Expression Inactivating Mechanisms of Major Histocompatibility Class I Alleles in *Melanogrammus aeglefinus*

by

Colleen Ryan

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Biology

Waterloo, Ontario, Canada, 2010

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Melanogrammus aeglefinus, commonly known as haddock, is a commercially important marine fish species closely related to cod. Preliminary investigations into the immune function of this species has revealed several unique and interesting features, including an unusually high number of expressed alleles of Major Histocompatibility (MH) Class I genes. The goal of this project was to examine the sequences of alleles, including the untranslated regions, for potential regulatory mechanisms which may limit the number of alleles expressed to the point of functional molecules.

Using a cDNA library from an individual haddock, a total of 22 unique alleles were isolated and sequenced, and three putative mechanisms for limiting expression were revealed. The first mechanism was the inversion of the open reading frame within the transcript. The second mechanism was the linking of the MH Class I transcript with the transcript of another gene. The third mechanism was non-classical substitutions at the nine amino acid residues involved in peptide anchoring. These three mechanisms represent novel ways of limiting expression and effectively reduced the number of alleles which could be expressed into functional classical MH Class I molecules.

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Brian Dixon for the opportunity to work on such an interesting project, as well as the direction, knowledge and support it took to help me complete this degree. I would also like to thank Dr. Kazuhiro Fujiki, who initiated work on haddock and isolated the first MH Class I sequences which were essential to the entire project, and Stephen DeBoer, who isolated additional clones. The other graduate students of the Dixon Lab and several summer and undergraduate students deserve thanks as well, for their support, advice, assistance and friendship which made this experience so enjoyable. Last, but not least, I would like to thank my family and friends, whose love and support was essential to me while I worked on this degree. In particular, I would like to thank my husband Rob, who patiently supported and encouraged me through the unexpected delays and missed deadlines to watch me immerge victorious in the end.

Table of Contents

List of Tables

List of Figures

1. Introduction

Melanogrammus aeglefinus, commonly known as haddock, is an economically important species of fish. Currently, haddock is caught in the wild rather than raised. Attempts to begin farming these fish are being made (Chambers and Howell, 2006; Treasurer et al., 2006), and one of the biggest hurdles to overcome when farming any species is disease (Bricknell et al., 2006). As a result, research investigating the immune function of these fish, as well as other related species, is increasing in importance. Many aspects of the fish immune system are in-line with the general features commonly seen. However, haddock demonstrate some striking differences from not only the well-studied mammalian immune system, but also from other fish species (Miller et al., 2002; Persson et al., 1999).

The focus of this research is on the Major Histocompatibility (MH) Class I molecules in haddock. Unlike most species, haddock have an unusually high number of alleles (Fujiki, unpublished data). It is unclear how many alleles are expressed and how their expression is regulated. Cloning and sequencing of the entire transcripts of all MH alleles will shed light on this unique feature of the haddock immune system.

1.1. General Overview of Immune Function

The immune system is a fundamental requirement for all species. It provides protection against invading pathogens as well as maintenance and healing of the body. Immune function jawed vertebrates, can be loosely divided into two interdependent sub-systems: the innate and the adaptive. The innate immune system is the initial line of defense. It has no specific recognition of antigen, and is present at all times. The adaptive immune response, however, is capable of highly specific recognition of antigens, and develops in response to antigen exposure. (Dixon and Stet, 2001; Watts et al., 2001).

The main cells of the adaptive immune system are the B-cells and T-cells. During their development, each of these cell types undergoes similar developmental processes wherein they develop highly specific antigen receptors which are used to drive the antigen-specific immune response (Larosa and Orange, 2008). Once activated, B-cells

are capable of secreting antibody, which is known as the B-cell receptor when bound to the B-cell surface (Schroeder and Cavacini, 2010). T-cells, on the other hand, do not produce a secreted equivalent to the T-cell receptor (TCR). Based on signals received during development in the thymus T-cells can become either cytotoxic T-cells (T_C) or helper T-cells (T_H) (Laky and Fowlkes, 2008; Picchietti et al., 2009). Cytotoxic T-cells, as their name suggests, are involved in killing infected cells (Fischer et al., 2006). Helper T-cells secrete molecules known as cytokines which are necessary to promote and direct other immune processes (Fietta and Delsante, 2009). The two types of T-cells can be easily identified by cell surface proteins. Cytotoxic T-cells have the co-receptor CD8 on their cell surface, whereas helper T-cells have the co-receptor CD4 on their cell surface (Laky and Fowlkes, 2008).

T-cell activation is a complicated process that begins with the interaction of the TCR with its specific antigen, yet TCRs cannot recognize antigen by itself. Rather, they require that the antigen be bound to a specialized molecule known as the Major Histocompatibility (MH) receptor (Costantini et al., 2008; Garcia, 1999). Cells of the body display MH molecule:antigen complexes on their cell surfaces with which T-cells are then able to interact.

1.2. Major Histocompatibility Molecules – Structure and Function

Major Histocompatibility molecules are a fundamental component to the adaptive immune system. As such they have highly conserved structure and function across all species possessing an adaptive immune system. These molecules are well studied in mammalian species and accurate parallels can be drawn between mammals and fish (Sullivan and Kim, 2008). There are, however, some differences. In mammals, and interestingly in the cartilaginous fishes, the genes encoding the MH molecules are contained within a tightly linked sets of genes referred to as the Major Histocompatibility Complexes (MHCs). Within these complexes are several genes, many of which encode products with immune system functions. As a result, MH molecules in these species are referred to as Major Histocompatibility Complex (MHC) molecules. In bony fishes, however, there is no apparent linkage of these genes resulting in the term MH instead

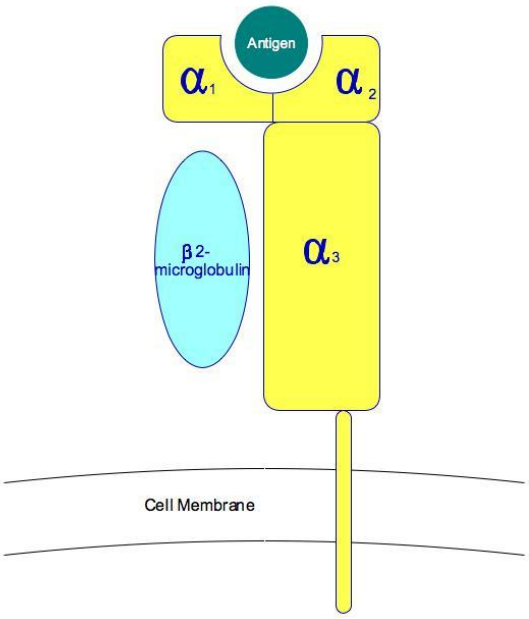
of MHC (Stet et al., 2003). The MHC and MH molecules themselves, however, maintain the same characteristics.

MHC molecules can be divided into two similar yet unique types, referred to as Class I and Class II. Both types of molecules are responsible for presenting peptide antigens on the cell-surface to T-cells. The types of molecules differ in the origin of the peptide antigen they present, the type of T-cell they interact with, and the cell types in which they are expresses. MHC Class I molecules present peptide antigens that are intracellular in origin, for example viral peptides, and interact with the T-cell co-receptor CD8 (Bjorkman and Parham, 1990). MHC Class II molecules present extracellular-derived peptide antigens, such as bacterial components, and interact with the T-cell co-receptor CD4 (Lanzavecchia, 1990; Sultmann et al., 1993). All nucleated cells express MHC Class I molecules, whereas only Antigen Presenting Cells (namely macrophages, dendritic cells and B-cells) express MHC Class II molecules (Dijkstra et al., 2003; Sultmann et al., 1993).

The basic structure of the MHC Class I molecule appears to be conserved across all species with an adaptive immune system (Janeway, 2001; Randelli et al., 2008; Stern and Wiley, 1994). Figure 1.1 shows a diagram of the structural features. The entire structure is composed of two non-covalently linked peptides: a heavy chain comprised of three sub-units α_1, α_2, and α_3, and a smaller peptide called β_2-microglobulin. Sub-units α_1 and α_2 form a “binding cleft” in which the antigen peptide is bound. The amino acid sequence of the binding cleft is in general highly variable, although nine specific residues have been identified as necessary for anchoring the peptide N- and C-termini to the binding grove, and therefore are highly conserved through most species (Shum et al., 1999). As a result of this variability, different MHC molecules are able to bind various peptides with differing affinities. Each MHC molecule is encoded by a unique allele.

As previously mentioned, the role of MHC Class I in the immune response is to present intracellular-derived antigens to CD8 T-cells. This process begins in the endoplasmic

Figure 1.1: Diagram of the structural features of MHC Class I. The complete molecule is made up of two non-covalently linked peptides, the β_2-microglobulin and the heavy chain. The heavy chain is comprised of three domains, α_1, α_2, and α_3. The α_1 and α_2 domains make up the binding cleft where peptide antigen is bound prior to cell surface expression.


Antigen
α1
α2
β2-microglobulin
α3
Cell Membrane

reticulum. Digested peptide fragments are actively transported into the endoplasmic reticulum and are combined, with the help of a series of chaperone proteins, with the immature MHC Class I molecules. These molecules are then transported to the cell surface where they can be “surveyed” by T-cells (Pamer and Cresswell, 1998). T-cell interaction requires that the bound peptide is a “match” for the highly specific TCR and the interaction of the MHC Class I molecule with the co-receptor CD8. Along with other ligand and co-receptor interactions between the two cells, a complex signaling cascade is initiated and results in the activation of the T-cell into an armed effector T-cell (Guselnikov et al., 2003; Nel, 2002; Park et al., 2002). In this case, the T-cells become cytotoxic T-cells, capable of inducing apoptosis of infected cells.

Major Histocompatibility molecules also play a vital role in T-cell selection. In the thymus, T-cells undergo a selection process which removes undesirable T-cells from the population. A review by Takahma outlines the details of this process. To summarize, developing T-cells interact through their TCRs with MHC:self-peptide complexes present on the surface of stromal cells in the thymus cortex. The strength of the interaction determines the fate of the T-cell. A weak interaction signals survival and the T-cells undergo further maturation, whereas a strong interaction signals apoptosis and the T-cells are deleted from the population. This process allows T-cells expressing TCRs capable of interacting with MHC molecules to continue maturing, while removing T-cells expressing TCRs against self-peptides. Only a small proportion of thymocytes survive to become mature T-cells.

In addition to the MHC molecules described above, there exists a group of molecules with similar structures referred to as non-classical MHC molecules. Although this set of molecules is not as well studied as the classical MHC molecules research suggests a diverse array of functions, both immunological and non-immunological. For example, a growing body of research has linked non-classical MHC molecules with tumor immune responses (Gomes et al., 2007). In rainbow trout, a specific non-classical MH molecule has been shown to impact resistance to Bacterial Cold Water Disease (Johnson et al., 2008). Hemochomatosis is an iron-metabolism disorder resulting from a defect in the gene coding for HFE-A, a non-classical MHC molecule (Ehrlich and Lemonnier, 2000).

A key difference between classical and non-classical molecules is the amino acid residues located at the nine conserved sites discussed earlier. Changes at two or more of these sites sufficiently changes the function of these molecules so that they are termed non-classical rather than classical MHC molecules (Shum et al., 1999).

1.3. Unique Immune System Features of Cod and Haddock

Cod and haddock possess several immune system features that are very unlike those of most other species. Natural antibodies, which react non-specifically to a multitude of antigens, are present in most fish species to a varying degree. In cod, however, these antibodies are extremely abundant. Furthermore, cod do not appear to produce any antigen-specific antibodies (Magnadottir et al., 2001). In haddock, no MH Class II genes have yet been identified, despite rigorous attempts. Also, the β_2-microglobulin is also unusual in that it is glycosylated (unpublished data). These findings lead to further study of the apparently very unusual immune system of these fish.

1.4. Allelic Frequency

The MHC Class I molecules are encoded by genes that are highly polymorphic and polygenic, that is to say there are a high number of alleles within the population and each individual possess several copies of the gene. For example, in humans there are over 100 alleles, and each individual has three genes (therefore, six expressed alleles). This results in billions of possible combinations of alleles. The population is therefore equipped with a wide range of alleles capable of participating in immune responses to a wide range of diseases. By increasing the number of expressed alleles, however, the T-cell selection process would become more stringent, leading to an increased number of T-cells being deleted. The dodecaploid species *Xenopus* has been shown to remove sections of chromosomal DNA to reduce the number of excess alleles (Sammut et al., 2002). This demonstrates the importance of limiting the number of expressed alleles.

A paper by Perrson *et al* published in 1999 investigated alleles of MC Class I in Atlantic cod (*Gadus morhua*), a species closely related to haddock. Using a combination of hybridization screening, PCR amplification and Southern blotting, an unusually high

number of unique MH Class I sequences were identified. In most teleost fish species studied an individual typically only has a few alleles of MH Class I. Cod, however, showed 17 genes in one individual. To validate this result, Southern blots of genomic DNA digested with the restriction enzyme *Ssp* I of four individuals were performed and confirmed the presence of many more alleles per individual than is usually seen in other fish species. The experiments performed however relied heavily on PCR, and the untranslated regions of the transcripts were not studied. As a result, possible regulatory mechanisms within the untranslated regions could not be investigated. In 2002, Miller et al. showed further evidence of the unusually high number of MH Class I genes present in cod by isolating an analyzing 83 MH Class I sequences from one individual.

1.5. Previous Work

Prior to the beginning of this project, much of the ground work had already been completed. Southern blotting of haddock DNA and the construction of a cDNA library were completed and screening of the library had begun.

Genomic DNA was collected from an individual haddock and aliquots were digested with four restriction enzymes. A Southern blot was performed, the results of which can be seen in Figure 1.2. The multiple bands seen are comparable with the results obtained from the cod experiments. Therefore, further study began. A cDNA library was constructed using mRNA from liver, spleen and head kidney of an individual haddock. The cDNA library was divided into 18 sub-libraries.

Based on MH Class I sequences from cod a set of degenerate primers were made and the cDNA library was screened using PCR. The products were sequenced and used to generate a probe, which was then used to screen the library. The four MH Class I sequences were initially found are shown in Figure 1.3.

Figure 1.2: Southern blot of haddock DNA for MH Class I genes. Four DNA samples, each digested with a different restriction enzyme, were compared. In each lane several bands are present, indicating the presence of multiple genes.

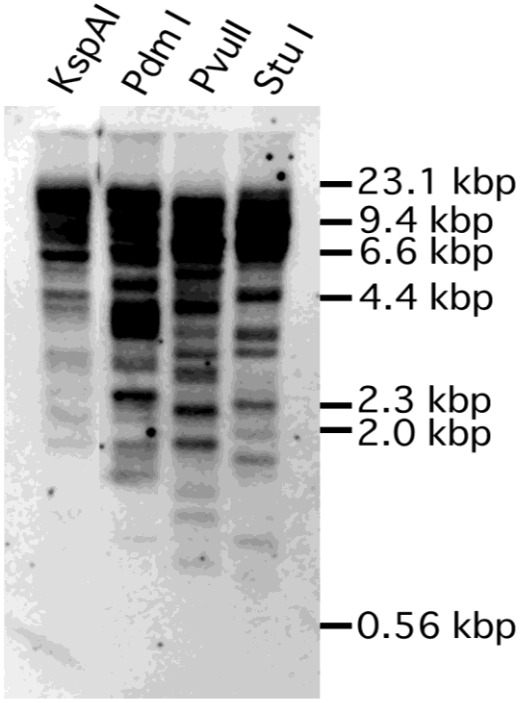

KspAI
Pdm I
PvuII
Stu I
23.1 kbp
9.4 kbp
6.6 kbp
4.4 kbp
2.3 kbp
2.0 kbp
0.56 kbp

Figure 1.3: Haddock MH Class I Sequences. The ORF of each sequence is shown in upper case letters. A) Sequence of Haddock UA pBK-CMV-1. This sequence is comprised of two sequences. The first is the sequence of the ribosomal protein L12. The second is that of the MH Class I sequence. B) Sequence of Haddock UA pBK-CMV-3. The antisense orientation of this sequence codes for an allele of MH Class I. C) Sequence of Haddock UA pBK-CMV-4. This sequence was used as the template for probe synthesis as well and the positive control for experiments. D) Sequence of Haddock UA pBK-CMV-6.

Haddock UA pBK-CMV-1

ggtaaacggtcgtcccgaatctgggttcatccgacaccatcaccaaccaagctgaacagagcactactaccagatcagc
atcatgcctcccaaattcgaccccagcgaaattaaagttgtgtacatgagatgtactggcggggaagttggtgccacctctg
cacttgctcccaaaattggacctctgggtctgtctcccaagaaagttggtgatgacatcgccaaggccaccggagactgg
aagggcctcaggatcaccgtgaagctgaccatccagaacagacaggctgtggttgaggttgttccctctgcatctgccctg
atcatcaaggctcttaaggagcccccacgtgaccggaagaaggtcaagaacatcaagcacagtgggtccgtgacgtttg
acgagatcgttgctatcgctcgtgttatgaggcatcgctctattgccagggaactctctggaaccatcaaggaaatccttggt
acctctcagtcagtaggctgcaccatcgatggccgcctcccccacgatgtcattgatgacatcaacagtggaaagattgaa
tgcccagcagagtaatgcttgacaataaaattgactaaattggagtcggagtcccgtctctgtggataatctcatcgagctgt
tcttctgaatacatcgatATGAAGGCGCTAATAGGGCTGCTGCTGCTGTTGGTCTTTGGTCA
CGGTGTGTCCTCACTGCTTCACTCTCTCCATTATTTCTACACGGGGTCGTCTGGAC
TCTCAACCTTCCCAGAGTTTGTTTCTGTGGGGATGGTGGATGGAGTTCAGTTTGAC
TACTATGACAGCAACACCCAGAAAACAGTCTTCAAACAGGACTGGATGGAGCAGG
TCACCAGAGGTGACCCCGACTACCTGGAGAGGAACACTGGAAGTGCTCAGAATGC
CCAGCAGGTCTTCAAAGTCAACATTGTGACTGCCAAGCAGCGCTTTAACCAGACAG
GAGGTGCCCATATAAATCAGAGGATGTATGGTTGTGAGTGGGATGATGAGGATGA
TTCTACTGAAGGTTATGACCAGCATGGTTATGATGGAGAGGACTTCATATCGTTGG
ACCTGAAGACCCTGACCTGGGTCGCTCCAGTACGTCAGGCCTTCGCCACCAAACA
GAAATGGGATCAGAATAAAGCTAGATTACAACATGATAAGAACTACTACACCAAGG
AGTGTGTTGATTGGCTGAAGAAGCACCTGGCCTATGGGAAGAGCACTCTGCTGAG
AACAGAGCGTCCGCGGGTGTCTCTGCTCCAGAGGAGCCCCTCCTCCCCAGTGGT
GTGCCATGCTACAGGCTTCTTCCCTGACAGGGTGGTGGTGCTCTGGAGGAGAGAC
GGTCAGGAGCTCCAGGAGCAGGTGGACCACGGGGAGGTCCTCCCCAACCACGAC
GGGACCTTCCAGGTCAGCGTGGACCTCAACCTCAAGGCCGTCCCACAGGAGGAC
TGGAGGAGGTACGAGTGTGTGGTCCAGCTGAGAGGCATCGAGGACATCTCCACC
CCCCTGGACCCCGCCCTCATCAGGACCAACGGGGGCAAGAGTGGCATCACCGTC
CCCATCATCATTGGGTTCGTTGTTCTCCTCCTCGCTGCTGCTGCTGCTGCTGTTGG
AGTCCTTCTGTACAAGAAGAGGAACGATTCACACAAGCAGCACAAACCAGTTGGTT
CTTCTGACACCAGCTCTGAGAACACAGAGGGGCAGAATCTGGCTCCTGAGGCCCG
ACCTCTGACAGAAGTTCGAAGTTAAatcatgaagagtcttctcacgttcagcttcctaaagcagggtggag
tggatcgcagttcatttataatccttacaaacaggcttgggacttcagctaactaaagatgctacatttgctaaattacaataa
catccatacttaactttattttgataacaatcatgaatctgaaagatctgtatatcaatgagatctctgtatttgtaatagatttgatt
aattctctggctttaacctgcaatgtttcactgacgcttggaaaccatgacgacacaaactcgccatgatggccattttgttttg
gttcctgccgcttcaagttataacctttgaatcatgttatagtgtttagtaatatttaagtttcaaataatggtcaaaggttctgaga
ctagtttttggacagtgagatccggtcaagtttgaatacaataatcgttaaattgtatgtttttttcttaaataatattatttgaattgca
tcgctacataggctaccacgatcgagctaggcagctttagccttaatccttttgttgtactattttctttgttctttggggaatttgatc
atagatactttggttattctgtcaataaagtacatccaaa

Haddock UA pBK-CMV-3

ggaaatattcatatctttattgagctgtatttgatgtacagttttcttcagacaagcagcaatacacaagagttttatgcactggtgagtatcc
acatacaagtccaacacaacacataataagttgttcatagcacataaagcacatgcgtgttgaatatgtacgtgtccgtgtatgtctttgc
ttgcatgagtttccgcatgcatgtatgaagtgaggtgtgtgtgtgtgtgtgtgtgagtgtgtgtgtgtgagatcaaaagctattgagtcccac
cattctgttgcagagaaagtgcatcaatttggcttcacacaaagaagtagcggtgtgtgtgtgtgtctgtgtgtgtgtgtgtgtgcgtgtgtgt
gtatgtatgtgtgggggtctgagagaagaagggggtaggggggagaaaatgaaggaatgatacataggggggacaaagacggag
agattagtcggcaggaagcgggtggttgtggagcgtgtctagctcctggccatggcggcggcgagggtctcgtagagcatgatgagc
ttggctctggcctcctcgaccaggggtctgccttggccttcagactctccacatgggggacacctgcttcttgatgccttcgatgtccag
ggagccggcggcggcctgggcgtcggtgaaggccttggtcagctgctccttgtactcctccacgtagggggaggctatggtcttgagg
ccctgcaggcgctcggtcagcttgccgcgcacggactccacgaggggcatgagggcggtcttggtctcctccacgttcttggacaccg
tggcggttcagctgcaaagtatctatgatcaaattccccaaagaacaaagaaaatagtacaacaaaaggattaaggtctaaagctgc
ctagctcgatcgtggtagcctatgtagcgatgcaattcaaataatattatttaagaaaaaacatacaatttaacgattattgtattcaaactt
gaccggatctcactgtccaaaaactagtctcagaacctttgaccattatttgaaacttaaatattactgaacactataacatgattcaaag
gttataacttgaagcggcaggaaccaaaacaaaatggccatcatggcgagtttgtgtcgtcatggtttccaagcgtcagtgaaacattg
caggttaaagccagagaattaatcaaatctattacaaatacagagatctaattgatttacagatctttcagattcatgattgttatcaaaat
aaagttaagtatggatgttattgtaatttagcaaatgtagcatctttagttagctgaagtcccaagcctgtttgtaaggattataaatgaactg
cgatccactccaccctgctttaggaagctgaacgtgagaagactcttcatgatTTAACTTCGAACTTCTGTCAGAGGT
CGGGCCTCAGGAGCCAGATTCTGCCCCTCTGTGTTCTCAGAGCTGGTGTCAGAAGAACCA
ACTGGTTTGTGCTGCTTGTGTGAATCGTTCCTCTTCTTGTACAGAAGGACTCCAACAGCAG
CAGCAGCAGCAGCGAGGAGGAGAACAACGAACCCAATGATGATGGGGACGGTGATGCCA
CTCTTGCCCCCGTTGGTCCTGATGAGGGCGGGGTCCAGGGGGGTGGAGATGTCCTCGAT
GCCTCTCAGCTGGACCACACACTCGTACCTCCTCCAGTCCTCCTGTGGGACGGCCTTGAG
GTTGAGGTCCACGCTGACCTGGAAGGTCCCGTCGTGGTTGGGGAGGACCTCCCCGTGGT
CCACCTGCTCCTGGAGCTCCTGACCGTCTCTCCTCCAGAACACCACCACCCTGTCAGGGA
AGAAGCCTGTAGCATGGCACACCACTGGGGAGGAGGGGCTCCTCTGGAGCAGAGACACC
CGCGGACGCTCTGTTCTCAGCAGAGTGCTCTTCCCATAGGACAGGTACTTCTTCAGCCAAT
CAACACACTCCTTGGTGTAGTAGTTCTTATCATGTTGTAATCTAGCTTTATCCTGATTCCATT
TCTGTTTGGTGGTGAAGGCCTGACGTACTGGAGCGACCCAGGTCAGGGTCTTCAGGTCCA
ACGATATGAAGTCCTCTCCATCATAACCATGCTGGTTATAACCATCAGTAGAATCATCCTCA
TCATCCCACTCACAACCATACATCCTCTGAGCCATGTGGGCACCTCCTGTCTGGTTAAAAC
GCTGCTTGGCAATCACAATGTTGACTTTGAAGCTCTGCTGGTGACCCTGAGATTTTCCAGT
TTCCCTCTCCAGGTAGGCGGGGTCATCTCTGGTGACCTGCTCCATCCAGTCCTGTTTGAG
GACCTTTTTCTGGGTGTTGCTGTCATAGTAGTTAAACTGAACTCCATCCACCATCCCCACAG
CAACAAACTCTGGGAAGGTTGAGAGTCCAGACGACGCCGTGTAGAAAAACTGCATAGAGT
GAAGCACTGAGGACACACCGTGACCAAAGACCAACAGCAGCAGCAGCCCTATTAGCGCCT
TCATatcgatgtattcagaagaacagctcgatgagattatccacagaggaagaactcaaccagtgaagatcaaccttctagacatg
gctgcagtgagctttaaactgttggtggagactgctgatggcaacaagagagctgaatacaaacctgcttcgcacgcccgccagtac
ctcggttgggcgttgaatgaatatgaagattcagtctcaaacttcagtttcaaaactgcaggcaaatttggattgccagcctatttgaccc
aggtgacttttgataataacacttttaccaaaactgagaacagcttctggacaattctggtcaaaaaccacaacatggtaacacctgttg
gactctacaactacgttgtccaggacagtgacgtggtcatcctgcgctacacgaccgggtgattctgactgtggcgccatcttcctgcag
gatcctgaattacactgagtgcctctcgacgctcctttaacatgaagattatttagtcttaaaatgtcaaatgttacaaatatgatcaattga
accgatataggtgaaacatataacggaataaaggattgtaattaaaaaaaaaaaaaaaaa

Haddock UA pBK-CMV-4

ggttgcgtctttgtggataatctcatcgagctattcttccgattacatcgatATGAAGGCGCTAATAGGGCTGCT
GTTGTTGGTCTTTGGTCACGATGTGTCCTCAGTGATTCACTCTCTGCAGATTTTCTA
CACGTCGTCTTCTGGACTCTCAACCTTCCCAGAGTTTGTGAGGGTTGAGATGGTGG
ATGAGGTTCAGACTGAGTACTATGACAGCAACACCCAGAGAATCATAACCAAACAG
GACTGGGTGGTCCAGGCCAACAGAGACAAAGTCCCCGACTACCTGGAGAGGGCG
ACTGAAAGCAGAAAGGTTAACCAGCAGGTCTTCAAAGGCAACATTGGGACCCTGA
AGCAGCGCTTTAACCAGACAGGAGGTGCCCACATTGTTCAGTGGATGTCTGGTTG
TGAGTGGGATGATGAGGATGATGCCACTGATGGTTATGAGCAGTATGGTTATGATG
GAGAGGACTTCCTATCGTTGGACCTGAAGACCCTGACCTGGGTCGCTCCAGTACA
TCAGGCTTTCACCACCAAACTGAGATGGGAACACAATAGAGCTTATAATCAAGCCA
AGAACAACTACTACACCAAGGAGTGTGTTGATTGGCTGAAGAAGTACCTGAGCTAC
GGGAAGAGCACTCTGCAGAGAACAGAGCGTCCGTGGGTGTCTCTGCTCCAGAGG
AGCCCCTCCTCCCCAGTGGTGTGCCATGCTACAGGCTTCTTCCCTAATAGGGGGG
TGGTGTTCTGGAGGAGAGACGGCCAGGAGCTCCAGGAGCAGGTGGACCACGGG
GAGGTCCTCCCCAACCACGACGGGACCTTCCAGGTCAGCGTGGACCTGAACCTAA
CGGCCGTCCCACAGGAGGACTGGAGGAGGTACGAGTGTGTGGTCCAGCTGAAAG
ACATCGAGGACATCTCCATCCTCCTGGACCCCGCCCACATCAGGACCAACGGGGC
TAAGACTGGTGTTGGAGGGGATGGAGGGGACAAGAGTCTCATCCTTGCTTTCATC
CTCACTGGTGTTGCTGTTGTTGTTGTTGCTGCTGTTGTTGGAGTCTTTCTGTACCG
GAAGAGGAACGATTCAGACAAGCGTCACAAACCACTTGGTTCTTCTGACACCAGCT
CTGAGAACACTGAGGTGCAGAAGCTGTCTCCTGAGGCCCAACCTCTGACCACAGG
TCAAAGTTAAatcatgaagaggcttgtcacgttcggcttcctaaaccaggatggagtggattgctggtcatttataatcc
ttacaatcaggcttgggacttcagctaactgcagagtctgcatttgctaaattacaattttcaaacttaactttattgtgataaca
atcatcaatgtgttagaagtgtcaatcaattacatctctctatttgtgatatatttgaataactcgctggcttaaacctgcaatcttt
cactgacgcttggaaaccatcgcaacacgacgtcgccatgatggccatttgggttccgccgcttcacgttataacctttaatt
atgttatattgtttagtaatatttaagtttcaaataatgttcaaagattctgaatcaagattttgacagtgagattttgtcaagtttca
atacaataatcgattaatttaaattgtttcattaaaatatattatttagcgttctgcaagctagaggattgcattgctacataggct
accacaatcctgctaggccgctttagcctcaattctttttgttgtattttctgtgttcttttgggaattggatagtatacttgggttattct
gtcaataaaatacataaaaattgaaaaaaaa

Haddock UA pBK-CMV-6

gcgtctctgtggataatctcatcgagctgttcttctgaatacatcgatATGAAGGCGCTAATAGGGCTGCTG
CTGCTGTTGGTCTTTGGTCACGGTGTGTCCTCAGTGCTTCACTCTATGCAGTATTT
CTACACGGCGTCGTCTGGACTCTCAGCCTTCCCAGAGTTTGTTGTTGTGGGGATG
GTGGATGAAGTTCAGATGGTTCACTATGACAGCAACAGTAAGAGAACAGTCCTCAA
ACAGGACTGGATGGAGCAGGTCACCAGAGGTGACCCCGCCTACCTGGAGAGGAA
CACTGGAAACTTTCTGGGTGCCCAGCAGGTCTTCAAAGGCAACATTGGGATTGCC
AAGCAGCGCTTTAACCAGACAGGAGGTACCCACTTATATCAGTGGATGTATGGTTG
TGAGTTGGATGATGAGGATGATTCTACTGATGGTTATCACCAGTATGGTTATGATG
GAGAGGACTTCATATCGTTGGACCTGAATACCCTGACCTGGGTCGCTCCAGTACG
TCAGGCCTTCCCCAACAAACAGAGATGGGATGGGGACAGAGCTAGGAATGAACGG
CATAAGCACTACTACACCAAGGAGTGTGTTGATTGGCTGAAGAAGCACCTGGCCTA
TGGGAAGAGCACTCTGCTGAGAACAGATCGTCCGCGGGTGTCTCTGCTCCAGAGG
AGCCCCTCCTCCCCAGTGGTGTGCCATGCTACAGGCTTCTTCCCTGACAGGGTGG
TGGTGTTCTGGAGGAGAGACGGTCAGGAGCTCCAGGAGCAGGTGGACCACGGGG
AGGTCCTCCCCAACCACGACGGGACCTTCCAGGTCAGCGTGGACCTCAACCTCAA
GGCCGTCCCACAGGAGGACTGGGGGAGGTACGAGTGTGTGGTCCAGCTGAGAGG
CATCGAGGACATCTCCACCCCCCTGGACCCCGCCCTCATCAGGACCAACGGGGG
CAAGAGTGGCATCACCGTCCCCATCATCATTGGGTTCGTTGTTCTCCTCCTCGCTG
CTGCTGCTGCTGCTGTTGGAGTCCTTCTGTCCAAGAAGAGGAACGATTCACACAA
GCAGCACAAACCAGTTGGTTCTTCTGACACCAGCTCTGAGAACACTGAGGGGCAG
AATCTGGCTCCTGAGGCCCGACCTCTGACAAAAGTTTGAagttaaatcatgaagagtcttctca
cgttcagcttcctaaagcagggtggagtggatcgcagttcatttataatccttacaaacaggcttgggacttcagctaactaa
agatgctatatttgcttaattacaataacattaactttattttgataacaatcatgaatctgaaagatctgtaaatcaattagatct
ctgtatttgtaatagatttgaataattcactggctttaacctgcaatgtttcactgacgcttggaaaccatgacaacataaactc
gccatgatggccatttttgttttggttcctgccgcttcaagttataacctttgaatcatgttatagtgtttagtaatatttaagtttcaaat
aatggtcaaagcttttgaaactagtttttggacagtgagattcggtcaagcttgaatacaataatcgttaaattgtgtgttttttctt
aaataatattatttgaattgcatcgctacataggctaccacgatcgagctaggcagctttagccttaatccttttgttgtactatttt
ctttgttctttggggaattttatcatagatactttggttattctgtcaataaagtacatccaaattgaaaaaaaaaaaaaaaaaa
a

Two of the sequences show interesting properties. The first sequence, Haddock UA pBK-CMV-1, appears to be attached to the sequence of a ribosomal protein L12. Joining these two sequences is a short “linker” sequence, which was not used when creating the library. The second sequence, Haddock UA pBK-CMV-3, has an inverted open reading frame (ORF). The ORF and portions of the 5’ and 3’ untranslated regions (UTRs) align with the other MH Class I alleles in the antisense orientation, yet the poly-A tail is in the sense orientation. Since functional protein would not be produced from either of these sequences, these anomalies could represent mechanisms used in this species to limit the number of expressed alleles. Figure 1.4 diagrams the structure of these alleles.

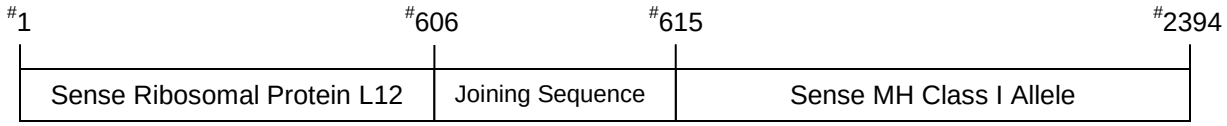
Using Haddock UA pBK-CMV-4 as a template for a probe, four additional clones were isolated using hybridization screening. These clones were sequenced as part of this project.

1.6. Purpose of Study

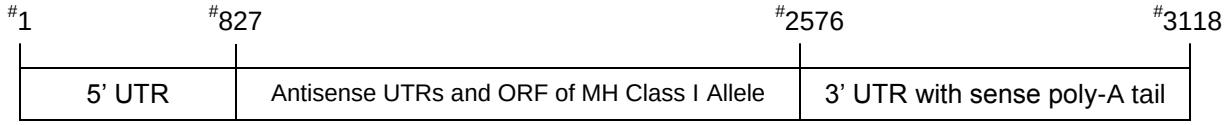
The goal of this project was to sequence MH Class I alleles in haddock and to investigate the UTRs of the mRNA transcripts for possible regulatory mechanisms. Preliminary research suggests that mechanisms are in place at the DNA level which limit the number of MH Class I alleles that are expressed. Hybridization and PCR screening were performed on clones of a haddock cDNA library and the sequences of the isolated clones were analyzed.

Figure 1.4: Schematic representation of the structure of Haddock UA pBK-CMV-1 and Haddock UA pBK-CMV-3. The sequence for Haddock UA pBK-CMV-1 appears to be attached to the sequence of the ribosomal protein L12. The ORF of Haddock UA pBK-CMV-3 is in the antisense orientation although the poly-A tail is in the sense orientation.

Haddock UA pBK-CMV-1



Haddock UA pBK-CMV-3



2. Materials and Methods

Please refer to Appendix I for explanation of the vector system used for the cDNA library.

2.1. PCR Screening of Haddock cDNA Library

2.1.1. Heat-treating Phage Sub-Libraries

Removal of the protein coat of the phage particles was necessary to allow PCR amplification of the desired sequences. This was accomplished by heat-treating aliquots of all sub-libraries. Aliquots of each sub-library were placed into separate tubes and boiled for 10 minutes. They were then allowed to cool to room temperature. Heat treated samples were stored after use at -20ºC.

2.1.2. PCR Screening

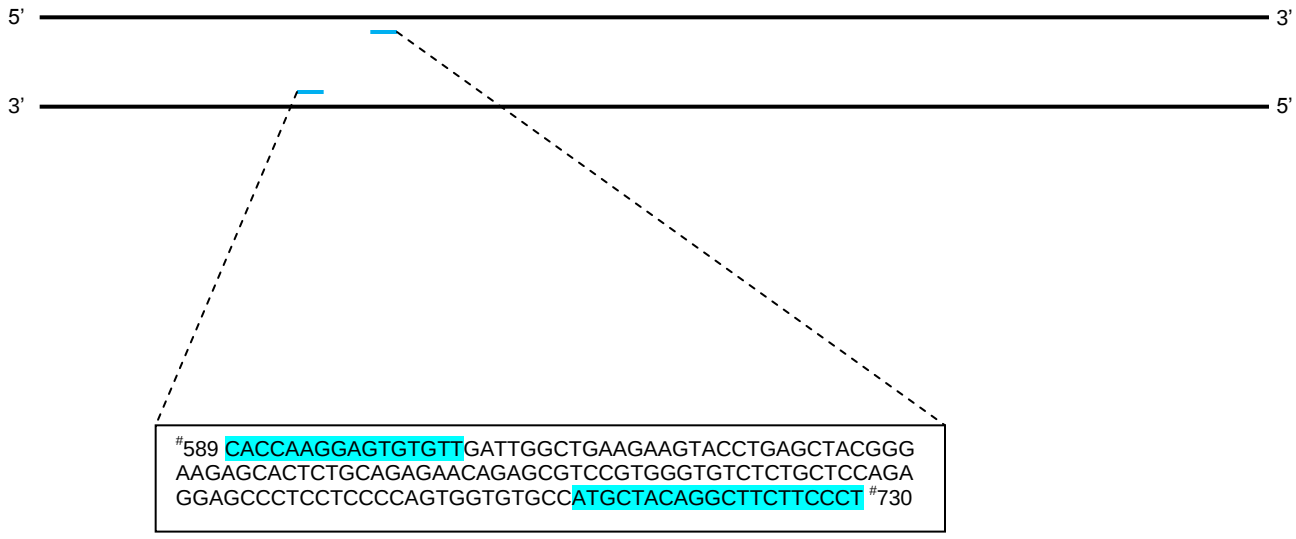
Each of the 18 sub-libraries of the haddock cDNA library were screened using haddock specific primers for MH Class I. The sense primer (5' CACCAAGGAGTGTGTTGATT 3') and antisense primer (5' AGGGAAGAAGCCTGTAGCAT 3') amplified expected fragments of 142 base pairs. Figure 2.1 shows the primer binding sites on a previously cloned haddock MH Class I sequence, Haddock UA p-BK-CMV-4. The fragments were amplified under the following conditions: 95ºC for 5 minutes, followed by 30 cycles of 95ºC for 30 seconds, 53ºC for 30 seconds and 72ºC for 30 seconds, followed by a 5 minute extension at 72ºC. The results of the PCR screens were run on 1.5% an agarose gel. Those sub-libraries with amplified fragments were considered to have clones containing MH Class I transcripts.

2.2. Maintenance of Cultures and Phage Stocks

2.2.1. *Escherichia coli* cultures XL1-Blue-MRF' and XLOLR

Two strains of *Escherichia coli*, XL1-Blue-MRF' and XLOLR, were used for experiments. These strains were provided as part of the ZapExpress® vector system used for the cDNA library. Both strains were stored long term as glycerol stocks (25% glycerol, 75% culture) at -80ºC. For short term use, both strains were maintained as streak plates on

Figure 2.1: Primer map of haddock specific primers "Haddock UA Sense" and "Haddock UA Antisense" with haddock MH Class I sequence, Haddock UA pBK-CMV-4.


5’
3’
3’
5’
#589 CACCAAGGAGTGTGTTGATTGGCTGAAGAAGTACCTGAGCTACGGG
AAGAGCACTCTGCAGAGAACAGAGCGTCCGTGGGTGTCTCTGCTCCAGA
GGAGCCCTCCTCCCCAGTGGTGTGCCATGCTACAGGCTTCTTCCCT #730

LB-tetracycline agar plates (0.5% yeast extract, 1% tryptone, 0.17M sodium chloride, 2% agar, 0.0015% tetracycline) and stored at 4ºC.

2.2.2. Helper Phage Stocks

The helper phage stocks provided as part of the ZapExpress® vector system was amplified and stored at -80ºC. To amplify the helper phage a 10mL culture of XL1-Blue-MRF' in a supplemented LB broth (0.5% yeast extract, 1% tryptone, 0.17M sodium chloride, 0.1M magnesium sulfate, 0.006M maltose) with an optical density at 600nm of 0.3 was inoculated with approximately 2.5 x 10^8 phage particles. The culture-phage mixture was incubated at 37ºC for 15 minutes without shaking followed by an overnight incubation at 37ºC with shaking at 225rpm. The mixture was then heated at 65ºC for 15 minutes and centrifuged at 10,000rpm for 5 minutes at room temperature. The supernatant, which contained the phage, was recovered. An aliquot was removed to titer the new helper phage stock. The new helper phage stock was prepared for long term storage by adding DMSO (final concentration 7%), aliquoting and storing at -80ºC.

2.3. Hybridization Screening of Haddock cDNA Library

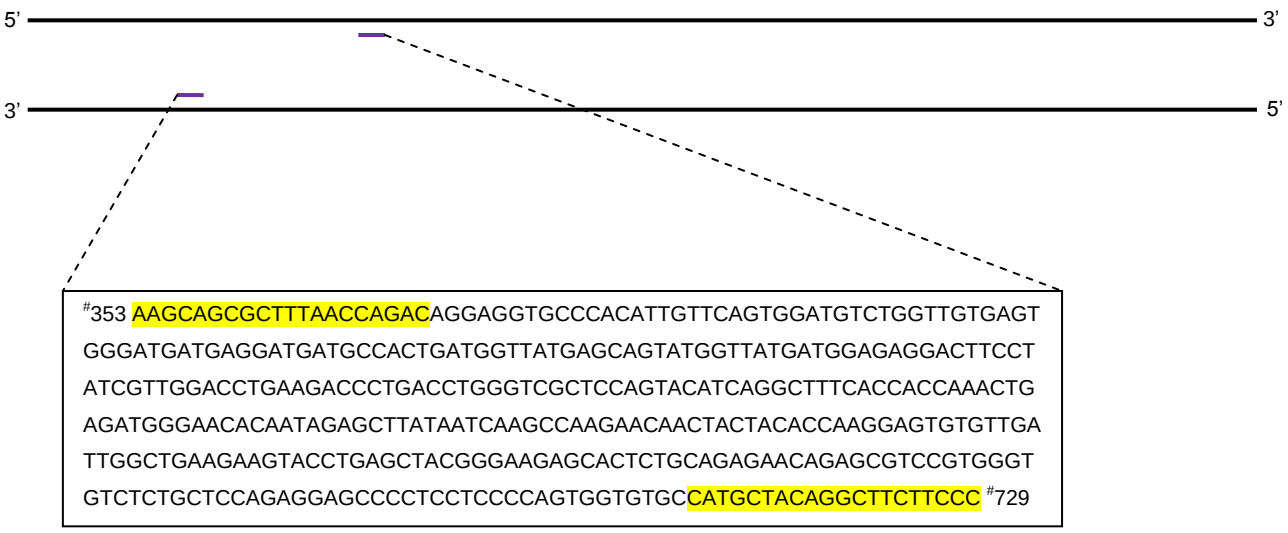
2.3.1. PCR Synthesis of DIG-Labeled Probe

A haddock MC Class I sequence was previously cloned into pBK-CMV. This clone, referred to as Haddock UA pBK-CMV-4, was used as the template for the DIG-labeled probe synthesis using PCR DIG Probe Synthesis Kit (Roche #11 636 090 910). The probe was amplified using a sense primer (5' AAGCAGCGYTTYAACCARAS 3') and an antisense primer (5' GGGTAGAARCCTGTMGCRTG 3') which gave an estimated product of 377 base pairs. Figure 2.2 shows the primer map as well as the probe sequence. The probe was synthesized under the following conditions: 95ºC for five minutes followed by 35 cycles of 95ºC for 30 seconds, 55ºC for 30 seconds, and 72ºC for 2 minutes, followed by a 5 minute extension at 72ºC. The products were run on a 1.5% agarose gel to verify proper synthesis.

2.3.2. Screening of library as phage

Two different methods were used to screen the library. Initially, the library was screened keeping the clones as phage. This is the method that was used previously.

Figure 2.2: Primer map of "Teleost UA univsense" and "Teleost UA univantisense" primers with Haddock UA p-BK-CMV-4 and probe sequence used for hybridization screening.


5’
3’
3’
5’
#353 AAGCAGCGCTTTAACCAGACAGGAGGTGCCCACATTGTTCAGTGGATGTCTGGTTGTGAGT
GGGATGATGAGGATGATGCCACTGATGGTTATGAGCAGTATGGTTATGATGGAGAGGACTTCCT
ATCGTTGGACCTGAAGACCCTGACCTGGGTCGCTCCAGTACATCAGGCTTTCACCACCAAACTG
AGATGGGAACACAATAGAGCTTATAATCAAGCCAAGAACAACTACTACACCAAGGAGTGTGTTGA
TTGGCTGAAGAAGTACCTGAGCTACGGGAAGAGCACTCTGCAGAGAACAGAGCGTCCGTGGGT
GTCTCTGCTCCAGAGGAGCCCCTCCTCCCCAGTGGTGTGCCATGCTACAGGCTTCTTCCC #729

However, problems with this method arose which required the clones be converted from phage to plasmid.

It was necessary to perform several rounds of screening. Phage diffuse through the agar beyond the boundaries of the plaque. To ensure an individual clone was isolated, the positive plaques collected were re-screened. The first round of screening was done using 150mm diameter plates with approximately 5000 pfu[1]/plate and a total of 10 plates. After hybridization and collection (see below) the positive isolates, which were collected individually, were re-screened (the second round of screening) using 100mm plates, a total of one plate per isolate. Positive isolates from the second round of screening were pooled and then underwent a third round of screening identical to the second round.

2.3.2.1. Plating

A culture of XL1-Blue-MRF' was grown in supplemented LB broth to an optical density at 600nm of 0.6 and infected with sub-library 16 phage. The phage were allowed to adsorb to the cells by incubation at 37ºC for 15 minutes. NZY top agar (2.2% (w/v) NZCYM powder [Gibco # M36550B], 0.7% (w/v) agar) at approximately 50ºC was added to the culture-phage mixture and then poured onto pre-warmed NZY agar (2.2% (w/v) NZCYM powder, 1.5% (w/v) agar) plates. Plates were incubated at 30ºC over night[2].

2.3.2.2. Hybridization Screening

Plates were removed from the incubator and allowed to cool at 4ºC for approximately two hours. Sterile nylon membrane was then aseptically laid on the surface of each plate. A 23G needle was then used to poke holes through the membrane and agar. A coloured Sharpie was used to mark the position of the holes on the bottom of the agar plate. The membranes were then peeled off the surface of the plate an allowed to dry, plaque-side up, for approximately 30 minutes. The plates were then stored in bags at

[1] Plaque Forming Unit (pfu).

[2] Optimal growth conditions were determined by growing plates at 37ºC, 30ºC, and room temperature overnight. Plaques on plates grown at 37ºC were larger than those grown at lower temperatures and were therefore less well isolated. Plaques on plates grown at room temperature were too small to be easily collected. The plaques on plates grown at 30ºC, however, were of sufficient size to be collected but were still well isolated, thereby limiting contamination from diffusion of phage from adjacent plaques.

4ºC. Once dry, the membranes were dipped, plaque-side up, in a series of solutions: denaturing solution (0.2M sodium hydroxide, 1.5M sodium chloride) for 20 seconds, followed by neutralizing solution (0.2M Tris-HCl pH 7.4, 0.3M sodium chloride, 0.03M sodium citrate) for five minutes, followed by wash solution (0.3M sodium chloride, 0.03 sodium citrate) for 5 minutes. The membranes were then allowed to dry, plaque-side up, for approximately one hour. Once dry, the membranes were exposed to UV, 12,000µJ for one minute, to fix the DNA to the membrane.

The next step was pre-hybridization. The prepared membranes were rolled up, plaque-side in, and placed into a hybridization bottle. Approximately 20mL of hybridization buffer (0.75M sodium chloride, 0.075M sodium citrate, 0.1% sarkosyl, 0.02% SDS, 1% nucleotide blocking solution) was added. The bottles placed in the hybridization chamber and incubated at 60ºC for two hours and six revolutions per minute.

Following the pre-hybridization was the hybridization. The hybridization buffer was removed from the bottles[3] and 6mL of probe solution (2µL of PCR probe synthesis product per milliliter of hybridization buffer) was added. The bottles were returned to the hybridization chamber and incubated at 60ºC over night at six revolutions per minute.

Once hybridization was complete the probe solution was recovered[4] and the membranes were washed. The first two washes were done in the hybridization chamber. The first wash used approximately 50mL 2X wash solution (0.3M sodium chloride, 0.03M sodium citrate) per bottle at less than 40ºC for 5 minutes and six revolutions per minute. The second wash used approximately 25mL 0.1X wash solution (0.015M sodium chloride, 0.0015M sodium citrate) per bottle at 60ºC for 30 minutes and 12 revolutions per minute. The membranes were then removed from the hybridization bottles and placed back-to-back (i.e. plaque-sides out) into dishes (i.e. two membranes per dish). The membranes were then washed twice with DIG Buffer I (0.1M maleic acid, 0.15M sodium chloride, pH 7.5) for five minutes per wash on an orbital shaker at approximately 15 rpm.

[3] Hybridization Buffer was collected and re-used several times. It was stored at 4ºC for short periods (one to two days) or at -20ºC for longer periods.
[4] The Probe Solution was re-used several times. With each successive use more probe was added.

Following these washes the membranes were incubated in DIG Buffer II (0.1M maleic acid, 0.15M sodium chloride, 1% (w/v) nucleotide blocking solution [Roche #11 096 176 001], autoclaved to dissolve the blocking solution) for 30 minutes on an orbital shaker at approximately 15 revolutions per minute. The membranes were then removed and placed back-to-back in Petri dishes containing antibody solution[5] (0.2µL anti-DIG-alkaline phosphatase antibody [Roche #11 093 274 910] per milliliter of DIG Buffer II). The antibody solution was re-used up to five times. Initially the incubation time was 30 minutes. With each successive use, however, the incubation time was increased by 15 minutes.

After incubation with antibody, the membranes were returned to the dishes was washed three times with DIG Buffer I for 15 minutes per wash on an orbital shaker at approximately 15 revolutions per minute. The membranes were then washed in DIG Buffer III (0.1M Tris-HCl pH 9.5, 0.1M sodium chloride, 0.05M magnesium chloride) for approximately three minutes.

The final step in the hybridization screening was the colour detection. This was done by placing the membranes, back-to-back, in Petri dishes containing colour-detecting solution[6](0.4mM nitro blue tetrazolium chloride, 0.4mM 5-bromo-4-chloro-3-indolyl phosphate toluidine salt in DIG Buffer III). The membranes were incubated in the dark at 37ºC until colour developed on the membranes. The solution could be used for up to five sets of two membranes, however, the length of incubation time increased with each successive set of membranes. Once colour developed the membranes were removed from the colour-detecting solution and placed in water.

2.3.2.3. Plaque Collection

Membranes were removed from the water and placed on cellophane on the bench top. An overhead transparency sheet was placed on top. Using a Sharpie, the holes poked

[5] The volume of antibody solution used depended on the size of membranes being screened. Large membranes used approximately 50mL per Petri dish, whereas small membranes used approximately 25mL per Petri dish.

[6] Colour-detecting solution was prepared immediately prior to use. The volumes used varied in a similar way to those of the antibody solution.

in the membrane were marked on the transparency sheet. Using a different colour Sharpie, the positive plaques were marked on the transparency sheet.

The transparency sheet was placed up-side-down on the bench top and the corresponding plate was placed overtop, with the poked holes aligning. Using sterile, truncated pipetteman tips, the positive plaques were “sucked-up” and put into a microfuge tube containing 500µL SM Buffer (0.1M sodium chloride, 0.008M magnesium sulfate, 0.05M Tris-HCl pH 7.5, 0.01% gelatin). Once all plaques were collected, 20µL of chloroform was added to each tube and the tubes were vortexed. Plaques were stored at -20ºC.

2.3.3. Screening of the library as plasmid

The conversion of the sub-libraries from phage to plasmid was done through the process of mass excision.

2.3.3.1. Mass Excision

Separate 50mL cultures of XL1-Blue-MRF’ and XLOLR were grown in supplemented LB broth. The cultures were pelleted at 5000rpm for 5 minutes and the pellets were re-suspended in 10mM magnesium sulfate to an optical density at 600nm of 1.0. The XL1-Blue-MRF’ culture was then inoculated with sub-library phage and helper phage (2 x 10^8 pfu sub-library phage and 2 x 10^{10} pfu helper phage into 2 x 10^9 cfu of culture) and incubated at 37ºC for 15 minutes. Following the incubation, 20mL supplemented LB broth was added and the mixture was incubated again for approximate 2.5 to 3 hours at 37ºC with shaking at 225rpm. The mixture was then heated at 70ºC for 20 minutes followed by centrifugation at 1000xg for 10 minutes. The supernatant was collected. A 100µL aliquot of the supernatant was combined with 20mL of the adjusted XLOLR culture and incubated at 37ºC for 15 minutes. Following the incubation 4mL of 6X NZY broth (12.12% (w/v) NZCYM powder) was added and the mixture was re-incubated at 37ºC for 45 minutes. An aliquot was removed to titer the stock and glycerol stocks were prepared and frozen at -20ºC.

2.3.3.2. Plating and Hybridization

Plates of the mass-excised sub-libraries were prepared by the spread plate method, up to 1000 cfu[7] per plate on 150mm diameter plates. The plates were incubated at 37ºC overnight.

The protocol for hybridization was identical to that outlined in Section 2.3.2.2 with one exception. After the membranes were removed, the plates were placed back at 37ºC and allowed to incubate over night. This was done because, as the membranes were peeled off, the colonies would peel off as well, leaving nothing to collect after the hybridization was complete. Re-incubating allowed for the re-growth of the colonies. Even with the re-incubation, however, not all colonies would re-grow. This source of error will be discussed in following sections.

2.3.3.3. Colony Collection

Colony collection was very similar to plaque collection. Transparency sheets were prepared in the same way they were for plaque collection. To collect the colonies, however, sterilized toothpicks were used to touch the positive colony and transfer the bacteria to a tube containing 5mL LB broth with Kanomycin (0.005%). These culture tubes were incubated overnight at 37ºC with shaking. Glycerol stocks were then prepared and stored at -20ºC. These cultures were also used for PCR screening and mini-preping (see below).

2.4. PCR Screening of Positive Clones

Cultures grown from positive hybridization isolates were used as template. The same primers and PCR conditions used for screening the haddock cDNA library (see section 2.3.1) were used to screen the positive isolates. Following the PCR reactions the products were run on a 1.5% agarose gel to verify.

[7] Colony Forming Unit (cfu).

2.5. Mini-prep of Positive Clones

The cultures which were confirmed positive by PCR were mini-prepped and the plasmid DNA was sent for sequencing. Mini-preps were done either by kit (Fermentas #K0502) or using the following protocol.

A volume of 1.5mL of each culture was spun down at 10,000rpm for 5 minutes. The supernatant was removed and the pellet was re-suspended by vortexing in 100µL of Solution 1 (0.05M glucose, 0.025M Tris-HCl pH 8.0, 0.1M EDTA). Next, 200µL of Solution 2 (0.2M sodium hydroxide, 1% SDS) was added. The mixtures were mixed by inverting 8 times and then placed on ice for five minutes. This was followed by the addition of 150µL of Solution 3 (3M potassium acetate, 2M acetic acid). The mixtures were mixed by inverting eight times and then placed on ice for five minutes. The mixtures were then spun down and the supernatant recovered. To the supernatant 20µL of 0.1% RNaseA [Sigma # R6513] was added and the samples were incubated at 37ºC for 30 minutes. After the incubation, 300µL of chilled isopropanol was added. The samples were inverted 15 times and placed on ice for at least 20 minutes. The samples were then centrifuged. The supernatant was discarded and the pellet was washed twice with chilled 70% ethanol. After removal of the supernatant from the final wash, the pellet was allowed to air dry completely before re-suspending in 20µL of water. The samples were then heated at 65ºC for 10 minutes before measuring the DNA concentration using the NanoDrop Spectrophotometer. If necessary, the samples were diluted to approximately 350ng/µL and the concentration re-measured.

2.6. Sequencing

Samples were sent for sequencing to TAGC Sequencing Facility at the Hospital for Sick Children and were prepared for sequencing according to the Facility's instructions. Samples of the four previously identified clones, A9, A16, B9 and C7, were also submitted. The samples were sequenced using primers for the T3 and T7 promotor regions, which flank the multiple cloning site (MCS) of pBK-CMV, and were provided by the sequencing facility. The T3 promotor lies upstream of the MCS and therefore sequencing using a primer to T3 gives the 5' end of the sequence. Sequencing using a

primer to T7 gives the 3' end of the sequence in the antisense orientation. To achieve a complete sequence the 5' sequence was aligned with the 3' sequence. In most cases, the sequences were sufficiently long that there was an overlap. As a result, the entire sequence of the clone was achieved.

Sequencing using primers to T3 and T7 did not always result in complete sequences. Several clones were too long to have the T3 sequence and T7 sequence overlap, and therefore, the middle of the sequence was missing. Also, in the case of Clone 17.3 the sequencing from the T7 primer was unsuccessful, despite repeated attempts. There was also difficulty obtaining a reliable sequence of Clone 1.2 from the T7 promotor primer. Therefore, specific primers were designed so that the entire transcript could be sequenced. Table 2.1 gives the sequences and primers used to complete the sequencing. For Clone 17.3 two specific forward primers were required to complete the sequencing. The first primer was made using the sequence from the T3 primer sequencing. A subsequent primer was made using the sequence from the first specific primer.

2.7. Analysis

Sequences were initially analyzed using the alignment editor BioEdit (Hall, 1999; http://www.mbio.ncsu.edu/BioEdit/bioedit.html). Each the four possible orientations (forward, reverse, complement, reverse complement) of each of the sequences were individually compared to the Haddock UA p-BK-CMV-4 sequence. Those sequences that showed significant alignment were aligned with each other in their alignment orientation using ClustalW as an accessory application through BioEdit.

Several of the sequences contained portions that did not align with the other MH Class I sequences. Using BLAST (NCBI) these portions were checked for similarity to sequences from other genes.

An alignment of the DNA and amino acid sequences were generated using ClustalW. The sequences of DNA that were translated were all in the alignment orientation. In the

Table 2.1: Allele specific primers used to complete sequencing. Sequencing 17.3 was unsuccessful from the T7 promotor primer. Therefore, only the 5' end of the sequence was known. As a result only forward primers were used to complete the sequencing. This required two sequence specific primers. The first was made from the sequence obtained from the T3 primer sequencing and the second primer was made from the sequence obtained from the first specific primer sequencing. Completion of the sequencing of Clone 1.2 only required the one specific forward primer. Clones C7, 16.16 and 16.22 were successfully sequenced by using only one set of sequence specific primers each. For these clones sequencing using the T3 and T7 primers was successful. Therefore, both forward and reverse specific primers could be made.

Sequence		First Primer Sequence (5’ to 3’)	Second Primer Sequence (5’ to 3’)
C7	Forward	CCGGGCATACCACCAGCACC	
	Reverse	AGAGGTCGGGCCTCAGGAGC	
1.2	Forward	TGATGAGGGCGGGGTCCAGG	
16.16	Forward	ACGGGTGGCGACTCTGCATC	
	Reverse	ACTCGACATGATGGCCGGACT	
16.22	Forward	GACGGGAGTTGGCGTTGGCA	
	Reverse	AGAGGTCGGGCCTCAGGAGC	
17.3	Forward	GCGACGTCTGACCCCCGTTG	TCCGACTCCAGGACGACGCC

case of sequences that contained portions belonging to other genes those portions were removed so that the sequences used were of the MH Class I sequences only.

Three neighbour-joining phylogenetic trees were generated using MEGA version 4 (Tamura et al., 2007). All alleles were in their alignment orientation, and as in the case of the DNA and amino acid sequence alignments, only the MH Class I portions of the sequences were used. The first tree was generated using the amino acid sequences of all identified haddock MH Class I alleles, six partial and 4 full length cod MH Class I alleles (Miller et al., 2002; Persson et al., 1999) and six full length rainbow trout MH Class I alleles (Kiryu et al., 2005). The tree used the Poisson correction model with complete deletion of gaps and missing data, and bootstrap values based on 1000 replicates. The second tree was generated using only full length alleles of haddock and cod. This tree used the same parameters as the first tree. The third tree was generated using the nucleotide sequences of the α_1 and α_2 domains of the full length haddock and cod alleles. This tree used the Jukes and Canter model with complete deletion of gaps, and bootstrap values were based on 1000 replicates.

3. Results

3.1. Screening

The haddock cDNA library that was used in this project was previously sub-divided into 18 sub-libraries. An initial screening of these sub-libraries was done using PCR to narrow down the sub-libraries containing MH Class I transcripts. The PCR used the haddock specific set of primer Figure 3.1 shows the results of this PCR. A total of nine sub-libraries showed a band of the appropriate size: sub-libraries 1, 2, 8, 12, 13, 15, 16, 17, and 18. Hybridization screening of all sub-libraries was done, except for sub-library 18 which had been previously screened by Stephen DeBoer.

Hybridization screening began on sub-library 16. Initially, the sub-library was screened as plaques, as described in section 2.3.2. During each round of screening it was noticed that there were two distinct plaque morphologies on the plates. The first was typical of the library phage while the second was typical of the helper phage. The first two rounds of screening appeared not to be affected by having to different types of phage growing. However, when the third round of screening began almost all of the second round phage stocks showed no growth of phage (i.e. there was confluent growth of bacteria with few to no plaques). Various attempts to re-grow the phage stocks were made. The first attempt was to amplify the phage stocks by growing them in an *E. coli* culture overnight prior to plating. When this failed the phage stocks from the first round of screening were re-plated, but these also showed no growth. One potential source of error was the presence of the second type of phage, assumed to be helper phage.

Acting on the assumption that the library phage had been previously contaminated with helper phage, the entire sub-library was converted from phage to plasmid through the process of mass excision, described in section 2.3.3.1. Hybridization screening continued and a total of 100 plates with approximately 500 to 1000 cfu per plate were screened. The positive colonies were collected, grown overnight and screened with PCR using the haddock specific primers. Figure 3.2 shows a typical PCR screen. The

Figure 3.1: Initial PCR screen of haddock cDNA library screen. Clearly visible bands at approximately 150bp can be seen from Sub-libraries 1, 2, 8, 12, 13, 16, 17 and 18 as well as the positive control. The positive control used Haddock UA pBK-CMV-4 for template DNA. In the negative control water was added instead of template DNA.

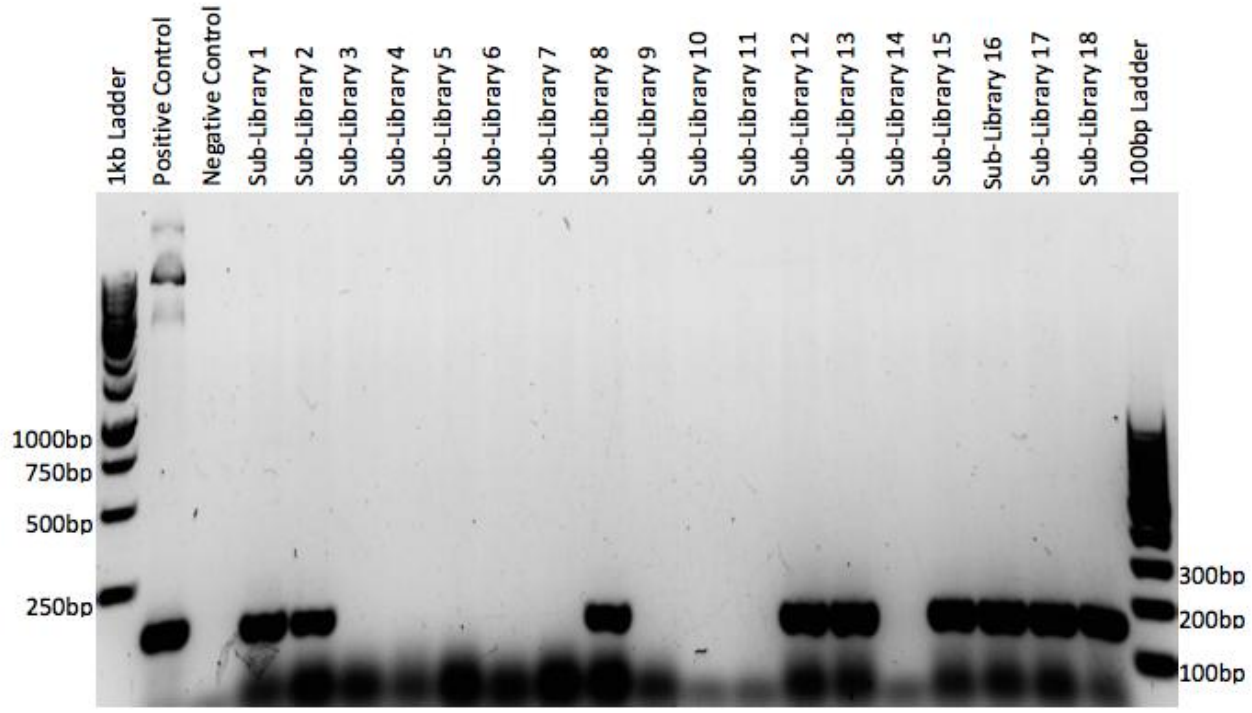
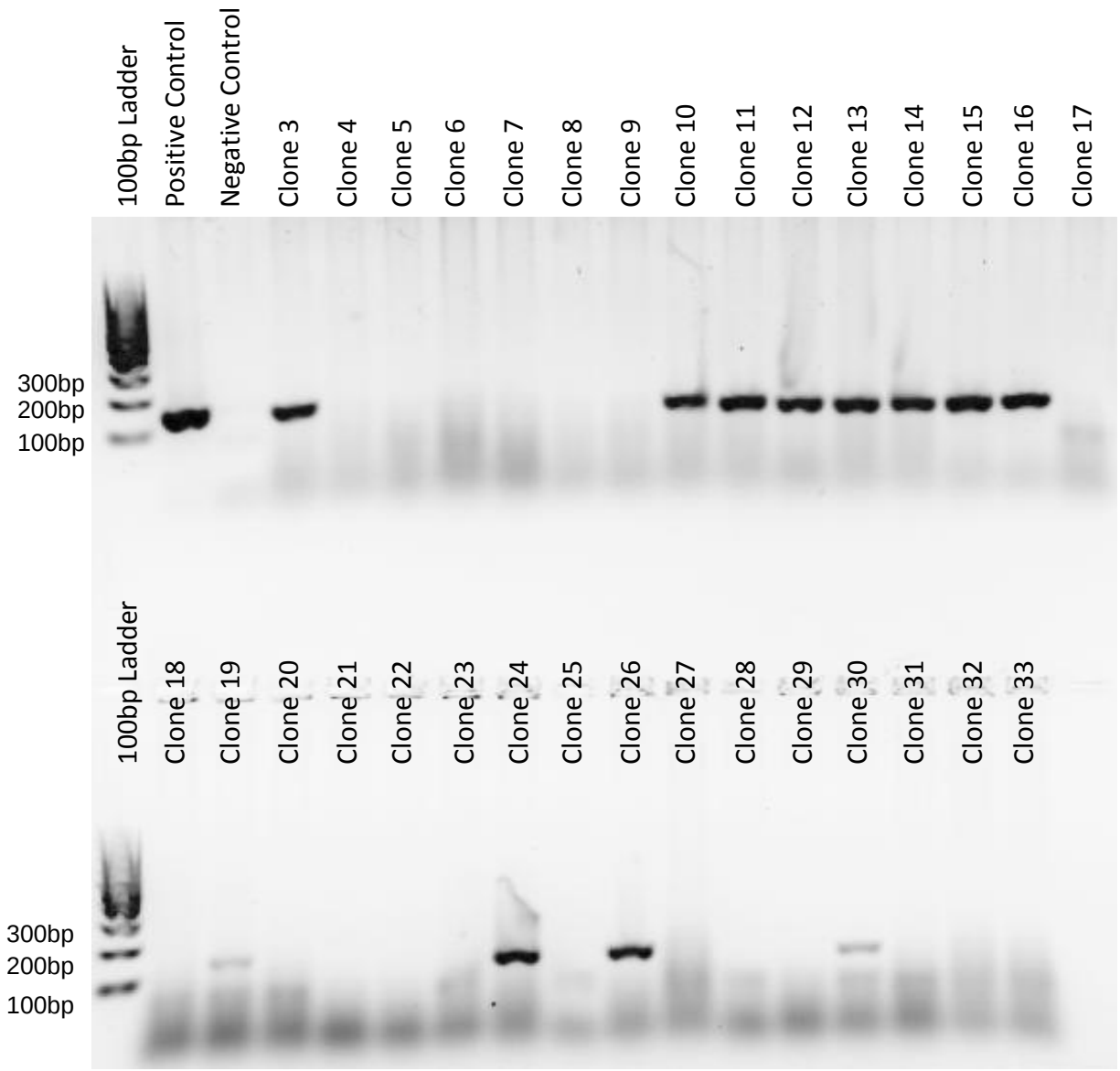

1kb Ladder
Positive Control
Negative Control
Sub-Library 1
Sub-Library 2
Sub-Library 3
Sub-Library 4
Sub-Library 5
Sub-Library 6
Sub-Library 7
Sub-Library 8
Sub-Library 9
Sub-Library 10
Sub-Library 11
Sub-Library 12
Sub-Library 13
Sub-Library 14
Sub-Library 15
Sub-Library 16
Sub-Library 17
Sub-Library 18
100bp Ladder
1000bp
750bp
500bp
250bp
300bp
200bp
100bp

Figure 3.2: PCR Screen of clones isolated from Sub-library 16. A total of 30 clones were collected from sub-library 16 were screened. Only isolates 3, 10, 11, 12, 13, 14, 15, 16, 19, 24, and 26 showed a positive result.


100bp Ladder
Positive Control
Negative Control
Clone 3
Clone 4
Clone 5
Clone 6
Clone 7
Clone 8
Clone 9
Clone 10
Clone 11
Clone 12
Clone 13
Clone 14
Clone 15
Clone 16
Clone 17
300bp
200bp
100bp
100bp Ladder
Clone 18
Clone 19
Clone 20
Clone 21
Clone 22
Clone 23
Clone 24
Clone 25
Clone 26
Clone 27
Clone 28
Clone 29
Clone 30
Clone 31
Clone 32
Clone 33
300bp
200bp
100bp

cultures which gave a positive PCR screen were then prepared for sequencing by mini-prepping the plasmid DNA.

Sub-libraries 1, 2, 8, 12, 13, 15 and 17 were screened in a similar fashion. Although not all sub-libraries showed the two morphologies of plaques, they were all converted from plaques to plasmid to maintain method consistency. These sub-libraries were screened on eight plates with approximately 500 cfu per plate.

3.2. Sequence Analysis

In addition to the four alleles previously sequenced, 18 unique alleles were obtained from all screened libraries. Four of the alleles obtained were from clones previously isolated by Stephen DeBoer. The alleles were initially divided into two groups based on their alignment orientation. The first group of 12 sequences aligned with Haddock UA pBK-CMV-4 in the sense orientation. These sequences are shown in Figure 3.3. The second group of six sequences aligned with Haddock UA pBK-CMV-4 in the reverse complement orientation and are shown in Figure 3.4. None of the sequences aligned in the reverse or complement orientations with the Haddock UA pBK-CMV-4 sequence.

3.2.1. Sequence Orientation

Since orientation of the ORF was a possible inactivation method for alleles it was necessary to confirm the directional cloning of transcripts during the cDNA library construction. Please refer to Appendix II for a description of the cDNA library construction method. Since the alleles were sequenced using primers to the vector it was possible to compare the 5' and 3' ends of the sequences to the vector sequences to verify directionality. The 5' end of a given sequence should contain vector sequence followed by the *EcoR* I adaptor sequence (5' GAATTCGGCACGAGG3'), and the 3' end should contain the *Xho* I site (CTCGAG) followed by vector sequence.

As expected, all clone sequences had 5' vector sequence at their 5' ends and 3' vector sequence at their 3' ends. However, two clones, B9 and 16.26 had no *EcoR* I adaptor

Figure 3.3 : MH Class I sequences isolated from haddock cDNA found in the sense orientation. Sequences are shown with the ORF in uppercase letters. A) Clone A9; B) Clone B9; C) C7; D) Clone 1.1; E) Clone 2.4; F) Clone 2.10; G) Clone 16.11; H) Clone 16.12; I) Clone 16.13; J) Clone 16.15; K) Clone 16.22; L) Clone 16.60.3

A9

tactcgcctgctgtcgacactagtggatccaaaggtttactcaagtgcgaagtcaaaaacggacagcataacataaaaacgtttaatc
aacatttttactcaactaaattgtttaatcttgaaataactcccaaaataatatccagaaatcatgaacttcattgtaggttttggttaacaaa
gtaaatccctcatgtaattaagctcttaacattgtttgctacaggctaacaaaggcttatcaagcctaaatgaggataactaaaagcgag
aaatcgctccgcgccccgtttggttcatcgtgacagatgtttctcagcaraaactgccaattaatttgctctcaaatccacctggaggattt
ggcatgactgactaataaaaaaacacattaagaataactattcttcagaatgcaagaattgatttaaactgtcctaaaaaattaaaaaa
cacaacaaaaccaagttgacatgtttttctacctccccaaccccttggtagagataaccctccctcgacagtcttcgacaggtcaggtca
gcgaggttccgaccccccgctgttggttcaaagcgaccaggaagtggctgttagcccaccctgagtggcggagctacaagaacca
ctccgtctcctctgacaaacagcatggggatgttccttttggtggacttgtacagctcctcgtacgtctcctcgtcgatctccaccgtggtca
ccgtctcctccacagcacccaggatcatgttcaggtgctggtcgtatgcatgcaatctgccgcggagctcccgatcgttcctcatcttgac
gtagatcctctcgtccaaactgagcctgatcaggtccaggggctcctccaccgtcggagttccgtctctgtggataatctcatcgagctgt
tcttctgaatacatcgatATGAAGGCGCTAATAGGGCTGCTGCTGCTGTTGGTCTTTGGTCACGGTG
TGTCCTCACTACTTCACTCTATGCAGTTTTTCTACACGGCGTCGTCTGGACTCTCAACCTTC
CCAGAGTTTGTTGTTGTTGGGATGGTGGATGGAGTTCAGATGGTTCACTATGACAGCAACA
GTAAGAGAGCCGTACCCAAACATGACGCGGCGGACCAGCTCTACAGAGATCATCCGGGTG
ACCTCGAGGGGGACACTGGAGGATTTCAGGCTGACCAGCAGGCCTTCAAAGCCAACATTG
GGATTGCCAAGCAGCGCTTTAACCAGACAGGAGGTACCCACATGGTTCAGTGGATGGTTG
GTTGTGAGTGGGATGATGAGGATGATTCTACTGATGGTTATAACCAGCATGGTTATGATGG
AGAGGACTTCATATCGTTGGACCTGAAGACCCTGACCTGGGTCGCTCCAGTAGACGAGGC
TTGTGCTACCAAACGGAGATGGGATGGGGATAGGGCCCGAAATCAATACTGGAAGAACTA
CTTCACCAAGGAGTGTGTTGATTGGCTGAAGAAGCTCCTGTCCTATGGGAAGAGCACTCTG
CTGAGAACAGATCGTCCGCGGGTGTCTCTGCTCCAGAGGAGCCCCTCCTCCCCAGTGGTG
TGCCATGCTACAGGCTTCTTCCCTGACAGGGTGGTGGTGTTCTGGAAGAGAGACGGTCAG
GAGCTCCAGGAGCAGGTGGACCACGGGGAGGTCCTCCCCAACCACGACGGGACCTTCCA
GGTCAGCGTGGACCTCAACCTCAAGGCCGTCCCACAGGAGGACTGGGGGAGGTACGAGT
GTGTGGTCCAGCTGAGAGGCATCGAGGACATCTCCACCCCCCTGGACCCGCCCCTCATCA
GGACCAACGGGGGCAAGAGTGGCATCACCGTCCCCATCATCATTGGGTTCGTTGTTCTCC
TCCTCGCTGCTGCTGCTGCTGCTGCTGTTGGAGTCCTTCTGTACAAGAAGAGGAACGATTC
ACACAAGCAGCACAAACCAGTTGATTCACACAAGCGACCCATATCAGTTGGTTCTTCTGAC
ACCAGCTCTGAGAACACTGAGGTGCAGAATCTGGATCCTGAAGCCCGACTTCTGACAGAA
GGTCGAAGTTAAatcatgaagagtcttctcacgttcagctccctaaagcagggtggagtggatcgcagttcatttataatccttac
aaacaggcttgggacttcagctaactaaagatgctatatttgcttaattacattaacatccatgctttattttgataaaaaacataaatccgt
aagtaaatcaataagtgtgaatcaattagatctctgtagcctatttgtgattgattaactcaatggattaattctgcaatctttcactgaccctt
ggaaaccatggcaacatgaccccgccattatggccattttgttttgcttctaactgtttcatcttataagcttttaattatgttatattgtttagtca
tattgaagtttcaaataatgttcaaagattctgaatgaagatttgggacagtgagattctgttgagtttgaatacaataatcattaaatgtaat
gtgtttccttgaaatatttaatttcagctagctagagaattgcatcgctacataggctaccacagtcaagctgttttagctttaatggtctttgtt
gtgttttctgtgttctttggggaattggatcatatatactttgcttattctgtcaataaatacacaaaaaagtacttctagagcggccgcgggc
ccatcgattccacccgggtgggtaccagtaaac

B9

caactgcggtcgacactagtggatccaaagctggagtcggagtcccgtctctgtggataatctcatcgagctgttcttctgaatacatcg
atATGAAGGCGCTAATAGGGCTGCTGCTGCTGTTGGTCTTTGGTCACGGTGTGTCCTCAGG
GCTTCACTCTCTCCATTATTTCTACACGGCGTCGTCTGGACTCTCAACCTTCCCAGAGTTTG
TTTCTGTGGGGATGGTGGATGGAGTTCAGATTGACTACTATGACAGCAACACCCAGAGAGT
AGTCCTCAAACAGGACTGGATGGAGCAGGTCACCAGAGGTGACCCCGCCTACCTGGAGA
GGGAAACTGGAATCAGAAAGGGTAACCAGCAGTTCTTCAAAGTCGGCGTTGGAACTCTGA
AGCAGCGCTTTAACCAGACAGGAGGTGCCCACATGGTTCAGGTCATGTATGGTTGTGAGT
GGGATGATGAGGATGATTCTACTGAAGGTTATGACCAGCTTGGTTATGATGGAGAGGACTT
CCTATCGTTGGACCTGAAGACCCTGACCTGGGTCGCTGCAGTACGTCAGGCCTTCACCAC
CAAACAGAGATGGGATGGGGACAGAGCTCAGAATGAACGGTATAAGCACTACTACACCAA
GGTCTGTGTTGATTGGCTGAAGAAGCACCTGGCCTATGGGAAGAGCACTCTGCTGAGAAC
AGATCGTCCGCGGGTGTCTCTGCTCCAGAGGAGCCCCTCCTCCCCAGTGGTGTGCCATGC
TACAGGCTTCTTCCCTGACAGGGTGGTGGTGTTCTGGAGGAGAGACGGTCAGGAGCTCCA
GGAGCAGGTGGACCACGGGGAGGTCCTCCCCAACCACGACGGGACCTTCCAGGTCAGCG
TGGACCTCATCCTCAAGGCCGTCCCACAGGAGGACTGGGGGAGGTACGAGTGTGTGGTC
CAGCTGAGAGGCATCGAGGACATCTCCACCCCCCTGGACCCCGCCCTCATCAGGACCAAC
AGGGGCAAGAGTGGCAGCACCGTCCCCATCATCATTGGGTTCGTTGTTCTCCTCCTCGCT
GCTGCTGCTGCTGTTGGAGTCTTTCTGTACAAGAAGAGGAACGATTCACACAAGCAGCACA
AACCAGTTGGTTCTTCTGACACCAGCTCTGAGAACACTGATGGGCAGAATCTGGCTCCTGA
GGCCCGACCTCTGACAGAAGTTTTAAGTTAAatcatgaagagtcttctcacgttcagcttcctaaagcagggtgga
gtggatcgcagttcatttataatccttacaaacaggcttgggacttcagctaactaaagatgctacatttgctaaattacaataacatccat
acttaactttattttgataacaatcatgaatctgaaagatctgtaaatcaattagatctctgtatttgtaatagatttgattaattctctggctttaa
cctgcaatgtttcactgacgcttggaaaccatgacgacacaaactcgccatgatggccattttgttttggttcctgccgcttcaagttataac
ctttgaatcatgttattgtgtttagtaatatttaagtttcaaataatggtcaaaggttctgagactagtttttggacagtgagatccggtcaagttt
gaatacaataaatgttaaattgtatgttttttcttaaataatattatttgaattgcatcgctacataggctaccacgatcgagctaggcagcttt
agccttaatccttttgttgtactattttctttgttctttggggaatttgatcatagatactggttattctgtcaataaagtacatccaaa

C7

gtgtcattgcgaaaactggtatacaaaatatttgtggggaattggcaacttctttaaacaaatttgatccctgtagtgaggcaaactaaca
gtacaaatacaaatgtgatcccttgaaccctatttttttttatttagatgacgctcaacaactcaaatgatggcaagactcaagcattgagg
gttacactttagtgatctcttcactagttgctgggtttgatgtttagtcgacctcttcgatggtggggccagatgagccaccgccgcctgggg
cagcacctccagcaccggggaagccaccgggcataccgtcgggcatgccaccgggcataccaccagcaccctggtacagcttgg
tgacgatggggttgcagaccttctccagctccttctgctggtgctcgtactcttccttctcggccgtctggttcttgtctagccagctgatgact
tcgttgcacttctccatgatcttctgcttgtcctcgtcgctgatcttgccggtgagcttctcatcctccacggtggacttcatgttgaaggcgta
cgactccagagagttcttggaggagaccttgtcacgctgggaatcatcttcagccttgtatttctcggcatcctgcaccatgcgctcgtcc
cgtctctgtggataatctcatcgagctgttcttctgaatacatcgatATGAAGGCGCTAATAGGGCTGCTGCTGCTG
TTGGTCTTTGGTCACGGTGTGTCCTCACTGCTTCACTCTATGCAGTATTTCTACACGGGGT
CGTCTGGACTCTCAACCTTCCCAGAGTTTGTTGTTGTTGGGATGGTGGATGGAGTTCAGAT
GGTTCACTATGACAGCAACACCCAGAGAACAGTCCTCAAACAAGACTGGATGAAGCAGGT
CACCAGAGGTGACCCCGCCTACCTGGAGAGGAACACTGGAAACTTTCAGGGTGCCCAGCA
GACCTTCAAAGTCGGCATTGGGACTCTGAGGCAGCGCTTTAACCAGACAGGAGGTACCCA
CTTATATCAGTTGATGTATGGTTGTGAGTTGGATGATGAGGATGATTCTACTGATGGTTATC
ACCAGTATGGTTATGATGGAGAGGACTTCCTATCGTTGGACCTGAAGACCCTGACCTGGGT
CGCTCCAGTACGTCAGGCCTTCCCCACCAAACAGAGATGGGATGGGGACAGAGCTTATAA
TGAACGGCAGAAGCACTACTACACCAAGGTCTGTGTTGATTGGCTGAAGAAGTACCTGTCC
TATGGGAAGAGCACTATGCTGAGAACAGAGCGTCCGCGGGTGTCTCTGCTCCAGAGGAGC
CCCTCCTCCCCAGTGGTGTGCCATGCTACAGGCTTCTTCCCTGACAGGGTGGTGGTGTTC
TGGAGGAGAGACGGTCAGGAGCTCCAGGAGCAGGTGGACCACGGGGAGGTCCTCCCCAA
CCACGACGGGACCTTCCAGGTCAGCGTGGACCTCAACCTCAAGGCCGTCCCACAGGAGG
ACTGGGGGAGGTACGAGTGTGTGGTCCAGCTGAGAGGCATCGAGGACATCTCCACCCCC
CTGGACCCCGCCCTCATCAGGACCAACTGGGGCAAGAGTGGCATCACCGTCCCCATCATC
ATTGGGTTCGTTGTTCTCCTCCTCGCTGCTGCTGCTGCTGCTGTTGGAGTCCTTCTGT
ACAAGAAGAGGAACGATTCACACAAGCAGCACAAACCAGTTGGTTCTTCTGACACCAGCTC
TGAGAACACTGAGGGGCAGAATCTGGCTCCTGAGGCCCGACCTCTGACAAAAGTTCAAAG
TTAAatgaagagtcttctcacgttcagcttcctaaagcagggtggagtggatcgcagttcatttataatccttacaaacaggcttggga
cttcagctaactaaagatgctatatttgctaaattacaataacatccatacttaactttattttgataacaatcatgaatctgaaagatctgta
aatcaattagatctctgtatttgtaacagatttgattaattctctggctttaacctgcaatgtttcactgacgcttggaaaccatgacaacaca
aactcgccatgatggccattttgttttggttcctgccgcttcaagttataaccttgaatcatgttatagtgtttagtaatatttaagtttcaaataa
tggtgaaaggttctgagactagtttttggacagtgagattcggtcaagtttgaatacaataatcgttaaattgtatgttttttctttaataatatta
tttgaattgcatcgctacataggctaccacgatcgagctaggcagctttagccttaatccttttgttggactattttctttgttctttgggaatttg
atcatagatactttggttattctgtcaataaagtacatccaaa

1.1

AGAGCCGTACCCAAACATGACACGGCGGACCAGTTCTACAGAGATCATCCGGGTGACCTC
GAGAGGGACACTGGAAACTTTCTGGGTTCCCAGCAGACCTACAAAGCCAACATTGGGATT
CTGAAGCAGCGCTTTAACCAGACAGGAGGTGCCCACATGTTTCAGTTCATGTATGGTTGTG
AGGTGGATGATGAGGATGATTCTACTGATGGTTATAGACAGTATGGTTATGATGGAGAGGA
CTTCCTATCGTTGGACCTGAAGTCCATGACCTGGGTCGCTCCAGTACGTCAGGCCTTCCC
CACCAAACAGAGATGGGATCAGGATAAAGCTGATCTACAATACATTAAGAACTACCAAACC
AAGGAGTGTGTTGATGGGCTGAAGAAGCTCCTGGCCTATGGGAAGAGCACTCTGCTGAGA
ACAGAGCGTCCGTGGGTGTCTCTGCTCCAGAGGAGCCCCTCCTCCCCAGTGGTGTGCCAT
GCTACAGGCTTCTTCCCTGACAGGGTGGTGGTGTTCTGGAGGAGAGACGGTCAGGATCTC
CAGGAGCAGGTGGACCACGGGGAGGTCCTCCCCAACCACGACGGGACCTTCCAGGTCAG
CGTGGACCTCAACCTCAAGGCCGTCCCACAGGAGGACTGGGGGAGGTACGAGTGTGTGG
TCCAGCTGAGAGGCATCGAGGACATCTCCACCCCCCTGGACCCCGCCCTCATCAGGACCA
ACGGGGGCAAGAGTGGCATCACCGTCCCCATCATCATTGGATTCGTTGTTCTCCTCCTCG
CTGCTGCTGCTGCTGCTGTTGGAGTCCTTCTGTCCAAGAAGTGGAACGATTCACACAAGCA
GCACAAACCAGTTGGTTCTTCTGACACCAGCTCTGAGAACACTGAGGGGCAGAACCTGGC
TCCTGAGGCCCGACCTCTGACAAAAGTTTGAagttaaatcatgaagagtcttctcacgttcagcttcctaaagcag
ggtggagtggatcgcagttcatttataatccttacaaacaggcttgggacttcagctaactaaagatgctatatttgcttaattacaataac
atccatacttaactttattttgataacaatcatgaatctgaaagatctgtaaatcaattagatctctgtatttgtaatagatttgattaattcactg
gctttaacctgcaatgtttcactgacgcttggaaaccatgacgacacaaactcgccatgatggccattttgttttggttcctgctgcttcaag
ttataacctttgaatcatgttatagtgtttagtaatatttaagtttcaaataatggtcaaaggttctgagactagtttttggacagtgagattcgg
tcaagcttgaatacaataatcgttaaattgtgtgttttttcttaaaaaatattatttgaattgcatcgctacataggctaccacgatcgagctag
gcagctttagcattaatccttttgttgtactattttctttgttctttggggaattttatcatagatactttggttattctgtcaataaagtacatcaaaa
ttgaaaaaa

2.4

GTTCAGTTCATGTATGGTTGTGAGTGGGATGATGAGGATGATTCTACTGATGGTTATGACC
AGTATGGTTATGATGGAGAGGACTTCATATCGTTGGACCTGAAGACCATGACCTGGGTCGC
TCCAGTACGTCAGGCCTTCACCACCAAACAGAGATGGAATCAGAATAAAGCTAGATTACAA
TACAATAAGAACTACTACACCAAGGTCTGTGTTGATTGGCTGAAGAAGTACCTGGCCTATG
GGAAGAGCACTCTGCTGAGAACAGATCGTCCGCGGGTGTCTCTGCTCCAGAGGAGCCCCT
CCTCCCCAGTGGTGTGCCATGCTACAGGCTTCTTCCCTGACAGGGTGGTGGTGTTCTGGA
GGAGAGACGGTCAGGAGCTCCAGGAGCAGGTGGACCACGGGGAGGTCCTCCCCAACCAC
GACGGGACCTTCCAGGTCAGCGTGGACCTCAACCTCAAGGCCGTCCCACAGGAGGACTG
GAGGAGGTACAAGTGTGTGGTCCAGCTGAGAGGCATCGAGGACATCTCCACCCCCTTGGA
CCCCGCCCTCATCAGGACCAACGGGGGCAAGAGTGGCATCACCGTCCCCATCATCATTGG
GTTCGTTGTTCTCCTCCTCGCTGCTGCTGCTGCTGCTGTTGGAGTCCTTCTGTACAAGAAG
AGGAACGATTCACACAAGCAGCACAAACCAGTTGGTTCTTCTGACACCAGCTCTGAGAACA
CTGAGGGGCAGAATCTGTCTCCTGAGGCCCGACCTCTGACAGAAGTTCAAAGTTAAatcatga
agagtcttctcacgttcagcttcctaaagcagggtggagtggatcgcagttcatttataatccttacaaacaggcttgggacttcagctaa
ctaaagatgctatatttgctaaattacaataacatccatacttaactttattttgataacagtcatgaatctgaaagatctgtaaatcaattag
atctctgtatttgtaatagatttgattaattcactggctttaacctgcaatgtttcactgacgcttggaaaccatgacaacacaaactcgccat
gatggccattttgttttggttcctgccgtttcaagttataacctttgaatcatgttatagtgtttagtaatatttaagtttcaaataatggtcaaagg
ttctgagactagtttttggacagtgagatccggtcaagtttgaatacaataatcgttaaattgtatgttttttcttaaataatattatttgaattgca
tcgctacataggctaccacgatcgagctaggtagcttcagccttaatccttttgttgtactattttctttgttctttggggaatttgatcatagata
ctttggttattctgtcaataaagtacatccaaa

2.10

GTCTTTGGTCACGATGTGTCCTCAGTGATTCACTCTCTGCGGATTTTCTACACGTCGTCTTC
TGGATTCTCAACCTTCCCAGAGTATGTGACAGTTCAGATGGTGGATGAGGTTCAGATTGAG
TACTATGACAGCAACACCCAGAGAATCATAACCAAACAGGACTGGGTGGTCCAGGCCAAC
AGATACAACATCCCCGACTACCTGGAGAGGTCAACTGAAAACAGAAAGGGTCAACAGCAG
GTCGCCAAAGTCAACATTGGGACCGCGAAGCAGCGCTTTAACCAGACAGGAGGTGCCCAC
ATTGTTCAGAGGATGTCTGGCTGTGAGTGGGATGATGAGGATGATACCACTGATGGTTATG
AGCAGTATGGTTATGATGGAGAGGACTTCCTATCGTTGGACCTGAAGACCCTGACCTGGG
TCGCTCCAGTACATCAGGCTCTCACCATCAAACTGAGATTGGATCACGATACAGCTTGGAA
ACAATACTTGAAGAACTTCTACACCAAGGAGTGTGTTGATTGGCTGAAGAAGTACCTGAGC
TACGGGAAGAGCACTCTGCAGAGAACAGAGCGTCCGCGGGTGTCTCTGCTCCAGAGGAG
CCCCTCCTCCCCAGTGGTGTGCCATGCTACAGGCTTCTTCCCTAACAGGGTGGTGGTGTT
CTGGAGGAGAGACGGCCAGGAGCTTCAGGAGCAGGTGGACCACGGGGAGGTCCTCCCC
AACCACGACGGGACCTTCCAGGTCAGCGTGGACCTGAACCTAACGGCCGTCCCACAGGA
GGACTGGGGGAGGTACGAGTGTGTGGTCCAGCTGAAAGGCATCGAGGACATCTCCATCCT
CCTGGACCCCGCCCACATCAGGACCAACGGGGGGGACAAGAGTCTCATCCTTGCTTTCAT
CCTCACTGGTGTTGCTGTTGTTGCTGTTTTTGTTGCTGTTGTTGTTGGAGTCTTTCTGTACC
GGAAGAGGAACGATTCAGACAAGCGTCACAAACCAGTTGGTTCTTCTGACACCAGCTCTGA
GAACACTGAGGTGCAGAAGCTGTCTCCTGAGGCCCAACCTCTGACCACTGGTCAAAGTTA
Aatcatgaagaggcttgtcacgttcggcttcctaaaccaggatggagtggatcgctggtcatttataatccttacaatcaggcttgggact
tcaaacctgcaatctttcactgacgcttggaaaccatagcaacacgacgtcgccatgatggccatttgggtcccgccgcttcacgtcat
aacctttaattatgttatattgttaagtaatatttaagtttcaaataatgttcaaagattctgaatcaagatttttgacagtgaggttttgtcaagt
ttcaatacaataatcgattaatttaagttgtttcattaaaatatattatttagcgttctgctagctagagaattgcattgctgcataggctaccac
aatcctgctaggccgctttagcctcaattctttttgttgtattttctgtgttctttggggaattggatagtatacttgggttattctgtcaataaaatg
cataaaaattgaaaaaaaaaaa

16.11

CTGAGCTGGGTCGCTCCAGTAGACGAGGCTTGTGCTACCAAACGGAGATGGGATGGGGA
TAGGGCCCGAAATCAATACTGGAAGAACTACTTCACCAAGGAGTGTGTTGATTGGCTGAAG
AAGCTCCTGTCCTATGGGAAGAGCACTCTGCTGAGAACAGATCGTCCGCGGGTGTCTCTG
CTCCAGAGGAGCCCCTCCTCCCCAGTGGTGTGCCATGCTACAGGCTTCTTCCCTGACAGG
GTGGTGGTGTTCTGGAAGAGAGACGGTCAGGAGCTCCAGGAGCAGGTGGACCACGGGGA
GGTCCTCCCCAACCACGACGGGACCTTCCAGGTCAGCGTGGACCTCAACCTCGAGGCCG
TCCCACAGGAGGACTGGGGGAGGTACGAGTGTGTGGTCCAGCTGAGAGGCATCGAGGAC
ATCTCCACCCCCCTGGACCCCGCCCTCATCAGGACCAACGGGGGCAAGAGTGGCATCAC
CGTCCCCATCATCATTGGGTTCGTTGTTCTCCTCCTCGCTGCTGCTGCTGCTGCTGTT
GGAGTCCTTCTGTACAAGAAGAGGAACGATTCACACAAGCAGCACAAACCAGTTGATTCAC
ACAAGCGACCCATATCAGTTGGTTCTTCTGACACCAGCTCTGAGAACACTGAGGTGCAGAA
TCTGGATCCTGAAGCCCGACTTCTGACAGAAGGTCGAAGTTAAatcatgaagagtcttctcacgttcagc
tccctaaagcagggtggagtggatcgcagttcatttataatccttacaaacaggcttgggacttcagctaactaaagatgctatatttgctt
aattacattaacatccatctttattttgataaaaaacataaatccgtaagtaaatcaataagtgtgaatcaattagatctctgtagcctatttg
tgattgattaactcaatggattaattctgcaatctttcactgacccttggaaaccatggcaacatgaccccgccattatggccattttgttttg
cttctaactgtttcatcttataagcttttaattatgttatattgtttagtcatattgaagtttcaaataatgttcaaagattctgaatgaagatttggg
acagtgagattctgttgagtttgaatacaataatcattaaatgtaatgtgtttccttgaaatatttaatttcagctagctagagaattgcatcgc
tacataggctaccacagtcaagctgttttagctttaatggtctttgttgtgttttctgtgttctttggggaattggatcatatatactttgcttattctgt
caataaatacacaaaaattgtgcctacgtaaaaaa

16.12

tggagtcggagtcccgtctctgtggataatctcatcgagctgttcttctgaagacatcgatATGAAGGCGCTAATAGGGCT
GCTGCTGCTGTTGGTCTTTGGTCACGGTGTGTCCTCAGTGCTTCACTCTATGCAGTTTTTCT
ACACGGCGTCGTCTGGACTCTCAACCTTCCCAGAGTTTGTTGTTGTTGGGATGGTGGATG
GAGTTCAGATGCTTCACTATGACAGCAACAGTAAGAGAGCCGTACCCAAACATGACGCGG
CGGACCAGCTCTTCAGAGATCATCCGGGTGACCTCGAGGGGGACACTGGAAACTTTCTGG
GTGCCCAGCAGATCTTCAAAGCCAACATTGGGACTCTGAAGCAGCGCTTTAACCAGACAG
GAGGTGCCCACATGGTTCAGTTCATGTATGGTTGTGAGTGGGATGATGAGGATGATTCTAC
TGATGGTTATAGACAGTATGGTTATGATGGAGAGGACTTCATATCGTGGGACATGAAGACC
ATGACCTGGGTCGCTCCAGTACGTCAGGCCTTCACCACCAAACAGAGATGGAATCAGAAT
AAAGCTGGTCTACAATACAGGAAGAACTACATCACCAAGGAGTGTGTTGATTGGCTGAAGA
AGTACCTGGCCTATGGGAAGAGCACTCTGCTGAGAACAGATCGTCCGCGGGTGTCTCTGC
TCCAGAGGAGCCCCTCCTCCCCAGTGGTGTGCCATGCTACAGGCTTCTTCCCTGACAGGG
TGGTGGTGTTCTGGAGGAGAGACGGTCAGGATCTCCAGGAGCAGGTGGACCACGGGGAG
GTCCTCCCCAACCACGACGGGACCTTCCAGGTCAGCGTGGACCTCAACCTCAAGGCCGTC
CCACAGGAGGACTGGAGGAGGTACCAGTGTGTGGTCCAGCTGAGAGGCATCGAGGACAT
CTCCACCCCCCTGGACCCCGCCCTCATCAGGACCAACGGGGGCAAGAGTGGCATCACCG
TCCCCATCATCATTGGGTTCGTTGTTCTCCTCCTCGCTGCTGCTGCTGCTGCTGTTGGAGT
CCTTCTGTACAAGAAGAGGAACGATTCACACAAGCAGCACAAACCAGTTGGTTCTTCTGAC
ACCAGCTCTGAGAACACTGAGGGGCAGAATCTGTCTCCTGAGGCCCGACCTCTGACAGAA
GTTCGAAGTTAAatcatgaagagtcttctcacgttcagcttcctaaagcagggaggagtggatcgcagttcatttataatccttac
aaacaggcttgggacttcagctaactaaagatgctatatttgctaaattacaataacatccatacttaactttattttgataacaatcatgaa
tctgaaagatctgtaaatcaatgagatctctgtatttgtaatagatttgattaattcactggctttaacctgcaatgtttcattgacgcttggaa
accatgacaacacaaactcgccatgatggccattttgttttggttcctgccgcttcaagttataacctttgaatcatgttatagtgtttagtaat
atttaagtttcaaataatggtcaaaggttctgagactagtttttggacagtgagattcggtcaagtttgaatacaataatcgttaaattgtgtg
ttttttcttaaataatattatttgaattgcatcgctacataggctaccacgatcgagctaggcagctttagccttaatccttttgttgtactattttct
ttgttctttggggtatttgatcatagatactttggttattctgtcaataaagtacatccaaattgaaaaaaaaaaaaaaaaaaa

16.13

GTCGCTCCAGTACGTCAGGCCTTCACCACCAAACAGAGATGGGATGGGGACAGAGCTCAG
AATGAACGGTATAAGCACTACCACACCAAGGAGTGTGTTGATTGGCTGAAGAAGCTCCTGG
CCTATGGGAAGAGCACTCTGCTGAGAACAGAGCGTCCGCGGGTGTCTCTGCTCCAGAGGA
GCCCCTCCTCCCCAGTGGTGTGCCATGCTACAGGCTTCTTCCCTGACAGGGTGGTGGTGT
TCTGGAGGAGAGACAGTCAGGAGCTCCAGGAGCAGGTGGACCACGGGGAGGTCCTCCCC
AACCACGACGGGACCTTCCAGGTCAGCGTGGACCTCAACCTCAAGGCCGTCCCACAGGA
GGACTGGAGGAGGTACCAGTGTGTGGTCCAGCTGAGAGGCATCGAGGACATCTCCACCC
CCCTGGACCCCGCCCTCATCAGGACCAACGGGGGCAAGAGTGGCATCACCGTCCCCATC
ATCATTGGGTTCGTTGTTCTCCTCCTCGCTGCTGCTGCTGCTGCTGTTGGAGTCCATCTGT
ACAAGAAGAGGAACGATTCACACAAGCAGCACAAACCAGTTGGTTCTTCTGACACCAGCTC
TGAGACCACTGAGGGGCAGAATCTGTCTCCTGAGGCCCGACCTCTGACAGAAGTTCGAAG
TTAAatcatgaagagtcttctcacgttcagcttcctaaagcagggtggagtggatcgcagttcatttataatccttacaaacaggcttgg
gacttcagctaactaaagatgctacatttgctaaattacaataacatccatacttaactttattttgataacaatcaggaatctgaaagatct
gtaaatcaatgagatctctgtatttgtaatagatttgattaattctctggctttaacctgcaatgtttcactgacgcttggaaaccatgacgac
acaaactcgccatgatggccattttgttttggttcctgccgcttcaagttataacctttgaatcatgttatagtgtttagtaatatttaagtttcaa
ataatggtcaaaggttctgagactagtttttggacagtgagatccggtcaagtttgaatacaataatcgttaaattgtatgttttttcttaaata
atattatttgaattgcatcgctacataggctaccacgatcaagctcggcagctttagccttaatccttttgttgtactattttctttgttctttgggg
aatttgatcatagatactttggttattctgtcaataaagtacatccaaattgaaaaaaaaaaaaaaaaaaa

16.15

gcgtctttgtggatattctcatcgagccttttttccgattacatcgatATGAAGGTGCTAATAGGGCTGCTGTTGTTGT
TCTTTGGTCACGATGTGTCCTCAGTGATTCACTCTGTGCAGGTTTTCTCCACGGCGTCGTC
TGGACTCTCAACCTTCCCTGAGTTTGTGATGGTTCAGATGGTGGATGAGGTTCAGGTTGAG
TACTATGACAGCAACACCCAGAGAATCATAACCAAACAGGACTGGGTGGACCAGGCCAAC
AGAGACAAAGTCCCCGACTACCTGGAGAAGGCAACTGAAAACAGAAAGGGTGAACAGCAG
GTCTTCAAACGCAACATTGGGACCCTGAAGCAGCGCTTTAACCAGACAGGAGGTGCCCAC
ATTGTTCAGAGGATGTTTGGTTGTGAGTTGGATGATGAGGATGATACCACTGATGGTTATG
AGCAGCACGGTTATGATGGAGAGGACTTCCTATCGTTGGACCTGAAGACCCTGACCTGGG
TCGCTCCAGTACATCAGGCTCTCACCACCAAACTGAGATGGGATCACGATACAGCCCGGA
AACAATACTTGAAGAACTTCTACACCAAGGAGTGTGTTGATTGGCTGAAGAAGTACCTGAG
CTACGGGAAGAGCACTCTGCAGAGAACAGAGCGTCCGCGGGTGTCTCTGCTCCAGAGGA
GCCCCTCCTCCCCAGTGGTGTGCCATGCTACAGGCTTCTTCCCTAACAGGGTGGTGGTGT
TCTGGAGGAGAGACGGCCAGGAGCTCCAGGAGCAGGTGGACCACGGGGAGGTCCTCCC
CAACCACGACGGGACCTTCCAGGTCAGCGTGGACCTGAACCTAACGGCCGTCCCACAGG
AGGACTGGGGGAAGTACGAGTGTGTGGTCCAGCTGAAAGGCATCGAGGACATCTCCATCC
TCCTGGACCCCGCCCACATCAGGACCAACGGGGGGGACAAGAGTCTCATCCTTGCTTTCA
TCCTCACTGGGGTTGCTGTTGTTGTTGTTGGTGCTGTTGTTGGAGTCTTTCTGTACTGGAA
GAGGAACGATTCAGACAAGCGTCACAAACCACTTGGTTCTTTCGACACCAGCTCTGAGAAC
ACTGAGGTGCAGAAGCCGTCTCCTGAGGCCCAACCTCTGACCACAGGTCAAAGTTAAatcat
gaagaggcttgtcacgttcggcttcctaaaccaggttggagtggatcgctggtcatttataatccttacaatcaggcttgggacttcagcta
actgcagagtctgcatttgctaaattacaattttcaaacttaactttattgtgataacaatcatcaatctgttagaagtgtcaatcaattacatc
tctctatttgtgatatatttgtataactcgctggcttaaacctgcaatatttcactgacgcttggaaaccatagcaacacgacgtcgccatga
tggccatttgggtcccgccgcttcacgttataaacttttaattatgttatattgtttagtaatatttaagtttcaaataatgttcaaagattctgaat
caagatttttgacagtgagattttgtcatgtttcaatacaataatcgattaatttaagttgtatcattaaaatatattatttagcgttctgctagcta
gaggattgcattgctgcataggctaccacaatcgtgctaggctgctttagcctcaattctttttgttgtattttctgtgttatttggggaattggat
agtatacttgggttattctgtcaataaaatacataaaaattgaaaaaaaaaaaaaaaaaaaaaa

16.22

cctcgtgccgaattcggcacgaggttttattccttgaactcgacagatggaacggtctgcacaggcagggccaaatcgagacgccca
attacttcttggtctcctcctccttggacaaggtcttgattatttcctctttcttggccgcgagacgttcctctctgcgcttgcgagcctccttagtc
ttggagcggcgagcctctgcttggtcagacaagagcttcttgcgggccttgtctgccttcagcttgtggatgtgttccattaggatacgcttg
ttcttgaacacgttacccttggccttaaggtacaggctgtggtacatgtgccggtcgatcttctttgattccctgtagcgacggagaagccg
acgcaagatcctcatgcggcgcatccagcacaacttctctggcatacgggcattggcagtacccctttctcttaccatagcccatgtgtcgt
cccttgcgacgtgcaagggtgttcttgcggcgtcgggcacgagaatgcacggtgacaggcttgcggatgatgagaccatccttaacc
agcttcctgatctgctgacgggagttggcgttggcaatctcattggtctcgttcgggtccaaccataccttcttcttgccacagcgcagaac
gctagaggcaagcctcttctggagcCTGAAGAAGTACCTGGCCTATGGGAAGAGCACTCTGCTGAGAA
CAGATCGTCCGCGGGTGTCTCTGCTCCAGAGGAGCCCCTCCTCCCCAGTGGTGTGCCATG
CTACAGGCTTCTTCCCTGACAGGGTGGTGGTGTTCTGGAGGAGAGACGGTCAGGAGCTCC
AGGAGCAGGTGGACCACGGGGAGGTCCTCCCCAACCACGACGGGACCTTCCAGGTCAGC
GTGGACCTCAACCTCAAGGCCGTCCCACAGGAGGACTGGAGGAGGTACGAGTGTGTGGT
CCAGCTGAGAGGCATCGAGGACATCTCCACCCCCCTGGACCCCGCCTACATCAGGACCAA
CGGGGGCAAGAGTGGCATCACCGTCCCCATCATCATTGGGTTCGTTGTTCTCCTCCTCGC
TGCTGCTGCTGCTGCTGTTGGAGTCCTTCTGTCCAAGAAGAGGAACGATTCACACAAGCA
GCACAAACCAGTTGGTTCTTCTGACACCAGCTCTGAGAACACTGAGGGGCAGAATCTGGC
TCCTGAGGCCCGACCTCTGACAAAAGTTTGGAGTTAAatcatgaagagtcttctcacgttcagcttcctaaag
cagggtggagtggatcgcagttcatttataatccttacaaacaggcttgggacttcagctaactaaagatgctatatttccttaattacaat
aacattaactttattttgataacaatcatgaatctgaaagatctgtaaatcaattagatctctgtatttgtaatagatttgaataattcactggct
ttaacctgcaatgtttcactgacgcttggaaaccatgacaacataaactccccatgatggccattttgttttggttcctgccgcttcaagttat
aacctttttaatcatgttatagtgtttagtaatatttaagtttcaaataatggtcaaaggttctgagactagtttttggacagtgagattcggtcaa
gcttgaatacaataatcgttaaattgtgtgttttttcttaaataatattatttgaattgcatcgctacataggctaccacgatcgagctaggcag
ctttggccttaatccttttgttggactattttctttgttctttggggaattttatcatagatactttgtttattctgtcaataaagtacatcaaaattgaa
aaaaaaaaaaaaaaaa

16.60.3

GTCGCTCCAGTACGTCAGGCCTTCACCACCAAACAGAGATGGGATGGGGACAGAGCTCAG
AATGAACGGTATAAGCACTACCACACCAAGGAGTGTGTTGATTGGCTGAAGAAGCTCCTGG
CCTATGGGAAGAGCACTCTGCTGAGAACAGAGCGTCCGCGGGTGTCTCTGCTCCAGAGGA
GCCCCTCCTCCCCAGTGGTGTGCCATGCTACAGGCTTCTTCCCTGACAGGGTGGTGGTGT
TCTGGAGGAGAGACAGTCAGGAGCTCCAGGAGCAGGTGGACCACGGGGAGGTCCTCCCC
AACCACGACGGGACCTTCCAGGTCAGCGTGGACCTCAACCTCAAGGCCGTCCCACAGGA
GGACTGGAGGAGGTACCAGTGTGTGGTCCAGCTGAGAGGCATCGAGGACATCTCCACCC
CCCTGGACCCCGCCCTCATCAGGACCAACGGGGGCAAGAGTGGCATCACCGTCCCCATC
ATCATTGGGTTCGTTGTTCTCCTCCTCGCTGCTGCTGCTGCTGCTGTTGGAGTCCATCTGT
ACAAGAAGAGGAACGATTCACACAAGCAGCACAAACCAGTTGGTTCTTCTGACACCAGCTC
TGAGACCACTGAGGGGCAGAATCTGTCTCCTGAGGCCCGACCTCTGACAGAAGTTCGAAG
TTAAatcatgaagagtcttctcacgttcagcttcctaaagcagggtggagtggatcgcagttcatttataatccttacaaacaggcttgg
gacttcagctaactaaagatgctacatttgctaaattacaataacatccatacttaactttattttgataacaatcaggaatctgaaagatct
gtaaatcaatgagatctctgtatttgtaatagatttgattaattctctggctttaacctgcaatgtttcactgacgcttggaaaccatgacgac
acaaactcgccatgatggccattttgttttggttcctgccgcttcaagttataacctttgaatcatgttatagtgtttagtaatatttaagtttcaa
ataatggtcaaaggttctgagactagtttttggacagtgagatccggtcaagtttgaatacaataatcgttaaattgtatgtttttttcttaaata
atattatttgaattgcatcgctacataggctaccacgatcaagctaggcagctttagccttaatccttttgttgtactattttctttgttctttgggg
aatttgatcatagatactttggttattctgtcaataaagtacatccaaattgaaaaaaaaaaaaaaaaaa

Figure 3.4: Sequences found in the reverse complement orientation. Sequences are shown with ORFs in uppercase letters. A) Clone A16; B) Clone 1.2; C) Clone 16.14; D) Clone 16.16; E) Clone 16.26; F) Clone 17.3

A16

ggcacgaggagtatctacgataaaattccccaaagaacaaagaaaatagtacaacaaaaggattaaggcaaaagctg
cctagcttgatcgtggtagcctatgtagcgatgcaattcaaataatattattaagaaaaaacatacaatttaacgattattgtattcaaactt
gaccgaatcccactgtccaaaaactagtctcagaacctttgaccattatttgaaacttaaatattactaaacactataacatgattcaaag
gttataacttgaagcggcaggaaccaaaacaaaatggccatcatggcgagtttgtgttgtcatggtttcaaagcgtcagtgaaacattg
caggttaaagccagtgaattaatcaaatctattacaaatacagagatctcattgatttacagatctttcagattcatgattgttatcaaaaca
aagttaagtatggatgttattgtaatttagcaaatatagcatctttagttagctgaagtcccaagcctgtttgtaaggattataaatgaactgc
gatccactccaccctgctttaggaagctgaacgtgagaagactcttcatgatTTAACTTTGAACTTCTGTCAGAGGTC
GGGCCTCAGGAGCCAAATTCTGCCCCTAAGTGTTCTCAGAGCTGGTGTCAGAAGAACCAA
CTGGTTTGTGCTGCTTGTGTGAATCGTTCCACTTCTTGTACAGAAGGACTCCAACAGCAGC
AGCGAGGAGGAGAACAACGAACCCAATGATGATGGGGACGGTGATGCCACTCTTGCCCC
CGTTGGTCCTGATGAGGGCGGGGTCCAGGGGGGTGGAGATGTCCTCGATGCCTCTCAGC
TGGACCACACACTCGTACCTCCCCCAGTCCTCCTGTGGGACGGCCTTGAGGTTGAGGTCC
ACGCTGACCTGGAAGGTCCCGTCGTGGTTGGGGAGGACCTCCCCGTGGTCCACCTGCTC
CTGGAGATCCTGACCGTCTCTCCTCCAGAACACCACCACCCTGTCAGGGAAGAAGCCTGT
AGCATGGCACACCACTGGGGAGGAGGGGCTCCTCTGGAGCAGAGACACCCGCGGACGCT
CTGTTCTCAGCAGAGTGCTCTTCCCATAGGCCAGGTACTTCTTCAGCCAATCAACACACTC
CTTGGTGTAGTAGTTCTTATTGTATTGTAGATCAGCTTTATTCTGATCCCATCTCTGTTTGGT
GGTGAAGGCCTGACGTACTGGAGCGACCCAGGTCAGGGTCTTCAGGTCCAACGATATGAA
GTCCTCTCCATCATAACCAAACTGGTCATAACCTTCAGTAGAATTATCCTCATCATCCCACT
CACAACCATACATCCTCTGAACCATGTGGGCACCTCCTGTCTGGTTAAAGCGCTGCTTGGC
AATCCCAATGTTGCCTTTGAAGGTCTGCTGGGCACCCAGAAAGCTTCCAGTTTCCCTCTTC
AGGTAGTCGGGGTCACCTCTGGTGACCTGCTCCATCCAGGCCTGTTTGAGGACTACTCTC
TGGGTGTTGCTGTCATAGTAGTCAAACTGAACTCCATCCACCATCCCAACAACAACAAACT
CTGGGAAGGTTGAGAGTCCAGACGACGCCGTGTAGAAATAATGGAGAGAGTGAAGCAGTG
AGGACACACCGTGACCAAAGACCAACAGCAGCAGCAGCCCTATTAGCGCCTTCATatcgatgt
attcagaagaacagctcgatgagattatccacagagacggg

1.2

ggcacgagggtactttattgacagaataaccaaagtatctatgatcaaattccccaaagaacaaagaaaatagtccaacaaaagga
ttaaggctaaagctgcctagctcgatcgtggtagcctatgtagcgatgcaattcaaataatattattaaagaaaaaacatacaatttaac
gattattgtattcaaacttgaccgaatctcactgtccaaaaactagtctcagaacctttgaccattatttgaaacttaaatattactaaacact
ataacatgattcaaaggttataacttgaagcggcaggaaccaaaacaaaatggccatcatggggagtttgtgttgtcatggtttccaag
cgtcagtgaaacattgcaggttaaagccagtgaattaatcaaatctattacaaatacagagatctaattgatttacagatctttcagagtc
atgattgttatcaaaataaagttatgtatggatgttattgtaatttagcaaatgtagcatctttagttagctgaagtcccaagcctgtttgtaag
gattataaatgaactgcgatccactccaccctgctttaggaagctgaacgtgagaagactcttcatgatTTAACTTTGAACTTT
TGTCAGAGGTCGGGCCTCAGGAGACAGATTCTGCCCCTCAGTGTTCTCAGAGCTGGTGTC
AGAAGAACCAACTGGTTTGTGCTGCTTGTGTGAATCGTTCCTCTTCTTGGACAGAAGGACT
CCAACAGCAGCAGCAGCAGCAGCGAGGAGGAGAACAACGAACCCAATGATGATGGGGAC
GGTGATGCCACTCTTGCCCCAGTTTTTCCTGATGAGGGCGGGGTCCAGGGGGGTGGAGA
TGCCCGCGATGCCTCTCAGCTGGACCACACACTCGTACCTCCTCCAGTCCTCCTGTGGGA
CGGCCTTGAGGTTGAGGTCCACGCTGACCTGGAAGGTCCCGTCGTGGTTGGGGAGGACC
TCCCCGTGGTCCACCTGCTCCTGGAGATCCTGACCGTCTCTCCTCCAGAACACCACCACC
CTGTCAGGGAAGAAGCCTGTAGCATGGCACACCACTGGGGAGGAGGGGCTCCTCTGGAG
CAGAGACACCCGCGGACGCTCTGTTCTCAGCAGAGTGCTCTTCCCATAGGCCAGGTACTT
CTTCAGCCAATCAACACACTCCTTGGTGTAGTAGTTCTTAGTGTATTGTAGAATAGCTTTAT
CCTGATTCAATCTCTGTTTGGTGGTGAAGGCCTGACGTACTGGAGCGACCCAGGTCAGGG
TCTTCAGGTCCAACGATAGGAAGTCCTCTCCATCATAACCATACTGGTGATAACCATCAGTA
GAATCATCCTCATCATCCCACTCACAACCATACATGTTCTGATATAAGTGGGTACCTCCTGT
CTGGTTAAAGCGCTGCTTGGCAATCCCAATGTTGGCTTTGAAGGTCTGCTGGTCACCCTGA
GCCCTTCCAGTGTCCCTCTCGAGTTCACCCGGATGATCTCTGTAGATCTGGTCCGCCGCG
TCATGTTTGGGTACGGCCCTCTTACTGTTGCTGTCATAGTGAACCATCTGAACTCCATCCA
CCATCCCCACAACAACAAACTCTGGGAAGGTTGAGAGTCCAGACGACGCCGTGTAGAAAT
ACTGCATAGAGTGAAGCACTGAGGACACACCGTGACCAAAGACCAACAGCAGCAGCAGCC
CTATTAGCGCCTTCATatcgatgtattcagaagaacagctcgatgagattatccacagagacgggactcccttcaacggct
cagtcttcgccgaactccagaggaaactcctgagatgtccctgctcacagagaaaagagagagagtcagagagaggaggggggg
agagagatagatggagaaagaaaaggaaagagagacacacatgatgagtgaagagacgaagagaggatcacttgagggtca
gaacaagtcggggagagtggtgattgcagcagcattactcactagctagtgactagtattgaatgcgcttatcagtctcgttgcgtatac
agggactgacacgacgatgcttctctccacattgtgagcgattcacgactgatcttgtcgagcgtcaattttgacgcaatgcagatgcac
cccatatgctgcaacacctttgctggtgatcgagtggttggagtggactctgaggtccttgcaggcccaggactgcatatgtgccaagat
gccagtgtctctcattcatgcacacatgcatgtctgagacacgcacggatctgtgtattcaaaattgtataatagaatatgaattatttttgtc
gtactgaagccactttgtacccaatcatgataacagattatggtatgtgatcaaaaggcctttcgtttttatttcttcttggtggctgccaattg
cggatgttcccatcaaagaagcaaacctgtgaatggtttcggtcactggaaaaatctacaggattatgtcgtcaacagcagattagattt
attttcatgttgggcatatctggcagtgtagatagtagcgtttaacttattcagagttgtcttattttgtcggttttttaactgtgatacatataattt
agctgaaggtctaacttgaaactagcttatttagatccccaatgtatccatagtgtaaatataccagaggattaaaaattgtttactgagg
aataaacgttgttaattattattttattaagactcattttagaagcagtgcagcaaattatttgaacattcatcattccacgctcaatatgcactt
actcaaaatacttggttcgacatttaatctgtctgctttttattgatgaggttagaaaatcctcccaaatagattctgcaactccaaaaaata
ggcagttggtaactcacaacttttagacaattaagaaaggccccatgatgtacatgtctattgtccccattaggaattccatgatgtgttat
gttttatttttatttggtatatgtgttttagtcctgtcatattgtattaacctaatgtggaattgcaattccttgttattcacatgtcatggtcatgcaca
ataaaacaatgtgccctcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa

16.14

ggcacgaggtattttattgacagaataacccaagtatactatccaattcccaaaagaacacagaaaatacaacaaagag
cactaaggctaaagcagcctagcacgattgtggtagcctatgttgtagcaatgcaatgctctagctagcagaacgctcaaataatatat
ttgaatgaaacacattaaattaatcgattattgtattgaaacttgaccgaatctcactttccaaaatcttgattcagaatttttgaacattatttg
aaacttaaatattacttaacaatataacataattaaaaggtttcaacgtgaagcggcggggcccaaatggccatcatggcgacgtcgt
gttgctatggtttccaagcgtcagtgaaagattgcaggtttaagccctcgagttattcaaatctatcacaaatagagagatgtaattgattg
acacttctaacagattgatgattgttatcacaataaagtcaagtttgaaaattgtaatttagcaaatgcagactctgcagttagctgaagtc
ccaagcctgattgtaaggattataaatgaacagcgatccactccatcctggtttaggaagccgaacgtgagaagcctcttcatgatTT
ATGCGGTCAGAGGTTGGTACTCAGCAGACAGCTTCTGCCCCTCAGTGTTCTCAGAACGCA
TGTCAGAACCGTTCCTCTTCTGGTACAGAAAGACTCCAACAACAGCAGCAACAACAACAAC
AGCAACACCAGTGAGGATGAAAGCAAGGATGAGACTCTTGTCCCCCCCCGTTGGTCCTGAT
GTGGGCGGGGTCCAGGAGGATGGAGATGTCCTCGATGCCTTTCAGCTGGACCACACACT
CGTACCTCCCCCAGTCCTCCTGTGGGACGGCCTTTAGGTTCAGGTCCACGCTGACCTGGA
AGGTCCCGTCGTGGTTGGGGAGGACCTCCCCGTGGTCCACCTGCTCCTCGAGCTCCTGG
CCGTCTCTCCTCCAGAACACCACCACCCTTTTAGGGAAGAAGCCTGTAGCATGGCACACC
ACTGGGGAGGAGGGGCTCCTCTGGAGCAGAGACACCCGCGGACGCTCTGTTCTCTGCAG
AGTGCTCTTCCCGTAGCTCAGGTACTTCTTCAGCCAATCAACACACTCCTTGGTGTATAAGT
TCTTCAAGTGTTGTTTATAAGCTGTATTCTGATCCAATCTCAGTTTGGTGGTGAGAGCCTGA
TGTACTGGAGCGACCCAGGTCAGGGTCTTCAGGTCCAACGATATGAAGTCCTCTCCATCAT
AACC

16.16

ggcacgaggatatatattaattatattatatgatggccaaatgatactactgcatagacgcatagggtgtaggtacctagtcc
acagaggcctttgctttatattgcgtacacctctccaaaggttgaacaatgcctaacggcaactgacggcctacccgcgtcttacctgct
atgacagacgagttgaccaagtttgcacatcttttaatttgtactacacatcttagagatcccagacttcacttagtttgcgacgcttggtag
agacccgacgtgcctacagcagcttaccatcagctgctctccgacccagaagggcaatataggctttagtgtgcattgcatggtaccat
gcactgctacaatgtgcttgatactacttaggccacggatatcccatcctcagggattttagatccgcaacatcttatgtatttgaagtgatt
attaggcgtacaatagtgttggcatacaatagaggttgagtatgcttgtagaaaaagtttgaagaaatactttgatgcaatgattgtatttg
aattgttatataagattcgactgagtaagttcctcagaatatttgtacttttcgaagaacaatttcttttgaattaaaatagatatttgaaaaaa
aaaaaaaaagttttcgtaaacaccttctttattcgtttcggtagcaaaagcgccagcttgacgttcatgaggtcctggtactcgcggatctg
gcgggccatgtcctgcttggctctctgcagggcatcctccaggtccttgatgcgggacttggcgtcgcgcacggccatctcgccgcgct
cctccgcctcggcgatctgggcctccaggtttccacgctgtcctttgaatccgtcgatctcggccgtcagcctctggatcatgcggttgag
gtcagcgatctcggtcttggtggacctcaggtcgtctccgtaccgggtggccgacgtctgcatctcctcgtacttggacttgtaccaggtct
cggcctcagcgcggctgcggttggcgatgtcctcgtactgggcgcgcacctccgccacgatggagtccatgtccagggtgcggctgtt
gtccatctccacgatgaccgacgtgtccttgatctggctctggagctcacgcagctcctcttcatagatctgcctgaggaagttgatctcgt
cggtcagactctccagcttggcctccagctcaaccttgttcatgtaggcctcgtccacgtccttcttgatgagcacaaagtcgagaataa
ccaaagtatctatgatcaaattccccaaagaacaaagaaaatagtccaacaaaggattaaggctaaagctgcctagctcgatcgtg
gtagcctatgtagcgatgcaattcaaataatatttcttaagaaaaaacacacaatttaacgattattgtattcaaacttgaccgaatctcac
tgtcaaaactagtctcagacttgaccattattgaactaatatactacactatacatgattcaaggtatactgaagcggcaggaaccaaa
accaaaaaaaaagatctatgtatagtccggccatcatgtcgagtttgtgtcgtcatggtttccaagcgtcagtgaaacattgcaggttaa
agccagtgaattaatcaaatctattacaaatacagagatctaattgatttacagatctttcagattcatgactgttatcaaaataaagttaa
gtatggatgttattgtaattaagcaaatatagcatctttagttagctgaagtcccaagcctgtttgtaaggattataaatgaactgcgatcca
ctccaccctgctttaggaagctgaacgtgagaagactcttcatgatTTAACTTTGAACTTCTGTCAGAGGTCGGGC
CTCAGGAGACAGATTCTGCCCCTCAGTGTTCTCAGAGCTGGTGTCAGAAGAACCAACTGG
TTTGTGCTGCTTGTGTGAATCGTTCCTCTTCTTGTAGAGAAGGACTCCAACAGCAGCAGCA
GCAGCAGCGAGGAGGAGAACAACGAACCCAATGATGATGGGGACGGTGATGCCACTCTT
GCCCCCGTTGGTCCTGATGAGGGCGGGGTCCAGGGGGGTGGAGATGTCCTCGATGCCTC
TCAGCTGGACCACACACTCGTACCTCCTCCAGTCCTCCTGTGGGACGGCCTTGAGGTTGA
GGTCCACGCTGACCTGGAAGGTCCCGTCGTGGTTGGGGAGGACCTCCCCGTGGTCCACC
TGCTCCTGGAGATCCTGACCGTCTCTCCTCCAGAACACCACCACCCTGTCAGGGAAGAAG
CCTGTAGCATGGCACACCACTGGGGAGGAGGGGCTCCTCTGGAGCAGAGACACCCGCGG
ACGCTCTGTTCTCAGCAGAGTGCTCTTCCCATAGGCCAGGTACTTCTTCAGCCAATCAACA
CACTCCTTGGTGATGTAGTGCTTCAG

16.26

tggatccaaagtattgacagaataaccaaagtatctatgatcaaattccccaaagaacaaagaaaatagtccaacaaaa
ggattaaggctaaagctgcctagctcgatcgtggtagcctatgtagcgatgcaattcaaataatattattaaagaaaaaacacacaattt
aacgattattgtattcaaacttcaccggatctcactgtccaaaaactagtctcagaacctttgaccattatttgaaacttaaatattactaaa
cactataacatgattcaaaggttataacttgaagcggcaggaaccaaaacaaaatggccatcatggcgagtttgtgtcgtcatggtttc
caagcgtcagtgaaacattgcaggttaaagccagagaattaatcaaatctattacaaatacagagatctcattgatttacagatctttca
gattcatgattgttatcaaaataaagttaagtattgatgttattgtaatttagcaaatatagcatctttagttagctgaagtcccaagcctgtttg
taaggattataaatgaactgcgatcccctccaccctgctttaggaagctgaacgtgagaagactcttcatgatttaactttgaacccctca
gtgttctcagagctggtgtcagaagaacctgggaagtagcagaatgggaggaggtgaagtcatgagctagatgtatctagatctatag
agctgtatagtccagtgagctagatgtaCTAAGCGTTCCTCTTCTTGTACAGAAGGACTCCAACAGCAGC
AGCAGCAGCAGCAGCGAGGGGGAGAACAACGAACCCAATGATGATGGGGACGGTGCTGC
CACTCTTGCCCCCGTTGGTCCTGATGAGGGCGGGGTCCAGGGGGGTGGAGATGTCCTCG
ATGCCTCTCAGCTGGACCACACACTCGTACCTCCTCCAGTCCTCCTGTGGGACGGTCTTG
AGGTTGAGGTCCACGCTGACCTGGAAGGTCCCGTCGTGGTTGGGGAGGACCTCCCCGTG
GTCCACCTGCTCCTGGAGCTCCTGACCGTCTCTCCTCCAGAACACCACCACCCTGTCAGG
GAAGAAGCCTGTAGCATGGCACACCACTGGGGAGGAGGGGCTCCTCTGGAGCAGAGACA
CCCGCGGACGCTCTGTTCTCAGCAGAGTGCTCTTCCCATAGGCCAGGTACTTCTTCAGCC
AATCAACACACTCCTTGGTGTAGTAGTGCTTATCGTATTGTAGATCAGCTTTATCCTGATTC
CATCTCTGTTTGGTGGTGAAGGCCTGACGTACTGGAGCGACCCAGGTCAGGGTCTTCAGG
TCAAACGATAGGAAGTCCTCTCCATCATAACCATACTGGTCATAACCATCAGTAGAATCATC
CTCATCATCCCACTCACAACCATACATGTTCTGATATAAGTGGGTACCTCCTGTCTGGTTAA
AGCGCTGCTTCAGATTCCCAATGTCGGCTTTGAAGGTCTGCTGGTCACCCTGAGAGATTGC
AGTGTTCCTCTCCAGGTAGGCGGGgagaaggccgaacggaaggtcgaccgaccgacgccaaagcgcagcac
gcgtccgctggtgaaggtgcccggctggaagttcaacgtcatgttgtagaacatgttcgccaccgccggagccggcgcctggttcgtg
aactgtgccgtcacgttgctggcgaccaggccgacgctggccgggccgaacgtgaagggcaggcctgcggtgcgcacgtcgaag
accagacccggcgtgctgggcacgacgacgttgccgccgcccggattgcccgtcgctgcattgaacgtgatgctggtgatgtagctg
gggccgttgtaggacaccgtgaaggcattggggtcggtgcgaccggtgacggagttgtcgcccaccagggtcaccgcaaccttgcc
gcggttgggcagtgcctgcgccaccgacgacgagctgtacgggtccaggtcgttcgggtaggtgctgttctccaggatggtgcgcattt
gcgccggtgtcaccgaaagtac

17.3

ccaagtatatttgatccaattccccaaagaacacagaaaatacaacaaagagcattacggctaaagcagcctagctcgagtgtggta
gccaacgtagcaatgcaatgctctagctagcagaactctaaaataatacatttgaaggaaacacattcaatttaacgaatattgtattca
aacttgacaaaatctcattgtccaaaattttaattcggaatctttaaacattatttaaaatgtaagtattactaaacaatataacataattaaa
aggttatagcatgaagcggtgggaaccaaaacaaaatggtttccaagcatcagcgaaagattggatgtttaagccagtgagttaatc
aaacagaacacaaatacagcgttttaatttactgacacttttctcacggattgatggttgctatcaaggcaaaataaagttaagtatgaat
gttattgtaattgagcaaatgcagactctgcagttagcccaggtcccaagcttgattgtaaggattataaatgaacagcgatccgctcca
ccctggtggaggaagatgaacgtgcgaagcctcttcaatTTAACTTTGACCTGTGGTCAGAGGTTGGGCCTC
CGGATCCTGCACCTCAGTGTTCTCAAAGCTGGTGTCAGAAGAACCAAGTGGTTTGTGACG
CTTATCTGAATCGTCCCTCTTGCAGTATACCAAGACTCCAGCAGTAGCAACAACACCAACA
ACAGCAACACCAGCGAGGATACCAGCAACAATACTGTGGGGGGGGACGTGCGACGTCTG
AGCCCCGTTGGTCCTGATGAGGGCGGGGTCCAGGGGGGTGGAGATGTCCTCGATGCCTC
TCAGCTCGACCACACACTCGTACCTCCCCCAGTCCTCCTGTGGGACGGCCATGAGGTTGA
GGTCCACGCTGACCTGGAAGGTCCCGTCGTGGTTGGGGAGGACCTCCCCGTGGTCCACC
TGCTCCTGGAGAACCTGACCGTCTCTCCTCCAGAACACCACCACCCTGTCAGTGTAGAAG
CCTGTAGCATGGCACACCACTGGGGAGGAGGGGCTCCTCTGGAGCAGAGACACCCGCGG
ACGGTCTGCCTGCAGAGTGCTCTTCCCATTGGCCAGTAGCTTCTTCAGCCAATCAACCTCC
TTGGTGTAGTAGTCCTTCTTCCTTTGTAAATCAGCTCCATCCTGATCCCATCTCTGTTTGGT
GGAGAAAGCCTGACGTACTGGAGCGACCCAGGTCTTGGTCTTCAGGTCCAACGATAGGAA
GTCCTCTCCATCATAACCATGCCGCTCAAAACCATCAGTAGAATCATCCTCATCATCCAAAT
CACAACCAAACGTCTCCTGATATACATGGACACGTTGTTTGTGGTTAAAGCTCTTCTTGGCA
TTCCCAAGGTTGGCTATGGAGCTCTGCTGGGCAACCAGAAAGCTTCTAGTTGTCGTCACCA
CGTAGGCCGGGTCATCTCTGGTGAACTTCTCCATCCAGTCCGGTTTGAGGAATATTCTCTG
GGTGTTGCTGTCGTAGTACTCAAACTGAACTCCATCCACCATCCCCACAGCAACAAACTCT
GGGAAGGTTGAGAGTCCAGACGACGCCGTGTAGAAATACTGCAGAGAGTGAGGCACTGA
GGACACACCGTGACAAAAGACCAACAGCAGCAACCCTATTAGCGCCTTCATatcgatgtattccga
ggaacaactcgatgagattatccacagagacgtataggctactccgactccaggacgacgcccaaatggttgaatggacagatgcc
gagcggaggcacatcgagacgatctggggcaagatcgacattgatgtctgcggaccactcgcgttgcagagatgcctgattgtgtatc
cgtggacgcagcgctacttcggtagctttggcgacctgagcaccgacgccgctatcgtgggaaaccccaaggtggccaagcacgg
cgtcgtggccctgaccggactgaagacggctctggaccacatggacgacatcaaggccacctacgctgccctgagcgtgctgcact
ccgagaaactgcacgtcgaccccgacaacttccgactgctgtgtgactgcctgacgattgtcgtcgccgggaagatggggaagaga
ctgagccccgagatgcaggctgcgtggcagaagtacctctccgcggtggtttccgccctggggagacagtaccactagaagatgtgt
gttgacgctagcgggggacaacagcaagctacaacttcagtgacaaacccgaaatggatgttttcatctgtcaataaaactattgttta
acagtcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
aaaaaaaaaaaa

sequence at either their 5' or 3' ends. Additionally, only six sequences, Clones 1.2, 16.12, 16.13, 16.22 and 16.60.3 had Xho I sites preceding the 3' vector sequence. Therefore, only these six sequences were confirmed to be in the correct orientation. Further investigation is needed to confirm that the remaining sequences were indeed directionally cloned.

Of the six clones which aligned in the reverse complement orientation, only two sequences had poly-A tails at the 3' end in the sense orientation. The structure of one of the clones, Clone 1.2, mimicked the structure of Haddock UA pBK-CMV-3 in that it appeared to have an inverted open reading frame. The 5' end of the sequence shows significant similarity with the antisense 3'UTR of Haddock UA pBK-CMV-4. The 3' end, however, was extremely long and did not align with any other sequences. It includes a polyadenylation signal and poly-A tail in the sense orientation. Clone 17.3 was the other sequence with a sense orientation poly-A tail. This sequence, however, appeared to be linked to the sequence of another gene and will be discussed in the following section.

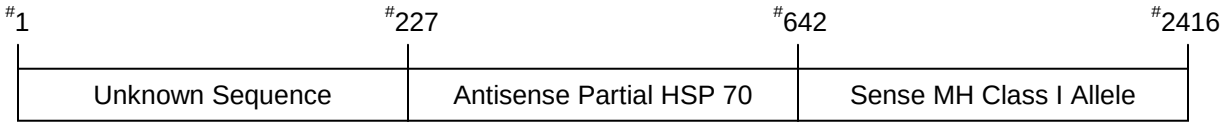
3.2.2. Linked Sequences

Five sequences contained portions, either at the 5' or 3' ends, which did not align with any of the other alleles. In four of these cases the MH Class I sequence appeared to be joined to the sequence of another gene. A diagram illustrating the sequence structure is shown in Figure 3.5.

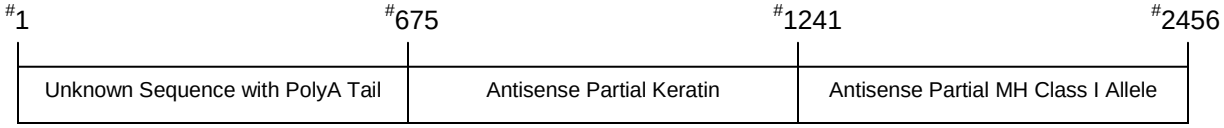
Clone C7 is 2416 base pairs (bp) in length. The first 226bp of sequence did not align with any other sequence. The sequence from 227 to 641bp aligned in the antisense orientation with a heat shock protein (HSP) 70 from several fish species including *Ctenopharyngodon idella* (grass carp) and *Hypophthalmichthys molitrix* (silver carp). This stretch of sequences does not include the entire HSP 70 sequence, however. The last of the sequence, from 632bp to 2416bp, aligned in the sense orientation with Haddock UA pBK-CMV-4and includes the complete 5' and 3' UTRs with the entire ORF.

Figure 3.5: Diagram illustrating the structure of "linked" sequences. A) C7; B) 16.16, C) 16.22; D) 16.26; E) 17.3

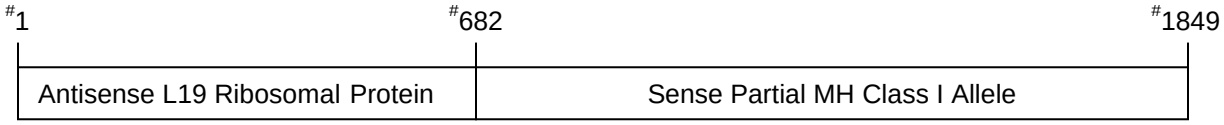
A) C7



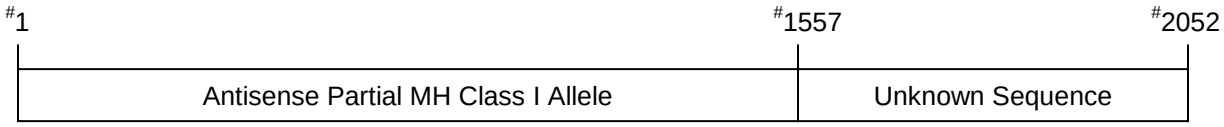
B) 16.16



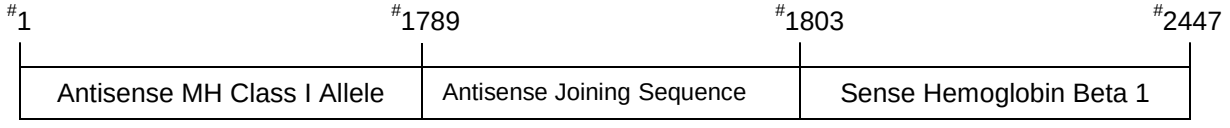
C) 16.22



D) 16.26



E) 17.3



Like Clone C7, Clone 16.16 was comprised of three segments of sequence. The first 674pb of the sequence did not show any significant similarity to any other sequences. This portion of the sequence, however, had a polyadenylation signal and poly-A tail.

The next segment of Clone 16.16 was from 675bp to 1241pb. This segment aligned in the antisense orientation with keratin sequences from several fish species including *Oncorhynchus mykiss* (rainbow trout), *Gasterosteus aculeatus* (stickleback), and *Ctenopharyngodon idella* (grass carp). Once again, this keratin sequence was only a partial sequence. The final segment was from 1242bp to 2456bp. This segment aligned with Haddock UA pBK-CMV-4 in the antisense orientation. Unlike Clone C7, though, this segment did not contain an entire ORF. Roughly 500bp of the 5' ORF were missing.

Clone 16.22 was divided into two sections. The 5' section contained a sequence which showed significant alignment in the antisense orientation with the ribosomal protein L19 of several fish species, including *Solea senegalensis* (sole), *Oncorhynchus masou* (cherry salmon), and *Salmo salar* (Atlantic salmon). The 3' section contained a partial sequence for an MH Class I allele. The first roughly 575 bp of ORF was missing. This sequence aligned with Haddock UA pBK-CMV-4 in the sense orientation.

Unlike the other clones with "linked" sequences, the non-aligning portion of Clone 16.26 did not show any significant similarity to other gene sequences. The 5' portion of the sequence was an MH Class I allele that aligned in the antisense orientation with Haddock UA pBK-CMV-4. Again, the first roughly 220bp of the ORF were missing. The 3' portion of the sequence was roughly 465bp in length.

Clone 17.3 was divided into two sections. The 5' section contained a complete ORF and both the 5' and 3' UTR of an MH Class I allele, which aligned with Haddock UA pBK-CMV-4 in the antisense orientation. A short sequence of 14bp linked this portion of the sequence with the sequence at the 3' end. In the antisense orientation this sequence is very similar to the sequence that linked the two gene segments seen in Haddock UA pBK-CMV-1. The 3' portion of the sequence contained a sequence highly similar to the hemoglobin beta 1 chain from *Gadus morhua* (cod) in the sense

orientation. Partial coding sequences of beta 1 globin from *Boreogadus saida* (polar cod) and *Arctogadus glacialis* (artic cod) also showed similarity to the 3' end in the sense orientation.

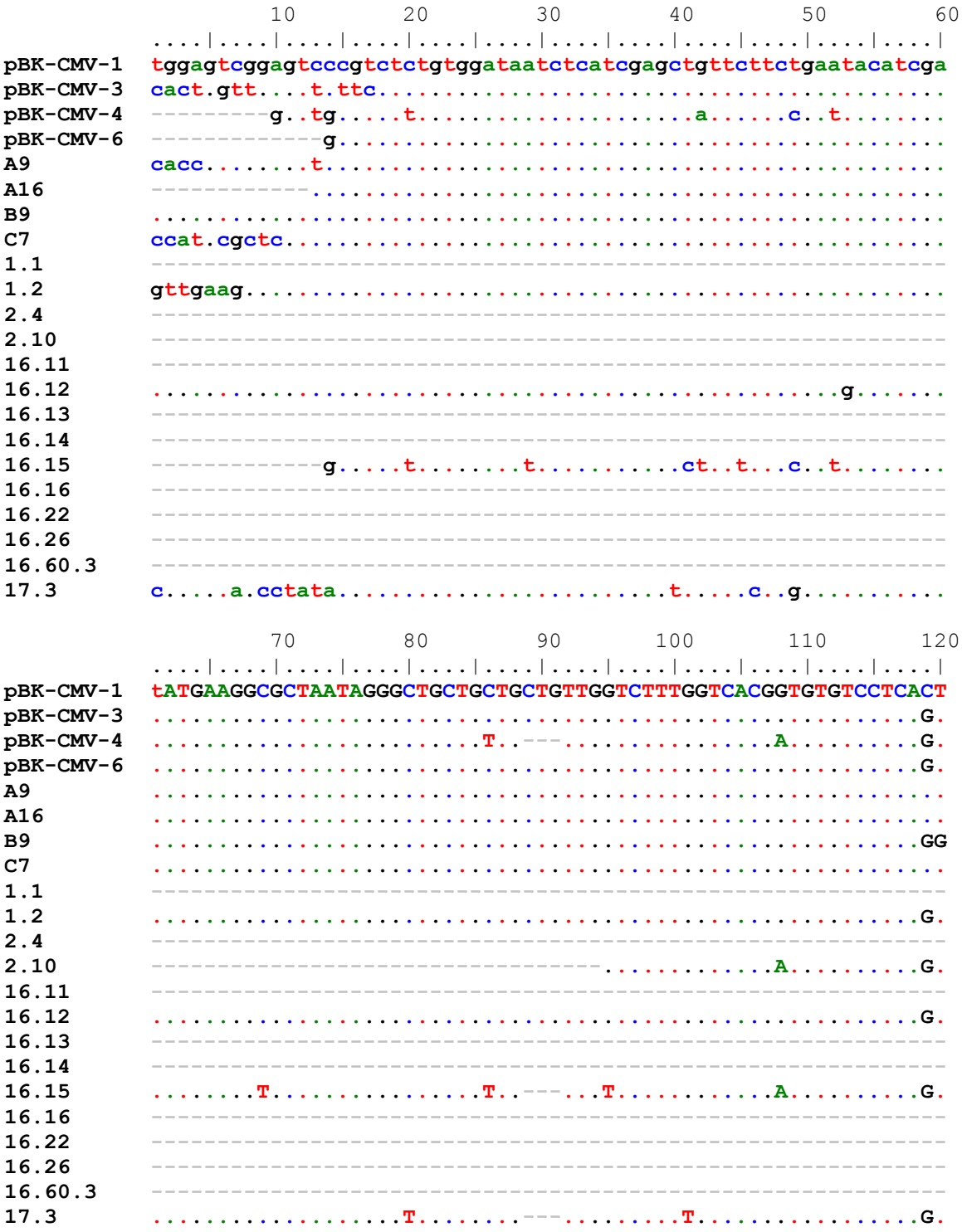
It is interesting to note that for Clones 16.22 and 17.3 the sequences are comprised of transcripts in both the sense and antisense orientations. Because of this, both strands of the cDNA would contain a poly-A tail. Since the poly-A tail was used to establish directional cloning the exact orientation of these sequences would still be in question.

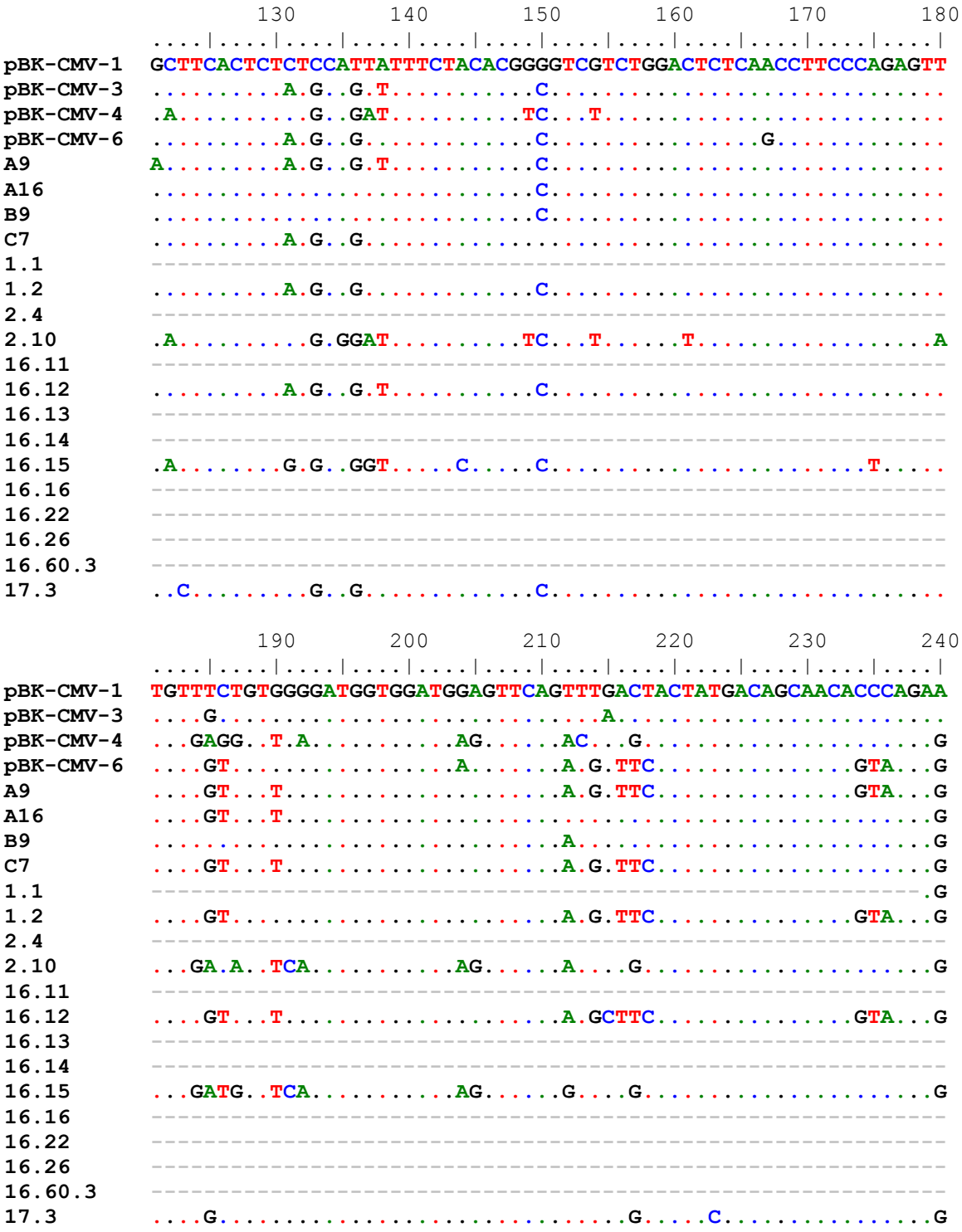
3.2.3. Sequence Alignments and Phylogenetic Tree

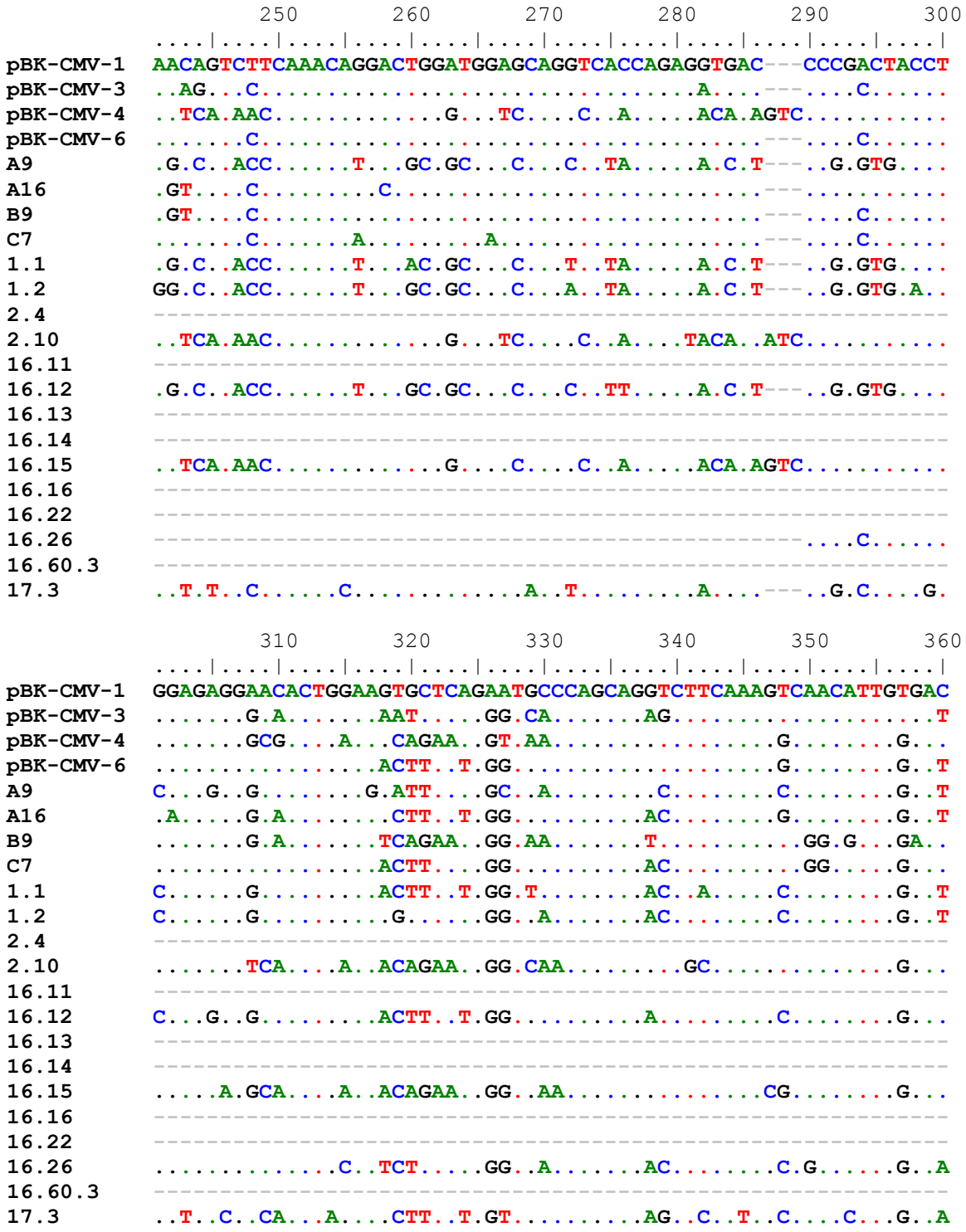
All Haddock MH Class I alleles, including those previously isolated and sequenced, were aligned with each other. To compare the alleles themselves non-MH Class I sequences were removed and the orientation of any antisense sequences was reversed. The DNA sequence alignment is shown in Figure 3.6. From this alignment it was seen that there were clusters of highly variable sequences. These were at base pairs #240 – 365, #398 – 410, and #531 – 594. Within the ORF there were two sites where certain alleles had large insert sequences. At bp #969 allele Haddock UA pBK-CMV-4 had a 30bp insert and allele 17.3 had a 21bp insert. At bp #1137 both alleles A9 and 16.11 had a 27bp insert. Finally, allele 16.26 had a stop codon at position # 1110 followed by a 93bp insert. Allele 16.14 had a 30bp deletion which flanked the insert at bp #1137. Additionally, 3 and six nucleotide deletions and additions were noted throughout the ORF. These are summarized in Table 3.1.

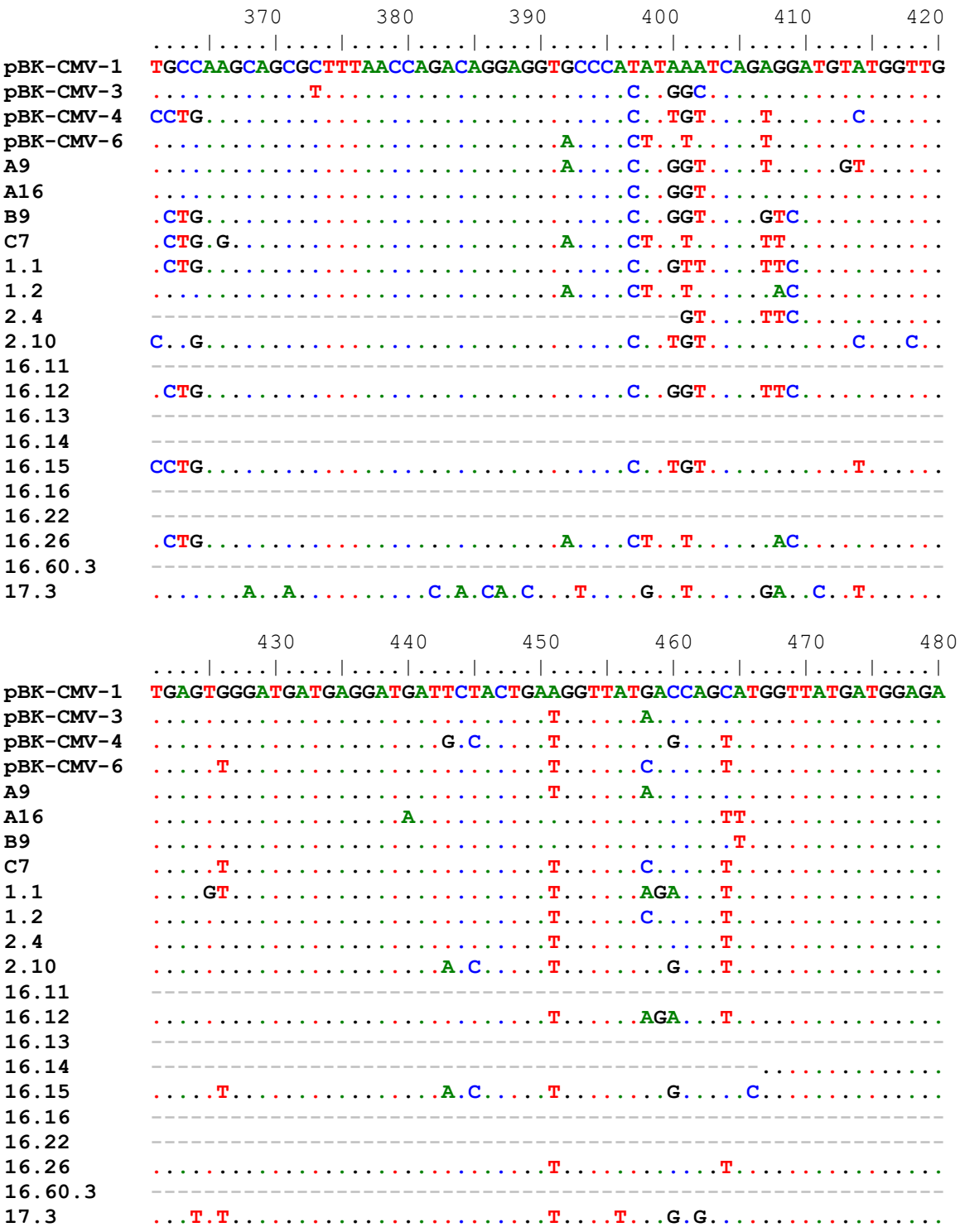
An alignment of the protein sequences was also generated using only the ORFs, and is shown in Figure 3.7. Included in this alignment were alleles from cod identified by Perrsons et al,1999 and Miller et al, 2002, and randomly selected rainbow trout MH Class I. As expected the haddock and cod alleles show significant similarity whereas the rainbow trout alleles show a very minimal number of conserved residues. There were several small groups of residues conserved between the species. These were at amino acid positions #104 – 108, #119 – 121, #136 – 141, #187 – 190, #214 – 224, #255 – 259 and #279 – 282. A potential glycosylation site (NXT/S) is present at position #106, which likely accounts for the conservation of residues in this area. The regions #136 –

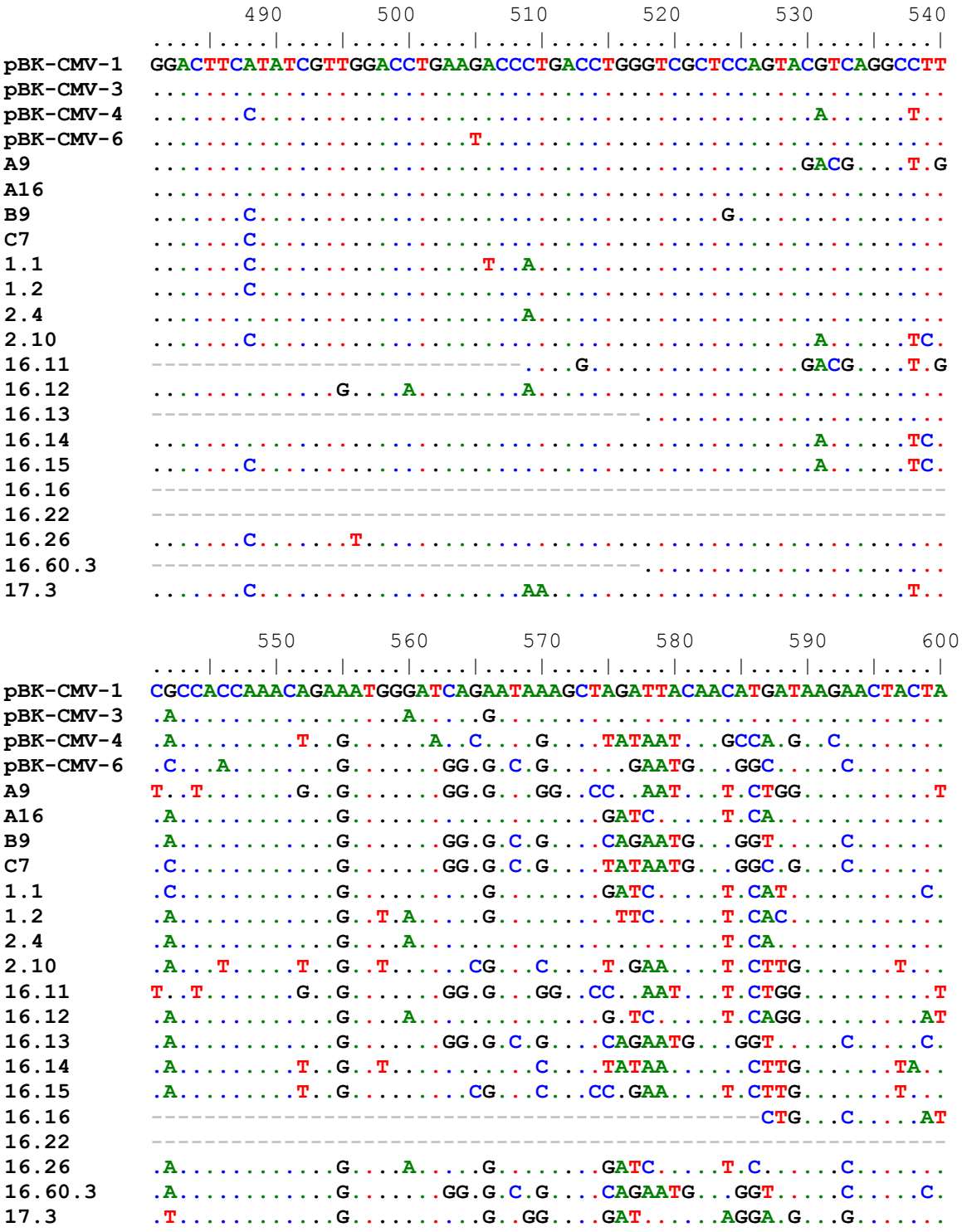
Figure 3.6: DNA sequence alignment of MH Class I sequences from isolated clones.

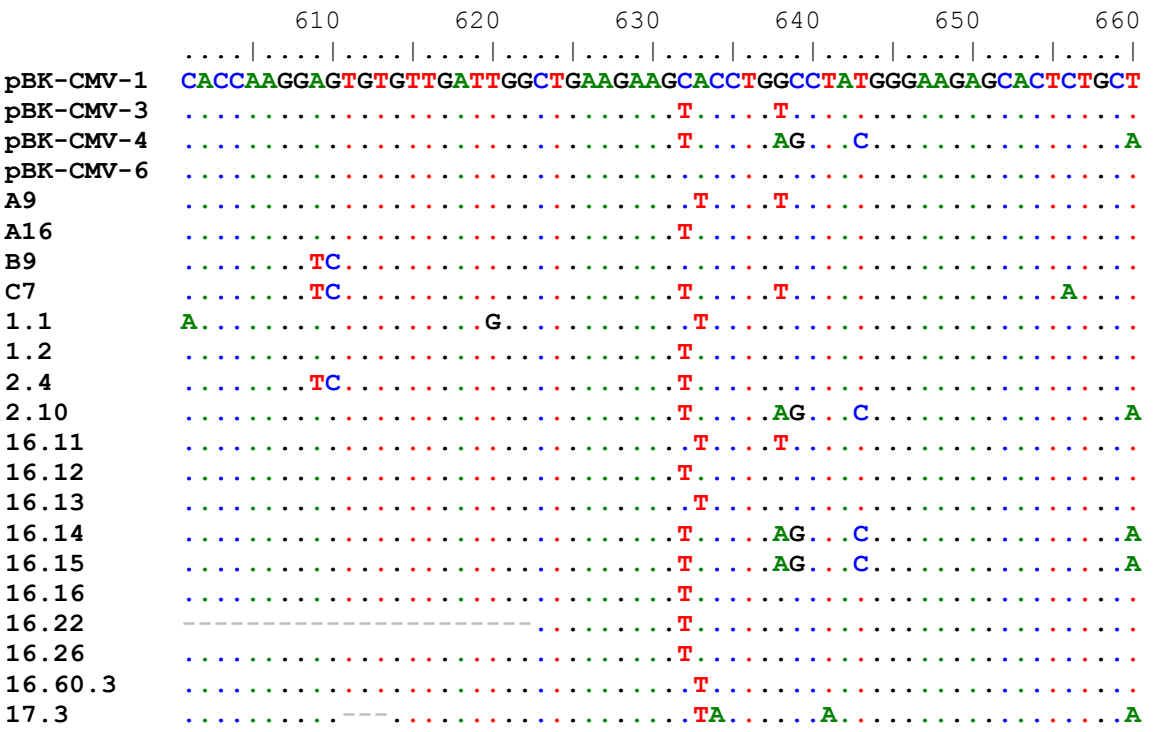

10 20 30 40 50 60
pBK-CMV-1 tggagtcggagtcccgtctctgtggataatctcatcgagctgttcttctgaatacatcga
pBK-CMV-3 cact.gtt....t.ttc.....
pBK-CMV-4 ---------g..tg.....t.....a......c..t.....
pBK-CMV-6 ----------g.....
A9 cacc.....t.....
A16 ----------.....
B9 .....
C7 ccat.cgctc.....
1.1 -----
1.2 gttgaag.....
2.4 -----
2.10 -----
16.11 -----
16.12 .....g.....
16.13 -----
16.14 -----
16.15 ----------g.....t.....t.....ct..t...c..t.....
16.16 -----
16.22 -----
16.26 -----
16.60.3 -----
17.3 c.....a.cctata.....t.....c..g.....
70 80 90 100 110 120
pBK-CMV-1 tATGAAGGCGCTAATAGGGCTGCTGCTGCTGTTGGTCTTTGGTCACGGTGTGTCCTCACT
pBK-CMV-3 .....G.
pBK-CMV-4 .....T..---.....A.....G.
pBK-CMV-6 .....G.
A9 .....
A16 .....
B9 .....GG
C7 .....
1.1 -----
1.2 .....G.
2.4 -----
2.10 -----.....A.....G.
16.11 -----
16.12 .....G.
16.13 -----
16.14 -----
16.15 .....T.....T..---...T.....A.....G.
16.16 -----
16.22 -----
16.26 -----
16.60.3 -----
17.3 .....T.....---.....T.....G.

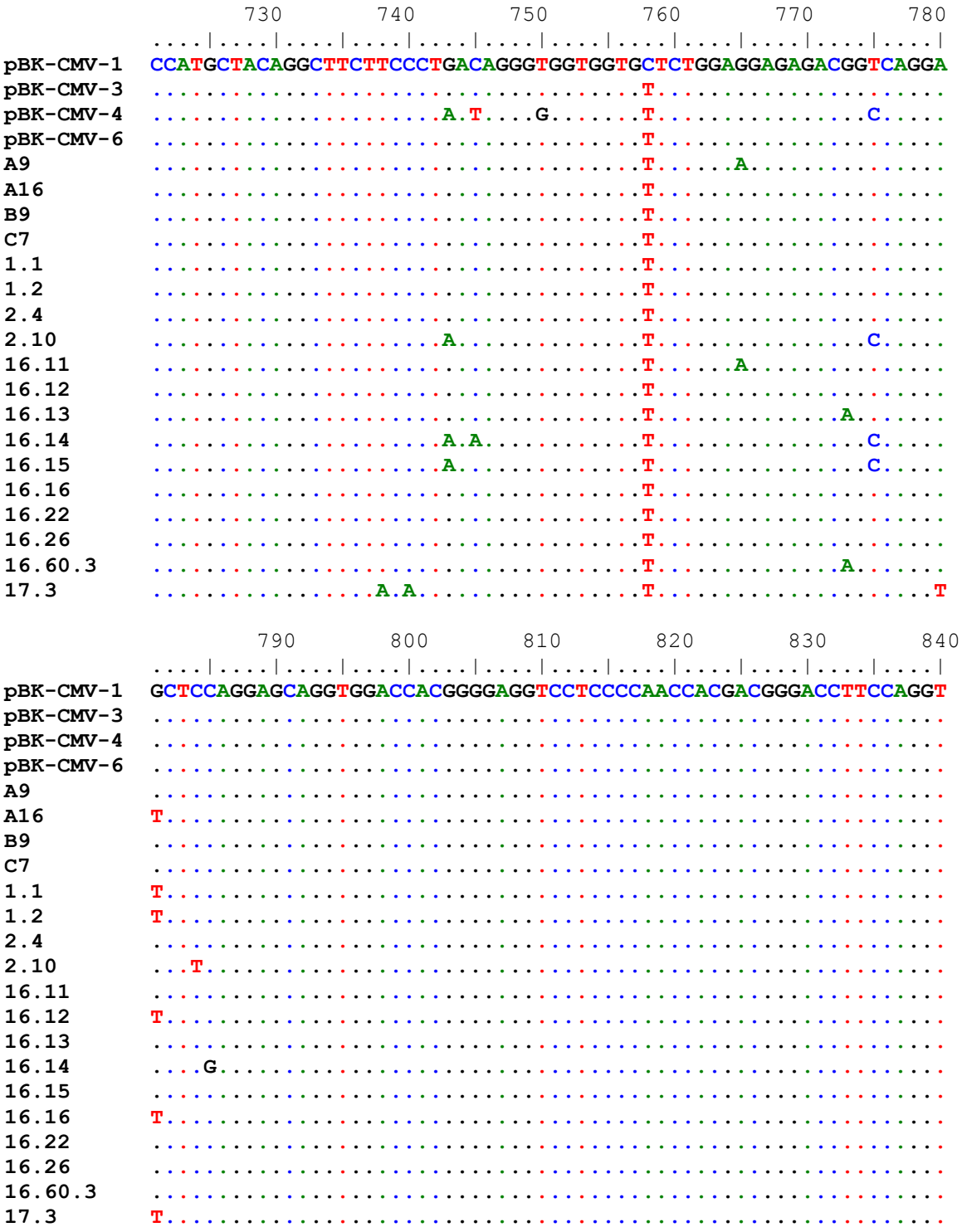

130 140 150 160 170 180
pBK-CMV-1 GCTTCACTCTCTCCATTATTTCTACACGGGGTCGTCTGGACTCTCAACCTTCCCAGAGTT
pBK-CMV-3 .........A.G..G.T..........C.........................
pBK-CMV-4 .A.........G..GAT.........TC...T.....................
pBK-CMV-6 .........A.G..G............C..............G..........
A9 A.........A.G..G.T..........C.....................
A16 ....C.....................
B9 ....C.....................
C7 .........A.G..G.....................
1.1 --
1.2 .........A.G..G............C.....................
2.4 --
2.10 .A.........G.GGAT.........TC...T.....T..............A
16.11 --
16.12 .........A.G..G.T..........C.....................
16.13 --
16.14 --
16.15 .A........G.G..GGT.....C.....C...................T.....
16.16 --
16.22 --
16.26 --
16.60.3 --
17.3 ..C.........G..G............C.....................
190 200 210 220 230 240
pBK-CMV-1 TGTTTCTGTGGGGATGGTGGATGGAGTTCAGTTTGACTACTATGACAGCAACACCCAGAA
pBK-CMV-3 ....G.....................A.....................
pBK-CMV-4 ...GAGG..T.A..........AG......AC...G..................G
pBK-CMV-6 ....GT..............A.......A.G.TTC..............GTA...G
A9 ....GT...T.....................A.G.TTC..............GTA...G
A16 ....GT...T.....................G
B9 ....A.....................G
C7 ....GT...T.....................A.G.TTC..................G
1.1 --G
1.2 ....GT.....................A.G.TTC..............GTA...G
2.4 --
2.10 ...GA.A..TCA..........AG......A....G..................G
16.11 --
16.12 ....GT...T.....................A.GCTTC..............GTA...G
16.13 --
16.14 --
16.15 ...GATG..TCA..........AG......G....G..................G
16.16 --
16.22 --
16.26 --
16.60.3 --
17.3 ....G.....................G.....C..............G

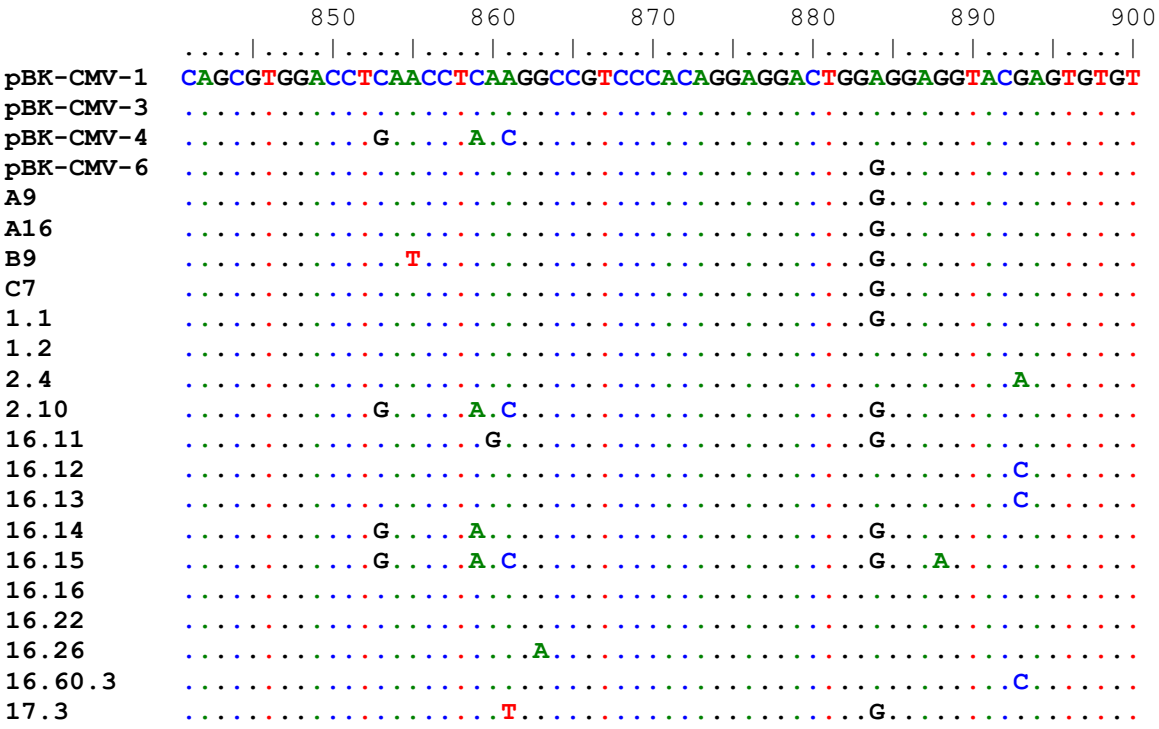

250 260 270 280 290 300
pBK-CMV-1 AACAGTCTTCAAACAGGACTGGATGGAGCAGGTCACCAGAGGTGAC---CCCGACTACCT
pBK-CMV-3 ..AG...C.....A....---....C......
pBK-CMV-4 ..TCA.AAC.....G...TC....C..A.....ACA.AGTC.....
pBK-CMV-6 .......C.....---....C......
A9 .G.C..ACC.....T...GC.GC...C...C..TA.....A.C.T----..G.GTG....
A16 .GT....C.....C.....---.....
B9 .GT....C.....---....C......
C7 .......C.....A.....A.....---....C......
1.1 .G.C..ACC.....T...AC.GC...C...T..TA.....A.C.T----..G.GTG....
1.2 GG.C..ACC.....T...GC.GC...C...A..TA.....A.C.T----..G.GTG.A..
2.4
2.10 ..TCA.AAC.....G...TC....C..A....TACA..ATC.....
16.11
16.12 .G.C..ACC.....T...GC.GC...C...C..TT.....A.C.T----..G.GTG....
16.13
16.14
16.15 ..TCA.AAC.....G....C....C..A.....ACA.AGTC.....
16.16
16.22
16.26 ....C......
16.60.3
17.3 ..T.T..C.....C.....A..T.....A....---..G.C....G.
310 320 330 340 350 360
pBK-CMV-1 GGAGAGGAACACTGGAAGTGCTCAGAATGCCCAGCAGGTCTTCAAAGTCAACATTGTGAC
pBK-CMV-3 .......G.A.....AAT.....GG.CA.....AG.....T
pBK-CMV-4 .......GCG....A...CAGAA..GT.AA.....G.....G...
pBK-CMV-6 .....ACTT..T.GG.....G.....G..T
A9 C...G..G.....G.ATT....GC..A.....C.....C.....G..T
A16 .A.....G.A.....CTT..T.GG.....AC.....G.....G..T
B9 .......G.A.....TCAGAA..GG.AA.....T.....GG.G...GA..
C7 .....ACTT....GG.....AC.....GG.....G...
1.1 C.....G.....ACTT..T.GG.T.....AC..A.....C.....G..T
1.2 C.....G.....G.....GG..A.....AC.....C.....G..T
2.4
2.10 .....TCA....A..ACAGAA..GG.CAA.....GC.....G...
16.11
16.12 C...G..G.....ACTT..T.GG.....A.....C.....G...
16.13
16.14
16.15 .....A.GCA....A..ACAGAA..GG..AA.....CG.....G...
16.16
16.22
16.26 .....C..TCT.....GG..A.....AC.....C.G.....G..A
16.60.3
17.3 ..T..C..CA...A....CTT..T.GT.....AG..C..T..C....C...G..A

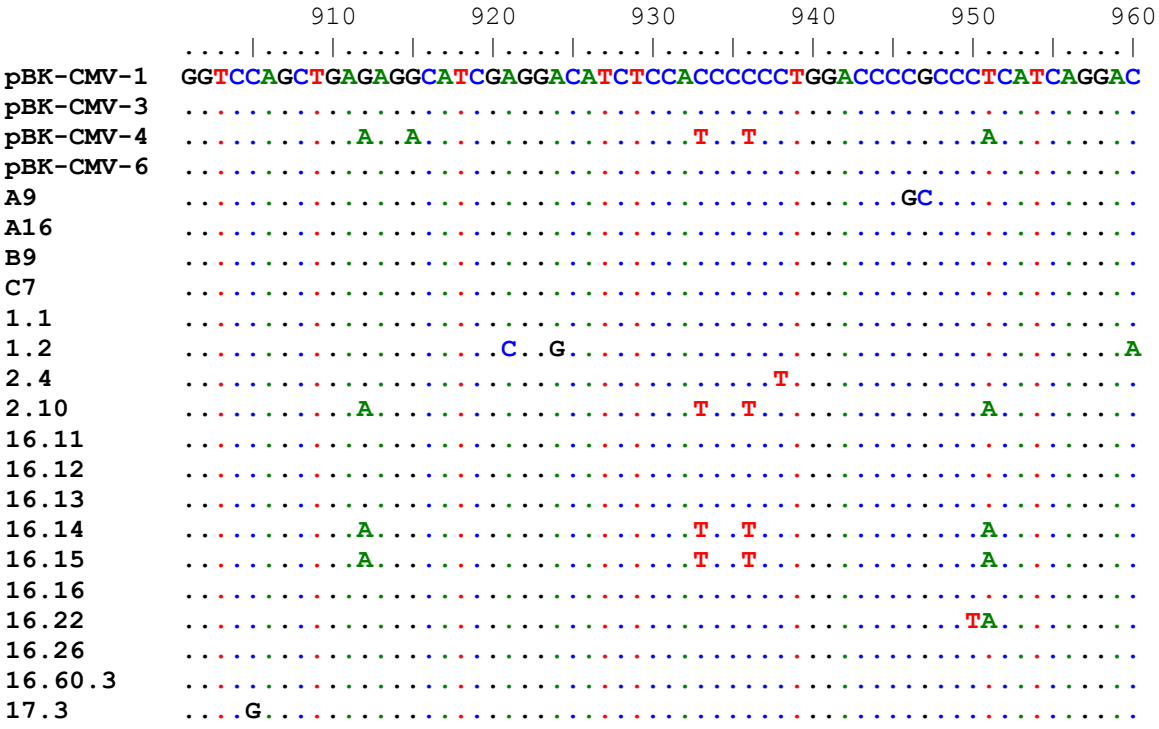

370 380 390 400 410 420
pBK-CMV-1 TGCCAAGCAGCGCTTTAACCAGACAGGAGGTGCCCATATAAATCAGAGGATGTATGGTTG
pBK-CMV-3 ..........T..........C..GGC..........
pBK-CMV-4 CCTG..........C..TGT....T......C......
pBK-CMV-6 ..........A....CT..T.....T..........
A9 ..........A....C..GGT....T.....GT......
A16 ..........C..GGT..........
B9 .CTG..........C..GGT....GTC..........
C7 .CTG.G..........A....CT..T.....TT..........
1.1 .CTG..........C..GTT....TTC..........
1.2 ..........A....CT..T......AC..........
2.4 ----------GT....TTC..........
2.10 C..G..........C..TGT..........C...C..
16.11 ----------
16.12 .CTG..........C..GGT....TTC..........
16.13 ----------
16.14 ----------
16.15 CCTG..........C..TGT..........T......
16.16 ----------
16.22 ----------
16.26 .CTG..........A....CT..T......AC..........
16.60.3 ----------
17.3 .......A..A..........C.A.CA.C...T....G..T.....GA..C..T......
430 440 450 460 470 480
pBK-CMV-1 TGAGTGGGATGATGAGGATGATTCTACTGAAGGTTATGACCAGCATGGTTATGATGGAGA
pBK-CMV-3 ..........T......A..........
pBK-CMV-4 ..........G.C.....T........G...T..........
pBK-CMV-6 .....T..........T......C.....T..........
A9 ..........T......A..........
A16 ..........A..........TT..........
B9 ....T..........
C7 .....T..........T......C.....T..........
1.1 ....GT..........T......AGA...T..........
1.2 ..........T......C.....T..........
2.4 ..........T..........T..........
2.10 ..........A.C.....T........G...T..........
16.11 ----------
16.12 ..........T......AGA...T..........
16.13 ----------
16.14 ----------..........
16.15 .....T..........A.C.....T........G.....C..........
16.16 ----------
16.22 ----------
16.26 ..........T..........T..........
16.60.3 ----------
17.3 ...T.T..........T....T...G.G..........

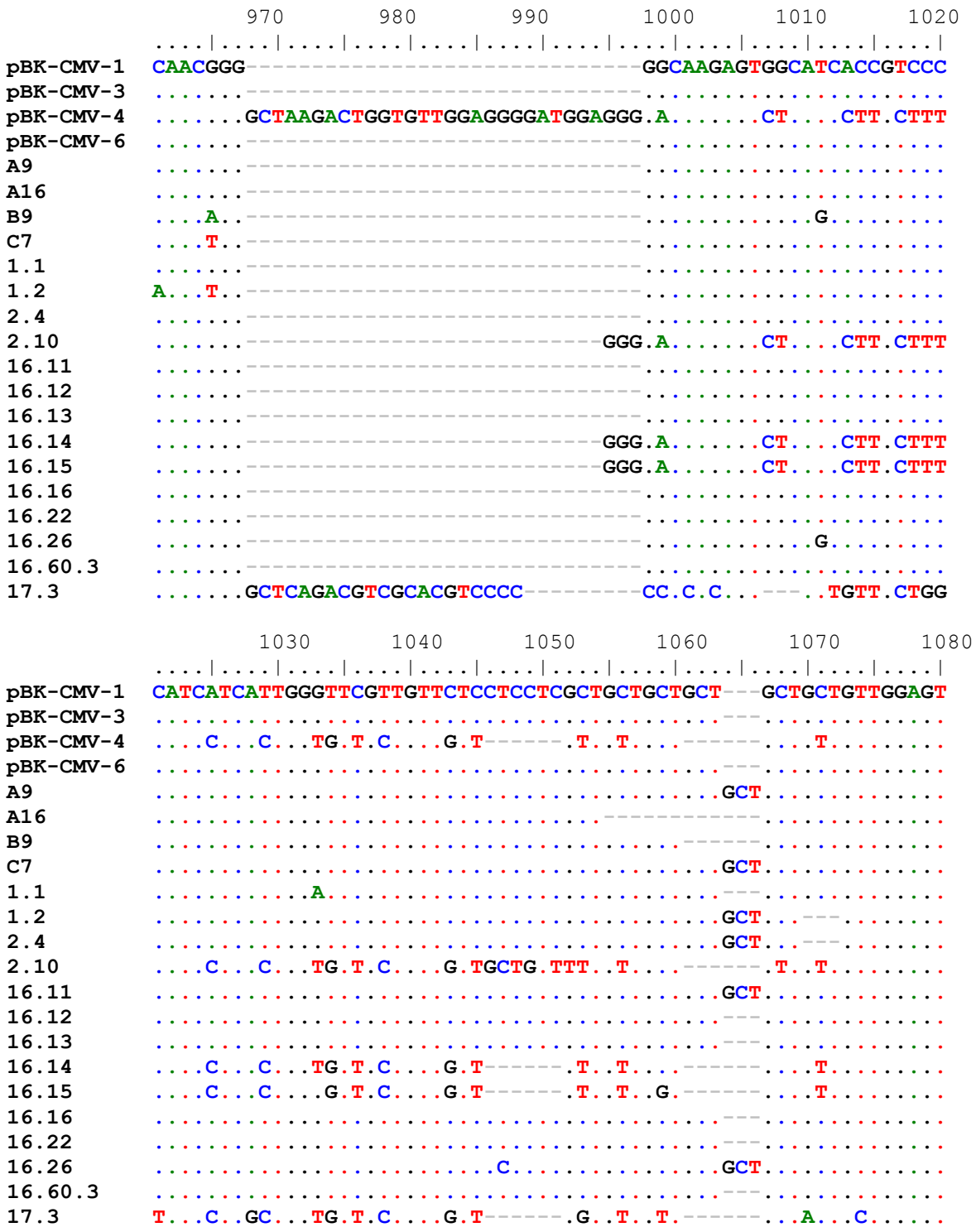

490 500 510 520 530 540
pBK-CMV-1 GGACTTCATATCGTTGGACCTGAAGACCCTGACCTGGGTCGCTCCAGTACGTCAGGCCTT
pBK-CMV-3
pBK-CMV-4
pBK-CMV-6
A9
A16
B9
C7
1.1
1.2
2.4
2.10
16.11
16.12
16.13
16.14
16.15
16.16
16.22
16.26
16.60.3
17.3
550 560 570 580 590 600
pBK-CMV-1 CGCCACCAAACAGAAATGGGATCAGAATAAAGCTAGATTACAACATGATAAGAACTACTA
pBK-CMV-3
pBK-CMV-4
pBK-CMV-6
A9
A16
B9
C7
1.1
1.2
2.4
2.10
16.11
16.12
16.13
16.14
16.15
16.16
16.22
16.26
16.60.3
17.3

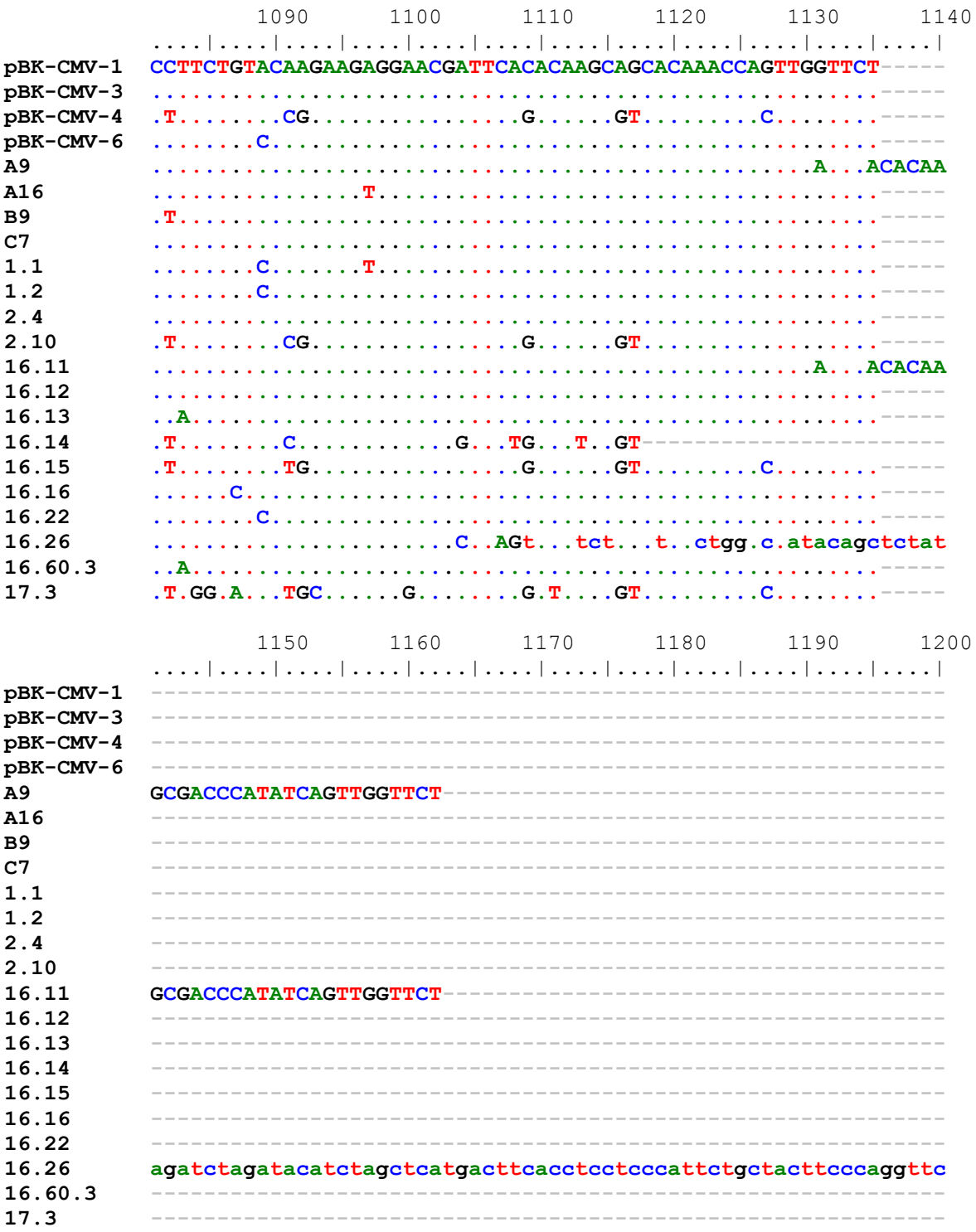

610 620 630 640 650 660
pBK-CMV-1 CACCAAGGAGTGTGTTGATTGGCTGAAGAAGCACCTGGCCTATGGGAAGAGCACTCTGCT
pBK-CMV-3 .....T.....T.....
pBK-CMV-4 .....T.....AG...C.....A
pBK-CMV-6 .....
A9 .....T....T.....
A16 .....T.....
B9 .....TC.....
C7 .....TC.....T.....T.....A....
1.1 A.....G.....T.....
1.2 .....T.....
2.4 .....TC.....T.....
2.10 .....T.....AG...C.....A
16.11 .....T....T.....
16.12 .....T.....
16.13 .....T.....
16.14 .....T.....AG...C.....A
16.15 .....T.....AG...C.....A
16.16 .....T.....
16.22 ----------.....T.....
16.26 .....T.....
16.60.3 .....T.....
17.3 .....---.....TA.....A.....A
670 680 690 700 710 720
pBK-CMV-1 GAGAACAGAGCGTCCGCGGGTGTCTCTGCTCCAGAGGAGCCCCTCCTCCCCAGTGGTGTG
pBK-CMV-3 .....
pBK-CMV-4 .....T.....
pBK-CMV-6 .....T.....
A9 .....T.....
A16 .....
B9 .....T.....
C7 .....
1.1 .....T.....
1.2 .....
2.4 .....T.....
2.10 .....
16.11 .....T.....
16.12 .....T.....
16.13 .....
16.14 .....
16.15 .....
16.16 .....
16.22 .....T.....
16.26 .....
16.60.3 .....
17.3 .---G....C.....

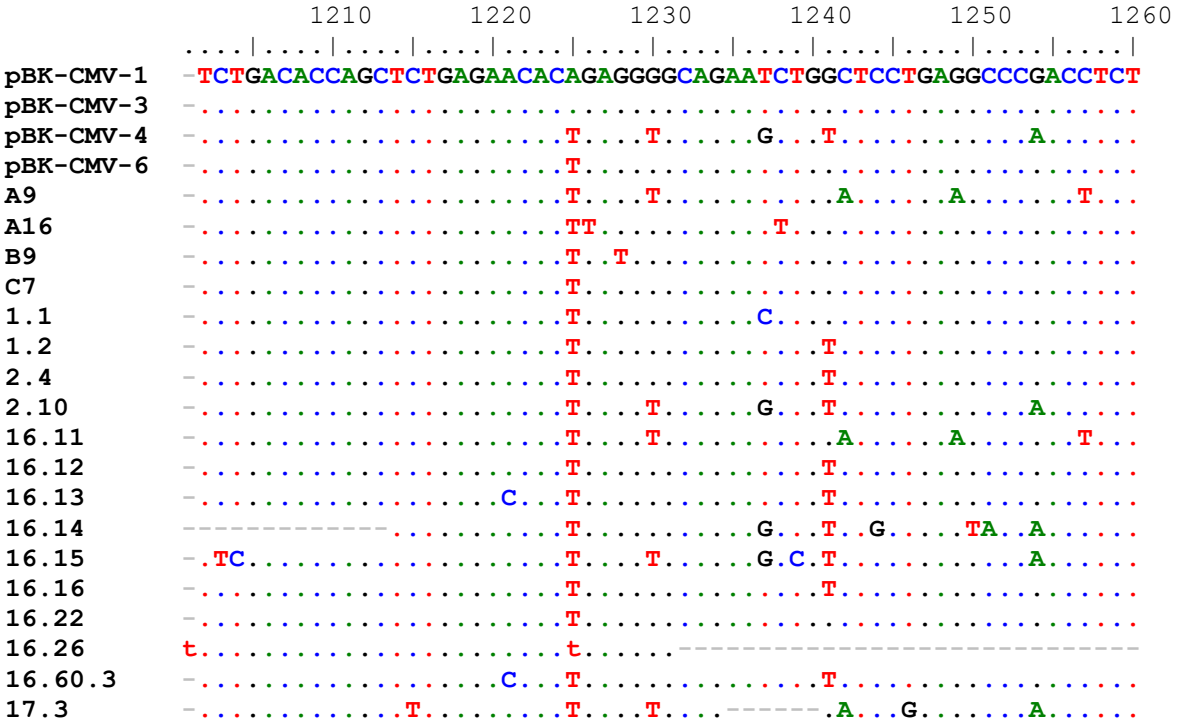

730 740 750 760 770 780
pBK-CMV-1 CCATGCTACAGGCTTCTTCCCTGACAGGGTGGTGGTGCTCTGGAGGAGAGACGGTCAGGA
pBK-CMV-3
pBK-CMV-4
pBK-CMV-6
A9
A16
B9
C7
1.1
1.2
2.4
2.10
16.11
16.12
16.13
16.14
16.15
16.16
16.22
16.26
16.60.3
17.3
790 800 810 820 830 840
pBK-CMV-1 GCTCCAGGAGCAGGTGGACCACGGGGAGGTCCTCCCCAACCACGACGGGACCTTCCAGGT
pBK-CMV-3
pBK-CMV-4
pBK-CMV-6
A9
A16
B9
C7
1.1
1.2
2.4
2.10
16.11
16.12
16.13
16.14
16.15
16.16
16.22
16.26
16.60.3
17.3

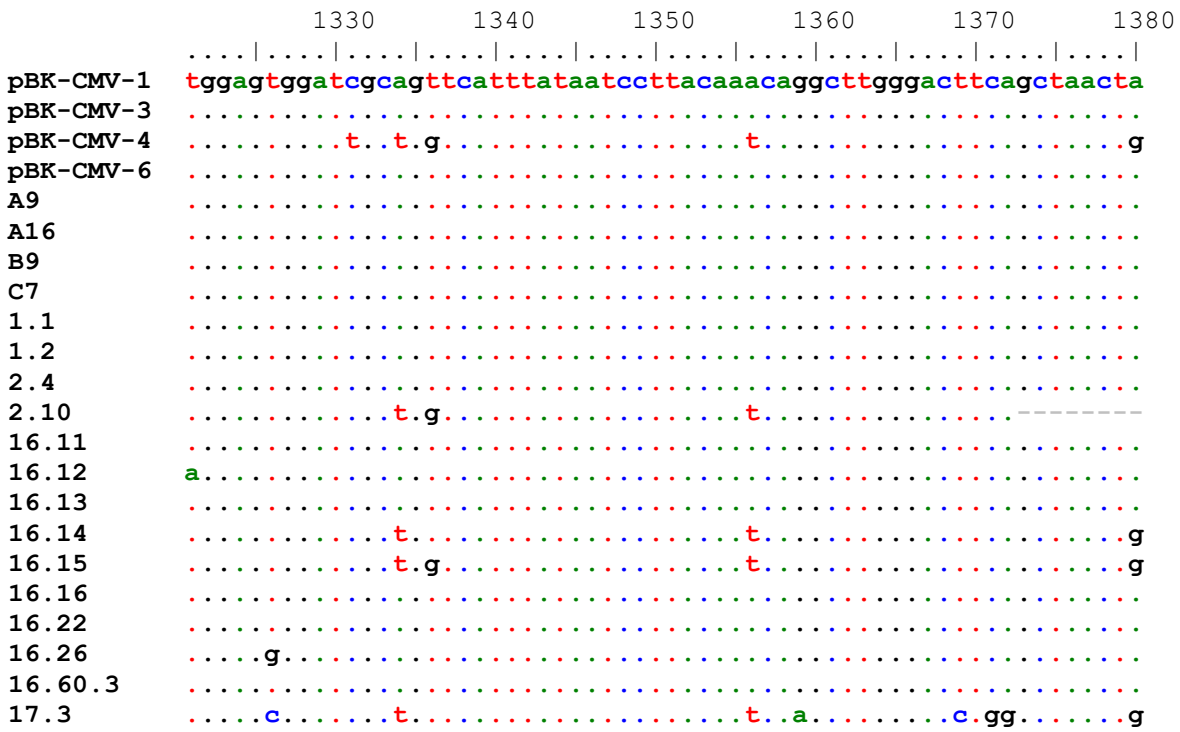

850 860 870 880 890 900
pBK-CMV-1 CAGCGTGGACCTCAACCTCAAGGCCGTCCCACAGGAGGACTGGAGGAGGTACGAGTGTGT
pBK-CMV-3
pBK-CMV-4
pBK-CMV-6
A9
A16
B9
C7
1.1
1.2
2.4
2.10
16.11
16.12
16.13
16.14
16.15
16.16
16.22
16.26
16.60.3
17.3

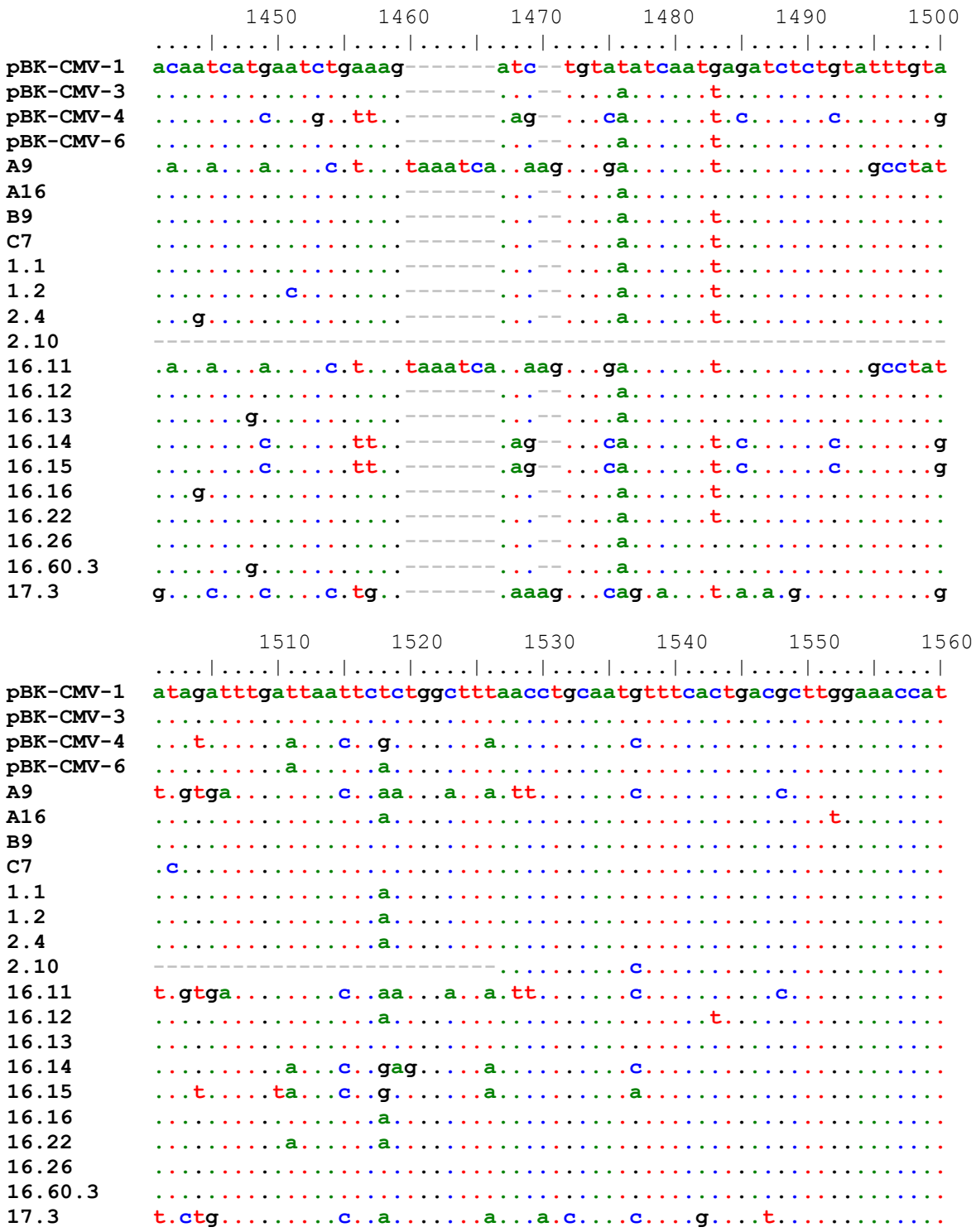

910 920 930 940 950 960
pBK-CMV-1 GGTCCAGCTGAGAGGCATCGAGGACATCTCCACCCCCCTGGACCCCGCCCTCATCAGGAC
pBK-CMV-3
pBK-CMV-4
pBK-CMV-6
A9
A16
B9
C7
1.1
1.2
2.4
2.10
16.11
16.12
16.13
16.14
16.15
16.16
16.22
16.26
16.60.3
17.3

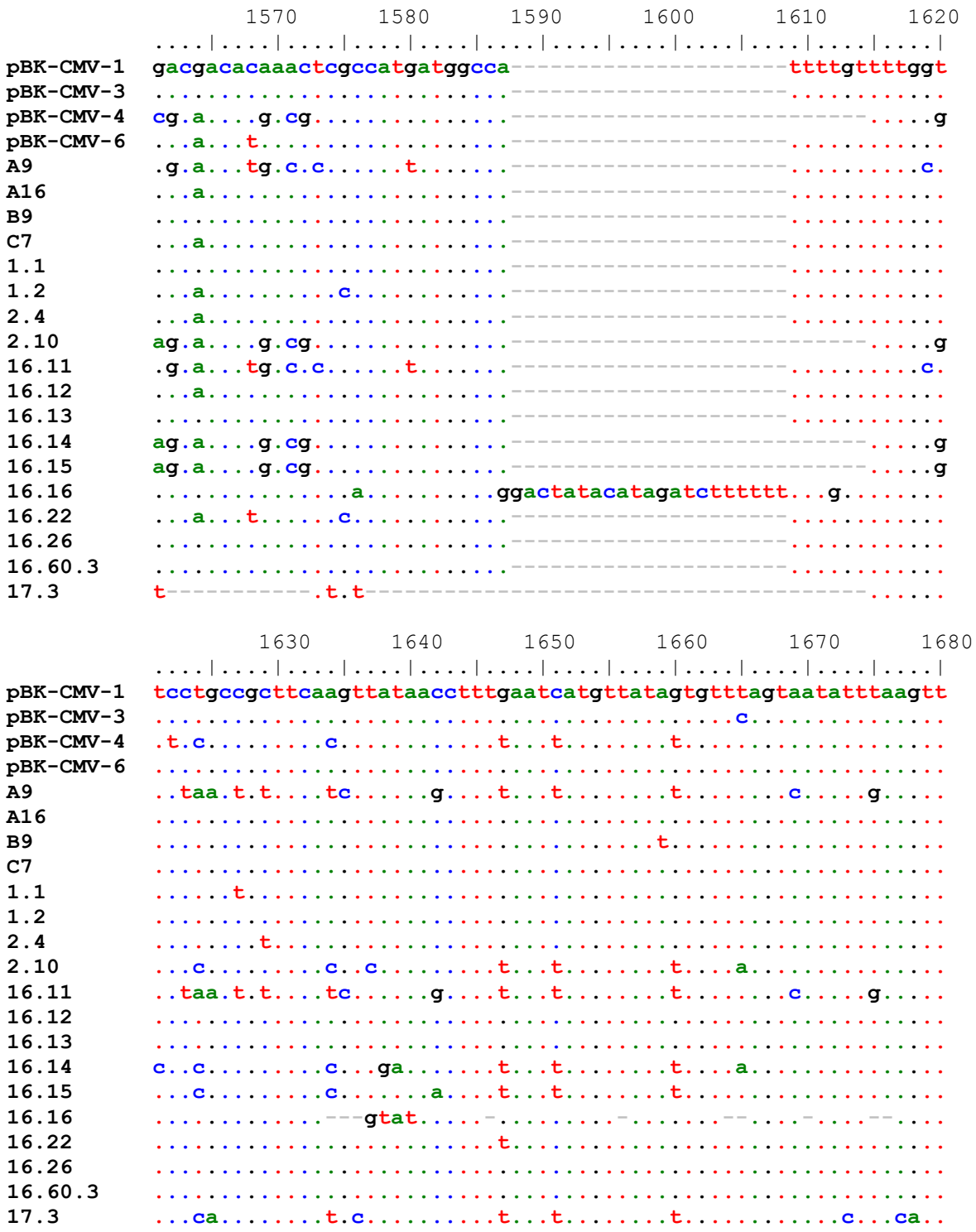

970 980 990 1000 1010 1020
pBK-CMV-1 CAACGGG----------GGCAAGAGTGGCATCACCGTCCC
pBK-CMV-3 ....----------.....................
pBK-CMV-4 .......GCTAAGACTGGTGTTGGAGGGGATGGAGGG.A.....CT....CTT.CTTT
pBK-CMV-6 ....----------.....................
A9 ....----------.....................
A16 ....----------.....................
B9 ....A..----------.............G.........
C7 ....T..----------.....................
1.1 ....----------.....................
1.2 A...T..----------.....................
2.4 ....----------.....................
2.10 .......----------GGG.A.....CT....CTT.CTTT
16.11 ....----------.....................
16.12 ....----------.....................
16.13 ....----------.....................
16.14 .......----------GGG.A.....CT....CTT.CTTT
16.15 .......----------GGG.A.....CT....CTT.CTTT
16.16 ....----------.....................
16.22 ....----------.....................
16.26 .......----------.............G.........
16.60.3 ....----------.....................
17.3 .......GCTCAGACGTCGCACGTCCCC---------CC.C.C...----..TGTT.CTGG
1030 1040 1050 1060 1070 1080
pBK-CMV-1 CATCATCATTGGGTTCGTTGTTCTCCTCCTCGCTGCTGCTGCT---GCTGCTGTTGGAGT
pBK-CMV-3 ....---..............
pBK-CMV-4 ....C...C...TG.T.C....G.T------.T..T....------....T.........
pBK-CMV-6 ....---..............
A9 ....GCT..............
A16 ....----------..............
B9 ....------..............
C7 ....GCT..............
1.1 ............A.....................................---..............
1.2 ....GCT...---.........
2.4 ....GCT...---.........
2.10 ....C...C...TG.T.C....G.TGCTG.TTT..T....------.T..T.........
16.11 ....GCT..............
16.12 ....---..............
16.13 ....---..............
16.14 ....C...C...TG.T.C....G.T------.T..T....------....T.........
16.15 ....C...C....G.T.C....G.T------.T..T..G.------....T.........
16.16 ....---..............
16.22 ....---..............
16.26 ........................C.............GCT..............
16.60.3 ....---..............
17.3 T...C..GC...TG.T.C....G.T------.G..T..T.------...A...C......

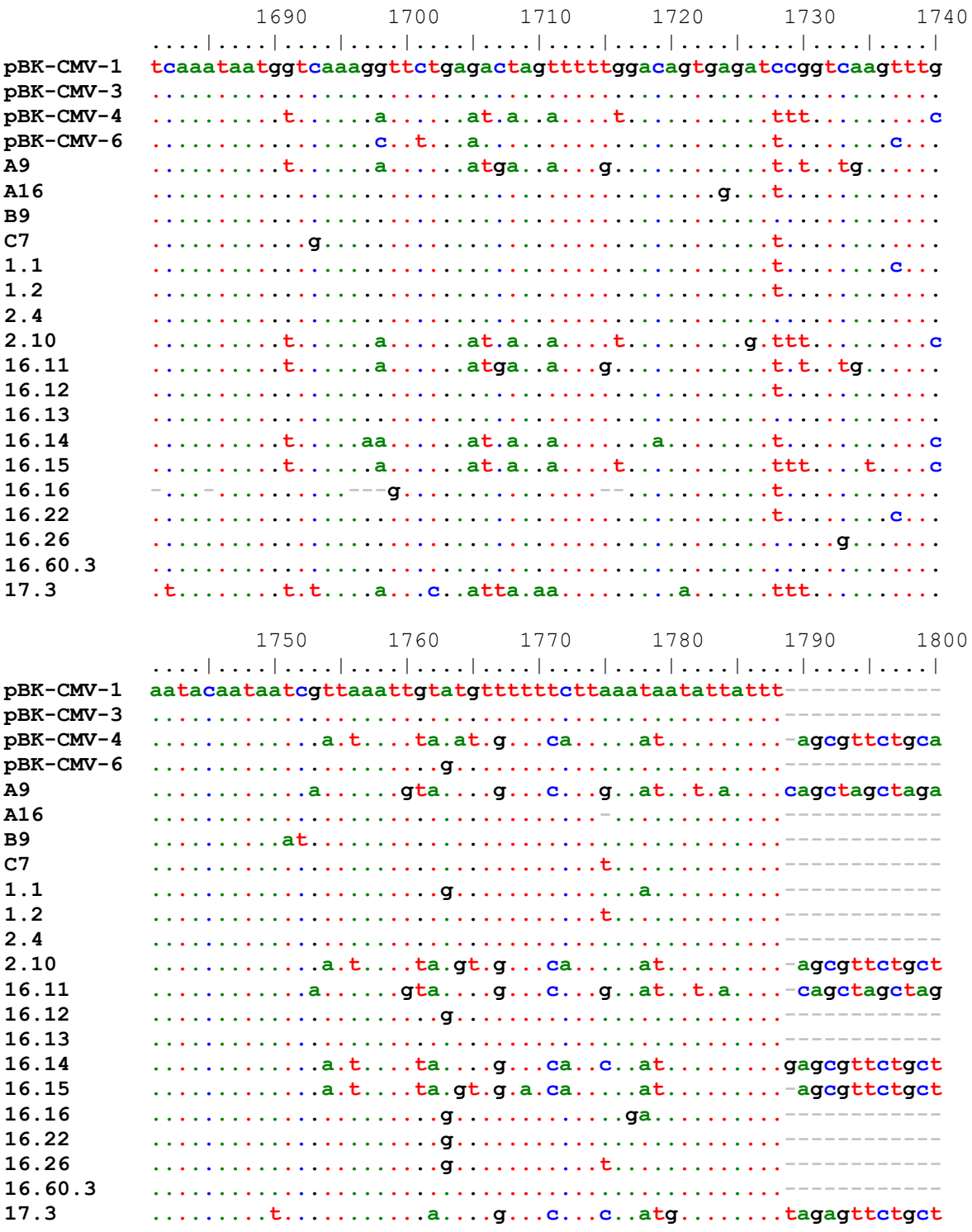

1090 1100 1110 1120 1130 1140
pBK-CMV-1 CCTTCTGTACAAGAAGAGGAACGATTCACACAAGCAGCACAAACCAGTTGGTTCT-----
pBK-CMV-3 .....-----
pBK-CMV-4 .T.....CG.....G.....GT.....C.....-----
pBK-CMV-6 .....C.....-----
A9 .....A...ACACAA
A16 .....T.....-----
B9 .T.....-----
C7 .....-----
1.1 .....C.....T.....-----
1.2 .....C.....-----
2.4 .....-----
2.10 .T.....CG.....G.....GT.....-----
16.11 .....A...ACACAA
16.12 .....-----
16.13 ..A.....-----
16.14 .T.....C.....G...TG...T..GT----------
16.15 .T.....TG.....G.....GT.....C.....-----
16.16 .....C.....-----
16.22 .....C.....-----
16.26 .....C..AGt...tct...t..ctgg.c.atacagctctat
16.60.3 ..A.....-----
17.3 .T.GG.A...TGC.....G.....G.T....GT.....C.....-----
1150 1160 1170 1180 1190 1200
pBK-CMV-1 ----------
pBK-CMV-3 ----------
pBK-CMV-4 ----------
pBK-CMV-6 ----------
A9 GCGACCCATATCAGTTGGTTCT----------
A16 ----------
B9 ----------
C7 ----------
1.1 ----------
1.2 ----------
2.4 ----------
2.10 ----------
16.11 GCGACCCATATCAGTTGGTTCT----------
16.12 ----------
16.13 ----------
16.14 ----------
16.15 ----------
16.16 ----------
16.22 ----------
16.26 agatctagatacatctagctcatgacttcacctcctcccattctgctacttcccaggttc
16.60.3 ----------
17.3 ----------


1210 1220 1230 1240 1250 1260
pBK-CMV-1 -TCTGACACCAGCTCTGAGAACACAGAGGGGCAGAATCTGGCTCCTGAGGCCCGACCTCT
pBK-CMV-3 -
pBK-CMV-4 -T....T.....G...T.....A.....
pBK-CMV-6 -T.....
A9 -T....T.....A.....A.....T...
A16 -TT.....T.....
B9 -T..T.....
C7 -T.....
1.1 -T.....C.....
1.2 -T.....T.....
2.4 -T.....T.....
2.10 -T....T.....G...T.....A.....
16.11 -T....T.....A.....A.....T...
16.12 -T.....T.....
16.13 -.....C...T.....T.....
16.14 ----------.....T.....G...T..G.....TA..A.....
16.15 -.TC.....T....T.....G.C.T.....A.....
16.16 -T.....T.....
16.22 -T.....
16.26 t.....t.....----------
16.60.3 -.....C...T.....T.....
17.3 -.....T.....T....T....------.A...G.....A.....
1270 1280 1290 1300 1310 1320
pBK-CMV-1 GACAGAAGTTCGAAGTTAAatcatgaagagtcttctcacgttcagcttcctaaagcaggg
pBK-CMV-3 .....
pBK-CMV-4 ...CAC..G..A.....g...g.....g.....c....a
pBK-CMV-6 ....A.....T.....
A9 .....G.....c.....
A16 .....A.....
B9 .....TT.....
C7 ....A.....A.....---.....
1.1 ....A.....T.....
1.2 ....A.....A.....
2.4 .....A.....
2.10 ...CACT.G..A.....g...g.....g.....c....a
16.11 .....G.....c.....
16.12 .....
16.13 .....
16.14 ...C.C.---------.....g.....g.....c....a
16.15 ...CAC..G..A.....g...g.....g.....c....t
16.16 .....A.....
16.22 ....A.....T.G.....
16.26 --------...a.....
16.60.3 .....
17.3 ...CAC..G..A.....--.....g....g.....t.....cc.c.....


1330 1340 1350 1360 1370 1380
pBK-CMV-1 tggagtggatcgcagttcatttataatccttacaaaacaggcttgggacttcagctaacta
pBK-CMV-3 .....
pBK-CMV-4 .....t..t.g.....t.....g
pBK-CMV-6 .....
A9 .....
A16 .....
B9 .....
C7 .....
1.1 .....
1.2 .....
2.4 .....
2.10 .....t.g.....t.....-------
16.11 .....
16.12 a.....
16.13 .....
16.14 .....t.....t.....g
16.15 .....t.g.....t.....g
16.16 .....
16.22 .....
16.26 .....g.....
16.60.3 .....
17.3 .....c.....t.....t..a.....c.gg.....g
1390 1400 1410 1420 1430 1440
pBK-CMV-1 aagatgctacatttgctaaattacaataacatccatacttaactttattttg-----ata
pBK-CMV-3 .....-----...
pBK-CMV-4 c...gt..g.....t---.t..a.....g..-----...
pBK-CMV-6 .....t.....t.....-----.....-----...
A9 .....t.....t.....t.....g.-----.....-----...
A16 .....t.....g.....-----...
B9 .....-----...
C7 .....t.....-----...
1.1 .....t.....t.....-----...
1.2 .....a.....-----...
2.4 .....t.....-----...
2.10 ----------
16.11 .....t.....t.....t.....------.....-----...
16.12 .....t.....-----...
16.13 .....-----...
16.14 c...gt..g.....t---.t..a....g.....g..-----...
16.15 c...gt..g.....t---.t..a.....g..-----...
16.16 .....t.....t.....-----...
16.22 .....t....c..t.....ta.-------.....-----...
16.26 .....t.....a.....-----...
16.60.3 .....-----...
17.3 c...gt..g.....c.....t.....ccttg...


1450 1460 1470 1480 1490 1500
pBK-CMV-1 acaatcatgaatctgaaag-------atc--tgtatatcaatgagatctctgtatttgta
pBK-CMV-3 ..................-------...--....a......t..................
pBK-CMV-4 ....c...g..tt..-------.ag--...ca......t.c......c.......g
pBK-CMV-6 ..................-------...--....a......t..................
A9 .a..a...a....c.t...taaatca..aag...ga......t..........gcctat
A16 ..................-------...--....a.....................
B9 ..................-------...--....a......t..................
C7 ..................-------...--....a......t..................
1.1 ..................-------...--....a......t..................
1.2 ....c........-------...--....a......t..................
2.4 ...g..............-------...--....a......t..................
2.10 ----------
16.11 .a..a...a....c.t...taaatca..aag...ga......t..........gcctat
16.12 ..................-------...--....a.....................
16.13 ....g..........-------...--....a.....................
16.14 ....c......tt..-------.ag--...ca......t.c......c.......g
16.15 ....c......tt..-------.ag--...ca......t.c......c.......g
16.16 ...g..............-------...--....a......t..................
16.22 ..................-------...--....a......t..................
16.26 ..................-------...--....a.....................
16.60.3 ....g..........-------...--....a.....................
17.3 g...c...c....c.tg..-------.aaag...cag.a...t.a.a.g.........g
1510 1520 1530 1540 1550 1560
pBK-CMV-1 atagatttgattaattctctggctttaacctgcaatgtttcactgacgcttggaaaccat
pBK-CMV-3 ....
pBK-CMV-4 ...t......a...c..g.......a..........c..........
pBK-CMV-6 ....a......a..........
A9 t.gtga........c..aa...a..a.tt.......c..........c..........
A16 ....a.....................t........
B9 ....
C7 .c..........
1.1 ....a..........
1.2 ....a..........
2.4 ....a..........
2.10 ----------.........c..........
16.11 t.gtga........c..aa...a..a.tt.......c..........c..........
16.12 ....a.....................t..............
16.13 ....
16.14 ....a...c..gag.....a..........c..........
16.15 ...t.....ta...c..g.......a..........a..........
16.16 ....a..........
16.22 ....a......a..........
16.26 ....
16.60.3 ....
17.3 t.ctg.........c..a.......a...a.c....c....g....t..........


1570 1580 1590 1600 1610 1620
pBK-CMV-1 gacgacacaaactcgccatgatggcca----------ttttgttttggt
pBK-CMV-3 ....----------............
pBK-CMV-4 cg.a....g.cg..............----------.....g
pBK-CMV-6 ...a...t..................----------............
A9 .g.a...tg.c.c.....t.....----------.....c.
A16 ...a.....................----------............
B9 ....----------............
C7 ...a.....................----------............
1.1 ....----------............
1.2 ...a.........c...........----------............
2.4 ...a.....................----------............
2.10 ag.a....g.cg..............----------.....g
16.11 .g.a...tg.c.c.....t.....----------.....c.
16.12 ...a.....................----------............
16.13 ....----------............
16.14 ag.a....g.cg..............----------.....g
16.15 ag.a....g.cg..............----------.....g
16.16 ....a.........ggactatacatagatctttttt...g.....
16.22 ...a...t.....c...........----------............
16.26 ....----------............
16.60.3 ....----------............
17.3 t----------.t.t----------.....
1630 1640 1650 1660 1670 1680
pBK-CMV-1 tcctgccgcttcaagttataacctttgaatcatgttatagtgtttagtaatatttaagtt
pBK-CMV-3 ....c.............
pBK-CMV-4 .t.c.........c...........t...t.......t.................
pBK-CMV-6 ....
A9 ..taa.t.t....tc.....g....t...t.......t.......c.....g.....
A16 ....
B9 ....t...................
C7 ....
1.1 ......t.....................................
1.2 ....
2.4 ....t.....................................
2.10 ...c.........c..c.........t...t.......t...a.............
16.11 ..taa.t.t....tc.....g....t...t.......t.......c.....g.....
16.12 ....
16.13 ....
16.14 c..c.........c...ga.......t...t.......t....a.............
16.15 ...c.........c.......a....t...t.......t.................
16.16 ....---gtat.....-.........-.......--....-....--....
16.22 ....t.....................
16.26 ....
16.60.3 ....
17.3 ...ca.........t.c..........t...t.......t...........c...ca..


1690 1700 1710 1720 1730 1740
pBK-CMV-1 tcaaataatggtcaaaggttctgagactagtttttggacagtgagatccggtcaagtttg
pBK-CMV-3
pBK-CMV-4
pBK-CMV-6
A9
A16
B9
C7
1.1
1.2
2.4
2.10
16.11
16.12
16.13
16.14
16.15
16.16
16.22
16.26
16.60.3
17.3
1750 1760 1770 1780 1790 1800
pBK-CMV-1 aatacaataatcgttaaattgtatgtttttcttaaataatattattt
pBK-CMV-3
pBK-CMV-4
pBK-CMV-6
A9
A16
B9
C7
1.1
1.2
2.4
2.10
16.11
16.12
16.13
16.14
16.15
16.16
16.22
16.26
16.60.3
17.3


1810 1820 1830 1840 1850 1860
pBK-CMV-1 -------gaattgcatcgctaca---taggctaccacgatcgagctaggcagctttag-c
pBK-CMV-3 -------.....------.....---.....a.
pBK-CMV-4 agctaga.g.....t.....---.....a...ct.....c.....-.
pBK-CMV-6 -------.....---.....-.
A9 -------.....---.....ag..a....-----.t.....-.
A16 -------.....---.....a.....t.-.
B9 -------.....---.....-.
C7 -------.....---.....-.
1.1 -------.....---.....-.
1.2 -------.....---.....-.
2.4 -------.....---.....t.....c..-.
2.10 agctaga.....t..g..---.....a...ct.....c.....-.
16.11 a------.....---.....ag..a....-----.t.....-.
16.12 -------.....---.....-.
16.13 -------.....---.....a....c.....-.
16.14 agctaga.c.....t.....aca.....a....t.....t.....-.
16.15 agctaga.g.....t..g..---.....a....t.....t.....-.
16.16 -------.....---.....-.
16.22 -------.....---.....g.-.
16.26 -------.....---.....-.
16.60.3 -------.....---.....a.....-.
17.3 agctaga.c.....t.....g---.t.....ac.....t.....-.
1870 1880 1890 1900 1910 1920
pBK-CMV-1 cttaatccttttgttgtactatttctttgttctttggggaatttgatcatagatactt
pBK-CMV-3 .....-.....
pBK-CMV-4 ..c...t.....---...tg.....g.....t.....g...--agt.....
pBK-CMV-6 .....-.....t.....
A9 t.....gg.c..t....---.g.....g.....g.....t.....
A16 .....-.....t...g.....-
B9 .....-.....-
C7 .....-....g.....t.....
1.1 a.....-.....t.....
1.2 .....-....g.....
2.4 .....-.....
2.10 ..c...t.....---...tg.....g.....g...--agt.....
16.11 t.....gg.c..---...tg.g.....g.....g.....t.....
16.12 .....-.....t.....
16.13 .....-.....
16.14 ....g.g..c..---...tg.....g.....t.....g...--agt.....
16.15 ..c...t.....---...tg.....g....a.....g...--agt.....
16.16 .....-.c...-....g.....
16.22 .....-....g.....t.....
16.26 .....-....g.....
16.60.3 .....-.....
17.3 .g....g..c..---...tg.....g.....g.....a.t.....

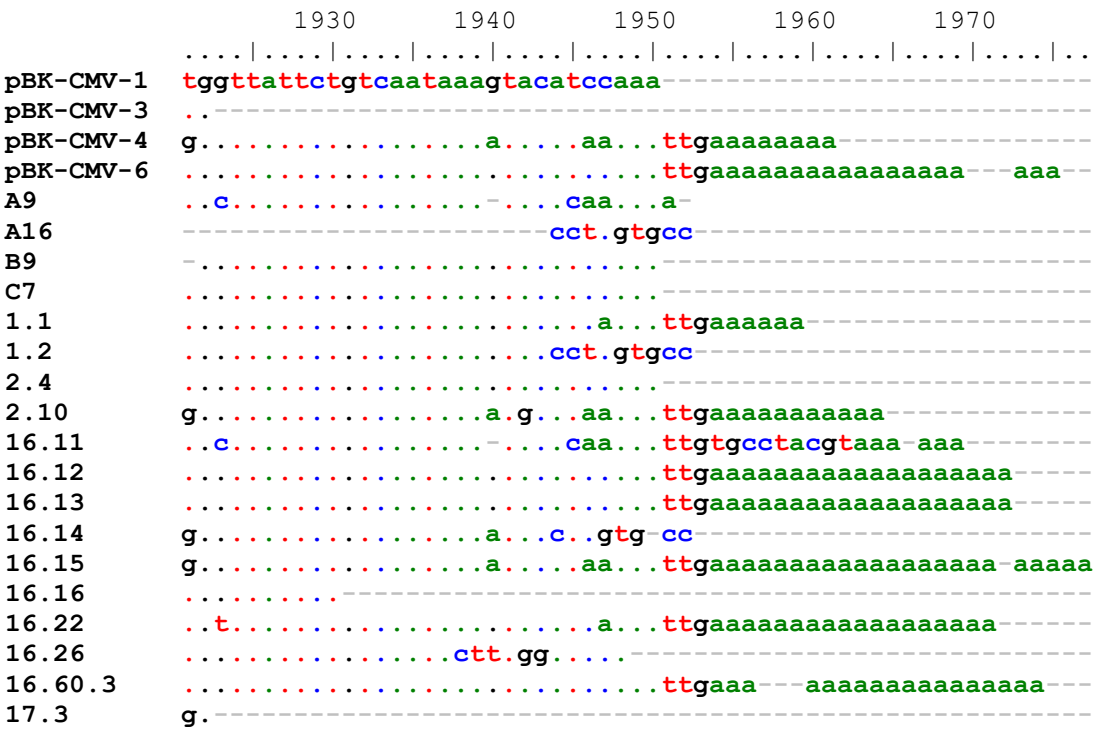

1930
1940
1950
1960
1970
pBK-CMV-1
pBK-CMV-3
pBK-CMV-4
pBK-CMV-6
A9
A16
B9
C7
1.1
1.2
2.4
2.10
16.11
16.12
16.13
16.14
16.15
16.16
16.22
16.26
16.60.3
17.3

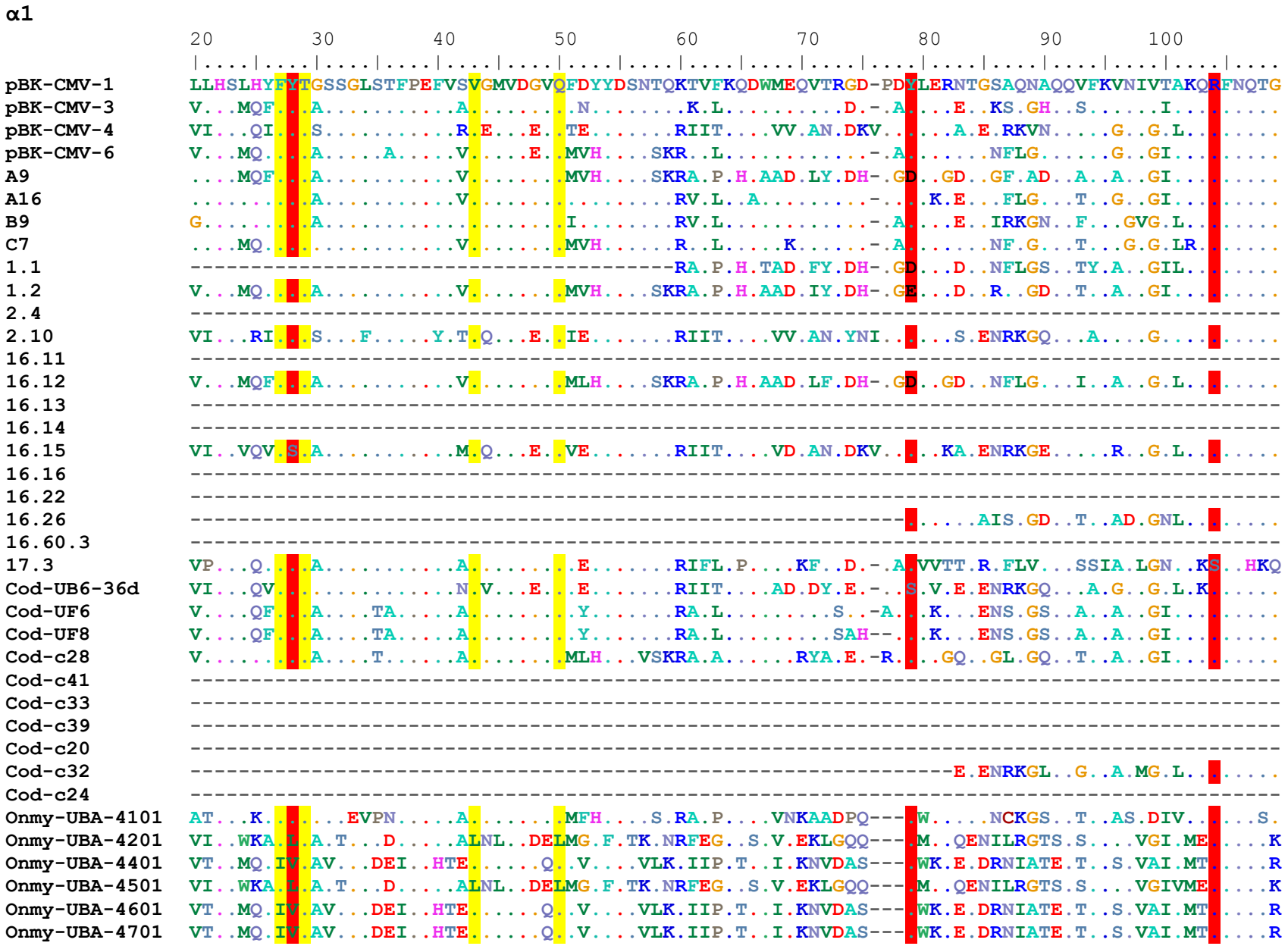
Table 3.1: Summary of the small nucleotide additions and deletions noted in the ORF of haddock MH Class I alleles.

Nucleotide Deletions							Nucleotide Additions	
#90–92	#612–614	#663–665	#1008–1010	#1047–1052	#1062-1064	#1236-1241	#288–290	#1065–1067
pBK-CMV-4	17.3	17.3	17.3	pBK-CMV-4	pBK-CMV-4	17.3	pBK-CMV-4	pBK-CMV-6
16.15				16.14	B9		2.10	C7
17.3				16.15	2.10		16.15	1.2
				17.3	16.14			2.4
					16.15			16.26
					17.3			

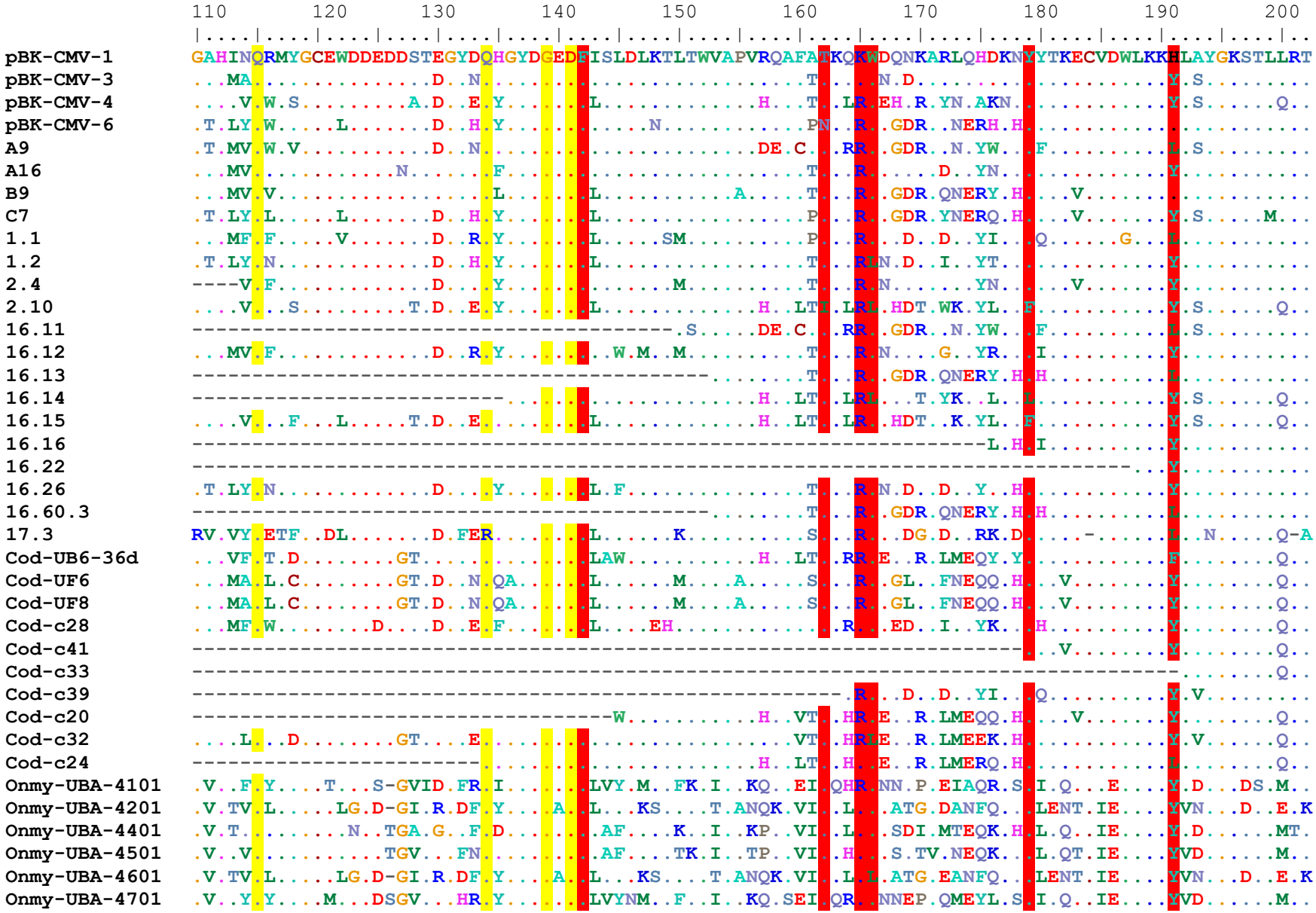
Figure 3.7: Amino acid alignment of haddock, cod and rainbow trout MH Class I allele sequences. Highlighted in yellow are the residues involved in binding β2-microglobulin. Highlighted in red are the residues essential for binding peptide.

Leader Peptide

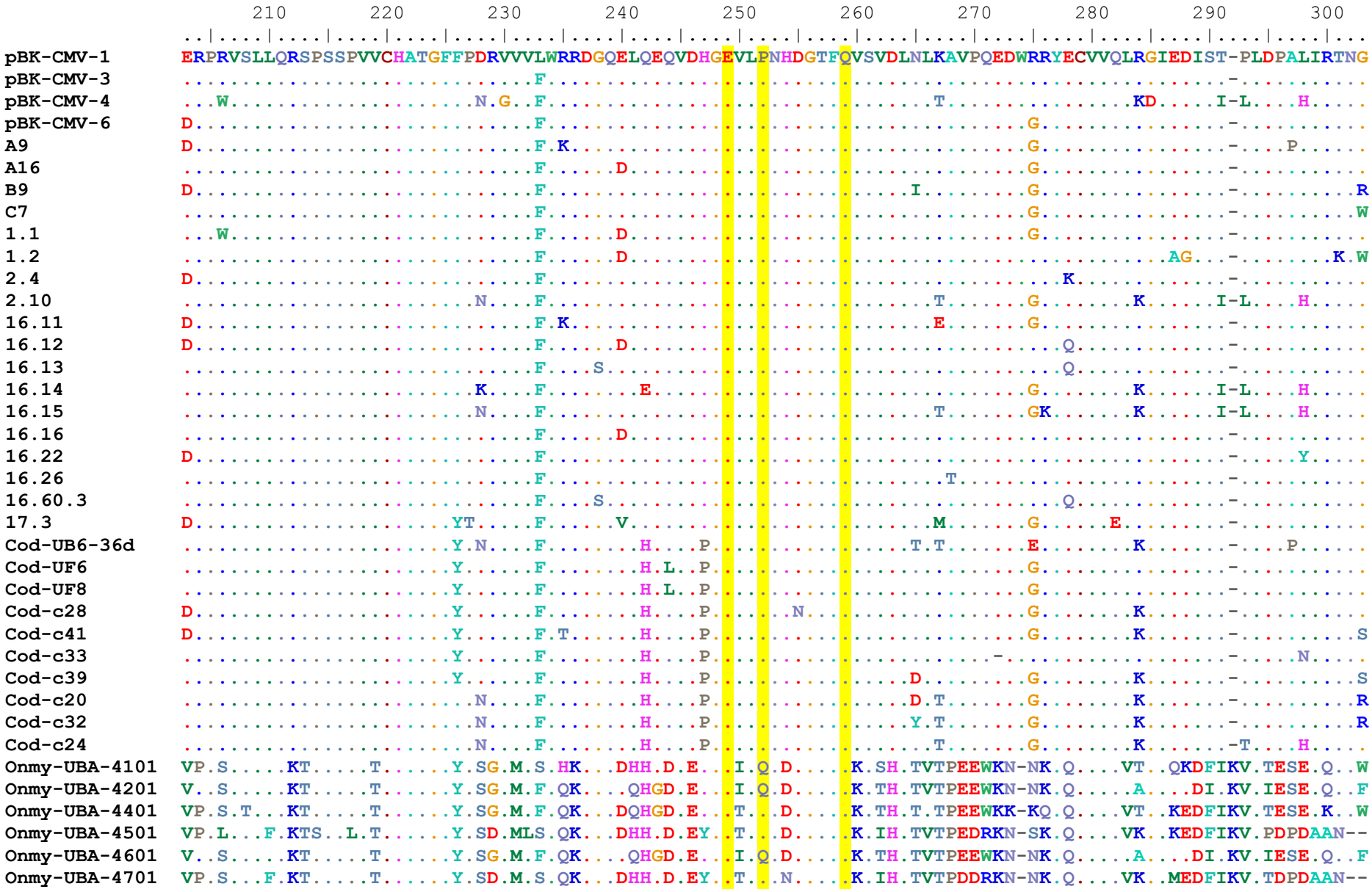
```
                        10
                ....|....|....|....
pBK-CMV-1       MKALIGLLLLLVFGHGVSS
pBK-CMV-3       ....................
pBK-CMV-4       .........-.....D...
pBK-CMV-6       ....................
A9              ....................
A16             ....................
B9              ....................
C7              ....................
1.1             -------------------
1.2             ....................
2.4             -------------------
2.10            ----------....D...
16.11           -------------------
16.12           ....................
16.13           -------------------
16.14           -------------------
16.15           ..V.....-.F..D...
16.16           -------------------
16.22           -------------------
16.26           -------------------
16.60.3         -------------------
17.3            .........-...C.....
Cod-UB6-36d     .........-.....D...
Cod-UF6         ..L.T....-.........
Cod-UF8         ..L.T....-......G..
Cod-c28         ....T....-.........
Cod-c41         -------------------
Cod-c33         -------------------
Cod-c39         -------------------
Cod-c20         -------------------
Cod-c32         -------------------
Cod-c24         -------------------
Onmy-UBA-4101   ..GI.L.V.GIGLL.TA.A
Onmy-UBA-4201   ..SF.L..G-IA-L.SS.A
Onmy-UBA-4401   .NG..L.VQGICLL.T..A
Onmy-UBA-4501   ..SF.L..G-IA-L.SS.A
Onmy-UBA-4601   ----------ICLL.T..A
Onmy-UBA-4701   .NG..L.VQGICLL.T..A
```

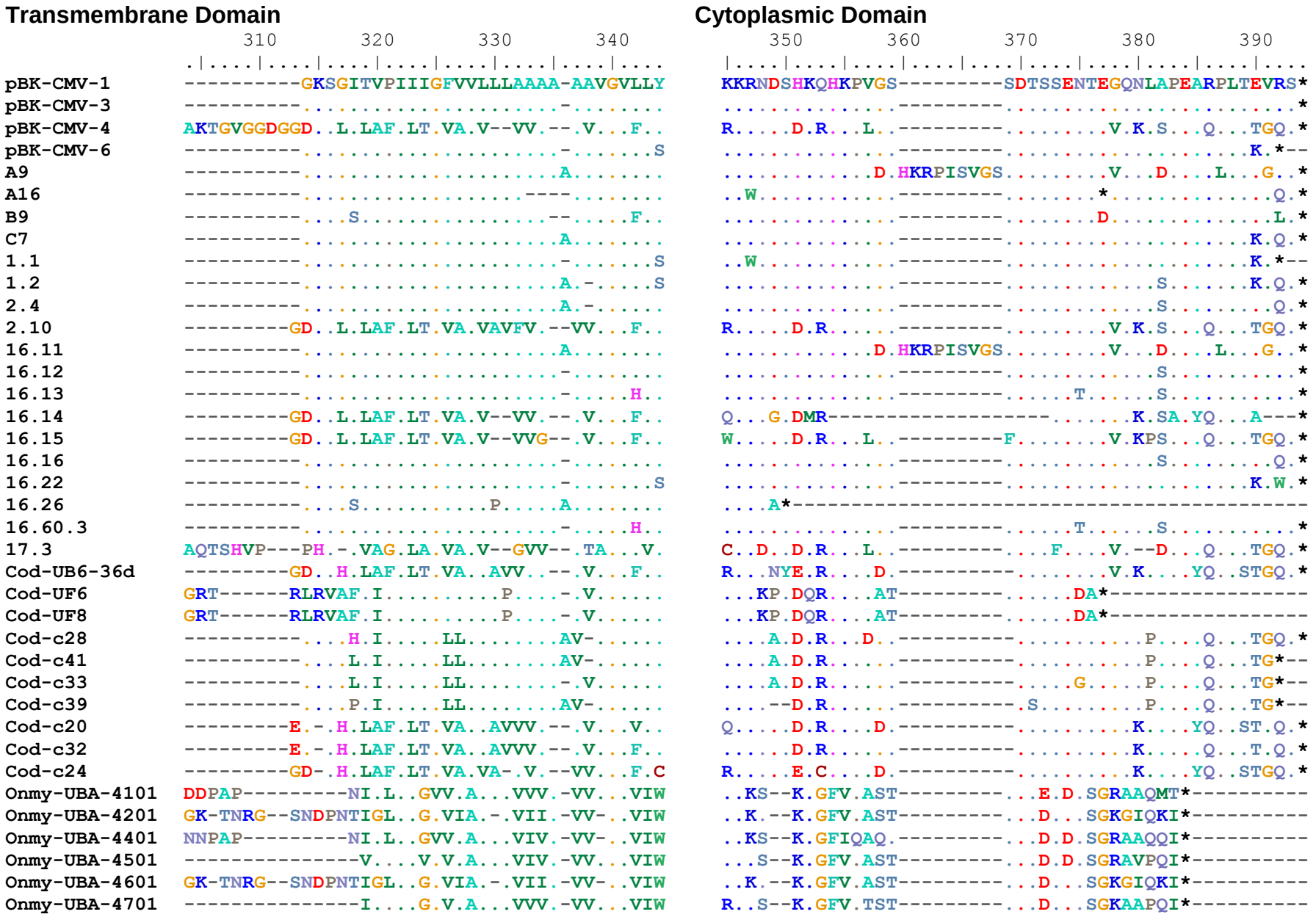

α1
20
30
40
50
60
70
80
90
100
pBK-CMV-1
pBK-CMV-3
pBK-CMV-4
pBK-CMV-6
A9
A16
B9
C7
1.1
1.2
2.4
2.10
16.11
16.12
16.13
16.14
16.15
16.16
16.22
16.26
16.60.3
17.3
Cod-UB6-36d
Cod-UF6
Cod-UF8
Cod-c28
Cod-c41
Cod-c33
Cod-c39
Cod-c20
Cod-c32
Cod-c24
Onmy-UBA-4101
Onmy-UBA-4201
Onmy-UBA-4401
Onmy-UBA-4501
Onmy-UBA-4601
Onmy-UBA-4701

α2


110 120 130 140 150 160 170 180 190 200
pBK-CMV-1 GAHINQRMYGCEWDDEDDSTEGYDQHGYDGEDDISLDLKTLTWVAPVRQAFATKQKWDQNKARLQHDKNYYTKECVDWLKKHLAYGKSTLLRT
pBK-CMV-3 ...MA.....D..N.....T....N.D.....Y.S.....
pBK-CMV-4 ...V.W.S.....A.D..E.Y.....L.....H...T..LR.EH.R.YN.AKN.....Y.S.....Q..
pBK-CMV-6 .T.LY.W.....L.....D..H.Y.....N.....PN..R..GDR..NERH.H.....
A9 .T.MV.W.V.....D..N.....DE.C...RR..GDR..N.YW...F.....L.S.....
A16 ...MV.....N.....F.....T...R.....D..YN.....Y.....
B9 ...MV.V.....L.....L.....A.....T...R..GDR.QNERY.H....V.....
C7 .T.LY.L.....L.....D..H.Y.....L.....P...R..GDR.YNERQ.H....V.....Y.S.....M...
1.1 ...MF.F.....V.....D..R.Y.....L.....SM.....P...R...D..D..YI...Q.....G...L.....
1.2 .T.LY.N.....D..H.Y.....L.....T...RLN.D..I..YT.....Y.....
2.4 ----V.F.....D....Y.....M.....T...R.N.....YN.....V.....Y.....
2.10 ....V...S.....T.D..E.Y.....L.....H..LTT.LRL.HDT.WK.YL..F.....Y.S.....Q..
16.11 ----------.S.....DE.C...RR..GDR..N.YW...F.....L.S.....
16.12 ...MV.F.....D..R.Y.....W.M..M.....T...R.N....G..YR...I.....Y.....
16.13 ----------.....T...R..GDR.QNERY.H.H.....L.....
16.14 ----------.....H..LT..LRL...T.YK..L..L.....Y.S.....Q..
16.15 ....V...F...L.....T.D..E.....L.....H..LT..LR..HDT..K.YL..F.....Y.S.....Q..
16.16 ----------L.H.I.....Y.....
16.22 ----------...Y.....
16.26 .T.LY.N.....D....Y.....L.F.....T...R.N.D..D..Y..H.....Y.....
16.60.3 ----------.....T...R..GDR.QNERY.H.H.....L.....
17.3 RV.VY.ETF..DL.....D.FER.....L.....K.....S...R...DG.D..RK.D.....-.....L..N.....Q-A
Cod-UB6-36d ...VF.T.D.....GT.....LAW.....H..LT..RR.E..R.LMEQY.Y.....F.....Q..
Cod-UF6 ...MA.L.C.....GT.D..N.QA.....L.....M....A.....S...R..GL..FNEQQ.H...V.....Y.....Q..
Cod-UF8 ...MA.L.C.....GT.D..N.QA.....L.....M....A.....S...R..GL..FNEQQ.H...V.....Y.....Q..
Cod-c28 ...MF.W.....D....D..E.F.....L....EH.....R...ED..I..YK...H.....Y.....Q..
Cod-c41 ----------...V.....Y.....Q..
Cod-c33 ----------.....Q..
Cod-c39 ----------.R...D..D..YI...Q.....Y.V.....
Cod-c20 ----------W.....H..VT..HR.E..R.LMEQQ.H....V.....Y.....Q..
Cod-c32 ....L...D.....GT....E.....VT..HRLE..R.LMEEK.H.....Y.V.....Q..
Cod-c24 ----------.....H..LT..H..E..R.LMERQ.H.....L.....Q..
Onmy-UBA-4101 .V..F.Y....T...S-GVID.FR.I.....LVY.M..FK.I..KQ..EI.QHR.NN.P.EIAQR.S.I.Q...E....Y.D...DS.M..
Onmy-UBA-4201 .V.TV.L.....LG.D-GI.R.DF.Y....A..L...KS....T.ANQK.VI..L...ATG.DANFQ...LENT.IE....YVN...D..E.K
Onmy-UBA-4401 .V.T.....N..TGA.G..F.D.....AF....K..I..KP..VI..L...SDI.MTEQK.H.L.Q..IE....Y.D.....MT.
Onmy-UBA-4501 .V..V.....TGV...FN.....AF....TK.I..TP..VI..H...S.TV.NEQK...L.QT.IE....YVD.....M..
Onmy-UBA-4601 .V.TV.L.....LG.D-GI.R.DF.Y....A..L...KS....T.ANQK.VI..L.L.ATG.EANFQ...LENT.IE....YVN...D..E.K
Onmy-UBA-4701 .V..Y.Y....M...DSGV...HR.Y.....LVYNM..F..I..KQ.SEI.QR...NNEP.QMEYL.S.I.Q..IE....YVD.....M..

α3


210 220 230 240 250 260 270 280 290 300
pBK-CMV-1 ERPRVSLLQRSPSSPVVCHATGFFPDRVVVLWRRDGQELQEQVDHGEVLPNHDGTFQVSVDLNLKAVPQEDWRRYECVVQLRGIEDIST-PLDPALIRTNG
pBK-CMV-3 ..........F.....-.....
pBK-CMV-4 ...W.....N.G..F.....T.....KD.....I-L...H.....
pBK-CMV-6 D.....F.....G.....-.....
A9 D.....F.K.....G.....-....P.....
A16 .....F.....D.....G.....-.....
B9 D.....F.....I.....G.....-.....R
C7 .....F.....G.....-.....W
1.1 ...W.....F.....D.....G.....-.....
1.2 .....F.....D.....AG...-.....K.W
2.4 D.....F.....K.....-.....
2.10 .....N....F.....T.....G.....K.....I-L...H.....
16.11 D.....F.K.....E.....G.....-.....
16.12 D.....F.....D.....Q.....-.....
16.13 .....F....S.....Q.....-.....
16.14 .....K....F.....E.....G.....K.....I-L...H.....
16.15 .....N....F.....T.....GK.....K.....I-L...H.....
16.16 .....F.....D.....-.....
16.22 D.....F.....-.....Y.....
16.26 .....F.....T.....-.....
16.60.3 .....F....S.....Q.....-.....
17.3 D.....YT.....F.....V.....M.....G.....E.....-.....
Cod-UB6-36d .....Y.N....F.....H....P.....T.T.....E.....K.....-....P.....
Cod-UF6 .....Y.....F.....H.L..P.....G.....-.....
Cod-UF8 .....Y.....F.....H.L..P.....G.....-.....
Cod-c28 D.....Y.....F.....H....P.....N.....G.....K.....-.....
Cod-c41 D.....Y.....F.T.....H....P.....G.....K.....-.....S
Cod-c33 .....Y.....F.....H....P.....-.....-.....N.....
Cod-c39 .....Y.....F.....H....P.....D.....G.....K.....-.....S
Cod-c20 .....N....F.....H....P.....D.T.....G.....K.....-.....R
Cod-c32 .....N....F.....H....P.....Y.T.....G.....K.....-.....R
Cod-c24 .....N....F.....H....P.....T.....G.....K.....-T....H.....
Onmy-UBA-4101 VP.S.....KT.....T.....Y.SG.M.S.HK...DHH.D.E...I.Q.D.....K.SH.TVTPEEWKN-NK.Q....VT..QKDFIKV.TESE.Q..W
Onmy-UBA-4201 V..S.....KT.....T.....Y.SG.M.F.QK....QHGD.E...I.Q.D.....K.TH.TVTPEEWKN-NK.Q.....A....DI.KV.IESE.Q..F
Onmy-UBA-4401 VP.S.T...KT.....T.....Y.SG.M.F.QK...DQHGD.E...T...D.....K.TH.T.TPEEWKK-KQ.Q....VT..KEDFIKV.TESE.K..W
Onmy-UBA-4501 VP.L...F.KTS..L.T.....Y.SD.MLS.QK...DHH.D.EY..T...D.....K.IH.TVTPEDRKN-SK.Q....VK..KEDFIKV.PDPDAAN--
Onmy-UBA-4601 V..S.....KT.....T.....Y.SG.M.F.QK....QHGD.E...I.Q.D.....K.TH.TVTPEEWKN-NK.Q.....A....DI.KV.IESE.Q..F
Onmy-UBA-4701 VP.S...F.KT.....T.....Y.SD.M.S.QK...DHH.D.EY..T...N.....K.IH.TVTPDDRKN-NK.Q....VK..MEDFIKV.TDPDAAN--


Transmembrane Domain
Cytoplasmic Domain
310
320
330
340
350
360
370
380
390
pBK-CMV-1
pBK-CMV-3
pBK-CMV-4
pBK-CMV-6
A9
A16
B9
C7
1.1
1.2
2.4
2.10
16.11
16.12
16.13
16.14
16.15
16.16
16.22
16.26
16.60.3
17.3
Cod-UB6-36d
Cod-UF6
Cod-UF8
Cod-c28
Cod-c41
Cod-c33
Cod-c39
Cod-c20
Cod-c32
Cod-c24
Onmy-UBA-4101
Onmy-UBA-4201
Onmy-UBA-4401
Onmy-UBA-4501
Onmy-UBA-4601
Onmy-UBA-4701

141 and #255 – 259 surround residues known to be important in β_2-microglobulin binding. The region at position #187 – 190 contains a residue important in peptide anchoring and is in close proximity to a cystine residue which forms a disulfide bond linking the α_2 and α_3 domains. It is interesting to note that one of the 3 nucleotide deletions noted earlier results in a deletion of this cysteine in allele 17.3. The regions at position #214 – 224 and #279 – 282 also contain cysteines involved in a disulfide bond between the α_2 and α_3 domains.

Several key residues have been highlighted within the alignment. Highlighted in yellow are residues known to be vital in binding β_2-microglobulin. Very few of these residues showed a substitution in the haddock and cod alleles. Highlighted in red in the protein alignment are the nine highly conserved residues responsible for anchoring the peptide termini to the binding grove. Several of the haddock and cod alleles had more than one substitution at these positions. These differences are summarized in Table 3.2. It is interesting to note that the lysine (K) at position #165 is not well conserved throughout the haddock and cod alleles, but was substituted with arginine (R). Furthermore, lysine and arginine are chemically very similar amino acids, which make it possible for this substitution to have little effect on antigen anchoring. Therefore, this substitution was ignored when tallying the number of substituted residues for each allele. The other substituted amino acids were also compared to the expected amino acids for chemical similarity. In alleles 2.10 and 16.15 a phenylalanine (F) was substituted for tyrosine at position 191. Although phenylalanine and tyrosine are both aromatic amino acids, phenylalanine is non-polar and tyrosine is polar due to its hydroxyl group. In this case, further study would be required to determine if this substitution would affect the antigen anchoring. For all other substitutions the substituted amino acids are chemically different from the expected amino acid, and therefore, these substitutions would have an impact on antigen anchoring. Only eight of the 13 full length alleles had fewer than 2 substitutions. Of these eight, two had inverted ORFs and two were joined with other gene sequences. This left only four alleles, pBK-CMV-4, A16, B9, and 16.12, with typical, conserved sequences for classical MH Class I.

Table 3.2: Summary of the amino acid residues substitutions and structural modifications of the haddock and cod alleles. Only full length alleles are shown since the partial alleles were missing the sequences containing all or some of the residues of interest. Expected amino acid residues, with positions indicated, were based on work by Shum et al., 1999. The number of differences of these residues does not include the arginie (R) substitution for lysine (K).

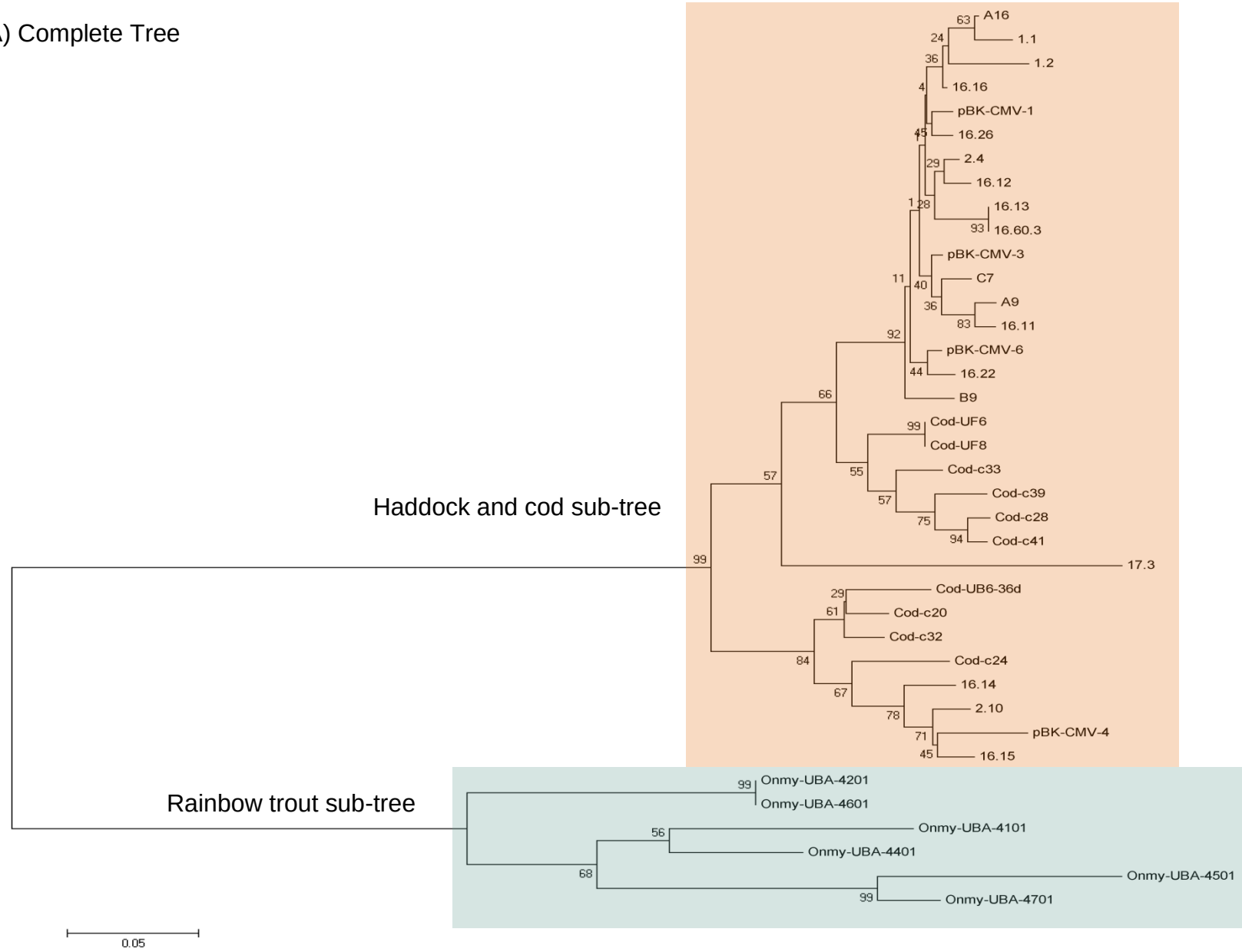
	Expected Amino Acid Residues									#	ORF	"Linked"
	Y28	Y79	R104	F142	T162	K165	W/L166	Y179	Y191	Diff.	Inversion	Sequence
pBK-CMV-1	-	-	-	-	-	-	-	-	H	1		√
pBK-CMV-3	-	-	-	-	-	-	-	-	-	0	√	
pBK-CMV-4	-	-	-	-	-	R	-	-	-	0		
pBK-CMV-6	-	-	-	-	N	R	-	-	H	2		
A9	-	D	-	-	-	R	-	-	L	2		
A16	-	-	-	-	-	R	-	-	-	0		
B9	-	-	-	-	-	R	-	-	H	1		
C7	-	-	-	-	-	R	-	-	-	0		√
1.2	-	E	-	-	-	R	-	-	-	1	√	
2.10	-	-	-	-	I	R	-	F	-	2		
16.12	-	D	-	-	-	R	-	-	-	1		
16.15	S	-	-	-	-	R	-	F	L	3		
17.3	-	-	S	-	-	R	-	-	L	2		
Cod-UB6-36d	-	-	-	-	-	R	-	-	F	1		
Cod-UF6	-	S	-	-	-	R	-	-	-	1		
Cod-UF8	-	-	-	-	-	R	-	-	-	0		
Cod-c28	-	-	-	-	-	-	-	-	-	0		

Using the amino acid sequence alignment a phylogenetic tree, shown in Figure 3.8, was constructed. The tree was comprised of two sub-trees: the rainbow trout alleles and the haddock and cod alleles. From the haddock and cod sub-tree three distinct groups of alleles could be seen. The first group contained the majority of the haddock alleles and no cod alleles. The second group contained only cod alleles. The third group contained both cod and haddock alleles. One haddock allele, 17.3, however did not group with any of the others. This tree was constructed using both full and partial alleles, some of which were missing most of the α_1 and α_2 domains. Since most of the allelic variances occur in these domains two additional trees of haddock and cod alleles were constructed.

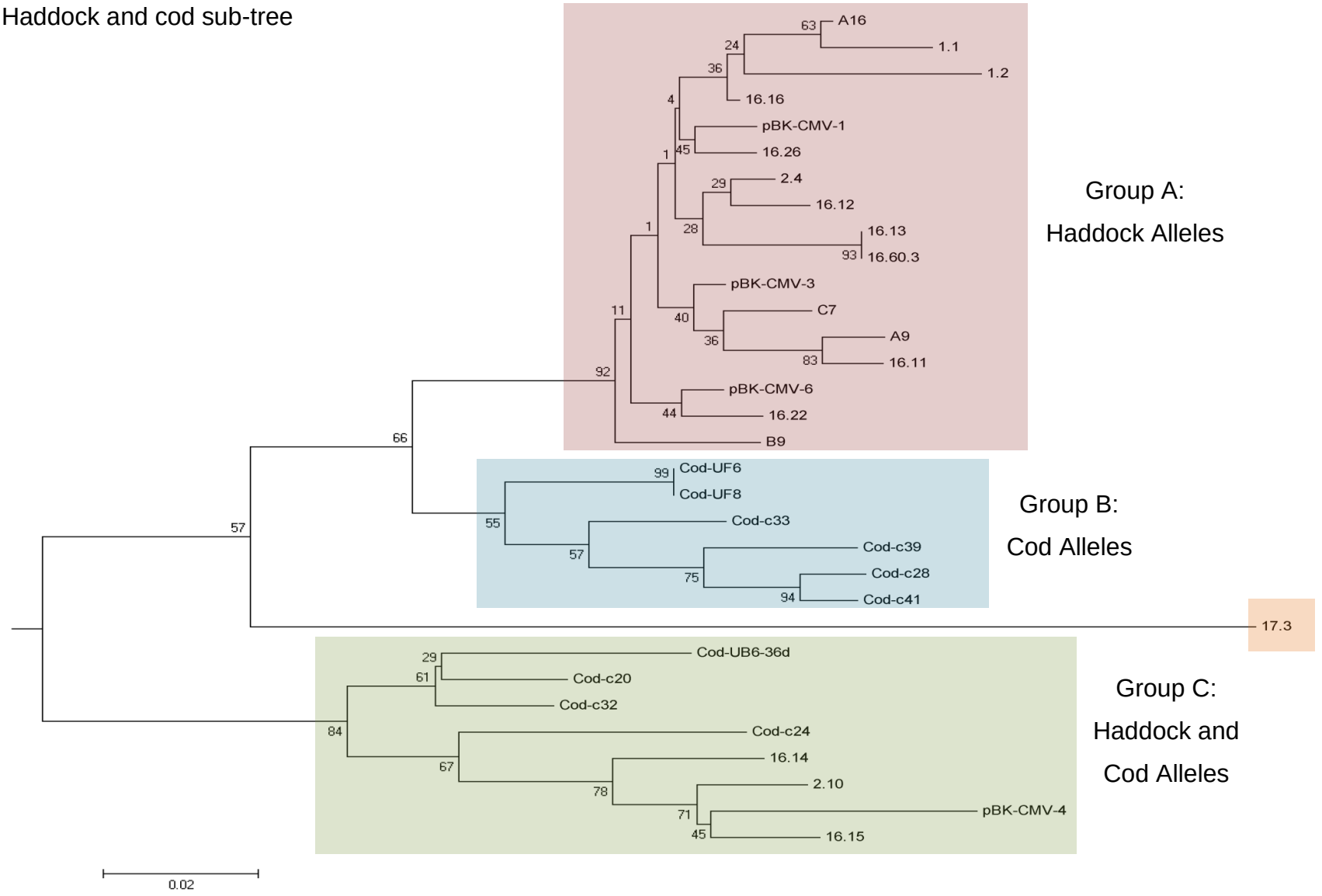
The first was constructed using the complete amino acid sequences of the full length alleles. This tree is shown in Figure 3.9. This tree mirrored the pattern found in the tree of all alleles in that there were distinct groups of haddock only, cod only and haddock and cod alleles, and allele 17.3 did not group with other alleles. To determine if there were any allelic lineages based on the α_1 and α_2 domains a second tree was constructed using the nucleotide sequence of the α_1 and α_2 domains and is shown in Figure 3.10. In this tree the alleles grouped quite differently than in the previously described trees in that the haddock and cod alleles did not group separately. The alleles were divided into three groups. The low bootstrap values of the first group, shown in purple, may be a result of the possibility that these alleles were actually part of distinct lineages which, due to the limited number of alleles available for analysis, were not clearly visible. The last two groups, however, had high bootstrap values indicating that these groups form distinct allelic lineages. Furthermore, the last group was identical to the Group C found in the other trees.

Figure 3.8: Neighbour-joining tree of haddock, cod and rainbow trout MH Class I alleles. The tree was constructed using both full and partial alleles. The Poisson correction model was used with complete deletion of gaps or missing data. Bootstrap values were based on 1000 replicates. A) The complete tree clearly shows the distance between the rainbow trout sequences and the haddock and cod sequences. B) An expanded view of the haddock and cod sub-tree. Within this sub-tree there are three groupings: a group of haddock only alleles, a group of cod only alleles and a group with both haddock and cod alleles. One allele from haddock, 17.3, did not fit into any of the groups.

A) Complete Tree



B) Haddock and cod sub-tree



Group A:
Haddock Alleles

Group B:
Cod Alleles

Group C:
Haddock and
Cod Alleles

Figure 3.9: Neighbour-joining tree of full length haddock and cod alleles. The amino acid sequences of the 13 full length haddock alleles as well as the four full length cod alleles were used. The tree was constructed using the Poisson correction model with complete deletion of gaps. Bootstrap values were based on 1000 replicates.

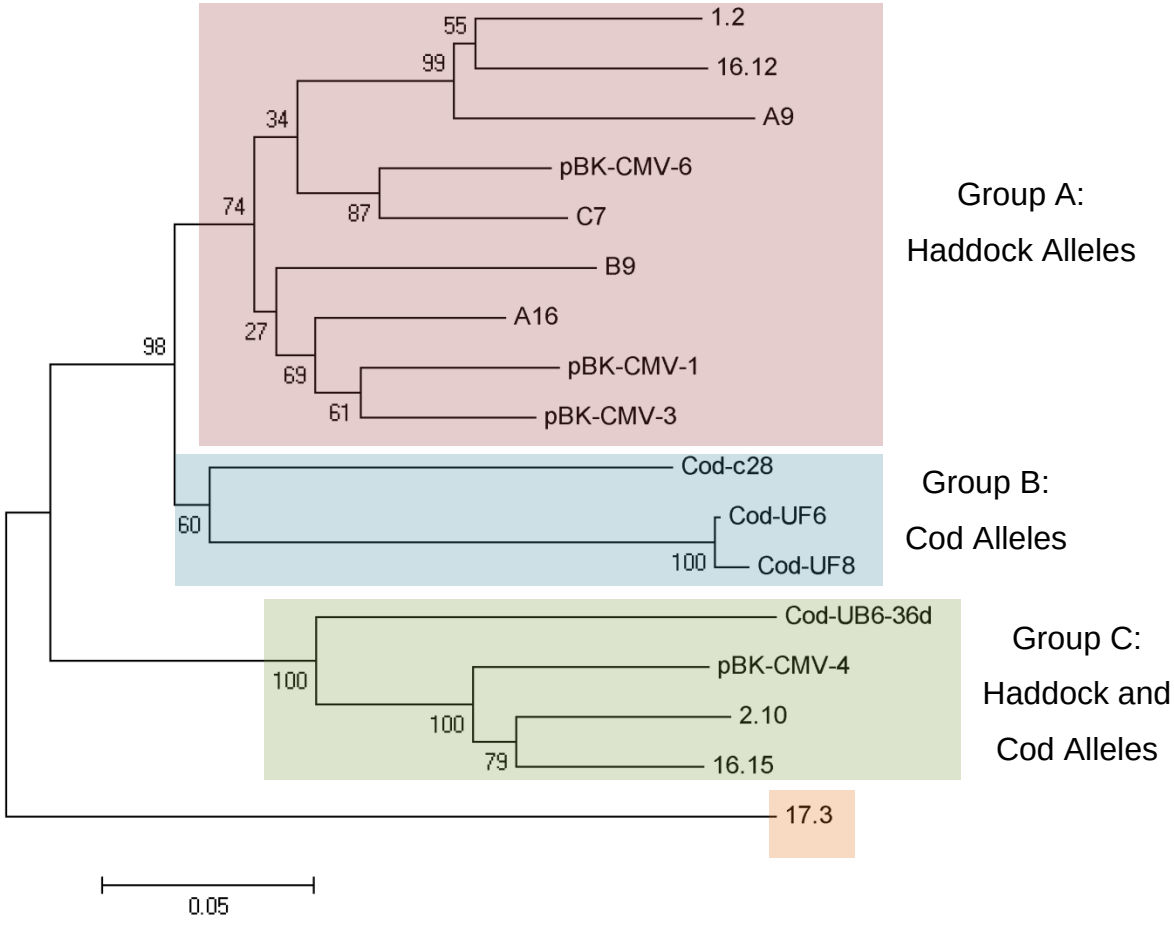
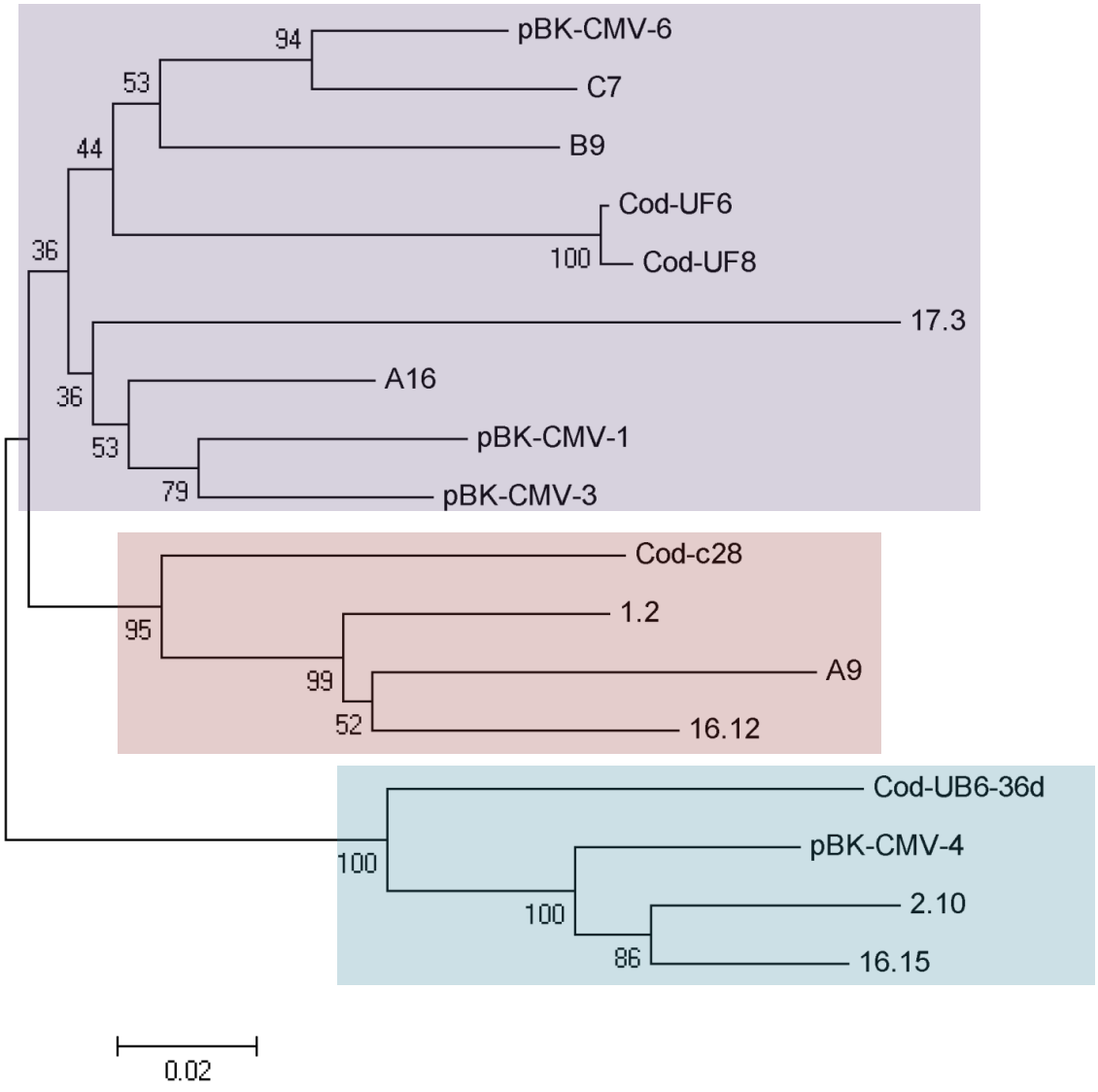

55
1.2
99
16.12
A9
34
pBK-CMV-6
74
87
C7
B9
A16
27
98
69
pBK-CMV-1
61
pBK-CMV-3
Group A:
Haddock Alleles
Cod-c28
60
Cod-UF6
100
Cod-UF8
Group B:
Cod Alleles
Cod-UB6-36d
100
pBK-CMV-4
100
2.10
79
16.15
Group C:
Haddock and
Cod Alleles
17.3
0.05

Figure 3.10: Neighbour-joining tree of the α_1 and α_2 domains of full length haddock and cod alleles. The nucleotide sequences of the α_1 and α_2 domains of the full length haddock and cod alleles were used in the Jukes and Canter model with complete deletion of gaps. Bootstrap values were based on 1000 replicates.


94
pBK-CMV-6
53
C7
B9
44
Cod-UF6
36
100
Cod-UF8
17.3
A16
36
53
pBK-CMV-1
79
pBK-CMV-3
Cod-c28
95
1.2
99
A9
52
16.12
Cod-UB6-36d
100
pBK-CMV-4
100
2.10
86
16.15
0.02

4. Discussion

In 1999, a study by Perrson et al. of cod MH Class I revealed an unusually high number of genes and consequently a high number of expressed alleles. These finding were reflected in another study by Miller et al. in 2002. This high number of expressed alleles presents a potential problem. As T-cells develop in the thymus, MH molecules are displayed on the surface of thymus cells and partake in the selection process which removes self-reactive and non-reactive T-cells from the population (Takahama, 2006). This process drives the need to establish a working balance between having too many and too few MH alleles. A sufficient number of alleles are needed to protect against a variety of diseases, while having too many alleles would lead to the deletion of too many developing T-cells.

Since haddock is a species closely related to cod experiments were undertaken to determine if these abnormalities in cod were also present in haddock. The results of these experiments showed that haddock also possessed an unusually high number of MH Class I genes and that these were expressed (unpublished data). This lead to experiments to determine what, if any, mechanisms were in place to limit the number of functional MH Class I molecules being expressed.

A total of 22 unique alleles have been isolated from an individual haddock and based on analysis of these sequences three mechanisms have been proposed to be responsible for reducing the number of alleles expressed into functional molecules. The first mechanism is inversion of the open reading frame and the second is linking the allele with another gene or transcript to produce a polycystronic-like sequence. An inverted ORF when read in the sense orientation would essentially be nonsense and would likely not begin with the necessary nucleotide sequence “ATG”. In the case of the polycystronic-like transcripts, the ribosome would dissociate from the mRNA at the end of the first gene product and the MH Class I sequence would never be translated. Therefore, both of these mechanisms would effectively prevent translation of the transcripts into functional protein. Additionally, the amino acid sequences of several of these alleles suggest that most of the alleles have mutations in amino acids critical for

classical MH class I antigen presentation function, making them into non classical MH molecules, and therefore, they may not be involved in T-cell selection or immune responses.

A total of seven alleles aligned in the antisense orientation. A distinction must be made, however, between the alleles with inverted ORFs and those in which the entire transcript was antisense. Alleles Haddock UA pBK-CMV-3 and 1.2, had antisense ORFs but sense poly-A tails. Furthermore, directional cloning of these alleles was verified by identification of the restriction enzyme sites and linker sequences used for the library creation. The remaining five alleles differ in that they did not have poly-A tails in the sense orientation, and the directional cloning could not be completely confirmed. Only the 5' Eco RI enzyme site and linker was found. Since this linker was added by blunt end ligation it could be added to both ends of the transcript and therefore, cannot be used solely to confirm directional cloning. It is possible that the antisense orientation of these five alleles was a result of erroneous directional cloning, and not the inversion mechanism seen in Haddock pBK-CMV-3 and 1.2.

Six alleles were found "linked" with sequences of other genes. Two of these alleles were notable for several reasons. Alleles Haddock UA pBK-CMV-1 and 17.3 both contained full sequences of an MH Class I allele and the other gene, ribosomal protein L12 and hemoglobin beta 1 respectively. Additionally, both of these alleles were linked with the other genes using a very similar sequence. This linking sequence was not part of any sequences used for creating the library. Since it is highly unlikely that the same sequence would be randomly joined in between the transcripts, the linked nature of these alleles appears to be another mechanism for inactivating allele expression. The remaining four alleles did not show a linking sequence of any kind. Furthermore, all but one were partial sequences. Since blunt end ligation was used it is possible that independent transcripts were ligated together during the library creation. Therefore, it is necessary to validate the linked nature of these sequences.

Both classical and non-classical MH Class I molecules play an important role in immune function but only classical alleles are thought to function in antigen presentation and T

cell selection. One key difference between classical and non-classical molecules are the residues at nine key positions within the α_1 and α_2 subunits. These nine residues are involved in anchoring peptide to the binding grove, and as shown by Shum et al. in 1999, are highly conserved throughout different species.

Analysis of these nine key residues within the MH Class I sequence has reveled another potential way haddock may prevent excessive T-cell loss during development. Five of the 13 full length alleles showed two or more substitutions at these positions. Functional analysis of MHC molecules in mammals shows that two or more substitutions of these residues are sufficient to change the function of the molecules from classical to non-classical (Shum et al., 1999). Therefore, it is highly likely that these five alleles in haddock are non-classical. As such, it is possible that they would not participate in T-cell selection in the thymus.

Residues involved in binding β_2-microglobulin as well cystine residues involved in disulfide bonds between the α_2 and α_3 sub-units were also examined. These residues showed a high degree of conservation, which indicates that they are not involved in limiting allele expression. One interesting exception is a missing cystine in allele 17.3. Since this allele would have a significantly different conformation its function would also differ significantly. The implications of this result require further study.

To summarize, 22 unique MH Class I alleles were isolated from an individual haddock. Eight alleles were found to have classical MH Class I sequences. Of these eight, two had inverted ORFs and two were linked with other gene sequences. Therefore, there are only four remaining classical MH Class I sequences which could be expressed into functional protein. This number reflects the number of alleles typically seen expressed in other fish species. Therefore, haddock appear to have a set of classical and non-classical MH Class I genes and use two novel mechanisms for limiting the expression of classical MH Class I alleles.

Prior to research on haddock, a dodecaploid *Xenopus* species was the only example of a species limiting its MH(C) Class I gene expression (Sammut et al., 2002). Due to its high ploidy it would be expected that these animals would have a high number of MHC

Class I alleles. However, Southern blot analysis and cDNA analysis revealed that individuals possessed a maximum of four alleles. This species, therefore, removes the DNA encoding additional MHC Class I genes to prevent the expression of too many alleles. This expression limiting method differs greatly from what has been shown in haddock. The three mechanisms of inverted ORFs, joined sequences and non-classical mutations therefore represent novel mechanisms used to limit the expression of MH Class I alleles.

Further research in this area is clearly warranted. There is clear evidence of three mechanisms for limiting the number of functional MH Class I molecules expressed by an individual. For two of these mechanisms not all of the experimental data is conclusive. The orientation of alleles A16, 16.14, 16.16, 16.22, 16.26 and 17.3 needs to be confirmed. Also, the linkage of alleles Haddock UA pBK-CMV-3, C7, 16.16, 16.22, 16.26, and 17.3 needs to be verified. Full length sequences of the partial clones are also required. In both cases, the questions concerning the validity of the results stems from the process of the library creation. Using cDNA from an individual haddock, PCR reactions with allele specific primers could confirm linkages. Verifying the orientation of alleles would require investigations of the genomic DNA.

Since haddock show such unusual properties for the regulation of their MH Class I genes the possibilities for future research are broad. It would be interesting to investigate whether the changes in orientation and the gene linking occur at the DNA or RNA level. What is the significance of the unknown sequences attached to several of the alleles? Are different alleles being expressed in different tissues? How does the high level of expression impact the number of T cells? Are there mechanisms that limit expression at the protein level? These are some of the questions that could be asked as we attempt to gain a deeper understanding of the immune function of this unusual and interesting species.

Refereces

Bjorkman, P.J., and Parham, P. (1990). Structure, function, and diversity of class I major histocompatibility complex molecules. Annu Rev Biochem *59*, 253-288.

Bricknell, I.R., Bron, J.E., and Bowden, T.J. (2006). Diseases of gadoid fish in cultivation: a review. ICES Journal of Marine Science *63*, 253-266.

Chambers, M.D., and Howell, W.H. (2006). Preliminary information on cod and haddock production in submerged cages off the coast of New Hampshire, USA. ICES Journal of Marine Science *63*, 385-392.

Costantini, S., Buonocore, F., and Facchiano, A.M. (2008). Molecular modelling of co-receptor CD8 alpha alpha and its complex with MHC class I and T-cell receptor in sea bream (Sparus aurata). Fish Shellfish Immunol *25*, 782-790.

Dijkstra, J.M., Kollner, B., Aoyagi, K., Sawamoto, Y., Kuroda, A., Ototake, M., Nakanishi, T., and Fischer, U. (2003). The rainbow trout classical MHC class I molecule Onmy-UBA*501 is expressed in similar cell types as mammalian classical MHC class I molecules. Fish Shellfish Immunol *14*, 1-23.

Dixon, B., and Stet, R.J. (2001). The relationship between major histocompatibility receptors and innate immunity in teleost fish. Dev Comp Immunol *25*, 683-699.

Ehrlich, R., and Lemonnier, F.A. (2000). HFE--a novel nonclassical class I molecule that is involved in iron metabolism. Immunity *13*, 585-588.

Fietta, P., and Delsante, G. (2009). The effector T helper cell triade. Riv Biol *102*, 61-74.

Fischer, U., Utke, K., Somamoto, T., Kollner, B., Ototake, M., and Nakanishi, T. (2006). Cytotoxic activities of fish leucocytes. Fish Shellfish Immunol *20*, 209-226.

Garcia, K.C. (1999). Molecular interactions between extracellular components of the T-cell receptor signaling complex. Immunol Rev *172*, 73-85.

Gomes, A.Q., Correia, D.V., and Silva-Santos, B. (2007). Non-classical major histocompatibility complex proteins as determinants of tumour immunosurveillance. EMBO Rep *8*, 1024-1030.

Guselnikov, S.V., Najakshin, A.M., and Taranin, A.V. (2003). Fugu rubripes possesses genes for the entire set of the ITAM-bearing transmembrane signal subunits. Immunogenetics *55*, 472-479.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series *41*, 95-98.

Janeway, C.A., P. Travers, M. Walprot, M. Shlomchik (2001). Immuno Biology, 5th edn (New York, NY: Garland Publishing).

Johnson, N.A., Vallejo, R.L., Silverstein, J.T., Welch, T.J., Wiens, G.D., Hallerman, E.M., and Palti, Y. (2008). Suggestive association of major histocompatibility IB genetic markers with resistance to bacterial cold water disease in rainbow trout (Oncorhynchus mykiss). Mar Biotechnol (NY) *10*, 429-437.

Kiryu, I., Dijkstra, J.M., Sarder, R.I., Fujiwara, A., Yoshiura, Y., and Ototake, M. (2005). New MHC class Ia domain lineages in rainbow trout (Oncorhynchus mykiss) which are shared with other fish species. Fish Shellfish Immunol *18*, 243-254.

Laky, K., and Fowlkes, B.J. (2008). Notch signaling in CD4 and CD8 T cell development. Curr Opin Immunol *20*, 197-202.

Lanzavecchia, A. (1990). Receptor-mediated antigen uptake and its effect on antigen presentation to class II-restricted T lymphocytes. Annu Rev Immunol *8*, 773-793.

Larosa, D.F., and Orange, J.S. (2008). 1. Lymphocytes. J Allergy Clin Immunol *121*, S364-369; quiz S412.

Magnadottir, B., Jonsdottir, H., Helgason, S., Bjornsson, B., Solem, S.T., and Pilstrom, L. (2001). Immune parameters of immunised cod (Gadus morhua L.). Fish Shellfish Immunol *11*, 75-89.

Miller, K.M., Kaukinen, K.H., and Schulze, A.D. (2002). Expansion and contraction of major histocompatibility complex genes: a teleostean example. Immunogenetics *53*, 941-963.

Nel, A.E. (2002). T-cell activation through the antigen receptor. Part 1: signaling components, signaling pathways, and signal integration at the T-cell antigen receptor synapse. J Allergy Clin Immunol *109*, 758-770.

Pamer, E., and Cresswell, P. (1998). Mechanisms of MHC class I--restricted antigen processing. Annu Rev Immunol *16*, 323-358.

Park, H., Zhou, H., Bengten, E., Wilson, M., Chinchar, V.G., Clem, L.W., and Miller, N.W. (2002). Activation of channel catfish (Ictalurus punctatus) T cells involves NFAT-like transcription factors. Dev Comp Immunol *26*, 775-784.

Persson, A.C., Stet, R.J., and Pilstrom, L. (1999). Characterization of MHC class I and beta(2)-microglobulin sequences in Atlantic cod reveals an unusually high number of expressed class I genes. Immunogenetics *50*, 49-59.

Picchietti, S., Guerra, L., Buonocore, F., Randelli, E., Fausto, A.M., and Abelli, L. (2009). Lymphocyte differentiation in sea bass thymus: CD4 and CD8-alpha gene expression studies. Fish Shellfish Immunol *27*, 50-56.

Randelli, E., Scala, V., Casani, D., Costantini, S., Facchiano, A., Mazzini, M., Scapigliati, G., and Buonocore, F. (2008). T cell receptor beta chain from sea bream

(Sparus aurata): molecular cloning, expression and modelling of the complexes with MHC class I. Mol Immunol *45*, 2017-2027.

Sammut, B., Marcuz, A., and Pasquier, L.D. (2002). The fate of duplicated major histocompatibility complex class Ia genes in a dodecaploid amphibian, Xenopus ruwenzoriensis. Eur J Immunol *32*, 2698-2709.

Schroeder, H.W., Jr., and Cavacini, L. (2010). Structure and function of immunoglobulins. J Allergy Clin Immunol *125*, S41-52.

Shum, B.P., Rajalingam, R., Magor, K.E., Azumi, K., Carr, W.H., Dixon, B., Stet, R.J., Adkison, M.A., Hedrick, R.P., and Parham, P. (1999). A divergent non-classical class I gene conserved in salmonids. Immunogenetics *49*, 479-490.

Stern, L.J., and Wiley, D.C. (1994). Antigenic peptide binding by class I and class II histocompatibility proteins. Structure *2*, 245-251.

Stet, R.J., Kruiswijk, C.P., and Dixon, B. (2003). Major histocompatibility lineages and immune gene function in teleost fishes: the road not taken. Crit Rev Immunol *23*, 441-471.

Sullivan, C., and Kim, C.H. (2008). Zebrafish as a model for infectious disease and immune function. Fish Shellfish Immunol *25*, 341-350.

Sultmann, H., Mayer, W.E., Figueroa, F., O'HUigin, C., and Klein, J. (1993). Zebrafish Mhc class II alpha chain-encoding genes: polymorphism, expression, and function. Immunogenetics *38*, 408-420.

Takahama, Y. (2006). Journey through the thymus: stromal guides for T-cell development and selection. Nat Rev Immunol *6*, 127-135.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol *24*, 1596-1599.

Treasurer, J.W., Sveier, H., Harvey, W., Allen, R., Cutts, C.J., Mazorra de Quero, C., and Ford, L. (2006). Growth, survival, diet, and on-growing husbandry of haddock Melanogrammus aeglefinus in tanks and netpens. ICES Journal of Marine Science *63*, March.

Watts, M., Munday, B.L., and Burke, C.M. (2001). Immune responses of teleost fish. Aust Vet J *79*, 570-574.

Appendix I

ZapExpress® Vector System

The ZapExpress Vector System makes use of two types of phage and two strains of *E. coli*, XL1-Blue-MRF' and XLOLR. The first phage, which will be referred to as "library phage", is a modified lambda phage and is the vector into which cDNA fragments are cloned. The second type of phage is referred to as "helper phage" and is used in the conversion of the lambda phage into the plasmid pBK-CMV.

The library phage are propagated in the *E. coli* strain XL1-Blue-MRF'. They grow with a typical lytic life-cycle and the plaques are small with complete lysis of host cells (plaques are completely clear). These phage can be converted to phagemids by a process called excision. Mass excision can be performed on an entire library, whereas a simple excision is done on an individual clone. The excision process is accomplished by co-infecting a culture of XL1-Blue-MRF' with both the library phage and the helper phage[8]. The helper phage acts through a double cross-over to remove the genes of the lambda phage necessary for lysis of the host cell at the end of its life cycle. The result is a phagemid, which contains the insert, that is capable of infecting but not lysing cells. This phagemid is used to infect the *E. coli* strain XLOLR. This strain of *E. coli* is resistant to lambda phage infection so that contamination from any remaining library phage is prevented. Once the phagemid infects XLOLR the DNA, including the insert, is replicated as a plasmid.

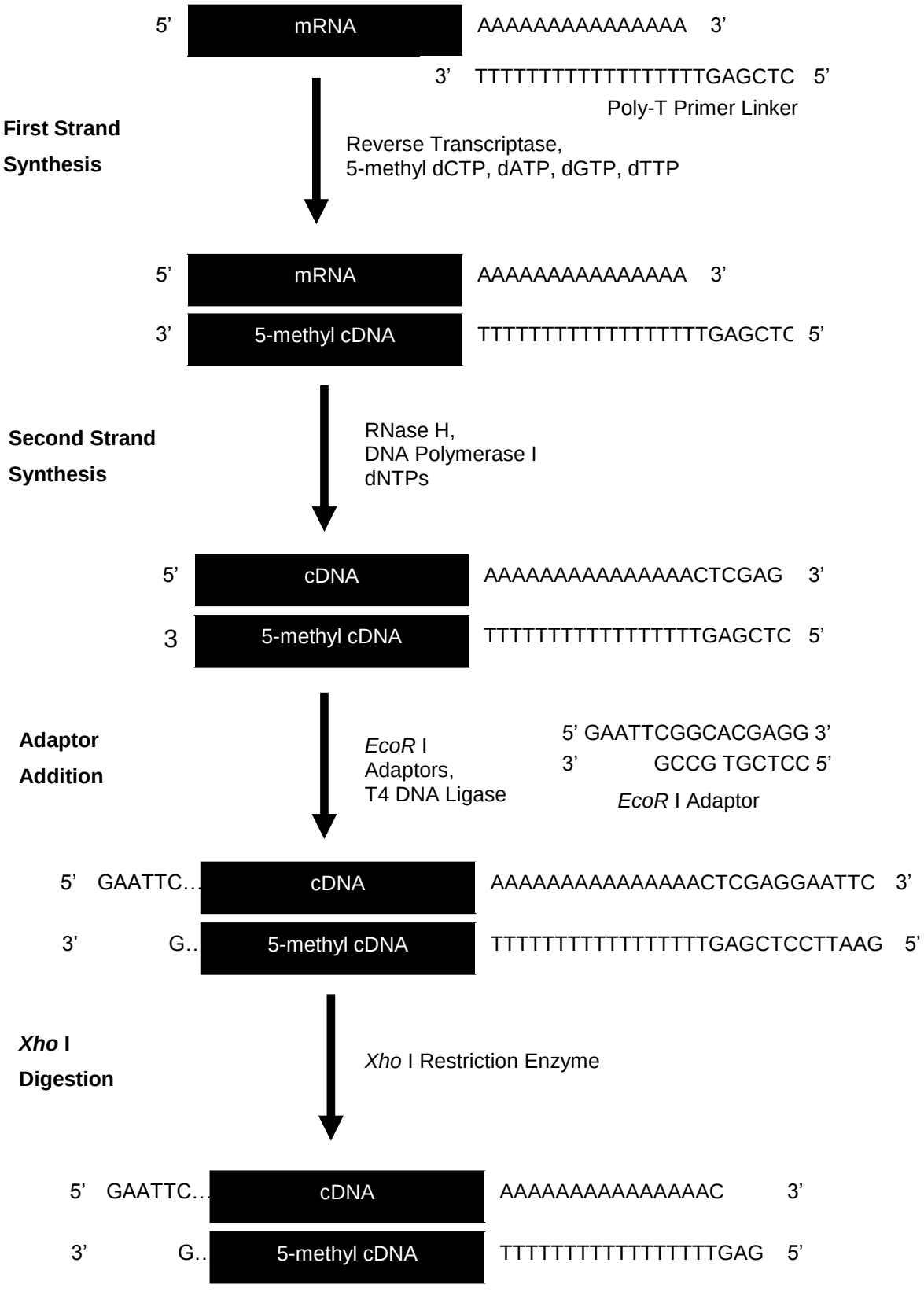
[8] The helper phage plaque morphology is different than that of the library phage. It is a larger plaque with incomplete lysis (i.e. opaque in appearance).

Appendix II

Zap Express® cDNA Library Construction Kit

The cDNA library used in this project was created using the Zap Express® cDNA Library Construction Kit. A poly-T primer linked with an Xho I restriction enzyme (REN) site is used during first strand synthesis. The poly-T primer binds to the poly-A tail of the mRNA transcripts and synthesis is carried out by a high-fidelity reverse transcriptase. In the first strand synthesis reaction 5-methyl dCTP is used in the mix of dNTPs so that the newly synthesized cDNA is protected against digestion with RENs. In the second strand synthesis reaction RNase H is added, which nicks the mRNA bound to the first cDNA strand. These nicked fragments act as primers for synthesis by DNA polymerase I. Synthesis is carried out using a regular mix of dNTPs so that the second cDNA strand can be cut with RENs. Following the second strand synthesis an EcoR I adaptor is blunt-end ligated using T4 DNA ligase. Finally, Xho I is used to cut the 3' end. The final product is a transcript with a 5' EcoR I site and a 3' Xho I site. The diagram in Figure A.1 was modified from the Zap Express cDNA Library Construction Kit Manual, and illustrates the process.

Figure A.1: Diagram of the method of the cDNA library construction.


5’ mRNA AAAAAAAAAAAAAAA 3’
3’ TTTTTTTTTTTTTTTTTTGAGCTC 5’
Poly-T Primer Linker
First Strand Synthesis
Reverse Transcriptase,
5-methyl dCTP, dATP, dGTP, dTTP
5’ mRNA AAAAAAAAAAAAAAA 3’
3’ 5-methyl cDNA TTTTTTTTTTTTTTTTTTGAGCTC 5’
Second Strand Synthesis
RNase H,
DNA Polymerase I
dNTPs
5’ cDNA AAAAAAAAAAAAAAACTCGAG 3’
3 5-methyl cDNA TTTTTTTTTTTTTTTTTTGAGCTC 5’
Adaptor Addition
EcoR I Adaptors,
T4 DNA Ligase
5’ GAATTCGGCACGAGG 3’
3’ GCCG TGCTCC 5’
EcoR I Adaptor
5’ GAATTC… cDNA AAAAAAAAAAAAAAACTCGAGGAATTC 3’
3’ G.. 5-methyl cDNA TTTTTTTTTTTTTTTTTTGAGCTCCTTAAG 5’
Xho I Digestion
Xho I Restriction Enzyme
5’ GAATTC… cDNA AAAAAAAAAAAAAAAC 3’
3’ G.. 5-methyl cDNA TTTTTTTTTTTTTTTTTTGAG 5’