

## Nuclear gene transformation in a dinoflagellate

Brittany N. Sprecher, Huan Zhang\*, Senjie Lin\*

Department of Marine Sciences, University of Connecticut, 1080 Shennecossett Rd, Groton, CT 06340

\* Address for correspondence; emails [senjie.lin@uconn.edu](mailto:senjie.lin@uconn.edu); [huan.zhang@uconn.edu](mailto:huan.zhang@uconn.edu)

## ABSTRACT

The lack of a robust gene transformation tool that allows functional testing of the vast number of nuclear genes in dinoflagellates has greatly hampered our understanding of fundamental biology in this ecologically important and evolutionarily unique lineage. Here we report the development of a dinoflagellate expression vector, an electroporation protocol, and successful expression of introduced genes in the dinoflagellate *Oxyrrhis marina*. This protocol, involving the use of Lonza's Nucleofector and a codon optimized antibiotic resistance gene, has been successfully used to produce consistent results in several independent experiments. It is anticipated that this protocol will be adaptable for other dinoflagellates and will allow characterization of many novel dinoflagellate genes.

# INTRODUCTION

As widely distributed primary producers, essential coral endosymbionts, and the greatest contributors of harmful algal blooms and biotoxins in the ocean, dinoflagellates are a diverse group of unicellular protists with great ecological significance, evolutionary uniqueness, and numerous cytological and genomic peculiarities. Dinoflagellates have immense and permanently condensed genomes with many chromosomes<sup>1-3</sup>; their genomes have a low protein-DNA ratio and histones are functionally replaced with dinoflagellate viral nuclear proteins<sup>4,5</sup>; there are high numbers of repetitive non-coding regions and gene copies, in some species up to ~5,000 copies, organized in tandem arrays<sup>6,7</sup>; only 5-30% of their genes are transcriptionally regulated<sup>7-9</sup>, and microRNAs seem to be the major gene regulating mechanism<sup>10</sup>; and they have undergone extreme plastid evolution, transferring a massive quantity of plastid genes to the nucleus in most of the autotrophic species<sup>11-13</sup>. However, the molecular underpinnings of these unusual features remain elusive. In attempts to address the gap of knowledge, an increasing amount of effort has been made in the last decade to analyze dinoflagellate transcriptomes<sup>14-27</sup> and genomes<sup>10,28-32</sup>. These experiments have provided not only extensive information on predicted genes and biological pathways, but also an even greater wealth of genes that have weak similarity to characterized proteins or no significant matches in databases. With the increasing volume of dinoflagellate transcriptomic and genomic data, the functional characterization of these novel genes has become a major bottleneck in translating system-level data into a mechanistic understanding of basic dinoflagellate biology, warranting the need for a dinoflagellate genetic transformation system.

Gene transformation attempts have been reported for dinoflagellates by three separate groups. Ten and Miller (1998)<sup>33</sup> utilized silicon carbide whiskers, polyethylene glycol (PEG), and vigorous

shaking to introduce foreign DNA into *Amphidinium* sp. and *Symbiodinium microadriaticum* with a success rate of ~1 ppm. Seventeen years later, Ortiz-Matamoros et al. used PEG, glass beads, shaking and, in some cases, co-incubation with *Agrobacterium tumefaciens* to transform foreign DNA into *Fugacium kawagutii* (formerly *Symbiodinium kawagutii*), *S. microadriaticum*, and an unclassified *Symbiodiniaceae* species<sup>34,35</sup>. Neither of these reports used codon optimized plasmids for dinoflagellate expression nor did they contain potential dinoflagellate promoters; moreover, both methods remain to be reproduced in other laboratories. In a recent study plasmids containing dinoflagellate minicircle DNA and an antibiotic resist gene were designed and introduced successfully through particle bombardment into the chloroplast genome of the dinoflagellate, *Amphidinium carterae*<sup>36</sup>. Here, we report a successful nuclear gene transformation method for the heterotrophic dinoflagellate, *Oxyrrhis marina*.

*O. marina* is a widespread and ecologically significant heterotrophic dinoflagellate. It is an established model species for both ecological and evolutionary research due its easy cultivable nature, extensive studies related to feeding behavior and nutrition, and its basal position in dinoflagellate phylogeny<sup>37–41</sup>. Although *O. marina* is an early branching dinoflagellate species, it still shares many of the peculiar biological characteristics described above and also retains more typical eukaryotic features that are lacking in later diverging dinoflagellate taxa; thus, it represents a good model for understanding dinoflagellate evolution<sup>41–43</sup>. In addition, *O. marina* has represented planktonic heterotrophs in experiments examining both how they feed and their nutritional value<sup>44–46</sup>. Through various studies as a prey species for copepods and rotifers, *O. marina* has been considered a trophic upgrade as they produce long-chain fatty acids, sterols, and

essential amino acids that phytoplankton alone cannot<sup>45,47,48</sup>. Their nutritional value lead to the proposition of using *O. marina* as nutraceuticals for humans and agriculture<sup>45</sup>.

Although *O. marina* lacks a published genome, several transcriptomic studies are available<sup>16,49–52</sup>, and the most exciting finding is *O. marina* possess a potential proton pumping rhodopsin with homology to proteorhodopsin<sup>50,51</sup>. Proteorhodopsin is a retinal protein/carotenoid complex that utilizes sunlight to pump protons across a membrane, a non-photosynthetic form of light harvesting<sup>53</sup>. Dinoflagellate species across the phylogenic tree have been found to possess proteorhodopsin homologs, allowing the translational study of this proteins function in *O. marina* to the other dinoflagellate species<sup>16,17,50</sup>. Therefore, having a genetic transformation system in place for *O. marina* will greatly excel our understanding of heterotrophic protist ecology, deepen our evolutionary understanding of dinoflagellates within their own branch and relative to other alveolates, allow exploration of the many predicted and novel dinoflagellate genes, and could tap into new industrial applications for *O. marina*, such as a food source or a potential alternative fuel. Additionally, since *O. marina* is a heterotrophic species, it is easier to detect the expression of introduced florescent proteins without interference from chlorophyll florescence, as in photoautotrophic species.

In this study, based on genomic and transcriptomic data from several dinoflagellates, we constructed a dinoflagellate expression system (named as DinoIII) that contains potential promoter and termination regions as well as important RNA elements. We incorporated a codon optimized rifampin resistance gene (DinoIII-*arrO*) and green fluorescent protein gene, *gfp*, (DinoIII-*gfp*) into DinoIII, and transformed this DNA as either PCR amplified fragments, excluding the plasmid

component for DinoIII-*arrO*, or linear DNA, using a restriction enzyme to digest DinoIII-*gfp*, into *O. marina* using Lonza 4D-Nucleofector™ X system (Basel, Switzerland), a gene transformation system enabling transfer of genes directly into the cells' nucleus<sup>54</sup>. We have been able to repeat transformation for antibiotic resistance several times and verified the presence of both the antibiotic resistance gene and green fluorescent protein several months after transfection.

## RESULTS

### Construction of dinoflagellate backbone expression vector

Initially, the RNA complex sequence from dinoflagellate the *Karenia brevis* (GenBank accession # FJ434727) was inserted into the pMD™19-T plasmid vector (Takara, Kusatsu, Shiga Prefecture, Japan) and was used as the vector's skeleton for a series of modifications. After the addition of

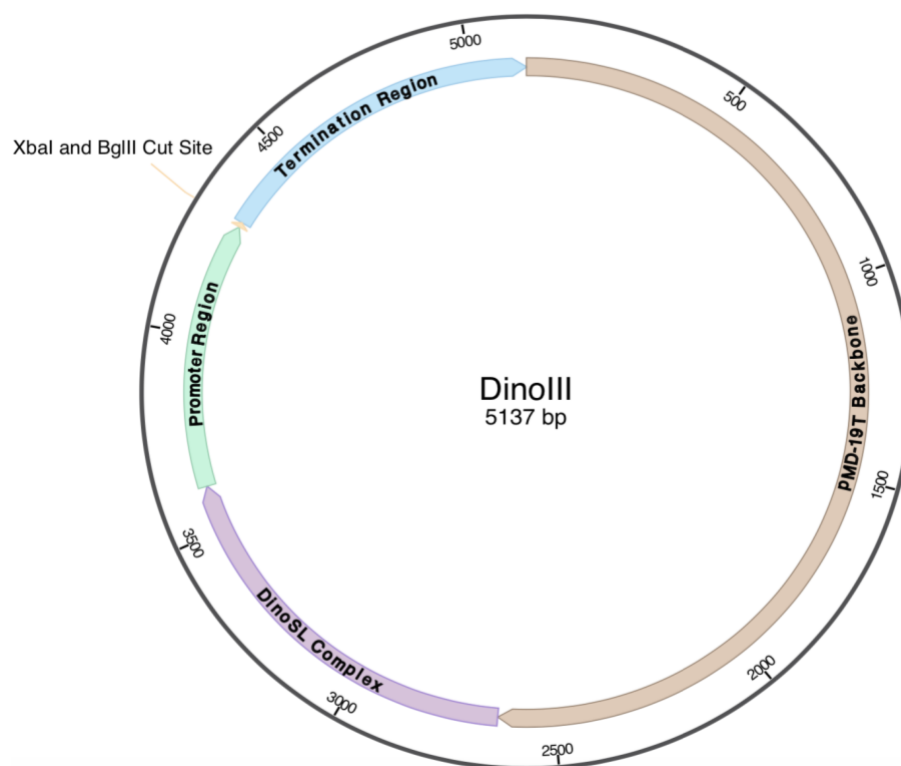


Fig. 1. Structure of DinoIII Expression Vector. The bacterial pMD-19 T-Vector portion is in brown, while others depict dinoflagellate elements, including DinoSL Complex region, Promoter Region, and Termination Region. XbaI and BglII Cut site, depicted with a line, that allow for easy gene incorporation in proper orientation.

more dinoflagellate elements, a functional dinoflagellate backbone vector was achieved, named DinoIII (5137bp; Fig. 1; Supplementary Tables 1-4).

### Transformation using Lonza 4D-Nucleofector™ X Unit system

With Lonza's 4D-Nucleofector™ X Unit, specifically designed for hard-to-transfect cell lines, we went through an extensive cell optimization protocol for *O. marina*, and identified seven adequate pulse code settings. We used these seven pulse codes for follow-up experiments (Table 1). Each pulse code had varying levels of success; for some *gfp* expression was observed, some had a higher percentage of cell survival rate (in which the cells were still growing and dividing after the wildtype NPC cells had died off), whereas others under antibiotic pressure grew to large enough populations to allow for RNA and DNA isolation. Taking all the data into consideration, the pulse codes that showed overall strongest performance were DS-137, DS-134, and DS-120.

Table 1. Adequate pulse code settings for transformation of DinoIII-*gfp* and DinoIII-*arrO* into *Oxyrrhis marina*

Pulse code	Number of Trials	Expression of <i>gfp</i>	Short-term Survival in Rifampin (Cells Survived > NPC)	Long-term Survival in Rifampin (Cells Survived > 1 month and RNA &/or DNA work was performed)
DS-137	12	Yes	66.67%	16.67%
DS-130	5	Yes	80.00%	0.00%
DS-138	9	No	66.67%	22.22%
DS-134	6	No	83.33%	33.33%
DS-150	9	No	66.67%	0.00%
ED-150	4	Weak	50.00%	0.00%
DS-120	12	Weak	66.67%	16.67%

Nevertheless, we recommend use of all seven settings in the first optimization tests for this species. For other algae to be studied, full optimization tests with the other available solutions should be utilized when using Lonza's 4D-Nucleofector™ X Unit.

## GFP Expression

The reporter expression vector, *DinoIII-gfp*, was introduced as linear DNA to *O. marina* cells and the presence of fluorescence was examined microscopically from the third day on. For several weeks we only observed a very dim green signal, but after three months the brightness of the green signal markedly increased for transformed cells using two of the pulse codes, DS-137 and DS-130 (video 1). Perhaps this increase was due to elevated expression after the transformed cells adapted to the new cell environment and/or was the result of accumulating the green fluorescent protein in one single area of the cells (Fig. 2). Less than 1% of the total *O. marina* population in the well was expressing the green fluorescent protein, making the isolation of this cell line challenging without a selection marker.

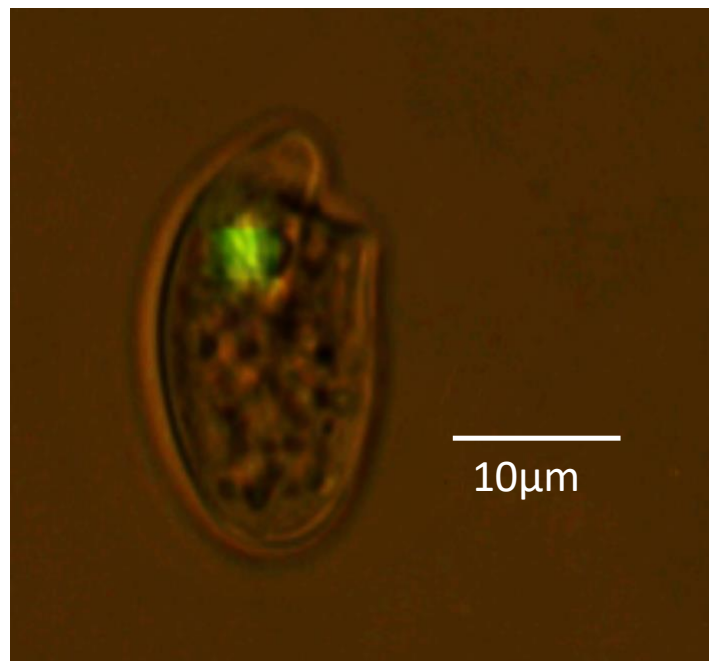


Fig. 2. Successful transformation of *Oxyrrhis marina* with a green fluorescence protein (GFP) gene (*gfp*). The green fluorescence indicates expression of the introduced *gfp*. Photo is a still image of *O. marina* with a video clip of a swimming *O. marina* cell.



## Rifampin Resistance as a selection marker

To facilitate purification of transformed *O. marina* cells, we screened several selection markers and found rifampin as the most suitable for *O. marina*. Rifampin is an antibiotic used to treat tuberculosis, leprosy, and Legionnaire's disease and its resistance in bacteria is due to Rifampin ADP-ribosyltransferase activity<sup>55</sup>. We introduced our codon-optimized homolog through our DinoIII vector (DinoIII-*arrO*) and obtained expression of *arrO*, which was verified in several ways. First, the transformed cell culture survived and grew while the wildtype died completely in rifampin-containing medium. Second, after approximately one month we isolated RNA and DNA from both the transformed cells cultured in rifampin-containing medium and a wildtype culture grown in rifampin-free growth medium, and performed reverse-transcription PCR (RT-PCR). We detected the expression of the resistance gene *arrO* only in the experimental treatment and not in the wildtype (Fig. 3A). In addition, we sequenced the PCR product and confirmed that it was *arrO*. Finally, the expression of *arrO* was detected from the cDNA synthesized using Oligo-dT as the primer (Fig. 3B), indicating the transcript of *arrO* was polyadenylated, a phenomenon best known for occurring mostly in eukaryotes mRNA.

Although the cells survived for more than one month, the population increased very slowly and did not seem healthy, probably due to low expression efficiency of the resistance gene. We attempted to increase the expression efficiency of our DinoIII vector by incorporating the intergenic region between *O. marina* rhodopsin tandem repeats, a potential promoter for this highly expressed protein. After introducing the new DinoIII-*arrO*-N PCR fragment into *O. marina* cells we saw an increase in growth rate under antibiotic selection and verified its expression, as reported

above, but this time three month after transfection (Fig. 3C). We still, when writing this manuscript, have the cell lines in culture.

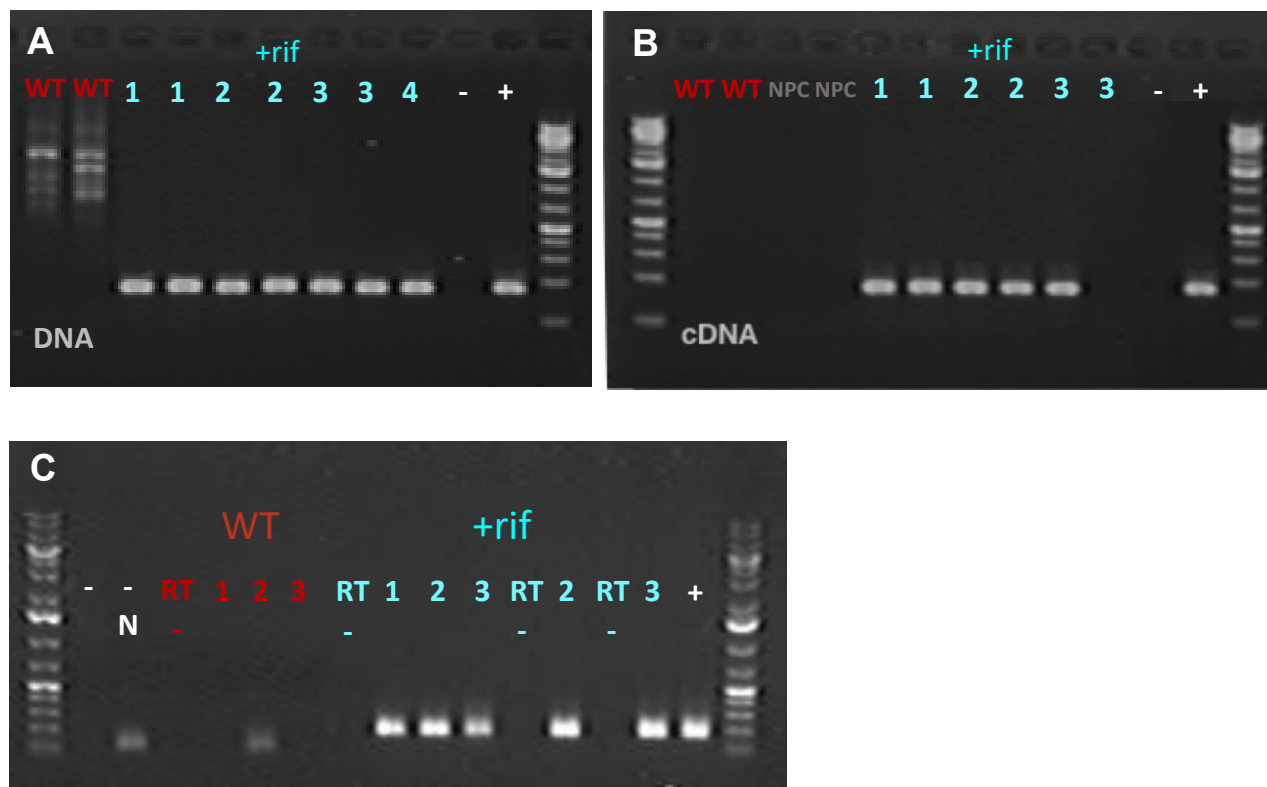


Fig. 3. Detection of *arrO* (rifampin resistance) gene and its expression in the transformed *Oxyrrhis marina* cells. A) *arrO* gene detected in genomic DNA of the transformed cells; B) *arrO* expression detected in the cDNAs of the transformed *O. marina*. C) *arrO-N* (*arrO* plus rhodopsin intergenic region) gene expression detected in the cDNAs of transformed *O. marina* after 3 months. +rif (in cyan), transformed cells; WT (in red), wild type; “-”, negative control, “- N”, negative nested PCR control; “RT -”, no reverse transcriptase negative control; “1” is N6 library; “2”, OdT Library; “3”, MdT library; “+”, plasmid positive control for *arrO-N*.

## DISCUSSION

In order to improve understanding of basic dinoflagellate biology, a gene transformation protocol is urgently needed to characterize the function of dinoflagellate genes, particularly the vast number of nuclear genes. A robust and reproducible protocol has been long-awaited. After testing multiple methods (including previously reported ones) and numerous conditions, we have found a passage and herein report a genome-targeted transformation method using a dinoflagellate *gfp* vector

(DinoIII-*gfp*) and two dinoflagellate rifampin resistance vectors (DinoIII-*arrO* and DinoIII-*arrO*-N) that were developed based on dinoflagellate genomic and transcriptomic data.

### Effectiveness of the promoter elements

Our efforts began with utilizing expression vectors from the previous reported dinoflagellate transformation<sup>34,35</sup>. This expression system utilized the plant CaMV 35S and *nos* promoters to drive expression of plasmids<sup>34,35</sup>. The *nos* and CaMV 35S promoters have been used extensively for plant transgenic studies, and the CaMV 35S promoter functions in both bacteria and animal systems<sup>56</sup> as well. These vectors contained the herbicide resistant gene, Basta (glufosinate), as well as several *gfp* fusion genes. Because *O. marina* is not sensitive to Basta, we were looking for green fluorescence as a marker for transformed cells, but it was not observed (results not shown). Therefore, we developed a series of dinoflagellate expression vectors based on existing dinoflagellate transcriptomic and genomic data, mirroring what was previously done for the two model alveolates, *Plasmodium falciparum* and *Tetrahymena thermophile*<sup>57,58</sup>.

In attempts to construct a vector that could drive expression of any gene in any dinoflagellate organism, we included dinoflagellate sequences from two different, phylogenetically separated species, *F. kawagutii* and *K. brevis*. The combination of the included elements (potential promoter, terminator, and RNA elements) were able to drive the expression, albeit at a low level, of the inserted genes in *O. marina*. In order to increase the expression level, we identified a G-rich intergenic region between the highly expressed rhodopsin genes, and incorporated it in our DinoIII-*arrO* vector, yielding visually higher cell survival under antibiotic pressure, indicative of stronger expression of the rifampin resistance gene. When comparing our intergenic region to the

luciferase tandem repeats<sup>59</sup>, our sequence is much shorter, only 70 base pairs compared to ~200-2000 and has no real sequence matches in public databanks. No proven promoter exists for dinoflagellates at this time and it is uncertain if the additional sequence contains a promoter. Previous research looking at the binding affinity of *Cryptocodinium cohnii* TATA-binding protein (TBP) homolog and *F. kawagutii* genomic content suggests that dinoflagellates have replaced the typical eukaryotic TATA box with a TTTT motif<sup>10,60</sup>, which is present 65 base pairs upstream from the start codon in our intergenic region and is also present in our “promoter” region with the closest motif 133 base pairs upstream from the start codon. Whether or not these sequences are important can be evaluated in future studies using our method.

### **Effectiveness of DNA introduction method**

*O. marina* is a naked dinoflagellate that had a difficult time withstanding the physical forces used in previously reported dinoflagellate transformation methods. Electroporation is a gentler method, allowing DNA to pass through temporary pores in an organism’s membrane and has been utilized in many organisms, but requires the removal of seawater and replacement with electroporation buffers, often unable to maintain the osmolality of marine organisms<sup>61</sup>. A new electroporation model, Lonza’s 4D-Nucleofector, provides a user with a score of built-in pulse settings and solutions that remove salts but help maintain dinoflagellates osmolality. The machine has been designed for rapid optimization of both buffer and electric pulse conditions, allowing delivery of nucleic acid substrates into the nucleus<sup>54</sup>. The nucleofector has been widely used on a variety of organisms and cell types and has recently been successfully used on two difficult to transfect marine protists, choanoflagellates and diplomonids<sup>62,63</sup>. For *O. marina*, seven pulse code settings (Table 1) and one solution, SG, allowed the expression of genes in DinoIII vectors. No one pulse

code performed the highest across all three of our different criteria. It is interesting to note that DS-138 is considered the weakest pulse setting and DS-120 is the highest of the seven and the remaining five are all in between. Unfortunately, these settings are proprietary and no correlation of the pulse settings can be extracted. Overall, if long-term or short-term expression is the goal, DS-134 is likely the ideal setting.

### **Identification of selection marker**

Although we were successful in taking microscopic videos of *O. marina* cells expressing the introduced green fluorescent protein, less than 1% of the population actually showed expression. To visualize the green fluorescent protein, cells need to be observed in the dark with blue light, and *O. marina* cells move very fast under the microscope (Video 1), making the isolation of this cell line challenging without a selection marker. The percentage of *O. marina* cells expressing *gfp* decreased with time and the green signal became weaker in four months and eventually was no longer visible. We also discovered that *O. marina* cells would give green-yellowish auto-fluorescence under blue light when fixed with Paraformaldehyde or any other commonly used fixatives, making it very challenging to take a clear image of the *gfp* expressed cell, a necessity to detect the exact location of GFP in the cell (Fig. 2).

Availability of an appropriate selection marker is crucial for yielding a useful transformed cell line. After extensive testing, rifampin was found to be effective for *O. marina*. Rifampin is a very strong pigmented antibiotic that appears to be very light sensitive. Due to this characteristic it is important to keep *O. marina* in lower light settings when under antibiotic selection, keep the cultures well fed, and continue to add new antibiotic medium to the transformed cell lines. Because

of the many strains of *O. marina*, it is important to test your species first to determine optimal antibiotic concentrations that can be used to select transformed cell lines, and apply an antibiotic cocktail that will reduce potential microbial communities in the culture.

Ideally a selection marker and a reporter gene can be located on the same plasmid allowing for dual expression. We attempted to put both the *arrO* and *gfp* genes within one single DinoIII vector in multiple arrangements (with or without stop codon in between, fused or not fused) but, unfortunately, the simultaneous expression of both genes was difficult to obtain. Future studies using *arrO* and *gfp* with the rhodopsin intergenic region in between could potentially get over this hurdle.

## Conclusion

Despite the proven challenges, we have developed a dinoflagellate expression system and successfully used it to express foreign genes in the dinoflagellate *Oxyrrhis marina*. Given the extensive studies on *O. marina*, its basal position in the dinoflagellate phylogenetic tree, its easy cultivable nature, and its wide acknowledgement as a model species for heterotrophic protists and dinoflagellates, the gene transformation tool developed here makes the species an even more valuable model. Having a genetic transformation system in place for *O. marina* will allow a deeper understanding of basic dinoflagellate biology. This report is the first stepping block to delve into dinoflagellates molecular biology using *O. marina* as the model and offers a dinoflagellate backbone vector with potential to work across the dinoflagellate phylogenetic tree.

## METHODS:

### Culturing *Oxyrrhis marina*

*Oxyrrhis marina* CCMP 1795 was grown at 20°C in autoclaved 0.22mm filtered seawater on a 14:10 hour light:dark cycle at a photon flux of  $\sim 100 \mu E m^{-2}s^{-1}$  and was fed *Dunaliella tertiolecta* CCMP1320 as prey<sup>51</sup>. Both species were purchased from the Provasoli-Guillard National Center of Marine Algae and Microbiota in West Boothby Harbor, Maine, USA.

### Constructing Dinoflagellate Expression Vectors

To optimize the utilization of our dinoflagellate expression system several regions were amplified from dinoflagellate genomes and were incorporated to serve as the vector backbone. The first region (974bp) comprises of DNA fragments from the dinoflagellate *Karenia brevis* including SL RNA, SRP RNA, several tRNAs, and U6<sup>64</sup>, which was named as DinoSL complex (Fig. 1; Supplementary Table 1). This region was PCR amplified with DinoSL and KbrSRP-U6R1 primer set (sequences and Tm in Table 2) using the high fidelity PrimeSTAR HS DNA Polymerase (Takara, Kusatsu, Shiga Prefecture, Japan) at 94°C for 1 min, 30 cycles at 95°C for 15s, 58°C for 30s, and 72°C for 1 min, and an additional elongation step at 72°C for 10min. The PCR product was run on 1% agarose to confirm the correct size, purified by passing through a DNA column (Zymo, Irvine, CA, USA), end-fixed, ligated into the pMD<sup>TM</sup>19-T plasmid vector (Takara), and transformed chemically into *Escherichia coli* competent cells. Ampicillin was used to select for colonies harboring the region and plasmids were isolated and sequenced to identify the best clone, named as pMD-Dino.

Table 2. Primers used in the present study

Primer Name	Sequence Information	Tm	Polymerase Used
SymkaLHC5FN1	GAGAACTAGTAAGTCCCGTGGCTGTCATATCTAG	68°C	Takara PrimeSTAR HS DNA Polymerase
SymLHC3_5R	GACTCCTGGCCGAGATCTTCTAGAGGCTCCGAAATTTGGTCTAAGCAC	68°C	Takara PrimeSTAR HS DNA Polymerase
SymLHC5_3F	CCAAATTTTCGGAGCCTCTAGAAGATCTCGGCCAGGAGTCACAGAAAACAAG	68°C	Takara PrimeSTAR HS DNA Polymerase
SymkaLHC3R1	TCTCTCGAATTCGGTGTGCTTGTAACCTTTTATC	68°C	Takara PrimeSTAR HS DNA Polymerase
DinoSL	NCCGTAGCCATTTTGGCTCAAG	58°C	Takara PrimeSTAR HS DNA Polymerase
KbrSRP-U6R1	CAGAGATCAAGACATGCTTCAGGAC	58°C	Takara PrimeSTAR HS DNA Polymerase
gfpNF2	AACTAGTATGGCTAGCAAAGGAGAAGAACTTTTC	5 cycles at 55°C and 25 at 62°C	Takara PrimeSTAR HS DNA Polymerase
gfpNR	TATGATCATCATTTGTAGAGCTCATCCATGCCA	5 cycles at 55°C and 25 at 62°C	Takara PrimeSTAR HS DNA Polymerase
arr2F	GAGAACTAGTATGGTGAAGGA	57°C	Takara PrimeSTAR HS DNA Polymerase
arr2R	TCTCTGATCACTAATCCTCG	57°C	Takara PrimeSTAR HS DNA Polymerase
OxyRhodF2	CACTACTTCMGNATCTTCAACTC	60°C	Takara PrimeSTAR HS DNA Polymerase
OxyrhodR	CAGAGGMACRGTCARCARCCARTC	60°C	Takara PrimeSTAR HS DNA Polymerase
Rhod_interspac erF	GAGAACTAGTAATTTTGGGAGTTGGGCT	57°C	Takara PrimeSTAR HS DNA Polymerase
Illu-DSL SymkaLHC3R1	TCTCTCGAATTCGGTGTGCTTGTAACCTTTTATC	68°C	Takara PrimeSTAR HS DNA Polymerase
arr2Q1F	TACCACGGAACCAAGGCGAACT	60°C	SsoAdvanced Universal SYBR Green Supermix
arr2Q1R	CCAAGCCAGACAGCGACATAGC	60°C	SsoAdvanced Universal SYBR Green Supermix



arr2Q1Fa	GAGATACCACGGAACCAAGGCGAACT	60°C	SsoAdvanced Universal SYBR Green Supermix
arr2Q1Ra	GAGACCAAGCCAGACAGCGACATAGC	60°C	SsoAdvanced Universal SYBR Green Supermix
MdT	TCAACGATACGCTACGTAACGTAATACGACTCACTATAGGGTTT TTTTTTTTTTTTTVN	42°C	Reverse Transcriptase

From the *F. kawagutii* genome sequence data<sup>10</sup>, we located the highly expressed light harvesting complex (LHC) gene. Its upstream “promoter” region (672bp; Supplementary Table 2) and downstream “termination” region (812bp; Supplementary Table 3) were PCR-amplified using the following primer sets: SymkaLHC5FN1 and SymLHC3\_5R for the “promoter” and SymLHC5\_3F and SymkaLHC3R1 for the “termination” region. All PCRs were performed at 94°C for 1 min, 25 cycles at 95°C for 15s, 68°C for 30s, and 72°C for 1 min, and 1 cycle of 72°C for 10 min. The sizes of the amplicons were checked by electrophoresis and DNA was purified by passing through a DNA column (Zymo).

SymLHC3\_5R and SymLHC5\_3F were designed to contain an overhang of either a portion of the “termination” region or the “promoter” region, respectfully, in order to link the two PCR products; thus, the two products were used in an equal molar ratio as template for the second PCR at 94°C for 1 min, 5 cycles without primers at 95°C for 15s, 68°C for 90s, 20 cycles with SymkaLHC5FN1 and SymkaLHC3R1 at 95°C for 15s, 68°C for 90s, and an extra elongation step of 72°C for 10 mins. The single product was checked by electrophoresis to verify the amplicon size, gel isolated, and digested with SpeI and EcoRI as SymkaLHC5FN1 had a SpeI site and SymkaLHC3R1 had an EcoRI site added to their 5’ ends for easy incorporation into pMD-Dino harboring the SL RNA, SRP RNA, several tRNAs, and U6 region. The pMD-Dino vector was digested with XbaI and

EcoRI and treated with alkaline phosphatase to avoid self-ligation. After 3.5 hours of digestion both products were purified by ethanol precipitation and ligated overnight in a 2:1 molar ratio (LHC product:vector) and transformed into competent *E. coli* cells. The colonies obtained were picked randomly, and plasmids were isolated and sequenced to identify the best clone harboring the correct DinoSL Complex-LHC region sequence, giving rise to the dinoflagellate expression vector backbone, DinoIII (5137bp; Fig. 1; Supplementary Table 4).

SymLHC3\_5R and SymLHC5\_3F primers were designed to also have an XbaI and BglII site in between the “Promoter” and “termination” regions so that a gene, either a reporter or an antibiotic resistant gene, could be inserted in the correct orientation. Accordingly, both a *gfp* gene and a rifampin resistance gene were incorporated into the expression vector, to yield DinoIII-*gfp* and DinoIII-*arrO*, respectively (Supplementary Table 5,6). For DinoIII-*gfp*, the crystal jelly *Aequorea victoria gfp* was amplified from the pGlo<sup>TM</sup> Plasmid (Bio-Rad, Hercules, CA, USA) using gfpNF2 and gfpNR at 94°C for 1 min, 5 cycles at 95°C for 15s, 55°C for 30s, and 72°C for 30s, 25 cycles at 95°C for 15s, 62°C for 30s, and 72°C for 30s, and an extra elongation step of 72°C for 10 mins. For *arrO*, a homolog to Rifampin ADP-ribosylating transferase from bacterium *Citrobacter freundii* was found on GenBank (accession # NC\_019991) and was codon-optimized for *O. marina* based on codon usage data from reported *O. marina* genes, and was synthesized through GeneArt (ThermoFisher Scientific, Waltham, MA, USA). Upon arrival the synthesized *arrO* was PCR amplified at 94°C for 1 min, 30 cycles at 95°C for 15s, 57°C for 30s, and 72°C for 30s, and an extra elongation step of 72°C for 10 mins with arr2F and arr2R primers. Both the *gfp* and *arrO* genes had a SpeI site at 5'-end and a BclI site at 3'-end; thus, after their PCR amplification, the single products were checked on a gel, passed through DNA columns to purify, and digested with

SpeI and BclII for 3.5 hours. At the same time DinoIII was digested with XbaI and BglII and treated with alkaline phosphates. After digestion, the *gfp* and *arrO* genes were ligated into DinoIII overnight in a 2:1 molar ratio, and were transformed into competent *E. coli* cells. Plasmids were isolated and sequenced to identify the clones containing correct sequences of DinoIII-*gfp* and DinoIII-*arrO*, respectively.

### Optimizing promoter region

To optimize the expression of transformed genes for *O. marina*, we set out to find a promoter region for their highly expressed proteorhodopsin genes (2-4 X10<sup>6</sup> copies/ng total RNA, twice that of mitochondrial *coxI*)<sup>51</sup>. To do this we used the OxyRhodF2 and OxyrhodR primer set<sup>51</sup>, under thermal cycle conditions with an extended extension time to favor long amplicons that cover two or more tandem repeats of the gene: 94°C for 1 min, 25 cycles at 95°C for 15s, 60°C for 30s, and 72°C for 90s, and an extra elongation step of 72°C for 10 min. Bands of ~800bp and ~1400bp were gel purified, cloned, and sequenced to identify a potential promoter. This yielded an intergenic region between rhodopsin tandem repeats, with the following sequence:

aattttgggagttgggctggaagatggggttggtggggatcgggggagaggtgactggtgtgtggtcgag. We added this sequence to the 5'-end of the *arrO* gene through GeneArt (ThermoFisher Scientific) and incorporated this *arrO*-N sequence into the DinoIII vector (DinoIII-*arrO*-N; Supplementary Table 7) as described above.

### Introducing DNA into *O. marina* using Lonza's Nucleofector

*O. marina* cultures were fed with *D. tertiolecta* three days before transformation in order to reach high cell densities. Taking advantage of the species photo-tactic behavior, *O. marina* cells were

concentrated using a flashlight, allowing cells to swim toward the light, consequently gathering only healthy cells from the culture. The cultures cell numbers were counted microscopically using a Sedgwick-Rafter counting chamber.

Electroporation was carried out using Lonza 4D-Nucleofector™ X Unit system in 16-well Nucleocuvette™ Strips using the manufacturer's SG and Supplemental 1 solutions. *DinoIII-gfp* was digested with EcoRI and introduced as a liner plasmid and *DinoIII-arrO* plasmid was PCR amplified, to produce linear fragments that only contained the dinoflagellate DNA portion. PCR was carried out using Illu-DSL and SymkaLHC3R1 primers at 94°C for 1 min, 25 cycles at 95°C for 15s, 68°C for 90s, and an extra elongation step of 72°C for 10 mins. The linear plasmid and PCR product were checked through electrophoresis, retrieved, and concentrated to 1 µg/µl using a Millipore Microcon DNA Fast Flow Column (Burlington, MA, USA). For each transformation well, 16.4 µl of solution SG, 3.6 µl of Supplemental 1 solution, and 2 µl of PCR product were used as transformation solution.

For every well, ca.  $2.5 \times 10^5$  cells were added and all the cells for the experiment (including the controls) were collected in 50mL tubes. The cells were centrifuged at 2500g for 3 mins, enough to form a pellet at the bottom of the tubes, and all but ~2-3mL of medium was removed. The cells were transferred into 1.5mL tubes and centrifuged at 900g for 2 mins and all remaining liquid was removed. The cells were re-suspended in the transformation solution and 22 µl were added to each well. After an initial optimization test the following electroporation settings were used for further experiments: DS-137, DS-130, DS-138, DS-134, DS-150, ED-150, DS-120, and no pulse controls (NPC).

Immediately after electroporation 80  $\mu$ l of the same seawater medium (SW) that *O. marina* was cultured on but with an antibiotic cocktail, AKS (100  $\mu$ g/ $\mu$ l ampicillin, 50  $\mu$ g/ $\mu$ l kanamycin, and 50  $\mu$ g/ $\mu$ l streptomycin), was added to each well. All of the volume was gently transferred into 24-well plates where each well already contained 1.4mL of the same SW+AKS medium. The transformed cells were allowed to recover for three days. For the DinoIII-*gfp* transformations, cells were examined microscopically under blue light for *gfp* expression. For DinoIII-*arrO* cells, 750  $\mu$ l were transferred to new 24-well plates and 750  $\mu$ l SW+AKS containing 450  $\mu$ g/ $\mu$ l of rifampin was added to both plates, so that the final concentration of rifampin was 225  $\mu$ g/ $\mu$ l. On the third day under antibiotic selection, *D. tertiolecta* in 225  $\mu$ g/ $\mu$ L rifampin medium was added to each well. New antibiotic solution was added every 3 weeks but the concentration was dropped down to 200  $\mu$ g/ $\mu$ L to allow for greater cell growth and *D. tertiolecta* in 225  $\mu$ g/ $\mu$ L rifampin medium was supplied whenever they were no longer detected in the medium.

### **Detecting the transformed gene and its expression**

Total DNA was isolated from both wildtype (WT) and +*arrO* *O. marina* cultures using our CTAB method<sup>65</sup> and total RNA was isolated using the Trizol-Chloroform method in combination with Zymo Quick RNA Miniprep Kit (Irvine, CA, USA)<sup>66</sup>. As the transformed cultures grew slowly under antibiotic pressure, only 200-400 cells were available. These cells were divided into two for DNA and collected on TSTP Isopore 3 $\mu$ m membrane filters (MilliporeSigma, Burlington, MA, USA). DNA and RNA were isolated as reported<sup>16,66</sup>.

Several first-strand cDNA preparations were made with RNA isolated using ImProm-II<sup>TM</sup> Reverse Transcriptase (Promega, Madison, WI, USA) following the manufacture's protocol with random hexamer (N6), oligo(dT)<sub>18</sub> (OdT), and modified OdT (MdT; Table 2) as the primers, respectively. If cell numbers were very low only OdT was used for cDNA synthesis to maximize cDNA production. Negative controls were included where no reverse transcriptase was added and was instead replaced with DEPC water.

PCR was performed using both the DNA and cDNA as templates with primer set arr2Q1Fa-arr2Q1Ra. Due to the low number of transformed cells used in DNA and RNA isolation, this PCR did not yield detectable amounts of products. The PCR products were then diluted 1000- and 10000-fold and used as template for a nested PCR with arr2Q1F-arr2Q1R as the primer (Table 2) for quantitative PCR. The products were run on a gel and sequenced directly to ensure the *arrO* gene was correctly amplified.

## Acknowledgements

We thank Dr. Lu Wang for her help in constructing DinoIII and Sarah Baseler for her assistance in screening selection markers. The work was supported by the Gordon and Betty Moore Foundation through Grant GBMF4980 (to SL and HZ) and the National Science Foundation Graduate Research Fellowship under Grant No. 1247393 (to BNS). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## References

1. Rill, R. L., Livolant, F., Aldrich, H. C. & Davidson, M. W. Electron microscopy of liquid crystalline DNA: direct evidence for cholesteric-like organization of DNA in dinoflagellate chromosomes. *Chromosoma* **98**, 280-286 (1989).
2. Hou, Y. & Lin, S. Distinct gene number-genome size relationships for eukaryotes and non-eukaryotes: Gene content estimation for dinoflagellate genomes. *PLoS One* **4**, e6978 (2009).
3. Chow, M. H., Yan, K. T. H., Bennett, M. J. & Wong, J. T. Y. Birefringence and DNA condensation of liquid crystalline chromosomes. *Eukaryot. Cell* **9**, 1577-1587 (2010).
4. Gornik, S. G. *et al.* Loss of nucleosomal DNA condensation coincides with appearance of a novel nuclear protein in dinoflagellates. *Curr. Biol.* **22**, 2303-2312 (2012).
5. Irwin, N. A. T. *et al.* Viral proteins as a potential driver of histone depletion in dinoflagellates. *Nat. Commun.* **9**, 1535 (2018).
6. Le, Q. H., Markovic, P., Hastings, J. W., Jovine, R. V. M. & Morse, D. Structure and organization of the peridinin-chlorophyll a-binding protein gene in *Gonyaulax polyedra*. *Mol. Gen. Genet.* **255**, 595-604 (1997).
7. Lin, S. Genomic understanding of dinoflagellates. *Res. Microbiol.* **162**, 551-569 (2011).
8. McLean, T. I. 'Eco-omics': a review of the application of genomics, transcriptomics, and proteomics for the study of the ecology of harmful algae. *Microb. Ecol.* **65**, 901-915 (2013).
9. Murray, S. A. *et al.* Unravelling the functional genetics of dinoflagellates: a review of approaches and opportunities. *Perspect. Phycol.* **3**, 37-52 (2016).
10. Lin, S. *et al.* The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene

- expression and coral symbiosis. *Science* **350**, 691-694 (2015).
11. Jackson, C. J., Gornik, S. G. & Waller, R. F. The mitochondrial genome and transcriptome of the basal dinoflagellate *Hematodinium* sp.: Character evolution within the highly derived mitochondrial genomes of dinoflagellates. *Genome Biol. Evol.* **4**, 59-72 (2012).
  12. Dorrell, R. G. & Howe, C. J. Integration of plastids with their hosts: Lessons learned from dinoflagellates. *Proc. Natl. Acad. Sci. USA* **112**, 10247-10254 (2015).
  13. Janouškovec, J. *et al.* Major transitions in dinoflagellate evolution unveiled by phylotranscriptomics. *Proc. Natl. Acad. Sci. USA* **114**, E171-E180 (2017).
  14. Erdner, D. L. & Anderson, D. M. Global transcriptional profiling of the toxic dinoflagellate *Alexandrium fundyense* using massively parallel signature sequencing. *BMC Genomics* **7**, 88 (2006).
  15. Nosenko, T. *et al.* Chimeric plastid proteome in the Florida ‘red tide’ dinoflagellate *Karenia brevis*. *Mol. Biol. Evol.* **23**, 2026-2038 (2006).
  16. Zhang, H. *et al.* Spliced leader RNA trans-splicing in dinoflagellates. *Proc. Natl. Acad. Sci. USA* **104**, 4618-4623 (2007).
  17. Lin, S., Zhang, H., Zhuang, Y., Tran, B. & Gill, J. Spliced leader-based metatranscriptomic analyses lead to recognition of hidden genomic features in dinoflagellates. *Proc. Natl. Acad. Sci. USA* **107**, 20033-20038 (2010).
  18. Moustafa, A. *et al.* Transcriptome profiling of a toxic dinoflagellate reveals a gene-rich protist and a potential impact on gene expression due to bacterial presence. *PLoS One* **5**, e9688 (2010).
  19. Toulza, E. *et al.* Gene expression in proliferating cells of the dinoflagellate *Alexandrium catenella* (Dinophyceae). *Appl. Environ. Microbiol.* **76**, 4521-4529 (2010).



20. Leggat, W., Yellowlees, D. & Medina, M. Recent progress in *Symbiodinium* transcriptomics. *J. Exp. Mar. Bio. Ecol.* **408**, 120-125 (2011).
21. Bayer, T. *et al.* *Symbiodinium* transcriptomes: Genome insights into the dinoflagellate symbionts of reef-building corals. *PLoS One* **7**, e35269 (2012).
22. Johnson, J. G., Morey, J. S., Neely, M. G., Ryan, J. C. & van Dolah, F. M. Transcriptome remodeling associated with chronological aging in the dinoflagellate, *Karenia brevis*. *Mar. Genomics* **5**, 15-25 (2012).
23. Zhang, H., Zhuang, Y., Gill, J. & Lin, S. Proof that dinoflagellate spliced leader (DinoSL) is a useful hook for fishing dinoflagellate transcripts from mixed microbial samples: *Symbiodinium kawagutii* as a case study. *Protist* **164**, 510-527 (2013).
24. Cooper, E. D., Bentlage, B., Gibbons, T. R., Bachvaroff, T. R. & Delwiche, C. F. Metatranscriptome profiling of a harmful algal bloom. *Harmful Algae* **37**, 75-83 (2014).
25. Gavelis, G. S., White, R. A., Suttle, C. A., Keeling, P. J. & Leander, B. S. Single-cell transcriptomics using spliced leader PCR: Evidence for multiple losses of photosynthesis in polykrikoid dinoflagellates. *BMC Genomics* **16**, 528 (2015).
26. Zhuang, Y., Zhang, H., Hannick, L. & Lin, S. Metatranscriptome profiling reveals versatile N-nutrient utilization, CO<sub>2</sub> limitation, oxidative stress, and active toxin production in an *Alexandrium fundyense* bloom. *Harmful Algae* **42**, 60-70 (2015).
27. Gong, W. *et al.* Molecular insights into a dinoflagellate bloom. *ISME J.* **11**, 439-452 (2017).
28. Jaeckisch, N. *et al.* Comparative genomic and transcriptomic characterization of the toxigenic marine dinoflagellate *Alexandrium ostenfeldii*. *PLoS One* **6**, e28012 (2011).
29. Shoguchi, E. *et al.* Draft assembly of the *Symbiodinium minutum* nuclear genome reveals

- dinoflagellate gene structure. *Curr. Biol.* **23**, 1399-1408 (2013).
30. Aranda, M. *et al.* Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Sci. Rep.* **6**, 39734 (2016).
31. Liu, H. *et al.* *Symbiodinium* genomes reveal adaptive evolution of functions related to coral-dinoflagellate symbiosis. *Commun. Biol.* **1**, 95 (2018).
32. Shoguchi, E. *et al.* Two divergent *Symbiodinium* genomes reveal conservation of a gene cluster for sunscreen biosynthesis and recently lost genes. *BMC Genomics* **19**, 458 (2018).
33. Ten Lohuis, M. R. & Miller, D. J. Genetic transformation of dinoflagellates (*Amphidinium* and *Symbiodinium*): Expression of GUS in microalgae using heterologous promoter constructs. *Plant J.* **13**, 427-435 (1998).
34. Ortiz-Matamoros, M. F., Villanueva, M. A. & Islas-Flores, T. Transient transformation of cultured photosynthetic dinoflagellates (*Symbiodinium* spp.) with plant-targeted vectors | Transformación de dinoflagelados fotosintéticos del género *Symbiodinium* en cultivo con vectores diseñados para plantas. *Ciencias Mar.* **41**, 21-32 (2015).
35. Ortiz-Matamoros, M. F. *et al.* Heterologous DNA uptake in cultured *Symbiodinium* spp. aided by *Agrobacterium tumefaciens*. *PLoS One* **10**, e0132693 (2015).
36. Nimmo, I. *et al.* Genetic transformation of the dinoflagellate chloroplast. *bioRxiv* **452144**, (2018).
37. Saldarriaga, J. F., McEwan, M. L., Fast, N. M., Taylor, F. J. R. & Keeling, P. J. Multiple protein phylogenies show that *Oxyrrhis marina* and *Perkinsus marinus* are early branches of the dinoflagellate lineage. *Int. J. Syst. Evol. Microbiol.* **53**, 355-365 (2003).
38. Zhang, H. & Lin, S. mRNA editing and spliced-leader RNA trans-splicing groups *Oxyrrhis*, *Noctiluca*, *Heterocapsa*, and *Amphidinium* as basal lineages of dinoflagellates.

- J. Phycol.* **44**, 703-711 (2008).
39. Montagnes, D. J. S. *et al.* An introduction to the special issue: *Oxyrrhis marina*, a model organism? *J. Plankton Res.* **33**, 549-554 (2011).
40. Montagnes, D. J. S. *et al.* *Oxyrrhis marina* growth, sex and reproduction. *J. Plankton Res.* **33**, 615-627 (2011).
41. Guo, Z., Zhang, H., Liu, S. & Lin, S. Biology of the marine heterotrophic dinoflagellate *Oxyrrhis marina*: current status and future directions. *Microorganisms* **1**, 33-57 (2013).
42. Slamovits, C. H., Saldarriaga, J. F., Larocque, A. & Keeling, P. J. The highly reduced and fragmented mitochondrial genome of the early-branching dinoflagellate *Oxyrrhis marina* shares characteristics with both apicomplexan and dinoflagellate mitochondrial genomes. *J. Mol. Biol.* **372**, 356-368 (2007).
43. Slamovits, C. H. & Keeling, P. J. Contributions of *Oxyrrhis marina* to molecular biology, genomics and organelle evolution of dinoflagellates. *J. Plankton Res.* **33**, 591-602 (2011).
44. Jeong, H. J. *et al.* Interactions among the toxic dinoflagellate *Amphidinium carterae*, the heterotrophic dinoflagellate *Oxyrrhis marina*, and the calanoid copepods *Acartia* spp. *Mar. Ecol. Prog. Ser.* **218**, 77-86 (2001).
45. Yang, Z., Jeong, H. J. & Montagnes, D. J. S. The role of *Oxyrrhis marina* as a model prey: current work and future directions. *J. Plankton Res.* **33**, 665-675 (2011).
46. Roberts, E. C. *et al.* Feeding in the dinoflagellate *Oxyrrhis marina*: linking behaviour with mechanisms. *J. Plankton Res.* **33**, 603-614 (2011).
47. Kleppel, G. S., Burkart, C. A. & Houchin, L. Nutrition and the regulation of egg production in the calanoid copepod *Acartia tonsa*. *Limnol. Oceanogr.* **43**, 1000-1007 (1998).

48. Lund, E. D., Chu, F. L. E., Harvey, E. & Adlof, R. Mechanism(s) of long chain n-3 essential fatty acid production in two species of heterotrophic protists: *Oxyrrhis marina* and *Gyrodinium dominans*. *Mar. Biol.* **155**, 23-26 (2008).
49. Lowe, C. D. *et al.* The transcriptome of the novel dinoflagellate *Oxyrrhis marina* (Alveolata: Dinophyceae): Response to salinity examined by 454 sequencing. *BMC Genomics* **12**, 519 (2011).
50. Slamovits, C. H., Okamoto, N., Burri, L., James, E. R. & Keeling, P. J. A bacterial proteorhodopsin proton pump in marine eukaryotes. *Nat. Commun.* **2**, 183 (2011).
51. Guo, Z., Zhang, H. & Lin, S. Light-promoted rhodopsin expression and starvation survival in the marine dinoflagellate *Oxyrrhis marina*. *PLoS One* **9**, e114941 (2014).
52. Lee, R. *et al.* Analysis of EST data of the marine protist *Oxyrrhis marina*, an emerging model for alveolate biology and evolution. *BMC Genomics* **15**, 122 (2014).
53. Beja, O. *et al.* Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science* **289**, 1902-1906 (2000).
54. Martinet, W., Schrijvers, D. M. & Kockx, M. M. Nucleofection as an efficient nonviral transfection method for human monocytic cells. *Biotechnol. Lett.* **25**, 1025-1029 (2003).
55. Baysarowich, J. *et al.* Rifamycin antibiotic resistance by ADP-ribosylation: Structure and diversity of Arr. *Proc. Natl. Acad. Sci. USA* **105**, 4886-4891 (2008).
56. Hull, R.; Covey, S. N.; Dale, P. Genetically modified plants and the 35S promoter: assessing the risks and enhancing the debate. *Microb. Ecol. Health Dis.* **12**, 1-5 (2000).
57. Kahn, R. W., Andersen, B. H. & Brunk, C. F. Transformation of *Tetrahymena thermophila* by microinjection of a foreign gene. *Proc. Natl. Acad. Sci. USA* **90**, 9295-9299 (1993).

58. Crabb, B. S. & Cowman, A. F. Characterization of promoters and stable transfection by homologous and nonhomologous recombination in *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. USA* **93**, 7289-7294 (1996).
59. Liu, L. & Hastings, J. W. Novel and rapidly diverging intergenic sequences between tandem repeats of the luciferase genes in seven dinoflagellate species. *J. Phycol.* **42**, 96-103 (2006).
60. Guillebault, D. *et al.* A new class of transcription initiation factors, intermediate between TATA box-binding proteins (TBPs) and TBP-like factors (TLFs), is present in the marine unicellular organism, the dinoflagellate *Cryptothecodinium cohnii*. *J. Biol. Chem.* **277**, 40881-40886 (2002).
61. Potter, H. & Heller, R. Transfection by electroporation in *Current. Protocols in Molecular Biology* (ed. Ausubel, F. M. *et al.*) 9.3.1-9.3.10 (John Wiley & Sons Inc, 2010).
62. Booth, D. S., Szmidt-Middleton, H. & King, N. Transfection of choanoflagellates illuminates their cell biology and the ancestry of animal septins. *Mol. Biol. Cell* **29**, 2969-3062 (2018).
63. Kaur, B. *et al.* Transformation of *Diplonema papillatum*, the type species of the highly diverse and abundant marine microeukaryotes Diplonemida (Euglenozoa). *Environ. Microbiol.* **20**, 1030-1040 (2018).
64. Zhang, H., Campbell, D. A., Sturm, N. R. & Lin, S. Dinoflagellate spliced leader RNA genes display a variety of sequences and genomic arrangements. *Mol. Biol. Evol.* **26**, 1757-1771 (2009).
65. Lin, S., Zhang, H. & Dubois, A. Low abundance distribution of *Pfiesteria piscicida* in Pacific and Western Atlantic as detected by mtDNA-18S rDNA real-time polymerase

- chain reaction. *J. Plankton Res.* **28**, 667-681 (2006).
66. Zhang, H. *et al.* An improved method for achieving high-quality RNA for copepod transcriptomic studies. *J. Exp. Mar. Bio. Ecol.* **446**, 57-66 (2013).

Video 1. The *Oxyrrhis marina* cell transformed with *gfp* showing green florescence under blue light. Video was taking using Olympus BX51 microscopic system with a DP74 Olympus camera under 100X magnification.

**Video in Supplementary Material**