The Ecology and Glycobiology of *Prymnesium parvum*

Ben Adam Wagstaff

This thesis is submitted in fulfilment of the requirements of the degree of Doctor of Philosophy at the University of East Anglia

Department of Biological Chemistry

John Innes Centre

Norwich

September 2017

Abstract

Prymnesium parvum is a toxin-producing haptophyte that causes harmful algal blooms (HABs) globally, leading to large scale fish kills that have severe ecological and economic implications.

A HAB on the Norfolk Broads, U.K, in 2015 caused the deaths of thousands of fish. Using optical microscopy and 16S rRNA gene sequencing of water samples, *P. parvum* was shown to dominate the microbial community during the fish-kill. Using liquid chromatography-mass spectrometry (LC-MS), the ladder-frame polyether prymnesin-B1 was detected in natural water samples for the first time. Furthermore, prymnesin-B1 was detected in the gill tissue of a deceased pike (*Exos lucius*) taken from the site of the bloom; clearing up literature doubt on the biologically relevant toxins and their targets.

Using microscopy, natural *P. parvum* populations from Hickling Broad were shown to be infected by a virus during the fish-kill. A new species of lytic virus that infects *P. parvum* was subsequently isolated, *Prymnesium parvum* DNA virus (PpDNAV-BW1). Morphological analysis and genome sequencing revealed PpDNAV-BW1 to belong to the Megaviridae family of algal viruses. We propose that viral lysis of *P. parvum* may act as a novel release mechanism for intracellular toxins.

The sialic acid, 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid (KDN) has recently been shown to be important in viral infections of microalgae. LC-MS was used to demonstrate that *P. parvum* contains KDN. Candidate sequences for KDN biosynthesis from *P. parvum* were cloned and expressed and shown to produce cytidine monophosphate-activated KDN (CMP-KDN). Using the newly characterized sequences in BLASTp analysis, we revealed that sialic acid biosynthesis is widespread amongst algae.

Using bioinformatics, NDP-β-L-rhamnose biosynthesis was explored in *P. parvum* and across other algae. We propose that the haptophytes have acquired bacterial TDP-β-L-rhamnose biosynthetic genes from horizontal gene transfer and subsequently passed them on to some dinoflagellate species. Sugar-nucleotide profiling of two representative algae support this proposition.

List of Contents

List of Figures

List of Tables

List of Supplementary Material

Acknowledgements

First, I would like to thank Rob Field for believing in my ability when I was a hapless undergraduate and offering me a PhD position. His guidance and support over the years have been invaluable and it is never a wrong time of the day to send him an email. I would particularly like to thank Rob for his help and support with scientific writing and writing grant proposals - I am sure I could not have learnt these skills alone. Most importantly though, I would like to thank Rob for his trust in me and his willingness to let me do the science I wanted to do.

Thanks to my secondary supervisor, Gill Malin, for her useful discussions surrounding phycology and algal virology and for the kind use of her lab whenever I needed it. A huge thank you to my collaborator from UEA, Jennifer Pratscher, who became my '*environmental microbiologist in shining armour*' when she joined the *Prymnesium* project. Without her input the project would not be where it is today. I also greatly appreciate the input and useful discussions had with Colin Murrell during the project, who contributed both his time and resources to helping both myself and Jennifer Pratscher along the way.

Other collaborators I would like to thank include Declan Schroeder and Willie Wilson from Plymouth, who welcomed me to the algal virus community with open arms. Thank you to Diane Saunders from the Earlham Institute for her help with organising *Prymnesium* meetings and her support with sequencing.

I would like to thank all industrial partners on the *Prymnesium* project including members of the Environment Agency, Broads Authority, Fishtrack, Norfolk Wildlife Trust, Natural England, and many others, who all worked together to make the Norfolk Broads a better place. A particular thank you to Steve Lane of the Environment Agency and John Currie of the Pike Anglers Club who have worked tirelessly to make sure this project succeeds - this project wouldn't be where it is today without both of your input.

Thanks to all the great facilities teams here at the JIC for their help and support – Elaine Barclay and Kim Findlay from bioimaging and Gerhard Saalbach, Lionel Hill and Paul Brett from mass spectrometry. Thank you to all past and present members of the Field group over the last 4 years for making it such a fantastic environment to work in. The breadth of projects, willingness to share, and unrivalled openness of the group has given me great memories and

made me the scientist I am today. I must give a special mention to Martin Rejzek for his brilliant supervision, trust, and his scientific curiosity that sparked the initial idea for the *Prymnesium* project. I would also like to say a massive thank you to Mike Rugen for all his help with 'mikerubiology' techniques and his friendship throughout my PhD - you made it a more enjoyable place to work. Another special thank you to Ed Hems who has been a great friend and colleague throughout. Our skillsets and interests complimented each other brilliantly, and I could always rely on Ed for help with anything.

Thanks to all my friends from Ipswich who frequently made sure I drunk until I stopped talking about science when I was home! Thanks to my parents, Teresa, and Kevin, who must be the best teachers of all. I don't know what you did to raise me the way you did, but if I can repeat even a small part of it with my future children I will be a happy man. None of this would have been possible without your love, support, and patience with me throughout university and my PhD. Thank you to Chloe for always being such a caring and thoughtful sister, qualities you've clearly inherited from mum! Finally, thank you to Nicola for always making sure I got out of bed with a smile on my face during the write-up process, for her love and support, and for frequently setting an example to me of how incredibly hard a person can work.

Declaration of Authorship

I declare that the work contained in this thesis, submitted by me for the degree of Doctor of Philosophy, is to the best of my knowledge my own original work, except where due reference is made.

Date

26/09/2017

Signed

Ben Adam Wagstaff

Abbreviations

“E	grid reference - Eastings
“N	grid reference - Northings
µmol m^{-2} s^{-1}	photon flux
16S	16S ribosomal RNA
18S	18S ribosomal RNA
28S	28S ribosomal RNA
Å	Ångstrom
aa	amino acid
AaV	*Aureococcus anophagefferens* virus
AcN	acetonitrile
ADP	adenosine diphosphate
Agly	aglycone
APMV	*Acanthamoeba polyphaga mimivirus*
Ara	arabinose
-ax	axial
bp	base pair
BSE	barley straw extract
CASH	Cryptophyta, Alveolata, Stramenopila and Haptophyta
CCAP	culture collection of algae and protozoa
CeV	*Chrysochromulina ericina* virus
CMP	cytidine monophosphate
CroV	*Cafeteria roenbergensis* virus
CsCl	caesium chloride
CTP	cytidine triphosphate
d	doublet
dd	doublet of doublets
ddd	doublet of doublet of doublets
DMB	1,2-diamino-4,5-methylenedioxybenzene
DNA polB	DNA Polymerase Beta
DOM	dissolved organic matter
dsDNA	double stranded DNA
DTT	dithiothreitol
EDTA	ethylenediaminetetra acetic acid

EGT	endosymbiotic gene transfer
EhV	*Emiliania huxleyi* virus
ELA	erythrocyte lysis assay
-eq	equitorial
ESI	electrospray ionization
EtOAc	ethyl acetate
FAA	fatty acid amide
g	relative centrifugal force
Gal	galactose
GDP	guanosine diphosphate
Glc	glucose
GlcNAc	*N*-acetyl glucosamine
GlcNAcA	2-acetimido-2-deoxy-glucuronic acid
H_2O_2	hydrogen peroxide
H_3PO_4	phosphoric acid
$H_4P_2O_7$	pyrophosphoric acid
HAB	harmful algal bloom
HaV	*Heterosigma akashiwo* virus
HCl	hydrochloric acid
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HeV	*Haptolina ericina* virus
HGT	horizontal gene transfer
His_6	polyhistidine tag
HPLC	high-performance liquid chromatography
HRMS	high resolution mass spectrometry
IPTG	isopropyl β-D-1-thiogalactopyranoside
iTOL	interactive tree of life
ITS	internal transcribed spacer
J	coupling constant
kDa	kilodalton
KDN	2-keto-3-deoxy-D-glycero-D-galacto-nononic acid
KDO	3-deoxy-D-manno-oct-2-ulosonic acid
LC_{50}	lethal concentration 50
LC-MS	liquid chromatography-mass spectrometry
LD_{50}	lethal dose 50

m	multiplet
m/z	mass to charge ratio
Man	mannose
ManN	mannosamine
ManNAc	*N*-acetyl mannosamine
MCP	Major capsid protein
MeOH	methanol
$MgCl_2$	magnesium chloride
$MgSO_4$	magnesium sulfate
MMETSP	Marine Microbial Eukaryote Transcriptome Sequencing Project
MRM	multiple reaction monitoring
MS	mass spectrometry
mw	molecular weight
NaCl	sodium chloride
NaOH	sodium hydroxide
NCBI	National Center for Biotechnology Information
NCLDV	nucleo-cytoplasmic large DNA virus
NDP	nucleotide diphosphate
Neu	neuraminic acid
Neu5Ac	*N*-acetyl neuraminic acid
NMP	nucleotide monophosphate
NMR	nuclear magnetic resonance
n-PrOH	*n*-propanol
OD	optical density
OLPV	organic lake phycodnavirus
OsO_4	osmium tetroxide
P	phosphate
p.i.	post infection
PAGE	pulsed-field gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
PEP	phosphoenolpyruvate
PFGE	pulsed-field gel electrophoresis
PGC	porous graphitic carbon

PgV	*Phaeocystis globosa* virus
PkV	*Prymnesium kappa* virus
PoV	*Pyramimonas orientalis* virus
PpDNAV	*Prymnesium parvum* DNA virus
PPi	inorganic pyrophosphate
ppm	parts per million
PpV	*Phaeocystis pouchetti* virus
Prym1	prymnesin-1
Prym2	prymnesin-2
PSU	practical salinity unit
PTFE	polytetrafluoroethylene
qPCR	quantitative polymerase chain reaction
qRT-PCR	quantitative real-time polymerase chain reaction
QTOF	quadrupole time-of-flight
Rha	rhamnose
RHM	plant trifunctional rhamnose biosynthetic protein
Rml	bacterial rhamnose biosynthetic protein
ROS	reactive oxygen species
RPM	revolutions per minute
s	singlet
SAX	strong anion exchange
SDS	sodium dodecyl sulfate
SPE	solid phase extraction
t	triplet
TDP	thymidine diphosphate
TEM	transmission electron microscopy
TFA	trifluoroacetic acid
TPWD	Texas Parks and Wildlife Department
Tris	trisaminomethane
UDP	uridine diphosphate
UEA	University of East Anglia
UER1	plant bifunctional rhamnose biosynthetic protein
UGD	UDP-glucose 4,6-dehydratase
UHPLC	ultra-high performance liquid chromatography
VLP	virus-like particle

Xyl	xylose
ρ	density

Publications Arising from this PhD

Copies of publications arising from this PhD can be found at the back of this thesis.

1. **Wagstaff, B**.; Vladu, I.; Barclay, J.; Schroeder, D.; Malin, G.; Field, R., Isolation and Characterization of a Double Stranded DNA Megavirus Infecting the Toxin-Producing Haptophyte *Prymnesium parvum*. *Viruses* **2017,** 9, (3), 40.

Author Contributions

Chapter 2 - Genetic and Metabolite Assessment of a Harmful Algal Bloom; *Prymnesium parvum* and its Toxins

Ben A. Wagstaff, Jennifer Pratscher, Elliot Brooks, Edward Hems, J. Colin Murrell, Robert A. Field

BAW, JP, JCM and RAF designed the study. All authors contributed to collection of water samples. BAW performed microscopy experiements with help from EH. BAW maintained all algal cultures. BAW performed toxin extraction experiments and analysis with help from EH. JP performed nucleic acid extractions, qPCR and sequencing with help from EB. JCM and RAF supervised the research. BAW wrote the chapter.

Chapter 3 - Isolation and Characterization of a Double Stranded DNA Megavirus Infecting the Toxin-Producing Haptophyte *Prymnesium parvum*

Ben A. Wagstaff, Iulia C. Vladu, J. Elaine Barclay, Declan C. Schroeder, Gill Malin, Robert A. Field

BAW and RAF designed the study. BAW performed all experiments, with help from ICV for chloroform sensitivity and culture maintenance, and JEB for thin-section preparation and EM imaing. DCS provided 14 *Prymnesium* strains. BAW, GM, DCS, and RAF analysed the data and reviewed parts of the chapter. BAW wrote the chapter.

Chapter 4 - Discovery and Characterization of *de novo* KDN Biosynthesis in the Toxic Haptophyte, *Prymnesium parvum*, Suggests Widespread Sialic Acid Biosynthesis Amongst Algae

Ben A. Wagstaff, Martin Rejzek, Robert A. Field

BAW and RAF designed the study. MR performed sugar-nucleotide profiling experiments, BAW performed all other experimental work. RAF supervised the work. BAW wrote the chapter.

Chapter 5 - NDP-β-L-rhamnose Biosynthesis Across the Algal Taxonomic Groups: an Evolutionary Perspective

Ben A. Wagstaff, Martin Rejzek, Lionel Hill, Sakonwan Kuhaudomlarp, Robert A. Field

BAW, MR and RAF designed the study. BAW maintained *P. parvum* cultures and SK maintained *E. gracilis* cultures. BAW did all bioinformatic analysis. MR did sugar-nucleotide profiling with help with LH. BAW, MR and RAF analysed the data. RAF supervised the work. BAW wrote the chapter with input from all other authors.

1 Thesis Introduction and Outlines

1.1 Introduction

1.1.1 Harmful algal blooms

Harmful algal blooms (HABs) are a rapid expansion of a phytoplankton population in an aquatic ecosystem [1]. Typically only involving one or a small number of phytoplankton species, these blooms are frequently harmful to the surrounding ecosystem either through production of algal toxins, mechanical damage to organisms, water hypoxia or other means [2]. Frequently, HABs cause damage to human and animal health, ecosystems, and aquatic organisms such as fish which in turn can cause significant economic damage. The frequency of HABs appears to have grown in recent years, and in turn there has been an increased focus from scientists and regulatory authorities to combat the negative impacts of HABs. Whilst the focus from governing bodies has been on mitigation or management, scientists have sought to investigate the causes of HABs and the toxin-producing species that cause them [3].

1.1.2 Prymnesium parvum

Commonly referred to as golden algae due to the fucoxanthin pigments contained in its chloroplasts, *Prymnesium parvum* is a unicellular, cosmopolitan, toxin-producing microalga belonging to the Haptophyta phyla. As a mixotroph, it can ingest dissolved organic matter (DOM) and feed on bacteria and protists in addition to obtaining energy through photosynthesis. Its successful cosmopolitan growth is in part due to the euryhaline nature of *P. parvum*, having been shown to tolerate water salinity levels ranging from 3 (just above freshwater) to 30 (sea water strength) practical salinity units (PSU) [4, 5].

1.1.2.1 Phylogeny

Although the phylogeny of protists is still undergoing rapid changes [6], *Prymnesium parvum* is currently classified as a member of the Chromista kingdom of life, Haptophyta phylum, Prymnesiophyceae class, Prymnesiales order and Prymnesiaceae family. As a member of the Haptophyta phylum, *P. parvum* is a close relative of the bloom-forming coccolithophore, *Emiliania huxleyi*, whose oceanic blooms are often so large that they can be seen from space [7]. Other members of the Prymnesiales, including *Chrysochromulina*, have previously been shown to form independent clades, with the *Prymnesium* genus occupying one independent clade and members of *Haptolina* forming a sister group to *Prymnesium* (Figure 1). This analysis by Edvardsen *et al* was based on a combination of nuclear 18S rRNA and partial 28S rRNA gene sequences, as well as plastid 16S rRNA ribosomal encoding DNA sequences [8].

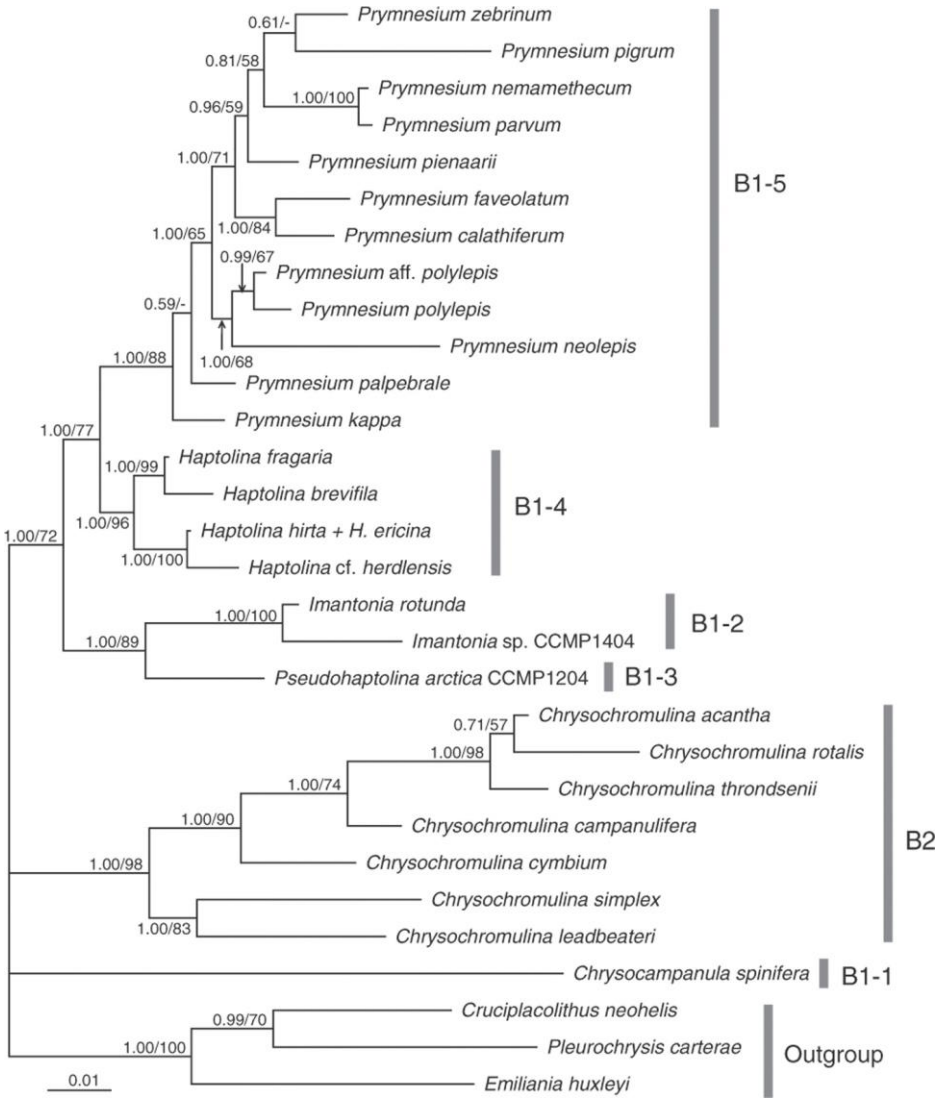


Figure 1 – Phylogenetic tree of members of the Prymnesiales based on nuclear 18S rRNA, partial 28S rRNA and plastid 16S rRNA gene sequences. Reprinted with permission from B. Edvardsen *et al*, European Journal of Phycology, 2011, 46(3), 202-228. Copyright 2011 Taylor & Francis.

1.1.2.2 Morphology

P. parvum is a unicellular microalga, ellipsoidal in shape, where the cell length is seen to range from 8-11 µm [9]. Unlike most toxin-producing microalgal blooms that are frequently referred to as "red-tide" blooms, *P. parvum* forms blooms that are golden in appearance due to the accessory pigment fucoxanthin in its chloroplasts (Figure 2B).

Although it is common for motile microalgae to contain just a single flagellum, *P. parvum* contains 2 flagella and 1 haptonema (Figure 2A). Whilst the flagella are used for motility in the water column, it has been proposed that the haptonema is used for attachment of prey in the

phagocytic process, aiding the organism in mixotrophy [10]. The nucleus of the organism is found between chloroplasts that line the inside of the cell membrane. Unusually (and often used for phylogenetic characterization), is the presence of two layers of scales on the cell surface. Scales are found in two forms, with either extremely inflexed rims or not (Figure 2B). These features, as well as a 2:1 flagella:haptonema ratio, are often used for phylogenetic analysis [11].

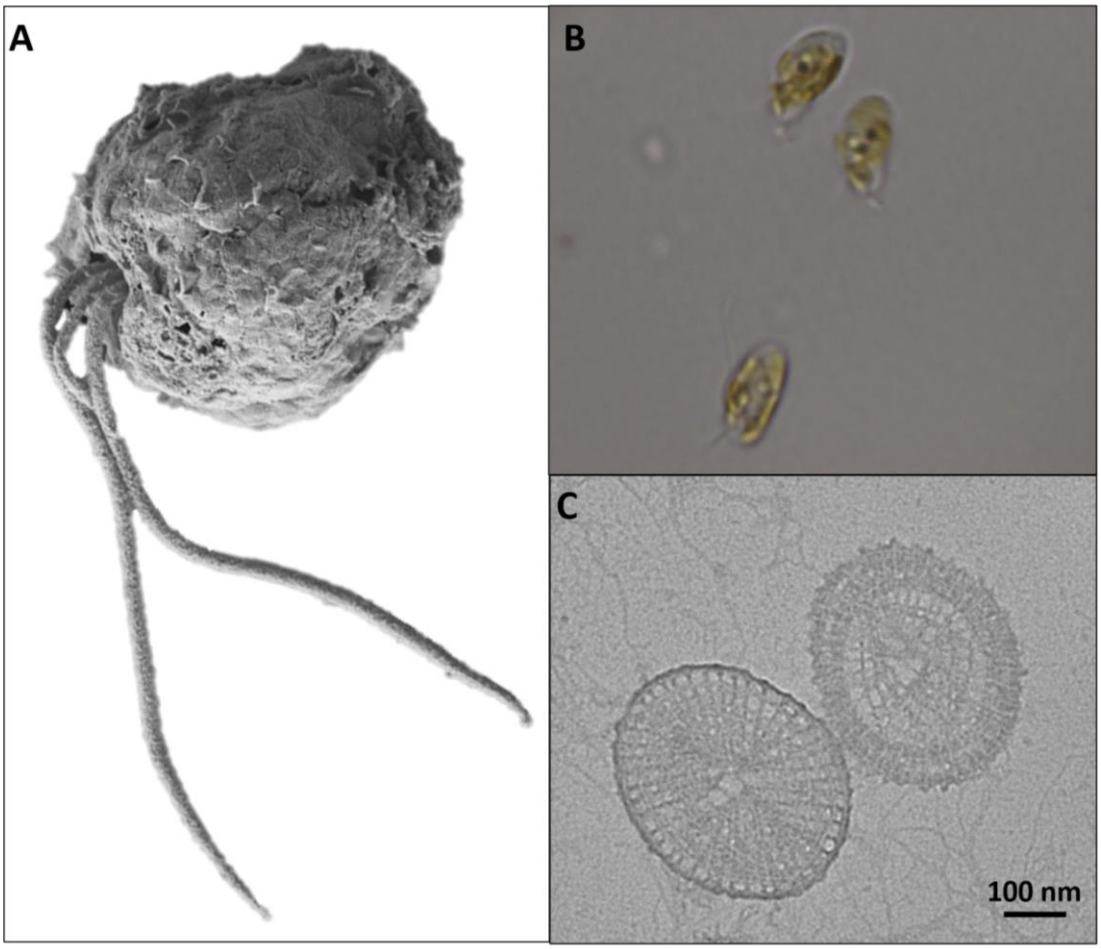


Figure 2 – Fine morphology of *Prymnesium parvum*. (A) *P. parvum* cell as observed by scanning electron microscopy (background digitally removed). Note the presence of 2 long flagella and a shorter haptonema. (B) 3 *P. parvum* cells observed by optical microscopy at 10,000x magnification showing the typical golden colour associated with blooms of the organism. (C) Scales of *P. parvum* observed by transmission electron microscopy (TEM) (scale bar represents 100 nm).

1.1.2.3 Life cycle

Current studies into the life cycle of *P. parvum* have resulted in ambiguous conclusions, with most theory being based on the work of Larsen *et al* in the 1990s [12, 13]. The current belief is

that, although *P. parvum* is eukaryotic, reproduction occurs via prokaryote-like binary fission, in a longitudinal manner, dividing approximately once per day [14].

A heteromorphic life cycle has been proposed for *P. parvum,* with both haploid and diploid stages. It was previously believed that *P. patelliferum* was a different species to *P. parvum* but work by Larsen and Medlin showed that they are one species using molecular phylogenetic analysis [12]. Follow up work showed that *P. parvum* exists in both haploid and diploid life stages, whilst *P. patelliferum* is always haploid [13]. This confusion over species assignment was eventually resolved with a plethora of mating experiments between *P. parvum* and *P. patelliferum* in their different life stages, when the authors showed that *P. patelliferum* (haploid) was unable to produce *P. parvum* (diploid or haploid) but combinations of *P. parvum* surprisingly produced both haploid *P. parvum* and *P. patelliferum* [13]. A life cycle has therefore been proposed that shows meiosis of diploid *P. parvum* to form haploid *P. parvum* and *P. patelliferum*, and syngamy (the merging of two haploid cells) to form diploid *P. parvum* (Figure 3).

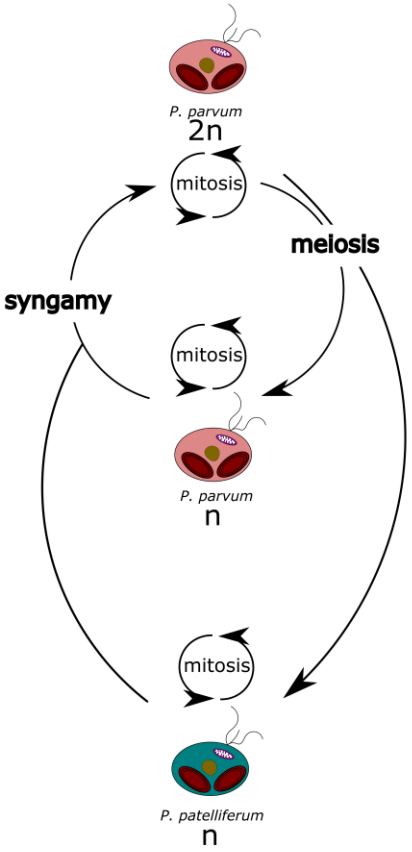


Figure 3 – Proposed life cycle of *P. parvum*/*P. patelliferum*. Redrawn from Larsen and Edvardsen, 1998 [13].

1.1.2.4 Geographic distribution of blooms

Toxic blooms of *P. parvum* have been reported worldwide (Figure 4), highlighting the ability of what was once considered a marine organism to thrive in a range of climates [15]. The organism has subsequently been classed as being both euryhaline and eurythermal, with examples of strains showing wide optimal salinity ranges from 8-25 PSU [16]. Other work has also shown various strains to tolerate temperatures from 5-30 ^{o}C [5], but tolerance to temperatures as low as 2 ^{o}C has been reported [17]. *P. parvum* blooms frequently occur in temperate and subtropical regions in low salinity, brackish inland waterways. Cell densities in blooms of *P. parvum* have been known to reach as high as 2 x 10^5 cells ml^{-1}, a figure only 10-fold lower than typically found in lab grown conditions [18].

Figure 4 – Reported global distribution of *P. parvum* blooms. Image reprinted under a Creative Commons Attribution 3.0 Unported (CC BY 3.0 - http://creativecommons.org/licenses/by/3.0/) from Manning, S. R.; La Clare, J. W., II. Prymnesins: Toxic Metabolites of the Golden Alga, *Prymnesium parvum* Carter (Haptophyta). Mar. Drugs 2010, 8, 678-704. Copyright 2010 by the authors; licensee Molecular Diversity Preservation International, Basel, Switzerland.

Although there is debate as to when the first fish kill due to *P. parvum* occurred, the first recorded kill is believed to have been in the Netherlands and reported by Liebert *et al* in 1920 [19]. Recorded kills due to the microalga on the Norfolk Broads, United Kingdom, started in 1969 [20] although records from the area showed fish kills with similar coloured water and fish phenotypes as early as 1894 [21]. Shilo and Aschner reported instances of *P. parvum* fish kills in Denmark and the Netherlands in 1938 and went on to discuss the many problems Israel had experienced with deadly blooms in the late 1940s that have since continued [17]. However,

the implication of *P. parvum* in fish mortality is global, with reports of mass kills from Scotland [22], Norway [23], Germany [24], Finland [25], China [26] and the USA [27], where it is a particular problem to the Texas aquaculture industry [28]. Since the first recorded fish kill due to *P. parvum* in Texas in 1985, Texas Parks and Wildlife Department (TPWD) have estimated approximately 34 million fish deaths due to blooms of the alga, a loss economically estimated at $13,000,000 [28].

1.1.3 *Prymnesium parvum* on the Norfolk Broads

The Norfolk Broads is a low-lying area that was excavated for peat and fuel prior to the 14th century. Subsequent flooding of the area from the 13th century onwards meant work was stopped but the digging of channels for resources such as roofing materials continued. This led to the formation of 63 broads across Norfolk and Suffolk, connected by channels that exist today [21]. The Broads, their channels, and the surrounding marshland acts as a haven for rare wildlife, and results in the Norfolk Broads being Britain's largest protected wetland, recently classified as a national park.

Many of the 63 Broads that span Norfolk and Suffolk are traversable by boat, and this has created a thriving tourism industry in the area, fuelled by angling and boating activities. However, in recent years the area has been tormented with environmental issues such as saline incursions and eutrophication, but of particular concern is the frequent recurrence of *P. parvum* blooms that leads to massive losses in fish stocks. Mass fish mortalities since the 1960s have been associated with *P. parvum*, with blooms since then occurring almost annually, although with no obvious seasonality. Bales *et al* discussed particularly large fish kills in 1969 and 1970, and smaller kills in 1973 and 1975 [20]. They claimed the older ecosystem, of particularly the Hicking Broad area and surrounding Upper Thurne area (Figure 5), changed from a largely charophyte-dominated state to phytoplankton-dominated by the mid-1970s. Their explanation for this was the sudden nesting of black-headed gulls (*Larus ridibundus*) and the resultant eutrophication brought about from gull guano. The gull guano was thought to have provided the necessary organic nutrients in the water for the microalgal species such as *P. parvum* to bloom. This theory was later supported by the observation that as the gulls dispersed the number of *P. parvum* cells declined.

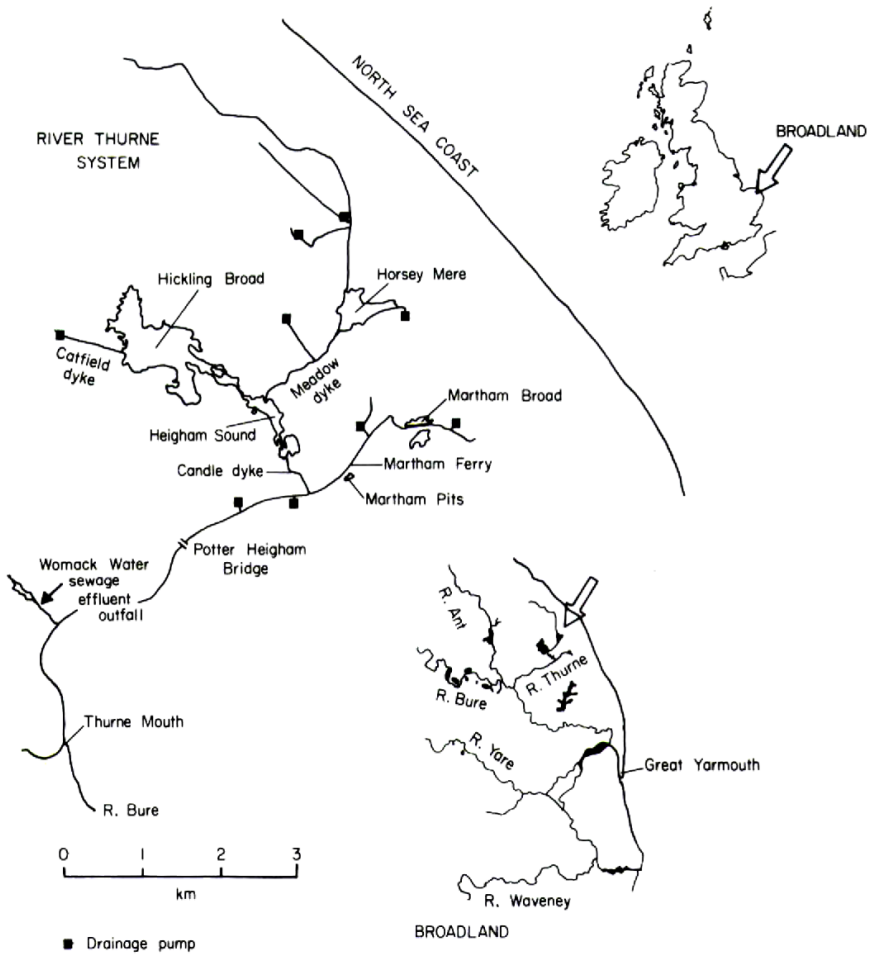


Figure 5 – Hickling Broad and the Upper Thurne area of the Norfolk Broads. Reprinted (adapted) with permission from P. Holdway *et al*, Freshwater Biology, 1978, 8, 295-311. Copyright 1977 John Wiley and Sons.

As well as the severe environmental impact of *P. parvum* blooms which can leave broads completely devoid of fish, the mass fish mortality also damages the economy. A recent report by the Broads Authority estimated that the tourism industry of the Broads contributes around £550 million annually to the local economy, a sum that is largely made up by the angling and boating activities that the Broads offers [29]. The loss of fish stocks through toxic *P. parvum* blooms threatens this revenue as was seen in Spring 2015 when multiple blooms of *P. parvum* across the Upper Thurne system left waterways littered with the unsightly appearance of dead fish (Figure 6). This bloom is discussed in more detail in Chapter 2 of this thesis.

Figure 6 – Dead fish on the waterways on the Upper Thurne area of the Norfolk Broads. Image captured by Martin Rejzek on 24th April 2015 during a toxic *P. parvum* bloom.

1.1.4 *Prymnesium parvum* toxins

Since *P. parvum* was first implicated in fish kills there has been debate on the toxic entity responsible, with some researchers suggesting it is a mixture of compounds rather than a single toxin [30]. It is generally believed, however, that the toxins produced by *P. parvum* are allelopathic, giving the organism a competitive advantage over other phytoplankton in the waterways [31]. The exact mechanism of the toxins in allelopathy is unknown, but inhibition of competitor growth, deterrence of grazers and lysis of prey are all strong possibilities. Despite the ambiguity, several toxic compounds have been extracted from *P. parvum* that include lipopolysaccharide-like compounds [32], proteolipid [33], galactoglycerolipids [34], fatty acid amides [35, 36], fatty acids [37], and the ladder-frame polyether prymnesins [38]. It is worth noting that some of these have seen been discredited by Blossom *et al* [39].

Because of the current ambiguity on the responsible toxin/s and the difficulty to measure these compounds analytically, toxicity assays are often performed on crude cell extracts using the erythrocyte lysis assay (ELA) technique [40]. This, however, is only a measure of intracellular toxicity so alternative assays also exist which measure extracellular toxicity and their effect on fish [41]. Although the exact whereabouts of *P. parvum* toxins and their mode of action is unknown, probably the most convincing work is by Remmel and Hambright who

proposed that toxins from *P. parvum* are intracellular and only released through contact with prey or by experimental procedures and natural causes of stress [42] (Figure 7).

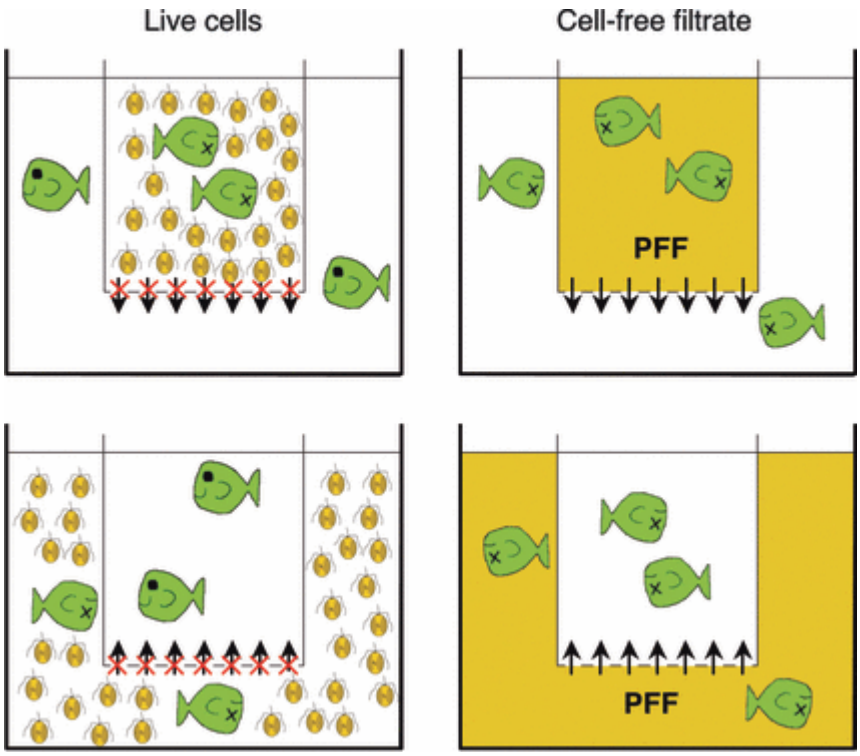
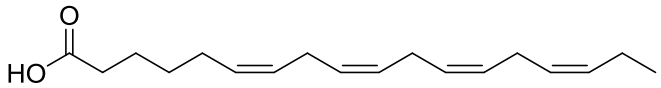


Figure 7 – Assessment of intracellular vs extracellular toxins by Remmel and Hambright. Experiment was designed to utilise permeable membranes to separate *P. parvum* cells from fish. Fish only died when in direct contact with live cells (contact micropredation) or when cells were lysed and the filtrate could pass through 0.2 µm membranes. This suggests toxins are intracellular and only released by contact or through breakdown of algal cells. PFF = Prymnesium free filtrate. Reprinted with permission from Remmel and Hambright, Ecol. Lett, 2012, 15(2), 126-132. Copyright John Wiley and Sons.

1.1.4.1 Fatty acids

Fatty acids have been implicated in the ichthyotoxicity of phytoplankton, cyanobacteria, and some lower eukaryotes for years, where it is believed that they act as allelochemicals [37, 43]. Henrikson *et al* have studied the role of fatty acids in *P. parvum* toxicity in some depth, and proposed that the ladder-frame polyether prymnesin-1 and -2 were not the biologically relevant toxins in field samples collected from the U.S. nor from laboratory samples [37]. Instead, a cocktail of fatty acids was collected from cultures, in which it was found that the level of toxicity positively correlated with the degree of unsaturation of the organic compounds. Interestingly, the level of fatty acids in lab grown cultures was dramatically higher

than in field collected samples, questioning the biological relevance of this class of compounds in *P. parvum* environmental toxicity. Despite this, Henrikson *et al* showed that stearidonic acid (Figure 8) isolated from *P. parvum* displayed a marked level of ichthyotoxicity, with LC_{50} values to fish of 21.9 ± 6.3 μM [37].



Stearidonic acid

Figure 8 – Primary fatty acid implicated in *P. parvum* toxicity.Henrikson *et al* [37].

The method of toxicity of fatty acids to algae is believed to be via membrane disruption, as shown by fluctuation in K^+ ion concentration in culture media when a mixture of cyanobacteria and chlorophytes were exposed to a solution of fatty acids; likely a result of K^+ ions leaking out from a disrupted membrane [44]. Exposure to unsaturated fatty acids resulted in higher extracellular levels of K^+ ions than exposure to saturated fatty acids, agreeing with other researchers that more unsaturated fatty acids are generally more toxic [42, 45].

1.1.4.2 Fatty acid amides

More recently, fatty acids amides (FAAs) have been implicated in *P. parvum* toxicity. Most of the work carried out in this area was by Bertin *et al*, who first showed the presence of FAAs in both lab and field cultures of *P. parvum* in 2012 [36]. In this work, seven primary FAAs were found from the toxic extracts of *P. parvum*, as well as one hydroxyamic acid (Figure 9). The group went on to show that, unlike fatty acids, these compounds accumulate to lethal levels in field cultures, and that an increase in pH and the presence of divalent cations increases the toxicity of the compounds [35].

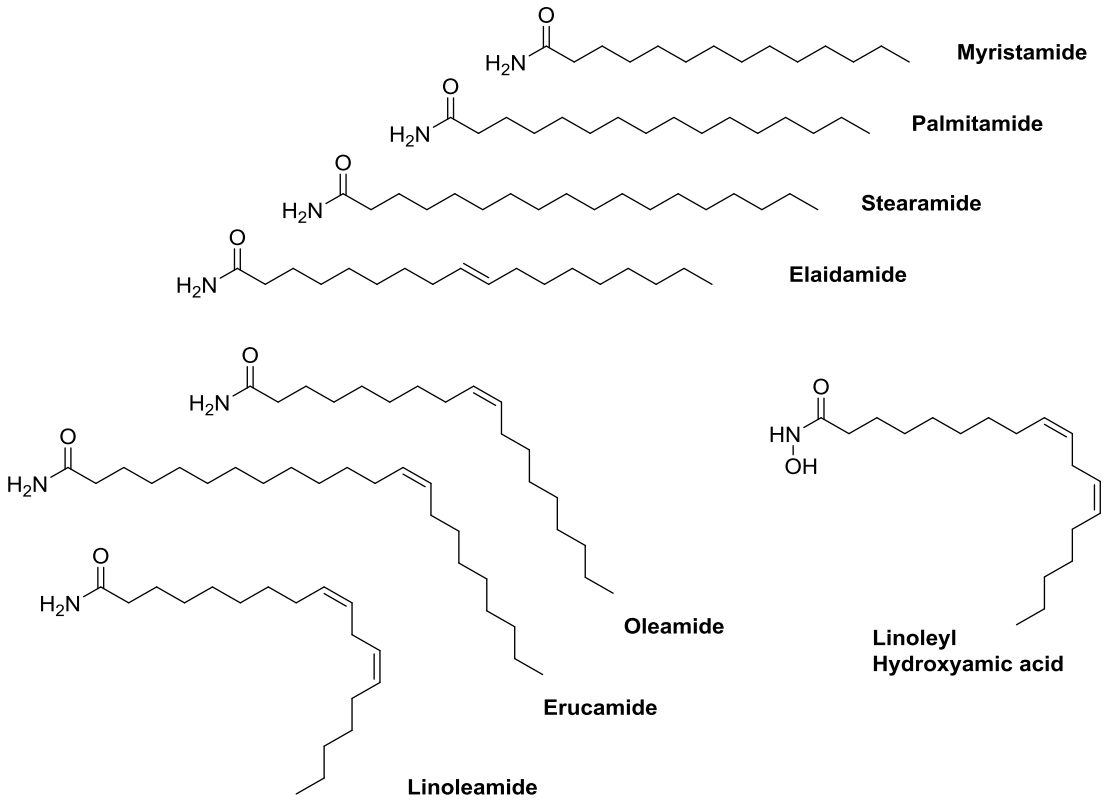


Figure 9 – Primary fatty acid amides and hydroxyamic acid isolated from *P. parvum*. Figure redrawn from Bertin *et al*, 2012 [36].

FAAs are abundant in nature [46] and there are previous examples of their isolation from microalgae and other marine organisms. Octadecanamide, as well as a suite of others, was isolated from the freshwater green alga, *Rhizoclonium hieroglyphicum* [47]. A particularly interesting class of compounds known as the malyngamides (derivatives of tetradecanoic acid) have been isolated from the cyanobacterium *Lyngbya majuscula* [48]. The unusual presence of a vinyl chloride in this natural product led to other researchers studying this organism, with Sitachitta *et al* later isolating grenadamide and its derivatives [49].

It is, however, FAAs abundance in nature that led to speculation as to whether these are the primary toxins in *P. parvum* toxicity. As well as the marine organisms listed previously, FAAs are produced by plants and are often considered essential components of oils. Oleamide, a type of primary FAA, was first discovered to naturally occur in human serum by Arafat *et al* [50]. Oleamide and several of its primary FAA derivatives are known to induce sleep [51], a symptom that would explain the lethargic symptoms observed in fish during blooms.

It was however the knowledge that oleamide and its derivatives are frequently used as slip agents in laboratory plastics [52] that made Blossom *et al* question the legitimacy of FAAs as

toxins of *P. parvum* [53]. The group could not detect oleamide in any of the 5 *P. parvum* strains examined, instead showing it could readily be extracted from laboratory plastics.

1.1.4.3 Prymnesins

In 1999 Igarashi *et al* reported isolation of two large, ladder-frame polyether toxins subsequently called prymnesin-1 and -2 [38]. In total, just 25 mg of the two compounds was isolated from 400 L of cell culture. The compounds were both shown to display haemolytic activity and to be ichthyotoxic at nanomolar concentrations, which for prymnesin-2 was shown to increase 100-fold from 300 nM to 3 nM in the presence of Ca^{2+} ions [54].

Structure elucidation of the two compounds showed them to be polycyclic polyether compounds with several unusual structural features (Figure 10). Firstly, both toxins are glycosylated; prymnesin-1 contains α-D-ribofuranose, α-L-arabinopyranose, and β-D-galactofuranose. Galactofuranose is found widely in prokaryotes, but also exists amongst lower eukaryotes. Its absence in mammals or plants has made its biosynthetic pathway a potential drug target [55, 56]. Unlike prymnesin-1, prymnesin-2 is glycosylated at just one position, with the extremely uncommon α-L-xylofuranose. Although xylose is found extensively throughout nature, sugars in their furanose ring form are typically less common, with very few reports of xylofuranose in natural products [57]. The additional sugar moieties in prymnesin-1 make it slightly more polar and it therefore elutes ahead of prymnesin-2 in reverse-phase chromatography [38].

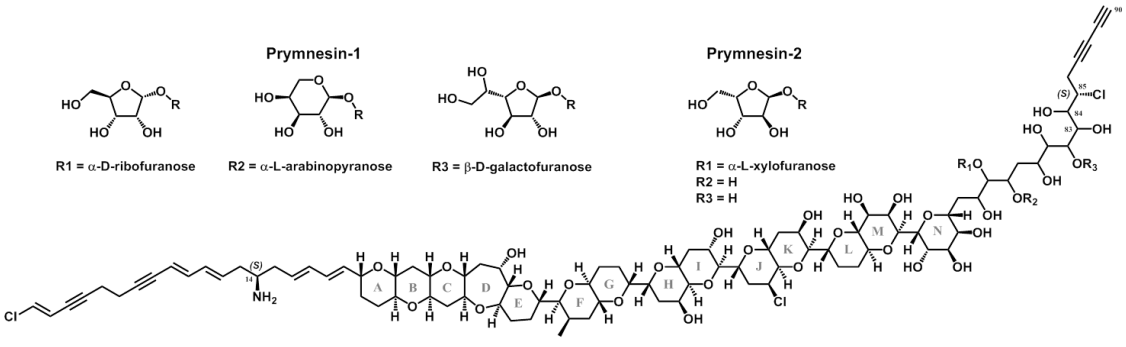


Figure 10 – Structure of prymnesin-1 and -2 first reported by Igarashi *et al* in 1996.

As well as glycosylation, these compounds contain other unusual structural features including chlorine atoms, alkyne moieties, a single amine, and a cyclic polyether backbone believed to be polyketide-derived. This polyether backbone has been observed in other algal natural product toxins, namely the ladder-frame polyether toxins from dinoflagellates that include

brevetoxin, yessotoxin, ciguatoxin and maitotoxin. The prymnesins, however, are the first example of ladder-frame polyether toxins produced outside of the Dinophyceae.

Although the toxicity of the prymnesins was unquestionable, the minute amounts produced in lab cultures and the fact that these compounds had never been detected in natural water samples led to speculation that they weren't the toxins responsible for fish deaths. Following a 14-year gap in literature reports of the toxins, Manning *et al* eventually reported on the detection of prymnesins-1 and -2 from lab cultures of *P. parvum* in 2013 [58], although there were still no reports of these compounds from natural water samples. However, it appears this inability to detect the prymnesins from both lab and natural samples may have been due to researchers looking for the wrong compounds; Rasmussen *et al* recently reported a chemodiversity of the prymnesins in 2016 when they identified numerous other types of prymnesins from different strains of *P. parvum* [59]. This new library of prymnesins had slight alterations in the structures of the toxins leading to different m/z values as observed by mass spectrometry.

1.1.4.4 Ladder-frame polyether toxins

Unicellular microalgae of the Dinoflagellata phylum are a diverse group of organisms that, like haptophytes, can thrive in both marine and more freshwater environments. Whilst some dinoflagellates are responsible for bioluminescence [60], others are known for forming toxic blooms often referred to as 'red-tide' blooms in which ladder-frame polyether toxins analogous to the prymnesins are frequently involved [61] (Figure 11). Because of the unique backbone that this family of toxins shares with the prymnesins, much speculation on the modes of action for the prymnesins is based on experimental data obtained for one or more of these dinoflagellate toxins.

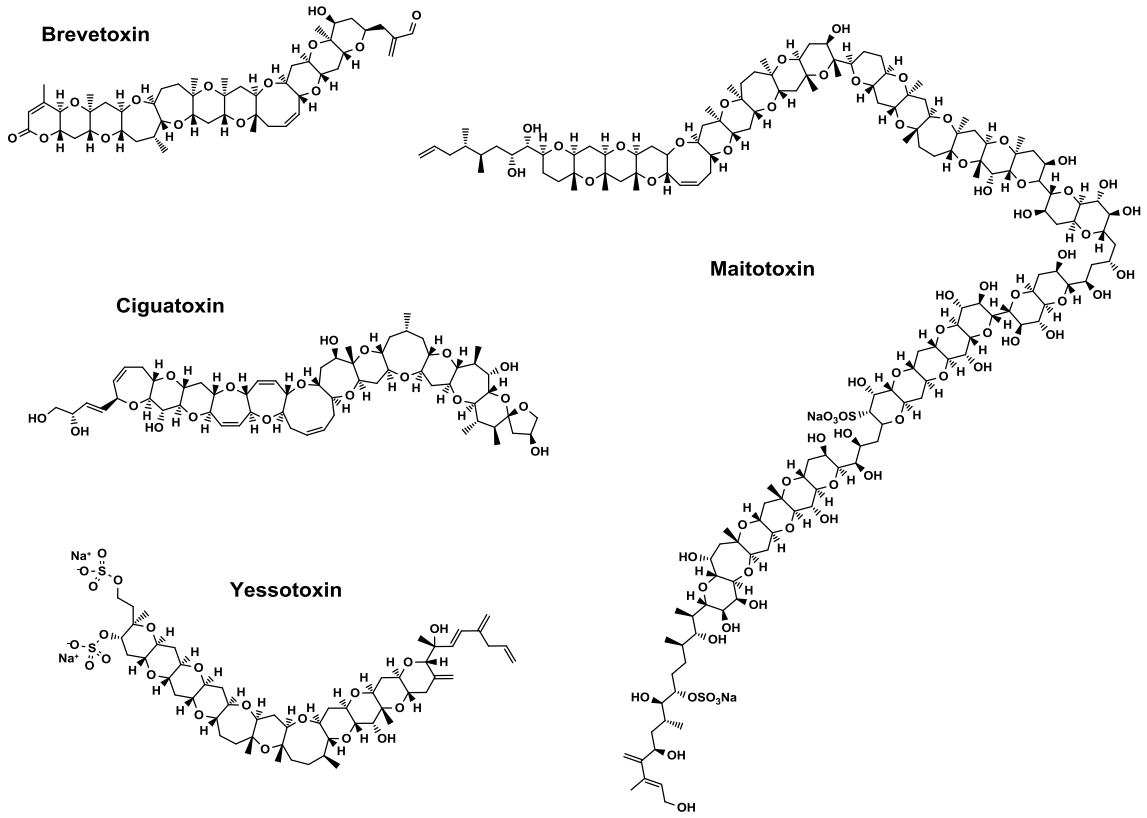


Figure 11 – Examples of other ladder-frame polyether toxins produced by dinoflagellates.

Brevetoxin, produced by many dinoflagellates of the *Karenia* genus, is the leading toxin for neurotoxic shellfish poisoning, and thus the levels in shellfish are now regulated by law [62]. Since its first isolation from *Karenia brevis* (formerly *Gymnodinium breve*) [63, 64], brevetoxin and many analogues have been found in several other dinoflagellate species [65]. Ciguatoxin, produced by the dinoflagellate *Gambeierdicus toxicus*, accumulates in fish and if ingested by humans causes ciguatera fish poisoning, a non-fatal condition presenting similar symptoms to food poisoning [66]. Maitotoxin, also produced by *G. toxicus*, is the largest non-protein based secondary metabolite currently known. Its activation of Ca^{2+} channels results in a remarkable LD_{50} of just 50 ng/kg in mice [67]. Yessotoxins are believed to induce ER-stress, triggered by its action as a potential ribotoxin [68], which has led to yessotoxin and its analogues being an interest in the targeting of cancer cells that resist normal apoptosis processes.

1.1.5 Factors affecting growth and toxicity

Although HABs can involve the blooming of one or more species, *P. parvum* blooms are frequently dominated by this species alone, with very few other species reaching high cellular concentrations. This suggests toxins produced by *P. parvum* act in an allelopathic nature, as

the modest growth rate alone wouldn't account for such a heterogeneous population [40]. Many environmental factors have been analysed for their effect on either growth of *P. parvum* or the toxicity of the organism. Whilst some correlations have been made, much work to date is largely inconclusive and the area remains a key research focus.

1.1.5.1 Nutrients

If the *Prymnesium* toxins are indeed allelopathic, one would expect production to be highest when nutrient levels are at their lowest; when the need for the organism to outcompete competitors is at its highest. This occurs during stationary phase of growth, when nutrients supplemented to the media have been largely consumed. Manning and La Claire [15] reported a reduction in ichthyotoxicity during the logarithmic growth phase when nutrient availability is high, compared to stationary phase when nutrient availability is low. This is further reinforced by work from Houdan *et al* who showed *P. parvum* cultures to be fatally toxic to *Artemia* sp. during stationary phase, but much less toxic during logarithmic phase [69]. Furthermore, work by others, in both lab cultures and field trials, has shown that in nutrient replete conditions toxicity is lowered or even non-existent [70, 71].

1.1.5.2 Salinity

P. parvum is euryhaline, meaning it can thrive in a range of salinities. There have been various reports suggesting that salinity has no impact on toxicity [5] but interestingly, Baker *et al* found toxicity to be at its highest at the extremities of salinity, when growth was limited [41]. This reinforces the idea that the toxins are allelopathic and possibly triggered as a stress response of the organism.

1.1.5.3 pH

A few studies have shown that some of the toxins released by *P. parvum* are ionisable, thus becoming increasingly more toxic as pH rises above 8.0 [72, 73]. In fact, a pH below 7 has been found to completely remove toxic properties, a study that was in agreement with findings that blooms of *P. parvum* occur between pH 7.2 to 9.3 [26]. Shilo also showed that whilst ichthyotoxicity rises as pH rises, hemolytic activity rises as pH falls; with cell lysis occurring as low as pH 5 [73]. Others reported water pH during blooms as high as 9.4 [25]. It was recently concluded that ichthyotoxicity occurs at a pH above 7, with maximum toxicity at pH 9 [15]. All these findings would suggest there is a strong correlation between pH and toxicity, in agreement with the idea that *P. parvum* toxins are ionisable. The basic amine on the polyether prymnesin-1 and -2 would support this argument.

1.1.5.4 Cofactors

Even though toxic extracts of *P. parvum* have hemolytic activity, ichthyotoxicity has been shown to require co-factors to 'activate' the toxins [16, 74]. Shilo established early on that the presence of divalent cations significantly enhanced the toxicity of *P. parvum* extracts [75] and this was later confirmed by several groups that showed an increase in toxicity in the presence of Na^{2+}, Ca^{2+} and Mg^{2+} ions [76]. This would agree with work on the polyether prymnesin toxins; when prymnesin-2 was first isolated Igarashi *et al* showed that it had an LD_{50} of 300 nM towards the minnow *Tanichthys albonubes*, but in the presence of Ca^{2+} this dropped to 3 nM [38]. This work would agree with the mode of action of other polyether natural products, such as monesin, that disrupt and influence the transport of ions across cell membranes [77]. These findings suggest a strong relationship between the presence of divalent cations and toxicity.

1.1.5.5 Light

P. parvum is known to grow best in moderate to low light conditions, but toxicity observations are often erratic. Larsen *et al* showed that 3 strains of *P. parvum* grew optimally at different light intensities, ranging from 65 µmol $m^{-2}s^{-1}$ for a strain from England, to 200 µmol $m^{-2}s^{-1}$ for an Australian strain [5]. In these cases, no relationship between toxicity and light exposure was observed. Other examples exist of strains that grow better in higher illumination, suggesting again an ability of the species to adapt to its environment [78]. Whilst some have debated that light plays a significant role on toxicity [79], others argue that there is no difference in light and dark grown cultures [80].

P. parvum is mixotrophic, meaning it can feed and acquire carbon in the absence of light. This may at first appear to suggest the organism doesn't require light to produce toxins, however, if toxins play a role in mixotrophy the reverse argument could be justified. Parnas *et al* showed that a complete loss of ichthyotoxicity was observed when cultures were grown under constant illumination, but hemolytic activity was retained [78]. This further suggests that more than one toxin is involved, and may also suggest that the ichthyotoxins are degraded, inactivated, or not even produced during over-exposure to light.

1.1.5.6 Temperature

As well as being euryhaline *P. parvum* is also eurythermal, meaning it can grow in a wide range of temperatures and climates worldwide [26]. Furthermore, there is no strong correlation between the time of year and a toxic *P. parvum* bloom occurring. Both facts make it difficult to

correlate temperature to toxicity. However, in laboratory cultures, a recent study showed that the stability of toxic extracts from *P. parvum* is indeed affected by temperature [39]. Exotoxins were highly unstable, losing their lytic activity completely after 3 days when stored at -20 ^{o}C, or 24 hours if stored at 4 ^{o}C. Conversely, intracellular toxins maintain toxicity if stored as a cell pellet at -20 ^{o}C, even after 90 days. The authors therefore suggested that intracellular toxins are either structurally different, or stored with protective vesicles [39]. Work on the effect of temperature on toxicity is inconclusive, although a general trend has been seen suggesting the instability of exotoxins of *P. parvum*.

1.1.5.7 Viruses and grazers

As well as the abiotic factors mentioned previously, viruses, grazers and other microbes have significant effects on the growth dynamics of algae in natural waters [81, 82]. The algal virology research field has boomed in recent years, with the discovery of *Acanthamoeba polyphaga mimivirus* in 2003 [83] bringing a new age in photosynthetic protist virology. It is widely accepted that algal viruses have significant effects on the control and regulation of algal population dynamics in ecosystems [84, 85], with viruses often responsible for sudden crashes in algal blooms. More recently, it has been found that viruses contribute significantly to global nutrient cycling [86, 87], with carbon shunting and even climate active gases such as dimethylsulfide (DMS) being released into the atmosphere because of viral lysis of algal blooms [88]. However, even though viruses have been isolated that infect toxin-producing algae [89], no studies have been undertaken showing how viruses impact toxicity of algal blooms. Prior to this thesis, no viruses had been isolated that infect *P. parvum*, although viruses infecting the non-toxic *Prymnesium kappa* have been isolated before [90].

Schwierzke *et al* showed using field experiments, that when viruses and grazers were removed via filtration, natural *P. parvum* populations increased [91]. They went on to show that grazers appeared to have a significant effect on *P. parvum* populations at the later stages of bloom development, and proposed that certain grazers are more resistant to *P. parvum* toxins than others. In particular, the rotifer *Notholca laurentiae* appeared to show a tolerance of *P. parvum* toxins that subsequently caused a bloom termination. This work, and other work by Roelke *et al* [92] appears to suggest that grazers exert a pressure on *P. parvum* that may result in termination of blooms if the grazing community shows tolerance to toxins produced by *P. parvum*.

1.1.6 Current monitoring and management strategies

The occurrence of HABs and algal toxins poses a severe threat to ecosystems, the economy and human health. There is therefore a great need to develop practical assays to detect toxins in both lab and field cultures. Currently animal bioassays are the detection method often used for algal toxins, although these come with a plethora of technical and ethical issues that must be overcome. On top of this, animal bioassays often take time: it is often the case that by the time toxin presence has been confirmed, fish populations are already devastated. Therefore, the emphasis in recent years has been to develop rapid, sensitive assays or management strategies that can be used in the field.

1.1.6.1 qPCR based monitoring

Although other methods do exist, the last decade has seen a boom in the number of molecular methods developed for the monitoring of toxic algal bloom species. Key advantages of methods such as quantitative real-time PCR include an unrivalled sensitivity and specificity [93]. qRT-PCR has been used to monitor many dinoflagellate species, but also haptophytes, diatoms and others. The *Karenia* genus that produces the polyether brevetoxins are one example of dinoflagellates with qRT-PCR assays established [94, 95], but assays for other toxin-producing species including the raphidophyte *Heterosigma* also exist [96].

qRT-PCR assays have also been applied to blooms of *P. parvum* with work by Galluzzi *et al* and Zamor *et al* proving successful in monitoring abundance of this organism [97, 98]. As is the case with all of these assays, however, incorporating them into regular monitoring programs remains key.

1.1.6.2 Clay flocculants

The use of clay flocculants has been widely used to mitigate HABs globally, and unlike many of the other methods is often used in larger water bodies [99]. However, like many of the described management methods, susceptibility to the clay varies between algal species. The use of clay on *P. parvum* was found to only be effective under certain conditions, and in most cases overall toxicity to aquatic organisms increased because of sediment toxicity brought about by the clay [100]. Although the use of clay has been shown to reduce algal cell numbers initially, after time cells and toxicity often return [101].

1.1.6.3 Barley Straw

The use of barley straw to control algal blooms is considered less environmentally damaging than alternative chemical approaches. Barley straw extract (BSE) is widely used as an inhibitor of algal growth [102, 103] although like clay, algal species often display a range of sensitivities to the extract. Treatment of ponds in which *P. parvum* was blooming with BSE showed no difference in toxicity to ponds that hadn't undergone treatment [104]. Grover *et al* elaborated on this, showing BSE does not inhibit growth or ichthyotoxicity of *P. parvum* in laboratory cultures [105].

1.1.6.4 Algaecides

Many strategies for the management of HABs involves the use of chemical algaecides, although their use in large water bodies is not usually feasible due to cost and availability of chemicals. Rodgers *et al* showed that out of three algaecides, Cutrine™-Plus (a copper-based algaecide) was the most effective at controlling *P. parvum* populations [106]. Tests in field populations showed that the algaecide was effective at controlling algal growth at concentrations non-toxic to fish, although like many other chemical algaecides, there is a risk of toxicity to other aquatic organisms.

1.1.6.5 Hydrogen peroxide

Although hydrogen peroxide had not been used prior to this thesis to combat *P. parvum* blooms, it has previously been used as a fast, effective management strategy to combat blooms of cyanobacteria. Cyanobacteria tend to be more sensitive to hydrogen peroxide than other eukaryotic phytoplankton [107], meaning even extremely low doses (~1 mg L^{-1}) will alleviate the effects of cyanobacteria blooms. Doses of just 2.3 mg L^{-1} were enough to control a dense bloom of the cyanobacteria *Planktothrix agardhii*, where the bacterial and toxin content in the water dropped by 99% In just a few days [108].

Although eukaryotic phytoplankton have higher tolerances for the ROS produced by hydrogen peroxide, there are still numerous examples where hydrogen peroxide has been successfully used to mitigate algal blooms. In 2012, Burson *et al* successfully terminated a dense bloom of the toxic dinoflagellate *Alexandrium ostenfeldii* [109]. The authors showed that treatment of a creek with 50 mg L^{-1} hydrogen peroxide reduced the bloom by 99.8% and cleared the water body of toxins responsible for paralytic shellfish poisoning. Most importantly, even with these higher concentrations of hydrogen peroxide, the authors noted that there were very few

negative effects on macroinvertebrates and fish, and that the microbial community quickly re-established itself after treatment was over.

1.1.6.6 Relocation of affected fish

Currently, one feasible but often time consuming strategy is the relocation of fish from toxic waters to safer waters. This was the case for the toxic bloom of *P. parvum* on the Norfolk Broads in 2015 when it was estimated ~600,000 fish were saved because of a mass relocation by the local environment agency and volunteers (Figure 12). Although an effective method, this method often costs a significant amount of money through the manpower required for such a task, so alternative management strategies are more desirable.

Figure 12 – Members of the environment agency and volunteers moving fish from toxic waters to safer waters. Hickling Broad, 2015. Image taken by Martin Rejzek.

1.2 Objectives and outline of the thesis

Despite the efforts of the community to research how abiotic factors influence the toxicity of *P. parvum* blooms, at the time of starting my PhD there were still many unanswered questions. The work discussed in this thesis sought to shed light on some of the fundamental mechanisms that effect bloom toxicity and fill in gaps that still exist in the wider literature. The ladder-frame polyether toxins known as the prymnesins are widely regarded as being the main toxic principles of *P. parvum* blooms, but the lack of detection of these compounds in natural water samples has made people question their significance. Furthermore, people are still unsure whether the main toxin responsible for fish deaths is intra or extracellular, and if it is intracellular - how it is released into the water?

Chapter 2 of this thesis analyses a toxic algal bloom across the Upper Thurne system of the Norfolk Broads in 2015 and identifies *P. parvum* as the species responsible through a combination of genetic analysis and metabolite analysis. The ladder-frame polyether toxins, the prymnesins, were detected in natural water for the first time and in the gill cells of a deceased pike, clarifying much literature doubt on both the toxin responsible and that the biological target is the gill cells – as seen for other polyether toxins such as the brevetxoins. To combat the current lack of early warning systems for following *Prymnesium* abundance we then developed a sensitive qPCR assay for the algae and have since recorded the changes in *P. parvum* populations across Hickling Broad over a 2-year period.

Samples taken during the toxic bloom of 2015 seemed to suggest the natural population of *P. parvum* was infected by an algal virus during the bloom. Chapter 3 analyses whole viral communities from Hickling Broad and goes on to describe the isolation and characterization of a new species of algal megavirus that infects *P. parvum*, *Prymnesium parvum* DNA virus (PpDNAV-BW1). The unexpected discovery of this virus led to a new hypothesis that algal viruses played important roles in toxic algal blooms – specifically that they may provide a release mechanism for intracellular algal toxins.

Chapters 4 and 5 focus on the glycobiology of algae and their viruses. 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid (KDN) is an uncommon sialic acid that has recently been reported to be involved in the viral infection of the haptophyte *Emiliania huxleyi* by *Emiliania huxleyi* virus (EhV). Prior to this work however, algae were not believed to be producers of sialic acids. Chapter 4 conclusively shows that *P. parvum* contains the sialic acid KDN and goes

on to show through expression of candidate enzymes how KDN is biosynthesised. Follow up bioinformatic analysis using these newly characterized genes suggests that sialic acid biosynthesis is much more widespread amongst algae than previously thought.

The 6-deoxy sugar, rhamnose, is an important monosaccharide found in structural polysaccharides, glycoproteins, and secondary metabolites across microbes, algae and plants. Furthermore, work in our group identified thymidine diphosphate (TDP) activated rhamnose as a key sugar in the viral infection process of *P. parvum*. Chapter 5 broadens the current literature knowledge about rhamnose biosynthesis using sugar nucleotide profiling techniques as well as a comprehensive bioinformatics analysis to show how algae produce UDP or TDP activated rhamnose, and where these biosynthetic capabilities evolved from.

Utilising the new knowledge discussed in this thesis, Chapter 6 explores recent work that is now being carried out to combat blooms of *P. parvum* on the Norfolk Broads. The sequencing of *Prymnesium parvum* DNA virus has allowed us to expand our current qPCR analysis to follow seasonal virus populations on Hickling Broad. As well as this, recent field trials that I championed have shown that low doses of hydrogen peroxide (H_2O_2) may be an effective management strategy of these deadly blooms. As a result, work practices for HAB management on the Norfolk Broads are being rewritten by the Environment Agency. The work outlined in this thesis was very much serendipitous, and Figure 13 summarises the timeline of events during the 4 years.

Oct 2013 – Apr 2014: Chemical synthesis of fatty acid amides as synthetic standards
Apr 2014 – Dec 2014: Culturing of *P. parvum* and detection of prymnesin-1 and -2 by LC-MS
Mar – Apr 2015: Toxic bloom on Hickling Broad – sample collection and first signs of viral infection
Apr 2015 – Mar 2016: Establishment of qPCR assay for *P. parvum*; discovery of PpDNAV
April 2016 – Nov 2016: Detection of novel prymnesin-B1 from water samples taken during bloom; first sequencing run of PpDNAV; discovery of KDN biosynthetic pathway in *P. parvum*
Dec 2016– Oct 2017: Discovery of TDP-rhamnose pathway in *P. parvum*; addition of PpDNAV to qPCR assay; first hydrogen peroxide field trials for management of *P. parvum*

Figure 13 - Approximate timeline of results arising from this thesis.

1.3 References

1. Hudnell, H. K., The state of U.S. freshwater harmful algal blooms assessments, policy and legislation. *Toxicon* **2010,** 55, (5), 1024-1034.
2. Smayda, T. J., What is a bloom? A commentary. *Limnol. Oceanogr.* **1997,** 42, (5), 1132-1136.
3. Anderson, D. M.; Cembella, A. D.; Hallegraeff, G. M., Progress in Understanding Harmful Algal Blooms: Paradigm Shifts and New Technologies for Research, Monitoring, and Management. *Ann. Rev. Mar. Sci.* **2012,** 4, (1), 143-176.
4. Granéli, E.; Edvardsen, B.; Roelke, D. L.; Hagström, J. A., The ecophysiology and bloom dynamics of *Prymnesium* spp. *Harmful Algae* **2012,** 14, 260-270.
5. Larsen, A.; Bryant, S.; Båmstedt, U., Growth rate and toxicity of *Prymnesium parvum* and *Prymnesium patelliferum*(haptophyta) in response to changes in salinity, light and temperature. *Sarsia* **1998,** 83, (5), 409-418.
6. Caron, D. A.; Alexander, H.; Allen, A. E.; Archibald, J. M.; Armbrust, E. V.; Bachy, C.; Bell, C. J.; Bharti, A.; Dyhrman, S. T.; Guida, S. M.; Heidelberg, K. B.; Kaye, J. Z.; Metzner, J.; Smith, S. R.; Worden, A. Z., Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nat. Rev. Micro.* **2017,** 15, (1), 6-20.
7. Merico, A.; Tyrrell, T.; Brown, C. W.; Groom, S. B.; Miller, P. I., Analysis of satellite imagery for *Emiliania huxleyi* blooms in the Bering Sea before 1997. *Geophys. Res. Lett.* **2003,** 30, (6).
8. Edvardsen, B.; Eikrem, W.; Throndsen, J.; Sáez, A. G.; Probert, I.; Medlin, L. K., Ribosomal DNA phylogenies and a morphological revision provide the basis for a revised taxonomy of the Prymnesiales (Haptophyta). *Eur. J. Phycol.* **2011,** 46, (3), 202-228.
9. Manton, I.; Leedale, G. F., Observations on the fine structure of *Prymnesium parvum* carter. *Archiv. Mikrobiol.* **1963,** 45, (3), 285-303.
10. Green, J. C.; Hibberd, D. J.; Pienaar, R. N., The taxonomy of *Prymnesium* (Prymnesiophyceae) including a description of a new cosmopolitan species, *P. Patellifera sp*. nov., and further observations on *P. parvum* N. carter. *Br. Phycol. J.* **1982,** 17, (4), 363-382.
11. Chang, F. H.; Ryan, K. G., *Prymnesium calathiferum* sp. nov. (Prymnesiophyceae), a new species isolated from Northland, New Zealand. *Phycologia* **1985,** 24, (2), 191-198.

12. Larsen, A.; Edvardsen, B., Relative ploidy levels in *Prymnesium parvum* and *P. patelliferum* (Haptophyta) analyzed by flow cytometry. *Phycologia* **1998,** 37, (6), 412-424.

13. Larsen, A.; Medlin, L. K., Inter- and intraspecific genetic variation in twelve *Prymesium* (haptophyceae) clones. *J. Phycol.* **1997,** 33, (6), 1007-1015.

14. Moestrup, Ø.; Thomsen, H. A., Taxonomy of toxic haptophytes (prymnesiophytes). In *Manual on Harmful Marine Microalgae*, Hallegraeff, G. M.; Anderson, D. M.; Cembella, A. D., Eds. *IOC Manuals and Guides – UNESCO* **1995**, 33, 319-338.

15. Manning, S. R.; La Claire, J. W., Prymnesins: Toxic Metabolites of the Golden Alga, *Prymnesium parvum* Carter (Haptophyta). *Mar. Drugs* **2010,** 8, (3), 678-704.

16. Larsen, A.; Eikrem, W.; Paasche, E., Growth and toxicity in *Prymnesium patelliferum* (Prymnesiophyceae) isolated from Norwegian waters. *Can. J. Bot.* **1993,** 71, (10), 1357-1362.

17. Shilo, M.; Aschner, M., Factors governing the toxicity of cultures containing the phytoflagellate *Prymnesium parvum* Carter. *J. Gen. Microbiol.* **1953,** 8, (3), 333-343.

18. Michaloudi, E.; Moustaka-Gouni, M.; Gkelis, S.; Pantelidakis, K., Plankton community structure during an ecosystem disruptive algal bloom of *Prymnesium parvum*. *J. Plankton Res.* **2009,** 31, (3), 301-309.

19. Liebert, F., Deerns, W. M., Onderzoek naar de oorzaak van een vischsterfte in den Polder Workumer-Niewland. **1920**, 1, 6.

20. Bales, M.; Moss, B.; Phillips, G.; Irvine, K.; Stansfield, J., The changing ecosystem of a shallow, brackish lake, Hickling Broad, Norfolk, U.K. II. Long-term trends in water chemistry and ecology and their implications for restoration of the lake. *Freshwat. Biol.* **1993,** 29, (1), 141-165.

21. Holdway, P. A.; Watson, R. A.; Moss, B., Aspects of the ecology of *Prymnesium parvum* (Haptophyta) and water chemistry in the Norfolk Broads, England. *Freshwat. Biol.* **1978,** 8, (4), 295-311.

22. Comin, F. A. a. F., X., Mass development of the phytoflagellate *Prymnesium parvum* Carter (Haptophyceae) in a coastal lagoon in the Ebro Delta. *Oecol. Aquat.* **1978,** 3, 3.

23. Kaartvedt, S.; Johnsen, T. M.; Aksnes, D. L.; Lie, U.; Svendsen, H., Occurrence of the toxic phytoflagellate *Prymnesium parvum* and associated fish mortality in a Norwegian fjord system. *Can. J. Fish. Aquat. Sci.* **1991,** 48, (12), 2316-2323.

24. Dietrich, W.; Hesse J. K., Local fish kill in a pond at the German North Sea coast associated with a mass development of *Prymnesium* sp. *Meeresforsch.* **1990,** 33, 104-106.

25. Lindholm, T.; Öhman, P.; Kurki-Helasmo, K.; Kincaid, B.; Meriluoto, J., Toxic algae and fish mortality in a brackish-water lake in Åland, SW Finland. *Hydrobiologia* **1999,** 397, 109-120.

26. Guo, M.; Harrison, P. J.; Taylor, F. J. R., Fish kills related to *Prymnesium parvum* N. Carter (Haptophyta) in the People's Republic of China. *J. Appl. Phycol.* **1996,** 8, (2), 111-117.

27. Roelke, D. L.; Barkoh, A.; Brooks, B. W.; Grover, J. P.; Hambright, K. D.; LaClaire, J. W.; Moeller, P. D. R.; Patino, R., A chronicle of a killer alga in the west: ecology, assessment, and management of *Prymnesium parvum* blooms. *Hydrobiologia* **2016,** 764, (1), 29-50.

28. Southard, G. M.; Fries, L. T.; Barkoh, A., *Prymnesium parvum*: The Texas Experience. *J. Am. Water. Resour. As.* **2010,** 46, (1), 14-23.

29. Authority, B., Broads Plan 2017. **2017**.

30. Shilo, M.; Sarig, S., Fish culture in warm water systems: problems and trends. *CRC Press* **1989**.

31. Smayda, T. J., Complexity in the eutrophication–harmful algal bloom relationship, with comment on the importance of grazing. *Harmful Algae* **2008,** 8, (1), 140-151.

32. Paster, Z. K., Pharmacology and mode of action of prymnesin. *Academic Press*: NY, **1973**, 241-263.

33. Dafni, Z.; Ulitzer, S.; Shilo, M., Influence of light and phosphate on toxin production and growth of *Prymnesium parvum*. *J. Gen. Microbiol.* **1972,** 70, (2), 199-207.

34. Kozakai, H.; Oshima, Y.; Yasumoto, T., Isolation and structural elucidation of hemolysin from the phytoflagellate *Prymnesium parvum*. *Agric. Biol. Chem.* **1982,** 46, (1), 233-236.

35. Bertin, M. J.; Zimba, P. V.; Beauchesne, K. R.; Huncik, K. M.; Moeller, P. D. R., The contribution of fatty acid amides to *Prymnesium parvum* Carter toxicity. *Harmful Algae* **2012,** 20, 117-125.

36. Bertin, M. J.; Zimba, P. V.; Beauchesne, K. R.; Huncik, K. M.; Moeller, P. D. R., Identification of toxic fatty acid amides isolated from the harmful alga *Prymnesium parvum* carter. *Harmful Algae* **2012,** 20, 111-116.

37. Henrikson, J. C.; Gharfeh, M. S.; Easton, A. C.; Easton, J. D.; Glenn, K. L.; Shadfan, M.; Mooberry, S. L.; Hambright, K. D.; Cichewicz, R. H., Reassessing the ichthyotoxin profile of cultured *Prymnesium parvum* (golden algae) and comparing it to samples collected from recent freshwater bloom and fish kill events in North America. *Toxicon* **2010,** 55, (7), 1396-404.

38. Igarashi, T.; Satake, M.; Yasumoto, T., Structures and partial stereochemical assignments for prymnesin-1 and prymnesin-2: potent hemolytic and ichthyotoxic glycosides isolated from the red tide alga *Prymnesium parvum*. *J. Am. Chem. Soc.* **1999,** 121, (37), 8499-8511.

39. Blossom, H. E.; Andersen, N. G.; Rasmussen, S. A.; Hansen, P. J., Stability of the intra- and extracellular toxins of *Prymnesium parvum* using a microalgal bioassay. *Harmful Algae* **2014,** 32, 11-21.

40. Granéli, E.; Johansson, N., Increase in the production of allelopathic substances by *Prymnesium parvum* cells grown under N- or P-deficient conditions. *Harmful Algae* **2003,** 2, (2), 135-145.

41. Baker, J. W.; Grover, J. P.; Brooks, B. W.; Ureña-Boeck, F.; Roelke, D. L.; Errera, R.; Kiesling, R. L., Growth and toxicity of *Prymnesium parvum* (haptophyta) as a function of salinity, light, and temperature. *J. Phycol.* **2007,** 43, (2), 219-227.

42. Remmel, E. J.; Hambright, K. D., Toxin-assisted micropredation: experimental evidence shows that contact micropredation rather than exotoxicity is the role of *Prymnesium* toxins. *Ecol. Lett.* **2012,** 15, (2), 126-132.

43. Collins, M., Algal Toxins. *Microbiological Rev.* **1978,** 42, (4), 20.

44. Wu, Q.; Kazantzis, M.; Doege, H.; Ortegon, A. M.; Tsang, B.; Falcon, A.; Stahl, A., Fatty acid transport protein 1 is required for nonshivering thermogenesis in brown adipose tissue. *Diabetes* **2006,** 55, (12), 3229-3237.

45. Luyen, Q. H.; Cho, J. Y.; Choi, J. S.; Kang, J. Y.; Park, N. G.; Hong, Y. K., Isolation of algal spore lytic C17 fatty acid from the crustose coralline seaweed *Lithophyllum yessoense*. *J. Appl. Phycol.* **2008,** 21, (4), 423.

46. Hannun, Y. A., Functions of ceramide in coordinating cellular responses to stress. *Science (New York, N.Y.)* **1996,** 274, (5294), 1855-9.

47. Dembitsky, V. M.; Shkrob, I.; Rozentsvet, O. A., Fatty acid amides from freshwater green alga *Rhizoclonium hieroglyphicum*. *Phytochemistry* **2000,** 54, (8), 965-967.

48. Wu, M.; Milligan, K. E.; Gerwick, W. H., Three new malyngamides from the marine cyanobacterium *Lyngbya majuscula*. *Tetrahedron* **1997,** 53, (47), 15983-15990.

49. Sitachitta, N.; Gerwick, W. H., Grenadadiene and grenadamide, cyclopropyl-containing fatty acid metabolites from the marine cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod.* **1998,** 61, (5), 681-4.

50. Arafat, E. S.; Trimble, J. W.; Andersen, R. N.; Dass, C.; Desiderio, D. M., Identification of fatty acid amides in human plasma. *Life Sci.* **1989,** 45, (18), 1679-1687.

51. Cravatt, B. F.; Lerner, R. A.; Boger, D. L., Structure Determination of an Endogenous Sleep-Inducing Lipid, cis-9-Octadecenamide (Oleamide): A synthetic approach to the chemical analysis of trace quantities of a natural product. *J. Am. Chem. Soc.* **1996,** 118, (3), 580-590.

52. McDonald, G. R.; Hudson, A. L.; Dunn, S. M.; You, H.; Baker, G. B.; Whittal, R. M.; Martin, J. W.; Jha, A.; Edmondson, D. E.; Holt, A., Bioactive contaminants leach from disposable laboratory plasticware. *Science (New York, N.Y.)* **2008,** 322, (5903), 917-917.

53. Blossom, H. E.; Rasmussen, S. A.; Andersen, N. G.; Larsen, T. O.; Nielsen, K. F.; Hansen, P. J., *Prymnesium parvum* revisited: Relationship between allelopathy, ichthyotoxicity, and chemical profiles in 5 strains. *Aquat. Toxicol.* **2014,** 157, 159-166.

54. Igarashi, T.; Satake, M.; Yasumoto, T., Prymnesin-2: A potent ichthyotoxic and hemolytic glycoside isolated from the red tide alga *Prymnesium parvum*. *J. Am. Chem. Soc.* **1996,** 118, (2), 479-480.

55. Misra, S.; Valicherla, G. R.; Mohd, S.; Gupta, J.; Gayen, J. R.; Misra-Bhattacharya, S., UDP-galactopyranose mutase, a potential drug target against human pathogenic nematode *Brugia malayi*. *Pathog. Dis.* **2016,** 74, (6).

56. Kizjakina, K.; Tanner, J. J.; Sobrado, P., Targeting UDP-galactopyranose mutases from eukaryotic human pathogens. *Curr. Pharm. Des.* **2013,** 19, (14), 2561-73.

57. Turnbull, W. B.; Stalford, S. A., Methylthioxylose–a jewel in the mycobacterial crown? *Org. Biomol. Chem.* **2012,** 10, (30), 5698-5706.

58. Manning, S. R.; La Claire Ii, J. W., Isolation of polyketides from *Prymnesium parvum* (Haptophyta) and their detection by liquid chromatography/mass spectrometry metabolic fingerprint analysis. *Anal. Biochem.* **2013,** 442, (2), 189-195.

59. Rasmussen, S. A.; Meier, S.; Andersen, N. G.; Blossom, H. E.; Duus, J. Ø.; Nielsen, K. F.; Hansen, P. J.; Larsen, T. O., Chemodiversity of ladder-frame prymnesin polyethers in *Prymnesium parvum*. *J. Nat. Prod.* **2016,** 79, (9), 2250-2256.

60. Valiadi, M.; Iglesias-Rodriguez, D., Understanding bioluminescence in dinoflagellates—How far have we come? *Microorganisms* **2013,** 1, (1), 3-25.

61. Wang, D. Z., Neurotoxins from marine dinoflagellates: A brief review. *Mar. Drugs* **2008,** 6, (2), 349-371.

62. Watkins, S. M.; Reich, A.; Fleming, L. E.; Hammond, R., Neurotoxic shellfish poisoning. *Mar. Drugs* **2008,** 6, (3), 431-55.

63. Lin, Y.-Y.; Risk, M.; Ray, S. M.; Van Engen, D.; Clardy, J.; Golik, J.; James, J. C.; Nakanishi, K., Isolation and structure of brevetoxin B from the "red tide" dinoflagellate

Ptychodiscus brevis (*Gymnodinium breve*). *J. Am. Chem. Soc.* **1981,** 103, (22), 6773-6775.

64. Baden, D. G., Marine food-borne dinoflagellate toxins. *Int. Rev. Cytol.* **1983,** 82, 99-150.
65. Yasumoto, T.; Murata, M., Marine toxins. *Chem. Rev.* **1993,** 93, (5), 1897-1909.
66. Yasumoto, T.; Murata, M., Polyether Toxins Involved in Seafood Poisoning. In *Marine Toxins*, American Chemical Society, **1990**, 418, 120-132.
67. Takahashi, M.; Ohizumi, Y.; Yasumoto, T., Maitotoxin, a Ca^{2+} channel activator candidate. *J. Biol. Chem.* **1982,** 257, (13), 7287-89.
68. Korsnes, M. S.; Røed, S. S.; Tranulis, M. A.; Espenes, A.; Christophersen, B., Yessotoxin triggers ribotoxic stress. *Toxicol. In Vitro*, **2014**, 28, (5), 975-981.
69. Houdan, A.; Bonnard, A.; Fresnel, J.; Fouchard, S.; Billard, C.; Probert, I., Toxicity of coastal coccolithophores (Prymnesiophyceae, Haptophyta). *J. Plankton Res.* **2004,** 26, (8), 875-883.
70. Johansson, N.; Granéli, E., Influence of different nutrient conditions on cell density, chemical composition and toxicity of *Prymnesium parvum* (Haptophyta) in semi-continuous cultures. *J. Exp. Mar. Biol. Ecol.* **1999,** 239, (2), 243-258.
71. Roelke, D. L.; Errera, R. M.; Kiesling, R.; Brooks, B. W.; Grover, J. P.; Schwierzke, L.; Urena-Boeck, F.; Baker, J.; Pinckney, J. L., Effects of nutrient enrichment on *Prymnesium parvum* population dynamics and toxicity: results from field experiments, Lake Possum Kingdom, USA. *Aquat. Microb. Ecol.* **2007,** 46, (2), 125-140.
72. McLaughlin, J. J. A., Euryhaline Chrysomonads: nutrition and toxigenesis in *Prymnesium parvum*, with notes on *Isochrysis galban*a and *Monochrysis lutheri*. *J. Protozool.* **1958,** 5, (1), 75-81.
73. Ulitzur, S.; Shilo, M., A sensitive assay system for determination of the ichthyotoxicity of *Prymnesium parvum*. *J. Gen. Microbiol.* **1964,** 36, (2), 161-169.
74. Yariv, J.; Hestrin, S., Toxicity of the extracellular phase of *Prymnesium parvum* cultures. *J. Gen. Microbiol.* **1961,** 24, (2), 165-175.
75. Shilo, M.; Rosenberger, R. F., Studies on the toxic principles formed by the chrysomonad *Pryminesium parvum* carter. *Ann. N.Y. Acad. Sci.* **1960,** 90, (3), 866-876.
76. Rijn, J. S., M., Environmental Factors in Fish Culture Systems. In *Fish Culture in Warm Water Systems: Problems and Trends*, Sarig, M. S.; Shilo, M., Ed. CRC Press, Inc.: Florida, **1989**; pp 163-177.
77. Pinkerton, M.; Steinrauf, L. K., Molecular structure of monovalent metal cation complexes of monensin. *J. Mol. Biol.* **1970,** 49, (3), 533-546.

78. Parnas, I.; Reich, K.; Bergmann, F., Photoinactivation of ichthyotoxin from axenic cultures of *Prymnesium parvum* Carter. *Appl. Microbiol.* **1962,** 10, 237-9.

79. Shilo, M., Formation and mode of action of algal toxins. *Bacteriol. Rev.* **1967,** 31, (3), 180-93.

80. Rahat, M.; Hochberg, A., Ethionine and Methionine Metabolism by the Chrysomonad Flagellate *Prymnesium parvum*. *J. Protozool.* **1971,** 18, (3), 378-382.

81. Wilhelm, S. W.; Coy, S. R.; Gann, E. R.; Moniruzzaman, M.; Stough, J. M. A., Standing on the shoulders of giant viruses: five lessons learned about large viruses infecting small eukaryotes and the opportunities they create. *PLoS Pathog.* **2016,** 12, (8).

82. Tillmann, U., Interactions between planktonic microalgae and protozoan grazers. *J. Eukaryot. Microbiol.* **2004,** 51, (2), 156-68.

83. La Scola, B.; Audic, S.; Robert, C.; Jungang, L.; de Lamballerie, X.; Drancourt, M.; Birtles, R.; Claverie, J. M.; Raoult, D., A giant virus in amoebae. *Science (New York, N.Y.)* **2003,** 299, (5615), 2033-2033.

84. Short, S. M., The ecology of viruses that infect eukaryotic algae. *Environ. Microbiol.* **2012,** 14, (9), 2253-71.

85. Brussaard, C. P. D., Viral control of phytoplankton populations—a review. *J. Eukaryot. Microbiol.* **2004,** 51, (2), 125-138.

86. Suttle, C. A., Marine viruses - major players in the global ecosystem. *Nat. Rev. Micro.* **2007,** 5, (10), 801-812.

87. Wilhelm, S. W.; Suttle, C. A., Viruses and nutrient cycles in the sea: viruses play critical roles in the structure and function of aquatic food webs. *Bioscience* **1999,** 49, (10), 781-788.

88. Malin, G.; Wilson, W. H.; Bratbak, G.; Liss, P. S.; Mann, N. H., Elevated production of dimethylsulfide resulting from viral infection of cultures of *Phaeocystis pouchetii*. *Limnol. Oceanogr.* **1998,** 43, (6), 1389-1393.

89. Nagasaki, K.; Yamaguchi, M., Isolation of a virus infectious to the harmful bloom causing microalga *Heterosigma akashiwo* (Raphidophyceae). *Aquat. Microb. Ecol.* **1997,** 13, (2), 135-140.

90. Johannessen, T. V.; Bratbak, G.; Larsen, A.; Ogata, H.; Egge, E. S.; Edvardsen, B.; Eikrem, W.; Sandaa, R.-A., Characterisation of three novel giant viruses reveals huge diversity among viruses infecting Prymnesiales (Haptophyta). *Virology* **2015,** 476, 180-188.

91. Schwierzke, L.; Roelke, D. L.; Brooks, B. W.; Grover, J. P.; Valenti, T. W.; Lahousse, M.; Miller, C. J.; Pinckney, J. L., *Prymnesium parvum* population dynamics during bloom

development: a role assessment of grazers and virus. *J. Am. Water. Resour. As.* **2010,** 46, (1), 63-75.

92. Roelke, D. L.; Schwierzke, L.; Brooks, B. W.; Grover, J. P.; Errera, R. M.; Valenti, T. W.; Pinckney, J. L., Factors influencing *Prymnesium parvum* population dynamics during bloom initiation: results from in-lake mesocosm experiments. *J. Am. Water. Resour. As.* **2010,** 46, (1), 76-91.
93. Antonella, P.; Luca, G., The quantitative real-time PCR applications in the monitoring of marine harmful algal bloom (HAB) species. *Environ. Sci. Pollut. Res.* **2013,** 20, (10), 6851-6862.
94. Casper, E. T.; Paul, J. H.; Smith, M. C.; Gray, M., Detection and quantification of the red tide dinoflagellate *Karenia brevis* by real-time nucleic acid sequence-based amplification. *Appl. Environ. Microbiol.* **2004,** 70, (8), 4727-4732.
95. Yuan, J.; Mi, T.; Zhen, Y.; Yu, Z., Development of a rapid detection and quantification method of *Karenia mikimotoi* by real-time quantitative PCR. *Harmful Algae* **2012,** 17, (Supplement C), 83-91.
96. Kamikawa, R.; Asai, J.; Miyahara, T.; Murata, K.; Oyama, K.; Yoshimatsu, S.; Yoshida, T.; Sako, Y., Application of a real-time PCR assay to a comprehensive method of monitoring harmful algae. *Microbes Environ.* **2006,** 21, (3), 163-173.
97. Galluzzi, L.; Bertozzini, E.; Penna, A.; Perini, F.; Pigalarga, A.; Graneli, E.; Magnani, M., Detection and quantification of *Prymnesium parvum* (Haptophyceae) by real-time PCR. *Lett. Appl. Microbiol.* **2008,** 46, (2), 261-266.
98. Zamor, R. M.; Glenn, K. L.; Hambright, K. D., Incorporating molecular tools into routine HAB monitoring programs: Using qPCR to track invasive Prymnesium. *Harmful Algae* **2012,** 15, (Supplement C), 1-7.
99. Shirota, A., Red tide problem and countermeasures. *Int. J. Aquacult. Fish. Technol.* **1989,** 1, 19-19.
100. Sengco, M. R.; Anderson, D. M., Controlling harmful algal blooms through clay flocculation. *J. Eukaryot. Microbiol.* **2004,** 51, (2), 169-172.
101. Hagström, J. A.; Sengco, M. R.; Villareal, T. A., Potential methods for managing *Prymnesium parvum* blooms and toxicity, with emphasis on clay and barley straw: a review. *J. Am. Water. Resour. As.* **2010,** 46, (1), 187-198.
102. Barrett, P. R. F.; Littlejohn, J. W.; Curnow, J., Long-term algal control in a reservoir using barley straw. *Hydrobiologia* **1999,** 415, 309-313.
103. Martin, D.; Ridge, I., The relative sensitivity of algae to decomposing barley straw. *J. Appl. Phycol.* **1999,** 11, (3), 285-291.

104. Barkoh, A.; Paret, J. M.; Lyon, D. D.; Begley, D. C.; Smith, D. G.; Schlechte, J. W., Evaluation of barley straw and a commercial probiotic for controlling *Prymnesium parvum* in fish production ponds. *N. Am. J. Aquacult.* **2008,** 70, (1), 80-91.

105. Grover, J. P.; Baker, J. W.; Urena-Boeck, F.; Brooks, B. W.; Errera, R. M.; Roelke, D. L.; Kiesling, R. L., Laboratory tests of ammonium and barley straw extract as agents to suppress abundance of the harmful alga *Prymnesium parvum* and its toxicity to fish. *Water Res.* **2007,** 41, (12), 2503-12.

106. Rodgers Jr, J. H.; Johnson, B. M.; Bishop, W. M., Comparison of three algaecides for controlling the density of *Prymnesium parvum*. *J. Am. Water. Resour. As.* **2010,** 46, (1), 153-160.

107. Drabkova, M.; Admiraal, W.; Marsalek, B., Combined exposure to hydrogen peroxide and light--selective effects on cyanobacteria, green algae, and diatoms. *Environ. Sci. Technol.* **2007,** 41, (1), 309-14.

108. Matthijs, H. C.; Visser, P. M.; Reeze, B.; Meeuse, J.; Slot, P. C.; Wijn, G.; Talens, R.; Huisman, J., Selective suppression of harmful cyanobacteria in an entire lake with hydrogen peroxide. *Water Res.* **2012,** 46, (5), 1460-72.

109. Burson, A.; Matthijs, H. C. P.; de Bruijne, W.; Talens, R.; Hoogenboom, R.; Gerssen, A.; Visser, P. M.; Stomp, M.; Steur, K.; van Scheppingen, Y.; Huisman, J., Termination of a toxic Alexandrium bloom with hydrogen peroxide. *Harmful Algae* **2014,** 31, (0), 125-135.

2 Genetic and Metabolite Assessment of a Harmful Algal Bloom; *Prymnesium parvum* and its Toxins

2.1 Abstract

Prymnesium parvum is a toxin-producing microalga that causes harmful algal blooms globally, frequently leading to massive fish kills that have adverse ecological and economic impacts. The dramatic effects observed on fish are suspected to be due to the ladder-frame polyether toxins prymnesin-1 and -2, but a lack of confirmation of the presence of these compounds in environmental samples has resulted in significant ambiguity about the toxic entity. A fish-killing bloom, suspected to be due to *P. parvum*, was reported in the Norfolk Broads, United Kingdom, in March 2015 at a site historically plagued by *Prymnesium* blooms. Here, we report confirmation that *P. parvum* dominated the microbial community during the bloom and show how the diversity of microbes changed after recovery of the system 18 months later. We also report, for the first time, the detection of the recently discovered prymnesin-B1 toxin [1] in natural water samples, and in gill tissue isolated from a dead pike (*Esox lucius*) taken from the site of the bloom. Following on from this work, we developed a sensitive quantitative real-time PCR assay for monitoring *P. parvum* and report algal population dynamics over a 20-month period on Hickling Broad.

2.2 Introduction

The frequency and distribution of harmful algal blooms (HABs) worldwide has increased in recent years due to a combination of eutrophication and climate change issues, often leading to serious ecological and economic impacts on native fish populations. The diverse division of microalgae known as the haptophytes are no exception to this scenario. This phylum includes the bloom-forming coccolithore, *Emiliania huxleyi*, whose oceanic blooms are so large they can often be seen from space [2]. HABs are not just a coastal issue however; blooms of the toxin-producing haptophyte, *Prymnesium parvum*, are more frequently associated with brackish, inland waters and aquaculture ponds, although this euryhaline cosmopolitan alga can be found in most water types worldwide [3]. The Norfolk Broads in East Anglia, England, have suffered from recurring fish-kills associated with *P. parvum* since the late 1960s [4], when a catastrophic bloom left many of the Broads almost completely devoid of fish. This was a devastating blow for the local economy of the area, which usually draws in tourists for the high-profile angling and boating activities that contribute over £500m annually to the Norfolk/Suffolk economy [5].

Ever since *P. parvum* was implicated in fish kills, there has been significant uncertainty about the toxic entities responsible for fish kills. A range of chemicals have been proposed to be involved in toxicity, including lipopolysaccharide-like compounds [6], galactoglycerolipids [7], fatty acid amides [8], fatty acids [9] and the ladder-frame polyether toxins prymnesin-1 and -2 [10, 11] (Figure 14). Some of these have since been ruled out [12], and more recently, the general consensus that the ladder-frame polyether toxins, prymnesin-1 and -2, are the primary toxins has been questioned. This questioning has largely been a result of the inability of numerous research groups to detect the toxins in either laboratory or field samples. Detection of prymnesin-1 and -2 *in vitro* was achieved in 2013 [13] for the first time since the original isolation in 1996. Recently, Rasmussen *et al* [1] have shown a large structural diversity of the prymnesin toxins amongst different strains of *Prymnesium*, and this likely contributed significantly to the failure to detect these compounds previously. Prior to the study reported in this thesis, however, these structurally impressive secondary metabolites had never been detected in wild, environmental samples.

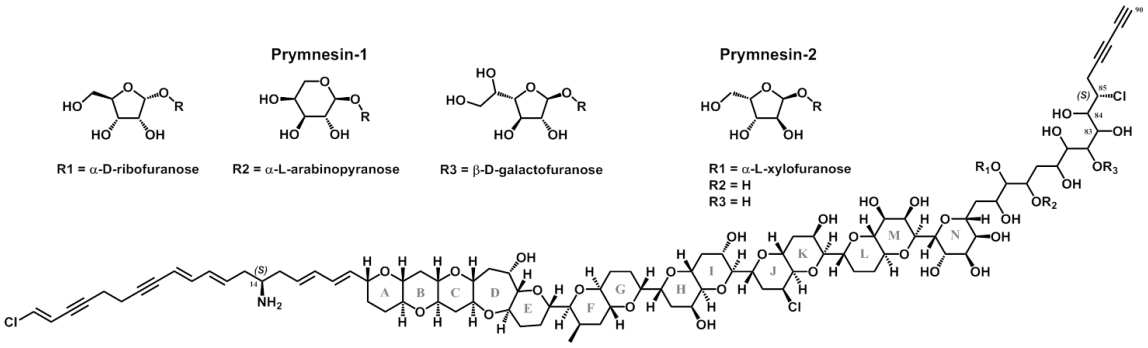


Figure 14 - Structure of prymnesin-1 and -2. Toxins were first isolated by Igarashi *et al* in 1996 [11].

To add to the ambiguity surrounding the metabolites involved in toxicity, very little is known about how *P. parvum* blooms affect or are affected by the surrounding microbial community. For other algal blooms, it has been shown that bacterial community compositions change significantly as blooms peak and decay [14, 15]. These changes may be the result of toxins produced by the alga, but are also likely a result of nutrient availability, with many researchers showing that phytoplankton blooms provide a plethora of dissolved organic material (DOM), produced by photosynthesis, to heterotrophic bacterial populations [16, 17]. As well as the photosynthetic DOM produced during a bloom, bacteria are also able to utilise the DOM released by dead algal cells following the demise of a bloom. Some bacteria are more equipped to do this and will actively break down dead algal cells [18, 19].

Monitoring for the onset of harmful algal blooms is one way to contribute to mitigating the drastic effects observed during toxic blooms. In recent years, molecular methods such as quantitative real-time PCR have been used to monitor toxin-producing algal species [20]. Some key examples of success lie with the dinoflagellates of the *Karenia* genus [21] and the raphidophyte *Heterosigma* sp. [22]. Efforts to monitor *P. parvum* using qRT-PCR have also found some success, with Galluzzi *et al* and Zamor *et al* showing how qRT-PCR can mirror and improve optical microscopy methods, which are currently the method of choice for monitoring *P. parvum* on the Norfolk Broads [23, 24].

In the present study, we explored the microbial composition on Hickling Broad during and after a harmful bloom in March 2015, and confirm that *P. parvum* dominated the brackish water microbiota during the mass fish mortality event. Ladder-frame polyether toxins produced by *P. parvum* were detected in natural water samples for the first time, and we were also able to detect these toxins in the gill cells of a dead pike at the site of the bloom; in

agreement with the speculated biological target as being the gill cell membranes [25]. Although a cocktail of toxins is still possible, these results strongly reinforce the suggested role of the polyether prymnesins in fish toxicity. Furthermore, we developed a sensitive qPCR-based assay for monitoring *P. parvum* in natural waters, and suggest that this may be an effective way of predicting the onset of toxic blooms of the *P. parvum*.

2.3 Results

2.3.1 Study site – fish kill March 2015

The Norfolk Broads are a low-lying area of navigable lakes spanning Norfolk and Suffolk, connected to the River Waveney, Yare, Bure and its' two tributaries, the Ant and Thurne. Whilst algal blooms pose an issue across several of the 63 Broads, blooms of *P. parvum* are somewhat confined to the upper Thurne system, frequently posing issues to Hickling Broad and the surrounding area (Figure 15).

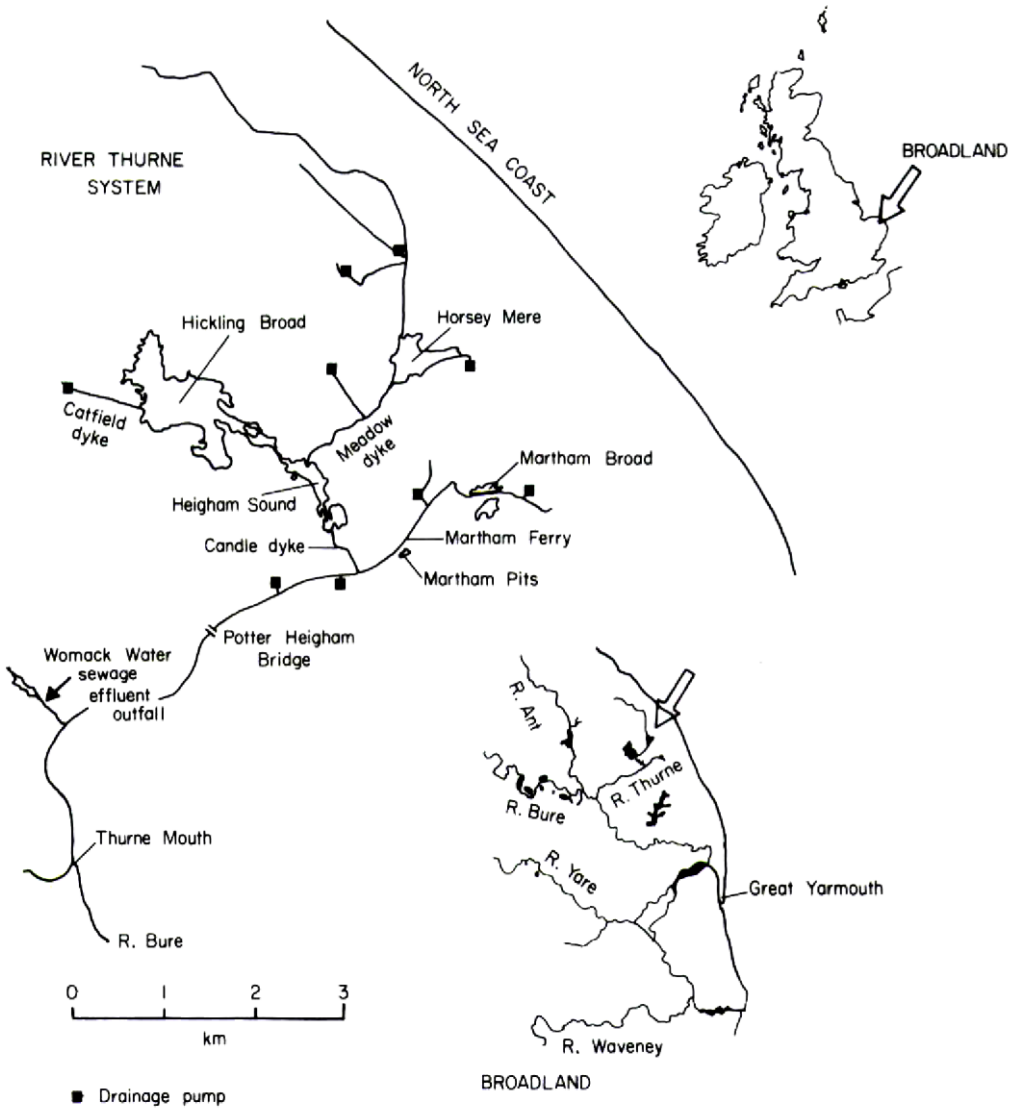


Figure 15 – Location of the Upper Thurne area of the Norfolk Broads. Reprinted (adapted) with permission from P. Holdway *et al*, Freshwater Biology, 1978, 8, 295-311. Copyright 1977 John Wiley and Sons.

To maintain consistent sampling, 11 regular sampling sites were set up on Hickling Broad as detailed in Figure 16. These sites were used for subsequent metabolite and genetic analysis of water samples. A relevant fish kill was first reported on 13th March 2015 when members of the public reported numerous distressed or dead fish across Hickling Broad. First inspection and water samples were taken on 17th March 2015, and locations of fish kills were recorded qualitatively: it was found that most fish deaths were confined to the Northwest of the Broad, closest to sampling sites 6 and 7, but spreading out as far as sample location 5 (Figure 16).

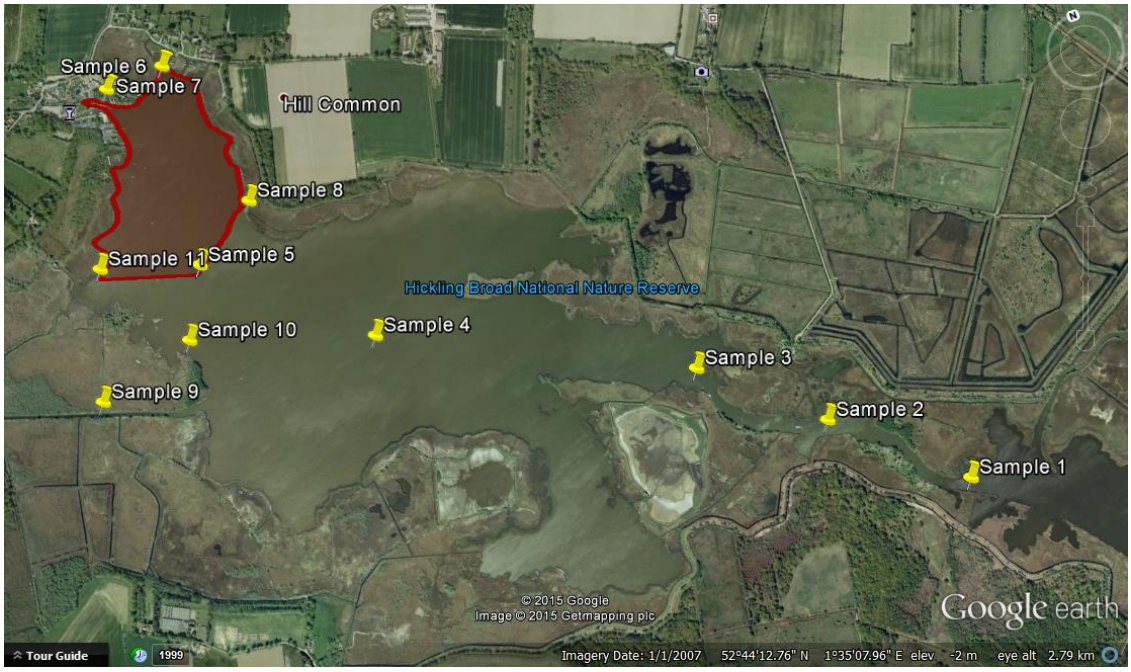


Figure 16 – Aerial view of Hickling Broad and the 11 sampling sites developed during this research. Most fish deaths were observed on the Northwest area of the Broad as denoted by a coloured red area. Google earth. Hickling Broad, Norfolk, Norwich, United Kingdom. 52°44'12.76"N, 1°35'07.96"E, eye alt 2.79 km. Copyrights: Google 2015, Getmapping plc 2015.

2.3.2 Optical microscopy of water samples

Optical microscopy of water samples taken from locations 1-11 during the harmful bloom qualitatively identified *P. parvum* as the cause of the harmful bloom. Many algal cells observed fitted typical morphological characteristics of *P. parvum* [26, 27]; cells were golden yellow in appearance, approximately 10 µm long, ellipsoidal in shape and were motile with 2 flagella and a smaller more centrally located haptonema (Figure 17).

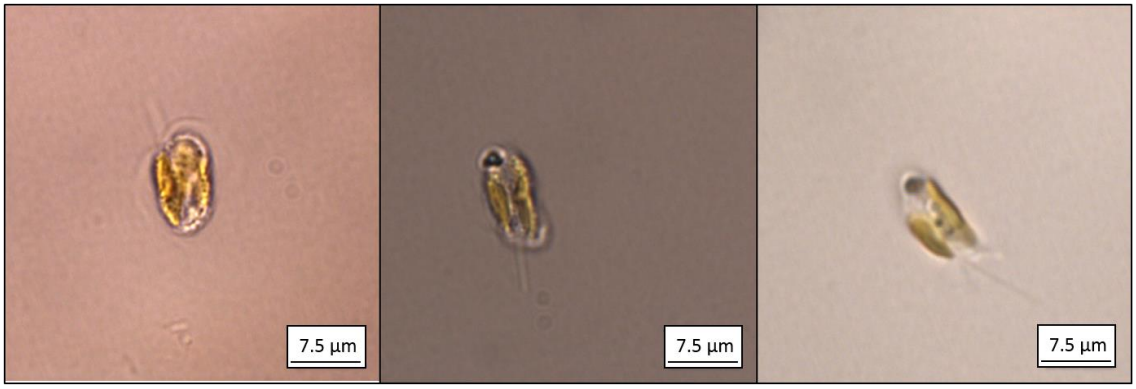


Figure 17 – Optical microscopy of *P. parvum* cells identified from water samples taken from Hickling Broad during a toxic algal bloom in March 2015. Cells display typical characteristics of *P. parvum* with lengths of approximately 10 µm, a golden yellow colour, ellipsoidal shapes and 2 flagella and 1 haptonema from one end of the cell. Scale bars are 7.5 µm.

2.3.3 Genetic analysis of microbial community

Water samples were taken from sample sites 1-11 on 17th March 2015 during a harmful bloom and then again during non-bloom conditions in September 2016. Bacterial and algal cells were pelleted by centrifugation and cell pellets were then provided to collaborating scientists at UEA (Dr Jennifer Pratscher) for nucleic acid extraction and sent to Mr DNA Lab (www.mrdnalab.com, Texas) for community analysis via 16S rRNA gene sequencing ( ).

P. parvum made up a significantly larger percentage of the total microbial community during bloom conditions (up to 44% in location 9) compared to non-bloom conditions (up to 2% in location 6), suggesting that it was responsible for the toxic bloom in March 2015. A much more diverse range of bacteria and algae were observed under non-bloom conditions, including cyanobacteria such as *Microcystis* sp., indicating the significant effect of the bloom on the microbial community. Although subtle differences in abundance can be seen, community compositions are generally very similar from location to location.

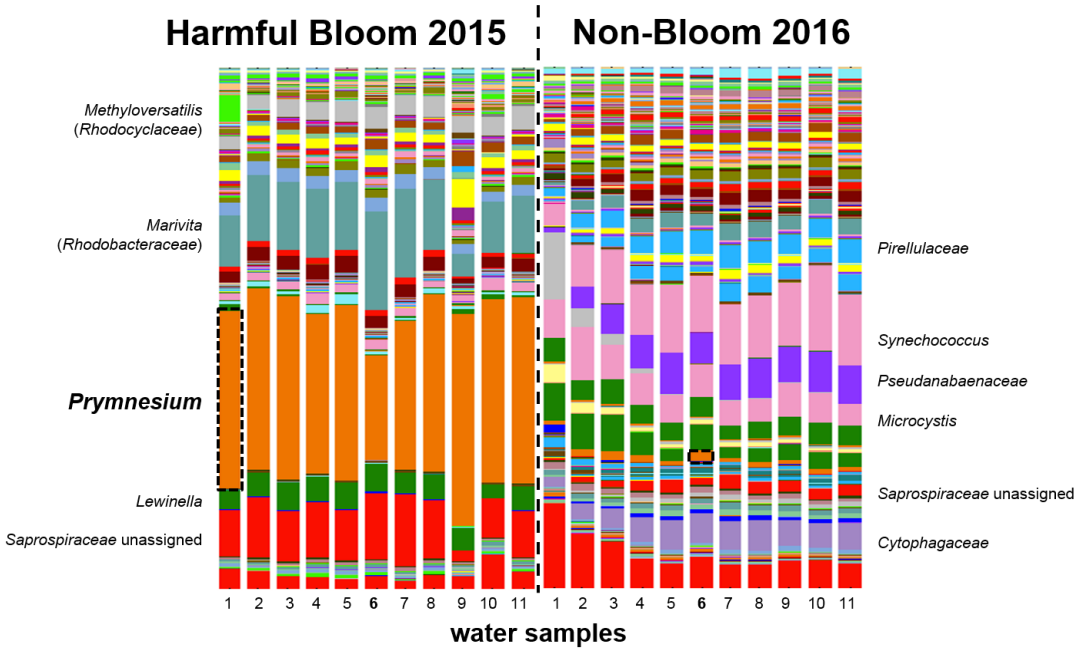


Figure 18 - Community analysis of water samples taken during a harmful bloom (left) and under non-bloom conditions (right). Total DNA extracted from water samples was sequenced using 16S sequencing and relative abundance of transcripts in each sample is reported as a percentage of the total community at 11 different sampling locations (y-axis). Each coloured bar represents a different family of bacteria or algae. Orange coloured *P. parvum* bars are indicated at location 1 with a black hashed border. *P. parvum* made up to 44% in some sampling locations during bloom conditions (location 9), and only up to 2% in non-bloom conditions (location 6).

Alongside *P. parvum*, the most dominant families observed during the harmful bloom in Spring 2015 include the gram-negative Methyloversatilis genus of the Rhodocyclaceae family, the Marivita genus of the Rhodobacteraceae family, the Lewinella genus of the Saprospiraceae family, and other unassigned members of the Saprospiraceae family. Although this analysis is based on 16S rRNA gene sequencing and will therefore exclude a significant proportion of eukaryotes that may exist in a waterway, it is interesting to note that eukaryotes with plastid 16S sequences are significantly less abundant than *P. parvum* when compared to prokaryotes, with all other major taxa found during the bloom being prokaryotic. Except for the Saprospiraceae family, the dominant families found in non-bloom conditions differ significantly from bloom conditions. Dominant families found under non-bloom conditions include the Pirellulaceae family, the Cytophagaceae family, and three dominant genera of cyanobacteria; Synechococcus, Pseudanabaena and Microcystis. These drastic changes to levels of cyanobacteria during non-bloom conditions warrant further investigation.

2.3.4 Detection of Prymnesin-1 & -2 in *P. parvum* 946/6

Intra and extracellular extractions of prymnesin-1 and -2 were performed on *P. parvum* 946/6 cultures. In general, crude prymnesins were extracted following a modified method of Manning *et al* [13] and analysed by LC-MS (Figure 19). Chromatograms were analysed for ions related to a range of prymnesin toxins [1]. An example of a typical isotope pattern seen for the aglycone form of both prymnesin toxins can be seen in Figure 20, where deconvolution of the m/z mass gives an error < 5 ppm. Prymnesin-1 and -2 could be found in both cell extracts and culture medium, and agreed well with fragments and adducts previously reported by Manning *et al* [13] (Table 1).

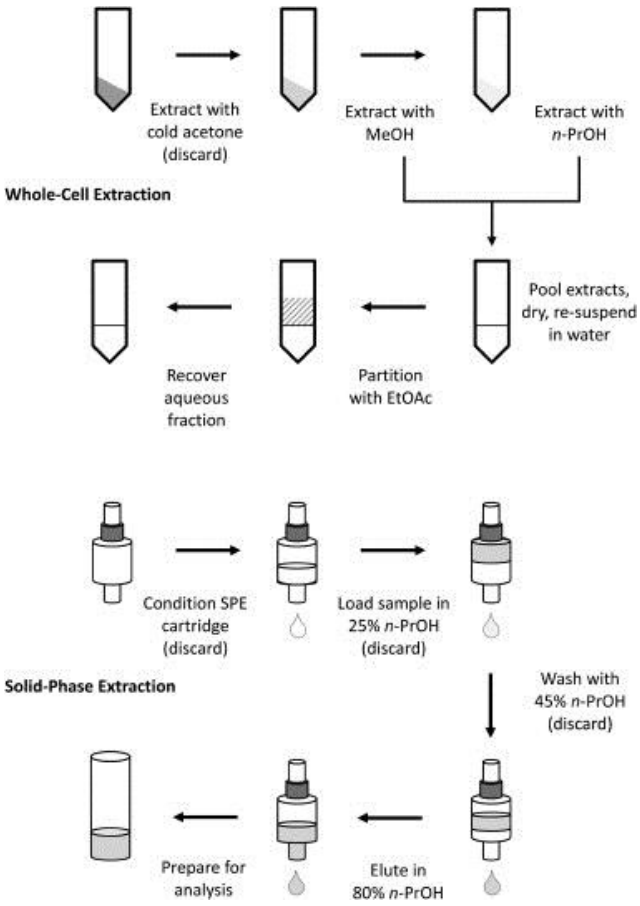


Figure 19 – Workflow for the extraction of prymnesins from *P. parvum.* Solvent conditions and extraction was followed as detailed by Manning and La Claire [13]. Image reprinted from Analytical Biochemistry, 442, 2, Shonna R. Manning and John W. La Claire II, Isolation of polyketides from *Prymnesium parvum* (Haptophyta) and their detection by liquid chromatography/mass spectrometry metabolic fingerprint analysis, 189-195, Copyright 2013, with permission from Elsevier.

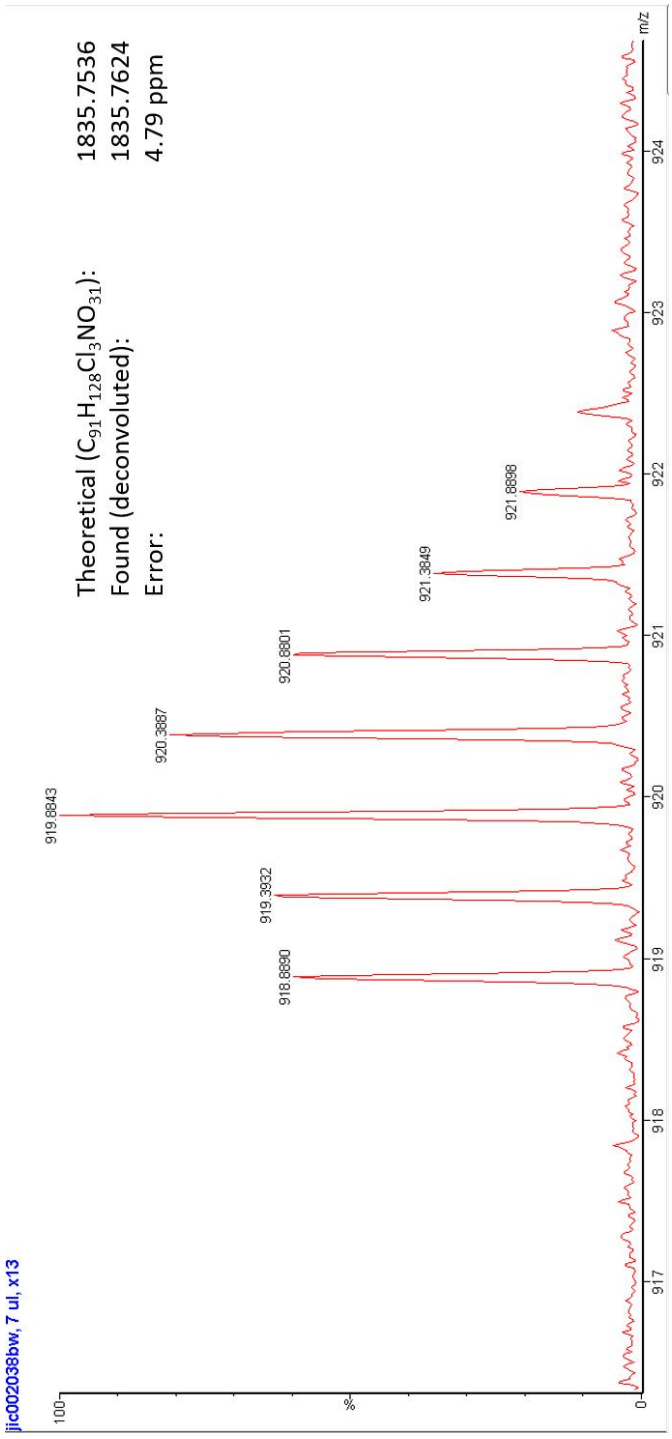


Figure 20 – Isotope pattern observed for aglycone form of the prymnesin toxins. Distribution of signals agrees well with that expected for the chlorinated prymnesin toxins [13]. Calculation of the deconvoluted mass from the m/z value of 918.8890 gives an accurate mass within 5 ppm of the expected mass.

m/z (2+)	Deconvoluted mass	Assignment
882.90	1763.79	Prym(agly) -2HCl
891.90	1781.79	Unknown
900.90	1800.79	Prym(agly) - HCl
918.90	1835.79	Prym(agly)
948.90	1895.79	Prym(agly) +K +Na -2H
957.90	1913.79	Prym(agly) - HCl + TFA
966.90	1931.79	Prym2 - HCl
984.90	1967.79	Prym2
1014.93	2027.85	Unknown
1023.96	2045.91	Prym2 -HCl + TFA
1032.97	2063.93	Prym1 - Hexose - HCl
1047.96	2093.91	Prym1 - Pentose - HCl
1050.93	2099.85	Prym1 - Hexose
1065.94	2129.87	Prym1 - Pentose
1093.50	2184.99	Unknown
1102.00	2201.99	Unknown
1114.00	2225.99	Prym1 - HCl
1123.50	2242.99	Prym1 - HCl +NH_4 -H
1131.98	2261.95	Prym1
1140.50	2276.99	Prym1 + NH_4 -H

Table 1 - Masses identified for prymnesin-1 and -2 and their adducts from cellular and extracellular preparations. Agly = aglycone form shared by both prymnesin-1 and -2. Masses highlighted red are those previously reported by Manning and La Claire [13]. Other masses reported represent novel masses with their suspected assignments.

2.3.5 Detection of Prymnesin-B1 in natural water samples and fish gills

For the detection of prymnesin toxins from water samples, 100 ml of water from locations 6 and 7 were extracted and analysed by LC-MS. Chromatograms were analysed for ions relating to prymnesin-1 and -2, as well as recently discovered prymnesins outlined by Rasmussen *et al* [1]. Of the prymnesins previously identified by Igarashi *et al* and Rasmussen *et al* [1, 10], only prymnesin-B1 was detected in locations 6, 7 and in the gill cells of a deceased pike (Figure 21**Error! Reference source not found.**). Isotope patterns matched those reported by Rasmussen *et al* (Figure 22), and although standards are not available retention times are in line with those previously reported for these toxins (elution at 70-80% acetonitrile under conditions reported by Manning and La Claire [13]). Toxins were not detectable in water samples taken in September 2016 under non-bloom conditions, suggesting high *P. parvum* cell counts correspond with detectable toxins in the waterways.

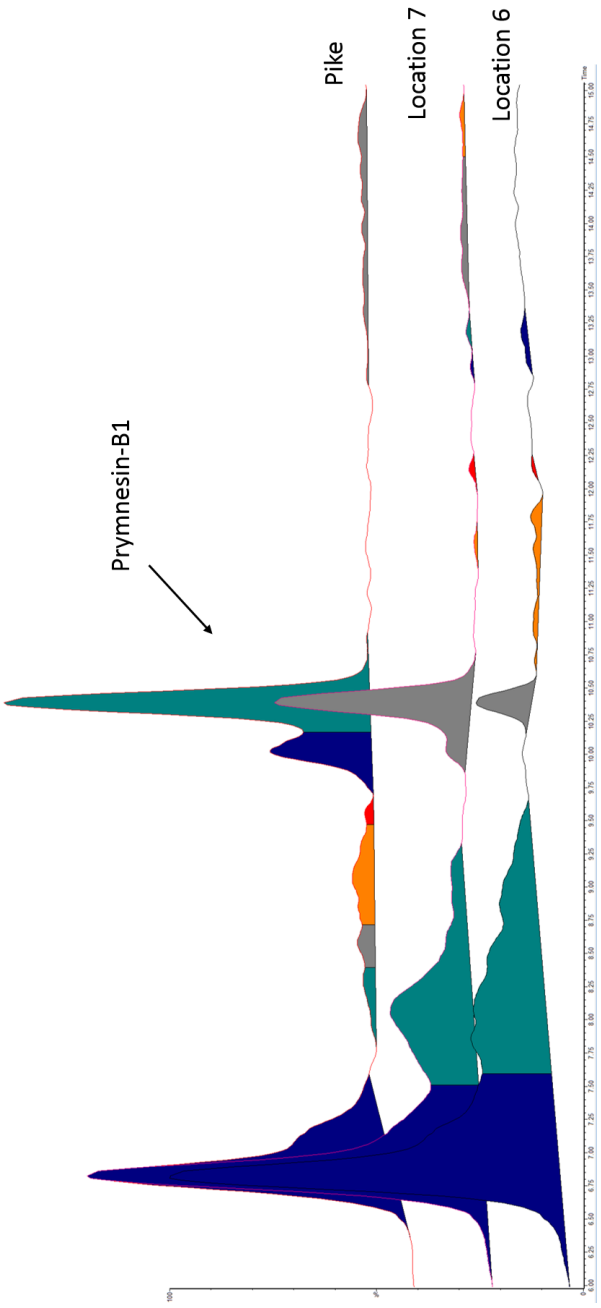


Figure 21– Detection of prymnesin-B1 from natural water samples and gill tissue. Extracted ion chromatograms (mass 828.896) for extracts prepared from water samples taken at location 6 & 7, and from the gill tissue of a dead pike taken from the site of the toxic-bloom. Peaks representing prymnesin-B1 are labelled. Earlier peaks found around 6 minutes retention time represent unrelated compounds.

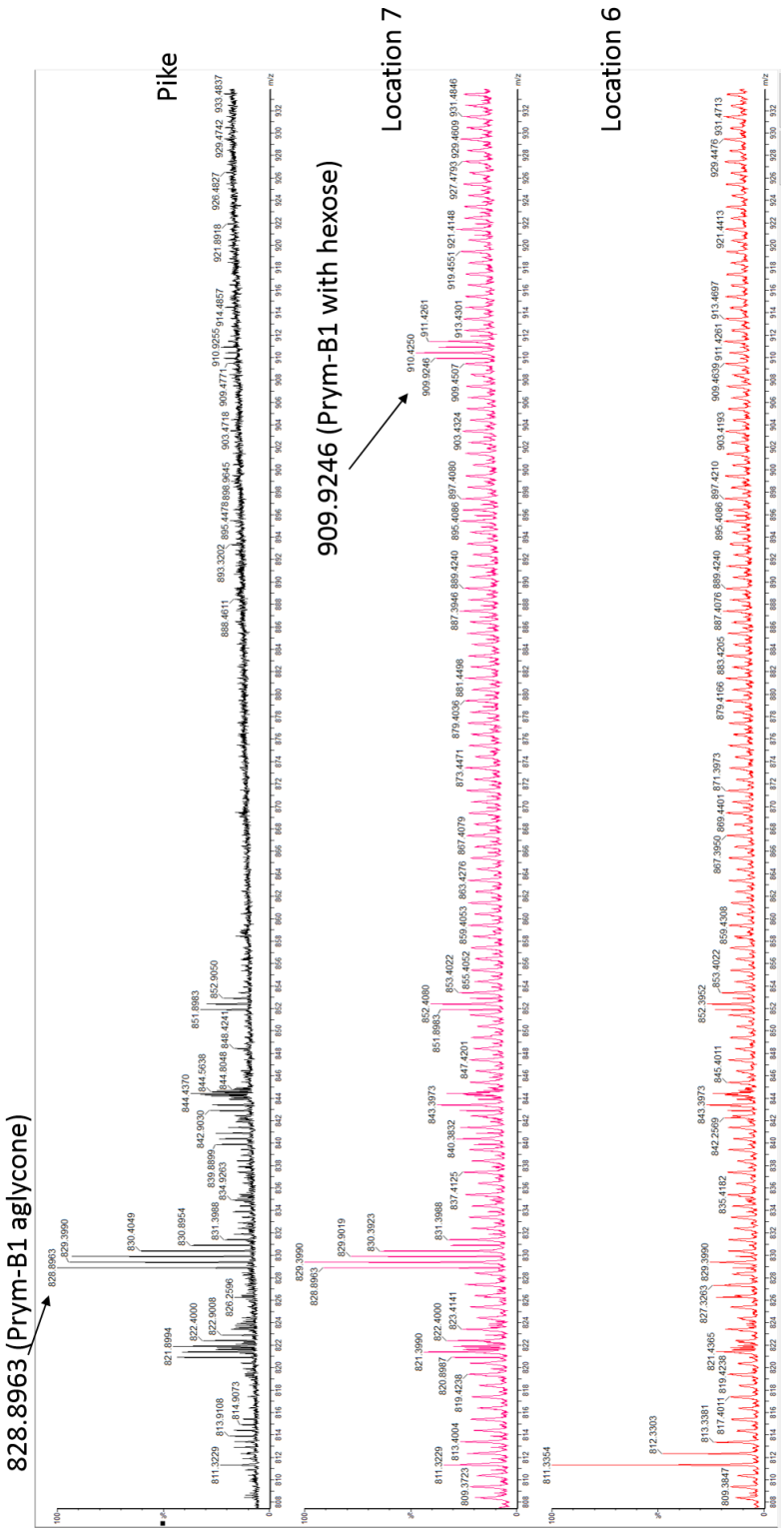


Figure 22 - Diagnostic ions found for prymnesin-B1 show typical isotope patterns seen for the chlorinated prymnesin compounds [13]. The detection of 828.8963 represents the aglycone form of prymnesin-B1 and has < 0.2 ppm error when compared to 828.8965 as reported by Rasmussen *et al* [1]. The detection of 909.9246 represents the glycosylated form of prymnesin-B1 and has 3.18 ppm error when compared to the calculated mass of prymnesin-B1 of 909.9211.

2.3.6 Monitoring of *P. parvum* by quantitative real-time PCR

To provide the local authorities (Environment Agency, Broads Authority) with a more sensitive monitoring method for *P. parvum* than optical microscopy, a quantitative real time PCR assay was developed. Using biweekly water samples taken from sampling locations 1 – 11 on Hickling Broad, seasonal populations dynamics were recorded for *P. parvum* specific ITS genes, using primers based on the work of Galluzzi *et al* and Zamor *et al* [23, 24].

Contradictory previous work on the Norfolk Broads suggests that *P. parvum* blooms have no obvious seasonality [4], but our work suggests that blooms occur primarily in late spring to summer, with blooms of the organism in July-August 2016 and June to September in 2017 (Figure 23). This may suggest a seasonality of HABs on the Norfolk Broads that correlates with temperature, light and other factors influenced by the seasons. Surprisingly though, these blooms did not reach the cell count levels observed in the Spring 2015 toxic bloom, where numbers reached more than 11,000,000 reads ml^{-1} at location 6 compared to ~1,950,000 reads ml^{-1} at the height of a non-toxic bloom in August 2016 (data not shown).

Blooms appear to rise and fall extremely fast, with blooms appearing to decline over 90% between 2-week sampling periods, which may suggest an unknown abiotic factor triggering bloom decline. Likewise, bloom progression appears to be rapid with periods between blooms (Jan 2016 – June 2016; Nov 2016 – May 2017) showing extremely low (but detectable) *P. parvum* abundance. This may suggest specific environmental factors are able to trigger rapid bloom progression, and future work correlating our qPCR data with abiotic factors such as pH, temperature, nutrient levels may help to understand what is causing *P. parvum* blooms in this area.

In agreement with the locality of fish deaths during the Spring 2015 bloom, *P. parvum* abundance seems to follow a trend to bloom to higher numbers at the north-west of Hickling Broad (Figure 16). Locations 5 and 6 appear to show consistently higher ITS reads during all blooms observed during this period, whilst location 9 appears to suffer less with *P. parvum* abundance (Figure 23). Other locations have sporadic data values, which may be improved by more regular sampling.

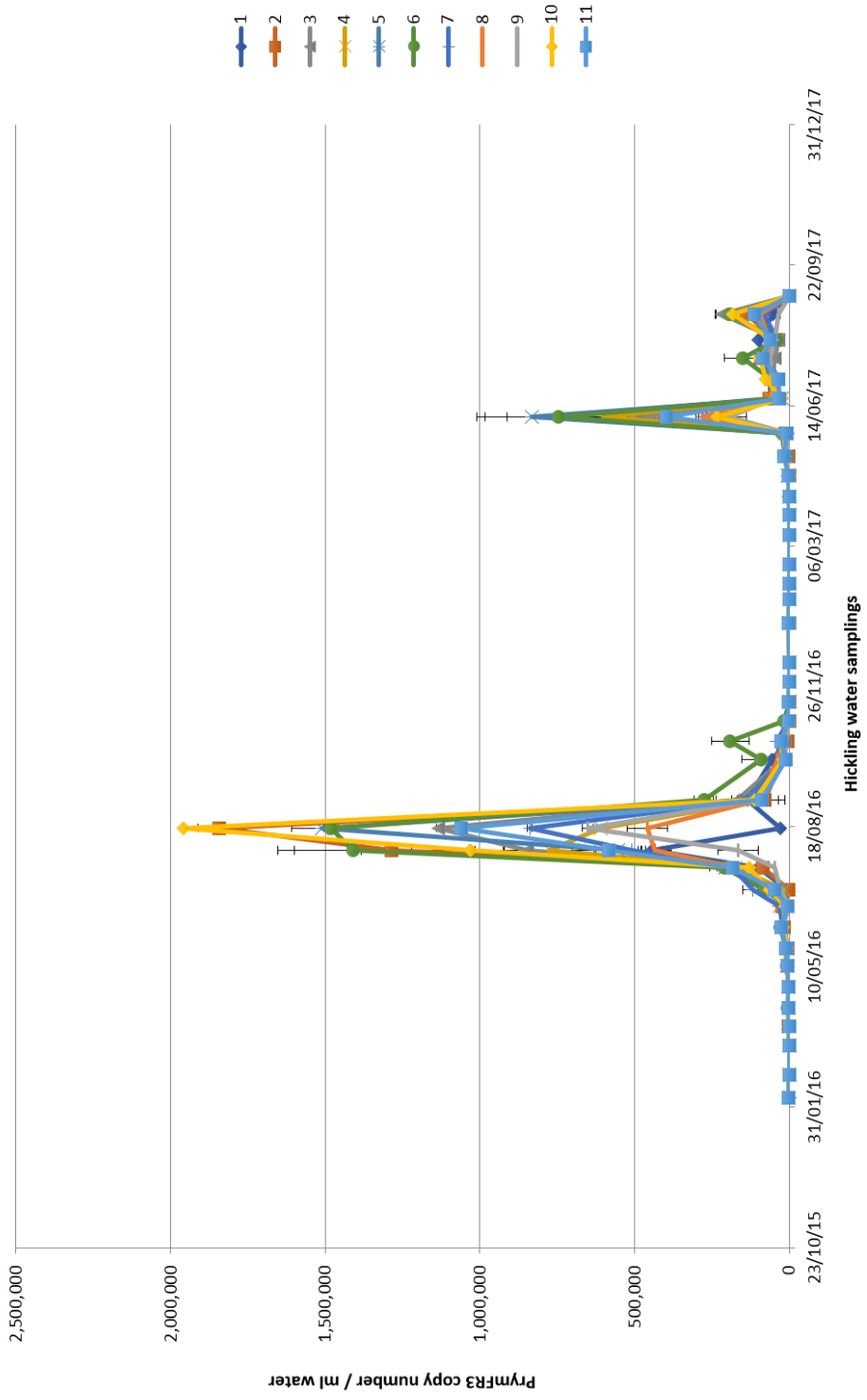


Figure 23 - Relative abundance of *P. parvum* on Hickling Broad across a 20-month period. Abundance is based on qRT-PCR ITS reads. Blooms of *P. parvum* can be observed throughout the summer months (June – September) in both 2016 and 2017. Data produced by Dr Jennifer Pratscher (University of East Anglia).

2.4 Discussion

This research was the result of an emergency circumstance in which a dense bloom of *P. parvum* resulted in the death of thousands of fish across the Upper Thurne area of the Norfolk Broads. More specifically, this toxic bloom threatened the native fish population on Hickling Broad, where it was estimated around 600,000 more fish were rescued by moving the fish to non-toxic waters. Previously, identification of blooms in the area relied on either anecdotal evidence or qualitative identification of *P. parvum* by optical microscopy. In this study, we showed using a combination of microscopy, genetic, and metabolite analysis that *P. parvum* was responsible for the toxic bloom. We showed that *P. parvum* dominated the microbial community during the bloom and that under non-bloom conditions, *P. parvum* represents a very small percentage of the microbial community. Moreover, we manage to detect the ladder-frame polyether prymnesins for the first time in natural water samples, and in the gill cells of a deceased pike. At the request of the local Environment Agency, we then developed a sensitive monitoring method for *P. parvum* based on quantitative real-time PCR, and use this to show *P. parvum* population dynamics over a 20-month period after the bloom occurred. Overall, these results confirm much literature doubt that the polyether prymnesin toxins are the relevant ichthytoxins and show that qPCR can be used as an effective monitoring tool for *P. parvum* in natural waters.

Currently, the primary method for analysing *P. parvum* blooms on the Norfolk Broads is optical microscopy, which has proven unsuccessful largely due to the insensitive nature of cell counting by such methods. To conclusively show that this bloom was caused by *P. parvum* we used a combination of optical microscopy and genetic analysis. Our optical microscopy qualitatively confirmed *P. parvum* was the culprit, where a large percentage of cells observed fitted morphological traits used to identify the species [26, 27]. Follow up genetic analysis based on 16S rRNA gene sequencing showed that *P. parvum* dominated the microbial community at the 11 sampling locations on Hickling Broad during the bloom (Figure 18). This dominance of the microbial ecosystem would agree with many descriptions of harmful algal blooms that state they are a rapid expansion of a phytoplankton population that typically involves one or a small number of species [28, 29]. As expected, the abundance of *P. parvum* in non-bloom conditions was only up to ~1-2%. Interestingly, however, the number of other species observed during bloom conditions is lower, showing the marked effect that a *P. parvum* bloom has on other microbes. This would agree with work by many that claims toxins of *P. parvum* are allelopathic [30, 31], acting to reduce growth or kill off competing microbes.

Despite the efforts of many, the main responsible ichthyotoxins in *P. parvum* blooms are still debated, with claims of galactoglycerolipids [7], fatty acid amides [8, 32], fatty acids [9] and others [6, 33]. A further type of toxins, the ladder-frame polyether prymnesins, have gathered the most attention due to their potent ichthyotoxicity with LD_{50} values towards fish in the low nm range [10, 11]. Their significance has however been questioned, with researchers failing to report detection of the toxins in lab cultures from 1999 when they were first isolated to 2013 when Manning and La Claire eventually reported a detailed extraction procedure for the toxins [13]. This was largely believed to be due to the small amounts produced by *P. parvum*, with Igarashi *et al* isolating only 25 mg from over 400 L of liquid culture [10]. More recently, Rasmussen *et al* reported on the discovery of a chemodiversity of the prymnesins, showing that different *Prymnesium* isolates produce slightly structurally modified versions of the toxins [1]. The authors go on to speculate that it is this, and not their low levels that has prevented the detection of these compounds previously. To shed light on this gap in knowledge we first showed that our lab strain of *P. parvum* (CCAP 946/6) produces the originally isolated forms of polyether prymnesin-1 and -2, with many of the fragments and adducts found by MS in agreement with those previously reported by Manning and La Claire (Table 1) [13]. Using a similar extraction and analysis method, we failed to detect prymnesin-1 and -2 in natural water samples, but instead were able to detect the recently reported prymnesin-B1 [1]. This would suggest that the *P. parvum* strain from the Norfolk Broads does not produce prymnesin-1 and -2, but instead produces the structurally similar prymnesin-B1. This significant finding would agree with the argument by Rasmussen *et al* that the previous lack of detection of these toxins may be due to a yet incomplete catalogue of prymnesin m/z values. On top of this, and unsurprisingly, toxins were not detected in water samples under non-bloom conditions. Furthermore, we detected prymnesin-B1 in the gill cells of a deceased pike (*Esox lucius*). This would agree with the speculated biological target of the prymnesins as being the gill cell membranes [25].

At the request of the local Environment Agency, we next sought to develop a sensitive monitoring method to follow *P. parvum* abundance in natural waters as a tool to predicting when a toxic bloom is likely to occur. The last decade has seen a boom in the use of molecular methods for this application, with particular success coming from the use of quantitative real-time PCR which offers unrivalled sensitivity and specificity [20]. Some examples of success stories include the monitoring of toxin-producing dinoflagellates such as the *Karenia* genus [21, 34], but Galluzzi *et al* and Zamor *et al* have also shown their value for *P. parvum* blooms

[23, 24]. In this study, we combined primers from both Galluzzi *et al* and Zamor *et al* to produce a sensitive qRT-PCR assay to monitor abundance of *P. parvum* ITS genes. Taking water samples for analysis approximately every 2 weeks for a period of 20 months, we have shown that blooms of *P. parvum* vary in their timing and intensity, but appear to show some seasonality, with blooms occurring late Spring and throughout the Summer of 2016 and 2017. This would contradict the past findings that blooms of *P. parvum* appear to show no seasonality across the Norfolk Broads [4]; a finding that was based on optical microscopy alone. Blooms appear to start and reach their decline within 3-4 weeks, which somewhat reflects growth properties seen in laboratory culture. However, interestingly, the sharp declines in cell numbers seen for the *P. parvum* blooms in this dataset may implicate algal viruses or other protozoan grazers in bloom demise [35, 36].

In conclusion, our results show that a toxic algal bloom on Hickling Broad in 2015 which caused the deaths of thousands of fish was caused by *P. parvum*. We show that the bloom was dominated by *P. parvum* alone, and that under non-bloom conditions *P. parvum* represents a small percentage of the microbial population. We detected the ladder-frame polyether prymnesin-B1 for the first time in natural water samples and in the gill cells of a deceased pike (*Exos lucius*). Finally, in an effort to combat future blooms of this organism across the Norfolk Broads, we developed a sensitive qRT-PCR assay and demonstrate how it can be used to monitor *P. parvum* population dynamics across a 20-month period.

2.5 Materials and Methods

2.5.1 *P. parvum* culture conditions

Prymnesium parvum (strain 946/6) was obtained from the Culture Collected of Algae and Protozoa (CCAP) (http://www.ccap.ac.uk/) and maintained in the recommended f/2 – Si media. Stock cultures were grown at 22 ^{o}C on a 14:10 light cycle with a light intensity of 100 µmol.m^{-2}.s^{-1} as previously described [37]. Under these conditions, 2-3 million cells ml^{-1} could be achieved by late log phase and 150 ml of culture at these cell densities were used for toxin extractions.

2.5.2 Study site – fish kill March 2015

A fish kill was first reported 13th March 2015 when members of the public contacted the environment agency concerned that several fish appeared to be in distress or dying on Hickling Broad, Norfolk. Hickling Broad represents an area of 122 hectares with average water depths just above 1 metre. To study the fish kill and likely algal bloom, sampling sites were set up to cover the majority of the Broad (Table 2). Water samples were collected from all sampling locations for toxin extractions on 17th March 2015 during a harmful bloom and then again during non-bloom conditions in September 2016.

Sample Point	Latitude	Longitude
1	52°43'46.91"N	1°35'57.05"E
2	52°43'56.20"N	1°35'39.72"E
3	52°44'4.66"N	1°35'23.75"E
4	52°44'19.12"N	1°34'39.49"E
5	52°44'29.04"N	1°34'17.54"E
6	52°44'47.62"N	1°34'12.55"E
7	52°44'48.02"N	1°34'21.73"E
8	52°44'33.22"N	1°34'27.76"E
9	52°44'19.97"N	1°33'57.14"E
10	52°44'22.73"N	1°34'12.43"E
11	52°44'31.74"N	1°34'2.94"E

Table 2 – Coordinates of sampling locations set up on Hickling Broad, Norfolk.

2.5.3 Optical microscopy

Water samples taken from Hickling Broad during the toxic bloom were visually analysed for the presence of *P. parvum* cells by pipetting 5 µl of fresh water onto glass slides and analysing on a light microscope (DM6000, Leica). Cells of *P. parvum* were identified through observation of characteristic traits (cell size, ellipsoidal shape, golden yellow colour, and the presence of 2 flagella and 1 haptonema).

2.5.4 Genetic analysis of microbial community

Water samples (100 ml) from each of sampling locations 1 – 11 (10 cm depth) were taken on 17th March 2015 during a harmful bloom, and again on 6th September 2016 under non-bloom conditions. Water samples were centrifuged at 3000 x *g* to pellet algal and bacterial cells and the cell pellets were stored at -78 ^{o}C prior to analysis. For genetic analysis, cell pellets were then provided to collaborating environmental scientist (Dr Jennifer Pratscher, UEA) for cDNA and DNA extraction. In brief, nucleic acids were extracted from the water biomass pellets using an SDS-based protocol. The cell pellets were added to Eppendorf tubes containing Lysing Matrix E beads (MP Biomedicals) and mixed with 1 ml of SDS extraction buffer. Cells were lysed in a FastPrep bead-beating system for 45 s at 6 m·s^{-1} and supernatants were subsequently extracted using phenol : chloroform : isoamyl alcohol (25:24:1) and chloroform : isoamyl alcohol (24:1). Nucleic acids were then precipitated with polyethylene glycol (PEG) 6000 solution (20%) and dissolved in 100 µl of nuclease-free water. 16S rRNA gene amplicon sequencing (515F/806r primers) for microbial community analysis was conducted by MR DNA (Texas, USA) on Illumina HiSeq on the DNA samples from all sampling locations.

2.5.5 Detection of Prymnesin-1 & -2 in *P. parvum* 946/6

For the extraction of crude prymnesins from inside algal cells, the general procedure outlined by Manning *et al* [13] was followed.

Cell pellets were lysed using cold acetone (4 ^{o}C) before being spun at 16,873 x g in a standard benchtop centrifuge. Chlorophyll containing supernatants were then removed. Acetone washes were repeated 2 more times. The resulting pellet was then extracted with MeOH and centrifuged as previously described. This was repeated a total of 3 times and the supernatants pooled after each round. The MeOH extraction procedure was then repeated with *n*-propanol and the supernatants were added to the MeOH supernatants. The resulting mixture was then dried *under vacuo* on a rotary evaporator and the remaining powder was re-suspended in water (~1-2ml). An equal volume of EtOAc was then added, and the solutions were mixed

using a vortex. The mixture was allowed to settle for 5 minutes on a bench, at which point the organic layer was removed. This defatting process was repeated a total of 3 times. The remaining aqueous layer was then loaded on a C18 cartridge (1 mg). The cartridge was washed at a flowrate of 1 ml min^{-1} with water for 5 column volumes, followed by washes with 20% *n*-propanol and 45% *n*-propanol each at 5 column volumes. The prymnesins were finally eluted using 80% *n*-propanol. The cartridge eluent was dried *under vacuo* and crude toxin extracts stored at -20 ^{o}C. Detection was performed via LC-MS using a QTOF MS system with a nano-spray ion source. Extracts were dissolved in an aqueous solution containing 0.1% TFA (mobile phase A). The analytes were resolved using mobile phase B (99.9% acetonitrile, 0.1% TFA) with an initial linear gradient from 2% to 50% acetonitrile over 10 minutes, followed by a second linear gradient from 50% to 90% acetonitrile over 30 minutes, on a C18 column.

Chromatograms were queried for double charged molecular ions corresponding to the backbone of prymnesin-1 and -2 (m/z 919.88), and where found corresponding peaks were manually analysed for other diagnostic ions previously reported for prymnesin-1 and -2 by Manning *et al* [13]. All data analysis was performed using MassLynx™ software (Waters).

2.5.6 Detection of Prymnesin-B1 in natural water samples and fish gills

For the extraction of prymnesins from natural water samples, 100 ml of water taken from locations 6 and 7 was centrifuged at 3000 x *g* to pellet cells and debris. The resulting supernatant was then passed through a 0.45 µM filter to remove excess algal/bacterial cells and debris before being passed through a 1 mg C18 cartridge at a flow rate of 1 ml min^{-1} to load the organic prymnesins onto the column. The cartridge was then washed with 5 column volumes of water to remove salts, before elution of organic toxins was achieved with 10 column volumes of 80% *n*-propanol.

The column eluent was then dried under vacuum using a rotary evaporator and re-suspended in a 1:1 mixture of H_2O:EtOAc (total volume 4 ml). The aqueous layer was subsequently defatted a total of 3 times by removal and re-addition of EtOAc, ensuring sufficient mixing of organic and aqueous layers each time. After removal of the last EtOAc, the remaining aqueous layer was dried under vacuum on a rotary evaporator. Dried extracts were re-suspended in 200 µl 0.1% TFA (mobile phase A) before being subject to LC-MS as outlined in section 2.5.5.

For the extraction of toxins from gill cells, a deceased pike (*Esox Lucius*) was taken from the site of the fish kill (close to sampling location 11) and the gill cells were removed by dissection

with a sharp scalpel (Figure 24). Gill cells (approximately 4 g) were ground in liquid nitrogen before being subject to an identical prymnesin extraction protocol as used for algal cells (Section 2.5.5).

Figure 24 – A dead pike collected from the toxic *P. parvum* bloom (A), and a close-up of its gills (B).

2.5.7 Monitoring *P. parvum* by quantitative real-time PCR

Water samples were collected from sampling locations 1-11 on Hickling Broad approximately once every 2 weeks for a 20-month period. Samples (100 ml) were obtained from a 10 cm depth and pelleted by centrifugation (3000 x *g*). Nucleic acids were extracted following the protocol in section 2.5.4. The abundance of *P. parvum* specific ITS genes was quantified by qPCR using previously published primers PrymF [23] and PrymR-3 [24] (Table 3). All assays were performed in a StepOnePlus Real-Time PCR system (Applied Biosystems). Respective standards were used, and controls were ran with water instead of extracted nucleic acid samples.

Primer name	Sequence 5' – 3'	Reference
PrymF	TGTCTGCCGTGGACTTAGTGCT	Galluzzi *et al* [23]
PrymR-3	ATGGCACAACGACTTGGT	Zamor *et al* [24]

Table 3 – List of primers used in this study.

2.6 References

1. Rasmussen, S. A.; Meier, S.; Andersen, N. G.; Blossom, H. E.; Duus, J. Ø.; Nielsen, K. F.; Hansen, P. J.; Larsen, T. O., Chemodiversity of ladder-frame prymnesin polyethers in *Prymnesium parvum*. *J. Nat. Prod.* **2016,** 79, (9), 2250-2256.
2. Merico, A.; Tyrrell, T.; Brown, C. W.; Groom, S. B.; Miller, P. I., Analysis of satellite imagery for *Emiliania huxleyi* blooms in the Bering Sea before 1997. *Geophys. Res. Lett.* **2003,** 30, (6).
3. Granéli, E.; Edvardsen, B.; Roelke, D. L.; Hagström, J. A., The ecophysiology and bloom dynamics of *Prymnesium* spp. *Harmful Algae* **2012,** 14, 260-270.
4. Holdway, P. A.; Watson, R. A.; Moss, B., Aspects of the ecology of *Prymnesium parvum* (Haptophyta) and water chemistry in the Norfolk Broads, England. *Freshwat. Biol.* **1978,** 8, (4), 295-311.
5. Authority, B., Broads Plan 2017. **2017**.
6. Paster, Z. K., Pharmacology and mode of action of prymnesin. *Academic Press*: NY, **1973**, 241-263.
7. Kozakai, H.; Oshima, Y.; Yasumoto, T., Isolation and structural elucidation of hemolysin from the phytoflagellate *Prymnesium parvum*. *Agric. Biol. Chem.* **1982,** 46, (1), 233-236.
8. Bertin, M. J.; Zimba, P. V.; Beauchesne, K. R.; Huncik, K. M.; Moeller, P. D. R., The contribution of fatty acid amides to *Prymnesium parvum* Carter toxicity. *Harmful Algae* **2012,** 20, 117-125.
9. Henrikson, J. C.; Gharfeh, M. S.; Easton, A. C.; Easton, J. D.; Glenn, K. L.; Shadfan, M.; Mooberry, S. L.; Hambright, K. D.; Cichewicz, R. H., Reassessing the ichthyotoxin profile of cultured *Prymnesium parvum* (golden algae) and comparing it to samples collected from recent freshwater bloom and fish kill events in North America. *Toxicon* **2010,** 55, (7), 1396-404.
10. Igarashi, T.; Satake, M.; Yasumoto, T., Structures and partial stereochemical assignments for prymnesin-1 and prymnesin-2: potent hemolytic and ichthyotoxic glycosides isolated from the red tide alga *Prymnesium parvum*. *J. Am. Chem. Soc.* **1999,** 121, (37), 8499-8511.
11. Igarashi, T.; Satake, M.; Yasumoto, T., Prymnesin-2: A potent ichthyotoxic and hemolytic glycoside isolated from the red tide alga *Prymnesium parvum*. *J. Am. Chem. Soc.* **1996,** 118, (2), 479-480.

12. Blossom, H. E.; Andersen, N. G.; Rasmussen, S. A.; Hansen, P. J., Stability of the intra- and extracellular toxins of *Prymnesium parvum* using a microalgal bioassay. *Harmful Algae* **2014,** 32, 11-21.

13. Manning, S. R.; La Claire Ii, J. W., Isolation of polyketides from *Prymnesium parvum* (Haptophyta) and their detection by liquid chromatography/mass spectrometry metabolic fingerprint analysis. *Anal. Biochem.* **2013,** 442, (2), 189-195.

14. Klindworth, A.; Mann, A. J.; Huang, S.; Wichels, A.; Quast, C.; Waldmann, J.; Teeling, H.; Glöckner, F. O., Diversity and activity of marine bacterioplankton during a diatom bloom in the North Sea assessed by total RNA and pyrotag sequencing. *Mar. Genomics* **2014,** 18, Part B, 185-192.

15. Landa, M.; Blain, S.; Christaki, U.; Monchy, S.; Obernosterer, I., Shifts in bacterial community composition associated with increased carbon cycling in a mosaic of phytoplankton blooms. *ISME J* **2016,** 10, (1), 39-50.

16. Smith, D. C.; Steward, G. F.; Long, R. A.; Azam, F., Bacterial mediation of carbon fluxes during a diatom bloom in a mesocosm. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **1995,** 42, (1), 75-97.

17. Larsson, U.; Hagström, A., Phytoplankton exudate release as an energy source for the growth of pelagic bacteria. *Mar. Biol.* **1979,** 52, (3), 199-206.

18. Bidle, K. D.; Azam, F., Accelerated dissolution of diatom silica by marine bacterial assemblages. *Nature* **1999,** 397, (6719), 508-512.

19. Smriga, S.; Fernandez, V. I.; Mitchell, J. G.; Stocker, R., Chemotaxis toward phytoplankton drives organic matter partitioning among marine bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **2016,** 113, (6), 1576-1581.

20. Antonella, P.; Luca, G., The quantitative real-time PCR applications in the monitoring of marine harmful algal bloom (HAB) species. *Environ. Sci. Pollut. Res.* **2013,** 20, (10), 6851-6862.

21. Casper, E. T.; Paul, J. H.; Smith, M. C.; Gray, M., Detection and quantification of the red tide dinoflagellate *Karenia brevis* by real-time nucleic acid sequence-based amplification. *Appl. Environ. Microbiol.* **2004,** 70, (8), 4727-4732.

22. Kamikawa, R.; Asai, J.; Miyahara, T.; Murata, K.; Oyama, K.; Yoshimatsu, S.; Yoshida, T.; Sako, Y., Application of a real-time PCR assay to a comprehensive method of monitoring harmful algae. *Microbes Environ.* **2006,** 21, (3), 163-173.

23. Galluzzi, L.; Bertozzini, E.; Penna, A.; Perini, F.; Pigalarga, A.; Graneli, E.; Magnani, M., Detection and quantification of *Prymnesium parvum* (Haptophyceae) by real-time PCR. *Lett. Appl. Microbiol.* **2008,** 46, (2), 261-266.

24. Zamor, R. M.; Glenn, K. L.; Hambright, K. D., Incorporating molecular tools into routine HAB monitoring programs: Using qPCR to track invasive Prymnesium. *Harmful Algae* **2012,** 15, (Supplement C), 1-7.

25. Manning, S. R.; La Claire, J. W., Prymnesins: toxic metabolites of the golden alga, *Prymnesium parvum* Carter (Haptophyta). *Mar. Drugs* **2010,** 8, (3), 678-704.

26. Manton, I.; Leedale, G. F., Observations on the fine structure of *Prymnesium parvum* carter. *Archiv. Mikrobiol.* **1963,** 45, (3), 285-303.

27. Chang, F. H.; Ryan, K. G., *Prymnesium calathiferum* sp. nov. (Prymnesiophyceae), a new species isolated from Northland, New Zealand. *Phycologia* **1985,** 24, (2), 191-198.

28. Hudnell, H. K., The state of U.S. freshwater harmful algal blooms assessments, policy and legislation. *Toxicon* **2010,** 55, (5), 1024-1034.

29. Smayda, T. J., What is a bloom? A commentary. *Limnol. Oceanogr.* **1997,** 42, (5), 1132-1136.

30. Smayda, T. J., Complexity in the eutrophication–harmful algal bloom relationship, with comment on the importance of grazing. *Harmful Algae* **2008,** 8, (1), 140-151.

31. Granéli, E.; Johansson, N., Increase in the production of allelopathic substances by *Prymnesium parvum* cells grown under N- or P-deficient conditions. *Harmful Algae* **2003,** 2, (2), 135-145.

32. Bertin, M. J.; Zimba, P. V.; Beauchesne, K. R.; Huncik, K. M.; Moeller, P. D. R., Identification of toxic fatty acid amides isolated from the harmful alga *Prymnesium parvum* carter. *Harmful Algae* **2012,** 20, 111-116.

33. Dafni, Z.; Ulitzer, S.; Shilo, M., Influence of light and phosphate on toxin production and growth of *Prymnesium parvum*. *J. Gen. Microbiol.* **1972,** 70, (2), 199-207.

34. Yuan, J.; Mi, T.; Zhen, Y.; Yu, Z., Development of a rapid detection and quantification method of *Karenia mikimotoi* by real-time quantitative PCR. *Harmful Algae* **2012,** 17, (Supplement C), 83-91.

35. Wilhelm, S. W.; Coy, S. R.; Gann, E. R.; Moniruzzaman, M.; Stough, J. M. A., Standing on the shoulders of giant viruses: five lessons learned about large viruses infecting small eukaryotes and the opportunities they create. *PLoS Pathog.* **2016,** 12, (8), e1005752.

36. Tillmann, U., Interactions between planktonic microalgae and protozoan grazers. *J. Eukaryot. Microbiol.* **2004,** 51, (2), 156-68.

37. Wagstaff, B.; Vladu, I.; Barclay, J.; Schroeder, D.; Malin, G.; Field, R., Isolation and characterization of a double stranded DNA megavirus infecting the toxin-producing haptophyte *Prymnesium parvum*. *Viruses* **2017,** 9, (3), 40.

3 Isolation and Characterization of a Double Stranded DNA Megavirus Infecting the Toxin-Producing Haptophyte *Prymnesium parvum*

Wagstaff, B.; Vladu, I.; Barclay, J.; Schroeder, D.; Malin, G.; Field, R., Isolation and Characterization of a Double Stranded DNA Megavirus Infecting the Toxin-Producing Haptophyte *Prymnesium parvum*. *Viruses* **2017,** 9, (3), 40.

3.1 Abstract

Prymnesium parvum is a toxin-producing haptophyte that causes harmful algal blooms globally, leading to large-scale fish kills that have severe ecological and economic implications. For the model haptophyte, *Emiliania huxleyi*, it has been shown that large dsDNA viruses play an important role in regulating blooms and therefore biogeochemical cycling, but much less work has been done looking at viruses that infect *P. parvum*, or the role that these viruses may play in regulating harmful algal blooms. In this study, we report the isolation and characterization of a lytic nucleo-cytoplasmic large DNA virus (NCLDV) collected from the site of a harmful *P. parvum* bloom. In subsequent experiments, this virus was shown to infect cultures of *Prymnesium* sp. and showed phylogenetic similarity to the extended Megaviridae family of algal viruses.

3.2 Introduction

The last two decades have seen a boom in the study of marine viruses and the role that they play in regulating both bacterial and unicellular eukaryote bloom dynamics [1,2]. Although phages and the bacteria that they infect have been studied for many years, the more recently discovered *Acanthamoeba polyphaga mimivirus* (APMV) and its *Megaviridae* relatives have brought about a new age in photosynthetic protist virology. It has recently been shown that dsDNA viruses infecting algae do not form monophyletic lineages [3], with divergence occurring even within the host division. A good example of this evolutionary divergence can be found in viruses that infect the coccolithophore *Emiliania huxleyi* (EhV) [4,5] and the prymnesiophyte *Phaeocystis globosa* (PgV) [6], which along with other algal viruses have been proposed to form an extended branch of the *Megaviridae* [7]. It is widely accepted that these viruses not only play a crucial role in ecosystem dynamics [8,9], but also contribute significantly to biogeochemical cycles [10,11]. A lesser studied impact, however, lies in the role that such viruses may play in the termination of toxic eukaryotic algal blooms. Lytic viruses that infect the toxic raphidophyte *Heterosigma akashiwo* have been extensively studied [12-19] but, because of the elusive nature of *H. akashwio* toxicity to fish, none of these studies sought to investigate the role of viral infection on levels of algal toxicity.

The toxin-producing haptophyte *Prymnesium parvum* forms dense blooms in marine, brackish and inland waters, devastating fish populations through the release of natural product toxins [20,21]. The haptophytes are a diverse division of microalgae that include the bloom-forming *Emiliania huxleyi* and *Phaeocystis globosa*, both of which play crucial roles in oceanic carbon and sulfur cycles [22,23]. Virus infection of these organisms has been studied in some detail, with the genome of the dsDNA *Phaeocystis globosa* virus (PgV-16T) being recently described [3]. From a metabolomics perspective, *Phaeocystis pouchetti* lysis by a strain-specific virus has been shown to cause substantial release of dimethyl sulphide and its major precursor dimethylsulphoniopropionate [24], an action that is believed to contribute significantly to the global sulfur cycle. Although much effort has gone into studying the relationship between *E. huxleyi* and its infecting viruses, viruses infecting toxin-producing algal species within the haptophyte family are much less well studied. These include the euryhaline species *Prymnesium* spp. and *Chrysochromulina* spp., whose blooms can often result in severe economic damage through loss of fish stocks [25,26]. Viruses that infect the non-toxic *P. kappa* have recently been described, but to date no viruses have been isolated and characterized that

infect the toxin-producing *P. parvum* species, even though Schwierzke *et al* have previously suggested a role for viruses in regulating natural *P. parvum* populations [27].

Hickling Broad is part of a network of broads that make up The Norfolk and Suffolk Broads, Britain's largest protected wetland. As a hotspot for angling and boating activities, the tourism industry from the Broads contributes ~£500 million per year to the local economy [28]. However, since the mid-1960s, this area has been plagued by almost annual blooms of *P. parvum,* often leading to the losses of thousands of fish crucial for these activities [29]. In this study, we first show that natural *P. parvum* populations appear to be infected by viruses. Following this we show that communities of algal viruses exist in the Norfolk Broads and we subsequently isolated a novel lytic virus of *P. parvum* 946/6, *Prymnesium parvum* DNA virus BW1 (henceforth referred to as PpDNAV), from the site of a recent harmful bloom event of this species in Norfolk, England. We show that the virus has a typical narrow host range; using morphological characterisation and phylogenetics, we also show that the virus lies in the recently described clade of algal megaviruses.

3.3 Results

3.3.1 Optical microscopy of natural *P. parvum* populations

Water samples taken during a bloom of *P. parvum* in Spring 2015 were analysed visually for virally infected cells. Many cells observed showed signs of distress (non-motile, erratic flapping of flagella against the cell), and in most cases (Figure 25), virus-like particles could be observed in the cytoplasm of the algal cells. When left for a 4-hour period, these cells were seen to have undergone a cell lysis event typical of that seen for lytic algal viruses, with membrane blebbing and loss of motility preceding cell lysis.

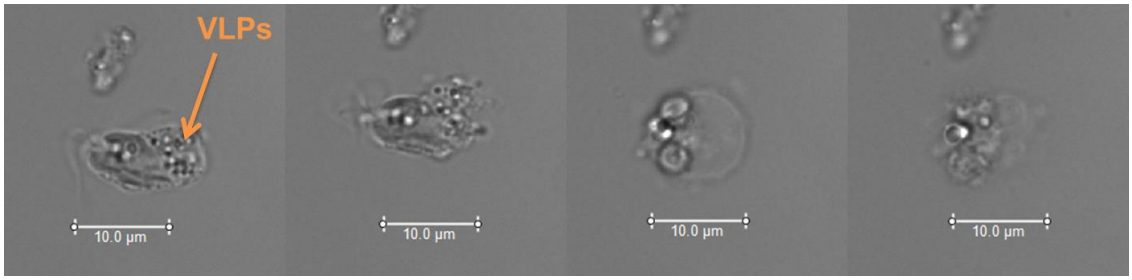


Figure 25 - A natural *P. parvum* cell from samples taken from Hickling Broad undergoing viral cell lysis. Images were taken throughout a 4 hour period and show (from left to right) a non-motile cell undergoing membrane blebbing before bursting and releasing the intracellular contents. Scale bar = 10 µm.

3.3.2 Isolation of whole viral communities

Large volume (10 L) water samples taken from Hickling Broad, Norfolk, were subject to a viral isolation technique in which rounds of filtration were used to select for viruses from a mixture of organisms. Epifluorescence microscopy confirmed removal of most bacteria and algae (Figure 26), and concentration of small VLP size particles that strained positively with SYBR Green.

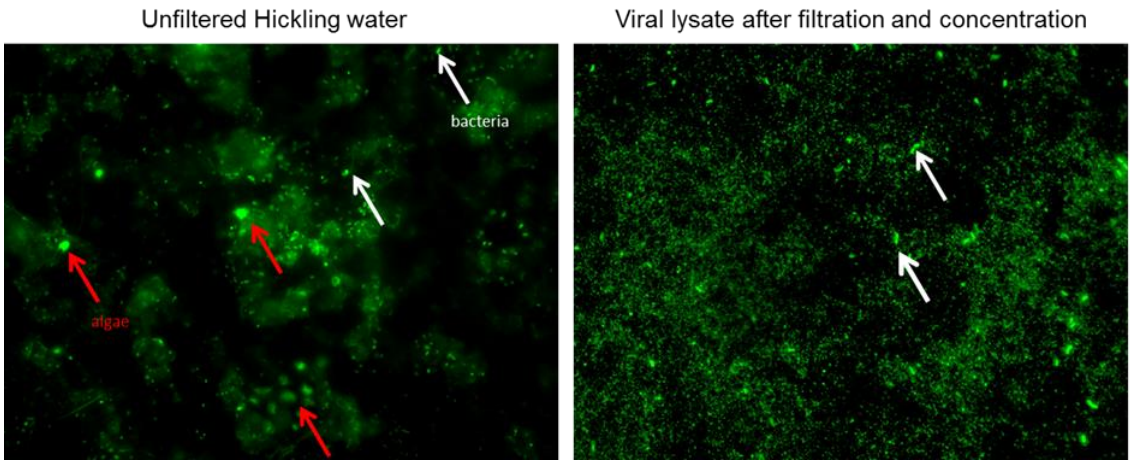


Figure 26 - Left; unfiltered water from Hickling Broad after treatment with SYBR Green and analysed using epifluorescence microscopy. Red arrows indicate algae, white arrows indicate bacteria. Right; white arrows indicate bacteria; smaller green dots are likely individual virus particles.

Filtered and concentrated viral communities were then subject to qualitative TEM analysis to investigate, by morphological analysis, the type of viruses present. As expected, the community contained a mixture of viruses including a mixture of bacteriophages (as observed by tail structures), ovoid-like viruses and icosahedral non-tailed viruses within the size range expected for algal viruses (Figure 27).

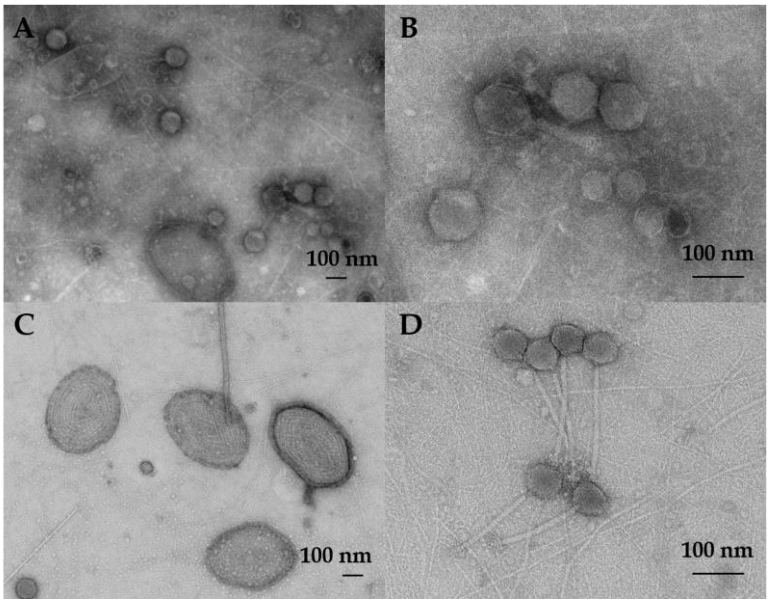


Figure 27 - A mixed virus population prepared from samples taken from Hickling Broad. A, B; icosahedral viruses, bacteriophages, larger ovoid shaped viruses. C; large ovoid-like VLPs. D; smaller bacteriophages.

An additional way to predict classification of algal viruses is based on their genome size. Therefore, viral communities were subject to pulsed field gel electrophoresis analysis (Figure 28), so that genome sizes could be predicted. Virus-plugs were prepared by embedding concentrated viral lysate in agarose. These plugs were then subjected to enzymatic digestion to release nucleic acids before being loaded on gels. A range of distinct bands could be observed, representing a variety of genome sizes. Under these conditions several distinct bands could be seen within the expected range for the *Phycodnaviridae* family of algal viruses (160 – 560 kbp) [30].

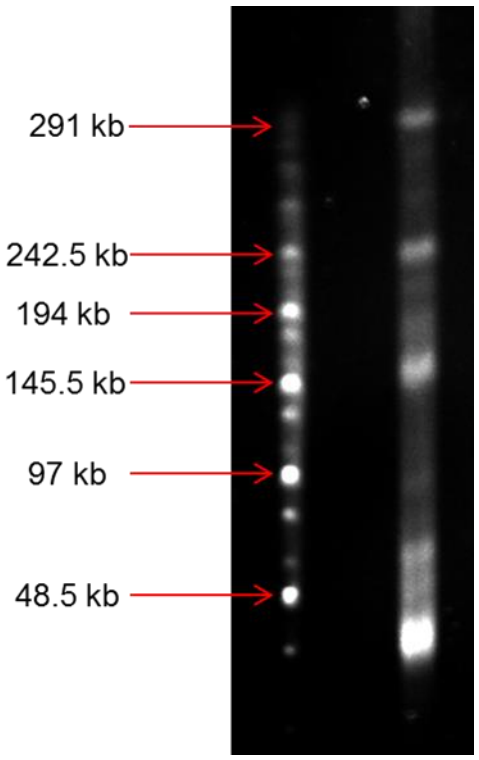


Figure 28 - Pulsed-field gel electrophoresis (PFGE) analysis of a mixed virus community isolated from Hickling Broad. Several distinct bands are visible from 48.5 kb to 291 kb. Ladder is Lambda ladder (Bio-Rad).

3.3.3 Isolation of lytic virus particles

PpDNAV isolation was conducted from water samples collected at Hickling Broad, Norfolk, England. Among four water samples from which viral lysates were prepared, lysis of *P. parvum* 946/6 occurred with three samples (Figure 29). Transmission electron micrographs of the viral lysates showed that icosahedral VLPs were present in all three samples, but samples 1 and 2 also contained significant levels of phage-like particles; we suspect that these were a result of infection with the low levels of bacteria that were present in the non-axenic *P. parvum* 946/6

cultures. To avoid further downstream separation of viruses, we chose to continue working with sample 4 only (sourced at Hickling Broad—52°44’19.12” N, Long—1°34’39.49” E), which appeared by TEM to be free of phages. After a triplicate dilution series, the resulting monoclonal viral lysate still lysed the host cells and TEM of thin-sectioned cells confirmed the presence of VLPs (Figure 30A,B); thereby fulfilling Koch’s postulates.

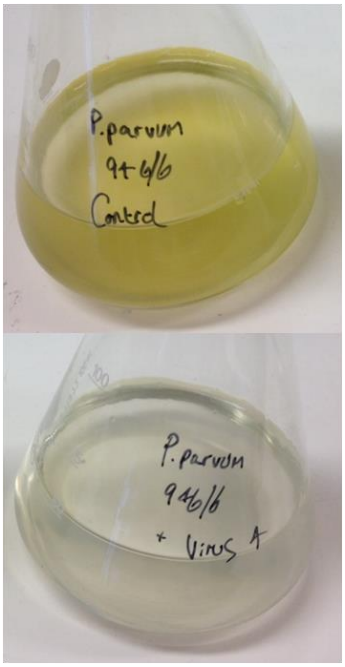


Figure 29 – Viral cell lysis as seen by culture clearing. (Top) control culture; (Bottom) ‘Cleared’ culture 96 h post viral infection.

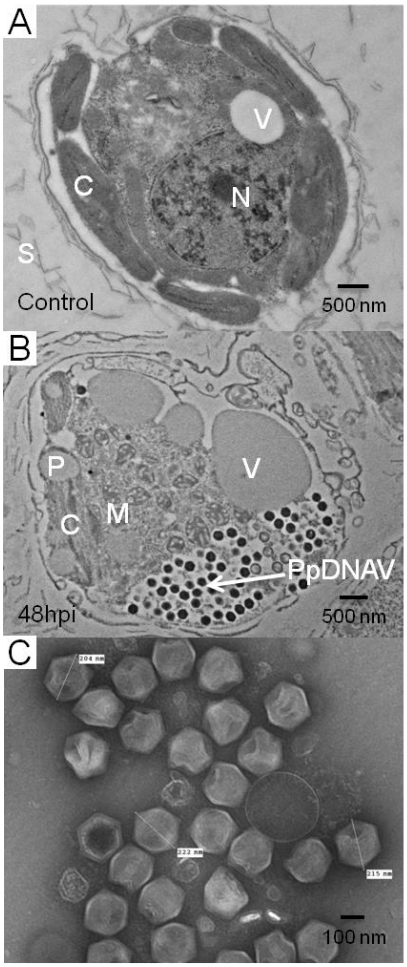


Figure 30 – Viral infection of *P. parvum* by PpDNAV observed by TEM. (A) Thin-sections of healthy *P. parvum* 946/6 cells; (B) Thin-sections of *P. parvum* 946/6 48 h post infection. (C) Free *Prymnesium parvum* DNA virus (PpDNAV) particles in culture supernatant 72 h post infection. C:chloroplast; V:contractile vacuole; N:nucleus; S:scales; M:mitochondria, P:pyrenoid.

3.3.4 Virus morphology, host range, and infectious properties

Transmission electron microscopy of isolated and intracellular (thin sectioned) viruses revealed an icosahedral capsid with an average diameter of 221 nm (*n* = 71) (Figure 30). Although no external viral lipid membrane was evident, some viral particles showed an internal white 'halo' between the capsid and the DNA of PpDNAV, suggestive of the virus having an internal membrane. The presence of a viral factory or viroplasm [31] in the host cytoplasm, and in some cases an imperfect vertex or a tail-like structure were also observed (Figure 31 and Figure 32). As seen in Figure 30B, establishment of a viral factory in the host cytoplasm also results in a loss of the nuclear envelope and therefore loss of the nucleus. This was observed in the majority of infected cells examined at 48 h p.i.

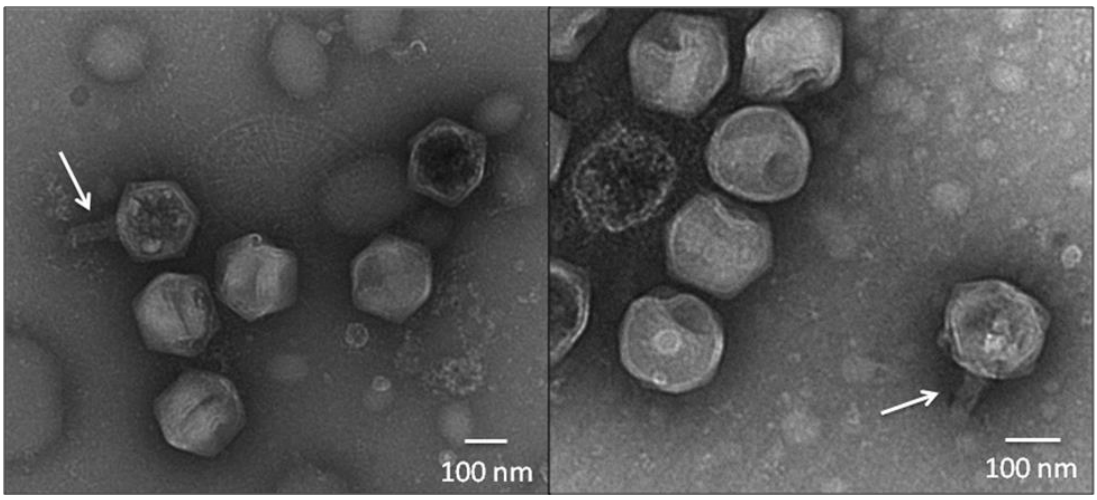


Figure 31 - Electron micrographs of negatively stained PpDNAV particles. A small number of particles appear to have a single imperfect vertex, likely representing a stargate structure (white arrow).

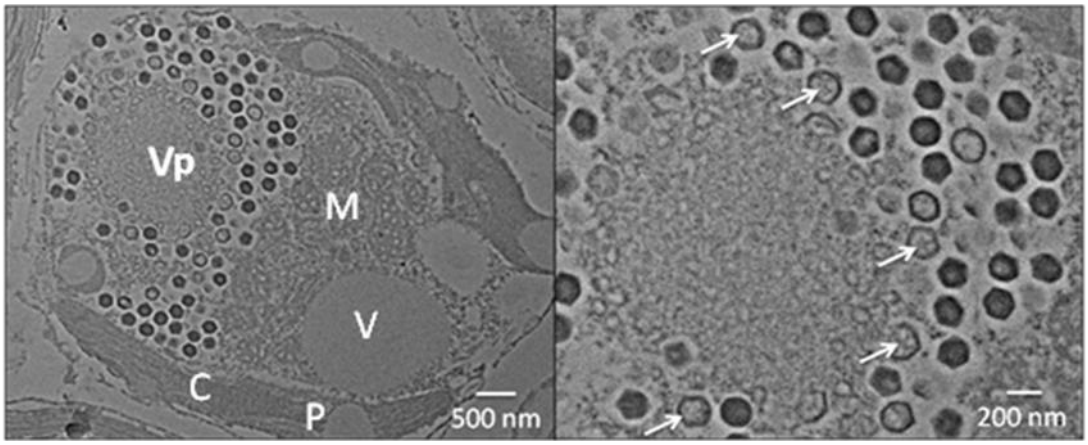


Figure 32 - Electron micrographs of infected *P. parvum* 946/6 48 h p.i. Left – Whole cell visually infected with many PpDNAV particles surrounding a viroplasm. Right – Magnified from left. Both mature and immature particles can be observed. 'Empty' particles are denoted by a white arrow. Vp:viroplasm; C:chloroplast; V:contractile vacuole; M:mitochondria, P:pyrenoid.

Fifteen different strains of *Prymnesium* were screened for sensitivity to PpDNAV (Table 4). PpDNAV was found to be sensitive to chloroform, whereby the chloroform-treated virus no longer caused lysis of *P. parvum* 946/6 (Figure 33). This supports the notion of a viral membrane in this system.

Figure 33 - Chloroform sensitivity assay. Left to right – *P. parvum* 946/6 + f/2 media; *P. parvum* 946/6 + PpDNAV; *P. parvum* 946/6 + Chloroform treated PpDNAV; *P. parvum* 946/6 + Chloroform treated f/2 media.

The lytic cycle of the virus was explored to determine both the incubation period and eclipse period (Figure 34). At 48 h p.i., the cells had clearly lost mobility and sedimented at the base of the culture flask. Re-suspension of the cells by shaking led to similar cell counts as seen at 24 h p.i., as determined by Coulter counting. The time before symptoms of viral infection, the incubation period, was therefore judged to be 24 h. The eclipse period reflects the time between infection and appearance of mature virus particles within the host; as new mature virions were first observed 48 h p.i., the eclipse period was judged to be 24–48 h. At 72 h p.i., the onset of cell lysis had occurred. PpDNAV appeared to lyse >95% of host cells by 120 h p.i., whilst uninfected control cultures continued to grow over the full course of the experiment.

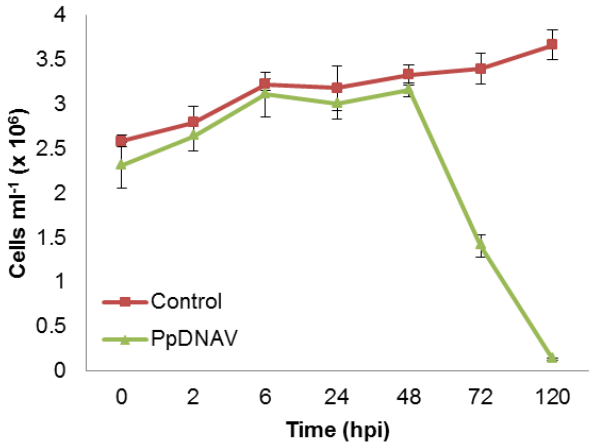


Figure 34 - PpDNAV infection cycle propagated on *P. parvum* 946/6. Graph shows the average number of algal cells in control cultures (squares) and PpDNAV infected cultures (circles). Error bars represent the standard error for triplicate cultures.

3.3.5 Genome sequencing and phylogenetic analysis

Predicted proteins from the initial genome assembly included the MCP1 protein (KY509047) and DNA polB (KY509048) which were used for phylogenetic analysis. The 525 aa sequence for MCP1 was found to have 91% sequence similarity to the major capsid protein 1 of *Phaeocystis globosa* virus (YP_008052475.1) and 84% similarity to MCP1 of Organic Lake Phycodnavirus 2 (ADX06358.1) with *E*-values of 0.0 in each case. This alignment allowed construction of a phylogenetic tree (Figure 35) that shows clustering with other megaviruses, including PgV-16T.

For DNA polB (KY509048), the 1281 aa sequence displayed 77% sequence similarity to DNA polB of PgV-16T (YP_008052566.1) and 64% similarity to DNA polB of Organic Lake phycodnavirus 2 (ADX06483.1). The resulting phylogenetic tree (Figure 36) shows a similar clustering of PpDNAV to the algal *Megaviridae* family, but also illustrates an obvious divergence between algal viruses that fall within the *Megaviridae* family and those that do not; with EhV-86 and *Heterosigma akashiwo* virus (HaV)-1 rightfully placed outside of the *Megaviridae* clade.

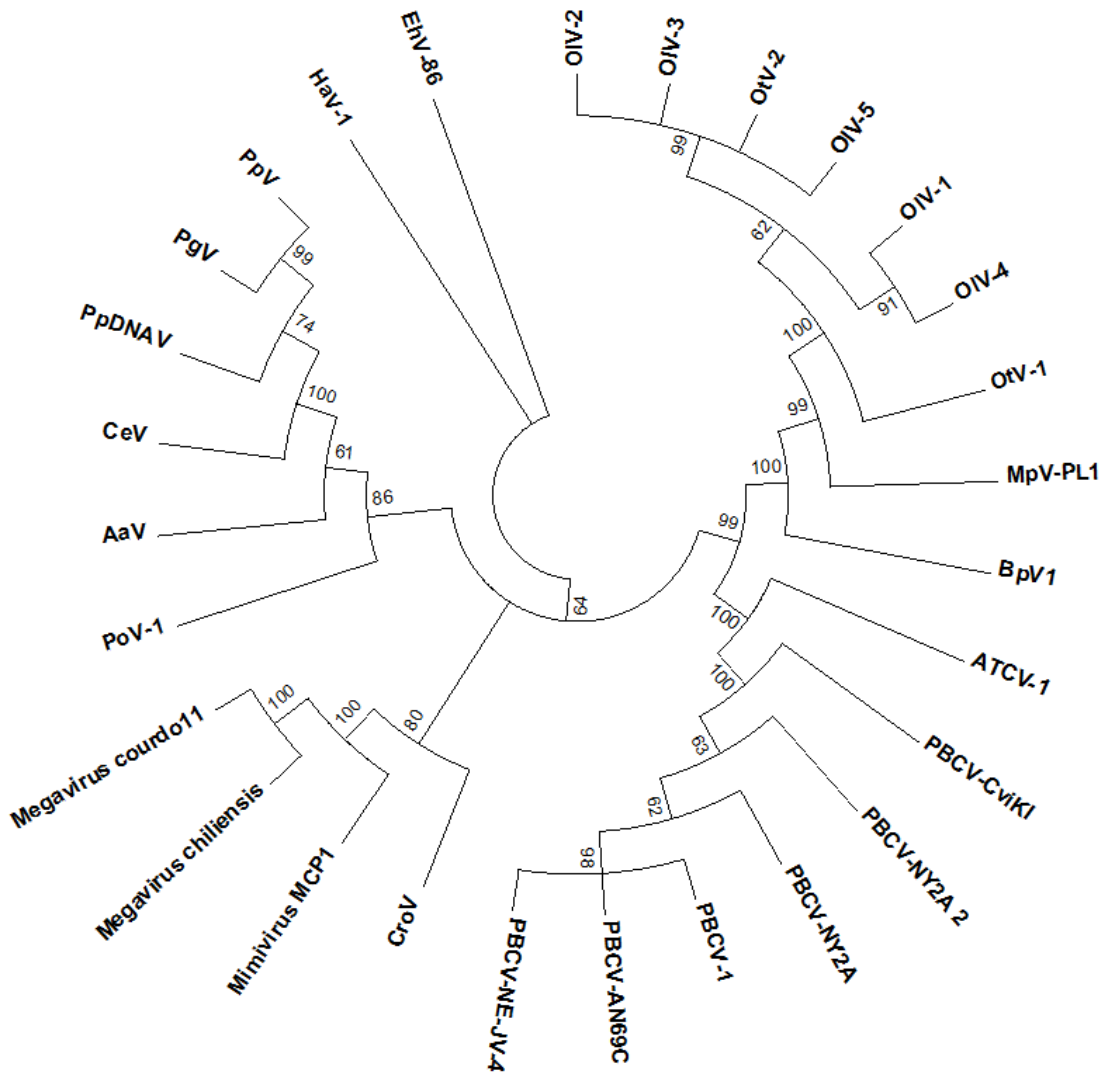


Figure 35 - Phylogenetic clustering of PpDNAV with other large algal *Megaviridae*. Alignment was performed using the default settings of MAFFT [32], and the Neighbour-joining method (midpoint-rooted) [33] was used to construct a tree from 28 viral MCP sequences using MEGA7 [34]. The final tree was based on 336 ungapped positions, 500 resampling permutations, and was collapsed for bootstrap values <50. The tree shows that PpDNAV clusters with the algal-infecting *Megaviridae*, separate to the *Phycodnaviridae*.

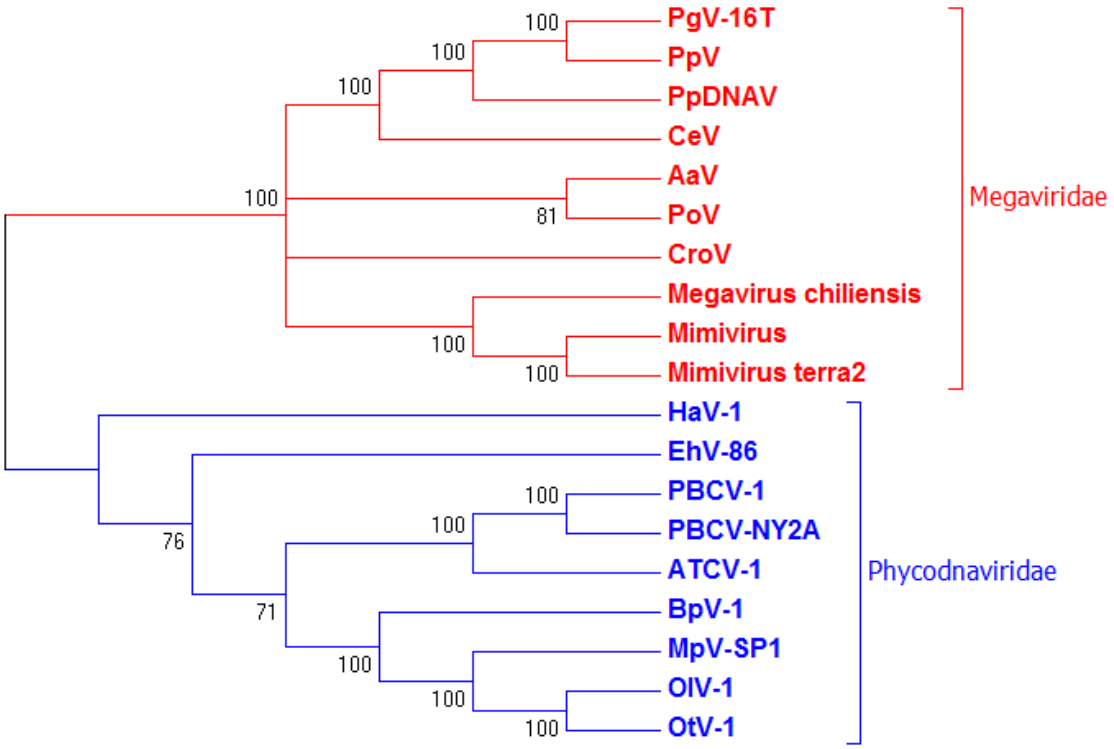


Figure 36 - Phylogenetic clustering of PpDNAV with other large algal *Megaviridae*. Alignment was performed using the default settings of multiple sequence alignment software version 7 (MAFFT) [32], and the neighbour-joining method (midpoint-rooted) [33] was used to construct a tree from 19 viral DNA Polymerase Beta (polB) sequences using Molecular Evolutionary Genetics Analysis version 7.0 (MEGA7) [34]. The final tree was based on 630 ungapped positions, 500 resampling permutations, and was collapsed for bootstrap values <50. The tree shows that PpDNAV clusters with the well-defined clade of *Megaviridae* and the algal-infecting *Megaviridae* (red), and not with the *Phycodnaviridae* (blue).

A preliminary genome assembly of PpDNAV-BW1 was found to be 523 kbp, although this is expected to change slightly when the assembly is complete. Currently, the genome is spread over 56 contigs and its overall A+T content is 72.4%. GeneMark™ identified 332 putative protein-coding sequences from the genome, which were then subject to BLASTp analysis (cutoff *E*-value < 10^{-5}) for manual annotation. Of the 332 genes, 273 (82%) shared highest homology to sequences from other viruses; 22 (7%) bacterial, 19 (6%) eukaryotic, 2 (1%) archaeal and 16 (5%) have no match to sequences in NCBI using these search settings (Figure 37).

Of the 273 sequences with highest homology to viral sequences, these are dominated by sequences found in the recently described *Phaeocystis globosa* Virus (PgV-16T) [3] (218, 80%).

Other viruses infecting the haptophyte *Phaeocystis* spp. share sequence homology, with 6 (2%) found with homology to PgV-12T, 2 (1%) found with homology to sequences from PgV-14T and 2 (1%) found with homology to sequences from PpV. Unsurprisingly, sequences are also found that share highest homology to viruses infecting other haptophytes; 18 (7%) share highest homology to sequences from CeV and 21 (8%) share highest homology to OLPV which are speculated to infect haptophytes [3]. 2 (1%) sequences share highest homology to another megavirus infecting an amoeba, CroV, and 4 (1%) have sequences found in other viruses (Figure 37).

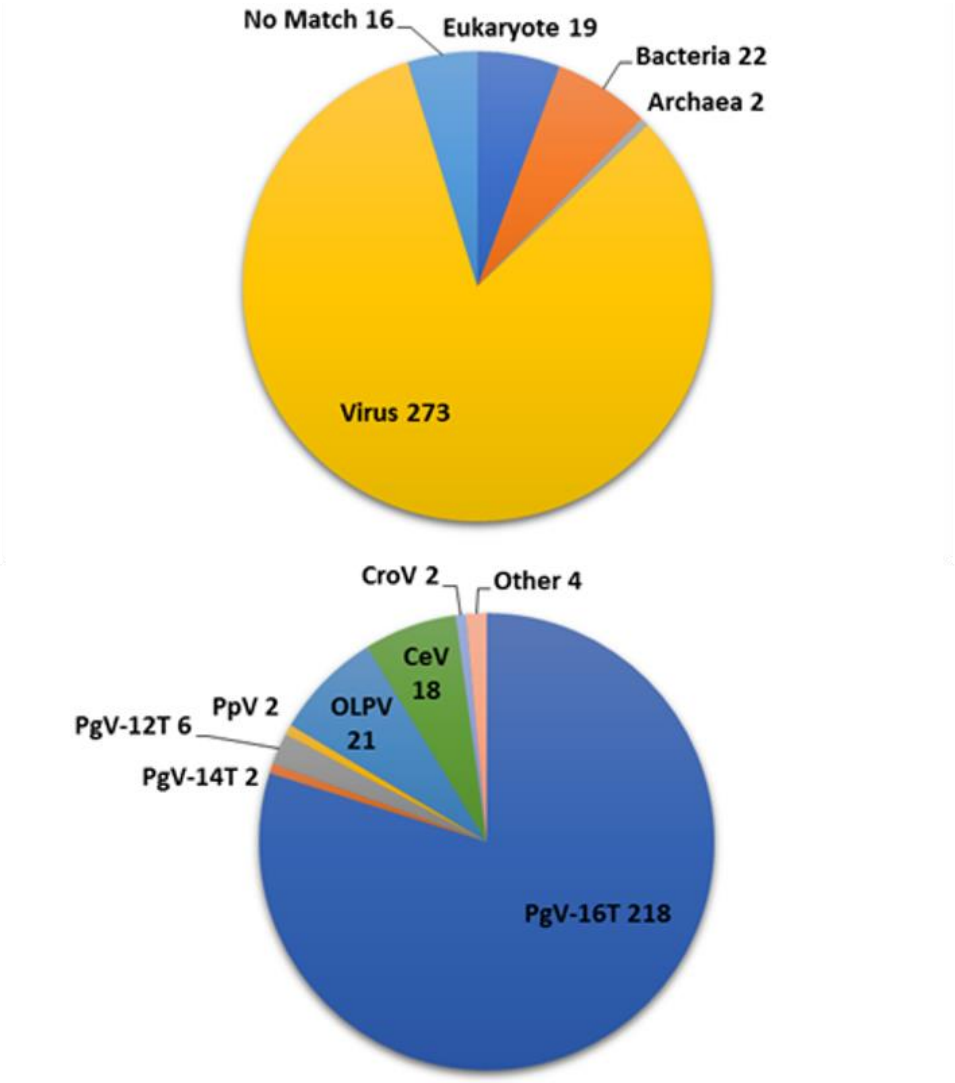


Figure 37 – Database similarity of the translated genome of PpDNAV-BW1. (Top) >75% of predicted proteins match those of viral sequences found in the NCBI NR protein database. (Bottom) Of the viral hits, >75% are most similar to proteins encoded in the genome of the recently described *Phaeocystis globosa* virus (PgV-16T).

3.4 Discussion

Haptophytes are abundant in marine waters but can also thrive in brackish inland waters. Whilst a significant amount of work has been done on the marine dwelling coccolithophore *Emiliania huxleyi* and its associated viruses, little work has looked at the toxin-producing members of the haptophytes. In the present study, we first showed qualitative evidence that natural populations of *P. parvum* are infected by lytic viruses in the Norfolk Broads. We next examined the viral communities in Hickling Broad and showed morphological and genomic evidence for the presence of algal viruses. Subsequently, we then isolated and characterized a novel megavirus, PpDNAV, from brackish inland waters where harmful blooms of *Prymnesium parvum* frequently occur [29]. We showed that this lytic virus was able to infect *P. parvum* 946/6, later expanded to five out of 15 *Prymnesium* strains tested. Morphological and phylogenetic analysis of two core dsDNA virus conserved genes suggests that this virus belongs to the extended *Megaviridae* family of algal-infecting viruses.

Transmission electron microscopy of negatively stained virus particles from a lysed culture supernatant revealed icosahedral capsids with an average diameter of 221 nm. Many particles appeared to have one imperfect vertex, with some showing material protruding from what appeared to be a stargate [31] (Figure 31). These likely represent particles in an advanced stage of packing or unpacking genetic material [35], and suggested early on that PpDNAV lies in the extended *Megaviridae* branch of algal viruses. Thin sections of infected *P. parvum* 946/6 cells showed evidence for a viroplasm as the site of replication, where empty capsids could be seen closer to the centre of the viroplasm (Figure 32). This further supported the inclusion of PpDNAV in the extended *Megaviridae* family [36]. The infectivity of PpDNAV is chloroform sensitive (Figure 33), and the lack of an obvious external lipid membrane observed by TEM may suggest that internal membrane/s are present; although chloroform sensitivity cannot always be used to confirm lipid membrane presence [37].

New mature virions were first observed by electron microscopy at 48 h p.i., so the eclipse period of the virus in infected algal cells was estimated to be 24–48 h. At 48 h, the algal cell count as recorded by Coulter counting was still the same as at 24 h, but a complete sedimentation of cells had occurred, suggesting a loss in motility and a likely shutdown of important cellular processes. By 72 h, a rapid decline in cell abundance could be observed, showing that the loss of motility precedes the host lysis event, as is seen for some other

flagellated algae [18]. In its natural environment, this may lead to accumulation of viral particles at the sediment surface rather than dispersed in the water column.

The algal host species specificity of PpDNAV was assessed against *P. parvum* 946/6, which had been kept in 7–8 PSU f/2 medium for two years, and 14 other *Prymnesium* strains which had been maintained in a full strength seawater medium. Initially, PpDNAV only infected *P. parvum* 946/6, but after ~6 months of sub-culturing of the other 14 strains in 7–8 PSU f/2 medium, the host range broadened to five out of the 15 strains. We speculate that the change in salinity contributed to the change in sensitivity to PpDNAV; recent work by Nedbalová *et al* [38] suggests that a change in membrane lipid composition in different salinities may account for this situation. This somewhat less restricted host range is similar to that found for *Haptolina ericina* virus (HeV RF02), *Prymnesium kappa* virus (PkV RF01) and *Prymnesium kappa* virus (PkV RF02) [39]. Taken together, this suggested that PpDNAV was a member of the algal *Megavirus* family [40].

Phylogenetic analysis using sequences for MCP1 and DNA polB of PpDNAV confirmed morphological findings, showing that PpDNAV clusters amongst the algal viruses belonging to the *Megaviridae* family, such as PgV-16T [3,6], *Chrysochromulina ericina* virus (CeV) [41], *Pyramimonas orientalis* virus (PoV) [42] and the recently reassigned *Aureococcus anophagefferens* virus (AaV) [43]. With the exception of *Emliania huxleyi* virus (EhV-86), which appears to branch independently, the close clustering of viruses infecting haptophytes, as well as the chlorella viruses clustering together, supports the notion that viruses co-evolve with their hosts [6,43,44].

The incomplete genome assembly of PpDNAV displays clear similarities to the genome recently described for *Phaeocystis globosa* virus (PgV-16T) [3]. The initial assembly was found to be 523 Kbp with an A+T% of 72.4%; the genome of PgV-16T is 459,984-bp with an A+T% of 68%. The homology of genes found in the genome shares a similar pattern to that found for PgV-16T. 6% of PpDNAV genes share highest homology to eukaryotic genes whilst 6% of PgV-16T genes share highest homology to eukaryotic genes; 7% of PpDNAV genes are most homologous to bacterial genes compared to 8% of PgV-16T genes and ~1% of genes have highest similarity to archael genes in both viruses. Whilst PgV-16T has 44% of genes with no matches and PpDNAV only 5% with no match, this can be explained by the significant number of genes from PpDNAV that are now also found in PgV-16T since it's genome has been added to the NCBI database; 80% of genes with highest sequence similarity to viral genes are most similar to PgV-16T genes.

The remaining 20% of viral genes are most similar to genes from other viruses infecting haptophytes, as is seen in PgV-16T.

Of the algal viruses compared in this study, only HaV-1 is known to infect a toxin producing host [12-15]. However, the toxic metabolites responsible for bloom toxicity are not established in *Heterosigma akashiwo*, making studies of viral impact on toxicity difficult. On the other hand, reports of toxic *P. parvum* metabolites are numerous and include fatty acids [45], glycerolipids [46] and very large ladder-frame polyether toxins, known collectively as the prymnesins [21,47–49]. Reports of cases of toxic and non-toxic blooms of *Prymnesium* and other harmful algal species [29] has led to speculation that an ecological trigger exists for toxicity. While efforts have been made to associate nutrients, pH and other conditions to bloom toxicity [21], the identity of the full spectrum of toxicity-causing agents remains to be establshed; there may conceivably be a role for viral infection in *Prymnesium* cell lysis and hence toxin release. We now have the opportunity to use this alga–virus system in clearing up some of these unanswered questions. Further studies into the effect of viral infection and host algal cell lysis on toxic bloom events need to be explored in order to fully understand the underlying mechanisms behind production and release of toxins from *Prymnesium*. In addition, as further sequences of algal viruses become available, new opportunities will open up for accurate monitoring of viral population fluctuations with respect their host. Furthermore, the increase in characterized viruses will provide more information when analysing metagenomic data sets such as those generated by the *Tara* Oceans expedition [50,51]. Hence the discovery and characterization of PpDNAV in this study will aid this burgeoning field of scientific endeavour.

3.5 Materials and Methods

3.5.1 *Prymnesium parvum* culture conditions

For choice of host cell, *P. parvum* 946/6 was obtained from the Culture Collection of Algae and Protozoa (CCAP—www.ccap.ac.uk). The additional 14 strains used for host range screening were obtained from the Marine Biological Association Culture Collection (https://www.mba.ac.uk/culture-collection/). Batch cultures were maintained at 22 °C on a 14:10 light cycle at 100 µmol·photons·m^{-2}·s^{-1}. Cultures were grown in f/2–Si medium at a salinity of 7–8 practical salinity unit (PSU). Under these conditions, cell densities of ~3 × 10^6 cells·mL^{-1} could be achieved after 12–16 days of growth.

3.5.2 Optical microscopy of natural *P. parvum* populations

Small volume (100 ml) water samples taken during a toxic bloom of *P. parvum* were visually analysed for the presence of *P. parvum* cells by pipetting 5 µl of fresh broads water onto glass slides and analyzing on an optical microscope (DM6000, Leica). Cells of *P. parvum* were identified by looking for characteristic traits (cell size, ellipsoidal shape, 2 flagella and 1 haptonema).

3.5.3 Isolation and analysis of whole viral communities

Large volume (10 L) water samples were taken from Hickling Broad, Norfolk, on 25th May 2015 shortly following a harmful bloom of *P. parvum*. This was subject to filtration through 1L 0.22 µm disposable filtration units (Nalgene™, ThermoFisher Sci, United Kingdom) to remove bacteria and algae before the filtrate was concentrated to 50 ml by tangential flow filtration against a 100 kDa mw-cutoff filter (Vivaflow™, Sartorius, United Kingdom).

The resulting viral preparation was then stained with SYBR Green I according the method outlined by Patel *et al* [52] and stained particles were visualized using a Leica DM6000 microscope with fitted filters for blue excitation and green emission. Viral communities pre- and post filtration and concentration were visualized to confirm removal of bacteria and algae. Transmission electron microscopy (TEM) analysis was performed on whole viral communities as described in section 3.5.5.

Viral agarose plugs were prepared and pulsed field gel electrophoresis (PFGE) was carried out according to the methods outlined by Johannessen *et al* [39]. In brief, viral preparations were pelleted by ultracentrifugation (150,000× *g*) before being re-suspended in SM buffer (0.1 M

NaCl, 8 mM $MgSO_4 \bullet 7H_2O$, 50 mM Tris-HCl, 0.005% Glycerin). Molten agarose (1.5%) was then added to a 1:1 ratio and the lysate was drawn up a 1 ml syringe and allowed to solidify. Small plugs (1 cm) were cut away and incubated overnight in a lysis buffer (1 mg/ml proteinase K, 1% SDS, 250 mM EDTA, pH 8). Plugs were then loaded onto agarose gels (1%) with molecular weight markers (CHEF lambda ladder - BioRad). Gels were ran in 0.5 x TBE buffer (89 mM Tris-NaOH, 89 mM boric acid, 2 mM EDTA, pH 8) at 6V for 18 hours at 14 ^{o}c. Pulses were set up to separate genomes up to 600 kbp (1-8 s separating genomes from 5-200 kbp, 20-40 s separating genomes from 50-600 kbp).

3.5.4 Isolation of lytic virus particles

PpDNAV was isolated from surface water samples taken at various locations on Hickling Broad, Norfolk, England on 9 February 2016. In brief, 4 × 100 mL water samples from various locations around the Broad were centrifuged at 3000× *g* and the supernatant subsequently filtered through 0.45 µm pore-size filters (Sartorius AG, Goettingen, Germany). The resulting solutions were then concentrated 100- to 200-fold using 100 kDa mw cut off spin filters (Amicon Ultra 15, Merck Millipore, Watford, United Kingdom) to give 0.5 to 1 mL of viral concentrate, which was stored at 4 °C in the dark until use. Small volumes (0.2 mL) of concentrate from each location were added to 1.8 mL of exponentially growing cultures of *P. parvum* 946/6. Blank culture medium was used as a control. Cultures were visually inspected for signs of cell lysis (culture clearing) after 7–10 days where the control cultures continued to grow. Culture clearing was then followed up by Transmission Electron Microscopy (TEM) analysis of the culture lysates. Clonal populations of PpDNAV were obtained by taking the supernatant of a lysed culture, and exhaustively diluting with media. These diluted samples (0.2 mL) were added to 1.8 mL of an exponentially growing culture of *P. parvum* 946/6. The highest dilution that still produced cell lysis after seven days was taken through to the next round. This was repeated at least three times and resulted in a population of PpDNAV free of morphologically different viruses, as judged by TEM.

3.5.5 Transmission electron microscopy

For TEM analysis of virus-like particles (VLPs), 2 mL of viral preparations were filtered through 0.45 µm filters and 10 µL of the filtrate was adsorbed onto a 400-mesh copper palladium grid with a carbon-coated pyroxylin support film before being negatively stained with 2% aqueous uranyl acetate [53]. The grids were viewed in a FEI Tecnai 20 transmission electron microscope (Eindhoven, The Netherlands) at 200 kV and digital TIFF images were taken with an AMT XR60B digital camera (Deben, Bury St Edmunds, UK).

For analysing intracellular VLPs, 1 mL of infected cultures of *P. parvum* 946/6 was taken at 24 and 48 h post-infection (p.i.). These were centrifuged at 3000× *g* to pellet algal cells and the supernatant was discarded. The pellet was washed twice with sterile medium. The pelleted cells were then resuspended in 2.5% (*v*/*v*) aqueous glutaraldehyde solution and left overnight. This suspension was then centrifuged at 3000× *g* to pellet the algal cells. Half the volume of the supernatant was then discarded and an equal volume of warm (60 °C) low gelling temperature agarose (Sigma Aldrich, Haverhill, United Kingdom) was added, before resuspension of the cells and placing on ice to solidify. The solidified samples were then put into 2.5% (*v*/*v*) glutaraldehyde with 0.05 M sodium cacodylate, pH 7.3 [54] and left overnight. Using a Leica EM TP machine (Leica Microsystems, Cambridge, United Kingdom), the samples were washed in 0.05 M sodium cacodylate and then post-fixed with 1% (*w*/*v*) OsO_4 in 0.05 M sodium cacodylate for 60 min at room temperature. After washing and dehydration with ethanol, the samples were gradually infiltrated with LR White resin (London Resin Company, London, United Kingdom) according to the manufacturer's instructions. After polymerization, the resulting material was sectioned with a diamond knife using a Leica EM UC6 ultramicrotome (Leica Microsystems). Ultrathin sections of approximately 90 nm were picked up on 200 mesh gold grids that had been coated in pyroxylin and carbon. The grids were then contrast-stained with 2% (*w*/*v*) uranyl acetate for 1 h and 1% (*w*/*v*) lead citrate for 1 min, washed in distilled water and air-dried. The grids were then viewed with a FEI Tecnai 20 transmission electron microscope (Eindhoven, The Netherlands) at 200 kV and digital TIFF images were produced.

3.5.6 Host specificity

Fifteen different strains of *Prymnesium* were tested in triplicate for signs of cell lysis by PpDNAV using the infection methodology described above. Cell lysis, as observed by culture clearing, was documented for five of the 15 strains tested (Table 4).

Genus/species	Strain code	Lysis with PpDNAV
Prymnesium parvum	946/6	+
Prymnesium parvum	94A	-
Prymnesium parvum	94C	+
Prymnesium parvum	579	-
Prymnesium patelliferum	527A	+
Prymnesium patelliferum	527C	+
Prymnesium patelliferum	527D	-
Prymnesium sp.	522	-
Prymnesium sp.	569	-
Prymnesium sp.	592	+
Prymnesium sp.	593	-
Prymnesium sp.	595	-
Prymnesium sp.	596	-
Prymnesium sp.	597	-
Prymnesium sp.	598	-

Table 4 - Host range of PpDNAV. + lysed culture, - culture not lysed.

3.5.7 Infection cycle

The lytic cycle of the PpDNAV was investigated by accurately recording algal cell abundance during an infection cycle. A late-log phase culture of *P. parvum* 946/6 was infected with PpDNAV (0.1% *v*/*v*) and triplicate aliquots (2 mL) were taken at various time points post infection (p.i.). These were diluted with 0.2 µm filtered seawater prior to counting using a Multisizer 3 Analyser (Beckman Coulter, High Wycombe, United Kingdom) fitted with a 100 µm aperture tube. The control culture continued to grow throughout the experiment, whilst the infected algal culture was lysed rapidly after 48 h.

3.5.8 Chloroform sensitivity

To test the virus sensitivity to chloroform, an adaptation of the method of Martínez Martínez *et al* was employed [55]. Briefly, 1 mL of 0.45 µm-filtered PpDNAV was added to an equivalent volume of chloroform and shaken vigorously for 5 min. The resulting mixture was then centrifuged at 4000× *g* in a benchtop centrifuge for 5 min to separate the organic and polar layers. The aqueous phase was transferred by pipetting to a clean microcentrifuge tube and

incubated at 37 °C for 1 h to remove residual chloroform. As a control, 1 mL of chloroform was added to 1 mL of f/2 medium. Chloroform-treated PpDNAV, chloroform-treated medium and untreated PpDNAV were added to *P. parvum* 946/6 as described above in the infectivity experiment protocol; signs of lysis, as judged by culture clearing, were recorded after one week.

3.5.9 Viral DNA extraction, sequencing, and phylogenetic analyses

For DNA extraction, 1 L of late log phase *P. parvum* 946/6 was infected with axenic PpDNAV (0.1% *v*/*v*). Lysis was allowed to occur over a period of five days, by which point almost all cells had been lysed. The culture was centrifuged at 6500× *g* to pellet cell debris, before being filtered through 0.22 µm filters to remove remaining cell debris or contaminating bacteria. The filtrate was incubated for 72 h with 100 µg/mL carbenicillin before being concentrated to 30 mL using 100 kDa mw cut-off spin filters. Ultracentrifugation at 150,000× *g* was used to pellet viral particles, and these were re-suspended in 2 mL of ρ = 1.4 CsCl and layered onto a CsCl gradient which was resolved at 150,000× *g* for 18 h. Fractions from ρ = 1.3 to ρ = 1.4 were pooled and DNA extracted using a PureLink Viral RNA/DNA Kit, according to the manufacturer's protocol.

An amount of 1 µg of purified viral DNA was then sent to The Earlham Institute, UK, for Illumina MiSeq sequencing (Illumina, Inc., San Diego, CA, USA) and assembly. A single paired-end (PE) library was prepared using a single lane of Illumina MiSeq to yield 250 bp paired end reads. The reads were then processed and trimmed using KAT (Kmer analysis toolkit) [56] and assembled using IDBA assembly tool [57]. The full set of assembled scaffolds was then aligned to NCBI nr/nt database using the diamond aligner [58] and sequences aligning to known Phycodnaviridae were filtered. This output was then analysed using the MEGAN metagenomic analysis package [59] (v5.11.3) to give a filtered assembly. The initial assembly was then analysed using GeneMarkS [60] which identified 332 protein-coding sequences. BLASTp analysis was then performed against the National Center for Biotechnology Information (NCBI) GenBank nonredundant (nr) protein sequence database [61] to identify major capsid protein and DNA Pol B candidates. Nucleic acid and amino acid sequences for the major capsid protein (MCP) and DNA Polymerase B (DNA polB) were submitted to Genbank with the accession codes KY509047 and KY509048, respectively.

Phylogenetic analysis was performed using the obtained sequences for MCP and DNA polB, as well as other related sequences from previously discovered algal viruses, identified using BLASTp. These sequences were aligned using the default settings of multiple sequence alignment software version 7 (MAFFT) [37], and trees were constructed from the neighbour-joining method [38] (midpoint-rooted) using Molecular Evolutionary Genetics Analysis version 7.0 (MEGA7) [39].

An initial genome assembly was annotated by manually blasting protein-coding sequences identified by GeneMarkS against the NCBI non-redundant protein database [61]. Where sequences had hits with *E*-value < 10^{-5}, the function and organism of the top hit was noted.

3.6 References

1. Thingstad, T.F. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol. Oceanogr.* **2000**, 45, (6), 1320–1328.
2. Fuhrman, J.A. Marine viruses and their biogeochemical and ecological effects. *Nature* **1999**, 399, (6736), 541–548.
3. Santini, S.; Jeudy, S.; Bartoli, J.; Poirot, O.; Lescot, M.; Abergel, C.; Barbe, V.; Wommack, K.E.; Noordeloos, A.A.; Brussaard, C.P.; *et al.* Genome of *Phaeocystis globosa* virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. *Proc. Natl. Acad. Sci. USA* **2013**, 110, (26), 10800–10805.
4. Wilson, W.H.; Tarran, G.A.; Schroeder, D.; Cox, M.; Oke, J.; Malin, G. Isolation of viruses responsible for the demise of an *Emiliania huxleyi* bloom in the English Channel. *J. Mar. Biol. Assoc. UK* **2002**, 82, (3), 369–377.
5. Schroeder, D.C.; Oke, J.; Malin, G.; Wilson, W.H. Coccolithovirus (Phycodnaviridae): Characterisation of a new large dsDNA algal virus that infects *Emiliana huxleyi*. *Arch. Virol.* **2002**, 147, (9), 1685–1698.
6. Brussaard, C.P.D.; Short, S.M.; Frederickson, C.M.; Suttle, C.A. Isolation and phylogenetic analysis of novel viruses infecting the phytoplankton *Phaeocystis globosa* (*Prymnesiophyceae*). *Appl. Environ. Microbiol.* **2004**, 70, (6), 3700–3705.
7. Moniruzzaman, M.; Gann, E.R.; LeCleir, G.R.; Kang, Y.; Gobler, C.J.; Wilhelm, S.W. Diversity and dynamics of algal Megaviridae members during a harmful brown tide caused by the pelagophyte, *Aureococcus anophagefferens*. *FEMS Microbiol. Ecol.* **2016**, 92, (5).
8. Short, S.M. The ecology of viruses that infect eukaryotic algae. *Environ. Microbiol.* **2012**, 14, (9), 2253–2271.
9. Brussaard, C.P.D. Viral control of phytoplankton populations—a review. *J. Eukaryot. Microbiol.* **2004**, 51, (2), 125–138.
10. Suttle, C.A. Marine viruses—Major players in the global ecosystem. *Nat. Rev. Microl.* **2007**, 5, (10), 801–812.
11. Wilhelm, S.W.; Suttle, C.A. Viruses and nutrient cycles in the sea: viruses play critical roles in the structure and function of aquatic food webs. *Bioscience* **1999**, 49, (10), 781–788.

12. Nagasaki, K.; Yamaguchi, M. Isolation of a virus infectious to the harmful bloom causing microalga *Heterosigma akashiwo* (*Raphidophyceae*). *Aquat. Microb. Ecol.* **1997**, 13, (2), 135–140.

13. Keizo, N.; Mineo, Y. Intra-species host specificity of HaV (*Heterosigma akashiwo* virus) clones. *Aquat. Microb. Ecol.* **1998**, 14, (1), 109–112.

14. Keizo, N.; Mineo, Y. Effect of temperature on the algicidal activity and the stability of HaV (*Heterosigma akashiwo* virus). *Aquat. Microb. Ecol.* **1998**, 15, (21), 211–216.

15. Lawrence, J.E.; Chan, A.M.; Suttle, C.A. A novel virus (HaNIV) causes lysis of the toxic bloom-forming alga *Heterosigma akashiwo* (*Raphidophyceae*). *J. Phycol.* **2001**, 37, (2), 216–222.

16. Lawrence, J.E.; Chan, A.M.; Suttle, C.A. Viruses causing lysis of the toxic bloom-forming alga *Heterosigma akashiwo* (*Raphidophyceae*) are widespread in coastal sediments of British Columbia, Canada. *Limnol. Oceanogr.* **2002**, 47, (2), 545–550.

17. Tai, V.; Lawrence, J.E.; Lang, A.S.; Chan, A.M.; Culley, A.I.; Suttle, C.A. Characterization of HaRNAV, a single-stranded RNA virus causing lysis of *Heterosigma akashiwo* (*Raphidophyceae*). *J. Phycol.* **2003**, 39, (2), 343–352.

18. Janice, E.L.; Curtis, A.S. Effect of viral infection on sinking rates of *Heterosigma akashiwo* and its implications for bloom termination. *Aquat. Microb. Ecol.* **2004**, 37, (1), 1– 7.

19. Lawrence, J.E.; Brussaard, C.P.D.; Suttle, C.A. Virus-specific responses of *Heterosigma akashiwo* to infection. *Appl. Environ. Microbiol.* **2006**, 72, (12), 7829–7834.

20. Granéli, E.; Edvardsen, B.; Roelke, D.L.; Hagström, J.A. The ecophysiology and bloom dynamics of *Prymnesium* spp. *Harmful Algae* **2012**, 14, 260–270.

21. Manning, S.R.; La Claire, J.W. Prymnesins: toxic metabolites of the golden alga, *Prymnesium parvum* Carter (*Haptophyta*). *Mar. Drugs* **2010**, 8, (3), 678–704.

22. Schoemann, V.; Becquevort, S.; Stefels, J.; Rousseau, V.; Lancelot, C. Phaeocystis blooms in the global ocean and their controlling mechanisms: A review. *J. Sea Res.* **2005**, 53, (1-2), 43–66.

23. Leblanc, K.; Hare, C.E.; Feng, Y.; Berg, G.M.; DiTullio, G.R.; Neeley, A.; Benner, I.; Sprengel, C.; Beck, A.; Sanudo-Wilhelmy, S.A.; *et al.* Distribution of calcifying and silicifying phytoplankton in relation to environmental and biogeochemical parameters during the late stages of the 2005 North East Atlantic Spring Bloom. *Biogeosciences* **2009**, 6, 2155–2179.

24. Malin, G.; Wilson, W.H.; Bratbak, G.; Liss, P.S.; Mann, N.H. Elevated production of dimethylsulfide resulting from viral infection of cultures of *Phaeocystis pouchetii*. *Limnol. Oceanogr.* **1998**, 43, (6), 1389–1393.

25. Edvardsen, B.; Paasche, E., Bloom dynamics and physiology of Prymnesium and Chrysochromulina. In: Anderson, D. M.; Cembella, A. D.; Hallegraeff, G. M., editors. Physiological Ecology of Harmful Algal Blooms. (NATO ASI Series G, vol. 41). Springer-Verlag, Berlin Heidelberg: 1998, 192-208.

26. Roelke, D.L.; Barkoh, A.; Brooks, B.W.; Grover, J.P.; Hambright, K.D.; LaClaire, J.W.; Moeller, P.D.R.; Patino, R. A chronicle of a killer alga in the west: ecology, assessment, and management of *Prymnesium parvum* blooms. *Hydrobiologia* **2016**, 764, (1), 29–50.

27. Schwierzke, L.; Roelke, D.L.; Brooks, B.W.; Grover, J.P.; Valenti, T.W.; Lahousse, M.; Miller, C.J.; Pinckney, J.L. *Prymnesium parvum* population dynamics during bloom development: a role assessment of grazers and virus. *J. Am. Water. Resour. Assoc.* **2010**, 46, (1), 63–75.

28. Authority, B., Broads Plan 2017. **2017**.

29. Holdway, P.A.; Watson, R.A.; Moss, B. Aspects of the ecology of *Prymnesium parvum* (Haptophyta) and water chemistry in the Norfolk Broads, England. *Freshwat. Biol.* **1978**, 8, (4), 295–311.

30. Wilson, W. H.; Van Etten, J. L.; Allen, M. J., The Phycodnaviridae: the story of how tiny giants rule the world. In *Lesser Known Large dsDNA Viruses*, Van Etten, J. L., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, **2009**, 1-42.

31. Den Boon, J.A.; Diaz, A.; Ahlquist, P. Cytoplasmic Viral Replication Complexes. *Cell Host Microbe* **2010**, 8, (1), 77–85.

32. Katoh, K.; Toh, H. Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinform.* **2008**, 9, (4), 286–298.

33. Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, 4, (4), 406–425.

34. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, 33, (7), 1870–1874.

35. Zauberman, N.; Mutsafi, Y.; Halevy, D.B.; Shimoni, E.; Klein, E.; Xiao, C.; Sun, S.; Minsky, A. Distinct DNA exit and packaging portals in the virus *Acanthamoeba polyphaga mimivirus*. *PLoS Biol.* **2008**, 6, (5), e114.

36. Mutsafi, Y.; Fridmann-Sirkis, Y.; Milrot, E.; Hevroni, L.; Minsky, A. Infection cycles of large DNA viruses: Emerging themes and underlying questions. *Virology* **2014**, 466–467, 3–14.

37. Feldman, H.A.; Wang, S.S. Sensitivity of various viruses to chloroform. *Proc. Soc. Exp. Biol. Med.* **1961**, 106, 736–738.

38. Nedbalová, L.; Střížek, A.; Sigler, K.; Řezanka, T. Effect of salinity on the fatty acid and triacylglycerol composition of five haptophyte algae from the genera Coccolithophora, Isochrysis and Prymnesium determined by LC-MS/APCI. *Phytochemistry* **2016**, 130, 64–76.

39. Johannessen, T.V.; Bratbak, G.; Larsen, A.; Ogata, H.; Egge, E.S.; Edvardsen, B.; Eikrem, W.; Sandaa, R. A. Characterisation of three novel giant viruses reveals huge diversity among viruses infecting *Prymnesiales* (*Haptophyta*). *Virology* **2015**, 476, 180–

40. Wilhelm, S.W.; Coy, S.R.; Gann, E.R.; Moniruzzaman, M.; Stough, J.M.A. Standing on the shoulders of giant viruses: five lessons learned about large viruses infecting small eukaryotes and the opportunities they create. *PLoS Pathog.* **2016**, 12, (8) e1005752.

41. Gallot-Lavallée, L.; Pagarete, A.; Legendre, M.; Santini, S.; Sandaa, R.-A.; Himmelbauer, H.; Ogata, H.; Bratbak, G.; Claverie, J.-M. The 474-kilobase-pair complete genome sequence of CeV-01B, a virus infecting *Haptolina* (*Chrysochromulina*) *ericina* (*Prymnesiophyceae*). *Genome Announc.* **2015**, 3, (6), e01413–e01415.

42. Sandaa, R.-A.; Heldal, M.; Castberg, T.; Thyrhaug, R.; Bratbak, G. Isolation and characterization of two viruses with large genome size infecting *Chrysochromulina ericina* (*Prymnesiophyceae*) and *Pyramimonas orientalis* (*Prasinophyceae*). *Virology* **2001**, 290, (2), 272–280.

43. Moniruzzaman, M.; LeCleir, G.R.; Brown, C.M.; Gobler, C.J.; Bidle, K.D.; Wilson, W.H.; Wilhelm, S.W. Genome of brown tide virus (AaV), the little giant of the Megaviridae, elucidates NCLDV genome expansion and host-virus coevolution. *Virology* **2014**, 466–467, 60–70.

44. Mirza, S.F.; Staniewski, M.A.; Short, C.M.; Long, A.M.; Chaban, Y.V.; Short, S.M. Isolation and characterization of a virus infecting the freshwater algae *Chrysochromulina parva*. *Virology* **2015**, 486, 105–115.

45. Henrikson, J.C.; Gharfeh, M.S.; Easton, A.C.; Easton, J.D.; Glenn, K.L.; Shadfan, M.; Mooberry, S.L.; Hambright, K.D.; Cichewicz, R.H. Reassessing the ichthyotoxin profile of cultured *Prymnesium parvum* (golden algae) and comparing it to samples collected from recent freshwater bloom and fish kill events in North America. *Toxicon* **2010**, 55, (7), 1396–1404.

46. Kozakai, H.; Oshima, Y.; Yasumoto, T. Isolation and structural elucidation of hemolysin from the phytoflagellate *Prymnesium parvum*. *Agric. Biol. Chem.* **1982**, 46, (1), 233–236.

47. Igarashi, T.; Satake, M.; Yasumoto, T. Prymnesin-2: a potent ichthyotoxic and hemolytic glycoside isolated from the red tide alga *Prymnesium parvum*. *J. Am. Chem. Soc.* **1996**, 118, (2), 479–480.

48. Igarashi, T.; Satake, M.; Yasumoto, T. Structures and partial stereochemical assignments for prymnesin-1 and prymnesin-2: potent hemolytic and ichthyotoxic glycosides isolated from the red tide alga *Prymnesium parvum*. *J. Am. Chem. Soc.* **1999**, 121, (37), 8499–8511.

49. Rasmussen, S.A.; Meier, S.; Andersen, N.G.; Blossom, H.E.; Duus, J.Ø.; Nielsen, K.F.; Hansen, P.J.; Larsen, T.O. Chemodiversity of ladder-frame prymnesin polyethers in *Prymnesium parvum*. *J. Nat. Prod.* **2016**, 79, (9), 2250–2256.

50. Roux, S.; Brum, J.R.; Dutilh, B.E.; Sunagawa, S.; Duhaime, M.B.; Loy, A.; Poulos, B.T.; Solonenko, N.; Lara, E.; Poulain, J.; *et al*. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **2016**, 537, (7622), 689–693.

51. Karsenti, E.; Acinas, S.G.; Bork, P.; Bowler, C.; De Vargas, C.; Raes, J.; Sullivan, M.; Arendt, D.; Benzoni, F.; Claverie, J.-M.; *et al*. A holistic approach to marine eco-systems biology. *PLoS Biol.* **2011**, 9, (10), e1001177.

52. Patel, A.; Noble, R. T.; Steele, J. A.; Schwalbach, M. S.; Hewson, I.; Fuhrman, J. A., Virus and prokaryote enumeration from planktonic aquatic environments by epifluorescence microscopy with SYBR Green I. *Nat. Protoc.* **2007,** 2, (2), 269-76.

53. Brenner, S.; Horne, R.W. A negative staining method for high resolution electron microscopy of viruses. *Biochim. Biophys. Acta* **1959**, 34, 103–110.

54. Gordon, G.B.; Miller, L.R.; Bensch, K.G. Fixation of tissue culture cells for ultrastructural cytochemistry. *Exp. Cell Res.* **1963**, 31, 440–443.

55. Martínez Martínez, J.; Boere, A.; Gilg, I.; van Lent, J.W.M.; Witte, H.J.; van Bleijswijk, J.D.L.; Brussaard, C.P.D. New lipid envelope-containing dsDNA virus isolates infecting *Micromonas pusilla* reveal a separate phylogenetic group. *Aquat. Microb. Ecol.* **2015**, 74, (1), 17–28.

56. GitHub. TGAC/KAT. [online] Available at: https://github.com/TGAC/KAT **2018** (Accessed 16 Feb. 2018).

57. IDBA project. [online] Available at: http://i.cs.hku.hk/~alse/hkubrg/projects/idba_ud/ **2018** (Accessed 16 Feb. 2018).

58. GitHub. Bbuchfink/diamond. [online] Available at: https://github.com/bbuchfink/diamond **2018** (Accessed 16 Feb. 2018).

59. Ab.inf.uni-tuebingen.de. MEGAN – Algorithms in bioinformatics. [online] Available at: http://ab.inf.uni-tuebingen.de/software/megan/ **2018** (Accessed 16 Feb. 2018).

60. Besemer, J.; Lomsadze, A.; Borodovsky, M. GeneMarkS: A self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **2001**, 29, (12), 2607–2618.

61. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, 215, (3), 403–410.

4 Discovery and Characterization of *de novo* KDN Biosynthesis in the Toxic Haptophyte, *Prymnesium parvum,* Suggests Widespread Sialic Acid Biosynthesis Amongst Algae

4.1 Abstract

Sialic acids are a family of more than 50 structurally distinct acidic carbohydrates found on the surface of all vertebrate cells. Frequently, sialic acids terminate the branches of glycan chains found on the surface of cells, exposing them to a range of interactions with the surrounding environment. Because of this, sialic acids often play important roles in host:pathogen interactions. The study of viruses that infect algae has boomed in recent years, but the molecular basis behind these viral infections remains unclear, with some speculation that sialic acids may play a role. However, the production of sialic acids by algae is largely unexplored. Here we report the *de novo* biosynthesis of the deaminated sialic acid, 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid (KDN), in the toxin-producing microalgae, *Prymnesium parvum*. Using biochemical methods we show that the alga contains CMP-KDN and that gene products of the algae convert Mannose-6-P to CMP-KDN; using bioinformatics we show that sialic acid biosynthesis is more widespread across the algae than previously thought, and may imply a key role for this acidic sugar in alga:virus infections.

4.2 Introduction

Sialic acids are acidic nine-carbon carbohydrates that are found cross kingdom [1] including on the surface of all vertebrate cells [2]. Frequently, sialic acids take up the terminal position on a glycan, exposing cell-surface sialic acids to a whole range of cell-cell interactions [3]. These interactions are most often exploited through protein-ligand interactions, where the proteins in question are the well-studied family of sialic acid binding lectins, and the sialic acids are the target ligand. Opportunistic pathogens, such as *E. coli* K1 or the influenza virus, often exploit this common biological interaction, where molecular mimicry can be used to evade host immunity, or recognition can be used to enhance virus binding [4, 5]. Alternatively, sialic acids can be modified in order to evade acquired immunity, and this has led to the discovery of over 50 different naturally occurring sialic acids [2]. The most common sialic acid in humans is *N*-acetyl neuraminic acid (Neu5Ac), first discovered by Klenk and Blix in the 1930s [6]. However other types of nine-carbon sialic acids also exist including the di-n-acetylated neuraminic acids, Pseudaminic acid and Legionaminic acid, as well as the deaminated neuraminic acid, 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid (KDN).

KDN was first discovered in 1986 in rainbow trout eggs [7] but since has been observed in salmon eggs [8], amphibian eggs [9-11], other organs of fish [12], as well as pathogenic bacteria [13, 14]. Its presence has also been observed at varying levels in different types of mammalian tissues [15, 16], although its presence in humans has since been shown to be a result of promiscuous Neu5Ac biosynthesis [17]. More recently, reports of KDN biosynthesis have been observed in the human gut symbiont *Bacteroides thetaiotaomicron* [18] (Figure 38), and to our knowledge the first report of sialic acids in microalgae when Fulton et al discovered a KDN-linked sphingolipid in the lipid rafts of the haptophyte *Emiliania huxleyi* [19, 20].

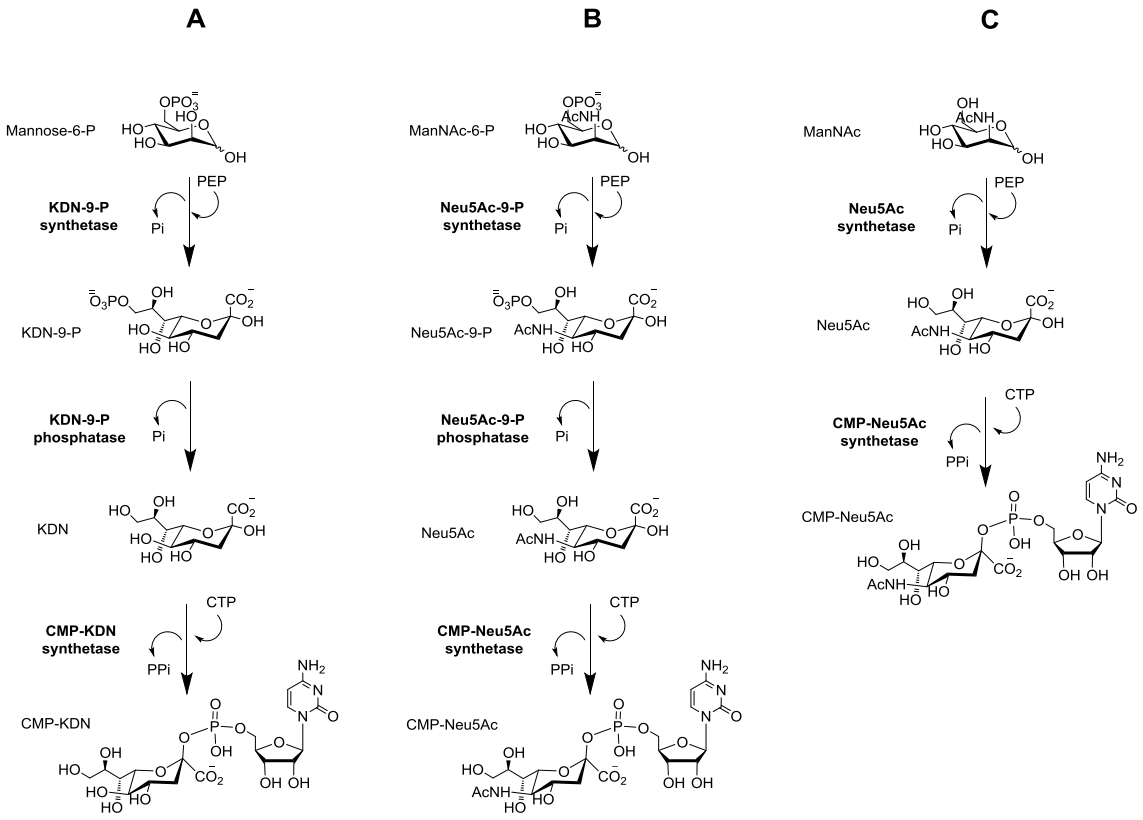


Figure 38 - Biosynthetic pathways of CMP-Neu5Ac in humans and bacteria (B&C, respectively) and CMP-KDN in *B. thetaiotaomicron* (A). Pathways for sialic acid biosynthesis in *B. thetaiotaomicron* and humans follows phosphorylation/dephosphorylation steps whereas many bacteria (C) do not utilise phosphorylated sugars as the starting point of the pathway, therefore not needing an additional step for phosphate removal.

The haptophytes are a widespread division of microalgae that play crucial roles in the ocean carbon and sulfur cycles [21, 22], but are also known to form blooms that are toxic to fish [23, 24]. The toxin-producing species in this family mainly belong to the *Prymnesium* and *Chrysochromulina* genus, the most studied species being *Prymnesium parvum*; which is known to cause harmful blooms that result in mass fish mortalities due to the release of extracellular toxins [24-28]. *P. parvum* is cosmopolitan, with blooms causing economic disaster on all continents except the Antarctica [29]. In evolutionary terms, *P. parvum* and the haptophyte family are derived from the red algal plastid, which was known to have undergone a significant amount of horizontal gene transfer with bacteria [30]. This is supported by a *P. parvum* transcriptome from The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) [31] that contains a plethora of both eukaryotic and prokaryotic genes that may aid the alga in allelopathy during nutrient limitation when competition is at its highest. Recent

work from our group has discovered a novel lytic virus that infects this alga and has been implicated in toxic blooms of this species [32].

Research into algal viruses has boomed in the last two decades; brought about by the discovery of *Acanthamoeba polyphaha mimivirus* in 2003 [33]. They have since been shown to play crucial roles in ecosystem dynamics [34] and biogeochemical cycles [35], where large scale lysis of algal populations can lead to changes in nutrient cycles in water systems. Although the effect of algal viruses has been studied in some detail, the molecular mechanisms behind these infections remains poorly understood. Some examples have suggested algal viruses utilise mammalian virus infection strategies, hijacking autophagy pathways [36] and utilising host sialic acids in their infection process [19, 20].

The recent reports by Fulton et al of a KDN-containing sphingolipid [19, 20] (Figure 39), the devastating effect of *P. parvum* on fish aquaculture [24], and the large genetic capacity of the haptophyte family made us question whether *P. parvum*, and more broadly the haptophytes, were capable of *de novo* sialic acid biosynthesis. Herein, we report a bioinformatics guided discovery of a biosynthetic pathway leading to cytidyl-activated KDN, as well as showing that multiple strains of *Prymnesium* all contain KDN. We show the nucleotide sugar, CMP-KDN, accumulates to significant levels in the cell. Moreover, using phylogenetic analysis we show that sialic acid biosynthesis is widespread amongst the haptophytes and dinoflagellates, members of the Alveolata phyla. Extending on this we show that biosynthesis of the structurally similar 2-keto-3-deoxyoctonic acid, KDO, is found scattered across algal groups. This work conclusively shows sialic acid biosynthesis in a unicellular eukaryote microalga and vastly expands the limited previous knowledge of the occurrence of this metabolic pathway in unicellular eukaryotes.

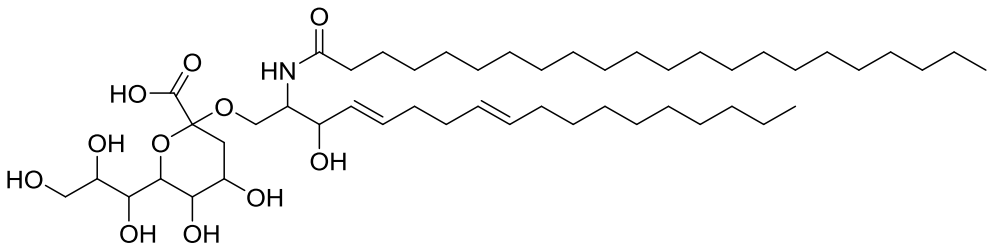


Figure 39 – Putative structure of a KDN-containing sphingolipid. Lipid originally isolated from the haptophyte, *Emiliania huxleyi* [19].

4.3 Results

4.3.1 DMB-HPLC analysis of sialic acids

To first determine the presence and type of sialic acids in *Prymnesium*, 15 strains were screened using the fluorescent dye 1,2-diamino-4,5-methylenedioxybenzene (DMB). Each strain was subjected to mild acid hydrolysis with hydrochloric acid to release free sialic acids. Following removal of insoluble cell debris, the resulting supernatants were reacted with DMB, before the labelled mixtures were injected onto an LC-MS system with in-line fluorescence. Peaks aligning with the internal standard Neu5Ac were not found (Figure 40A) but peaks aligning with the internal standard KDN could be observed in all 15 strains (Figure 40B). The predicted peaks for KDN in strains of *Prymnesium* contained identical masses to those seen for the KDN internal standard (Figure 41) and aligned well with reported literature values for DMB-KDN adducts. Masses corresponding to the DMB-adducts of other sialic acids could not be observed in any of the strains upon manual inspection [37]. Taken together, these results strongly suggested *Prymnesium* strains can produce the deaminated sialic acid, KDN.

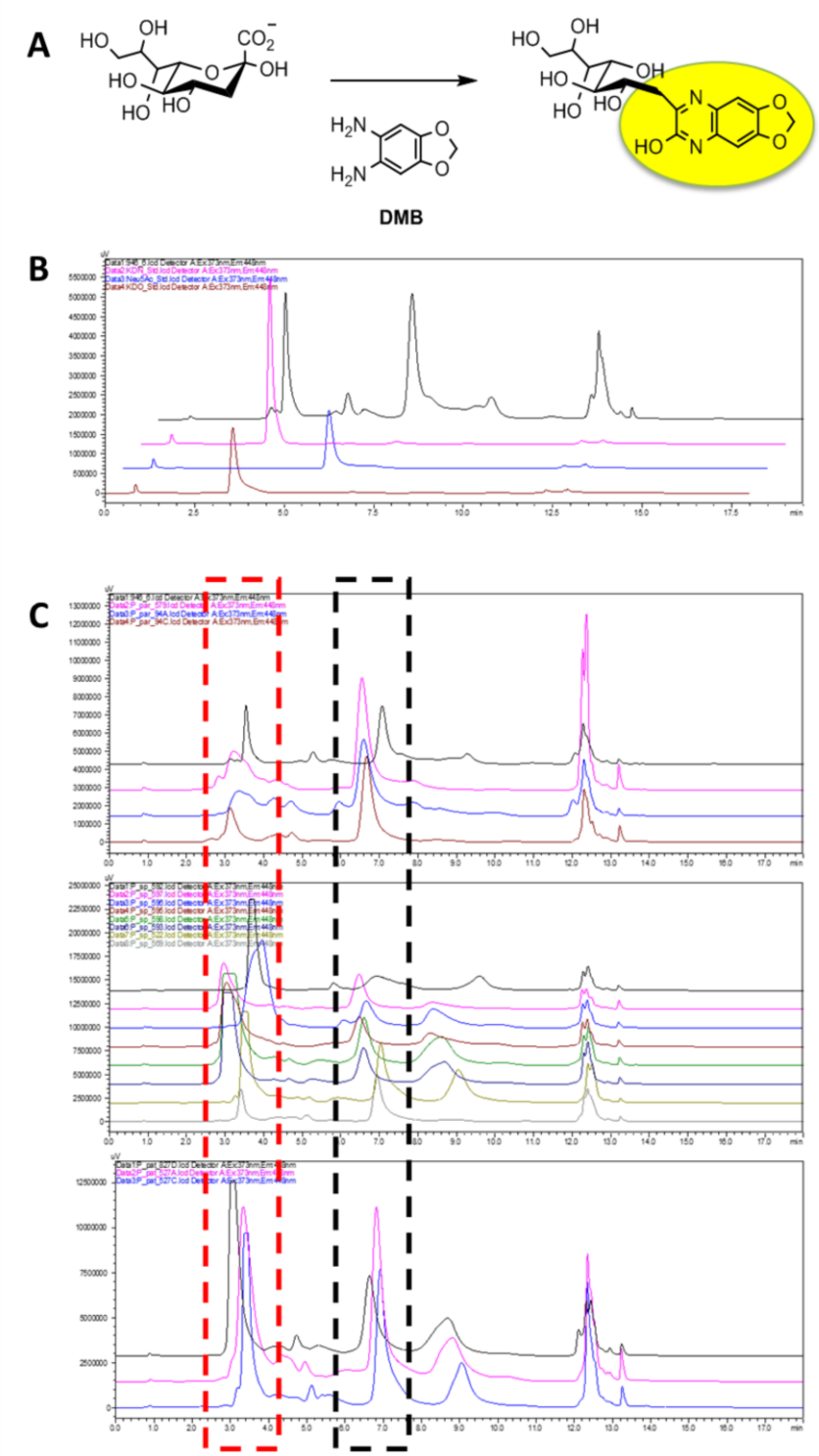


Figure 40 - DMB-HPLC labelling of sialic acids in 15 strains of *Prymnesium*. (A) General scheme for DMB labelling of KDN. Yellow circle represents a fluorescent product. (B) Brown – KDO standard. Blue – Neu5Ac standard. Pink – KDN standard. *P. parvum* 946/6 contains peaks aligning with KDN/KDO and are confirmed to be KDN by subsequent m/s analysis (Figure 41). (C) Top – *P. parvum* strains. Middle – *P. sp.* strains. Bottom – *P. patelliferum* strains. Peaks around 2.6-3.2 minutes (Red hashed line) correspond to DMB-KDN and are confirmed by identification of corresponding DMB-KDN masses (Figure 41). Other peaks (Black hashed line) are unknown but do not relate to masses of any known DMB-sialic acid adducts or their acetylated derivatives.

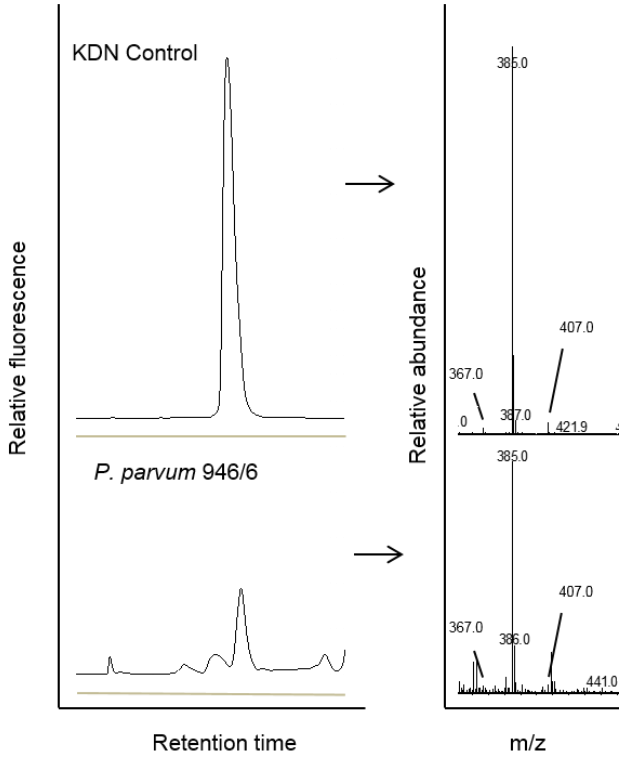


Figure 41 – *P. parvum* 946/6 contains KDN. Left – HPLC analysis of DMB-derivatized sialic acids from whole cell extracts of *P. parvum* show peaks aligning to KDN standards. Right – ESI-MS signals from peaks aligning with KDN match those of signals seen for KDN standards.

4.3.2 Sugar nucleotide profiling of CMP-KDN

Monosaccharides are frequently activated inside the cell by attachment to nucleotide-phosphate bases. Whilst hexoses and pentoses are most commonly activated with uridine diphosphate (UDP), thymidine diphosphate (TDP), guanidine diphosphate (GDP) or adenine diphosphate (ADP), sialic acids are activated with cytidine monophosphate (CMP) [2]. Therefore, we next sought to establish the levels of intracellular CMP-KDN in *Prymnesium parvum* 946/6. Whole cells were extracted using ethanol in late-log phase using a modified method of Turnock *et al* [38]. Following a partitioning between water and butan-1-ol, the samples were then subjected to solid phase extraction (SPE) using EnviCarb graphitised carbon columns following previously established methods [39]. Based on work by Pabst *et al* [40], samples were then analysed and intracellular levels of CMP-KDN quantified using LC-MS/MS. A detailed protocol for the extraction and quantification of sugar nucleotides can be found in Rejzek *et al* [41].

CMP-KDN produced in this study was used as an internal standard to generate multiple reaction monitoring (MRM) transitions of the precursor and fragments and to determine

retention times. Where in doubt, co-injection of cell extracts with internal standards was used to confirm analyte identification. Using this methodology, a strong signal for CMP-KDN could be observed in extracts of *P. parvum* 946/6. Using peak area integration and by comparison with a second internal standard, GDP-α-D-Glc, levels of intracellular CMP-KDN were estimated to be 206.6 ± 72.1 pmol per gram of wet cell pellet (Table 5). Concentrations of the other sugar nucleotides used in this study for comparison were found to range between 71.2 ± 17.4 and 328.2 ± 101.0 pmol per gram of cell pellet, showing that CMP-KDN makes up a key sugar nucleotide in *P. parvum* glycobiology.

Sugar Nucleotide Species	Relative Retention Time [min]	MRM transitions	Fragment	Concentration [pmol/g of cell pellet]
CMP-α-D-KDN	0.8	572 → 322 572 → 97	$[NMP-H]^-$ $[H_3PO_4-H]^-$	206.6 ± 72.1*
UDP-α-D-Glc	1	565 → 323 565 → 79	$[NMP-H]^-$ $[H_3PO_4-H_3O]^-$	328.2 ± 101.0*
UDP-α-D-Gal	0.92	565 → 323 565 → 159	$[NMP-H]^-$ $[H_4P_2O_7-H_3O]^-$	71.2 ± 17.4*
GDP-α-D-Glc (added internal standard)	1.56	604 → 362 604 → 241	$[NMP-H]^-$ $c[Glc-1-P-H-H_2O]^-$	1540.0**

Table 5 – Relative retention times and MRM transitions of sugar nucleotides. Sugar nucleotides examined in this study (CMP-α-D-KDN) and others shown for comparison (UDP-α-D-Glc, UDP-α-D-Gal and internal standard GDP-α-D-Glc). *The data are mean of 3 biological replicates, ± indicates standard error. ** GDP-α-D-Glc (1540 pmol / g wet pellet) was added to enable quantification of other sugar nucleotides.

4.3.3 Identification of putative CMP-KDN biosynthesis transcripts in *P. parvum*

Protein sequences involved in CMP-KDN biosynthesis in *B. thetaiotaomicron* VPI-5482 [18] were used to identify transcripts from *P. parvum* with high sequence similarity. Subsequently, protein sequences from the *P. paruvm* transcriptome with high similarity to KDN-9-P synthase (CAMPEP_0191217894) and CMP-transferase (CAMPEP_0191219004) were identified. No sequences were found with high sequence similarity (*E*-values ≤ $1E^{-05}$) to the second enzyme of

the pathway, KDN-9-P phosphatase. Further analysis of the CMP-transferase sequence identified a putative transmembrane domain at the C-terminus of the protein. No sequence alignments corresponding to transmembrane domains were seen in the KDN-9-P synthase.

4.3.4 Cloning and expression of *P. parvum* KDN-9-P synthase (CAMPEP_0191217894), *B. thetaiotaomicron* KDN-9-P phosphatase and *P. parvum* CMP-transferase (CAMPEP_0191219004)

Sequences with high sequence similarity to KDN-9-P synthase and CMP-transferase from *P. parvum* (CAMPEP_0191217894, CAMPEP_0191219004), along with KDN-9-P phosphatase from *B. thetaiotaomicron* VPI-5482 (ExPASy accession number: Q8A712), were codon optimized for expression in *E. coli* and synthesized using IDT's gBlock gene fragment service. Sequences were synthesized with In-Fusion™ cloning overhangs which meant they could be cloned directly into pOPINF without the need for any rounds of PCR amplification.

KDN-9-P synthase expressed in high levels (24 mg/L of *E. coli*) (Figure 42), but caused problems following affinity chromatography. Purified protein showed a tendency to precipitate if left more than 2 days. This issue was resolved by including DTT (5 mM) into the protein lysis and extraction buffer, as well as diluting purified fractions after affinity chromatography.

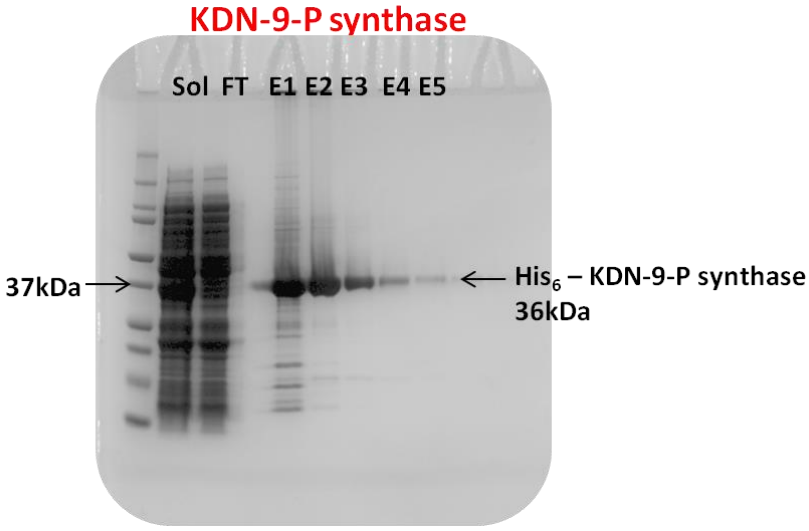


Figure 42 - SDS-PAGE analysis of KDN-9-P synthase expressed in *E. coli* BL21 Codon+. Ladder = Precision Plus Protein™ Dual Color Ladder. Sol = soluble protein; FT = column flow through; E1-5 = elution fractions.

Although expression levels of *B. thetaiotaomicron* VPI-5482 KDN-9-P phosphatase were largely insoluble under these conditions, low milligram levels of soluble protein could be purified with no issues (Figure 43).

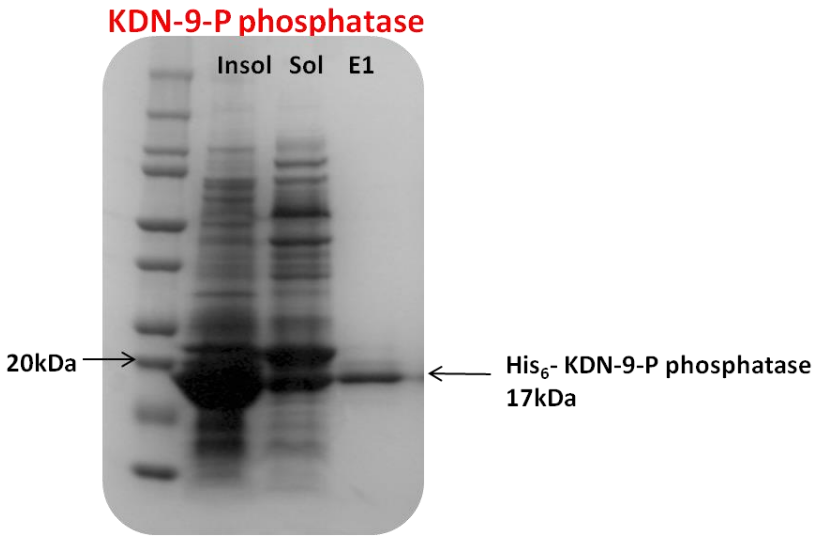


Figure 43 - SDS-PAGE analysis of KDN-9-P phosphatase expressed in *E. coli* BL21 Codon+. Ladder = Precision Plus Protein™ Dual Color Ladder. Insol = insoluble protein; Sol = soluble protein; E1 = elution fraction.

Initial attempts to express full length CMP-transferase in *E. coli* were unsuccessful, resulting in insoluble protein of the expected mass. Further analysis of the sequence using Phobius identified a C-terminal transmembrane domain which was subsequently truncated using PCR before cloning and expression was re-attempted. Truncated CMP-transferase (AA 1-233) was cloned and expressed and resulted in good levels of soluble protein of the expected mass (Figure 44).

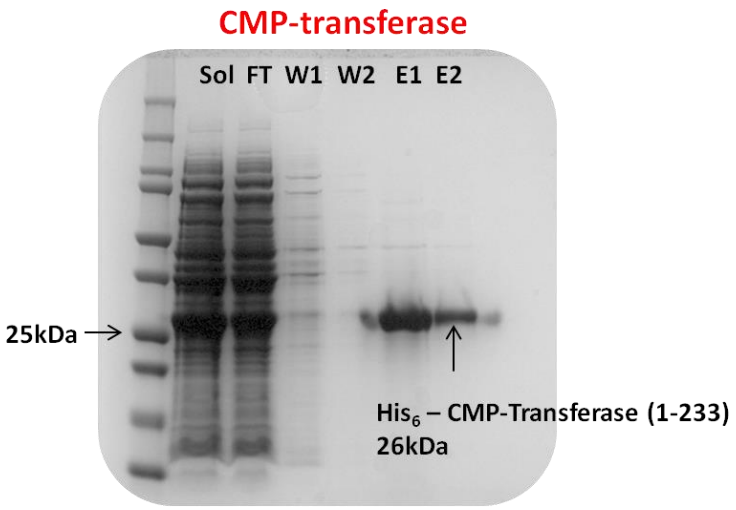


Figure 44 - SDS-PAGE analysis of CMP-transferase expressed in *E. coli* BL21 Codon+. Ladder = Precision Plus Protein™ Dual Color Ladder. Sol = soluble protein; FT = column flow through; W1-2 = wash fractions; E1-2 = elution fractions.

4.3.5 CAMPEP_0191217894 encodes a functional KDN-9-P synthase

For characterization of CAMPEP_0191217894, a combination of ^{1}H NMR, ^{31}P NMR and ESI-MS were used. Based on sequence similarity to *B. thetaiotaomicron* VPI-5482 BT1714 [18], the reaction was anticipated to produce KDN-9-P from phosphoenol pyruvate (PEP) and mannose-6-phosphate. Reaction solutions contained 38 µg of CAMPEP_0191217894, 2 mM $MgCl_2$, 8 mM of PEP and 10 mM of mannose-6-phosphate, buffered in 50 mM HEPES (pD 7.5). Reactions were monitored at room temperature and overnight by ^{1}H NMR, whereby formation of the axial and equatorial H_3 protons could be easily followed (Figure 45). For ^{31}P NMR, consumption of mannose-6-phosphate and PEP could be readily followed (Figure 46); coinciding with the depletion of these two peaks, formation of two new peaks representative of inorganic phosphate and KDN-9-P could be observed. As with ^{1}H NMR values, ^{31}P NMR values corresponding to KDN-9-P aligned well with previously published values [18].

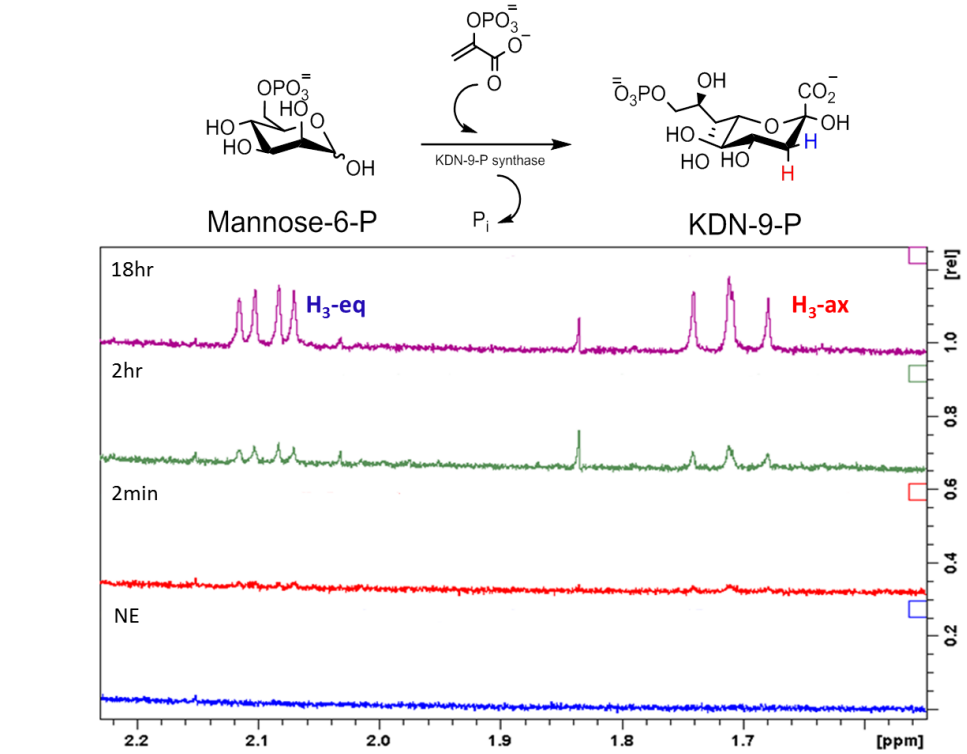


Figure 45 – Reaction catalysed by CAMPEP_0191217894 as followed by ^{1}H NMR. Reactions were monitored overnight by ^{1}H NMR by following formation of axial and equatorial H_3 protons on KDN-9-P.

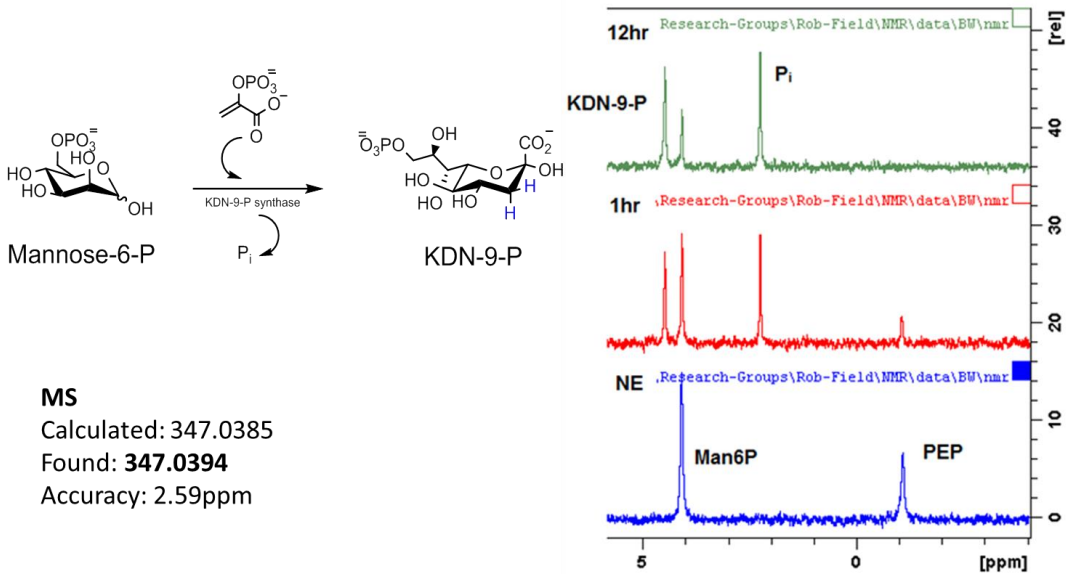


Figure 46 - Reaction catalysed by CAMPEP_0191217894 as followed by ^{31}P NMR. Blue – Before enzyme addition; two signals representing mannose-6-P and PEP can be seen. Red – 1 hour after enzyme addition; two new signals representing KDN-9-P and P_i can be seen. Green – 12 hours after enzyme addition; Reaction completion as seen by consumption of all PEP. ESI-MS of the reaction products confirmed formation of KDN-9-P.

The specificity of the enzyme was then examined by screening with a range of alternative sugar substrates known to produce sialic acid products (Table 6). Reaction conditions were the same as previously described and sugar substrates were used in a slight excess to ensure that PEP consumption could be attributed to complete conversion by ^{31}P NMR. ESI-MS of the reaction mixtures after 18 hours was carried out to look for the respective sialic acid products. In each case, conversion was only seen when mannose-6-P was used as the starting sugar substrate, which had gone to completion at the 18-hour time point. Masses attributed to other sialic acid products could not be detected in any reaction mixture, nor was any level of conversion observed by ^{31}P NMR, suggesting a strong enzyme specificity for KDN-9-P production.

Substrate	**Predicted Product**	**^{31}P NMR conversion at 18 hr (%)**	**ESI-MS presence of product at 18 hr**
Man-6-P	KDN-9-P	100	Y
ManNAc-6-P	Neu5Ac-9-P	0	N
Ara-5-P	KDO-8-P	0	N
Man	KDN	0	N
ManNAc	Neu5Ac	0	N
ManN	Neu	0	N
Ara	KDO	0	N

Table 6 – Specificity of KDN-9-P synthase for mannose-6-P as determined by ^{31}P NMR and ESI-MS of the reaction mixture after 18 hours.

In addition to following the enzymatic reaction progress by NMR and MS techniques, a colorimetric phosphate-release assay was employed (BioMol Green, Enzo) to establish kinetic parameters. This assay monitored the reaction by following release of the bi-product, inorganic phosphate. For determination of K_m values for both PEP and mannose-6-P, reactions were set up to include 100 µM of one substrate whilst varying the concentration of the second. The steady state kinetic values determined at 25 ^{o}C using this method are k_{cat} = 14.91 min^{-1}, V_{max} = 30.41 ± 1.94 pmol min^{-1}, $^{(PEP)}K_m$ = 18.31 ± 1.83 µM (at 100 µM mannose-6-P) and $^{(mannose\text{-}6\text{-}P)}K_m$ = 63.63 ± 10.51 µM (at 100 µM PEP) (Figure 47).

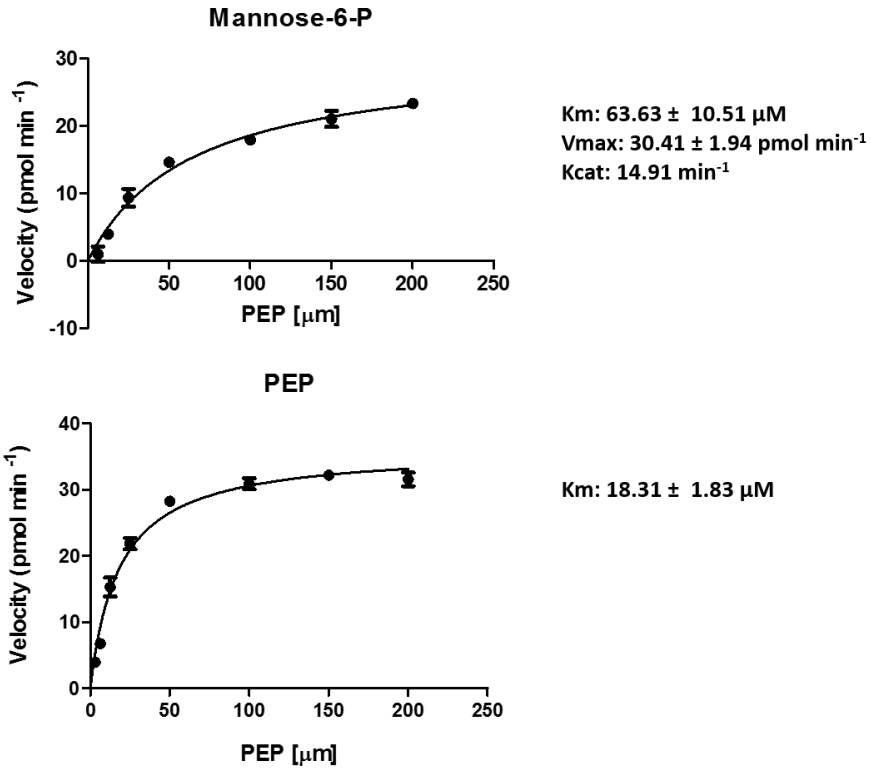


Figure 47 - Kinetic analysis of KDN-9-P synthase with PEP and mannose-6-P as substrates. In each case, when one substrates concentration was varied (x-axis), the other was kept constant (100 μM). Values reported are the result of 3 repeats and standard error is shown where applicable.

4.3.6 KDN-9-P phosphatase accepts the product of CAMPEP_0191217894 (KDN-9-P synthase)

To further confirm the identity of the KDN-9-P synthase product and to generate substrate (KDN) for CMP-transferase, a coupled assay was employed utilizing both *P. parvum* KDN-9-P synthase and *B. thetaiotaomicron* VPI-5482 KDN-9-P phosphatase. KDN-9-P synthase was added to an excess of PEP and mannose-6-P and left for 3 hours to produce KDN-9-P and the bi-product, inorganic phosphate. KDN-9-P phosphatase was then added and the reaction was left for a further 3 hours. The loss of KDN-9-P after addition of KDN-9-P phosphatase by ^{31}P NMR clearly showed that the product of KDN-9-P synthase is accepted by *B. thetaiotaomicron* VPI-5482 KDN-9-P phosphatase (Figure 48).

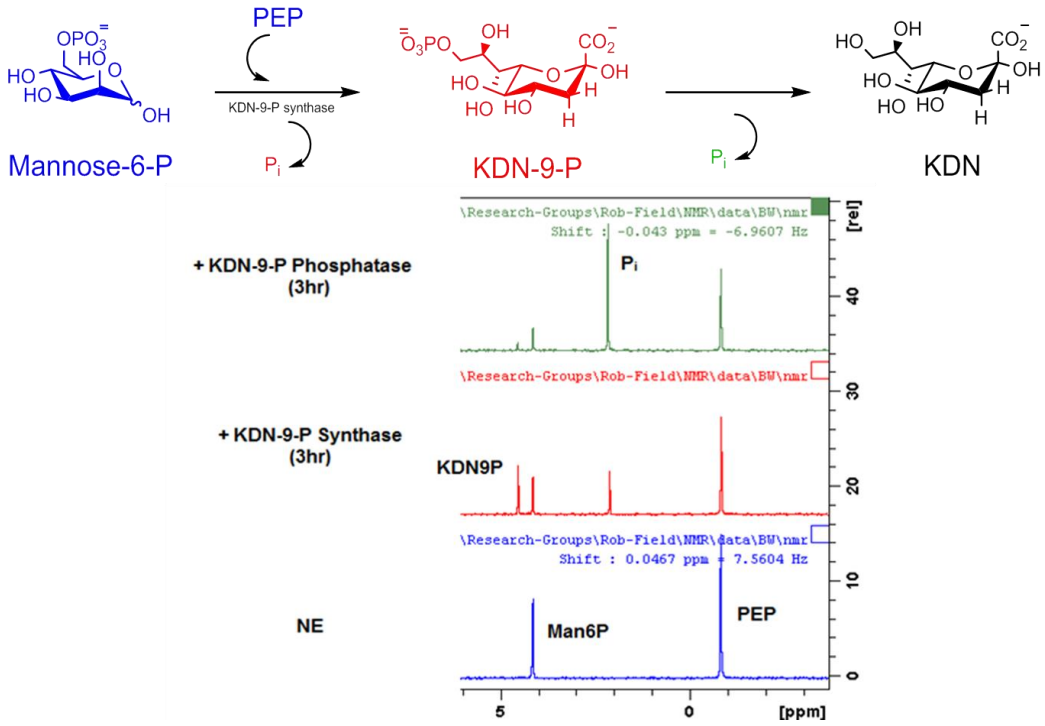


Figure 48 – ^{31}P NMR time course showing KDN-9-P synthase and KDN-9-P phosphatase activity. Incubation of the reaction product of KDN-9-P synthase with KDN-9-P phosphatase from *B. thetaiotaomicron* VPI-5482 leads to consumption of KDN-9-P and thus produces more inorganic phosphate.

4.3.7 CAMPEP_0191219004 encodes a functional CMP-transferase

Following cleavage of phosphate from KDN-9-P to produce KDN, it was speculated that the last enzyme in the pathway would be a CMP-transferase to active KDN. For this reason, cytidine 5'-triphosphatase (CTP) and KDN (Produced by KDN-9-P phosphatase or commercial – Sigma Aldrich, United Kingdom) were used as substrates in a buffered solution containing 100 mM Tris-HCl, pH 9.0, 10 mM $MgCl_2$. Reactions were monitored at 25 ^{o}C by ^{31}P NMR to follow reaction progress (Figure 49), where loss of signals for CTP and subsequent formation of signals for CMP-KDN and the bi-product PPi could be observed. When the reaction was seen to have gone to completion the product was purified by strong anion exchange (SAX), before being subjected to ^{1}H NMR, ^{31}P NMR, and ESI-MS analysis. All NMR values aligned well with reported literature values [18] and MS confirmed the expected identity of CMP-KDN.

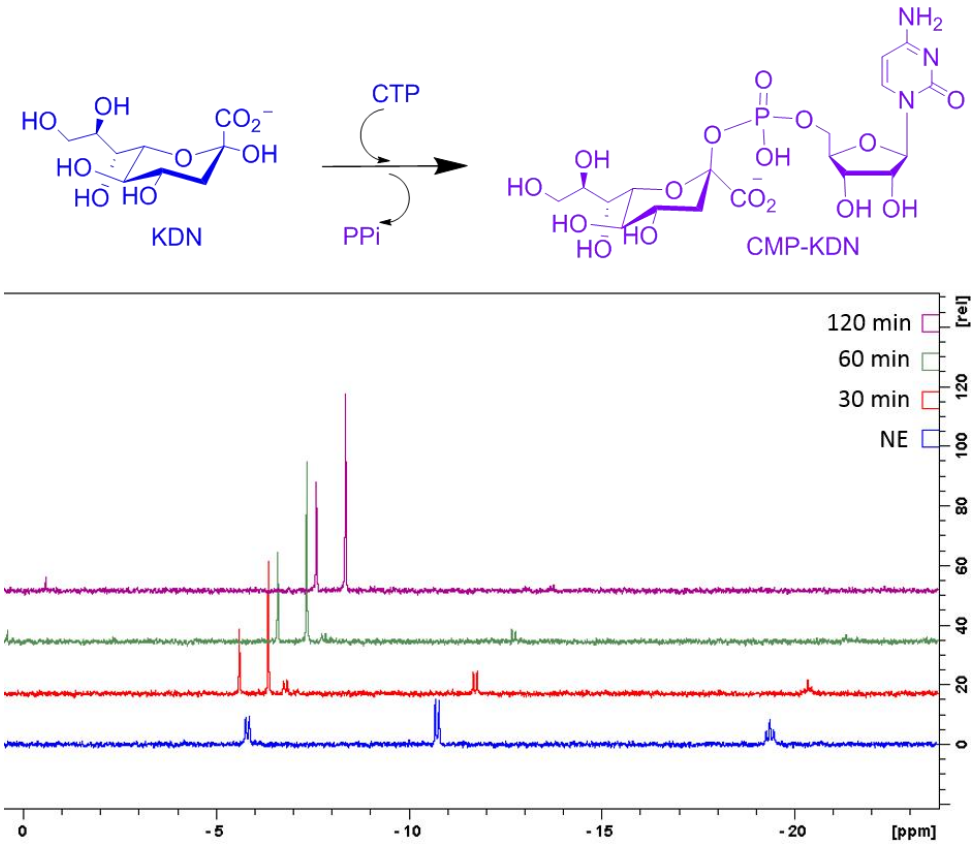


Figure 49 - Reaction catalysed by CAMPEP_0191219004 (CMP-transferase). Reactions were monitored over 120 minutes by ^{31}P NMR. Blue (NE) shows 3 phosphorus signals from the starting material CTP. Over the 120-minute time course two new singlets appear for PPi (-5.37 ppm) and CMP-KDN (-4.61 ppm). The reaction reaches near completion at this point (~98% by peak area integration). Spectrums are offset by 0.1 ppm for ease of visualization.

4.3.8 Identification of sialic acid synthase and CMP-transferase genes across the algal groups

To identify algal transcripts involved in sialic acid biosynthesis across the algal groups, BLASTp searches were carried out using newly discovered sequences for KDN-9-P synthase and CMP-transferase from *P. parvum* as consensus sequences. BLASTp searches were performed against 160 translated transcriptomes from MMETSP [31] or genomes from NCBI [42] where available. It was quickly noted that several other algae including the green alga *Pyrammimonas parkae* contained sequences with higher sequence similarity to KDO-8-P synthase. KDO-8-P synthase produces the structurally similar 2-keto-3-deoxyoctonic acid, KDO, and the enzyme is structurally distinct by sequence to sialic acid synthases that produce the C9 sialic acids KDN or Neu5Ac [43]. Therefore, the KDO-8-P synthase transcript (CAMPEP_0191499076) and CMP-transferase transcript (CAMPEP_0191478336) from *P. parkae* were also used as consensus sequences to query the nucleic acid databases.

As expected, most species of algae that contained sequences with high sequence identity to either the C9-like sialic acid synthases or C8-like 2-keto-3-deoxyoctonic acid synthases also encoded CMP-transferases to active said sugars (Figure 50). Only 15 of 160 algal strains examined had one of the two parts of the pathway. Sequences involved in sialic acid or KDO biosynthesis could be found across most algal groups, except for the glaucophytes, excavates and rhizaria where only *Chlorarachnion reptans* contained sialic acid biosynthetic machinery. Of the 160 databases examined, 48 contained sequences with high sequence similarity to sialic acid synthases such as the KDN-9-P synthase from *P. parvum*. Of these 48, only 7 are found outside of the Haptophyta and Dinoflagellata phyla, with all but one haptophyte found to contain homologues to this enzyme. Homologues to KDO-8-P synthase are more scattered but appear to be particularly frequent in green algae and cryptophytes (46% and 53% abundant, respectively). Sequences for both KDO and C9 sialic acid synthesis appear to be mutually exclusive with some exceptions in the haptophytes.

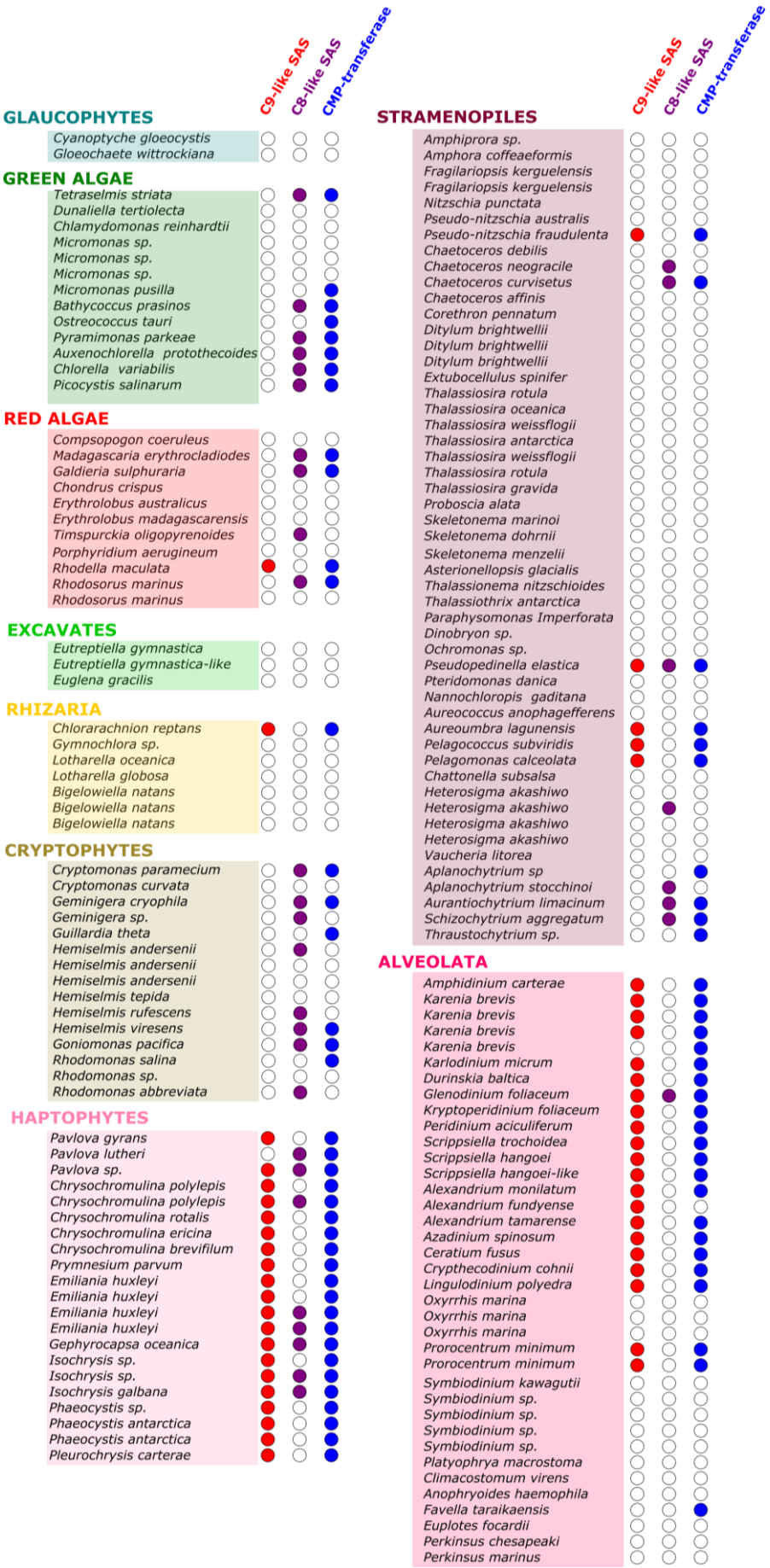

C9-like SAS
C8-like SAS
CMP-transferase
GLAUCOPHYTES
Cyanoptyche gloeocystis
Gloeochaete wittrockiana
GREEN ALGAE
Tetraselmis striata
Dunaliella tertiolecta
Chlamydomonas reinhardtii
Micromonas sp.
Micromonas sp.
Micromonas sp.
Micromonas pusilla
Bathycoccus prasinos
Ostreococcus tauri
Pyramimonas parkeae
Auxenochlorella protothecoides
Chlorella variabilis
Picocystis salinarum
RED ALGAE
Compsopogon coeruleus
Madagascaria erythrocladiodes
Galdieria sulphuraria
Chondrus crispus
Erythrolobus australicus
Erythrolobus madagascarensis
Timspurckia oligopyrenoides
Porphyridium aerugineum
Rhodella maculata
Rhodosorus marinus
Rhodosorus marinus
EXCAVATES
Eutreptiella gymnastica
Eutreptiella gymnastica-like
Euglena gracilis
RHIZARIA
Chlorarachnion reptans
Gymnochlora sp.
Lotharella oceanica
Lotharella globosa
Bigelowiella natans
Bigelowiella natans
Bigelowiella natans
CRYPTOPHYTES
Cryptomonas paramecium
Cryptomonas curvata
Geminigera cryophila
Geminigera sp.
Guillardia theta
Hemiselmis andersenii
Hemiselmis andersenii
Hemiselmis andersenii
Hemiselmis tepida
Hemiselmis rufescens
Hemiselmis viresens
Goniomonas pacifica
Rhodomonas salina
Rhodomonas sp.
Rhodomonas abbreviata
HAPTOPHYTES
Pavlova gyrans
Pavlova lutheri
Pavlova sp.
Chrysochromulina polylepis
Chrysochromulina polylepis
Chrysochromulina rotalis
Chrysochromulina ericina
Chrysochromulina brevifilum
Prymnesium parvum
Emiliania huxleyi
Emiliania huxleyi
Emiliania huxleyi
Emiliania huxleyi
Gephyrocapsa oceanica
Isochrysis sp.
Isochrysis sp.
Isochrysis galbana
Phaeocystis sp.
Phaeocystis antarctica
Phaeocystis antarctica
Pleurochrysis carterae
C9-like SAS
C8-like SAS
CMP-transferase
STRAMENOPILES
Amphiprora sp.
Amphora coffeaeformis
Fragilariopsis kerguelensis
Fragilariopsis kerguelensis
Nitzschia punctata
Pseudo-nitzschia australis
Pseudo-nitzschia fraudulenta
Chaetoceros debilis
Chaetoceros neogracile
Chaetoceros curvisetus
Chaetoceros affinis
Corethron pennatum
Ditylum brightwellii
Ditylum brightwellii
Ditylum brightwellii
Extubocellulus spinifer
Thalassiosira rotula
Thalassiosira oceanica
Thalassiosira weissflogii
Thalassiosira antarctica
Thalassiosira weissflogii
Thalassiosira rotula
Thalassiosira gravida
Proboscia alata
Skeletonema marinoi
Skeletonema dohrnii
Skeletonema menzelii
Asterionellopsis glacialis
Thalassionema nitzschioides
Thalassiothrix antarctica
Paraphysomonas Imperforata
Dinobryon sp.
Ochromonas sp.
Pseudopedinella elastica
Pteridomonas danica
Nannochloropis gaditana
Aureococcus anophagefferens
Aureoumbra lagunensis
Pelagococcus subviridis
Pelagomonas calceolata
Chattonella subsalsa
Heterosigma akashiwo
Heterosigma akashiwo
Heterosigma akashiwo
Heterosigma akashiwo
Vaucheria litorea
Aplanochytrium sp
Aplanochytrium stocchinoi
Aurantiochytrium limacinum
Schizochytrium aggregatum
Thraustochytrium sp.
ALVEOLATA
Amphidinium carterae
Karenia brevis
Karenia brevis
Karenia brevis
Karenia brevis
Karlodinium micrum
Durinskia baltica
Glenodinium foliaceum
Kryptoperidinium foliaceum
Peridinium aciculiferum
Scrippsiella trochoidea
Scrippsiella hangoei
Scrippsiella hangoei-like
Alexandrium monilatum
Alexandrium fundyense
Alexandrium tamarense
Azadinium spinosum
Ceratium fusus
Crypthecodinium cohnii
Lingulodinium polyedra
Oxyrrhis marina
Oxyrrhis marina
Oxyrrhis marina
Prorocentrum minimum
Prorocentrum minimum
Symbiodinium kawagutii
Symbiodinium sp.
Symbiodinium sp.
Symbiodinium sp.
Symbiodinium sp.
Platyophrya macrostoma
Climacostomum virens
Anophryoides haemophila
Favella taraikaensis
Euplotes focardii
Perkinsus chesapeaki
Perkinsus marinus

Figure 50 - Coulson plot showing the distribution of sialic acid biosynthesis pathways in algae. A total of 160 transcriptomes or genomes were analysed for the presence of C9 or C8 sialic acid synthases, as well as the corresponding CMP-transferase enzymes. Where a transcript was found for a given gene, a filled circle can be found. For C9 sialic acid synthases, circles are filled red. For C8 sialic acid synthases, circles are filled purple. For CMP-transferase genes, circles are filled blue. Where the same species is mentioned more than once, difference strains have been analysed. A full list of strains used in this study, as well as sources of transcriptome, genome and sequence identifiers can be found in Supplementary Material 1.

4.3.9 Phylogenetic analysis of algal sialic acid biosynthesis machinery

Maximum likelihood phylogenetic trees were constructed to compare sialic acid synthase/KDO synthase genes (Figure 51) and CMP-transferase genes (Figure 52) across the algal groups. Figure 51 shows that sequences with higher sequence identity to *P. parvum* KDN-9-P synthase (bottom clade on tree) do not form a monophyletic lineage with sequences more homologous to KDO-8-P synthase-like sequences. The clade with higher homology to KDN-9-P synthase from *P. parvum* is dominated by sequences from the Alveolata and Haptophyta, whilst sequences with higher homology to KDO-8-P synthase are scattered throughout the algal groups. CMP-transferase sequences show a similar trend (Figure 52), with a clade at the bottom half of the tree showing higher homology to the CMP-KDN synthase from *P. parvum* and a clade at the top showing higher homology to the CMP-transferase sequence from *P. parkae* (CAMPEP_0191478336). As with the sialic acid synthases, the former clade is dominated by transcripts from the Alveolata and Haptophyta, whilst the clade with higher homology to the CMP-transferase from *P. parkae* contains a mixture of algae from most groups examined.

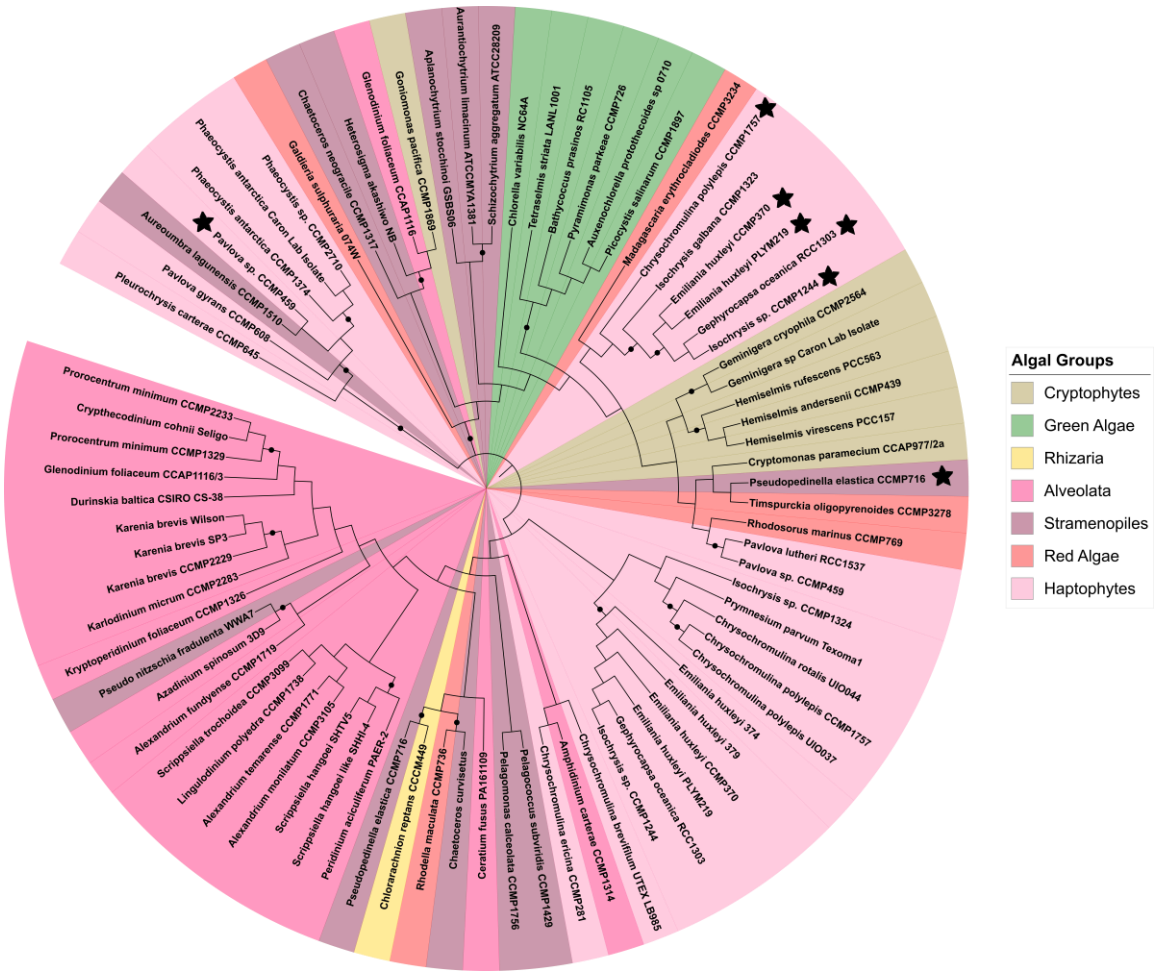


Figure 51 - Phylogenetic clustering of sialic acid synthases across the algal groups. Sequences with high homology to *P. parvum* KDN-9-P synthase (CAMPEP_0191217894) can be found on the bottom half of the tree. Sequences with higher homology to *P. parkae* KDO-8-P synthase (CAMPEP_0191499076) are found on the top half of the tree. Organisms with a star symbol contain two sequences in the tree. Alignment was performed using the default settings of MAFFT [44], and an unrooted maximum-liklihood phylogenetic tree was produced using 80 sequences. The tree was drawn using MEGA7 [45] and iTOL [46] and the final tree was based on 35 ungapped amino acid positions and 100 resampling permutations. Branches with bootstrap support >50% are labelled with a black circle.

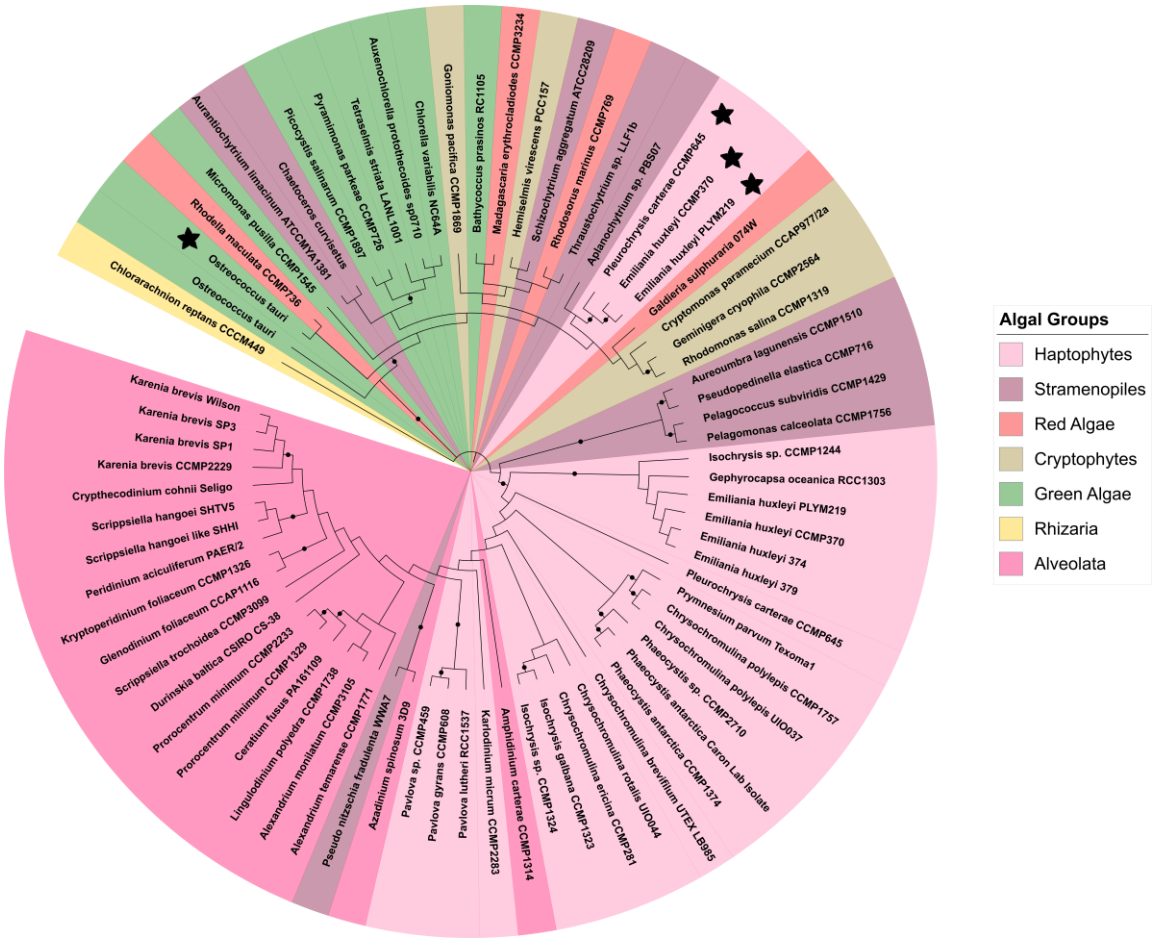


Figure 52 - Phylogenetic clustering of CMP-transferases across the algal groups. Sequences with high homology to *P. parvum* CMP-KDN synthase (CAMPEP_0191219004) can be found on the bottom half of the tree. Sequences with higher homology to *P. parkae* CMP-transferase (CAMPEP_0191478336) are found on the top half of the tree. Organisms with a star symbol contain two sequences in the tree. Alignment was performed using the default settings of MAFFT [44], and an unrooted maximum-liklihood phylogenetic tree was produced using 74 sequences. The tree was drawn using MEGA7 [45] and iTOL [46] and the final tree was based on 66 ungapped amino acid positions and 100 resampling permutations. Branches with bootstrap support >50% are labelled with a black circle.

4.4 Discussion

Whilst sialic acid biosynthesis has been explored in many kingdoms of life [1], the biosynthesis (or even presence) of sialic acids in algae is largely unknown. Here we provide the first biochemical and bioinformatic evidence that the toxin-producing microalgae, *Prymnesium parvum,* produces the de-aminated sialic acid, KDN, *de novo*. Liquid chromatography and mass spectrometry highlighted the presence of KDN in 15 different *Prymnesium* strains and independent LC-MS analysis identified CMP-KDN as the activated form of the sugar. This was supported by identification of sequences in a publicly available transcriptome *P. parvum* (MMETSP - Texoma1) which have high sequence identity to known KDN-9-P synthase and CMP-transferase genes. The corresponding proteins were successfully expressed in *E. coli* and synthesis of CMP-KDN from Man-6-P was shown. Using these sequences, 160 algal nucleic acid databases from MMETSP [31] or NCBI [42] were screened for sialic acid biosynthesis machinery. Sequences with homology to the structurally related KDO-8-P synthase were found scattered across the algal groups (Figure 50); sequences with similarity to the KDN-9-P synthase from *P. parvum* were abundant amongst the haptophytes and dinoflagellates. These bioinformatic findings are consistent with the little that is currently known about sialic acids or KDO in algae; KDO has been observed in the green alga *Tetraselmis striata* [47] and KDN observed in the haptophyte *Emiliania huxleyi* [20].

Although a clear homolog of KDN-9-P phosphatase couldn't be found, the pathway is presumed to follow a phosphorylation/dephosphorylation pathway such as seen in *B. theta* [18] (Figure 38) as the KDN-9-P synthase does not convert mannose to KDN. This lack of KDN-9-P phosphatase homolog is not surprising however; the functional redundancy of sugar phosphatases has been reported previously [48]. What is more surprising is the micromolar value observed for K_m with respect to KDN-9-P synthase and its substrate, Man-6-P (Table 7). This low Michaelis Constant suggests that the sialic acid synthase from *P. parvum* has a higher affinity for its substrate than other previously studied sialic acid synthases do for their substrates, such as those observed in Table 7. This may in part be explained by the specificity *P. parvum* KDN-9-P synthase shows for Man-6-P (Table 6); a trait that is not observed in other sialic acid synthases that show some level of promiscuity [49]. Also surprising is the higher k_{cat} value observed for this enzyme (14.91 min^{-1}); most enzymes in this family have k_{cat} values closer to 1 min^{-1} [18]. These ambiguities warrant further investigation.

Host Organism	Substrate	K_m [mM]
P. parvum (this study)	Man-6-P	0.065
B. thetaiotaomicron [18]	Man-6-P	1.4
H. sapiens [49]	Man-6-P	2.62
H. sapiens [49]	ManNAc-6-P	1.04
N. meningitidis [49]	ManNAc	11.6
C. jejuni [50]	ManNAc	17.6

Table 7 - K_m values for sugar substrates used by KDN-9-P synthase in this study, and other previously characterized sialic acid synthase enzymes.

Using the newly discovered sequences from *P.* parvum, 160 nucleic acid databases of other microalgae where screened for sialic acid or KDO biosynthetic machinery. Sequences with homology to genes involved in KDO biosynthesis were found spread across the algal groups, whilst sialic acid synthase homologs were found to be more confined to the Haptophyta and Alveolata phlya. Although there are exceptions, the biosynthesis of KDO and sialic acids appears to be mutually exclusive, suggesting these compounds may fill the role of one another; it has previously been suggested that genes involved in sialic acid biosynthesis have been "reinvented" from the KDO pathway, which share common ancestral genetic origins [1]. Some haptophytes contain both sialic acid and KDO biosynthetic machinery which may be a result of the extensive endosymbiotic gene transfer that has occurred within the algal groups [51]. Phylogenetic trees created to study the evolutionary distance between KDO/sialic acid synthases and CMP-transferases show similar traits. Both trees show two main groups of sequences, showing differences between the KDO and sialic acid pathways. One extra group exists for the sialic acid synthase tree (Figure 51) which contains solely haptophyte sequences. This group is more closely related to the KDO group of sequences, which may suggest these sequences are involved in the biosynthesis of alternative C8 or C7 structurally related 2-keto-3-deoxyoctonic or heptonic acids [2].

The importance of sialic acids in host:pathogen interactions is well documented [52], with a key example being the binding of the human Influenza virus to Neu5Ac residues on human epithelial cells [4]. More recently, viruses have been shown to play crucial roles in the population dynamics of algae [53], where they influence both ecosystem dynamics [34] and biogeochemical cycles [35]. The *Emiliania huxleyi* viruses (EhV) are probably the best studied of the algal viruses [54], and recently Fulton et al discovered key roles for a novel host sialic acid

lipid in viral infection of this alga [19, 20]. Although KDO had been reported in algae previously, C9 sialic acids were unknown to this group of organisms prior to this work. Our work conclusively demonstrates that the toxin producing microalgae, *P. parvum*, is capable of *de novo* KDN biosynthesis. The discovery of these new sequences was used to show that sialic acid biosynthesis occurs extensively throughout the haptophytes and dinoflagellates. This work was also able to highlight the distribution of biosynthetic genes for the structurally related KDO, across the algal groups. The widespread occurrence of these molecules and the ever-increasing number of algal viruses being discovered suggests that sialic acids may have key roles to play in early eukaryote host-pathogen interactions.

4.5 Materials and methods

4.5.1 Growth and maintenance of *Prymnesium* cultures

Cultures of *Prymnesium* were grown in f/2 -Si medium (7-8 practical salinity units) at 22 °C on a 14:10 light cycle as previously described [32]. Stock cultures were made axenic by treatment with Carbenicillin (100 µg/ml) as judged by optical microscopy. For metabolomic extractions, cells were harvested at late-log phase of growth which often represented ~2 – 3 x 10^6 cells ml^{-1}. A full list of strains used in this study can be seen below (Table 8).

Strain Identifier	Genus	Species	Date Collected	Isolated by	Area Collected
94A	*Prymnesium*	*parvum*	1969	R. Waters	Pacific: NE
94C	*Prymnesium*	*parvum*	Jan 1978	J.C. Green	Mediterranean Sea
522	*Prymnesium*	sp.			
527C	*Prymnesium*	*patelliferum*	1970	R. Waters	Pacific: NE
527A	*Prymnesium*	*patelliferum*	Aug 69	R. Waters	Pacific: NE
527D	*Prymnesium*	*patelliferum*		U. Tillmann	North Sea
569	*Prymnesium*	sp.	Jan 78	J.C. Green	Mediterranean Sea; Libya, coastal
579	*Prymnesium*	*parvum*			Irish Sea; Menai Straits
592	*Prymnesium*	sp.	Jul 78	Hällfors	Baltis Sea, Finland
593	*Prymnesium*	sp.	Jul 78	Hällfors	Baltis Sea, Finland
595	*Prymnesium*	sp.	Jul 78	Hällfors	Baltic Sea, Finland
596	*Prymnesium*	sp.	Aug 78	Hällfors	Baltic Sea, Finland
597	*Prymnesium*	sp.	Jul 78	Hällfors	Baltic Sea, Finland
598	*Prymnesium*	sp.	Jul 78	S & G Hällfors	Baltic Sea, Finland
946/6	*Prymnesium*	*parvum*	1953	Droop	Isle of Cumbrae, Scotland, UK

Table 8- A complete list of strains of *Prymnesium* used in this study.

4.5.2 DMB-HPLC analysis of sialic acids

Free sialic acids were liberated from cells by re-suspending 10 ml of pelleted *Prymnesium* cells in hydrochloric acid (200 µl, 0.1 M) at 80 °C for 1 hour. Insoluble cell debris was then pelleted by centrifugation and soluble material as well as standards for KDN (Sigma Aldrich, Haverhill, UK) and Neu5Ac were then derivatized with DMB per the manufacturer's instructions (Sialic acid fluorescence labelling kit, Clontech, Takara Bio Europe SAS, Saint-Germain-en-Laye, France). Control reactions were set up in parallel with HCl to act as a blank background.

Derivatized samples (7 µl) were then injected into a Nexera/Prominence UHPLC system (Shimadzu) equipped with an in-line fluorescence detector (excitation: 373 nm, emission: 448 nm) and a LCMS-2020 single quadrupole MS set to scan between 100 – 600 m/z. Samples were ran on a Kinetex 2.6 µm EVO C18 100 Å LC column (50 x 2.1 mm). Buffer A was an aqueous MeOH, H_2O buffer (7:93) and buffer B was an organic AcN, MeOH buffer (93:7). The gradient was set up so that fluorescent DMB-sialic acid adducts eluted within the first 10 mins and was set as follows: 0-10 min 5% B, 10-12.5 min 100% B, 12.5-15 min 5% B, 15-22 min 5% B. The samples were ran at a flow rate of 0.2 ml min^{-1} and peaks observed were compared to retention times of KDN, KDO and Neu5Ac standards. Previously reported mass values and relative retention times for sialic acids and their *O*-acetyl derivatives were also used for comparison [37].

4.5.3 Sugar nucleotide profiling of CMP-KDN

Intracellular sugar nucleotides were extracted and analysed using previously published methods [41]. CMP-KDN was produced using commercial KDN (Sigma Aldrich, Haverhill, United Kingdom) and recombinant CMP-transferase produced in this study to act as an internal standard. Details of the enzymatic reaction conditions can be found in section 4.5.8 of this chapter. Other sugar nucleotides used in this study for comparison were UDP-α-D-Glc, UDP-α-D-Gal, and GDP-α-D-Glc as an internal standard for quantification. Internal standard was used to find relative retention times and MRM transitions for CMP-KDN (Table 5). Co-injection of cell extracts with internal standards confirmed identity of peaks.

4.5.4 Identification of putative CMP-KDN biosynthesis transcripts in *P. parvum*

For the identification of transcripts involved in sialic acid biosynthesis from *P. parvum*, BLASTp [55] analysis was carried out against a publicly available transcriptome of *P. parvum* (Texoma1 – Marine Microbial Eukaryote Transcriptome Sequencing Project, MMETSP [31]). Protein sequences involved in CMP-KDN biosynthesis from *Bacteroidetes thetaiotaomicron* VPI-5482 [18] were used as consensus sequences (ExPASy accession numbers: Q8A712, Q8A711 and Q8A710). Any hits with *E*-values ≤ $1E^{-10}$ were then analysed manually for conserved domains before being assigned as a hit. Hits were subsequently analysed for transmembrane domains and signal peptides using the online tool Phobius [56] (http://phobius.sbc.su.se/).

4.5.5 Cloning and expression of *P. parvum* KDN-9-P synthase and CMP-transferase

Transcripts identified with high sequence similarity to KDN-9-P synthase (CAMPEP_0191217894) and CMP-transferase (CAMPEP_0191219004), along with KDN-9-P phosphatase from *B. thetaiotaomicron* VPI-5482 (ExPASy accession number: Q8A712), were codon optimized for expression in *E. coli* using IDTdna's codon optimization software (https://www.idtdna.com/CodonOpt). The resulting sequences were then synthesized with overhangs for In-Fusion™ cloning into pOPINF vector [57] using IDT's gBlock gene fragment synthesis service. A full list of sequences used in this study can be found in Supplementary Material 2.

gBlock fragments were then cloned into pOPINF vectors using In-Fusion™ cloning kit (Clontech, UK) per the manufacturer's instructions. The resulting plasmids were then transformed into Stellar competent cells, before being propagated and extracted (Miniprep kit, Qiagen, UK). Positively transformed plasmids were identified by size comparison to a non-transformed pOPINF control using gel electrophoresis. Plasmids containing the gBlock sequences were then finally transformed into BL21 CodonPlus™ (Agilent Technologies, Fisher Scientific, UK) for protein expression. For all proteins, 1L of *E. coli* cells were grown to an OD of 0.6 at 37 ^{o}C before being transferred to 18 ^{o}C for 1 hour. Induction was performed with 0.5 mM of IPTG and cells were left at 18 ^{o}C overnight. Proteins were extracted in the following buffer: 50 mM Tris-HCl pH 7.5, 0.5 M NaCl, 20 mM imidazole, protease inhibitor cocktail (Sigma) 1/100 v/v, 2 mg DNase. KDN-9-P synthase also required addition of 5 mM DTT to prevent precipitation of the enzyme due to oxidation. All proteins were purified using Nickel affinity chromatography, and fractions judged to be >95% pure by SDS-PAGE were pooled for further analysis.

4.5.6 CAMPEP_0191217894 encodes a functional KDN-9-P synthase

Reaction mixtures contained 38 µg of CAMPEP_0191217894, 2 mM $MgCl_2$, 8 mM of phosphoenolpyruvate (PEP) and 10 mM of mannose-6-phosphate, buffered in 50 mM HEPES (pD 7.5). A no-enzyme control time point was taken prior to addition of enzyme. Once enzymes were added, time points were recorded for ^{1}H and ^{31}P NMR. ^{1}H NMR signals representing H_3-eq (~2.1 ppm) and H_3-ax (~1.7 ppm) were monitored over the time course for formation of KDN-9-phosphate. ^{31}P NMR was used to monitor reaction progress - loss of phosphate signal at -1 ppm representing PEP showed reaction completion. Once the reaction had reached completion, the resulting mixture was analysed by ESI-MS for the presence of

KDN-9-P mass. HRMS, ESI negative: calculated: 347.0385, found: 347.0394. A range of other sugars were screened as substrates using the same methodology. All sugar substrates were purchased from Sigma Aldrich UK or Carbosynth.

To establish kinetic parameters, a colorimetric phosphate-release assay was employed (BioMol Green, Enzo). This assay followed reaction progress by monitoring realise of the bi-product of the reaction, inorganic phosphate. Reactions mixtures included a fixed concentration of one substrate (100 µM) whilst varying the other (3.125 µM, 6.25 µM, 12.5 µM, 25 µM, 50 µM, 100 µM, 200 µM) in a buffered solution containing 50 mM HEPES (pH 7.5), 2 mM $MgCl_2$ and CAMPEP_0191217894 (Final conc = 4.725 µg/ml). The steady state kinetic values determined at 25 ^oC using this method are k_{cat} = 14.91 min^{-1}, V_{max} = 30.41 ± 1.94 pmol min^{-1}, $^{(PEP)}K_m$ = 18.31 ± 1.83 µM (at 100 µM mannose-6-P) and $^{(mannose\text{-}6\text{-}P)}K_m$ = 63.63 ± 10.51 µM (at 100 µM PEP) (Figure 47). Values and graphs were produced using GraphPad Prism (Version 7).

4.5.7 KDN-9-P phosphatase accepts the product of CAMPEP_0191217894

A coupled assay was employed to confirm that the product of the KDN-9-P synthase enzyme was accepted by *B. thetaiotaomicron* VPI-5482 KDN-9-P phosphatase. Reaction conditions were essentially the same as described in section 4.5.6, but an excess of PEP was used with respect to Man-6-P (10 mM and 8 mM, respectively). A no enzyme ^{31}P NMR time point was acquired before 38 µg of CAMPEP_0191217894 (KDN-9-P synthase) was added. The reaction was left for 3 hours at which point formation of KDN-9-P and inorganic phosphate was observed (Figure 48). KDN-9-P phosphatase from *B. thetaiotaomicron* VPI-5482 was then added to the reaction mixture (4 µg) and the reaction allowed to progress a further 3 hours.

4.5.8 CAMPEP_0191219004 encodes a functional CMP-transferase

Reaction mixtures contained 20 µg of CAMPEP_0191219004, 10 mM $MgCl_2$, 8 mM of cytidine triphosphate (CTP) and 10 mM of commercial KDN (Sigma Aldrich, UK) buffered in 100 mM TRIS-HCl (pH 9). A ^{31}P NMR no enzyme time point was taken before addition of CAMPEP_0191219004. The reaction was seen to go to completion when signals for CTP (-5.83 ppm, doublet, -10.75 ppm, doublet, -19.38 ppm, triplet) had diminished. At this point, CMP-KDN was purified using strong anion exchange as outlined in Wagstaff *et al* [58]. ^1H NMR, ^{31}P NMR and MS values aligned well with previously described literature [18, 59] (Figure 53). ^1H

NMR (D_2O) ppm: 7.9 (d, *J* = 7.5 Hz, 1H), 6.04 (d, *J* = 7.5 Hz, 1H), 5.91 (d, *J* = 4.5 Hz, 1H), 4.27 (m, 1H), 4.22 (t, *J* = 4.8 Hz, 1H), 4.16 (d, *J* = 5.3 Hz, 3H), 4.01 (dd, *J* = 10 Hz, *J* = 1 Hz, 1H), 3.95 (ddd, *J* = 11.5 Hz, *J* = 9.4 Hz, *J* = 4.9 Hz, 1H), 3.85 (m, 2H), 3.72 – 3.45 (m, 16H, buffer protons and H-5,7,8 of KDN), 2.36 (dd, *J* = 13 Hz, 1H), 1.52 (ddd, *J* = 14 Hz, *J* = 13.5 Hz, *J* = 5.7, 1H). ^{31}P NMR (H_2O) ppm: -4.61 (s). HRMS, ESI negative: calculated: 572.1129, found: 572.1135.

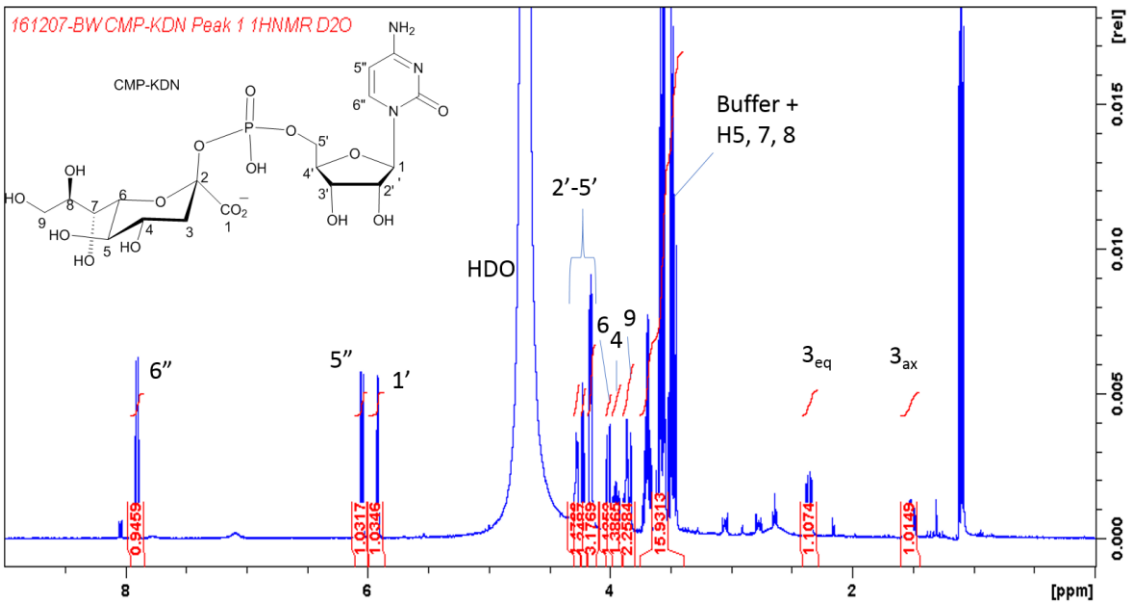


Figure 53 – ^1H NMR spectrum of CMP-KDN with proton assignments. Impurities are from residual buffer that eluted off the SAX column. Only protons from C5, 7 and 8 are covered by buffer signals.

4.5.9 Identification of sialic acid synthase and CMP-transferase genes across the algal groups

KDN-9-P synthase (CAMPEP_0191217894) and CMP-transferase (CAMPEP_0191219004) protein sequences from *P. parvum* Texoma1 were used as consensus sequences in BLASTp analysis of other algal transcriptomes from the MMETSP database [31]. The green alga, *Pyrammimonas parkae*, was found to have low sequence identity homologues to both *Prymnesium* transcripts, instead displaying high sequence similarity to a characterized KDO-8-P synthase from *Haemophilus influenzae* and CMP-KDO synthase from *Pseudomonas aeruginosa* (44% sequence identity - PDB: 1O60_A and 41% sequence identity - PDB: 4XWI_A respectively). These new transcripts were used in addition to *P. parvum*'s to query a total of 160 nucleic acid databases from other algal groups. A full list of databases used can be found in Supplementary Material 1. Sequences displaying *E*-values ≤ $1E^{-10}$ were subject to manual inspection of conserved domains before being assigned as a hit.

4.5.10 Phylogenetic analysis of algal sialic acid biosynthesis machinery

Two independent sets of algal sequences identified in BLASTp analysis (sialic acid synthases and CMP-transferases) were aligned using the default settings of MAFFT [44]. Maximum-likelihood phylogenetic trees were constructed using MEGA7 [45] and drawn using iTOL [46].

4.6 References

1. Varki, A.; Schauer, R., Sialic acids. In *Essentials of Glycobiology*, Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor Laboratory Press The Consortium of Glycobiology Editors, La Jolla, California.: Cold Spring Harbor (NY), **2009**.
2. Angata, T.; Varki, A., Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. *Chem. Rev.* **2002,** 102, (2), 439-69.
3. Schauer, R., Sialic acids as regulators of molecular and cellular interactions. *Curr. Opin. Struct. Biol.* **2009,** 19, (5), 507-514.
4. Stencel-Baerenwald, J. E.; Reiss, K.; Reiter, D. M.; Stehle, T.; Dermody, T. S., The sweet spot: defining virus-sialic acid interactions. *Nat. Rev. Micro.* **2014,** 12, (11), 739-749.
5. Gagneux, P.; Varki, A., Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology* **1999,** 9, (8), 747-55.
6. Blix, G., Über die Kohlenhydratgruppen des Submaxillarismucins. *Hoppe-Seyler´ s Zeitschrift für physiologische Chemie* **1936,** 240, (1-2), 43-54.
7. Nadano, D.; Iwasaki, M.; Endo, S.; Kitajima, K.; Inoue, S.; Inoue, Y., A naturally occurring deaminated neuraminic acid, 3-deoxy-D-glycero-D-galacto-nonulosonic acid (KDN). Its unique occurrence at the nonreducing ends of oligosialyl chains in polysialoglycoprotein of rainbow trout eggs. *J. Biol. Chem.* **1986,** 261, (25), 11550-11557.
8. Iwasaki, M.; Inoue, S.; Troy, F. A., A new sialic acid analogue, 9-*O*-acetyl-deaminated neuraminic acid, and alpha -2,8-linked *O*-acetylated poly(N-glycolylneuraminyl) chains in a novel polysialoglycoprotein from salmon eggs. *J. Biol. Chem.* **1990,** 265, (5), 2596-2602.
9. Strecker, G.; Wieruszeski, J.-M.; Michalski, J.-C.; Alonso, C.; Boilly, B.; Montreuil, J., Characterization of Lex, Ley and A Ley antigen determinants in KDN-containing *O*-linked glycan chains from *Pleurodeles waltlii* jelly coat eggs. *FEBS Lett.* **1992,** 298, (1), 39-43.
10. Strecker, G.; Wieruszeski, J.-M.; Michalski, J.-C.; Alonso, C.; Leroy, Y.; Boilly, B.; Montreuil, J., Primary structure of neutral and acidic oligosaccharide-alditols derived from the jelly coat of the *Mexican axolotl*. *Eur. J. Biochem.* **1992,** 207, (3), 995-1002.
11. Plancke, Y.; Wieruszeski, J.-M.; Alonso, C.; Boilly, B.; Strecker, G., Structure of four acidic oligosaccharides from the jelly coat surrounding the eggs of *Xenopus laevis*. *Eur. J. Biochem.* **1995,** 231, (2), 434-439.
12. Kimura, M.; Hama, Y.; Sumi, T.; Asakawa, M.; Rao, B. N.; Horne, A. P.; Li, S. C.; Li, Y. T.; Nakagawa, H., Characterization of a deaminated neuraminic acid-containing

glycoprotein from the skin mucus of the loach, *Misgurnus anguillicaudatus*. *J. Biol. Chem.* **1994,** 269, (51), 32138-43.

13. Knirel, Y. A.; Kocharova, N. A.; Shashkov, A. S.; Kochetkov, N. K.; Mamontova, V. A.; Solov'eva, T. F., Structure of the capsular polysaccharide of *Klebsiella ozaenae* serotype K4 containing 3-deoxy-D-glycero-D-galacto-nonulosonic acid. *Carbohydr. Res.* **1989,** 188, 145-155.
14. Shashkov, A. S.; Tul'skaya, E. M.; Evtushenko, L. I.; Denisenko, V. A.; Ivanyuk, V. G.; Stomakhin, A. A.; Naumova, I. B.; Stackebrandt, E., Cell wall anionic polymers of *Streptomyces* sp. MB-8, the causative agent of potato scab. *Carbohydr. Res.* **2002,** 337, (21–23), 2255-2261.
15. Inoue, S.; Kitajima, K.; Inoue, Y., Identification of 2-keto-3-deoxy-D-glycero-D-galactonononic acid (KDN, Deaminoneuraminic Acid) residues in mammalian tissues and human lung carcinoma cells: chemical evidence of the occurrence of KDN glycoconjugates in mammals. *J. Biol. Chem.* **1996,** 271, (40), 24341-24344.
16. Inoue, S.; Lin, S.-L.; Chang, T.; Wu, S.-H.; Yao, C.-W.; Chu, T.-Y.; Troy, F. A.; Inoue, Y., Identification of free deaminated sialic acid (2-keto-3-deoxy-D-glycero-D-galacto-nononic Acid) in human red blood cells and its elevated expression in fetal cord red blood cells and ovarian cancer cells. *J. Biol. Chem.* **1998,** 273, (42), 27199-27204.
17. Hao, J.; Vann, Willie F.; Hinderlich, S.; Sundaramoorthy, M., Elimination of 2-keto-3-deoxy-D-glycero-D-galacto-nonulosonic acid 9-phosphate synthase activity from human *N*-acetylneuraminic acid 9-phosphate synthase by a single mutation. *Biochem. J* **2006,** 397, (Pt 1), 195-201.
18. Wang, L.; Lu, Z.; Allen, K. N.; Mariano, P. S.; Dunaway-Mariano, D., Human symbiont *Bacteroides thetaiotaomicron* synthesizes 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid (KDN). *Chem. Biol.* **2008,** 15, (9), 893-897.
19. Rose, S. L.; Fulton, J. M.; Brown, C. M.; Natale, F.; Van Mooy, B. A. S.; Bidle, K. D., Isolation and characterization of lipid rafts in *Emiliania huxleyi*: a role for membrane microdomains in host–virus interactions. *Environ. Microbiol.* **2014,** 16, (4), 1150-1166.
20. Fulton, J. M.; Fredricks, H. F.; Bidle, K. D.; Vardi, A.; Kendrick, B. J.; DiTullio, G. R.; Van Mooy, B. A. S., Novel molecular determinants of viral susceptibility and resistance in the lipidome of *Emiliania huxleyi*. *Environ. Microbiol.* **2014,** 16, (4), 1137-1149.
21. Schoemann, V.; Becquevort, S.; Stefels, J.; Rousseau, V.; Lancelot, C., Phaeocystis blooms in the global ocean and their controlling mechanisms: a review. *J. Sea Res.* **2005,** 53, (1–2), 43-66.

22. Leblanc, K.; Hare, C. E.; Feng, Y.; Berg, G. M.; DiTullio, G. R.; Neeley, A.; Benner, I.; Sprengel, C.; Beck, A.; Sanudo-Wilhelmy, S. A.; Passow, U.; Klinck, K.; Rowe, J. M.; Wilhelm, S. W.; Brown, C. W.; Hutchins, D. A., Distribution of calcifying and silicifying phytoplankton in relation to environmental and biogeochemical parameters during the late stages of the 2005 North East Atlantic Spring Bloom. *Biogeosciences* **2009,** 6, (10), 2155-2179.

23. Edvardsen, B.; Paasche, E., Bloom dynamics and physiology of Prymnesium and Chrysochromulina. In: Anderson, D. M.; Cembella, A. D.; Hallegraeff, G. M., editors. Physiological Ecology of Harmful Algal Blooms. (NATO ASI Series G, vol. 41). Springer-Verlag, Berlin Heidelberg: **1998**, 192-208.

24. Roelke, D. L.; Barkoh, A.; Brooks, B. W.; Grover, J. P.; Hambright, K. D.; LaClaire, J. W.; Moeller, P. D. R.; Patino, R., A chronicle of a killer alga in the west: ecology, assessment, and management of *Prymnesium parvum* blooms. *Hydrobiologia* **2016,** 764, (1), 29-50.

25. Yariv, J.; Hestrin, S., Toxicity of the extracellular phase of *Prymnesium parvum* cultures. *J. Gen. Microbiol.* **1961,** 24, (2), 165-175.

26. Kaartvedt, S.; Johnsen, T. M.; Aksnes, D. L.; Lie, U.; Svendsen, H., Occurrence of the toxic phytoflagellate *Prymnesium parvum* and associated fish mortality in a Norwegian fjord system. *Can. J. Fish. Aquat. Sci.* **1991,** 48, (12), 2316-2323.

27. Guo, M.; Harrison, P. J.; Taylor, F. J. R., Fish kills related to *Prymnesium parvum* N. Carter (Haptophyta) in the People's Republic of China. *J. Appl. Phycol.* **1996,** 8, (2), 111-117.

28. Southard, G. M.; Fries, L. T.; Barkoh, A., *Prymnesium parvum*: The Texas Experience. *J. Am. Water. Resour. Assoc.* **2010,** 46, (1), 14-23.

29. Granéli, E.; Edvardsen, B.; Roelke, D. L.; Hagström, J. A., The ecophysiology and bloom dynamics of *Prymnesium* spp. *Harmful Algae* **2012,** 14, 260-270.

30. Janouškovec, J.; Liu, S.-L.; Martone, P. T.; Carré, W.; Leblanc, C.; Collén, J.; Keeling, P. J., Evolution of red algal plastid genomes: ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers. *PLOS ONE* **2013,** 8, (3), e59001.

31. Keeling, P. J.; Burki, F.; Wilcox, H. M.; Allam, B.; Allen, E. E.; Amaral-Zettler, L. A.; Armbrust, E. V.; Archibald, J. M.; Bharti, A. K.; Bell, C. J.; Beszteri, B.; Bidle, K. D.; Cameron, C. T.; Campbell, L.; Caron, D. A.; Cattolico, R. A.; Collier, J. L.; Coyne, K.; Davy, S. K.; Deschamps, P.; Dyhrman, S. T.; Edvardsen, B.; Gates, R. D.; Gobler, C. J.; Greenwood, S. J.; Guida, S. M.; Jacobi, J. L.; Jakobsen, K. S.; James, E. R.; Jenkins, B.; John, U.; Johnson, M. D.; Juhl, A. R.; Kamp, A.; Katz, L. A.; Kiene, R.; Kudryavtsev, A.; Leander, B. S.; Lin, S.; Lovejoy, C.; Lynn, D.; Marchetti, A.; McManus, G.; Nedelcu, A. M.; Menden-Deuer, S.; Miceli, C.; Mock, T.; Montresor, M.; Moran, M. A.; Murray, S.;

Nadathur, G.; Nagai, S.; Ngam, P. B.; Palenik, B.; Pawlowski, J.; Petroni, G.; Piganeau, G.; Posewitz, M. C.; Rengefors, K.; Romano, G.; Rumpho, M. E.; Rynearson, T.; Schilling, K. B.; Schroeder, D. C.; Simpson, A. G. B.; Slamovits, C. H.; Smith, D. R.; Smith, G. J.; Smith, S. R.; Sosik, H. M.; Stief, P.; Theriot, E.; Twary, S. N.; Umale, P. E.; Vaulot, D.; Wawrik, B.; Wheeler, G. L.; Wilson, W. H.; Xu, Y.; Zingone, A.; Worden, A. Z., The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* **2014,** 12, (6), e1001889.

32. Wagstaff, B.; Vladu, I.; Barclay, J.; Schroeder, D.; Malin, G.; Field, R., Isolation and characterization of a double stranded DNA megavirus infecting the toxin-producing haptophyte *Prymnesium parvum*. *Viruses* **2017,** 9, (3), 40.
33. La Scola, B.; Audic, S.; Robert, C.; Jungang, L.; de Lamballerie, X.; Drancourt, M.; Birtles, R.; Claverie, J. M.; Raoult, D., A giant virus in amoebae. *Science (New York, N.Y.)* **2003,** 299, (5615), 2033.
34. Short, S. M., The ecology of viruses that infect eukaryotic algae. *Environ. Microbiol.* **2012,** 14, (9), 2253-71.
35. Suttle, C. A., Marine viruses - major players in the global ecosystem. *Nat. Rev. Micro.* **2007,** 5, (10), 801-812.
36. Schatz, D.; Shemi, A.; Rosenwasser, S.; Sabanay, H.; Wolf, S. G.; Ben-Dor, S.; Vardi, A., Hijacking of an autophagy-like process is critical for the life cycle of a DNA virus infecting oceanic algal blooms. *New Phytol.* **2014,** 204, (4), 854-63.
37. Klein, A.; Diaz, S.; Ferreira, I.; Lamblin, G.; Roussel, P.; Manzi, A. E., New sialic acids from biological sources identified by a comprehensive and sensitive approach: liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS) of SIA quinoxalinones. *Glycobiology* **1997,** 7, (3), 421-432.
38. Turnock, D. C.; Ferguson, M. A., Sugar nucleotide pools of *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania major*. *Eukaryot. Cell* **2007,** 6, (8), 1450-1463.
39. Räbinä, J.; Mäki, M.; Savilahti, E. M.; Järvinen, N.; Penttilä, L.; Renkonen, R., Analysis of nucleotide sugars from cell lysates by ion-pair solid-phase extraction and reversed-phase high-performance liquid chromatography. *Glycoconjugate J.* **2001,** 18, (10), 799-805.
40. Pabst, M.; Grass, J.; Fischl, R.; Léonard, R.; Jin, C.; Hinterkörner, G.; Borth, N.; Altmann, F., Nucleotide and nucleotide sugar analysis by liquid chromatography-electrospray ionization-mass spectrometry on surface-conditioned porous graphitic carbon. *Anal. Chem.* **2010,** 82, (23), 9782-9788.

41. Rejzek, M.; Hill, L.; Hems, E. S.; Kuhaudomlarp, S.; Wagstaff, B. A.; Field, R. A., Profiling of sugar nucleotides. *Methods Enzymol.* **2017**, 597, 209-39.

42. Coordinators., N. R., Database resources of the National center for biotechnology information. *Nucleic Acids Res.* **2017,** 45, (D1), D12-d17.

43. Radaev, S.; Dastidar, P.; Patel, M.; Woodard, R.W. and Gatti, D.L., Structure and mechanism of 3-deoxy-D-manno-octulosonate 8-phosphate synthase. *J. Biol. Chem* **2000**, 275, (13), 9476-9484.

44. Katoh, K.; Toh, H., Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinformatics* **2008,** 9, (4), 286-98.

45. Kumar, S.; Stecher, G.; Tamura, K., MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016,** 33, 1870-1874.

46. Letunic, I.; Bork, P., Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **2016,** 44, (W1), W242-5.

47. Becker, D.; Becker, B.; Satir, P.; Melkonian, M., Isolation, purification, and characterization of flagellar scales from the green flagellate *Tetraselmis striata* (Prasinophyceae). *Protoplasma* **1990,** 156, (1), 103-112.

48. Daughtry, K. D.; Huang, H.; Malashkevich, V.; Patskovsky, Y.; Liu, W.; Ramagopal, U.; Sauder, J. M.; Burley, S. K.; Almo, S. C.; Dunaway-Mariano, D.; Allen, K. N., The structural basis for divergence of substrate specificity and biological function within HAD phosphatases in lipopolysaccharide and sialic acid biosynthesis. *Biochemistry* **2013,** 52, (32), 5372-5386.

49. Hao, J.; Balagurumoorthy, P.; Sarilla, S.; Sundaramoorthy, M., Cloning, expression, and characterization of sialic acid synthases. *Biochem. Biophys. Res. Commun.* **2005,** 338, (3), 1507-14.

50. Sundaram, Appavu K.; Pitts, L.; Muhammad, K.; Wu, J.; Betenbaugh, M.; Woodard, Ronald W.; Vann, Willie F., Characterization of *N*-acetylneuraminic acid synthase isoenzyme 1 from *Campylobacter jejuni*. *Biochem. J* **2004,** 383, (Pt 1), 83-89.

51. Keeling, P. J., The endosymbiotic origin, diversification and fate of plastids. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **2010,** 365, (1541), 729-748.

52. Varki, A.; Gagneux, P., Multifarious roles of sialic acids in immunity. *Ann. N.Y. Acad. Sci.* **2012,** 1253, (1), 16-36.

53. Wilhelm, S. W.; Coy, S. R.; Gann, E. R.; Moniruzzaman, M.; Stough, J. M. A., Standing on the shoulders of giant viruses: five lessons learned about large viruses infecting small eukaryotes and the opportunities they create. *PLoS Path.* **2016,** 12, (8), e1005752.

54. Wilson, W. H.; Tarran, G. A.; Schroeder, D.; Cox, M.; Oke, J.; Malin, G., Isolation of viruses responsible for the demise of an *Emiliania huxleyi* bloom in the English Channel. *J. Mar. Biol. Assoc. U.K.* **2002,** 82, (3), 369-377.

55. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic local alignment search tool. *J. Mol. Biol.* **1990,** 215, (3), 403-10.

56. Kall, L.; Krogh, A.; Sonnhammer, E. L., A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **2004,** 338, (5), 1027-36.

57. Berrow, N. S.; Alderton, D.; Sainsbury, S.; Nettleship, J.; Assenberg, R.; Rahman, N.; Stuart, D. I.; Owens, R. J., A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* **2007,** 35, (6), e45-e45.

58. Wagstaff, B. A.; Rejzek, M.; Pesnot, T.; Tedaldi, L. M.; Caputi, L.; O'Neill, E. C.; Benini, S.; Wagner, G. K.; Field, R. A., Enzymatic synthesis of nucleobase-modified UDP-sugars: scope and limitations. *Carbohydr. Res.* **2015,** 404, (Supplement C), 17-25.

59. Terada, T.; Kitazume, S.; Kitajima, K.; Inoue, S.; Ito, F.; Troy, F. A.; Inoue, Y., Synthesis of CMP-deaminoneuraminic acid (CMP-KDN) using the CTP:CMP-3-deoxynonulosonate cytidylyltransferase from rainbow trout testis. Identification and characterization of a CMP-KDN synthetase. *J. Biol. Chem.* **1993,** 268, (4), 2640-8.

4.7 Supplementary Material

Kingdom	Phylum	Class	Family	Genus	Species	Strain	Database and identifier	Sequence
GLAUCOPHYTES	Glaucophyta	Glaucocystophyceae	Glaucosphaeraceae	Cyanoptyche	gloeocystis	SAG4.97	MMETSP1086	
	Glaucophyta	Glaucophyceae	Gloeochaetaceae	Gloeochaete	wittrockiana	SAG46.84	MMETSP1089	
GREEN ALGAE	Chlorophyta	Chlorodendrophyceae	Chlorodendraceae	Tetraselmis	striata	LANL1001	MMETSP0817, MMETSP0818, MMETSP0819, MMETSP0820	
	Chlorophyta	Chlorophyceae	Dunaliellaceae	Dunaliella	tertiolecta	CCMP1320	MMETSP1126, MMETSP1127, MMETSP1128	
	Chlorophyta	Chlorophyceae	Chlamydomonadaceae	Chlamydomonas	reinhardtii	CC-503 cw92 mt+	https://www.ncbi.nlm.nih.gov/nuccore/158276217	
	Chlorophyta	Mamiellophyceae	Mamiellaceae	Micromonas	sp.	RCC472	MMETSP1084, MMETSP1387	
	Chlorophyta	Mamiellophyceae	Mamiellaceae	Micromonas	sp.	NEPCC29	MMETSP1386, MMETSP1082	
	Chlorophyta	Mamiellophyceae	Mamiellaceae	Micromonas	sp.	CCMP2099	MMETSP1390, MMETSP0802	
	Chlorophyta	Mamiellophyceae	Mamiellaceae	Micromonas	pusilla	CCMP1545	https://www.ncbi.nlm.nih.gov/genome/?term=txid564608[Organism:noexp]	XP_003062615.1
	Chlorophyta	Mamiellophyceae	Mamiellaceae	Bathycoccus	prasinos	RC1105	https://www.ncbi.nlm.nih.gov/genome/12309	XP_007509032.1, XP_007511142
	Chlorophyta	Mamiellophyceae	Mamiellaceae	Ostreococcus	tauri		https://www.ncbi.nlm.nih.gov/genome/373	OUS41656.1, XP_003083551.1
	Chlorophyta	Pyramimonadophyceae	Halosphaeraceae	Pyramimonas	parkeae	CCMP726	MMETSP0058, MMETSP0059	CAMPEP_0191499076, CAMPEP_0191478336

	Chlorophyta	Trebouxiophyceae	Chlorellaceae	Auxenochlorella	protothecoides	sp 0710	http://www.ncbi.nlm.nih.gov/genome?LinkName=nuccore_genome&from_uid=667612142	XP_011395670.1, XP_011396550.1
	Chlorophyta	Trebouxiophyceae	Chlorellaceae	Chlorella	variabilis	NC64A	https://www.ncbi.nlm.nih.gov/genome/694	XP_005845017.1, XP_005843072.1
	Chlorophyta	Unknown	Unknown	Picocystis	salinarum	CCMP1897	MMETSP1159, MMETSP0807	CAMPEP_0190741864, CAMPEP_0190748198
RED ALGAE	Rhodophyta	Compsopogonophyceae	Compsopogonaceae	Compsopogon	coeruleus	SAG 36.94	MMETSP0312	
	Rhodophyta	Compsopogonophyceae	Erythrotrichiaceae	Madagascaria	erythrocladioides	CCMP3234	MMETSP1450	CAMPEP_0198335420, CAMPEP_0198324100
	Rhodophyta	Cyanidiophyceae	Galdieriaceae	Galdieria	sulphuraria	074W	http://www.ncbi.nlm.nih.gov/genome/405	XP_005706679.1, XP_005706001.1
	Rhodophyta	Florideophyceae	Gigartinaceae	Chondrus	crispus	Stackhouse	http://www.ncbi.nlm.nih.gov/genome/12106	
	Rhodophyta	Porphyridiophyceae	Porphyridiaceae	Erythrolobus	australicus	CCMP3124	MMETSP1353	
	Rhodophyta	Porphyridiophyceae	Porphyridiaceae	Erythrolobus	madagascarensis	CCMP3276	MMETSP1354	
	Rhodophyta	Porphyridiophyceae	Porphyridiaceae	Timspurckia	oligopyrenoides	CCMP3278	MMETSP1172	CAMPEP_0182450842
	Rhodophyta	Rhodellophyceae	Porphyridiaceae	Porphyridium	aerugineum	SAG 1380-2	MMETSP0313	
	Rhodophyta	Rhodellophyceae	Rhodellaceae	Rhodella	maculata	CCMP736	MMETSP0167, MMETSP0314	CAMPEP_0191526874, CAMPEP_0191530618
	Rhodophyta	Rhodellophyceae	Stylonemataceae	Rhodosorus	marinus	CCMP 769	MMETSP0011	CAMPEP_0113968256, CAMPEP_0113970242
	Rhodophyta	Rhodellophyceae	Stylonemataceae	Rhodosorus	marinus	UTEX LB 2760	MMETSP0315	
EXCAVATES	Euglenozoa	Euglenophyceae	Eutreptiaceae	Eutreptiella	gymnastica	NIES-381	MMETSP0039	

	Euglenozoa	Euglenophyceae	Eutreptiaceae	Eutreptiella	gymnastica-like	CCMP1594	MMETSP0809, MMETSP0810, MMETSP0811	
	Euglenozoa	Euglenophyceae	Euglenaceae	Euglena	gracilis	??	??	
RHIZARIA	Cercozoa	Chlorarachniophyceae	Chlorarachniaceae	Chlorarachnion	reptans	CCCM449	MMETSP0109	CAMPEP_0114524360, CAMPEP_0114518112
	Cercozoa	Chlorarachniophyceae	Chlorarachniaceae	Gymnochlora	sp.	CCMP2014	MMETSP0110	
	Cercozoa	Chlorarachniophyceae	Chlorarachniaceae	Lotharella	oceanica	CCMP622	MMETSP0040	
	Cercozoa	Chlorarachniophyceae	Chlorarachniaceae	Lotharella	globosa	CCCM811	MMETSP0111, MMETSP0112	
	Cercozoa	Chlorarachniophyceae	Chlorarachniaceae	Bigelowiella	natans	CCMP623	MMETSP1052	
	Cercozoa	Chlorarachniophyceae	Chlorarachniaceae	Bigelowiella	natans	CCMP 2755	MMETSP0045	
	Cercozoa	Chlorarachniophyceae	Chlorarachniaceae	Bigelowiella	natans	CCMP1259	MMETSP1054	
CRYPTOPHYTES	Cryptophyta	Cryptophyceae	Cryptomonadaceae	Cryptomonas	paramecium	CCAP977/2a	MMETSP0038	CAMPEP_0113666334, CAMPEP_0113694150
	Cryptophyta	Cryptophyceae	Cryptomonadaceae	Cryptomonas	curvata	CCAP979/52	MMETSP1050	
	Cryptophyta	Cryptophyceae	Geminigeraceae	Geminigera	cryophila	CCMP2564	MMETSP0799	CAMPEP_0179434534, CAMPEP_0179449568
	Cryptophyta	Cryptophyceae	Geminigeraceae	Geminigera	sp.	Caron Lab Isolate	MMETSP1102	CAMPEP_0173083546
	Cryptophyta	Cryptophyceae	Geminigeraceae	Guillardia	theta	CCMP 2712	MMETSP0046	CAMPEP_0113805578
	Cryptophyta	Cryptophyceae	Hemiselmidaceae	Hemiselmis	andersenii	CCMP439	MMETSP1041	CAMPEP_0172040998
	Cryptophyta	Cryptophyceae	Hemiselmidaceae	Hemiselmis	andersenii	CCMP1180	MMETSP1042	
	Cryptophyta	Cryptophyceae	Hemiselmidaceae	Hemiselmis	andersenii	CCMP441	MMETSP1043	
	Cryptophyta	Cryptophyceae	Hemiselmidaceae	Hemiselmis	tepida	CCMP443	MMETSP1355	
	Cryptophyta	Cryptophyceae	Hemiselmidaceae	Hemiselmis	rufescens	PCC563	MMETSP1357	CAMPEP_0173463728
	Cryptophyta	Cryptophyceae	Hemiselmidaceae	Hemiselmis	viresens	PCC157	MMETSP1356	CAMPEP_0173417114, CAMPEP_0173401560

	Cryptophyta	Cryptophyceae	Goniomonodaceae	Goniomonas	Pacifica	CCMP1869	MMETSP0107, MMETSP0108	CAMPEP_0188542586, CAMPEP_0188462444
	Cryptophyta	Cryptophyceae	Pyrenomonadaceae	Rhodomonas	salina	CCMP1319	MMETSP1047	CAMPEP_0172107506
	Cryptophyta	Cryptophyceae	Pyrenomonadaceae	Rhodomonas	sp.	CCMP768	MMETSP1091, MMETSP1389	
	Cryptophyta	Cryptophyceae	Pyrenomonadaceae	Rhodomonas	abbreviata	Caron Lab Isolate	MMETSP1101	CAMPEP_0181320480
HAPTOPHYTES	Haptophyta	Pavlovophyceae	Pavlovaceae	Pavlova	gyrans	CCMP608	MMETSP1466	CAMPEP_0206063936, CAMPEP_0206044920
	Haptophyta	Pavlovophyceae	Pavlovaceae	Pavlova	lutheri	RCC1537	MMETSP1463	CAMPEP_0205992236, CAMPEP_0205975006
	Haptophyta	Pavlovophyceae	Pavlovaceae	Pavlova	sp.	CCMP459	MMETSP1139, MMETSP1140, MMETSP1381	CAMPEP_0190501384, CAMPEP_0190500156, CAMPEP_0190493934
	Haptophyta	Prymnesiophyceae	Prymnesiaceae	Chrysochromulina	polylepis	UIO037	MMETSP0286	CAMPEP_0115789522, CAMPEP_0115804416
	Haptophyta	Prymnesiophyceae	Prymnesiaceae	Chrysochromulina	polylepis	CCMP1757	MMETSP0143, MMETSP0145, MMETSP0146, MMETSP0147	CAMPEP_0193774398, CAMPEP_0193733264, CAMPEP_0193714344
	Haptophyta	Prymnesiophyceae	Prymnesiaceae	Chrysochromulina	rotalis	UIO044	MMETSP0287	CAMPEP_0115848686, CAMPEP_0115849336
	Haptophyta	Prymnesiophyceae	Prymnesiaceae	Chrysochromulina	ericina	CCMP281	MMETSP1096	CAMPEP_0181225404, CAMPEP_0181185074, CAMPEP_0181203974
	Haptophyta	Prymnesiophyceae	Prymnesiaceae	Chrysochromulina	brevifilum	UTEX LB 985	MMETSP1094	CAMPEP_0174694040, CAMPEP_0174722522
	Haptophyta	Prymnesiophyceae	Prymnesiaceae	Prymnesium	parvum	Texoma1	MMETSP0006, MMETSP0007, MMETSP0008, MMETSP0815, MMETSP0814	CAMPEP_0191217894, CAMPEP_0191219004
	Haptophyta	Prymnesiophyceae	Noelaerhabdaceae	Emiliania	huxleyi	374	MMETSP1006, MMETSP1007, MMETSP1008, MMETSP1009	CAMPEP_0187593462, CAMPEP_0187593926
	Haptophyta	Prymnesiophyceae	Noelaerhabdaceae	Emiliania	huxleyi	379	MMETSP0994, MMETSP0995, MMETSP0996, MMETSP0997	CAMPEP_0187638932, CAMPEP_0187642372

	Haptophyta	Prymnesiophyceae	Noelaerhabdaceae	Emiliania	huxleyi	PLY M219	MMETSP1150, MMETSP1151, MMETSP1152, MMETSP1153	CAMPEP_0187784104, CAMPEP_0187739266, CAMPEP_0187774796, CAMPEP_0187752722
	Haptophyta	Prymnesiophyceae	Noelaerhabdaceae	Emiliania	huxleyi	CCMP370	MMETSP1154, MMETSP1155, MMETSP1156, MMETSP1157	CAMPEP_0187689422, CAMPEP_0187675512, CAMPEP_0187666690, CAMPEP_0187690812
	Haptophyta	Prymnesiophyceae	Noelaerhabdaceae	Gephyrocapsa	oceanica	RCC1303	MMETSP1363, MMETSP1364, MMETSP1365, MMETSP1366	CAMPEP_0188186124, CAMPEP_0188175510, CAMPEP_0188184854,
	Haptophyta	Prymnesiophyceae	Isochrysidaceae	Isochrysis	sp.	CCMP1324	MMETSP1129, MMETSP1130, MMETSP1131, MMETSP1132	CAMPEP_0188822040, CAMPEP_0188824840,
	Haptophyta	Prymnesiophyceae	Isochrysidaceae	Isochrysis	sp	CCMP1244	MMETSP1090, MMETSP1388	CAMPEP_0188737208, CAMPEP_0188802600, CAMPEP_0188741594
	Haptophyta	Prymnesiophyceae	Isochrysidaceae	Isochrysis	galbana	CCMP1323	MMETSP0944, MMETSP0943, MMETSP0595	CAMPEP_0193653704, CAMPEP_0193664480, CAMPEP_0193650476
	Haptophyta	Prymnesiophyceae	Phaeocystaceae	Phaeocystis	Sp	CCMP2710	MMETSP1162	CAMPEP_0118810330, CAMPEP_0118810770
	Haptophyta	Prymnesiophyceae	Phaeocystaceae	Phaeocystis	antarctica	Caron Lab Isolate	MMETSP1100	CAMPEP_0172966702, CAMPEP_0172994118
	Haptophyta	Prymnesiophyceae	Phaeocystaceae	Phaeocystis	antarctica	CCMP1374	MMETSP1444	CAMPEP_0198171844, CAMPEP_0198162072
	Haptophyta	Prymnesiophyceae	Pleurochrysidaceae	Pleurochrysis	carterae	CCMP645	MMETSP1136, MMETSP1137, MMETSP1138	CAMPEP_0190795712, CAMPEP_0190762420, CAMPEP_0190796590
STRAMENOPILES	Ochrophyta	Bacillariophyceae	Amphipleuraceae	Amphiprora	sp.	CCMP467	MMETSP0725, MMETSP0726, MMETSP0727, MMETSP0724	
	Ochrophyta	Bacillariophyceae	Catenulaceae	Amphora	coffeaeformis	CCMP127	MMETSP0316, MMETSP0317, MMETSP0318	

	Ochrophyta	Bacillariophyceae	Bacillariaceae	Fragilariopsis	kerguelensis	L26-C5	MMETSP0733, MMETSP0734, MMETSP0735, MMETSP0736	
	Ochrophyta	Bacillariophyceae	Bacillariaceae	Fragilariopsis	kerguelensis	L2-C3	MMETSP0906, MMETSP0907, MMETSP0908, MMETSP0909	
	Ochrophyta	Bacillariophyceae	Bacillariaceae	Nitzschia	punctata	CCMP561	MMETSP0744, MMETSP0745, MMETSP0746, MMETSP0747	
	Ochrophyta	Bacillariophyceae	Bacillariaceae	Pseudo-nitzschia	australis	10249 10 AB	MMETSP0139, MMETSP0142, MMETSP0140, MMETSP0141	
	Ochrophyta	Bacillariophyceae	Bacillariaceae	Pseudo-nitzschia	fraudulenta	WWA7	MMETSP0850, MMETSP0851, MMETSP0852, MMETSP0853	CAMPEP_0199821422, CAMPEP_0199816092
	Ochrophyta	Bacillariophyceae	Chaetocerotaceae	Chaetoceros	debilis	MM31A-1	MMETSP0149, MMETSP0150	
	Ochrophyta	Bacillariophyceae	Chaetocerotaceae	Chaetoceros	neogracile	CCMP1317	MMETSP0751, MMETSP0752, MMETSP0753, MMETSP0754	CAMPEP_0200995064
	Ochrophyta	Bacillariophyceae	Chaetocerotaceae	Chaetoceros	curvisetus		MMETSP0716, MMETSP0717, MMETSP0718, MMETSP0719	CAMPEP_0187051316, CAMPEP_0187072830
	Ochrophyta	Bacillariophyceae	Chaetocerotaceae	Chaetoceros	affinis	CCMP159	MMETSP0088, MMETSP0090, MMETSP0091, MMETSP0092	
	Ochrophyta	Bacillariophyceae	Corethraceae	Corethron	pennatum	L29A3	MMETSP0169, MMETSP0171	
	Ochrophyta	Bacillariophyceae	Lithodesmiaceae	Ditylum	brightwellii	GSO103	MMETSP1002, MMETSP1005	
	Ochrophyta	Bacillariophyceae	Lithodesmiaceae	Ditylum	brightwellii	GSO104	MMETSP1010, MMETSP1012, MMETSP1013	
	Ochrophyta	Bacillariophyceae	Lithodesmiaceae	Ditylum	brightwellii	GSO105	MMETSP0998, MMETSP1001	
	Ochrophyta	Bacillariophyceae	Cymatosiraceae	Extubocellulus	spinifer	CCMP396	MMETSP0699, MMETSP0697, MMETSP0698,	
	Ochrophyta	Bacillariophyceae	Thalassiosiraceae	Thalassiosira	rotula	CCMP3096	MMETSP0403, MMETSP0404	

	Ochrophyta	Bacillariophyceae	Thalassiosiraceae	Thalassiosira	oceanica	CCMP1005	MMETSP0970, MMETSP0971, MMETSP0972, MMETSP0973	
	Ochrophyta	Bacillariophyceae	Thalassiosiraceae	Thalassiosira	weissflogii	CCMP1010	MMETSP0898, MMETSP0899, MMETSP0900, MMETSP0901, MMETSP1407, MMETSP1408, MMETSP1405, MMETSP1406, MMETSP1415, MMETSP1416, MMETSP1417, MMETSP1418, MMETSP1419, MMETSP1420, MMETSP1421, MMETSP1422, MMETSP1409, MMETSP1410, MMETSP1411, MMETSP1412, MMETSP1413, MMETSP1414,	
	Ochrophyta	Bacillariophyceae	Thalassiosiraceae	Thalassiosira	antarctica	CCMP982	MMETSP0902, MMETSP0903, MMETSP0904, MMETSP0905	
	Ochrophyta	Bacillariophyceae	Thalassiosiraceae	Thalassiosira	weissflogii	CCMP1336	MMETSP0878, MMETSP0879, MMETSP0880, MMETSP0881	
	Ochrophyta	Bacillariophyceae	Thalassiosiraceae	Thalassiosira	rotula	GSO102	MMETSP0910, MMETSP0911, MMETSP0912, MMETSP0913	
	Ochrophyta	Bacillariophyceae	Thalassiosiraceae	Thalassiosira	gravida	GMp14c1	MMETSP0492, MMETSP0493, MMETSP0494	
	Ochrophyta	Bacillariophyceae	Rhizosoleniaceae	Proboscia	alata	PI-D3	MMETSP0174, MMETSP0176	
	Ochrophyta	Bacillariophyceae	Skeletonemaceae	Skeletonema	marinoi	skelA	MMETSP0920, MMETSP0918	
	Ochrophyta	Bacillariophyceae	Skeletonemaceae	Skeletonema	dohrnii	SkelB	MMETSP0562, MMETSP0563	
	Ochrophyta	Bacillariophyceae	Skeletonemaceae	Skeletonema	menzelii	CCMP793	MMETSP0603, MMETSP0604	
	Ochrophyta	Bacillariophyceae	Fragilariaceae	Asterionellopsis	glacialis	CCMP134	MMETSP0705, MMETSP0706, MMETSP0707, MMETSP0708	
	Ochrophyta	Bacillariophyceae	Thalassionemataceae	Thalassionema	nitzschioides	L26-B	MMETSP0156, MMETSP0158	

	Ochrophyta	Bacillariophyceae	Thalassionemataceae	Thalassiothrix	antarctica	L6-D1	MMETSP0152, MMETSP0154	
	Ochrophyta	Chrysophyceae	Synuraceae	Paraphysomonas	Imperforata	PA2	MMETSP0103, MMETSP0104	
	Ochrophyta	Chrysophyceae	Dinobryaceae	Dinobryon	sp.	UTEXLB2267	MMETSP0019, MMETSP0020, MMETSP0812	
	Ochrophyta	Chrysophyceae	Ochromonadaceae	Ochromonas	sp.	CCMP1393	MMETSP0004, MMETSP0005	
	Ochrophyta	Dictyochophyceae	Pedinellaceae	Pseudopedinella	elastica	CCMP716	MMETSP1068, MMETSP1097	CAMPEP_0191399824, CAMPEP_0191445174, CAMPEP_0191393194
	Ochrophyta	Dictyochophyceae	Pedinellaceae	Pteridomonas	danica	PT	MMETSP0101, MMETSP0102	
	Ochrophyta	Eustigmatophyceae	Eustigmataceae	Nannochloropis	gaditana	B-31	http://www.ncbi.nlm.nih.gov/genome/11691?genome_assembly_id=53301	
	Ochrophyta	Pelagophyceae	Pelagomonodaceae	Aureococcus	anophagefferens	CCMP1850	MMETSP0914, MMETSP0915, MMETSP0916, MMETSP0917	
	Ochrophyta	Pelagophyceae	Sarcionochrysidaceae	Aureoumbra	lagunensis	CCMP1510	MMETSP0890, MMETSP0891, MMETSP0892, MMETSP0893	CAMPEP_0186712708, CAMPEP_0186709398
	Ochrophyta	Pelagophyceae	Pelagomonodaceae	Pelagococcus	subviridis	CCMP1429	MMETSP0882, MMETSP0883, MMETSP0884, MMETSP0885	CAMPEP_0190566250, CAMPEP_0190534626,
	Ochrophyta	Pelagophyceae	Pelagomonodaceae	Pelagomonas	calceolata	CCMP1756	MMETSP0888, MMETSP0889, MMETSP0886, MMETSP0887	CAMPEP_0199710348, CAMPEP_0199707362
	Ochrophyta	Raphidophyceae	Chattonellaceae	Chattonella	subsalsa	CCMP2191	MMETSP0947, MMETSP0948, MMETSP0949, MMETSP0950	
	Ochrophyta	Raphidophyceae	Chattonellaceae	Heterosigma	akashiwo	CCMP2393	MMETSP0292, MMETSP0294, MMETSP0295, MMETSP0296	
	Ochrophyta	Raphidophyceae	Chattonellaceae	Heterosigma	akashiwo	NB	MMETSP0416, MMETSP0414, MMETSP0415	CAMPEP_0200281528
	Ochrophyta	Raphidophyceae	Chattonellaceae	Heterosigma	akashiwo	CCMP3107	MMETSP0409, MMETSP0410, MMETSP0411	

	Ochrophyta	Raphidophyceae	Chattonellaceae	Heterosigma	akashiwo	CCMP 452	MMETSP0894, MMETSP0895, MMETSP0896, MMETSP0897	
	Ochrophyta	Xanthophyceae	Vaucheriaceae	Vaucheria	litorea	CCMP2940	MMETSP0945, MMETSP0946	
	Labyrinthista	Labyrinthulea	Thraustochytriaceae	Aplanochytrium	sp	PBS07	MMETSP0954, MMETSP0955, MMETSP0956, MMETSP0957	CAMPEP_0200285312
	Labyrinthista	Labyrinthulea	Thraustochytriaceae	Aplanochytrium	stocchinoi	GSBS06	MMETSP1347, MMETSP1348, MMETSP1349, MMETSP1346	CAMPEP_0200742152
	Labyrinthista	Labyrinthulea	Thraustochytriidae	Aurantiochytrium	limacinum	ATCCMYA-1381	MMETSP0959, MMETSP0960, MMETSP0961, MMETSP0958	CAMPEP_0186630566, CAMPEP_0186630590
	Labyrinthista	Labyrinthulea	Thraustochytriaceae	Schizochytrium	aggregatum	ATCC28209	MMETSP0962, MMETSP0963, MMETSP0964, MMETSP0965	CAMPEP_0191617876, CAMPEP_0191593346
	Labyrinthista	Labyrinthulea	Thraustochytriaceae	Thraustochytrium	sp.	LLF1b	MMETSP0198, MMETSP0199	CAMPEP_0193125154
ALVEOLATA	Dinoflagellata	Dinophyceae	Gymnodiniaceae	Amphidinium	carterae	CCMP1314	MMETSP0399, MMETSP0259, MMETSP0258, MMETSP0398C	CAMPEP_0186418116, CAMPEP_0186410580
	Dinoflagellata	Dinophyceae	Gymnodiniaceae	Karenia	brevis	CCMP2229	MMETSP0027, MMETSP0029, MMETSP0030, MMETSP0031	CAMPEP_0188924570, CAMPEP_0188846816
	Dinoflagellata	Dinophyceae	Gymnodiniaceae	Karenia	brevis	Wilson	MMETSP0202, MMETSP0201, MMETSP0648, MMETSP0649	CAMPEP_0189347872, CAMPEP_0189457996
	Dinoflagellata	Dinophyceae	Gymnodiniaceae	Karenia	brevis	SP3	MMETSP0527, MMETSP0528	CAMPEP_0189224940, CAMPEP_0189232232
	Dinoflagellata	Dinophyceae	Gymnodiniaceae	Karenia	brevis	SP1	MMETSP0573, MMETSP0574	CAMPEP_0189040674
	Dinoflagellata	Dinophyceae	Gymnodiniaceae	Karlodinium	micrum	CCMP2283	MMETSP1015, MMETSP1016, MMETSP1017	CAMPEP_0200829246, CAMPEP_0200797158
	Dinoflagellata	Dinophyceae	Peridiniaceae	Durinskia	baltica	CSIRO CS-38	MMETSP0117, MMETSP0116	CAMPEP_0199932972, CAMPEP_0199938892

	Dinoflagellata	Dinophyceae	Peridiniaceae	Glenodinium	foliaceum	CCAP 1116/3	MMETSP0118, MMETSP0119	CAMPEP_0188250780, CAMPEP_0188433956, CAMPEP_0188276828
	Dinoflagellata	Dinophyceae	Peridiniaceae	Kryptoperidinium	foliaceum	CCMP 1326	MMETSP0121, MMETSP0120	CAMPEP_0189652728, CAMPEP_0189889440
	Dinoflagellata	Dinophyceae	Peridiniaceae	Peridinium	aciculiferum	PAER-2	MMETSP0370, MMETSP0371	CAMPEP_0190614556, CAMPEP_0190676276
	Dinoflagellata	Dinophyceae	Peridiniaceae	Scrippsiella	trochoidea	CCMP3099	MMETSP0270, MMETSP0271, MMETSP0272	CAMPEP_0192123156, CAMPEP_0192103400
	Dinoflagellata	Dinophyceae	Peridiniaceae	Scrippsiella	Hangoei	SHTV-5	MMETSP0359, MMETSP0360, MMETSP0361	CAMPEP_0191783558, CAMPEP_0191797954
	Dinoflagellata	Dinophyceae	Peridiniaceae	Scrippsiella	hangoei-like	SHHI-4	MMETSP0367, MMETSP0368, MMETSP0369	CAMPEP_0199299354, CAMPEP_0199190796
	Dinoflagellata	Dinophyceae	Goniodomataceae	Alexandrium	monilatum	CCMP3105	MMETSP0095, MMETSP0096, MMETSP0097, MMETSP0093	CAMPEP_0200565566, CAMPEP_0200553256
	Dinoflagellata	Dinophyceae	Goniodomataceae	Alexandrium	fundyense	CCMP1719	MMETSP0196C, MMETSP0347	CAMPEP_0185987224
	Dinoflagellata	Dinophyceae	Goniodomataceae	Alexandrium	tamarense	CCMP1771	MMETSP0382, MMETSP0384, MMETSP0378, MMETSP0380	CAMPEP_0186245008, CAMPEP_0186327612
	Dinoflagellata	Dinophyceae	Unknown	Azadinium	spinosum	3D9	MMETSP1036, MMETSP1037, MMETSP1038	CAMPEP_0186815290, CAMPEP_0186817968
	Dinoflagellata	Dinophyceae	Ceratiaceae	Ceratium	fusus	PA161109	MMETSP1075, MMETSP1074	CAMPEP_0199430866, CAMPEP_0199436922
	Dinoflagellata	Dinophyceae	Crypthecodiniacea	Crypthecodinium	cohnii	Seligo	MMETSP0323, MMETSP0325, MMETSP0326, MMETSP0324	CAMPEP_0193892292, CAMPEP_0193855486
	Dinoflagellata	Dinophyceae	Gonyaulacaceae	Lingulodinium	polyedra	CCMP 1738	MMETSP1032, MMETSP1033, MMETSP1034, MMETSP1035	CAMPEP_0190037028, CAMPEP_0190068998
	Dinoflagellata	Dinophyceae	Oxyrrhinaceae	Oxyrrhis	marina	CCMP1795	MMETSP0452_2, MMETSP0451_2C	
	Dinoflagellata	Dinophyceae	Oxyrrhinaceae	Oxyrrhis	marina	Unknown	MMETSP0468, MMETSP0469, MMETSP0470, MMETSP0471	
	Dinoflagellata	Dinophyceae	Oxyrrhinaceae	Oxyrrhis	marina	LB1974	MMETSP1424, MMETSP1425, MMETSP1426	

	Dinoflagellata	Dinophyceae	Prorocentraceae	Prorocentrum	minimum	CCMP1329	MMETSP0053, MMETSP0055, MMETSP0057, MMETSP0056	CAMPEP_0190910698, CAMPEP_0191025674
	Dinoflagellata	Dinophyceae	Prorocentraceae	Prorocentrum	minimum	CCMP2233	MMETSP0267, MMETSP0268, MMETSP0269	CAMPEP_0191110956, CAMPEP_0191181992
	Dinoflagellata	Dinophyceae	Symbiodiniaceae	Symbiodinium	kawagutii	CCMP2468	MMETSP0132_2, MMETSP0133_2, MMETSP0134_2, MMETSP0135_2	
	Dinoflagellata	Dinophyceae	Symbiodiniaceae	Symbiodinium	sp.	CCMP2430	MMETSP1115, MMETSP1116, MMETSP1117	
	Dinoflagellata	Dinophyceae	Symbiodiniaceae	Symbiodinium	sp.	Mp	MMETSP1122, MMETSP1123, MMETSP1124, MMETSP1125	
	Dinoflagellata	Dinophyceae	Symbiodiniaceae	Symbiodinium	sp.	C1	MMETSP1367, MMETSP1369	
	Dinoflagellata	Dinophyceae	Symbiodiniaceae	Symbiodinium	sp.	C15	MMETSP1370, MMETSP1371	
	Ciliophora	Colopdea	Platyophryidae	Platyophrya	macrostoma	WH	MMETSP0127	
	Ciliophora	Heterotrichea	Climacostomidae	Climacostomum	virens	Stock W-24	MMETSP1397	
	Ciliophora	Oligohymenophorea	Orchitophryidae	Anophryoides	haemophila	AH6	MMETSP1018	
	Ciliophora	Oligotrichea	Ptychocylididae	Favella	taraikaensis	Fe Narragansett Bay	MMETSP0434, MMETSP0436	CAMPEP_0199855288
	Ciliophora	Spirotrichea	Euplotidae	Euplotes	focardii	TN1	MMETSP0205, MMETSP0206	
	Perkinsozoa	Perkinsea	Perkinsidae	Perkinsus	chesapeaki	ATCC PRA-65	MMETSP0925, MMETSP0924C	
	Perkinsozoa	Perkinsea	Perkinsidae	Perkinsus	marinus	ATCC 50439	MMETSP0923, MMETSP0922	

Supplementary Material 1 – Species examined in this study with their respective nucleic acid database identifiers and resulting sequence identifiers for KDN biosynthesis.

Codon optimized KDN-9-P Synthase – *E. coli*

ATG AGC GCC AAA AAA CAG AAA GTA GAC GCC GCT CCA GCG CCT ATT GTA TAT CAC GAA CCG AAA GTG ATG GCG GAG ATT GGA TGT AAC CAT ATG GGC GAT CTG GAA ATC GCA AAG GAG TTA CTG ACG CTG GCA AAA CAG GCG GGT GCA GAG TAT GGG AAA TTT CAG AAA CGG AAT CCA AAA GAA CTG CTT ACG GTG GAG CAG TAC GCG GCC CCA CAC CCG AAT CCG CGT AAT AGC TAT GGT GAT ACA TAT GGC GCG CAT CGC GAG TTT TTG GAG TTC ACT ATT GAG CAG CAT GCG GAA TTG AAG AAG CAC TGC GAG AAA ATT GGC TTA GGT TAC TCG TGT AGC GTA TGG GAC ATG ACT TCT GCG AAA GAA ATT GCG TCG ATT AAC CCG GAC CTG ATC AAA GTT GGC TCG CCC AGC AAC CAA CAT TGG GAG ATG CAG AAA ATC CTC CGT GAC GAA TAC AGC GGG GAC GTT CAC ATC TCC ACG GGT ATG ACT ACA AAA GAA GAA ATT GAG AAG ATC GTG CAA TTT TGG GAG GAG GGA AAA GGT GAT GCC AAA AAT CGG CTG GTA TTA TAT AAT TGC ACC AGC GGT TAT CCG GTC CCG TTT GAA GAT GTT TGC CTG CTG GAG CTC CGT GAA CTC CAC GCC CTG TAT GCT GGT CGC GTG AAG CAT TTA GGG TTC TCT GGG CAC CAC CTG GGT ATC GCC ATT GAT ATT GCA GCA TAT GCC CTC GGC GCC ACA TGG AAC GAA CGC CAT TTC ACC AAA GAT CGC ACT TGG AAA GGA ACA GAT CAT GCT GCG AGT CTG GAA CCA GCG GGC TTG AGC AAA CTG TGC CGT GAT CTG AAG GCG ACT TGG AAA TGC ATG AGC ACT AAG AAA ACC GAA ATC CTG CCG ATC GAA AGC GAG CAG CGC GCC AAG CTG AAA TGG GGT TGC TAT AAC GCT AGC AAA GTG GTG AAG TAA

Modified KDN-9-P synthase with pOPIN overhangs and no start/stop codon ordered from IDT gBlock service:

AAGTTCTGTTTCAGGGCCCG AGC GCC AAA AAA CAG AAA GTA GAC GCC GCT CCA GCG CCT ATT GTA TAT CAC GAA CCG AAA GTG ATG GCG GAG ATT GGA TGT AAC CAT ATG GGC GAT CTG GAA ATC GCA AAG GAG TTA CTG ACG CTG GCA AAA CAG GCG GGT GCA GAG TAT GGG AAA TTT CAG AAA CGG AAT CCA AAA GAA CTG CTT ACG GTG GAG CAG TAC GCG GCC CCA CAC CCG AAT CCG CGT AAT AGC TAT GGT GAT ACA TAT GGC GCG CAT CGC GAG TTT TTG GAG TTC ACT ATT GAG CAG CAT GCG GAA TTG AAG AAG CAC TGC GAG AAA ATT GGC TTA GGT TAC TCG TGT AGC GTA TGG GAC ATG ACT TCT GCG AAA GAA ATT GCG TCG ATT AAC CCG GAC CTG ATC AAA GTT GGC TCG CCC AGC AAC CAA CAT TGG GAG ATG CAG AAA ATC CTC CGT GAC GAA TAC AGC GGG GAC GTT CAC ATC TCC ACG GGT ATG ACT ACA AAA GAA GAA ATT GAG AAG ATC GTG CAA TTT TGG GAG GAG GGA AAA GGT GAT GCC AAA AAT CGG CTG GTA TTA TAT AAT TGC ACC AGC GGT TAT CCG GTC CCG TTT GAA GAT GTT TGC CTG CTG GAG CTC CGT GAA CTC CAC GCC CTG TAT GCT GGT CGC GTG AAG CAT TTA GGG TTC TCT GGG CAC CAC CTG GGT ATC GCC ATT GAT ATT GCA GCA TAT GCC CTC GGC GCC ACA TGG AAC GAA CGC CAT TTC ACC AAA GAT CGC ACT TGG AAA GGA ACA GAT CAT GCT GCG AGT CTG GAA CCA GCG GGC TTG AGC AAA CTG TGC CGT GAT CTG AAG GCG ACT TGG AAA TGC ATG AGC ACT AAG AAA ACC GAA ATC CTG CCG ATC GAA AGC GAG CAG CGC GCC AAG CTG AAA TGG GGT TGC TAT AAC GCT AGC AAA GTG GTG AAG TAAAGCTTTCTAGACCAT

Codon optimized KDN-9-P phosphatase – *E. coli*

ATGAAAGAAA TCAAATTGAT TCTGACCGAC ATCGATGGTG TTTGGACGGA CGGTGGAATG TTCTACGATC AGACGGGTAA CGAATGGAAA AAATTTAACA CTTCTGATTC TGCCGGTATT TTCTGGGCAC ATAACAAAGG AATTCCGGTG GGCATCCTGA CCGGAGAAAA GACGGAAATC GTCCGTCGCC GTGCGGAAAA ACTGAAGGTC GATTATCTGT TCCAAGGTGT AGTTGATAAA TTATCAGCGG CCGAGGAACT GTGCAACGAG CTTGGTATCA ATTTGGAACA GGTCGCCTAC ATTGGTGATG ATTTAAACGA TGCCAAACTT TTGAAACGCG TGGGTATCGC TGGTGTACCT GCGTCAGCGC CCTTCTACAT TCGTCGCCTG TCAACGATCT TTTTGGAAAA ACGGGGCGGC GAAGGTGTGT TTCGCGAATT TGTTGAAAAA GTTCTGGGTA TCAATCTGGA GGATTTTATT GCTGTCATCC AATG

Modified KDN-9-P phosphatase with pOPIN overhangs and no start/stop codon:

AAGTTCTGTTTCAGGGCCCG AAAGAAA TCAAATTGAT TCTGACCGAC ATCGATGGTG TTTGGACGGA CGGTGGAATG TTCTACGATC AGACGGGTAA CGAATGGAAA AAATTTAACA CTTCTGATTC TGCCGGTATT TTCTGGGCAC ATAACAAAGG AATTCCGGTG GGCATCCTGA CCGGAGAAAA GACGGAAATC GTCCGTCGCC GTGCGGAAAA ACTGAAGGTC GATTATCTGT TCCAAGGTGT AGTTGATAAA TTATCAGCGG CCGAGGAACT GTGCAACGAG CTTGGTATCA ATTTGGAACA GGTCGCCTAC ATTGGTGATG ATTTAAACGA TGCCAAACTT TTGAAACGCG TGGGTATCGC TGGTGTACCT GCGTCAGCGC CCTTCTACAT TCGTCGCCTG TCAACGATCT TTTTGGAAAA ACGGGGCGGC GAAGGTGTGT TTCGCGAATT TGTTGAAAAA GTTCTGGGTA TCAATCTGGA GGATTTTATT GCTGTCATCC AA TAAAGCTTTCTAGACCAT

Codon optimized CMP-KDN synthase – E. coli

ATG ACC GTT TGG CAT CCG GTA CCT GAG GTA CGT ATT GTA GCG GTA ATT CCG GCA CGT GGT GGC AGC GTT TCG ATT CCC CGG AAA AAC ATT AAG CCT CTG GCG GGC CGC CCG CTG ATC GAT TGG GTC ATC AAA CCG GCG CTG CAC TGC GGG ATT TTT ACC GAT GTA TAC GTG AGC ACC GAC GAT GAT GCT ATC GCG AGC GTC GCT GAA AAA TGT GGC GCC AAA GTG CAT CGG CGT GAT CCG GCC ACG GCG ACC GCT ACG GCC ACC ACC GAG TCT GCG CTG CTT GAC TTC GCG CAG TCA CAC GGT GAC TTT GAC GTA CTG TGT CTT ATT CAA GCA ACC TCC CCG TTT ATT ACC CCT CGC GAT CTG ATT AAC GGC TGG GAA TTA ATG CGC GCC ATG GAA GCC GAT AGC CTC GTA ACC GCG GTG CGT GCG CAT CGC TTC CTT TGG CAG GTT GAC AAA GAT ACA GGT CTT GCG AAA GCG AAA AAC TAT GAC CCA CTG AAA CGC CCG CGC CGT CAG GAC TGG GAT GGG GAA CTG GTG GAG AAT GGC GCT TTT TAC ATG ACC ACC AAA GCA TGC TTA GAG AAA CAT AAA TGT CGC CTC GGG GAA AAG ATG GTC CTG CTG GAG ATG GAA GAG CAT ACG TTT ACT GAA CTG GAT TCG TTA GTA GAC TGG CAG ATC GTG ACC AAT ATG ACC GAA AAT TAC GGT TAC TGG CCG CCG CGT AAC TGG GGT GAA GCC GCG TCC TCC TCA GCC CGT CCG GAC GCG GCC AAG ATC GTA TTG TGT GCG CTG GGC GTT CTG GCT CTG GGT CTG TCG ATT GGA CGT ATG AGC AAA TAA

Modified CMP-KDN synthase with pOPIN overhangs and no start/stop codon:

AAGTTCTGTTTCAGGGCCCG ACC GTT TGG CAT CCG GTA CCT GAG GTA CGT ATT GTA GCG GTA ATT CCG GCA CGT GGT GGC AGC GTT TCG ATT CCC CGG AAA AAC ATT AAG CCT CTG GCG GGC CGC CCG CTG ATC GAT TGG GTC ATC AAA CCG GCG CTG CAC TGC GGG ATT TTT ACC GAT GTA TAC GTG AGC ACC GAC GAT GAT GCT ATC GCG AGC GTC GCT GAA AAA TGT GGC GCC AAA GTG CAT CGG CGT GAT CCG GCC ACG GCG ACC GCT ACG GCC ACC ACC GAG TCT GCG CTG CTT GAC TTC GCG CAG TCA CAC GGT GAC TTT GAC GTA CTG TGT CTT ATT CAA GCA ACC TCC CCG TTT ATT ACC CCT CGC GAT CTG ATT AAC GGC TGG GAA TTA ATG CGC GCC ATG GAA GCC GAT AGC CTC GTA ACC GCG GTG CGT GCG CAT CGC TTC CTT TGG CAG GTT GAC AAA GAT ACA GGT CTT GCG AAA GCG AAA AAC TAT GAC CCA CTG AAA CGC CCG CGC CGT CAG GAC TGG GAT GGG GAA CTG GTG GAG AAT GGC GCT TTT TAC ATG ACC ACC AAA GCA TGC TTA GAG AAA CAT AAA TGT CGC CTC GGG GAA AAG ATG GTC CTG CTG GAG ATG GAA GAG CAT ACG TTT ACT GAA CTG GAT TCG TTA GTA GAC TGG CAG ATC GTG ACC AAT ATG ACC GAA AAT TAC GGT TAC TGG CCG CCG CGT AAC TGG GGT GAA GCC GCG TCC TCC TCA GCC CGT CCG GAC GCG GCC AAG ATC GTA TTG TGT GCG CTG GGC GTT CTG GCT CTG GGT CTG TCG ATT GGA CGT ATG AGC AAATAAAGCTTTCTAGACCAT

Modified truncated CMP-KDN synthase with pOPIN overhangs and no start/stop codon ordered from IDT gBlock service:

AAGTTCTGTTTCAGGGCCCG ACC GTT TGG CAT CCG GTA CCT GAG GTA CGT ATT GTA GCG GTA ATT CCG GCA CGT GGT GGC AGC GTT TCG ATT CCC CGG AAA AAC ATT AAG CCT CTG GCG GGC CGC CCG CTG ATC GAT TGG GTC ATC AAA CCG GCG CTG CAC TGC GGG ATT TTT ACC GAT GTA TAC GTG AGC ACC GAC GAT GAT GCT ATC GCG AGC GTC

GCT GAA AAA TGT GGC GCC AAA GTG CAT CGG CGT GAT CCG GCC ACG GCG ACC GCT ACG GCC ACC ACC GAG TCT GCG CTG CTT GAC TTC GCG CAG TCA CAC GGT GAC TTT GAC GTA CTG TGT CTT ATT CAA GCA ACC TCC CCG TTT ATT ACC CCT CGC GAT CTG ATT AAC GGC TGG GAA TTA ATG CGC GCC ATG GAA GCC GAT AGC CTC GTA ACC GCG GTG CGT GCG CAT CGC TTC CTT TGG CAG GTT GAC AAA GAT ACA GGT CTT GCG AAA GCG AAA AAC TAT GAC CCA CTG AAA CGC CCG CGC CGT CAG GAC TGG GAT GGG GAA CTG GTG GAG AAT GGC GCT TTT TAC ATG ACC ACC AAA GCA TGC TTA GAG AAA CAT AAA TGT CGC CTC GGG GAA AAG ATG GTC CTG CTG GAG ATG GAA GAG CAT ACG TTT ACT GAA CTG GAT TCG TTA GTA GAC TGG CAG ATC GTG ACC AAT ATG ACC GAA AAT TAC GGT TAC TGGTAAAGCTTTCTAGACCAT

Supplementary Material 2 – Sequences used in this study for protein expression. Underlined sequences are overhangs added to the protein sequences for direct cloning into pOPINF vectors using In-Fusion ™ cloning.

5 NDP-β-L-rhamnose Biosynthesis Across the Algal Taxonomic Groups: an Evolutionary Perspective

5.1 Abstract

The 6-deoxy sugar, rhamnose, is an important monosaccharide found in structural polysaccharides, glycoproteins, and secondary metabolites across microbes, algae and plants. The activated sources of L-rhamnose for carbohydrate polymer synthesis are the sugar nucleotides, thymidine 5'-diphospho-β-L-rhamnose (TDP-β-L-Rha) or uridine 5'-diphospho-β-L-rhamnose (UDP-β-L-Rha). Whilst the biosynthesis of these two sugar nucleotides has been studied in some detail in bacteria and plants, no work yet has looked at the production of L-rhamnose in algae. In this work we broaden the current literature knowledge surrounding L-rhamnose biosynthesis using sugar nucleotide profiling techniques as well as a comprehensive bioinformatics analysis to show how algae produce UDP or TDP activated rhamnose. Taking our findings, we propose a likely evolutionary history of rhamnose biosynthesis across the algal groups.

5.2 Introduction

The 6-deoxy sugar rhamnose (Rha) is found in glycoproteins, structural polysaccharides and secondary metabolites across the microbes, algae and plants, but is absent in the animals. Most commonly found as the L enantiomer, L-Rha is present in the capsules or cell walls of bacteria [1, 2], fungi [3], and plants [4]; it is also present in the lesser-studied glycans of viruses [5]. Previous work has also found L-Rha in fungal and bacterial glycans that play crucial roles in host-pathogen interactions [6, 7]. Due to the absence of L-Rha biosynthesis capability in animals, and the essentiality of this sugar for virulence of numerous pathogens, the biosynthetic microbial pathway for L-Rha production has drawn interest as a potential therapeutic target [8, 9].

The activated sources of L-Rha for carbohydrate polymer biosynthesis are thymidine 5'-diphospho-β-L-rhamnose (TDP-β-L-Rha) or uridine 5'-diphospho-β-L-rhamnose (UDP-β-L-Rha), which are produced biosynthetically from thymidine 5'-diphospho-α-D-glucose (TDP-α-D-Glc) or uridine 5'-diphospho-α-D-glucose (UDP-α-D-Glc), respectively (Figure 54). In bacteria, many examples have shown that TDP-β-L-Rha is produced from TDP-α-D-Glc by the action of three independent enzymes [10]; RmlB (a 4,6-dehydratase), RmlC (a 3,5-epimerase) and RmlD (a 4-reductase). RmlB, in particular, is a central player in the biosynthesis of glycosylated natural products: its' product, TDP-6-deoxy-L-*lyxo*-hexulose, is subject to numerous enzymatic processes that produce diverse sugar nucleotide products [11, 12]. The Rml enzymes have been studied in some detail, with crystal structures of all three having been solved [13-15]. More recently, enzymes from plants [16, 17], fungi [6] and even viruses [18] have been shown to synthesise UDP-β-L-Rha from UDP-α-D-Glc. These enzymes are structurally distinct from their bacterial orthologues, with multiple enzymatic activities found on individual proteins. For plants, the dehydratase, epimerase and reductase activities are often found on one large protein (RHM) [16], but there are also instances of the three enzymatic activities being shared over two proteins (UGD/UER1) [19]. Fungi and viruses also contain homologs of the bifunctional epimerase-reductase UER1 plant system, often showing substrate specificity for the uridine activated sugars over the thymidine alternatives [6, 18].

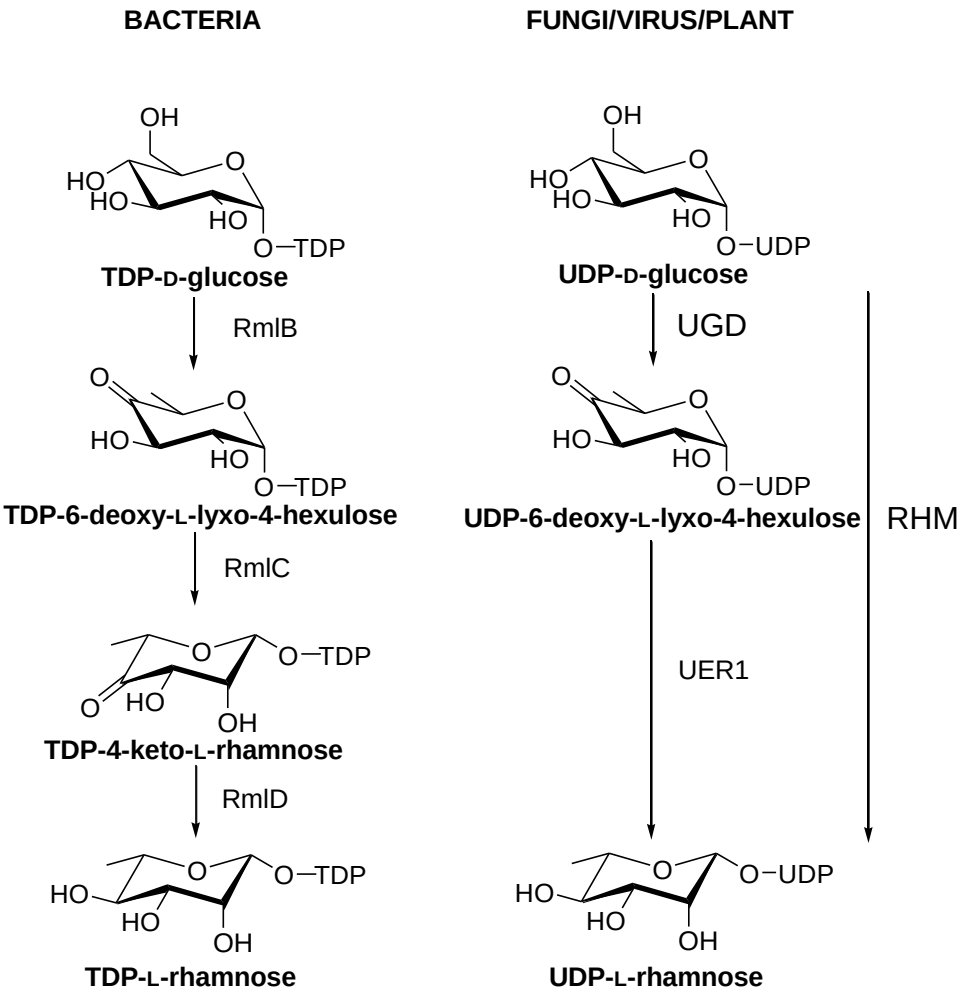


Figure 54 - Biosynthesis of NDP-β-L-Rha in bacteria, fungi, viruses and plants. In bacteria, three independent enzymes catalyse dehydration, epimerization and reduction steps to yield TDP-β-L-Rha – RmlB, RmlC and RmlD, respectively. In plants, fungi and viruses these three enzymatic activities are found on multi-functional enzymes. UER1 from *Arabadopsis thaliana* has been shown to be bifunctional and contain both 3,5-epimerase and 4-reductase activities. RHM proteins from *A. thaliana* has been shown to catalyse both of the previous steps, including the initial 4,6-dehydration step to form the keto-sugar.

Although the biosynthesis of L-Rha has been studied in some detail in bacteria, fungi and plants, previously no work has been done looking at the biosynthesis of L-Rha in the diverse algal groups, even though its presence has been noted in structural polysaccharides of macroalgae [20], and in the surface glycans and pellicle of the green microalgae *Euglena gracilis* [21-23]. Recent work by O'Neill *et al* identified L-Rha catabolic genes in *E. gracilis* but didn't explore the metabolism of L-Rha in any detail [24]. Evolutionarily distinct algae derived from the red algal plastid have also been found to contain L-Rha, with early work identifying

the sugar in cell preparations from the haptophytes *Isochrysis galbana* and *Prymnesium parvum* [25]. Furthermore, independent research carried out in our research group has identified a potential role for TDP-β-L-Rha in viral infection of *P. parvum*, showing a drastic increase in levels of the sugar nucleotide during viral infection (Figure 55).

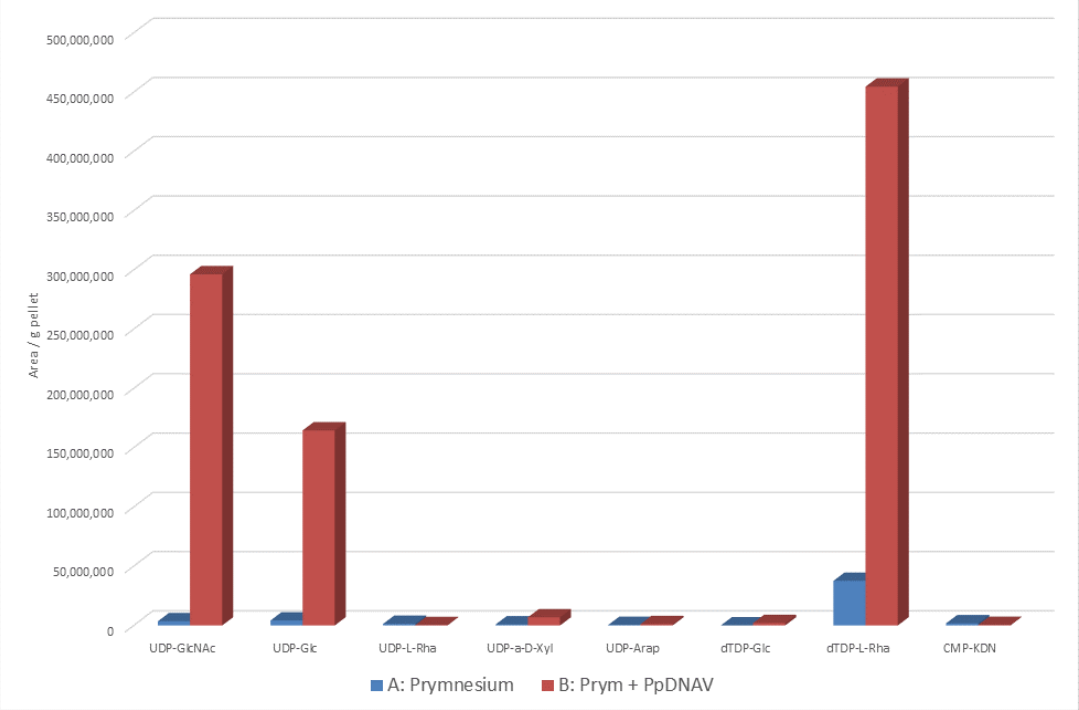


Figure 55 – Evidence for the involvement of TDP-β-L-Rha in viral infection of *P. parvum*. The abundance of 8 sugar nucleotides (x-axis) was monitored in control cultures (blue) and cultures infected by *Prymnesium parvum* DNA virus (blue). A significant increase of TDP-β-L-Rha, UDP-GlcNAc, and UDP-Glc was observed in cultures 48 hours post infection.

Despite the reported occurrence of L-Rha in the algae, it is not known how algae produce L-Rha, which nucleotides they use to activate L-Rha, or how and where algae acquired their L-Rha biosynthetic machinery in evolutionary terms. To increase understanding of L-Rha biosynthesis in algae we had three main objectives in this work: (1) to evaluate the occurrence of enzymes involved in L-Rha across the algal groups, using previously characterized bacterial and plant enzymes known to be involved in L-Rha biosynthesis; (2) to analyse the evolution of the algal pathways with respect to previously characterized bacterial and plant pathways; and (3) to discover whether algae activate cytosolic L-Rha using thymidine diphosphate (TDP) or uridine diphosphate (UDP). By mining recently available algal transcriptomes and genomes we show that algal L-Rha biosynthesis is closely related to other eukaryotic pathways, with exceptions in the haptophytes, fucoxanthin-containing dinoflagellates and rhizarians. By profiling intracellular sugar-nucleotides of a representative euglenid, *Euglena gracilis*, and

haptophyte, *Prymnesium parvum*, we support our bioinformatic findings by showing a preference for TDP-activated L-Rha in *P. parvum,* and UDP-activated L-Rha in *E. gracilis.* Using these findings, we evaluate potential routes for the evolution of the NDP-β-L-Rha pathway in the algae.

5.3 Results

Algae are photosynthetic eukaryotic organisms belonging to a wide range of taxonomic groups. They arose following endosymbiotic acquisition of a cyanobacterium by a heterotrophic protist [26, 27]. This led to the emergence of the glaucophyte, red algal and green algal lineages by a process referred to as the primary endosymbiosis event. Other eukaryotic lineages including the cryptophytes, alveolates (e.g. dinoflagellates), stramenopiles, haptophytes, euglenids and chlorarachniophytes subsequently obtained their plastids through a secondary endosymbiosis event. The euglenid and chlorarachniophyte plastids are of green algal origin, whilst the cryptophyte, alveolate, stramenopile and haptophyte (collectively termed CASH) plastids are believed to be of red algal origin [27]. This has since been followed by tertiary endosymbiosis events in the Dinoflagellata phylum, whereby the dinoflagellates have replaced their peridinin-containing plastids with plastids of other algae [28]. One such example of this is the acquisition of the fucoxanthin-containing haptophyte plastid by the Gymnodiniaceae family of dinoflagellates [29, 30]. With each endosymbiosis event, a massive amount of lateral gene transfer occurred from the engulfed symbiont to the host nuclear genome by a process known as endosymbiotic gene transfer (EGT). This in turn has led to the enormous genetic and therefore physiological diversity observed in the algal groups [31].

5.3.1 Distribution of L-Rha biosynthetic genes in algae

In order to identify algal transcripts involved in NDP-β-L-Rha biosynthesis, BLAST searches were performed using reference sequences from bacteria and plants known to be involved in these pathways. BLASTp searches were carried out against the putative proteins assembled and translated from the algal transcriptomes and genomes using protein sequences for RmlC (NP_217982.1), UER1 (NP_564806.1) and RHM (NP_177978.1). We found that in most instances the bacterial pathway (represented by RmlC or RmlC/D Fusion) and the plant pathway (represented by UER1 or RHM) were mutually exclusive, with the exception of 15 dinoflagellates found in the Alveolata superphylum (Figure 56). In general, our findings suggest a larger support for the eukaryotic pathway amongst algal groups than the bacterial pathway.

For the algae derived from primary endosymbiosis (i.e. glaucophytes, red algae and green algae), no homologs of bacterial RmlC were found. Of the two glaucophyte transcriptomes examined in this study, both contained one homolog of UER1, suggesting a plant-like biosynthesis of L-Rha in this phylum. Of the green algae examined, six out of thirteen

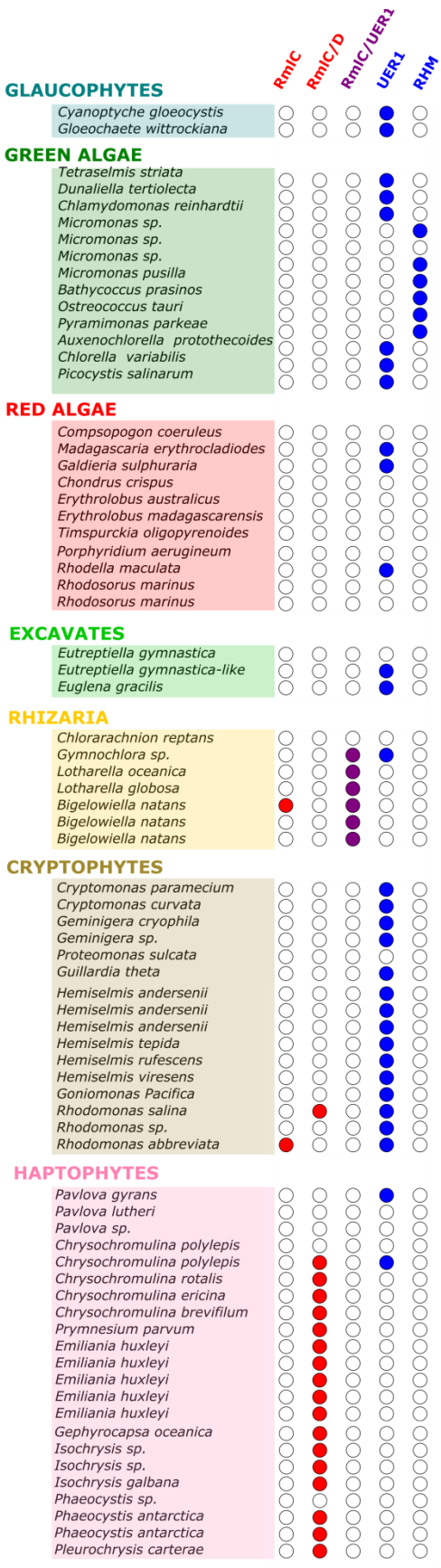
contained a UER1 homolog, and six of the remaining seven contained a homolog of the trifunctional RHM. No hits were observed for one *Micromonas* sp. strain examined. The lack of bacterial RmlC homologs and abundance of plant UER1 or RHM homologs would also support a plant-like biosynthesis of L-Rha in this taxonomic group. Of the eleven red algae examined, only three contained homologs to UER1, with no homologs to RmlC or RHM being observed, which may suggest a large-scale loss of L-Rha biosynthesis in this group.

Amongst algae derived from secondary endosymbiosis with a green algal symbiont (i.e. excavates and rhizarians), two out of three excavates examined contained UER1 homologs, with no instances of RmlC homologs. No hits were observed for *Eutreptiella gymnastica*. This suggests a plant-like L-Rha biosynthesis pathway again, although a larger sample size would be more desirable. In Rhizaria, an unexpected recurrent transcript was found that was a fusion between bacterial RmlC and plant UER1. Of the seven rhizarians examined, six contained this RmlC/UER1 putative fusion protein. In addition to this transcript, *Gymnochlora* sp. also contained a standalone UER1 homolog, and one strain of *Bigelowiella natans* also contained a standalone RmlC homolog. No hits for L-Rha biosynthesis were found for *Chlorarachnion reptans*. Rhizaria therefore appear to combine both bacterial and plant-like machinery for L-Rha biosynthesis.

Amongst algae derived from secondary endosymbiosis with a red-alga (i.e. CASH), the cryptophytes largely contain UER1 homologs, with fifteen out of sixteen examined containing a UER1 homolog. Of these fifteen, *Rhodomonas abbreviata* also contains an RmlC homolog, and *Rhodomonas salina* also contains a homolog of an RmlC/D fusion. *Proteomonas sulcata* contains no hits for L-Rha biosynthesis. The abundance of UER1 homologs and lack of RmlC homologs in this group would suggest a plant-like biosynthesis of L-Rha in the cryptophytes. Of the twenty-two haptophytes examined, only 2 contained homologs to either UER1 or RHM. Unexpectedly, seventeen of the twenty-two examined contained a putative fusion protein between bacterial RmlC and RmlD. No hits were found for L-Rha biosynthesis in *Pavlova lutheri*, *Pavlova* sp., one strain of *Chrysochromulina polyepsis* and *Phaeocystis* sp.. Taken together, this suggests a bacterial-like pathway for L-Rha biosynthesis in the haptophytes. The stramenopiles displayed a similar pattern to the cryptophytes, with forty-three of the fifty-one strains examined containing a homolog to UER1. Of these forty-three *Amphiprora* sp. and *Chaetoceros neogracile* also contained an RmlC/D fusion homolog, and *Asterionellopsis glacialis* and *Pseudopedinella elastica* also contained a standalone RmlC homolog. *Pseudo-nitzschia fraudulenta* contained only a standalone RmlC homolog, and seven strains examined

contained no hits to known L-Rha biosynthetic enzymes. This consistent abundance of UER1 homologs would also support a plant-like L-Rha biosynthesis pathway in the stramenopiles. Unlike the other groups, a clear mix of bacterial and plant-like L-Rha biosynthesis machinery is observed for the alveolates. Of the thirty-eight strains examined, twenty-nine had homologs to UER1, RHM, or both, suggesting plant-like pathways are present consistently in this superphylum. However, bacterial pathways also appear to be consistently present, with sixteen out of thirty-eight strains having a homolog to either standalone RmlC or an RmlC/D fusion. Interestingly, six out of six dinoflagellates examined from the Gymnodiniaceae family (*Amphidinium, Karenia* and *Karlodinium* genera) all contained an RmlC/D fusion; the possible origin of this fusion protein is discussed later. The dinoflagellates, which make up a subgroup of the alveolates, represent a phylum that has undergone extensive endosymbiosis events, and this may explain the abundance of both bacterial and plant-like L-Rha biosynthesis pathways in this group.

To conclude, these findings suggest that most algae use exclusively the plant-like pathway for L-Rha biosynthesis. One exception is the haptophytes, which appear to exclusively use the bacterial pathway. The rhizaria also differ, with the appearance of a fusion between bacterial RmlC and plant UER1. And finally, some members of the Alveolata, such as the Gymnodiniaceae family, which in addition to the plant-like pathway, contain the bacterial pathway.


RmlC
RmlC/D
RmlC/UER1
UER1
RHM
GLAUCOPHYTES
Cyanoptyche gloeocystis
Gloeochaete wittrockiana
GREEN ALGAE
Tetraselmis striata
Dunaliella tertiolecta
Chlamydomonas reinhardtii
Micromonas sp.
Micromonas sp.
Micromonas sp.
Micromonas pusilla
Bathycoccus prasinos
Ostreococcus tauri
Pyramimonas parkeae
Auxenochlorella protothecoides
Chlorella variabilis
Picocystis salinarum
RED ALGAE
Compsopogon coeruleus
Madagascaria erythrocladiodes
Galdieria sulphuraria
Chondrus crispus
Erythrolobus australicus
Erythrolobus madagascarensis
Timspurckia oligopyrenoides
Porphyridium aerugineum
Rhodella maculata
Rhodosorus marinus
Rhodosorus marinus
EXCAVATES
Eutreptiella gymnastica
Eutreptiella gymnastica-like
Euglena gracilis
RHIZARIA
Chlorarachnion reptans
Gymnochlora sp.
Lotharella oceanica
Lotharella globosa
Bigelowiella natans
Bigelowiella natans
Bigelowiella natans
CRYPTOPHYTES
Cryptomonas paramecium
Cryptomonas curvata
Geminigera cryophila
Geminigera sp.
Proteomonas sulcata
Guillardia theta
Hemiselmis andersenii
Hemiselmis andersenii
Hemiselmis andersenii
Hemiselmis tepida
Hemiselmis rufescens
Hemiselmis viresens
Goniomonas Pacifica
Rhodomonas salina
Rhodomonas sp.
Rhodomonas abbreviata
HAPTOPHYTES
Pavlova gyrans
Pavlova lutheri
Pavlova sp.
Chrysochromulina polylepis
Chrysochromulina polylepis
Chrysochromulina rotalis
Chrysochromulina ericina
Chrysochromulina brevifilum
Prymnesium parvum
Emiliania huxleyi
Emiliania huxleyi
Emiliania huxleyi
Emiliania huxleyi
Emiliania huxleyi
Gephyrocapsa oceanica
Isochrysis sp.
Isochrysis sp.
Isochrysis galbana
Phaeocystis sp.
Phaeocystis antarctica
Phaeocystis antarctica
Pleurochrysis carterae

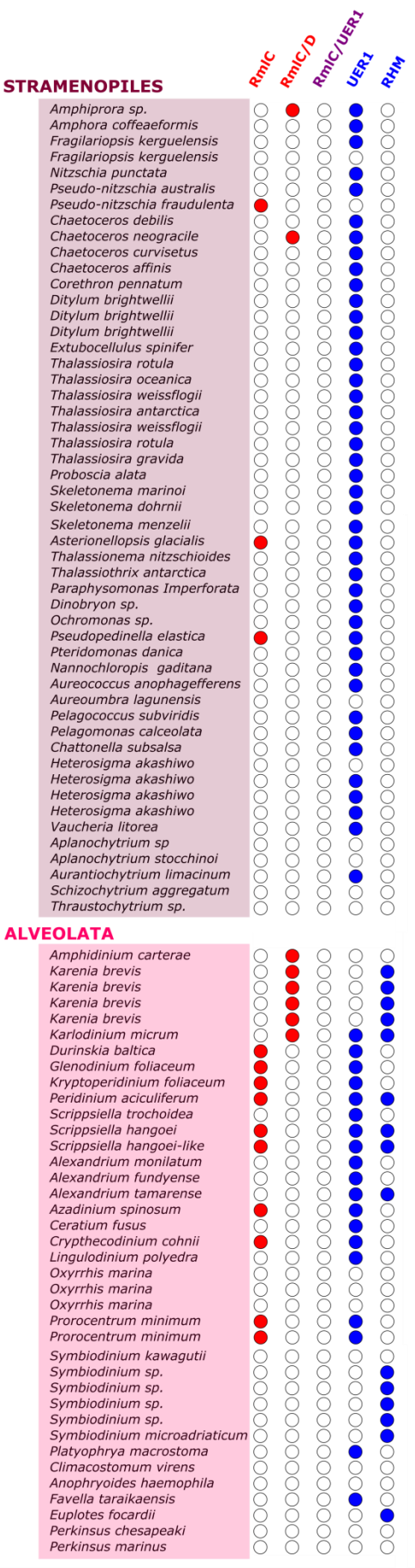

RmlC
RmlC/D
RmlC/UER1
UER1
RHM
STRAMENOPILES
Amphiprora sp.
Amphora coffeaeformis
Fragilariopsis kerguelensis
Fragilariopsis kerguelensis
Nitzschia punctata
Pseudo-nitzschia australis
Pseudo-nitzschia fraudulenta
Chaetoceros debilis
Chaetoceros neogracile
Chaetoceros curvisetus
Chaetoceros affinis
Corethron pennatum
Ditylum brightwellii
Ditylum brightwellii
Ditylum brightwellii
Extubocellulus spinifer
Thalassiosira rotula
Thalassiosira oceanica
Thalassiosira weissflogii
Thalassiosira antarctica
Thalassiosira weissflogii
Thalassiosira rotula
Thalassiosira gravida
Proboscia alata
Skeletonema marinoi
Skeletonema dohrnii
Skeletonema menzelii
Asterionellopsis glacialis
Thalassionema nitzschioides
Thalassiothrix antarctica
Paraphysomonas Imperforata
Dinobryon sp.
Ochromonas sp.
Pseudopedinella elastica
Pteridomonas danica
Nannochloropis gaditana
Aureococcus anophagefferens
Aureoumbra lagunensis
Pelagococcus subviridis
Pelagomonas calceolata
Chattonella subsalsa
Heterosigma akashiwo
Heterosigma akashiwo
Heterosigma akashiwo
Heterosigma akashiwo
Vaucheria litorea
Aplanochytrium sp
Aplanochytrium stocchinoi
Aurantiochytrium limacinum
Schizochytrium aggregatum
Thraustochytrium sp.
ALVEOLATA
Amphidinium carterae
Karenia brevis
Karenia brevis
Karenia brevis
Karenia brevis
Karlodinium micrum
Durinskia baltica
Glenodinium foliaceum
Kryptoperidinium foliaceum
Peridinium aciculiferum
Scrippsiella trochoidea
Scrippsiella hangoei
Scrippsiella hangoei-like
Alexandrium monilatum
Alexandrium fundyense
Alexandrium tamarense
Azadinium spinosum
Ceratium fusus
Crypthecodinium cohnii
Lingulodinium polyedra
Oxyrrhis marina
Oxyrrhis marina
Oxyrrhis marina
Prorocentrum minimum
Prorocentrum minimum
Symbiodinium kawagutii
Symbiodinium sp.
Symbiodinium sp.
Symbiodinium sp.
Symbiodinium sp.
Symbiodinium microadriaticum
Platyophrya macrostoma
Climacostomum virens
Anophryoides haemophila
Favella taraikaensis
Euplotes focardii
Perkinsus chesapeaki
Perkinsus marinus

Figure 56 - Coulson plot showing the distribution of L-Rha biosynthesis pathways in algae. A total of 163 transcriptomes or genomes were analysed for the presence of NDP-β-L-Rha biosynthesis genes from bacterial (RmlC) or plant (UER1 or RHM) pathways. Where a transcript was identified for a given gene, a filled circle can be found. For bacterial pathways (i.e. RmlC or RmlC/D fusion) circles are filled red. For plant pathways (i.e. UER1 or RHM) circles are filled blue. When a fusion of bacterial RmlC and plant UER1 is observed, circles are filled purple. Multiple mention of the same species name means that different strains have been analysed. For a full list of transcriptome, genome and corresponding sequence identifiers, along with strains used in this study, refer to Supplementary Material 3.

5.3.2 Phylogenetic analysis and evolutionary implications

A maximum likelihood tree of representative RmlC, RmlC/D, UER1, and RHM homologs was constructed using previously acquired algal sequences, along with sequences from bacteria and plants. Bacterial sequences, including RmlC/D fusions, clearly do not form monophyletic lineages with plant-like UER1 or RHM sequences (Figure 57). Some eukaryotic lineages including the haptophytes and fucoxanthin-containing dinoflagellates (i.e. Gymnodiniaceae) contain RmlC/D fusions that branch with bacterial RmlC or RmlC/D sequences (shown in red). UER1 or RHM homologs from other algal groups including the cryptophytes, alveolates (peridinin-containing dinoflagellates), stramenopiles, excavates, glaucophytes, green algae and red algae branch with plant sequences from *Arabidopsis thaliana* and *Brassica napus*. Relationships between major taxonomic groups within the plant-like part of the tree (blue) are poorly resolved (bootstrap values > 50% shown), but dinoflagellates can be seen scattered throughout, likely due to their complex evolution which has involved multiple endosymbiotic events.

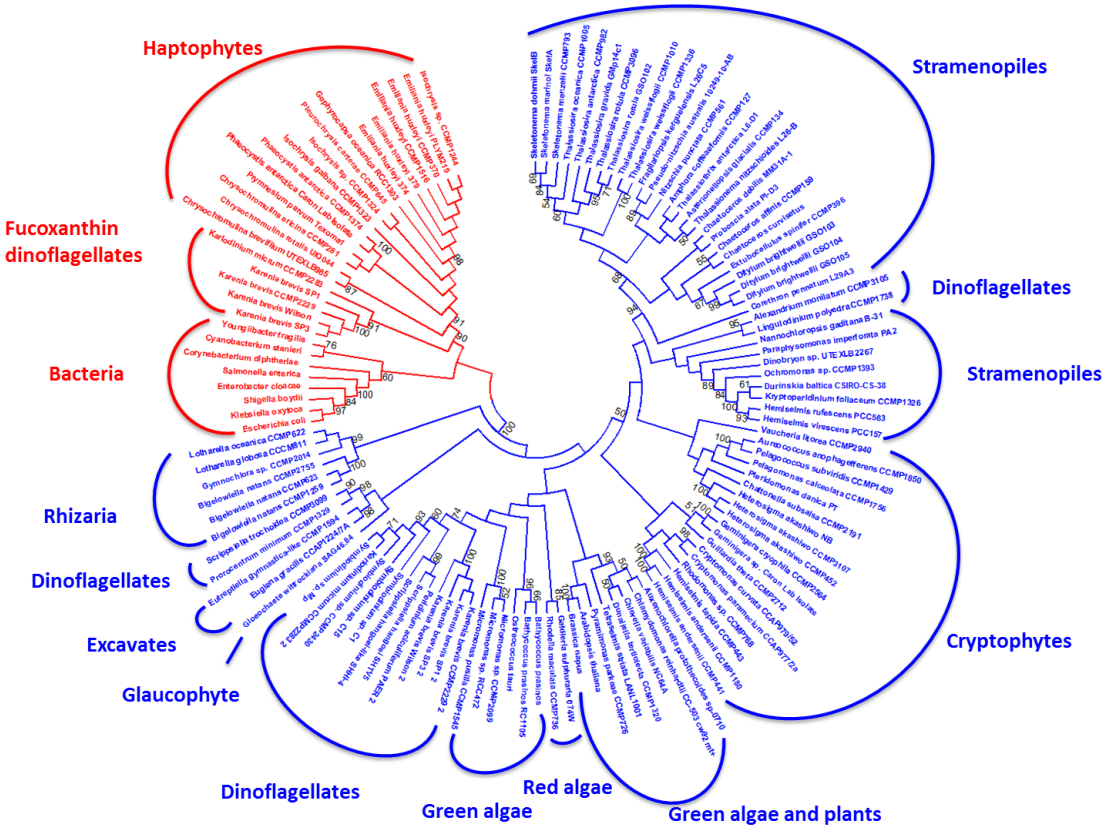


Figure 57 - Phylogenetic clustering of NDP-β-L-Rha biosynthetic machinery. Protein sequences with homology to bacterial RmlC are highlighted in red, whilst sequences with homology to plant UER1 or RHM are highlighted blue. Alignment was performed using the default settings of MAFFT [32], and an unrooted maximum-likelihood phylogenetic tree was produced using 122 sequences from algae, bacteria and plants. The tree was drawn using MEGA7 [33]. The final tree was based on 135 ungapped amino acid positions, 100 resampling permutations and only branches with bootstrap support >50% are labelled. A detailed list of sequences used to create this tree can be found in Supplementary Material 3.

5.3.3 Sugar nucleotide profiling

We next sought to examine the presence and the levels of TDP- and/or UDP-β-L-Rha in the haptophyte, *P. parvum,* and the euglenid, *E. gracilis*. Our bioinformatics based analysis had suggested that *P. parvum* contained the bacterial pathway for L-Rha biosynthesis, likely utilizing TDP as the activating nucleoside diphosphate; whilst *E. gracilis* appeared to contain the plant pathway for L-Rha biosynthesis, likely utilizing UDP adducts. We also looked at the presence of TDP- and/or UDP-α-D-Glc, the likely biosynthetic precursors of the corresponding L-Rha derivatives. As neither TDP-β-L-Rha or UDP-β-L-Rha were commercially available at the

time of this study, these two compounds were chemically synthesized and kindly provided by Martin Rejzek for use as authentic standards for LC-MS/MS experiments.

5.3.4 Quantification of intracellular levels of NDP-β-L-Rha in algal cells

Axenic cultures of algal cells were grown under sterile conditions, at 22 °C, and on a 14-on:10-off light cycle. The cells were harvested between mid to late-log phase and at the same time of day to avoid differences in sugar nucleotide levels due to the differences in growth phase. For *E. gracilis,* this represented ~6 days of growth, whilst for *P. parvum* late-log phase was usually achieved after ~14 days of growth. Cold ethanol (70%) was used to bring about cell lysis and to extract the target metabolites under very mild conditions [34], thus minimising degradation of the labile sugar nucleotides. Moreover, ethanol efficiently precipitates and inactivates cytosolic enzymes and prevents undesired enzymatic degradation. The samples were subjected to solid phase extraction (SPE) using EnviCarb graphitised carbon column [35] followed by LC-MS/MS based on a method by Pabst *et al* [36]. Authentic standards of sugar nucleotides were used to generate MRM transitions and to determine retention times. Where in doubt, co-injection of samples with standards was used to further confirm analyte identification. Internal standards [guanosine 5'-diphospho-α-D-glucose (GDP-α-D-Glc) for *P. parvum*, uridine 5'-diphospho-2-acetamido-2-deoxy-α-D-glucuronic acid (UDP-α-D-GlcNAcA) for *E. gracilis*, were used for quantification and allowed direct comparison of relative sugar nucleotide levels between *E. gracilis* and *P. parvum* (Figure 58).

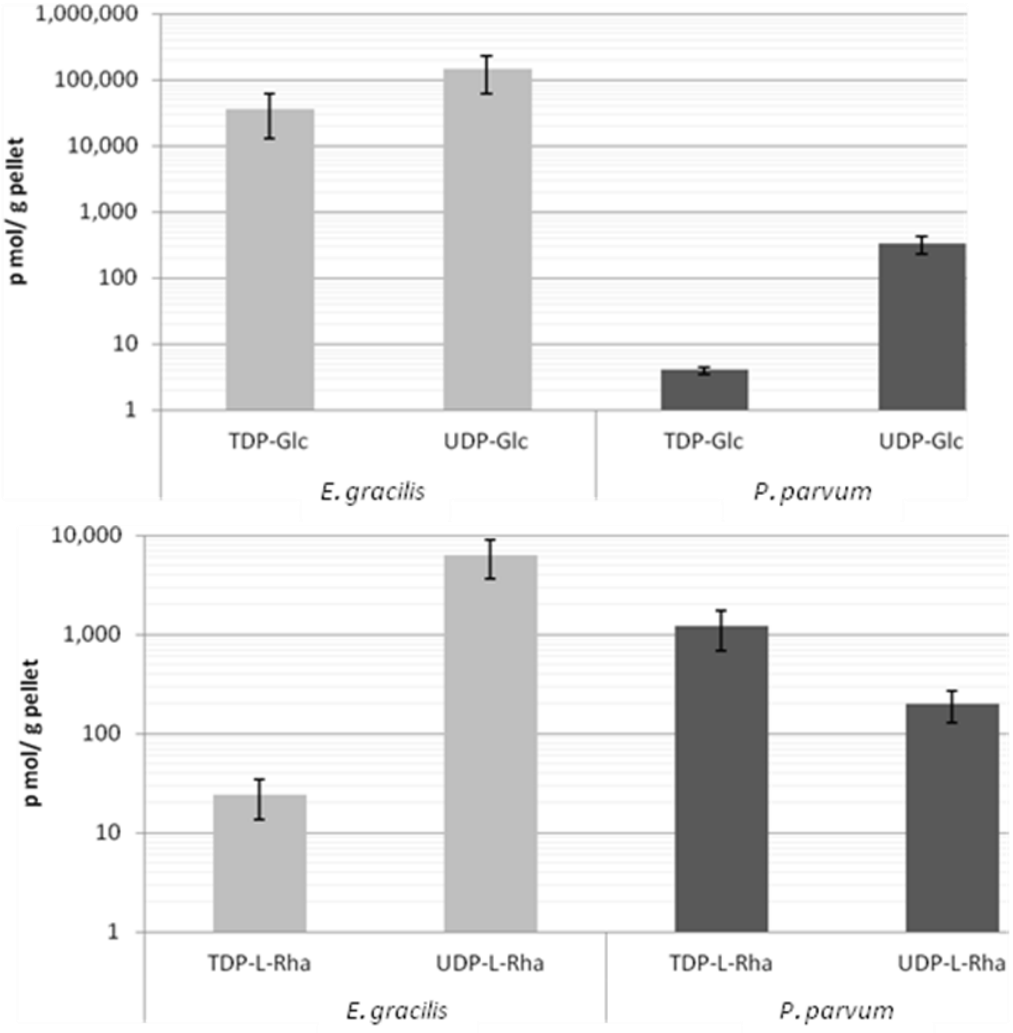


Figure 58 - Assessment of levels of TDP or UDP-activated glucose and L-Rha in *E. gracilis* and *P. parvum*. Concentrations of Top) TDP-α-D-Glc and UDP-α-D-Glc, Bottom) TDP-β-L-Rha and UDP-β-L-Rha.

LC-MS/MS results from biological triplicate show target NDP-sugars ranging from low pmol to mid nmol levels per gram of algal cells (Figure 58). *E. gracilis* contains approximately 4-fold more UDP-α-D-Glc than TDP-α-D-Glc, at the mid nmol range. Whilst levels of both TDP-α-D-Glc and UDP-α-D-Glc were lower in *P. parvum* at the low to high pmol range, levels of TDP-α-D-Glc were significantly lower, with UDP-α-D-Glc ~82 fold more abundant in *P. parvum*. These results appeared to have little correlation with the levels of activated L-Rha; *E. gracilis* contained approximately 262 times more UDP-β-L-Rha than TDP-β-L-Rha and *P. parvum* contained almost 6 times more TDP-β-L-Rha than UDP-β-L-Rha. For both organisms, levels of activated L-Rha ranged from 24 pmol to 6.3 nmol/g pellet. These results support our bioinformatic analysis, which indicates a bacterial-like L-Rha biosynthesis pathway in *P. parvum* and a plant-like L-Rha biosynthesis pathway in *E. gracilis*.

5.4 Discussion

The biosynthesis of L-Rha has been well described for bacteria [37], fungi [6, 38], plants [16, 17], and even viruses [18], so it is surprising that no effort has been previously made to decipher L-Rha biosynthetic pathways in algae, a previously reported source of L-Rha [20, 21, 25]. To fill this gap in knowledge, we examined the distribution of the bacterial RmlC enzyme and the plant UER1 and RHM enzymes across the algal groups. For this analysis 163 transcriptomes from MMETSP [39] or genomes from NCBI were used that represented a diverse mixture of all algal groups. RmlC was chosen as it is the only enzyme specific to the L-Rha pathway, with RmlB and RmlD paralogues found in alternative sugar nucleotide biosynthetic pathways [11, 40]. It is important to note that for this analysis, lack of transcripts could be due to lack of expression under the experimental growth conditions and doesn't necessarily equate to lack of gene in the genome of the organism. Equally with genomic analysis, lack of genes could be due to insufficient read depth.

We discovered that most algal groups utilize primarily a plant-like biosynthesis of L-Rha, with transcripts for UER1 and RHM identified throughout the glaucophytes, red algae, green algae, excavates, cryptophytes, alveolates and stramenopiles. In contrast, we found that the haptophytes show very little evidence for plant-like L-Rha biosynthesis, and instead operate a bacterial-like Rml biosynthesis pathway, with transcripts for a fusion of RmlC and RmlD abundant throughout. This fusion protein may represent a good example of gene fusion in early eukaryotes as discussed by Yin *et al* [41]. The Rhizaria are also an exception, with an unexpected fusion between bacterial RmlC and plant UER1 found throughout. The biochemical function of this fusion is unknown, but it is conceivable that the Rhizaria replaced redundant 3,5-epimerase functionality of UER1 with the corresponding epimerase functionality of bacterial RmlC. Although individual instances of bacterial RmlC homologs are found scattered across all algal groups, they appear to be more apparent in the Alveolata, with sixteen out of thirty-eight strains examined containing a homolog of either RmlC or an RmlC/D fusion. In addition, transcripts corresponding to the trifunctional RHM are more abundant in this superphylum. Interestingly, like the haptophytes, the Gymnodiniaceae family of dinoflagellates all contain RmlC/D fusions. This increased genetic diversity in the Alveolata could be due to the presence of tertiary or even quaternary endosymbiosis events found in the Dinoflagellata phylum [42].

5.4.1 Phylogenetic analysis and evolutionary implications

To investigate the evolutionary origin of both the plant-like UER1/RHM sequences and bacterial RmlC sequences, a maximum likelihood phylogenetic tree was constructed (Figure 57). The tree shows a clear divergence between bacterial- and plant-like pathways, with the haptophyte and fucoxanthin-containing dinoflagellate RmlC/D sequences branching with the corresponding bacterial sequences. All other groups of algae appear to have the plant-like UER1 or RHM machinery. This broad distribution of UER1 or RHM-like sequences would support an ancient evolutionary origin of this gene. On the other hand, the RmlC/D fusion is only observed consistently in the haptophytes and Gymnodiniaceae, and branches closely with bacterial sequences. This would suggest a bacterial origin of these sequences and may suggest a case of horizontal gene transfer (HGT) in the haptophytes or Gymnodiniaceae; although we cannot discount the possibility that this gene was present in the last common eukaryotic ancestor and subsequently lost in all other groups of algae (however unlikely this may be). The tree also shows that RmlC/D sequences from Gymnodiniaceae branch more closely with the haptophytes than bacteria. This suggests that one instance of HGT occurred to eukaryotes, rather than two independent instances to the haptophytes and Gymnodiniaceae. Given that the Gymnodiniaceae plastids are known to have come from tertiary endosymbiosis with haptophytes [29, 30], it seems likely that endosymbiotic gene transfer (EGT) of RmlC/D from haptophytes to this family of dinoflagellates has occurred. The additional presence of the UER1/RHM biosynthetic machinery in the Gymnodiniaceae supports the previous two propositions. The absence of the plant-like machinery in the haptophytes would suggest the loss of plant-like machinery sometime after secondary endosymbiosis. The acquisition of the bacterial Rml pathway may have occurred to replace the lost plant-like pathway, or alternatively if the bacterial pathway was acquired before UER1/RHM gene loss then gene loss of the plant-like pathway may be explained by genome reduction in this group.

Figure 57 also shows a close branching of sequences from the Rhizaria with the plant L-Rha biosynthetic pathway, even though all sequences from the Rhizaria contain a C-terminal RmlC fusion. The absence of RmlC in any green algae examined in this study would suggest that an independent HGT event occurred that incorporated RmlC into the genome of the Rhizaria after secondary endosymbiosis. We cannot, however, discount the possibility that the green algal symbiont had obtained this RmlC via HGT prior to secondary endosymbiosis, and passed it onto the Rhizaria via EGT. These evolutionary propositions are illustrated below (Figure 59).

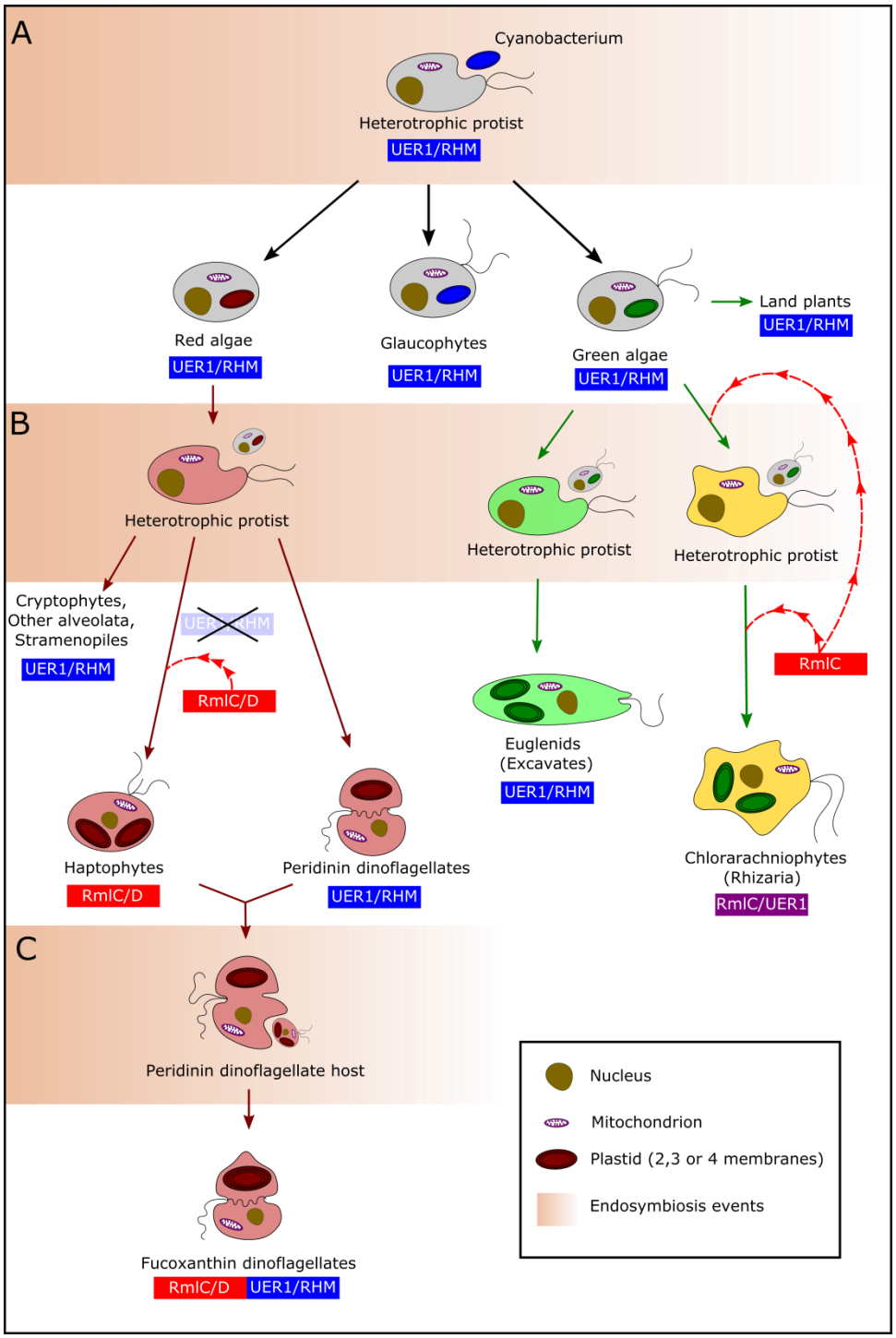


Figure 59 - A proposed evolutionary model of NDP-β-L-Rha biosynthesis in photosynthetic eukaryotes. The presence of the plant-like UER1 or RHM machinery is denoted by a blue box under the group name. The presence of bacterial RmlC or RmlC/D is denoted by a red box under the group name. The presence of a bacterial RmlC and plant UER1 fusion is denoted by a purple box under the group. A dashed red arrow indicates a likely horizontal gene transfer acquisition of a bacterial gene. A – Primary endosymbiosis between a heterotrophic protist host and cyanobacterial symbiont, leading to emergence of red algae, glaucophytes and green algae. B – Secondary endosymbiosis between a heterotrophic protist host and common red algal ancestor leading to the CASH group of algae; Two

separate endosymbiosis events with different green algae leading to the Excavates and Chlorarachniophytes. C – Example tertiary endosymbiosis event between a peridin dinoflagellate host and a haptophyte symbiont leading to the fucoxanthin-containing dinoflagellates (Gymnodiniaceae). Nuclei are represented by a brown circular shape; Mitochondria are represented by purple outlined ovals; Plastids are coloured according to their algal origin (i.e. red if derived from a red alga, green if derived from green algae). The colours of the cytoplasm in algae derived from secondary or tertiary endosymbiosis are coloured roughly per the scheme seen in Figure 56.

5.4.2 Quantification of glucose and rhamnose sugar nucleotides in *E. gracilis* and *P. parvum*

Two representative algae, *P. parvum* and *E. gracilis*, were selected to validate our bioinformatics findings. *P. parvum* is a haptophyte that was found in these studies to encode the bacterial RmlC/D pathway for L-Rha biosynthesis, whilst *E. gracilis* was found to encode the plant UER1 pathway. Given that enzymes of the bacterial pathway have been shown to exhibit some level of *in vitro* specificity for TDP-β-L-Rha formation [43, 44], we expected that *P. parvum* might contain higher levels of TDP-β-L-Rha than the UDP adduct. Conversely, as UER1 or RHM homologs of the plant pathway have been shown to exhibit some *in vitro* specificity for UDP activated L-Rha production [16, 18], we also expected that *E. gracilis* might contain higher levels of UDP-β-L-Rha than the TDP adduct. To evaluate these prospects, we set out to quantify the intracellular levels of NDP-β-L-Rha metabolites in the two-algal species.

Whilst UDP-α-D-Glc and TDP-α-D-Glc are commercially available, UDP-β-L-Rha and TDP-β-L-Rha were not. Instead, Martin Rejzek (Research Assistant – Rob Field group) kindly provided these compounds which had been chemically synthesized prior to this study.

As with other porous graphitic carbon sugar nucleotide profiling methods [36], the more hydrophobic nature of the thymidine with respect to uridine led to good separation of the corresponding TDP- and UDP-sugar species. More importantly though, separation was also achieved for L-Rha and glucose species bearing the same nucleotide base. One possibility for the difference in choice of activating base between the two algae was that the availability of the starting compound for the L-Rha biosynthetic pathways, either UDP-α-D-Glc or TDP-α-D-Glc. For this reason, we also investigated the levels of these two sugar nucleotides. *E. gracilis* was found to have ~4 times more UDP-α-D-Glc than TDP-α-D-Glc, whilst *P. parvum* contained approximately ~82 times more UDP-α-D-Glc than TDP-α-D-Glc. *E. gracilis* was then found to have ~280 times more UDP-β-L-Rha than TDP-β-L-Rha, supporting our bioinformatics-based

assignment of this species to the plant UER1/RHM pathway. On the other hand, *P. parvum* was found to contain ~6 times more TDP-β-L-Rha than UDP-β-L-Rha, supporting our assignment of *P. parvum* to the bacterial Rml TDP-β-L-Rha biosynthesis pathway. Interestingly though, both species displayed appreciable levels of both TDP and UDP activated L-Rha. This would support *in vitro* work on both pathways that shows reduced, but still measurable, activities for the alternative nucleotide in each pathway [16, 17, 44].

Turnock *et al* [34] showed that *Trypanosoma cruzi* produces the UDP- activated form of L-Rha, showing the pathway precursor, UDP-α-D-Glc, is ~28 times more abundant than UDP-β-L-Rha in this species. This figure is strikingly similar to the data we present here for *E. gracilis*, an evolutionary neighbour of the trypanosomes [46, 47], which has ~23 times more UDP-α-D-Glc than UDP-β-L-Rha (Figure 58). Interestingly, for *P. parvum*, TDP-β-L-Rha is ~250 x more abundant than TDP-α-D-Glc, suggesting an as yet undefined but crucial role for L-Rha in *Prymnesium* glycobiology. These findings are somewhat at odds with the early findings of Marker [25], who claimed that L-Rha is only present in trace amounts in the extracellular cell preparations of *P. parvum*, with none observed inside the cell. For *E. gracilis*, initial work by Barras and Stone [21], among others [22], confirmed the presence of L-Rha in the pellicle and mucus of *E. gracilis*. This work was later built upon by Nakano *et al* [23], who showed L-Rha to be the most abundant monosaccharide in the pellicle of *E. gracilis*. The lack of correlation between relative levels of activated glucose to L-Rha sugar nucleotides in turn supports our bioinformatic findings that *E. gracilis* and *P. parvum* have evolved independent pathways for L-Rha biosynthesis.

5.5 Conclusion

Although a sugar nucleotide profiling of the other organisms discussed in this study is warranted, based on our combined use of bioinformatics analysis and LC-MS/MS profiling, we show that L-Rha biosynthetic genes are widespread throughout the diverse algal groups. We show that the plant UER1/RHM pathways are abundant throughout the glaucophytes, red algae, green algae, excavates, cryptophytes, alveolates and stramenopiles but are surprisingly absent from haptophytes, which contain a fusion of RmlC and RmlD from the bacterial pathway. Using phylogenetic analysis, it seems likely that a HGT event from bacteria led to the haptophytes acquiring this RmlC/D bacterial machinery, and propose that it was then passed on to the Gymnodiniaceae through EGT to the dinoflagellate host. Using sugar nucleotide profiling we confirm our bioinformatics findings by showing that the haptophyte *P. parvum* contains mainly TDP-β-L-Rha, while the excavate *E. gracilis* contains primarily UDP-β-L-Rha. Although *in vitro* analysis of enzyme specificities is warranted in each case, this is often difficult due to poor expression of target enzymes, or lack of availability of substrates, which has often been the case for UDP-6-deoxy-L-*lyxo*-4-hexulose [17]. We show here that the combined use of bioinformatics and LC-MS/MS based profiling provides an alternative way to answer questions surrounding the evolution of enzymatic pathways associated with rhamnose-based sugar nucleotide production.

5.6 Experimental

TDP-α-D-Glc, UDP-α-D-Glc and GDP-α-D-Glc were obtained commercially from Sigma Aldrich (Haverhill, UK). UDP-α-D-GlcNAcA was prepared as previously described [48], and TDP-β-L-Rha and UDP-β-L-Rha were provided as gifts by Martin Rejzek (Research Assistant of Rob Field group).

5.6.1 *Euglena gracilis* axenic *cell culture* [46]

Euglena gracilis var. *saccharophila* Klebs (strain 1224/7a) was obtained from the Culture Collection of Algae and Protozoa (CCAP) (http://www.ccap.ac.uk/). Stock culture was treated with antibiotics according to a method suggested by CCAP to produce an axenic culture (https://www.ccap.ac.uk/documents/Antibiotic_treatment.pdf) with small modifications to the antibiotic components (only Cefotaxime, Carbenicillin and Kanamycin were used). The stock culture was treated with 0%, 0.5% and 1% of the antibiotic mixture in the recommended 1x EG + 1x JM (EG: *Euglena gracilis* medium, JM: Jaworski's medium) media for *Euglena gracilis* and subsequently inoculated into fresh 1x EG + 1x JM media at the following time intervals: 24, 48 and 72 hours. The culture was examined by microscopy and plating on 1x EG + 1x JM agar to confirm the production of an axenic culture.

Batch cultures (3 biological replicates) were grown essentially as described before [24]. In brief, cells were grown at 22 °C on a 14:10 light cycle with a light intensity of 100 $\mu mol.m^{-2}.s^{-1}$. Mid-log phase (OD_{600} = 1.1 in about 6 days) cultures were harvested. Cells were pelleted by centrifugation (6,750 x *g* for 20 min at 4 °C). The pellet was re-suspended in ice-cold phosphate-buffered saline (PBS, 200 ml) and centrifuged again (6,750 x *g* for 20 min at 4 °C). The pellet was transferred into tared centrifuge vial (Oak Ridge) using PBS (25 ml), centrifuged (6,750 x *g* for 20 min at 4 °C) and the supernatant was carefully decanted before weighing out the wet pellet. UDP-α-D-GlcNAcA was added to the cell pellet as internal standard (1.46 nmol / g wet pellet). The cells were lysed straight away without flash freezing and/or cold storage.

5.6.2 *Prymnesium parvum* axenic cell culture

Prymnesium parvum (strain 946/6) was obtained from the Culture Collection of Algae and Protozoa (CCAP) (http://www.ccap.ac.uk/) and maintained in the recommended f/2 –Si media. Stock cultures were treated with Carbenicillin (100 µg/ml) in order to obtain axenic cultures, which were judged to be axenic by optical microscopy. Batch cultures (3 biological repeats) were grown at 22 °C on a 14:10 light cycle with a light intensity of 100 $\mu mol.m^{-2}.s^{-1}$, as previously described [49]. Under these conditions, cell densities of ~3 x 10^6 cells ml^{-1} could be

achieved after 12-16 days of growth. Cells were pelleted by centrifugation (6,748 x *g*, 20 min, 4 °C). The pellet was transferred into tared centrifuge vial (Oak Ridge) using ice-cold phosphate-buffered saline (PBS, 20 ml), centrifuged (12 857 x *g*, 20 min, 4 °C) to give pellet. GDP-α-D-Glc was added (1.54 nmol / g wet pellet). The cells were lysed straight away without flash freezing and/or cold storage.

5.6.3 Sugar-nucleotide extraction and profiling [50]

Pelleted cells of *E. gracilis* and *P. parvum* containing known amount of the appropriate internal standards were lysed with cold (-20 °C) 70% ethanol (20 ml) in an ice bath for 1 h with occasional shaking / vortexing. The cell debris was removed by centrifugation (28 928 x *g*, 20 min, 4 °C) and the supernatant was transferred into a glass round-bottom flask (100 ml). Ethanol was evaporated at reduced pressure and ambient temperature and the aqueous residue was freeze dried. At this stage the sample can be stored at -80 °C for any length of time before the next step.

Lipophilic components were removed by partitioning the sample between water and butan-1-ol [34]. The sample was dissolved in 9% aqueous butan-1-ol (3 x 2 ml) and transferred into a glass vial (10 ml volume). The solution was extracted with 90% butan-1-ol (3 x 2 ml) to remove lipids (but also chlorophyll, and insoluble polysaccharides such as paramylon forming middle layer). The bottom layer was collected and extracted again. Centrifugation was used to speed up the separation of the layers (200 - 800 x *g*, 4 °C, 5 min). The clear aqueous layer was collected and freeze dried in a pear-shaped flask (foaming may appear under vacuum). Samples were stored at -80 °C before the next step.

SPE of sugar nucleotides was performed essentially as described by Rabina and co-workers [35]. A graphitised carbon column (EnviCarb, Supelco, 250 mg, 3 ml) was conditioned by washing with 80% aqueous acetonitrile containing 0.1 % trifluoroacetic acid (3 ml) followed by water (2 ml). The sample was dissolved in ammonium bicarbonate (5 mM, 500 µl) and applied on the SPE column. The column was washed with water (2 ml), followed by 25% aqueous acetonitrile (2 ml), and 50 mM triethylammonium acetate buffer (pH 7.0, 2 ml). Finally, the sugar nucleotides were eluted with 50 mM triethylammonium acetate buffer pH 7.0 containing 25% acetonitrile (1.5 ml). The sample was filtered using 0.45 µm disc filters (PTFE) and freeze dried. Samples were stored at -80 °C prior to LC-MS/MS analysis.

LC-MS/MS profiling of sugar nucleotides was performed on a Xevo TQ-S tandem quadrupole mass spectrometer (Waters) operated in MRM mode coupled to an Acquity UPLC. ESI-MS/MS

analysis was performed in negative ion mode using a source with a capillary voltage of 1.5 kV, 500 °C desolvation temperature, 1000 $l.hr^{-1}$ desolvation gas, 150 $l.hr^{-1}$ cone gas, and 7 bar nebulizer pressure. MRM transitions for sugar nucleotide standards in negative ESI mode were generated using IntelliStart software (Table 9). Samples (10 µM) were introduced at 10 µl/min combined with a flow from the HPLC pump typical of an LC run. Once LC retention times of standards have been established, the mass transitions were collected in time-windows centred on the relevant peaks, to avoid collecting excessive numbers of transitions simultaneously. MassLynx software (Waters) was used to collect, to analyse and to process data.

Liquid chromatography separation of sugar nucleotides was achieved on a surface-conditioned PGC column (Hypercarb, Thermo Scientific, dimensions 1 x 100 mm, particle size 5 µm) equipped with a column guard (Hypercarb, 5 µm, 1 x 10 mm). Sugar nucleotides were eluted using mobile phase A: formic acid 0.3% brought to pH 9.0 with ammonia and mobile phase B: acetonitrile using the following multistep gradient at a flow rate 80 µl/min: 0 min: 2% B; 20 min: 15% B; 26 min: 50% B; 27 min: 90% B; 30 min: 90% B; 31 min: 2% B; 50 min: 2% B. Available sugar nucleotide standards (10 µM) were injected (5 µl) to determine retention time (Table 9).

Sugar Nucleotide	Relative Retention time	MRM transitions	Fragment
UDP-α-D-Glc	1.00	565 → 323 565 → 79	$[NMP-H]^-$ $[H_3PO_4-H_3O]^-$
UDP-α-D-GlcNAcA	0.89	620 → 403 620 → 159	$[NDP-H]^-$ $[H_4P_2O_7-H_3O]^-$
UDP-β-L-Rha	0.84	549 → 323 549 → 159	$[NMP-H]^-$ $[H_4P_2O_7-H_3O]^-$
TDP-α-D-Glc	1.39	563 → 321 563 → 241	$[NMP-H]^-$ $[Glc-1-P-H-H_2O]^-$
TDP-β-L-Rha	1.35	547 → 321 547 → 225	$[NMP-H]^-$ $c[Rha-1-P-H-H_2O]^-$
GDP-α-D-Glc	1.56	604 → 362 604 → 241	$[NMP-H]^-$ $c[Glc-1-P-H-H_2O]^-$

Table 9 - Relative retention times and MRM transitions of sugar nucleotides standards

Limit of detection was determined to be 10 fmol on column using a serial dilution of UDP-α-D-Glc. Samples of extracted sugar nucleotides were reconstituted in buffer A (25 µL) and injected (5 µl, 20 % of total) using partial loop injection. Analysis of 3 biological replicates was performed. Where in doubt, co-injection of sample with appropriate standard sugar nucleotide was used for positive identification. Data processing was performed using MassLynx (Waters) software. Although between runs there were significant differences in absolute retention times of standards, relative retentions were reasonably reproducible (Table 9). To ensure maximum retention time (R_f) stability, after a batch of samples, the PGC column had to be regenerated and its performance was tested using UDP-α-D-Glc as a standard. The regeneration and column performance steps were performed using a standard HPLC (Ultimate 3000, Dionex) system with UV detection at 265 nm. Column performance was tested first before regeneration steps at flow rate 80 µl/min by injecting UDP-α-D-Glc (5 µL, 10 µM) standard and elution using mobile phases A and B as mentioned earlier. The column was then washed for 3 hrs with mobile phase C (acetonitrile 80%, water 20 %, TFA 0.1%) [51] followed by water (5 column volumes). Next, the PGC column was reduced at flow rate 80 µl/min with freshly prepared sodium sulphite (100 mM) for 24 h [36] followed by MQ water (5 column volumes). The column was then washed at 80 µl/min with high acetonitrile (90% B, 10% A) for

30 min and equilibrated with 2% B, 98 % A for 10 min. The column performance was tested again by injecting UDP-α-D-Glc (5 µl, 10 µM) standard. The column was stored in 2% B, 98 % A.

5.6.4 Bioinformatic analysis

For the identification of NDP-β-L-Rha biosynthetic pathways, BLASTp [52] analysis was carried out against the transcriptomes (MMETSP) or genomes (NCBI) of representative algae from all algal groups Supplementary Material 3. Protein sequences for RmlC (NP_217982.1), UER1 (NP_564806.1) and RHM1 (NP_177978.1) were used as consensus sequences. Hits with *E*-values ≤ 1E-10 were then manually analysed for conserved domains before being assigned as a hit.

For phylogenetic analysis of NDP-β-L-Rha biosynthesis hits, multiple sequence alignments were generated using the default settings of MAFFT [32], including additional sequences from bacteria and plants. Regions of poor alignment were inspected for manually and their respective sequences were removed. An unrooted maximum likelihood tree was then generated using MEGA7 with 100 bootstraps. The final tree was based on 135 ungapped amino acid residues and was made up of 122 sequences.

5.7 References

1. Ma, Y.; Stern, R. J.; Scherman, M. S.; Vissa, V. D.; Yan, W.; Jones, V. C.; Zhang, F.; Franzblau, S. G.; Lewis, W. H.; McNeil, M. R., Drug targeting *Mycobacterium tuberculosis* cell wall synthesis: genetics of dTDP-rhamnose synthetic enzymes and development of a microtiter plate-based screen for inhibitors of conversion of dTDP-glucose to dTDP-rhamnose. *Antimicrob. Agents Chemother.* **2001,** 45, (5), 1407-16.
2. Dong, C.; Beis, K.; Giraud, M. F.; Blankenfeldt, W.; Allard, S.; Major, L. L.; Kerr, I. D.; Whitfield, C.; Naismith, J. H., A structural perspective on the enzymes that convert dTDP-D-glucose into dTDP-L-rhamnose. *Biochem. Soc. Trans.* **2003,** 31, (Pt 3), 532-6.
3. Pettolino, F.; Sasaki, I.; Turbic, A.; Wilson, S. M.; Bacic, A.; Hrmova, M.; Fincher, G. B., Hyphal cell walls from the plant pathogen *Rhynchosporium secalis* contain (1,3/1,6)-β-D-glucans, galacto- and rhamnomannans, (1,3;1,4)-β-D-glucans and chitin. *FEBS J.* **2009,** 276, (14), 3698-3709.
4. Ridley, B. L.; O'Neill, M. A.; Mohnen, D., Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry* **2001,** 57, (6), 929-67.
5. Piacente, F.; Gaglianone, M.; Laugieri, M.; Tonetti, M., The autonomous glycosylation of large DNA viruses. *Int. J. Mol. Sci.* **2015,** 16, (12), 26169.
6. Martinez, V.; Ingwers, M.; Smith, J.; Glushka, J.; Yang, T.; Bar-Peled, M., Biosynthesis of UDP-4-keto-6-deoxyglucose and UDP-rhamnose in pathogenic fungi *Magnaporthe grisea* and *Botryotinia fuckeliana*. *J. Biol. Chem.* **2012,** 287, (2), 879-892.
7. Jiang, X. M.; Neal, B.; Santiago, F.; Lee, S. J.; Romana, L. K.; Reeves, P. R., Structure and sequence of the rfb (O antigen) gene cluster of *Salmonella serovar* typhimurium (strain LT2). *Mol. Microbiol.* **1991,** 5, (3), 695-713.
8. Yamashita, Y.; Shibata, Y.; Nakano, Y.; Tsuda, H.; Kido, N.; Ohta, M.; Koga, T., A novel gene required for rhamnose-glucose polysaccharide synthesis in *Streptococcus mutans*. *J. Bacteriol.* **1999,** 181, (20), 6556-6559.
9. Mistou, M.-Y.; Sutcliffe, I. C.; van Sorge, N. M., Bacterial glycobiology: rhamnose-containing cell wall polysaccharides in Gram-positive bacteria. *FEMS Microbiol. Rev.* **2016,** 40, (4), 464-479.
10. Giraud, M. F.; Naismith, J. H., The rhamnose pathway. *Curr. Opin. Struct. Biol.* **2000,** 10, (6), 687-96.
11. Thibodeaux, C. J.; Melançon, C. E.; Liu, H. W., Natural-product sugar biosynthesis and enzymatic glycodiversification. *Angew. Chem. Int. Ed.* **2008,** 47, (51), 9814-9859.

12. Singh, S.; Phillips Jr, G. N.; Thorson, J. S., The structural biology of enzymes involved in natural product glycosylation. *Nat. Prod. Rep.* **2012,** 29, (10), 1201-1237.
13. Allard, S. T.; Giraud, M. F.; Whitfield, C.; Graninger, M.; Messner, P.; Naismith, J. H., The crystal structure of dTDP-D-Glucose 4,6-dehydratase (RmlB) from *Salmonella enterica* serovar Typhimurium, the second enzyme in the dTDP-L-rhamnose pathway. *J. Mol. Biol.* **2001,** 307, (1), 283-95.
14. Blankenfeldt, W.; Kerr, I. D.; Giraud, M. F.; McMiken, H. J.; Leonard, G. A.; Whitfield, C.; Messner, P.; Graninger, M.; Naismith, J. H., Variation on a theme of SDR: dTDP-6-deoxy-L-lyxo-4-hexulose reductase (RmlD) shows a new Mg^{2+}-dependent dimerisation mode in a well-known enzyme family. *Structure* **2002,** 10, 773-786.
15. Dong, C.; Major, L. L.; Srikannathasan, V.; Errey, J. C.; Giraud, M. F.; Lam, J. S.; Graninger, M.; Messner, P.; McNeil, M. R.; Field, R. A.; Whitfield, C.; Naismith, J. H., RmlC, a C3' and C5' carbohydrate epimerase, appears to operate via an intermediate with an unusual twist boat conformation. *J. Mol. Biol.* **2007,** 365, (1), 146-159.
16. Oka, T.; Nemoto, T.; Jigami, Y., Functional analysis of Arabidopsis thaliana RHM2/MUM4, a multidomain protein involved in UDP-D-glucose to UDP-L-rhamnose conversion. *J. Biol. Chem.* **2007,** 282, (8), 5389-403.
17. Han, X.; Qian, L.; Zhang, L.; Liu, X., Structural and biochemical insights into nucleotide–rhamnose synthase/epimerase–reductase from Arabidopsis thaliana. *Biochim. Biophys. Acta* **2015,** 1854, (10, Part A), 1476-1486.
18. Chothi, M. P.; Duncan, G. A.; Armirotti, A.; Abergel, C.; Gurnon, J. R.; Van Etten, J. L.; Bernardi, C.; Damonte, G.; Tonetti, M., Identification of an L-rhamnose synthetic pathway in two nucleocytoplasmic large DNA viruses. *J. Virol.* **2010,** 84, (17), 8829-8838.
19. Watt, G.; Leoff, C.; Harper, A. D.; Bar-Peled, M., A bifunctional 3,5-epimerase/4-keto reductase for nucleotide-rhamnose synthesis in Arabidopsis. *Plant Physiol.* **2004,** 134, (4), 1337-46.
20. Lahaye, M.; Robic, A., Structure and functional properties of ulvan, a polysaccharide from green seaweeds. *Biomacromolecules* **2007,** 8, (6), 1765-74.
21. Barras, D. R.; Stone, B. A., The chemical composition of the pellicle of *Euglena gracilis* var. *bacillaris*. *Biochem. J* **1965,** 97, 14 - 15.
22. Cogburn, J. N.; Schiff, J. A., Purification and properties of the mucus of *Euglena Gracilis* (Euglenophyceae). *J. Phycol.* **1984,** 20, (4), 533-544.
23. Nakano, Y.; Urade, Y.; Urade, R.; Kitaoka, S., Isolation, purification, and characterization of the pellicle of *Euglena gracilis*. *J. Biochem.* **1987,** 102, (5), 1053-1063.

24. O'Neill, E. C.; Trick, M.; Hill, L.; Rejzek, M.; Dusi, R. G.; Hamilton, C. J.; Zimba, P. V.; Henrissat, B.; Field, R. A., The transcriptome of *Euglena gracilis* reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. *Molecular BioSystems* **2015,** 11, (10), 2808-2820.

25. Marker, A., Extracellular carbohydrate liberation in the flagellates *Isochrysis galbana* and *Prymnesium parvum*. *J. Mar. Biol. Assoc. U.K.* **1965,** 45, (03), 755-772.

26. Keeling, P. J., The endosymbiotic origin, diversification and fate of plastids. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **2010,** 365, (1541), 729-748.

27. Chan, C. X.; Bhattacharya, D., The origin of plastids. *Nature Education* **2010,** 3, (9), 84.

28. Yoon, H. S.; Hackett, J. D.; Van Dolah, F. M.; Nosenko, T.; Lidie, K. L.; Bhattacharya, D., Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. *Mol. Biol. Evol.* **2005,** 22, (5), 1299-1308.

29. Gabrielsen, T. M.; Minge, M. A.; Espelund, M.; Tooming-Klunderud, A.; Patil, V.; Nederbragt, A. J.; Otis, C.; Turmel, M.; Shalchian-Tabrizi, K.; Lemieux, C.; Jakobsen, K. S., Genome evolution of a tertiary dinoflagellate plastid. *PLOS ONE* **2011,** 6, (4), e19132.

30. Tengs, T.; Dahlberg, O. J.; Shalchian-Tabrizi, K.; Klaveness, D.; Rudi, K.; Delwiche, C. F.; Jakobsen, K. S., Phylogenetic analyses indicate that the 19' hexanoyloxy-fucoxanthin-containing dinoflagellates have tertiary plastids of haptophyte origin. *Mol. Biol. Evol.* **2000,** 17, (5), 718-729.

31. Timmis, J. N.; Ayliffe, M. A.; Huang, C. Y.; Martin, W., Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **2004,** 5, (2), 123-135.

32. Katoh, K.; Toh, H., Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinformatics* **2008,** 9, (4), 286-98.

33. Humar, S.; Stecher, G.; Tamura, K., MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016,** 33, 1870-1874.

34. Turnock, D. C.; Ferguson, M. A., Sugar nucleotide pools of *Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania major*. *Eukaryot. Cell* **2007,** 6, (8), 1450-1463.

35. Räbinä, J.; Mäki, M.; Savilahti, E. M.; Järvinen, N.; Penttilä, L.; Renkonen, R., Analysis of nucleotide sugars from cell lysates by ion-pair solid-phase extraction and reversed-phase high-performance liquid chromatography. *Glycoconjugate J.* **2001,** 18, (10), 799-805.

36. Pabst, M.; Grass, J.; Fischl, R.; Léonard, R.; Jin, C.; Hinterkörner, G.; Borth, N.; Altmann, F., Nucleotide and nucleotide sugar analysis by liquid chromatography-electrospray

ionization-mass spectrometry on surface-conditioned porous graphitic carbon. *Anal. Chem.* **2010,** 82, (23), 9782-9788.

37. Mäki, M.; Renkonen, R., Biosynthesis of 6-deoxyhexose glycans in bacteria. *Glycobiology* **2004,** 14, (3), 1R-15R.

38. Santhanam, P.; Boshoven, J. C.; Salas, O.; Bowler, K.; Islam, M. T.; Saber, M. K.; van den Berg, G. C. M.; Bar-Peled, M.; Thomma, B. P. H. J., Rhamnose synthase activity is required for pathogenicity of the vascular wilt fungus *Verticillium dahliae*. *Mol. Plant Pathol.* **2017,** 18, (3), 347-362.

39. Keeling, P. J.; Burki, F.; Wilcox, H. M.; Allam, B.; Allen, E. E.; Amaral-Zettler, L. A.; Armbrust, E. V.; Archibald, J. M.; Bharti, A. K.; Bell, C. J.; Beszteri, B.; Bidle, K. D.; Cameron, C. T.; Campbell, L.; Caron, D. A.; Cattolico, R. A.; Collier, J. L.; Coyne, K.; Davy, S. K.; Deschamps, P.; Dyhrman, S. T.; Edvardsen, B.; Gates, R. D.; Gobler, C. J.; Greenwood, S. J.; Guida, S. M.; Jacobi, J. L.; Jakobsen, K. S.; James, E. R.; Jenkins, B.; John, U.; Johnson, M. D.; Juhl, A. R.; Kamp, A.; Katz, L. A.; Kiene, R.; Kudryavtsev, A.; Leander, B. S.; Lin, S.; Lovejoy, C.; Lynn, D.; Marchetti, A.; McManus, G.; Nedelcu, A. M.; Menden-Deuer, S.; Miceli, C.; Mock, T.; Montresor, M.; Moran, M. A.; Murray, S.; Nadathur, G.; Nagai, S.; Ngam, P. B.; Palenik, B.; Pawlowski, J.; Petroni, G.; Piganeau, G.; Posewitz, M. C.; Rengefors, K.; Romano, G.; Rumpho, M. E.; Rynearson, T.; Schilling, K. B.; Schroeder, D. C.; Simpson, A. G. B.; Slamovits, C. H.; Smith, D. R.; Smith, G. J.; Smith, S. R.; Sosik, H. M.; Stief, P.; Theriot, E.; Twary, S. N.; Umale, P. E.; Vaulot, D.; Wawrik, B.; Wheeler, G. L.; Wilson, W. H.; Xu, Y.; Zingone, A.; Worden, A. Z., The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLOS Biology* **2014,** 12, (6), e1001889.

40. Liu, H. W.; Thorson, J. S., Pathways and mechanisms in the biogenesis of novel deoxysugars by bacteria. *Annu. Rev. Microbiol.* **1994,** 48, 223-56.

41. Yin, Y.; Huang, J.; Gu, X.; Bar-Peled, M.; Xu, Y., Evolution of plant nucleotide-sugar interconversion enzymes. *PLOS ONE* **2011,** 6, (11), e27995.

42. Morden, C. W.; Sherwood, A. R., Continued evolutionary surprises among dinoflagellates. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** 99, (18), 11558-11560.

43. Teramoto, M.; Zhang, Z.; Shizuma, M.; Kawasaki, T.; Kawarabayasi, Y.; Nakamura, N., The thermostable enzyme genes of the dTDP-L-rhamnose synthesis pathway (rmlBCD) from a thermophilic archaeon. In *Advances in Applied Biotechnology*, Petre, M., Ed. InTech: Rijeka, **2012**, p Ch. 12.

44. Feng, L.; Shou, Q.; Butcher, Rebecca A., Identification of a dTDP-rhamnose biosynthetic pathway that oscillates with the molting cycle in *Caenorhabditis elegans*. *Biochem. J.* **2016,** 473, (11), 1507-1521.

45. Sabesan, S.; Neira, S., Synthesis of glycosyl phosphates and azides. *Carbohydr. Res.* **1992,** 223, 169-185.

46. O'Neill, E. C.; Trick, M.; Henrissat, B.; Field, R. A., Euglena in time: Evolution, control of central metabolic processes and multi-domain proteins in carbohydrate and natural product biochemistry. *Perspectives in Science* **2015,** 6, 84-93.

47. Schwartzbach, S.; Shigeoka, S., Euglena: biochemistry, cell and molecular biology. Springer: 2017; Vol. 979.

48. Rejzek, M.; Mukhopadhyay, B.; Wenzel, C. Q.; Lam, J. S.; Field, R. A., Direct oxidation of sugar nucleotides to the corresponding uronic acids: TEMPO and platinum-based procedures. *Carbohydr. Res.* **2007,** 342, (3–4), 460-466.

49. Wagstaff, B.; Vladu, I.; Barclay, J.; Schroeder, D.; Malin, G.; Field, R., Isolation and characterization of a double stranded DNA megavirus infecting the toxin-producing haptophyte *Prymnesium parvum*. *Viruses* **2017,** 9, (3), 40.

50. Rejzek, M.; Hill, L.; Hems, E. S.; Kuhaudomlarp, S.; Wagstaff, B. A.; Field, R. A., Profiling of sugar nucleotides. *Methods Enzymol.* **2017**, 597, 209-39.

51. Behmuller, R.; Forstenlehner, I. C.; Tenhaken, R.; Huber, C. G., Quantitative HPLC-MS analysis of nucleotide sugars in plant cells following off-line SPE sample preparation. *Anal. Bioanal. Chem.* **2014,** 406, (13), 3229-3237.

52. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic local alignment search tool. *J. Mol. Biol.* **1990,** 215, (3), 403-10.

5.8 Supplementary Information

Kingdom	Phylum	Class	Family	Genus	Species	Strain	Database and identifier	Sequence
GLAUCOPHYTES	Glaucophyta	Glaucocystophyceae	Glaucosphaeraceae	Cyanoptyche	gloeocystis	SAG4.97	MMETSP1086	CAMPEP_0196658872
	Glaucophyta	Glaucophyceae	Gloeochaetaceae	Gloeochaete	wittrockiana	SAG46.84	MMETSP1089	CAMPEP_0184350920, CAMPEP_0184349520
GREEN ALGAE	Chlorophyta	Chlorodendrophyceae	Chlorodendraceae	Tetraselmis	striata	LANL1001	MMETSP0817, MMETSP0818, MMETSP0819, MMETSP0820	CAMPEP_0200914898
	Chlorophyta	Chlorophyceae	Dunaliellaceae	Dunaliella	tertiolecta	CCMP1320	MMETSP1126, MMETSP1127, MMETSP1128	CAMPEP_0187373064
	Chlorophyta	Chlorophyceae	Chlamydomonadaceae	Chlamydomonas	reinhardtii	CC-503 cw92 mt+	https://www.ncbi.nlm.nih.gov/nuccore/158276217	XP_001695032.1
	Chlorophyta	Mamiellophyceae	Mamiellaceae	Micromonas	sp.	RCC472	MMETSP1084, MMETSP1387	
	Chlorophyta	Mamiellophyceae	Mamiellaceae	Micromonas	sp.	NEPCC29	MMETSP1386, MMETSP1082	
	Chlorophyta	Mamiellophyceae	Mamiellaceae	Micromonas	sp.	CCMP2099	MMETSP1390, MMETSP0802	CAMPEP_0190189836
	Chlorophyta	Mamiellophyceae	Mamiellaceae	Micromonas	pusilla	CCMP1545	https://www.ncbi.nlm.nih.gov/genome/?term=txid564608[Organism:noexp]	XP_003057320.1
	Chlorophyta	Mamiellophyceae	Mamiellaceae	Bathycoccus	prasinos	RC1105	https://www.ncbi.nlm.nih.gov/genome/12309	XP_007514013.1, XP_007508076.1
	Chlorophyta	Mamiellophyceae	Mamiellaceae	Ostreococcus	tauri		https://www.ncbi.nlm.nih.gov/genome/373	CEG01892.1
	Chlorophyta	Pyramimonadophyceae	Halosphaeraceae	Pyramimonas	parkeae	CCMP726	MMETSP0058, MMETSP0059	CAMPEP_0191497618, CAMPEP_0191502286
	Chlorophyta	Trebouxiophyceae	Chlorellaceae	Auxenochlorella	protothecoides	sp 0710	http://www.ncbi.nlm.nih.gov/genome?LinkName=nuccore_genome&from_uid=667612142	XP_011395817.1
	Chlorophyta	Trebouxiophyceae	Chlorellaceae	Chlorella	variabilis	NC64A	https://www.ncbi.nlm.nih.gov/genome/694	XP_005851985.1

	Chlorophyta	Unknown	Unknown	Picocystis	salinarum	CCMP1897	MMETSP1159, MMETSP0807	CAMPEP_0190751704
RED ALGAE	Rhodophyta	Compsopogonophyceae	Compsopogonaceae	Compsopogon	coeruleus	SAG 36.94	MMETSP0312	
	Rhodophyta	Compsopogonophyceae	Erythrotrichiaceae	Madagascaria	erythrocladiodes	CCMP3234	MMETSP1450	CAMPEP_0198320684
	Rhodophyta	Cyanidiophyceae	Galdieriaceae	Galdieria	sulphuraria	074W	http://www.ncbi.nlm.nih.gov/genome/405	XP_005707655.1
	Rhodophyta	Florideophyceae	Gigartinaceae	Chondrus	crispus	Stackhouse	http://www.ncbi.nlm.nih.gov/genome/12106	
	Rhodophyta	Porphyridiophyceae	Porphyridiaceae	Erythrolobus	australicus	CCMP3124	MMETSP1353	
	Rhodophyta	Porphyridiophyceae	Porphyridiaceae	Erythrolobus	madagascarensis	CCMP3276	MMETSP1354	
	Rhodophyta	Porphyridiophyceae	Porphyridiaceae	Timspurckia	oligopyrenoides	CCMP3278	MMETSP1172	
	Rhodophyta	Rhodellophyceae	Porphyridiaceae	Porphyridium	aerugineum	SAG 1380-2	MMETSP0313	
	Rhodophyta	Rhodellophyceae	Rhodellaceae	Rhodella	maculata	CCMP736	MMETSP0167, MMETSP0314	CAMPEP_0191515942, CAMPEP_0191515916
	Rhodophyta	Rhodellophyceae	Stylonemataceae	Rhodosorus	marinus	CCMP 769	MMETSP0011	
	Rhodophyta	Rhodellophyceae	Stylonemataceae	Rhodosorus	marinus	UTEX LB 2760	MMETSP0315	
EXCAVATES	Euglenozoa	Euglenophyceae	Eutreptiaceae	Eutreptiella	gymnastica	NIES-381	MMETSP0039	
	Euglenozoa	Euglenophyceae	Eutreptiaceae	Eutreptiella	gymnastica-like	CCMP1594	MMETSP0809, MMETSP0810, MMETSP0811	CAMPEP_0200405894
	Euglenozoa	Euglenophyceae	Euglenaceae	Euglena	gracilis	??	??	light_m.86199
RHIZARIA	Cercozoa	Chlorarachniophyceae	Chlorarachniaceae	Chlorarachnion	reptans	CCCM449	MMETSP0109	CAMPEP_0114518010
	Cercozoa	Chlorarachniophyceae	Chlorarachniaceae	Gymnochlora	sp.	CCMP2014	MMETSP0110	CAMPEP_0167746174, CAMPEP_0167745746
	Cercozoa	Chlorarachniophyceae	Chlorarachniaceae	Lotharella	oceanica	CCMP622	MMETSP0040	CAMPEP_0170191836

	Cercozoa	Chlorarachniophyceae	Chlorarachniaceae	Lotharella	globosa	CCCM811	MMETSP0111, MMETSP0112	CAMPEP_0190181758
	Cercozoa	Chlorarachniophyceae	Chlorarachniaceae	Bigelowiella	natans	CCMP623	MMETSP1052	CAMPEP_0169532188, CAMPEP_0169546672,
	Cercozoa	Chlorarachniophyceae	Chlorarachniaceae	Bigelowiella	natans	CCMP 2755	MMETSP0045	CAMPEP_0114200706
	Cercozoa	Chlorarachniophyceae	Chlorarachniaceae	Bigelowiella	natans	CCMP1259	MMETSP1054	CAMPEP_0169599456
CRYPTOPHYTES	Cryptophyta	Cryptophyceae	Cryptomonadaceae	Cryptomonas	parameciu m	CCAP977/2a	MMETSP0038	CAMPEP_0113696002
	Cryptophyta	Cryptophyceae	Cryptomonadaceae	Cryptomonas	curvata	CCAP979/52	MMETSP1050	CAMPEP_0172168610, CAMPEP_0172173114
	Cryptophyta	Cryptophyceae	Geminigeraceae	Geminigera	cryophila	CCMP2564	MMETSP0799	CAMPEP_0179481402, CAMPEP_0179458582
	Cryptophyta	Cryptophyceae	Geminigeraceae	Geminigera	sp.	Caron Lab Isolate	MMETSP1102	CAMPEP_0173087898, CAMPEP_0173097438
	Cryptophyta	Cryptophyceae	Geminigeraceae	Proteomonas	sulcata	CCMP704	MMETSP1049	
	Cryptophyta	Cryptophyceae	Geminigeraceae	Guillardia	theta	CCMP 2712	MMETSP0046	CAMPEP_0113812340
	Cryptophyta	Cryptophyceae	Hemiselmidaceae	Hemiselmis	andersenii	CCMP439	MMETSP1041	CAMPEP_0172041234
	Cryptophyta	Cryptophyceae	Hemiselmidaceae	Hemiselmis	andersenii	CCMP1180	MMETSP1042	CAMPEP_0169429296
	Cryptophyta	Cryptophyceae	Hemiselmidaceae	Hemiselmis	andersenii	CCMP441	MMETSP1043	CAMPEP_0172046830
	Cryptophyta	Cryptophyceae	Hemiselmidaceae	Hemiselmis	tepida	CCMP443	MMETSP1355	CAMPEP_0174941776
	Cryptophyta	Cryptophyceae	Hemiselmidaceae	Hemiselmis	rufescens	PCC563	MMETSP1357	CAMPEP_0173449620, CAMPEP_0173437524
	Cryptophyta	Cryptophyceae	Hemiselmidaceae	Hemiselmis	viresens	PCC157	MMETSP1356	CAMPEP_0173408906, CAMPEP_0173411170
	Cryptophyta	Cryptophyceae	Goniomonodaceae	Goniomonas	Pacifica	CCMP1869	MMETSP0107, MMETSP0108	CAMPEP_0188509280
	Cryptophyta	Cryptophyceae	Pyrenomonadaceae	Rhodomonas	salina	CCMP1319	MMETSP1047	CAMPEP_0172097468, CAMPEP_0172099434
	Cryptophyta	Cryptophyceae	Pyrenomonadace	Rhodomonas	sp.	CCMP768	MMETSP1091, MMETSP1389	CAMPEP_0191543004

			ae					
	Cryptophyta	Cryptophyceae	Pyrenomonadaceae	Rhodomonas	abbreviata	Caron Lab Isolate	MMETSP1101	CAMPEP_0181338540, CAMPEP_0181298928, CAMPEP_0181313830, CAMPEP_0181297718
HAPTOPHYTES	Haptophyta	Pavlovophyceae	Pavlovaceae	Pavlova	gyrans	CCMP608	MMETSP1466	CAMPEP_0206039264, CAMPEP_0206053400
	Haptophyta	Pavlovophyceae	Pavlovaceae	Pavlova	lutheri	RCC1537	MMETSP1463	
	Haptophyta	Pavlovophyceae	Pavlovaceae	Pavlova	sp.	CCMP459	MMETSP1139, MMETSP1140, MMETSP1381	CAMPEP_0190524078,
	Haptophyta	Prymnesiophyceae	Prymnesiaceae	Chrysochromulina	polylepis	UIO037	MMETSP0286	
	Haptophyta	Prymnesiophyceae	Prymnesiaceae	Chrysochromulina	polylepis	CCMP1757	MMETSP0143, MMETSP0145, MMETSP0146, MMETSP0147	CAMPEP_0193719730, CAMPEP_0193726698
	Haptophyta	Prymnesiophyceae	Prymnesiaceae	Chrysochromulina	rotalis	UIO044	MMETSP0287	CAMPEP_0115854040
	Haptophyta	Prymnesiophyceae	Prymnesiaceae	Chrysochromulina	ericina	CCMP281	MMETSP1096	CAMPEP_0181246850, CAMPEP_0181201634,
	Haptophyta	Prymnesiophyceae	Prymnesiaceae	Chrysochromulina	brevifilum	UTEX LB 985	MMETSP1094	CAMPEP_0174727498
	Haptophyta	Prymnesiophyceae	Prymnesiaceae	Prymnesium	parvum	Texoma1	MMETSP0006, MMETSP0007, MMETSP0008, MMETSP0815, MMETSP0814	CAMPEP_0191228776,
	Haptophyta	Prymnesiophyceae	Noelaerhabdaceae	Emiliania	huxleyi	374	MMETSP1006, MMETSP1007, MMETSP1008, MMETSP1009	CAMPEP_0187581196
	Haptophyta	Prymnesiophyceae	Noelaerhabdaceae	Emiliania	huxleyi	379	MMETSP0994, MMETSP0995, MMETSP0996, MMETSP0997	CAMPEP_0187642360
	Haptophyta	Prymnesiophyceae	Noelaerhabdaceae	Emiliania	huxleyi	PLY M219	MMETSP1150, MMETSP1151, MMETSP1152, MMETSP1153	CAMPEP_0187777132
	Haptophyta	Prymnesiophyceae	Noelaerhabdaceae	Emiliania	huxleyi	CCMP370	MMETSP1154, MMETSP1155, MMETSP1156, MMETSP1157	CAMPEP_0187665496
	Haptophyta	Prymnesiophyceae	Noelaerhabdaceae	Emiliania	huxleyi	CCMP1516	http://www.ncbi.nlm.nih.gov/genome/2	XP_005785625.1
	Haptophyta	Prymnesiophyceae	Noelaerhabdaceae	Gephyrocapsa	oceanica	RCC1303	MMETSP1363, MMETSP1364, MMETSP1365, MMETSP1366	CAMPEP_0188174236

	Haptophyta	Prymnesiophyceae	Isochrysidaceae	Isochrysis	sp.	CCMP1324	MMETSP1129, MMETSP1130, MMETSP1131, MMETSP1132	CAMPEP_0188830390
	Haptophyta	Prymnesiophyceae	Isochrysidaceae	Isochrysis	sp	CCMP1244	MMETSP1090, MMETSP1388	CAMPEP_0188768992
	Haptophyta	Prymnesiophyceae	Isochrysidaceae	Isochrysis	galbana	CCMP1323	MMETSP0944, MMETSP0943, MMETSP0595	CAMPEP_0193694840, CAMPEP_0193669402
	Haptophyta	Prymnesiophyceae	Phaeocystaceae	Phaeocystis	Sp	CCMP2710	MMETSP1162	
	Haptophyta	Prymnesiophyceae	Phaeocystaceae	Phaeocystis	antarctica	Caron Lab Isolate	MMETSP1100	CAMPEP_0172959570
	Haptophyta	Prymnesiophyceae	Phaeocystaceae	Phaeocystis	antarctica	CCMP1374	MMETSP1444	CAMPEP_0198174474
	Haptophyta	Prymnesiophyceae	Pleurochrysidaceae	Pleurochrysis	carterae	CCMP645	MMETSP1136, MMETSP1137, MMETSP1138	CAMPEP_0190806002
STRAMENOPILES	Ochrophyta	Bacillariophyceae	Amphipleuraceae	Amphiprora	sp.	CCMP467	MMETSP0725, MMETSP0726, MMETSP0727, MMETSP0724	CAMPEP_0186506366, CAMPEP_0186488760
	Ochrophyta	Bacillariophyceae	Catenulaceae	Amphora	coffeaeformis	CCMP127	MMETSP0316, MMETSP0317, MMETSP0318	CAMPEP_0186538782
	Ochrophyta	Bacillariophyceae	Bacillariaceae	Fragilariopsis	kerguelensis	L26-C5	MMETSP0733, MMETSP0734, MMETSP0735, MMETSP0736	CAMPEP_0188139216
	Ochrophyta	Bacillariophyceae	Bacillariaceae	Fragilariopsis	kerguelensis	L2-C3	MMETSP0906, MMETSP0907, MMETSP0908, MMETSP0909	
	Ochrophyta	Bacillariophyceae	Bacillariaceae	Nitzschia	punctata	CCMP561	MMETSP0744, MMETSP0745, MMETSP0746, MMETSP0747	CAMPEP_0199327438
	Ochrophyta	Bacillariophyceae	Bacillariaceae	Pseudo-nitzschia	australis	10249 10 AB	MMETSP0139, MMETSP0142, MMETSP0140, MMETSP0141	CAMPEP_0199647432
	Ochrophyta	Bacillariophyceae	Bacillariaceae	Pseudo-nitzschia	fraudulenta	WWA7	MMETSP0850, MMETSP0851, MMETSP0852, MMETSP0853	CAMPEP_0199791592
	Ochrophyta	Bacillariophyceae	Chaetocerotaceae	Chaetoceros	debilis	MM31A-1	MMETSP0149, MMETSP0150	CAMPEP_0200879214
	Ochrophyta	Bacillariophyceae	Chaetocerotaceae	Chaetoceros	neogracile	CCMP1317	MMETSP0751, MMETSP0752, MMETSP0753, MMETSP0754	CAMPEP_0201017622, CAMPEP_0201002392, CAMPEP_0200999620
	Ochrophyta	Bacillariophyceae	Chaetocerotaceae	Chaetoceros	curvisetus		MMETSP0716, MMETSP0717, MMETSP0718, MMETSP0719	CAMPEP_0187046754
	Ochrophyta	Bacillariophyceae	Chaetocerotaceae	Chaetoceros	affinis	CCMP159	MMETSP0088, MMETSP0090, MMETSP0091, MMETSP0092	CAMPEP_0187024338

	Ochrophyta	Bacillariophyceae	Corethraceae	Corethron	pennatum	L29A3	MMETSP0169, MMETSP0171	CAMPEP_0200351556, CAMPEP_0200360040
	Ochrophyta	Bacillariophyceae	Lithodesmiaceae	Ditylum	brightwellii	GSO103	MMETSP1002, MMETSP1005	CAMPEP_0187315252
	Ochrophyta	Bacillariophyceae	Lithodesmiaceae	Ditylum	brightwellii	GSO104	MMETSP1010, MMETSP1012, MMETSP1013	CAMPEP_0193980186
	Ochrophyta	Bacillariophyceae	Lithodesmiaceae	Ditylum	brightwellii	GSO105	MMETSP0998, MMETSP1001	CAMPEP_0187336034
	Ochrophyta	Bacillariophyceae	Cymatosiraceae	Extubocellulus	spinifer	CCMP396	MMETSP0699, MMETSP0697, MMETSP0698,	CAMPEP_0200501572
	Ochrophyta	Bacillariophyceae	Thalassiosiraceae	Thalassiosira	rotula	CCMP3096	MMETSP0403, MMETSP0404	CAMPEP_0192952656
	Ochrophyta	Bacillariophyceae	Thalassiosiraceae	Thalassiosira	oceanica	CCMP1005	MMETSP0970, MMETSP0971, MMETSP0972, MMETSP0973	CAMPEP_0192905742, CAMPEP_0192933258
	Ochrophyta	Bacillariophyceae	Thalassiosiraceae	Thalassiosira	weissflogii	CCMP1010	MMETSP0898, MMETSP0899, MMETSP0900, MMETSP0901, MMETSP1407, MMETSP1408, MMETSP1405, MMETSP1406, MMETSP1415, MMETSP1416, MMETSP1417, MMETSP1418, MMETSP1419, MMETSP1420, MMETSP1421, MMETSP1422, MMETSP1409, MMETSP1410, MMETSP1411, MMETSP1412, MMETSP1413, MMETSP1414,	CAMPEP_0193043134
	Ochrophyta	Bacillariophyceae	Thalassiosiraceae	Thalassiosira	antarctica	CCMP982	MMETSP0902, MMETSP0903, MMETSP0904, MMETSP0905	CAMPEP_0200088110
	Ochrophyta	Bacillariophyceae	Thalassiosiraceae	Thalassiosira	weissflogii	CCMP1336	MMETSP0878, MMETSP0879, MMETSP0880, MMETSP0881	CAMPEP_0193083268
	Ochrophyta	Bacillariophyceae	Thalassiosiraceae	Thalassiosira	rotula	GSO102	MMETSP0910, MMETSP0911, MMETSP0912, MMETSP0913	CAMPEP_0192998562
	Ochrophyta	Bacillariophyceae	Thalassiosiraceae	Thalassiosira	gravida	GMp14c1	MMETSP0492, MMETSP0493, MMETSP0494	CAMPEP_0200699684
	Ochrophyta	Bacillariophyceae	Rhizosoleniaceae	Proboscia	alata	PI-D3	MMETSP0174, MMETSP0176	CAMPEP_0200159498
	Ochrophyta	Bacillariophyceae	Skeletonemaceae	Skeletonema	marinoi	skelA	MMETSP0920, MMETSP0918	CAMPEP_0192233268
	Ochrophyta	Bacillariophyceae	Skeletonemaceae	Skeletonema	dohrnii	SkelB	MMETSP0562, MMETSP0563	CAMPEP_0192137106
	Ochrophyta	Bacillariophyceae	Skeletonemaceae	Skeletonema	menzelii	CCMP793	MMETSP0603, MMETSP0604	CAMPEP_0192272568
	Ochrophyta	Bacillariophyceae	Fragilariaceae	Asterionellopsis	glacialis	CCMP134	MMETSP0705, MMETSP0706, MMETSP0707,	CAMPEP_0199870378,

							MMETSP0708	CAMPEP_0199870466
	Ochrophyta	Bacillariophyceae	Thalassionemataceae	Thalassionema	nitzschioides	L26-B	MMETSP0156, MMETSP0158	CAMPEP_0200190772
	Ochrophyta	Bacillariophyceae	Thalassionemataceae	Thalassiothrix	antarctica	L6-D1	MMETSP0152, MMETSP0154	CAMPEP_0200962748
	Ochrophyta	Chrysophyceae	Synuraceae	Paraphysomonas	Imperforata	PA2	MMETSP0103, MMETSP0104	CAMPEP_0190482658
	Ochrophyta	Chrysophyceae	Dinobryaceae	Dinobryon	sp.	UTEXLB2267	MMETSP0019, MMETSP0020, MMETSP0812	CAMPEP_0187279028, CAMPEP_0187277876
	Ochrophyta	Chrysophyceae	Ochromonadaceae	Ochromonas	sp.	CCMP1393	MMETSP0004, MMETSP0005	CAMPEP_0190272254
	Ochrophyta	Dictyochophyceae	Pedinellaceae	Pseudopedinella	elastica	CCMP716	MMETSP1068, MMETSP1097	CAMPEP_0191451212, CAMPEP_0191405316
	Ochrophyta	Dictyochophyceae	Pedinellaceae	Pteridomonas	danica	PT	MMETSP0101, MMETSP0102	CAMPEP_0193806330
	Ochrophyta	Eustigmatophyceae	Eustigmataceae	Nannochloropis	gaditana	B-31	http://www.ncbi.nlm.nih.gov/genome/11691?genome_assembly_id=53301	EWM28830.1
	Ochrophyta	Pelagophyceae	Pelagomonodaceae	Aureococcus	anophagefferens	CCMP1850	MMETSP0914, MMETSP0915, MMETSP0916, MMETSP0917	CAMPEP_0186659954, CAMPEP_0186691924
	Ochrophyta	Pelagophyceae	Sarcionochrysidaceae	Aureoumbra	lagunensis	CCMP1510	MMETSP0890, MMETSP0891, MMETSP0892, MMETSP0893	
	Ochrophyta	Pelagophyceae	Pelagomonodaceae	Pelagococcus	subviridis	CCMP1429	MMETSP0882, MMETSP0883, MMETSP0884, MMETSP0885	CAMPEP_0190535384
	Ochrophyta	Pelagophyceae	Pelagomonodaceae	Pelagomonas	calceolata	CCMP1756	MMETSP0888, MMETSP0889, MMETSP0886, MMETSP0887	CAMPEP_0199675904
	Ochrophyta	Raphidophyceae	Chattonellaceae	Chattonella	subsalsa	CCMP2191	MMETSP0947, MMETSP0948, MMETSP0949, MMETSP0950	CAMPEP_0187151886
	Ochrophyta	Raphidophyceae	Chattonellaceae	Heterosigma	akashiwo	CCMP2393	MMETSP0292, MMETSP0294, MMETSP0295, MMETSP0296	
	Ochrophyta	Raphidophyceae	Chattonellaceae	Heterosigma	akashiwo	NB	MMETSP0416, MMETSP0414, MMETSP0415	CAMPEP_0200245898
	Ochrophyta	Raphidophyceae	Chattonellaceae	Heterosigma	akashiwo	CCMP3107	MMETSP0409, MMETSP0410, MMETSP0411	CAMPEP_0188619588
	Ochrophyta	Raphidophyceae	Chattonellaceae	Heterosigma	akashiwo	CCMP 452	MMETSP0894, MMETSP0895, MMETSP0896, MMETSP0897	CAMPEP_0188660716

	Ochrophyta	Xanthophyceae	Vaucheriaceae	Vaucheria	litorea	CCMP2940	MMETSP0945, MMETSP0946	CAMPEP_0199157760
	Labyrinthista	Labyrinthulea	Thraustochytriaceae	Aplanochytrium	sp	PBS07	MMETSP0954, MMETSP0955, MMETSP0956, MMETSP0957	
	Labyrinthista	Labyrinthulea	Thraustochytriaceae	Aplanochytrium	stocchinoi	GSBS06	MMETSP1347, MMETSP1348, MMETSP1349, MMETSP1346	
	Labyrinthista	Labyrinthulea	Thraustochytriidae	Aurantiochytrium	limacinum	ATCCMYA-1381	MMETSP0959, MMETSP0960, MMETSP0961, MMETSP0958	CAMPEP_0186647428
	Labyrinthista	Labyrinthulea	Thraustochytriaceae	Schizochytrium	aggregatum	ATCC28209	MMETSP0962, MMETSP0963, MMETSP0964, MMETSP0965	
	Labyrinthista	Labyrinthulea	Thraustochytriaceae	Thraustochytrium	sp.	LLF1b	MMETSP0198, MMETSP0199	
ALVEOLATA	Dinoflagellata	Dinophyceae	Gymnodiniaceae	Amphidinium	carterae	CCMP1314	MMETSP0399, MMETSP0259, MMETSP0258, MMETSP0398C	CAMPEP_0186428018, CAMPEP_0186426112
	Dinoflagellata	Dinophyceae	Gymnodiniaceae	Karenia	brevis	CCMP2229	MMETSP0027, MMETSP0029, MMETSP0030, MMETSP0031	CAMPEP_0188846322, CAMPEP_0188912736
	Dinoflagellata	Dinophyceae	Gymnodiniaceae	Karenia	brevis	Wilson	MMETSP0202, MMETSP0201, MMETSP0648, MMETSP0649	CAMPEP_0189492014, CAMPEP_0189388456
	Dinoflagellata	Dinophyceae	Gymnodiniaceae	Karenia	brevis	SP3	MMETSP0527, MMETSP0528	CAMPEP_0189309980, CAMPEP_0189259884
	Dinoflagellata	Dinophyceae	Gymnodiniaceae	Karenia	brevis	SP1	MMETSP0573, MMETSP0574	CAMPEP_0189168478, CAMPEP_0189051810
	Dinoflagellata	Dinophyceae	Gymnodiniaceae	Karlodinium	micrum	CCMP2283	MMETSP1015, MMETSP1016, MMETSP1017	CAMPEP_0200794146, CAMPEP_0200809940, CAMPEP_0200812546, CAMPEP_0200785114
	Dinoflagellata	Dinophyceae	Peridiniaceae	Durinskia	baltica	CSIRO CS-38	MMETSP0117, MMETSP0116	CAMPEP_0200004706, CAMPEP_0200061890, CAMPEP_0199929530
	Dinoflagellata	Dinophyceae	Peridiniaceae	Glenodinium	foliaceum	CCAP 1116/3	MMETSP0118, MMETSP0119	CAMPEP_0188311894, CAMPEP_0188271626, CAMPEP_0188383902, CAMPEP_0188247870, CAMPEP_0188407728
	Dinoflagellat	Dinophyceae	Peridiniaceae	Kryptoperidiniu	foliaceum	CCMP 1326	MMETSP0121, MMETSP0120	CAMPEP_0189696278,

	a			m				CAMPEP_0189642352, CAMPEP_0189900696, CAMPEP_0189771244
	Dinoflagellata	Dinophyceae	Peridiniaceae	Peridinium	aciculiferum	PAER-2	MMETSP0370, MMETSP0371	CAMPEP_0190670536, CAMPEP_0190612054,
	Dinoflagellata	Dinophyceae	Peridiniaceae	Scrippsiella	trochoidea	CCMP3099	MMETSP0270, MMETSP0271, MMETSP0272	CAMPEP_0191964902, CAMPEP_0191966966
	Dinoflagellata	Dinophyceae	Peridiniaceae	Scrippsiella	Hangoei	SHTV-5	MMETSP0359, MMETSP0360, MMETSP0361	CAMPEP_0191865414, CAMPEP_0191802586, CAMPEP_0191847402
	Dinoflagellata	Dinophyceae	Peridiniaceae	Scrippsiella	hangoei-like	SHHI-4	MMETSP0367, MMETSP0368, MMETSP0369	CAMPEP_0199174568, CAMPEP_0199204792, CAMPEP_0199287414
	Dinoflagellata	Dinophyceae	Goniodomataceae	Alexandrium	monilatum	CCMP3105	MMETSP0095, MMETSP0096, MMETSP0097, MMETSP0093	CAMPEP_0200518424, CAMPEP_0200664972
	Dinoflagellata	Dinophyceae	Goniodomataceae	Alexandrium	fundyense	CCMP1719	MMETSP0196C, MMETSP0347	CAMPEP_0185978494
	Dinoflagellata	Dinophyceae	Goniodomataceae	Alexandrium	tamarense	CCMP1771	MMETSP0382, MMETSP0384, MMETSP0378, MMETSP0380	CAMPEP_0186347800, CAMPEP_0186194818
	Dinoflagellata	Dinophyceae	Unknown	Azadinium	spinosum	3D9	MMETSP1036, MMETSP1037, MMETSP1038	CAMPEP_0186738362, CAMPEP_0186849724
	Dinoflagellata	Dinophyceae	Ceratiaceae	Ceratium	fusus	PA161109	MMETSP1075, MMETSP1074	CAMPEP_0199470176, CAMPEP_0199448012
	Dinoflagellata	Dinophyceae	Crypthecodiniacea	Crypthecodinium	cohnii	Seligo	MMETSP0323, MMETSP0325, MMETSP0326, MMETSP0324	CAMPEP_0193873420, CAMPEP_0193920440
	Dinoflagellata	Dinophyceae	Gonyaulacaceae	Lingulodinium	polyedra	CCMP 1738	MMETSP1032, MMETSP1033, MMETSP1034, MMETSP1035	CAMPEP_0190033118, CAMPEP_0189981950
	Dinoflagellata	Dinophyceae	Oxyrrhinaceae	Oxyrrhis	marina	CCMP1795	MMETSP0452_2, MMETSP0451_2C	
	Dinoflagellata	Dinophyceae	Oxyrrhinaceae	Oxyrrhis	marina	Unknown	MMETSP0468, MMETSP0469, MMETSP0470, MMETSP0471	
	Dinoflagellata	Dinophyceae	Oxyrrhinaceae	Oxyrrhis	marina	LB1974	MMETSP1424, MMETSP1425, MMETSP1426	
	Dinoflagellata	Dinophyceae	Prorocentraceae	Prorocentrum	minimum	CCMP1329	MMETSP0053, MMETSP0055, MMETSP0057, MMETSP0056	CAMPEP_0190887074, CAMPEP_0190973906,

								CAMPEP_0190881924
	Dinoflagellata	Dinophyceae	Prorocentraceae	Prorocentrum	minimum	CCMP2233	MMETSP0267, MMETSP0268, MMETSP0269	CAMPEP_0191171648, CAMPEP_0191085234, CAMPEP_0191101348
	Dinoflagellata	Dinophyceae	Symbiodiniaceae	Symbiodinium	kawagutii	CCMP2468	MMETSP0132_2, MMETSP0133_2, MMETSP0134_2, MMETSP0135_2	
	Dinoflagellata	Dinophyceae	Symbiodiniaceae	Symbiodinium	sp.	CCMP2430	MMETSP1115, MMETSP1116, MMETSP1117	CAMPEP_0192483534
	Dinoflagellata	Dinophyceae	Symbiodiniaceae	Symbiodinium	sp.	Mp	MMETSP1122, MMETSP1123, MMETSP1124, MMETSP1125	CAMPEP_0192623488
	Dinoflagellata	Dinophyceae	Symbiodiniaceae	Symbiodinium	sp.	C1	MMETSP1367, MMETSP1369	CAMPEP_0199632616, CAMPEP_0199584150
	Dinoflagellata	Dinophyceae	Symbiodiniaceae	Symbiodinium	sp.	C15	MMETSP1370, MMETSP1371	CAMPEP_0192427984
	Dinoflagellata	Dinophyceae	Symbiodiniaceae	Symbiodinium	microadriaticum	CCMP2467	https://www.ncbi.nlm.nih.gov/genome/?term=txid2951[orgn]	OLP79990.1
	Ciliophora	Colopdea	Platyophryidae	Platyophrya	macrostoma	WH	MMETSP0127	CAMPEP_0176435236, CAMPEP_0176475592, CAMPEP_0176458050
	Ciliophora	Heterotrichea	Climacostomidae	Climacostomum	virens	Stock W-24	MMETSP1397	
	Ciliophora	Oligohymenophorea	Orchitophryidae	Anophryoides	haemophila	AH6	MMETSP1018	
	Ciliophora	Oligotrichea	Ptychocylididae	Favella	taraikaensis	Fe Narragansett Bay	MMETSP0434, MMETSP0436	CAMPEP_0199840474
	Ciliophora	Spirotrichea	Euplotidae	Euplotes	focardii	TN1	MMETSP0205, MMETSP0206	CAMPEP_0187815918
	Perkinsozoa	Perkinsea	Perkinsidae	Perkinsus	chesapeaki	ATCC PRA-65	MMETSP0925, MMETSP0924C	
	Perkinsozoa	Perkinsea	Perkinsidae	Perkinsus	marinus	ATCC 50439	MMETSP0923, MMETSP0922	

Supplementary Material 3 – List of organisms used in this study with respective nucleic acid database identifiers, and sequence identifiers found for NDP-rhamnose biosynthesis.

6 Key Findings and Recent Developments

6.1 Key findings of this thesis

6.1.1 Ichthyotoxins responsible for toxic Prymnesium blooms

One aim of this thesis was to fill some fundamental gaps in knowledge that still existed surrounding physiology of *Prymnesium parvum* blooms and their associated toxicity towards aquatic organisms. A key objective was to clear up much literature ambiguity surrounding the toxins responsible for fish deaths. Given the numerous claims of toxic compounds released by *P. parvum,* which included galactoglycerolipids [1], fatty acids [2], fatty acid amides [3, 4], and ladder-frame polyether prymnesins [5, 6], at the time of starting this thesis one could only speculate on the biologically relevant toxin(s). However, the potent ichthyotoxic LD_{50} values observed for the polyether prymnesins [6] and their structural similarity to other known algal toxins [7] led to the hypothesis that the prymnesins had a key role in toxic *P. parvum* blooms – even if researchers had failed to report the detection of these compounds since their original isolation in 1996-1999.

A major advancement in the prymnesin research field came in 2013 when Manning and La Claire reported a detailed extraction and analysis protocol for the detection of prymnesins from cultures of *P. parvum* [8]. Using the methods outlined in that paper, we were able to show that our laboratory strain of *P. parvum* (CCAP 946/6) produced the polyether prymnesins-1 and -2 (Chapter 2). Work by Blossom *et al* in 2014 then further supported our hypothesis that the prymnesins played a key role, having shown that fatty acid amides and others were not biologically relevant toxins in *P. parvum* ichthytoxicity [9]. Using a combination of optical microscopy and genetic analysis, we then showed that a harmful algal bloom in 2015 on the Norfolk Broads, United Kingdom, was caused by *P. parvum* (Chapter 2). At the time of the bloom, we were unable to detect prymnesin-1 or -2 in water samples or gill tissue preparations, making us question our initial hypothesis. However, Rasmussen *et al* would go on to report a structural diversity of the prymnesins in 2016, outlining new m/z values for novel prymnesin toxins including prymnesin-B1 [10]. Using these new findings, we were able to detect prymnesins for the first time in natural water samples, having detected prymnesin-B1 in water samples from the site of the 2015 bloom. Furthermore, we were able to detect prymnesin-B1 in the gill tissue of a deceased pike. Taken together, these findings strongly support the assignment of the polyether prymnesins as the biologically relevant ichthytoxins, although a cocktail of toxins which include the prymnesins cannot be ruled out.

6.1.2 Method of toxin release – implication for algal viruses

Because of the ambiguity in the nature of the *Prymnesium*-associated ichthytoxins, their localisation (i.e. intra- vs extra-cellular) and mode of toxicity was unknown when this thesis was started. Many toxicity assays were previously based on the erythrocyte lysis assay (ELA) [11] but this assay may have also been assessing other toxic compounds produced by *P. parvum* and not those specific to fish. Probably the most convincing work at the time of starting this thesis was by Remmel and Hambright who, in 2012, showed that ichthytoxins from *P. parvum* were intracellular and that they were only released through contact with prey or by natural causes of stress to the algal cells [12]. Importantly, the toxicity studies in this work used live 10-14-day old fathead minnows (*Pimephales promelas*) rather than other standard bioassays.

In Chapter 3, we showed that natural populations of *P. parvum* during the toxic bloom of 2015 on Hickling Broad appeared to be infected by an algal virus, and subsequently isolated and characterized a new species of Prymnesium-infecting virus, named *Prymnesium parvum* DNA virus (PpDNAV-BW1). Using cell culture techniques, we showed that PpDNAV-BW1 is lytic; causing lysis of more than 95% of the *P. parvum* population 120-hours post infection. Morphological and genetic analysis of this virus placed it into the Megaviridae family of algal viruses. Given that other algal viruses had previously been reported to be involved in the release of intracellular metabolites, such as dimethylsulfide [13], we immediately formed the hypothesis after discovery of PpDNAV-BW1 that it may be involved in the release of the intracellular prymnesin toxins (Figure 60). This would agree with the work of Remmel and Hambright, who proposed toxin release through natural causes of stress [12] and of La Claire *et al* who propose that prymnesins are not actively secreted but are likely excreted passively or by cell-lysis events [14]. We propose that at some point during bloom formation, natural populations of *P. parvum* may become infected by lytic viruses. This infection ultimately leads to a sudden lysis event in the algal population, causing release of new viral progeny and dissolved organic matter (DOM) from the cells. Importantly, we propose that this DOM includes the toxic prymnesins, raising extracellular levels above the threshold to be toxic to aquatic life. We also suggest that resuspension of sediment may be a cause for increased instances of viral infection, as the viral particles are re-suspended in the waterbody and more likely to encounter *P. parvum* cells (Figure 60). Given the toxin extraction and analysis protocols described in Chapter 2 of this thesis, follow-up work in this area will look at the

effect of viral infection on the production of prymnesin toxins, and of viral lysis on extracellular prymnesin levels.

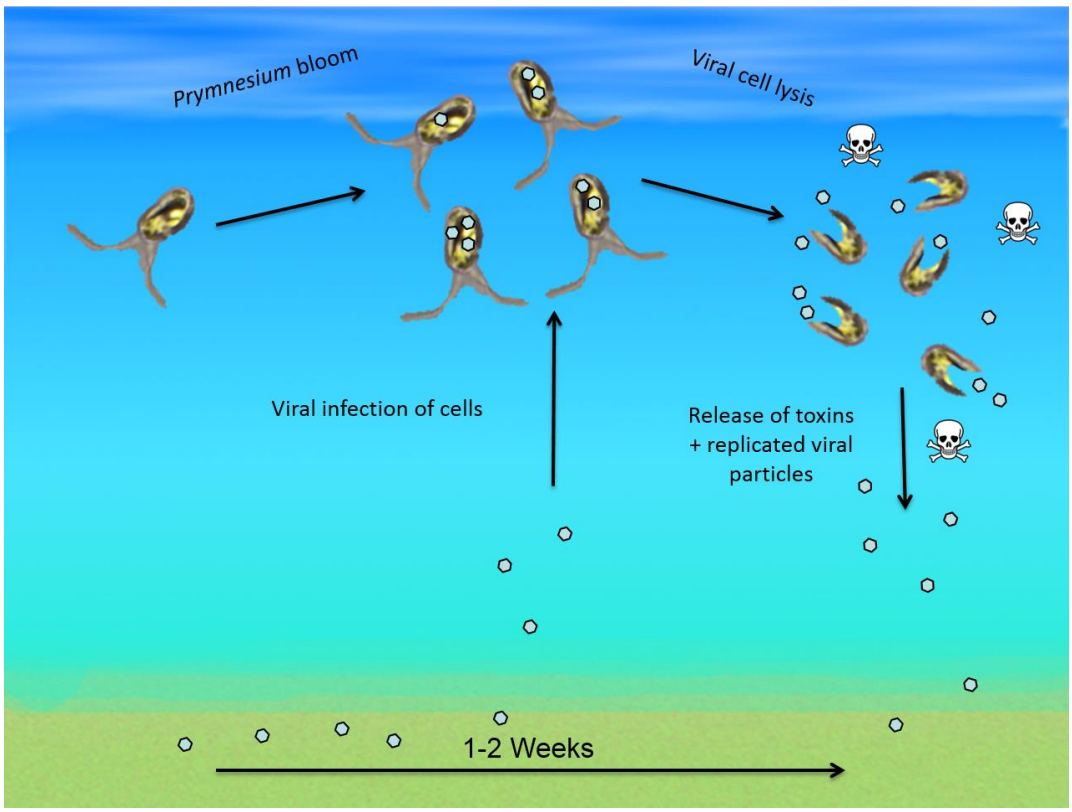


Figure 60 – Proposed mechanism of viral lysis-mediated toxin release in a *P. parvum* bloom. Grey hexagons represent viral particles. Background image adapted under a Creative Commons Attribution-No Derivative Works 3.0 Licence (https://creativecommons.org/licenses/by-nd/3.0/) with permission of the author (kazza234).

6.1.3 Sialic acid biosynthesis in algae and potential roles in viral infection

After entering the world of algal viruses following the discovery of PpDNAV-BW1, our interest was turned to the molecular basis behind viral infections of algae. Research from Israel had recently shown that the uncommon sialic acid, 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid (KDN) played a key role in infection of the haptophyte *Emiliania huxleyi* by *Emiliania huxleyi* virus (EhV) [15]. However, prior to this work, reports of sialic acids in algae were sparse; to our knowledge sialic acids had not been reported in algae at all. Given the importance of sialic acids in a suite of host:pathogen interactions [16, 17], influenza virus

infection of the lungs being the archetypal example, and their recent implications in algal virus infections, we then sought to investigate the presence of sialic acids in *P. parvum*.

Using a range of LC-MS techniques, Chapter 4 showed that *P. parvum* contains the cytidine monophosphate (CMP)-activated form of KDN, CMP-KDN. Using a publicly available transcriptome of *P. parvum* we then identified sialic acid biosynthetic genes from *P. parvum*. These genes were successfully cloned and expressed in *E. coli*, and the biosynthetic capabilities of the corresponding enzymes were analysed using a range of analytical techniques that included NMR, MS, and colorimetric assays. As expected, these assays confirmed that *P. parvum* produces CMP-KDN starting from mannose-6-phosphate, in a similar fashion to that seen for *Bacteroides thetaiotaomicron* [18]. Using a colorimetric phosphate release assay, we demonstrated that KDN-9-P synthase from *P. parvum* has a more efficient k_{cat} and K_m value than many other enzymes of this family, a feature that warrants further investigation.

Furthermore, using these newly discovered and characterized sequences, we used BLASTp analysis [19] to query more than 150 algal nucleic acid databases from NCBI [20] and MMETSP [21]. Using this analysis, we showed that sialic acid biosynthesis is much more widespread amongst algae than previously thought. Phylogenetic trees were created and display distinct clades of algal sialic acid synthases with similarity to KDN-9-P synthase from *P. parvum*, which we speculate separate sialic acid biosynthesis from the biosynthesis of the structurally similar acidic sugar, KDO. Taken together, the findings in Chapter 4 show that sialic acid biosynthesis is abundant throughout the algae, and coupled with recent work by Fulton *et al* which implicates KDN in viral infection of *E. huxleyi* [15], may suggest sialic acids play crucial roles in algal virus interactions.

6.1.4 The algal monosaccharide L-rhamnose has a complicated evolutionary origin

During our bioinformatic analysis of sialic acid biosynthesis in *P. parvum* we came across a range of other interesting sugar-nucleotide biosynthetic enzymes. One transcript that caught our attention translated to a fusion protein of RmlC and RmlD – two bacterial enzymes involved in L-rhamnose biosynthesis [22]. Rhamnose is an important monosaccharide that is found in structural polysaccharides across microbes, algae, and plants, but not in animals. The

biosynthesis of nucleotide activated rhamnose has been extensively studied in bacteria [23] and plants [24], and it is generally accepted that most prokaryotes activate rhamnose as thymidine diphosphate (TDP) adducts, whilst most eukaryotes activate rhamnose as uridine diphosphate (UDP) adducts. Prior to our study however, it was unknown how algae produce rhamnose and which activated sugar nucleotide they produce. Furthermore, independent work from our group identified rhamnose as being an important sugar in the viral infection process of *P. parvum*.

In Chapter 5, we used a bioinformatics-guided approach to discover rhamnose biosynthetic genes across the algal groups using nucleic acid databases from NCBI [20] and MMETSP [21]. We found that whilst most algal groups appear to encode the 'plant-like' UDP-β-L-Rha pathway, the haptophytes (including *P. parvum*) and some dinoflagellates encode the 'bacterial-like' TDP-β-L-Rha pathway. To support these findings, we took two representative algae from each pathway; *P. parvum* which should produce TDP-β-L-Rha, and *Euglena gracilis* that encoded the 'plant-like' pathway and should produce UDP-β-L-Rha. We profiled the sugar nucleotides of these species using LC-MS and showed that, as expected, *P. parvum* contains more TDP-β-L-Rha than the UDP adduct, and *E. gracilis* contains more UDP-β-L-Rha than the TDP adduct. Taking these findings together, we proposed an evolutionary model for L-rhamnose biosynthesis across the algal groups, suggesting that the haptophytes lost the 'plant-like' pathway sometime after secondary endosymbiosis. We then proposed a horizontal gene transfer event from bacteria, whereby the haptophytes obtained the 'bacterial-like' pathway. Finally, a tertiary endosymbiosis event with the peridinin-containing dinoflagellates led to the fucoxanthin-containing dinoflagellates having both the bacterial-like pathway (from *P. parvum*) and the plant-like pathway (from their dinoflagellate ancestor).

6.2 Recent development of practical applications

This thesis was carried out as a BBSRC-funded iCASE studentship in conjunction with the Environment Agency. With that in mind, some key objectives of this thesis involved the generation of practical methods for monitoring and managing blooms of *P. parvum* on the Norfolk Broads.

6.2.1 Detection of *P. parvum* and PpDNAV-BW1

In Chapter 2 we report the uses of qPCR in monitoring *P. parvum* populations across Hickling Broad over a 20 month period, and propose that this could be used to predict when blooms of *P. parvum* are likely to occur. However, as we describe earlier, we suspect that viral lysis of natural populations of *P. parvum* are a cause of toxin release (Figure 60), meaning that monitoring of *P. parvum* populations alone is not enough to predict *toxic* blooms of this species. Since the discovery and sequencing of PpDNAV-BW1, we have worked to expand our qPCR analysis to include monitoring of PpDNAV-BW1. With significant input from collaborating scientist Dr Jennifer Pratscher (UEA), we have recently developed a specific and sensitive qPCR assay for PpDNAV-BW1 based on sequences encoding the major capsid protein (MCP). Using the same nucleic acid samples that were previously used to monitor algal abundance, we have been able to produce a seasonal profile of virus abundance with relation to its algal host (Figure 61). The findings from this dataset display typical Lotka-Volterra predator:prey population dynamics [25], suggesting a key role for this virus in control of the native *P. parvum* population in Hickling Broad. As is seen in Figure 61, increases in abundance of viral transcripts directly correlate with decreases in abundance of algal transcripts. We propose that the regular use of these qPCR assays will allow the Environment Agency to predict when a toxic bloom of *P. parvum* is likely to occur.

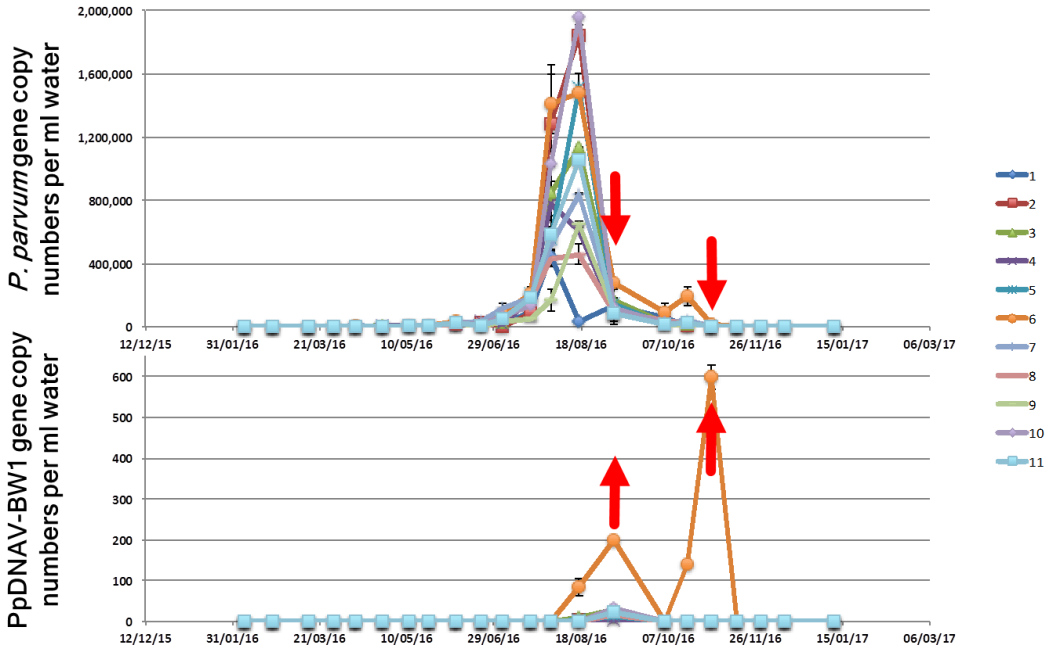


Figure 61 – Seasonal abundances of *P. parvum* and PpDNAV. (Top) Abundance of *P. parvum* as determined by copies of internal transcribed spacer (ITS) sequence. (Bottom) Abundance of PpDNAV-BW1 as determined by copies of MCP1 sequences based on the following primers: MCP1-1F 5'-TGTCTGCCGTGGACTTAGTGCT-3', MCP1-1R 5'-ATGGCACAACGACTTGGT-3' (unpublished). Decreases in algal transcripts coincide with sharp increases in viral transcripts (red arrows).

6.2.2 Control of *P. parvum* blooms using hydrogen peroxide

Alongside the work documented in this thesis, I championed, with the help of many collaborating scientists and government agencies, field trials for the use of hydrogen peroxide in the management of *P. parvum* blooms on Hickling Broad. Previous work has shown hydrogen peroxide to be an effective algaecide in the management of toxic algal blooms, but reports of its use in the U.K are sparse [26, 27]. Furthermore, work carried out by myself showed in a laboratory setting that low concentrations of H_2O_2 (>20 mg L^{-1}) were effective at reducing levels of both *P. parvum* and its polyether prymnesin toxins over a 48-hour period. To build on this, we sought to assess whether H_2O_2 was effective at reducing levels of *P. parvum* on Hickling Broad. Importantly, we wanted to show that the doses required to kill *P. parvum* (and presumably other phytoplankton) were not toxic to macroinvertebrates such as water fleas (*Daphnia* sp.), or fish. Field trials were subsequently set up at Whispering Reeds Boatyard, Hickling Broad, where H_2O_2 was applied to a small area at a final concentration of 30-40 mg L^{-1} over a 4-hour period (Figure 62).

Figure 62 – Hydrogen peroxide field trials for management of *P. parvum* blooms. Image taken on 29/06/2017 by Martin Rejzek at Whispering Reeds Boatyard on Hickling Broad, United Kingdom.

Water samples were then taken at regular intervals following the peroxide application and nucleic acids were extracted. Peroxide concentrations were followed closely throughout the treatment using QUANTOFIX® Peroxide 100 test strips (Sigma Aldrich, U.K). Using the qPCR assay outlined in Chapter 2 of this thesis, we were able to show that areas treated with peroxide showed a marked loss in *P. parvum* cell numbers over the next 24 hours (even under non-bloom steady state conditions at the time of the trial), compared to control locations that were not sprayed with peroxide (Figure 63). *P. parvum* numbers returned to normal non-bloom background levels in all locations 96 hours after treatment. Importantly no adverse effects on macroinvertebrates or fish were noted throughout the field trial. The success of this trial has now led the Environment Agency to re-design protocols for the practical management of *P. parvum* blooms, with a view to incorporating H_2O_2 in their waterways management strategies.

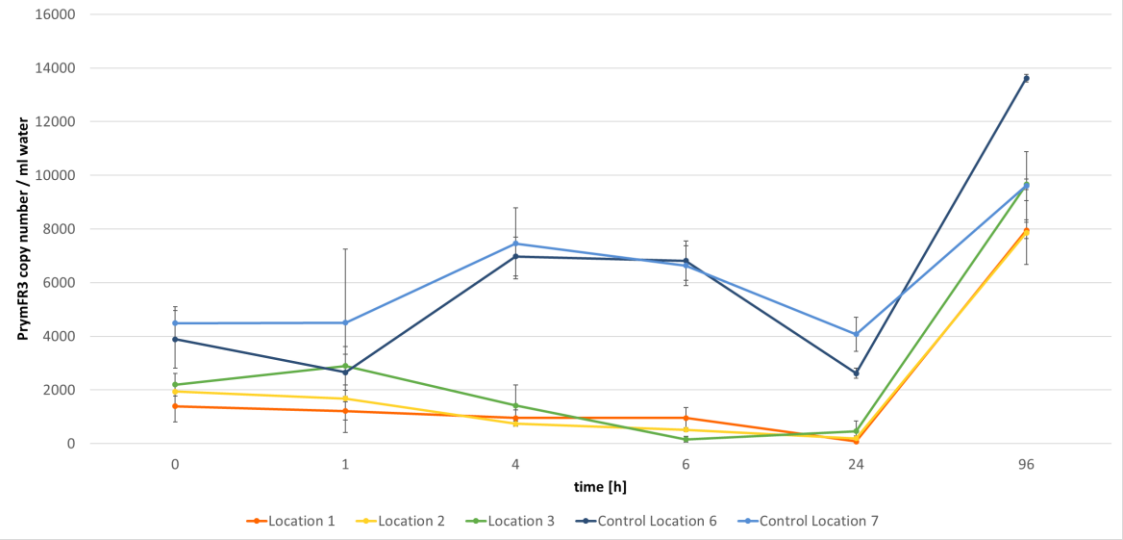


Figure 63 - *P. parvum* specific qPCR quantification for hydrogen peroxide trial – Whispering Reeds Boatyard, Hickling Broad (29/06/2017). The abundance of *P. parvum* was monitored at 3 locations treated with hydrogen peroxide at 30 mg L^{-1} doses (location 1, 2, 3) and at 2 control locations (Control location 6, 7) over a 96-hour period after H_2O_2 application. Whilst P. parvum abundance rose slightly at control locations, locations treated with peroxide saw levels of P. parvum drop over a 24-hour period. By 96-hours post application all locations saw levels rise to around 100,000 copies ml^{-1}.

Overall, the findings presented in this thesis reveal new insights into the bloom dynamics of *Prymnesium parvum* and its' glycobiology. The discovery of a novel species of virus highlights an importance of algal viruses in regulation of toxic blooms of *P. parvum*, and may represent a globally significant mechanism for toxin-release from harmful algae. This new model system of a toxin-producing alga and lytic virus can now be explored to address many key questions that have emerged during this study.

6.3 References

1. Kozakai, H.; Oshima, Y.; Yasumoto, T., Isolation and structural elucidation of hemolysin from the phytoflagellate *Prymnesium parvum*. *Agric. Biol. Chem.* **1982,** 46, (1), 233-236.
2. Henrikson, J. C.; Gharfeh, M. S.; Easton, A. C.; Easton, J. D.; Glenn, K. L.; Shadfan, M.; Mooberry, S. L.; Hambright, K. D.; Cichewicz, R. H., Reassessing the ichthyotoxin profile of cultured *Prymnesium parvum* (golden algae) and comparing it to samples collected from recent freshwater bloom and fish kill events in North America. *Toxicon* **2010,** 55, (7), 1396-404.
3. Bertin, M. J.; Zimba, P. V.; Beauchesne, K. R.; Huncik, K. M.; Moeller, P. D. R., The contribution of fatty acid amides to *Prymnesium parvum* Carter toxicity. *Harmful Algae* **2012,** 20, 117-125.
4. Bertin, M. J.; Zimba, P. V.; Beauchesne, K. R.; Huncik, K. M.; Moeller, P. D. R., Identification of toxic fatty acid amides isolated from the harmful alga *Prymnesium parvum* carter. *Harmful Algae* **2012,** 20, 111-116.
5. Igarashi, T.; Satake, M.; Yasumoto, T., Structures and partial stereochemical assignments for prymnesin-1 and prymnesin-2: potent hemolytic and ichthyotoxic glycosides isolated from the red tide alga *Prymnesium parvum*. *J. Am. Chem. Soc.* **1999,** 121, (37), 8499-8511.
6. Igarashi, T.; Satake, M.; Yasumoto, T., Prymnesin-2: a potent ichthyotoxic and hemolytic glycoside isolated from the red tide alga *Prymnesium parvum*. *J. Am. Chem. Soc.* **1996,** 118, (2), 479-480.
7. Rasmussen, S. A.; Andersen, A. J. C.; Andersen, N. G.; Nielsen, K. F.; Hansen, P. J.; Larsen, T. O., Chemical diversity, origin, and analysis of phycotoxins. *J. Nat. Prod.* **2016,** 79, (3), 662-673.
8. Manning, S. R.; La Claire Ii, J. W., Isolation of polyketides from *Prymnesium parvum* (Haptophyta) and their detection by liquid chromatography/mass spectrometry metabolic fingerprint analysis. *Anal. Biochem.* **2013,** 442, (2), 189-195.
9. Blossom, H. E.; Rasmussen, S. A.; Andersen, N. G.; Larsen, T. O.; Nielsen, K. F.; Hansen, P. J., *Prymnesium parvum* revisited: Relationship between allelopathy, ichthyotoxicity, and chemical profiles in 5 strains. *Aquat. Toxicol.* **2014,** 157, 159-166.
10. Rasmussen, S. A.; Meier, S.; Andersen, N. G.; Blossom, H. E.; Duus, J. Ø.; Nielsen, K. F.; Hansen, P. J.; Larsen, T. O., Chemodiversity of ladder-frame prymnesin polyethers in *Prymnesium parvum*. *J. Nat. Prod.* **2016,** 79, (9), 2250-2256.

11. Granéli, E.; Johansson, N., Increase in the production of allelopathic substances by *Prymnesium parvum* cells grown under N- or P-deficient conditions. *Harmful Algae* **2003,** 2, (2), 135-145.
12. Remmel, E. J.; Hambright, K. D., Toxin-assisted micropredation: experimental evidence shows that contact micropredation rather than exotoxicity is the role of Prymnesium toxins. *Ecol. Lett.* **2012,** 15, (2), 126-132.
13. Malin, G.; Wilson, W. H.; Bratbak, G.; Liss, P. S.; Mann, N. H., Elevated production of dimethylsulfide resulting from viral infection of cultures of *Phaeocystis pouchetii*. *Limnol. Oceanogr.* **1998,** 43, (6), 1389-1393.
14. La Claire Ii, J. W.; Manning, S. R.; Talarski, A. E., Semi-quantitative assay for polyketide prymnesins isolated from *Prymnesium parvum* (Haptophyta) cultures. *Toxicon* **2015,** 102, 74-80.
15. Fulton, J. M.; Fredricks, H. F.; Bidle, K. D.; Vardi, A.; Kendrick, B. J.; DiTullio, G. R.; Van Mooy, B. A. S., Novel molecular determinants of viral susceptibility and resistance in the lipidome of *Emiliania huxleyi*. *Environ. Microbiol.* **2014,** 16, (4), 1137-1149.
16. Schauer, R., Sialic acids as regulators of molecular and cellular interactions. *Curr. Opin. Struct. Biol.* **2009,** 19, (5), 507-514.
17. Varki, A.; Schauer, R., Sialic Acids. In *Essentials of Glycobiology*, Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor Laboratory Press. The Consortium of Glycobiology Editors, La Jolla, California.: Cold Spring Harbor (NY), 2009.
18. Wang, L.; Lu, Z.; Allen, K. N.; Mariano, P. S.; Dunaway-Mariano, D., Human symbiont *Bacteroides thetaiotaomicron* synthesizes 2-keto-3-deoxy-D-glycero-D-galacto-nononic acid (KDN). *Chem. Biol.* **2008,** 15, (9), 893-897.
19. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic local alignment search tool. *J. Mol. Biol.* **1990,** 215, (3), 403-10.
20. Coordinators., N. R., Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **2017,** 45, (D1), D12-d17.
21. Keeling, P. J.; Burki, F.; Wilcox, H. M.; Allam, B.; Allen, E. E.; Amaral-Zettler, L. A.; Armbrust, E. V.; Archibald, J. M.; Bharti, A. K.; Bell, C. J.; Beszteri, B.; Bidle, K. D.; Cameron, C. T.; Campbell, L.; Caron, D. A.; Cattolico, R. A.; Collier, J. L.; Coyne, K.; Davy, S. K.; Deschamps, P.; Dyhrman, S. T.; Edvardsen, B.; Gates, R. D.; Gobler, C. J.; Greenwood, S. J.; Guida, S. M.; Jacobi, J. L.; Jakobsen, K. S.; James, E. R.; Jenkins, B.; John, U.; Johnson, M. D.; Juhl, A. R.; Kamp, A.; Katz, L. A.; Kiene, R.; Kudryavtsev, A.;

Leander, B. S.; Lin, S.; Lovejoy, C.; Lynn, D.; Marchetti, A.; McManus, G.; Nedelcu, A. M.; Menden-Deuer, S.; Miceli, C.; Mock, T.; Montresor, M.; Moran, M. A.; Murray, S.; Nadathur, G.; Nagai, S.; Ngam, P. B.; Palenik, B.; Pawlowski, J.; Petroni, G.; Piganeau, G.; Posewitz, M. C.; Rengefors, K.; Romano, G.; Rumpho, M. E.; Rynearson, T.; Schilling, K. B.; Schroeder, D. C.; Simpson, A. G. B.; Slamovits, C. H.; Smith, D. R.; Smith, G. J.; Smith, S. R.; Sosik, H. M.; Stief, P.; Theriot, E.; Twary, S. N.; Umale, P. E.; Vaulot, D.; Wawrik, B.; Wheeler, G. L.; Wilson, W. H.; Xu, Y.; Zingone, A.; Worden, A. Z., The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* **2014,** 12, (6), e1001889.

22. Giraud, M. F.; Naismith, J. H., The rhamnose pathway. *Curr. Opin. Struct. Biol.* **2000,** 10, (6), 687-96.
23. Dong, C.; Beis, K.; Giraud, M. F.; Blankenfeldt, W.; Allard, S.; Major, L. L.; Kerr, I. D.; Whitfield, C.; Naismith, J. H., A structural perspective on the enzymes that convert dTDP-D-glucose into dTDP-L-rhamnose. *Biochem. Soc. Trans.* **2003,** 31, (Pt 3), 532-6.
24. Oka, T.; Nemoto, T.; Jigami, Y., Functional analysis of *Arabidopsis thaliana* RHM2/MUM4, a multidomain protein involved in UDP-D-glucose to UDP-L-rhamnose conversion. *J. Biol. Chem.* **2007,** 282, (8), 5389-403.
25. Yorke, J. A.; Anderson, W. N., Predator-prey patterns. *Proc. Natl. Acad. Sci. U.S.A.* **1973,** 70, (7), 2069-2071.
26. Burson, A.; Matthijs, H. C. P.; de Bruijne, W.; Talens, R.; Hoogenboom, R.; Gerssen, A.; Visser, P. M.; Stomp, M.; Steur, K.; van Scheppingen, Y.; Huisman, J., Termination of a toxic Alexandrium bloom with hydrogen peroxide. *Harmful Algae* **2014,** 31, (0), 125-135.
27. Matthijs, H. C.; Visser, P. M.; Reeze, B.; Meeuse, J.; Slot, P. C.; Wijn, G.; Talens, R.; Huisman, J., Selective suppression of harmful cyanobacteria in an entire lake with hydrogen peroxide. *Water Res.* **2012,** 46, (5), 1460-72.

Research Papers

1. **Wagstaff, B**.; Vladu, I.; Barclay, J.; Schroeder, D.; Malin, G.; Field, R., Isolation and Characterization of a Double Stranded DNA Megavirus Infecting the Toxin-Producing Haptophyte *Prymnesium parvum*. *Viruses* **2017,** 9, (3), 40.