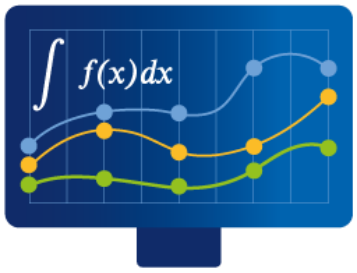
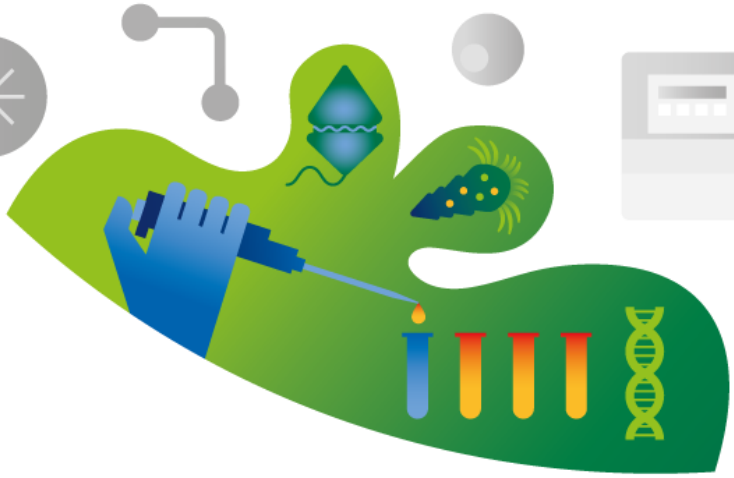
# Interpretative Toolkit To Enable The Use Of Genomic Data Aligned With Physiological Status For Representative Mixoplankton



MIXOTROPH.ORG
MIXITIN

MixITiN Project Report D3.7

H2020-MSCA-ITN
Bringing marine ecology into 21st century


Training next generation marine ecologists in the mixotroph paradigm

# MixITiN

**Project no. 766327**

# Interpretative toolkit to enable the use of genomic data aligned with physiological status for representative mixoplankton species

Konstantinos Anestis, Joost Samir Mansour, Maira Maselli, Per Juel Hansen, Fabrice Not, Uwe John

## Acknowledgements

*Project **MixITiN** has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 766327. This document reflects only the author's view; the REA and the European Commission are not responsible for any use that may be made of the information it contains.*

## How to cite this work

Anestis K, Mansour JS, Maselli M, Hansen PJ, Not F, John U (2022) Interpretative toolkit to enable the use of genomic data aligned with physiological status for representative mixoplankton species. Publisher: Zenodo. https://doi.org/10.5281/zenodo.5520904.

## Copyright © 2022 Anestis, Mansour, Maselli, Hansen, Not and John



## Updates

Please check for any updates in the latest version of this publication by visiting the url given in 'How to cite this work', above. If you note any errors, please contact the lead author and/or project co-ordinator. Thank you.

## Lead author contact

Uwe John
Department of Ecological Chemistry
Alfred-Wegener-Institute Helmholtz Centre for Polar and Marine Research
Bremerhaven
Germany.
Email: uwe.john@awi.de

## Project Co-ordinator contact

Aditee Mitra
School of Earth and Environmental Sciences
Cardiff University
Cardiff CF10 3AT
Wales, U.K.
Email: MitraA2@cardiff.ac.uk

## Author affiliations

**Konstantinos Anestis**, Department of Ecological Chemistry, Alfred-Wegener-Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany.

**Joost Samir Mansour,** CNRS and Sorbonne University, UMR7144 Adaptation and Diversity in Marine Environment (AD2M) Laboratory, Ecology of Marine Plankton Team, Station Biologique de Roscoff, France.

**Maira Maselli**, Marine Biological Section, University of Copenhagen, DK-3000, Helsingør, Denmark.

**Per Juel Hansen**, Marine Biological Section, University of Copenhagen, DK-3000, Helsingør, Denmark.

**Fabrice Not,** CNRS and Sorbonne University, UMR7144 Adaptation and Diversity in Marine Environment (AD2M) Laboratory, Ecology of Marine Plankton Team, Station Biologique de Roscoff, France.

**Uwe John**, Department of Ecological Chemistry, Alfred-Wegener-Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany.

# Table of Contents

# 1 Executive Summary

- In *Prymnesum parvum* (CM), Prymnesin production and the expression of the biosynthesis pathway associated polyketide synthases have been studied and metabolic trade-offs have been elucidated with transcriptomics.
- Mixotrophy in the kleptoplastidic and karyokleptic *Strombidium* cf *basimorphum* (GNCM) has been studied and detailed by a combinational approache of single-cell transcriptomics, qPCR analysis and fluorescence in-situ hybridisation (FISH).
- The combined approach of qPCR and FISH appears to be a promising tool to investigate the potential retention of prey nuclei in NCM.
- Culture based studies allow to take samples of organisms with a known feeding history making easier the link of physiology and genomic.
- Metabolic interactions of symbiont and Acantharia (eSNCM, host) of endosymbiotic mixoplankton can be explored using single-cell transcriptomics, and linked to single-cell physiological approaches (such as NanoSIMS).

# 2 Introduction

Mixotrophy is the combination of autotrophy and heterotrophy and in particular phototrophy and phagotrophy expressed within a single organism (Flynn et al., 2019). The term mixoplankton describes planktonic protists that express, or have potential to express both, phototrophy and phagotrophy. The functional classification of mixoplankton is based on how the cell incorporates photosynthesis. For mixotrophic protists (*i.e.* mixoplankton), the classification and the hypothesised evolutionary path is illustrated in figure 1 (Mansour and Anestis, 2021). Mitra et al. (2016) introduced the general grouping of either *constitutive mixoplankton* (*CM*) or *non-constitutive mixoplankton* (*NCM*). A CM has an inherent capability for both phototrophy and phagotrophy. Constitutive mixoplankton are found in most microeukaryotic lineages including green algae, euglenophytes, cryptophytes, chrysophytes, haptophytes, and dinoflagellates. NCM which are defined by the need to acquire their photosynthetic ability through external means and are found mostly among ciliates, dinoflagellates, Foraminifera, and Radiolaria. Phototrophic activity in NCM can be broadly achieved by three methods which further divide NCM into sub-groups: 1) those that steal chloroplasts (kleptoplasts) of (any) prey (*generalist non-constitutive mixoplankton, GNCM*); 2) those that steal chloroplasts from a specific type of prey *(plastidic specialist non-constitutive mixoplankton pSNCM*) and 3) those that harbour photosynthetically active endosymbionts (*endosymbiotic specialist non-constitutive mixoplankton*, *eSNCM*) (acquired phototrophy reviewed in Stoecker et al., 2009).

Approaches to study species functionality and responses to environmental parameters are predominantly done using cultures. This becomes more complex for mixoplankton and specifically NCM, as they are characterised by their interactions to other organisms.

The presence of other eukaryotes is in many cases (and especially for NCM) is obligatory in order to sustain a culture, if at all possible. This is a main obstacle, which could result in issues or ambiguity in downstream bioinformatics work and generally limits current knowledge on NCM. To tackle this, novel cultivation-independent approaches like single-cell sampling and analyses can be an outcome to study NCM. Such approaches will help improve our knowledge about the genomic aspects of interactions of kleptoplastic/karyokleptic and endosymbiotic mixotrophs, and, when unknown, could hint at physiological potential (Burns et al., 2018; Jimenez et al., 2020).

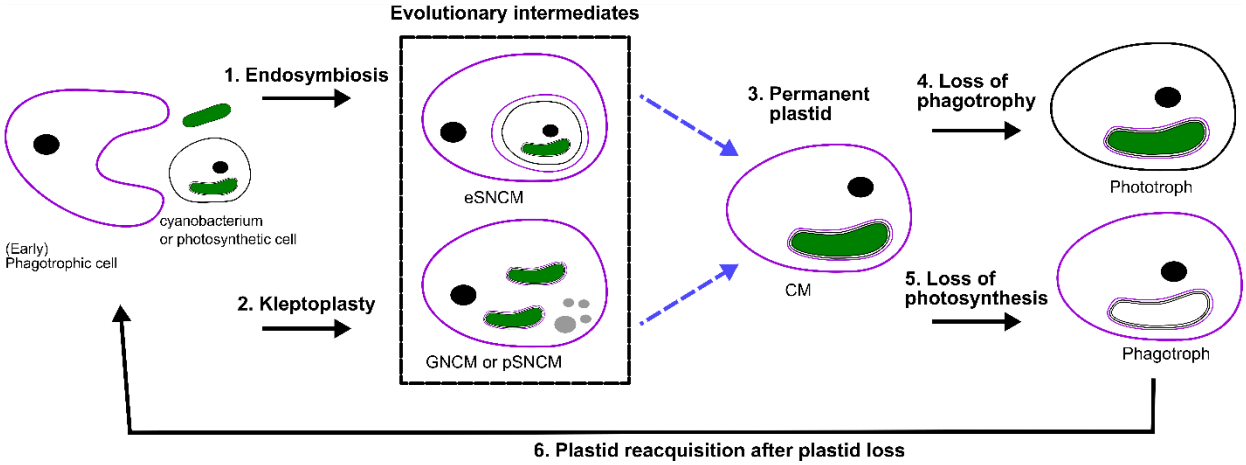



*Figure 1. Perspective of protist plastid evolution and the role of mixotrophy. Conceptual illustration of plastid evolution depicting non-constitutive mixoplankton as hypothetical intermediate stages to permanent plastids. The cell nucleus is shown as a black circle, kleptokaryons as grey circles, and photosynthetic plastids in green. While complex permanent plastids gained by secondary or tertiary endosymbiosis are bound by three or four membranes, respectively as illustrated by the number of lines around the plastid. In secondary and higher plastid acquisition, the photosynthetic cell can be either a CM or phototroph with primary, secondary or higher-order plastid. Illustrated here is secondary plastid acquisition with the host membrane in purple. Blue dashed lines are the hypothetical route toward permanent plastid retention. (Reproduced from Mansour, J. S., and Anestis, K. (2021) under CC BY-4.0).*

A major challenge in studies examine mixotrophy at the cellular bases and understanding the regulative processes of metabolic needs and photosynthetic activities is linking physiological potential to rate processes. For this reason, there is an emerging need to combine both eco-physiological parameters and genomics data in order to study the

contribution of both phagotrophy and phototrophy in cellular physiology. To achieve that, we have made an effort to create integrative datasets which include sampling for genomic analyses in single-cell and batch cultures, aligned with physiological parameters for the studied nutritional status. This document gives an overview of how this was achieved, as well as the generated datasets and sample parameters which are available for future studies.

Transcriptomic studies of grazing are usually based on whole culture RNA extractions. As a result, information comes from a mix of cells that feed on prey. However, feeding is not a synchronized process and possibly only a small portion of the total cell population engages in phagotrophy at any instant. Performing single-cell isolations from cultures allows us to select only for cells that actively fed and thus contained a distinct food vacuole. Additionally, molecular markers are used for prey nuclei detection and improve the understanding of role and function of kleptoplasty in planktonic protist.

Before proceeding to the establishment of interpretative tools combining both physiological and genomic data, it is important to highlight the parameters important for studying processes of mixotrophy (here mixoplanktonic activities). The main question is how much the photosynthesis and ingestion contribute respectively to the energetic budget of the cells. Thus, the photosynthetic rates should be measured in conjunction with the ingestion rates. The $^{14}C$ technique is an effective tool for measuring photosynthetic activity (Nielsen 1952; Rivkin and Seliger 1981) while ingestion rates are usually measured via prey disappearance during short incubations (Heinbokel 1978). Dissolved inorganic nutrients concentration in the media should be monitored during the incubation to estimate the cellular up-take and excretion rates. These measurements, together with the measurements of the elemental composition (C:N:P) of the organisms' biomass would allow a complete description of nutrients dynamics in the protists. These

parameters have been measured on cultures of the CM *P. parvum* and the NCM *S. basimorphum.* Both bulk and single-cell samples have been taken for genomic analysis, allowing us to gain understanding of process in both population and single-cell level.

# 3 Approaches and methodology

## 3.1 Toxin biosynthesis in *Prymnesium parvum* (CM)

The CM *Prymnesium parvum* was used in order to study physiological processes such as toxin production as well as the dynamics of phagotrophy under different conditions (such as salinity and phosphorous availability). To provide new insights on toxin biosynthesis mechanisms, we compared the prymnesin (toxin) profile, the presence of polyketide synthase genes (PKSs), and the metabolic and cellular functions of nine *P. parvum* strains. We chose three strains for each of the prymnesin types, A, B and C. Transcripts encoding for PKS were used to assess the phylogeny of the KS domain and the relationship between prymnesin type and PKS genes. Furthermore, we measured the cellular toxin content of all nine strains, relating patterns between toxin types and cellular toxin contents with the presence of PKS genes. To understand the potential molecular and metabolic trade-offs of toxin production in *P. parvum,* we performed gene expression variance analysis to detect genes showing expression patterns that relate to the cellular toxin content (**Error! Reference source not found.**). Detailed description of the m ethodology is provided in Anestis et al. (2021).

## *3.2* Eco-physiology of *Prymnesium parvum* (CM)

Cultures of the *P. parvum* (strain UIO-223) were grown under different salinity and phosphorus availability. The default medium used was K-medium. Low salinity was achieved by diluting filter-sterilised seawater (salinity 30) with MilliQ water to obtain a

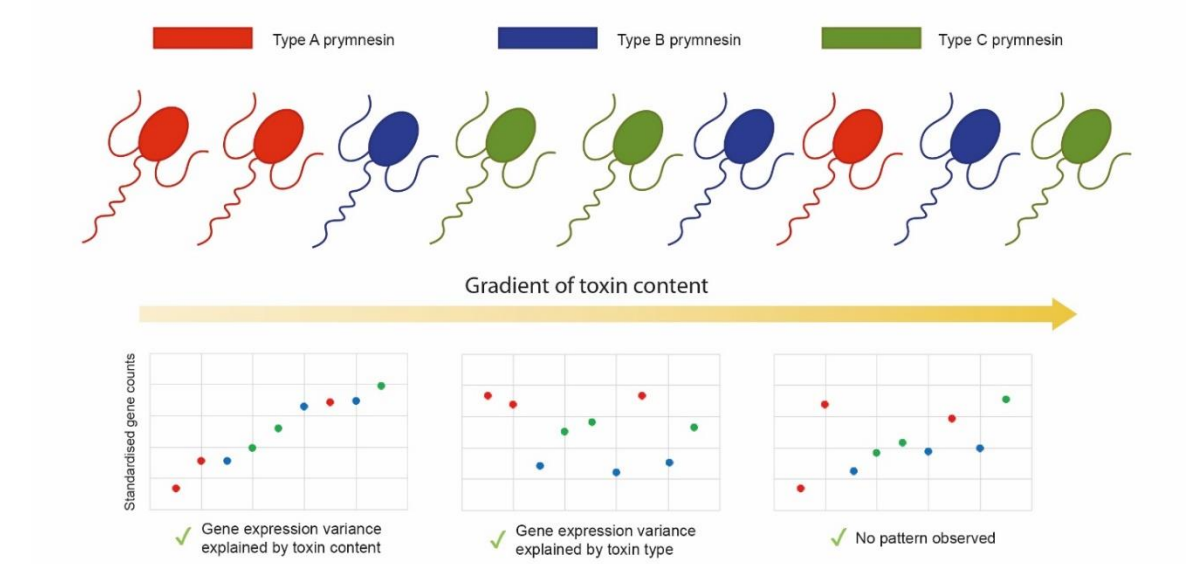



*Figure 2. Conceptual illustration of the approach followed to analyse toxin content related gene variance. Strains of various toxin type (Type A, B and C) and toxin content were used. Gene expression variance could be explained by either toxin content amount, prymnesin type or neither of these parameters (residuals).*

salinity of 5. All essential nutrients contained in K-medium were added, but in low phosphorus cultures, $PO_4^-$ was added at a final concentration of 2.4 μM. The concentration of $PO_4^-$ for the low phosphorus cultures was decided upon preliminary work and in accordance with pre-existing literature.

Mixotrophy in *P. parvum* was tested using *Teleaulax acuta* as prey. Incubations were set up using *P. parvum* cells at middle exponential phase for the non-limiting phosphorus condition and stationary phase for those in phosphorus deficiency. The initial cell concentration for the incubation was $30\times10^3$ cells $mL^{-1}$ and the ratio between *P. parvum* and prey was 1:1. To estimate phagotrophy, the incubations were sampled at different time points and cells were fixed using Lugol. The samples were examined under light microscopy (Axio Vert.A1 Microscope equipped with a Colibri 7 light source Zeiss), and the number of *P. parvum* cells containing a food vacuole were estimated.

## 3.3 Mixotrophy in the kleptoplastidic and karyokleptic *Strombidium* cf *basimorphum* (NCM)

We used culture samples of the NCM ciliate *Strombidium* cf *basimorphum* (isolated and cultured as in Maselli et al. 2020) to study the dynamics of the kleptoplasts and the relative contribution of phagotrophy vs phototrophy. S*trombidium* cf *basimorphum* was cultured with saturating amount of prey (T0), allowed to deplete the prey (T1-T2) and then starved for several days (T3), after which the cultures were re-fed (T4-T5) (Figure 3). Physiological parameters (growth rate, ingestion rates, nutrients up-take and excretion rates, photosynthetic rates, chlorophyll content and CNP content) were measured and single cell samples were taken for transcriptomic analysis. Samples of the bulk cultures were taken at the different time points to investigate the potential retention of prey genetic material in the ciliates via qPCR and FISH (methods are described in Maselli et al., 2021).

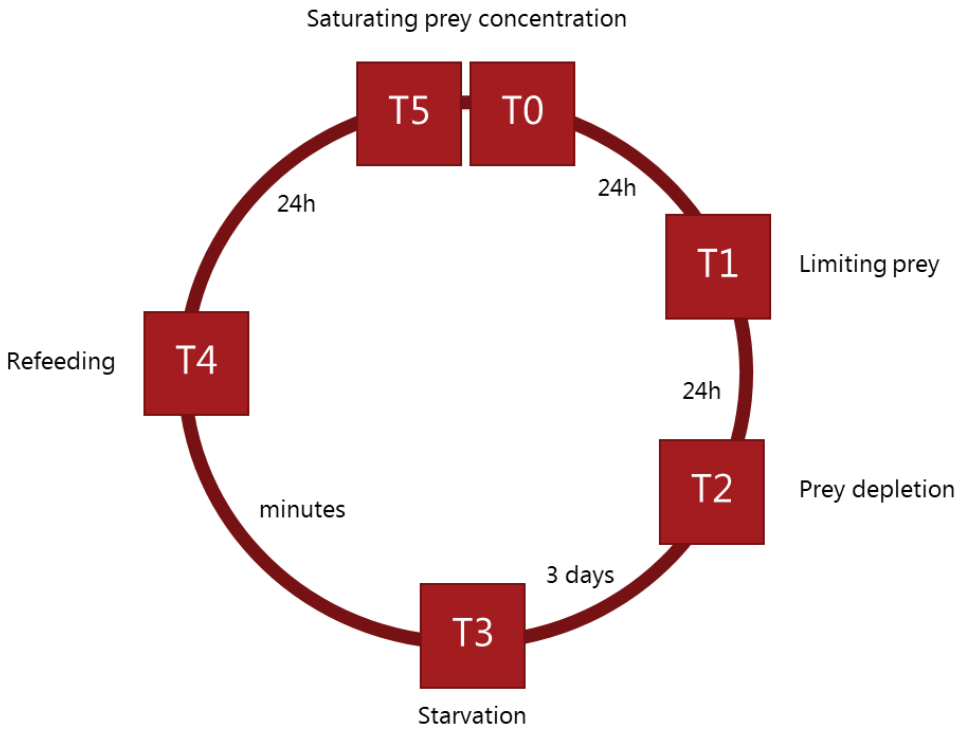



*Figure 3. Sampling strategy over 5 sampling points (T0-T5) to study physiology and perform single cell transcriptomics and batch culture transcriptomics in cultures of the NCM S. basimorphum in different nutritional status.*

To validate the presence of prey transcripts in the ciliate as an indicator of active nuclei, nucleomorph, and/or plastid genome, single-cell transcriptomics were performed. Eight single cells were isolated from the experimental cultures after 4 days of prey starvation. Each cell was individually picked with a drawn Pasteur pipette, washed three times by transferring it into clean filter-sterilised medium. Detailed protocols for the single-cell RNA isolation and library preparation are available at protocol.io (see Mansour et al., 2021b and Mansour et al., 2021a). Moreover, the full methodology we used is described in Maselli et al., 2021.

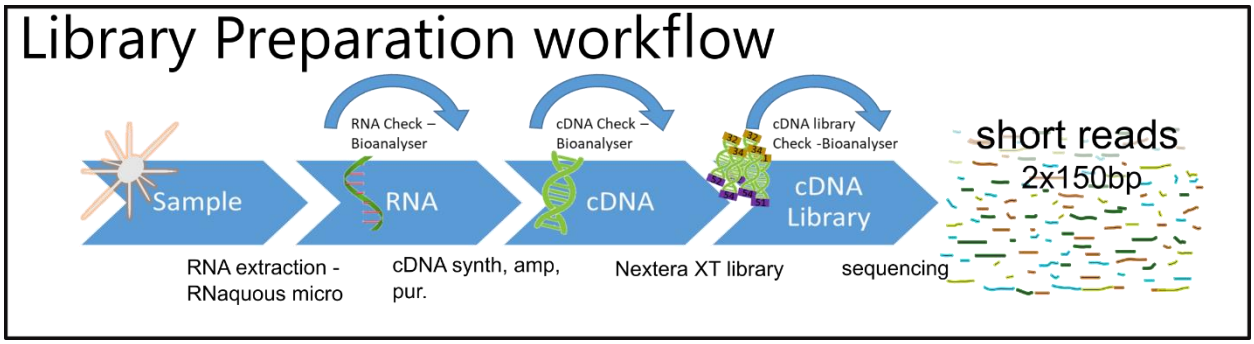



*Figure 4. cDNA library preparation workflow from sample to transcriptome sequencing. All major steps are indicated including a quality check by Bioanalyzer.*

## 3.4 Nutrient uptake capability of Acantharia (eSNCM)

For uncultivatable endosymbiotic mixoplankton (eSNCM) cultivation-independent single-cell approaches are especially important. In addition, the MixITiN proposal as well as various recent studies (Leles et al., 2017, 2019; Faure et al., 2019) have further highlighted the importance of establishing effective protocols for their sampling. Sampling using slow horizontal plankton net-tows are proven to be the best compromise between ease of sampling and keeping cell quality for subsequent physiological investigations (Graham et al., 1976; Mansour et al., 2021c). Detailed protocols for single-cell sampling

have been reported in deliverable D3.8 and are publicly available at protocol.io (dx.doi.org/10.17504/protocols.io.bqvrmw56).

Acantharia were collected to investigate their carbon and nitrogen uptake capability and nutrient transfer dynamics between symbiont and host using molecular and chemical imaging techniques. Isolated single cells were transferred to experimental conditions and incubated in a controlled environment of 23 °C with light at 191-194 µmol photons $m^{-2}$ $s^{-1}$. Experimental treatments consisted of dark and light incubations of specimen in filter-sterilized seawater (FSW, 0.22-µm-pore-size) enriched with $^{13}C$ and/or $^{15}N$ (either nitrate or ammonium). The nutrient treatments are as follows: 1) 1 mM $NaH^{13}CO_3$; 2) 1 mM $NaH^{13}CO_3$ + 1 µM $^{15}NO_3$; or 3) 1 mM $NaH^{13}CO_3$ + 0.4 µM $^{15}NH_4$. Acantharia were incubated for five different durations (Figure 5), each time-point consisting of a separate incubation. For each treatment and time point 30 cells were incubated. At T4 the cells were transferred to fresh FSW, through an intermediate transfer step in FSW (by means of rinsing), to start the chase incubation period. Between 15 and 21 Acantharia were chemically fixed at each time point in 1% glutaraldehyde + 2% paraformaldehyde, for subsequent NanoSIMS analysis. Simultaneously, at each timepoint 7 Acantharia (in a max 5 µL volume of SW) were individually deposited in 100 µL lysis buffer (RNAqueous kit), immediately flash frozen in liquid nitrogen, and stored at -80 °C for transcriptomics analyses. Detailed protocols for the single-cell RNA isolation and library preparation are available at protocol.io (see Mansour et al., 2021a and Mansour et al., 2021b)

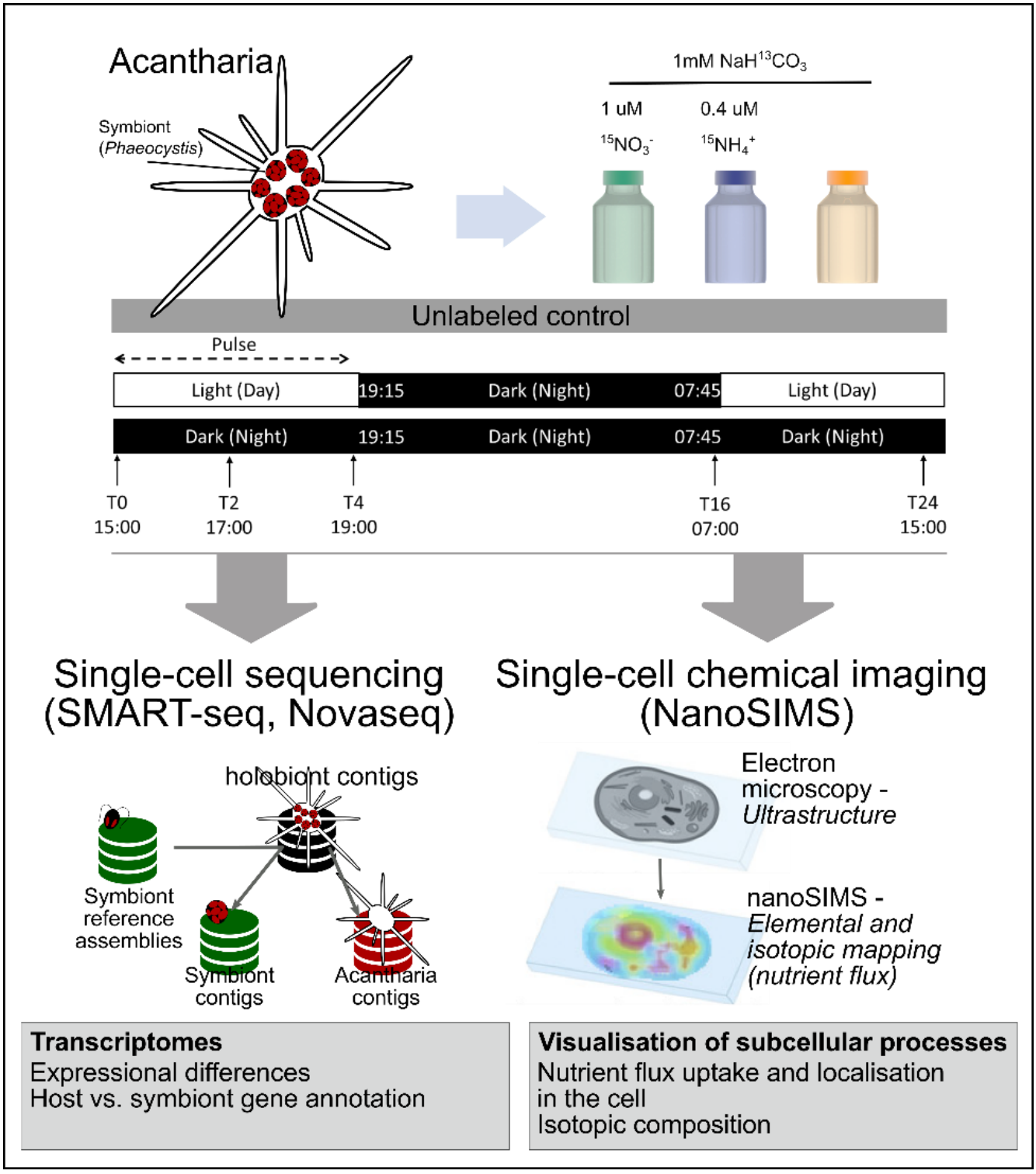



*Figure 5. Experimental setup to investigate the effect of inorganic nutrients on Acantharia. Related gene expression can be linked to visualised (NanoSIMS) uptake of nutrients.*

# 4 Linking genomic data to the experimental framework

## *4.1* Toxin production and molecular trade-offs in *Prymnesium parvum*

The role of polyketide synthases in the biosynthesis of marine secondary metabolites has been an emerging topic in marine molecular ecology. Dinoflagellates are the main protist group has been studied for the molecular basis of toxin biosynthesis, and has shed light on the diversity and presence of PKSs in this potentially toxic and phylogenetically diverse group (Monroe and Van Dolah, 2008; Kohli et al., 2016). Toxin production and uptake in *P. parvum* are strictly connected as toxins facilitate prey lysis and thus connected to its mixotrophic behaviour (Skovgaard and Hansen, 2003; Tillmann, 2003).

In *P. parvum*, the total number of PKS-related contigs across the nine strains varied from 37 to 109. The KS domain is highly conserved was used for phylogenetic purposes (John et al., 2008). The resulting phylogeny is in accordance with previous studies that involved KS domains from a wide range of protists (John et al., 2008; Kohli et al., 2016). The *P. parvum* ketoacyl synthase (KS) domains formed a well-supported haptophyte specific clade. Within this group, the *P. parvum* KS domains dominated in three clades. Two of those clades, A and C, were well supported with both of them having a bootstrap value of 100. Besides having *P. parvum* specific KS clades, no clear relation to the prymnesin type produced by the strains was found. This indicates the presence of phylogenetically similar core genes that are involved in the biosynthesis of prymnesins and are independent of the prymnesin type. Moreover, ‘unusual’ PKS domains were described and they highlight the high complexity of polyketide biosynthesis. These findings deserve more attention and need to be combined with biochemical approaches in order to gain

deeper understanding of gene to product relationships. The first step has been achieved by providing a detail description of all PKS contigs found in *P. parvum* as well as the functional organization of PKS domains.

The gradient of toxin content across the nine strains was compared to gene expression data, providing a direct link between a physiological parameter and transcriptomic data. The production of prymnesins imply cellular and metabolic cost *via* the investment of carbon and energy resources. The toxin content was used to explain the expression variance of 6,335 genes. The pathways of further interest involved in metabolic pathways, as they are the ones more related to toxin biosynthesis. A general downregulation of the cell's metabolism was observed, with 1,892 transcripts showing a negative correlation to toxin content and 631 transcripts showing a positive correlation (Anestis et al. 2021).

The ability to produce prymnesin characterizes all *P. parvum* strains studied up to date (Binzer et al., 2019). However, in the natural environment, within populations the presence of strains of various toxicity has ecological implications and attributes different advantages on individual cells and the collective levels, the population (Tillmann and Hansen, 2009, John et al. 2015). Highly toxic strains are important for outcompeting other organisms and to avoid grazers (Donk and Ianora, 2011). This has been shown for toxigenic species (John et al., 2002; Tillmann and Hansen, 2009) and a mutual facilitation of toxic strains for non-toxic ones has been postulated (John et al., 2015).

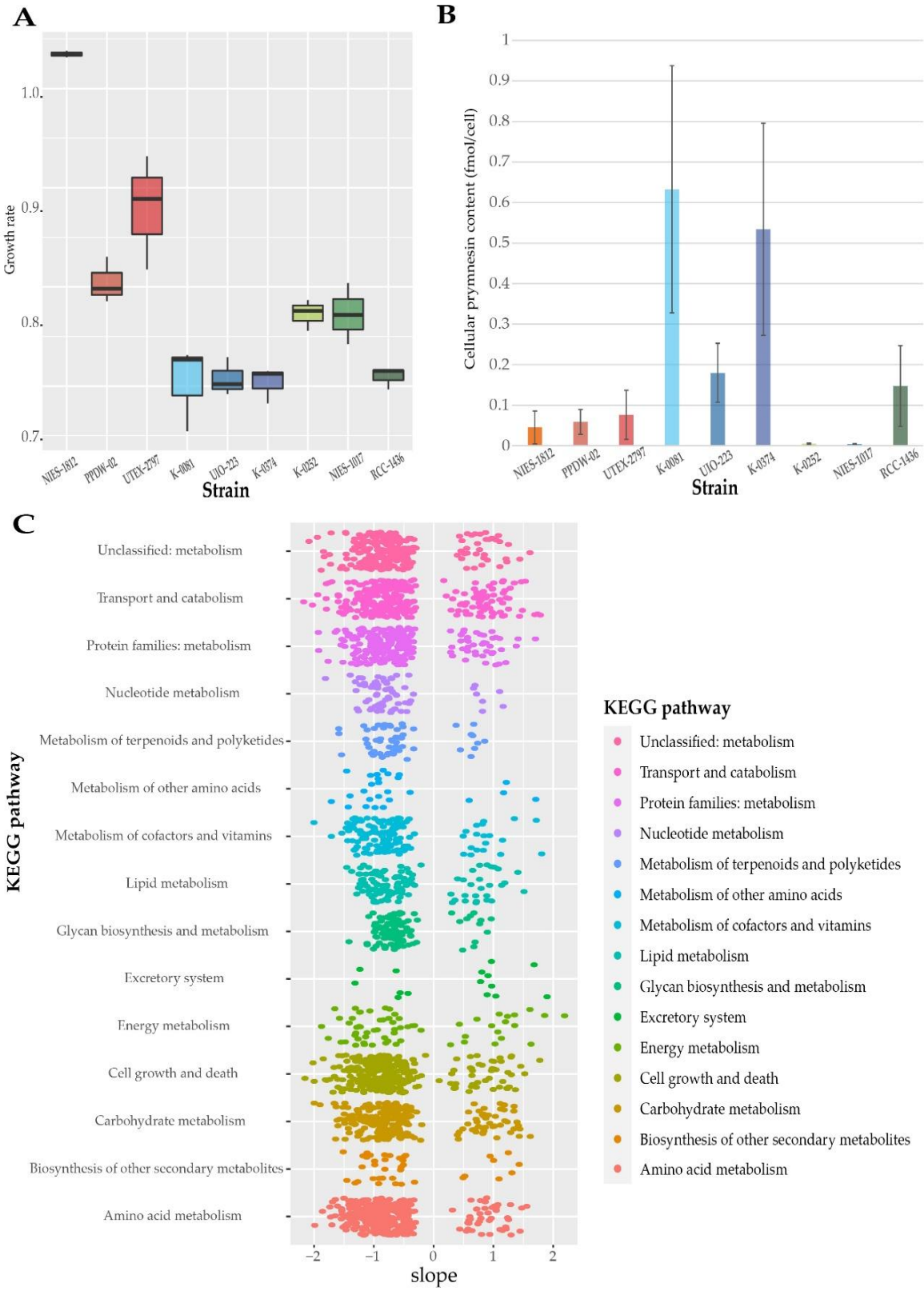



*Figure 6. Growth rates ($d^{-1}$) of the nine Prymnesium parvum strains (A) and their corresponding cellular toxin content expressed as fmol $cell^{-1}$ (B). Chart of transcripts for which gene expression variance was explained by the cellular toxin content (C). The x axis indicates the slope of the correlation with <0 showing a negative correlation and >0 a positive correlation. (Figure from Anestis et al., 2021 under CC BY-NC-ND 4.0.*

## *4.2* Eco-physiology of *Prymnesium parvum*

Studying phagotrophy in CM is challenging due to general low rates of phagotrophy. For example, in *P. parvum* cultures only a subset of cells feed on prey, which in some case can be 10-15% of the total cells. A *P*. *parvum* population is not homogeneous in regard to their response to the presence of prey, there can be cells that feed, cells that already contain a food vacuole (the digestion status can also be variable), and cells that do not show any response to the presence of prey. Applying batch culture approaches inevitably leads to an under representation of molecular pathways related to phagotrophy. To tackle this, it is important to incorporate novel methods that allow us to study transcriptomics of cells of a certain nutritional status, single-cell transcriptomics is a powerful tool towards this direction. We combined approaches that include both batch culture and single-cell transcriptomics (Figure 7). Library generation methodology of *P. parvum* single–cell total RNA is reported in Mansour et al., 2021a and Mansour et al., 2021b.

Transcriptomic data from single cells needs to be aligned with batch culture transcriptomes in order to compensate potential biased of single-cell transcriptomics over a general reference transcriptome representing the majority of expected genes of an organism. Such approaches have not been applied in protist research and can help answer many questions:

1. Can single-cell transcriptomics be insightful about molecular process in protists?
2. Is it able to compare single-cell and batch culture transcriptomes at level of molecular processes?
3. Are single-cell transcriptomics accurate enough and provide replicability of experiments and outcome?

4. What are the most efficient bioinformatics tools for analysing single-cell transcriptomic data?

Within MixITiN, we generated datasets that will combine traditional and novel transcriptomic approaches, with the objective to contribute to answering these technical and conceptual questions.

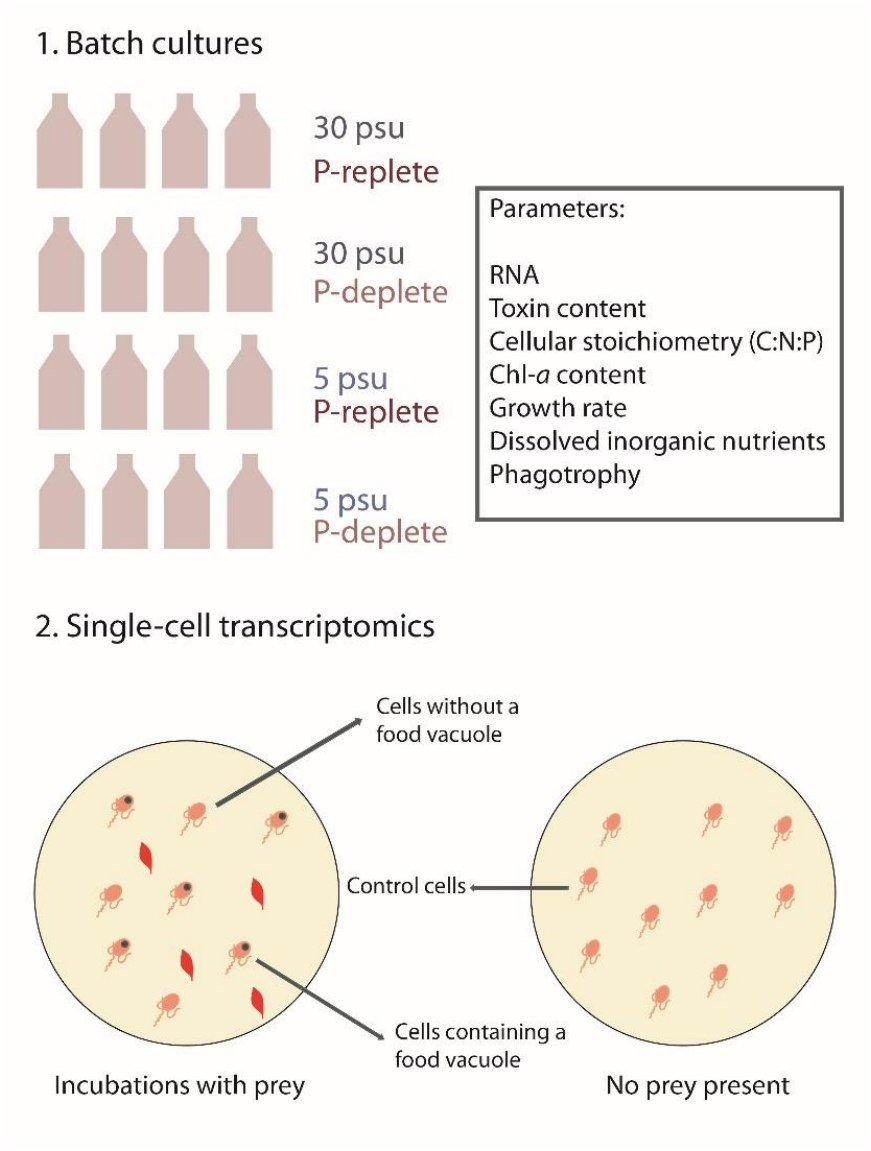



*Figure 7. Conceptual illustration of the approach followed for studying the eco-physiology of Prymnesium parvum. The conditions that were tested included different salinity (5 and 30) and phosphorus (P) availability (replete and deplete). The cultures were sampled for various parameters and single cell transcriptomics were used for selected cells of a certain nutritional status.*

## 4.3 Kleptoplastid dynamics in the mixotrophic *Strombidium*. cf. *basimorphum*

Unlike a permanent plastid, a stolen plastid or kleptoplast is transient. Retention time and functionality of the kleptoplasts are dynamic and can be from days to months. Both the ability of the host to control the upkeep of the plastid, and the inherent stability of the original plastid can affect the retention time (Green et al., 2005). Genomic approaches aim to help to disentangle heterotrophic processes such as phagotrophy from osmotrophy and phototrophic-induced anabolic activities. This information will then support cellular and metabolic modelling by better elucidating the physiological mechanisms and quantifying their importance in different scenarios.

Application of 'omics approaches to the groups of NCMs offers the potential to understand the evolutionary processes and establishment of permanent plastids. The increasing interest of the scientific community towards studying kleptoplastidic mixoplankton indicates the importance of such model organism in order to investigate the transition from heterotrophy to phototrophy with permanently established chloroplasts. General mechanistic understanding is needed to disentangle the dependency of the plastid from its own regulative processes and how the "host"/predator can control photosynthetic and other essential plastid processes for its own benefit. The presence of photosynthesis-related genes in kleptoplastidic mixoplankton could indicate a transition point towards heterotroph over mixo- to phototrophy, while the phylogenetic placement of those genes can provide crucial information about the selectivity and evolutionary trajectory of prey/plastid as well as the spectrum of preferred plastids by kleptoplastidic organisms (Hehenberger et al., 2019; Hongo et al., 2019).

This ciliate studied here (*S. basimorphum*) can maintain and keep chloroplasts functionality unaltered for several days when chloroplasts are not replaced via the

ingestion of prey. However, cellular chlorophyll content decreases to about half in starved cells compared to well fed cells. Despite that, *S. basimorphum* demonstrated to exploit chloroplasts for photosynthesis even more efficiently when starved of prey compared to when it is actively feeding (Figure 8).

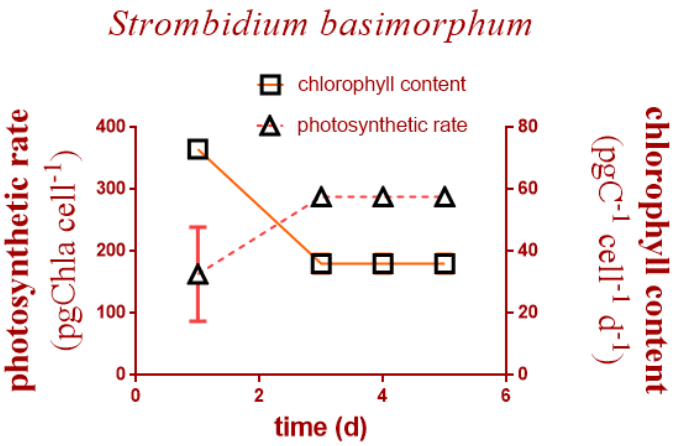



*Figure 8. Strombidium cf. basimorphum chlorophyll content and photosynthetic rates when actively feeding (day 1) and after prey depletion and starvation (subsequent time points; day 3, 4 and 5).*

Despite the fact that kleptoplasts seems to be better exploited when the ciliate does not ingest prey, *S.* cf *basimorhpum* is not able to survive as a pure autotroph in absence of prey and cultures decline in number as soon as prey get depleted (Figure 9).

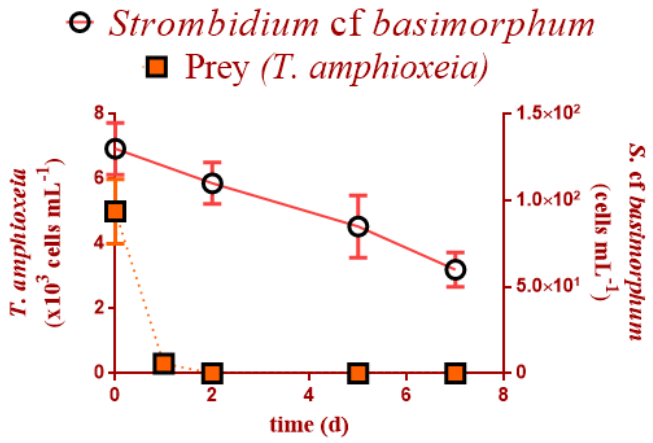



*Figure 9. Growth performance of Strombidium cf. basimorphum (circles) and Teleaulax amphioxeia (squares) over a seven day feeding experiment. S. cf. basimorphum decline in cell number right after prey depletion (day 2).*

Carbon acquired via photosynthesis is apparently not sufficient to sustain the energetic requirements of the ciliate which additionally looks to only rely on prey ingestion to obtain nutrients other than carbon. Indeed, no nitrogen or phosphorous uptake was observable in this species, as $NO_3^-$ and $PO_4^-$ concentration in the media didn´t change over time along the experiment (Figure 9).

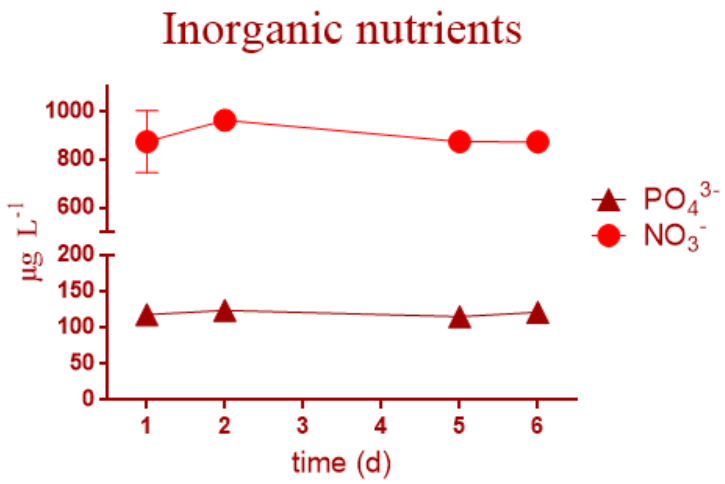



*Figure 10. Inorganic nutrients concentration (µg $L^{-1}$) during the experiment*

Laborious experimental work to understand physiological responses of a studied species towards different parameters are important prerequisites before proceeding to genomics-based studies. We used a wide range of molecular techniques to study kleptoplast and kleptokaryon retention in *S. basimorphum.*

Quantitative polymerase chain reaction assays can be used to detect the presence of prey genetic material in DNA extracted from mixoplanktonic grazers and provide a semi-quantitative estimation of its concentration (Figure 11). If performed on cultures with a known feeding history would give indication about the potential retention of prey genetic material, which could be relevant in understanding the molecular mechanisms that stands behind the retention of functional chloroplasts in NCM organisms. Prey nuclear and nucleomorph 28S rDNA were detected in the DNA extracted from the ciliate using the

qPCR assays. The relative concentration of these prey genes was lower in ciliates subjected to prey deprivation.

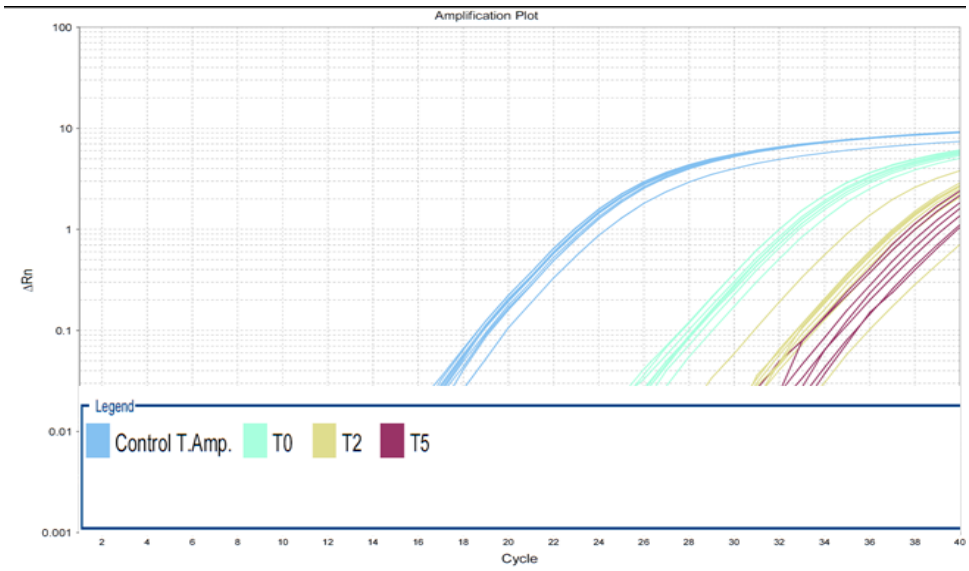



*Figure 11. Amplification plots of the qPCR assays on the Teleaulax amphioxeia nuclear 28S D2 USE b) on DNA extracted from T. amphioxeia (control) and Strombidium cf. basimorphum at different time points over prey deprivation and starvation.*

The fluorescent signals obtained upon hybridization of ciliates samples with a fluorescent in situ hybridisation (FISH) probe for prey rRNA seem comparable among individuals (single cells) sampled at different time points (Figure 12). Individual cells that contained the labelled genetic material as wells as individual cells that did not were found in all samples. The ribosomes observed via FISH could have been sequestered from the prey together with chloroplasts or actively transcribed from the prey nuclear gene. Fluorescent in situ hybridisation can be used for the same purpose as qPCR. However, different from qPCR, this technique does not allow the quantification of prey DNA in the sample. FISH is instead useful to discriminate for presence/absence of prey DNA within individual cells of the same population.

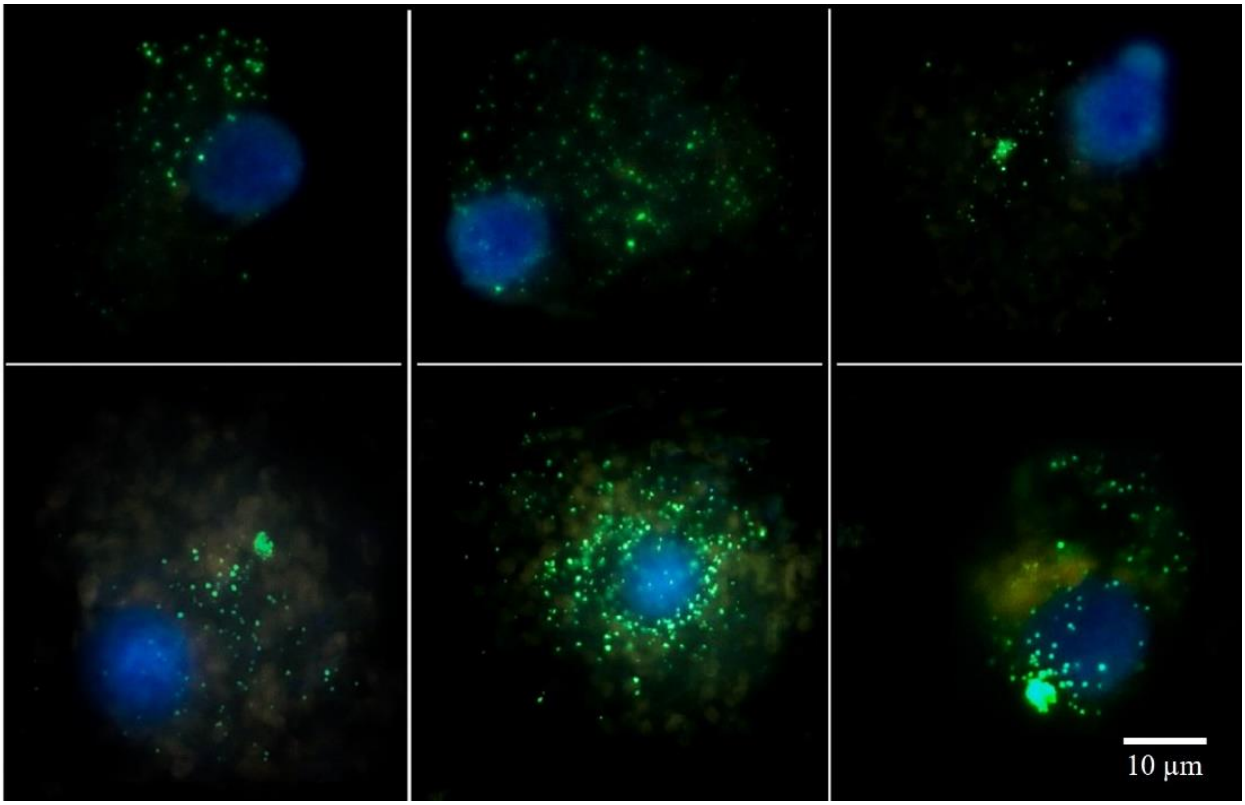



*Figure 12. Micrographs of Strombidium cf. basimorphum (ciliate) cells in different nutritional status; fed of Teleaulax (prey) (top: well fed, bottom: prey starved) hybridized with a probe for prey rRNA. Blue: ciliates nuclei (DAPI stained), orange: kleptoplasts, green: prey rRNA probe (Alexa488).*

Single-cell transcriptomics revealed the retention of *Teleaulax* genetic material by *S.* cf. *basimorphum.* A total of 282 transcripts of prey nuclear and chloroplast origin were present in starved cells. Among the 100 most expressed genes of prey origin there were transcripts encoded in the chloroplast genome (Figure 13). Chloroplast genes included photosystem I and II apoproteins, subunits and cytochromes. Moreover, we detected prey nuclear encoded genes involved in other metabolic processes such as amino acid biosynthesis and degradation. Genetic information pathways involved genes related to the transcription and translation of the prey nucleus within the host.

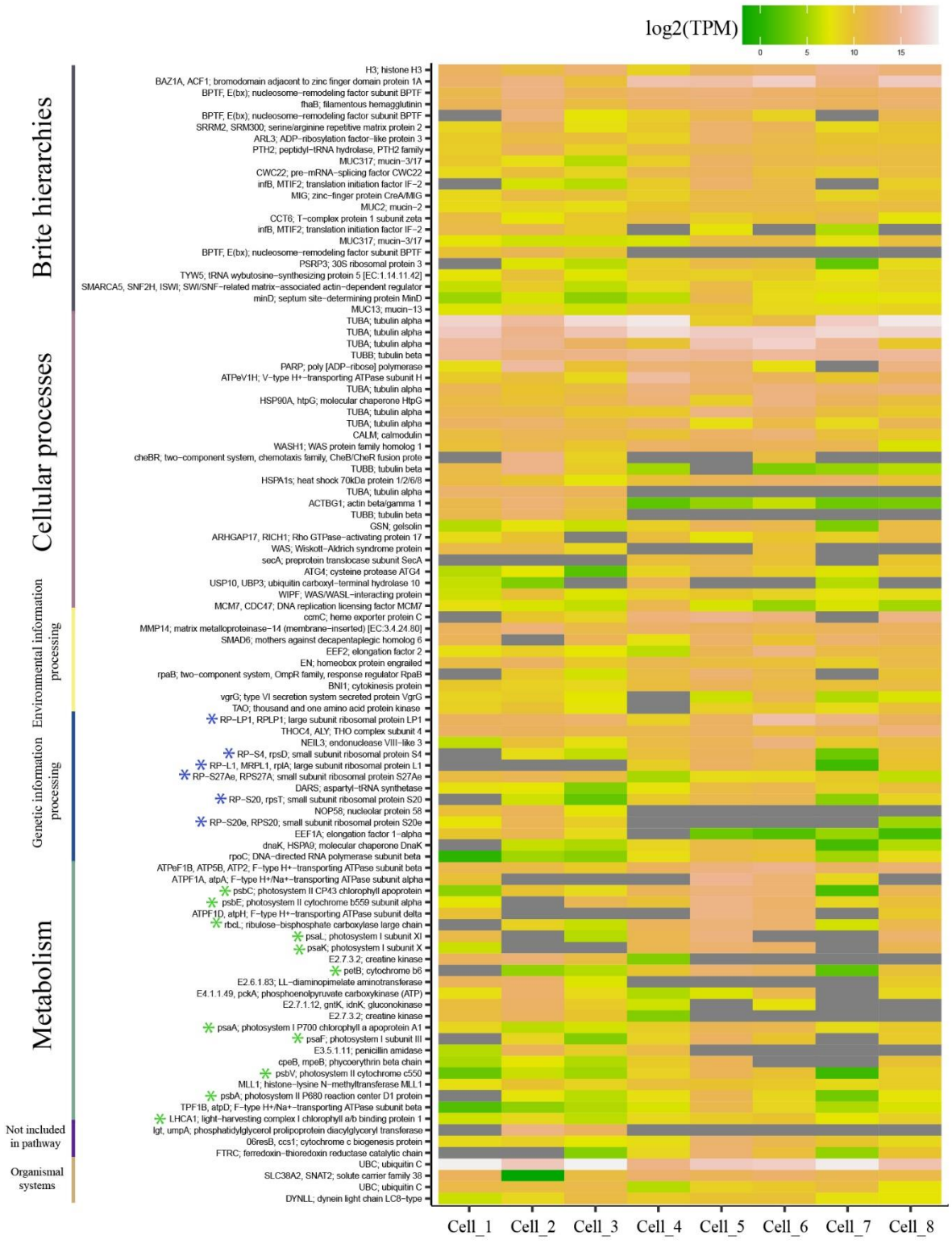



*Figure 13. Heatmap representing the 100 most expressed transcripts of prey origin and their corresponding pathways according to Kyoto Encyclopedia of Genes and Genomes. The expression values are provided for each separate cell and are shown as transcripts per million (TPM) (Reproduced from rom Maselli et al., 2021 under CC BY 4.0).*

Based on these observations, S*trombidium* cf *basimorphum* is evidently dependent on prey ingestion to sustain itself. However, the photosynthetic rates measured in the ciliate suggests potential for regulation of the photosynthetic activity of the sequestered chloroplasts in relation to the ciliate's nutritional status. Kleptoplastidic *Strombidium spp.* are thought to lack genes associated with chloroplasts functionality. Molecular methods are needed to investigate the potential role of genetic material of prey origin as well as the responses of the ciliate.

The retention of prey nuclei and the further analyses of their transcriptional activity within the ciliate will provide important information about the benefit of the ciliate by retaining the nuclei of the prey. The kleptoplastidic ciliate *Mesodinium sp.* does not only steal the plastid (kleptoplasts) from its cryptophyte prey but also the nucleus (kleptokaryon; Hansen et al., 2013, 2016). The kleptokaryon is transcriptionally active and could account for approximately half of the total transcriptome of the ciliate (Altenburger et al., 2020). The kleptokaryon, in this case, expresses genes for both maintaining the chloroplasts and synthesizing metabolites for which *Mesodinium* lacks the genetic toolkit (Lasek-Nesselquist et al., 2015; Kim et al., 2016). However, no photosynthesis-related genes were found to be transcribed by the genome of the ciliate (Altenburger et al., 2020).

*Dinophysis* (Dinoflagellata) are known to sequester plastids of cryptophyte origin, but not the nucleus (Park et al., 2014). *Dinophysis* harbour several nuclear transcripts which are involved in photosynthesis-related processes including plastid maintenance or pigment biosynthesis (Hongo et al., 2019).

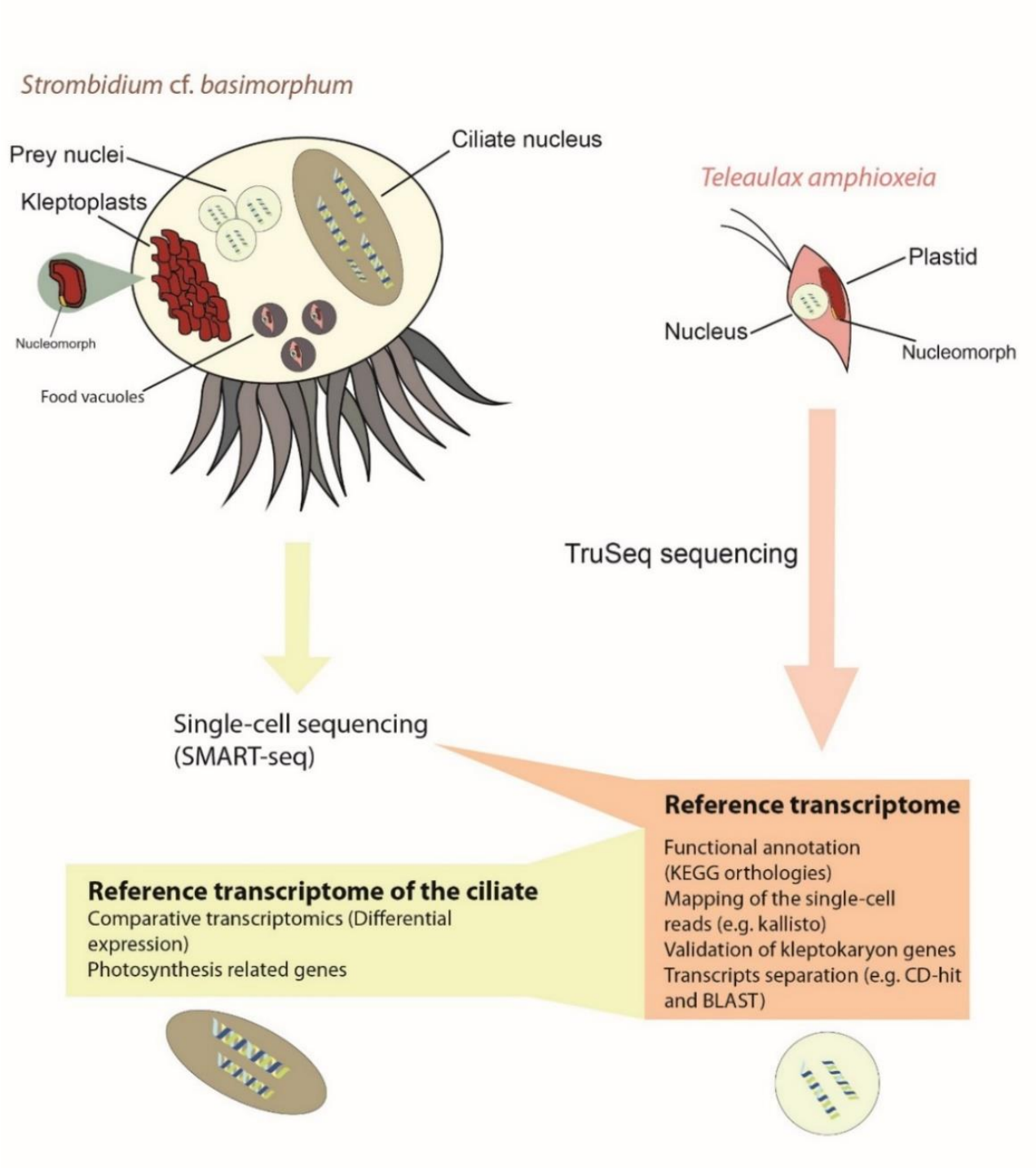



*Figure 14. Conceptual illustration of the approach followed for elucidating molecular processes in Strombidium cf. basimorphum feeding on the cryptophyte Teleaulax amphioxeia.*

*S*. cf. b*asimorphum,* apart from retaining prey nuclei, it also has genes related to photosynthetic processes. We detected genes related to photosynthetic pathways. Two isoforms of PetH (ferredoxin--NADP+ reductase) were present in almost all cells (46 out

of 47 cells) and showed high expression values. This gene is part of the photosynthetic electron transport system and its presence indicates that they facilitate the maintenance of the kleptoplasts. A gene involved in electron transport, PetF (ferredoxin), was also present alongside petC (cytochrome b6-f complex iron-sulfur subunit) which is part of the cytochrome b6/f complex. The photosystem II cytochrome b559 subunit alpha (psbE) and photosysytem II reaction center protein K (psbK) were also present in the ciliate transcriptome but both detected in only 2 cells, indicating low transcription level of these proteins. However, the rest of the genes involved in these pathways were not present in the transcriptome of the ciliate, which could be result of either absence from the ciliate genome or low transcription levels.

Multiple genes involved in porphyrin and chlorophyll metabolism are also present. These genes involve magnesium chelatase subunits. We found genes encoding for chelatase subunit D, G and H (chlD, chlD, chlG), which were present in multiple cells and with high expression values. Genes for other chelatase subunits were also found but their expression levels were very low. However, it confirmed their presence in the transcriptome of the ciliate.

All these findings are similar to the case of *Dinophysis*, which does not retain prey nuclei itself, but does inherently harbour photosynthesis related genes (Hongo et al., 2019). *S.* cf. *basimorphum* could thus similarly be considered an intermediate case, given its ability to use the kleptokaryon of the prey and at the same time own photosynthetic genes which are connected to the functionality of the kleptoplasts.

## 4.4 Acantharian nutrient acquisition capabilities

A total of 2.1 billion 2x100bp paired-end reads were generated by Novaseq S1 flowcell sequencing. 19.4 to 116.2 million sequence reads per sample. Bioinformatic quality

filtering with Trimmomatic removed only 0.003% read pairs, SortmeRNA filtered out 328 million reads (15%) leaving 1.8 billion prospected paired mRNA reads, 4.8 – 70.2 million read per sample. *De novo* assembly generated 2,108,782 transcripts totalling 2,649,628,289 bp, with a N50 of 2073 and an Ex90N50 of 3200. The GC content of the transcripts exhibited a bimodial distribution (Figure 15), low GC% transcripts are assumed to be transcripts from the Acantharia host, while the high GC% transcripts from the algal symbiont (Liu et al., 2019). The holobiont transcriptome assembly includes complete sequences for 96.9% of searched eukaryotic BUSCOs. More complete transcriptomes contain more full-length BUSCOs, which are well-conserved genes for the representative genomes of each BUSCO group. BUSCOs provide a method to quantitatively assess the quality of a transcriptome in terms of gene content.

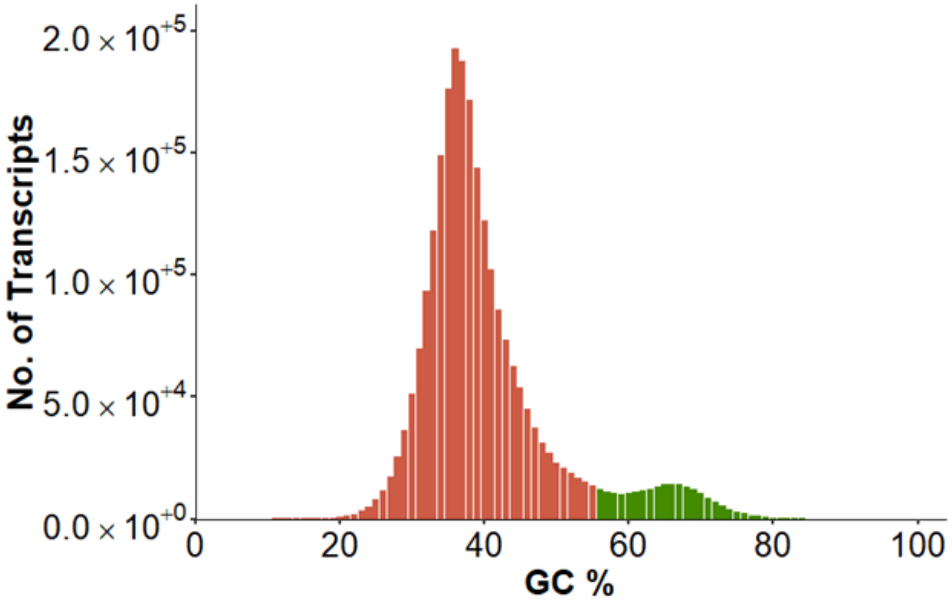



*Figure 15. GC content of the Acantharia holobiont de-novo assembly, indicated in red and green, respectively, are presumed Acantharia and symbiont GC% ranges.*

All the assembled sequences were searched against a custom database containing four of *Phaeocytis* transcriptomes (i.e. METDB_00327, METDB_00333, METDB_00329, and https://github.com/maggimars/PcordataSymbiosisDGE/blob/2b7836d2f6ebf9a5d80cbcc

e22988f161428e459/pc_euk_seqs.fasta) using MegaBLAST. All hits with bitscores > 90 were considered *Phaeocytis* transcripts. The hits were separated out of our assembly. Only using the Blast results to filter out the symbiont of the holobiont transcriptome results in a GC-content profile for the presumed host transcripts that still contains a small high-GC-content tail. For that reason, we additionally use a Gaussian Mixed Model (GMM) to filter out those contigs, i.e. contigs with 99% certainty of belonging to the high GC content model. The GC% of the transcripts assigned to either the symbiont or host is shown in Figure 16. The split assembly had 91% and 89% complete eukaryotic BUSCOs for the contigs assigned to *Phaeocystis* and those assigned to Acantharia respectively.

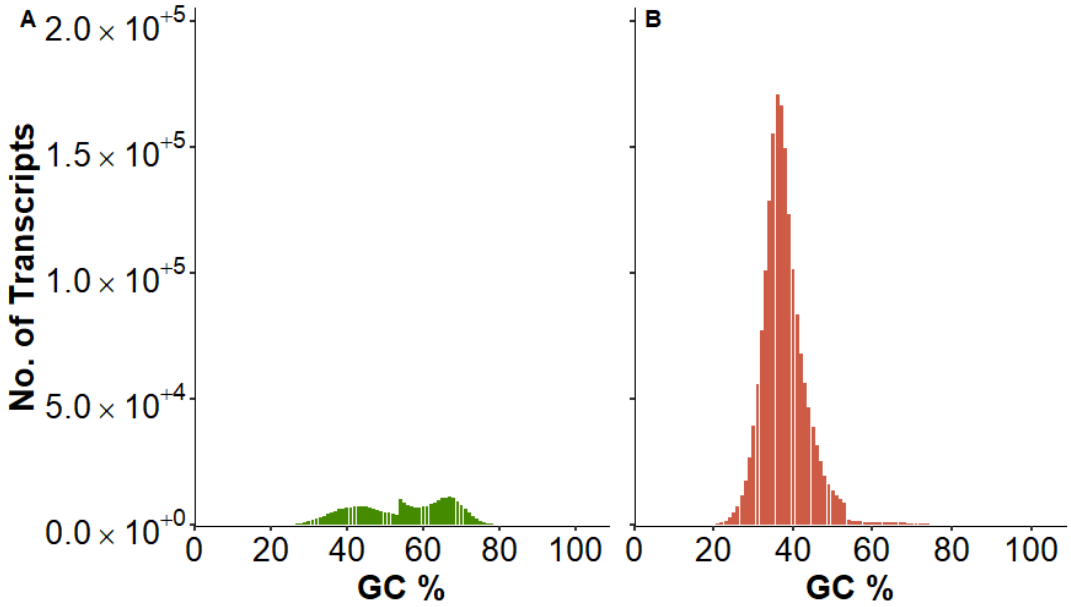



*Figure 16. GC content of the de novo Assembly split into (**A**) Phaeocytis contigs, and (**B**) Non-Phaeocytis (i.e. presumed Acantharia contigs), by means of BLASTn and GMM results.*

Our combined experimental setup for transcriptomics and chemical imaging aimed to corroborate the capabilities in C and N uptake of Acantharia from both a molecular and physiological angle (Figure 5). With the chemical imaging approach, we aimed to follow nutrient transport and localization over time using stable isotopes and NanoSIMS. Thereby we aimed to visualize and quantify carbon uptake, incorporation, and photosynthate translocation between symbionts and host over time, and the effect of

nitrogen ($NO_3^-$ or $NH_4^+$) thereon. The transcriptomics approach is utilised to assess gene expressional difference under the same treatments and time series. Thereby investigating transcriptional difference, adaptability/plasticity of nutrient uptake when increased nutrients are present (eutrophication). As well as, N Metabolism adjustment and capabilities of symbiont and/or hosts). Separation of the holobiont transcriptome allows us additionally to detangle specific processes, such as those involved in carbon and nitrogen metabolism, that are due to host or symbiont transcriptional activity.

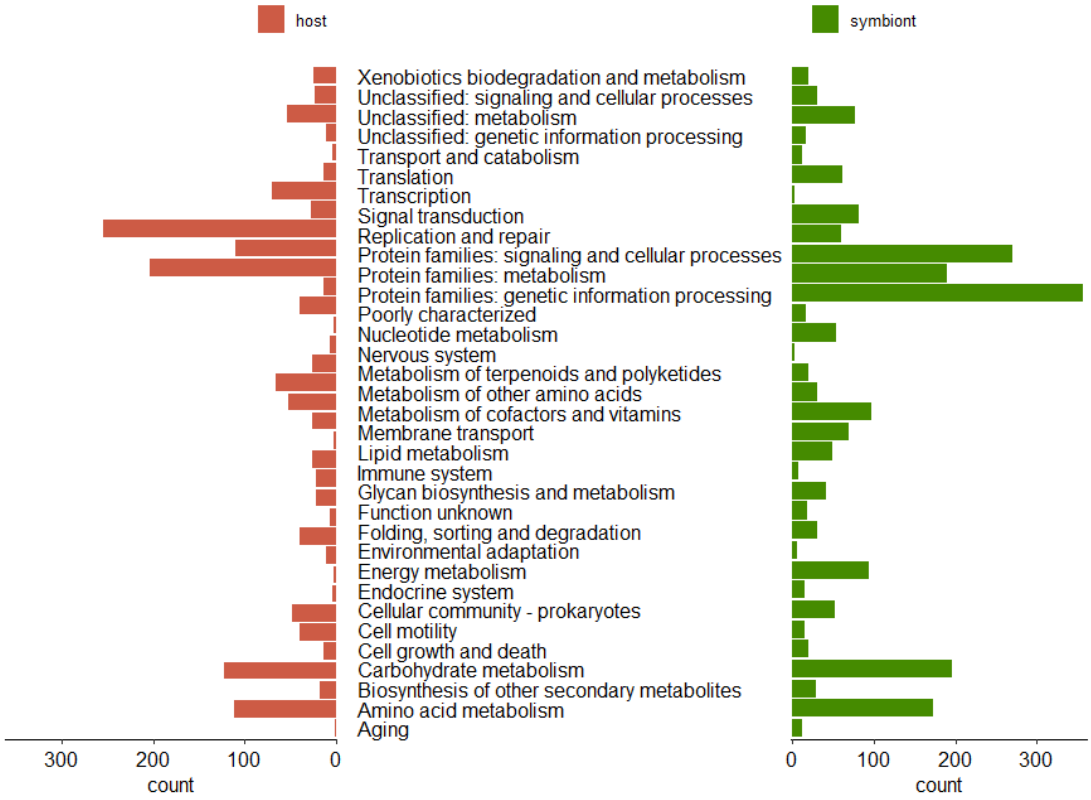



*Figure 17. K- Number annotated transcript counts for symbiont and host transcripts and the relevant KEGG pathways.*

For example, a preliminary investigation of KEGG pathway annotations indicated that most transcripts associated with energy metabolism are linked to the symbiont partition of the transcriptome. Notably, the only nitrogen metabolism enzymes linked to the host are involved in ammonium metabolism, i.e. glutamate dehydrogenase and glutamine

synthetase, whereas enzymes involved in nitrate metabolism are also found for the symbiont. The number of unique KEGG annotation for the symbiont transcriptome is, however, much more than for the host, 2214 versus 1511 KEGGs for symbiont and host respectively (**Figure 17)**. Even though, the host transcriptome is considerably larger with 1,803,192 transcripts to 305,580 transcripts of the symbiont, a lower annotation rate was indeed expected the host transcriptome. Nonetheless, such indications of absence or presence of enzymatic pathways allow us to hypothesis the role of the host in said pathway, and combine it with the interpretation of chemical imaging for better interpretation.

## 4.5 Potential for integrating genomic information into explicit mathematical models

One major goal of the project was to develop novel tools for interpreting the degree of mixotrophy in populations by integrating physiological and genomic data linked with individual cell-based models. Batch culture experiments and single-cell analytical approaches were applied to investigate key physiological processes associated with phagotrophy and/or kleptoplasty with time-series resolution. Physiological parameters, such as rates of feeding, growth, photosynthesis, and stoichiometric relationships were determined and interpreted with respect to detailed patterns of gene expression. The data from physiological and the grazer/prey interaction experiments, i.e., with the haptophyte CM *Prymnesium* and to a lesser extent with the NCM ciliate *Strombidium,* and eSNCM endosymbiotic Acantharia were analyzed at the metabolic process level generated from the functionally annotated genes and their expression patterns. The first level functional annotation analysis reveals only instantaneous metabolic activity (“transient physiological status”) but does not capture the dynamics of the metabolic processes necessary for modelling rates of gene expression. In essence, such genomic data are descriptive and hypothesis building rather than revealing the quantitative shifts in rates for the processes

under study. The crucial steps for parameterizing the data for direct incorporation into dynamic mathematical models could therefore not be achieved. The exact recording of the change from one physiological status to the next and regulatory mechanisms underlying such processes have been done.. The time-dependence linking the rates of physiological processes and the respective gene expression can be reflected in the transcriptome, but this level of analysis could not be achieved within the scope of this project. Only the unambiguous linkage of physiology, genomics and time will provide the data matrix to complement the modelling of mixotrophy in the target organisms. Nevertheless, the goals of using the genomic data from CM as well as NCMs to separate cellular physiological from evolutionary mechanisms and processes were partially achieved in the project, e.g., by revealing the active prey nucleus of *Teleaulax* in *Strombidium* via the starvation experiments. Furthermore, first results on the decoupling of eco-physiological from evolutionary processes was achieved with *Prymnesium*, whereby key processes for mixotrophy, like phagotrophy, lysosome production, toxicity induction, etc. were inducible at the cellular level. Inducibility of "mixotrophy" was also possible (under certain environmental stimuli) when no food was offered. This inducible response to various stimuli (e.g. nutrients and salinity gradients), even partially without the key stimulus of food suggests an evolutionary process that persists even in cultures after several years without mixotropic nutrition (Anestis et al., in prep).

# 5 Conclusions and future perspectives

In this project, we were able to show that the combination of sophisticated cultivation approaches from batch culture to single cell approaches, together with genomics and tuned bioinformatics tools, can provide insights into biosynthetic pathways and metabolic trade-offs of representative mixoplankton species. Transcriptomic data was aligned with parameters such as toxin content in the CM *P. parvum,* providing insights about molecular

metabolic trade-offs of toxin biosynthesis. Moreover, the experimental planning and data obtained to investigate physiology and mixotrophy in *P. parvum* will provide substantial information about the efficiency of using novel approaches to study mixotrophy as well as further use for modelling purpose.

We were able to investigate the process of prey digestion and kleptoplast enslavement in *S.* cf *basimorphum*. In the future, the data will need to be further analysed for metabolic and enzymatic processes to support modelling approaches with biosynthetic details and potential rates for cellular processes. On a higher level, applying our approaches to more different species of mixoplankton will elucidate the evolutionary pathways from CMs to GNCMs to SNCMs and probably finally to an endosymbiotic/permanent plastid.

# 6 References

Altenburger, A., Cai, H., Li, Q., Drumm, K., Kim, M., Zhu, Y., et al. (2020). Limits to the cellular control of sequestered cryptophyte prey in the marine ciliate *Mesodinium rubrum*. ISME J. 15, 1056–1072.

Anestis, K., Kohli, G.S., Wohlrab, S., Varga, E., Larsen, T.O., Hansen, P.J., John, U. (2021). Polyketide synthase genes and molecular trade-offs in the ichthyotoxic species *Prymnesium parvum*. Sci. Total Environ. 795.

Andrews, S. (2010). FastQC A quality control tool for high throughput sequence data. Babraham Bioinforma.

Binzer, S.B., Svenssen, D.K., Daugbjerg, N., Alves-de-Souza, C., Pinto, E., Hansen, P.J., et al.,2019. A-, B- and C-type prymnesins are clade specific compounds and chemotaxo-nomic markers in *Prymnesium parvum*. Harmful Algae 81, 10–1

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., et al. (2017). A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. Cell Rep. 18, 762–776.

Burns, J. A., Pittis, A. A., and Kim, E. (2018). Gene-based predictive models of trophic modes suggest Asgard archaea are not phagocytotic. Nat. Ecol. Evol. 2, 697–704.

Donk, E. Van, Ianora, A., 2011. Induced defences in marine and freshwater phytoplankton:a review. Hydrobiologia 668, 3–19.

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048.

Faure, E., Not, F., Benoiston, A.-S., Labadie, K., Bittner, L., and Ayata, S.-D. (2019). Mixotrophic protists display contrasted biogeographies in the global ocean. ISME J. 13, 1072–1083.

Flynn, K. J., Mitra, A., Anestis, K., Anschütz, A. A., Calbet, A., Ferreira, G. D., et al. (2019). Mixotrophic protists and a new paradigm for marine ecology: where does plankton research go now? J. Plankton Res.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652.

Graham, L. B., Colburn, A. D., and Burke, J. C. (1976). A new, simple method for gently collecting planktonic protozoa. Limnol. Oceanogr. 21, 336–341.

Green, B. J., Fox, T. C., and Rumpho, M. E. (2005). Stability of isolated algal chloroplasts that participate in a unique mollusc/kleptoplast association. Symbiosis 40, 31–40.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512.

Hansen, P. J., Nielsen, L. T., Johnson, M., Berge, T., and Flynn, K. J. (2013). Acquired phototrophy in *Mesodinium* and *Dinophysis* - a review of cellular organization, prey selectivity, nutrient uptake and bioenergetics. Harmful Algae 28, 126–139.

Hansen, P. J., Ojamäe, K., Berge, T., Trampe, E. C. L., Nielsen, L. T., Lips, I., et al. (2016). Photoregulation in a kleptochloroplastidic dinoflagellate, *Dinophysis acuta*. Front. Microbiol. 7:785.

Hehenberger, E., Gast, R. J., and Keeling, P. J. (2019). A kleptoplastidic dinoflagellate and the tipping point between transient and fully integrated plastid endosymbiosis. Proc. Natl. Acad. Sci. U. S. A. 116, 17934–17942.

Heinbokel, J. F., Diego, S. and Jolla, L. (1978) Studies on the functional role of tintinnids in the Southern California Bight. I. Grazing and growth rates in laboratory cultures. Mar. Biol., 189, 177–189.

Hongo, Y., Yabuki, A., Fujikura, K., and Nagai, S. (2019). Genes functioned in kleptoplastids of *Dinophysis* are derived from haptophytes rather than from cryptophytes. Sci. Rep. 9, 1–11.

Jimenez, V., Burns, J., Le Gall, F., Not, F., and Vaulot, D. (2020). No evidence of phago-

mixotropy in *Micromonas polaris*, the dominant picophytoplankton species in the Arctic. J. Phycol. 57, 435-446.

John, U., Tillmann, U., Medlin, L.K., 2002. A comparative approach to study inhibition ofgrazing and lipid composition of a toxic and non-toxic clone of *Chrysochromulina polylepis* (Prymnesiophyceae). Harmful Algae 1, 45–57

John, U., Tillmann, U., Hülskötter, J., Alpermann, T.J., Wohlrab, S., Van de Waal, D.B., 2015.Intraspecific facilitation by allelochemical mediated grazing protection within a toxi-genic dinoflagellate population. Proc. R. Soc. B Biol. Sci. 282

Katoh, K., Kuma, K. I., Toh, H., and Miyata, T. (2005). MAFFT version 5: Improvement in accuracy of multiple sequence alignment. Nucleic Acids Res. 33, 511–518.

Kim, G. H., Han, J. H., Kim, B., Han, J. W., Nam, S. W., Shin, W., et al. (2016). Cryptophyte gene regulation in the kleptoplastidic, karyokleptic ciliate *Mesodinium rubrum*. Harmful Algae 52, 23–33.

Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28, 3211–3217.

Lasek-Nesselquist, E., Wisecaver, J. H., Hackett, J. D., and Johnson, M. D. (2015). Insights into transcriptional changes that accompany organelle sequestration from the stolen nucleus of *Mesodinium rubrum*. BMC Genomics16:805.

Leles, S. G., Mitra, A., Flynn, K. J., Stoecker, D. K., Hansen, P. J., Calbet, A., et al. (2017). Oceanic protists with different forms of acquired phototrophy display contrasting biogeographies and abundance. Proc. R. Soc. B Biol. Sci. 284, 20170664.

Leles, S. G., Mitra, A., Flynn, K. J., Tillmann, U., Stoecker, D., Jeong, H. J., et al. (2019). Sampling bias misrepresents the biogeographical significance of constitutive mixotrophs across global oceans. Glob. Ecol. Biogeogr. 28, 418–428.

Liu, Z., Hu, S. K., Campbell, V., Tatters, A. O., Heidelberg, K. B., and Caron, D. A. (2017). Single-cell transcriptomics of small microbial eukaryotes: Limitations and potential. ISME J. 11, 1282–1285.

Mansour, J., Anestis, K., Not, F., and John, U. (2021a). cDNA library preparation from total RNA extracts of Single-cell marine protists (e.g. Acantharia, *Strombidium basimorphum*, and *Prymnesium parvum*) for transcriptome sequencing V.2. *protocols.io*.

Mansour, J., Anestis, K., Not, F., and John, U. (2021b). Single-cell total RNA extraction from marine protists (e.g. Acantharia, *Strombidium* cf *basimorphum*, and *Prymnesium parvum*) V.2.

Mansour, J. S., Norlin, A., Llopis Monferrer, N., L'Helguen, S., and Not, F. (2021c). Carbon and nitrogen content to biovolume relationships for marine protist of the Rhizaria lineage (Radiolaria and Phaeodaria). Limnol. Oceanogr. 66, 1703–1717.

Mansour, J.S., and Anestis, K. (2021d). Eco-evolutionary perspectives on mixoplankton. Front. Mar. Sci.

Maselli, M., Anestis, K., Klemm, K., Hansen, P.J., John, U. (2021). Retention of prey genetic material by the kleptoplastidic ciliate *Strombidium* cf. *basimorphum*. Front. Microbiol.

Mitra, A., Flynn, K. J., Tillmann, U., Raven, J. A., Caron, D., Stoecker, D. K., et al. (2016). Defining planktonic protist functional groups on mechanisms for energy and nutrient acquisition: Incorporation of diverse mixotrophic strategies. Protist 167, 106–120.

Park, M. G., Kim, M., and Kim, S. (2014). The acquisition of plastids/phototrophy in heterotrophic dinoflagellates. Acta Protozool. 53, 39–50.

Rivkin, R. B. and Seliger, H. H. (1981) Liquid scintillation counting for 14C uptake of single algal cells isolated from natural samples. Limnol. Oceanogr., 26, 780–785.

Schoener, D. M., and McManus, G. B. (2012). Plastid retention, use, and replacement in a kleptoplastidic ciliate. Aquat. Microb. Ecol. 67, 177–187.

Song, Y., Milon, B., Ott, S., Zhao, X., Sadzewicz, L., Shetty, A., et al. (2018). A comparative analysis of library prep approaches for sequencing low input translatome samples. BMC Genomics 19, 1–16.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Steemann Nielsen, E. 1952. The use of radioactive carbon (C14) for measuring organic production in the sea. J. Conseil, Perm. Intern. Exploration Mer, 18: 117-140.

Stoecker, D. K., Silver, M. W., Michaels, A. E., and Davis, L. H. (1988). Obligate mixotrophy in *Laboea strobila*, a ciliate which retains chloroplasts. Mar. Biol. 99, 415–423.

Stoecker, D. K., Johnson, M. D., De Vargas, C., and Not, F. (2009). Acquired phototrophy in aquatic protists. Aquat. Microb. Ecol. 57, 279–310.

Tillmann, U., Hansen, P.J., 2009. Allelopathic effects of *Alexandrium tamarense* on otheralgae: evidence from mixed growth experiments. Aquat. Microb. Ecol. 57, 101–112.