

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Student Research Projects, Dissertations, and
Theses - Chemistry Department

Chemistry, Department of

3-30-2006

A COMPARATIVE GENOMIC ANALYSIS OF CHLORELLA NC64A VIRUS NY-2A AND CHLORELLA Pbi VIRUS MT325 FROM THE FAMILY PHYCODNAVIRIDAE

Lisa A. Fitzgerald

University of Nebraska-Lincoln, lisa.fitzgerald@nrl.navy.mil

Follow this and additional works at: <https://digitalcommons.unl.edu/chemistrydiss>

 Part of the Chemistry Commons

Fitzgerald, Lisa A., "A COMPARATIVE GENOMIC ANALYSIS OF CHLORELLA NC64A VIRUS NY-2A AND CHLORELLA Pbi VIRUS MT325 FROM THE FAMILY PHYCODNAVIRIDAE" (2006). *Student Research Projects, Dissertations, and Theses - Chemistry Department*. 1.
<https://digitalcommons.unl.edu/chemistrydiss/1>

This Article is brought to you for free and open access by the Chemistry, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Student Research Projects, Dissertations, and Theses - Chemistry Department by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

A COMPARATIVE GENOMIC ANALYSIS OF *CHLORELLA* NC64A VIRUS NY-2A
AND *CHLORELLA* Pbi VIRUS MT325 FROM THE FAMILY *PHYCODNAVIRIDAE*

by

Lisa A. Fitzgerald

A DISSERTATION

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillment of Requirements
For the Degree of Doctor of Philosophy

Major: Chemistry

Under the Supervision of Professor James L. Van Etten

Lincoln, Nebraska

May 2006

A COMPARATIVE GENOMIC ANALYSIS OF *CHLORELLA* NC64A VIRUS NY-2A
AND *CHLORELLA* Pbi VIRUS MT325 FROM THE FAMILY *PHYCODNAVIRIDAE*

Lisa A. Fitzgerald, Ph.D.

University of Nebraska, 2006

Adviser: James L. Van Etten

The family *Phycodnaviridae* consists of a morphologically similar but genetically diverse group of large dsDNA viruses which infect both fresh and marine water eukaryotic algae. Two viruses, the 370 kb *Chlorella* NC64A virus NY-2A and the 313 kb *Chlorella* Pbi virus MT325, from the family *Phycodnaviridae*, genus *Chlorovirus*, were sequenced, analyzed, and compared to the prototype chlorella virus PBCV-1. The NY-2A genome, the largest chlorella viral genome sequenced to date, contains 886 open reading frames (ORFs) of 65 codons or larger and encodes 404 putative proteins and 7 tRNAs. The MT325 genome, the smallest chlorella viral genome sequenced to date, contains 845 ORFs and encodes 331 putative proteins and 10 tRNAs. The protein-encoding genes are evenly distributed on both strands, and the intergenic space is minimal. Approximately 50% of the viral gene products have been identified, including some which are the first of their type to be detected in a virus. Domain scatter plots revealed that NY-2A and MT325 are mosaics of both prokaryotic- and eukaryotic-like proteins. A comparison of the three chlorella viruses sequenced, NY-2A, MT325, and

the previously sequenced PBCV-1, revealed that ~75% of the viral encoded proteins are homologous. However, 64% of these homologs are classified as unknowns based on a lack of sequence similarity to proteins in public databases.

In addition to the genomic annotations, I conducted an extensive comparison of the chloroviruses to other phycodnavirus members as well as other large dsDNA viruses. Surprisingly, only 10 (~3%) of the chlorovirus genes are encoded by all six of the sequenced phycodnaviruses.

TABLE OF CONTENTS

TITLE PAGE.....	i
ABSTRACT.....	ii
TABLE OF CONTENTS.....	iv
LIST OF FIGURES AND TABLES.....	viii
LIST OF ABBREVIATIONS.....	xi
ACKNOWLEDGMENTS.....	xiii
DEDICATION.....	xiv

CHAPTER 1

INTRODUCTION AND BACKGROUND: *PHYCODNA VIRIDAE-* LARGE DNA ALGAL VIRUSES

1.1 INTRODUCTION.....	1
1.2 DISCOVERY OF ALGAL VIRUSES.....	2
1.3 GENERAL CHARACTERISTICS AND TAXONOMY.....	4
1.4 CHLORELLA VIRUSES.....	7
1.5 NATURAL HISTORY OF THE CHLORELLA VIRUSES.....	8
1.6 <i>PARAMECIUM BURSARIA</i> CHLORELLA VIRUS TYPE-1.....	15
1.6.1 The Life Cycle.....	15
1.6.2 PBCV-1 Genome.....	17
1.6.3 PBCV-1-encoded Proteins.....	18
1.7 DIVERSITY OF CHLOROVIRUS GENOMES.....	20
1.8 PHYCODNA VIRUS GENES ARE LIKELY VERY OLD.....	22
1.9 CONCLUSION.....	24
1.10 REFERENCES.....	26

CHAPTER 2

GENOMIC ANALYSES OF *CHLORELLA* NC64A VIRUS NY-2A AND *CHLORELLA* Pbi VIRUS MT325

2.1 INTRODUCTION.....	37
2.2 EXPERIMENTAL METHODS.....	38
2.2.1 Viral DNA Isolation and Sequencing.....	38
2.2.2 Genomic Sequence Analysis.....	39
2.2.2.1 Initial Analysis of Viral Genomes.....	39
2.2.2.2 Identification of Putative tRNAs.....	39
2.2.2.3 Identification and Naming of Open Reading Frames.....	39
2.2.2.4 Identification of Major versus Minor Open Reading Frames.....	40
2.2.2.5 Characterization of Open Reading Frames.....	42
2.2.2.6 Analysis with the Non-redundant Database at NCBI.....	43
2.2.2.7 Analysis with the Protein Families Database.....	43
2.2.2.8 Identification of Clusters of Orthologous Groups.....	43
2.2.2.9 Identification of Inteins and Introns.....	43
2.3 RESULTS.....	45
2.3.1 Analysis and Annotation of the NY-2A Genome.....	45
2.3.1.1 Identification of Putative tRNAs.....	47
2.3.1.2 Identification of Major and Minor Open Reading Frames.....	47
2.3.1.3 Characterization of Open Reading Frames.....	53
2.3.1.4 Annotation of NY-2A Genome.....	59
2.3.1.5 Identification of Inteins and Introns.....	61
2.3.1.6 Identification of Gene Families.....	62
2.3.2 Analysis and Annotation of the MT325 Genome.....	62
2.3.2.1 Identification of Putative tRNAs.....	64
2.3.2.2 Identification of Major and Minor Open Reading Frames.....	65
2.3.2.3 Characterization of Open Reading Frames.....	70
2.3.2.4 Annotation of MT325 Genome.....	76
2.3.2.5 Identification of Inteins and Introns.....	78
2.3.2.6 Identification of Gene Families.....	78
2.4 DISCUSSION.....	79
2.4.1 DNA Replication and Repair-Associated Proteins.....	82
2.4.2 Transcription-Associated Proteins.....	83
2.4.3 Protein Synthesis, Modification, and Degradation.....	84

2.4.4 Nucleotide Metabolism-Associated Proteins.....	85
2.4.5 Protein Kinases, Phosphatases, and Channel Proteins.....	86
2.4.6 Sugar- and Lipid-Manipulating Proteins.....	87
2.4.7 Cell Wall-Degrading Enzymes.....	88
2.4.8 Restriction-Modification Enzymes.....	88
2.4.9 Integration and Transposition Enzymes.....	88
2.4.10 Miscellaneous Proteins.....	89
2.5 REFERENCES.....	91
2.6 ADDENDUM.....	96
2.6.1 Termini of the PBCV-1, NY-2A, and MT325 Genomes.....	96
2.6.1.1 Termini of the PBCV-1 Genome.....	96
2.6.1.2 Termini of the NY-2A Genome.....	99
2.6.1.3 Termini of the MT325 Genome.....	102
2.6.2 NY-2A Genome.....	105
2.6.2.1 General Characteristics of Open Reading Frames.....	105
2.6.2.2 Analysis of Open Reading Frames.....	128
2.6.2.3 Results from Analysis with the Protein Families (Pfam) Database..	162
2.6.2.4 Results of Analysis with Clusters of Orthologous Groups (COGs)..	166
2.6.3 MT325 Genome.....	169
2.6.3.1 General Characteristics of Open Reading Frames.....	169
2.6.3.2 Analysis of Open Reading Frames.....	190
2.6.3.3 Results from Analysis with the Protein Families (Pfam) Database..	222
2.6.3.4 Results of Analysis with Clusters of Orthologous Groups (COGs)..	225

CHAPTER 3

GENOMIC COMPARISONS OF THREE CHLORELLA VIRUSES, PBCV-1, NY-2A, AND MT325

3.1 INTRODUCTION.....	228
3.2 EXPERIMENTAL METHODS.....	229
3.2.1 Comparing the Orientations of the Chlorella Viral Genomes.....	229
3.2.2 Genomic Comparisons.....	229
3.2.3 Domain and Kingdom Scatter Plots.....	230
3.3 RESULTS.....	230
3.3.1 Comparison of the Orientations of the Chlorella Viral Genomes.....	230

3.3.2 Comparison of the Chlorella Virus Open Reading Frames.....	234
3.3.3 Domain and Kingdom Scatter Plots.....	238
3.4 DISCUSSION.....	239
3.5 REFERENCES.....	248
3.6 ADDENDUM.....	250
3.6.1 Homologous ORFs in Chlorella Viruses.....	250
3.6.1.1 PBCV-1 Homologous ORFs.....	250
3.6.1.2 NY-2A Homologous ORFs.....	258
3.6.1.3 MT325 Homologous ORFs.....	268
3.6.2 Domain and Kingdom Scatter Plots.....	277
3.6.2.1 PBCV-1 Domain and Kingdom Scatter Plots.....	277
3.6.2.2 NY-2A Domain and Kingdom Scatter Plots.....	281
3.6.2.3 MT325 Domain and Kingdom Scatter Plots.....	285

CHAPTER 4

COMPARATIVE ANALYSES OF *CHLOROVIRUSES*, *PHYCODNAVIRIDAE*, AND NUCLEO-CYTOPLASMIC LARGE DNA VIRUSES

4.1 INTRODUCTION.....	289
4.2 BACKGROUND.....	290
4.2.1 Characteristics and Taxonomy of Sequenced Phycodnaviruses.....	290
4.2.1.1 Coccolithoviruses.....	290
4.2.1.2 Phaeoviruses.....	291
4.2.2 Nucleo-Cytoplasmic Large DNA Viruses.....	292
4.3 EXPERIMENTAL METHODS.....	294
4.3.1 Comparisons of Phycodnavirus Metabolic Domains.....	294
4.3.2 Analysis of NCLDV Core Genes.....	294
4.4 RESULTS AND DISCUSSION.....	295
4.4.1 Comparisons of the Sequenced Phycodnavirus Genomes.....	295
4.4.2 Comparisons of Phycodnavirus Gene Content.....	297
4.4.3 NCLDV Comparative Studies.....	302
4.5 REFERENCES.....	306

TABLE OF FIGURES AND TABLES

Description	Page
CHAPTER 1	
Figure 1.1 - Phylogenetic tree of DNA polymerase gene fragments from members of the family <i>Phycodnaviridae</i> and other large dsDNA viruses.....	5
Table 1.1 - Taxonomy and general characteristics of some phycodnaviruses.....	6
Figure 1.2 - <i>P. bursaria</i> filled with <i>Chlorella</i> NC64A.....	9
Figure 1.3 - Infection of <i>Chlorella</i> strain NC64A by PBCV-1.....	16
Figure 1.4 - PBCV-1 Partial Gene Map.....	19
CHAPTER 2	
Table 2.1 - Functional Proteins Encoded by PBCV-1.....	41
Figure 2.1 - Genomic Dot Plots Comparing the Genomes of PBCV-1 to NY-2A....	46
Table 2.2 - NY-2A tRNAs.....	47
Table 2.3 - NY-2A ORFs Reclassified from a Major to a Minor.....	50
Table 2.3a - NY-2A ORFs Changed from a Major to a Minor Based on Criterion (ii).....	50
Table 2.3b - NY-2A ORFs Changed from a Major to a Minor Based on Criterion (ii).....	51
Table 2.3c - NY-2A ORFs Changed from a Major to a Minor Based on Criterion (iv).....	51
Table 2.4 - NY-2A ORFs Reclassified from a Minor to a Major.....	52
Table 2.4a - NY-2A ORFs Changed from a Minor to a Major Based on Criterion (v).....	52

Table 2.4b - NY-2A ORFs Changed from a Minor to a Major Based on Criterion (vi).....	52
Figure 2.2 - Orientation of NY-2A ORFs.....	55
Figure 2.3 - Size of NY-2A ORFs.....	56
Figure 2.4 - Analysis of the pIs from the NY-2A ORFs.....	57
Figure 2.5 - Intergenic Space Between Major NY-2A ORFs.....	58
Table 2.5 - NY-2A Analysis with the NR Database.....	59
Table 2.6 - NY-2A COGs.....	60
Figure 2.6 - Genomic Dot Plots Comparing the Genomes of PBCV-1 to MT325....	63
Table 2.7 - MT325 tRNAs.....	64
Table 2.8 - MT325 ORFs Reclassified from a Major to a Minor.....	67
Table 2.8a - MT325 ORFs Changed from a Major to a Minor Based on Criterion (ii).....	67
Table 2.8b - MT325 ORFs Changed from a Major to a Minor Based on Criterion (iii).....	68
Table 2.8c - MT325 ORFs Changed from a Major to a Minor Based on Criterion (iv).....	68
Table 2.9 - NY-2A ORFs Reclassified from a Minor to a Major.....	69
Table 2.9a - MT325 ORFs Changed from a Minor to a Major Based on Criterion (v).....	69
Table 2.9b - MT325 ORFs Changed from a Minor to a Major Based on Criterion (vi).....	70
Figure 2.7 - Orientation of MT325 ORFs.....	72
Figure 2.8 - Size of MT325 ORFs.....	73
Figure 2.9 - Analysis of the pIs from the MT325 ORFs.....	74
Figure 2.10 Intergenic Space Between Major MT325 ORFs.....	75
Table 2.10 - MT325 Analysis with the NR Database.....	76
Table 2.11 - MT325 COGs.....	77

Figure 2.11 NY-2A Partial Gene Map.....	80
Figure 2.12 MT325 Partial Gene Map.....	81

CHAPTER 3

Figure 3.1 - Genomic Dot Plots of Chlorella Viruses Based on DNA Analysis (blastn).....	231
Figure 3.2 - Genomic Dot Plots of Chlorella Viruses Based on a Translational Analysis (tblastx).....	232
Figure 3.3 - Dot Plots of the NY-2A or the MT325 genomes.....	233
Figure 3.4 - Genomic Lineups of the Three Sequenced Chlorella Viral Genomes...	234
Figure 3.5 - Genomic ORF Comparisons of Chlorella Viruses.....	237
Figure 3.6 - Domain Scatter Plots of the Chlorella Viruses.....	238
Figure 3.7 - Genomic Dot Plots of MT325 and a NC64A Virus.....	240
Table 3.1 - Comparisom of NC64A Viral ORFs to the MT325 Viral ORFs.....	241
Figure 3.8 - Comparisom of Putative Proteins Encoded by Chlorella Viruses.....	244

CHAPTER 4

Table 4.1 - Genome Data of Sequenced Phycodnaviruses.....	295
Figure 4.1 - Selected ORFs in the Sequenced Phycodnavirus Genomes.....	301
Figure 4.2 - Comparisom of NCLDV Core Genes.....	303

LIST OF ABBREVIATIONS

ASFV.....	African swine fever virus
BLAST.....	Basic Local Alignment Search Tool
blosum.....	Blocks substitution matrix
CbV-PW1.....	<i>Chrysochromulina brevifilum</i> virus PW1
COG.....	Cluster of Orthologous Groups
CVM-1.....	Chlorella virus Marburg
dCMP deaminase.....	deoxycytidylate deaminase
dsDNA.....	double-stranded DNA
dUTPase.....	dUTP pyrophosphatase
EhV-86.....	<i>Emiliania huxleyi</i> virus 86
EsV-1.....	<i>Ectocarpus siliculosus</i> virus
FirrV-1.....	<i>Feldmannia irregularis</i> virus
GCG.....	Genetics Computer Group
GDP.....	Guanosine monophosphate
HaV01.....	<i>Heterosigma akashiwo</i> virus 01
InBase.....	Intein Database and Registry
IS.....	Insertion Sequence
kb.....	kilobases
KcV.....	Chlorella virus potassium channel
LPS.....	Lipopolysaccharide
m5C.....	5-methylcytosine
m6A.....	N ⁶ -methyladenine
Mb.....	Megabases
MpV-SP1.....	<i>Micromonas pusilla</i> virus SP1

mRNA.....	Messenger RNA
NCBI.....	National Center for Biotechnology Information
NCLDV.....	Nucleo-Cytoplasmic Large DNA Viruses
NR.....	Non-redundant
NTP.....	Nucleoside 5'-triphosphates
ODC.....	Ornithine Decarboxylase
ORF.....	Open reading frame
p.i.	Post-infection
PBCV-1.....	<i>Paramecium bursaria</i> chlorella virus-1
PCNA.....	Proliferating Cell Nuclear Antigen
PDG.....	Pyrimidine dimer-specific glycosylase
Pfam.....	Protein Families
PFU.....	Plaque forming units
pI.....	Isoelectric point
SKP.....	Seventeen kilodalton protein
TFIIB.....	Transcription factor IIB
TFIIS.....	Transcription factor IIS
TFIIS.....	Transcription factor IIS
TIGR.....	The Institute for Genomic Research
tRNA.....	Transfer RNA
UDP.....	Uridine diphosphate
vSET.....	PBCV-1 SET domain-containing protein

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Jim Van Etten for taking me into his lab and teaching me only a fraction of what he knows. Thanks for always believing in me and keeping me on the right track. You are not only an amazing scientist but also a great role model!

I would also like to thank Dr. Roy French, Dr. David Hage, and Dr. Larry Parkhurst for taking time to serve on my graduate committee. Furthermore, I would like to thank Dr. Mark Griep for not only serving as a committee member but helping me through all of the qualifying exams. You provided valuable assistance during my RUI and OPO.

A very special thank you goes to my family. I wouldn't be the person I am today without the love and support of each member. My parents allowed me to grow up in a strong Christian family home and I've had two perfect examples to follow. Mom and Dad, I love you! Believe it or not, but you have sent me 97 e-cards and 144 motivational e-mails while I was writing (bet you didn't think I was keeping them all!) Even though they aren't here to see me graduate, I'd like to thank my grandparents for all their love and for being positive role models. Cheb you're amazing!! You are always willing to listen to me no matter when I call. B__ you definitely know how to have a good time! Thanks to you I can now say I've survived a hurricane. I love you guys!!!

There are several very close friends that I have to thank:

- Steve Schuetz- Starting with the best Halloween party I've ever attended to being able to just visit for New Years, we always know how to have a great time! I enjoy all of our late night conversations and I would have to say that you are the best friend that anyone could have...THANKS for always being there for me!!!
- Justin Biffinger- Even though we had our ups and downs, thanks for all of the good times! Especially all the times that you were there to help make me see the brighter side of life!
- Amy Ziems- Countless times I've come into your lab to vent! You're a great listener and are always there to help me talk through situations. I'm glad that you finally thought I was friendly enough to approach!
- Lindsay Salem- What can I say...you have helped me through some of my toughest times! You always know how to put a smile back on my face!

A big thank you goes out to the Van Etten lab members (past and present), especially for putting up with me the past few months! Lab Mates: Irina Agarkova, Philip Boucher (you're awesome), Tae-Jin Choi, Garry Duncan, David Dunigan, Mike Graves, Tiara Harms (partner in crime), James Gurnon (always willing to help), Ming Kang (never afraid to speak his mind), Paola Valbuzzi (amazing tour guide...thanks for showing me around Italy), Giane Yanai, Varvara Yashchenko, and Yuanzheng Zhang (a.k.a. my man).

DEDICATED TO MY FAMILY
FOR ALL THEIR LOVE AND SUPPORT

Mom and Dad
Jake, Julie, Janae, Jaycee, and Jayla Adams
David, Tricia, and the future Erin Fitzgerald



CHAPTER 1

INTRODUCTION AND BACKGROUND: **PHYCODNAVIRIDAE- LARGE DNA ALGAL VIRUSES**

1.1 INTRODUCTION

Members and potential members of the *Phycodnaviridae* family are a genetically diverse, but morphologically similar, group of viruses which infect both fresh and marine water eukaryotic algae. The family name is derived from its members' two defining characteristics: i) "phyco" from their algal hosts, and ii) "dna" from their double-stranded DNA (dsDNA) genomes (1). The phycodnaviruses are among the viriplankton which have recently been recognized as important ecological elements in aqueous environments (2). For example, phycodnaviruses, along with other viruses, play important roles in the dynamics of algal blooms, nutrient cycling, algal community structure, and possibly gene transfer between organisms (3-6).

The ease with which new aquatic viruses are found suggests that they may represent the greatest source of uncharacterized genetic diversity on Earth (7). These new aquatic viruses, including the phycodnaviruses, are constantly being discovered as more diverse environmental samples are collected. Ongoing metagenomic studies using massive DNA sequencing also indicates a greater viral diversity than could have been imagined just a few years ago (7, 8). The genetic diversity that exists in the phycodnaviruses, albeit with only a few completed sequences, indicates that the limits of diversity are nowhere in sight. Our appreciation for this diversity is limited, in part, because phycodnaviruses have only been identified in about 0.1% of the ~40,000 known

eukaryotic algal species (9). Paradoxically, it may be that both biological and genetic diversity are the unifying themes of the phycodnaviruses.

Accumulating genetic evidence suggests that members from the family *Phycodnaviridae* are ancient. These phycodnaviruses, together with the poxviruses, iridoviruses, African swine fever virus (ASFV), and the recently discovered Mimivirus share a common evolutionary ancestor which may have arisen at the point of eukaryogenesis, approximately 2.0-2.7 billion years ago (10, 11). They all share nine gene products and at least two of these viral families encode an additional 41 homologous gene products (10, 12, 13). Collectively, these viruses are referred to as nucleocytoplasmic large DNA viruses (NCLDV) (13).

The genomes of a few *Phycodnaviridae* family members have been sequenced, and this information is contributing to our understanding of the relationships of these viruses with other family members as well as with other large DNA viruses and cellular organisms. This chapter focuses on the discovery and characterization of the *Phycodnaviridae* family as well as in depths look into the genus *Chlorovirus* and the prototype virus, *Paramecium bursaria* chlorella virus-1 (PBCV-1).

1.2 DISCOVERY OF ALGAL VIRUSES

Since the early 1970s, viruses or virus-like particles have been reported in at least 44 taxa of eukaryotic algae, which composes 10 of the 14 algal classes (14). However, most of the early reports described isolated accounts of microscopic observations, and the virus particles were not characterized further for a variety of reasons. This situation changed in the early 1980s with the discovery of large dsDNA-containing viruses that

infect and replicate in certain strains of unicellular, eukaryotic, exsymbiotic chlorella-like green algae (also referred to as zoochlorellae) and the more recent identification of tractable viral systems in brown algae.

In 1978, Kawakami & Kawakami (15) described the appearance of large (180-nm diameter), lytic viruses in zoochlorellae after the algae were released from the protozoan *Paramecium bursaria*. No virus particles were detected in zoochlorellae growing symbiotically inside the paramecium. Independently, lytic viruses were described in zoochlorellae isolated from the green coelenterate *Hydra viridis* (16, 17) and also from *P. bursaria* (18). Consistent with the previously-described report, viruses appeared only after the zoochlorellae were separated from their host. Fortunately, the zoochlorellae from *P. bursaria* had previously been cultured free from their host, and further experiments revealed that these cultured algae serve as hosts for many closely related viruses. These lytic chlorella viruses can be produced in large quantities and assayed by plaque formation using standard bacteriophage techniques (19, 20).

In addition to the fresh water *Chloroviruses*, a few additional large icosahedral, dsDNA-containing viruses which infect marine algae are currently being studied. At least one member of the remaining five genera from the family *Phycodnaviridae* is under examination. The genus *Coccolithovirus* includes the *Emiliania huxleyi* virus 86 (EhV-86) as its type species and this virus is the largest phycodnavirus sequenced to date (21). The *Phaeoviruses* are viruses that infect filamentous brown algae, *Phaeophyceae*. There are two members of this genus which have been sequenced and are under current investigation. One virus member infects *Ectocarpus* sp. (EsV-1) (22) and the other infects *Feldmannia* sp. (FirrV-1) (23). The remaining three genera, *Prasinovirus*,

Prymnesiovirus, and *Raphidovirus*, are not as well characterized as the other three *Phycodnaviridae* genera, but each genera contains at least one member which is currently under investigation. These include viruses which infect the unicellular algae, *Micromonas pusilla* (MpV viruses) (24, 25), *Chrysochomulina brevifilum* (CbV-PW1 virus) (26), and *Heterosigma akashiwo* (HaV viruses) (27, 28). The chlorella viruses and the marine algal viruses are found throughout nature and have been isolated from both fresh water and seawater collected from around the world.

1.3 GENERAL CHARACTERISTICS AND TAXONOMY

Phycodnaviruses are large (mean diameter of 160+-45 nm) icosahedron structures which encapsidate large (160 to 560 kb) dsDNA genomes. These viruses have an internal membrane that is required for infection. Phylogenetic analyses of the δ-DNA polymerases of the phycodnaviruses indicate that these viruses are more closely related to each other than to other dsDNA viruses and that they form a monophyletic group, suggesting a common ancestor (Figure 1.1) (29). However, these viruses can be divided into six clades which correlate with their hosts, and each has been given genus status. Often the genera can be distinguished by additional properties (e.g., lytic vs. lysogenic life styles or linear vs. circular genomes) (29). For example, members of the genus *Chlorovirus* infect fresh water algae, whereas, members of the other five genera (*Coccolithovirus*, *Phaeovirus*, *Prasinovirus*, *Prymnesiovirus*, and *Raphidovirus*) infect marine algae. The general characteristics of the six genera have been described (30), and Table 1.1 summarizes selected properties of the genera.

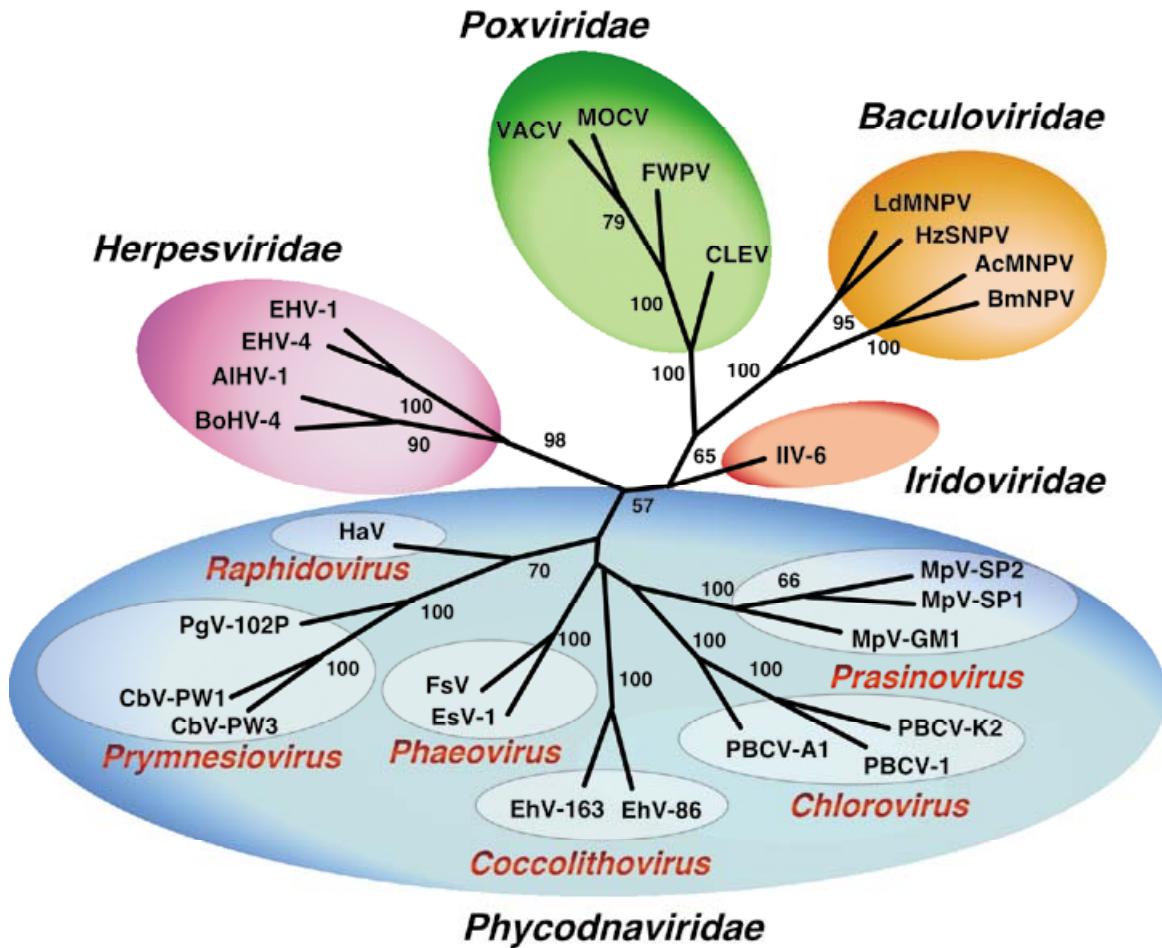


Figure 1.1 – Phylogenetic tree of DNA polymerase gene fragments from members of the family *Phycodnaviridae* and other large dsDNA viruses. Abbreviations are: *Phycodnaviridae*: HaV, Heterosigma akashiwo virus; CbV-xx, viruses that infect Chrysochromulina brevifilum; PgV-102P, Phaeocystis globosa virus 102 (Plymouth); FsV, Feldmannia sp. virus; EsV-1, Ectocarpus siliculosus virus 1; EhV-xx, viruses that infect *Emiliania huxleyi*; MpV-xx, viruses that infect *Micromonas pusilla*; PBCV-xx, viruses that infect *Chlorella* NC64A; *Herpesviridae*: EHV-1, *Equine herpesvirus* 1; EHV-4, *Equine herpesvirus* 4; BoHV-4, *Bovine herpesvirus* 4; AIHV-1, *Alcelaphine herpesvirus* 1; *Poxviridae*: VACV, *Vaccinia* virus; MOCV, *Molluscum contagiosum* virus; FWPV, *Fowl poxvirus*; CLEV, *Choristoneura biennis entomopoxvirus*; *Baculoviridae*: AcMNPV, *Autographa californica nucleopolyhedrovirus*; BmNPV, *Bombyx mori nucleopolyhedrovirus*; HzSNPV, *Helicoverpa zea nucleopolyhedrovirus*; LdMNPV, *Lymantria dispar nucleopolyhedrovirus*; *Iridoviridae*: IIV-6, *Invertebrate iridescent virus* 6. Figure was taken from Wilson et al., 2005 (1).

Table 1.1 Taxonomy and general characteristics of some phycodnaviruses

Genus ^a	Type Species ^a	Known Host Range ^a	Source	Particle Diameter (nm)	Genome Size (kbp)
<i>Chlorovirus</i>	<i>Paramaecium bursaria</i> chlorella virus-1 (PBCV-1)	<i>Chlorella NC64A</i> <i>Chlorella Pbi</i> <i>Hydra viridis</i>	FW	190	313-370
<i>Coccolithovirus</i>	<i>Emiliania huxleyi</i> virus 86 (EhV-86)	<i>Emiliania huxleyi</i>	MW	160 - 200	407-415
<i>Phaeovirus</i>	<i>Ectocarpus siliculosus</i> virus-1 (EsV-1)	<i>Phaeophyceae</i> <i>Ectocarpus siliculosus</i> <i>Ectocarpus fasciculatus</i> <i>Feldmannia simplex</i> <i>Feldmannia irregularis</i> <i>Feldmannia species</i> <i>Hinckssia hinckssiae</i> <i>Myriotrichia clavaeformis</i> <i>Pilayella littoralis</i>	MW	130 – 200	160-340
<i>Prasinovirus</i>	<i>Microonas pusilla</i> virus SP1 (MpV-SP1)	<i>Micromonas pusilla</i> <i>Pyramimonas orientalis</i>	MW	115 - 200	200 - 560
<i>Prymnesiovirus</i>	<i>Chrysomonulina brevifilum</i> virus PW1 (CbV-PW1)	<i>Haptophyceae (aka</i> <i>Prymnesiophyceae)</i> <i>Chrysomonulina brevifilum</i> <i>Chrysomonulina strobilus</i> <i>Chrysomonulina globosa</i>	MW	120 - 160	485 - 510
<i>Raphidovirus</i>	<i>Heterosigma akashiwo</i> virus 01 (HaV01)	<i>Heterosigma akashiwo</i>	MW	202	294

a, data abstracted from <http://www.ncbi.nlm.nih.gov/ICTVdb/Ictv/index.htm>
 FW - fresh water; MW - marine/coastal water

1.4 CHLORELLA VIRUSES

Chlorella viruses are large (1900 Å along the fivefold axes), icosahedral, plaque-forming, dsDNA viruses which infect certain unicellular, chlorella-like green algae. The chlorella viruses are classified in the genus *Chlorovirus* which consists of four species (31): (i) viruses that infect *Chlorella* NC64A (NC64A viruses), (ii) viruses that infect *Chlorella* Pbi (Pbi viruses), (iii) viruses that infect symbiotic chlorella in the coelenterate *Hydra viridis* and (iv) the newly discovered group of viruses that infect endosymbiotic *Chlorella* algae of the helizoon *Acanthocystis turfacea* (32). The chlorella host *Hydra viridis* has not been cultured free of the virus, and, consequently, the virus can be isolated only from chlorella cells freshly released from the hydra.

Algae included in the genus *Chlorella* are small, spherical or ellipsoidal, unicellular, nonmotile, asexually reproducing green algae and are among the most widely distributed and frequently encountered algae on earth (33, 34). Chlorella species have a rigid cell wall and typically have a single chloroplast, which sometimes contains a pyrenoid body. They have a simple developmental cycle and reproduce by mitotic division. Vegetative cells increase in size and, depending on the species and the environmental conditions, divide into two, four, or eight or more progeny that are released via the rupture or enzymatic digestion of the parental wall. Electrophoretic karyotyping of several chlorella isolates, including *Chlorella* NC64A, indicates that these algae contain multiple chromosomes ranging in size from ~1 Mb to ~6.5 Mb and have a total genome size of ~40 Mb.

Algae assigned to the genus *Chlorella* are more heterogeneous than their simple morphology suggests. Because of this diversity, characteristics observed in one

Chlorella species or isolate may not be observed in other species. This degree of heterogeneity is illustrated by the following two examples: (i) The G + C content of the nuclear DNA of *Chlorella* species ranges from 43% to 79% (35). However, most isolates assigned to the same species have similar G + C content. The PBCV-1 host, *Chlorella* NC64A, has a G + C content of 67%. The genome of *Chlorella* NC64A also contains methylated nucleotides (21% of the cytosines are 5-methylcytosine (m5C) and 0.6% of the adenines are N6-methyladenine (m6A)) (36). (ii) The cell wall polysaccharides of the *Chlorella* species vary widely (37-39) even among isolates assigned to the same species (39). Unexpectedly, lipopolysaccharide (LPS)-like components were identified recently in the cell walls of some chlorella isolates (40). This finding was unexpected given that LPS is typically only found in Gram-negative bacteria.

Most chlorella species are naturally free-living. However, some live as hereditary endosymbionts with freshwater and, to lesser extent, marine animals (41). The only known hosts for the *Chlorella* viruses are symbiotic chlorella, some of which can be cultured.

1.5 NATURAL HISTORY OF THE CHLORELLA VIRUSES

Several unexpected discoveries demonstrating the diversity of the chlorella viruses and their genomes are discussed below. Because the natural history of these viruses is poorly understood, major advances in our understanding of the biology of these viruses are needed before we can fully appreciate and explain their evolutionary origin.

The two known cultured hosts for the chlorella viruses, *Chlorella* NC64A and *Chlorella* Pbi, typically exist as hereditary endosymbionts in green isolates of the



Figure 1.2 – *P. bursaria* filled with the symbiotic alga (*Chlorella* NC64A)

protozoan *P. bursaria* (Figure 1.2). Interestingly, zoochlorella are resistant to viral infection when they are in the symbiotic relationship with the ciliate (42). In this symbiotic unit, algae are enclosed individually in perialgal vacuoles surrounded by a host-derived membrane (43). The initial establishment of a successful symbiotic relationship and the long-term maintenance of this symbiosis require that the algae resist digestion by the paramecium. Reassociation

studies with different *Chlorella* species and alga-free *P. bursaria* indicate that only the original symbiotic algae is able to readily reestablish symbiosis with the ciliate (43) while other chlorellae are digested. Although there have been numerous studies and considerable speculation on the factor(s) which allow the *P. bursaria* to distinguish chlorellae that are suitable for symbiosis from those that are not suitable, this specific recognition factor(s) remains unidentified.

Chlorella NC64A and *Chlorella* Pbi were originally isolated from American and European *P. bursaria* strains, respectively. Viruses that infect *Chlorella* NC64A can neither infect nor attach to *Chlorella* Pbi and vice versa (42). Because the viruses can distinguish between these two *Chlorella* isolates, it was hypothesized that the viral receptor might also serve as the recognition factor for the paramecium. However, this hypothesis is incorrect since both *Chlorella* NC64A and *Chlorella* Pbi are able to reestablish stable symbiotic relationships with either American or European isolates of *P. bursaria*.

NC64A viruses have been isolated from fresh water collected in the United States (44-46), China (47), Japan, Brazil (48), Australia (JL Van Etten, J Rohozinski, unpublished results), Argentina, and Israel (Y Zhang, M Nelson, JL Van Etten, unpublished results). Pbi viruses, while initially found in fresh water collected in Europe (49, 50), have more recently been isolated from water collected in Australia, Canada, northern United States (Minnesota, Wisconsin, and Montana), and in higher altitudes of the western United States (M Nelson, Y Zhang, JL Van Etten, unpublished results). Therefore, the initial assumption that Pbi viruses were limited to Europe and the NC64A viruses were limited to the Americas and Eastern Asia is incorrect. In fact, two water samples collected in Australia contained both NC64A and Pbi viruses (JL Van Etten, J Rohozinski, unpublished results). Thus, the most important factors influencing the distribution of NC64A and Pbi viruses are probably latitude and altitude.

Typically, the chlorella virus titer in nature is 1–100 plaque forming units (PFU)/ml, but titers as high as 100,000 PFU/ml have been obtained. The natural concentration of chlorella viruses is not static but rather fluctuates with the seasons, with the highest titers typically found in late Spring (46, 48). However, a water sample collected from an ice-covered pond in Ohio contained a titer of >1000 PFU/ml (M. Nelson, unpublished results).

The high titers of chlorella viruses in some indigenous waters are surprising given that they are constantly exposed to solar radiation which should damage the viral DNA and inactivate the virus. For instance, inactivation of bacteriophages and cyanophages occurs at a rate of 0.4–0.8% per hour in full sunlight (51, 52). The chlorella viruses, however, have apparently adapted to exposure to solar radiation by utilizing two

independent DNA repair systems (53). First, PBCV-1 encodes a pyrimidine dimer-specific glycosylase (PDG) that initiates the repair of UV-induced thymidine dimers. This DNA repair system functions in both light and dark. Second, PBCV-1 uses its host's photolyase to repair UV-induced thymidine dimers. Therefore, PBCV-1 can replicate whenever it infects a suitable host in either day- or nighttime. The *pdg* gene, which encodes this enzyme, is expressed early after virus infection. While the PDG enzyme was expected to be packaged in the virion and accompany the virus DNA into the host where it could initiate DNA repair, attempts to detect the PDG protein in PBCV-1 virions were unsuccessful.

PBCV-1 also encodes two other enzymes that may aid in its survival in direct sunlight. First, the virus encodes a putative Cu/Zn superoxide dismutase that protects DNA from reactive oxygen species. Second, the viral polyamine spermine has been shown recently to function as a free-radical scavenger, thereby reducing DNA damage (54). These findings possibly explain why PBCV-1 encodes four polyamine biosynthetic enzymes (homospermidine synthase, N-carbamoylputrescine amidohydrolase, O-methyltransferase, and ornithine decarboxylase), however, these enzymes are not packaged in the virion.

The PBCV-1-encoded PDG repair enzyme has an additional interesting feature. It cleaves both *cis-syn* and *trans-syn-II* cyclobutane pyrimidine dimers, whereas its bacteriophage T4 homolog endonuclease V cleaves only the *cis-syn* cyclobutane pyrimidine isomer (55). This finding prompted the examination of *pdg* genes from 41 other NC64A viruses isolated from diverse geographic regions. All the chlorella viruses examined contained the *pdg* gene (53). However, the genes from 15 of the 41 viruses

contained a 98-nucleotide perfectly conserved intron. Another 4 viruses contained an 81-nucleotide intron at the same position as the 98-nucleotide intron, and one contained an intron that differed from the other three introns by one nucleotide. Surprisingly, the *pdg* gene exons were less conserved than their corresponding introns. The introns contain 5'-AG/GTATGT and 3'-TTGCAG/AA splice site sequences, which are characteristic of nucleus-located, spliceosomally processed pre-mRNA introns.

There was no obvious geographic correlation between the *pdg* intron-containing and intron-lacking viruses. Of the 98-nucleotide intron-containing viruses, 13 were collected throughout the United States between 1983 and 1985. Viruses isolated from Australia in 1995 and Argentina in 1996 contained an identical intron. Interestingly, some water samples contained both 98-nucleotide intron-containing and intron-lacking viruses. For example, water samples collected in Massachusetts, North Carolina, Alabama, Illinois, and California had both intron-containing and intron-lacking viruses. The perfectly conserved 98-nucleotide intron sequence of 15 viruses and the near perfectly conserved 81-nucleotide intron sequence of another 4 viruses imply that either the intron was acquired recently or there is strong selective pressure to maintain its DNA sequence once it is present within the *pdg* gene. The abilities of intron-containing and intron-lacking viruses to repair UV-damaged DNA in the dark are indistinguishable (56). These findings contradict the widely accepted dogma that intron sequences are more variable than exon sequences.

Similar to the sporadic appearance of the highly conserved intron within the *pdg* gene, the PBCV-1 transcription factor TFIIS-like gene contains a 400-nucleotide self-splicing group IB intron (57). The group IB introns were originally discovered in two

related NC64A viruses, CVU1 and CVB11 (58). The intron in CVU1 has 98% nucleotide identity and is in the same position in a TFIIS-like gene as the PBCV-1 intron. The CVB11 intron has 80% nucleotide identity and is located within an unidentified open reading frame. Three other NC64A viruses isolated in Japan at the same time as the CVU1 and CVB11 viruses lack the intron. It is interesting that two NC64A viruses, PBCV-1 isolated in the United States in 1981 and CVU1 isolated in Japan around 1990, have nearly identical introns located at the same position within the same gene, while this same intron can be found in a different gene of a different virus (CVB11) or be absent in other Japanese chlorella viruses.

Recently, several hundred additional NC64A viruses isolated in Japan have been screened for the presence of the self-splicing intron. The results demonstrated that the intron is present in ~8% of these isolates (59). It was inserted within the TFIIS-like gene in ~60% of the isolates and, in the other ~40%, the intron was within the same unidentified reading frame gene as described for CVB11. In a few of the viruses, the major capsid protein gene contained the intron, and a few of the viruses had two copies of the intron. Nucleotide sequence analysis of these introns and their flanking regions indicated that the intron sequences are under strong selective pressure possibly mediated by the exons (i.e. introns within the same gene had >99% sequence identity, whereas introns within different genes were only 72-78% identical).

The variable levels of methylated bases observed in the genomes of different chlorella viruses are another example of the natural diversity of these viruses. The levels of methylated bases in the viral genomes range from 0.1% 5mC and undetectable levels of 6mA to 45% 5mC and 37% 6mA (60). The concentrations of 5mC and 6mA in the

genomes have been used as a criterion for classifying viruses into groups. Similar to the situation with the introns described above, there is no obvious geographic correlation for the methylation levels of viral DNAs. Typically, viruses isolated from one water sample have similar levels of methylated bases. However, exceptions are common. For example, four of six plaques originally picked from a water sample collected in New York fell into different classes. The level of methylation in these viruses ranged from 0.4% 5mC and no detectable 6mA for virus NY-2C to 45% 5mC and 37% 6mA for virus NY-2A.

To date, the PBCV-1 genome is the only chlorella virus genome to be completely sequenced, annotated, and published. However, two additional chlorella viruses have recently been sequenced, which is the subject of this dissertation, and several more are underway. A detailed physical map of virus CVK2 is available (K Nishida, Y Kimura, T Kawasaki, M Fujie, T Yamada, published results, (61)), and portions of several other NC64A viruses have also been sequenced. Comparisons of these viruses have resulted in the unexpected finding that "anonymous" genes are often either inserted between or deleted from otherwise colinear genes. Two examples illustrate this phenomenon. The chitosanase gene, a late gene, has been characterized in the NC64A virus CVK2 (62). PBCV-1 encodes the same chitosanase gene, but it is flanked by two ORFs that also exist in CVK2. However, CVK2 encodes an ORF that is absent in PBCV-1 and is inserted immediately downstream of its chitosanase gene. Likewise, an ORF which is absent in PBCV-1 is inserted between two otherwise colinear genes in the NC64A virus SC-1A (63). While the significance of the presence and/or absence of these viral genes are not understood, this finding indicates that the total number of genes encoded by the chlorella

virus group is greater than that of any one isolate. This phenomenon is similar to the above-described situation of introns within certain viral genes.

1.6 PARAMECIUM BURSARIA CHLORELLA VIRUS TYPE-1

1.6.1 The Life Cycle

PBCV-1 infects its host by attaching rapidly, specifically, and irreversibly to the external surface of the algal cell wall (64). Attachment always occurs at a virus vertex, possibly with hair-like appendages (65). The determinants for host range are associated with attachment specificity. As mentioned previously, NC64A viruses do not attach to *Chlorella* Pbi cells and Pbi viruses do not attach to *Chlorella* NC64A cells. After the virus is attached, the host cell wall is degraded at the viral attachment site (Figure 1.3A) (64). Unlike other phycodnaviruses, the chloroviruses encode enzymes that are involved in polysaccharide degradation, which may play a role in attachment-associated cell wall digestion. Following host cell wall degradation, the internal membrane of the virus most likely fuses with the membrane of the host cell. This fusion results in viral DNA and virion-associated protein entry into the host cell and leaves an empty capsid on the host cell surface (Figure 1.3B). Infection leads to a rapid depolarization of the host cell membrane (66, 67), which is hypothesized to be caused by a virus-encoded potassium ion channel (called Kcv) located in the internal virion membrane. Presumably, this ion channel is activated when the viral and host membranes fuse. Such rapid depolarization may aid in the release of viral DNA into the host cell and/or restrict subsequent infection of the host cell by additional viruses.

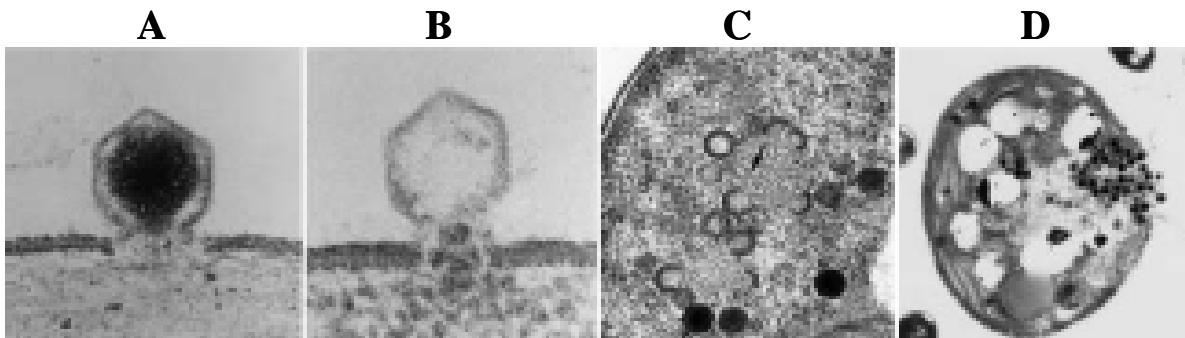


Figure 1.3 – Infection of *Chlorella* strain NC64A by PBCV-1 **A.** Digestion of the wall at the point of attachment. **B.** The contents of the virus are released into the cell. **C.** Cytoplasmic virus assembly centers at 3-4 p.i. **D.** Release of progeny virus by lysis of host cell at 6-8 p.i.

Circumstantial evidence suggests that, during an infection, viral DNA and its associated proteins quickly move to the nucleus where early transcription is detected within 5-10 minutes post-infection (p.i.) (68). Interestingly, data indicate that within minutes of infection, degradation of host chromosomal DNA begins, an effect likely mediated by virion-associated site-specific endonucleases (Agarkova, et al., manuscript in preparation). This degradation may be involved in the recycling of nucleotides from the host cell DNA into the newly replicating viral DNA as well as in the inhibition of host transcription.

In the immediate-early phase of infection, the host cell is reprogrammed to transcribe viral RNAs. While little is known as to how this is mediated, chromatin remodeling may be involved. PBCV-1 encodes a 119 amino acid, SET domain-containing protein (referred to as vSET) that di-methylates Lys27 of histone 3 (69). vSET is packaged in the PBCV-1 virion, and accumulating evidence suggests that vSET may be involved in the repression of host transcription following PBCV-1 infection (Manzur et al., unpublished results).

Viral DNA replication begins 60-90 minutes p.i. and is followed by transcription of late viral genes (68, 70). Ultrastructural studies of PBCV-1-infected chlorella suggest that the host cell nuclear membrane remains intact, at least during the early stages of viral replication (71). However, a functional host cell nucleus is not essential for viral replication since PBCV-1 can replicate, albeit poorly and with a small burst size, in UV-irradiated host cells (72). Approximately 2-3 hours p.i., assembly of viral capsids begins in localized regions in the cytoplasm, called virus assembly centers, which become prominent at 3-4 hours p.i. (Figure 1.3C) (71). By five hours p.i., the cytoplasm is filled with infectious progeny virus particles [~1000 particles/cell, (73)], and, by 6-8 hours p.i., localized lysis of the host cell releases those progeny (Figure 1.3D). Of the progeny released, 25-50% of the virus particles are infectious (i.e. each infected cell yields ~350 plaque-forming units) (73). Intact, infectious PBCV-1 particles accumulate inside the host 30-40 minutes prior to release. Some chlorella viruses have replication cycles longer than that of PBCV-1. For example, NC64A virus NY-2A has an 18 hour replication cycle and, consequently, forms smaller plaques than PBCV-1

1.6.2 PBCV-1 Genome

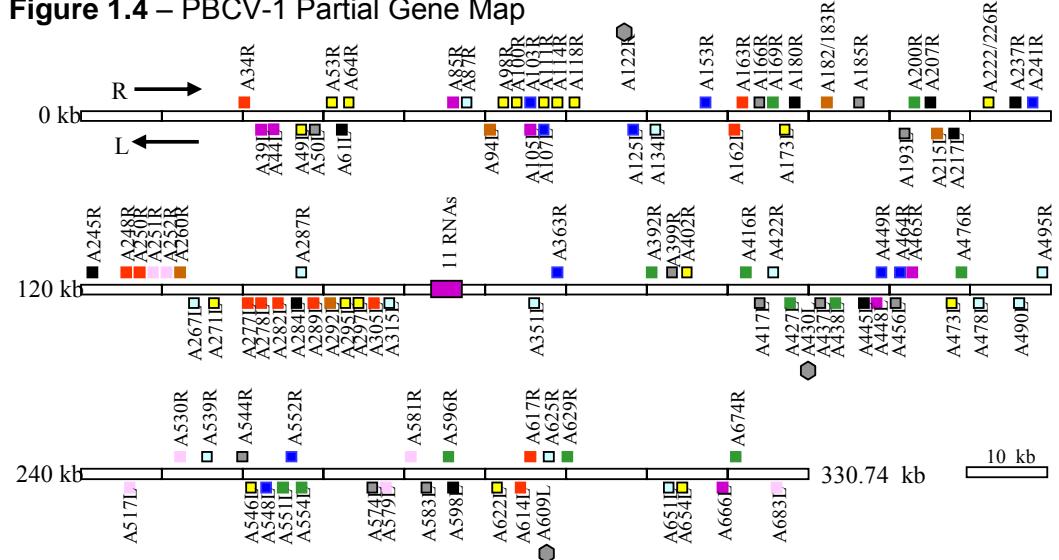
The PBCV-1 genome is a linear, 330 kb, non-permuted dsDNA with 35 nucleotide-long, partially base-paired, covalently closed hairpin termini (74, 75). A 2,221 bp inverted repeat region is present adjacent to each termini (76), but the rest of the genome is primarily single copy DNA sequence. The PBCV-1 genome contains ~700 open reading frames (ORFs) of 65 or more codons, and ~370 of these ORFs likely encode proteins. The putative protein-encoding genes are evenly distributed on both

strands, and the intergenic space is minimal, 275 ORFs are separated by less than 100 nucleotides. Surprisingly, there is a 1788-bp region located near the middle of the genome which contains 11 tRNA genes.

Approximately half of the putative 370 PBCV-1 gene products have been identified and a partial gene map is shown in Figure 1.4. They consist of both prokaryotic- and eukaryotic-like proteins, many of which have not been associated with viruses before. Eighty-four ORFs have homologs within PBCV-1, forming 26 groups. The size of these groups ranges from two to six members.

1.6.3 PBCV-1-encoded Proteins

Many PBCV-1-encoded enzymes are either the smallest or among the smallest proteins of their class. In addition, homologous genes in the chloroviruses can differ in nucleotide sequence by as much as 50%, which translates into amino acid differences of 30 to 40%. Therefore, comparative gene sequence analyses can identify conserved amino acids in proteins as well as regions that tolerate amino acid changes. Their small sizes and the finding that many virus-encoded proteins are “user friendly” have resulted in the biochemical and structural characterization of several PBCV-1 enzymes, such as i) the smallest eukaryotic ATP-dependent DNA ligase (77), which is the subject of intensive mechanistic and structural studies [(78) and references cited therein], ii) the smallest type II DNA topoisomerase (79), which cleaves dsDNAs ~30 times faster than the human type II DNA topoisomerase (80), and is, therefore, being used as a model enzyme for the study of the topoisomerase II DNA cleavage process, iii) an RNA guanylyltransferase (81) that was the first enzyme of its type to have its crystal structure resolved (82, 83), iv) a small

Figure 1.4 – PBCV-1 Partial Gene Map

DNA Replication, Recombination, and Repair		DNA Restriction/Modification		Integration and Transposition	
A468R	Archaeo-eukaryotic primase	A251R	Adenine methyltransferase	A134L	GIY-YIG endonuclease
A565R	ATPase (Chromosome segregation)	A581R	Adenine methyltransferase	A287R	GIY-YIG endonuclease
A392R	ATPase (DNA packaging)	A517L	Cytosine methyltransferase	A315L	GIY-YIG endonuclease
A561L	ATPase (DNA repair)	A530R	Cytosine methyltransferase	A351L	GIY-YIG endonuclease
A554/56/65R	ATPase (PP-loop)	A683L	Cytosine methyltransferase	A495R	GIY-YIG endonuclease
A544R	ATP-dependent DNA ligase	A252R	DNA restriction endonuclease	A539R	GIY-YIG endonuclease
A437L	DNA binding protein	A579L	DNA restriction endonuclease	A651L	GIY-YIG endonuclease
A583L	DNA Topoisomerase II	A582L	DNA restriction endonuclease	A87R	HNH endonuclease
A166R	Exonuclease	A593L	DNA restriction endonuclease	A267L	HNH endonuclease
A456L	Helicase-Superfamily III	A622L	DNA restriction endonuclease	A354R	HNH endonuclease
A193L	PCNA	A614L	DNA restriction endonuclease	A422R	HNH endonuclease
A574L	PCNA	A609L	DNA restriction endonuclease	A478L	HNH endonuclease
A50L	Pyrimidine dimer-specific glycosylase	A623L	DNA restriction endonuclease	A490L	HNH endonuclease
A417L	Replication factor C	A626L	DNA restriction endonuclease	A366L	Transposase
A399R	RNase H	A629R	DNA restriction endonuclease	A625R	Transposase
A185R	δ DNA polymerase	A476R	DNA restriction endonuclease		
Sugar and Lipid Manipulation		A427L	DNA restriction endonuclease	Transcription	
A426R	6-phosphofructokinase (?)	A674R	DNA restriction endonuclease	A612L	Histone H3, Lys 27 methylase
A254R	Acetyltransferase	A683L	DNA restriction endonuclease	A103R	mRNA guanylyltransferase
A53R	D-lactate dehydrogenase	A666L	DNA restriction endonuclease	A449R	RNA triphosphatase
A297L	Fructose-2,6 bisphosphatase	A654L	DNA restriction endonuclease	A464R	RNase III
A295L	Fucose synthase	A621L	DNA restriction endonuclease	A153R	Superfamily II helicase
A111/114R	Fucosyl-Glycosyltransferase	A162L	DNA restriction endonuclease	A241R	Superfamily II helicase
A118R	GDP-D-mannose dehydratase	A287R	DNA restriction endonuclease	A363R	Superfamily II helicase
A100R	Glucosamine synthetase	A315L	DNA restriction endonuclease	A189/192R	SWI/SNF chromatin remodeling complex
A526R	Glycerocephosphodiesterase (?)	A351L	DNA restriction endonuclease	A548L	SWI/SNF helicase
A49L	Glycerocephosphoryl diesterase	A495R	DNA restriction endonuclease	A107L	Transcription factor TFIIIB
A064R	Glycosyltransferase	A465R	DNA restriction endonuclease	A552R	Transcription factor TFIIID
A219/222/226R	Glycosyltransferase	A646L	Translation elongation factor-3	A125L	Transcription factor TFIIS
A473L	Glycosyltransferase	A666L	Translation elongation factor-3	A482R	VLTF2-type transcription factor
A546L	Glycosyltransferase	A105L	Ubiquitin C-terminal hydrolase		
A98R	Hyaluronan synthase	A604L	Zn metallopeptidase	Miscellaneous	
A271L	Lysophospholipase			A445L	ABC transporter protein
A654L	N-acetyltransferase	11 tRNAs		A638R	Agmatine iminohydrolase
A173L	Patatin-like phospholipase			A284L	Amidase
A609L	UDP-glucose 4-dehydrogenase			A245R	Cu/Zn-superoxide dismutase
Cell Wall Degradation				A180R	Fibronectin binding protein
A181/182R	Chitinase			A598L	Histidine decarboxylase
A260R	Chitinase			A237R	Homoserpinidase synthase
A292L	Chitosanase			A217L	Monoamine oxidase
A215L	β & α 1,4 linked glucuronic lyase			A78R	N-carbamoylput. amidohydrolase
A94L	β-1,3-glucanase			A61R	O-methyltransferase
Signalining				A207R	Ornithine/Arginine decarboxylase
A305L	Dual specificity phosphatase			A246R	Pathogenesis-related protein
A162L	Glutamate receptor				
A163R	Ligand-gated channel protein				
A250R	Potassium channel protein				
A34R	Serine/Threonine protein kinase				
A248R	Serine/Threonine protein kinase				
A277L	Serine/Threonine protein kinase				
A278L	Serine/Threonine protein kinase				
A282L	Serine/Threonine protein kinase				
A289L	Serine/Threonine protein kinase				
A614L	Serine/Threonine protein kinase				
A617R	Serine/Threonine protein kinase				

prolyl-4-hydroxylase that converts Pro-containing peptides into hydroxyl-Pro-containing peptides in a sequence-specific fashion (84), and v) the smallest protein (94 amino acids) to form a functional K⁺ ion channel (85). These minimalist enzymes may represent precursors of contemporary proteins.

1.7 DIVERSITY OF CHLOROVIRUS GENOMES

Not all PBCV-1 genes are required for viral replication in the laboratory. For example, four spontaneously-derived PBCV-1 mutants were isolated that contain 27- to 37-kb deletions involving the left end of the 330-kb genome (86). Two of these mutants have deletions beginning at nucleotide coordinates 4.9 or 16 kb and ending at 42 kb. In total, these two deleted regions, which probably resulted from aberrant recombination, encode 28 putative proteins. The other two mutants, which probably arose from aberrant non-homologous recombination, lack the entire left terminal 37-kb of the PBCV-1 genome, including the 2.2 kb terminal inverted-repeat region. Instead, the left terminus of these mutant viruses is replaced by an inverted 7.7- or 18.5-kb copy of the right end of the PBCV-1 genome. These deleted regions encode 26 single-copy open reading frames, 23 of which are common to those deleted in the two mutant viruses described above. Taken together, ~40 kb of single-copy DNA encoding 31 ORFs at the left end of the genome, or 12% of the PBCV-1 genome, is unnecessary for PBCV-1 replication in the laboratory. However, it should be noted that replication of these PBCV-1 deletion mutants is attenuated (i.e. their burst sizes are about half that of the wild type virus).

The occurrence of these deletion mutants also indicates that the size of the inverted terminal repeats of the PBCV-1 genome can vary. This conclusion is consistent

with the finding that the size and sequence of the inverted repeat region is not conserved among chlorella viruses (76, 87). The lack of conservation is somewhat surprising given that one might predict that the genome termini would be essential for viral DNA replication and/or packaging. These results also indicate that the viral DNA packaging machinery can tolerate significant differences in genome size (e.g., the largest deletion in PBCV-1 creates a genome of ~302 kb, whereas chlorella virus NY-2A has a genome of ~370 kb). Similar large deletions have been reported in the left terminus of the chlorella virus CVK1 genome (88).

Large insertions have also been observed in the genomes of chlorella viruses. Comparisons of the PBCV-1 and CVK2 genomes indicate an ~15 kb region in the left terminus of PBCV-1 that is absent in the CVK2 genome. Additionally, the CVK2 genome contains a 22.2 kb insertion in this region that contains 5 copies of a homolog of the PBCV-1 glycoprotein Vp260; this inserted sequence is absent in the PBCV-1 genome (89, 90).

The sizes and locations of the deletions and transpositions in the chlorella viruses resemble those observed in poxviruses (91) and ASFV (92) deletion mutants. Like PBCV-1, poxviruses and ASFV genomes have inverted terminal repeats and covalently-closed hairpin ends. Models explaining the generation of the deletions and deletion/transpositions in the poxvirus genomes (91, 93) may be relevant to analogous changes observed in the chlorella viruses.

In summary, considerable variation exists among the genomes of different chloroviruses, and the total number of genes in the chlorovirus gene pool exceeds that of any single isolate. The different sizes of the chlorovirus genomes and the large deletions

and insertions detected in the various mutant viruses suggest that dynamic and frequent rearrangements of viral genomes occur in natural environments. The fact that the left end of the chlorella viral genome is tolerant of deletions, insertions, and rearrangements suggests that a recombinational “hotspot” in this region may allow viruses to exchange genes among one another and possibly with their host(s).

1.8 PHYCODNAVIRUS GENES ARE LIKELY VERY OLD

As mentioned in the introduction, accumulating evidence suggests that at least some of the phycodnavirus genes, and by implication the viruses themselves, have a long evolutionary history possibly dating back to the time of prokaryotic and eukaryotic organism separation (ca. 2.5 – 3.0 billion years ago) (10, 11, 94). This evidence includes the following: i) Phylogenetic analysis of the δ-DNA polymerases of phycodnaviruses places its origin near the origin of all eukaryotic δ-DNA polymerases (11); ii) Phylogenetic analyses of several other PBCV-1 gene products (ornithine decarboxylase, potassium ion channel protein, and GDP-D- mannose-4, 6 dehydratase) place these viral gene products near the base of these genetic trees; iii) Many PBCV-1 encoded proteins, such as histone H3-lysine²⁷ dimethylase, ornithine decarboxylase, DNA topoisomerase type II, and the potassium ion channel protein, are among the smallest of their type and may represent the ancestors of their larger relatives. Traditionally, it is believed that viruses acquire genes from their hosts, and that the gene products would become less complex with time. However, the reverse seems possible here (i.e. the precursor proteins were simpler, and during evolution the proteins added domains to develop specialized roles, such as gaining regulatory capabilities, in more complex organisms); iv) Some

PBCV-1-encoded enzymes are more flexible than their homologs in higher eukaryotic organisms. For example, some viral enzymes carry out multiple functions while more 'advanced' organisms require separate enzymes to accomplish the same tasks. One interpretation of this finding is that these viral proteins may be the progenitor enzymes, and, therefore, they are more precocious than their highly evolved eukaryotic homologs in which two separate enzymes are required. This dual functionality of some PBCV-1 enzymes is not the result of gene fusion. Examples of such dual functioning enzymes include the following: a) ornithine decarboxylase, which decarboxylates arginine more efficiently than ornithine (95); b) dCMP deaminase, which deaminates dCMP as well as dCTP and dCDP (Zhang et al., in preparation). The catalytic site of the PBCV-1 dCMP deaminase appears to be more flexible than that of other organismal homologs. [Note: the Km of the viral enzyme is similar for dCMP and dCTP but it is higher than the Km of dCMP deaminases from other organisms.]; c) GDP-D-mannose 4,6 dehydratase, which catalyzes not only the formation of GDP-4-keto-6-deoxy-D-mannose, an intermediate in the synthesis of GDP-L-fucose, but can also reduce the same intermediate to GDP-D-rihamnose (96). We expect that other phycodnavirus enzymes have dual functions, a finding consist with their presumed ancient origin; v) The genomes of phycodnaviruses are mosaics of both prokaryotic (e.g. the genes which encode a DNA methyltransferases and restriction endonucleases) and eukaryotic genes (e.g. the genes which encode a mRNA capping enzymes). The viruses could have acquired these genes over time from their hosts, or, perhaps these genes already existed at the time that prokaryotic and eukaryotic organisms separated. If the latter possibility is true, it implies that the progenitor organism(s) lost certain genes as it evolved into the viruses; vi) Finally, one of

the earliest eukaryotic cells could have resembled a single-celled alga (97), and viruses which infected this progenitor may have co-evolved with its host.

1.9 CONCLUSION

Sequence analyses of phycodnaviruses suggest that its family members possibly display more sequence diversity than members of any other virus family. World-wide, there are 10^3 to 10^5 phycodnaviruses per ml of water, a prediction based on an estimate of 10^7 virus-like particles per ml of seawater and that 0.01-1% of these particles are phycodnaviruses (98, 99). Given the current estimates of fresh and oceanic waters having a volume of 10^{24} mL, and assuming that (i) the recombination and mutation frequencies of the viruses are similar to their eukaryotic host ($\sim 10^{-6}$ nucleotide changes per replication cycle), (ii) a genome size of 10^5 kb, (iii) infection incidences of 10%, (iv) a 10% sequence change resulting in the creation of a new virus, and (v) one infection cycle per day, we can estimate that 10^{19} new phycodnaviruses could be generated each day. Many of these viruses would be defective. Nevertheless, this estimation suggests that phycodnaviruses may be capable of exploring sequence space more extensively than any other eukaryotic virus. This may be primarily due to the vast numbers of available hosts in fresh and marine/coastal waters.

Despite the large genetic diversity detected among the sequenced phycodnaviruses, phylogenetic analyses of the δ -DNA polymerases (2, 11, 29) and the superfamily of archeao-eukaryotic primases (13) indicate that the phycodnaviruses group into a monophyletic clade within the NCLDVs. A recent study using eight concatenated core NCLDV genes also indicates that the phycodnaviruses cluster together and are

members of the NCLDV “superfamily” (100). However, it is also obvious that the identification of phycodnaviruses has just begun. Metagenomic studies, such as DNA sequencing of samples taken from the Sargasso Sea (101), indicate that many of the gene products resemble those of the phycodnaviruses. More specifically, 100 of the 213 PBCV-1 major capsid protein BLAST hits are to the unknown environmental sequences from the Sargasso Sea project. Many of these “unknowns” are more similar to the PBCV-1 major capsid protein than the major capsid proteins of other known phycodnaviruses. These results are interesting considering that PBCV-1 infects a fresh water alga, whereas the Sargasso Sea, which is located in the mid-Atlantic Ocean, is saltwater.

Phylogenetic studies of individual phycodnavirus genes also suggest that the viruses, or at least some of their genes, are evolutionary very old. Thus, studies on the phycodnaviruses, as well as other NCLDV members, may reveal information regarding the evolution of genes and genomes. Some evolutionary biologists have suggested that large dsDNA viruses, like the phycodnaviruses, may be the origin of the nucleus in eukaryotic cells (11, 102, 103), while, other biologists have suggested that the NCLDVs may be members of a fourth domain of life (10). Even if these intriguing hypotheses are shown to be incorrect, the study of the phycodnaviruses will continue to reveal many unexpected and exciting phenomena.

The annotations and genomic studies of two newly sequenced chloroviruses, NY-2A and MT325, are the main focal points of this dissertation. Following an extensive analysis of these two new chlorovirus genomes, NY-2A and MT325, a comparative study between the *Chloroviruses*, the *Phycodnaviridae*, and the NCLDVs is explored.

1.10 REFERENCES

- (1) Wilson, W. H., Van Etten, J. L., Schroeder, D. S., Nagasaki, K., Brussaard, C., Delaroque, N., Bratbak, G., and Suttle, C. (2005) Phycodnaviridae, in *Virus Taxonomy: Classification and Nomenclature of Viruses* (Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U., and Ball, L. A., Eds.) pp 163-175, Elsevier Academic Press, San Diego.
- (2) Chen, F., and Suttle, C. A. (1996) Evolutionary relationships among large double-stranded DNA viruses that infect microalgae and other organisms as inferred from DNA polymerase genes. *Virology* 219, 170.
- (3) Westbroek, P., Van Hinte, J. E., Brummer, G. J., Veldhuis, M., Brownlee, C., Green, J. C., Harris, R., and Heimdal, B. R. (1994) *Emiliania huxleyi* as a key to bioshpere-geosphere interactions, in *The Haptophyte Algae* (Green, J. C., Ed.) pp 321-334, Clarendon Press, Oxford.
- (4) Westbroek, P., Brown, C. W., Van Bleijswijk, J., Brownlee, C., Brummer, G. J., Conte, M., Egge, J., Fernadex, E., Jordan, R., Knappertsbusch, M., Stefels, J., Veldhuis, M., Van der Val, P., and Young, J. (1993) A model system approach to biological climate forcing--the example of *Emiliania huxleyi*, in *Global Planetary Change* pp 27-46.
- (5) Charlson, R. J., Lovelock, J. E., Andreae, M. O., and Warren, S. G. (1987) Oceanic phytoplankton, atmospheric sulfur, cloud albedo and climate. *Nature (London, United Kingdom)* 326, 655-61.
- (6) Malin, G., Liss, P. S., and Turner, S. M. (1994) Dimethyl sulfide production and atmospheric consequences, in *The Haptophyte Algae* (Green, J. C., Ed.) pp 303-320, Clarendon Press, Oxford.
- (7) Hambly, E., and Suttle, C. A. (2005) The viriosphere, diversity, and genetic exchange within phage communities. *Curr. Opin. Microbiol.* 8, 444-450.
- (8) Wommack, K. E., and Colwell, R. R. (2000) Viriplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* 64, 69.
- (9) Guiry, M. D., and Rindi, F. (2005), World-wide electronic publication, National University of Ireland, Galway.

- (10) Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Claverie, J. M. (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* 306, 1344-1350.
- (11) Villarreal, L. P., and DeFilippis, V. R. (2000) A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. *J. Virol.* 74, 7079.
- (12) Iyer, L. M., Balaji, S., Koonin, E. V., and Aravind, L. (2006) Evolutionary genomics of Nucleo-Cytoplasmic Large DNA Viruses. *Virus Research* in press.
- (13) Iyer, L. M., Aravind, L., and Koonin, E. V. (2001) Common origin of four diverse families of large eukaryotic DNA viruses. *J. Virol.* 75, 11720.
- (14) Van Etten, J. L., Lane, L. C., and Meints, R. H. (1991) Viruses and viruslike particles of eukaryotic algae. *Microbiological Reviews* 55, 586-620.
- (15) Kawakami, H., and Kawakami, N. (1978) Behavior of a virus in a symbiotic system, *Paramecium bursaria-zoochlorella*. *J. Protozool.* 25, 217.
- (16) Meints, R. H., Van Etten, J. L., Kuczmarski, D., Lee, K., and Ang, B. (1981) Viral infection of the symbiotic chlorella-like alga present in *Hydra viridis*. *Virology* 113, 698.
- (17) Van Etten, J. L., Meints, R. H., Burbank, D. E., Kuczmarski, D., Cuppels, D. A., and Lane, L. C. (1981) Isolation and characterization of a virus from the intracellular green alga symbiotic with *Hydra viridis*. *Virology* 113, 704-11.
- (18) Van Etten, J. L., Meints, R. H., Kuczmarski, D., Burbank, D. E., and Lee, K. (1982) Viruses of symbiotic chlorella-like algae isolated from *Paramecium bursaria* and *Hydra viridis*. *Proc. Natl. Acad. Sci. USA* 79, 3867.
- (19) Tessman, I. (1985) Genetic recombination of the DNA plant virus PBCV1 in a Chlorella-like alga. *Virology* 145, 319-22.
- (20) Van Etten, J. L., Burbank, D. E., Kuczmarski, D., and Meints, R. H. (1983) Virus infection of culturable chlorella-like algae and development of a plaque assay. *Science* 219, 994.

- (21) Wilson, W. H., Schroeder, D. C., Allen, M. J., Holden, M. T., Parkhill, J., Barrell, B. G., Churcher, C., Hamlin, N., Mungall, K., Norbertczak, H., Quail, M. A., Price, C., Rabbinowitsch, E., Walker, D., Craigon, M., Roy, D., and Ghazal, P. (2005) Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. *Science* 309, 1090-2.
- (22) Muller, D. G., Kawai, H., Stache, B., and Lanka, S. (1990) A virus infection in the marine brown alga *Ectocarpus siliculosus* (Phaeophyceae). *Bot. Acta* 103, 72.
- (23) Henry, E. C., and Meints, R. H. (1992) A persistent virus infection in *Feldmannia* (Phaeophyceae). *J. Phycol.* 28, 517.
- (24) Cottrell, M. T., and Suttle, C. A. (1991) Wide spread occurrence and clonal variation in viruses which cause lysis of a cosmopolitan, eukaryotic marine phytoplankton, *Micromonas pusilla*. *Mar. Ecol. Prog. Ser.* 78, 1.
- (25) Mayer, J. A., and Taylor, F. J. R. (1979) A virus which lyses the marine nanoflagellate *Micromonas pusilla*. *Nature* 281, 299.
- (26) Brussaard, C. P., Short, S. M., Frederickson, C. M., and Suttle, C. A. (2004) Isolation and phylogenetic analysis of novel viruses infecting the phytoplankton *Phaeocystis globosa* (Prymnesiophyceae). *Appl Environ Microbiol* 70, 3700-5.
- (27) Nagasaki, K., Ando, M., Itakura, S., Imai, I., and Ishida, Y. (1994) Virus-like particles in *Heterosigma akashiwo* Raphidophyceae: a possible red tide disintegration mechanism. *Mar. Biol.* 119, 307.
- (28) Nagasaki, K., Ando, M., Itakura, S., Imai, I., and Ishida, Y. (1994) Viral mortality in the final stage of *Heterosigma akashiwo* Raphidophyceae red tide. *J. Plankton Res.* 16, 1595.
- (29) Wilson, W. H., Van Etten, J. L., Schroeder, D. S., Nagasaki, K., Brussaard, C., Delaroche, N., Bratbak, G., and Suttle, C. (2005) *Phycodnaviridae*, Elsevier Academic Press, San Diego.
- (30) Brussaard, C. P. D. (2004) Viral control of phytoplankton populations - a review. *J. Eukaryot. Microbiol.* 51, 125-138.

- (31) Van Etten, J. L., Burbank, D. E., Joshi, J., and Meints, R. H. (1984) DNA synthesis in a Chlorella-like alga following infection with the virus PBCV-1. *Virology 134*, 443-9.
- (32) Bubeck, J. A., and Pfitzner, A. J. (2005) Isolation and characterization of a new type of chlorovirus that infects an endosymbiotic Chlorella strain of the heliozoon *Acanthocystis turfacea*. *J Gen Virol 86*, 2871-7.
- (33) Shihra, I., and Krauss, R. (1965) *Chlorella Physiology and Taxonomy of Forty-One Isolates.*, Maryland Press, College Park, MD.
- (34) Oh-Hama, T., and Miyachi, S. (1988) *Micro-algal Biotechnology*, Cambridge University Press, Cambridge.
- (35) Hellmann, V., and Kessler, E. (1974) Physiologische und Biochemische Beitrage zur Taxonomie der gattung Chlorella. VIII. Die Basenzusammensetzung der DNS. *Arch. Microbiol. 95*, 311.
- (36) Van Etten, J. L., Schuster, A. M., Girton, L., Burbank, D. E., Swinton, D., and Hattman, S. (1985) DNA methylation of viruses infecting a eukaryotic Chlorella-like green alga. *Nucleic Acids Research 13*, 3471-8.
- (37) Loos, E., and Meindl, D. (1982) Composition of the cell wall of Chlorella fusca. *Planta 156*, 270.
- (38) Takeda, H. (1988) Classification of chlorella strains by cell wall sugar composition. *Phytochemistry 27*, 3822.
- (39) Yamada, T., and Sakaguchi, K. (1982) Comparative studies on chlorella cell walls: induction of protoplast formation. *Arch. Microbiol. 132*, 1013.
- (40) Royce, C. L., and Pardy, R. L. (1996) Endotoxin-like properties of an extract from a symbiotic, eukaryotic chlorella-like green alga. *J. Endotoxin Res. 3*, 437.
- (41) Trench, R. K. (1979) The cell biology of plant-animal symbiosis. *Annu. Rev. Plant Physiol. 30*, 485.

- (42) Reisser, W., Burbank, D. E., Meints, R. H., Becker, B., and Van Etten, J. L. (1991) Viruses distinguish symbiotic Chlorella spp. of Paramecium bursaria. *Endocytobiosis Cell Res.* 7, 245.
- (43) Reisser, W. (1992) pp 746, Biopress, Bristol, UK.
- (44) Schuster, A. M., Burbank, D. E., Meister, B., Skrdla, M. P., Meints, R. H., Hattman, S., Swinton, D., and Van Etten, J. L. (1986) Characterization of viruses infecting a eukaryotic Chlorella-like green alga. *Virology* 150, 170-7.
- (45) Van Etten, J. L., Burbank, D. E., Schuster, A. M., and Meints, R. H. (1985) Lytic viruses infecting a Chlorella-like alga. *Virology* 140, 135-43.
- (46) Van Etten, J. L., Van Etten, C. H., Johnson, J. K., and Burbank, D. E. (1985) A survey for viruses from freshwater that infect a eukaryotic chlorella-like green alga. *Appl. Environ. Microbiol.* 49, 1326.
- (47) Zhang, Y., Burbank, D. E., and Van Etten, J. L. (1988) Chlorella viruses isolated in China. *Appl. Environ. Microbiol.* 54, 2170.
- (48) Yamada, T., Higashiyama, T., and Fukuda, T. (1991) Screening of natural waters for viruses which infect chlorella cells. *Appl. Environ. Microbiol.* 57, 3433.
- (49) Reisser, W., Becker, B., and Klein, T. (1986) Studies on ultrastructure and host range of a chlorella attacking virus. *Protoplasma* 135, 162.
- (50) Reisser, W., Burbank, D. E., Meints, S. M., Meints, R. H., Becker, B., and Van Etten, J. L. (1988) A comparison of viruses infecting two different chlorella-like green algae. *Virology* 167, 143.
- (51) Suttle, C. A., and Chan, A. M. (1994) Dynamics and distribution of cyanophages and their effect on marine Synechococcus spp. *Appl. Environ. Microbiol.* 60, 3167.
- (52) Suttle, C. A., and Chen, F. (1992) Mechanisms and rates of decay of marine viruses in seawater. *Appl. Environ. Microbiol.* 58, 3721.

- (53) Furuta, M., Schrader, J. O., Schrader, H. S., Kokjohn, T. A., Nyaga, S., McCullough, A. K., Lloyd, R. S., Burbank, D. E., Landstein, D., Lane, L., and Van Etten, J. L. (1997) Chlorella virus PBCV-1 encodes a homolog of the bacteriophage T4 UV damage repair gene denV. *Applied and Environmental Microbiology* 63, 1551-1556.
- (54) Ha, H. C., Sirisoma, N. S., Kuppusamy, P., Zweier, J. L., Woster, P. M., and Casero, R. A. (1998) The natural polyamine spermine functions directly as a free radical scavenger. *Proc. Natl. Acad. Sci. USA* 95, 11140.
- (55) McCullough, A. K., Romberg, M. T., Nyaga, S., Wei, Y., Wood, T. G., Taylor, J.-S., Van Etten, J. L., Dodson, M. L., and Lloyd, R. S. (1998) Characterization of a novel cis-syn and trans-syn-II pyrimidine dimer glycosylase/AP lyase from a eukaryotic algal virus, *Paramecium bursaria chlorella virus-1*. *Journal of Biological Chemistry* 273, 13136-13142.
- (56) Sun, L., Li, Y., McCullough, A. K., Wood, T. G., and Lloyd, R. S. (2000) Intron conservation in a UV-specific DNA repair gene encoded by chlorella viruses. *J. Mol. Evol.* 50, 82.
- (57) Li, Y., Lu, Z., Burbank, D. E., Kutish, G. F., Rock, D. L., and Van Etten, J. L. (1995) Analysis of 43 kb of the Chlorella virus PBCV-1 330-kb genome: map positions 45 to 88. *Virology* 212, 134-50.
- (58) Yamada, T., Tamura, K., Aimi, T., and Songsri, P. (1994) Self-splicing group I introns in eukaryotic viruses. *Nucleic Acids Res.* 22, 2532.
- (59) Nishida, K., Suzuki, S., Kimura, Y., Nomura, N., Fujie, M., and Yamada, T. (1998) Group I introns found in chlorella viruses: biological implications. *Virology* 242, 319.
- (60) Van Etten, J. L., Lane, L. C., and Meints, R. H. (1991) Viruses and viruslike particles of eukaryotic algae. *Microbiol. Rev.* 55, 586.
- (61) Nishida, K., Kimura, Y., Kawasaki, T., Fujie, M., and Yamada, T. (1999) Genetic variation of Chlorella viruses: variable regions localized on the CVK2 genomic DNA. *Virology* 255, 376-384.
- (62) Yamada, T., Hiramatsu, S., Songsri, P., and Fujie, M. (1997) Alternative

- expression of a chitosanase gene produces two different proteins in cells infected with chlorella virus CVK2. *Virology* 230, 361.
- (63) Que, Q., and Van Etten, J. L. (1995) Characterization of a protein kinase gene from two Chlorella viruses. *Virus Research* 35, 291-305.
- (64) Meints, R. H., Lee, K., Burbank, D. E., and Van Etten, J. L. (1984) Infection of a Chlorella-like alga with the virus, PBCV-1: ultrastructural studies. *Virology* 138, 341-346.
- (65) Van Etten, J. L., Lane, L. C., and Meints, R. H. (1991) Viruses and viruslike particles of eukaryotic algae. *Microbiol. Rev.* 55, 586-620.
- (66) Mehmel, M., Rothermel, M., Meckel, T., Van Etten, J. L., Moroni, A., and Thiel, G. (2003) Possible function for virus encoded K⁺ channel Kcv in the replication of chlorella virus PBCV-1. *FEBS Lett.*
- (67) Frohns, F., Kasmann, A., Kramer, D., Schafer, B., Mehmel, M., Kang, M., Van Etten, J. L., Gazzarrini, S., Moroni, A., and Thiel, G. (2006) Potassium Ion Channels of Chlorella Viruses Cause Rapid Depolarization of Host Cells during Infection. *J Virol* 80, 2437-44.
- (68) Schuster, A. M., Girton, L., Burbank, D. E., and Van Etten, J. L. (1986) Infection of a Chlorella-like alga with the virus PBCV-1: transcriptional studies. *Virology* 148, 181-189.
- (69) Manzur, K. L., Farooq, A., Zeng, L., Plotnikova, O., and Koch, A. W. (2003) A dimeric viral SET domain methyltransferase specific to Lys27 of histone H3. *Nat. Struct. Biol.* 10, 187.
- (70) Van Etten, J. L., Burbank, D. E., Joshi, J., and Meints, R. H. (1984) DNA synthesis in a Chlorella-like alga following infection with the virus PBCV-1. *Virology* 134, 443-449.
- (71) Meints, R. H., Lee, K., and Van Etten, J. L. (1986) Assembly site of the virus PBCV-1 in a Chlorella-like green alga: ultrastructural studies. *Virology* 154, 240-245.

- (72) Van Etten, J. L., Burbank, D. E., and Meints, R. H. (1986) Replication of the algal virus PBCV-1 in UV-irradiated Chlorella. *Intervirology* 26, 115-120.
- (73) Van Etten, J. L., Burbank, D. E., Xia, Y., and Meints, R. H. (1983) Growth cycle of a virus, PBCV-1, that infects Chlorella-like algae. *Virology* 126, 117-125.
- (74) Girton, L. E., and Van Etten, J. L. (1987) Restriction site map of the Chlorella virus PBCV-1 genome. *Plant Mol. Biol.* 9, 247-257.
- (75) Zhang, Y., Strasser, P., Grabherr, R., and Van Etten, J. L. (1994) Hairpin loop structure at the termini of the chlorella virus PBCV-1 genome. *Virology* 202, 1079-1082.
- (76) Strasser, P., Zhang, Y. P., Rohozinski, J., and Van Etten, J. L. (1991) The termini of the Chlorella virus PBCV-1 genome are identical 2.2-kbp inverted repeats. *Virology* 180, 763-769.
- (77) Ho, C. K., Van Etten, J. L., and Shuman, S. (1997) Characterization of an ATP-dependent DNA ligase encoded by Chlorella virus PBCV-1. *J. Virol.* 71, 1931-1937.
- (78) Sriskanda, V., and Shuman, S. (2002) Role of nucleotidyltransferase motifs I, III and IV in the catalysis of phosphodiester bond formation by chlorella virus DNA ligase. *Nucleic Acids Res.* 30, 903.
- (79) Lavrukhin, O. V., Fortune, J. M., Wood, T. G., Burbank, D. E., and Van Etten, J. L. (2000) Topoisomerase II from chlorella virus PBCV-1. *J. Biol. Chem.* 275, 6915.
- (80) Fortune, J. M., Lavrukhin, O. V., Gurnon, J. R., Van Etten, J. L., Lloyd, R. S., and Osheroff, N. (2001) Topoisomerase II from chlorella virus PBCV-1 has an exceptionally high DNA cleavage activity. *J. Biol. Chem.* 276, 24401.
- (81) Ho, C. K., Van Etten, J. L., and Shuman, S. (1996) Expression and characterization of an RNA capping enzyme encoded by chlorella virus PBCV-1. *J. Virol.* 70, 6658.
- (82) Hakansson, K., Doherty, A. J., Shuman, S., and Wigley, D. B. (1997) X-ray

- crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes. *Cell* 89, 543.
- (83) Hakansson, K., and Wigley, D. B. (1998) Structure of a complex between a cap analogue and mRNA guanylyl transferase demonstrates the structural chemistry of RNA capping. *Proc. Natl. Acad. Sci. USA* 95, 1505.
- (84) Eriksson, M., Myllyharju, J., Tu, H., Hellman, M., and Kivirikko, K. I. (1999) Evidence for 4-hydroxyproline in viral proteins: characterization of a viral prolyl 4-hydroxylase and its peptide substrates. *J. Biol. Chem.* 274, 22131.
- (85) Plugge, B., Gazzarrini, S., Nelson, M., Cerana, R., Van Etten, J. L., Derst, C., DiFrancesco, D., Moroni, A., and Thiel, G. (2000) A potassium channel protein encoded by chlorella virus PBCV-1. *Science* 287, 1641-1644.
- (86) Landstein, D., Burbank, D. E., Nietfeldt, J. W., and Van Etten, J. L. (1995) Large deletions in antigenic variants of the chlorella virus PBCV-1. *Virology* 214, 413.
- (87) Yamada, T., and Higashiyama, T. (1993) Characterization of the terminal inverted repeats and their neighboring tandem repeats in the Chlorella CVK1 virus genome. *Mol. Gen. Genet.* 241, 554-563.
- (88) Songsri, P., Hiramatsu, S., Fujie, M., and Yamada, T. (1997) Proteolytic processing of Chlorella virus CVK2 capsid proteins. *Virology* 227, 252-254.
- (89) Nishida, K., Kimura, Y., Kawasaki, T., Fujie, M., and Yamada, T. (1999) Genetic variation of Chlorella viruses: Variable regions localized on the CVK2 genomic DNA. *Virology* 255, 376-384.
- (90) Chuchird, N., Nishida, K., Kawasaki, T., Fujie, M., Usami, S., and Yamada, T. (2002) A variable region on the chlorovirus CVK2 genome contains five copies of the gene for Vp260, a viral-surface glycoprotein. *Virology* 295, 289.
- (91) Turner, P. C., and Moyer, R. W. (1990) The molecular pathogenesis of poxviruses. *Curr. Top. Microbiol. Immunol.* 163, 125.
- (92) Blasco, R., de la Vega, I., Almazan, F., Aguero, M., and Vinuela, E. (1989) Genetic variation of African swine fever virus: variable regions near the ends of

- the viral DNA. *Virology* 173, 251-257.
- (93) Shchelkunov, S. N., and Totmenin, A. V. (1995) Two types of deletions in orthopoxvirus genomes. *Virus Genes* 9, 231.
- (94) Gu, X. (1997) The age of the common ancestor of eukaryotes and prokaryotes: statistical inferences. *Mol. Biol. Evol.* 14, 861-866.
- (95) Shah, R., Coleman, C. S., Mir, K., Baldwin, J., Van Etten, J. L., Grishin, N. V., Pegg, A. E., Stanley, B. A., and Phillips, M. A. (2004) Paramecium bursaria chlorella virus-1 encodes an unusual arginine decarboxylase that is a close homolog of eukaryotic ornithine decarboxylases. *J Biol Chem* 279, 35760-7.
- (96) Tonetti, M., Zanardi, D., Gurnon, J. R., Fruscione, F., and Armirotti, A. (2003) Paramecium bursaria Chlorella virus encodes two enzymes involved in the biosynthesis of GDP-L-fructose and GDP-D-rhamnose. *J. Biol. Chem.* 278, 21559.
- (97) Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G., and Bhattacharya, D. (2004) A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* 21, 809-818.
- (98) Bergh, O., Borsheim, K. Y., Bratbak, G., and Heldal, M. (1989) High abundance of viruses found in aquatic environments. *Nature* 340, 467.
- (99) Proctor, L. M., and Fuhrman, J. A. (1990) Viral mortality of marine bacteria and cyanobacteria. *Nature* 343, 60-62.
- (100) Allen, M. J., Schroeder, D. C., Holden, M. T., and Wilson, W. H. (2005) Evolutionary history of the Coccothioviridae. *Mol. Biol. Evol. in press.*
- (101) Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y. H., and Smith, H. O. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66-74.
- (102) Pennisi, E. (2004) Evolutionary biology. The birth of the nucleus. *Science* 305,

766-8.

- (103) Bell, P. J. (2001) Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? *J. Mol. Evol.* 53, 251-256.

CHAPTER 2

GENOMIC ANALYSES OF *CHLORELLA* NC64A VIRUS NY-2A AND *CHLORELLA* Pbi VIRUS MT325

2.1 INTRODUCTION

Hundreds of plaque-forming viruses capable of infecting *Chlorella* NC64A have been isolated from fresh water collected throughout the world. Like PBCV-1, all of these viruses are composed of many structural proteins and a large (at least 300 kb) dsDNA genome with an ~40% G+C content. They are polyhedron-shaped, with diameters of 190 nm and can be distinguished from each other by their plaque size, antiserum reactivity, DNA restriction patterns, and the nature and abundance of methylated bases in their genome (1). Each viral genome contains 5-methylcytosine (5mC) at a level which varies from 0.1-45% of the total cytosines, and many contain N⁶-methyladenine (6mA) at levels from 1.45-37% of the total adenines. Because the level of methylation varies greatly from one virus to another, it provides a unique system for studying the effect of base methylation on protein/DNA interactions. While the sequence of the chlorella virus PBCV-1 increased our knowledge of large dsDNA algal viruses, it revealed how much is still unknown. Undoubtedly, the genomic sequences of additional chlorella viruses will lead to the identification of new genes whose functions will be both scientifically interesting and possibly commercially valuable. In addition, the information gained from these sequenced genomes will advance the study of the evolutionary history of not only algal viruses but also that of other families of large dsDNA viruses, such as the

nucleocytoplasmic large DNA viruses (NCLDV). Therefore, in addition to the genome of the chlorella prototype virus PBCV-1, two other chlorella viruses were selected to be sequenced.

The genome of NY-2A, a virus which infects *Chlorella* NC64A, was chosen to be sequenced for two reasons. First, it has the largest genome (~370 kb) of the 36 partially-characterized *Chlorella* NC64A viruses. Second, it has a genome which is heavily methylated relative to that of the prototype virus PBCV-1 (45% 5mC, 37% 6mA and 1.9% 5mC, 1.5% 6mA, respectively)(2). The second genome selected for sequencing was virus MT325, a *Chlorella* Pbi virus. The viruses that infect *Chlorella* Pbi, isolated from a European isolate of *P. bursaria*, were first detected in European fresh waters (3, 4). The Pbi viruses, like the NC64A viruses, have large polyhedron shapes (with diameters of 190 nm) and are composed of many structural proteins and large (at least 300 kb) dsDNA genomes containing methylated bases. They are serologically distinct from the NC64A viruses. Their genomes have a higher G+C content (46%) than the NC64A viruses (40%). Additionally, the Pbi viruses neither infect nor attach to *Chlorella* NC64A and vice versa (5).

In this chapter, the genomes of NY-2A and MT325, two Chlorella viruses which infect different hosts, will be explored.

2.2 EXPERIMENTAL METHODS

2.2.1 Viral DNA Isolation and Sequencing

The production and purification of PBCV-1 and the isolation of its genomic DNA was performed as previously described (1, 6, 7). The NY-2A host, *Chlorella* NC64A,

and the MT325 host, *Chlorella Pbi*, were grown on MBBM and FES medium, respectively. NY-2A and MT325 viruses were produced and purified and their viral DNAs were isolated using the methods and protocols developed for PBCV-1. Purified genomic DNA was sequenced and assembled at The Institute for Genomic Research (TIGR).

2.2.2 Genomic Sequence Analysis

2.2.2.1 Initial Analysis of Viral Genomes

Genomic dot plots were created to determine the orientation of the viral genomes relative to the PBCV-1 genome. The NY-2A and MT325 genomes were individually plotted against the PBCV-1 genome using blastn (DNA vs. DNA) and tblastx (translation vs. translation). Similarities between the two genomes with E-values $<10^{-5}$ were represented.

2.2.2.2 Identification of Putative tRNAs

Putative tRNA genes were identified using the tRNAscan-SE program developed by Lowe and Eddy at Washington University School of Medicine in 1997 (8).

2.2.2.3 Identification and Naming of Open Reading Frames

A potential protein-coding region, or Open Reading Frame (ORF), was defined as a continuous stretch of DNA that translates into a polypeptide initiated by an ATG translation start codon and extends for 65 or more codons using the standard genetic code. The ORF Finder program (http://bioinformatics.org/sms/orf_find.html) was used to identify all potential ORFs which met this criterion. The ORFs are numbered consecutively starting at the beginning of the genome (as determined by alignment with

the PBCV-1 genome). The letter R or L following the number indicates that the orientation of the putative ORF is either left-to-right or right-to-left, respectively.

In the PBCV-1 genome, the first *Chlorella* NC64A virus sequenced, putative ORFs were designated with either an “A” or an “a” depending on whether the ORF was classified as a major or minor ORF, respectively. The criteria for the classification of major/minor ORFs will be described below. To avoid confusion between different sequenced viral genomes, this ORF naming scheme was extended. For the NY-2A viral genome, the second *Chlorella* NC64A viral genome sequenced, putative ORFs were designated either “B” (for major ORFs) or “b” (for minor ORFs). When the third *Chlorella* NC64A viral genome is sequenced, its putative ORFs will be designated with a “C”, and this naming will continue for all subsequent *Chlorella* NC64A virus genomes. For the MT325 genome, the first *Chlorella* Pbi viral genome sequenced, each putative ORF was designated with an “M”. This letter was chosen to distinguish it from viruses which infect *Chlorella* NC64A (i.e. PBCV-1 and NY-2A) and because it is in the middle of the alphabet. When the genome of the next *Chlorella* Pbi virus is sequenced, its putative ORFs will be designated with the letter “N”.

2.2.2.4 Identification of Major versus Minor Open Reading Frames

Putative ORFs were designated to be either major or minor based on several criteria. It was previously reported that, for five transcribed PBCV-1 genes, the 50 bases preceding the start codon were at least 70% A+T (9). Since then, several functional PBCV-1 ORFs have been characterized. Table 2.1 lists 44 functional proteins encoded by the PBCV-1 genome and the percentages of A+T nucleotides of the 50 bases upstream of their start codons. While the A+T percentages range from 50-88%, only four proteins

Table 2.1 Functional Proteins Encoded by PBCV-1

ORF	Description	% A+T
A34R	Ser/Thr protein kinase	82
A50L	Pyrimidine dimer-specific glycosylase	82
A78R	N-carbamoylput. amidohydrolase	74
A85R	Prolyl 4-hydroxylase	66
A94L	β -1, 3-glucanase	76
A98R	Hyaluronan synthase	78
A100R	Glucosamine synthetase	72
A103R	mRNA guanyltransferase	84
A118R	GDP-D-mannose dehydratase	74
A166R	Exonuclease	80
A169R	Aspartate transcarbamylase	74
A181R	Chitinase	80
A207R	Ornithine decarboxylase	86
A215L	β & α 1,4 linked glucuronic lyase	82
A237R	Homospermidine synthase	76
A245R	Cu/Zn-superoxide dismutase	80
A248R	Ser/Thr protein kinase	70
A250R	Potassium channel protein	82
A251R	Adenine DNA methylase	82
A252R	DNA restriction endonuclease	74
A260R	Chitinase	78
A277L	Ser/Thr protein kinase	70
A278L	Ser/Thr protein kinase	80
A282L	Ser/Thr protein kinase	60
A284L	Amidase	82
A289L	Ser/Thr protein kinase	74
A292L	Chitosanase	80
A295L	Fucose synthetase	86
A438L	Glutaredoxin	86
A449R	RNA triphosphatase	70
A464R	RNase III	50
A517L	Cytosine DNA methylase	88
A530R	Cytosine DNA methylase	78
A544R	ATP-dependent DNA ligase	64
A551L	dUTP pyrophosphatase	70
A579L	DNA restriction endonuclease	82
A581R	Adenine DNA methylase	82
A583L	DNA topoisomerase II	82
A596R	dCMP deaminase	78
A609L	UDP-glucose dehydrogenase	82
A612L	Histone H3, Lys 27 methylase	80
A614L	Ser/Thr protein kinase	86
A638R	Agmatine iminohydrolase	80
A674R	Thymidylate synthase X	70

Table 2.1 Functional Proteins Encoded by PBCV-1. Forty-four functional proteins encoded by the PBCV-1 genome are listed along with their percentage of A+T nucleotides of the 50 bases upstream of their ATG start codon.

have a $< 70\%$ A+T composition of the 50 nucleotides upstream of their start codon.

Therefore, any ORF in which the 50 nucleotides upstream of its start codon has a $\geq 70\%$ A+T composition was initially classified as a major ORF.

Two characteristics changed an initially-designated major ORF to a minor ORF. First, if the major ORF resided within a larger major ORF, regardless of orientation, it was reclassified as a minor ORF. Second, if the major ORF significantly overlapped with a larger major ORF, regardless of direction, it was reclassified as a minor ORF. Two criteria were used to classify an ORF which has $< 70\%$ A+T composition of the 50 nucleotides upstream of the start codon as a major ORF. First, based on the analyses with public databases (Non-redundant database at NCBI, Protein Families, or Clusters of Orthologous Groups), any ORF with an identified homolog was classified as a major ORF regardless of the A+T composition of the 50 nucleotides upstream of its start codon. Second, if a minor ORF resided between, but did not overlap with, two major ORFs, it was reclassified as a major ORF.

2.2.2.5 Characterization of Open Reading Frames

DNA and protein sequences were initially analyzed using the Genetics Computer Group (GCG) Wisconsin Package, v. 10.3 (10). For each ORF, the following general characteristics were determined: the A+T content of the 50 bases upstream from the ATG start codon, the frame in which the ORF was encoded, the G+C content of the gene, and the number of amino acids, the molecular weight, the pI, and the extinction coefficient of the putative encoded protein.

2.2.2.6 Analysis with the Non-redundant Database at NCBI

Each ORF identified was used in a search for homologs using the protein-protein BLAST (blastp) program (11) against the non-redundant (NR) protein databases at NCBI. The criteria used to search the NR database were as follows: Expect = 10; Word Size = 3; Scoring matrix = blosum62; and E-value < 0.001. For simplicity, only the top three subjects for each query are shown in the addendum.

2.2.2.7 Analysis with the Protein Families Database

The Protein Families (Pfam) database is a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains and families (12). Each putative identified ORF was scanned for potential functional attributes using Pfam version 18.0 (12). An E-value of < 0.001 was used for this analysis.

2.2.2.8 Identification of Clusters of Orthologous Groups

Clusters of Orthologous Groups (COGs) of proteins were generated by comparing protein sequences encoded by complete genomes representing major phylogenetic lineages. Each COG consists of individual proteins or groups of paralogs from at least 3 lineages and, thus, corresponds to an ancient conserved domain (13, 14). Every identified ORF was scanned to determine if it belonged to a particular COG. The E-value cutoff for this analysis was < 0.001.

2.2.2.9 Identification of Inteins and Introns

The identification of intervening sequences (introns and inteins) within protein coding genes were initially inferred by the presence of either out-of-frame or in-frame sequence insertions compared to their homologs in the public databases. The

characterization was then refined by searching for the consensus signals and features known to be associated with such intervening sequences. The identification of inteins was performed using InBase (<http://www.neb.com/neb/inteins.html>), the Intein Database and Registry (15). The identification of introns was based on comparisons with the two types of introns found in PBCV-1. The first type of intron, which is present in the DNA polymerase gene from the Chlorella virus PBCV-1, contains a 5' AG/GUGAGU and 3' UGCAG/UU splice-site sequence (16). Another nuclear-located, spliceosomal processed pre-mRNA intron is located the pyrimidine dimmer-specific glycosylase (PDG) encoded by several chlorella viruses. The introns in the *pdg* gene have 5'AG/GTATGT and 3' TTGCAG/AA splice site sequences (17). The second type of intron, Group 1B introns, contain four conserved sequence elements: the P element (AAATTGCGGGAA), the Q element (AATCCGCAGC), the R element (GTTCAACGACTAAA), and the S element (AAGATATAAGTCT) (18).

2.3 RESULTS

2.3.1 Analysis and Annotation of the NY-2A Genome

The NY-2A genome is the largest *Chlorovirus* genome sequenced to date. It consists of 368,683 nucleotides. The genomes of the two alga viruses which infect the same host, *Chlorella* NC64A, each contain identical inverted repeats. After this conserved region, the sequence diverges. Alignment of the 5' terminus and the inverted 3' terminus of a genome is depicted in addendum 2.6.1. The termini of the PBCV-1 genome contain identical inverted repeats of 2,221 nucleotides (19). Furthermore, the termini of the NY-2A genome also contains identical inverted repeats, however it should be noted that there is only a 52% identity between the inverted repeats of PBCV-1 and NY-2A. The alignment of the NY-2A inverted repeats is slightly different than the termini of the PBCV-1 genome. This difference in the termini is probably due to how the libraries were constructed. For example, there are an additional 42 nucleotides on the 3' terminus before the beginning of the identical inverted repeats. Following this addition, there are 2,186 nucleotides from the 5' terminus which contain identical inverted repeats with the 3' terminus. However, at nucleotide 1,587 there is an insert of 21 nucleotides in the 3' inverted terminus. A high sequence identity following the 2,186 nucleotide inverted repeat of the NY-2A termini continues for approximately 100 nucleotides and then the sequence diverges. The average G+C content of the genome sequence is 40.7%, a value similar to the 40.0% G+C content of the PBCV-1 genome. To determine the orientation of the NY-2A genome relative to the genome of the prototype virus PBCV-1, we compared the NY-2A genome with that of PBCV-1 using a genomic dot plot. Based on these results (Figure 2.1), an orientation co-linear with the PBCV-1 genome was

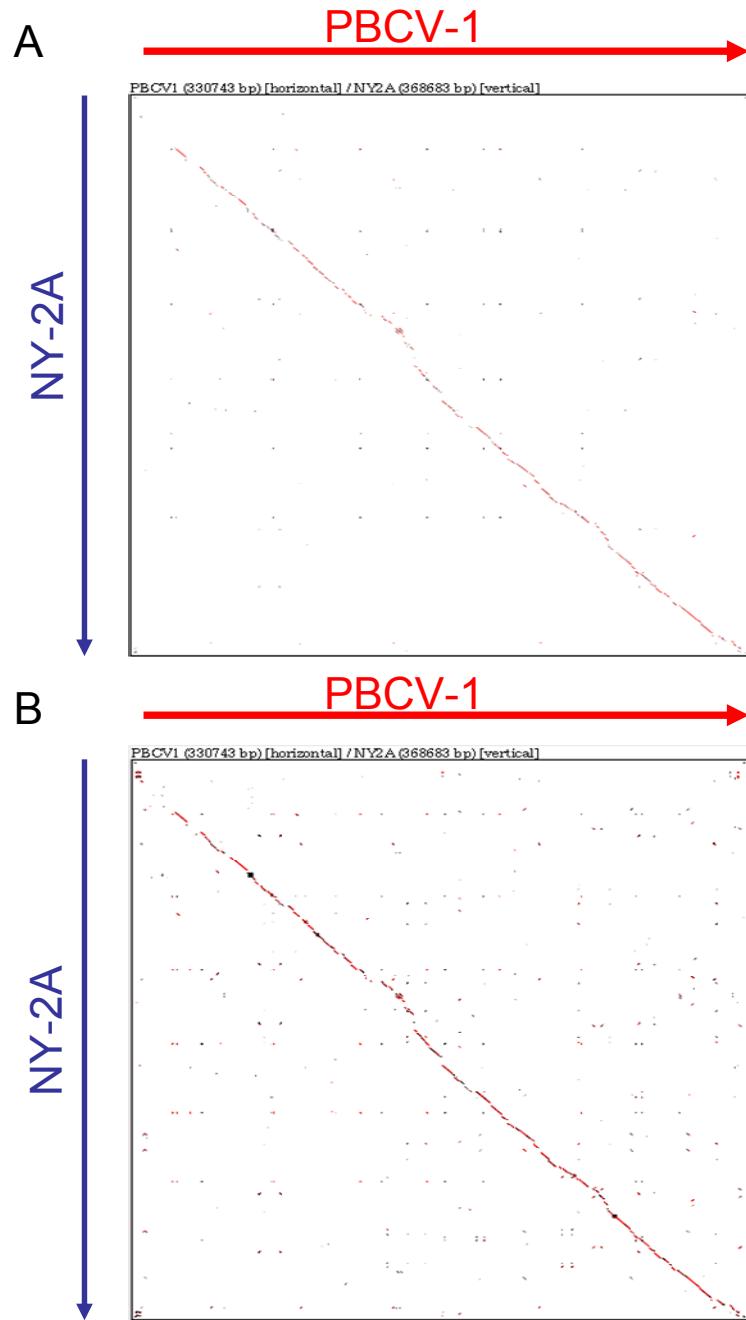


Figure 2.1 Genomic Dot Plots Comparing the Genomes of PBCV-1 to NY-2A (A) blastn (DNA vs. DNA) and (B) tblastx (translation vs. translation) of the PBCV-1 genome to the NY-2A genome. The dots represent similarities with E-value $<10^{-5}$. The red dots represent similarity > 200 bp and the black dots represent similarity ≤ 200 bp.

established. This alignment demonstrates a high degree of co-linearity between the genomes of these two alga viruses which infect the same host, *Chlorella* NC64A.

2.3.1.1 Identification of Putative tRNAs

The 368-kb genome of NY-2A was analyzed using the tRNAscan-SE program. Seven putative tRNAs, which were centrally clustered within the genome, were identified. One tRNA, tRNA^{Tyr}, contains a predicted 12 nucleotide intron located at nucleotide 195331 to 195343. Table 2.2 lists the 7 tRNAs encoded by the NY-2A genome and the frequency of codon usage in the tRNAs relative to that in all ORFs. The percentage with which a certain amino acid was coded for by a particular tRNA as compared to all tRNAs for that particular amino acid was examined. For example, 2 of the 6 leucine-coded tRNAs in the NY-2A genome account for 37% of all the leucines encoded by all of the ORFs. In general, each tRNA represents 26-42% of the total usage of the particular amino acid for which it codes.

Table 2.2 NY-2A tRNAs						
tRNA #	tRNA Begin	tRNA End	Anti-Codon	tRNA Type	% Codon Usage/ Total Codons	% Codon Usage/AA tRNA Type
1	194698	194778	TAA	Leu	1.0	14.5
2	194921	195005	CAA	Leu	1.6	22.4
3	195029	195100	TCT	Arg	1.6	34.3
4	195201	195272	GTT	Asn	2.5	41.8
5	195295	195379	GTA	Tyr	1.4	34.8
6	195382	195456	CTT	Lys	2.8	34.3
7	195482	195554	AAC	Val	1.7	25.7

2.3.1.2 Identification of Major and Minor Open Reading Frames

The criteria used for the identification and classification of ORFs are described in the experimental methods. Briefly, a putative protein-coding region, or ORF, was

defined as a continuous stretch of DNA that translates into a polypeptide initiated by an ATG translation start codon and extends for 65 or more codons. Using these criteria, 886 ORFs were identified in the 368 kb genome of NY-2A. The ORF names were based on three criteria. First, the ORF name begins with either a “B”, for a major ORF, or a “b”, for a minor ORF. Second, the ORFs were numbered consecutively in the order in which they appeared in the genome which was aligned with the PBCV-1 genome. Third, the letter R or L following the ORF number indicates that the putative transcript runs either left-to-right or right-to-left, respectively.

The 886 ORFs of the NY-2A genome were classified as either major or minor ORFs based on the following considerations: (i) Any ORF in which the A+T content of the 50 bases upstream of the start codon is $\geq 70\%$ was originally classified as a major ORF. Using this criterion, 418 ORFs were initially classified as major ORFs. Three criteria determined the reclassification of a major ORF, which has a $\geq 70\%$ A+T content of the 50 bases upstream of the start codon, to a minor ORF (Table 2.3 shown is the major ORF grouped together with the ORF that was changed from a major to a minor).
(ii) Any major ORF which resided within a larger major ORF of opposite polarity was reclassified as a minor ORF. As depicted in Table 2.3a, 42 ORFs were reclassified from major to minor ORFs using this criterion. Of these 42 ORFs, 26 resided entirely within the larger major ORF, and the remaining 16 resided almost entirely within the larger ORF (i.e. 2-52 nucleotides of their coding region extended outside the larger major ORF).
(iii) Any major ORF which resided within a larger major ORF of the same polarity was reclassified as a minor ORF. As depicted in Table 2.3b, 8 ORFs were reclassified from major to minor ORFs based on this criterion. As illustrated by the tables, the majority of

the major ORFs reclassified as a minor ORF based on the above criteria, were of the opposite polarity. (iv) Any major ORF which overlapped with a larger major ORF, regardless of its orientation was reclassified as a minor ORF. As depicted in Table 2.3c, 8 ORFs were reclassified from major to minor ORFs based on this criterion. Two criterions were examined to reclassify a minor ORF, which has a < 70% A+T content of the 50 nucleotides upstream of the start codon, to a major ORF (Table 2.4). (v) Based on the analysis of the ORFs using the non-redundant, Pfam or COG databases, 9 ORFs (depicted in Table 2.4a) which were initially classified as minor ORFs were reclassified as major ORFs because they were determined to encode a putative functional protein. (vi) Any minor ORF which resided between, but did not overlap with, two major ORFs was reclassified as a major ORF. As depicted in Table 2.4b, 35 ORFs were reclassified from minor to major ORFs based on this criterion.

Table 2.3 NY-2A ORFs Reclassified from a Major to a Minor

Table 2.3a Changes from a Major to a Minor ORF based on (ii)					
	ORF	Location	% A+T	AA	
1	B6L	2382 .. 2960	84%	192	22 B503L
	b7R	2490 .. 2786	70%	98	b502R
2	B16L	6237 .. 7343	84%	368	23 B522L
	b17R	6534 .. 6791	70%	85	b521R
3	B27L	11787 .. 12722	78%	311	24 B525L
	b26R	11772 .. 12023	76%	83	b528R
4	B80L	41010 .. 42947	76%	645	25 B533L
	b81R	41058 .. 41318	70%	86	b535R
5	B110L	51336 .. 51869	78%	177	26 B571L
	b111R	51462 .. 51689	76%	75	b572R
6	B110L	51336 .. 51869	78%	177	27 B598L
	b112R	51647 .. 51871	74%	74	b599R
7	B150L	66997 .. 67869	90%	290	28 B606L
	b153R	67662 .. 67859	72%	65	b605R
8	B159R	69422 .. 72025	82%	867	29 B606L
	b162L	71198 .. 71629	74%	143	b607R
9	B235L	104004 .. 104315	70%	103	30 B611L
	b234R	103983 .. 104264	84%	93	b610R
10	B246R	108041 .. 108649	76%	202	31 B619L
	b247L	108161 .. 108376	74%	71	b620R
11	B258R	112905 .. 116741	84%	1278	32 B638R
	b259L	113161 .. 113388	76%	75	b639L
12	B264L	117560 .. 118018	86%	152	33 B641R
	b265R	117803 .. 118021	70%	72	b642L
13	B298L	129994 .. 131112	72%	372	34 B667L
	b299R	130109 .. 130354	70%	81	b668R
14	B305R	131942 .. 133465	90%	507	35 B723L
	b309L	133079 .. 133471	76%	130	b722R
15	B331R	141226 .. 142155	74%	309	36 B772L
	b333L	141867 .. 142088	70%	73	b773R
16	B354L	149115 .. 149909	92%	264	37 B781L
	b355R	149131 .. 149385	80%	84	b782R
17	B365L	153977 .. 155794	84%	605	38 B792R
	b364R	153945 .. 154205	70%	86	b793L
18	B365L	153977 .. 155794	84%	605	39 B805R
	b366R	154349 .. 154567	70%	72	b808L
19	B378L	160393 .. 162330	78%	645	40 B816L
	b379R	160441 .. 160701	70%	86	b817R
20	B465R	196060 .. 197199	76%	379	41 B869R
	b467L	196840 .. 197178	76%	112	b870L
21	B469L	197983 .. 199194	72%	403	42 B878L
	b470R	198020 .. 198232	74%	70	b877R

Table 2.3a NY-2A ORFs changed from a major to a minor based on criterion (ii)

Forty-two ORFs were reclassified from a major to a minor ORF based on criterion (ii), which states that if a major ORF resides within a larger major ORF of the opposite polarity it will be reclassified as a minor ORF. The table lists the ORFs which met this criterion along with their location, % A+T 50 nucleotides upstream, and the number of amino acids encoded.

Table 2.3b Changes from a Major to a Minor ORF based on (iii)

	ORF	Location		% A+T	AA
1	B27L	11787	..	12722	78% 311
	b30L	12287	..	12532	74% 81
2	B226L	101915	..	102754	82% 279
	b229L	102510	..	102791	70% 93
3	B525L	219471	..	220100	82% 209
	b528L	219952	..	220152	84% 66
4	B587R	247296	..	247763	76% 155
	b586R	247244	..	247480	86% 78
5	B593R	248586	..	248927	76% 113
	b594R	248630	..	248884	74% 84
6	B756L	313106	..	315055	96% 649
	b758L	314017	..	314268	70% 83
7	B777L	323249	..	324310	78% 353
	b779L	323845	..	324315	78% 156
8	B805R	333334	..	334356	76% 340
	b807R	333596	..	333802	74% 68

Table 2.3c Changes from a Major to a Minor ORF based on (iv)

	ORF	Location		% A+T	AA
1	B113R	51897	..	52244	70% 115
	b114L	51933	..	52130	72% 65
2	B346L	146651	..	147400	80% 249
	b345R	146532	..	146747	76% 71
3	B371L	158499	..	159338	78% 279
	b373R	159166	..	159450	70% 94
4	B556R	233711	..	234835	88% 374
	b557R	234738	..	234977	78% 79
5	B559L	234845	..	235477	81% 210
	b558L	234766	..	235011	74% 81
6	B579R	243061	..	243315	82% 84
	b578L	242975	..	243175	74% 66
7	B850L	354158	..	355333	74% 391
	b851R	355248	..	355454	74% 68
8	B874L	363017	..	364117	84% 366
	b873R	362956	..	363207	70% 83

Table 2.3b NY-2A ORF changed from a major to a minor ORF based on criterion (iii). Eight ORFs were reclassified from a major to a minor ORF based on criterion (iii), which states that if a major ORF resides within a larger major ORF of the same polarity it will be reclassified as a minor ORF. **Table 2.3c NY-2A ORF changed from a major to a minor ORF based on criterion (iv).** Eight ORFs were reclassified from a major to a minor ORF based on criterion (iv), which states that if a major ORF overlapped with larger ORFs, regardless of direction, it was reclassified as a minor. These tables list the ORFs which met this criterion along with the location, % A+T 50 nucleotides upstream, and the number of amino acids the ORF encoded.

Table 2.4 NY-2A ORFs Reclassified from a Minor to a Major.

Table 2.4a Changes from a Minor to a Major ORF based on (v)			
	ORF	Description	Source
1	B253R	DNA polymerase	NR, Pfam
2	B271R	Cytosine deaminase	NR, Pfam, COG
3	B288L	β & α 1,4 linked glucuronic lyase	NR
4	B393L	Chitosanase	NR, Pfam
5	B430L	Dual specificity phosphatase	NR, Pfam, COG
6	B446R	HNH endonuclease	NR, Pfam
7	B543L	Adenine methyltransferase	NR
8	B711L	Transposase	NR, Pfam, COG
9	B774R	Adenine methyltransferase	NR, Pfam, COG

Table 2.4a NY-2A ORF changed from a minor to a major ORF based on criterion (v).
Nine ORFs were reclassified from a minor to a major based on criterion (v), which states that if an analysis of an ORF against the NR, Pfam, or COG databases revealed that the ORF had a putative function it would be reclassified as a major ORF.

Table 2.4b Changes from a Minor to a Major ORF based on (vi)			
	ORF		ORF
1	B001L	19	B438R
2	B003R	20	B441R
3	B004R	21	B457L
4	B075L	22	B461R
5	B087L	23	B494R
6	B099R	24	B496L
7	B108L	25	B560L
8	B109R	26	B566R
9	B130R	27	B590R
10	B154L	28	B662R
11	B256R	29	B734R
12	B272L	30	B741L
13	B324L	31	B811R
14	B342R	32	B813L
15	B361R	33	B883L
16	B408R	34	B884L
17	B411L	35	B886R
18	B431L		

Table 2.4b NY-2A ORF changed from a minor to a major ORF based on criterion (vi).
Thirty-five ORFs were reclassified from a minor to a major based on criterion (vi), which states that if a minor ORF resided between two major ORFs, without overlapping, it was reclassified as a major ORF.

2.3.1.3 Characterization of Open Reading Frames

Using the above-described criteria, 404 of the 886 open-reading frames (ORFs) were classified as major ORFs. For each ORF, the GCG software was used to identify several general characteristics and properties: the nucleotide composition of the ORF, the A+T content of the 50 nucleotides upstream of the ORF, the frame in which the putative protein was encoded, the number of amino acids in the encoded protein, and the putative encoded protein's molecular weight, pI, and extinction coefficient. These general characteristics are summarized in addendum 2.6.2.1. Following the classification and characterization of every ORF, relative genomic analyses of the ORFs were conducted. In Figure 2.2, the orientation of the ORFs, relative to either the frame in which they were encoded or to their overall direction within the NY-2A genome, are depicted. In general, the directions and the frames in which the ORFs were encoded were evenly represented, regardless of ORF classification. The size of the putative encoded protein (e.g. the number of encoded amino acids) for each ORF is shown in Figure 2.3. Over 70% of the minor ORFs encoded proteins containing less than 100 amino acids. According to the Release 48.4 of 08-Nov-05 from the UniProt/KB/Swiss-Prot database (us.expasy.org/sprot/relnotes/relstat.html), the average encoded protein size for the 197228 sequence entries is 362 amino acids. Only a small fraction of these entries encoded putative proteins of less than 100 amino acids, a finding which supports our classification of these ORFs as minor. The analysis of the pIs of the putative proteins encoded by the ORFs is depicted in Figure 2.4. Despite a basic trend in pIs observed for the major ORFs, a peak at pI 4.5 was also observed. Basic proteins are likely to be associated with the virion where they may function to neutralize the positively-charged

genomic DNA. The functions of the proteins corresponding to the peak at pI 4.5 are involved in several metabolic processes (i.e., exonuclease, SKP-1, PCNA, and arginine decarboxylase). Figure 2.5 illustrates the intergenic space between the major ORFs. The majority of the ORFs were separated by less than 100 nucleotides. Together, these figures illustrate several characteristics of the ORFs encoded by the NY-2A viral genome. In general, the major ORFs are evenly distributed along the genome with very little intergenic space.

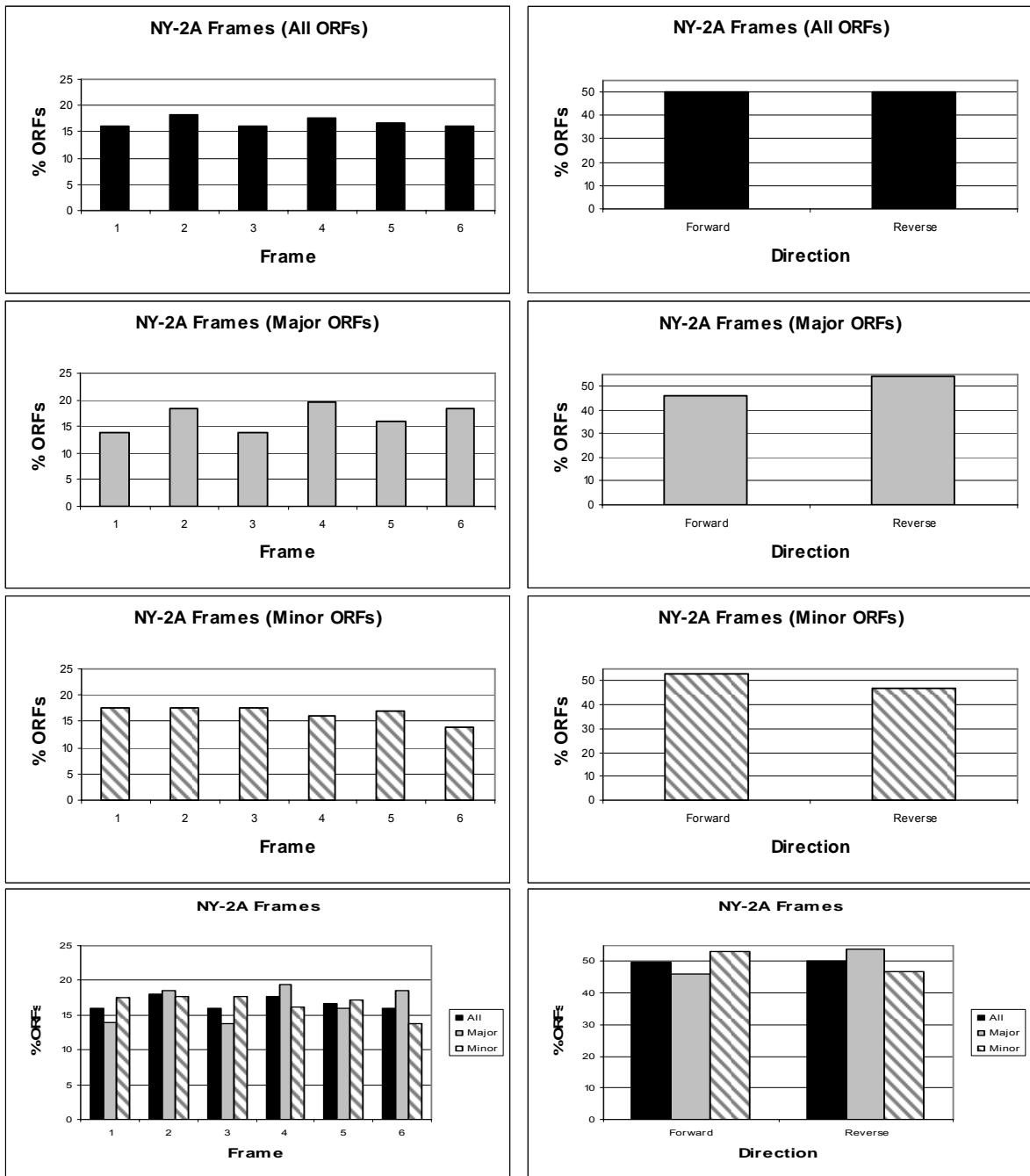


Figure 2.2 Orientation of NY-2A ORFs relative to either the frame in which they were encoded or to their overall direction.

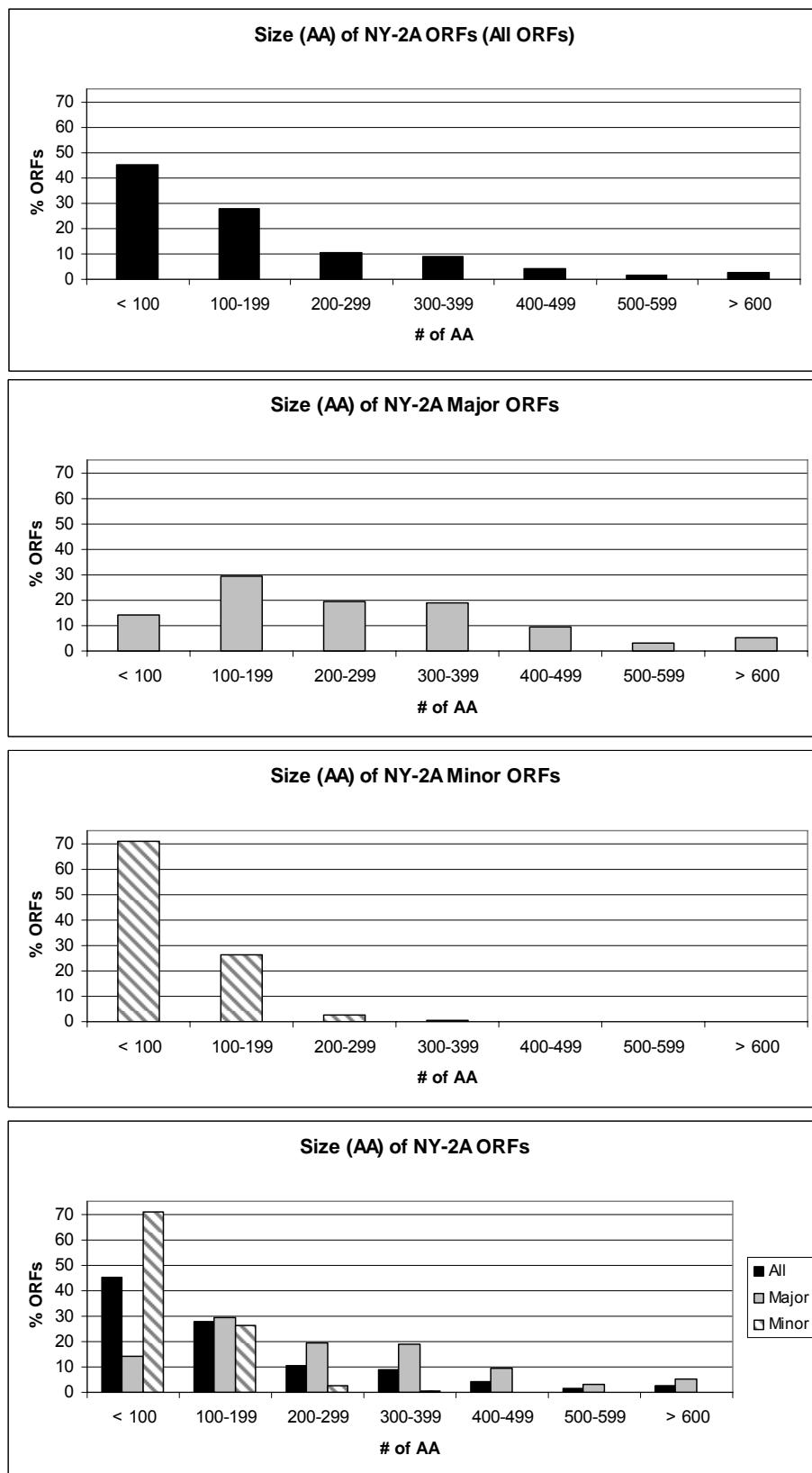


Figure 2.3 Size (# Amino Acids) of the NY-2A ORFs

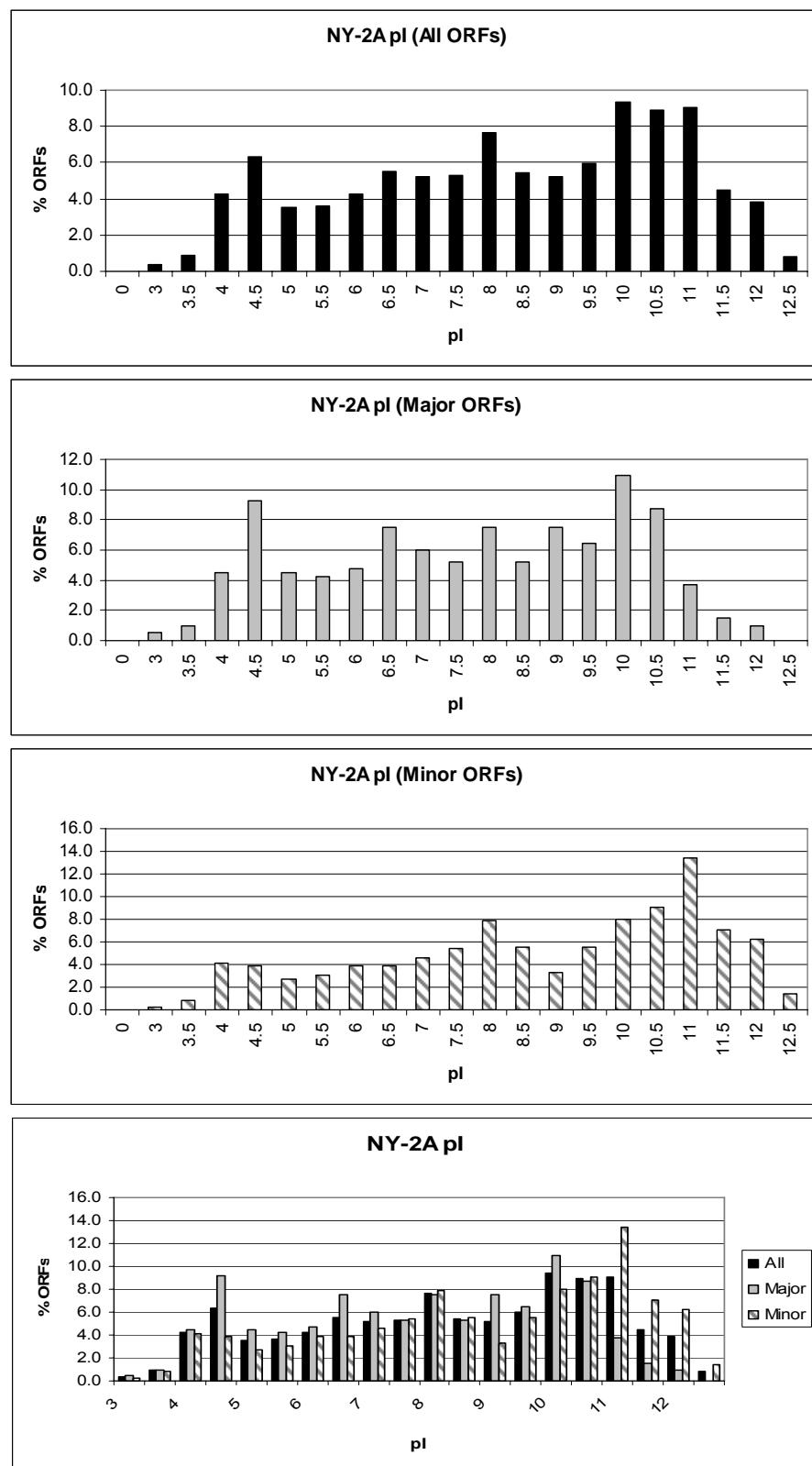


Figure 2.4 Analysis of the pIs from the NY-2A ORFs

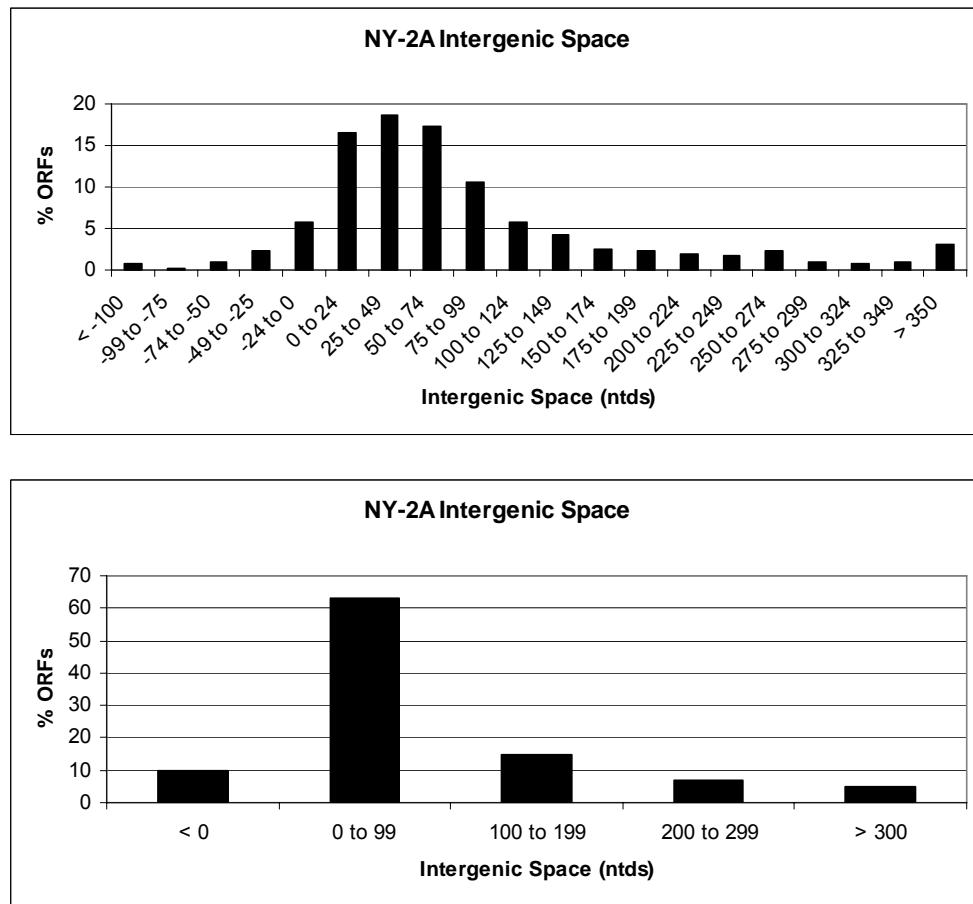


Figure 2.5 Intergenic space between major NY-2A ORFs

2.3.1.4 Annotation of NY-2A Genome

Every ORF was run against the non-redundant (NR) database at NCBI using the criteria described in the experimental methods. The top three hits for each ORF are listed in addendum 2.6.2.2. In summary, Table 2.5 shows that 76% of the 482 minor ORFs have no hits to the NR database and 21% of the minor ORFs only have one hit to the PBCV-1 genome. The remaining 15 minor ORFs which did have a hit in the NR database were either predicted to be membrane proteins or encoded a protein rich in a particular amino acid (e.g. Pro-rich proteins). These results support our classification of these ORFs as minor. Overall, there were 680, or 77%, of NY-2A ORFs which have either no homolog or only a homolog in the PBCV-1 genome. A number of the remaining 191 major ORFs had significant hits in the NR database.

Table 2.5 NY-2A Analysis with the NR Database			
	NY-2A ORFs		
	Minor (482 ORFs)	Major (404 ORFs)	Total (886 ORFs)
No Hit Found	364/76%	42/10%	406/46%
PBCV-1 Homologues	103/21%	171/42%	274/31%

Table 2.5 NY-2A analysis with the NR database. Shown are the number and percentage of NY-2A ORFs with either no hit to an entry in the NR database or only a hit to the PBCV-1 genome.

The Pfam database was used to identify conserved domains within the NY-2A ORFs. Addendum 2.6.2.3 lists the 132 ORFs which were identified as encoding a putative protein containing a Pfam domain. Clusters of Orthologous Groups were detected in 90 major ORFs, which are listed in addendum 2.6.2.4. These 90 ORFs are classified into different metabolic domains (Table 2.6). Twenty of these ORFs matched a COG classified as “poorly characterized” even though they are conserved in at least three

Table 2.6 NY-2A COGs	COGs
Information storage and processing	39
DNA replication, recombination and repair	27
RNA processing and modification	1
Transcription	2
Transcription/ DNA replication, recombination, and repair	2
Translation, ribosomal structure and biogenesis	1
Translation, ribosomal structure and biogenesis/ Nucleotide transport and metabolism	6
Cellular processes	22
Cell division and chromosome partitioning	1
Cell envelope biogenesis, outer membrane	11
Cell envelope biogenesis, outer membrane/ Carbohydrate and metabolism	1
Posttranslational modification, protein turnover, chaperones	5
Signal transduction mechanisms	1
Signal transduction mechanisms / Transcription / DNA replication, recombination, and repair	3
Metabolism	9
Amino acid transport and metabolism	3
Carbohydrate transport and metabolism	2
Energy production and conversion	1
Lipid metabolism	1
Nucleotide transport and metabolism/ Translation, ribosomal structure and biogenesis	1
Secondary metabolites biosynthesis, transport and catabolism	1
Poorly characterized	20
General function prediction only	16
Function unknown	4
TOTAL COGs	90

genomes and, therefore, they are probably significant. Of the remaining 70 ORFs which matched a COG, over half are involved in information storage and processing, such as DNA replication and transcription, and RNA processing. Given the life cycle of this large DNA virus, the importance of these functions for the virus is not surprising. Several other metabolic domains were also detected within the ORFs of the NY-2A genome. The results of the NR, Pfam, and COG databases strengthen our classification of major and minor ORFs. Furthermore, only the ORFs classified as major had a significant hit in either a domain or had a homolog in one or more of these databases.

2.3.1.5 Identification of Inteins and Introns

The Intein Database and Registry (InBase) was used to identify two inteins within the NY-2A ORFs. The NY-2A genome is not the first viral genome to encode more than one intein. However, of the three chlorella viruses sequenced, NY-2A is the only genome to encode an intein. The α -subunit of ribonucleotide reductase (B832R) contains an intein called CIV RIR1 (E value = 5E-66). The CIV RIR1 intein, was the first intein discovered in a dsDNA eukaryotic virus, Chilo iridescent virus (15). A second intein, Lpe Helicase (E value = 1E-31), exists in a putative helicase (B508R). In addition to these two intervening sequences, three introns are present in the NY-2A genome. The DNA polymerase gene (B249/253R) contains a 5'AG/GUGAGU and 3'UGCAG/UU splice-site sequence (20). The *pdg* gene has a 5'AG/GTATGT and 3'-TTGCAG/AA splice site sequence which is characteristic of nuclear-located, spliceosomal processed pre-mRNA introns. This gene contains an 81-nucleotide intron sequence which is also present in another 4 chlorella viruses implying strong selective pressure to maintain the

DNA sequence of the intron when it is in the *pdg* gene. Furthermore, the NY-2A genome contains a small intron (12 bp) in the tRNA^{Tyr}.

2.3.1.6 Identification of Gene Families

One hundred forty-eight of the NY-2A ORFs resemble 1 or more other NY-2A ORFs forming 32 families. 18 families have two members, 1 family has three members, 1 family has four members, 1 family has five members, 3 families have six members, 3 families have seven members, 2 families have eight members, 1 family has twelve members, 1 family has sixteen members, and 1 family has seventeen members.

2.3.2 Analysis and Annotation of the MT325 Genome

The MT325 genome, the smallest of the three *Chlorovirus* genomes sequenced to date, contains 314,335 nucleotides. Like the termini of PBCV-1 and NY-2A, the MT325 genome contains identical inverted repeats, however, they are not both located at the genomes' termini. The 100% identical inverted repeat is 1749 nucleotides long. The MT325 genome is unique when compared to the NC64A virus genomes due to the location of the inverted repeat at the 3' end. This repeat is located 1785 nucleotides upstream from the 3' termini (Addendum 2.6.1.3). The average G+C content of the genome is 45.3%, a level higher than that observed in the *Chlorella* NC64A-infecting viruses, PBCV-1 and NY-2A (40-41%). To determine the orientation of the MT325 genome relative to that of the prototype virus PBCV-1, a comparison of the genomic dot plots for the PBCV-1 and MT325 genomes was used. Based on these results (Figure 2.6), the orientation allowing the best co-linear alignment was determined. Unlike the

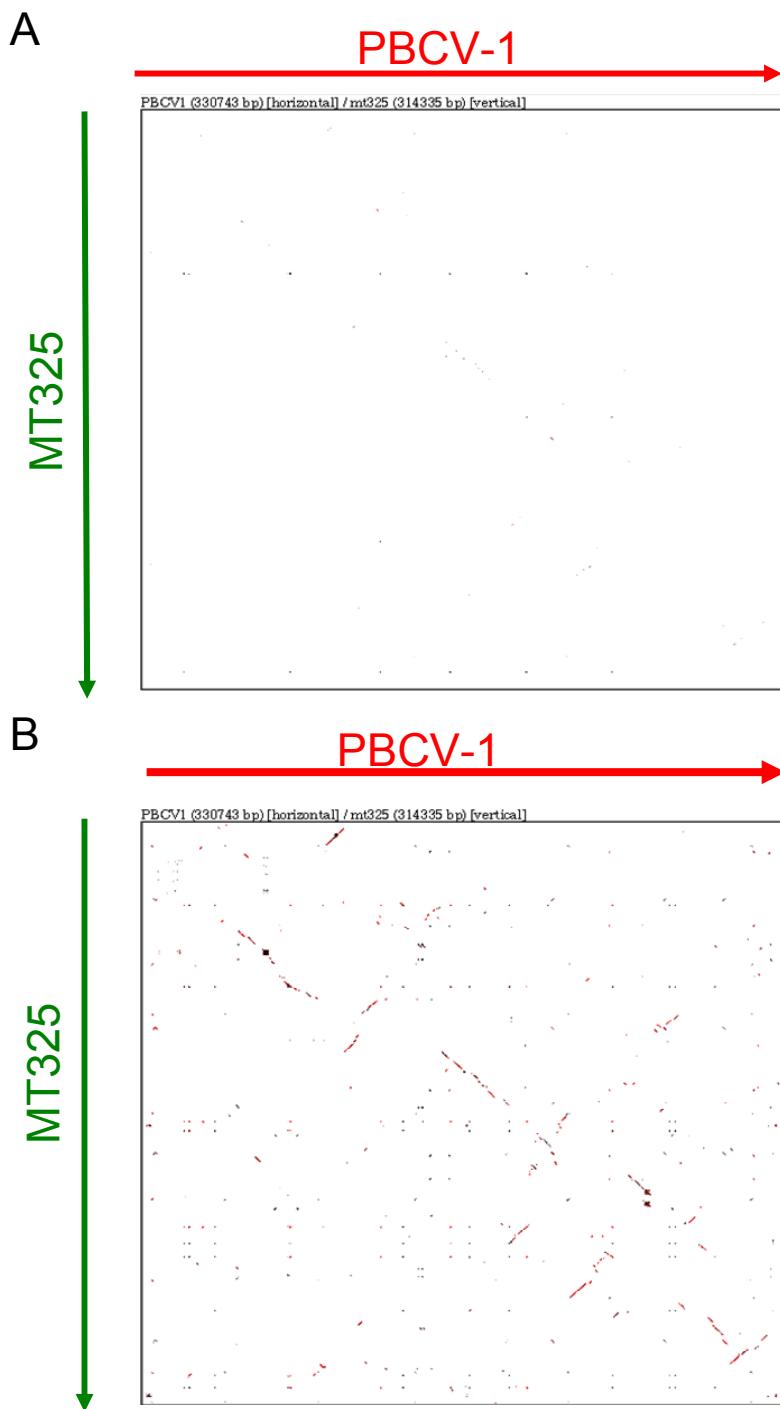


Figure 2.6 Genomic Dot Plots Comparing the Genomes of PBCV-1 to MT325 (A) blastn (DNA vs. DNA) and (B) tblastx (translation vs. translation) of the PBCV-1 genome to the MT325 genome. The dots represent similarities with E-value $<10^{-5}$. The red dots represent similarity > 200 bp and the black dots represent similarity ≤ 200 bp.

genomic dot plots comparing the two *Chlorella* NC64A-infecting viruses, there is only slight co-linearity between the prototype *Chlorovirus* PBCV-1 and the MT325 genomes. As depicted in Figure 2.6, there is essentially no similarity between the PBCV-1 and MT325 genomes at the DNA level ($E\text{-value} < 10^{-5}$). However, the tblastx dot plot identifies the best possible orientation for MT325 genome that allows for marginal co-linearity with the genome of PBCV-1.

2.3.2.1 Identification of Putative tRNAs

The 314 kb MT325 genome was analyzed using the tRNAscan-SE program. Ten putative tRNAs were identified. One tRNA, tRNA^{Tyr}, contains a 12 nucleotide intron from nucleotide 135213 to 135225. The insertion of an intron in the tyrosine tRNA (anticodon GTA) also exists in the PBCV-1 and NY-2A genomes. Table 2.7 lists the 10 tRNAs encoded by the MT325 genome and the frequency of codon usage within the tRNAs relative to that within all genomic ORFs. The percentage with which a certain amino acid was coded for by a particular tRNA as compared to all tRNAs for that

Table 2.7 MT325 tRNAs						
tRNA #	tRNA Begin	tRNA End	Anti-Codon	tRNA Type	% Codon Usage/ Total Codons	% Codon Usage/A tRNA Type
1	134515	134587	TAT	Ile	1.6	25.1
2	134612	134695	TAA	Leu	0.5	6.3
3	134720	134792	GAA	Phe	2.2	48.5
4	134816	134888	TCT	Arg	1.0	22.6
5	134892	134962	TCC	Gly	1.9	27.4
6	134986	135059	GTT	Asn	3.0	52.5
7	135082	135153	GTT	Asn	3.0	52.5
8	135176	135261	GTA	Tyr	1.6	44.1
9	135264	135336	CTT	Lys	4.5	62.1
10	135498	135569	CGT	Thr	1.0	16.2

particular amino acid was examined. For example, the MT325 genome encodes 2 tRNA^{Asn}, accounting for 53% of all the asparagines encoded by all of the ORFs. In general, each tRNA represents 6-62% of the total usage of the particular amino acid for which it codes. The frequency in the usage of the MT325 tRNAs is significantly larger than the range that occurs in the *Chlorella* NC64A viruses, PBCV-1 and NY-2A.

2.3.2.2 Identification of Major and Minor Open Reading Frames

The criteria used for the identification and classification of ORFs have been described and summarized previously. Using these criteria, 845 ORFs, or potential protein-coding regions, were identified in the MT325 genome. The ORF names were based on three criteria. First, the ORF name begins with either an “M”, for a major ORF, or an “m”, for a minor ORF. Second, the ORFs are numbered consecutively in the order in which they appear in the genome as aligned with the PBCV-1 genome. Third, the letter R or L following the ORF number indicates that the putative transcript runs either left-to-right or right-to-left, respectively.

The 845 ORFs of the MT325 genome were classified as either major or minor ORFs based on the following criteria. As described previously, any ORF in which the A+T content of the 50 bases upstream of the start codon is $\geq 70\%$ was originally classified as a major ORF. Using this criterion, 307 ORFs were initially classified as major ORFs. Three criteria determined the reclassification of a major ORF to a minor ORF (Table 2.8 shown is the major ORF grouped together with the ORF that was changed from a major to a minor). First, any major ORF which resides within a larger major ORF of opposite polarity was reclassified as a minor ORF. As depicted in Table 2.8a, 31 ORFs were reclassified from major to minor ORFs using this criterion. Of these 31 ORFs, 23 resided

entirely within the larger major ORF, and the remaining 8 resided almost entirely within the larger ORF (i.e. 4-63 nucleotides of their coding region extended outside the larger major ORF). Second, any major ORF which resides within a larger major ORF of the same polarity was reclassified as a minor ORF. As depicted in Table 2.8b, 5 ORFs were reclassified from major to minor ORFs based on this criterion. As illustrated by Tables 2.8a and 2.8b, all but 5 of the reclassified ORFs resided within major ORFs of the opposite polarity. Third, any major ORF which significantly overlapped with a larger major ORF, regardless of its orientation, was reclassified as a minor ORF. As depicted in Table 2.8c, 3 ORFs were reclassified from major to minor ORFs based on this criterion. Two characteristics were examined to reclassify a minor ORF, which has a < 70% A+T content of the 50 nucleotides upstream of the start codon, to a major ORF (Table 2.9). First, based on the analysis of the ORFs using the non-redundant, Pfam and COG databases, 34 ORFs (depicted in Table 2.9a) which were initially classified as minor ORFs were reclassified as major ORFs because they were determined to encode a putative functional protein. Unlike the NY-2A genome, which infects *Chlorella* NC64A, several MT325 ORFs were reclassified as a major ORF based on such analyses. This unusual finding could be explained if the promoters of MT325 ORFs either have a lower A+T content than traditionally observed or are located in a region other than the 50 nucleotides upstream of the start codon. Second, any minor ORF which resides between but does not overlap with two major ORFs was reclassified as a major ORF. As depicted in Table 2.9b, 29 ORFs were reclassified from minor to major ORFs based on this criterion.

Table 2.8 MT325 ORFs Reclassified from a Major to a Minor

Table 2.8a Changes from a Major to a Minor ORF based on (ii)									
	ORF	Location	% A+T	AA		ORF	Location	% A+T	AA
1	M47R	22096 .. 26496	76%	1467		M460R	177267 .. 178019	76%	251
	m53L	24687 .. 24923	72%	79		m461L	177327 .. 177809	80%	161
2	M55R	26539 .. 30921	84%	1461		M460R	177267 .. 178019	76%	251
	m60L	30757 .. 30984	80%	76		m462L	177671 .. 177892	70%	74
3	M69R	35403 .. 36698	44%	432		M476L	184220 .. 184825	66%	202
	m70L	36332 .. 36685	80%	118		m477R	184604 .. 184816	76%	71
4	M71R	36740 .. 41230	82%	1497		M535L	209244 .. 211856	78%	871
	m77L	39955 .. 40155	70%	67		m534R	209205 .. 209474	72%	90
5	M83R	43344 .. 43598	78%	85		M535L	209244 .. 211856	78%	871
	m84L	43356 .. 43646	76%	97		m536R	209245 .. 209604	72%	120
6	M88R	45967 .. 46605	76%	213		M586L	225321 .. 226085	88%	255
	m89L	45978 .. 46601	74%	208		m585R	225310 .. 225513	72%	68
7	M97L	49549 .. 50625	84%	359		M607L	231219 .. 231719	80%	167
	m98R	49636 .. 50043	74%	136		m608R	231258 .. 231482	70%	75
8	M140L	65487 .. 65990	74%	168		M616L	234347 .. 235330	84%	328
	m142R	65562 .. 65795	72%	78		m617R	234588 .. 234998	70%	137
9	M180L	82906 .. 83379	92%	158		M626L	237798 .. 238229	78%	144
	m182R	83120 .. 83344	76%	75		m625R	237774 .. 238019	78%	82
10	M253L	104888 .. 106765	86%	626		M657L	248729 .. 249244	72%	172
	m254R	105328 .. 105609	70%	94		m658R	248739 .. 249017	72%	93
11	M272L	111928 .. 113304	54%	459		M686L	257446 .. 258678	70%	411
	m273R	111941 .. 112186	82%	82		m685R	257437 .. 257700	76%	88
12	M278R	113482 .. 114282	70%	267		M794R	294905 .. 295714	80%	270
	m280L	114136 .. 114363	78%	76		m795L	295085 .. 295699	72%	205
13	M283R	114950 .. 116107	76%	386		M797R	295768 .. 296901	76%	378
	m287L	115906 .. 116160	80%	85		m800L	296483 .. 296908	82%	142
14	M417L	165593 .. 167392	76%	600		M803L	297839 .. 299338	70%	500
	m418R	165741 .. 166007	70%	89		m804R	297846 .. 298169	74%	108
15	M449L	173706 .. 174017	76%	104		M820R	304499 .. 304954	78%	152
	m450R	173707 .. 173970	78%	88		m821L	304520 .. 304786	72%	89
16	M451L	174043 .. 175368	72%	442					
	m452R	174059 .. 174325	76%	89					

Table 2.8a MT325 ORFs changed from a major to a minor based on criterion (ii) Thirty-one ORFs were reclassified from a major to a minor ORF based on criterion (ii), which states that if a major ORF resides within a larger major ORF of the opposite polarity it will be reclassified as a minor ORF. The table lists the ORFs which met this criterion along with their location, % A+T 50 nucleotides upstream, and the number of amino acids encoded.

Table 2.8b Changes from a Major to a Minor ORF based on (iii)

	ORF	Location		% A+T	AA
1	M289R	117311	..	118414	72%
	m290R	117315	..	117527	72%
2	M367L	143740	..	144396	80%
	m368L	144146	..	144406	80%
3	M488L	187886	..	188632	80%
	m490L	188398	..	188607	74%
4	M618L	235418	..	235642	80%
	m619L	235443	..	235658	76%
5	M735L	274810	..	275268	70%
	m736L	274995	..	275201	76%

Table 2.8c Changes from a Major to a Minor ORF based on (iv)

	ORF	Location		% A+T	AA
1	M186R	84369	..	85928	74%
	m185R	84298	..	84501	76%
2	M444R	172564	..	172911	74%
	m441L	172141	..	172878	70%
3	M770R	287716	..	288117	82%
	m771R	287942	..	288343	70%

Table 2.8b MT325 ORF changed from a major to a minor ORF based on criterion (iii). Five ORFs were reclassified from a major to a minor ORF based on criterion (iii), which states that if a major ORF resides within a larger major ORF of the same polarity it will be reclassified as a minor ORF. **Table 2.8c MT325 ORF changed from a major to a minor ORF based on criterion (iv).** Three ORFs were reclassified from a major to a minor ORF based on criterion (iv), which states that if a major ORF overlapped with larger ORFs, regardless of direction, it was reclassified as a minor. These tables list the ORFs which met this criterion along with the location, % A+T 50 nucleotides upstream, and the number of amino acids the ORF encoded.

Table 2.9 MT325 ORFs Reclassified from a Minor to a Major.

Table 2.9a Changes from a Minor to a Major ORF based on (v)			
	ORF	Description	Source
1	M10L	Deoxycytidylate deaminase	Pfam, COG
2	M26L	2-hydroxyacid dehydrogenase	Pfam, COG
3	M34L	Thymidylate synthase	Pfam, COG
4	M37R	Glutamine amidotransferase	Pfam, COG
5	M91R	Glycosyl hydrolase	Pfam
6	M103L	Carbon-nitrogen hydrolase	Pfam, COG
7	M124L	Glycosyl hydrolase	Pfam, COG
8	M139L	Transcription factor TFIIB	Pfam, COG
9	M221L	Protein kinase	Pfam, COG
10	M264L	dUTPase	Pfam, COG
11	M272L	Helicase	Pfam, COG
12	M302L	AAA ATPase	COG
13	M319L	33 kDa translation protein	NR
14	M341R	33 kDa translation protein	NR
15	M354R	Cellulase	NR
16	M359L	Cytosine DNA methylase	Pfam, COG
17	M389R	ATPase PP-loop family	Pfam, COG
18	M457R	ATP binding motif	NR
19	M476L	Non-histone chromosomal protein	Pfam
20	M530L	Deoxycytidylate deaminase	Pfam, COG
21	M531L	Ribonucleotide reductase	NR
22	M543R	Protein kinase	Pfam, COG
23	M570L	RNase H	Pfam, COG
24	M627L	Pyrimidine dimer DNA glycosylase	Pfam
25	M653L	Ribonucleotide reductase	Pfam, COG
26	M672L	RNase III	Pfam, COG
27	M674R	Poxvirus D5 protein-like	Pfam, COG
28	M697L	PCNA	Pfam, COG
29	M742R	ABC transporter	Pfam, COG
30	M758R	Acetyltransferase	Pfam
31	M807R	Skp1 family protein	Pfam, COG
32	M822R	Lipase	Pfam, COG
33	M829L	Ankyrin repeat	Pfam, COG
34	M833L	Ankyrin repeat	Pfam, COG

Table 2.9a MT325 ORF changed from a minor to a major ORF based on criterion (v).
Thirty-four ORFs were reclassified from a minor to a major based on criterion (v), which states that if an analysis of an ORF against the NR, Pfam, or COG databases revealed that the ORF had a putative function it would be reclassified as a major ORF.

Table 2.9b Changes from a Minor to a Major ORF based on (vi)			
	ORF		ORF
1	M3L	16	M454R
2	M93L	17	M472R
3	M116L	18	M515R
4	M122R	19	M569L
5	M136L	20	M577L
6	M148L	21	M647L
7	M200L	22	M706L
8	M237L	23	M715R
9	M298R	24	M740L
10	M315L	25	M773L
11	M346L	26	M817R
12	M361R	27	M828R
13	M411L	28	M838L
14	M421L	29	M845L
15	M435L		

Table 2.9b MT325 ORF changed from a minor to a major ORF based on criterion (vi).
Twenty-nine ORFs were reclassified from a minor to a major based on criterion (vi), which states that if a minor ORF resided between two major ORFs, without overlapping, it was reclassified as a major ORF.

2.3.2.3 Characterization of Open Reading Frames

Using the above-described criteria, 331 of the 845 open-reading frames (ORFs) were classified as major ORFs. For each ORF, the GCG software was used to identify several general characteristics and properties: the nucleotide composition of the ORF, the A+T content of the 50 nucleotides upstream of the ORF, the frame in which the putative protein was encoded, the number of amino acids in the encoded protein, and the putative encoded protein's molecular weight, pI, and extinction coefficient. These general characteristics are summarized in addendum 2.6.3.1. Following the classification and characterization of every ORF, relative genomic analyses of the ORFs were performed. In Figure 2.7, the orientation of the ORFs, relative to either the frame in

which they were encoded or to their overall direction within the MT325 genome, are depicted. In general, the directions and the frames in which the ORFs were encoded were slightly skewed to the reverse orientation. The size of the putative encoded protein (e.g. the number of encoded amino acids) for each ORF is shown in Figure 2.8. Over 60% of the minor ORFs encoded proteins containing less than 100 amino acids. The analysis of the pIs of the putative proteins encoded by the ORFs is depicted in Figure 2.9. Despite a basic trend in pIs observed for the major ORFs, a peak at pI 4.5 was also observed. Basic proteins are likely to be associated with the virion where they function to neutralize the positively-charged genomic DNA. However, the function of the proteins corresponding to the peak at pI 4.5 varies (i.e., cellulose, protein kinase, and beta-1,3 glucanase). Figure 2.10 illustrates the intergenic space between the major ORFs. The majority of the ORFs were separated by less than 100 nucleotides. Together, these figures illustrate several characteristics of the ORFs encoded by the MT325 viral genome. In general, the major ORFs are evenly distributed along the genome with very little overlap.

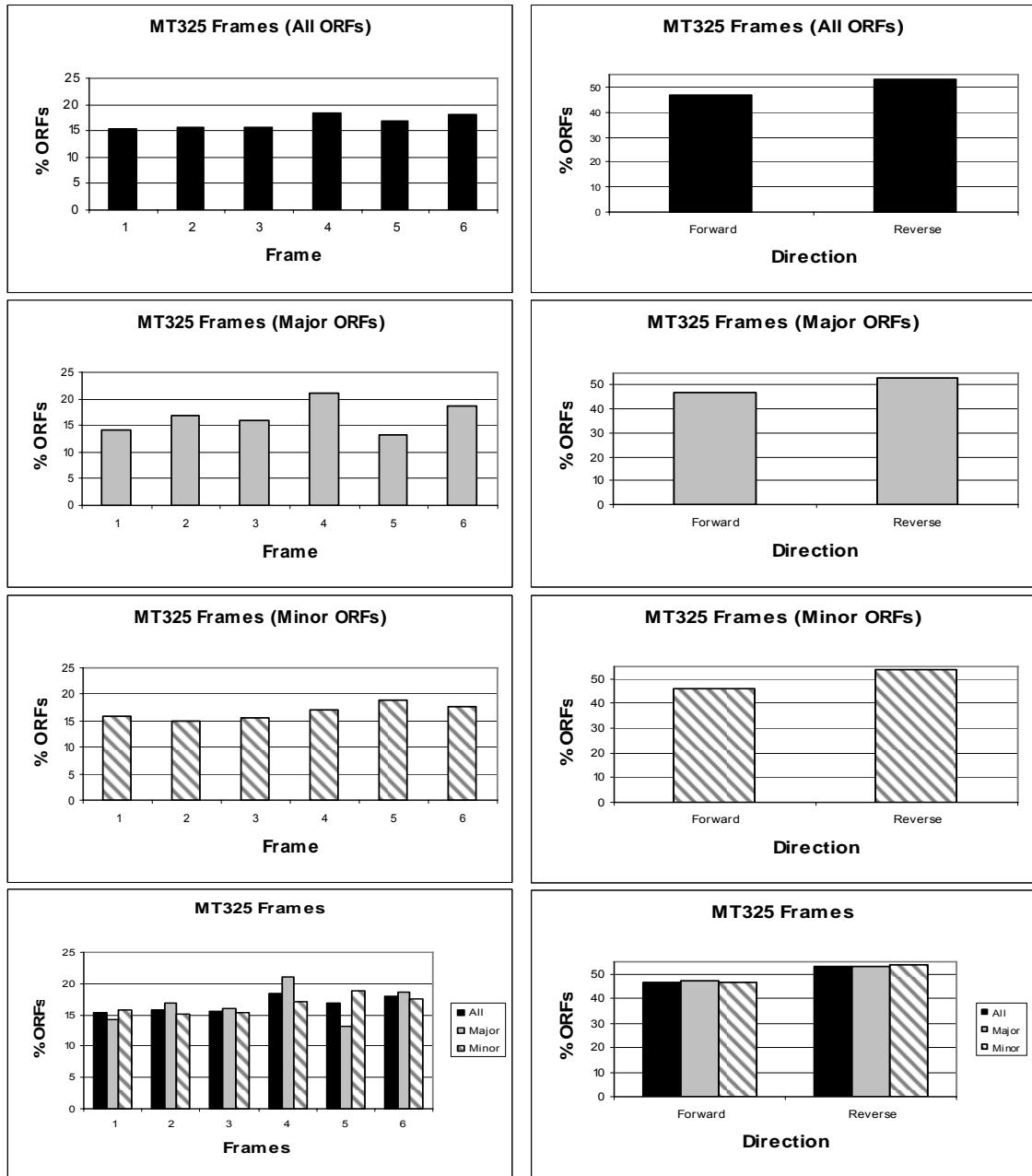


Figure 2.7 Orientation of MT325 ORFs relative to either the frame in which they were encoded or to their overall direction.

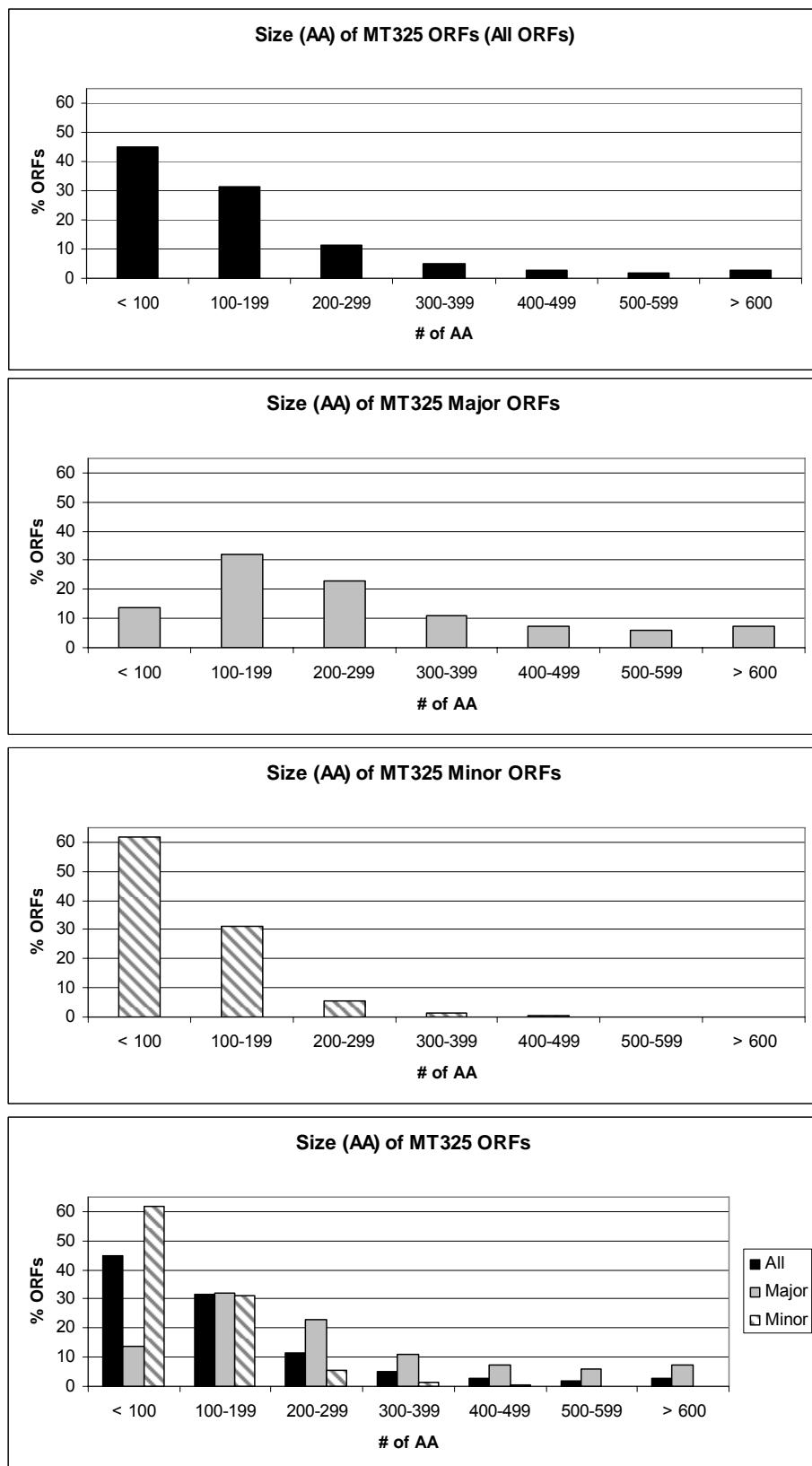


Figure 2.8 Size (# Amino Acids) of the MT325 ORFs

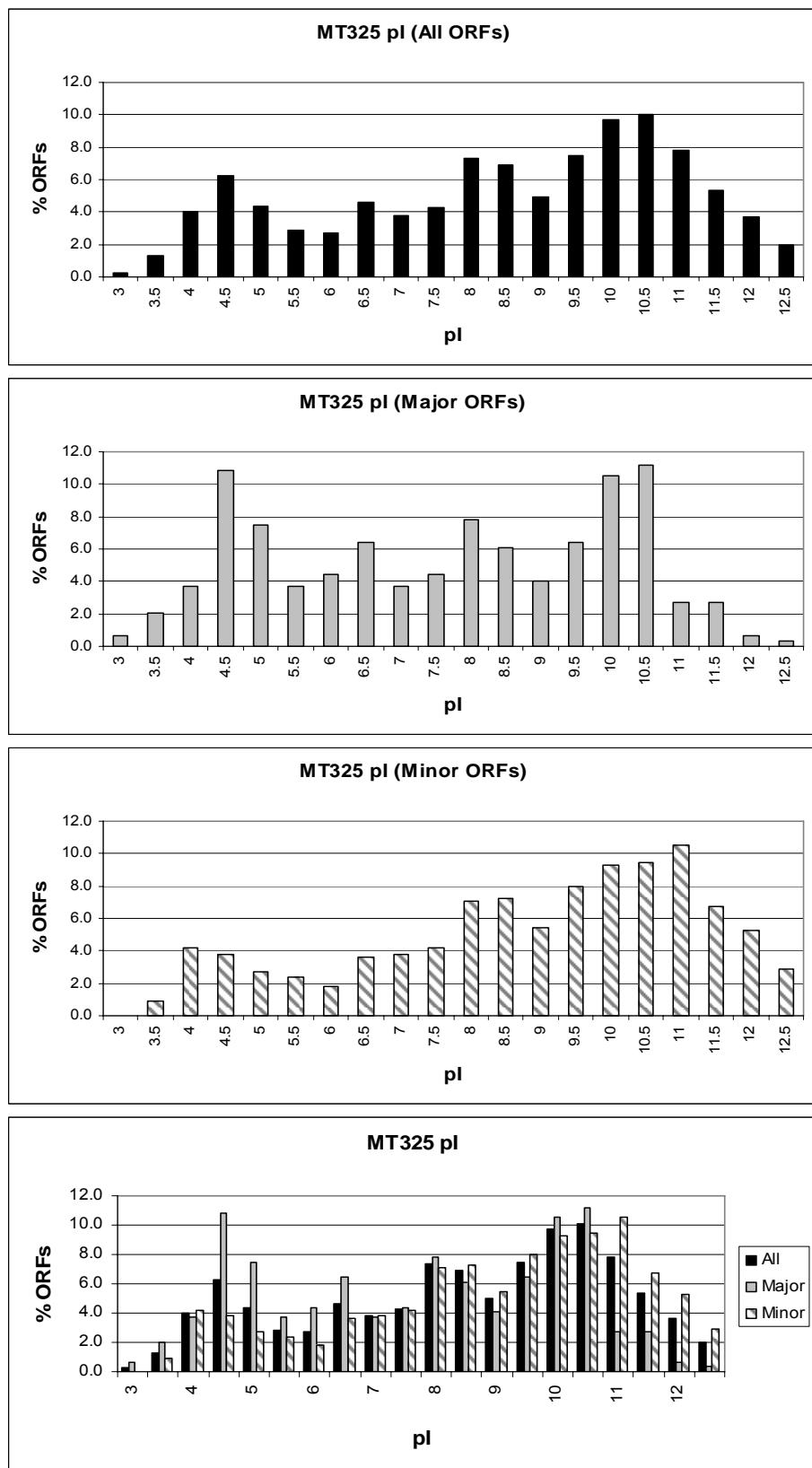


Figure 2.9 Analysis of the pIs from the MT325 ORFs

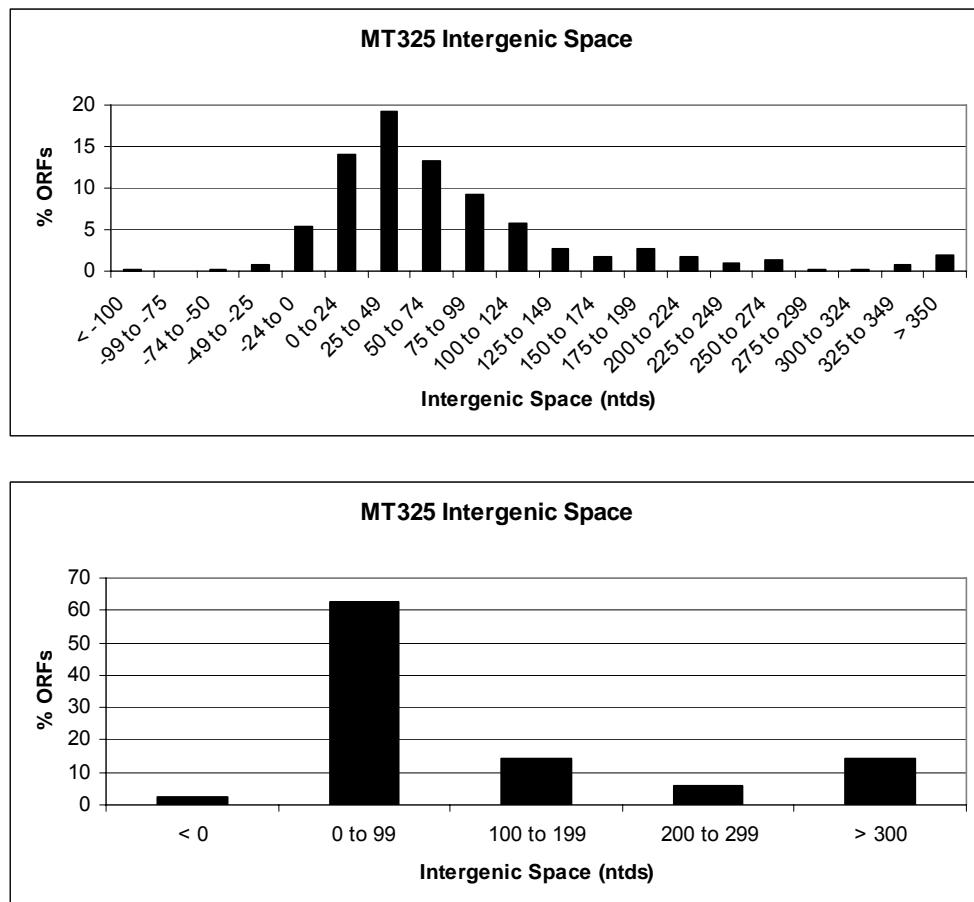


Figure 2.10 Intergenic space between major MT325 ORFs

2.3.2.4 Annotation of MT325 Genome

Every ORF was run against the non-redundant (NR) database at NCBI using the criteria described in the experimental methods. The top three hits for each ORF are listed in addendum 2.6.3.2. In summary, Table 2.10 shows that 96% of the minor ORFs either have no hits in the NR database or it only hits an ORF in the PBCV-1 genome. The remaining 18 minor ORFs which did have a hit in the NR database were either membrane proteins or encoded a protein rich in a particular amino acid (e.g. Pro-rich proteins). These results support our classification of these ORFs as minor. Overall, there were 689, or 82%, of MT325 ORFs which have either no homolog or only a homolog in the PBCV-1 genome; a number slightly higher than the NY-2A genome. Many of the remaining 138 major ORFs had significant hits in the NR database.

Table 2.10 MT325 Analysis with the NR Database			
	MT325 ORFs		
	Minor (514 ORFs)	Major (331 ORFs)	Total (845 ORFs)
No Hit Found	463/90%	52/16%	515/61%
PBCV-1 Homologues	33/6%	141/43%	174/21%

Table 2.10 MT325 analysis with the NR database. Shown are the number and percentage of MT325 ORFs with either no hit to an entry in the NR database or only a hit to the PBCV-1 genome.

The Pfam database was used to identify conserved domains within the MT325 ORFs. Addendum 2.6.3.3 lists the 101 ORFs which were identified as encoding a putative protein containing a Pfam domain. Clusters of Orthologous Groups were detected in 68 major ORFs, which are listed in addendum 2.6.2.4. These 68 ORFs are classified into different metabolic domains (Table 2.11). Fourteen of these ORFs

Table 2.11 MT325 COGs	COGs
Information storage and processing	15
DNA replication, recombination and repair	7
RNA processing and modification	1
Transcription	4
Transcription/ DNA replication, recombination, and repair	2
Translation, ribosomal structure and biogenesis/ Nucleotide transport and metabolism	1
Cellular processes	24
Cell division and chromosome partitioning	1
Cell envelope biogenesis, outer membrane	13
Inorganic ion transport and metabolism	2
Posttranslational modification, protein turnover, chaperones	4
Signal transduction mechanisms	1
Signal transduction mechanisms / Transcription / DNA replication, recombination, and repair	3
Metabolism	15
Amino acid transport and metabolism	3
Carbohydrate transport and metabolism	3
Energy production and conversion/ Coenzyme metabolism/ General function prediction only	1
Nucleotide transport and metabolism	6
Secondary metabolites biosynthesis, transport and catabolism	2
Poorly characterized	14
General function prediction only	12
Function unknown	2
TOTAL COGs	68

matched a COG classified as “poorly characterized” even though they are conserved in at least three genomes and, therefore, must be important. There are 54 ORFs remaining which matched a COG. Unlike the NY-2A genome, the MT325 genome contained more COGs involved in cellular processes than information storage and metabolism. The difference in the types of COGs found within these genomes might be related to the host that they infect. Several other metabolic domains were also detected within the ORFs of the MT325 genome. The results of the NR, Pfam, and COG databases strengthen our classification of major and minor ORFs. Furthermore, only the major ORFs either had a significant hit in a domain or had a homolog in one or more of these databases.

2.3.2.5 Identification of Inteins and Introns

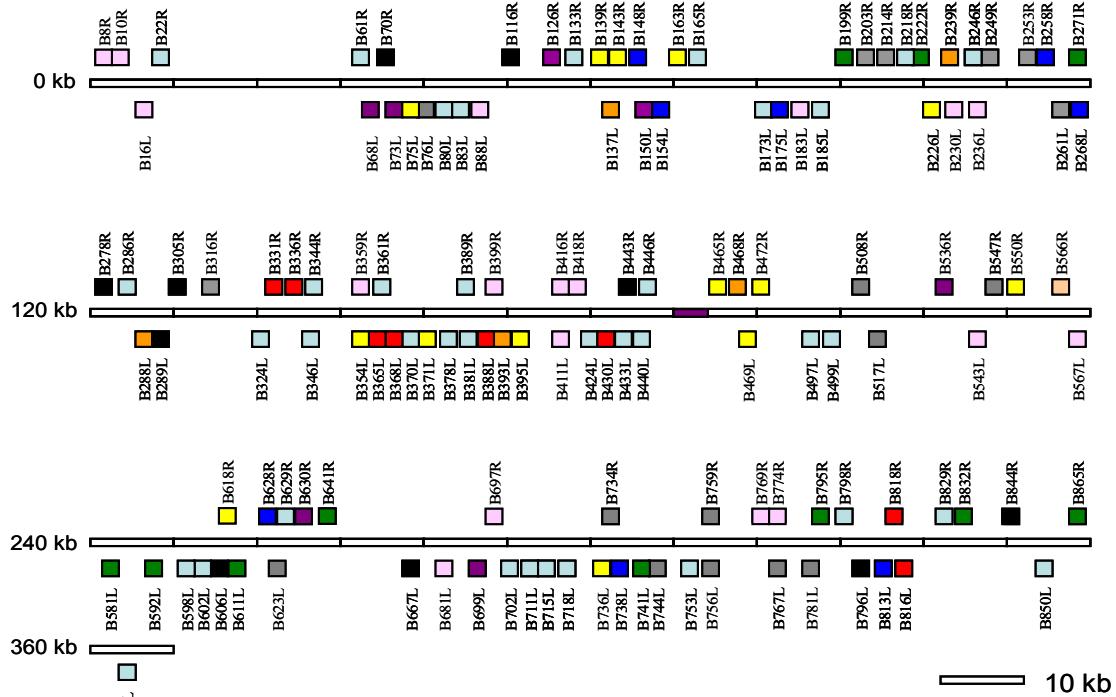
The Intein Database and Registry (InBase) was used to determine if an MT325 ORF had the presence of an intein. Unlike the NY-2A genome, which encodes two inteins, no inteins were detected in the MT325 genome. Furthermore, the MT325 ORFs were scanned using the conserved intron sequences from the PBCV-1 genome. Every ORF was additionally compared to their homologs to determine the presence of an intervening sequence. The only intervening sequence detected in the MT325 genome was a small intron (12 bp) in the tRNA^{Tyr}. This intron is found in all of the tRNA^{Tyr} encoded by the three sequenced Chlorella viruses.

2.3.2.6 Identification of Gene Families

Ninety-one of the MT325 ORFs resemble 1 or more other MT325 ORFs forming 24 families. 15 families have two members, 2 families have three members, 2 families have five members, 2 families have seven members, 1 family has eight members, 1 family has nine members, and 1 family has fourteen members.

2.4 DISCUSSION

Sequencing of the NY-2A and MT325 chlorella viral genomes revealed several new and important findings. The 368,683-bp NY-2A genome, the largest chlorella viral genome sequenced to date, contains 886 ORFs of 65 codons or larger and likely encodes 404 different proteins. The 314,335-bp MT325 genome, the smallest chlorella viral genome sequenced to date, contains 845 ORFs and likely encodes 331 different proteins. In both genomes, the putative protein-encoding genes are evenly distributed on both strands, and the intergenic space is minimal. Approximately half of the putative gene products have been potentially identified; they consist of both prokaryotic- and eukaryotic-like proteins. Figure 2.11 and Figure 2.12 illustrate a partial gene map of genomes of the chlorella viruses NY-2A and MT325, respectively, as annotated using the NR, Pfam, and COG databases. The partial gene map demonstrates that the major ORFs are found throughout the entire genome and are found in both orientations.

Figure 2.11 – NY-2A Partial Gene Map

DNA Replication, Recombination, and Repair	
B633R	Archaeo-eukaryotic primase
B759R	ATPase (Chromosome segregation)
B536R	ATPase (DNA packaging)
B756L	ATPase (DNA repair)
B744L	ATPase (PP-loop)
B734R	ATP-dependent DNA ligase
B591L	DNA binding protein
B781L	DNA Topoisomerase II
B214R	Exonuclease
B623L	Helicase-Superfamily III
B261L	PCNA
B767L	PCNA
B76L	Pyrimidine dimer-specific glycosylase
B571L	Replication factor C
B547R	RNase H
B249/253R	δ DNA polymerase

Sugar and Lipid Manipulation	
B580R	6-phosphofructokinase (?)
B139R	Chitin synthase
B472R	Chitin synthase
B371L	Choloylglycine hydrolase
B419L	Fructose-2,6 bisphosphatase
B395L	Fucose synthase
B159R	Fucosyl-Glycosyltransferase
B163R	GDP-D-mannose dehydratase
B143R	Glucosamine synthetase
B692R	Glycerophosphodiesterase (?)
B75L	Glycerophosphoryl diesterase
B618R	Glycosyltransferase
B736L	Glycosyltransferase
B550R	Lipoprotein lipase
B354L	Lysophospholipase
B853L	N-acetyltransferase
B226L	Patatin-like phospholipase
B469L	Polysaccharide deacetylase
B465R	UDP-glucose 6-dehydrogenase

Miscellaneous	
B606L	ABC transporter protein
B844R	Arginine iminohydrolase
B371L	Amidase
B796L	Histidine decarboxylase
B305R	Homoserpidine synthase
B289L	Monoamine oxidase
B116R	N-carbamoylput. amidohydrolase
B278R	Ornithine/Arginine decarboxylase
B322R	Pathogenesis-related protein

Signaling	
B430L	Dual specificity phosphatase
B336R	Potassium channel protein
B331R	Serine/Threonine protein kinase
B365L	Serine/Threonine protein kinase
B368L	Serine/Threonine protein kinase
B388L	Serine/Threonine protein kinase
B816L	Serine/Threonine protein kinase
B818R	Serine/Threonine protein kinase

DNA Restriction/Modification	
B10R	Adenine methyltransferase
B16L	Adenine methyltransferase
B230L	Adenine methyltransferase
B236L	Adenine methyltransferase
B359R	Adenine methyltransferase
B399R	Adenine methyltransferase
B416R	Adenine methyltransferase
B418R	Adenine methyltransferase
B543L	Adenine methyltransferase
B567L	Adenine methyltransferase
B774R	Adenine methyltransferase
B8R	Cytosine methyltransferase
B88L	Cytosine methyltransferase
B411L	Cytosine methyltransferase
B566R	Cytosine methyltransferase
B681L	Cytosine methyltransferase
B697R	Cytosine methyltransferase
B769R	Cytosine methyltransferase
B361R	DNase restriction endonuclease
B183L	Methyltransferase

Nucleotide Metabolism	
B222R	Aspartate transcarbamylase
B271R	Cytosine deaminase
B795R	dCMP deaminase
B741L	dUTP pyrophosphatase
B592L	Glutaredoxin
B457L	NTP pyrophosphohydrolase
B832R	Ribo. Reductase (large subunit)
B641R	Ribo. Reductase (small subunit)
B581L	Thioredoxin
B865R	Thymidylate synthase X

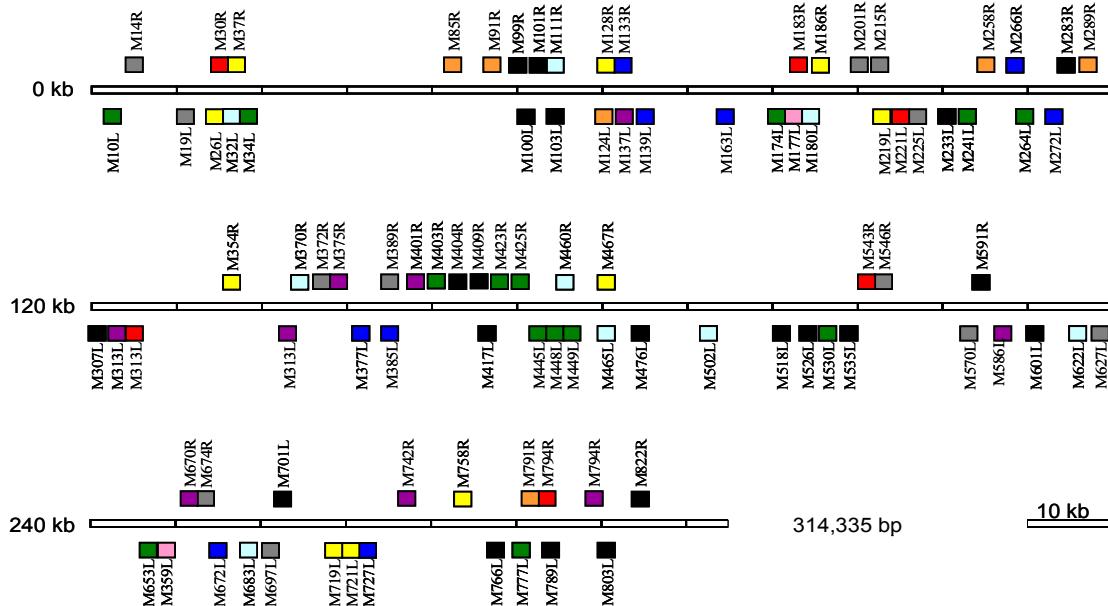
Transcription	
B813L	Histone H3, Lys 27 methylase
B268L	Histone H3, Lys 27 methylase
B148R	mRNA guanylyltransferase
B612R	RNA triphosphatase
B628R	RNAse III
B203R	Superfamily II helicase
B316R	Superfamily II helicase
B505R	Superfamily II helicase
B258R	SWI/SNF chromatin remodeling complex
B738L	SWI/SNF helicase
B154L	Transcription factor TFIIB
B743R	Transcription factor TFIID
B175L	Transcription factor TFIIS
B647R	VLTF2-type transcription factor

Cell Wall Degradation	
B239R	Chitinase
B393L	Chitosanase
B288L	β & α, 1,4 linked glucuronic lyase
B468R	β & α, 1,4 linked glucuronic lyase
B137L	β-1,3-glucanase

Integration and Transposition	
B39L	GIY-YIG endonuclease
B185L	GIY-YIG endonuclease
B206L	GIY-YIG endonuclease
B246R	GIY-YIG endonuclease
B286R	GIY-YIG endonuclease
B346L	GIY-YIG endonuclease
B361R	GIY-YIG endonuclease
B389R	GIY-YIG endonuclease
B433L	GIY-YIG endonuclease
B440L	GIY-YIG endonuclease
B497L	GIY-YIG endonuclease
B499L	GIY-YIG endonuclease
B602L	GIY-YIG endonuclease
B629R	GIY-YIG endonuclease
B724R	GIY-YIG endonuclease
B850L	GIY-YIG endonuclease
B22R	HNH endonuclease
B61R	HNH endonuclease
B133R	HNH endonuclease
B165R	HNH endonuclease
B173L	HNH endonuclease
B199R	HNH endonuclease
B218R	HNH endonuclease
B324L	HNH endonuclease
B370L	HNH endonuclease
B424L	HNH endonuclease
B446R	HNH endonuclease
B598L	HNH endonuclease
B718L	HNH endonuclease
B747L	HNH endonuclease
B753L	HNH endonuclease
B798R	HNH endonuclease
B805R	HNH endonuclease
B878L	HNH endonuclease
B83L	Resolvase
B381L	Resolvase
B715L	Resolvase
B80L	Transposase
B334R	Transposase
B378L	Transposase
B522L	Transposase
B702L	Transposase
B711L	Transposase
B829R	Transposase

Protein Synthesis, Modification, and Degradation	
B73L	ATPase (AAA+ class)
B126R	Prolyl-4-hydroxylase
B611L	Protein disulfide isomerase
B68L	SKP-1 protein
B630R	Thiol oxidoreductase
B699L	Ubiquitin
B150L	Ubiquitin C-terminal hydrolase
B803L	Zn metallopeptidase
7 tRNAs	

Figure 2.12 – MT325 Partial Gene Map



DNA Replication, Recombination, and Repair	
M664L	Archaeo-eukaryotic primase
M253L	ATPase (Chromosome segregation)
M586L	ATPase (DNA packaging)
M258L	ATPase (DNA repair)
M389R	ATPase (PP-loop)
M476L	DNA binding protein
M546R	DNA Topoisomerase II
M215R	Exonuclease
M674R	Helicase-Superfamily III
M14R	PCNA
M697L	PCNA
M627L	Pyrimidine dimer-specific glycosylase
M430L	Replication factor C
M570L	RNase H
M19L	δ DNA polymerase

Nucleotide Metabolism	
M10L	Cytosine deaminase
M530L	dCMP deaminase
M425R	Deoxyuridine kinase
M264L	dUTP pyrophosphatase
M241L	Glutaredoxin
M423R	Glutaredoxin
M335L	NTP pyrophosphohydrolase
M174L	Nucleotidyl transferase
M777L	Ribo. Reductase (large subunit)
M653L	Ribo. Reductase (small subunit)
M445L	Thioredoxin
M448L	Thioredoxin
M449L	Thioredoxin
M34L	Thymidylate synthase X

Integration and Transposition	
M32L	G1Y-YIG endonuclease
M111R	G1Y-YIG endonuclease
M180L	G1Y-YIG endonuclease
M288L	G1Y-YIG endonuclease
M370R	G1Y-YIG endonuclease
M460R	G1Y-YIG endonuclease
M465L	G1Y-YIG endonuclease
M502L	G1Y-YIG endonuclease
M517R	G1Y-YIG endonuclease
M683L	G1Y-YIG endonuclease
M69R	HNH endonuclease
M93L	HNH endonuclease
M578R	HNH endonuclease
M622L	HNH endonuclease
M844R	Transposase

Sugar and Lipid Manipulation	
M444R	6-phosphofructokinase (?)
M354R	Cellulase precursor
M186R	Cellulase synthase
M26L	D-lactate dehydrogenase
M100L	Fructose-2,6 bisphosphatase
M467R	Fucosyl-Glycosyltransferase
M37R	Glucosamine synthetase
M501R	Glycerophosphodiesterase (?)
M186R	Glycosyltransferase
M491R	Glycosyltransferase
M721L	Glycosyltransferase
M128R	Hyaluronan synthase
M758R	N-acetyltransferase
M219L	Patatin-like phospholipase
M719L	UDP-glucose 6-dehydrogenase

Protein Synthesis, Modification, and Degradation	
M302L	ATPase (AAA+ class)
M363L	ATPase (AAA+ class)
M401R	Prolyl 4-hydroxylase
M403R	Protein disulfide isomerase
M807R	SKP-1 protein
M670R	Thiol oxidoreductase
M375R	Translation elongation factor-3
M742R	Translation elongation factor-3
M137L	Ubiquitin C-terminal hydrolase
10 tRNAs	

Transcription	
M727L	Histone H3, Lys 27 methylase
M385L	Histone-like protein
M133R	mRNA guanylyltransferase
M399L	RNA triphosphatase
M672L	RNase III
M201R	Superfamily II helicase
M225L	Superfamily II helicase
M372R	Superfamily II helicase
M15L	SWI/SNF chromatin remodeling complex
M272L	SWI/SNF helicase
M139L	Transcription factor TFIIB
M266R	Transcription factor TFIID
M163L	Transcription factor TFIIS
M635R	VLTF2-type transcription factor

Cell Wall Degradation	
M85R	Chitinase
M791R	Chitinase
M91R	Chitosanase
M289R	β & α 1,4 linked glucuronic lyase
M124L	β-1,3-glucanase

Miscellaneous	
M404R	ABC transporter protein
M766L	Agmatine iminohydrolase
M101R	Amidase
M518L	Autotransporter adhesin
M526L	Autotransporter adhesin
M535L	Ca ²⁺ transporting ATPase
M99R	Cu/Zn-superoxide dismutase
M789L	Fibronectin binding protein
M601L	Histidine decarboxylase
M233L	Homoserine synthase
M283R	Monoamine oxidase
M103L	N-carbamoylput. amidohydrolase
M307L	Ornithine/Arginine decarboxylase
M822R	Pathogenesis-related protein

Signaling	
M30R	Aquaglyceroporin
M313L	Dual specificity phosphatase
M183R	Potassium channel protein
M143L	Serine/Threonine protein kinase
M221L	Serine/Threonine protein kinase
M543R	Serine/Threonine protein kinase
M729L	Serine/Threonine protein kinase
M733L	Serine/Threonine protein kinase
M794R	Serine/Threonine protein kinase

DNA Restriction/Modification	
M359L	Cytosine methyltransferase
M177L	Methyltransferase

2.4.1 DNA Replication and Repair-Associated Proteins

The NY-2A and MT325 genomes contain several ORFs that are involved in either DNA replication, recombination, or repair, such as a B-family DNA polymerase, a type-II DNA topoisomerase, a superfamily II RNA helicase, two ORFs that resemble sliding clamp processivity factor (PCNA) proteins, and a pyrimidine-specific glycosylase.

The gene encoding the NY-2A DNA polymerase (B249/253R) contains an 86-nucleotide intron with 5'-AG/GUGAGU and 3'UGCAG/UU splice site sequences, as well as a possible branch point UCAC sequence. These sequences are hallmarks of spliceosomally-spliced mRNA introns. Experiments indicate that the DNA polymerase genes from 38 other NC64A viruses also have an intron at the same position (19). Interestingly, the DNA polymerase gene from MT325 (M19L), a Pbi virus, lacks this intron.

NY-2A encodes an ATP-dependent DNA ligase (B734R). Several other viruses, including vaccinia virus, ASFV, and bacteriophages T3, T4, and T7 encode a homologous ligase. This enzyme joins two DNA fragments by catalyzing the formation of an internucleotide phosphodiester bond. They range in size from 269 amino acids (in *Haemophilus influenza*) (23) to 1070 amino acids (in *Xenopus laevis*) (24). PBCV-1 encodes the smallest functional eukaryotic ATP-dependent DNA ligase (298 amino acids) (25-28). The NY-2A ligase which is 323 amino acids, is slightly larger.

A type-II DNA topoisomerase, which has >40% amino acid identity with type-II topoisomerases from several eukaryotic organisms, is encoded by both the NY-2A (B781L) and MT325 (M546R) genomes. These enzymes are ATP-dependent and function to thread a dsDNA segment through a transient double-strand break (29).

PBCV-1 encodes one of the smallest, characterized type II DNA topoisomerase, 1061 amino acids (30). It cleaves dsDNAs ~30 times faster than the human type II DNA topoisomerase (31). However, the smallest characterized topoisomerase II enzyme is encoded by CVM-1 (32). The CVM-1 type II enzyme is only 1058 amino acids in length and displays DNA cleavage activity that is approximately 50-fold greater than that of human topoisomerase II alpha. The NY-2A type-II DNA topoisomerase has 90% amino acid identity and is the same size as its PBCV-1 homolog (1061 amino acids), while the putative type-II DNA topoisomerase encoded by the MT325 genome is the same size as the CVM-1 enzyme (1058 amino acids) with only 67% amino acid identity to the PBCV-1 topoisomerase.

Similar to the PBCV-1 genome, the genomes of the NY-2A and MT325 viruses encode proteins that resemble PCNA-like proteins from other organisms. The NY-2A (B261L and B767L) and MT325 (M14R and M697L) proteins are more similar to their homologs from other organisms than they are to each other. This finding suggests these viruses independently acquired their PCNA genes. PCNA interacts with proteins involved in not only DNA replication but also DNA repair and post-replicative processing, such as DNA methylases and DNA transposases (33). Because these viruses encode proteins involved in both DNA repair and DNA methylation, their two PCNAs may serve different functions in their respective viral life cycles.

2.4.2 Transcription-Associated Proteins

In the three sequenced chlorella viral genomes, no recognizable RNA polymerase or a RNA polymerase component has been detected. This observation supports the idea

that infecting viral DNAs are targeted to the nucleus and that host RNA polymerase(s) initiate viral transcription, possibly in conjunction with virion-packaged transcription factors. NY-2A and MT325 encode at least three transcription factor-like elements: TFIIB (B154L, M139L), TFIID (B743R, M266R), and TFIIS (B175L, M163L). A mRNA capping enzyme (B148R, M133R) and a RNase III enzyme (B628R, M672L) that is presumably involved in the processing of viral mRNAs are also present in both viruses.

In the immediate-early phase of infection, the host is reprogrammed to transcribe viral RNAs. Very little is known as to how this occurs, but chromatin remodeling may be involved. PBCV-1 encodes a 119 amino acid SET domain containing protein (referred to as vSET) that di-methylates Lys27 in histone 3 (34). vSET is packaged in the PBCV-1 virion and accumulating evidence indicates that vSET could be involved in repression of host transcription following PBCV-1 infection (Manzur et al., unpublished results). Homologs to the PBCV-1 vSET are found in NY-2A and MT325 genomes. These proteins (B813L and M727L) are involved in the methylation of Lys27 of the histone H3 complex. Furthermore, the NY-2A genome encodes a second protein that contains a SET domain (B268L). In addition to these methylases, both genomes encode a SWF/SNF family helicase (B738L, M272L) and a SWI/SNF chromatin remodeling complex (B258R, M15L).

2.4.3 Protein Synthesis, Modification and Degradation

Chlorella viruses are the first viruses shown to encode a translation elongation factor. MT325 contains two ORFs which encode proteins homologous to the translation elongation factor from PBCV-1. The 901-amino-acid long MT325 translation elongation

factor (M742R) has 65% amino acid identity with its PBCV-1 homolog and is also quite similar to translation elongation factors from other organisms. In contrast, the smaller translation elongation factor (M375R-179 amino acids) encoded by MT325 is only similar to the N-terminus of its PBCV-1 homolog and is not significantly similar to any other proteins in the NR database. Interestingly, NY-2A does not encode a translation elongation factor.

The NY-2A and MT325 genomes contain ORFs that encode putative proteins similar to those involved in post-translational modification, prolyl-4-hydroxylase (B126R, M401R), or protein-degrading events, such as the ubiquitin C-terminal hydrolase (B150L, M137L). Both genomes also encode multiple AAA family ATPases, which are involved in the extraction and degradation of membrane proteins.

Finally, both viral genomes encode transfer RNAs. NY-2A encodes 7 tRNAs: 2 for Leu and 1 for each Arg, Asn, Lys, Tyr, and Val. The Tyr-tRNA gene is predicted to have a 12-nucleotide intron. The MT325 genome encodes 10 tRNAs: 2 for Asn and 1 for each Arg, Gly, Ile, Leu, Lys, Phe, Thr, and Tyr. Like the genes from the two NC64A viruses, the Tyr-tRNA MT325 gene is predicted to also contain an intron. Codon usage analyses of a few viral-encoded proteins indicate a strong correlation between the abundance of the virally-encoded tRNAs and their usage in viral protein.

2.4.4 Nucleotide Metabolism-Associated Proteins

The NY-2A and MT325 genomes encode several enzymes involved in nucleotide metabolism. These enzymes are important given that the DNA concentration in viral-infected cells increases rapidly following infection (35). Therefore, large quantities of

dNTPs must be synthesized for all of the DNA to be replicated. Both viral genomes encode the small and large subunits of ribonucleotide reductase, thioredoxin, glutaredoxin, dUTP pyrophosphatase (dUTPase), deoxycytidylate (dCMP) deaminase, and a thymidylate synthase.

PBCV-1 is the first known virus to encode a functional aspartate transcarbamylase, the key regulatory enzyme in the *de novo* biosynthesis of pyrimidines (36). NY-2A (B222R) also codes for such an enzyme which is 85% amino acid identical to its PBCV-1 homolog. MT325 does not encode such an enzyme.

NY-2A and MT325, like many large dsDNA viruses, encode both ribonucleotide reductase subunits. The NY-2A ribonucleotide reductase large subunit (B832R) contains an intein. This is the first description of a virally-encoded intein found among the Chlorella viruses.

The MT325 genome encodes two glutaredoxins (M241L, M423R) and three thioredoxins (M445L, M448L, and M449L), while the NY-2A genome encodes only one glutaredoxin (B592L) and one thioredoxin (B581L). These enzymes are likely important for viral DNA synthesis since large quantities of dNTPs (that cannot be accounted for simply by recycling dNTPs from host DNA) are needed following infection.

2.4.5 Protein Kinases, Phosphatases, and Channel Proteins

Both the NY-2A and the MT325 genomes encode similar protein kinases and a putative protein that resembles a dual-specificity phosphatase. MT325 (M143L, M221L, M543R, M729L, M733L and M794R) encodes six Ser/Thr protein kinases. Additionally, NY-2A also encodes six (B331R, B365L, B368L, B388L, B816L, and B818R) Ser/Thr

protein kinases. As compared with other viruses, this large number of viral-encoded proteins involved in phosphorylation/dephosphorylation is unusual given that this is a common mechanism for regulating many cellular processes. Therefore, it is likely that the viral-encoded proteins are involved in one or more signal transduction pathways that are important for viral replication.

The algal viruses are the first viruses known to encode K⁺ channel proteins (37). Like PBCV-1, NY-2A (B336R) and MT325 (M183R) encode a functional K⁺ channel protein. MT325 is unique because it also encodes another functional channel protein, an aquaglyceroporin (M30R).

2.4.6 Sugar- and Lipid-Manipulating Proteins

Surprisingly, the PBCV-1 genome contains 12 ORFs with high identity to enzymes involved in manipulating sugars, synthesizing polysaccharides, or metabolizing lipids. Several of these ORFs have homologs in NY-2A and MT325. Therefore, phycodnaviruses are unusual in that, of the three chlorella viruses sequenced, they all encode enzymes involved in sugar metabolism. Two NY-2A encoded enzymes, GDP-D-mannose dehydratase (B163R) and fucose synthase (B395L), comprise a three-step pathway that converts GDP-D-mannose to GDP-L-fucose. Fucose, a rare sugar, is present in the glycan attached to the major capsid protein. MT325 encodes three enzymes, glucosamine: fructose-6-phosphate amidotransferase (M37R), UDP-glucose dehydrogenase (M719L), and hyaluronan synthase (M128R), which are involved in the synthesis of hyaluronan, a linear polysaccharide composed of alternating β 1,4 glucuronic acid and β 1,3-N-acetylglucosamine residues (38, 39). NY-2A also encodes glucosamine

fructose-6-phosphate amidotransferase (B143R) and UDP-glucose dehydrogenase (B465R), however there is no homolog to hyaluronan synthase.

2.4.7 Cell Wall-Degrading Enzymes

NY-2A contains five ORFs which encode putative proteins that may be involved in degrading *Chlorella* cell walls. These proteins include a chitinase (B239R), a chitosanase (B393L), a β -1, 3-glucanase (B137L), and two β & α 1, 4 linked glucuronic lyases (B288L, B468R). MT325 encodes all of the cell wall-degrading enzymes that are present in the NY-2A genome (M85R, M91R, M124L, and M289R respectively). In addition, MT325 encodes an additional chitinase (M791R).

2.4.8 Restriction-Modification Enzymes

Chlorella viral genomes contain different levels of 5mC and 6mA (40). Therefore, it is not surprising that these viruses encode multiple 5mC and 6mA DNA methyltransferases. The genome of NY-2A is heavily methylated (45% 5mC and 37% 6mA), and, as a result, the genome encodes 18 DNA methyltransferases: 11 adenine methyltransferases, 7 cytosine methyltransferases. The MT325 genome encodes only two DNA methyltransferases, one of which is specific for methylating cytosines.

2.4.9 Integration and Transposition Enzymes

Even though NY-2A replicates with a lytic life cycle, there are numerous proteins that could be involved in manipulating its genome. There are 7 ORFs which encode proteins classified as putative transposases (B80L, B334R, B378L, B522L, B702L,

B711L, and B829R) based on significant homology to proteins in the NR, Pfam, and COG databases. Surprisingly, 3 of these transposases have internal resolvases (B83L, B381L, and B715L). In addition to the transposases and resolvases, NY-2A contains 34 ORFs with motifs found in homing endonucleases. Homing endonucleases are rare DNA-cleaving enzymes typically encoded by introns and inteins. These endonucleases are classified into four families (41). Sixteen of the NY-2A ORFs are members of the GIY-YIG family, and eighteen are members of the HNH family. It is unknown if the putative NY-2A-encoded transposases, resolvaes, and homing endonucleases are functional and/or if they play an essential role in the virus life cycle. The NY-2A virus encodes several proteins that could facilitate DNA rearrangements either within or between viruses, and possibly with its host. Unlike NY-2A, MT325 does not encode a resolvase. However, MT325 encodes a transposase (M844R). Furthermore, MT325 does encode 14 homing endonucleases; ten are members of the GIY-YIG family and four are members of the HNH family.

2.4.10 Miscellaneous Proteins

The NY-2A and MT325 genomes are predicted to encode several additional enzymes, including the polyamine-metabolizing enzymes ornithine decarboxylase (ODC) (B278R, M307L), N-carbamoylputrescine amidohydrolase (B116R, M103L), and homospermidine synthase (B305R, M233L). ODC catalyzes the key regulatory step in polyamine biosynthesis, the decarboxylation of ornithine to putrescine (42). The PBCV-1 ODC is the smallest ODC characterized to date (372 codons) (43). Interestingly, this enzyme decarboxylates arginine more efficiently than it decarboxylates ornithine. NY-

2A and MT325 both encode a 372 codon ODC, with 86 and 63% amino acid identity, respectively, with the PBCV-1 homolog. It will be interesting to examine whether these enzymes are bi-functional like its PBCV-1 homolog. The homospermidine synthase enzyme synthesizes homospermidine from two molecules of putrescine (44). The NY-2A homolog is 94% amino acid identical with the functional PBCV-1-encoded homospermidine synthase, whereas the MT325 homolog is only 69% identical. Homospermidine, a rare polyamine, is not the major polyamine in the virions. The significance of a virally-encoded pathway that synthesizes homospermidine is unknown.

MT325 (M99R) also encodes a putative protein with 76% amino acid identity with the PBCV-1-encoded Cu/Zn superoxide dismutase. This enzyme functions to convert superoxide radical anions into molecular oxygen and hydrogen peroxide (45). Presumably, the MT325-encoded enzyme reduces superoxide accumulation during sunlight.

The genome of NY-2A, a virus which infects *Chlorella* NC64A, and the genome of MT325, a virus which infects *Chlorella* Pbi, are the largest and smallest chlorella viruses sequenced to date. The annotation of these algal virus genomes led to the identification of new genes whose functions are scientifically interesting (e.g., aquaglyceroporin) or of possibly health value (e.g., DNA topoisomerase II). Additionally, the information gained from these sequenced genomes provides insight into the evolutionary history of algal viruses.

2.5 REFERENCES

- (1) Van Etten, J. L., Lane, L. C., and Meints, R. H. (1991) Viruses and viruslike particles of eukaryotic algae. *Microbiol Rev* 55, 586-620.
- (2) Schuster, A. M., Burbank, D. E., Meister, B., Skrdla, M. P., Meints, R. H., Hattman, S., Swinton, D., and Van Etten, J. L. (1986) Characterization of viruses infecting a eukaryotic Chlorella-like green alga. *Virology* 150, 170-7.
- (3) Reisser, W., Becker, B., and Klein, T. (1986) Studies on ultrastructure and host range of a chlorella attacking virus. *Protoplasma* 135, 162.
- (4) Reisser, W., Burbank, D. E., Meints, S. M., Meints, R. H., Becker, B., and Van Etten, J. L. (1988) A comparison of viruses infecting two different Chlorella-like green algae. *Virology* 167, 143-9.
- (5) Reisser, W., Burbank, D. E., Meints, R. H., Becker, B., and Van Etten James, L. (1991) Viruses distinguish symbiotic *Chlorella* spp. of *Paramecium bursaria*. *Endocytobiosis Cell Research* 7, 245-257.
- (6) Van Etten, J. L., Burbank, D. E., Xia, Y., and Meints, R. H. (1983) Growth cycle of a virus, PBCV-1, that infects Chlorella-like algae. *Virology* 126, 117-25.
- (7) Van Etten, J. L., Meints, R. H., Burbank, D. E., KuczmarSKI, D., Cuppels, D. A., and Lane, L. C. (1981) Isolation and characterization of a virus from the intracellular green alga symbiotic with *Hydra viridis*. *Virology* 113, 704-11.
- (8) Lowe, T. M., and Eddy, S. R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25, 955-64.
- (9) Schuster, A. M., Graves, M., Korth, K., Ziegelbein, M., Brumbaugh, J., Grone, D., and Meints, R. H. (1990) Transcription and sequence studies of a 4.3-kbp fragment from a ds-DNA eukaryotic algal virus. *Virology* 176, 515-23.
- (10) Group, G. C. (2002), Accelrys, San Diego, CA.

- (11) Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J Mol Biol* 215, 403-10.
- (12) Finn, R. D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, E. L., and Bateman, A. (2006) Pfam: clans, web tools and services. *Nucleic Acids Res* 34, D247-51.
- (13) Tatusov, R. L., Koonin, E. V., and Lipman, D. J. (1997) A genomic perspective on protein families. *Science* 278, 631-7.
- (14) Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., and Natale, D. A. (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4, 41.
- (15) Perler, F. B. (2002) InBase: the Intein Database. *Nucleic Acids Res* 30, 383-4.
- (16) Grabherr, R., Strasser, P., and Van Etten, J. L. (1992) The DNA polymerase gene from chlorella viruses PBCV-1 and NY-2A contains an intron with nuclear splicing sequences. *Virology* 188, 721.
- (17) Sun, L., Li, Y., McCullough, A. K., Wood, T. G., Lloyd, R. S., Adams, B., Gurnon, J. R., and Van Etten, J. L. (2000) Intron conservation in a UV-specific DNA repair gene encoded by chlorella viruses. *Journal of Molecular Evolution* 50, 82-92.
- (18) Li, Y., Lu, Z., Burbank, D. E., Kutish, G. F., Rock, D. L., and Van Etten, J. L. (1995) Analysis of 43 kb of the Chlorella virus PBCV-1 330-kb genome: map positions 45 to 88. *Virology* 212, 134-50.
- (19) Strasser, P., Zhang, Y., Rohozinski, J., and Van Etten, J. L. (1991) The termini of the chlorella virus PBCV-1 genome are identical 2.2-kbp inverted repeats. *Virology* 180, 763-9.
- (20) Grabherr, R., Strasser, P., and Van Etten, J. L. (1992) The DNA polymerase gene from Chlorella viruses PBCV-1 and NY-2A contains an intron with nuclear splicing sequences. *Virology* 188, 721-31.

- (21) Strasser, P., Zhang, Y. P., Rohozinski, J., and Van Etten, J. L. (1991) The termini of the chlorella virus PBCV-1 genome are identical 2.2-kbp inverted repeats. *Virology 180*, 763-9.
- (22) Yamada, T., and Higashiyama, T. (1993) Characterization of the terminal inverted repeats and their neighboring tandem repeats in the chlorella CVK1 virus genome. *Mol. Gen. Genet. 241*, 554.
- (23) Cheng, C., and Shuman, S. (1997) Characterization of an ATP-dependent DNA ligase encoded by *Haemophilus influenzae*. *Nucleic Acids Res. 25*, 1369.
- (24) Lepetit, D., Thiebaud, P., Aoufouchi, S., Prigent, C., Guesne, R., and Theze, N. (1996) The cloning and characterization of a cDNA encoding *Xenopus laevis* DNA ligase I. *Gene 172*, 273.
- (25) Ho, C. K., Van Etten, J. L., and Shuman, S. (1997) Characterization of an ATP-dependent DNA ligase encoded by chlorella virus PBCV-1. *J. Virology 71*, 1931.
- (26) Sriskanda, V., and Shuman, S. (1998) Specificity and fidelity of strand joining by Chlorella virus DNA ligase. *Nucleic Acids Research 26*, 3536-3541.
- (27) Sriskanda, V., and Shuman, S. (1998) Chlorella virus DNA ligase: nick recognition and mutational analysis. *Nucleic Acids Research 26*, 525-531.
- (28) Sriskanda, V., and Shuman, S. (1998) Mutational analysis of chlorella virus DNA ligase: catalytic roles of domain I and motif VI. *Nucleic Acids Res. 26*, 4618.
- (29) Roca, J. (1995) The mechanisms of DNA topoisomerases. *Trends Biochem. Sci. 20*, 156.
- (30) Lavrukhin, O. V., Fortune, J. M., Wood, T. G., Burbank, D. E., Van Etten, J. L., Osheroff, N., and Lloyd, R. S. (2000) Topoisomerase II from Chlorella virus PBCV-1: Characterization of the smallest known type II topoisomerase. *Journal of Biological Chemistry 275*, 6915-6921.
- (31) Fortune, J. M., Lavrukhin, O. V., Gurnon, J. R., Van Etten, J. L., Lloyd, R. S., and Osheroff, N. (2001) Topoisomerase II from Chlorella virus PBCV-1 has an

- exceptionally high DNA cleavage activity. *Journal of Biological Chemistry* 276, 24401-24408.
- (32) Dickey, J. S., Choi, T.-J., Van Etten, J. L., and Osheroff, N. (2005) Chlorella Virus Marburg Topoisomerase II: High DNA Cleavage Activity as a Characteristic of Chlorella Virus Type II Enzymes. *Biochemistry* 44, 3899-3908.
- (33) Warbrick, E. (2000) The puzzle of PCNA's many partners. *BioEssays* 22, 997.
- (34) Manzur, K. L., Farooq, A., Zeng, L., Plotnikova, O., Koch, A. W., Sachchidanand, and Zhou, M.-M. (2003) A dimeric viral SET domain methyltransferase specific to Lys27 of histone H3. *Nature Structural Biology* 10, 187-196.
- (35) Van Etten, J. L., Burbank, D. E., Joshi, J., and Meints, R. H. (1984) DNA synthesis in a Chlorella-like alga following infection with the virus PBCV-1. *Virology* 134, 443-9.
- (36) Landstein, D., Mincberg, M., Arad, S., and Tal, J. (1996) An early gene of the Chlorella virus PBCV-1 encodes a functional aspartate transcarbamylase. *Virology* 221, 151-158.
- (37) Kang, M., Moroni, A., Gazzarrini, S., DiFrancesco, D., Thiel, G., Severino, M., and Van Etten, J. L. (2004) Small potassium ion channel proteins encoded by chlorella viruses. *Proceedings of the National Academy of Sciences of the United States of America* 101, 5318-5324.
- (38) DeAngelis, P. L., Wei, J., Graves, M. V., Burbank, D. E., and Van Etten, J. L. (1997) Hyaluronan synthase of chlorella virus PBCV-1. *Science (Washington, D. C.)* 278, 1800-1803.
- (39) Landstein, D., Graves, M. V., Burbank, D. E., DeAngelis, P., and Van Etten, J. L. (1998) Chlorella virus PBCV-1 encodes functional glutamine:fructose-6-phosphate amidotransferase and UDP-glucose dehydrogenase enzymes. *Virology* 250, 388-396.
- (40) Van Etten, J. L., Lane, L. C., and Meints, R. H. (1991) Viruses and viruslike particles of eukaryotic algae. *Microbiological Reviews* 55, 586-620.

- (41) Belfort, M., and Roberts, R. J. (1997) Homing endonucleases: keeping the house in order. *Nucleic Acids Res.* 25, 3379.
- (42) Davis, R. H., Morris, D. R., and Coffino, P. (1992) Sequestered end products and enzyme regulation: the case of ornithine decarboxylase. *Microbiol. Rev.* 56, 280.
- (43) Morehead Tiara, A., Gurnon James, R., Adams, B., Nickerson Kenneth, W., Fitzgerald Lisa, A., and Van Etten James, L. (2002) Ornithine decarboxylase encoded by chlorella virus PBCV-1. *Virology* 301, 165-75.
- (44) Kaiser, A., Vollmert, M., Tholl, D., Graves, M. V., Gurnon, J. R., Xing, W., Liseck, A. D., Nickerson, K. W., and Van Etten, J. L. (1999) Chlorella virus PBCV-1 encodes a functional homospermidine synthase. *Virology* 263, 254-262.
- (45) Bannister, J. V., Bannister, W. H., and Rotilio, G. (1987) Aspects of the structure, function, and applications of superoxide dismutase. *CRC Rev. Biochem.* 22, 111.

2.6 ADDENDUM

2.6.1 Termini of the PBCV-1, NY-2A, and MT325 Genomes

Gap alignment of the 5' termini and the inverted 3' termini of each genome.

2.6.1.1 *Termini of the PBCV-1 genome*

```

Percent Similarity: 99.775    Percent Identity: 99.775
Match display thresholds for the alignment(s):
| = IDENTITY
: = 5
. = 1

5'PBCV-1 terminus x 3'PBCV-1 inverted terminus

1 GGGAGAACAGGTGGGATTGACAGTGGTAAATGTGTTGACCACGAGTAAA 50
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
1 CCCAGAACAGGTGGGATTGACAGTGGTAAATGTGTTGACCACGAGTAAA 50

51 AACAGGGCCCGGAAGCGGGCTATATAGAAGAGCGCAAGAAGAACACATA 100
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
51 AACAGGGCCCGGAAGCGGGCTATATAGAAGAGCGCAAGAAGAACACATA 100

101 AGGAGAGTTATTTGATTGGGCAAATCGCTGGCAAAATTGGCAAAATTTC 150
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
101 AGGAGAGTTATTTGATTGGGCAAATCGCTGGCAAAATTGGCAAAATTTC 150

151 TTCAATAGTTTTGTATCAAATAGCGAAAATATTTTTTCTCAAAAAGT 200
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
151 TTCAATAGTTTTGTATCAAATAGCGAAAATATTTTTTCTCAAAAAGT 200

201 TTTTGACTGGTTAGCGTAAACTATTTAGTTCTCATTTATGAGTTTA 250
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
201 TTTTGACTGGTTAGCGTAAACTATTTAGTTCTCATTTATGAGTTTA 250

251 TCGGAGTTGGTAATAAATCTCACAAAACCTCTAAGGACAAACTCTGGCAGA 300
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
251 TCGGAGTTGGTAATAAATCTCACAAAACCTCTAAGGACAAACTCTGGCAGA 300

301 AATCCTAACGGAGAACATTAAGAGTTCGTGGCTCTGGCTCGTATCGA 350
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
301 AATCCTAACGGAGAACATTAAGAGTTCGTGGCTCTGGCTCGTATCGA 350

351 CAAACGAGGTCTTCGGTTGTATATGTTCTCTGGCAATCACTAACGGCCGC 400
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
351 CAAACGAGGTCTTCGGTTGTATATGTTCTCTGGCAATCACTAACGGCCGC 400

401 CCGGTGTCATCTGCCGCATCACGAAGTACCCCCGAAACCTCAGGTGGTAT 450
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
401 CCGGTGTCATCTGCCGCATCACGAAGTACCCCCGAAACCTCAGGTGGTAT 450

451 GTAAACAGAGTATACAAGCTAACGGACAATTCTCGTACAGAAATACTAA 500
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
451 GTAAACAGAGTATACAAGCTAACGGACAATTCTCGTACAGAAATACTAA 500

501 GGACATAATTGTTATCTATTCTTTAAGTCCGTATTCTGCTTGTCATA 550
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
501 GGACATAATTGTTATCTATTCTTTAAGTCCGTATTCTGCTTGTCATA 550

551 ATTTATTGGTATAATTTCAGTATTCCGATACTCTTCGGCAAAACTTTCTCA 600
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
551 ATTTATTGGTATAATTTCAGTATTCCGATACTCTTCGGCAAAACTTTCTCA 600

601 AATCATCTAACGAAATTCTTAGTATTCTAAATGACATATATAATTG 650
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
601 AATCATCTAACGAAATTCTTAGTATTCTAAATGACATATATAATTG 650

651 TTTGGTCCTCTCTTACTTCATAACCTCTCAATCCTTTGGTGTAT 700
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|

```

651 TTTGGTCCTCCTTACTTCATAACCTCTCAATCCTTTGGTGTAT 700
 701 AGGAAAAATGGTAGGGCGAAAACGATCTGGTAAGTGTGATTATTCAACA 750
 701 AGGAAAAATGGTAGGGCGAAAACGATCTGGTAAGTGTGATTATTCAACA 750
 751 ACATATTACTATCTTATCCCCTCGTCAATTATAATACATATTATCAGT 800
 751 ACATATTACTATCTTATCCCCTCGTCAATTATAATACATATTATCAGT 800
 801 TTCACTCTACATCCTTGCTTCTGTCTGGTTTATATTCTCTAAATT 850
 801 TTCACTCTACATCCTTGCTTCTGTCTGGTTTATATTCTCTAAATT 850
 851 CAACAAATCAAATTCCCAATCGGATCTCCTCTTAGAAGAACCCCC 900
 851 CAACAAATCAAATTCCCAATCGGATCTCCTCTTAGAAGAACCCCC 900
 901 AAATCAACGCCGTTATATTAGAACCTAATTATTACCTAATCTCGCGTGG 950
 901 AAATCAACGCCGTTATATTAGAACCTAATTATTACCTAATCTCGCGTGG 950
 951 TTAAACATCTACGGTATGTCTGTAGTATTGTCATTGTGTGCTACCTAT 1000
 951 TTAAACATCTACGGTATGTCTGTAGTATTGTCATTGTGTGCTACCTAT 1000
 1001 ATAGTAGTATCCTGTATCTCTATGAACCTCTCATATATACAGAGAGCCG 1050
 1001 ATAGTAGTATCCTGTATCTCTATGAACCTCTCATATATACAGAGAGCCG 1050
 1051 TTGTGTCACTCATAATTAGGTAATCTCAAAGACATAGGACAAAATA 1100
 1051 TTGTGTCACTCATAATTAGGTAATCTCAAAGACATAGGACAAAATA 1100
 1101 CTCAAATGGTGTGATATGCTGAAATAATCTCGTATCGACAAACATCATCG 1150
 1101 CTCAAATGGTGTGATATGCTGAAATAATCTCGTATCGACAAACATCATCG 1150
 1151 TTGGTGATAATCTCACAAACTCTAACGGACAAACTCTGGCAGAAATCCT 1200
 1151 TTGGTGATAATCTCACAAACTCTAACGGACAAACTCTGGCAGAAATCCT 1200
 1201 AAGGAGAACATTAAGAGTTCTGTGGTCTCTGGTCTGTATCGACAAACG 1250
 1201 AAGGAGAACATTAAGAGTTCTGTGGTCTCTGGTCTGTATCGACAAACG 1250
 1251 GGGTCGTCGGTTGTATACGAAACTAACGGACACACAGGAATTAAATC 1300
 1251 GGGTCGTCGGTTGTATACGAAACTAACGGACACACAGGAATTAAATC 1300
 1301 AAGACATTATTAAGCACAGCACGACAGCACGACGATGTTGTTAAACAC 1350
 1301 AAGACATTATTAAGCACAGCACGACAGCACGACGATGTTGTTAAACAC 1350
 1351 ACTCACACCGAACATTATGCCGCAGCGCTACCTTGGGCTCCGCTCCTG 1400
 1351 ACTCACACCGAACATTATGCCGCAGCGCTACCTTGGGCTCCGCTCCTG 1400
 1401 GTCGAACATGGGCTCCGCTCCGCTAAATCGCTACACTCCGCTA 1450
 1401 GTCGAACATGGGCTCCGCTCCGCTAAATCGCTACACTCCGCTA 1450
 1451 CGCTACGGGCTTGTCACGCCGCCCTACGCTCCGCTACGCTTCGCTCA 1500
 1451 CGCTACGGGCTTGTCACGCCGCCCTACGCTCCGCTACGCTTCGCTCA 1500
 1501 GCCAGTGAATTGAGACTACCTGTGGTGTAGTTATATAATGGTAAATAC 1550
 1501 GCCAGTGAATTGAGACTACCTGTGGTGTAGTTATATAATGGTAAATAC 1550
 1551 CTATATATACAGGTCTATTACATATAATGCGCTGGACTTTTAGTTGT 1600
 1551 CTATATATACAGGTCTATTACATATAATGCGCTGGACTTTTAGTTGT 1600
 1601 ATACTACTTTGTATAAAACCTATAAGTTACATGTGTTTACATATAATG 1650
 1601 ATACTACTTTGTATAAAACCTATAAGTTACATGTGTTTACATATAATG 1650
 1651 CGCTTGCAACTTTTACATCATCAACATCTTGAAACATATAATGCGCT 1700

```

1651 ||||||||| 1700
1651 CGCTTGCAACTTTATACATCATCACACATCTGAAAACATATAATGCGCT 1700
1701 TGCAAACTTGTATATCATTAAAGATATCAAAAACATATAATGCGCTTGCG 1750
1701 TGCAAACTTGTATATCATTAAAGATATCAAAAACATATAATGCGCTTGCG 1750
1751 ATTTTATATATCATCAACTTGAAGAGTGTACATATAATGCGCTTG 1800
1751 ATTTTATATATCATCAACTTGAAGAGTGTACATATAATGCGCTTG 1800
1801 CAAACTTGTATATCATTAAAGATGTCTGGTCAAATGACACTTTTATTT 1850
1801 CAAACTTGTATATCATTAAAGATGTCTGGTCAAATGACACTTTTATTT 1850
1851 CAAGTTTAAAGTTTACTTTTTACTTTCCATTCACTTGTCA 1900
1851 CAAGTTTAAAGTTTACTTTTTACTTTCCATTCACTTGTCA 1900
1901 ACGGGCTCAATAGTTCGGGGTGTTGACCACAAACACGGTGCCGACGCG 1950
1901 ACGGGCTCAATAGTTCGGGGTGTTGACCACAAACACGGTGCCGACGCG 1950
1951 GTACAGCTGTATGTTCGGTGAAGAACCTCTTAATCGACTGAATCGAAC 2000
1951 GTACAGCTGTATGTTCGGTGAAGAACCTCTTAATCGACTGAATCGAAC 2000
2001 TAGTCTCTGGGCCAGTGGGGGTCGATCGGCTCGTGGTCTCAACC 2050
2001 TAGTCTCTGGGCCAGTGGGGGTCGATCGGCTCGTGGTCTCAACC 2050
2051 ACGATGGCTTGTCTGGTGAAGAGGTTCTTGATGGCGAAATCATTGT 2100
2051 ACGATGGCTTGTCTGGTGAAGAGGTTCTTGATGGCGAAATCATTGT 2100
2101 GAATGCTCTGGTAAAGATACTCATCTACAACGGAACCGTATTATAC 2150
2101 GAATGCTCTGGTAAAGATACTCATCTACAACGGAACCGTATTATAC 2150
2151 CTACGGCGTCATTCATCGTCATTCATCGTCATTTAAATCGTC 2200
2151 CTACGGCGTCATTCATCGTCATTCATCGTCATTTAAATCGTC 2200
2201 ATTCGGGTCAAATGACAAA 2221
2201 ATTCGGGTCAAATGACAAA 2221

```

2.6.1.2 Termini of the NY-2A genome

Percent Similarity: 99.217 Percent Identity: 99.217

Match display thresholds for the alignment(s) :

	=	IDENTITY
:	=	5
.	=	1

5'NY-2A terminus x 3'NY-2A inverted terminus

```

1 .....TGACACCA 8
1 AATCTTATAATCGATTCGCCACACGACCATCTGGTCAAATGACACCA 50
9 CACAATTCACTGGCCACGAGTAGAACAGGGCTCGTGGTCGGAGTTATAT 58
51 CACAATTCACTGGCCACGAGTAGAACAGGGCTCGTGGTCGGAGTTATAT 100
59 ACAAGATCGTAATAAGAAATAACTAAGGAGAGTTATGTGATTGGGCAAAT 108
101 ACAAGATCGTAATAAGAAATAACTAAGGAGAGTTATGTGATTGGGCAAAT 150

```

109 CTC TGAC AAA ATT CCT CGA TAG TTT CT TCA AA AT A GCG GAG AAAA 158
 |||||
 151 CTC TGAC AAA ATT CCT CGA TAG TTT CT TCA AA AT A GCG GAG AAAA 200
 |||||
 159 AAA ACT ATT TTATA ATT TTG GTT AGCG TAA ACT ATT TTG TTCT 208
 |||||
 201 AAA ACT ATT TTATA ATT TTG GTT AGCG TAA ACT ATT TTG TTCT 250
 |||||
 209 ATT TTT GTG GTT GTG AGA GAT CGT AGGA AGT CT CA GAG AT TTCT AAGG 258
 |||||
 251 ATT TTT GTG GTT GTG AGA GAT CGT AGGA AGT CT CA GAG AT TTCT AAGG 300
 |||||
 259 ACA AA CT CTG ACATA AA ACTA AGG ACAGT AT AT GG CT TG AT CG AC AA 308
 |||||
 301 ACA AA CT CTG ACATA AA ACTA AGG ACAGT AT AT GG CT TG AT CG AC AA 350
 |||||
 309 ATT CT GTT GAT GG TAT AT GG CA AA CT AAGG TC GC CT CG CA CT GC 358
 |||||
 351 ATT CT GTT GAT GG TAT AT GG CA AA CT AAGG TC GC CT CG CA CT GC 400
 |||||
 359 TACT CGC GAAG CAC ACCT GAG ACT TCT GT GG TAT GTACA AGC TCG CCT 408
 |||||
 401 TACT CGC GAAG CAC ACCT GAG ACT TCT GT GG TAT GTACA AGC TCG CCT 450
 |||||
 409 CGT AT CGAC AA ATT CT GTT GTG AT A CGT CA AA CT AAGG TCG C 458
 |||||
 451 CGT AT CGAC AA ATT CT GTT GTG AT A CGT CA AA CT AAGG TCG C 500
 |||||
 459 CTC GC AT CTG CT ACT CGC GAAG CAC ACCT GAG ACT TCT GT GG TAT GT A 508
 |||||
 501 CTC GC AT CTG CT ACT CGC GAAG CAC ACCT GAG ACT TCT GT GG TAT GT A 550
 |||||
 509 CAAG CT CGAG GT CA ATT CGT CGA TA CA AC AT AT GAC CTT TT GG TAT CT G 558
 |||||
 551 CAAG CT CGAG GT CA ATT CGT CGA TA CA AC AT AT GAC CTT TT GG TAT CT G 600
 |||||
 559 TT CT GA AT CTA AT CT GG TT CAG A ACT GAG TCA TATT TT ACAC GG AG TA 608
 |||||
 601 TT CT GA AT CTA AT CT GG TT CAG A ACT GAG TCA TATT TT ACAC GG AG TA 650
 |||||
 609 TAT GG TCT TG TAT CGAC AA ATT CCT GG TT GAT GG TGC AT A TACT AT AT GG 658
 |||||
 651 TAT GG TCT TG TAT CGAC AA ATT CCT GG TT GAT GG TGC AT A TACT AT AT GG 700
 |||||
 659 TACA AAC ACA AGG ACAG C CT CTG AC AT GT TT GT AT AA ACT GAC AC AG CA 708
 |||||
 701 TACA AAC ACA AGG ACAG C CT CTG AC AT GT TT GT AT AA ACT GAC AC AG CA 750
 |||||
 709 CGC ACA AC AC AT AGC AC CG CG CAG C AC GG CCT AT TT GG CG GT CAT CT G CT A 758
 |||||
 751 CGC ACA AC AC AT AGC AC CG CG CAG C AC GG CCT AT TT GG CG GT CAT CT G CT A 800
 |||||
 759 CTT CGC GAAG CAC ACCT GAG ACT TCT AT GT GG TAT AT AA AGT TAT AT AC AA 808
 |||||
 801 CTT CGC GAAG CAC ACCT GAG ACT TCT AT GT GG TAT AT AA AGT TAT AT AC AA 850
 |||||
 809 GCT CGAG GT CA ATT CGT CGA TATA AC AT AT GAC CTT TT GG TAT CT GT TC 858
 |||||
 851 GCT CGAG GT CA ATT CGT CGA TATA AC AT AT GAC CTT TT GG TAT CT GT TC 900
 |||||
 859 TGA AT CTA AT CT GG TT CAG A AC AG AGT CA TATT TT ACAC GG AGT GAT C 908
 |||||
 901 TGA AT CTA AT CT GG TT CAG A AC AG AGT CA TATT TT ACAC GG AGT GAT C 950
 |||||
 909 AGT CT CG TAT CGAC AA ATT CCT GG TT GAT GG TGC AT A TACT AT AT GG TAC 958
 |||||
 951 AGT CT CG TAT CGAC AA ATT CCT GG TT GAT GG TGC AT A TACT AT AT GG TAC 1000
 |||||
 959 AAAC ACA AGG ACAG C CT CTG AC AT GT TT GT AT AA ACT GAC AC AG C AC GC 1008
 |||||
 1001 AAAC ACA AGG ACAG C CT CTG AC AT GT TT GT AT AA ACT GAC AC AG C AC GC 1050
 |||||
 1009 ACA AC AC AC AGC AC CG CAC AA AC AT AGC AC CG CG CAG C AC GG CCT ATT TT GG 1058
 |||||
 1051 ACA AC AC AC AGC AC CG CAC AA AC AT AGC AC CG CG CAG C AC GG CCT ATT TT GG 1100
 |||||
 1059 CGGT CAT CTG CT ACG GAA CG CT AT CTG AGC CCT T CAT GT GG TAT GT AC AA 1108
 |||||
 1101 CGGT CAT CTG CT ACG GAA CG CT AT CTG AGC CCT T CAT GT GG TAT GT AC AA 1150

1109 GCTCGAGGTCAATT CGTCGATAACACATATGACTCTGTTT CGAACATAT 1158
 |||||
 1151 GCTCGAGGTCAATT CGTCGATAACACATATGACTCTGTTT CGAACATAT 1200
 |||||
 1159 TCTGGTT CAGAACAGAGTCATATTTTACACGGTCTGATCAGTCTCGTA 1208
 |||||
 1201 TCTGGTT CAGAACAGAGTCATATTTTACACGGTCTGATCAGTCTCGTA 1250
 |||||
 1209 TCGACAAATTATT CGAAGTACATGTAATTCTGTTGAAACACAAGG 1258
 |||||
 1251 TCGACAAATTATT CGAAGTACATGTAATTCTGTTGAAACACAAGG 1300
 |||||
 1259 ACAGACTCTCATCATAGACAAATTATGTCCTCATGGTCTTGTTG 1308
 |||||
 1301 ACAGACTCTCATCATAGACAAATTATGTCCTCATGGTCTTGTTG 1350
 |||||
 1309 AGACATATAGCGAGCATAACACGCACAGCATGGCAATGGAGGGTTAAAA 1358
 |||||
 1351 AGACATATAGCGAGCATAACACGCACAGCATGGCAATGGAGGGTTAAAA 1400
 |||||
 1359 TGCACATGT CACAAAGTTCACTCATGACGAACATCATGCCG CAGGCGCTG 1408
 |||||
 1401 TGCACATGT CACAAAGTTCACTCATGACGAACATCATGCCG CAGGCGCTG 1450
 |||||
 1409 TCCTTGGCGCTGCTCCATACGCGT GTT CGACCGTGTGATTT CGTCGT 1458
 |||||
 1451 TCCTTGGCGCTGCTCCATACGCGT GTT CGACCGTGTGATTT CGTCGT 1500
 |||||
 1459 ACACAAAGTTCACTCATGACGAACATCATGCCG CAGGCGCTGTCCTGGG 1508
 |||||
 1501 ACACAAAGTTCACTCATGACGAACATCATGCCG CAGGCGCTGTCCTGGG 1550
 |||||
 1509 CGCTGCTCCATACGCGT GTT CGACCGTGTGATTT CGTCGTACACCGAG 1558
 |||||
 1551 CGCTGCTCCATACGCGT GTT CGACCGTGTGATTT CGTCGTACACCGAG 1600
 |||||
 1559 ATATCTCGGGATTCTCGGGATTCTATGAT..... 1587
 |||||
 1601 ATATCTCGGGATTCTCGGGATTCTATGATTT CGAACGCTGTTTACCTAT 1650
 |||||
 1588 TTTCGAAGCTGTTTACCTATAATGAGATTGCACTTTGGTGATCTACT 1637
 |||||
 1701 TTTCGAAGCTGTTTACCTATAATGAGATTGCACTTTGGTGATCTACT 1750
 |||||
 1638 AGATTCCGAACAAAGTATCAGTT CGAACGCTGTTTACCTATAATGAGAT 1687
 |||||
 1751 AGATTCCGAACAAAGTATCAGTT CGAACGCTGTTTACCTATAATGAGAT 1800
 |||||
 1688 TGCAC TTTTGGTGATCTACTAGATT CGAACAAAGTATCAGTT CGAAG 1737
 |||||
 1801 TGCAC TTTTGGTGATCTACTAGATT CGAACAAACTATCAGTT CGAAG 1850
 |||||
 1738 CTGTTTACCTATAATGAGATTGCACTTTGGTGATCTACTAGATTCCG 1787
 |||||
 1851 CTGTTTACCTATAATGAGATTGCACTTTGGTGATCTACTAGATTCCG 1900
 |||||
 1788 AACAAAGTATCAGTT CGAACGCTGTTTACCTATAATGAGATTGCACTT 1837
 |||||
 1901 AACAAAGTATCAGTT CGAACGCTGTTTACCTATAATGAGATTGCACTT 1950
 |||||
 1838 TTGGTGATCTACTAGATT CGAACAAAGTATCAGTT CGAACGCTGTTTA 1887
 |||||
 1951 TTGGTGATCTACTAGATT CGAACAAAGTATCAGTT CGAACGCTGTTTA 2000
 |||||
 1888 CCTATAATGAGATTGCACTTTGGTGATCATTAGATGTCCGGGTCAA 1937
 |||||
 2001 CCTATAATGAGATTGCACTTTGGTGATCATTAGATGTCCGGGTCAA 2050
 |||||
 1938 TGACACTGTTACTT CGAACATT TAAAAGTTTCAATGTCTCGTTAAC 1987
 |||||
 2051 TGACACTGTTACTT CGAACATT TAAAAGTTTCAATGTCTCGTTAAC 2100
 |||||
 1988 GATTAGTCAACGATCTCAATAGTT CGGGGGT GTTGACCACAAACACGGT 2037
 |||||
 2101 GATTAGTCAACGATCTCAATAGTT CGGGGGT GTTGACCACAAACACGGT 2150

```

2038 GCCAACCGGGTACAAACCCACGTGCTCGTGAAGAAACGCTTGACAGACT 2087
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2151 GCCAACCGGGTACAAACCCACGTGCTCGTGAAGAAACGCTTGACAGACT 2200
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2088 CGAAGACCGATCCGTTGACGGACCCGCTGCAACACACTCGATAGGCTCG 2137
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2201 CGAAGACCGATCCGTTGACGGACCCGCTGCAACACACTCGATAGGCTCG 2250
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2138 ACGGTATCGACCAGAACGGCTTGTGCTGGTGAAGAGATTCTTGATGGA 2187
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2251 ACGGTATCGACCAGAACGGCTTGTGCTGGTGAAGAGATTCTTGATGGC 2300
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2188 AGACATCATGGTAATGCTACTATATGTGATACAGCTGACGCGTGTGACA 2237
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2301 AGACATCATGGTAACGCTACTATATGTGATACAGATGACGCATGTGACA 2350
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2238 TATATACACACACAGAACAGTGTCAATTTCAGGATCAAATGACTAACTGT 2287
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2351 TATATACACACACAGAACAGTGTCAATTTCAGGATCAAATGACAATTCCA 2400
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2288 TTAAAAAAACATC 2300
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2401 GGATCAAATGACT 2413

```

2.6.1.3 Termini of the MT325 genome

```

Percent Similarity: 100.000    Percent Identity: 100.000

Match display thresholds for the alignment(s):
| = IDENTITY
: = 5
. = 1

5' MT325 terminus x 3' MT325 inverted terminus

1 GTCCTATGGGCTGCGCTCCCTGGTTGACGCTCCGCTGCGCTGCGC 50
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1785 GTCCTATGGGCTGCGCTCCCTGGTTGACGCTCCGCTGCGCTGCGC 1834
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 TCGACCAGCAAATGATTGGTGTATGTGTTATACATATTGGCAATC 100
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1835 TCGACCAGCAAATGATTGGTGTATGTGTTATACATATTGGCAATC 1884
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 TGTACATAATACTTGCAAATTCATATAAAAGTACTCTTAAATATAAT 150
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1885 TGTACATAATACTTGCAAATTCATATAAAAGTACTCTTAAATATAAT 1934
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 ACGCGGGGCATATACTTATAAAGAAAAGGACCTTTAATATAATAATGTA 200
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1935 ACGCGGGGCATATACTTATAAAGAAAAGGACCTTTAATATAATAATGTA 1984
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
201 CTTCAAATATAATAACGGGGCTCATTTTAATATTAAAAGACCTTTAAT 250
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1985 CTTCAAATATAATAACGGGGCTCATTTTAATATTAAAAGACCTTTAAT 2034
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251 TGTAATTACACCTTAAACATTACGGGGTCAAAATTAAATACGGG 300
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2035 TGTAATTACACCTTAAACATTACGGGGTCAAAATTAAATACGGG 2084
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 GGTCAATTTCACACTAAAAGAGCTTTAATGTGAATTCAAATACGCG 350
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2085 GGTCAATTTCACACTAAAAGAGCTTTAATGTGAATTCAAATACGCG 2134
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
351 GGGCAAAGAATCATAAGGAAAAGAAATTAAATTAGGGAGTAGGGGTACAAA 400
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2135 GGGCAAAGAATCATAAGGAAAAGAAATTAAATTAGGGAGTAGGGGTACAAA 2184
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
401 GCGGGATGCAACTTCCGCGTTGAGAGTCGCGTGTGCAAGGCAGTCAGT 450
||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2185 GCGGGATGCAACTTCCGCGTTGAGAGTCGCGTGTGCAAGGCAGTCAGT 2234

```

451 TTTGCTTGGAACTTCCCAGTCTCACGGCAGACTCTGTTCTACGAAGCC 500
 |||||
 2235 TTTGCTTGGAACTTCCCAGTCTCACGGCAGACTCTGTTCTACGAAGCC 2284
 |||||
 501 CAGTCCTTTGGCCAACTCGTTGTACGTGGCTAGTCTTCGAGAAC 550
 |||||
 2285 CAGTCCTTTGGCCAACTCGTTGTACGTGGCTAGTCTTCGAGAAC 2334
 |||||
 551 CTCTTCCCTGCGTCGAGTCGATCGGCATACTTTTTCTATGAGGTCA 600
 |||||
 2335 CTCTTCCCTGCGTCGAGTCGATCGGCATACTTTTTCTATGAGGTCA 2384
 |||||
 601 AGTTTTTGGCAAACCTCTTCTCTGTTCTTTCTTGCTCTGGAGGTC 650
 |||||
 2385 AGTTTTTGGCAAACCTCTTCTCTGTTCTTTCTTGCTCTGGAGGTC 2434
 |||||
 651 ATTGATCATGGCCCCGAGTACAGCGTTGCGACTGCAACTCTTGTGGG 700
 |||||
 2435 ATTGATCATGGCCCCGAGTACAGCGTTGCGACTGCAACTCTTGTGGG 2484
 .
 .
 .
 701 ACTCCGTGACGATGTTCAGTTGTCCTGGAACCTCCTGAGTCTACGGCA 750
 |||||
 2485 ACTCCGTGACGATGTTCAGTTGTCCTGGAACCTCCTGAGTCTACGGCA 2534
 |||||
 751 GACTCTGTTCTACGAAGCCAAGTGTGTTGGCAACTCTTGTACGTG 800
 |||||
 2535 GACTCTGTTCTACGAAGCCAAGTGTGTTGGCAACTCTTGTACGTG 2584
 |||||
 801 GCCTAGTCTTCGATAATCTCTTTCCCTGCGGTGAGTCTATCGGAG 850
 |||||
 2585 GCCTAGTCTTCGATAATCTCTTTCCCTGCGGTGAGTCTATCGGAG 2634
 |||||
 851 ACTCTGTTTACGACGCCAACGTGTTGGCAACTCGTTGTCTACGTGG 900
 |||||
 2635 ACTCTGTTTACGACGCCAACGTGTTGGCAACTCGTTGTCTACGTGG 2684
 .
 .
 .
 901 CCTAGTCTTCGATAATCTCTTTCCCTGCGGTGAGTCTATCGGCATA 950
 |||||
 2685 CCTAGTCTTCGATAATCTCTTTCCCTGCGGTGAGTCTATCGGCATA 2734
 .
 .
 .
 951 CTTTTTTCTATGAGGTCAAGTTTCCGCAAACCTCTCCTGTTTTT 1000
 |||||
 2735 CTTTTTTCTATGAGGTCAAGTTTCCGCAAACCTCTCCTGTTTTT 2784
 .
 .
 .
 1001 CTTTGCCTGAGTGGAGAGAACCACTTCAGCTCATTGATCTGATAC 1050
 |||||
 2785 CTTTGCCTGAGTGGAGAGAACCACTTCAGCTCATTGATCTGATAC 2834
 .
 .
 .
 1051 ATATATTTTCTCCAATTGACGTACTTCGCAATATGATTCTTTCT 1100
 |||||
 2835 ATATATTTTCTCCAATTGACGTACTTCGCAATATGATTCTTTCT 2884
 .
 .
 .
 1101 GTTTCTCGAGATTCTTAAGGTTTTACATAATGTTTCAATTAAACA 1150
 |||||
 2885 GTTTCTCGAGATTCTTAAGGTTTTACATAATGTTTCAATTAAACA 2934
 .
 .
 .
 1151 GTCCATGACTTGTCTTGATGTTCTCCCGCTCGCGCTCGGTCAAGGGCGC 1200
 |||||
 2935 GTCCATGACTTGTCTTGATGTTCTCCCGCTCGCGCTCGGTCAAGGGCGC 2984
 .
 .
 .
 1201 CATCCCCCTTGCTGCCTTGTCTGGCCGCCAGGGCTCCACAAATTAA 1250
 |||||
 2985 CATCCCCCTTGCTGCCTTGTCTGGCCGCCAGGGCTCCACAAATTAA 3034
 .
 .
 .
 1251 AAGCGTCCATCTCGTGGTCAGCGTGTACTGGTCTCGCCGGCGGGGGC 1300
 |||||
 3035 AAGCGTCCATCTCGTGGTCAGCGTGTACTGGTCTCGCCGGCGGGGGC 3084

1301 GCAGGTGCCTTGCGCCTTGGCTTGGCAGCCAAATTGCCCTATGCC 1350
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
 3085 GCAGGTGCCTTGCGCCTTGGCTTGGCAGCCAAATTGCCCTATGCC 3134
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
 1351 GGCAAAAGTAGCTCTGAACACCGAGACCTTCTCGGGCGCAATCACAGGCT 1400
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
 3135 GGCAAAAGTAGCTCTGAACACCGAGACCTTCTCGGGCGCAATCACAGGCT 3184
 ||||| ||||| ||||| ||||| ||||| |||||
 1401 CCTCAATGGCGCCGGCACCTCGACGTCAACGTGGTAATGCTTCTTACTC 1450
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
 3185 CCTCAATGGCGCCGGCACCTCGACGTCAACGTGGTAATGCTTCTTACTC 3234
 ||||| ||||| ||||| ||||| ||||| |||||
 1451 TGCTCGGGCATCTCGACCTTGGCATTGTACAGCGTCGGTGTGCTCGGAGAA 1500
 ||||| ||||| ||||| ||||| ||||| ||||| |||||
 3235 TGCTCGGGCATCTCGACCTTGGCATTGTACAGCGTCGGTGTGCTCGGAGAA 3284
 ||||| ||||| ||||| ||||| ||||| |||||
 1501 GAACTGCTTGATGGAAACCATTGTTGAGTGGGTGTTGTTGTTGTT 1550
 ||||| ||||| ||||| ||||| ||||| |||||
 3285 GAACTGCTTGATGGAAACCATTGTTGAGTGGGTGTTGTTGTTGTT 3334
 ||||| ||||| ||||| |||||
 1551 GATTGTGATTGTTGGTAGATGGTGGTGGTGGAGGAGATGTTCTTGT 1600
 ||||| ||||| ||||| ||||| ||||| |||||
 3335 GATTGTGATTGTTGGTAGATGGTGGTGGTGGAGGAGATGTTCTTGT 3384
 ||||| ||||| ||||| ||||| |||||
 1601 CACAGGTGTCCCTTTATACACAATTCTCGATATGGCATTACCAAGGA 1650
 ||||| ||||| ||||| ||||| ||||| |||||
 3385 CACAGGTGTCCCTTTATACACAATTCTCGATATGGCATTACCAAGGA 3434
 ||||| ||||| ||||| ||||| |||||
 1651 CAAACGACCGTAACCCGGGTGTTCTGGTCAAATGACACTCAAATTTT 1700
 ||||| ||||| ||||| ||||| ||||| |||||
 3435 CAAACGACCGTAACCCGGGTGTTCTGGTCAAATGACACTCAAATTTT 3484
 ||||| ||||| ||||| ||||| |||||
 1701 GCGTTTTAAAAAAGGATAAGGATACCCCTTAATCTTATCAAATGTT 1749
 ||||| ||||| ||||| ||||| ||||| |||||
 3485 GCGTTTTAAAAAAGGATAAGGATACCCCTTAATCTTATCAAATGTT 3533

2.6.2 NY-2A Genome

2.6.2.1 General Characteristics of Open Reading Frames

NY-2A ORF	ORF Location		% A+T (-50 ntts from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
B1L	1156	875	62%	4	45	70	59	69	84	282	93	10442	4.7	2860
b2R	1094	1318	46%	2	37	67	41	43	74	225	74	8632	7.5	8310
B3R	1474	1710	52%	1	44	52	50	55	80	237	78	8908	4.6	5240
B4R	1894	2202	68%	1	46	86	68	75	80	309	102	11765	6.9	5750
b5L	2238	1990	66%	5	51	55	66	61	67	249	82	8911	5	1310
B6L	2960	2382	84%	6	39	169	104	123	183	579	192	22527	4.3	31890
b7R	2490	2786	70%	3	39	97	60	55	85	297	98	11940	6.9	15280
B8R	3196	4191	82%	1	40	348	187	211	250	996	331	37020	7.2	9290
b9L	3809	3486	62%	6	34	90	62	49	123	324	107	12108	8.9	5240
B10R	4258	5169	88%	1	38	332	157	188	235	912	303	35231	8.4	64900
b11R	4401	4646	64%	3	41	78	45	55	68	246	81	9617	10	16500
b12L	4938	4660	62%	5	39	65	61	49	104	279	92	10957	5.8	7770
B13L	6175	5147	86%	4	36	346	149	219	315	1029	342	40120	6.8	41540
b14R	5639	5842	66%	2	36	66	41	32	65	204	67	8179	7.7	17930
b15R	5727	5948	66%	3	35	73	43	34	72	222	73	8111	11	5120
B16L	7343	6237	84%	6	37	360	180	234	333	1107	368	42158	6.7	32640
b17R	6534	6791	70%	3	36	82	52	41	83	258	85	9951	8.2	5810
B18L	8812	7394	86%	4	45	402	299	341	377	1419	472	52453	6.2	24540
b19R	7503	7760	62%	3	41	73	53	53	79	258	85	9542	12	2620
b20R	7886	8092	60%	2	45	59	46	47	55	207	68	8040	9.5	2620
b21R	8099	8407	62%	2	47	77	80	64	88	309	102	11883	5.1	2770
B22R	8915	10120	80%	2	44	404	260	267	275	1206	401	47049	5.7	76930
b23L	10098	9817	58%	5	41	59	57	59	107	282	93	10451	11	5240
B24L	11693	10095	72%	6	43	475	288	394	442	1599	532	58479	7.5	41130
b25R	11545	11775	60%	1	31	59	34	37	101	231	76	8822	5.6	9650
b26R	11772	12023	76%	3	44	75	58	53	66	252	83	9548	6.8	5810
B27L	12722	11787	78%	6	42	258	228	164	286	936	311	34169	4.1	25130
b28L	12034	11798	52%	4	45	62	51	56	68	237	78	8797	12	5780
b29R	11941	12282	54%	1	48	99	43	120	80	342	113	10710	5.5	0
b30L	12532	12287	74%	4	38	68	37	56	85	246	81	9240	9	7120
B31R	12817	13512	78%	1	46	209	161	157	169	696	231	25841	11	21800
b32L	13424	13224	56%	6	49	44	43	55	59	201	66	7377	12	60
B33R	13725	14489	84%	3	36	295	136	139	195	765	254	28893	8.4	16820
B34L	18948	14491	78%	5	43	1324	919	1004	1211	4458	1485	154847	5.3	126010
b35L	15050	14853	56%	6	42	54	42	42	60	198	65	7533	11	3930
b36R	16436	16717	52%	2	45	75	71	56	80	282	93	10259	4.5	30

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
b37R	16808	17017	54%	2	45	46	52	43	69	210	69	7532	4.1	5720
b38R	17756	17974	68%	2	45	60	48	51	60	219	72	8116	4.9	12750
B39L	20232	18976	82%	5	35	462	185	254	356	1257	418	48781	7	66350
B40L	24708	20293	84%	5	45	1271	936	1043	1166	4416	1471	152141	5	150700
b41R	21050	21364	54%	2	40	93	71	56	95	315	104	11413	7.2	5300
b42R	22389	22670	64%	3	47	77	83	50	72	282	93	11856	12	16050
b43R	22421	22843	54%	2	46	116	120	73	114	423	140	15311	6.3	5780
b44R	23459	23812	60%	2	50	74	97	81	102	354	117	12726	4.8	7030
b45R	23930	24175	60%	2	48	55	56	63	72	246	81	8900	4.4	1310
b46R	24182	24589	56%	2	47	95	101	89	123	408	135	14730	4.2	3870
B47L	29110	24749	72%	4	43	1305	879	993	1185	4362	1453	150566	5.2	126690
b48L	26325	26110	66%	5	40	60	39	47	70	216	71	8649	8.1	10870
b49R	26791	26988	56%	1	46	54	59	32	53	198	65	8147	12	12180
b50R	26811	27245	62%	3	46	114	123	79	119	435	144	15583	4.1	5780
b51R	27309	27839	68%	3	46	138	138	105	150	531	176	18676	4.1	1460
b52R	27861	28163	62%	3	50	61	82	69	91	303	100	10797	4.2	1340
b53R	28332	28577	60%	3	46	57	53	61	75	246	81	8932	4.4	1310
b54R	28584	28991	56%	3	46	93	100	89	126	408	135	14621	4.5	5150
B55L	32010	29149	88%	5	40	887	511	637	827	2862	953	96122	7.2	48990
b56R	29612	29986	58%	2	38	108	77	64	126	375	124	13367	3.9	60
b57R	30320	30520	62%	2	42	50	46	38	67	201	66	7110	3.8	90
b58R	31124	31528	54%	2	39	117	91	68	129	405	134	15085	6.9	3840
B59R	32108	33343	92%	2	46	345	287	276	328	1236	411	46389	6.2	78240
b60L	32920	32648	56%	4	43	79	52	66	76	273	90	10945	9.4	22250
B61R	33410	34609	76%	2	42	425	221	285	269	1200	399	47155	7.1	71210
b62L	33873	33601	54%	5	43	61	50	67	95	273	90	10315	11	8310
b63L	34102	33848	62%	4	42	51	64	42	98	255	84	9874	11	1370
b64L	34235	33999	58%	6	41	52	66	30	89	237	78	8997	9.5	5180
b65R	34224	34460	56%	3	43	84	46	56	51	237	78	9096	11	7000
b66L	34551	34312	68%	5	44	49	58	48	85	240	79	8904	11	5180
b67L	34994	34635	56%	6	36	131	59	69	101	360	119	14572	11	15990
B68L	35422	34961	78%	4	40	162	90	93	117	462	153	17740	4.5	8310
b69L	35893	35483	56%	4	60	48	80	165	118	411	136	12299	4.4	1310
B70R	35493	36794	78%	3	48	392	349	277	284	1302	433	46282	11	43930
b71R	36206	36433	64%	2	50	60	51	62	55	228	75	9126	8.1	24190
b72L	36671	36414	50%	6	38	73	49	48	88	258	85	9940	8.1	8400
B73L	38602	36791	84%	4	39	637	350	354	471	1812	603	68713	6.8	37670
B74R	38665	39036	82%	1	39	124	76	68	104	372	123	13999	6.7	6490
B75L	39713	39027	64%	6	36	247	125	125	190	687	228	26157	8.2	17330
B76L	40053	39703	80%	5	38	129	63	72	87	351	116	13655	9.9	20340

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
B77R	40140	40433	86%	3	32	102	52	43	97	294	97	11094	10	12860
B78L	40916	40455	74%	6	40	160	86	97	119	462	153	17496	7.7	25490
b79R	40524	40889	68%	3	40	89	78	70	129	366	121	14745	8.6	32580
B80L	42947	41010	76%	6	43	649	380	444	465	1938	645	74189	11	63890
b81R	41058	41318	70%	3	44	63	67	49	82	261	86	9914	11	5720
b82L	41518	41315	52%	4	37	78	34	41	51	204	67	8186	11	2620
B83L	42942	42256	80%	5	45	217	157	155	158	687	228	25944	9.2	16620
b84R	42395	42676	60%	2	48	68	67	68	79	282	93	10236	11	7090
b85R	42643	42987	54%	1	42	84	65	79	117	345	114	13325	7.1	7090
B86L	44356	43352	80%	4	37	333	186	189	297	1005	334	38699	6.2	31300
B87L	44805	44401	68%	5	46	121	75	112	97	405	134	14896	4.2	1280
B88L	45914	44883	72%	6	42	321	192	237	282	1032	343	39141	8	43280
b89R	45133	45390	56%	1	38	58	49	48	103	258	85	9618	11	12750
b90R	45616	45861	58%	1	43	72	59	46	69	246	81	8611	11	1370
B91L	46574	45942	88%	6	46	176	145	149	163	633	210	23884	4.6	34940
b92R	46155	46460	48%	3	45	86	66	72	82	306	101	11906	4.8	16680
b93R	46342	46545	50%	1	43	63	41	47	53	204	67	7426	12	30
B94R	46650	47411	90%	3	44	232	157	180	193	762	253	29733	5.7	52240
b95L	46999	46778	52%	4	41	56	55	37	74	222	73	8198	11	10810
b96L	47248	47030	64%	4	48	48	48	57	66	219	72	7960	11	8280
b97L	47417	47106	66%	6	47	84	70	76	82	312	103	12208	5.3	12750
b98R	47125	47397	50%	1	48	72	69	61	71	273	90	10282	12	12750
B99R	47430	48368	62%	3	47	228	247	197	267	939	312	34094	4.1	23850
b100L	47933	47523	32%	6	48	110	92	106	103	411	136	15466	5.1	7120
b101L	48129	47830	46%	5	51	86	65	89	60	300	99	11742	11	10900
b102L	48497	48021	60%	6	43	150	97	109	121	477	158	18277	6.9	9620
B103R	48399	49460	74%	3	40	361	210	210	281	1062	353	42122	7.7	36360
B104R	49523	50467	78%	2	46	284	199	233	229	945	314	36072	7.7	33150
b105L	50272	49583	50%	4	46	164	172	145	209	690	229	26574	6.3	24130
b106R	49716	49985	54%	3	43	84	47	69	70	270	89	10580	12	1280
b107L	51327	50461	62%	5	38	307	166	161	233	867	288	34169	6	54770
B108L	50779	50576	66%	4	39	66	41	38	59	204	67	8321	11	19150
B109R	50858	51085	68%	2	41	51	39	55	83	228	75	8713	12	12720
B110L	51869	51336	78%	6	36	213	92	98	131	534	177	20852	4.9	14710
b111R	51462	51689	76%	3	34	48	39	38	103	228	75	9215	7.3	5180
b112R	51647	51871	74%	2	40	57	44	45	79	225	74	8972	12	12720
B113R	51897	52244	70%	3	42	89	80	67	112	348	115	13350	8.9	7800
b114L	52130	51933	72%	6	47	58	40	53	47	198	65	7803	12	7000
B115L	52417	52160	82%	4	43	71	50	60	77	258	85	9669	7.7	3840
B116R	52496	53392	70%	2	46	251	197	212	237	897	298	33196	5.5	43390

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
B117L	53903	53409	78%	6	33	196	74	87	138	495	164	19735	4.4	29360
B118R	54289	55131	74%	1	38	301	157	165	220	843	280	32433	7.2	20070
b119L	55073	54864	52%	6	40	42	51	34	83	210	69	7757	7.7	1280
B120R	55378	56112	84%	1	46	248	152	189	146	735	244	27410	9.4	11640
b121L	55797	55408	58%	5	45	81	91	83	135	390	129	15077	9.6	10960
B122L	56702	56124	78%	6	42	205	112	129	133	579	192	22334	7.6	22930
b123R	56358	56663	62%	3	42	76	74	53	103	306	101	11831	4.8	5810
B124L	57325	56762	76%	4	39	189	100	118	157	564	187	22309	8.2	18610
b125R	56984	57241	64%	2	37	73	50	46	89	258	85	9981	10	12180
B126R	57440	58132	80%	2	40	230	115	165	183	693	230	26649	4.9	31360
b127L	57923	57678	50%	6	41	57	52	49	88	246	81	9809	10	20400
b128L	58104	57907	52%	5	46	40	52	39	67	198	65	7039	11	2590
b129L	58339	58136	64%	4	45	65	45	47	47	204	67	8130	12	7000
B130R	58211	58696	62%	2	43	128	106	101	151	486	161	18260	4.3	29450
b131R	58365	58583	62%	3	43	60	52	43	64	219	72	8256	12	5750
b132L	58974	58738	52%	5	40	57	52	43	85	237	78	8910	12	5150
B133R	58776	60110	80%	3	42	472	272	291	300	1335	444	51928	9.7	69390
b134L	59114	58803	64%	6	42	64	75	55	118	312	103	12581	7.8	22310
b135L	59214	59014	54%	5	41	43	40	43	75	201	66	7495	6.3	8400
B136L	61382	60111	86%	6	41	460	262	255	295	1272	423	48541	11	45260
B137L	62395	61433	82%	4	36	297	175	173	318	963	320	36868	4.2	90810
b138R	61694	61939	66%	2	41	75	55	47	69	246	81	9885	12	19660
B139R	62474	64000	86%	2	37	455	264	298	510	1527	508	58790	8.9	106150
b140L	62904	62707	66%	5	37	57	40	33	68	198	65	6944	11	1310
b141L	63313	62987	62%	4	37	99	69	52	107	327	108	12675	7.2	10990
b142L	64323	63955	42%	5	40	109	87	59	114	369	122	14433	6.9	12180
B143R	64124	65911	76%	2	46	516	395	430	447	1788	595	65547	5.9	45560
b144L	64525	64283	58%	4	49	56	56	64	67	243	80	9312	11	5720
b145L	65092	64769	52%	4	39	91	70	56	107	324	107	13058	8	9620
b146L	65668	65345	50%	4	49	78	91	68	87	324	107	12913	10	4020
b147L	66322	65978	60%	4	41	90	69	73	113	345	114	12773	9.7	9740
B148R	65997	66962	82%	3	42	296	191	213	266	966	321	37156	7.6	34400
b149L	66692	66303	66%	6	42	115	82	80	113	390	129	15404	6.2	19720
B150L	67869	66997	90%	5	39	291	168	169	245	873	290	33857	7	43930
b151R	67201	67524	62%	1	40	92	59	69	104	324	107	12413	6.2	14090
b152R	67392	67625	58%	3	40	63	47	46	78	234	77	9391	9.2	4080
b153R	67662	67859	72%	3	40	53	38	42	65	198	65	8102	11	10840
B154L	68926	67910	64%	4	40	335	206	201	275	1017	338	38140	8	15750
b155R	68327	68542	54%	2	44	56	41	53	66	216	71	8475	8.2	9710
b156R	68717	68920	52%	2	32	61	27	38	78	204	67	8207	4.2	10900

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
B157L	69356	68865	78%	6	41	164	96	106	126	492	163	19370	9.1	18520
b158R	69009	69380	56%	3	42	100	87	70	115	372	123	14529	8	10900
B159R	69422	72025	82%	2	42	766	508	594	736	2604	867	99779	6.1	150910
b160L	70413	70192	50%	5	43	60	44	51	67	222	73	8001	8.9	2680
b161R	70743	71042	54%	3	42	89	53	72	86	300	99	11533	8.5	11500
b162L	71629	71198	74%	4	44	120	98	92	122	432	143	16696	6.1	15370
B163R	72054	73085	70%	3	44	304	201	258	269	1032	343	38744	6.4	45980
b164L	72869	72429	50%	6	42	121	109	78	133	441	146	17220	7.8	4020
B165R	73136	74515	80%	2	41	472	275	295	338	1380	459	53301	9.3	66830
b166L	73854	73648	64%	5	47	44	50	47	66	207	68	7203	7.2	1430
b167R	73818	74015	58%	3	41	63	38	44	53	198	65	7454	11	2560
B168R	74535	74849	72%	3	40	112	59	68	76	315	104	12460	11	27880
b169L	74781	74557	68%	5	43	57	57	40	71	225	74	8637	7.3	7030
B170R	74904	79064	86%	3	46	1157	943	973	1088	4161	1386	144434	4.8	97720
b171L	77620	77414	68%	4	48	56	51	48	52	207	68	7443	4.3	0
b172L	78367	78164	52%	4	43	59	47	41	57	204	67	7414	4.4	7030
B173L	80294	79050	80%	6	28	534	145	200	366	1245	414	48626	10	66830
b174L	80976	80671	66%	5	46	100	64	76	66	306	101	11612	11	16680
B175L	81274	80696	72%	4	44	196	115	139	129	579	192	22608	9.9	47660
b176R	81260	81484	54%	2	38	71	39	47	68	225	74	8756	13	12660
B177R	81304	82023	74%	1	44	235	146	169	170	720	239	26205	10	21770
b178L	81902	81693	60%	6	45	47	48	46	69	210	69	7419	4.4	60
B179L	83145	82024	80%	5	39	400	193	239	290	1122	373	43101	9.3	63760
b180L	83063	82863	56%	6	37	82	32	43	44	201	66	7785	12	9530
B181L	83646	83239	70%	5	44	118	97	83	110	408	135	15940	5.3	16050
b182R	83419	83664	52%	1	41	66	45	55	80	246	81	9355	6.8	30
B183L	84405	83731	88%	5	35	249	114	123	189	675	224	26259	4.9	20400
b184R	84338	85081	56%	2	33	258	138	110	238	744	247	28194	10	20480
B185L	85553	85059	80%	6	38	189	99	88	119	495	164	19412	10	14710
b186R	85081	85332	58%	1	33	64	35	49	104	252	83	9100	11	1280
B187R	85628	86065	78%	2	43	180	97	93	68	438	145	16408	12	1280
B188R	86122	86415	82%	1	44	75	62	68	89	294	97	11556	11	11020
b189L	86411	86193	58%	6	47	70	51	51	47	219	72	8494	12	1280
B190L	86605	86303	84%	4	41	99	59	64	81	303	100	11805	10	23560
b191L	86982	86737	56%	5	41	60	46	56	84	246	81	9785	9.6	15310
B192R	86869	90096	90%	1	45	1110	788	679	651	3228	1075	117502	11	79690
b193L	88266	87613	50%	5	45	147	138	158	211	654	217	25487	8.1	31300
b194L	89385	88372	50%	5	43	202	208	229	375	1014	337	40258	8.3	38800
b195L	89784	89554	62%	5	38	41	40	48	102	231	76	9345	8.7	14830
b196L	90176	89835	62%	6	56	54	62	129	97	342	113	10995	8.8	7060

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
B197L	90498	90103	76%	5	42	133	80	86	97	396	131	14998	4.7	3930
b198R	90137	90364	54%	2	40	53	50	42	83	228	75	8438	8.4	3870
B199R	90758	91789	70%	2	43	351	217	231	233	1032	343	40225	9.9	54200
b200L	91679	91434	60%	6	42	63	54	50	79	246	81	10051	8	15310
B201R	92170	92538	84%	1	33	127	53	70	119	369	122	13526	8.9	10930
b202L	93004	92606	62%	4	41	116	92	71	120	399	132	14801	11	7740
B203R	92613	93995	80%	3	43	430	287	309	357	1383	460	51877	8.6	30820
b204L	93509	93297	64%	6	48	53	44	58	58	213	70	8331	9.5	15280
b205R	93739	93957	48%	1	41	82	44	45	48	219	72	8762	11	13940
B206L	94835	93996	80%	6	37	350	110	197	183	840	279	31756	10	57300
B207L	95322	94987	82%	5	47	102	76	81	77	336	111	12205	4	9530
B208L	95646	95359	72%	5	41	91	53	65	79	288	95	11378	3.4	19060
B209R	95821	96132	72%	1	48	110	72	79	51	312	103	11746	11	0
b210R	96129	96326	60%	3	31	66	25	37	70	198	65	7841	9.2	9560
B211L	96657	96319	80%	5	39	124	60	72	83	339	112	12739	5.7	60
B212L	97112	96666	92%	6	37	176	76	91	104	447	148	17372	10	19060
b213R	96790	97032	64%	1	40	50	53	43	97	243	80	8835	4.8	2560
B214R	97201	98007	86%	1	45	238	166	199	204	807	268	31316	4.5	48430
B215R	98046	98549	92%	3	45	155	110	117	122	504	167	18402	4.8	22190
b216L	98691	98428	50%	5	42	71	56	56	81	264	87	10544	11	8340
b217R	98550	98771	60%	3	41	77	41	49	55	222	73	8447	12	5750
B218R	98605	99807	74%	1	44	408	246	287	262	1203	400	47203	6.2	68620
b219L	99210	98752	60%	5	47	101	108	106	144	459	152	17792	9.4	21000
b220L	99643	99194	56%	4	43	101	110	85	154	450	149	18180	10	11700
b221L	99444	99229	66%	5	44	46	56	39	75	216	71	8521	11	3870
B222R	99830	100795	70%	2	44	286	204	225	251	966	321	36606	6.3	21170
b223L	100651	100367	62%	4	41	82	65	53	85	285	94	11355	6.9	1310
B224R	100805	101911	78%	2	47	323	245	275	264	1107	368	40379	10	56590
b225R	101320	101607	56%	1	51	76	75	72	65	288	95	11387	12	12720
B226L	102754	101915	82%	4	43	269	168	189	214	840	279	30630	8.1	17360
b227L	102443	102189	56%	6	43	72	54	55	74	255	84	10094	11	10930
b228R	102495	102713	58%	3	48	47	61	45	66	219	72	7508	5.5	1340
b229L	102791	102510	70%	6	45	92	57	69	64	282	93	11600	12	10960
B230L	103948	102800	84%	4	42	369	237	245	298	1149	382	43397	6.6	16200
b231R	103109	103456	64%	2	43	99	76	75	98	348	115	13638	8.2	5180
b232R	103233	103439	58%	3	43	55	44	46	62	207	68	7749	12	1280
b233R	103730	103942	48%	2	40	52	47	38	76	213	70	8367	8.4	7710
b234R	103983	104264	84%	3	38	70	50	56	106	282	93	10893	7.8	7120
B235L	104315	104004	70%	6	37	115	61	55	81	312	103	12058	9.4	9650
B236L	105293	104319	82%	6	40	324	188	200	263	975	324	37284	7.7	27660

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
b237R	104784	105035	58%	3	39	66	54	45	87	252	83	10158	11	2680
b238R	104938	105144	62%	1	38	60	42	36	69	207	68	7816	6.2	6400
B239R	105405	107828	70%	3	47	724	539	603	558	2424	807	88361	8.8	116250
b240L	105932	105714	56%	6	51	52	53	59	55	219	72	8791	12	8280
b241L	106345	105950	56%	4	49	95	109	86	106	396	131	13926	11	5180
b242L	106218	106021	46%	5	48	48	49	46	55	198	65	7761	11	8430
b243R	107021	107257	60%	2	47	66	47	65	59	237	78	9690	12	22310
b244L	107501	107229	52%	6	45	53	72	52	96	273	90	10700	7.4	7770
b245L	107728	107525	52%	4	47	40	55	41	68	204	67	7077	5.3	1340
B246R	108041	108649	76%	2	33	243	84	120	162	609	202	23665	10	37670
b247L	108376	108161	74%	4	42	46	60	30	80	216	71	8634	12	2620
b248L	109088	108882	48%	6	43	52	42	47	66	207	68	7361	12	1310
B249R	108892	111174	70%	1	46	674	476	583	550	2283	760	86967	8.2	102570
b250R	109008	109313	66%	3	45	98	63	74	71	306	101	11567	6.5	8310
b251L	110856	110617	50%	5	50	59	65	54	62	240	79	9280	12	30
b252L	111274	111068	64%	4	48	42	51	49	65	207	68	7243	12	1280
B253R	111273	111719	54%	3	40	162	86	95	104	447	148	17282	9.6	3900
b254L	111590	111300	58%	6	43	73	63	62	93	291	96	11554	8.5	10930
B255R	111779	112039	76%	2	42	81	48	61	71	261	86	10381	6.5	24750
B256R	112097	112879	60%	2	40	274	146	167	196	783	260	29717	5.1	38950
b257R	112119	112532	54%	3	42	143	82	90	99	414	137	16214	12	11020
B258R	112905	116741	84%	3	47	1320	885	908	724	3837	1278	142004	12	44090
b259L	113388	113161	76%	5	51	55	59	57	57	228	75	8877	11	7030
b260R	113323	113526	50%	1	43	66	53	35	50	204	67	8126	11	1370
B261L	117526	116738	84%	4	42	244	164	164	217	789	262	29687	4.7	12980
b262R	117035	117442	62%	2	43	115	85	90	118	408	135	16275	6.2	6550
b263R	117285	117488	58%	3	47	56	49	46	53	204	67	7827	11	8310
B264L	118018	117560	86%	4	39	125	80	100	154	459	152	17347	7.6	30620
b265R	117803	118021	70%	2	42	62	46	45	66	219	72	8313	7.9	6490
b266R	117834	118031	64%	3	41	59	43	39	57	198	65	7212	6.5	3840
B267R	118059	118361	82%	3	39	98	67	52	86	303	100	11010	11	9530
B268L	118935	118363	74%	5	45	171	100	159	143	573	190	21805	8.6	16410
b269R	118429	118770	54%	1	43	85	90	56	111	342	113	13192	9.8	2710
b270L	119236	118967	58%	4	40	70	53	55	92	270	89	10826	10	7830
B271R	119006	119443	66%	2	42	145	93	90	110	438	145	16486	10	10930
B272L	119752	119450	68%	4	43	86	65	66	86	303	100	10750	11	13940
B273L	120114	119776	90%	5	50	82	76	94	87	339	112	12174	5.1	10930
B274R	120178	120828	90%	1	41	203	137	129	182	651	216	24067	5.1	15930
b275L	120539	120327	64%	6	45	50	43	53	67	213	70	7142	9.5	2560
b276L	121069	120794	44%	4	42	76	59	58	83	276	91	10100	4.8	3840

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
B277R	120837	121457	70%	3	48	200	150	145	126	621	206	22638	12	9530
B278R	121519	122637	88%	1	45	329	237	262	291	1119	372	41924	4.6	41600
b279L	121992	121726	64%	5	44	62	61	56	88	267	88	10437	5.1	3960
b280L	122077	121778	52%	4	44	73	70	61	96	300	99	12014	10	13520
b281L	122298	122038	46%	5	48	60	58	67	76	261	86	9876	5.9	3960
b282L	122941	122699	32%	4	55	28	74	59	82	243	80	8644	9.1	180
B283R	122739	123410	76%	3	53	204	137	221	110	672	223	25831	4.8	15220
B284L	123853	123407	78%	4	42	138	95	92	122	447	148	16592	4.7	15220
B285L	124281	123880	70%	5	42	144	74	93	91	402	133	15172	7.2	2560
B286R	124399	125490	80%	1	37	407	178	225	282	1092	363	42042	5.8	55630
b287L	124767	124495	60%	5	41	69	70	41	93	273	90	10696	11	2620
B288L	126376	125456	58%	4	43	297	164	231	229	921	306	34116	9.7	53400
B289L	127526	126378	88%	6	37	407	183	242	317	1149	382	44120	10	42730
b290R	127120	127386	64%	1	39	60	57	46	104	267	88	10011	9.4	3870
B291L	127983	127567	74%	5	38	148	76	81	112	417	138	16026	10	21620
B292L	128227	128000	80%	4	39	55	38	50	85	228	75	8399	7.9	14030
b293R	128201	128842	58%	2	36	194	101	131	216	642	213	24248	9.8	24270
b294R	128500	128802	62%	1	39	82	54	63	104	303	100	11741	8.5	15340
b295L	128706	128509	60%	5	37	68	40	33	57	198	65	7394	9.4	8280
B296L	128851	128615	76%	4	38	77	46	44	70	237	78	9779	9.1	9710
B297R	128897	129991	76%	2	40	376	211	232	276	1095	364	41803	5.2	62450
B298L	131112	129994	72%	5	40	389	205	248	277	1119	372	43122	10	60040
b299R	130109	130354	70%	2	42	59	67	36	84	246	81	9105	6.5	2560
b300R	130765	130974	56%	1	42	50	46	42	72	210	69	8279	10	8340
b301L	131445	131209	64%	5	38	60	47	44	86	237	78	9721	10	14000
B302R	131221	131556	78%	1	39	124	67	63	82	336	111	13384	10	28020
B303L	131879	131553	84%	6	42	105	72	65	85	327	108	12653	11	6460
b304R	131655	131894	60%	3	43	62	45	59	74	240	79	9396	5.6	6970
B305R	131942	133465	90%	2	44	496	322	347	359	1524	507	57118	9.5	102310
b306L	132505	132086	64%	4	42	97	93	83	147	420	139	16462	6.1	9590
b307R	132274	132570	58%	1	46	95	59	79	64	297	98	11647	12	8340
b308L	132988	132746	56%	4	46	54	52	59	78	243	80	9421	9.6	16590
b309L	133471	133079	76%	4	45	107	93	82	111	393	130	16213	8.4	19150
B310R	133532	134680	80%	2	44	366	231	277	275	1149	382	43271	8.3	33390
b311L	134132	133836	56%	6	42	63	78	48	108	297	98	11590	8	13550
b312L	134378	134157	56%	6	49	53	49	59	61	222	73	8906	11	1400
b313L	134482	134231	56%	4	49	60	57	67	68	252	83	9383	4.3	5810
B314L	135118	134681	82%	4	43	143	88	99	108	438	145	16944	7	15960
b315R	134867	135148	58%	2	40	70	55	58	99	282	93	11172	8.7	5180
B316R	135269	137446	86%	2	43	695	466	466	551	2178	725	82708	6.7	81720

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
b317L	136156	135803	58%	4	43	78	67	86	123	354	117	13666	5.2	7060
b318L	136446	136156	58%	5	49	77	80	62	72	291	96	10472	11	8340
b319L	136869	136609	66%	5	44	61	60	55	85	261	86	9935	8.1	9590
b320L	136906	136646	68%	4	45	59	61	56	85	261	86	9951	9.5	3870
b321L	137268	136966	60%	5	41	84	62	63	94	303	100	11249	10	12090
B322R	137470	138399	76%	1	51	286	264	210	170	930	309	33066	11	24990
B323R	138490	139632	82%	1	37	382	193	234	334	1143	380	43577	6.1	29710
B324L	141277	139790	56%	4	42	506	289	338	355	1488	495	57849	8.8	79640
b325R	139886	140239	62%	2	43	80	76	75	123	354	117	13790	6.5	2740
b326R	140005	140226	56%	1	44	49	49	48	76	222	73	8788	10	7770
b327L	140436	140146	68%	5	42	98	54	67	72	291	96	11148	12	7090
b328R	140472	140717	68%	3	46	53	61	52	80	246	81	8952	8.1	5840
b329R	140783	141154	58%	2	44	80	94	70	128	372	123	14323	10	5180
b330R	140919	141176	66%	3	45	56	63	52	87	258	85	9506	6.5	3900
B331R	141226	142155	74%	1	39	313	187	176	254	930	309	36241	6.8	29360
b332L	142200	141820	56%	5	41	96	75	83	127	381	126	15611	10	36250
b333L	142088	141867	70%	6	41	53	44	47	78	222	73	8029	11	2620
B334R	142342	143640	78%	1	43	433	282	281	303	1299	432	49997	11	52990
b335R	142847	143107	60%	2	46	70	57	64	70	261	86	10130	10	14060
B336R	143748	144032	78%	3	42	72	61	58	94	285	94	10516	8.1	8310
b337L	144046	143828	50%	4	44	68	45	52	54	219	72	7896	8.7	1340
b338R	143894	144289	60%	2	49	78	105	91	122	396	131	15934	12	17900
B339L	144764	144033	78%	6	47	242	150	191	149	732	243	27091	9.4	10360
b340R	144567	144785	56%	3	37	58	48	32	81	219	72	8567	8.2	3960
B341R	145096	145326	94%	1	35	103	36	44	48	231	76	8922	8.4	8280
B342R	145380	145880	68%	3	36	175	82	99	145	501	166	19570	10	21650
b343L	145670	145416	58%	6	35	67	49	40	99	255	84	10505	9.6	3930
B344R	145914	146525	86%	3	27	250	72	94	196	612	203	24958	9.4	37100
b345R	146532	146747	76%	3	30	68	27	37	84	216	71	8554	7.4	13460
B346L	147400	146651	80%	4	36	304	120	150	176	750	249	28613	11	37700
b347R	147171	147380	66%	3	41	45	46	40	79	210	69	7949	9.2	4050
B348L	148125	147478	84%	5	46	211	116	180	141	648	215	24357	8.8	70510
B349R	148156	148359	86%	1	42	47	45	40	72	204	67	7720	9.6	14030
B350L	149086	148337	82%	4	42	258	149	163	180	750	249	27961	9.7	37470
b351R	148536	148859	50%	3	40	74	63	67	120	324	107	11853	8.8	3930
b352R	148738	148968	62%	1	40	54	52	41	84	231	76	8917	8.4	6490
b353R	148965	149195	66%	3	37	65	48	38	80	231	76	8793	7.7	3840
B354L	149909	149115	92%	6	44	238	164	182	211	795	264	30216	6.7	20400
b355R	149131	149385	80%	1	44	66	52	61	76	255	84	9506	12	12720
b356R	149313	149918	52%	3	44	165	143	121	177	606	201	23987	6.3	28060

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
B357L	150365	149946	86%	6	41	130	79	93	118	420	139	15823	11	15220
B358R	150699	151376	76%	3	36	245	109	136	188	678	225	27128	4.6	35910
B359R	151439	152248	78%	2	40	257	145	181	227	810	269	31031	8.4	28830
b360L	151701	151480	66%	5	43	51	50	45	76	222	73	7954	7.9	2650
B361R	152245	153897	64%	1	36	645	304	295	409	1653	550	62900	11	68250
b362L	152549	152337	64%	6	39	51	43	40	79	213	70	7768	7.3	1310
b363L	153719	153279	68%	6	41	85	60	122	174	441	146	15324	3.8	5150
b364R	153945	154205	70%	3	40	52	51	54	104	261	86	9585	7.3	3900
B365L	155794	153977	84%	4	44	626	384	412	396	1818	605	68916	11	70100
b366R	154349	154567	70%	2	43	39	42	52	86	219	72	8862	10	49480
b367R	155445	155726	58%	3	43	66	66	55	95	282	93	10256	3.6	8310
B368L	157355	155850	70%	6	46	487	338	348	333	1506	501	55852	11	54170
b369L	156439	156236	54%	4	50	58	59	43	44	204	67	7695	12	90
B370L	158428	157379	82%	4	42	380	206	233	231	1050	349	40458	10	57240
B371L	159338	158499	78%	6	43	276	189	175	200	840	279	30624	8.6	12860
b372R	158851	159144	48%	1	44	63	55	73	103	294	97	10278	8.1	9590
b373R	159166	159450	70%	1	42	71	65	54	95	285	94	10983	11	6430
b374L	159688	159350	60%	4	40	114	62	75	88	339	112	12758	9.8	5150
b375L	160259	159720	68%	6	37	206	100	102	132	540	179	21714	7.3	36280
b376L	160212	159943	62%	5	37	94	51	49	76	270	89	11085	12	3840
B377R	159965	160177	70%	2	38	60	37	44	72	213	70	8308	12	1400
B378L	162330	160393	78%	5	43	649	380	444	465	1938	645	74185	11	63890
b379R	160441	160701	70%	1	44	63	67	49	82	261	86	9914	11	5720
b380L	160901	160698	52%	6	37	78	34	41	51	204	67	8186	11	2620
B381L	162325	161639	80%	4	45	217	157	155	158	687	228	25944	9.2	16620
b382R	161778	162059	60%	3	48	68	67	68	79	282	93	10236	11	7090
b383R	162026	162370	54%	2	42	84	65	79	117	345	114	13309	7.1	7090
b384L	163226	162789	54%	6	43	113	103	86	136	438	145	15886	4.8	9650
B385R	162793	163911	74%	1	43	366	212	269	272	1119	372	41924	9.3	46400
b386L	163670	163389	52%	6	46	66	75	56	85	282	93	10509	8.1	8310
b387L	163806	163585	68%	5	45	52	55	46	69	222	73	8841	11	12090
B388L	164705	163899	80%	6	42	252	169	172	214	807	268	30152	6.8	18640
B389R	164757	165632	80%	3	39	335	140	199	202	876	291	33847	9.1	61170
b390L	165197	164805	66%	6	41	97	94	66	136	393	130	15522	7.7	17930
b391L	165036	164815	58%	5	42	55	52	41	74	222	73	8905	8.9	8340
b392L	165564	165280	64%	5	38	58	66	43	118	285	94	11410	8	7770
B393L	166590	165622	68%	5	44	325	207	219	218	969	322	36661	9.8	62850
b394R	165839	166255	58%	2	47	88	101	96	132	417	138	15506	11	10900
B395L	167521	166559	70%	4	43	319	206	212	226	963	320	36181	6.9	30730
b396R	166583	166786	68%	2	43	50	42	45	67	204	67	7923	6.1	2560

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
B397L	168970	167549	80%	4	40	477	271	292	382	1422	473	56137	4.5	92240
b398R	168014	168241	60%	2	38	58	41	45	84	228	75	9122	5.8	17270
B399R	169101	169883	80%	3	38	262	134	161	226	783	260	30965	6.7	39660
b400L	169541	169284	66%	6	38	73	58	40	87	258	85	9923	7.5	14030
B401L	170423	169941	74%	6	37	174	90	90	129	483	160	19026	10	21110
B402L	171327	170512	76%	5	37	280	152	153	231	816	271	32142	6.3	28760
B403R	171458	172438	72%	2	38	322	180	195	284	981	326	37809	9.1	59690
B404R	172455	172853	86%	3	45	129	87	91	92	399	132	15150	12	10810
b405L	172778	172569	56%	6	43	38	48	43	81	210	69	8129	9	1400
b406L	173638	172892	56%	4	43	193	179	142	233	747	248	29947	12	56930
b407R	172923	173882	64%	3	45	288	198	234	240	960	319	36487	11	4080
B408R	172928	174238	66%	2	46	392	284	316	319	1311	436	48634	8.5	59400
b409R	173722	174120	60%	1	52	102	110	99	88	399	132	15851	12	7060
b410L	174082	173792	48%	4	53	63	73	80	75	291	96	10947	5.7	60
B411L	175347	174235	64%	5	42	355	224	246	288	1113	370	42137	9	66100
b412R	174400	174630	54%	1	35	61	40	42	88	231	76	9153	10	12150
b413R	174647	174910	68%	2	45	60	59	61	84	264	87	9674	10	12750
b414R	175160	175372	42%	2	42	50	48	42	73	213	70	7339	11	1340
b415L	175647	175369	58%	5	37	93	54	49	83	279	92	10660	11	2590
B416R	175415	176533	70%	2	37	362	190	223	344	1119	372	42556	8.5	28800
b417L	176281	176030	58%	4	35	80	54	35	83	252	83	9736	9.9	2650
B418R	176591	177745	80%	2	40	377	210	249	319	1155	384	44216	8.4	48460
B419L	178256	177705	88%	6	39	174	112	101	165	552	183	21610	10	30590
b420L	178083	177886	62%	5	43	54	37	48	59	198	65	7723	8.7	2710
B421L	178974	178291	78%	5	43	230	127	164	163	684	227	25866	6.7	13430
b422L	178819	178421	56%	4	45	135	81	97	86	399	132	16357	11	21120
b423R	178624	178854	56%	1	44	47	57	45	82	231	76	8991	5.9	3900
B424L	180056	179001	78%	6	37	387	182	209	278	1056	351	41185	10	66800
b425R	179072	179386	58%	2	40	78	69	57	111	315	104	12435	7.3	14770
B426L	180802	180170	74%	4	40	215	137	115	166	633	210	24078	5.2	20430
b427L	180756	180466	56%	5	40	100	57	59	75	291	96	10795	13	7000
b428L	181027	180818	60%	4	32	54	36	32	88	210	69	7975	11	5120
B429R	180858	181103	76%	3	35	99	42	44	61	246	81	9852	7.5	5120
B430L	181638	181093	68%	5	45	194	119	127	106	546	181	20417	11	21770
B431L	181923	181660	68%	5	41	86	45	64	69	264	87	10597	4.1	19150
B432L	182286	181975	92%	5	42	103	61	69	79	312	103	11973	4.3	13370
B433L	183229	182408	86%	4	32	366	112	149	195	822	273	30938	10	32580
B434L	183809	183297	88%	6	46	163	128	106	116	513	170	18454	8.4	14680
B435L	184818	184072	72%	5	50	190	198	172	187	747	248	27584	5.1	26030
B436R	185022	185399	66%	3	42	102	85	73	118	378	125	14355	8.6	7890

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
B437L	185347	185117	78%	4	46	62	45	61	63	231	76	8354	8.9	8370
B438R	185480	185686	64%	2	34	62	31	40	74	207	68	8317	9.4	11610
b439R	185526	185843	58%	3	37	102	52	66	98	318	105	12534	10	19060
B440L	186653	185835	78%	6	35	342	114	174	189	819	272	30868	9.7	43390
B441R	186800	186997	68%	2	45	69	39	50	40	198	65	7379	11	5120
b442L	187558	187109	54%	4	50	93	65	162	130	450	149	13786	4.9	2590
B443R	187137	188429	74%	3	47	389	334	273	297	1293	430	48140	11	95820
b444R	187222	187596	54%	1	54	104	145	58	68	375	124	14390	11	9650
b445L	188739	188440	74%	5	43	73	63	67	97	300	99	11786	4.3	15310
B446R	188488	189726	62%	1	42	433	243	273	290	1239	412	48719	7.3	72520
b447R	188648	188848	62%	2	43	65	48	39	49	201	66	7614	11	2560
b448L	189490	189176	48%	4	40	78	67	58	112	315	104	12355	6.1	14800
b449L	189953	189654	40%	6	38	71	64	50	115	300	99	11406	6.8	9590
B450R	189733	190080	72%	1	44	134	74	78	62	348	115	12813	10	6970
B451R	190094	190450	82%	2	48	83	79	92	103	357	118	12804	8.4	13460
b452R	190383	190754	50%	3	42	93	68	87	124	372	123	15639	10	38100
B453L	190988	190452	96%	6	41	182	116	103	136	537	178	20648	5.4	15960
b454R	190762	190983	62%	1	44	48	49	48	77	222	73	7771	11	1280
B455L	192365	191064	78%	6	46	404	282	323	293	1302	433	48401	4.7	18490
b456R	191807	192175	48%	2	48	79	99	79	112	369	122	14528	11	13430
B457L	193000	192401	62%	4	42	213	125	124	138	600	199	23226	9.2	19150
B458L	194109	193033	90%	5	42	340	221	236	280	1077	358	42419	9.6	69450
b459R	193084	193284	64%	1	42	48	50	35	68	201	66	7741	6.2	2560
B460R	194141	194467	84%	2	43	101	62	77	87	327	108	12106	8.4	3870
B461R	195038	195238	46%	2	46	52	44	49	56	201	66	7350	11	14030
B462R	195636	195845	76%	3	40	51	40	43	76	210	69	8081	6.5	9710
B463L	195924	195718	74%	5	38	75	38	40	54	207	68	7965	10	1340
b464L	196239	196006	58%	5	41	57	51	44	82	234	77	8846	4.6	6520
B465R	196060	197199	76%	1	45	318	263	250	309	1140	379	41980	5.3	23880
b466L	196973	196764	52%	6	45	65	46	49	50	210	69	7668	11	60
b467L	197178	196840	76%	5	43	93	70	76	100	339	112	13299	8	4020
B468R	197225	197980	76%	2	43	234	127	198	197	756	251	27537	8.3	48880
B469L	199194	197983	72%	5	40	346	232	251	383	1212	403	46108	4.7	65020
b470R	198020	198232	74%	2	38	66	40	41	66	213	70	7690	10	2560
b471R	198286	198522	50%	1	39	74	45	48	70	237	78	9536	11	15220
B472R	199284	200867	80%	3	42	433	321	350	480	1584	527	60491	9.4	106950
b473L	199573	199370	58%	4	49	58	50	49	47	204	67	7292	10	6430
b474L	199871	199653	54%	6	37	54	46	35	84	219	72	8468	4.8	10930
b475L	199876	199667	56%	4	38	52	45	35	78	210	69	7833	9.6	5240
b476R	200290	200511	54%	1	45	51	49	51	71	222	73	8105	8.2	2680

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
B477R	200923	202254	78%	1	35	489	200	272	371	1332	443	51079	7.4	41010
b478L	201560	201327	52%	6	39	63	48	44	79	234	77	8675	11	8280
b479L	201818	201615	62%	6	33	65	43	24	72	204	67	7607	10	8280
B480L	203450	202287	80%	6	46	316	277	257	314	1164	387	43025	8.2	91000
b481L	202819	202556	58%	4	54	59	94	48	63	264	87	10764	12	1310
b482R	202586	202930	58%	2	52	85	70	110	80	345	114	13268	11	5150
b483L	203374	203159	60%	4	49	61	45	61	49	216	71	8015	4.8	7060
b484L	203601	203203	64%	5	43	127	78	94	100	399	132	15907	12	50820
B485L	204187	203471	74%	4	40	232	105	184	196	717	238	28072	10	12300
b486R	203549	203764	60%	2	39	61	56	29	70	216	71	8602	10	2560
B487L	204708	204289	76%	5	43	113	92	90	125	420	139	16081	4.1	12920
B488L	206441	204786	78%	6	49	492	396	419	349	1656	551	60638	8.9	53510
b489R	205122	205568	52%	3	50	97	117	105	128	447	148	16752	12	11470
b490L	205678	205436	48%	4	51	70	54	69	50	243	80	9335	12	1310
b491R	206097	206435	60%	3	46	70	86	70	113	339	112	13488	11	27970
B492R	206482	206817	82%	1	42	92	84	57	103	336	111	13208	10	12120
b493L	206785	206543	68%	4	43	75	42	63	63	243	80	9137	4.3	15220
B494R	206855	207403	68%	2	45	172	133	115	129	549	182	20794	9.2	9020
b495L	207124	206912	44%	4	48	50	45	57	61	213	70	8186	5.5	5840
B496L	207712	207413	56%	4	45	104	64	71	61	300	99	11353	10	10240
B497L	208940	208185	74%	6	38	255	98	186	217	756	251	28362	10	49050
B498R	209126	209494	70%	2	49	105	91	89	84	369	122	14876	6.1	33030
B499L	210510	209473	74%	5	43	378	187	255	218	1038	345	39655	10	36360
b500R	209963	210295	50%	2	47	56	88	68	121	333	110	12242	11	16590
b501R	210076	210411	54%	1	43	65	85	58	128	336	111	13067	11	7770
b502R	210590	210802	70%	2	48	45	42	60	66	213	70	7943	8.5	15340
B503L	211246	210623	88%	4	49	203	173	131	117	624	207	23363	3.7	32430
b504L	211098	210886	66%	5	53	73	58	54	28	213	70	7707	12	5720
b505R	210940	211146	48%	1	49	29	50	51	77	207	68	8920	13	1490
B506L	212324	211311	84%	6	39	360	189	206	259	1014	337	38565	4.3	13460
b507R	211564	212292	60%	1	38	181	145	130	273	729	242	26362	4.6	17900
B508R	212392	217230	82%	1	43	1618	1019	1081	1121	4839	1612	179397	11	110200
b509L	212669	212451	62%	6	46	49	46	54	70	219	72	8112	9	150
b510L	212878	212678	58%	4	45	45	37	53	66	201	66	8484	8.4	22280
b511L	213222	212935	60%	5	46	67	71	62	88	288	95	10838	6.9	9620
b512L	214583	214362	62%	6	36	56	41	39	86	222	73	8306	8.4	3930
b513L	214842	214639	54%	5	43	43	50	37	74	204	67	7789	7.7	5240
b514L	215348	214980	66%	6	44	85	84	79	121	369	122	13589	8.2	3960
b515L	215996	215751	60%	6	43	54	53	53	86	246	81	8910	4.7	3930
b516L	216333	215836	56%	5	45	120	118	106	154	498	165	19661	10	26270

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
b517L	216826	216590	50%	4	42	52	52	48	85	237	78	9126	8	3930
b518L	217097	216810	50%	6	57	43	89	76	80	288	95	10028	3.1	1280
B519L	217629	217231	74%	5	39	147	73	83	96	399	132	15165	4.9	3960
b520R	217366	217638	58%	1	40	66	58	50	99	273	90	10867	11	9020
b521R	217670	217954	72%	2	42	52	70	50	113	285	94	10556	7.9	4050
B522L	218432	217674	88%	6	43	265	148	181	165	759	252	28632	8.9	12950
b523R	218670	218972	58%	3	48	63	80	64	96	303	100	11264	13	60
B524L	219443	218676	80%	6	44	262	136	201	169	768	255	28697	9.8	75870
B525L	220100	219471	82%	6	46	181	148	139	162	630	209	24565	11	35080
b526R	219582	220115	60%	3	46	136	119	125	154	534	177	20544	11	7060
b527R	219832	220038	58%	1	45	56	48	46	57	207	68	7733	11	5150
b528L	220152	219952	84%	5	40	55	45	35	66	201	66	8131	11	12690
B529R	220244	221701	94%	2	42	449	301	316	392	1458	485	55622	6.6	52440
b530L	220800	220429	46%	5	44	89	88	75	120	372	123	14001	11	13460
b531L	221479	221246	58%	4	39	63	46	45	80	234	77	9220	8.7	10870
b532R	221970	222647	82%	3	48	149	143	185	201	678	225	23310	7.4	5240
B533L	223579	221981	90%	4	44	526	341	355	377	1599	532	57886	7.6	71820
b534R	222934	223149	54%	1	45	44	56	41	75	216	71	8457	12	1340
b535R	223332	223586	78%	3	34	72	41	46	96	255	84	9427	11	1280
B536R	223657	224421	78%	1	42	228	143	177	217	765	254	29145	9.5	23020
B537R	224436	224642	78%	3	34	57	28	43	79	207	68	7582	6.5	5180
b538L	224818	224618	62%	4	39	61	44	35	61	201	66	8078	6.1	2590
B539R	224639	225019	70%	2	39	118	70	78	115	381	126	14882	3.8	13400
B540R	225075	225323	82%	3	41	83	48	55	63	249	82	9410	10	30
B541L	225909	225457	78%	5	33	164	86	65	138	453	150	18220	6.8	10240
B542R	226129	227025	80%	1	37	339	155	181	222	897	298	34690	10	10330
B543L	227791	226982	66%	4	39	270	137	175	228	810	269	31137	8.3	21830
b544R	227430	227630	72%	3	40	55	41	40	65	201	66	7351	5.7	1340
b545R	227716	227922	54%	1	35	67	41	32	67	207	68	8178	8.4	3930
B546L	228364	228008	86%	4	48	82	93	77	105	357	118	12963	10	14000
B547R	228437	229018	82%	2	44	192	115	140	135	582	193	21770	8.1	21680
B548R	229041	229406	76%	3	41	142	59	90	75	366	121	13903	9.9	2620
B549R	229428	230264	76%	3	44	248	182	183	224	837	278	31746	6.1	21200
B550R	230410	231108	76%	1	44	215	145	164	175	699	232	26042	8	27400
b551L	230873	230562	62%	6	46	75	79	66	92	312	103	11189	5	8310
B552R	231143	231436	80%	2	41	113	57	64	60	294	97	11478	9.7	6970
B553R	231469	232032	88%	1	43	194	94	150	126	564	187	21799	3.1	25600
B554R	232061	233632	82%	2	48	498	482	280	312	1572	523	55629	11	15450
b555L	233373	233164	60%	5	41	54	42	44	70	210	69	7185	4.5	1280
B556R	233711	234835	88%	2	34	418	152	231	324	1125	374	44108	7.5	52350

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
b557R	234738	234977	78%	3	37	80	41	47	72	240	79	9000	10	5240
b558L	235011	234766	74%	5	37	70	46	45	85	246	81	9630	11	5210
B559L	235477	234845	92%	4	45	180	145	137	171	633	210	23551	8.9	19920
B560L	236307	235507	64%	5	39	277	147	168	209	801	266	29776	4.7	16530
b561R	235907	236275	58%	2	36	111	68	64	126	369	122	13675	11	2590
B562L	236642	236265	74%	6	35	125	69	63	121	378	125	14945	4.6	19060
b563L	236923	236717	54%	4	45	58	44	49	56	207	68	7928	12	12690
B564R	236732	237268	82%	2	46	148	128	119	142	537	178	20260	6.5	29250
b565R	236938	237183	62%	1	48	67	60	59	60	246	81	9562	12	18410
B566R	237261	238310	58%	3	38	342	172	226	310	1050	349	39452	8.4	19440
B567L	239412	238297	86%	5	35	377	179	212	348	1116	371	42587	8.7	14920
B568L	240166	239456	82%	4	41	235	167	124	185	711	236	26383	9	16500
b569L	240525	240127	68%	5	36	139	71	73	116	399	132	15170	10	13520
B570R	240239	240457	86%	2	44	52	49	47	71	219	72	8002	11	2560
B571L	241762	240458	78%	4	37	458	214	268	365	1305	434	49992	6.8	40950
b572R	241059	241331	82%	3	39	70	58	49	96	273	90	10142	12	90
B573L	242013	241786	84%	5	36	74	42	39	73	228	75	8627	7.2	1340
b574R	241925	242158	66%	2	33	79	39	38	78	234	77	9240	10	5150
B575R	242046	242342	88%	3	39	105	46	70	76	297	98	11059	9.8	20430
B576R	242579	243037	76%	2	35	167	78	84	130	459	152	17927	4.8	17270
b577L	243362	242955	66%	6	28	99	71	44	194	408	135	16596	7.6	14680
b578L	243175	242975	74%	4	30	52	36	25	88	201	66	7867	7.9	10360
B579R	243061	243315	82%	1	29	130	22	51	52	255	84	9919	11	45520
B580R	243363	243713	74%	3	35	139	68	55	89	351	116	13439	4.9	1340
B581L	244069	243710	80%	4	33	130	67	53	110	360	119	13721	6.5	8370
B582L	244350	244120	86%	5	48	66	54	56	55	231	76	8407	4.1	2560
B583L	245741	244380	86%	6	32	478	199	242	443	1362	453	53284	5.3	64670
b584L	246026	245796	56%	6	45	60	59	45	67	231	76	9523	9.6	14090
B585L	247180	245867	80%	4	54	321	425	288	280	1314	437	48356	7.4	63100
b586R	247244	247480	86%	2	41	74	47	49	67	237	78	9951	11	9050
B587R	247296	247763	76%	3	45	149	98	112	109	468	155	17171	8.5	9140
b588L	247502	247305	56%	6	45	44	45	44	65	198	65	7552	4.3	2650
b589R	247529	247729	60%	2	48	63	42	54	42	201	66	7964	12	10870
B590R	247745	247957	64%	2	54	48	60	56	49	213	70	7811	12	5750
B591L	248300	247989	78%	6	49	95	67	87	63	312	103	11195	11	1280
B592L	248563	248327	86%	4	39	83	42	50	62	237	78	9164	11	5180
B593R	248586	248927	76%	3	37	102	56	71	113	342	113	13308	7.4	37500
b594R	248630	248884	74%	2	41	71	44	60	80	255	84	10178	7.7	20370
B595L	249499	249086	70%	4	46	103	84	107	120	414	137	15535	4.4	19660
B596R	249630	250556	84%	3	38	318	165	186	258	927	308	35494	5.4	16710

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
B597L	250971	250654	76%	5	38	119	62	60	77	318	105	11898	5	1280
B598L	252214	251015	76%	4	45	390	247	290	273	1200	399	46693	6.5	76840
b599R	251031	251234	78%	3	41	44	45	39	76	204	67	7602	10	6430
b600L	251379	251140	58%	5	43	79	49	54	58	240	79	8992	12	7060
b601R	251179	251385	62%	1	45	53	47	46	61	207	68	8237	8	6490
B602L	253070	252426	74%	6	42	229	120	154	142	645	214	24183	9	43990
b603L	252727	252491	60%	4	42	84	45	54	54	237	78	8989	8.2	1400
b604L	252864	252640	58%	5	47	78	42	63	42	225	74	9184	12	39860
b605R	253105	253413	72%	1	38	91	69	47	102	309	102	11576	11	12090
B606L	254543	253155	92%	6	43	432	263	330	364	1389	462	52741	6.7	37040
b607R	253570	253776	70%	1	44	51	61	31	64	207	68	7451	9.5	2590
b608R	253773	253973	64%	3	44	55	43	46	57	201	66	7867	8.8	2650
b609R	254050	254247	56%	1	42	54	44	40	60	198	65	7220	12	2590
b610R	254593	254820	92%	1	39	62	50	39	77	228	75	8873	10	2590
B611L	254928	254608	84%	5	41	100	56	75	90	321	106	12540	11	7060
B612R	255088	255654	78%	1	38	208	109	109	141	567	188	22206	7.8	23500
b613L	255578	255321	68%	6	40	55	50	54	99	258	85	9528	9	2710
B614R	255923	256663	90%	2	29	299	87	127	228	741	246	29189	11	35620
b615L	256808	256599	60%	6	31	61	43	23	83	210	69	8332	8.7	10870
b616L	257115	256909	50%	5	46	55	55	41	56	207	68	8303	8.2	11560
B617L	258269	256956	84%	6	55	317	437	285	275	1314	437	48351	7.4	63100
B618R	258376	259188	92%	1	26	340	85	123	265	813	270	32680	5.6	44190
B619L	260121	259258	82%	5	41	299	195	157	213	864	287	31183	4.7	14740
b620R	259287	259514	78%	3	44	49	43	58	78	228	75	8505	5.5	1400
b621R	259760	260131	48%	2	36	96	61	73	142	372	123	13095	7	2680
b622L	260375	260133	62%	6	43	76	52	52	63	243	80	9625	11	15310
B623L	262114	260150	84%	4	42	623	378	439	525	1965	654	75532	5.6	99750
b624R	260319	260528	52%	3	48	54	50	50	56	210	69	7793	11	2590
b625R	260550	260747	62%	3	47	42	53	40	63	198	65	7053	12	12690
b626R	260597	260926	56%	2	42	89	77	60	104	330	109	13172	5.2	12270
b627L	260838	260629	68%	5	45	69	38	56	47	210	69	8276	11	24070
B628R	262453	263259	82%	1	41	240	150	179	238	807	268	30484	5.8	21050
B629R	263456	264319	82%	2	34	352	119	174	219	864	287	34604	9.6	32520
B630R	264343	264696	84%	1	38	122	67	66	99	354	117	13609	10	29250
B631L	265646	264708	90%	6	37	331	159	188	261	939	312	36747	6.8	37730
b632R	265053	265295	68%	3	39	65	50	45	83	243	80	9468	5.6	10930
B633R	265781	267112	80%	2	41	447	253	295	337	1332	443	50873	9	40410
b634L	267000	266800	60%	5	46	44	49	44	64	201	66	7021	11	1280
b635L	267178	266945	60%	4	35	62	40	43	89	234	77	9377	11	2650
B636R	267189	267791	76%	3	47	178	138	145	142	603	200	22682	4.5	19120

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
b637R	267598	267804	60%	1	45	68	50	44	45	207	68	8112	13	1370
B638R	267825	268346	82%	3	35	192	76	109	145	522	173	20872	9.6	24870
b639L	268357	268100	78%	4	36	56	53	40	109	258	85	9810	8.8	7060
b640L	268734	268315	60%	5	43	113	85	94	128	420	139	16228	8.3	18200
B641R	268408	269382	78%	1	44	293	211	215	256	975	324	37496	4.6	52780
b642L	269022	268771	70%	5	50	60	69	58	65	252	83	9644	5.2	9560
b643R	268842	269084	52%	3	45	70	51	59	63	243	80	9554	13	7000
B644L	269768	269490	78%	6	47	67	42	90	80	279	92	9800	11	26030
B645L	270471	269797	84%	5	37	246	102	147	180	675	224	25999	4.8	15040
b646R	270265	270519	58%	1	33	69	45	40	101	255	84	10270	9.9	12150
B647R	270545	271192	88%	2	41	223	136	128	161	648	215	24900	10	27490
b648L	271204	270692	68%	4	42	120	100	115	178	513	170	20337	4.3	41910
b649R	271189	271401	58%	1	39	53	46	37	77	213	70	8320	9.6	2620
B650L	271671	271204	86%	5	36	171	80	87	130	468	155	18686	9.8	23610
B651R	271706	272140	88%	2	41	161	85	94	95	435	144	17055	11	18490
B652R	272197	273150	82%	1	50	263	265	212	214	954	317	34876	5.5	18490
b653L	273025	272699	48%	4	57	59	85	102	81	327	108	12229	11	1370
b654L	273393	273154	52%	5	40	74	43	52	71	240	79	8995	11	3870
B655R	273188	273418	80%	2	45	59	51	52	69	231	76	8423	4.6	3840
B656L	273983	273390	70%	6	40	211	123	113	147	594	197	23264	7.2	16670
B657L	274583	274023	84%	6	36	205	94	109	153	561	186	22127	9.7	23700
b658L	274835	274638	62%	6	37	37	38	35	88	198	65	7891	9.5	12120
B659L	274938	274708	70%	5	38	42	47	40	102	231	76	8977	4	5210
b660R	274930	275160	66%	1	43	75	49	50	57	231	76	8838	11	5120
b661L	275246	275004	68%	6	48	54	59	57	73	243	80	9262	5.7	5750
B662R	275157	275585	56%	3	46	137	99	100	93	429	142	16227	7.8	13640
B663R	275618	276700	86%	2	38	421	198	213	251	1083	360	42233	8.3	39040
b664L	276274	276029	58%	4	47	47	65	51	83	246	81	9314	7.2	2770
b665L	277079	276645	58%	6	38	105	107	57	166	435	144	16445	12	3840
B666R	276730	277149	78%	1	39	150	59	104	107	420	139	15130	10	18440
B667L	278216	277206	86%	6	43	329	251	188	243	1011	336	36876	5.3	31420
b668R	277504	277911	72%	1	54	57	74	148	129	408	135	12577	7.9	2560
b669L	277873	277562	58%	4	58	96	128	53	35	312	103	12635	13	2560
B670L	278536	278249	86%	4	38	92	58	50	88	288	95	11119	9.1	12090
B671L	279393	278554	82%	5	35	318	136	161	225	840	279	32352	9.6	30730
B672L	280914	279457	70%	5	42	460	283	326	389	1458	485	55250	5.9	39490
b673R	279838	280053	54%	1	38	62	41	40	73	216	71	8678	7.8	9650
b674R	280183	280392	54%	1	41	67	41	45	57	210	69	8382	10	2620
b675R	280283	280762	48%	2	43	117	118	89	156	480	159	17659	11	2650
b676R	280793	281071	60%	2	29	99	41	41	98	279	92	10818	11	6400

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
B677R	281091	281480	72%	3	40	134	76	81	99	390	129	15212	7.9	17990
b678L	281578	281276	66%	4	36	82	52	58	111	303	100	11483	11	2620
b679L	281763	281491	68%	5	38	78	47	58	90	273	90	10290	11	12660
B680L	282061	281744	80%	4	33	123	51	55	89	318	105	12430	11	7680
B681L	283318	282266	88%	4	42	332	197	242	282	1053	350	39715	8	66180
b682R	282509	282757	62%	2	39	66	52	44	87	249	82	10101	9.3	13370
b683R	283064	283303	56%	2	49	64	66	51	59	240	79	8851	7.2	8310
B684L	283920	283375	80%	5	42	180	117	115	134	546	181	20208	11	17780
B685L	284553	283945	94%	5	42	182	105	150	172	609	202	22942	6.3	24300
b686R	284131	284337	58%	1	37	62	43	34	68	207	68	8387	8.7	5180
B687R	284606	285112	92%	2	48	144	118	127	118	507	168	18787	9.7	24780
B688L	286084	285119	72%	4	47	297	217	234	218	966	321	36794	8	38840
b689R	285314	285589	46%	2	51	57	72	69	78	276	91	10216	4.5	1370
b690R	285433	285654	46%	1	47	47	48	56	71	222	73	8922	12	14030
b691R	285722	286021	54%	2	42	67	76	50	107	300	99	12305	12	9590
B692R	286188	286628	72%	3	50	134	110	111	86	441	146	16276	7.3	20370
b693R	286618	286872	50%	1	38	94	45	51	65	255	84	9949	6	8370
B694R	286652	286939	72%	2	40	112	55	60	61	288	95	11329	11	6400
b695L	287587	287012	52%	4	43	155	143	107	171	576	191	20831	8.6	6730
B696R	287026	287253	74%	1	46	57	39	67	65	228	75	8206	9.4	17810
B697R	287217	288326	90%	3	43	320	218	257	315	1110	369	41929	7	58680
b698L	288061	287837	60%	4	39	71	43	44	67	225	74	8348	7.5	2590
B699L	288473	288237	82%	6	45	84	53	54	46	237	78	8751	7.7	1280
B700L	288718	288521	78%	4	43	55	46	39	58	198	65	7518	6.5	15220
B701L	288990	288751	84%	5	48	60	58	56	66	240	79	8693	10	8310
B702L	290554	289256	74%	4	44	435	289	277	298	1299	432	50007	11	55550
b703R	290057	290287	62%	2	40	52	43	50	86	231	76	8999	8.2	1460
B704L	290797	290588	88%	4	24	94	34	16	66	210	69	8737	10	19200
B705R	290861	292453	88%	2	45	473	372	349	399	1593	530	57738	4.5	62020
b706L	292254	292051	54%	5	43	50	45	43	66	204	67	7435	11	9560
b707R	292215	292508	56%	3	47	84	74	65	71	294	97	11404	12	2680
b708R	292258	292653	62%	1	46	103	95	87	111	396	131	15249	10	29280
b709L	292505	292272	58%	6	50	53	54	63	64	234	77	8526	7.3	1430
B710L	292665	292450	78%	5	39	76	38	47	55	216	71	8273	4.5	8310
B711L	294182	292995	66%	6	41	407	214	277	290	1188	395	46127	11	51110
b712R	293043	293303	70%	3	44	63	67	49	82	261	86	9914	11	5720
b713L	293503	293300	52%	4	37	78	34	41	51	204	67	8186	11	2620
B714L	294931	294131	78%	4	44	256	178	178	189	801	266	29811	11	12810
B715L	294926	294240	80%	6	45	217	157	155	158	687	228	25944	9.2	16620
b716R	294379	294660	60%	1	48	68	67	68	79	282	93	10236	11	7090

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
b717R	294627	294971	54%	3	42	84	65	79	117	345	114	13309	7.1	7090
B718L	295930	294968	78%	4	36	373	139	203	248	963	320	37234	7.2	47230
B719L	296234	295989	74%	6	42	77	52	52	65	246	81	9171	11	9530
B720L	297005	296259	86%	6	43	271	150	168	158	747	248	28196	9.2	7830
b721R	296554	296796	46%	1	50	47	61	61	74	243	80	8609	12	7000
b722R	297145	297354	74%	1	42	49	51	37	73	210	69	8053	11	3900
B723L	297952	297173	82%	4	39	279	136	166	199	780	259	30655	6.2	25670
B724R	298088	298609	80%	2	42	176	104	113	129	522	173	19988	10	12860
B725L	302168	298593	74%	6	48	1041	732	967	836	3576	1191	121561	6.4	85160
b726R	298958	299227	56%	2	47	60	71	55	84	270	89	10442	12	9560
b727R	299509	299775	54%	1	43	71	74	42	80	267	88	9959	9.4	1280
b728R	300382	300612	44%	1	49	52	63	51	65	231	76	8528	4.9	2560
b729L	300654	300424	58%	5	50	62	52	64	53	231	76	9355	8.9	21030
b730R	300703	300975	58%	1	51	58	79	59	77	273	90	9783	4.3	1280
b731L	301192	300992	48%	4	47	49	42	53	57	201	66	7723	12	6970
b732L	301942	301712	54%	4	54	58	46	79	48	231	76	9237	12	12720
b733R	302090	302290	52%	2	37	56	37	38	70	201	66	7667	6.8	5210
B734R	302287	303258	60%	1	46	287	213	232	240	972	323	37261	7.3	26860
b735L	303135	302806	66%	5	48	75	96	62	97	330	109	12999	9.6	5690
B736L	304417	303200	84%	4	42	363	219	292	344	1218	405	45548	6.5	47920
b737R	304424	304693	58%	2	41	63	50	62	95	270	89	10169	4.4	120
B738L	305824	304448	78%	4	44	417	295	311	354	1377	458	53263	8	50250
b739R	304868	305149	56%	2	44	64	67	57	94	282	93	10972	6.6	1340
b740R	305522	305758	58%	2	49	62	58	59	58	237	78	9197	7.1	1280
B741L	306328	305903	68%	4	47	104	94	106	122	426	141	14860	4.7	7680
b742L	306626	306411	58%	6	29	73	34	29	80	216	71	8896	9.5	2650
B743R	306447	307400	86%	3	38	329	189	174	262	954	317	36298	9.4	17360
B744L	308889	307375	80%	5	40	465	292	316	442	1515	504	59721	7	80410
b745R	307583	307798	58%	2	40	63	53	33	67	216	71	7777	11	1310
b746R	308440	308667	56%	1	45	69	44	59	56	228	75	8596	9.4	8340
B747L	309892	308963	86%	4	30	383	119	157	271	930	309	37315	9.6	52590
B748L	311175	309973	82%	5	41	357	250	243	353	1203	400	45365	5.2	42760
b749L	310397	310170	68%	6	41	71	48	45	64	228	75	8705	11	8280
b750R	310396	310734	60%	1	40	102	61	76	100	339	112	14019	11	24900
B751L	311864	311229	76%	6	42	220	123	146	147	636	211	23212	10	10840
b752R	311491	311919	50%	1	41	105	89	88	147	429	142	15790	8.3	3990
B753L	313024	311882	80%	4	43	386	225	266	266	1143	380	43546	8.9	72280
b754R	311928	312419	64%	3	43	118	117	96	161	492	163	18549	10	14110
b755R	312549	312842	58%	3	43	62	65	62	105	294	97	10793	7.3	6550
B756L	315055	313106	96%	4	43	683	380	468	419	1950	649	71423	10	96240

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
b757R	313468	313773	64%	1	45	79	85	54	88	306	101	12257	9.5	19210
b758L	314268	314017	70%	5	52	74	76	56	46	252	83	9613	12	1310
B759R	315098	317125	94%	2	44	670	410	484	464	2028	675	73891	7.5	112300
b760L	316659	316249	62%	5	49	89	122	79	121	411	136	15664	7.6	21030
B761L	317559	317122	82%	5	34	166	74	74	124	438	145	16640	10	10840
B762L	318099	317560	72%	5	31	191	70	99	180	540	179	21599	5.7	33120
B763L	318503	318123	80%	6	40	116	60	93	112	381	126	14495	5.1	8730
b764L	319044	318472	64%	5	38	143	116	102	212	573	190	21974	11	16680
B765R	318578	318925	82%	2	43	141	69	79	59	348	115	12976	12	0
B766R	318937	319482	72%	1	40	186	103	118	139	546	181	20765	7.1	29310
B767L	320220	319483	80%	5	40	245	139	157	197	738	245	27913	4.2	12860
b768R	319808	320032	68%	2	40	56	46	45	78	225	74	8006	6.7	2560
B769R	320318	321406	84%	2	43	334	220	248	287	1089	362	41126	7.8	47920
b770L	320661	320422	62%	5	43	64	55	49	72	240	79	8678	11	1310
B771L	321914	321408	90%	6	40	168	96	108	135	507	168	19006	8.6	22960
B772L	322374	321973	84%	5	43	150	79	92	81	402	133	15328	11	28650
b773R	321985	322185	84%	1	40	39	45	36	81	201	66	7958	9	14060
B774R	322209	323246	58%	3	40	319	211	206	302	1038	345	40185	8.5	45920
b775R	322472	322717	60%	2	42	91	53	51	51	246	81	10044	12	8280
b776R	323143	323379	62%	1	41	70	49	48	70	237	78	9172	12	5210
B777L	324310	323249	78%	4	46	318	233	258	253	1062	353	40155	7.7	45960
b778R	323613	323954	54%	3	47	82	84	77	99	342	113	12461	10	5240
b779L	324315	323845	78%	5	42	155	84	114	118	471	156	18061	11	2680
b780R	323873	324250	62%	2	42	97	91	69	121	378	125	14862	8.7	31190
B781L	327552	324367	76%	5	46	959	672	779	776	3186	1061	120398	8.6	124160
b782R	324577	324861	72%	1	42	74	60	59	92	285	94	11086	11	30
b783R	324644	324940	58%	2	42	81	66	59	91	297	98	11188	12	3960
b784R	324877	325545	64%	1	46	160	177	134	198	669	222	25861	7.9	22340
b785R	325757	326059	50%	2	42	88	66	62	87	303	100	11569	11	8280
b786R	326026	326223	64%	1	48	48	43	52	55	198	65	7511	7.5	90
b787R	326189	326431	58%	2	52	48	69	58	68	243	80	8691	12	1280
b788R	326701	327339	50%	1	45	156	158	130	195	639	212	24493	6.1	19480
b789R	326918	327244	52%	2	46	78	78	71	100	327	108	12115	11	3930
b790L	327290	327060	68%	6	42	81	45	53	52	231	76	8809	11	1310
B791R	327785	327994	82%	2	36	63	43	33	71	210	69	7856	4.3	10840
B792R	328107	328853	76%	3	48	207	211	144	185	747	248	27555	6.2	10810
b793L	328864	328544	72%	4	50	78	52	109	82	321	106	10789	7.3	13970
b794R	328771	328992	58%	1	37	77	43	39	63	222	73	8486	12	1340
B795R	328889	329317	72%	2	38	157	69	95	108	429	142	16027	7.7	20490
B796L	330418	329318	90%	4	40	338	207	238	318	1101	366	41990	7.4	40890

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
b797R	329840	330394	64%	2	41	150	119	108	178	555	184	21831	5.3	21150
B798R	330531	331472	84%	3	38	356	157	197	232	942	313	36634	8.1	53760
b799R	331312	331545	62%	1	38	86	38	52	58	234	77	8586	11	1340
B800R	331566	331838	84%	3	42	87	59	57	70	273	90	10180	10	7060
B801L	332243	331839	72%	6	40	140	81	82	102	405	134	15467	4.7	11520
B802R	332353	332682	70%	1	42	100	63	75	92	330	109	13103	6.1	23500
B803L	333220	332873	80%	4	39	112	70	65	101	348	115	13079	10	1370
b804L	333525	333316	68%	5	33	35	36	34	105	210	69	8637	9.6	9560
B805R	333334	334356	76%	1	40	385	183	227	228	1023	340	40539	6.2	38230
b806L	333814	333371	54%	4	36	82	88	70	204	444	147	17494	6.5	21290
b807R	333596	333802	74%	2	37	86	32	44	45	207	68	7963	11	12150
b808L	334041	333766	70%	5	47	69	73	57	77	276	91	10549	5.2	8340
b809L	334305	334060	62%	5	43	63	60	45	78	246	81	9346	7.9	2680
b810L	335041	334358	58%	4	38	222	136	121	205	684	227	25361	11	9050
B811R	334668	334865	40%	3	27	61	28	26	83	198	65	7884	11	2560
B812R	334878	336053	82%	3	37	394	206	227	349	1176	391	45702	6.1	35430
B813L	336508	336149	64%	4	43	112	64	90	94	360	119	13440	9.3	19120
b814L	336462	336256	60%	5	44	59	37	55	56	207	68	8086	10	9650
b815R	336299	336532	52%	2	40	72	59	35	68	234	77	9426	11	10840
B816L	338000	336522	82%	6	41	511	324	289	355	1479	492	55745	11	42230
b817R	337681	337956	74%	1	42	71	60	57	88	276	91	9909	8.4	2650
B818R	338069	339022	76%	2	35	345	148	183	278	954	317	36799	9.9	21110
B819L	339378	339025	70%	5	39	129	64	73	88	354	117	13543	3.9	3840
B820L	340051	339413	88%	4	41	245	112	147	135	639	212	24651	4.4	17810
b821R	339599	339853	62%	2	42	44	67	40	104	255	84	10005	9.8	2620
b822R	339622	339864	58%	1	42	43	63	39	98	243	80	9857	11	9050
B823L	340341	340090	84%	5	35	74	46	41	91	252	83	9640	9.4	16560
B824L	340719	340366	88%	5	38	102	65	71	116	354	117	12938	10	23500
B825L	342338	340776	82%	6	45	443	359	343	418	1563	520	58194	5.8	95530
B826L	342592	342389	74%	4	40	69	37	45	53	204	67	7582	8.7	180
b827R	342463	342708	60%	1	34	89	44	39	74	246	81	9983	11	9050
B828R	342623	342988	78%	2	36	122	61	71	112	366	121	13556	10	24750
B829R	343057	344355	74%	1	43	431	269	288	311	1299	432	49945	11	55550
b830L	343739	343515	48%	6	45	56	52	49	68	225	74	8395	10	15340
B831R	344451	345758	70%	3	42	431	270	275	332	1308	435	48557	11	46640
B832R	345824	349135	84%	2	44	1006	683	775	848	3312	1103	124554	8.2	132050
b833R	346107	346331	46%	3	38	77	48	37	63	225	74	8494	11	7000
b834R	346425	346679	48%	3	47	63	58	63	71	255	84	9367	10	1370
b835L	346759	346529	60%	4	49	61	66	47	57	231	76	8644	9.5	7000
b836R	347103	347372	66%	3	41	82	52	60	76	270	89	10546	11	12660

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
b837L	348868	348413	50%	4	50	99	109	119	129	456	151	17572	4.9	16650
b838R	348898	349293	54%	1	44	121	84	89	102	396	131	15300	8.4	28030
B839R	349163	349525	74%	2	40	111	72	74	106	363	120	13217	8.4	5150
B840L	349933	349526	74%	4	41	121	74	92	121	408	135	15756	8.5	33120
B841R	349976	350233	80%	2	43	73	48	62	75	258	85	9901	9.8	20400
B842R	350288	350503	88%	2	32	85	33	37	61	216	71	8172	5.9	2560
B843R	350574	350999	86%	3	39	134	85	81	126	426	141	16475	9.2	9020
B844R	351043	352122	84%	1	42	355	219	238	268	1080	359	40961	5.5	61000
b845R	351293	351532	58%	2	43	73	43	60	64	240	79	8873	12	8250
b846L	351821	351597	62%	6	38	57	45	41	82	225	74	8316	11	10870
B847R	352141	353601	70%	1	43	550	326	309	276	1461	486	55508	12	37100
b848L	352899	352657	62%	5	47	47	54	60	82	243	80	9154	11	9590
B849R	353640	354155	88%	3	43	160	114	109	133	516	171	19007	7.3	6430
B850L	355333	354158	74%	4	34	455	159	241	321	1176	391	44881	10	56190
b851R	355248	355454	74%	3	31	60	42	23	82	207	68	8214	10	6430
B852R	355481	355852	86%	2	34	144	59	66	103	372	123	14341	7.3	16080
B853L	356456	355863	82%	6	40	194	123	112	165	594	197	22618	6.5	16820
b854R	355888	356208	66%	1	40	90	66	62	103	321	106	12116	7.9	10330
b855L	356272	356057	64%	4	40	69	46	41	60	216	71	8136	11	1400
b856R	356289	356534	60%	3	37	68	38	52	88	246	81	9957	11	4020
B857L	356897	356487	86%	6	38	148	77	79	107	411	136	15548	4.6	14680
B858L	357602	357048	82%	6	46	162	106	148	139	555	184	20644	3.5	17270
B859L	358136	357621	78%	6	38	152	102	95	167	516	171	19590	11	17840
B860L	358655	358191	84%	6	41	142	95	94	134	465	154	17455	6.4	6400
b861L	358591	358394	60%	4	42	59	47	36	56	198	65	7676	8.7	6460
B862L	359162	358707	82%	6	39	155	85	94	122	456	151	17198	6.9	5120
b863R	358777	359001	56%	1	38	64	47	39	75	225	74	8273	6.4	2620
b864R	358998	359288	66%	3	36	85	53	52	101	291	96	11534	8.1	23620
B865R	359447	360097	80%	2	43	211	149	133	158	651	216	24862	8	37500
b866R	360033	360353	58%	3	48	72	82	73	94	321	106	12384	12	14030
B867L	360849	360094	80%	5	44	261	151	181	163	756	251	28389	9.4	12920
b868R	360109	360753	56%	1	45	133	159	134	219	645	214	25349	8.7	26180
B869R	361094	362188	70%	2	39	387	193	229	286	1095	364	41459	11	46010
b870L	361566	361165	70%	5	41	101	81	83	137	402	133	14343	6	2710
B871R	362347	362844	76%	1	38	158	77	114	149	498	165	18577	4.5	14800
b872L	363026	362793	58%	6	29	97	42	27	68	234	77	9057	10	9020
b873R	362956	363207	70%	1	38	74	43	52	83	252	83	9940	7.9	16590
B874L	364117	363017	84%	4	43	322	216	255	308	1101	366	40136	5.9	32140
b875R	363278	363751	56%	2	45	126	117	97	134	474	157	17693	6.6	6910
b876R	363453	363683	64%	3	46	60	56	50	65	231	76	8272	11	11440

NY-2A ORF	ORF Location		% A+T (<50 ntds from start codon)	Frame	% G+C (gene)	# of Each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
b877R	364224	364451	74%	3	40	50	49	42	87	228	75	8722	4.4	180
B878L	365432	364230	74%	6	44	411	241	286	265	1203	400	47079	6.9	73770
b879R	364671	365132	58%	3	45	103	112	98	149	462	153	18132	9.7	8430
B880L	366205	365498	72%	4	46	194	148	178	188	708	235	25794	5.7	10510
b881R	365789	366217	66%	2	46	108	106	90	125	429	142	15911	5.9	5570
b882R	366333	366581	64%	3	51	54	65	62	68	249	82	8911	5	1310
B883L	366677	366369	68%	6	47	85	69	75	80	309	102	11765	6.9	5750
B884L	367168	366932	52%	4	44	52	50	55	80	237	78	8908	4.6	5240
b885L	367548	367324	46%	5	37	67	41	43	74	225	74	8632	7.5	8310
B886R	367486	367767	62%	1	45	70	59	69	84	282	93	10442	4.7	2860

2.6.2.2 Analysis of Open Reading Frames Against the Non-redundant Database

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
B1L	No Hit Found			
b2R	No Hit Found			
B3R	No Hit Found			
B4R	identical to Chlorella virus PBCV-1 terminal repeat ORF A3R, corresponds to GenBank Accession Number M55319 [Paramecium bursaria Chlorella virus 1]	2.7E-23	NP_049045	61%
b5L	identical to PBCV-1 terminal repeat ORF, corresponds to GenBank Accession Number M55319 [Paramecium bursaria Chlorella virus 1]	1.4E-19	NP_049046	64%
B6L	No Hit Found			
b7R	No Hit Found			
B8R	putative DNA methylase [Escherichia coli]	5.1E-33	CAD33713	34%
	putative DNA methylase [Escherichia coli]	2.5E-32	CAD33712	31%
	C-5 cytosine-specific DNA methylase [Clostridium thermocellum ATCC 27405]	6.9E-30	ZP_00504119	34%
b9L	No Hit Found			
B10R	DNA methylase N-4/N-6 [Crocospaera watsonii WH 8501]	2.7E-62	ZP_00515061	42%
	unknown [Shigella flexneri bacteriophage V]	4.6E-22	AAL89450	30%
	putative DNA adenine methyltransferase encoded by prophage CP-933O [Escherichia coli O157:H7 EDL933]	1.9E-20	AAG56134	31%
b11R	No Hit Found			
b12L	No Hit Found			
B13L	hypothetical protein PSSM4_047 [Cyanophage P-SSM4]	1.7E-07	YP_214608	32%
	hypothetical protein [Thermoplasma acidophilum]	1.6E-05	CAC11914	28%
b14R	No Hit Found			
b15R	No Hit Found			
B16L	DNA adenine methyltransferase [Chlorella virus NY2A]	0.0E+00	AAC03124	96%
	site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) CviBIII - Chlorella virus CV-NC1A	2.0E-161	CAA29835	74%
	DNA adenine methyltransferase [Chlorella virus SC-1A]	3.3E-100	AAC57945	50%
b17R	No Hit Found			
B18L	NACHT domain protein, putative [Aspergillus fumigatus Af293]	5.8E-10	EAL87814	35%
	hypothetical protein [Homo sapiens]	1.5E-34	CAI56716	30%
	ankyrin 3 isoform 1 [Homo sapiens]	4.6E-15	NP_066267	27%
b19R	No Hit Found			
b20R	No Hit Found			
b21R	No Hit Found			
B22R	A354R [Paramecium bursaria Chlorella virus 1]	6.3E-31	NP_048711	39%
	endonuclease [Tetrahymena thermophila]	3.8E-04	AAL73456	31%
b23L	No Hit Found			
B24L	NACHT domain protein, putative [Aspergillus fumigatus Af293]	4.1E-23	EAL87814	28%
	SecG [Dictyostelium discoideum]	2.7E-14	XP_637214	32%
	hypothetical protein AN1130.2 [Aspergillus nidulans FGSC A4]	8.6E-13	EAA66248	29%
b25R	No Hit Found			
b26R	No Hit Found			
B27L	No Hit Found			
b28L	No Hit Found			
b29R	No Hit Found			
b30L	No Hit Found			
B31R	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
b32L	No Hit Found			
B33R	A9R [Paramecium bursaria Chlorella virus 1]	4.1E-84	NP_048357	82%
	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	4.4E-22	NP_048807	32%
	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	9.9E-22	NP_049005	33%
B34L	Vp260 like protein [Chlorella virus]	0.0E+00	BAB83467	80%
	Vp260 like protein [Chlorella virus]	0.0E+00	BAB83469	34%
	Vp260 like protein [Chlorella virus]	0.0E+00	BAB83468	35%
b35L	No Hit Found			
b36R	No Hit Found			
b37R	No Hit Found			
b38R	No Hit Found			
B39L	No Hit Found			
B40L	Vp260 like protein [Chlorella virus]	0.0E+00	BAB83468	76%
	Vp260 like protein [Chlorella virus]	0.0E+00	BAB83469	72%
	Vp260 like protein [Chlorella virus]	0.0E+00	BAB83470	65%
b41R	No Hit Found			
b42R	No Hit Found			
b43R	No Hit Found			
b44R	No Hit Found			
b45R	No Hit Found			
b46R	No Hit Found			
B47L	Vp260 like protein [Chlorella virus]	0.0E+00	BAB83469	62%
	Vp260 like protein [Chlorella virus]	0.0E+00	BAB83468	62%
	Vp260 like protein [Chlorella virus]	0.0E+00	BAB83470	63%
b48L	No Hit Found			
b49R	No Hit Found			
b50R	No Hit Found			
b51R	hepatitis A virus cellular receptor 1 long form [Cercopithecus aethiops]	5.5E-04	AAC39774	34%
	hepatitis A virus cellular receptor 1 short form [Cercopithecus aethiops]	5.5E-04	AAC39773	34%
	hepatitis A virus cellular receptor 1 long form [Cercopithecus aethiops]	7.2E-04	AAC39772	29%
b52R	No Hit Found			
b53R	No Hit Found			
b54R	No Hit Found			
B55L	Vp260 like protein [Chlorella virus]	0.0E+00	BAB83471	82%
	Vp260 like protein [Chlorella virus]	6.6E-174	BAB83467	39%
	Vp260 like protein [Chlorella virus]	8.7E-174	BAB83468	40%
b56R	No Hit Found			
b57R	No Hit Found			
b58R	No Hit Found			
B59R	putative capsid protein [Chlorella virus]	1.6E-117	AAC27494	83%
	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052 [Paramecium bursaria Chlorella virus 1]	2.5E-75	NP_048787	43%
	major capsid protein MCP1 [Chlorella virus]	7.4E-75	BAA76601	43%
b60L	No Hit Found			
B61R	A354R [Paramecium bursaria Chlorella virus 1]	2.9E-36	NP_048711	40%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081 [Paramecium bursaria Chlorella virus 1]	5.9E-13	NP_048779	29%
b62L	No Hit Found			
b63L	No Hit Found			
b64L	No Hit Found			
b65R	No Hit Found			
b66L	No Hit Found			
b67L	A37L [Paramecium bursaria Chlorella virus 1]	6.2E-44	NP_048385	77%
B68L	contains ATP/GTP-binding motif A; similar to Dictyostelium FP21 glycoprotein, corresponds to Swiss-Prot Accession Number P52285 [Paramecium bursaria Chlorella virus 1]	4.8E-41	NP_048387	55%
	SKP1 [Nicotiana tabacum]	2.3E-19	AAT99735	37%
	SKP1-like protein [Nicotiana clevelandii]	5.2E-19	AAC63273	36%
b69L	contains Gly-rich motifs GAGLGTGF (5X); similar to Arabidopsis Gly-rich protein, corresponds to Swiss-Prot Accession Number P27843 [Paramecium bursaria Chlorella virus 1]	6.0E-07	NP_048388	77%
B70R	contains Pro-rich Px motif, PAPK (8X); similar to Thermoproteus virus protein TPX, corresponds to Swiss-Prot Accession Number P19275 [Paramecium bursaria Chlorella virus 1]	2.9E-08	NP_048389	93%
	similar to PBCV-1 ORF A41R, corresponds to GenBank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	3.5E-139	NP_048519	79%
	P0481E12.18 [Oryza sativa (japonica cultivar-group)]	2.8E-11	NP_916095	27%
b71R	a43R [Paramecium bursaria Chlorella virus 1]	6.1E-39	NP_048391	93%
b72L	a172L [Paramecium bursaria Chlorella virus 1]	5.0E-17	NP_048520	51%
B73L	contains ATP/GTP-binding site motif A; contains DDC/GAD/HDC/TyrDC pyridoxal phosphate attachment site; similar to yeast BCS1 protein, corresponds to Swiss-Prot Accession Number P32839 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048392	89%
	hypothetical protein, conserved [Trypanosoma cruzi]	2.2E-20	EAN88268	25%
	hypothetical protein, conserved [Trypanosoma cruzi]	3.8E-20	EAN97512	24%
B74R	A48R [Paramecium bursaria Chlorella virus 1]	1.0E-57	NP_048396	90%
B75L	similar to Escherichia coli glycerophosphoryl diester phosphodiesterase, corresponds to Swiss-Prot Accession Number P10908 [Paramecium bursaria Chlorella virus 1]	9.5E-116	NP_048397	94%
	226aa long hypothetical glycerophosphoryl diester phosphodiesterase [Sulfolobus tokodaii str. 7]	7.7E-25	BAB65669	34%
	glycerophosphoryl diester phosphodiesterase [Bacillus clausii KSM-K16]	1.2E-20	YP_174990	34%
B76L	pyrimidine dimer-specific glycosylase [Chlorella virus NYs-1]	6.6E-62	AAD33382	99%
	PBCV-1 pyrimidine dimer-specific glycosylase [Paramecium bursaria Chlorella virus 1] >gi 624073 gb AAC96418.1 PBCV-1 pyrimidine dimer-specific glycosylase [Paramecium bursaria Chlorella virus 1]	4.3E-53	NP_048398	88%
	pyrimidine dimer-specific glycosylase [Chlorella virus NE-8A]	1E-52	AAD33375	87%
B77R	No Hit Found			
B78L	contains type I hydrophobic transmembrane region and ATP/GTP binding motif [Paramecium bursaria Chlorella virus 1]	1.2E-76	NP_048399	91%
	hypothetical protein Bcep02006353 [Burkholderia fungorum LB400]	7.2E-37	ZP_00279033	48%
	conserved hypothetical protein [Chromobacterium violaceum ATCC 12472]	4.1E-32	AAQ58722	48%
b79R	a52R [Paramecium bursaria Chlorella virus 1]	4.1E-32	NP_048400	74%
B80L	putative transposase [Chlorella virus]	5.2E-55	AAU06281	31%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
	similar to Synechocystis transposase, corresponds to GenBank Accession Number D90909 [Paramecium bursaria Chlorella virus 1]	8.9E-55	NP_048981	31%
	putative transposase [Acanthamoeba polyphaga mimivirus]	5.4E-12	YP_143208	24%
b81R	No Hit Found			
b82L	No Hit Found			
B83L	putative resolvase [Acanthamoeba polyphaga mimivirus]	3.6E-24	YP_143125	41%
	putative resolvase [Acanthamoeba polyphaga mimivirus]	4.0E-23	YP_142434	39%
	putative resolvase [Acanthamoeba polyphaga mimivirus]	1.3E-21	YP_142457	37%
b84R	No Hit Found			
b85R	No Hit Found			
B86L	No Hit Found			
B87L	No Hit Found			
B88L	M.CviAII cytosine DNA methyltransferase [Paramecium bursaria Chlorella virus 1]	1.4E-86	NP_048873	50%
	M.CviAIV cytosine DNA methyltransferase [Paramecium bursaria Chlorella virus 1]	1.3E-76	NP_048886	48%
	cytosine methyltransferase [Paramecium bursaria Chlorella virus NYs-1]	1.6E-74	AAC64006	45%
b89R	No Hit Found			
b90R	a518R [Paramecium bursaria Chlorella virus 1]	6.0E-15	NP_048874	62%
B91L	No Hit Found			
b92R	No Hit Found			
b93R	No Hit Found			
B94R	No Hit Found			
b95L	No Hit Found			
b96L	No Hit Found			
b97L	No Hit Found			
b98R	No Hit Found			
B99R	contains Pro-rich Px motif EPSPEPxP (5X), and PEST sequence; similar to trypanosome procyclin precursor, corresponds to Swiss-Prot Accession Number P08469 [Paramecium bursaria Chlorella virus 1]	5.9E-52	NP_048415	85%
b100L	No Hit Found			
b101L	No Hit Found			
b102L	contains Gly-rich Gx motif LGGGLG (5X); contains type I hydrophobic transmembrane sequence; contains protein splicing signature; similar to Arabidopsis Gly-rich protein, corresponds to Swiss-Prot Accession Number P27483 [Paramecium bursaria Chlorella virus 1]	2.5E-22	NP_048417	60%
B103R	similar to Mycoplasma hypothetical protein MG366, corresponds to Swiss-Prot Accession Number P47606 [Paramecium bursaria Chlorella virus 1]	8.9E-172	NP_048419	82%
	A63L [Paramecium bursaria Chlorella virus 1]	9.2E-36	NP_048411	36%
	COG1196: Chromosome segregation ATPases [Xylella fastidiosa Dixon]	8.1E-16	ZP_00039080	41%
B104R	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	2.7E-134	NP_048920	73%
	A154L [Paramecium bursaria Chlorella virus 1]	8.0E-134	NP_048502	70%
	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903 [Paramecium bursaria Chlorella virus 1]	1.8E-125	NP_048477	68%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
b105L	a155R [Paramecium bursaria Chlorella virus 1]	1.0E-16	NP_048503	60%
b106R	No Hit Found			
b107L	A75L [Paramecium bursaria Chlorella virus 1]	2.1E-146	NP_048423	87%
B108L	No Hit Found			
B109R	No Hit Found			
B110L	hypothetical protein [Plasmodium falciparum 3D7]	1.9E-04	NP_704578	29%
	unnamed protein product [Podospora anserina]	3.3E-04	CAD60740	27%
b111R	No Hit Found			
b112R	No Hit Found			
B113R	No Hit Found			
b114L	a76L [Paramecium bursaria Chlorella virus 1]	2.2E-09	NP_048424	81%
B115L	contains type 1 hydrophobic transmembrane region [Paramecium bursaria Chlorella virus 1]	1.3E-17	NP_048425	70%
B116R	contains ATP/GTP-binding site motif A; similar to rat beta-alanine synthetase, corresponds to Swiss-Prot Accession Number Q03248 [Paramecium bursaria Chlorella virus 1]	1.5E-163	NP_048426	93%
	Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase [Azotobacter vinelandii AvOP]	8.3E-77	ZP_00417184	48%
	putative carbon-nitrogen hydrolase [Yersinia pseudotuberculosis IP 32953]	1.1E-76	YP_071711	50%
B117L	encodes Asp/Lys rich sequence [Paramecium bursaria Chlorella virus 1]	3.3E-07	NP_048628	32%
	contains phenyl group binding site (CAAX box) [Paramecium bursaria Chlorella virus 1]	5.3E-05	NP_048438	40%
B118R	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	8.1E-47	NP_049003	46%
	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	3.3E-16	NP_049005	65%
	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	4.8E-15	NP_048807	68%
b119L	No Hit Found			
B120R	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	3.2E-107	NP_048629	77%
	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	8.7E-105	NP_048807	77%
	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.3E-103	NP_049005	75%
b121L	a80L [Paramecium bursaria Chlorella virus 1]	2.1E-15	NP_048428	56%
	similar to PBCV-1 ORF a80L, encoded by GenBank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	3.9E-06	NP_048808	55%
B122L	A81L [Paramecium bursaria Chlorella virus 1]	1.9E-75	NP_048429	71%
b123R	No Hit Found			
B124L	A84L [Paramecium bursaria Chlorella virus 1]	1.6E-87	NP_048432	82%
b125R	No Hit Found			
B126R	PBCV-1 prolyl 4-hydroxylase [Paramecium bursaria Chlorella virus 1]	8.0E-94	NP_048433	77%
	hypothetical protein Raeut03003807 [Ralstonia eutropha JMP134]	1.9E-15	ZP_00351069	35%
	prolyl 4-hydroxylase alpha subunit [Bacillus cereus G9241]	2.2E-14	ZP_00238502	29%
b127L	No Hit Found			
b128L	a86L [Paramecium bursaria Chlorella virus 1]	1.0E-22	NP_048434	85%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
b129L	No Hit Found			
B130R	A138R [Paramecium bursaria Chlorella virus 1]	8.7E-05	NP_048486	24%
b131R	No Hit Found			
b132L	No Hit Found			
B133R	A87R [Paramecium bursaria Chlorella virus 1]	8.8E-162	NP_048435	65%
	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081 [Paramecium bursaria Chlorella virus 1]	1.2E-30	NP_048779	34%
	A354R [Paramecium bursaria Chlorella virus 1]	1.1E-05	NP_048711	25%
b134L	No Hit Found			
b135L	No Hit Found			
B136L	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563 [Paramecium bursaria Chlorella virus 1]	1.5E-150	NP_048441	94%
	hypothetical protein A93L - Chlorella virus PBCV-1	1.2E-99	T17583	92%
	a91L [Paramecium bursaria Chlorella virus 1]	5.5E-57	NP_048439	85%
B137L	1,3(4)-beta-glucanase [uncultured murine large bowel bacterium BAC 14]	1.9E-37	AAX16367	39%
	beta-1,3-glucanase A1 precursor - Bacillus circulans >gi 142973 gb AAA22474.1	6.1E-36	JQ0420	38%
	endo-beta-1,3-glucanase [Bacillus clausii KSM-K16]	1.4E-35	YP_174203	38%
b138R	No Hit Found			
B139R	chitin synthase [Chlorella virus]	4.2E-78	BAB83509	38%
	hypothetical protein FG06550.1 [Gibberella zeae PH-1]	8.3E-42	EAA78335	27%
	hypothetical protein FG10619.1 [Gibberella zeae PH-1]	3.2E-41	EAA68628	27%
b140L	No Hit Found			
b141L	No Hit Found			
b142L	No Hit Found			
B143R	PBCV-1 glucosamine synthetase [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048448	85%
	glutamine:fructose-6-phosphate amidotransferase GFAT [Chlorella virus]	0.0E+00	BAD15299	85%
	glucosamine--fructose-6-phosphate aminotransferase [isomerizing] [Bordetella parapertussis 12822]	2.0E-154	NP_886345	49%
b144L	a101L [Paramecium bursaria Chlorella virus 1]	1.3E-14	NP_048449	65%
b145L	No Hit Found			
b146L	No Hit Found			
b147L	a102L [Paramecium bursaria Chlorella virus 1]	2.7E-34	NP_048450	64%
B148R	PBCV-1 mRNA guanylyltransferase [Paramecium bursaria Chlorella virus 1]	4.8E-150	NP_048451	76%
	Chain B, Structure Of Guanylylated Mrna Capping Enzyme Complexed With Gtp	2.4E-149	1CKN_B	76%
	mRNA capping enzyme family protein [Arabidopsis thaliana]	1.1E-16	NP_974263	27%
b149L	a104L [Paramecium bursaria Chlorella virus 1]	8.4E-17	NP_048452	43%
B150L	contains ubiquitin carboxy-terminal hydrolase active sites; similar to human ubiquitin carboxy-terminal hydrolase, corresponds to Swiss-Prot Accession Number Q09879 [Paramecium bursaria Chlorella virus 1]	1.1E-123	NP_048453	68%
	ENSANGP0000018711 [Anopheles gambiae str. PEST]	3.4E-11	XP_312004	24%
	CG14619-PE, isoform E [Drosophila melanogaster]	9.8E-11	AAN09565	27%
b151R	a106R [Paramecium bursaria Chlorella virus 1]	1.4E-19	NP_048454	48%
b152R	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
b153R	No Hit Found			
B154L	similar to Pyrococcus woesei factor TFIIB homolog, corresponds to GenBank Accession Number X70668 [Paramecium bursaria Chlorella virus 1]	1.2E-114	NP_048455	66%
	putative transcription initiation factor IIB [Acanthamoeba polyphaga mimivirus]	2.0E-08	YP_142604	25%
	transcription initiation factor TFIIB Sua7p; ZnR+2cyclins [Cryptosporidium parvum]	2.6E-08	EAK89316	27%
b155R	a108R [Paramecium bursaria Chlorella virus 1]	3.7E-09	NP_048456	42%
b156R	No Hit Found			
B157L	A109L [Paramecium bursaria Chlorella virus 1]	1.3E-45	NP_048457	84%
b158R	No Hit Found			
B159R	A114R [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048462	87%
	A111R [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048459	94%
	hypothetical protein [Desulfotalea psychrophila LSv54]	4.5E-20	CAG34747	32%
b160L	a112L [Paramecium bursaria Chlorella virus 1]	3.8E-33	NP_048460	97%
b161R	a116R [Paramecium bursaria Chlorella virus 1]	1.1E-24	NP_048464	74%
b162L	a117L [Paramecium bursaria Chlorella virus 1]	6.2E-44	NP_048465	63%
B163R	GDP-D-mannose 4,6 dehydratase [Paramecium bursaria Chlorella virus 1]	6.7E-177	AAO67555	88%
	GDP-mannose-4,6-dehydratase [Vibrio vulnificus YJ016]	1.2E-109	NP_933142	59%
	GDP-D-mannose dehydratase [Yersinia pseudotuberculosis IP 32953]	1.6E-109	YP_069550	59%
b164L	No Hit Found			
B165R	A87R [Paramecium bursaria Chlorella virus 1]	1.2E-161	NP_048435	62%
	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081 [Paramecium bursaria Chlorella virus 1]	9.9E-31	NP_048779	29%
	HNH endonuclease [Acanthamoeba polyphaga mimivirus]	3.5E-04	YP_142599	37%
b166L	No Hit Found			
b167R	No Hit Found			
B168R	A121R [Paramecium bursaria Chlorella virus 1]	1.3E-41	NP_048469	75%
	Tlr 6Fp protein [Tetrahymena thermophila]	7.0E-16	AAL73467	44%
	Tlr 6Fp protein [Tetrahymena thermophila]	2.0E-15	AAL73477	43%
b169L	No Hit Found			
B170R	PBCV-1 Vp260 protein [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048470	55%
	glycoprotein Vp260	1.2E-14	AAA86307	31%
	contains a leucine zipper motif; similar to Bacillus subtilis phage P2A preneck appendage protein, corresponds to Swiss-Prot Accession Number P07537 [Paramecium bursaria Chlorella virus 1]	2.2E-144	NP_048471	78%
b171L	No Hit Found			
b172L	No Hit Found			
B173L	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081 [Paramecium bursaria Chlorella virus 1]	5.4E-09	NP_048779	27%
	A87R [Paramecium bursaria Chlorella virus 1]	1.5E-06	NP_048435	27%
	putative HNH homing endonuclease [bacteriophage bIL170]	1.9E-06	AAC27227	30%
b174L	a124L [Paramecium bursaria Chlorella virus 1]	7.3E-29	NP_048473	73%
B175L	contains a zinc ribbon domain; similar to Chlorella virus CVU1 TFIIS-like transcription factor, corresponds to GenBank Accession Number D29631 [Paramecium bursaria Chlorella virus 1]	2.6E-88	NP_048472	91%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
	transcription elongation factor SII [Chlorella virus]	3.4E-88	BAA04187	91%
	transcription elongation factor TFIIS homolog - Chlorella virus CV-U1	6.4E-87	S47662	90%
b176R	No Hit Found			
B177R	A127R [Paramecium bursaria Chlorella virus 1]	8.1E-116	NP_048475	84%
b178L	a128L [Paramecium bursaria Chlorella virus 1]	5.4E-24	NP_048476	81%
B179L	No Hit Found			
b180L	No Hit Found			
B181L	A131L [Paramecium bursaria Chlorella virus 1]	2.1E-52	NP_048479	70%
b182R	a132R [Paramecium bursaria Chlorella virus 1]	1.4E-11	NP_048480	44%
B183L	hypothetical protein [Caulobacter crescentus CB15]	4.4E-17	AAK23092	31%
	Methyltransferase FkbM [Crocospphaera watsonii WH 8501]	2.8E-11	ZP_00517616	28%
	hypothetical protein glr0593 [Gloeobacter violaceus PCC 7421]	4.0E-10	NP_923539	26%
b184R	similar to Synechocystis 3-pyrophosphohydrolase, corresponds to GenBank Accession Number D90911 [Paramecium bursaria Chlorella virus 1]	1.0E-108	NP_048481	95%
	unknown [Arabidopsis thaliana]	2.9E-23	AAM64943	32%
	expressed protein [Arabidopsis thaliana]	2.9E-23	AAD20906	33%
B185L	similar to bacteriophage T4 intron-associated endonuclease, corresponds to Swiss-Prot Accession Number P13299 [Paramecium bursaria Chlorella virus 1]	2.2E-67	NP_048482	76%
	A315L [Paramecium bursaria Chlorella virus 1]	1.5E-07	NP_048671	34%
	PBCV-1 33kd peptide [Paramecium bursaria Chlorella virus 1]	1.1E-05	NP_048641	35%
b186R	No Hit Found			
B187R	a136R [Paramecium bursaria Chlorella virus 1]	2.1E-60	NP_048484	81%
B188R	A137R [Paramecium bursaria Chlorella virus 1]	9.4E-16	NP_048485	66%
b189L	No Hit Found			
B190L	A139L [Paramecium bursaria Chlorella virus 1]	4.4E-34	NP_048487	76%
b191L	No Hit Found			
B192R	PBCV-1 surface protein [Paramecium bursaria Chlorella virus 1]	6.2E-91	NP_048488	51%
	surface protein [Chlorella virus CVK2]	1.1E-87	BAD12236	50%
	surface protein [Paramecium bursaria Chlorella virus 1]	1.5E-89	BAD22850	51%
b193L	a144L [Paramecium bursaria Chlorella virus 1]	3.7E-10	NP_048492	43%
	hypothetical protein a141L - Chlorella virus PBCV-1	3.7E-10	T17632	43%
b194L	a144L [Paramecium bursaria Chlorella virus 1]	4.4E-24	NP_048492	45%
	hypothetical protein a144L - Chlorella virus PBCV-1	4.4E-24	T17635	45%
	hypothetical protein a141L - Chlorella virus PBCV-1	5.5E-06	T17632	30%
b195L	Phe-, Gly-rich protein: RCGF 3X, GCGF 11X, RSGF 5X, GSGF 2X [Paramecium bursaria Chlorella virus 1]	3.2E-24	NP_048495	71%
	glycine tyrosine-rich protein a147L - Chlorella virus PBCV-1	3.2E-24	T17638	71%
	a144L [Paramecium bursaria Chlorella virus 1]	1.5E-05	NP_048492	49%
b196L	No Hit Found			
B197L	A150L [Paramecium bursaria Chlorella virus 1]	2.0E-42	NP_048498	80%
b198R	No Hit Found			
B199R	A354R [Paramecium bursaria Chlorella virus 1]	3.5E-24	NP_048711	36%
	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081 [Paramecium bursaria Chlorella virus 1]	1.8E-12	NP_048779	29%
	A87R [Paramecium bursaria Chlorella virus 1]	3.3E-06	NP_048435	32%
b200L	No Hit Found			
B201R	A151R [Paramecium bursaria Chlorella virus 1]	4.2E-61	NP_048499	94%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
	A148R [Paramecium bursaria Chlorella virus 1]	3.2E-08	NP_048496	36%
b202L	contains prokaryotic membrane lipoprotein lipid attachment site [Paramecium bursaria Chlorella virus 1]	4.2E-16	NP_048500	70%
B203R	similar to phage T5 helicase, corresponds to Swiss-Prot Accession Number P11107 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048501	78%
	VV A18 helicase [Acanthamoeba polyphaga mimivirus]	4.7E-49	YP_142750	32%
	EsV-1-66 [Ectocarpus siliculosus virus]	3.8E-46	AAK14489	33%
b204L	No Hit Found			
b205R	No Hit Found			
B206L	A315L [Paramecium bursaria Chlorella virus 1]	6.1E-55	NP_048671	43%
	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	6.1E-55	NP_049007	43%
	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440 [Paramecium bursaria Chlorella virus 1]	5.8E-37	NP_048851	65%
B207L	A157L [Paramecium bursaria Chlorella virus 1]	1.2E-39	NP_048505	78%
B208L	A158L [Paramecium bursaria Chlorella virus 1]	1.5E-16	NP_048506	55%
B209R	A161R [Paramecium bursaria Chlorella virus 1]	1.2E-10	NP_048509	51%
b210R	No Hit Found			
B211L	A165L [Paramecium bursaria Chlorella virus 1]	2.0E-42	NP_048513	74%
	ORF23/BdrA [Borrelia burgdorferi]	8.1E-04	AAF29795	29%
	hypothetical protein DDB0219478 [Dictyostelium discoideum]	8.1E-04	XP_636056	35%
B212L	A165L [Paramecium bursaria Chlorella virus 1]	1.6E-36	NP_048513	65%
b213R	No Hit Found			
B214R	PBCV-1 exonuclease [Paramecium bursaria Chlorella virus 1]	1.7E-131	NP_048514	79%
	unknown protein [Arabidopsis thaliana]	6.2E-17	AAO64106	34%
	F12A21.19 [Arabidopsis thaliana]	6.2E-17	AAG28903	34%
B215R	A168R [Paramecium bursaria Chlorella virus 1]	2.4E-64	NP_048516	72%
b216L	No Hit Found			
b217R	No Hit Found			
B218R	A354R [Paramecium bursaria Chlorella virus 1]	2.9E-36	NP_048711	39%
	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081 [Paramecium bursaria Chlorella virus 1]	4.7E-10	NP_048779	25%
b219L	No Hit Found			
b220L	No Hit Found			
b221L	No Hit Found			
B222R	PBCV-1 aspartate transcarbamylase [Paramecium bursaria Chlorella virus 1]	1.2E-156	NP_048517	85%
	aspartate carbamoyltransferase [Solanum tuberosum]	2.7E-76	CAC85728	49%
	Aspartate carbamoyltransferase 3, chloroplast precursor (Aspartate transcarbamylase 3) (ATCase 3)	3.0E-75	AAB67857	51%
b223L	No Hit Found			
B224R	similar to PBCV-1 ORF A41R, corresponds to GenBank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	4.3E-172	NP_048519	82%
	contains Pro-rich Px motif, PAPK (8X); similar to Thermoproteus virus protein TPX, corresponds to Swiss-Prot Accession Number P19275 [Paramecium bursaria Chlorella virus 1]	9.9E-145	NP_048389	82%
	PBCV-1 surface protein [Paramecium bursaria Chlorella virus 1]	2.1E-09	NP_048488	54%
b225R	a43R [Paramecium bursaria Chlorella virus 1]	5.6E-08	NP_048391	44%
B226L	similar to E. coli hypothetical protein, corresponds to Swiss-Prot Accession Number P39407 [Paramecium bursaria Chlorella virus 1]	4.2E-144	NP_048521	90%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
	Patatin-like phospholipase family [Bacillus cereus G9241]	1.0E-17	ZP_00240206	33%
	phospholipase, patatin family [Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough]	5.1E-17	YP_009106	34%
b227L	No Hit Found			
b228R	No Hit Found			
b229L	No Hit Found			
B230L	Modification methylase CviRI (Adenine-specific methyltransferase CviRI) (M.CviRI)	1.9E-175	P52284	79%
	site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) CviBIII - Chlorella virus CV-NC1A	1.0E-54	CAA29835	34%
	DNA adenine methyltransferase [Chlorella virus SC-1A]	3E-49	AAC57945	31%
b231R	No Hit Found			
b232R	No Hit Found			
b233R	No Hit Found			
b234R	No Hit Found			
B235L	A687R [Paramecium bursaria Chlorella virus 1]	1.4E-24	NP_049043	71%
B236L	DNA adenine methyltransferase [Chlorella virus NY2A]	1.9E-173	AAC03127	94%
	M.CviAII adenine DNA methyltransferase [Paramecium bursaria Chlorella virus 1]	4.3E-138	NP_048600	77%
	site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) CviAII - Chlorella virus PBCV-1	1.6E-137	S27901	76%
b237R	No Hit Found			
b238R	a251aL [Paramecium bursaria Chlorella virus 1]	3.7E-12	NP_048601	71%
B239R	vChti-1 [Chlorella virus]	0.0E+00	BAA78554	83%
	PBCV-1 chitinase [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048529	82%
	putative sugar hydrolase [Streptomyces avermitilis MA-4680]	8.1E-40	BAC72964	38%
b240L	No Hit Found			
b241L	No Hit Found			
b242L	No Hit Found			
b243R	No Hit Found			
b244L	No Hit Found			
b245L	No Hit Found			
B246R	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.5E-55	NP_049007	55%
	A315L [Paramecium bursaria Chlorella virus 1]	7.3E-55	NP_048671	55%
	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440 [Paramecium bursaria Chlorella virus 1]	2.4E-42	NP_048851	55%
b247L	No Hit Found			
b248L	No Hit Found			
B249R	DNA-directed DNA polymerase (EC 2.7.7.7) - Chlorella virus CV-NY-2A	0.0E+00	P30320	94%
	DNA polymerase [Chlorella virus K2]	0.0E+00	BAA35142	83%
	PBVC-1 DNA polymerase [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048532	83%
b250R	No Hit Found			
b251L	No Hit Found			
b252L	a188L [Paramecium bursaria Chlorella virus 1]	5.1E-22	NP_048535	83%
B253R	PBVC-1 DNA polymerase [Paramecium bursaria Chlorella virus 1]	6.7E-70	NP_048532	86%
	DNA polymerase [Chlorella virus K2]	7.4E-69	BAA35142	84%
	B-family DNA polymerase [Heterosigma akashiwo virus 01]	6E-15	BAE06251	34%
b254L	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
B255R	No Hit Found			
B256R	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440 [Paramecium bursaria Chlorella virus 1]	4.0E-05	NP_048851	31%
b257R	No Hit Found			
B258R	similar to SWI/SNF chromatin remodeling complex subunit OSA2 [Homo sapiens], corresponds to accession number 22597106 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048536	59%
	proline-rich protein A189R - Chlorella virus PBCV-1	4.6E-16	T17679	30%
	hypothetical protein A192R - Chlorella virus PBCV-1	7.8E-56	T17682	48%
b259L	No Hit Found			
b260R	No Hit Found			
B261L	similar to human PCNA, corresponds to Swiss-Prot Accession Number P12004 [Paramecium bursaria Chlorella virus 1]	2.2E-144	NP_048540	97%
	PREDICTED: similar to ENSANGP0000012272 [Apis mellifera]	8.3E-35	XP_395519	30%
	PREDICTED: proliferating cell nuclear antigen [Pan troglodytes]	2.4E-34	XP_514499	29%
b262R	No Hit Found			
b263R	a194R [Paramecium bursaria Chlorella virus 1]	1.3E-25	NP_048541	80%
B264L	A196L [Paramecium bursaria Chlorella virus 1]	1.7E-70	NP_048543	79%
b265R	a198R [Paramecium bursaria Chlorella virus 1]	3.9E-14	NP_048545	47%
b266R	a197R [Paramecium bursaria Chlorella virus 1]	1.6E-18	NP_048544	77%
B267R	A199R [Paramecium bursaria Chlorella virus 1]	1.0E-30	NP_048546	84%
B268L	zinc finger (MYND type) family protein / SET domain-containing protein [Arabidopsis thaliana]	6.4E-07	NP_849969	42%
	putative SET-domain transcriptional regulator [Arabidopsis thaliana]	6.4E-07	AAD03568	42%
	unnamed protein product [Tetraodon nigroviridis]	1.9E-06	CAG07000	37%
b269R	No Hit Found			
b270L	No Hit Found			
B271R	contains cytidine and deoxycytidine deaminase Zn-binding region signature [Paramecium bursaria Chlorella virus 1]	1.1E-48	NP_048547	75%
	FirrV-1-A29 [Feldmannia irregularis virus a]	6.7E-06	AAR26853	29%
	related to cytidine and deoxycytidylate deaminase family protein [Desulfotalea psychrophila LSv54]	1.6E-04	CAG36205	32%
B272L	A201L [Paramecium bursaria Chlorella virus 1]	2.6E-26	NP_048548	63%
B273L	A202L [Paramecium bursaria Chlorella virus 1]	8.0E-52	NP_048549	85%
B274R	A203R [Paramecium bursaria Chlorella virus 1]	6.0E-93	NP_048550	92%
b275L	a204L [Paramecium bursaria Chlorella virus 1]	1.1E-27	NP_048551	91%
b276L	No Hit Found			
B277R	A205R [Paramecium bursaria Chlorella virus 1]	2.0E-42	NP_048552	53%
B278R	PBCV-1 arginine decarboxylase [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048554	86%
	ornithine decarboxylase [Trypanosoma brucei gambiense]	1.3E-72	AAD02222	41%
	Ornithine decarboxylase (ODC)	1.3E-72	P07805	41%
b279L	No Hit Found			
b280L	No Hit Found			
b281L	No Hit Found			
b282L	No Hit Found			
B283R	KAEKA (6X), SDDD (7X) [Paramecium bursaria Chlorella virus 1]	2.5E-12	NP_048555	32%
B284L	A213L [Paramecium bursaria Chlorella virus 1]	1.0E-65	NP_048560	80%
B285L	A214L [Paramecium bursaria Chlorella virus 1]	9.1E-43	NP_048561	79%
B286R	hypothetical protein PF10_0099 [Plasmodium falciparum 3D7]	5.7E-04	NP_700573	26%
b287L	No Hit Found			
B288L	PBCV-1 alginate lyase [Paramecium bursaria Chlorella virus 1]	4.1E	NP_048562	67%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
		111		
	vAL-1 [Chlorella virus]	7.8E-110	BAB19127	67%
	alginate lyase [Chlorella virus]	1.0E-109	BAA83789	66%
B289L	similar to bovine monoamine oxidase, corresponds to Swiss-Prot Accession Number P21398 [Paramecium bursaria Chlorella virus 1]	9.6E-175	NP_048564	84%
	orf1 [Chlorella virus]	1.6E-174	BAA83788	84%
b290R	No Hit Found			
B291L	A227L [Paramecium bursaria Chlorella virus 1]	6.8E-51	NP_048575	78%
	MobD.6 hypothetical protein [Enterobacteria phage T4]	2.8E-04	AAD42598	31%
B292L	A229L [Paramecium bursaria Chlorella virus 1]	7.3E-24	NP_048577	68%
b293R	A230R [Paramecium bursaria Chlorella virus 1]	2.4E-83	NP_048578	79%
b294R	No Hit Found			
b295L	No Hit Found			
B296L	No Hit Found			
B297R	No Hit Found			
B298L	contains ATP/GTP-binding motif A [Paramecium bursaria Chlorella virus 1]	1.4E-167	NP_048579	76%
	similar to Chlorella virus PBCV-1 ORF A231L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	2.1E-81	NP_048983	45%
b299R	No Hit Found			
b300R	a232R [Paramecium bursaria Chlorella virus 1]	2.2E-12	NP_048580	58%
b301L	No Hit Found			
B302R	A233R [Paramecium bursaria Chlorella virus 1]	3.1E-43	NP_048581	73%
B303L	A234L [Paramecium bursaria Chlorella virus 1]	3.9E-38	NP_048582	66%
b304R	a235R [Paramecium bursaria Chlorella virus 1]	1.8E-11	NP_048583	46%
B305R	PBCV-1 homospermidine synthase [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048585	89%
	Homospermidine synthase [Pelodictyon phaeoclathratiforme BU-1]	1.2E-83	ZP_00589757	39%
	homospermidine synthase [Methanosarcina acetivorans str. C2A]	9.8E-75	AAM05046	35%
b306L	No Hit Found			
b307R	No Hit Found			
b308L	No Hit Found			
b309L	No Hit Found			
B310R	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	2.2E-41	NP_048920	54%
	A154L [Paramecium bursaria Chlorella virus 1]	1.4E-45	NP_048502	52%
	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903 [Paramecium bursaria Chlorella virus 1]	9.8E-127	NP_048477	60%
b311L	No Hit Found			
b312L	No Hit Found			
b313L	a155R [Paramecium bursaria Chlorella virus 1]	7.8E-15	NP_048503	52%
B314L	A239L [Paramecium bursaria Chlorella virus 1]	2.8E-28	NP_048587	54%
b315R	No Hit Found			
B316R	contains ATP-GTP binding motif; similar to <i>Saccharomyces cerevisiae</i> antiviral protein SKI2, corresponds to Swiss-Prot Accession Number P35207 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048589	92%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
	ATP-DEPENDENT RNA HELICASE (SKI2 FAMILY) [Encephalitozoon cuniculi GB-M1]	3.2E-85	CAD25317	33%
	Mtr4p like SKI family SFII helicase [Cryptosporidium parvum]	7.2E-85	EAK89441	38%
b317L	a242L [Paramecium bursaria Chlorella virus 1]	6.9E-43	NP_048590	74%
b318L	No Hit Found			
b319L	No Hit Found			
b320L	No Hit Found			
b321L	No Hit Found			
B322R	Pro-rich, PAPK (20X); similar to Arabidopsis anter-specific Pro-rich protein, corresponds to Swiss-Prot Accession Number P40602 [Paramecium bursaria Chlorella virus 1]	6.8E-85	NP_048594	96%
	hevein-like protein precursor [Arabidopsis thaliana]	5.0E-11	AAF04912	37%
	PR-4 type protein [Vitis vinifera]	1.9E-10	AAC33732	38%
B323R	4 ankyrin repeats; similar to Drosophila melanogaster ankyrin, corresponds to GenBank Accession Number L35601 [Paramecium bursaria Chlorella virus 1]	1.2E-121	NP_048596	61%
	Guanosine polyphosphate pyrophosphohydrolases/synthetases homolog [Rickettsia felis URRWXCal2]	1.4E-19	YP_246397	31%
	GA14074-PA [Drosophila pseudoobscura]	4.7E-04	EAL29245	22%
B324L	A87R [Paramecium bursaria Chlorella virus 1]	6.4E-148	NP_048435	61%
	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081 [Paramecium bursaria Chlorella virus 1]	1.0E-28	NP_048779	33%
	HNH endonuclease [Acanthamoeba polyphaga mimivirus]	2.7E-05	YP_142599	24%
b325R	No Hit Found			
b326R	No Hit Found			
b327L	No Hit Found			
b328R	No Hit Found			
b329R	No Hit Found			
b330R	No Hit Found			
B331R	protein kinase A248R [Chlorella virus]	2.8E-171	AAU06280	96%
	protein kinase A248R [Chlorella virus]	9.9E-169	AAU06282	94%
	protein kinase A248R [Chlorella virus]	3.5E-158	AAU06275	95%
b332L	No Hit Found			
b333L	a249L [Paramecium bursaria Chlorella virus 1]	3.8E-04	NP_048598	65%
B334R	putative transposase [Chlorella virus]	0.0E+00	AAU06281	99%
	similar to Synechocystis transposase, corresponds to GenBank Accession Number D90909 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048981	85%
	putative transposase [Acanthamoeba polyphaga mimivirus]	7.0E-23	YP_142458	30%
b335R	No Hit Found			
B336R	potassium channel protein [Chlorella virus MA-1D]	7.7E-47	AAQ16138	95%
	potassium channel protein [Chlorella virus IL-5-2s1]	3.8E-46	AAQ16141	94%
	PBCV-1 K+ ion channel protein	9E-45	NP_048599	90%
b337L	No Hit Found			
b338R	No Hit Found			
B339L	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	1.8E-113	NP_048629	81%
	A79R [Paramecium bursaria Chlorella virus 1]	2.9E-108	NP_048427	83%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	2.2E-100	NP_048807	74%
b340R	a648L [Paramecium bursaria Chlorella virus 1]	3.5E-07	NP_049004	47%
B341R	hypothetical protein [Oenothera berteriana]	6.4E-04	CAA45895	37%
	A253R [Paramecium bursaria Chlorella virus 1]	6.9E-56	NP_048604	70%
B342R	hypothetical protein A253R [Chlorella virus]	5.8E-55	AAU06294	69%
	hypothetical protein A253R [Chlorella virus]	1.0E-54	AAU06289	69%
b343L	a252aL [Paramecium bursaria Chlorella virus 1]	1.3E-17	NP_048603	66%
B344R	A261R [Paramecium bursaria Chlorella virus 1]	7.4E-55	NP_048615	93%
b345R	No Hit Found			
	PBCV-1 33kd peptide [Paramecium bursaria Chlorella virus 1]	1.0E-68	NP_048641	56%
	A315L [Paramecium bursaria Chlorella virus 1]	1.2E-43	NP_048671	45%
B346L	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.6E-40	NP_049007	41%
b347R	No Hit Found			
B348L	A262L [Paramecium bursaria Chlorella virus 1]	4.2E-46	NP_048616	83%
	A263L [Paramecium bursaria Chlorella virus 1]	2.9E-15	NP_048617	53%
B349R	a264R [Paramecium bursaria Chlorella virus 1]	3.3E-21	NP_048618	68%
B350L	A265L [Paramecium bursaria Chlorella virus 1]	8.0E-69	NP_048619	72%
b351R	No Hit Found			
b352R	a266R [Paramecium bursaria Chlorella virus 1]	2.8E-07	NP_048620	47%
b353R	No Hit Found			
	A271L [Paramecium bursaria Chlorella virus 1]	5.9E-65	NP_048625	74%
B354L	lipase, putative [Thermotoga maritima MSB8]	3.8E-11	NP_229151	25%
	putative lipase [Arabidopsis thaliana]	1.9E-10	AAK93696	26%
b355R	a270R [Paramecium bursaria Chlorella virus 1]	5.6E-16	NP_048624	64%
b356R	a272R [Paramecium bursaria Chlorella virus 1]	1.3E-11	NP_048626	52%
	A273L [Paramecium bursaria Chlorella virus 1]	8.8E-62	NP_048627	86%
B357L	unknown [Acanthamoeba polyphaga mimivirus]	5.2E-22	YP_142507	44%
	Protein of unknown function DUF305 [Mesorhizobium sp. BNC1]	1.1E-06	EAN05440	33%
B358R	encodes Asp/Lys rich sequence [Paramecium bursaria Chlorella virus 1]	4.9E-56	NP_048628	60%
	hypothetical protein SalaDRAFT_0836 [Sphingopyxis alaskensis RB2256]	6.8E-40	ZP_00579458	35%
B359R	DNA adenine methyltransferase [Chlorella virus NY2A]	2.2E-22	AAC03123	35%
	putative TYPE II DNA MODIFICATION ENZYME (METHYLTRANSFERASE) [Helicobacter pylori J99]	2.1E-17	NP_223729	34%
b360L	No Hit Found			
B361R	CviPII top-strand DNA nicking endonuclease [Chlorella virus]	1.2E-33	AAV84098	37%
	Type II restriction enzyme CviJI (Endonuclease CviJI) (R.CviJI)	5.3E-13	P52283	30%
b362L	No Hit Found			
b363L	No Hit Found			
b364R	a281R [Paramecium bursaria Chlorella virus 1]	3.1E-06	NP_048635	51%
	similar to bovine cylycin I, corresponds to Swiss-Prot Accession Number P35662 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048632	92%
B365L	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048636	58%
	RPQT-like (9x) [Paramecium bursaria Chlorella virus 1]	8.7E-89	NP_048970	50%
b366R	a279R [Paramecium bursaria Chlorella virus 1]	4.6E-39	NP_048633	97%
b367R	a280R [Paramecium bursaria Chlorella virus 1]	2.3E-43	NP_048634	94%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
B368L	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048636	89%
	similar to bovine cyligin I, corresponds to Swiss-Prot Accession Number P35662 [Paramecium bursaria Chlorella virus 1]	3.4E-173	NP_048632	68%
	RPQT-like (9x) [Paramecium bursaria Chlorella virus 1]	9.3E-78	NP_048970	42%
b369L	a283L [Paramecium bursaria Chlorella virus 1]	6.1E-31	NP_048637	92%
B370L	A354R [Paramecium bursaria Chlorella virus 1]	3.4E-75	NP_048711	58%
	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081 [Paramecium bursaria Chlorella virus 1]	5.4E-12	NP_048779	28%
	putative endodeoxyribonuclease [Bacteriophage r1t]	3.9E-10	NP_695069	34%
B371L	PBCV-1 amidase [Paramecium bursaria Chlorella virus 1]	1.3E-137	NP_048638	86%
	Choloylglycine hydrolase [Bacillus licheniformis ATCC 14580]	2.5E-32	YP_081289	34%
	choloylglycine hydrolase family protein [Brucella abortus biovar 1 str. 9-941]	1.3E-31	YP_222152	36%
b372R	No Hit Found			
b373R	a285R [Paramecium bursaria Chlorella virus 1]	9.2E-24	NP_048639	91%
b374L	Chain A, 4ank: A Designed Ankyrin Repeat Protein With Four Identical Consensus Repeats	1.0E-06	1N0R_A	43%
	ankyrin repeat protein mbp3_5 [synthetic construct]	1.0E-06	AAQ93811	34%
	NACHT domain protein, putative [Aspergillus fumigatus Af293]	8.2E-04	EAL87814	32%
b375L	No Hit Found			
b376L	No Hit Found			
B377R	No Hit Found			
B378L	putative transposase [Chlorella virus]	5.2E-55	AAU06281	31%
	similar to Synechocystis transposase, corresponds to GenBank Accession Number D90909 [Paramecium bursaria Chlorella virus 1]	8.9E-55	NP_048981	31%
	putative transposase [Acanthamoeba polyphaga mimivirus]	5.4E-12	YP_143208	24%
b379R	No Hit Found			
b380L	No Hit Found			
B381L	putative resolvase [Acanthamoeba polyphaga mimivirus]	3.9E-24	YP_143125	41%
	putative resolvase [Acanthamoeba polyphaga mimivirus]	4.3E-23	YP_142434	39%
	putative resolvase [Acanthamoeba polyphaga mimivirus]	1.4E-21	YP_142457	37%
b382R	No Hit Found			
b383R	No Hit Found			
b384L	No Hit Found			
B385R	A286R [Paramecium bursaria Chlorella virus 1]	8.2E-163	NP_048640	72%
b386L	No Hit Found			
b387L	No Hit Found			
B388L	similar to PBCV-1 serine/threonine protein kinase, corresponds to GenBank Accession Number U14660 [Paramecium bursaria Chlorella virus 1]	1.6E-102	NP_048643	68%
	protein kinase A248R [Chlorella virus]	7.5E-31	AAU06282	34%
	similar to PBCV-1 serine/threonine protein kinase, corresponds to GenBank Accession Number U14660 [Paramecium bursaria Chlorella virus 1]	9.8E-31	NP_048631	33%
B389R	PBCV-1 33kd peptide [Paramecium bursaria Chlorella virus 1]	4.7E-77	NP_048641	57%
	A315L [Paramecium bursaria Chlorella virus 1]	9.5E-54	NP_048671	45%
	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	4.3E-38	NP_049007	40%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
b390L	ORF245 [Chlorella virus]	3.0E-22	BAA20344	53%
b391L	No Hit Found			
b392L	No Hit Found			
B393L	PBCV-1 chitosanase [Paramecium bursaria Chlorella virus 1]	5.2E-152	NP_048646	80%
	vChta-1 [Chlorella virus]	3.2E-149	BAA20342	78%
	Chitosanase precursor	8.7E-14	P48846	29%
b394R	a293R [Paramecium bursaria Chlorella virus 1]	1.3E-41	NP_048647	69%
B395L	GDP-4-keto-6-deoxy-D-mannose epimerase/reductase [Paramecium bursaria Chlorella virus 1]	2.4E-157	AAO67556	84%
	PBCV-1 fucose synthase [Paramecium bursaria Chlorella virus 1]	2.0E-156	NP_048649	84%
	NAD-dependent epimerase/dehydratase [Methanosarcina barkeri str. fusaro]	9.9E-103	ZP_00544248	62%
b396R	No Hit Found			
B397L	Asp-rich [Paramecium bursaria Chlorella virus 1]	1.1E-29	NP_048725	27%
b398R	No Hit Found			
B399R	DNA adenine methyltransferase [Chlorella virus NY2A]	2.5E-148	AAC03126	96%
	Modification methylase CviBI (Adenine-specific methyltransferase CviBI) (M.CviBI)	7.5E-137	Q01511	85%
	M1.BstSEI [Geobacillus stearothermophilus]	5.5E-47	AAX89133	43%
b400L	No Hit Found			
B401L	No Hit Found			
B402L	No Hit Found			
B403R	A328L [Paramecium bursaria Chlorella virus 1]	1.7E-20	NP_048684	29%
B404R	A296R [Paramecium bursaria Chlorella virus 1]	1.2E-31	NP_048650	77%
b405L	No Hit Found			
b406L	No Hit Found			
b407R	a156L [Paramecium bursaria Chlorella virus 1]	1.9E-05	NP_048504	45%
B408R	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	7.5E-65	NP_048920	61%
	A154L [Paramecium bursaria Chlorella virus 1]	1.4E-31	NP_048502	59%
	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903 [Paramecium bursaria Chlorella virus 1]	5.4E-47	NP_048477	52%
b409R	No Hit Found			
b410L	a155R [Paramecium bursaria Chlorella virus 1]	1.5E-24	NP_048503	62%
B411L	M.CviAII cytosine DNA methyltransferase [Paramecium bursaria Chlorella virus 1]	3.7E-83	NP_048873	44%
	cytosine methyltransferase [Paramecium bursaria Chlorella virus NYs-1]	1.7E-80	AAC64006	45%
	Modification methylase CviJI (Cytosine-specific methyltransferase CviJI) (M.CviJI)	7.3E-79	P36216	43%
b412R	No Hit Found			
b413R	No Hit Found			
b414R	a518R [Paramecium bursaria Chlorella virus 1]	2.8E-04	NP_048874	43%
b415L	No Hit Found			
B416R	DNA adenine methyltransferase [Chlorella virus SC-1A]	5.1E-88	AAC57943	44%
	DNA adenine methyltransferase [Chlorella virus NY2A]	1.6E-49	AAC03124	31%
	DNA adenine methyltransferase [Chlorella virus SC-1A]	3.8E-43	AAC57945	28%
b417L	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
B418R	DNA methyltrasferase [Helicobacter pylori]	9.1E-72	CAB92310	42%
	M.Hpy188I [Helicobacter pylori]	4.5E-71	AAF74028	42%
	SNF2 family protein [Streptococcus agalactiae 2603V/R]	1.2E-10	AAN00153	35%
B419L	A297L [Paramecium bursaria Chlorella virus 1]	2.3E-80	NP_048651	81%
b420L	No Hit Found			
B421L	A298L [Paramecium bursaria Chlorella virus 1]	6.6E-101	NP_048652	77%
	hypothetical protein Bcen2424DRAFT_1981 [Burkholderia cenocepacia HI2424]	4.3E-07	ZP_00463717	27%
	hypothetical protein Bcepao3002567 [Burkholderia cepacia R18194]	5.6E-07	ZP_00216135	27%
b422L	No Hit Found			
b423R	No Hit Found			
B424L	A354R [Paramecium bursaria Chlorella virus 1]	1.3E-69	NP_048711	53%
	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081 [Paramecium bursaria Chlorella virus 1]	5.6E-09	NP_048779	28%
	A87R [Paramecium bursaria Chlorella virus 1]	1.6E-08	NP_048435	25%
b425R	No Hit Found			
B426L	A301L [Paramecium bursaria Chlorella virus 1]	9.8E-53	NP_048655	62%
b427L	No Hit Found			
b428L	a303L [Paramecium bursaria Chlorella virus 1]	3.9E-06	NP_048657	44%
B429R	A304R [Paramecium bursaria Chlorella virus 1]	1.9E-16	NP_048658	60%
B430L	similar to human protein Tyr-phosphatase, corresponds to GenBank Accession Number U27193 [Paramecium bursaria Chlorella virus 1]	1.0E-88	NP_048659	92%
	Lateral-signal-induced phosphatase protein 1 [Caenorhabditis elegans]	1.4E-13	AAB88308	34%
	hypothetical protein C05B10.1 - Caenorhabditis elegans	1.4E-13	T32494	34%
B431L	A306L [Paramecium bursaria Chlorella virus 1]	6.3E-28	NP_048660	86%
B432L	A308L [Paramecium bursaria Chlorella virus 1]	1.7E-17	NP_048663	72%
	a309L [Paramecium bursaria Chlorella virus 1]	1.1E-11	NP_048664	71%
B433L	A315L [Paramecium bursaria Chlorella virus 1]	1.3E-70	NP_048671	52%
	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.9E-53	NP_049007	43%
	PBCV-1 33kd peptide [Paramecium bursaria Chlorella virus 1]	7.0E-48	NP_048641	42%
B434L	A310L [Paramecium bursaria Chlorella virus 1]	7.0E-83	NP_048665	85%
B435L	PBCV-1 33kDa translation peptide [Paramecium bursaria Chlorella virus 1]	7.8E-117	NP_048667	90%
	EsV-1-76 [Ectocarpus siliculosus virus]	4.0E-12	AAK14499	29%
	FirrV-1-II [Feldmannia irregularis virus a]	7.1E-09	AAR26966	29%
b436R	No Hit Found			
B437L	A313L [Paramecium bursaria Chlorella virus 1]	5.4E-19	NP_048669	65%
B438R	No Hit Found			
b439R	A314R [Paramecium bursaria Chlorella virus 1]	1.6E-26	NP_048670	71%
B440L	A315L [Paramecium bursaria Chlorella virus 1]	2.0E-66	NP_048671	50%
	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.2E-52	NP_049007	44%
	PBCV-1 33kd peptide [Paramecium bursaria Chlorella virus 1]	3.6E-44	NP_048641	41%
b441R	A89R [Paramecium bursaria Chlorella virus 1]	6.3E-28	NP_048437	92%
b442L	contains Gly-rich motifs GAGLGTGF (5X); similar to Arabidopsis Gly-rich protein, corresponds to Swiss-Prot Accession Number P27843 [Paramecium bursaria Chlorella virus 1]	9.7E-05	NP_048388	53%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
B443R	A318R [Paramecium bursaria Chlorella virus 1]	6.3E-117	NP_048674	91%
	PAPK (17X); similar to PBCV-1 ORF A41R, corresponds to Genbank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	9.1E-23	NP_048672	77%
b444R	No Hit Found			
b445L	No Hit Found			
B446R	A354R [Paramecium bursaria Chlorella virus 1]	9.8E-35	NP_048711	37%
	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081 [Paramecium bursaria Chlorella virus 1]	2.1E-08	NP_048779	26%
	endonuclease [Tetrahymena thermophila]	8.1E-05	AAL73456	30%
b447R	No Hit Found			
b448L	No Hit Found			
b449L	No Hit Found			
B450R	A320R [Paramecium bursaria Chlorella virus 1]	1.1E-29	NP_048676	57%
B451R	A321R [Paramecium bursaria Chlorella virus 1]	4.9E-33	NP_048677	60%
b452R	No Hit Found			
B453L	A322L [Paramecium bursaria Chlorella virus 1]	1.0E-50	NP_048678	57%
b454R	No Hit Found			
B455L	A324L [Paramecium bursaria Chlorella virus 1]	3.5E-155	NP_048680	72%
	FirrV-1-B22 [Feldmannia irregularis virus a]	1.7E-13	AAR26897	27%
	EsV-1-103 [Ectocarpus siliculosus virus]	2.8E-11	AAK14521	22%
b456R	No Hit Found			
B457L	A326L [Paramecium bursaria Chlorella virus 1]	6.8E-90	NP_048682	77%
	hypothetical protein DDB0218133 [Dictyostelium discoideum]	2.0E-04	XP_647817	32%
	hypothetical protein lpp2168 [Legionella pneumophila str. Paris]	3.4E-04	YP_124480	30%
B458L	A328L [Paramecium bursaria Chlorella virus 1]	4.2E-161	NP_048684	75%
b459R	No Hit Found			
B460R	A329R [Paramecium bursaria Chlorella virus 1]	2.9E-33	NP_048685	68%
B461R	No Hit Found			
B462R	No Hit Found			
B463L	No Hit Found			
b464L	No Hit Found			
B465R	hypothetical protein FG06548.1 [Gibberella zae PH-1]	2.9E-59	EAA78333	39%
	hypothetical protein FG03171.1 [Gibberella zae PH-1]	3.7E-54	EAA72911	38%
	hypothetical protein FG11489.1 [Gibberella zae PH-1]	5.0E-51	EAA78448	40%
b466L	No Hit Found			
b467L	No Hit Found			
B468R	PBCV-1 alginate lyase [Paramecium bursaria Chlorella virus 1]	4.1E-28	NP_048562	36%
	alginate lyase [Chlorella virus]	3.4E-27	BAA83789	35%
	vAL-1 [Chlorella virus]	2.2E-26	BAB19127	35%
B469L	GA14716-PA [Drosophila pseudoobscura]	5.5E-35	EAL33604	33%
	ENSANGP00000011077 [Anopheles gambiae str. PEST]	7.2E-35	XP_317336	31%
	CG17905-PA [Drosophila melanogaster]	8.0E-34	NP_609806	31%
b470R	No Hit Found			
b471R	No Hit Found			
B472R	chitin synthase [Chlorella virus]	0.0E+00	BAB83509	93%
	hypothetical protein FG06550.1 [Gibberella zae PH-1]	1.9E-44	EAA78335	29%
	hypothetical protein FG10619.1 [Gibberella zae PH-1]	6.0E-43	EAA68628	28%
b473L	No Hit Found			
b474L	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
b475L	No Hit Found			
b476R	No Hit Found			
B477R	4 ankyrin repeats; similar to <i>Drosophila melanogaster</i> ankyrin, encoded by GenBank Accession Number L35601 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048686	94%
	PREDICTED: similar to CG7462-PB, isoform B [<i>Apis mellifera</i>] ankyrin - fruit fly (<i>Drosophila melanogaster</i>)	1.6E-06 6.8E-13	XP_395788 AAC37208	28% 26%
b478L	No Hit Found			
b479L	a331L [Paramecium bursaria Chlorella virus 1]	1.0E-30	NP_048687	98%
B480L	PLPRNLLL (4X), SPPPSKP (3X) [Paramecium bursaria Chlorella virus 1]	2.3E-115	NP_048689	85%
	a332L [Paramecium bursaria Chlorella virus 1]	2.6E-58	NP_048688	97%
	ENSANGP0000018413 [Anopheles gambiae str. PEST]	2.5E-13	XP_321236	34%
b481L	PLPRNLLL (4X), SPPPSKP (3X) [Paramecium bursaria Chlorella virus 1]	1.2E-07	NP_048689	100%
b482R	No Hit Found			
b483L	No Hit Found			
b484L	No Hit Found			
B485L	A339L [Paramecium bursaria Chlorella virus 1]-1	2.6E-21	NP_048695	43%
	A337L [Paramecium bursaria Chlorella virus 1]	5.7E-21	NP_048693	64%
b486R	No Hit Found			
B487L	A341L [Paramecium bursaria Chlorella virus 1]	1.4E-59	NP_048697	82%
B488L	A342L [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048699	85%
b489R	a344R [Paramecium bursaria Chlorella virus 1] 1	5.7E-29	NP_048701	55%
b490L	a345L [Paramecium bursaria Chlorella virus 1]	1.8E-14	NP_048702	64%
b491R	No Hit Found			
B492R	No Hit Found			
b493L	A87R [Paramecium bursaria Chlorella virus 1]	6.2E-12	NP_048435	58%
B494R	A348R [Paramecium bursaria Chlorella virus 1]	3.7E-56	NP_048705	71%
b495L	a347L [Paramecium bursaria Chlorella virus 1]	9.2E-16	NP_048704	52%
B496L	URF14.2 [Chlorella virus]	4.9E-41	BAA22202	88%
	URF14.2 [Chlorella virus]	3.5E-39	BAA22201	83%
	URF14.2 [Chlorella virus]	7.9E-39	BAA22199	82%
B497L	KKD (6X), mixed charge [Paramecium bursaria Chlorella virus 1]	1.4E-15	NP_048708	46%
	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.9E-09	NP_049007	27%
	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440 [Paramecium bursaria Chlorella virus 1]	2.0E-06	NP_048851	35%
B498R	a350R [Paramecium bursaria Chlorella virus 1]	6.8E-43	NP_048707	62%
	unnamed protein product [Candida glabrata CBS138]	7.3E-05	CAG59956	34%
	hypothetical protein DDB0188233 [Dictyostelium discoideum]	9.5E-05	XP_636443	32%
B499L	KKD (6X), mixed charge [Paramecium bursaria Chlorella virus 1]	1.5E-123	NP_048708	65%
	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440 [Paramecium bursaria Chlorella virus 1]	3.0E-07	NP_048851	39%
	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	7.3E-06	NP_049007	34%
b500R	No Hit Found			
b501R	No Hit Found			
b502R	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
B503L	Asp/Glu rich; DAEDDDIYxxET (2X) negative charge cluster [Paramecium bursaria Chlorella virus 1]	4.8E-97	NP_048709	84%
	unknown [Acanthamoeba polyphaga mimivirus]	5.1E-06	YP_142843	24%
b504L	No Hit Found			
b505R	No Hit Found			
B506L	A357L [Paramecium bursaria Chlorella virus 1]	2.0E-88	NP_048714	68%
	a359L [Paramecium bursaria Chlorella virus 1]	1.8E-17	NP_048716	65%
b507R	No Hit Found			
B508R	similar to chicken vitellogenin II, corresponds to Swiss-Prot Accession Number P02845 [Paramecium bursaria Chlorella virus 1]	2.5E-107	NP_048720	38%
	putative helicase [Listonella pelagia phage phiHSIC]	6.9E-33	YP_224270	29%
	vacuolar membrane H-ATPase [Saccharomyces cariocanus]	5.7E-19	CAC86344	28%
b509L	No Hit Found			
b510L	No Hit Found			
b511L	No Hit Found			
b512L	No Hit Found			
b513L	No Hit Found			
b514L	No Hit Found			
b515L	No Hit Found			
b516L	No Hit Found			
b517L	No Hit Found			
b518L	No Hit Found			
B519L	A366L [Paramecium bursaria Chlorella virus 1]	1.2E-10	NP_048723	34%
b520R	No Hit Found			
b521R	No Hit Found			
B522L	A366L [Paramecium bursaria Chlorella virus 1]	9.2E-89	NP_048723	62%
b523R	a377R [Paramecium bursaria Chlorella virus 1]	9.3E-16	NP_048734	54%
B524L	A378L [Paramecium bursaria Chlorella virus 1]	4.0E-95	NP_048735	74%
	DNA binding protein [Chlorella virus]	4.1E-07	BAA11342	51%
	A565R [Paramecium bursaria Chlorella virus 1]	3.1E-07	NP_048921	50%
B525L	A379L [Paramecium bursaria Chlorella virus 1]	4.1E-104	NP_048736	86%
b526R	a380R [Paramecium bursaria Chlorella virus 1]	1.5E-33	NP_048737	71%
b527R	a381R [Paramecium bursaria Chlorella virus 1]	9.6E-29	NP_048738	89%
b528L	No Hit Found			
B529R	similar to PBCV-1 major capsid protein, corresponds to Swiss-Prot Accession Number P30328 [Paramecium bursaria Chlorella virus 1]	1.8E-134	NP_048740	89%
	cytochrome C family heme-binding site signature [Paramecium bursaria Chlorella virus 1]	9.3E-43	NP_048739	63%
	major capsid protein Vp54 [Chlorella virus]	8.4E-36	BAA22198	34%
b530L	No Hit Found			
b531L	No Hit Found			
b532R	Gly-rich, AGLG (9x); similar to herpesvirus hypothetical protein 5, corresponds to GenBank Accession Number S43071 [Paramecium bursaria Chlorella virus 1]	3.2E-39	NP_048743	66%
B533L	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472 [Paramecium bursaria Chlorella virus 1]	1.4E-15	NP_048741	31%
	DNA binding protein [Chlorella virus]	5.7E-25	BAA11343	38%
	A565R [Paramecium bursaria Chlorella virus 1]	8.3E-24	NP_048921	38%
b534R	a388R [Paramecium bursaria Chlorella virus 1]	7.3E-05	NP_048745	75%
b535R	a391R [Paramecium bursaria Chlorella virus 1]	5.6E-16	NP_048748	63%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
B536R	contains ATP/GTP-binding site motif A; similar to frog virus 3 ATPase, corresponds to GenBank Accession Number M80551 [Paramecium bursaria Chlorella virus 1]	8.2E-109	NP_048749	78%
	AGB-1 [Heterosigma akashiwo virus]	1.2E-38	BAB69884	44%
	FirrV-1-A12 [Feldmannia irregularis virus a]	8.9E-31	AAR26836	38%
B537R	No Hit Found			
b538L	No Hit Found			
B539R	A394R [Paramecium bursaria Chlorella virus 1]	2.8E-44	NP_048751	65%
B540R	A395R [Paramecium bursaria Chlorella virus 1]	5.4E-32	NP_048752	81%
B541L	A396L [Paramecium bursaria Chlorella virus 1]	6.9E-19	NP_048753	71%
B542R	No Hit Found			
B543L	DNA adenine methyltransferase [Chlorella virus NY2A]	2.7E-153	AAC03123	98%
	hypothetical protein SalaDRAFT_0836 [Sphingopyxis alaskensis RB2256]	3.5E-28	ZP_00579458	31%
	putative type II DNA methylase protein [Mycoplasma mobile 163K]	1.7E-22	YP_015792	42%
b544R	No Hit Found			
b545R	No Hit Found			
B546L	A398L [Paramecium bursaria Chlorella virus 1]	1.1E-51	NP_048755	84%
B547R	A399R [Paramecium bursaria Chlorella virus 1]	8.0E-77	NP_048756	73%
B548R	similar to PBCV-1 ORF A214, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	5.7E-42	NP_048757	75%
B549R	A401R [Paramecium bursaria Chlorella virus 1]	7.0E-136	NP_048758	88%
	hypothetical protein bll5851 [Bradyrhizobium japonicum USDA 110]	5.9E-42	NP_772491	39%
	hypothetical protein GOX0246 [Gluconobacter oxydans 621H]	7.3E-32	YP_190685	31%
B550R	A402R [Paramecium bursaria Chlorella virus 1]	3.3E-111	NP_048759	85%
b551L	No Hit Found			
B552R	A403R [Paramecium bursaria Chlorella virus 1]	1.2E-42	NP_048760	92%
B553R	Glu-, Asn-rich [Paramecium bursaria Chlorella virus 1]	2.0E-21	NP_048761	87%
B554R	Pro-, Lys-rich, PAPK (30x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472 [Paramecium bursaria Chlorella virus 1]	1.3E-122	NP_048762	77%
b555L	No Hit Found			
B556R	No Hit Found			
b557R	No Hit Found			
b558L	No Hit Found			
B559L	A407L [Paramecium bursaria Chlorella virus 1]	6.9E-91	NP_048764	77%
B560L	A408L [Paramecium bursaria Chlorella virus 1]	3.6E-94	NP_048765	82%
	A410L [Paramecium bursaria Chlorella virus 1]	7.4E-15	NP_048767	40%
	EsV-1-42 [Ectocarpus siliculosus virus]	4.8E-14	AAK14468	34%
b561R	a409R [Paramecium bursaria Chlorella virus 1]	2.8E-12	NP_048766	80%
B562L	A410L [Paramecium bursaria Chlorella virus 1]	9.6E-53	NP_048767	90%
	A408L [Paramecium bursaria Chlorella virus 1]	2.7E-15	NP_048765	37%
	EsV-1-42 [Ectocarpus siliculosus virus]	1.2E-05	AAK14468	30%
b563L	No Hit Found			
B564R	A412R [Paramecium bursaria Chlorella virus 1]	9.0E-87	NP_048769	86%
b565R	No Hit Found			
B566R	Type II restriction-modification system methylation subunit [Bacillus cereus ATCC 14579]	2.6E-35	AAP07928	36%
	putative DNA methylase [Escherichia coli]	7.7E-35	CAD33713	35%
	modification methylase M.NspHI [Nostoc sp. ATCC 29106]	2.9E-34	AAC97192	35%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
B567L	DNA adenine methyltransferase [Chlorella virus SC-1A]	0.0E+00	AAC57943	85%
	Modification methylase CviRI (Adenine-specific methyltransferase CviRI) (M.CviRI)	2.6E-52	P52284	35%
	DNA adenine methyltransferase [Chlorella virus NY2A]	3.8E-51	AAC03124	34%
B568L	Gln-rich, QQQQM(4x); similar to human transcription factor TFIID, corresponds to Swiss-Prot Accession Number P20226 [Paramecium bursaria Chlorella virus 1]	1.7E-73	NP_048770	61%
b569L	a415L [Paramecium bursaria Chlorella virus 1]	1.2E-23	NP_048772	80%
B570R	A414R [Paramecium bursaria Chlorella virus 1]	1.1E-29	NP_048771	87%
B571L	A417L [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048774	76%
	replication factor C, 140 kDa subunit [Plasmodium yoelii yoelii]	6.6E-05	EAA16369	20%
	replication factor C, large subunit [Methanococcus maripaludis S2]	1.9E-04	NP_987442	27%
b572R	No Hit Found			
B573L	A420L [Paramecium bursaria Chlorella virus 1]	6.5E-25	NP_048777	77%
b574R	No Hit Found			
B575R	A421R [Paramecium bursaria Chlorella virus 1]	2.7E-31	NP_048778	63%
	A139L [Paramecium bursaria Chlorella virus 1]	8.8E-06	NP_048487	36%
B576R	A423R [Paramecium bursaria Chlorella virus 1]	1.5E-58	NP_048780	73%
b577L	No Hit Found			
b578L	No Hit Found			
B579R	No Hit Found			
B580R	A426R [Paramecium bursaria Chlorella virus 1]	9.3E-48	NP_048783	79%
B581L	contains thioredoxin active site-like sequence; similar to Synechocystis thioredoxin-like protein, corresponds to Swiss-Prot Accession Number P52232 [Paramecium bursaria Chlorella virus 1]	4.9E-49	NP_048784	75%
	GA21460-PA [Drosophila pseudoobscura]	1.9E-05	EAL29786	25%
	Thioredoxin [Chlorobium phaeobacteroides BS1]	2.5E-05	ZP_00533498	29%
B582L	Pro- and Glu-rich, PENPEV (10x); similar to Streptococcus B antigen, corresponds to Swiss-Prot Accession Number P27951 [Paramecium bursaria Chlorella virus 1]	2.2E-04	NP_048785	78%
B583L	A429L [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048786	77%
	ankyrin repeat family protein [Arabidopsis thaliana]	3.5E-12	NP_192255	25%
	hypothetical protein [Arabidopsis thaliana]	1.5E-10	CAB77831	24%
b584L	No Hit Found			
B585L	major capsid protein MCP1 [Chlorella virus]	0.0E+00	BAA76601	90%
	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048787	89%
	major capsid protein Vp54 [Chlorella virus]	0.0E+00	BAA22198	87%
b586R	No Hit Found			
B587R	Lys-rich [Paramecium bursaria Chlorella virus 1]	1.9E-45	NP_048789	74%
b588L	No Hit Found			
b589R	No Hit Found			
B590R	No Hit Found			
B591L	similar to Methanothrix chromosomal protein MC1A, corresponds to Swiss-Prot Accession Number P15251 [Paramecium bursaria Chlorella virus 1]	1.7E-30	NP_048794	67%
B592L	similar to E. coli glutaredoxin, corresponds to Swiss-Prot Accession Number P37687 [Paramecium bursaria Chlorella virus 1]	1.3E-30	NP_048795	79%
B593R	A439R [Paramecium bursaria Chlorella virus 1]	1.2E-44	NP_048796	72%
b594R	No Hit Found			
B595L	A441L [Paramecium bursaria Chlorella virus 1]	4.2E-53	NP_048798	73%
B596R	A443R [Paramecium bursaria Chlorella virus 1]	5.6E-	NP_048800	85%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
		148		
B597L	A444L [Paramecium bursaria Chlorella virus 1]	7.4E-45	NP_048801	89%
	A354R [Paramecium bursaria Chlorella virus 1]	1.2E-37	NP_048711	38%
B598L	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081 [Paramecium bursaria Chlorella virus 1]	3.1E-14	NP_048779	31%
b599R	No Hit Found			
b600L	No Hit Found			
b601R	No Hit Found			
	A315L [Paramecium bursaria Chlorella virus 1]	1.3E-63	NP_048671	61%
B602L	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.8E-62	NP_049007	57%
	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440 [Paramecium bursaria Chlorella virus 1]	7.0E-38	NP_048851	58%
b603L	No Hit Found			
b604L	No Hit Found			
b605R	No Hit Found			
	similar to Clostridium pasteurianum ORF, corresponds to GenBank Accession Number Z28353 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048802	89%
B606L	488aa long conserved hypothetical protein [Sulfolobus tokodaii str. 7]	1.2E-31	BAB66733	30%
	universally conserved protein [Sulfolobus acidocaldarius DSM 639]	2.2E-30	YP_256267	28%
b607R	a446R [Paramecium bursaria Chlorella virus 1]	2.6E-18	NP_048803	88%
b608R	No Hit Found			
b609R	a447R [Paramecium bursaria Chlorella virus 1]	1.7E-28	NP_048804	95%
b610R	No Hit Found			
	contains cytochrome C family heme-binding site signature; similar to maize protein disulphide isomerase, correspond to Swiss-Prot Accession Number P52588 [Paramecium bursaria Chlorella virus 1]	2.7E-47	NP_048805	82%
B611L	PREDICTED: hypothetical protein XP_532876 [Canis familiaris]	1.8E-06	XP_532876	40%
	protein disulfide isomerase, putative [Trypanosoma cruzi]	2.3E-06	EAN99322	30%
B612R	PBCV-1 RNA triphosphatase [Paramecium bursaria Chlorella virus 1]	3.6E-87	NP_048806	82%
b613L	No Hit Found			
B614R	No Hit Found			
b615L	No Hit Found			
b616L	No Hit Found			
	major capsid protein MCP1 [Chlorella virus]	0.0E+00	BAA76601	87%
B617L	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048787	87%
	major capsid protein Vp54 [Chlorella virus]	0.0E+00	BAA22198	85%
	putative N-acetylglucosaminyltransferase [Bdellovibrio bacteriovorus HD100]	2.1E-28	NP_968551	32%
B618R	hypothetical protein FG05624.1 [Gibberella zeae PH-1]	4.2E-13	EAA75195	30%
	hypothetical protein MG04516.4 [Magnaporthe grisea 70-15]	4.6E-12	EAA50757	28%
B619L	A454L [Paramecium bursaria Chlorella virus 1]	1.0E-140	NP_048811	86%
b620R	a453R [Paramecium bursaria Chlorella virus 1]	6.5E-17	NP_048810	64%
b621R	a455R [Paramecium bursaria Chlorella virus 1]	3.8E-41	NP_048812	72%
b622L	No Hit Found			
	contains ATP/GTP-binding site motif A [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048813	91%
B623L	EsV-1-109 [Ectocarpus siliculosus virus]	3.9E-74	AAK14527	32%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
	FirrV-1-B27 [Feldmannia irregularis virus a]	1.1E-68	AAR26902	33%
b624R	a457R [Paramecium bursaria Chlorella virus 1]	3.3E-13	NP_048814	100%
b625R	No Hit Found			
b626R	No Hit Found			
b627L	a458L [Paramecium bursaria Chlorella virus 1]	2.6E-18	NP_048815	60%
B628R	similar to Bacillus ribonuclease III, corresponds to Swiss-Prot Accession Number P51833 [Paramecium bursaria Chlorella virus 1]	3.5E-137	NP_048820	88%
	ribonuclease III [Chlorobium tepidum TLS]	8.3E-30	NP_662993	36%
	Ribonuclease III [Pelodictyon phaeoclastratiforme BU-1]	1.1E-29	ZP_00590199	35%
B629R	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440 [Paramecium bursaria Chlorella virus 1]	3.4E-11	NP_048851	32%
	A315L [Paramecium bursaria Chlorella virus 1]	1.3E-10	NP_048671	34%
	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	4.8E-10	NP_049007	45%
B630R	PBCV-1 thiol oxidoreductase [Paramecium bursaria Chlorella virus 1]	8.4E-49	NP_048821	73%
	putative thiol oxidoreductase [Acanthamoeba polyphaga mimivirus]	2.3E-14	YP_142722	38%
	thiol oxidoreductase E10R [Acanthamoeba polyphaga mimivirus]	5.9E-10	YP_142950	38%
B631L	A467L [Paramecium bursaria Chlorella virus 1]	1.9E-164	NP_048823	91%
	Hypothetical protein [Amsacta moorei entomopoxvirus]	2.0E-07	NP_065022	27%
b632R	No Hit Found			
B633R	A468R [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048824	81%
	FirrV-1-A46 [Feldmannia irregularis virus a]	9.8E-04	AAR26870	33%
b634L	a469L [Paramecium bursaria Chlorella virus 1]	1.8E-06	NP_048825	46%
b635L	No Hit Found			
B636R	A470R [Paramecium bursaria Chlorella virus 1]	4.1E-79	NP_048826	74%
	UKCH-2 [Heterosigma akashiwo virus]	1.5E-17	BAB69883	36%
	FirrV-1-A23 [Feldmannia irregularis virus a]	7.4E-04	AAR26847	29%
b637R	No Hit Found			
B638R	A471R [Paramecium bursaria Chlorella virus 1]	2.5E-83	NP_048827	84%
	unknown [Acanthamoeba polyphaga mimivirus]	1.3E-26	YP_142861	47%
	FirrV-1-A5 [Feldmannia irregularis virus a]	1.0E-15	AAR26829	32%
b639L	a472L [Paramecium bursaria Chlorella virus 1]	2.9E-17	NP_048828	60%
b640L	contains prokaryotic membrane lipoprotein lipid attachment site; similar to vaccinia virus F-ORF-C 11.6 kD protein, corresponds to Swiss-Prot Accession Number P20561 [Paramecium bursaria Chlorella virus 1]	3.3E-08	NP_048833	60%
	WSSV244 [shrimp white spot syndrome virus]	5.2E-06	AAL89112	38%
B641R	contains ribonucleotide reductase (RR) signature; similar to tobacco RR small subunit, corresponds to Swiss-Prot Accession Number P49730 [Paramecium bursaria Chlorella virus 1]	1.0E-171	NP_048832	89%
	ribonucleotide reductase small subunit [Arabidopsis thaliana]	1.2E-108	AAO62422	59%
	putative ribonucleotide reductase small subunit [Arabidopsis thaliana]	2.7E-108	AAM51287	58%
b642L	No Hit Found			
b643R	No Hit Found			
B644L	Gly-rich [Paramecium bursaria Chlorella virus 1]	4.1E-24	NP_048836	87%
B645L	A481L [Paramecium bursaria Chlorella virus 1]	4.9E-117	NP_048837	85%
b646R	No Hit Found			
B647R	A482R [Paramecium bursaria Chlorella virus 1]	1.4E-107	NP_048838	86%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
	FirrV-1-I5 [Feldmannia irregularis virus a]	1.4E-09	AAR26970	36%
	FirrV-1-B14 [Feldmannia irregularis virus a]	2.6E-08	AAR26889	39%
b648L	No Hit Found			
b649R	No Hit Found			
B650L	A484L [Paramecium bursaria Chlorella virus 1]	2.6E-74	NP_048840	83%
B651R	A485R [Paramecium bursaria Chlorella virus 1]	2.7E-63	NP_048841	83%
	unknown [Acanthamoeba polyphaga mimivirus]	3.7E-04	YP_142801	26%
B652R	Pro-,Gln-rich [Paramecium bursaria Chlorella virus 1]	3.4E-124	NP_048844	76%
	unknown [Acanthamoeba polyphaga mimivirus]	4.5E-07	YP_142771	28%
b653L	No Hit Found			
b654L	No Hit Found			
B655R	A491R [Paramecium bursaria Chlorella virus 1]	2.2E-28	NP_048847	77%
B656L	A492L [Paramecium bursaria Chlorella virus 1]	3.5E-83	NP_048848	80%
	A493L [Paramecium bursaria Chlorella virus 1]	1.1E-49	NP_048849	57%
B657L	A493L [Paramecium bursaria Chlorella virus 1] &	4.5E-82	NP_048849	81%
	A492L [Paramecium bursaria Chlorella virus 1]	6.9E-59	NP_048848	62%
b658L	No Hit Found			
B659L	No Hit Found			
b660R	No Hit Found			
b661L	No Hit Found			
B662R	Lys-, Arg-rich; contains eukaryotic putative RNA-binding region RNP-1 signature; similar to PBCV-1 ORF A267L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	2.2E-12	NP_048834	37%
	Lys-, Glu-rich [Paramecium bursaria Chlorella virus 1]	1.4E-11	NP_048846	38%
	unknown [Acanthamoeba polyphaga mimivirus]	2.3E-14	YP_142777	41%
B663R	A494R [Paramecium bursaria Chlorella virus 1]	1.9E-169	NP_048850	80%
	PBCV1-A494R-like protein [Acanthamoeba polyphaga mimivirus]	1.4E-34	YP_142783	41%
	EsV-1-101 [Ectocarpus siliculosus virus]	2.2E-32	AAK14519	41%
b664L	No Hit Found			
b665L	Pro-rich, IPPPNMSLPLS (3x) [Paramecium bursaria Chlorella virus 1]	1.3E-09	NP_048854	85%
B666R	Gly-rich, KGKDIGGG (4x) [Paramecium bursaria Chlorella virus 1]	6.1E-15	NP_048853	36%
B667L	A501L [Paramecium bursaria Chlorella virus 1]	4.0E-17	NP_048857	92%
	Pro-, Ser-rich [Paramecium bursaria Chlorella virus 1]	6.4E-07	NP_048856	58%
b668R	No Hit Found			
b669L	No Hit Found			
B670L	A502L [Paramecium bursaria Chlorella virus 1]	2.5E-32	NP_048858	89%
B671L	contains prokaryotic membrane lipoprotein lipid attachment site [Paramecium bursaria Chlorella virus 1]	2.4E-136	NP_048859	83%
B672L	A505L [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048861	80%
b673R	a507R [Paramecium bursaria Chlorella virus 1]	1.2E-15	NP_048863	55%
b674R	a507R [Paramecium bursaria Chlorella virus 1]	1.1E-11	NP_048863	59%
b675R	a508R [Paramecium bursaria Chlorella virus 1]	3.9E-23	NP_048864	80%
b676R	a510R [Paramecium bursaria Chlorella virus 1]	9.9E-18	NP_048866	87%
B677R	tripartite motif protein 50 [Rattus norvegicus]	1.1E-05	NP_851594	36%
	Tripartite motif protein 50 [Rattus norvegicus]	1.1E-05	AAH85684	36%
	CG10981-PA, isoform A [Drosophila melanogaster]	2.5E-05	AAF51944	43%
b678L	No Hit Found			
b679L	A519L [Paramecium bursaria Chlorella virus 1]	1.3E-30	NP_048875	80%
B680L	A520L [Paramecium bursaria Chlorella virus 1]	1.0E-41	NP_048876	86%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
B681L	CviPII m5C DNA methyltransferase [Chlorella virus]	0.0E+00	AAV84097	92%
	M.CviAII cytosine DNA methyltransferase [Paramecium bursaria Chlorella virus 1]	9.6E-86	NP_048873	47%
	cytosine methyltransferase [Paramecium bursaria Chlorella virus NYs-1]	2.8E-85	AAC64006	48%
b682R	No Hit Found			
b683R	No Hit Found			
B684L	contains Gln-rich, neutral zinc metallopeptidase, zinc binding region signature [Paramecium bursaria Chlorella virus 1]	5.2E-69	NP_048877	76%
B685L	contains Gln-rich, neutral zinc metallopeptidase, zinc binding region signature [Paramecium bursaria Chlorella virus 1]	8.5E-80	NP_048877	76%
	metal-dependent hydrolase [Acanthamoeba polyphaga mimivirus]	6.7E-08	YP_142679	29%
	136R [Chilo iridescent virus]	8.2E-06	AAB94453	38%
b686R	No Hit Found			
B687R	A523R [Paramecium bursaria Chlorella virus 1]	4.3E-77	NP_048879	83%
B688L	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.0E-136	NP_048920	72%
	A154L [Paramecium bursaria Chlorella virus 1]	5.0E-06	NP_048502	43%
	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903 [Paramecium bursaria Chlorella virus 1]	2.9E-126	NP_048477	68%
b689R	a155R [Paramecium bursaria Chlorella virus 1]	1.6E-15	NP_048503	61%
b690R	No Hit Found			
b691R	No Hit Found			
B692R	A526R [Paramecium bursaria Chlorella virus 1]	3.0E-59	NP_048882	84%
b693R	No Hit Found			
B694R	A527R [Paramecium bursaria Chlorella virus 1]	1.3E-20	NP_048883	85%
b695L	a518R [Paramecium bursaria Chlorella virus 1]	9.0E-25	NP_048874	63%
	a529L [Paramecium bursaria Chlorella virus 1]	4.6E-21	NP_048885	81%
	a685R [Paramecium bursaria Chlorella virus 1]	5.5E-06	NP_049041	50%
B696R	No Hit Found			
B697R	M.CviAIV cytosine DNA methyltransferase [Paramecium bursaria Chlorella virus 1]	1.4E-167	NP_048886	85%
	M.CviAII cytosine DNA methyltransferase [Paramecium bursaria Chlorella virus 1]	1.9E-116	NP_048873	59%
	cytosine methyltransferase [Paramecium bursaria Chlorella virus NYs-1]	2.5E-79	AAC64006	48%
b698L	No Hit Found			
B699L	ubiquitin extension protein (UbCEP52) [Chlamydomonas reinhardtii]	2.9E-33	CAA43216	94%
	polyubiquitin [Volvox carteri f. nagariensis]	2.9E-33	CAA52290	94%
	Ubiquitin	2.9E-33	P14624	94%
B700L	A531L [Paramecium bursaria Chlorella virus 1]	3.7E-20	NP_048887	76%
B701L	A532L [Paramecium bursaria Chlorella virus 1]	1.2E-18	NP_048888	58%
B702L	putative transposase [Chlorella virus]	0.0E+00	AAU06281	93%
	similar to Synechocystis transposase, corresponds to GenBank Accession Number D90909 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048981	86%
	putative transposase [Acanthamoeba polyphaga mimivirus]	1.4E-23	YP_142458	31%
b703R	No Hit Found			
B704L	unnamed protein product [Mus musculus]	1.9E-05	BAB32292	50%
	hypothetical protein PFL2170c [Plasmodium falciparum 3D7]	5.6E-05	NP_701794	50%
	DNA repair protein-like-related [Plasmodium yoelii yoelii]	4.8E-04	EAA22469	28%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
B705R	A533R [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048889	86%
	unknown protein	0.0E+00	AAA66400	86%
	ORF1 protein - Chlorella virus PBCV-1	0.0E+00	A46352	86%
b706L	No Hit Found			
b707R	No Hit Found			
b708R	No Hit Found			
b709L	No Hit Found			
B710L	A535L [Paramecium bursaria Chlorella virus 1]	1.0E-32	NP_048891	90%
B711L	putative transposase [Chlorella virus]	5.4E-51	AAU06281	32%
	similar to Synechocystis transposase, corresponds to GenBank Accession Number D90909 [Paramecium bursaria Chlorella virus 1]	5.9E-50	NP_048981	33%
	putative transposase [Acanthamoeba polyphaga mimivirus]	9.2E-11	YP_143208	28%
b712R	No Hit Found			
b713L	No Hit Found			
B714L	No Hit Found			
B715L	putative resolvase [Acanthamoeba polyphaga mimivirus]	3.8E-24	YP_143125	41%
	putative resolvase [Acanthamoeba polyphaga mimivirus]	4.2E-23	YP_142434	39%
	putative resolvase [Acanthamoeba polyphaga mimivirus]	1.4E-21	YP_142457	37%
b716R	No Hit Found			
b717R	No Hit Found			
B718L	A354R [Paramecium bursaria Chlorella virus 1]	1.2E-15	NP_048711	36%
	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081 [Paramecium bursaria Chlorella virus 1]	4.0E-11	NP_048779	31%
	A87R [Paramecium bursaria Chlorella virus 1]	1.5E-10	NP_048435	32%
B719L	A536L [Paramecium bursaria Chlorella virus 1]	5.1E-14	NP_048892	66%
B720L	A366L [Paramecium bursaria Chlorella virus 1]	3.8E-55	NP_048723	61%
b721R	No Hit Found			
b722R	No Hit Found			
B723L	A537L [Paramecium bursaria Chlorella virus 1]	3.1E-87	NP_048893	60%
B724R	A539R [Paramecium bursaria Chlorella virus 1]	4.2E-78	NP_048895	80%
	similar to bacteriophage T4 intron-associated endonuclease, corresponds to Swiss-Prot Accession Number P13299 [Paramecium bursaria Chlorella virus 1]	1.6E-05	NP_048482	56%
	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440 [Paramecium bursaria Chlorella virus 1]	1.1E-04	NP_048851	34%
B725L	A540L [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048896	51%
	unknown protein	9.6E-101	AAA66404	64%
	Hep_Hag [Chlorobium phaeobacteroides BS1]	1.5E-13	ZP_00532602	24%
b726R	No Hit Found			
b727R	No Hit Found			
b728R	No Hit Found			
b729L	No Hit Found			
b730R	No Hit Found			
b731L	No Hit Found			
b732L	No Hit Found			
b733R	No Hit Found			
B734R	Chain A, New Crystal Structure Of Chlorella Virus Dna Ligase-Adenylate	5.8E-143	1P8L_A	80%
	PBCV-1 DNA ligase [Paramecium bursaria Chlorella virus 1]	1.7E-142	NP_048900	81%
	Chain A, Crystal Structure Of Chlorella Virus Dna Ligase-Adenylate	5.4E-	1FVI_A	81%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
b735L	No Hit Found	141		
B736L	similar to Streptomyces glucosyltransferase protein, corresponds to GenBank Accession Number AB005901 [Paramecium bursaria Chlorella virus 1]	8.9E-150	NP_048902	81%
	hypothetical protein A546L - Chlorella virus PBCV-1	3.5E-138	T18048	80%
	putative glycosyltransferases [Bacteriophage RM 378]	6.4E-07	NP_835600	34%
b737R	No Hit Found			
B738L	similar to Caenorhabditis transcription activator, corresponds to Swiss-Prot Accession Number P41877 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048904	85%
	COG0553: Superfamily II DNA/RNA helicases, SNF2 family [Streptococcus pneumoniae TIGR4]	1.2E-41	ZP_00404276	30%
	Snf2 family protein [Streptococcus pneumoniae TIGR4]	1.2E-41	NP_345973	30%
b739R	a549R [Paramecium bursaria Chlorella virus 1]	2.8E-12	NP_048905	41%
b740R	a550R [Paramecium bursaria Chlorella virus 1]>gi 2447139 gb AAC96997.1 a550R [Paramecium bursaria Chlorella virus 1] >gi 7461228 pir T18052 DEAH box protein a550R - Chlorella virus PBCV-1	8.8E-14	NP_048906	44%
B741L	similar to tomato dUTP pyrophosphatase, corresponds to GenBank Accession Number S40549 [Paramecium bursaria Chlorella virus 1]	5.4E-56	NP_048907	83%
	deoxyuridine triphosphatase [Paramecium bursaria Chlorella virus IL3A]	3.9E-54	AAW51452	80%
	deoxyuridine triphosphatase [Paramecium bursaria Chlorella virus SH6A]	1.1E-53	AAW51453	81%
b742L	No Hit Found			
B743R	similar to Sulfolobus TATA-binding protein, corresponds to GenBank Accession Number S55311'; [Paramecium bursaria Chlorella virus 1]	4.3E-127	NP_048908	80%
	hypothetical protein A552R - Chlorella virus PBCV-1	6.5E-107	T18054	84%
B744L	similar to MesJ cell cycle protein [Chlorobium tepidum], corresponds to accession number AAG12421 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048910	88%
	ATPase homolog A554L - Chlorella virus PBCV-1	2.7E-133	T18056	85%
	hypothetical protein A557L - Chlorella virus PBCV-1	4.1E-49	T18059	85%
b745R	No Hit Found			
b746R	No Hit Found			
B747L	A267L [Paramecium bursaria Chlorella virus 1]	1.6E-46	NP_048621	42%
	Lys-, Glu-rich [Paramecium bursaria Chlorella virus 1]	1.2E-09	NP_048846	39%
	Lys-, Arg-rich; contains eukaryotic putative RNA-binding region RNP-1 signature; similar to PBCV-1 ORF A267L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.3E-22	NP_048834	28%
B748L	similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048914	93%
	contains aminoacyl-tRNA synthetase class-II signature [Paramecium bursaria Chlorella virus 1]	2.4E-115	NP_048359	52%
	similar to PBCV-1 major capsid protein, corresponds to Swiss-Prot Accession Number P30328 [Paramecium bursaria Chlorella virus 1]	1.8E-86	NP_048358	41%
b749L	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
b750R	No Hit Found			
B751L	A559L [Paramecium bursaria Chlorella virus 1]	1.1E-72	NP_048915	71%
b752R	a560R [Paramecium bursaria Chlorella virus 1]	2.8E-12	NP_048916	65%
B753L	A267L [Paramecium bursaria Chlorella virus 1]	1.3E-35	NP_048621	32%
	Lys-, Glu-rich [Paramecium bursaria Chlorella virus 1]	3.4E-07	NP_048846	32%
B753L	Lys-, Arg-rich; contains eukaryotic putative RNA-binding region RNP-1 signature; similar to PBCV-1 ORF A267L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	3.9E-19	NP_048834	30%
b754R	No Hit Found			
b755R	No Hit Found			
B756L	DNA binding protein [Chlorella virus]	2.1E-72	BAA11342	58%
	similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305 [Paramecium bursaria Chlorella virus 1]	2.1E-19	NP_048917	100%
	A565R [Paramecium bursaria Chlorella virus 1]	2.1E-19	NP_048921	100%
b757R	No Hit Found			
b758L	No Hit Found			
B759R	DNA binding protein [Chlorella virus]	0.0E+00	BAA11343	78%
	A565R [Paramecium bursaria Chlorella virus 1]	2.5E-23	NP_048921	100%
	DNA binding protein [Chlorella virus]	2.9E-56	BAA11342	98%
b760L	a566L [Paramecium bursaria Chlorella virus 1]	9.2E-40	NP_048922	88%
B761L	A567L [Paramecium bursaria Chlorella virus 1]	1.6E-52	NP_048923	71%
B762L	A568L [Paramecium bursaria Chlorella virus 1]	4.9E-72	NP_048924	73%
B763L	A570L [Paramecium bursaria Chlorella virus 1]	1.4E-59	NP_048926	82%
b764L	No Hit Found			
B765R	A571R [Paramecium bursaria Chlorella virus 1]	1.1E-48	NP_048927	86%
	Arg-rich [Paramecium bursaria Chlorella virus 1]	9.5E-08	NP_048792	48%
B766R	A572R [Paramecium bursaria Chlorella virus 1]	1.8E-93	NP_048928	86%
B767L	similar to Periwinkle PCNA, corresponds to GenBank Accession Number X55052 [Paramecium bursaria Chlorella virus 1]	1.1E-115	NP_048930	83%
	proliferating cell nuclear antigen [Populus nigra]	1.8E-33	Q9MAY3	33%
	SPATULA-like [Oryza sativa (japonica cultivar-group)]	4.4E-32	XP_468284	32%
b768R	No Hit Found			
B769R	cytosine methyltransferase [Paramecium bursaria Chlorella virus NYs-1]	0.0E+00	AAC64006	99%
	Modification methylase CviJI (Cytosine-specific methyltransferase CviJI)	2.6E-142	P36216	66%
	nonfunctional M.CviAV cytosine DNA methyltransferase [Paramecium bursaria Chlorella virus 1]	3.8E-141	NP_049039	65%
b770L	a518R [Paramecium bursaria Chlorella virus 1]	1.0E-09	NP_048874	53%
B771L	A575L [Paramecium bursaria Chlorella virus 1]	9.6E-85	NP_048931	89%
B772L	A577L [Paramecium bursaria Chlorella virus 1]	6.3E-36	NP_048933	84%
b773R	No Hit Found			
B774R	PBCV-1 M.CviAI methylase [Paramecium bursaria Chlorella virus 1]	1.2E-133	NP_048937	87%
	N6 adenine-specific DNA methyltransferase, D12 class [Clostridium thermocellum ATCC 27405]	2.5E-46	ZP_00510571	42%
	N6 adenine-specific DNA methyltransferase, D12 class [Ferroplasma acidarmanus Fer1]	6.8E-44	EAM94529	44%
b775R	No Hit Found			
b776R	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
B777L	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.7E-143	NP_048920	67%
	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903 [Paramecium bursaria Chlorella virus 1]	7.8E-136	NP_048477	66%
	A154L [Paramecium bursaria Chlorella virus 1]	5.4E-129	NP_048502	62%
b778R	No Hit Found			
b779L	No Hit Found			
b780R	No Hit Found			
B781L	PBCV-1 DNA topoisomerase II [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048939	89%
	topoisomerase II [Chlorella virus Marburg 1]	0.0E+00	AAU95770	66%
	DNA TOPOISOMERASE II [Encephalitozoon cuniculi GB-M1]	0.0E+00	CAD25222	48%
b782R	No Hit Found			
b783R	contains prenyl group binding site (CAAX) box [Paramecium bursaria Chlorella virus 1]	2.8E-44	NP_048940	91%
b784R	a585R [Paramecium bursaria Chlorella virus 1]	2.3E-34	NP_048941	82%
b785R	No Hit Found			
b786R	a587R [Paramecium bursaria Chlorella virus 1]	3.2E-24	NP_048943	81%
b787R	No Hit Found			
b788R	a588R [Paramecium bursaria Chlorella virus 1]	6.3E-23	NP_048944	50%
b789R	No Hit Found			
b790L	a589L [Paramecium bursaria Chlorella virus 1]	9.1E-11	NP_048945	52%
B791R	A592R [Paramecium bursaria Chlorella virus 1]	5.1E-24	NP_048948	94%
B792R	a593R [Paramecium bursaria Chlorella virus 1]	8.9E-12	NP_048949	37%
b793L	a595L [Paramecium bursaria Chlorella virus 1] &	3.5E-10	NP_048951	67%
b794R	No Hit Found			
B795R	similar to Vibrio fischeri dCMP deaminase, corresponds to Swiss-Prot Accession Number P33968 [Paramecium bursaria Chlorella virus 1]	2.6E-66	NP_048952	83%
	deoxycytidylate deaminase [Bacteriophage phi JL001]	1.8E-27	YP_223954	47%
	COG2131: Deoxycytidylate deaminase [Magnetospirillum magnetotacticum MS-1]	1.9E-21	ZP_00052863	39%
B796L	similar to tomato histidine decarboxylase, corresponds to Swiss-Prot Accession Number P54772 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048954	84%
	putative histidine decarboxylase [Oryza sativa (japonica cultivar-group)]	1.7E-56	AAP51789	38%
	COG0076: Glutamate decarboxylase and related PLP-dependent proteins [Nostoc punctiforme PCC 73102]	5.0E-56	ZP_00106716	37%
b797R	No Hit Found			
B798R	A267L [Paramecium bursaria Chlorella virus 1]	1.2E-44	NP_048621	36%
	Lys-, Glu-rich [Paramecium bursaria Chlorella virus 1]	9.8E-07	NP_048846	35%
	Lys-, Arg-rich; contains eukaryotic putative RNA-binding region RNP-1 signature; similar to PBCV-1 ORF A267L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	8.3E-06	NP_048834	35%
b799R	No Hit Found			
B800R	A601R [Paramecium bursaria Chlorella virus 1]	7.6E-26	NP_048957	70%
B801L	A602L [Paramecium bursaria Chlorella virus 1]	7.2E-16	NP_048958	73%
B802R	A603R [Paramecium bursaria Chlorella virus 1]	1.2E-39	NP_048959	74%
B803L	A604L [Paramecium bursaria Chlorella virus 1]	2.7E-15	NP_048960	36%
b804L	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
B805R	Lys-, Arg-rich; contains eukaryotic putative RNA-binding region RNP-1 signature; similar to PBCV-1 ORF A267L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	2.4E-81	NP_048834	45%
	Lys-, Glu-rich [Paramecium bursaria Chlorella virus 1]	4.4E-80	NP_048846	44%
	unknown [Acanthamoeba polyphaga mimivirus]	1.6E-16	YP_142777	27%
b806L	hypothetical protein with signal peptide [Cryptosporidium parvum]	6.4E-04	EAK89294	30%
b807R	No Hit Found			
b808L	No Hit Found			
b809L	No Hit Found			
b810L	A605L [Paramecium bursaria Chlorella virus 1]	1.9E-63	NP_048961	77%
	a606L [Paramecium bursaria Chlorella virus 1]	5.6E-15	NP_048962	63%
B811R	No Hit Found			
B812R	contains 3 ankyrin repeat-like elements; similar to Drosophila ankyrin, corresponds to GenBank Accession Number L35601 [Paramecium bursaria Chlorella virus 1]	6.8E-107	NP_048963	82%
	A608R [Paramecium bursaria Chlorella virus 1]	1.5E-61	NP_048964	72%
	ankyrin repeat protein, putative [Entamoeba histolytica HM-1:IMSS]	7.5E-05	XP_650635	26%
B813L	PBCV-1 histone H3-Lys 27 methyltransferase (vSET) [Paramecium bursaria Chlorella virus 1]	2.1E-52	NP_048968	78%
	Nuclear protein SET [Pelodictyon phaeoclathratiforme BU-1]	1.0E-09	ZP_00588496	33%
	hypothetical protein CT0952 [Chlorobium tepidum TLS]	1.0E-09	NP_661845	35%
b814L	No Hit Found			
b815R	a613R [Paramecium bursaria Chlorella virus 1]	2.6E-18	NP_048969	52%
B816L	RPQT-like (9x) [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048970	73%
	similar to bovine cylcin I, corresponds to Swiss-Prot Accession Number P35662 [Paramecium bursaria Chlorella virus 1]	6.7E-81	NP_048632	50%
	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	9.7E-72	NP_048636	44%
b817R	a616R [Paramecium bursaria Chlorella virus 1]	2.0E-26	NP_048972	63%
B818R	similar to Variola virus orf E10L, corresponds to Swiss-Prot Accession Number P33801 [Paramecium bursaria Chlorella virus 1]	5.4E-146	NP_048973	79%
	S/T protein kinase, similar to Paramecium bursaria chlorella virus 1 A617R [Acanthamoeba polyphaga mimivirus]	6.9E-16	YP_142754	30%
	380R [Chilo iridescent virus]	7.6E-07	AAK82240	27%
B819L	A618L [Paramecium bursaria Chlorella virus 1]	1.6E-39	NP_048974	83%
B820L	A619L [Paramecium bursaria Chlorella virus 1]	2.4E-54	NP_048975	56%
b821R	No Hit Found			
b822R	No Hit Found			
B823L	similar to Synechocystis orf 90, corresponds to GenBank Accession Number D90902 [Paramecium bursaria Chlorella virus 1]	1.1E-37	NP_048976	89%
	A635R [Paramecium bursaria Chlorella virus 1]	7.3E-13	NP_048991	48%
B824L	A621L [Paramecium bursaria Chlorella virus 1]	1.2E-55	NP_048977	88%
B825L	similar to Simulium iridescent virus capsid protein, corresponds to Swiss-Prot Accession Number P22166 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048978	94%
	putative capsid protein [Chlorella virus]	0.0E+00	AAC27493	73%
	hypothetical major capsid protein [Heterosigma akashiwo virus 01]	9.5E-25	BAE06835	69%
B826L	similar to frog ubiquitin-like fusion protein, corresponds to Accession Number JN0673 [Paramecium bursaria Chlorella virus 1]	3.5E-31	NP_048979	95%
	putative protein [Arabidopsis thaliana]	1.4E-08	CAB81349	62%
	putative zinc finger protein [Oryza sativa (japonica cultivar-group)]	1.2E-07	XP_469956	51%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
b827R	No Hit Found			
B828R	A624R [Paramecium bursaria Chlorella virus 1]	9.4E-53	NP_048980	86%
	conserved hypothetical protein [Listeria monocytogenes str. 1/2a F6854]	3.5E-07	ZP_00234461	28%
	hypothetical protein lin0598 [Listeria innocua Clip11262]	6.0E-07	NP_469941	28%
B829R	similar to Synechocystis transposase, corresponds to GenBank Accession Number D90909 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048981	100%
	putative transposase [Chlorella virus]	0.0E+00	AAU06281	85%
	transposase [Nostoc sp. PCC 7120]	9.4E-20	BAB78230	30%
b830L	a626L [Paramecium bursaria Chlorella virus 1]	5.5E-40	NP_048982	100%
B831R	similar to Chlorella virus PBCV-1 ORF A231L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048983	93%
	contains ATP/GTP-binding motif A [Paramecium bursaria Chlorella virus 1]	3.7E-80	NP_048579	45%
B832R	similar to Schizosaccharomyces ribonucleotide reductase M1 chain, corresponds to Swiss-Prot Accession Number P36602 [Paramecium bursaria Chlorella virus 1]	1.1E-178	NP_048985	91%
	Protein splicing (intein) site [Crocospaera watsonii WH 8501]	3.8E-128	ZP_00514074	31%
	ribonucleoside-diphosphate reductase large chain homolog - Chilo iridescent virus	1.1E-127	AAB94427	34%
b833R	a630R [Paramecium bursaria Chlorella virus 1]	5.4E-19	NP_048986	67%
b834R	No Hit Found			
b835L	No Hit Found			
b836R	No Hit Found			
b837L	No Hit Found			
b838R	No Hit Found			
B839R	A633R [Paramecium bursaria Chlorella virus 1]	8.0E-52	NP_048989	82%
b840L	A634L [Paramecium bursaria Chlorella virus 1]	7.3E-69	NP_048990	91%
B841R	A635R [Paramecium bursaria Chlorella virus 1]	2.6E-42	NP_048991	97%
	similar to Synechocystis orf 90, corresponds to GenBank Accession Number D90902 [Paramecium bursaria Chlorella virus 1]	4.9E-12	NP_048976	43%
B842R	A636R [Paramecium bursaria Chlorella virus 1]	4.6E-15	NP_048992	73%
B843R	A637R [Paramecium bursaria Chlorella virus 1]	8.7E-62	NP_048993	82%
B844R	PBCV-1 Agmatine iminohydrolase [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_048994	96%
	hypothetical protein A638R - Chlorella virus PBCV-1	0.0E+00	T18140	96%
	hypothetical protein lmo0038 [Listeria monocytogenes EGD-e]	2.8E-104	NP_463571	52%
b845R	a640R [Paramecium bursaria Chlorella virus 1]	1.3E-36	NP_048996	96%
b846L	a641L [Paramecium bursaria Chlorella virus 1]	5.9E-34	NP_048997	95%
B847R	Gln-rich; KQQ (6X) [Paramecium bursaria Chlorella virus 1]	3.6E-79	NP_048999	63%
	A642R [Paramecium bursaria Chlorella virus 1]	1.8E-25	NP_048998	100%
b848L	No Hit Found			
B849R	A644R [Paramecium bursaria Chlorella virus 1]	8.8E-33	NP_049000	80%
	unknown [Acanthamoeba polyphaga mimivirus]	2.2E-07	YP_142763	33%
B850L	PBCV-1 33kd peptide [Paramecium bursaria Chlorella virus 1]	3.0E-06	NP_048641	28%
	GIY-YIG catalytic domain containing protein; possible intron encoded endonuclease [Bacillus cereus E33L]	4.0E-06	YP_085011	35%
	I-TevI homing endonuclease [Enterobacteria phage T4]	2.0E-05	AAD42521	32%
b851R	No Hit Found			
B852R	A645R [Paramecium bursaria Chlorella virus 1]	6.1E-55	NP_049001	81%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
B853L	A654L [Paramecium bursaria Chlorella virus 1]	1.3E-98	NP_049010	86%
	COG0454: Histone acetyltransferase HPA2 and related acetyltransferases [Trichodesmium erythraeum IMS101]	4.9E-08	ZP_00324949	29%
	EG:BACR25B3.5 [Drosophila melanogaster]	5.0E-05	CAB72290	25%
b854R	No Hit Found			
b855L	a655L [Paramecium bursaria Chlorella virus 1]	3.7E-09	NP_049011	54%
b856R	No Hit Found			
B857L	A656L [Paramecium bursaria Chlorella virus 1]	3.1E-43	NP_049012	71%
B858L	A659L [Paramecium bursaria Chlorella virus 1]	3.5E-47	NP_049015	77%
B859L	A662L [Paramecium bursaria Chlorella virus 1]	9.3E-75	NP_049018	77%
	unknown protein [Oryza sativa (japonica cultivar-group)]	1.7E-12	BAD87006	31%
	unknown protein [Arabidopsis thaliana]	2.8E-10	AAM13270	25%
B860L	similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank Accession Number D16505 [Paramecium bursaria Chlorella virus 1]	1.7E-41	NP_049020	59%
	similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank Accession Number D16505 [Paramecium bursaria Chlorella virus 1]	3.8E-33	NP_049021	48%
b861L	No Hit Found			
B862L	similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank Accession Number D16505 [Paramecium bursaria Chlorella virus 1] Chlorella virus PBCV-1	2.9E-70	NP_049021	86%
	similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank Accession Number D16505 [Paramecium bursaria Chlorella virus 1]	1.6E-31	NP_049020	52%
b863R	No Hit Found			
b864R	No Hit Found			
B865R	similar to Synechocystis ORF s111635, corresponds to GenBank Accession Number D90903 [Paramecium bursaria Chlorella virus 1]	2.4E-102	NP_049030	83%
	thymidilate synthase [Bacteriophage S-PM2]	7.1E-62	CAF34258	58%
	Td [Cyanophage P-SSM4]	1.9E-59	YP_214756	59%
b866R	No Hit Found			
B867L	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	2.0E-115	NP_048629	80%
	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	2.1E-109	NP_048807	76%
	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	7.0E-97	NP_049005	66%
b868R	a80L [Paramecium bursaria Chlorella virus 1]	1.6E-29	NP_048428	49%
	similar to PBCV-1 ORF a80L, encoded by GenBank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	1.7E-20	NP_048808	53%
	similar to PBCV-1 ORF a80L, corresponds to GenBank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	3.3E-19	NP_048630	57%
B869R	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	0.0E+00	NP_049032	87%
	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	6.3E-11	NP_048636	84%
	similar to bovine cylcin I, corresponds to Swiss-Prot Accession Number P35662 [Paramecium bursaria Chlorella virus 1]	3.4E-04	NP_048632	61%
b870L	a677L [Paramecium bursaria Chlorella virus 1]	7.2E-24	NP_049033	84%

ORF	PROTEIN HITS	E-VALUE	Accession Number	% ID
B871R	No Hit Found			
b872L	No Hit Found			
b873R	No Hit Found			
B874L	contains 10 ankyrin-like repeats; similar to human ankyrin, corresponds to Swiss-Prot Accession Number P16157 [Paramecium bursaria Chlorella virus 1]	2.4E-151	NP_049038	71%
	contains 4 ankyrin repeats; similar to reticulocyte ankyrin, corresponds to Swiss-Prot Accession Number P16157 [Paramecium bursaria Chlorella virus 1]	6.1E-14	NP_048353	48%
	NACHT domain protein, putative [Aspergillus fumigatus Af293]	9.2E-18	EAL87814	32%
b875R	a6L [Paramecium bursaria Chlorella virus 1]	1.4E-09	NP_048354	48%
b876R	No Hit Found			
b877R	No Hit Found			
B878L	A354R [Paramecium bursaria Chlorella virus 1]	2.9E-36	NP_048711	38%
	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081 [Paramecium bursaria Chlorella virus 1]	3.7E-15	NP_048779	29%
	hypothetical phage protein [Streptococcus pyogenes MGAS8232]	2.2E-04	AAL98037	32%
b879R	No Hit Found			
B880L	contains 4 ankyrin repeats; similar to reticulocyte ankyrin, corresponds to Swiss-Prot Accession Number P16157 [Paramecium bursaria Chlorella virus 1]	1.4E-77	NP_048353	71%
	contains 10 ankyrin-like repeats; similar to human ankyrin, corresponds to Swiss-Prot Accession Number P16157 [Paramecium bursaria Chlorella virus 1]	2.4E-32	NP_049038	50%
	SecG [Dictyostelium discoideum]	2.7E-07	XP_637214	34%
b881R	a6L [Paramecium bursaria Chlorella virus 1]	5.1E-06	NP_048354	48%
b882R	identical to PBCV-1 terminal repeat ORF, corresponds to GenBank Accession Number M55319 [Paramecium bursaria Chlorella virus 1]	2.8E-20	NP_049046	67%
B883L	identical to Chlorella virus PBCV-1 terminal repeat ORF A3R, corresponds to GenBank Accession Number M55319 [Paramecium bursaria Chlorella virus 1]	2.7E-23	NP_049045	61%
B884L	No Hit Found			
b885L	No Hit Found			
B886R	No Hit Found			

2.6.2.3 Results from analysis with the Protein Families (Pfam) Database

NY-2A Pfam			
ORF	Pfam	E-value	Description
B8R	PF00145.7	1.4e-60	C-5 cytosine-specific DNA methylase
B10R	PF01555.8	2.2e-28	DNA methylase
B18L	PF00023.17	1.2e-13	Ankyrin repeat
B22R	PF01844.12	2.4e-07	HNH endonuclease
B24L	PF00023.17	9.2e-09	Ankyrin repeat
B33R	PF07150.1	3.5e-11	Protein of unknown function (DUF1390)
B34L	PF06598.1	7.8e-13	Chlorovirus glycoprotein repeat
B40L	PF06598.1	5.0e-15	Chlorovirus glycoprotein repeat
B47L	PF06598.1	1.3e-15	Chlorovirus glycoprotein repeat
B55L	PF06598.1	3.5e-15	Chlorovirus glycoprotein repeat
B59R	PF04451.2	4.2e-93	Iridovirus major capsid protein
B61R	PF01844.12	1.5e-04	HNH endonuclease
B68L	PF03931.4	2.6e-16	Skp1 family, tetramerisation domain
	PF01466.8	5.8e-35	Skp1 family, dimerisation domain
B73L	PF00004.17	4.8e-14	ATPase family associated with various cellular activities (AAA)
B75L	PF03009.7	5.1e-61	Glycerophosphoryl diester phosphodiesterase family
B76L	PF03013.4	2.8e-63	Pyrimidine dimer DNA glycosylase
B80L	PF07282.1	1.3e-11	Putative transposase DNA-binding domain
B83L	PF00376.12	7.5e-04	MerR family regulatory protein
	PF00239.10	1.5e-14	Resolvase, N terminal domain
	PF00145.7	6.3e-25	C-5 cytosine-specific DNA methylase
B116R	PF00795.11	3.7e-47	Carbon-nitrogen hydrolase
B118R	PF07150.1	2.3e-05	Protein of unknown function (DUF1390)
B120R	PF07150.1	1.9e-184	Protein of unknown function (DUF1390)
B126R	PF03171.9	7.6e-14	2OG-Fe(II) oxygenase superfamily
B133R	PF01844.12	6.6e-08	HNH endonuclease
B137L	PF00722.9	3.1e-07	Glycosyl hydrolases family 16
B139R	PF03142.4	6.3e-06	Chitin synthase
B143R	PF00310.10	3.2e-24	Glutamine amidotransferases class-II
	PF01380.11	3.7e-29	SIS domain
B148R	PF03919.5	2.7e-24	mRNA capping enzyme, C-terminal domain
	PF01331.9	5.8e-59	mRNA capping enzyme, catalytic domain
B150L	PF00443.17	3.8e-35	Ubiquitin carboxyl-terminal hydrolase
B154L	PF08271.1	3.6e-04	TFIIB zinc-binding
B163R	PF01370.10	1.3e-94	NAD dependent epimerase/dehydratase family
B165R	PF01844.12	1.2e-07	HNH endonuclease
B170R	PF06598.1	4.0e-15	Chlorovirus glycoprotein repeat
B173L	PF01844.12	3.1e-05	HNH endonuclease
B175L	PF07500.3	5.7e-28	Transcription factor S-II (TFIIS), central domain
	PF01096.8	9.7e-10	Transcription factor S-II (TFIIS)
B185L	PF01541.13	5.6e-12	GIY-YIG catalytic domain
B203R	PF04851.5	2.0e-10	Type III restriction enzyme, res subunit
B206L	PF07453.2	5.7e-05	NUMOD1 domain
	PF01541.13	2.0e-12	GIY-YIG catalytic domain

NY-2A Pfam			
ORF	Pfam	E-value	Description
B218R	PF01844.12	8.8e-07	HNH endonuclease
B222R	PF00185.12	1.7e-34	Aspartate/ornithine carbamoyltransferase, Asp/Orn binding domain
	PF02729.9	3.9e-51	Aspartate/ornithine carbamoyltransferase, carbamoyl-P binding domain
B226L	PF01734.11	1.3e-46	Patatin-like phospholipase
B236L	PF02086.5	9.7e-84	D12 class N6 adenine-specific DNA methyltransferase
B239R	PF00704.16	1.1e-07	Glycosyl hydrolases family 18
	PF00553.9	3.8e-05	Cellulose binding domain
B246R	PF01541.13	1.8e-09	GIY-YIG catalytic domain
	PF07453.2	5.3e-09	NUMOD1 domain
B249R	PF00136.10	2.8e-46	DNA polymerase family B
	PF03104.8	8.9e-111	DNA polymerase family B, exonuclease domain
B253R	PF00136.10	7.7e-29	DNA polymerase family B
B261L	PF00705.8	2.5e-17	Proliferating cell nuclear antigen, N-terminal domain
	PF02747.5	4.1e-11	Proliferating cell nuclear antigen, C-terminal domain
B268L	PF01753.8	1.9e-09	MYND finger
B271R	PF00383.12	4.7e-06	Cytidine and deoxycytidylate deaminase zinc-binding region
B278R	PF00278.11	1.9e-16	Pyridoxal-dependent decarboxylase, C-terminal sheet domain
	PF02784.6	2.3e-75	Pyridoxal-dependent decarboxylase, pyridoxal binding domain
B286R	PF01541.13	8.7e-05	GIY-YIG catalytic domain
B305R	PF06408.1	0	Homospermidine synthase
B316R	PF08148.1	9.5e-04	DSHCT (NUC185) domain
	PF00270.16	2.0e-27	DEAD/DEAH box helicase
B322R	PF00967.7	1.4e-68	Barwin family
B323R	PF00023.17	3.1e-08	Ankyrin repeat
B324L	PF01844.12	1.5e-08	HNH endonuclease
B331R	PF00069.14	4.6e-52	Protein kinase domain
B334R	PF07282.1	5.4e-23	Putative transposase DNA-binding domain
	PF01385.9	2.0e-04	Probable transposase
B336R	PF07885.3	5.7e-08	Ion channel
B339L	PF07150.1	1.9e-178	Protein of unknown function (DUF1390)
B346L	PF01541.13	4.9e-15	GIY-YIG catalytic domain
	PF07453.2	9.7e-10	NUMOD1 domain
B357L	PF03713.3	3.7e-18	Domain of unknown function (DUF305)
B370L	PF01844.12	3.3e-04	HNH endonuclease
B371L	PF02275.7	5.2e-28	Linear amide C-N hydrolases, cholylglycine hydrolase family
B378L	PF07282.1	1.3e-11	Putative transposase DNA-binding domain
B381L	PF00376.12	7.5e-04	MerR family regulatory protein
	PF00239.10	1.5e-14	Resolvase, N terminal domain
B388L	PF00069.14	6.8e-50	Protein kinase domain
B389R	PF01541.13	3.8e-14	GIY-YIG catalytic domain
	PF07453.2	3.0e-09	NUMOD1 domain
B393L	PF01374.8	1.1e-110	Glycosyl hydrolase family 46
B395L	PF01370.10	3.7e-91	NAD dependent epimerase/dehydratase family
B399R	PF02086.5	3.5e-63	D12 class N6 adenine-specific DNA methyltransferase
B411L	PF00145.7	3.3e-26	C-5 cytosine-specific DNA methylase
B418R	PF07669.1	5.9e-04	Eco57I restriction endonuclease

NY-2A Pfam			
ORF	Pfam	E-value	Description
B430L	PF00782.9	4.6e-56	Dual specificity phosphatase, catalytic domain
B433L	PF07453.2	1.1e-05	NUMOD1 domain
	PF01541.13	4.2e-11	GIY-YIG catalytic domain
B440L	PF01541.13	3.1e-11	GIY-YIG catalytic domain
B446R	PF01844.12	1.5e-07	HNH endonuclease
B465R	PF03720.4	2.9e-04	UDP-glucose/GDP-mannose dehydrogenase family, UDP binding domain
	PF03721.4	2.4e-11	UDP-glucose/GDP-mannose dehydrogenase family, NAD binding domain
	PF00984.9	5.7e-09	UDP-glucose/GDP-mannose dehydrogenase family, central domain
B469L	PF01522.10	1.1e-13	Polysaccharide deacetylase
B472R	PF03142.4	9.8e-06	Chitin synthase
B477R	PF00023.17	6.9e-09	Ankyrin repeat
B480L	PF01607.12	3.0e-06	Chitin binding Peritrophin-A domain
B499L	PF01541.13	1.6e-09	GIY-YIG catalytic domain
B529R	PF04451.2	8.9e-15	Iridovirus major capsid protein
B533L	PF01607.12	1.0e-05	Chitin binding Peritrophin-A domain
B547R	PF00075.12	8.7e-24	RNase H
B566R	PF00145.7	1.1e-56	C-5 cytosine-specific DNA methylase
B585L	PF04451.2	6e-231	Iridovirus major capsid protein
B592L	PF00462.11	5.8e-06	Glutaredoxin
B598L	PF01844.12	9.8e-06	HNH endonuclease
B602L	PF07453.2	6.6e-04	NUMOD1 domain
	PF01541.13	7.9e-11	GIY-YIG catalytic domain
B606L	PF03109.6	1.2e-42	ABC1 family
B617L	PF04451.2	2.5e-225	Iridovirus major capsid protein
B618R	PF04724.3	3.0e-16	Glycosyltransferase family 17
B623L	PF03288.5	1.6e-05	Poxvirus D5 protein-like
B628R	PF00636.13	2.4e-41	RNase3 domain
	PF00035.12	3.6e-17	Double-stranded RNA binding motif
B629R	PF01541.13	3.0e-11	GIY-YIG catalytic domain
B630R	PF04777.2	4.4e-36	Erv1 / Alr family
B641R	PF00268.11	2.2e-128	Ribonucleotide reductase, small chain
B677R	PF00097.12	9.3e-05	Zinc finger, C3HC4 type (RING finger)
B681L	PF00145.7	1.4e-16	C-5 cytosine-specific DNA methylase
B697R	PF00145.7	3.2e-18	C-5 cytosine-specific DNA methylase
B699L	PF00240.12	1.5e-40	Ubiquitin family
B702L	PF01385.9	4.3e-04	Probable transposase
	PF07282.1	5.3e-23	Putative transposase DNA-binding domain
B711L	PF07282.1	1.3e-11	Putative transposase DNA-binding domain
B715L	PF00376.12	7.5e-04	MerR family regulatory protein
	PF00239.10	1.5e-14	Resolvase, N terminal domain
B720L	PF07150.1	4.8e-05	Protein of unknown function (DUF1390)
B734R	PF01068.10	2.3e-25	ATP dependent DNA ligase domain
B738L	PF00176.12	1.9e-25	SNF2 family N-terminal domain
	PF00271.18	7.8e-16	Helicase conserved C-terminal domain
B741L	PF00692.9	9.9e-50	dUTPase
B744L	PF01171.10	3.2e-14	PP-loop family

NY-2A Pfam			
ORF	Pfam	E-value	Description
B748L	PF04451.2	3.1e-189	Iridovirus major capsid protein
B753L	PF07453.2	2.4e-07	NUMOD1 domain
B767L	PF00705.8	2.1e-14	Proliferating cell nuclear antigen, N-terminal domain
	PF02747.5	2.5e-07	Proliferating cell nuclear antigen, C-terminal domain
B769R	PF00145.7	4.2e-64	C-5 cytosine-specific DNA methylase
B774R	PF02086.5	2.2e-83	D12 class N6 adenine-specific DNA methyltransferase
B781L	PF00521.9	2.7e-71	DNA gyrase/topoisomerase IV, subunit A-DNA topoisomerase II
	PF00204.12	7.9e-12	DNA gyrase B-DNA topoisomerase II
	PF02518.13	8.6e-07	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase-DNA topo II
B795R	PF00383.12	1.1e-31	Cytidine and deoxycytidylate deaminase zinc-binding region
B796L	PF00282.8	2.3e-14	Pyridoxal-dependent decarboxylase conserved domain
B798R	PF07453.2	2.2e-06	NUMOD1 domain
B813L	PF00856.16	1.7e-05	SET domain
B825L	PF04451.2	1.8e-277	Iridovirus major capsid protein
B826L	PF01428.6	1.3e-22	AN1-like Zinc finger
B829R	PF01385.9	7.7e-05	Probable transposase
	PF07282.1	8.8e-24	Putative transposase DNA-binding domain
B832R	PF00317.10	2.1e-31	Ribonucleotide reductase, all-alpha domain
	PF03477.5	7.8e-21	ATP cone domain
	PF02867.4	6.1e-211	Ribonucleotide reductase, barrel domain
B844R	PF04371.4	5.5e-182	Porphyromonas-type peptidyl-arginine deiminase
B850L	PF01541.13	1.3e-12	GIY-YIG catalytic domain
B865R	PF02511.5	3.9e-101	Thymidylate synthase complementing protein
B867L	PF07150.1	6.3e-185	Protein of unknown function (DUF1390)
B874L	PF00023.17	2.5e-08	Ankyrin repeat
B878L	PF01844.12	4.3e-07	HNH endonuclease
B880L	PF00023.17	1.2e-09	Ankyrin repeat

2.6.2.4 Results of analysis with Clusters of Orthologous Groups (COGs)

NY-2A COGs			
ORF	E-value	COG	Description
Information Storage and Processing			
Transcription			
B628R	6.E-43	COG0571	dsRNA-specific ribonuclease
B154L	2.E-14	COG1405	Transcription initiation factor TFIIB, Brf1 subunit/TFIIB
Translation, ribosomal structure and biogenesis			
B418R	4.E-09	COG2890	Methylase of polypeptide chain release factors
RNA processing and modification			
B148R	1.E-20	COG5226	mRNA capping enzyme, guanylyltransferase (alpha) subunit
DNA replication, recombination, and repair			
B781L	e-107	COG0187	Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), B subunit
B566R	2.E-51	COG0270	Site-specific DNA methylase
B8R	2.E-46	COG0270	Site-specific DNA methylase
B769R	1.E-28	COG0270	Site-specific DNA methylase
B411L	5.E-25	COG0270	Site-specific DNA methylase
B681L	7.E-25	COG0270	Site-specific DNA methylase
B697R	1.E-24	COG0270	Site-specific DNA methylase
B88L	3.E-24	COG0270	Site-specific DNA methylase
B547R	2.E-10	COG0328	Ribonuclease HI
B774R	5.E-44	COG0338	Site-specific DNA methylase
B399R	2.E-35	COG0338	Site-specific DNA methylase
B249R	9.E-68	COG0417	DNA polymerase elongation subunit (family B)
B261L	3.E-11	COG0592	DNA polymerase sliding clamp subunit (PCNA homolog)
B767L	1.E-08	COG0592	DNA polymerase sliding clamp subunit (PCNA homolog)
B702L	3.E-18	COG0675	Transposase and inactivated derivatives
B829R	4.E-18	COG0675	Transposase and inactivated derivatives
B334R	1.E-17	COG0675	Transposase and inactivated derivatives
B378L	4.E-09	COG0675	Transposase and inactivated derivatives
B80L	4.E-09	COG0675	Transposase and inactivated derivatives
B711L	9.E-09	COG0675	Transposase and inactivated derivatives
B10R	4.E-13	COG0863	DNA modification methylase
B734R	3.E-12	COG1793	ATP-dependent DNA ligase
B381L	7.E-35	COG2452	Predicted site-specific integrase-resolvase
B715L	7.E-35	COG2452	Predicted site-specific integrase-resolvase
B83L	7.E-35	COG2452	Predicted site-specific integrase-resolvase
B236L	5.E-53	COG3392	Adenine-specific DNA methylase
B316R	8.E-99	COG4581	Superfamily II RNA helicase
Translation, ribosomal structure and biogenesis/ Nucleotide transport and metabolism			
B641R	8.E-64	COG0208	Ribonucleotide reductase, beta subunit
B832R	1.E-81	COG0209	Ribonucleotide reductase, alpha subunit
B222R	2.E-84	COG0540	Aspartate carbamoyltransferase, catalytic chain
B741L	2.E-35	COG0756	dUTPase
B865R	1.E-33	COG1351	Predicted alternative thymidylate synthase

NY-2A COGs			
ORF	E-value	COG	Description
B795R	7.E-24	COG2131	Deoxycytidylate deaminase
Transcription / DNA replication, recombination, and repair			
B738L	2.E-37	COG0553	Superfamily II DNA/RNA helicases, SNF2 family
B203R	2.E-31	COG1061	DNA or RNA helicases of superfamily II
Cellular Processes			
Posttranslational modification, protein turnover, chaperones			
B73L	2.E-09	COG0464	ATPases of the AAA+ class
B592L	1.E-06	COG0695	Glutaredoxin and related proteins
B630R	1.E-09	COG5054	Mitochondrial sulfhydryl oxidase involved in the biogenesis of cytosolic Fe/S proteins
B68L	2.E-24	COG5201	SCF ubiquitin ligase, SKP1 component
B150L	5.E-06	COG5533	Ubiquitin C-terminal hydrolase
Cell envelope biogenesis, outer membrane			
B736L	2.E-11	COG0438	Glycosyltransferase
B143R	e-180	COG0449	Glucosamine 6-phosphate synthetase, contains amidotransferase and phosphosugar isomerase domains
B465R	4.E-56	COG0677	UDP-N-acetyl-D-mannosaminuronate dehydrogenase
B443R	2.E-09	COG0810	Periplasmic protein TonB, links inner and outer membranes
B667L	3.E-09	COG0810	Periplasmic protein TonB, links inner and outer membranes
B70R	2.E-08	COG0810	Periplasmic protein TonB, links inner and outer membranes
B163R	e-139	COG1089	GDP-D-mannose dehydratase
B139R	9.E-12	COG1215	Glycosyltransferases, probably involved in cell wall biogenesis
B472R	5.E-06	COG1215	Glycosyltransferases, probably involved in cell wall biogenesis
B371L	1.E-45	COG3049	Penicillin V acylase and related amidases
B258R	4.E-06	COG3064	Membrane protein involved in colicin uptake
Cell division and chromosome partitioning			
B744L	3.E-20	COG0037	Predicted ATPase of the PP-loop superfamily implicated in cell cycle control
Signal transduction mechanisms			
B430L	3.E-08	COG2453	Predicted protein-tyrosine phosphatase
Cell envelope biogenesis, outer membrane / Carbohydrate transport and metabolism			
B395L	1.E-40	COG0451	Nucleoside-diphosphate-sugar epimerases
Signal transduction mechanisms / Transcription / DNA replication, recombination, and repair			
B331R	7.E-23	COG0515	Serine/threonine protein kinase
B388L	2.E-19	COG0515	Serine/threonine protein kinase
B816L	8.E-07	COG0515	Serine/threonine protein kinase
Metabolism			
Carbohydrate transport and metabolism			
B469L	3.E-11	COG0726	Predicted xylanase/chitin deacetylase
B137L	4.E-12	COG2273	Beta-glucanase/Beta-glucan synthetase
Nucleotide transport and metabolism / Translation, ribosomal structure and biogenesis			
B271R	3.E-08	COG0590	Cytosine/adenosine deaminases
Energy production and conversion			

NY-2A COGs			
ORF	E-value	COG	Description
B75L	4.E-29	COG0584	Glycerophosphoryl diester phosphodiesterase
Amino acid transport and metabolism			
B278R	2.E-52	COG0019	Diaminopimelate decarboxylase
B796L	2.E-31	COG0076	Glutamate decarboxylase and related PLP-dependent proteins
B844R	e-108	COG2957	Peptidylarginine deiminase and related enzymes
Secondary metabolites biosynthesis, transport, and catabolism			
B305R	4.E-86	COG5310	Homospermidine synthase
Lipid metabolism			
B354L	3.E-19	COG2267	Lysophospholipase

Poorly Characterized			
General function prediction only			
B116R	3.E-47	COG0388	Predicted amidohydrolase
B606L	8.E-47	COG0661	Predicted unusual protein kinase
B365L	8.E-06	COG0661	Predicted unusual protein kinase
B880L	2.E-16	COG0666	FOG: Ankyrin repeat
B18L	3.E-16	COG0666	FOG: Ankyrin repeat
B874L	3.E-16	COG0666	FOG: Ankyrin repeat
B24L	1.E-14	COG0666	FOG: Ankyrin repeat
B477R	2.E-07	COG0666	FOG: Ankyrin repeat
B374L	8.E-07	COG0666	FOG: Ankyrin repeat
B226L	1.E-18	COG1752	Predicted esterase of the alpha-beta hydrolase superfamily
B813L	1.E-08	COG2940	Proteins containing SET domain
B623L	3.E-12	COG3378	Predicted ATPase
B416R	2.E-10	COG4106	Trans-aconitate methyltransferase
B16L	4.E-12	COG4123	Predicted O-methyltransferase
B567L	4.E-10	COG4123	Predicted O-methyltransferase
B230L	6.E-10	COG4123	Predicted O-methyltransferase
Function Unknown			
B847R	2.E-06	COG4487	Uncharacterized protein conserved in bacteria
B828R	4.E-06	COG4852	Predicted membrane protein
B322R	2.E-06	COG5373	Predicted membrane protein
B554R	8.E-06	COG5373	Predicted membrane protein

2.6.3 MT325 Genome

2.6.3.1 General Characteristics of Open Reading Frames

ORF	ORF Location		% A+T (-50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
M1L	1052	381	72%	6	47	241	158	161	112	672	223	26252	6.7	9620
m2R	406	1011	64%	1	48	94	150	138	224	606	201	21985	11.1	42480
M3L	1521	1144	66%	5	58	94	103	117	64	378	125	13987	9.9	27880
m4R	1345	1569	38%	1	52	42	54	64	65	225	74	8031	7	90
M5L	2294	1752	80%	6	42	188	96	131	128	543	180	20422	10.2	26060
m6L	1984	1772	56%	4	45	66	44	52	51	213	70	8218	11.4	7030
M7L	2839	2327	70%	4	44	145	108	118	142	513	170	19445	6.1	7830
m8R	2751	3122	54%	3	42	110	74	81	107	372	123	14084	9.9	10930
M9L	3736	2927	80%	4	45	236	173	189	212	810	269	31503	6	52270
M10L	4146	3787	66%	5	45	104	76	86	94	360	119	13198	10	9620
M11L	4524	4234	82%	5	43	75	63	62	91	291	96	10635	10.5	12090
M12R	4552	5022	88%	1	42	144	88	109	130	471	156	17893	7.7	41400
m13L	5323	5069	52%	4	48	62	54	69	70	255	84	9563	6.5	19660
M14R	5076	5870	74%	3	48	205	207	172	211	795	264	29727	5	18580
M15L	9734	5874	78%	6	50	1178	975	942	766	3861	1286	141486	11.5	45370
m16R	8002	8274	58%	1	47	54	65	64	90	273	90	9630	10.8	2680
m17L	9445	9143	52%	4	46	101	74	64	64	303	100	11874	12.3	1340
m18R	9473	9685	56%	2	52	28	63	47	75	213	70	7725	7.2	7120
M19L	12489	9766	76%	5	48	699	586	731	708	2724	907	102971	7.9	1E+05
m20R	10550	10783	56%	2	47	62	65	46	61	234	77	8548	10.3	9530
m21L	10985	10758	46%	6	49	58	45	66	59	228	75	8825	12.9	11470
m22R	12278	12505	56%	2	53	57	56	64	51	228	75	8223	10.5	1340
M23L	12664	12443	82%	4	41	63	48	42	69	222	73	8574	8.7	10960
M24L	12885	12682	72%	5	31	56	29	35	84	204	67	7955	9	9680
m25R	12808	13548	76%	1	42	242	146	168	185	741	246	27970	8	19380
M26L	14628	13588	64%	5	53	245	281	267	248	1041	346	38209	7.1	18670
m27L	13919	13680	56%	6	54	59	68	62	51	240	79	8882	12.5	28510
m28R	13883	14080	50%	2	51	45	51	49	53	198	65	7804	11.1	20910
m29R	13984	14634	52%	1	53	163	174	173	141	651	216	23948	7.7	21150
M30R	14798	15610	78%	2	49	158	212	184	259	813	270	29490	7.2	26150
m31L	15378	14908	54%	5	49	149	114	117	91	471	156	17528	12.2	39290
M32L	16521	15832	78%	5	44	240	123	183	144	690	229	26055	10.1	33800
m33R	16234	16554	62%	1	45	73	82	64	102	321	106	11945	6.2	3870
M34L	17228	16584	64%	6	50	164	180	144	157	645	214	24468	8.6	38750
m35R	16704	16922	56%	3	47	55	48	56	60	219	72	7877	9.7	7090
m36R	17212	17418	42%	1	41	54	37	48	68	207	68	8218	11.7	10870
M37R	17363	19144	68%	2	48	458	431	422	471	1782	593	64752	6	39130

ORF	ORF Location		% A+T (-50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
m38L	17670	17446	50%	5	47	62	50	56	57	225	74	8356	5	17130
m39L	18114	17797	56%	5	44	91	72	69	86	318	105	11261	8.7	8310
m40L	18421	18065	58%	4	45	99	78	81	99	357	118	13663	9.9	10930
m41L	18699	18382	40%	5	50	74	82	78	84	318	105	11220	6.5	8310
M42R	19188	22058	84%	3	44	832	609	657	773	2871	956	96590	5.8	43270
m43L	19522	19253	52%	4	45	75	68	54	73	270	89	9826	4.5	1280
m44L	19975	19670	56%	4	43	79	59	73	95	306	101	11051	4.4	2590
m45L	21577	21278	62%	4	38	85	64	51	100	300	99	10862	4.5	5720
m46R	21776	21976	66%	2	52	41	58	47	55	201	66	7550	10.8	9530
M47R	22096	26499	76%	1	44	1302	951	972	1179	4404	1467	151714	6.6	1E+05
m48L	22622	22182	58%	6	43	113	95	95	138	441	146	16109	4.3	9560
m49L	22928	22629	56%	6	42	91	59	66	84	300	99	10982	8.4	2590
m50L	23345	23043	50%	6	47	73	76	65	89	303	100	10829	4.7	1340
m51L	24366	24073	54%	5	47	78	80	58	78	294	97	10642	4.3	6550
m52L	24653	24363	62%	6	46	79	68	65	79	291	96	10452	3.9	7060
m53L	24923	24684	72%	6	45	71	51	57	61	240	79	8474	8.4	3870
m54L	25580	25371	58%	6	38	62	38	42	68	210	69	7720	9.2	13940
M55R	26539	30924	84%	1	45	1216	938	1054	1178	4386	1461	149620	5.6	1E+05
m56L	27371	26625	52%	6	46	191	173	173	210	747	248	26896	4.7	9050
m57R	26894	27190	52%	2	47	79	71	68	79	297	98	11612	6.9	13970
m58L	27707	27486	58%	6	47	49	56	49	68	222	73	7951	4.2	1310
m59L	28397	28053	50%	6	42	102	81	65	97	345	114	12389	4.6	90
m60L	30984	30754	80%	5	45	58	54	50	69	231	76	8620	7.8	12240
M61R	31058	35461	78%	2	46	1233	949	1060	1162	4404	1467	150569	4.9	1E+05
m62L	31584	31144	58%	5	47	106	112	97	126	441	146	15830	4	13940
m63L	32307	31630	56%	5	47	163	170	152	193	678	225	23961	4.7	17840
m64L	32916	32572	50%	5	43	101	83	66	95	345	114	12443	4.4	90
m65L	34177	33803	68%	4	45	112	88	79	96	375	124	14507	12.2	17930
m66L	34977	34777	56%	5	50	43	45	55	58	201	66	6987	9.5	5690
m67L	35511	35080	64%	5	45	123	103	91	115	432	143	15511	10.1	9050
m68L	35587	35114	58%	4	45	130	111	101	132	474	157	18212	8.4	28950
m69R	35403	36701	44%	3	46	416	276	327	280	1299	432	49569	9.8	92870
M70L	36685	36329	80%	4	46	71	89	77	120	357	118	13451	9.6	15280
M71R	36740	41233	82%	2	45	1279	1031	1001	1183	4494	1497	156170	5.3	1E+05
m72L	37356	36862	62%	5	48	124	127	110	134	495	164	17829	10.8	1370
m73L	38289	38023	60%	5	49	60	66	66	75	267	88	9773	10.2	7000
m74L	38931	38662	56%	5	46	78	64	61	67	270	89	9736	4.4	5750
m75L	39174	38965	48%	5	45	51	52	42	65	210	69	7938	4.9	30
m76R	39316	39567	50%	1	47	64	69	49	70	252	83	9923	8.9	28570
m77L	40155	39952	70%	5	41	55	32	51	66	204	67	7004	9.4	7000
M78R	41286	42518	88%	3	50	321	335	284	293	1233	410	46454	6.3	79460

ORF	ORF Location		% A+T (-50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
m79R	41369	41665	48%	2	51	77	80	70	70	297	98	11475	12	17160
m80L	42909	42481	56%	5	45	71	89	104	165	429	142	16160	9.6	23650
M81R	42549	43172	80%	3	50	222	160	154	88	624	207	23404	11.5	20370
m82L	43103	42708	54%	6	54	47	117	98	134	396	131	14509	8.1	7510
M83R	43344	43601	78%	3	48	69	69	56	64	258	85	9761	7.5	16770
m84L	43646	43353	76%	6	47	76	62	77	79	294	97	10345	11	24040
M85R	43716	45287	76%	3	49	440	412	354	366	1572	523	57630	9.7	98410
M86R	45344	45934	80%	2	47	159	162	113	157	591	196	22099	8.1	21680
m87L	45838	45350	66%	4	49	130	101	138	120	489	162	18200	10.1	21120
M88R	45967	46608	76%	1	47	165	131	168	178	642	213	23541	4.7	15340
m89L	46601	45975	74%	6	47	176	165	128	158	627	208	22819	11.9	9620
M90L	47179	46682	84%	4	47	145	117	119	117	498	165	18421	12.2	12090
M91R	47173	48204	58%	1	43	334	213	226	259	1032	343	39378	9.1	65440
m92L	48155	47940	68%	6	39	63	45	40	68	216	71	8213	11	5690
M93L	49565	48432	68%	6	44	375	226	275	258	1134	377	43655	9.1	70730
m94L	49089	48859	56%	5	46	75	43	64	49	231	76	8967	10	6490
m95R	49042	49248	56%	1	50	35	57	46	69	207	68	7666	7.5	7030
m96R	49139	49408	56%	2	52	43	75	65	87	270	89	9819	8.3	1640
M97L	50625	49546	84%	5	45	336	242	241	261	1080	359	39958	8.7	48310
m98R	49636	50046	74%	1	46	102	99	89	121	411	136	15837	9.5	7830
M99R	50664	51173	80%	3	44	163	93	131	123	510	169	18218	8.2	7860
M100L	51754	51176	84%	4	41	187	130	109	153	579	192	22989	7.2	38810
M101R	51793	52632	90%	1	44	256	175	194	215	840	279	31106	6.3	18610
m102L	52463	52125	60%	6	48	84	82	82	91	339	112	11770	8.4	8340
M103L	53704	52814	66%	4	51	210	216	239	226	891	296	32913	7	45210
m104L	53598	53266	30%	5	46	83	69	83	98	333	110	12337	12.8	11410
m105R	53501	53743	46%	2	50	58	61	60	64	243	80	8770	10.6	2650
m106R	53592	53807	54%	3	46	58	43	56	59	216	71	8281	11.7	13970
M107R	53804	54712	74%	2	34	361	141	170	237	909	302	36037	10.7	42050
m108L	54210	53920	68%	5	37	68	63	44	116	291	96	10933	8.5	2740
M109R	55071	55520	74%	3	45	119	91	113	127	450	149	16947	8.4	17870
m110L	55378	55172	66%	4	48	44	56	44	63	207	68	7586	11.5	12690
M111R	55571	56290	72%	2	41	265	115	181	159	720	239	27297	9.9	36360
m112L	56110	55805	60%	4	42	44	81	46	135	306	101	11818	8.1	5840
m113L	56421	56128	54%	5	41	94	60	60	80	294	97	11313	9.3	13400
M114L	56971	56291	84%	4	48	198	143	182	158	681	226	24792	8.6	61690
m115R	56850	57083	56%	3	36	74	39	45	76	234	77	9169	10.3	5330
M116L	57762	56995	54%	5	46	238	179	175	176	768	255	29093	9.4	38780
m117R	57070	57552	62%	1	48	99	110	121	153	483	160	18270	8.7	17250
m118R	57197	57400	60%	2	50	38	47	54	65	204	67	7616	11.3	24040
m119R	57772	58008	66%	1	42	55	48	52	82	237	78	8654	9.5	1310

ORF	ORF Location		% A+T (-50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
M120L	58145	57786	76%	6	40	127	76	67	90	360	119	13636	4.6	5120
M121R	58225	58506	86%	1	51	72	81	63	66	282	93	10133	9.4	9680
M122R	58523	58948	64%	2	41	167	79	96	84	426	141	16342	11.3	6970
m123L	59143	58922	52%	4	38	82	42	43	55	222	73	8217	8.4	5150
M124L	60186	59140	68%	5	48	272	247	254	274	1047	348	38757	4.8	85180
m125R	59183	59389	60%	2	49	56	48	53	50	207	68	7609	9.2	10810
m126L	59546	59346	42%	6	42	56	34	50	61	201	66	7633	10.4	14000
m127R	59969	60169	46%	2	46	56	46	46	53	201	66	6978	11.6	1310
M128R	60357	62042	76%	3	48	363	373	435	515	1686	561	64125	7.4	2E+05
m129L	61436	60561	58%	6	50	233	237	198	208	876	291	32014	7.6	19810
m130R	60982	61287	58%	1	51	68	80	77	81	306	101	10867	12	11500
m131L	61450	61178	66%	4	49	72	73	61	67	273	90	10383	10.7	5240
m132R	61348	61671	50%	1	45	62	75	71	116	324	107	11635	10.7	17310
M133R	62232	63191	74%	3	51	212	237	255	256	960	319	35899	8.3	23110
m134L	63080	62340	56%	6	52	202	203	185	151	741	246	27342	10.3	8340
m135L	63466	63206	56%	4	49	57	58	70	76	261	86	9384	11.4	7060
M136L	63572	63213	68%	6	48	93	85	89	93	360	119	13801	8.1	17360
M137L	64469	63630	74%	6	44	247	203	168	222	840	279	31700	5.2	36220
m138R	63993	64253	52%	3	46	65	50	71	75	261	86	9859	8.2	15400
M139L	65483	64545	66%	6	45	269	221	201	248	939	312	34924	7.5	15720
M140L	65990	65484	74%	6	47	150	122	118	117	507	168	19352	9.1	27340
m141R	65554	65790	64%	1	51	56	64	58	59	237	78	8664	11.9	9590
m142R	65562	65798	72%	3	53	55	65	60	57	237	78	8802	7.2	21060
M143L	67855	66038	78%	4	44	622	371	429	396	1818	605	67782	10.9	66800
m144R	66095	66403	66%	2	37	62	69	46	132	309	102	12243	8	26690
m145R	66456	66689	58%	3	47	50	44	66	74	234	77	8011	4.2	1310
m146L	67062	66832	54%	5	46	74	50	56	51	231	76	8598	10.7	19750
m147R	67584	67787	48%	3	46	43	51	42	68	204	67	7446	11.1	2620
M148L	68171	67881	68%	6	45	88	80	52	71	291	96	10593	8.4	5150
M149R	68260	68733	80%	1	45	133	113	102	126	474	157	17620	5	10300
M150R	68760	69074	78%	3	41	118	52	78	67	315	104	12364	11	24750
m151L	68990	68787	60%	6	45	37	54	37	76	204	67	7940	10.4	8310
M152R	69108	74003	86%	3	46	1334	1033	1213	1316	4896	1631	167106	5.1	1E+05
m153L	70066	69113	46%	4	45	263	228	206	257	954	317	35138	6.1	17270
m154R	69889	70086	62%	1	46	53	44	48	53	198	65	7313	6.5	12660
m155L	70945	70079	52%	4	49	219	226	197	225	867	288	31695	4.5	9020
m156L	71659	70988	50%	4	47	170	165	149	188	672	223	24747	4.5	7710
m157L	72478	72137	58%	4	48	85	86	77	94	342	113	12140	5.8	1340
m158L	73897	73514	54%	4	40	104	89	63	128	384	127	13995	7.9	9530
m159L	74736	73981	60%	5	41	174	179	131	272	756	251	28064	9.9	7710
M160R	74021	75310	72%	2	42	450	248	294	298	1290	429	48921	10.9	37670

ORF	ORF Location		% A+T (-50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
m161L	75058	74840	56%	4	45	49	44	55	71	219	72	8380	9.5	9620
m162L	75287	75087	58%	6	41	50	47	35	69	201	66	7832	8.3	10960
M163L	75853	75311	76%	4	47	172	114	139	118	543	180	20781	9.3	40750
m164L	75555	75358	62%	5	51	61	44	56	37	198	65	7299	11.5	11470
m165R	75447	75650	48%	3	45	49	51	41	63	204	67	7295	11.1	1280
M166R	75958	76620	70%	1	44	221	140	152	150	663	220	24357	10.3	31300
M167L	77379	76621	74%	5	41	268	147	163	181	759	252	28461	8.5	12980
m168R	77002	77307	54%	1	41	69	62	64	111	306	101	12234	8.1	19210
m169R	77057	77335	54%	2	39	66	52	56	105	279	92	10279	5.7	7060
M170R	77524	77892	86%	1	46	102	96	74	97	369	122	13842	8.1	6460
M171L	78147	77893	76%	5	64	34	108	56	57	255	84	8671	3.5	1280
m172R	77947	78165	56%	1	65	45	49	94	31	219	72	8083	7.3	6970
M173R	78230	78682	78%	2	43	137	88	107	121	453	150	17358	4.8	33120
M174L	80034	78658	72%	5	44	411	294	314	358	1377	458	51957	5.9	51190
m175R	78679	78900	50%	1	42	67	48	46	61	222	73	8287	10.1	7060
m176R	79213	79437	60%	1	44	56	50	49	70	225	74	8518	5.9	15250
M177L	81483	80071	80%	5	43	438	310	303	362	1413	470	53949	5	37220
M178L	82188	81526	74%	5	40	211	142	126	184	663	220	25945	7.3	23670
M179L	82706	82221	78%	6	41	151	113	86	136	486	161	18287	11.1	29220
M180L	83379	82903	92%	5	39	176	89	98	114	477	158	18511	10.3	18580
m181R	83089	83286	64%	1	38	53	39	36	70	198	65	7735	9.4	2560
m182R	83120	83347	76%	2	40	62	49	43	74	228	75	8983	9.4	23530
M183R	83430	83717	92%	3	43	79	58	66	85	288	95	10292	7.8	8250
M184L	84274	83933	76%	4	43	103	73	73	93	342	113	12956	10	17870
m185R	84298	84504	76%	1	43	63	46	43	55	207	68	7611	9	15370
M186R	84369	85931	74%	3	50	332	425	356	450	1563	520	58806	8.6	1E+05
m187L	84637	84398	58%	4	46	75	50	61	54	240	79	9090	10.4	7680
m188R	85442	85702	50%	2	49	56	80	47	78	261	86	9635	7.6	9680
m189L	86056	85850	48%	4	35	59	27	46	75	207	68	7769	10.5	5150
M190R	85998	89207	88%	3	50	924	855	756	675	3210	1069	113276	10.8	54400
m191R	86062	86259	50%	1	52	44	56	47	51	198	65	7378	12.5	60
m192L	86518	86285	38%	4	47	54	50	61	69	234	77	8530	11.9	15220
m193L	86978	86646	52%	6	47	85	69	86	93	333	110	12209	8.5	8430
m194R	86944	87168	50%	1	51	49	67	47	62	225	74	8175	10.2	11380
m195L	87611	87255	52%	6	52	65	97	89	106	357	118	13099	6.9	12930
m196L	87908	87675	52%	6	49	56	57	57	64	234	77	8517	7.6	15400
m197L	88675	88472	54%	4	54	31	49	61	63	204	67	6605	4.1	60
m198L	89035	88760	50%	4	55	48	41	110	77	276	91	7907	3.6	0
m199R	88777	89043	54%	1	55	73	107	40	47	267	88	9996	12.4	1280
M200L	89586	89275	66%	5	46	86	66	76	84	312	103	11554	5.2	3930
M201R	89691	91055	74%	3	45	386	286	334	359	1365	454	50600	8.4	33950

ORF	ORF Location		% A+T (-50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
m202L	90305	90033	62%	6	42	79	65	51	78	273	90	10583	10.9	12270
m203L	90361	90110	52%	4	43	78	56	53	65	252	83	9178	9.4	12750
M204L	91463	91110	82%	6	53	95	80	106	73	354	117	12748	3.9	8250
m205R	91243	91449	48%	1	52	47	57	50	53	207	68	7274	5.7	11470
M206L	91741	91520	72%	4	48	68	46	61	47	222	73	8892	3.3	10240
m207L	92549	91953	64%	6	47	114	129	152	202	597	198	21899	9.5	29340
M208L	92392	91991	70%	4	51	61	84	123	134	402	133	14348	10.1	27940
m209R	92076	92378	54%	3	53	86	82	80	55	303	100	10929	11.1	2560
M210L	92872	92585	80%	4	47	93	67	69	59	288	95	10894	10	30
m211R	92723	92980	44%	2	39	51	48	52	107	258	85	9934	5.9	18020
m212R	92799	93011	50%	3	36	48	35	42	88	213	70	8299	8.2	9080
M213L	93392	92934	78%	6	43	160	102	97	100	459	152	17760	10.2	17810
m214L	93214	92999	68%	4	45	73	49	49	45	216	71	8386	11.2	8370
M215R	93462	94301	78%	3	48	229	187	216	208	840	279	32166	5.1	38270
m216L	94192	93986	56%	4	49	40	62	40	65	207	68	7518	11.3	9620
m217L	94725	94321	50%	5	49	95	93	106	111	405	134	14249	11.1	8310
M218R	94325	94810	84%	2	50	127	123	120	116	486	161	17491	4.7	20910
M219L	95631	94807	82%	5	43	257	173	183	212	825	274	30277	9.9	17270
m220R	95282	95599	56%	2	48	63	82	72	101	318	105	11593	10.6	1400
M221L	96553	95651	64%	4	42	229	161	221	292	903	300	34091	4.5	30020
m222R	95790	96116	56%	3	44	100	80	63	84	327	108	11754	11	14030
m223R	95852	96349	62%	2	42	162	116	93	127	498	165	19120	10.7	24160
m224L	96651	96370	66%	5	43	80	57	64	81	282	93	10282	10.4	14060
M225L	98791	96644	72%	4	48	572	490	546	540	2148	715	80083	7.2	73300
m226R	96669	97109	62%	3	47	114	109	97	121	441	146	16117	11.6	19630
m227R	97242	97445	56%	3	50	49	65	37	53	204	67	7657	10.5	7060
m228L	98010	97804	42%	5	46	53	42	54	58	207	68	7835	10.8	12750
m229R	98004	98411	56%	3	51	106	100	110	92	408	135	15085	10.3	19690
m230R	98541	98753	64%	3	48	47	52	51	63	213	70	7903	5.6	14000
m231R	98648	98860	50%	2	43	53	47	45	68	213	70	8371	9.1	9710
M232R	98893	99207	74%	1	36	118	52	62	83	315	104	12385	10.6	13370
M233L	100741	99233	72%	4	47	429	333	370	377	1509	502	56421	9.7	89620
m234R	99504	99704	48%	3	48	56	54	43	48	201	66	7385	11.1	8250
m235R	99846	100058	48%	3	52	44	60	51	58	213	70	7783	9.6	14000
m236R	100367	100783	62%	2	40	114	93	73	137	417	138	15832	8.2	8310
M237L	101090	100761	64%	6	47	100	72	83	75	330	109	12661	5.2	16500
M238R	101294	101653	78%	2	46	94	89	76	101	360	119	13220	10.7	6430
m239L	101646	101320	56%	5	48	87	73	84	83	327	108	12057	8.4	21000
M240L	101915	101643	76%	6	37	102	51	51	69	273	90	11405	10.5	21620
M241L	102200	101937	84%	6	36	90	38	57	79	264	87	10152	4.8	6520
M242L	102804	102232	72%	5	41	180	125	109	159	573	190	20889	10.1	12150

ORF	ORF Location		% A+T (-50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
M243R	102829	103065	84%	1	40	62	41	53	81	237	78	8682	6.5	6460
m244R	103086	103478	66%	3	43	123	85	85	100	393	130	14801	10	10810
m245L	103420	103187	50%	4	45	62	55	50	67	234	77	8875	10.7	8250
M246L	103535	103293	74%	6	41	55	51	48	89	243	80	9501	7.9	9650
M247R	103586	104338	74%	2	46	240	175	173	165	753	250	28318	8.2	16820
m248R	103875	104156	54%	3	50	90	73	68	51	282	93	10393	11.8	1400
M249R	104467	104883	88%	1	44	131	96	88	102	417	138	15411	9.3	8310
m250L	104786	104514	56%	6	44	73	63	57	80	273	90	9824	11.8	5750
m251R	104705	105115	62%	2	48	107	113	84	107	411	136	15209	8.2	24280
m252L	104988	104788	48%	5	47	52	38	56	55	201	66	7565	11.9	22790
M253L	106765	104885	86%	4	47	591	433	448	409	1881	626	67331	8.3	86170
m254R	105328	105612	70%	1	52	63	66	82	74	285	94	10948	8.4	14090
m255L	105987	105772	66%	5	48	67	38	66	45	216	71	7608	5	0
m256R	106227	106430	56%	3	40	39	38	44	83	204	67	7670	9	5840
m257R	106575	106814	62%	3	37	65	49	40	86	240	79	8743	10.8	2650
M258R	106836	108623	88%	3	49	505	400	470	413	1788	595	64075	10.1	62430
m259R	107614	107817	58%	1	52	58	63	44	39	204	67	7647	12.2	30
m260R	107861	108205	48%	2	56	81	86	106	72	345	114	13240	10.5	35720
m261R	108103	108390	60%	1	45	80	50	79	79	288	95	10770	8.5	22880
m262L	108508	108275	56%	4	44	54	64	39	77	234	77	8580	10.6	7060
M263R	108657	109313	78%	3	43	211	132	152	162	657	218	24291	10.2	22220
M264L	109732	109310	66%	4	50	90	89	122	122	423	140	14801	4.8	7680
m265L	110071	109823	62%	4	43	78	53	54	64	249	82	9248	10.3	90
M266R	109853	110674	84%	2	44	210	175	186	251	822	273	30420	8.4	3930
m267L	110331	110116	52%	5	41	70	55	34	57	216	71	7952	4.6	5750
m268R	110517	110732	52%	3	44	55	55	39	67	216	71	8511	11.9	24070
M269L	111880	110678	86%	4	44	370	292	242	299	1203	400	44553	5.7	45830
m270R	111648	111845	58%	3	41	46	44	37	71	198	65	7317	7.3	1310
m271L	111986	111705	52%	6	38	98	54	53	77	282	93	11394	10.7	15400
M272L	113304	111925	54%	5	47	398	308	343	331	1380	459	51473	8.8	36990
m273R	111941	112189	82%	2	45	61	62	51	75	249	82	9538	9.9	11500
m274R	112244	112471	46%	2	46	53	52	53	70	228	75	8243	6.4	17130
m275R	112616	112900	52%	2	48	65	74	62	84	285	94	10895	11.9	16500
m276L	113591	113301	56%	6	47	80	65	71	75	291	96	10834	11.1	7060
m277L	113988	113374	50%	5	52	160	150	167	138	615	204	22757	8.6	9650
M278R	113482	114285	70%	1	53	165	230	197	212	804	267	29093	4.1	23850
m279R	113597	113962	42%	2	55	71	96	104	95	366	121	13679	12.6	22940
m280L	114363	114133	78%	5	42	62	42	56	71	231	76	8537	7.3	8400
M281L	114563	114366	76%	6	51	32	38	63	65	198	65	7037	12.2	12690
M282R	114658	114918	80%	1	46	53	57	63	88	261	86	9835	10.9	26630
M283R	114950	116110	76%	2	42	355	203	283	320	1161	386	44284	9.8	35850

ORF	ORF Location		% A+T (~50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
m284L	115387	115022	66%	4	44	86	97	63	120	366	121	14400	9.5	9680
m285L	115368	115102	58%	5	45	62	72	47	86	267	88	9894	12	8280
m286L	115960	115679	68%	4	40	88	64	50	80	282	93	10740	10.2	15250
m287L	116160	115903	80%	5	37	84	53	42	79	258	85	9532	10.9	16500
M288L	117169	116111	80%	4	31	424	158	174	303	1059	352	40609	10.1	57330
M289R	117311	118417	72%	2	45	344	219	279	265	1107	368	40588	10.2	64810
m290R	117315	117530	72%	3	46	64	43	56	53	216	71	8139	8.5	11560
m291L	117744	117436	52%	5	45	67	71	69	102	309	102	11318	5	16560
m292L	117991	117608	68%	4	50	80	106	85	113	384	127	14384	8.7	23650
m293L	118314	118066	58%	5	42	62	58	46	83	249	82	9087	10.2	8310
M294R	118438	119091	70%	1	53	119	148	198	189	654	217	23979	9	32590
m295L	119073	118453	54%	5	53	178	185	146	112	621	206	23412	10.5	26290
m296L	119251	118898	46%	4	49	92	104	68	90	354	117	12946	8.8	1490
m297L	119390	119160	62%	6	46	47	55	51	78	231	76	8273	4.8	1400
M298R	119184	119657	62%	3	46	158	103	116	97	474	157	18096	5.8	10810
M299R	119736	120185	72%	3	41	129	93	93	135	450	149	16647	4.5	7680
m300L	120203	119895	58%	6	40	98	66	57	88	309	102	11860	9.5	3990
m301L	120447	120217	42%	5	53	68	30	92	41	231	76	9521	0	11440
M302L	121198	120230	64%	4	53	296	184	329	160	969	322	36432	4	8310
m303R	120462	121244	44%	3	51	125	232	171	255	783	260	26934	12.5	22880
m304L	120771	120529	48%	5	45	76	40	70	57	243	80	9697	0	19630
m305R	120878	121219	48%	2	55	42	107	81	112	342	113	12348	6.5	11710
m306R	121304	121786	60%	2	48	122	126	106	129	483	160	17674	5.4	27970
M307L	122464	121346	86%	4	46	310	252	267	290	1119	372	41790	5.5	48600
M308L	123252	122545	84%	5	46	235	162	164	147	708	235	25934	11.6	6400
m309R	122771	123259	48%	2	45	101	104	118	166	489	162	17762	11.8	8280
M310L	123954	123292	90%	5	49	177	163	159	164	663	220	24035	4.8	15930
m311R	123590	123817	56%	2	54	52	59	65	52	228	75	7703	10.6	1280
M312R	124001	124339	90%	2	49	74	83	84	98	339	112	12025	4.5	6460
M313L	124897	124340	86%	4	44	205	124	120	109	558	185	20825	10.9	20430
M314L	125260	125003	76%	4	46	70	52	67	69	258	85	10247	3.7	21710
M315L	125720	125382	68%	6	42	105	53	90	91	339	112	13006	4.9	10810
m316R	125397	125699	58%	3	42	84	80	47	92	303	100	11764	10	3960
M317L	126306	125800	80%	5	46	146	134	99	128	507	168	18117	8.1	15990
m318R	126300	126713	54%	3	53	99	111	108	96	414	137	14709	11.1	8280
M319L	127075	126350	62%	4	56	163	216	188	159	726	241	26794	5.3	23470
m320R	126788	127183	50%	2	51	77	70	131	118	396	131	13961	4.7	19750
m321L	127242	127018	58%	5	45	86	73	29	37	225	74	8169	11.3	6970
M322L	127493	127239	70%	6	44	71	52	60	72	255	84	9373	8.2	7090
M323R	127579	127818	78%	1	48	63	47	67	63	240	79	9150	5	9530
M324R	127875	128993	72%	3	39	300	214	228	377	1119	372	41244	4.7	42910

ORF	ORF Location		% A+T (-50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
m325L	129162	128965	62%	5	36	63	40	31	64	198	65	7329	10.6	3840
M326R	129017	129370	76%	2	40	130	66	77	81	354	117	12850	10.7	9530
m327L	129461	129201	52%	6	38	63	55	43	100	261	86	10034	10.3	15310
M328R	129391	129732	80%	1	50	74	77	95	96	342	113	11845	4.9	5210
M329L	130346	129846	84%	6	45	159	110	115	117	501	166	19229	5.5	27370
m330R	129967	130341	62%	1	45	85	94	74	122	375	124	13545	6.2	8280
m331R	130163	130546	64%	2	49	77	96	92	119	384	127	14295	11	23590
M332L	131644	130376	78%	4	53	318	347	328	276	1269	422	46790	5.1	8960
m333L	131493	131140	40%	5	51	96	90	89	79	354	117	13037	11.8	19750
M334R	131667	132266	72%	3	42	156	141	112	191	600	199	23257	10.5	27400
m335L	132326	131691	68%	6	42	203	116	151	166	636	211	24703	9	33920
m336R	132317	132766	66%	2	45	139	119	85	107	450	149	17790	11.1	9620
M337L	133396	132338	76%	4	46	260	205	278	316	1059	352	40348	9.8	60400
m338R	132869	133273	52%	2	48	118	106	87	94	405	134	15824	9.1	12840
m339R	133035	133406	56%	3	44	109	91	72	100	372	123	14049	8.8	14030
m340L	134014	133565	50%	4	50	83	80	147	140	450	149	15782	4.4	14120
M341R	133727	134452	66%	2	57	163	224	187	152	726	241	26778	5.3	23470
m342L	134328	134089	38%	5	57	47	63	74	56	240	79	8597	9.5	7000
m343L	134957	134691	54%	6	56	53	78	71	65	267	88	10004	11.7	21030
M344L	136141	135617	78%	4	41	159	87	129	150	525	174	20191	6.8	14740
m345L	136343	136146	56%	6	37	67	33	41	57	198	65	7481	9.4	12750
M346L	136605	136186	64%	5	45	118	92	97	113	420	139	15951	4.6	14680
m347R	136321	136635	66%	1	45	79	67	76	93	315	104	11829	8.4	2650
M348L	138420	136741	90%	5	54	425	457	453	345	1680	559	60602	9.3	48360
m349R	136790	137074	62%	2	55	52	73	84	76	285	94	9873	11.4	11500
m350L	137303	136968	38%	6	54	85	88	92	71	336	111	12565	12.1	19660
m351R	137080	137361	46%	1	55	62	78	77	65	282	93	10301	7.2	19840
m352L	137918	137448	52%	6	58	97	151	122	101	471	156	16876	10.7	7090
m353L	138776	138429	52%	6	45	100	65	92	91	348	115	13426	9.1	32500
M354R	138534	139850	68%	3	48	341	333	303	340	1317	438	48896	4.5	67880
m355R	138625	138912	60%	1	53	69	81	72	66	288	95	10689	11.3	12750
m356R	138988	139194	44%	1	48	55	46	53	53	207	68	7482	12	120
M357L	140281	139898	86%	4	51	108	106	89	81	384	127	14449	10.2	14080
m358R	140365	140838	78%	1	40	172	101	89	112	474	157	18718	9.9	36900
M359L	141751	140723	64%	4	50	264	240	276	249	1029	342	38474	8.3	63770
m360R	140726	141082	66%	2	46	101	95	70	91	357	118	13253	12.1	1430
M361R	141459	141674	54%	3	55	49	62	57	48	216	71	7642	11.3	1340
M362L	142574	141933	88%	6	55	159	179	173	131	642	213	23841	3.7	35560
M363L	143695	142670	72%	4	45	296	167	290	273	1026	341	38819	4	10870
m364R	142845	143195	56%	3	43	106	83	68	94	351	116	12808	11.7	7000
m365L	143505	143170	60%	5	44	114	25	124	73	336	111	14532	0	7000

ORF	ORF Location		% A+T (-50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
m366R	143202	143645	60%	3	45	98	157	41	148	444	147	14330	6.4	30
M367L	144396	143737	80%	5	52	162	183	159	156	660	219	25120	4.8	17510
m368L	144406	144143	80%	4	48	74	77	49	64	264	87	10087	9.6	9680
m369L	144817	144422	60%	4	48	83	113	78	122	396	131	14769	7.5	1430
M370R	144446	145249	80%	2	47	272	156	219	157	804	267	30019	10.2	39460
m371L	145139	144933	46%	6	42	38	49	38	82	207	68	8056	8.1	19780
M372R	145289	148684	78%	2	47	1034	803	802	757	3396	1131	124386	10.9	85040
m373L	147588	147244	50%	5	45	72	87	67	119	345	114	12770	11	10840
m374L	147919	147719	44%	4	46	46	46	46	63	201	66	7230	4.2	3870
M375R	148717	149256	76%	1	30	195	78	83	184	540	179	21200	4.8	12890
M376R	149243	149833	78%	2	30	244	87	90	170	591	196	22944	10.9	22330
M377L	150512	149898	72%	6	52	147	169	152	147	615	204	23517	10.9	34310
m378L	150262	149984	46%	4	52	69	75	71	64	279	92	9915	11.6	5810
M379R	150719	151159	78%	2	44	140	89	107	105	441	146	17152	4.4	22990
m380L	150985	150776	62%	4	43	51	49	42	68	210	69	7990	8.7	9590
M381R	151222	152640	84%	1	44	460	325	296	338	1419	472	53699	6.4	49820
m382L	151760	151494	50%	6	46	63	61	63	80	267	88	10263	11.3	22250
m383L	152087	151887	58%	6	43	50	38	48	65	201	66	7012	8.2	120
m384L	152546	152346	64%	6	38	59	31	45	66	201	66	7640	8.7	22250
M385L	154232	152649	74%	6	47	510	347	392	335	1584	527	59189	11.1	52230
m386R	152797	153087	58%	1	45	67	63	67	94	291	96	10845	8	13520
m387R	153385	154278	58%	1	47	163	229	195	307	894	297	33022	6.5	15930
M388R	154334	155041	76%	2	41	227	113	175	193	708	235	27143	4.6	55220
M389R	155125	156624	68%	1	45	448	339	337	376	1500	499	57654	8.5	57450
m390L	155619	155377	54%	5	50	54	48	73	68	243	80	9265	5.1	14030
m391L	156030	155728	64%	5	47	76	79	64	84	303	100	11266	6.5	16650
m392L	156173	155940	50%	6	44	61	55	47	71	234	77	8748	9.5	17870
m393R	155957	156178	54%	2	44	67	44	53	58	222	73	8627	10.8	28540
m394L	156620	156309	58%	6	40	78	71	55	108	312	103	11564	7.3	60
M395L	156892	156629	76%	4	42	60	50	60	94	264	87	9641	10.4	19630
m396R	156675	156944	56%	3	37	97	58	43	72	270	89	10041	11.4	2560
M397R	157055	157819	80%	2	41	260	149	168	188	765	254	28998	8.3	18070
m398R	157206	157544	56%	3	42	115	72	70	82	339	112	12841	11.4	6520
M399L	158394	157843	74%	5	50	171	136	139	106	552	183	20780	7.7	23500
m400L	158861	158493	64%	6	44	93	90	74	112	369	122	13543	10.4	10900
M401R	158497	159171	74%	1	46	197	147	163	168	675	224	25100	8	31810
m402L	159108	158842	56%	5	48	68	65	62	72	267	88	10296	10.5	11440
M403R	159213	159533	80%	3	49	75	75	81	90	321	106	11843	10.1	8400
M404R	159584	160975	86%	2	47	371	320	338	363	1392	463	52175	6	25010
m405L	160671	160354	56%	5	49	77	82	75	84	318	105	11780	11.8	2590
m406L	161079	160882	64%	5	33	59	27	39	73	198	65	7659	11.1	5150

ORF	ORF Location		% A+T (~50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
M407R	161042	161383	86%	2	44	114	71	81	76	342	113	13228	5	2560
m408L	162340	161597	58%	4	50	96	136	238	274	744	247	23927	4.2	5120
M409R	161649	163058	84%	3	47	485	376	287	262	1410	469	50356	10.8	57990
m410L	163253	163035	48%	6	45	49	42	56	72	219	72	8078	9.7	19720
M411L	163494	163069	68%	5	49	96	104	103	123	426	141	15821	4.4	16530
m412R	163120	163335	54%	1	49	68	58	48	42	216	71	8023	6.3	90
M413L	164611	163610	70%	4	50	226	213	292	271	1002	333	36900	6.6	29370
m414R	163833	164216	56%	3	54	96	124	84	80	384	127	13142	9.3	330
m415R	163984	164430	50%	1	52	111	135	98	103	447	148	16958	12	33690
M416R	164684	165580	72%	2	46	257	204	210	226	897	298	33173	5.3	22250
M417L	167392	165590	76%	4	47	541	478	378	406	1803	600	64525	8.5	38470
m418R	165741	166010	70%	3	44	61	64	56	89	270	89	10162	4.4	10900
m419R	166383	166706	50%	3	56	70	54	128	72	324	107	9402	8.7	1340
m420R	166897	167295	56%	1	53	67	74	136	122	399	132	15374	12.1	9620
M421L	167746	167414	68%	4	42	86	76	64	107	333	110	12635	6.5	27940
m422L	168066	167668	50%	5	44	98	99	78	124	399	132	15708	10.6	38780
M423R	167767	168027	78%	1	46	81	55	66	59	261	86	9970	10.8	1400
m424L	168014	167778	68%	6	49	52	62	53	70	237	78	8947	11.9	1370
M425R	168038	168598	74%	2	44	160	107	138	156	561	186	21496	4.9	26830
m426L	168393	168196	50%	5	43	62	50	35	51	198	65	7193	4.9	5150
M427R	168662	169720	80%	2	47	332	222	274	231	1059	352	39679	8.7	22960
m428L	168933	168709	46%	5	44	37	50	50	88	225	74	8258	7.7	2770
m429L	169269	169069	50%	5	48	45	57	40	59	201	66	7307	8	9620
M430L	170803	169721	70%	4	46	315	253	249	266	1083	360	40027	8.3	21830
m431R	169803	170048	54%	3	43	66	56	51	73	246	81	9411	11.4	8250
m432R	170094	170342	52%	3	43	60	56	52	81	249	82	9027	10.3	8280
m433R	170566	170784	56%	1	51	54	49	63	53	219	72	8552	12.6	22790
m434R	170706	171068	60%	3	44	96	64	94	109	363	120	13909	5.9	13460
M435L	171179	170835	54%	6	40	107	79	58	101	345	114	13066	10.5	1340
m436R	170959	171363	60%	1	42	121	79	92	113	405	134	15334	10.5	24810
M437R	171162	171416	70%	3	44	79	55	58	63	255	84	10357	11.3	14060
M438L	171644	171399	72%	6	40	62	59	40	85	246	81	9007	7.9	14030
m439L	171801	171592	58%	5	38	67	45	35	63	210	69	7870	10.6	6400
M440R	171671	172141	78%	2	45	139	96	114	122	471	156	17495	4.9	14710
m441L	172878	172138	70%	5	41	146	169	135	291	741	246	28122	8.6	22480
m442R	172162	172524	82%	1	39	173	66	77	47	363	120	13654	11	48080
m443L	172616	172218	52%	6	42	58	94	72	175	399	132	15720	10.9	2560
M444R	172564	172914	74%	1	45	103	68	90	90	351	116	13306	4.8	1310
M445L	173265	172915	74%	5	48	95	69	100	87	351	116	13194	5.8	8340
m446R	172973	173239	52%	2	49	67	75	57	68	267	88	9937	10.4	12750
m447R	173255	173527	48%	2	37	95	62	39	77	273	90	10726	9.4	6460

ORF	ORF Location		% A+T (-50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
M448L	173669	173331	76%	6	40	89	57	79	114	339	112	12691	5.8	2620
M449L	174017	173703	76%	6	43	105	63	73	74	315	104	11731	8.1	5180
m450R	173707	173973	78%	1	45	62	65	55	85	267	88	9470	10.1	9650
M451L	175368	174040	72%	5	42	395	286	278	370	1329	442	51064	5.2	55660
m452R	174059	174328	76%	2	44	73	56	63	78	270	89	10278	10.8	16530
m453R	175006	175221	60%	1	38	64	44	39	69	216	71	8469	8.6	15430
M454R	175405	175731	68%	1	38	92	64	60	111	327	108	12898	9.6	24960
M455R	175778	176059	82%	2	47	78	71	61	72	282	93	10778	10.7	9530
m456L	176463	176068	52%	5	48	101	100	89	106	396	131	15408	11.1	19180
M457R	176074	177204	62%	1	50	322	256	312	241	1131	376	42426	10.1	56200
m458L	176799	176530	44%	5	50	54	73	62	81	270	89	10129	11	20940
m459L	176915	176640	42%	6	54	60	80	69	67	276	91	10411	7.8	12210
M460R	177267	178022	76%	3	36	291	99	170	196	756	251	28856	10.1	25580
m461L	177809	177324	80%	6	38	126	113	71	176	486	161	18842	10.8	14890
m462L	177892	177668	70%	4	38	40	53	33	99	225	74	8116	4.2	7090
M463L	179395	178094	86%	4	60	267	472	303	260	1302	433	47649	6.9	49160
m464R	179187	179438	42%	3	44	60	50	62	80	252	83	9780	9.4	24840
M465L	180410	179547	80%	6	44	311	148	236	169	864	287	32624	9	32520
m466R	180176	180391	46%	2	47	44	59	43	70	216	71	8156	9.7	14060
M467R	180530	183079	76%	2	52	579	660	667	644	2550	849	95586	6.4	1E+05
m468L	181119	180898	56%	5	52	53	58	57	54	222	73	8735	10.3	25350
m469L	181264	180974	42%	4	49	84	77	67	63	291	96	10877	11.8	1280
m470L	181909	181706	48%	4	55	47	56	57	44	204	67	7564	10.7	17070
m471L	183057	182827	64%	5	48	61	61	51	58	231	76	8082	11.7	12690
M472R	183130	183588	68%	1	48	124	87	134	114	459	152	16606	8.2	5240
m473L	183499	183269	50%	4	50	45	71	44	71	231	76	8278	7.9	9740
m474L	183918	183589	66%	5	48	52	89	70	119	330	109	12267	10.2	11440
M475R	183617	183967	78%	2	49	128	74	99	50	351	116	12728	11.9	0
M476L	184825	184217	66%	4	47	189	144	141	135	609	202	22830	10.5	12210
m477R	184604	184819	76%	2	49	36	50	56	74	216	71	7673	10.6	3840
M478L	185179	184859	72%	4	44	89	79	63	90	321	106	12214	4.9	7000
M479R	185275	185769	72%	1	46	135	116	111	133	495	164	18672	5.1	5270
m480L	185780	185547	58%	6	43	60	49	51	74	234	77	8472	11.3	7030
m481L	185980	185771	50%	4	49	53	42	61	54	210	69	8076	11.9	7030
M482R	185802	186341	70%	3	48	141	126	134	139	540	179	20330	7.5	29190
m483R	186328	186609	52%	1	44	90	66	57	69	282	93	10659	11.3	15280
M484L	187082	186342	88%	6	47	203	194	156	188	741	246	26920	9.6	19060
M485R	187170	187865	80%	3	47	203	168	161	164	696	231	25815	4.8	15250
m486L	187726	187433	42%	4	50	64	67	80	83	294	97	10789	5	5840
m487R	187873	188394	58%	1	59	108	176	134	104	522	173	20717	4.1	1E+05
M488L	188632	187883	80%	4	55	185	188	224	153	750	249	26090	12.9	3840

ORF	ORF Location		% A+T (-50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
m489R	187941	188483	40%	3	58	108	175	138	122	543	180	19134	8	5840
m490L	188607	188395	74%	5	43	73	49	43	48	213	70	8411	11.9	3870
M491R	188838	189986	74%	3	52	299	297	295	258	1149	382	43961	8	60520
m492L	189190	188972	46%	4	48	48	47	58	66	219	72	8398	12.2	13970
m493L	189269	189003	60%	6	52	52	61	77	77	267	88	9348	3.7	2620
m494L	189635	189429	56%	6	56	50	62	54	41	207	68	7424	8.4	17100
M495R	190020	190649	78%	3	46	188	170	122	150	630	209	23347	8.6	14800
M496L	191314	190697	84%	4	45	196	134	146	142	618	205	22857	8.2	24300
M497R	191377	191913	88%	1	53	138	155	129	115	537	178	19620	8.6	27460
m498L	191788	191480	54%	4	54	61	79	89	80	309	102	12356	9.9	35780
m499R	191600	191803	40%	2	53	47	53	56	48	204	67	7015	7.3	5780
m500L	191928	191731	60%	5	48	55	38	58	47	198	65	6823	9.7	60
M501R	191937	192383	70%	3	49	129	108	110	100	447	148	16445	8.4	22930
M502L	193151	192414	78%	6	45	265	133	197	143	738	245	27604	9.8	35080
m503R	192663	192926	66%	3	45	40	74	45	105	264	87	9503	7.9	120
M504R	193217	193498	74%	2	41	101	52	63	66	282	93	11054	10.7	7710
m505R	193511	193780	68%	2	43	68	58	58	86	270	89	10511	6.8	10960
M506L	193756	193544	76%	4	49	59	54	50	50	213	70	7976	10.4	8310
M507L	194020	193781	78%	4	53	58	71	57	54	240	79	8651	9.9	5750
M508R	194269	195852	82%	1	48	464	414	349	357	1584	527	57116	5	53060
m509L	195125	194829	54%	6	48	71	67	76	83	297	98	10360	9.3	3870
m510L	195823	195554	62%	4	51	62	67	71	70	270	89	10288	8.8	31810
M511L	196067	195849	72%	6	43	71	39	56	53	219	72	8111	4.3	7030
M512L	196357	196118	76%	4	49	64	58	59	59	240	79	9144	10.9	22190
m513R	196352	196681	58%	2	40	103	63	68	96	330	109	13034	11.3	10300
m514L	197128	196382	68%	4	39	237	134	161	215	747	248	28607	5.3	21770
M515R	196755	197240	64%	3	37	129	107	74	176	486	161	17936	9.4	9110
m516L	197407	197180	54%	4	45	49	46	56	77	228	75	8402	6.8	2590
m517R	197246	197752	68%	2	48	141	122	123	121	507	168	19025	10.4	6490
M518L	201296	197727	76%	6	48	970	695	1007	898	3570	1189	120258	4.7	75630
m519L	198025	197810	38%	4	46	62	36	63	55	216	71	8026	12	22880
m520R	199333	199671	58%	1	42	91	89	55	104	339	112	12586	10.2	5120
m521L	201067	200870	50%	4	47	56	35	59	48	198	65	7478	11.2	18380
M522L	203211	201364	74%	5	45	562	412	416	458	1848	615	68416	8.6	59140
m523L	202412	202158	54%	6	45	69	64	52	70	255	84	9652	8.5	7000
m524R	203196	203429	52%	3	38	52	47	43	92	234	77	9148	8.1	14150
M525L	204063	203302	74%	5	41	284	147	169	162	762	253	28488	7.7	15540
M526L	207559	204263	76%	4	47	983	678	880	756	3297	1098	110867	6.6	74890
m527R	205326	205583	60%	3	47	60	72	50	76	258	85	9178	5.6	2590
m528R	206769	207017	40%	3	50	44	62	62	81	249	82	9366	4	3870
m529R	207508	207717	54%	1	37	66	46	32	66	210	69	8311	11	5120

ORF	ORF Location		% A+T (~50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
M530L	208054	207620	68%	4	45	140	99	97	99	435	144	16138	5.4	16650
M531L	209217	208120	66%	5	43	349	189	286	274	1098	365	41271	8.5	33390
m532R	208354	208563	48%	1	46	48	55	41	66	210	69	7720	7.3	1370
m533R	208901	209158	62%	2	43	53	61	49	95	258	85	9402	7.9	1400
m534R	209205	209477	72%	3	44	81	52	69	71	273	90	10035	4.2	1370
M535L	211856	209241	78%	6	52	643	691	657	625	2616	871	95332	6.5	69430
m536R	209245	209607	72%	1	44	120	69	91	83	363	120	13508	11	12780
m537R	209650	210066	44%	1	54	97	109	116	95	417	138	14806	10.1	7060
m538R	210184	210417	48%	1	55	47	68	61	58	234	77	8141	9.9	90
m539R	210439	210876	50%	1	51	94	98	127	119	438	145	16354	7.5	25350
m540L	211104	210826	50%	5	55	52	80	73	74	279	92	10459	8.3	12330
m541R	210931	211458	54%	1	55	125	154	134	115	528	175	19175	10.9	20970
m542R	211603	211884	56%	1	49	63	61	76	82	282	93	10212	10.8	2560
M543R	212005	212823	68%	1	45	230	183	183	223	819	272	30496	8.8	17870
m544L	212294	212094	58%	6	43	52	44	42	63	201	66	7408	10.4	12720
m545L	212862	212650	64%	5	37	56	39	39	79	213	70	8226	7.7	12810
M546R	212942	216118	72%	2	50	767	781	796	833	3177	1058	117661	9.6	1E+05
m547R	213534	213746	50%	3	47	42	42	59	70	213	70	8061	8.8	29820
m548L	214314	213916	56%	5	56	89	107	117	86	399	132	14427	12.5	1280
m549L	214851	214561	50%	5	53	58	72	82	79	291	96	10866	10.4	19660
m550L	215851	214859	58%	4	50	259	259	233	242	993	330	36800	7.8	8490
m551L	215751	215539	56%	5	46	51	50	49	63	213	70	7775	12.4	1340
m552L	216399	215959	56%	5	38	109	88	80	164	441	146	17292	9.4	28680
M553R	216182	216592	82%	2	41	166	77	90	78	411	136	15898	11	28650
m554L	216622	216272	64%	4	40	64	77	63	147	351	116	13757	9.7	19240
m555L	216794	216495	58%	6	41	70	69	54	107	300	99	11133	9.4	10840
M556R	216631	217155	70%	1	47	151	115	131	128	525	174	19110	9.9	25520
M557L	218777	217152	90%	6	49	472	481	315	358	1626	541	56441	10.4	19830
m558R	218073	218723	54%	3	59	98	108	274	171	651	216	28187	11.7	2E+05
m559R	218182	218799	50%	1	57	93	93	262	170	618	205	16940	3.6	2620
M560L	219451	218900	88%	4	46	166	93	162	131	552	183	20995	3.2	20510
m561R	218995	219363	54%	1	48	80	120	56	113	369	122	14411	12.6	1430
m562R	219308	219766	54%	2	43	110	99	98	152	459	152	17061	8	6610
M563L	219757	219491	72%	4	50	79	66	67	55	267	88	10050	10.3	6970
M564L	220484	219798	74%	6	52	159	181	177	170	687	228	24868	6.9	17840
m565R	220197	220547	48%	3	48	94	77	90	90	351	116	12698	7.4	8430
m566R	220465	220710	38%	1	45	66	52	59	69	246	81	8735	6.4	3900
M567L	221450	220578	82%	6	53	196	246	217	214	873	290	32584	5.5	30080
m568R	220749	221105	48%	3	55	77	98	98	84	357	118	12605	5.1	1370
M569L	221845	221492	68%	4	48	114	72	97	71	354	117	13085	10	2560
M570L	222368	221874	68%	6	51	134	126	126	109	495	164	18237	9.9	15930

ORF	ORF Location		% A+T (-50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
m571R	221876	222112	72%	2	53	49	68	58	62	237	78	9063	8.8	18560
M572R	222471	222827	78%	3	49	92	84	90	91	357	118	12884	10.4	13940
m573R	222557	222787	52%	2	48	51	54	58	68	231	76	8549	11	16590
M574L	223289	222828	72%	6	50	169	107	124	62	462	153	17498	11.7	1280
m575R	222990	223190	58%	3	51	30	59	43	69	201	66	7657	7.8	2710
m576R	223021	223308	52%	1	50	41	83	62	102	288	95	10471	8.6	18560
M577L	223684	223313	64%	4	41	110	70	83	109	372	123	14264	10.1	2590
M578R	223670	224296	70%	2	44	196	130	145	156	627	208	23717	8.2	23290
m579L	224482	224249	44%	4	43	61	46	54	73	234	77	8951	8.8	7060
M580L	224861	224289	82%	6	45	164	122	136	151	573	190	21461	3.8	17270
m581R	224324	224608	58%	2	47	72	73	61	79	285	94	11129	9.6	17900
m582R	224481	224774	52%	3	48	72	75	67	80	294	97	11723	10.7	31840
m583L	225246	224908	68%	5	45	105	72	81	81	339	112	12615	4.5	12150
M584R	224914	225222	84%	1	45	77	74	66	92	309	102	12137	10.3	26690
m585R	225310	225516	72%	1	43	47	53	35	72	207	68	7666	7.3	5720
M586L	226085	225318	88%	6	45	214	148	201	205	768	255	29031	9.2	26180
m587R	225640	225972	50%	1	48	88	86	75	84	333	110	12464	11	15310
m588R	225711	225956	58%	3	48	63	70	49	64	246	81	9832	7.6	18440
m589L	225961	225752	56%	4	49	53	42	60	55	210	69	8334	11.7	36790
m590L	226414	226160	56%	4	41	69	56	49	81	255	84	10380	8.5	23880
M591R	226168	227928	92%	1	50	490	432	457	382	1761	586	61373	6.8	64850
m592L	226406	226179	62%	6	43	57	52	45	74	228	75	7908	11.7	0
m593L	226879	226568	48%	4	50	68	84	73	87	312	103	12106	5.9	28060
m594L	227162	226686	60%	6	50	107	142	98	130	477	158	16875	7.4	8400
m595R	226710	226961	60%	3	53	71	56	77	48	252	83	9667	10.7	27940
m596L	227504	227187	64%	6	52	67	80	85	86	318	105	10900	11.1	2560
m597R	227549	227785	64%	2	67	61	96	62	18	237	78	9370	12.1	30
m598L	228408	228193	48%	5	38	56	35	46	79	216	71	8175	6.4	9650
M599R	228277	229140	78%	1	52	196	305	143	220	864	287	30573	6.5	5180
m600L	229103	228672	58%	6	57	107	54	193	78	432	143	13337	4.5	5690
M601L	230203	229124	76%	4	42	319	206	249	306	1080	359	40296	7.4	52180
m602R	229665	229877	54%	3	40	54	48	38	73	213	70	7776	4.7	6970
m603R	229877	230149	64%	2	44	60	79	42	92	273	90	10141	9	8400
m604R	230175	230438	60%	3	41	74	52	57	81	264	87	10143	10	9590
M605L	230627	230223	80%	6	44	102	88	92	123	405	134	15193	4.9	5150
M606R	230722	231039	78%	1	48	82	72	80	84	318	105	11995	6	16530
M607L	231719	231216	80%	6	44	144	109	112	139	504	167	18475	11	3870
m608R	231258	231485	70%	3	43	72	49	49	58	228	75	8552	11.2	3900
m609L	232244	231795	52%	6	40	146	96	86	122	450	149	16807	10.7	20400
M610L	232680	232159	74%	5	43	154	118	107	143	522	173	20119	6	12830
m611L	233193	232744	60%	5	36	85	85	78	202	450	149	17372	9.5	34430

ORF	ORF Location		% A+T (-50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
M612R	232763	233836	78%	2	42	389	226	224	235	1074	357	42322	8	38530
m613L	233551	233303	50%	4	41	65	47	54	83	249	82	9897	10.6	8340
m614L	234221	233910	64%	6	48	72	101	48	91	312	103	10727	9.6	1280
M615R	233914	234342	72%	1	46	114	78	121	116	429	142	14985	10.1	8370
M616L	235330	234344	84%	4	48	269	271	204	243	987	328	35317	4.8	26210
m617R	234588	235001	70%	3	54	79	74	151	110	414	137	12171	10.5	0
M618L	235642	235415	80%	4	49	54	61	51	62	228	75	8174	4.3	3840
m619L	235658	235440	76%	6	46	59	56	45	59	219	72	8059	9.6	8430
M620L	236651	235689	84%	6	55	225	309	220	209	963	320	34619	5.2	18490
m621R	236035	236256	40%	1	56	45	63	62	52	222	73	7560	12.5	6970
M622L	237725	236676	78%	6	39	365	173	235	277	1050	349	40424	9.7	49020
m623R	236877	237317	52%	3	42	112	110	74	145	441	146	16740	10.2	14150
m624L	237337	237104	60%	4	43	73	33	67	61	234	77	9003	11.3	11440
M625R	237774	238022	78%	3	48	54	63	56	76	249	82	9215	6.8	10810
m626L	238229	237795	78%	6	46	139	96	105	95	435	144	16889	11	14650
M627L	238682	238248	52%	6	48	127	102	105	101	435	144	16778	10.2	20340
M628L	239312	239037	80%	6	59	49	65	97	65	276	91	9597	10.7	27880
M629L	240038	239340	80%	6	47	197	139	187	176	699	232	26197	4.6	8610
m630R	239412	239960	56%	3	46	137	145	110	157	549	182	20849	10.1	18500
M631L	241199	240102	70%	6	44	314	218	270	296	1098	365	41834	10.8	40140
m632R	240321	240941	58%	3	44	183	144	128	166	621	206	24642	10.4	31330
m633R	240856	241293	58%	1	44	106	112	81	139	438	145	16722	9.3	10870
M634R	241316	241591	70%	2	45	89	53	71	63	276	91	10826	4.5	20370
M635R	241632	242279	72%	3	48	196	160	153	139	648	215	24418	9.3	33150
m636L	242219	241980	56%	6	47	48	55	57	80	240	79	8977	8.3	12870
M637L	242741	242274	86%	6	40	141	79	110	138	468	155	18316	6.8	17240
M638R	242798	244504	88%	2	47	519	436	365	387	1707	568	61761	10.4	56020
m639R	243064	243261	62%	1	50	61	55	44	38	198	65	7679	6.9	7060
M640L	243558	243076	58%	5	47	104	91	136	152	483	160	16919	4.3	12120
m641L	243676	243470	50%	4	47	46	42	56	63	207	68	8027	8.8	32380
m642L	243786	243565	62%	5	48	47	50	56	69	222	73	7562	8.4	1370
m643L	243970	243710	58%	4	49	58	57	70	76	261	86	9839	6.7	23650
m644L	244353	244099	52%	5	54	42	61	77	75	255	84	8332	4.2	7000
m645R	244473	244871	66%	3	37	112	110	36	141	399	132	15773	11.2	16500
m646L	244759	244532	68%	4	37	76	14	71	67	228	75	9036	11.4	7060
M647L	245060	244569	64%	6	42	142	64	145	141	492	163	18895	3.6	21710
m648L	245910	245239	54%	5	48	125	175	145	227	672	223	24607	4.7	5180
M649R	245243	246652	80%	2	46	458	307	335	310	1410	469	53136	10.9	48930
m650L	246234	246022	64%	5	40	51	40	45	77	213	70	7649	9.6	3900
M651R	246879	247631	84%	3	44	227	159	176	191	753	250	28262	8.3	18070
m652L	247565	247155	68%	6	50	94	104	100	113	411	136	15258	10.1	6520

ORF	ORF Location		% A+T (-50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
M653L	248608	247628	66%	4	52	232	250	258	241	981	326	37172	4.5	41370
m654R	247737	248303	56%	3	53	143	158	140	126	567	188	21002	10.6	28090
m655L	248307	248041	46%	5	53	52	65	77	73	267	88	9768	12.2	210
m656R	248276	248584	48%	2	50	74	69	87	79	309	102	11360	5.5	7120
M657L	249244	248726	72%	4	44	167	112	115	125	519	172	20086	10	17870
m658R	248739	249020	72%	3	48	57	72	64	89	282	93	10388	10.8	1370
m659L	249484	249278	46%	4	61	40	72	55	40	207	68	7713	12.5	17160
M660L	249878	249282	74%	6	59	127	183	167	120	597	198	22210	4.6	17840
m661R	249468	249794	28%	3	57	66	100	87	74	327	108	11501	6.7	5780
m662R	249604	249807	42%	1	56	45	56	59	44	204	67	7525	12.2	13970
m663L	250168	249911	52%	4	44	99	63	51	45	258	85	9430	10.8	3930
M664L	251214	249919	78%	5	47	399	284	326	287	1296	431	48624	8.5	37850
m665R	250297	250518	62%	1	49	53	62	46	61	222	73	8492	8.6	2770
m666R	250307	250549	62%	2	49	58	66	53	66	243	80	8885	4.6	15250
M667R	251310	252233	74%	3	46	271	185	242	226	924	307	35055	7.1	26290
m668L	252326	251325	44%	6	46	250	261	202	289	1002	333	37718	8.9	25730
m669L	252400	252107	46%	4	47	71	69	70	84	294	97	10876	10.6	16560
M670R	252261	252611	70%	3	51	91	93	87	80	351	116	13163	10.5	25410
m671L	252679	252401	44%	4	48	70	66	67	76	279	92	10620	12.3	10840
M672L	253412	252612	62%	6	49	220	183	208	190	801	266	29614	6.6	24300
m673R	252619	252861	64%	1	52	51	72	54	66	243	80	8850	8.5	12780
M674R	253807	255768	68%	1	49	502	463	491	506	1962	653	75010	5.2	93550
m675L	254829	254320	48%	5	48	142	119	127	122	510	169	19041	10.5	9590
m676L	254668	254468	56%	4	50	46	52	49	54	201	66	7207	6.7	17190
m677L	255539	254907	52%	6	50	167	168	149	149	633	210	23332	10.4	23650
m678R	255023	255286	54%	2	50	58	56	77	73	264	87	9355	9.9	5930
m679R	255756	256118	50%	3	40	130	64	80	89	363	120	13416	10.4	6670
M680R	255802	256677	82%	1	49	253	224	206	193	876	291	31257	5	12180
m681L	256491	256264	46%	5	51	52	51	66	59	228	75	8245	8	12750
m682L	256816	256607	56%	4	43	52	36	55	67	210	69	7907	7.3	9680
M683L	257390	256686	76%	6	39	259	119	158	169	705	234	26930	10.8	19920
m684R	256735	257037	62%	1	36	72	68	41	122	303	100	11391	7.1	2650
m685R	257437	257703	76%	1	55	69	69	78	51	267	88	9271	7.1	2590
M686L	258678	257443	70%	5	42	353	303	218	362	1236	411	44515	4.1	33350
m687L	257723	257481	54%	6	56	51	74	62	56	243	80	8778	12.5	24040
m688R	257588	257968	42%	2	49	100	59	129	93	381	126	12964	4.7	11410
m689L	259013	258786	48%	6	43	65	58	40	65	228	75	8357	8.9	1370
M690R	258822	259514	82%	3	50	181	183	161	168	693	230	25407	10.7	34490
m691L	259232	259014	60%	6	50	51	56	53	59	219	72	7859	9.7	5750
m692L	260188	259511	42%	4	35	223	110	129	216	678	225	26390	7.6	17840
M693R	259590	259796	70%	3	35	60	38	34	75	207	68	7881	11.8	8280

ORF	ORF Location		% A+T (-50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
M694L	260573	260175	82%	6	49	103	89	105	102	399	132	15045	6.4	10010
M695R	260638	261162	80%	1	48	148	122	128	127	525	174	19818	7.2	29280
m696L	261340	261026	48%	4	43	80	77	60	98	315	104	11864	12.4	13970
M697L	261989	261159	62%	6	44	223	178	190	240	831	276	30770	4.2	10330
m698R	261490	261711	62%	1	46	65	53	49	55	222	73	8054	9.4	8250
M699L	262481	261990	80%	6	43	143	89	121	139	492	163	18189	10.9	20970
M700L	263013	262510	80%	5	44	161	117	106	120	504	167	19585	10.1	19090
M701R	263097	264131	74%	3	51	278	290	234	233	1035	344	37915	8.5	95440
m702R	263252	263503	50%	2	51	62	64	65	61	252	83	10081	11.7	36850
m703L	263952	263530	60%	5	53	85	79	144	115	423	140	15928	7.3	41290
M704R	264221	264751	80%	2	46	160	136	110	125	531	176	19644	11.7	19060
m705L	265052	264708	52%	6	46	91	65	95	94	345	114	12689	11.9	7060
M706L	266241	264838	66%	5	51	369	371	342	322	1404	467	51911	8.4	17010
m707R	264862	266202	68%	1	52	302	326	366	347	1341	446	50164	5.2	34430
m708L	265403	265125	46%	6	51	75	71	71	62	279	92	10271	10.8	7030
m709L	265694	265458	42%	6	52	65	75	48	49	237	78	8487	9.5	18380
m710R	265685	265894	50%	2	49	55	43	60	52	210	69	7833	12.5	17160
m711L	266105	265722	54%	6	52	92	113	88	91	384	127	14366	12.1	3900
m712R	265925	266326	46%	2	48	107	95	96	104	402	133	14941	10.7	55230
m713R	266286	267020	64%	3	44	215	153	173	194	735	244	27641	5.5	34430
M714L	266984	266685	72%	6	46	71	75	62	92	300	99	11582	8.4	16590
M715R	266983	267306	66%	1	41	93	68	64	99	324	107	12722	6.8	15930
m716L	267588	267301	68%	5	39	80	55	56	97	288	95	10823	10.4	13460
M717R	267359	268522	82%	2	41	368	251	225	320	1164	387	43763	6.7	39780
m718R	268194	268403	60%	3	41	60	42	44	64	210	69	7788	10.2	16620
M719L	269688	268519	76%	5	47	312	297	255	306	1170	389	43205	5.2	21910
m720R	269695	270039	58%	1	44	118	76	75	76	345	114	13296	8.3	2800
M721L	271560	269746	76%	5	45	416	396	426	577	1815	604	69546	7.9	1E+05
m722L	270314	270075	58%	6	48	42	62	53	83	240	79	8739	8.2	10990
m723R	270224	270478	60%	2	48	82	53	69	51	255	84	9946	11.9	10810
m724R	270551	270778	60%	2	49	61	70	42	55	228	75	8089	11.6	7000
m725R	270778	271287	44%	1	46	136	125	108	141	510	169	19092	11.4	20970
m726R	271199	271462	52%	2	43	94	59	55	56	264	87	10099	6.8	25350
M727L	271992	271633	78%	5	43	111	69	87	93	360	119	13074	10.4	10840
m728R	271756	271959	64%	1	48	56	52	45	51	204	67	7784	11.5	7060
M729L	273776	272025	84%	6	46	567	418	385	382	1752	583	65557	11.6	47920
m730R	272215	272436	68%	1	53	43	52	65	62	222	73	8097	4	19690
m731R	272950	273156	60%	1	42	55	45	42	65	207	68	7676	12.1	11410
m732L	273939	273742	60%	5	27	63	29	24	82	198	65	7596	6.4	3900
M733R	273860	274810	68%	2	38	306	174	186	285	951	316	37045	9.7	24920
m734L	274105	273863	82%	4	40	64	55	43	81	243	80	9276	9.6	2680

ORF	ORF Location		% A+T (-50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
M735L	275268	274807	70%	5	43	138	92	108	124	462	153	17288	4	5120
m736L	275201	274992	76%	6	43	64	36	55	55	210	69	7982	12.9	12660
m737R	275006	275212	54%	2	41	54	51	34	68	207	68	7203	11	2560
M738L	275902	275228	76%	4	40	233	105	168	169	675	224	25231	4.2	13400
m739L	276077	275790	48%	6	41	75	62	57	94	288	95	11218	10.8	8400
M740L	276247	275954	66%	4	47	67	68	69	90	294	97	11139	8.6	19120
M741L	276682	276332	74%	4	48	81	68	101	101	351	116	12704	8.7	26660
M742R	276753	279458	68%	3	54	651	741	713	601	2706	901	99479	6.2	89800
m743L	277075	276821	48%	4	51	64	62	67	62	255	84	9211	11.8	5750
m744L	277130	276849	42%	6	52	67	69	78	68	282	93	10646	5.6	15280
m745L	277792	277583	44%	4	50	48	57	47	58	210	69	7812	12.4	8250
m746L	278098	277802	46%	4	54	69	78	82	68	297	98	10799	9.6	18410
m747L	279611	279399	56%	6	54	40	72	44	57	213	70	7839	11.4	18410
M748L	281025	279460	80%	5	49	386	431	336	413	1566	521	57927	5.3	93060
m749R	279997	280476	46%	1	48	140	112	117	111	480	159	17779	10.2	60
m750R	281024	281278	50%	2	38	73	52	46	84	255	84	9749	11.3	9560
m751R	281164	281472	62%	1	43	93	61	72	83	309	102	11368	7	3990
M752R	281297	281656	74%	2	41	115	65	81	99	360	119	13503	9.8	29160
M753L	282564	281659	86%	5	46	253	198	219	236	906	301	33932	10.8	20400
m754R	281720	282268	58%	2	47	145	141	115	148	549	182	20562	11.5	29160
m755L	282855	282595	54%	5	49	71	68	60	62	261	86	9346	9.3	2560
M756R	282608	283159	86%	2	48	149	120	147	136	552	183	20182	3.5	12210
M757R	283302	283736	82%	3	44	144	92	100	99	435	144	16319	6.8	23500
M758R	283779	284372	64%	3	43	159	129	124	182	594	197	21904	5.1	11610
m759L	284429	284064	68%	6	41	105	77	73	111	366	121	13716	8.2	15340
m760L	284676	284329	52%	5	44	82	79	74	113	348	115	13328	7.9	22400
m761R	284444	285034	56%	2	46	176	132	137	146	591	196	21929	11.1	21060
M762L	285067	284546	82%	4	48	124	126	122	150	522	173	19325	9.7	10240
M763L	286537	285095	72%	4	48	463	328	365	287	1443	480	54550	11.3	52320
m764R	285340	285690	40%	1	48	43	88	82	138	351	116	12054	11.4	7000
m765R	286458	286763	50%	3	41	81	65	61	99	306	101	11497	8.8	8340
M766L	287657	286566	76%	6	47	265	214	296	317	1092	363	41124	5.5	57990
m767R	286794	287159	58%	3	47	95	99	72	100	366	121	13629	10.8	14000
m768L	287392	287156	60%	4	50	50	37	81	69	237	78	8795	9.1	7150
m769R	287562	287774	48%	3	36	81	41	35	56	213	70	8317	10.3	12150
M770R	287716	288120	82%	1	42	123	90	79	113	405	134	15339	9.6	21680
m771R	287942	288346	70%	2	42	113	94	77	121	405	134	15316	10.4	15310
m772L	288195	287995	58%	5	43	63	34	53	51	201	66	7466	7	5810
M773L	288496	288113	68%	4	41	127	77	80	100	384	127	14775	8.6	15420
m774R	288510	288710	64%	3	43	53	43	44	61	201	66	7647	10	2680
m775L	288923	288567	68%	6	45	91	83	78	105	357	118	13204	9.3	15250

ORF	ORF Location		% A+T (-50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
m776L	288978	288703	48%	5	42	73	62	54	87	276	91	10641	9.5	24900
M777L	291250	288953	70%	4	53	524	614	598	562	2298	765	85429	7.9	93100
m778R	289113	289319	54%	3	51	52	56	49	50	207	68	7758	9.6	24070
m779R	289217	290182	54%	2	54	225	266	253	222	966	321	35383	7.7	16770
m780R	289629	289871	52%	3	53	53	68	62	60	243	80	8855	12	7120
m781R	290315	290692	44%	2	54	92	103	102	81	378	125	13024	10	1400
m782R	290843	291055	62%	2	51	50	41	68	54	213	70	7386	4.2	120
m783R	291039	291293	46%	3	47	68	59	60	68	255	84	9702	8.8	90
m784R	291512	291964	54%	2	47	122	118	97	116	453	150	16783	9.7	19900
M785L	292819	291518	78%	4	47	347	289	328	338	1302	433	48688	11	43510
m786L	291834	291586	50%	5	49	58	56	67	68	249	82	9154	12.3	5750
m787R	292539	292775	58%	3	51	41	61	59	76	237	78	8636	11.8	7090
m788L	292928	292713	52%	6	43	69	33	60	54	216	71	8079	11	14030
M789L	293169	292837	72%	5	46	99	66	88	80	333	110	12093	9.9	9560
m790R	293075	293308	50%	2	38	80	47	43	64	234	77	9082	10.4	26630
M791R	293278	294840	78%	1	48	475	369	384	335	1563	520	56959	7.3	68000
m792L	293804	293379	52%	6	50	85	109	105	127	426	141	15178	11.2	14060
m793L	294470	294057	64%	6	49	91	100	104	119	414	137	15029	8.5	22310
M794R	294905	295717	80%	2	46	234	177	200	202	813	270	30666	6.5	16050
m795L	295699	295082	72%	4	46	156	146	141	175	618	205	23599	8.2	9740
m796L	295692	295483	68%	5	47	52	44	54	60	210	69	7689	12.4	1280
M797R	295768	296904	76%	1	53	311	256	344	226	1137	378	39914	9	80520
m798L	296146	295811	52%	4	50	64	101	67	104	336	111	12490	8.6	9770
m799L	296174	295842	60%	6	52	64	103	69	97	333	110	11650	7.7	12120
m800L	296908	296480	82%	4	58	88	139	109	93	429	142	16809	9.8	30300
M801R	297080	297835	86%	2	41	264	150	158	184	756	251	28633	7.6	15510
m802L	297748	297251	64%	4	42	120	103	108	167	498	165	19347	6.9	27490
M803L	299338	297836	70%	4	47	429	394	317	363	1503	500	53360	10.8	50930
m804R	297846	298172	74%	3	46	81	79	71	96	327	108	12009	7.4	8340
m805L	298566	298243	52%	5	48	96	78	78	72	324	107	12558	12	8340
m806R	298743	299018	54%	3	55	58	41	111	66	276	91	7839	4.1	0
M807R	299385	299852	68%	3	45	132	106	103	127	468	155	17752	4.6	9560
m808L	299794	299567	46%	4	43	59	46	52	71	228	75	8362	11.3	7030
m809L	300075	299878	62%	5	47	60	38	56	44	198	65	7513	12.2	30
M810R	299894	300145	70%	2	46	62	64	52	74	252	83	9479	10.6	6970
M811L	300904	300146	80%	4	46	211	180	168	200	759	252	29823	5.3	31270
m812R	300587	301036	48%	2	40	138	87	94	131	450	149	17651	10	24810
M813R	301197	301883	78%	3	49	166	159	175	187	687	228	25820	4.9	36250
M814L	302784	302209	72%	5	49	170	136	144	126	576	191	21819	10.6	19090
m815R	302491	302691	48%	1	46	44	52	41	64	201	66	7545	10.9	2590
m816R	302543	302773	40%	2	44	53	52	50	76	231	76	8359	11.9	5720

ORF	ORF Location		% A+T (-50 ntds from start codon)	Frame	% G+C (Gene)	# of each Nucleotide				Total Ntds	AA	MW	pI	ϵ
	Begin	End				A	C	G	T					
M817R	302866	303333	62%	1	44	133	93	115	127	468	155	18714	6.8	19830
M818R	303386	303769	82%	2	41	121	78	79	106	384	127	14320	4.2	6580
M819R	303786	304226	74%	3	43	138	91	98	114	441	146	16766	10.2	19690
M820R	304499	304957	78%	2	40	143	87	97	132	459	152	18364	6.3	10270
m821L	304786	304517	72%	4	44	66	66	53	85	270	89	10145	10	1340
M822R	305057	306856	68%	2	48	545	455	411	389	1800	599	65967	10.9	1E+05
m823L	305535	305104	66%	5	55	49	97	142	144	432	143	13299	8.8	1370
m824L	305839	305420	66%	4	45	107	96	95	122	420	139	16290	8.6	23710
m825L	306579	306358	40%	5	44	55	45	52	70	222	73	8112	8.8	1370
m826L	307347	306853	54%	5	33	186	79	86	144	495	164	19180	9.6	15370
m827L	307655	307443	50%	6	48	40	45	57	71	213	70	7765	6.8	7030
M828R	307542	308213	64%	3	51	175	157	187	153	672	223	25744	7.7	30730
M829L	309223	308345	66%	4	51	202	210	240	227	879	292	32820	6.7	37620
m830R	308525	308812	48%	2	52	74	84	67	63	288	95	10947	11	12780
m831R	308602	308865	50%	1	53	62	75	64	63	264	87	10138	9.7	5270
m832R	309017	309244	54%	2	48	63	54	55	56	228	75	8778	8.9	8370
M833L	310061	309255	68%	6	60	167	239	242	159	807	268	28812	5.8	41530
m834L	309526	309284	50%	4	62	34	67	84	58	243	80	9020	9.2	35660
m835R	309618	309974	42%	3	60	57	100	113	87	357	118	13225	5.1	38220
m836R	309842	310264	42%	2	52	76	96	123	128	423	140	16554	12.5	31160
m837L	310068	309862	62%	5	57	52	61	58	36	207	68	7844	12.6	90
M838L	310686	310312	46%	5	50	103	101	87	84	375	124	13323	8.6	1430
m839R	310328	310768	60%	2	51	99	102	123	117	441	146	16480	5.1	12780
m840L	311207	310983	38%	6	52	42	54	64	65	225	74	8031	7	90
m841R	311031	311408	66%	3	58	94	103	117	64	378	125	13987	9.9	27880
M842R	311500	312171	72%	1	47	241	158	161	112	672	223	26252	6.7	9620
m843L	312146	311541	64%	6	48	94	150	138	224	606	201	21985	11.1	42480
M844R	312868	313599	78%	1	43	241	145	173	173	732	243	27061	8.5	10360
M845L	313927	313709	68%	4	49	61	50	58	50	219	72	8513	6.3	9650

2.6.3.2 Analysis of Open Reading Frames Against the Non-redundant Database

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
M1L	RepA / Rep+ protein KID [Moorella thermoacetica ATCC 39073]	1.21E-06	ZP_00575949	25%
	hypothetical malaria antigen [Plasmodium falciparum 3D7]	1.74E-05	NP_705165	25%
	hypothetical protein BGP110 [Borrelia garinii PBi]	6.62E-05	YP_161395	27%
m2R	No Hit Found			
m3L	No Hit Found			
m4R	No Hit Found			
M5L	A81L [Paramecium bursaria Chlorella virus 1]	1.47E-23	NP_048429	46%
m6L	No Hit Found			
M7L	A84L [Paramecium bursaria Chlorella virus 1]	1.70E-12	NP_048432	37%
m8R	No Hit Found			
M9L	A75L [Paramecium bursaria Chlorella virus 1]	4.56E-60	NP_048423	42%
M10L	contains cytidine and deoxycytidine deaminase Zn-binding region signature [Paramecium bursaria Chlorella virus 1]	6.02E-39	NP_048547	62%
	FirrV-1-A29 [Feldmannia irregularis virus a]	1.15E-05	AAR26853	27%
	cytosine deaminase [Chlamydia trachomatis D/UW-3/CX]	7.43E-05	NP_220365	34%
M11L	A199R [Paramecium bursaria Chlorella virus 1]	2.19E-12	NP_048546	45%
M12R	A196L [Paramecium bursaria Chlorella virus 1]	4.05E-46	NP_048543	56%
m13L	a194R [Paramecium bursaria Chlorella virus 1]	9.90E-13	NP_048541	49%
M14R	similar to human PCNA, corresponds to Swiss-Prot Accession Number P12004 [Paramecium bursaria Chlorella virus 1]	1.15E-100	NP_048540	68%
	PREDICTED: similar to proliferating cell nuclear antigen [Canis familiaris]	1.38E-37	XP_534355	32%
	proliferative cell nuclear antigen [Zea mays]	3.07E-37	CAA55669	34%
M15L	similar to SWI/SNF chromatin remodeling complex subunit OSA2 [Homo sapiens]	0	NP_048536	41%
	proline-rich protein A189R - Chlorella virus PBCV-1	3.82E-18	T17679	30%
	hypothetical protein A192R - Chlorella virus PBCV-1	1.66E-13	T17682	31%
m16R	No Hit Found			
m17L	No Hit Found			
m18R	No Hit Found			
M19L	PBVC-1 DNA polymerase [Paramecium bursaria Chlorella virus 1]	0	NP_048532	73%
	DNA polymerase [Chlorella virus K2]	0	BAA35142	73%
	DNA polymerase >gi 281076 pir B42543 DNA-directed DNA polymerase (EC 2.7.7.7) - Chlorella virus CV-NY-2A	0	P30320	72%
m20R	No Hit Found			
m21L	No Hit Found			
m22R	No Hit Found			
M23L	No Hit Found			
M24L	No Hit Found			
m25R	A9R [Paramecium bursaria Chlorella virus 1]	1.99E-56	NP_048357	57%
	A177R [Paramecium bursaria Chlorella virus 1]	3.78E-31	NP_048525	34%
	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.10E-30	NP_049005	33%

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
M26L	contains D-isomer specific 2-hydroxyacid dehydrogenase signature; similar to E. coli D-lactate dehydrogenase, corresponds to Swiss-Prot Accession Number P52643 [Paramecium bursaria Chlorella virus 1]	2.91E-127	NP_048401	67%
	hypothetical protein MG08725.4 [Magnaporthe grisea 70-15]	5.58E-62	EAA51203	39%
	related to D-lactate dehydrogenase [Neurospora crassa]	1.68E-58	CAE81937	41%
m27L	No Hit Found			
m28R	contains type 1 hydrophobic transmembrane region [Paramecium bursaria Chlorella virus 1]	3.26E-08	NP_048404	59%
m29R	a55L [Paramecium bursaria Chlorella virus 1]	6.79E-12	NP_048403	24%
M30R	Glycerol uptake facilitator protein, GLPF [Clostridium acetobutylicum ATCC 824]	3.31E-34	AAK79288	34%
	glycerol uptake facilitator protein [Lactobacillus sakei]	8.45E-30	AAU85390	33%
	glycerol uptake facilitator protein [Clostridium perfringens str. 13]	9.34E-29	BAB80632	34%
m31L	No Hit Found			
M32L	A315L [Paramecium bursaria Chlorella virus 1]	9.86E-52	NP_048671	48%
	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	2.97E-48	NP_049007	46%
	PBCV-1 33kd peptide [Paramecium bursaria Chlorella virus 1]	1.74E-40	NP_048641	45%
m33R	No Hit Found			
M34L	similar to Synechocystis ORF s111635, corresponds to GenBank Accession Number D90903 [Paramecium bursaria Chlorella virus 1]	3.58E-74	NP_049030	61%
	thymidilate synthase [Bacteriophage S-PM2]	2.58E-56	CAF34258	54%
	Thymidylate synthase complementing protein [Pelodictyon phaeoclastratiforme BU-1] >gi 68244234 gb EAN26426.1 Thymidylate synthase complementing protein [Pelodictyon phaeoclastratiforme BU-1]	4.40E-56	ZP_00588145	51%
m35R	No Hit Found			
m36R	No Hit Found			
M37R	glutamine:fructose-6-phosphate amidotransferase GFAT [Chlorella virus]	0	BAD15299	61%
	PBCV-1 glucosamine synthetase [Paramecium bursaria Chlorella virus 1]	0	NP_048448	61%
	glucosamine--fructose-6-phosphate aminotransferase [isomerizing] [Bordetella parapertussis 12822]	3.50E-143	NP_886345	47%
m38L	No Hit Found			
m39L	No Hit Found			
m40L	No Hit Found			
m41L	No Hit Found			
M42R	Vp260 like protein [Chlorella virus]	0	BAB83471	71%
	Vp260 like protein [Chlorella virus]	0	BAB83467	43%
	Vp260 like protein [Chlorella virus]	3.83E-36	BAB83468	34%
m43L	No Hit Found			
m44L	No Hit Found			
m45L	No Hit Found			
m46R	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
M47R	Vp260 like protein [Chlorella virus]	0	BAB83470	68%
	Vp260 like protein [Chlorella virus]	0	BAB83469	56%
	Vp260 like protein [Chlorella virus]	0	BAB83468	54%
m48L	No Hit Found			
m49L	No Hit Found			
m50L	No Hit Found			
m51L	No Hit Found			
m52L	No Hit Found			
m53L	No Hit Found			
m54L	No Hit Found			
M55R	Vp260 like protein [Chlorella virus]	0	BAB83469	58%
	Vp260 like protein [Chlorella virus]	0	BAB83470	62%
	Vp260 like protein [Chlorella virus]	0	BAB83468	56%
m56L	No Hit Found			
m57R	No Hit Found			
m58L	No Hit Found			
m59L	mucin - rat (fragment) >gi 205546 gb AAA41642.1 mucin	0.000838	A39321	32%
m60L	No Hit Found			
M61R	Vp260 like protein [Chlorella virus]	0	BAB83469	58%
	Vp260 like protein [Chlorella virus]	0	BAB83470	62%
	Vp260 like protein [Chlorella virus]	0	BAB83468	56%
m62L	No Hit Found			
m63L	No Hit Found			
m64L	No Hit Found			
m65L	No Hit Found			
m66L	No Hit Found			
m67L	No Hit Found			
m68L	No Hit Found			
m69R	A354R [Paramecium bursaria Chlorella virus 1]	3.34E-49	NP_048711	42%
	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081 [Paramecium bursaria Chlorella virus 1]	0.000147	NP_048779	24%
M70L	No Hit Found			
M71R	Vp260 like protein [Chlorella virus]	0	BAB83467	62%
	Vp260 like protein [Chlorella virus]	0	BAB83469	34%
	Vp260 like protein [Chlorella virus]	0	BAB83468	33%
m72L	No Hit Found			
m73L	No Hit Found			
m74L	No Hit Found			
m75L	No Hit Found			
m76R	No Hit Found			
m77L	No Hit Found			
M78R	putative capsid protein [Chlorella virus CVG-1]	3.72E-135	AAC27494	99%
	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052 [Paramecium bursaria Chlorella virus 1]	4.47E-80	NP_048787	45%
	major capsid protein MCP1 [Chlorella virus]	2.90E-79	BAA76601	44%
m79R	No Hit Found			
m80L	No Hit Found			
M81R	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
m82L	No Hit Found			
M83R	No Hit Found			
m84L	A201L [Paramecium bursaria Chlorella virus 1]	2.66E-26	NP_048548	78%
M85R	PBCV-1 chitinase [Paramecium bursaria Chlorella virus 1]	2.86E-178	NP_048613	75%
	chitinase [Metarhizium flavoviride]	3.28E-57	CAB44709	36%
	hypothetical protein FG03591.1 [Gibberella zeae PH-1]	2.13E-56	EAA73155	36%
M86R	A301L [Paramecium bursaria Chlorella virus 1]	4.84E-08	NP_048655	32%
m87L	No Hit Found			
M88R	A298L [Paramecium bursaria Chlorella virus 1]	1.12E-37	NP_048652	45%
m89L	No Hit Found			
M90L	A296R [Paramecium bursaria Chlorella virus 1]>gi 1181459 gb AAC96664.1 A296R [Paramecium bursaria Chlorella virus 1]>gi 7461892 pir T17793 serine/arginine-rich protein A296R - Chlorella virus PBCV-1	1.68E-14	NP_048650	64%
M91R	vChta-1 [Chlorella virus]	8.10E-114	BAA20342	64%
	PBCV-1 chitosanase [Paramecium bursaria Chlorella virus 1]	1.53E-112	NP_048646	63%
	chitosanase [Amycolatopsis sp. CsO-2]	5.05E-15	BAA94840	30%
m92L	a293R [Paramecium bursaria Chlorella virus 1]	9.29E-08	NP_048647	56%
m93L	A354R [Paramecium bursaria Chlorella virus 1]	2.58E-47	NP_048711	42%
	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081 [Paramecium bursaria Chlorella virus 1]	1.75E-11	NP_048779	29%
	A87R [Paramecium bursaria Chlorella virus 1]	2.52E-10	NP_048435	32%
m94L	No Hit Found			
m95R	No Hit Found			
m96R	No Hit Found			
M97L	A286R [Paramecium bursaria Chlorella virus 1]	6.05E-99	NP_048640	53%
m98R	No Hit Found			
M99R	contains Cu/Zn superoxide dismutase signatures 1 and 2; similar to Neurospora Cu/Zn-superoxide dismutase, corresponds to Swiss-Prot Accession Number P07509 [Paramecium bursaria Chlorella virus 1]	4.32E-69	NP_048593	76%
	unnamed protein product [Yarrowia lipolytica CLIB99]	1.27E-44	CAG79443	61%
	Superoxide dismutase [Cu-Zn]	2.16E-44	Q751L8	59%
M100L	A297L [Paramecium bursaria Chlorella virus 1]	6.99E-49	NP_048651	53%
	Fructose-2,6 bisphosphatase. phosphoglycerate mutase family protein, putative [Aspergillus fumigatus Af293]	4.29E-06	EAL88370	25%
	Fructose-2,6 bisphosphatase. phosphoglycerate mutase family protein, putative [Aspergillus fumigatus Af293]>gi 70986692 ref XP_748836.1 phosphoglycerate mutase family protein [Aspergillus fumigatus Af293]	6.20E-05	EAL86798	24%
M101R	PBCV-1 amidase [Paramecium bursaria Chlorella virus 1]	2.03E-90	NP_048638	57%
	COG3049: Penicillin V acylase and related amidases [Oenococcus oeni PSU-1]	1.02E-33	ZP_00318651	34%
	Choloylglycine hydrolase [Bacillus cereus ATCC 14579]>gi 30019170 ref NP_830801.1 Choloylglycine hydrolase [Bacillus cereus ATCC 14579]	5.09E-33	AAP08002	34%
m102L	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
M103L	contains ATP/GTP-binding site motif A; similar to rat beta-alanine synthetase, corresponds to Swiss-Prot Accession Number Q03248 [Paramecium bursaria Chlorella virus 1]	3.10E-116	NP_048426	68%
	COG0388: Predicted amidohydrolase [Streptococcus pneumoniae TIGR4]	1.11E-73	ZP_00403767	47%
	putative carbon-nitrogen hydrolase [Vibrio parahaemolyticus RIMD 2210633]	1.90E-73	NP_798153	49%
m104L	No Hit Found			
m105R	No Hit Found			
m106R	No Hit Found			
M107R	A243R [Paramecium bursaria Chlorella virus 1]	4.03E-135	NP_048591	76%
m108L	No Hit Found			
M109R	A253R [Paramecium bursaria Chlorella virus 1]	4.84E-12	NP_048604	35%
	hypothetical protein A253R [Chlorella virus]	4.84E-12	AAU06294	36%
	hypothetical protein A253R [Chlorella virus]	1.08E-11	AAU06289	34%
m110L	No Hit Found			
M111R	A315L [Paramecium bursaria Chlorella virus 1]	3.47E-50	NP_048671	45%
	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	7.24E-48	NP_049007	44%
	PBCV-1 33kd peptide [Paramecium bursaria Chlorella virus 1]	6.81E-30	NP_048641	39%
m112L	No Hit Found			
m113L	No Hit Found			
M114L	A262L [Paramecium bursaria Chlorella virus 1]>gi 1181425 gb AAC96630.1 A262L [Paramecium bursaria Chlorella virus 1] >gi 7461406 pir T17759 hypothetical protein A262L - Chlorella virus PBCV-1	8.28E-27	NP_048616	47%
m115R	No Hit Found			
m116L	A265L [Paramecium bursaria Chlorella virus 1]	1.49E-49	NP_048619	56%
m117R	No Hit Found			
m118R	No Hit Found			
m119R	No Hit Found			
M120L	No Hit Found			
M121R	No Hit Found			
m122R	a136R [Paramecium bursaria Chlorella virus 1]	4.25E-08	NP_048484	34%
m123L	No Hit Found			
M124L	PBCV-1 beta-1,3 glucanase [Paramecium bursaria Chlorella virus 1]	4.11E-113	NP_048442	57%
	Glucan endo-1,3-beta-glucosidase A1 precursor ((1->3)-beta-glucan endohydrolase) ((1->3)-beta-glucanase A1)>gi 142973 gb AAA22474.1 beta-1,3-glucanase A1 (glcA)	4.22E-25	P23903	33%
	beta-1,3-glucanase [Bacillus circulans]	2.73E-24	AAC60453	33%
m125R	No Hit Found			
m126L	No Hit Found			
m127R	No Hit Found			
M128R	hyaluronan synthase [Paramecium bursaria Chlorella virus 1]	0	AAD26641	70%
	PBCV-1 hyaluronic acid synthetase [Paramecium bursaria Chlorella virus 1]	0	NP_048446	70%

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
	hyaluronan synthase [Paramecium bursaria Chlorella virus 1]	0	AAD26643	70%
m129L	No Hit Found			
m130R	No Hit Found			
m131L	No Hit Found			
m132R	No Hit Found			
M133R	PBCV-1 mRNA guanylyltransferase [Paramecium bursaria Chlorella virus 1]	2.13E-89	NP_048451	53%
	Chain B, Structure Of Guanylylated Mrna Capping Enzyme Complexed With Gtp	1.06E-88	1CKN_B	53%
	putative RNA guanylyltransferase [infectious spleen and kidney necrosis virus]	3.59E-12	AAL98788	28%
m134L	No Hit Found			
m135L	No Hit Found			
m136L	A645R [Paramecium bursaria Chlorella virus 1]	1.03E-14	NP_049001	35%
M137L	contains ubiquitin carboxy-terminal hydrolase active sites; similar to human ubiquitin carboxy-terminal hydrolase, corresponds to Swiss-Prot Accession Number Q09879 [Paramecium bursaria Chlorella virus 1]	1.67E-68	NP_048453	45%
	ubiquitin carboxyl-terminal hydrolase, putative [Entamoeba histolytica HM-1:IMSS]	7.40E-08	XP_654739	26%
	ubiquitin specific protease 2 [Gallus gallus]	2.63E-05	NP_990257	28%
m138R	No Hit Found			
M139L	similar to Pyrococcus woesei factor TFIIB homolog, corresponds to GenBank Accession Number X70668 [Paramecium bursaria Chlorella virus 1]	2.89E-51	NP_048455	36%
	Sua7p: Transcription initiation factor IIB (Swiss Prot accession number P29055) [Saccharomyces cerevisiae]	4.09E-05	AAB68135	24%
	YPR086W [Saccharomyces cerevisiae]	4.09E-05	AAT93251	24%
M140L	A109L [Paramecium bursaria Chlorella virus 1]	5.74E-29	NP_048457	58%
m141R	No Hit Found			
m142R	No Hit Found			
M143L	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	6.15E-135	NP_048636	45%
	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662 [Paramecium bursaria Chlorella virus 1]	9.51E-128	NP_048632	50%
	RPQT-like (9x) [Paramecium bursaria Chlorella virus 1]	5.46E-83	NP_048970	47%
m144R	No Hit Found			
m145R	No Hit Found			
m146L	No Hit Found			
m147R	No Hit Found			
m148L	No Hit Found			
M149R	similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank Accession Number D16505 [Paramecium bursaria Chlorella virus 1]	2.79E-34	NP_049020	52%
	similar to Chlorella virus CVK2 unknown ORF, corresponds to GenBank Accession Number D16505 [Paramecium bursaria Chlorella virus 1]	2.37E-25	NP_049021	40%
M150R	A121R [Paramecium bursaria Chlorella virus 1]	1.05E-27	NP_048469	73%
	Tlr 6Fp protein (mobile genetic element) [Tetrahymena thermophila]	1.09E-08	AAL73473	44%

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
	Tlr 6Fp protein (mobile genetic element) [Tetrahymena thermophila]	1.86E-08	AAL73467	42%
m151L	No Hit Found			
M152R	PBCV-1 Vp260 protein [Paramecium bursaria Chlorella virus 1]	5.73E-67	NP_048470	31%
	glycoprotein Vp260	9.26E-25	AAA86307	43%
	contains a leucine zipper motif; similar to Bacillus subtilis phage P2A peneck appendage protein, corresponds to Swiss-Prot Accession Number P07537 [Paramecium bursaria Chlorella virus 1]	2.34E-44	NP_048471	45%
m153L	unnamed protein product [Yarrowia lipolytica CLIB99]	2.63E-07	CAG83749	32%
	hepatitis A virus cellular receptor 1 long form [Cercopithecus aethiops]	2.02E-07	AAC39772	29%
	hepatitis A virus cellular receptor 1 short form [Cercopithecus aethiops]	2.02E-07	AAC39771	29%
m154R	No Hit Found			
m155L	unnamed protein product [Yarrowia lipolytica CLIB99]	0.000179	CAG83749	35%
	hepatitis A virus cellular receptor 1 long form [Cercopithecus aethiops]	0.000105	AAC39772	26%
	hepatitis A virus cellular receptor 1 short form [Cercopithecus aethiops]	0.000105	AAC39771	26%
m156L	unnamed protein product [Yarrowia lipolytica CLIB99]	4.15E-07	CAG83749	31%
	unknown [Arabidopsis thaliana]	0.000429	AAM64739	28%
	Hypothetical protein CBG21638 [Caenorhabditis briggsae]	0.000429	CAE74002	27%
m157L	No Hit Found			
m158L	No Hit Found			
m159L	a281R [Paramecium bursaria Chlorella virus 1]	1.56E-11	NP_048635	49%
M160R	similar to Chlorella virus PBCV-1 ORF A282L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	6.23E-56	NP_049032	38%
	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	9.74E-09	NP_048636	52%
	similar to bovine cylicin I, corresponds to Swiss-Prot Accession Number P35662 [Paramecium bursaria Chlorella virus 1]	3.47E-06	NP_048632	68%
m161L	No Hit Found			
m162L	No Hit Found			
M163L	transcription elongation factor SII [Chlorella virus]	4.23E-63	BAA04187	62%
	contains a zinc ribbon domain; similar to Chlorella virus CVU1 TFIIS-like transcription factor, corresponds to GenBank Accession Number D29631 [Paramecium bursaria Chlorella virus 1]	1.61E-62	NP_048472	62%
	transcription elongation factor TFIIS homolog - Chlorella virus CV-U1	4.68E-62	S47662	61%
m164L	No Hit Found			
m165R	No Hit Found			
M166R	A127R [Paramecium bursaria Chlorella virus 1]	9.71E-78	NP_048475	65%
M167L	A9R [Paramecium bursaria Chlorella virus 1]	4.51E-51	NP_048357	53%
	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.85E-20	NP_049005	29%

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.91E-17	NP_048807	26%
m168R	No Hit Found			
m169R	No Hit Found			
M170R	A48R [Paramecium bursaria Chlorella virus 1]	1.52E-18	NP_048396	39%
M171L	No Hit Found			
m172R	No Hit Found			
M173R	hypothetical protein SAR11_0565 [Candidatus Pelagibacter ubique HTCC1062]	9.07E-19	YP_265989	40%
M174L	putative nucleotidyltransferase [Bacteroides fragilis NCTC 9343]	1.33E-51	CAH09375	49%
	lipopolysaccharide biosynthesis protein, putative [Campylobacter lari RM2100]	9.56E-50	ZP_00368786	49%
	dTDP-glucose pyrophosphorylase [Magnetospirillum magneticum AMB-1]	3E-44	BAE48960	44%
m175R	No Hit Found			
m176R	No Hit Found			
M177L	Methyltransferase FkbM [Solibacter usitatus Ellin6076]	9.95E-10	ZP_00520036	30%
M178L	hypothetical protein ZMO1468 [Zymomonas mobilis subsp. mobilis ZM4]	1.69E-05	AAV90092	37%
M179L	A662L [Paramecium bursaria Chlorella virus 1]	1.01E-53	NP_049018	64%
	unknown protein [Arabidopsis thaliana]	1.53E-09	AAM13270	24%
	unknown protein [Oryza sativa (japonica cultivar-group)]	3.42E-09	BAD87006	25%
M180L	similar to bacteriophage T4 intron-associated endonuclease, corresponds to Swiss-Prot Accession Number P13299 [Paramecium bursaria Chlorella virus 1]	1.42E-33	NP_048482	53%
	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	3.64E-05	NP_049007	35%
	A315L [Paramecium bursaria Chlorella virus 1]>gi 1181478 gb AAC96683.1 A315L [Paramecium bursaria Chlorella virus 1]	8.12E-05	NP_048671	26%
m181R	No Hit Found			
m182R	No Hit Found			
M183R	potassium channel protein [Chlorella virus IL-2B]	2.13E-15	AAQ16135	47%
	PBCV-1 K+ ion channel protein [Paramecium bursaria Chlorella virus 1]	2.78E-15	NP_048599	48%
	potassium channel protein [Chlorella virus NY2A]	2.78E-15	AAQ16140	46%
M184L	A139L [Paramecium bursaria Chlorella virus 1]	1.86E-24	NP_048487	55%
m185R	No Hit Found			
M186R	similar to Acetobacter cellulose synthase, corresponds to Swiss-Prot Accession Number P21877 [Paramecium bursaria Chlorella virus 1]	0	NP_048829	72%
	hypothetical protein [Neurospora crassa]	7.11E-121	XP_328932	45%
	Cellulose synthase (UDP-forming) [Frankia sp. Ccl3]	1.23E-72	EAM84974	35%
m187L	No Hit Found			
m188R	No Hit Found			
m189L	No Hit Found			
M190R	surface protein [Chlorella virus]	1.45E-07	BAE02830	48%
	PBCV-1 surface protein [Paramecium bursaria Chlorella virus 1]	0.000196	NP_048488	42%

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
	surface protein [Paramecium bursaria Chlorella virus 1]	0.000571	BAD22850	43%
m191R	No Hit Found			
m192L	No Hit Found			
m193L	No Hit Found			
m194R	No Hit Found			
m195L	No Hit Found			
m196L	No Hit Found			
m197L	No Hit Found			
m198L	No Hit Found			
m199R	No Hit Found			
m200L	A150L [Paramecium bursaria Chlorella virus 1]	4.31E-13	NP_048498	39%
M201R	similar to phage T5 helicase, corresponds to Swiss-Prot Accession Number P11107 [Paramecium bursaria Chlorella virus 1]	7.97E-174	NP_048501	66%
	EsV-1-66 [Ectocarpus siliculosus virus]	2.66E-44	AAK14489	34%
	VV A18 helicase [Acanthamoeba polyphaga mimivirus]	1.05E-40	YP_142750	30%
m202L	No Hit Found			
m203L	No Hit Found			
M204L	A157L [Paramecium bursaria Chlorella virus 1]	5.34E-19	NP_048505	44%
m205R	No Hit Found			
M206L	No Hit Found			
m207L	No Hit Found			
M208L	No Hit Found			
m209R	No Hit Found			
M210L	A165L [Paramecium bursaria Chlorella virus 1]	2.20E-12	NP_048513	47%
m211R	No Hit Found			
m212R	No Hit Found			
M213L	A165L [Paramecium bursaria Chlorella virus 1]	6.44E-17	NP_048513	38%
m214L	No Hit Found			
M215R	PBCV-1 exonuclease [Paramecium bursaria Chlorella virus 1]	9.46E-80	NP_048514	60%
	EsV-1-64 [Ectocarpus siliculosus virus]	4.62E-18	AAK14487	29%
	unknown protein [Arabidopsis thaliana]	1.54E-13	AAO64106	32%
m216L	No Hit Found			
m217L	No Hit Found			
M218R	A168R [Paramecium bursaria Chlorella virus 1]	1.13E-28	NP_048516	54%
M219L	similar to E. coli hypothetical protein, corresponds to Swiss-Prot Accession Number P39407 [Paramecium bursaria Chlorella virus 1]	3.47E-87	NP_048521	58%
	Patatin-like phospholipase family [Bacillus cereus G9241]	2.99E-22	ZP_00240206	37%
	hypothetical protein PBPRB1472 [Photobacterium profundum SS9]	8.43E-17	YP_133138	30%
m220R	No Hit Found			
M221L	protein kinase A248R [Chlorella virus]	4.82E-48	AAU06282	38%
	protein kinase A248R [Chlorella virus]	8.23E-48	AAU06280	38%
	protein kinase A248R [Chlorella virus]	8.23E-48	AAU06275	38%
m222R	a249L [Paramecium bursaria Chlorella virus 1]	1.36E-06	NP_048598	44%
m223R	No Hit Found			
m224L	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
M225L	contains ATP-GTP binding motif; similar to <i>Saccharomyces cerevisiae</i> antiviral protein SKI2, corresponds to Swiss-Prot Accession Number P35207 [Paramecium bursaria Chlorella virus 1]	0	NP_048589	51%
	DEAD/DEAH box helicase, putative [Entamoeba histolytica HM-1:IMSS]	4.19E-85	XP_652111	41%
	PREDICTED: similar to ENSANGP00000020973 [Apis mellifera]	6.69E-83	XP_624031	41%
m226R	No Hit Found			
m227R	No Hit Found			
m228L	No Hit Found			
m229R	No Hit Found			
m230R	No Hit Found			
m231R	No Hit Found			
M232R	No Hit Found			
M233L	PBCV-1 homospermidine synthase [Paramecium bursaria Chlorella virus 1]	0	NP_048585	70%
	Homospermidine synthase [Pelodictyon phaeoclathratiforme BU-1]	3.66E-90	ZP_00589757	39%
	homospermidine synthase [Methanosc礼ina acetivorans str. C2A]	1.82E-81	AAM05046	37%
m234R	a238L [Paramecium bursaria Chlorella virus 1]	2.10E-15	NP_048586	66%
m235R	No Hit Found			
m236R	No Hit Found			
m237L	No Hit Found			
M238R	A234L [Paramecium bursaria Chlorella virus 1]	2.01E-18	NP_048582	43%
m239L	No Hit Found			
M240L	A233R [Paramecium bursaria Chlorella virus 1]	4.20E-16	NP_048581	47%
M241L	Hypothetical protein CBG01566 [Caenorhabditis briggsae]	4.41E-05	CAE58431	32%
	EsV-1-128 [Ectocarpus siliculosus virus]	0.000286	AAK14546	35%
M242L	A230R [Paramecium bursaria Chlorella virus 1]	2.05E-37	NP_048578	40%
M243R	No Hit Found			
m244R	A227L [Paramecium bursaria Chlorella virus 1]	2.05E-39	NP_048575	61%
	MobD.6 hypothetical protein [Enterobacteria phage T4]	9.61E-05	AAD42598	27%
m245L	a228R [Paramecium bursaria Chlorella virus 1]	2.84E-12	NP_048576	51%
M246L	No Hit Found			
M247R	A9R [Paramecium bursaria Chlorella virus 1]	1.25E-61	NP_048357	64%
	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	4.79E-21	NP_049005	34%
	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	6.91E-20	NP_048807	32%
m248R	No Hit Found			
M249R	A567L [Paramecium bursaria Chlorella virus 1]	1.13E-13	NP_048923	38%
m250L	No Hit Found			
m251R	No Hit Found			
m252L	No Hit Found			
M253L	A565R [Paramecium bursaria Chlorella virus 1]	1.34E-15	NP_048921	60%
	DNA binding protein [Chlorella virus]	4.17E-09	BAA11343	54%

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472 [Paramecium bursaria Chlorella virus 1]	1.59E-08	NP_048741	51%
m254R	No Hit Found			
m255L	No Hit Found			
m256R	No Hit Found			
m257R	No Hit Found			
M258R	similar to Chlorella virus CVK2 DNA binding protein, corresponds to GenBank Accession Number D78305 [Paramecium bursaria Chlorella virus 1]	9.24E-35	NP_048917	39%
	DNA binding protein [Chlorella virus]	1.13E-32	BAA11342	38%
	alginate lyase [Chlorella virus]	3.53E-18	BAA83789	37%
m259R	No Hit Found			
m260R	No Hit Found			
m261R	No Hit Found			
m262L	No Hit Found			
M263R	A559L [Paramecium bursaria Chlorella virus 1]	3.12E-12	NP_048915	28%
M264L	deoxyuridine triphosphatase [Paramecium bursaria Chlorella virus IL3A]	4.21E-56	AAW51452	76%
	similar to tomato dUTP pyrophosphatase, corresponds to GenBank Accession Number S40549 [Paramecium bursaria Chlorella virus 1]	5.50E-56	NP_048907	75%
	deoxyuridine triphosphatase [Paramecium bursaria Chlorella virus SH6A]	2.73E-55	AAW51453	75%
m265L	No Hit Found			
M266R	similar to Sulfolobus TATA-binding protein, corresponds to GenBank Accession Number S55311 [Paramecium bursaria Chlorella virus 1]	8.04E-44	NP_048908	43%
	hypothetical protein A552R - Chlorella virus PBCV-1	1.47E-37	T18054	43%
	TATA-box binding protein (TBP) [Acanthamoeba polyphaga mimivirus]	7.35E-05	YP_142807	31%
m267L	No Hit Found			
m268R	No Hit Found			
M269L	similar to Chlorella virus PBCV-1 ORF A11L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	3.58E-143	NP_048914	60%
	contains aminoacyl-tRNA synthetase class-II signature [Paramecium bursaria Chlorella virus 1]	4.42E-109	NP_048359	49%
	similar to PBCV-1 major capsid protein, corresponds to Swiss-Prot Accession Number P30328 [Paramecium bursaria Chlorella virus 1]	1.81E-86	NP_048358	40%
m270R	No Hit Found			
m271L	No Hit Found			
M272L	similar to Caenorhabditis transcription activator, corresponds to Swiss-Prot Accession Number P41877 [Paramecium bursaria Chlorella virus 1]	2.39E-125	NP_048904	52%
	helicase [Oceanobacillus iheyensis HTE831]	4.04E-40	NP_693997	32%
	swf/snf helicase [Chlamydophila pneumoniae TW-183]	2.89E-38	AAP98807	31%
m273R	No Hit Found			
m274R	No Hit Found			
m275R	No Hit Found			
m276L	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
m277L	No Hit Found			
M278R	No Hit Found			
m279R	No Hit Found			
m280L	No Hit Found			
M281L	No Hit Found			
M282R	A519L [Paramecium bursaria Chlorella virus 1]	4.87E-20	NP_048875	60%
M283R	orf1 [Chlorella virus]	9.53E-154	BAA83788	68%
	similar to bovine monoamine oxidase, corresponds to Swiss-Prot Accession Number P21398 [Paramecium bursaria Chlorella virus 1]	1.68E-150	NP_048564	67%
m284L	No Hit Found			
m285L	No Hit Found			
m286L	No Hit Found			
m287L	No Hit Found			
M288L	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440 [Paramecium bursaria Chlorella virus 1]	0.000931	NP_048851	37%
M289R	PBCV-1 alginate lyase [Paramecium bursaria Chlorella virus 1]	1.18E-113	NP_048562	67%
	vAL-1 [Chlorella virus]	2.63E-113	BAB19127	57%
	alginate lyase [Chlorella virus]	2.23E-112	BAA83789	56%
m290R	No Hit Found			
m291L	No Hit Found			
m292L	No Hit Found			
m293L	No Hit Found			
M294R	No Hit Found			
m295L	No Hit Found			
m296L	No Hit Found			
m297L	No Hit Found			
m298R	A214L [Paramecium bursaria Chlorella virus 1]	1.21E-29	NP_048561	45%
M299R	A213L [Paramecium bursaria Chlorella virus 1]	7.18E-32	NP_048560	52%
m300L	No Hit Found			
m301L	No Hit Found			
M302L	No Hit Found			
m303R	No Hit Found			
m304L	No Hit Found			
m305R	No Hit Found			
m306R	No Hit Found			
M307L	PBCV-1 arginine decarboxylase [Paramecium bursaria Chlorella virus 1]	6.99E-138	NP_048554	63%
	ornithine decarboxylase [Lotus corniculatus var. japonicus]	9.63E-71	CAE02644	41%
	ornithine decarboxylase [Nicotiana tabacum]	4.04E-69	CAA71498	40%
M308L	A205R [Paramecium bursaria Chlorella virus 1]	1.04E-43	NP_048552	48%
m309R	No Hit Found			
M310L	A203R [Paramecium bursaria Chlorella virus 1]	3.95E-71	NP_048550	63%
m311R	a204L [Paramecium bursaria Chlorella virus 1]	9.50E-16	NP_048551	65%
M312R	A202L [Paramecium bursaria Chlorella virus 1]	1.29E-25	NP_048549	53%

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
M313L	similar to human protein Tyr-phosphatase, corresponds to GenBank Accession Number U27193 [Paramecium bursaria Chlorella virus 1]	3.51E-63	NP_048659	66%
	Zgc:77247 [Danio rerio]	1.89E-16	AAH67137	35%
	PREDICTED: similar to KIAA1700 protein [Canis familiaris]	1.22E-15	XP_543810	35%
M314L	A306L [Paramecium bursaria Chlorella virus 1]	3.06E-22	NP_048660	84%
m315L	A308L [Paramecium bursaria Chlorella virus 1]	7.37E-05	NP_048663	41%
m316R	No Hit Found			
M317L	A310L [Paramecium bursaria Chlorella virus 1]	3.35E-45	NP_048665	54%
m318R	No Hit Found			
m319L	PBCV-1 33kDa translation peptide [Paramecium bursaria Chlorella virus 1]	2.19E-92	NP_048667	71%
	EsV-1-76 [Ectocarpus siliculosus virus]	2.83E-15	AAK14499	31%
	FirrV-1-I1 [Feldmannia irregularis virus a]	2.83E-07	AAR26966	28%
m320R	No Hit Found			
m321L	No Hit Found			
M322L	A313L [Paramecium bursaria Chlorella virus 1]	0.000377	NP_048669	32%
M323R	A314R [Paramecium bursaria Chlorella virus 1]	2.30E-30	NP_048670	83%
M324R	membrane protein, putative [Dehalococcoides ethenogenes 195]	6.13E-17	YP_181494	26%
	hypothetical protein [Pyrococcus furiosus DSM 3638]	3.15E-13	AAL80842	24%
	unknown [Methanothermobacter thermautotrophicus str. Delta H]	3.48E-12	AAB85442	27%
m325L	No Hit Found			
M326R	A320R [Paramecium bursaria Chlorella virus 1]	3.83E-17	NP_048676	37%
	DNA binding protein [Chlorella virus]	0.000285	BAA11342	31%
m327L	No Hit Found			
M328R	A321R [Paramecium bursaria Chlorella virus 1]	4.92E-09	NP_048677	35%
M329L	A322L [Paramecium bursaria Chlorella virus 1]	5.16E-27	NP_048678	41%
m330R	No Hit Found			
m331R	No Hit Found			
M332L	A324L [Paramecium bursaria Chlorella virus 1]	6.19E-109	NP_048680	56%
	FirrV-1-B22 [Feldmannia irregularis virus a]	3.61E-16	AAR26897	31%
	EsV-1-103 [Ectocarpus siliculosus virus]	1.42E-12	AAK14521	28%
m333L	No Hit Found			
M334R	No Hit Found			
m335L	A326L [Paramecium bursaria Chlorella virus 1]	8.13E-63	NP_048682	60%
m336R	No Hit Found			
M337L	A328L [Paramecium bursaria Chlorella virus 1]	7.48E-70	NP_048684	38%
m338R	No Hit Found			
m339R	No Hit Found			
m340L	No Hit Found			
m341R	PBCV-1 33kDa translation peptide [Paramecium bursaria Chlorella virus 1]	7.52E-93	NP_048667	71%
	EsV-1-76 [Ectocarpus siliculosus virus]	2.83E-15	AAK14499	31%
	FirrV-1-I1 [Feldmannia irregularis virus a]	1.66E-07	AAR26966	28%
m342L	No Hit Found			
m343L	No Hit Found			
M344L	A337L [Paramecium bursaria Chlorella virus 1]	2.93E-10	NP_048693	48%
m345L	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
m346L	No Hit Found			
m347R	No Hit Found			
M348L	A342L [Paramecium bursaria Chlorella virus 1]	0	NP_048699	68%
m349R	No Hit Found			
m350L	No Hit Found			
m351R	No Hit Found			
m352L	No Hit Found			
m353L	No Hit Found			
M354R	cellulase precursor [Alicyclobacillus acidocaldarius]	6.09E-06	CAD86595	29%
	putative cellulase [uncultured bacterium]	0.000256	CAF22222	28%
m355R	No Hit Found			
m356R	No Hit Found			
M357L	URF14.2 [Chlorella virus]	5.29E-35	BAA22200	66%
	URF14.2 [Chlorella virus]	6.91E-35	BAA22202	66%
	URF14.2 [Chlorella virus]	9.02E-35	BAA22201	65%
m358R	a350R [Paramecium bursaria Chlorella virus 1]	1.34E-36	NP_048707	50%
	hypothetical protein DDB0188233 [Dictyostelium discoideum]	0.00023	XP_636443	28%
M359L	M.CviAII cytosine DNA methyltransferase [Paramecium bursaria Chlorella virus 1]	3.49E-125	NP_048873	63%
	M.CviAIV cytosine DNA methyltransferase [Paramecium bursaria Chlorella virus 1]	3.61E-122	NP_048886	62%
	CviPII m5C DNA methyltransferase [Chlorella virus]	2.14E-82	AAV84097	48%
m360R	No Hit Found			
m361R	a518R [Paramecium bursaria Chlorella virus 1]	3.18E-24	NP_048874	76%
	a529L [Paramecium bursaria Chlorella virus 1]	9.61E-05	NP_048885	64%
M362L	Asp/Glu rich; DAEDDDIYxxET (2X) negative charge cluster [Paramecium bursaria Chlorella virus 1]	1.11E-59	NP_048709	59%
	unknown [Acanthamoeba polyphaga mimivirus]	2.05E-05	YP_142843	33%
M363L	A357L [Paramecium bursaria Chlorella virus 1]	2.49E-38	NP_048714	52%
m364R	No Hit Found			
m365L	No Hit Found			
m366R	No Hit Found			
M367L	A131L [Paramecium bursaria Chlorella virus 1]	5.03E-10	NP_048479	36%
m368L	No Hit Found			
m369L	No Hit Found			
M370R	A315L [Paramecium bursaria Chlorella virus 1]	1.49E-63	NP_048671	49%
	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	4.21E-58	NP_049007	47%
	PBCV-1 33kd peptide [Paramecium bursaria Chlorella virus 1]	2.74E-49	NP_048641	42%
m371L	No Hit Found			
M372R	similar to chicken vitellogenin II, corresponds to Swiss-Prot Accession Number P02845 [Paramecium bursaria Chlorella virus 1]	0	NP_048720	61%
	A360R [Paramecium bursaria Chlorella virus 1]	8.08E-65	NP_048717	59%
	A361R [Paramecium bursaria Chlorella virus 1]	4.19E-29	NP_048718	72%
m373L	No Hit Found			
m374L	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
M375R	translation elongation factor EF-3 homolog - Chlorella virus CVK2 >gi 454301 dbj BAA03956.1 translation elongation factor-3 [Chlorella virus]	1.01E-08	A48779	39%
	A646L [Paramecium bursaria Chlorella virus 1]	1.23E-06	NP_049002	41%
M376R	No Hit Found			
M377L	A379L [Paramecium bursaria Chlorella virus 1]	8.91E-40	NP_048736	47%
m378L	No Hit Found			
M379R	A486L [Paramecium bursaria Chlorella virus 1]	6.85E-19	NP_048842	39%
m380L	No Hit Found			
M381R	similar to PBCV-1 major capsid protein, corresponds to Swiss-Prot Accession Number P30328 [Paramecium bursaria Chlorella virus 1]	2.50E-77	NP_048740	53%
	hypothetical major capsid protein [Heterosigma akashiwo virus 01]	1.34E-38	BAE06835	34%
	major capsid protein MCP1 [Chlorella virus]	1.59E-31	BAA76601	33%
m382L	No Hit Found			
m383L	No Hit Found			
m384L	No Hit Found			
M385L	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563 [Paramecium bursaria Chlorella virus 1]	3.04E-10	NP_048441	45%
	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	1.08E-15	NP_048636	52%
	a91L [Paramecium bursaria Chlorella virus 1]	8.23E-24	NP_048439	50%
m386R	No Hit Found			
m387R	a281R [Paramecium bursaria Chlorella virus 1]	2.01E-14	NP_048635	40%
M388R	hypothetical protein PSSM4_035 [Cyanophage P-SSM4] >gi 61563781 gb AAX46836.1 hypothetical protein PSSM4_035 [Cyanophage P-SSM4]	1.33E-06	YP_214596	29%
M389R	similar to MesJ cell cycle protein [Chlorobium tepidum]	1.11E-147	NP_048910	53%
	ATPase homolog A554L - Chlorella virus PBCV-1	9.57E-83	T18056	55%
	hypothetical protein A557L - Chlorella virus PBCV-1	7.97E-21	T18059	52%
m390L	No Hit Found			
m391L	No Hit Found			
m392L	No Hit Found			
m393R	No Hit Found			
m394L	No Hit Found			
M395L	No Hit Found			
m396R	No Hit Found			
M397R	A9R [Paramecium bursaria Chlorella virus 1]	1.17E-54	NP_048357	57%
	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.22E-19	NP_048807	26%
	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.76E-18	NP_049005	28%
m398R	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
M399L	PBCV-1 RNA triphosphatase [Paramecium bursaria Chlorella virus 1] >gi 1620121 gb AAC96817.1 PBCV-1 RNA triphosphatase [Paramecium bursaria Chlorella virus 1] >gi 7461588 pir T17952 hypothetical protein A449R - Chlorella virus PBCV-1	1.44E-45	NP_048806	55%
m400L	No Hit Found			
M401R	PBCV-1 prolyl 4-hydroxylase [Paramecium bursaria Chlorella virus 1] >gi 1131429 gb AAC96453.1 PBCV-1 prolyl 4-hydroxylase [Paramecium bursaria Chlorella virus 1]	1.79E-58	NP_048433	56%
	hypothetical protein Raeut03003807 [Ralstonia eutropha JMP134]	5.28E-18	ZP_00351069	34%
	HYPOTHETICAL PROTEIN [Ralstonia solanacearum]	3.42E-17	CAD16521	34%
m402L	No Hit Found			
M403R	contains cytochrome C family heme-binding site signature; similar to maize protein disulphide isomerase, correspond to Swiss-Prot Accession Number P52588 [Paramecium bursaria Chlorella virus 1]	3.72E-28	NP_048805	56%
	transglutaminase [Brugia malayi]	7.04E-11	AAQ23042	37%
	protein disulfide isomerase ER-60 [Takifugu rubripes]	1.20E-10	AAF78087	36%
M404R	similar to Clostridium pasteurianum ORF, corresponds to GenBank Accession Number Z28353 [Paramecium bursaria Chlorella virus 1]	3.90E-176	NP_048802	64%
	COG0661: Predicted unusual protein kinase [Trichodesmium erythraeum IMS101]	4.49E-31	ZP_00327806	26%
	hypothetical protein syc0672_c [Synechococcus elongatus PCC 6301]	6.49E-30	YP_171382	27%
m405L	a446R [Paramecium bursaria Chlorella virus 1]	2.13E-07	NP_048803	60%
m406L	No Hit Found			
M407R	A444L [Paramecium bursaria Chlorella virus 1]	6.20E-12	NP_048801	35%
m408L	No Hit Found			
M409R	No Hit Found			
m410L	No Hit Found			
m411L	A441L [Paramecium bursaria Chlorella virus 1]	2.83E-28	NP_048798	46%
m412R	No Hit Found			
M413L	GA14074-PA [Drosophila pseudoobscura]	4.12E-14	EAL29245	34%
	hypothetical protein AN8019.2 [Aspergillus nidulans FGSC A4]	6.13E-18	EAA59641	37%
	GH01626p [Drosophila melanogaster]	6.97E-38	AAM11327	38%
m414R	hypothetical protein WwSim0361 [Wolbachia endosymbiont of Drosophila simulans]	1.04E-06	ZP_00372892	42%
	hypothetical protein WwSim0771 [Wolbachia endosymbiont of Drosophila simulans]	3.34E-05	ZP_00372930	39%
	PREDICTED: similar to seven transmembrane helix receptor [Mus musculus]	9.73E-05	XP_144682	33%
m415R	No Hit Found			
M416R	A443R [Paramecium bursaria Chlorella virus 1]	1.44E-44	NP_048800	34%
M417L	surface protein [Chlorella virus]	7.72E-05	BAE02830	39%
	PBCV-1 surface protein [Paramecium bursaria Chlorella virus 1]	0.0005	NP_048488	32%
	surface protein [Chlorella virus CVK2]	0.000383	BAD12236	36%
m418R	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
m419R	No Hit Found			
m420R	No Hit Found			
m421L	A439R [Paramecium bursaria Chlorella virus 1]	6.25E-28	NP_048796	61%
m422L	No Hit Found			
M423R	similar to E. coli glutaredoxin, corresponds to Swiss-Prot Accession Number P37687 [Paramecium bursaria Chlorella virus 1] probable peroxiredoxin/glutaredoxin family protein [Chromobacterium violaceum ATCC 12472] GrxC, glutaredoxin 3 [Brucella abortus biovar 1 str. 9-941]	7.77E-18 1.16E-05 1.52E-05	NP_048795 AAQ59708 YP_222529	55% 39% 32%
m424L	No Hit Found			
M425R	contains ATP/GTP-binding site motif A; similar to Bacillus subtilis 24.1 kDa protein, corresponds to Swiss-Prot Accession Number P37530 [Paramecium bursaria Chlorella virus 1] 143R [Chilo iridescent virus] deoxynucleoside kinases [Lymphocystis disease virus - isolate China]	2.10E-47 1.12E-16 8.91E-14	NP_048773 AAB94460 YP_073536	51% 32% 33%
m426L	No Hit Found			
M427R	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903 [Paramecium bursaria Chlorella virus 1] A154L [Paramecium bursaria Chlorella virus 1] similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.04E-71 1.84E-68 3.84E-66	NP_048477 NP_048502 NP_048920	45% 41% 40%
m428L	No Hit Found			
m429L	No Hit Found			
M430L	A417L [Paramecium bursaria Chlorella virus 1]	1.33E-53	NP_048774	40%
m431R	No Hit Found			
m432R	No Hit Found			
m433R	No Hit Found			
m434R	No Hit Found			
m435L	A420L [Paramecium bursaria Chlorella virus 1]	1.86E-11	NP_048777	48%
m436R	A421R [Paramecium bursaria Chlorella virus 1]	1.90E-16	NP_048778	53%
M437R	No Hit Found			
M438L	No Hit Found			
m439L	No Hit Found			
M440R	A423R [Paramecium bursaria Chlorella virus 1]	3.69E-15	NP_048780	37%
m441L	No Hit Found			
m442R	No Hit Found			
m443L	No Hit Found			
M444R	A426R [Paramecium bursaria Chlorella virus 1]	3.14E-19	NP_048783	42%
M445L	contains thioredoxin active site-like sequence; similar to Synechocystis thioredoxin-like protein, corresponds to Swiss-Prot Accession Number P52232 [Paramecium bursaria Chlorella virus 1] THIOREDOXIN TRXC (TRX) (MPT46) [Mycobacterium tuberculosis H37Rv] thioredoxin h2 - spinach (fragments)	1.97E-21 2.34E-06 3.99E-06	NP_048784 NP_218431 S15137	44% 32% 37%
m446R	No Hit Found			
m447R	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
M448L	contains thioredoxin active site-like sequence; similar to Synechocystis thioredoxin-like protein, corresponds to Swiss-Prot Accession Number P52232 [Paramecium bursaria Chlorella virus 1]	1.34E-06	NP_048784	28%
M449L	contains thioredoxin active site-like sequence; similar to Synechocystis thioredoxin-like protein, corresponds to Swiss-Prot Accession Number P52232 [Paramecium bursaria Chlorella virus 1]	3.89E-06	NP_048784	34%
m450R	No Hit Found			
M451L	A429L [Paramecium bursaria Chlorella virus 1]	5.50E-55	NP_048786	34%
	ankyrin repeat family protein [Arabidopsis thaliana]	2.77E-06	AAY78871	27%
	ankyrin-like protein [Arabidopsis thaliana]	1.37E-05	BAB03143	23%
m452R	No Hit Found			
m453R	No Hit Found			
m454R	No Hit Found			
M455R	A601R [Paramecium bursaria Chlorella virus 1]	7.58E-18	NP_048957	58%
	A656L [Paramecium bursaria Chlorella virus 1]	0.000214	NP_049012	42%
m456L	No Hit Found			
m457R	contains ATP/GTP-binding motif A [Paramecium bursaria Chlorella virus 1]	3.16E-138	NP_048579	64%
	similar to Chlorella virus PBCV-1 ORF A231L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.22E-81	NP_048983	44%
m458L	No Hit Found			
m459L	No Hit Found			
M460R	putative GIY-YIG endonuclease [Bacillus sp. BSG40]	2.10E-08	CAC51107	31%
	GIY-YIG catalytic domain containing protein; possible intron encoded endonuclease [Bacillus cereus E33L]	5.18E-07	YP_085011	34%
	GIY ND1 i4 grp IB protein b [Podospora anserina]	2.84E-05	CAA38813	35%
m461L	No Hit Found			
m462L	No Hit Found			
M463L	major capsid protein Vp49 [Chlorella virus]	0	AAC27492	88%
	PBCV-1 major capsid protein Vp54, corresponds to GenBank Accession Number M85052 [Paramecium bursaria Chlorella virus 1]	0	NP_048787	77%
	major capsid protein Vp54 [Chlorella virus]	0	BAA22198	76%
m464R	No Hit Found			
M465L	A315L [Paramecium bursaria Chlorella virus 1]	2.69E-61	NP_048671	46%
	PBCV-1 33kd peptide [Paramecium bursaria Chlorella virus 1]	1.53E-56	NP_048641	44%
	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.30E-47	NP_049007	40%
m466R	No Hit Found			
M467R	A114R [Paramecium bursaria Chlorella virus 1]	0	NP_048462	64%
	A111R [Paramecium bursaria Chlorella virus 1]	3.04E-138	NP_048459	59%
	possible glycosyltransferase [Haemophilus ducreyi 35000HP]	2.26E-16	AAP95345	35%
m468L	No Hit Found			
m469L	No Hit Found			
m470L	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
m471L	No Hit Found			
m472R	Lys-rich [Paramecium bursaria Chlorella virus 1]	1.10E-24	NP_048789	46%
m473L	No Hit Found			
m474L	No Hit Found			
M475R	A571R [Paramecium bursaria Chlorella virus 1]	1.15E-37	NP_048927	79%
	Arg-rich [Paramecium bursaria Chlorella virus 1]	2.11E-07	NP_048792	61%
	A205R [Paramecium bursaria Chlorella virus 1]	0.000831	NP_048552	57%
M476L	similar to Methanothrix chromosomal protein MC1A, corresponds to Swiss-Prot Accession Number P15251 [Paramecium bursaria Chlorella virus 1]	1.34E-32	NP_048794	70%
	Chromosomal protein MC1a	1.65E-06	P15249	44%
	Chromosomal protein MC1b	1.83E-05	P15250	45%
m477R	No Hit Found			
M478L	A410L [Paramecium bursaria Chlorella virus 1]	1.46E-24	NP_048767	49%
	A408L [Paramecium bursaria Chlorella virus 1]	5.21E-14	NP_048765	33%
	EsV-1-42 [Ectocarpus siliculosus virus]	1.16E-05	AAK14468	32%
M479R	A411R [Paramecium bursaria Chlorella virus 1]	3.67E-14	NP_048768	38%
m480L	No Hit Found			
m481L	No Hit Found			
M482R	A412R [Paramecium bursaria Chlorella virus 1]	5.65E-44	NP_048769	50%
m483R	No Hit Found			
M484L	Gln-rich, QQQM(4x); similar to human transcription factor TFIID, corresponds to Swiss-Prot Accession Number P20226 [Paramecium bursaria Chlorella virus 1]	4.32E-35	NP_048770	38%
M485R	A408L [Paramecium bursaria Chlorella virus 1]	4.08E-53	NP_048765	54%
	EsV-1-42 [Ectocarpus siliculosus virus]	2.77E-09	AAK14468	28%
	A410L [Paramecium bursaria Chlorella virus 1]	6.82E-08	NP_048767	31%
m486L	No Hit Found			
m487R	hypothetical protein Bmal10_01004924 [Burkholderia mallei 10399]	8.67E-07	ZP_00433134	29%
	COG3210: Large exoproteins involved in heme utilization or adhesion [Magnetospirillum magnetotacticum MS-1]	0.000106	ZP_00054046	31%
	PREDICTED: similar to Hypothetical protein CBG20540 [Rattus norvegicus]	0.000687	XP_344436	20%
M488L	COG4991: Uncharacterized protein with a bacterial SH3 domain homologue [Burkholderia fungorum LB400]	6.23E-05	ZP_00283849	38%
	hsp70-like protein [Plasmodium chabaudi]	6.64E-15	AAK14819	46%
	procollagen type I alpha 2 chain [Danio rerio]	0.000106	CAC51030	32%
m489R	Extraembryonic, spermatogenesis, homeobox 1-like [Homo sapiens]	2.06E-09	AAH42633	42%
	ESX1L [Homo sapiens]	1.38E-05	AAM62141	37%
	COG4991: Uncharacterized protein with a bacterial SH3 domain homologue [Burkholderia fungorum LB400]	1.02E-08	ZP_00283849	37%
m490L	No Hit Found			
M491R	No Hit Found			
m492L	No Hit Found			
m493L	No Hit Found			
m494L	No Hit Found			
M495R	A407L [Paramecium bursaria Chlorella virus 1]	4.05E-83	NP_048764	70%
	unknown [Acanthamoeba polyphaga mimivirus]	0.000287	YP_142822	24%

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
M496L	contains Gln-rich, neutral zinc metallopeptidase, zinc binding region signature [Paramecium bursaria Chlorella virus 1]	1.81E-56	NP_048877	63%
	metal-dependent hydrolase [Acanthamoeba polyphaga mimivirus]	1.40E-08	YP_142679	32%
	136R [Chilo iridescent virus]	0.000123	AAB94453	36%
M497R	A523R [Paramecium bursaria Chlorella virus 1]	1.23E-59	NP_048879	74%
m498L	No Hit Found			
m499R	No Hit Found			
m500L	No Hit Found			
M501R	A526R [Paramecium bursaria Chlorella virus 1]	8.82E-30	NP_048882	52%
M502L	A315L [Paramecium bursaria Chlorella virus 1]	7.49E-64	NP_048671	55%
	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	4.25E-59	NP_049007	49%
	PBCV-1 33kd peptide [Paramecium bursaria Chlorella virus 1]	9.52E-43	NP_048641	45%
m503R	No Hit Found			
M504R	A527R [Paramecium bursaria Chlorella virus 1]	1.93E-13	NP_048883	62%
m505R	No Hit Found			
M506L	A531L [Paramecium bursaria Chlorella virus 1]	0.000165	NP_048887	53%
M507L	A532L [Paramecium bursaria Chlorella virus 1]	4.93E-19	NP_048888	61%
M508R	A533R [Paramecium bursaria Chlorella virus 1]	1.51E-118	NP_048889	59%
	unknown protein	5.74E-118	AAA66400	58%
	ORF1 protein - Chlorella virus PBCV-1	7.50E-118	A46352	58%
m509L	No Hit Found			
m510L	No Hit Found			
M511L	A535L [Paramecium bursaria Chlorella virus 1]	2.88E-09	NP_048891	47%
M512L	A536L [Paramecium bursaria Chlorella virus 1]	2.92E-09	NP_048892	41%
m513R	No Hit Found			
m514L	A537L [Paramecium bursaria Chlorella virus 1]	2.50E-22	NP_048893	32%
m515R	No Hit Found			
m516L	No Hit Found			
m517R	A539R [Paramecium bursaria Chlorella virus 1]	8.25E-44	NP_048895	52%
	similar to bacteriophage T4 intron-associated endonuclease, corresponds to Swiss-Prot Accession Number P13299 [Paramecium bursaria Chlorella virus 1]	1.99E-05	NP_048482	53%
M518L	A540L [Paramecium bursaria Chlorella virus 1]	1.56E-58	NP_048896	29%
	unknown protein	3.63E-31	AAA66404	47%
	possible T4-like proximal tail fiber [Cyanophage P-SSM2]	9.92E-13	YP_214529	21%
m519L	No Hit Found			
m520R	No Hit Found			
m521L	No Hit Found			
M522L	contains Pro-rich Px motif, PAPK (19X); similar to Arabidopsis anter-specific Pro-rich protein, corresponds to Swiss-Prot Accession Number P40602 [Paramecium bursaria Chlorella virus 1]	2.92E-07	NP_048383	26%
m523L	No Hit Found			
m524R	No Hit Found			
M525L	A9R [Paramecium bursaria Chlorella virus 1]	8.01E-56	NP_048357	63%

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	7.06E-28	NP_049005	32%
	similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	8.08E-24	NP_048807	32%
M526L	A540L [Paramecium bursaria Chlorella virus 1]	7.66E-20	NP_048896	27%
	unknown protein	4.78E-30	AAA66404	39%
	Hep_Hag [Chlorobium phaeobacteroides BS1]	1.50E-07	ZP_00533161	29%
m527R	a541R [Paramecium bursaria Chlorella virus 1]	2.60E-05	NP_048897	40%
m528R	No Hit Found			
m529R	No Hit Found			
M530L	similar to Vibrio fischeri dCMP deaminase, corresponds to Swiss-Prot Accession Number P33968 [Paramecium bursaria Chlorella virus 1]	2.57E-52	NP_048952	66%
	deoxycytidylate deaminase [Bacteriophage phi JL001]	3.74E-27	YP_223954	51%
	COG2131: Deoxycytidylate deaminase [Magnetospirillum magnetotacticum MS-1]	1.47E-23	ZP_00052863	45%
m531L	A154L [Paramecium bursaria Chlorella virus 1]	4.88E-11	NP_048502	50%
	similar to Chlorella virus PBCV-1 ORF A154L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	3.93E-69	NP_048920	44%
	similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903 [Paramecium bursaria Chlorella virus 1]>gi 1131473 gb AAC96497.1 similar to E. coli ribonucleoside-triphosphate reductase, corresponds to Swiss-Prot Accession Number P28903 [Paramecium bursaria Chlorella virus 1]>gi 7461272 pir T17619 hypothetical protein A129R - Chlorella virus PBCV-1	5.67E-68	NP_048477	40%
m532R	No Hit Found			
m533R	No Hit Found			
m534R	No Hit Found			
M535L	putative calcium-transporting ATPase 8, plasma membrane-type [Oryza sativa (japonica cultivar-group)]	6.09E-150	XP_483341	40%
	hypothetical protein DDB0188438 [Dictyostelium discoideum]	3.02E-149	XP_636219	40%
	Ca2+-transporting ATPase-like protein [Arabidopsis thaliana]	1.27E-147	CAB79748	39%
m536R	No Hit Found			
m537R	hypothetical protein CaO19.6069 [Candida albicans SC5314]	8.69E-06	EAL00130	51%
	unknown [Schistosoma japonicum]	0.000365	AAX24226	48%
m538R	No Hit Found			
m539R	No Hit Found			
m540L	No Hit Found			
m541R	No Hit Found			
m542R	No Hit Found			
M543R	similar to PBCV-1 serine/threonine protein kinase, corresponds to GenBank Accession Number U14660 [Paramecium bursaria Chlorella virus 1]	2.25E-30	NP_048643	34%
	unnamed protein product [Candida glabrata CBS138]	1.24E-20	CAG59101	31%

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
	unnamed protein product [Yarrowia lipolytica CLIB99]	3.38E-18	CAG80377	29%
m544L	No Hit Found			
m545L	No Hit Found			
M546R	topoisomerase II [Chlorella virus Marburg 1] PBCV-1 DNA topoisomerase II [Paramecium bursaria Chlorella virus 1] DNA TOPOISOMERASE II [Encephalitozoon cuniculi GB-M1]	0 0 0	AAU95770 NP_048939 CAD25222	91% 64% 47%
m547R	No Hit Found			
m548L	No Hit Found			
m549L	a586R [Paramecium bursaria Chlorella virus 1]	1.29E-12	NP_048942	60%
m550L	No Hit Found			
m551L	contains prenyl group binding site (CAAX) box [Paramecium bursaria Chlorella virus 1]	3.91E-14	NP_048940	58%
m552L	No Hit Found			
M553R	A577L [Paramecium bursaria Chlorella virus 1]	3.54E-23	NP_048933	65%
m554L	No Hit Found			
m555L	No Hit Found			
M556R	A375R [Paramecium bursaria Chlorella virus 1]	4.47E-51	NP_048732	58%
M557L	Pro-, Lys-rich, PAPK (30x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472 [Paramecium bursaria Chlorella virus 1] surface protein [Chlorella virus] Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472 [Paramecium bursaria Chlorella virus 1]	1.18E-49 8.86E-05 0.000116	NP_048762 BAE02830 NP_048741	39% 35% 30%
m558R	Hypothetical protein CBG24243 [Caenorhabditis briggsae] PREDICTED: similar to Coagulation factor VII (Serum prothrombin conversion accelerator) [Pan troglodytes] unknown [Schistosoma japonicum]	5.95E-08 0.000681 3.86E-07	CAE56520 XP_509745 AAX27666	53% 20% 37%
m559R	No Hit Found			
M560L	Glu-, Asn-rich [Paramecium bursaria Chlorella virus 1]	6.51E-14	NP_048761	64%
m561R	No Hit Found			
m562R	No Hit Found			
M563L	A403R [Paramecium bursaria Chlorella virus 1]	5.34E-19	NP_048760	52%
M564L	A402R [Paramecium bursaria Chlorella virus 1]	9.74E-60	NP_048759	48%
m565R	No Hit Found			
m566R	No Hit Found			
M567L	A401R [Paramecium bursaria Chlorella virus 1] hypothetical protein bll5851 [Bradyrhizobium japonicum USDA 110] hypothetical protein GOX0246 [Gluconobacter oxydans 621H]	2.54E-91 4.57E-32 1.42E-25	NP_048758 NP_772491 YP_190685	61% 35% 33%
m568R	No Hit Found			
m569L	similar to PBCV-1 ORF A214, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	9.43E-16	NP_048757	37%
M570L	A399R [Paramecium bursaria Chlorella virus 1] hypothetical protein AN5188.2 [Aspergillus nidulans FGSC A4] unnamed protein product [Drosophila miranda]	8.41E-35 2.16E-06 3.68E-06	NP_048756 EAA62369 CAA41925	54% 39% 34%
m571R	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
M572R	A398L [Paramecium bursaria Chlorella virus 1]	7.67E-26	NP_048755	49%
m573R	No Hit Found			
M574L	hypothetical protein [Chlorella virus]	2.36E-11	AAU06284	32%
m575R	No Hit Found			
m576R	No Hit Found			
m577L	A395R [Paramecium bursaria Chlorella virus 1]	1.05E-22	NP_048752	66%
M578R	Lys-, Arg-rich; contains eukaryotic putative RNA-binding region RNP-1 signature; similar to PBCV-1 ORF A267L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	5.48E-40	NP_048834	43%
	Lys-, Glu-rich [Paramecium bursaria Chlorella virus 1]	0.000127	NP_048846	29%
	unknown [Acanthamoeba polyphaga mimivirus]	1.51E-05	YP_142777	23%
m579L	No Hit Found			
M580L	A394R [Paramecium bursaria Chlorella virus 1]	6.20E-18	NP_048751	44%
m581R	No Hit Found			
m582R	No Hit Found			
m583L	A394R [Paramecium bursaria Chlorella virus 1]	2.45E-08	NP_048751	50%
M584R	No Hit Found			
m585R	No Hit Found			
M586L	contains ATP/GTP-binding site motif A; similar to frog virus 3 ATPase, corresponds to GenBank Accession Number M80551 [Paramecium bursaria Chlorella virus 1]	6.13E-112	NP_048749	78%
	AGB-1 [Heterosigma akashiwo virus]	1.22E-35	BAB69884	37%
	FirrV-1-A12 [Feldmannia irregularis virus a]	2.30E-34	AAR26836	41%
m587R	No Hit Found			
m588R	No Hit Found			
m589L	No Hit Found			
m590L	No Hit Found			
M591R	Lys-, Pro-rich, PAPK (10x); similar to wheat Pro-, Lys-rich protein, corresponds to GenBank Accession Number X52472 [Paramecium bursaria Chlorella virus 1]	8.52E-17	NP_048741	31%
	A565R [Paramecium bursaria Chlorella virus 1]	1.78E-06	NP_048921	27%
	DNA binding protein [Chlorella virus]	8.00E-07	BAA11343	28%
m592L	PREDICTED: similar to PDZ domain containing 3 isoform a [Canis familiaris]	8.92E-06	XP_536513	58%
	unknown [Arabidopsis thaliana]	3.39E-05	AAN60242	53%
	At2g10940/F15K19.1 [Arabidopsis thaliana]	3.39E-05	AAN18126	53%
m593L	No Hit Found			
m594L	a389R [Paramecium bursaria Chlorella virus 1]	1.02E-07	NP_048746	53%
m595R	No Hit Found			
m596L	No Hit Found			
m597R	Hypothetical protein CBG21394 [Caenorhabditis briggsae]	0.000284	CAE73829	28%
	repetin [Mus musculus]	0.000371	NP_033126	72%
	Hypothetical protein CBG07098 [Caenorhabditis briggsae]	1.51E-05	CAE62907	26%
m598L	No Hit Found			
M599R	No Hit Found			
m600L	No Hit Found			
M601L	similar to tomato histidine decarboxylase, corresponds to Swiss-Prot Accession Number P54772 [Paramecium bursaria Chlorella virus 1]	1.93E-129	NP_048954	62%

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
	histidine decarboxylase (EC 4.1.1.22) - Klebsiella planticola >gi 27151767 sp P28578 DCHS_KLEPL Histidine decarboxylase (HDC)	1.20E-54	B40004	37%
	histidine decarboxylase	1.20E-54	AAA25071	37%
m602R	No Hit Found			
m603R	No Hit Found			
m604R	No Hit Found			
M605L	A602L [Paramecium bursaria Chlorella virus 1]	7.99E-07	NP_048958	47%
M606R	A603R [Paramecium bursaria Chlorella virus 1]	1.86E-19	NP_048959	43%
M607L	No Hit Found			
m608R	No Hit Found			
m609L	No Hit Found			
M610L	A492L [Paramecium bursaria Chlorella virus 1]	3.50E-24	NP_048848	38%
	A493L [Paramecium bursaria Chlorella virus 1]	4.59E-16	NP_048849	42%
m611L	No Hit Found			
M612R	A494R [Paramecium bursaria Chlorella virus 1]	4.01E-111	NP_048850	56%
	PBCV1-A494R-like protein [Acanthamoeba polyphaga mimivirus]	6.11E-35	YP_142783	34%
	EsV-1-101 [Ectocarpus siliculosus virus]	1.46E-28	AAK14519	38%
m613L	No Hit Found			
m614L	No Hit Found			
M615R	No Hit Found			
M616L	A501L [Paramecium bursaria Chlorella virus 1]	8.66E-09	NP_048857	48%
	Pro-, Ser-rich [Paramecium bursaria Chlorella virus 1]	7.33E-08	NP_048856	63%
m617R	No Hit Found			
M618L	A491R [Paramecium bursaria Chlorella virus 1]	2.27E-09	NP_048847	38%
m619L	No Hit Found			
M620L	Pro-,Gln-rich [Paramecium bursaria Chlorella virus 1]	1.25E-97	NP_048844	61%
	unknown [Acanthamoeba polyphaga mimivirus]	2.68E-07	YP_142771	27%
m621R	No Hit Found			
M622L	similar to Bacteriophage SP01 gene 31 intron, corresponds to Swiss-Prot Accession Number P34081 [Paramecium bursaria Chlorella virus 1]	0.000185	NP_048779	31%
	putative endonuclease [Bacteriophage phi AT3]	1.37E-15	YP_025062	38%
	HNH endonuclease [Acanthamoeba polyphaga mimivirus]	0.000109	YP_142599	33%
m623R	No Hit Found			
m624L	No Hit Found			
M625R	No Hit Found			
m626L	A485R [Paramecium bursaria Chlorella virus 1]	1.80E-43	NP_048841	57%
M627L	pyrimidine dimer-specific glycosylase [Chlorella virus]	1.21E-39	AAD33374	59%
	PBCV-1 pyrimidine dimer-specific glycosylase [Paramecium bursaria Chlorella virus 1]	1.58E-39	NP_048398	59%
	pyrimidine dimer-specific glycosylase [Chlorella virus]	2.06E-39	AAD33369	58%
M628L	Gly-rich [Paramecium bursaria Chlorella virus 1]	2.62E-18	NP_048836	59%
M629L	A481L [Paramecium bursaria Chlorella virus 1]	1.20E-52	NP_048837	44%
	P0551A11.21 [Oryza sativa (japonica cultivar-group)]	0.000462	NP_908582	48%
	ring-H2 zinc finger protein-like [Oryza sativa (japonica cultivar-group)]	0.000462	BAD45600	48%
m630R	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
M631L	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563 [Paramecium bursaria Chlorella virus 1]	8.25E-43	NP_048441	43%
	a91L [Paramecium bursaria Chlorella virus 1]	9.16E-26	NP_048439	50%
	hypothetical protein A93L - Chlorella virus PBCV-1	9.48E-23	T17583	45%
m632R	No Hit Found			
m633R	a281R [Paramecium bursaria Chlorella virus 1]	9.64E-05	NP_048635	57%
M634R	A452L [Paramecium bursaria Chlorella virus 1]	7.41E-05	NP_048809	37%
M635R	A482R [Paramecium bursaria Chlorella virus 1]	3.65E-50	NP_048838	44%
	FirrV-1-I5 [Feldmannia irregularis virus a]	1.83E-09	AAR26970	39%
	FirrV-1-B14 [Feldmannia irregularis virus a]	5.33E-09	AAR26889	38%
m636L	No Hit Found			
M637L	A484L [Paramecium bursaria Chlorella virus 1]	7.44E-45	NP_048840	54%
M638R	PBCV-1 surface protein [Paramecium bursaria Chlorella virus 1]	1.19E-07	NP_048488	29%
	surface protein [Paramecium bursaria Chlorella virus 1]	1.19E-07	BAD22850	29%
	surface protein [Chlorella virus CVK2]	2.66E-07	BAD12236	28%
m639R	No Hit Found			
m640L	No Hit Found			
m641L	No Hit Found			
m642L	No Hit Found			
m643L	No Hit Found			
m644L	No Hit Found			
m645R	No Hit Found			
m646L	No Hit Found			
m647L	No Hit Found			
m648L	a281R [Paramecium bursaria Chlorella virus 1]	3.51E-06	NP_048635	55%
M649R	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563 [Paramecium bursaria Chlorella virus 1]	1.09E-32	NP_048441	32%
	a91L [Paramecium bursaria Chlorella virus 1]	1.80E-27	NP_048439	52%
	hypothetical protein A93L - Chlorella virus PBCV-1	1.53E-10	T17583	29%
m650L	No Hit Found			
M651R	A9R [Paramecium bursaria Chlorella virus 1]	9.21E-65	NP_048357	67%
	similar to PBCV-1 ORF A79R, corresponds to GenBank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	1.10E-17	NP_048629	30%
	A79R [Paramecium bursaria Chlorella virus 1]	1.22E-16	NP_048427	28%
m652L	No Hit Found			
M653L	contains ribonucleotide reductase (RR) signature; similar to tobacco RR small subunit, corresponds to Swiss-Prot Accession Number P49730 [Paramecium bursaria Chlorella virus 1]	2.25E-142	NP_048832	75%
	putative ribonucleotide reductase small subunit [Arabidopsis thaliana]	9.54E-101	AAM51287	56%
	ribonucleotide reductase small subunit [Arabidopsis thaliana]	6.19E-100	AAO62422	55%
m654R	unnamed protein product [Kluyveromyces lactis NRRL Y-1140]	2.45E-19	CAG98466	45%
	unknown [Saccharomyces cerevisiae]	1.39E-06	AAL79317	41%
	CPXV051A protein [Cowpox virus]	2.01E-05	AAP41052	53%
m655L	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
m656R	No Hit Found			
M657L	A471R [Paramecium bursaria Chlorella virus 1]	6.01E-53	NP_048827	52%
	unknown [Acanthamoeba polyphaga mimivirus]	2.46E-22	YP_142861	40%
	FirrV-1-A5 [Feldmannia irregularis virus a]	1.18E-16	AAR26829	31%
m658R	No Hit Found			
m659L	No Hit Found			
M660L	A470R [Paramecium bursaria Chlorella virus 1]	9.55E-60	NP_048826	60%
	UKCH-2 [Heterosigma akashiwo virus]	6.92E-18	BAB69883	40%
m661R	No Hit Found			
m662R	No Hit Found			
m663L	No Hit Found			
M664L	A468R [Paramecium bursaria Chlorella virus 1]	2.34E-119	NP_048824	57%
	FirrV-1-A46 [Feldmannia irregularis virus a]	0.000557	AAR26870	33%
	EsV-1-45 [Ectocarpus siliculosus virus]	2.26E-05	AAK14471	23%
m665R	No Hit Found			
m666R	No Hit Found			
M667R	A467L [Paramecium bursaria Chlorella virus 1]	1.34E-109	NP_048823	62%
	Hypothetical protein [Amsacta moorei entomopoxvirus]	1.02E-08	NP_065022	33%
m668L	No Hit Found			
m669L	No Hit Found			
M670R	PBCV-1 thiol oxidoreductase [Paramecium bursaria Chlorella virus 1]	6.29E-44	NP_048821	67%
	putative thiol oxidoreductase [Acanthamoeba polyphaga mimivirus]	1.20E-10	YP_142722	36%
	unnamed protein product [Yarrowia lipolytica CLIB99]	5.03E-09	CAG81498	33%
m671L	a466L [Paramecium bursaria Chlorella virus 1]	1.18E-10	NP_048822	52%
M672L	similar to Bacillus ribonuclease III, corresponds to Swiss-Prot Accession Number P51833 [Paramecium bursaria Chlorella virus 1]	9.53E-95	NP_048820	66%
	Ribonuclease III [Pelodictyon phaeoclathratiforme BU-1]	2.02E-28	ZP_00590199	37%
	Ribonuclease III [Chlorobium phaeobacteroides BS1]	3.45E-28	ZP_00532592	34%
m673R	No Hit Found			
M674R	contains ATP/GTP-binding site motif A [Paramecium bursaria Chlorella virus 1]	0	NP_048813	65%
	EsV-1-109 [Ectocarpus siliculosus virus]	5.61E-73	AAK14527	36%
	FirrV-1-B27 [Feldmannia irregularis virus a]	7.58E-70	AAR26902	30%
m675L	a460R [Paramecium bursaria Chlorella virus 1]	8.03E-07	NP_048817	31%
m676L	No Hit Found			
m677L	a459R [Paramecium bursaria Chlorella virus 1]	6.88E-22	NP_048816	72%
	a457R [Paramecium bursaria Chlorella virus 1]>gi 1620129gb AAC96825.1 a457R [Paramecium bursaria Chlorella virus 1] >gi 7461597 pir T17960 hypothetical protein a457R - Chlorella virus PBCV-1	7.63E-13	NP_048814	60%
m678R	No Hit Found			
m679R	No Hit Found			
M680R	A454L [Paramecium bursaria Chlorella virus 1]	2.02E-96	NP_048811	64%
m681L	No Hit Found			
m682L	No Hit Found			
M683L	similar to PBCV-1 ORF A315L, corresponds to GenBank Accession Number M74440 [Paramecium bursaria Chlorella virus 1]	2.31E-35	NP_048851	41%

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
	similar to Chlorella virus PBCV-1 ORF A315L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	7.98E-20	NP_049007	35%
	A315L [Paramecium bursaria Chlorella virus 1]	5.17E-19	NP_048671	36%
m684R	No Hit Found			
m685R	No Hit Found			
M686L	contains Pro-rich Px motif EPSPEPxP (5X), and PEST sequence; similar to trypanosome procyclin precursor, corresponds to Swiss-Prot Accession Number P08469 [Paramecium bursaria Chlorella virus 1]	2.78E-05	NP_048415	25%
m687L	No Hit Found			
m688R	No Hit Found			
m689L	No Hit Found			
M690R	No Hit Found			
m691L	No Hit Found			
m692L	A568L [Paramecium bursaria Chlorella virus 1]	1.96E-12	NP_048924	27%
M693R	No Hit Found			
M694L	A570L [Paramecium bursaria Chlorella virus 1]	1.08E-43	NP_048926	66%
M695R	A572R [Paramecium bursaria Chlorella virus 1]	1.34E-63	NP_048928	65%
m696L	No Hit Found			
M697L	similar to Periwinkle PCNA, corresponds to GenBank Accession Number X55052 [Paramecium bursaria Chlorella virus 1]	4.05E-59	NP_048930	45%
	Hypothetical protein CBG13430 [Caenorhabditis briggsae]	5.34E-27	CAE67843	31%
	proliferating Cell Nuclear Antigen homolog (29.0 kD) (pcn-1) [Caenorhabditis elegans]	6.97E-27	NP_500466	30%
m698R	No Hit Found			
M699L	No Hit Found			
M700L	A575L [Paramecium bursaria Chlorella virus 1]	4.17E-24	NP_048931	36%
M701R	PLPRNLLL (4X), SPPPSKP (3X) [Paramecium bursaria Chlorella virus 1]	1.59E-93	NP_048689	70%
	a332L [Paramecium bursaria Chlorella virus 1]	3.51E-16	NP_048688	74%
	ENSANGP00000018413 [Anopheles gambiae str. PEST]	4.45E-11	XP_321236	29%
m702R	No Hit Found			
m703L	No Hit Found			
M704R	contains Gln-rich, neutral zinc metallopeptidase, zinc binding region signature [Paramecium bursaria Chlorella virus 1]	1.35E-50	NP_048877	60%
m705L	No Hit Found			
m706L	a507R [Paramecium bursaria Chlorella virus 1]	1.02E-14	NP_048863	35%
m707R	A505L [Paramecium bursaria Chlorella virus 1]	7.13E-127	NP_048861	58%
m708L	No Hit Found			
m709L	No Hit Found			
m710R	No Hit Found			
m711L	a506R [Paramecium bursaria Chlorella virus 1]	5.50E-16	NP_048862	62%
m712R	No Hit Found			
m713R	contains prokaryotic membrane lipoprotein lipid attachment site [Paramecium bursaria Chlorella virus 1]	4.58E-45	NP_048859	40%
M714L	No Hit Found			
m715R	A502L [Paramecium bursaria Chlorella virus 1]	1.15E-21	NP_048858	58%
m716L	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
M717R	contains 3 ankyrin repeat-like elements; similar to Drosophila ankyrin, corresponds to GenBank Accession Number L35601 [Paramecium bursaria Chlorella virus 1]	1.42E-48	NP_048963	46%
	A608R [Paramecium bursaria Chlorella virus 1]	1.18E-18	NP_048964	34%
	ankyrin repeat protein E4_2 [synthetic construct]	2.08E-07	AAO25691	32%
m718R	No Hit Found			
M719L	PBCV-1 UDP-glucose dehydrogenase [Paramecium bursaria Chlorella virus 1]	7.43E-130	NP_048965	59%
	unknown [Pasteurella multocida subsp. multocida str. Pm70]	7.22E-117	AAK02860	54%
	UDP-glucose dehydrogenase [Pasteurella multocida]	1.61E-116	AAC67251	54%
m720R	hypothetical protein a225L - Chlorella virus PBCV-1	4.44E-05	T17716	39%
	a225L [Paramecium bursaria Chlorella virus 1]	4.44E-05	NP_048573	39%
M721L	similar to cellulose synthase catalytic subunit (UDP-forming) [Brucella melitensis]	0	NP_048569	69%
	hypothetical protein A222R - Chlorella virus PBCV-1	5.17E-150	T17712	76%
	cellulose synthase [Agrobacterium tumefaciens str. C58]	3.90E-81	NP_533811	42%
m722L	No Hit Found			
m723R	a224L [Paramecium bursaria Chlorella virus 1]	2.78E-15	NP_048572	56%
m724R	hypothetical protein a223aL - Chlorella virus PBCV-1	8.62E-09	T17714	43%
	a223aL [Paramecium bursaria Chlorella virus 1]	3.27E-08	NP_048571	41%
m725R	hypothetical protein a221L - Chlorella virus PBCV-1	1.51E-05	T17711	37%
m726R	No Hit Found			
M727L	PBCV-1 histone H3-Lys 27 methyltransferase (vSET) [Paramecium bursaria Chlorella virus 1]	6.45E-33	NP_048968	57%
	Nuclear protein SET [Pelodictyon phaeoclathratiforme BU-1]	3.44E-10	ZP_00588496	36%
	hypothetical protein PFL0690c [Plasmodium falciparum 3D7]	1.89E-08	NP_701503	42%
m728R	No Hit Found			
M729L	RPQT-like (9x) [Paramecium bursaria Chlorella virus 1]	3.22E-133	NP_048970	45%
	similar to bovine cyligin I, corresponds to Swiss-Prot Accession Number P35662 [Paramecium bursaria Chlorella virus 1]	1.20E-79	NP_048632	36%
	similar to PBCV-1 ORF A34R, corresponds to GenBank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	1.47E-77	NP_048636	38%
m730R	No Hit Found			
m731R	No Hit Found			
m732L	No Hit Found			
M733L	similar to Variola virus orf E10L, corresponds to Swiss-Prot Accession Number P33801 [Paramecium bursaria Chlorella virus 1]	4.68E-81	NP_048973	49%
	S/T protein kinase, similar to Paramecium bursaria chlorella virus 1 A617R [Acanthamoeba polyphaga mimivirus]	5.28E-08	YP_142754	28%
	Hypothetical protein Y73B6A.1 [Caenorhabditis elegans]	4.18E-05	AAF36040	25%
m734R	No Hit Found			
M735L	A618L [Paramecium bursaria Chlorella virus 1]	1.34E-14	NP_048974	55%
m736L	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
m737R	No Hit Found			
M738L	No Hit Found			
m739L	No Hit Found			
m740L	similar to Synechocystis orf 90, corresponds to GenBank Accession Number D90902 [Paramecium bursaria Chlorella virus 1]	1.57E-10	NP_048976	34%
	A635R [Paramecium bursaria Chlorella virus 1]	6.81E-06	NP_048991	30%
M741L	A621L [Paramecium bursaria Chlorella virus 1]	1.50E-29	NP_048977	52%
M742R	Chlorella virus CVK2 translation elongation factor-3 homolog, refer to GenBank Accession Number D16505 [Paramecium bursaria Chlorella virus 1]	0	NP_049022	65%
	translation elongation factor EF-3 homolog - Chlorella virus CVK2 >gi 454301 dbj BAA03956.1 translation elongation factor-3 [Chlorella virus]	0	A48779	62%
	unnamed protein product [Candida glabrata CBS138]	0	CAG58023	48%
m743L	No Hit Found			
m744L	No Hit Found			
m745L	No Hit Found			
m746L	No Hit Found			
m747L	No Hit Found			
M748L	putative capsid protein [Chlorella virus CVG-1]	0	AAC27493	97%
	similar to Simulium iridescent virus capsid protein, corresponds to Swiss-Prot Accession Number P22166 [Paramecium bursaria Chlorella virus 1]	0	NP_048978	72%
	hypothetical major capsid protein [Heterosigma akashiwo virus 01]	1.63E-24	BAE06835	67%
m749R	No Hit Found			
m750R	No Hit Found			
m751R	No Hit Found			
M752R	A624R [Paramecium bursaria Chlorella virus 1]	8.72E-30	NP_048980	52%
	conserved hypothetical protein [Psychrobacter cryohalolentis K5]	1.50E-05	EA011866	33%
	conserved hypothetical protein [Listeria monocytogenes str. 1/2a F6854]	1.50E-05	ZP_00234461	26%
M753L	a91L [Paramecium bursaria Chlorella virus 1]	3.40E-09	NP_048439	34%
	similar to Chlamydia histone-like protein, corresponds to GenBank Accession Number D71563 [Paramecium bursaria Chlorella virus 1]	0.00043	NP_048441	29%
m754R	No Hit Found			
m755L	No Hit Found			
M756R	A659L [Paramecium bursaria Chlorella virus 1]	2.24E-14	NP_049015	39%
M757R	A656L [Paramecium bursaria Chlorella virus 1]	5.43E-10	NP_049012	42%
M758R	A654L [Paramecium bursaria Chlorella virus 1]	5.01E-53	NP_049010	51%
	COG0454: Histone acetyltransferase HPA2 and related acetyltransferases [Trichodesmium erythraeum IMS101]	7.57E-09	ZP_00324949	26%
	PREDICTED: similar to arylalkylamine N-acetyltransferase [Apis mellifera]	5.43E-07	XP_392876	30%
m759L	No Hit Found			
m760L	No Hit Found			
m761R	No Hit Found			
M762L	A644R [Paramecium bursaria Chlorella virus 1]	1.86E-25	NP_049000	72%
	unknown [Acanthamoeba polyphaga mimivirus]	0.000181	YP_142763	28%

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
M763L	Gln-rich; KQQ (6X) [Paramecium bursaria Chlorella virus 1]	3.84E-57	NP_048999	47%
	A642R [Paramecium bursaria Chlorella virus 1]	1.25E-07	NP_048998	68%
	proline rich protein 2 (predicted) [Rattus norvegicus]	0.00013	NP_001013229	38%
m764R	No Hit Found			
m765R	No Hit Found			
M766L	PBCV-1 Agmatine iminohydrolase [Paramecium bursaria Chlorella virus 1]	3.95E-138	NP_048994	63%
	hypothetical protein A638R - Chlorella virus PBCV-1	2.03E-126	T18140	62%
	hypothetical protein lmo0038 [Listeria monocytogenes EGD-e]	3.37E-105	NP_463571	52%
m767R	No Hit Found			
m768L	No Hit Found			
m769R	No Hit Found			
M770R	A634L [Paramecium bursaria Chlorella virus 1]	1.44E-40	NP_048990	58%
m771R	No Hit Found			
m772L	No Hit Found			
m773L	No Hit Found			
m774R	No Hit Found			
m775L	A633R [Paramecium bursaria Chlorella virus 1]	8.73E-38	NP_048989	61%
m776L	No Hit Found			
M777L	similar to Schizosaccharomyces ribonucleotide reductase M1 chain, corresponds to Swiss-Prot Accession Number P36602 [Paramecium bursaria Chlorella virus 1]	0	NP_048985	78%
	hypothetical protein [Gallus gallus]	0	CAG31174	51%
	RRM1 protein [Xenopus laevis]	0	AAH46846	51%
m778R	No Hit Found			
m779R	No Hit Found			
m780R	No Hit Found			
m781R	a631L [Paramecium bursaria Chlorella virus 1]	8.87E-06	NP_048987	41%
m782R	No Hit Found			
m783R	No Hit Found			
m784R	No Hit Found			
M785L	similar to Chlorella virus PBCV-1 ORF A231L, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.20E-131	NP_048983	54%
	contains ATP/GTP-binding motif A [Paramecium bursaria Chlorella virus 1]	8.19E-80	NP_048579	45%
m786L	No Hit Found			
m787R	No Hit Found			
m788L	No Hit Found			
M789L	similar to Streptococcus pyogenes fibronectin protein, corresponds to GenBank Accession Number L28919 [Paramecium bursaria Chlorella virus 1]	7.64E-26	NP_048528	52%
	Protein of unknown function DUF814:Fibronectin-binding A, N-terminal [Clostridium thermocellum ATCC 27405]	1.71E-09	ZP_00504390	35%
	fibronectin-binding protein, putative [Thermotoga maritima MSB8]	1.45E-08	NP_228174	32%
m790R	No Hit Found			
M791R	vChti-1 [Chlorella virus]	1.24E-40	BAA78554	36%
	PBCV-1 chitinase [Paramecium bursaria Chlorella virus 1]	3.27E-41	NP_048529	37%

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
	Cellulose-binding, bacterial type [Frankia sp. EAN1pec]	2.49E-57	ZP_00570566	42%
m792L	No Hit Found			
m793L	No Hit Found			
M794R	protein kinase A248R [Chlorella virus] protein kinase A248R [Chlorella virus] protein kinase A248R [Chlorella virus]	2.88E-62 2.88E-62 4.15E-61	AAU06280 AAU06275 AAU06282	48% 48% 47%
m795L	No Hit Found			
m796L	No Hit Found			
M797R	A378L [Paramecium bursaria Chlorella virus 1] QI74 protein [Hypocrea lixii]	8.37E-30 0.000159	NP_048735 CAA64974	56% 22%
m798L	No Hit Found			
m799L	No Hit Found			
m800L	No Hit Found			
M801R	A9R [Paramecium bursaria Chlorella virus 1] similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1] similar to PBCV-1 ORF A275R, encoded by GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	7.36E-62 3.81E-26 1.23E-24	NP_048357 NP_049005 NP_048807	63% 32% 30%
m802L	No Hit Found			
M803L	contains Pro-rich Px motif, PAPK (8X); similar to Thermoproteus virus protein TPX, corresponds to Swiss-Prot Accession Number P19275 [Paramecium bursaria Chlorella virus 1] similar to PBCV-1 ORF A41R, corresponds to GenBank Accession Number U17055 [Paramecium bursaria Chlorella virus 1] P0481E12.18 [Oryza sativa (japonica cultivar-group)]	4.12E-118 6.81E-113 1.41E-09	NP_048389 NP_048519 NP_916095	65% 63% 24%
m804R	No Hit Found			
m805L	No Hit Found			
m806R	No Hit Found			
M807R	contains ATP/GTP-binding motif A; similar to Dictyostelium FP21 glycoprotein, corresponds to Swiss-Prot Accession Number P52285 [Paramecium bursaria Chlorella virus 1] kinetochore (SKP1p)-like protein [Arabidopsis thaliana] Skp1p-like protein [Arabidopsis thaliana]	3.59E-23 3.59E-23 1.36E-22	NP_048387 CAB80138 CAB80164	36% 40% 38%
m808L	No Hit Found			
m809L	No Hit Found			
M810R	No Hit Found			
M811L	No Hit Found			
m812R	No Hit Found			
M813R	No Hit Found			
M814L	No Hit Found			
m815R	No Hit Found			
m816R	No Hit Found			
m817R	No Hit Found			
M818R	No Hit Found			
M819R	No Hit Found			
M820R	No Hit Found			
m821L	No Hit Found			

ORF	PROTEIN HITS	E-VALUE	ACCESSION NUMBER	% ID
M822R	A318R [Paramecium bursaria Chlorella virus 1]	1.38E-62	NP_048674	58%
	PAPK (17X); similar to PBCV-1 ORF A41R, corresponds to Genbank Accession Number U17055 [Paramecium bursaria Chlorella virus 1]	2.02E-13	NP_048672	62%
	COG2755: Lysophospholipase L1 and related esterases [Rubrivivax gelatinosus PM1]	0.000293	ZP_00244751	31%
m823L	No Hit Found			
m824L	No Hit Found			
m825L	a319L [Paramecium bursaria Chlorella virus 1]	1.02E-09	NP_048675	54%
m826L	No Hit Found			
m827L	No Hit Found			
m828R	A88R [Paramecium bursaria Chlorella virus 1]	4.28E-12	NP_048436	35%
m829L	hypothetical protein AN8019.2 [Aspergillus nidulans FGSC A4]	3.11E-12	EAA59641	38%
	Hypothetical protein CBG09456 [Caenorhabditis briggsae]	4.65E-08	CAE64680	34%
	PREDICTED: similar to CG7462-PB, isoform B [Apis mellifera]	5.15E-07	XP_392578	32%
m830R	No Hit Found			
m831R	No Hit Found			
m832R	No Hit Found			
m833L	contains 10 ankyrin-like repeats; similar to human ankyrin, corresponds to Swiss-Prot Accession Number P16157 [Paramecium bursaria Chlorella virus 1]	5.24E-16	NP_049038	36%
	contains 4 ankyrin repeats; similar to reticulocyte ankyrin, corresponds to Swiss-Prot Accession Number P16157 [Paramecium bursaria Chlorella virus 1]	1.98E-39	NP_048353	37%
	ankyrin repeat protein [Aspergillus fumigatus Af293]	2.06E-12	EAL84954	37%
m834L	No Hit Found			
m835R	No Hit Found			
m836R	No Hit Found			
m837L	No Hit Found			
m838L	No Hit Found			
m839R	No Hit Found			
m840L	No Hit Found			
m841R	No Hit Found			
M842R	RepA / Rep+ protein KID [Moorella thermoacetica ATCC 39073]	1.21E-06	ZP_00575949	25%
	hypothetical malaria antigen [Plasmodium falciparum 3D7]	1.74E-05	NP_705165	25%
	hypothetical protein BGP110 [Borrelia garinii PBi]	6.62E-05	YP_161395	27%
m843L	No Hit Found			
M844R	A366L [Paramecium bursaria Chlorella virus 1]	5.04E-36	NP_048723	38%
	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	9.90E-08	NP_049005	30%
	similar to Chlorella virus PBCV-1 ORF A450R, corresponds to GenBank Accession Number U42580 [Paramecium bursaria Chlorella virus 1]	1.87E-06	NP_049003	41%
m845L	No Hit Found			

2.6.3.3 Results from analysis with the Protein Families (Pfam) Database

MT325 Pfam			
ORF	Pfam	E-value	Description
M10L	PF00383.12	4.5e-05	Cytidine and deoxycytidylate deaminase zinc-binding region
M14R	PF00705.8	8.9e-24	Proliferating cell nuclear antigen, N-terminal domain
	PF02747.5	6.8e-15	Proliferating cell nuclear antigen, C-terminal domain
M19L	PF00136.10	1.9e-143	DNA polymerase family B
	PF03104.8	4.4e-79	DNA polymerase family B, exonuclease domain
M25R	PF07150.1	5.7e-27	Protein of unknown function (DUF1390)
M26L	PF00389.18	3.9e-15	D-isomer specific 2-hydroxyacid dehydrogenase, catalytic domain
	PF02826.6	1.3e-60	D-isomer specific 2-hydroxyacid dehydrogenase, NAD binding domain
M30R	PF00230.8	1.5e-16	Major intrinsic protein
M32L	PF01541.13	2.1e-05	GIY-YIG catalytic domain
	PF07453.2	1e-06	NUMOD1 domain
M34L	PF02511.5	1.9e-80	Thymidylate synthase complementing protein
M37R	PF00310.10	3.5e-29	Glutamine amidotransferases class-II
	PF01380.11	8.2e-27	SIS domain
M42R	PF06598.1	8.4e-16	Chlorovirus glycoprotein repeat
M47R	PF06598.1	1.5e-14	Chlorovirus glycoprotein repeat
M55R	PF06598.1	9.7e-16	Chlorovirus glycoprotein repeat
M61R	PF06598.1	6.1e-15	Chlorovirus glycoprotein repeat
M71R	PF06598.1	4.6e-13	Chlorovirus glycoprotein repeat
M78R	PF04451.2	4.4e-93	Iridovirus major capsid protein
M85R	PF00704.16	2e-86	Glycosyl hydrolases family 18
	PF01607.12	6e-06	Chitin binding Peritrophin-A domain
M91R	PF01374.8	1.5e-98	Glycosyl hydrolase family 46
M99R	PF00080.8	1.5e-76	Copper/zinc superoxide dismutase (SODC)
M101R	PF02275.7	4.9e-25	Linear amide C-N hydrolases, choloylglycine hydrolase family
M103L	PF00795.11	2.3e-49	Carbon-nitrogen hydrolase
M111R	PF01541.13	7.4e-05	GIY-YIG catalytic domain
	PF07453.2	1.9e-07	NUMOD1 domain
M124L	PF00722.9	2.3e-06	Glycosyl hydrolases family 16
M128R	PF03142.4	0.00064	Chitin synthase
M133R	PF01331.9	2.4e-35	mRNA capping enzyme, catalytic domain
	PF03919.5	3.4e-08	mRNA capping enzyme, C-terminal domain
M137L	PF00443.17	5.1e-21	Ubiquitin carboxyl-terminal hydrolase
M139L	PF00382.9	0.0042	Transcription factor TFIIB repeat
M152R	PF06598.1	2.9e-12	Chlorovirus glycoprotein repeat
M163L	PF01096.8	5.4e-26	Transcription factor S-II (TFIIS)
	PF07500.3	8.5e-11	Transcription factor S-II (TFIIS), central domain
M167L	PF07150.1	3.6e-11	Protein of unknown function (DUF1390)
M174L	PF00483.12	2.5e-05	Nucleotidyl transferase
M174L	PF00702.13	1.1e-24	haloacid dehalogenase-like hydrolase
M180L	PF01541.13	1.8e-07	GIY-YIG catalytic domain
M183R	PF07885.3	0.00016	Ion channel
M186R	PF00535.14	1.6e-15	Glycosyl transferase family 2

MT325 Pfam			
ORF	Pfam	E-value	Description
M201R	PF04851.5	4.5e-08	Type III restriction enzyme, res subunit
M219L	PF01734.11	9.2e-41	Patatin-like phospholipase
M221L	PF00069.14	1.8e-49	Protein kinase domain
M225L	PF00270.16	4.2e-30	DEAD/DEAH box helicase
	PF00271.18	6.4e-09	Helicase conserved C-terminal domain
	PF08148.1	0.00081	DSHCT (NUC185) domain
M233L	PF06408.1	1.6e-243	Homospermidine synthase
M241L	PF00462.11	2.4e-06	Glutaredoxin
M247R	PF07150.1	2.5e-10	Protein of unknown function (DUF1390)
M264L	PF00692.9	4.1e-50	dUTPase
M266R	PF00352.11	0.024	Transcription factor TFIID (or TATA-binding protein, TBP)
M269L	PF04451.2	1.6e-145	Iridovirus major capsid protein
M272L	PF00176.12	9.8e-21	SNF2 family N-terminal domain
	PF00271.18	7.6e-15	Helicase conserved C-terminal domain
M307L	PF00278.11	2.5e-17	Pyridoxal-dependent decarboxylase, C-terminal sheet domain
	PF02784.6	4.2e-65	Pyridoxal-dependent decarboxylase, pyridoxal binding domain
M313L	PF00782.9	7.6e-42	Dual specificity phosphatase, catalytic domain
M359L	PF00145.7	2.6e-21	C-5 cytosine-specific DNA methylase
M370R	PF01541.13	0.00067	GIY-YIG catalytic domain
	PF07453.2	1.1e-07	NUMOD1 domain
M372R	PF00271.18	0.0046	Helicase conserved C-terminal domain
M381R	PF04451.2	1.2e-13	Iridovirus major capsid protein
M389R	PF01171.10	7.1e-25	PP-loop family
M397R	PF07150.1	1.2e-10	Protein of unknown function (DUF1390)
M401R	PF03171.9	2.4e-10	2OG-Fe(II) oxygenase superfamily
M403R	PF00085.8	4e-07	Thioredoxin
M404R	PF03109.6	3.6e-39	ABC1 family
M413L	PF00023.17	1.9e-13	Ankyrin repeat
M423R	PF00462.11	1.7e-09	Glutaredoxin
M425R	PF01712.8	5.1e-20	Deoxyribonucleoside kinase
M445L	PF00085.8	0.0006	Thioredoxin
M448L	PF00085.8	0.0046	Thioredoxin
M460R	PF01541.13	0.013	GIY-YIG catalytic domain
M463L	PF04451.2	7.5e-224	Iridovirus major capsid protein
M465L	PF01541.13	8.8e-05	GIY-YIG catalytic domain
	PF07453.2	0.00072	NUMOD1 domain
M467R	PF00535.14	9.3e-05	Glycosyl transferase family 2
M476L	PF05854.1	4.5e-08	Non-histone chromosomal protein MC1
M502L	PF01541.13	2.3e-07	GIY-YIG catalytic domain
	PF07453.2	9.3e-05	NUMOD1 domain
M525L	PF07150.1	2.7e-19	Protein of unknown function (DUF1390)
M530L	PF00383.12	8.5e-27	Cytidine and deoxycytidylate deaminase zinc-binding region
M535L	PF00122.8	3.7e-50	E1-E2 ATPase
	PF00689.9	3.6e-05	Cation transporting ATPase, C-terminus
	PF00690.14	1.9e-06	Cation transporter/ATPase, N-terminus

MT325 Pfam			
ORF	Pfam	E-value	Description
	PF00702.13	3.6e-16	haloacid dehalogenase-like hydrolase
M543R	PF00069.14	3.2e-41	Protein kinase domain
M546R	PF00204.12	1.4e-09	DNA gyrase B
	PF00521.9	3.2e-68	DNA gyrase/topoisomerase IV, subunit A
	PF02518.13	5.1e-05	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase
M570L	PF00075.12	1.8e-18	RNase H
M601L	PF00282.8	5.2e-15	Pyridoxal-dependent decarboxylase conserved domain
M622L	PF01844.12	4.9e-08	HNH endonuclease
M627L	PF03013.4	8.2e-71	Pyrimidine dimer DNA glycosylase
M651R	PF07150.1	1.5e-09	Protein of unknown function (DUF1390)
M653L	PF00268.11	9.5e-121	Ribonucleotide reductase, small chain
M670R	PF04777.2	4.4e-30	Erv1 / Alr family
M672L	PF00035.12	4.1e-12	Double-stranded RNA binding motif
	PF00636.13	1.1e-44	RNase3 domain
M674R	PF03288.5	1.6e-06	Poxvirus D5 protein-like
M683L	PF01541.13	0.0097	GIY-YIG catalytic domain
M697L	PF00705.8	4.4e-06	Proliferating cell nuclear antigen, N-terminal domain
	PF02747.5	3.1e-07	Proliferating cell nuclear antigen, C-terminal domain
M701R	PF01607.12	1.9e-07	Chitin binding Peritrophin-A domain
	PF03067.5	0.00021	Chitin binding domain
M717R	PF00023.17	0.012	Ankyrin repeat
M719L	PF00984.9	1.2e-27	UDP-glucose/GDP-mannose dehydrogenase family, central domain
	PF03720.4	8.8e-06	UDP-glucose/GDP-mannose dehydrogenase family, UDP binding domain
	PF03721.4	1.2e-37	UDP-glucose/GDP-mannose dehydrogenase family, NAD binding domain
M721L	PF00535.14	1.8e-05	Glycosyl transferase family 2
M727L	PF00856.16	2.6e-07	SET domain
M742R	PF00005.14	1.3e-48	ABC transporter
	PF02985.9	0.24	HEAT repeat
M748L	PF04451.2	2.7e-286	Iridovirus major capsid protein
M758R	PF00583.12	3.4e-05	Acetyltransferase (GNAT) family
M766L	PF04371.4	1.7e-184	Porphyromonas-type peptidyl-arginine deiminase
M777L	PF00317.10	4.6e-28	Ribonucleotide reductase, all-alpha domain
	PF02867.4	7.1e-253	Ribonucleotide reductase, barrel domain
	PF03477.5	2.1e-15	ATP cone domain
M789L	PF05670.2	4.7e-23	Domain of unknown function (DUF814)
M794R	PF00069.14	1.5e-62	Protein kinase domain
M801R	PF07150.1	1.2e-17	Protein of unknown function (DUF1390)
M807R	PF01466.8	2.6e-14	Skp1 family, dimerisation domain
	PF03931.4	2.8e-18	Skp1 family, tetramerisation domain
M822R	PF00657.11	0.00015	GDSDL-like Lipase/Acylhydrolase
M829L	PF00023.17	3e-12	Ankyrin repeat
M833L	PF00023.17	3.8e-09	Ankyrin repeat
M844R	PF07150.1	0.00017	Protein of unknown function (DUF1390)

2.6.3.4 Results of analysis with Clusters of Orthologous Groups (COGs)

MT325 COGs			
ORF	E-value	COG	Description
Information Storage and Processing			
Transcription			
M672L	2.E-44	COG0571	dsRNA-specific ribonuclease
M789L	1.E-16	COG1293	Predicted RNA-binding protein homologous to eukaryotic snRNP
M139L	3.E-16	COG1405	Transcription initiation factor TFIIB, Brf1 subunit/Transcription initiation factor TFIIB
M163L	1.E-10	COG1594	DNA-directed RNA polymerase, subunit M/Transcription elongation factor TFIIS
RNA processing and modification			
M133R	1.E-19	COG5226	mRNA capping enzyme, guanylyltransferase (alpha) subunit
DNA replication, recombination, and repair			
M546R	e-100	COG0187	Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), B subunit
M359L	5.E-25	COG0270	Site-specific DNA methylase
M570L	6.E-10	COG0328	Ribonuclease HI
M19L	3.E-99	COG0417	DNA polymerase elongation subunit (family B)
M697L	3.E-10	COG0592	DNA polymerase sliding clamp subunit (PCNA homolog)
M14R	3.E-09	COG0592	DNA polymerase sliding clamp subunit (PCNA homolog)
M225L	e-106	COG4581	Superfamily II RNA helicase
Nucleotide transport and metabolism / Translation, ribosomal structure and biogenesis			
M10L	7.E-09	COG0590	Cytosine/adenosine deaminases
Transcription / DNA replication, recombination, and repair			
M272L	7.E-40	COG0553	Superfamily II DNA/RNA helicases, SNF2 family
M201R	2.E-34	COG1061	DNA or RNA helicases of superfamily II

Cellular Processes			
Posttranslational modification, protein turnover, chaperones			
M423R	8.E-09	COG0695	Glutaredoxin and related proteins
M241L	5.E-06	COG0695	Glutaredoxin and related proteins
M670R	8.E-10	COG5054	Mitochondrial sulphhydryl oxidase involved in the biogenesis of cytosolic Fe/S proteins
M807R	8.E-21	COG5201	SCF ubiquitin ligase, SKP1 component
Cell envelope biogenesis, outer membrane			
M37R	e-177	COG0449	Glucosamine 6-phosphate synthetase, contains amidotransferase and phosphosugar isomerase domains
M467R	4.E-06	COG0463	Glycosyltransferases involved in cell wall biogenesis
M409R	3.E-10	COG0810	Periplasmic protein TonB, links inner and outer membranes
M822R	3.E-10	COG0810	Periplasmic protein TonB, links inner and outer membranes
M591R	6.E-07	COG0810	Periplasmic protein TonB, links inner and outer membranes
M803L	1.E-06	COG0810	Periplasmic protein TonB, links inner and outer membranes
M701R	4.E-06	COG0810	Periplasmic protein TonB, links inner and outer membranes
M417L	6.E-06	COG0810	Periplasmic protein TonB, links inner and outer membranes
M719L	8.E-99	COG1004	Predicted UDP-glucose 6-dehydrogenase
M186R	2.E-32	COG1215	Glycosyltransferases, probably involved in cell wall biogenesis
M721L	5.E-25	COG1215	Glycosyltransferases, probably involved in cell wall biogenesis
M128R	2.E-19	COG1215	Glycosyltransferases, probably involved in cell wall biogenesis

MT325 COGs			
ORF	E-value	COG	Description
M101R	3.E-44	COG3049	Penicillin V acylase and related amidases
Inorganic ion transport and metabolism			
M535L	e-153	COG0474	Cation transport ATPase
M99R	6.E-30	COG2032	Cu/Zn superoxide dismutase
Cell division and chromosome partitioning			
M389R	1.E-23	COG0037	Predicted ATPase of the PP-loop superfamily implicated in cell cycle control
Signal transduction mechanisms			
M313L	2.E-08	COG2453	Predicted protein-tyrosine phosphatase
General function prediction only / Signal transduction mechanisms / Transcription / DNA replication, recombination, and repair			
M221L	1.E-27	COG0515	Serine/threonine protein kinase
M794R	2.E-27	COG0515	Serine/threonine protein kinase
M543R	5.E-25	COG0515	Serine/threonine protein kinase

Metabolism			
Carbohydrate transport and metabolism			
M30R	5.E-32	COG0580	Glycerol uptake facilitator and related permeases (Major Intrinsic Protein Family)
M124L	1.E-11	COG2273	Beta-glucanase/Beta-glucan synthetase
M85R	6.E-58	COG3325	Chitinase
Nucleotide transport and metabolism			
M653L	3.E-63	COG0208	Ribonucleotide reductase, beta subunit
M777L	e-152	COG0209	Ribonucleotide reductase, alpha subunit
M264L	1.E-36	COG0756	dUTPase
M34L	4.E-33	COG1351	Predicted alternative thymidylate synthase
M425R	4.E-20	COG1428	Deoxynucleoside kinases
M530L	1.E-25	COG2131	Deoxyctydylate deaminase
Energy production and conversion / Coenzyme metabolism / General function prediction only			
M26L	8.E-68	COG1052	Lactate dehydrogenase and related dehydrogenases
Amino acid transport and metabolism			
M307L	4.E-50	COG0019	Diaminopimelate decarboxylase
M601L	1.E-36	COG0076	Glutamate decarboxylase and related PLP-dependent proteins
M766L	e-109	COG2957	Peptidylarginine deiminase and related enzymes
Secondary metabolites biosynthesis, transport, and catabolism			
M283R	2.E-06	COG1233	Phytoene dehydrogenase and related proteins
M233L	4.E-82	COG5310	Homospermidine synthase

Poorly Characterized			
General function prediction only			
M103L	4.E-47	COG0388	Predicted amidohydrolase
M742R	3.E-75	COG0488	ATPase components of ABC transporters with duplicated ATPase domains
M174L	1.E-28	COG0637	Predicted phosphatase/phosphohexomutase
M404R	8.E-45	COG0661	Predicted unusual protein kinase

MT325 COGs			
ORF	E-value	COG	Description
M833L	1.E-20	COG0666	FOG: Ankyrin repeat
M413L	8.E-20	COG0666	FOG: Ankyrin repeat
M829L	3.E-11	COG0666	FOG: Ankyrin repeat
M219L	6.E-23	COG1752	Predicted esterase of the alpha-beta hydrolase superfamily
M727L	8.E-06	COG2940	Proteins containing SET domain
M674R	5.E-13	COG3378	Predicted ATPase
M302L	3.E-07	COG5271	AAA ATPase containing von Willebrand factor type A (vWA) domain
M363L	4.E-07	COG5271	AAA ATPase containing von Willebrand factor type A (vWA) domain
Function Unknown			
M616L	3.E-07	COG5373	Predicted membrane protein
M557L	8.E-06	COG5373	Predicted membrane protein

CHAPTER 3

GENOMIC COMPARISONS OF THREE CHLORELLA VIRUSES, PBCV-1, NY-2A, AND MT325

3.1 INTRODUCTION

The genus *Chlorovirus* is composed of a group of viruses with very large and complex DNA genomes. These viruses are members of the family *Phycodnaviridae*. PBCV-1 was the first *Phycodnaviridae*, and until recently the only *Chlorovirus*, genome sequenced (1-5). Therefore, it has been classified as the prototype virus for this family. To better understand the *Chlorovirus* genus, the genomes of two additional viruses, NY-2A (an NC64A virus) and MT325 (a Pbi virus), were sequenced. The chlorella viruses have linear DNA genomes ranging in length from 314 to 368 kb and a G+C content of 40 or 45%. In Chapter 2, the annotations of the newly sequenced NY-2A and MT325 genomes were described in detail. This chapter focuses on the comparison of the genomes of the three sequenced chlorella viruses. First, the relative orientations of the chlorella viral genomes will be discussed. Second, the examination of similar genomic sequences, which has led to the identification of several gene homologs in the three sequenced chloroviruses, will be discussed. The abundance of gene homologs within this genus indicates the chloroviruses have a common ancestor. Finally, gene acquisition will be discussed. A comparison of the ORFs from these chlorella viruses with ORFs from other organisms suggests that the chlorella viruses either acquired their genes from organisms in all three domains of life or could have perhaps been involved in

eukaryogenesis, the putative process by which a complex DNA virus served as the ancestor of the eukaryotic nucleus (6).

3.2 EXPERIMENTAL METHODS

3.2.1 Comparing the Orientations of the Chlorella Viral Genomes

To determine the orientation of the sequenced viral genomes relative to the genome of the prototype chlorella virus PBCV-1, we used genomic dot plots. The PBCV-1, NY-2A, and MT325 genomes were individually plotted against the other chlorella viral genomes using blastn (DNA vs. DNA) and tblastx (translation vs. translation). Similarities with E-values $<10^{-5}$ are represented. Additionally, a genomic lineup was created from the largest chlorella virus, NY-2A. Each NY-2A ORF was used to search for homologs in either the PBCV-1 or the MT325 genome. A line was drawn between detected homologs. If the homologs were in the same orientation, the line was blue, and if the homologs were translated in opposite orientations, the line was red.

3.2.2 Genomic Comparisons

In Chapter 2, the classification of the major/minor ORFs and the annotations of each were discussed in detail. Briefly, every ORF was analyzed using the NCBI NR database, the Protein Families database (Pfam), and the Clusters of Orthologous Groups (COGs). ORFs were assigned as homologs if blastp analysis resulted in an E-value < 0.05 .

3.2.3 Domain and Kingdom Scatter Plots

Every NY-2A and MT325 ORF defined in Chapter 2 and the previously defined PBCV-1 ORFs were used to create domain and kingdom scatter plots. The results of a blastp comparison of the ORFs from these three chloroviruses with the proteins in the NCBI NR database were used to create these domain scatter plots. The highest score from the Eukaryota, Bacteria, and Archaea domains were plotted. In this analysis, the blastp cutoff was E-value < 0.001. However, the scores, rather than the E-values, were used to construct the plots. Furthermore, the domain Eukaryota was split into its kingdoms: Metazoa, Protist, Plant, and Fungi. The same criteria used for the domain analysis were used for the analysis of the Eukaryota kingdoms.

3.3 RESULTS

3.3.1 Comparison of the Orientations of the Chlorella Viral Genomes

To determine the orientation of the newly sequenced NY-2A and MT325 chlorella viral genomes relative to the genome of the prototype virus PBCV-1, genomic dot plots were created. Figure 3.1 shows the dot plots of the three chlorella viral genomes at the DNA level based on blastn (DNA vs. DNA) analysis. The dots represent similarities with E-value < 10^{-5} . The red dots correspond to similarities greater than 200 bp, and the black dots correspond to similarities less than 200 bp. Figure 3.2 shows the dot plots of the three chlorella viral genomes at the translated sequence levels based on tblastx (translation vs. translation) analysis. The criteria used in the tblastx analysis were the same as that used for the blastn analysis. As expected due to degeneracy, the tblastx dot plots revealed a higher degree of similarity than the blastn dot plots. These results are

consistent with our initial conclusion that there is co-linearity between the viruses which infect the same host, *Chlorella* NC64A. However, when the Pbi virus, MT325, is compared to either of the NC64A viruses, PBCV-1 or NY-2A, only a small number of similarities at the DNA level are observed. There was a significant increase in the number of similarities detected between the Pbi virus and the NC64A viruses at the translation level. However, there is still very little co-linearity detected between the viruses which infect different hosts.

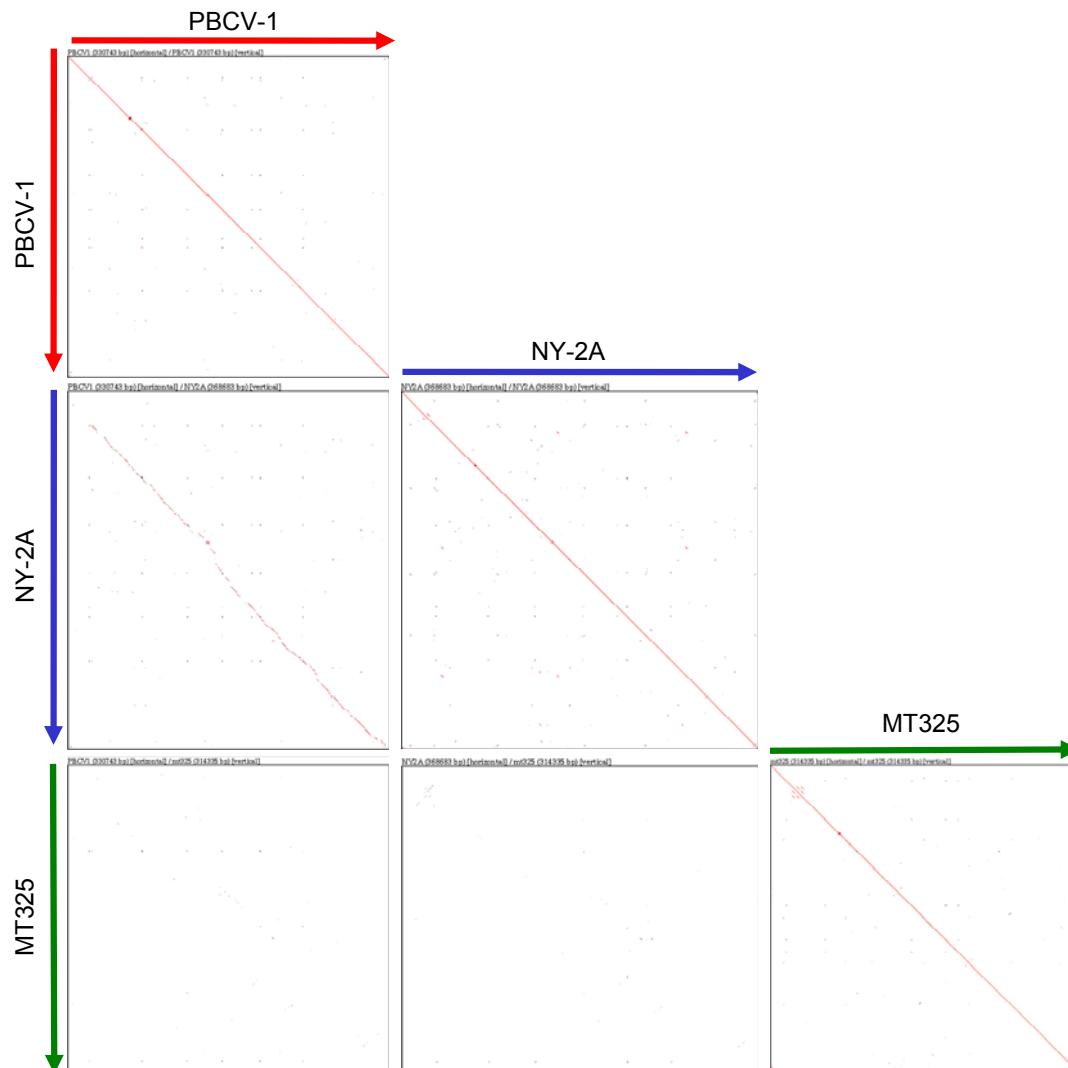


Figure 3.1 Genomic dot plots of chlorella viruses based on a DNA analysis (blastn)

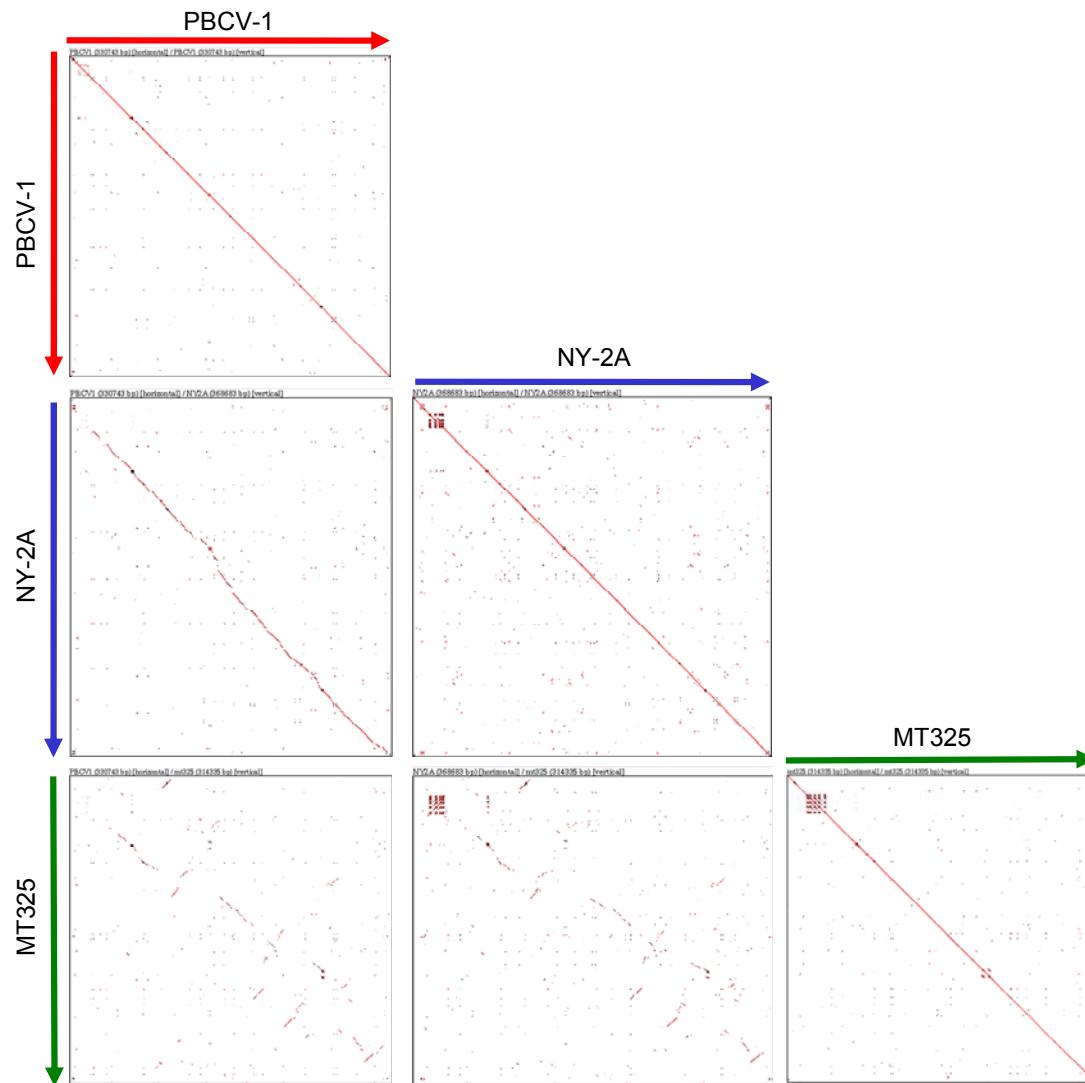


Figure 3.2 Genomic dot plots of chlorella viruses based on a translational analysis (tblastx)

The cluster of dots in the top left corner of three tblastx plots (NY-2A vs. NY-2A, NY-2A vs. MT325, and MT325 vs. MT325) represents repeats within the left end termini of the viral genomes. There are four repeats in the NY-2A genome (Figure 3.3A), whereas the MT325 genome (Figure 3.3B) encodes five repeats. These clusters of dots,

which correspond to Vp260-like proteins, are more abundant in the left end terminus of the NY-2A and MT325 genomes than the PBCV-1 genome.

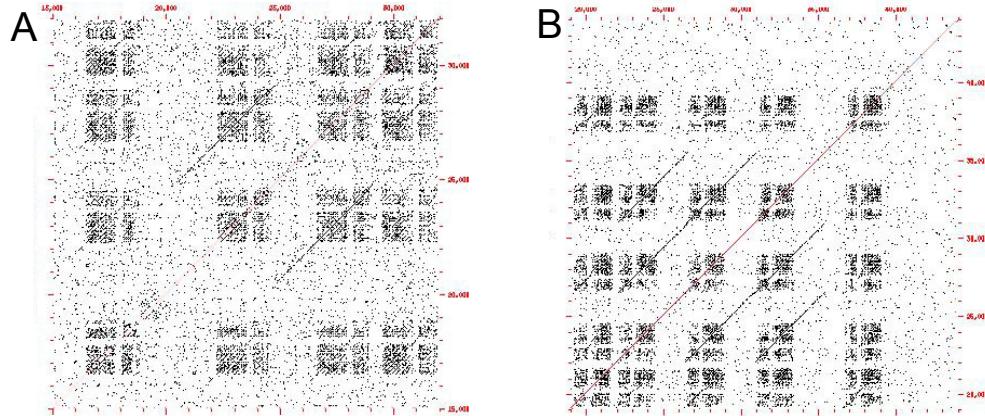


Figure 3.3 Dot plots of the NY-2A or the MT325 genome to itself. Shown are the Vp260-like protein repeats located in the left end termini of the NY-2A (A) and MT325 (B) viral genomes.

Figure 3.4 shows the genomic lineup of the three sequenced chlorella viral genomes. The largest genome, that of NY-2A, was used as a point of reference and, therefore, is placed in the middle. The ORFs from the NY-2A genome were compared to the ORFs from either the PBCV-1 or the MT325 genomes. When a homolog was detected, a line was drawn between the ORFs. If the ORFs were in the same orientation, the line representing their homology was blue, and if the ORFs were in the opposite orientation, the line representing their homology was red. As with the dot plots, the genomic lineup revealed that there is a high degree of co-linearity between the two NC64A viral genomes, PBCV-1 and NY-2A. Additionally, the genomic lineup of the 2 NC64A viral genomes revealed that the number of co-linear homologs is more condensed near the right terminus when compared to the left terminus. This lack of homology in the

left end termini of the genomes exemplifies the diversity between chlorovirus genomes as mentioned in Chapter 1. Furthermore, the dot plots and the genomic lineup showed that there were many homologous ORFs between NY-2A and MT325 but that their orientations differed.

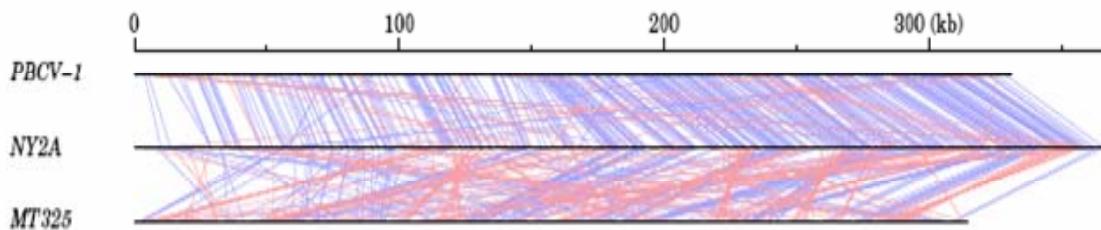


Figure 3.4 Genomic lineups of the three sequenced chlorella viral genomes

3.3.2 Comparison of the Chlorella Virus Open Reading Frames

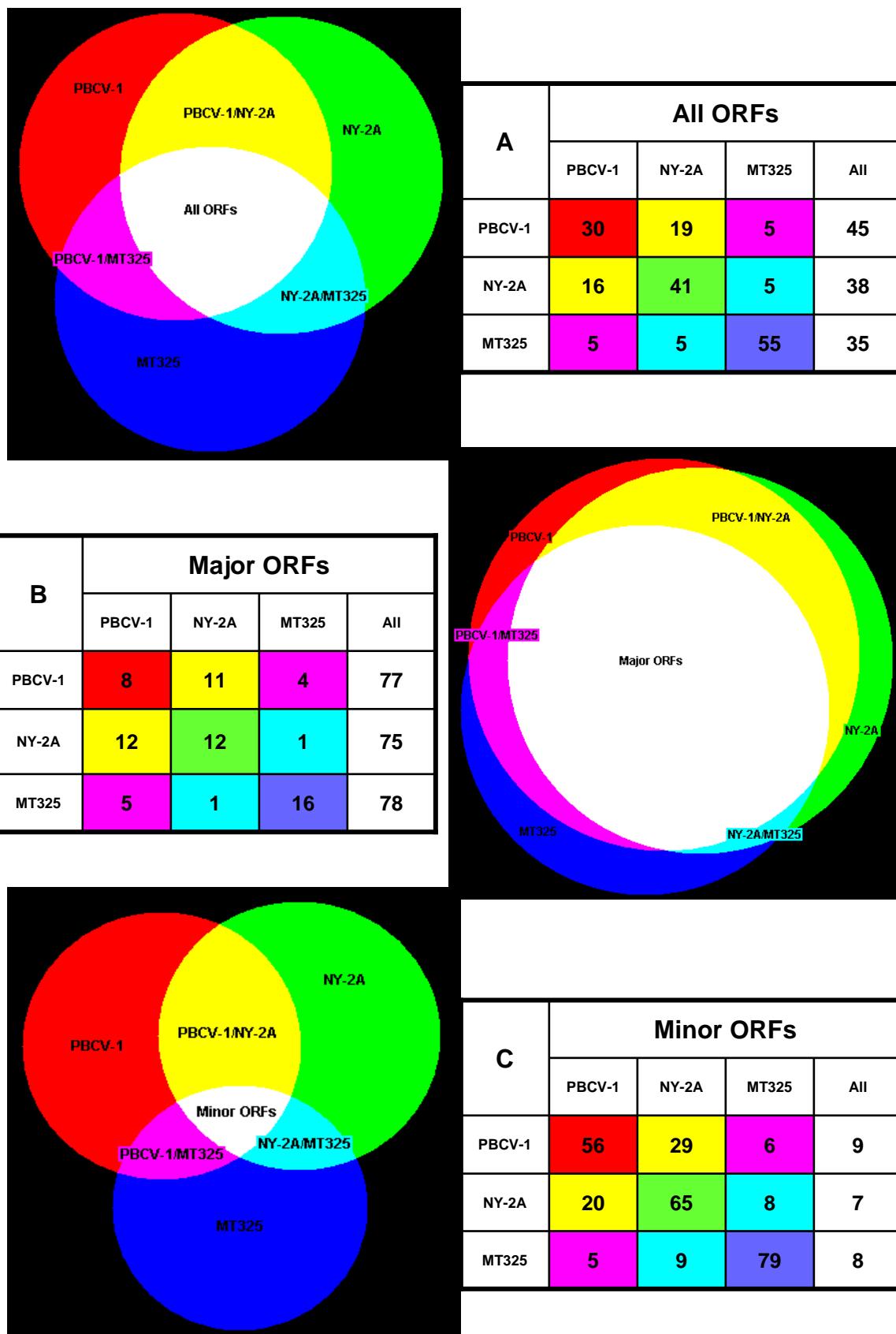
Of the 2419 ORFs (defined as a stretch of DNA that would translate into at least 65 codons) identified in the genomes of the three chlorella viruses, only 1102 (46%) were classified as major, i.e. as probably encoding a protein. The remaining 1317 ORFs (54%) were classified as minor. Every chlorella virus ORF was examined to establish homology amongst the chlorella viruses. ORFs were assigned as homologs if blastp analysis resulted in an E-value < 0.05. Addendum 3.6.1 lists every ORF along with their homologs and putative function when detected. Figure 3.5 shows the percentage of the ORFs in each genome as compared to the other two chlorella viruses. PBCV-1, NY-2A, and MT325 ORFs are represented as red, green, and blue circles, respectively.

The first diagram (Figure 3.5A) represents all of the ORFs identified in the three genomes. Thirty-five to 45% of the ORFs encoded by the chlorella viral genomes had homologs in all three of the chlorella viruses. Nineteen percent of the ORFs in the PBCV-1 genome and 16% of the ORFs in the NY-2A genome are only homologous between the two NC64A viruses. In contrast, only 5% of the ORFs were homologous between a NC64A (PBCV-1 or NY-2A) and the Pbi (MT325) virus.

An analysis of the major (B) or the minor (C) ORFs from the three chlorella viral genomes is depicted in Figure 3.5. When the percentages of the major ORFs are compared to the percentages of all of the ORFs from a genome, there is an increase in the number of homologous ORFs detected among the three chlorella viruses. Seventy-five to seventy-eight percent of the major ORFs from the chlorella viral genomes were homologous. This finding implies that 75% of the major ORFs from the chlorella viral genomes were either acquired from their last common ancestor and/or are essential for the viral life cycle and, therefore, are under pressure to be conserved. Eighty-seven to eighty-eight percent of the major ORFs from the two NC64A viruses, PBCV-1 or NY-2A, are homologous. Eighty-one percent of the major ORFs from the PBCV-1 genome are homologous with ORFs from the MT325 genome, while only 76% of the major ORFs from the NY-2A genome are homologous with ORFs from the MT325 genome. The results of this analysis indicate that the percentage of homology between these three chlorella viral genomes is relatively equivalent. However, the major ORFs from the PBCV-1 genome are slightly more homologous to the ORFs from the MT325 genome than is the homology between the NY-2A and the MT325 genomes. Overall, only a small percentage (8-16%) of the major ORFs from a given viral genome is unique. It is

not surprising that the minor ORFs have significantly different homology percentages. In contrast to the major ORFs, 56-79% of the minor ORFs from a given genome are unique. Only 7-9% of the minor ORFs from a given genome are homologous among all chlorella viruses. This is further confirmation that the major ORFs are largely conserved and may play a significant role in the life cycle of the chlorella viruses, as opposed to majority of the minor ORFs which are largely unique and are not homologous to any other sequences in the public databases.

Figure 3.5 Genomic ORF comparisons of chlorella viruses. The red, green, and blue circles represent the chlorella viruses PBCV-1, NY-2A, and MT325 respectively. The colored boxes correlate to the genome or combination of genomes that they represent. The number inside the box is the percentage of the genomic ORFs as they correspond to a certain classification. (A) All ORFs, (B) Major ORFs, or (C) Minor ORFs are shown separately to compare the differences. Proportional Venn Diagrams were constructed using VennCircles (7).



3.3.3 Domain and Kingdom Scatter Plots

An analysis of the domain scatter plots revealed that the three chloroviruses encode genes which are present in every domain of life: Eukaryota, Bacteria, and Archaea. However, when comparing the highest scores from one domain to another, all three chlorella viral ORFs had more representatives in the Eukaryotes than in Bacteria or Archaea (Figure 3.6). Of these three domains, the Archaea domain typically had lower scores when compared to the other domains. Nevertheless, there are still some ORFs which are representatives of the Archaea domain.

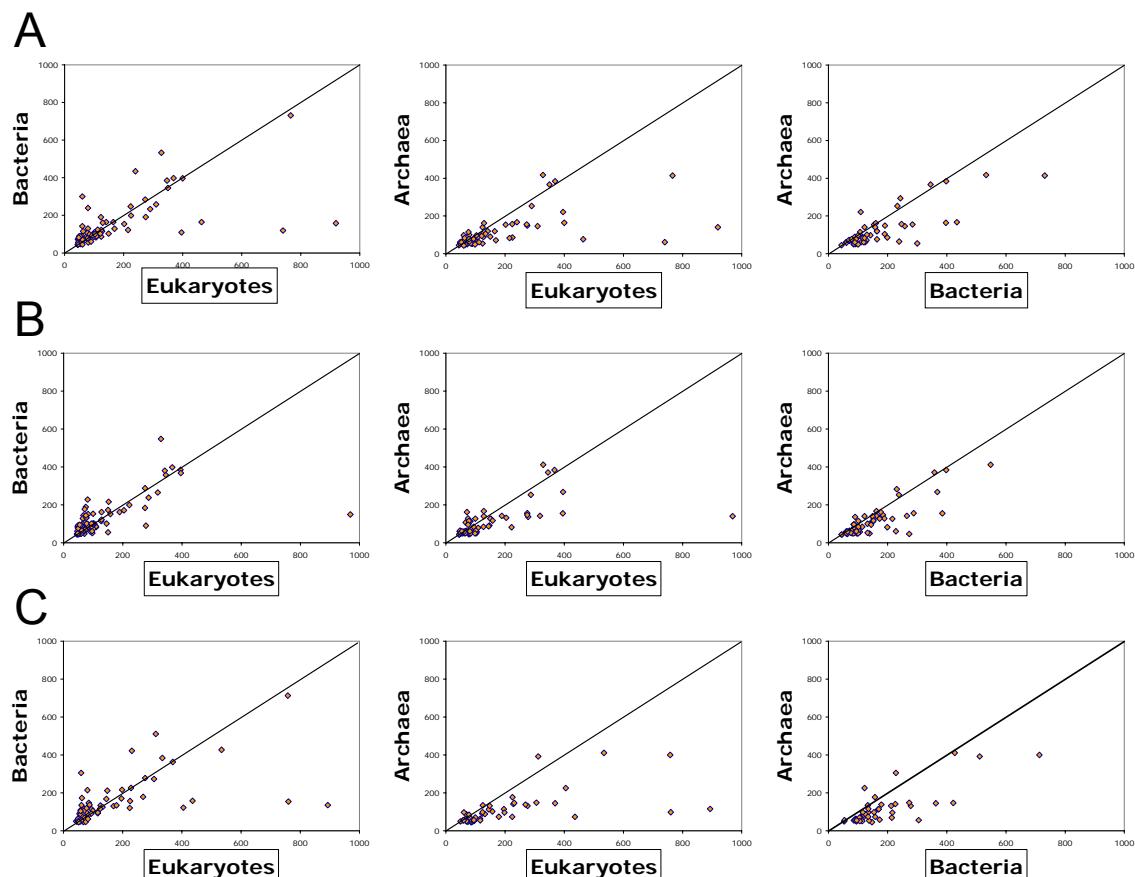


Figure 3.6 Domain scatter plots of the chlorella viruses, (A) PBCV-1, (B) NY-2A, and (C) MT325. The highest score that an ORF received when compared to each domain of life were plotted.

Comparisons with the kingdoms within the Eukaryotic domain revealed a similar pattern for all three chlorella viruses. The three viral genomes represented the kingdoms within the eukaryotic domain in a similar order: Metazoa > Fungi > Protist > Plant. Addendum 3.6.2 shows all of the plots between the three domains and four kingdoms.

3.4 DISCUSSION

To better understand the members of the genus *Chlorovirus*, the genomes of two additional chlorella viruses, NY-2A (a NC64A virus) and MT325 (a Pbi virus), were sequenced and compared with the sequence and annotation of the genome of the prototype virus, PBCV-1. Genomic dot plots revealed that the two NC64A viruses, PBCV-1 and NY-2A, were nearly co-linear. MT325, a Pbi virus, was orientated to try to find the best possible relationship with the NC64A viruses. The translation level comparison of MT325 to the NC64A viruses revealed that approximately 30-35% of the MT325 genome was co-linear with the genomes of the NC64A viruses (Figure 3.7, red ovals). However, there are segments of the MT325 genome that have undergone genomic rearrangements. For example, approximately 15-20% of the MT325 genome appears to be ‘co-linear’ with a NC64A virus genome, however these genomic segments are in the reverse orientation (Figure 3.7, blue ovals). Typically, these “inverted” sequences code for minor ORFs with no known function when compared to the public databases. However, there are some major ORFs in this region which, based on sequence homology, encode proteins homologous to the public databases (i.e. arginine/ornithine decarboxylase). In addition to the “inverted” co-linear segments of a Pbi virus genome

compared with a NC64A virus genome, there are two significant inverted segments that should be noted (Figure 3.7 green ovals). For example, a DNA polymerase, a PCNA, and a cytosine deaminase are in opposite orientations when comparing MT325 to a NC64A virus. In the NC64A viruses this set of putative proteins is encoded a third of the way into the genome. However, the MT325 genome codes for these proteins in the first 5% of its genome. A few examples of the ORFs within these three classifications are listed in Table 3.1. The functional significance of these MT325 “inverted” genomic regions has not yet been determined and will be the subject of further examination.

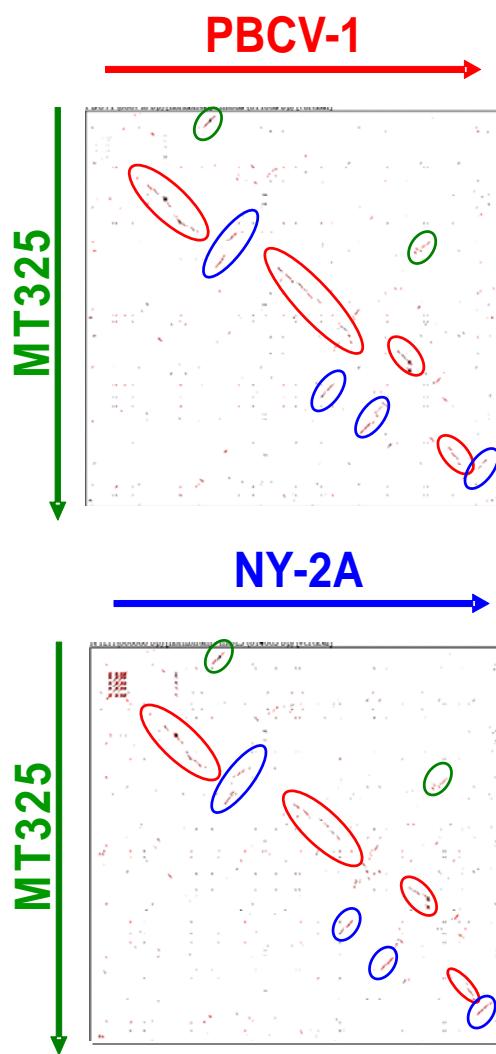


Figure 3.7 Genomic dot plots of MT325 and a NC64A virus, PBCV-1 or NY-2A. The red ovals represent the ~30% co-linearity between the viral genomes. The blue ovals represent the “inverted” co-linear sequences, and the green ovals represent “inverted” sequences which are non-linear to a NC64A virus.

Table 3.1

PBCV-1		NY-2A			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
Red Ovals								
A94L	364	B137L	320	31%	M124L	348	57%	β-1,3-glucanase
A103R	330	B148R	321	76%	M133R	319	54%	mRNA guanylyltransferase
A105L	284	B150L	290	68%	M137L	279	45%	Ubiquitin C-terminal hydrolase
A107L	290	B154L	338	66%	M139L	312	38%	Transcription factor TFIIB
A125L	180	B175L	192	91%	M163L	180	71%	Transcription factor TFIIS
A134L	165	B185L	164	89%	M180L	158	47%	Homing endonuclease GIY-YIG
A153R	459	B203R	460	81%	M201R	454	67%	Superfamily II helicase
A166R	268	B214R	268	79%	M215R	279	59%	Exonuclease
A173L	288	B226L	279	90%	M219L	274	57%	Patatin-like phospholipase
Blue Ovals								
A426R	114	B580R	116	79%	M444R	116	42%	6-phosphofructokinase (?)
A427L	119	B581L	119	76%	M445L	116	44%	Thioredoxin
A429L	476	B583L	453	77%	M451L	442	34%	Ankyrin Repeats
A430L	437	B585L	437	95%	M463L	433	79%	Major Capsid Protein
A437L	103	B591L	103	74%	M476L	202	70%	DNA binding protein
A612L	119	B813L	119	78%	M727L	119	57%	Histone H3, Lys 27 methylase
A614L	577	B816L	492	44%	M729L	583	51%	Serine/Threonine protein kinase
A617R	321	B818R	317	79%	M733R	316	48%	Serine/Threonine protein kinase
A526R	129	B692R	146	84%	M501R	148	52%	Glycerophosphodiesterase (?)
A539R	173	B724R	173	80%	M517R	168	50%	Homing endonuclease GIY-YIG
A540L	1176	B725L	1191	59%	M518L	1189	33%	Autotransporter adhesin
Green Ovals								
A185R	913	B249R	760	87%	M019L	907	73%	δ DNA polymerase
A189/192R	1299	B258R	1278	71%	M015L	1286	45%	SWI/SNF chromatin remodeling complex
A193L	262	B261L	262	97%	M014R	264	68%	PCNA
A200R	118	B271R	145	75%	M010L	119	62%	Cytosine deaminase
A392R	258	B536R	254	87%	M586L	255	83%	ATPase (DNA packaging)
A399R	194	B547R	193	73%	M570L	164	51%	RNase H
A402R	227	B550R	232	85%	M564L	228	49%	Lipoprotein lipase
A629R	771	B832R	1103	90%	M777L	765	79%	Ribo. Reductase (large subunit)
A638R	359	B844R	359	96%	M766L	363	62%	Agmatine iminohydrolase
A654L	197	B853L	197	86%	M758R	197	51%	N-acetyltransferase

Table 3.1 Comparison of NC64A (PBCV-1 or NY-2A) viral ORFs to the MT325 viral ORFs. Red Ovals correspond to ORFs which are transcribed in the same direction and are linear from the left end termini to the right end termini of the viral genomes. These homologs are located in similar positions within the viral genomes. Blue Ovals correspond to the MT325 inverted regions. The ORFs from a NC64A virus are encoded in one direction, whereas the homologous ORF from the MT325 virus is encoded in the opposite direction. Additionally, the NC64A ORFs are linear from the left end termini to the right end termini whereas the MT325 homologs are oriented in the opposite direction. These homologous ORFs are located in similar positions within the viral genomes. Green Ovals correspond to homologs which have the same characteristics of the blue ovals, however the locations of these homologs are not found within the same location. Listed are the ORFs, the number of amino acids encoded, and the % identity to the PBCV-1 ORF. The ovals correspond to those shown in Figure 3.7.

The genomic dot plots also revealed a cluster of dots located within the left terminus for two of the chlorella viruses, NY-2A and MT325. These clusters are associated with the Vp260-like proteins, which are large proteins consisting of ~1500 amino acids. In NY-2A and MT325, there are 4-5 repeats of these proteins. These repeats are located in the left end termini of the genomes and comprise 1-2% of the total genomes.

The identification, classification, and annotation of the NY-2A and MT325 ORFs were presented in Chapter 2. The ORFs from these two newly sequenced viral genomes were compared with the ORFs from the PBCV-1 genome. Approximately 75% of the major ORFs were homologous within all three of the sequenced chlorovirus genomes. Interestingly, 64% of these homologous major ORFs are classified as unknown based on a lack of sequence similarity found in public databases. This result suggests that these ORFs must have some significant function in the viral life cycle or they would have been deleted over time. The remaining 36% of the major ORFs that are homologous among the three chlorella viruses have been assigned a putative function and are listed in Figure 3.8. As expected, several of these chlorella virus homologs are involved in DNA replication, recombination, and repair. For example, a δ DNA polymerase, a DNA

topoisomerase II, a DNA-binding protein, an exonuclease, a superfamily III helicase, an archaeo-eukaryotic primase, 2 PCNAs, an RNase H, a pyrimidine dimmer-specific glycosylase, a replication factor C, and 4 types of ATPases exist in all of the chlorella viruses. Nucleotide metabolism is essential for the virus to replicate its large DNA genome. Therefore, it is not surprising that several putative proteins involved in nucleotide metabolism are found in all three of the chlorella viruses. These proteins include the large and small subunits of ribonucleotide reductase, thioredoxin, glutaredoxin, NTP pyrophosphohydrolase, thymidylate synthase, dUTP pyrophosphatase, cytosine deaminase, and dCMP deaminase. Additionally, several of the homologous proteins coded for by the three chlorella viruses are involved in transcription. For example, there are 3 different transcription factors (TFIIB, TFIID, and TFIIS), a histone H3 Lys27 dimethylase, a SWI/SNF helicase, 3 superfamily II helicases, RNA triphosphatase, RNase III, mRNA guanylyltransferase, and a VLTF2-type transcription factor. There are a set of homologous cytosine DNA methylases encoded by these three viruses which are likely involved in the modification of newly replicated genomic DNA. In addition, there are 3 types of homologous putative proteins involved in viral integration and transposition encoded by all three of the chlorella viruses. These proteins include 7 GIY-YIG homing endonucleases, 4 HNH homing endonucleases, and a transposase. Furthermore, there are 6 Ser/Thr protein kinases, a potassium channel protein, and a dual specificity phosphatase which are encoded by the genomes of PBCV-1, NY-2A, and MT325 viruses. Four putative protein homologs are also involved in cell wall degradation: a chitinase, a chitosanase, a β -1, 3-glucanase, and a β & α 1, 4 linked glucuronic lyase. Nine putative proteins encoded by all three of these viruses are

involved in sugar and lipid manipulation. These proteins include a 6-phosphofructokinase, a fructose-2, 6 bisphosphatase, a fucosyltransferase, a glucosamine synthetase, a glycerophosphodiesterase, a glycosyltransferase, a lipoprotein lipase, a patatin phospholipase, and an UDP-glucose dehydrogenase. There are 5 putative proteins (prolyl 4-hydroxylase, protein disulfide isomerase, SKP1 protein, thiol oxidoreductase, and ubiquitin C-terminal hydrolase) involved in protein synthesis, modification, and degradation. The chlorella viruses have an array of miscellaneous putative protein homologs common to all three of the chlorella viruses. For example, all of the chlorella viruses encode an amidase, a monoamine oxidase an ABC transporter protein, an agmatine iminohydrolase, 6 ankyrin repeats, 2 major capsid proteins and four enzymes involved in polyamine biosynthesis (an ornithine\arginine decarboxylase, a histidine decarboxylase, a homospermidine synthase, and a carbon-nitrogen hydrolase).

Eight percent of the major ORFs of the PBCV-1 genome are unique and only 7 of the 31 ORFs have been assigned a putative function. These functions include an acetyltransferase, an O-methyltransferase, a glycosyltransferase, a glutamate receptor, a ligand-gated channel protein, and 2 DNA restriction endonucleases. The remaining 24 major unique PBCV-1 ORFs are classified as unknowns. Twelve percent of the major ORFs of the NY-2A genome and 16% of the major ORFs of the MT325 genome are unique. Seventeen of the 50 major ORFs unique to NY-2A and 14 of the 54 major ORFs unique to MT325 have been assigned putative functions. The remaining major unique ORFs are classified as unknowns.

The annotation of the three chlorella viruses revealed that some of the putative proteins are eukaryotic-like while others are prokaryotic-like. Domain scatter plots were

used to demonstrate that the three chloroviruses encode genes which are present in every domain of life. However, there was an apparent trend with the three chlorella viruses. Overall, the ORFs were more eukaryotic-like than prokaryotic-like. There was also a small number of ORFs which closely resembled those found in the Archaea domain. Additionally, the eukaryotic domain was further divided into its four eukaryotic kingdoms. Once again, a similar trend was observed with each of the three sequenced chlorella viral genomes: Metazoa > Fungi > Protist > Plant.

As mentioned in Chapter 1, accumulating evidence indicates that at least some of the phycodnavirus genes, and by implication the viruses themselves, have a long evolutionary history, possibly dating back to the time that prokaryotic and eukaryotic organisms separated (ca. 2.0 – 2.7 billion years ago) (8-10). The domain scatter plots confirmed that the chlorovirus genomes are mosaics of prokaryotic and eukaryotic genes. These chloroviruses encode an amazing variety of proteins, some of which are very prokaryotic-like (the DNA methyltransferases and restriction endonucleases) and others that are very eukaryotic-like (the mRNA capping enzymes). The genes encoding these proteins could have been acquired from their hosts over time, but another possible explanation for their presence is that these genes existed at the time that prokaryotic and eukaryotic organisms separated. This would imply that the progenitor organism(s) lost genes as it was captured by various eukaryotic organisms and evolved into viruses. Some evolutionary biologist have suggested one of the earliest eukaryotic cells could have been a single celled alga (11), and a virus of this progenitor may have co-evolved with its hosts. Although, the proposal of the antiquity of the phycodnavirus proteins is speculative (although not without support), additional viral genomes are currently being

sequenced that may provide additional information about the evolutionary origin of these viruses.

3.5 REFERENCES

- (1) Li, Y., Lu, Z., Burbank, D. E., Kutish, G. F., Rock, D. L., and Van Etten, J. L. (1995) Analysis of 43 kb of the chlorella virus PBCV-1 330 kb genome: map position 45 to 88. *Virology* 212, 134.
- (2) Lu, Z., Li, Y., Zhang, Y., Kutish, G. F., Rock, D. L., and Van Etten, J. L. (1995) Analysis of 45 kb of DNA located at the left end of the chlorella virus PBCV-1 genome. *Virology* 206, 339.
- (3) Kutish, G. F., Li, Y., Lu, Z., Furuta, M., Rock, D. L., and Van Etten, J. L. (1996) Analysis of 76 kb of the chlorella virus PBCV-1 330-kb genome: map positions 182 to 258. *Virology* 223, 303.
- (4) Lu, Z., Li, Y., Que, Q., Kutish, G. F., Rock, D. L., and Van Etten, J. L. (1996) Analysis of 94 kb of the chlorella virus PBCV-1 330-kb genome: map positions 88 to 182. *Virology* 216, 102.
- (5) Li, Y., Lu, Z., Sun, L., Ropp, S., Kutish, G. F., Rock, D. L., and Van Etten, J. L. (1997) Analysis of 74 kb of DNA located at the right end of the 330-kb chlorella virus PBCV-1 genome. *Virology* 237, 360.
- (6) Bell, P. J. (2001) Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? *J Mol Evol* 53, 251-6.
- (7) Chow, S., and Ruskey, F. (2004) Drawing Area-Proportional Venn and Euler Diagrams. *Proceedings of Graph Drawing 2003* 2912, 466-477.
- (8) Villarreal, L. P., and DeFilippis, V. R. (2000) A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. *J. Virol.* 74, 7079.
- (9) Gu, X. (1997) The age of the common ancestor of eukaryotes and prokaryotes: statistical inferences. *Mol Biol Evol* 14, 861-6.
- (10) Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Claverie, J. M. (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* 306, 1344-50.

- (11) Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G., and Bhattacharya, D. (2004) A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* 21, 809-818.

PBCV-1			NY-2A			MT325			Description
ORF	AA		ORF	AA	%ID	ORF	AA	%ID	
A2L	183								
A3R	141	B883L	102	57%					
A5R	262	B880L	235	61%	M833L	268	39%	Ankyrin Repeats	
A7L	186	B024L	532	31%	M833L	268	42%	Ankyrin Repeats	
A8L	137	B874L	366	40%	M833L	268	36%	Ankyrin Repeats	
A9R	175	B33R	254	82%	M247R	250	70%		
A10R	401	B748L	400	41%	M269L	400	41%	Major Capsid Protein	
A11L	403	B748L	400	52%	M269L	400	49%		
A14R	1369	B040L	1471	23%	M042R	956	31%		
A18L	1335	B055L	953	31%	M152R	1631	26%	VP260 like protein	
A25L	401	B055L	953	25%	M042R	956	26%		
A29L	820	B055L	953	26%	M042R	956	30%		
A34R	308	B365L	605	24%	M729L	583	26%	Serine/Threonine protein kinase	
A35L	549	B192R	1075	33%	M522L	615	33%		
A37L	105	B067L	119	77%					
A39L	151	B068L	153	66%	M807R	155	36%	SKP-1 protein	
A41R	412	B070R	433	96%	M803L	500	66%	Periplasmic protein TonB	
A44L	599	B073L	603	89%				ATPase (AAA+ class)	
A48R	123	B074R	123	90%	M170R	122	38%		
A49L	219	B075L	228	94%				Glycerophosphoryl diesterase	
A50L	141	B076L	116	88%	M627L	144	69%	Pyrimidine dimer-specific glycosylase	
A51L	203	B078L	153	91%					
A53R	363				M026L	346	70%	D-lactate dehydrogenase	
A57R	544								
A60L	234								
A61L	209							O-methyltransferase	
A63L	232	B103R	353	36%					
A64R	638							Glycosyltransferase	
A67R	309	B099R	312	75%	M686L	411	27%		
A71R	354	B103R	353	82%					
A75L	280	B107L	288	87%	M009L	269	40%		
A77L	91	B115L	85	63%					
A78R	298	B116R	298	97%	M103L	296	68%	N-carbamoylput. amidohydrolase	
A79R	229	B339L	243	83%	M025R	246	33%		
A81L	189	B122L	192	73%	M005L	180	39%		
A84L	188	B124L	187	82%	M007L	170	33%		
A85R	242	B126R	230	81%	M401R	224	50%	Prolyl-4-hydroxylase	
A87R	456	B133R	444	65%	M093L	377	33%	HNH endonuclease	
A88R	257				M828R	223	27%		
A89R	68	B441R	65	92%					
A90R	154								
A92/93L	359	B136L	423	90%	M631L	365	39%		
A94L	364	B137L	320	31%	M124L	348	57%	β-1,3-glucanase	
A98R	568				M128R	561	73%	Hyaluronan synthase	
A100R	595	B143R	595	87%	M037R	593	61%	Glucosamine synthetase	
A103R	330	B148R	321	76%	M133R	319	54%	mRNA guanylyltransferase	
A105L	284	B150L	290	68%	M137L	279	45%	Ubiquitin C-terminal hydrolase	
A107L	290	B154L	338	66%	M139L	312	38%	Transcription factor TFIIB	

PBCV-1			NY-2A			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID		
A109L	108	B157L	163	84%	M140L	168	58%		
A111/114R	389	B159R	867	94%	M467R	849	58%	Fucosyl-Glycosyltransferase	
A118R	345	B163R	343	88%				GDP-D-mannose dehydratase	
A121R	97	B168R	104	75%	M150R	104	71%		
A122R	1032	B170R	1386	45%	M152R	1631	44%		
A123R	311	B170R	1386	43%	M152R	1631	39%		
A125L	180	B175L	192	91%	M163L	180	71%	Transcription factor TFIIS	
A127R	245	B177R	239	82%	M166R	220	63%		
A129R	358	B408R	436	53%	M427R	352	44%		
A130R	105								
A131L	136	B181L	135	69%	M367L	219	38%		
A133R	207	B184R	247	95%					
A134L	165	B185L	164	89%	M180L	158	47%	GIY-YIG endonuclease	
A135L	182								
A137R	72	B188R	97	66%					
A138R	273	B130R	161	25%	M278R	267	24%		
A139L	108	B190L	100	77%	M184L	113	55%		
A140/145R	1130	B192R	1075	52%	M190R	1069	45%		
A148R	110	B201R	122	38%					
A150L	152	B197L	131	60%	M200L	103	39%		
A151R	135	B201R	122	94%					
A153R	459	B203R	460	81%	M201R	454	67%	Superfamily II helicase	
A154L	347	B408R	436	61%	M531L	365	51%		
A157L	110	B207L	111	77%	M204L	117	55%		
A158L	104	B208L	95	55%	M560L	183	34%		
A161R	123	B209R	103	39%					
A162L	411							Glutamate receptor	
A163R	433							Ligand-gated channel protein	
A165L	279	B212L	148	75%	M213L	152	38%		
A166R	268	B214R	268	79%	M215R	279	59%	Exonuclease	
A168R	166	B215R	167	85%	M218R	161	52%		
A169R	323	B222R	321	85%				Aspartate transcarbamylase	
A171R	387	B224R	368	80%	M803L	500	64%		
A173L	288	B226L	279	90%	M219L	274	57%	Patatin-like phospholipase	
A175R	69								
A177R	246	B120R	244	60%	M025R	246	35%		
A180R	108				M789L	110	59%	Fibronectin binding protein	
A181/182R	830	B239R	807	36%	M791R	520	63%	Chitinase	
A185R	913	B249R	760	87%	M019L	907	73%	δ DNA polymerase	
A189/192R	1299	B258R	1278	71%	M015L	1286	45%	SWI/SNF chromatin remodeling complex	
A193L	262	B261L	262	97%	M014R	264	68%	PCNA	
A196L	152	B264L	152	86%	M012R	156	62%		
A199R	101	B267R	100	80%	M011L	96	46%		
A200R	118	B271R	145	75%	M010L	119	62%	Cytosine deaminase	
A201L	94	B272L	100	69%	M084L	97	72%		
A202L	113	B273L	112	85%	M312R	112	53%		
A203R	216	B274R	216	92%	M310L	220	68%		
A205R	206	B277R	206	76%	M308L	235	59%		

PBCV-1			NY-2A			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID		
A207R	372	B278R	372	86%	M307L	372	63%	Ornithine/Arginine decarboxylase	
A208R	311	B283R	223	59%					
A213L	148	B284L	148	80%	M299R	149	52%		
A214L	135	B285L	133	79%	M298R	157	50%		
A215L	321	B288L	306	72%	M289R	368	69%	β & α 1,4 linked glucuronic lyase	
A217L	394	B289L	382	86%	M283R	386	73%	Monoamine oxidase	
A219/222/226R	677				M721L	604	70%	Glycosyltransferase	
A227L	137	B291L	138	78%	M244R	130	62%		
A229L	77	B292L	75	66%	M556R	174	32%		
A230R	196	B293R	213	79%	M242L	190	40%		
A231L	383	B298L	372	78%	M457R	376	63%		
A233R	112	B302R	111	73%	M240L	90	48%		
A234L	108	B303L	108	66%	M238R	119	51%		
A237R	518	B305R	507	94%	M233L	502	69%	Homospermidine synthase	
A239L	111	B314L	145	54%					
A241R	725	B316R	725	92%	M225L	715	50%	Superfamily II helicase	
A243R	302				M107R	302	76%		
A245R	187				M099R	169	76%	Cu/Zn-superoxide dismutase	
A246R	288	B322R	309	68%	M822R	599	60%	Pathogenesis-related protein	
A247R	383	B323R	380	61%	M413L	333	24%	Ankyrin Repeats	
A248R	308	B331R	309	74%	M794R	270	44%	Serine/Threonine protein kinase	
A250R	94	B336R	94	90%	M183R	95	49%	Potassium channel protein	
A251R	326	B236L	324	80%				Adenine methyltransferase	
A252R	342							DNA restriction endonuclease	
A253R	150	B342R	166	82%	M109R	149	34%		
A254R	158							Acetyltransferase	
A256L	411								
A257L	361								
A259L	174								
A260R	484				M085R	523	77%	Chitinase	
A261R	136	B344R	203	83%					
A262L	123	B348L	215	83%	M114L	226	49%		
A263L	141	B348L	215	79%					
A265L	170	B350L	249	88%	M116L	255	65%		
A267L	314	B747L	308	41%	M578R	208	34%	HNH endonuclease	
A271L	159	B354L	264	80%				Lysophospholipase	
A273L	401	B358R	225	63%					
A275R	252	B267L	251	80%	M025R	246	33%		
A277L	303	B388L	268	32%	M794R	270	33%	Serine/Threonine protein kinase	
A278L	610	B365L	605	72%	M143L	605	45%	Serine/Threonine protein kinase	
A282L	569	B368L	501	94%	M143L	605	49%	Serine/Threonine protein kinase	
A284L	279	B371L	279	86%	M101R	279	57%	Amidase	
A286R	378	B385R	372	75%	M097L	359	50%		
A287R	251	B389R	291	62%	M465L	287	51%	GIY-YIG endonuclease	
A289L	283	B388L	268	65%	M794R	270	33%	Serine/Threonine protein kinase	
A292L	328	B393L	322	78%	M091R	343	64%	Chitosanase	
A295L	317	B395L	320	84%				Fucose synthase	
A297R	333	B419L	183	81%	M100L	192	53%	Fructose-2,6 bisphosphatase	

PBCV-1		NY-2A			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
A298L	225	B421L	227	77%	M088R	213	42%	
A301L	241	B426L	210	52%	M086R	196	32%	
A304R	78	B429R	81	73%				
A305L	204	B430L	181	92%	M313L	185	67%	Dual specificity phosphatase
A306L	86	B431L	87	80%	M314L	85	69%	
A308L	86	B432L	103	74%	M315L	112	41%	
A310L	170	B434L	170	85%	M317L	168	53%	
A312L	238	B435L	248	90%	M341R	241	73%	
A313L	71	B437L	76	77%	M322L	84	42%	
A314R	80	B439R	105	73%	M323R	79	84%	
A315L	246	B433L	273	58%	M370R	267	57%	GIY-YIG endonuclease
A316R	225	B443R	430	79%	M822R	599	64%	
A318R	211	B443R	430	91%	M822R	599	56%	
A320R	139	B450R	115	70%	M326R	117	37%	
A321R	119	B451R	118	66%	M328R	113	39%	
A322L	176	B453L	178	67%	M329L	166	42%	
A324L	453	B455L	433	76%	M332L	422	54%	
A326L	209	B457L	199	74%	M335L	211	59%	NTP pyrophosphohydrolase
A328L	355	B458L	358	75%	M337L	352	39%	
A329R	96	B460R	108	68%				
A330R	432	B477R	443	96%	M413L	333	36%	Ankyrin Repeats
A331L	299	B480L	387	86%	M701R	344	70%	
A337L	78	B485L	238	64%	M344L	174	48%	
A339L	191	B485L	238	48%				
A341L	135	B487L	139	82%	M346L	139	26%	
A342L	576	B488L	551	89%	M348L	559	69%	
A348R	159	B494R	182	72%				
A349L	141	B496L	99	94%	M357L	127	73%	
A351L	358	B499L	345	72%	M111R	239	27%	GIY-YIG endonuclease
A352L	207	B503L	207	95%	M362L	213	73%	
A354R	237	B370L	349	60%	M069R	432	44%	HNH endonuclease
A357L	276	B506L	337	71%	M363L	341	42%	
A360R	241	B508R	1612	32%	M372R	1131	64%	
A361R	97	B508R	1612	40%	M372R	1131	72%	
A363R	811	B508R	1612	60%	M372R	1131	61%	Superfamily II helicase
A366L	255	B522L	252	64%	M844R	243	39%	Transposase (?)
A368L	443	B397L	473	27%				
A373R	152							
A375R	173	B292L	75	41%	M556R	174	63%	
A378L	249	B524L	255	81%	M797R	378	40%	
A379L	207	B525L	209	85%	M377L	204	44%	
A382R	157	B529R	485	73%	M381R	472	33%	
A383R	301	B529R	485	89%	M381R	472	53%	
A384L	456	B533L	532	42%	M591R	586	62%	
A390L	153	B533L	532	62%	M591R	586	48%	
A392R	258	B536R	254	87%	M586L	255	83%	ATPase (DNA packaging)
A394R	121	B539R	126	65%	M580L	190	44%	
A395R	82	B540R	82	81%	M577L	123	66%	

PBCV-1		NY-2A			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
A396L	70	B541L	150	69%	M820R	152	30%	
A397R	152							
A398L	118	B546L	118	84%	M572R	118	62%	
A399R	194	B547R	193	73%	M570L	164	51%	RNase H
A400R	118	B548R	121	81%	M569L	117	45%	
A401R	277	B549R	278	89%	M567L	290	64%	
A402R	227	B550R	232	85%	M564L	228	49%	
A403R	93	B552R	97	92%	M563L	88	52%	
A404R	132	B553R	187	84%	M560L	183	63%	
A405R	496	B554R	523	75%	M557L	541	29%	
A407L	210	B559L	210	86%	M495R	209	70%	
A408L	277	B560L	266	83%	M485R	231	50%	
A410L	110	B562L	125	90%	M478L	106	48%	
A411R	170				M479R	164	37%	
A412R	179	B564R	178	88%	M482R	179	49%	
A413L	244	B568L	236	71%	M484L	246	42%	
A414R	93	B570R	72	87%	M477R	71	58%	
A416R	188				M425R	186	50%	
A417L	429	B571L	434	78%	M430L	360	38%	Deoxyribonucleoside kinase
A420L	70	B573L	75	77%	M435L	114	65%	Replication factor C
A421R	98	B575R	98	73%	M436R	134	53%	
A422R	342	B165R	459	31%	M622L	349	31%	HNH endonuclease
A423R	157	B576R	152	71%	M440R	156	36%	
A424R	114							
A426R	114	B580R	116	79%	M444R	116	42%	6-phosphofructokinase (?)
A427L	119	B581L	119	76%	M445L	116	44%	Thioredoxin
A428L	145	B582L	76	62%				
A429L	476	B583L	453	77%	M451L	442	34%	Ankyrin Repeats
A430L	437	B585L	437	95%	M463L	433	79%	Major Capsid Protein
A431L	66							
A432R	176	B587R	155	74%	M472R	152	46%	
A435R	73	B764L	190	55%	M474L	109	52%	
A437L	103	B591L	103	74%	M476L	202	70%	DNA binding protein
A438L	78	B592L	78	79%	M423R	86	55%	Glutaredoxin
A439R	112	B593R	113	83%	M421L	110	62%	
A440L	88							
A441L	137	B595L	137	82%	M411L	141	53%	
A443R	308	B596R	308	85%	M416R	298	35%	
A444L	104	B597L	105	91%	M407R	113	40%	
A445L	462	B606L	462	91%	M404R	463	64%	ABC transporter protein
A448L	106	B611L	106	82%	M403R	106	56%	Protein disulfide isomerase
A449R	193	B612R	188	82%	M399L	183	54%	RNA triphosphatase
A450R	249	B687L	251	75%	M025R	246	32%	
A452L	79				M634R	91	38%	
A454L	289	B619L	287	88%	M680R	291	62%	
A456L	654	B623L	654	91%	M674R	653	65%	Helicase-Superfamily III
A462R	68							
A464R	275	B628R	268	88%	M672L	266	66%	RNase III

PBCV-1			NY-2A			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID		
A465R	118	B630R	117	72%	M670R	116	67%	Thiol oxidoreductase	
A467L	312	B631L	312	96%	M667R	307	62%		
A468R	443	B633R	443	82%	M664L	431	56%	Archaeo-eukaryotic primase	
A470R	203	B636R	200	79%	M660L	198	65%		
A471R	173	B638R	173	84%	M657L	172	52%		
A473L	517				M186R	520	74%	Glycosyltransferase	
A476R	324	B641R	324	89%	M653L	326	77%	Ribo. Reductase (small subunit)	
A478L	310	B805R	340	50%	M578R	208	43%	HNH endonuclease	
A480L	93	B644L	92	87%	M628L	91	55%		
A481L	224	B645L	224	85%	M629L	232	45%		
A482R	215	B647R	215	84%	M635R	215	43%	VLTF2-type transcription factor	
A484L	155	B650L	155	83%	M637L	155	54%		
A485R	148	B651R	144	83%	M626L	144	57%		
A486L	152				M379R	146	38%		
A488R	317	B652R	317	81%	M620L	320	65%		
A490L	310	B805R	340	47%	M578R	208	30%	HNH endonuclease	
A491R	76	B655R	76	77%	M618L	75	47%		
A492L	189	B656L	197	86%	M610L	173	37%		
A493L	191	B657L	186	86%	M610L	173	39%		
A494R	360	B663R	360	85%	M612R	357	57%		
A495R	221	B206L	279	54%	M032L	229	45%	GIY-YIG endonuclease	
A497R	146	B666R	139	77%	M615R	142	52%		
A500L	95	B667L	336	69%	M616L	328	40%		
A501L	75	B667L	336	92%	M616L	328	54%		
A502L	95	B670L	95	87%	M715R	107	54%		
A503L	304	B671L	279	83%	M713R	244	37%		
A505L	484	B672L	485	80%	M707R	446	56%		
A512R	804								
A517L	344	B697R	369	59%	M359L	342	63%	Cytosine methyltransferase	
A519L	82	B679L	90	90%	M282R	86	60%		
A520L	100	B680L	105	86%					
A521L	392	B685L	202	77%	M496L	205	64%		
A523R	171	B687R	168	83%	M497R	178	74%		
A526R	129	B692R	146	84%	M501R	148	52%	Glycerophosphodiesterase (?)	
A527R	105	B694R	95	88%	M504R	93	63%		
A530R	335	B697R	369	85%	M359L	342	65%	Cytosine methyltransferase	
A531L	67	B700L	65	75%	M506L	70	52%		
A532L	79	B701L	79	89%	M507L	79	60%		
A533R	374	B705R	530	90%	M508R	527	60%		
A535L	71	B710L	71	90%	M511L	72	52%		
A536L	73	B719L	81	72%	M512L	79	52%		
A537L	265	B723L	259	60%	M514L	248	31%		
A539R	173	B724R	173	80%	M517R	168	50%	GIY-YIG endonuclease	
A540L	1176	B725L	1191	59%	M518L	1189	33%	Autotransporter adhesin	
A543L	87	B725L	1191	79%	M518L	1189	41%		
A544R	298	B734R	323	81%				ATP-dependent DNA ligase	
A546L	321	B736L	405	81%				Glycosyltransferase	
A548L	458	B738L	458	85%	M272L	459	51%	SWI/SNF helicase	

PBCV-1			NY-2A			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID		
A551L	141	B741L	141	83%	M264L	140	74%	dUTP pyrophosphatase	
A552R	270	B743R	317	84%	M266R	273	42%	Transcription factor TFIID	
A554S/556S/57L	498	B744L	504	89%	M389R	499	52%	ATPase (PP-loop)	
A558L	400	B748L	400	93%	M269L	400	61%		
A559L	213	B751L	211	75%	M263R	218	35%		
A561L	649	B756L	649	90%	M258R	595	45%	ATPase (DNA repair)	
A564L	351	B408R	436	64%	M531L	365	44%		
A565R	576	B759R	675	89%	M253L	626	48%	ATPase (Chromosome segregation)	
A567L	152	B761L	145	64%	M249R	138	43%		
A568L	179	B762L	179	73%	M692L	225	29%		
A570L	128	B763L	126	88%	M694L	132	65%		
A571R	116	B765R	115	86%	M475R	116	74%		
A572R	181	B766R	181	86%	M695R	174	65%		
A574L	264	B767L	245	83%	M697L	276	44%	PCNA	
A575L	168	B771L	168	89%	M700L	167	35%		
A577L	106	B772L	133	83%	M553R	136	79%		
A579L	183							DNA restriction endonuclease	
A581R	265	B774R	345	86%				Adenine methyltransferase	
A583L	1061	B781L	1061	90%	M546R	1058	67%	DNA Topoisomerase II	
A590L	313								
A592R	69	B791R	69	94%					
A594R	146								
A596R	142	B795R	142	83%	M530L	144	66%	dCMP deaminase	
A598L	363	B796L	366	84%	M601L	359	61%	Histidine decarboxylase	
A601R	101	B800R	90	57%	M445R	93	64%		
A602L	128	B801L	134	73%	M605L	134	47%		
A603R	105	B802R	109	73%	M606R	105	43%		
A604L	134	B803L	115	40%				Zn metallopeptidase	
A605L	158	B810L	227	85%	M609L	149	43%		
A607R	229	B812R	391	82%	M717R	387	46%		
A608R	152	B812R	391	72%	M717R	387	33%		
A609L	389	B465R	379	25%	M719L	389	59%	UDP-glucose 6-dehydrogenase	
A612L	119	B813L	119	78%	M727L	119	57%	Histone H3, Lys 27 methylase	
A614L	577	B816L	492	44%	M729L	583	51%	Serine/Threonine protein kinase	
A617R	321	B818R	317	79%	M733R	316	48%	Serine/Threonine protein kinase	
A618L	131	B819L	117	77%	M735L	153	46%		
A619L	237	B820L	212	64%	M738L	224	25%		
A620L	83	B823L	83	89%	M740L	97	55%		
A621L	117	B824L	117	88%	M741L	116	52%		
A622L	520	B825L	520	94%	M748L	521	73%		
A623L	67	B826L	67	95%					
A624R	121	B828R	121	97%	M752R	119	52%		
A625R	433	B829R	432	100%				Transposase	
A627R	445	B831R	435	97%	M785L	433	55%		
A628L	97								
A629R	771	B832R	1103	90%	M777L	765	79%	Rib. Reductase (large subunit)	
A633R	120	B839R	120	87%	M775L	118	61%		
A634L	134	B840L	135	91%	M770R	134	58%		

PBCV-1		NY-2A			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
A635R	85	B841R	85	97%	M740L	97	40%	
A636R	96	B842R	71	54%				
A637R	141	B843R	141	95%				
A638R	359	B844R	359	96%	M766L	363	62%	Agmatine iminohydrolase
A642R	74	B847R	486	100%	M763L	480	63%	
A643R	269	B847R	486	78%	M763L	480	79%	
A644R	87	B849R	171	80%	M762L	173	73%	
A645R	123	B852R	123	81%	M136L	119	40%	
A646L	141				M375R	179	42%	Translation elongation factor-3
A647R	189	B118R	280	53%	M525L	253	53%	
A649R	259	B120R	244	73%	M525L	253	35%	
A651L	230	B206L	279	60%	M502L	245	53%	Homing endonuclease G1Y-Y1G
A654L	197	B853L	197	86%	M758R	197	51%	N-acetyltransferase
A656L	120	B857L	136	71%	M757R	144	38%	
A658R	95							
A659L	180	B858L	184	81%	M756R	183	58%	
A662L	171	B859L	171	77%	M179L	161	64%	
A664L	142	B860L	154	58%	M149R	157	51%	
A665L	171	B862L	151	86%	M149R	157	40%	
A666L	918				M742R	901	65%	Translation elongation factor-3
A672R	211	B018L	472	35%	M413L	333	34%	
A674R	216	B265R	216	87%	M034L	214	60%	Thymidylate synthase X
A676R	375	B869R	364	83%	M160R	429	60%	
A678R	385	B480L	387	34%	M701R	344	39%	
A682L	368	B874L	366	70%	M833L	268	38%	
A683L	367	B769R	362	64%	M359L	342	42%	Cytosine methyltransferase
A686L	162							
A687R	75	B235L	103	72%	M826L	164	44%	
A689L	141	B883L	102	57%				
A691R	183							
PBCV-1		NY-2A			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
a1L	89							
a4L	78	B882R	82	68%				
a6L	72	B875R	157	34%				
a12R	97							
a13L	113							
a15L	182							
a16L	230							
a17L	66							
a19R	216							
a20R	84							
a21R	155							
a22R	109							
a23R	74							
a24R	120							
a26R	198	B050R	144	32%	M072L	164	29%	
a27L	116	B034L	1485	29%	M061R	1467	31%	

PBCV-1		NY-2A			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
a28R	89							
a30R	178							
a31R	91							
a32R	192							
a33R	79							
a36R	113							
a38R	79							
a40L	129	B069L	136	77%				
a42L	88							
a43R	75	B071R	75	93%	M558R	216	52%	
a45R	70							
a46R	166							
a47L	67							
a52R	83	B079R	121	90%				
a54L	107							
a55L	131							
a56L	104							
a58L	162							
a59L	74							
a62R	123							
a65L	81							
a66L	112							
a68L	100							
a69L	146	B102L	158	60%				
a70L	88							
a72L	90							
a73L	69							
a74L	67							
a76L	103	B114L	65	72%				
a80L	135	B868R	214	49%				
a82R	71							
a83R	90							
a86L	80	B128L	65	80%				
a91L	126	B136L	423	87%	M631L	365	53%	
a95R	72							
a96R	67							
a97R	71							
a99L	80							
a101L	87	B144L	80	65%				
a102L	168	B147L	114	64%				
a104L	96	B149L	129	43%	M134L	246	28%	
a106R	132	B151R	107	48%				
a108R	71	B155R	71	42%				
a110L	102							
a112L	97	B160L	73	97%				
a113L	94							
a115L	76							
a116R	81	B161R	99	74%				

PBCV-1		NY-2A			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
a117L	130	B162L	143	63%				
a119L	133							
a120L	90				M151L	67	50%	
a124L	84	B174L	101	73%	M164L	65	39%	
a126R	66							
a128L	86	B178L	69	81%				
a132R	90	B182R	81	44%				
a136R	146	B187R	145	81%	M122R	141	38%	
a142L	133							
a143L	67							
a144L	498	B194L	337	50%				
a146L	84							
a147L	260	B195L	76	70%				
a149L	136							
a152L	90	B202L	132	73%				
a155R	94	B410L	96	63%				
a156L	110	B407R	319	47%				
a159R	110							
a160L	164							
a164L	134							
a167L	67							
a170L	89							
a172L	106	B072L	85	50%				
a174L	65							
a176L	77	B136L	423	35%	M143L	605	33%	
a178L	81	B868R	214	36%				
a179L	109							
a183L	76							
a184L	116							
a186L	84							
a187L	247							
a188L	77	B252L	68	83%				
a190L	72							
a194R	82	B263R	67	89%	M013L	84	49%	
a195R	82							
a197R	95	B266R	65	77%				
a198R	74	B265R	72	47%				
a204L	79	B275L	70	91%	M311R	75	65%	
a206L	85							
a209R	114							
a210L	163							
a211R	107							
a212R	68							
a216R	107				M291L	102	72%	
a218L	121							
a220L	129							
a223aL	92				M724R	75	62%	
a223R	70							

PBCV-1		NY-2A			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
a224L	88				M723R	84	60%	
a225L	90							
a228R	76				M245L	77	52%	
a232R	87	B300R	69	58%				
a235R	83	B304R	79	46%				
a236L	87							
a238L	114				M234R	66	66%	
a240L	65							
a242L	117	B317L	117	80%				
a244L	77							
a246aR	102							
a249L	102	B333L	73	56%	M222R	108	44%	
a251aL	141	B238R	68	61%				
a252aL	84	B343L	84	67%				
a255R	65							
a258R	83							
a259aR	80							
a259R	66							
a260aR	65							
a264R	67	B349R	67	68%				
a266R	83	B352R	76	47%				
a268R	71							
a269R	68							
a270R	75	B355R	84	75%				
a272R	67	B356R	201	52%				
a276L	94	B868R	214	57%				
a279R	72	B366R	72	97%				
a280R	93	B367R	93	94%	M147R	67	36%	
a281R	239	B364R	86	59%	M387R	297	57%	
a283L	99	B369L	67	92%				
a285R	71	B373R	94	90%				
a288L	88	B347R	69	77%				
a290R	135							
a291R	89							
a293R	168	B394R	138	71%				
a294R	74							
a299R	172				M089L	208	41%	
a300R	76	B423R	76	42%				
a302R	140							
a303L	112	B428L	69	69%				
a307aR	80							
a307R	78							
a309L	78	B432L	103	71%				
a311R	83							
a312aR	70							
a317L	154							
a319L	84				M825L	73	48%	
a323R	132				M330R	124	38%	

PBCV-1		NY-2A			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
a325R	74							
a327R	108							
a331L	67	B479L	67	98%				
a332L	109	B480L	387	96%	M701R	344	74%	
a334R	110	B482R	114	93%				
a335L	74							
a336R	85							
a338R	89	B486R	71	45%				
a340R	83							
a341R	65							
a343R	86							
a344R	115	B489R	148	62%				
a345L	96	B490L	80	64%				
a346L	75							
a347L	70	B495L	70	52%				
a350R	122	B498R	122	69%	M358R	157	50%	
a353R	68							
a355L	72							
a356R	107							
a358R	77							
a359L	86	B506L	337	78%				
a362L	81							
a364L	67							
a365L	96							
a367R	91							
a369R	80				M288L	352	31%	
a370R	82							
a371R	69							
a372L	76							
a374L	87							
a376R	84							
a377R	91	B523R	100	73%				
a380R	101	B526R	177	71%				
a381R	85	B527R	68	89%				
a385L	76							
a386R	259	B532R	225	81%				
a387R	175							
a388R	78	B534R	71	75%				
a389R	102	B534R	71	48%	M594L	158	54%	
a391R	95	B535R	84	54%				
a393L	65							
a406L	66							
a409R	82	B561R	122	80%				
a415L	69	B569L	132	80%				
a418R	79							
a419R	70							
a425L	67							
a433R	87	B589R	66	40%				

PBCV-1		NY-2A			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
a434L	72							
a436L	69	B765R	115	73%	M475R	116	65%	
a442R	131							
a446R	101	B607R	68	88%	M405L	105	60%	
a447R	99	B609R	65	95%				
a451L	101	B868R	214	53%				
a453R	89	B620R	75	64%				
a455R	147	B621R	123	79%				
a457R	76	B624R	69	####	M677L	210	60%	
a458L	83	B627L	69	60%				
a459R	100				M677L	210	72%	
a460R	78				M675L	169	31%	
a463L	192							
a466L	91				M671L	92	52%	
a469L	76	B634L	66	69%				
a472L	90	B639L	85	76%	M658R	93	42%	
a474R	105							
a475R	90							
a477L	73	B640L	139	60%				
a479R	108	B658L	65	40%				
a483L	71							
a487R	65							
a489R	122							
a496L	70							
a498L	124	B665L	144	58%	M614L	103	64%	
a499L	77							
a504R	96							
a506R	70				M711L	127	62%	
a507R	186	B673R	71	55%	M706L	467	34%	
a508R	77	B675R	159	80%				
a509R	101							
a510R	67	B676R	92	87%				
a511L	73							
a513R	120							
a514L	144							
a515L	203							
a516R	68							
a518R	100	B695L	191	62%	M361R	71	76%	
a522R	90							
a524L	107							
a525R	66							
a528R	94							
a529L	72	B695L	191	80%	M361R	71	64%	
a534R	105	B705R	530	95%	M508R	527	76%	
a538L	65							
a541R	96	B728R	76	45%	M527R	85	48%	
a542R	72							
a545L	73							

PBCV-1		NY-2A			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
a547R	77							
a549R	81	B739R	93	41%				
a550R	136	B740R	78	44%				
a553L	71							
a555R	113							
a560R	104	B752R	142	72%				
a562R	65							
a563L	82	B758L	83	67%				
a566L	90	B760L	136	96%				
a569R	84							
a573L	69							
a576R	73							
a578L	67							
a580R	73							
a582L	78							
a584R	98	B783R	98	91%	M551L	70	58%	
a585R	79	B784R	222	82%	M550L	330	37%	
a586R	75	B785R	100	85%	M549L	96	60%	
a587R	114	B786R	65	81%				
a588R	116	B788R	212	50%				
a589L	76	B790L	76	52%				
a591L	69							
a593R	132	B792R	248	60%	M599R	287	49%	
a595L	84							
a597L	98							
a599R	179	B797R	184	72%				
a600R	82							
a606L	114	B810L	227	59%				
a610R	91							
a611R	74							
a613R	87	B815R	77	57%	M728R	67	32%	
a615R	66							
a616R	93	B817R	91	63%				
a626L	74	B830L	74	100%				
a630R	79	B833R	74	74%				
a631L	89				M781R	125	41%	
a632L	72							
a639L	132							
a640R	79	B845R	79	96%				
a641L	85	B846L	74	95%				
a648L	69	B340R	72	50%				
a650L	87							
a652R	75							
a653R	65							
a655L	106	B855L	71	52%				
a657L	82							
a660R	110							
a661R	102							

PBCV-1		NY-2A			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
a663R	79							
a667R	96							
a668R	65							
a669R	83							
a670R	122					M744L	93	43%
a671L	97							
a673L	68							
a675L	122							
a677L	76	B870L	133	84%				
a679L	94							
a680R	87							
a681R	74							
a684R	106							
a685R	80	B695L	191	54%				
a688L	71							
a690R	78	B882R	82	68%				
a692R	89							

NY-2A			PBCV-1			MT325			Description
ORF	AA		ORF	AA	%ID	ORF	AA	%ID	
B1L	93								
B3R	78								
B4R	102	A689L	141	61%					
B6L	192								
B8R	331	A517L	344	32%	M359L	342	31%	Cytosine methyltransferase	
B10R	303								Adenine methyltransferase
B13L	342								
B16L	368								Adenine methyltransferase
B18L	472	A682L	368	35%	M833L	268	31%		
B22R	401	A354R	237	41%	M069R	432	44%	HNH endonuclease	
B24L	532	A682L	368	31%	M833L	268	31%		
B27L	311	A67R	309	36%	M686L	411	43%		
B31R	231				M690R	230	54%		
B33R	254	A9R	175	82%	M247R	250	64%		
B34L	1485	A122R	1032	22%	M071R	1497	64%		
B39L	418				M288L	352	42%	GIY-YIG endonuclease	
B40L	1471	A122R	1032	29%	M061R	1467	61%		
B47L	1453	A122R	1032	32%	M061R	1467	61%		
B55L	953	A122R	1032	36%	M042R	956	76%		
B59R	411	A430L	437	43%	M078R	410	86%		
B61R	399	A354R	237	40%	M069R	432	40%	HNH endonuclease	
B68L	153	A39L	151	63%	M807R	155	31%	SKP-1 protein	
B70R	433	A41R	412	93%	M803L	500	67%	Periplasmic protein TonB	
B73L	603	A44L	599	89%				ATPase (AAA+ class)	
B74R	123	A48R	123	90%	M170R	122	37%		
B75L	228	A49L	219	94%				Glycerophosphoryl diesterase	
B76L	116	A50L	141	88%	M627L	144	69%	Pyrimidine dimer-specific glycosylase	
B77R	97								
B78L	153	A51L	203	91%					
B80L	645	A625R	433	32%				Transposase	
B83L	228							Resolvase	
B86L	334	A546L	321	36%					
B87L	134								
B88L	343	A517L	344	50%	M359L	342	49%	Cytosine methyltransferase	
B91L	210								
B94R	253								
B99R	312	A67R	309	85%	M686L	411	32%		
B103R	353	A71R	354	82%					
B104R	314	A564L	351	73%	M531L	365	45%		
B108L	67								
B109R	75								
B110L	177								
B113R	115								
B115L	85	A77L	91	70%	M810R	83	40%		
B116R	298	A78R	298	97%	M103L	296	67%	N-carbamoylput. amidohydrolase	
B117L	164	a274R	263	32%					
B118R	280	A647R	189	46%	M525L	253	57%		
B120R	244	A275R	252	77%	M025R	246	34%		

NY-2A		PBCV-1			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
B122L	192	A81L	189	71%	M005L	180	43%	
B124L	187	A84L	188	82%	M007L	170	33%	
B126R	230	A85R	242	82%	M401R	224	50%	Prolyl-4-hydroxylase
B130R	161	A138R	273	24%	M278R	267	50%	
B133R	444	A87R	456	65%	M622L	349	28%	HNH endonuclease
B136L	423	A92/93L	359	94%	M631L	365	44%	
B137L	320	A94L	364	32%	M124L	348	29%	β -1,3-glucanase
B139R	508	A98R	568	24%				Chitin synthase
B143R	595	A100R	595	87%	M037R	593	59%	Glucosamine synthetase
B148R	321	A103R	330	76%	M133R	319	53%	mRNA guanylyltransferase
B150L	290	A105L	284	68%	M137L	279	43%	Ubiquitin C-terminal hydrolase
B154L	338	A107L	290	66%	M139L	312	36%	Transcription factor TFIIB
B157L	163	A109L	108	84%	M140L	168	59%	
B159R	867	A114R	485	87%	M467R	849	62%	Fucosyl-Glycosyltransferase
B163R	343	A118R	345	88%				GDP-D-mannose dehydratase
B165R	459	A87R	456	62%	M622L	349	31%	HNH endonuclease
B168R	104	A121R	97	75%	M150R	104	70%	
B170R	1386	A122R	1032	57%	M152R	1631	50%	
B173L	414	A422R	342	27%	M093L	377	26%	HNH endonuclease
B175L	192	A125L	180	91%	M163L	180	68%	Transcription factor TFIIS
B177R	239	A127R	245	84%	M166R	220	61%	
B179L	373	A495R	221	33%	M288L	352	29%	
B181L	135	A131L	136	70%	M367L	219	29%	
B183L	224							Methyltransferase
B185L	164	A134L	165	89%	M180L	158	46%	GIY-YIG endonuclease
B187R	145	A136R	146	81%	M122R	141	34%	
B188R	97	A137R	72	66%				
B190L	100	A139L	108	76%	M184L	113	56%	
B192R	1075	A140/A145R	1130	68%	M190R	1069	42%	
B197L	131	A150L	152	80%	M200L	103	41%	
B199R	343	A354R	237	36%	M093L	377	38%	HNH endonuclease
B201R	122	A151R	135	94%				
B203R	460	A153R	459	81%	M201R	454	64%	Superfamily II helicase
B206L	279	A651L	230	47%	M502L	245	48%	GIY-YIG endonuclease
B207L	111	A157L	110	78%	M204L	117	56%	
B208L	95	A158L	104	55%				
B209R	103	A161R	123	59%	M209R	100	32%	
B211L	112	A165L	279	74%	M210L	95	36%	
B212L	148	A165L	279	75%	M213L	152	39%	
B214R	268	A166R	268	79%	M215R	279	55%	Exonuclease
B215R	167	A168R	166	85%	M218R	161	50%	
B218R	400	A354R	237	39%	M069R	432	41%	HNH endonuclease
B222R	321	A169R	323	85%				Aspartate transcarbamylase
B224R	368	A171R	387	85%	M803L	500	65%	
B226L	279	A173L	288	90%	M219L	274	58%	Patatin-like phospholipase
B230L	382							Adenine methyltransferase
B235L	103	A687R	75	71%	M826L	164	40%	
B236L	324	A251R	326	81%				Adenine methyltransferase

NY-2A		PBCV-1			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
B239R	807	A181/A182R	830	84%	M791R	520	40%	Chitinase
B246R	202	A651L	230	55%	M502L	245	54%	GIY-YIG endonuclease
B249/253R	760	A185R	913	86%	M019L	907	69%	6 DNA polymerase
B255R	86				M288L	352	32%	
B256R	260	A495R	221	31%	M288L	352	30%	
B258R	1278	A189/192R	1299	71%	M015L	1286	36%	SW/SNF chromatin remodeling complex
B261L	262	A193L	262	97%	M014R	264	68%	PCNA
B264L	152	A196L	152	86%	M012R	156	56%	
B267R	100	A199R	101	84%	M011L	96	48%	
B268L	190							Histone H3, Lys 27 methylase
B271R	145	A200R	118	75%	M010L	119	66%	Cytosine deaminase
B272L	100	A201L	94	72%	M084L	97	72%	
B273L	112	A202L	113	85%	M312R	112	53%	
B274R	216	A203R	216	92%	M310L	220	68%	
B277R	206	A205R	206	74%	M308L	235	68%	
B278R	372	A207R	372	86%	M307L	372	61%	Ornithine/Arginine decarboxylase
B283R	223	A208R	311	50%				
B284L	148	A213L	148	80%	M299R	149	50%	
B285L	133	A214L	135	79%	M298R	157	46%	
B286R	363	A495R	221	35%	M288L	352	31%	GIY-YIG endonuclease
B288L	306	A215L	321	77%	M289R	368	66%	β & α , 1,4 linked glucuronic lyase
B289L	382	A217L	394	88%	M283R	386	72%	Monoamine oxidase
B291L	138	A227L	137	78%	M244R	130	55%	
B292L	75	A229L	77	68%	M556R	174	38%	
B296L	78							
B297R	364	A495R	221	37%	M288L	352	30%	
B298L	372	A231L	383	79%	M457R	376	64%	
B302R	111	A233R	112	73%	M240L	90	47%	
B303L	108	A234L	108	66%	M238R	119	46%	
B305R	507	A237R	518	94%	M233L	502	69%	Homospermidine synthase
B310R	382	A564L	351	54%	M427R	352	44%	
B314L	145	A239L	111	54%	M122R	141	32%	
B316R	725	A241R	725	92%	M225L	715	50%	Superfamily II helicase
B322R	309	A246R	288	64%	M803L	500	69%	Pathogenesis-related protein
B323R	380	A247R	383	61%	M413L	333	32%	
B324L	495	A87R	456	61%	M622L	349	29%	HNH endonuclease
B331R	309	A248R	308	74%	M794R	270	49%	Serine/Threonine protein kinase
B334R	432	A625R	433	85%				Transposase
B336R	94	A250R	94	90%	M183R	95	46%	Potassium channel protein
B339L	243	A275R	252	81%	M025R	246	35%	
B341R	76							
B342R	166	A253R	150	82%	M109R	149	34%	
B344R	203	A261R	136	93%				
B346L	249	A287R	251	56%	M502L	245	46%	GIY-YIG endonuclease
B348L	215	A262L	123	83%	M114L	226	50%	
B349R	67	a264R	67	68%				
B350L	249	A265L	170	88%	M116L	255	65%	
B354L	264	A271L	159	80%				Lysophospholipase

NY-2A			PBCV-1			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID		
B357L	139	A273L	138	86%					
B358R	225	a274R	263	65%					
B359R	269							Adenine methyltransferase	
B361R	550							Adenine methyltransferase	
B365L	605	A278L	610	97%	M143L	605	43%	Serine/Threonine protein kinase	
B368L	501	A282L	569	93%	M143L	605	65%	Serine/Threonine protein kinase	
B370L	349	A354R	237	58%	M093L	377	46%	HNH endonuclease	
B371L	279	A284L	279	86%	M101R	279	58%	Choloylglycine hydrolase	
B377R	70								
B378L	645	A625R	433	32%				Transposase	
B381L	228							Resolvase	
B385R	372	A286R	378	75%	M097L	359	49%		
B388L	268	A289L	283	69%	M794R	270	35%	Serine/Threonine protein kinase	
B389R	291	A287R	251	61%	M465L	287	51%	GIY-YIG endonuclease	
B393L	322	A292L	328	80%	M091R	343	64%	Chitosanase	
B395L	320	A295L	317	84%				Fucose synthase	
B397L	473	A368L	443	27%					
B399R	260	A581R	265	34%				Adenine methyltransferase	
B401L	160								
B402L	271				M093L	377	30%		
B403R	326	A328L	355	30%	M337L	352	28%		
B404R	132	A296R	156	77%	M090L	165	55%		
B408R	436	A564L	351	61%	M427R	352	35%		
B411L	370	A517L	344	44%	M359L	342	44%	Cytosine methyltransferase	
B416R	372							Adenine methyltransferase	
B418R	384							Adenine methyltransferase	
B419L	183	A297L	177	81%	M100L	192	56%	Fructose-2,6 bisphosphatase	
B421L	227	A298L	225	77%	M088R	213	39%		
B424L	351	A354R	237	53%	M093L	377	44%	HNH endonuclease	
B426L	210	A301L	241	62%	M086R	196	36%		
B429R	81	A304R	78	73%					
B430L	181	A305L	204	92%	M313L	185	68%	Dual specificity phosphatase	
B431L	87	A306L	86	86%	M314L	85	75%		
B432L	103	A308L	86	72%	M315L	112	39%		
B433L	273	A315L	246	52%	M465L	287	52%	GIY-YIG endonuclease	
B434L	170	A310L	170	85%	M317L	168	55%		
B435L	248	A312L	238	90%	M319L	241	71%		
B437L	76	A313L	71	77%	M322L	84	38%		
B438R	68								
B440L	272	A315L	246	50%	M465L	287	53%	GIY-YIG endonuclease	
B441R	65	A89R	68	92%					
B443R	430	A318R	211	91%	M822R	599	59%	Periplasmic protein TonB	
B446R	412	A354R	237	37%	M093L	377	42%	HNH endonuclease	
B450R	115	A320R	139	70%	M326R	117	30%		
B451R	118	A321R	119	65%	M328R	113	33%		
B453L	178	A322L	176	66%	M329L	166	43%		
B455L	433	A324L	453	81%	M332L	422	55%		
B457L	199	A326L	209	77%	M335L	211	61%	NTP pyrophosphohydrolase	

NY-2A			PBCV-1			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID		
B458L	358	A328L	355	75%	M337L	352	38%		
B460R	108	A329R	96	68%					
B461R	66								
B462R	69								
B463L	68								
B465R	379	A609L	389	27%	M719L	389	29%	UDP-glucose 6-dehydrogenase	
B468R	251	A215L	321	36%	M289R	368	35%	β & α 1,4 linked glucuronic lyase	
B469L	403							Polysaccharide deacetylase	
B472R	527	A98R	568	21%				Chitin synthase	
B477R	443	A330R	432	94%	M413L	333	36%		
B480L	387	A333L	299	85%	M701R	344	66%		
B485L	238	A339L	191	48%	M344L	174	38%		
B487L	139	A341L	135	82%	M346L	139	26%		
B488L	551	A342L	576	90%	M348L	559	70%		
B492R	111								
B494R	182	A348R	159	71%					
B496L	99	A349L	141	93%	M357L	127	65%		
B497L	251	A351L	358	46%	M111R	239	30%	GIY-YIG endonuclease	
B498R	122	a350R	122	69%	M358R	157	50%		
B499L	345	A351L	358	72%	M111R	239	28%	GIY-YIG endonuclease	
B503L	207	A352L	207	95%	M362L	213	74%		
B506L	337	A357L	276	70%	M363L	341	39%		
B508R	1612	A363R	811	62%	M372R	1131	32%	Superfamily II helicase	
B519L	132	A366L	255	38%					
B522L	252	A366L	255	64%	M844R	243	35%	Transposase	
B524L	255	A378L	249	83%	M797R	378	42%		
B525L	209	A379L	207	86%	M377L	204	47%		
B529R	485	A383R	301	89%	M381R	472	52%		
B533L	532	A384L	456	42%	M591R	586	64%		
B536R	254	A392R	258	87%	M586L	255	79%	ATPase (DNA packaging)	
B537R	68								
B539R	126	A394R	121	65%	M580L	190	43%		
B540R	82	A395R	82	81%	M577L	123	69%		
B541L	150	A396L	70	71%	M820R	152	31%		
B542R	298								
B543L	269							Adenine methyltransferase	
B546L	118	A398L	118	84%	M572R	118	59%		
B547R	193	A399R	194	73%	M570L	164	48%	RNase H	
B548R	121	A400R	118	81%	M569L	117	46%		
B549R	278	A401R	277	89%	M567L	290	64%		
B550R	232	A402R	227	85%	M564L	228	48%	Lipoprotein lipase	
B552R	97	A403R	93	92%	M563L	88	51%		
B553R	187	A404R	132	87%	M560L	183	63%		
B554R	523	A405R	496	77%	M557L	541	40%		
B556R	374	A495R	221	33%	M288L	352	27%		
B559L	210	A407L	210	86%	M495R	209	71%		
B560L	266	A408L	277	86%	M485R	231	51%		
B562L	125	A410L	110	90%	M478L	106	49%		

NY-2A		PBCV-1			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
B564R	178	A412R	179	87%	M482R	179	54%	
B566R	349	A517L	344	32%	M359L	342	32%	Cytosine methyltransferase
B567L	371							Adenine methyltransferase
B568L	236	A413L	244	73%	M484L	246	44%	
B570R	72	A414R	93	87%	M477R	71	59%	
B571L	434	A417L	429	78%	M430L	360	37%	Replication factor C
B573L	75	A420L	70	77%	M435L	114	61%	
B575R	98	A421R	98	63%	M436R	134	43%	
B576R	152	A423R	157	73%	M440R	156	36%	
B579R	84							
B580R	116	A426R	114	79%	M444R	116	42%	6-phosphofructokinase (?)
B581L	119	A427L	119	75%	M445L	116	39%	Thioredoxin
B582L	76	A428L	145	78%				
B583L	453	A429L	476	77%	M451L	442	33%	
B585L	437	A430L	437	95%	M463L	433	79%	
B587R	155	A432R	176	74%	M472R	152	43%	
B590R	70							
B591L	103	A437L	103	75%	M476L	202	70%	DNA binding protein
B592L	78	A438L	78	79%	M423R	86	54%	Glutaredoxin
B593R	113	A439R	112	83%	M421L	110	59%	
B595L	137	A441L	137	82%	M411L	141	52%	
B596R	308	A443R	308	85%	M416R	298	34%	
B597L	105	A444L	104	89%	M407R	113	40%	
B598L	399	A354R	237	38%	M093L	377	39%	HNH endonuclease
B602L	214	A315L	246	61%	M032L	229	60%	GIY-YIG endonuclease
B606L	462	A445L	462	91%	M404R	463	65%	ABC transporter protein
B611L	106	A448L	106	82%	M403R	106	48%	Protein disulfide isomerase
B612R	188	A449R	193	82%	M399L	183	51%	RNA triphosphatase
B614R	246							
B617L	437	A430L	437	93%	M463L	433	79%	
B618R	270							Glycosyltransferase
B619L	287	A454L	289	89%	M680R	291	61%	
B623L	654	A456L	654	91%	M674R	653	64%	Helicase-Superfamily III
B628R	268	A464R	275	88%	M672L	266	65%	RNase III
B629R	287	A495R	221	32%	M111R	238	35%	GIY-YIG endonuclease
B630R	117	A465R	118	73%	M670R	116	60%	Thiol oxidoreductase
B631L	312	A467L	312	96%	M667R	307	63%	
B633R	443	A468R	443	81%	M664L	431	55%	Archaeo-eukaryotic primase
B636R	200	A470R	203	84%	M660L	198	68%	
B638R	173	A471R	173	84%	M657L	172	51%	
B641R	324	A476R	324	89%	M653L	326	74%	Ribo. Reductase (small subunit)
B644L	92	A480L	93	87%	M628L	91	50%	
B645L	224	A481L	224	85%	M629L	232	44%	
B647R	215	A482R	215	86%	M635R	215	45%	VLTF2-type transcription factor
B650L	155	A484L	155	83%	M637L	155	54%	
B651R	144	A485R	148	83%	M626L	144	56%	
B652R	317	A488R	317	80%	M620L	320	60%	
B655R	76	A491R	76	77%	M618L	75	45%	

NY-2A		PBCV-1			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
B656L	197	A492L	189	86%	M610L	173	38%	
B657L	186	A493L	191	87%	M610L	173	36%	
B659L	76							
B662R	142	A478L	310	37%	M578R	208	36%	
B663R	360	A494R	360	84%	M612R	357	59%	
B666R	139	A497R	146	65%	M615R	142	45%	
B667L	336	A501L	75	92%	M616L	328	44%	Periplasmic protein TonB
B670L	95	A502L	95	89%	M715R	107	65%	
B671L	279	A503L	304	83%	M713R	244	34%	
B672L	485	A505L	484	81%	M707R	446	54%	
B677R	129							
B680L	105	A520L	100	86%				
B681L	350	A517L	344	47%	M359L	342	47%	Cytosine methyltransferase
B684L	181	A521L	392	79%	M704R	176	55%	
B685L	202	A521L	392	76%	M496L	205	66%	
B687R	168	A523R	171	83%	M497R	178	73%	
B688L	321	A564L	351	72%	M427R	352	44%	
B692R	146	A526R	129	84%	M501R	148	56%	Glycerophosphodiesterase (?)
B694R	95	A527R	105	85%	M504R	93	54%	
B696R	75	a528R	94	50%				
B697R	369	A530R	335	85%	M359L	342	66%	Cytosine methyltransferase
B699L	78							Ubiquitin
B700L	65	A531L	67	76%	M506L	70	38%	
B701L	79	A532L	79	89%	M507L	79	68%	
B702L	432	A625R	433	86%				Transposase
B704L	69							
B705R	530	A533R	374	90%	M508R	527	65%	
B710L	71	A535L	71	90%	M511L	72	52%	
B711L	395	A625R	433	36%				Transposase
B714L	266	A625R	433	42%				
B715L	228							Resolvase
B718L	320	A354R	237	36%	M093L	377	35%	HNH endonuclease
B719L	81	A536L	73	66%	M512L	79	63%	
B720L	248	A366L	255	61%	M844R	243	36%	
B723L	259	A537L	265	60%	M514L	248	26%	
B724R	173	A539R	173	80%	M517R	168	50%	GIY-YIG endonuclease
B725L	1191	A540L	1176	61%	M526L	1098	38%	Autotransporter adhesin
B734R	323	A544R	298	81%				ATP-dependent DNA ligase
B736L	405	A546L	321	81%				Glycosyltransferase
B738L	458	A548L	458	85%	M272L	459	51%	SWI/SNF helicase
B741L	141	A551L	141	83%	M264L	140	74%	dUTP pyrophosphatase
B743R	317	A552R	270	84%	M266R	273	38%	Transcription factor TFIID
B744L	504	A554S/556S/57L	498	89%	M389R	499	52%	ATPase (PP-loop)
B747L	309	A267L	314	42%	M578R	208	29%	HNH endonuclease
B748L	400	A558L	400	93%	M269L	400	62%	
B751L	211	A559L	213	76%	M263R	218	31%	
B753L	380	A267L	314	32%	M578R	208	33%	HNH endonuclease
B756L	649	A561L	649	100%	M258R	595	46%	ATPase (DNA repair)

NY-2A		PBCV-1			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
b49R	65				M045L	99	46%	
b50R	144	a26R	198	38%	M045L	99	46%	
b51R	176	a26R	198	29%	M059L	114	47%	
b52R	100				M050L	100	70%	
b53R	81				M056L	248	63%	
b54R	135				M062L	146	68%	
b56R	124				M045L	99	62%	
b57R	66				M045L	99	42%	
b58R	134	a26R	198	32%	M044L	101	60%	
b60L	90							
b62L	90							
b63L	84							
b64L	78							
b65R	78							
b66L	79				M070L	118	44%	
b67L	119	A37L	105	77%				
b69L	136	a40L	129	77%				
b71R	75	a43R	75	93%				
b72L	85	a172L	106	51%				
b79R	121	a52R	83	90%				
b81R	86							
b82L	67							
b84R	93							
b85R	114							
b89R	85							
b90R	81	a518R	100	62%	M361R	71	60%	
b92R	101							
b93R	67							
b95L	73							
b96L	72							
b97L	103							
b98R	90							
b100L	136							
b101L	99							
b102L	158	a69L	146	60%				
b105L	229	a155R	94	60%				
b106R	89	a156L	110	36%				
b107L	288	A75L	280	87%	M009L	269	40%	
b111R	75							
b112R	74							
b114L	65	a76L	103	81%				
b119L	69							
b121L	129	a80L	135	56%				
b123R	101							
b125R	85							
b127L	81							
b128L	65	a86L	80	85%				
b129L	67							

NY-2A		PBCV-1			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
b131R	72							
b132L	78							
b134L	103							
b135L	66							
b138R	81							
b140L	65							
b141L	108							
b142L	122							
b144L	80	a101L	87	65%				
b145L	107				M040L	118	34%	
b146L	107							
b147L	114	a102L	168	64%				
b149L	129	a104L	96	43%	M134L	246	23%	
b151R	107	a106R	132	48%				
b152R	77							
b153R	65							
b155R	71	a108R	71	42%				
b156R	67							
b158R	123							
b160L	73	a112L	97	97%				
b161R	99	a116R	81	74%				
b162L	143	a117L	130	63%				
b164L	146							
b166L	68							
b167R	65							
b169L	74							
b171L	68							
b172L	67							
b174L	101	a124L	84	73%	M164L	65	44%	
b176R	74							
b178L	69	a128L	86	81%				
b180L	66							
b182R	81	a132R	90	44%				
b184R	247	A133R	207	95%				
b186R	83							
b189L	72							
b191L	81							
b193L	217	a144L	498	43%				
b194L	337	a144L	498	47%				
b195L	76	a147L	260	71%				
b196L	113							
b198R	75							
b200L	81							
b202L	132	a152L	90	70%				
b204L	70							
b205R	72							
b210R	65							
b213R	80							

NY-2A		PBCV-1			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
b216L	87							
b217R	73							
b219L	152							
b220L	149							
b221L	71							
b223L	94							
b225R	95	a43R	75	53%				
b227L	84							
b228R	72				M220R	105	58%	
b229L	93							
b231R	115							
b232R	68							
b233R	70							
b234R	93							
b237R	83							
b238R	68	a251aL	141	71%				
b240L	72							
b241L	131	a183L	76	48%	M793L	137	71%	
b242L	65							
b243R	78							
b244L	90							
b245L	67				M793L	137	50%	
b247L	71							
b248L	68				M022R	75	45%	
b250R	101							
b251L	79							
b252L	68	a188L	77	83%				
b254L	96							
b257R	137							
b259L	75							
b260R	67							
b262R	135							
b263R	67	a194R	82	89%	M013L	84	49%	
b265R	72	a198R	74	47%				
b266R	65	a197R	95	77%				
b269R	113							
b270L	89							
b275L	70	a204L	79	91%	M311R	75	69%	
b276L	91							
b279L	88							
b280L	99							
b281L	86							
b282L	80							
b287L	90							
b290R	88				M285L	88	71%	
b293R	213	A230R	196	79%	M242L	190	40%	
b294R	100							
b295L	65							

NY-2A		PBCV-1			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
b299R	81					M459L	91	75%
b300R	69	a232R	87	58%	M456L	131	32%	
b301L	78							
b304R	79	a235R	83	46%				
b306L	139				M236R	138	36%	
b307R	98							
b308L	80							
b309L	130							
b311L	98							
b312L	73							
b313L	83	a155R	94	52%				
b315R	93							
b317L	117	a242L	117	80%				
b318L	96							
b319L	86							
b320L	86							
b321L	100				M226R	146	39%	
b325R	117							
b326R	73							
b327L	96							
b328R	81							
b329R	123							
b330R	85							
b332L	126							
b333L	73	a249L	102	65%				
b335R	86							
b337L	72							
b338R	131							
b340R	72	a648L	69	47%				
b343L	84	a252aL	84	66%				
b345R	71							
b347R	69	a288L	88	77%				
b351R	107				M118R	67	59%	
b352R	76	a266R	83	47%				
b353R	76							
b355R	84	a270R	75	75%				
b356R	201	a272R	67	52%				
b360L	73							
b362L	70							
b363L	146							
b364R	86	a281R	239	51%	M387R	297	56%	
b366R	72	a279R	72	97%				
b367R	93	a280R	93	94%	M147R	67	38%	
b369L	67	a283L	98	92%				
b372R	97				M102L	112	52%	
b373R	94	a285R	71	91%				
b374L	112	A682L	368	31%	M413L	333	31%	
b375L	179							

NY-2A		PBCV-1			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
b376L	89							
b379R	86							
b380L	67							
b382R	93							
b383R	114							
b384L	145							
b386L	93							
b387L	73				M098R	136	34%	
b390L	130							
b391L	73							
b392L	94							
b394R	138	a293R	168	69%				
b396R	67							
b398R	75							
b400L	85							
b405L	69							
b406L	248							
b407R	319	a156L	110	45%				
b409R	132							
b410L	96	a155R	94	62%				
b412R	76							
b413R	87							
b414R	70	a518R	100	54%				
b415L	92							
b417L	83							
b420L	65							
b422L	132							
b423R	76	a300R	76	42%				
b425R	104							
b427L	96							
b428L	69	a303L	112	59%				
b436R	125							
b439R	105	A314R	80	71%	M323R	79	66%	
b442L	149	a40L	129	53%				
b444R	124							
b445L	99							
b447R	66							
b448L	104							
b449L	99							
b452R	123							
b454R	73	a323R	132	54%	M330R	124	60%	
b456R	122							
b459R	66							
b464L	77							
b466L	69							
b467L	112							
b470R	70							
b471R	78							

NY-2A		PBCV-1			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
b473L	67							
b474L	72							
b475L	69							
b476R	73							
b478L	77							
b479L	67	a331L	67	98%				
b481L	87	A333L	299	100%				
b482R	114	a334R	110	94%				
b483L	71							
b484L	132							M702R
b486R	71	a338R	89	45%				
b489R	148	a344R	115	62%				
b490L	80	a345L	96	64%				
b491R	112							
b493L	80	A87R	456	58%				
b495L	70	a347L	70	52%				
b500R	110							
b501R	111							
b502R	70							
b504L	70							
b505R	68							
b507R	242							M364R
b509L	72							
b510L	66							
b511L	95							
b512L	73							
b513L	67							
b514L	122							
b515L	81							
b516L	165							
b517L	78							
b518L	95							
b520R	90							
b521R	94							
b523R	100	a377R	91	72%				
b526R	177	a380R	101	71%				
b527R	68	a381R	85	89%				
b528L	66							
b530L	123							M382L
b531L	77							
b532R	225	a386R	259	81%				
b534R	71	a388R	78	75%				
b535R	84	a391R	95	63%				
b538L	66							
b544R	66							
b545R	68							
b551L	103							
b555L	69							

NY-2A		PBCV-1			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
b557R	79							
b558L	81							
b561R	122	a409R	82	80%				
b563L	68							
b565R	81							
b569L	132	a415L	69	80%				
b572R	90							
b574R	77							
b577L	135							
b578L	66							
b584L	76							
b586R	78							
b588L	65							
b589R	66							
b594R	84							
b599R	67							
b600L	79							
b601R	68							
b603L	78							
b604L	74							
b605R	102							
b607R	68	a446R	101	88%	M405L	105	54%	
b608R	66							
b609R	65	a447R	99	95%				
b610R	75							
b613L	85							
b615L	69							
b616L	68							
b620R	75	a453R	89	64%				
b621R	123	a455R	147	78%				
b622L	80							
b624R	69	a457R	76	####	M677L	210	59%	
b625R	65				M677L	210	72%	
b626R	109							
b627L	69	a458L	83	60%				
b632R	80				M668L	333	25%	
b634L	66	a469L	76	63%				
b635L	77							
b637R	68							
b639L	85	a472L	90	76%	M658R	93	45%	
b640L	139	a477L	73	60%	M656R	102	40%	
b642L	83							
b643R	80							
b646R	84							
b648L	170							
b649R	70							
b653L	108							
b654L	79							

NY-2A		PBCV-1			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
b658L	65							
b660R	76							
b661L	80							
b664L	81							
b665L	144	a498L	124	85%				
b668R	135							
b669L	103							
b673R	71	a507R	186	55%				
b674R	69	a507R	186	59%	M706L	467	39%	
b675R	159	a508R	77	80%	M708L	92	48%	
b676R	92	a510R	67	87%				
b678L	100							
b679L	90	A519L	82	90%	M282R	86	60%	
b682R	82							
b683R	79							
b686R	68							
b689R	91	a155R	94	61%				
b690R	73							
b691R	99							
b693R	84							
b695L	191	a518R	100	63%	M361R	71	64%	
b698L	74							
b703R	76							
b706L	67							
b707R	97							
b708R	131							
b709L	77							
b712R	86							
b713L	67							
b716R	93							
b717R	114							
b721R	80							
b722R	69							
b726R	89							
b727R	88							
b728R	76	a541R	96	53%	M527R	85	41%	
b729L	76							
b730R	90				M156L	223	27%	
b731L	66							
b732L	76							
b733R	66							
b735L	109							
b737R	89							
b739R	93	a549R	81	41%				
b740R	78	a550R	136	44%				
b742L	71							
b745R	71							
b746R	75				M390L	80	37%	

NY-2A		PBCV-1			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
b749L	75							
b750R	112							
b752R	142	a560R	104	65%				
b754R	163							
b755R	97							
b757R	101							
b758L	83	a563L	82	45%				
b760L	136	a566L	90	97%				
b764L	190	A435R	73	55%	M474L	109	60%	
b768R	74							
b770L	79	a518R	100	53%	M361R	71	51%	
b773R	66				M554L	116	58%	
b775R	81							
b776R	78							
b778R	113				M533R	85	40%	
b779L	156	a156L	110	32%				
b780R	125							
b782R	94				M550L	330	35%	
b783R	98	a584R	98	91%	M551L	70	57%	
b784R	222	a585R	79	82%	M550L	330	29%	
b785R	100	a586R	75	85%	M549L	96	62%	
b786R	65	a587R	114	81%				
b787R	80				M548L	132	67%	
b788R	212	a588R	116	50%				
b789R	108							
b790L	76	a589L	76	52%				
b793L	106	a595L	84	67%				
b794R	73							
b797R	184	a599R	179	72%	M603R	90	32%	
b799R	77							
b804L	69							
b806L	147							
b807R	68							
b808L	91							
b809L	81							
b810L	227	A605L	158	85%	M609L	149	43%	
b814L	68							
b815R	77	a613R	87	52%				
b817R	91	a616R	93	63%				
b821R	84							
b822R	80							
b827R	81							
b830L	74	a626L	74	100%				
b833R	74	a630R	79	74%				
b834R	84							
b835L	76				M781R	125	37%	
b836R	89							
b837L	151				M779R	321	32%	

NY-2A		PBCV-1			MT325			Description
ORF	AA	ORF	AA	%ID	ORF	AA	%ID	
b838R	131							
b845R	79	a640R	79	96%				
b846L	74	a641L	85	95%				
b848L	80							
b851R	68							
b854R	106							
b855L	71	a655L	106	54%				
b856R	81							
b861L	65							
b863R	74							
b864R	96							
b866R	106							
b868R	214	a80L	135	49%				
b870L	133	a677L	76	84%	M159L	251	31%	
b872L	77							
b873R	83							
b875R	157	a6L	72	34%				
b876R	76							
b877R	75							
b879R	153							
b881R	142	a6L	72	35%				
b882R	82	a690R	78	67%				
b885L	74							

MT325		PBCV-1			NY-2A			Description
ORF	AA	ORF	AA	% ID	ORF	AA	%ID	
M1L	223							
M3L	125							
M5L	180	A81L	189	46%	B122L	192	44%	
M7L	170	A84L	188	37%	B124L	187	37%	
M9L	269	A75L	280	42%	B107L	288	41%	
M10L	119	A200R	118	62%	B271R	145	66%	Cytosine deaminase
M11L	96	A199R	101	45%	B267R	100	48%	
M12R	156	A196L	152	59%	B264L	152	56%	
M14R	264	A193L	262	68%	B261L	262	68%	PCNA
M15L	1286	A189/192R	1299	48%	B258R	1278	53%	SWI/SNF chromatin remodeling complex
M19L	907	A185R	913	74%	B249R	760	70%	6 DNA polymerase
M23L	73							
M24L	67							
M26L	346	A53R	363	70%				D-lactate dehydrogenase
M30R	270							Aquaglyceroporin
M32L	229	A315L	246	51%	B602L	214	55%	GIY-YIG endonuclease
M34L	214	A674R	216	61%	B865R	216	62%	Thymidylate synthase X
M37R	593	A100R	595	61%	B143R	595	60%	Glucosamine synthetase
M42R	956	A122R	1032	34%	B055L	953	77%	
M47R	1467	A122R	1032	30%	B040L	1471	59%	
M55R	1461	A122R	1032	24%	B040L	1471	60%	
M61R	1467	A122R	1032	33%	B040L	1471	61%	
M69R	432	A354R	237	42%	B424L	351	45%	HNH endonuclease
M71R	1497	A122R	1032	31%	B034L	1485	65%	
M78R	410	A430L	437	46%	B059R	411	86%	
M81R	207							
M83R	85							
M85R	523	A260R	484	77%	B480L	387	48%	Chitinase
M86R	196	A301L	241	32%	B426L	210	30%	
M88R	213	A298L	225	45%	B421L	227	41%	
M90L	165	A296R	156	64%	B404R	132	55%	
M91R	343	A292L	328	64%	B393L	322	61%	Chitosanase
M93L	377	A354R	237	42%	B370L	349	46%	HNH endonuclease
M97L	359	A286R	378	53%	B385R	372	52%	
M99R	169	A245R	187	76%				Cu/Zn-superoxide dismutase
M100L	192	A297L	177	53%	B419L	183	56%	Fructose-2,6 bisphosphatase
M101R	279	A284L	279	57%	B371L	279	58%	Amidase
M103L	296	A78R	298	68%	B116R	298	67%	N-carbamoylput. amidohydrolase
M107R	302	A243R	302	76%				
M109R	149	A253R	150	35%	B342R	166	35%	
M111R	239	A315L	246	47%	B440L	272	53%	GIY-YIG endonuclease
M114L	226	A262L	123	47%	B348L	215	50%	
M116L	255	A265L	170	65%	B350L	249	65%	
M120L	119	A647R	189	33%	B118R	280	32%	
M121R	93							
M122R	141	a136R	146	38%	B187R	145	34%	
M124L	348	A94L	364	58%	B137L	320	28%	β-1,3-glucanase
M128R	561	A98R	568	73%				Hyaluronan synthase
M133R	319	A103R	330	54%	B148R	321	53%	mRNA guanylyltransferase
M136L	119	A645R	123	35%	B852R	123	35%	

MT325			PBCV-1			NY-2A			Description
ORF	AA	ORF	AA	% ID	ORF	AA	% ID		
M137L	279	A105L	284	45%	B150L	290	44%	Ubiquitin C-terminal hydrolase	
M139L	312	A107L	290	39%	B154L	338	37%	Transcription factor TFIIB	
M140L	168	A109L	108	58%	B157L	163	57%		
M143L	605	A282L	569	45%	B365L	605	75%	Serine/Threonine protein kinase	
M148L	96								
M149R	157	A664L	142	52%	B860L	154	49%		
M150R	104	A121R	97	73%	B168R	104	67%		
M152R	1631	A122R	1032	44%	B170R	1386	31%		
M160R	429	A676R	375	40%	B869R	364	42%		
M163L	180	A125L	180	71%	B175L	192	68%	Transcription factor TFIIS	
M166R	220	A127R	245	65%	B177R	239	63%		
M167L	252	A9R	175	53%	B033R	254	48%		
M170R	122	A48R	123	39%	B074R	123	39%		
M171L	84								
M173R	150								
M174L	458							Nucleotidyl transferase	
M177L	470								
M178L	220								
M179L	161	A662L	171	64%	B859L	171	60%		
M180L	158	A134L	165	53%	B185L	164	50%	GIY-YIG endonuclease	
M183R	95	A250R	94	48%	B336R	94	46%	Potassium channel protein	
M184L	113	A139L	108	55%	B190L	100	57%		
M186R	520	A473L	517	74%				Cellulase synthase	
M190R	1069	A140/A145R	1130	37%	B192R	1075	51%		
M200L	103	A150L	152	39%	B197L	131	41%		
M201R	454	A153R	459	66%	B203R	460	63%	Superfamily II helicase	
M204L	117	A157L	110	51%	B207L	111	49%		
M206L	73								
M208L	133	a160L	164	30%					
M210L	95	A165L	279	47%	B211L	112	43%		
M213L	152	A165L	279	38%	B212L	148	39%		
M215R	279	A166R	268	60%	B214R	268	57%	Exonuclease	
M218R	161	A168R	166	54%	B215R	167	53%		
M219L	274	A173L	288	58%	B226L	279	58%	Patatin-like phospholipase	
M221L	300	A248R	308	38%	B331R	309	38%	Serine/Threonine protein kinase	
M225L	715	A241R	725	51%	B316R	725	51%	Superfamily II helicase	
M232R	104								
M233L	502	A237R	518	70%	B305R	507	69%	Homospermidine synthase	
M237L	109								
M238R	119	A234L	108	47%	B303L	108	42%		
M240L	90	A233R	112	47%	B302R	111	46%		
M241L	87	A438L	78	30%	B592L	78	30%	Glutaredoxin	
M242L	190	A230R	196	40%	B293R	213	40%		
M243R	78								
M246L	80								
M247R	250	A9R	175	70%	B033R	254	65%		
M249R	138	A567L	152	38%	B761L	145	37%		
M253L	626	A565R	576	47%	B759R	675	50%	ATPase (Chromosome segregation)	
M258R	595	A561L	649	41%	B756L	649	42%	ATPase (DNA repair)	
M263R	218	A559L	213	37%	B751L	211	32%		

MT325			PBCV-1			NY-2A			Description
ORF	AA	ORF	AA	% ID	ORF	AA	% ID		
M264L	140	A551L	141	75%	B741L	141	74%	dUTP pyrophosphatase	
M266R	273	A552R	270	43%	B743R	317	39%	Transcription factor TFIID	
M269L	400	A558L	400	61%	B748L	400	62%		
M272L	459	A548L	458	52%	B738L	458	51%	SWI/SNF helicase	
M278R	267	A138R	273	23%	B130R	161	50%		
M281L	65								
M282R	86	A519L	82	60%	B679L	90	60%		
M283R	386	A217L	394	73%	B289L	382	70%	Monoamine oxidase	
M288L	352	A495R	221	37%	B039L	418	43%	GIY-YIG endonuclease	
M289R	368	A215L	321	72%	B288L	306	64%	β & α , 1,4 linked glucuronic lyase	
M294R	217								
M298R	157	A214L	135	50%	B285L	133	46%		
M299R	149	A213L	148	52%	B284L	148	50%		
M302L	322							ATPase (AAA+ class)	
M307L	372	A207R	372	63%	B278R	372	61%	Ornithine/Arginine decarboxylase	
M308L	235	A205R	206	48%	B277R	206	48%		
M310L	220	A203R	216	65%	B274R	216	65%		
M312R	112	A202L	113	53%	B273L	112	53%		
M313L	185	A305L	204	66%	B430L	181	68%	Dual specificity phosphatase	
M314L	85	A306L	86	84%	B431L	87	78%		
M315L	112	A308L	86	41%	B432L	103	41%		
M317L	168	A310L	170	54%	B434L	170	56%		
M319L	241	A312L	238	71%	B435L	248	70%		
M322L	84	A313L	71	42%	B437L	76	57%		
M323R	79	A314R	80	83%	B439R	105	66%		
M324R	372								
M326R	117	A320R	139	37%	B450R	115	30%		
M328R	113	A321R	119	41%	B451R	118	33%		
M329L	166	A322L	176	46%	B453L	178	46%		
M332L	422	A324L	453	58%	B455L	433	56%		
M335L	211	A326L	209	60%	B457L	199	58%	NTP pyrophosphohydrolase	
M337L	352	A328L	355	38%	B458L	358	38%		
M341R	241	A312L	238	71%	B435L	248	70%		
M344L	174	A337L	78	48%	B485L	238	32%		
M346L	139	A341L	135	25%	B487L	139	25%		
M348L	559	A342L	576	68%	B488L	551	68%		
M354R	438							Cellulase precursor	
M357L	127	A349L	141	73%	B496L	99	65%		
M359L	342	A517L	344	64%	B697R	369	67%	Cytosine methyltransferase	
M361R	71	a518R	100	76%	B695L	191	63%		
M362L	213	A352L	207	70%	B503L	207	70%		
M363L	341	A357L	276	52%	B506L	337	54%	ATPase (AAA+ class)	
M367L	219	A131L	136	36%	B181L	135	27%		
M370R	267	A315L	246	49%	B440L	272	52%	GIY-YIG endonuclease	
M372R	1131	A363R	811	61%	B508R	1612	33%	Superfamily II helicase	
M375R	179	A646L	141	41%				Translation elongation factor-3	
M376R	196								
M377L	204	A379L	207	47%	B525L	209	50%		
M379R	146	A486L	152	39%					
M381R	472	A383R	301	53%	B529R	485	54%		

MT325			PBCV-1			NY-2A			Description
ORF	AA	ORF	AA	% ID	ORF	AA	% ID		
M385L	527	A92/93L	359	48%	B136L	423	41%	Histone-like protein	
M388R	235								
M389R	499	A554/556/557L	498	53%	B744L	504	52%	ATPase (PP-loop)	
M395L	87								
M397R	254	A9R	175	57%	B033R	254	50%		
M399L	183	A449R	193	55%	B612R	188	52%	RNA triphosphatase	
M401R	224	A85R	242	56%	B126R	230	57%	Prolyl-4-hydroxylase	
M403R	106	A448L	106	56%	B611L	106	48%	Protein disulfide isomerase	
M404R	463	A445L	462	64%	B606L	462	65%	ABC transporter protein	
M407R	113	A444L	104	35%	B597L	105	36%		
M409R	469							Periplasmic protein TonB	
M411L	141	A441L	137	53%	B595L	137	52%		
M413L	333	A682L	368	35%	B018L	472	33%		
M416R	298	A443R	308	34%	B596R	308	34%		
M417L	600	A140/145R	1130	29%	B533L	532	39%	Periplasmic protein TonB	
M421L	110	A439R	112	61%	B593R	113	59%		
M423R	86	A438L	78	55%	B592L	78	54%	Glutaredoxin	
M425R	186	A416R	188	51%				Deoxyribonucleoside kinase	
M427R	352	A129R	358	45%	B310R	382	43%		
M430L	360	A417L	429	40%	B571L	434	40%	Replication factor C	
M435L	114	A420L	70	65%	B573L	75	61%		
M437R	84								
M438L	81								
M440R	156	A423R	157	37%	B576R	152	37%		
M444R	116	A426R	114	42%	B580R	116	42%	6-phosphofructokinase	
M445L	116	A427L	119	44%	B581L	119	38%	Thioredoxin	
M448L	112	A427L	119	28%	B581L	119	27%	Thioredoxin	
M449L	104	A427L	119	34%				Thioredoxin	
M451L	442	A429L	476	34%	B583L	453	33%		
M454R	108								
M455R	93	A601R	101	58%	B800R	90	55%		
M457R	376	A231L	383	64%	B298L	372	63%		
M460R	251	A495R	221	29%	B850L	391	34%	GIY-YIG endonuclease	
M463L	433	A430L	437	80%	B585L	437	79%		
M465L	287	A315L	246	46%	B440L	272	50%	GIY-YIG endonuclease	
M467R	849	A114R	485	67%	B159R	867	63%	Fucosyl-Glycosyltransferase	
M472R	152	A432R	176	46%	B587R	155	42%		
M475R	116	A571R	116	79%	B765R	115	77%		
M476L	202	A437L	103	70%	B591L	103	70%	DNA binding protein	
M478L	106	A410L	110	49%	B562L	125	50%		
M479R	164	A411R	170	38%					
M482R	179	A412R	179	50%	B564R	178	51%		
M484L	246	A413L	244	42%	B568L	236	45%		
M485R	231	A408L	277	54%	B560L	266	52%		
M488L	249							Glycosyltransferase	
M491R	382								
M495R	209	A407L	210	70%	B559L	210	71%		
M496L	205	A521L	392	63%	B685L	202	66%		
M497R	178	A523R	171	74%	B687R	168	73%		
M501R	148	A526R	129	52%	B692R	146	55%	Glycerophosphodiesterase (?)	

MT325			PBCV-1			NY-2A			Description
ORF	AA	ORF	AA	% ID	ORF	AA	% ID		
M502L	245	A315L	246	55%	B206L	279	56%	GIY-YIG endonuclease	
M504R	93	A527R	105	62%	B694R	95	55%		
M506L	70	A531L	67	53%	B700L	65	45%		
M507L	79	A532L	79	61%	B701L	79	60%		
M508R	527	A533R	374	60%	B705R	530	65%		
M511L	72	A535L	71	51%	B710L	71	51%		
M512L	79	A536L	73	48%	B719L	81	56%		
M515R	161								
M517R	168	A539R	173	52%	B724R	173	51%	GIY-YIG endonuclease	
M518L	1189	A540L	1176	31%	B725L	1191	41%	Autotransporter adhesin	
M522L	615	A35L	549	26%					
M525L	253	A9R	175	68%	B033R	254	58%		
M526L	1098	A540L	1176	30%	B725L	1191	41%	Autotransporter adhesin	
M530L	144	A596R	142	66%	B795R	142	64%	dCMP deaminase	
M531L	365	A154L	347	50%	B310R	382	44%		
M535L	871							Ca2+ transporting ATPase	
M543R	272	A289L	283	34%	B388L	268	32%	Serine/Threonine protein kinase	
M546R	1058	A583L	1061	67%	B781L	1061	67%	DNA Topoisomerase II	
M553R	136	A577L	106	74%	B772L	133	71%		
M556R	174	A375R	173	63%	B292L	75	38%		
M557L	541	A405R	496	31%	B554R	523	40%		
M560L	183	A404R	132	64%	B553R	187	62%		
M563L	88	A403R	93	52%	B552R	97	51%		
M564L	228	A402R	227	48%	B550R	232	48%	Lipoprotein lipase	
M567L	290	A401R	277	61%	B549R	278	60%		
M569L	117	A400R	118	45%	B548R	121	46%		
M570L	164	A399R	194	54%	B547R	193	51%	RNase H	
M572R	118	A398L	118	62%	B546L	118	59%		
M574L	153	a136R	146	41%	B258R	1278	28%		
M577L	123	A395R	82	66%	B540R	82	69%		
M578R	208	A478L	310	43%	B805R	340	39%	HNH endonuclease	
M580L	190	A394R	121	44%	B539R	126	43%		
M584R	102								
M586L	255	A392R	258	83%	B536R	254	81%	ATPase (DNA packaging)	
M591R	586	A384L	456	62%	B533L	532	62%	Periplasmic protein TonB	
M599R	287	a593R	132	60%	B792R	248	53%		
M601L	359	A598L	363	62%	B796L	366	59%	Histidine decarboxylase	
M605L	134	A602L	128	47%	B801L	134	37%		
M606R	105	A603R	105	43%	B802R	109	39%		
M607L	167								
M610L	173	A492L	189	38%	B656L	197	40%		
M612R	357	A494R	360	59%	B663R	360	59%		
M615R	142	A497R	146	48%	B666R	139	42%		
M616L	328	A501L	75	48%	B667L	336	52%		
M618L	75	A491R	76	47%	B655R	76	45%		
M620L	320	A488R	317	65%	B652R	317	64%		
M622L	349	A422R	342	31%	B370L	349	35%	HNH endonuclease	
M625R	82								
M627L	144	A50L	141	69%	B076L	116	68%	Pyrimidine dimer-specific glycosylase	
M628L	91	A480L	93	59%	B644L	92	56%		

MT325		PBCV-1			NY-2A			Description
ORF	AA	ORF	AA	% ID	ORF	AA	%ID	
m194R	74							
m195L	118							
m196L	77							
m197L	67							
m198L	91							
m199R	88							
m202L	90							
m203L	83							
m205R	68							
m207L	198							
m209R	100	A161R	123	40%				
m211R	85							
m212R	70							
m214L	71							
m216L	68							
m217L	134							
m220R	105				B228R	72	58%	
m222R	108	a249L	102	44%				
m223R	165							
m224L	93							
m226R	146				B321L	100	40%	
m227R	67							
m228L	68							
m229R	135							
m230R	70							
m231R	70							
m234R	66	a238L	114	66%				
m235R	70							
m236R	138				B306L	139	36%	
m239L	108							
m244R	130	A227L	137	61%	B291L	138	53%	
m245L	77	a228R	76	51%				
m248R	93							
m250L	90							
m251R	136							
m252L	66							
m254R	94							
m255L	71							
m256R	67							
m257R	79							
m259R	67							
m260R	114							
m261R	95							
m262L	77							
m265L	82							
m267L	71							
m268R	71							
m270R	65							
m271L	93							
m273R	82							

MT325		PBCV-1			NY-2A			Description
ORF	AA	ORF	AA	% ID	ORF	AA	%ID	
m274R	75							
m275R	94							
m276L	96							
m277L	204							
m279R	121							
m280L	76							
m284L	121							
m285L	88							B290R 88 71%
m286L	93							
m287L	85							
m290R	71							
m291L	102							
m292L	127							
m293L	82							
m295L	206							
m296L	117							
m297L	76							
m300L	102							
m301L	76							
m303R	260							
m304L	80							
m305R	113							
m306R	160							
m309R	162							
m311R	75	a204L	79	65%	B275L	70	69%	
m316R	100							
m318R	137							
m320R	131							
m321L	74							
m325L	65							
m327L	86							
m330R	124	a323R	132	40%				
m331R	127							
m333L	117							
m334R	199							
m336R	149							
m338R	134							
m339R	123							
m340L	149							
m342L	79							
m343L	88							
m345L	65							
m347R	104							
m349R	94							
m350L	111							
m351R	93							B489R 148 32%
m352L	156							
m353L	115							
m355R	95							
m356R	68							

MT325		PBCV-1			NY-2A			Description
ORF	AA	ORF	AA	% ID	ORF	AA	%ID	
m358R	157	a350R	122	50%	B498R	122	50%	
m360R	118							
m364R	116				B507R	242	45%	
m365L	111							
m366R	147							
m368L	87							
m369L	131							
m371L	68							
m373L	114							
m374L	66							
m378L	92							
m380L	69							
m382L	88				B530L	123	58%	
m383L	66							
m384L	66							
m386R	96							
m387R	297	a281R	239	45%	B364R	86	55%	
m390L	80				B746R	75	37%	
m391L	100							
m392L	77							
m393R	73							
m394L	103							
m396R	89							
m398R	112							
m400L	122							
m402L	88							
m405L	105	a446R	101	60%	B607R	68	54%	
m406L	65							
m408L	247							
m410L	72							
m412R	71							
m414R	127							
m415R	148							
m418R	89							
m419R	107							
m420R	132							
m422L	132							
m424L	78							
m426L	65							
m428L	74							
m429L	66							
m431R	81							
m432R	82							
m433R	72							
m434R	120							
m436R	134	A421R	98	53%	B575R	98	49%	
m439L	69							
m441L	246							
m442R	120							
m443L	132							

MT325		PBCV-1			NY-2A			Description
ORF	AA	ORF	AA	% ID	ORF	AA	%ID	
m446R	88							
m447R	90							
m450R	88							
m452R	89							
m453R	71							
m456L	131				B300R	69	32%	
m458L	89							
m459L	91				B299R	81	72%	
m461L	161							
m462L	74							
m464R	83				B530L	123	39%	
m466R	71							
m468L	73							
m469L	96							
m470L	67	a115L	76	33%				
m471L	76							
m473L	76							
m474L	109	A435R	73	52%	B764L	190	60%	
m477R	71	A414R	93	59%	B570R	72	60%	
m480L	77							
m481L	69							
m483R	93							
m486L	97							
m487R	173							
m489R	180							
m490L	70							
m492L	72							
m493L	88							
m494L	68							
m498L	102							
m499R	67							
m500L	65							
m503R	87							
m505R	89							
m509L	98							
m510L	89							
m513R	109							
m514L	248	A537L	265	32%	B723L	259	27%	
m516L	75							
m519L	71							
m520R	112	a541R	96	43%	B728R	76	37%	
m521L	65							
m523L	84							
m524R	77							
m527R	85	a541R	96	46%	B728R	76	44%	
m528R	82	a541R	96	41%				
m529R	69							
m532R	69							
m533R	85				B778R	113	39%	

MT325		PBCV-1			NY-2A			Description
ORF	AA	ORF	AA	% ID	ORF	AA	%ID	
m534R	90							
m536R	120							
m537R	138							
m538R	77							
m539R	145							
m540L	92							
m541R	175							
m542R	93							
m544L	66							
m545L	70							
m547R	70							
m548L	132				B787R	80	67%	
m549L	96	a586R	75	60%	B785R	100	62%	
m550L	330	a585R	79	37%	B784R	222	29%	
m551L	70	a584R	98	58%	B783R	98	57%	
m552L	146							
m554L	116				B773R	66	58%	
m555L	99							
m558R	216	a42L	88	45%				
m559R	205							
m561R	122							
m562R	152							
m565R	116							
m566R	81							
m568R	118							
m571R	78							
m573R	76							
m575R	66							
m576R	95							
m579L	77							
m581R	94							
m582R	97							
m583L	112	A394R	121	50%	B539R	126	45%	
m585R	68							
m587R	110							
m588R	81							
m589L	69							
m590L	84							
m592L	75				B087L	134	30%	
m593L	103							
m594L	158	a389R	102	54%				
m595R	83							
m596L	105							
m597R	78							
m598L	71							
m600L	143							
m602R	70							
m603R	90				B797R	184	32%	
m604R	87							
m608R	75							

MT325		PBCV-1			NY-2A			Description
ORF	AA	ORF	AA	% ID	ORF	AA	%ID	
m609L	149	A605L	158	43%	B810L	227	43%	
m611L	149							
m613L	82							
m614L	103							
m617R	137							
m619L	72							
m621R	73							
m623R	146							
m624L	77							
m626L	144	A485R	148	57%	B651R	144	56%	
m630R	182							
m632R	206							
m633R	145	a281R	239	53%	B364R	86	68%	
m636L	79							
m639R	65							
m640L	160							
m641L	68							
m642L	73							
m643L	86							
m644L	84							
m645R	132							
m646L	75							
m648L	223	a281R	239	55%	B364R	86	53%	
m650L	70							
m652L	136							
m654R	188							
m655L	88							
m656R	102				B640L	139	40%	
m658R	93				B639L	85	45%	
m659L	68							
m661R	108							
m662R	67							
m663L	85							
m665R	73							
m666R	80							
m668L	333				B632R	80	25%	
m669L	97							
m671L	92	a466L	91	52%				
m673R	80							
m675L	169	a460R	78	31%				
m676L	66							
m677L	210	a459R	100	72%	B625R	65	67%	
m678R	87							
m679R	120							
m681L	75							
m682L	69							
m684R	100							
m685R	88							
m687L	80							
m688R	126							

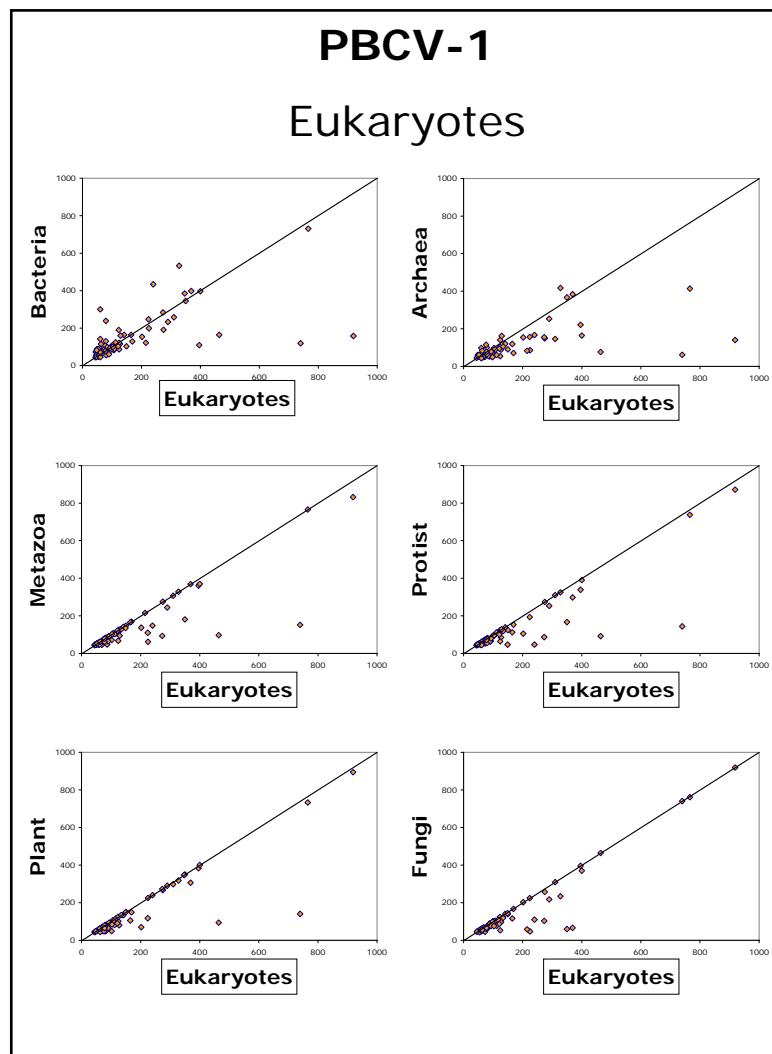
MT325		PBCV-1			NY-2A			Description
ORF	AA	ORF	AA	% ID	ORF	AA	%ID	
m689L	75							
m691L	72							
m692L	225	A568L	179	27%	B762L	179	32%	
m696L	104							
m698R	73							
m702R	83				B484L	132	45%	
m703L	140							
m705L	114							
m707R	446	A505L	484	58%	B672L	485	55%	
m708L	92				B675R	159	48%	
m709L	78							
m710R	69							
m711L	127	a506R	70	62%				
m712R	133							
m713R	244	A503L	304	40%	B671L	279	37%	
m716L	95							
m718R	69							
m720R	114	a225L	90	39%				
m722L	79							
m723R	84	a224L	88	58%				
m724R	75	a223aL	92	62%				
m725R	169							
m726R	87							
m728R	67	a613R	87	32%				
m730R	73							
m731R	68							
m732L	65							
m734R	316	A617R	321	49%	B818R	317	49%	
m736L	69							
m737R	68							
m739L	95							
m743L	84							
m744L	93	a670R	122	42%				
m745L	69							
m746L	98							
m747L	70							
m749R	159							
m750R	84							
m751R	102							
m754R	182							
m755L	86							
m759L	121							
m760L	115							
m761R	196							
m764R	116							
m765R	101							
m767R	121							
m768L	78							
m769R	70							
m771R	134							

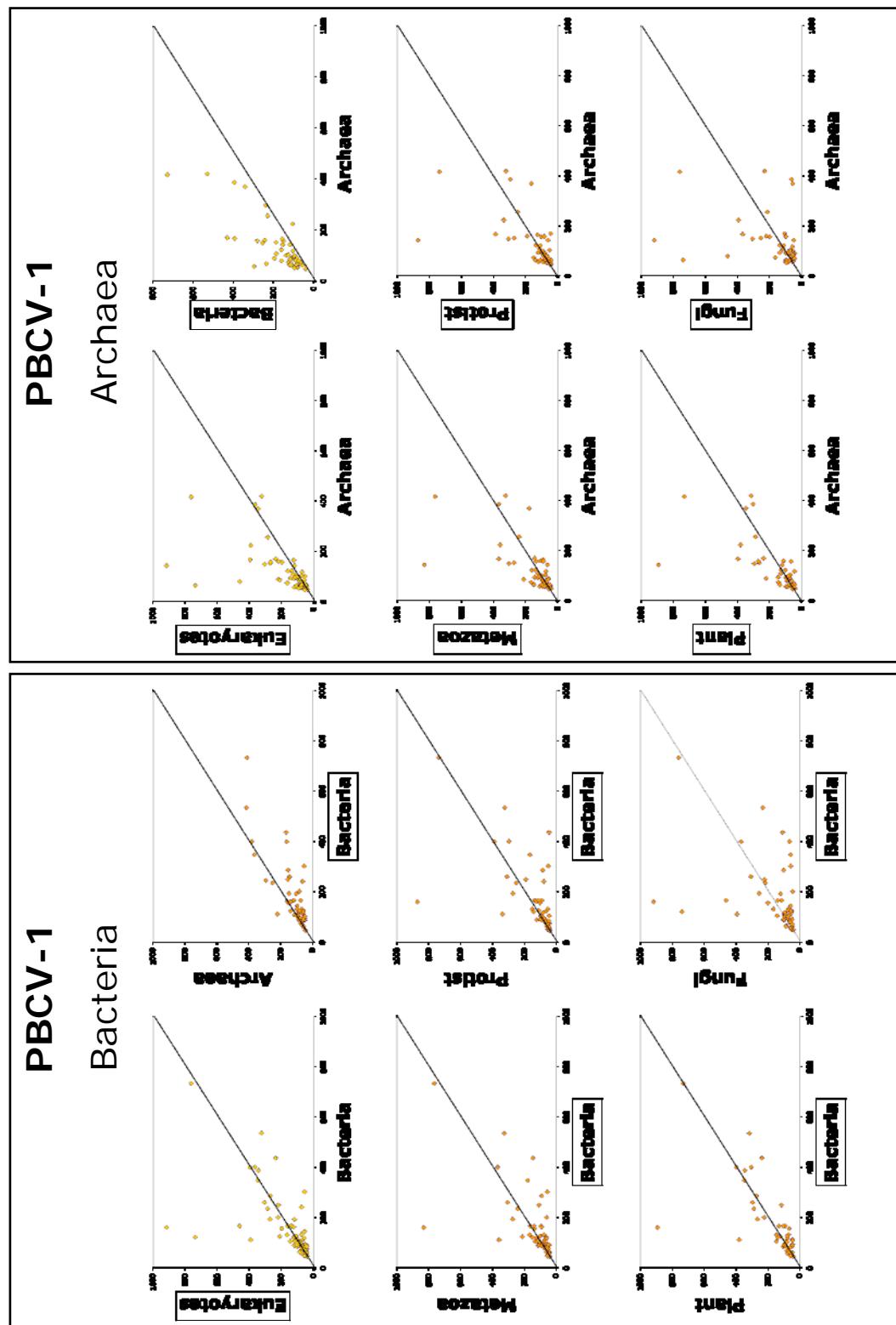
MT325		PBCV-1			NY-2A			Description
ORF	AA	ORF	AA	% ID	ORF	AA	%ID	
m772L	66							
m774R	66							
m775L	118	A633R	120	61%	B839R	120	56%	
m776L	91							
m778R	68							
m779R	321				B837L	151	32%	
m780R	80							
m781R	125	a631L	89	41%	B835L	76	37%	
m782R	70							
m783R	84							
m784R	150							
m786L	82							
m787R	78							
m788L	71							
m790R	77	a179L	109	68%				
m792L	141							
m793L	137				B241L	131	71%	
m795L	205							
m796L	69							
m798L	111							
m799L	110							
m800L	142							
m802L	165							
m804R	108							
m805L	107							
m806R	91							
m808L	75							
m809L	65							
m812R	149							
m815R	66							
m816R	76							
m821L	89							
m823L	143							
m824L	139							
m825L	73	a319L	84	54%				
m826L	164	A687R	75	29%	B235L	103	30%	
m827L	70							
m830R	95							
m831R	87							
m832R	75							
m834L	80							
m835R	118							
m836R	140							
m837L	68							
m839R	146							
m840L	74							
m841R	125							
m843L	201							

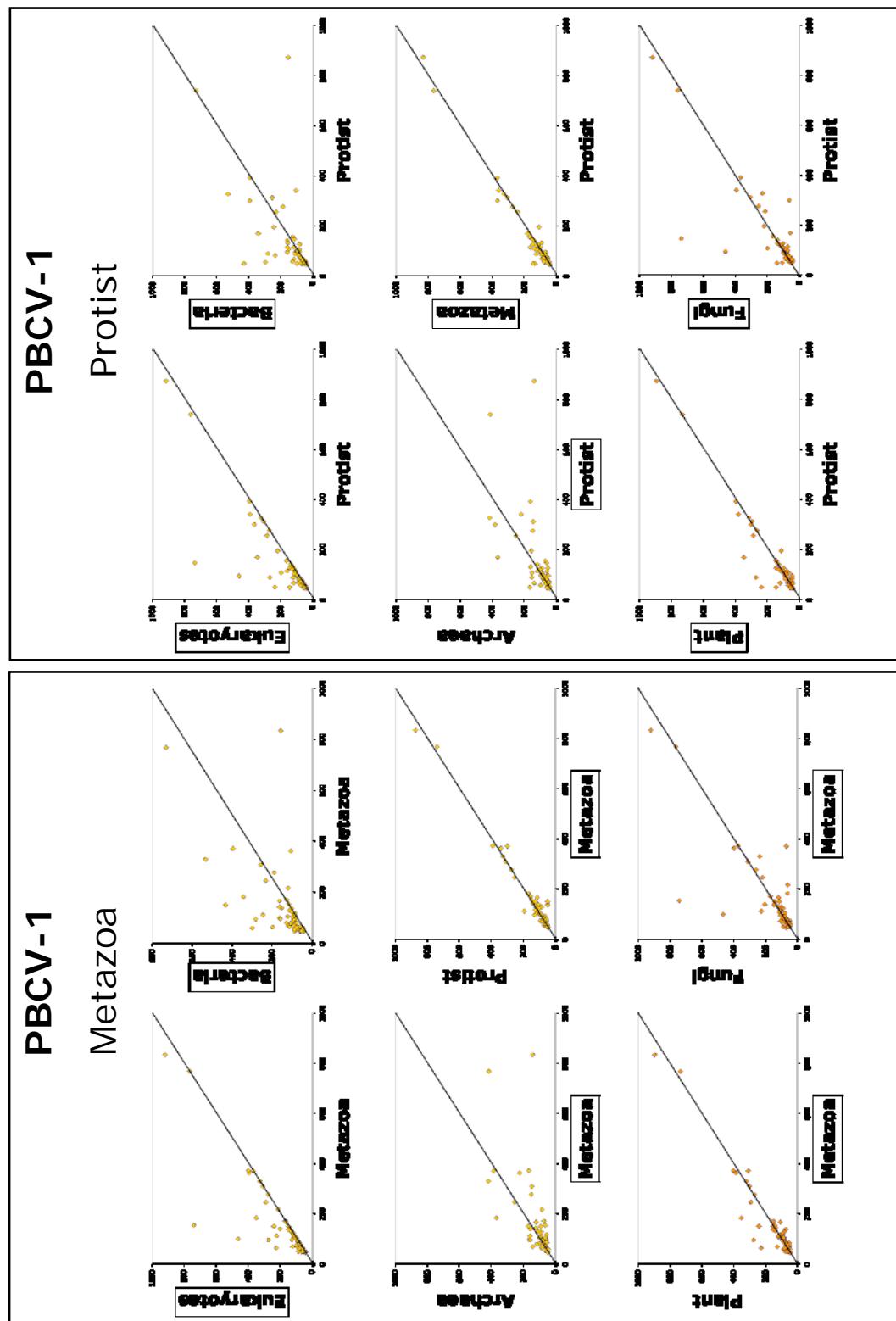
3.6.2 Domain and Kingdom Scatter Plots

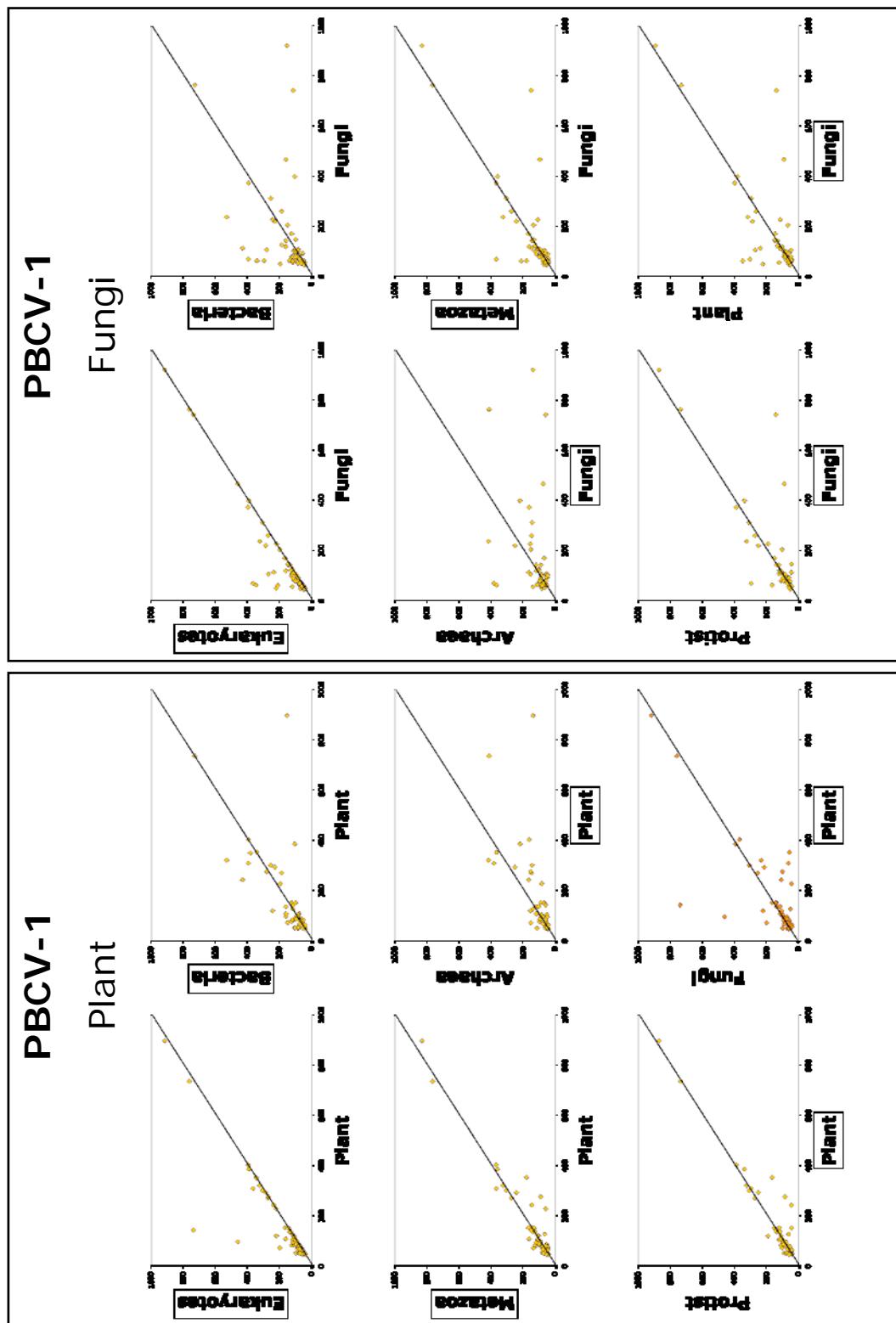
The results of the blastp comparison of the chlorella virus ORFs with the proteins in the NCBI NR database were used to create these domain and kingdom scatter plots. The plots between the three domains and four kingdoms (from the Eukaryotic domain) are presented. Every graph has boxed either the domain or kingdom name which has the highest similarity.

3.6.2.1 PBCV-1 Domain and Kingdom Scatter Plots

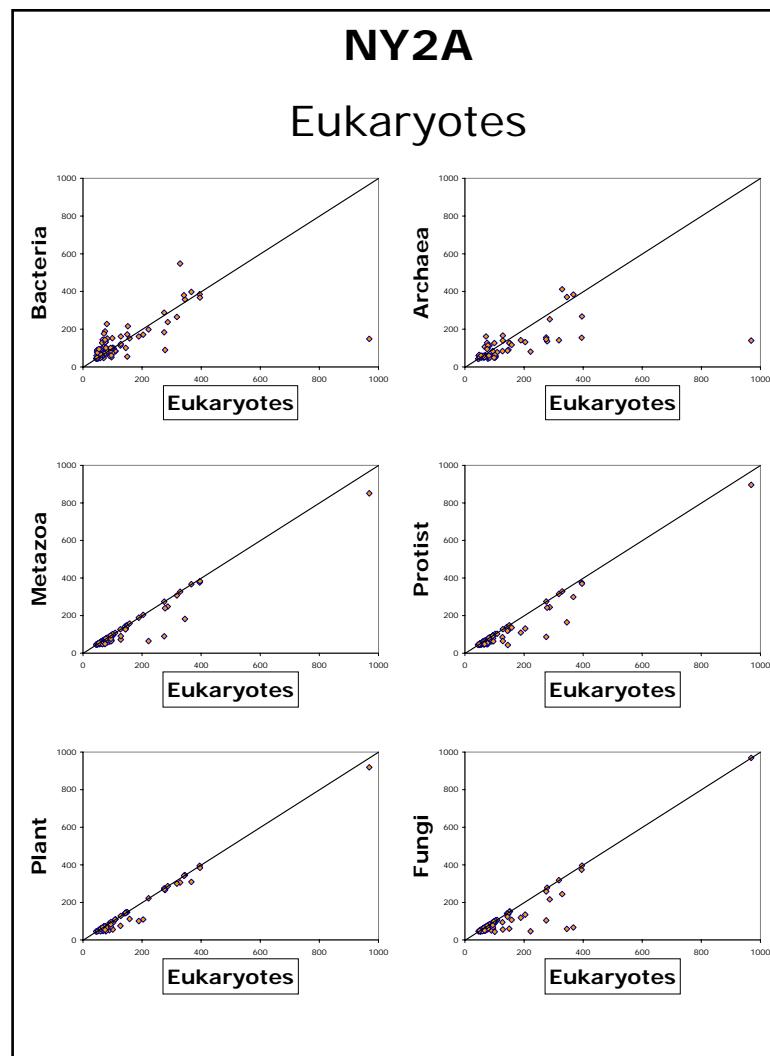


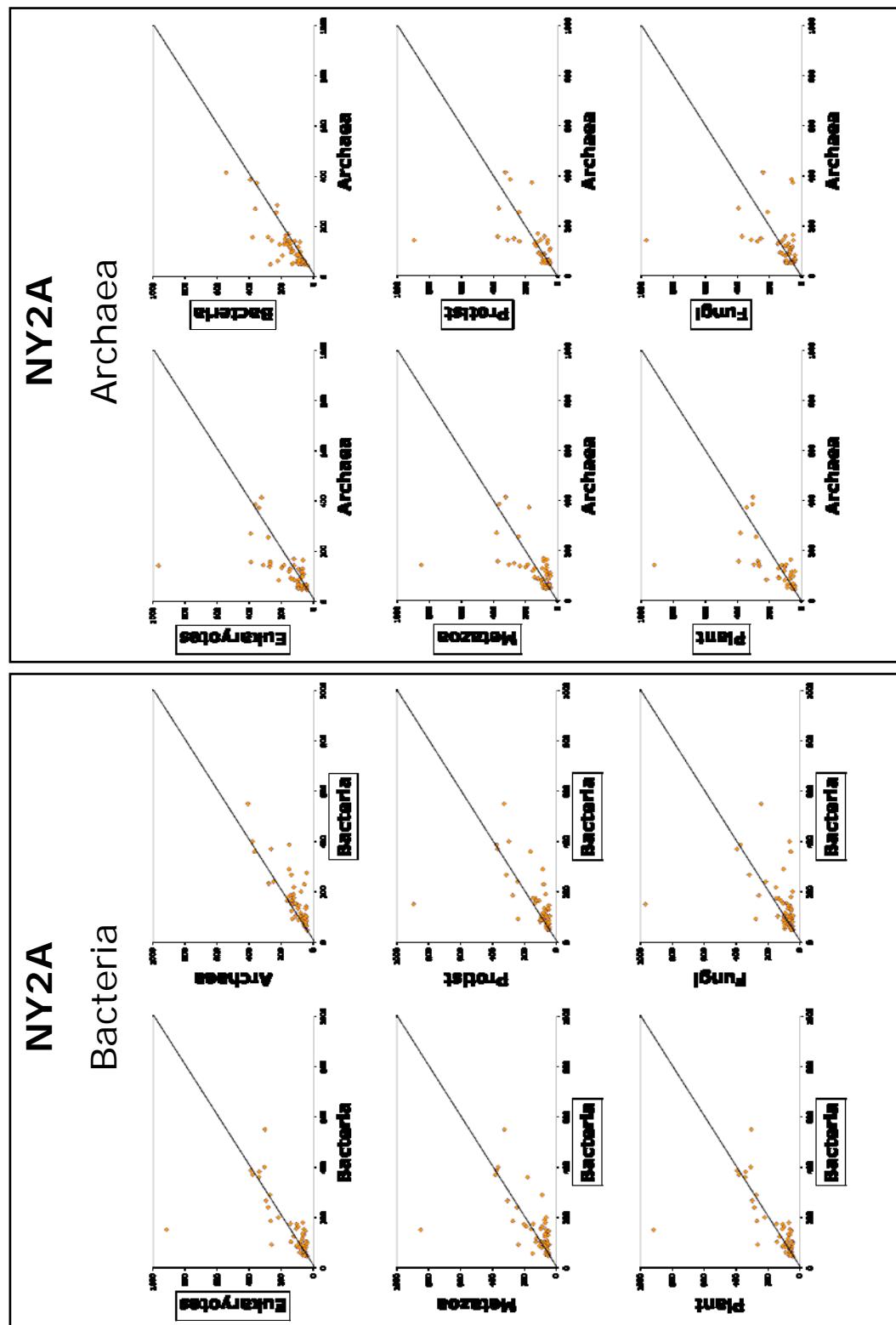


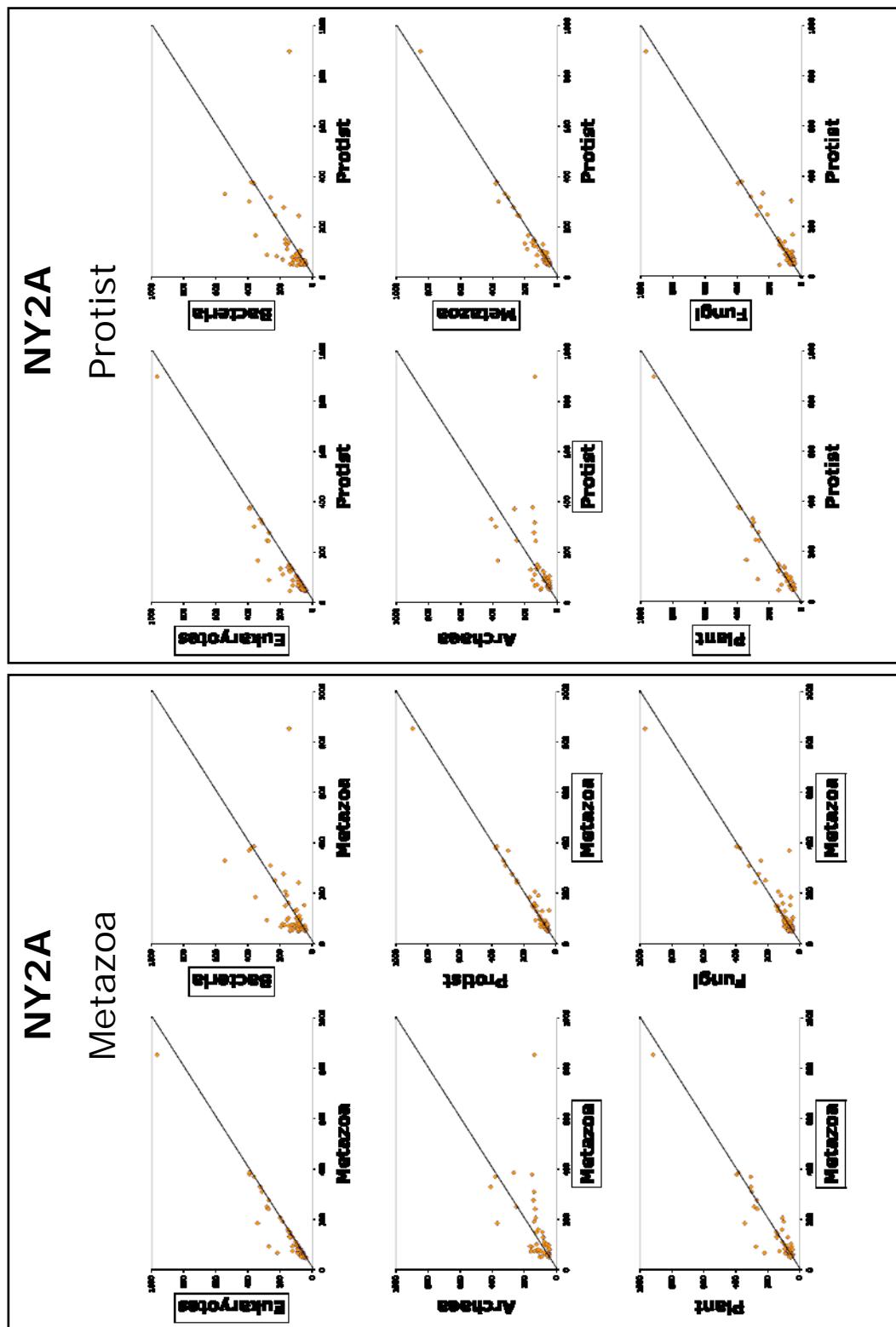


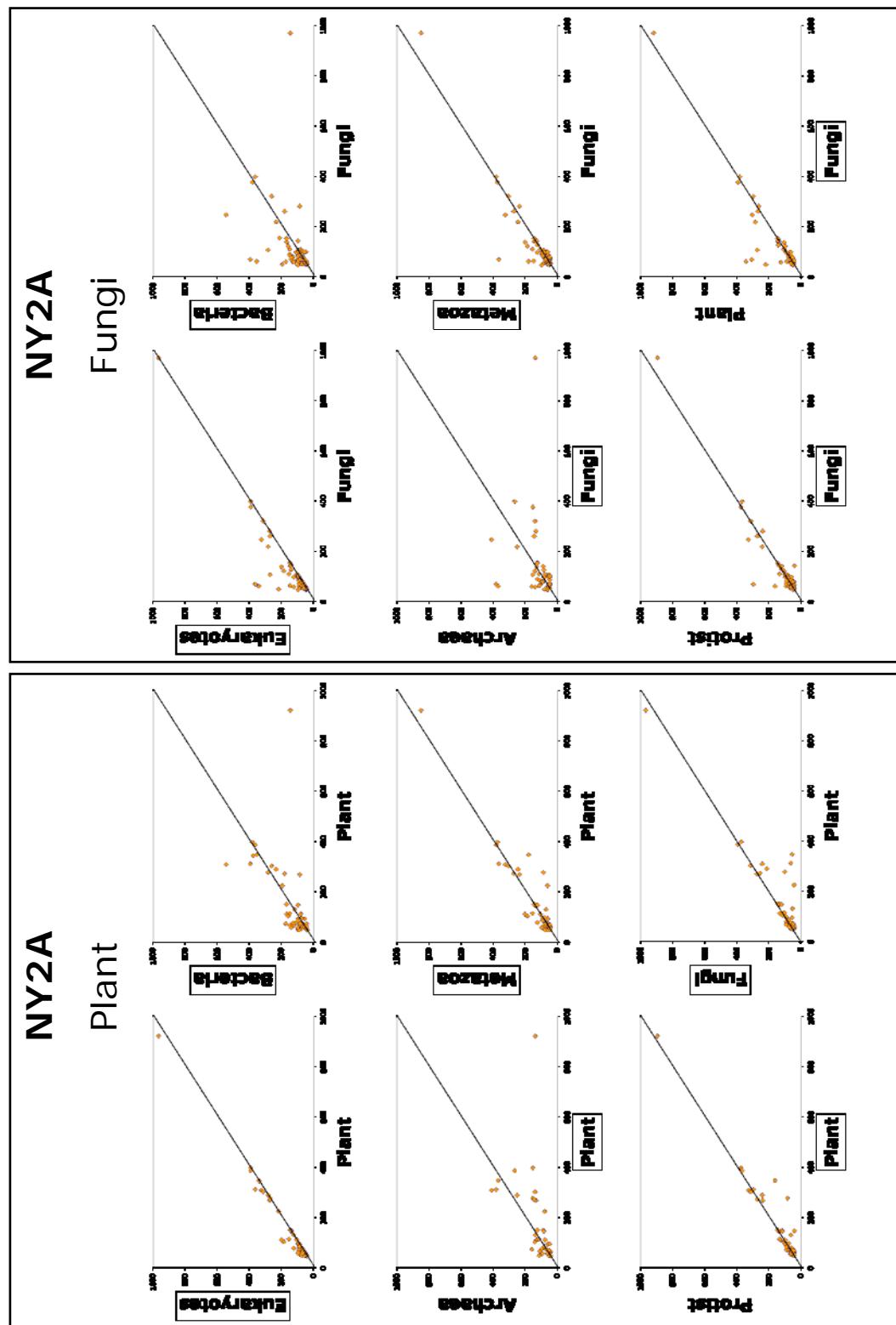


3.6.2.2 NY-2A Domain and Kingdom Scatter Plots

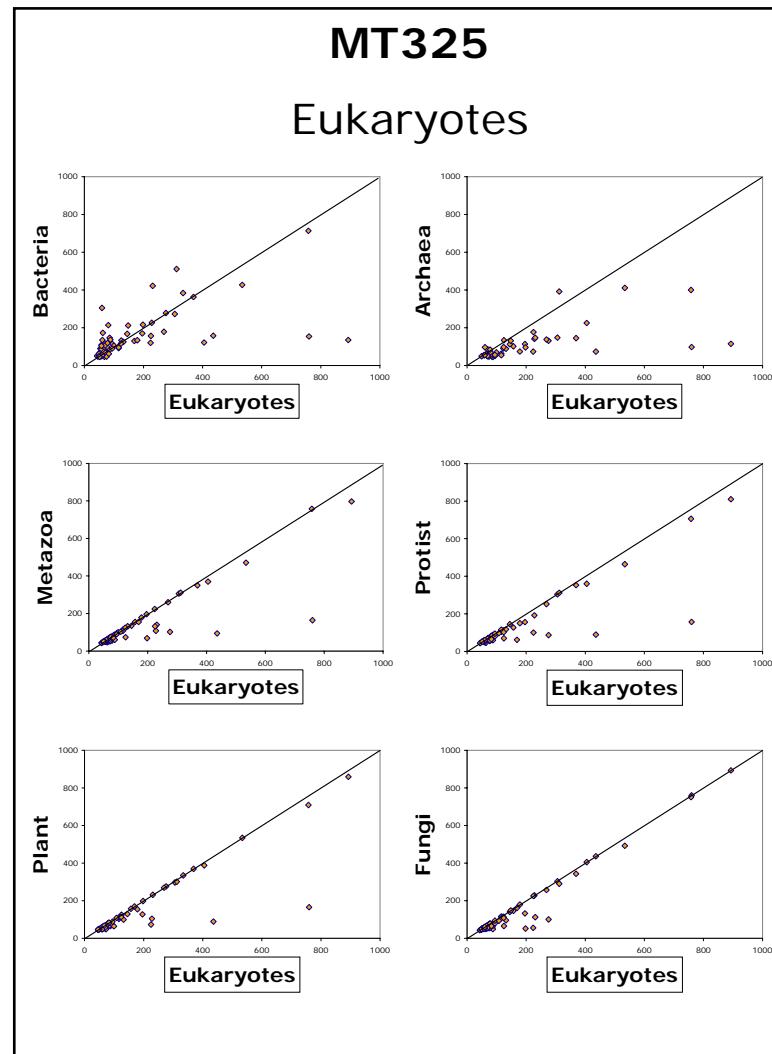


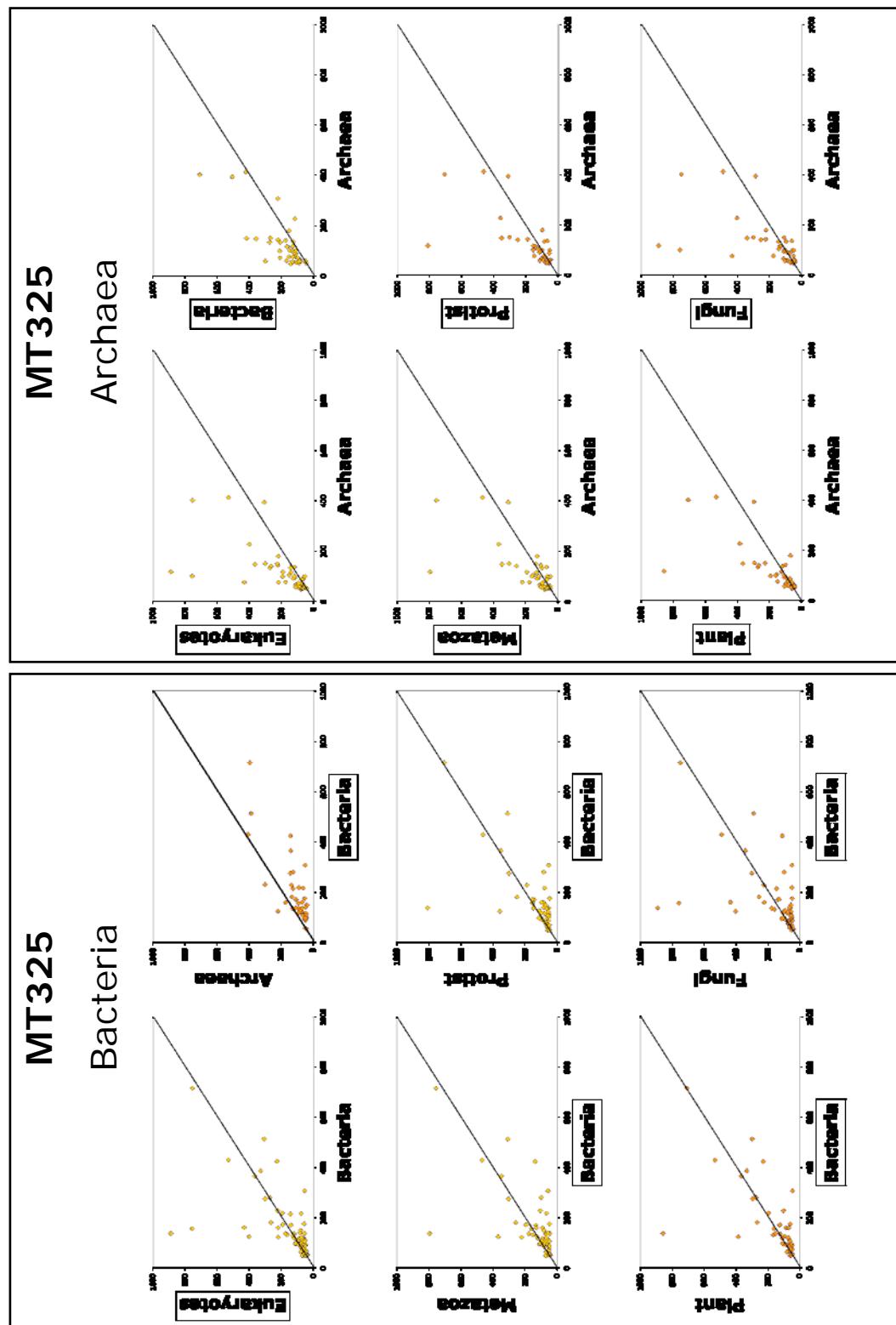


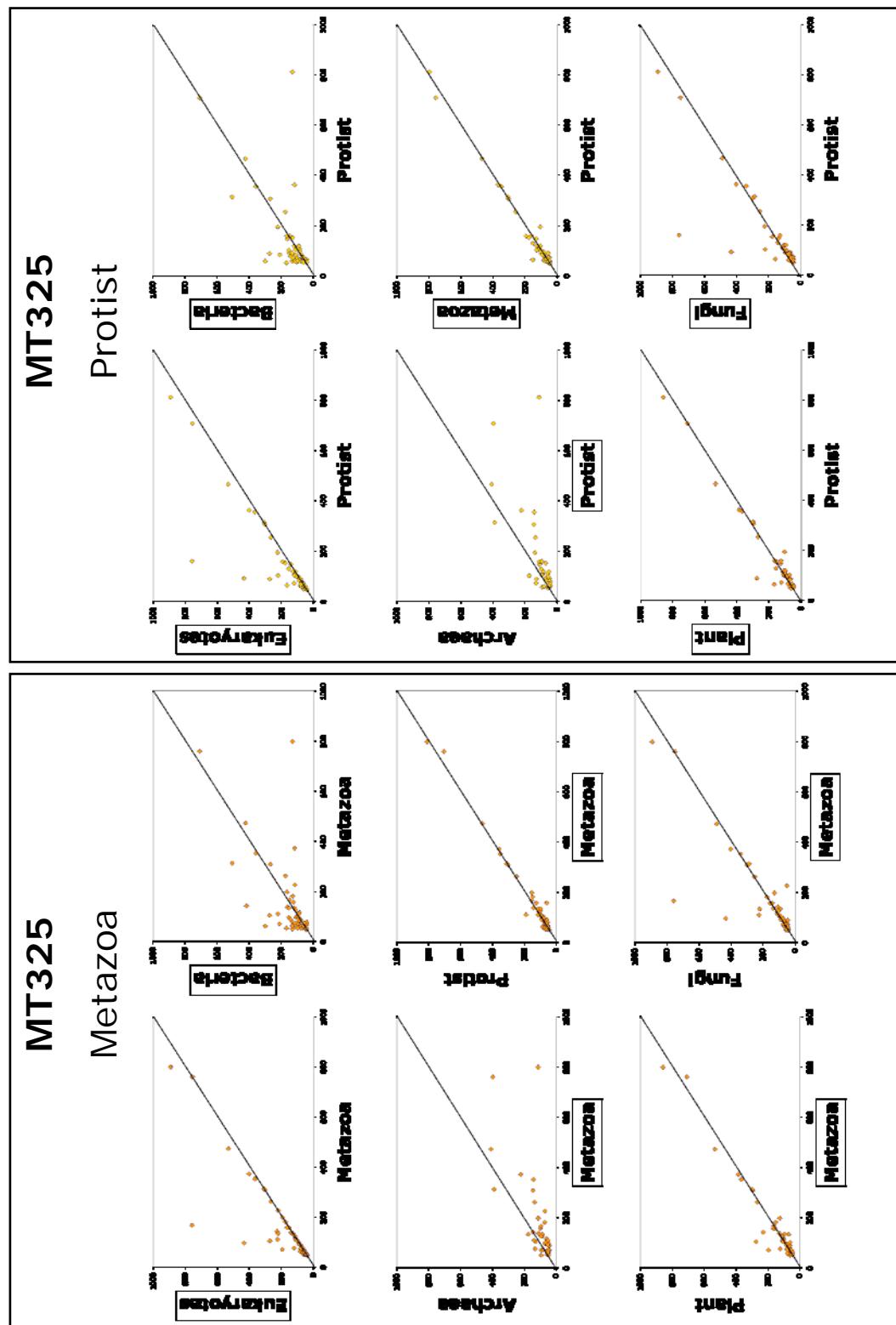


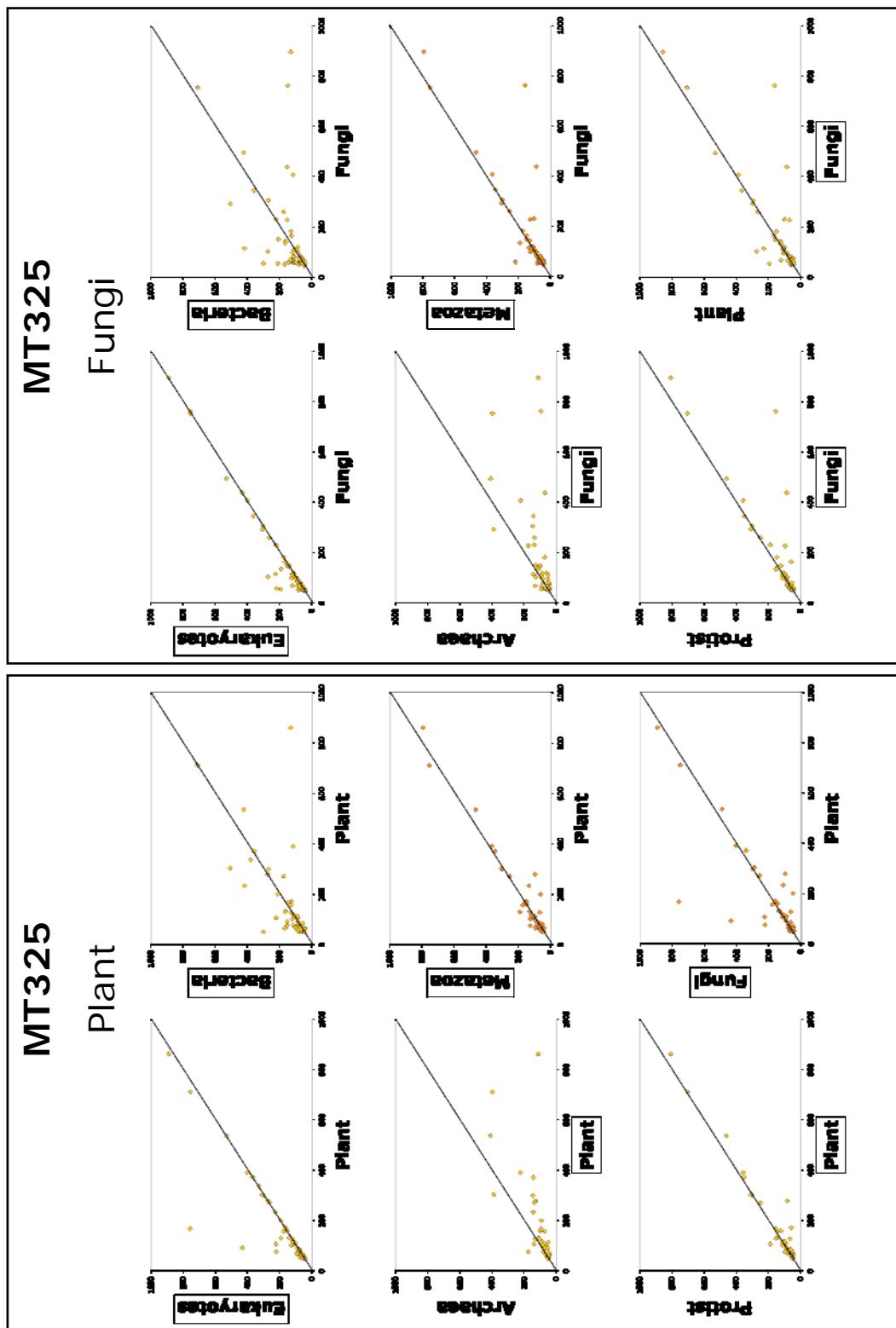


3.6.2.3 MT325 Domain and Kingdom Scatter Plots









CHAPTER 4

COMPARATIVE ANALYSES OF CHLOROVIRUSES, *PHYCODNAVIRIDAE*, AND NUCLEO-CYTOPLASMIC LARGE DNA VIRUSES

4.1 INTRODUCTION

The family *Phycodnaviridae* is composed of a diverse group of large icosahedral, dsDNA viruses that infect algae. The genomes of these viruses range from 160 to 560 kb. The family consists of six genera which were based initially on host range. This classification was later supported by sequence comparisons. The family is monophyletic with branches for each genus, but the phycodnaviruses have evolutionary roots connecting them with several other families of large DNA viruses, referred to as the nucleo-cytoplasmic large DNA viruses (NCLDV). The phycodnaviruses have varied genome structures, some with large regions of non-coding sequence and others with regions of single-stranded DNA. Typically, phycodnaviruses encode hundreds of genes. Genome analyses have revealed an excess of 1000 unique genes, with only 10 homologous genes, which have been assigned a putative function, found among all three genera of the phycodnaviruses sequenced to date. Thus, the gene diversity of these viruses far exceeds their number of “core” genes. While little is known about the replication of these viruses, their infection of phytoplankton have serious global consequences, such as alterations in geochemical cycling and weather patterns. Six genomes of viruses from three genera of the family *Phycodnaviridae* have been sequenced, and, in this chapter, I compare those sequenced genomes and genes that they

encode. Following this comparison, I will also perform a comparative study of the phycodnaviruses and the NCLDVs.

4.2 BACKGROUND

4.2.1 Characteristics and Taxonomy of Sequenced Phycodnaviruses

The genome of at least one member in three genera of the family *Phycodnaviridae* (*Chlorovirus*, *Coccolithovirus*, and *Phaeovirus*) has been sequenced. The most characterized genera, the *Chloroviruses*, have been discussed in detail in previous chapters and, therefore, will not be described here.

4.2.1.1 Coccolithoviruses

The *Coccolithovirus* genus includes viruses that infect the marine coccolithophorid, *Emiliania huxleyi*. The *Emiliania huxleyi* type 86 virus (EhV-86) is the prototype member of this genus. The genome sequence of EhV-86 was reported in 2005 and represents the largest phycodnavirus genome sequenced to date (1). The circular EhV-86 genome is 407 kb in length and is predicted to contain 472 protein-encoding genes, only 66 (14%) of which have detectable homologs in Genbank (1). Unexpectedly, the virus contains several genes never detected before in a virus, such as four gene products involved in sphingolipid biosynthesis (sterol desaturase, serine palmitoyltransferase, transmembrane fatty acid elongation protein, and lipid phosphate phosphatase) and two genes which encode desaturases. EhV-86 also encodes 8 proteases.

The EhV-86 genome has a G+C content of 40% and has three distinct families of repeats of non-coding and coding sequences (1). One family is proposed to be involved in gene regulation. A second family is proposed to be involved in calcium binding which

may play a role in virus adsorption and/or release. The third family is proposed to be involved in the initiation of DNA replication.

4.2.1.2 Phaeoviruses

The *Phaeovirus* genus includes viruses that infect marine filamentous brown algae in the class Phaeophyceae. The *Ectocarpus siliculosus* virus (EsV-1) is the prototype member of this genus. Its genome sequence was reported in 2001 and was the largest sequenced virus at that time. The EsV-1 genome is 335 kb in length and has a 52% G+C content with low levels of methylated cytosines (1% 5mC) and adenines (3% 6mA) (2, 3). It contains several regions of tandem repeats which make up ~12% of the total genome (2). In addition to these repeats, the EsV-1 genome also contains several regions of single stranded DNA of various lengths (4, 5). The structure of the EsV-1 genome is still unclear. Pulsed-field gel electrophoresis, DNA restriction mapping, and electron microscopy experiments indicate that it is a circular molecule (4, 6, 7). However, DNA sequencing results generated one 335 kb linear contig with termini containing almost perfect inverted repeats of 1.8 and 1.6 kb (2). Attempts to link the termini (i.e., to complete the circle) with PCR using primers specific for the inverted repeat regions have been unsuccessful. Therefore, the inverted repeats appear to represent the ends of a linear molecule. These conflicting results can be explained if the complementary inverted terminal repeats anneal with each other to generate a cruciform structure that effectively closes the DNA circle (2).

Analysis of the EsV-1 genome identified 231 major ORFs of 65 codons or more. Forty-eight percent of these major ORFs had homology to other proteins in the public databases. EsV-1 differs from other algal viruses because it encodes six putative hybrid

His-protein kinases, including one that is located in the virion internal membrane (2).

Hybrid His-protein kinases are members of a large family of proteins involved in two signaling systems that serve as stimulus-response coupling mechanisms in many organisms (8). Because of the life cycle of EsV-1, its not surprising that its genome also encodes proteins involved in viral integration/transposition (2). Additionally, EsV-1 encodes several proteins involved in DNA replication, nucleotide metabolism, transcription, and sugar/lipid metabolism.

In 2003, the genomic sequence of a second phaeovirus, *Feldmannia irregularis* (FirrV-1), was published (9). The FirrV-1 genome consists of a linear 180 kb DNA molecule and many smaller DNA fragments of 10-170 kb. One hundred fifty-six ORFs have been identified in the FirrV-1 genome, and 93 of these are structurally related to genes in the EsV-1 genome (9). While well over one-third of the FirrV-1 genes have no homologs in the EsV-1 genome, the EsV-1 genome contains about 75 more genes than the FirrV-1 genome.

The genomic sequencing of several other phycodnaviruses is either currently underway or will soon be. The previously sequenced genomes have provided many unexpected results and have identified a number of proteins unique to this viral family. Therefore, the sequencing of additional genomes of genera representatives could potentially reveal clues toward the evolutionary source(s) of the family *Phycodnaviridae*.

4.2.2 Nucleo-Cytoplasmic Large DNA Viruses

In 2001, a comprehensive comparative analysis of the proteins encoded by large eukaryotic DNA viruses revealed evidence of a common ancestry. These viruses

comprised a monophyletic assemblage and were called Nucleo-Cytoplasmic Large DNA Viruses (NCLDV). This group was originally composed of four virus families: poxviruses, asfarviruses, iridoviruses, and phycodnaviruses (10). They received their name because of the location of their replication and the fact that they have large dsDNA genomes. NCLDVs either replicate exclusively in the host cell's cytoplasm or begin their life cycle in the host cell's nucleus but complete the process in the cytoplasm. These viruses do not significantly depend on their host's replication or transcription systems. Such independence is consistent with the fact that all of these viruses encode several conserved proteins associated with key life-cycle processes, for example DNA polymerases, helicases and DNA clamps for DNA replication, transcription factors involved in transcription initiation and elongation, ATPase pumps for DNA packaging, and chaperones involved in the capsid assembly (10). Since viral proteins tend to evolve faster than their cellular counterparts, the identification of homologous relationships can sometimes be difficult to detect. Therefore, this study analyzed sets of orthologous rather than homologous viral proteins. Once a set of orthologs was detected within the NCLDVs, they were analyzed to determine their relationship, if any, to non-NCLDV proteins. Using this system, genes which were conserved within the NCLDV families but were not found in non-NCLDV-encoding genes were classified as "core genes." These core genes were subsequently divided into 4 groups: Group I (genes conserved in all NCLDV), Group II (genes conserved in all four families of NCDLV but missing in one or more lineages with families), Group III (genes conserved in three families of NCLDV), and Group IV (genes conserved in two families of NCLDV). This study identified three classes of clusters. The first class contains 9 clusters of apparent

orthologs shared by all NCLDVs. The second class, which has 8 clusters, is represented in all families (although missing in one or more species). The last, and least conserved class is composed of 14 clusters that are conserved in all but one family (10). These results suggest that all NCLDV families have descended from a common ancestor that had a complex gene repertoire and was capable of independently completing its replication cycle. Recently, the NCLDV core genes have been re-examined in light of the increase in available NCLDV genomic sequences, including the largest virus sequenced to date: the Mimivirus (11). The mimiviridae has been added to the original four families of the NCLDVs. This new analysis has identified an additional 10 core proteins, resulting in a total of 41 proteins traceable to the last common NCLDV ancestor (Iyer, *et. al.*, in press).

4.3 EXPERIMENTAL METHODS

4.3.1 Comparisons of Phycodnavirus Metabolic Domains

To identify potential homologies between the sequenced phycodnavirus genomes, their ORFs were examined. The classification of the ORFs were divided into their metabolic domains, and blastp was used to determine homology, E-value < 0.01.

4.3.2 Analysis of NCLDVs Core Genes

The identification of the NCLDV core genes was based on the known PBCV-1 core genes (10). Using the NCBI database, the PBCV-1 core genes were used to determine the degree of conservation within the NCLDVs. The program blastp, using an E-value < 0.1, was performed on the different NCLDV viral families. This analysis

provided information on the degree of homology, or the lack of homology, of the various NCLDV families' core genes with the PBCV-1 core genes. Only the top hits within a viral family were used for the analysis.

4.4 RESULTS AND DISCUSSION

4.4.1 Comparisons of the Sequenced Phycodnavirus Genomes

The genomes of the six sequenced phycodnaviruses from three different genera (the *Chloroviruses* PBCV-1, NY-2A, and MT325, the *Phaeovirus* EsV-1 and FirrV-1, and the *Coccolithovirus* EhV-86) each have distinguishing structural and functional characteristics (Table 4.1). One common characteristic is their large dsDNA genomes. The chlorovirus genomes are linear molecules with hairpin ends. EsV-1 has a linear genome with apparent sticky ends and FirrV-1 has a linear genome with the ends uncharacterized. The EhV-86 genome is circular. The chloroviruses and EhV-86 replicate autonomously as episomes, while EsV-1 and FirrV-1 integrate into the host chromosome and re-activate in response to developmental cues.

Table 4.1 Genome data of sequenced phycodnaviruses						
Genus	Virus	Host	Genome Size (bp)	% G+C Content	Major ORFs	tRNAs
Chlorovirus	PBCV-1	<i>Chlorella</i> NC64A	330,743	40.0	366	11
	NY-2A	<i>Chlorella</i> NC64A	368,683	40.7	394	7
	MT325	<i>Chlorella</i> Pbi	314,335	45.3	333	10
Phaeovirus	EsV-1	<i>Ectocarpus siliculosus</i>	335,593	51.7	231	0
	FirrV-1	<i>Feldmannia irregularis</i>	191,667	ND	ND	ND
Coccolithovirus	EhV-86	<i>Emiliania huxleyi</i>	407,339	40.2	472	5

ND-not determined

Several common features exist between two of the three genera of the phycodnaviruses. First, EsV-1 and the chloroviruses genome sizes are similar (330-335 kb), whereas the EhV-86 genome is larger (407 kb). Second, the chloroviruses and EsV-1 genomes are linear and contain inverted terminal repeats. In the PBCV-1 genome, the inverted repeat regions (2.2 kb) are 100% identical. In contrast, in the EsV-1 genome, the terminal repeats consist of short non-identical regions. Third, methylated bases are common in chlorella virus genomes, but are present at low levels in the EsV-1 genome (4). The methylation status of the EhV-86 genome is unknown.

There are three features common to both the EhV-86 and the chlorovirus genomes. First, both the EhV-86 and the chlorovirus genomes encode tRNAs while, the phaeoviruses do not. The chlorovirus tRNAs are clustered together in the center of the genomes. There are 7 to 11 tRNAs per chlorella virus genome. The EhV-86 genome encodes 5 tRNAs. Four of these tRNAs are located in the first 4 kb (10%) of the genome, and one tRNA is located in the middle of the genome. Second, both the chlorella virus and EhV-86 genomes have a G+C content of 40-45%, a value which is lower than the 52% G+C content of the EsV-1 genome. Third, both the chlorovirus and the EhV-86 genomes contain introns. The PBCV-1 genome contains three introns, the EhV-86 genome has two introns, and the EsV-1 and FirrV-1 genomes have no known introns.

The genomes of two marine algal viruses, EsV-1 and EhV-86, share one common feature. Both genomes contain large regions of repeated sequences in addition to the EsV-1 genome terminal repeats. The EsV-1 repeat regions comprise ~12% of the genome. An additional 22% of the EsV-1 genome apparently lacks coding potential. Therefore, only ~67% of the EsV-1 genome contains protein-encoding genes. As

mentioned above, the EhV-86 genome contains three families of distinctly different repeated sequences that are found throughout its genome (12).

Comparison of the phycodnavirus genomes and the genes which they encode demonstrates that there is no co-linearity of the common genes found in the three genera of viruses. However, this finding is not surprising given that very little co-linearity exists between two viruses of the same genera that infect different isolates of chlorella (*Chlorella* NC64A and *Chlorella* Pbi).

4.4.2 Comparisons of Phycodnavirus Gene Content

The genomes of EhV-86, the phaeoviruses, and the chlorella viruses encode many putative proteins based on sequence homologies detected in public databases (2, 13, 14). However, many of their ORFs do not have any detectable homologies in the public databases and are, therefore, classified as unknowns. One hundred forty-nine putative proteins encoded by these genomes were arranged by their metabolic domains (Figure 4.1). Some ORFs encode proteins with the same functions, but their evolution and origin are unclear. For example, the large subunit of replication factor C is present in the genomes of the chloroviruses and phaeoviruses, yet the sequence similarities of these ORFs fall below the threshold ($E = 0.01$) of the BLAST search results. Nevertheless, Figure 4.1 represents a functional grouping of these proteins, and, in most cases, the relationships are homologous.

Only 10 (~7%) of these common proteins are encoded by all six phycodnaviruses. It is not surprising that many of these 10 common proteins are involved in important functions in the viral life cycle (i.e. δ -DNA polymerase, the large and small subunits of

ribonucleotide reductase, PCNA, superfamily III helicases, and the newly recognized archaeo-eukaryotic primases) (10). Another common ORF among all six phycodnaviruses is one which encodes the major capsid protein. These ORFs have amino acid sequence identities of 26-30%. Overall, the amino acid identities among the common protein homologs range from 19-53%.

The chlorella viruses and EsV-1 encode 28 ORF homologs that are absent in EhV-86. Seven of these homologs are listed in Figure 4.1, including two homologs which are conserved in the three chloroviruses and the two phaeoviruses. The remaining 21 homologs have unknown functions. The average amino acid sequence identity among the unknown ORFs is 30% with a range of 22-37%. The chlorella viruses and EhV-86 have 12 homologs that are absent in EsV-1. Interestingly, most of the chlorella virus and EsV-1 homologs have unknown functions while, only two of the 12 chlorella virus and EhV-86 homologs have an unknown function.

One unexpected outcome of the phycodnavirus sequencing projects was the discovery of several virus-encoded gene products which could contribute to DNA rearrangements and/or transposition. The chlorella viruses encode proteins that resemble bacterial transposases. In addition, these viruses encode several ORFs that contain motifs found in homing endonucleases (15). For example, the recently sequenced chlorovirus NY-2A genome encodes 7 putative transposases and 33 putative homing endonucleases. EsV-1 also encodes two ORFs that resemble bacterial transposases and two ORFs with a homing endonuclease motif. EhV-86 encodes one ORF that resembles a homing endonuclease.

The EsV-1 genome has three large dispersed repeat regions (termed R1, R2, and R3) that contain ORFs which are not considered to be genes because they lack upstream AT-rich sequence elements. These ORFs do not match any proteins in the public databases. Interestingly, the R1 and R2 repeats are located downstream of genes encoding putative bacterial IS4 family transposases [IS, insertion sequence] (16). The R1- and R2-transposase units are bracketed by imperfect inverted repeats of 20 bp that resemble the ends of bacterial IS4 elements. Furthermore, R1 and R2 frame a genomic ~29 kb segment which forms a transposon-like structure. Like bacterial transposons, the EsV-1 transposon could confer advantages to its host.

Therefore, sequence analysis of the phycodnavirus genomes revealed several proteins that could be involved in DNA rearrangements, either within or between viral genomes. These enzymes could contribute to the apparent plasticity of the viral genomes. However, in our laboratory, the chlorella virus genomes are stable with only a few spontaneous mutants occurring. However, it should be noted that none of the putative phycodnavirus transposases and homing endonucleases have been tested for function.

Diversity appears to be a reoccurring theme in the genomes of the members of the family *Phycodnaviridae*. The phycodnaviruses have different life cycles (lytic vs. lysogeny), genomic structures (linear vs. circular), and location of hosts (marine vs. fresh waters). Apart from these differences, there are also several similarities. For example, all phycodnaviruses infect algal hosts and they all have large dsDNA genomes. Regardless of these similarities and differences, the sequenced genomes of a few of these viruses have provided valuable information that is contributing to our understanding of

the relationships of the virus members within the family as well as with other large DNA viruses and cellular organisms.

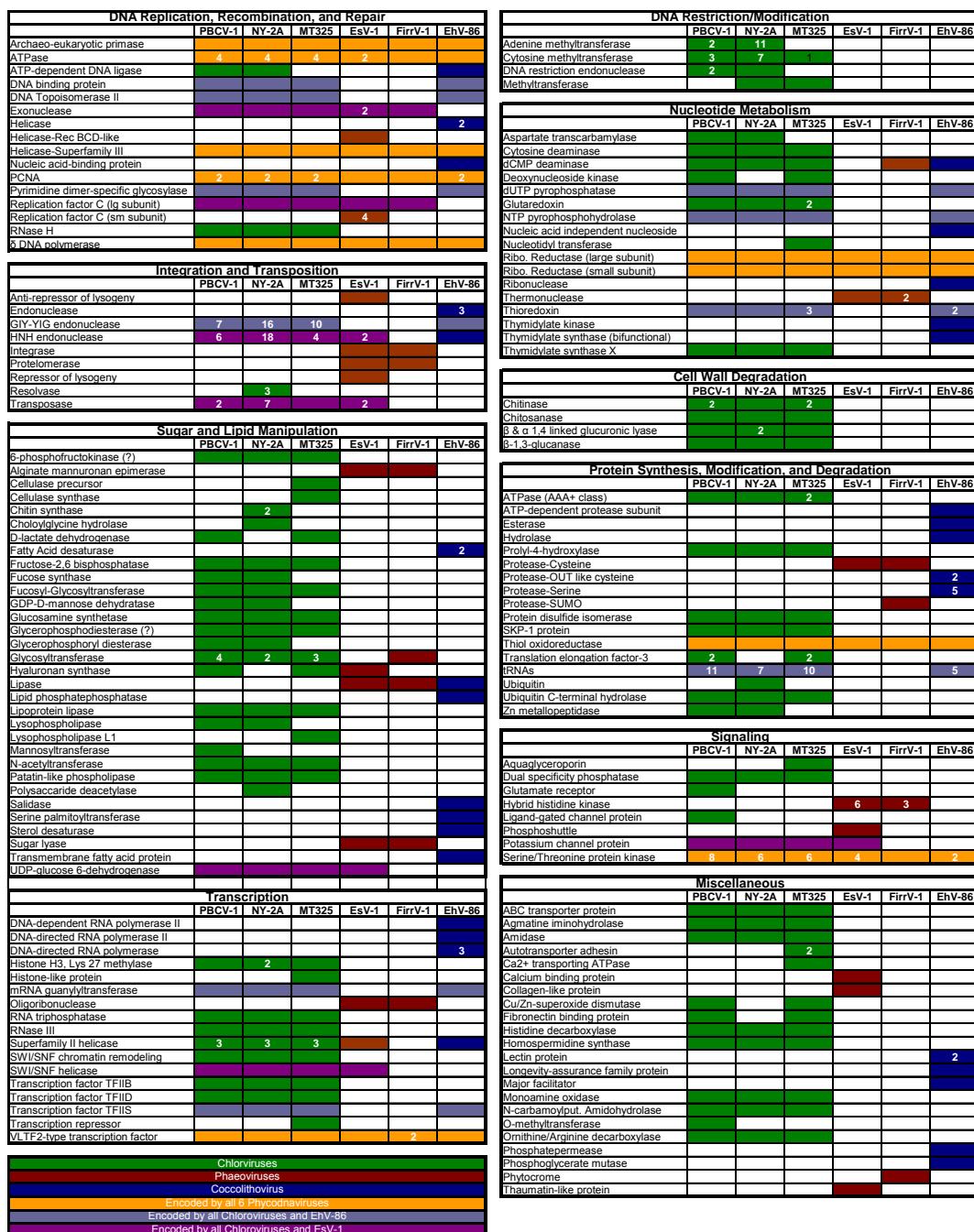


Figure 4.1 Selected ORFs in the sequenced phycodnavirus genomes arranged by their metabolic domains. If a genome encodes a putative protein more than once, the value in the box indicates the number of genes of this type per genome. Color-coding is indicated on the figure and is used to depict the relationship between the viruses. Green, Red, and Blue indicates proteins that are encoded by the genus *Chlorovirus*, *Phaeovirus*, and the *Coccolithovirus* EhV-86, respectively. Orange indicates proteins that are encoded by all six sequenced phycodnaviruses; Teal indicates proteins that are encoded by the chloroviruses and EhV-86; and Purple indicates proteins that are encoded by the chloroviruses and EsV-1.

4.4.3 NCLDV Comparative Studies

In 2001, a class of dsDNA viruses that were evolutionarily conserved was recognized. This group of viruses, now called NCLDVs, consisted of four viral families: poxviruses, asfarviruses, iridoviruses, and phycodnaviruses. Iyer and colleagues reported a comprehensive comparative study of the sequenced NCLDV genomes. Orthologous core proteins were identified, and 31 of these proteins were present in at least three of the four viral families examined. This original study included only two phycodnaviruses, PBCV-1 and EsV-1. However, since this study was published, the genomes of several additional NCLDVs have been sequenced. There are now four more phycodnavirus genomes which have been sequenced. These include NY-2A, MT325, FirrV-1, and EhV-86. Furthermore, in 2004, the largest virus sequenced to date, the 1.2 Mb Mimivirus, was published. This virus is the only member of a new viral family, the *Mimiviridae*, which is now the fifth viral family of the NCLDVs. With the addition of these new NCLDV viral genomes, a re-examination of the NCLDV orthologous core proteins was performed.

The new comparative analysis of the orthologous core proteins was based on the original core proteins that were identified in the best-characterized phycodnavirus, PBCV-1. However, instead of comparing the orthologous proteins identified within the genomes of the NCLDVs, we analyzed the putative homologous relationships that these proteins have with their PBCV-1 homologs.

From the original analysis, a PBCV-1 protein was identified in every Group I and II ortholog, half of the Group III orthologs, and 11 of the 30 Group IV orthologs, for a total of 37 core proteins. The current analysis compared these 37 PBCV-1 core proteins

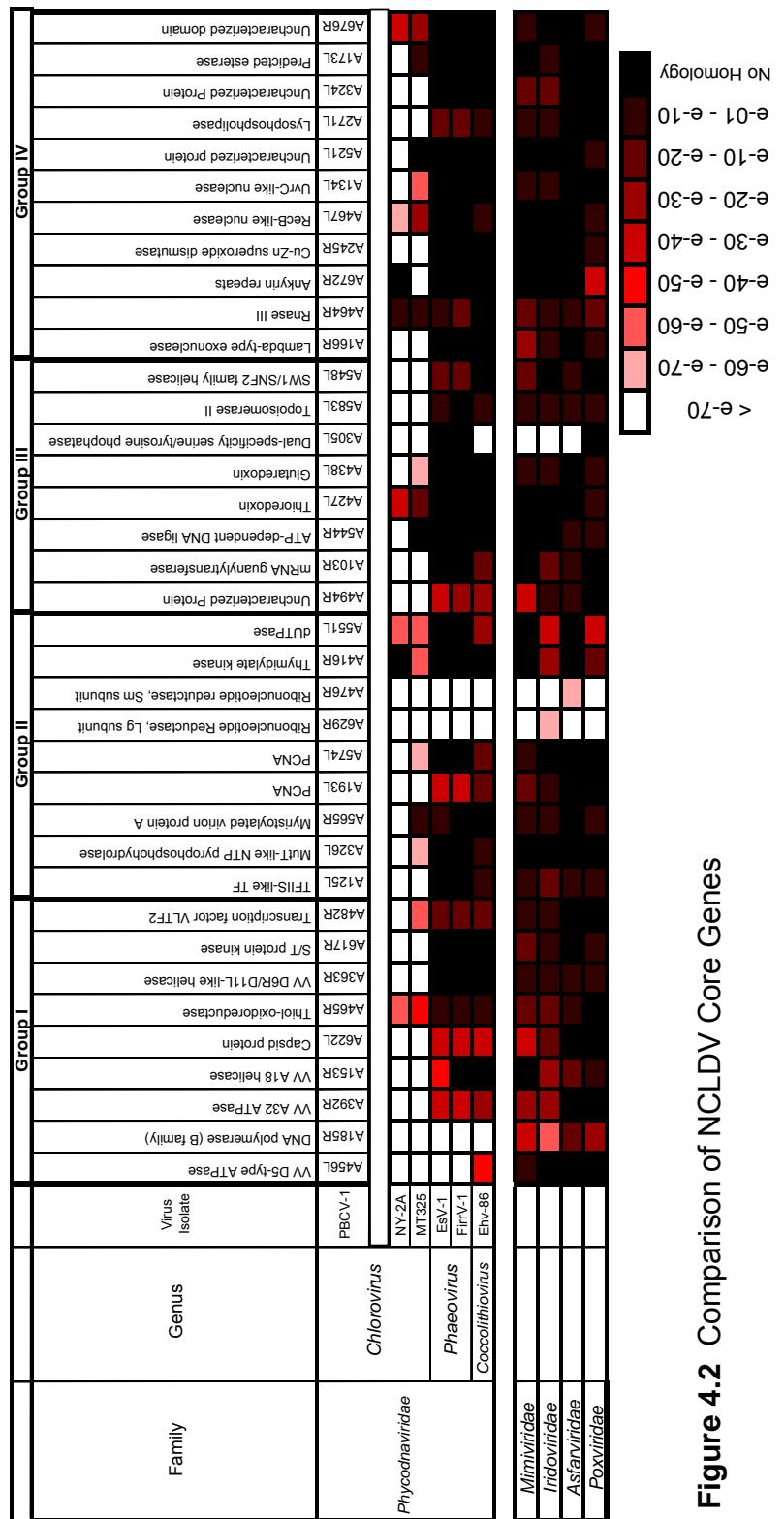


Figure 4.2 Comparison of NCLDV Core Genes

with proteins from the other phycodnaviruses and the 4 other NCLDV viral family members. Only the top hit with either a phycodnavirus or viral NCLDV family member is reported, and its E-value is noted in Figure 4.2. A spectrum of red denotes the homology that a protein from another NCLDV family member has with the PBCV-1 core protein, with a white box showing the highest degree of homology and a dark maroon box showing the lowest degree of homology. The black boxes represent no homology detected with the PBCV-1 protein.

The results presented in Figure 4.2 are not surprising. With few exceptions, there was a high degree of similarity among the chloroviruses (NY-2A and MT325 with PBCV-1). The highest degree of the similarity in the chloroviruses was observed with the NY-2A virus, a chlorella virus which infects the same host as PBCV-1, followed by the MT325 virus, a chlorella virus which infects a different host than PBCV-1. The degree of similarity, although still significant, decreases when comparing the *Phaeoviruses* and the *Coccolithovirus* with PBCV-1. However, a significant decrease in similarity is detected with the other 4 families of the NCLDVs. In fact, if the original classifications of the NCLDV groups were based on homologs rather than orthologs, several core proteins would have been overlooked.

These NCLDV comparative studies illustrate a different degree of similarity between the NCLDVs relative to the PBCV-1 core proteins. The phycodnaviruses, as expected, have a higher degree of similarity to the PBCV-1 core proteins than do the other NCLDV families. For example, the DNA polymerases from the phycodnaviruses have a higher degree of similarity than the DNA polymerases from the other NCLDVs, a finding which confirms that this family forms a monophyletic group apart from the other

families. Based on the color scheme, it is easy to observe that, of the remaining NCLDV families, the mimiviridae has the highest degree of similarity with the phycodnaviruses, followed by the iridoviruses, asfarvirus, and poxviruses. This result is in agreement with a phylogenetic tree of the NCLDVs built based on eight conserved proteins, the NCLDV group I proteins.

In conclusion, the family *Phycodnaviridae* is a genetically diverse group of viruses which infect both fresh and marine water eukaryotic algae. Six phycodnavirus genomes have been sequenced, and this information is contributing to our understanding of the relationships of these viruses with other family members as well as with the other NCLDVs. Sequence analyses of the phycodnaviruses suggest that its family members possibly display more sequence diversity than do members of any other virus family. This viral family consists of six genera that were based initially on host range. However, phylogenetic comparisons with the DNA polymerase supports the classification of the six genera. From the six sequenced genomes of the phycodnaviruses, there are over 1000 unique genes, only 10 putative functional proteins are homologous among the three genera. Despite these differences, the family is monophyletic with branches for each genus. The phycodnaviruses are evolutionarily related to a family of large DNA viruses (the NCLDVs) as shown by comprehensive comparative analyses. Every sequenced phycodnavirus has provided unexpected results. This is largely due to the diversity of this viral family which provides a wealth of unique proteins, some which are the first of their type to be detected in a virus. As a result, genomic sequencing of representatives of additional genera will produce many more new and interesting surprises.

4.5 REFERENCES

- (1) Wilson, W. H., Schroeder, D. C., Allen, M. J., Holden, M. T., Parkhill, J., Barrell, B. G., Churcher, C., Hamlin, N., Mungall, K., Norbertczak, H., Quail, M. A., Price, C., Rabbinowitsch, E., Walker, D., Craigon, M., Roy, D., and Ghazal, P. (2005) Complete genome sequence and lytic phase transcription profile of a Coccothovirus. *Science* 309, 1090-2.
- (2) Delaroche, N., Müller, D. G., Bothe, G., Pohl, T., Knippers, R., and Boland, W. (2001) The complete DNA sequence of the *Ectocarpus siliculosus* virus genome. *Virology* 287, 112.
- (3) Lanka, S. T. J., Klein, M., Ramsperger, U., Müller, D. G., and Knippers, R. (1993) Genome structure of a virus infecting the marine brown alga *Ectocarpus siliculosus*. *Virology* 193, 802.
- (4) Lanka, S. T. J., Klein, M., Ramsperger, U., Müller, D. G., and Knippers, R. (1993) Genome structure of a virus infecting the marine brown alga *Ectocarpus siliculosus*. *Virology* 193, 802-811.
- (5) Klein, M., Lanka, S., Müller, D., and Knippers, R. (1994) Single-stranded regions in the genome of the *Ectocarpus siliculosus* virus. *Virology* 202, 1076.
- (6) Kapp, M. (1998) Viruses infecting marine brown algae. *Virus Genes* 16, 111-117.
- (7) Müller, D. G., Kapp, M., and Knippers, R. (1998) Viruses in marine brown algae, in *Adv. Virus Res.* pp 49-67.
- (8) Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2000) Two-component signal transduction. *Annu. Rev. Biochem.* 69, 183-215.
- (9) Delaroche, N., Boland, W., Müller, D. G., and Knippers, R. (2003) Comparisons of two large phaeoviral genomes and evolutionary implications. *J Mol Evol* 57, 613-22.
- (10) Iyer, L. M., Aravind, L., and Koonin, E. V. (2001) Common origin of four diverse families of large eukaryotic DNA viruses. *J. Virol.* 75, 11720.

- (11) Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Claverie, J. M. (2004) The 1.2-megabase genome sequence of Mimivirus. *Science 306*, 1344-50.
- (12) Allen, M. J., Schroeder, D. C., and Wilson, W. H. (2005) Preliminary characterisation of repeat families in the genome of EhV-86, a giant algal virus that infects the marine microalga *Emiliania huxleyi*. *Arch. Virol. in press*.
- (13) Wilson, W. H., Schroeder, D. C., Allen, M. J., Holden, M. T., Parkhill, J., Barrell, B. G., Churcher, C., Hamlin, N., Mungall, K., Norbertczak, H., Quail, M. A., Price, C., Rabinowitsch, E., Walker, D., Craigon, M., Roy, D., and Ghazal, P. (2005) Complete genome sequence and lytic phase transcription profile of a Coccothovirus. *Science 309*, 1090-1092.
- (14) Li, Y., Lu, Z., Sun, L., Ropp, S., Kutish, G. F., Rock, D. L., and Van Etten, J. L. (1997) Analysis of 74 kb of DNA located at the right end of the 330-kb chlorella virus PBCV-1 genome. *Virology 237*, 360-377.
- (15) Kowalski, J. C., Belfort, M., Stapleton, M. A., Holpert, M., and Dansereau, J. T. (1999) Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. *Nucleic Acids Res.* 27, 2115.
- (16) Rezsohazy, R., Hallet, B., Delcour, J., and Mahillon, J. (1993) The IS4 family of insertion sequences: evidence for a conserved transposase motif. *Mol. Microbiol.* 9, 1283-1295.