Characterization of exocellular DNA in the oligotrophic ocean

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI’I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

OCEANOGRAPHY

MAY 2021

By:
Morgan Donerly Linney

Dissertation Committee:
David M. Karl, Chairperson
Edward F. DeLong
Grieg F. Steward
Angelicque E. White
Edward G. Ruby
Acknowledgments:

I will begin by saying thank you David Karl, my PhD advisor. I feel incredibly grateful to have learned from him. Not only have I learned about oceanography, but I’ve learned what makes a fierce and caring leader. One day after a trans-Pacific cruise I went into his office to ask questions about science, and nearly 7 years later am leaving with a degree. Thank you for teaching me to think big, I am honored to have worked with and learned from the Karl lab.

I am fortunate to have an excellent dissertation committee, and would like to thank Grieg Steward, Angel White, Ed DeLong, and Ned Ruby. Thank you for your thoughtful comments and invaluable advise. I would also like to thank all current and past BEACH (Karl) Lab members and the HOT program. Thank you for your mentorship and teaching me the value (and joy) of teamwork. Special thanks to Sam Wilson for keeping me on track and believing in me when I did not.

In addition to my academic advisors, I feel incredibly grateful for all of my friends the years in Hawaii. These were formative years, and I thank you for sticking with me and growing with me. I met many of you at the weekly SOEST TGs – the exhale at the end of a long work week, a time to gather across disciplines.

Most of all, I would like to thank my family – my mom, Sue Donerly for answering phone calls at all hours and being there when I needed, my dad, Elwood Linney, for writing a morning email every day to me and my sister since we left home, and my big sister Caitlin Linney, you are my best friend – I love you all. I feel incredibly lucky to have such a close family. I wish I could better express what you all mean to me.

If you are reading this as a student, or a forever student just know that inner negative voice is louder, but smaller than any other persons’. Do your best to set aside your insecurities, they are often the only thing holding you back from success and simply enjoying life.

Funding for this dissertation was generously provided by the Simons Foundation (Award ID 329108 to Dave Karl and Ed DeLong)
Abstract: Approximately half of the DNA in the open ocean is present outside of living cells. Together, this exocellular DNA (or dissolved DNA; D-DNA) is comprised of truly dissolved “free” DNA (F-DNA), virion encapsidated DNA, and DNA inside of membrane vesicles. It is ubiquitous in nature yet its sources, sinks, and ecological characteristics are largely unknown. One reason for the uncertainty is the methodological limitation that precludes a distinction among the three pools that comprise D-DNA. Using a novel method that provides complete separation of vesicles, viruses, and F-DNA, the first fully sequenced open ocean water-column profile of exocellular DNA was obtained. Euphotic zone F-DNA (75-125 m) contained mostly bacterial and viral sequences, with bacteria dominating in the mesopelagic zone (500-1000 m). A high proportion of mesopelagic zone (500 and 1000 m) F-DNA sequences appeared to originate from surface waters, including a large amount of DNA contributed by high-light Prochlorococcus ecotypes. These results indicate the composition of F-DNA in different regions of the water-column (euphotic and mesopelagic) and suggest potential mechanisms for dissolved organic matter cycling and export. Experiments designed to examine the dynamics of F-DNA suggest that it is produced by viral lysis of microbial cells, and may be consumed by heterotrophic bacteria and protists. Collectively, this research provides novel insights into the microbial origins and dynamics of F-DNA in the open ocean.
# Table of Contents

Acknowledgments ........................................................................................................... ii
Abstract .......................................................................................................................... iii
List of Tables .................................................................................................................. vi
List of Figures ............................................................................................................... vii

Chapter 1: Exocellular DNA in the sea, research objectives ............................................. 1
1.1 Oligotrophic ocean ecosystems are driven by marine microorganisms ............. 2
1.2 Dissolved DNA: a labile constituent of dissolved organic matter ................. 4
1.3 Dynamics of D-DNA in the ocean ....................................................................... 6
1.4 Challenges in isolating D-DNA constituents .................................................... 9
1.5 Research objectives ............................................................................................. 12
1.6 References ............................................................................................................ 14
1.7 Figures .................................................................................................................. 20

Chapter 2: A method for characterizing dissolved DNA and its application to the North Pacific Subtropical Gyre ................................................................. 24
2.1 Abstract .................................................................................................................. 25
2.2 Introduction ............................................................................................................ 26
2.3 Method ..................................................................................................................... 29
  2.3.1 Collection and prefiltration ........................................................................ 29
  2.3.2 Ultrafiltration (tangential flow and/or centrifugal) .................................... 30
  2.3.3 Density gradient separation ....................................................................... 30
  2.3.4 Quantification of isolated DNA ................................................................. 32
  2.3.5 Fragment analysis ...................................................................................... 33
2.4 Assesment .............................................................................................................. 33
  2.4.1 Transmission electron microscopy ........................................................... 34
  2.4.2 Epifluorescence microscopy .................................................................... 35
  2.4.3 DNase treatment ...................................................................................... 36
  2.4.4 Protein quantification .............................................................................. 37
  2.4.5 Method validation ..................................................................................... 37
    2.4.5.1 Prefiltration ..................................................................................... 38
    2.4.5.2 Internal standard curve and recovery ............................................ 39
    2.4.5.3 Low-molecular weight filtrate DNA ............................................ 40
    2.4.5.4 Field application .......................................................................... 41
2.5 Discussion .............................................................................................................. 42
2.6 Comments and recommendations ...................................................................... 46
2.7 References .............................................................................................................. 47
2.8 Acknowledgments ............................................................................................... 51
Chapter 3: Microbial sources of exocellular DNA

3.1 Abstract...........................................................................................................62
3.2 Introduction.......................................................................................................63
3.3 Methods...........................................................................................................66
  3.3.1 Sample collection.......................................................................................66
  3.3.2 Molecular weight determination of environmental F-DNA..................67
  3.3.3 DNA extraction and purification.................................................................67
  3.3.4 Metagenomic library preparation...............................................................68
  3.3.5 DNA sequence analysis and annotation.....................................................69
  3.3.6 Comparison to cellular and viral datasets collected from Station ALOHA.................................................................70
3.4 Results.............................................................................................................71
  3.4.1 Microbial composition of exocellular free DNA and other D-DNA pools....................................................................................71
  3.4.2 Depth of origin of D-DNA throughout the water column.......................75
  3.4.3 Size distributions of environmental F-DNA through the water column....................................................................................76
  3.4.4 Comparing vesicle, viral, and F-DNA fractions by non-metric multidimensional scaling.................................................................77
3.5 Discussion........................................................................................................79
3.6 References.......................................................................................................88
3.7 Acknowledgments.........................................................................................95
3.8 Figures............................................................................................................96
3.9 Supplementary.............................................................................................99

Chapter 4: Impacts of protistan grazing and viral lysis on exocellular DNA and ATP production

4.1 Abstract........................................................................................................111
4.2 Introduction...................................................................................................111
4.3 Methods.......................................................................................................115
  4.3.1 Experimental system.................................................................................115
  4.3.2 Intracellular nutrient and DNA calculations..........................................116
  4.3.3 Exocellular DNA measurements.............................................................117
  4.3.4 Dissolved ATP.........................................................................................119
  4.3.5 Viral abundances and quantitative PCR (qPCR)....................................119
4.4 Results..................................................................................................................120
  4.4.1 Minimal exocellular DNA and ATP produced by growing
  Prochlorococcus (Control)................................................................................120
  4.4.2 Exocellular DNA and ATP released by viral-lysis of
  Prochlorococcus................................................................................................121
  4.4.3 Grazer consumption of Prochlorococcus mortality products.....123
4.5 Discussion.............................................................................................................125
4.6 References............................................................................................................131
4.7 Acknowledgments.................................................................................................136
4.8 Figures..................................................................................................................137
4.10 Equation..............................................................................................................144

Chapter 5: Conclusions and future directions.......................................................145
List of Tables

Chapter 2:

Table 2.1  Comparison of methods to extract dissolved DNA (D-DNA) from seawater.................................................................52

Table 2.2  Microscopy and molecular characterizations of the CsCl gradient pools.................................................................54

Chapter 3:

Table 3.S1  Total number of gene counts for each dissolved DNA sample (vesicles, viruses, and free DNA)........................................102

Table 3.S2  Proportion of Prochlorococcus HL sequences present in each dissolved metagenomic DNA sample (vesicles, viruses, and free DNA/FDNA)...............................................................103

Table S3  Microscopy and molecular characterizations of the cesium chloride gradient pools..........................................................104
List of Figures

Chapter 1:

Figure 1.1 Water masses present at Station ALOHA in the North Pacific Subtropical Gyre…………………………………………………………………………………20

Figure 1.2 Karl & Church (2014) schematic illustrating the habitat characteristics of Station ALOHA…………………………………………………………………………………21

Figure 1.3 Mende et al. (2019) figure representing the abundances of different microorganisms found in the water-column at Station ALOHA…………………22

Figure 1.4 Constituents of dissolved DNA (D-DNA)………………………………23

Chapter 2:

Figure 2.1 Transmission electron micrographs of the dominant structures observed in D-DNA pools ……………………………………………………………………………………55

Figure 2.2 Molecular size-dependent recovery of ladder and environmental F-DNA………………………………………………………………………………………………………56

Figure 2.3 Molecular weight distribution for environmental F-DNA recovered from three depths at Station ALOHA (75, 125, and 500 m)…………………………57

Figure 2.4 Depth profiles of various pools of dissolved DNA constituents collected from the North Pacific Subtropical Gyre…………………………………………58

Figure 2.S1 Habitat characteristics of Station ALOHA during sample collection……………………………………………………………………………………………………59

Figure 2.S2 Typical examples of D-DNA distributions measured across 0.5 mL fractions of cesium chloride density gradients……………………………………60

Chapter 3:
Figure 3.1  Taxonomic annotation of metagenomic sequences from the three dissolved DNA constituents collected from the North Pacific Subtropical Gyre...........................................................................................................96

Figure 3.2  Annotation of D-DNA metagenomic sequences suggests their probable depths of origin.........................................................................................................................97

Figure 3.3  Fragment analysis of F-DNA samples from the North Pacific Subtropical Gyre........................................................................................................................................98

Figure 3.S1  Diversity of host-specified phages contributing to the three dissolved DNA pools.......................................................................................................................................105

Figure 3.S2A-B  Bray-Curtis dissimilarity based non-metric multidimensional scaling (NMDS) of D-DNA fractions....................................................................................................................106

Figure 3.S3  Nutrients and cell abundance depth profiles at Station ALOHA........................................................................................................................................107

Chapter 4:

Figure 4.1  Schematic representation of total D-DNA and F-DNA quantification for all samples........................................................................................................................................137

Figure 4.2  Time-course of D-DNA and F-DNA quantified over the time course of MED4 control.....................................................................................................................................138

Figure 4.3  Time-course of D-ATP quantified over the time course of MED4 control........................................................................................................................................139

Figure 4.4  Time-course of D-DNA and F-DNA quantified over the time course of MED4 challenged by a virus (P-SSP7)..............................................................................................................140

Figure 4.5  Time-course of D-ATP quantified over the time course of MED4 challenged by a virus (P-SSP7)........................................................................................................................................141
Figure 4.6 Time-course of D-DNA and F-DNA quantified over the time course combined grazer- and viral-driven mortality of MED4 (“combination 1”)……………………………………………………………………………………………………………………142

Figure 4.7 Time-course of D-DNA and F-DNA quantified over the time course combined grazer- and viral-driven mortality of MED4 (“combination 2”)……………………………………………………………………………………………………………………143
Chapter 1: Exocellular DNA in the sea, research objectives
Morgan D. Linney

1Daniel K. Inouye Center for Microbial Oceanography: Research and Education (C-MORE) and Department of Oceanography, University of Hawaii at Manoa, Honolulu, HI 96822 USA

Keywords: microbial oceanography, dissolved organic matter, dissolved organic carbon, North Pacific Ocean, DNA, dissolved DNA, exocellular DNA, free DNA, marine viruses, extracellular vesicles
1.1 Oligotrophic ocean ecosystems are driven by marine microorganisms

In the open ocean, nutrient-deplete subtropical gyres comprise 40% of the earth’s surface (Karl and Church 2014), and are inhabited by diverse marine microorganisms: microeukaryotes, viruses, archaea, and bacteria. Microorganisms comprise 90% of the oceanic biomass (Cavicchioli et al. 2019) and play integral roles in supporting nearly half of global photosynthesis (Field et al. 1998) and the cycling of biocritical nutrients (carbon, nitrogen, and phosphorus).

Representative of many oligotrophic ocean systems, the North Pacific Subtropical Gyre (NPSG) is the largest contiguous biome on earth, and the home to the Hawaii Ocean Time-series (HOT) research site, Station ALOHA (A Long-term Oligotrophic Habitat Assessment; 22° 45'N, 158° 00'W). Over thirty years of biogeochemical research has been conducted at Station ALOHA, and provides a wealth of interdisciplinary baseline data: hydrographic data and its implications for thermohaline circulation, metabolic balances, nutrient cycling, cultured microbial isolates, complex microbial interactions, and export dynamics. More recently, in-depth gene catalogs of both cellular (Mende et al. 2019) and viral communities (Luo et al. 2020) have been generated, which have broadened our understanding of the distributional patterns of diverse microbial communities in the open ocean.

Station ALOHA exhibits relatively low seasonality (Church et al. 2013), and sampling expands the entire water column (4,740 m), and is characterized by distinct
water masses (Figure 1.1). Surface waters are warm >24°C and low in inorganic nutrients (nitrate, nitrite, and phosphate), high in turbulent mixing (within the 0-50 m), with sunlight reaching to ~200 m (Figure 1.2). Within this sunlit region, photosynthetic microorganisms form the base of the marine food web, fueling sequentially higher trophic levels with fixed carbon. In this region, the 26-year average primary production ($^{14}$C-based measurements) is 536.8 ($\pm$135) mg C m$^{-2}$ d$^{-1}$ (Karl et al. 2021), with approximately 5.4 ($\pm$2.1)% of the depth-integrated primary production is exported out of the euphotic zone (Karl et al. 2021). Below the sunlit ocean, temperatures become colder at depth (<5°C), inorganic nutrients are higher, and organic nutrients are limiting to microbial growth. The microbial diversity within the stratified water-column at Station ALOHA is reflective of these distinct regions (Figure 1.3). Photosynthetic Prochlorococcus dominates the sunlit zone, heterotrophic Pelagibacter ubique (SAR11) is abundant throughout the entire water column, as well as, abundant marine archaea some of which play vital roles in ammonia oxidation (Church et al. 2010). Both Prochlorococcus and SAR11 are comprised of distinct clades dependent on adaptation to different environmental conditions, (e.g. temperature, light-availability, phage resistance, and nutrient sources) (Johnson et al. 2006; Grote et al. 2012). Even with thorough genetic characterizations there are still many novel and uncharacterized microorganisms with unknown metabolisms and ecological functions.
With the wealth of oceanographic research conducted at Station ALOHA, it provides an excellent backdrop from which to pursue new investigations into microbial dynamics. It is here that this thesis research was conducted, providing a new dataset to include alongside decades of oceanographic research – together helping us to learn more about the complexities of microbial ecology.

1.2 Dissolved DNA: a labile constituent of dissolved organic matter

Dissolved organic carbon (DOC) is one of the largest reservoirs of cycled carbon on Earth. Approximately 660 Gt carbon is in the form of DOC, and is comparable to the reservoir of atmospheric carbon dioxide (Hansell et al. 2009). Consequently, the composition and cycling of DOC has implications for climate change and the inventory of carbon on Earth. Dissolved DNA (operationally defined as DNA that passes through a 0.1 or 0.2 µm filter) is a labile component of DOC. It exists as: cell-free DNA (F-DNA), and as viruses and extracellular vesicles (Figure 1.4). The availability of D-DNA has both ecological and evolutionary significance, as a component of all living cells and one of the most important biological molecules – it is the blueprint for life. Additionally, DNA has an approximate C:N:P ratio of 10:4:1, making it a source of nitrogen and phosphorus, which are often limiting to primary producers in open-ocean ecosystems (Karl & Church 2014).
D-DNA has been documented in a variety of both marine and freshwater environments. D-DNA concentrations have a similar concentration pattern as DOC, decreasing with distance from shore and depth (DeFlaun et al. 1987; Karl & Bailiff 1989; Weinbauer et al. 1993; Weinbauer et al. 1995). Utilizing an ethanol precipitation method DeFlaun et al. (1987) found oceanic concentrations in the Gulf of Mexico to range from 0.20 to 1.9 µg L⁻¹. By a different method utilizing cetyltrimethylammonium bromide (CTAB) to precipitate DNA, Karl & Bailiff (1989) found between 0.56-1.4 µg L⁻¹ of D-DNA in the North Pacific Subtropical Gyre (NPSG). In both of these open ocean systems D-DNA concentrations have been found to range from a quarter to a half of the particulate DNA fraction (DeFlaun et al. 1987; Karl & Bailiff 1989). In other words, nearly half the DNA in the ocean is found outside of living cells. D-DNA concentrations are typically much higher in coastal systems, and has been measured in the Bombay Harbor (Minear 1972), Kaneohe Bay (Karl & Bailiff 1989), Northern Adriatic Sea (Turk et al. 1992), Tampa Bay (DeFlaun et al. 1987), and Bransfield Strait (Bailiff & Karl 1991). For estuarine environments the range of concentrations that have been reported are 5-45 µg L⁻¹, and 2-15µg L⁻¹ for coastal habitats (DeFlaun et al. 1987; Karl & Bailiff 1989; Boehme et al. 1993). D-DNA has also been documented as a component of marine sediments, and may comprise the majority of total DNA in sediments (Dell’Anno & Danovaro 2005; Torti et al. 2015). Based on a method that pre-treated sediments with DNase enzymes, Dell’Anno et al. (2002) found 90% of the total DNA to be
enzymatically hydrolysable. If this holds true for all marine sediments this could be the largest reservoir of DNA in the world’s oceans, accounting for an estimated 0.3-0.45 Gt of DNA in deep-sea sediments (Dell’Anno & Danovaro 2005). Extracellular DNA throughout diverse marine environments, has the potential to play important roles in gene transfer and biogeochemical cycles.

1.3 Dynamics of D-DNA in the ocean

D-DNA can be produced by viral lysis (Weinbauer et al. 1993; Alonso et al. 2000; Brum 2005; Riemann et al. 2008), programmed cell death/autolysis, protistan grazing (Turk et al. 1992; Alonso et al. 2000), exposure to cytotoxic agents (Torti et al. 2015), and as exudates from growing bacteria (Paul et al. 1990). For instance, Turk et al. (1992) used $^{32}$P labelled DNA to study the cycling of D-DNA. They found increased production of D-DNA in the presence of nanoflagellates that were presumably grazing on bacteria. In turn, the labelled D-DNA was then consumed by bacteria and cycled into inorganic phosphorus. This study illustrates how D-DNA can be involved in the turnover of phosphorus between organic and inorganic pools, but did not compare this to the possible fate of nitrogen and carbon originating from D-DNA. Furthermore, in a P-limited estuary of the Baltic Sea, Riemann et al. (2008) found viruses to be an important mediator in the production of D-DNA and the recycling of carbon and phosphorus. Throughout the estuary D-DNA produced by viral lysis was then available
to bacterioplankton, and comprised an estimated 20-46% and 14-32% of their phosphate and carbon demands, respectively. These calculations are based on a number of assumptions regarding nutrient content of cells and growth rates, which may not be accurate for a mixed microbial community. They found that about a quarter of the total D-DNA pool was produced by viral lysis of bacterioplankton cells (Riemann et al. 2008). These studies suggest that multiple mechanisms may lead to the accumulation of D-DNA.

Following production, D-DNA may be converted to single stranded, linear polynucleotides during bacterial transformation (Frisher et al. 1990), or hydrolyzed by nucleases and utilized as a nutrient source or nucleotide precursor (Paul et al. 1987; Turk et al. 1992). Accordingly, Maeda & Taga (1973) reported the presence of deoxyribonuclease activity in samples taken from Tokyo Bay seawater and marine sediments. Another study by Paul et al. (1987) documented the hydrolysis of DNA in the marine environment both by cell-associated and extracellular nucleases. The hydrolyzed products were followed using radiotracers and found to cycle through the particulate and dissolved DNA fractions. In open-ocean systems a majority (50%) of the DNase activity was found in the <0.45 µm fraction, and only 10% in the dissolved <0.2 µm fraction suggesting more cell-associated nucleases in these environments. This is supported by Hollibaugh & Azam (1983) who found that physical association between bacteria and proteins was necessary for degradation and thereby uptake of hydrolyzed...
products. This finding suggests it is an energetic disadvantage for planktonic organisms to utilize exoenzymes due to enzyme dilution, however exoenzymes may be more efficient in environments where molecular diffusion is restricted (i.e. sediments and particles). Later on, Lennon (2007) found that distinct microbial communities originally collected from Woods Hole, MA were able to be cultivated on low molecular weight (LMW) and high molecular weight (HMW) DNA, and is suggested to be due to differences in distinct microbial abilities to produce nucleases. This may also indicate a succession of microbial communities that consume D-DNA. Together, these studies exemplify possible mechanisms for the hydrolysis and cycling of D-DNA.

Turnover of the D-DNA pool is dependent on the above mentioned production and uptake mechanisms, however, few studies have investigated the dynamics of D-DNA. Utilizing [³H]thymidine additions to seawater, Paul et al. (1987) estimated turnover time to be 22.7-146 days in the Gulf of Mexico. This is assuming that the only production of D-DNA is from heterotrophically growing bacterioplankton, as thymidine is not taken up by autotrophs (Fuhrman & Azam 1980). Turk et al. (1992) found a difference in turnover between P-limited north Adriatic Sea water (0.4 µg L⁻¹ hr⁻¹ degradation rate) and N-limited southern California water (0.002 µg L⁻¹ hr⁻¹). Labelled DNA turned over more rapidly in the P-limited environment than the N-limited one. Also in the P-limited environment, inorganic phosphate values increased, which was coupled to DNA turnover. Brum (2005) found that different components of the D-DNA
pool turned over at dissimilar rates at Station ALOHA. Enzymatically hydrolyzable DNA (proxy for F-DNA) turned over within 0.97-6.2 hours while viral D-DNA took 9.6-24 hours, suggesting that different components of the D-DNA pool may vary in their production, bioavailability, and turnover.

1.4 Challenges in isolating D-DNA constituents

More recent studies of D-DNA dynamics demonstrate the importance in making separate quantification and turnover measurements of D-DNA components. Viruses and vesicles that are found in the <0.1 µm fraction were only discovered in the ocean in 1979 and 2014, respectively (Torrella & Morita 1979; Biller et al. 2014). Earlier studies of D-DNA separated fractions by molecular weight (DeFlaun et al. 1987) and using 0.05 µm filters (Karl & Bailiff 1989), but were not able to distinguish F-DNA from viruses and vesicles. These discoveries redefine the definition of D-DNA and requires a distinction among these fractions, as the colloidal fractions originate from different sources and are suggested to have different dynamics (Brum 2005).

F-DNA may persist in the ocean due to common DNA modifications: protein binding, phosphorothioation, histones, and methylation. These alterations could make F-DNA less accessible to hydrolytic enzymes and degradation. A possible candidate for F-DNA is phosphorothioated DNA which is resistant to many nuclease activities and present in some marine bacteria (Eckstein 2000; Wang et al. 2007). Phosphorothioates
are used in medicine as an essential component of therapeutic antisense oligonucleotides to control gene expression (Eckstein 2014). Phosphorothioated DNA has been found in *Pelagibacter ubique* which is one of the most abundant microbes in open-ocean systems (Morris et al. 2002). Phosphorothioation is found in prokaryotes from diverse habitats making it possible this is a pervasive trait in microbes, and is proposed to be part of a restriction-modification system providing defense against foreign DNA (Wang et al. 2011). Additionally, protein-bound DNA is another candidate for F-DNA. Both prokaryotic and eukaryotic cells contain genome packaging proteins which help assemble and compress their genetic material. Eukaryotic cells and some archaea package their DNA around histone proteins, which create nucleosome structures. Lacking a nucleus, prokaryotes supercoil their DNA which works to package their genetic material and regulate gene expression. A host of proteins (e.g. nucleoid-associated HU, topoimerase I, and more) help coil the DNA and maintain its structure. Once microbial cells are lysed or grazed in the ocean and their DNA is presumably released into the environment the folding proteins may still be bound to DNA – possibly impeding degradation. Both HU proteins and histones have been shown to protect DNA from DNase I digestion (Van Holde et al. 1980; Mukherjee et al. 2008). Histones and HU proteins have also been shown to protect DNA from environmental stressors such as - UV radiation, oxidative stress, and acid degradation (Ljungman & Hanawalt 1992; Enright et al. 1992; Olenick et al. 1994; Wang et al. 2012; Takata et al.
2013; Wang & Maier 2015; Almarza et al. 2015). Other modifications along with these may extend the residence of F-DNA, perhaps making F-DNA less bioavailable than previously thought.

Previously, the total D-DNA pool has been quantified by a variety of methods. Many of which involved precipitating DNA by ethanol (DeFlaun et al. 1986) or with CTAB (Karl & Bailiff 1989). When precipitated by these methods, DNA from viral particles can contribute to the D-DNA signal, causing difficulty in separately quantifying and characterizing D-DNA components. Brum et al. (2004) measured the different components of D-DNA by estimating the contribution of viral DNA to the total signal, and using DNase to estimate the free DNA (enzymatically hydrolysable) contributions. Paul et al. (1991) adapted a method using vortex flow filtration to concentrate large volumes of viruses and free DNA. From this concentrate they enumerated the amount of viruses and estimated the DNA contribution based on the average DNA per viral particle. However, no direct measurements of the various D-DNA pools have been made yet. By adopting methods used to concentrate viruses (Brum et al. 2013) and vesicles (Biller et al. 2014) involving tangential flow (or crossflow) ultrafiltration, the DNA concentrate can then be separated from viruses by density gradient ultracentrifugation (Lawrence & Steward 2010). The purified fractions can then be analyzed and characterized independently. As of yet there has been no study that has specifically separated and identified the F-DNA. As a common component of DOM,
accounting for an approximate 1-5% of the global oceanic dissolved nutrients (C, N, P), F-DNA may play an important role in cycling labile organic matter, transferring genetic information, and providing limiting nutrients to marine microorganisms.

1.5 Research objectives

The motivation for my dissertation research was to characterize F-DNA alongside the other D-DNA constituents, to better understand the microbial sources and dynamics. This research was conducted at Station ALOHA in the North Pacific Subtropical Gyre. In Chapter 2, I developed and applied a method that concentrates and separates vesicles, viruses, and F-DNA. This separation allows for the direct examination of all three known D-DNA constituents. Next, in Chapter 3, we sequenced F-DNA alongside the other D-DNA constituents to investigate the distinct microbial sources of D-DNA. These are the first reported F-DNA metagenomic sequences. From this analysis we also found genetic evidence of SAR11 vesicles. Finally, in Chapter 4, we investigate the dynamics producing F-DNA over the course of a mortality experiment. In this experiment we measured F-DNA production following Prochlorococcus mortality by viruses and grazers. This provided insight into the microbial dynamics surrounding F-DNA production and consumption by heterotrophic bacteria. Each of these components were collaborative efforts. Together, my
collaborators and I hope to have provided novel insights into the microbial sources and dynamics of F-DNA in the open ocean.
1.6 References


Grabowski, E., R. M. Letelier, E. A. Laws, and D. M. Karl. 2019. Coupling carbon and


**Figure 1.1.** Water masses present at Station ALOHA in the North Pacific Subtropical Gyre. (a) Temperature (°C) and salinity at Station ALOHA illustrates distinct water masses (data obtained via the Hawaii Ocean Time-series HOT-DOGS application; University of Hawai‘i at Mānoa. National Science Foundation Award # 1756517) (b) Table describing figure water masses (data from Lukas & Santiago-Mandujano 2008).

<table>
<thead>
<tr>
<th>Full name</th>
<th>Formation Site</th>
<th>Depth range (m)</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPTW North Pacific Tropical Water</td>
<td>Central subtropical gyre, sometimes at ALOHA during ENSO</td>
<td>100-140 m</td>
<td>24.3-24.7 σθ, 21-25°C, 34.9-35.4</td>
</tr>
<tr>
<td>ESMW Eastern North Pacific Subtropical Mode Water</td>
<td>25-30°N, 135-140°W</td>
<td>140-200 m</td>
<td>24-25.4 σθ, 12-21°C, 34.8-35.4</td>
</tr>
<tr>
<td>SSM Shallow Salinity Minima</td>
<td>35-50°N, 145-160°W</td>
<td>200-500 m</td>
<td>25.1-26.2 σθ, 11-17°C, 34.4-35.1</td>
</tr>
<tr>
<td>NPW North Pacific Intermediate Water</td>
<td>Western boundary of North Pacific</td>
<td>500-770 m</td>
<td>26.8 σθ, 5-7°C, 34.1-34.2</td>
</tr>
<tr>
<td>AAIW Antarctic Intermediate Water</td>
<td>Mid-latitude southwest Pacific</td>
<td>770-4000 m</td>
<td>27.2 σθ, &lt;2.5°C, 34.5</td>
</tr>
</tbody>
</table>
Figure 1.2. Karl & Church (2014) schematic illustrating the habitat characteristics of Station ALOHA.
Figure 1.3. Mende et al. (2019) figure representing the abundances of different microorganisms found in the water-column at Station ALOHA
Figure 1.4. Constituents of dissolved DNA (D-DNA). Dissolved DNA is operationally defined as DNA that flows through a filter (typically 0.1 or 0.2 µm). It may include DNA inside of encapsidated viruses, extracellular vesicles, and as truly “free” DNA (F-DNA).
Chapter 2: A method for characterizing dissolved DNA and its application to the North Pacific Subtropical Gyre

Morgan D. Linney

Coauthors: Christopher R. Schvarcz, Grieg F. Steward, Edward F. DeLong, David M. Karl
Affiliation: Daniel K. Inouye Center for Microbial Oceanography: Research and Education (C-MORE) and Department of Oceanography, University of Hawaii at Manoa, Honolulu, HI 96822

2.1 Abstract: Dissolved DNA (D-DNA) is a ubiquitous component of dissolved organic matter in aquatic systems. It is operationally defined as the DNA that passes a membrane filter and thus includes pools of truly dissolved “free” DNA (F-DNA), virion encapsidated DNA, DNA within membrane vesicles, and possibly other bound forms, each with different sources and lability. We investigated whether filtration (< 0.1 µm), concentration by tangential flow ultrafiltration (> 30 kDa), and fractionation in an equilibrium buoyant density gradient could be used to discriminate the mass contributions of the different pools of filterable DNA in seawater. Spike-in experiments with a known range of DNA standards (75–20,000 bp) indicated that this method results in high recoveries of F-DNA (68–86%) with minimal degradation. Application of the fractionation method to seawater samples collected from the oligotrophic North Pacific Ocean followed by analysis of fractions (epifluorescence and electron microscopy, DNase digestion) suggested that the low-density fractions (1.30–1.35 g mL⁻¹) were dominated by vesicle-like particles, mid-density fractions (1.45–1.55 g mL⁻¹) by virus-like particles, and high-density fractions (1.60–1.70 g mL⁻¹) by F-DNA. The estimated concentration of DNA that is either F-DNA, in viruses, or in vesicles was 0.13, 0.14, and 0.08 µg L⁻¹ in the euphotic zone and 0.09, 0.04, and 0.03 µg L⁻¹ in the mesopelagic zone. The approach described should be useful for more detailed investigations of the abundance, dynamics, and sources of DNA in the distinct pools that comprise filterable DNA in aquatic environments.
2.2 Introduction

Dissolved DNA (D-DNA)—a highly labile and nutrient-rich component of dissolved organic material (DOM)—is ubiquitous in all aquatic habitats investigated to date, including freshwater rivers and lakes (Pillai and Ganguly 1972), coastal marine and estuarine systems (DeFlaun et al. 1986), sediments (Dell’Anno et al. 2002), and the open ocean (Karl and Bailiff 1989; Brum 2005). Concentrations of D-DNA often represent a substantial fraction of the total DNA (DeFlaun et al. 1987; Karl and Bailiff 1989), especially in the open ocean. Spatial patterns of D-DNA concentrations in the ocean are similar to those of total microbial biomass and total DOM, decreasing with distance from shore and with depth (DeFlaun et al. 1987; Karl and Bailiff 1989; Weinbauer et al. 1993; Weinbauer et al. 1995). The concentration and bioavailability of D-DNA has both evolutionary and ecological significance. Double stranded D-DNA has the potential to alter the genetic make-up of microbial cells by the process of natural transformation (Hermansson and Linberg 1994) or, if hydrolyzed, can serve as a source of nitrogen and phosphorus (Jørgensen and Jacobsen 1996), macronutrients that are often limiting to microorganisms in open ocean ecosystems (Karl and Church 2014). As a consequence, there has been considerable interest in developing methods to investigate the nature and dynamics of this component of the larger DOM pool.

Methods to study D-DNA have usually employed 0.2 µm pore size filters to separate particulate from the dissolved fractions (DeFlaun et al. 1986; Karl and Bailiff
This operational definition means that D-DNA is a heterogeneous mixture of truly dissolved or “free” DNA (F-DNA), membrane-enclosed DNA in the form of filterable bacteria (Ghuneim et al. 2018) and extracellular vesicles (Biller et al. 2014), and encapsidated DNA in the form of the many small viruses that dominate aquatic ecosystems (Wommack and Colwell 2000). Depending on the relative contributions of the various D-DNA constituents and their size spectra, the total D-DNA measured could vary as a function of the filter pore size used to remove particulates. In prior investigations, estimates of D-DNA concentration were relatively insensitive to the choice of pre-filter. Negligible differences were seen, for example, between samples filtered through 0.1 µm vs. 0.2 µm membranes (DeFlaun et al. 1986) or among samples filtered through membranes or glass-fiber filters with pore sizes that ranged from 0.05 to 0.7 µm (Karl and Bailiff 1989). This suggests that bacteria capable of passing a 0.2 µm filter (e.g., MacDonell and Hood 1982) were not a major contributor to the D-DNA measurements. However, these comparisons were conducted using coastal seawater. The relative importance of 0.2 µm-filterable bacteria in lower productivity open ocean waters could be greater. To ensure the complete exclusion of small cells and the clean isolation of F-DNA, a 0.1 µm filter was adopted for this study. Other filter sizes may be considered in more productive ecosystems.

Previous studies have separated D-DNA by molecular weight (DeFlaun et al. 1987) or by using 0.05 µm pore size filters (Karl and Bailiff 1989), but they did not
isolate F-DNA or distinguish viral DNA from vesicle DNA. Brum (2005) found that different components of the D-DNA pool have varying turnover times in the surface ocean. Enzymatically hydrolyzable DNA (by 80 units mL\(^{-1}\) bovine pancreas DNase I) turned over within 0.1–6.2 hours while viral D-DNA took 9.6–24 hours, suggesting that these different components of the D-DNA pool may vary in their bioavailability, turnover, and source-sink pathways. A recent report of abundant, small (50-250 nm) DNA-containing membrane vesicles in the ocean (Biller et al. 2014) revealed the presence of yet another constituent of the D-DNA pool that is distinct in its source and composition, further demonstrating the importance in making separate quantification and turnover measurements of D-DNA components.

Herein, we describe the evaluation and application of a method to resolve, recover, quantify, and characterize three distinct, DNA-containing components in the < 0.1 µm fraction of seawater. To achieve this, D-DNA in 0.1 µm-filtered seawater was concentrated by tangential flow ultrafiltration and constituents were partitioned based on differences in their buoyant density using density gradient ultracentrifugation, a protocol commonly applied for isolating viruses and vesicles (Lawrence and Steward 2010; Brum et al. 2013a; Biller et al. 2014). DNA from individual density fractions was then quantified, as well as characterized using enzymatic and microscopic techniques to confirm the identity of each component. The molecular weight and quantity of F-DNA varied at different depths in the water column. The results suggest that the
amount of DNA inside of viral particles is comparable to that found in F-DNA for seawater from the North Pacific Subtropical Gyre (NPSG), which is similar to previous reports of D-DNA pools (Jiang and Paul 1995; Brum 2005).

2.3 Method

A method was applied to characterize three distinct constituents of D-DNA: viral DNA, vesicular DNA, and F-DNA. Past methods employed chemical agents to precipitate DNA from seawater, but these methods are expected to lyse viral particles (DeFlaun et al. 1986; Karl and Bailiff 1989) and vesicles, thus preventing confident discrimination among the D-DNA constituents (Table 2.1). The main steps of our method include (1) seawater sample collection and prefiltration (0.1 µm) to remove particulate DNA, (2) ultrafiltration (30 kDa nominal molecular weight limit) (NMWL) to concentrate D-DNA, (3) CsCl density gradient separation of different pools of D-DNA, and (4) characterization of isolated gradient fractions.

2.3.1 Collection and prefiltration

All samples for this research were collected from near Station ALOHA (A Long-Term Oligotrophic Habitat Assessment; 22° 45'N, 158° 00'W), an open-ocean research site in the NPSG located 100 km north of Oahu, Hawaii. Since 1988, the Hawaii Ocean Time-series (HOT) program has conducted near-monthly biogeochemical and hydrographical
measurements at Station ALOHA (Karl and Church 2014) and documented water mass properties (Figure 2.S1; Sabine et al. 1995). For this method, seawater samples were collected using a Niskin® bottle rosette equipped with a CTD and transferred to 20 L acid-washed polycarbonate carboys. Up to 50 L of seawater were collected at each sample location. Each sample was filtered through a double-layered 0.1 µm polyethersulfone (PES) capsule filter (Supor Acropak™ 1000; Pall) by gravity, directly from the Niskin® bottle. After collection and prefiltration, samples were immediately concentrated by ultrafiltration.

2.3.2 Ultrafiltration (tangential flow and/or centrifugal)

The filtrate was concentrated using a tangential flow ultrafiltration system (TFF; Millipore Pellicon 2; 0.1 m² Ultracei membrane, 30 kDa NMWL) to reduce the sample volume from 20–50 L to 100–250 mL, then further concentrated using a centrifugal ultrafilter (Centricon 70-Plus, 30 kDa NMWL; Millipore). This produced a final total sample volume of 1–2 mL. Centrifugal ultrafiltration alone may suffice for smaller volume samples collected from more productive regions.

2.3.3 Density gradient separation

For all samples, vesicles, viruses, and F-DNA were separated and purified based on differences in buoyant densities using a cesium chloride (CsCl) equilibrium buoyant
density gradient. The gradient was prepared by sequentially adding three density layers to the ultracentrifuge tube: 1.6, 1.45, and 1.2 g mL⁻¹ CsCl in SM buffer (100 mM sodium chloride, 8 mM magnesium sulfate, 50 mM Tris-HCl pH 7.5), having a homogenous CsCl density of about 1.40 g mL⁻¹. The concentrated sample was gravimetrically adjusted to a density of 1.45 g mL⁻¹ by adding concentrated CsCl solution (1.9 g mL⁻¹ in SM buffer) and was added as the second layer of the density gradient (Lawrence and Steward 2010). The samples were then spun in an ultracentrifuge (Beckman Coulter Optima XPN-80) in a swinging bucket rotor assembly (SW 41 Ti) for 72 hours at 4 °C at 30,000 rpm (154,000 x g). Samples were removed and separated into 0.5 mL fractions using a fraction collector (Auto Densi-Flow, Labconco). Marine viruses have previously been shown to migrate to CsCl densities between 1.45–1.55 g mL⁻¹ (Lawrence and Steward 2010), while DNA has been reported around 1.63–1.76 g mL⁻¹ (Wells and Larson 1972; Lueders et al. 2004). Marine microbial vesicles have been reported from 1.19–1.35 g mL⁻¹ in CsCl (Choi et al. 2015; Kwon et al. 2019), congruent with previous reports of membrane-enclosed DNA migrating around 1.30 g mL⁻¹ (Anderson et al. 1966). The density of each fraction was determined gravimetrically and the volume measured using a positive displacement pipet. The fractions corresponding to three distinct peaks in D-DNA were combined into three separate “pools” for subsequent analyses: (1) low-density DNA (1.30–1.35 g mL⁻¹), (2) medium-density DNA (1.40–1.55 g mL⁻¹), and (3) high-density DNA (1.60–1.70 g mL⁻¹).
CsCl was exchanged with TE buffer (10 mM Tris-HCl, 1 mM disodium EDTA, pH 7.5) for the low- and high-density pools and with SM buffer for the medium-density viral pool, by centrifugal ultrafiltration (Amicon Ultra-0.5 Centrifugal Filter Unit, 10 kDa NMWL; Millipore). To perform the buffer exchange, 0.5 mL fractions are added to the centrifugal device and concentrated to about 0.1 mL. This concentrate is then reconstituted to 0.5 mL in the appropriate buffer and concentrated again to 0.1 mL. This is performed a total of three times to rinse the fraction of remaining salts and effectively exchange the buffer.

2.3.4 Quantification of isolated DNA

DNA in each of the three pools was quantified using a Qubit 2.0 fluorometer following the Qubit dsDNA HS Assay (ThermoFisher Scientific). While the Qubit fluorometer and proprietary buffers and fluorochromes were used in this study, other similar reagents may be used. Prior to quantification, DNA from the vesicular (low-density) and viral (medium-density) pools were extracted following the whole community and viral nucleic acid extraction methods, respectively, as described by Mueller et al. (2014). Both extraction methods have three primary steps: (1) lysis and digestion by a lysis buffer with proteinase K, (2) salt-induced protein precipitation with ammonium acetate and, (3) nucleic acid precipitation and recovery. The whole community extraction (utilized here to extract nucleic acids from vesicles) employs a
lysozyme step prior to exposure to the lysis buffer and proteinase K, which is not required for the viral nucleic acid extraction. Pools may be further investigated by additional nucleic acid assays, fluorometric protein quantification, epifluorescence microscopy, fragment analysis, or transmission electron microscopy (TEM).

2.3.5 Fragment analysis

To assess the molecular weight of recovered F-DNA, concentrated samples were analyzed by capillary electrophoresis (Fragment Analyzer™ Automated CE System; Advanced Analytical Technologies, Incorporated) with a 33 cm capillary using the High Sensitivity Genomic DNA Analysis Kit. Samples were run following the manufacturer’s instructions (protocol DNF-488-33).

2.4 Assessment

The following experiments were conducted to evaluate the effectiveness of this method for isolating F-DNA and to characterize the distribution of different D-DNA components in the density gradient fractions. The samples subjected to these experiments were collected from the euphotic (5, 75, and 125 m), upper mesopelagic (150 and 250 m), and lower mesopelagic (350, 500, and 1000 m) zones of the water column. For all samples, DNA concentrations exhibited maxima within three distinct density ranges of the D-DNA cesium chloride gradient and were partitioned into three
pools: (1) low-density DNA, 1.30–1.35 g mL\(^{-1}\); (2) medium-density DNA, 1.40–1.55 g mL\(^{-1}\); and (3) high-density F-DNA, 1.60–1.70 g mL\(^{-1}\) (Figure 2.S2A–B). The amount of DNA in between these three pools (1.20 – 1.30, 1.35-1.40, and 1.55-1.60 g mL\(^{-1}\)) together accounted for <10% of the sum of all density fractions (total D-DNA). All homogenized samples representing each of these density ranges were analyzed by TEM (morphology), epifluorescence microscopy (virus-like particle enumeration), DNA quantification before and after treatment with DNase I, and protein quantification. D-DNA characterizations are reported for four separate depth profiles at Station ALOHA, collected in November 2017, April 2018, May 2018, and June 2019.

2.4.1 Transmission electron microscopy

Electron microscopy of negatively stained samples was routinely used to characterize the density gradient pools. A small subsample (4 µL) of concentrated, buffer-exchanged fractions was applied for 45 seconds to carbon-coated Formvar-coated copper grids (Electron Microscopy Sciences) that were rendered hydrophilic by glow discharge. Grids were then washed with 10 µL of ultrapure water (Barnstead, NanoPure), negatively stained with 2.5% uranyl acetate for 45 seconds, washed with 10 µL of ultrapure water again, and allowed to air dry. Material on the grids was examined by TEM (Hitachi HT7700) at 65,000–100,000x magnification at 100 kV. High-resolution images were captured with a 2048 x 2048 CCD camera (AMT XR-41B). For all three D-
DNA cesium chloride gradient pools (low-, medium-, and high-density), the first 50 randomly observed objects were used to assess the dominant structures in each sample.

The low-density DNA pools were dominated by filamentous (10 x 500–600 nm) and coccoid (30–100 nm diameter) structures (Figure 2.1). The medium-density pools contained a variety of virus-like morphotypes (tailed and non-tailed). The emphasis of observations was not on characterizing viral morphologies, but on examining the F-DNA (high-density) pool to determine whether viruses or vesicles were present. No viruses were observed in the 1.60–1.70 g mL\(^{-1}\) D-DNA pool. For all samples, this pool was dominated by narrow structures of approximately the same width (6–8 nm) but with varying length. These results suggest that the dominant structures in the low- and medium-density pools are likely encapsulated, whereas the high-density D-DNA pool is dominated by linear structures (Table 2.2).

2.4.2 Epifluorescence microscopy

After the density and nucleic acid content of each D-DNA pool was determined, DNA-containing virus-like particles (VLPs) were enumerated by epifluorescence microscopy (Suttle and Fuhrman 2010). A volume containing at least 50 ng of DNA from each pool was diluted in ultrapure water to 1 mL. After dilution, samples were gently filtered (≤13 kPa) onto a 0.02 μm Anodisc (Whatman) filter. Subsequently, the
filter was stained with 200-fold diluted PicoGreen reagent (ThermoFisher Scientific) which is highly specific to double-stranded DNA. Following staining, phenylenediamine was applied to a glass slide and the filter, to prevent fading. VLPs were enumerated at 1000x magnification using a Nikon Eclipse 90i microscope. No VLPs were observed in any of the high-density pools (1.60–1.70 g mL⁻¹). By comparison, VLPs were found in both the low-density and medium-density D-DNA pools (Table 2.2). These data support the TEM observations that the low- and medium-density pools may be dominated by encapsulated DNA particles (e.g., vesicles and/or viruses), whereas these are absent from the high-density pool.

2.4.3 DNase treatment

To test the effectiveness of DNase I in hydrolyzing the three D-DNA pools, subsamples (~50 ng DNA from each pool) were treated with 5 U mL⁻¹ DNase I (ThermoFisher Scientific from Escherichia coli cells, RNase-free, EN0521) for 30 minutes at 37 ºC. The reaction was terminated by the addition of 50 mM EDTA, followed by incubation at 65 ºC for 10 minutes. To assist in confirming which pools contained F-DNA, DNA concentrations were measured before and after DNase I treatment using a Qubit 2.0 fluorometer (Table 2.2). Buffer samples (negative controls) and triplicate controls of 50 ng dsDNA ladder (ThermoFisher Scientific GeneRuler 1kb DNA Ladder, SM0312) were also analyzed. In low- and medium-density pools the loss of DNA was
minimal following DNase I treatment (89–101% and 88–98% DNA remaining, respectively). In all F-DNA samples the fluorometrically measured DNA was reduced to 3–8%, suggesting that a major portion of this DNA pool is susceptible to DNase digestion and may not be protected by an organic (lipid or protein) or inorganic (e.g., clay particles) structure.

2.4.4 Protein quantification

Viral particles are often encapsulated by a protein coat, and the concentration of protein associated with this coat can be quantified using fluorescent dyes. To measure the concentration of protein and DNA in separated viral and F-DNA pools, fluorometric protein and dsDNA quantifications were performed (following the Qubit Protein and the Qubit dsDNA High Sensitivity Assays, respectively). The low- and medium-density pools for all depths had a much lower average DNA:protein ratio than that for the F-DNA pools (Table 2.2), indicating that the contents of the low- and medium-density pools are dominated by both protein and DNA structures, whereas the F-DNA pools are proportionately higher in DNA.

2.4.5 Method validation

To validate the method, we investigated the effect of prefiltration (double-layered 0.1 μm) on D-DNA concentrations, analyzed the recovery of DNA of varying molecular
weights, and conducted a field experiment to assess the concentration of each D-DNA pool with respect to depth.

2.4.5.1 Prefiltration

Since it is possible that F-DNA may be an artifact of microbial cell breakage during prefiltration, concentrations were quantified before and after the prefiltration of Prochlorococcus marinus strain MIT9301 cells (2 × 10⁸ cells mL⁻¹) through a double-layered 0.1 µm capsule filter. P. marinus is a small (0.5–0.7 µm) cyanobacterium that is ubiquitous in tropical and subtropical marine ecosystems (Chisholm et al. 1992; Partensky et al. 1999). These cells were carefully added to seawater collected from 1000 m at Station ALOHA to achieve a final concentration of 6 × 10⁵ cells mL⁻¹. Deep seawater was used in this experiment because it contains a minimal concentration of D-DNA. The concentration of added cells used in this experiment is comparable to total microbial surface concentrations at Station ALOHA and about two times higher than surface P. marinus cell counts, therefore exposing the capsule filter to a similar cell load. Prior to the addition of the P. marinus cells, the concentration of F-DNA in this 1000 m seawater was found to be 0.09 (±0.01) µg L⁻¹, using the previously described method. Cultured cells, containing approximately 1 µg L⁻¹ total DNA were carefully added to 10 L of seawater, then filtered at 50 mL min⁻¹ through the capsule filter. The filtrate was collected, concentrated by tangential flow ultrafiltration, and F-DNA was
subsequently measured using the method described herein. After the prefiltration of *P. marinus*, F-DNA was 0.08 (±0.02) µg L⁻¹, a value that was not significantly different from the initial F-DNA concentration. Viral and vesicle DNA were also measured but were undetectable before and after filtration. Therefore, the prefiltration step used to separate dissolved from particulate DNA did not result in the artifactual production of D-DNA, a finding in accordance with previous tests using natural communities prefiltered through 0.2 µm filters (DeFlaun et al. 1986, Beebee 1991). When sampling biomes that are known to have many delicate cell assemblages other concentration methods, such as gravity or reverse filtration may also be considered.

2.4.5.2 Internal standard curve and recovery

To assess the recovery of DNA with variable molecular weights using the method described herein, an internal standard curve (0, 0.2, 0.4, 0.6, 0.8, and 1 µg L⁻¹) was created in duplicate by the addition of varying amounts of 75–20,000 bp (≈4.8 × 10⁴ to 1.3 × 10⁷ g mol⁻¹) DNA ladder (GeneRuler 1 kb Plus DNA Ladder, ThermoFisher Scientific) to 5 L of 0.1 µm-filtered seawater from 1000 m at Station ALOHA, followed by processing with tangential flow ultrafiltration, centrifugal ultrafiltration, and density gradient separation. After collection of the high-density F-DNA pool (1.60–1.70 g mL⁻¹) from the CsCl density gradient, the sample was exchanged into TE buffer by centrifugal ultrafiltration. Absolute recovery was determined by measuring mass-
standardized fluorescence with a Qubit 2.0 fluorometer following the Qubit dsDNA HS Assay (ThermoFisher Scientific). The recovery of the ladder DNA from 30 kDa ultrafiltration and subsequent density gradient separation was 68–86% (Figure 2.2A). The size and relative recoveries of fragments of different sizes were determined from migration rate and fluorescence intensity of each band using capillary electrophoresis as described above. Because of possible non-linearity in the relationship between mass and fluorescence, the fragment-based recoveries (Figure 2.2B) can only be considered estimates of relative recovery. All molecular weights from the ladder were recovered with little to no shearing (Figure 2.2C). For comparison, three environmental F-DNA samples were measured using fragment analysis (Figure 2.3) and illustrate the molecular sizes of F-DNA that can be recovered from an open-ocean ecosystem.

2.4.5.3 Low-molecular weight filtrate DNA

Using a tangential flow filter with a 30 kDa nominal molecular weight cutoff may lead to the loss of DNA <50 bp. To assess whether low-molecular weight DNA was lost in the tangential flow ultrafiltrate (permeate), the permeate was processed using the cetyltrimethylammonium bromide (CTAB)-precipitation method of Karl and Bailiff (1989). Two-liter filtrate samples from four depths (5, 75, 125, and 500 m) at Station ALOHA collected on HOT-297 in November 2017, were investigated in triplicate. Following this procedure, 0.05 (±0.01), 0.12 (±0.03), and 0.14 (±0.02) µg DNA L⁻¹ were
recovered in the < 30 kDa permeate from 5, 75, and 125 m, respectively. These values account for 25, 38, and 45% of the F-DNA and 10, 17, and 27% of the total D-DNA collected from these three depths following the method described in this manuscript. The recovered DNA in the permeate from the 500 m sample was undetectable (<0.01 µg L\(^{-1}\)).

2.4.5.4 Field application

To verify the efficacy of this method, we collected samples at or near Station ALOHA on four separate cruises (HOT-297 November 2017, FK180310 April 2018, HOT-302 May 2018, HOT-312 June 2019) and processed the samples using the method described above. All samples were concentrated by tangential flow ultrafiltration (30 kDa NMWL) and separated by buoyant density centrifugation. Low-density encapsulated DNA, medium-density encapsulated DNA, and high-density F-DNA were found to contribute on average 15, 46, and 39% of the total D-DNA, respectively, with ranges of 8–43%, 20–64%, and 23–66% (Figure 2.4). Within these four depth profiles, medium-density viral assemblages and F-DNA are the primary contributors to total D-DNA concentrations (0.03–0.32 and 0.02–0.31 µg L\(^{-1}\), respectively). On one occasion in May 2018, the total D-DNA (0.73 µg L\(^{-1}\)) collected from 50 m was dominated by the low-density vesicle pool (0.33 µg L\(^{-1}\)). This was not observed for any other depth or profile, so is considered to be an outlier. F-DNA, the
medium-density viral pools, and total D-DNA peaked at 75 m (0.14–0.30, 0.24–0.31, 0.40–72 µg L\(^{-1}\), respectively) on two cruises when this depth was measured (November 2017 and June 2019).

2.5 Discussion

This method provides a means not only to discriminate among, but also to recover, distinct components of operationally defined D-DNA. The non-destructive concentration and recovery of these materials means that each pool can be further investigated to determine its distinctive characteristics. Downstream analysis of vesicular, viral, and F-DNA (e.g. nucleic acid base or sequence analyses) can provide insight into the microbial composition of D-DNA. Furthermore, biogeochemical analyses of low-molecular weight D-DNA utilizing CTAB (Karl and Bailiff 1989) can yield insights into the F-DNA size distribution and molecular weight, thereby providing estimates of nitrogen and phosphorus associated with this recovered DNA, and consequences for nutrient limitation in the open ocean. Few studies have specifically isolated and characterized F-DNA, so its sources and physicochemical properties (e.g., stoichiometry, possible modifications, taxonomic affiliations, etc.) are largely unknown. Beebee (1991) used ultracentrifugation to separate viruses and soluble DNA (F-DNA). The pellet was examined by TEM, and was found to be enriched in virus-like particles. The soluble DNA in the supernatant was purified by chromatography, and the
molecular weight was measured using gel electrophoresis. They reported that the pelleted “viral” pool was the primary contributor to D-DNA collected from different aquatic environments (78% of D-DNA in pondwater and 85% in seawater). Jiang and Paul (1995) utilized a similar method of ultracentrifugation to separate the bound and soluble DNA pools, and subsequently characterized the pools using hybridization with rRNA gene-targeted probes. They found that soluble DNA (F-DNA) comprised half of the D-DNA, while viral DNA and “other” bound DNA contributed the other half. While these methods provide important datasets, differential centrifugation offers a less complete separation of D-DNA components than density gradient ultracentrifugation. Additionally, the current method provides more complete recovery of F-DNA (68-86%) than Beebee (1991; 36%) and Jiang and Paul (1995; 65%). As a common component of DOM, F-DNA may play important roles in cycling labile substrates, transferring genetic information, and providing limiting nutrients to marine microorganisms.

Previous studies have identified a fraction of D-DNA that is resistant to DNase enzymes (Jiang and Paul 1995; Brum 2005). In this study, treatment with DNase was used to confirm the identity of the F-DNA constituent collected from the CsCl density gradient. The small residual after digestion (3–8%) suggests that only a minor fraction of the F-DNA pool is DNase-resistant and that, enigmatically, most of the F-DNA is in a form that should be very readily degraded or taken up by microorganisms. The
dominance of this labile form of DNA must therefore be sustained by continuous production.

The DNase-resistant portion of the F-DNA may represent DNA that is modified (e.g. phosphorothioation or methylation) or associated with proteins (e.g., histones or other DNA binding proteins). One source of phosphorothioated DNA, which is resistant to many nuclease activities, is produced by *Pelagibacter ubique* (Eckstein 2000; Wang et al. 2007), one of the most abundant microorganisms in open-ocean systems (Morris et al. 2002). Protein binding may also confer DNase resistance. Both prokaryotic and eukaryotic cells contain genome packaging proteins which help organize and compress their genetic material. Once microbial cells are lysed or grazed and their DNA is released into the environment, these proteins may still be bound to DNA, possibly impeding degradation. Both bacterial binding proteins and histones have been shown to protect DNA from DNase I digestion (Van Holde et al. 1980; Mukherjee et al. 2008). Histones and bacterial binding proteins have also been shown to protect DNA from environmental stressors such as ultraviolet radiation, oxidative stress, and degradation by acid (Enright et al. 1992; Ljungman and Hanawalt 1992; Oleinick et al. 1994; Wang et al. 2012; Takata et al. 2013; Almarza et al. 2014; Wang and Maier 2015). These modifications and others may extend the residence times of DNase-resistant F-DNA, altering its dynamics and bioavailability. However, the low
concentrations of the DNase-resistant F-DNA suggest that even this material is turning over relatively rapidly in the ecosystem.

The distribution of dissolved DNA components shown here for Station ALOHA follows a pattern consistent with DOM, in which concentrations are higher at the surface and decrease with depth. The 75 m peaks observed in November 2017 and June 2019 of F-DNA, the medium-density viral pools, and total D-DNA correspond to maximum abundances in Prochlorococcus (van den Engh et al. 2017), viral abundances (Brum 2005), and cell-associated cyanophage (DeLong et al. 2006). The cause of this 75 m maximum is still an area of active research. In all 1000 m samples, D-DNA components were significantly lower than at the surface, with F-DNA dominating. The dynamic concentration ranges of each of these dissolved DNA components suggest there may be different zones of net production and net utilization with depth.

The depth distribution patterns of D-DNA observed in this study are similar to those observed previously in the North Pacific Subtropical Gyre (Karl and Bailiff 1989; Brum 2005). However, absolute concentrations observed during this current study are approximately 50% lower. This may be due, in part, to the efficiency of recovery of our method, which we estimate ranged from 68–86%. One known source of loss in this method was the passage of low-molecular-weight DNA through the ultrafilter during concentration, which accounted for 10-27% of total D-DNA. Higher concentrations reported in previous studies could also have resulted from the use of larger pore size
prefilters (0.2 µm), which may inadequately remove some of the smallest and most abundant prokaryotes (e.g., *Pelagibacter ubique*). The use of double-layered 0.1 µm polyethersulfone capsule filters in this study would have more effectively removed the cellular fraction, minimizing contributions of living cells to the D-DNA. While this would also exclude viruses > 0.1 µm in size, their relative abundance in most oceanic regions appears to be low (Brum et al. 2013b). The concentrations quantified in this study are comparable to surface particulate DNA concentrations recovered from glass fiber or 0.2 µm polycarbonate filters: 0.75–2.38 µg L⁻¹ in the equatorial Pacific Ocean (Winn and Karl 1986) and 0.8–5.9 µg L⁻¹ in the Eastern Mediterranean Sea (Dell’Anno et al. 1999).

**2.6 Comments and recommendations:** The method described in this paper should be applicable for aquatic field research in many different environments, both marine and freshwater systems. Total DNA mass required for downstream analysis (e.g., sequence analysis or further physical or chemical characterization) must be considered before determining the volume to be collected and concentrated. In water with higher biomass, one may consider reducing the volume concentrated and possibly omitting the tangential flow ultrafiltration step.
2.7 References


Mueller, J. A., A. I. Culley, and G. F. Steward. 2014. Variables influencing extraction of nucleic acids from microbial plankton (viruses, bacteria, and protists) collected on


2.8 Acknowledgments: We thank the Hawaii Ocean Time-series team, captain and crew of the R/V Kilo Moana, R/V Ka‘imikai-O-Kanaloa, and R/V Falkor for their help at sea. We thank T. M. Weatherby at the Biological Electron Microscope Facility at the University of Hawaii, Mānoa for her assistance capturing images using the transmission electron microscope. We thank A. Burger for performing the Fragment Analyzer assays. This work was supported by the Simons Collaboration of Ocean Processes and Ecology (Award ID 329108 to DMK and EFD).
**Table 2.1 Comparison of methods to extract dissolved DNA (D-DNA) from seawater**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Procedure</th>
<th>Advantages</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pillai and Ganguly 1972</td>
<td>Filtration, barium sulfate precipitation, centrifugation, sodium chloride elution, hydrolysis by hydrochloric acid, UV nucleotide quantification</td>
<td>• DNA and RNA quantification</td>
<td>• Acid and base additions damage nucleic acids and prevent downstream analyses (e.g., microscopy, molecular weight estimates, and sequence analysis)</td>
</tr>
<tr>
<td>DeFlaun et al. 1986</td>
<td>Filtration, ethanol precipitation, centrifugation, dialysis,</td>
<td>• Collects a range of low and high molecular weight DNA</td>
<td>• Does not isolate D-DNA constituents efficiently because ethanol can lyse viral particles contributing to fluorescent quantification of F-DNA.</td>
</tr>
<tr>
<td>Karl and Bailiff 1989</td>
<td>Filtration, dilution of seawater 1:1 with EDTA, CTAB precipitation, filtration</td>
<td>• Precipitates DNA and RNA • Simple and reproducible • Recovery of low molecular weight DNA</td>
<td>• Does not separate D-DNA constituents efficiently because EDTA can lyse viral particles, overestimating F-DNA • Difficult to remove CTAB-DNA precipitate from GFF for downstream analyses</td>
</tr>
<tr>
<td>Beebee 1991</td>
<td>Filtration, ultracentrifugation to purify viruses and chromatography to purify free DNA, observation of pelleted viruses by transmission electron microscopy</td>
<td>• Simple method to separate D-DNA pools without any chemical additions or numerous steps</td>
<td>• Free DNA can pellet and skew the value of viral DNA • Vesicles and small cells are likely to contribute to the pelleted DNA</td>
</tr>
<tr>
<td>Paul et al. 1991</td>
<td>Filtration, large volume concentration by vortex flow filtration, enumeration of viruses and estimation of viral DNA contribution, estimation of F-DNA by DNase hydrolysis</td>
<td>• Allows for the collection and concentration of D-DNA from large volumes</td>
<td>• Indirect estimate of D-DNA components (viral DNA and F-DNA) • No separation of D-DNA components for downstream analyses</td>
</tr>
<tr>
<td>Brum et al. 2004</td>
<td>Centrifugal ultrafiltration, DNase hydrolysis to estimate F-DNA contribution</td>
<td>• Simple and reproducible • Only requires small volumes to quantify D-DNA components</td>
<td>• In oligotrophic systems, small volumes do not produce enough DNA mass for downstream experiments</td>
</tr>
<tr>
<td>Study</td>
<td>Methods</td>
<td>Advantages</td>
<td>Disadvantages</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Brum et al. 2013a            | Filtration, tangential flow ultrafiltration, centrifugal ultrafiltration, density gradient ultracentrifugation | • Suitable for investigating morphology and genomic data of marine viruses | • F-DNA was not the focus of this study but could likely be isolated using the method described  
|                              |                                                                         |                                                                             | • In oligotrophic systems, large volumes required  
|                              |                                                                         |                                                                             | • Time-consuming                                                               |
| Biller et al. 2014           | Filtration of large volumes >50–100 L, tangential flow ultrafiltration, ultracentrifugation, Optiprep density gradient | • Suitable for collecting vesicles  
|                              |                                                                         | • Optiprep used for collecting membrane-bound vesicles | • In oligotrophic systems, large volumes required |
| Present study                | Filtration of large volumes, tangential flow ultrafiltration (TFF), centrifugal ultrafiltration, density gradient ultracentrifugation | • Directly separates D-DNA constituents (e.g., vesicles, viruses, and F-DNA)  
|                              |                                                                         | • Recovers 75–2000 bp dsDNA using 30 kDa filter  
|                              |                                                                         | • All D-DNA constituents suitable for downstream analyses (e.g., microscopy, molecular weight estimates, and sequence analysis)  
|                              |                                                                         | • When low masses are sufficient or in eutrophic ecosystems, TFF can be eliminated | • Low molecular weight DNA (<75 bp) may be lost, depending on the nominal molecular weight limit of the ultrafilters used  
|                              |                                                                         |                                                                             | • In oligotrophic systems, large volumes required |

- **In less oligotrophic systems**, can utilize large centrifugal concentration units (e.g., Centricon 70-Plus) to acquire more DNA mass.
- Indirectly quantifies D-DNA components.
- Does not determine molecular weight of DNA.

**Filtration, tangential flow ultrafiltration, ultracentrifugation, density gradient ultracentrifugation**

- Suitable for collecting vesicles
- Optiprep used for collecting membrane-bound vesicles
- In oligotrophic systems, large volumes required
Table 2.2 Microscopy and molecular characterizations of the CsCl gradient pools

<table>
<thead>
<tr>
<th>Pool</th>
<th>Dominant contents (examined by TEM)</th>
<th>Epifluorescence VLP detection</th>
<th>% DNA remaining post-DNase Treatment</th>
<th>DNA:protein range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-Density (1.30–1.35 g mL⁻¹)</td>
<td>Filamentous structures, round viruses or vesicles</td>
<td>VLP detected</td>
<td>89–100</td>
<td>1.2–2.3</td>
</tr>
<tr>
<td>Medium-Density (1.40–1.55 g mL⁻¹)</td>
<td>Viruses (various morphologies)</td>
<td>VLP detected</td>
<td>88–98</td>
<td>4.3–6.8</td>
</tr>
<tr>
<td>High-Density (1.60–1.70 g mL⁻¹)</td>
<td>Linear structures (DNA)</td>
<td>VLP absent</td>
<td>3–8</td>
<td>30.6–42.1</td>
</tr>
</tbody>
</table>
Figure 2.1 Transmission electron micrographs of the dominant structures observed in D-DNA pools collected from the low-density (vesicles), medium-density (viruses), and high-density (F-DNA) fractions, which correspond to distinct D-DNA peaks that were consistently observed in cesium chloride gradients.
Figure 2.2 Molecular size-dependent recovery of ladder and environmental F-DNA. (A) Average percent recovery for total ladder DNA, measured fluorometrically. (B) Recovery of ladder DNA (75, 200, 300, 400, 700, 1000, 2000, 3000, 4000, 7000, 10000, 20000 bp) using the method described herein. Five concentrations of ladder DNA standards (0.2, 0.4, 0.6, 0.8, and 1 µg L⁻¹) were added to separate samples of 0.1 µm-filtered seawater collected from 1000 m at Station ALOHA and processed with this method using 30 kDa ultrafiltration. Capillary electrophoresis results from the fragment analyzer reveal little to no degradation of the recovered DNA ladder standards. LM (lower marker) is the low-molecular weight standard that is added to calibrate and align all samples. (C) Percent recovery of the same ladder DNA, the resulting recoveries were calculated based on uncalibrated fluorescence intensity and are thus intended to indicate only relative differences in recovery among bands. Bands at 500, 1500, and 5000 were not included in the analysis because they contained higher mass than the other bands and relative recoveries would be confounded by non-linearity in the fluorescence-mass relationship.
Figure 2.3 Molecular weight distribution for environmental F-DNA recovered from three depths at Station ALOHA (75, 125, and 500 m). Relative fluorescence units are plotted as a function of DNA fragment sizes as determined by capillary electrophoresis from 10 bp to the limit of resolution at approximately 40,000 bp.
Figure 2.4 Depth profiles of various pools of dissolved DNA constituents collected from the North Pacific Subtropical Gyre. “F-DNA” represents the high-density free DNA pool (blue), “LDV” represents the low-density pool containing vesicular and viral DNA (red), “MDV” represents the medium-density DNA pool dominated by viruses (yellow), and Total D-DNA is the summation of these three pools (black).
Figure 2.5.1 Habitat characteristics of Station ALOHA during sample collection. Depth profile shows average of from the same four expeditions (November 2017, April 2018, May 2018, June 2019) dissolved DNA samples were collected from for oxygen (yellow; µmol/kg), temperature (green; °C), and nitrate + nitrite (blue; µmol/kg). Error bars represent standard deviation of four the four cruises.
Figure 2.S2 Typical examples of D-DNA distributions measured across 0.5 mL fractions of cesium chloride density gradients. Samples represent environmental F-DNA recovered from Station ALOHA seawater collected in the euphotic (50 m; A) and mesopelagic (250 m; B) zones, following concentration by ultrafiltration and separation by equilibrium buoyant density gradient ultracentrifugation, as described in the main text. The low-density, medium density, and high-density D-DNA fractions are pooled conservatively at 1.30–1.35, 1.40–1.55, and 1.60–1.70 g mL⁻¹, respectively. The remaining D-DNA typically contributes <10% of the total D-DNA signal.
Chapter 3: Microbial sources of exocellular DNA in the ocean

Morgan D. Linney

Coauthors: John M. Eppley, Anna E. Romano, Elaine Luo, Edward F. DeLong, David M. Karl
Affiliation: Daniel K. Inouye Center for Microbial Oceanography: Research and Education (C-MORE) and Department of Oceanography, University of Hawaii at Manoa, Honolulu, HI 96822
Data accessibility: Genomic data and raw reads are available under the NCBI BioProject accession number PRJNA727670
3.1 Abstract: Dissolved DNA (D-DNA) is operationally defined as DNA that passes through a membrane filter (0.1 µm). It is composed of vesicles, viruses, and exocellular free DNA (F-DNA) and is ubiquitous in all aquatic systems, although the sources, sinks, and ecological consequences are largely unknown. Using a method that provides separation of vesicles, viruses, and F-DNA, we compared open ocean depth profiles of DNA associated with each fraction. Pelagibacter-like DNA dominated the vesicle fractions for all samples examined over a depth range of 75 - 500 m. The viral DNA fraction was composed predominantly of myovirus-like and podovirus-like DNA, and contained the highest proportion of unannotated sequences. Euphotic zone F-DNA (75-125 m) contained mostly bacterial and viral sequences, with bacteria dominating in the mesopelagic (500-1000 m). A high proportion of mesopelagic (500 and 1000 m) F-DNA sequences appeared to originate from surface waters, including a large amount of DNA contributed by high-light Prochlorococcus species. Throughout the water-column, but especially in the mesopelagic, the composition of F-DNA sequences was not always reflective of co-occurring microbial communities that inhabit the same sampling depth. These results reveal the composition of F-DNA in different regions of the water-column (euphotic and mesopelagic zones), with implications for dissolved organic matter cycling and export (by way of sinking particles and/or migratory zooplankton) as a delivery mechanism.
3.2 Introduction

Nearly half of the DNA in the open ocean is present outside of living organisms (Holm-Hansen et al. 1968; Winn & Karl 1986; DeFlaun et al. 1987; Dell’Anno et al. 1999). It is contained inside protein-encapsidated viruses, extracellular vesicles, and as the naked molecule itself, free DNA (F-DNA). Together this DNA that passes a membrane filter (0.1 or 0.2 µm) is dissolved (D-DNA), the cycling of labile DNA in the ocean has consequences for global nutrient cycling, export, and potential for genetic exchange. DNA in the ocean has been studied from multiple oceanographic perspectives (e.g., studying environmental DNA for fisheries management, sequencing cellular DNA to identify taxonomic relationships and quantifying population diversity, quantifying viral diversity and dynamics), making the overlapping terminology challenging to disentangle. While – environmental DNA (eDNA) research may have similar objectives of holistically studying an ecosystem - by collecting the remnants of large macro-organisms (e.g., scales, mucus, and reproductive cells) to uncover which organisms are present in a given study site (Djurhuus et al. 2020) - it employs the same collection method as cellular analyses (Mende et al. 2019). DNA such as this that is captured on a filter has been termed “particulate” DNA (P-DNA; Winn & Karl 1986), distinguishing it from DNA that passes a filter (D-DNA; Karl & Bailiff 1989). D-DNA can be thought of as predominantly “microbial eDNA,” the manifestation of microbial dynamics (e.g., viral lysis, autolysis, and grazing) and activity in the ocean.
Subtropical gyres comprise 40% of the earth’s surface, yet surface waters are deplete of essential elements like nitrogen and phosphorus (8), highlighting D-DNA (with a C:N:P ratio of 10:4:1) as a nutrient-rich molecule in this system. It can be taken up as polydeoxyribonucleotides (>10 kb) (Jeffrey et al. 1990; Paul et al. 1991; Frischer et al. 1994), as well as hydrolyzed and broken down to nucleotides and nucleic acid bases (Turk et al. 1992; Jørgensen & Jacobsen 1996; Lennon 2007). Alternatively, the genomic information encoded by D-DNA has the potential to be integrated by natural transformation (Hermansson & Linberg 1994).

Streamlined genomes are common in oligotrophic systems (Swan et al. 2013) and are found in many of the numerically dominant microorganisms (e.g., photoautotrophic Prochlorococcus, heterotrophic Pelagibacter, UCYN-A, as well as others) inhabiting the North Pacific Subtropical Gyre (Giovannoni et al. 2005; Tripp et al. 2010; Swan et al. 2013; Biller et al. 2014b). While these microorganisms have an advantage in a relatively stable and nutrient-deplete environment (Braakman et al. 2017), they are more vulnerable to extinction in the face of environmental changes than organisms with larger genomes (Bentkowski et al. 2015). Therefore, genetic exchange is essential to the survival of microorganisms like Prochlorococcus and Pelagibacter, evident by variable genomic islands and large pangenomes that are thought to have emerged by horizontal gene transfer (HGT; Coleman et al. 2006; Grote et al. 2012). However, the common mechanisms (e.g., conjugative systems,
plasmids, or transposons) of HGT have yet to be confirmed in many of these organisms. In oligotrophic environments, D-DNA may be a source of nutrients or an agent of intracellular genetic exchange.

One of the largest ecosystems on earth, the North Pacific Subtropical Gyre (15°N to 35°N latitude and 135°E to 135°W) is home to one of the most well-studied water columns (Karl & Church 2014) and exhibits relatively low seasonality (Church et al. 2013). Over thirty years of biogeochemical research provides a wealth of baseline datasets: hydrographic observations (and its implications for thermohaline circulation), metabolic balances, nutrient cycling, cultured microbial isolates, complex microbial interactions, and export dynamics. More recently, gene catalogs of both cellular and viral communities throughout the water column at Station ALOHA have been reported (Mende et al. 2017), broadening understanding of the distributional patterns of diverse microbial and viral communities in the open ocean.

While vesicles (Biller et al. 2014a) and viruses, two of the three D-DNA constituents, have been previously examined in open ocean systems, the biological composition and depth origins of F-DNA is relatively uncharacterized. Primarily due to prior D-DNA isolation methods that utilized low-volume filtration (Brum et al. 2004) and chemical precipitation (DeFlaun et al. 1986; Karl & Bailiff 1989) to isolate bulk D-DNA, these methods pose challenges in acquiring enough DNA (>0.1 µg) to assemble metagenomic libraries for sequencing, leaving at least 25% of the open ocean DNA
inventory uncharacterized (Winn & Karl 1986; Brum 2005; Linney et al. 2021). The average proportion of D-DNA that is either F-DNA, in viruses, or in vesicles is, 45, 35, and 20%, respectively (Linney et al. 2021).

To investigate all three D-DNA pools, and their microbial sources throughout the water column, we analyzed all three D-DNA pools throughout the water column of the North Pacific Subtropical Gyre. Employing a method (Linney et al. 2021) that separates the three pools by density, DNA from vesicle, virus, and F-DNA fractions was sequenced and compared. This study provides the first genetic characterization of the F-DNA pool, alongside the other D-DNA pools, and reveals viral lysis as a plausible source of F-DNA, and export as a delivery mechanism of DNA to mesopelagic depths.

3.3 Methods

3.3.1 Sample collection

Seawater samples were collected during three cruises near Station ALOHA (22°45′N, 158°00′W), R/V Kilo Moana cruise HOT297 (November 2017), R/V Falkor cruise FK180310 (April 2018), and R/V Ka’imikai-O-Kanaloa cruise HOT302 (May 2018) using standard Niskin-type bottles. Dissolved DNA samples (D-DNA; vesicles, viruses, and free DNA) were collected and separated following Linney et al. (2021). This method includes three primary steps: (1) prefiltration (~20 L) directly from the Niskin-
type bottles through a double-layered 0.1 µm polyethersulfone (PES) capsule filter (Pall Acropak 1000 cm²), (2) concentration by tangential flow ultrafiltration (Millipore Pellicon 2; 0.1 m² Ultracel membrane, 30 kDa nominal molecular weight limit; NMWL) down to ~1 mL, (3) D-DNA separation by density gradient ultracentrifugation in cesium chloride. Following separation, D-DNA constituents are buffer exchanged with either TE (vesicles and F-DNA: 10 mM Tris-HCl, 1 mM disodium EDTA, pH 7.5) or SM (viruses: 100 mM sodium chloride, 8 mM magnesium sulfate, 50 mM Tris-HCl pH 7.5) buffers. This method ensures that all D-DNA constituents are collected from the same sample, enabling comparison.

3.3.2 Molecular weight determination of environmental F-DNA

The size distributions of isolated F-DNA was analyzed by capillary electrophoresis (Fragment Analyzer™ Automated CE System; Advanced Analytical Technologies, Incorporated) with a 33 cm capillary using the High Sensitivity Genomic DNA Analysis Kit. Samples were run following the manufacturer’s instructions (protocol DNF-488-33). Digital sample peaks (electropherograms) were generated by ProSize 2.0 software.

3.3.3 DNA extraction and purification
DNA was extracted from viruses, vesicles, and exocellular “free” DNA (100 µL each sample) by first lysing and digesting membrane in sucrose lysis buffer (final concentrations: 40 mM EDTA, 50 mM Tris (pH 8.3) and 0.75 M sucrose, and lysozyme (0.5 mg ml\(^{-1}\) final concentration), at 37°C for 30 min. Proteinase K (0.8 mg ml\(^{-1}\) final concentration) and SDS (0.8% final concentration) were added, and the lysate was further incubated at 55°C for 2 hours. The DNA was purified using a Chemagen MSM I instrument with the Saliva DNA CMG-1037 kit (Perkin Elmer, Waltham, MA). DNA and sequencing library sample quantity and quality were assessed by PicoGreen dsDNA quantitation (Invitrogen, Waltham MA), and by capillary electrophoresis using a Fragment Analyzer™ Automated CE System, protocol DNF-488-33.

3.3.4 Metagenomic library preparation

D-DNA (vesicles, viruses, or free DNA) metagenomic libraries were prepared using a semi-automated EPmotion instrument with TruSeq Nano DNA library preparation kit (15041110), DNA input per sample 2 ng/µL sheared to an average size of 350 bp utilizing a Covaris M220 focused ultrasonicator following manufacturer’s recommendations with modifications to shear time to target 350 bp, and Microtube-50 AFA fiber tubes. (For some selected samples that had bimodal DNA size distributions two sequencing libraries were prepared, one with DNA shearing, and one without DNA shearing). Metagenome sequencing libraries were prepared using Illumina’s TruSeq
Nano LT library preparation kit, and sequenced on an Illumina Nextseq500 system, using V2 high output 300 cycle reagent kit, with addition of 1% of a PhiX control.

3.3.5 DNA sequence analysis and annotation

Sequenced D-DNA reads were filtered and trimmed for quality using the iu-filter-quality-minoche tool (Minoche et al. 2011) from illumina-utils (Eren et al. 2013). Quality filtered reads were identified and counted by mapping against the ALOHA 2.0 metagenomic reference gene catalog (Mende et al. 2017, Luo et al. 2020) using lastal (Kielbasa et al. 2011). The output of read mapping against the ALOHA 2.0 gene catalog generated a gene count table for each D-DNA dataset. Gene counts in all samples were normalized by the total number of mapped genes. The family-level taxonomic abundance was measured by the number of gene counts assigned to each microbial family. The relative abundances of quality controlled D-DNA gene counts were used to assess variation among D-DNA sample types (vesicle, virus, and F-DNA) and across the water column (75-1000 m). The most abundant ("major") bacterial and viral families were determined by a contribution of >0.5% (cutoff) to all D-DNA samples, and were compiled into proportional gene count tables. These assessments revealed the microbial sources of the D-DNA samples, as well as the known viral hosts and Prochlorococcus ecotype proportion.
To estimate the proportion of D-DNA sequences from each metagenomic sample that originated from a particular water column depth (0-4000 m), we compared the normalized coverage the ALOHA 2.0 catalog (Mende et al. 2017, Luo et al. 2020) had at each depth in the ALOHA 2.0 survey to the normalized coverage of each gene in our samples. First we calculated the probability that a sequence matching a given gene originated at a given depth as the ratio of the coverage of that gene at that depth to the total coverage of the gene. Next, for all depths, we multiplied that probability by the normalized coverage for each gene in our metagenomic D-DNA samples to get the portion of that coverage that likely originated at each depth. The coverages, now portioned by probable depth origin, were then aggregated by domain (Bacteria, Archaea, Eukaryote, Viral, or unknown) to better understand potential source patterns.

3.3.6 Comparison to cellular and viral datasets collected from Station ALOHA

All D-DNA gene count tables (vesicles, viruses, and free DNA) were combined and mapped to previously reported cellular (Mende et al. 2017) and viral (Luo et al. 2020) ALOHA 2.0 metagenomic reference gene catalogs collected on multiple Hawaii Ocean Time-series (HOT) research cruises, from the same depths at Station ALOHA (75, 100, 125, 250, 500, 1000 m). Statistical analyses were conducted to compare D-DNA with cellular and viral HOT samples. Read counts per taxonomic clade were
normalized by calculating their proportions relative to the total number of mapped
gene counts per sample. Normalized values were compiled in a proportional order-
level count table and square root transformed. Two-dimensional ordination methods
were used on normalized count tables to compare D-DNA datasets to cellular and viral
samples from the same depths. To visualize distances, non-metric multidimensional
scaling (NMDS) plots were generated using the metaMDS function and Bray-Curtis
distance matrices constructed from normalized gene counts in the VEGAN R package.

3.4 Results

3.4.1 Microbial composition of exocellular free DNA and other D-DNA pools

Summed across six F-DNA samples collected in the North Pacific Subtropical
Gyre, the majority of annotated genes in DNA-based metagenomic libraries were
derived from bacteria and viruses (Figure 3.1), with minimal contributions from Eukarya
and Archaea (<1.0 and 1.1%, of all annotated sequences, respectively). The proportion
of F-DNA virus sequences in surface waters ranged 10-35% above or at the deep
chlorophyll maximum (75-125 m), about the same as the proportion of Bacteria (16-
34%) over the same depth range. However, in the mesopelagic zone (250-1000 m), the
proportion of F-DNA sequences from viruses was lower (2-16%) , whereas the
proportion of bacterial-derived sequences was higher 38-45%. Across all F-DNA
metagenomic libraries, the microbial assemblage was dominated (69-85% of family-
level annotated sequences) by three main taxonomic groups: Pelagibacteraceae, Myoviridae, and Prochlorococcus (Figure 3.1); all of which have been documented as abundant microorganisms in the North Pacific Subtropical Gyre (Mende et al. 2019; Brum et al. 2013; Luo et al. 2020). Ubiquitous heterotrophs of the family Pelagibacteraceae contributed an average of 23% of F-DNA sequences over all depths sampled, with a range of 2-46%. Prochlorococcus represented 43, 42, and 53% of F-DNA sequences collected from 125, 500, and 1000 m, respectively, and represented an average 23% of family-annotated sequences. Of these Prochlorococcus sequences (125, 500, and 1000 m) more than 85% were from the high-light (HL) ecotypes (Table 3.S2). The highest proportion of Myoviridae F-DNA sequences was observed in the upper euphotic zone, 45% (75 m) and 69% (100 m). Of the viral-derived sequences constituting the F-DNA samples, the majority (>90% of annotated viral sequences with known hosts) were dominated by viruses known to infect Synechococcus (38% ±4.6), Prochlorococcus (31% ±4.8), Pelagibacter (16% ±7.2), and other Cyanobacteria (7.9% ±1.3) (Figure 3.S1). Synechococcus viruses peaked at the deep chlorophyll maximum on two occasions (43%). Other viruses known to infect SAR116 (5.4% ±2.2) and Vibrio (1.1% ±0.6), were detected at much lower proportions in the F-DNA. F-DNA-derived Vibrio sequences were highest in mesopelagic samples (250; 2.3%, 500; 1.0%, 1000; 1.4%), but were only <1% in euphotic samples.
Among other taxonomic groups contributing to metagenomic F-DNA libraries, Archaeal contribution was minimal (ranged 0.18-2.03% of annotated sequences for all samples). Of those, the most were most highly similar to those of ammonia-oxidizing Thaumarchaeota (0.02-1.6%), which were most prevalent at depths greater than 100 m. Of the three domains contributing to F-DNA metagenomes, Eukarya was the lowest (0.28-1.45%). The taxonomic families that were most abundant included heterokont Pelagomonadaceae, coccolithophore Noelaerhabdaceae, and Bathycoccaceae.

DNA from the vesicle D-DNA samples collected throughout the euphotic and mesopelagic zones (75-500 m; Figure 3.1) was also sequenced and compared. Summed across all samples, Bacteria contributed overwhelmingly to these metagenomic libraries (77% ±22 all annotated sequences). Viruses, Archaea, and Eukaryota contributed an average of 18.4, 3.7, 0.6%, respectively to all annotated sequences (Table 3.S1). Of the bacterial sequences, at the family-level sequences were heavily dominated by Pelagibacteraceae (81% ±8 of family-level annotated sequences), with only 34% ±12 of sequences left unannotated, the lowest of all three D-DNA fractions. Other bacterial sequences that contributed to the vesicle samples, include those from Rhodospirillaceae (2.9%), Rhodobacteraceae (2.7%), Flavobacteriaceae (1.0%), and Prochlorococcus (0.5%). Viral sequences in the vesicle metagenomic DNA libraries were highest in the upper euphotic zone and were dominated by both Podoviruses (7%) and Myoviruses (6%).
Averaged across all virus fraction metagenomic libraries, these samples were 
dominated by annotated sequences derived from viruses (Figure 3.1), consistent with 
previous reports of this D-DNA fraction utilizing transmission electron micrographs and 
epiﬂuorescence analyses (Linney et al. 2021). The viral metagenomic libraries had the 
lowest number of recovered sequences, many of which were novel and unannotated 
(61-73% unannotated across all samples). Recovered genes ranged from 13,080,771-
15,551,348, and averaged at 13,822,361 (±1,071,408) (Table 3.S1). For all samples, 
viral family-level annotated sequences were nearly split between Myoviruses (23-38%) 
and Podoviruses (19-32%), with minimal contributions from Siphoviruses (4-7%). Of the 
metagenomic libraries contributing to the viral libraries, Prochlorococcus phages, other 
Cyanophages and Pelagibacter phages were the most abundant (Figure 3.S1). 
Synechococcus and Prochlorococcus phages dominated the euphotic samples (75-125 
m). At the deep chlorophyll maximum, Synechococcus and Cyanophage sequences in 
the virus fraction metagomomes peaked. In mesopelagic samples, Pelagibacter phages 
and Vibro phages increased in proportion in the virus fraction. These depths are 
consistent with both cellular host and as well as virus abundances previously reported 
at Station ALOHA (Aylward et al. 2017; Mende et al. 2019; Luo et al. 2020).

Overall, the metagenomic libraries developed from the three dissolved DNA 
pools were distinct with respect to their microbial DNA compositions. The vesicle 
fraction was primarily dominated by a single taxonomic family (Pelagibacteraceae)
across all depths, the viral fraction was dominated by bacteriophages, and lastly the exocellular F-DNA pool had both bacterial and viral derived DNA. While DNA from the former two pools have been previously described by metagenomic analyses, the composition DNA in the F-DNA fraction has not been previously reported.

3.4.2 Depth of origin of D-DNA throughout the water column

To infer the depths of origin of different D-DNA fractions, we mapped DNA sequences against a depth-resolved microbial gene catalogue from Station ALOHA (Figure 3.2; Mende et al. 2017; Luo et al. 2020). The objective was to determine whether genes from the D-DNA fractions matched Station ALOHA genes recovered from the same sampling depths as the D-DNA, or whether the D-DNA was potentially transported from other depths.

The viral samples were most similar to Station ALOHA annotated genes, that matched the depth at which they were collected. This was particularly evident in the mesopelagic zone samples collected from 250-500 m (Figure 3.2). In these samples 25% of the genes were derived from their respective collection depths, with only 13% from the euphotic zone (5-200 m). Viral samples collected from the DCM had high contributions from typical DCM depths (100-175 m; 29-36%), as well as neighboring upper euphotic (5-75 m; 15%) and upper mesopelagic zones (200-250 m; 13%), with
only 1% from the lower mesopelagic zone (500-1000 m). Of the three D-DNA fractions, the viral samples had the highest average of genes with unknown depths (43% ±4.5%).

In contrast to the viral samples, the vesicle and F-DNA samples appeared to contain both autochthonous and allochthonous DNA (Figure 3.2). In the euphotic zone samples (75-125 m), as expected, sequences were dominated (>50%) by surface-derived DNA (5-200 m), with minimal mesopelagic zone contributions (<10%). However, in mesopelagic zone vesicle and F-DNA samples (250-1000 m), genes originated primarily from the upper euphotic zone (5-75 m; >30%), and to a lesser extent (<20%) the depth from which they were collected. Of these mesopelagic zone samples, the shallowest F-DNA sample (250 m) had the most depth-diverse genes, originating from depths throughout the euphotic and mesopelagic zones (5-500 m).

3.4.3 Size distributions of environmental F-DNA through the water column

The size spectra of recovered F-DNA was measured by capillary electrophoresis, following density gradient separation and buffer exchange. Seven F-DNA samples collected throughout the euphotic and mesopelagic zones (5-1000 m) were measured to assess the degradation of samples and molecular weight distributions prior to sequencing (Figure 3.3). Samples collected in the upper euphotic zone (5-100 m) had a distinct peak (<5000 bp peak width) of high molecular weight F-DNA (HMW; referred to here as >1,000 bp) and a lower proportion of low molecular weight (LMW; <1,000
bp) F-DNA, ranging between 24-38%, compared to mesopelagic zone (250-1000 m) samples which ranged between 33-65%. In these upper euphotic zone samples, there was a high proportion of F-DNA 1000-40,000 bp (62-73%), whereas in mesopelagic zone samples this HMW F-DNA tended to be lower (48% average, 35-66%). Lower euphotic (125 m) and mesopelagic zone samples (250-1000 m) tended to have a broader range of F-DNA sizes, suggesting that this DNA may have been more degraded. In mesopelagic zone samples, the HMW F-DNA decreased, the maximum peaks of HMW-DNA in samples 250-1000 m were less distinct (>10,000 bp peak width) and there was more F-DNA between peaks. In samples 5-100 m, <25% of the DNA was <350 bp. At 1000 m the peaks were unpronounced, suggesting a notable level of degradation in this deep sample.

3.4.4 Comparing vesicle, viral, and F-DNA fractions by non-metric multidimensional scaling

To compare all dissolved DNA fractions (vesicles, viruses, and F-DNA) with each other and previously reported Station ALOHA viral and P-DNA metagenomic sequences, two-dimensional ordination methods were employed. Bray-Curtis dissimilarity based non-metric multidimensional scaling (NMDS) of dissolved DNA and the Station ALOHA gene catalogue (Figure 3.S2A-B, respectively; 0.02 µm filtered “viral” and 0.2 µm filtered “cellular” communities) microbial communities were
compared on family-level annotated metagenomes. This comparison confirmed that the viral D-DNA samples collected in this study were similar in composition to previously characterized Station ALOHA viroplankton communities recovered from the same respective depths (Figure 3.S2A, stress = 0.11). As for F-DNA sequences, upper euphotic zone samples (75-100 m) clustered with their respective Station ALOHA viral samples, whereas lower euphotic (125 m) and mesopelagic zone (250-1000 m) samples clustered together and not with their respective sample depths. Similarly, vesicle D-DNA did not cluster with any Station ALOHA viral samples.

The same D-DNA sequences were compared to the cellular microbial community genes utilizing NMDS (stress = 0.07; Figure 3.S2B). From this analysis two distinct cellular communities emerged, euphotic (75-125 m) and mesopelagic (250-1000 m) Station ALOHA samples clustered together, in general consistent with previous reports (Mende et al. 2019). Viral D-DNA did not cluster with any Station ALOHA cellular communities. Similarly, F-DNA fractions did not cluster with their respective Station ALOHA depths, deep F-DNA samples (500 and 1000 m) clustered with cellular communities filtered from 75 m, revealing a potential cellular origin of this F-DNA. Upper mesopelagic (250 m) F-DNA clustered nearest cellular metagenomes collected from 125 m. Surface (75 m) vesicle D-DNA clustered with euphotic zone cellular communities, whereas lower euphotic (125 m) and mesopelagic (500 m) vesicles samples clustered with mesopelagic Station ALOHA cellular samples. D-DNA
samples that had high proportions of viral sequences (Figure 3.1) clustered together (Figure 3.S2B; viral D-DNA, 100 m F-DNA, and total dissolved DNA from 100 and 250 m).

3.5 Discussion

Describing the various forms of D-DNA in the ocean is important in order to better understand the diversity of marine life and microbial dynamics. Here we show that D-DNA is comprised of three distinct pools: vesicles, viruses, and exocellular F-DNA. Recent work has characterized viral communities found throughout the water column (Aylward et al. 2017; Luo et al. 2017; Luo et al. 2020) as well as vesicles as a potential mode of mobile gene element transfer (Hackl et al. 2020). The vesicle D-DNA pool was anticipated to be well represented by DNA from the cyanobacterium Prochlorococcus (Biller et al. 2014a; Biller et al. 2017), yet Prochlorococcus appeared to contribute to only 0.5-1% of the annotated sequences. Instead, the vesicle-derived DNA sequences were dominated by DNA derived from the ubiquitous heterotrophic bacterial family Pelagibacteraceae. The predominance of Pelagibacter in the vesicle pool may have several explanations. This DNA fraction could be derived in part from intact ultra-small Pelagibacter cells (<0.1 µm), or from Pelagibacter vesicles. It has been recently documented that structures resembling vesicles were produced by Pelagibacter (Zhao et al. 2017; Morris et al. 2020). There were several differences in
our study design compared to previous vesicle studies in marine plankton (Biller et al. 2014a; Biller et al. 2017), that may account in part for these results. These differences include our use of a 0.1 µm prefilter (compared to 0.2 µm prefilters used in previous studies), and our use of CsCl density gradients, compared to Iodixanol gradients (Biller et al. 2014a; Biller et al. 2017). Consistent with previous virome reports from the North Pacific Subtropical Gyre (Luo et al. 2017; Luo et al. 2020), the virus fraction was dominated by viruses related to those that infect Prochlorococcus and Pelagibacter (Figure 3.S1).

Our results reveal the first metagenomic characterization of F-DNA, alongside two other D-DNA pools, vesicles and viruses. The importance of further characterizing the F-DNA fraction is compounded by recent discoveries of the other D-DNA pools (viruses and vesicles) as potential vectors of genetic exchange in the marine environment (Hackl et al. 2020) and cyanobacterial pili capable of utilizing exogenous DNA (Taton et al. 2020; Aguilo-Ferretjans et al. 2021).

Exocellular microbial-derived DNA is not unique to the open ocean (DeFlaun et al. 1987, Karl & Bailiff 1989; Brum 2005; Linney et al. 2021). Terminology may vary, but F-DNA has also been reported in marine sediments (Torti et al. 2015) and terrestrial soils (Pietramellara et al. 2009), even contributing 90% of the total DNA pool in marine sediments (Dell’Anno et al. 2002). Vibrio cholerae are known to take up exogenous DNA from dead cells, which is thought to shape antibiotic resistance, surface
colonization, and intercellular communication (Ellison et al. 2018). Cell-free DNA has also been documented in the human bloodstream, in some cases it originates from tumor cells, and can be used as a noninvasive cancer diagnosis (Stewart & Tsui 2018). Free nucleic acids in the form of viroids (RNA) are even capable of causing infection in higher plants (Adkar-Purusothama & Perreault 2020). Across microbial systems, a variety of functions have been attributed to cell-free DNA, exemplifying the vast evolutionary and ecological potential F-DNA may have in the open ocean.

In the open ocean, F-DNA accounts for 25-50% of total D-DNA (Brum et al. 2004; Brum 2005; Linney et al. 2021), and has been shown to be rapidly consumed by microorganisms (Paul et al. 1987; Turk et al. 1992; Brum 2005; Lennon 2007). Whether it is used primarily as a nutrient source or for genetic exchange remains largely unknown. Previous work comparing the turnover of D-DNA pools, indicates that F-DNA is taken up more quickly than DNA inside of viruses (Brum 2005; Riemann et al. 2008), suggesting it may be more readily available than structurally enclosed D-DNA. Our investigation into F-DNA reveals that there are different microbial sources for this material at the surface (75-125 m) versus the mesopelagic (250-1000 m), implicating distinct ecological and evolutionary consequences.

At the surface, F-DNA sequences were dominated by viral and bacterial sequences – a possible manifestation of active viral lysis. Viral lysis is thought to be the dominant mode of viral replication at the surface, while there is evidence (viral gene
markers – integrase, repressor protein cl, and excisionase) that temperate phages may be more prevalent below the deep chlorophyll maximum (Luo et al. 2017). At depth, cell abundances and primary production are lower than at the surface (Karl & Church 2014), providing less energy and fewer hosts for viral replication (Moebus 1996; Weinbauer et al. 2003; Luo et al. 2020). This depth-distinction in viral replication at Station ALOHA coincides with the marked decrease in viral-derived F-DNA sequences in the mesopelagic, the same depths fewer free phage sequences (0.02 µm filtered) are observed (Luo et al. 2017). Conversely, when cells are lysed, the cell and viral machinery contents are exposed and spill into the open ocean. Thus all DNA that was enclosed by the cellular membrane, both viral and cellular DNA, may become F-DNA after lysis. Indeed, viral lysis has been documented as a production mechanism of both total D-DNA (Alonso et al. 2000) and F-DNA (Brum 2005; Riemann et al. 2009). Furthermore, an experiment documenting the production of free ribosomes following viral lysis of Synechococcus supports the possibility that viral lysis may be an important process introducing other free nucleic acids into the marine environment (Zhong et al. 2016). Interrupted phage packaging has not yet been explicitly described in marine systems, but there is evidence of unpackaged DNA from nuclease treatments following cyanophage infection of dominant picocyanobacteria (Baran et al. 2018; Zborowsky & Lindell 2019), as well as in other microbial systems (Liljemark & Anderson 1970; Powell et al. 1992). Together, our results suggest that viral lysis of bacterial cells at the surface
may be important to the production of F-DNA in the open ocean. Protozoan grazing (Turk et al. 1992; Alonso et al. 2000) and cell exudates (Paul et al. 1987; Paul et al. 1990) have also been shown to result in increased D-DNA and might also play significant roles in the production of F-DNA in the open ocean. More research, separating the D-DNA pools, is needed to uncover the source and dynamics behind these production mechanisms.

The high proportion of surface-originated sequences in mesopelagic F-DNA sequences suggests downward export of either cells or free-DNA, as a delivery mechanism (Figure 3.2). The exact mechanisms of F-DNA delivery to the mesopelagic have yet to be documented. Possible explanations can be inferred from previous works, including: sediment trap and water column metagenomic analyses (Fontanez et al. 2015; Mende et al. 2019), flux calculations of sinking particle disaggregation and degradation (Richardson & Jackson 2007; Collins et al. 2015), and migratory zooplankton (Hannides et al. 2013). All of these mechanisms begin by defining the source of F-DNA, which presumably originates from autochthonous particulate DNA (P-DNA; includes cellular DNA, detrital DNA, and eDNA), reflecting the microbial communities present at the collection depth, or from allochthonous P-DNA. Recently, surface-originated P-DNA (as filtered cells) has been documented in the mesopelagic. One such report found Prochlorococcus sequences filtered from recovered sediment traps deployed at Station ALOHA (Fontanez et al. 2015), and was hypothesized to be
due to entrainment on particles or in fecal pellets. Additionally, a recent investigation of core microbiome populations at Station ALOHA consistently found high-light *Prochlorococcus* ecotypes originating from the upper euphotic, at 500-1000 m in the P-DNA at Station ALOHA (Mende et al. 2019). Consistent with this observation, our work demonstrates that high-light *Prochlorococcus* also dominate the F-DNA metagenomes at similar depths (Table 3.S2), implicating a surface water export.

With this evidence of surface P-DNA exported from the surface, rapid surface turnover of free DNA, and mobilization by migratory zooplankton, it is presumed that the majority of DNA leaves the surface in the particulate phase, rather than dissolved. Following the sinking of P-DNA, it is theorized that particles disaggregate biologically and/or mechanically (Collins et al. 2015), and remineralization occurs by free-living or suspended microorganisms in the water column rather than microbial decomposition on sinking particles (Karl et al. 1988). This theory stems from studies investigating carbon flux, nutrient content of particles, on-particle metabolic activity, and accompanying models. Measurements of biomass production (particulate ATP) demonstrated there was a net loss of living material on particles (Karl et al. 1988). From this and accompanying rate measurements it was concluded that sinking particles were not likely to be sites of active microbial decomposition. Similarly, Collins et al. (2015) compared on-particle metabolic activity to the metabolic demand of the water column itself. They found that the measurement of on-particle remineralization could not
account for particle flux attenuation, suggesting that microbial decomposition occurs in
the water column rather than on particles.

By estimating the flux of particulate DNA out of the euphotic zone, delivery of
DNA to mesopelagic depths (by route of sinking particles) can be estimated
(Supplementary equation 3.1). This calculation assumes: (1) P-DNA becomes F-DNA by
sinking particle disaggregation, rather than on-particle degradation (Collin et al. 2015);
(2) the flux conforms to a Martin Curve (Martin et al. 1987); (3) adheres to export
efficiency and flux attenuation values determined from decades of production and
export analyses at Station ALOHA (Grabowski et al. 2019); and (4) steady-state. This
calculation predicts that the daily flux of DNA at 500 m is approximately 12-24% of the
concentrations of F-DNA reported at 500 m and 3-8% at 1000 m (Supplementary
equation 3.1; Brum 2005; Linney et al. 2021). This calculation is consistent with our
finding of 19% and 10% of the 500 m and 1000 m F-DNA annotated sequences were
derived from surface-originated *Prochlorococcus*. Despite their small size,
picophytoplankton like *Prochlorococcus* that dominate primary production in
oligotrophic oceans are theorized to contribute proportional carbon export values
(Richardson & Jackson 2007). This suggests that carbon exported from the surface
ocean is likely to be directly or indirectly derived from *Prochlorococcus* and other
primary producers, therefore surface-derived sequences are likely to contribute
significantly to mesopelagic F-DNA.
However, analyses revealed that F-DNA from 500 and 1000 m had more surface-derived sequences than the Martin curve and field data (Karl et al. 1996; Hebel & Karl 2001) predicts, 67 and 46% euphotic-derived (0-200 m) of sequences, respectively. If we assume that gene annotations accurately reflect D-DNA depth origins, a significant fraction of surface DNA remains unaccounted for - solely by sinking particles - suggesting other mechanisms may supplement F-DNA delivery to the mesopelagic. Delivery and F-DNA standing stocks may also depend on depth-dependent turnover rates of F-DNA, which are still not well constrained. Another important consideration is water mass controls on the supply of surface F-DNA. It is theorized based upon decades of salinity and temperature measurements at Station ALOHA (Lukas & Santiago-Mandujano 2008) that the water column is comprised of distinct water masses. From this, it is hypothesized that surface waters in the northwestern Pacific supply the North Pacific Intermediate Water which manifests at Station ALOHA around 500-770 m (Talley 1993). This water mass is estimated to have been in contact with the atmosphere only 30-60 years ago according to a study (Bullister et al. 2006) that used transient tracers to estimate apparent water mass age at Station ALOHA. This provides evidence for other additional delivery mechanisms of surface DNA to the mesopelagic.

Other possible mechanisms delivering surface DNA to the mesopelagic may be more episodic in nature, for example: the passage and disaggregation of organic
matter like the summer export pulse (Karl et al. 2012; Poff et al. 2021), migratory zooplankton (Hannides et al. 2013), and aggregates formed by picophytoplankton transparent exopolymer particles (Cruz & Neuer 2019) and/or clay-ballasted particles (Deng et al. 2015). Recent analyses of abyssal sinking particles (collected at 4000 m) at Station ALOHA found that the organic carbon-specific energy was relatively high and energy-replete suggesting that deep organic carbon may be surface-derived (Grabowski et al. 2019). Together, our findings and past reports highlight that surface and mesopelagic microbial communities may be more vertically connected than previously thought. Investigations of the mechanisms delivering F-DNA and other labile DOM constituents, is critical for expanding our understanding of water column dynamics and the biological carbon pump.

Our research utilized a new method (Linney et al. 2021) for separating ecologically important pools of D-DNA, and preliminary analyses of DNA sequences found in water column profiles of the different fractions of vesicles, viruses, and exocellular F-DNA. The results highlight the potential ecological contributions of D-DNA with respect to the cycling of limiting nutrients (nitrogen and phosphorus) in the open-ocean, as well as a viable vector of genetic exchange. While these results have important implications, more research into the dynamics of F-DNA is required to understand how this molecule is transformed from the particulate phase, its connection
to microbially-mediated food webs, and its capacity to carry surface microbial genes to mesopelagic depths.

3.6 References:


3.7 Acknowledgments: We thank the Hawaii Ocean Time-series team, captain and crew of the R/V Kilo Moana, R/V Kaʻimikai-O-Kanaloa, and R/V Falkor for their help at sea. We thank A. Burger for performing the Fragment Analyzer assays. This work was supported by the Simons Collaboration of Ocean Processes and Ecology (Award ID 329108 to DMK and EFD).
Figure 3.1. Taxonomic annotation of metagenomic sequences from the three dissolved DNA constituents collected from the North Pacific Subtropical Gyre. (A) Domain-level taxonomic composition and proportion of unannotated sequences for each sample. (B) Family-level taxonomic composition of three dissolved DNA constituents (vesicles, viruses, and exocellular free DNA). The low-density vesicle constituents were collected from three euphotic and mesopelagic depths (75, 125, and 500 m). Viral fractions were collected from five depths (75, 100, 125, 250, and 500 m) in the North Pacific Subtropical Gyre. Exocellular F-DNA samples were collected from six depths (75, 100, 125, 250, 500, and 1000 m).
Figure 3.2. Annotation of D-DNA metagenomic sequences suggests their probable depths of origin. Best sequence matches to the ALOHA 2.0 gene catalogue (and their corresponding sampling depths) were used to assign the probable depth of origin to individual D-DNA metagenomic sequence reads. The size of the pie chart is proportional to the total number of D-DNA metagenomic reads, whose best match is to Station ALOHA gene originating from a given corresponding water depth.
Figure 3.3. Fragment analysis of F-DNA samples from the North Pacific Subtropical Gyre. Relative fluorescence unit of each sample used to normalize and calculate proportion. Molecular weight values shown from 75-50,000 basepairs according to Agilent DNF-488-0500 protocol sizing range. Samples collected above the deep chlorophyll maximum (125 m) had more distinct peaks, samples below had a larger range of sizes indicative of degradation.
3.9 Supplementary Material

**Title:** Microbial sources of exocellular DNA in the ocean

**Authors:** Morgan D. Linney¹, John M. Eppley¹, Anna E. Romano¹, Elaine Luo¹, Edward F. DeLong¹, David M. Karl¹

**Affiliations:**
¹Department of Oceanography, Daniel K. Inouye Center for Microbial Oceanography: Research and Education (C-MORE), University of Hawai’i at Mānoa, Honolulu, HI 96822, USA.

*Corresponding author. Email: mlinney@hawaii.edu

**Document Contents:**
1. Supplementary Text (Results)
2. Table S1: Total number of gene counts for each dissolved DNA sample (vesicles, viruses, and free DNA)
3. Table S2: Proportion of *Prochlorococcus* HL sequences present in each dissolved metagenomic DNA sample (vesicles, viruses, and free DNA/FDNA)
4. Table S3. Microscopy and molecular characterizations of the cesium chloride gradient pools
5. Figure S1. Diversity of host-specified phages contributing to the three dissolved DNA pools.
6. Figure S2A-B. Bray-Curtis dissimilarity based non-metric multidimensional scaling (NMDS) of D-DNA fractions
7. Figure S3. Nutrients and cell abundance depth profiles at Station ALOHA
8. Supplementary Equation

**Supplementary Spreadsheets:**
1. DatasetS1: All family-level annotated taxon counts for all dissolved metagenomic DNA (vesicles, viruses, and free DNA/FDNA) samples
2. DatasetS2: Major bacterial and viral families (cutoff: major family contributes ≥0.65% of all DDNA samples) contributing to dissolved metagenomic DNA (vesicles, viruses, and free DNA/FDNA) gene counts. Data used in main text (as proportions).
3. DatasetS3: Host-associated phage gene counts for each dissolved metagenomic DNA sample (vesicles, viruses, and free DNA/FDNA) annotated from the ALOHA 2.0 catalog metadata

4. DatasetS4: Depth-assigned gene counts for each dissolved metagenomic DNA sample type (vesicles, viruses, and free DNA/FDNA) annotated from the ALOHA 2.0 catalog metadata. Data used in main text (as proportions).

Supplementary Text (Results):

A. Habitat characteristics.

Dissolved DNA samples were collected on cruises in the North Pacific Subtropical Gyre to Station ALOHA (22°45’N, 158°W): (i) November 2017, HOT-297 (Figure S3), (ii) April 2018, FK180310, and (iii) May 2018, HOT-302. For cruise (i), inorganic nutrients (N+N and phosphate) were typical of conditions at Station ALOHA ranging in the euphotic zone (0-200 m), 0.01-3.31 and 0.03-0.31 µM (N+N and phosphate, respectively). In the mesopelagic zone (500-1000 m) values ranged, 32.49-41.92 and 2.44-3.04 µM (N+N and phosphate, respectively). Fluorometric chlorophyll a ranged 0.01-0.26 µg L⁻¹ in the euphotic zone, consistent with historical averages. In the upper euphotic zone (5-100 m), Prochlorococcus cell abundances were consistent with previous Station ALOHA measurements (1.10-1.83 × 10⁵ cells mL⁻¹), and declined sharply in the lower euphotic zone (125-175 m; 0.01-0.30 × 10⁵ cells mL⁻¹). Cell abundances of heterotrophic bacteria were consistent with historical euphotic ranges (1.78-5.02 × 10⁵ cells mL⁻¹).
B. Confirmation of exocellular free DNA isolation

The method (Linney et al. 2021) applied to isolate vesicles, viruses, and free DNA was validated by a number of experimental tests. In brief, the method concentrates large volumes of prefiltered (0.1 µm) seawater by ultrafiltration (30 kDa), then separates D-DNA pools by density gradient ultracentrifugation. This allows for the direct examination of these three components. Following this separation these three pools were subjected to a number of experiments to assess the efficacy of this method. These experiments involved: visualization analyses (epifluorescence and transmission electron microscopy), macromolecular quantification (fluorescence quantification of proteins, RNA, and DNA), and DNase treatment. From these examinations it was determined that F-DNA was present in the most dense (1.6-1.7 g mL$^{-1}$ in cesium chloride) fraction of the density gradient. Visual examinations of this F-DNA isolate, confirmed the presence of DNA and the absence of viruses or vesicles. There was also a near complete (92-97%) degradation of fluorescently quantified DNA signal in this isolate following DNase treatment (Table S3). The utilization of independent evaluation methods helped confirm its application for F-DNA isolation, and warranted further investigations to identify its biological composition alongside the other D-DNA pools.
Table 3.S1. Total number of gene counts for each dissolved DNA sample (vesicles, viruses, and free DNA).

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Sample Collection Depth (m)</th>
<th>Dissolved DNA Type</th>
<th>Total read counts</th>
<th>Total annotated genes</th>
<th>Total family-level annotated gene counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ves75</td>
<td>75</td>
<td>Vesicle</td>
<td>12747332</td>
<td>12255903</td>
<td>6657660</td>
</tr>
<tr>
<td>Ves125</td>
<td>125</td>
<td>Vesicle</td>
<td>15397229</td>
<td>15166035</td>
<td>11550100</td>
</tr>
<tr>
<td>Ves500</td>
<td>500</td>
<td>Vesicle</td>
<td>16694964</td>
<td>15979742</td>
<td>11152500</td>
</tr>
<tr>
<td>Vir75</td>
<td>75</td>
<td>Virus</td>
<td>14362092</td>
<td>13970978</td>
<td>5420090</td>
</tr>
<tr>
<td>Vir100</td>
<td>100</td>
<td>Virus</td>
<td>13363478</td>
<td>13080771</td>
<td>5031730</td>
</tr>
<tr>
<td>Vir125</td>
<td>125</td>
<td>Virus</td>
<td>14031653</td>
<td>13693086</td>
<td>4942910</td>
</tr>
<tr>
<td>Vir250</td>
<td>250</td>
<td>Virus</td>
<td>13206072</td>
<td>12815625</td>
<td>3645800</td>
</tr>
<tr>
<td>Vir500</td>
<td>500</td>
<td>Virus</td>
<td>16163098</td>
<td>15551348</td>
<td>4223860</td>
</tr>
<tr>
<td>FDNA75</td>
<td>75</td>
<td>FDNA</td>
<td>12427745</td>
<td>11879884</td>
<td>6651250</td>
</tr>
<tr>
<td>FDNA100</td>
<td>100</td>
<td>FDNA</td>
<td>7737206</td>
<td>7585269</td>
<td>3862660</td>
</tr>
<tr>
<td>FDNA125</td>
<td>125</td>
<td>FDNA</td>
<td>14886720</td>
<td>11639558</td>
<td>6976200</td>
</tr>
<tr>
<td>FDNA250</td>
<td>250</td>
<td>FDNA</td>
<td>10650078</td>
<td>8509298</td>
<td>5337180</td>
</tr>
<tr>
<td>FDNA500</td>
<td>500</td>
<td>FDNA</td>
<td>14638768</td>
<td>13600011</td>
<td>9109520</td>
</tr>
<tr>
<td>FDNA1000</td>
<td>1000</td>
<td>FDNA</td>
<td>14653482</td>
<td>9800715</td>
<td>6138520</td>
</tr>
<tr>
<td>Sample Name</td>
<td>Sample Collection Depth (m)</td>
<td>Dissolved DNA Type</td>
<td>Sample Location</td>
<td>Cruise Number</td>
<td>% Prochlorococcus</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------</td>
<td>--------------------</td>
<td>-----------------</td>
<td>---------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Ves75</td>
<td>75</td>
<td>Vesicle</td>
<td>(22° 49’N, 158°00’W)</td>
<td>HOT297</td>
<td>0.9561485561</td>
</tr>
<tr>
<td>Ves125</td>
<td>125</td>
<td>Vesicle</td>
<td>(22° 49’N, 158°00’W)</td>
<td>HOT297</td>
<td>0.1446403620</td>
</tr>
<tr>
<td>Ves500</td>
<td>500</td>
<td>Vesicle</td>
<td>(22° 49’N, 158°00’W)</td>
<td>HOT297</td>
<td>0.2351377189</td>
</tr>
<tr>
<td>Vir75</td>
<td>75</td>
<td>Virus</td>
<td>(22° 49’N, 158°00’W)</td>
<td>HOT297</td>
<td>8.0577307866</td>
</tr>
<tr>
<td>Vir100</td>
<td>100</td>
<td>Virus</td>
<td>(22° 17’N, 157°01’W)</td>
<td>FK180310</td>
<td>5.1901582892</td>
</tr>
<tr>
<td>Vir125</td>
<td>125</td>
<td>Virus</td>
<td>(22° 49’N, 158°00’W)</td>
<td>HOT297</td>
<td>5.4472308710</td>
</tr>
<tr>
<td>Vir250</td>
<td>250</td>
<td>Virus</td>
<td>(22° 17’N, 157°01’W)</td>
<td>FK180310</td>
<td>6.1427366645</td>
</tr>
<tr>
<td>Vir500</td>
<td>500</td>
<td>Virus</td>
<td>(22° 49’N, 158°00’W)</td>
<td>HOT297</td>
<td>7.2196579772</td>
</tr>
<tr>
<td>FDNA75</td>
<td>75</td>
<td>FDNA</td>
<td>(22° 49’N, 158°00’W)</td>
<td>HOT297</td>
<td>4.8454270686</td>
</tr>
<tr>
<td>FDNA100</td>
<td>100</td>
<td>FDNA</td>
<td>(22° 17’N, 157°01’W)</td>
<td>FK180310</td>
<td>7.0830119712</td>
</tr>
<tr>
<td>FDNA125</td>
<td>125</td>
<td>FDNA</td>
<td>(22° 49’N, 158°00’W)</td>
<td>HOT297</td>
<td>41.3596521946</td>
</tr>
<tr>
<td>FDNA250</td>
<td>250</td>
<td>FDNA</td>
<td>(22° 17’N, 157°01’W)</td>
<td>FK180310</td>
<td>6.6518196999</td>
</tr>
<tr>
<td>FDNA500</td>
<td>500</td>
<td>FDNA</td>
<td>(22° 49’N, 158°00’W)</td>
<td>HOT297</td>
<td>40.89583484</td>
</tr>
<tr>
<td>FDNA1000</td>
<td>1000</td>
<td>FDNA</td>
<td>(22° 49’N, 158°00’W)</td>
<td>HOT302</td>
<td>30.02699432</td>
</tr>
</tbody>
</table>

**Table 3.S2.** Proportion of *Prochlorococcus* HL sequences present in each dissolved metagenomic DNA sample (vesicles, viruses, and free DNA/FDNA).
Table 3.53. Microscopy and molecular characterizations of the cesium chloride gradient pools. The recovery of DNA from each pool had a range of 68-86%.

<table>
<thead>
<tr>
<th>Pool</th>
<th>Density in cesium chloride</th>
<th>Dominant contents (examined by TEM)</th>
<th>Epifluorescence VLP detection</th>
<th>% DNA remaining post-DNase Treatment</th>
<th>DNA:protein range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vesicles</td>
<td>(1.30–1.35 g mL(^{-1}))</td>
<td>Filamentous structures, round viruses or vesicles</td>
<td>VLP detected</td>
<td>89–100</td>
<td>1.2–2.3</td>
</tr>
<tr>
<td>Viruses</td>
<td>(1.40–1.55 g mL(^{-1}))</td>
<td>Viruses (various morphologies)</td>
<td>VLP detected</td>
<td>88–98</td>
<td>4.3–6.8</td>
</tr>
<tr>
<td>Free DNA (F-DNA)</td>
<td>(1.60–1.70 g mL(^{-1}))</td>
<td>Linear structures (DNA)</td>
<td>VLP absent</td>
<td>3–8</td>
<td>30.6–42.1</td>
</tr>
</tbody>
</table>
Figure 3.S1. Diversity of host-specified phages contributing to the three dissolved DNA pools. Values are proportions of genes derived from viruses annotated to known hosts. The dominant phage hosts were determined by a contribution of >0.05% (cutoff) to all dissolved metagenomic gene counts, resulting in eight associated hosts across the three D-DNA pools.
Figure 3.52. Comparison of D-DNA samples with cellular and viral datasets collected from Station ALOHA. Bray-Curtis dissimilarity based non-metric multidimensional scaling (NMDS) of D-DNA samples (vesicles, viruses, and free DNA “FDNA”) and ALOHA 2.0 gene catalogue (“HOT”) family-level annotated abundances. Ellipsoids represent 95% CI around the centroid. Total “DDNA” samples were samples collected by the same ultrafiltration methods, but were not separated by density. (A) Comparison of viral ALOHA 2.0 gene catalogue (viral “HOT” dataset) with D-DNA samples. Surface free DNA (75 and 100 m) cluster with viral HOT samples from the same depth, while deeper (125-1000 m) F-DNA samples cluster together with the vesicle samples. “Virus” D-DNA samples cluster with viral “HOT” dataset from the same depths. (b) Analysis with cellular ALOHA 2.0 gene catalogue (“HOT”) and D-DNA samples. Virus-enriched D-DNA samples (“Virus”, total “DDNA”, and free DNA from 100 m) cluster together, and separately from cellular HOT datasets. Deeper free DNA (500-1000 m) samples cluster with cellular HOT datasets from the upper euphotic (75-100m).
Figure 3.S3. Nutrients and cell abundance depth profiles at Station ALOHA in November. Depth profiles of environmental data at Station ALOHA during the sampling period (2017-2019). (a) N+N depth profile (b) phosphate (c) Prochlorococcus and heterotrophic bacteria cell abundances (d) fluorometric chlorophyll.
Supplementary Equation:

1. **Calculate flux (F_{eu}) of particulate DNA out of the euphotic zone:**
   \[
   F_{eu} = \text{Net primary production (NPP)} \times \text{export efficiency (ef)} \times P:\text{DNA}:\text{PC}
   \]
   \[
   \text{NPP} = 500 \text{ mgC m}^{-2} \text{ day}^{-1} \quad (1)
   \]
   \[
   \text{ef} = 0.055 \quad (1)
   \]
   \[
   P:\text{DNA} = 0.2 \text{ mg m}^{-2} \text{ (Value at 100 m in North Pacific from 2)}
   \]
   \[
   \text{PC} = 1.5 \text{ mg m}^{-2} \text{ (Approximate value at 100 m acquired from HOTDOGS)}
   \]
   *DNA-Carbon and ef unitless*

2. **Calculate flux (F_z) at depth (z):**
   \[
   F_z = \text{Flux (euphotic)} \times (\text{depth/depth}_0)^{(-b)} \quad (3)
   \]
   
   Depth = \text{z (500 m and 1000 m where F-DNA samples were sequences, see Figure 2 in main text)}
   
   Depth_0 = \text{net productivity depth, 150 m at Station ALOHA}
   
   b = \text{flux attenuation for organic carbon, 0.89 (1)}
   
   *b is unitless*

3. **Source of F-DNA to water column at z is flux divergence of value from Step 2 (F_z):**
   \[
   \text{div} = (-b/z)^*(F_z)
   \]
   *units will be in mg m^{-3} day, which is the same as } \mu g L^{-1} day^{-1}*

4. **Calculate % of value at 500 m and 1000 m, respectively**
   
   500 m = 0.02 \mu g L^{-1} \quad (4, 5)
   
   1000 m = 0.01 \mu g L^{-1} \quad (4)

**Assumptions:**

1. Particulate DNA become F-DNA by disaggregation rather than degradation on particles (6)
2. Flux conforms to a Martin Curve (3)

**References:**


Chapter 4: Impacts of protistan grazing and viral lysis on exocellular DNA and ATP production

Morgan D. Linney¹

Co-authors: Svetlana Goldin², Karin M. Björkman¹, Tara Clemente¹, Dave A. Caron³, and Debbie Lindell²

Affiliations: ¹Daniel K. Inouye Center for Microbial Oceanography: Research and Education (C-MORE) and Department of Oceanography, University of Hawaii at Manoa, Honolulu, HI 96822, ²Faculty of Biology, Technion-Israel Institute of Technology 3200003 Haifa, Israel, ³Department of Biological Sciences, University of Southern California, Los Angeles, CA, 90089

Author contributions: DL and DAC designed research. MDL, SG, KMB, TC, DL, and DAC conducted research. MDL, SG, KMB, DL, DAC analyzed data.

Note: The (*) contents of the methods and results described herein are the unpublished materials belonging to Debbie Lindell and Dave Caron. Interpretations of results written by Morgan D. Linney.
4.1 Abstract:
In this study, two mortality agents - a podovirus (P-SSP7) and a chrysophyte grazer (Paraphysomonas bandaiensis) - were used to comparatively and study the mortality of a photoautotroph (Prochlorococcus MED4) and the production of F-DNA. Viral-driven mortality resulted in >50× more D-DNA (7.5±3.3 ng mL⁻¹) following viral lysis, than MED4 growth in the absence of viruses. Grazer-driven mortality yielded less D-DNA (1.1±0.4 ng mL⁻¹). Quantitative PCR (qPCR) revealed that phage DNA contributed >70% of the targeted F-DNA following viral lysis in all viral treatments - both solely and in the presence of the grazer. Our work shows that both Prochlorococcus growth and mortality produce F-DNA, but to varying degrees. Viral lysis contributed the highest concentrations of F-DNA and total D-DNA, while grazer activity produced minimal concentrations and reduced the overall accumulation of total D-DNA, F-DNA, and dissolved ATP (D-ATP). These analyses reveal key mortality dynamics that produce exocellular DNA and ATP, which may influence the microbial loop and sustain bacterial production.

4.2 Introduction:
Prochlorococcus is the most abundant photosynthetic microorganism in the open-ocean (Partensky et al. 1999) and populations play fundamental roles in the marine environment, both in life (conversion of light energy to biological energy) and
death (mediating the flux of carbon). *Prochlorococcus* populations are numerically dominant from about 40°N to 40°S, throughout the upper water column down to depths reaching about 200 m (Partensky et al. 1999). As a primary producer, *Prochlorococcus* fixes carbon dioxide, which goes on to fuel higher and higher trophic levels. On average, the global abundance of *Prochlorococcus* is about $3 \times 10^{27}$ cells, and accounts for an estimated 4 Gt C yr$^{-1}$ of ocean net primary production (Flombaum et al. 2013) supporting approximately a third of the global oxygen production (Johnson et al. 2006).

The abundance, global mass, and ecosystem modulation of *Prochlorococcus* corroborates that the products of their mortality is equally important—although debatably, more difficult to constrain. The mortality of *Prochlorococcus*, results in the production of dissolved organic carbon which may be used as a nutrient source by other microorganisms (Zhao et al. 2019), fueling 10-40% of bacterial production (Bertilsson et al. 2005). *Prochlorococcus* mortality is mediated by two dominant agents: cyanophages and protistan grazers (Christaki et al. 2001; Hirose et al. 2008; Ribalet et al. 2015; Mruwat et al. 2020), however the degree to which either viral lysis or grazing, solely or interactively mediates the release of dissolved *Prochlorococcus*-derived material into the marine environment remains unknown. The frequency and composition of the production of dissolved material is critical to constrain in order to
reveal the mechanisms, and role of these processes in producing DOC, and thereby mediating the microbial loop.

Both viral lysis and sloppy protistan grazing can shunt the biological carbon pump at the surface (0-200 m) by recycling dissolved material, rather than sinking out of the euphotic zone as particles. It has been documented that dissolved DNA (D-DNA) can be produced by both viral lysis and protistan grazing (Turk et al. 1992; Alonso et al. 2000). D-DNA is operationally defined as DNA that flows through a filter (typically 0.1-0.2 µm) and includes DNA inside of viruses, vesicles, and “free” DNA (F-DNA). Brum (2005) and Riemann et al. (2009) showed that both total D-DNA and F-DNA production were correlated with increased viral production in the North Pacific Subtropical Gyre and Baltic Sea, respectively. D-DNA production has also been documented from growing heterotrophic bacteria, without any mortality agents present (Paul et al. 1987) and by autolysis (Alonso et al. 2000). The rapid turnover and biological availability of dissolved adenosine triphosphate (D-ATP) in the sea suggest it may also be produced under similar mortality dynamics (Björkman & Karl 2001).

The concentration and bioavailability of D-DNA and D-ATP has both evolutionary and ecological significance. D-DNA is a highly labile (Turk et al. 1992; Brum 2005; Riemann et al. 2009) constituent of dissolved organic matter, and has been found in all aquatic environments that has been investigated (Pillai and Ganguly 1972; DeFlaun et al. 1986; Karl & Bailiff 1989; Linney et al. 2021 – Chapter 1). It can be used
by other microorganisms as a nutrient source (Turk et al. 1992; Jørgensen & Jacobsen 1996; Lennon 2007), or even taken up as high molecular weight DNA (>5,000 bp; Jeffrey et al. 1990; Frischer et al. 1994).

While *Prochlorococcus* is considered a unified “species” (>97% 16S rRNA similarity), the group is comprised of genetically distinct clades (or ecotypes) (Biller et al. 2015) that are physiologically unique. Ecotype differentiation has been shown to be dependent on environmental conditions, (e.g. temperature, light, phage resistance, and nutrient sources), and is attributed to hypervariable genomic islands (Coleman et al. 2006) and a vast pangenome (~80,000 collective of genes harbored by all *Prochlorococcus* clades) (Kettler et al. 2007; Biller et al. 2014). Recent work, found mobile genomic elements (tycheposons) were enriched in vesicles and viruses, suggesting a possible mode of genetic transfer in *Prochlorococcus* (Hackl et al. 2021). These tycheposons can carry functional genes (e.g. nutrient acquisition and phage defense) and are hypothesized to contribute to the diversity of these genomic islands, thus potentially facilitating the evolution of *Prochlorococcus*.

The presence of *Prochlorococcus* F-DNA sequences in the same D-DNA pool as tycheposons, has important implications as another potential mode of genomic transfer in the marine environment. F-DNA comprises up to 50% of the total D-DNA pool (along with vesicles and viruses), and 25% of the total DNA inventory in the North Pacific at Station ALOHA (Chapter 1). If *Prochlorococcus* is utilizing mobile elements in
2/3 of the D-DNA constituents (vesicles and viruses) to propagate genomic variability, there may also be mobile elements in F-DNA. Together these findings implicate the varied roles (e.g. as a nutrient or genetic source) D-DNA constituents play in the marine environment. Highlighting the importance of further revealing the sources and mechanisms producing these components in an oligotrophic ocean (Swan et al. 2013; Giovannoni et al. 2005; Tripp et al. 2010; Biller et al. 2014). To investigate the microbial production of exocellular DNA, we performed a comparative analysis to determine whether predator-prey interactions could be a significant source of D-DNA. We conducted culture-based experiments with Prochlorococcus and two mortality agents (a virus and a grazer), independently and in combination to assess any synergistic dynamics. Over the course of 48 hours, we measured D-DNA, F-DNA, and D-ATP concentrations, and identified the microbial sources of F-DNA under different mortality conditions. Two diel cycles were chosen as the time course because it was expected that mortality of Prochlorococcus would occur well-within this time course by both mortality agents. Our analyses reveal important mechanisms influencing the production of dissolved organic matter and the accumulation of exocellular DNA and ATP. These assessments have implications for how the mortality of an ecologically relevant and abundant microorganism is as important to open-ocean ecosystem as its photosynthetic roles, and illustrates effective tools for measuring these dynamics in lab- and field- based experiments.
4.3 Method (* = see Appendix for extended version)

*4.3.1 Experimental System (abbrev.)

Four experimental treatments were conducted in biological triplicates to assess the growth and mortality responses of MED4 (non-axenic, grown in Pro99 seawater-based medium; Moore et al. 2007) to two mortality agents: (1) MED4 only control, (2) MED4 with cyanophage P-SSP7, (3) MED4 with P. bandaiensis (non-axenic), and (4) MED4 with both mortality agents. P-SSP7 is a double-stranded DNA T7-like podovirus from clade B, specifically infecting MED4 (Sullivan et al. 2003). Starting concentrations of MED4 cells, heterotrophic bacterial cells, grazers, and viruses were: ~1.5-1.7 × 10⁷ cells mL⁻¹, ~1 × 10⁶ cells mL⁻¹, ~9 × 10⁶ phage mL⁻¹, ~6 × 10³ grazers mL⁻¹, respectively across the experimental treatments. MED4 and grazer cell abundances were enumerated every 2 hours, phage and heterotrophic bacteria were enumerated every 6 hours, and sampled for exocellular DNA (D-DNA and F-DNA) and ATP (D-ATP) every 12 hours. Exocellular DNA and ATP were collected following filtration through a glass-fiber filter (GFF 25 mm), with the filtrate collected into a 250 mL polyethylene bottle to separate cells from dissolved material. The filtrate from each sample was further processed to quantify exocellular DNA and ATP concentrations and to characterize the microbial F-DNA sources by quantitative PCR analyses. All microorganisms were enumerated using an Influx flow cytometer.
4.3.2 Intracellular Nutrient and DNA Calculations

Across all experiments, MED4 cellular carbon, nitrogen, and phosphorus concentrations were calculated from MED4 abundances. These estimations were used to approximate the concentration of organic nutrients released following MED4 cell mortality. Elemental stoichiometry was estimated from known MED4 elemental stoichiometry (Bertilsson et al. 2003) across the time-course to estimate the concentrations of intracellular nutrients and the subsequent supply of extracellular C, N, and P following MED4 population decimation. Phosphorus-replete values were used, with the following conversion factors: $3.8 \times 10^{-9}$, $6.7 \times 10^{-10}$, $3.2 \times 10^{-11}$ µM C, N, and P cell$^{-1}$.

DNA concentrations of MED4, P-SSP7, P. bandaiensis, were estimated based on known genome sizes and experimental abundances. Calculations are described in detail in Equation 1. Each genome size was converted to a molecular weight using known genomes sizes and known molecular weights of purine and pyrimidine bases. This provided a conversion factor for each microorganisms that could be used to estimate DNA mass from all the experimental treatments. This calculation is an estimate and does not account for changes in polyploidy. Each treatment had a mixed assemblage of heterotrophic bacteria with varying proportions across each experiment. Therefore, the concentration of DNA from heterotrophic bacteria was the average
genome size of the characterized heterotrophic microbial components (3.5 Mbp, nearly double MED4) measured in each experimental treatment (Kearney pers. communication). These values were used to approximate intracellular DNA concentrations for each microorganism.

4.3.3 Exocellular DNA measurements

Exocellular DNA samples were processed using a low-volume (≤5 mL) modified method (Figure 4.1) developed by Brum et al. (2004). In this method, centrifugal ultrafiltration is used to quantify low concentrations of D-DNA and F-DNA, allowing for high sensitivity measurements (0.01 ng mL⁻¹ detection limit). The concentration of F-DNA is calculated as the difference between the total D-DNA sample, and the DNase-treated sample; therefore, for each time point two sample types are collected (DNase-treated and total D-DNA). In this study, D-DNA and DNase-treated 5 mL subsamples were collected from the same filtrate (GFF prefiltered). Due to volume restrictions, the total D-DNA was a single measurement for each time point, while DNase-treated samples were conducted in triplicate.

Following prefiltration, total D-DNA samples were concentrated by centrifugal ultrafiltration (Amicon <10 kDa centrifugal units) and buffer exchanged in to Tris-EDTA (TE) buffer (pH 8). Following desalting with TE, the sample was concentrated to ~0.1
mL and quantified by fluorescence using the Qubit HS dsDNA Assay kit (ThermoFisher Scientific) following the manufacturer’s instructions.

DNase-treated samples were processed similarly to total D-DNA samples, with a few additional steps. Degradation of F-DNA was conducted under conditions known to fully digest F-DNA while keeping encapsidated phages intact (Baran et al. 2018). To do this, prefiltered (GFF) samples were incubated with DNase I (5 U mL$^{-1}$) at 37°C for 60 minutes. 50 mM EDTA (pH 8) was used to inactivate the enzyme. Following degradation, DNase-treated samples are concentrated and buffer exchanged the same way as total D-DNA samples (Figure 4.1). Triplicate DNase-treated samples are averaged and subtracted from the total D-DNA value to obtain a concentration of F-DNA.

4.3.4 Dissolved ATP

To assess the release of other dissolved material throughout the experiments, D-ATP was analyzed alongside exocellular DNA. D-ATP samples were analyzed using a volume-modified method of Björkman and Karl (2001) which uses firefly bioluminescence to quantify ATP. This method employs the MAGIC (MAGnesium-Induced Coprecipitation; Karl & Tien 1992) procedure to concentrate dissolved phosphorus-containing molecules (like ATP). Each sample was split into duplicate 50 mL samples and concentrated using 1N NaOH to initiate MAGIC precipitation of D-
ATP. Following precipitation, the pellet was dissolved in 0.5 N HCl after which the volume was adjusted to 0.5 mL with deionized water. Samples were processed for all replicates of the MED4 control and viral experimental replicates.

*4.3.5 Viral abundances and Quantitative PCR (qPCR)

Real-time qPCR was used to quantify P-SSP7 viral abundances and determine the microbial sources (MED4, *P. bandaiensis*, and P-SSP7) contributing to the F-DNA samples. qPCR was conducted before and after DNase treatment (as described above) to assess sources vulnerable to degradation, thus defining “F-DNA” for this study. To enumerate encapsidated viral counts, P-SSP7 specific primers (Zborowsky & Lindell 2019) were utilized across the time course. qPCR DNA concentrations were calculated based on known genome sizes of MED4 (1,657,990 bp), *P. bandaiensis* (31,189,705 bp), and P-SSP7 (44,970 bp). One gene count was assumed to equate to an entire genome.

4.4 Results (* = see Appendix for extended version with cell abundances)

*4.4.1 Minimal exocellular DNA and ATP produced by growing Prochlorococcus

The time course of *Prochlorococcus* MED4 growth without either agent of mortality served as an experimental control (Appendix Figure 1a). Following 48 hours of a consistent light-dark cycle and replete nutrients, the non-axenic culture reached
calculated intracellular C, N, and P concentrations of 161, 28, and 1.3 µM, respectively, compared to 65, 11, and 0.5 µM at the start (0 hours) of the experiment. Note, these are estimations based on experimental abundances and published MED4 cellular stoichiometry (Bertilsson et al. 2003), and do not account for changes in cell size. Background heterotrophic bacteria were minimal (approximately 1-5% of MED4 cell abundances) with low variability across the control treatments (Appendix Figure 1a). Although MED4 cell-associated DNA reached ~75 ng mL⁻¹ at 48 hours (Appendix Figure 3a), the growth of MED4 did not result in net production of exocellular ATP or DNA. Total D-DNA and D-ATP remained relatively constant in the controls, with average concentrations of 0.17 (±0.09) ng mL⁻¹ (Figure 4.2) and 0.08 (±0.01) ng mL⁻¹ (Figure 4.3), respectively, with F-DNA below the detection limit of the Qubit 4.0 fluorometer (Figure 4.2). However, qPCR is more sensitive and determined Prochlorococcus as the only measured source of F-DNA (Figure 4.2; note that F-DNA from heterotrophic bacteria was not measured by qPCR).

*4.4.2 Exocellular DNA and ATP released by viral-lysis of Prochlorococcus

Viral-driven mortality of Prochlorococcus produced 50× and 5× more D-DNA and D-ATP, respectively (Figure 4.4; D-DNA, Figure 4.5; D-ATP), compared to the concentrations produced from the same MED4 culture without being challenged by P-SSP7 (Figure 4.2; D-DNA, Figure 4.3; D-ATP). The initial phage and MED4
concentrations were 9.30 (±1.03) × 10^6 phage mL\(^{-1}\) (Appendix Figure 2a) and 1.64 (±0.05) × 10^7 cells mL\(^{-1}\) (Appendix Figure 1b), respectively, when P-SSP7 and MED4 were first introduced. Viral lysis began around 6 hours, but took effect on the MED-4 population by 24 hours, at which time cell abundances dropped and remained low until the end of the experiment (36-48 hours; <10% of the starting population).

Estimates based on MED4 cell stoichiometry, predict ~22, 5, 1 μM C, N, P was released extracellularly in the first day. By this time, encapsidated phage counts (Appendix Figure 2a) increased to 2.47 × 10^8 (±5.53 × 10^7) gene copies (phages) mL\(^{-1}\) similar to the phage F-DNA gene copies at this time point 2.66 × 10^8 (±4.59 × 10^7) copies mL\(^{-1}\).

Across all mortality conditions, the highest concentrations of D-DNA, F-DNA, and D-ATP coincided with the time course of viral lysis of MED4 cells. Concentrations of D-ATP, D-DNA, and F-DNA peaked to 0.34 (±0.03) ng mL\(^{-1}\), 7.50 (±3.3) ng mL\(^{-1}\), 4.9 (±3.3) ng mL\(^{-1}\), respectively at 24 hours (Figure 4.4; D-DNA and F-DNA, Figure 4.5; D-ATP). In other words, concentrations were 5x, 20x, and >100x higher than before viral lysis (0 hrs) for D-ATP, D-DNA, and F-DNA respectively. qPCR determined that both P-SSP7 and MED4 contributed to the F-DNA pool during infection (0-24 hr; Figure 4.4), but following viral lysis the majority of the targeted (qPCR measured) F-DNA originated from the phage (>80%), and increased in proportion across the time course, and eventually became the sole measured source of F-DNA (100%) by 36 hours (Figure
4.4). By 30 hours, encapsidated phage counts was $3.09 \times 10^8$ (±$2.18 \times 10^7$) mL$^{-1}$, and outnumbered phage F-DNA gene copies $2.55 \times 10^8$ mL$^{-1}$ at this time point (Appendix 2a), but with high variability (±$1.33 \times 10^8$) across triplicate experiments. Following 30 hours, encapsidated phage gene copies continued to outnumber phage F-DNA gene copies (Appendix Figure 2a).

Throughout the time course, populations of heterotrophic bacteria were ~10× higher than throughout the MED4 control (Appendix Figure 1b). Heterotrophic bacteria cell counts increased by 24 hours following viral-driven mortality of MED4 populations, and coincided with the 24 hour peak in exocellular DNA (Figure 4.4) and ATP (Figure 4.5). By 36 hours, exocellular DNA and ATP had rapidly declined (4.09±1.87, 2.17±1.60, and 0.07±0.03 ng mL$^{-1}$, D-DNA, F-DNA, and D-ATP, respectively) to <50% of the 24 hour peak. By the end of the experiment (48 hours) D-ATP reached the average control concentration (0.08±0.01 ng mL$^{-1}$; Figure 4.4), while D-DNA (1.63±0.70 ng mL$^{-1}$) and F-DNA (0.55±0.96 ng mL$^{-1}$) were still 5× and >20× higher than at the beginning of the viral-driven mortality experiment (0 hours). This decline in exocellular DNA and ATP corresponded with increasing concentrations in heterotrophic bacteria cell abundances (Appendix Figure 1b).

*4.4.3 Grazer consumption of Prochlorococcus mortality products*
Starting abundances in the combined virus and grazer mortality experiments were, $1.72 \pm 0.11 \times 10^7$ MED4 cells mL$^{-1}$, $6.3 \pm 2.5 \times 10^3$ grazers mL$^{-1}$, and $8.62 \pm 3.44 \times 10^7$ phage mL$^{-1}$ respectively (Appendix Figure 1c). However, the anticipated “grazer-only” experiment was contaminated with P-SSP7 viruses ($\sim 7.7 \times 10^7$ phage mL$^{-1}$), thus no comparison of combination and “grazer-only” experiments can be made. Therefore what was “grazer-only” will be termed from hereon as “combination 1” (Appendix Figure 1c) and the other combination experiment will be “combination 2” (Appendix Figure 1d). In both combination experiments heterotrophic bacteria cell counts were similar to those observed when MED4 was not challenged by a mortality agent (Appendix Figure 1a). However, heterotrophic bacteria populations declined across the time course in combination 1 (Appendix Figure 1c), while there was more observed variability in abundances in combination 2 (Appendix Figure 1d).

Estimates based on MED4 cell stoichiometry, predict 66, 12, 0.6 uM cell-associated C, N, P was lost (either in the particulate and/or dissolved phase) following MED4 mortality in the first day. Intracellular grazer DNA went from an estimated 0.2 (0 hours) to $\sim 2$ ng DNA mL$^{-1}$ by 24 hours in both combination experiments, and coincided with MED4 intracellular decrease from 31 to 0 ng DNA mL$^{-1}$ over the same time course (Appendix Figure 3).

Although combined grazer- and viral-driven MED4 mortality resulted in a marked decrease in estimated intracellular MED4 DNA (Appendix Figure 3), D-DNA
and F-DNA did not accumulate in either combination experiments (Figure 4.6 & 4.7).

Both total D-DNA and F-DNA were more variable across biological replicates, than was observed in the control (Figure 4.2) and viral-driven (Figure 4.4) mortality conditions. In both combination experiments, D-DNA (1.34±0.10 and 1.95±0.43 ng mL\(^{-1}\), combination 1 and 2, respectively) and F-DNA (0.31±0.21 and 0.67±0.45 ng mL\(^{-1}\), combination 1 and 2, respectively) concentrations peaked at 12 hours. In both combination experiments, total D-DNA was >500% higher than in the control, but was still lower than in the viral-driven mortality experiments.

In combination 2, qPCR revealed P-SSP7 as the dominant contributor to F-DNA, while MED4 contributed <5% of all qPCR quantified F-DNA from 0-12 hours (Figure 4.7). By 12 hours, encapsidated phage counts were \(1.77 \times 10^8 \, (\pm 4.34 \times 10^7) \, \text{mL}^{-1}\), compared to phage F-DNA gene copies \(6.11(\pm 1.43) \times 10^7 \, \text{mL}^{-1}\) at this time point (Appendix 2b). Throughout the combined grazer- and viral-driven MED4 mortality experiment (combination 2), encapsidated phage gene copies were higher than free phage gene copies by an order of magnitude (\(\times 10^1 \, \text{gene copies mL}^{-1}\)). By 24 hours P-SSP7 was the sole contributor to F-DNA (Figure 4.7). Despite the high abundance of grazers (Appendix Figure 1d), <1% of the qPCR measured F-DNA was attributable to the grazer across the time course in combination 2.

4.5 Discussion
While it is well established that photoautotrophs like *Prochlorococcus* supply organic substrates to fuel heterotrophic production, the pathways producing biologically relevant molecules like F-DNA and D-ATP are unconstrained. Here, we show - using culture-based experiments - that both MED4 growth and mortality can result in the extracellular release of these molecules, but at different magnitudes. Exponentially growing MED4 cultures yielded background levels of D-DNA and D-ATP. Viral lysis yielded the highest concentrations of dissolved material, while combined viral- and grazer-driven mortality yielded less dissolved material. Heterotrophic bacterial abundances were highest following viral lysis. These findings support the hypothesis that where viral lysis is more active in the open ocean, bacterial production may be higher, fueling the microbial loop, whereas, grazer-driven mortality of *Prochlorococcus* may yield lower concentrations of dissolved material. The high concentrations of F-DNA following viral lysis also has implications for the time course in which it is viable for genetic exchange, or perhaps providing nutrients for surviving microbial cells.

Exponentially growing MED4 cultures yielded low concentrations of exocellular DNA and ATP, in some cases undetectable. Past studies found dissolved DNA, RNA, and protein production coincided with the growth of heterotrophic bacteria (Paul et al. 1987), compared to these values, photoautotrophic organisms like MED4 may be more conservative in the release of dissolved material. Similarly, Zhong et al. (2016) found
that a growing heterotroph (Vibrio) produced free ribosomes, while a growing photoautotroph (Synechococcus) did not yield measurable free ribosomes. During unchallenged growth, F-DNA was undetectable by the fluorometric method (detection limit = 0.01 ng mL⁻¹) but was detected using qPCR which has an amplification step leading to higher sensitivity. Additionally, qPCR DNA concentrations are estimations based on gene counts extrapolated to entire genomes (one gene count equate to an entire genome), thus this measure may artifactually be higher than other quantification methods. The documentation of total D-DNA yields from MED4 growth alone is a novel finding. The lack of F-DNA detection by the fluorometric method could mean this D-DNA is protected from DNase activity, perhaps enclosed in extracellular vesicles. A similar study using high-performance liquid chromatography showed that, unchallenged MED4 exudates may be less bioavailable (determined by total dissolved amino acid yields and subsequent bacterial production on exudates) than viral lysates (Xiao et al. 2021).

While combined viral- and grazer-driven mortality had higher yields of F-DNA than the control, the concentrations were variable. Compared to the wealth of knowledge about the environmental forces impacting Prochlorococcus, less is known about the impact of predatory grazers. Grazers, like, P. bandaiensis (~5 µm diameter; Takahashi 1976) play vital roles at the base of the marine food web, by connecting globally important picocyanobacteria to higher trophic levels (Caron & Finlay 1994;
Sherr & Sherr 1994). In this study, grazer-driven mortality did not yield high concentrations of D-DNA or F-DNA. The low concentrations of dissolved material may suggest that *P. bandaiensis* may be efficient at consuming MED4 products (including DNA), and may not be a “sloppy” grazer. This is consistent with *P. bandaiensis* employing food vacuoles to digest its prey, rather than mechanical feeding used by copepods and other mesozooplanktonic grazers (Møller 2007). The rapid decimation of MED4 in both combination experiments could suggest grazing dynamics are not influenced by the viral infection, however, until “grazer-only” experiments are conducted this cannot be stated conclusively. In combination experiments, the high proportion of viral sequences in the F-DNA may suggest that if the grazers digest infected cells they may not have completely digested all of the viral contents (i.e. viral DNA). While this grazer yielded less D-DNA and F-DNA, it is possible that other types of grazers may have different mechanisms contributing varying magnitudes of dissolved material.

Another important consideration posed by the rapid decimation of MED4 in grazer treatments suggests that future studies should collect samples with a greater time resolution (every 2-4 hours) to fully characterize the production of dissolved material. Here, we collected samples every 12 hours, by the timepoint (T=12) MED4 cell abundances were driven down to 10% of the starting population. These first 12 hours are critical collection times, and ought to be sampled at least every three hours.
to assess whether there are rapid increases in dissolved material following initial contact between MED4 and grazer.

In treatments with viruses, there were high proportions of viral F-DNA following viral lysis. Unpackaged phage DNA has not been explicitly documented in marine microbial systems, but this study provides documentation of the microbial sources of F-DNA following viral lysis. In most viruses, the mechanisms of viral DNA synthesis, capsid production, and packaging are independent processes controlled by different genes and enzymatic processes (Karhu et al. 2007). Additionally, significant amounts of metabolic energy (1 ATP molecule for every 2 basepairs in phi29; Guo et al. 1987) are required to package DNA into capsids, suggesting that once the infected cells become energy depleted not viral DNA may be packaged – resulting in empty capsids and viral F-DNA at the time of viral lysis. The accumulation of unpackaged viral DNA in P-SSP7 has also been linked to the mispackaging of host DNA into capsids (Laurenceau et al. 2020). Or perhaps, it is more advantageous for viruses to synthesize excess DNA to ensure packaging, therefore yielding high concentrations of unpackaged phage DNA following viral lysis. However, the exact mechanisms behind this observation are yet to be determined, all we know at this time is that there is measurable unpackaged P-SSP7 DNA following the lysis of MED4 cells. Thus, when viral lysis is active, more viral F-DNA may be released into the environment. This is consistent with the finding that viral F-DNA sequences are higher in proportion within the euphotic zone (Chapter 2) in the
North Pacific Subtropical Gyre – the same depths where viral lysis has been shown to be a dominant mode of viral replication (Luo et al. 2017). Further research is needed in order to identify the mechanisms leading to unpackaged phage DNA, as this is likely a significant process contributing F-DNA to the environment.

Identifying the mechanisms, rates, and dynamics of F-DNA production and uptake are critical in order to better understand its role in the oligotrophic ocean. Here we measured the production of exocellular DNA and ATP associated with the growth and mortality of an ecologically-important photoautotroph. Further work employing radio- or fluorescently-labelled DNA alongside size fractionation of targeted microorganisms (predator and prey) could help determine the flow of exocellular DNA and ATP through viral- and/or grazer-driven mortality.
References:


**Acknowledgments:** We would like to thank the mortality team for their assistance and excellent teamwork, the Technion and the Lindell Lab for hosting. This work was supported by the Simons Foundation.
Filter (GFF) sample to exclude cells and collect exocellular nucleotides

Quantification of total D-DNA

5 mL filtrate concentrated to ~100 μL by centrifugal ultrafiltration (10kDa)

Wash concentrated sample in Tris-EDTA Buffer

Recover ~100 μL DNA sample

Quantify total D-DNA on Qubit using dsDNA HS kit

Quantification of F-DNA

Add DNase I (5 U mL⁻¹) to 5 mL filtrate to hydrolyze F-DNA

Add EDTA to inactivate DNase I enzyme

Concentrate to ~100 μL by centrifugal ultrafiltration (10kDa)

Wash concentrated sample in Tris-EDTA Buffer

Recover ~100 μL DNA sample

Quantify remaining DNA on Qubit using dsDNA HS kit

\[ [\text{F-DNA}] = [\text{Total D-DNA}] - [\text{DNase treated D-DNA}] \]

**Figure 1.** Schematic representation of total D-DNA and F-DNA quantification for all samples. [F-DNA] is determined by the difference between the [total-DNA] and the [DNase treated D-DNA] samples.
Figure 4.2. Time-course of D-DNA (“Total DDNA”, black) and F-DNA (“Free DNA”, red) quantified over the time course of MED4 control (no mortality agents added). Error bars represent standard deviation of the error across triplicate experiments. Pie charts depict proportion of microbial source qPCR measured F-DNA (collected by Lindell Lab). Proportional qPCR F-DNA mass (ng mL\(^{-1}\)) is estimated based on measured gene copies. Each gene copy is equated to the mass of a whole genome based on each targeted microorganism (MED4 “Prochlorococcus”, P. bandaiensis “grazer”, P-SSP7 “phage”) known genome size.
Figure 4.3. Time-course of D-ATP quantified over the time course of MED4 control (no mortality agents added). Error bars represent standard deviation of the error across triplicate experiments.
Figure 4.4. Time-course of D-DNA (“Total DDNA”, black) and F-DNA (“Free DNA”, red) quantified over the time course of MED4 challenged by a virus (P-SSP7). Error bars represent standard deviation of the error across triplicate experiments. Pie charts depict proportion of microbial source qPCR measured F-DNA (collected by Lindell Lab). Proportional qPCR F-DNA mass (ng mL\(^{-1}\)) is estimated based on measured gene copies. Each gene copy is equated to the mass of a whole genome based on each targeted microorganism (MED4 “Prochlorococcus”, or P-SSP7 “phage”) known genome size.
Figure 4.5. Time-course of D-ATP quantified over the time course of MED4 challenged by a virus (P-SSP7). Error bars represent standard deviation of the error across triplicate experiments.
Figure 4.6. Time-course of D-DNA ("Total DDNA", black) and F-DNA ("Free DNA", red) quantified over the time course combined grazer- and viral- driven mortality of MED4 ("combination 1"). Error bars represent standard deviation of the error across triplicate experiments.
Figure 4.7. Time-course of D-DNA (“Total DDNA”, black) and F-DNA (“Free DNA”, red) quantified over the time course combined grazer- and viral- driven mortality of MED4 (“combination 2”). Error bars represent standard deviation of the error across triplicate experiments. Pie charts depict proportion of microbial source qPCR measured F-DNA (collected by Lindell Lab). Proportional qPCR F-DNA mass (ng mL⁻¹) is estimated based on measured gene copies. Each gene copy is equated to the mass of a whole genome based on each targeted microorganism (MED4 “Prochlorococcus”, P. bandaiensis “grazer”, P-SSP7 “phage”) known genome size.
Equation 1:
1. **P-DNA conversion factor calculations for MED4, P-SSP7, and *Paraphysomonas bandaiensis***. Conversion factor used in combination with abundances to estimate DNA concentrations across the four mortality conditions.

1. **Known genome sizes:**
   - MED4 = 1657990 bp (Rocap et al. 2003)
   - P-SSP7 = 44970 bp (Sullivan et al. 2005)
   - *P. bandaiensis* = 31189705 bp (Caron et al. 2002)
   - *HB* = 3,500,000 bp (average)
2. **Convert to moles (using Avogadro’s #):**
   - MED4 = 2.75e-18 molDNA cell⁻¹
   - P-SSP7 = 7.47e-20 molDNA phage⁻¹
   - *P. bandaiensis* = 5.18e-17 molDNA flagellate⁻¹
   - HB = 5.8e-18 molDNA cell⁻¹
3. **Conversion factor (grams DNA/cell, phage, flagellate, 650 g/mol bp):**
   - MED4 = 1.79e-15 gDNA cell⁻¹
   - P-SSP7 = 4.85e-17 gDNA phage⁻¹
   - *P. bandaiensis* = 3.37e-14 gDNA flagellate⁻¹
   - HB = 3.8e-15 gDNA cell⁻¹

**Considerations:**
- Conversion factors do not account for changes in cell size
- Heterotrophic bacteria is a mixed community ranging in genome size from 2-5 Mbp cell⁻¹, 3.5 Mbp was used as an average value
Chapter 5: Conclusions and future directions
Morgan D. Linney

1Daniel K. Inouye Center for Microbial Oceanography: Research and Education (C-MORE) and Department of Oceanography, University of Hawaii at Manoa, Honolulu, HI 96822 USA
This work investigating free DNA (F-DNA) is presented in the context of exocellular nucleotide dynamics, which consists of not only F-DNA, but also DNA inside viruses, and subcellular vesicles. Together this DNA passes a filter (typically 0.1 or 0.2 µm) and is operationally defined as dissolved DNA (D-DNA). Altogether D-DNA in the oceans is equivalent in magnitude to intracellular DNA. D-DNA in the ocean is not a new discovery, its presence in the marine environment has been enigmatic since it was first identified as a component of dissolved organic matter in 1972 (Pillai & Ganguly 1972). Following its discovery, D-DNA was investigated in bulk to answer questions surrounding nutrient cycling (DeFlaun et al. 1986; Karl & Bailiff 1989). But, recent technological advances have provided additional perspectives of exocellular DNA, as a possible vector of genetic transfer and its importance in sustaining microbial evolution. Decades since it was first measured, armored with novel approaches this dissertation is among a growing body of literature on exocellular DNA that is rapidly advancing. In a way, we are just beginning to understand the microbial diversity and dynamics of exocellular DNA in the oligotrophic ocean.

Prior to this work, it was unclear whether F-DNA was truly present in the open ocean or whether it was an experimental artifact of DNA enclosed in marine viruses or subcellular vesicles. Much of the work done on D-DNA was pursued before marine viruses (Torrella & Morita 1979; Bergh et al. 1989) and cyanobacterial vesicles (Biller et al. 2014) were found to be abundant in the open ocean. Therefore, in previous studies
the entire D-DNA pool was measured (DeFlaun et al. 1986; Karl & Bailiff 1989), with F-DNA quantified by difference or inference rather than direct quantification (Brum et al. 2005; Riemann et al. 2009). A central motivation of this dissertation was to identify a method that separates these three known D-DNA constituents (viruses, vesicles, and F-DNA), in order to investigate them directly.

In Chapter 2, this method was presented, and subsequently published in Limnology and Oceanography: Methods (Linney et al. 2021). In brief, the method consists of concentrating prefiltered seawater (0.1 µm) by tangential ultrafiltration, then separates vesicles, viruses, and F-DNA by density. DNA is more dense than protein-enclosed viruses, and viruses are more dense than lipid-enclosed vesicles. This method allows for the isolation of each constituent. Following separation, the three pools were subjected to a number of experiments to assess the efficacy of this method. These experiments involved: visualization analyses (epifluorescence and transmission electron microscopy), macromolecular quantification (fluorescence quantification of proteins, RNA, and DNA), and DNase treatment. From these examinations it was determined that F-DNA was present in the most dense fraction of the density gradient.

Following the application of this method, Chapter 3 describes the microbial sources of these three D-DNA constituents, providing the first fully sequenced water column profile of D-DNA. Interestingly, two distinct “groups” of F-DNA samples emerged (euphotic zone: 75-100 m; and, mesopelagic zone: 125-1000 m). These two
groups were distinguished by the microbial groups that comprised them as well as by fragment analyses. These results may imply different F-DNA production mechanisms at different water-column depths. Specifically, the euphotic zone samples had high proportions of viral F-DNA and were representative of the microbial communities present at those depths, while mesopelagic zone samples had a higher proportion of bacterial F-DNA as well as a significant contribution of surface originated F-DNA. These findings have implications for export as a delivery mechanism of F-DNA to depth, as well as viral lysis as plausible source of F-DNA at the surface.

Among the F-DNA sequences were high proportions of the abundant photoautotroph, *Prochlorococcus*. Following field measurements, Chapter 4 describes laboratory experiments with two mortality agents - a virus (P-SSP7) and a grazer (*Paraphysomonas bandaiensis*) - were used to comparatively and synergistically measure the mortality rates of *Prochlorococcus* (MED4) and the production of F-DNA. These experiments revealed that high concentrations of exocellular nucleotides (total D-DNA, F-DNA, and D-ATP) coincided with viral lysis. qPCR revealed that >80% of the F-DNA originated from the virus, with <20% from *Prochlorococcus* host. This result has a number of implications. One important consideration is that the host’s DNA is typically degraded when a cell is infected, and the system is hijacked to synthesize viral DNA rather than host DNA. This could yield a lower proportion of host to viral DNA. The other implication from this result is the possibility of incomplete phage packaging.
Why would the host be lysed before all the DNA was packaged? In some bacteriophages, the mechanisms of viral DNA synthesis, capsid production, and packaging are independent processes controlled by distinct genes and enzymatic processes (Karhu et al. 2007). Additionally, significant amounts of ATP energy (1 molecule/2 basepairs in phi29; Guo et al. 1987) are required to package DNA into capsids, suggesting that once the infected cell runs out of energy not all capsids may be packaged – resulting in empty capsids and/or unpackaged DNA at the time of viral lysis. Conversely, in the grazer-driven mortality experiments F-DNA did not accumulate, suggesting *P. bandaiensis* may be efficient at consuming *Prochlorococcus*. This finding is consistent with *P. bandaiensis* employing food vacuoles to digest its prey, rather than mechanical feeding used by copepods and other mesozooplanktonic grazers (Møller 2007).

This study is among a field of research on exocellular DNA dynamics that is rapidly advancing. Back when this field was first emerging, it became evident that D-DNA was a ubiquitous component of DOM, and was presented in all aquatic environments that were investigated, freshwater rivers and lakes (Pillai and Ganguly 1972), the open ocean (Karl & Bailiff 1989), coastal systems (DeFlaun et al. 1986), and in sediments (Dell’Anno et al. 2002). Initial investigations into D-DNA described the concentration ranges of D-DNA, as well as the dynamics of D-DNA production (Turk et al. 1992; Alonso et al. 2000) and utilization (Lennon et al. 2007). Now, decades later
genomics and metagenomics are providing additional insights into the genetic origins and potential of exocellular DNA and its importance in sustaining microbial evolution. Recent work found evidence of mobile genetic elements enriched in \textit{Prochlorococcus} vesicles and viruses (Hackl et al. 2020). The presence of \textit{Prochlorococcus} F-DNA sequences in the same exocellular DNA pool as tycheposons, has important implications as another potential mode of genomic transfer in the marine environment. F-DNA comprises up to 50% of the total dissolved DNA pool (along with vesicles and viruses), and 25% of the total DNA inventory in the North Pacific at Station ALOHA (Linney et al. 2021).

Additionally, research has documented cyanobacterial pili capable of utilizing exocellular DNA (Taton et al. 2020), as well as to prevent sinking and predation (Aguilo-Ferretjans et al. 2021). From this research, patterns of pili utilization emerged, specifically, \textit{Synechococcus elongatus} pili dynamics were shown to follow a circadian rhythm (Taton et al. 2020). Another possible pattern of exocellular DNA uptake could be following a mortality event, like in Chapter 4. In a non-marine microbial system, \textit{Vibrio cholerae} was shown to use exocellular DNA following the mortality of other \textit{V. cholerae} cells (Ellison et al. 2018). With the release of F-DNA following viral-driven mortality of \textit{Prochlorococcus}, this DNA may provide a source of nutrients or nucleotides to surviving \textit{Prochlorococcus} cells, or other microorganisms.
While this dissertation research has hopefully played a role in moving the field of microbial exocellular DNA research forward, there are still important questions left to be answered:

1. Is exocellular DNA used primarily as a nutrient source a vector of genetic exchange, or both? How might this vary among heterotrophs or autotrophs?
2. What is the environmental turnover of F-DNA, and how might production and uptake dynamics vary throughout the water-column?
3. What export mechanisms contribute to delivering surface DNA to mesopelagic depths?
4. Could exocellular DNA serve as a vector of infection (e.g. viroid)?

Some experimental approaches to questions (1) and (2) could involve radiolabeling (e.g. $^{33}$P) known DNA sequences and tracking whether it was taken up by cells. In the field this could be measured by incorporation into filtered cells in incubation experiments to calculate environmental turnover. In the lab, radiolabeled DNA could be paired with qPCR measurements of genomic DNA to identify whether the DNA was incorporated. However, there are limitations to documenting these dynamics as we are just beginning to recognize distinct patterns of exocellular DNA uptake (e.g. circadian rhythms, mortality events, etc.)
Additionally, to investigate the export mechanisms (3) of exocellular DNA, F-DNA on sinking particles could be pursued to measure the proportion of surface DNA degraded at depth to yield F-DNA below the euphotic zone. This might involve measuring concentrations of particulate and F-DNA in sediment traps to calculate on-particle degradation and the flux out of the euphotic zone. It would be important to have these measurements alongside water-column respiration rates and respiration rates of bacteria attached to sinking particles. This would help in identifying the processes that may transform particulate DNA into F-DNA. Additional observations of migratory zooplankton would be important in constraining these export mechanisms.

Identifying potential viroids (4) could be done from a computational perspective, but would be expensive and challenging to pursue observationally (e.g. microscopy, fluorescent quantification, etc.) as there is no evidence yet of their existence in a dsDNA form or in the marine environment. But in a genetically reduced ocean, could infective viruses devolved or evolved from viroids?

While scientific research inevitably provides more questions than it answers, the hope of this dissertation research has been to provide a re-evaluation of exocellular DNA constituents and significance in the marine environment. While the ocean was our platform for studying exocellular DNA, it is relevant to many fields of biology as it is ubiquitous in nature. Examining microbial processes through the lens of exocellular DNA, has provided insights into how labile DOM may be produced at the surface (by
viral lysis), and the strength of the biological pump in exporting and transforming organic matter. With the combination of past D-DNA investigations and current pursuits we are beginning to answer critical questions regarding the ecological and evolutionary significance of exocellular DNA in aquatic environments.

References:


