi L, Riccardi M, Vanucci S, Ciminiello P, et al. Toxin Levels and Profiles in Microalgae from the North-Western Adriatic Sea-15 Years of Studies on Cultured Species. *Mar Drugs*. 2012;10(1):140-62.
27. Powers L, Creed IF, Trick CG. Sinking of *Heterosigma akashiwo* results in increased toxicity of this harmful algal bloom species. *Harmful Algae*. 2012;13:95-104.
28. Haque SM, Onoue Y. Effects of salinity on growth and toxin production of a noxious phytoflagellate, *Heterosigma akashiwo* (Raphidophyceae). *Botanica Marina*. 2002;45(4):356-63.
29. Twiner MJ, Dixon SJ, Trick CG. Extracellular organics from specific cultures of *Heterosigma akashiwo* (Raphidophyceae) irreversibly alter respiratory activity in mammalian cells. *Harmful Algae*. 2004;3(2):173-82.
30. Wang LP, Yan T, Zhou MJ. Impacts of HAB species *Heterosigma akashiwo* on early development of the scallop *Argopecten irradians* Lamarck. *Aquaculture*. 2006;255(1-4):374-83.
31. Seong KA, Jeong HJ, Kim S, Kim GH, Kang JH. Bacterivory by co-occurring red-tide algae, heterotrophic nanoflagellates, and ciliates on marine bacteria in the Korean waters. *Marine Ecology Progress Series*. 2006;322:85-97.
32. Jeong HJ, Seong KA, Kang NS, Yoo YD, Nam SW, Park JY, et al. Feeding by raphidophytes on the cyanobacterium *Synechococcus*. *Aquatic Microbial Ecology*. 2010;58:181-95.
33. Jeong HJ. Mixotrophy in red-tide algae raphidophytes. *Journal of Eukaryotic Microbiology*. 2011;58:215-22.
34. Masuda I, Kamikawa R, Ueda M, Oyama K, Yoshimatsu S, Inagaki Y, et al. Mitochondrial genomes from two red tide forming raphidophycean algae *Heterosigma akashiwo* and *Chattonella marina* var. *marina*. *Harmful Algae*. 2011;10(2):130-7.
35. Cattolico RA, Jacobs MA, Zhou Y, Chang J, Duplessis M, Lybrand T, et al. Chloroplast genome sequencing analysis of *Heterosigma akashiwo* CCMP452 (West Atlantic) and NIES293 (West Pacific) strains. *BMC Genomics*. 2008;9:211.
36. Haley ST, Alexander H, Juhl AR, Dyhrman ST. Transcriptional response of the harmful raphidophyte *Heterosigma akashiwo* to nitrate and phosphate stress. *Harmful Algae*. 2017;68:258-70.
37. Ji N, Lin L, Li L, Yu L, Zhang Y, Luo H, et al. Metatranscriptome analysis reveals environmental and diel regulation of a *Heterosigma akashiwo* (raphidophyceae) bloom. *Environ Microbiol*. 2018;20(3):1078-94.
38. Burns JA, Pittis AA, Kim E. Gene-based predictive models of trophic modes suggest Asgard archaea are not phagocytotic. *Nat Ecol Evol*. 2018;2(4):697-704.
39. Higashi A, Fujitani Y, Nakayama N, Tani A, Ueki S. Selective growth promotion of bloom-forming raphidophyte *Heterosigma akashiwo* by a marine bacterial strain. *Harmful Algae*. 2016;60:150-6.
40. Smeds L, Kunstner A. ConDeTri—a content dependent read trimmer for Illumina data. *PLoS ONE*. 2011;6(10):e26314.
41. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 2013;8(8):1494-512.
42. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-60.
43. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-20.

44. Eddy SR, editor Multiple alignment using hidden Markov models. Proceedings of the International Conference on Intelligent Systems for Molecular Biology; 1995.
45. Bock NA, Charvet S, Burns J, Gyaltshen Y, Rozenberg A, Duhamel S, et al. Experimental identification and in silico prediction of bacterivory in green algae. *Isme J.* 2021;15(7):1987-2000.
46. Shoguchi E, Shinzato C, Kawashima T, Gyoja F, Mungpakdee S, Koyanagi R, et al. Draft assembly of the Symbiodinium minutum nuclear genome reveals dinoflagellate gene structure. *Curr Biol.* 2013;23(15):1399-408.

Hosted file

Table1_fastq.docx available at <https://authorea.com/users/453501/articles/551316-de-novo-assembly-and-inferred-functional-annotation-of-the-transcriptome-of-heterosigma-akashiwo>

Hosted file

Table2_v2.docx available at <https://authorea.com/users/453501/articles/551316-de-novo-assembly-and-inferred-functional-annotation-of-the-transcriptome-of-heterosigma-akashiwo>

Hosted file

Table3_v3.docx available at <https://authorea.com/users/453501/articles/551316-de-novo-assembly-and-inferred-functional-annotation-of-the-transcriptome-of-heterosigma-akashiwo>

Hosted file

Table4_G0.docx available at <https://authorea.com/users/453501/articles/551316-de-novo-assembly-and-inferred-functional-annotation-of-the-transcriptome-of-heterosigma-akashiwo>

Hosted file

Table5_pfam.docx available at <https://authorea.com/users/453501/articles/551316-de-novo-assembly-and-inferred-functional-annotation-of-the-transcriptome-of-heterosigma-akashiwo>

Fig 1

Fig2

Fig 3

Fig 4