CHARACTERIZATION OF VIRUSES CAUSING LYSIS OF A TOXIC BLOOM-FORMING ALGA, *HETEROSIGMA AKASHIWO* (RAPHIDOPHYCEAE)

by

VERA TAI

BSc. (Hon.) University of New Brunswick, 1998

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES
Department of Botany

We accept this thesis as conforming to the required standard

The University of British Columbia

October 2001

© Vera Tai, 2001
In presenting this thesis in partial fulfilment of the requirements for an advanced
degree at the University of British Columbia, I agree that the Library shall make it
freely available for reference and study. I further agree that permission for extensive
copying of this thesis for scholarly purposes may be granted by the head of my
department or by his or her representatives. It is understood that copying or
publication of this thesis for financial gain shall not be allowed without my written
permission.

Department of _Botany_

The University of British Columbia
Vancouver, Canada

Date _October 26, 2001_
ABSTRACT

Two viruses, HaRNAV and HaNIV, infecting the toxic bloom-forming alga *Heterosigma akashiwo* (Hada) Hada ex Hada et Chihara were characterized. HaRNAV is a 25 nm, single-stranded RNA virus with a genome size of approximately 9100 nucleotides. The virus particle has a lipid component as indicated by its sensitivity to chloroform and contains at least 5 structural proteins ranging in apparent size from 24 to 34 kDa. This is the first report of a single-stranded RNA virus infecting a phytoplankton. HaNIV is a 50 nm, double-stranded DNA virus with a 38.5 kb genome. HaNIV particles are also sensitive to chloroform and the capsid contains a major 38 kDa protein and a less abundant 20.5 kDa protein. HaRNAV and HaNIV caused lysis of a different range of strains, but both predominantly caused lysis of strains from the Northeast Pacific. Because different viruses infected different host strains, the complexity of virus-host interactions in the environment is demonstrated. Over 90% of the HaNIV genome was sequenced and is represented as two contigs, 17 765 and 15 858 base pairs in length. Open reading frames identified from the sequences had similarities with terminase, lysozyme, recombination, and structural proteins of bacteriophages from the *Myoviridae*, *Podoviridae*, and *Siphoviridae*. The homology of HaNIV ORFs with proteins from diverse bacteriophages suggests that the evolution of HaNIV involved numerous horizontal gene transfer events. HaNIV was also shown to be latent. Latency may be an important survival strategy for phytoplankton viruses. Irrefutable evidence proving that the host of HaNIV particles is *Heterosigma akashiwo* and not a bacterium, however, is imperative before the ecological and evolutionary implications can be determined.
TABLE OF CONTENTS

Abstract ii
Table of Contents iii
List of Tables v
List of Figures vi
Acknowledgements viii

Chapter 1. Introduction to the bloom dynamics and harmful effects of *Heterosigma akashiwo* 1

1.1. Harmful Algal Blooms, a Definition 1
1.2. Harmful Effects of *Heterosigma akashiwo* 3
1.3. Bloom Dynamics of *Heterosigma akashiwo* 4
1.4. Viruses in Marine Environments 8
1.5. Literature Cited 11

Chapter 2. Characterization of HaRNAV and HaNIV, novel single-stranded RNA and double-stranded DNA viruses causing lysis of *Heterosigma akashiwo* 19

2.1. Introduction 19
2.2. Materials and Methods 22
  2.2.1. Host cells 22
  2.2.3. Virus isolates 22
  2.2.4. Growth conditions 23
  2.2.5. Amplification and purification of viruses 23
  2.2.6. Transmission electron microscopy 25
  2.2.7. Chloroform sensitivity 25
  2.2.8. Structural proteins 26
  2.2.9. Nucleic acid extraction and analysis 26
  2.2.10. Hybridization assay 28
2.3. Results 29
  2.3.1. Morphology 29
  2.3.2. Chloroform sensitivity 31
  2.3.3. Structural proteins 32
  2.3.4. Nucleic acid type and genome size 33
  2.3.5. Hybridization 35
  2.3.6. Host range 36
2.4. Discussion 36
  2.4.1. HaRNAV 38
  2.4.2. HaNIV 40
  2.4.3. Host range and ecological significance 42
2.5. Literature Cited 45
Chapter 3. Genomic sequences from HaNIV show homology to bacteriophage genes

3.1. Introduction 55
3.2. Materials and Methods 57
   3.2.1. Construction of HaNIV clone library 57
   3.2.2. Sequencing and assembly of contigs 58
   3.2.3. Sequence analysis 59
   3.2.4. Detection of the HaNIV genome in the host, Heterosigma akashiwo 61
3.3. Results 62
   3.3.1. Sequence analysis 62
   3.3.2. Phylogenetic analysis 67
   3.3.3. Analysis of the host genome 68
3.4. Discussion 72
   3.4.1. ORF 16A and 18A possibly encode terminase subunits 72
   3.4.2. ORF 18B shares similarities with phage lysozymes 73
   3.4.3. ORF 1B is homologous with NinB sequences 74
   3.4.4. ORF 21A may be a structural protein 75
   3.4.5. Phylogeny 76
   3.4.6. Algal virus or bacteriophage? 78
   3.4.6. HaNIV may be latent 81
3.5. Literature Cited 83

Appendix A. Sequence of HaNIV contig A 92

Appendix B. Sequence of HaNIV contig B 98
Table 2.1. Sensitivity of various *Heterosigma akashiwo* strains to HaNIV and HaRNAV. + indicates the virus caused cell lysis, - indicates no lysis compared to a control culture, * indicates a discrepancy from the result reported by Lawrence et al. (2001). NEPCC 522 was the strain used for virus isolation.

Table 3.1. GenBank accession numbers for protein sequences used for phylogenetic analyses. * indicates that the assigned function for the protein is putative.

Table 3.2. ORFs predicted from HaNIV sequence. The letters A and B denote the contig from which the ORF was identified. The coordinates describe the location of the ORF from the first nucleotide of the first start codon with a ribosome binding site (RBS) to the last nucleotide of a stop codon. If a RBS was not detected, then the ORF is described from the first encountered start codon, regardless of a RBS. The reading frame is shown in parentheses. * indicates a GTG start codon, otherwise it is ATG. The length of the translated ORF is in amino acids (aa). Matches show significant matches (P < 0.005) for the appropriate ORF. For the highest ranking match, GenBank accession numbers for the protein database precede the description of the protein and e-values and identities are provided. Identities show the percentage of identical amino acids in the homologous region (length shown in parentheses).
| Figure 1.1. | Bloom dynamics of *Heterosigma akashiwo*. Initiation, growth, and termination phases are described in the text. | 5 |
| Figure 1.2. | Potential role of viruses in bloom dynamics. (A) Viruses are rapidly transmitted through a dense monospecific population compared to (B), a population of low density. Rapid, species-specific virus transmission might explain the sudden dissipation of *Heterosigma akashiwo* blooms. | 7 |
| Figure 2.1. | Electron micrographs of virus particles causing lysis of *Heterosigma akashiwo*. A) HaNIV particle negatively stained with UA. B) HaRNAV from the lower band in the sucrose gradient, stained with PTA. Scale bars = 50 nm. | 30 |
| Figure 2.2. | Growth of *Heterosigma akashiwo* in the presence of (A) HaNIV and (B) HaRNAV with and without chloroform. For both graphs, open symbols indicate chloroform treatments with the appropriate virus added (○) and without (△). Solid symbols indicate control treatments (no chloroform) with virus added (●) and without (▲). The arrow indicates when viruses were added to the cell cultures. In each case, values are the mean ± SD, n = 8. | 31 |
| Figure 2.3. | Structural proteins of *Heterosigma akashiwo* viruses. Proteins obtained from purified HaNIV (lane 1) and HaRNAV from the upper band (lane 2) and lower band (lane 3) of the sucrose gradient. Molecular weight markers (lane 4). The single 18 kDa capsid protein from tobacco mosaic virus was used as a control (lane 5). The proteins were stained with Coomassie Blue. | 32 |
| Figure 2.4. | Restriction enzyme digests of HaNIV DNA. HaNIV DNA was digested with Hind III (lane 1) and Hinc II (lane 2). Bacteriophage λ DNA digested with Hind III (lane 3) was used as a molecular weight marker. | 33 |
| Figure 2.5. | Genomic RNA isolated from HaRNAV particles (lane 3). dsRNA from bacteriophage φ6 (lane 1) and ssRNA from TMV (lane 2) were used as controls. RNA standards were run in lane 4. | 34 |
| Figure 2.6. | Hybridization of HaNIV DNA with DNA from A) non-infected *Heterosigma akashiwo*, (B) HaNIV infected *H. akashiwo*, (C) a resistant strain of *H. akashiwo*, and (D) unlabeled HaNIV. DNA was blotted onto the membrane in serial dilutions. | 35 |
Figure 3.1. Location of HaNIV ORFs along contigs A and B. Black arrows indicate ORFs with homologies to proteins in GenBank. The arrowhead indicates the direction of transcription.

Figure 3.2. Unrooted parsimony tree of viral LSU terminases and HaNIV ORF 18A. The consensus tree is shown. Numbers at branch nodes indicate percentage bootstrap support from 100 replicates. Values less than 50 % are not shown. Any significant support from the neighbor-joining analysis is shown in parentheses.

Figure 3.3. Unrooted parsimony tree of lysozyme amino acid sequences and HaNIV ORF 18B. The consensus tree is shown. Numbers at branch nodes indicate percentage bootstrap support from 100 replicates. Values less than 50 % are not shown. Any significant support from the neighbor-joining analysis is shown in parentheses.

Figure 3.4. PCR analysis of DNA from infected and non-infected *Heterosigma akashiwo*. PCR products with HaNIV primers using template DNA from non-infected *H. akashiwo* (lane 1), infected *H. akashiwo* (lane 2), and HaNIV particles used as a positive control (lane 4). DNA was not amplified from DNA isolated from a resistant strain of *H. akashiwo* (lane 3). The negative control was run in lane 5 and molecular weight markers in lane 6.
ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Curtis Suttle, for providing me with a challenging and rewarding place to work for my Master’s degree. And along with my committee members, Dr. Max (F.J.R.) Taylor and Dr. Jim Hudson, thank you for supporting my research and providing critical and helpful advice when asked.

My research would not have been successful without the help and advice of Amy Chan, Dr. Janice Lawrence, Steven Short, and Dr. Andrew Lang. Thank you Amy, in addition to managing the lab, for helping me get my project off the ground. Thank you Janice for many insightful conversations, improving my culturing techniques, and taking the time to work with me on the electron microscope. Thank you Steve for helping me with cloning techniques, PCR, sequence analysis, and all things molecular. And for the helpful suggestions made during the late stages of my thesis, thank you Andrew.

I would like to thank Elaine Humphries and the EM lab at UBC for the use of their facilities. Thank you to the group at the Nucleic Acid Protein Services (NAPS) lab at UBC for their prompt and helpful service. For funding, thank you to the Natural Sciences and Engineering Research Council (NSERC).

Finally, thank you to those who have made the lab an enjoyable place work and helped me out whenever needed. Thanks to Sean Brigden, André Comeau, and Alice Ortmann. Thank you to Tanya St. John for being an endless source of friendship and curing my soccer itch. And thanks to Alex Culley, Cindy Frederickson, Nicole McLearn, Karen Reid, Maarjte Verschuur, Eric Chu, George Huang, Michelle Lin, Cydney Nielsen, Nicole Radziminski, Nathan Taylor, Darrell Yip, and Derek Yuen.
CHAPTER 1. Introduction to the bloom dynamics and harmful effects of

*Heterosigma akashiwo*

Harmful algal blooms (HABs) have occurred at least since biblical times based on a passage in Exodus "...all the waters that were in the river were turned to blood. And the fish that were in the river died, and the river stank, and the Egyptians could not drink the water of the river" (Exodus 7: 20-21). These blooms of phytoplankton are noted for their striking colour, hence they are more commonly referred to as red tides, and their toxic effects to humans and aquatic animals. Of the estimated 3400-4100 species of phytoplankton, about 300 have been reported to form blooms including diatoms, dinoflagellates, silicoflagellates, prymnesiophytes, and raphidophytes, but only 60-80 of these are known to be harmful or toxic (Sournia 1995). Included in the group of harmful bloom-forming phytoplankton is a member of the Raphidophyceae, *Heterosigma akashiwo* (Hada) Hada ex Hada et Chihara, the subject of this thesis.

**Harmful Algal Blooms, a Definition**

What constitutes a harmful algal bloom is not strictly defined and is better understood when broken down into two parts: what is "harmful" and what is a "bloom". Traditionally, phytoplankton are considered to be harmful if they threaten public health or aquaculture industries, but they also have negative in situ ecological effects, such as allelopathy, on organisms not directly related to human interests (Smayda 1997). Public
health usually is threatened by toxin-producing phytoplankton. Shell-fish poisoning is the most common cause due to the accumulation of toxins in shell-fish as they feed on phytoplankton. In the Pacific Northwest, paralytic shell-fish poisoning (PSP), caused by saxitoxins produced by dinoflagellates from the genus *Alexandrium*, amnesic shell-fish poisoning (ASP), caused by domoic acid produced by diatoms from the genus *Pseudo-nitzschia*, and diarrhetic shell-fish poisoning (DSP), caused by okadaic acid produced by dinoflagellates from the genus *Dinophysis*, are of concern (Taylor et al. 1994, Horner et al. 1997). Although harmless to the shell-fish, the presence of these toxins renders them poisonous to humans and forces the closure of shell-fish farms.

Harmful phytoplankton species also adversely affect fish aquaculture. Fish mortalities due to phytoplankton result from several mechanisms: physical damage to the gills, by the barbed spines of *Chaetoceros* spp. for example; asphyxiation due to oxygen depletion following blooms, such as with *Ceratium* species; gas-bubble trauma caused by extreme oxygen supersaturation of seawater from algal photosynthesis; and chemical toxicity due to the release of toxins, including brevetoxins, hemolysins, cytotoxins, or ammonia produced by species such as *Karenia brevis*, *Gyrodinium aureolum*, *Chrysochromulina polylepsis*, and *Noctiluca scintillans* (Black 1991, Smayda 1997).

Blooms are usually linked to harmful phytoplankton events because dense populations, in sum, have more cells to cause physical damage, use more oxygen, or produce a greater abundance of toxins. A toxic event and the occurrence of a bloom, however, are not mutually dependent. Ingestion of only a few cells of a toxic phytoplankton can result in the death of fish larvae (Smayda 1997). In addition, not all blooms are harmful. Blooms are often described as significant increases in population
size measured as abundance (i.e. cells·L⁻¹). Defining "significant increases in population size", however, is problematic because maximal, potential bloom levels can vary depending on geography, season, and species (Smayda 1997). Nor is it useful terminology because toxicity can result from high or low biomass bloom events and high or low cell abundances (Smayda 1997). Despite these conceptual difficulties, monitoring the presence of harmful species, the abundance of toxins, and understanding the factors that regulate the occurrence, toxicity, and bloom dynamics of specific species are important strategies for mitigating the negative effects of phytoplankton.

Harmful Effects of *Heterosigma akashiwo*

*Heterosigma akashiwo* is considered a harmful phytoplankton species because of its devastating impact on finfish aquaculture. A member of the Raphidophyceae, *H. akashiwo* is a 10-25 μm, yellow-brown, oval-shaped, unicellular flagellate species that lacks a cell wall (Honjo 1993, Taylor et al. 1994, Smayda 1998). This organism has also been referred to as *Heterosigma carterae* (Hulbert) Taylor and has been mistaken for *Olisthodiscus luteus* Carter in some publications (Taylor 1992, Throndsen 1996). *H. akashiwo* forms blooms that cause fish kills of cultured salmon, yellow-tail, and red sea bream in coastal seas of temperate and sub-arctic areas of both Northern and Southern hemispheres (Chang et al. 1990, Honjo 1993, Taylor and Haigh 1993). In Japan, damage caused by *H. akashiwo* blooms has amounted to 2 billion yen (about $25 million) over a period of 16 years (1972-1987) (Honjo 1993). In British Columbia, from 1986 to 1991,
*H. akashiwo* was responsible for losses of greater than $12 million to the finfish aquaculture industry (Taylor 1993, Taylor and Horner 1994). It is not known how *Heterosigma akashiwo* causes fish mortalities. The production of reactive oxygen species such as superoxide and hydroxyl radicals resulting in damage to fish gill lamellae has been suggested (Yang et al. 1995). *H. akashiwo* is also believed to produce low amounts of neurotoxins that are chemically similar to brevetoxins produced by *Karenia brevis* (Khan et al. 1997). This result is consistent with an earlier study by Black et al. (1991) who proposed that the death of juvenile chinook salmon (*Oncorhynchus tshawytscha*) exposed to a *H. akashiwo* bloom was likely due to a toxic agent rather than deoxygenation, oxygen supersaturation, or physical gill damage. In addition, it was found that *H. akashiwo* was toxic only when bacteria were present in culture experiments with salmonids (Carrasquero-Verde 1999). The specific mechanism of *H. akashiwo* fish toxicity remains unclear, however, and further research is required.

**Bloom Dynamics of *Heterosigma akashiwo***

Blooms of *Heterosigma akashiwo* occur in cycles of initiation, growth, and termination (Fig. 1.1). The blooms originate from an over-wintering population of non-motile vegetative cells and cysts in ocean sediments (Yamochi 1984, Honjo 1993). The blooms are initiated by the release of motile cells from the resting stages triggered by a rise in temperature above 10°C (Taylor and Haigh 1993, Smayda 1998, Imai and Itakura 1999). Once released, for successful bloom initiation, the motile cells must overcome competition from other species and obtain sufficient nutrients for cell growth. Because
Figure 1.1. Bloom dynamics of *Heterosigma akashiwo*. Initiation, growth, and termination phases are described in the text.

*H. akashiwo* cysts can germinate under low-light conditions, *H. akashiwo* has an advantage over competing species such as the diatom *Skelotonema costatum* which prefers light for germination (Smayda 1998, Imai and Itakura 1999). *H. akashiwo* is also known to produce allelopathic substances which inhibit the growth of competitors and deter grazers (Pratt 1966, Verity and Stoecker 1982, Egloff 1986, Uye and Takamatsu 1990).

For growth, *Heterosigma akashiwo* requires inorganic nutrients such as phosphate and nitrate, vitamins, and trace metals, especially iron and manganese (Honjo 1993, Smayda 1998). As a flagellated cell, *H. akashiwo* obtains nutrients from deeper waters at night and photosynthesizes at the surface during the day through diurnal vertical
migrations (Watanabe et al. 1988). The daily upward and downward migrations allow *H. akashiwo* to take advantage of stratified environments where nutrients limit growth in the euphotic zone, but are at higher concentrations at depth (Honjo 1993, Smayda 1997, Smayda 1998). Stratification due to fresh water run-off from rivers may also trigger blooms through the introduction of iron and other micronutrients into coastal systems (Honjo 1993, Taylor and Haigh 1993).

Once a bloom of *Heterosigma akashiwo* has developed, it can last several days (Nagasaki et al. 1996) to several months (Taylor and Haigh 1993). The blooms appear as red-brown discolourations of the water and can reach magnitudes of 200 million cells·L\(^{-1}\) (Taylor et al. 1994). Not all blooms reach harmful levels, however. A preliminary estimate suggests that fish will begin to die at 12 million cells·L\(^{-1}\), but this value varies depending on the species, size of the fish, and environmental factors (Taylor et al. 1994).

The dissipation or termination of *Heterosigma akashiwo* blooms often occurs very suddenly and may be triggered by several factors. Physical factors include the breakdown of stratification by wind-induced turbulence (Taylor and Haigh 1993). Nutrient limitation may also be involved (Smayda 1998). Increasing evidence, however, indicates that biological factors, such as bacteria and viruses, can have a significant role in *H. akashiwo* bloom termination (Nagasaki et al 1994a, b, Kim et al. 1998). As pathogens, bacteria and viruses can kill or cause lysis of *H. akashiwo* resulting in a reduction of *H. akashiwo* cells and the disintegration and dissipation of the bloom. In addition, viruses are typically species or strain specific and would be rapidly transmitted through a dense monospecific population as in bloom situations (Fig. 1.2). Although the specific triggers are not known, during bloom termination, motile *H. akashiwo* cells cease vertical
migrations, form cysts or remain as non-motile vegetative cells, and sink, coming to rest in the sediments (Itakura et al. 1996).

As described above, *Heterosigma akashiwo* bloom dynamics are characterized by several phases: initiation, rapid and exponential growth, maintenance, and termination.

The factors that regulate the length, perseverance, severity, and dissipation of *H. akashiwo* blooms, however, are not well understood.

Figure 1.2. Potential role of viruses in bloom dynamics. (A) Viruses are rapidly transmitted through a dense monospecific population compared to (B), a population of low density. Rapid, species-specific virus transmission might explain the sudden dissipation of *Heterosigma akashiwo* blooms.
Viruses in marine environments

There is considerable interest in understanding the factors regulating *H. akashiwo* bloom dynamics because of the damaging effects of *H. akashiwo* blooms to aquaculture fish. Viruses are an important factor to consider because, in addition to the possibility of directly causing bloom termination by cell lysis, they may also influence bloom distribution, length, and severity. Viruses may affect other ecological processes as well such as nutrient cycling (Gobler et al. 1997, Wilhelm and Suttle 1999), primary productivity, community structure, and genetic transfer (Suttle et al. 1990, Bratbak et al. 1994).

Until recently, viruses in aquatic environments were believed to be low in abundance and ecologically unimportant (Wiggins and Alexander 1985). These estimates were based on culture assays with various bacterial hosts. Thus, there is little historical data regarding marine phytoplankton viruses. Most reports of phytoplankton viruses result from incidental observations of virus-like particles inside cells using transmission electron microscopy (reviewed in van Etten et al. 1991). Total direct virus counts, however, based on transmission electron and epifluorescence microscopy, have revealed that viruses generally range from $10^6$-$10^9$·ml$^{-1}$ in aquatic environments and are frequently 10 fold higher than the bacterial abundance (Bergh et al. 1989, Proctor and Fuhrman 1990, Suttle et al. 1990, Bratbak et al. 1994). These estimates are $10^3$-$10^7$ times higher than those based on culture assays (Bergh et al. 1989).

Although in aquatic environments most viruses are bacteriophages (viruses infecting bacteria), viruses infecting eukaryotic phytoplankton are also a component of
the aquatic community. To date, eukaryotic phytoplankton viruses have been isolated which infect *Aureococcus anophagefferens* (Gastrich et al. 1998), *Chrysochromulina* sp. (Suttle and Chan 1995), *Heterosigma akashiwo* (Nagasaki and Yamaguchi 1997, Lawrence et al. 2001), *Micromonas pusilla* (Mayer and Taylor 1979, Waters and Chan 1982, Cottrell and Suttle 1991), and *Phaeocystis pouchetii* (Jacobsen et al. 1996). Viruses have also been implicated in the bloom termination of *Emiliania huxleyi* (Bratbak et al. 1993, Bratbak et al. 1996).

Two viruses have been isolated and described which infect *Heterosigma akashiwo*. In Japan, a large, 200 nm *H. akashiwo* virus (HaV) was isolated (Nagasaki and Yamaguchi 1997). This virus is very similar to viruses found replicating in cells isolated during the termination of a *H. akashiwo* bloom (Nagasaki 1994a, b). In British Columbia, a small, 30 nm virus infecting *Heterosigma akashiwo* was isolated from the Strait of Georgia (Lawrence et al. 2001). This virus, named HaNIV for *Heterosigma akashiwo* nuclear inclusion virus, differs from HaV. In addition to size, HaNIV forms crystalline arrays as it assembles inside the nucleus, as opposed to the cytoplasmic assembly of HaV (Lawrence et al. 2001). These significant differences indicate that HaNIV is unrelated to HaV. Other studies have shown that isolates of *H. akashiwo* viruses vary in the range of host strains that they infect (Nagasaki and Yamaguchi 1998, Tarutani et al. 2000). Together, this evidence indicates diversity among *H. akashiwo* viruses, yet other than TEM studies and studies investigating the impact of these viruses on the host, there is little information about the viruses themselves and their ecological role, particularly in bloom dynamics.
This thesis examines two unrelated viruses isolated from the Strait of Georgia, British Columbia that cause lysis of *Heterosigma akashiwo*. In contrast to previous studies that examined the effects of *H. akashiwo* viruses on their host, the viruses themselves are the targets of investigation. Chapter 2 describes the characterization of the two viruses, HaRNAV and HaNIV. The characterization of HaRNAV has resulted in the discovery of the first single-stranded RNA (ssRNA) phytoplankton virus. The name HaRNAV, meaning *Heterosigma akashiwo* RNA virus, is derived from the nature of its genome. In addition, the characterization of HaNIV demonstrates the diversity of double-stranded DNA (dsDNA) algal viruses. Chapter 3 provides the genome sequence of HaNIV. The genome sequence was not completed, but is presented as two contigs, 17 780 and 15 858 base pairs in length. These sequences represent over 90 % of the genome. The sequence has revealed open reading frames from HaNIV that are homologous with genes from diverse bacterial viruses, thus providing evidence of horizontal gene transfer. It was also demonstrated that HaNIV is latent, *i.e.* the virus does not replicate immediately upon infection of its host, but its genome is integrated into the host's DNA or carried extrachromosomally.

The characterization of *Heterosigma akashiwo* viruses will hopefully result in a better understanding of the factors regulating *H. akashiwo* blooms. Virus characterization can lead to the development of new tools and techniques for monitoring the effects of viruses in the natural environment. In addition, this thesis has demonstrated the diversity of viruses in marine environments and the need for further exploration of marine microbes.
LITERATURE CITED


CHAPTER 2. Characterization of HaRNAV and HaNIV, novel single-stranded
RNA and double-stranded DNA viruses causing lysis of

*Heterosigma akashiwo*

**INTRODUCTION**

Viruses generally consist of a nucleic acid genome of either DNA or RNA
capsidated by a protein shell. Viruses are unable to perform any metabolic functions
on their own. Thus, in order to express genes and replicate, they must infect a host
organism and take over the cellular machinery. Although simple in structure, these
entities can cause illness and sometimes death of their host. For unicellular organisms,
such as bacteria and phytoplankton, a lytic viral infection, in which progeny viruses are
produced and released by cell lysis, always results in cell death.

There are greater than 4000 known virus species (Murphy et al. 1995) infecting a
diversity of organisms including archaeabacteria, bacteria, fungi, plants, insects, and
animals. A virus species is defined as “a polythetic [based on multiple criteria] class of
viruses that constitutes a replicating lineage and occupies a particular ecological niche”
(Mayo 1996). The actual number of viruses is undoubtedly much higher and it is
generally believed that there are several viruses specific for every species of living
organism (Fenner 1974, Fuhrman 1999). Viruses have been observed in almost every
group of algae, but few of these viruses have been isolated and characterized.
Viruses can be distinguished from one another based on several physical, antigenic, and biological properties. Primarily, viruses are characterized by the morphology of the viral particle and the nature of the viral genome (Fenner 1974). Other characteristics used to describe viruses include the mass, density, and stability of the virion, the number and size of structural and non-structural proteins, the presence of lipids or carbohydrates, genome organization, the site of replication, assembly, and maturation, host range, mode of transmission, and geographic distribution (Murphy 1996). These characteristics are used as a basis for virus classification.

The classification of viruses follows a hierarchical scheme and is governed by the International Committee for the Taxonomy of Viruses (ICTV). Virus species are grouped into genera, families, and occasionally orders based on certain common characters such as biochemical composition, replication strategy, and nature of the particle structure. This classification scheme was used initially as an organizational system and, unlike the classification of organisms, was not meant to imply any phylogenetic relationships. However, information based on genome sequence data and similarities in gene order and arrangement has provided the means for phylogenetic analyses. Three orders of viruses have recently been established to reflect phylogeny. They are the Caudovirales - the tailed bacteriophage group (Maniloff and Ackermann 1998), the Mononegavirales - a negative stranded ssRNA virus group (Pringle 1991), and the Nidovirales - a positive stranded ssRNA virus group (Cavanagh 1997).

Viral isolates that infect phytoplankton are typically large (100-200 nm) double-stranded DNA viruses. A number of these isolates, infecting Micromonas pusilla (Mayer and Taylor 1979, Waters and Chan 1982, Cottrell and Suttle 1991), Chrysochromulina
spp. (Suttle and Chan 1995), and the brown alga *Ectocarpus siliculosis* (Müller et al. 1990), have been included in the virus family *Phycodnaviridae* (Pringle 1999). This family was established based on viruses that infect an exsymbiotic *Chlorella*-like alga (van Etten 1995, van Etten and Meints 1999). A DNA polymerase gene-based phylogeny suggests that the *Phycodnaviridae* is monophyletic and most closely related to herpes viruses (Chen and Suttle 1996). However, the *Phycodnaviridae* shares structural and genomic features with the iridoviruses, vaccinia virus (a pox virus), and the unclassified African Swine Fever virus (van Etten and Meints 1999). Other viruses infecting eukaryotic phytoplankton, including *Aureococcus anophagefferens* (Gastrich et al. 1998), *Emiliania huxleyi* (Bratbak et al. 1993, Bratbak et al. 1996), *Heterosigma akashiwo* (Nagasaki and Yamaguchi 1997), and *Phaeocystis pouchetii* (Jacobsen et al. 1996), have not been assigned to a family but are morphologically similar to the *Phycodnaviridae*.

In addition to *Phycodnaviridae*-like particles, other types of viruses have been reported which infect microalgae. Milligan and Cosper (1994) and Garry et al. (1998) described a phage-like particle that infects *Aureococcus anophagefferens*, although some concern has been expressed that this may be a contaminating bacteriophage (virus infecting bacteria) (Suttle 2000). More recently, Lawrence et al. (2001) have described a virus (HaNIV) that infects *Heterosigma akashiwo* which is much smaller (30 nm) than viruses belonging to the *Phycodnaviridae* and forms crystalline arrays in the nucleus. This chapter provides further characterization of HaNIV in terms of morphology, structure, and genome type. As well, a new virus, *Heterosigma akashiwo* RNA virus (HaRNAV), is described. It is the first single-stranded (ss) RNA virus reported to cause lysis of phytoplankton. The characterization of HaNIV and HaRNAV provided
information to compare the two novel viruses with known eukaryotic algal viruses and other virus families.

MATERIALS AND METHODS

Host cells. *Heterosigma akashiwo* (NEPCC 522), obtained from the Northeast Pacific Culture Collection at the University of British Columbia, was used for the isolation and amplification of the viruses. This strain was treated with 500 μg·ml⁻¹ ampicillin, 500 μg·ml⁻¹ kanamycin, and 100 μg·ml⁻¹ gentamicin or 500 μg·ml⁻¹ ampicillin, 250 μg·ml⁻¹ streptomycin, and 500 μg·ml⁻¹ kanamycin in an attempt to render the cultures axenic (Cottrell and Suttle 1993). Only the latter treatment was successful.

To test the host range of the viruses, fourteen additional strains of *Heterosigma akashiwo* (Table 2.1) were tested as potential hosts. In addition to the 12 strains used by Lawrence et al. (2001), two strains, H93616 and H94608 (Nagasaki and Yamaguchi 1998), were tested in this study.

Virus isolates. Viruses were isolated and cloned as described previously (Suttle et al. 1991). Briefly, 200 L water samples were collected and filtered to remove plankton and most bacteria, but leaving the virus-size fraction. The filtered sea-water was concentrated using an Amicon spiral wound ultrafiltration cartridge with a 30 kDa cut-off. The concentrated virus-size fraction was added in a dilution series to exponentially-growing cultures of *Heterosigma akashiwo*. Samples positive for lytic agents of *H. akashiwo* caused cell lysis or a decrease in cell growth. Clonal isolates from these positive samples were obtained by repeatedly diluting the lytic agents to extinction.
Inoculating an exponentially growing culture of *H. akashiwo* with the end-point dilution resulted in the replication and amplification of a single lytic agent of *H. akashiwo*. The amplified lytic agents (i.e. cell lysates) were filtered through a 0.22 μm pore-size GV Durapore® filter (Millipore) and stored at 4 °C.

HaNIV was isolated in 1997 from water collected in Discovery Passage, BC, Canada (Lawrence et al. 2001). HaRNAV was isolated from the Fraser River Plume (49° 04.749'N X 123° 23.076'W) in the Strait of Georgia, BC on August 22, 1996. The isolate originated from a 200 L water sample collected at the chlorophyll maximum (depth of 6.7 m).

**Growth conditions.** Experiments for initial TEM characterization, chloroform sensitivity assay, and host range study were conducted with 5 ml cultures of exponentially growing *Heterosigma akashiwo*. The experiments were performed in triplicate, with the exception of the chloroform sensitivity assay, which was done in replicates of eight. All cultures were grown in F/2 enriched seawater (Guillard 1975) (30%/o) supplemented with 10 nM sodium selenite. The cells were incubated at 20 °C with a light:dark cycle of 14:10 using 260 or 200 μmol·m⁻²·s⁻¹ photosynthetically active radiation. Daily measurements of *in vivo* chlorophyll fluorescence (Turner Designs fluorometer) were used as a measure of cell growth. In each assay 1 % or 0.1 % (v/v) of a 0.22 μm filtered cell lysate containing the viruses was added to the host cell culture. A 90 % decrease in fluorescence compared to control cultures was considered an indicator of cell lysis.

**Amplification and purification of viruses.** Ten to fifteen 800 ml or 2 L cultures of *Heterosigma akashiwo* were grown exponentially in 1 L Erlenmeyer or 2.5 L low form
culture flasks (Fisher), respectively. Growth was monitored by *in vivo* fluorescence of subsamples of the cultures. After 5-7 days, 0.1 % (v/v) of a 0.22 μm filtered cell lysate was added to each flask and incubated until cell lysis occurred within 7 days. The lysates from each culture were pooled, filtered, ultrafiltered (30 kD cut-off), and concentrated as described by Cottrell and Suttle (1991) with the following modifications. The retentate following ultrafiltration was centrifuged at 4100 X g for 30 min to remove remaining bacteria and the supernatant centrifuged for 5 h at 108 000 X g to pellet the viruses. The virus pellet was softened and resuspended in approximately 75 μl of the supernatant. The resuspended pellets were pooled and incubated with 1 U·μl⁻¹ DNase I and 0.1 U·μl⁻¹ RNase A for 3 h at room temperature to remove nucleic acids not contained within virus particles.

Sucrose gradients were used to purify the virus particles. For HaNIV, sucrose step gradients (10, 20, 30, and 40 %) were prepared in F/2 + Se medium and left to linearize overnight at 4 °C. Sucrose gradients from 5 to 40 % were made up in 50 mM Tris (pH 7.6) for HaRNAV. The virus suspensions were layered over the linear sucrose gradients (200 - 300 μl of viruses per gradient) and centrifuged at 50 000 X g in an SW-40 Beckman rotor for 1 h to purify HaNIV or 3 h for HaRNAV. Bands were visualized by shining a light directly over the sucrose gradient and removed by inserting a 23-gauge needle with a syringe into the tube just below the band. For HaNIV, there was a single band at 3.3 cm from the top, whereas the HaRNAV preparation resulted in bands at 2.2 and 4.0 cm. The purified viruses (~ 1 ml) were placed in a 1.5 ml screw-capped microcentrifuge tube. The tube was topped up with 0.22 μm filtered F/2 + Se or 50 mM Tris (pH 7.6), floated in a 14 ml centrifuge tube filled with water, spun for 3 h at 115 000
X g (SW-40 rotor), the supernatant removed, and the procedure was repeated. The virus pellets were softened with 50 μl of supernatant at 4 °C and resuspended in 100 μl media or 50 mM Tris.

To confirm that a lytic agent of *Heterosigma akashiwo* was purified, the viruses were diluted 1000-fold and added back to exponentially growing cultures of *H. akashiwo*.

**Transmission electron microscopy.** For initial characterization by TEM, 0.22 μm filtered lysates were spun for 3 h at 115 000 X g onto carbon and formvar coated, glow-discharged, 400-mesh copper TEM grids. The grids were stained with 1 % uranyl acetate (UA) (w/v) for 20 sec and excess stain was wicked off with filter paper.

Purified virus particles provided a cleaner sample with less debris for TEM characterization. In addition, the purification of the particles was confirmed by comparing the morphology of these particles to particles from cell lysates before and after virus purification. The purified virus particles were prepared for TEM by placing 10 μl of the purified virus sample on top of a TEM grid. The viruses were left to adsorb to the grid for 30 min at room temperature. The virus drop was removed with filter paper and the grid was stained with either 1 % UA or 2 % phosphotungstic acid (PTA) (pH 5.5) as described above.

The grids were viewed with a Zeiss 10C-transmission electron microscope at an accelerating voltage of 80 kV. Sizes of virus particles were determined from negatives using a scale lupe calibrated with a 2160 waffle-grating replica.

**Chloroform sensitivity.** An aliquot (500 μl) of 0.22 μm filtered lysate was vigorously shaken with an equal volume of chloroform for 5 min. Chloroform was removed by centrifugation at 4100 X g for 5 min and the aqueous phase containing the
viruses was recovered and left for 6 h at room temperature to let any remaining chloroform evaporate. The same procedure was completed with 500 μl of media as a control. The chloroform-treated viruses, chloroform-treated media, and non-treated viruses were added to exponentially growing *Heterosigma akashiwo* cells as described above.

**Structural proteins.** Purified virus particles (3-10 μl) were heated for 5 min at 100 °C. A 12 % SDS/polyacrylamide separating gel with a 4 % stacking gel was prepared following the protocol of Ausubel et al. (1992). The samples were electrophoresed at 10 mA through the stacking gel and then at 15 mA through the separating gel. Tobacco mosaic virus (TMV) was used as a control. The gel was fixed with 50 % v/v methanol and 10 % v/v acetic acid and stained overnight with 0.5 % Coomassie Blue. Protein molecular weight standards (Gibco BRL) ranging from 14 300 to 200 000 MW were used for size calibration and the gel was analyzed with Gel-Pro analyzer software (version 3.1.00.00 for Windows 95.NT, Media Cybernetics).

**Nucleic acid extraction and analysis.** Purified HaNIV (500 μl) was incubated for 15 min at 68 °C with 0.1 % (w/v) SDS and 5 mM EDTA. The lysed particles were extracted once with an equal volume of phenol, twice with an equal volume of phenol: chloroform (1:1), and once with an equal volume of chloroform. The nucleic acids were precipitated with 2 volumes of 100 % ethanol and incubated at -20 °C overnight. This preparation was spun for 25 min at 17 500 X g, the supernatent removed, and the nucleic acid pellet washed with 70 % ethanol. The pellet was air-dried, resuspended in water, and stored at -20 °C.
Nucleic acids from HaRNAV were obtained using the QIAamp Viral RNA kit (Qiagen) following the manufacturer's protocol and stored at -20 °C.

Nucleic acid concentration was quantified by absorbance at 260 nm. Nucleic acid type was determined by digesting the viral nucleic acids with 0.07 µg·µl⁻¹ DNase I in 0.1X SSC (1X = 0.15 M NaCl and 0.015 M Na₃citrate-2H₂O, pH 7.0), 0.07 µg·µl⁻¹ RNase A in 0.01X SSC, and 0.07 µg·µl⁻¹ RNase A in 2X SSC. Double stranded (ds) RNA is not digested by RNase A under high salt conditions (Semancik et al. 1973). Bacteriophage λ dsDNA, bacteriophage φ6 dsRNA, and TMV ssRNA were used as controls. The nucleic acids were electrophoresed through a 0.75 % agarose gel, stained with 0.5 µL·µl⁻¹ ethidium bromide, and visualized with a UV transilluminator (Fisher).

The genome size of HaNIV was estimated by digesting 0.7 µg of genomic DNA with the restriction enzymes Hind III (1 U·µl⁻¹) and Hinc II (2 U·µl⁻¹) at 37 °C for 3 h using the manufacturer's (New England Biolabs) recommended buffers. The digested DNA was electrophoresed and visualized as described above. λ DNA digested with Hind III (Gibco BRL) was used to calibrate the size of the digested fragments and the size of the bands was determined using Gel-Pro Analyzer software.

For HaRNAV, the whole genomic RNA was run on a 0.75 % denaturing formaldehyde gel alongside a 0.24-9.5 kb RNA ladder (Gibco BRL), bacteriophage φ6 dsRNA genome, and TMV ssRNA genome. The RNA samples and formaldehyde gel were prepared and run following standard protocols (Sambrook et al. 1989).

The genomic RNA from HaRNAV particles was analyzed to determine if the genome was poly-A tailed. HaRNAV RNA was run through an Oligotex™ mRNA column (Qiagen) following the manufacturer's protocols. These columns isolate poly-A
tailed RNAs. All other nucleic acids are washed through the column. The poly-A tailed RNAs are then eluted from the column with water. Nucleic acids that were washed through the column were re-isolated using the QIAamp Viral RNA kit (Qiagen) following the manufacturer's protocol. All nucleic acids were examined on a 0.75 % formaldehyde gel alongside RNA size standards as described above.

**Hybridization assay.** Exponentially growing *Heterosigma akashiwo* cultures were infected with 1 % (v/v) of an HaNIV 0.22 filtered lysate and harvested 6 days post infection, before the cells lysed. Cells from infected and non-infected cultures were pelleted by centrifugation (1500 X g for 10 min), resuspended with media, and centrifuged again. The cells were resuspended in 400 μl of water and lysed in a solution containing 0.1 M Tris, pH 8, 0.05 M EDTA, 0.2 M NaCl, and 1 % SDS (Saunders 1993). DNA was extracted using a standard phenol and chloroform: isoamyl alcohol protocol, similar to the one described for HaNIV, precipitated with ethanol, and resuspended in water. Dilutions of DNA from infected and non-infected *H. akashiwo*, ranging from 30 ng to 10 pg, were dot-blotted (VacuDot-VS, American Bionetics) onto a nylon membrane. HaNIV DNA and *H. akashiwo* DNA from a strain resistant to HaNIV (NEPCC 522 treated with ampicillin, streptomycin, and kanamycin) were also blotted onto the membrane as positive and negative controls, respectively. The membrane was dried for 2 h at 80 °C and the DNA was cross-linked with UV light (UV Crosslinker, Fisher) to the membrane.

A biotinylated probe was constructed from HaNIV DNA by nick translation (BioNick™ Labeling System) following the manufacturer's protocol (Gibco\BRL). Hybridization was conducted at 42 °C overnight in 35 X 300 mm tubes (Robbins
Scientific). The hybridization solution contained 25% formamide, 6X SSPE (1.08 M NaCl, 60 mM NaH₂PO₄, 60 mM EDTA, pH 7.4), 5X Denhardt's solution (0.1% w/v Ficoll, 0.1% w/v polyvinylpyrrolidone, 0.1% w/v BSA), 1% SDS, and 100 ng·ml⁻¹ probe. The membrane was washed and the probe was detected by chemiluminescence using the PhotoGene Nucleic Acid Detection System (Gibco/BRL) following the manufacturer's protocols. The only modification to the protocol was that the initial wash was conducted at 50 °C rather than 65 °C. The signal was visualized by exposing the membrane to BioMax Light film (Kodak) for 30 min.

RESULTS

Based on TEM observations, purified HaNIV particles were indistinguishable from those found in cell lysates before purification. HaNIV is 50 nm in diameter and hexagonal in shape (Fig. 2.1A) suggesting icosahedral symmetry. The HaNIV particles have densely staining circular cores and an extension at one vertex of the capsid that extends 13 nm from the edge of the core. The five other vertices are 7 nm away from the edge of the core. The extension was not visible in all HaNIV particles, likely depending on the orientation of the particle. Enveloped particles were not observed. The purified particles were able to cause lysis of exponentially growing cultures of Heterosigma akashiwo and virus-like particles (VLPs) of the same morphology were recovered from the cell lysate (data not shown).

For HaRNAV, the virion is 25 nm in diameter and polygonal in shape (Fig. 2.1B). The three dimensional particle is probably icosahedral, but appeared more spherical in
shape than HaNIV. The particles did not appear to be enveloped. The particles were negatively stained, but the stain also completely penetrated some particles suggesting that these particles may be "empty" and devoid of nucleic acids (Hemida and Murant 1989, Hemida et al. 1989). From the sucrose gradient, bands at 2.2 and 4.0 cm contained morphologically similar VLPs (data not shown). However, only particles from the lower band consistently caused lysis of *Heterosigma akashiwo* (data not shown) suggesting that viruses in the top band were incompletely assembled. Purified particles, VLPs from cells infected with purified VLPs, and VLPs from cell lysates before purification were indistinguishable by TEM.

Figure 2.1. Electron micrographs of virus particles causing lysis of *Heterosigma akashiwo*. A) HaNIV particle negatively stained with UA. B) HaRNAV from the lower band in the sucrose gradient, stained with PTA. Scale bars = 50 nm.
Chloroform sensitivity

After inoculation with chloroform-treated HaNIV and HaRNAV, *H. akashiwo* cultures continued to increase in fluorescence while cultures inoculated with untreated viruses rapidly decreased in fluorescence (Fig. 2.2) indicating that both HaRNAV and HaNIV are chloroform sensitive.

Figure 2.2. Growth of *Heterosigma akashiwo* in the presence of (A) HaNIV and (B) HaRNAV with and without chloroform. For both graphs, open symbols indicate chloroform treatments with the appropriate virus added (○) and without (△). Solid symbols indicate control treatments (no chloroform) with virus added (●) and without (▲). The arrow indicates when viruses were added to the cell cultures. In each case, values are the mean ± SD, n = 8.
Structural proteins

The size and number of structural proteins in the virus particles were determined by polyacrylamide gel electrophoresis (PAGE). HaNIV particles contained at least two structural proteins (Fig. 2.3, lane 1). A 38.0 kDa protein was stained more intensely than a second 20.5 kDa protein.

HaRNAV contains at least five proteins, 33.9, 29.0, 26.1, 24.6, and 24.0 kDa (Fig. 2.3, lane 3). Virus particles collected from the upper band in the sucrose gradient had an additional 39.0 kDa protein (Fig. 2.3, lane 2).

Figure 2.3. Structural proteins of Heterosigma akashiwo viruses. Proteins obtained from purified HaNIV (lane 1) and HaRNAV from the upper band (lane 2) and lower band (lane 3) of the sucrose gradient. Molecular weight markers (lane 4). The single 18 kDa capsid protein from tobacco mosaic virus was used as a control (lane 5). The proteins were stained with Coomassie Blue.
Nucleic acid type and genome size

Nucleic acids from HaNIV were not digested by RNase A, but were digested with DNase I (results not shown) indicating that the HaNIV genome is double-stranded DNA. The HaNIV genome was digested into 12 and 15 fragments by Hind III and Hinc II, respectively (Fig. 2.4). After estimating the size of the fragments, the genome of HaNIV was determined to be approximately 37,000 base pairs (bp) based on the Hind III digest and 40,000 bp based on Hinc II. Averaging these estimates results in an approximate genome size of 38,500 bp.

Figure 2.4. Restriction enzyme digests of HaNIV DNA. HaNIV DNA was digested with Hind III (lane 1) and Hinc II (lane 2). Bacteriophage λ DNA digested with Hind III (lane 3) was used as a molecular weight marker.
Nucleic acids from HaRNAV were digested by RNase A in 0.01X and 2X SSC, but not with DNase A (results not shown). This result indicates that the genome of HaRNAV is single-stranded RNA. Based on gel electrophoresis, the length of the HaRNAV genome was estimated to be approximately 9100 nucleotides (Fig. 2.5). The genomic RNA was not retained by the poly-A tail purification column, but was recovered from the wash (data not shown). Therefore, the HaRNAV genome does not contain a poly-A tail.

Figure 2.5. Genomic RNA isolated from HaRNAV particles (lane 3). dsRNA from bacteriophage φ6 (lane 1) and ssRNA from TMV (lane 2) were used as controls. RNA standards were run in lane 4.
Hybridization

HaNIV DNA hybridized to *Heterosigma akashiwo* DNA from infected cells, but not to DNA from non-infected cells (Fig. 2.6).

![Hybridization Gel](image)

Figure 2.6. Hybridization of HaNIV DNA with DNA from A) non-infected *Heterosigma akashiwo*, (B) HaNIV infected *H. akashiwo*, (C) a resistant strain of *H. akashiwo*, and (D) unlabeled HaNIV. DNA was blotted onto the membrane in serial dilutions.
Host range

Host range results of HaNIV and HaRNAV are reported in Table 2.1. HaNIV caused lysis of 6 of the 7 strains of *Heterosigma akashiwo* from the Northeast Pacific, the resistant strain being NWFSC 503 (Table 2.1). HaNIV was unable to cause lysis of the strains from the Japanese or Northwest Atlantic coasts.

HaRNAV caused lysis of 2 strains of *Heterosigma akashiwo* (H93616, H94608) from Japan and 3 strains (NEPCC 102, NEPCC 522, NEPCC 764) from the Northeast Pacific (Table 2.1). Japanese strains NIES 4 and NIES 293, the Northeast Pacific strains NEPCC 278, NEPCC 625, NWFSC 500, and NWFSC 503, as well as all strains from the Northwest Atlantic coast were resistant to HaRNAV.

**DISCUSSION**

A number of significant findings are reported in this study. Demonstrated, for the first time, is the existence of a single-stranded (ss) RNA virus (HaRNAV) that infects a microalga. As well, the partial characterization of this virus and a novel double-stranded (ds) DNA virus (HaNIV) has been provided. The characterization of HaRNAV and HaNIV describes two types of viruses that have never before been reported for algal species. Both these viruses infect *Heterosigma akashiwo*, an ecologically and economically important toxic bloom-forming alga. Therefore, there are at least 3 different viruses that infect and cause lysis of this organism. These results and their implications are discussed in detail below.
Table 2.1. Sensitivity of various *Heterosigma akashiwo* strains to HaNIV and HaRNAV. + indicates the virus caused cell lysis, - indicates no lysis compared to a control culture, * indicates a discrepancy from the result reported by Lawrence et al. (2001). NEPCC 522 was the strain used for virus isolation.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Origin of Isolation</th>
<th>HaNIV</th>
<th>HaRNAV</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Northwest Pacific</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEPCC 102</td>
<td>English Bay, BC, Canada</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>NEPCC 278</td>
<td>Marietta, WA, USA</td>
<td>*<em>+</em></td>
<td>-</td>
</tr>
<tr>
<td>NEPCC 522</td>
<td>Jericho Beach, BC, Canada</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>NEPCC 625</td>
<td>Genoa Bay, BC, Canada</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>NEPCC 764</td>
<td>Kyoquot Sound, BC, Canada</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>NWFSC 500</td>
<td>Kalalock, WA, USA</td>
<td>*<em>+</em></td>
<td>-</td>
</tr>
<tr>
<td>NWFSC 503</td>
<td>Manchester, WA, USA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Atlantic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEPCC 560</td>
<td>Long Island Sound, NY, USA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NEPCC 630</td>
<td>Long Island Sound, NY, USA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CCMP 525</td>
<td>Long Island Sound, NY, USA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>UTEX LB 2005</td>
<td>York River, Gloucester Point, VA, USA</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Japanese</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIES 4 (type strain)</td>
<td>Fukuyama Bay, Hiroshima, Japan</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NIES 293</td>
<td>Onagawa Bay, Miyagi, Japan</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>H93616</td>
<td>Hiroshima Bay, Japan</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>H94608</td>
<td>Hiroshima Bay, Japan</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>
HaRNAV. HaRNAV is the first ssRNA virus found to infect phytoplankton. Another ssRNA virus infects the green alga *Chara corallina* (Gibbs et al. 1975, Skotnicki et al. 1976), however, it is rod shaped with similarities to the furoviruses (van Etten et al. 1991). HaRNAV is polyhedral and is more similar in size and shape to plant viruses belonging to the families Bromoviridae, Comoviridae, Sequiviridae, and Tombusviridae, the invertebrate virus families Nodaviridae and Tetraviridae, the vertebrate virus family Picornaviridae, and the bacteria viruses of the Leviridae (Murphy et al. 1995). However, except for the Sequiviridae and Picornaviridae, HaRNAV differs from these virus families in genome size and the number of structural proteins. Compared to HaRNAV, which has a single 9100 nucleotide genome and at least 5 structural proteins, these virus families typically have a smaller (5.5 kb for Tetraviridae, 4-4.7 kb for Tombusviridae, 3.5-4.3 kb for Leviridae) or segmented (Bromoviridae, Comoviridae, Nodaviridae) genomes and have one (Bromoviridae, Leviridae, Tetraviridae, Tombusviridae), two (Comoviridae), or three (Nodaviridae, Tetraviridae) capsid proteins (Murphy et al. 1995).

HaRNAV is most similar to the Sequiviridae. Viruses belonging to the Sequiviridae are 30 nm in diameter, have a genome size ranging from 9-12 kb, and have at least 3 capsid proteins ranging from 22.5 to 34 kDa (Hemida and Murant 1989, Hemida et al. 1989, Reavy et al. 1993). In addition, like HaRNAV, the genome of some sequiviruses is not polyadenylated (Turnbull-Ross et al. 1992, Mayo et al. 1995). It is not known if sequiviruses are sensitive to organic solvents like HaRNAV, but lipids have not been reported for these viruses. Sensitivity to chloroform and other organic solvents is an indicator of a lipid component in the structure of the virus (Fenner et al. 1974).
Viruses belonging to the families *Flaviviridae* and *Togaviridae* are sensitive to organic solvents and have lipid envelopes in addition to having a similar genome size to HaRNAV (Murphy et al. 1995). These viruses, however, are larger (40-60 nm for *Flaviviridae* and 70 nm for *Togaviridae*) than HaRNAV. Although HaRNAV was sensitive to chloroform, a lipid envelope was not obvious in TEM observations of HaRNAV particles (Fig. 2.1B). The lipid component may reside inside the protein shell, like viruses belonging to the *Phycodnaviridae* (van Etten and Meints 1999). Viruses of the dsDNA family *Corticoviridae* are also sensitive to organic solvents, but do not have an envelope. Instead, a lipid bilayer forms between the inner and outer shell of the virus particles (Silbert et al. 1969, Ackermann 1995).

The double-band pattern in the sucrose gradients of HaRNAV has been observed in other systems, including anthriscus yellows virus (AYV) belonging to the *Sequiviridae* (Hemida and Murant 1989, Hemida et al. 1989). The AYV particles from the upper zone of the sucrose gradient were shown to be "empty" protein shells. Our results suggest a similar explanation for the upper band of particles from HaRNAV. The structural similarity of the particles from the upper and lower bands, based on PAGE, indicates that they are particles of the same virus (Fig. 2.3). However, because particles from the upper band were non-infectious, consistently penetrated by the negative stain, and because nucleic acids could not be isolated from them (data not shown), these particles are considered to be incompletely assembled HaRNAV. The additional 39.0 kDa protein obtained from particles in the upper band may represent an immature protein that needs to be cleaved to complete the particle or may be a scaffolding protein.
Although HaRNAV shares many characteristics with the *Sequiviridae*, further analysis, especially of the genome organization, is necessary to determine the relation of HaRNAV to known families of ssRNA viruses. Compared to known algal viruses, however, HaRNAV is unique.

**HaNIV.** HaNIV, although a dsDNA virus, does not resemble previously characterised algal viruses including a *Phycodnaviridae*-like virus infecting *Heterosigma akashiwo* isolated from Japan (Nagasaki and Yamaguchi 1997). HaNIV is a polyhedral, 50 nm virus with a 38.5 kb dsDNA genome and at least 2 structural proteins. The virus is also known to assemble within the nucleus (Lawrence et al. 2001). In contrast, other dsDNA algal viruses isolated and characterised are 100-200 nm icosahedral viruses assembling in the cytoplasm with 100-380 kb genomes (van Etten and Meints 1999).

In addition to differing from known algal viruses, HaNIV is also unique among all known dsDNA viruses. Few icosahedral dsDNA viruses are smaller than 60 nm and possess a small number of structural proteins. Although it is common for RNA viruses to have one or two structural proteins, most icosahedral dsDNA viruses rarely contain fewer than 10 proteins.

The *Papovaviridae*, a family of viruses infecting vertebrates, is structurally similar to HaNIV. These viruses are icosahedral, 40-55 nm in diameter and contain 3 structural proteins (Frisque et al. 1995). As with HaNIV, these viruses also assemble in the nucleus and form crystalline arrays (Howatson 1973). However, because the genome of papovaviruses is only about 5 kb (Frisque et al. 1995) it is unlikely that HaNIV is related to the *Papovaviridae*. 
Viruses with a similar dsDNA genome size to HaNIV include vertebrate viruses belonging to the *Adenoviridae* and bacteriophages of the *Podoviridae*. Adenovirus particles are typically 80-110 nm in diameter with fibres protruding from the capsid and contain approximately 40 different proteins (Russell et al. 1995). These viruses are also known to assemble in the nucleus and form crystalline arrays of progeny virus particles or protein subunits (Horne 1973, Russell et al. 1995). Podoviruses are around 60 nm in diameter, contain 12 proteins, and have a small tail (Ackermann and Dubow 1995).

Despite having some similarities with known dsDNA virus families, HaNIV differs significantly from these families in either genome size or structure. In addition, HaNIV is chloroform sensitive whereas most small, icosahedral, dsDNA viruses are resistant to chloroform. Sequence analysis of the HaNIV genome is needed to further assess the relationship of HaNIV to other viruses.

The size of the free virus particles of HaNIV differs from the intracellular particles described by Lawrence et al. (2001). Intracellularly, the particles are ca. 30 nm in diameter and are organized into crystalline arrays. In contrast, extracellular HaNIV particles are 50 nm in diameter. Because of the size difference between the intracellular and extracellular particles, it was important to demonstrate that the purified virus particles were the same as those observed intracellularly. This was shown by hybridizing genomic DNA from purified HaNIV particles to DNA from infected *H. akashiwo* cells (Fig. 2.6).

The difference in size between intracellular and extracellular HaNIV is likely because the intracellular particles are probably not mature viruses. The particles could be empty capsids awaiting final completion, in which the particles would expand in size...
before they are released from the cell, as occurs with T4 phages (Black et al. 1994). The particles could also be excess viral proteins formed into paracrystalline masses, as observed with adenovirus infection (Henry et al. 1971, Russell et al. 1995).

**Host range and ecological significance.** The characterisation of HaNIV and HaRNAV demonstrates the diversity of viruses infecting microalgae. This diversity has significant implications for our understanding of virus-host interactions in the environment and the effect of these interactions in bloom dynamics and other ecological processes. To assess the interactions of HaNIV and HaRNAV with *Heterosigma akashiwo*, a host range study, which tests the specificity of the viruses' pathogenic effect, was conducted.

Similar to other viruses that infect *Heterosigma akashiwo* (Nagasaki and Yamaguchi 1998, Nagasaki et al. 1999, Tarutani et al. 2000) and other microalgae (Sahlsten 1998), HaNIV and HaRNAV are strain-specific. We tested the viruses against host strains from distant geographic areas (Japan, the Northeast Pacific, and the Northwest Atlantic) unlike previous host range studies (Nagasaki and Yamaguchi 1998, Nagasaki et al. 1999, Tarutani et al. 2000) that tested host strains from the same geographic area. Although HaNIV and HaRNAV are unrelated viruses, both were isolated from the Strait of Georgia, BC and mostly caused lysis of host strains from the Northeast Pacific, the same geographical location as their isolation. Both viruses were unable cause lysis of *H. akashiwo* strains originating from the Northwest Atlantic and HaNIV caused lysis of host strains from the Northeast Pacific only.

The host range of HaNIV was similar to that reported by Lawrence et al. (2001) except for two discrepancies (Table 2.1). In the present study, *Heterosigma akashiwo*
strains NWFSC 500 and NEPCC 278, both isolated from Washington State, lysed when challenged with HaNIV and the lytic effect was propagated. The reason for the differences in host range observed between the present study and that of Lawrence et al. (2001) are unknown.

HaRNAV caused lysis of fewer *Heterosigma akashiwo* strains from the Northeast Pacific than HaNIV, but was able to lyse two strains from Japan. Its pathogenic effect, however, was restricted to Pacific host strains.

Nagasaki et al. (1999) did not find a relationship between the location of virus isolation and the isolation location of susceptible host strains. In other words, the area of host or virus isolation was not a good predictor of host range. However, virus isolates were tested only against strains isolated from various areas of Japan; therefore, their result does not exclude the possibility of large-scale geographic patterns affecting host range.

On a smaller spatial scale, the variation in host specificity of HaNIV and HaRNAV demonstrates the complexity of virus-host interactions in the natural environment. Although HaNIV and HaRNAV were originally isolated off the same strain of *Heterosigma akashiwo*, each virus infected a different subset of host strains. Therefore, in a given geographical area with a diversity of host strains, multiple viruses likely co-exist, each with a different range of host specificity. This virus community may collectively determine the composition of host strains or the strains that are able to bloom in that area.

Tarutani et al. (2000) have shown that *H. akashiwo* cells differ in virus susceptibility as a bloom progresses. Cells isolated during the bloom were most
susceptible to viral lysis, while cells isolated as the bloom collapsed were more resistant. This result suggests that the composition of *H. akashiwo* strains changes throughout a bloom and the virus community may be regulating this change. As suggested in many other studies (Suttle et al. 1990, Waterbury and Valois 1993, Hennes et al. 1995, Fuhrman 1999, van Hannen et al. 1999, Wommack and Colwell 2000), viruses may be an important factor in determining aquatic microbial community composition, not only among species, but intra-specifically as well.

The discovery of HaNIV and HaRNAV demonstrates the diversity of the virus community in the Strait of Georgia, BC and the diversity of viruses infecting a single isolate of *Heterosigma akashiwo*. The characterization of these two novel *H. akashiwo* viruses has shown that these viruses are unrelated to each other, to a *H. akashiwo* virus isolated from Japan (Nagasaki et al. 1994a, b, Nagasaki and Yamaguchi 1997), or to other known algal viruses.

The diversity of *H. akashiwo* viruses and the variation in their host ranges suggests that in the natural environment the interaction of diverse viruses with diverse host strains is complex. In order to understand the role of viruses in the bloom dynamics of *H. akashiwo* or any other phytoplankton species, we must understand the composition of the virus community, the variation of the community in time and space, and the interaction of the individual viruses with their hosts. The characterisation of phytoplankton viruses will enable us to develop methods to study these properties of the virus community in the natural environment and unravel the complex ecological role of viruses in phytoplankton bloom dynamics.
LITERATURE CITED


CHAPTER 3. Genomic sequences from HaNIV show homology to bacteriophage genes

INTRODUCTION

Genome analysis using sequence data is an essential component of virus characterization. Sequence data provides information on genome organization, coding and regulatory regions, and the types of proteins and enzymes produced by the virus. The information is also invaluable for phylogenetic analyses, evolutionary studies, and understanding the process of viral infection and replication. Because viruses have relatively small genomes compared to other living organisms, many complete genome sequences have been determined and they are increasingly easier to obtain with our growing technological capabilities.

The generation of sequence data has greatly advanced our understanding of evolutionary relationships among viruses. Single gene phylogenies, often based on sequence data from polymerase or capsid genes, have been useful for inferring relationships among both DNA (for example, Chen and Suttle 1996, Knopf 1998, Tidona et al. 1998) and RNA viruses (for example, Goldbach and deHaan 1994, Ward 1993). Complete genome sequences, however, are an obviously better data set for evolutionary analysis. Rather than inferring the evolution of a virus based on one or two conserved genes, with a complete virus genome sequence, the evolution of the genome as a whole can be examined at the nucleotide level. The evolution of viruses can be assessed based
on genome organization, gene order, and the occurrence of gene modules in addition to individual gene sequences. Among the bacteriophages for example, gene order is more conserved than nucleotide or amino acid sequence (Ackermann 1998) and through comparative genomics the mosaic nature of phage genomes (Lucchini et al. 1999, Juhala et al. 2000, Mediavilla et al. 2000, vander Byl and Kropinski 2000, Desiere et al. 2001) is revealed. Regions of clear homology are interspersed with unrelated regions, therefore demonstrating that phages have likely evolved by extensive horizontal gene transfer and the exchange of gene cassettes from a common ancestral pool of genes (Hendrix et al. 1999).

Most complete genome sequences have been obtained from well-studied animal, plant, and bacteria viruses and few sequences have been reported for viruses from marine environments. The complete genome sequence has been obtained for a marine virus infecting a *Roseobacter* isolate (Roseophage SIO1) (Rohwer et al. 2000). Roseophage SIO1 was identified as a member of the *Podoviridae* phage family based on physical characteristics and also shared genetic similarities with the podoviruses T3 and T7. Other than this study, however, complete genome sequences have not been determined for marine viruses and only a few phylogenetic studies have been conducted (Chen and Suttle 1996, Lee et al. 1998).

As a contribution to the growing number of complete viral genome sequences and to further our understanding of marine and algal virus diversity and evolution, the genome sequence of HaNIV, a virus causing lysis of *Heterosigma akashiwo*, was determined. This is the first genome sequence reported for a virus with a phytoplankton host. Over 90% of the genome represented in two contigs was sequenced. The
annotation of these sequences, the function of putative open reading frames (ORFs), and the genetic relationships of HaNIV to other viruses are discussed in this chapter. The sequence information was also used to show that HaNIV might be a latent virus.

MATERIALS AND METHODS

HaNIV particles were amplified from cultures of *Heterosigma akashiwo* inoculated with the virus, concentrated by ultracentrifugation, and purified by 10 to 40 % sucrose gradients as described in the previous chapter. Genomic DNA from purified HaNIV particles was obtained, also described in Chapter 2, using a standard phenol:chloroform extraction method and precipitating the DNA with ethanol.

**Construction of HaNIV clone library.** The genomic DNA was digested with either the restriction enzymes Hind III or Hinc II (New England Biolabs). pGEM-3Zf(+) plasmid vectors (Promega) were also linearized by cutting with either of the same restriction enzymes. The ends of the linearized plasmid DNA were dephosphorylated with calf intestinal alkaline phosphatase (Gibco/BRL). The digested HaNIV DNA fragments were ligated with T4 DNA ligase (Gibco/BRL) into the linearized, dephosphorylated plasmid vectors. The recombinant plasmids were transformed into Epicurian Coli® XL1-Blue super competent cells (Stratagene) or Library efficiency® DH5α™ competent cells (Gibco/BRL). Transformants were selected by growing the cells on Luria-Bertani (LB) agar plates containing 100 μg·ml⁻¹ ampicillin and 100 μl of 2 % 5-bromo-4-chloro-3-inodyl-β-D-galactopyranoside (X-gal) and 20 μl of 20 mg·ml⁻¹ isopropyl-1-thio-β-D-galactopyrano-side (IPTG) spread onto the hardened plates.
Transformed cells with recombinant plasmids formed white colony forming units (CFUs) whereas cells containing plasmids without an insert formed blue CFUs. Individual white CFUs were picked for overnight cultures grown in LB liquid media containing 100 μg·ml⁻¹ ampicillin. Plasmid DNA was obtained from the overnight cultures by alkaline lysis following standard protocols (Sambrook et al. 1989) or with a QiaPrep® Spin Miniprep Kit (Qiagen).

The presence and size of the inserted DNA fragment in the plasmid DNA was confirmed by digesting the recombinant plasmid with the same restriction enzyme used to linearize the plasmid. Thus, the inserted fragment was cut out of the plasmid DNA. The digested DNA was run on a 0.8 % agarose gel, stained with ethidium bromide, and visualized with an UV transilluminator. The *E. coli* clones were identified based on the size of the fragment carried by the plasmid vector. DNA from plasmids with different sized inserts were chosen for sequencing.

**Sequencing and assembly of contigs.** Sequencing reactions were performed by thermal cycling using ABI PRISM BigDye Terminators (PE Applied Biosystems, Inc.). Following the reaction, excess dye terminators were removed with Centri-Sep columns (Princeton Separations Inc.) following the manufacturer's protocols. The sequence was obtained by automated sequencing using ABI 373 Stretch or ABI Prism 277 sequencers (Nucleic Acid Protein Services (NAPS) facility at the University of British Columbia). Initially, the inserted DNA fragment was sequenced from the plasmid DNA in forward and reverse directions by using sequencing primers (-21M13 and M13R from NAPS) complementary to opposite sides of the multiple cloning site on the pGEM plasmid. Subsequent sequencing of the insert was accomplished by designing sequencing primers
from the ends of the sequence information previously obtained (i.e. by primer walking). Sequences were analyzed and assembled into contiguous sequences (contigs) by eye using BioEdit, version 5.0.7 (Hall 1999, available from the web at http://www.mbio.ncsu.edu/BioEdit/bioedit.html).

Many sequences could not be aligned to one another following fragment cloning and sequencing because not all the fragments were cloned. To obtain overlapping sequence and to resolve gaps in the sequence, PCR primers were designed from the ends of the contigs. Using HaNIV genomic DNA as the template and different pairs of primers for PCR reactions, DNA was amplified when primers annealed to adjacent locations on the genome. The PCR products were cloned and sequenced or sequenced directly as described above.

**Sequence analysis.** Open reading frames (ORFs) were identified from the genomic sequence by translating in all six reading frames using any start codon, but ending at stop codons (BioEdit). Potential start codons (ATG or GTG) and ribosome binding sites (a series of As and Gs, 1-10 nucleotides preceding the start codon) were identified by eye. Translations greater than 75 amino acids were identified as putative ORFs. PSI(\(\psi\))-BLAST (Altschul et al. 1997, http://www.ncbi.nlm.nih.gov/BLAST) was used to search for possible homologies of HaNIV ORFs to known protein sequences in GenBank.

Phylogenetic trees were calculated for HaNIV ORFs homologous to lysozymes and the large subunit of terminase. HaNIV ORFs were aligned to protein sequences from bacterial phages, prophages, herpesviruses, and vertebrates obtained from GenBank (Table 3.1) using Clustal W (Thompson et al. 1994, available with BioEdit).
Table 3.1. GenBank accession numbers for protein sequences used for phylogenetic analyses. * indicates that the assigned function for the protein is putative.

<table>
<thead>
<tr>
<th>Virus or species</th>
<th>LSU terminase</th>
<th>Lysozyme</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Myoviridae</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enterobacteriophage T4</td>
<td>NP_049776</td>
<td>NP_049736</td>
</tr>
<tr>
<td>Enterobacteriophage P1</td>
<td>P27753</td>
<td>CAA61013</td>
</tr>
<tr>
<td>Enterobacteriophage P2</td>
<td>NP_046758</td>
<td>NP_046765</td>
</tr>
<tr>
<td>Enterobacteriophage Mu (μ)</td>
<td></td>
<td>AAF01099</td>
</tr>
<tr>
<td><strong>Podoviridae</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enterobacteriophage T3</td>
<td>P10310</td>
<td>P20331</td>
</tr>
<tr>
<td>Enterobacteriophage T7</td>
<td>P03694</td>
<td>P00806</td>
</tr>
<tr>
<td>Enterobacteriophage P22</td>
<td>AAF75044</td>
<td>AAF75040</td>
</tr>
<tr>
<td>Bacteriophage PS34</td>
<td>AAF09706 *</td>
<td></td>
</tr>
<tr>
<td>Bacteriophage PS119</td>
<td>AAF09710</td>
<td></td>
</tr>
<tr>
<td><strong>Siphoviridae</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacteriophage lambda (λ)</td>
<td>NP_040581</td>
<td>NP_040645</td>
</tr>
<tr>
<td>Bacteriophage 21</td>
<td>AAA32340</td>
<td>AAA32350</td>
</tr>
<tr>
<td>Bacteriophage HK022</td>
<td>AAF30354 *</td>
<td>NP_037696</td>
</tr>
<tr>
<td>Bacteriophage HK97</td>
<td>AAF31098 *</td>
<td>AAF31145</td>
</tr>
<tr>
<td>Bacteriophage HK620</td>
<td>AAK28891 *</td>
<td>NP_112069 *</td>
</tr>
<tr>
<td><strong>Unclassified</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacteriophage APSE-1</td>
<td>NP_050979 *</td>
<td>AAF03956 *</td>
</tr>
<tr>
<td><strong>Prophages</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Xyella fastidiosa (1)</td>
<td>AAF83318 *</td>
<td></td>
</tr>
<tr>
<td>In Xyella fastidiosa (2)</td>
<td>AAF84373 *</td>
<td></td>
</tr>
<tr>
<td>In Xyella fastidiosa (3)</td>
<td>AAF85113 *</td>
<td></td>
</tr>
<tr>
<td>In Escherichia coli (1)</td>
<td>BAB38385 *</td>
<td></td>
</tr>
<tr>
<td>CP-933V in Escherichia coli</td>
<td>AAG57223 *</td>
<td></td>
</tr>
<tr>
<td>DLP12 in Escherichia coli</td>
<td>P78285 *</td>
<td></td>
</tr>
<tr>
<td>In Haemophilus influenzae</td>
<td>P44184 *</td>
<td></td>
</tr>
<tr>
<td><strong>Herpesviruses</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HSV11</td>
<td>P04295</td>
<td></td>
</tr>
<tr>
<td>HSV7</td>
<td>AAC40780</td>
<td></td>
</tr>
<tr>
<td><strong>Vertebrates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gallus gallus (chicken)</td>
<td></td>
<td>1LYZ</td>
</tr>
<tr>
<td>Anser anser (goose)</td>
<td></td>
<td>P00718</td>
</tr>
</tbody>
</table>
alignments were replicated 100 times by bootstrapping and analyzed using the Neighbor-joining and Parsimony algorithms found in Phylip (Phylogenetic Inference Package; version 3.5c; distributed by J. Felsenstein at http://evolution.genetics.washington.edu/phylip.html). Neighbor-joining trees were calculated from a distance matrix obtained with the protein distance algorithm of Phylip. Trees were viewed using TreeView, version 1.6.5 (Page 1996, http://taxonomy.zoology.gla.ac.uk/rod/treeview.html).

**Detection of the HaNIV genome in the host, *Heterosigma akashiwo***. DNA was obtained from *Heterosigma akashiwo* cells infected and non-infected with HaNIV as described in the previous chapter. DNA from a strain of *H. akashiwo* resistant to HaNIV was also obtained. Cells from infected and non-infected cultures were pelleted by centrifugation (1500 X g for 10 min), resuspended with media, and centrifuged again. DNA was extracted using a standard phenol and chloroform: isoamyl alcohol protocol, precipitated with ethanol, and resuspended in water. DNA from infected and non-infected cells were used as templates for PCR reactions with HaNIV-specific primers. The primers used for PCR amplification were Hc13AF (5'-CAC ATC GCC CAT GAT TCA) and Hc13R (5'-TGA AGC AGG TCG ATC TAG) and annealed to positions 2412 and 2999, respectively on contig B (Fig. 3.1, Appendix B). HaNIV DNA was used as a control. Aliquots of the PCR products were analyzed on a 1.5 % agarose gel, stained with ethidium bromide, and visualized with a UV transilluminator. The remaining PCR products were cleaned with the QiaQuick® PCR purification kit following the manufacturer's protocol and sequenced directly as described above.
RESULTS

Following the cloning procedure, ten of the thirteen Hind III fragments were cloned and three fragments, approximately 2000, 1000, and 500 bp in length, were not. Because the Hinc II digestion resulted in blunt ended fragments (Hind III digestion left 5' overhangs), the ligation of Hinc II fragments with the vector was not as successful as with Hind III fragments. Thus, only ten of the fifteen fragments were cloned. The two largest fragments and three fragments ranging from 1500 to 2000 bp are missing from the clone library.

Once the cloned fragments were sequenced, overlapping regions were found between sequences from Hind III and Hinc II fragments. Eight contigs resulted from this process. A combination of PCR and direct sequencing successfully identified sequences not represented in the clone library. The eight contigs were aligned into two longer sequences 17 765 and 15 858 bp in length and are identified as contigs A and B, respectively. Due to time constraints the genome sequence of HaNIV was not completed. Direct sequencing by primer walking is continuing in order to align contigs A and B and to sequence the ends of the genome. Although the genome sequence is incomplete, the sequence of contigs A and B, representing approximately 94 % of the genome, was analyzed and annotated for this chapter.

Sequence analysis. The GC content of contigs A and B were 52.01 % and 51.48 %, respectively.

From contigs A and B, 76 putative ORFs greater than 75 amino acids in length were identified. All ORFs were numbered sequentially along with the letters A or B to
denote the contig from which they were identified. The location and orientation of the
ORFs are diagrammed in Figure 3.1 and their coordinates on the contigs are listed in
Table 3.2. The translated amino acid sequences of the putative ORFs were submitted to
the PSI-BLAST program at the National Center for Biotechnology Information (NCBI)
for the identification of homologs.

Similarities were found to lysozymes and the small subunit (SSU) and large
subunit (LSU) of terminases from bacteriophages and prophages (Table 3.2). ORF 16A
was homologous to the terminase small subunit from bacteriophage HK620. ORF 18A,
located directly downstream of ORF 16A, encoded a protein that was most similar to the
terminase B (LSU) protein from enterobacteriophage P1. Other hits to ORF 18A
included a hypothetical protein from a prophage region of *Haemophilus influenzae* and
the C-terminus of putative LSU terminases from bacteriophages HK620 and APSE-1.
ORF 18B was most similar to a lysozyme sequence from a prophage region of *Xyella
fastidiosa*. This ORF also showed homology to lysozymes from bacteriophages PS119
and P22 and to putative lysozymes from bacteriophage PS34 (gp19) and prophages of
*Escherichia coli*.

ORF 1B was similar to the NinB proteins of bacteriophage 21, phage lambda,
phage P22, and gp61 of phage HK97. The homologous region spanned the entire length
of ORF 1B, but did not include the first 50 amino acids of known NinB amino acid
sequences. Homologies were also detected to predicted proteins of unknown function
from the bacteria *Mesorhizobium loti* (ORF 7A) and *Legionella pneumophila* (ORF 21A)
(Table 3.2). Re-iterating the PSI-BLAST search with ORF 21A and the unknown protein
Figure 3.1. Location of HaNIV ORFs along contigs A and B. Black arrows indicate ORFs with homologies to proteins in GenBank. The arrowhead indicates the direction of transcription.
Table 3.2. ORFs predicted from HaNIV sequence. The letters A and B denote the contig from which the ORF was identified. The coordinates describe the location of the ORF from the first nucleotide of the first start codon with a ribosome binding site (RBS) to the last nucleotide of a stop codon. If a RBS was not detected, then the ORF is described from the first encountered start codon, regardless of a RBS. The reading frame is shown in parentheses. * indicates a GTG start codon, otherwise it is ATG. The length of the translated ORF is in amino acids (aa). Matches show significant matches (P < 0.005) for the appropriate ORF. For the highest ranking match, GenBank accession numbers for the protein database precede the description of the protein and e-values and identities are provided. Identities show the percentage of identical amino acids in the homologous region (length shown in parentheses).

<table>
<thead>
<tr>
<th>ORF</th>
<th>Coordinates</th>
<th>Length (aa)</th>
<th>RBS</th>
<th>Matches</th>
<th>Psi-Blast e-value</th>
<th>Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contig A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1A</td>
<td>1A 446 - 670 (2)</td>
<td>75</td>
<td>-</td>
<td>BAB50597, unknown protein, <em>Mesorhizobium loti</em></td>
<td>$2 \times 10^{-7}$</td>
<td>30 (132)</td>
</tr>
<tr>
<td>2A</td>
<td>670 - 1359 (1)</td>
<td>230</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3A</td>
<td>1409 - 1777 (2)</td>
<td>123</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4A</td>
<td>1768 - 2127 (1)</td>
<td>120</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5A</td>
<td>1946 - 2212 (2)</td>
<td>89</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6A</td>
<td>2093 - 2344 (-2)</td>
<td>84</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7A</td>
<td>2127 - 2531 (3)</td>
<td>135</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8A</td>
<td>2624 - 3409 (2)</td>
<td>256</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9A</td>
<td>2906 - 3448 (-2)</td>
<td>181</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10A</td>
<td>3012 - 3281 (-3)</td>
<td>90</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11A</td>
<td>3413 - 4024 (2)</td>
<td>204</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12A</td>
<td>4425 - 4886 (3)</td>
<td>154</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13A</td>
<td>4568 - 4879 (-2)</td>
<td>104</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14A</td>
<td>5004 - 5405 (3)</td>
<td>134</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15A</td>
<td>5108 - 5569 (-2)</td>
<td>130</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16A</td>
<td>5415 - 5867 (3)</td>
<td>151</td>
<td>-</td>
<td>AF335538, Terminase small subunit, bacteriophage HK 620</td>
<td>$6 \times 10^{-12}$</td>
<td>34 (134)</td>
</tr>
<tr>
<td>17A</td>
<td>5560 - 5811 (1)</td>
<td>84</td>
<td>-</td>
<td>P27753, Terminase B protein, bacteriophage P1; also matches hypothetical protein H11410, <em>Haemophilus influenzae</em>; P18, bacteriophage APSE-1; terminase large subunit, bacteriophage HK620</td>
<td>$6 \times 10^{-9}$</td>
<td>25 (303)</td>
</tr>
<tr>
<td>18A</td>
<td>5870 - 7345 (2)</td>
<td>492</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19A</td>
<td>7272 - 7796 (3)</td>
<td>175</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20A</td>
<td>7799 - 8099 (2)</td>
<td>97</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21A</td>
<td>8213 - 9793 (2)</td>
<td>527</td>
<td>-</td>
<td>CAC33475, unknown protein in <em>Legionella pneumophila</em> serogroup 1, 30kb stable genetic element</td>
<td>$4 \times 10^{-5}$</td>
<td>19 (379)</td>
</tr>
<tr>
<td>22A</td>
<td>10104 - 10766 (3)</td>
<td>221</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23A</td>
<td>10785 - 11792 (3)</td>
<td>336</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.2. continued

<table>
<thead>
<tr>
<th>ORF</th>
<th>Coordinates</th>
<th>Length (aa)</th>
<th>RBS Matches</th>
<th>Matches</th>
<th>Psi-Blast e-value</th>
<th>Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>24A</td>
<td>11808 - 12260 (3)</td>
<td>151</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25A</td>
<td>11836 - 12201 (1) *</td>
<td>122</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26A</td>
<td>12263 - 12622 (-2) *</td>
<td>120</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27A</td>
<td>12468 - 13124 (3) *</td>
<td>219</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28A</td>
<td>12537 - 12902 (-1) *</td>
<td>122</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29A</td>
<td>13120 - 15543 (1) *</td>
<td>808</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30A</td>
<td>13148 - 13405 (2)</td>
<td>86</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31A</td>
<td>14566 - 14889 (-3)</td>
<td>108</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32A</td>
<td>15550 - 16131 (1)</td>
<td>194</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33A</td>
<td>16138 - 17763 (1) *</td>
<td>542</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34A</td>
<td>16166 - 16636 (2)</td>
<td>157</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35A</td>
<td>16272 - 16493 (-1)</td>
<td>74</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36A</td>
<td>16694 - 16921 (2)</td>
<td>76</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37A</td>
<td>16968 - 17195 (-1) *</td>
<td>76</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Contig B**

| IB  | 184 - 423 (1) * | 80 | - | CAB39988, NinB protein, bacteriophage 21; also matches gp61, bacteriophage HK97; NinB, bacteriophage phage lambda; NinB, bacteriophage P22 |

<table>
<thead>
<tr>
<th>ORF</th>
<th>Coordinates</th>
<th>Length (aa)</th>
<th>RBS Matches</th>
<th>Matches</th>
<th>Psi-Blast e-value</th>
<th>Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2B</td>
<td>423 - 893 (3)</td>
<td>157</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>640 - 903 (1)</td>
<td>88</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4B</td>
<td>811 - 1116 (-1)</td>
<td>102</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5B</td>
<td>893 - 1702 (2)</td>
<td>270</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6B</td>
<td>1926 - 2153 (3) *</td>
<td>76</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7B</td>
<td>2182 - 2436 (1)</td>
<td>85</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8B</td>
<td>2364 - 2843 (3)</td>
<td>160</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9B</td>
<td>3179 - 3415 (2)</td>
<td>79</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10B</td>
<td>3732 - 3986 (3)</td>
<td>85</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11B</td>
<td>4083 - 4778 (3)</td>
<td>232</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12B</td>
<td>4828 - 5205 (1)</td>
<td>126</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13B</td>
<td>5397 - 5678 (3)</td>
<td>94</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14B</td>
<td>5653 - 6051 (-1)</td>
<td>129</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15B</td>
<td>5989 - 6411 (1)</td>
<td>141</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16B</td>
<td>6033 - 6281 (-2)</td>
<td>83</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17B</td>
<td>6166 - 6525 (-1)</td>
<td>120</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18B</td>
<td>6480 - 7490 (-2)</td>
<td>337</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ORF</th>
<th>Coordinates</th>
<th>Length (aa)</th>
<th>RBS Matches</th>
<th>Matches</th>
<th>Psi-Blast e-value</th>
<th>Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>19B</td>
<td>6836 - 7099 (2) *</td>
<td>88</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20B</td>
<td>7465 - 8769 (-1)</td>
<td>435</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21B</td>
<td>7481 - 7768 (2) *</td>
<td>96</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22B</td>
<td>7888 - 8241 (1)</td>
<td>118</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23B</td>
<td>8351 - 8641 (-3) *</td>
<td>97</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24B</td>
<td>8730 - 9164 (-2)</td>
<td>145</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25B</td>
<td>8743 - 8985 (1)</td>
<td>81</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.2. continued

<table>
<thead>
<tr>
<th>ORF</th>
<th>Coordinates</th>
<th>Length (aa)</th>
<th>RBS</th>
<th>Matches</th>
<th>Psi-Blast e-value</th>
<th>Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>26B</td>
<td>9168 – 9587 (-2)</td>
<td>140</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27B</td>
<td>9605 - 12346 (-3)</td>
<td>914</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28B</td>
<td>10560 - 10793 (3)</td>
<td>75</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29B</td>
<td>10993 - 11232 (-1)</td>
<td>80</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30B</td>
<td>11291 - 11521 (2)</td>
<td>77</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31B</td>
<td>12517 - 14022 (-1)</td>
<td>512</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32B</td>
<td>12937 - 13599 (1)</td>
<td>221</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33B</td>
<td>13038 - 13328 (-2) *</td>
<td>97</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34B</td>
<td>13475 - 13699 (2)</td>
<td>75</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35B</td>
<td>13728 - 13953 (1)</td>
<td>76</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36B</td>
<td>13974 - 14324 (3) *</td>
<td>117</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37B</td>
<td>14031 - 15704 (-2)</td>
<td>558</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38B</td>
<td>14254 - 14499 (1)</td>
<td>82</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39B</td>
<td>14991 - 15407 (3) *</td>
<td>139</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40B</td>
<td>15019 - 15507 (1) *</td>
<td>163</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

from *L. pneumophila* resulted in significant hits to the head-tail joining protein of bacteriophages T7, phiYeO3-12, and T3.

**Phylogenetic analyses.** Phylogenetic analyses were conducted with HaNIV ORFs homologous to LSU terminases and lysozymes. Generally, the parsimony analyses gave better, more robust results with stronger bootstrap support than neighbor-joining, therefore, for each analysis, the parsimony tree is shown (Figs. 3.2 and 3.3). The poor neighbor-joining results occurred because many of the distances calculated between the operational taxonomic units (OTUs) were infinite.

Based on the parsimony analysis of ORF 18A with LSU terminases from bacteriophages, prophages, and herpesviruses, ORF 18A grouped strongly with the LSU terminases from bacteriophages P1 and P22 (90 % bootstrap support) (Fig. 3.2). The herpesvirus terminases grouped strongly together (100 %) and in a cluster including LSU
terminases from bacteriophage lambda, 21, and a prophage in *Xyella fastidiosa* (96 %). Bacteriophage 21 and lambda paired together (100 %), as did T3 and T7, and HK022 and HK97. The association of terminases from HK620, APSE-1, and a prophage region in *Haemophilus influenzae* was also supported (87 %).

The analysis of lysozyme sequences (Fig. 3.3) included lysozymes from chicken (*Gallus gallus*) and goose (*Anser anser*) in addition to viral ones. The chicken and goose lysozymes grouped within the phage lysozymes although strong relationships were not found. ORF 18B grouped with prophage lysozymes in *Xyella fastidiosa* (83 %). Lysozymes from the siphoviruses lambda, HK022, HK97, and HK620 grouped together with strong support (100 %) with the P2 lysozyme as a sister to this group (100 %). The fifth siphovirus lysozyme included in the analysis, from phage 21, grouped with the lysozyme from prophage DLP12 in *Escherichia coli* on the opposite end of the tree (99 %). Lysozymes from phage 21 and DLP12 clustered with those from PS119 and PS34 (100 %). Other strongly supported groups were lysozymes from T3 and T7 (100 %) and APSE-1 and P22 (100 %).

**Analysis of the host genome.** PCR with HaNIV-specific primers successfully amplified the expected 600 bp product from the DNA of infected and non-infected *Heterosigma akashiwo* (Fig. 3.4). Amplification was consistently stronger with template from infected cells. The sequences of these products were identical to HaNIV sequence (data not shown).
Figure 3.2. Unrooted parsimony tree of viral LSU terminases and HaNIV ORF 18A. The consensus tree is shown. Numbers at branch nodes indicate percentage bootstrap support from 100 replicates. Values less than 50% are not shown. Any significant support from the neighbor-joining analysis is shown in parentheses.
Figure 3.3. Unrooted parsimony tree of lysozyme amino acid sequences and HaNIV ORF 18B. The consensus tree is shown. Numbers at branch nodes indicate percentage bootstrap support from 100 replicates. Values less than 50 % are not shown. Any significant support from the neighbor-joining analysis is shown in parentheses.
Figure 3.4. PCR analysis of DNA from infected and non-infected *Heterosigma akashiwo*. PCR products with HaNIV primers using template DNA from non-infected *H. akashiwo* (lane 1), infected *H. akashiwo* (lane 2), and HaNIV particles used as a positive control (lane 4). DNA was not amplified from DNA isolated from a resistant strain of *H. akashiwo* (lane 3). The negative control was run in lane 5 and molecular weight markers in lane 6.
DISCUSSION

By using the genome sequence of HaNIV, potential open reading frames were identified and compared to genes of other viruses. HaNIV ORFs were similar to several known bacteriophage proteins and these similarities provide insights into the evolution and biology of HaNIV.

**ORF 16A and 18A possibly encode terminase subunits.** The translated sequences of ORFs 16A and 18A were homologous to the small and large subunit of phage terminase enzymes, respectively. Similar to phage genomes (Skorupski et al. 1992, Loessner et al. 2000, van der Byl and Kropinski 2000, Desiere et al. 2001), the SSU ORF was located upstream of the LSU ORF. The terminase (or pacase) enzyme complex is responsible for cutting DNA concatemers and packaging DNA into viral capsids (reviewed in Black 1989). Concatemers are tandem arrays of mature phage DNA covalently joined end to end. Bacteriophages use various strategies to ensure that a full complement of genetic information from the concatemer is packaged into each progeny prohead. For λ, T3, and T7 phages, terminase cuts the DNA concatemers at specific DNA sequences (for lambda these are called *cos* sites), thus the mature DNA in each virion has identical ends (Hohn 1983). For T4 phages, the cleavage of DNA concatemers does not depend on sequence. The initial cut is random and the second cut occurs when the virion head is completely filled with DNA ("headful" cleavage) (Streisinger et al. 1967). This results in virions with terminally redundant, circularly permuted DNA. For P1 and P22 phages, the initial cut in the DNA concatemer occurs at a specific sequence, designated *pac*, but the second cut is non-specific and occurs when the virion is filled.
(Tye et al. 1974). Packaging into successive proheads continues processively from this second cut in a headful manner. This results in viral genomes that have non-random terminally redundant circularly permuted ends.

The presence of terminase-like ORFs in the HaNIV genome suggests that DNA packaging of the genome may be similar to the strategies of known bacteriophages. ORFs 16A and 18A were most similar to the SSU terminase of HK620 (a siphovirus) and terminase B (LSU) of coliphage P1 (a myovirus), respectively, therefore HaNIV may have a DNA packaging strategy similar to these two viruses. The function of the HK620 and P1 terminase complexes cannot be compared, however, because the DNA packaging strategy of HK620 is not known. Because the terminase complex of HaNIV has components of terminases similar to two different viruses, it would be interesting to test the functional complementation of the three terminases. In addition, determining the ends of the HaNIV genome will be useful for inferring a DNA packaging mechanism for HaNIV.

**ORF 18B shares similarities with phage lysozymes.** The lysozyme amino acid sequences of many phages were homologous to HaNIV ORF 18B. Lysozymes are responsible for degradation of the host cell wall late in replication so that progeny phages can be released. ORF 18B was most similar to putative lysozymes from prophages in *Xyella fastidiosa* clone 9a5c. *X. fastidiosa* is a bacterium responsible for various diseases in economically important plants. Knowledge of prophages in the genome of *X. fastidiosa* was evident following analysis of its complete genome sequence (Simpson et al. 2000). In addition to prophage lysozymes, ORF 18B was homologous to lysozymes
encoded by gene 19 of the bacteriophages PS119, PS34, and P22, all members of the
*Podoviridae*.

There are several genes responsible for cell lysis during the late stages of phage replication. The lysis enzymes of P22 (gp13, gp15, and gp19) function analogously to those of bacteriophage λ (proteins S, Rz, and R, respectively) despite the absence of noticeable amino acid sequence similarities (Fig. 3.3, Rennell and Poteete 1985, Casjens et al. 1989). For P22, and similarly with λ, pores in the inner membrane are formed by gp13 through which gp19, the main lysozyme, can exit. The lysozymes produced by gene 19, also known as muramidases (λ produces transglycosylases), degrade the peptidoglycan (or murein) layer. The function of gp15 is not completely understood, but it is thought to complete peptidoglycan degradation or disrupt cell membranes in the presence of divalent cations (Casjens et al. 1989).

*Heterosigma akashiwo* does not have a rigid cell wall like bacteria, but it does have a glycocalyx surrounding its cell membrane. The glycocalyx is formed by mucus secretion and is composed of acidic, complex carbohydrates (presumably of hyaluronic acid) and a neutral protein-carbohydrate complex of glycoproteins (Honjo 1993). It may be possible that ORF 18B is a lysozyme responsible for degrading the glycocalyx during lytic viral replication.

**ORF 1B is homologous with NinB sequences.** The Nin region of bacteriophage λ is a region of eight, mostly unknown ORFs between the DNA replication gene P and the regulator of late transcription, gene Q (Hendrix et al. 1983). The region is so named because of the N-independent (Nin) phenotype that results from the nin5 deletion mutation which covers this region. The product of gene N is responsible for activating
early gene expression by antagonizing transcription termination and is required for λ to
grow lytically (Hendrix et al. 1983). In pN- mutants, an additional nin5 mutation allows
lytic λ replication because of the deletion of a transcription terminator (Hendrix et al.
1983). Thus, early genes are transcribed without a functional N gene, i.e. an N-
independent phenotype.

HaNIV ORF 1B is homologous to NinB (also called orf) of λ and to similar
proteins from bacteriophage 21, HK97, P22, HK620, 933W, VT2-Sa, and prophages
from Escherichia coli. Most of these proteins are similar in length, 145-148 amino acids
(aa) long, with the exception of the proteins from HK620 (136 aa), VT2-Sa (181 aa), and
the considerably shorter ORF 1B from HaNIV (80 aa).

The λ NinB is one of the few nin proteins for which a function has been inferred.
It is a trans-acting factor involved in λ recombination (Sawitzke and Stahl 1994). In
addition, Juhala et al. (2000) propose that the function of the entire nin region, is to
provide a collection of genes enabling a phage to adapt to a particular ecological niche.
This is an interesting hypothesis to pursue considering that HaNIV is a marine virus and
other viruses with nin proteins are enterobacteriophages. However, ORFs surrounding
ORF 1B did not show homology to other nin proteins and further analysis is necessary in
order to determine the function of this region.

**ORF 21A may be a structural protein.** ORF 21A was most homologous to an
unknown protein from an unstable genetic element of Legionella pneumophila, a human
bacterial pathogen (Lüneberg 2001). Both of these ORFs showed homology to the head-
tail joining proteins of bacteriophages T7, φYeO3-12, and T3, all Podoviridae.
Morphologically, HaNIV resembles podoviruses, but has a small extension at one apex of
the capsid instead of a distinct tail (see Chapter 2, Fig. 2.1A). Because ORF 21A has homologies with head-tail joining proteins, ORF 21A may encode a structural protein involved with its capsid extension. The head-tail joining protein of T3 is 58.6 kDa (Beck et al. 1989) and it is expected that ORF 21A encodes a protein of similar size. However, a structural protein of this size was not detected from HaNIV particles (see Chapter 2, Fig. 2.3) likely because the protein is a minor component. Another potential structural similarity between HaNIV and podoviruses involves the major capsid protein. The major structural protein of HaNIV is 38 kDa and, although podovirus particles contain about 12 proteins ranging from 13 to 150 kDa, they also have a major capsid protein of 38 kDa (Ackermann and Dubow 1995). Unfortunately, the ORF encoding the major capsid protein could not be identified from the HaNIV genome sequences. Identifying the amino acid sequences of the HaNIV capsid proteins will be required to confirm any structural similarities between HaNIV and podoviruses.

**Phylogeny.** HaNIV encodes several putative proteins that are homologous to bacteriophage proteins. However, the phages that encode these proteins do not belong to a single group, but span a diversity of bacteriophages. HaNIV ORFs 16A, 18A, 21A, 1B, and 18B showed the greatest similarity to proteins from a siphovirus (HK620), a myovirus (P1), podoviruses (T3 and T7), a siphovirus (phage 21), and prophages (in *Xyella fastidiosa*), respectively. Much like the genome of most phages, the HaNIV genome is probably a mosaic of genes with different evolutionary origins.

The phylogenetic analyses performed in the present study were used to demonstrate the mosaicism of HaNIV and other viruses. In each analysis, proteins from the same phages (myoviruses T4, P1, and P2, podoviruses T3, T7, P22, and APSE-1,
siphoviruses lambda, 21, HK97, HK022 and HK620 and HaNIV) were included. In
general, viruses belonging to the same family did not group together. Podoviruses T3
and T7 and siphoviruses HK97 and HK022 were the only pairs of viruses to consistently
branch together. In addition, aside from these pairings, branches supported in one
analysis, were not by the other.

The lack of sequence homology among proteins from similar phages despite
having analogous gene function is well recognized (Ackermann 1998). For example, as
in Fig. 3.2 and other studies (Smith et al. 1999, Desiere et al. 2001), some bacteriophage
terminases are more similar to herpesviruses than to other phages belonging to the same
family. Lysozymes from T4 and P22 are known to share similar 3-dimensional structures
and enzymatic functions to animal and plant lysozymes, but have variable amino acid
sequences (Weaver et al. 1985, Holm and Sander 1994). It has also been shown that the
lysozymes of T4, P22, and λ are interchangeable in complementation assays even though
they have different enzymatic activities and primary structure (Rennell and Poteete
1985). The difference in amino acid sequences among proteins with analogous functions
could be due either to horizontal gene transfer or the evolution of the protein from an
ancestral gene originating prior to the divergence of prokaryotes and eukaryotes (Grütter

Although some bacteriophage genes may be ancestral, horizontal gene transfer,
involving the exchange of modules or blocks of genes between viruses, has been
described as the major mode of bacteriophage evolution (Ackermann 1998). This mode
of evolution was recognized because of the genetic mosaicism observed in phage
genomes (Hendrix et al. 1999). Thus, due to the extent of horizontal gene transfer among
viruses, phylogenetic analyses of protein sequences can be used to infer evolutionary relationships among viruses, but not their origin. Instead of sequence comparisons, genome organization, gene order, and protein conformation and function may be better indicators of viral evolution. These types of analyses are currently driving our knowledge of the evolution of viruses.

For HaNIV, because the function of most of the ORFs is unknown, gene order and organization could not be detected. Of the ORFs for which a function was putatively determined, a structural gene (ORF 21A) followed the terminase genes (ORF 16A and 18A). This gene organization is also found in many bacteriophages (Pfister et al. 1998, Desiere et al. 1999, Lucchini et al. 1999, Juhala et al. 2000). The genomes of most bacteriophages are organized into gene clusters of related function (Ackermann 1998). These clusters include packaging, head morphogenesis, tail morphogenesis, lysis, lysogeny, and replication genes. To determine if the HaNIV genome is similarly organized, the function of the putative HaNIV ORFs must be identified.

**Algal virus or bacteriophage?** The sequence of the HaNIV genome revealed homologies that were unexpected for a virus infecting a eukaryotic phytoplankton. Five ORFs showed homology to proteins from bacterial prophages and bacteriophages. Because the virus replicates within a eukaryotic host, it was expected that HaNIV ORFs would be homologous to genes from viruses infecting eukaryotes. There are two possible explanations for this result. HaNIV is a virus of *Heterosigma akashiwo*, but carries genes reminiscent of prokaryotic phages. Alternatively, the virus may infect a bacterium and not *H. akashiwo*. 
The homologies of HaNIV ORFs with phage proteins is not completely surprising considering the prevalence of horizontal gene transfer in the evolution of viruses. As Hendrix et al. (1999) explain, all viruses have access to a global pool of gene sequences and the extent of horizontal gene transfer between viruses of phylogenetically distant hosts depends on the number of individual steps through "phylogenetic space" required to bring a particular sequence and virus together. Hendrix et al. (1999) were referring exclusively to bacteriophages, but the presence of diverse phage-like genes indicating the mosaicism of the HaNIV genome suggests that, if *Heterosigma akashiwo* is the host of HaNIV, horizontal gene transfer may be much more extensive than previously considered. In addition, because the similarities between HaNIV and bacteriophage genes are still recognizable, the gene sequences are either highly conserved or the gene transfer occurred relatively recently and the "steps" through "phylogenetic space" must have been very large.

Common ancestry could also explain the sequence homology between HaNIV and bacteriophages. Homology due to common ancestry between viruses of phylogenetically distant hosts has been demonstrated previously. Between animal and bacterial viruses, the homology was not recognized through sequence similarities, however, but phenotypically through the solved structures of capsid proteins (Benson et al. 1999). The homology in protein structure between animal and bacterial viruses is believed to be due to common ancestry (Hendrix 1999). Homologies, both morphological and genetic, are also noted between viruses from the two domains of prokaryotes, the Bacteria and Archea, and are probably best explained as the result of common ancestry (Pfister et al. 1998, Hendrix 1999).
The alternative explanation for the phage homology, that the isolated particles of HaNIV are bacteriophages, cannot be completely discounted. HaNIV particles were not isolated from an axenic culture of *Heterosigma akashiwo*, so it is possible that bacteriophages were isolated inadvertently. Although the hybridization assay discussed in the previous chapter indicates that infection is associated with *H. akashiwo* cells, there could have been contaminating bacteria. Thus, if the HaNIV particles are bacteriophages, then the bacterial host must have a physical and physiological connection with *H. akashiwo*. There must be a physical relationship because the infected bacteria must have pelleted with *H. akashiwo* during the host DNA isolation procedure of the hybridization assay. The centrifuge spin was designed to pellet *H. akashiwo* cells and leave bacterial cells in the supernatant, but bacteria may have been caught in the thick glycocalyx of *H. akashiwo*. Secondly, there must be a physiological relationship because virus inoculation results in the lysis of *H. akashiwo*. Therefore, a viral infection of the bacteria must negatively affect the growth of *H. akashiwo*. Unfortunately, ecological relationships between bacteria and phytoplankton, especially symbiotic or obligate, are poorly understood. The few studies relating bacteria and *H. akashiwo* concern algicidal bacteria (Kim et al. 1998, Yoshinaga et al. 1998) and the association of bacteria with *H. akashiwo* toxicity (Carrasquero-Verde 1999).

Further evidence showing that HaNIV is a virus of *H. akashiwo* will have to be obtained from more complex techniques such as *in situ* hybridization with transmission electron microscopy or *in situ* PCR. Despite the ambiguity of the host, however, the homology of HaNIV ORFs to phage enzymes provides insights into the replication process and evolution of this virus.
HaNIV may be latent. The PCR and sequence analysis of non-infected *Heterosigma akashiwo* DNA indicated that at least a portion of the HaNIV genome was present (Fig. 3.4). This result suggests that HaNIV is latent, but it is not known if the genome is integrated into the host's genome or if it is extrachromosomal. An integrase gene, which encodes an enzyme responsible for integrating phage genomes such as λ's into its host, was not detected from the HaNIV genome but the gene may have diverged from the λ-type integrases.

The discovery of a latent virus that can cause lysis of *Heterosigma akashiwo* has important ecological implications. If viral infection is an important factor in the bloom termination of *H. akashiwo* (Nagasaki et al. 1994a, b), progeny viruses would be released in the water column. These particles would be subject to numerous processes causing the loss and decay of viral infectivity. The decay of viral infectivity is caused mostly by solar radiation, but virus particles are also lost due to protozoan grazing and attachment with particulates (Suttle and Chen 1992, Gonzalez and Suttle 1993). Because of these mechanisms that effectively eliminate viral infectivity, other strategies must be in place to ensure the virus' survival. Latency may be a survival mechanism for viruses when host abundance is low (Steward and Levin 1984) such as following the termination of a bloom or during the winter.

The prevalence of latency has been investigated for marine bacteria (Jiang and Paul 1994, Jiang and Paul 1996, Weinbauer and Suttle 1996, Weinbauer and Suttle 1999). Latency, however, has not been studied in phytoplankton. If HaNIV is latent, this virus, together with *Heterosigma akashiwo*, may provide a system to further investigate the impact of latency in viral and phytoplankton ecology.
In summary, two main results were obtained from the use of HaNIV genome sequence data. Putative HaNIV ORFs showed homology to bacteriophage genes suggesting an evolutionary connection between HaNIV and bacteriophages. The sequence information was also used to demonstrate that HaNIV might be latent. The full implications of this study, however, await further data. Although over 90% of the genome has been sequenced, the entire genome sequence should be completed to rigorously analyze gene order, genome organization, and regulatory regions. Additionally, because of the importance of latency in our understanding of phytoplankton ecology, it is imperative that we irrefutably determine the host of HaNIV and its relationship with *Heterosigma akashiwo*. 
LITERATURE CITED


APPENDIX A. Sequence of HaNIV contig A

```
10 20 30 40 50
GTCGACACAT GCCTAAAAAT ACTTGCCGAG TTTAACCTTG ACGGACCTAAG
60
ATCCGACTT CGTCGACAGT TTGTCTGAAC AGGCGGGCG GCACATTTC CAACATGAA
120
GAGTTGGGTG TCTGCAAGGC GCAGGCGGCG AAGGTCGAGC AAGGCTGACGA
180
GGCGTGGCAG ACGGATCGGC TGATGACCTT CAAAGACCTA CAGGTACGAT
240
TGAACATGTC TTAAGTGGGC CCAGGTGACT GAAATGGGC CGCGACAGCA
300
GCGGCGAAGA AGGCGTCGTT CACGAGCTGA CCAGGTCGCC GTCTGAACGC
360
AG(T)GGGACAT CAGGAGTGGC TCAGCTGAGC TCGGCGTGTG CTACGCTCAG
420
GAGGCCCGAG GAGGACCGAA CAGTGACGCT TGCAGTGACT CAGCTGACTT
480
GTGGCAAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
540
GGGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
600
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
660
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
720
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
780
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
840
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
900
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
960
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
1020
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
1080
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
1140
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
1200
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
1260
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
1320
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
1380
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
1440
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
1500
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
1560
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
1620
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
1680
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
1740
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
1800
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
1860
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
1920
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
1980
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
2040
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
2100
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
2160
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
2220
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
2280
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
2340
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
2400
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
2460
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
2520
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
2580
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
2640
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
2700
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
2760
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
2820
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
2880
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
2940
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
3000
GGCGGAGGC TCGGCGGCGA CCGGCGTGGC TCGGCGAGGA CGCGACTGAG
3060
```
CCCTTTCGCT TTTCTGCGAC GCAGACAGTT ATGCTGGAGT TTGGCGAGGC ATATATCCGG 13380
TTCCATAGCT TTACCAAAAG GCCAGAAGCT GCTGGTATCG CTTCTGGGTT ATGATGCTAT 13397
GGTGTTCTTG TTGCAGCAAC GGTTTTGACG CCAACGACGG GCGTTGATGC TTTTGATGCA 13444
GGGTTAACTT ACGACCAAGG CGATATTGTT ACCGAAAGCA GCAAGACGTG GTACGCGGTT 13501
AGCGACGTAC CAACGGGCCA TACGCCAAGC TCAAACGAAT ATGGCTCTGC TCCAGTCGTT 13558
TCTGCAACAT GGTGTGCAGT CAGTGCACGG AATGCGTCAA TTCCATCTGG TTATACGTTC 13615
TACGGCACTT CATTGCCGGA AGATGTCGGA GTTGGTGACA AGGCGGCAAT ATCCAGAACA 13672
ACTTACACGT TTGATGAGGG CGCTGGCGGC AGCTTAGAAC CGGACGTGGT TCTTGTGGAA 13729
AGCACGGTTT ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 13786
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 13843
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 13900
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 13957
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 14014
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 14071
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 14128
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 14185
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 14242
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 14299
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 14356
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 14413
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 14470
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 14527
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 14584
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 14641
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 14698
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 14755
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 14812
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 14869
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC 14926
GAGTTGGGGA TAGTTTATGA ATGATGCTTA TGATGGCGTC GCTAGTTCTG CGTCAGCGGG TTACTGGTAC
CTGGATCAA AAGCGGTGCG AGCGATGGGC GAGCTTTTGT CGGTCAGAT TAGTATCTTG 16800
CTGGATGCCG ATCGAATTTG TGATGCCGCA AAGGTTGTTG AGGCCACGCGGG GGTAATGGTC 16860
TCGCGGATAG TTGCAAGAGGG GATGGCCACCC CGGTGATAGA AGAAAGCCTT TACAGGCGAT 16920
GTGAAAGGGCA AGGGCGGGGA TGGTCTGGAT GTGCGCGGAG GTGACTATGG CACGCTCTTG 16980
AAATGGGCGG GGTAGATCGA AAACCTGAA ATGCAAGAGG CGGTTAAGTGC TCGCTGGAAT 17040
ACGCTAAAG AGCGCGATAC GCCGAGAGAG GCAGCGAGAGGC CACGAGATGC CACGGTCACG 17100
GCGTGGGAAC GTCTTGAGCA GGGTGGTTTG GTGATGCCG TGCCCGTGAG TGCCCGGAGG 17160
CAGCTAGGAC CCGAACACCG GATAAGAGTG CGCAATGGGG AGGAAGCGCA GGCAGAAACG 17220
GATGATCCGA AGCGCGAGAC GGATTTGCAA ACTTACAATG ACATATGACG CTTCCTTGAT 17280
GCCGCGATGT CGCAGAGGTC TTTGAGTCTC CTGAAATGCGA ACGCTGATGA GGTTTACAGG 17340
GTCGATTGATA AAGCCTTACC CACCAAGGTT AGTTCGTCGA TGGATGATGA CGGTGCAACG 17400
GTPAAGAGGTG CGCCAGCGTG AAGCAGGTC ATCACAAATG CTTAGATCG TTCAGGCGTG 17460
AACGAAAGTA ACAAGGCGG GCTACTTCTG GCCATGACG ATGCAGGCAG GCAGTATCAG 17520
CTGAGAGCTG GCAAGAGGCC TTCTGAGTAC GCGGTGAGAA AAGCATTTGT GGCCTGCGCG 17580
GCTAAATCGA AGGACGACG TCTTTGTCGG TTTAATGACA CGACAGGCGG GGCCTTGGAG 17640
ATTGAAGAGA TTGGGAGGCT CGCGGACCGG CATGTGCAGG CGTTTTAAC TTGGTCTGGA 17700
AACACCGGATA AAGTTACTAC GATTCCCCAA GACAAATTGA ATACATCTTG CTCAGATGCCG 17760
ATGAC 17765
APPENDIX B. Sequence of HaNIV contig B

AAGCTGACCG CAGACGAGGC TGGGTGACCT CAACGCAACG AGAGCAGCTG TTATGAGCGC

10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300
CTACTCCAGC GCCACCGGTC TCTTCGCGCT ACCGAAGCGT GATAGACGTT AGGGTCATGT
CCTCAGGTAT AGCCCACGTA ATTTGGGTCT TGCCCGTTAC TGCAAAGCTG TAAGCATATG
GCCCGCCGTC ACCATATGTT TGTGAAGTAT CGAGGGCAGT CAGCGTAGGC CCGGTTATCA
CGGCATTAGC AAGAAGTGGG AAAGAGCGCG CGCCAATAGC CCCTCGCGGC GCAATTTGAG
CCTGCGCCAG AACGGCGGCG TAGACCCACG GACCAAGATA GTTGCCTGGA TGATTGTAGC
CAGACCCAGC CCCATCCCCT AGCGCGAACA ACTGACGATC TTTGACTTGT GGCGTATAAG
CGCGCGGACG CGCATACAGC ATTTTTACCG CCGTGACGAA AGCTGCGCCA CTCGCATCTG
TAACCTGTGG TGTAAACCAG CCGCCGTAGC GATAAAGAAC CTCCGAAAAC AGGGGAGCTA
CCGTCTGCCC AAAACCATTT CCAGTTGTAC TTAGCCAGAC TGAAATACGG TTGCCGCTAA
TTTCTATATA TAGCTCAACA TCGGATGCAA ACCATGTACC CTGTTGCAAG TTTTGATAAT
TGTCAGCCGC TCCAGCTGGA CCTGACGATA AGCGCCAGCG AATGTTTCCG CCCGCTGTCT
TAGAGATACG CAACCAATCG AGCGGACCTG CGCCGACCTG AATAACGACT GAAGCATCAG
GGTCTGTTGT GTCATCAACT AATTCTGACA CGTCCAAAGA TACCCGATAC AAGCAATCAA
TGACTTCTTT TGAGCTCAGT ACAAAGGAAA CTAGCGTTTG CGGTGGCTTC AATATTCGGC GCATCTGTCG
TGTAATCGTT ATCTGATGC ATGGGCGCGT ATGAGCGAAC CAGACCGCAC GCCTGATCTC
GCGTGTCCTG CTCGGATCGA ACACCGTATT GACTGTCTGT CGGGTCAAGA GCTGGTCTGA
CGCATGTCGC TAAAACAGAT TGTGGCAGCG TTTCCAGATT GCTAGCCCAG CTTATAGCTG
CGTCTAAATC GTCTGGATCA AGCGTAGCTC CATGGTTCAT GCCAAACGAA AATATAATAA
TATCGGGACT TAGTTGCTCA ACCAAGTCTT GCCATGACTG AGTTCCATCA GTCCACCACT
CAACCCCTGA CGGCTGATTA CCGCTACCGG GCGCAGCAGC ATTAAAGTTA CTGTATGTTT
TGCCGCCAAT AGCGCGGTTA TAAAATGTAA AACCACCCTG TCCAAGTTGC TCAGATAACC
ATTGTTCAAT CGCGGCCGAC AACATATCAC CGCGCCCGTT CGCATTGGCG AGATAAGTCG
AAATGGAGTC ACCACACGC ACCAAACAGC GGTCTTCAGT AAAAGTGTTT ATAGAGCATG
TTTCTGGGTT ATGCTGCAGT GCATCAACCG GAATGACCTG CTTTCGATAA ATTCCAGTGA
GCGCGCCACT ACCAAAGAAA CTAGCGTTTG CGGTGGCTTC AATATTCGGC GCATCTGTCG
TGACAGCGCT GCCCGTCGCG CCTGTCGCAA CATATCGCGC CAAGTTAATT TCACCTGTGT
CTAGATAGCG AAAAGCGCCC ACTGTGCCGG ACTTGTCGCT AGTGAGCGTA AACCAGACAC
CGCCATCACC TTCGCTAGAT GTAACCTCAT CGGCGCTGAC ATTAGATGAC TGGTCACCTG
CTTCCCATAC GCACAAGCCT TCGCCGCCAT CACCCGTTGC GTCACGGCCT GACAATATAA
AACCCTGCCC AGCAACAGGC GTATAGGTTA AGCCTCGCAA CTCCGTCTTC GTCTTAACCG
CGGGCGCAGC CCAATTTATT GCGCTAGCAA GAGTTCCGTC TTCAATCGCC CCGGAAAGCG
TAACAATCGC GGCAGCAATC CCCGAAGTAA GGATAACCTC CGCATCAATA TCAGAGACAG
CGGAAACGTC CGCAATGAAT GCAGTTACAA GCAGTTACAA GCAGTTACAA GCAGTTACAA
GAGTGGACTG CCAAGCAGCA ACTGTCGGCG ACTTGTGCCT AGTGACGCCT AGTGGACGCCT
CGCATCAGCT GCAGCTAGCT GCAGCTAGCT GCAGCTAGCT GCAGCTAGCT GCAGCTAGCT
TGGGCGCTGT AACCATTGTC ACATTACCGC CCGGACTTGC CTCTTGGTCC GCGTTTAAAG
TTGACTGTGTA ATTCGTTGTA ATCGTCAGAA CAGTTTGACG ATCAGACTCC TCACGGATAA
CTTGCAAATC CGCCTGAGAA AAACATTTAA ACGAAAAGGA AAATACAGTC GTTGAGCCGT
TACCCGTTGG CCGCGTCTTT CCGTCCATTT CTACGTGCTT TCCTCTGTTC TCTGTCGGC
ATATCGGGATG AGCTGCCTTG TAGATGCTTA AACCATGCCT ATGAGCTCCA TATGTTCTCA
CTCTCCATTG AGCTGCCTTG TAGATGCTTA AACCATGCCT ATGAGCTCCA TATGTTCTCA
CTCTCCATTG AGCTGCCTTG TAGATGCTTA AACCATGCCT ATGAGCTCCA TATGTTCTCA
CTCTCCATTG AGCTGCCTTG TAGATGCTTA AACCATGCCT ATGAGCTCCA TATGTTCTCA
CCGGCTCTGC  ACACGCGCTTT  TCCTAATCGG  CTCATGCGCC  TCCGGTTAAG  TGGTTCTCTC  13320
ACGGCTACGC  ATGAAGACCG  ATTTTTCGTT  AATCCTGCTTC  GATGCCTTGA  TCTAGCTGAAG  13380
CCGGCCGTCG  CAATATCCCG  ACGCCGCGCC  GGCCGAATC  CATACGGGCG  13440
AGAATAGCTG  TAGCCGCGCTT  ACGGCTTAAA  ATAGTTGTCG  GATGTCTTTA  13500
CAGACGGTAC  GCCGCTATCG  CACATGCTCC  TGGGCGGCG  GCCGGTCAAG  13560
ACCTCGCTCG  CGGCGCGTTA  CCCGGTTTCA  ATGTCTTGCG  GCCGCCATAT  13620
CCCATATTCA  GGCGCATTAA  TTTCAGACGAT  GGTGTTTCAA  AATGGCGAAG  13680
AGGACGACGC  AAGAGCGATG  CCGAGCACG  TCTCTCGGCT  GCCGCGATAT  13740
AGTCGATCG  AGACGCACAA  GGGCTTACG  ATTCAGCTTTA  CGCTCTTCGA  AAAAAACCGG  13800
ATTGGTCGAG  GCCCCGCGAT  CCTAATACGT  GACCCTGCCG  TTCCATCTCG  AATTCRCGTC  13860
CTGGGCTGAT  CGGATCTGAA  TTGGGCGCAT  TATTTCTACG  GAAATAGCC  13920
GCCCGAATAT  GTCCGCAACT  TCGTCTGAGC  AGATACGGC  TGCACCTTTG  TTTAGTGAAC  13980
CCCTTCTGA  CGATAAGCCT  CGCAGACGCC  AATCGGCAC  ATCTGGCATA  GCAATCGCAA  14040
AGTCGATCG  CTGCCATCGG  GCGAGCTGT  CCAAGAAGTC  GTGCTCTATC  GCCTTGCTGC  14100
GCCGCCATAT  GCCGACACCG  GACAGACCAG  GATCGAGGTG  GTGCTCTTTA  CAGAAAGTCC  14160
CCCACTGTTA  GCGCACTCAG  GACTGCAGCA  ATAAACGCGA  CATACTCAGC  CTCACATCAG  14220
AAGGTGCTTT  CACTGGCGCT  CAAATGGCCT  TCCGCTGTAT  TGGTCAGTTT  CTGCATCTGA  14280
TTCCAGAGAT  GGCTACTGCC  GACCTGCGAT  CTTGAGCGCC  TCTGCTGCTC  CAGAATCCTG  14340
GGTGTCTCAA  ATGCCTCATG  ATCTGTCGCA  ACCCTGCGCC  TTTCCAGAGA  CTTGAGCGCC  14400
AATGGGCAGC  GCTAAGACCA  AACATCAAGG  TCCGGTACCG  TGCTCTGTTG  AATTCRCGTC  14460
AGGTGATGAT  GGGCTGGTTCC  GCGCTTTTCC  TGTCCCTCGAT  GATTCCTATC  AACATCAAGG  14520
GCCAGCTTTT  GCCGCCATCG  CGAGAGCTGC  TGCAATGCGT  TTTCCAGAGA  CTTGAGCGCC  14580
AGCTGTCGTG  CTCTTCTTAC  GCGATCGCGA  AAACCTCCTG  ATTCAGCTTC  AGGCTCTGCG  14640
CGACGACGCG  GACGCTCGAT  GCCGCCACAA  AGGACGCGCC  CTCCGCTTCG  GCCAGCACCA  14700
TCACGGGACT  GTGCAATGCC  AAACATCTCC  AGAGCTTGCA  TCAACATGGC  TGGCTCTGCG  14760
GCCTCTGGCA  GCTTGATCGC  CTGGTACTGT  TTGCTGATAT  AGCGAGGCGC  TATCTCTTC  14820
GGTGTGTGTG  CTTGATGCCG  GCACAGCTTG  GCAAGACGCA  TGAATTCGTC  TTTCCAGAGA  14880
TCAACGGGCT  TAAGACGCGC  GCCGCCCTTC  CTGATCTGCC  GCATGCTCTG  ATCGGTNTAG  14940
CGGACGACGC  CATCCCAACT  GCCGCCGCTTT  GCCGCCGCTA  GCCGCCCTCT  TGAGAGATTG  15000
CCGGCCGTTA  TCACAGGCTA  GCCCGTCTAT  AATGGGCTTG  ATGCACCTAT  GCCGCCGCTA  15060
CCGGGTCGCC  CTTCCGGCGG  AATGCCAGCC  AATGGAATAG  CTTGCTTACG  AGCCTTTTCA  15120
GCCAATGCGT  TGCGATCCGG  TTTCCGGCGG  ATGCTTTCGG  AGCCTGGTCC  GCCGGTCTCC  15180
ATTGCTTCA  CTCACACGTC  GCATGGGCCT  AGAAGCTGGC  CTTAGCTTTA  GCCCGATGTC  15240
TTCAGAGCG  TCTAGAGAC  GTAAGATCTC  GCCCTTCCCT  GGTCATGGCA  CTCTAGTCC  15300
ATTCATCTTG  CGGCAACAAA  ACCATCTCTC  GCCGGCAACA  TAACAGGATG  TCTGCTGCTC  15360
TCCGAGCGTG  GCTGTTGCTG  ACCGGCTGCG  GAGTGGCCCG  CATCCGACG  GCCGCCGCCC  15420
TCACAACGCG  CACCAGCCGG  TTATTTATCT  GGCATTCTG  CATCTGCGC  ACGGACCATG  15480
TCTGCAGTCG  GCTGCTCTGC  TTTCCATATA  GGCTTCTCTC  GCCTGGCCCG  AATGGGCTTG  15540
ATATGCTTTT  TAAACAGGCT  GC CTAGCGACG  ACAGCGCCGG  CTTCGGGCCG  GTGTTGTCGCA  15600
TACTTTCATAGCGGC  CTCAGTCGAT  CTGCTTGCAT  ACTCCGTAAT  GCCGGCGGCA  15660
TGACGACCTG  GGCACAGCTG  AAATCAGGCC  TCAGTAGCGA  AGATCGAGTG  AAAGATGCTG  15720
CAGGCCTCCG  GCTGCTGCTG  GCCGATCGGC  CCCCTTAAAC  GCCGGTGACG  ATCGTGTAAT  15780
CAGTGCGGAG  ACTATCGGAA  CCGGGAGCTA  AGCAGCAGGC  TGGTGCCGAC  GCCGGCGGCG  15840
AAATATGCTT  CAAGAGCTT  15858