# *An investigation into the existence of Cyprinid Herpesvirus 3 encoded microRNAs*

A thesis submitted for the degree of Ph.D

By

Owen Donohoe B.Sc. (Hons)

The work in this thesis was carried out under the supervision of
Dr. Dermot Walls (DCU, Dublin)
Dr. Kathy Henshilwood (Marine Institute, Galway)
Keith Way (Cefas, UK)

School of Biotechnology
Dublin City University
July 2013

## Declaration of ownership

*I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of PhD. is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.*

Signed: _____ ID No.: 54053087 Date: _____

## Acknowledgements

I would like to express my profound gratitude to my supervisors Dr. Kathy Henshilwood (Marine Institute), Dr. Dermot Walls (DCU) and Keith Way (Cefas, UK) whose knowledge, advice, support and patience and at every stage of the project was a tremendous help and instrumental in ensuring that the research reached its full potential.

To Kathy for supporting the introduction of the Fellowship Program, taking me on in the first place, making sure that I always had everything I needed in terms of materials/consumables and for also actively encouraging the exploration of new technology as a means driving the research further. Thanks for making sure that I give talks on the project on a regular basis, allowing me to develop the public speaking skills that are so vital to all researchers today. Thanks for also constantly arranging things in the background and making sure I stayed on track over the last few years. In addition, I would like to thank both you and your husband Bill for the great nights out that many of us spent you in Galway/Oranmore.

Thanks to Dermot for so enthusiastically accepting our invitation to take on the role of Academic Supervisor for the project. Your experience and expertise in this field provided us with many valuable insights and you proved to be a wonderful source of ideas regarding the direction of the project right from the start. Indeed, it was one of your early suggestions that guided us down the microRNA route in the first place, and proved to be a real winner. Also I am very grateful for your help in the final stages of the project, the tremendous effort you made to get through my (rather lengthy) thesis chapters within the time-frame available and for ensuring that the viva occurred so promptly after the thesis submission date. I will always be grateful to you for making this project what it was and I look forward to getting you that pint or two some night in Dublin.

I would also like to thank to Keith for spotting the need for a project on CyHV-3, specifically the need for identification of CyHV-3 latency associated transcripts, thus getting the whole project going. Thanks to both you and your colleagues in Cefas for being a fantastic source of help in getting the project off the ground, providing the

the necessary cells, viral stocks, protocols and technical support to get me started. You also continued to be a great help at later stages of the project, in providing the use of your lab facilities and the use of some of your *in vivo* samples. You, your colleagues and your family always made me feel very welcome over in Weymouth which is something that meant a lot to me. In addition, I'd like to thank you for your eclectic taste in music, introducing me to many new artists that may have taken me much longer to discover otherwise.

I am also very grateful to my colleagues in the Marine Institute for all the help that they gave me along the way. Thanks to my FHU team leader Dr. Neil Ruane for allowing me to be flexible regarding my dual role as Molecular Analyst and Research Fellow within the FHU and for making sure that I attended the EAFP conference in Split. Thanks to Dr. Stephen McCleary for teaching me how to use the sequencer and for being a great housemate. Thanks to Dave Swords and Lorraine McCarthy for getting me up and running with the cell culture and for allowing me to borrow consumables when mine were low. Thanks to Michelle Geary for her great sense of humour, it was great to share a partition with you. Thanks to Cathy Hickey for the little tips and bits of advice which she dispensed on a regular basis, for letting me borrow things form the bacteriology lab. Thanks to Evelyn Collins for all her baking which I thoroughly enjoyed over the last few years, maybe I'll give it a go myself now. Thanks to my Section Manager, Fiona Geoghegan who was a great help in ensuring that I had everything that I needed. Thanks to Dr. Margaret Rae for suggesting that I use the BioLuminizer in her lab to visualise my northern blots, which and allowed me to obtain much better results than I would have otherwise obtained with my X-Ray-film and make-shift darkroom approach. To my fellow Research Fellows, the soon-to-be Drs John Flannery and Paulina Rajko-Nenow who I shared the labs with, it was always great to know that there was someone else in the same boat and going though the same trials and tribulations, thank you for all your wonderful cooking and all the happy times I spent in your home. And also thanks to everyone else I knew in the Marine Institute, it was a pleasure sharing an office, lab, coffee dock with you and you all made working there a really enjoyable experience, which I'll never forget.

To my parents, Una and Eugene, I will be forever grateful to you for all your hard work and for doing whatever you could to give us all the opportunity to get an education. To Mam for all the sandwiches, lasagne and home-baked bread you made that collectively got me through school, college and my PhD, thanks, I couldn't have done it all without them. To Dad who always appeared to show an interest in how my research was going, and who was always keen to offer his scientific advice, thanks for looking after things when I needed it most, it meant a lot to me. To my younger brother Cyril for also supporting both my sisters any I during college, it would have been very difficult for us all without your help, thanks mate. To the boys in Moate, who I have not been able to go out with as often as I would have liked in recent years due to other commitments, cheers for not slagging me too much, now that I'm done, things will be different. I might even get to watch an early kick-off match in Bo's.

## Abstract: An investigation into the existence of *Cyprinid herpesvirus 3* encoded microRNAs

*Cyprinid herpesvirus 3* (CyHV-3), a member of the *Alloherpesviridae* family, is a cause of mass mortalities in carp (*Cyprinus carpio carpio)* and koi (*Cyprinus carpio koi*) worldwide. It is difficult to detect the low levels of viral genomic DNA present in latently infected (carrier) fish. Recent evidence shows that latent herpesviral gene transcription is often confined to non-protein-coding genes including microRNAs (miRNAs). This study set out to determine if CyHV-3 encoded its own miRNAs and to investigate their suitability as biomarkers for the diagnosis of latent CyHV-3 infections. Extensive bioinformatic analysis of the CyHV-3 genome suggested the presence of non-conserved precursor-miRNA (pre-miRNA) genes. High throughput deep RNA sequencing of size-selected small RNA fractions from *in vitro* CyHV-3 lytic infections resulted in the identification of potential miRNAs, associated isomiRs and miRNA–offset RNA (moRNAs), a new functional class of small RNAs related to miRNAs. DNA microarray analysis, northern blotting and stem-loop RT-qPCR were used to definitively identify some of these as novel viral miRNAs. In all, seven CyHV-3 pre-miRNAs were identified and three of these were conclusively detected *in vivo* by stem-loop RT-qPCR. Tentative CyHV-3 mRNA targets for some of these miRNAs were also identified, thus offering insights into the possible role of these novel viral genes during infection. Potential homologues of some of these CyHV-3 pre-miRNAs were identified in the two closely related viruses CyHV-1 and -2, indicating that other members of the *Alloherpesviridae* family may also encode miRNAs. This is the first report of miRNA and miRNA-offset-RNAs being produced by a member of the *Alloherpesviridae* family. Collectively, these findings may help further our understanding of the biology of these economically important viruses and they may also serve as biomarkers for the identification of latently infected fish.

## Oral Presentations

**O. Donohoe**, K. Way, D. Stone, K. Henshilwood & D. Walls "Investigation into latency associated with Koi Herpesvirus infection in carp (Cyprinus carpio)". The 8th International Symposium on Viruses of Lower Vertebrates (ISVLV) Santiago de Compostela, Spain. April, 2010.

**O. Donohoe** & K. Way, D. "Overview of Koi Herpesvirus (KHV)". Annual Fish Health Unit Seminar, Marine Institute, Galway, October 2010.

**O. Donohoe** "Cloning of small RNAs from Koi herpesvirus infected cells" Marine Environment & Food Safety Services Research Day, Marine Institute, Galway, March 2011.

**O. Donohoe,** K. Way, K. Henshilwood, D. Stone and D. Walls "An Investigation into the existence of koi herpesvirus derived microRNAs". 15th European Association of Fish Pathologists (EAFP) conference. Split (Croatia), September, 2011.

**O. Donohoe, "**Studies on KHV latency: An investigation in the existence of KHV derived microRNAs" Lunch-time Seminar Series, Cefas Laboratories, Weymouth, UK, October 2012.

**O. Donohoe** and D. Walls "Koi Herpesvirus (KHV) Latency: "An investigation in the existence of KHV derived microRNAs". DCU School of Biotechnology's Annual Research Day, January 2013.

# Abbreviations

3’ - 3 prime

3’ UTR – 3 prime untranslated region

5' - 5 prime

5' UTR – 5 prime untranslated region

ATP - adenosine triphosphate

amol - attomoles

bp - base pairs

CCB - common carp brain cells

CSHMM - context sensitive Markov model

Ct - cycle threshold

CyHV-3 - *Cyprinid herpesvirus 3*

CyHVs - cyprinid herpesviruses

DNA - deoxyribonucleic acid

dNTP - deoxyribonucleotide triphosphate

DMSO - dimethyl sulfoxide

dsDNA - double stranded DNA

EtOH - ethanol

fmol - femtomole

FBS - fetal bovine serum

g - gravitational force

HP - hairpin

HV - herpesvirus

Hz - hertz

HCl - hydrochloric acid

kbp – kilobase pairs

LB - Luria-Bertani medium

MHP - main hairpin

$TCID_{50}$ - median tissue culture infective dose

mRNA - messenger RNA

µm - micrometer

miRNA - microRNA

µg - microgram

µL - microlitre

moRNA - microRNA-offset RNA

mA - milliamp

mg - milligrams

mL - millilitres

mM - millimolar

MEM - minimum essential media

MFE - minimum free energy

mins - minutes

M - molar

MBG - molecular biology grade

nm - nanometer

ng - nanograms

nmol - nanomoles

nt - nucleotides

ORF - open reading frame

PBS - phosphate buffered saline

pmol - picomoles

PAGE - polyacrylamide gel electrophoresis

PolyA - polyadenylation

PEG - polyethylene glycol

PCR - polymerase chain reaction

pre-miRNA - precursor miRNA

pri-miRNA - primary miRNA

RV - relative value

RT - reverse transcription

RT-qPCR - reverse transcription quantitative PCR

rpm - revolutions per minute

RNA - ribonucleic Acid

rRNA - ribosomal RNA

RISC - RNA-induced silencing complex

sec - seconds

ssRNA - single stranded RNA

NaOAc - sodium acetate

NaOH - sodium hydroxide

SHP - subsidiary hairpin

TEMED - tetramethylethylenediamine

tRNA - transfer RNA

Tris-HCl - tris(hydroxymethyl)aminomethane

U - units

V - volts

WC - window count

# Table of contents

# 1 Introduction

## 1.1 Herpesvirales

The order *Herpesvirales* is a large group of disease-causing viruses that are widely distributed in nature. Members of this order are commonly referred to as herpesviruses (HVs) and can give rise to both lytic and latent infections in animal hosts. The name comes from the Greek work "herpin" meaning "to creep" in reference to the reoccurring nature of the diseases caused by these viruses. This ability to establish long-term latent infections is undoubtedly a significant factor that has enabled HVs to become some of the most prevalent viral pathogens known. To date, over 200 have been identified in mammalian, avian, reptilian, piscine, amphibian and bivalve hosts, with each virus being highly evolved to replicate within a specific or narrow range of host species (Davison, 2002; Roizmann, 2001). These highly adapted and ubiquitous pathogens significantly affect human health (Arvin et al., 2007). Although in most cases, both long term latent infections and periodic lytic reactivation can be relatively benign (at worst, undesirable), in some circumstances human HVs can cause life threatening disease, particularly in the case of immunocompromised individuals (Ljungman, 1993). Some non-human HVs are much more pathogenic, causing highly contagious disease resulting in acute mass mortality among susceptible hosts. Viral latency presents a huge problem in this context as it allows these pathogens to persist in populations for long periods in the absence of any symptoms and results in significant economic loss upon reactivation. Such viruses are a major problem for certain sectors of the agriculture and aquaculture industries (Atkins et al., 2011; Bowland and Shewen, 2000; Pokorova et al., 2005) and for these reasons have been the main focus of research in terms of non-human HVs.

Due to substantial phylogenetic differences, *Herpesvirales* is split into 3 families: (i) *Herpesviridae,* consisting of mammalian, avian and reptilian herpesviruses (ii) *Alloherpesviridae* consisting of piscine and amphibian herpesviruses and (iii) *Malacoherpesviridae* consisting of two known bivalve herpesviruses. Despite the

significant phylogenetic distance between these families (Davison et al., 2009) all HVs retain common distinctive features that differentiates them from other viral orders (Ackermann, 2004). All HVs include a linear double-stranded DNA genome from 120-295 kbp in size (Aoki et al., 2007; Davison et al., 2009). Reports suggest that this may in fact be arranged into the shape of a torus (Zhou et al., 1999). The nucleic acid is surrounded by (and associated with) a distinct T=16 icosahedral capsid of ~100 nm in diameter consisting of different types of capsomers (major capsid proteins, triplex proteins and small capsomer interacting proteins). Capsids also feature a small portal located at one vertex for HV DNA entry and exit (Cardone et al., 2007; Rochat et al., 2011). The capsid is surrounded by a proteinaceous matrix called the tegument. This is surrounded by a viral envelope containing glycoproteins. These glycoproteins are shorter and much more numerous than those present on many other enveloped viruses (Wildy and Watson, 1962). Mature enveloped virons range from 120-260 nm with size depending on tegument thickness, which can vary depending on where the viron is in the infected cell (Falke et al., 1959) (Figure 1.1).

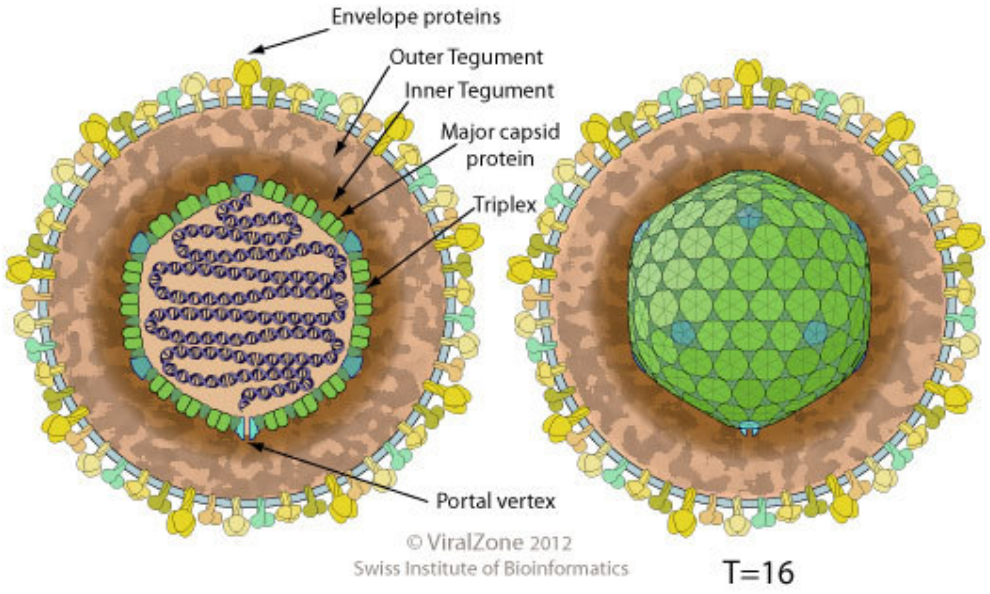



**Figure 1.1 Diagram of typical herpesvirus viron structure**
Left: cross-section of viron on the left showing all of the layers described in Section 1.1. Right: illustration of a T=16 icosahedral capsid structure. Figure from expasy.org (2013)

In addition to these distinct morphological attributes HVs also share 4 distinctive biological traits: (i) They encode a large array of genes involved in nucleic acid metabolism and synthesis e.g. thymidine kinase, thymidylate synthase, dUTPase,

ribonucleotide reductase, DNA polymerase, primase and helicase; (ii) Viral DNA replication and capsid assembly occurs in the nucleus with further processing to yield mature virons occuring in the cytoplasm; (iii) Viral replication and the release of new infective viral progeny results in destruction of the host cell; (iv) All HVs can also establish latent infections although this only occurs in specific types of cells (Roizmann, 2001).

The exact types of cells in which specific HVs can establish latent infections (tissue tropism) is important and can be indicative of their general biology. This aspect of HV biology has been best studied in members of the Herpesviridae family, particularly in human in HVs. In broad terms, these HVs display 3 distinct types of tissue tropism during latency and accordingly members of the Herpesviridae are split into 3 subfamilies on this basis. Members of the *Alphaherpesvirinae* subfamily establish latent infections primarily in sensory ganglia. Typically these HVs also have shorter lytic replication cycles than other HVs, causing fast efficient destruction of infected cells with rapid spread. It is difficult for these reasons to establish in vitro models of latency for these viruses, whereas in contrast latency is established quite readily in the case of in vivo models (Cohrs and Gilden, 2001). HVs of the *Betaherpesvirinae* subfamily are maintained in lymphoreticular cells, kidney, secretory glands, and other tissues. Notably these viruses display relatively slow replication cycles in vitro. The in vitro cytopatic effects (CPE) associated with these viruses includes a notable cellular enlargement (cytomegalia) before lysis (Roizmann 2001). Members of the *Gammaherpesvirinae* subfamily establish latency in the lymphoid tissue. In vivo, these viruses usually infect either T or B lymphocytes and are capable of transforming such cells (Küppers, 2003). In contrast to members of the *alphaherpesvirinae* and *Betaherpesvirinae* groups, the establishment of a latent infection is almost always the default pathway in vitro (Speck and Ganem, 2010). A summary of all human HVs as categorized by subfamily is given in Table 1.1.

**Table 1.1 Summary of human HVs and their tissue tropism**
Viruses that are also the prototype for their respective families are marked in bold. Table adapted from Penkert and Kalejta (2011)

| Family | Virus | Productive (Lytic) Replication | Site of Latency |
|---|---|---|---|
| A | ***Herpes Simplex 1* (HSV-1)** | Epithelial and keratinocyte | Neurons |
| | *Herpes Simplex 2* (HSV-2) | Epithelial and keratinocyte | Neurons |
| | *Varicella zoster virus* (VSV) | Epithelial, keratinocyte, T cell, sebocyte, monocyte, endothelial, Langerhans and peripheral blood mononuclear cells | Neurons |
| β | ***Human cytomegalovirus* (HCMV)** | Macrophage, dendritic, endothelial, smooth muscle, epithelial and fibroblast cells | $CD34^+$ hematopoietic stem cell and monocytes |
| | *Human herpesvirus 6* (HHV-6) | T cells | T cells |
| | *Human herpesvirus 7* (HHV-7) | T cells | T cells |
| Γ | ***Epstein–Barr virus* (EBV)** | B cell and epithelial cells | B cells |
| | *Kaposi's sarcoma-associated herpesvirus* (KSHV) | Lymphocytes | B cells |

## 1.2 Herpesvirus life cycle

Every HV studied to date has the ability to establish both lytic and latent infections and the factors that dictate what kind of infection takes place vary for each HV. These include host cell type, host cell differentiation status and host cell interaction with elements of the immune system. These factors and the general characteristics of the HV life cycle are described below.

### 1.2.1 Lytic replication

The lytic replication strategies of members of the three *Herpesviridae* subfamilies all display common characteristics (Johnson and Baines, 2011). The infection process begins with attachment to the host cell surface. Glycoproteins in the viral envelope (of which there are usually about a dozen per HV) bind to cell surface molecules. At least 3 types of conserved HV glycoproteins appear to be essential for this process in *Herpesviridae*, namely glycoproteins B, H and L (gB, gH, and gL). Additional

glycoproteins involved in cell surface interaction vary between HVs and may be the determinants of host cell specificity. On this note, gD and gO have been found to play specific roles in the *Alphaherpesvirinae* (Ligas and Johnson, 1988) and *Betaherpesvirinae* (Huber and Compton, 1998) respectively. Most HVs primarily bind weakly (and reversibly) to heparan sulphate glycosaminoglycans (GAG) on cell surfaces (Shukla and Spear, 2001), a notable exception to this is EBV which analogously binds to complement receptor 2 (CR2) on the surface of its host cells (Szakonyi et al., 2006). All HVs enter host cells through fusion between the viral envelope and the cell membrane. Although the initial nature of cell surface interaction varies between HVs, they all rely on very similar mechanisms involving viral gB, gH and gL (Heldwein and Krummenacher, 2008) of which the latter form a gH/gL dimer (Peng et al., 1998). Fusion results in the introduction of the viral capsid to the cytoplasm. Capsids are then guided towards the nucleus by latching onto a network of microtubules (Granzow et al., 1997). Upon binding to a nuclear pore, the viral genome is injected from the capsid into the nucleus. If conditions are permissible for lytic replication, the synthesis of new infective viral particles begins (alternatively a latent infection is established at this stage, with subsequent lytic infections occurring upon reactivation see Section 1.2.2). This process proceeds in a coordinated manner. During lytic replication, viral gene expression occurs as a classic regulatory cascade that is conserved across all HVs. This involves the initial expression of immediate early (IE) genes which mainly encode gene *trans*-activators. This is followed by early (E) genes, whose expression is dependent on expression of IE *trans*-activator genes. In the final phase, late (L) genes that code for structural proteins involved in assembly of new viron components i.e. capsids, tegument and envelope glycoproteins. In the case of most HVs, transcription of IE genes relies on ready made IE *trans*-activators that were incorporated into the tegument and also released into the cell upon viral entry (Penkert and Kalejta, 2011; Roizmann, 2001) (for more detail see Section 1.2.2.). Genome replication takes place in the early stages of lytic replication. This occurs by a rolling-circle mechanism that results in long head-to-tail genome concatemers (Jacob et al., 1979). Concatemerized genomes are separated by cleavage based on specific signals at the genomic termini and then packaged into newly formed capsids via the portal vertex (Cardone et al., 2007; Johnson and Baines, 2011; Rochat et al., 2011). While in the nucleus, capsids are coated with an initial tegument layer. In order to continue the process of viron

maturation, immature virons must exit the nucleus and access the cytoplasm. Nuclear egress takes place in two steps. Firstly the inner nuclear membrane (INM) is partially dismantled allowing immature virons to become enveloped in the INM as they enter the perinuclear space (Scott and O'Hare, 2001). Fusion of this temporary viral envelope with the outer nuclear layer results in de-envelopment and release of non-enveloped immature virons into the cytoplasm. Once in the cytoplasm, additional layers of tegument are acquired, after which the viron enters the Golgi and *trans*-Golgi-network by budding and becoming re-enveloped in the process. Newly enveloped virons are then transported to the cell membrane in cytoplasmic vesicles. These vesicles fuse with the plasma membrane releasing the mature enveloped virons. Several subtly different models of this maturation process have been proposed however the model described here is supported by a significant amount of independent experimental data (Johnson and Baines, 2011; Roizmann, 2001).

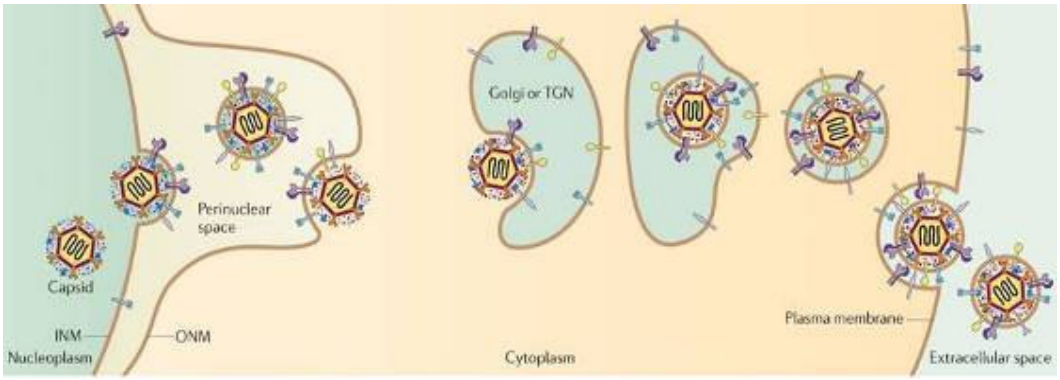



**Figure 1.2 Model of herpesvirus egress**
Figure from Johnson and Baines (2011).

### 1.2.2 Latency

Latency refers to a type of infection where the HVs go dormant or silent within the host. This results in the establishment life-long infections, from which lytic infections may periodically reactivate under permissive conditions. During this stage HV genomes are maintained in the nucleus as non-integrated closed circular episomes, however low frequency chromosomal integration has been observed in some HVs (Morissette and Flamand, 2010). During latency, genes involved in lytic replication are turned off; instead viral gene expression is limited to a specific subset of genes involved in the long-term maintenance of latency. This takes place through tight regulation of other viral genes, immune-evasion and periodic viral genome

replication and partitioning where necessary. The fact that these latent genomes still retain the capacity to undergo productive lytic replication is an important aspect of latency that makes it distinct from abortive infections. However reactivation may may never result in illness or any detectable symptoms. This occurs when a small subset of latently infected cells undergo lytic reactivation resulting in asymptomatic shedding of low levels of active viral particles (stimulating an on-going immune response) and it is something that has been observed in all 3 *Herpesviridae* subfamilies (Hadinoto et al., 2009; Ling et al., 2003; Tronstein E, 2011). Latency is typically described as having 3 phases: establishment, maintenance, and reactivation. A fourth distinctive phase occurring between latency and reactivation, termed "animation", has been proposed to describe the key event that causes the transition between latency and reactivation (Penkert and Kalejta, 2011). Some of the main molecular events corresponding to these stages of latency have been described (even if not fully understood). These processes have been best studied in the prototypical members of each of the *Herpesviridae* subfamilies (indicated in Table 1.1). The molecular basis of latency in each type of HV is slightly different, although gene regulation through remodelling of chromatin appears to be a common feature in *Herpesviridae* (Bloom et al., 2010; Sinclair, 2010; Tempera and Lieberman, 2010). Using the prototype as a representative, the following section gives a brief description of the molecular basis of latency (establishment, maintenance, animation and reactivation) in each *Herpesviridae* subfamily.

#### 1.2.2.1 *Alphaherpesvirinae: HSV-1*

It is believed that HSV-1 enters latency as a consequence of IE genes not being expressed. These genes are normally activated by the viral *trans*-activator VP16. Its presence does not rely on any initial viral gene expression as it is a component of the tegument layer (O'Hare and Goding, 1988; Stern et al., 1989). Activation of IE genes requires the entry of VP16 into the nucleus and is facilitated by the host protein HCF (La Boissiere et al., 1999). However, in neurons, HCF is sequestered in the cytoplasm due to association with the protein Zhangfei which is selectively expressed in neurons (Akhova et al., 2005). As a result, VP16 is prevented from entering the nucleus in neurons, HSV-1 IE genes are not activated and latency is established. As HSV-1 establishes latency in non-dividing cells, it does not need to

undergo genome replication or partitioning. Transcription during latency is almost exclusively limited to a no non-coding transcript called *LAT* which is responsible for the maintenance of latent infections (Stroop et al., 1984). This transcript has been associated with cell survival through the blocking of apoptosis (Perng et al., 2000). It also functions as a primary microRNA (pri-miRNA) giving rise to several mature microRNAs (miRNAs) one of which silences spurious expression of the IE gene *ICP0* (Umbach and Cullen, 2010). *In vivo*, latency may also be maintained through extra-cellular factors such as $CD8^+$ T-cell mediated inhibition of HSV-1 reactivation. $CD8^+$ T-cells surround latently infected neurons and release of cytotoxic granuales into these cells. These granales contain several proteases, one of which, granzyme B (GrB), can directly cleave the HSV-1 IE protein IPC4 which is essential for initiation of the lytic stage. *In vivo* experiments suggest that VP16 is implicated in both animation and reactivation phases and that this process is triggered by cellular stress. However, the cellular changes that occur to facilitate VP16 transcription and (presumably) disassociation of HCF from Zhangfei are unknown (Thompson et al., 2009).

#### 1.2.2.2 *Betaherpesvirinae: HCMV*

HCMV latency is established following repression of the major immediate-early enhancer-promoter (MIEP) by the host cell proteins Daxx and HDACs coupled with an enigmatic DAC-independent mechanism, which together create and maintain a transcriptionally repressive chromatin structure on the HCMV genome (Saffert et al., 2010; Woodhall et al., 2006). In the establishment of lytic infections, this intrinsic cellular defense mechanism mediated by Daxx, is normally overcome by HCMV tegument-derived IE gene trans-activator phosphoprotein-71 (pp71) (Hofmann et al., 2002; Saffert and Kalejta, 2006). However, during latency this process is inhibited by Natural Killer (NK) cells which introduce granzyme-M (GrM) into infected cells (in a similar manner to the role of $CD8^+$ T-cells during HSV-1 infections, see Section 1.2.2.1). Once introduced to infected cells, GrM directly cleaves pp71 after $Leu^{439}$, thus this process is likley to facilite both establishment and maintenance of latency (Domselaar et al., 2010). The maintenance of latency is aided by immunosuppression through the expression of a HCMV-encoded functional homologue of the cytokine interleukin-10 (IL-10) (Jenkins et al., 2008). This acts to down-regulate major

histocompatibility complex class II (MHC-II) expression on the surface of cells. Although it has been shown that this can aid the establishment and maintenance of latency, it is not essential (Cheung et al., 2009). Also, it is not known if latently infected cells divide, thus it is unclear if latent HCMV undergoes genome replication or if partitioning is required during latency. It has been suggested that miRNAs may be also involved in the maintenance of latency, but so far HCMV miRNA expression has only been investigated during lytic infections (Meshesha et al., 2012). Reactivation is known to be dependent on cellular differentiation (Hertel et al., 2003; Söderberg-Nauclér et al., 2001) and although the underlying molecular events corresponding to this are unknown, both animation and reactivation phases may take place as a result of the cessation of pp71 cleavage by GrM (Domselaar et al., 2010) allowing pp71 to counteract the effects of Daxx. However, studies have been unable to detect *de novo* expression of pp71 prior to IE gene expression during reactivation (Reeves and Sinclair, 2010).

#### 1.2.2.3 *Gammaherpesvirinae: EBV*

Here, and unlike HSV-1 and HCMV, where inhibition of IE gene expression is a key element in the establishment of latent infections, the expression of IE genes is actually required for the establishment of EBV latency. Latency is the default fate of EBV upon infection of primary B lymphocytes, transforming them into proliferating immortalized lymphoblastoid cell lines (LCLs) (Speck and Ganem, 2010). As with HSV-1 and HCMV, viral IE expression may be initiated by EBV tegument proteins (although in the case of HSV-1 and HCMV it is for the purposes of establishing lytic infections). The EBV tegument protein BNRF1 has been shown to disrupt repressive Daxx-mediated chromatin remodeling on the EBV genome allowing expression of the IE genes required for the establishment of latency (Tsai et al., 2011). Up to 11 gene products contribute to this transformation process. In particular, 2 EBV genes *BALF1* and *BHRF1* (both viral homologues of a human anti-apoptotic gene B-cell lymphoma-2), have been shown to be vital to the initiation of this process, although they are not involved in long-term maintenance (Altmann and Hammerschmidt, 2005). Z protein, a product of the IE gene *BZLF1* (and a homologue of human AP-1) has been directly implicated in the proliferation of host cells (Kalla et al., 2010) (although it is not apparent how Z avoids repression by cellular protein Zeb at this

stage). As latently infected cells proliferate, DNA replication and partitioning is required during EBV latency. Both of these processes are mediated by the EBV protein EBNA-1 (Sears et al., 2004). Non-coding RNAs also play a role in maintaining latency. For example miRNAs from the 3' UTR of the *BHRF1* locus have been shown to inhibit apoptosis and promote cell proliferation (Seto et al., 2010). *EBERs* allow latently infected cells to be resistant to PKR-mediated apoptosis (Iwakiri and Takada, 2010) and miRNAs derived from the *BART* transcript are involved in regulation of EBV Latent membrane Protein-1 (LMP-1) expression which itself is implicated in inducing cell growth (Lo et al., 2007). EBV lytic reactivation occurs when memory B cells differentiate into plasma cells as a result of immunogenic stimulation (Amon and Farrell, 2005; Laichalk and Thorley-Lawson, 2005). At a molecular level, EBV genome methylation and host cell factors appear to be two key elements in the transition from latent to lytic EBV infections. During latency, the EBV genome becomes extensively methylated (Kalla et al., 2010), with the exception of loci involved in latency (Chau and Lieberman, 2004). Despite the fact that loci associated with lytic reactivation become methylated, they can still be activated by the EBV protein Z, which preferentially activates methylated loci (Bhende et al., 2004; Kalla et al., 2010). However, in order for this to occur it requires the down-regulation of Zeb, a cellular repressor of the Z-protein (Kraus et al., 2003; Yu et al., 2007). It has been postulated that this down-regulation of Zeb happens during memory B cell differentiation (Penkert and Kalejta, 2011) and is the animation step required for reactivation of a lytic infection.

## 1.3 Herpesviruses that infect fish

Members of the *Alloherpesviridae* family infect fish and amphibians. Currently, there are 14 known members of the *Alloherpesviridae* family that infect fish. The majority of known fish HVs have been identified in economically important fish species such as salmon, carp, catfish, sturgeon, eel and cod (Table 1.2). Some of these can be highly pathogenic e.g. CyHV-3 (Pokorova et al., 2005) and IcHV-1 (Wolf and Darlington, 1971) although others are less pathogenic (depending on host age) causing tumour-like proliferations e.g. CyHV-1 (Sano et al., 1991). A phylogenetic comparison of these fish HVs using full-length terminase gene

sequences suggested the existence of two separate clades within the *Alloherpesviridae* family (Waltzek et al., 2009). On the basis of phylogenetic similarity, ten of these HVs that infect fish have been tentatively assigned to three distinct genera *Cyprinivirus*, *Ictalurivirus* and *Salmonivirus* (ICTV, 2013). Details of the fourteen *Alloherpesviridae* that infect fish, their clades, genera, hosts and effects are given in Table 1.2. In addition to these HVs, electron microscopy of diseased fish tissues has resulted in the identification of many more viruses that have been tentatively described as HVs. This is mainly on the basis of HV-like morphological features and such viruses need to be further characterized before being classified as members of the *Alloherpesviridae* family (Hanson et al (2011).

**Table 1.2 List of known HVs that infect fish**
Table adapted from Hanson et al (2011)

| Virus name | Clade | Genus | Alternative name(s) | Host(s) | Disease |
|---|---|---|---|---|---|
| *Anguillid herpesvirus 1* (AngHV-1)* | 1 | *Cyprinivirus* | HV anguillae (HVA) | Japanese eel (*Anguilla japonica*), European eel (*Anguilla Anguilla*) | Hemorrhages of skin, fins, gills, liver |
| *Cyprinid herpesvirus 1* (CyHV-1) | 1 | *Cyprinivirus* | HV cyprini, carp pox HV, carp HV (CHV) | Common carp (*Cyprinus carpio carpio*), Koi (*C. Caroio koi*) | High losses in fry-exopthalmia hemorrhages, survivors have papillomas |
| *Cyprinid herpesvirus 2* (CyHV-2) | 1 | *Cyprinivirus* | Goldfish hematopoietic necrosis virus (GFHNV) | Goldfish (*Carassius auratus*) | High mortality all ages. Necrosis of hematopoietic tissue, spleen, pancreas, intestine |
| *Cyprinid herpesvirus 3* (CyHV-3) | 1 | *Cyprinivirus* | Koi HV (KHV), carp nephritis and gill necrosis virus (CNGV) | Common carp (*C. carpio carpio*), Koi (*C. Caroio koi*) | Gill inflammation, hyperplasia, and necrosis, hematopoietic tissue necrosis, high mortality,18–26 ℃, all ages |
| *Ictalurid herpesvirus 1* (IcHV-1) | 2 | *Ictalurivirus* | Channel catfish virus (CCV), Channel catfish herpesvirus | Channel catfish (*Ictalurus punctatus*) | Kidney, liver and intestinal necrosis, hemorrhages, high mortality in young fish at above 27 ℃ |
| *Ictalurid herpesvirus 2* (IcHV-2) | 2 | *Ictalurivirus* | Ictalurus melas HV (IcmHV) | Black bullhead (*Ameiurus melas*) | Kidney necrosis, hemorrhages, high mortality all ages |
| *Acipenserid herpesvirus 1* (AciHV-1) | 2 | Not Assigned | White sturgeon HV 1 | White sturgeon (*Acipenser transmontanus*) | diffuse dermatitis, high losses in juveniles |
| *Acipenserid herpesvirus 2* (AciHV-2) | 2 | *Ictalurivirus* | White sturgeon HV 2 | White sturgeon | Epithelial hyperplasia |
| *Salmonid herpesvirus 1* (SalHV-1) | 2 | *Salmonivirus* | HV salmonis (HPV) Steelhead herpesvirus (SHV) | Rainbow trout (*Oncorhynchus mykiss*) | Mild disease low losses at 10 ℃. Adults- Virus shedding in ovarian fluid. No signs of disease. |
| *Salmonid herpesvirus 2* (SalHV-2) | 2 | *Salmonivirus* | Oncorhynchus masou virus (OMV) | Cherry salmon (*O. masou*), Coho salmon (*O. kisutch*), Sockeye salmon (*O. nerka*), chum salmon (*O. keta*), rainbow trout, | Viremia, external hemorrhages expthalmia, hepatic necrosis with high losses in young. Survivors oral papillomas, virus shed in ovaran fluid |
| *Salmonid herpesvirus 3* (SalHV-3) | 2 | *Salmonivirus* | Epizootic epitheliotropic disease virus (EEDV) | Lake trout (*Salvelinus namaycush*), lake trout × brook trout (*S. fontinalis*) hybrids | Epithelial hyperplasia, hypertrophy, hemorrhages on eye and jaw. High losses in juveniles at 6–15 ℃ |
| *Gadid herpesvirus 1* (GaHV-1) | Not Assigned | Not Assigned | Atlantic cod herpesvirus (ACHV) | Atlantic cod (*Gadus morhua*) | Hypertophy of cells in gills. High losses in adults. |
| *Pilchard herpesvirus* | Not Assigned | Not Assigned | | Australian pilchard (*Sardinops sagax*) | Acute losses with gill inflammation, epithelial hyperplasia and hypertrophy |
| *Percid herpesvirus 1* (PeHV-1) | Not Assigned | Not Assigned | HV vitreum, walleye HV | Walleye (*Stizostedion vitreum*) | diffuse epidermal hyperplasia |

*Genus classification for AngHV-1 updated from Hanson et al (2011) based on data from van Beurden et al (2012; 2010)

Members of the *Alloherpesviridae* family are only distantly related to those of the *Herpesviridae* family. In fact there is only one gene (terminase, involved in genome packaging) that is convincingly conserved among all known members of the order *Herpesvirales* (Davison, 2002; Roizmann, 2001). As more complete genomes become available for members of the *Alloherpesviridae* family, it has become ever more apparent that it represents a much more divergent family when compared to the *Herpesviridae*. In the *Herpesviridae* family there are 43 genes conserved among all members. These genes are referred to as core genes and are essential to the fundamental biological processes of these HVs (Davison et al., 2002). By contrast, only 12 core genes have been identified as being in some way conserved among members of the *Alloherpesviridae* family (Davison et al., 2012). Yet, remarkably HVs from both the *Herpesviridae* and *Alloherpesviridae* families share the same distinctive structural, biological and biochemical characteristics described in Section 1.1 (Hanson et al., 2011). All fish HVs have virons that are morphologically similar (in terms of capsid architecture, the presence of a tegument and virion envelope) to that of *Herpesviridae*. The use of mass spectrometry to catalogue the structural proteins present in mature virons of IcHV-1 (Davison and Davison, 1995), CyHV-3 (Michel et al., 2010b) and AngHV-1 (Van Beurden et al., 2011) have revealed that the types of structural proteins present in *Alloherpesviridae* virons are similar to those of the *Herpesviridae*. The complete genomes of five fish HVs have been sequenced: these include IcHV-1 (Davison, 1992), CyHV-3 (Aoki et al., 2007), AngHV-1 (Van Beurden et al., 2010), CyHV-1 and CyHV-2 (Davison et al., 2012). All of these viruses have genome structures that are the same as *Herpesviridae* Type A, consisting of 1 unique region flanked by direct terminal repeats. However, as with members of the *Herpesviridae* family, not all members of the *Alloherpesviridae* family have the same genomic structure. Although not fully sequenced yet, the genome structure of SalHV-1 is different, but does resemble a *Herpesviridae*-type D structure (Davison, 1998), consisting of two unique regions (one long and one short) with the short unique region flanked by inverted repeats (Figure 1.3).

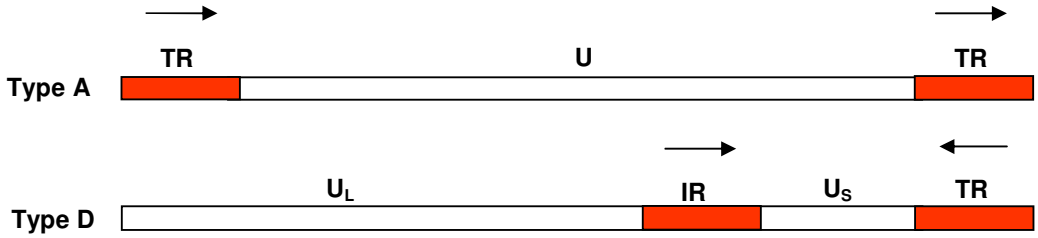



**Figure 1.3 Herpesvirus genome structures Types A and D**
Type A consists of 1 unique region (U) flanked by direct terminal repeats (TR). Type D consists of two unique regions, one long ($U_L$) and one short ($U_S$) where the short unique region is flanked by two inverted repeat regions, one internal (IR) and one terminal (TR)

For members of the *Herpesviridae* family, gene expression takes place in a temporal fashion during productive infections. To date, the gene expression kinetics of 3 members of the *Alloherpesviridae* family have been investigated, these include IcHV-1 (Huang and Hanson, 1998; Stingley and Gray, 2000), CyHV-3 (Ilouze et al., 2012a) and AngHV-1 (Van Beurden et al., 2013). These studies have all shown that gene expression in these viruses follows a similar temporal expression pattern. Like all HVs, members of the *Alloherpesviridae* family have been shown to establish latent (or at least latent-like) infections. The exact molecular basis of this has not been thoroughly investigated for any member of this family. Preliminary molecular characterisation of latency has involved the detection of viral DNA in healthy survivors of viral challenge. This has been demonstrated with CyHV-3 (Eide et al., 2011b; Gilad et al., 2004) IcHV-1 (Boyle and Blackwell, 1991; Gray et al., 1999) and CyHV-1 (Sano et al., 1993). In addition, viral reactivation has been convincingly demonstrated in the case CyHV-3 (St-Hilaire et al., 2005).

Unlike HVs of homeothermic hosts (i.e. mammals and birds) HVs infecting poikilothermic hosts (i.e. fish) experience radical temperature changes. Temperature fluctuation can cause significant changes in host fish for many reasons e.g. natural adaptive response to seasonal change, physiological stress in response to unseasonal temperatures (either too low or too high). As described in section 1.2.2 the viral-host interaction is an important factor that dictates the lifecycle of HVs. Changes in host cells or to host systems that interact with them may have important consequences with respect to viral life cycle. Viruses with narrow host ranges like HVs are highly evolved to live with their hosts. If hosts are pre-disposed to being affected by temperature changes encountered in their natural habitats, it is likely that HVs

infecting them will have life cycles that respond to this also. On this note, rather unsurprisingly, temperature is an important factor in the pathogenicity of these HVs. Some members of the *Alloherpesviridae* will only cause mortality at relatively warm temperatures, such as CyHV-3 which has a permissive water temperature range of 16-28°C (Gilad et al., 2003) and IcHV1 which has a slightly higher temperature range of 25-33 °C (Wolf and Darlington, 1971). By comparison, disease associated with other *Alloherpesviridae* members occurs at much lower temperatures e.g. SalHV-1 (5-10 °C) (Wolf et al., 1978) and CyHV1 (<15°C) (Roberts, 2011). Outside of these permissive temperature ranges, some of these viruses have been shown to persist by establishing latent infections, which subsequently reactivate when permissive temperatures are encountered again. This kind of response to temperature is something that has been best studied in CyHV3 (Gilad et al., 2004; St-Hilaire et al., 2005), although work into the underlying molecular processes involved is still in its infancy (Section 1.4.10.2)

## 1.4 *Cyprinid herpesvirus 3* (CyHV-3)

*Cyprinid herpesvirus 3* (CyHV-3) is a highly contagious virus that causes acute mass mortalities in populations of common carp (*Cyprinus carpio carpio*) and ornamental koi (*Cyprinus carpio koi*). The virus is also commonly referred to as Koi Herpesvirus (KHV) and for a period of time it was referred to as carp interstitial nephritis and gill necrosis virus (CNGV) in publications from certain research groups (Pikarsky et al., 2004; Ronen et al., 2003). During a typical outbreak, mortality rates among affected fish populations are usually over 80% (Haenen et al., 2004; Hedrick et al., 2000; Ilouze et al., 2006; Perelberg et al., 2003). Described by R.P. Hendrick as "the worst and most rapidly spreading" fish virus he had ever encountered, CyHV-3 is now recognized as a significant problem for both the common carp and koi culture industries (Pearson, 2004).

### 1.4.1 Discovery, spread and current global distribution

The first confirmed outbreaks of CyHV-3 occurred in 1998. The first of these began in May of that year on several carp farms in north-western Israel. This was followed by a separate outbreak three months later in an ornamental koi retail facility in eastern USA. In both cases, fish of all ages were affected in water temperatures of

22-23°C and the same causative agent (morphologically similar to HVs) was isolated from infected tissue samples from both locations (Hedrick et al., 2000). The vast geographical distance between the locations of these two principle outbreaks suggested that the virus had already spread to many more regions prior to its initial description by Hedrick et al. (2000). Indeed, retrospective analysis of archived samples suggested that the same virus may have been encountered in the UK as far back as 1996 (Haenen et al., 2004; Walster, 1999; Way, 2008; Way et al., 2004a) and in Germany in 1997 (Bretzinger et al., 1999).

By 2001, CyHV-3 outbreaks were reported in 90% of carp farms in Israel (Perelberg et al., 2003). Since these initial epizootics, additional outbreaks of the virus have been confirmed worldwide (Pokorova et al., 2005). The fast global spread of CyHV-3 is more than likely heavily connected with the worldwide fish trade, specifically the unregulated transport and trading of koi (Pokorova et al., 2005; Way, 2012; Whittington and Chong, 2007). As of 2012, CyHV-3 outbreaks have been confirmed in 28 countries (see Table 1.3). It is possible that outbreaks have occurred in many more regions but remain unreported or undetected. Two isolated cases of CyHV-3 infections have been confirmed in the Republic of Ireland. The first occurred in imported koi, and was restricted to a single domestic pond. The second was identified through a routine health inspection on a consignment of imported koi by the resident border inspection patrol (McCleary et al., 2011). In both cases, all fish associated with these batches were culled and the virus was contained. Effective management of these two cases ensured that no outbreak occurred. As a result, the Republic of Ireland retains a CyHV-3 free status and as such is not listed in Table 1.3.

Apart from Israel, some of the most severe outbreaks to date have occurred in Japan and Indonesia (Bondad-Reantaso et al., 2007; Lio-Po, 2011; Sano et al., 2004; Way, 2008). The spread of CyHV-3 has had major economic impact with devastating losses reported in both intensive and extensive carp and koi culture facilities. However cases have not been confined to aquaculture, with many outbreaks also reported in wild populations (Denham, 2003; Garver et al., 2010; Grimmett et al., 2006; Kempter and Bergmann, 2007; Uchii et al., 2009) and fisheries (Taylor et al., 2010; Way, 2008). The emergence of CyHV-3 is an important issue for cyprinid

aquaculture. Consequently, the World Organisation for Animal Health or Office International des Epizooties (OIE), as it is more commonly refferred to as, has listed CyHV-3 as a notifiable disease and outbreaks must be reported to this organisation. Within the European Union, CyHV-3 is now listed as a non-exotic disease. Under the Aquatic Animal Health Directive (2006/88/EC) each EU member state is required to follow the appropriate diagnostic methods and ensure appropriate controls are in place to minimise the spread of CyHV-3 to, or within their respective jurisdictions. If the disease status of a country is unknown, extensive surveillance programs must be implemented before it can be declared free of CyHV-3. As of yet there are no vaccines approved for use within the EU. Currently, attempts to monitor for CyHV-3 involve the use of several highly sensitive and robust PCR (Bercovier et al., 2005; Gilad et al., 2004) and serological-based diagnostic techniques (Adkison et al., 2005) in order to confirm the presence of KHV during or after high-mortality outbreak, allowing appropriate action to be taken.

**Table 1.3 List of countries with confirmed CyHV-3 outbreaks**
Based on a combination of information available in Pokorova et al. (2005) and Way (2012)

| Europe | Asia | America | Africa |
|---|---|---|---|
| Austria | Indonesia | USA | South Africa |
| Belgium | Israel | Canada | |
| Denmark | Japan | | |
| England | Taiwan | | |
| France | Thailand | | |
| Germany | China | | |
| Italy | South Korea | | |
| Luxemburg | Malaysia | | |
| The Netherlands | Singapore | | |
| Switzerland | | | |
| Poland | | | |
| Czech Republic | | | |
| Romania | | | |
| Slovenia | | | |
| Spain | | | |
| Sweden | | | |

### 1.4.2 Hosts

#### 1.4.2.1 *Susceptible fish species and economic impact*

CyHV-3 primarily causes acute mass mortality among common carp and koi. These are two very economically important fish species. The popular hobby of keeping and trading high value ornamental koi represents a significant portion of a global ornamental fish industry (Balon, 1995; David et al., 2004), which is itself estimated to be indirectly worth around US$15 billion (fao.org, 2013a). Common carp (*Cyprinus carpio carpio*) is one of the most widely cultivated fish species for human consumption, with an estimated 3.4 million metric tonnes being produced annually. As of 2010, this industry was valued at over US$4.5 billion (fao.org, 2010). As of 2002, common carp represented nearly 14% of global fresh water aquaculture output, with most (~70%) production occurring in China (fao.org, 2013b). In many Asian regions, farmed carp represents an important source of protein for many low-income families. Carp aquaculture provides both direct and indirect employment in the same regions. As with other farmed fish species, threats to the long-term feasibility of carp aquaculture are becoming ever more relevant in the context of global food security (Liao and Chao, 2009).

Ghost carp (hybrids of common carp and koi) are also susceptible to the disease. Hybrids of common carp and koi with non-susceptible cyprinid species have shown mixed results in terms of resistance. Goldfish-koi hybrids show reduced mortality rates (35-42%) compared to crucian carp (*Carassius carassius*)-koi hybrids (100%) (Bergmann et al., 2010). Furthermore, Goldfish (*Carassius auratus* )-common carp hybrids were shown to be quite resistant to CyHV-3 infection with mortality rates of ~5% under permissive conditions. PCR showed that ~50% of them remained positive for CyHV-3 DNA for up to 25 days following infection, however it is not clear if viral replication was occurring in these survivors (Hedrick et al., 2006).

#### 1.4.2.2 *Vector Species*

It is possible that other aquatic species, that are not susceptible to CyHV-3, play important roles in transmission.

#### 1.4.2.2.1 Fish vector species

Goldfish co-habited with CyHV-3 infected koi have been shown to be resistant to the virus yet test positive for CyHV-3 DNA (Sadler et al., 2008). In addition to goldfish, CyHV-3 DNA has also been detected in other cyprinid species such as ide (*Leuciscus idus*) and grass carp (*Ctenopharyngodon idella*) (Bergmann et al., 2009) and non-cyprinid species such as ornamental catfish (*Ancistrus sp.*) (Bergmann et al., 2009), Russian sturgeon (*Acipenser gueldenstaedtii*) and Atlantic sturgeon (*A. oxyrinchus*) (Kempter et al., 2009). In addition the transmission of CyHV-3 to naïve carp (without the development of disease) from several wild species (gudgeon, pike, rudd, ruffe, and tench) previously co-habited with suspected latently infected carp has been observed (Fabian et al., 2013).

#### 1.4.2.2.2 Invertebrate vector species

It is known that filter feeding aquatic invertebrates can accumulate bacterial and viral particles without suffering any adverse effects. In some cases viral particles accumulated in this way remain infectious for long periods e.g. norovirus in oysters (Lowther et al., 2008). CyHV-3 DNA has been detected in swan mussels, (*Anodonta cygnea*) in Poland (Kielpinski et al., 2010) and in lake plankton, specifically *Rotifera sp.* in Japan (Minamoto et al., 2011). Both of these observations suggest that natural mechanisms exist within habitats that may result in the concentration of virus before contact with carp or koi. It is unclear how concentration of the virus in these vector species affects the stability of viral particles outside the host, how long they remain infective or if vector species facilitate viral replication.

### 1.4.3 Clinical signs and mortality

During a typical outbreak, morbidity among infected populations is usually ~100%. Clinically infected fish exhibit general lethargy and appetite loss. At later stages fish may also show loss of equilibrium. One of the principal external signs of clinical infection is swollen, pale or patchy gills showing severe necrotic damage. Other variable non-specific external signs can include skin lesions, mucus hypersecretion, enophthalmia (sunken eyes), hyperemia (increased blood flow to specific tissues), hemorrhaging of the epidermis and fins and fin erosion. Examples of these are

displayed in Figure 1.4. Internally, severe nephritis can be one of the most prominent internal pathological changes. Organs such as liver, spleen, gastro-intestinal tract show necrosis, numerous microscopic lesions and other forms of histopathological damage (Figure 1.5) thus demonstrating the systemic nature of CyHV-3 infection (Hedrick et al., 2005, 2000; Miyazaki et al., 2008a; Walster, 1999). The course of disease is quite rapid and mortality can occur in as little as 7-12 days following exposure under ideal conditions (Pikarsky et al., 2004). Under such optimum conditions mortality rates are usually >80% and 100% mortality is not uncommon (Haenen et al., 2004; Hedrick et al., 2000; Ilouze et al., 2006; Perelberg et al., 2003).

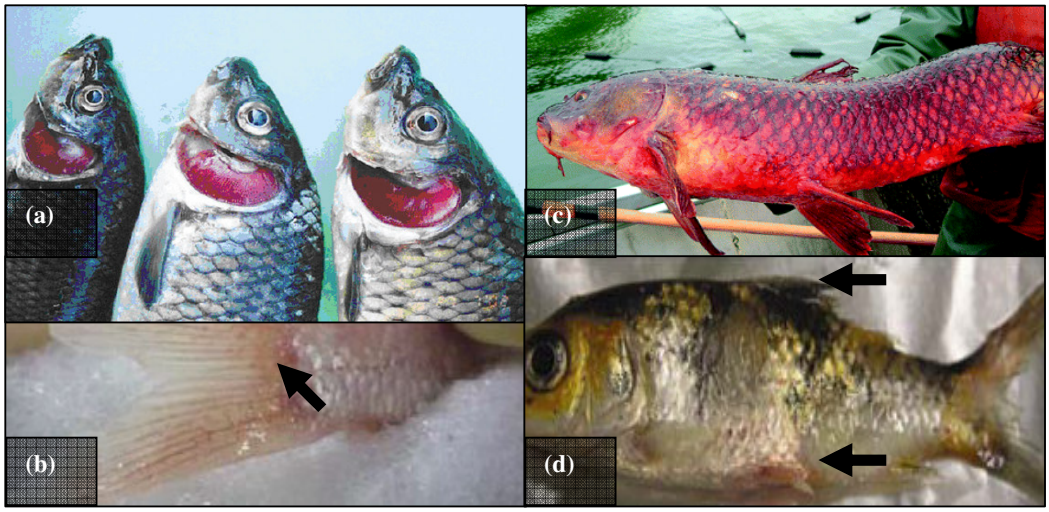



**Figure 1.4 Examples of some common clinical signs associated with disease caused by CyHV-3**
(a) Gills displaying varying degrees of necrosis Lio-Po (2011), (b) hyperemia occurring at the base of the caudal fin from Michel et al. (2010a) (c) widespread epidermal hemorrhageing from Garver et al (2010), (d) herpetic lesions and erosion of fins Michel et al. (2010a).

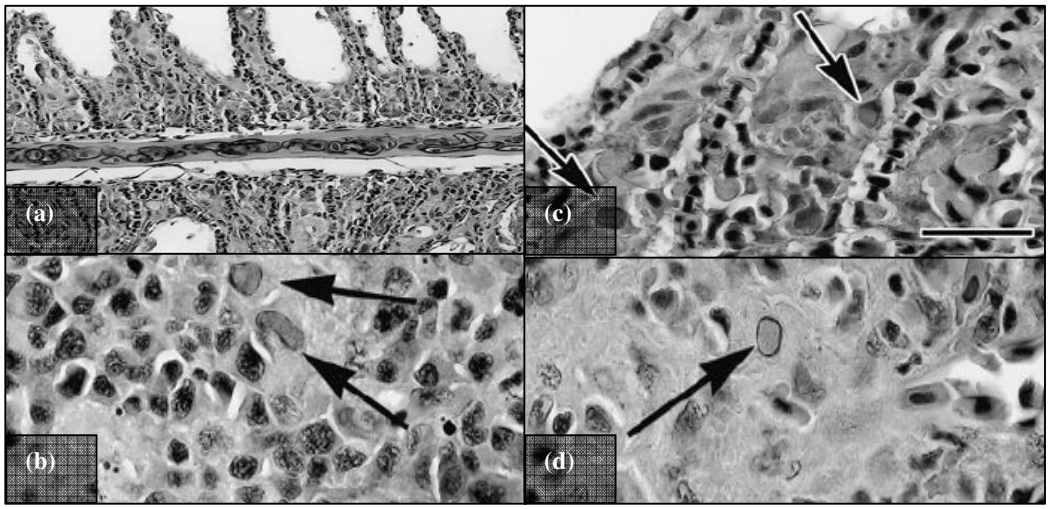



**Figure 1.5 Examples of visible pathology from microscopic lesions**
(a) Hyperplasia in gill, (b) nuclei with intranuclear inclusions in kidney, (c) nuclei with intranuclear inclusions in branchial epithelium, (d) chromatin margination of spleen stromal cell. Images are from Hedrick et al. (2000).

### 1.4.4 Factors in the emergence of disease

The disease caused by CyHV-3 (mostly referred to as KHV-disease or KHVD) is highly virulent and contagious, resulting in acute mass mortality amongst the host species. Despite its virulence, high mortality outbreaks can only occur within a permissive water temperature range of 16-28$^{o}$C with an optimum range of 22-25$^{o}$C (Gilad et al., 2004, 2003; Hedrick et al., 2000; Perelberg et al., 2003; St-Hilaire et al., 2005; Way, 2007). As a result, in some climates clinical signs of the disease and mass mortality only occur in spring and summer. This also coincides with the time of year when fish gather for spawning, which occurs when water temperatures reach 17-18°C (fao.org, 2013b). Outside the permissive temperature, symptoms appear to subside and mortality decreases. Despite this, viral DNA can still be detected while fish are kept at low temperatures (Gilad et al., 2004; St-Hilaire et al., 2005). Furthermore, clinical signs and mortality can resume once fish are exposed to permissive temperatures again (St-Hilaire et al., 2005). CyHV-3 can cause disease in both juvenile and adult fish (Bretzinger et al., 1999; Hedrick et al., 2000; Sano et al., 2004; Terhune et al., 2004). Carp have been shown to remain unaffected while at the larval sage (Ito et al., 2007). However, they do become susceptible once they mature past this stage to become fry. In fact, co-habitation studies have shown that at this stage (fry: 1-3 months, 2.5-6 g) carp are more susceptible to disease caused by CyHV-3 than any other stage of their development (Perelberg et al., 2003). Other

factors such as population density and stress (due to infections with other pathogens, poor water quality, transport etc.) may also contribute towards the development of disease associated with CyHV-3.

### 1.4.5 Transmission

The main mode of transmission is horizontal with virulent virus being shed from infected fish, through faeces, urine and gill and skin mucous (Pokorova et al., 2005). Fish become infected through encountering waterborne virus or though direct physical contact with infected fish. Virus shed via faeces is more likely to accumulate in the sediment. Plankton vectors that concentrate virus may also settle in the sediment when they become inactive or die. This accumulation of virus in the sediment can result in there being up to 1000 times more viral load in the sediment when compared to water in the same location (Honjo et al., 2012) and may result in fish becoming infected while stirring or burrowing in the sediment as part of their normal feeding behaviour. Also, the indiscriminate filter feeding mechanisms of bivalves (e.g. swan muscles) may concentrate both viral particles and plankton vectors themselves (Minamoto et al., 2011). Thus carp may become infected through feeding on such vector species.

Virus shed into water remains infective in water for at least 4 hours at permissive temperatures of 22-25°C, explaining the highly contagious nature of the disease (Perelberg et al., 2003). At lower temperatures of 15°C, viral infectivity can persist in river water for three days and for at least seven days in sterile river water. This suggests that the local microbial ecosystem has an effect on viral stability and transmission (Shimizu et al., 2006). It is also possible that other factors may act to keep the virus stable in the environment. Active viral particles may be preserved in faeces for long periods of time (Dishon et al., 2005) meaning that it is plausible that faecal-borne viral particles that have accumulated in sediment are preserved for long periods after outbreaks during non-permissive temperatures. The extent to which virus that has accumulated in sediment or bivalve vectors contributes to overall transmission has not been established. It is likely that most transmission occurs through direct contact with infected fish or water-borne virus. No cases of vertical

transmission from infected parents to eggs have been observed but such events cannot be ruled out.

Transmission of CyHV-3 from clinically infected fish during outbreaks may not be the only way in which the virus is spread. In one case, viral DNA was detected in a Japanese river up to 4 months prior to an outbreak while temperatures (9-11°C) were far too low for the virus to cause disease (Haramoto et al., 2007), however the infectivity of the virus detected is unknown. The presence of viral DNA in river water long before an outbreak during non-permissive temperatures may occur as a result of the virus being shed by other fish vector species (Section 1.4.2.2 ) or by latently infected carp that may periodically shed low levels of virus, as is characteristic of latency with other HVs (Section 1.2.2). From studies of carp populations in Japan, researchers have suggested that the presence of anti-CyHV-3 antibody in healthy carp long after disease outbreaks (Uchii et al., 2009), may be caused by such periodic low-level reactivation of latent virus continuously boosting the animal's immune response. This may also result in these fish acting as reservoirs for infectious virus particles at low non-permissive temperatures (<16°C) without developing disease themselves. At such temperatures, naive carp that subsequently become infected would not immediately develop disease either.

The same study by Uchii et al. (2009) also provided an interesting insight into factors affecting the transmission and persistence of CyHV-3 in the wild. In this study, the authors investigated the distribution of CyHV-3 among wild carp populations in a Japanese lake two years after a high mortality CyHV-3 outbreak (two sporadic low level mortality outbreaks had occurred in the two intervening years). It was found that CyHV-3 DNA could be detected in only 6% of small fish (≤300 mm) but at 31% of fish greater than 300 mm. Also, a significant number of carp tested positive for antibodies to CyHV-3. Of these antibody-positive fish, 45% tested negative for viral DNA by PCR suggesting that viral load was reduced beyond the limit of detection for the assay used. Furthermore, 11% of fish (one of which was <300 mm) tested positive by PCR but showed no sign of antibodies. This suggested that it was not long since these fish were infected as it had previously been shown in carp that the antibody response to CyHV-3 only begins to rise to detectable levels after 14 days following exposure, reaching a peak at 21 days after which it plateaus for at least a

year (Adkison et al., 2005). In summary, fish greater than 300 mm were more likely to be: (i) infected (possibly latently) on the basis of testing positive for CyHV-3 DNA and anti-CyHV-3 antibody and (ii) recently infected on the basis of testing positive for CyHV-3 DNA and negative for anti-CyHV-3 antibody. A possible reason for greater prevalence among fish larger than 300 mm and in particular among fish greater than 400 mm (which accounted for 69% of those positive for anti-CyHV-3 antibody) may be because of spawning activity. Carp first start to spawn at a size of ~360 mm. this suggests that, in the wild, it is generally sexually mature carp that are more likely to become infected because of more frequent interactions with a high number of other individuals.

### 1.4.6 Viral entry and replication

The virus may enter its host through many routes. Early studies suggested that gills may be the initial point of entry into the hosts (Gilad et al., 2004; Miyazaki et al., 2008a; Pikarsky et al., 2004). This was based on a the high levels of CyHV-3 found in gills at the early stages of infection and the fact that the disease is mainly characterised by the presence of gill lesions and gill necrosis. However later studies showed that the virus can easily enter hosts via the skin and that this may indeed be the predominant portal of entry and that the virus can also replicate in the skin (Costes et al., 2009). Paradoxically, further studies showed that the epidermal mucus layer prevents viral entry through the skin (Raj et al., 2011) but suggested that localized inconsistencies or loss of mucus layers may facilitate viral entry. Recent evidence suggests that fish may also become infected following feeding on contaminated material via infection of the pharyngeal periodontal mucosa (Fournier et al., 2012).

Once in its host cell, electron microscopy has revealed that CyHV-3 replication takes place in a similar manner to other HVs and that it involves two distinct envelopments (Miwa et al., 2007; Miyazaki et al., 2008a). Viron replication and maturation is consistent with the general HV model (described earlier in Section 1.2.1.). Infected cells generally show intranuclear inclusion bodies, marginal hyperchromatosis and an increase in filamentous nucleoproteins (Miyazaki et al., 2008a)

CyHV-3 can be detected by PCR in various organs (gill, liver, gut, spleen, kidney and brain) and mucus as early as one day after initial exposure (Gilad et al., 2004). Interestingly, a recent study showed that CyHV-3 replication in the intestine upregulates the expression of claudin-2, -3c, -11, and -23 (Syakuri et al., 2013). The virus replicates most efficiently in the kidneys where it induces the most prominent pathological changes. These changes can be observed in as little as 2 days following initial exposure (Pikarsky et al., 2004). The quick systematic spread of the virus makes it difficult to resolve a clear picture of the sequential spread from the initial sites of infection. The rapid transfer to the kidneys and subsequent localisation of CyHV-3 in white blood cells may then allow the virus to quickly spread to other organs. Rapid replication in the gills and release into the water may promote the transmission of CyHV-3 in a fashion that is analogous to that of mammalian respiratory viruses (Ilouze et al., 2011; Pikarsky et al., 2004). It has been suggested that the pathological damage sustained to the gill, kidneys and gut during infection results in loss of osmoregulatory functions in these organs, ultimately leading to death (Gilad et al., 2004).

### 1.4.7 Diagnostic methods

#### 1.4.7.1 *Cell culture*

CyHV-3 can replicate in several established cell lines. Cell lines from host species include, koi fin cells (KF-1) (Hedrick et al., 2000), KF-101 (Lin et al., 2013) and common carp brain cells (CCB) (Neukirch et al., 1999). Interestingly, CyHV-3 has been shown to replicate in cell lines derived from resistant cyprinid species such as silver carp (*Hypophthalmichthys molitrix*) (Tol/FL cells) and Goldfish (Au cells) (Davidovich et al., 2007). However, other cell lines derived from common carp have been shown to be somewhat restrictive to CyHV-3 growth. Epithelioma papulosum cyprinid cells (EPC cells) were initially reported to be resistant to CyHV-3 (Davidovich et al., 2007). Later the same group reported that CyHV-3 does replicate in these cells but that this only occurs three weeks post-inoculation and leads only to low titres of virus (their unpublished results; Ilouze et al., 2011). These results indicate that CyHV-3 can replicate in host species other than carp and koi and

suggest that it is possible for CyHV-3 to infect other cyprinid hosts. Based on the restrictiveness of EPCs, additional factors other than the presence of appropriate host cell receptors may dictate whether or not CyHV-3 will replicate in specific cells. Such additional factors may include the types of innate immune responses in cells. While *in vitro* culture of CyHV-3 on susceptible cell lines presents a useful way to study the virus, these methods are not sensitive or reliable enough for routine use in screening for or diagnosis of CyHV-3 in fish tissue. In addition, virus cultivation cannot be used as a confirmatory method (Haenen et al., 2004; Way, 2012).

#### 1.4.7.2 *Antibody detection*

Anti-CyHV-3 antibodies produced by hosts in response to CyHV-3 challenge can be detected using an enzyme linked immunosorbent assay (ELISA) developed by Adkison et al (2005). As expected, the ability to produce high levels of anti-CyHV-3 antibodies does result in protection against CyHV-3 infection, especially fish that have survived CyHV-3 exposure (Perelberg et al., 2008). This is quite useful as it allows the identification of healthy fish that have been exposed to CyHV-3 at some point and which have more than likely survived as carriers. Due to insufficient knowledge of the immune response in carp (especially in relation to how long anti-CyHV-3 antibodies remain at detectable levels), it is difficult to definitively evaluate the potential usefulness of targeting anti-CyHV-3 antibodies as part of routine screening programs for CyHV-3 infection. There is some preliminary data available regarding this: for example it has been shown that in 15–50g fish the anti-CyHV-3 antibody levels remain detectable for 280 days (Perelberg et al., 2008) and in larger fish ranging from 100-200g anti-CyHV-3 antibodies were shown to remain at detectable levels for at least a year after exposure (Adkison et al., 2005). However, as there are many more variables (water temperature, length of time exposed to specific temperature, viral titre used etc) that will affect the magnitude and length of immune response, it is hard to directly compare results from different controlled experiments. The analysis of antibody responses to CyHV-3 in wild populations has also given mixed results, but these results suggest that a long-term response is possible. Carp tested in fisheries in the UK were still positive for anti-CyHV-3 antibody up to one year following large outbreaks (Taylor et al., 2010) and wild carp in Japan have tested positive up to two years after outbreaks (Uchii et al., 2009). However, in all

cases it is not known if the anti-CyHV-3 antibodies detected are present as a result of the initial exposure to CyHV-3 or the result of a later immune response to low level CyHV-3 reactivation and shedding. In cases where anti-CyHV-3 antibodies have declined beyond detection since initial viral exposure, fish still remain resistant to the virus, most likely due to the rapid response of memory B-cells. For this reason the absence of detectable antibodies does not mean that fish have not been previously exposed or are not harbouring latent virus. In this way, testing for the presence of anti-CyHV-3 antibodies might result in false negatives. In addition, due to the presence of shared antigens, anti-CyHV-3 antibodies are known to cross-react with anti-CyHV-1 antibodies (Adkison et al., 2005) thus potentially leading to false positive results.

#### 1.4.7.3 *Antigen Detection*

ELISA methods to detect CyHV-3 antigens have also been developed (Dishon et al., 2005; Pikarsky et al., 2004). These assays may be suitable for detecting of CyHV-3 during clinical infections when there are high levels of virus but may not be sensitive enough to enable screening for carriers in healthy populations (Way, 2012). Furthermore, if no viral proteins are expressed during latency (or if protein expression is restricted), antigen detection methods would not be suitable.

#### 1.4.7.4 *Polymerase Chain reaction (PCR)*

There have been several conventional and real-time PCR methods used to detect CyHV-3. The most widely used and sensitive conventional PCR methods include those developed by Bercovier et al. (2005) and Yuasa et al (2005). The most widely used real-time PCR method is the method developed by Gilad et al. (2004). These are the most sensitive methods used to diagnose and confirm CyHV-3 clinical infections. Because of their higher sensitivity they are also ideal for screening healthy populations for latent carriers.

#### 1.4.7.5 *Most reliable method for screening for CyHV-3 carriers*

It appears that assaying for anti-CyHV-3 antibodies or CyHV-3 DNA are the most suitable methods for screening healthy fish in order to detect potential latent carriers.

However it is not clear which of these approaches is most appropriate. There have been cases where anti-CyHV-3 antibodies have been detected in healthy fish in the absence of any CyHV-3 DNA (Uchii et al., 2009). Conversely there have been cases where CyHV-3 DNA has been detected in the absence of any anti-CyHV-3 antibodies (Eide et al., 2011b; Uchii et al., 2009). Ideally both tests should be used in unison, however the extra advantages offered by PCR in terms of reduced labour, higher sensitivity, increased specificity and the ability to verify results with southern blotting and/or DNA sequencing has resulted in PCR being the most popular approach. It is for these reasons also that PCR it is the method recommended by the OIE when screening for CyHV-3 carriers (Way, 2012).

### 1.4.8 Control

#### 1.4.8.1 *Vaccination*

There has been some work done regarding the vaccination of carp against CyHV-3-related disease. The first approach to this involved creating "naturally resistant" fish. This was done by first exposing carp to CyHV-3 at a permissive temperature for 3-5 days and then moving them to a non-permissive temperature at or above 30$^{o}$C for 30 days to promote survival. When these fish were subsequently transferred to open-air ponds (where water temperatures fell within the permissive temperature range) and re-challenged with CyHV-3 the mortality rate was reduced to 39% compared to a control group (kept at 22$^{o}$C and not exposed to the virus) which showed 82% mortality. These "naturally resistant" fish also showed high levels of anti-CyHV-3 antibody and remained resistant for periods of at least 6-12 months. In addition the fish showed no sign of viral DNA or infectious virus. Despite its simplicity and effectiveness, this approach still resulted in ~40% mortality. The high cost (especially in cold climates) involved in raising water temperatures to 30$^{o}$C for long periods of time and the fact that it increases the risk of other disease makes this approach quite impractical to execute on a routine basis. In addition, there is a possibility these fish could become latently infected as a result of the initial exposure and that this may reactivate at a later date during permissive temperatures. No latent CyHV-3 has not been detected in such fish to date, although this possibility needs to be investigated further (Ronen et al., 2003).

Other studies have focused on generating live attenuated CyHV-3 strains for the purpose of inducing host immunity (Perelberg et al., 2005). In these experiments, it was found that if carp were exposed to the attenuated strain (by bathing) and kept at a permissive temperature, this resulted in near 100% resistance to subsequent challenge with non-attenuated strains. This method of vaccination conferred resistance to CyHV-3 for up to 8 months (Perelberg et al., 2008). Other approaches such as oral vaccines have also been developed. These consist of formalin-inactivated CyHV-3 (at that stage, CyHV-3 antigens) entrapped in liposomes. This approach has been shown to work reasonably well resulting in lower mortality rates ranging from 23% to 35% (Miyazaki et al., 2008b; Yasumoto et al., 2006). More recently, work has also commenced on a DNA vaccine using a CyHV-3 glycoprotein to stimulate an immune response. The use of this vaccine has been reported to result in a reduced mortality rate of 3.3% (Nuryati et al., 2012).

#### 1.4.8.2 *Resistance breeding*

As an alternative way to control CyHV-3, some studies have focused on the selection of CyHV-3-resistant carp strains for crossbreeding with CyHV-3-sensitive domesticated strains in order to evaluate the resistance of the resulting progeny (Shapira et al., 2005; Zak et al., 2007). Interestingly, Shapira et al (2005) found that crossing the domesticated strain Dor-70 (D) with a the wild-type Sassan (S) resulted in progeny that showed a 60% survival rate when challenged with CyHV-3. Other research into genetic factors likely to influence resistance to CyHV-3 include the identification of carp SNPs associated with the innate immune response (Kongchum et al., 2010) and the identification of different polymorphisms in the carp major histocompatibility (MH) class II B gene that are associated with both susceptibility and resistance to CyHV-3 (Rakus et al., 2009).

### 1.4.9 Characterisation of CyHV-3

#### 1.4.9.1 *Morphology*

Initial investigations of KHV outbreaks revealed a virus with characteristics consistent with HVs. Similar morphological characteristics included an inner capsid displaying icosahedral symmetry surrounded by a thread-like tegument and an additional host derived viral envelope on mature virons, giving an overall diameter of 170-230nm (Hedrick et al., 2000). Further morphological characterization supported these findings and showed the presence of two distinct envelopments (Miwa et al., 2007) (Figure 1.6) consistent with morphology and replication of other HVs (described earlier in Section 1.1 and 1.2.1, respectively).

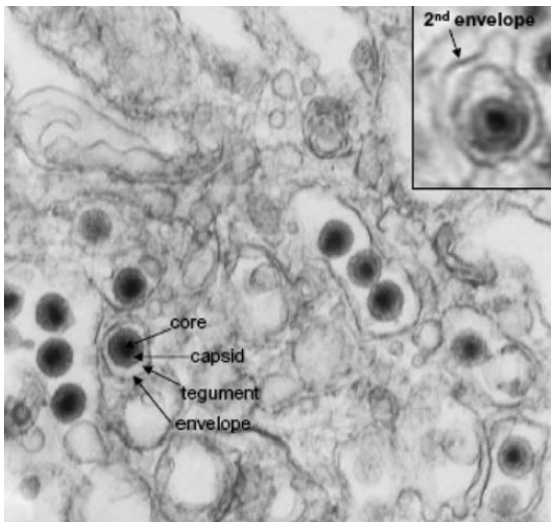



**Figure 1.6 Electron micrograph of mature CyHV-3 virons in an infected carp cell**
The main features of the mature viron (the core, capsid, tegument and envelope) are indicated in the bottom left of the image. An example of a mature viron with a second envelope can be seen in the top right corner of the image. This is more than likely to be a cytoplasmic vesicle involved in the transport of mature virons to the cell membrane for release. This image was taken from Ilouze et al. (2011)

#### 1.4.9.2 *Phylogenetic Analysis*

Early phylogenetic analysis based on limited sequence data showed little evidence that CyHV-3 was related to mammalian or avian herpesviruses (HVs). Strangely, amino acid sequences for some putative genes showed more similarity to genes from other distantly related dsDNA viruses from the *Poxviridae*, *Adenoviridae*

and *Baculoviridae* families (Hutoran et al., 2005). Due to a perceived lack of compelling phylogenetic evidence to suggest its inclusion as a member of the *Herpesviridae* family, the virus was sometimes just referred to as carp interstitial nephritis and gill necrosis virus or CNGV (based on the principal clinical signs) rather than Koi herpesvirus (KHV) (as it was more commonly known at the time). Subsequently, sequencing of the most phylogenetically relevant genes (terminase sub-unit and DNA polymerase) from this virus and other lower vertebrate HVs, (particularly CyHV-1 and CyHV-2) allowed more meaningful phylogenetic analyses be carried out. This showed that the virus was not closely related to mammalian and avian HVs but was more closely related to the other known Cyprinid HVs (CyHV-1 and CyHV-2) and (albeit much more distantly) to other lower vertebrate viruses such as IcHV-1 and ranid HV-1 (RaHV-1) (Waltzek et al., 2005). Thus, the classification issue was resolved and it was confirmed that the initial description of the virus by Hedrick *et al.* (2000) as a member of the *Herpesviridae* family (as it was known then before re-structuring, see Section 1.3 for details) was correct, prompting the proposal for its inclusion in this family under the formal name CyHV-3 (Waltzek et al., 2005; Way et al., 2004b). Subsequent sequencing of the entire CyHV-3 genome (Aoki et al., 2007) and other genomes such as CyHV-1, CyHV-2 (Davison et al., 2012) and Ang-HV1 (Van Beurden et al., 2010) has since confirmed the initial findings that CyHV-1, -2 and -3 are all closely related and in addition that all of them are more closely related to AngHV-1 than to any other member of the *Alloherpesviridae* family.

Hutoran et al (2005) originally observed that CyHV-3 contained some genes that were more closely related to distantly related dsDNA viruses outside the order *Herpesvirales*. Further analysis of the protein-coding sequences of several genes such as ribonucleotide reductase (RNR) thymidine kinase (TK) showed that these were more similar to the same genes present in the *Poxviridae* family (Ilouze et al., 2006). A thymidylate monophosphate kinase (TmpK) gene encoded by CyHV-3 but not found in any other HVs was also most closely related to a TmpK gene found in the *Poxviridae* family. In addition, CyHV-3 was also found to encode a form of Serpin (or serine protease inhibitor) that is very similar to B22R which is found exclusively in members of the *Poxviridae* family. It was speculated that the presence of these genes in the CyHV-3 genome may be the result of gene acquisition through

horizontal gene transfer. It us unclear how thi scould have occurred as both types of viruses replicate within different cellular compartments, i.e Poxviruses replicate in the cytoplasm and HVs replicate in the nucleus. However, in rare occasions where host cells may be co-infected with both types of virus, breakdown of the nuclear membrane (during mytosis or due to the effects of viral infecton) may bring these different viruses into close contact during replication. These results have also prompted speculation that CyHV-3 may be an ancestor of a particularly ancient viral family from which several other viral families evolved, including the *Herpesvirales* order and the *Poxviridae* family (Ilouze et al., 2006). Interestingly, sequencing of the genomes for CyHV-1 and CyHV- 2 revealed that of the non-HV genes, TmpK (ORF140 in CyHV-3) was found in both these viruses but B22R (ORF139 in CyHV-3) was not present in CyHV-1 (Davison et al., 2012).

#### 1.4.9.3 *Genome and genes*

Sequencing of the CyHV-3 genome also confirmed earlier estimates of the large genome size (Hutoran et al., 2005), showing that it was 295 kb in size and therefore larger than any other known HV genome (Aoki et al., 2007). This large genome size makes CyHV-3 an ideal model for mutagenesis of large DNA viruses and this has been demonstrated by Costes et al (2008). The genome contains one unique region flanked by direct terminal repeats (each ~22 kb in length), resembling a Type-A structure (Section 1.3). Aoki et al (2007) also predicted the existence of 156 protein-coding genes, which includes 5 gene families (Figure 1.7), all of which have now been shown to be transcribed into mRNA (Ilouze et al., 2012a). Of these 156 genes, 120 have orthologues in CyHV-1 and CyHV-2 and 55 have orthologues in AngHV-1. In addition, twelve of them (referred to as core genes) are conserved across all members of the *Alloherpesviridae* family (Figure 1.7) (Davison et al., 2012). Interestingly all three CyHVs contain telomere-like repeats (based on the element TTAGGG). These are located close to the ends of the terminal repeats on the three viral genomes. Telomere-like repeats are also present in other HVs such as Mareks disease virus (MDV-1) and HHV-6A, HHV-6B and HHV-7 and may be involved in the integration of these viral genomes into the telomere regions of host genomes (Arbuckle et al., 2010; Kaufer et al., 2011; Morissette and Flamand, 2010).

During lytic replication, CyHV-3 genes are expressed in a coordinated manner similar to other herpesviruses (Section 1.2.1). Analysis of CyHV-3 gene expression in the early stages of *in vitro* infections has resulted in the identification of 15 immediate early (IE) genes, 112 early (E) genes and 22 late (L) genes. Seven of the 156 ORFs remain unclassified and require additional analysis to determine at what stage of the process they are expressed (Ilouze et al., 2012a). Using liquid chromatography tandem mass spectrometry-based approaches to analyze the protein composition of mature CyHV-3 virons, Michel et al (2010b) identified 40 CyHV-3 genes that encode structural proteins. Using bioinformatic methods, 17 of these structural proteins were sub-classified into 3 capsid, 2 tegument and 13 envelope proteins. The sequences of 2 of these envelope proteins (ORF25 and ORF65) and a predicted CyHV-3 membrane protein (ORF 116) have been compared between four CyHV-3 strains (U, I, J and K) and found to contain several significant deletions and insertions in some strains, most notably in the K strain (Han et al., 2013). Some individual CyHV-3 genes and gene products have been characterised in more detail. These include TK (Bercovier et al., 2005; Costes et al., 2008), viral IL-10 homologue (Sunarto et al., 2012), ORF112 (Tomé et al., 2013) and ORF81 (Rosenkranz et al., 2008). To date no non-coding CyHV-3 genes have been identified.

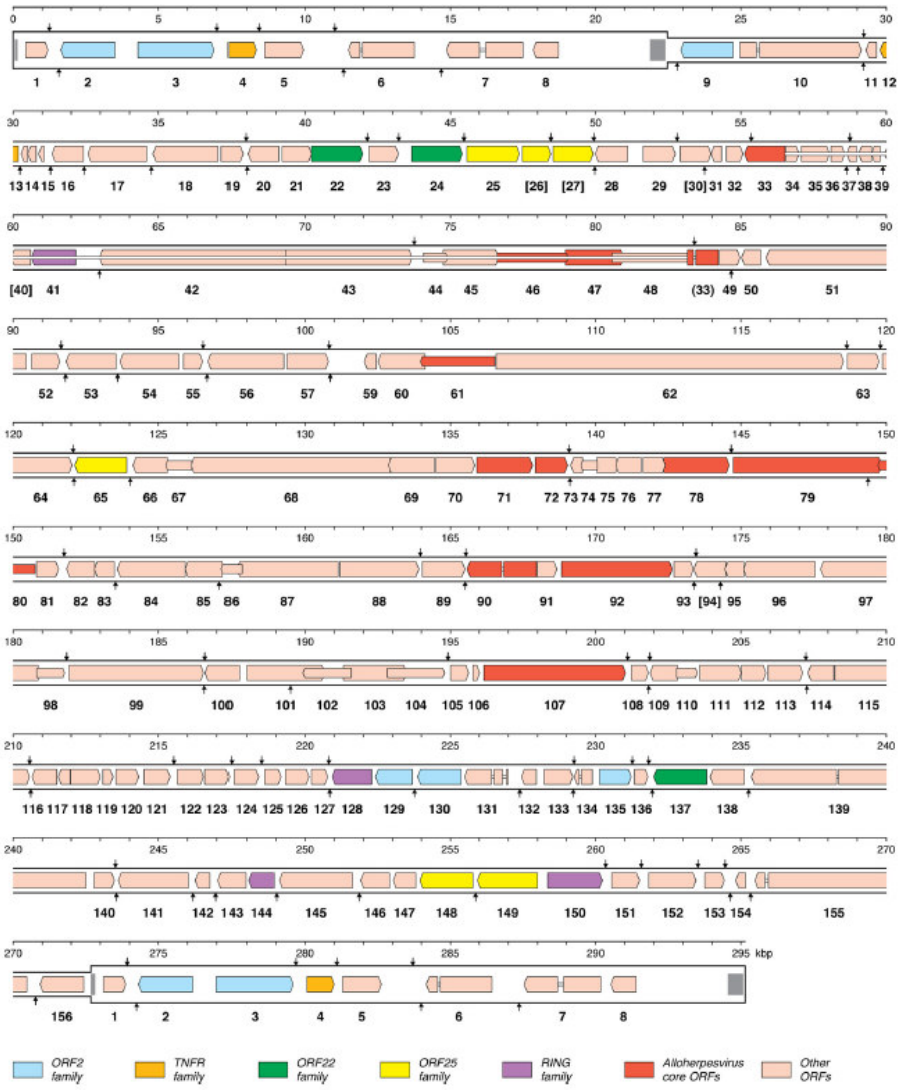



**Figure 1.7 CyHV-3 genome map showing all predicted ORFs, gene families and *Alloherpesviridae* core genes**
Diagram taken from Davison et al (2012) (updated version of map provided by Aoki et al (2007)).

### 1.4.10 Latency

Like other HVs, it is likely that CyHV-3 also establishes long-term latent infections that may reactivate under the appropriate conditions. Experimental infection studies and field investigations suggest that CyHV-3 latency is temperature-dependent. Outside the permissive temperature range, mortality associated with outbreaks decreases but low levels of CyHV-3 DNA can still be found in healthy survivors long after initial exposure (Gilad et al., 2004). CyHV-3 can be found even in the absence of any apparent outbreaks (Pokorova et al., 2007). In addition, despite the presence of CyHV-3 DNA in many organs of healthy fish (and reminiscent latent

infections established by other HVs), infectious virus cannot be isolated from the same tissues (Eide et al., 2011a, 2011b). Furthermore, lytic infections can be reactivated in latent carriers (causing mortality) by introducing them to temperatures within the CyHV-3 permissive temperature range. CyHV-3 reactivation has been demonstrated in this manner in both common carp (St-Hilaire et al., 2005) and koi (Eide et al., 2011b). A detailed description of some of the experiments that have shown that CyHV-3 can establish a latent infection is given below.

#### 1.4.10.1 *Experimental evidence CyHV-3 latency and reactivation*

There has been several large-scale *in vivo* infection experiments conducted since 2000. The results of these experiments have revealed a lot about the possible nature of CyHV-3 and especially the significance of water temperature on mortality rates. Thus the optimum water temperature range for a CyHV-3 infection to take hold has been established to be 16-28$^{o}$C. However it is experiments involving infection outside this optimum temperature range that have yielded the most interesting results and have given a greater insight into the possible nature of CyHV-3 latency.

One such study has shown that bath exposure to the virus at lower non-permissive temperatures such as 13$^{o}$C results in no symptoms or mortality. Subsequent transfer of the same fish to a permissive temperature of 23$^{o}$C, 30 days after initial exposure, resulted in acute death and high mortality (~80%). However the same study showed that fish kept at extended periods of time (64 days in this case) at 13$^{o}$C do not experience any mortality when moved to a permissive temperature of 23$^{o}$C (Gilad et al., 2003). In a separate study by St-Hilaire et al. (2005), fish were initially exposed to CyHV-3 at a permissive temperature of 22$^{o}$C for just 2 hours before being transferred to non-permissive water temperatures of 12$^{o}$C for 125 days. Unlike the results reported by Gilad, *et al.* (2003) above, in this case transferring the fish to a permissive temperature of 23$^{o}$C after a prolonged period at the non-permissive temperature did result in mortality of these fish (although mortality was reduced to ~50%) and 100% mortality among naïve fish that were cohabited with these fish at 23$^{o}$C. This suggested that the introduction of a short initial exposure to CyHV-3 (2h in this case) at the permissive temperature before extended periods at a non-

permissive temperature was enough for the virus to establish an infection capable of remaining dormant at low temperatures and reactivating at permissive temperatures.

Also in the same study, exposure to CyHV-3 at permissive temperatures for long periods (21$^{o}$C for 15 days) resulted in high mortality outbreaks (95%) with a small number of fish surviving due to subsequent lowering of the temperature to 12$^{o}$C (to promote survival). These fish showed no reactivation of the virus or infection of cohabiting naive fish that were introduced as the temperature was raised to 23$^{o}$C. This was in spite of the fact that the fish were also stressed by injections of the stress hormone cortisol. One of the survivor fish from the original outbreak did test positive for anti-CyHV-3 antibodies. A similar experiment was later conducted where fish were exposed to virus leading to another relatively high mortality outbreak, this time at an initial temperature of 18$^{o}$C for 20 days (later lowered to 11$^{o}$C to promote survival). Survivors were split into 2 groups in separate tanks and the temperature was raised to 23$^{o}$C. Reactivation of CyHV-3 in response to the raised temperature only occurred in one of these groups. Failure of clinical signs to reappear did not necessarily mean that the virus was not present, as interestingly in the group where no reactivation occurred, anti-CyHV-3 antibody was detected in one of the naïve fish that co-habited with the survivors. The only possible source of CyHV-3 in this tank would have been the co-habiting survivors of the previous outbreak. This suggested that the virus may have been actively shed by the survivor fish in the absence of clinical symptoms i.e. the hallmarks of a persistent/latent infection. The use of a re-circulation system in this tank may have helped retain any virus particles shed, increasing the chances of contact with naïve fish. Despite the fact that this took place at a permissive temperature, it did not result in a high mortality outbreak among naïve fish that had picked up the virus. This evidence certainly suggests that survivors may act as asymptomatic carriers and shed active virus particles that are at least capable of causing an immune response in naïve fish as shown by the production of anti-CyHV-3 antibodies.

The experimental studies on described above on CyHV-3 reactivation were done using carp. Reactivation has also been demonstrated in koi under similar circumstances. In this case six koi known to have had previous CyHV-3 infections or exposure, were re-introduced to CyHV-3 at permissive temperatures. CyHV-3 DNA

could be detected in the faecal secretions and gill swabs from these fish only when the temperature was raised but not beforehand. In addition, infectious virus could be isolated from the same sources only after the temperature was raised. This reactivation of CyHV-3 resulted in two mortalities. Interestingly, one these mortalities had been a survivor of both a suspected CyHV-outbreak in 1998 and confirmed CyHV-3 outbreak in 2003 (10 years prior to reactivation) (Eide et al., 2011b).

It has been shown that virus remains viable in water for approximately 4 hours (Perelberg et al., 2003). Hence in the experiments mentioned above, the clinical signs that appeared in fish, following the switch to a permissive temperature, were not caused by virus particles remaining in the immediate environment after the initial exposure, as in all cases a switch to the permissive temperature occurred long after 4 hours. In fact of the studies mentioned, the least time that had elapsed before a switch to the permissive temperature was 30 days (Gilad et al., 2003). Therefore it can be concluded from these studies that keeping fish at low temperatures after or during initial exposure must result in an infection of some kind at this stage, albeit in the absence or reduction of mortality and clinical signs. It is this initial dormant infection that can be the only source of reactivated virus at the permissive temperature later on. As shown, clinical signs *may or may not* re-appear and cause repeat outbreaks when transferred to permissive temperatures. These studies suggest that this may depend on the nature of the initial exposure with respect to temperature and time. It is highly unlikely that exposed fish survive better at lower temperatures due to temporary suppression of the virus by the adaptive immune response as it is well documented that the immune response in carp is at its lowest at these temperatures. The necessary magnitude of response required to fight off infection is much more likely to happen at higher temperatures (Le Morvan et al., 1998, 1996).

Whether or not the survival of CyHV-3 infected fish at low temperatures represents some form of temporary latency in response to low temperatures is something that has yet to be confirmed. Initiation of latency during cold temperatures may be an advantageous evolutionary adaptation for successful viruses of poikilothermic hosts whose immune system declines in response to cold temperatures. The establishment of a dormant infection at this stage would act to prevent the virus completely killing

off its immuno-suppressed host, allowing the host to survive the winter and act as a reservoir for infectious virus particles and emergence of the virus in spring when hosts start to interact in high numbers during spawning.

*In vivo* studies conducted at the other end of the non-permissive temperature scale (i.e. >28$^{o}$C) also show interesting results. As mentioned in Section 1.4.8.1 studies have shown that vaccination of fish can be achieved through exposure to CyHV-3 at a permissive temperature followed by transfer to a non-permissive temperature of 30$^{o}$C for 30 days. Resistant fish generated in this manner showed no sign of viral DNA or infectious virus at permissive temperatures. In addition cohabitation of "naturally resistant" fish with naïve fish did not lead to disease transmission (Ronen et al., 2003). Whether or not the virus remains present in some capacity in these "naturally resistant" fish is something that remains to be seen. The immune response in carp is at its most potent during such high temperatures (Le Morvan et al., 1998). This correlates with a decrease in CyHV-3 mortality above ~25$^{o}$C. It would be an advantage for a virus to go silent at high temperatures and stopping viral gene expression would prevent the display of viral antigens on infected cells and would act as a way of evading detection by the host's immune system while it is at its peak.

Dishon et al. (2007) produced an excellent *in vitro* simulation of the conditions used to create "naturally resistant" fish. Their study showed that a dormant persistent infection in the common carp brain cell line (CCBs) can be maintained at the higher non-permissive temperature of 30$^{o}$C. No viral replication occurred at this temperature. Importantly this study showed that virus was not destroyed at the non-permissive temperatures but instead reactivated again once cultures were reverted to permissive temperatures. However, the ability to reactivate virus ceased following 70 days at 30$^{o}$C indicating the infection may have become abortive at this point. It is unclear if an abortive infection (i.e. an infection where no viral DNA synthesis or new viral particles are produced) could occur *in vivo*, under similar conditions, in the case of "naturally resistant" fish.

This study by Dishon et al. (2007) presented an interesting *in vitro* model that may have facilitated the study of CyHV-3 latency/persistence. However, attempts to replicate this experiment in the manner described in this publication in our laboratory

and in our collaborating lab (Cefas, UK) did not show the same results. Instead CPE progressed at a much slowere rate and in some cases the negative control cells began to detach from the flask surface, indicating that the even in the absence of viral infection the cells were not well suited to culture at this elevated temperature. This *in vitro* model was not utilised in any other study until it was reproduced again by the same group for a follow-up study on gene expression (Ilouze et al., 2012b). Notably in this second study, and based on some further evidence presented therein, the infection established at 30$^{o}$C was now explicitly described as an abortive infection rather than a persistent infection as described in the first study.

It is clear from both the *in vitro* and *in vivo* studies described above that the characteristics of CyHV-3 infection do change in response to non-permissive temperatures, resulting in the establishment of a dormant infection involving the cessation of viral replication. During this stage, although viral genomic DNA can be detected, no infectious virus can be isolated. Furthermore, in some cases lytic replication can be reactivated upon re-exposure to permissive temperatures. This is certainly not unlike the pattern of behaviour displayed by other HVs during latent infections. CyHV-3 can therefore be described as a virus that is capable of establishing latent infection and that both establishment and reactivation are temperature-dependent. The molecular basis of these observations remains to be elucidated however.

#### 1.4.10.2 *CyHV-3 gene expression during latency*

Genes involved in lytic infections are switched off during latency. Instead, the virus relies on a subset of genes involved in the long-term maintenance of latent infections. There is not much information on CyHV-3 gene expression during latency but from the small volume of work that has been done in this area it is clear that the temperature changes described above also cause dramatic changes to CyHV-3 gene expression (most notably the down-regulation of genes involved in lytic replication) Some of these observations are outlined in this section.

In addition to establishing an apparent *in vitro* model of a CyHV-3 latent/persistent infection by incubating cells at 30$^{o}$C Dishon et al. (2007) (refer to Section 1.4.10.1)

also monitored the expression levels of viral genes throughout this infection. Initially 20 genes were monitored and all were found to be down-regulated to undetectable levels. In a later study another similar infection was set up at 30$^{o}$C and gene expression monitoring was expanded to include all 156 genes (Ilouze et al., 2012b) (at this point the infection was described as an abortive infection). The results showed that viral transcription patterns changed as a result of transfer to the non-permissive temperature. Some genes that had earlier been reported to be expressed during lytic replication (Ilouze et al., 2012a) were not expressed at all and the expression of other viral genes gradually decreased beyond the limit of detection. Not all genes were down-regulated at the same rate, and some genes (notably ORF115 and ORF116) were expressed for a longer time than any of the others. However as this infection was described as an abortive infection its relevance to CyHV-3 latency is unclear at this stage.

In terms of CyHV-3 gene expression (or lack of) during latency *in vivo*, Eide et al (2011b) showed that at lower permissive temperatures, while CyHV-3 DNA could be found in leucocytes from healthy fish, the same samples were negative for mRNA for the viral DNA polymerase and major capsid protein (MCP). This suggested that the CyHV-3 genome present was not undergoing replication. In another recent study, the expression of the same two genes along with those encoding helicase, IL-10 homologue and intercapsomeric triplex protein (ITP) were also monitored. Gene expression was monitored during lytic replication, latency and reactivation of CyHV-3 *in vivo* (Sunarto et al., 2012). All genes were expressed at high levels during lytic replication and at moderate levels during reactivation. Like Eide et al (2011b), this study found that the genes encoding the viral DNA polymerase, MCP and helicase were not expressed during latency. However, mRNAs from the IL-10 homologue and ITP were present at low levels during latency. It is unclear what role the CyHV-3 late gene ITP would play during latency and it is not known if it is even translated at that stage. The presence of the CyHV-3 IL-10 homologue may facilitate immune evasion during latency, in the same way as the HCMV IL-10 homologue although its expression is not essential for latency in HCMV (Section 1.2.2). It should be noted that not all HV IL-10 homologs are expressed in this manner e.g. the EBV IL-10 homologue is only expressed during lytic infections (Hsu et al., 1990). With this in mind it is also possible that these transcripts were just down regulated at a slower

rate than the other three transcripts assessed in this study and if fish were kept at 11ºC for longer than 24 days before sampling it is possible that expression of these genes would not be detected. Nevertheless, to date mRNA for ITP and IL-10 are the only CyHV-3 genes found to be transcribed during latency *in vivo*. In addition, based on the role of its homologue in HCMV it is plausible that CyHV-3 IL-10 plays an active role in helping to facilitate the maintenance of CyHV-3 latency.

#### **1.4.10.3** ***Tissue tropism***

CyHV-3 DNA can be routinely detected in healthy hosts by PCR. If this DNA represents latent genome, the results do not point to any particular tissue tropism during latency as viral genome has been reported in a wide variety of organs in healthy survivors including brain, eye, spleen, gills hematopoietic kidney, trunk kidney, and intestine (Eide et al., 2011a; Gilad et al., 2004). In addition some studies suggest that CyHV-3 DNA may persist in the brain longer than any other organs (Gilad et al., 2004; Yuasa and Sano, 2009). Apart from organs, CyHV-3 DNA can also be found in leukocytes. In addition to being non-lethal, recent studies suggest that testing leucocytes may in fact be the most reliable and sensitive way to detect latent CyHV-3 in both carp and koi (Eide et al., 2011a, 2011b; Xu et al., 2013).

#### **1.4.10.4** ***Viral levels during latency***

In all cases CyHV-3 DNA can only be detected at low levels in healthy carriers (Eide et al., 2011b; Gilad et al., 2004). However this low level of virus is consistent with the levels of other HVs during latency (Sawtell et al., 1998; Wang et al., 2005a). Even in leucocytes, the estimated level of latent genomes ranged from only 2-150 copies per microgram of total DNA (Eide et al., 2011b).

#### **1.4.10.5** ***Difficulties associated with diagnosing latent CyHV-3 infections***

Assaying for viral DNA is considered the most reliable/sensitive method for detecting latent CyHV-3 carriers (see Section 1.4.7). Experimental infection studies have shown it to be very likely that the low level CyHV-3 DNA detected in long term healthy survivors is indeed latent viral genome. Although such findings certainly support the idea that CyHV-3 can establish a latent infection, it is the

additional evidence such as reactivation of disease or the production of infectious viral particles upon introduction to permissive temperatures that prove that viral DNA initially detected at non-permissive temperatures represents latent viral genome. While CyHV-3 DNA likely to represent latent genome can be detected in survivors, currently there is no way to conclusively distinguish between latency and an abortive infection without reactivating infectious virus. Therefore in the strictest sense, the detection of low level viral DNA *alone* cannot be used to confirm latency. In addition, because the CyHV-3 genome is present in such low levels (as with latent infections of HVs) and because the levels of CyHV-3 genome detected in the organs of survivors tend to decrease over time after initial exposure (Gilad et al., 2004) there is a possibility that PCR could give false negatives when testing carrier fish. Therefore, although PCR is currently the most suitable method available assaying for latent CyHV-3, it may not always be reliable.

#### **1.4.10.6** ***An alternative diagnostic target: Viral RNA transcripts***

Analysis of CyHV-3 gene transcription during *in vitro* and *in vivo* infections at high and low non-permissive temperatures has shown a complete absence of or gradual down-regulation of known viral protein-coding genes (Dishon et al., 2007; Eide et al., 2011b; Ilouze et al., 2012b; Sunarto et al., 2012). Although two CyHV-3 genes, namely the viral Il-10 homologue and ITP, have been identified as being potentially expressed during latency (Sunarto et al., 2012), it is not clear whether their expression is gradually down-regulated (like all other CyHV-3 protein-coding genes) or if they both remain stably expressed during latency.

The identification of more genes likely to be expressed during CyHV-3 latency would not only be quite useful in terms of the molecular characterisation of this stage of the viral life cycle but might also be useful targets for the diagnosis of latent carriers. Currently the most reliable method to identify potential latent carriers of CyHV-3 is the detection of viral genomic DNA. However as outlined in Section 1.4.10.5, detection of viral DNA alone cannot confirm latency. If viral gene expression could be detected it would imply that the virus is still active in some capacity and capable of reactivation unlike an abortive infection. In addition to allowing confirmation of latency (as opposed to an abortive infection) targeting RNA

transcripts (instead of DNA) may make for more sensitive assays. In theory RNA transcripts from genes expressed during latency should be much more abundant than the low levels of viral genome from which they are transcribed. Being present in higher copy numbers, these RNA transcripts should be much more easily detectable and might reduce the occurrences of false positives.

There are several examples of protein-coding genes being expressed by HVs during latency e.g. EBNAs by EBV and a homologue of the cytokine interleukin-10 (IL-10) by HCMV (although not essential for latency). Usually however, HVs also express non-coding transcripts during this stage (or sometimes only non-coding transcripts at this stage e.g. HSV-1). These are sometimes just referred to as latency-associated transcripts (or LATs) and these are involved in the regulation of other genes. Examples include the LAT transcript in HSV-1 (Stroop et al., 1984) EBERs in EBV (Lerner et al., 1981).

One particular class of LATs that are particularly prominent during HV latency are miRNAs. They are involved in the regulation of both viral and host genes and provide an ideal way for the virus to maintain a cellular homeostasis that facilitates latency without using (or limiting the use of) potentially immunogenic viral proteins. Unsurprisingly most HV miRNAs identified to date have been found in latently infected cells although many are also expressed during lytic infections (although mainly with different expression levels). These non-coding transcripts are described in more detail in Section 1.5

## 1.5 MiRNAs

In recent years many HVs have been shown to encode miRNAs. MiRNAs are short non-coding RNA transcripts of typically ~22 nucleotides (nt) in length. They were first discovered in *Caenorhabditis elegans* (Lee et al., 1993) and were subsequently shown to have important roles in gene regulation through targeted cleavage or translational repression of specific mRNAs (Ambros, 2001; Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001). These transcripts are mostly encoded in intergenic regions, although they can occur antisense to protein-coding genes,

within 3' and 5' UTRs, introns or in rare cases overlapping with protein-coding sequences (Cai et al., 2005).

### 1.5.1 Biogenesis

MiRNAs are derived from a series of cleavage steps from much longer primary miRNA transcripts (pri-miRNAs) (Lee et al., 2002). Pri-miRNAs are usually transcribed by RNA polymerase II. The terminals of these primary transcripts are modified by the addition of 5' 7-methyl guanylate cap and 3' polyA tail (Cai et al., 2004; Lee et al., 2004). Certain parts of the pri-miRNA form distinct hairpin secondary structures, these structures are recognised and cleaved by the RNAse III enzyme Drosha and its co-factor DGCR8 (collectively referred to as the microprocessor complex). Cleavage occurs ~20-30 nt from the hairpin terminal loop, releasing a precursor-miRNA hairpin (pre-miRNA) that is typically ~60-70 nt in length and displays a characteristic 2 nt 3' overhang (Fukunaga et al., 2012; Lee et al., 2003) (Figure 1.8). These are then transported to the cytoplasm by Exportin-5 (Yi et al., 2003). Once in the cytoplasm, pre-miRNAs are cleaved near the terminal loop by another RNase III enzyme called Dicer that works in association with TRBP. This second cleavage step results in the release of a mature miRNA duplex present on the stem of the precursor, again this usually results in a 2 nt 3'overhang (Chendrimada et al., 2005; Ji, 2008; Koscianska et al., 2011).

Both of these strands from the mature duplex can become incorporated into a multi-protein complex known as the RNA-induced silencing complex (RISC), a key component of which is the Argonaute protein. In addition to facilitating miRNA processing, TRBP also recruits Argonaute 2 (Ago2) to the mature miRNAs bound by Dicer (Chendrimada et al., 2005). Usually one of these strands becomes stably incorporated in the RISC more frequently than the other with preference given to the strand with the lowest 5' terminus base-pairing stability (Khvorova et al., 2003; Schwarz et al., 2003; Tomari et al., 2007). This strand is referred to as the mature miRNA, guide strand or major form while the other strand that is not usually incorporated is known as the mature miRNA* (pronounced "miRNA star"), passenger strand or minor form. These strands will be referred to as the major form and minor form respectively from this point onwards. Once incorporated into the

RISC the miRNA can then carry out targeted mRNA silencing. This biogenesis pathway is illustrated in Figure 1.8 (a). Although most miRNA biogenesis occurs in this fashion, non-canonical miRNA biogenesis pathways have been discovered. Most notably, some pre-miRNA-hairpins occurring in introns can be processed independently of Drosha through the combined action of splicing machinery and lariat-debranching enzyme. Pre-miRNA hairpins that are processed in this fashion are referred to as "mirtrons" (Okamura et al., 2007). Less abundant variations of miRNAs are also normally present; these share the same core sequence as the miRNAs but display varying degrees of terminal heterogeneity, these are classed as isomiRs (Morin et al., 2008). These may occur as a result of inconsistencies in Dicer and Drosha processing of mature miRNAs from the pre-miRNAs (Ruby et al., 2007, 2006; Wu et al., 2009) and the nature of the heterogeneity may be influenced by pre-miRNA structural features (Starega-Roslan et al., 2011). These transcripts can also be incorporated into the RISC. In some cases they may be present at physiologically relevant levels and thus may also contribute significantly to gene silencing (Cloonan et al., 2011).

### 1.5.2 Mode of gene silencing

Once incorporated into the RISC, mature miRNAs can carry out gene specific silencing. Because gene silencing occurs through complementary base pairing between the miRNAs and specific regions on target mRNAs, each miRNA is capable of post-transcriptional regulation of a specific subset of mRNAs. These specific mRNA regions are referred to as target sites. This base pairing does not need to be fully complementary although in plants, miRNAs generally display a high degree of complementary base pairing with target sites, which are usually within the ORFs of these mRNAs (Voinnet, 2009). By contrast, perfect or near perfect base pairing between animal miRNAs and their target miRNAs are rare. In addition animal miRNA target sites are usually located within 3' UTRs as opposed to ORFs. In this case, mRNA target sites are generally perfectly complementary to positions 2-7 on the miRNA. This is known as the seed region and it is the primary determinant of animal miRNA target specificity. Additional base pairing between the target site and miRNA at positions 13-16 on the miRNA is also sometimes important and this acts to supplement or compensate for weak seed pairing stability (Garcia et al., 2011;

Grimson et al., 2007). These and other factors that are important for target specificity are described later in Section 1.8.

Ultimately target mRNAs are silenced through action of various components of RISC. This occurs by either endonucleolytic cleavage or mRNA repression (enhanced mRNA degredation/turnover or translational inhibition). Endonucleolytic cleavage occurs when miRNAs are complementary to mRNA targets and is carried out by Ago2. MRNA repression occurs when there is only a match to the seed region (normally the case in animal cells). This can be carried out by any of the several other Argonaute proteins (Gu and Kay, 2010; Su et al., 2009). Often mRNA repression involves the aggregation of RISC bound mRNA into translationally inactive cellular processing bodies (p-bodies) leading to degradation (Beckham and Parker, 2008; Eulalio et al., 2008; Huntzinger and Izaurralde, 2011).

### 1.5.3 MiRBase

MiRBase is an online database of all published miRNA sequences. Each miRNA entry is represented by its precursor name. Details of mature miRNAs derived from each pre-miRNA, a summary caption and associated publications are available for each pre-miRNA entry. Recent versions of miRBase have also included details of isomiRs, alignment signatures (alignments of miRNAs and isomiRs and their respective positions on pre-miRNAs) and read counts from deep sequencing experiments, although these details are usually only available for widely researched miRNA entries (Kozomara and Griffiths-Jones, 2011). Each miRBase entry is named according to a specific naming system. All pre-miRNAs start with a 3-4 letter abbreviation of the species of origin, followed by "-mir-" followed by a number, ascending in order of genome position or in order of discovery (sometimes this number may be preceded by a letter). MiRNAs use the same name as the precursor they are processed from except "-mir-" is replaced with "-miR-" Where the two miRNAs are annotated on a given pre-miRNA, both miRNA names end with either -5p or -3p depending on the arm from which they are derived. For example the first miRNAs discovered from HSV-1 are in miRBase under the precursor name hsv-mir-H1 and the miRNAs are named hsv-miR-H1-5p and hsv-miR-H1-3p. Exceptions to

these rules include miRNAs named prior to establishment of the nomenclature system, e.g. let-7a.

## 1.6 Viral miRNAs

Many eukaryotic miRNAs are involved in fundamental cellular processes and hence have been found to be heavily evolutionarily conserved across many distantly related metazoans (Lee et al., 2007; Wheeler et al., 2009). Viruses of these organisms are often found to encode homologues of host genes as a result of horizontal gene acquisition events during their evolution. Considering the ubiquitous presence of miRNAs in metazoan genomes, it is unsurprising that many viruses have also acquired this method of post-transcriptional regulation. Since their first discovery on viral genomes (Pfeffer et al., 2004) miRNA-coding genes have been found in 30 different viruses. Details of all currently known viral miRNAs are given in Table 1.4

### 1.6.1 Differences to eukaryotic miRNAs

Most viral miRNAs have been found to be no different from cellular miRNAs in their biogenesis and in the processes by which they carry out gene regulation and as such, they rely entirely on the use of host cell miRNA biogenesis and silencing machinery to function. Despite the many similarities to host miRNAs there are some differences that have given further insights into the biology of miRNAs in terms of biogenesis, evolution and the extent of their prevalence outside metazoans.

#### 1.6.1.1 *Biogenesis of viral miRNAs*

All known viral miRNAs are processed using cellular RNA processing machinery. However some viruses deviate from the canonical miRNA biogenesis pathway outlined earlier in Section 1.5.1. Most pri-miRNAs are transcribed by RNA polymerase II, although some viral pre-miRNAs have been shown to be transcribed by RNA polymerase III. Examples include, pri-miRNAs from murine γ-herpesvirus 68 (MHV-68) which are transcribed alongside viral tRNA-like transcripts (Bogerd et al., 2010; Reese et al., 2010) and miRNAs from Bovine leukaemia virus (BLV) (Kincaid et al., 2012). Furthermore, in both these cases the processing of pre-miRNAs from these pri-miRNAs takes place in a non-canonical fashion that is

independent of Drosha cleavage. In the case of MHV68, pre-miRNAs are processed from tRNA-like transcripts by host tRNase Z. In BLV, pri-miRNAs do not require processing by Drosha and are directly processed by Dicer i.e. the pri-miRNAs also function as pre-miRNAs without a need for prior Drosha processing. In addition, Herpesvirus saimiri (HVS) encodes pre-miRNAs on the 3' ends of other non-coding RNAs called HSURs. Mature HSURs and pre-miRNAs are processed from the same primary transcript by the host cell Integrator-Complex (Cazalla et al., 2011). All of the biogenesis pathways described here are shown in Figure 1.8.

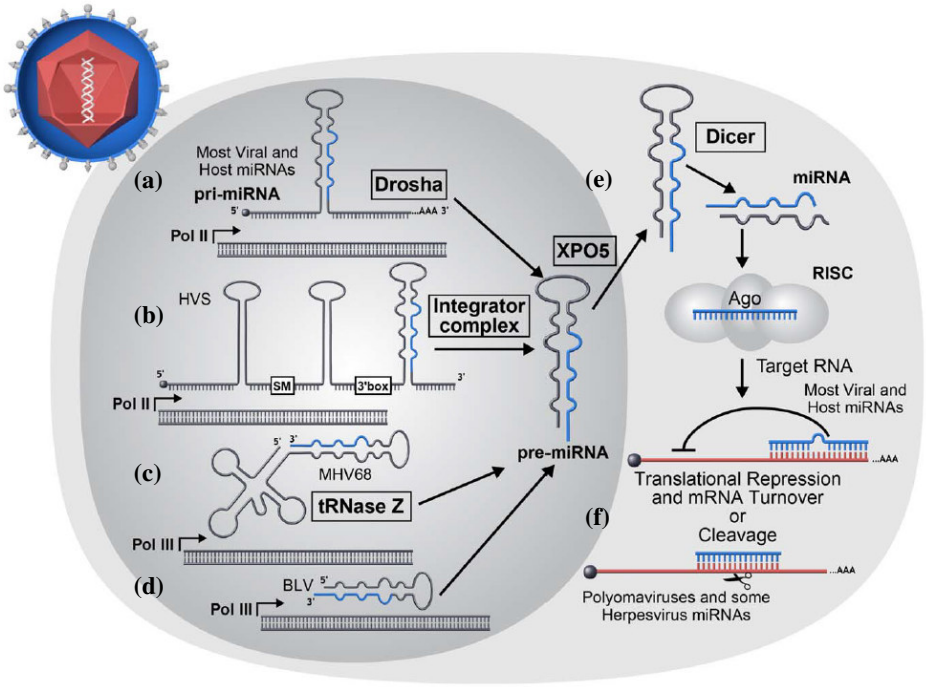



**Figure 1.8 Canonical and non- canonical viral miRNA biogenesis pathways**
The nucleus is represented by the dark grey shape on the left. It is surrounded by the cytoplasm in light grey. (a) The canonical miRNA biogenesis pathway used by most cellular and viral miRNAs is described in Section 1.5 (b) HVS pre-miRNA forming after 3' box on primary HSUR transcript and processed by cellular integrator complex (c) MHV68 pre-miRNA forming on the 3' end of a viral tRNA-like transcript (transcribed by RNA polymerase III) and processed by tRNase Z (d) BLV pre-miRNA (pri-miRNA) directly transcribed as a short discrete transcriptional unit by RNA polymerase III (e) Export of pre-miRNA from nucleus by Exprotin5 (XPO5), Dicer processing releasing mature miRNA duplex (with 3' overhangs) and incorporation into RISC (f) Targeted regulation of specific mRNAs. Top: imperfect complementary pairing to target site leading to translational repression of target mRNA, Bottom: perfect complementary base pairing to target site leading to cleavage of target mRNA. The diagram was adapted from Kincaid and Sullivan (2012).

#### 1.6.1.2 *Prevalence among viral families*

Even though miRNAs are ubiquitous among eukaryotic organisms, current knowledge suggests that they may not be equally prevalent among all viral families. Most known viral miRNAs occur in dsDNA viruses (Table 1.4). There may be several reasons why this is the case. Firstly, in theory, miRNA genes encoded in an ssRNA + sense genome may form the appropriate pre-miRNA hairpin structures without a need for transcription, allowing parts of the genome to act as substrate for Drosha. Thus, the presence of miRNA genes in RNA viral genomes may result in genome cleavage and may be disadvantageous for such viruses. Secondly most RNA viruses replicate in the cytoplasm (some exceptions include *Orthomyxoviridae* and *Retroviridae*) and therefore do not have access to the microprocessor complex involved in canonical miRNA biogenesis pathways as these are housed in the nuclei of host cells. Unlike most RNA viruses retroviruses replicate in the nucleus and in addition they also have a DNA stage, thus these are the only RNA viruses that have been shown to produce miRNAs (Kincaid et al., 2012; Ouellet et al., 2008). As outlined in Section 1.6.1.1 BLV pre-miRNAs are directly transcribed by RNA Polymerase III. In this case the correct pre-miRNA structure is only formed in this context. The same sequence does not fold into the same hairpin in the setting of the full genome or when present in longer RNA polymerase II transcribed mRNAs, thus allowing BLV to circumvent cleavage of the viral genome while still at its RNA stage and cleavage of overlapping mRNAs. It has been demonstrated that the insertion of a miRNA gene into influenza A virus - a member of the *Orthomyxoviridae* family, results in efficient miRNA processing and expression (Varble et al., 2010). In another interesting experiment, an EBV pre-miRNA was inserted into the genome of tick-borne encephalitis virus (TBEV), a member of the *Flaviviridae* family. Despite the fact that TBEV is an RNA virus that replicates in the cytoplasm, it was demonstrated that it was possible for miRNAs to be processed from the pre-miRNA, but at a low level (Rouha et al., 2010). This process did involve Drosha and the observed inefficiency of miRNA processing may be caused by the separation of the miRNA from the microprocessor components in the nucleus. It is possible that low level miRNA processing did occur as a result of breakdown of the nuclear membrane during mitosis, thus allowing temporary access to the miRNA processing machinery.

Despite the fact that it is possible for some RNA viruses to encode miRNAs, they seem to be much more prevalent in dsDNA viruses, such as members of the *Polyomaviridae* family and HVs in particular. In fact the majority of all known viral miRNAs occur in HVs. This may be due to the fact that they have larger genomes, but it should also be noted that most of them occur as part of miRNA gene clusters that occupy only short stretches of their respective genomes, meaning that the presence of multiple miRNA genes is not necessarily a result of their having larger genomes. One of the main reasons for the high prevalence of miRNAs among HVs may be latency. All known HVs have the ability to establish latent infections (Roizmann, 2001). MiRNAs appear to be closely associated with latency and this is reflected in the fact that the majority of known HV miRNAs have initially been indentified in latently infected cells. Latency represents a unique transcriptional state for the HVs, where they only rely on a subset of genes in order to maintain latency on a long-term basis. It is thought that the utilization of miRNAs during latency in particular is an evolutionary adaption of herpesviruses most likely driven by the fact that miRNAs provide an ideal way for the virus to maintain a cellular homeostasis that facilitates latency without using (or limiting the use of) potentially immunogenic viral proteins. This idea of miRNAs being connected with HV latency seems quite plausible. However, studies on HV genomes have concluded that 3 human HVs, HHV-3, HHV-6 and HHV-7 are not likely to encode miRNAs. These viruses all establish latent infections and it remains unclear why they would have not adapted similar methods of gene regulation (Pfeffer et al., 2005b). Indeed deep sequencing of small RNAs from HHV-3 latently infected ganglia failed to identify any potential HHV-3 miRNAs (Umbach et al., 2009a). Contrary to predictions however, a recent study was able to identify a low number of HHV-6B encoded miRNAs in latently infected cell cultures (Tuddenham et al., 2012) although these have been removed from latest Release of miRBase (Table 1.4). Many viral miRNAs expressed during latency are also expressed (albeit at different levels) during lytic infections (Cai et al., 2006, 2005; Cui et al., 2006; Pfeffer et al., 2005b; Reese et al., 2010; Umbach and Cullen, 2010; Zhu et al., 2010b). The biological relevance of this is not clear and this may just occur as a consequence of an overall increase in the levels transcription from viral genomes or read-through transcripts during lytic infections. Indeed, many viral miRNAs do have specific roles during lytic infections (Section 1.6.3).

**Table 1.4 Details of all viral miRNA entries in the current release of miRBase (Release 20 – June 2013)**
All virus families represented in this table are dsDNA viruses with the exception of *Retroviridae*. N.I. = Not Investigated.

| Family | Virus | Known Pre-miRNAs | Known miRNAs | Infection | |
|---|---|---|---|---|---|
| | | | | Lytic | Latent |
| ***Alphaherpesvirinae*** | *Anatid herpesvirus 1* (AHV-1) | 24 | 33 | Yes | N.I |
| | *Bovine herpesvirus 1* (BoHV-1) | 10 | 12 | Yes | N.I. |
| | *Herpes B virus* (HBV) | 3 | 3 | Yes | N.I. |
| | *Herpes Simplex Virus 1* (HSV-1) | 17 | 26 | Yes | Yes |
| | *Herpes Simplex Virus 2* (HSV-2) | 18 | 24 | Yes | Yes |
| | *HV of turkeys* (HVT) | 17 | 28 | Yes | N.I. |
| | *Infectious laryngotracheitis virus* (ILHV) | 7 | 10 | Yes | N.I. |
| | *Marek's disease virus type 1* (MDV-1) | 14 | 26 | Yes | Yes |
| | *Marek's disease virus type 2* (MDV-2) | 18 | 36 | Yes | Yes |
| | *Pseudorabies virus* (PSV) | 13 | 13 | Yes | N.I. |
| ***Betaherpesvirinae*** | *Human cytomegalovirus* (HCMV) | 11 | 17 | Yes | N.I. |
| | *Human herpesvirus 6B* (HHV-6B)[a] | 4 | 8 | Yes | N.I. |
| | *Mouse cytomegalovirus* (MCHV) | 18 | 29 | Yes | N.I. |
| ***Gammaherpesvirinae*** | *Epstein-Barr virus* (EBV) | 25 | 44 | Yes | Yes |
| | *Herpesvirus saimiri strain A11* (HSV) | 3 | 6 | N.I. | Yes |
| | *Kaposi sarcoma-associated HV* (KHSV) | 13 | 25 | Yes | Yes |
| | *Mouse gammaherpesvirus 68* (MHV68) | 15 | 28 | Yes | Yes |
| | *Rhesus lymphocryptovirus* (rLCV) | 36 | 68 | N.I. | Yes |
| | *Rhesus monkey rhadinovirus* (RRV) | 7 | 11 | Yes | Yes |
| ***Polyomaviridae*** | *BK polyomavirus* (BKV) | 1 | 2 | Yes | N.I. |
| | *JC polyomavirus* (JCV) | 1 | 2 | Yes | N.I. |
| | *Merkel cell polyomavirus* (MCV) | 1 | 2 | Yes | N.I. |
| | *Simian virus 40* (SV40) | 1 | 2 | Yes | N.I. |
| | *Murine polyomavirus* (PyV)[b] | 1 | 2 | Yes | N.I. |
| | *Simian agent 12* (SA12)[c] | 1 | 2 | Yes | N.I. |
| **Unclassified: Related to both *Polyomaviridae* and *Papillomaviridae*** | *Bandicoot papillomatosis carcinomatosis virus type 1* (BPCV-1) | 1 | 1 | Yes | N.I. |
| | *Bandicoot papillomatosis carcinomatosis virus type 2* (BPCV-2) | 1 | 1 | Yes | N.I. |
| ***Ascoviridae*** | *Heliothis virescens ascovirus* (HvAV)[d] | 1 | 1 | Yes | N.I. |
| ***Retroviridae*** | *Bovine leukemia virus* (BLV) | 5 | 8 | Persistent | |
| | *Human immunodeficiency virus 1* (HIV-1)[e] | 3 | 4 | Persistent | |

[a]HHV-6B entry changed from "4" pre-miRNAs in miRBase Release 19 to "0" pre-miRNAs in Release 20 (the reason for this change was not disclosed). [b]PyV miRNAs not submitted to miRBase taken from Sullivan et al (2009), [c]SA12 miRNAs not submitted to miRBase taken from Cantalupo et al (2005), [d]HVaV miRNAs not submitted to miRBase taken from Hussain et al. (2008) [e]Two HIV miRNA entries may be deleted from future releases of miRBase due to lack of sufficient experimental evidence

## 1.6.2 Evolutionary conservation

In contrast to eukaryotic miRNAs, most viral miRNAs show no evolutionary conservation with host or other viral miRNAs. This is true even where the genomic location is conserved between related viruses e.g. the miRNA clusters in Kaposi's sarcoma-associated herpesvirus (KSHV) and related rhesus rhadinovirus (RRV)

(Walz et al., 2010) and miRNAs in Marek's disease viruses-1 and -2 (Grey et al., 2008; Yao et al., 2007). Notable exceptions appear to only be limited to very closely related viruses e.g. 22 miRNAs shared between EBV and rLCV (Riley et al., 2010) and 9 pre-miRNAs shared between RRV and Japanese macaque herpesvirus (JMHV) (Walz et al., 2010). There are several miRNAs from HSV-1 and HSV-2 that show conservation in genomic location and significant sequence homology particularly in their seed regions but no identical miRNAs (Jurak et al., 2010). Sequence conservation seems to be more prevalent in the *Polyomaviridae* family with miRNAs shared between BKV and JCV (Seo et al., 2008) and BPCV-1 and BPCV-2 (Chen et al., 2011) respectively. Although some miRNAs are conserved, this does not necessarily imply that their respective pre-miRNA sequences (Figure 1.9) are conserved suggesting that in these cases it is only the miRNAs themselves that are under selective pressure (Takane and Kanai, 2011).

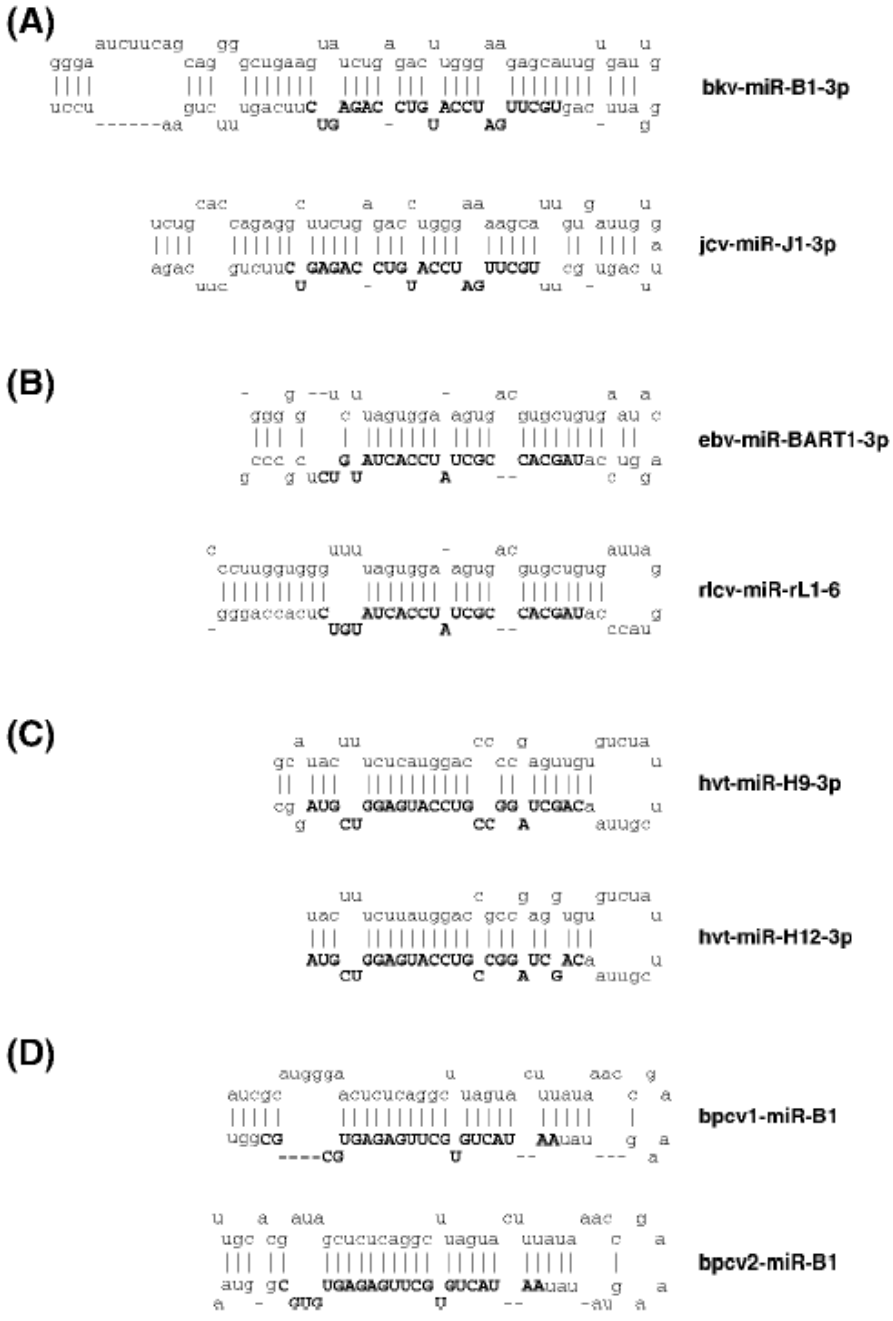



**Figure 1.9 Examples of miRNAs conserved between different viruses or that are present in different pre-miRNAs in the same virus**

Conserved miRNAs are highlighted in bold on their respective pre-miRNAs. All miRNAs are named following the system described in 1.5.3. For each miRNA the virus of origin is stated in the prefix e.g. ebv-miR-BART-1-3p is from Epstein-Barr Virus. (A) Conserved miRNA between JCV and BKV (B) Conserved miRNA between EBV and rLCV (C) MiRNA present on 2 different pre-miRNAs on the HVT genome (D) Conserved miRNA between BPCV-1 and BPCV-2. Diagram adapted from Tankane and Kanai (2011).

### 1.6.3 Functions of viral miRNAs

The functions of most viral miRNAs are either unknown or have only been tentatively predicted. Some have been investigated however, in particular in the case of miRNAs from EBV, HSV-1 and KSHV. In the context of both lytic and latent infections, viral miRNAs have been found to carry out three main types of functions (i) promotion of cell survival (ii) viral immune evasion, (iii) to control or limit lytic replication. Viral miRNAs carry out these functions through the regulation of both host and viral protein-coding genes. While this is comprehensively reviewed elsewhere (Grundhoff and Sullivan, 2011; Kincaid and Sullivan, 2012; Skalsky and Cullen, 2010), a few key examples illustrating the general functions of viral miRNAs are outlined below.

Cell survival is important for the establishment and maintenance of long-term latent infections and usually involves viral regulation of host cell factors that promote cell death. For example, during EBV latency, ebv-mir-BART-5 down-regulates the host cell pro-apoptotic protein (PUMA) (Choy et al., 2008) and other EBV miRNAs from the BART cluster target the cellular *Bim* gene which encodes a Bcl-2 interacting mediator of cell death. KSHV miRNAs have also been shown to promote cell survival by targeting the apoptosis initiator TWEAKR (Abend et al., 2010) and late apoptotic caspase 3 (Suffert et al., 2011). MiRNAs from KSHV, EBV and HCMV have all been shown to regulate the pro-apoptotic gene *BCLAF1*. This is interesting as none of these viral miRNAs are conserved between these three viruses. Instead this may represent a case of convergent evolution indicting that the regulation of this gene is important in the lifecycle of all three viruses (Lee et al., 2012; Riley et al., 2012; Ziegelbauer et al., 2009).

The use of miRNAs to suppress viral proteins capable of eliciting an immune response during latency is advantageous to viruses (Sullivan, 2008). This can also be useful during productive lytic infections. All members of the *Polyomaviridae* family (investigated so far) have been found to express a single miRNA that is antisense to a viral early gene encoding a protein called T-antigen (Cantalupo et al., 2005; Seo et al., 2009, 2008; Sullivan et al., 2009, 2005). It is hypothesized that down-regulation of this gene helps evade cytotoxic T-cell-mediated cell lysis. While this was shown to be the case in SV40 (Sullivan et al., 2005), down-regulation of the T antigen by

PyV miRNA did not enhance cytotoxic T-cell evasion under the experimental conditions used (Sullivan et al., 2009). In a similar case to the one mentioned earlier, there is yet more evidence of convergent evolution in the function of miRNAs from EBV, KSHV and HCMV. A set of miRNAs from these 3 viruses has been shown to down-regulate the major histocompatibility complex class I-related chain B (*MICB*) mRNA. This encodes a ligand for natural killer (NK) cells and its down-regulation in infected cells leads to reduced NK mediated destruction of infected cells (Nachmani et al., 2009; Stern-Ginossar et al., 2007).

As well as evading the immune response viral miRNAs can also act to prolong the longevity of infected cells and limit the lytic cycle. Many viral miRNAs that are expressed during latency have been shown to suppress viral genes that are involved in lytic infections and therefore act to maintain latency by preventing lytic cycle initiation. Some examples of HSV-1 and EBV miRNAs that function in this manner were described in Section 1.2.2. Expression of the HSV-1IE early gene *ICP0* is prevented by the miRNA hsv1-miR-H2-5p. This HSV-1 miRNA is unusual as it is situated antisense to the *ICP0* ORF in the HSV genome; as a result, the miRNA is perfectly complementary to a target site within the *ICP0* coding region. Such an occurrence is rare in animal cells but is more common in viruses and may be a consequence of their small genome size. A homologue of this miRNA also exists in HSV-2. This was originally designated HSV-2-miR-III but for convenience is now designated hsv2-miR-H2-5p in line with its homologue in HSV-1. Although this miRNA is slightly divergent in sequence, its genomic location is conserved, hence it also targets the HSV-2 *ICP0* homologue (Jurak et al., 2010). KSHV also uses miRNAs to maintain latency, for example the miRNA miR-K12-9-5p targets the 3' UTR of an mRNA for the KSHV Replication and Transcription Activator protein (RTA), reducing the occurrence of spontaneous lytic reactivation in latently infected cells (Bellare and Ganem, 2009; David, 2010). EBV miRNA miR-BART2 down-regulates the viral DNA polymerase BALF5 during latency, presumably acting to inhibit the transition to lytic replication (Barth et al., 2008). As well as regulating lytic genes during latency, viral miRNAs can also regulate lytic genes during lytic infections. During HCMV lytic infections the miRNA miR-UL112-1 regulates the viral trans-activator gene *IE-72* (interestingly it has 2 target sites in its 3' UTR). It is likely the regulation of this gene late in lytic replication is necessary to ensure the

correct temporal sequence of viral gene expression during this process (Grey et al., 2007). Notably this is the same HCMV miRNA that targets the *MICB* gene (mentioned earlier) and is a good example of a miRNA that regulates both cellular and viral genes.

Some viral miRNAs have been shown to possess the exact same seed regions as cellular miRNAs and are therefore capable of regulating at least some of the same transcripts. This is the case for oncogenic viruses such as MDV-1, KSHV and BLV. MDV-1 and KSHV encode orthologues of miR-155 and BLV encodes a mir-29 orthologue. In all cases the viral miRNA seed regions are exact matches to the cellular miRNAs, implying that they have the same targets (Boss et al., 2011; Kincaid et al., 2012; Zhao et al., 2011). Strikingly, these particular cellular miRNAs are known to directly contribute to oncogenesis when over-expressed and unsurprisingly all of the viral orthologues mentioned have been shown to independently contribute to the same process.

## 1.7 Identification of novel viral miRNAs

Given their roles in the regulation of other genes, the identification of novel miRNA genes (either conserved or non-conserved) can significantly contribute to or change our understanding of the molecular basis of many biological processes. The identification of novel miRNAs usually involves a mixture of bioinformatics and experimental work. In the following section these approaches are described and critically evaluated in the context of this study.

### 1.7.1 Prediction of novel viral miRNAs

The use of bioinformatics to predict novel miRNAs in genomes can be very useful and if approached properly, the predictions can serve as a solid basis from which to proceed with experimental approaches to novel miRNA identification. The approach to novel miRNA prediction depends on the nature of the organism in question and the objectives of the study, but broadly speaking the methods used are dictated by whether the study requires the identification of conserved or non-conserved miRNAs.

#### 1.7.1.1 *De Novo prediction of non-evolutionarily conserved miRNAs*

Based on observations in other viruses (Section 1.6) it is more likely that most novel viral miRNAs are species-specific, and in such cases *de novo* approaches to novel miRNA prediction are the most appropriate approached to use. *De novo* miRNA prediction does not rely on any evolutionary conservation and involves searching for miRNAs in a more indirect manner. This usually involves scanning genomes and predicting the occurrence of RNA secondary structures that loosely resemble pre-miRNAs. However, genomes produce many random RNA hairpin structures for many reasons and only a small number of these may function as pre-miRNAs, it is therefore important that these can be distinguished from non-pre-miRNA hairpins in some way. Unfortunately, the structural properties of *bone fide* pre-miRNAs are difficult to define in a manner that would allow them to be definitively distinguished from non-pre-miRNA-like hairpins or pseudo-pre-miRNAs. This means that unfortunately it is not currently possible to develop a "perfect" algorithm that would allow definitive distinction of real pre-mRNAs from background noise. Thus, approaches to *de novo* pre-miRNA prediction could be largely described as being based on a "loose" form of pattern recognition. These are generally rule-based approaches that take into account general structural properties of known pre-miRNAs. In the development of such prediction methods, the ideal structural properties of pre-miRNAs are usually defined through analysis of sets of *bone fide* pre-miRNAs, referred to as 'training data'. Properties analysed include hairpin length, terminal-loop size, minimum free energy, the extent of base-pairing in stems, bulge frequency, bulge size and bulge location, sequence composition, sequence complexity, repeat elements and internal and inverted repeats and other biologically relevant features (Bentwich, 2005). The exact composition of the training data used to develop methods may have some effect on performance or suitability in certain circumstances. This is important as training data can vary considerably in content, size and organisms of origin. Hence methods may only be as useful as the quality and suitability of the training data that is used as part of their respective development processes. Once pre-miRNA structural properties are defined in a way that allows *bone fide* miRNAs to be distinguished from non-pre-miRNA-like hairpins to an acceptable degree (checked by using appropriate negative training data), algorithms are developed to carryout high throughput analysis of the same properties among predicted hairpins. To distinguish between low and high likelihood pre-miRNA

predictions, each candidate is usually scored. Scores are usually determined by assigning values to each structural property based on a predefined scoring system. This is where most *de novo* prediction methods differ from one another. Different methods commonly assign different weights to different structural properties. An ideal way to ascertain the significance scores for given hairpins is to impose cut-off values, allowing elimination of low likelihood pre-miRNA predictions. The establishment and implementation of cut-off values relevant to specific studies is important in order to ensure correct interpretation of the significance of given predictions. The use of cut-off values is a delicate balance between increasing accuracy (lowering false positive rates) and maintaining sensitivity (not eliminating real pre-miRNAs from predictions), and it is best to have some acceptable way of measuring of the effects of various cut-off values on both of these aspects. Several methods that allow high throughput *de novo* prediction of pre-miRNAs in genomes have been developed. Details of several of these methods are given in Table 1.5.

**Table 1.5 List of methods used for *de novo* prediction of miRNAs in genomes**

| Method | Description | Reference |
|---|---|---|
| Bentwich-Method | Involves, predicting hairpins in genomes, filtering predicted hairpins based on overlap with other known RNA transcripts and scoring based on stability and structure (Not publically available) | Bentwich et al (2005) |
| Pfeffer-SVM-Method | Uses a support- vector machine (SVM) learning algorithm to predict pre-miRNAs in genomes (Not publically available) | Pfeffer et al (2005b) |
| VMir | Predicts pre-miRNAs in viral genomes, assigns scores to hairpins based on structural consistency with pre-miRNAs and stability. | Grundhoff et al (2006) |
| MiRank | Predicts hairpins using RNAfold. Uses random-walk machine learning algorithm to assess hairpins. Requires a small amount of known miRNAs from the genome in question in order to predict novel miRNAs | Xu et al (2008) |
| SSC Profiler | Scans genomes for hairpins. Uses a Hidden Markov Model based method to assess hairpins. Only for available for use with human genome. | Oulas et al (2009) |
| NOVOMIR | Searches for hairpins in genomes using RNAfold. Evaluates and scores predicted precursors based on consistency with plant miRNAs. | Teune and Steger (2010) |

The programs listed in Table 1.5 mainly consist of a series of algorithms (a pipeline) that predicts hairpins in a genome followed by scoring of these hairpins in terms of their structural consistency with *bone-fide* pre-miRNAs. Two of these methods, Pfeffer-SVM-Method and VMir have been successfully used to predict viral pre-miRNAs, however Pfeffer-SVM-Method is not publically available. The Bentwich-Method also seems quite suitable to viral miRNA prediction although like the Pfeffer-SVM method, it is not publically available and for the reasons described above SSC profiler and, miRANK are not suitable (Table 1.5).

VMir was specifically designed for the prediction of novel pre-miRNAs in viral genomes. It also generates particularly useful graphical and raw data output that can be easily utilised by other downstream methods in the novel miRNA identification process. It works by sliding an analysis window of predefined size over sections of genome. This window advances along the genome in predefined increments. All RNA secondary structures formed by the sequence in each window are predicted using the in-built RNAfold algorithm. Any hairpins that form are identified and

scored. The scoring system is based on structural features and similarity to *bone fide* pre-miRNA structures (see Grundhoff et al. (2006) for a more detailed account of this scoring system). Hairpins from different analysis windows are compared and classified as main, subsidiary or repeated Hairpins (MHPs, SHPs, or RHPs, respectively). They are then grouped into local and/or repeat families. Local families consist of 2 or more hairpins forming at the same location and sharing the same core loop sequence but differing in stem length. These are simply variations of essentially the same hairpin that have formed differently in separate analysis windows. These differences in the predicted structure occur due to slight changes in the flanking sequences available for folding calculations in each analysis window. Within each local family, the longest hairpin is designated as the MHP and shorter variations of this hairpin (identified in different analysis windows) are designated SHPs. Repeat groups consist of hairpins located at different parts of the genome that have the same sequence and structure. Within these repeat groups, the hairpin located nearest to the 5' end of the genome (based on the forward strand) is designated as the MHP of the group and all other repeats of this are designated RHPs. Individual hairpins that are not part of local or repeat groups are by default classified as MHPs. Each MHP is assigned an individual alphanumerical ID e.g. MD*** for regions on the forward strand and MR*** for regions on the reverse strand.

The graphical output consists of a chart indicating the position of each predicted MHP (as a representative of each local hairpin family). Genome position is represented on the X-axis and the VMir score of each MHP is represented by the Y-axis. Selecting individual MHPs will bring up a graphical representation of its structure in a window on the right hand side and includes information such as score, rank, window counts, SHPs, hairpin lengths, RNA sequence and individual hairpin structures in dot bracket-notation and in plain text format. To improve the quality of predictions VMir also allows results to be filtered based on several parameters:

- Score: The higher the score, the more pre-miRNA-like the hairpin. Setting the score filters high eliminates non-pre-miRNA-like hairpins from the results

- Min/Max Hairpin Size: A lot of results fall outside the sizes ranges of *bone fide* pre-miRNA molecules. Setting appropriate size filters can get rid of these.

- Window Count (WC): This is the number of windows in which a specific MHP or SHP was detected in during analysis. Adjacent windows will fold differently due to differing base composition at the beginning and ends of the sequence being analysed. A higher WC gives more confidence that the hairpin represents a stable structure within its specific local sequence context. The WC of a specific MHP or SHP is known as the absolute-WC. The combined values for absolute-WC for the MHP and all SHPs in a local HP family is known as the relative-WC and this acts as an indication of the overall stability of a given MHP site. Results can be filtered based on relative-WC, thus removing any unstable HPs on the basis that pre-miRNAs should be more stable than random HPs from other RNA transcripts.

A significant advantage in using VMir is the fact that it also has integrated array design tool that will generate user specified probes (and control probes) to detect potential miRNAs derived from pre-miRNAs that it has predicted, thus aiding high throughput experimental evaluation of its predictions. In addition it will also accept imports of raw fluorescence data from subsequent DNA microarray experiments and carry out automated analysis of the data in order to identify positive signals for miRNAs from predicted precursors.

#### 1.7.1.2 *Pre-miRNA classifiers*

Some methods of miRNA prediction concentrate on the classification of proposed pre-miRNAs that have been predicted by other means i.e. high throughput hairpin prediction in genomes (as described in section 1.7.1.1 ) or experimentally identified pre-miRNA candidates. Although classifiers do not predict pre-miRNAs themselves they do provide a useful way to filter prediction results. This is important sometimes because while most *de novo* prediction methods score hairpins, usually they do not classify them as pre-miRNAs or pseudo-pre-miRNAs, whereas classifiers will do this and in addition will usually provide associated probability scores. As expected, pre-

miRNAs classified as having a high probability of being *bone-fide* pre-miRNAs also score highly using *de novo* prediction methods, although classifiers generally analyse proposed pre-miRNAs in a much more thorough manner and also employ statistical methods. Pre-miRNA classifiers are ideal methods to use to support conclusions from *de novo* pre-miRNA prediction programs or experimentally identified pre-miRNAs. Details of several of pre-miRNA classifiers are given in Table 1.6.

**Table 1.6 List of pre-miRNA classifier methods**

| Method | Description | Reference |
|---|---|---|
| Triplet-SVM | See text below | (Xue et al., 2005) |
| MiPred | See text below | (Jiang et al., 2007) |
| miPred | SVM (support vector machine) based approach where classification based on 23 global and intrinsic features of pre-miRNAs | (Ng and Mishra, 2007) |
| microPred | SVM based approach similar to miPred but used much higher quality training data sets | (Batuwita and Palade, 2009) |
| CSHMM-Method | Uses of a context-sensitive-hidden-Markov-model to estimate the probability that a given hairpin is a real pre-miRNA. | (Agarwal et al., 2010) |
| miRD | Uses 2 strategies to (A) to classify multi-stem pre-miRNA and (B) to classify standard single-stem pre-miRNAs using 59 features based on sequence and structural composition | (Zhang et al., 2011b) |

In theory, any of these are suitable for classification of viral pre-miRNA candidates. However some of these methods have been directly compared to each other in terms of performance and the results of these comparisons may be useful when it comes to selecting what method(s) are most suitable for particular studies. A comparison of MiPred and Triplet-SVM showed that MiPred performed better at identifying pseudo-pre-miRNAs, correctly classifying 96% of them compared to Triplet-SVM (86%) (Sinha et al., 2009). MiPred takes advantage of a novel machine learning algorithm known as 'Random Forest' that takes into account local contiguous structure-sequence composition. This is based on earlier observations that the local sequence and structural features (especially the stem) are important pre-miRNA for distinguishing between genuine pre-miRNAs and random hairpins. There appears to be a subtle pattern to the type and frequency of sub-structures and sequence motifs that are most likely to be present in pre-miRNA stems. In addition these features will have a tendency to make pre-miRNAs more stable than other types of hairpins (Bonnet et al., 2004b). Before Jiang et al, (2007) developed MiPred, Xue et al,

(2005) had developed a triplet-SVM to analyse local contiguous sequence structure composition of pre-miRNAs. The approach involved structural analysis of hairpins by focusing on 3 adjacent stem nucleotides at a time, referring to these as triplets. Each triplet was described in terms of its sequence and structure, with the sequence defined by the middle nucleotide and the structure defined as either "(((", "((.", "(..", "(.(", ".((", ".(.", "..(" or "..." with "." meaning unpaired and "(" meaning paired. As there are 4 different middle nucleotide possibilities there are 32 possible sequence-structure triplets denoted as "A(((" "C(((" etc… In this case a type of machine learning algorithm known as a 'support-vector-machine' was used to study a training data set consisting of bone-fide pre-miRNAs and pseudo-pre-miRNAs and take into account patterns of triplets present. This was the basis of the triplet-SVM method (Xue et al., 2005). To develop MiPred, Jiang et al, (2007) further built on the idea behind triplet-SVM by using an improved machine learning method known as Random-Forest (introduced already). This was also combined with the use of p-value of randomisation tests to assess the statistical significance of the low minimum free energy (MFE) displayed by potential pre-miRNAs. This involves randomly shuffling the sequence 1000 times, measuring the MFE of new structures formed and analysing the distribution of MFE values to assign a p-value to the MFE of proposed pre-miRNAs. The output from MiPred consists of classification of a predicted pre-miRNA as "pre-miRNA", "pseudo-pre-miRNA" or not a "pre-miRNA" with an associated percent confidence and a p-value of randomisation. Interestingly Jiang et al, (2007) also showed this p-value of randomisation to be the most important feature for distinguishing between real and pseudo pre-miRNAs.

A separate performance comparison between the CSHMM-Method and miPred (not to be confused with MiPred) showed that the CSHMM-method performed slightly better at identifying pseudo-pre-miRNAs (98% vs. 98.4%) (Agarwal et al., 2010). In the CSHMM-method, classification is based on the use of a context-sensitive-hidden-Markov-model to estimate the probability that a given hairpin is a real pre-miRNA. Its output consists of the predicted pre-miRNA structure and probability score. This is very different the Random-Forest machine learning approach used by MiPred, although both methods appear to be able to adequately differentiate between real and pseudo pre-miRNAs and may be best used in combination for identifying genuine viral pre miRNAs.

### 1.7.2 Experimental detection of novel viral miRNAs.

Ultimately, conclusive evidence of the existence of predicted miRNA genes can only be found through experimental evaluation. Common approaches to this are described below.

#### 1.7.2.1 *Small RNA Sequencing*

Sequencing of small RNAs from cells or tissue has provided an ideal way to identify novel non-conserved miRNAs. MiRNAs make up a very small percentage of total RNA (Masotti et al., 2009). Therefore, in order to improve the coverage of small RNAs and reduce background noise attributed to non-miRNA transcripts, samples are generally specifically enriched for miRNAs before sequencing. This is done by size fractionation through PAGE, where generally only RNA within the miRNA size range (17-25nt) is isolated.

Sample preparation involves an initial step to convert unknown single-stranded RNA (ssRNA) transcripts into a double-stranded DNA (dsDNA) sequencing library. This can be done in a number of steps regardless of the sequence of the ssRNAs present in the sample. It involves the ligation of specific adaptors to the 5' and 3' ends of all ssRNA species followed by RT-PCR using primers specific to the adaptors. T4 RNA ligase 1 is commonly used to carry out ligation of these adaptors. It will readily ligate ssRNA strands in the presence of Adenosine triphosphate (ATP) provided one strand has a 3' hydroxyl group (OH) (acceptor) and the other has a 5' phosphate group ($PO_4$) (donor). T4 RNA ligase 1 becomes adenylated by ATP resulting in the covalent attachment of adenosine monophosphate (AMP) to the enzyme. This AMP is then transferred to the $PO_4$ group of the donor. T4 ligase 1 then catalyses the attack of the adenylated donor by the OH group of the acceptor, resulting in the formation of a new phosphodiester bond and release of the AMP (Ho et al., 2004). For correct sample preparation it is vital that adaptor ligation occurs correctly, i.e. specific ligation of 5' and 3' adaptors to their respective ends only, without concatamerization or circularisation. The key to achieving this is through selective, reversible blocking of adaptor/RNA terminals (making them refractory to ligation). There are two approaches to this. Both involve carrying out 5' and 3' adaptor ligation

separately and are described by Pfeffer et al (2005a). The first approach, referred to as the Basic Protocol, involves dephosphorylation of ssRNA with alkaline phosphatase to remove 5' donor groups and then carrying out 3'adaptor ligation using an adaptor with a blocked 3' OH group. The resulting products (only consisting of ssRNA with 3' ends ligated to 3' adaptors) are then separated from non-ligated adaptors, the 5' donor group is re-introduced to the ssRNA through phosphorylation using T4 polynucleotide kinase and then 5' adaptor ligation takes place using an adaptor containing no 5'donor group. This results in ligation of the 5' ends of ssRNA to 5' adaptors only. The second protocol, referred to as the Alternate Protocol, involves 3' adaptor ligation in the absence of ATP using a 5' pre-adenylated 3' adaptor (again with a blocked OH group on the 3' end). As the adapter is pre-adenylated there is no need for ATP and the 3' adaptor can only ligate to the 3' end of the ssRNA. This circumvents the need for dephosphorylation and re-phosphorylation of the ssRNA. This process in outlined in Figure 1.10.

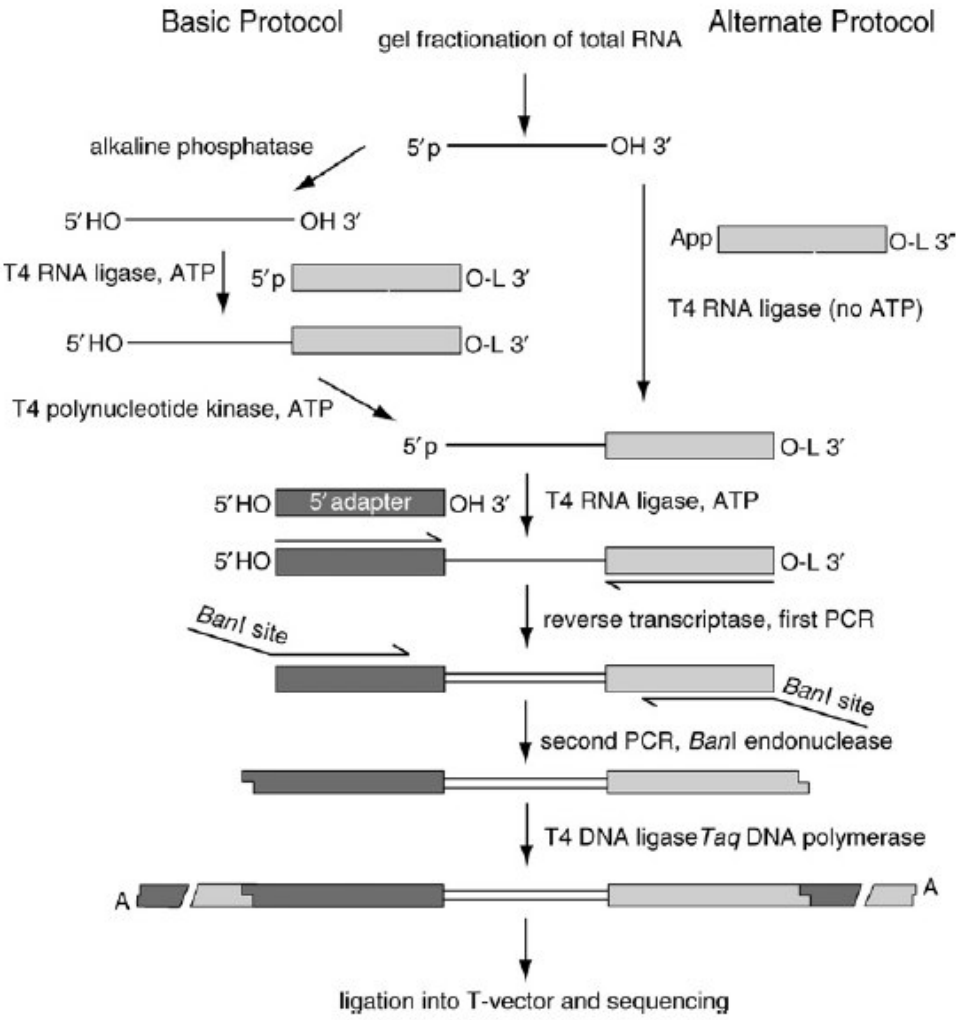


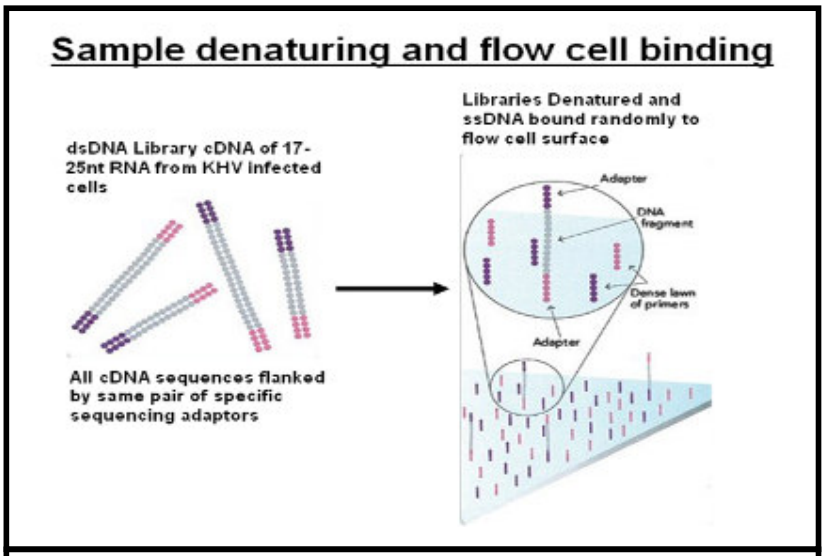

**Figure 1.10 Sequencing Library preparation**
Diagram taken from Pfeffer et al (2005a)

Although this can be done using T4 RNA ligase, a mutant of T4 RNA ligase, known as 'T4-Rnl2-truncated', is much more efficient at joining 5' pre-adenylated adaptors to 3' ends of ssRNA and in addition, unlike T4 RNA ligase 1, it does not have a tendency to adenylate the 5' end of ssRNA and therefore largely eliminates unwanted low-level ligation of 3' adaptors to 5' ends of ssRNA species (Viollet et al., 2011).

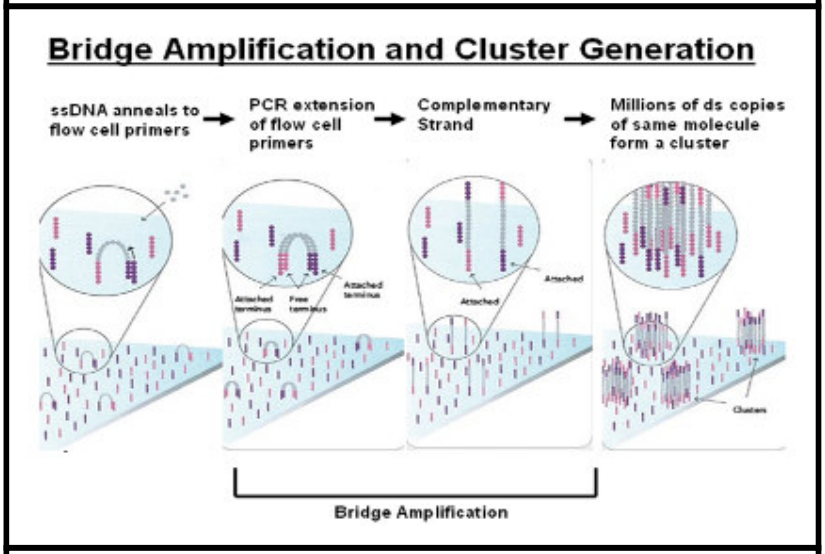
Small RNA sequences can be determined following their conversion to double-stranded cDNA. Initially, cloning of such products followed by low throughput dye terminator sequencing was a popular approach to miRNA discovery and it was through this approach that the first viral miRNAs were discovered (Pfeffer et al.,

2004). In recent times, high throughput deep sequencing has become the favoured approach to novel miRNA discovery due to the shear depth of coverage that it offers. It allows and unprecedented sensitivity in the profiling of RNA transcripts making it theoretically possible to detect every RNA species present in a given sample within a short time period. There are three main deep sequencing technologies in widespread use: the Illumina Genome Analyzer, the Applied Biosystems SOLiD Sequencer and the Roche/454 FLX Pyrosequencer These are comprehensively reviewed elsewhere (Mardis, 2008). The most popular technology used for miRNA discovery is Illumina Genome Analyzer (Bentley et al., 2008). This offers very high coverage of approx 20-25 million reads per sample lane and has been extensively used to identify novel herpesvirus miRNAs from infected cells (Glazov et al., 2010; Jurak et al., 2010; Meshesha et al., 2012; Reese et al., 2010; Riley et al., 2010; Umbach and Cullen, 2010; Umbach et al., 2010). Using Illumina technology for miRNA sequencing, samples are prepared in a similar manner to that described by Pfeffer et al (2005a) (using the Alternate protocol). The sequences flanking the small RNAs in the resulting dsDNA sequencing libraries are highly specific and facilitate the subsequent process of high throughput sequencing. This process is outlined in Figure 1.11.

(a)


Sample denaturing and flow cell binding
dsDNA Library cDNA of 17-25nt RNA from KHV infected cells
All cDNA sequences flanked by same pair of specific sequencing adaptors
Libraries Denatured and ssDNA bound randomly to flow cell surface
Adapter
DNA fragment
Dense lawn of primers
Adapter

(b)


Bridge Amplification and Cluster Generation
ssDNA anneals to flow cell primers
PCR extension of flow cell primers
Complementary Strand
Millions of ds copies of same molecule form a cluster
Bridge Amplification

(c)

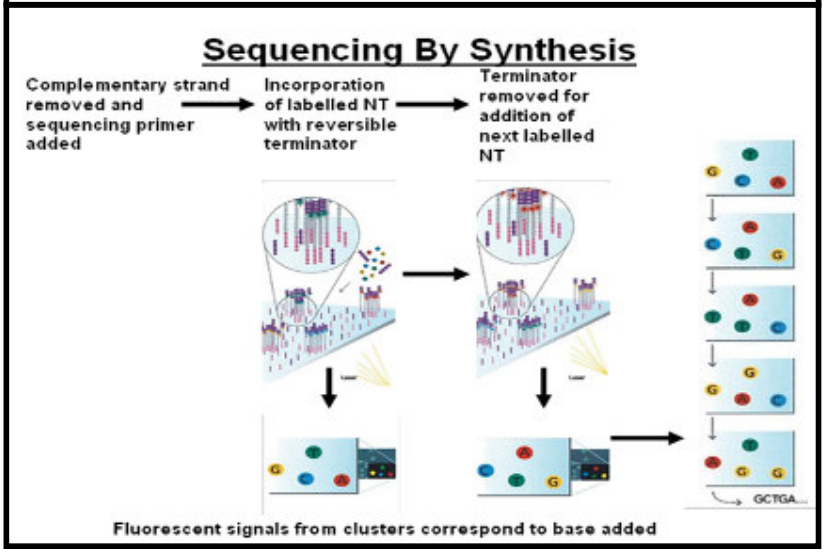

Sequencing By Synthesis
Complementary strand removed and sequencing primer added
Incorporation of labelled NT with reversible terminator
Terminator removed for addition of next labelled NT
Fluorescent signals from clusters correspond to base added

**Figure 1.11 Illumina sequencing process**
(a) Sequencing libraries are denatured and one end of individual ssDNA molecules becomes bound randomly all over the surface of the flow cell. The flow cell surface is also coated with a dense lawn of primers (also bound to flow cell on one end) of which they are 2 types, each complementary to one of the adaptor sequences. (b) The ssDNA template strands are allowed to anneal to these flow cell primers, and then these primers are extended through PCR to create a complementary strand. This is known as "bridge-amplification" and this process continues until the area immediately around the original template strand becomes populated with millions of double-stranded copies of this molecule to form a cluster. At the same time, millions of other individual clusters (representing other template strands with different sequences) are also formed all over the flow cell surface. (c) All complementary strands are removed by chemical cleavage of one flow cell primer (in this diagram it is the blue one). Then a sequencing primer (complementary to the remaining portion of the blue adaptor on the unbound end of the template strands) is added. Sequencing occurs through the extension of this sequencing primer with the incorporation of labelled nucleotides. Each nucleotide has a reversible terminator, so extension stops after one nucleotide is inserted at which point a high resolution image of the flow cell is taken. Each cluster gives a fluorescent signal corresponding to the base being added thus allowing identification of the first base added in each cluster. The terminator dyes are then removed allowing the addition of another fluorescent nucleotide and the flow cell is imaged again allowing the identification of the next base in each cluster. This process continues resulting in the generation of a series of images, each recording what base was incorporated in each cluster at each extension step. 'Basecalling' is done using an algorithm that monitors the change in emission colour from each cluster in each image/extension step, revealing the sequence of the ssDNA molecule that is generated in each cluster. Diagrams adapted from Illumina.com (2010)

### 1.7.2.2 *Analysis of Deep sequencing data*

Once small RNAs are sequenced they are mapped to a genome, using high throughput mapping software. Novel non-conserved miRNAs can then be identified by checking for the presence of predicted pre-miRNAs overlapping the same genomic loci as these small RNAs in the absence of any overlap with any other RNA species such as mRNAs, rRNAs, tRNAs etc. Small RNAs mapping to the stem in close proximity to the loop are designated as either 5' or 3' 'miRNAs' depending on the arm of the stem to which they map.

In addition to identifying the exact sequence of miRNAs derived from predicted pre-miRNA stems, sequencing can also supply additional levels of relevant information. For example, depending on the depth of coverage and expression levels of the transcripts themselves, it is often the case that many more small RNAs (other than miRNAs) will map to the same pre-miRNA, although the miRNAs will be the most abundant. IsomiRs associated with each miRNA (lesser abundant variations of the miRNAs display varying degrees of 5' and 3' terminal heterogeneity) can also be detected. Additional pre-miRNA-derived transcripts may also be observed mapping

to the stem immediately adjacent to the miRNAs, these are referred to as microRNA-offset-RNAs (moRNAs) and are a separate functional class of small RNA distinct from miRNAs (Shi et al., 2009). If miRNAs are identified from both arms of the pre-miRNA, this allows the prediction of the intermediate mature-miRNA-duplex structure, which should have characteristic 3' overhangs. Quantitative information regarding the frequency of each transcript in the data is also quite useful and allows designation of major form/minor form strands from a given pre-miRNA among other things.

All of these additional layers of information beyond simply identifying the method mature miRNA sequences themselves are very important. They help distinguish between miRNAs and other small transcripts in the sequencing data and are especially relevant to the process of identifying novel non-conserved miRNAs from such data. For this, it is important to also take into account the features of all the small RNAs that have mapped to pre-miRNAs and also the features of the pre-miRNA themselves as a whole. With this in mind, recent studies have aimed to define suitable criteria for assessing such features in order to aid the correct interpretation of miRNA and non-miRNA signals in deep sequencing data (Chiang et al., 2010; Friedländer et al., 2008; Hendrix et al., 2010). As a result, a general consensus has started to materialize. These were summarized by Kozomara and Griffiths-Jones (2011):

1. Putative miRNAs must occur multiple times in the data (typically read counts of 10-20 are used) and ideally should be found in more than one experiment.

2. Putative miRNAs should map to regions that are predicted to give rise to stable pre-miRNA-like secondary structures (ideally this would also incorporate the use of information about predicted pre-miRNAs from *de novo* prediction methods and pre-miRNA classifiers). Reads mapping to many regions in the genome should be eliminated.

3. Putative miRNAs should not overlap with any other RNA species such as mRNAs, rRNAs, tRNAs etc as it is difficult to prove that they do not represent degradation fragments from such transcripts.

4. IsomiRs of miRNAs may display heterogeneity on both ends relative to the start and end positions of miRNAs although their 5' ends should display significantly less heterogeneity.

5. Ideally both 5' and 3' arm miRNAs should be identified from a predicted hairpin and should form a mature-miRNA-duplex that displays the appropriate 3' overhangs.

In addition to the points mentioned above, Kozomara and Griffiths-Jones (2011) highlighted the fact that alignment signatures of miRNAs mapping to pre-miRNAs are distinctly different from that of non-miRNA small RNAs mapping to other parts of the genome. This is because pre-miRNAs are cleaved up in a very specific manner by RNase III enzymes Dicer and Drosha. This results in reads mapping to the pre-miRNA sequence in a "stack-like" pattern, with each stack representing a distinct type of small RNA derived from processing of the pre-miRNA i.e. 5' and 3'miRNAs and associated isomiRs (and 5' moRNA and/or 3'moRNAs if present), with little or no reads mapping to the loop. In addition more pronounced 3' end heterogeneity should be apparent in such alignment profiles. By contrast, small RNA reads that represent random degradation products from larger transcripts are more likely to map to genomes in the form of overlapping reads offset across the entire length of the degraded transcript due to the fact that their degradation is an entirely random process (Figure 1.12)

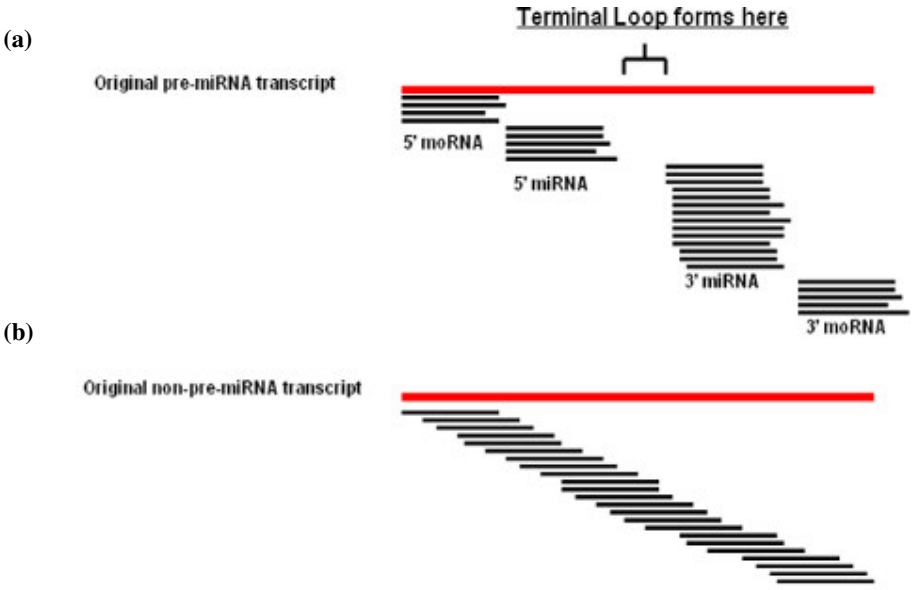



**Figure 1.12 Examples of miRNA-like and non-miRNA-like alignment signatures**
(a) miRNA-like alignment signature (b) non-miRNA-like alignment signature

Deep sequencing generates large amount of data, for example each flow cell from an Illumina genome analyser will produce ~25 million reads. While deep sequencing is a very useful tool, analysing such data for the presence of signals associated with novel non-conserved miRNAs is an enormous task in the context of eukaryotic organisms in contrast to viral genomes, which are much more manageable. Due to this, researchers have found it necessary to develop automated methods to "mine" deep sequencing data for novel non-conserved miRNAs. In recent years several algorithms have been developed to address this problem. They all start off by assessing the genomic locus of mapped small RNAs for the presence of pre-miRNA like hairpins. They differ however in the methods they use to classify small RNAs as being genuine miRNAs or not. They work in similar ways to *de novo* prediction methods (discussed earlier) but in this case the algorithms only assess sequences flanking loci that have small RNAs mapping to them. Some also cross-reference these small RNAs with known mature miRNAs and pre-miRNAs in order to infer the existence of new evolutionarily conserved miRNAs, or known miRNAs from the organism in question. Details of several automated miRNA identification methods are given in Table 1.7.

**Table 1.7 List of methods for the automated identification of miRNAs from deep sequencing data**

| Method | Approach | Reference |
|---|---|---|
| MiRDeep | See below | Friedländer et al. (2008) Updated to miRDeep2 by Friedländer et al. (2012) |
| miRExpress | Aligns reads to known miRNAs from same organism (i.e. identifying known miRNAs) and miRNAs form other organisms (i.e. identifying or novel conserved miRNAs) | Wang et al. (2009) |
| Mireap | See below | No publication associated with its development. First used by Zhang et al (2009). Wang et al, (2011) referred to it as "developed" by them. Always referenced as http://sourceforge.net/projects/Mireap/ (2013) |
| MIReNA | Uses combinatorial rules to identify pre-miRNAs from mapped reads. | Mathelier and Carbone (2010) |
| miRTRAP | Takes into account features of miRNA biogenesis. In addition, it takes into account the prevalence of small RNAs from anti-sense loci on the basis that some miRNAs will have either perfectly matched antisense RNAs, or none, but never offset antisense small RNAs. It also takes into account reads from flanking loci, as the presence of moRNAs can support miRNA predictions in some genomes (in particular *Caenorhabditis elegans*). | Hendrix et al.(2010) |
| mirTools | Classifies small RNA reads from deep sequencing as known miRNAs, ncRNA, mRNA, genomic repeats, annotates known miRNAs present, predicts new miRNAs using miRDeep core algorithm. Very useful for well-annotated genomes. | Zhu et al., (2010a) |
| DSAP | Web based tool. Filters out reads derived from non-miRNA known ncRNA and miRNAs Allows comparative miRNA expression analysis and cross species comparative analysis. Clustered sequences (possible miRNAs and isomiRs) that do not match to ncRNA or known miRNAs can be downloaded and analysed further elsewhere. | (Huang et al., 2010) |
| miRanalyzer | Finds novel miRNAs by first filtering reads through MirBase and RefSeq. Then maps remaining reads to genome, checks for pre-miRNAs and used Random Forest model to predict probability of candidates being real miRNAs. Also takes into account evidence from multiple experiments to build consensus on miRNA and pre-miRNA annotation. | Hackenberg et al., (2011) |

These kinds of programs differ slightly in their approach and performance in terms of identifying known and novel miRNAs from deep sequencing data. The performances of several of these programs have been directly compared following analysis of the same datasets (Li et al., 2012). The programs included in this comparison study included miRDeep, Mireap, MIReNA, miRTRAP, miRanalyser, miRExpress, DSAP and mirTools. Overall it was found that Mireap and miRDeep were the most suitable for detection of novel miRNA signals in sequencing data from vertebrates (Li et al., 2012).

The miRDeep core algorithm was designed to accurately identify miRNAs within deep sequencing data. It looks for stacks of reads (miRNAs and isomiRs) mapping to the same genomic loci and excises the flanking sequence for further analysis. MiRDeep employs a probabilistic model of miRNA biogenesis to assess miRNA candidates. It takes into account read frequency, the positions occupied by small RNAs on predicted pre-miRNAs and pre-miRNA secondary structure. Firstly it recognises that miRNA alignment signatures should take the form of a stack-like pattern and that all small RNAs derived from genuine pre-miRNAs need to fall within one of these stacks (it tolerates a small amount of deviation from this model due to expected inconsistencies in miRNA processing). It takes into account features of pre-miRNA structure such as the extent of base-pairing between the mature and star miRNAs. In a similar way to MiPred (Jiang et al., 2007) it also assess the statistical significance of the MFE displayed by the pre-miRNAs by assigning a p-value of randomisation to it. If candidate miRNA genes do not fit the model they are discarded and remaining sites are scored based in their overall consistency with the model. In addition, MiRDeep also estimates the false positive rate by carrying out statistical analysis of the classifications made by the mirDeep core algorithm. A new improved version of miRDeep, called miRDeep2 has also recently been developed (Friedländer et al., 2012). One of the improvements includes the recognition that micro-RNA-offset RNAs (moRNAs) may also cause extra stacks in the alignment signature if present. Steps have been taken to ensure that their presence does not interfere with the way that the algorithm defines the boundaries of potential pre-miRNAs which was initially based on scanning for 1-3 stacks of reads mapped to the genome (i.e. 5' and 3' miRNA and possible loop). Specifically the improvement stops the occurrence of asymmetrically excised pre-miRNA sequences that were a

consequence of moRNAs not being taken into account (having not been discovered at the time) in the initial model. It is important to note that miRDeep does not identify moRNAs; it just ensures that they do not interfere with appropriate sequence excision if present. More information on small improvements to on the first version of MiRDeep are available in Friedländer et al. (2012).

Mireap is also designed to look for novel miRNA signals within data from deep sequencing. Although it performed quite well in this respect when compared to other programs there is no information available as to how it was developed or how it works (Li et al., 2012). Despite this, it has a proven track record and has been used in many other studies since it was first used by Zhang et al (2009)

An important factor in novel miRNA annotation is read frequency. Where miRNA enriched samples are used, it is expected that the most abundant reads in such samples are miRNAs. Therefore if transcripts map to predicted pre-miRNAs in a miRNA-like pattern the presence of high read counts (or read counts over background levels) strongly support annotation of these genes as miRNAs (Cullen, 2011; Kozomara and Griffiths-Jones, 2011). Although the qualitative and quantitative data from deep sequencing is ideal for identifying novel miRNAs, deep sequencing has a tendency to suffer from enzymatic bias (Hafner et al., 2011; Sorefan et al., 2012). This is mainly due to the fact that sample preparation involves a series of enzymatic steps, each introducing its own bias towards or against specific transcripts depending on sequence content and resulting in over- or under representation respectively, in the sequencing data. As small RNA abundance is an important aspect of miRNA identification in sequencing data it is important that quantitative observations from deep sequencing experiments are verified using methods that are more quantitatively reliable (e.g. RT-qPCR) or are free from enzymatic bias (e.g. DNA microarray hybridization).

#### 1.7.2.3 *DNA microarray hybridization*

DNA microarray hybridization is a very useful high throughput approach to analyse gene expression. There are two varieties of arrays, spotted and oligonucleotide. Spotted arrays are made by first synthesising oligonucleotides or cDNAs

complementary to target transcripts followed by their deposition or "spotting" onto the array surface. In the case of oligonucleotide arrays, cDNA probes are synthesised *in-situ* on the array surface. (Pease et al.(1994) and Nuwaysir et al.(2002)). Hybridizations with oligonucleotide arrays are much more sensitive and for this reason were used in this study (see diagram of oligonucleotide array and detection process in Figure 1.13). In studies on gene expression, samples can consist of cDNA (RNA samples reverse-transcribed for labelling purposes or in order to make them more stable), amplified cDNA or aRNA (amplified RNA) generated through *in vitro* transcription from cDNA due to the incorporation of a T7 promoter into the PolyT RT primers. Labelling can be done through the direct incorporation of labelled nucleotides during cDNA or aRNA preparation. Alternatively, the incorporation of modified nucleotides amenable to subsequent labelling is also a popular method (Do and Choi, 2007). Labels are usually cyanine dyes such as Cy3 and Cy5. Cy3 fluorescence occurs in the yellow-green spectrum (emission peak ~570 nm) and Cy5 fluorescence occurs in the red spectrum (emission peak ~670nm).

DNA microarrays are an effective and sensitive tool for high throughput expression profiling of miRNAs (Baskerville and Bartel, 2005; Liu et al., 2004; Miska et al., 2004; Thomson et al., 2004) as well as miRNA discovery (Barad et al., 2004; Bentwich et al., 2005) including viral miRNA discovery (Grundhoff et al., 2006; Yao et al., 2012). The general sample preparation methods mentioned above work well for mRNA analysis, however they are not directly applicable to miRNAs due to the latter's small size and lack of a common tag (such as a polyA tail). This has been overcome through ligation of 3' adaptors containing T7 promoters to miRNAs (Barad et al., 2004) allowing RT and subsequent generation of labelled complementary RNA (cRNA). Methods involving direct enzymatic labelling of miRNAs have also been developed (Shingara et al., 2005).

When using DNA microarrays to detect miRNAs, it is best to use RNA that is enriched for miRNAs i.e. RNA <200nt in size (Barad et al., 2004; Bentwich et al., 2005). This enrichment is necessary because miRNAs make up a very small percentage of total RNA (Masotti et al., 2009), and using such samples allows definitive distinguishing of miRNA signals from background. This is particularly important where novel non-conserved miRNA exist as their expression levels are

unknown and may be low. In such experiments probes are usually designed on the basis of non-conserved pre-miRNA predictions (Bentwich et al., 2005; Grundhoff et al., 2006). As pre-miRNAs are readily processed by cellular miRNA processing systems, fully intact pre-miRNAs can often be in extremely low abundance and it has been shown that DNA microarray hybridization is not a suitable method for detecting such transcripts even where samples were enriched for both miRNA and pre-miRNAs (Barad et al., 2004; Grundhoff et al., 2006). Conversely, the smaller fully processed mature miRNAs derived from these pre-miRNAs can accumulate in higher levels once incorporated into the RISC and are therefore much more abundant and more readily detectable on arrays. For this reason, where novel non-conserved pre-miRNAs are predicted and investigated by array analysis, probes are always designed to detect theoretical mature miRNAs derived from them, thus they are simply complementary to the stems of predicted pre-miRNAs (Bentwich et al., 2005; Grundhoff et al., 2006; Sullivan and Grundhoff, 2007). VMir is an example of a program that is used both to predict pre-miRNAs in viral genomes and also to design appropriate array probes in this manner as a means of facilitating high throughput experimental evaluation of predictions (Grundhoff et al., 2006). Importantly it designs control probes for such experiments and also analyses exported raw fluorescence data from subsequent experiments.

Pre-miRNAs will generally not contribute to signals from such probes and so in some cases RNA within the pre-miRNA size range is even excluded in order to prepare samples that are specifically enriched for miRNAs. In fact, samples consisting of RNA <40 nt in size (Shingara et al., 2005) or even RNA exclusively within the miRNA size range of 17-25 nt (Grundhoff et al., 2006; Sullivan and Grundhoff, 2007) are best for use in miRNA expression analysis and/or miRNA discovery using arrays.

Modifying RNA samples through enzymatic means prior to analysis will introduce certain biases into the results obtained, especially if PCR is involved. However sometimes this is a necessary part of the preparation process e.g. where sample quantity is limited (Livesey, 2003). If the latter is not an issue, enzymatic bias can be avoided by direct labelling of RNA (not aRNA or cRNA) through chemical methods where dyes are covalently attached to nucleotides (Wiegant et al., 1999). This

approach is more straightforward and flexible as it is applicable to both mRNA and miRNA. If enzymes can be avoided then array hybridization presents us with an ideal high throughput way to verify quantitative information generated by other methods that may suffer from enzymatic bias, such as deep sequencing.

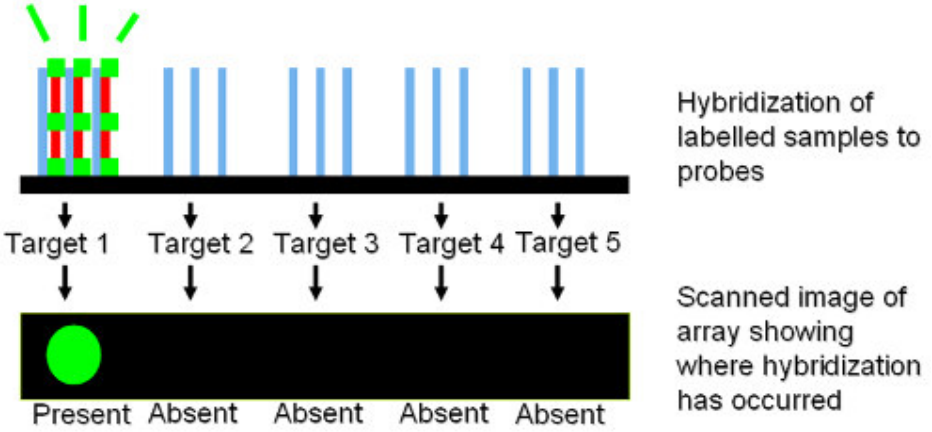



**Figure 1.13 The DNA microarray hybridization process**
Labelled RNA/cDNA/aRNA is in red, labels are in green and oligonucleotide probes are in blue

### 1.7.2.4 *Northern blotting*

Although a less sensitive and lower throughput technique when compared to deep sequencing or DNA microarray hybridization, northern blotting remains one of the best ways to conclusively demonstrate the presence of novel miRNAs in a biological sample. This is due to the fact that it reveals the size of all RNA transcripts containing the target miRNA sequence across all RNA size ranges. Detection of *bone fide* miRNAs through northern blotting potentially allows the detection of two other distinct transcripts associated with miRNAs, namely the pre-miRNA and the pri-miRNA. All of these thee transcripts may be detected simultaneously in the same sample and appear as three discrete bands, in their expected size ranges. Typically it is not possible to estimate the size of the pri-miRNA on the same gel as the miRNA and pre-miRNA (as the gels used are designed to resolve lower molecular weight RNA) although it is possible to compare the size of pre-miRNA and miRNA bands to their predicted sizes. The miRNA should be detected as a band in the 17-25 nt size range and the pre-miRNA should be typically in the 50-100 nt size range. Importantly, the presence of discrete bands in the absence of smearing indicates that the small RNAs being targeted are not the result of random degradation of larger

RNA molecules. Sometimes the miRNA will be detected alone in the absence of pre-miRNAs or pri-miRNAs. Sometimes, somewhat counterintuitive to what would be expected, it is also possible to detect pre-miRNAs in the absence of miRNAs (Cui et al., 2006; Munson and Burch, 2012). Also in addition to the ~22 nt miRNA it is sometimes possible to detect additional bands (within the 17-25nt range), representing highly abundant isomiRs of the miRNA (Cai et al., 2005).

## 1.8 MiRNA Target prediction

MiRNA target prediction is a useful way of inferring miRNA function and often acts as a basis for laboratory studies. The basic requirement of a potential target site is that it has a perfect complementary match to the seed region. Although looking for seed pairing is important, this criterion is not stringent enough, and on its own results in a high rate of false positives. Based on the characteristics of experimentally verified miRNA target sites, it has become apparent that it is necessary to also take into account other subtle criteria in order to improve target site prediction. These criteria relate to both the target sites themselves and surrounding sequence and include features such as seed pairing stability, 3' pairing contribution (pattern of base pairing outside the seed), local AU content, target site abundance and even position in within a 3' UTR (Garcia et al., 2011). Assessing potential target sites using these additional criteria is key to improving predictions and to help distinguish genuine miRNA target sites from background noise. Most miRNA target site prediction methods now take such additional features into account for this purpose and details of several of these methods are outlined in Table 1.8.

**Table 1.8 List of miRNA target prediction methods**

| Method | Approach | Reference |
|---|---|---|
| DIANA-microT | Slides a window along 3' UTRs, at 1 nt increments; in each window it calculates the stability of miRNA binding (allowing mismatches) and compares it to the stability of a perfectly complementary sequence. It searches for miRNA/mRNA duplexes with canonical structure (i.e. 7-9 nt match in 5' miRNA, 6 nt match to seed, central bulge and supplementary pairing at the 3' in miRNA) It also estimates signal-to-noise ratio based on randomized miRNA sequences, give a probability score to each result and takes into account evolutionary conservation of target sites. | Kiriakidou et al. (2004) |
| miRanda | Looks for seed complementarity, seed mismatches and gaps. It examines miRNA/mRNA duplex stability and the extent of compensatory base pairing outside seed regions. It adds extra weight to predictions for miRNAs with multiple target sites in the same 3' UTR. Also takes into account evolutionary conservation of target sites. | John et al. (2004) |
| TargetScan | See below | Lewis et al. (2005), updated by Grimson et al. (2007) and Garcia et al. (2011) |
| PicTar | Looks for highly complementary regions in highly conserved parts of 3' UTRs. It examines the stability of miRNA/mRNA duplexes and scores results using a Hidden Markov Model, favouring miRNAs with multiple target sites in the same 3' UTRs and takes into account the conservation of target sites. | Krek et al. (2005) |
| RNA22 | Unlike other methods RNA22 first locates potential miRNA binding sites in 3' UTRs by looking for "hot-spots" that display certain patterns and then matches this site to an appropriate targeting miRNA and makes decisions on appropriate matches based on min/max base-paired/in-paired nucleotides in potential duplexes and stability. Prediction does not rely on conservation. | Miranda et al. (2006) |
| PITA | See Below | Kertesz et al. (2007) |

Although many approaches to miRNA target prediction have been developed, the high degree of heterogeneity in miRNA/mRNA interactions means that there is no perfect model to accurately predict all types of miRNA target sites. However, some have been shown to work quite well. TargetScan (Garcia et al., 2011; Grimson et al.,

2007; Lewis et al., 2005) is a popular "rule-based" approach for predicting miRNA target sites that also takes into account target site conservation, making it directly applicable to finding target sites for eukaryotic miRNAs. Using the web interface, the user is restricted to searching for targets to known miRNAs in 3' UTRs of a small variety of model organisms. Custom miRNA (i.e. novel miRNA) target searches can also be done, but searches are restricted to 3' UTRs from the same small range of model organisms. This approach is not suited to looking for targets for a high number of novel miRNAs in a custom dataset of 3' UTRs. However the source code for the algorithm is freely available allowing complete customisation of the process to suit the needs of specific studies. In addition, an analysis can also be done without needing to take conservation into account. Overall therefore TargetScan is a flexible method that can be applied to wide range of studies. It ranks predicted sites by assigning a context score to them based on how well they conform to the ideal characteristics of target sites. Experimental evaluation of predictions has shown good correlation between context score and protein down-regulation (Baek et al., 2008)

There are two main algorithms in the TargetScan pipeline, namely TargetScan60 and TargetScan60-Context-Scores. The output from the former is used as input for the latter. TargetScan60 performs a straightforward search for target sites in 3'UTRs based on matches to seed regions. In its simplest form, a match consists of a sequence complementary to the six bases of a seed region (bases 2-7). Sites with such straightforward seed matches are known as 6-mer sites. Sites complementary to bases 2-8 are considered more significant than 6-mer sites and are referred to as 7-mer-m8 sites. In *bone-fide* target sites, there seems to be a preference for the presence of a conserved adenine on the 3' end (opposite position 1 on the miRNA), sites with straightforward seed matches that also have this feature are called 7-mer-A1 sites. Furthermore, 7-mer-m8 sites that also have an A at their 3' ends are known as 8-mer-1A sites. Understandably some of these types of target sites perform better than others and as such they display the following hierarchy of site efficacy: 8mer-1A > 7mer-m8 > 7mer-1A > 6mer (Grimson et al., 2007). Six-mer sites are by far the least effective and are typically identified by chance more frequently than other types of sites. For this reason, TargetScan60 disregards 6-mer sites in searches. The output consists of information such as the name, site-type (8mer-1A, 7mer-m8 or 7mer-A1) and position of the 3' UTR containing the possible target site. The output can then be

used for a more detailed analysis by another algorithm called TargetScan-Context-Scores.

The second program in the pipeline is called TargetScan60-Context-Scores. This algorithm takes the predicted target sites from TargetScan60 and analyses them further based on the stability of the pairing between the miRNA and target site as well as taking into account its local sequence context. Sites are given a context score and ranked based on their consistency with the characteristics of *bone fide* miRNA target sites. Characteristics that contribute to the context score include.

- Seed-Pairing-Stability (SPS): There is a correlation between greater SPS and target site efficacy and this is taken into account when ranking potential target sites (Garcia et al., 2011).

- 3' Pairing Contribution: Additional contributions to overall miRNA-target stability are made through base pairing outside the seed region. There is a pattern to this type of base pairing. It has a tendency to be contiguous, uninterrupted by bulges, wobbles or mismatches. It also seems to mainly involve bases at positions 13–16 on the miRNA in particular. In fact, pairing at these positions is found to be more important for efficacy than pairing anywhere else in the 3' end of the miRNA, (Grimson et al., 2007). Base pairing at the 3' end can supplement SPS or compensate for weak SPS. When scoring predictions TargetScan60-Context-Scores looks specifically for these kinds of patterns of base pairing at the 3' end of miRNAs.

- Local AU content: There is a strong preference for high AU content in sequences immediately flanking the miRNA target site with this high AU content tailing off ~30 nt upstream and downstream of the target site (Grimson et al., 2007)

- Position in 3' UTR: Target sites tend to distributed in a specific pattern within 3' UTRs with most known target sites clustered at both ends of 3' UTRs rather than in the centre (Hon and Zhang, 2007). Despite the fact that many

target sites can be predicted in the centres of 3' UTRs, sites occurring near the beginning (excluding the first 15 nt) or near the end of 3' UTRs are generally much more effective and therefore score higher (Grimson et al., 2007).

- Target Abundance (TA): This is the number of times that the same target site occurs in other 3' UTRs. The lower the TA the better, as miRNAs will not be spread as thinly between many target transcripts which may reduce their overall efficacy. This is subjective of course, considering differences in (and fluctuations of) miRNA expression levels (Garcia et al., 2011).

- Seed match type: Site type also contributes to the score based on the following order of efficacy -8mer-1A > 7mer-m8 > 7mer-A1

The output from TargetScan60-Context-Scores consists of a summary of each target site in terms of how is scores with regard to the characteristics described above. The sum of all of these scores is called the context+ score with a lower value here meaning a better quality match.

Although TargetScan looks for many important attributes when predicting potential miRNA target sites allowing it to make high quality predictions, sometimes high confidence target sites do not have the expected mRNA knockdown efficacy when tested experimentally while other lower ranked sites might perform better. This is often due to target site accessibility (or lack thereof). This is because 3' UTRs are capable of forming complex RNA secondary structures. Sometimes the target site can be incorporated into a stable secondary structure, or sometimes if not then the formation of complex structures surrounding the target can physically prevent the RISC from accessing the site. Another method of miRNA target prediction known as PITA addresses this problem by taking target site accessibility into account . It works by initially identifying target sites based on seed region matches. It does this by simply looking for 6-8 nt matches starting from position 2 on the miRNA (base-pair wobbles are allowed as an option). Once a target site is identified the secondary structures that it may form (and secondary structures of regions 70 nt upstream and downstream) are then predicted and the minimum free energies of these structures

are calculated. The minimum free energy of the miRNA-target site base pairing is also calculated. The algorithm then compares the potential energy lost by opening these secondary structures ($\Delta G_{Open}$) to the potential energy gained by binding of the miRNA to the target site instead ($\Delta G_{Duplex}$). This is done by calculating $\Delta G_{Duplex}$-$\Delta G_{Open}$ and the difference is referred to as $\Delta\Delta G$. If this value is still negative then there is more energy to be gained by miRNA-target base pairing and thus this is the more energetically favourable outcome, see Figure 1.14 below.

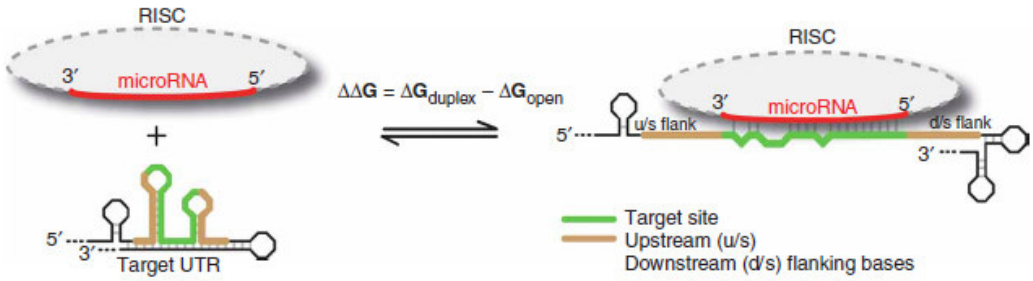



**Figure 1.14 Principles behind miRNA target prediction using PITA**
Diagram illustrating the transition between unbound miRNA/mRNA duplex and bound miRNA/mRNA duplex. The free energy gained from this transition is calculated using the equation shown in the middle. Diagram taken from (Kertesz et al., 2007)

The PITA output consists of a summary of each predicted target site including co-ordinates, minimum free energies of structures formed, target-miRNA pairing and $\Delta\Delta G$. The lower the $\Delta\Delta G$ the more energetically favourable the miRNA-target interaction is compared to the formation or persistence of structures at or around the target site. If a miRNA has several target sites on the same 3' UTR, PITA calculates the overall miRNA-UTR interaction score for that site. The overall miRNA-UTR interaction score is equal to as $\Delta\Delta G$ if a miRNA targets only one site on a given 3' UTR.

## 1.9 Aims of this study

In recent years viral miRNAs have been shown be important regulators of both host and viral genes during both lytic and latent infections (Kincaid and Sullivan, 2012) . In particular, they have been found to be expressed by some members of the *Herpesviridae* family. Current evidence suggests that the presence of miRNA-coding genes in some family members may be due to their distinct biological characteristics

(Section 1.6.1.2). Although a distantly related grouping, members of the *Alloherpesviridae* family display many of the same biological characteristics as the *Herpesviridae* (Section 1.3). Thus, it is likely that they also encode miRNAs of their own. Based on the general lack of miRNA conservation among viruses, any such miRNA-coding sequences are also likely to be non-conserved, although they may have analogous functions (Section 1.5). Interestingly, it is equally possible that they have completely different functions and act in different ways to promote and/or control various stages of the viral life cycle. In light of this, a study to identify miRNAs encoded by members of the *Alloherpesviridae* family is therefore justified and may help further our understanding of the biology of these viruses.

CyHV-3 is a highly pathogenic member of the *Alloherpesviridae*. Its ability to establish latent infections in response to non-permissive temperatures has been well-documented, although the characterisation of the molecular basis of CyHV-3 latency is still in its infancy (Section 1.4.10.). Due to their strong association with latency in other HVs, the identification of CyHV-3-encoded miRNAs might provide useful insights into how CyHV-3 modulates viral and/or host gene expression during latency. It is likely that such transcripts would also have roles during lytic infections, thus also furthering our understanding of the molecular basis of CyHV-3 infections in general. Knowledge of CyHV-3 transcripts expressed during latency may enable the development of more sensitive and suitable diagnostic methods to identify latent carriers (Section 1.4.10.6).

The first step towards this is the identification of CyHV-3-encoded miRNAs and hence this was the main aim of this study. The primary samples used here to identify CyHV-3 miRNAs consisted of RNA prepared from *in vitro* lytic infections, facilitating the identification of miRNAs expressed during the virus production cycle and that may also remain expressed during latency (as is the case with most other HV miRNAs).

First, the presence of pre-miRNA-coding sequences were predicted following analysis of the CyHV-3 genome, thus providing a theoretical basis for the study. These predictions were then experimentally evaluated using various methods to identify and characterise the candidate CyHV-3 miRNAs. These methods included

deep sequencing, DNA microarray hybridization, northern blotting and reverse transcription real-time PCR (RT-qPCR). The end goal of this study was to use RT-qPCR to test for the expression of these miRNAs in various tissues from lytically infected fish and also in long-term healthy survivors of high mortality outbreaks (i.e. fish that are likely to be latent carriers). Finally, CyHV-3 miRNAs were further characterized in terms of sequence conservation and potential functions. The workflow followed is outlined in Figure 1.15.

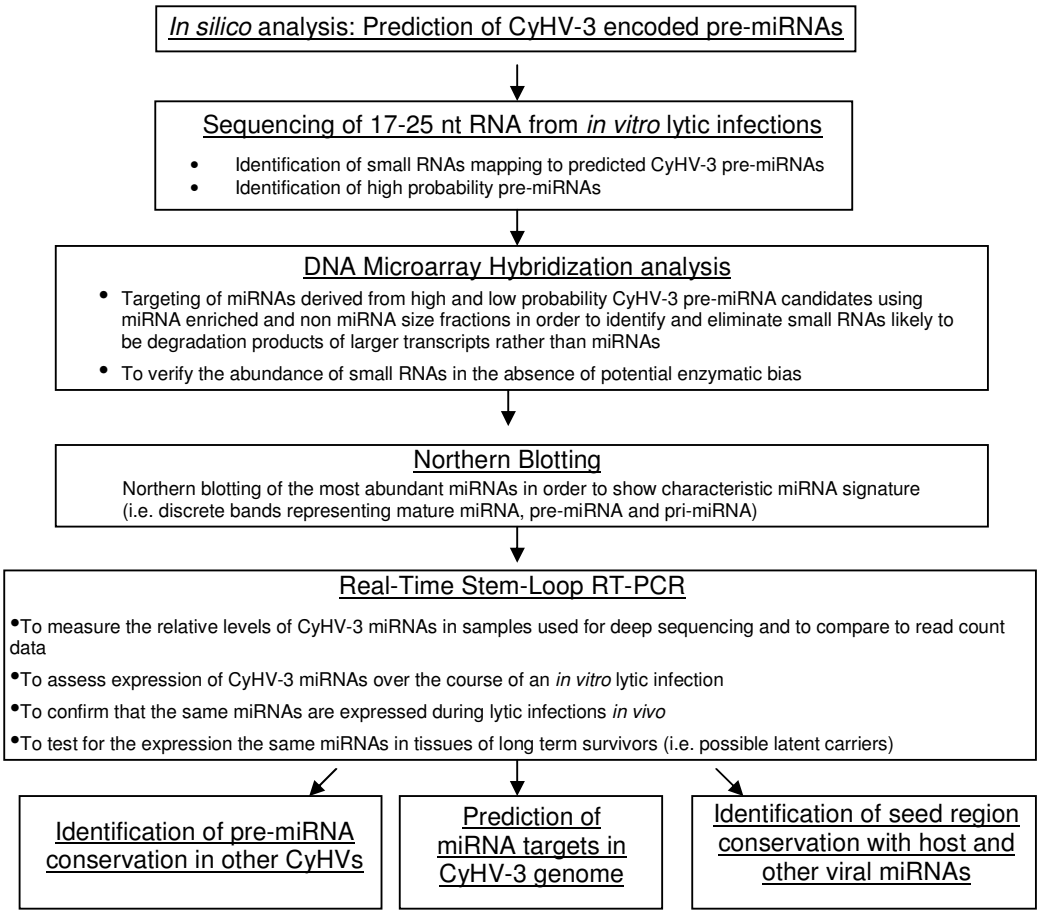



**Figure 1.15 Outline of the workflow used for the identification and characterisation of CyHV-3 miRNAs.**

# 2 Materials and methods

## 2.1 Reagents

10 bp DNA Ladder (Invitrogen 10821-015)
10 x Tris/Borate/EDTA (TBE) Buffer (Sigma, 93290-5L)
10x Tris Acetate-EDTA buffer (Sigma, T8280-1L)
2.5% Trypsin (Gibco, 15090-046)
2x Gel loading buffer (Ambion, AM8547)
40% Acrylamide 19 acryl: 1 bis-acryl (Ambion, AM9022)
50 mM Magnesium Chloride (Promega)
5M NaCl (Ambion, AM9760G)
Acetelated BSA (Promega, R3961)
Agarose (Sigma, A9539)
Ammonium persulphate (APS) (Sigma, A3678-100G)
BigDye® Terminator v3.1 Cycle Sequencing Kit (Life Technologies, 4337455)
BrightStar® BioDetect™ Kit (Ambion, AM1930)a
Chloroform (Lab-Scan, A07C11X)
Custom 5' Biotin labelled DNA probe (IDT)
Custom TaqMan Small RNA Assays, (Applied Biosystems, 4440418)
Deoxynucleotide Triphosphates (dNTPs) 100mM (Promega, U1240)
Dextran sulfate (Sigma, D8906)
Dimethyl sulfoxide (DMSO) ≥99.9% (Sigma, D8418-50ML)
Distilled Water
DL-Dithiothreitol (DTT) 100mM (Promega, P1171)
DyeEx 2.0 Spin Kit (Qiagen, 63204)
DynaMarker Prestain Marker for Small RNA Plus (Biodynamics, DM253S)
Ethidium Bromide 10 mg / mL (Sigma, E1510-10ML)
ExoSAP-IT (Affymetrix 78250)
Foetal Bovine Serum, Non-USA origin, sterile-filtered, cell culture tested (Sigma, F7524)
Formamide (Sigma, F9037)
Gel loading buffer (Sigma, G2526)
GlutaMAX™-I Supplement, 200 mM (Gibco, 35050-038)
GoTaq Flexi DNA Polymerase (Promega, M8306)
Hi-Di Formamide (Applied Biosystems, 4311320)
IPTG (Sigma, I6758)
Label IT® miRNA Labeling Kit, Version 2 (Mirus, Bio 9510)
Linear Acrylamide 5 mg / mL (Ambion, AM9520)
MBG Water (Sigma, W4502)
MegaClear Kit (Ambion, AM1908)
Minimum Essential Medium (MEM) Eagle in Earle's BSS, with 2,2g/L NaHCO3, 25 mM HEPES (1x) (Sigma, M72780)
miRNeasy Mini Kit (Qiagen, 217004)
Molecular biology Grade (MBG) Ethanol (Sigma, E7-148-1GA)
Molecular Biology Grade (MBG) Water (Sigma, W4502)
Non Essential Amino Acids (100X) (Gibco, 11140-035)
Penicillin-Streptomycin mixture, 5,000 units of penicillin (base) and 5,000 µg of streptomycin (base)/ mL (Gibco,15070-063)

pGEM®-T Easy Vector System II (Promega, M8306)
Phosphate buffered saline tablets (BR0014G, Oxoid)
Phusion High-Fidelity DNA Polymerase 2 U/ µL (Finnzymes, F-530S)
Polyethylene glycol (PEG) solution, 8,000, ~50% (Sigma, 83271-100ML-F)
QIAzol Lysis Reagent (Qiagen, 79306)
RNAlater (Ambion, AM7020)
RNase-Free DNase Set (Qiagen, 79254)
RNaseZap (Ambion, AM9780)
RNasin Plus RNase Inhibitor (Promega, N2615)
RNeasy MinElute Cleanup Kit (Qiagen, 74204)
Sodium acetate anhydrous (Sigma, S2889)
SuperScript III RT (Invitrogen, 18080-093)
T4 RNA Ligase I 10 U / µL (New England Biolabs,
T4 RNA Ligase II 200 U / µL (New England Biolabs, M0242S)
TaqMan Universal PCR Master mix (Applied Biosystems, 4364340)
TaqMan® miRNA Reverse Transcription Kit (Applied Biosystems, 4366596)
Tetramethylethylenediamine (TEMED) ( Sigma, T7024)
Tween 20 (Sigma, P9416-100ML)
Ultra Pure 1 M Tris-HCI, pH 8.0 (Invitrogen, 15568-025)
Ultra Pure Water
ULTRAhyb®-Oligo buffer (Ambion, AM8663)
UltraPure™ 10% SDS Solution (Invitrogen,
UltraPure™ 10% SDS Solution (Invitrogen, 24730-020)
UltraPure™ 20X SSC (Invitrogen 15557-044) or UltraPure™ 20X SSPE (Invitrogen, 15591-043)b
UltraPure™ 20X SSPE (Invitrogen, 15591-043)
Urea (Ambion, AM9902)
Versene 0.2 g/L (Gibco, 15040-033)
X-Gal (Sigma, B4252)

## 2.2 Equipment

0.45 µm syringe filter
0.5 ml thin walled PCR tubes
1 L Duran Flasks
1.8 mL Micro-centrifuge tubes
10 mL Sterile Syringe
10 x 10 cm vertical gel electrophoresis apparatus
100 mL Duran Flasks
1000 mL Corning® 0.22 µm vacuum filter system (cellulose acetate membrane) (Sigma, CLS430516)
15-22°C Incubators
2 L Conical Flask
2 mL Micro-centrifuge Safe-Lock tubes
20 mL sterile universals
-20°C Freezer
20-gauge needles
-20ºC Freezer
4ºC Fridge

5 mL pipette
5 mL sterile syringe
50 mL Sterile universals
7 mm stainless steel beads (Qiagen)
75cm$^2$ Cell Culture Flasks (Sigma, CLS430641-100EA)
-80ºC Freezer
ABI PRISM® 310 Genetic Analyzer
Applied Biosystems 7500 Real-Time PCR System
Autoclave (Touchclave)
Balance
BioLuminizer (CAMAG)
Blot-box 9,1x6,6cm 3-5ml capacity (VWR, 700-0249)
BrightStar®-Plus Positively Charged Nylon Membrane (Ambion, AM10102)
Cryovials (Nunc)
Electroblotting module and cassettes
Electronic sealer
Fibre Pads
Fume hood (Chem Flow 2000 Standard)
Gel Tank, Casts, Combs
Heating Block
Laminar Flow Cabinet (Heraeus Hera Safe 1200x650x630)
Masking Tape
Microwave Oven
Nanodrop Spectrometer
NovaRay Scanner (Alpha Innotech)
Nucleic acid and nuclease free pipette tips
Pipeteboy
Power Supply
Refrigerated centrifuge (Sorvall Legend RT)
Rocker
Rocker Platform
Sealable Hybridization Bags, Whatman® 20.3 x 25.4 cm (VWR, 89027-004)
Sterile forceps
Sterile scalpel
Sterile scissors
Tecan HS 400™ Pro Hybridization Station
Thermocycler
TissueLyser (Qiagen)
UV Transilluminator
Vacuum Pump and tubing
Water Bath
Water purification System (Millipore, Milli-Q plus)
Weighing Boats
Whatman 3MM Filter Paper 46 cm x 57cm (VWR, 514-8013)

## 2.3 Cell Culture

### 2.3.1 Preparation of cell culture mixtures/solutions

#### 2.3.1.1 *Cell Culture Media*

Cell culture media was prepared as per Table 2.1

**Table 2.1 Cell vulture media components**

| Component | Volume |
|---|---|
| Minimum Essential Medium (MEM) Eagle in Earle's BSS, with 2,2g/L NaHCO3, 25 mM HEPES (1x) | 500 mL |
| Foetal Bovine Serum, Non-USA origin, sterile-filtered, cell culture tested | 50 mL |
| Non Essential Amino Acids (100X) | 5 mL |
| Gibco GlutaMAX™-I Supplement, 200 mM | 5 mL |

The components were mixed in autoclaved 2 L conical flask. If necessary pH was adjusted by eye by adding 0.22 μM syringe filtered 1M NaOH or HCl until a neutral red colour persisted. The mixture was poured into a 1000 mL 0.22 μm vacuum filter system and using vacuum pump the mixture to be drawn through the filter into a sterile collection flask. Filtered media was incubated at 27°C for 3 days for sterility testing, and stored at 4°C immediately before being use 5 mL of Penicillin-Streptomycin mixture was added to the media to protect against possible bacterial contamination post filtration. Media was kept at 4°C when not in use.

### 2.3.2 Cells

Common Carp brain cells were originally obtained from the Collection of Cell Lines in Veterinary Medicine (CCLV) at the Friedrich-Loeffler Federal Research Institute for Animal Health, Greifswald Germany. They were received on cell passage number 55 at Cefas Weymouth Laboratory, Weymouth UK where they were maintained before being received at the Marine Institute, Galway Ireland on passage number 95.

### 2.3.3 Cell culture protocol

#### 2.3.3.1 *Aseptic Technique*

All cell culture work including preparation of mixtures and reagents were prepared in a laminar flow cabinet. All internal surfaces of the cabinet, reagent containers and equipment were cleaned with 70% EtOH before and after use.

#### 2.3.3.2 *Maintenance of Cell Line*

The CCB cell line was maintained in 75cm$^2$ cell culture flasks. Flasks were split every 3-4 weeks. Splitting was done in a laminar flow cabinet under aseptic conditions. Old media was aspirated off into the waste beaker. Cells were rinsed using 5 mL PBS to remove any residual media. PBS was removed and cells were trypsinized using 1 mL 0.1% trypsin. Trypsin was deactivated by the addition of 3 mL cell culture media. Cell clumps were broken up by pipetting. For a 1:4 split, 3 mL of suspended cells were removed (and either disposed of or used to seed new flasks, see Section 2.3.3.3) 24 mL of cell culture media was added to the 1 mL of suspended cells remaining in the flask. Flasks were put at 27°C for 24 hours to (give the cells a temporary growth boost) and were subsequently maintained at 20°C until split again.

#### 2.3.3.3 *Seeding new cell culture flasks (Up-Scaling cell stocks)*

Of the 3 mL of suspended cells removed from flasks during splitting, 1 mL was added to 3 new flasks and 24 mL of cell culture media was added to each flask. Flasks were put at 27°C for 24 hours to (give the cells a temporary growth boost) and were subsequently maintained at 20°C until split again.

## 2.4 Viral Culture

### 2.4.1 Virus

CyHV-3 used in this study was an isolate referred to as H361 supplied by Cefas Weymouth Laboratory, Weymouth, UK. It was received at passage 3 in the form of

an infected Common Carp Brain cell culture. The infected cell culture was received at the Marine Institute at 5 days post infection (dpi) and subsequently kept at 22°C until 17 d.p.i. when extensive cytopatic effects (CPE) could be observed. At this point viral supernatant was collected and stored.

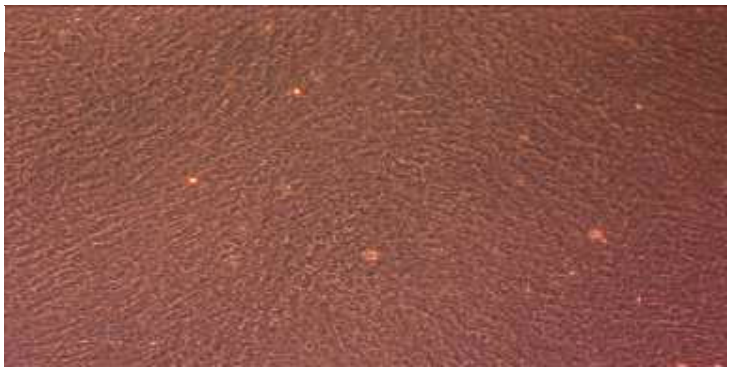
### 2.4.2 Viral culture protocol

Aliquots of viral supernatant were taken from storage at -80°C and thawed out on ice. 250-500 µL of viral supernatant was used to inoculate 1 day old flasks of CCB cells (~75% confluent after 24 hours at 27°C). Once inoculated, flasks were rocked gently to disperse the virus throughout the cell culture media. To establish a lytic infection, inoculated flasks were kept at 22°C until CPE appeared at ~7-12 d.p.i. (Figure 2.1).

### 2.4.3 Viral Supernatant collection

Once widespread CPE was visible within infected cultures the media was taken off and placed in a 20 mL sterile universal tube. This was spun down at 200 x g for 10 min at 4°C in order to pellet any suspended cells and cellular debris. The supernatant was drawn up into a 10 mL sterile syringe and filtered through a 0.45 µm syringe filter into a new 20 mL sterile universal. The supernatant was split into 1 mL aliquots in 1.8 mL cryovial tubes and stored at -80°C. These were used as sources of more infectious virus for subsequent viral culture.

**(a)**



**(b)**

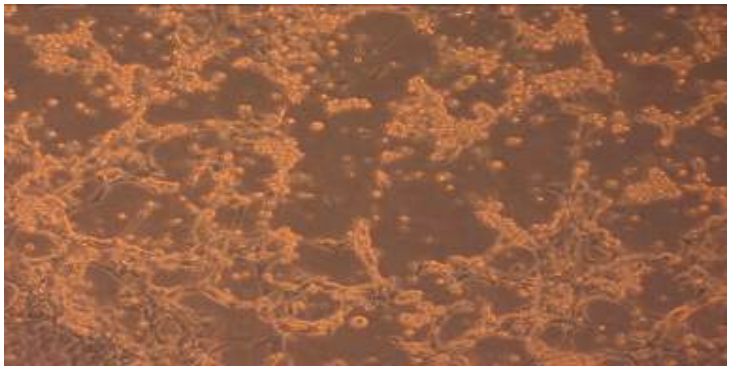


**Figure 2.1 An example of the Cytopatic Effects (CPE) caused by CyHV-3 during an *in vitro* lytic infection.**
(a) Example of healthy monolayer of CCB cells (b) example of widespread CPE characterized by the vacuolation and rounding up of cells and the formation plaques as large patches of the monolayer dethatch from the flask surface.

## 2.5 RNA Preparation

### 2.5.1 Preparation of RNA extraction mixtures/solutions

#### 2.5.1.1 *3 M Sodium Acetate*

To prepare this, 24.61 g of NaOAc anhydrous was added to 100 mL MBG water and allowed to dissolve at RT.

#### 2.5.1.2 *10 mM Tris-HCl*

To prepare this, 500 µL 1 M Tris-HCl was added to 49.5 mL MBG Water

### 2.5.2 RNA Preparation from biological matrices

#### 2.5.2.1 *Sample Lysis*

***Cells***

Cells were trypsinized and media was added as outlined in Section 2.3.3.2. The cell suspension was transferred to a 20 mL sterile universal and cells were pelleted by centrifuging at 300 x g for 5 min. Supernatant was aspirated off and residual media was removed by washing pellet with 3 mL PBS. PBS was replaced with 700 µL QIAzol Lysis Reagent and the pellet was broken up by pipetting. The lysate was homogenised by passing it through a 20-gauge needle several times using a 5 mL syringe (this was done 4-5 times or until the lysate became less viscous).

***Stabilized tissue***

25 mg portions were excised from RNAlater stabilised tissue samples using a sterile disposable scalpel and forceps and placed in sterile 2 mL Safe-Lock micro centrifuge tubes. 700 µL QIAzol Lysis Reagent and one autoclaved 7 mm stainless steel were added to each sample. To homogenize samples they were shook at a 20 Hz for 2 min in the Qiagen TissueLyser.

#### 2.5.2.2 *Phase separation*

Lysed samples were transferred to 1.8 mL micro-centrifuge tube and left at room temperature for 5 min. 140 µL of chloroform was added, the mixture was shook vigorously to mix well and left at room temperature for 3 min and centrifuged at 12000 x g at 4ºC for 15 min. This separated the sample into 3 phases. The upper aqueous layer was transferred to new micro-centrifuge tubes. The aqueous layer should have a volume of ~ 350 µL, however in this study, in order to avoid disturbing the lower phases only 300 µL was taken and brought forward to RNA extraction.

#### 2.5.2.3 *RNA extraction*

Four hundred and fifty microlitres (1.5 volumes) of 100% EtOH was added to the aqueous layer and mixed thoroughly. 700 µL of the sample was pipetted into an RNeasy Mini spin column and this was centrifuged at 8000 x g for 15 s at, the flow through was discarded and the collection tube was reused. This was repeated until the entire sample had gone through the column. 700 µL Buffer RWT was added to the column and it was centrifuged at 8000 x g for 15 s, the flow through was discarded and the collection tube was reused. 500 µL Buffer RPE was added to the column and it was centrifuged at 8000 x g for 15 s. The flow through was discarded and the collection tube was reused. This step was repeated again but spun for 2 min the second time and the flow through and collection tube was discarded. The column was placed in a new collection tube and spun at 20,000 x g for 1 min, to remove traces of Buffer RPE. The column was transferred to a 1.8 mL micro-centrifuge tube and 50 µL 10 mM Tris-HCl was added to the column. This was incubated at room temperature for 1 min and total RNA was eluted by centrifuging at 8000 x g for 1 min.

### 2.5.3 RNA size fractionation

The pre-stained RNA marker was seperated in a polyacrylamide gel as outlined in Section 2.6.2. This marker contains bands of 20, 30, 40, 50, 75 and 100 bps in length.

Sterile forceps and scalpels used for gel excision were pre-soaked in RNaseZap and rinsed in MBG water before use. Gel sections containing the size fraction of interest were identified and excised using the pre-stained marker in the adjacent lane as a guide.

### 2.5.4 RNA extraction from polyacrylamide gels

The excised gel sections were placed in a 1.8 mL microcentrifuge tube containing 1 mL of 1 M NaCl. The gel was shredded up in solution using a scissors (pre-soaked in RNaseZap and rinsed in MBG water before use). Once the mixture reached a slurry-like consistency it was pipetted into a 15 mL sterile universal containing 9 mL of 1 M NaCl. To elute RNA from the gel fragments, this mixture was incubated on a rocker for 8 hours (or overnight) at 4ºC. The Eluted RNA was separated from the gel debris by centrifuging it at for 5 min at 2,000 x g. The supernatant was transferred to a 50 mL sterile universal. This contained most of the RNA and was placed at 4ºC for later use. In order to elute the remaining traces of RNA from the gel debris, 2 mL 1 M NaCl was added to the remaining gel pellet. This was left rocking at room temperature for 1 hour and centrifuged. The supernatant was collected and added to the supernatant collected earlier. In order to precipitate the RNA 1.5 µL of 5 mg/ml linear acrylamide and 18 mL 100% MBG EtOH (final concentration of 60%) was added with thorough vortexing after each addition. The mixture was put through a MegaClear column and the precipitated RNA was captured on the RNA binding filter. As the MegaClear column had a capacity of ~500 µL and the sample volume was 30 mL it was more practical to draw the mixture through the column using a vacuum rather than a series of centrifugation steps. A vacuum extraction apparatus was constructed by removing a plunger from a 5 mL syringe and replacing it with a MegaClear column which forms a seal. Tubing was fed through a whole created in the bottom of a 15 mL universal; this was attached to the needle holder on the syringe barrel, which was placed into the 15 mL universal allowing it to act as a holder for the apparatus allowing it to be placed on a standard tube rack. The other end of the tubing was attached to a vacuum, allowing flow through to be pumped into a separate waste container, see Figure 2.2. The sample was applied to the column slowly using a 5 mL pipette. After the entire mixture was drawn through, an additional 4 mL of 80% MBG EtOH was also drawn through to wash the column.

The column was placed in 1.5 mL centrifuge tube, 500 µL of 80% MBG EtOH was added and it was centrifuged at 5,000 x g for 1 min. The flow through was discarded. This was centrifuged at 10,000 x g for 1 min to remove traces of EtOH. The column was put in a new collection tube and 50 µL of 10 mM Tris-HCl (heated to 95ºC) was applied to the filter. This was incubated at room temperature for 2 min and centrifuged at 10,000 x g for 1 min. This elution step was repeated again to give a total elution volume of ~100 µL. The large elution volume was necessary to ensure that the entire filter was properly wetted.

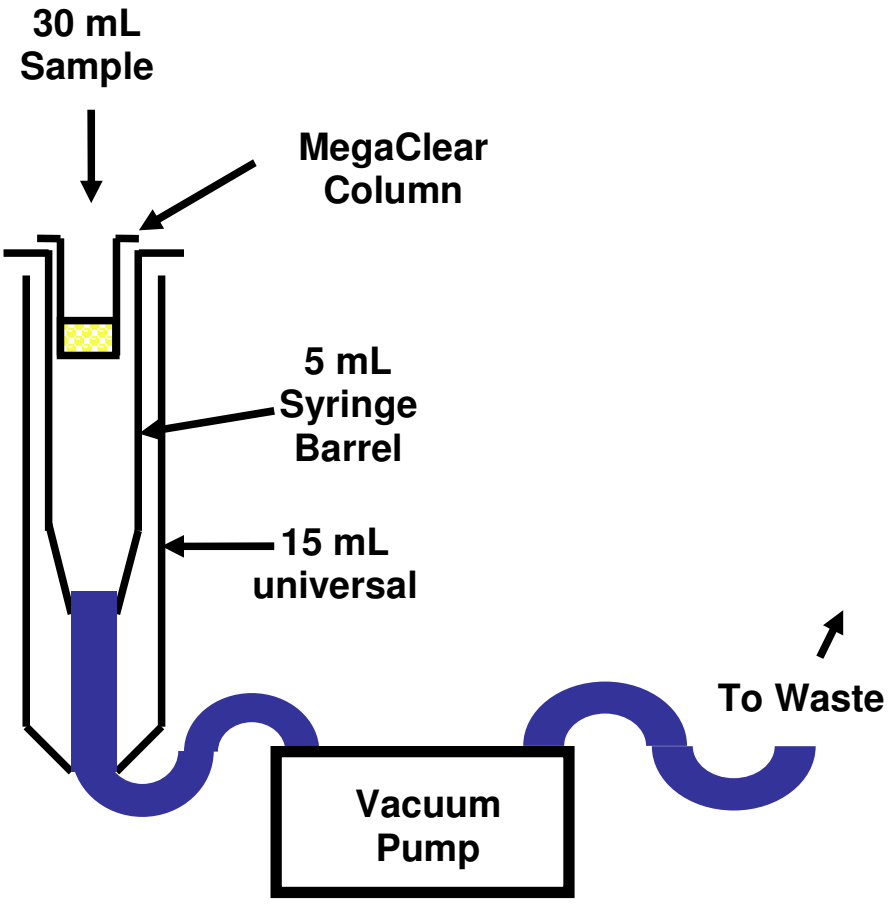



**Figure 2.2 Diagram of vacuum extraction apparatus**

#### 2.5.4.1 *Ethanol Precipitation and Concentration of RNA*

The volume of purified RNA sample were made up to ~100 µL (if not already at this volume) using 10 mM Tris-HCl. The following components were added to the samples in this order with thorough vortexing in after each addition: 11 µLof 3 M NaOAc, 1 µL of 5 mg/ mL Linear Acrylamide and 333 µL 100 % MBG EtOH. Samples were incubated at - 80ºC for 1.5 hours or kept there at this point for long term storage. For subsequent re-suspension, samples were removed from storage at -80ºC and immediately centrifuged at 20,000 x g for 30 min at 4ºC. The supernatant was removed and the pellets were rinsed with 1 mL 80 % MBG EtOH by inverting the tube several times and centrifuged again at 16,000 x g for 10 min at 4ºC. All supernatant was removed and remaining of EtOH were eliminated by leaving the tube open to air dry at room temperature for 5-10 min. RNA pellets were resuspended in the desired volume of 10mM Tris- HCl.

## 2.6 Polyacrylamide Gel Electrophoresis (PAGE)

### 2.6.1 Preparation of mixtures/solutions for PAGE

#### 2.6.1.1 *10% APS*

5 g of APS was dissolved in 50 mL MBG water in a 50 ml universal. This was stored at 4ºC until needed.

#### 2.6.1.2 *1 x TBE Buffer*

100 mL of 10 X TBE was diluted with 900 mL distilled water in a 1 L Duran flask. This was stored at room temperature until needed.

#### 2.6.1.3 *7 M Urea-15% acrylamide mixture*

The following components were added to a 50 mL sterile universal in a fume-hood

**Table 2.2 7 M Urea-15% acrylamide mixture components**

| Component | Amount | Final Conc. |
|---|---|---|
| Urea | 5.88 g* | 6.93 M |
| 10 x TBE | 1.4 mL | 1 x |
| 40% Acrylamide | 5.2 mL | 14.9% |
| MBG Water | 2.7 mL | N / A |

*The volume of the 5.88 g of urea accounted for ~5.7 mL, bringing the total volume to ~14 mL.

The urea was dissolved by placing the mixture in a 45ºC water bath for ~15-20 min. Once the urea was dissolved the mixture was split up into 7 mL aliquots and stored at 4ºC until needed.

## 2.6.2 PAGE protocol

The vertical gel casting apparatus was assembled, ensuring that a proper seal was created at the bottom. In a fume-hood, 14 µL TEMED and 35 µL 10% APS was added to 7mL of the 7 M Urea-15% acrylamide mixture, with thorough vortexing in between each addition. This was immediately added to the gel casting apparatus and a comb was added. This required ~5 mL of urea-acrylamide mixture. The remainder of the gel mixture was used as an indicator for complete polymerisation. Once complete, the gel was placed in the gel tank, the inner buffer was chamber was filled with 1 x TBE and the comb was removed. The wells were flushed out with 1 x TBE using a syringe.1 x TBE buffer was poured into the buffer chamber of the gel tank (typically until just past the bottom of the gel). Gels were pre-run at 200 V for 30 min. Samples were thawed on ice and 1 volume of 2x gel loading buffer was added. The samples were heated to 95ºC for 5 min and put on ice for 1 min. Typically 5 µL of pre-stained miRNA marker was used along side samples (if applicable), this did not require the addition of 2x loading buffer but was also heated and cooled in the same fashion as the samples. Wells were flushed out again with 1 x TBE using a syringe. Up to 20 µL of sample was slowly added to each well. Gels were run at 200 V for 60-90 minutes.

## 2.7 Preparation of Deep-Sequencing Library

The strategy used for library preparation in this study was adapted from the Illumina Small RNA v1.5 protocol. The modified in-house strategy took longer to execute but it was significantly lower in cost as it circumvented the need to purchase an expensive library preparation kit from Illumina. The main difference was the use of a different 3' adapter. For this, a relatively in-expensive pre-adenylated 3' adaptor (IDT small RNA cloning linker 1) was used. As this adaptor did not have the same sequence as the Illumina 3' adaptor, the appropriate 3' adaptor sequences (compatible with Illumina flow cell primers during Illumina sequencing) were added in after the IDT pre-adenylated adapter by incorporating them into the PCR primers, see Figure 2.3 and Figure 2.4. In addition, PAGE purification between ligation steps and RT combined with using lower amounts of adaptor resulted in cleaner sequencing libraries that did not require agarose gel purification prior to sequencing. This resulted in slightly longer products than those of conventional Illumina small RNA sequencing libraries although this posed no compatibility issues with the downstream sequencing process. Illumina sequencing was carried out by TrinSeq, the TCD Genome Sequencing Laboratory based in Institute of Molecular Medicine, St James' Hospital Dublin.

## 2.7.1 Adaptors and primers

Adaptors and primers were synthesized by IDT (Table 2.3)

**Table 2.3 Details of adaptors and primers used in deep-sequencing library preparation**

| Name | Type | Sequence |
|---|---|---|
| 3' Adaptor | DNA | 5' /5rApp /CTGTAGGCACCATCAAT/3ddC/ 3' * |
| 5' Adaptor | RNA | 5' GUUCAGAGUUCUACAGUCCGACGAUC 3' |
| Reverse Transcription Primer | DNA | 5' ATTGATGGTGCCTACAG 3' |
| Forward Primer | DNA | 5' AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA 3' |
| Reverse Primer | DNA | 5'CAAGCAGAAGACGGCATACGAATTGATGGTGCCTACAG 3' |
| 5'pTag Forward primer | DNA | 5' GCTATGACCATGATTACGCCAA 3' |
| 3'pTag Reverse primer | DNA | 5' TGTAAAACGACGGCCAGTGAA 3' |

*3' Adaptor terminal modifications:/5rApp/ = 5' Adenylation containing a pyrophosphate linkage. This acts as substrate for T4 RNA Ligase II in the absence of ATP. /3ddC/ = 3' dideoxy-Cytosine (ddC) base. This blocks ligation on this end.

## 2.7.2 Preparation of mixtures/solutions

### 2.7.2.1 *Adaptors*

Lyophilized adaptors were first centrifuged at 5000 x g for 1 min to ensure all adaptors were removed from the sides of the tube. A master stock was prepared by re-suspending adaptors to a concentration of 0.1mM using 10mM Tris-HCl. Working stocks of 6.6 pmol / µL were prepared from these by diluting them 1/15 in 10mM Tris-HCl. Adaptor stocks were stored at -80 ºC.

### 2.7.2.2 *Primers*

Lyophilized adaptors were first centrifuged at 5000 x g for 1 min to ensure all adaptors were removed from the sides of the tube. A master stock was prepared by re-suspending these using 10mM Tris-HCl to a concentration of 1 mM. Working

stocks of 10 pmol / µL were prepared from these by diluting them 1/100 in 10mM Tris-HCl. These were stored at -20ºC.

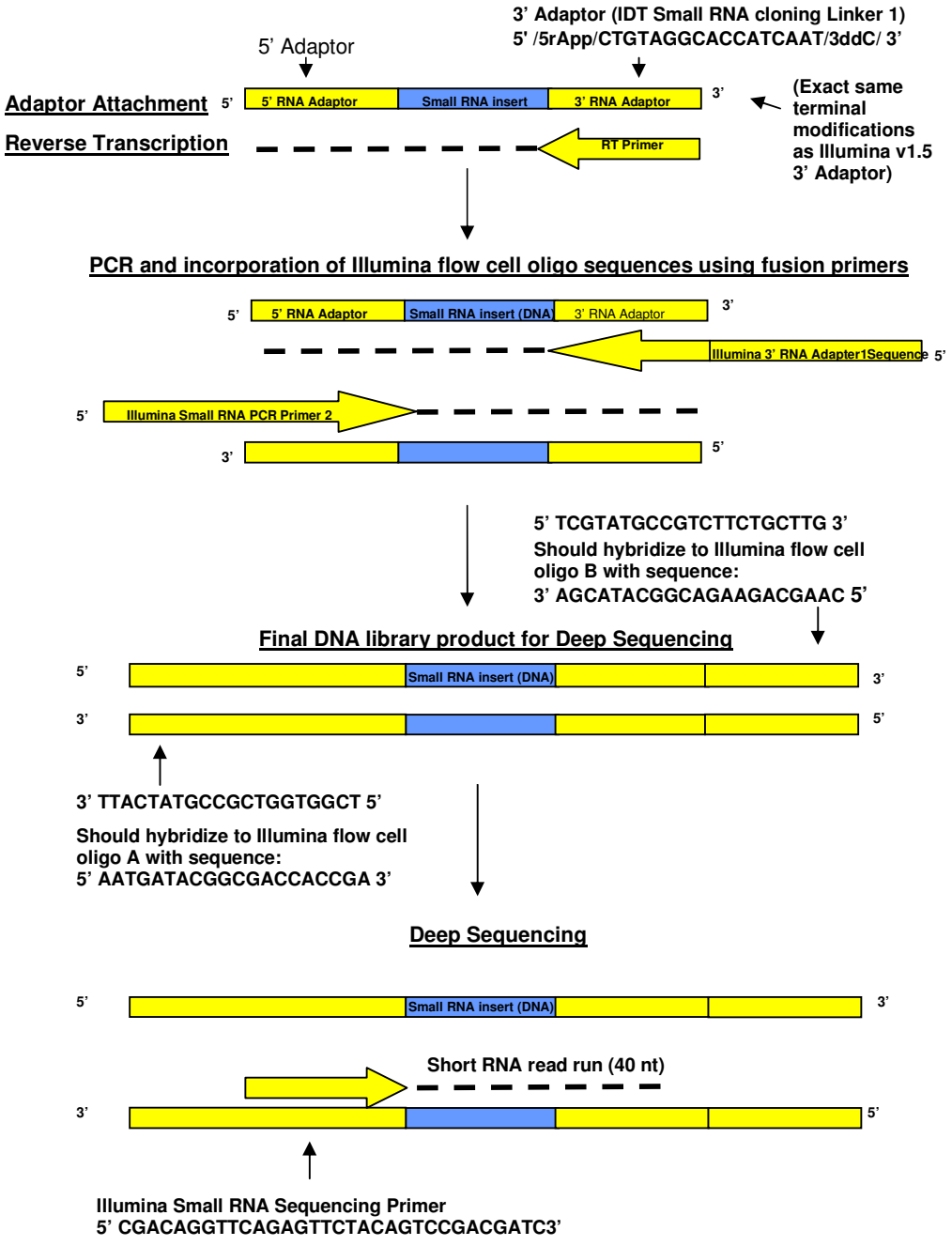



**Figure 2.3 Summary of in-house small RNA deep-sequencing library preparation**

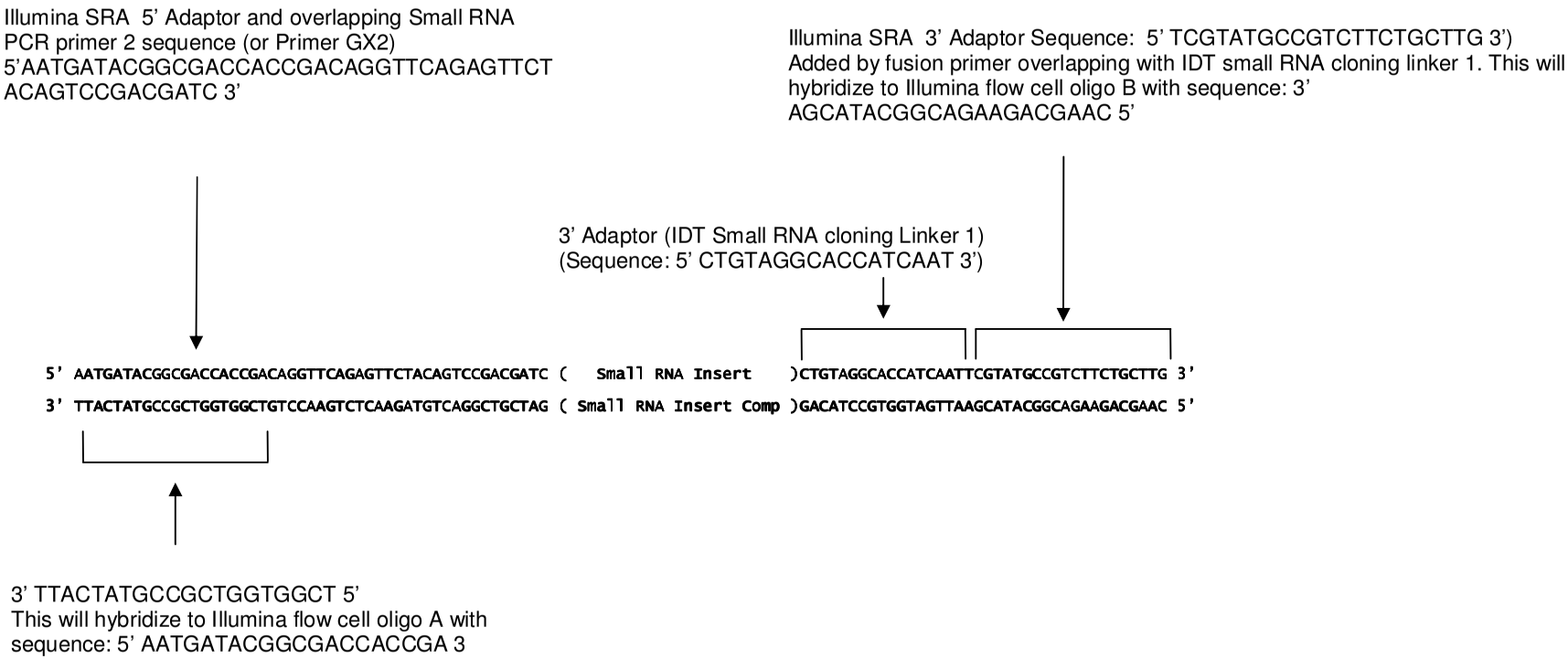



**Figure 2.4 Complete sequence of dsDNA library after in-house small RNA sequencing library preparation**

### 2.7.2.3 *3' Ligation Reaction mix*

This was made up in a 0.5 mL thin walled PCR tube immediately before use.

**Table 2.4 Details of 3' ligation reaction mix**

| Component | Stock Conc. | Volume | Final Conc. (in 20 µL) |
|---|---|---|---|
| T4 Ligase 2 | 200 U / µL | 1 µL | 10 U / µL |
| T4 Ligase II Buffer | 10x | 2 µL | 1x |
| MgCl2 | 50mM | 1.6 µL | 4mM |
| DMSO | 100% v/v | 2 µL | 10% v/v |
| PEG 8000 | 50% | 2.9 µL | 7.25% v/v |
| RNasin | 40 U/ µL | 1 µL | 1 U / µL |
| **Total** | | 10.5 µL | |

### 2.7.2.4 *5' Ligation Reaction mix*

This was made up in a 0.5 mL thin walled PCR tube immediately before use

**Table 2.5 Details of 5' ligation reaction mix**

| Component | Stock Conc. | Volume | Final Conc. (in 20 µL) |
|---|---|---|---|
| T4 Ligase 1 | 20 U/ µL | 1 µL | 1 U / µL |
| T4 Ligase I Buffer | 10x | 2 µL | 1x |
| DMSO | 100% v/v | 2 µL | 10% v/v |
| PEG 8000 | 50% | 4.5 µL | 11.25% v/v |
| RNasin | 40 U / µL | 1 µL | 2 U / µL |
| **Total** | | 10.5 µL | |

### 2.7.2.5 *Library Reverse transcription (RT) reaction mix*

This was made up in a 0.5 mL thin walled PCR tube and stored at -20ºC until needed

**Table 2.6 Details of library preparation RT reaction mix**

| Component | Stock Conc. | Volume | Final Conc. (in 20 µL) |
|---|---|---|---|
| dNTPs | 10 mM | 1 µL | 1 mM |
| SuperScript III RT | 200 U / µL | 1 µL | 20 U / µL |
| SS III RT Buffer | 5x | 4 µL | 1x |
| DTT | 0.1 M | 1 µL | 0.05 M |
| RNasin | 40 U/ µL | 1 µL | 2 U / µL |
| **Total** | | 8 µL | |

### 2.7.2.6 *Library amplification PCR master mix*

This was prepared in a 0.5 mL thin walled PCR tube and stored at -20ºC until needed.

**Table 2.7 Details of library amplification PCR master mix**

| Component | Stock Conc. | Volume | Final Conc. (in 50 µL) |
|---|---|---|---|
| 5X Phusion HF Buffer | 5x | 10 µL | 1x |
| Forward Primer | 10uM | 1 µL | 0.2uM |
| Reverse Primer | 10uM | 1 µL | 0.2uM |
| DNTPs | 25mM | 0.5 µL | 250uM |
| Phusion DNA Polymerase | 2 U / µL | 0.5 µL | 0.02 U / µL |
| **Total** | | 13 µL | |

#### 2.7.2.7 *Library validation PCR master mix*

This was prepared in a 0.5 mL thin walled PCR tube and stored at -20ºC until needed.

**Table 2.8 Details of library amplification PCR master mix**

| Component | Stock Conc. | Volume | Final Conc. (in 50 µL) |
|---|---|---|---|
| 5X Green GoTaq® Flexi Buffer | 5x | 10 µL | 1x |
| $MgCl_2$ | 25 mM | 5 µL | 2.5 |
| dNTPs | 25 mM | 0.5 µL | 1.25 |
| Forward Primer | 10 µM | 1 µL | 0.2 pMol / µL |
| Reverse Primer | 10 µM | 1 µL | 0.2 pMol / µL |
| GoTaq Flexi DNA Polymerase | 5 U / µL | 0.25 µL | 0.025 U / µL |
| MBG H2O | N/A | 29.75 µL | N/A |
| **Total** | | 47.5 µL | |

#### 2.7.2.8 *1 X TAE Buffer*

One Litre of 10x TAE Buffer was added to 9 L of distilled water in a dispenser and kept at room temperature.

#### 2.7.2.9 *Agarose Gels*

Immediately before use, either 1.2 g (2%) or 2.4 g (4%) of agarose was added to 60ml 1x TAE buffer and heated in a microwave until the agarose was fully dissolved. In a fume hood, 3 µL of 10 mg / mL ethidium bromide was added to the gel and mixed well by swirling.

#### 2.7.2.10 *DNA Ladders*

There were 2 sizes of DNA ladder used in this study, 10bp and 100bp. These were all supplied by Invitrogen and come in a concentration of 1 µg/µl and each tube contained 50 µL. Ready to use tock solutions of all DNA ladders were prepared by adding 160 µL 6x gel loading dye and 790 µL 1x TAE Buffer. Ladders were stored at at -20ºC.

#### 2.7.2.11 *50 mg/mL X-Gal*

To prepare this, 500 mg of X-Gal was dissolved in 10 ml distilled water in a Duran flask. This was wrapped in thin foil and stored at 20°C.

#### 2.7.2.12 *100 mM IPTG*

To prepare this, 238 mg of IPTG was dissolved in 10 ml distilled water and stored at -20°C.

#### 2.7.2.13 *Ligation reaction mix*

This was prepared immediately before use.

**Table 2.9 Ligation reaction mix**

| Component | Stock Conc. | Volume | Final Conc. (in 20 µL) |
|---|---|---|---|
| 2X Rapid Ligation Buffer | 2x | 5 µL | 1x |
| pGEM®-T Easy Vector | 50 ng / µL | 1 µL | 5 ng / µL |
| Library validation PCR product | N/A | 3 µL | N/A |
| T4 DNA Ligase | 3 Weiss units/ µl | 1 µL | 0.3 Weiss units/ µl |
| **Total** | | 10 µL | |

#### 2.7.2.14 *Colony PCR mix*

This was prepared in a 0.5 mL thin walled PCR tube and stored at -20ºC until needed.

| Component | Stock Conc. | Volume | Final Conc. (in 50 µL) |
|---|---|---|---|
| 5X Green GoTaq® Flexi Buffer | 5x | 10 µL | 1x |
| $MgCl_2$ | 25 mM | 5 µL | 2.5 |
| dNTPs | 25 mM | 0.5 µL | 1.25 |
| 5'pTag Forward primer | 10 µM | 1 µL | 0.2uM |
| 3'pTag Reverse primer | 10 µM | 1 µL | 0.2uM |
| GoTaq Flexi DNA Polymerase | 5 U / µL | 0.25 µL | 0.025 U / µL |
| MBG H2O | N/A | 32.25. µL | N/A |
| **Total** | | 50 µL | |

#### 2.7.2.15 *Dye terminator PCR mix*

The dye terminator PCR mix was prepared immediately before use as per Table 2.10.

**Table 2.10 Dye terminator PCR mix**

| Component | Stock Conc. | Volume | Final Conc. (in 50 µL) |
|---|---|---|---|
| BigDye Terminator v3. Reaction mix | 5x | 4 µL | 1x |
| BigDye Terminator v3.1 Buffer (5X) | 5x | 4 µL | 1x |
| 5'pTAG Forward primer | 10 µM | 4µL | 0.2uM |
| MBG H2O | N/A | 4. µL | N/A |
| **Total** | | 16 µL | |

### 2.7.3 RNA library preparation protocol

#### 2.7.3.1 *17-25 nt RNA size fractionation*

Size fractionation and subsequent gel extraction of 17-25 nt RNA was carried out as per Section 2.5.3 and 2.5.4. For this particular application the gel was run at 200 V for 60 min. A section of gel was excised from the sample lane that corresponded to half way between the 20 and 30 nt bands and the same distance again below the 20 nt marker on the pre-stained RNA marker in the adjacent lane. Extracted RNA was concentrated by EtOH precipitation and resuspended (as per section 2.5.4.1) in a volume of 9.5 µL 10mM Tris HCl. In order to see if RNA size fractionation was successful before proceeding with adaptor ligation, 1.5 µL of the RNA sample was used to measure the RNA quantity on a Nanodrop spectrometer. Libraries were prepared from 2 different infections i) 400 ng 17-25 nt RNA from cells infected with a 1 mL of 2.7x $10^4$ $TCID_{50}$ / mL H361 isolate (18 d.p.i.) at and ii) 72 ng from a 17-25 nt RNA from cells infected with 1 mL of 5x $10^5$ $TCID_{50}$ / mL N076 isolate (14 d.p.i.) (less RNA in N076 sample as it was extracted from a 25$cm^2$ flask compared to 75 $cm^2$ flask for the H361 isolate).

#### 2.7.3.2 *3' Adaptor Ligation*

The remaining 8 µL of concentrated RNA was mixed with 1.5 µL of 3' adaptor working stock. The RNA was denatured by heating to 70ºC for 2 min, put on ice for 1 min and added to a thawed 3' ligation reaction. The reaction was incubated at 22ºC for 2 hours. The ligation product was separated from unused adaptors and RNA by size fractionation. The gel was ran at 200 V for 90 min. The ligation products were expected to be ~34-42 nt in length, therefore section of gel was excised from the sample lane that corresponded to little under half way between the 30 and 40 nt bands and little over the 40 nt band on the pre-stained RNA marker in the adjacent lane. Ligation products were extracted from the gel slice and concentrated, resuspended in volume of 9.5 µL 10mM Tris HCl. In order to check if 3' ligation, and gel extraction was successful before proceeding with 5' adaptor ligation, 1.5 µL of the RNA sample was used to measure the RNA quantity on a Nanodrop spectrometer.

#### 2.7.3.3 *5' Adaptor Ligation*

The remaining 8 µL of concentrated 3' ligation product ligated to the 5' adaptor in the same manner as above, using the 5' adaptor ligation mix, which was incubated at 37ºC for 1 hour. The 5' ligation products were expected to be ~60-68 nt in length, therefore a section of gel was excised from the sample lane that corresponded to between the 60 and 75 nt bands on the pre-stained RNA marker in the adjacent lane. The product was extracted and checked as per Section 2.7.3.2.

#### 2.7.3.4 *Reverse Transcription*

The remaining 8 µL of concentrated 5' ligation product was mixed with 1 µL of RT primer working stock and 3 µL MBG water. This was denatured by heating to 70ºC for 2 minutes, put on ice for 1 min and added to an RT mix. This was heated on a thermocycler at 55ºC for 90 min followed by 70ºC for 15 min and held at 4ºC.

#### 2.7.3.5 *Amplification of sequencing library by PCR*

All 20 µL of the RT reaction was added to a library amplification PCR master mix, and 17 µL of MBG water. This was placed in a thermocycler and subjected to the following cycling conditions: 1 cycle of 98ºC for 30 sec, 12 cycles of 98ºC for 10 sec, 60ºC for 30 sec, and 72ºC for 15 sec followed by a final extensions step of 72ºC for 10 min (the amount of cycles was kept to minimum in order to keep amplification bias to minimum).

### 2.7.4 Validation of amplified sequencing library

It was important that no non-specific products were generated in sequencing library preparation. The final dsDNA product should be 104-113 bp in length. This was verified by agarose gel electrophoresis. The library itself was not run on gels, in-order to conserve it. Instead, the quality of the library was assessed indirectly by using 2.5 µL of it as template for another PCR reaction. This involved using the same primers with a non-high fidelity DNA polymerase (Taq) that also introduced TA overhangs. The introduction A-overhangs also facilitated downstream TA cloning of the PCR products for more in dept validation of the library.

#### 2.7.4.1 *Gel electrophoresis*

For this PCR 2.5 µL of dsDNA library was added to the library-validation-PCR-master mix. This was placed in a thermocycler and subjected to the following cycling conditions. 1 cycle of 94ºC for 5 min, 15 cycles of 95ºC for 1min, 55ºC for 1 min and 72ºC for 1 min followed by a final extension step of 72ºC for 10 min (In this PCR the amount of cycles was kept to minimum primarily in order to avoid the generation of non-specific products due to over-cycling). For gel electrophoresis, the ends of a gel cast were sealed with masking tape, 4% agarose gel was poured into it and a comb of appropriates size was put in place. Once set, the masking tape was removed and the gel it was placed in a gel tank which was filled with 1x TAE buffer until the gel was fully submerged. The comb was removed and 10 µL of PCR product (mixed with 2 µL of 6x gel loading buffer). One lane was loaded with 10 µL DNA ladder. The gels were run at 110 V for 60 min and visualized using a UV transilluminator and a CCD camera.

#### 2.7.4.2 *Cloning*

A ligation reaction as set up as per Table 2.9, components were mixed by pipetting. This was incubated for 1 hour 22ºC kept at 4ºC overnight for maximum efficiency. Ligation reaction was centrifuged and 2 µl was added to a new microcentrifuge tube on ice. JM109 High Efficiency Competent Cells (a component of the pGEM®-T Easy Vector System II kit) were removed from storage at -80ºC and thawed in ice for 5 min. These were gently mixed (by flicking the tube) and 50 µl of cells were added to the 2 µl ligation product. This was gently mixed (by flicking the tube) and left on ice for 20 min. Cells were heat shocked by placing them in a water bath at 42°C for 45 – 50 sec and returned to ice for 2 min. Cells were removed from ice and 950 µl room temperature LB medium was added. Cells were incubated for 1.5 hours at 37°C with shaking (~150rpm). LB-ampicillin plates were dried 37°C for 10 minutes after which 100 µl of 100 mM IPTG and 20 µl of 50 mg/ml X-Gal were spread onto them. Plates were warmed at 37°C for 30 min to allow absorption of IPTG. Cells were removed from incubation and 100 µl of culture was spread onto /ampicillin/IPTG/X-Gal plate. These were incubated overnight (16 – 24 hours) at 37°C. DNA from white colonies (cells carrying plasmids with inserts from the sequencing library) was transferred to PCR reactions by using a sterile loop to gently touch the edges of the colonies followed by vigorously mixing the loop in a colony-PCR-mix. This was subjected to the following cycling conditions: 1 cycle of 96ºC for 10 min, 30 cycles of 94ºC for 1min, 37ºC for 1 min and 72ºC for 1 min followed by a final extension step of 72ºC for 10 min. These PCR products were ran on 2% agarose gels as per section 2.7.4.1. Colony-PCRs showing PCR products of 312-320 bp in size (104-112 insert plus a total of 208 nt of flanking pGEM®-T Easy Vector sequence) were selected for dye-terminator sequencing.

#### 2.7.4.3 *Dye terminator sequencing*

In order to clean up colony PCR reactions before dye terminator sequencing, 4 µl of ExoSAP-IT was added to 10 µl of positive colony PCR product. This was incubated at 37°C for 15 min (degrading remaining primers and nucleotides) followed by incubation at 80°C for 15 min to inactivate ExoSAP-IT. Four micro-liters of cleaned up colony PCR product was added to dye-terminator PCR mix. This was subjected to the following reaction conditions: 25 cycles of 96ºC for 30 sec, 45ºC for 15 sec and

60ºC for 1 min followed by a final extension step of 72ºC for 10 min. In order to remove un-incorporated dye terminators from dye terminator PCRs before sequencing they were cleaned up using the DyeEx 2.0 Spin Kit. Columns were prepared by gently vortexing them (to re-suspend the resin), opening the cap a quarter turn and snapping the bottom closure. Empty columns were centrifuged for 3min at 750 x g. they were placed in collection tubes and 20 µl of dye terminator PCR product was slowly added to the centre of the gel bed surface. These were centrifuged for 3min at 750 x g. The collection tubes containing purified samples were incubated at 80ºC for 30 min in order to dry the sample. Samples were resuspended in 12.5 µl Hi-Di Formamide, heated to 90ºC for 2 mins and vortexed well before being applied to the ABI PRISM 310 Genetic Analyzer sample tray.

## 2.8 Array Hybridization

### 2.8.1 Probe Design

Probes were designed using the probe design tool in VMir. These were designed to detect proposed and theoretical mature miRNAs derived from the 5' and 3' arms of 2914 predicted pre-miRNAs These included both high and low probability pre-miRNA candidates from manual analysis of the deep-sequencing results and predicted pre-miRNAs that did not feature in the results at all.

### 2.8.2 Controls

#### 2.8.2.1 *Control Probes*

Each miRNA-probe was included in duplicate and control probes were also designed

***Mismatch probes:***

Each miRNA-probe had an associated mismatch control containing one mismatch with its target every 5 nt. This allowed estimation of non-specific signal occurring as a result of cross hybridization due to sequence similarity (Figure 2.5).

***Scrambled controls:***

Each miRNA-probe had an associated scrambled control generated by random shuffling of probe sequence. (Figure 2.5).

***Flanking controls***

Probes targeting sequences (30 nt in length) flanking the target regions (25 nt upstream of the 5' arm target and 25 nt downstream of the 3' arm target) were also included on the array. These controls were not part of the predicted miRNA target sequence and were intended to account for non-specific signals caused by hybridization of small degradation products of larger transcripts which overlap predicted pre-miRNA loci. If such overlapping degradation products hybridized to probes targeting putative miRNAs they would also hybridize to the associated flanking controls to the same extent. (Figure 2.5).

#### 2.8.2.2 *Positive Control targets*

Probes were also designed to detect positive control targets. Positive control probes also had associated mismatch and scrambled controls. These are described below.

- **Endogenous controls:**

  Probes targeting highly expressed *Cyprinus carpio* (host) miRNAs from deep sequencing were also included as endogenous controls. These acted as a control for labelling and acted as an indicator of the quality of the miRNA enriched RNA samples used.

- **Exogenous spike-in control:**

  All labelling reactions were spiked with a defined amount of synthetic miRNA. This consisted of an RNA oligo with sequence GUGCUCACUCUCUUCUGUCG, based on the Arabidopsis thaliana miRNA ath-miR156g (MirBase Ac. MIMAT0001012). This acted as a positive control for the labelling reaction and hybridization and for all samples regardless of whether they contained miRNAs or not.

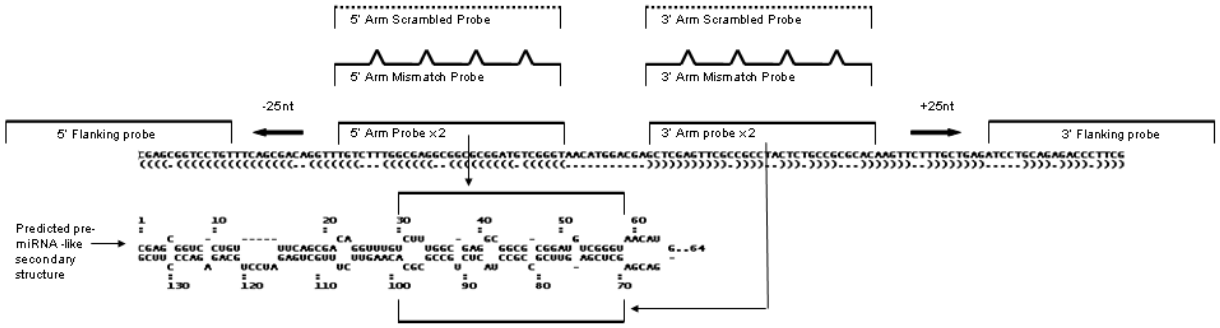



**Figure 2.5 Illustration of probe design showing control-probe targets**
On top: linear sequence of predicted pre-miRNA. On bottom: Secondary structure of predicted pre-miRNA

### 2.8.2.3 *Types of samples used*

- **S1: CyHV-3 positive miRNA enriched**

  This sample consisted of 17-25 nt RNA from CyHV-3 infected CCB cells. Probes targeting CyHV-3 miRNAs should show positive signals in this sample.

- **S2: CyHV-3 positive non-miRNA**

  This sample consisted of 26-35 nt RNA from the same CyHV-3 infected CCB cells as sample 1. This RNA was derived from the sample used in deep sequencing except it consists of small RNA just outside the RNA size range. Probes targeting genuine miRNAs should not also show signals in this sample.

- **S3: CyHV-3 Negative miRNA enriched sample**

  This sample consisted of 17-25 nt RNA from non-infected CCB cells. Probes targeting genuine CyHV-3 miRNAs should not also show signals in this sample, in this way it acted as an extra negative control allowing us to identify NSH to host RNA.

### 2.8.3 Preparation of array hybridization mixtures/solutions

#### 2.8.3.1 *Exogenous spike-in control*

The exogenous spike-in control, consisting of an RNA oligo with sequence GUGCUCACUCUCUUCUGUCG, based on the *Arabidopsis thaliana* miRNA ath-miR156g (MirBase Ac. MIMAT0001012), was synthesized by IDT. A master stock was prepared by re-suspending lyophilised the pellet to a concentration of 1 nmol/ µL (i.e. 1 µL per nmol present) using 10mM Tris-HCl. This was diluted to $10^{-7}$ to make a working stock of 100 amol/ µL (or ~7 pg / µL). This was split up into 5 µL aliquots and stored at -80°C.

#### 2.8.3.2 *Label IT Reagent*

Label IT Reagent arrived as dried pellet and was resuspended before use by adding 22 µL Reconstitution Solution (supplied Label IT® miRNA Labeling Kit, Version 2). This was mixed by vortexing and centrifuged to ensure all solution was collected at the bottom of the tube.

#### 2.8.3.3 *Labelling reaction mixture*

This was prepared immediately before use. Components were added in order as they appear in Table 2.11 in accordance with the Label IT® miRNA Labeling Kit Version 2 protocol.

**Table 2.11 Labelling reaction mix**

| Component | Stock Conc. | Volume | Final Conc. |
|---|---|---|---|
| MBG H2O | N/A | 33 µL | N/A |
| Labelling Buffer M | 10x | 10 µL | 1x |
| RNA sample (~400ng brought to 35 µL in 10mM Tris HCl) | ~11 ng/ µL | 35 µL | ~4 ng / µL |
| Exogenous-Spike-in control | 100 amol/ µL | 2 µL | 2 amol/ µL |
| Label IT® Reagent | 25x | 20 µL | 5x* |
| **Total** | | 100 µL | |

*The final concentration of Label IT® Reagent was increased to 5x (which was optional according to the Label IT® miRNA Labeling Kit, Version 2 protocol). This results in maximum labelling density and was done to reduce the possibility of false negatives that may occur if some target transcripts were present in very low levels.

### 2.8.3.4 *50% w/v dextran sulphate*

50% w/v dextran sulphate working stock was prepared by adding 0.5g dextran sulphate to 1000 µL MBG H2O. Due to the highly viscous consistency this could not be adequately mixed by vortexing. Instead the mixture was adequately mixed by placing the micro-centrifuge tube the TissueLyser and shaking it at a 20 Hz for 2 min. Bubbles were removed by centrifugation. This was stored at room temperature until needed.

### 2.8.3.5 *Hybridization buffer mix*

This was prepared immediately before use and mixed by vortexing (Table 2.12).

**Table 2.12 Hybridization buffer mix**

| Component | Stock Conc. | Volume | Final Conc. |
|---|---|---|---|
| SSPE | N/A | 17.5 µL | N/A |
| Formamide | 100% | 28 µL | 20x |
| Tween-20 | 1% | 1.4 µL | 0.01% |
| Acetylated BSA | 1 mg / ml | 1.4 µL | 0.01 mg / mL |
| Dextran Sulphate | 50% w/v | 18.5 µL | ~6.5% w/v |
| SDS | 10% w/v | 28 µL | 2% |
| Labelled RNA (~400ng eluted in 45 µL Elution buffer) | ~8.9 ng / µL | 44.2 µL | ~2.69 ng / µL |
| **Total** | | 140 µL | |

#### 2.8.3.6 *Wash buffers*

Wash buffers were made up in sufficient enough volumes to allow for a flow rate of 11 mL / min during washing steps when using the Tecan HS 400™ Pro Hybridization Station. Once prepared, they were stored at room temperature. Wash buffers were prepared as per Table 2.13

**Table 2.13 Array wash buffer details**

| **Wash buffer 1** | |
|---|---|
| **2x SSPE 0.1% SDS (Low Stringency)** | |
| **Component** | **Volume** |
| 20x SSPE | 100 mL |
| 10% SDS | 10 mL |
| H2O | 890 mL |
| **Wash buffer 2** | |
| **1x SSPE (Intermediate Stringency)** | |
| **Component** | **Volume** |
| 20x SSPE | 50 mL |
| H2O | 950 mL |
| **Wash buffer 3** | |
| **0.1x SSPE (High Stringency)** | |
| **Component** | **Volume** |
| 20x SSPE | 5 mL |
| H2O | 995 mL |

### 2.8.4 RNA Labelling and purification

RNA was size fractionated as per Section 2.5.3. Approximately 400ng RNA was selected from each sample and made up to a volume of 35 µL in 10mM Tris HCl. This was stored at - 20°C until labelling. The labelling reaction mixture (including RNA sample) was prepared as per Section 2.8.3.3. This was incubated for 1.5 hours at 37°C. The labelling reaction was stopped by adding 10 µL 10x Stop Reagent and briefly stored on ice until purification. Samples were prepared for purification by adding 10 µL Column Binding Buffer and 250 µL 100 % MBG EtOH and mixed by vortexing. Samples were added to purification columns and centrifuged at 11, 000 x g for 30 sec. The flow through was discarded and 600 µL Wash buffer was added to the column and it was centrifuged at 11, 000 x g for 30 sec. The flow through was discarded and 200 µL Wash buffer was added to the column and it was centrifuged at 11, 000 x g for 2 min. The column was transferred to a 1.8 mL eppendorf tube and 45

μL Elution buffer was applied to the membrane and it was centrifuged at 11, 000 x g for 1 min. The purified labelled RNA was stored at -20°C until used for array hybridization.

### 2.8.5 Array Hybridization, washing and drying

The Tecan HS 400™ Pro Hybridization Station was prepared by running the wash program which involved pumping MBG H2O through all channels and chambers in the system. Each wash buffer flask was assigned and connected to a different flow channel, this information and the experiment conditions were programmed into the hybridization station. The experimental conditions are given in Table 2.14

**Table 2.14 Array hybridization and washing conditions**

| **Hybridization program** | | | |
|---|---|---|---|
| **Duration** | **Temperature** | **Agitation** | **Extra settings** |
| 14 hours | 42℃ | Medium | High viscosity mode "on" |
| **Slide washing and Drying program** | | | |
| **Wash buffer** | | **Temperature** | **Duration** |
| Wash Buffer 1 (Channel 1) | | 30℃ | 5 min |
| Wash Buffer 2 (Channel 2) | | 22℃ | 5 min |
| Wash Buffer 3 (Channel 3) | | 22℃ | 2 min |
| $N_2$ Drying | | RT | 5 min |

The hybridization buffer (including labelled RNA) was prepared as per Section 2.8.3.5. This was heated to 65°C for 5 mins and kept on ice for 5 mins. The arrays were placed in the Tecan 400 hybridization chamber with the array side (same side as barcode) facing up. The chamber was closed and sealed. The hybridization, washing and drying programs were executed. After system initialization, the samples were injected into the hybridization chambers (when prompted by the display) and the hybridization program commenced automatically followed by washing and drying programs.

### 2.8.6 Array Scanning and acquisition of fluorescence data

Arrays were scanned using NovaRay Scanner (Alpha Innotech). Arrays were placed on the slide tray with the barcode facing up and at the flat end of the slide holder (this resulted in the arrays being scanned in the correct orientation). Using the visual

preview mode, the scanning area, focus settings were manually defined for each scan. All arrays were scanned using an 8 sec exposure and resolution of 5 μm. The resulting images were analyzed using the ArrayVision software package v8.0. To facilitate this, the GAL file (GenePix Array List file) sent with the arrays was imported into ArrayVision. This contained information on the layout and identity of each feature on the array allowing each feature to be correctly identified and analyzed. The results were exported as a tab delimited text files consisting of probe names in one column with florescence intensity data in the other.

### 2.8.7 Analysis of fluorescence data

#### 2.8.7.1 *Identification of miRNA-like signals*

The results files were imported into VMir for automated analysis of the data for identification of miRNA-like signals. Each probe targeting a miRNA on the 5'or 3' arm of a predicted pre-miRNA had its own associated control probes and in turn each predicted pre-miRNA loci had its own flanking controls. Separate probes targeting miRNAs from the same pre-miRNA and their associated control probes were automatically recognized based on the nomenclature system used allowing them to be analysed together to give an overall score for each predicted pre-miRNA.

VMir first calculated the median intensity of all of these control probes in order to estimate the background. Only probes with intensities that rose above this minimum background level were considered for further analysis by VMir. Each miRNA-probe was present in duplicate; if the signals from these probes were above the background level their mean was calculated. If any of the associated controls for these individual probes were also above background-levels, this indicated that at least some of the signal for the miRNA-probes may be attributed to non-specific hybridization (NSH). The extent to which the control probes were above background was used for the estimation of the degree of NSH taking place. As a result the amount by which the control was above background (i.e. control intensity minus background) was subtracted from the mean of the miRNA-probe replicates. The resulting value was referred to as the "corrected-value" for the miRNA-probe. Flanking controls that

were above background were treated in the same way, but unlike scrambled and mismatch controls, they were not specific to individual 5' or 3' arm miRNA-probes but instead act as controls for the pre-miRNA loci as a whole and as such, even though they were not replicates of each other, their values (i.e. intensity – minus minimum background) were always averaged before subtraction from mean miRNA-probe values. All fluorescence data was analysed in this fashion except in the case of samples consisting 26-35 nt CyHV-3 RNA and 17-25 nt CCB, where the flanking probes were ignored in order to avoid possible under-estimation of miRNA-probe signals from such samples. For each probe it was *only* the control type with the highest value that was used for correction. The corrected-value of the best performing miRNA-probe for each pre-miRNA hairpin (either 5' or 3' arm probe) was used to represent the value for that predicted pre-miRNA. Only corrected-values that were still above background were considered positive by VMir.

In this study, all corrected-values from VMir were expressed as a multiple of the background. This was referred to as the "relative-value" (RV). RVs from VMir that were higher than background were not immediately taken as positive. Separate cut-off values were established for each hybridization. To establish these cut-offs, the level of variability in the intensities of the control probes was established by calculating the average absolute-deviation from the background (i.e. background previously estimated by VMir using the median intensity of the same control probes). Cut-off values were defined as 1 absolute deviation above the background-level. In this study *only* pre-miRNAs with RVs that were above the established cut-off values for their respective hybridizations were considered positive. Positive signals from the Sample 1 (17-25 nt CyHV-3 RNA) were deemed to be potentially attributed to CyHV-3 miRNAs if they showed no corresponding positive signals from Samples 2 (26-35 nt CyHV-3 RNA) and 3 (17-25 nt CCB RNA). The processes involved in analysis of fluorescence data are outlined in Figure 2.6

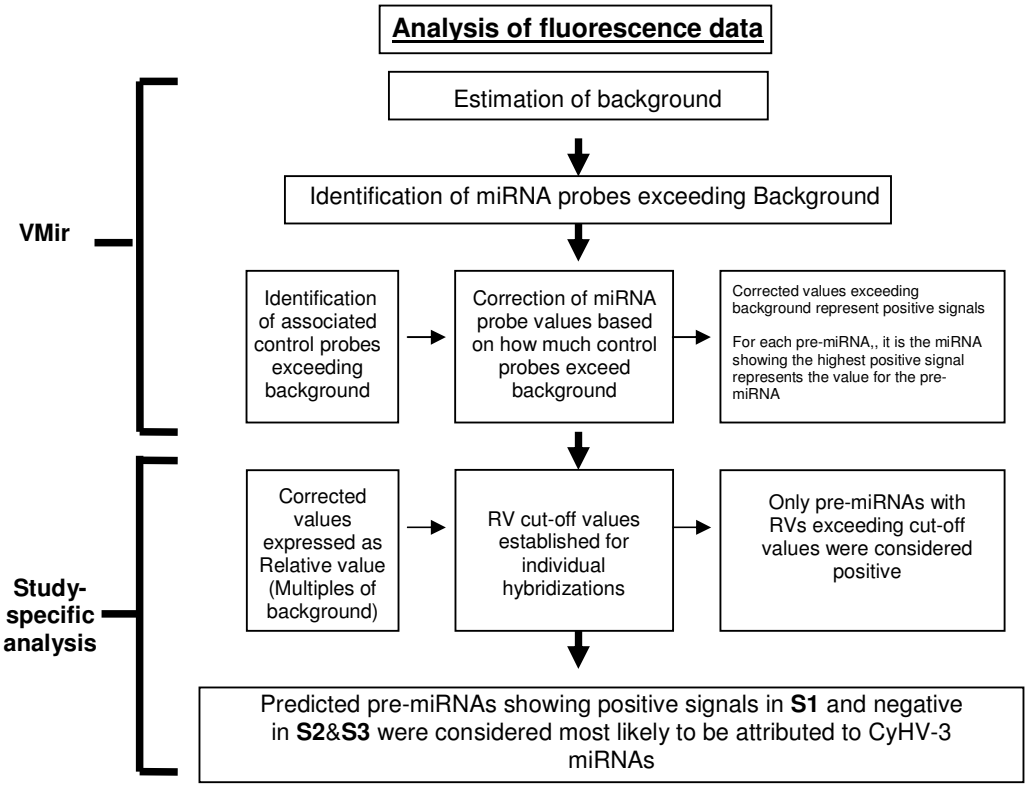



**Figure 2.6 Outline of processes involved in analysis of fluorescence data**

#### 2.8.7.2 *Investigation of new provisional pre-miRNA candidates.*

Pre-miRNAs, outside of the 21 pre-miRNA candidates identified in deep sequencing, were classified as new provisional pre-miRNA candidates and were further investigated by searching the deep sequencing data for any low frequency transcripts (read counts <10) that mapped to them. This was done using SeqMap in a similar manner to the process outlined in Section 2.14.1 and all alignment signatures were assessed as per Section 2.14.3.4.

## 2.9 Northern Blotting

### 2.9.1 Samples

Total RNA from CyHV-3 N076 *in vitro* lytic infections (separate infection to what was used in deep sequencing) and non-infected CCB cells was extracted as per Section 2.5.2. Polyacrylamide gels were ran (200V for 60 mins) as per Section 2.6.

Each gel contained 2 lanes of RNA from infected cells and 2 from non-infected cells (Figure 1.1.9.1) with 10 µg RNA in each. RNA was electroblotted onto nylon membranes as per Section 2.9.2. After cross-linking, nylon membranes were split into 2 pieces and each was probed separately for CyHV-3 putative miRNAs and *Cyprinus carpio* let-7a respectively. The gel loading layout and work flow are illustrated in Figure 1.1.9.1.

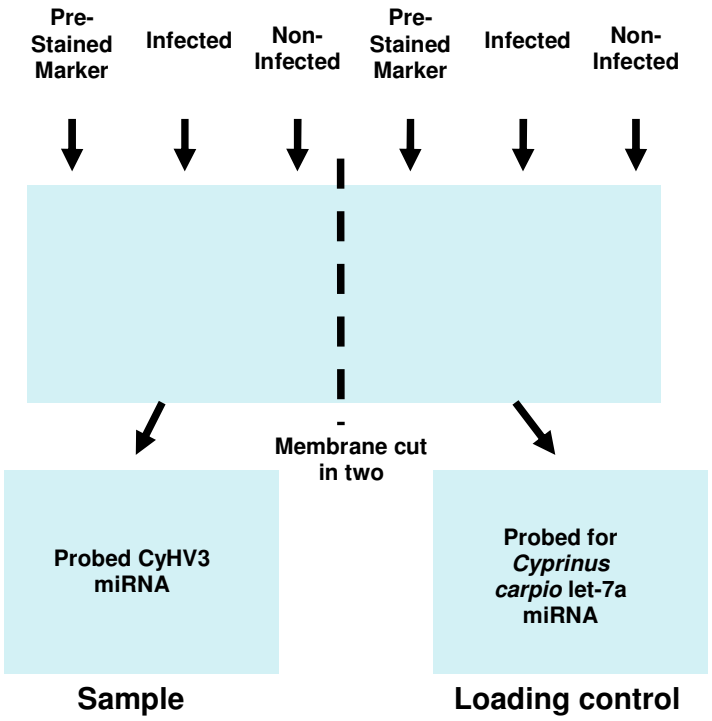



**Figure 2.7 PAGE sample layout and workflow for electroblotting**

## 2.9.2 Electroblotting

Before electroblotting, 2x fibre pads, 2x sheets of filter paper, a nylon membrane and the gels were soaked (15-30 sec) in 0.5x TBE Buffer. These were used to assemble a gel-sandwich in an open electro blotting cassette, see Figure 2.8. Once assembled, the electroblotting cassette was closed, locked and placed in the electroblotting module in the gel tank. The gel tank was filled up to the mark maximum with 1x TBE buffer, and electroblotting was carried out using 400 mA for 30 mins. This was enough time to remove all of the RNA, and loading dye from the gel. Cross-linking was done by baking membranes in an oven at 80$^{o}$C for 4 hours. If not being used

immediately the blotted membrane was placed in a 50ml sterile universal tube (rolled up, not overlapping, sample side facing the inwards) and stored at -20$^{o}$C.

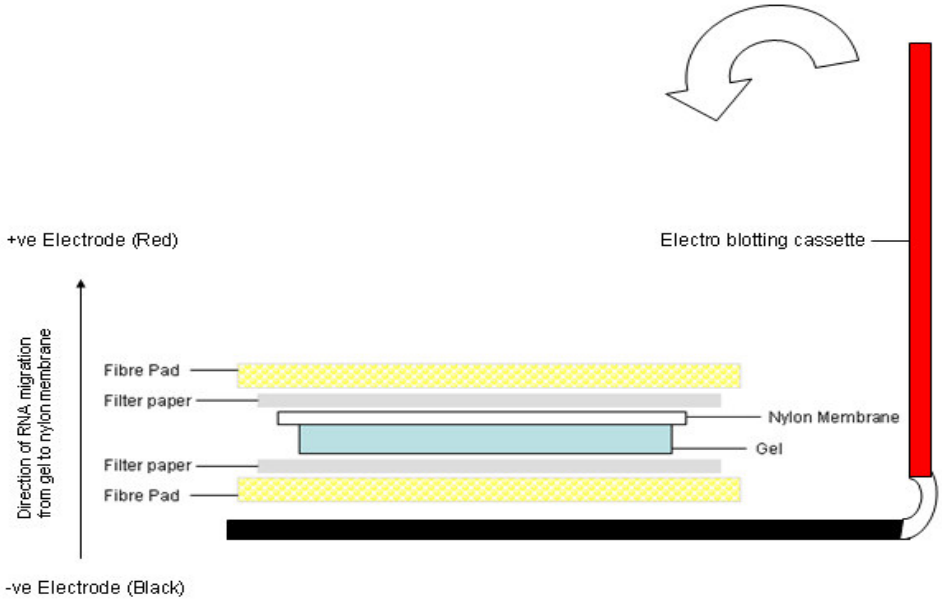



**Figure 2.8 Diagram of gel sandwich assembly**

## 2.9.3 Probes

Probes targeting the CyHV-3 putative miRNAs were synthesized by IDT. In addition, as a loading control *Cyprinus carpio* Let-7a miRNA was also targeted. All probes were 5' biotin-labelled; see probe sequences in Table 2.15.

**Table 2.15 List of probe sequences**

| Assay Name | Probe Sequence |
|---|---|
| CyHV-3 MR5057 3' miRNA | 5' TCGTCGACAGCGGCCGCAATTT 3' |
| CyHV-3 MD11776 3' miRNA | 5' GCCAACACCCTCACGCAGGCCT 3' |
| CyHV-3 MD1111 3' miRNA | 5'ACCACAACAGACTGAGACCATCCT 3' |
| CyHV-3 MD1111 5' miRNA | 5' TGATGATGATGAGCCTGCTGCT 3' |
| CyHV-3 MR5075 3' miRNA | 5' CCTTGCAGTTGATACTAGCACCGT 3' |
| *Cyprinus carpio* Let-7a miRNA | 5' AACTATACAACCTACTACCTCA 3' |

### 2.9.4 Preparation of mixtures/solutions for hybridization washing and detection

#### 2.9.4.1 *Biotin labelled probes*

Lyophilized probe was first centrifuged at 5000 x g for 1 min to ensure all probe was removed from the sides of the tube. Each probe was then resuspended to a final concentration of 1 nmol / µL using 10mM Tris-HCl and stored at – 20$^{o}$C.

#### 2.9.4.2 *Pre-hybridization buffer*

ULTRAhyb®-Oligo buffer was removed from storage at 4$^{o}$C and preheated at in a water bath 68$^{o}$C for at least 30 mins immediately before use to get rid of any precipitants.

#### 2.9.4.3 *Diluted probe for hybridization*

Immediately before use, 1 µL of probe (1 nmol/µL) was diluted in 1 mL pre-heated ULTRAhyb-Oligo buffer and mixed well by pipetting.

#### 2.9.4.4 *Low Stringency Wash Buffer*

If precipitation was present it was removed by heating to 37$^{o}$C until precipitants dissolved. There were 3 washing steps in total. Each washing step required 20 mL of Low Stringency Wash Buffer. 3x 50 mL universals containing 20 mL of Low Stringency Wash Buffer were prepared. Two of these were cooled to room temperature (~22$^{o}$C) and one was kept in a water bath at 37$^{o}$C.

#### 2.9.4.5 *1x Wash buffer (for detection)*

The 5X stock solution of Wash Buffer was pre-heated at 37$^{o}$C for 15-20 mins (varied depending on volume) or until precipitants dissolved. There were 5 wash steps in total. Each wash step required 20 mL of 1x Wash Buffer. 5x 50 mL universals each containing 20 mL of 1x Wash Buffer were prepared by diluting 4 ml Wash Buffer in 16 mL distilled water. This was prepared as needed and not stored.

#### 2.9.4.6 *Blocking Buffer*

The 1X Blocking Buffer was pre-heated at 37$^{o}$C for 15-20 mins (varied depending on volume) or until precipitants dissolved There were 4 blocking steps in total. Each blocking step required 20 mL of 1x Blocking Buffer. 4x 50 mL universals each containing 20 mL of 1x Blocking Buffer were prepared. Buffer was cooled to room temperature before use.

#### 2.9.4.7 *Diluted Streptavidin linked Alkaline Phosphatase (Strep-AP)*

1 µL of Streptavidin linked Alkaline Phosphatase was diluted in 10 mL Blocking Buffer and mixed well by vortexing. This was prepared as needed and not stored.

#### 2.9.4.8 *1x Assay Buffer*

There were 2 wash steps with Assay Buffer in total. Each wash step required 20 mL of 1x Assay Buffer. 2 x 50 mL universals each containing 20 mL 1x Assay Buffer were prepared by diluting 2 ml 10x Assay Buffer in 18 mL distilled water. This was prepared as needed and not stored.

#### 2.9.4.9 *CDP-Star (Substrate)*

This was removed from storage at 4$^{o}$C and 5 mL was dispensed 10 mins before use allowing it to heat to RT. This was stored in the dark until used.

### 2.9.5 Hybridization and washing

A blotted membrane was removed from storage at -20$^{o}$C, placed in blot-box (9.1x 66 cm). For pre-hybridization, 9 ml pre-heated ULTRAhyb®-Oligo buffer was added. The buffer was thoroughly washed over all areas of the membrane manually and the blot-box was placed on a rocking platform and incubated at 37$^{o}$C for 30min. The pre-hybridized membrane was removed from the 37$^{o}$C incubator and 1 ml diluted probe was quickly added using a pipette. This was immediately mixed manually by washing the buffer over membrane several times. The blot-box was placed back on a

rocking platform and incubated at 37°C for 8-12 hours. After hybridization the membrane was put in a new blot-box and incubated for 5 min with 20 mL Low Stringency Wash Buffer at room temperature on a rocking platform. This was repeated again in a new blot-box (blot-box from hybridization rinsed out with distilled water) and a 3rd time in another new blot-box for 2min at 37°C using Low Stringency Wash Buffer pre-heated to 37°C.

### 2.9.6 Chemiluminescent Detection

This was done using components from the BrightStar® BioDetect™ Kit (Ambion, AM1930) All incubation steps in the detection protocol took place on a rocking platform at room temperature. The blot-box was changed and washed with distilled water between all incubations. Buffers were only used once and disposed of after each incubation step. The details of each incubation step are given in Table 2.16

**Table 2.16 Details of incubation steps involved in northern blotting probe detection procedure.**

| Incubation Step | Description | Buffer | Duration |
|---|---|---|---|
| 1 | Wash-Step 1 | 1x Wash buffer | 5 min |
| 2 | Wash-Step 2 | 1x Wash buffer | 5 min |
| 3 | Blocking-Step 1 | Blocking Buffer | 5 min |
| 4 | Blocking-Step 2 | Blocking Buffer | 5 min |
| 5 | Blocking-Step 3 | Blocking Buffer | 30 min |
| 6 | Strep-AP Incubation | Diluted Strep-AP | 30 min |
| 7 | Blocking-Step 4 | Blocking Buffer | 10 min |
| 8 | Wash-Step 3 | 1 x Wash Buffer | 10 min |
| 9 | Wash-Step 4 | 1 x Wash Buffer | 10 min |
| 10 | Wash-Step 5 | 1 x Wash Buffer | 10 min |
| 11 | Reaction Prep 1 | 1x Assay Buffer | 2 min |
| 12 | Reaction Prep 2 | 1x Assay Buffer | 2 min |
| 13 | Substrate Addition | CDP-Star | 5 min |

After incubation step 13, the membrane was lightly blotted on filter paper without letting the membrane dry (to get rid of excess CDP-Star). The membrane was placed in translucent hybridization folder which was then heat sealed using an electronic

sealer. Chemiluminescent signals were visualised and recorded electronically on a BioLuminizer (CAMAG), using 10 minute exposures.

## 2.10 Stem-loop RT-qPCR

Due to their short lengths (~17-25 nt), TaqMan RT-PCR analysis of miRNAs involves a slightly different approach known as Stem-loop RT-qPCR. This was a two step process, see both steps outlined in Figure 2.9.

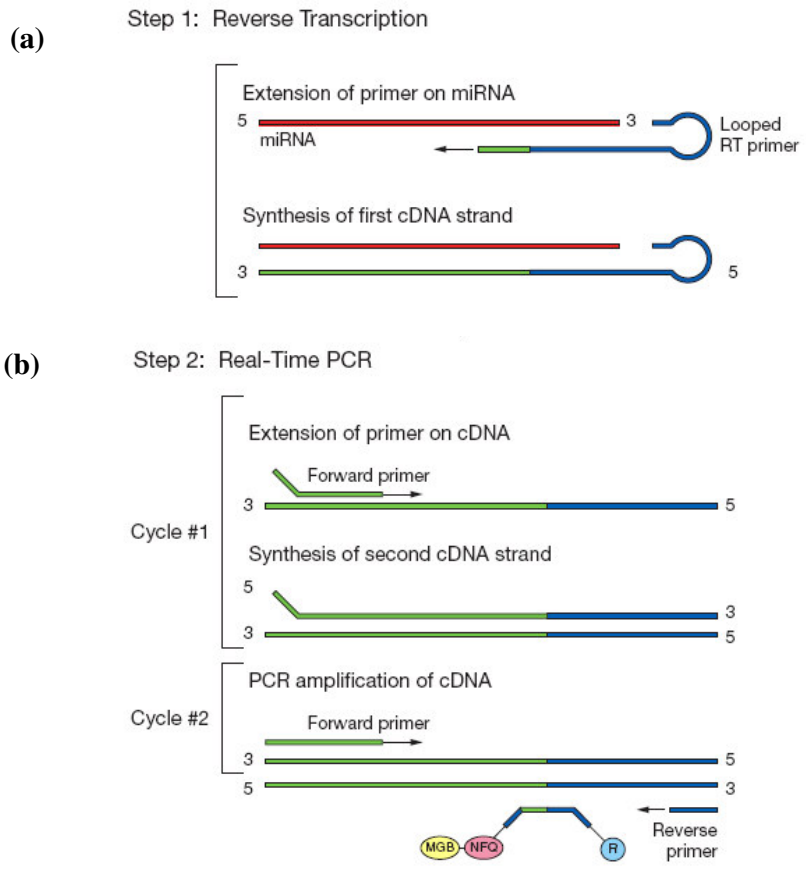



**Figure 2.9 Stem-loop RT-qPCR process**
(a) RT Step: A portion of the RT-primer not specific for the target miRNA forms a stem-loop secondary structure. The presence of this structure prevents the primer from annealing to any larger transcripts containing the same target sequence, as a result in can only anneal to the mature miRNA during the RT step. (b) Real-time PCR: Due to the higher temperatures in the PCR stage the stem-loop denatures. This creates enough space for two small primers and a probe (with one of the primers targeting the denatured stem-loop sequence) allowing TaqMan PCR to be carried out.

### 2.10.1 Primers and probes

Custom probes were designed by submitting target mature miRNA sequences to the Custom TaqMan Small RNA assay Design Tool in the Applied Biosystems website (https://www5.appliedbiosystems.com/tools/smallrna/). The appropriate primers and probes were designed using this online tool however due to the proprietary design process used the sequences of these primer and probes were not disclosed. A list of mature miRNA target sequences are displayed in Table 2.17.

**Table 2.17 List of CyHV-3 miRNA sequences targeted by Stem-loop RT-qPCR**

| Assay Name | Small RNA target Sequence |
|---|---|
| MR5057 3' miRNA | 5' AAAUUGCGGCCGCUGUCGACGA 3' |
| MD11776 3'miRNA | 5' AGGCCUGCGUGAGGGUGUUGGC 3' |
| MD1111 3'miRNA | 5' AGGAUGGUCUCAGUCUGUUGUGGU 3' |
| MD1111 5' miRNA | 5' AGCAGCAGGCUCAUCAUCAUCA 3' |
| MR5075 3' miRNA | 5' ACGGUGCUAGUAUCAACUGCAAGG 3' |
| MR5075 5' miRNA | 5' UUGUAGUUGGUACAACUGCCUGC 3' |
| MD11410 5' miRNA | 5' ACGCGUAGGCCGUGUCCACCUC 3' |
| MD9812 5' miRNA | 5' AGGCCGUCGGCCUUCAUGAUG 3' |

The host miRNA Let-7a (5' UGAGGUAGUAGGUUGUAUAGUU 3') was present at similarly high levels in sequencing experiments. It was therefore used as both an endogenous positive control and a reference gene. As the sequence was found to be highly conserved across most eukaryotic organisms and exactly the same as human Let-7a, it was not necessary to purchase a custom assay to detect this target as a pre-existing human let-7a assay was already available to for purchase.

## 2.10.2 Preparation mixtures/solutions for Stem-loop RT-qPCR

### 2.10.2.1 *Reverse transcription (RT) Reaction Mix*

This was scaled up as required and stored at -20ºC until needed.

**Table 2.18 RT reaction mix components**

| Component | Stock Conc. | Volume | Final Conc. (in 15 µL) |
|---|---|---|---|
| dNTPs | 100mM | 0.15 µL | 1 mM |
| MultiScribe™ RT | 50 U/ µL | 1 µL | 3.33 U/ µL |
| RT Buffer | 10x | 1.5 µL | 1 x |
| RNase Inhibitor | 20 U/ µL | 0.19 µL | 0.25 U/ µL |
| MBG $H_2O$ | N/A | 0.16 µL | N/A |
| **Total** | | 3 µL | |

### 2.10.2.2 *Stem-loop-RT-Primer preparation*

This was scaled up as required and stored at -20ºC until needed. 0.1 x primers were used with *in vitro* samples and 1x primers were used to test *in vivo* samples (see Section 5.1 for reasoning).

**Table 2.19 RT primer preparation using 1x or 0.1x final concentration of RT Primer)**

| Component | Stock Conc. | Volume | Final Conc. (in 15 µL) |
|---|---|---|---|
| 5x RT Primer* | 5x | 3/0.3 µL | 1/0.1x |
| MBG $H_2O$ | N/A | 8/10.7 µL | N/A |
| **Total** | | 11 µL | |

*Where possible, RT reactions were multiplexed, thus several primers were added to this mix and $H_2O$ volumes were adjusted accordingly

#### 2.10.2.3 *PCR-Reaction-Mix (using 1 x or 0.1x final concentration of primer-probe)*

This was scaled up as required and stored at -20ºC until needed. 0.1 x primers were used with *in vitro* samples and 1x primers were used to test *in vivo* samples (see Section 5.1 for reasoning).

**Table 2.20 Stem-loop RT-qPCR mix**

| Component | Stock Conc. | Volume | Final Conc. (in 20 µL) |
|---|---|---|---|
| TaqMan Mix | 2x | 10 µL | 1 x |
| 20x Primer-Probe | 20x | 1/0.1 µL | 1/0.1x |
| cDNA | ~ | 1.3 µL | ~ |
| Nuclease Free Water | N/A | 7.7/8.6 µL | N/A |
| **Total** | | 20 µL | |

### 2.10.3 Stem-loop RT-qPCR Protocol

Total RNA was prepared as per section 2.5.2. RNA concentration was measured using Nanodrop. RNA samples were thawed on ice, 1 µL was taken from each sample and denatured by heating to 70ºC for 5 min in a thin-wall PCR tube and left on ice for 1 min. RT-Reaction-mix and RT-Primer were thawed, 3 µL RT reaction mix and 11 µL RT-Primer (either 1x or 0.1x) were added to the denatured RNA and mixed by pipetting. RT was carried out using following reaction conditions: 30 min at 16 °C, 30 min at 42°C and 5 min at 85 °C. The resulting cDNA was either stored at -20ºC until required or kept on ice temporarily for immediate use in PCR. PCR-Reaction-mix (either 1x or 0.1x final primer concentration) was thawed, vortexed and 18.7 µL was added to individual sample wells on a 96-well plate followed by 1.3 µL cDNA. The plate was sealed and centrifuged at 1000 x g for 1 min. PCR was carried out on an Applied Biosystems 7500 real-time PCR machine using the following PCR conditions: 1 cycle of 2 min at 50°C, 1 cycle of 10 min at 95°C, 40 cycles of 15 sec at 95°C and 60 sec at 60°C (both extension and annealing were carried out at this temperature).

### 2.10.4 Analysis of Real-time PCR data

All PCR data was analysed using either one of the following techniques. Technique 1 was used where the levels of the same miRNA target were being compared (i.e. where taking efficiency into account does not change the relative levels). Technique 2 was used where levels different miRNAs were being compared (where each assay may have a significantly different reaction efficiency).

#### 2.10.4.1 *Technique 1*

Results were analysed using the Applied Biosystems 7500 Software (version 2.0.1). The assay efficiency was taken to be 100% (i.e.1) and individual Ct values were transformed to linear scale expression quantities using Equation 1 and compared to each other.

**Equation 1:** $2^{-Ct}$ (i.e. $(E_{ff} + 1)^{-Ct}$ (where $E_{ff}$ = % Efficiency)

#### 2.10.4.2 *Technique 2*

The raw amplification data, normalized to internal passive reference dye ROX (i.e. Rn data) was recorded using Applied Biosystems 7500 Software. To calculate the PCR efficiency in each individual well, Rn data from each well was exported to an Excel file and reformatted as outlined in. Figure 2.10 This reformatted data was imported into a program called LinReg PCR version 2013.0 (Ruijter et al., 2009).

**(a)**

| | Column 1 | Column 2 | Column 3 | Column 4 | Column 5 |
|---|---|---|---|---|---|
| Row 1 | Well | Cycle | Target Name | Rn | |
| Row 2 | A1 | 1 | MD9812 | 0.107637 | |
| Row 3 | A2 | 2 | MD9812 | 0.109796 | |
| Row 4 | A3 | 3 | MD9812 | 0.109980 | |
| Row 5 | A4 | 4 | MD9812 | 0.110514 | |
| Row 6 | A5 | 5 | MD9812 | 0.110302 | |

**(b)**

| | Column 1 | Column 2 | Column 3 | Column 4 | Column 5 |
|---|---|---|---|---|---|
| Row 1 | | | 1 | 2 | 3 |
| Row 2 | Sample 1_Assay_Name | | 0.10764 | 0.107637 | 0.109980 |
| Row 3 | Sample 2_Assay_Name | | 0.706 | 0.711646 | 0.713133 |
| Row 4 | Sample 3_Assay_Name | | 0.70803 | 0.711791 | 0.715243 |
| Row 5 | Sample 4_Assay_Name | | 0.72205 | 0.724081 | 0.726755 |
| Row 6 | Sample 5_Assay_Name | | 0.70424 | 0.708158 | 0.710543 |

**Figure 2.10 Reformatting of exported raw Rn values for LinReg-PCR**
(a) Format of Rn data exported from Applied Biosystems 7500 Software (b) Rn data reformatted for LinReg PCR. The reformatted data had cycle number in row 1 starting at column 3 with the corresponding fluoresce values for each sample at each cycle in the same positions in the rows underneath. The sample name was in the format Sample_Name_Assay_Name (Assay Name = Amplicon Name in LinReg PCR)

In LinReg PCR each amplicon-group was defined based on the value for "Assay_Name" in the "Amplicon Groups" tab. In "User Settings" tab, the options for excluding samples without amplification and without plateau phases (i.e. negatives) were selected. Also efficiency outliners were defined as samples with efficiencies that deviate by more that 10% from median efficiency for their respective amplicon groups. As a result these outliners were excluded when calculating mean efficiency for their respective amplicon-groups. LinReg-PCR also used the amplification data to estimate an appropriate fluorescence threshold for each amplicon group and indicated Ct values based on group thresholds. Using the mean efficiency ($E_{ff}$) for all samples in each amplicon-group (i.e. each assay), the threshold ($N_{n)}$ and the Ct value (n) for each sample LinReg PCR calculated an arbitrary value ($N_0$) representing the initial target molecule copy numbers in a given sample. These calculations were based on the Equation 3 which was derived from Equation 2, see equations and descriptions in Figure 2.11.

**Equation 2:** $N_n = N_0*(E_{ff} + 1)^n$ ⟶ **Equation 3:** $N_0 = N_n*(E_{ff} + 1)^{-n}$

**Figure 2.11 Description of Equations used for relative quantification in Stem-loop PCR**
Equation 2 describes the exponential amplification of target molecules by PCR where Nn was an arbitrary value (i.e. fluorescence) representing target molecule copy number at cycle n, $N_0$ was an arbitrary value representing target molecule copy number at Cycle 0 (i.e. initial target molecule copy number in sample), Eff refers to PCR reaction efficiency and n represents the number of cycles. This can be rearranged to form Equation 3 which allows $N_0$ to be calculated using values for Nn, Eff and n.

For comparison of separate PCRs derived from a common multiplex RT reaction (where multiple RT primers were used in the same RT reaction to compare levels of different miRNAs in a single sample) no normalization was required. In such cases, the relative levels of different miRNAs were established by directly comparing $N_0$ values. Where levels of miRNAs were compared between separate samples, the $N_0$ values for each target miRNA were normalized by dividing them by $N_0$ values for the reference gene within the same sample (in this case the reference gene was the host gene let-7a).

## 2.11 Real time PCR

CyHV-3 and host DNA levels were measures by real time PCR. The assays used for this were based on those developed by Gilad et al (2004).

### 2.11.1 Preparation mixtures/solutions for real time PCR

#### 2.11.1.1 *Primers*

Lyophilized primers were first centrifuged at 5000 x g for 1 min to ensure all primers were removed from the sides of the tube. A master stock was prepared by re-suspending primers to a concentration of 1 mM using 10mM Tris-HCl. Working stocks of 10 pmol / µL were prepared from these by diluting them 1/100 in 10mM Tris-HCl. Primer stocks were stored at -20 ºC.

#### 2.11.1.2 *Probes*

Custom probes were ordered from Applied Biosystems. Working stocks were prepared of 10 pmol/ µL were prepared by diluting master stocks 1/10 in 10mM Tris-HCl. Primer stocks were stored at -20 ºC.

**Table 2.21 List of CyHV-3 miRNA sequences targeted by Stem-loop RT-qPCR**

| Target | Primer Name | Sequence |
|---|---|---|
| CyHV-3 Genomic DNA | KHV86f | 5' GACGCCGGAGACCTTGTG 3' |
| | KHV163r | 5' CGGGTTCTTATTTTTGTCCTTGTT 3' |
| | KHV109p | FAM 5'CTTCCTCTGCTCGGCGAGCACG 3' MGB |
| *Cyprinus carpio* Genomic DNA | KoiGluc162f | 5' ACTGCGAGTGGAGACACATGAT 3' |
| | KoiGluc230r | 5' TCAGGTGTGGAGCGGACAT 3' |
| | KoiGluc185p | VIC 5' AAGCCAGTGTCAAAAGCCAGTGTCAA 3' MGB |

#### 2.11.1.3 *PCR-Reaction-Mix (using 1 x or 0.1x final concentration of primer-probe)*

This was scaled up as required and stored at -20ºC until needed. 0.1 x primers were used with *in vitro* samples and 1x primers were used to test *in vivo* samples (see Section 5.1 for reasoning).

**Table 2.22 Stem-loop RT-qPCR mix**

| Component | Stock Conc. | Volume | Final Conc. (in 20 µL) |
|---|---|---|---|
| TaqMan Mix | 2x | 12.5 µL | 1 x |
| Forward primer | 10 pmol/ µL | 1.5 µL | 300 nM |
| Reverse primer | 10 pmol/ µL | 1.5 µL | 300 nM |
| Probe | 10 pmol/ µL | 0.5 µL | 100 nM |
| Nuclease Free Water | N/A | 4 µL | N/A |
| **Total** | | 20 µL | |

### 2.11.2 Real-time PCR protocol

PCR-Reaction-mix was thawed, vortexed and 20 µL was added to individual sample wells on a 96-well plate followed by 5 µL of DNA (1000 ng) The plate was sealed and centrifuged at 1000 x g for 1 min. PCR was carried out on an Applied Biosystems 7500 real-time PCR machine using the following PCR conditions: 1 cycle of 2 min at 50°C, 1 cycle of 10 min at 95°C, 40 cycles of 15 sec at 95°C and 60 sec at 60°C (both extension and annealing were carried out at this temperature). As the relative levels of different targets were not being compared, there was no need to take reaction efficiency into account. All PCR data was analysed using Technique 1.() CyHV-3 genomic DNA levels were normalised to host cell DNA levels by dividing CyHV-3 linear scale expression values by the corresponding values for Cyprinus carpio genomic DNA in the same sample.

## 2.12 De Novo Prediction of CyHV-3 pre-miRNAs

### 2.12.1 De Novo prediction of pre-miRNAs in viral genomes

Pre-miRNAs were predicted in viral genomes using VMir Analyser (v2.3) (Grundhoff et al., 2006). Viral genomes analysed included CyHV-3 (DQ657948.1) genomes of 6 other herpesviruses known to encode a combined total of 77 pre-miRNAs (details are available in Supplementary File 3.1 T.1) for viruses and pre-miRNAs) Each genome was analysed using the same settings; Genome Conformation: Linear, Window size: 500, Step size: 10, Minimum HP size: Any, Maximum HP size: Any, Min. HP Score: Any. Max. Size Calc.Score: 100, Max. Size Assign SHPs: 50, Max. Size Assign RHPs: 50. Results were viewed using VMir Viewer (v1.5).

### 2.12.2 Establishment of filters based on characteristics of known viral pre-miRNAs

VMir pre-miRNA predictions from 6 other herpesviruses genomes known to encode pre-miRNAs were screened to see if these known pre-miRNAs were correctly predicted in each viral genome. The pre-miRNAs searched for were based on sequences listed in miRBase at the time of the study (MirBase release 13). This was

done using the "Find Sequence" tool in VMir Viewer. Once identified, relevant attributes such as size, Relative Window Count (WC) and VMir Score were recorded for each correctly predicted pre-miRNA. For each virus, the minimum, maximum and average values for these attributes were calculated. These were used to establish the values typically displayed by bone-fide herpesvirus pre-miRNAs present in VMir output. These were used to establish relevant cut-off values that could be used to filter VMir pre-miRNA predictions from the CyHV-3 genome that did not have similar characteristics to known viral pre-miRNAs. To aid in identification of the optimum cut-off value, the sensitivity of potential cut-off values were estimated by calculation the percentage of the correctly predicted pre-miRNAs passing.

### 2.12.3 Further assessment of miRNAs passing cut-off values

After imposing the cut-offs values each remaining predicted pre-miRNA was further assessed based on genome position, any predictions occurring in ORFs were eliminated. This was done by searching for the pre-miRNA sequences in the CyHV-3 genome map using SeqBuilder. Any predicted pre-miRNAs occurring outside ORFs were analysed in more detail using MiPred (see implementation in section 2.14.3.2. In addition to MHPs, SHPs were analysed also. Through this analysis each remaining predicted pre-miRNA was classified as either a real pre-miRNA, a pseudo pre-miRNA or not pre-miRNA like at all.

## 2.13 Mapping small RNA deep sequencing reads to the genome

Quality filtered reads from deep sequencing were collapsed into a sequence tag file. This was a tab delimited text file consisting of a non redundant list of unique reads sequenced in one column and their associated read-count in a second column. Before mapping reads to the CyHV-3 genome, they were pre-processed. Using Excel, sequences containing "Ns" were removed using filter tool. Using a formula for counting the amount of characters in individual cells, the length of each read was compiled and using the sorting function, transcripts with lengths less than 15 nt or greater than 26nt were removed. Each sequence was given unique ID and read count information was also incorporated onto the end of each ID separated by an underscore using the concatenate function in Excel. The file was converted to fasta using the perl script tabtofasta.pl (see Section 2.19.2 for use) and any remaining 3'

adaptor sequences were removed using perl script clip_adapters.pl (see section 2.19.2 for use of clip_adapters.pl) These reads were mapped to the CyHV-3 genome with no mismatches using SeqMap (Jiang and Wong, 2008). It acted as a useful source of data for several other aspects of this study and was referred to as the Genome-SeqMap-Output and is available in Supplementary File 4.1 T.1-T.4.

## 2.14 Non-automated identification of CyHV-3 pre-miRNAs from deep sequencing data

### 2.14.1 Mapping highly-expressed CyHV-3 derived small RNAs to VMir predicted CyHV-3 pre-miRNAs

Viral transcripts with read counts ≥10 were retrieved from the Genome-SeqMap-Output. All highly expressed CyHV-3 derived transcripts were mapped to the 5' and 3' arm sequences of VMir predicted CyHV-3 pre-miRNAs and CyHV-3 ORFs using SeqMap (both sets of reverence sequences were first converted to fasta as per Section 2.13).

### 2.14.2 Selection of pre-miRNA candidates from mapping data

The data from the mapping CyHV-3 derived small RNAs to CyHV-3 protein coding sequences and predicted pre-miRNA sequences was combined with the Genome-SeqMap-Output in order to identify highly abundant RNAs, from non-coding regions mapping to the arms of predicted pre-miRNAs. For this, the Genome-SeqMap-Output was separated into reads mapping to forward and reverse strands and both groups were sorted in order of 5'starting position. This allowed easy manual identification of large stacks of overlapping reads mapping to the same locus (i.e. many reads with similar 5'start positions offset by ~ +/- 4 nt relative to each other). Stacks containing transcripts with read counts ≥10, mapping to predicted pre-miRNAs and not mapping to ORFs (all this information was contained within separate columns), were selected for further inspection. The pre-miRNA positions of the most abundant transcript in a given stack (putative miRNA) were manually inspected in VMir Viewer to check if they occurred on the stem adjacent to the terminal loop (i.e. roughly consistent with the model of miRNA biogenesis). The pre-

miRNA sequence was also retrieved from the VMir output and its genomic position was checked on the CyHV-3 genome map in SeqBuilder to make sure that none of the extended sequence overlapped with protein coding sequences. The sequencing data was checked for further stacks or individual transcripts mapping to the other arm of the same predicted pre-miRNA being inspected (usually the next/previous transcripts in the list as sequencing data was arranged in order of 5' starting position). Other reads with read counts ≥10, not occurring in stacks but mapping to predicted pre-miRNAs outside ORFs were also inspected in the same way. This allowed the identification of potential miRNAs (most abundant), isomiRs and moRNAs (microRNA-offset-RNAs) from predicted pre-miRNAs. Predicted pre-miRNAs that were identified as (i) having small RNAs mapped to their 5' and 3' arms adjacent to the terminal loop (including at least one read with a read count ≥10) and (ii) not overlapping with any ORFs were noted as provisional pre-miRNA candidates. Only those provisional pre-miRNA candidates that could be identified in more than one experiment, due to the same pattern of small RNA mapping (even if none of these transcripts had a read count ≥10 in the second experiment) were defined as pre-miRNA candidates. These were subjected to more in dept analysis.

### 2.14.3 In depth analysis of pre-miRNA candidates

This was done through the methods outlined below

#### 2.14.3.1 *Mature miRNA duplex structure assessment*

The structures of these pre-miRNA candidates were copied (in text form) from VMir Viewer and the positions of the proposed 5' and 3' miRNAs mapping to their stems were highlighted. Once the miRNAs were highlighted, the predicted mature miRNA duplex structure was inspected for the presence of 3' overhangs, a characteristic of Drosha and Dicer processing.

#### 2.14.3.2 *Pre-miRNA structure assessment*

The structure of each pre-miRNA candidate was also analyzed using two different pre-miRNA classifiers, MiPred and CSHMM-Method. MiPred was accessed at http://www.bioinf.seu.edu.cn/miRNA/ and CSHMM-Method was accessed at

http://web.iitd.ac.in/~sumeet/mirna/. In all cases the sequence analysed was that of the most stable version to form within its local sequence context (as predicted by VMir) i.e. either the MHPs or SHPs of the predicted pre-miRNA that showed the highest absolute-WC and containing the full length putative miRNA sequences. In the case of pre-miRNAs also containing candidate moRNAs, the most stable versions containing the combined full length miRNA and moRNA sequences were also analysed. Through these methods, pre-miRNA candidates were classified either pre-miRNA-like or non-pre-miRNA-like.

#### 2.14.3.3 *Visualisation of mapped CyHV-3 reads and enrichment quantification*

SeqMonk (Version 0.16.0) was used to visualize the Genome-SeqMap-Output created in Section 2.13. It accepts genomes in EMBL format only. However, most viral genomes (including CyHv-3) were only available in GenBank format. As a result the GenBank file for the CyHV-3 genome (DQ6579481.1) was downloaded and converted to EMBL using the online tool Readseq (http://www.ebi.ac.uk/Tools/sfc/readseq/). Once converted to EMBL, the file was saved as a DAT file and placed in the SeqMonk genome directory. The structure of the accession line in the EMBL file was also modified so that it could be recognized by SeqMonk. This involved introducing 5 additional fields separated by colons. The values used for all 6 fields were as follows:

**1.** Simply the text "chromosome"
**2.** Assembly name/Accession number
**3.** Chromosome number (Just referred to as "Genome" for CyHV-3)
**4.** The starting base of this sequence (normally 1)
**5.** The last base of this sequence (normally the sequence length)
**6.** The direction of the sequence (should always be 1)

Based on this, the accession line from the CyHV-3 EMBL file was changed as outlined below:

```
AC      DQ657948; (original accession line)
AC      chromosome:DQ657948.1:Genome:1:295146:1; (modifications accession line)
```

The Genome-SeqMap-Output from each experiment containing information for each mapped read such as start position, end position, strand (+/-) and chromosome (just referred to as "genome" for CyHV-3 (corresponding with field 3 in the modified accession line of the genome file) was saved as a tab delimited text file for import into SeqMonk. To display quantitative information SeqMonk derives read count information from the mapped data by counting the amount of rows containing identical mapping info (same positions, strand and chromosome). As a result the mapped data was first un-collapsed i.e. modified so that each row was repeated, with the number of repeats equal to the value in the read count column. The data was modified in this way using a script called fastx_uncollapser (see Section 2.19.2 for use) by defining the column in each row containing read count values.

#### 2.14.3.4 *Inspection of miRNA alignment signatures*

All the candidate pre-miRNA sequences, pre-miRNA linear structures (in dot-bracket notation) and reads mapped to each candidate pre-miRNA were retrieved and consolidated in .txt files. Mapped reads were manually aligned to the pre-miRNA sequences and the linear structures. This allowed detailed view of alignment signatures to ascertain if they fitted the model of miRNA biogenesis as outlined by Kozomara and Griffiths-Jones (2011).

#### 2.14.3.5 *IsomiR end heterogeneity assessment*

IsomiRs of putative miRNAs from pre-miRNA candidates that showed miRNA-like alignment profiles were assessed in terms of their degree of 5' and 3' heterogeneity relative to the putative miRNAs themselves. For each putative miRNA, the start and end positions of their associated isomiRs (defined as transcripts that had ≥60% of their bases overlapping with the miRNA) were expressed in terms of their distance (nt) from the miRNA start and end positions. IsomiRs that were <19nt in length or those representing <0.1% of the combined miRNA and isomiR read count (for the miRNA in question) were not included in analysis. For each isomiR, the start and end positions (relative to miRNA start and end positions) were calculated e.g. taking the 5' end of the miRNA to be 0, and isomiRs with 5' positions that were 1 nt before the miRNA were assigned an offset value of -1 for its 5' end (absolute values were

used in calculations) to indicate the degree of 5' end heterogeneity. The 3' ends of all isomiRs were assigned offset values in the same manner. The offset-values (representing the degree of 5' and 3' heterogeneity) for all isomiRs of a given miRNA were compared in three ways. The number unique isomiRs showing 5' and 3' end heterogeneity respectively were counted and compared using the "Unique 3' End: 5' End Off-set IsomiR Ratio" (Ratio-1), where ratios greater than 1 indicated that there were more unique isomiRs displaying 3' end heterogeneity. The overall degree of heterogeneity displayed by 5' and 3' ends of all unique isomiRs relative to the start and end positions of the miRNA was established by calculating the average off-set values for both ends of all isomiRs (even if one end was not off-set) and comparing them using the "3' End: 5' End Average Absolute Off-set Ratio" (Ratio-2), where ratios greater than 1 indicated that there was a greater degree of 3' end heterogeneity. Finally, the overall degree of 5' and 3' end processing consistency among isomiRs was established by comparing the combined read count of all isomiRs offset at the 5' end to that of all isomiRs offset at the 3' end using the "3' End :5' End Processing Consistency Ratio" (Ratio-3).

#### 2.14.3.6 *Identification of high probability pre-miRNA genes*

The results from in-dept analysis of pre-miRNA candidates were used to identify high probability pre-miRNAs from the pre-miRNA candidates. This was done on the basis of compliance with the pre-miRNA identification criteria outlined below. This criteria was based on the criteria outlined by Kozomara and Griffiths-Jones (2011) ( described in Section 1.7.2.2) with some additions to improve stringency.

**Pre-miRNA identification criteria**

1. At least one of the putative miRNAs from the same precursor must have a read-count of ≥10 or more in at least one infection.
2. The pre-miRNA candidate must not map to a protein coding region.
3. The putative miRNAs must be derived from the stems of the pre-miRNA candidate hairpin structure in close proximity to loop.
4. Each putative miRNA must also be accompanied by major/minor strand in *both* infections.

5. The same transcript (or close isomiR differing by a maximum of +/- 1nt on the 5' end) must be identified as the putative miRNA (dominant read from a given pre-miRNA arm) in *both* infections.
6. The pre-miRNA candidate must be classified as a real pre-miRNA by both MiPred and CSHMM-Method.
7. The pre-miRNA candidate structure must have an MFE < -25 kcal/mol.
8. The proposed mature-miRNA-duplex must display 3' overhangs (or may have a maximum of one blunt end) in *both* infections.
9. Alignments of putative miRNAs/moRNAs pre-miRNAs must resemble miRNA-like alignment signatures in *both* infections.
10. For a given pre-miRNA candidate, the isomiRs of at least one of the putative miRNAs must have a Ratio-1 value that is ≥1 in both infections and a Ratio-3 value that is ≥1 in at least one infection.
11. Reads mapped to pre-miRNA candidate loci must collectively form discrete stacks and must be enriched for mapped reads compared to flanking genomic regions.

## 2.15 Automated identification of CyHV-3 pre-miRNAs from deep sequencing data

Due to the benefits of combining output from different methods (as discussed in Section 2.14.3.6), two methods for automated identification of novel miRNAs from deep sequencing data were used. These were MirDeep2 and Mireap and their use in this study is described below

### 2.15.1 MiRDeep2

The process of identifying miRNAs using miedeep2.pl involved the used of two other programs to prepare the appropriate input files. Although the deep sequencing data had already been mapped to the CyHV-3 genome using SeqMap, the mapping file used by mirdeep2.pl needed to be in .arf format. Another script called mapper.pl generated mapped reads in this format. However, mapper.pl required the reference genome to be first broken up into cluster files with a Burrows-Wheeler index. This is a form of compression and is necessary in order to keep the memory footprint of the

mapping process small when working with large genomes. Memory footprint is not an issue when dealing with genomes in size range of herpesviruses, but the genome file was still prepared in this way purely for compatibility purposes This was carried out using a script called Bowtie. All of these steps are outlined below.

1. **Genome Indexing (using Bowtie):** The input file required by Bowtie consisted of CyHV-3 (Ac. DQ657948.1) genome in fasta format. Using the "build" function the genome was indexed into 6 cluster files all recognized by a defined common prefix in the file names (see Section 2.19.2 for use of Bowtie).These files could now be collectively used as a reference for mapper.pl.

2. **Mapping (using mapper.pl):** Information from the Genome-SeqMap-Output generated in Section 2.13 was used to make a tab delimited .txt file with CyHV-3 derived reads and read counts in one column and sequences in the other, see format below.

```
READ_ID1_xRead Count   NNNNNNNNNNNNNNN
READ_ID2_xRead Count   NNNNNNNNNNNNNNN
READ_ID3_xRead Count   NNNNNNNNNNNNNNN
READ_ID4_xRead Count   NNNNNNNNNNNNNNN
```

   This tab delimited text file was converted into fasta using the script tabtofasta.pl. The indexed genome clusters generated by Bowtie and the new fasta file containing the reads were both used as input by mapper.pl (see Section 2.19.2 for use of mapper.pl) which mapped these reads to the CyHv-3 genome and generated output in .arf format which could be read by mirdeep2.pl .

3. Novel miRNA identification (using mirdeep2.pl): The output .arf file from mapper.pl, the fasta file containing the reads and the fasta file originally used to generate the indexed genome were used as input for mirdeep2.pl (see Section 2.19.2 for use of mirdeep2.pl). The resulting output consisted of pre-miRNA and miRNA candidates, their scores, read signatures and coverage plots.

### 2.15.2 Mireap

Unlike mirdeep2.pl, Mireap required mapped reads in a basic format. For this, the information from the Genome-SeqMap-Output generated in Section 2.13 was used to make a tab delimited .txt file with read ID, chromosome (with the value used corresponding to the fasta identifier for the reference genome) and genome positions see format below.

```
READ ID        Chromosome     Start   End
READ_ID1       KHV_U          NNN     NNN
READ_ID1       KHV_U          NNN     NNN
READ_ID1       KHV_U          NNN     NNN
```

The information in the Genome-SeqMap-Output was also used to make a tab delimited .txt file with CyHV-3 derived reads and read counts in one column (separated by an underscore) and sequences in the other, see format below.

```
READ_ID1_Read Count    NNNNNNNNNNNNNNN
READ_ID2_Read Count    NNNNNNNNNNNNNNN
READ_ID3_Read Count    NNNNNNNNNNNNNNN
READ_ID4_Read Count    NNNNNNNNNNNNNNN
```

This tab delimited text file was converted into fasta using the script tabtofasta.pl Also the CyHV-3 genome was also required in fasta format. These 3 files were used as input for Mireap (see Section 2.19.2 for use of Mireap). The resulting output consisted of pre-miRNA and miRNA candidates and their associated read signatures

### 2.15.3 Filtering and comparison of results from two automated methods

Before comparison, results were filtered by elimination of predictions occurring ORFs (using genome map in Seqbuilder) and elimination of predictions with no corresponding minor strand. Remaining predicted pre-miRNAs from the two methods were renamed based on their VMir name (all were also predicted by VMir). Predictions that were not common to both MirDeep2 and Mireap were eliminated. These were compared to the results obtained from manual identification of miRNAs from the deep sequencing data.

## 2.16 Comparison of CyHV-3 seed regions to known host and viral miRNA seed regions

All common carp (*Cyprinus Carpio carpio*) and viral miRNAs retrieved from MiRBase (Release 18) in fasta format. To facilitate manipulation in Excel, this fasta file was converted to tab delimited format using fastatotab.pl (see Section 2.19.2 for use of fastatotab.pl). In Excel, columns containing miRNA sequences were broken up by inserting column breaks between characters 1-2 and 8-9, thus isolating the seed sequence (positions 2-8) in a separate column. The remaining parts of the miRNA sequences were deleted. This file (now containing miRNA IDs and seed regions in two separate columns) was saved as a tab delimited text file and converted back into fasta format using tabtofasta.pl (see Section 2.19.2 for use of tabtofasta.pl). The CyHV-3 miRNA seed regions were formatted in the same way in a separate fasta file. CyHV-3 seed regions mapped to viral and common carp seed regions from MiRBase using SeqMap allowing 1 mismatch (see Section 2.19.2 for use of SeqMap). MiRNAs with matching seeds were also compared in full length in Excel.

## 2.17 Search for CyHV-3 miRNA homology in other CyHVs

All CyHV-3 miRNAs and pre-miRNAs were aligned to CyHV-1 (Ac. JQ815364.1) and CyHV-2 (Ac. JQ815363.1) genomes using BLAST. For more specific homology searches, genomic regions from other CyHV genomes that were directly orthologous to regions containing miRNAs in the CyHV-3 genome were selected. This was done by checking if the ORFs occurring opposite all CyHV-3 miRNAs had orthologs in either of the other two CyHVs. This was checked by referring to Davison et al. (2012), and all ORFs opposite to CyHV-3 high probability pre-miRNAs were found to have orthologs in CyHV-2 and CyHV-3 with the exception of ORF-7 (opposite MD1111) which had no ortholog in CyHV-1. The sequences of these orthologs in other CyHVs were retrieved from GenBank and reverse complemented (using ReCo RoKo tool) to obtain the sequence of the non-coding regions from other CyHVs that were directly orthologous to the non-coding regions in CyHV-3 that encoded miRNAs. CyHV-3 pre-miRNA sequences were aligned to their respective orthologous non-coding regions in CyHV-1 and CyHV-2 using BLAST i.e. two sequences were aligned using blastn program available online

(http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch). High quality matches (with E-values <1E-8) were aligned to CyHV-3 pre-miRNAs using Multalin (http://multalin.toulouse.inra.fr/multalin/). Both of these methods performed the same task but BLAST gave more informative statistical output while Multalin gave more visually informative output. In essence, the best features of both approaches were combined.

## 2.18 MiRNA Target Prediction

In this study 2 programs were used to predict miRNA targets for CyHV-3 miRNAs in 3' UTRs of CyHV-3 genes.

### 2.18.1 Retrieval of 3' UTR sequences

CyHV-3 3' UTR sequences were retrieved from Seqbuilder. As no CyHV-3 3' UTRs were annotated at the time of this study, for practical purposes the 3; UTRs for each ORF were defined as the sequence between the stop codons and the nearest downstream polyA signal. There are more ORFs than polyA signals in the CyHV-3 genome. As a consequence, many ORFs occur in the 3' UTRs of upstream ORFs. Where this occurred, the 2 overlapping 3' UTRs were included in analysis. UTRs that also contained ORFs had the names of these ORF(s) were mentioned in their 3' UTR names, see example in Figure 2.12.

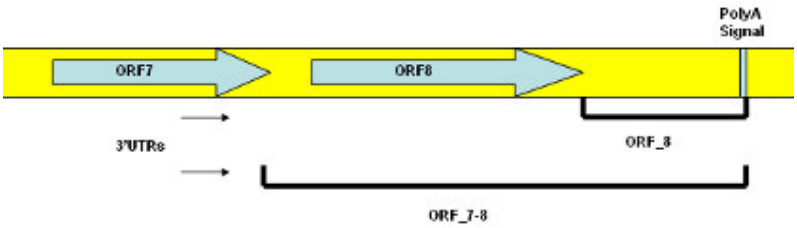



**Figure 2.12 Illustration of ORF naming system**
In this diagram, the 3' UTR for ORF7 overlaps with the 3'UTR and coding region for ORF8. Hence the retrieved 3' UTR sequence for ORF7 was named ORF_7-8.

### 2.18.2 TargetScan

1. **targetscan_60.pl:** The file containing the miRNAs for target prediction consisted of a tab delimited .txt file with information such as: miRNA ID, miRNA sequence (positions 2-8) and species ID (this was only relevant to studies taking evolutionary conservation into account, for the purposes of this study a value of 1 was used for this field for each miRNA), see format below.

```
miRNA_ID1       NNNNNNNN        1
miRNA_ID2       NNNNNNNN        1
miRNA_ID3       NNNNNNNN        1
miRNA_ID4       NNNNNNNN        1
```

   The file consisting of the 3' UTRs was formatted in a similar way except species ID was in the middle column (a value of 1 was used for all 3' UTRs, corresponding to the value used for all miRNAs above), see format below.

```
3'UTR_ID1       1       NNNNNNNNNNNNNNN.............
3'UTR_ID2       1       NNNNNNNNNNNNNNN.............
3'UTR_ID3       1       NNNNNNNNNNNNNNN.............
3'UTR_ID4       1       NNNNNNNNNNNNNNN.............
```

   These two files were used as input for the script targetscan_60.pl (see Section 2.19.2 for use of targetscan_60.pl). The output consisted of information such as name of 3' UTR containing possible target site and a description of the site type (see Section 1.8 for details). This was used as input for targetscan_60_context_scores.pl

2. **targetscan_60_context_scores.pl:** The output from targetscan_60.pl, the tab delimited .txt file containing 3'UTR sequences and an extra dab delimited .txt file containing full length miRNA sequences were used as input for this script, see format for this extra file below. Again, the value used for species ID was 1, corresponding to what was used in all other input files. MirBase_ID was not applicable to this study as none of the miRNA sequences that were used were in MirBase.

```
miRNA_family_ID         Species_ID      MiRBase_ID      Mature_sequence
miRNA_ID1               1               n/a             NNNNNNNNNNNNNNNNNNNN
miRNA_ID2               1               n/a             NNNNNNNNNNNNNNNNNNNN
miRNA_ID3               1               n/a             NNNNNNNNNNNNNNNNNNNN
miRNA_ID4               1               n/a             NNNNNNNNNNNNNNNNNNNN
```

**targetscan_60_context_scores.pl** was implemented as per see Section 2.19.2. The output consisted of a summary of each predicted target site in the form of a breakdown of to how they scored in terms of their consistency with the characteristics of known miRNA target sites. This was all summed up in the form of a context+ score. A percentile rank column was also added to the output in order to help assess the significance of individual context+ scores.

### 2.18.3 PITA

The input for PITA consisted of 2 fasta files containing full length miRNAs and 3' UTRs respectively. The analysis was implemented as per Section 2.19.2. A percentile rank column was also added to the output in order to help assess the significance of specific ΔΔG scores.

### 2.18.4 Comparison of TargetScan and PITA results

The top ranking predicted target sites from both TargetScan and PITA were compared. A prediction was considered probable if ranked within the top 10 TargetScan and PITA predictions for CyHV-3 UTRs.

## 2.19 Bioinformatics scripts used

### 2.19.1 General Summary

The bioinformatics scripts used in this study comprised of mixture of Linux based and Windows based scripts, see general details of all scripts used in Table 2.23. For convenience Linux was ran (in the form of Ubuntu 12.04 LTS) on a virtual machine within Windows XP using Oracle VM virtualBox manager. A shared folder was set up as a bridge between the two operating systems. This folder was used to carry out all data analysis and as such it was also the default location for all output files as allowed easy sharing of input and output data between the two operating systems.

**Table 2.23 Details of all bioinformatics scripts used in this study**

| Script | Used for | Implemented via |
|---|---|---|
| Tabtofasta.pl | High throughput formatting | Linux terminal |
| Fastatotab.pl | High throughput formatting | Linux terminal |
| Clip_adapters.pl | High throughput formatting | Linux terminal |
| fastx_uncollapser | High throughput formatting | Linux terminal |
| seqmap.exe | Read Mapping | Windows command prompt |
| Bowtie | Genome-indexing, miRDeep | Linux terminal |
| mapper.pl | Read Mapping, miRDeep | Linux terminal |
| Mirdeep2.pl | Novel miRNA identification | Linux terminal |
| Mireap.pl | Novel miRNA identification | Linux terminal |
| Targetscan_60.pl | miRNA target prediction | Windows command prompt |
| Targetscan_60_context_scores.pl | miRNA target prediction | Windows command prompt |
| Pita_prediction.pl | miRNA target prediction | Linux terminal |

### 2.19.2 Implementation of scripts

In all cases, input files were located in the same directories as the scripts used to analyse them. All input files were tab delimited text files (.txt) or fasta files (.fa). No spaces were allowed in any file names as this was incompatible with the use of command lines to implement analysis. Table 2.24 contains information as to how each script was implemented and what specific options were used in this study.

**Table 2.24 (1/2) Implementation of bioinformatics scripts used in this study**

| High throughput formatting tools | |
|---|---|
| **Script** | **Usage** |
| Tabtofasta.pl | `tabtofasta.pl input_tab_delimited_file_name.txt > converted_fasta_file_name` |
| Fastatotab.pl | `fastatotab.pl input_fasta_file_name.txt > converted_tab_delimited_file_name` |
| Clip_adapters.pl | `clip_adapters.pl inpurt_reads_file_name.fa NNNNNADAPTORSEQUENCENNNNNNN > clipped_reald_output.fa` |
| fastx_uncollapser | `fastx_uncollapser -c (column number containing read counts) -v -i Input_filr_name.txt -o output_file_name.txt` |
| **Mapping tools** | |
| **Script** | **Usage** |
| Seqmap | `seqmap-1.0.12-windows 0 ipnut_reads_file_name.fa input_genome_file_name.fa mapped_reads_output_file_name /output_all_matches`<br>Options used:<br>`0` : Number of mismatches allowed = 0<br>`/output_all_matches` : Output format |
| mapper.pl (Used with miRDeep) | `mapper.pl ipnut_reads_file_name.fa -c -l 17 -p bowtie_output_cluster_file_name_prefix -s processed_reads_File_name.fa -t mapped_reads_output_file_name.arf -v`<br>Options used<br>`-c` : input file is fasta format<br>`-l 17` : Removes sequences with lengths >17nt from the reads input reads files in case they interfere with miRNA signals<br>`-s` : Defines name of new reads file with reads >17 removed<br>`-v` : progress report |
| **Genome Indexing tools** | |
| **Script** | **Usage** |
| Bowtie (Used with miRDeep) | `bowtie-build input_genome_file_name.fa output_cluster_file_name_prefix` |

**Table 2.24 (2/2) Implementation of bioinformatics scripts used in this study**

| **Automated miRNA identification tools** | |
|---|---|
| **Script** | **Usage** |
| Mirdeep2.pl | `miRDeep2.pl processed_reads_File_name.fa input_genome_file.fa mapped_reads_output_file_name.arf none none none 2> process_report_log_file_name.log`<br>Options used:<br>**none**: No list of known miRNAs from genome analysed included.<br>**none**: No list of known miRNAs from closely related species included<br>**none**: No list of known pre-miRNAs from genome analysed included. |
| Mireap.pl | `Perl Mireap.pl -i input_reads_file_name.fa -m mapped_reads_file_name.txt -r input_genome_File_name.fa -o /specified_directory/specified_output_folder_name` |
| **MiRNA target prediction tools** | |
| **Script** | **Usage** |
| targetscan_60.pl | `targetscan_60.pl miRNA_seed_list_file_name.txt 3'UTR_list_file_name.txt targetscan_60_output_file_name.txt` |
| targetscan_60_context_scores.pl | `targetscan_60_context_scores.pl full_lenght_miRNA_list_File_name.txt 3'UTR_list_file_name.txt targetscan_60_output_file_name.txt targetscan_60_context_scores_output_file_name.txt` |
| Pita_prediction.pl | `./pita_prediction.pl -utr 3'UTR_list_file_name.fa -mir full_lenght_miRNA_list_file.fa -prefix output_folder_file_name_prefix -m 6;0,7;0,8;0`<br>Options used:<br>**-m 6;0,7;0,8;0:** This indicated that no mismatches were allowed for all lengths of seed region matches (by default one GU wobble was allowed in 7nt and 8nt seed matches. |

# 3 De Novo Prediction of Pre-miRNA-coding genes on the CyHV-3 genome

In order to search for theoretical evidence to suggest that the CyHV-3 genome could encode miRNAs, the genome was analysed using VMir (Grundhoff et al., 2006). This predicted all hairpin structures occurring on the genome and scored them in Reterms of their consistency with general pre-miRNA structural characteristics. Predictions most likely to be genuine were then selected based on the characteristics of known viral pre-miRNAs that were correctly predicted by VMir. The strategy used in this study is outlined below in Figure 3.1

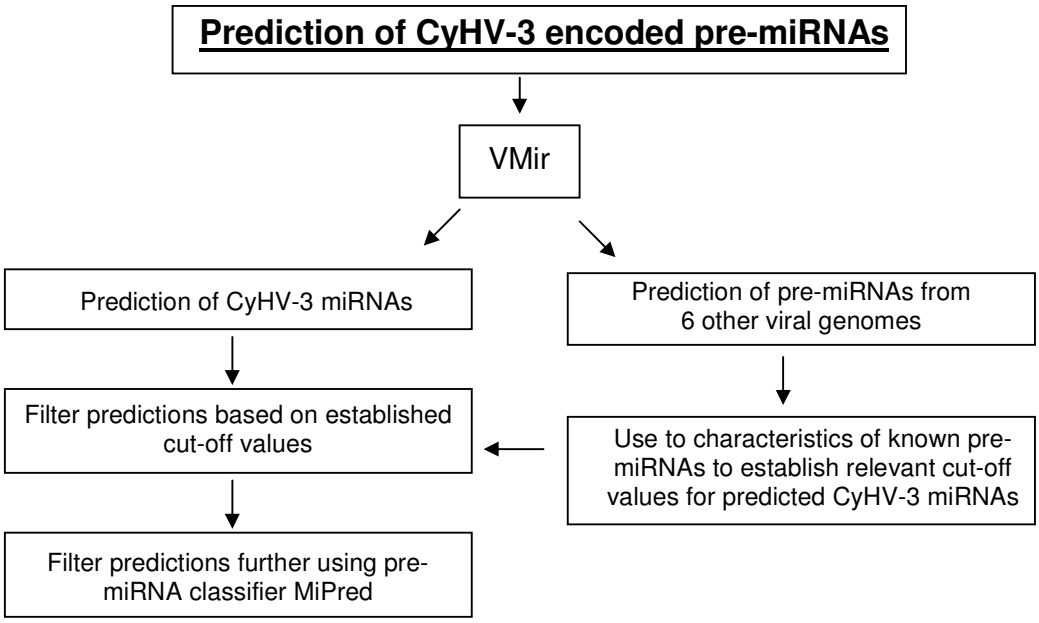



**Figure 3.1 Outline of strategy used for prediction of CyHV-3 pre-miRNAs**

## 3.1 Initial predictions

Each pre-miRNA prediction was represented by a main hairpin (MHP) and shorter variations of this MHP referred to as subsiduary hairpins (SHPs) that were predicted to form in different analysis windows (described in more detail in Section 1.7.1.1). VMir predicted 43,921 MHPs in the CyHV-3 genome. On average each MHP had an average of 1.56 associated SHP variations, resulting in a total of 69,084 individual

pre-miRNAs when these length variants are also counted. A visual representation of these predictions is displayed in Figure 3.2.

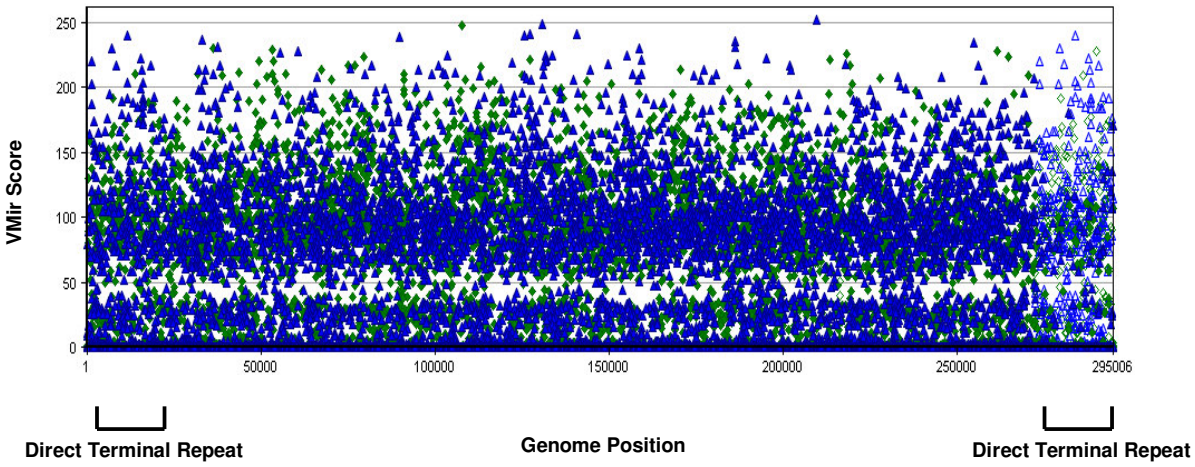



**Figure 3.2 Output from VMir analysis of the CyHV-3 genome**
MHP positions on the forward and reverse strands of the CyHV-3 genome are represented by solid blue triangles and green diamonds respectively. Repeated MHPs are represented by open shapes. As expected, hairpin structures in the left terminal repeat also occur in the right terminal repeat, hence hairpins in the right terminal repeat are denoted here as open shapes.

## 3.2 Identification of relevant cut-off values for predicted pre-miRNAs

In order to establish relevant cut-off values for pre-miRNAs predicted in the CyHV-3 genome, VMir was also used to predict pre-miRNAs on six other herpesvirus genomes known to encode pre-miRNAs. Details of viral genomes analyzed and known pre-miRNAs that were predicted in these genomes are available in Supplementary File 3.1 T.1. The prediction statistics for each of these six viruses are compared to the prediction statistics from CyHV-3 (Table 3.1).

**Table 3.1 Summary of VMir pre-miRNA predictions and statistics on all viral genomes analysed in this study**

Note: The number of known pre-miRNAs from each virus was based on information available from miRBase (Release 13) at the start of this study. Several of these viruses have since had more miRNA genes identified; up-to-date information is available in Table 1.4.

| Virus | EBV | HSV-1 | HCMV | KSHV | MDV-1 | MGHV68 | Mean | CyHV-3 |
|---|---|---|---|---|---|---|---|---|
| **Genome Size (kb)** | 184.11 | 152.26 | 235.64 | 137.50 | 177.87 | 119.45 | 167.80 | 295.00 |
| **Predicted MHPs** | 27042 | 23518 | 34615 | 19187 | 23699 | 16463 | 24087 | 43921 |
| **SHPs** | 43189 | 36976 | 54162 | 30637 | 36023 | 25173 | 37693 | 69084 |
| **Ave. SHPs Per MHP** | 1.60 | 1.57 | 1.56 | 1.60 | 1.52 | 1.53 | 1.56 | 1.57 |
| **MHPs per Kb** | 146.9 | 154.5 | 146.9 | 139.5 | 133.2 | 137.8 | 143.14 | 148.88 |
| **Pre-miRNAs known at time of analysis (MirBase release 13)** | **25** | **6** | **11** | **12** | **14** | **9** | | **N/A** |

All 77 known pre-miRNA genes (based on miRBase Release 13) on these six viral genomes were successfully predicted by VMir. A summary of the sizes, VMir Scores and relative-WC values (referred to as just WC) of all correctly predicted pre-miRNAs from each virus are shown in Table 3.2. These values are based on the MHPs of each predicted pre-miRNA. No correctly predicted pre-miRNA had a score lower than 112 and most (55%) were over the average value of 186. Some pre-miRNAs did have WC values at or close to the lowest possible value of 1, although the majority (89%) had WC values >40.

**Table 3.2 Characteristics of VMir predicted MHPs corresponding to known pre-miRNAs on the other six viral genomes**

| Viral genome | EBV | HSV-1 | HCMV | KSHV | MDV-1 | MGHV68 | Mean |
|---|---|---|---|---|---|---|---|
| **Correctly predicted of known pre-miRNAs** | 25/25 | 6/6 | 11/11 | 12/12 | 14/14 | 9/9 | |
| | | | | | | | |
| Min VMir Score | 152.2 | 124 | 134.3 | 120.5 | 191.6 | 112.3 | 139.15 |
| Max VMir Score | 233.5 | 237.3 | 244.7 | 267.9 | 276.3 | 167.1 | 237.80 |
| **Mean** | **196.06** | **195.25** | **187.76** | **188.63** | **209.00** | **140.72** | **186.236** |
| | | | | | | | |
| Min MHP length | 78 | 80 | 66 | 51 | 69 | 47 | 65.17 |
| Max MHP length | 192 | 140 | 181 | 164 | 139 | 93 | 151.50 |
| **Mean** | **112.36** | **105.50** | **108.73** | **117.17** | **101.27** | **66.56** | **101.930** |
| | | | | | | | |
| Min WC | 43 | 2 | 38 | 11 | 33 | 1 | 21.33 |
| Max WC | 49 | 47 | 48 | 48 | 48 | 48 | 48.00 |
| **Mean** | **45.92** | **24.00** | **44.45** | **40.25** | **44.64** | **40.00** | **39.877** |
| | | | | | | | |

## 3.3 Testing and application of cut-off values

The information in Table 3.2 was used to establish relevant cut-off values for filtering out the least likely CyHV-3 pre-miRNA predictions from Figure 3.2.

### 3.3.1 Pre-miRNA length cut-off values

The mean minimum and maximum pre-miRNA lengths displayed by correctly predicted pre-miRNAs in the 6 other viruses were 65 bp and 152 bp respectively, as shown in Table 3.2. Using these values as a guide, it was decided that low stringency cut-offs of 50 bp and 200 bp for minimum and maximum pre-miRNA lengths would be applied to pre-miRNA predictions from CyHV-3. These cut-off values were slightly outside the values displayed by correctly predicted pre-miRNAs in the six other viruses but they were selected in order to avoid eliminating genuine pre-miRNA predictions that were outliners in terms of minimum and maximum MHP size.

### 3.3.2 WC and VMir Score cut-off testing

Based on the data presented in Table 3.2, the suitability of combinations of relevant cut-offs for WC and VMir Scores were tested. The suitability of each combination of cut-off values was assessed based on percentage sensitivity. These results are shown in Table 3.3. Using the mean WC and VMir Scores as cut-off values gave the worst sensitivity, in theory only allowing the prediction of 49.6% of the 77 correctly predicted known pre-miRNAs. Using the mean WC and lowering the VMir score cut-off to the mean minimum value improved the sensitivity to 77.9%. Lowering the WC to the minimum mean only increased the sensitivity to 81.8%. The fact that 89% of correctly predicted pre-miRNAs had WC values over 40 indicated that it may be important to use a WC cut-off of 40. This explains why lowering the WC cut-off value below 40 did not significantly increase sensitivity. This was also reflected in the difference in percentage sensitivity between using mean minimum WC and mean VMir Score as cut-offs and vice versa (53.2% vs. 77.9%). Thus, the WC cut-off was defined as 40. In order to improve the sensitivity, without reducing the cut-off value for WC below the mean, the cut-off value for VMir Score was further reduced to 112, in line with the lowest score among the correctly predicted pre-miRNAs. This

resulted in the best sensitivity of 83.1% (that is to say, these cut-off values still allowed the detection of 83% of the 77 known pre-miRNAs on the other 6 viral genomes).

**Table 3.3 Estimated sensitivity of different combinations of VMir Score and WC cut-off values**

| Cut-offs used | WC Cut-off | VMir Score Cut-off | Estimated % Sensitivity | Predicted CyHV-3 pre-miRNAs post-filtering |
|---|---|---|---|---|
| Mean WC and VMir Scores | 40 | 186 | 49.4 | 49 |
| Mean Min WC and Mean VMir Score | 21 | 186 | 53.2 | 142 |
| Mean WC and Mean Min.VMir Score | 40 | 139 | 77.9 | 192 |
| Mean WC and Lowest VMir Score in data | 40 | 112 | 83.1 | 303 |
| Mean Min. WC and VMir Scores | 21 | 139 | 81.8 | 563 |

Applying any of these combinations of cut-off values to the 43,921 predicted CyHV-3 pre-miRNAs had the effect of eliminating most of them (Table 3.3). In general, the higher the sensitivity, the less predictions were eliminated, suggesting a possible reduction in accuracy as sensitivity increased. Despite this, the cut-off values displaying the best sensitivity in this case did *not* also show the least accuracy. Using the cut-off values WC: 40, VMir Score: 112 resulted in the highest sensitivity (83.1%, as mentioned earlier) and eliminated all but 303 pre-miRNA predictions. By contrast, the next most sensitive cut-offs of WC: 21, VMir Score: 139 actually left 563 predictions. This suggested that using cut-off values of WC: 40, VMir Score: 112 resulted in more accurate predictions than using WC: 21, VMir Score: 139 despite also being more sensitive. As a result, these were the cut-off values (WC: 40, VMir Score: 112) that were ultimately used to eliminate low probability predictions on the CyHV-3 genome. The effect of applying these cut-off values to filter out low probability pre-miRNA predictions in the CyHV-3 genome is displayed in Figure 3.3. The effect of applying the same cut-off values to the VMir predictions from the six other viral genomes and comparison to CyHV-3 in this regard are shown in Table 3.4.

**Table 3.4 Summary of VMir pre-miRNA predictions passing established cut-offs values**

| Filter Applied : Min Score: 112, Min MHP size: 45, Max MHP size 200, Min WC: 40 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Viral Genome** | **EBV** | **HSV** | **HCMV** | **KSHV** | **MDV-1** | **MGHV-68** | **Mean** | **CyHV-3** |
| Predicted pre-miRNAs after application of cut-off values | 173 | 195 | 252 | 111 | 143 | 106 | **163.33** | **303** |
| % Reduction in Predicted MHPs due to filter | 99.36 | 99.17 | 99.27 | 99.42 | 99.40 | 99.36 | **99.33** | **99.31** |
| Filtered MHP sites per Kb post filter | 0.94 | 1.28 | 1.07 | 0.81 | 0.80 | 0.89 | **0.96** | **1.03** |
| Real miRNAs post filter | 25/25 | 2/6 | 10/11 | 10/12 | 10/14 | 7/9 | | |

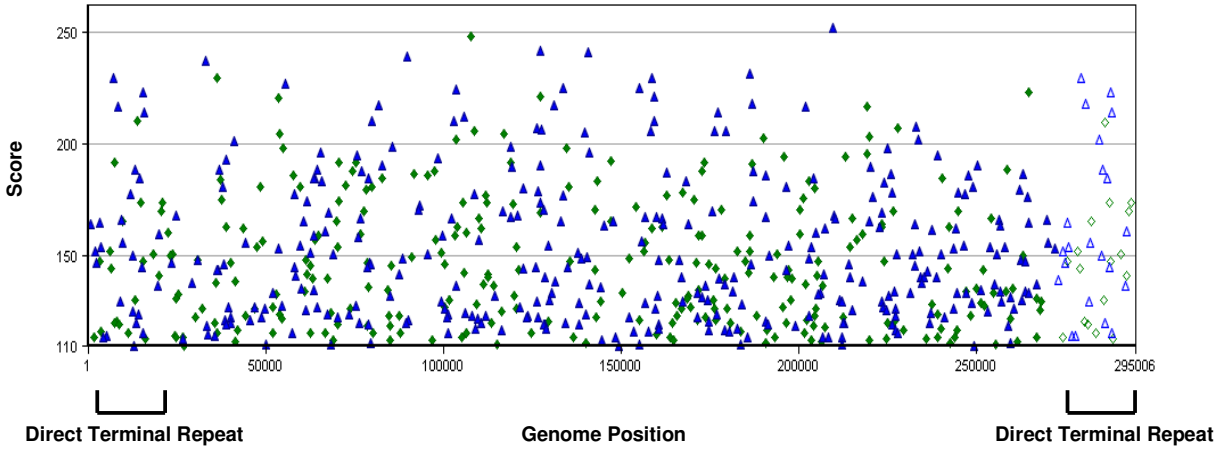



**Figure 3.3 VMir pre-miRNA predictions on the CyHV-3 genome after cut-off values**
MHP positions in the forward and reverse orientations are represented by solid blue triangles and green diamonds respectively. Repeated MHPs are represented by open shapes. As expected hairpin structures in the left terminal repeat also occur in the right terminal repeat, hence hairpins in the right terminal repeat are presented here as open shapes.

## 3.4 Further assessment of pre-miRNA predictions:

Any predictions occurring in ORFs were eliminated. This resulted in the elimination of 104 predictions. The remaining 199 were structurally analysed using the pre-miRNA-classifier MiPred. This involved analysis of both MHPs and SHPs as shorter versions hairpins of the same hairpin may fold slightly differently or may be more stable in the local sequence context (i.e. may have different values for WC). The inclusion of SHPs resulted in the analysis of 597 more pre-miRNA structures (i.e. length variations of MHPs). Out of a total of 796 predicted pre-miRNAs analysed, 488 were classified as real pre-miRNAs. In total there were 155 hairpin sites where either the MHP or at least 1 SHP was classified as a real pre-miRNA. These results are summarized below (Table 3.5).

**Table 3.5 Summary of MiPred analysis on VMir predicted CyHV-3 pre-miRNAs**

| Classification | No. of Hairpins | Percentage |
|---|---|---|
| Not a pre-miRNA-like | 138 | 17% |
| Pseudo pre-miRNA | 170 | 22% |
| Real pre-miRNA | 488 | 61% |
| 155 sites (77%) contain at least 1 "Real" pre-miRNA | | |

# 4 Experimental evaluation of pre-miRNA predictions

In order to experimentally evaluate VMir and MiPred CyHV-3 pre-miRNA predictions (Section 3), RNA from an *in vitro* CyHV-3 lytic infection was analysed using a variety of methods. CyHV-3 putative miRNAs/pre-miRNAs were initially identified by deep sequencing of small RNAs from CyHV-3-infected cells followed by identification of highly abundant small RNAs mapping to VMir predicted pre-miRNAs. In order to verify the RNA size range distribution of these small RNAs and to rule out the possibility of over-representation in the deep sequencing data (due to potential enzymatic bias) the same RNA was re-tested using array hybridization, targeting miRNAs from a range of high and low probability miRNA candidates. High probability candidates were also later analysed by northern blotting in order to check if they gave bands of sizes that were indicative of genuine miRNAs.

## 4.1 Deep sequencing of small RNAs from *in vitro* CyHV-3 infections

In order to identify small RNAs mapping to predicted pre-miRNAs, high throughput Illumina cDNA sequencing was carried out on 17-25 nt RNA from two separate *in vitro* infections of CCB cells using two different CyHV-3 isolates, H361 and N076. This approach allowed the detection of putative miRNAs derived from stems of predicted pre-miRNAs. Furthermore, sequencing from two different infections allowed the assessment of consistency in small RNA profiles between different flasks infected with different isolates. This strategy used to identify putative miRNAs from pre-miRNA candidates is outlined below in Figure 4.1

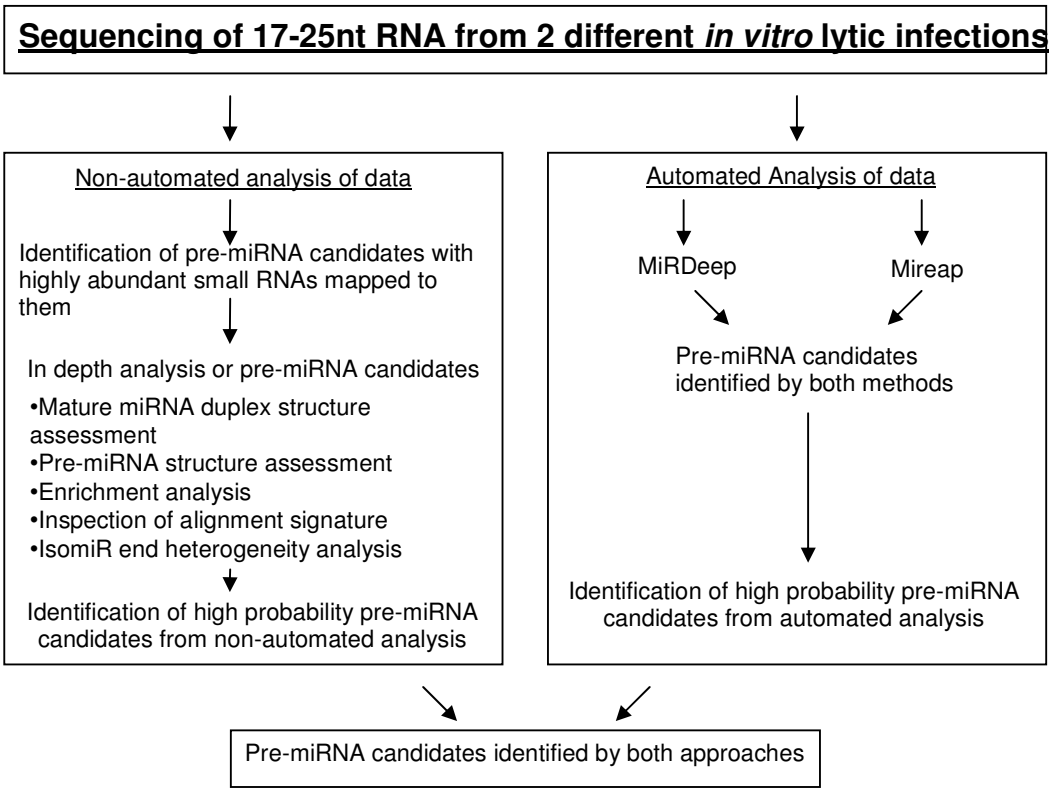



**Figure 4.1 Outline of workflow for identification of high probability CyHV-3 pre-miRNAs through analysis of small RNA deep sequencing data**

## 4.1.1 Deep sequencing library validation

### 4.1.1.1 *Quality check of amplified library*

It was important that no non-specific products were present in the deep sequencing library. This was verified by agarose gel electrophoresis. Based on the adapter and primer sequences using in this study, a successfully prepared sequencing library (consisting of amplified dsDNA versions of 17-25 nt ssRNA now flanked by full-length adaptor sequences) should consist of only dsDNA ranging from 104-112 bp in length. The vast majority of these products should be in the 107-109 bp size range on the basis that the 17-25 nt samples consisted predominantly of miRNAs which are 20-22 nt in length. This should be visible on gels as a single strong band with no additional (non-specific) products.

In the initial optimization of the in-house deep sequencing library preparation process (using a modified version of the Illumina Small RNA v1.5 protocol described in Section 2.7), 150 pmol of each adaptor was used without PAGE purification of 5' ligation products before the RT step (similar to the Illumina Small RNA v1.5 protocol). While this generated the correct product of 107-109 bp in size, it also generated much more abundant shorter non-specific products of ~90 bp in length. This smaller non-specific product was likely to be amplified adaptor-dimers which should be 87 bp in length (Figure 4.2 (a) Lib 1). Repeating this process and including PAGE purification of 5' ligation products before the RT step reduced the amount of adaptor-dimers but did not eliminate them (Figure 4.2 (a) Lib 2). It became apparent that excess pre-adenylated 3' adaptors from the 3' ligation step were still present after PAGE purification and were entering the 5' adapter ligation reactions to form adapter-dimers. It is likely that at this stage, adapter-dimers form much easier than the intended 5' ligation product due to the pre-adenylation of the 3' adaptor. These shorter products may also be much more efficiently amplified during PCR due to their small size and enzymatic bias. This may have been the cause of the result shown in Figure 4.2 (a) Lib 1. PAGE purification of 5' ligation products before the RT step went some way towards reducing the amount of adapter-dimers entering the RT step, although some of them still persisted (Figure 4.2 (a) Lib 2.). It was found that this problem could easily be eliminated by repeating the process used to prepare Lib 2 and using less adaptor (10 pmol compared to 150 pmol). This resulted in the elimination of adaptor-dimers and a much stronger band representing the correct library amplification product. This method was used to successfully prepare cDNA libraries from size-selected small RNA from both the H361 and N076 infections. These amplified libraries (displaying a single band at ~110 bp) are displayed in Figure 4.2 (a) and (b).

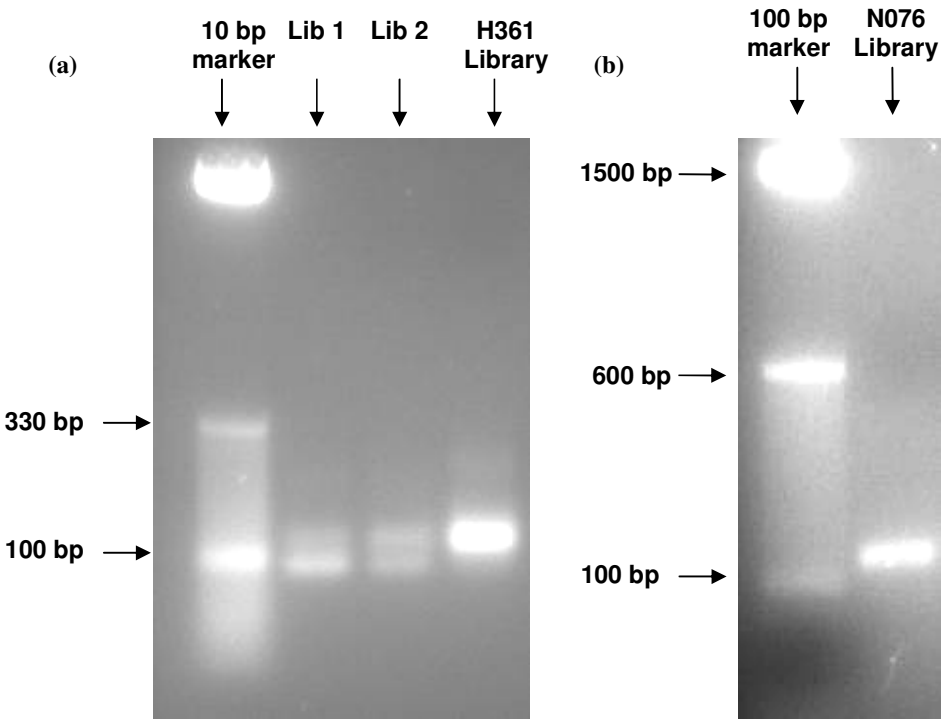
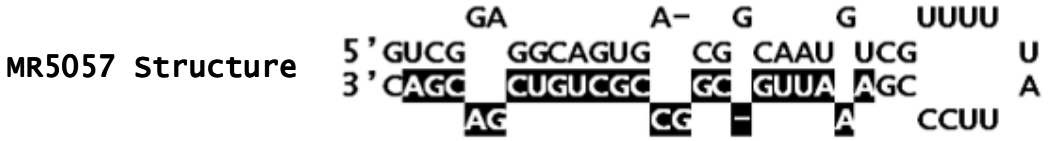



**Figure 4.2 Validation of Deep-sequencing libraries by gel electrophoresis.**
(a) Lib 1 and 2 show additional smaller non-specific-products ~90 bp. H361 library shows the correct 104-112 bp product (b) N076 library shows the correct 104-112 bp product

#### 4.1.1.2 *Assessment of content of amplified libraries*

In order to verify that the libraries were enriched for miRNAs over any other type of small RNA, PCR product was cloned into pGEM-T Easy vectors and used to transform highly competent *E. coli* JM109 cells. Dye terminator sequencing was then used to analyse inserts in these vectors. In total, 40 clones were sequenced. This analysis confirmed the correct assembly of the Illumina-compatible flanking sequences. Most of the sequences matched evolutionarily conserved eukaryotic miRNAs i.e. novel conserved carp cell miRNAs. This confirmed that the library was specifically enriched for miRNAs and suitable for deep sequencing analysis. In addition, one putative CyHV-3 miRNA was identified from dye terminator sequencing. This was identified on the basis that it mapped to the stem of a predicted CyHV-3 pre-miRNA (MR5057). This putative miRNA and its predicted pre-miRNA are shown in Figure 4.3.

(a)




(b)

miRNA sequence 5’ AAAUUGCGGCCGCUGUCGACGA 3’

(c)

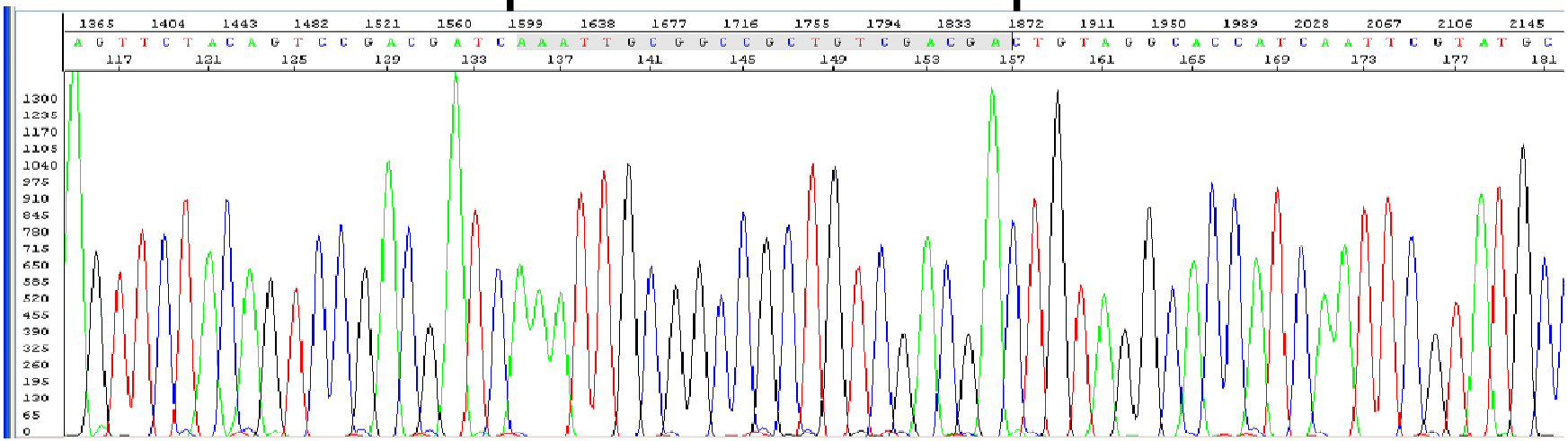


**Figure 4.3 Identification of a putative CyHV-3 miRNA through dye terminator sequencing of an insert (from amplified cDNA library) cloned into pGEM-T Easy Vector**

(a) Structure of predicted pre-miRNA MR5057 (b) Sequence of mature miRNA from 3’ arm of MR5057 (c) Electropherogram with putative miRNA sequence highlighted.

### 4.1.2 Mapping small RNA deep sequencing data to the CyHV-3 genome

Quality filtered and processed reads from deep sequencing were mapped to the CyHV-3 genome. This revealed that the vast majority of transcripts sequenced from both infections were from the host and that only a small percentage were of viral origin (see Table 4.1 and Figure 4.4). In total there were 212,399 and 1,601,110 CyHV-3 transcripts sequenced from the H361 and N076 infections respectively.

**Table 4.1 Overall deep sequencing yields from H361 and N076 infections.**

| Infection | Total Yield (Read Count) | CyHV-3 Total Read Count | CyHV-3 Total Read Count as a % of total yield |
|---|---|---|---|
| H361 | 12,016,247 | 212,399 | 1.77% |
| N076 | 23,903,173 | 1,601,110 | 6.70% |

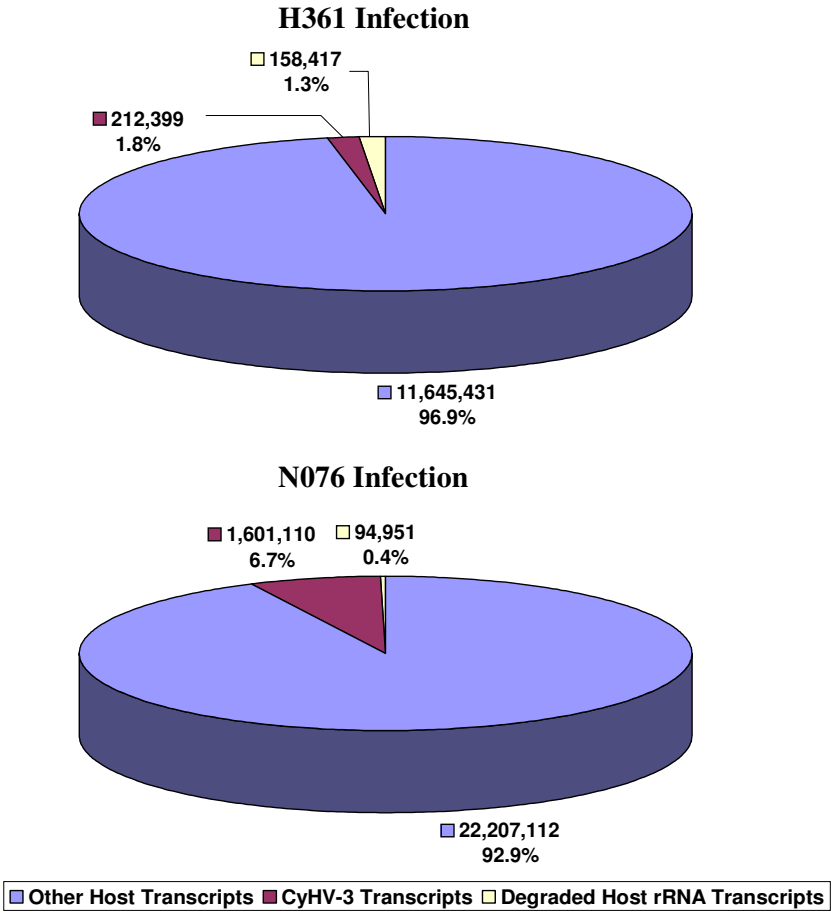



**Figure 4.4 A breakdown of the origin of reads (viral/host) from both infections**
This shows the percentage of transcripts sequenced that were of viral and host origin and specifically those of host rRNA origin. The latter give an indication as to the extent of overall RNA degradation.

Coverage plots of the mapped sequencing data are displawed in Figure 4.5. It can be seen that the entire genome was well-represented and that overall there was good agreement between the two infections both in terms of qualitative and quantitative data. In general however, the read counts for individual transcripts from the N076 infection were significantly higher than the corresponding read counts for the same transcripts from the H361 infection. It can also be seen from the quantitative data (Figure 4.5) that there are discrete parts of the genome that have many more reads mapping to them than other parts. However, in both infections the vast majority of transcripts were detected only once i.e. they had a read count of 1, thus the median read count from both infections was equal to 1. A breakdown of unique reads and frequency ranges from both infections is also shown (Table 1.6 (a) and (b)). Details of all individual reads mapping to the CyHV-3 genome including start and end positions, length and read counts are displayed in Supplementary File 4.1 T.1-T.4.

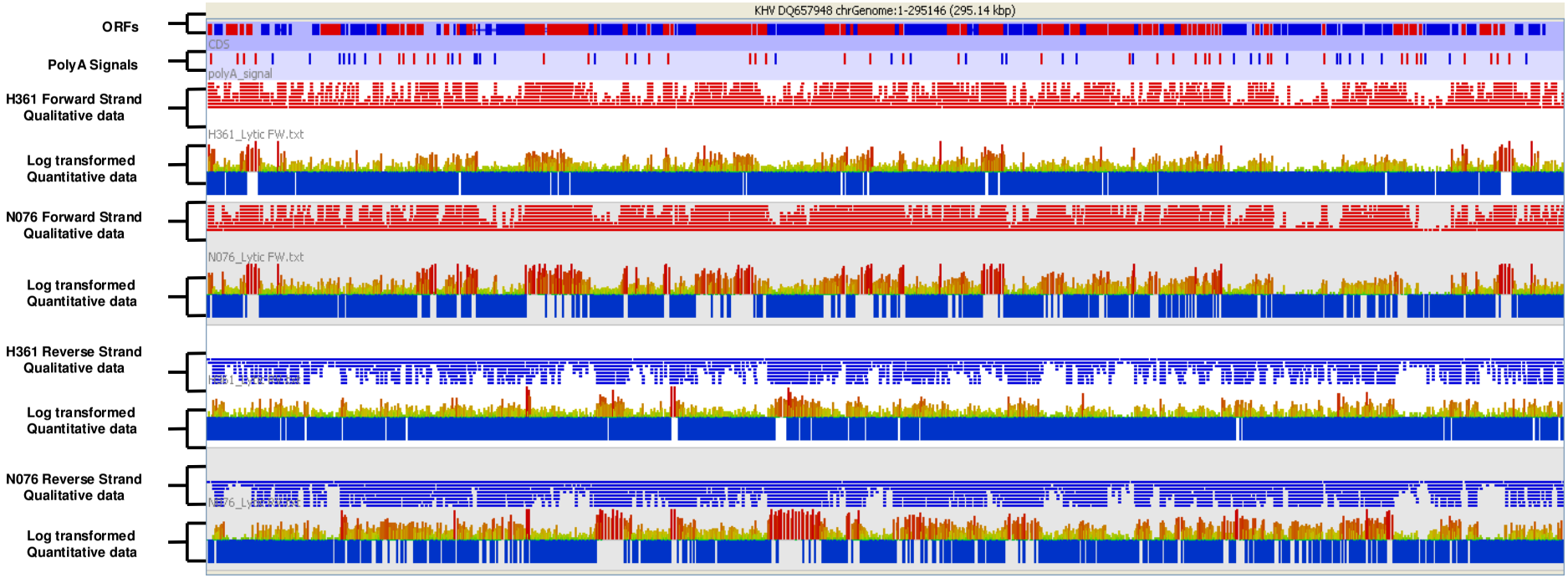



**Figure 4.5 Coverage plots of sequence data mapped to the CyHV-3 genome**
Reads mapping to the forward and reverse strands are shown on separate data tracks for both infections. Reads mapping to forward strands are in red (always above axis) and reads mapping to reverse strands are in blue (always below axis). The log transformed quantitative information (i.e. read counts) is displayed directly under each respective qualitative data track. The highest read counts are in red and the lowest are in blue and are below the axis. ORFs on the forward strand are in red and ORFs on the reverse strand are in blue. Their associated PolyA signals are indicated using the same colour scheme.

**Table 4.2 A breakdown of unique small RNA reads sequenced in terms of frequency ranges**
Numbers of unique transcripts sequenced and their frequencies from (a) the H361 infection and (b) the N076 infection

**(a)**

| H361 infection | | | | | |
|---|---|---|---|---|---|
| **Total Unique-Transcripts** | | | **CyHV-3 Unique-Transcripts** | | |
| **Read count range** | **Number of unique-transcripts** | **Percentage of total unique-transcripts** | **Read count range** | **Number of CyHV-3 unique-transcripts** | **Percentage of total unique-transcripts** |
| >999 | 563 | 0.16% | >999 | 9 | 0.003% |
| 100-999 | 2424 | 0.71% | 100-999 | 41 | 0.012% |
| 10-99 | 15033 | 4.38% | 10-99 | 581 | 0.169% |
| 5-9 | 14411 | 4.20% | 5-9 | 1032 | 0.301% |
| 2-4 | 55944 | 16.30% | 2-4 | 6331 | 1.845% |
| 1 | 254799 | 74.25% | 1 | 22223 | 6.476% |
| **Total number of unique-transcripts** | 343174 | 100.00% | **Total number of CyHV-3 unique-transcripts** | 30217 | 8.81% |

**(b)**

| N076 infection | | | | | |
|---|---|---|---|---|---|
| **Total Unique-Transcripts** | | | **CyHV-3 Unique-Transcripts** | | |
| **Read count range** | **Number of unique transcripts** | **Percentage of total unique transcripts** | **Read count range** | **Number of CyHV-3 unique transcripts** | **Percentage of total unique transcripts** |
| >999 | 499 | 0.15% | >999 | 19 | 0.007% |
| 100-999 | 1847 | 0.54% | 100-999 | 218 | 0.083% |
| 10-99 | 11811 | 3.44% | 10-99 | 3598 | 1.374% |
| 5-9 | 12038 | 3.51% | 5-9 | 5033 | 1.923% |
| 2-4 | 44595 | 12.99% | 2-4 | 4172 | 1.594% |
| 1 | 190998 | 55.66% | 1 | 59539 | 22.743% |
| **Total number of unique transcripts** | 261788 | 100.00% | **Total number of CyHV-3 unique transcripts** | 72579 | 27.724% |

The 17-25 nt RNA samples were specifically enriched for miRNAs and therefore miRNAs should be among the most abundant transcripts present in the sample. Transcripts with read counts ≥10 were considered significantly more abundant than the median read count of 1 and as such, miRNAs were expected to be among this subset of highly-expressed transcripts. Such highly-expressed transcripts only represented ~2% and ~5% of unique CyHV-3 transcripts identified in the H361 and

N076 infections respectively (Table 4.2 (a) and (b) and Table 4.3). However the combined read counts of these highly-expressed transcripts accounted for the vast majority of transcripts sequenced from both infections, representing 79.29% and 93.25% of all transcripts sequenced in the H361 and N076 infections respectively (Table 4.3). Unique transcripts were further divided into those mapping to coding-regions and non-coding regions of the genome. These details are relevant in this study because in addition to being of high abundance in these samples, miRNAs are expected to be mainly derived from non-coding regions of the genome. Figure 4.6 shows that highly expressed transcripts mapping to non-coding regions only represented a small subset of unique transcripts, ~1% in both infections. However, Figure 4.7 shows that the combined read counts of these highly abundant transcripts actually constituted the vast majority of CyHV-3 transcripts sequenced in both infections. This indicated that these non-protein-coding regions of the CyHV-3 genome were quite transcriptionally active and that this may have been due to the expression of miRNAs. In 17-25 nt size fractionated RNA samples, miRNAs should be among a small subset of highly expressed CyHV-3 transcripts from non-coding regions. Therefore for the purposes of novel miRNA discovery, transcripts from this subset were inspected further.

**Table 4.3 Details of highly-expressed unique CyHV-3 transcripts sequenced.**

| Infection | Unique CyHV-3 transcripts | Highly expressed unique CyHV-3 transcripts (read counts ≧10) | Percentage of unique CyHV-3 transcripts that are highly expressed transcripts | Combined read count of unique CyHV-3 high-highly transcripts | Percentage of total CyHV-3 read count |
|---|---|---|---|---|---|
| H361 | 30,217 | 631 | 2.09% | 168,421 | 79.29% |
| N076 | 72,579 | 3,835 | 5.28% | 1,492,985 | 93.25% |

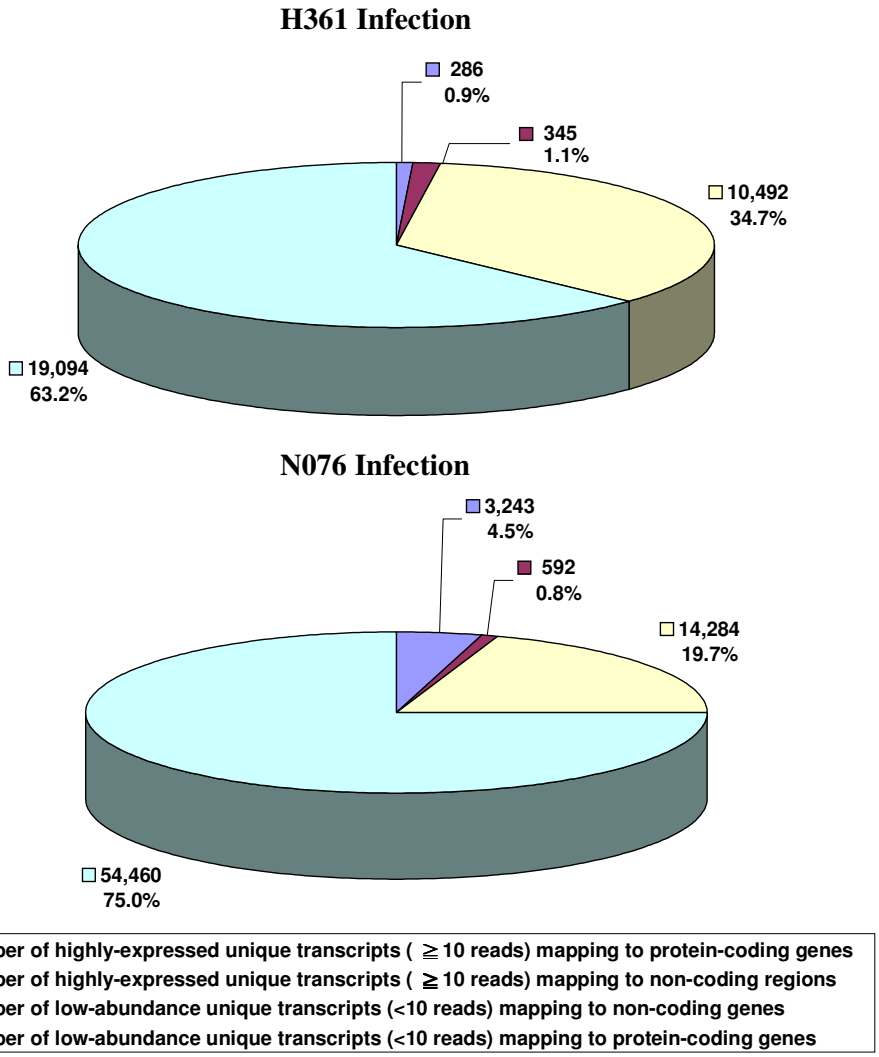



**Figure 4.6 A breakdown of the number of highly expressed /low-abundance unique CyHV-3 transcripts sequenced in terms of their genomic regions.**

High-abundance unique transcripts (≧10 reads) mapping to both protein-coding and non-coding regions represented a small subset of unique CyHV-3 transcripts. Most unique transcripts were present low abundance (<10 reads) and were mapped to protein-coding regions.

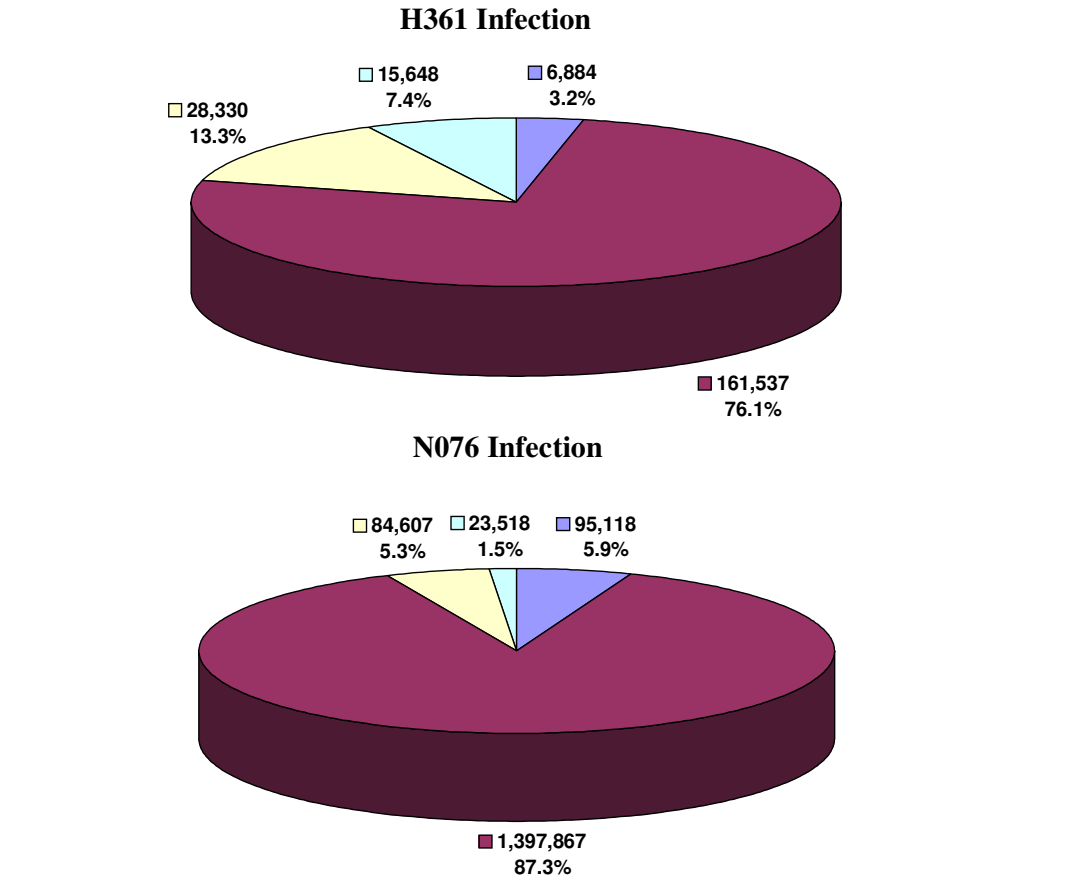



**Figure 4.7 A breakdown of the combined read counts of highly-expressed /low-abundance unique CyHV-3 transcripts in terms of their genomic regions.**
High-abundance unique transcripts (≧10 reads) mapping to non-coding regions account for the vast majority of CyHV-3 transcripts sequenced in both samples.

### 4.1.3 Non-automated identification of miRNAs from deep sequencing data

#### 4.1.3.1 *Initial identification of pre-miRNA candidates*

By mapping CyHV-3 transcripts to CyHV-3 protein-coding sequences and predicted pre-miRNA sequences it was possible to identify highly-expressed transcripts (read counts ≥10) mapping to predicted pre-miRNAs occurring in non-coding parts of the CyHV-3 genome. These provisional pre-miRNA candidates (i.e. VMir predicted pre-CyHV-3 miRNAs with sequenced small RNAs mapping to them) were checked for the presence of putative miRNAs mapping to both arms and adjacent to the terminal loop (only one of which needed to be highly-expressed i.e. a read count ≥10). In order that a provisional pre-miRNA candidate be considered for further analysis, it was essential that the same pattern of small RNA mapping was observed in both infections. This at least suggested that some of these patterns were caused by a precise repeated pattern of miRNA processing from pre-miRNAs as opposed to the random degradation of larger non-coding transcripts. This strategy resulted in the identification of 21 pre-miRNA candidates (no longer considered provisional candidates) for further analysis. Nineteen of these occurred in 4 discrete clusters (defined as containing pre-miRNA candidates within 10 kb of each other). Details of these pre-miRNA candidates are given in Table 4.4 and their genomic positions can be seen in Figure 4.8.

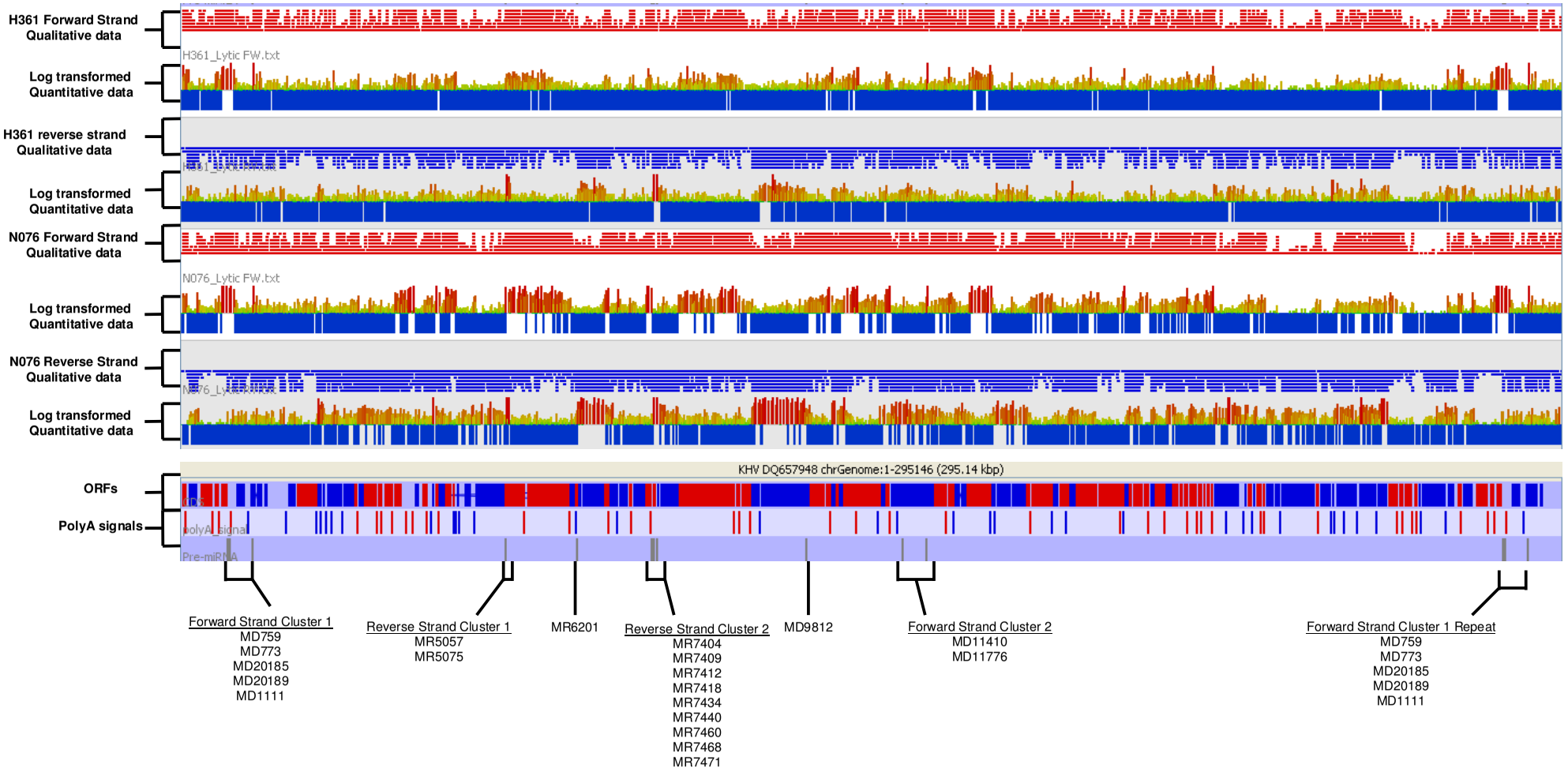



**Figure 4.8 Locations of pre-miRNA candidates on the CyHV-3 genome and their comparison to small RNA sequencing data.**
ORFs on the forward strand are in red and ORFs on the reverse strand are in blue. Their associated PolyA signals are coloured in the same way

**Table 4.4 (Part 1/3) List of Pre-miRNA candidates identified on the forward strand of the CyHV-3 genome (GenBank Ac No. DQ6579481.1)**
The co-ordinates of the pre-miRNA candidates are based on MHPs predicted by VMir

| Forward Strand | | | | | | |
|---|---|---|---|---|---|---|
| **Clustered/ non-clustered** | **Pre-miRNA candidate name** | **Start** | **End** | **VMir Score** | **WC Relative** | **Pre-miRNA candidate sequence 5' to 3'** |
| **Forward Strand Cluster 1** | MD759 | 10265 | 10325 | 93.9 | 28 | GGUGUGGGGGCCCAGCAAAGCC ACGUGCUAGAGGCUGAGGGCAG GCUGAGGUCCCUGAUUC |
| | MD773 | 10531 | 10574 | 119.7 | 1 | CGCUUGGGGCUGCCGUCUGCUG UUCCCGAUGGGGCUCUGGUGCG |
| | MD20185 | 10754 | 10812 | 62.9 | 15 | GGUUACUCGUAAUGGUUACAUA UAUUACAUAUAUUACAUAUCAUA CUAUACGGUUUACU |
| | MD20189 | 11006 | 11054 | 33.4 | 1 | UGGUCUUUGGAAGAAACGAAUA AACAAGCGAUUUAACUUUGGGG GUCCG |
| | MD1111 | 15432 | 15603 | 222.9 | 42 | UCAGCAGUCUCUUGUGGGUCUC UGUCUGCAGUGGACAUCACCUC CUCCUCACCAGCAUCAGCAGCA GGCUCAUCAUCAUCAUCAGAGU CAGAGAUGAGGAUGGUCUCAGU CUGUUGUGGUGGUGGUGUGGA CUUGGCUGCACGGCGCUGCUCA CGGCUUGCAAGACGGCUGA |
| **Forward Strand non-clustered** | MD9812 | 133905 | 134011 | 224.8 | 46 | GGCACAGCGUGCGGGCCCUUUC CGGCAGGCCGUCGGCCUUCAUG AUGUAGACGCCGGACUCGUCGC GAAGGUCGACGGCGCCCGGCGA GCGCGGCCUCACUCCGGCC |
| **Forward Strand Cluster 2** | MD11410 | 154441 | 154515 | 141.8 | 37 | GCCCACCGAGGACGCGUAGGCC GUGUCCACCUCGUUGAUCUUGG AGAUGUCCACGGUGUACCCGGC CUCGGCGGC |
| | MD11776 | 159561 | 159696 | 210.2 | 45 | GCCCAGCUGAUCCUUGGACUCU GGCACGCCGUUGAUACCGGUCA GCACCCUCCCGUAGAACUUGUC CACAAAGUUGCAGCAGGCCUGC GUGAGGGUGUUGGCCGGCACGU GCUUGUUGGGCGUCCACACGCC GGGC |

**Table 4.4 (Part 2/3) List of Pre-miRNA candidates identified on the forward strand of the CyHV-3 genome (GenBank Ac No. DQ6579481.1)**
The co-ordinates of the pre-miRNA candidates are based on MHPs predicted by VMir

| Reverse Strand | | | | | | |
|---|---|---|---|---|---|---|
| **Clustered/ non-clustered** | **Pre-miRNA candidate name** | **Start** | **End** | **VMir Score** | **WC Relative** | **Pre-miRNA candidate sequence 5' to 3'** |
| **Reverse Strand Cluster 1** | MR5057 | 69618 | 69677 | 114.2 | 2 | GUCGGAGGCAGUGACGGCAAUG UCGUUUUUAUUCCCGAAAUUGC GGCCGCUGUCGACGAC |
| | MR5075 | 69875 | 70015 | 165.4 | 47 | GGUUCUUGCUUGGUUGACUGCG AGGAAGGCAGCACGGGGGUGGG CGUCUUUGUAGUUGGUACAACU GCCUGCUUCACGGUGCUAGUAU CAACUGCAAGGGUGCCCGUUCC ACUCUUCUCCGCUGGAUCGACC UUGACGGCC |
| **Reverse Strand non-clustered** | MR6201 | 84889 | 84939 | 123 | 45 | GACGGCGCUGUGCCAGAAGGGC UUUUCCAGCGCUUCGAGCGCGU CGUCGUC |
| **Reverse Strand Cluster 2** | MR7404 | 100908 | 100958 | 76.3 | 9 | ACGUUGUGGGGGACGGUUCAGA UCACCAACGAACCUCCGAUCACA UUUUGU |
| | MR7409 | 100949 | 101022 | 115.4 | 11 | CCGGCGACUCGUGUGGGACGUG AUGAGCGGCGAGCCCCGCGACG GCUCGUCCACACACAUCCUCAC GUUGUGGG |
| | MR7412 | 101011 | 101055 | 76.5 | 1 | ACGGCUCGCCUGGGCUCAGAAG CGGGACUAGGACCGGCGACUCG U |
| | MR7418 | 101072 | 101119 | 86.9 | 18 | CAACCGGCCGCGGACCCUGGUC CCUCAGACCGGUAGACGUCGCC GUUG |
| | MR7434 | 101223 | 101300 | 172.5 | 8 | AGGUUGCUCGCCGUCGCCGUCU CGGCUUGGAGGUACCUGCGGGC ACGUCGAGACGGACGGCUCGGG CCAGCGCUGCCU |

**Table 4.4 (Part 3/3) List of Pre-miRNA candidates identified on the forward strand of the CyHV-3 genome (GenBank Ac No. DQ6579481.1)**
The co-ordinates of the pre-miRNA candidates are based on MHPs predicted by VMir

| Reverse Strand | | | | | | |
|---|---|---|---|---|---|---|
| **Clustered/non-clustered** | **Pre-miRNA Name** | **Start** | **End** | **VMir Score** | **WC Relative** | **Pre-miRNA candidate sequence 5' to 3'** |
| **Reverse Strand Cluster 2 (Continued)** | MR7440 | 101364 | 101508 | 115.9 | 12 | UCUCACGCGUGACGGAUGGGU GUGCGACGGUCGCUGCGACUA CUCGCCAACCUCCUCACGUCA GA |
| | MR7460 | 101591 | 101658 | 144.5 | 17 | GGGUCAGAGAGACGGCUCACU CGGUAGGGAGCGUCCGACGUC GUGCGGGCCACCGUCGCGGG GACAUGGGUUCGGGUCCCGCC GCCGAAAGCGAGCUGGAUCCA UCCCGGUCUUCACCUCGGCCA AAUGGACUCCGUCUCUCGAUU GUCGUGACGACUCUGCUACUC GGGUGCACCGAAAACCUCUAU GGCCC |
| | MR7468 | 101668 | 101718 | 33 | 1 | GGCACAGCGUGCGGGCCCUUU CCGGCAGGCCGUCGGCCUUCA UGAUGUAGACGCCGGACUCGU CGCGAAGGUCGACGGCGCCCG GCGAGCGCGGCCUCACUCCGG CC |
| | MR7471 | 101717 | 101781 | 130.1 | 42 | GCCCACCGAGGACGCGUAGGC CGUGUCCACCUCGUUGAUCUU GGAGAUGUCCACGGUGUACCC GGCCUCGGCGGC |
| | MR7476 | 101817 | 102009 | 130.7 | 37 | GCCCAGCUGAUCCUUGGACUC UGGCACGCCGUUGAUACCGGU CAGCACCCUCCCGUAGAACUU GUCCACAAAGUUGCAGCAGGC CUGCGUGAGGGUGUUGGCCG GCACGUGCUUGUUGGGCGUC CACACGCCGGGC |

Details of putative miRNAs which were the basis for identification of these pre-miRNA candidates are given in Table 4.5. It can be seen from the read counts in Table 4.5 that some of these are extremely abundant, for example the transcripts mapping to pre-miRNA candidates MR5057, MD11776, MD773 and MD1111. In particular the putative miRNA from the 3' arm of MR5057 was very highly expressed. It was the most abundant CyHV-3 transcript in both infections accounting for 29% and 72% of all CyHV-3 transcripts in the H361 and N076 infections respectively. Overall, it was the 35th and 4th most abundant transcript sequenced in the H361 and N076 infections respectively, making it even more abundant that 99.9%

of host transcripts sequenced in both infections. In addition to miRNAs, 6 pre-miRNAs were also found to have additional transcripts occurring immediately adjacent to a putative miRNA. These are a new class of pre-miRNA derivatives called microRNA-offset-RNAs (or moRNAs). Some of these were very highly expressed. For example the most abundant putative moRNA detected was the 3' putative moRNA from MD11776 in the N076 infection (3,171 reads). In both infections, the 5' putative moRNA from MR5075 was more abundant than the putative miRNA on the same arm. Conversely, the putative moRNAs from MD9812 only occurred once in both infections. Details of read counts for all putative miRNAs and moRNAs are given in Table 4.5 and their positions on their respective candidate pre-miRNA secondary structures are shown in Figure 4.10. Overall it can be seen from Table 4.5 that read counts were generally higher from the N076 infection when compared to the H361 infection. These levels are also directly compared in Figure 1.12. It was also observed that the dominant isomiRs at some miRNA loci were slightly different from one infection to the next. Out of the 42 putative miRNAs, 21 (50%) are represented by slightly different isomiRs (putative isomiRs at this stage) in each infection. IsomiRs are variants or miRNAs that are slightly different in length and occur due to inconsistencies in miRNA processing (Section 1.5.1). Despite this, the majority of putative miRNAs (32 of them, or 73.8%) still retained the same 5' end in both infections. Of those that did not share the same 5' ends, 4 of them (9.5%) were offset by just 1 nt and 7 of them (16.7%) were offset by >1nt at their 5'ends. In addition, for one pre-miRNA candidate, MD1111, the major form switched from the miRNA on the 3' arm in the H361 infection to the miRNA on the 5' arm in the N076 infection.

**Table 4.5 (Part 1/3) List of putative miRNAs and moRNAs from pre-miRNA candidates and their read counts**

| Pre-miRNA candidate name | Arm | Infection | Sequence | Same isomiR is miRNA in both infections? | IsomiR present in both infections | Read Count |
|---|---|---|---|---|---|---|
| MD759 | 5' miRNA | H361 | GUGGGGGCCCAGCAAAGCCACGU | No | No | 1 |
| | | N076 | GGGGGCCCAGCAAAGCCAC | | No | 3 |
| | 3' miRNA | H361 | GAGGGCAGGCUGAGGUCCCUGAUU | No | Yes | 15 |
| | | N076 | AGGCUGAGGUCCCUGAUU | | No | 70 |
| MD773 | 5' miRNA | H361 | AGGGGGCCGUCGCUUGGGGCUGCC | No | No | 1 |
| | | N076 | CGCUUGGGGCUGCCGUCU | | No | 7 |
| | 3' miRNA | H361 | AUGGGGCUCUGGUGCGCCACGAGU | Yes | Yes | 53006 |
| | | N076 | AUGGGGCUCUGGUGCGCCACGAGU | | Yes | 7314 |
| MD20185 | 5' miRNA | H361 | AUGGUUACUCGUAAUGGUUA | No. Same 5' End | Yes | 36 |
| | | N076 | AUGGUUACUCGUAAUGGUU | | Yes | 453 |
| | 3' miRNA | H361 | UACUAUACGGUUUACUUGUUUA | No. 5' end off-set +/- 1nt | No | 2 |
| | | N076 | ACUAUACGGUUUACUUGUUU | | No | 5 |
| MD20189 | 5' miRNA | H361 | AAUGGUCUUUGGAAGAAAC | No. Same 5' End | Yes | 107 |
| | | N076 | AAUGGUCUUUGGAAGAAACGAAU | | Yes | 950 |
| | 3' miRNA | H361 | AACUUUGGGGGUCCGGU | No. Same 5' End | Yes | 2 |
| | | N076 | AACUUUGGGGGUCCGG | | Yes | 9 |
| MD1111 | 5' miRNA | H361 | AGCAGCAGGCUCAUCAUCAUCA | Yes | Yes | 567 |
| | | N076 | AGCAGCAGGCUCAUCAUCAUCA | | Yes | 1055 |
| | 3' miRNA | H361 | AGGAUGGUCUCAGUCUGUUGUGGU | Yes | Yes | 2711 |
| | | N076 | AGGAUGGUCUCAGUCUGUUGUGGU | | Yes | 72 |
| | 5' moRNA | H361 | AUCACCUCCUCCUCACCAGC | No. 5' end off-set +/- 1nt | No | 3 |
| | | N076 | UCACCUCCUCCUCACCAGC | | No | 4 |
| | 3' moRNA | H361 | UGGUGGUGUGGACUUGGC | No. Same 5' End | Yes | 38 |
| | | N076 | UGGUGGUGUGGACUUGGCU | | Yes | 85 |
| MD9812 | 5' miRNA | H361 | AGGCCGUCGGCCUUCAUGAUG | No. Same 5' End | Yes | 23 |
| | | N076 | AGGCCGUCGGCCUUCAUGAUGU | | Yes | 7 |
| | 3' miRNA | H361 | GUCGCGAAGGUCGACGGCGCC | No. Same 5' End | Yes | 1 |
| | | N076 | GUCGCGAAGGUCGACGGCGCC | | Yes | 5 |
| | 5' moRNA | H361 | CGUGCGGGCCCUUUCCGGC | Yes | Yes | 1 |
| | | N076 | CGUGCGGGCCCUUUCCGGC | | Yes | 1 |
| | 3' moRNA | H361 | N/A | N/A | N/A | N/A |
| | | N076 | CGGCGAGCGCGGCCUCACUCC | | N/A | 1 |
| MD11410 | 5' miRNA | H361 | ACGCGUAGGCCGUGUCCACCUC | Yes | Yes | 10 |
| | | N076 | ACGCGUAGGCCGUGUCCACCUC | | Yes | 26 |
| | 3' miRNA | H361 | GAUGUCCACGGUGUACCCGGC | No. 5' end off-set +/- 1nt | No | 4 |
| | | N076 | AUGUCCACGGUGUACCCGGCC | | No | 2 |

**Table 4.5 (Part 2/3) List of putative miRNAs and moRNAs from pre-miRNA candidates and their read counts**

| Pre-miRNA candidate name | Arm | Infection | Sequence | Same isomiR is miRNA in both infections? | IsomiR present in both infections | Read Count |
|---|---|---|---|---|---|---|
| MD11776 | 5' miRNA | H361 | CAGCACCCUCCCGUAGAACUUG | Yes | Yes | 301 |
| | | N076 | CAGCACCCUCCCGUAGAACUUG | | Yes | 1126 |
| | 3' miRNA | H361 | AGGCCUGCGUGAGGGUGUUGGC | No. Same 5' End | Yes | 2114 |
| | | N076 | AGGCCUGCGUGAGGGUGUUGG | | Yes | 10396 |
| | 5' moRNA | H361 | ACUCUGGCACGCCGUUGAUACCGGU | No | Yes | 18 |
| | | N076 | UCUGGCACGCCGUUGAUACCGGU | | Yes | 85 |
| | 3' moRNA | H361 | CGGCACGUGCUUGUUGGGCG | Yes | Yes | 60 |
| | | N076 | CGGCACGUGCUUGUUGGGCG | | Yes | 3171 |
| MR5057 | 5' miRNA | H361 | GGAGGCAGUGACGGCAAUGUCGUU | Yes | Yes | 15 |
| | | N076 | GGAGGCAGUGACGGCAAUGUCGUU | | Yes | 149 |
| | 3' miRNA | H361 | AAAUUGCGGCCGCUGUCGACGA | Yes | Yes | 62596 |
| | | N076 | AAAUUGCGGCCGCUGUCGACGA | | Yes | 1166058 |
| MR5075 | 5' miRNA | H361 | UUGUAGUUGGUACAACUGCCUGC | Yes | Yes | 12 |
| | | N076 | UUGUAGUUGGUACAACUGCCUGC | | Yes | 95 |
| | 3' miRNA | H361 | ACGGUGCUAGUAUCAACUGCAAGG | Yes | Yes | 125 |
| | | N076 | ACGGUGCUAGUAUCAACUGCAAGG | | Yes | 217 |
| | 5' moRNA | H361 | AGCACGGGGGUGGGCGUCU | Yes | Yes | 26 |
| | | N076 | AGCACGGGGGUGGGCGUCU | | Yes | 122 |
| | 3' moRNA | H361 | N/A | N/A | N/A | N/A |
| | | N076 | N/A | | N/A | N/A |
| MR6201 | 5' miRNA | H361 | UGACGGCGCUGUGCCAGAAGGG | No. 5' end off-set +/- 1nt | Yes | 2 |
| | | N076 | AUGACGGCGCUGUGCCAGAAGGG | | Yes | 6 |
| | 3' miRNA | H361 | AGCGCUUCGAGCGCGUCGUCGUCUU | Yes | Yes | 6 |
| | | N076 | AGCGCUUCGAGCGCGUCGUCGUCUU | | Yes | 22 |
| MR7404 | 5' miRNA | H361 | UGUGGGGGACGGUUCAGAUC | No. Same 5' End | Yes | 91 |
| | | N076 | UGUGGGGGACGGUUCAGAU | | Yes | 128 |
| | 3' miRNA | H361 | AACGAACCUCCGAUCACAUUU | Yes | Yes | 205 |
| | | N076 | AACGAACCUCCGAUCACAUUU | | Yes | 1033 |
| MR7409 | 5' miRNA | H361 | GGGACGUGAUGAGCGGCGAGC | Yes | Yes | 124 |
| | | N076 | GGGACGUGAUGAGCGGCGAGC | | Yes | 973 |
| | 3' miRNA | H361 | GACGGCUCGUCCACACACAUCCUC | No. Same 5' End | No | 2 |
| | | N076 | GACGGCUCGUCCACACACAUCCU | | No | 12 |
| MR7412 | 5' miRNA | H361 | AGACGGCUCGCCUGGGCUCAGAA | No. Same 5' End | No | 1 |
| | | N076 | AGACGGCUCGCCUGGGCUC | | Yes | 23 |
| | 3' miRNA | H361 | AGGACCGGCGACUCGUGU | Yes | Yes | 14 |
| | | N076 | AGGACCGGCGACUCGUGU | | Yes | 81 |

**Table 4.5 (Part 3/3) List of putative miRNAs and moRNAs from pre-miRNA candidates and their read counts**

| Pre-miRNA candidate name | Arm | Infection | Sequence | Same isomiR is miRNA in both infections | IsomiR present in both infections | Read Count |
|---|---|---|---|---|---|---|
| MR7418 | 5' miRNA | H361 | AACCGGCCGCGGACCCUGGUC | No. Same 5' End | No | 2 |
| | | N076 | AACCGGCCGCGGACCCUGGUCC | | Yes | 12 |
| | 3' miRNA | H361 | AGACCGGUAGACGUCGCCGUU | Yes | Yes | 66 |
| | | N076 | AGACCGGUAGACGUCGCCGUU | | Yes | 130 |
| MR7434 | 5' miRNA | H361 | GCUCGCCGUCGCCGUCUCGG | No | Yes | 2 |
| | | N076 | UUGCUCGCCGUCGCCGUCUCGGC | | Yes | 14 |
| | 3' miRNA | H361 | CGAGACGGACGGCUCGGGC | Yes | Yes | 56 |
| | | N076 | CGAGACGGACGGCUCGGGC | | Yes | 508 |
| MR7440 | 5' miRNA | H361 | AUCCCCGGCGGUCCAC | Yes | Yes | 13 |
| | | N076 | AUCCCCGGCGGUCCAC | | Yes | 154 |
| | 3' miRNA | H361 | GUUCCGCGGGACGCCUUCGU | Yes | Yes | 15 |
| | | N076 | GUUCCGCGGGACGCCUUCGU | | Yes | 67 |
| | 5' moRNA | H361 | AGGGUCACGGCGUUGGU | No. 5' end off-set +/- 1nt | Yes | 38 |
| | | N076 | CAGGGUCACGGCGUUGGU | | Yes | 58 |
| | 3' moRNA | H361 | CCGAUGAACGGUUGAUGAU | Yes | Yes | 280 |
| | | N076 | CCGAUGAACGGUUGAUGAU | | Yes | 70 |
| MR7460 | 5' miRNA | H361 | AUCCGGCGAGCGAUCUCUGCC | No. 5' end off-set +/- 1nt | Yes | 5 |
| | | N076 | UCCGGCGAGCGAUCUCUGCC | | Yes | 126 |
| | 3' miRNA | H361 | AGGGAUCGGGGCACGCGGAAU | Yes | Yes | 13 |
| | | N076 | AGGGAUCGGGGCACGCGGAAU | | Yes | 46 |
| MR7468 | 5' miRNA | H361 | ACUCACGGGGAGCCGAGUCUGA | No | No | 1 |
| | | N076 | UCACGGGGAGCCGAGUCU | | No | 86 |
| | 3' miRNA | H361 | AAGCGCUUCACGGGCUCCAGGGU | No | No | 3 |
| | | N076 | ACGGGCUCCAGGGUCUCGC | | Yes | 32 |
| MR7471 | 5' miRNA | H361 | ACGCGUGACGGAUGGGUGUGC | No | Yes | 20 |
| | | N076 | GACGGAUGGGUGUGCGACGGUCGCU | | Yes | 69 |
| | 3' miRNA | H361 | CGCCAACCUCCUCACGUCAGAC | Yes | Yes | 197 |
| | | N076 | CGCCAACCUCCUCACGUCAGAC | | Yes | 1152 |
| MR7476 | 5' miRNA | H361 | ACCGUCGCGGGGACAUGGGUU | Yes | Yes | 218 |
| | | N076 | ACCGUCGCGGGGACAUGGGUU | | Yes | 880 |
| | 3' miRNA | H361 | CCCGGUCUUCACCUCGGC | Yes | Yes | 61 |
| | | N076 | CCCGGUCUUCACCUCGGC | | Yes | 431 |
| | 5' moRNA | H361 | AGGGAGCGUCCGACGUCGUGC | Yes | Yes | 301 |
| | | N076 | AGGGAGCGUCCGACGUCGUGC | | Yes | 297 |
| | 3' moRNA | H361 | GAUUGUCGUGACGACUCUGC | Yes | Yes | 48 |
| | | N076 | GAUUGUCGUGACGACUCUGC | | Yes | 314 |

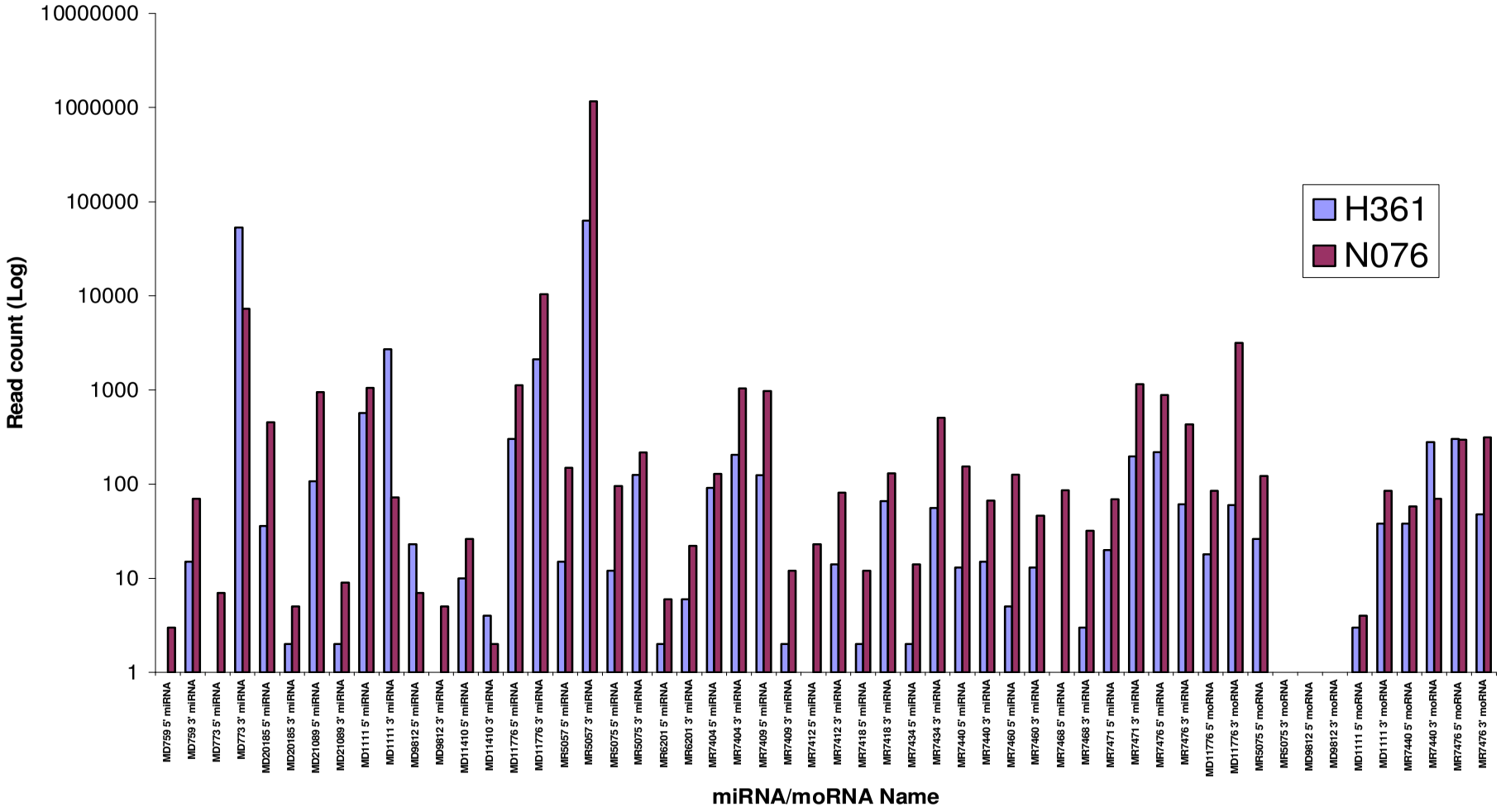



**Figure 4.9 Putative miRNA/moRNA read count comparison between the CyHV-3 H361 and N076 infections**

#### 4.1.3.2 *In depth analysis of pre-miRNA candidates*

Only pre-miRNA candidates were selected for more in depth assessment so as to determine the likelihood that these predictions represented genuine pre-miRNAs. This involved analysing the characteristics of both the pre-miRNAs themselves and all of the small RNAs mapping to their stems in order to ascertain if they conformed to the characteristics of *bone-fide* miRNAs.

#### 4.1.3.3 *Mature miRNA duplex structure assessment*

During miRNA biogenesis, the processing of miRNAs from pre-miRNAs typically results in the formation of mature-miRNA-duplexes with characteristic 3' overhangs. Mature miRNA duplexes were predicted by highlighting the putative miRNA sequence's pre-miRNA candidate secondary structures. These were inspected for the presence of 1-4 nt 3'end overhangs (or a maximum of one blunt end) in both infections. Out of the 21 pre-miRNA candidates, 9 of them (42.9%) were found to meet these criteria in both H361and N076 infections. The results of this analysis are shown in Figure 4.10. These structures can also be viewed in Supplementary File 4.1 T.5.

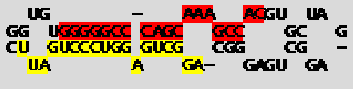
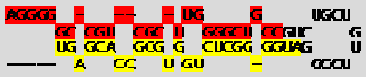
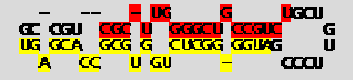
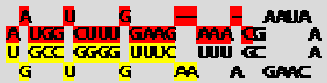
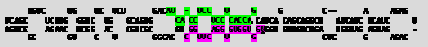
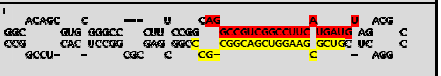
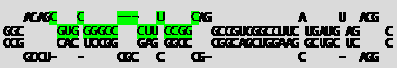
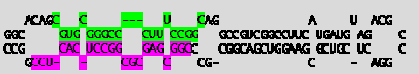
| Candidate Pre-miRNA name | Pre-miRNA candidate and mature -miRNA-duplex structure from H361 infection | Pre-miRNA candidate and mature -miRNA-duplex structure from N076 infection |
|---|---|---|
| MD759 |  1-4 nt 3' end overhang? 5' miRNA: Yes 3' miRNA: Yes |  1-4 nt 3' end overhang? 5' miRNA: No 3' miRNA: No |
| MD773 |  1-4 nt 3' end overhang? 5' miRNA: No 3' miRNA: No |  1-4 nt 3' end overhang? 5' miRNA: Yes 3' miRNA: No |
| MD20185 |  1-4 nt 3' end overhang? 5' miRNA: No 3' miRNA: No |  1-4 nt 3' end overhang? 5' miRNA: No 3' miRNA: No |
| MD20189 |  1-4 nt 3' end overhang? 5' miRNA: Yes 3' miRNA: Blunt |  1-4 nt 3' end overhang? 5' miRNA: Yes 3' miRNA: Blunt |
| MD1111 |  1-4 nt 3' end overhang? 5' miRNA: Yes 3' miRNA: Yes | |
| MD1111 moRNA |  1-4 nt 3' end overhang? 5' moRNA: No 3' moRNA: No |  1-4 nt 3' end overhang? 5' moRNA: No 3' moRNA: No |
| MD9812 |  1-4 nt 3' end overhang? 5' miRNA: Yes 3' miRNA: Yes |  1-4 nt 3' end overhang? 5' miRNA: Yes 3' miRNA: Yes |
| MD9812 moRNAs |  1-4 nt 3' end overhang? N/A |  1-4 nt 3' end overhang? 5' moRNA: Yes 3' moRNA: Yes |

**Figure 4.10 (Part 1/4) Structures of CyHV-3 pre-miRNA candidates and predicted mature miRNA-duplexes, based on putative miRNA sequences**
5' miRNAs are red, 3' miRNAs are yellow, 5' moRNAs in green and 3' moRNAs in purple.

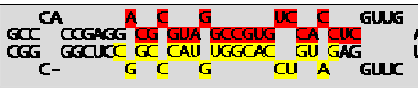
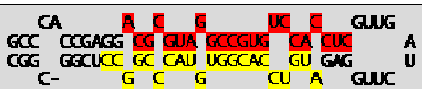
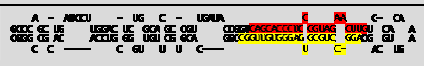
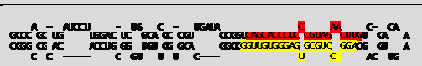
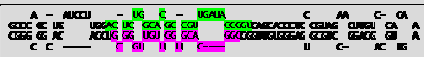
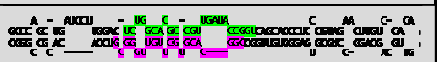
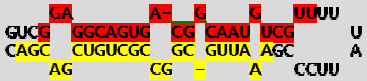
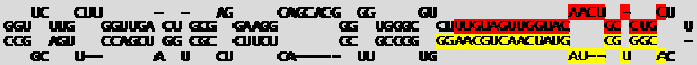
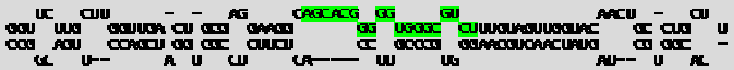
| Candidate Pre-miRNA name | Pre-miRNA candidate and mature -miRNA-duplex structure from H361 infection | Pre-miRNA candidate and mature -miRNA-duplex structure from N076 infection |
|---|---|---|
| MD11410 |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Yes<br>3' miRNA: Yes |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Yes<br>3' miRNA: Yes |
| MD11776 |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Yes<br>3' miRNA: Yes |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Yes<br>3' miRNA: Yes |
| MD11776 moRNAs |  <br>1-4 nt 3' end overhang?<br>5' moRNA: Yes<br>3' moRNA: No |  <br>1-4 nt 3' end overhang?<br>5' moRNA: Yes<br>3' moRNA: Yes |
| MR5057 |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Yes<br>3' miRNA: Yes | |
| MR5075 |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Blunt<br>3' miRNA: Yes | |
| MR5075 moRNAs |  <br>1-4 nt 3' end overhang?<br>N/A | |
| MR6201 |  <br>1-4 nt 3' end overhang?<br>5' miRNA: No<br>3' miRNA: Blunt |  <br>1-4 nt 3' end overhang?<br>5' miRNA: No<br>3' miRNA: Yes |

**Figure 4.10 (Part 2/4) Structures of CyHV-3 pre-miRNA candidates and predicted mature miRNA-duplexes, based on putative miRNA sequences**
MiRNAs are red, 3' miRNAs are yellow, 5' moRNAs in green and 3' moRNAs in purple.

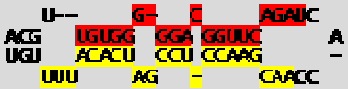
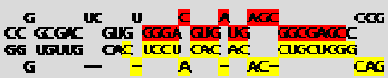
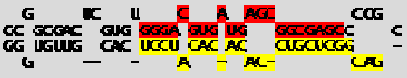
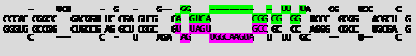
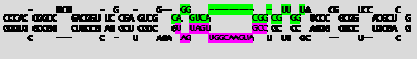
| Candidate Pre-miRNA name | Pre-miRNA candidate and mature -miRNA-duplex structure from H361 infection | Pre-miRNA candidate and mature -miRNA-duplex structure from N076 infection |
|---|---|---|
| MR7404 |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Yes<br>3' miRNA: Yes |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Yes<br>3' miRNA: Yes |
| MR7409 |  <br>1-4 nt 3' end overhang?<br>5' miRNA: No<br>3' miRNA: Yes |  <br>1-4 nt 3' end overhang?<br>5' miRNA: No<br>3' miRNA: Blunt |
| MR7412 |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Yes<br>3' miRNA: Blunt |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Blunt<br>3' miRNA: Blunt |
| MR7418 |  <br>1-4 nt 3' end overhang?<br>5' miRNA: No<br>3' miRNA: Blunt |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Blunt<br>3' miRNA: Blunt |
| MR7434 |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Yes<br>3' miRNA: No |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Yes<br>3' miRNA: No |
| MR7440 |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Blunt<br>3' miRNA: No | |
| MR7440 moRNAs |  <br>1-4 nt 3' end overhang?<br>5' moRNA: No<br>3' moRNA: Blunt |  <br>1-4 nt 3' end overhang?<br>5' moRNA: No<br>3' moRNA: No |

**Figure 4.10 (Part 3/4) Structures of CyHV-3 pre-miRNA candidates and predicted mature miRNA-duplexes, based on putative miRNA sequences**
5' miRNAs are red, 3' miRNAs are yellow, 5' moRNAs in green and 3' moRNAs in purple.

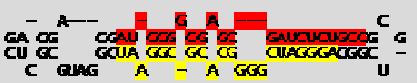
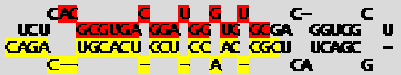
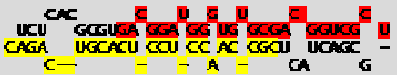
| Pre-miRNA | miRNA Duplex from H361 Infection | miRNA Duplex from N076 Infection |
|---|---|---|
| MR7460 |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Yes<br>3' miRNA: Blunt |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Yes<br>3' miRNA: Yes |
| MR7468 |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Blunt<br>3' miRNA: Blunt |  <br>1-4 nt 3' end overhang?<br>5' miRNA: No<br>3' miRNA: No |
| MR7471 |  <br>1-4 nt 3' end overhang?<br>5' miRNA: No<br>3' miRNA: No |  <br>1-4 nt 3' end overhang?<br>5' miRNA: No<br>3' miRNA: No |
| MR7476 |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Yes<br>3' miRNA: No | |
| MR7476 moRNAs |  <br>1-4 nt 3' end overhang?<br>5' miRNA: Yes<br>3' miRNA: No | |

**Figure 4.10 (Part 4/4) Structures of CyHV-3 pre-miRNA candidates and predicted mature miRNA-duplexes, based on putative miRNA sequences**
5' miRNAs are red, 3' miRNAs are yellow, 5' moRNAs in green and 3' moRNAs in purple.

#### 4.1.3.4 *Assessment of Pre-miRNA structures*

Two separate pre-miRNA classifiers, namely MiPred and the CSHMM-Method, were used in order to assess the structures of pre-miRNA candidates. These two classifiers performed well in performance comparison studies elsewhere (Agarwal et al., 2010; Sinha et al., 2009) and were therefore deemed to be most suitable for differentiating between pseudo-pre-miRNAs and genuine pre-miRNAs. Details of their performance and how they work are outlined in Section 1.7.1.2. In the case of each pre-miRNA candidate, it was the most stable version of the hairpin predicted (by VMir) to contain the full length miRNAs that was analyzed. This was based on the absolute-WC of the predicted MHP and SHPs (if any) from VMir analysis. WC is a measure of the stability of a specific hairpin structure within its local sequence context. Predicted length-variants with higher absolute WC values are more likely to form stable pre-miRNAs within the local sequence context than those with lower absolute WC values. These were therefore the most relevant variants to analyse for the purposes of pre-miRNA candidate classification. Where putative moRNAs were identified and where these sequences were not included in the most stable SHP (pre-miRNA variant) containing the miRNAs, the next most SHP also containing the moRNA sequence was also analysed. These longer variants were taken into account because previous studies indicated that moRNAs are included as part of the original pre-miRNA structure (Bortoluzzi et al., 2012, 2011).

In order for a pre-miRNA to be considered classified as a real pre-miRNA in this study, it needed to be classified as real by both methods. In addition, where longer variants were also used to account for moRNAs, both variants also needed to be classified as real pre-miRNAs by both methods. Fourteen out of the 21 pre-miRNA candidates were classified as real pre-miRNAs based on these criteria (Table 4.6). A more detailed description of the results including details of sequences used, MiPred Prediction Confidence and the CSHMM Prediction likelihood are available in Supplementary File 4.1. T.6

**Table 4.6 CyHV-3 pre-miRNA candidate structural analysis using pre-miRNA classifiers**
The pre-miRNA classifiers used were MiPred and the CSHMM-Method. Relative WC info was also included for comparative purposes

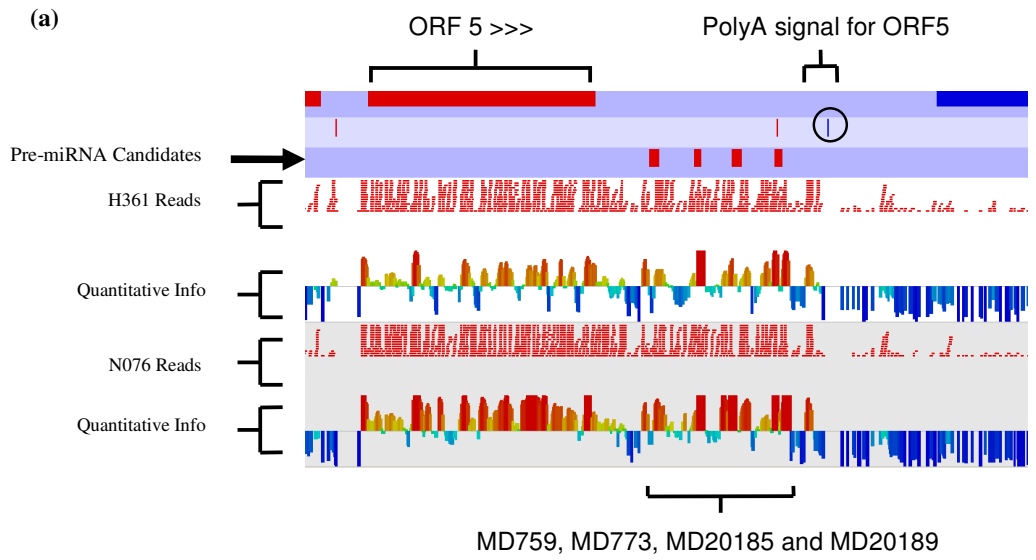
| Pre-miRNA Name | Absolute WC | Relative WC | MFE kcal/Mol | p-value (shuffle times:1000) | MiPred Result | CSHMM Result | Classified as Real Pre-miRNA by both methods |
|---|---|---|---|---|---|---|---|
| **MD759** | 17 | 28 | -28.1 | N/A | Not Real | Not Real | No |
| **MD773** | 1 | 1 | -17.8 | N/A | Not Real | Not Real | No |
| **MD20185** | 1 | 15 | -6.4 | N/A | Not Real | **Real** | No |
| **MD20189** | 1 | 1 | -7.6 | N/A | Not Real | Not Real | No |
| **MD1111** | 31 | 42 | -43.6 | 0.001 | **Real** | **Real** | **Yes** |
| **MD1111 (+ moRNAs)** | 5 | 42 | -45.4 | 0.002 | **Real** | **Real** | **Yes** |
| **MD9812** | 14 | 46 | -37.2 | 0.004 | **Real** | **Real** | **Yes** |
| **MD9812 (+ moRNAs)** | 1 | 46 | -56.2 | 0.044 | **Real** | **Real** | **Yes** |
| **MD11410** | 29 | 37 | -33.0 | 0.002 | **Real** | **Real** | **Yes** |
| **MD11776** | 17 | 45 | -40.4 | 0.002 | **Real** | **Real** | **Yes** |
| **MD11776 (+ moRNAs)** | 10 | 45 | -57.5 | 0.002 | **Real** | **Real** | **Yes** |
| **MR5057** | 2 | 2 | -26.9 | 0.004 | **Real** | **Real** | **Yes** |
| **MR5075** | 28 | 47 | -40.2 | 0.001 | **Real** | **Real** | **Yes** |
| **MR5075 (+ moRNAs)** | 14 | 47 | -58.7 | 0.001 | **Real** | **Real** | **Yes** |
| **MR6201** | 45 | 45 | -27.9 | 0.005 | **Real** | **Real** | **Yes** |
| **MR7404** | 5 | 9 | -27.9 | 0.012 | **Real** | Not Real | No |
| **MR7409** | 1 | 11 | -27.6 | 0.015 | **Real** | **Real** | **Yes** |
| **MR7412** | 1 | 1 | -15.3 | N/A | Not Real | Not Real | No |
| **MR7418** | 18 | 18 | -19.9 | N/A | Not Real | **Real** | No |
| **MR7434** | 9 | 8 | -38.7 | 0.106 | Not Real | **Real** | No |
| **MR7440 (+ moRNAs)** | 3 | 12 | -70.8 | N/A | Not Real | Not Real | No |
| **MR7460** | 16 | 17 | -29.1 | 0.014 | **Real** | **Real** | **Yes** |
| **MR7468** | 1 | 1 | -16.1 | N/A | Not Real | Not Real | No |
| **MR7471** | 36 | 42 | -25.1 | 0.013 | **Real** | **Real** | **Yes** |
| **MR7476 (+ moRNAs)** | 14 | 37 | -95.4 | N/A | Not Real | **Real** | No |

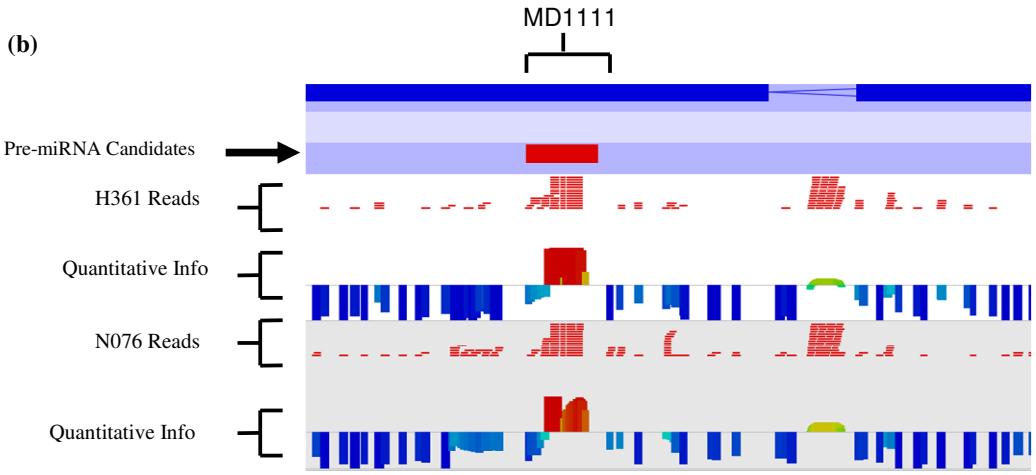
#### 4.1.3.5 *Enrichment quantification of putative CyHV-3 miRNAs*

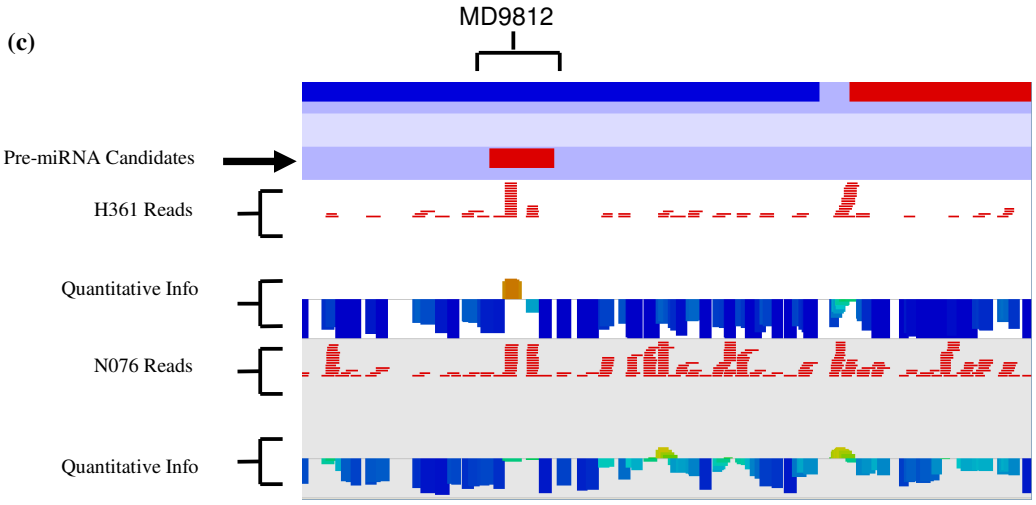
If the miRNA-enriched RNA sample that was sequenced from the CyHV-3 infections contained CyHV-3 miRNAs, then the mapping of this sequencing data to the CyHV-3 genome should reveal high concentrations of unique reads (stacks of miRNAs and isomiRs) that map to discrete pre-miRNA loci. Notably, this should be in stark contrast to local non-coding genomic regions surrounding these pre-miRNA loci. In addition, individual transcripts within these stacks should show much higher read counts when compared to transcripts that map to surrounding regions. In this study, candidate pre-miRNA loci were inspected for these kinds of characteristics using SeqMonk. This facilitated the visual analysis of transcripts that mapped to pre-miRNA candidate loci and surrounding loci, thus enabling a comparison to be made between the two infections. The use of the enrichment-quantification function in SeqMonk also facilitated the identification of loci that were specifically enriched in terms of read counts relative to surrounding genomic regions. To visualise this information, the genome file was first opened and then the mapping data from forward and reverse strands of the CyHV-3 genome were imported separately for each experiment so that they could be displayed on separate data tracks. The quantitative information displayed represented a log2 ratio of the observed base density in the region divided by the overall base density on the same data track.

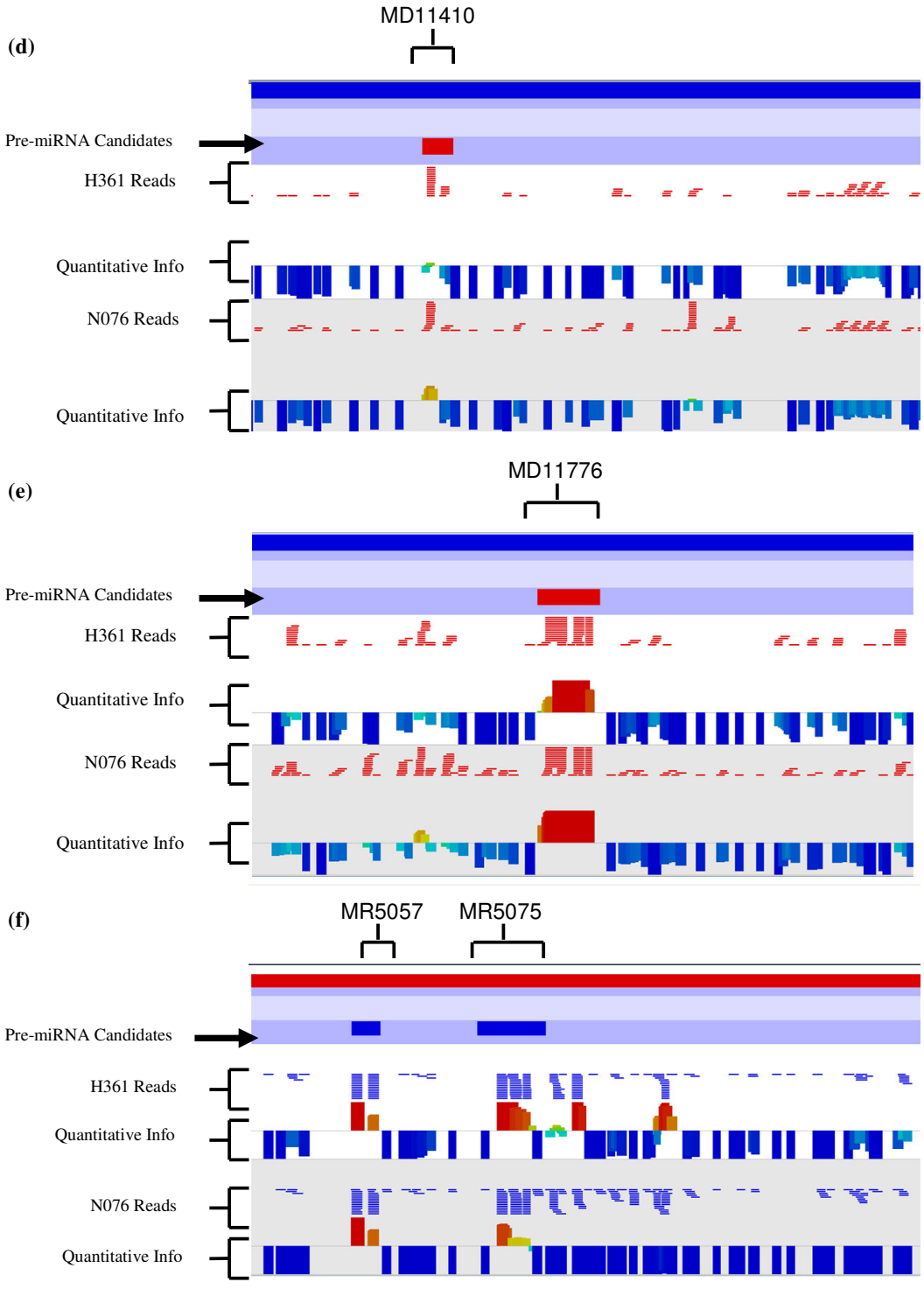
Out of the 21 pre-miRNA candidate loci, six of them were consistent with the characteristics outlined above: these were MD1111, MD9812, MD11410, MD11776, MR5057 and MR5075, (Figure 4.11 (b)-(f)). All of these are enriched when compared to their surrounding genomic regions, with the exception of MD9812 in the N076 infection. However, even though there is no enrichment in reads mapping to the MD9812 locus in the N076 infection it can be seen from the quantitative information that read counts for this locus are less depleted than that of the surrounding genomic region (Figure 4.11 (c)). Unlike these six pre-miRNA candidates, the remaining 15 pre-miRNA candidate loci (Figure 4.11(a), (g) and (h)) do not show individual discrete stacks of highly expressed transcripts. They are all parts of a continuous (in some cases overlapping) stretch of read stacks. None of the remaining fifteen pre-miRNA loci (Figure 4.11 (a), (g) and (h)) are further enriched in terms of read count relative to their surrounding genomic regions. This aspect is more apparent in Figure

4.12 where basic read count quantification (rather than enrichment quantification) is displayed for these five pre-miRNA candidate loci. Although putative miRNAs from MR6201 have comparable read counts to MD9812 and MD11410 (Figure 4.9 and Table 4.5), both of these are enriched (Figure 4.11 (c) and (d)) whereas MR6201 is not (Figure 4.11 (g)). This is because it is not as enriched relative to other transcripts that map to its surrounding genomic region. The ten pre-miRNA candidates in Reverse Cluster 2 (Figure 4.8) that are represented in Figure 4.11 (h) and Figure 4.12 (c) form a large continuous read stack which itself is highly enriched in terms of read count when compared to the surrounding genomic region. However the stack consists mainly of overlapping reads and there are no clear boundaries between each individual pre-miRNA candidate. In addition, regions of this stack that do not apparently contain pre-miRNA candidates are equally enriched in terms of read counts. Notably it is apparent from Figure 4.11 (a), (g), and (h) and Figure 4.12 that all fifteen pre-miRNA candidate loci appear to be located within three theoretical 3' UTRs. This is because they are all located between the stop codons and PolyA signals for upstream protein-coding genes. In contrast none of the other six pre-miRNA candidate loci that show the characteristics of pre-miRNAs occur within any apparent 3' UTR.


(a)
ORF 5 >>>
PolyA signal for ORF5
Pre-miRNA Candidates
H361 Reads
Quantitative Info
N076 Reads
Quantitative Info
MD759, MD773, MD20185 and MD20189


(b)
MD1111
Pre-miRNA Candidates
H361 Reads
Quantitative Info
N076 Reads
Quantitative Info


(c)
MD9812
Pre-miRNA Candidates
H361 Reads
Quantitative Info
N076 Reads
Quantitative Info


(d)
MD11410
Pre-miRNA Candidates
H361 Reads
Quantitative Info
N076 Reads
Quantitative Info
(e)
MD11776
Pre-miRNA Candidates
H361 Reads
Quantitative Info
N076 Reads
Quantitative Info
(f)
MR5057
MR5075
Pre-miRNA Candidates
H361 Reads
Quantitative Info
N076 Reads
Quantitative Info

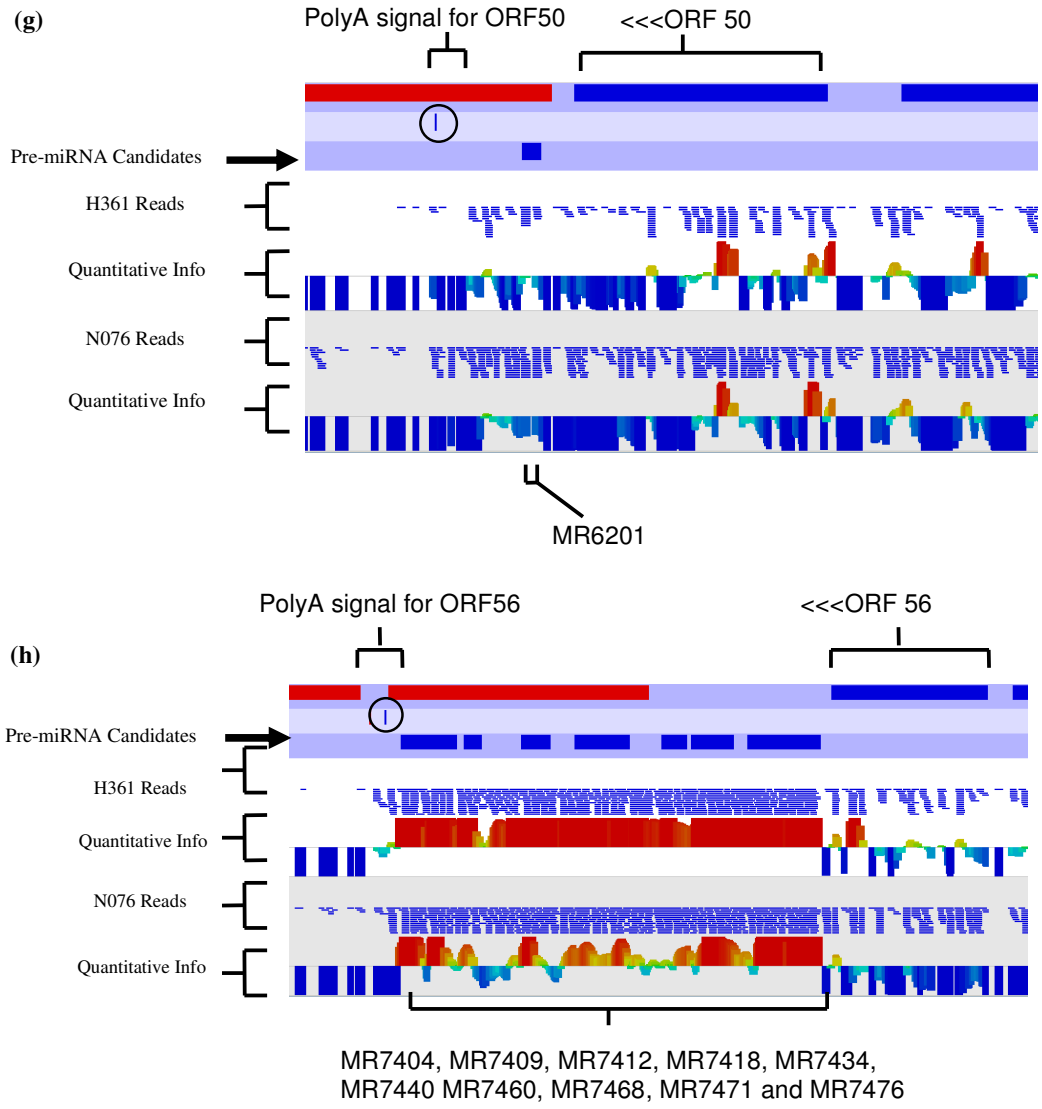



**Figure 4.11 Enrichment quantification of CyHV-3 putative miRNAs using SeqMonk**
Reads mapping to the forward strand are shown in red, those mapping to the reverse strand are in blue. Protein coding genes and PolyA signals on the forward strand are red; those on the reverse strand are blue and indicated using circles. Loci that are enriched in terms of read counts relative to the surrounding genomic region appear as positive values (i.e. above the axis) in the quantitative information tracks.

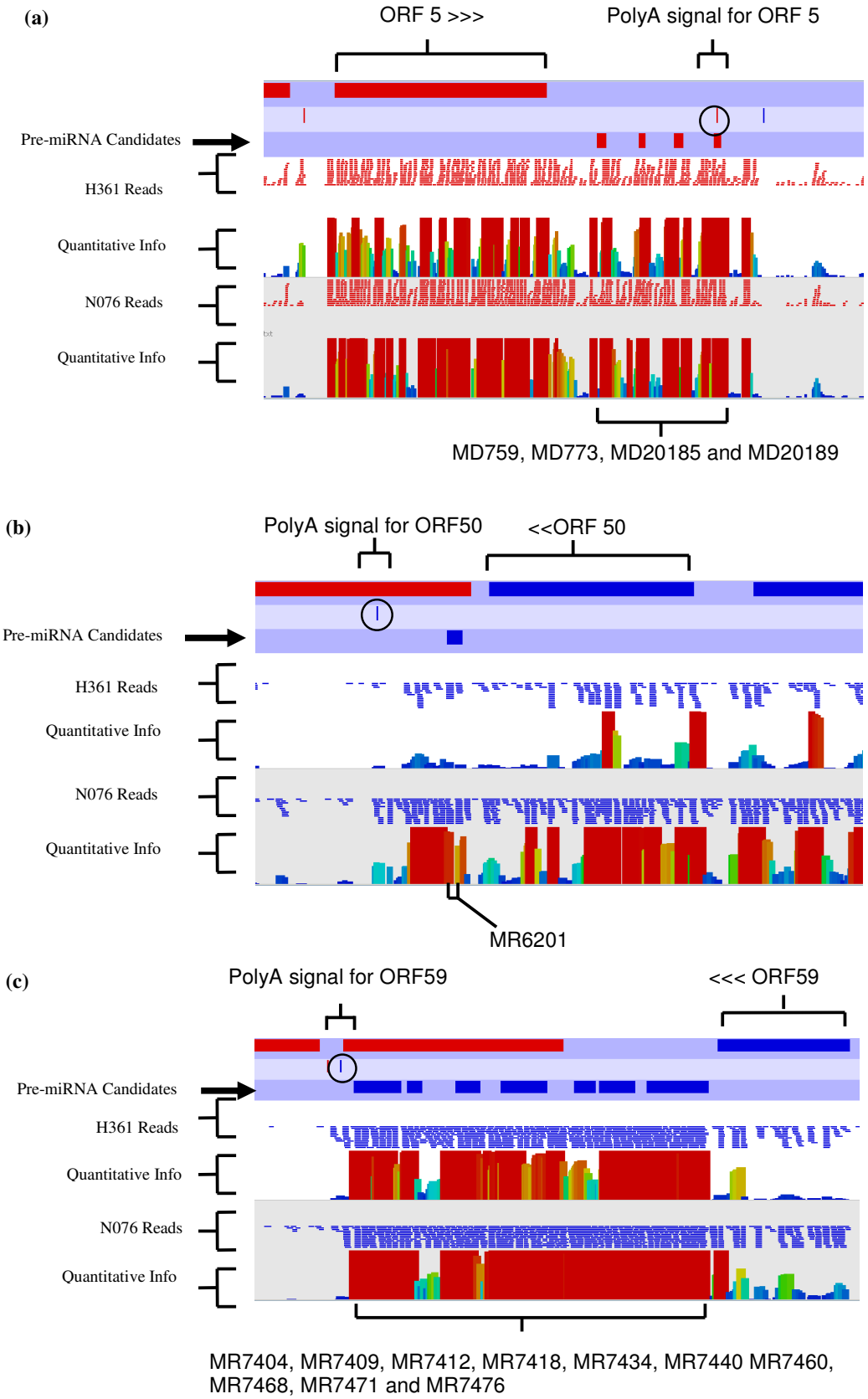



**Figure 4.12 Read count quantification (using SeqMonk) of pre-miRNA candidate loci showing non-miRNA-like patterns of read enrichment**

Reads mapping to the forward strand are in red, those mapping to the reverse strand are in blue. Protein coding genes and PolyA signals on the forward strand are red; those on the reverse strand are blue and indicated using circles.

#### 4.1.3.6 *Inspection of miRNA alignment signatures*

In order to establish if the small RNAs mapping to pre-miRNA candidates did so in a way that was consistent with the model of miRNA biogenesis, the alignment signature of each pre-miRNA candidate was assessed. The ideal miRNA alignment signature was outlined by Kozomara and Griffiths-Jones (2011) and is discussed in Section 1.7.2.2 (for an example see Figure 1.12).

In this study, alignment profiles that were either closely or loosely consistent with that of pre-miRNAs as described by Kozomara and Griffiths-Jones (2011) were identified as having a miRNA-like alignment signature. Seven of the 21 pre-miRNA candidates had alignment signatures that were considered miRNA-like (based on the deep sequencing data obtained from both infections). Signatures for the 7 miRNA-like and 14 non-miRNA-like alignments from the H361 and N076 infections are available as Supplementary File 4.2. Five of these pre-miRNA candidates (MR5057, MD9812, MD11410, MR6201 and MD11776) displayed strong miRNA-like alignment signatures. These all contained discrete stacks of reads representing either putative miRNAs or moRNAs. Notably, the alignment signature for MD11776 was particularly consistent with the ideal miRNA alignment signature. Transcripts were observed from five distinct regions of the MD11776 loci representing the 2 miRNAs, 2 moRNAs and low level terminal loop sequences. The signature observed for MD11776 resembles that of a "five-phased-precursor" recently described in several publications (Berezikov et al., 2011; Bortoluzzi et al., 2012, 2011) (see Figure 4.13).

The alignment profiles for MD1111 and MR5075 were slightly less consistent with the ideal miRNA alignment profile compared to other pre-miRNAs (MR5057, MD9812, MD11410, MR6201 and MD11776), although they certainly displayed the basic features of miRNA-like patterns. Importantly the signatures for MD1111 and MR5075 were very different from that of the remaining 14 pre-miRNAs. The alignment signatures for the other 14 pre-miRNA candidates consisted of overlapping reads offset across the entire pre-miRNA sequence. These signatures are not consistent with the model of miRNA biogenesis and are in stark contrast to the

miRNA-like alignment signatures displayed by MD1111, MD9812, MD11410, MD11776, MR5057, MR5075 and MR6201 (Supplementary File 4.2)

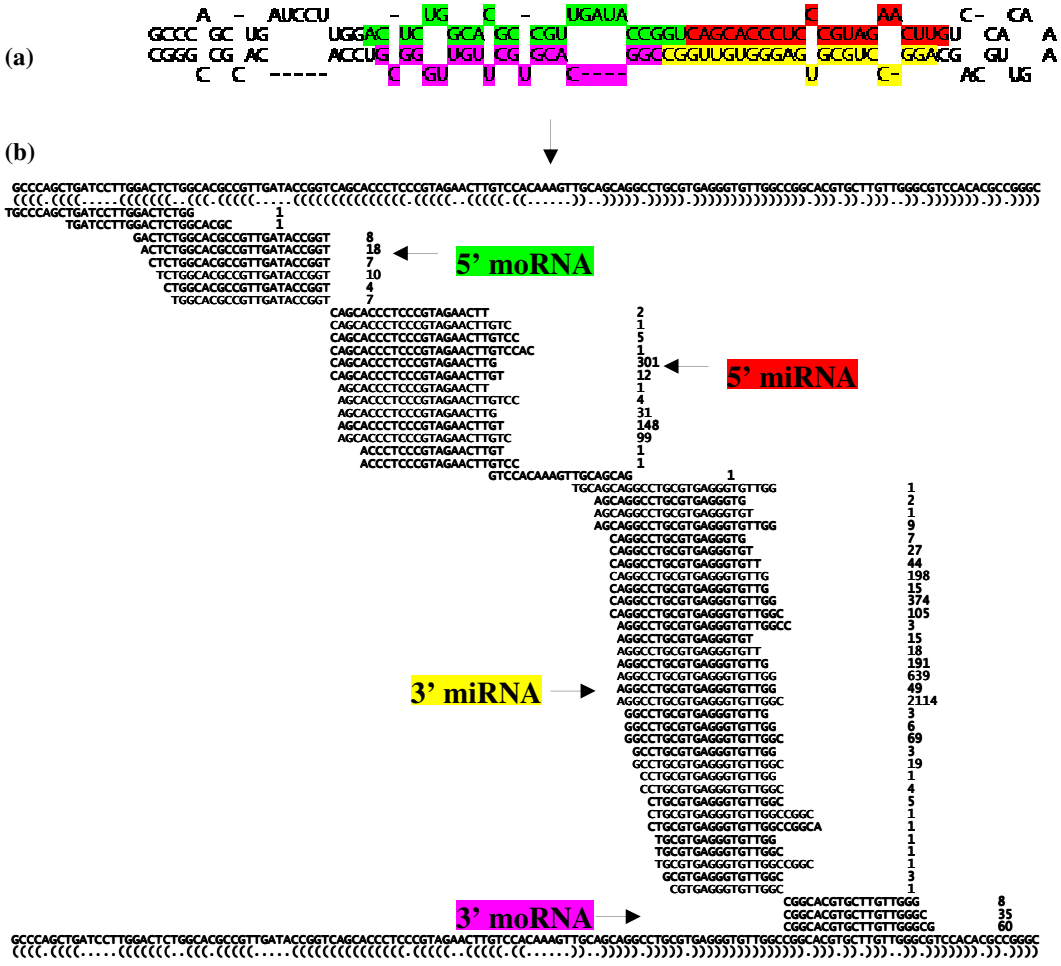



**Figure 4.13 "Five-phased precursor" miRNA-like alignment signature displayed by pre-miRNA candidate MD11776**

(a) Secondary structure of pre-miRNA candidate MD11776 with miRNA and moRNA reads highlighted on the stem (b) Alignment signature of small RNAs to pre-miRNA candidate MD11776. This is a good example of a site that displays all of the different types of distinct functional groups that are expected to be derived from a single pre-miRNA (miRNAs, moRNAs, terminal loop and associated isomers). Based on the read count information, it can be seen that it is the dominant read at each locus that is designated as the miRNA/moRNA with the rest designated as isomiRs/ moRNA isomers

#### 4.1.3.7 *IsomiR end heterogeneity analysis*

Due to the depth of coverage offered by deep sequencing it was possible to detect accompanying isomiRs for the putative miRNAs identified. These variants are all off-set by 1 or more nucleotides on either the 5' or 3' ends (or both) relative to the putative miRNAs. While heterogeneity can be observed at both ends, isomiRs generally tend to display more 3'end heterogeneity overall. The degree of end heterogeneity displayed by both 5' and 3' ends of isomiRs of putative miRNAs identified in this study were compared. Only putative miRNAs from pre-miRNA candidates showing miRNA-like alignment profiles were considered for this. IsomiRs from separated infections were analysed separately, resulting in 28 different sets of isomiRs being analyzed.

Most unique isomiRs from each putative miRNA were off-set at the 3' end i.e. most values (60.7%) for Ratio-1 were >1 (Table 4.7). Although 32.1% of putative miRNAs had similar numbers 5' offset isomiRs (i.e. values for Ratio-1 were equal to 1), 78% of these particular putative miRNAs and 89.3% of total putative miRNAs displayed a greater degree of 3' end heterogeneity among their isomiRs i.e. values for Ratio-2 were >1 (Table 4.7), meaning that offset values for 3' ends of unique isomiRs were generally higher.

The measure the overall degree of consistency 5' end processing consistency among these isomiRs it was also necessary to take read count into account using Ratio-3. This indicated that the isomiRs from 67.9.% of putative miRNAs showed more 5' end processing consistency and that 10.7% showed no significant difference to 3' end processing consistency (Table 4.7).

Overall, the isomiRs of the putative miRNAs from the pre-miRNA candidate MD1111 showed the least 5' end processing consistency (i.e. lowest values for Ratio-3). The same isomiRs were also the only ones to have values for Ratio-1 and Ratio-2 that were <1. However, isomiRs from the 5' arm miRNA of MD1111 did have good values for Ratio-2 in both infections and also showed more overall 5' end consistency among isomiRs in the H361 infection (Ratio-3 >1). The results of this analysis are summarized in Table 4.7 and displayed in Table 4.8 and the datasets used to calculate

these ratios (i.e. offset values for both ends of each individual isomiR) are available in Supplementary File 4.3 T.1-T.14.

**Table 4.7 Summary of isomiR end heterogeneity analysis**

| | Ratio-1 | Ratio-2 | Ratio-3 |
|---|---|---|---|
| **More 3' Arm Heterogeneity** | 60.7% | 89.3% | 67.9% |
| **More 5' Arm Heterogeneity** | 7.1% | 7.1% | 21.4% |

**Table 4.8 (Part 1/4) Analysis of isomiR end heterogeneity for putative miRNAs from CyHV-3 pre-miRNA candidates**
MiRNAs with 3' End: 5' End ratios that are <1 are highlighted in bold

| Pre-miRNA | miRNA | Infection | miRNA end | Transcripts ≧19nt representing ≧0.1% of combined miRNA and isomiR read count from respective strands | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Number of unique isomiRs | | Average absolute offset (nt) displayed by all unique isomiRs | | Read counts of IsomiRs as a % of the combined miRNA and isomiR read count from respective strands | |
| | | | | Number of unique isomiRs offset at each end | **Unique 3' End :5' End Off-set IsomiR Ratio (Ratio-1)** Should be >1 if there are more unique isomiRs displaying 3' end heterogeneity | Average offset (nt) at each end | **3' End :5' End Average Absolute Off-set Ratio (Ratio-2)** Should be >1 if there is a greater degree of 3' end heterogeneity overall | Read counts of isomiRs off-set at each end as a % of total read count | **3' End :5' End Processing Consistency Ratio (Ratio-3)** Should be >1 if 5' end processing is more consistent than at the 3' end processing |
| MD1111 | 5' Arm | H361 | 5' End | 4 | **0.75** | 0.80 | 1.25 | 1.36% | 1.88 |
| | | | 3' End | 3 | | 1.00 | | 2.54% | |
| | | N076 | 5' End | 2 | 1.50 | 1.00 | 1.50 | 1.58% | **0.12** |
| | | | 3' End | 3 | | 1.50 | | 0.19% | |
| | 3' Arm | H361 | 5' End | 6 | 1.00 | 1.80 | **0.61** | 2.96% | **0.85** |
| | | | 3' End | 6 | | 1.10 | | 2.50% | |
| | | N076 | 5' End | 3 | **0.33** | 1.60 | **0.13** | 27.27% | **0.37** |
| | | | 3' End | 1 | | 0.20 | | 10.00% | |
| MD9812 | 5' Arm | H361 | 5' End | 1 | 6.00 | 0.86 | 3.67 | 1.69% | 36.00 |
| | | | 3' End | 6 | | 3.14 | | 61.02% | |
| | | N076 | 5' End | 2 | 1.00 | 1.30 | 1.15 | 15.38% | 4.00 |
| | | | 3' End | 2 | | 1.50 | | 61.54% | |
| | 3' Arm | H361 | 5' End | 3 | 1.00 | 0.50 | 1.00 | 50.00% | 1.00 |
| | | | 3' End | 3 | | 0.50 | | 50.00% | |
| | | N076 | 5' End | 2 | 1.50 | 0.40 | 2.00 | 40.00% | 1.00 |
| | | | 3' End | 3 | | 0.80 | | 40.00% | |

**Table 4.8 (Part 2/4) Analysis of isomiR end heterogeneity for putative miRNAs from CyHV-3 pre-miRNA candidates**
MiRNAs with 3' End: 5' End ratios that are <1 are highlighted in bold

| Pre-miRNA | miRNA | Infection | miRNA end | Transcripts ≧19nt representing ≧0.1% of combined miRNA and isomiR read count from respective strands | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Number of unique isomiRs | | Average absolute offset (nt) displayed by all unique isomiRs | | Read counts of IsomiRs as a % of the combined miRNA and isomiR read count from respective strands | |
| | | | | Number of unique isomiRs offset at each end | **Unique 3' End :5' End Off-set IsomiR Ratio (Ratio-1)** Should be >1 if there are more unique isomiRs displaying 3' end heterogeneity | Average offset (nt) at each end | **3' End :5' End Average Absolute Off-set Ratio (Ratio-2)** Should be >1 if there is a greater degree of 3' end heterogeneity overall | Read counts of isomiRs off-set at each end as a % of total read count | **3' End :5' End Processing Consistency Ratio (Ratio-3)** Should be >1 if 5' end processing is more consistent than at the 3' end processing |
| MD11410 | 5' Arm | H361 | 5' End | 1 | 2.00 | 0.25 | 3.00 | 5.88% | 6.00 |
| | | | 3' End | 2 | | 0.75 | | 35.29% | |
| | | N076 | 5' End | 6 | 1.00 | 2.40 | 1.25 | 29.79% | 1.21 |
| | | | 3' End | 6 | | 3.00 | | 36.17% | |
| | 3' Arm | H361 | 5' End | 1 | 1.00 | 2.00 | 1.50 | 20.00% | 4.00 |
| | | | 3' End | 1 | | 3.00 | | 80.00% | |
| | | N076 | 5' End | 1 | 2.00 | 0.33 | 9.00 | 33.33% | 2.00 |
| | | | 3' End | 2 | | 3.00 | | 66.67% | |
| MD11776 | 5' Arm | H361 | 5' End | 7 | 1.57 | 1.00 | 1.80 | 46.95% | **0.96** |
| | | | 3' End | 11 | | 1.80 | | 45.30% | |
| | | N076 | 5' End | 4 | 1.43 | 0.44 | 3.25 | 40.51% | 1.13 |
| | | | 3' End | 7 | | 1.44 | | 45.61% | |
| | 3' Arm | H361 | 5' End | 14 | 1.07 | 1.10 | 1.36 | 22.68% | 1.80 |
| | | | 3' End | 15 | | 1.50 | | 40.71% | |
| | | N076 | 5' End | 10 | 1.00 | 0.85 | 1.18 | 32.17% | 1.04 |
| | | | 3' End | 10 | | 1.00 | | 33.54% | |

**Table 4.8 (Part 3/4) Analysis of isomiR end heterogeneity for putative miRNAs from CyHV-3 pre-miRNA candidates**
MiRNAs with 3' End: 5' End ratios that are <1 are highlighted in bold

| Pre-miRNA | miRNA | Infection | miRNA end | Transcripts ≧19nt representing ≧0.1% of combined miRNA and isomiR read count from respective strands | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Number of unique isomiRs | | Average absolute offset (nt) displayed by all unique isomiRs | | Read counts of IsomiRs as a % of the combined miRNA and isomiR read count from respective strands | |
| | | | | Number of unique isomiRs offset at each end | **Unique 3' End :5' End Off-set IsomiR Ratio (Ratio-1)** Should be >1 if there are more unique isomiRs displaying 3' end heterogeneity | Average offset (nt) at each end | **3' End :5' End Average Absolute Off-set Ratio (Ratio-2)** Should be >1 if there is a greater degree of 3' end heterogeneity overall | Read counts of isomiRs off-set at each end as a % of total read count | **3' End :5' End Processing Consistency Ratio (Ratio-3)** Should be >1 if 5' end processing is more consistent than at the 3' end processing |
| MR5057 | 5' Arm | H361 | 5' End | 1 | 2.00 | 0.25 | 5.00 | 7.14% | 5.50 |
| | | | 3' End | 2 | | 1.25 | | 39.29% | |
| | | N076 | 5' End | 4 | 1.75 | 1.00 | 1.70 | 3.67% | 14.92 |
| | | | 3' End | 7 | | 1.70 | | 54.80% | |
| | 3' Arm | H361 | 5' End | 2 | 2.50 | 0.30 | 3.67 | 1.16% | 18.14 |
| | | | 3' End | 5 | | 1.10 | | 20.97% | |
| | | N076 | 5' End | 4 | 1.00 | 0.60 | 1.67 | 1.35% | 7.83 |
| | | | 3' End | 4 | | 1.00 | | 10.56% | |
| MR5075 | 5' Arm | H361 | 5' End | 6 | 1.33 | 1.80 | 1.39 | 18.75% | 3.89 |
| | | | 3' End | 8 | | 2.50 | | 72.92% | |
| | | N076 | 5' End | 11 | 1.18 | 1.50 | 1.20 | 9.84% | 4.63 |
| | | | 3' End | 13 | | 1.80 | | 45.60% | |
| | 3' Arm | H361 | 5' End | 12 | 1.25 | 1.58 | 1.65 | 42.16% | **0.96** |
| | | | 3' End | 15 | | 2.60 | | 40.67% | |
| | | N076 | 5' End | 16 | 1.00 | 1.90 | 1.32 | 49.15% | **0.91** |
| | | | 3' End | 16 | | 2.50 | | 44.72% | |

**Table 4.8 (Part 4/4) Results of isomiR end heterogeneity analysis for putative miRNAs from CyHV-3 pre-miRNA candidates**
MiRNAs with 3' End: 5' End ratios that are <1 are highlighted in bold

| Pre-miRNA | miRNA | Infection | miRNA end | Transcripts ≧19nt representing ≧0.1% of combined miRNA and isomiR read count from respective strands | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Number of unique isomiRs | | Average absolute offset (nt) displayed by all unique isomiRs | | Read counts of IsomiRs as a % of the combined miRNA and isomiR read count from respective strands | |
| | | | | Number of unique isomiRs offset at each end | **Unique 3' End :5' End Off-set IsomiR Ratio (Ratio-1)** Should be >1 if there are more unique isomiRs displaying 3' end heterogeneity | Average offset (nt) at each end | **3' End :5' End Average Absolute Off-set Ratio (Ratio-2)** Should be >1 if there is a greater degree of 3' end heterogeneity overall | Read counts of isomiRs off-set at each end as a % of total read count | **3' End :5' End Processing Consistency Ratio (Ratio-3)** Should be >1 if 5' end processing is more consistent than at the 3' end processing |
| MR6201 | 5' Arm | H361 | 5' End | 1 | 3.00 | 0.50 | 2.00 | 20.00% | 3.00 |
| | | | 3' End | 3 | | 1.00 | | 60.00% | |
| | | N076 | 5' End | 12 | 1.25 | 1.80 | 1.33 | 61.90% | 1.19 |
| | | | 3' End | 15 | | 2.40 | | 73.81% | |
| | 3' Arm | H361 | 5' End | 2 | 1.00 | 0.80 | 1.25 | 22.22% | 1.00 |
| | | | 3' End | 2 | | 1.00 | | 22.22% | |
| | | N076 | 5' End | 5 | 1.20 | 1.40 | 1.50 | 15.15% | 2.00 |
| | | | 3' End | 6 | | 2.10 | | 30.30% | |

#### 4.1.3.8 *Identification of high probability pre-miRNA candidates*

By analysing the candidate pre-miRNAs in the manner described above, it was possible to identify high probability pre-miRNA candidates that complied with the criteria outlined in Section 2.14.3.6. Pre-miRNA candidates that did not comply with all of the criteria were eliminated. By default, all pre-miRNA candidates complied with points 1-4 as these were the criteria used for their initial identification. The main reason for eliminating most pre-miRNA candidates was non-compliance with points 11, 9, 8 and 6 (in descending order of effectiveness). In total, six out of the 21 pre-miRNA candidates MD1111, MD9812, MD11410, MD11776, MR5057 and MR5075 complied with all the pre-miRNA identification criteria outlined earlier (Section 2.14.3.6). As a result, these six pre-miRNA candidates were considered to be high probability CyHV-3 pre-miRNAs. The results of this assessment are summarized in Table 4.9

**Table 4.9 A summary of all CyHV-3 pre-miRNA candidates in terms of compliance with pre-miRNA identification criteria.**
Pre-miRNA candidates that complied with all of the criteria are highlighted in grey

| Pre-miRNA Name | 1. Highly expressed small RNAs | 2. From non-coding region | 3. Putative miRNAs adjacent to stem | 4. Putative major and minor form detected | 5. MiRNA isoform stability | 6. Classified as real pre-miRNA | 7. Pre-miRNA structure has MFE <-20 kcal/mol | 8. Mature-miRNA-duplex 3' overhangs | 9. MiRNA-like alignment signature | 10. Greater 3' end heterogeneity among isomiRs | 11. Discrete, read-enriched locus | Compliance with all points |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MD759 | **Yes** | **Yes (3' UTR)** | **Yes** | **Yes** | No | No | **Yes** | No | No | N/A | No | No |
| MD773 | **Yes** | **Yes (3' UTR)** | **Yes** | **Yes** | No | No | No | No | No | N/A | No | No |
| MD20185 | **Yes** | **Yes (3' UTR)** | **Yes** | **Yes** | Yes | No | No | No | No | N/A | No | No |
| MD20189 | **Yes** | **Yes (3' UTR)** | **Yes** | **Yes** | Yes | No | No | No | No | N/A | No | No |
| MD1111 | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** |
| MD9812 | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** |
| MD11410 | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** |
| MD11776 | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** |
| MR5057 | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** |
| MR5075 | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** |
| MD6201 | **Yes** | **Yes (3' UTR)** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | No | **Yes** | **Yes** | No | No |
| MR7404 | **Yes** | **Yes (3' UTR)** | **Yes** | **Yes** | **Yes** | No | **Yes** | **Yes** | No | N/A | No | No |
| MR7409 | **Yes** | **Yes (3' UTR)** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | No | No | N/A | No | No |
| MR7412 | **Yes** | **Yes (3' UTR)** | **Yes** | **Yes** | **Yes** | No | No | No | No | N/A | No | No |
| MR7418 | **Yes** | **Yes (3' UTR)** | **Yes** | **Yes** | **Yes** | No | No | No | No | N/A | No | No |
| MR7434 | **Yes** | **Yes (3' UTR)** | **Yes** | **Yes** | No | No | **Yes** | No | No | N/A | No | No |
| MR7440 | **Yes** | **Yes (3' UTR)** | **Yes** | **Yes** | **Yes** | No | **Yes** | No | No | N/A | No | No |
| MR7460 | **Yes** | **Yes (3' UTR)** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | No | N/A | No | No |
| MR7468 | **Yes** | **Yes (3' UTR)** | **Yes** | **Yes** | No | No | No | No | No | N/A | No | No |
| MR7471 | **Yes** | **Yes (3' UTR)** | **Yes** | **Yes** | No | **Yes** | **Yes** | No | No | N/A | No | No |
| MR7476 | **Yes** | **Yes (3' UTR)** | **Yes** | **Yes** | **Yes** | No | **Yes** | No | No | N/A | No | No |

### 4.1.4 Automated identification of miRNAs from deep sequencing data

In order to support the findings made following non-automated identification of miRNAs and to possibly identify additional high probability pre-miRNA candidates, the same deep sequencing data was also analysed using two automated methods of miRNA identification, namely MirDeep and Mireap. The choice of methods was based on the results of a performance comparison study in which these two methods were found to be most suitable for predicting novel miRNAs (see section 1.7.2.2 and Li et al (2012)). The decision to use both of these methods together was based on other findings in the same comparison study by Li et al (2012) which suggested that the best practice is to use more than one method to support the classification of specific loci as pre-miRNAs

#### 4.1.4.1 *Automated identification of CyHV-3 miRNAs using MirDeep and Mireap*

The use of two different automated methods to detect miRNAs from two different deep sequencing experiments resulted in 4 sets of results. (Table 4.10, Table 4.11, Table 4.12, and Table 4.13). The pre-miRNA sequences, miRNA sequences and alignment signatures for these results can be seen in Supplementary File 4.4. All of these were then cross-compared in order to identify pre-miRNA candidates that were common to multiple sets of results. A summary of the pre-miRNA candidates identified and the numbers that were common to different sets of results can be seen in Table 4.14. More pre-miRNA candidates were identified by Mireap than by miRDeep. This was because Mireap identified many pre-miRNA candidates in the absence of a suitable minor form derived from the opposite arm of the pre-miRNA candidates. Both methods identified pre-miRNA candidates within protein-coding regions. Consistent with the criteria used for non-automated identification of pre-miRNA candidates, any candidates identified using the automated methods that occurred within protein-coding genes and those with no accompanying minor forms were eliminated from further analysis. MiRDeep also assigns scores to all pre-miRNA candidates that it identifies, allowing ranking in terms of how likely they were to be genuine pre-miRNAs. Using miRDeep, the top 6 and top 3 ranked candidates from the H361 (Table 4.10) and N076 (Table 4.11) infections respectively were common to

both sets of results (ignoring candidates within protein-coding regions, marked in red) (summarized in Table 4.15). These pre-miRNA candidates also have the highest read counts of all candidates from non-coding regions in their respective infections. Li et al. (2012) concluded that it should be possible to identify most genuine pre-miRNAs from deep sequencing data by more than one method. Interestingly, the six candidates mentioned above were also the *only* candidates to be independently identified using Mireap, with the exception of one pre-miRNA candidate from a protein-coding region (Table 4.15). Due to the fact that these six candidates were different from the rest in this respect, and on the basis that identification by multiple methods supports pre-miRNA classification, they were deemed to be high-probability pre-miRNA candidates. Furthermore, five of the six were also among those deemed to be high probability pre-miRNA candidates from non-automated analysis (Table 4.9). This indicated that elimination of candidates *not* identified by both methods had the effect of retaining only high-quality pre-miRNA candidates. Incidentally, inspection of the eliminated pre-miRNA candidates revealed that none of them conformed to the pre-miRNA identification criteria (data not shown; alignment signatures can be seen in Supplementary File 4.4). In addition to pre-miRNA candidates that were identified by both automated methods, those found in both infections by the one method only are shown in Table 4.15. Only one such pre-miRNA candidate (MD11704) was identified. MD11704 conformed to all of the pre-miRNA identification criteria from Section 2.14.3.6 except Point 1 as the putative miRNAs were not highly expressed and did not have read counts ≥10 in either infection (this candidate is examined in more detail, for other reasons, in Section 6.1). Although all of the putative miRNAs from all high probability pre-miRNA candidates were present in each infection, not all of the pre-miRNA candidates were identified in both infections by the automated methods. For example MD11410 was not identified by Mireap in the N076 infection, MD9812 was not identified in the N076 infection by miRDeep and both MD1111 and MR6201 were only identified in the H361 infection by both methods (Table 4.15). Also, it is unclear as to why MR5075, a high probability pre-miRNA candidate from non-automated analysis, was not identified at all using the automated methods.

**Table 4.10 Pre-miRNA candidates predicted by miRDeep from the CyHV-3 H361 infection**

| Provisional ID (assigned by miRDeep) | Equivalent VMir Prediction ID | Rank | miRDeep2 score | Total Read Count (Including isomiRs) |
|---|---|---|---|---|
| KHVU_4678 | MR5057 | 1 | 40967 | 80360 |
| KHVU_3683 | MD11111 | 2 | 3523.6 | 6908 |
| KHVU_193 | MD11111 (repeat) | 3 | 3523.6 | 6908 |
| KHVU_2055 | MD11776 | 4 | 2306.6 | 4521 |
| KHVU_6996 | MR18527 | 5 | 96.2 | 195 |
| KHVU_2519 | MD14318 | 6 | 89.9 | 181 |
| KHVU_1710 | MD9812 | 7 | 28.7 | 60 |
| KHVU_4945 | MR6545 | 8 | 12.2 | 29 |
| KHVU_1990 | MD11410 | 9 | 8.7 | 21 |
| KHVU_5500 | MD9701 | 10 | 7.5 | 21 |
| KHVU_4878 | MR6201 | 11 | 4.9 | 15 |
| KHVU_4908 | MR6373 | 12 | 4.3 | 12 |
| KHVU_1288 | MD7493 | 13 | 3 | 3 |
| KHVU_6071 | MR13096 | 14 | 2.6 | 58 |
| KHVU_2189 | MD12474 | 15 | 2.2 | 175 |
| KHVU_4505 | MR4130 | 16 | 2.2 | 11 |
| KHVU_1380 | MD7990 | 17 | 2.1 | 77 |
| KHVU_6794 | MR17381 | 18 | 0.5 | 9 |
| KHVU_3991 | MR1328 | 19 | 0.4 | 16 |
| KHVU_7469 | MR1328 (repeat) | 20 | 0.4 | 16 |
| KHVU_2367 | MD13496 | 21 | 0 | 5 |

| Key | Explanation |
|---|---|
| | Previously deemed to be high-probability pre-miRNA candidate from non-automated analysis of deep sequencing data |
| | New pre-miRNA candidate from automated analysis of deep sequencing data |
| | Previously deemed to be low probability pre-miRNA candidate from non-automated analysis of deep sequencing data |
| | Occurs within protein-coding region (eliminated) |

**Table 4.11 Pre-miRNA candidates predicted by miRDeep from the CyHV-3 N076 infection**

| Provisional ID (assigned by miRDeep) | Equivalent VMir Prediction ID | Rank | miRDeep2 score | Total Read Count (Including isomiRs) |
|---|---|---|---|---|
| KHVU_5130 | MR5057 | 1 | 674758.2 | 1323514 |
| KHVU_2337 | MD11776 | 2 | 14677 | 28785 |
| KHVU_1442 | Not Predicted | 3 | 298.9 | 593 |
| KHVU_7440 | Not Predicted | 4 | 80.7 | 164 |
| KHVU_840 | MD4173 | 5 | 74 | 150 |
| KHVU_6024 | MR9635 | 6 | 56.8 | 119 |
| KHVU_1146 | MD5909 | 7 | 47.3 | 98 |
| KHVU_2258 | MD11410 | 8 | 22.5 | 48 |
| KHVU_6494 | MR12001 | 9 | 21 | 46 |
| KHVU_8115 | MR159 | 10 | 9 | 30 |
| KHVU_4183 | MR159 (repeat) | 11 | 9 | 30 |
| KHVU_2722 | MD13747 | 12 | 5.6 | 16 |
| KHVU_1084 | MD5586 | 13 | 2.5 | 516 |
| KHVU_7144 | MR15193 | 14 | 2.4 | 2 |
| KHVU_1590 | MD7990 | 15 | 2.1 | 155 |
| KHVU_2930 | MD14795 | 16 | 1.9 | 52 |
| KHVU_8057 | MR19774 | 17 | 1.9 | 142 |
| KHVU_6960 | MR14203 | 18 | 1.9 | 9 |
| KHVU_8040 | Not Predicted | 19 | 0 | 17 |

| Key | Explanation |
|---|---|
| | **Previously deemed to be high-probability pre-miRNA candidate from non-automated analysis of deep sequencing data** |
| | **New pre-miRNA candidate from automated analysis of deep sequencing data** |
| | **Occurs within protein-coding region (eliminated)** |

**Table 4.12 Pre-miRNA candidates predicted by Mireap from the CyHV-3 H361 infection**

| Provisional ID (assigned by Mireap) | Equivalent VMir Prediction ID |
|---|---|
| m0001 | MD1111 |
| m0002 | MD3150 |
| m0003 | MD4188 |
| m0004 | MD5234 |
| m0008 | MD8620 |
| m0009 | MD9812 |
| m0010 | Not predicted |
| m0011 | MD11410 |
| m0012 | MD11704 |
| m0013 | MD11776 |
| m0015 | MD13551 |
| m0019 | MD1111 (repeat) |
| m0020 | MR1635 |
| m0021 | MR2459 |
| m0022 | MR2741 |
| m0023 | MD2877 |
| m0024 | MR3613 |
| m0027 | MR5057 |
| m0030 | MR5552 |
| m0031 | MR6201 |
| m0034 | MR7191 |
| m0035 | MR7627 |
| m0042 | MR11068 |
| m0050 | MR17129 |
| m0051 | None |
| m0053 | MR17383 |
| m0054 | MR18408 |
| m0055 | MR19207 |

| Key | Explanation |
|---|---|
| | Previously deemed to be high-probability pre-miRNA candidate from non-automated analysis of deep sequencing data |
| | New pre-miRNA candidate from automated analysis of deep sequencing data |
| | Previously deemed to be low probability pre-miRNA candidate from non-automated analysis of deep sequencing data |
| | Occurs within protein-coding region (eliminated) |

**Table 4.13 Pre-miRNA candidates predicted by Mireap from the CyHV-3 N076 infection**

| Provisional ID (assigned by Mireap) | Equivalent VMir Prediction ID |
|---|---|
| M0001 | MD8080 |
| M0002 | MD9812 |
| M0003 | MD11704 |
| M0004 | MD11776 |
| M0007 | MD16282 |
| M0008 | Not predicted |
| M0009 | MR5057 |
| M0011 | MR6853 |
| M0014 | Not predicted |
| M0016 | Not predicted |
| M0018 | Not predicted |
| M0019 | MR15158 |

| Key | Explanation |
|---|---|
| | **Previously deemed to be high-probability pre-miRNA candidate from non-automated analysis of deep sequencing data** |
| | **New pre-miRNA candidate from automated analysis of deep sequencing data** |
| | **Occurs within protein-coding region (eliminated)** |

**Table 4.14 Summary of results from automated identification of miRNAs from sequencing data**

| Method | | miRDeep | | Mireap | |
|---|---|---|---|---|---|
| Infection | | H361 Infection | N076 Infection | H361 Infection | N076 Infection |
| Total Predicted Pre-miRNAs | | 19 (+ 2repeat) | 18 (+1 repeat) | 55 (28 have no minor form) | 21 (9 have no minor form) |
| Protein-coding regions | Total number of pre-miRNA candidates found in protein-coding regions | 10 | 9 | 15 (with minor form) | 5 (with minor form) |
| | Found by both miRDeep and Mireap in either infection | 1 | 0 | 1 | 0 |
| | Found in both infections by same method | 1 | 1 | 0 | 0 |
| | Found in different infection by different method | 0 | 0 | 0 | 0 |
| | Found in both infections by both methods | 0 | 0 | 0 | 0 |
| Non-protein-coding regions | Total amount of pre-miRNA candidates in non-protein-coding regions | 9 (+2 repeats) | 9 (+1 repeat) | 12 (with minor form +1 repeat) | 7 (with minor form) |
| | Found by both miRDeep and Mireap in either infection | 6 | 3 | 6 | 3 |
| | Found in both infections by same method | 3 | 3 | 4 | 4 |
| | Found in different infection by different method | 3 | 2 | 3 | 3 |
| | Found in both infections by both methods | 2 | 2 | 2 | 2 |
| | Not found in both infections by the same method or in either infection by different methods. These were all also analysed using MiPred and the CSHMM. None were classified as real pre-miRNAs | 3 | 6 | 5 | 3 |

**Table 4.15 Summary of CyHV-3 pre-miRNA candidates from non-coding regions predicted by both automated methods or in both infections by the same method.**

| Predicted pre-miRNA name | Set of results that pre-miRNA candidates were identified in | | | | Found by both methods | Found in different infections by one or both methods |
|---|---|---|---|---|---|---|
| | miRDeep H361 Infection | miRDeep N076 Infection | Mireap H361 Infection | Mireap N076 Infection | | |
| MR5057 | Yes | Yes | Yes | Yes | Yes | Yes |
| MD11776 | Yes | Yes | Yes | Yes | Yes | Yes |
| MD11410 | Yes | Yes | Yes | No | Yes | Yes |
| MD9812 | Yes | No | Yes | Yes | Yes | Yes |
| MD1111 | Yes | No | Yes | No | Yes | No |
| MR6201 | Yes | No | Yes | No | Yes | No |
| MD11704 | No | No | Yes | Yes | No | Yes |

| Key | Explanation |
|---|---|
| | Previously deemed to be high-probability pre-miRNA candidate from non-automated analysis of deep sequencing data |
| | New pre-miRNA candidate from automated analysis of deep sequencing data |
| | Previously deemed to be low probability pre-miRNA candidate from non-automated analysis of deep sequencing data |

### 4.1.5 Summary -identification of pre-miRNA candidates from deep sequencing data

The combined list of high-probability candidates from both non-automated and automated analysis gave a total of seven high probability candidates (Table 4.16). MR6201 was not deemed to be a high probability candidate following non-automated analysis due to non-conformance with points 8 and 11 in the pre-miRNA identification criteria and for this reason it was eliminated even though it was identified by both automated methods. MR5075 was identified as a high probability pre-miRNA following non-automated analysis although it was not identified from the same deep sequencing data using the automated methods. Due to its conformance with the pre-miRNA identification criteria and because of the fact that other pre-miRNA candidates identified by both automated methods were not identified in both infections (even though they were present in both infections) it was decided to retain MR5075 in the final list of high probability pre-miRNA candidates. This resulted in a total of six high probability pre-miRNA candidates for further characterisation by several other methods (Table 4.16).

**Table 4.16 Comparison of high-probability pre-miRNA candidates following non-automated and automated analysis of deep sequencing data.**
Those highlighted in dark-grey were classified as the most likely pre-miRNA candidates overall following both non-automated and automated analysis of the deep sequencing data

| Pre-miRNA name | High- probability miRNA candidate in: | |
|---|---|---|
| | Non-automated approach | Automated approach |
| MR5057 | Yes | Yes |
| MD11776 | Yes | Yes |
| MD11410 | Yes | Yes |
| MD9812 | Yes | Yes |
| MD1111 | Yes | Yes |
| MR5075 | Yes | No |
| MR6201 | No | Yes |

## 4.2 DNA Microarray analysis

In order to support the findings made following small RNA deep sequencing (Section 4.1) the same RNA sample was also analysed by DNA microarray hybridization. Size-selected small RNAs, from both miRNA inclusive and non-inclusive size ranges were tested to see if probes targeting putative miRNAs from high probability pre-miRNAs gave positive signals in the miRNA size range only. This was on the basis that mature miRNAs are specifically enriched in the 17-25 nt size range. Also, because DNA array hybridization can be carried out in a non-enzymatically biased manner, it was also used to test if 'highly abundant' small RNAs that were detected by deep sequencing were in fact merely over-represented in the data due to enzymatic bias. In addition, the high throughput nature of the method facilitated targeting of many more theoretical CyHV-3 miRNAs that may have been under represented in the deep sequencing data for the same reasons. Endogenous and exogenous positive controls were also employed throughout. A work-flow diagram for this is outlined in Figure 4.14.

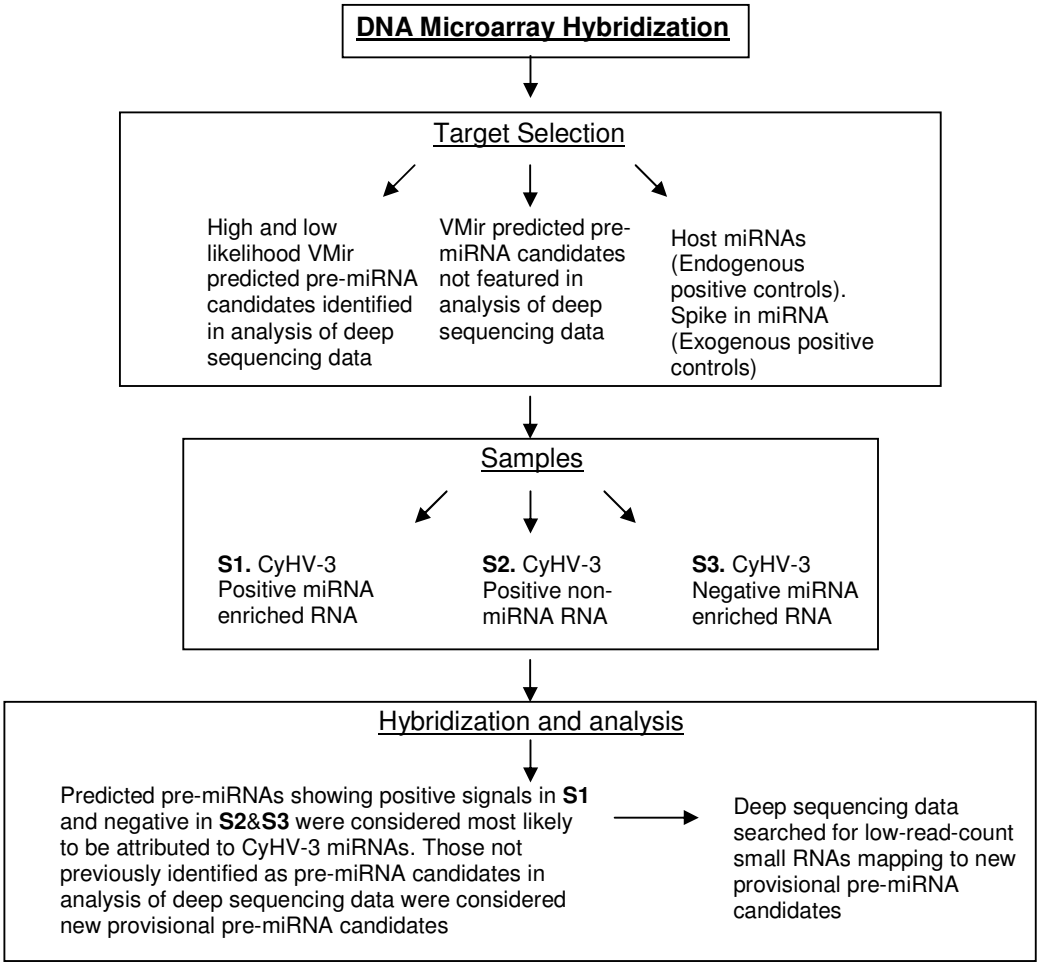



**Figure 4.14 Outline of workflow for DNA Microarray hybridization**

## 4.2.1 Positive controls

Oligonucleotide probes targeting transcripts known to be present or absent in specific samples were also included on the array. These controls were included in order to confirm that the samples were prepared correctly and fit for purpose. There were two types of positive control: endogenous and exogenous. Their detection and absence, respectively, in specific samples were ideal indicators of correct RNA size fractionation, labelling/purification and overall RNA quality.

### 4.2.1.1 *Endogenous positive controls*

Probes targeting five *Cyprinus carpio* (host) miRNAs that were identified following analysis of deep sequencing data, namely, ccr-miR-143, ccr-miR-21, ccr-miR-22a,

ccr-let-7a and ccr-et-7b were included on the array. MiRNAs should only be present in the 17-25 nt RNA samples (S1 and S3) and not in the 26-35 nt RNA sample (S2). If size fractionation was done correctly, probes for these miRNAs should give positive results in S1 and S3 and not in S2. The hybridization intensities for endogenous positive control probes in all samples are shown in Table 4.17. In S1 and S3, all replicate probes for all miRNA targets display much higher intensities than the background for their respective hybridizations and much higher intensities than their associated mismatch and scrambled control probes. The intensity values for probes targeting ccr-miR-21, ccr-miR-22a, ccr-let-7a and ccr-et-7b are either at saturation (showing a maximum value of 65535) or close to saturation. The intensity values for probes targeting ccr-miR-143 are not as high but are still significantly over background levels. The RVs for all host miRNA targets are all above the cut-off values for S1 and S3 respectively. Conversely, for S2, the same probes display lower intensity values that are much closer to background levels and none of their respective RVs are over the cut-off value for the S2 hybridization.

#### 4.2.1.2 *Exogenous Positive controls*

A synthetic miRNA was spiked into all labelling reactions. Twenty-four replicate probes targeting this exogenous control were included on the array. As expected, all of these probes showed high intensity values. In both S1 and S3 these probes displayed intensity values that were either at or close to saturation and had RVs of 37.7 and 35.9 respectively. While the intensities for these probes in the S2 sample were also quite high they were not as high as the corresponding values from S1 and S2 resulting in a lower RVs of 14.0. These results are shown in Table 4.18.

**Table 4.17 Hybridization intensities for endogenous positive control probes in all samples**

| | Endogenous Positive controls | | | |
|---|---|---|---|---|
| | **Probe Type** | **S1. H361 17-25 nt (Background 1661) RV cut-off: 1.8** | **S2. H361 26-35 nt (Background 1302) RV cut-off: 1.6** | **S3. CCB 17-25 nt (Background 1813) RV cut-off: 1.9** |
| **ccr-miR-143** | Probe Replicate 1 | 34873 | 1562 | 65535 |
| | Probe Replicate 2 | 33151 | 1159 | 65535 |
| | Probe Replicate 3 | 32374 | 1537 | 65535 |
| | Probe Replicate 4 | 18272 | 2770 | 65535 |
| | Mismatch Control 1 | 2094 | 2827 | 1962 |
| | Mismatch Control 2 | 1491 | 1072 | 2110 |
| | Scrambled Control 1 | 1778 | 1668 | 1836 |
| | Scrambled Control 2 | 2118 | 1997 | 1416 |
| | **RV** | **17.8** | **1.3** | **36.0** |
| **ccr-miR-21** | Probe Replicate 1 | 65535 | 3694 | 65535 |
| | Probe Replicate 2 | 65535 | 1202 | 65535 |
| | Probe Replicate 3 | 65535 | 2656 | 65535 |
| | Probe Replicate 4 | 65535 | 3283 | 65535 |
| | Mismatch Control 1 | 1241 | 890 | 1141 |
| | Mismatch Control 2 | 1595 | 1050 | 1419 |
| | Scrambled Control 1 | 2047 | 3103 | 2541 |
| | Scrambled Control 2 | 1596 | 3098 | 1178 |
| | **RV** | **39.5** | **1.4** | **36.1** |
| **ccr-miR-22a** | Probe Replicate 1 | 65535 | 3334 | 65535 |
| | Probe Replicate 2 | 65535 | 2451 | 65535 |
| | Probe Replicate 3 | 65535 | 2439 | 65535 |
| | Probe Replicate 4 | 65535 | 4204 | 64385 |
| | Mismatch Control 1 | 2394 | 2057 | 2127 |
| | Mismatch Control 2 | 1951 | 1398 | 2437 |
| | Scrambled Control 1 | 1172 | 2991 | 1142 |
| | Scrambled Control 2 | 1167 | 3002 | 891 |
| | **RV** | **39.3** | **1.5** | **35.7** |
| **ccr-let-7a** | Probe Replicate 1 | 65535 | 1196 | 65535 |
| | Probe Replicate 2 | 63437 | 1204 | 65535 |
| | Probe Replicate 3 | 65535 | 1172 | 65535 |
| | Probe Replicate 4 | 65535 | 1717 | 65535 |
| | Mismatch Control 1 | 1420 | 868 | 305 |
| | Mismatch Control 2 | 2270 | 3268 | 1934 |
| | Scrambled Control 1 | 1883 | 2373 | 2672 |
| | Scrambled Control 2 | 840 | 949 | 1765 |
| | **RV** | **39.1** | **1.0** | **35.9** |
| **ccr-et-7b** | Probe Replicate 1 | 65535 | 1609 | 65535 |
| | Probe Replicate 2 | 65535 | 3493 | 65535 |
| | Probe Replicate 3 | 64193 | 1039 | 65535 |
| | Probe Replicate 4 | 59354 | 902 | 65535 |
| | Mismatch Control 1 | 2498 | 4627 | 2674 |
| | Mismatch Control 2 | 2331 | 1472 | 997 |
| | Scrambled Control 1 | 1408 | 2928 | 1329 |
| | Scrambled Control 2 | 1897 | 2246 | 2531 |
| | **RV** | **38.0** | **1.4** | **36.1** |

**Table 4.18 Hybridization intensities for exogenous positive control probes in all samples**

| Exogenous Positive Control: Synthetic ath-miR156g spike-in | | | |
|---|---|---|---|
| **Probe** | **S1. H361 17-25 nt (Background 1661) RV cut-off: 1.8** | **S2. H361 26-35 nt (Background 1302) RV cut-off: 1.6** | **S3. CCB 17-25 nt (Background 1813) RV cut-off: 1.9** |
| Probe Replicate 1 | 65535 | 15124 | 65535 |
| Probe Replicate 2 | 65535 | 18907 | 65535 |
| Probe Replicate 3 | 65535 | 18176 | 65535 |
| Probe Replicate 4 | 65535 | 19045 | 65535 |
| Probe Replicate 5 | 64538 | 20077 | 65535 |
| Probe Replicate 6 | 60049 | 16909 | 65535 |
| Probe Replicate 7 | 65535 | 24482 | 64147 |
| Probe Replicate 8 | 65535 | 14581 | 65535 |
| Probe Replicate 9 | 56062 | 12409 | 65535 |
| Probe Replicate 10 | 65535 | 16358 | 65535 |
| Probe Replicate 11 | 65535 | 20918 | 65535 |
| Probe Replicate 12 | 65535 | 23303 | 65535 |
| Probe Replicate 13 | 65535 | 20780 | 65535 |
| Probe Replicate 14 | 65535 | 13954 | 65535 |
| Probe Replicate 15 | 65535 | 18211 | 65535 |
| Probe Replicate 16 | 65535 | 25818 | 64178 |
| Probe Replicate 17 | 65535 | 16520 | 65535 |
| Probe Replicate 18 | 65535 | 17377 | 65535 |
| Probe Replicate 19 | 65535 | 23901 | 65535 |
| Probe Replicate 20 | 65535 | 23016 | 65535 |
| Probe Replicate 21 | 65535 | 20919 | 65535 |
| Probe Replicate 22 | 61562 | 15595 | 65535 |
| Probe Replicate 23 | 65535 | 17661 | 65535 |
| Probe Replicate 24 | 65535 | 15858 | 65535 |
| Mismatch Control Probe Replicate 1 | 1982 | 2824 | 1247 |
| Mismatch Control Probe Replicate 2 | 1160 | 502 | 1343 |
| Mismatch Control Probe Replicate 3 | 2199 | 3128 | 2116 |
| Mismatch Control Probe Replicate 4 | 1159 | 826 | 1418 |
| Scrambled Control Probe Replicate 1 | 1372 | 482 | 1830 |
| Scrambled Control Probe Replicate 2 | 10425 | 1539 | 729 |
| Scrambled Control Probe Replicate 3 | 2296 | 2075 | 2379 |
| Scrambled Control Probe Replicate 4 | 1104 | 578 | 1921 |
| **RV** | **37.7** | **14.0** | **35.9** |

## 4.2.2 Analysis of twenty-one CyHV-3 pre-miRNA candidates

In Section 4.1.3.1, twenty-one CyHV-3 pre-miRNA candidates were identified. Six of these were deemed to be high probability pre-miRNA candidates. Assuming that these are real then small RNAs (putative miRNAs) processed from these pre-miRNAs should be specifically enriched in the 17-25 nt RNA and not present in non-miRNA size fractions such as that covering 26-35 nt. Conversely, small RNAs mapping to low probability pre-miRNA candidates are more likely to be degradation products from

larger transcripts, therefore these target sequences should be evenly distributed across many RNA size ranges and not specifically enriched in the 17-25 nt size fraction. In order to test this assumption and to support the conclusions from Section 4.1, arrays containing probes targeting putative miRNAs from these pre-miRNAs were hybridized to non-enzymatically labelled RNA from samples S1, S2 and S3 (the latter used as a negative control).

Putative miRNAs from four out of the six high probability pre-miRNAs were detected in S1. These were detected from both arms of pre-miRNA candidates MR5057 and MD11776; putative miRNAs were detected from only one arm of pre-miRNA candidates MD1111 and MR5075. There were no significant corresponding signals in S2 and S3 for any of these 4 pre-miRNAs, indicating that these target sequences were specifically enriched in the miRNA RNA size fraction and that signals in this sample were not the result of non-specific hybridization to host small RNAs in the miRNA size range. The hybridization intensities and RVs for miRNA probes from these 4 pre-miRNAs are shown in Table 4.19.

The RVs (in S1) for these pre-miRNAs were then compared to the corresponding read counts obtained following deep sequencing analysis of the same sample from the H361 infection (see Figure 4.15). As with the deep sequencing data, the putative miRNA from the 3' arm of MR5057 was the most highly expressed putative miRNA from these four pre-miRNA candidates. The RVs suggested that there was less of a difference in expression levels between the major and minor forms derived from both MR5057 and MD11776. Only one miRNA was detected for MD1111 and MR5075. Although the read counts from both infections indicated that it was the miRNA from the 3' arm that was the major form derived from pre-miRNA candidate MR5075 (Table 4.5), only the miRNA from the 5' arm was detected on the array. In the case of MD1111, it was not clear from the deep sequencing data which miRNA was the major form as this was different in both infections (Table 4.5) and it was only the 5' miRNA that was detected on the array. The array results for MD1111 and MR5075 are the opposite of those obtained from deep sequencing analysis of the same sample (H361) with regard to the apparent major forms derived from these precursors, thus it suggests that the putative miRNA from the 3' ends of these pre-miRNA candidates was over represented in the sequencing data.

**Table 4.19 Hybridization intensities for probes targeting the 4 high probability pre-miRNA candidates**

The RV for each pre-miRNA represents that of the miRNA (5' or 3' arm) with the highest signal

| | Probe | S1. H361 17-25 nt (Background 1661) RV cut-off: 1.8 | S2. H361 26-35 nt (Background 1302) RV cut-off: 1.6 | S3. CCB 17-25 nt (Background 1813) RV cut-off: 1.9 |
|---|---|---|---|---|
| MR5057 | 5' Arm Replicate 1 | 8056 | 2457 | 2089 |
| | 5' Arm Replicate 2 | 7648 | 1637 | 1636 |
| | 3' Arm Replicate 1 | 35107 | 2849 | 1135 |
| | 3' Arm Replicate 2 | 46549 | 1568 | 1141 |
| | 5' Scrambled Control | 1419 | 2605 | 3467 |
| | 3' Scrambled Control | 1077 | 1692 | 1706 |
| | 5' Mismatch Control | 2494 | 3200 | 2726 |
| | 3' Mismatch Control | 5399 | 3400 | 9507 |
| | 5' Flanking control | 935 | 720 | 1084 |
| | 3' Flanking control | 1315 | 1080 | 1793 |
| | **RV** | **22.3** | **1.0** | **1.0** |
| MD11776 | 5' Arm Replicate 1 | 6702 | 1127 | 1131 |
| | 5' Arm Replicate 2 | 6150 | 1261 | 669 |
| | 3' Arm Replicate 1 | 14531 | 1960 | 2413 |
| | 3' Arm Replicate 2 | 14535 | 2055 | 2665 |
| | 5' Scrambled Control | 2182 | 985 | 1834 |
| | 3' Scrambled Control | 6787 | 1806 | 9576 |
| | 5' Mismatch Control | 1723 | 1854 | 1992 |
| | 3' Mismatch Control | 331 | 476 | 1208 |
| | 5' Flanking control | 1345 | 1776 | 1601 |
| | 3' Flanking control | 2493 | 3488 | 2783 |
| | **RV** | **5.6** | **1.2** | **1.0** |
| MD1111 | 5' Arm Replicate 1 | 2898 | 1033 | 1164 |
| | 5' Arm Replicate 2 | 4095 | 1704 | 2554 |
| | 3' Arm Replicate 1 | 3338 | 2221 | 1609 |
| | 3' Arm Replicate 2 | 3656 | 1920 | 1696 |
| | 5' Scrambled Control | 1862 | 1389 | 1322 |
| | 3' Scrambled Control | 5250 | 1132 | 3772 |
| | 5' Mismatch Control | 1135 | 458 | 1307 |
| | 3' Mismatch Control | 2261 | 2036 | 1543 |
| | 5' Flanking control | 845 | 718 | 756 |
| | 3' Flanking control | 2348 | 1118 | 1854 |
| | **RV** | **2.0** | **1.0** | **1.0** |
| MR5075 | 5' Arm Replicate 1 | 4851 | 2799 | 2430 |
| | 5' Arm Replicate 2 | 4247 | 1015 | 1928 |
| | 3' Arm Replicate 1 | 2018 | 675 | 1365 |
| | 3' Arm Replicate 2 | 1660 | 432 | 1757 |
| | 5' Scrambled Control | 1947 | 988 | 2127 |
| | 3' Scrambled Control | 2038 | 840 | 1570 |
| | 5' Mismatch Control | 2133 | 2957 | 2671 |
| | 3' Mismatch Control | 1549 | 682 | 2082 |
| | 5' Flanking control | 3233 | 2038 | 3189 |
| | 3' Flanking control | 1851 | 616 | 1528 |
| | **RV** | **2.2** | **1.0** | **1.0** |

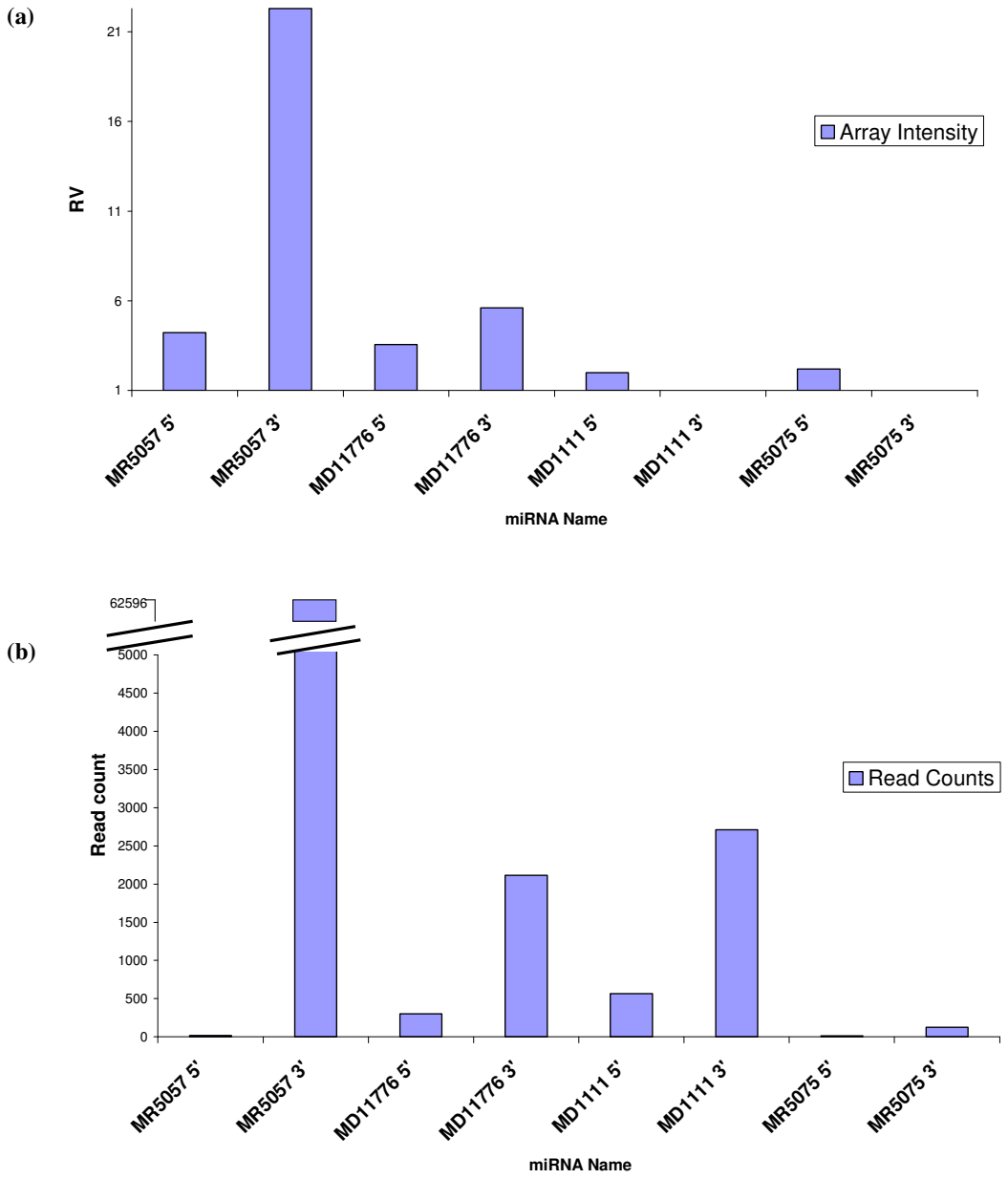



**Figure 4.15 (a) RVs for probes targeting putative miRNAs from the four high probability pre-miRNAs that showed positive signals on DNA microarray hybridization (b) Read counts corresponding to miRNAs from the same four pre-miRNAs in the H361 infection.**

Out of the fifteen low probability pre-miRNA candidates identified in non-automated analysis of the deep sequencing data, small RNAs from ten of them were also detected in S1. Nine of these also gave positive signals in S2 and one also gave a low signal in S3. The RVs for these fifteen low probability candidates and the six high probability pre-miRNAs in all three samples are compared in Table 4.20. Interestingly all of the pre-miRNAs from Reverse Cluster 2 gave signals in S2, even the three pre-miRNAs

(MR7412, MR7418 and MR7440) from this cluster that showed no signals in S1. MiRNA probes for these three precursors did display high signals in S1, but high signals from corresponding flanking control probes suggested that these miRNA probe signals were caused by larger overlapping transcripts. Therefore the signals from the miRNA probes were sufficiently corrected for this, eliminating false positive signals in S1 for these three pre-miRNAs. All of the other probes in Reverse Cluster 2 also had high signals (RVs) from flanking probes in S1 but not enough to bring their corrected values below the cut-off point for this hybridization. Hence the remaining seven pre-miRNAs in the same cluster still scored positive signals in S1.

In summary, miRNA probes for twelve of the fifteen low probability pre-miRNAs identified from the non-automated analysis of deep sequencing showed signals in S2 (26-35 nt RNA, outside the size range of miRNAs, pre-miRNAs and pri-miRNAs). These signals are likely to be caused by hybridization to larger RNA degradation products containing the same target sequence. In contrast, none of the high probability pre-miRNAs showed positive signals in S2. These results are largely what would be expected based on the observations from the deep sequencing data, and thus these results support the main findings from that experiment. Interestingly, the signals from the low probability miRNA candidate MD20189 did give signals that were of a similar pattern to those of the high probability pre-miRNA candidates (i.e. positive signal in S1 with no corresponding positive signal in S2 or S3), however, due to a complete lack of conformance with the pre-miRNA identification criteria earlier (Table 4.9), MD20189 it was not considered to be a pre-miRNA.

.

**Table 4.20 DNA microarray analysis of pre-miRNA candidates**
The RV for each pre-miRNA represents that of the miRNA (5' or 3' arm) with the highest signal. RVs that exceed the cut-off value for a given hybridization are highlighted in bold

| Pre-miRNA | Compliance with miRNA identification criteria | S1. H361 17-25 nt Cut-off: 1.8 times Background | | S2. H361 26-35nt Cut-off: 1.6 times Background | | S3. CCB 17-25 nt Cut-off: 1.9 times Background | | |
|---|---|---|---|---|---|---|---|---|
| | | RV | Percentile Rank[a] | RV | Percentile Rank[a] | RV | Percentile Rank[a] | CyHV-3 miRNA-like Signals |
| MD759 | No | **9.6** | **98.3** | **2.4** | **96.6** | 1.0 | 26.4 | No |
| MD773 | No | **11.6** | **98.9** | **1.9** | **87.7** | 1.0 | 26.4 | No |
| MD20185 | No | 1.5 | 83.5 | 1.0 | 19.0 | 0.9 | 13.1 | No |
| MD20189 | No | **5.1** | **96.2** | 1.0 | 19.0 | 1.2 | 78.0 | Yes |
| MD11410 | Yes | 0.9 | 9.5 | 0.8 | 5.2 | 0.6 | 0.5 | No |
| MD11776 | Yes | **5.6** | **96.8** | 1.2 | 61.2 | 1.0 | 26.4 | Yes |
| MD1111 | Yes | **2.0** | **89.9** | 1.0 | 52.8 | 1.0 | 61.7 | Yes |
| MD9812 | Yes | 1.0 | 18.9 | 1.0 | 19.0 | 1.0 | 26.4 | No |
| MR5057 | Yes | **22.3** | **99.7** | 1.0 | 19.0 | 1.0 | 26.4 | Yes |
| MR5075 | Yes | **2.2** | **91.1** | 1.0 | 19.0 | 1.0 | 26.4 | Yes |
| MR7404 | No | **11.1** | **98.7** | **1.9** | **91.0** | 1.6 | 88.2 | No |
| MR7409 | No | **11.2** | **98.8** | **6.8** | **99.2** | 1.0 | 26.4 | No |
| MR7412 | No | 1.00* | 18.9 | **2.1** | **94.2** | 1.1 | 67.7 | No |
| MR7418 | No | 1.00* | 18.9 | **4.7** | **98.6** | 1.0 | 26.4 | No |
| MR7434 | No | **4.1** | **95.3** | **1.6** | **84.9** | 1.0 | 26.4 | No |
| MR7440 | No | 1.00* | 18.9 | **3.5** | **98.3** | 1.1 | 72.4 | No |
| MR7460 | No | **6.3** | **97.3** | **10.0** | **99.6** | 1.0 | 26.4 | No |
| MR7468** | No | **30.5** | **99.8** | **6.6** | **99.1** | **2.2** | **92.5** | No |
| MR7471 | No | **12.4** | **98.9** | **2.7** | **97.6** | 1.0 | 26.4 | No |
| MR7476 | No | **15.7** | **99.2** | **6.7** | **99.2** | 1.0 | 26.4 | No |
| MR6201 | No | 1.5 | 85.3 | 1.0 | 19.0 | 1.1 | 65.4 | No |

*Probes for these actually gave significant values, but high values in the associated flanking control probes meant that after correction these were cancelled out. ** Despite correction for non-specific hybridization, the positive signals for MR7468 in the CCB cells suggest that some of the signal from the 17-25 nt RNA (H361 infection) may be due to non-specific hybridization to host transcripts, but this only accounts for a small proportion of the positive signal.

MiRNAs should be the most abundant transcripts in the 17-25 nt RNA sample used for deep sequencing and thus the pre-miRNA candidates listed in Table 4.20 were primarily identified on the basis of them having highly abundant small RNAs mapping to them. As these specific sequences could potentially be over-represented in the data due to enzymatic bias and because other sequences could be under-represented for the same reasons, re-analysing the same sample using array hybridization provided an ideal way to investigate this possibility in the absence of enzymatic bias. These arrays were designed to target high probability and low probability miRNA candidates from the deep sequencing experiments and also theoretical small RNAs derived from predicted pre-miRNAs that were not represented

or were ignored in the sequencing data. The small RNAs derived from the 21 pre-miRNA candidates in Table 4.20 were among the most abundant identified following the sequencing of small RNAs from the H361 infection. In agreement with this, for the S1 sample, the percentile ranks for the positive signals (RVs over a cut-off value of 1.6) among these twenty-one pre-miRNAs are quite high, ranging from 89.9 to 99.2 (average of 97.1). This indicated that regardless of whether these small RNAs were derived from high or low probability pre-miRNA candidates, in the absence of any enzymatic bias, most were still shown to be among the most abundant transcripts present in the sample. These percentile rank values refer to their rank among the results for all 2,914 other CyHV-3 pre-miRNAs targeted in the same sample.

In total, 318 of these other pre-miRNAs also showed positive signals. However, this figure represents the number of positive signals before elimination of positive results (positive signals in S3) and elimination of other false positives caused by array surface artefacts. It was necessary to eliminate such signals in order to facilitate the identification of new provisional pre-miRNA candidates based on array data. Of these 318 other positive pre-miRNAs, 136 of them were eliminated for having corresponding positive signals in S3. The array spots for the remaining 182 positives were visually inspected to ensure that they were not caused by array surface artefacts (i.e. “blobs” caused by dust, salt precipitants etc. which would produce erroneous signals). This resulted in the elimination of 111 high but artificial positive signals. (spots corresponding to positive signals from S3 were inspected in the same manner before elimination). In Table 4.21, the RVs from the remaining 71 positives were compared to the RVs of the fourteen positives from the 21 pre-miRNA candidates obtained from the sequencing of small RNAs

**Table 4.21 (Part 1/2) Summary of all positive array signals from S1 that were not attributed to CCB transcripts or array surface artefacts**

| Name | RV | Genomic region | Description | Percentile Rank |
|---|---|---|---|---|
| MR7468 | **30.5** | Non-Coding | Low Probability Pre-miRNA from Deep Sequencing | 100 |
| MR7462 | 22.6 | Non-Coding | High signal in 26-35nt RNA | 98.8 |
| MR5057 | **22.3** | Non-Coding | High Probability Pre-miRNA from Deep Sequencing | 97.6 |
| MR7476 | **15.7** | Non-Coding | Low Probability Pre-miRNA from Deep Sequencing | 96.4 |
| MR12034 | 14.3 | Non-Coding | Novel provisional Pre-miRNA Candidate | 95.2 |
| MR7471 | **12.4** | Non-Coding | Low Probability Pre-miRNA from Deep Sequencing | 94 |
| MD8620 | 11.8 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 92.8 |
| MD773 | **11.6** | Non-Coding | Low Probability Pre-miRNA from Deep Sequencing | 91.6 |
| MR7409 | **11.2** | Non-Coding | Low Probability Pre-miRNA from Deep Sequencing | 90.4 |
| MR7404 | **11.1** | Non-Coding | Low Probability Pre-miRNA from Deep Sequencing | 89.2 |
| MR14980 | 9.8 | Protein-Coding | High signal in 26-35nt RNA | 88 |
| MD759 | **9.6** | Non-Coding | Low Probability Pre-miRNA from Deep Sequencing | 86.9 |
| MR8901 | 8.4 | Non-Coding | High signal in 26-35nt RNA | 85.7 |
| MR7410 | 8.4 | Non-Coding | High signal in 26-35nt RNA | 84.5 |
| MD18707 | 7.5 | Non-Coding | Novel provisional Pre-miRNA Candidate | 83.3 |
| MR7460 | **6.3** | Non-Coding | Low Probability Pre-miRNA from Deep Sequencing | 82.1 |
| MD11776 | **5.7** | Non-Coding | High Probability Pre-miRNA from Deep Sequencing | 80.9 |
| MD20189 | **5.1** | Non-Coding | Low Probability Pre-miRNA from Deep Sequencing | 79.7 |
| MR5537 | 4.9 | Non-Coding | Novel provisional Pre-miRNA Candidate | 78.5 |
| MR8063 | 4.7 | Non-Coding | Novel provisional Pre-miRNA Candidate | 77.3 |
| MR15569 | 4.7 | Non-Coding | Novel provisional Pre-miRNA Candidate | 76.1 |
| MR9420 | 4.5 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 75 |
| MD16363 | 4.2 | Non-Coding | Novel provisional Pre-miRNA Candidate | 73.8 |
| MR7434 | **4.1** | Non-Coding | Low Probability Pre-miRNA from Deep Sequencing | 72.6 |
| MR5055 | 3.9 | Non-Coding | Novel provisional Pre-miRNA Candidate | 71.4 |
| MR11140 | 3.9 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 70.2 |
| MR5123 | 3.8 | Non-Coding | Novel provisional Pre-miRNA Candidate | 69 |
| MR5420 | 3.8 | Non-Coding | High signal in 26-35nt RNA | 67.8 |
| MD8281 | 3.7 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 66.6 |
| MD2897 | 3.3 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 65.4 |
| MD9074 | 3.2 | Non-Coding | Novel provisional Pre-miRNA Candidate | 64.2 |
| MR9157 | 3.0 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 63 |
| MR15858 | 3.0 | Non-Coding | Novel provisional Pre-miRNA Candidate | 61.9 |
| MD12312 | 3.0 | Non-Coding | Novel provisional Pre-miRNA Candidate | 60.7 |
| MD8950 | 2.8 | Non-Coding | Novel provisional Pre-miRNA Candidate | 59.5 |
| MR9187 | 2.8 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 58.3 |
| MD19009 | 2.7 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 57.1 |
| MD11118 | 2.6 | Non-Coding | Novel provisional Pre-miRNA Candidate | 55.9 |
| MR5512 | 2.6 | Non-Coding | Novel provisional Pre-miRNA Candidate | 54.7 |
| MD8952 | 2.5 | Non-Coding | Novel provisional Pre-miRNA Candidate | 53.5 |
| MR6677 | 2.5 | Non-Coding | Novel provisional Pre-miRNA Candidate | 52.3 |
| MR13051 | 2.5 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 51.1 |
| MD10606 | 2.5 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 50 |
| MD4873 | 2.5 | Non-Coding | Novel provisional Pre-miRNA Candidate | 48.8 |
| MD718 | 2.4 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 47.6 |
| MR9158 | 2.4 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 46.4 |
| MD9743 | 2.3 | Non-Coding | Novel provisional Pre-miRNA Candidate | 45.2 |
| MD9852 | 2.3 | Non-Coding | Novel provisional Pre-miRNA Candidate | 44 |
| MD1559 | 2.3 | Non-Coding | Novel provisional Pre-miRNA Candidate | 42.8 |

| Key | Description |
|---|---|
| | Previously deemed to be high probability pre-miRNA candidate following non-automated analysis of deep sequencing data. No corresponding signal in 26-35nt. |
| | Previously deemed to be low probability pre-miRNA candidate following non-automated analysis of deep sequencing data. Corresponding signal in 26-35nt. Most likely degradation product |
| | Previously deemed to be low probability pre-miRNA candidate following non-automated analysis of deep sequencing data. No corresponding signal in 26-35nt |
| | VMir Predicted Pre-miRNA. Not identified in deep sequencing. Corresponding signal in 26-35nt |
| | VMir Predicted Pre-miRNA. Not identified in deep sequencing. No corresponding signal in 26-35nt. New provisional pre-miRNA candidate |

**Table 4.21 (Part 2/2) Summary of all positive array signals from S1 that were not attributed to CCB transcripts or array surface artefacts**

| Name | RV | Genomic region | Description | Percentile Rank |
|---|---|---|---|---|
| MD695 | 2.2 | Protein-Coding | High signal in 26-35nt RNA | 41.6 |
| MD2718 | 2.2 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 40.4 |
| MD18142 | 2.2 | Non-Coding | Novel provisional Pre-miRNA Candidate | 39.2 |
| MR5075 | 2.2 | Non-Coding | High Probability Pre-miRNA from Deep Sequencing | 38 |
| MR8070 | 2.2 | Non-Coding | Novel provisional Pre-miRNA Candidate | 36.9 |
| MR1556 | 2.2 | Non-Coding | Novel provisional Pre-miRNA Candidate | 35.7 |
| MR8599 | 2.1 | Non-Coding | Novel provisional Pre-miRNA Candidate | 34.5 |
| MR12580 | 2.1 | Non-Coding | Novel provisional Pre-miRNA Candidate | 33.3 |
| MD665 | 2.1 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 32.1 |
| MR4453 | 2.1 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 30.9 |
| MR18968 | 2.0 | Non-Coding | Novel provisional Pre-miRNA Candidate | 29.7 |
| MR3949 | 2.0 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 28.5 |
| MR16865 | 2.0 | Non-Coding | Novel provisional Pre-miRNA Candidate | 27.3 |
| MR10689 | 2.0 | Non-Coding | Novel provisional Pre-miRNA Candidate | 26.1 |
| MD1111 | 2.0 | Non-Coding | High Probability Pre-miRNA from Deep Sequencing | 25 |
| MR8192 | 2.0 | Non-Coding | Novel provisional Pre-miRNA Candidate | 23.8 |
| MD18477 | 2.0 | Non-Coding | Novel provisional Pre-miRNA Candidate | 22.6 |
| MR1281 | 1.9 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 21.4 |
| MD7271 | 1.9 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 20.2 |
| MD16522 | 1.9 | Non-Coding | Novel provisional Pre-miRNA Candidate | 19 |
| MD5586 | 1.9 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 17.8 |
| MR8069 | 1.9 | Non-Coding | Novel provisional Pre-miRNA Candidate | 16.6 |
| MD5645 | 1.9 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 15.4 |
| MR17045 | 1.9 | Non-Coding | Novel provisional Pre-miRNA Candidate | 14.2 |
| MR3935 | 1.9 | Non-Coding | Novel provisional Pre-miRNA Candidate | 13 |
| MR4138 | 1.9 | Non-Coding | Novel provisional Pre-miRNA Candidate | 11.9 |
| MD5470 | 1.9 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 10.7 |
| MD710 | 1.9 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 9.5 |
| MD18849 | 1.8 | Non-Coding | Novel provisional Pre-miRNA Candidate | 8.3 |
| MD16282 | 1.8 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 7.1 |
| MD9246 | 1.8 | Non-Coding | Novel provisional Pre-miRNA Candidate | 5.9 |
| MD18702 | 1.8 | Non-Coding | Novel provisional Pre-miRNA Candidate | 4.7 |
| MD11829 | 1.8 | Non-Coding | Novel provisional Pre-miRNA Candidate | 3.5 |
| MD9506 | 1.8 | Non-Coding | Novel provisional Pre-miRNA Candidate | 2.3 |
| MD6507 | 1.8 | Non-Coding | Novel provisional Pre-miRNA Candidate | 1.1 |
| MD3906 | 1.8 | Protein-Coding | Novel provisional Pre-miRNA Candidate | 0 |

Of the 14 pre-miRNA candidates from sequencing that also gave positive results in array hybridization analysis of S1, seven of them feature in the top ten and twelve of them feature in the top twenty highest positive signals from array hybridization (Table 4.21). This indicated that most of them were not just over-represented in the deep sequencing data due to enzymatic bias. While many of the pre-miRNA candidates (both high and low probability) identified in deep sequencing gave strong signals, it is important to note that the signals from MR5075 and MD1111 were ranked 53$^{rd}$ and 64$^{th}$ respectively out of the 85 positives.

### 4.2.3 Identification and analysis of new provisional CyHV-3 pre-miRNA candidates from array hybridization

The remaining pre-miRNAs showing positive signals were analysed further. Out of the 71 signals, six of them were eliminated due to corresponding signals in S2. The other 65 pre-miRNAs were deemed to display CyHV-3 miRNA-like signals (i.e. positive in S1 and negative in S2 and S3) making them provisional pre-miRNA candidates. These candidates were investigated further by searching the sequencing data (from both infections) for evidence to support the existence of miRNAs derived from these pre-miRNAs i.e. transcripts with low read counts that were potentially under represented in the data due to enzymatic bias. Originally such transcripts would not have been considered as they were not mapped to predicted pre-miRNAs complying with Point 1 of the pre-miRNA identification criteria as applied during non-automated analysis (i.e. at least one small RNA from the pre-miRNA needed to have a read count >10, in at least 1 infection), however positive signals from probes targeting them suggested that their actual expression levels may have been higher. Alternatively, these transcripts may have been identified as putative miRNAs by the automatic methods, but if the same pre-miRNA candidate was not identified in both infections or by both algorithms, these candidates would have been ignored.

The sequencing data revealed that many of these 65 pre-miRNAs showing CyHV-3 miRNA-like signals did indeed have small RNAs mapping to them. However, other than non-compliance with Points 1 and 2 (must be from non-coding region), none of them displayed even a moderate compliance with the other points in the list of miRNA identification criteria, mainly showing non-compliance with Points 3 (not derived from stems), 4, (minor forms not present) and 9 (no miRNA-like signature).

### 4.2.4 Summary of array hybridization results

In summary, no new provisional pre-miRNA candidates identified by array analysis could be supported by corresponding low level signals from the sequencing data and therefore these were not investigated further. All four of the high probability pre-miRNA candidates (MR5057, MD11776, MD1111 and MR5075) that were detected in S1 showed CyHV-3 miRNA-like signals (no corresponding signals in S2 or S3).

The two high probability pre-miRNAs that showed no signals (MD9812 and MD11410) also displayed the lowest miRNA read counts in deep sequencing and therefore these miRNAs may have been beyond the limit of detection for array hybridization.

The fourteen pre-miRNA candidates identified in deep sequencing that were also the detected in array hybridization had RVs that were generally among the highest of all the positives. This indicated that even in the absence of enzymatic bias, the small RNAs derived from these pre-mRNA candidates were generally still among the most highly expressed transcripts detected in the S1 sample, regardless of whether they were from high or low probability candidates. This was largely in-line with the observations from the deep sequencing data, indicating they were not over represented in the sequencing data due to enzymatic bias. In the cases of the high probability pre-miRNA candidates MR5057 and MD11776 from deep sequencing, the results from array hybridization agreed well with the sequencing data and supported their annotation as genuine pre-miRNAs. The results for MD1111 and MR5075 on the array differed from the deep sequencing data, in terms of their expression levels and the dominant miRNA derived from them, implying the need to investigate this further by other methods.

The results from array hybridization supported the elimination of low probability pre-miRNA candidates from further analysis. This indicated the criteria used to distinguish between low and high probability pre-miRNA candidates in the analysis of the deep sequencing data was well founded. Therefore the remaining high probability pre-miRNA candidates were now referred to as “pre-miRNAs” and accordingly putative miRNAs derived from them were now referred to as miRNAs

## 4.3 Northern blotting

Northern blotting analysis was also carried out in order to support the annotation of high probability pre-miRNA candidates as genuine CyHV-3 pre-miRNAs. As with DNA array hybridization, probes were designed to target miRNAs derived from the stems of selected pre-miRNA candidates. Unlike previous experiments, total RNA was used thus allowing the detection of all transcripts containing the same target

sequences across all RNA size ranges. The sample in question was from a new *in vitro* lytic infection using the N076 CyHV-3 isolate (separate infection to that used in deep sequencing). The miRNAs initially targeted included the major forms (as indicated from deep sequencing and array hybridization) from pre-miRNA candidates MR5057 and MD11776 and both putative miRNAs from pre-miRNA candidate MD1111 (as it was unclear which miRNA was the major form). As northern blotting was not as sensitive as deep sequencing or array hybridization, it was only the most abundant putative miRNAs that were also detected in array hybridization that were selected for investigation by northern blotting, with a view to extending northern blotting analysis to additional targets if all initial targets were successfully detected.

No bands were detected when probes for MD11776 or MD1111 were used (data not shown). However the MR5057 probe did show positive hybridization signals (Figure 4.16). Two discrete transcripts were detected, one at ~22 nt and the other at ~56 nt. These sizes correspond to the sizes expected for the processed miRNA (22 nt) identified by deep sequencing (Section 4.1) and the MR5057 pre-miRNA (56 nt) inferred from VMir pre-miRNA structure prediction (Section 3) and the boundaries of the 5' and 3' arm miRNAs (Section 4.1). Importantly, the same signals were not seen in the total RNA from non-infected CCB cells. The detection of the host miRNA ccr-let-7a in lanes from both infected and non-infected cells confirmed equal loading and validated the negative result obtained with the non-infected sample. Higher molecular weight signals on the blots may be the result of probe binding to pri-mRNA transcripts.

This mature miRNA from MR5057 was shown to be the most highly expressed and most readily detectable CyHV-3 miRNA in the RNA deep sequencing and array hybridization experiments, and unsurprisingly it was the only CyHV-3 miRNA that was detectable by northern blotting Although present in the sample (as later confirmed by stem-loop RT-qPCR, data not shown) the levels of the miRNAs from pre-miRNA MD11776 and MD1111 were too low to be detected by northern blotting.

MR5057 occurs only 257 bp from MR5075 (based on 5' start positions) (Table 4.4), meaning that both pre-miRNAs may occur on the same pri-miRNA. Their miRNAs may be processed at different efficiencies (based on the differences in expression

levels) but their pre-miRNAs may be processed at similar rates. It was for this reason the same sample was also probed for the 3' miRNA from MR5075 to see if a pre-miRNA band could at least be detected, however no signals were detected for this target either (results not shown).

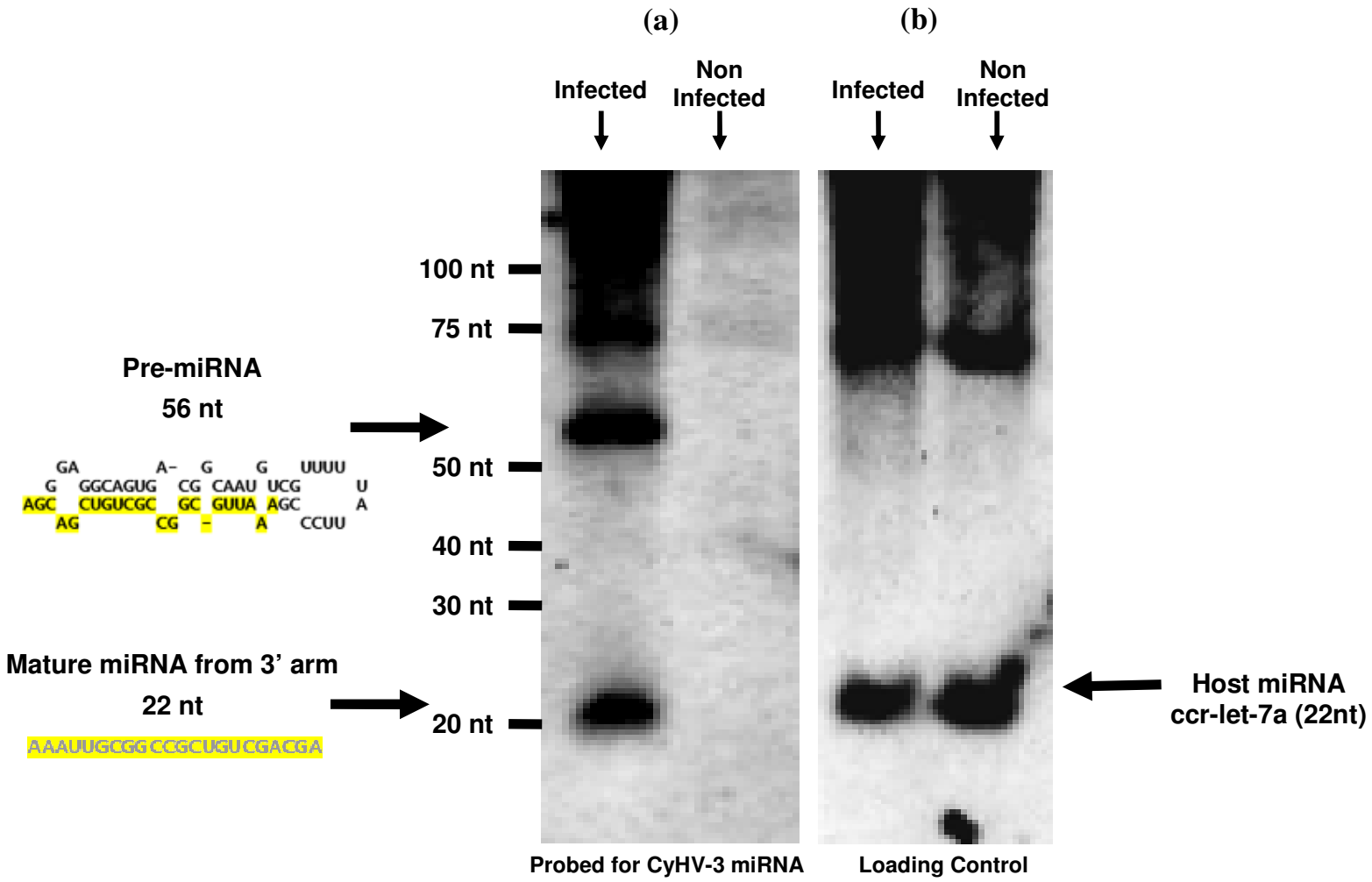



**Figure 4.16 (a) Detection of CyHV-3 pre-miRNA MR5057 (56 nt) and 3' arm mature miRNA (22 nt) (b) Loading control: detection of host miRNA ccr-let-7a (22 nt)**

# 5 Diagnostic assay development and application

Stem-loop RT-qPCR assays were developed in order to accurately and sensitively measure the relative expression levels of miRNAs (both high and low level) derived from all CyHV-3 pre-miRNAs. Stem-loop RT-qPCR is a TaqMan–based assay that is specific for mature miRNAs due to the use of a special type of RT primer known as a stem-loop RT primer (Chen et al., 2005). The process is explained in Figure 2.9 Assays were designed to detect the major forms (as determined by sequencing/array hybridization) from all 6 CyHV-3 pre-miRNAs. Assays were also designed to detect both miRNAs from MD1111 and MR5075 as it was not clear from previous experiments which miRNAs were in fact the major forms derived from these pre-miRNAs.

## 5.1 Stem Loop RT-qPCR assay testing and optimization

Before sample testing could begin, experimental parameters such as primer/probe concentration and input RNA had to be optimised to see how they affected each assay in terms of sensitivity and specificity. The strategy for stem-loop RT-qPCR optimization is outlined in Figure 4.14.

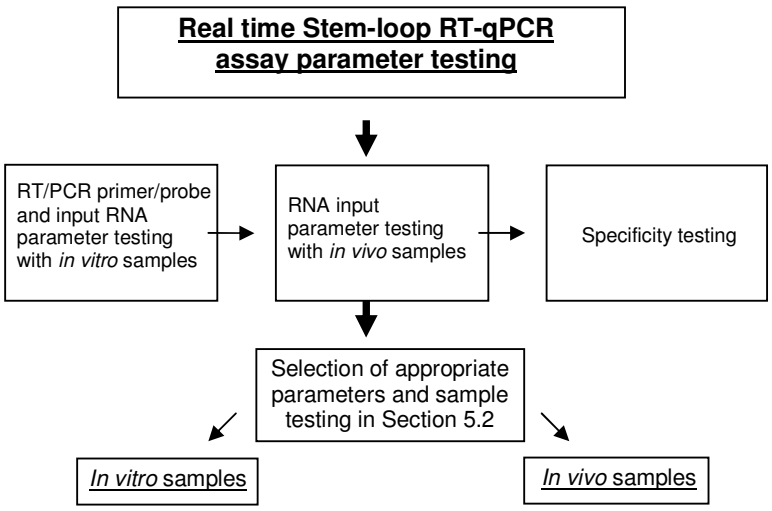



**Figure 5.1 Outline of strategy for stem-loop RT-qPCR parameter testing and sample testing**

### 5.1.1 Assay Parameter testing

Before the assays were used to test experimental samples, the primer and probe concentrations were varied in order to establish the most appropriate assay conditions for each sample type. An endogenous control assay for the carp miRNA let-7a was used as a proxy for all stem-loop RT-qPCR assays (just referred to as RT-qPCR from this point onwards).

#### 5.1.1.1 *Primer/probe and input RNA parameter testing with in vitro samples*

The performance of RT-qPCR assays was first tested using different final concentrations (0.1x and 1x) of RT primer and PCR Primer/Probe mix and different input quantities of total RNA from CCB cells (see combinations of parameters used with let-7a assay Table 5.1). The resulting Ct values corresponding to each different RT primer/PCR Primer-Probe and input RNA combination are shown in Table 5.1. These Ct values were compared using Technique 1 (Section 2.10.4.1) and these results are shown in Figure 5.2.

**Table 5.1 Parameter combinations used for let-7a RT-qPCR testing with *in vitro* samples**

| Primer Parameter set | RT-Primer | PCR-Primer- Probe | Input RNA | Result (Ct) |
|---|---|---|---|---|
| 1 | 1x | 1x | 1000 ng | 16.37 |
| 1 | 1x | 1x | 100 ng | 15.40 |
| 1 | 1x | 1x | 10 ng | 17.22 |
| 2 | 1x | 0.1x | 1000 ng | 18.77 |
| 2 | 1x | 0.1x | 100 ng | 17.92 |
| 2 | 1x | 0.1x | 10 ng | 19.52 |
| 3 | 0.1x | 1x | 1000 ng | 18.97 |
| 3 | 0.1x | 1x | 100 ng | 18.81 |
| 3 | 0.1x | 1x | 10 ng | 20.19 |
| 4 | 0.1x | 0.1x | 1000 ng | 21.42 |
| 4 | 0.1x | 0.1x | 100 ng | 20.97 |
| 4 | 0.1x | 0.1x | 10 ng | 22.54 |

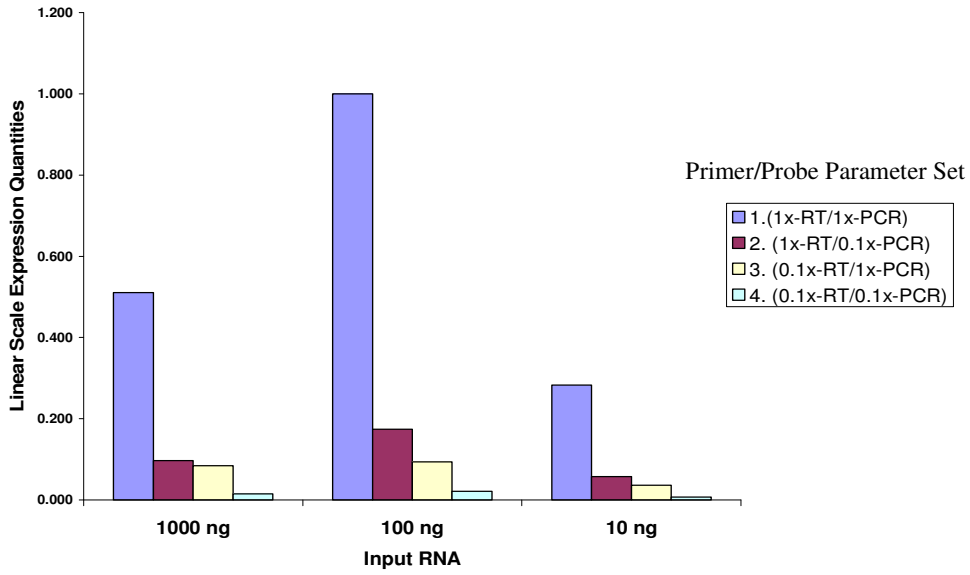



**Figure 5.2 Results for primer/probe and input RNA parameter testing with *in vitro* samples**

The strongest signal (i.e. lowest Ct value) was obtained using Primer/Probe parameter set 1 (1x RT primer/1x PCR primer-probe) with 100 ng of input total RNA (Figure 5.3). In addition, Primer/Probe parameter set 1 (1x RT primer/1x PCR primer-probe) performed best for all input RNA quantities. Conversely, Primer/Probe perameter set 4 (0.1x RT primer/0.1x PCR primer-probe) performed the worst for all input RNA quantities. In terms of input RNA quantity, the use of 100 ng input RNA gave the lowest Ct values with all primer parameter sets. Conversely, the use of 10 ng input RNA performed worst with all primer sets. Despite the fact that it performed the worst, the use of primer parameter set 4 and in combination with 10 ng input of RNA still gave a Ct of 22.54 (Table 5.1) indicating that these parameters were still more than adequate to detect the host miRNA ccr-let-7a and therefore also putative CyHV-3 miRNAs that are expressed at much lower levels.

#### 5.1.1.2 *Input RNA quantity testing with in vivo samples*

RNA input testing with *in vitro* samples indicated that 100 ng RNA consistently gave the best results. However, *in vivo* samples are very different to cell culture-derived samples and could possibly contain different amounts or types of inhibitors that could persist through RNA extraction, meaning that the RNA input quantities required for optimum results with *in vivo* samples could differ from that of *in vitro* samples. In

order to test this, 27 RNA samples from four different *Cyprinus carpio* tissues (brain, gill, gut and head-kidney) were tested for host let-7a. The RNA concentrations in these samples ranged from 29.3-4415.3 ng/mL and each sample was tested at neat, 1/10 and 1/100 concentrations. In order to save on consumables, this was tested using 0.1x RT primer/0.1x PCR primer-probe, which had already been shown to be more than adequate (Section 5.1.1.1). Sample details and the resulting Ct values obtained following testing are shown in Table 5.2. Samples that had RNA input quantities comparable to those tested with *in vitro* samples in Section 5.1.1.1 are highlighted in bold. These Ct values were compared using Technique 1 (Section 2.10.4.1) and the results of this comparison are shown in Figure 5.3

**Table 5.2 Let-7a RT-qPCR input RNA quantity testing using RNA derived from *in vivo* samples**
Samples highlighted in bold had RNA input quantities that were comparable to those tested with RNA derived from *in vitro* samples in Section 5.1.1.1

| Sample details | | | | Results (Ct) | | |
|---|---|---|---|---|---|---|
| **Sample ID** | **Details** | **260/280** | **Neat Input RNA** | **Neat** | **1/10** | **1/100** |
| **1** | **Fish21-1-HK** | **2** | **916.5 ng** | **22.34** | **21.71** | **23.64** |
| **2** | **Fish21-1-GILL** | **2** | **824.5 ng** | **22.50** | **21.61** | **24.21** |
| 3 | Fish21-1-BRAIN | 2 | 1431.9 ng | 20.17 | 19.47 | 22.57 |
| 4 | Fish21-1-GUT | 1.9 | 624 ng | 23.07 | 22.62 | 23.64 |
| 5 | Fish-4-1-HK | 1.8 | 3890.6 ng | 24.97 | 24.20 | 25.40 |
| 6 | Fish-4-1-GILL | 1.5 | 4415.3 ng | 24.28 | 25.93 | 26.40 |
| 7 | Fish-4-1-BRAIN | 1 | 541 ng | 21.05 | 20.99 | 23.24 |
| 8 | Fish-4-1-GUT | 1.5 | 29.3 ng | 24.26 | 27.41 | 30.91 |
| **9** | **Fish-15-1-HK** | **2** | **1148.7 ng** | **22.35** | **22.39** | **24.45** |
| 10 | Fish-15-1-GILL | 2 | 603.8 ng | 21.96 | 21.66 | 23.23 |
| **11** | **Fish-15-1-BRAIN** | **1.9** | **1203.4 ng** | **21.70** | **21.92** | **23.85** |
| 12 | Fish-15-1-GUT | 2 | 1132.4 ng | 23.49 | 22.42 | 23.97 |
| 13 | Fish-3-1-HK | 2 | 2296.7 ng | 22.69 | 22.54 | 24.47 |
| **14** | **Fish-3-1-GILL** | **2** | **1178.1 ng** | **23.66** | **23.67** | **23.93** |
| 15 | Fish-3-1-BRAIN | 2 | 447.1 ng | 20.94 | 21.29 | 23.10 |
| 16 | Fish-3-1-GUT | 2 | 2058 ng | 23.92 | 24.38 | 25.03 |
| 17 | Fish-16-1-HK | 1.7 | 4130 ng | 25.15 | 24.78 | 25.21 |
| 18 | Fish-16-1-GILL | 2 | 1456.2 ng | 24.27 | 24.56 | 24.96 |
| 19 | Fish-16-1-BRAIN | 2 | 3029.3 ng | 21.70 | 22.03 | 22.70 |
| 20 | Fish-16-1-GUT | 1.9 | 3710.9 ng | 25.28 | 25.75 | 26.02 |
| 21 | Fish11-1-HK | 1.9 | 3629.6 ng | 23.88 | 23.79 | 24.96 |
| 22 | Fish11-1-GILL | 1.7 | 152.2 ng | 28.08 | 26.57 | 35.93 |
| 23 | Fish11-1-BRAIN | 2 | 581.7 ng | 22.63 | 22.52 | 24.85 |
| 24 | Fish11-1-GUT | 2 | 1403.3 ng | 23.22 | 23.36 | 24.22 |
| 25 | Fish-8-1-HK | 1.8 | 3798 ng | 26.94 | 27.93 | 30.25 |
| **26** | **Fish-8-2-GILL** | **1.9** | **1193.5 ng** | **27.47** | **22.73** | **26.36** |
| **27** | **Fish-8-3-Brain** | **1.9** | **1066.7 ng** | **23.07** | **26.76** | **22.54** |

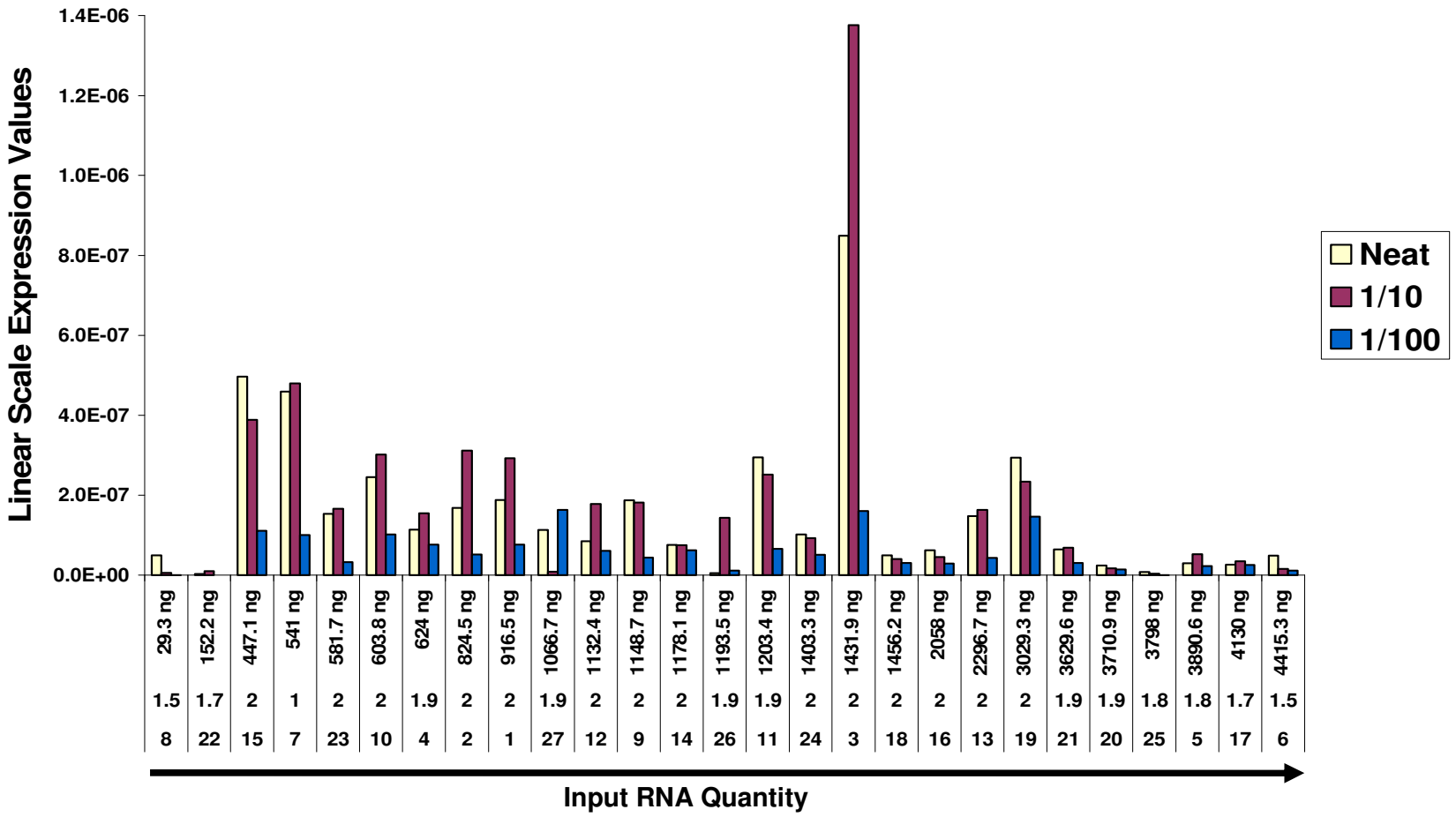



**Figure 5.3 Let-7a RT-qPCR input RNA quantity testing with total RNA prepared from *in vivo* samples**
RNA samples originating from brain tissue and Fish-21 gave higher signals on average and are marked with * and x respectively

It was evident from results in Figure 5.3 that there was no clear relationship between the neat input RNA quantity and assay performance, nor was there a correlation to A260/280 values. This suggested that the performance of the assay with each individual sample may be dependent on other factors intrinsic to each individual sample such as inhibitors or a matrix effect. On this note there was a correlation between Ct value and tissue type. Samples from brain tissue gave lower Ct signals on average, compared to other tissues (marked with * in Figure 5.3). In addition organs from Fish-21 (marked with x in Figure 5.3) gave higher signals, on average, than samples from other fish. The lowest signal obtained for each sample was always the 1/100 dilution (with the exception of sample 27). The strongest signal for each sample was obtained from the Neat or the 1/10 dilution and generally there was not much difference between them (Figure 5.3). This was true even for samples that had RNA input quantities that were comparable to the quantities used *in vitro* (section 5.1.1.1; marked in bold in Table 5.2) As these were the same samples that were to be tested for CyHV-3 miRNAs, these results suggested that it was best to test both Neat and 1/10 dilutions when carrying out this testing (Section 5.2.2).

#### **5.1.1.3** ***Specificity testing for CyHV-3 miRNA RT-qPCR assays***

Higher primer levels were seen to result in greater sensitivity (Section 5.1.1.1). Lower concentrations such as 0.1x RT primer/0.1x PCR primer-probe were also demonstrated to be more than adequate for detecting host (Section 5.1.1.1 ) and CyHV-3 miRNAs *in vitro* (initial testing for CyHV-3 miRNA assays not shown, see other results later in Section 5.2.1). However, it was assumed that CyHV-3 target transcripts would be present at lower levels in host tissue, and that therefore it was preferable to use 1x RT primer/1x PCR primer-probe when testing *in vivo* samples. If the levels of target transcript were low in tissues, very high Ct values would be expected (~ 35). Even though it results in more sensitive assays, the use of high primer concentrations increases the chances of low-level non-specific signals that would also result in high Ct values (~35). In order to identify assays that have a tendency (due to their primer sequences) to give low non-specific signals when using high primer concentrations, all CyHV-3 miRNA assays were first tested on RNA from non-infected CCB cells using high quantities of input RNA (1000 ng) and primer-

probe (1x). The same tests were also carried out using low amounts of RT primer/1x PCR primer-probe (0.1x), which were less likely to result in non-specific signals. The results from these specificity tests are to be seen in Table 5.3.

**Table 5.3 Ct values for CyHV-3 miRNA assay specificity testing (using high and low primer concentrations)**
The assays that showed non-specific signals at higher primer concentrations are highlighted

| Assay | 0.1x-RT/0.1x-PCR (Low) | | 1x-RT/1x-PCR (High) | |
|---|---|---|---|---|
| | Replicate 1 | Replicate 2 | Replicate 1 | Replicate 2 |
| MR5057 3' | Negative | Negative | Negative | Negative |
| MD11776 3' | Negative | Negative | Negative | Negative |
| MD1111 3' | Negative | Negative | 36.13 | 36.57 |
| MD1111 5' | Negative | Negative | 35.27 | 35.98 |
| MR5075 3' | Negative | Negative | Negative | Negative |
| MR5075 5' | Negative | Negative | Negative | Negative |
| MD11410 5' | Negative | Negative | 35.32 | 35.79 |
| MD9812 5' | Negative | Negative | Negative | 36.12 |

None of the CyHV-3 miRNA assays gave non-specific signals when 0.1x RT primer/0.1x PCR primer-probe (low concentration) was used. When 1x RT primer/1x PCR primer-probe (high concentration) was used, the assays for MD1111 3', MD11776, MD11410 and MD9812 showed low-level signals (~35). This may have been because some of the primers for these assays have higher GC content and therefore have higher affinities for both specific and non-specific targets during the annealing step. In order to quickly test this theory, RNA from non-infected CCB cells was tested again using 1x RT primer/1x PCR primer-probe (high concentration) at 60°C (same as before), 65°C and 70°C. As a positive control for each annealing temperature RNA from CyHV-3 infected cells was also tested. The inclusion of positive controls also provided a way to test the effects of higher annealing temperatures on the sensitivity of these assays, which could be used to investigate the feasibility of running all assays at high annealing temperatures to avoid non-specific signals at higher primer concentrations (the performance of the assay for the host miRNA was also tested in this regard). The results of these tests are to be seen in Table 5.4.

**Table 5.4 Ct values for CyHV-3 miRNA assay specificity testing at higher annealing temperatures**
The CyHV-3 assays that showed non-specific signals in non-infected CCB RNA are highlighted

| Sample | | Infected | | | Non-infected | | |
|---|---|---|---|---|---|---|---|
| Annealing Temperature | | 60℃ | 65℃ | 70℃ | 60℃ | 65℃ | 70℃ |
| Host Let-7a* | Replicate 1 | 25.93 | Negative | Negative | 24.13 | Negative | Negative |
| | Replicate 2 | 25.44 | Negative | Negative | 24.33 | Negative | Negative |
| MR5057 | Replicate 1 | 15.87 | 21.2 | 38.37 | Negative | Negative | Negative |
| | Replicate 2 | 18.44 | 21.98 | 36.99 | Negative | Negative | Negative |
| MD11776 | Replicate 1 | 26.97 | 28.97 | Negative | Negative | Negative | Negative |
| | Replicate 2 | 23.36 | 27.77 | Negative | Negative | Negative | Negative |
| MD1111 3' | Replicate 1 | 22.74 | 28.86 | Negative | 32.55 | Negative | Negative |
| | Replicate 2 | 22.43 | 28.99 | Negative | 28.34 | Negative | Negative |
| MD1111 5' | Replicate 1 | 25.52 | 22.52 | 33.17 | Negative | 38.33 | Negative |
| | Replicate 2 | 25.82 | 22.48 | 34.97 | 35.65 | 35.86 | Negative |
| MR5075 5' | Replicate 1 | 25.70 | 35.12 | Negative | 36.30 | Negative | Negative |
| | Replicate 2 | 26.61 | 34.98 | Negative | Negative | Negative | Negative |
| MR5075 3' | Replicate 1 | 27.72 | 34.98 | Negative | Negative | Negative | Negative |
| | Replicate 2 | 25.26 | 32.26 | Negative | Negative | Negative | Negative |
| MD9812 | Replicate 1 | 26.88 | 32.91 | Negative | 33.86 | Negative | Negative |
| | Replicate 2 | 26.73 | 32.97 | Negative | 35.12 | Negative | Negative |
| MD11410 | Replicate 1 | 25.96 | 26.46 | 37.17 | 35.72 | Negative | Negative |
| | Replicate 2 | 25.91 | 26.43 | 37.00 | 36.05 | 37.04 | Negative |

.* Host Let-7a* was test using 0.1x primer/probe.

As with the previous round of testing (Table 5.3) the same four assays gave low levels of non-specific signals in non-infected RNA at 60°C, although in this experiment the assay for MR5075 5' also showed similar signals (in one replicate). Using the RNA from non-infected cells, only two of these assays (MD11115' and MD11410) still showed positive signals at 65°C and none showed positive signals at 70°C. This indicated that the primer/probe sets for MD1111 5' and MD11410 may have had the strongest affinity to non-specific targets. All of the CyHV-3 miRNA targets could be detected in the RNA from infected CCB cells at 60°C. All of them could also be detected at 65°C although the sensitively was reduced, with the exception of the assay for MD1111 5' which actually improved significantly. The assay for the host miRNA gave no signals at this temperature. At 70°C only the assays for MD11115' and MD11410 and MR5057 gave positive signals, although in all cases the sensitively was

reduced. This further supports the idea that the primers for MD1111 5' and MD11410 have higher affinities for both their specific and non-specific targets.

In conclusion, the use of 1x RT primer/1x PCR primer-probe (high concentration) at 60°C gave the best sensitivity with all assays. Therefore these are the most suitable parameters to use for testing RNA *in vivo* samples where the CyHV-3 miRNA levels may be very low. However, the same parameters were also most likely to give non-specific signals with five of these assays (MD1111 3', MD1111 5', MD11410, MD9812 and MR5075 5'). The same non-specific signals are likely to be seen in RNA from tissue samples. If so, such low non-specific signals will be indistinguishable from low-level signals from CyHV-3 miRNAs present in the same tissues. Thus, due to a tendency to produce non-specific signals these five assays may not be suitable for testing for CyHV-3 miRNAs *in vivo*. In contrast, the assays for MR5057, MD11176 and MR5075 3' do not appear to suffer from these problems. These non-specific signals can be easily eliminated by using lower primer concentrations (0.1x RT primer/0.1x PCR primer-probe). This does reduce the overall assay sensitivity but it is more than adequate for testing samples known to have high levels of target transcript i.e. during *in vitro* lytic infections. Thus, all of these assays could be used to detect and compare levels of CyHV-3 miRNA *in vitro* without any specificity problems.

In summary, it was found that it was best to use low primer concentrations (0.1x RT primer/0.1x PCR primer-probe) for all assays when testing *in vitro* (cell culture-derived) samples and high primer concentrations (1x RT primer/1x PCR primer-probe) when testing *in vivo* samples, in which case, the assays for MR5057, MD11176 and MR5075 are the most suitable to use.

## 5.2 Sample Testing

The RT-qPCR assays were first used to verify the relative CyHV-3 miRNA expression levels that were observed in the deep sequencing data (which was measured as read counts) by using them to re-test one of the RNA samples used (from the H361 infection). They were also ideal for comparing the levels observed in deep sequencing to the levels observed over the course of a second *in vitro* lytic infection (separate to N076 infection used for deep sequencing and northern blotting) in order to establish if the same relative expression levels are typical of this kind of *in vitro* CyHV-3 infection. Lastly RT-qPCR provided most suitable way to test for the same miRNAs *in vivo*. The strategy used for the testing of samples is outlined in Figure 5.4.

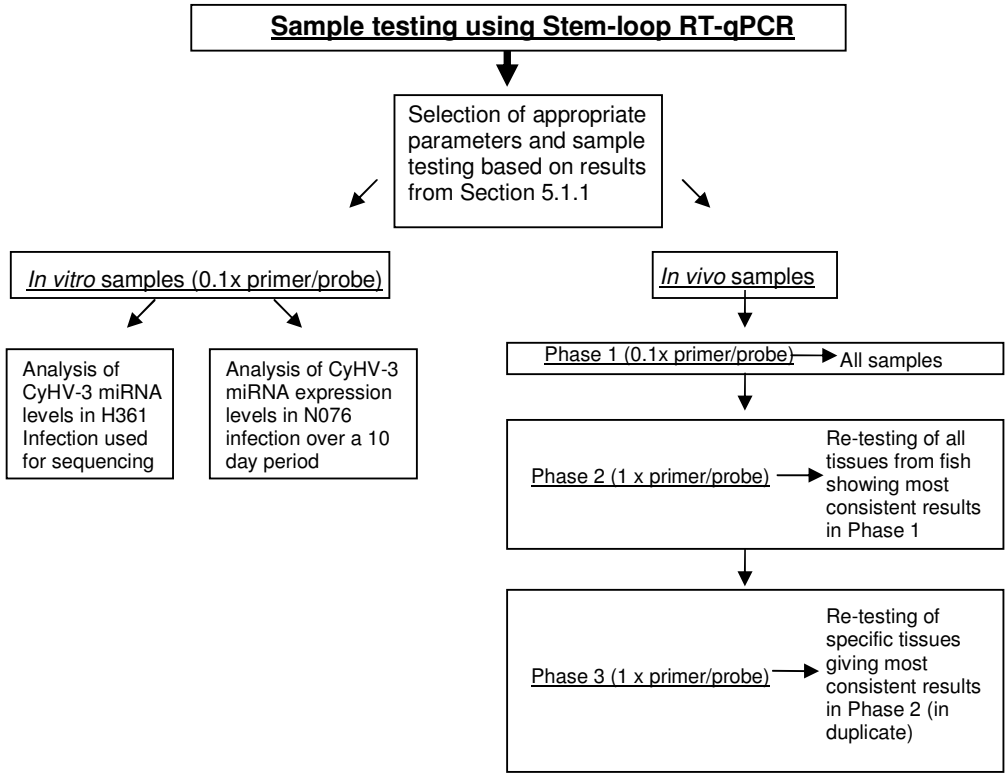



**Figure 5.4 Outline of strategy used to test samples for CyHV-3 miRNA expression**

### 5.2.1 Testing of RNA from infected cells (*in vitro* infections)

All *in vitro* samples were tested by RT-qPCR using 0.1x RT-Primer/0.1x PCR primer-probe concentrations.

#### 5.2.1.1 *RT-qPCR analysis of relative CyHV-3 miRNA levels in the RNA used for deep sequencing (H361 infection)*

A replicate aliquot of the same RNA sample used for deep sequencing analysis (H361 *in vitro* infection) was tested for all CyHV-3 miRNAs. One thousand nanograms of total RNA was used in one multiplex RT reaction (containing RT primers for all CyHV-3 miRNA targets), followed by separate real-time PCRs for each target. As different PCRs could display different PCR efficiencies, the relative levels of each target transcript were calculated using Technique 2 (Section 2.10.4.2) and then compared to each other (see Table 5.5 and Figure 5.5 (a)). These were also compared to the read count levels that were obtained in the RNA deep sequencing experiment from the same sample (Table 5.5 and Figure 5.5(b)).

**Table 5.5 RT-qPCR measurement of CyHV-3 miRNA levels in RNA prepared following CyHV-3 H361 infection of CCB cells used previously for small RNA deep sequencing**
The corresponding read counts obtained from the same sample in deep sequencing are also included. All of these levels refer to the major form miRNA from each CyHV-3 pre-miRNA unless otherwise stated.

| Target | Replicate | Ct Value | Efficiency | $N_0$ | Mean $N_0$ | Read counts |
|---|---|---|---|---|---|---|
| **MR5057** | 1 | 19.84 | 1.99 | 2.14E-08 | 2.39E-08 | 62596 |
| | 2 | 20.16 | | 2.65E-08 | | |
| **MD11776** | 1 | 24.78 | 1.93 | 4.51E-09 | 4.80E-09 | 2114 |
| | 2 | 24.97 | | 5.09E-09 | | |
| **MD1111 5'** | 1 | 24.67 | 2.01 | 1.55E-09 | 1.53E-09 | 567 |
| | 2 | 24.30 | | 1.52E-09 | | |
| **MD1111 3'** | 1 | 24.75 | 1.94 | 1.74E-09 | 2.00E-09 | 2711 |
| | 2 | 24.72 | | 2.25E-09 | | |
| **MR5075 5'** | 1 | 25.07 | 1.91 | 1.05E-09 | 1.08E-09 | 12 |
| | 2 | 24.85 | | 1.11E-09 | | |
| **MR5075 3'** | 1 | 27.92 | 1.87 | 2.87E-09 | 3.09E-09 | 125 |
| | 2 | 27.82 | | 3.31E-09 | | |
| **MD9812** | 1 | 27.85 | 1.96 | 1.70E-09 | 1.60E-09 | 23 |
| | 2 | 28.00 | | 1.49E-09 | | |
| **MD11410** | 1 | 28.58 | 1.92 | 3.64E-10 | 3.47E-10 | 10 |
| | 2 | 28.38 | | 3.30E-10 | | |

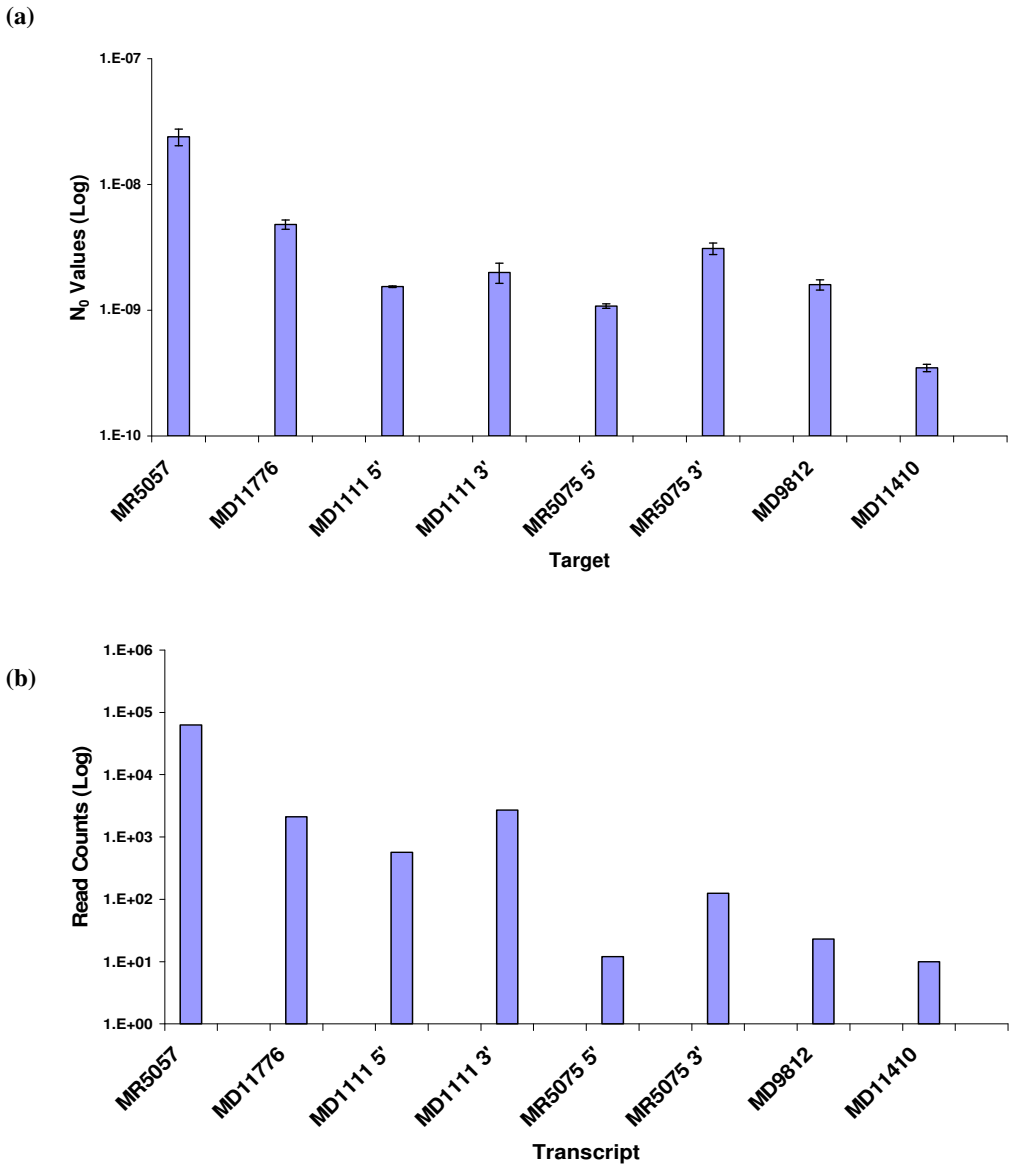



**Figure 5.5 Relative levels of CyHV-3 miRNAs in RNA from the H361 infection**
Relative expression levels as measured by (a) stem loop RT-qPCR (error bars represent standard deviation) and (b) deep sequencing (read counts). Unless otherwise stated, all of these levels refer to the major form miRNA from each CyHV-3 pre-miRNA

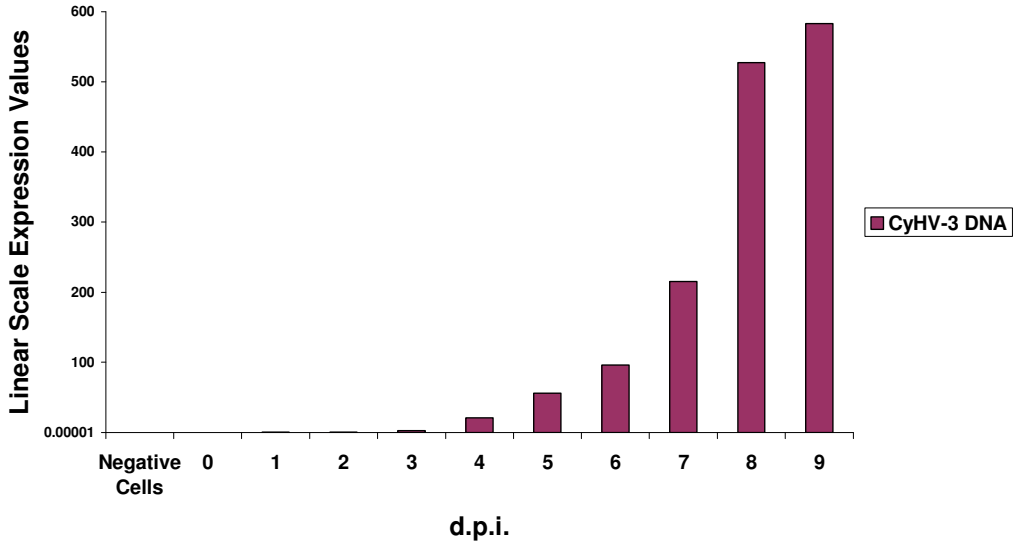
It can be seen that the low concentration of primer/probe (0.1x) used was sufficient to detect all target transcripts, in fact the highest Ct value was 28.58 (Table 5.5). As with all previous experiments, the 3' major form derived from MR5057 was the most highly expressed transcript present (Figure 5.5 (a)). The relative levels of all other

transcripts targeted generally correlated well with the corresponding read counts obtained following deep sequencing of the same sample (Figure 5.5 (a) and (b)). The most notable difference between the two experiments is the levels of both miRNAs from the pre-miRNA MD1111. The RT-qPCR results indicated that both miRNAs were actually present in much lower levels relative to the other targets, indicating that it may have been over-represented in deep sequencing data (Figure 5.5 (a) and (b)). Interestingly, contrary to the results from DNA microarray hybridization, the RT-qPCR results are in general agreement with the deep sequencing results in relation to the most abundant miRNAs from MD1111 and MR5075, indicating that the miRNA from the 3' arms of both of these pre-miRNAs are indeed the major forms.

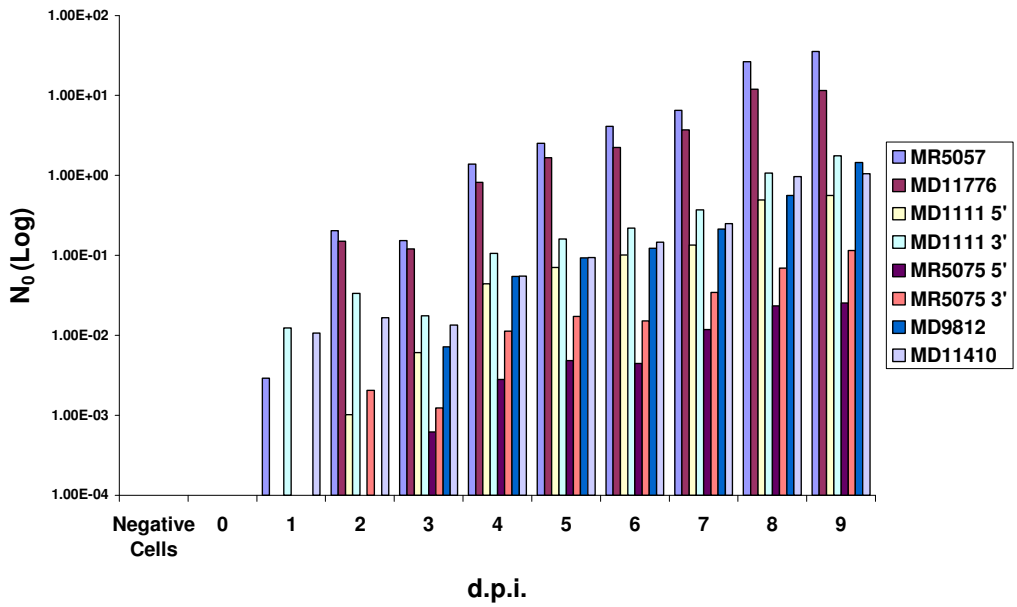
#### 5.2.1.2 *CyHV-3 miRNA expression over the course of an in vitro infection*

In order to show that the positive signals observed following CyHV-3 infection of CCB cells (Section 5.2.1.1) were indeed products of productive viral replication, the levels of the CyHV-3 viral miRNAs were measured over the course of an *in vitro* lytic infection, from the point of inoculation to the development of widespread CPE. CyHV-3 genomic DNA levels were also measured in the same cells over the same period. To do this, 11 replicate 75 $cm^2$ cell culture flasks were prepared by seeding them from a common trypsinized CCB cell culture suspension. Ten of these were inoculated with CyHV-3 (N076 isolate) and a lytic infection was established. The remaining non-inoculated flask was kept under the same conditions, serving as a negative control. One flask was taken for total RNA extraction every 24 hours. Relative levels were calculated using Technique 2 and normalized to host miRNA ccr-let-7a. Details, of the Ct values, efficiencies and other values used to calculate $N_0$ values are available in Supplementary File 5.1 T.1. CyHV-3 DNA levels from this experiment are shown in Figure 5.6 (a) and all CyHV-3 miRNA levels are to be seen in Figure 5.6 (b). The results for specific groups of targets are also compared (using linear scales) in Figure 5.6 (c)-(f).

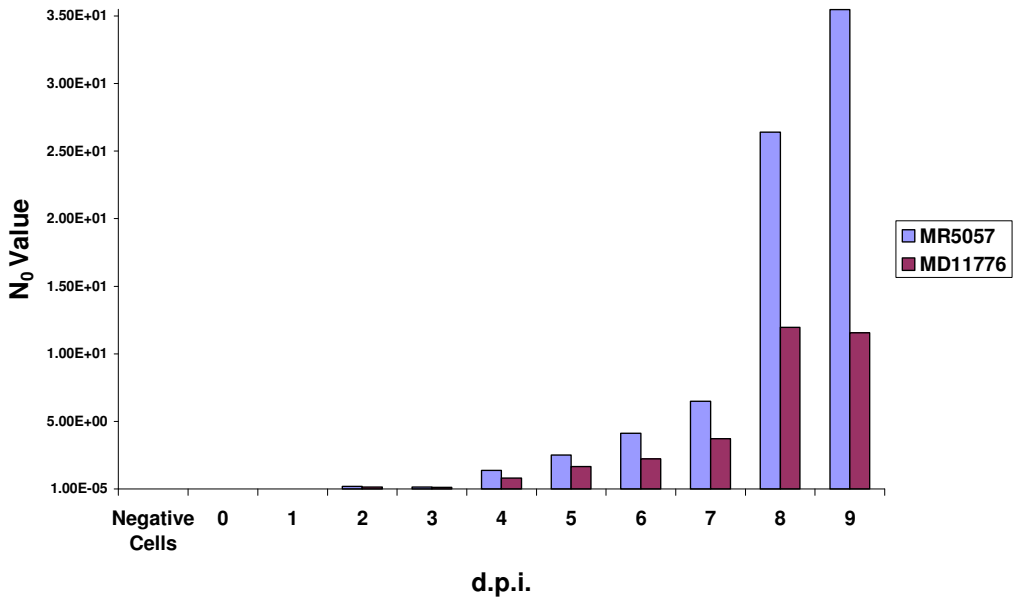
(a)


Linear Scale Expression Values
600
500
400
300
200
100
0.00001
Negative Cells
0
1
2
3
4
5
6
7
8
9
d.p.i.
CyHV-3 DNA

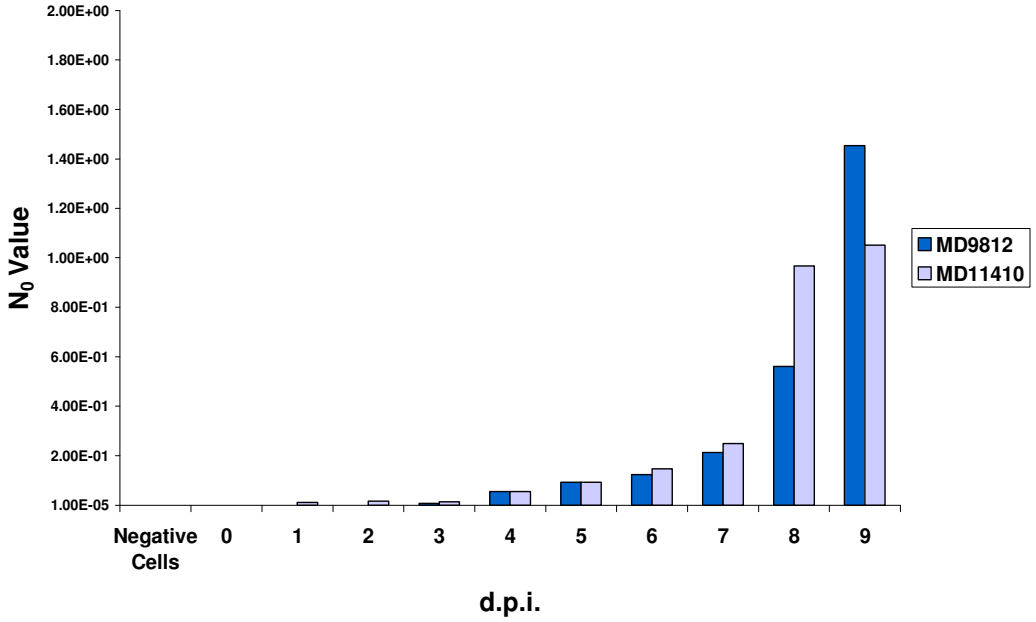
(b)


N0(Log)
1.00E+02
1.00E+01
1.00E+00
1.00E-01
1.00E-02
1.00E-03
1.00E-04
Negative Cells
0
1
2
3
4
5
6
7
8
9
d.p.i.
MR5057
MD11776
MD1111 5'
MD1111 3'
MR5075 5'
MR5075 3'
MD9812
MD11410

(c)


3.50E+01
3.00E+01
2.50E+01
2.00E+01
1.50E+01
1.00E+01
5.00E+00
1.00E-05
N0 Value
MR5057
MD11776
Negative Cells
0
1
2
3
4
5
6
7
8
9
d.p.i.

(d)


2.00E+00
1.80E+00
1.60E+00
1.40E+00
1.20E+00
1.00E+00
8.00E-01
6.00E-01
4.00E-01
2.00E-01
1.00E-05
N0 Value
MD9812
MD11410
Negative Cells
0
1
2
3
4
5
6
7
8
9
d.p.i.

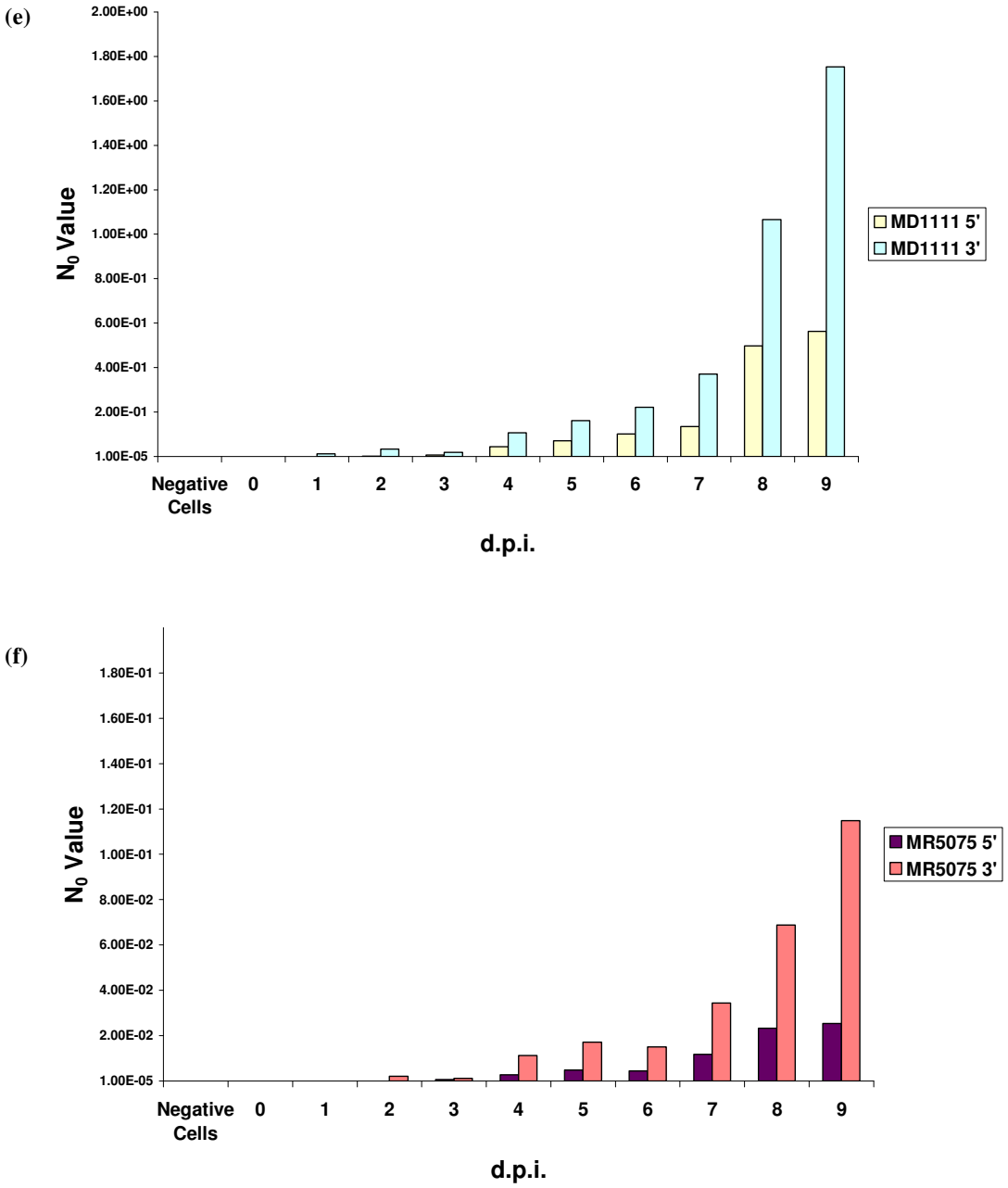



**Figure 5.6 Relative levels of CyHV-3 miRNA levels over the course of lytic infection *in vitro***
(a) CyHV-3 DNA levels normalized to CCB cell DNA (b) Relative expression levels of CyHV-3 miRNA targets (log scale) normalized to host miRNA ccr-let-7a. (c)- (f) Comparative expression levels of specific pairings of CyHV-3 miRNA targets (linear scales).

The establishment of a CyHV-3 lytic infection resulted in the development of widespread CPE at 8-9 d.p.i. (data not shown). The levels of CyHV-3 DNA progressively increased from 0-9 d.p.i., indicating that CyHV-3 replication was occurring (Figure 5.6 (a)). During the same period there was also a progressive increase in the levels of all CyHV-3 miRNAs (Figure 5.6 (b)).

Only three CyHV-3 miRNAs were detectable at 1 d.p.i. (MR5057, MD1111 3' and MD11410). All targets could be detected by 2 d.p.i. (although not all were visible with the scale used in Figure 5.6 (b)). At all stages from 2 d.p.i. onwards, MR5057 was consistently the most abundant target followed by MD11776 and MD1111 3'. Notably, both miRNAs from the MR5075 were present at lower levels throughout the infection relative to other targets (Figure 5.6 (b)) compared to in the H361 sample (Figure 5.5 (a)).

The difference in expression levels between MR5057 and the next highest target MD11776 increased dramatically at 8 and 9 d.p.i. (Figure 5.6 (c)) this corresponded to a similar increase in viral DNA levels at the same time point (Figure 5.6 (a)). The same trend was observed in the expression levels of all other targets (Figure 5.6 (d)-(f)) although not as pronounced. In fact, MD11410 only marginally increased from 8-9 d.p.i. compared to all other targets (Figure 5.6 (d)). The miRNA from the 3' arms of MD1111 and MR5075 were always present at higher levels than the miRNAs from the 5' arms of these pre-miRNAs (Figure 5.6 (e)-(f)) providing conclusive evidence that these miRNAs must be the major forms derived these pre-miRNAs.

### 5.2.2 Testing of tissues from CyHV-3-infected fish (*in vivo* infections)

#### 5.2.2.1 *Samples tested*

Samples were sourced and prepared by our collaborators at Cefas Laboratories, Weymouth, UK.

**RNA from tissues of a lytically infected fish**

This fish was originally a survivor of a high mortality CyHV-3 outbreak in a UK lake in the summer of 2006. A lytic infection was subsequently reactivated when the fish was kept at a permissive temperature. This indicated that other long-term survivors from the same outbreak were also likely to be latent carriers. Three samples were taken from this fish, these were gill, brain and head-kidney. It was possible that the same miRNAs expressed during the *in vitro* CyHV-3 lytic infection were also present in these samples. Hence this fish acted as a positive control for the experiment and was referred to as Fish-8.

**RNA from tissues of six healthy survivors of a CyHV-3 outbreak**

These six fish were survivors of the same high-mortality CyHV-3 outbreak as Fish-8. Hence, it was possible that these fish were also latent carriers. Unlike Fish-8, lytic infections were not reactivated in these fish, thus it was possible that they were latently infected at the time of sampling. Four samples were taken from each fish, these included gill, brain, gut and head-kidney. These six fish were referred to as Fish-4, -3, -15, -16 -11 and -21.

**Controls**

Positive RT-qPCR controls consisted of RNA from CyHV-3 *in vitro* lytic infections. Negative tissue controls consisted of RNA extracted from gill, brain, gut and head-kidney of CyHV-3 free carp sampled in lakes within the Republic of Ireland.

#### 5.2.2.2 *Phase 1: Preliminary Screening of all samples*

All samples were initially tested using 0.1x RT-Primer/0.1x PCR primer-probe concentrations in order to see if such RT-qPCR parameters were sufficient to detect CyHV-3 miRNAs. Samples were tested at three concentrations: Neat, 1/10 and 1/100 (the neat concentrations of all samples are given in Table 5.2). This also acted as a low-cost way to screen samples, allowing tentative selection of those most likely to give positive results, and thus appropriate for retesting using 1x RT-Primer/PCR primer-probe concentrations. In this phase, samples were only tested for 3 viral targets MR5057, MD11776 and MD1111 3' (Table 5.6).

**Table 5.6 Phase 1: Testing of all *in vivo* samples using 0.1x RT-Primer/0.1x PCR primer-probe concentrations**

| Sample | ID | Description | Let-7a | | | MD1111 3' | | | MR5057 | | | MD11776 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1/100 | 1/10 | Neat | 1/100 | 1/10 | Neat | 1/100 | 1/10 | Neat | 1/100 | 1/10 | Neat |
| Healthy Long Term Survivors | 1 | Fish-21-1-HK | 23.64 | 21.71 | 22.34 | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| | 2 | Fish-21-1-GILL | 24.21 | 21.61 | 22.50 | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| | 3 | Fish-21-1-BRAIN | 22.57 | 19.47 | 20.17 | Negative | **39.10** | **35.63** | Negative | **36.98** | Negative | Negative | **36.94** | Negative |
| | 4 | Fish-21-1-GUT | 23.64 | 22.62 | 23.07 | Negative | Negative | Negative | Negative | **37.48** | Negative | Negative | Negative | Negative |
| | 5 | Fish-4-1-HK | 25.40 | 24.20 | 24.97 | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| | 6 | Fish-4-1-GILL | 26.40 | 25.93 | 24.28 | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| | 7 | Fish-4-1-BRAIN | 23.24 | 20.99 | 21.05 | Negative | Negative | **35.98** | Negative | Negative | Negative | Negative | Negative | Negative |
| | 8 | Fish-4-1-GUT | 30.91 | 27.41 | 24.26 | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| | 9 | Fish-15-1-HK | 24.45 | 22.39 | 22.35 | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| | 10 | Fish-15-1-GILL | 23.23 | 21.66 | 21.96 | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| | 11 | Fish-15-1-BRAIN | 23.85 | 21.92 | 21.70 | Negative | Negative | **34.95** | Negative | Negative | Negative | Negative | Negative | Negative |
| | 12 | Fish-15-1-GUT | 23.97 | 22.42 | 23.49 | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| | 13 | Fish-3-1-HK | 24.47 | 22.54 | 22.69 | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| | 14 | Fish-3-1-GILL | 23.93 | 23.67 | 23.66 | Negative | Negative | **38.07** | Negative | Negative | Negative | Negative | Negative | Negative |
| | 15 | Fish-3-1-BRAIN | 23.10 | 21.29 | 20.94 | Negative | **36.98** | **36.14** | Negative | Negative | Negative | Negative | Negative | Negative |
| | 16 | Fish-3-1-GUT | 25.03 | 24.38 | 23.92 | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| | 17 | Fish-16-1-HK | 25.21 | 24.78 | 25.15 | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| | 18 | Fish-16-1-GILL | 24.96 | 24.56 | 24.27 | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| | 19 | Fish-16-1-BRAIN | 22.70 | 22.03 | 21.70 | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| | 20 | Fish-16-1-GUT | 26.02 | 25.75 | 25.28 | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| | 21 | Fish-11-1-HK | 24.96 | 23.79 | 23.88 | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| | 22 | Fish-11-1-GILL | 35.93 | 26.57 | 28.08 | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| | 23 | Fish-11-1-BRAIN | 24.85 | 22.52 | 22.63 | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| | 24 | Fish-11-1-GUT | 24.22 | 23.36 | 23.22 | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| Fish with Reactivated Lytic Infection | 25 | Fish-8-1-HK | 30.25 | 27.93 | 26.94 | Negative | Negative | Negative | **35.96** | Negative | Negative | **36.95** | Negative | Negative |
| | 26 | Fish-8-2-GILL | 26.36 | 22.73 | 27.47 | Negative | **35.60** | Negative | Negative | Negative | Negative | Negative | Negative | Negative |
| | 27 | Fish-8-3-Brain | 22.54 | 26.76 | 23.07 | **36.95** | Negative | **39.95** | Negative | **35.20** | Negative | Negative | Negative | Negative |

| Key | Description |
|---|---|
| | Host miRNA assay |
| | Using 1x PCR primer-probe, this assay gave low non-specific signals in non-infected CCB cells (Section 5.1) thus it was likely to also give low-level false positives with tissue samples |
| | Using 1x primer-probe this assay gave ***no*** low non-specific signals in non-infected CCB cells (Section 5.1) |

The positive RT-qPCR and negative tissue controls gave the expected results. The assay for MD1111 3' yielded more positive signals than MR5057 and MD11776 (Table 5.6). Positive test readings for MD1111 3' were distributed across samples from several different fish, whereas the positives for MR5057 and MD11776 were confined to Fish 21 (a possible latent carrier) and Fish 8 (the lytically infected fish). Due to the fact that the assay for MD1111 was previously shown to have a tendency to give low-level non-specific signals (even if only demonstrated using 1x primer concentrations with *in vivo* samples in Section 5.1.1.3) and because there was no discernable pattern to the positives (unlike MR5057 and MD11776), the results from MD1111 were ignored. The positive results for MR5057 and MD11776 suggested that Fish 21 and Fish 8 (Fish-8 was expected to be positive) were the two fish most likely to also test positive when retested using higher RT-qPCR primer concentrations in Phase 2, thus all tissues from these fish were selected for further testing.

#### **5.2.2.3** ***Phase 2: Re-testing of individual fish identified as most likely to give positive results***

All organs from individual fish showing the most consistent results in Phase 1 were re-tested using the 1x RT-Primer/1x PCR primer-probe formulation which had earlier been shown give a more sensitive test. In this phase of testing, all assays were used in order to confirm that they (i.e. assays for MD1111 5', MD1111 3' MR5075 5', MD9812 and MD11410) were not suitable for use due to a tendency to low give non-specific signals when 1x RT-Primer/1x PCR primer-probe was used.

The positive RT-qPCR and negative tissue controls gave the expected results. The majority of samples tested using these five assays gave low-level positive test results (66/70) (Table 5.7). This was in stark contrast to the assays for MR5057, MD11776 and MR5075 3' (Table 5.7). It is highly likely that the positive results for MD1111 5', MD1111 3' MR5075 5', MD9812 and MD11410 were indeed false positives. Although there was no way to confirm this, it was decided that it was best to ignore these results.

Assays for MR5057, MD11776 and MR5075 3' gave positive test results with tissue samples 3, 4, 26 and 27 (Table 5.7). Although the Ct values were high, they were

significantly lower than any signals from Phase 1 testing. Thus, these samples were selected for re-testing in Phase 3.

**Table 5.7 Phase 2: Testing of samples from fish that were identified in Phase 1 as being most likely to give positive results using 1x RT-Primer/1x PCR primer-probe concentrations**

| | | | Let-7a | | MD1111 5' | | MD1111 3' | | MR5075 5' | | MD9812 | | MD11410 | | MR5057 | | MD11776 | | MR5075 3' | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Description** | **ID number** | **Sample** | **1/10** | **Neat** | **1/10** | **Neat** | **1/10** | **Neat** | **1/10** | **Neat** | **1/10** | **Neat** | **1/10** | **Neat** | **1/10** | **Neat** | **1/10** | **Neat** | **1/10** | **Neat** |
| **Healthy Long-Term Survivor** | 1 | Fish-21-1-HK | **23.39** | **22.93** | **32.93** | **30.88** | **34.22** | **33.45** | Neg | Neg | **34.96** | **33.98** | **35.98** | **32.76** | Neg | Neg | Neg | Neg | Neg | Neg |
| | 2 | Fish-21-1GILL | **23.25** | **23.08** | **31.74** | **30.14** | **38.59** | **33.00** | **36.85** | **38.14** | **33.61** | **33.28** | **33.71** | **31.06** | Neg | Neg | Neg | Neg | Neg | Neg |
| | 3 | Fish-21-1-BRAIN | **21.51** | **20.95** | **31.72** | **30.11** | **33.74** | **34.99** | **38.20** | **37.06** | **34.23** | **33.35** | **35.94** | **33.79** | Neg | Neg | Neg | Neg | Neg | **37.13** |
| | 4 | Fish-21-GUT | **24.36** | **24.02** | **30.91** | **29.93** | **32.72** | **33.73** | **37.24** | Neg | **31.56** | **32.48** | **33.96** | **32.63** | **34.03** | Neg | Neg | Neg | Neg | Neg |
| **Fish with Reactivated Lytic Infection** | 25 | Fish-8-1-HK | **28.35** | **28.78** | **31.34** | **32.69** | **35.88** | **31.66** | **36.86** | **38.56** | **32.18** | **33.87** | **31.79** | **32.38** | Neg | Neg | Neg | Neg | Neg | Neg |
| | 26 | Fish-8-2-GILL | **27.58** | **27.86** | **30.65** | **30.09** | **34.64** | **31.03** | **36.62** | **35.89** | **32.59** | **33.07** | **31.33** | **31.24** | **31.58** | **32.96** | Neg | **36.97** | **37.08** | **34.66** |
| | 27 | Fish-8-3-Brain | **23.72** | **23.77** | **30.05** | **29.57** | **32.79** | **32.34** | Neg | **36.31** | **32.88** | **34.04** | **32.49** | **32.21** | Neg | Neg | Neg | Neg | Neg | **38.70** |

| Key | Description |
|---|---|
| | Host miRNA assay |
| | Using 1x PCR primer-probe, this assay gave low non-specific signals in non-infected CCB cells (Section 5.1) thus it was likely to also give low-level false positives with tissue samples |
| | Using 1x primer-probe this assay gave ***no*** low non-specific signals in non-infected CCB cells (Section 5.1) |

#### 5.2.2.4 *Phase 3: Re-testing of individual tissue samples that were most likely to give positive results*

Specific samples showing the most consistent results from Phase 2 were retested in duplicate using 1x RT-Primer/PCR primer-probe in order to confirm these results. RNA samples were tested at both neat and 1/10 dilutions. Due to possible low-level non-specific signals with five of the assays, only assays for MR5057, MD11776 and MR5075 3' were used in Phase 3. The results are shown in Table 5.8 and the relative levels of each target are compared in Figure 5.7. The estimated relative levels of these targets were compared using Technique 1. The different PCR efficiencies were not taken into account as most of the Ct values were low, thus the corresponding amplification curves were not suitable for calculating these with high confidence. Relative levels were normalised using the corresponding levels of host miRNA ccr-let-7a in the same samples (levels based on Ct values in Table 5.7). Only samples showing (i) amplification in each duplicate and (ii) an average Ct value that was <35 were deemed positive. The only positive test results in Phase 3 therefore were obtained from samples 4 (from possible latent carrier) and sample 26 (from fish with lytic infection).

The positive RT-qPCR and negative tissue controls gave the expected results. At least one replicate from each of the 4 samples gave a positive result. Sample 26 tested positive for all 3 targets using both dilutions and in all replicates (Table 5.8). This sample consisted of RNA from the gill of a fish experiencing a reactivated lytic infection. MR5057 was present at the highest levels followed by MD11776 and MR5075 and this was the case for both dilutions of sample 26 (Figure 5.7). The relative levels of these 3 targets in the *in vivo* lytic infection corresponded to the relative levels for the same targets from *in vitro* lytic infections (Figure 5.5 (a) Figure 5.6 (b)). The Ct values associated with MR5057 in Phase 3 were quite low compared to all other positives and were the only positive signals from Phase 3 that had Ct values <32.

The only other sample to score positives in all replicates was Sample 4, although MR5057 was the only assay to test positive in that case. This RNA sample was from the gut of a long-term healthy survivor of a previous high mortality CyHV-3 outbreak. The levels of these miRNAs in sample 4 and sample 26 are directly compared (not

normalised) in Figure 5.7. This result showed that the MR5057 levels are low in Sample 4 (possible latent carrier) relative to Sample 26 (lytic infection). When these samples are normalised to levels of host miRNA in the same samples (Figure 5.8) the levels of MR5057 in Sample 4 are shown to be even lower relative to its level in Sample 26.

**Table 5.8 Phase 3: testing of samples that were identified in Phase 2 as most likely to give reproducible positive results using 1x RT-Primer/1x PCR primer-probe.**

| Description | ID | Sample | Dilution | Replicate | MR5057 | MD11776 | MR5075 |
|---|---|---|---|---|---|---|---|
| Healthy Long-Term Survivor | 3 | Fish-21-1-BRAIN | 1/10 | 1 | Negative | Negative | Negative |
| | | | | 2 | Negative | Negative | Negative |
| | | | Neat | 1 | Negative | Negative | Negative |
| | | | | 2 | Negative | Negative | **37.97** |
| | 4 | Fish-21-1-GUT | 1/10 | 1 | **34.46** | Negative | Negative |
| | | | | 2 | **33.88** | Negative | Negative |
| | | | Neat | 1 | **34.23** | Negative | Negative |
| | | | | 2 | **34.79** | Negative | Negative |
| Fish with Reactivated Lytic Infection | 26 | Fish-8-2-GILL | 1/10 | 1 | **30.66** | **35.02** | **36.38** |
| | | | | 2 | **30.52** | **33.95** | **37.11** |
| | | | Neat | 1 | **30.45** | **37.09** | **34.77** |
| | | | | 2 | **29.61** | **32.38** | **33.42** |
| | 27 | Fish-8-3-BRAIN | 1/10 | 1 | Negative | Negative | Negative |
| | | | | 2 | Negative | **37.06** | Negative |
| | | | Neat | 1 | Negative | **36.65** | Negative |
| | | | | 2 | Negative | Negative | Negative |

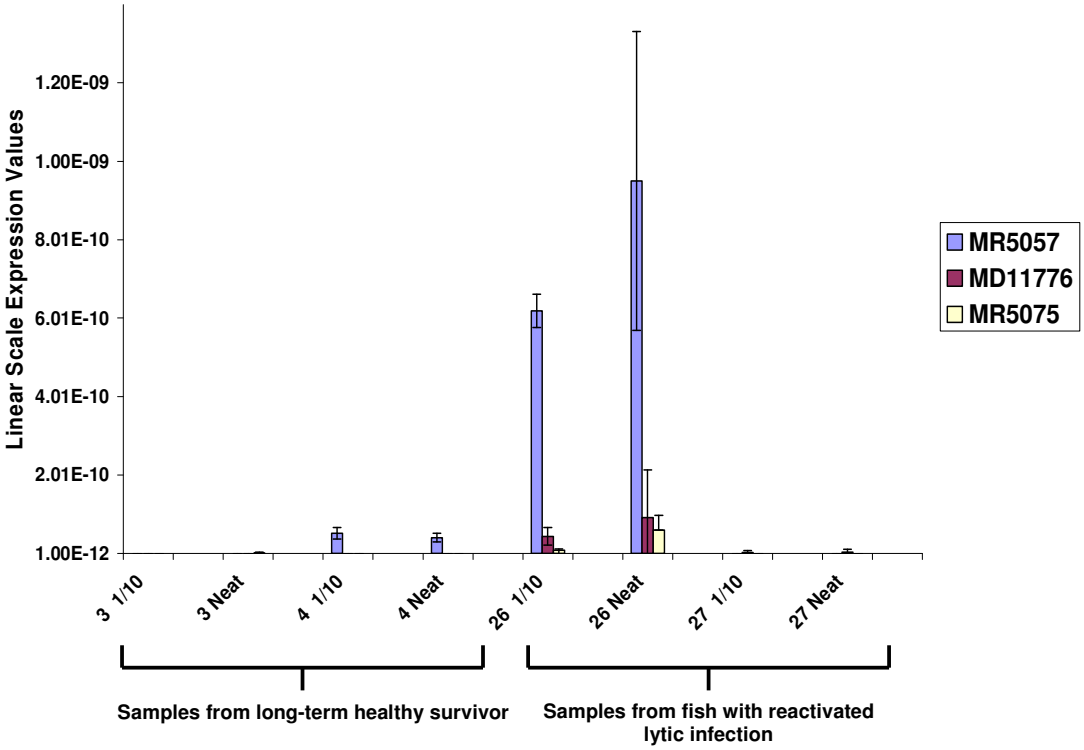


**Figure 5.7 Comparison of the relative levels of CyHV-3 miRNAs (not normalized) from tissue samples tested in Phase 3**
These expression values are based on the average of the duplicates. Error bars represent standard deviation between duplicates.

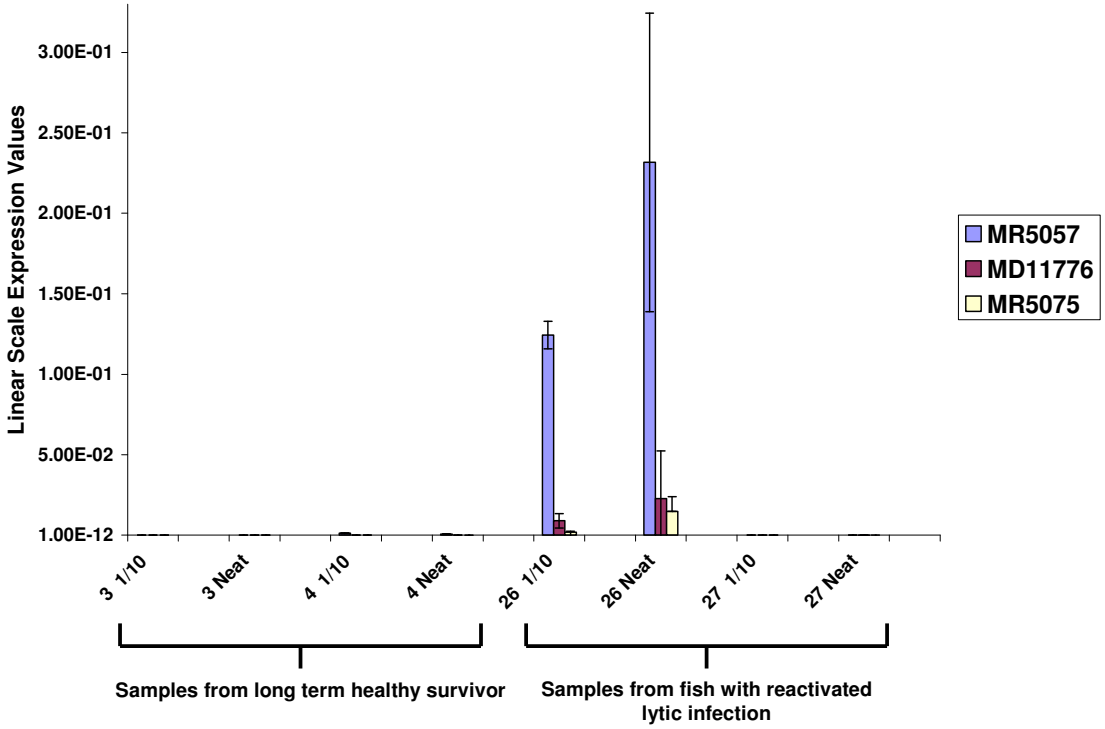



**Figure 5.8 Comparison of the relative levels of CyHV-3 miRNAs (normalized) from tissue samples tested in Phase 3 testing**
Expression values are based on the average of the duplicates and were normalized to levels of ccr-let-7a in the same samples. Error bars represent standard deviation between duplicates

# 6 Investigation into CyHV-3 miRNA sequence conservation

## 6.1 Identification of potential CyHV-3 pre-miRNA and/or miRNA homologues in other CyHV genomes

MiRNA sequences are generally not conserved between viruses. Viral miRNA conservation and/or high sequence similarity is quite rare and to date this has only been observed between closely related viruses (Section 1.6.2). The closest related viruses to CyHV-3 are CyHV-1 and CyHV-2. Incidentally, these 2 viral genomes were recently sequenced and found to contain many homologues of CyHV-3 genes and so the possibility that there were homologues of CyHV-3 miRNA genes on these genomes was thus investigated.

BLAST searches were carried out for high probability CyHV-3 pre-miRNAs in the other CyHV genomes. This was initially confined to non-coding regions that were directly orthologous to non-coding regions of the CyHV-3 genome that encoded pre-miRNAs. Only two of these CyHV-3 pre-miRNAs showed matches in these regions, namely to MD1111 and MD11776 (Table 6.1). The match to MD1111 was only partial and only present in CyHV-2, but this match was to the region of the pre-miRNA coding for the 5' arm miRNA (Figure 6.1 (a)). Matches to MD11776 were found in both CyHV-1 and CyHV-2 and consisted of the entire pre-miRNA sequence (Figure 6.1 (b)).

The search was then expanded to include the entire genomes of CyHV-1 and CyHV-2. This time the E-values for the MD1111 and MD11776 matches described above were higher. The highest E-Value in both searches was the partial match to the MD1111 pre-miRNA in CyHV-2. This went from 1E-8 in the first BLAST search (region-specific) to 2E-6 in the second BLAST search (entire genome) (Table 6.1). This served as a basis for establishing a relevant cut-off point for E-Values of 1E-5. In the second search (based on the entire genomes of CyHV-1 and CyHV-2) many other CyHV-3 pre-miRNAs showed partial matches to short stretches of sequences on several parts of the CyHV-2 and CyHV-3 genomes, however none of these matches showed E-values <1E-5.

As a result of the good matches observed to pre-miRNAs MD1111 and MD11776 it was deemed relevant to check if there were any other pre-miRNA candidates that occurred in the same regions but that had been eliminated in Section 4.1.3 or 4.1.4 due to low read counts. This was because any other CyHV-3 pre-miRNA candidates occurring in the same region were also likely to show some sequence conservation, thus supporting their annotation as pre-miRNAs despite their low read counts. One such candidate, MD11704 was found to occur very close to MD11776 (~1 kb). It was identified in both infections by Mireap (Section 4.1.4.1) and also complied with most of the pre-miRNA identification criteria (details in Supplementary File 6.1 and Table 6.3 (a)-(f)), but was eliminated due to low read counts. Incidentally BLAST analysis revealed a possible homologue to that pre-miRNA candidate in both CyHV-1 and

CyHV-2 and both matches had very low E-values (Table 6.1). In addition, the match spanned the entire MD11704 pre-miRNA sequence (Figure 6.1 (c))

With the exception of the first 3 nucleotides, the 5' miRNA sequence from MD1111 was perfectly conserved in the candidate CyHV-2 homologue (Figure 6.1 (a)). The remainder of the pre-miRNA was not well conserved unlike the general genomic location (i.e. ORF7 in CyHV-2). The seed region of the 3' miRNA sequence in MD11776 is conserved in CyHV-1 and CyHV-2 and the remainder of the miRNA is also conserved in both viruses with the exception of 2 bases (Figure 6.1 (b)) The sequence similarity across the remainder of the pre-miRNA was high (78% for CyHV-1 and 87% for CyHV-2). The seed region for the 5' miRNA from MD11704 was also conserved in CyHV-1 and CyHV-2 and the remainder of the miRNA was also well conserved with 2 and 3 mismatches to CyHV-1 and CyHV-2 respectively. As was the case with CyHV-3, the homologues for both MD11776 and MD11704 are located in a non-coding region of the genome opposite ORF87 in both CyHV1 and CyHV-2.

VMir was also used to predict pre-miRNAs in the CyHV-1 and CyHV-2 genomes (described in Section 3). One of the potential CyHV-3 pre-miRNA homologues was also predicted by VMir, namely the homologue of MD11776 in the CyHV-2 genome and this matched perfectly to the homologue identified through BLAST. In addition this predicted pre-miRNA had a high VMir score of 203 and a high relative WC of 39. Furthermore, analysis of the predicted pre-miRNA structure using pre-miRNA classifiers (as in Section 4.1.3.2) classified this predicted pre-miRNA as a real miRNA. Conversely, the same analysis of the equivalent homologue in the CyHV-1 genome suggested that it was not a real pre-miRNA. None of the other sequences for possible pre-miRNA homologues were classed as real pre-miRNAs (Table 6.2)

**Table 6.1 BLAST searches for CyHV-3 pre-miRNA sequence homologues on other CyHV genomes**

| Sequences Aligned | | Results from first BLAST to orthologous non-coding regions | | | | Results from BLAST to whole genome |
|---|---|---|---|---|---|---|
| CyHV-3 pre-miRNA | Viral genomic region | Score | E-Value | Identities | Gaps | E-Value |
| **MD1111** | CyHV-2 non-coding region opposite ORF7 | 44.6 bits(48) | 1.00E-08 | 24/24(100%) | 0/24(0%) | 2.00E-06 |
| **MD11410** | CyHV-1 non-coding region opposite ORF84 | No Significant results | | | | |
| | CyHV-2 non-coding region opposite ORF84 | | | | | |
| **MD11776** | CyHV-1 non-coding region opposite ORF87 | 104 bits(114) | 2.00E-26 | 95/120(79%) | 0/120(0%) | 2.00E-24 |
| | CyHV-2 non-coding region opposite ORF87 | 159 bits(176) | 3.00E-43 | 114/131(87%) | 0/131(0%) | 2.00E-41 |
| **MD11704** | CyHV-1 non-coding region opposite ORF87 | 69.8 bits(76) | 2.00E-16 | 58/71(82%) | 0/71(0%) | 2.00E-14 |
| | CyHV-2 non-coding region opposite ORF87 | 86.0 bits(94) | 3.00E-21 | 65/77(84%) | 0/71(0%) | 2.00E-19 |
| **MD9812** | CyHV-1 non-coding region opposite ORF69 | No Significant results | | | | |
| | CyHV-2 non-coding region opposite ORF69 | | | | | |
| **MR5057** | CyHV-1 non-coding region opposite ORF43 | No Significant results | | | | |
| | CyHV-2 non-coding region opposite ORF43 | | | | | |
| **MR5075** | CyHV-1 non-coding region opposite ORF43 | No Significant results | | | | |
| | CyHV-2 non-coding region opposite ORF43 | | | | | |

**(a)**

| Sequence details | Genome Position | Sequence Alignment |
|---|---|---|
| CyHV-2 (Possible MD11111 homologue) | 9641 | GAAACATGTTTCTTTTTCTTCTTCT-TAGCAGGCTCATCATCATCATCAGACAAATCAACAACAGCCTTAGACTTCTTCTTCTTATCTTTCTTTGACTTTTTAGA |
| CyHV-3 MD1111 Pre-miRNA | 15468 | ATCACCT-CCTCCTCACCAGCATCAGCAGCAGGCTCATCATCATCATCAGA---GTCAGAGATGAGGATGGTCTCAGTCTGTTGTGGTGGTGGTGTGGACTTGGCT |
| CyHV-3 MD1111 5' moRNA | | TCACCT-CCTCCTCACCAGC |
| CyHV-3 MD1111 5' miRNA | | AGCAGCAGGCTCATCATCATCA |
| CyHV-3 MD1111 3' miRNA | | AGGATGGTCTCAGTCTGTTGTGGT |
| CyHV-3 MD1111 3' moRNA | | TGGTGGTGTGGACTTGGCT |

**(b)**

| Sequence details | Genome Position | Sequence Alignment |
|---|---|---|
| CyHV-1 (Possible MD11776 homologue) | 143064 | AGGCATCCCGTTCACATCAGTGAGGACTTTTCCGTAGAACTTGTCGACAAAGTTGCAGCAGGCCTGGGTGAGGGTGTTGGCAGGAACATGTTTGTTTGGTAGCCA |
| CyHV-2 (Possible MD11776 homologue) | 149470 | TGGCATCCCGTTGATGCCGGTGAGAACCTTCCCGTAGAACTTGTCGACAAAGTTGCAGCAGGCCTGTGTGAGGTTGTTGGCCGGTACGTGTTTGTTGGGATGCCA |
| CyHV-3 MD11776 Pre-miRNA | 159580 | TCTGGCACGCCGTTGATACCGGTCAGCACCCTCCCGTAGAACTTGTCCACAAAGTTGCAGCAGGCCTGCGTGAGGGTGTTGGCCGGCACGTGCTTGTTGGGCG |
| CyHV-3 MD11776 5' moRNA | | TCTGGCACGCCGTTGATACCGG |
| CyHV-3 MD11776 5' miRNA | | CAGCACCCTCCCGTAGAACTTG |
| CyHV-3 MD11776 3' miRNA | | AGGCCTGCGTGAGGGTGTTGGC |
| CyHV-3 MD11776 3' moRNA | | CGGCACGTGCTTGTTGGGCG |

**(c)**

| Sequence details | Genome Position | Sequence Alignment |
|---|---|---|
| CyHV-1 (Possible MD11704 homologue) | 142069 | AGAGTCCCGACCTTGTGGTTGAGGGTGCGCGTACTAAAGTACGTGTGGCCATACGCGTTGAGGCTGTCCAC |
| CyHV-2 (Possible MD11704 homologue) | 148472 | AGAGTCCCCACCTTGTGGTTCAGCGTACGAGTGCTAAAGTACGTGTGTCCGTAGGCGCACAGGGAGTCCACGTT |
| CyHV-3 MD11704 | 158568 | CAGCGTCCCCACCTTGTGGTTGAGCGTCCGCGTCGAAAAGTACGTGTGGCCGTACGCACACAGGGTGTCCACGTTGCC |
| 5' miRNA CyHV-3 | | ACCTTGTGGTTGAGCGTCCGCG |
| 3' miRNA CyHV-3 | | TGGCCGTACGCACACAGGGTGT |

**Figure 6.1 Alignment of CyHV-3 pre-miRNAs to possible homologues in CyHV-1 and CyHV-3 genomes**
(a) Alignment to of MD1111 to possible CyHV-2 homologue. Alignment of MD11776 (b) and MD11704 (c) to possible CyHV-1 and CyHV-2 homologues

**Table 6.2 Structural assessment of possible CyHV-3 pre-miRNA homologues from CyHV-1 and CyHV-2, carried out as per Section 4.1.3.2 using pre-miRNA classifiers**
The only possible CyHV-3 pre-miRNA homologue that was also predicted by VMir analysis of the CyHV-1 and CyHV-2 genomes was the candidate MD11776 homologue in CyHV-2. Hence, as this homologue was predicted by VMir, the SHP with the highest absolute WC was analyzed. All other sequences used are based on the alignments listed in Figure 6.1

| Pre-miRNA | Pre-miRNA sequence used for analysis | Abs. WC | Rel. WC | MiPred | CSHMM | Classified as Real Pre-miRNA |
|---|---|---|---|---|---|---|
| **CyHV-1 Possible MD1111 homologue** | GAAACATGTTTCTTTTTCTTCTTCTTAGCAGGCTCATCATCATCATCAGACAAATCAAC AACAGCCTTAGACTTCTTCTTCTTATCTTTCTTTGACTTTTTAGA | **N/A** | **N/A** | **No** | **No** | **No** |
| **CyHV-1 Possible MD11776 homologue** | AGGCAUCCCGUUCACAUCAGUGAGGACUUUUCCGUAGAACUUGUCGACAAAGUUG CAGCAGGCCUGGGUGAGGGUGUUGGCAGGAACAUGUUUGUUUGGUAGCCA | **N/A** | **N/A** | **No** | **No** | **No** |
| **CyHV-2 Possible MD11776 homologue** | GGCATCCCGTTGATGCCGGTGAGAACCTTCCCGTAGAACTTGTCGACAAAGTTGCAG CAGGCCTGTGTGAGGTTGTTGGCCGGTACGTGTTTGTTGGGATGCC | **22** | **32** | **Yes** | **Yes** | **Yes** |
| **CyHV-1 Possible MD11704 homologue** | AGAGUCCCGACCUUGUGGUUGAGGGUGCGCGUACUAAAGUACGUGUGGCCAUACG CGUUGAGGCUGUCCAC | **N/A** | **N/A** | **No** | **No** | **No** |
| **CyHV-2 Possible MD11704 homologue** | AGAGUCCCCACCUUGUGGUUCAGCGUACGAGUGCUAAAGUACGUGUGUCCGUAGG CGCACAGGGAGUCCACGUU | **N/A** | **N/A** | **No** | **No** | **No** |

**Table 6.3 Summary of assessment of pre-miRNA candidate MD11704**
(a) Putative miRNAs and read counts (b) mature miRNA duplex assessment (c) VMir prediction details (d) pre-miRNA classification (e) end heterogeneity (f) summary

**(a)**

| Pre-miRNA | Arm | Infection | Sequence | Same isomiR is miRNA in both infections | IsomiR present in both infections | Read Count |
|---|---|---|---|---|---|---|
| MD11704 | 5' miRNA | H361 | ACCTTGTGGTTGAGCGTCCGC | No | Yes | 1 |
| | | N076 | ACCTTGTGGTTGAGCGTCCGCG | | No | 5 |
| | 3' miRNA | H361 | TGGCCGTACGCACACAGGGTGT | Yes | Yes | 3 |
| | | N076 | TGGCCGTACGCACACAGGGTGT | | Yes | 7 |

**(c)**

| Clustered/non-clustered | Pre-miRNA | Start | End | VMir Score | Rel. WC | Pre-miRNA candidate sequence 5' to 3' |
|---|---|---|---|---|---|---|
| **Reverse Strand Cluster 2** | MD11704 | 158475 | 158714 | 207 | 44 | CUGAUUGAAGCGCGCGAUGGCCUCCUUGGGGAUCUCGUAGAAGAGCGGUAUCUCAACUGUCGUGCCGUUGAUAG UGAUGCCGUCCAGGUUGGGCAGCGUCCCCACCUUGUGGUUGAGCGUCCGCGUCGAAAAGUACGUGUGGCCGUAC GCACACAGGGUGUCCACGUUGCCGUUGGGUCGCAUGAACUCGGCGGUGAAGAUGUUGGGCAUGGUGUUCUUGGU GUCCUCCUGGUUGUCCGGCAUCAGGUAG |

**(d)**

| Pre-miRNA | Pre-miRNA sequence used for analysis | Abs. WC | Rel. WC | MFE (kcal/mol) | MiPred | CSHMM | Classified as Real Pre-miRNA by both methods |
|---|---|---|---|---|---|---|---|
| **MD11704** | GGCAGCGUCCCCACCUUGUGGUUGAGCGUCCGCGUC GAAAAGUACGUGUGGCCGUACGCACACAGGGUGUCC ACGUUGCC | 20 | 44 | -39.40 | Real | Real | Yes |

(e)

| Pre-miRNA | miRNA | Ratio-1* | | Ratio-2* | | Ratio-3* | |
|---|---|---|---|---|---|---|---|
| | | H361 | N076 | H361 | N076 | H361 | N076 |
| MD11704 | 5' miRNA | No isomiRs | 2.0 | No isomiRs | 1.0 | No isomiRs | 2.0 |
| | 3' miRNA | No isomiRs | 1.0 | No isomiRs | 1.0 | No isomiRs | 1.0 |

*** Very few isomiRs were detected: may require a greater number of isomiRs in order to carry out meaningful end heterogeneity analysis**

(f)

| Pre-miRNA Name | 1. Highly expressed small RNAs | 2. From non-coding region | 3. Putative miRNAs adjacent to stem | 4. Putative major and minor form detected | 5. MiRNA isoform stability | 6. Classified as real pre-miRNA | 7. Pre-miRNA structure has MFE <-25 | 8. Mature-miRNA-duplex 3' overhangs | 9. MiRNA-like alignment signature | 10. Greater 3' end heterogeneity among isomiRs * | 11. Discrete, read-enriched locus | Compliance with all points |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **MD11704** | No | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **Yes** | **No** | No | No |

## 6.2 CyHV-3 miRNA seed region conservation

The seed regions of miRNAs are the primary determinants of miRNA target specificity and for this reason seed regions are the most likely regions to be conserved between otherwise divergent miRNAs. MiRNAs sharing the same seed regions are likely to target the same mRNA and thus have the same function. In order to see if any CyHV-3 miRNAs shared seed regions with any known viral and/or host miRNAs, SeqMap was used to compare seed regions from all miRNAs from the six high probability pre-miRNAs to all known viral and host seeds (allowing only 1 mismatch).

### 6.2.1 CyHV-3 seed region matches to viral miRNAs

Six miRNAs from five of the CyHV-3 pre-miRNA had either full or partial matches (1 mismatch) to seed regions of other viral miRNAs. Two had full seed matches, one of these (MD11776 5') had a full match to miRNAs from 4 different viruses (one of these viral miRNA is actually conserved between EBV and rLCV). The most common matches are to miRNAs from EBV, rLCV and MDV-1 (Table 6.4).

**Table 6.4 CyHV-3 seed region matches to known viral miRNAs**
Seed regions (2-7) are marked in bold. U/C or G/A bases occurring at equivalent positions on different miRNAs are indicated with an "x". This is important in this context because despite the differences in sequence, the nucleotides at these positions on both miRNAs can in theory still base pair with the same nucleotides on what are possibly the same the target sites due to GU-wobbles.

| MiRNA | Alignment (Seed regions in bold) |
|---|---|
| | M**MMM MM**M M |
| **MD1111 3'** | A**GGAUGG**UCUCAGUCUGUUGUGGU |
| **blv-miR-B5-5p** | A**GGAAGG**UUGUGGCUCAGAGGU |
| U/C Bases in both can pair with G in Target Site | x x xxx |
| G/A Bases in both can pair with U in Target Site | X x |
| | |
| | **MMMMM**M M M M M M M |
| **MD1111 3'** | A**GGAUGG**UCUCAGUCUGUUGUGGU |
| **hhv6b-miR-Ro6-4-5p** | C**GGAUGG**AGCCCGCUCGGUCUCG |
| U/C Bases in both can pair with G in Target Site | x xxx |
| G/A Bases in both can pair with U in Target Site | |
| | **MM MMM**M MM M |
| **MD11776 3'** | A**GGCCUG**CGUGAGGGUGUUGGC |
| **mdv1-miR-M2-3p** | C**GGACUG**CCGCAGAAUAGCUU |
| U/C Bases in both can pair with G in Target Site | X |
| G/A Bases in both can pair with U in Target Site | xx x |
| | **MMMMM** M |
| **MD11776 5'** | C**AGCACC**CUCCCGUAGAACUUG |
| **blv-miR-B4-3p** | U**AGCACC**ACAGUCUCUGCGCCUUU |
| U/C Bases in both can pair with G in Target Site | x xx |
| G/A Bases in both can pair with U in Target Site | x |
| | |
| | **MMMMM** M M M |
| **MD11776 5'** | C**AGCACC**CUCCCGUAGAACUUG |
| **ebv-miR-BART1-3p** | U**AGCACC**GCUAUCCACUAUGUC |
| U/C Bases in both can pair with G in Target Site | xx x x x |
| G/A Bases in both can pair with U in Target Site | |
| | |
| | **MMMMM** MMM M MM |
| **MD11776 5'** | C**AGCACC**CUCCCGUAGAACUUG |
| **mdv2-miR-M21-3p** | G**AGCACC**ACGCCGAUGGACGGAGA |
| U/C Bases in both can pair with G in Target Site | x |
| G/A Bases in both can pair with U in Target Site | x x |
| | |
| | **MMMMM** M M M |
| **MD11776 5'** | C**AGCACC**CUCCCGUAGAACUUG |
| **rlcv-miR-rL1-6-3p** | U**AGCACC**GCUAUCCACUAUGUC |
| U/C Bases in both can pair with G in Target Site | X x |
| G/A Bases in both can pair with U in Target Site | |
| | |
| | **MM MMM**M M M |
| **MR5057 3'** | A**AAUUGC**GGCCGCUGUCGACGA |
| **rlcv-miR-rL1-32-3p** | U**AAAUGC**GAGCAGUAGUAGGCG |
| U/C Bases in both can pair with G in Target Site | X |
| G/A Bases in both can pair with U in Target Site | x x x xx |
| | |
| | **M MMMM**M M M MM |
| **MR5075 3'** | A**CGGUGC**UAGUAUCAACUGCAAGG |
| **hvsa-miR-HSUR2-5p** | U**CUGUGC**UCUCAAGGCUUUAAA |
| U/C Bases in both can pair with G in Target Site | X |
| G/A Bases in both can pair with U in Target Site | x |
| | |
| | M**MMMMM** MMM |
| **MD9812 5'** | A**GGCCGU**CGGCCUUCAUGAUG |
| **hsv2-miR-H23-5p** | A**GGCCGU**GGAGCUUGCCAGC |
| U/C Bases in both can pair with G in Target Site | x |
| G/A Bases in both can pair with U in Target Site | x xx |
| | |
| | M**MMM M**M M M M |
| **MD9812 5'** | A**GGCCGU**CGGCCUUCAUGAUG |
| **mdv1-miR-M12-5p** | A**GGCCCU**CCGUAUAAUGUAAAUGU |
| U/C Bases in both can pair with G in Target Site | X |
| G/A Bases in both can pair with U in Target Site | x |

### 6.2.2 CyHV-3 seed region matches to host miRNAs

Three CyHV-3 5'arm miRNAs (all minor forms) had matches to multiple host miRNAs, although many multiple matches were to members of the same miRNA families that share the same seed sequence. MR5057 5' (minor strand) was found to have a partial match (1 mismatch) to the seed region of one of the most abundant cellular miRNAs ccr-Let-7a and members of its miRNA family including Let-7b, Let 7g and Let-7i (Table 6.5). In addition it also had matches to positions 8 and 9 on these miRNAs. MD1111 5' had perfect matches (positions 1-7) to mir-103 and mir-107 (same family) and with mir-457a and mir-457b (same family). MD11776 5' also had perfect matches to miRNAs mir-29a and mir-29b (Table 6.5). All of these also tended to have additional matches outside the seed regions, and interestingly these were usually from bases 14-17 (Table 6.5). This is relevant in this context because in addition to base pairing between the seed and target site, base pairing outside the seed, especially with bases at these positions, can sometimes be important for miRNA-target duplex stability (Section 1.8).

**Table 6.5 CyHV-3 seed region matches to known host miRNAs**

Seed regions (2-7) are marked in bold. U/C or G/A bases occurring at equivalent positions on different miRNAs are indicated with an "x".

| MiRNA | Alignment (Seed regions in bold) |
|---|---|
| | **MMMM M**MM MM |
| **MR5057 5'** | G**GAGGCA**GUGACGGCAAUGUCGUU |
| **ccr-let-7a** | U**GAGGUA**GUAGGUUGUAUAGUU |
| U/C Bases in both can pair with G in Target Site | **X** x |
| G/A Bases in both can pair with U in Target Site | x x |
| | |
| | **MMMM M**MM M |
| **MR5057 5'** | G**GAGGCA**GUGACGGCAAUGUCGUU |
| **ccr-let-7b** | U**GAGGUA**GUAGGUUGUGUGGUU |
| U/C Bases in both can pair with G in Target Site | **X** x |
| G/A Bases in both can pair with U in Target Site | xx x |
| | |
| | **MMMM M**MM MM |
| **MR5057 5'** | G**GAGGCA**GUGACGGCAAUGUCGUU |
| **ccr-let-7g** | U**GAGGUA**GUAGUUUGUAUAGUU |
| U/C Bases in both can pair with G in Target Site | **X** x x |
| G/A Bases in both can pair with U in Target Site | Xx x |
| | |
| | **MMMM M**MM |
| **MR5057 5'** | G**GAGGCA**GUGACGGCAAUGUCGUU |
| **ccr-let-7i** | U**GAGGUA**GUAGUUUGUGCUGU |
| U/C Bases in both can pair with G in Target Site | **X** x x x |
| G/A Bases in both can pair with U in Target Site | x x |
| | |
| | M**MMMMMM** M |
| **MD1111 5'** | A**GCAGCA**GGCUCAUCAUCAUCA |
| **ccr-miR-103** | A**GCAGCA**UUGUACAGGGCUAUGA |
| U/C Bases in both can pair with G in Target Site | x |
| G/A Bases in both can pair with U in Target Site | x x |
| | |
| | M**MMMMMM** M M |
| **MD1111 5'** | A**GCAGCA**GGCUCAUCAUCAUCA |
| **ccr-miR-107** | A**GCAGCA**UUGUACAGGGCUAUC |
| U/C Bases in both can pair with G in Target Site | x |
| G/A Bases in both can pair with U in Target Site | x |
| | |
| | **MMMM**MMMM MM |
| **MD1111 5'** | A**GCAGCA**GGCUCAUCAUCAUCA |
| **ccr-miR-214** | U**ACAGCA**GGCACAGACAGG |
| U/C Bases in both can pair with G in Target Site | |
| G/A Bases in both can pair with U in Target Site | **X** x |
| | |
| | M**MMMMMM** MM M |
| **MD1111 5'** | A**GCAGCA**GGCUCAUCAUCAUCA |
| **ccr-miR-457a** | A**GCAGCA**CAUCAAUAUUGGC |
| U/C Bases in both can pair with G in Target Site | x x |
| G/A Bases in both can pair with U in Target Site | x x |
| | |
| | M**MMMMMM** MM M |
| **MD1111 5'** | A**GCAGCA**GGCUCAUCAUCAUCA |
| **ccr-miR-457b** | A**GCAGCA**CAUAAAUACUGGAG |
| U/C Bases in both can pair with G in Target Site | x |
| G/A Bases in both can pair with U in Target Site | x x |
| | |
| | **MMMMM** M MM |
| **MD11776 5'** | C**AGCACC**CUCCCGUAGAACUUG |
| **ccr-miR-29a** | U**AGCACC**AUUUGAAAUCGGUUA |
| U/C Bases in both can pair with G in Target Site | X xx |
| G/A Bases in both can pair with U in Target Site | x x x |
| | |
| | **MMMMM** M M M |
| **MD11776 5'** | C**AGCACC**CUCCCGUAGAACUUG |
| **ccr-miR-29b** | U**AGCACC**AUUUGAAAUCAGUGUU |
| U/C Bases in both can pair with G in Target Site | X xx |
| G/A Bases in both can pair with U in Target Site | x |

# 7 CyHV-3 miRNA target prediction

Viral mRNA target site predictions were made for miRNAs derived from all high probability CyHV-3 pre-miRNAs. Target sites were predicted for both major and minor form miRNAs as it is possible for both to be functionally active**.** This process involved the use two very different target site prediction methods, TargetScan and PITA. The process of miRNA target site prediction used in this study is outlined below in Figure 7.1. The target site predictions for MR5057 and MD11776 are described in this section. The results for other miRNAs are given in the Supplementary File 7.1.

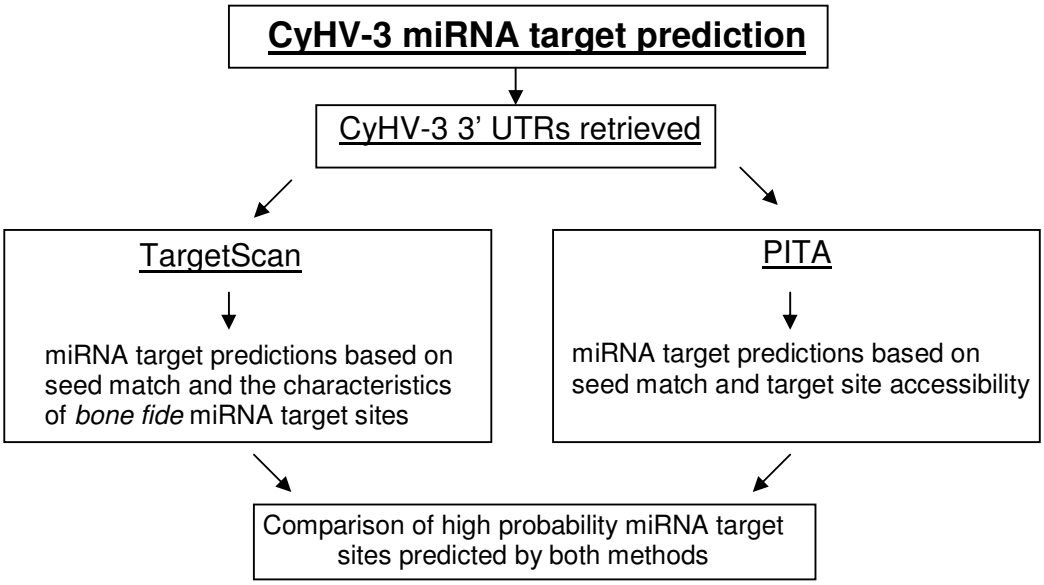



**Figure 7.1 Outline of process involved in CyHV-3 miRNA target site prediction**

## 7.1 Predicted CyHV-3 targets for the 5' arm miRNA from precursor MR5057

TargetScan and PITA predicted 45 and 106 potential target sites respectively for this miRNA. Details of all these predicted target sites are available in Supplementary File 7.2 T.1-T.3. The top ranked site from TargetScan was located in the 3' UTR of ORF117. The same site was also the top ranked site by PITA (Table 7.1). The seed

match is a 8mer-1a site (Figure 7.2) and occurs near the start of the 3' UTR of ORF117, from position 39-46 (Figure 7.3). Bases at positions 9-11, 14-15 and 22 on the miRNA are also complementary to the 3' UTR (Figure 7.2).

**Table 7.1 Comparison of highest ranking TargetScan and PITA results for MR5057 5' miRNA target site prediction**

| MR5057 5' CyHV-3 Targets | | | | | | |
|---|---|---|---|---|---|---|
| **Best TargetScan Prediction** | **3' UTR** | **Seed Match Position** | **Context Score** | **Rank** | **Total Sites** | **Percentile** |
| | ORF_116-117 | 39-46 | -0.381 | 1 | 45 | 100 |
| **Results for the same site predicted by PITA** | | | **Δ ΔG** | **Rank** | **Total Sites** | **Percentile** |
| | | | -22.41 | 1 | 106 | 100 |

```
ORF-117-116 3' UTR 5'  UCACAUCCGCUUUCACUGCCUCA 3'
MR5057 5' MiRNA    3' UUGCUGUAACGGCAGUGACGGAGG 5'
```

**Figure 7.2 Alignment of MR5057 5' miRNA to its highest ranking predicted target site in Table 7.1**
Seed matches are highlighted in red. Other elements of the 8mer-1a target site are highlighted in green (nucleotide on target site base pairing to position 8 on miRNA) and purple (A on target site opposite position 1 on miRNA). All base pairing between the miRNA and target site are highlighted in bold.


**Figure 7.3 Genomic location of highest ranking predicted target site for MR5057 5'miRNA (from Table 7.1)**

## 7.2 Predicted CyHV-3 targets for the 3' arm miRNA from precursor MR5057

TargetScan and PITA predicted 2 and 22 potential target sites respectively for this miRNA (details of all these sites are given in Supplementary File 7.2 T.4-T.6). The top ranked site from TargetScan is in the 3' UTR of ORF123. The same site was also the top ranked site by PITA (Table 7.2). The seed match is a 7mer-m8 site and occurs at position 22-28 on the 3' UTR of ORF123. Bases at positions 9, 13, 15 and 20 on the miRNA are also complementary to the 3' UTR (Figure 7.4). This 3' UTR

overlaps with the 3' UTR of the upstream ORF122 (as ORF123 actually occurs in the 3' UTR of ORF122) (Figure 7.5).

**Table 7.2 Comparison of highest ranking TargetScan and PITA results for MR5057 3' miRNA target site prediction**

| MR5057 3' CyHV-3 Targets | | | | | | |
|---|---|---|---|---|---|---|
| **Best TargetScan Prediction** | **3' UTR** | **Seed Match Position** | **Context Score** | **Rank** | **Total Sites** | **Percentile** |
| | ORF-123 | 22-28 | -0.139 | 1 | 2 | 100 |
| **Results for the same site predicted by PITA** | | | **Δ ΔG** | **Rank** | **Total Sites** | **Percentile** |
| | | | -13.65 | 1 | 22 | 100 |

```
ORF-123 3' UTR          5'GGCAUACUCUGAACCCGCAAUUG 3'
MR5057 3' MiRNA         3' AGCAGCUGUCGCCGGCGUUAAA 5'
```

**Figure 7.4 Alignment of MR5057 3' miRNA to its highest ranking predicted target site in Table 7.2**
Seed matches are highlighted in red. Other elements of the 7mer-m8 target site are highlighted in green (nucleotide on target site base pairing to position 8 on miRNA). All base pairing between the miRNA and its target site are highlighted in bold.

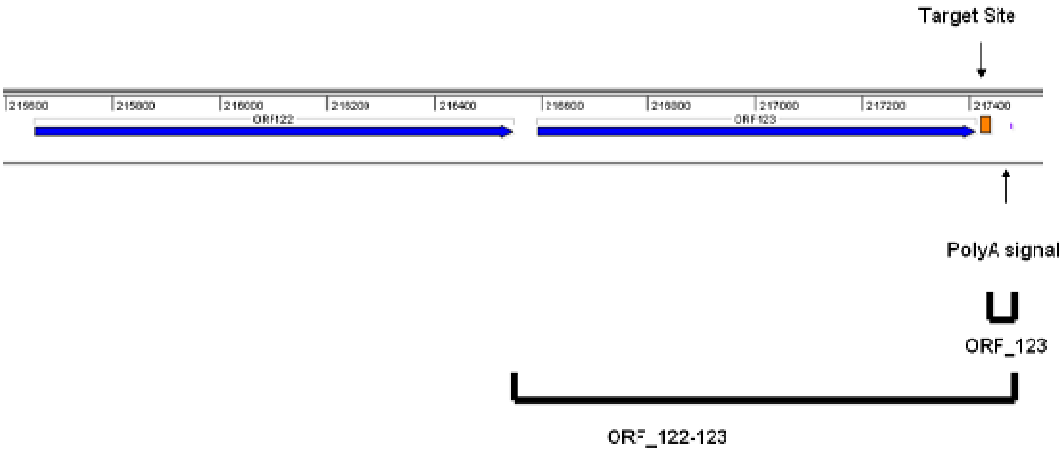



**Figure 7.5 Genomic location of the highest ranking predicted target site for MR5057 3'miRNA (from Table 7.2)**

## 7.3 Predicted CyHV-3 targets for the 5' arm miRNA from precursor MD11776

TargetScan and PITA predicted 18 and 81 potential target sites respectively for this miRNA (details of these sites are given in Supplementary File 7.2 T.7-T.9). The top ranked predicted target site from TargetScan was located in the 3' UTR of ORF117.

The ΔΔG value for this site, as calculated by PITA was a positive value, indicating that this site was not accessible. The most accessible site as deemed by PITA was in the 3' UTR of ORF 129, which is still within the top 10 ranked predictions from TargetScan (Table 7.3). The seed match is 7mer-m8 and occurs near the end of the 3' UTR from position 1305-1312. Bases at positions 12, 14, 16, 18-19 and 21-22 on the miRNA are also complementary to the 3' UTR (Figure 7.6) This occurs within the coding region of ORF128 (which itself occurs entirely within the 3' UTR of ORF129) (Figure 7.7).

**Table 7.3 Comparison of highest ranking TargetScan and PITA results for MD11776 5' miRNA target site prediction**

Due to its high ranking in both sets of results the target site in the 3' UTR of ORF 128-129 was identified as most likely CyHV-3 target site for the MD11776 5' miRNA

| Best TargetScan Prediction | | | | | | |
|---|---|---|---|---|---|---|
| **Best TargetScan Prediction** | **3' UTR** | **Seed Match Position** | **Context Score** | **Rank** | **Total Sites** | **Percentile** |
| | ORF 107 | 29-35 | -0.222 | 1 | 18 | 100 |
| **Results for the same site predicted by PITA** | | | **ΔΔG** | **Rank** | **Total Sites** | **Percentile** |
| | | | 0.62 | 64 | 81 | 21.2 |
| **Best PITA Prediction** | | | | | | |
| **Best PITA Prediction** | **3' UTR** | **Seed Match Position** | **ΔΔG** | **Rank** | **Total Sites** | **Percentile** |
| | ORF 128-129 | 1305-1312 | -10.37 | 1 | 81 | 100 |
| **Results for the same site predicted by TargetScan** | | | **Context Score** | **Rank** | **Total Sites** | **Percentile** |
| | | | -0.123 | 7 | 18 | 64.7 |
| **TargetScan Prediction (from 10 top ranked) with best corresponding PITA Prediction** | | | | | | |

```
ORF_128-129 3' UTR      5' ACAUGUCCAAGGUCGGGGUGCUG 3'
MD11776 5' miRNA        3'  GUUCAAGAUGCCCUCCCACGAC
```

**Figure 7.6 Alignment of MD11776 5' miRNA to its most likely CyHV-3 target site (in 3' UTR of ORF 128-129) based on data in Table 7.3**

Seed matches are highlighted in red. Other elements of the 7mer-m8 target site are highlighted in green (nucleotide on target site base pairing to position 8 on miRNA). All base pairing between the miRNA and its target site are highlighted in bold.

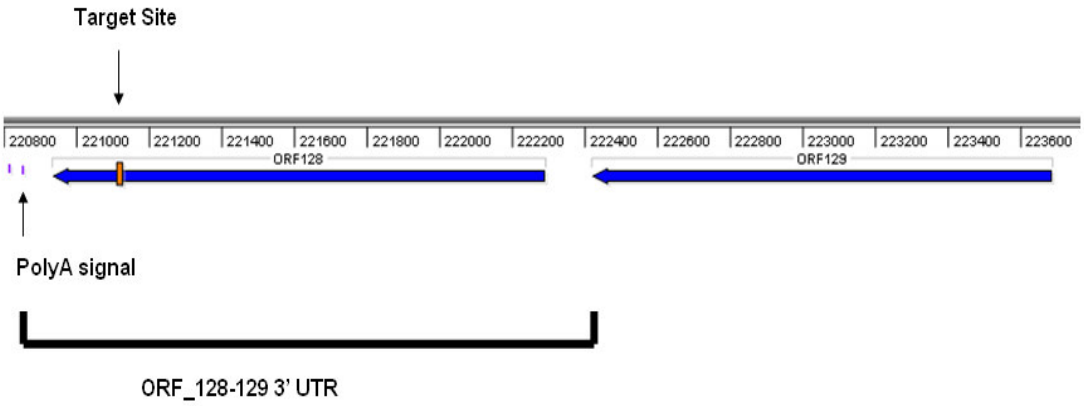



**Figure 7.7 Genomic location of the most likely CyHV-3 target site for MD11776 5'miRNA based on the data in Table 7.3**

## 7.4 Predicted CyHV-3 targets for the 3' arm miRNA from precursor MD11776

TargetScan and PITA predicted 63 and 231 potential target sites respectively for this miRNA (details of these sites are given in Supplementary File 7.2 T.10-T.11). The top ranked site from TargetScan was in the 3' UTR of ORF84. This was also the 2nd highest ranked prediction made by PITA. Conversely the highest ranked site in the PITA prediction is outside the top 10 ranked sites in TargetScan (Table 7.4). The close correlation between the two methods regarding the target site in ORF84 indicates that it is a much higher quality prediction. The seed match is an 8mer-1a and occurs near the end of the 3' UTR of ORF87 at position 2291-2298. Bases at positions 9, 14-16 and 21 on the miRNA are also complementary to the 3' UTR. This extra base pairing outside the seed region is based on the prediction of an asymmetric bulge forming in a miRNA-3'UTR duplex (Figure 7.8) The 3' UTR appears to be shared with ORF86 and 85 and overlaps with the end of the coding region of ORF84 (Figure 7.9).

**Table 7.4 Comparison of highest ranking TargetScan and PITA results for MD11776 3' miRNA target site prediction**

Due to its high ranking in both sets of results the target site in the 3' UTR of ORF 84-85 was identified as the most likely CyHV-3 target site for the MD11776 3' miRNA.

| Best TargetScan Prediction | | | | | | |
|---|---|---|---|---|---|---|
| **Best TargetScan Prediction** | **3' UTR** | **Seed Match Position** | **Context Score** | **Rank** | **Total Sites** | **Percentile Rank** |
| | ORF_84-85 | 2291-2298 | -0.319 | 1 | 63 | 100 |
| **Results for the same site predicted by PITA** | | | **Δ ΔG** | **Rank** | **Total Sites** | **Percentile** |
| | | | -24.55 | 2 | 231 | 98.2 |
| **Best PITA Prediction** | | | | | | |
| **Best PITA Prediction** | **3' UTR** | **Seed Match Position** | **Δ ΔG** | **Rank** | **Total Sites** | **Percentile** |
| | ORF_66-67 | 365-371 | -25.61 | 1 | 231 | 100 |
| **Results for the same site predicted by PITA** | | | **Context Score** | **Rank** | **Total Sites** | **Percentile** |
| | | | -0.044 | 17 | 63 | 74.1 |

```
ORF 84-85 3'UTR      5'   CACCAACCAGUUCGUGCAGGCCA  3'
MD11776 3' miRNA     3'   CGGUUGUGGGAGUG--CGUCCGGA  5'
```

**Figure 7.8 Alignment of MD11776 3' miRNA to its most likely CyHV-3 target site (in the 3' UTR of ORF 84-85); based on data in Table 7.4**

Seed matches are highlighted in red. Other elements of the 8mer-1a target site are highlighted in green (nucleotide on target site base pairing to position 8 on miRNA) and purple (A on target site opposite position 1 on miRNA). All base pairing between the miRNA and target- site are highlighted in bold.

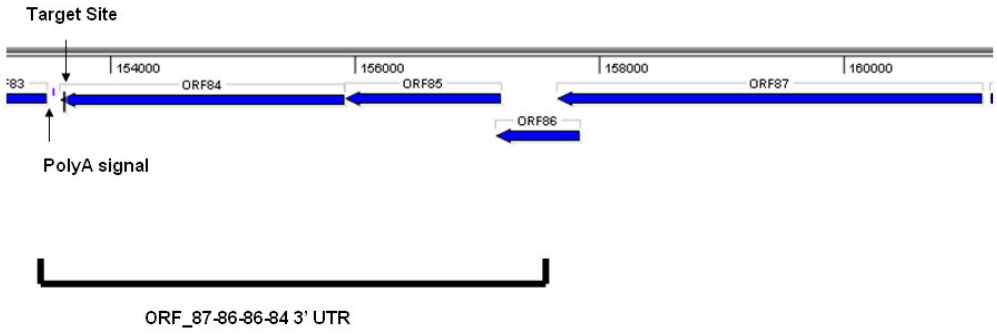



**Figure 7.9 Genomic location of the most likely Cyhv-3 target site for MD11776 3'miRNA based on data from Table 7.4**

# 8 Discussion

The aims of this study were to investigate the miRNA coding potential of CyHV-3. This investigation was conducted in a stepwise manner through a series of interdependent experiments. Each experiment was designed to closely scrutinize observations from the preceding experiment in order to support or oppose tentative conclusions that were inferred and to act as the basis for proceeding with further investigations. The results of these experiments are discussed below.

## 8.1 De Novo Prediction of Pre-miRNAs on the CyHV-3 genome

Most viral miRNAs are not evolutionarily conserved. Where viral miRNA conservation or sequence similarity does occur, it is only confined to a few miRNAs in closely related viral species. Therefore successful prediction of novel viral miRNAs requires the use of *de novo* miRNA prediction methods. The principles of *de novo* miRNA prediction (and descriptions of several such methods) are outlined in Section 1.7.1.1. VMir was specifically designed for *de novo* prediction of pre-miRNAs in viral genomes (Grundhoff et al., 2006) and due to its free availability, informative output and successful track record it was chosen for this study. This method is outlined in more detail in Section 1.7.1.1 .

In order to distinguish predicted CyHV-3 pre-miRNAs most likely to be genuine from background noise, cut-off values were established for the three most relevant attributes associated with these predictions which included predicted pre-miRNA size, VMir-score and relative window count (WC). It was important that the cut-off values used were relevant. Appropriate cut-off values for size range were easily established based on the size range of *bone-fide* pre-miRNAs. However, VMir score and WC are attributes that exist entirely within the context of VMir analysis, and although it was apparent that the higher these values were, the more "pre-miRNA-like" the prediction was, it was not apparent what lower limit cut-off values should be used. To avoid selecting arbitrary cut-off values, other viral genomes known to encode bone-fide miRNAs were also analyzed in order to see how known pre-miRNAs were interpreted in terms of VMir-score and WC, thus facilitating the

establishment of relevant cut-off values for these attributes for the CyHV-3 predicted pre-miRNAs.

Elimination of candidates based on VMir Score and WC cut-off values resulted in elimination of ~99% of predictions, leaving 303 predictions. Further elimination based on genomic location (i.e. elimination of those predicted to occur in ORFs) and MiPred classification resulted in a final list of 155 predicted CyHV-3 pre-miRNA. These results suggested that the CyHV-3 genome contained 155 sites that when theoretically transcribed into RNA could give rise to secondary structures that were consistent with pre-miRNAs in terms of structure and stability.

The assessment of the CyHV-3 genome in this manner was an important pre-requisite to experimental work. Although miRNA genes appear to be quite prevalent in HVs compared to other kinds of viruses, some HV genomes such as HHV-3 (VSV), HHV-6, and HHV-7 have in fact been predicted to lack any probable pre-miRNAs unlike others such as HSV-1, HCMV, EBV, KSHV and MHV68 (Pfeffer et al., 2005b). To date there have been no experimental investigations into the existence of HHV-7 encoded miRNAs and deep sequencing has failed to detect any HHV-3 miRNAs (Umbach et al., 2009a). Intriguingly, four pre-miRNAs were reported in HHV-6 (Tuddenham et al., 2012) although they are no longer present in the latest release of miRBase (Table 1.4). Previous pre-miRNA predictions in these three genomes were made using a method referred to as the Pfeffer-SVM-Method, which is not publicly available. As expected, the VMir prediction statistics from these three genomes (HHV-3, HHV-6 and HHV-7) showed significant differences to those from the other 6 viruses known to encode miRNAs. Applying the same cut-off values to predictions on these 3 genomes resulted in an average of 0.61 pre-miRNA predictions per kbp (data not shown) compared to an average of 0.96 for the six other viruses and a value of 1.03 for the CyHV-3 genome (Table 3.4). Importantly, this showed that the VMir prediction statistics for the CyHV-3 genome were in line with those of the six other viral genomes known to encode pre-miRNAs and that this was very different picture from other viral genomes previously predicted not to encode pre-miRNAs. Overall, this provided a solid theoretical basis to proceed with experimental investigations into the existence of CyHV-3 encoded miRNAs.

The cut-off values used allowed the detection of 83.1% of the 77 known miRNAs in the other six viral genomes analyzed. Based on the values for the prediction attributes for the seven CyHV-3 miRNAs proposed here (Table 8.1), four out of the seven were present in the remaining 303 filtered predictions resulting in a sensitivity of 57.1% for the CyHV-3 predictions when using these cut-off values (incidentally, all seven would have passed the subsequent genomic location and MiPred elimination steps). All seven exceeded the VMir-score cut-off value. Two of them, MD11410 and MR5057, did not exceed the WC cut-off value. Although MD11410 was close with a value of 37, the WC value for MR5057 was quite low and in stark contrast to the others (the average WC value for these six was 43.5). In spite of this, out of the seven proposed miRNAs, MR5057 has the strongest experimental evidence in support of its annotation as a pre-miRNA. MD11704 was not in the original 303 predictions either due to its unusually large size. In hindsight, there should have been no upper limit put on pre-miRNA size. In theory the length of a pre-miRNA stem should not make a difference in terms of its ability to as to act as substrate for Drosha while still part of the pri-miRNA. If the upper length limit is ignored, five of the seven proposed miRNAs would have been predicted, resulting in a sensitivity of 71.4% which is closer to the sensitivity that these cut-off values gave for the other 77 viral pre-miRNAs (which would remain at 83.1% even after removing the upper limit on pre-miRNA size). Overall, the prediction attributes for the 7 proposed CyHV-3 pre-miRNAs are very similar to those of the 77 other viral pre-miRNAs from the other six viruses (Table 8.1), thus further supporting their annotation as genuine pre-miRNAs.

**Table 8.1 Summary of prediction attributes of seven proposed CyHV-3 pre-miRNAs**

| Attribute | VMir Score | WC | MHP Size | Pass cut-off values? |
|---|---|---|---|---|
| **Cut-off values** | **>112** | **>40** | **45-200 nt** | |
| MD1111 | 229.0 | 42 | 172 | **Yes** |
| MD9812 | 224.8 | 46 | 107 | **Yes** |
| MD11410 | 141.8 | 37 | 75 | No (WC too low) |
| MD11704 | 207.0 | 44 | 250 | No (MHP size too big) |
| MD11776 | 210.2 | 45 | 136 | **Yes** |
| MR5057 | 114.2 | 2 | 60 | No (WC too low) |
| MR5075 | 165.4 | 47 | 141 | **Yes** |
| Average | **184.62** | **37.6** | **134.4** | |
| Average for 77 pre-miRNAs from other six viruses* | **186.23** | **39.9** | **101.9** | |

*Six other viral genomes were EBV, MDV-1, HSV-1, KSHV, MGHV68 and HCMV. Further details of the genomes used and predictions are is Supplementary File 3.1.

The CyHV-1 and CyHV-2 genomes had not yet been sequenced when this study commenced, thus pre-miRNA prediction was done using VMir. The recent sequencing of these two genomes now makes it possible to potentially predict additional novel miRNAs that are shared (or at least show some sequence similarity) between these two genomes and CyHV-3. Several methods use pair-wise genome alignments to improve predictions by identifying predicted pre-miRNAs that display stretches of sequence similarity with predicted pre-miRNAs in other genomes (i.e. possible conserved miRNAs, seed regions etc.). Such sequence similarities between CyHV-1, CyHV-2 and CyHV-3 demonstrated in this study using BLAST (i.e. MD1111, MD11776 and MD11704 in CyHV-1 and 2; Section 6). This suggests that the use of a *de novo* pre-miRNA prediction approach that takes into account potential low level sequence homology between predicted pre-miRNAs in these viruses may result in the identification of novel pre-miRNA candidates. Methods that may be suitable to carry out this kind of analysis include EVOfold (Pedersen et al., 2006), miRseeker (Lai et al., 2003), MIRFINDER (Bonnet et al., 2004a)

## 8.2 Experimental evaluation of pre-miRNA predictions

The primary samples used in experiments to identify CyHV-3 encoded miRNAs consisted of RNA from *in vitro* lytic infections. This was because RNA from *in vivo* latent infections was not considered suitable for the identification of novel viral miRNAs. This decision was made on the basis that other studies have shown that viral miRNAs are hard to detect in latently infected tissues. Typically, not all known miRNAs can be detected in such samples. This may be because there are lower proportions of infected cells in tissue samples compared to *in vitro* samples resulting in a higher level of host background noise in the sequencing data, and so less viral miRNAs would be featured in the data. This means that viral miRNAs that are identified in such samples typically have very low read counts compared to *in vitro* infections (Amen and Griffiths, 2011; Umbach et al., 2009b). While deep sequencing of small RNAs from latent tissue samples may be suitable for cataloguing the presence of known viral miRNAs, due to the high background noise and the possible absence of associated low abundance transcripts (i.e. minor forms on opposite arm of pre-miRNAs and isomiRs) used to support the presence of novel miRNAs, *in vivo* samples (both lytic and latent) were deemed not suitable for the identification of novel viral miRNAs. Due to the fact that higher proportions of cells are infected during *in vitro* infections, RNA from such experiments is more enriched for viral transcripts in general, making *in vitro* infections much more suitable for the identification of novel viral miRNAs.

### 8.2.1 Deep sequencing of small RNAs from *in vitro* CyHV-3 infections

In order to initially identify putative CyHV-3 encoded miRNAs, the initial experimental stages of this study involved the indirect identification of CyHV-3 pre-miRNAs. Specifically, this involved analyzing deep sequencing data for small RNAs (17-25 nt) that mapped to the CyHV-3 genome in a manner that was consistent with the model of miRNA biogenesis. This allowed putative miRNAs to be distinguished from non-miRNA transcripts also present in the same samples. Small RNAs were sequenced from two different CyHV-3 infections. This was important because unlike random degradation products, small RNAs that are generated through relatively specific and repeatable processes such as miRNA biogenesis should show more

consistency between different infections when compared to random degradation products.

The percentage of total reads that were of CyHV-3 origin was ~ 2 and ~ 7% in the H361 and N076 infections respectively. While this percentage was quite low in both infections, similar results have been observed elsewhere. For example, viral small RNAs constituted ~ 2% of reads sequenced from marmoset T-cells infected with HVS (Cazalla et al., 2011) and MCMV reads from infected bone marrow macrophages represented as little as 0.5% of total reads. However the same study with MCMV observed that ~ 30% of transcripts sequenced from mouse embryonic fibroblasts infected with MCMV were of viral origin (Buck et al., 2007), demonstrating that although lytic infections were established in both cells types, differences in experimental conditions can have a significant effect on the proportion of viral transcripts present. In addition to cell type, several other parameters such as the amount of virus used, temperature, time and medium composition may also have effects on the proportion of viral transcripts present in RNA samples. In this regard, similar variations were also found between different HCMV *in vitro* infection conditions (Dunn et al., 2005; Pfeffer et al., 2005b). Even though CyHV-3 culture conditions were kept constant between the two infections used in this study, both flasks were inoculated with significantly different amounts of different isolates, comprising 1 mL of 2.7x $10^4$ $TCID_{50}$ / mL and 1 mL 5.0x $10^5$ $TCID_{50}$ / mL of the H361 and N076 isolates respectivley. Also, RNA was also extracted at different time points based on CPE progression. RNA was extracted from cells only when visible CPE was widespread in the infected flasks. This was done in order to ensure that lytic replication (and hence viral gene transcription) was taking place in the maximum amount of cells, thus reducing cellular RNA background noise. However, it was reasoned that it was important not to allow CPE to progress too far as increased levels of cellular RNA turnover in infected cells (Glaunsinger and Ganem, 2006) may have the effect of increasing non-miRNA background noise in the 17-25 nt RNA size range. The N076 infection took place in smaller flask than the H361 infection ($25cm^2$ vs. $75cm^2$), therefore viral concentrations in the medium were higher by comparison and CPE had progressed to what was deemed to be a satisfactory level at an earlier time point post inoculation (14 vs. 18 d.p.i.). In

summary it is likely that this different sampling time-point, the different quantities of virus used and different isolates used are responsible for the differences in the proportion of viral transcripts observed in the deep sequencing data from both of these infections.

The 17-25nt RNA samples should have been specifically enriched for miRNAs and as expected the majority of the reads, ~ 58 % and ~ 76 % from H361 and N076 infections respectively, consisted of known host miRNAs. Interestingly a significant proportion of the remainder consisted of novel host miRNAs and previously undetected minor forms derived from currently known host pre-miRNAs (data not shown as it relates to a separate study). This loosely correlated with the proportions of reads derived from seven newly discovered CyHV-3 pre-miRNAs, which (including associated isomiRs) represented 42% and 85% of all CyHV-3 reads from the H361 and N076 isolates respectively.

These seven pre-miRNAs were discovered through a combination of manual interpretation of the sequencing data (i.e. non-automated approaches focusing on a small subset of highly abundant transcripts mapping to predicted pre-miRNAs in non-coding regions) followed by the use of automated methods in interpretation of the same data (i.e. analysis of the complete dataset using special algorithms to identify CyHV-3 pre-miRNAs). The findings from both of these approaches are discussed separately in the following sections.

### 8.2.2 Non-automated identification of miRNAs from deep sequencing data

In order for a small RNA to be initially considered as a putative miRNA in this study (i) it had to map to a predicted pre-miRNA from a non-coding region, (ii) be accompanied by a corresponding major/minor form in both infections and (iii) at least one of them had to have a read count >10 in at least one infection. Due to the read count criteria, only six of the seven CyHV-3 pre-miRNAs were initially identified using the non-automated approach (MD11704 was not). All miRNAs identified from these six CyHV-3 pre-miRNAs were present in both infections;

however, some were not always the dominant small RNA at these loci i.e. different isomiRs of the same miRNAs dominant in both infections. This was the case for 4 out of the 12 miRNAs from these six CyHV-3 pre-miRNAs. However, this is a common observation in deep sequencing miRNA data. Often the dominant miRNAs at specific loci are found to differ slightly from the reference miRNA recorded in miRBase and are mainly offset at the 3' end. This was found to be the case with miRNAs from several viruses e.g. HCMV (Meshesha et al., 2012), MGHV (Reese et al., 2010), KSHV (Umbach and Cullen, 2010) HSV-1 (Umbach et al., 2009a) and EBV (Chen et al., 2010) and in other organisms (Kuchenbauer et al., 2008; Morin et al., 2008; Reid et al., 2008) and may be regulated in a tissue-specific manner (Fernandez-Valverde et al., 2010; Lee et al., 2010). Out of the 4 CyHV-3 miRNAs that differed between both infections, only one of them differed at the 5' end (3'arm miRNA from MD11410) and this was only offset by 1 nt. This one case represented 8% of all CyHV-3 miRNAs. This rate was low relative to that of putative miRNAs from other pre-miRNA candidates later determined not to be genuine, in which case 17% of them were found to have 5' end variations, the majority of which represented cases where the 5' end differed by >1 nt. This demonstrated that the 5' ends had more of a tendency to remain unchanged among the dominant isomiRs observed in both infections compared to other non-miRNA small RNAs present in the same samples. This makes sense as the 5' ends of miRNAs contain the seed regions (from positions 2-7) and are important from a functional perspective, therefore this part of the miRNA would need to remain relatively unchanged in the dominant isomiRs (instead they differ at the 3' ends). Conversely, for non-miRNA small RNAs most likely generated through random degradation, there would be no reason for them to display such a bias at their 5' ends. In this study, if such differences were observed between infections, in order to be considered a high probability pre-miRNA, the dominant isomiRs (i.e. miRNAs) from both arms were only allowed to differ by a maximum of +/- 1nt at the 5'end (pre-miRNA identification criteria). Although it has been shown elsewhere that the dominant isomiRs may differ by up to 4 nt at their 5' ends in different infections e.g. the MGHV miRNA mgv-miR-M1-2-3p (Reese et al., 2010), a maximum allowance of +/- 1nt of a difference was still maintained in this study. This may be a slightly conservative approach to novel pre-miRNA identification but it did ensure selection of only those pre-miRNA candidates that

showed the highest degrees of consistency (in terms of putative miRNA processing) between the two infections.

Of the six genuine CyHV-3 pre-miRNAs identified, the mature miRNAs representing the major and minor forms were consistent between both infections with one exception, the miRNAs from MD1111. In that case the major form switched from being the 3' arm miRNA in the H361 infection to the 5' arm in the N076 infection. This "switching" between major and minor forms is not uncommon and has been shown to occur in other organisms. Such changes may occur at certain developmental stages (Jagadeeswaran et al., 2010) and in some cases may also be determined by tissue type (Ro et al., 2007). Such switches have also been shown to occur at different stages of viral life cycles, for example with MGHV the 3' arm miRNA from the pre-miRNA mgv-mir-M1-2 was shown to be the major form in latently infected B cells (Pfeffer et al., 2005b) but switched to the 5' arm in lytically infected fibroblasts (Reese et al., 2010). The same has also been shown to occur between two different lytic infections as demonstrated with HCMV pre-miRNA hcmv-miR-UL22A which switched from 5' arm (Pfeffer et al., 2005b) to the 3' arm (Meshesha et al., 2012). Although such switches in major forms between lytic and latent infections may occur as part of overall changes in viral transcription programs between the two stages - possibly highlighting different roles for the same pre-miRNA in each stage - it is difficult to imagine why such shifts would occur within the same stages of viral life cycles. However this may be down to experimental conditions, it stands to reason that if cellular pre-miRNAs regularly switch major forms in different tissue types, the same may also occur when different studies use different cell types, which often happens.

Four out of the six genuine CyHV-3 pre-miRNAs identified using this approach were also found to have additional transcripts derived from stems of the same pre-miRNAs adjacent to the proposed miRNAs. This suggested the presence of viral moRNAs. These are a recently discovered class of small RNAs originally described by Shi et al (2009). Since then evidence has emerged to suggest their presence in other organisms (Langenberger et al., 2009) and interestingly in several other HVs such as HSV-1, HSV-2 (Jurak et al., 2010), KSHV (Lin et al., 2010), HCMV (Meshesha et al., 2012), RRV (Umbach et al., 2010), PRV (Wu et al., 2012) and

AHV-1 (Yao et al., 2012). In fact some small RNAs originally identified as viral miRNAs have been re-classified as pre-miRNAs due to their positions in stems. Examples of these former PRV (Wu et al., 2012) and RRV (Umbach et al., 2010) miRNAs. Based on its position on the BoHV-1 pre-miRNA bhv1-mir-B2 (Glazov et al., 2010), the currently annotated "5' miRNA" the may be more correctly described as a moRNA, however, this has not been acknowledged in any literature to date. In most cases the 3' ends of the 5' moRNAs and the 5' ends of the 3' moRNAs occur immediately adjacent to the mRNAs on the same arms of the pre-miRNA (Figure 4.10) suggesting that both classes of small RNA arise from a common pre-miRNA cleavage event (Berezikov et al., 2011; Bortoluzzi et al., 2012; Shi et al., 2009). It is not clear whether the latter is carried out by Drosha or Dicer. Some studies have suggested that either enzyme can process moRNAs from pre-miRNA and that this may depend on the cell type as exemplified by a model suggested for the biogenesis of PRV moRNAs (Wu et al., 2012). Some moRNAs have been shown to span annotated Drosha cleavage sites and overlap with miRNAs by up to 8 nt suggesting that some moRNAs may arise from non-canonical Drosha cleavage followed by mutually exclusive miRNA or moRNA processing from the same pre-miRNAs (Bortoluzzi et al., 2012). The 3' arm moRNA from MD1111 overlaps the adjacent miRNA by 1 nt and may be processed in this fashion.

Even though the samples sequenced were significantly enriched for CyHV-3 miRNAs, 58% and 15 % CyHV-3 transcripts from non-coding regions in the H361 and N076 infections respectively did *not* represent miRNAs or isomiRs. In particular there were significant numbers of highly enriched CyHV-3 small RNAs that mapped to predicted pre-miRNAs in 3' UTRs. Although viral miRNAs have been found to occur in the 3' UTRs of other viruses such as KSHV (Cai et al., 2005) and MCHV (Buck et al., 2007), none of these CyHV-3 transcripts from 3' UTRs were found to be miRNAs. This demonstrates that although the identification of highly abundant small RNAs mapping to predicted pre-miRNAs in non-coding regions is a good starting point for identifying novel miRNA genes in deep sequencing data, it is important to examine other features of pre-miRNA candidates in more detail in order to find further evidence to support their annotation as pre-miRNAs.

The structural characteristics of pre-miRNAs are quite important. It is these characteristics that distinguish them from random hairpins and allows them to act as substrates for Drosha and Dicer (Han et al., 2006; Tsutsumi et al., 2011). In order to classify the VMir predicted CyHV-3 pre-miRNA structure as either pre-miRNA-like or not they were all analyzed by MiPred and the CSHMM-Method. In almost all cases there were several variations of each pre-miRNA, all differing in stem length and absolute WC value. The longest variant was referred to as the MHP and any shorter variants were referred to as SHPs. Different variants of the same pre-miRNA can be classified differently by pre-miRNA classifiers such as MiPred. This was evident after further assessment of VMir predictions in Section 3.4 (Supplementary File 3.1 T.10). It was important therefore that the most relevant variant was analysed. Naturally, the selection process focused on variants that were most likely to form stable hairpin structures. Generally shorter variants had higher absolute WCs, meaning that they had more of a tendency to form in many more analysis windows than longer variants, indicating that they were more stable within the local sequence context. This does not relate to structural stability (MFE) which was generally much lower (indicating greater stability) for longer pre-miRNA variants. However, high structural stability displayed by longer variants was not as relevant if there was less of a chance of them forming in the local sequence context (i.e. as they had a tendency to have lower absolute WCs). It was important that the local sequence context was taken into account, even though there is currently no information on the pri-miRNA sequences for any CyHV-3 pre-miRNAs, such flanking sequences will almost certainly be present alongside the pre-miRNA on the pri-miRNA transcript. Therefore it was best to analyse the variant that was more likely to form within this local sequence context. Hence, for each pre-miRNA candidate, it was the most stable variant (in terms of absolute WC) containing the full-length miRNAs that was analyzed. Although longer pre-miRNAs are more thermodynamically stable, for all high-probability CyHV-3 pre-miRNA candidates these shorter variants were still found to have low MFE values (<-25 kcal/mol) that were also found to be statistically significant ($p < 0.05$) by MiPred (Table 4.6 and Supplementary File 4.1 T.5). In addition the same CyHV-3 pre-miRNAs were found to have local contiguous structure-sequence composition that was consistent with that of *bone-fide* pre-miRNAs by both MiPred and the CSHMM-Method, (these methods are outlined in Section 1.7.1.2).

In cases where pre-miRNAs were also predicted to give rise to moRNAs, the most stable variant (in terms of absolute-WC) also containing full length moRNAs was analysed. This was on the basis that if these are genuine moRNAs, these longer predicted pre-miRNA candidate variants (containing these additional moRNA sequences) should inevitably exist at some stage of the miRNA biogenesis process. Indeed two different variants of the same pre-miRNA (normal and longer versions containing unprocessed moRNAs) have been detected by northern blotting (Wu et al., 2012). If this is the case, these longer variants would also be expected to be classified as "real pre-miRNAs" by pre-miRNA classifiers on the basis that they should be recognised by components involved in the miRNA biogenesis process. (Bortoluzzi et al., 2012). It is possible that such longer variants act as substrates for either Drosha or Dicer some stage at least for initiation of miRNA biogenesis if not for the purposes of moRNA processing. For this reason it was deemed that the analysis of longer variants was relevant, where applicable. Moreover this may even be a more accurate representation of the pre-miRNA structure, or represent a temporary intermediate form. Interestingly, additional analysis of longer variants of high probability CyHV-3 pre-miRNAs also gave the same results as shorter variants (i.e. all were classified as real pre-miRNAs by both methods with low and statistically significant MFE). By contrast, 73% of other pre-miRNA candidates later deemed not to be genuine pre-miRNAs were *not* classified as pre-miRNAs by both methods.

Processing of pre-miRNAs by Drosha and Dicer typically results in a mature miRNA duplex with a 2 nt 3' overhang. This is characteristic of the activities of all RNase III enzymes (Ji, 2008). In order to see if the putative miRNAs from pre-miRNA candidates conformed to this characteristic, all the predicted miRNA duplexes were assessed by simply highlighting their positions on the predicted pre-miRNA structure. However because of the fact that the most dominant isomiR on each arm is designated as the mature miRNA, this creates a small problem when assessing miRNA duplex structures.

All miRNAs and isomiRs are initially part of a miRNA duplex with 2 nt 3' end overhangs. However, if a particular small RNA is identified as the dominant isomiR

(i.e. miRNA) on one arm of a given pre-miRNA, this does not necessarily mean that its original partner strand will be detected as the dominant isomiR mapping to the opposite arm of the pre-miRNA (in the sequencing data). Instead the dominant isomiR on the opposite arm may originate from a subset of duplexes that were cleaved at slightly different positions (i.e. isomiRs of the miRNA). This could be down to several reasons, such as differences in efficiency of RISC incorporation (i.e. there may be a preference for some isomiRs over others) quantitative inaccuracies intrinsic to deep sequencing (i.e. a bias for some sequences over others). Thus, inspection of the predicted mature miRNA duplex structure by highlighting the positions of dominant isomiRs in the pre-miRNA structure (as used in this study) may not always show a mature miRNA duplex with 2 nt 3' overhangs. This will only occur if the isomiRs designated as the mature miRNAs (i.e. dominant isomiRs) on both strands are actually generated by the exact same cutting events. Therefore, to allow for this, predicted miRNA duplexes were instead assessed for the presence of 1-4 nt 3' overhangs. This allowed for small deviations from the model of miRNA biogenesis while still allowing the identification of a strong bias towards the presence of overhangs on the 3' ends. In order for a pre-miRNA to be considered a likely pre-miRNA in this study, bias towards the presence of 3' overhangs had to be observed in both infections (one blunt end was allowed but no 5' overhangs). All of the miRNA duplexes from the six CyHV-3 pre-miRNAs met these criteria; in contrast, only 20% of the other pre-miRNA candidates that were later determined not to be genuine also met these criteria.

Based on miRNA duplexes, the ends of all six CyHV-3 pre-miRNA hairpins displayed 3' overhangs on both ends with the exception of MR5075 which consistently displayed onle blunt end in both infections (Figure 4.10). Interestingly the 3' end of the 5' miRNA and 5' end of the 3' miRNA (i.e. the Dicer cleavage sites) are actually situated on the loop (both overlap the loop by 1 nt) as opposed to the stem of the pre-miRNA. This could represent a non-canonical cleavage site for Dicer resulting in non-canonical processing. In addition, there are other examples of annotated pre-miRNAs in miRBase with mature miRNA duplexes that are predicted to have blunt ends, for example in HSV-1 the pre-miRNA hsv1-mir-H5 (Jurak et al., 2010), indicating that there are precedents for this phenomenon.

In addition to mature miRNAs, the fact that deep sequencing facilitated the comprehensive profiling of other less abundant small RNAs derived from these predicted pre-miRNAs (i.e. isomiRs and other pre-miRNA derivatives such as moRNAs) was also quite important and very useful in terms of novel pre-miRNA identification. This is because it is the characteristics of these additional transcripts that provide the additional information that can be used to either support or oppose the annotation of identified pre-miRNA candidates as being genuine or not.

Unlike small RNAs that represent random degradation products from mRNA transcripts, small RNAs derived from miRNA or moRNA biogenesis are generated from specific cleavage events. This is reflected in their alignment signatures (Figure 1.12 (a)). This "miRNA-like" alignment signature was noted elsewhere to act as an important way to support high confidence miRNA annotation (Kozomara and Griffiths-Jones, 2011). In fact in the present study this was the most effective way of differentiating between CyHV-3 pre-miRNAs and pre-miRNA candidates that were later deemed not to be genuine. Inspection of signatures showed that all of the six genuine CyHV-3 pre-miRNAs showed this kind of signature, but only one (7%) of the other fifteen pre-miRNA candidates did. Although this is something that is quite important and useful, in most other studies concerned with novel viral miRNA identification, the alignment signatures of small RNAs to viral pre-miRNAs were either not commented on at all or not provided. Notable exceptions to this trend are studies published by Zhu et al (2010b) and Wu et al (2012).

By its nature, the kind of pattern described by Kozomara and Griffiths-Jones (2011) is hard to define precisely. In addition, these kinds of alignment signatures will vary for different pre-miRNAs or even for the same pre-miRNA from different samples. In some situations, trying to describe a given alignment pattern as either "like" or "unlike" what is recommended in the reference above can be subjective, although this was not the case for all alignments in this study. Seven of the twenty-one pre-miRNA candidates had alignment signatures that were distinctly miRNA-like. This included all of the six genuine pre-miRNA candidates. However the signatures for MD1111 and MR5075 deviated to a small degree from the "ideal" signatures (Supplementary File 4.2). The signature for MD1111 was quite elongated. Individual

stacks representing miRNAs and moRNAs are not as well defined due to some overlap between transcripts in the miRNA and moRNA stacks (overlaps were confined to low abundance isomiRs/moRNA isomers). In the alignment signature for MR5075 the stacks representing the miRNAs and moRNAs are more clearly defined than those of MD1111 (as they are higher) however again there is a similar overlap between the 5'moRNA and 5' miRNA stacks. There is also overlap between some low abundance isomiRs in both miRNA stacks with these overlaps spanning the terminal loop sequence. In the signatures for both MD1111 and MR5075 there are also additional low level transcripts mapping to the pre-miRNA candidates on the 5' side of the 5' moRNAs. It is possible that these just represent degradation fragments of the pre-miRNA or even the original primary transcript. However, overall alignment signatures for both MD1111 and MR5075 were still considered miRNA-like due to the fact that instances of moRNAs overlapping miRNAs (by up to 8 nt) and additional transcripts adjacent to moRNAs have been observed elsewhere on *bone-fide* pre-miRNAs (Bortoluzzi et al., 2012; Zhang et al., 2010).

Two out of the six CyHV-3 pre-miRNAs (MD1111 and MD11776) also showed additional low count reads mapping to the terminal loop. It is possible that this may simply represent pre-miRNA degraded fragments of the pre-miRNA loop and thus more likely to be present in the data as consequence of shear depth of coverage rather than being of any biological relevance. It may also be argued that detection of loop fragments helps infer the existence of a complete stem-loop structure (Friedländer et al., 2008). Pre-miRNA loop fragments have been identified at low level before from other viral pre-miRNAs, for example in KSHV (Umbach and Cullen, 2010). Strangely and quite counter intuitively, sometimes these loop fragments can actually be more abundant than the mature miRNAs derived from the same pre-miRNA, for example in PRV (Wu et al., 2012). Mature miRNAs accumulate to higher levels than pre-miRNA fragments due to incorporation into RISC (protecting them from immediate degradation). The high level of loop fragments in the case of PRV suggested that it may also be incorporated into the RISC. Indeed it has been demonstrated elsewhere that such events do occur, and that some pre-miRNAs can give rise to three distinct types of functional miRNAs: major form, minor form and the newly described "loop-miR" (Winter et al., 2013).

In recent years it has become apparent that isomiRs tend to display more much more 3' end heterogeneity compared to 5' end heterogeneity (Ruby et al., 2007). More than likely, most isomiRs retain the same functions as the miRNA as the seed regions remain unaltered and indeed this may be the very reason why there is a strong bias towards 3' end heterogeneity. Viral isomiRs also tend show this characteristic and this has been noted in several studies (Chen et al., 2010; Meshesha et al., 2012; Riley et al., 2010; Umbach and Cullen, 2010; Wu et al., 2012; Yao et al., 2012; Zhu et al., 2010b). Although noted, it has generally not been presented in any great detail in any publications. Instead these observations are commonly described in terms of collective end heterogeneity for all isomiRs recovered during sequencing (Meshesha et al., 2012; Umbach and Cullen, 2010), shown for one pre-miRNA (Wu et al., 2012) or shown for all without any corresponding values (Amen and Griffiths, 2011; Riley et al., 2010; Zhu et al., 2010b). In the present study it was reasoned that it would be useful to quantitatively gauge the amount of 5' and 3' end heterogeneity displayed by isomiRs of each individual miRNA, to establish if there was bias towards 3' end heterogeneity among individual sets of isomiRs for each miRNA. This was done on the basis that it could be used as evidence to support the annotation of novel CyHV-3 miRNAs. IsomiRs representing <0.1% of the combined miRNA and isomiR read count (for the miRNA in question) were less likely to be biologically relevant and were eliminated from this analysis. In order to avoid the risk of including partially degraded isomiRs which could possibly skew the results, transcripts <19nt in length were also eliminated from analysis as per Riley et al (2010). Only isomiRs from the seven pre-miRNAs showing miRNA-like alignment signatures were analyzed.

Ratio-1 was used in order to compare the number of unique isomiRs in each set that displayed 5' end heterogeneity to the numbers that displayed 3' heterogeneity. Ratio-2 was used in order to compare the degree of heterogeneity in 5' ends to that of the 3'ends (in terms of average offsets on each end) for each set of isomiRs. Overall it was found that most isomiR sets were composed of isomiRs that were offset at the 3'end and in almost all cases the degree 3' heterogeneity was much more pronounced within each isomiR set. This suggested that there was a clear bias towards 3' end heterogeneity within each set of isomiRs. However, even if isomiRs displayed a higher degree of 3' end heterogeneity, it is possible that unique isomiRs displaying

5’ end heterogeneity (albeit a lower degree of heterogeneity) could be more numerous in read count, meaning that overall more 5’ end processing events deviated from the miRNA 5’ end position. In order to investigate this, a third ratio was also used in which the read counts of each unique isomiR were taken into consideration. (Ratio-3). This showed that 5’ processing was much more consistent within most sets of isomiRs. Although some did not show much difference between the two ends in this respect. One reason for this may be that in these cases most isomiRs display both 5’ and 3’ end heterogeneity. Therefore many of the same isomiRs contribute to the total read counts for both 5’ and 3’ end heterogeneity. Overall, the use of these ratios indicated that the most individual sets of isomiRs were biased towards displaying more 3’ end heterogeneity. This is very consistent with the characteristics of isomiRs of *bone-fide* miRNAs which also generally display a greater amount of 3’ end heterogeneity.

Unlike the other CyHV-3 pre-miRNAs, the isomiRs from MD1111 were biased towards displaying more 5’ end heterogeneity. Although it is more unusual, isomiRs of some miRNAs can show more 5’ end heterogeneity, examples of this include mmu-mir-341 in mice and hsa-mir-126 in humans (Lee et al., 2010). In addition the amount of MD1111 isomiRs that were offset at the 5’ ends generally only represented a small percentage of total reads in each infection (median of 2.3 % for each miRNA from each infection), indicating that the majority of 5’ end processing was consistent, although not as consistent as 3’ end processing. For other CyHV-3 miRNAs ~40% total reads showed 5’ heterogeneity (e.g. MD11776 and MR5075), however other known miRNAs have been shown to display similarly high proportions of 5’ heterogeneity such as the KSHV miRNA miR-K10-3p (Umbach and Cullen, 2010). The same study also observed more 3’ heterogeneity in miRNAs from the 5’ arms of pre-miRNAs. In such cases the 3’ ends are defined by Dicer processing. Interestingly, the results seen for CyHV-3 miRNAs in this study correlate with this. Values for Ratio-3 for 5’ miRNAs are on average 2 times higher than the same values for 3’ arm miRNAs (for which 3’ ends are processed by Drosha). It is important to note that there is an unusually large Ratio-3 value for the 5’ arm miRNA from MD9812 in the H361 infection, which could distort these results, however if this is eliminated then the average Ratio-3 values for 5’ arms are still marginally higher. Overall only one pre-miRNA candidate showed 5’ end heterogeneity in

>50% of reads processed from one of its miRNA loci. This was MR6201 which turned out to be the only pre-miRNA candidate analyzed that was later deemed not to be a genuine pre-miRNA.

.

The results from each stage of in depth analysis of the CyHV-3 pre-miRNA candidates were compiled and the observed characteristics for each CyHV-3 pre-miRNA candidate assessed in terms of how they conformed to the characteristics of *bone-fide* pre-miRNAs using the pre-miRNA identification criteria. This systematic approach to the identification of miRNA-like characteristics among pre-miRNA candidates showed a clear distinction between different candidates. It provided significant evidence to support the annotation of six high probability CyHV-3 pre-miRNAs as *bone-fide* pre-miRNAs. It also provided a solid experimental basis for describing some pre-miRNA candidates as low-probability pre-miRNAs.

Out of the six pre-miRNAs identified here four of them occur as part of two clusters. Furthermore the seventh pre-miRNA identified later in this study also occurs in one of these three clusters. This means that most of these pre-miRNAs share the same pri-miRNAs, although they may be processed at different rates accounting for differences in miRNA read counts. This is similar to the distribution patterns of pre-miRNAs in other viruses. In most viruses these genes are heavily clustered e.g. in EBV, KSHV although in some viruses these may be non-clustered e.g. HCMV.

#### **8.2.2.1** ***Automated identification of miRNAs from deep sequencing data***

The non-automated approach taken to identify CyHV-3 miRNAs focused on pre-miRNA candidates from non-coding regions with highly abundant putative miRNAs (read count >10). However these sequences only represented a small subset of unique reads present in the sequencing data from both infections. In order to explore the possibility that within the data, there were other CyHV-3 pre-miRNAs that produced miRNAs of lower abundance the entire data set was reanalysed using algorithms specifically developed to indentify miRNA signals within sequencing data. This also allowed the identification of pre-miRNA candidates within protein-coding genes. Although such candidates from protein-coding genes were ultimately not considered,

this presented an ideal way to verify that they showed a general lack of consistency and poor compliance with the characteristics of *bone-fide* pre-miRNAs thus supporting the elimination form analysis in the non-automated approach.

This strategy involved using two different methods to identify pre-miRNAs, namely miRDeep and Mireap (Section 4.1.4). Although quite useful, these kinds of methods are likely to give a lot of false positives. In addition, different results can be obtained from the same data depending on which method is used in analysis. Nonetheless, this observation may be used to improve predictions. The results from the comparison study carried out by Li et al. (2012) suggested that it was best to combine results from different methods to increase accuracy in novel miRNA prediction. They tested the performance of multiple methods in a task that involved prediction of known miRNAs from the same deep sequencing datasets. While none of the methods were able to identify all known miRNAs within each dataset, the vast majority of known miRNAs identified by all methods used. Only a very small minority of known miRNAs were not found by multiple methods. This indicated that in most cases, independent identification of pre-miRNAs by multiple methods acts to support predictions. Based on this, it was reasoned here that the best approach to automated identification of novel CyHV-3 pre-miRNAs was to carry out this analysis using more than one method.

As expected, analysis of the deep sequencing data using both methods mainly resulted in false positives. Only six pre-miRNA candidates (a fraction of the total number) were independently identified by both automated methods, thus these six were deemed the most likely to be genuine. Strikingly, five out of these six candidates were also included in the six high probability CyHV-3 pre-miRNAs identified earlier using the non-automated approach. This supported the theory that elimination of candidates *not* identified by both methods had the effect of retaining high quality pre-miRNA candidates only. This agreement between the automated and non-automated approaches indicated that the methods used to distinguish between low and high probability pre-miRNAs in both approaches was well-founded.

Despite the fact that the automated approaches identified similar amounts of (and in some cases more) pre-miRNA candidates in protein-coding regions compared to non-coding regions, only one candidate from a protein-coding region was identified by both automated methods and only one was found in both infections by the same method (Table 4.14). Ignoring the fact that these were from protein-coding regions, neither of these two pre-miRNA candidates were classified as "Real Pre-miRNAs" by MiPred or the CSHMM-Method (data not shown). This lack of repeated identification by the same automated method (i.e. in both infections) or independent identification by different automated methods (in either infection) supports the theory that miRNA-signals identified in protein-coding regions are more likely to be caused by sporadic background noise rather than genuine miRNAs.

Unlike the non-automated approach, not all of the high probability pre-miRNAs from the automated approach were identified in both infections. As all of the miRNAs from these pre-miRNAs were still abundant in both infections, the only reason for them to be identified in one infection and not in another would be a slight change in the alignment signatures. This may be due to the presence or absence of extra isomiRs from one infection to the next. Inspection of the alignment signatures for such candidates (i.e. those not identified in both infections by the automated methods) revealed that while they did change marginally from one infection to the next, a miRNA-like alignment signature was still present in both infections. It is likely that it is these subtle changes between infections that result in some pre-miRNAs not being identified in both infections by the automated methods. Both miRDeep and Mireap identified more high probability pre-miRNA candidates in the H361 infection compared to the N076 infection, suggesting that overall there were better miRNA signals in the H361 infection.

MR5075 was the only high probability pre-miRNA identified in the non-automated approach that was not identified by either automated method. To investigate if this could have been caused by aberrant isomiRs in the alignment signature, such reads were removed from the data followed by reanalysis by MiRDeep. However MR5075 was not identified as a pre-miRNA by MiRDeep regardless of how many aberrant isomiRs were removed. During the analysis process miRDeep creates interim data for use by different algorithms at different stages of the analysis process. One such

interim dataset consists of a list of provisional pre-miRNA candidate sequences excised from the genome based on stacks of reads mapping to them. The sequence for MR5075 (or at least a variation of it) was found within the interim datasets for both infections, indicating that MR5075 was indeed initially identified as a pre-miRNA candidate by miRDeep at one stage of the process, but later eliminated for some reason.

MiRDeep also scored each pre-miRNA identified, allowing candidates to be ranked in terms of how likely they were to be genuine. The miRDeep2 scores associated with each pre-miRNA showed an interesting trend. All of the high probability pre-miRNAs from the non-automated approach that were also identified by MiRDeep were also the top ranked candidates from non-coding regions in both infections. This refers to their ranking before the elimination of candidates not also identified by Mireap, demonstrating further agreement between the non-automated and automated approaches.

MirDeep also estimated a signal-to-noise ratio based on the statistical significance of scores. As it bases the identification of pre-miRNAs on a combination of structural features and small alignment signatures, it works under the assumption that the two are explicitly related, i.e. that the unique structural features of pre-miRNAs act as substrates of Drosha and Dicer resulting a specific type of cleavage leading to a characteristic alignment signature. To test the significance of the relationship between the pre-miRNA structure and alignment signature within results, miRDeep simulates a null-hypothesis whereby there is no relationship between pre-miRNA structure and small RNA alignment signatures. This is done by taking proposed miRNAs and isomiR sequences from one precursor and aligning them to another precursor selected randomly from the data. Alignments to the new precursor are based entirely on position (not sequence) i.e. they are placed in the same positions on the new pre-miRNA (relative to the 5' end). These are termed controls and are analysed by the miRDeep core algorithm in the same way as real alignment signatures and a new score is given. This is repeated 100 times and the average amount of randomly matched pre-miRNA/small RNAs that pass the score cut-off points after each of the 100 permutations is taken as the amount of pre-miRNA

candidates that are expected to avoid elimination by chance alone (Friedländer et al., 2008). These control results were used by MiRDeep as a basis to estimate the signal-to-noise ratio for a range of given score cut-off points (Supplementary File 4.1.4 T.1-T.2).

In the H361 infection the lowest scoring high probability CyHV-3 pre-miRNA (from the non-automated approach) also identified by miRDeep was MD11410 which had a score of 8.7. MiRDeep estimated that for pre-miRNAs with scores >8 (i.e. cut-off point), the signal-to-noise-ratio should be 3.7. Out of eight pre-miRNA candidates identified by miRDeep with scores >8, only five were considered likely represent real miRNA signals (on the basis that they were also identified by Mireap and classed as high probability candidates in the non-automated approach). The remaining three were candidates from protein-coding regions. Taking the latter to be noise, this meant that the real signal-to-noise-ratio among candidates with scores >8 was 1.6 (i.e. 5/3). This suggested that the signal-to-noise ratio was over estimated by a factor of 2.3. Testing the signal-to-noise ratio for the N076 infection in the same manner revealed a similar degree of overestimation. This suggested that true miRNA signals above this cut-off point were statistically lower than expected and that maybe some of the eliminated pre-miRNAs (in this case all from protein-coding regions) that had significant scores, warranted further investigation. For this reason the alignment signatures for these candidates were inspected (Supplementary File 4.1.4 T.1-T.2) and those showing miRNA-like signatures were further analysed by MiPred and the CSHMM-method. However, this analysis revealed that none of these candidates were miRNA-like. This observation, combined with the fact that none of these particular candidates from non-coding regions were identified in both infections or by both automated methods, was considered as conclusive evidence to classify these candidates as background noise despite their high scores.

Most pre-miRNAs have low folding energies compared to non-RNA hairpin structures (Bonnet et al., 2004b), hence assessment of pre-miRNA candidates in this regard is quite important. In a similar way to MiPred (described in Section 1.7.1.2) miRDeep also assessed the statistical significance of the folding energies displayed by pre-miRNA candidates. While all of the five candidates that were also identified

by the non-automated approach were earlier found to have statistically significant folding energies by MiPred, conversely most of them were found not to have significant folding energies by MiRDeep (Supplementary File 4.1.4 T.1-T.2). This was a very significant discrepancy between the two methods that required further investigation. Closer examination of the two methods revealed that while both methods used the exact same algorithm, Randfold (Workman and Krogh, 1999), to assess the significance of folding energy, the approach used in each case was slightly different. Both methods randomly shuffle the pre-miRNA sequence 1000 times, to assess the distribution of folding energies. The difference however is that miRDeep uses simple mononucleotide shuffling (Friedländer et al., 2008), whereas MiPred uses dinucleotide shuffling (Jiang et al., 2007) for this analysis. Workman and Krogh (1999) found that dinucleotide shuffling is more relevant in this context as the stability of the RNA secondary structures depends dinucleotide stacking energies. Thus it was concluded that MiPred provided more valid conclusions regarding the statistical significance of pre-miRNA stability and hence the equivalent results from MiRDeep were ignored.

### 8.2.3 Summary of the identification of pre-miRNA candidates from deep sequencing data

The results from automated and non-automated approaches to the analysis of sequencing data were largely in agreement with each other regarding the most likely CyHV-3 pre-miRNAs. This indicated that the elimination of candidates through (i) in depth analysis of pre-miRNA candidates in the non-automated approach and (ii) elimination of pre-miRNAs not identified by both methods used taking the automated approach, were the best ways to filter results from each respective approach used. Overall combining the two approaches resulted in the identification of seven high probability CyHV-3 pre-miRNAs.

In both cases pre-miRNA candidates occurring in protein-coding regions were not considered. Generally miRNAs do not occur within such regions, thus ignoring unique reads that mapped to protein-coding regions ensured that the identification of to pre-miRNAs using the non-automated approach was confined to genomic regions most likely to encode miRNAs. Also, because the majority of unique reads mapped

to protein-coding regions in both infections, it was rather impractical to identify all possible pre-miRNA candidates in these regions using non-automated approaches, even if most unique reads could be ignored for having read counts <10. However this could be more easily done using the automated methods, although, as expected this revealed no high probability candidates from protein-coding-regions, indicating that it was correct to ignore them in non-automated analysis. This does not rule out the possibility that CyHV-3 does encode pre-miRNAs within coding regions, and it is known that other viruses such as HCMV (Buck et al., 2007) and KSHV (Cai et al., 2005) encode pre-miRNAs within coding regions. In such cases the same molecule can act as both an mRNA and a pri-miRNA resulting in two different possible fates for the same molecule. In some cases the miRNA may be inefficiently processed from the mRNA allowing most of the mRNA to be transported to the cytoplasm before Drosha cleavage (Cai et al., 2005). However identifying miRNA signals from protein-coding regions in deep sequencing data from is much more difficult and especially so if the mRNA in question is highly expressed. Numerous random mRNA degradation products interfere with the interpretation of important signals in data beyond the miRNAs themselves, such as the alignment signature. This may be especially relevant during lytic infections, when viral transcription and RNA turnover levels increase. As exemplified by this study, for the same reasons it is difficult to assess the alignment signatures of pre-miRNA candidates occurring in 3' UTRs. Despite this, it is still unlikely that many other candidates were genuine pre-miRNAs due to a general lack of conformance with other characteristics of pre-miRNAs. Notably out of 15 of these candidates occurring in 3' UTRs, many of them did have MFE values <-25 kcal/mol and for those that did, none (with the exception of MR6201) were considered statistically significant by MiPred.

It is likely that increased levels of RNA turnover during the lytic infection significantly contributed to the amount of non-miRNA background noise in the sequencing data. Other studies have not commented in any detail on the measures taken to identify viral miRNAs from background noise in such samples, and if CyHV-3 miRNAs were identified based on the same criteria used in other studies it would have resulted in many more pre-miRNA candidates, with many of them not actually being genuine miRNAs. Some studies have even proposed the existence of pre-miRNAs in the absence of minor forms and isomiRs or even if the major forms

have extremely low read counts, are not detected in multiple sequencing experiments and subsequently not detected by PCR (Glazov et al., 2010). However it is possible that there may be low levels of background noise observed in these studies, and that the majority of the few viral transcripts detected might map to a handful of predicted pre-miRNAs. This may especially be the case where viral miRNAs are identified in latently infected cell lines where most genes expressed during lytic infections are inactivated with the exception of miRNAs. This may also explain why many other studies on viral miRNA identification to not involve deep sequencing from multiple infections, however it is something that is not explicitly clear from the literature. By contrast, the stringent criteria used for identifying pre-miRNAs in this study made for a considerably more conservative approach towards pre-miRNA identification, but it was nonetheless necessary to ensure that only genuine pre-miRNAs were identified.

Although the results from the automated and non-automated approaches were quite similar, both methods had their advantages and disadvantages. The automated approach allowed the identification of one additional CyHV-3 pre-miRNA, namely MD11704. This pre-miRNA was not considered in the non-automated approach due to the fact that its miRNA read counts were too low. However the non-automated approach did allow the identification of MR5075, which was not identified using the automated approach. The in-depth analysis carried out on the miRNA candidates in the non-automated approach allowed the elimination of MR6201 despite the fact that is was detected by both automated methods, demonstrating that overall the non-automated approach was more effective at distinguishing between low and high probability pre-miRNA candidates. Overall the results from both approaches complemented each other quite well. The use of both in unison was ultimately better than the use of one approach alone, resulting in the identification of seven high probability CyHV-3 pre-miRNAs rather than six (i.e. had either method been used alone). Although more labour intensive, the non-automated approach had much more accuracy and was probably best for this reason. Importantly this approach facilitated the identification of moRNAs, something that was not possible with the automated approach.

The discovery of CyHV-3 encoded moRNAs was not anticipated but it is a very interesting aside. Although their biogenesis is intricately connected with that of miRNAs and sometimes they can be even more abundant than them, it was only recently with the advent of deep sequencing, that the presence of moRNAs was fully appreciated. However they are still far from understood. Somewhat surprisingly, their abundance is not linked to whether they occur adjacent to the major form or minor form miRNA on the pre-miRNA. In fact moRNAs processed from the from 5' arm of the pre-miRNAs are generally more dominant in samples (Taft et al., 2010; Umbach et al., 2010). This suggests that while their biogenesis may indeed by linked to miRNAs, the rates of processing are not necessarily interdependent and may be independently regulated. The CyHV-3 miRNAs identified in this study showed no clear dominance of moRNAs from one arm over the other (two with 5' and two with dominant 3' arm moRNAs)

The levels of moRNAs present in samples do not generally correlate with the levels of mature miRNAs processed from the same pre-miRNAs (Berezikov et al., 2011). In some cases moRNAs have been shown to be present at much lower levels than miRNAs from the same pre-miRNAs, as with KSHV (Umbach and Cullen, 2010), RRV (Umbach et al., 2010) and in other organisms (Berezikov et al., 2011; Langenberger et al., 2009). In other cases some have been found to be more prevalent than miRNAs from the same pre-miRNA, as demonstrated in HSV-2 (Jurak et al., 2010). This dominance of either moRNAs or miRNAs from the same pre-miRNA can also fluctuate depending on the circumstances, for example, such fluctuations were observed at different developmental stages (Shi et al., 2009). It is possible that the dominant classes of small RNAs from viral pre-miRNAs could fluctuate in the same manner under different experimental conditions. Experimental conditions that result in the absence of detectable miRNAs and the dominance of moRNAs may explain why some viral moRNAs were originally classified as miRNAs. In the case of the CyHV-3 moRNAs that were discovered in this study, all of them were significantly less abundant than miRNAs from the same pre-miRNAs although some were more abundant than their adjacent miRNAs for example the 5' moRNA from MR5075 and the 3' moRNA from MD1111 (but only in one infection).

The functions of moRNAs are still unknown. Unlike miRNAs it appears that they localise in the nucleus rather than in the cytoplasm (Taft et al., 2010). Despite this, one study did demonstrate that moRNAs may be incorporated into the RISC (Umbach et al 2010). In that study, the expression of the RRV moRNA moR-rR1-3-5p correlated with a 3-fold down regulation of a reporter gene containing a fully complementary artificial target site. Its effect on the reporter gene was moderate compared to the miRNA processed from the same pre-miRNA in the same cells. However, it did demonstrate that moRNAs may be incorporated into the RISC but possibly at a lower level. While these findings are interesting, observations made elsewhere indicating that moRNAs accumulate in the nucleus (Taft et al., 2010) suggest that they may not always be incorporated into the RISC. Ultimately, the role of this new class of small RNAs is currently unclear and requires more investigation.

Much of the in depth assessment of the miRNA signals in the non-automated approach was based on quantitative information. It was an important part of the assessment of small RNAs (putative miRNAs and isomiRs) to determine if they collectively displayed fundamental miRNA-like attributes. For example, quantitative information was central to the identification of the major and minor forms derived from each pre-miRNA, and the dominant isomiRs in each infection. This of course had knock-on effects on the nature of the predicted mature miRNA duplexes and assessment of 5' end processing consistency (using Ratio-3). More importantly, it was the quantitative data that was primarily used to identify CyHV-3 miRNAs as transcripts that were distinct from background noise i.e. they were required to have had a read counts ≥10. It is clear reliable quantitative information is vital in order to carry out valid analyses of this nature.

It is important to acknowledge that although deep sequencing is a useful tool it does suffer from enzymatic bias, which may cause some sequences to be over-represented or under-represented in the data. This occurs as sample preparation involves a series of enzymatic steps (i.e. 3' adapter ligation, 5' adapter ligation, RT and PCR) and biases introduced at each step can have a cumulative effect. It has been shown that the ligation steps in particular can be the source of considerable bias which varies depending on the combination of enzymes and adapter sequences used (Hafner et al.,

2011; Sorefan et al., 2012). Sample preparation can even suffer from gel extraction bias (Quail et al., 2008). In addition it has been shown that results will vary significantly due to protocols used in library preparation (Linsen et al., 2009; Toedling et al., 2012). In fact, as a result of widespread use of the Illumina sample preparation protocol, the majority of sequences present in miRBase are strongly biased towards sequences that are preferred by the Illumina adapters (Sorefan et al., 2012).

There is ample evidence of this happening following deep sequencing of miRNAs and this most often becomes apparent when miRNA expression is also analysed using non-enzymatic methods or methods that offer more quantitative precision. For example transcripts with low read counts can sometimes be easily detected by northern blotting while others with much higher read counts show faint signals (Reese et al., 2010) and the same observations have also been made when using PCR to verify miRNA expression levels (Meshesha et al., 2012).

The potential for sequencing bias is something that needs to be considered in studies such as the one described here. As highlighted earlier quantitative information is central to miRNA identification from deep sequencing data, in fact automated methods for miRNA identification such as miRDeep attach a lot of weight to significantly high read counts when making scoring predictions (Friedländer et al., 2008). Under controlled experimental conditions it has been shown that bias alone will cause many transcripts to display dramatically higher than expected read counts. For example Sorefan et al (2012) found that >50,000 different transcripts that were expected to appear only once in a specific dataset actually appeared more than 10 times. In theory it is possible some small RNAs detected in this study (especially at levels close to the cut-off of 10) could have been over-represented. Therefore it was important to take the quantitative information obtained from deep sequencing as a rough guide only until the levels of these small RNAs were verified using appropriate additional methods. For these reasons, the most likely CyHV-3 miRNAs identified here were investigated further in order to verify their presence in infected cells and to estimated their relative expression levels.

## 8.3 DNA Microarray Analysis

MiRNAs should be the most abundant transcripts in the 17-25 nt RNA sample used for deep sequencing and thus these pre-miRNA candidates identified in the non-automated approach were primarily identified on the basis of highly abundant small RNAs mapping to them. As these specific sequences could potentially be over-represented in the data due to enzymatic bias and because other sequences could be under-represented for the same reasons, analysing the same sample using array hybridization provided an ideal way to investigate this possibility. Probes were designed to detect putative miRNAs from all high and low probability pre-miRNA candidates. In theory it was possible that miRNAs from other VMir predicted pre-miRNAs were not present or ignored due to under representation of small RNAs mapping to them. For this reason, additional probes designed to detect theoretical miRNAs from the highest scoring pre-miRNAs predicted by VMir were included on the array.

If the read counts observed in deep sequencing were not the result of over representation, regardless of whether these small RNAs are from high or low probability pre-miRNA candidates, probes targeting those identified as the most abundant transcripts should nonetheless show the highest signals in array hybridization using the same sample. This was found to be the case for most of the 15 low probability and 2 high probability pre-miRNAs in the data, indicating that these were actually among the most abundant transcripts in the sample, as previously indicated by the deep sequencing data.

In addition, the majority (12/15) of low probability pre-miRNAs identified following non-automated analysis gave signals using the S2 as hybridization probe (26-35 nt RNA from the H361 infection). This supported the conclusions from seep sequencing that these were likely to be degradation products from larger non-miRNA transcripts. More specifically, these were all likely to be degraded fragments from 3' UTRs to which they mapped. Therefore as expected, other fragments of the same 3' UTRs containing the same target sequences were detectable in other RNA size ranges with the same probes.

In agreement with the deep sequencing data, the highest signals seen among the high probability candidates were from MR5057 and MD11776. In both cases the signals from the major and minor strands were also in agreement with the deep sequencing data. The read counts for the 3' miRNA from MD1111 were at similar levels to the 3' miRNA from MD11776 (in fact higher), which was not detected on the array. This suggests that (i) this was over-represented in the deep sequencing data and (ii) it was the 5' arm that was the major form expressed in the H361 infection. The results for MR5075 also suggested that contrary to the deep sequencing data, it was the 5' arm miRNA that was the major form of this pre-miRNA.

The two remaining, high probability candidates identified by the non-automated approach had much lower read counts in the sequencing data, suggesting their levels were beyond the limits of detection in array hybridization.

A cut-off point had to be established in order to distinguish significant from non-significant results. RVs above background were not immediately taken as positive either. A lower limit cut-off for RVs was first established for each hybridization. Therefore for a signal to be considered positive in a given hybridization its RV must also have been above the cut-off value for that hybridization. The cut-off values used in this study were based on the level of variability in the signals for the control probes used to estimate background, which was established by calculating their average absolute deviation from this background level. The cut-off values were then defined as one average absolute deviation above these background levels. Hence, pre-miRNAs with RVs that were greater than these cut-off values (even after correction) were considered positive.

Using cut-off values based on the intrinsic variability within the dataset is similar to the approach used by Grundhoff *et al* (2006), except in that case the cut-off values were 1.5-1.75 standard deviations above the background. In this study to detect CyHV-3 miRNAs, the cut-offs were based on the average absolute deviation instead of the standard deviation. This was because this is a more robust statistic that is less sensitive to being distorted by major outliners. This is because when calculating the standard deviation, the differences are squared, so outliners are weighted more heavily. This is not a problem when using the absolute deviation and therefore it is

better suited to the analysis of data such as this where outliners (mainly caused by artefacts on the array surface) are likely to be encountered, thus avoiding over-estimation of the true level of variability. The cut-off values were defined as one average absolute deviation above background so as to avoid imposing higher arbitrary cut-off values. (i.e. 1.5 absolute deviations above background or higher). At the very least, RVs that were greater than one absolute standard deviation above this cut-off value (even after correction) were seen to give signals that were higher than the intrinsic level of variation present in the data and therefore significantly different from background levels. However, further investigation of the discrepancies between the array and deep sequencing suggested that this cut-off may have been too low.

As highlighted earlier it was only the highest level miRNAs from deep-sequencing that were (i) also detected at the highest level on the array and (ii) in agreement with the deep sequencing data regarding major and minor forms. The low level signals from MD1111 and MR5075 that were not in agreement with the deep sequencing results were investigated further by re-analyzing the same sample by RT-qPCR, targeting both miRNAs (5' and 3' arm miRNAs) from these two pre-miRNAs. The results from RT-qPCR were in agreement with the sequencing data showing that the miRNAs from the 3' arms of MD1111 and MR5075 were the major forms derived from these pre-miRNAs in this infection. Thus, the observation of low-level signals from the 5' arms of both of these pre-miRNAs in the absence of positive signals from the 3' arms was not a true reflection of the levels of target present. However, somewhat in agreement with array hybridization the RT-qPCR results also indicated that the miRNA from the 3' arm of MD1111 was over-represented in the sequencing data to begin with and was actually present at a much lower level than MD11776, confirming why there was no signal from the probe targeting this miRNA. These RT-qPCR results are discussed in more detail later.

Although the signals from the probes targeting the 5' arm miRNAs from MD1111 and MR5075 were both above the cut-off point, the fact that they were both also relatively low suggests that these signals may just have been due to background noise. The establishment of a cut-off value was supposed to eliminate such background noise, however in hindsight this may have been set too low. This may also explain why there were also so many apparent signals for novel provisional

miRNA candidates for which there was no corresponding evidence in the deep sequencing data.

In conclusion, DNA microarray hybridization using miRNA-containing and non-miRNA-containing size fractions proved to be a very useful high throughput method of distinguishing between miRNA and non-miRNA small RNAs. However the experimental conditions used here only resulted in the detection of highly expressed miRNAs such as those from MR5057, MD11776 and all highly expressed host miRNAs. The methodology used was simply not sensitive enough for low-level miRNA detection. The use of molecular crowding agents such as dextran sulphate in hybridization buffers dramatically increases the sensitivity of array hybridizations (Ku et al., 2004). Indeed, the inclusion of dextran sulphate in the hybridization buffer (final conc. 6.5% w/v) in this experiment did allow the detection both major and minor forms derived from two CyHV-3 miRNAs, which is in stark contrast to similar earlier hybridizations carried out in the absence of dextran sulphate (data not shown). It is possible that increasing the levels of dextran sulphate and/or input RNA probe might improve the sensitivity in such an experiment. It may also be useful to combine this approach with a higher more stringent cut-off point in order to adequately eliminate background noise. Overall it can be concluded that the highest level miRNAs identified in deep sequencing were not over-represented due to an experimental bias and that most high level non-miRNA small RNAs were indeed degraded RNA.

## 8.4 Northern blotting

The detection of discrete miRNA signals using northern blotting is probably the ideal way to prove the existence of proposed miRNAs. This is because it can be used to simultaneously detect and display all of the transcripts that contain the target sequence. However northern blotting is not as sensitive as deep sequencing or array hybridization. The mature miRNA from MR5057 was shown to be the most highly expressed and most readily detectable CyHV-3 miRNA in the deep sequencing and array hybridization experiments and unsurprisingly it was the only one that was detectable by northern blotting. The specific signals obtained for MR5057 are characteristic of miRNAs and corresponded perfectly with results from VMir, deep

sequencing (pre-miRNA and miRNA length) and array hybridization (RNA size range distribution). These observations are enough to confirm that MR5057 is a genuine CyHV-3 encoded pre-miRNA.

The high read counts and array signals, suggest that this miRNA it is efficiently processed by Dicer. The pre-miRNA nonetheless does accumulate in sufficiently high levels to remain detectable by northern blotting. This may be due to (i) high levels of pri-miRNA expression, (ii) highly efficient Drosha processing of the pre-miRNA or (iii) both. Alternatively, Dicer processing of this pre-miRNA may not be efficient at all and the high levels of processed mature miRNAs present may simply be a consequence of high cellular levels of the MR5057 pre-miRNA. Regardless of how efficient the MR5057 pre-miRNA processing is, the results suggested that its processing must be more efficient than that of the pre-miRNA for MR5075. Given their close proximity to each other on the genome (257 bp separating the two 5' start positions), it is likely that they share the same pri-miRNA, yet a band for the MR5075 pre-miRNA could not be detected. The difference in the levels of miRNAs derived from these two pre-miRNAs (observed in all other experiments in this study) and the relative levels of their pre-miRNAs as observed by northern blotting, implies that they must be processed at very different rates however. It is not uncommon for this to occur, for example in EBV, despite the fact that that all miRNAs in the BART cluster are likely to be transcribed as part of a single pri-miRNA, the levels of these miRNAs within cells can differ by as much as 50-fold (Pratt et al., 2009).

Although present in the sample (as later tested by real time RT-qPCR, data not shown) the levels of miRNAs from pre-miRNA candidates MD11776, MD1111 and MR5075 were too low to be detected by northern blotting, and ideally should be characterised by more sensitive methods. This demonstrates that although useful, northern blotting is not always sensitive enough to detect low-level transcripts. Some modifications may be enough to enable detection here however. In these experiments, 10 μg of total RNA was loaded per lane and increasing that amount to 20-100 μg may have led to the detection of miRNAs and possibly pre-miRNAs from some of these transcripts. In particular it would be interesting to repeat this with MD11776, firstly as it's the next most abundant candidate and therefore the most likely of the remaining targets to be detectable, and secondly because it may be

possible to detect two isoforms of the pre-miRNA i.e. an additional larger version containing un-processed moRNAs as has been previously shown for a PRV pre-miRNA (Wu et al., 2012).

## 8.5 Diagnostic assay development and use

Due to the possibility of enzymatic bias in deep sequencing the expression levels of CyHV-3 miRNAs needed to be verified by other methods. It was evident from previous experiments that many of the miRNAs derived from CyHV-3 pre-miRNAs were present at very low levels and that verification of the expression levels for these miRNAs required a different more sensitive approach. A sensitive method such as RT-qPCR was ideal for this. Due to the small size of miRNAs and the need to avoid non-specific signals from pre-miRNAs or pri-miRNAs (which contain the same target sequence) or even from genomic DNA, the approach used for detection of miRNAs by RT-qPCR is slightly different. This approach is known as Stem-loop RT-PCR (Chen et al., 2005) (Figure 2.9)

However, due to the proprietary methods used in primer-design, the primer sequences were not disclosed by the manufacturer. In addition the assays themselves were very expensive and for this reason they were only purchased in the smallest sizes. However reducing the amount of primer/probe used per reaction assay was an ideal way to ensure that there was enough to meet the needs of this present study. It was important to first assess the effects of this deviation from the manufacturer's recommendations. As expected using 0.1x primer/probe reduced the sentitivity, but it was still more than adequate to detect host and viral miRNAs *in vitro.* The use of 1x CyHV-3 miRNA primer/probe was confined to *in vivo* samples where it was more likely that more sensitivity was needed as a consequence of lower proportions of infected/non-infected cells. For some CyHV-3 assays, it was found that the use of 1x primer/probe resulted in low level non-specific signals in RNA from negative CCB cells (this did not occur using 0.1x primer/probe). Subsequent testing involving different annealing temperatures revealed that this may have been due to the primer/probe sequence i.e. possibly due to higher GC content. This was something to bear in mind when using some of these assays to test for CyHV-3 miRNAs *in vivo* as

low level non-specific signals would be indistinguishable from low level signals caused by the presence of low level CyHV-3 miRNAs.

## 8.6 *In vitro* sample testing

In order to assess whether the relative levels of CyHV-3 miRNAs observed in deep sequencing were caused by enzymatic bias, the same RNA sample was retested using RT-qPCR. This retest showed that the relative levels observed in deep sequencing were more or less accurate except for the miRNAs from MD1111 which may have been over represented due to enzymatic bias. This also revealed that the major/minor form designation from deep sequencing was correct. Interestingly, sequencing from the N076 experiment showed that the level for the miRNAs from MD1111 was greatly reduced, compared to most other miRNAs, although it also showed that the major form had switched. If these sequencing results from the N076 infection are taken to be reasonably accurate (as exemplified by the RT-qPCR results in the H361 infection) and assuming no additional bias was introduced in the second experiment (i.e. on the basis of the same sequence content in both infections with respect to adaptors and small RNAs present and the same methodology), it is plausible that there was a shift in major/minor form in this infection. It would have been interesting to also test RNA from the same N076 infection by RT-pPCR, although this was not possible as there was no more of that particular sample remaining.

The measurement of CyHV-3 miRNA expression over the course of an *in vitro* lytic infection (using RNA from a new infection with the N076 isolate) showed that the rise in their expression levels directly correlated with the rise in viral DNA levels over the same time period, thus definitively linking these positive signals for CyHV-3 miRNAs to viral replication. The increase in the expression levels over the course of the infection is presumably linked to an overall increase in the levels of viral transcripts as more cells become infected. This experiment also showed that the relative expression levels observed in the previous experiments (i.e. deep sequencing using the H361 isolate) were indeed typical of this kind of CyHV-3 *in vitro* infection regardless of what isolate was used. MR5057 and MD11776 are consistently expressed at the highest levels as determined by RT-qPCR. For this reason it makes

sense that these were the only ones detected on the array, and why MR5057 was the only one detected by northern blotting. In agreement with the deep sequencing and RT-qPCR data for the H361 infection, the 3' mRNAs from MR5075 and MD1111 were consistently found to be the major forms derived from these pre-miRNAs at all time points over the course of the *in vitro* infection using the N076 isolate. These were the only pre-miRNAs to have both major and minor forms targeted by RT-qPCR. The fact that the miRNAs from the 3' arms were always present in higher levels compared to those from the 5' arms suggests that this type of expression pattern must arise from a precise, repeated process rather than random RNA degradation. Thus, these observations support the annotation of MD1111 and MR5075 as genuine pre-miRNAs.

## 8.7 *In vivo* sample testing

While *in vitro* infections are very useful for studying viruses, naturally they are not fully representative of the situation *in vivo*, where many additional factors may come into play. Most notably, elements of the immune system are absent. This may be significant, as sometimes the nature of viral interaction with the immune response can have profound effects on the behaviour of a virus *in vivo* as described in Section 1.2.2. For this reason it was important to investigate whether or not the same miRNAs were expressed during a lytic *in vivo* infection and if the relative expression levels were similar to what was observed *in vitro*. To do this it was necessary to use RNA from lytically infected tissue. Due to possible non-specific signals from five out of the eight CyHV-3 miRNA assays ultimately only three CyHV-3 miRNAs were targeted in Phase 3, namely the assays targeting the major forms derived from MR5057, MD11776 and MR5075. Out of the three samples from this fish only the gill gave positive results. All three targets were detected. As somewhat expected the levels of CyHV-3 miRNA were low compared to the levels *in vitro*. Interestingly, although low by comparison, the relative levels of these 3 miRNAs did mirror the relative levels of the same miRNAs *in vitro*. This confirmed that at least three of the CyHV-3 miRNAs identified the lytic infections *in vitro* are also expressed *in vivo*. It is unclear why the same miRNAs did not show positive signals in the brain or kidney. This is especially surprising in the case of the kidney as CyHV-3 replicates

most efficiently in this organ (Pikarsky et al., 2004). One reason is that the levels were beyond the limit of detection. While the levels of the positive signals in the gill were quite low, it does not necessarily mean that the expression levels were low and could in fact be a consequence of low level virus in this tissue (or at least in the section taken for RNA sampling). On this note it would be interesting to have extracted both RNA and DNA from the same homogenate, thus allowing this to be investigated. However, only extracted RNA samples were available for this study.

The lytic infection in Fish-8 was actually a reactivated lytic infection. The fish was a long-term survivor of a previous high mortality outbreak in a lake in the UK. As expected introduction to a permissive water temperature resulted in the reactivation of a lytic infection, demonstrating that this fish was indeed a latent carrier. This suggested that other fish from the same lake that were also survivors of the same outbreak may also have been latent carriers. As miRNAs in other HVs are usually also expressed during latency, it is possible that CyHV-3 encoded miRNAs are also expressed at this stage. For reasons described earlier (Section 1.4.10.6) these may be more detectable than low levels of viral DNA during latency and therefore more ideal as diagnostic targets for the diagnosis of latent infections. In order to explore this possibility, six additional fish that were long-term survivors of the same outbreak as Fish-8 were also tested for CyHV-3 miRNAs. RNA from gill, gut, brain and kidney samples were tested from these fish. The only samples that were selected for Phase-3 testing were the brain and gut samples from Fish-21. The gut samples showed positive signals for MR5057 in both sample dilutions. Although the positive signals were quite consistent they were very low. If this fish was latently infected the result suggested that the CyHV-3 miRNAs targeted were not readily detectable and that there could be other CyHV-3 miRNAs that were expressed at higher levels during latency. However this was just a quick exploratory experiment using a small sample size. Crucially the fish that were used were not definitively shown to be latently infected. It was possible that the positive signal represents low level reactivation. The only way to definitively confirm that specific fish are latent carriers is to reactivate a lytic infection. Unfortunately, once reactivation occurs the same fish are no longer useful subjects for the study of latency.

There are other ways to approach such an experiment however. Recent findings indicate that testing leucocytes may in fact be the most reliable and sensitive way to detect latent CyHV-3 (Eide et al., 2011a, 2011b; Xu et al., 2013). This also presents a way to carry out non-lethal sampling from potential latent carriers before confirming latency through reactivation. Blood sampled from fish confirmed to have been latent carriers (at the time of sampling) would be ideal test material for use in an investigation into CyHV-3 miRNA expression during latency. In addition DNA could also be extracted from the same samples allowing a direct comparison between the levels of viral DNA and RNA present, thus offering an ideal way to evaluate the potential benefits of targeting latency associated miRNAs instead of genomic DNA.

## 8.8 Investigation into CyHV-3 miRNA sequence conservation

### 8.8.1 Identification of potential CyHV-3 pre-miRNA and/or miRNA homologues in other CyHV genomes

Viral miRNA conservation and/or high sequence similarity is quite rare and to date, this has only been observed between closely related viruses. For this reason the CyHV-1 and CyHV-2 genomes were searched for sequences that were homologous to CyHV-3 pre-miRNAs. BLAST searches revealed significant matches to three CyHV-3 pre-miRNA sequences in these two other viruses. The pre-miRNAs in question were MD1111, MD11776 and MD11704 (only considered a genuine CyHV-3 pre-miRNA at this point in the study). Interestingly, these were all located in orthologous non-coding regions of these three viral genomes.

As highlighted in Section 1.6.2, it was possible that if there was some pre-miRNA sequence conservation between these viruses, it would be mainly confined to the regions of the pre-miRNA encoding the mature miRNAs. Indeed in all three such cases found for CyHV-3 pre-miRNAs the highest sequence similarity was in the regions encoding the mature miRNAs. In fact for MD1111, the only sequence similarity was to the region encoding the 5' miRNA. In addition this match was only in the CyHV-2 genome. Unlike all other CyHV-3 miRNAs, there are no non-coding regions in the CyHV-1 genome that are orthologous to the non-coding region

encoding MD1111 in the CyHV-3 genome. This is not unusual as there are several other genes present in the CyHV-3 and CyHV-2 genomes that are also absent in the CyHV-1 genome (Davison et al., 2012).

Matches for the other CyHV-3 pre-miRNAs were more widespread across the pre-miRNA sequences although the regions showing the highest similarity were those encoding the mature miRNAs, namely the 3' arm miRNA for MD11776 and the 5' arm miRNA for MD11704. The regions encoding the MD11776 moRNAs did not show the same levels of sequence similarity indicating that it is possible that the miRNA sequences are under more selective pressure. If these are genuine miRNA orthologues in these three viruses, then it would make sense were it to be the seed regions that were most conserved and indeed this is the case with the 3' arm miRNA for MD11776 and the 5' arm miRNA for MD11704 with their respective seed regions being perfectly conserved. The least conserved seed region is that of the 5' arm miRNA for MD1111 5' which showed two mismatches to the CyHV-2 genome (all other cases of sequence differences in seed regions show a maximum of one mismatch with each virus). Such differences in the seed regions may suggest that while there may be an evolutionary relationship between these miRNA sequences, the may nonetheless have divergent functions or that their targets may also have diverged.

Notably all of these CyHV-3 pre-miRNAs and potential homologues are found in non-coding regions opposite the same orthologous protein-coding genes in all viruses. While it is interesting that these pre-miRNA sequences may be conserved due to some functional importance it is also possible that it is in fact the protein-coding gene on the opposite DNA strand that is under selective pressure. However, the remaining four CyHV-3 pre-miRNAs also occur in non-coding regions opposite orthologous protein-coding genes but do not show any signs of conservation. This suggests that it is possible that the sequence similarities observed for MD1111, MD11776 and MD11704 may not have simply been a consequence of their genomic location.

The only potential CyHV-3 pre-miRNA homologue in CyHV-1 and CyHV-2 that was also independently predicted by VMir was the MD11776 homologue in CyHV-

2. This had both a high VMir score and a high WC value in addition to being the only potential CyHV-3 pre-miRNA homologue to be classified as a pre-miRNA by MiPred and the CSHMM-Method. None of the other potential CyHV-3 pre-miRNA homologue sequences were deemed to fold into pre-miRNA-like structures, possibly due to differences in sequence.

Although the potential MD11776 homologue in CyHV-2 does form a pre-miRNA stem-loop structure, this does not automatically mean that the miRNA is functionally active. The pre-miRNA secondary structure is an extremely important factor in determining the efficiency with which a pre-miRNAs are processed by Dicer and Drosha, or if they are even processed at all. Even small mutations in pre-miRNA sequences in different strains of the same viruses can render pre-miRNAs functionally inactive. Several mutations in the KSHV pre-miRNA kshv-mir-K9 of the strain persistently infecting BC-3 cells has rendered this gene inactive compared to the same gene in the KSHV strain infecting BS-1 cells (Kuchenbauer et al., 2008). Strikingly a single point mutation in the EBV pre-miRNA BHRF1-3 in the Raji strain has disrupted pre-miRNA structure enough to inhibit its maturation (Pratt et al., 2009).

### 8.8.2 CyHV-3 seed region matches to viral miRNAs

The seed regions of miRNAs are the primary determinants of function, thus miRNAs from different viruses that share the same seed sequence may also have conserved target sites on orthologous genes. CyHV-3 miRNA seed regions were compared to seed regions from other known viral miRNAs as a way of identifying other miRNAs that carry out the same functions during infections. The most common matches were to miRNAs from EBV, rLCV and MDV-1. These are also viruses with some of the greatest complements of known miRNAs (68 in rLCV, 44 in EBV and 26 in MDV-1), so statistically there is a greater likelihood that matches to miRNAs from these viruses are due to chance, although it may not necessarily be the case. Interestingly two miRNAs (full match to MD11776 5'and partial to MD1111 3') had matches to seed regions from two BLV miRNAs. BLV has a low number of miRNAs (8 known) and it does seem that a match to two of these is less likely to occur by chance. Moreover it is not a member of the order *Herpesvirales*; it is a member of the

*Retroviridae* family. This does seem quite unusual, although CyHV-3 has already been shown to have a very unusual genome, in that it contains several genes e.g. (TmpK and B22R-like genes) not found in any other herpesviruses but instead genes that are strongly associated with phylogenetically distant viruses, in particular members of *Poxviridae, Iridoviridae* families and it also has proteins that show significant similarity to proteins from African swine fever virus (Ilouze et al., 2006). The phylogenic process behind the formation of this unusual genome is not clear, however it does present a plausible theory that CyHV-3 may also have also acquired miRNA genes from other distantly related viruses outside the order *Herpesvirales*. On the other hand, the presence of the same seed sequences does not necessarily mean that there is any evolutionary relationship between any of these miRNAs. Instead it may occur as a result of convergent evolution.

Regardless of the reasons behind the observed similarities in seed regions, it is possible miRNAs with the same seeds have orthologous functions. Therefore, as roles for these other viral miRNAs are elucidated this may provide insights into the possible roles of these CyHV-3 miRNAs that possess the same seed regions. Interestingly the BLV miRNA blv-miR-B4-3p (with identical seed region to MD11776 5' miRNA) down-regulates the tumor suppressor gene *HBP1*. This function was originally proposed due to the seed region sequence similarity with the host oncogenic miRNA miR-29a (Kincaid et al., 2012), demonstrating that viral miRNAs that share seed regions with host miRNAs may also have similar functions, allowing the virus to tap into host miRNA regulatory networks. Similarly miRNAs from other oncogenic viruses such as MDV-1 and KSHV encode orthologues of miR-155. (Boss et al., 2011; Zhao et al., 2011). As the MD11776 5' arm miRNA has an identical seed region with that of the *Cyprinus carpio* miR-29a and it too may have the same function as its host cell's orthologue. However as CyHV-3 is not an oncogenic virus, it is not clear how down regulating an orthologous tumor suppressor gene would be an advantage to CyHV-3, although it is possible that *Cyprinus carpio* miR-29a (and hence the MD11776 5' arm miRNA) carries out a different function (the same may be said for the host protein). Furthermore, under the experimental conditions used in this study, MD11776 5' arm miRNA was shown to be the minor form derived from this pre-miRNA and would not expected to be present in physiologically relevant levels during a lytic infection.

## 8.9 CyHV-3 miRNA target prediction

The high levels of some of the CyHV-3 miRNAs may indicate that they have significant biological roles during *in vitro* infections. The final part of this study was concerned with the prediction of targets for some of these miRNAs. It is mostly the major form that is thought to be the biologically active miRNA form of the duplex. However, this is not always true. The minor form is still always incorporated into the RISC at lower levels and sometimes it is present at physiologically relevant levels depending on the miRNA genes in question (Okamura et al., 2008). Alternatively both strands may accumulate to similar levels (Kim et al., 2009) and the major/minor form can switch depending on the circumstances (Jagadeeswaran et al., 2010; Reese et al., 2010; Ro et al., 2007). For these reasons it was decided to investigate possible viral mRNA targets for both the major and minor forms derived from CyHV-3 pre-miRNAs. Target site prediction was carried out for all CyHV-3 miRNAs (Supplementary File 7.1), but it was only the predictions for miRNAs from MR5057 and MD11776 that were evaluated in more detail (Section 7). This was because MR5057 and MD11776 had the most evidence to support their annotation as genuine pre-miRNAs.

MiRNAs generally target the 3' UTRs of mRNAs. In rare cases target sites can be found within ORFs (and even 5' UTRs). Both of these pre-miRNAs occur opposite protein-coding genes, thus these are obvious potential targets for these miRNAs and they also were also deemed to be high probability targets by PITA and TargetScan (results not shown). Target sites that map to within ORFs are generally not as functionally active as target sites in 3' UTRs however and this is believed to be mainly due to interference from translation (Grimson et al., 2007; Gu et al., 2009; Lin and Ganem, 2011).

Predicting target sites in 3' UTRs based on complementary matches the miRNA seed regions not so straightforward. There are other subtle criteria that need to be taken into account. This is important as miRNA target prediction by its nature gives lots of false positives, but looking for these additional criteria provides an ideal way to distinguish high-quality miRNA target predictions from background noise (Section 1.8). Two very different approaches, TargetScan and PITA were used for miRNA

prediction (Section 7). While TargetScan takes into account many characteristics of target sites, it does not address target site accessibility. Conversely PITA bases its ranking entirely on target site accessibility and acts as a way to also check TargetScan predictions for target site accessibility, thus further improving the quality of the results.

TargetScan and PITA were in complete agreement regarding the most likely target sites for the miRNAs from MR5057. For the 5' miRNA from MR5057, the 3' UTR from ORF117 was predicted to be the most likely target site by both methods. The site itself occurs near the start of the 3' UTR for ORF117 (Figure 7.3). This 3' UTR also contains ORF116 (hence the 3' UTR referred to as ORF_116-117 in the analysis). This target site is possibly in the 5' UTR of ORF116, inferring that the miRNA may not have any real effect on the expression of ORF116 (as target sites that occur outside 3' UTRs are generally less efficient).

The most likely target site for the 3' miRNA from MR5057 was predicted to be the 3' UTR of ORF122 /ORF123. In fact it was the only target site identified by TargetScan. It occurs near the start of the 3' UTR ORF123 which overlaps with the end of the 3'UTR of ORF122 (Figure 7.5). The location of this target site (i.e. in the 3' UTR of both ORFs) suggests that the miRNA could potentially regulate both of these transcripts.

Both target sites for miRNAs from MD11776 occur in 3' UTRs that also overlap with ORFs (Figure 7.7 and Figure 7.9). MiRNA target sites do not usually occur within ORFs, but this may be simply unavoidable in viral genomes where a lot of gene overlap occurs due to limitations on space. A target site for the KSHV miRNA hshv-miR-K5 also occurs in a similar genomic context (Lin and Ganem, 2011). However it is important to note that overlapping genes may not necessarily share the same primary RNA transcripts but instead may be under the control of different promoters. Therefore the target site may sometimes be transcribed as part of a UTR and as part of an ORF at other times, depending on which promoter is active. If this is the case with the CyHV-3 target sites for MD11776 miRNAs, then miRNA repression of the gene upstream of the UTR should not be subject to interference from the process of translation. The genes upstream from this the target site for

MD11776 3' are ORFs 87 (early gene), 85 (late gene) and 86 (late gene), however there are no details available as to their functions. The gene upstream of the target site for MD11776 is *ORF129* (early gene) and its function is also unknown.

In contrast there is more information available on the function of the genes targeted by the miRNAs from MR5057. The 5' miRNA target site is in the 3' UTR of *ORF117*. The encoded protein is predicted to have a hydrophobic region (Aoki et al., 2007) that is likely to represent a transmembrane domain, hence it may form part of the viral envelope. Also it was recently identified as an early gene (Ilouze et al., 2012a). *ORF123*, one of the candidate targets for the MR5057 3' miRNA has been identified as a deoxyuridine triphosphatase (Aoki et al., 2007) and therefore likely to be involved in nucleic acid metabolism and part of the viral DNA replication process. It was also identified as a viron protein (Michel et al., 2010b) also found in other herpesviruses such as HCMV (Varnum et al., 2004) and RRV (O'Connor and Kedes, 2006) and recently identified as an early gene (Ilouze et al., 2012a).

As the protein encoded by the early gene ORF117 may be incorporated into the viral envelope it is possible that the continuous expression of this protein is surplus to requirement once viral maturation is complete. However due to the fact that it is predicted to be targeted by the 5' miRNA (minor form) from MR5057, it is possible that this miRNA has no real effect on the expression of this gene unless the levels of this miRNA are caused to increase under other circumstances.

Deoxyuridine triphosphatase is involved in nucleic acid metabolism -specifically the processing of dUTP to dUMP, the precursor for thymine nucleotides. Therefore it is likely that the expression of *ORF123* (in particular) is excess to requirement once viral genomic replication is complete. It is tempting to speculate that the transcription of the MR5057 pri-miRNA may be regulated in a temporal manner ensuring that there are sufficiently high levels of the 3' miRNA from MR5057 processed to down-regulate expression of *ORF123* once it is no longer required. Therefore this miRNA may contribute to a dynamic and collective process that ensures correct temporal progression of viral gene expression during lytic infections.

CyHV-3 miRNAs can also target host miRNAs, and ideally carp 3' UTRs should also be assessed for possible CyHV-3 miRNA target sites. Unfortunately at the time of this study there was no compiled list of carp 3' UTRs available. However, there was a list of 18,233 3'UTRs for the related species *Danio rerio* (zebrafish) available from the UTR sequence database (utrdb.ba.itb.cnr.it/). As these two genomes show a high degree of similarity (based on initial data from the carp genome project) (Xu et al., 2011), these zebrafish 3' UTRs were used to as a proxy for Carp 3' UTRs in CyHV-3 miRNA target prediction with the initial aim of identifying high-likelihood predictions in the zebrafish genome and then searching for their conservation in the carp genome once the latter became sufficiently annotated. While significant progress has been made towards this in recent years (Henkel et al., 2012; Zhang et al., 2011a), it was not sufficiently annotated for these purposes before the end of this study. Therefore target site predictions in zebrafish 3' UTRs could not be checked for conservation in the corresponding carp 3' UTRs.

In is interesting that two very different approaches to miRNA target prediction gave identical results regarding the most likely CyHV-3 mRNA targets for the miRNAs from MR5057. The target sites that were found to conform most closely to the characteristics of *bone-fide* target sites using TargetScan were also the same sites that were found to be the most accessible using PITA. This is in agreement with the idea that genomes will tend to evolve in a way that ensures that important miRNA target sites remain physically accessible and available for base pairing with RISC bound miRNAs. These results regarding target prediction provide a solid basis for further experimental investigation.

# 9 Conclusion

This study has gathered experimental evidence to propose the existence of at least seven CyHV-3 encoded pre-miRNAs. All of these proposed pre-miRNAs were initially predicted to exist as part of the initial analysis of the CyHV-3 genome in order to determine its pre-miRNA coding potential prior to the commencement of experimental investigation. All of the pre-miRNAs and small RNAs (miRNAs and associated isomiRs) that map to these precursors display explicit indicative characteristics of genuine pre-miRNAs and miRNAs such as alignment signature,

miRNA duplex, stable pre-miRNA and specific enrichment in miRNA enriched samples. They also possess more subtle miRNA-like characteristics such as a general bias towards more 3' end heterogeneity among isomiRs and tendency to form discrete regions of read enrichment upon visualization of mapped data. Importantly all of these characteristics were shown to be consistent over two separate infections using two different CyHV-3 isolates.

The 3' miRNA from MR5057 was shown in all experiments to be the most abundant CyHV-3 miRNA and was the only miRNA detected by northern blotting providing conclusive evidence that MR5057 is a genuine pre-miRNA. Only miRNAs from MR5057 and MD11776 were detected by array hybridization. Even though miRNAs from MD11776 could not be detected by northern blotting, array hybridization showed that like the host miRNAs and the miRNAs from MR5057, the miRNAs from MD11776 were specifically enriched in the miRNA enriched sample only and not in the non-RNA sample from the same infection, indicating that this proposed pre-miRNA is more than likely genuine. Of all the CyHV-3 pre-miRNAs, MD11776 also showed the most signs of sequence and genomic location conservation in the two other CyHVs and in one of these (CyHV-2) it was the only potential homologue that was also likely to give rise to a functional pre-miRNA. The fact that this gene shows signs of conservation, in particular the region containing the major form miRNA, suggests that it has an important function as a miRNA, further supporting its annotation as a genuine pre-miRNA.

The miRNAs from the five remaining pre-mRNAs were present at low levels, therefore in addition to strong evidence from deep sequencing, miRNAs from these pre-miRNAs could only be detected through RT-qPCR, except for MD11704 which was not targeted by RT-qPCR. The fact that the expression levels of the major and minor forms derived from MD1111 and MR5075 were found to show consistency (i.e. that the major form was always present at higher levels), strongly supports the annotation of MD1111 and MR5075 as genuine miRNAs. Overall the use of RT-qPCR for relative quantification of CyHV-3 miRNA expression levels revealed a consistent expression profile for these CyHV-3 miRNAs during lytic infections both *in vitro* and *in vivo*, suggesting their expression is subject to specific underlying regulatory process. As similar expression profiles were found for these miRNAs *in*

*vitro* and *in vivo* it is highly likely that this profile would also be observed during lytic infections and across a broad range of experimental conditions.

In conclusion MR5057 is a genuine CyHV-3 encoded pre-miRNA. Based in the evidence gathered in this study, the six remaining proposed CyHV-3 pre-miRNAs are more than likely also genuine pre-miRNAs. Due to the fact that the miRNAs from MD11776 were the only other CyHV-3 miRNAs also detected by array hybridization, ultimately there is more evidence to indicate that these small RNAs are specifically enriched in the 17-25 nt RNA size range. Therefore out of the six remaining proposed pre-miRNAs, MD11776 is the most likely to be a genuine pre-miRNA. The detection of all of these transcripts by northern blotting would have been sufficient to confirm that they are in fact genuine miRNAs. The failure to detect any CyHV-3 miRNAs by northern blotting, other than the most abundant one, suggests that this was purely due to the sensitivity of the method. As discussed earlier, it may be possible to detect miRNAs at lower levels by repeating northern blotting using increased quantities of RNA. Due to their strong association with latency in other viruses, CyHV-3 miRNAs represent possible diagnostic markers for latent infections. The small exploratory experiment carried out at the end of this study, using samples from healthy fish, may not have been adequate enough to fully evaluate the potential usefulness of CyHV-3 encoded miRNAs in this context. As discussed earlier, for several reasons, carrying out a similar experiment using RNA samples from leucocytes may be a more suitable approach and may provide more clear cut answers in relation to the expression of these miRNAs during latency *in vivo.*

In accordance with the nomenclature system in use in miRBase the official names that have been suggested for the CyHV-3 pre-miRNAs identified in this study are shown in Table 9.1.

**Table 9.1 Official names suggested for all CyHV-3 pre-miRNAs identified in this study**
The names are based on the nomenclature used in miRBase. Pre-miRNAs are numbered based on their position on the CyHV3 genome (KHV-U strain Ac. DQ657948.1) Mature miRNAs take the same name as the pre-miRNA they are derived from except -mir is changed to –miR and the name ends with either -5p or -3p depending on depending on whether the miRNA id derived from the 5' or 3' arm.

| Names used in this study | Proposed names in accordance with miRBase miRNA nomenclature system |
|---|---|
| **MD1111** | CyHV3-mir-1 |
| **MD9812** | CyHV3-mir-2 |
| **MD11410** | CyHV3-mir-3 |
| **MD11704** | CyHV3-mir-4 |
| **MD11776** | CyHV3-mir-5 |
| **MR5057** | CyHV3-mir-6 |
| **MR5075** | CyHV3-mir-7 |

MiRNAs are important players in post-transcriptional regulation and are vital for many fundamental cellular processes in almost all eukaryotic organisms. Their unique characteristics also make them ideal tools for viruses, providing a genetically economic way to modulate both viral and host gene expression without stimulating an immune response. As with miRNAs from other viruses, it is likely that the CyHV-3 miRNAs identified in this study have specific functions during lytic infections. This may be particularly true of the miRNAs that are consistently highly expressed such as the 3' miRNA from MR5057 (CyHV3-miR-6-3p). In addition, a target for CyHV3-miR-6-3p was tentatively predicted at the end of this study giving a solid theoretical basis with which to proceed with further characterisation of this miRNA in terms of its function.

Although the functions of moRNAs are not yet known the additional discovery that CyHV-3 encodes moRNAs is intriguing. Together, the identification of both CyHV-3 encoded miRNAs and moRNAs opens up new lines of inquiry into the molecular basis of CyHV-3 infections. The detection of latent carriers will continue to play a key role in monitoring (and possibly controlling) the spread of this virus in many regions and the diagnostic potential of CyHV-3 encoded miRNAs in this context, can now be thoroughly evaluated. Furthermore, the identification of potential CyHV-3 pre-miRNA homologues in CyHV-1 and CyHV-2 provides interesting openings for investigations on the existence of miRNAs encoded by these viruses and other members of the *Alloherpesviridae* family.

# 10 Bibliography


Abend, J.R., Uldrick, T., Ziegelbauer, J.M., 2010. Regulation of tumor necrosis factor-like weak inducer of apoptosis receptor protein (TWEAKR) expression by Kaposi's sarcoma-associated herpesvirus microRNA prevents TWEAK-induced apoptosis and inflammatory cytokine expression. J. Virol. 84, 12139–12151.

Ackermann, M., 2004. Herpesviruses: a brief overview. Methods Mol. Biol. 256, 199–219.

Adkison, M.A., Gilad, O., Hedrick, R.P., 2005. An Enzyme Linked Immunosorbent Assay (ELISA) for Detection of Antibodies to the Koi Herpesvirus (KHV) in the Serum of Koi Cyprinus carpio. Fish Pathology 40, 53–62.

Agarwal, S., Vaz, C., Bhattacharya, A., Srinivasan, A., 2010. Prediction of novel precursor miRNAs using a context-sensitive hidden Markov model (CSHMM). BMC Bioinformatics 11, S29.

Akhova, O., Bainbridge, M., Misra, V., 2005. The neuronal host cell factor-binding protein Zhangfei inhibits herpes simplex virus replication. J. Virol. 79, 14708–14718.

Altmann, M., Hammerschmidt, W., 2005. Epstein-Barr virus provides a new paradigm: a requirement for the immediate inhibition of apoptosis. PLoS Biol. 3, e404.

Ambros, V., 2001. microRNAs: tiny regulators with great potential. Cell 107, 823–826.

Amen, M.A., Griffiths, A., 2011. Identification and Expression Analysis of Herpes B Virus-Encoded Small RNAs. J. Virol. 85, 7296–7311.

Amon, W., Farrell, P.J., 2005. Reactivation of Epstein-Barr virus from latency. Rev. Med. Virol. 15, 149–156.

Aoki, T., Hirono, I., Kurokawa, K., Fukuda, H., Nahary, R., Eldar, A., Davison, A.J., Waltzek, T.B., Bercovier, H., Hedrick, R.P., 2007. Genome sequences of three koi herpesvirus isolates representing the expanding distribution of an emerging disease threatening koi and common carp worldwide. Journal of virology 81, 5058–65.

Arbuckle, J.H., Medveczky, M.M., Luka, J., Hadley, S.H., Luegmayr, A., Ablashi, D., Lund, T.C., Tolar, J., De Meirleir, K., Montoya, J.G., Komaroff, A.L., Ambros, P.F., Medveczky, P.G., 2010. The latent human herpesvirus-6A genome specifically integrates in telomeres of human chromosomes in vivo and in vitro. Proc. Natl. Acad. Sci. U.S.A. 107, 5563–5568.

Arvin, A., Campadelli-Fiume, G., Mocarski, E., Moore, P.S., Roizman, B., Whitley, R., Yamanishi, K. (Eds.), 2007. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge University Press, Cambridge.

Atkins, K.E., Read, A.F., Savill, N.J., Renz, K.G., Walkden-Brown, S.W., Woolhouse, M.E., 2011. Modelling Marek's Disease Virus (MDV) infection: parameter estimates for mortality rate and infectiousness. BMC Veterinary Research 7, 70.

Baek, D., Villén, J., Shin, C., Camargo, F.D., Gygi, S.P., Bartel, D.P., 2008. The impact of microRNAs on protein output. Nature 455, 64–71.

Balon, E.K., 1995. Origin and domestication of the wild carp, Cyprinus carpio: from Roman gourmets to the swimming flowers. Aquaculture 129, 3–48.

Barad, O., Meiri, E., Avniel, A., Aharonov, R., Barzilai, A., Bentwich, I., Einav, U., Gilad, S., Hurban, P., Karov, Y., Lobenhofer, E.K., Sharon, E., Shiboleth, Y.M., Shtutman, M., Bentwich, Z., Einat, P., 2004. MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. Genome research 14, 2486–94.

Baskerville, S., Bartel, D.P., 2005. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA 11, 241–247.

Batuwita, R., Palade, V., 2009. microPred: effective classification of pre-miRNAs for human miRNA gene prediction. Bioinformatics 25, 989–995.

Beckham, C.J., Parker, R., 2008. P-bodies, stress granules and viral life cycles. Cell Host Microbe 3, 206–212.

Bellare, P., Ganem, D., 2009. Regulation of KSHV lytic switch protein expression by a virus-encoded microRNA: an evolutionary adaptation that fine-tunes lytic reactivation. Cell Host Microbe 6, 570.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., et al., 2008. Accurate Whole Human Genome Sequencing using Reversible Terminator Chemistry. Nature 456, 53–59.

Bentwich, I., 2005. Prediction and validation of microRNAs and their targets. FEBS Lett. 579, 5904–5910.

Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., Sharon, E., Spector, Y., Bentwich, Z., 2005. Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet 37, 766–70.

Bercovier, H., Fishman, Y., Nahary, R., Sinai, S., Zlotkin, A., Eyngor, M., Gilad, O., Eldar, A., Hedrick, R.P., 2005. Cloning of the koi herpesvirus (KHV) gene encoding thymidine kinase and its use for a highly sensitive PCR based diagnosis. BMC microbiology 5, 13.

Berezikov, E., Robine, N., Samsonova, A., Westholm, J.O., Naqvi, A., Hung, J.-H., Okamura, K., Dai, Q., Bortolamiol-Becet, D., Martin, R., Zhao, Y., Zamore, P.D., Hannon, G.J., Marra, M.A., Weng, Z., Perrimon, N., Lai, E.C., 2011. Deep annotation of Drosophila melanogaster microRNAs yields insights into their processing, modification, and emergence. Genome Res. 21, 203–215.

Bergmann, S.M., Sadowski, J., Kiełpiński, M., Bartłomiejczyk, M., Fichtner, D., Riebe, R., Lenk, M., Kempter, J., 2010. Susceptibility of koi x crucian carp and koi x goldfish hybrids to koi herpesvirus (KHV) and the development of KHV disease (KHVD). J. Fish Dis. 33, 267–272.

Bergmann, S.M., Schütze, H., Fischer, U., Fichtner, D., Riechardt, M., Meyer, K., Schrudde, D., Kempter, J., 2009. Detection of koi herpes virus (KHV) genome in apparently healthy fish. Bulletin of the European Association of Fish Pathologists 29, 145–152.

Bhende, P.M., Seaman, W.T., Delecluse, H.-J., Kenney, S.C., 2004. The EBV lytic switch protein, Z, preferentially binds to and activates the methylated viral genome. Nat. Genet. 36, 1099–1104.

Bloom, D.C., Giordani, N.V., Kwiatkowski, D.L., 2010. Epigenetic regulation of latent HSV-1 gene expression. Biochim. Biophys. Acta 1799, 246–256.

Bogerd, H.P., Karnowski, H.W., Cai, X., Shin, J., Pohlers, M., Cullen, B.R., 2010. A mammalian herpesvirus uses non-canonical expression and processing mechanisms to generate viral microRNAs. Mol Cell 37, 135.

Bondad-Reantaso, M.G., Sunarto, A., Subasinghe, R.P., 2007. Managing the koi herpesvirus disease outbreak in Indonesia and the lessons learned. Developments in biologicals 129, 21–8.

Bonnet, E., Wuyts, J., Rouze, P., Van de Peer, Y., 2004a. Detection of 91 potential conserved plant microRNAs in Arabidopsis thaliana and Oryza sativa identifies important target genes. Proc Natl Acad Sci U S A 101, 11511–11516.

Bonnet, E., Wuyts, J., Rouzé, P., Van de Peer, Y., 2004b. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. Bioinformatics 20, 2911–2917.

Bortoluzzi, S., Biasiolo, M., Bisognin, A., 2011. MicroRNA-offset RNAs (moRNAs): by-product spectators or functional players? Trends Mol Med 17, 473–474.

Bortoluzzi, S., Bisognin, A., Biasiolo, M., Guglielmelli, P., Biamonte, F., Norfo, R., Manfredini, R., Vannucchi, A.M., 2012. Characterization and discovery of novel miRNAs and moRNAs in JAK2V617F-mutated SET2 cells. Blood 119, e120–130.

Boss, I.W., Nadeau, P.E., Abbott, J.R., Yang, Y., Mergia, A., Renne, R., 2011. A Kaposi's Sarcoma-Associated Herpesvirus-Encoded Ortholog of MicroRNA miR-155 Induces Human Splenic B-Cell Expansion in NOD/LtSz-scid IL2R?null Mice▽. J Virol 85, 9877–9886.

Bowland, S.L., Shewen, P.E., 2000. Bovine respiratory disease: commercial vaccines currently available in Canada. Can Vet J 41, 33–48.

Boyle, J., Blackwell, J., 1991. Use of polymerase chain reaction to detect latent channel catfish virus. Am. J. Vet. Res. 52, 1965–1968.

Bretzinger, A., Fischer-scherl, T., Oumouna, M., Hoffmann, R., Truyen, U., 1999. Mass mortalities in koi carp, Cyprinus carpio, associated with gill and skin disease. Bull. Eur. Assoc. Fish Pathol. 19, 182–185.

Buck, A.H., Santoyo-Lopez, J., Robertson, K.A., Kumar, D.S., Reczko, M., Ghazal, P., 2007. Discrete Clusters of Virus-Encoded MicroRNAs Are Associated with Complementary Strands of the Genome and the 7.2-Kilobase Stable Intron in Murine Cytomegalovirus. J Virol 81, 13761–13770.

Cai, X., Hagedorn, C.H., Cullen, B.R., 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA 10, 1957–1966.

Cai, X., Lu, S., Zhang, Z., Gonzalez, C.M., Damania, B., Cullen, B.R., 2005. Kaposi's sarcoma-associated herpesvirus expresses an array of viral microRNAs in latently infected cells. Proc Natl Acad Sci U S A 102, 5570–5.

Cai, X., Schafer, A., Lu, S., Bilello, J.P., Desrosiers, R.C., Edwards, R., Raab-Traub, N., Cullen, B.R., 2006. Epstein-Barr virus microRNAs are evolutionarily conserved and differentially expressed. PLoS pathogens 2, e23.

Cantalupo, P., Doering, A., Sullivan, C.S., Pal, A., Peden, K.W.C., Lewis, A.M., Pipas, J.M., 2005. Complete Nucleotide Sequence of Polyomavirus SA12. J Virol 79, 13094–13104.

Cardone, G., Winkler, D.C., Trus, B.L., Cheng, N., Heuser, J.E., Newcomb, W.W., Brown, J.C., Steven, A.C., 2007. Visualization of the Herpes Simplex Virus Portal in situ by Cryo-electron Tomography. Virology 361, 426–434.

Cazalla, D., Xie, M., Steitz, J.A., 2011. A Primate Herpesvirus Uses the Integrator Complex to Generate Viral MicroRNAs. Mol Cell 43, 982–992.

Chau, C.M., Lieberman, P.M., 2004. Dynamic chromatin boundaries delineate a latency control region of Epstein-Barr virus. J. Virol. 78, 12308–12319.

Chen, C., Ridzon, D.A., Broomer, A.J., Zhou, Z., Lee, D.H., Nguyen, J.T., Barbisin, M., Xu, N.L., Mahuvakar, V.R., Andersen, M.R., Lao, K.Q., Livak, K.J., Guegler, K.J., 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic acids research 33, e179.

Chen, C.J., Kincaid, R.P., Seo, G.J., Bennett, M.D., Sullivan, C.S., 2011. Insights into Polyomaviridae microRNA function derived from study of the bandicoot papillomatosis carcinomatosis viruses. J. Virol. 85, 4487–4500.

Chen, S.-J., Chen, G.-H., Chen, Y.-H., Liu, C.-Y., Chang, K.-P., Chang, Y.-S., Chen, H.-C., 2010. Characterization of Epstein-Barr Virus miRNAome in Nasopharyngeal Carcinoma by Deep Sequencing. PLoS ONE 5, e12745.

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., Shiekhattar, R., 2005. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature 436, 740–744.

Cheung, A.K.L., Gottlieb, D.J., Plachter, B., Pepperl-Klindworth, S., Avdic, S., Cunningham, A.L., Abendroth, A., Slobedman, B., 2009. The role of the human cytomegalovirus UL111A gene in down-regulating CD4+ T-cell recognition of latently infected cells: implications for virus elimination during latency. Blood 114, 4128–4137.

Chiang, H.R., Schoenfeld, L.W., Ruby, J.G., Auyeung, V.C., Spies, N., Baek, D., Johnston, W.K., Russ, C., Luo, S., Babiarz, J.E., Blelloch, R., Schroth, G.P., Nusbaum, C., Bartel, D.P., 2010. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev. 24, 992–1009.

Choy, E.Y.-W., Siu, K.-L., Kok, K.-H., Lung, R.W.-M., Tsang, C.M., To, K.-F., Kwong, D.L.-W., Tsao, S.W., Jin, D.-Y., 2008. An Epstein-Barr virus-encoded microRNA targets PUMA to promote host cell survival. J. Exp. Med. 205, 2551–2560.

Cloonan, N., Wani, S., Xu, Q., Gu, J., Lea, K., Heater, S., Barbacioru, C., Steptoe, A.L., Martin, H.C., Nourbakhsh, E., Krishnan, K., Gardiner, B., Wang, X., Nones, K., Steen, J.A., Matigian, N.A., Wood, D.L., Kassahn, K.S., Waddell, N., Shepherd, J., Lee, C., Ichikawa, J., McKernan, K., Bramlett, K., Kuersten, S., Grimmond, S.M., 2011. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. Genome Biology 12, R126.

Costes, B., Fournier, G., Michel, B., Delforge, C., Raj, V.S., Dewals, B., Gillet, L., Drion, P., Body, A., Schynts, F., Lieffrig, F., Vanderplasschen, A., 2008. Cloning of the koi herpesvirus genome as an infectious bacterial artificial chromosome demonstrates that disruption of the thymidine kinase locus induces partial attenuation in Cyprinus carpio koi. Journal of virology 82, 4955–64.

Costes, B., Raj, V.S., Michel, B., Fournier, G., Thirion, M., Gillet, L., Mast, J., Lieffrig, F., Bremont, M., Vanderplasschen, A., 2009. The Major Portal of Entry of Koi Herpesvirus in Cyprinus carpio Is the Skin. J Virol 83, 2819–2830.

Cui, C., Griffiths, A., Li, G., Silva, L.M., Kramer, M.F., Gaasterland, T., Wang, X.J., Coen, D.M., 2006. Prediction and identification of herpes simplex virus 1-encoded microRNAs. J. Virol. 80, 5499–508.

Cullen, B.R., 2011. Viruses and microRNAs: RISCy interactions with serious consequences. Genes Dev. 25, 1881–1894.
David, L., Rothbard, S., Rubinstein, I., Katzman, H., Hulata, G., Hillel, J., Lavi, U., 2004. Aspects of red and black color inheritance in the Japanese ornamental (Koi) carp (Cyprinus carpio L.). Aquaculture 233, 129–147.
David, R., 2010. Viral infection: miRNAs help KSHV lay low. Nat Rev Micro 8, 158–159.
Davidovich, M., Dishon, A., Ilouze, M., Kotler, M., 2007. Susceptibility of cyprinid cultured cells to cyprinid herpesvirus 3. Arch. Virol. 152, 1541–1546.
Davison, A.J., 1992. Channel catfish virus: a new type of herpesvirus. Virology 186, 9–14.
Davison, A.J., 1998. The Genome of Salmonid Herpesvirus 1. J Virol 72, 1974–1982.
Davison, A.J., 2002. Evolution of the herpesviruses. Vet. Microbiol. 86, 69–88.
Davison, A.J., Dargan, D.J., Stow, N.D., 2002. Fundamental and accessory systems in herpesviruses. Antiviral Res. 56, 1–11.
Davison, A.J., Davison, M.D., 1995. Identification of structural proteins of channel catfish virus by mass spectrometry. Virology 206, 1035–1043.
Davison, A.J., Eberle, R., Ehlers, B., Hayward, G.S., McGeoch, D.J., Minson, A.C., Pellett, P.E., Roizman, B., Studdert, M.J., Thiry, E., 2009. The order Herpesvirales. Arch. Virol. 154, 171–177.
Davison, A.J., Kurobe, T., Gatherer, D., Cunningham, C., Korf, I., Fukuda, H., Hedrick, R.P., Waltzek, T.B., 2012. Comparative genomics of carp herpesviruses. J. Virol.
Denham, K., 2003. Koi herpesvirus in wild fish. Vet. Rec. 153, 507.
Dezulian, T., Remmert, M., Palatnik, J.F., Weigel, D., Huson, D.H., 2006. Identification of plant microRNA homologs. Bioinformatics 22, 359–360.
Dishon, A., Davidovich, M., Ilouze, M., Kotler, M., 2007. Persistence of cyprinid herpesvirus 3 in infected cultured carp cells. Journal of virology 81, 4828–36.
Dishon, A., Perelberg, A., Bishara-Shieban, J., Ilouze, M., Davidovich, M., Werker, S., Kotler, M., 2005. Detection of Carp Interstitial Nephritis and Gill Necrosis Virus in Fish Droppings. Appl Environ Microbiol 71, 7285–7291.
Do, J.H., Choi, D.-K., 2007. cDNA Labeling Strategies for Microarrays Using Fluorescent Dyes. Engineering in Life Sciences 7, 26–34.
Domselaar, R. van, Philippen, L.E., Quadir, R., Wiertz, E.J.H.J., Kummer, J.A., Bovenschen, N., 2010. Noncytotoxic Inhibition of Cytomegalovirus Replication through NK Cell Protease Granzyme M-Mediated Cleavage of Viral Phosphoprotein 71. J Immunol 185, 7605–7613.
Dunn, W., Trang, P., Zhong, Q., Yang, E., Van Belle, C., Liu, F., 2005. Human cytomegalovirus expresses novel microRNAs during productive viral infection. Cell. Microbiol. 7, 1684–1695.
Eide, K., Miller-Morgan, T., Heidel, J., Bildfell, R., Jin, L., 2011a. Results of total DNA measurement in koi tissue by Koi Herpes Virus real-time PCR. J. Virol. Methods 172, 81–84.
Eide, K.E., Miller-Morgan, T., Heidel, J.R., Kent, M.L., Bildfell, R.J., Lapatra, S., Watson, G., Jin, L., 2011b. Investigation of koi herpesvirus latency in koi. J. Virol. 85, 4954–4962.
Eulalio, A., Huntzinger, E., Izaurralde, E., 2008. Getting to the Root of miRNA-Mediated Gene Silencing. Cell 132, 9–14.

expasy.org, 2013. . ViralZone: Roseolovirus. URL http://viralzone.expasy.org/all_by_species/181.html

Fabian, M., Baumer, A., Steinhagen, D., 2013. Do wild fish species contribute to the transmission of koi herpesvirus to carp in hatchery ponds? Journal of Fish Diseases 36, 505–514.

Falke, D., Siegert, R., Vogell, W., 1959. [Electron microscopic findings on the problem of double membrane formation in herpes simplex virus]. Arch Gesamte Virusforsch 9, 484–496.

fao.org, 2010. Yearbook of Fishery Statistics. Food and Agriculture Organization of the United Nations.

fao.org, 2013a. FAO Fisheries & Aquaculture - Ornamental fish [WWW Document]. FAO Fisheries & Aquaculture - Ornamental fish. URL http://www.fao.org/fishery/topic/13611/en

fao.org, 2013b. Cultured Aquatic Species Information Programme Cyprinus carpio (Linnaeus, 1758) [WWW Document]. FAO Fisheries & Aquaculture Cyprinus carpio. URL http://www.fao.org/fishery/culturedspecies/Cyprinus_carpio/en

Fernandez-Valverde, S.L., Taft, R.J., Mattick, J.S., 2010. Dynamic isomiR regulation in Drosophila development. RNA 16, 1881–1888.

Fournier, G., Boutier, M., Stalin Raj, V., Mast, J., Parmentier, E., Vanderwalle, P., Peeters, D., Lieffrig, F., Farnir, F., Gillet, L., Vanderplasschen, A., 2012. Feeding Cyprinus carpio with infectious materials mediates cyprinid herpesvirus 3 entry through infection of pharyngeal periodontal mucosa. Vet Res 43, 6.

Friedländer, M.R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., Rajewsky, N., 2008. Discovering microRNAs from deep sequencing data using miRDeep. Nat. Biotechnol. 26, 407–415.

Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W., Rajewsky, N., 2012. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. 40, 37–52.

Fukunaga, R., Han, B.W., Hung, J.-H., Xu, J., Weng, Z., Zamore, P.D., 2012. Dicer partner proteins tune the length of mature miRNAs in flies and mammals. Cell 151, 533–546.

Garcia, D.M., Baek, D., Shin, C., Bell, G.W., Grimson, A., Bartel, D.P., 2011. Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs. Nature structural & molecular biology 18, 1139–46.

Garver, K.A., Al-Hussinee, L., Hawley, L.M., Schroeder, T., Edes, S., LePage, V., Contador, E., Russell, S., Lord, S., Stevenson, R.M.W., Souter, B., Wright, E., Lumsden, J.S., 2010. Mass Mortality Associated with Koi Herpesvirus in Wild Common Carp in Canada. jwildlifedis 46, 1242–1251.

Gilad, O., Yun, S., Adkison, M.A., Way, K., Willits, N.H., Bercovier, H., Hedrick, R.P., 2003. Molecular comparison of isolates of an emerging fish pathogen, koi herpesvirus, and the effect of water temperature on mortality of experimentally infected koi. The Journal of general virology 84, 2661–7.

Gilad, O., Yun, S., Zagmutt-Vergara, F.J., Leutenegger, C.M., Bercovier, H., Hedrick, R.P., 2004. Concentrations of a Koi herpesvirus (KHV) in tissues of experimentally infected Cyprinus carpio koi as assessed by real-time TaqMan PCR. Diseases of aquatic organisms 60, 179–87.

Glaunsinger, B.A., Ganem, D.E., 2006. Messenger RNA turnover and its regulation in herpesviral infection. Adv. Virus Res. 66, 337–394.

Glazov, E.A., Horwood, P.F., Assavalapsakul, W., Kongsuwan, K., Mitchell, R.W., Mitter, N., Mahony, T.J., 2010. Characterization of microRNAs encoded by the bovine herpesvirus 1 genome. J. Gen. Virol. 91, 32–41.

Granzow, H., Weiland, F., Jons, A., Klupp, B.G., Karger, A., Mettenleiter, T.C., 1997. Ultrastructural analysis of the replication cycle of pseudorabies virus in cell culture: a reassessment. J Virol 71, 2072–2082.

Gray, W.L., Williams, R.J., Jordan, R.L., Griffin, B.R., 1999. Detection of channel catfish virus DNA in latently infected catfish. J. Gen. Virol. 80 ( Pt 7), 1817–1822.

Grey, F., Hook, L., Nelson, J., 2008. The functions of herpesvirus-encoded microRNAs. Med Microbiol Immunol 197, 261–7.

Grey, F., Meyers, H., White, E.A., Spector, D.H., Nelson, J., 2007. A Human Cytomegalovirus-Encoded microRNA Regulates Expression of Multiple Viral Genes Involved in Replication. PLoS Pathog 3.

Grimmett, S.G., Warg, J.V., Getchell, R.G., Johnson, D.J., Bowser, P.R., 2006. An unusual koi herpesvirus associated with a mortality event of common carp Cyprinus carpio in New York State, USA. Journal of wildlife diseases 42, 658–62.

Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P., Bartel, D.P., 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Molecular cell 27, 91–105.

Grundhoff, A., Sullivan, C.S., 2011. Virus-encoded microRNAs. Virology 411, 325–343.

Grundhoff, A., Sullivan, C.S., Ganem, D., 2006. A combined computational and microarray-based approach identifies novel microRNAs encoded by human gamma-herpesviruses. RNA (New York, N.Y 12, 733–50.

Gu, S., Jin, L., Zhang, F., Sarnow, P., Kay, M.A., 2009. Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. Nat. Struct. Mol. Biol. 16, 144–150.

Gu, S., Kay, M.A., 2010. How do miRNAs mediate translational repression? Silence 1, 11.

Hackenberg, M., Rodríguez-Ezpeleta, N., Aransay, A.M., 2011. miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments. Nucleic Acids Res. 39, W132–138.

Hadinoto, V., Shapiro, M., Sun, C.C., Thorley-Lawson, D.A., 2009. The dynamics of EBV shedding implicate a central role for epithelial cells in amplifying viral output. PLoS Pathog. 5, e1000496.

Haenen, O.L.M., Way, K., Bergmann, S.M., Ariel, E., 2004. The emergence of koi herpesvirus and its significance to European aquaculture. Bull. Eur. Assoc. Fish Pathol. 24, 293–307.

Hafner, M., Renwick, N., Brown, M., Mihailović, A., Holoch, D., Lin, C., Pena, J.T.G., Nusbaum, J.D., Morozov, P., Ludwig, J., Ojo, T., Luo, S., Schroth, G., Tuschl, T., 2011. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. RNA.

Han, J., Lee, Y., Yeom, K.-H., Nam, J.-W., Heo, I., Rhee, J.-K., Sohn, S.Y., Cho, Y., Zhang, B.-T., Kim, V.N., 2006. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell 125, 887–901.

Han, J.E., Kim, J.H., Renault, T., Choresca, C., Shin, S.P., Jun, J.W., Park, S.C., 2013. Identifying the Viral Genes Encoding Envelope Glycoproteins for Differentiation of Cyprinid herpesvirus 3 Isolates. Viruses 5, 568–576.

Hanson, L., Dishon, A., Kotler, M., 2011. Herpesviruses that Infect Fish. Viruses 3, 2160–2191.

Haramoto, E., Kitajima, M., Katayama, H., Ohgaki, S., 2007. Detection of koi herpesvirus DNA in river water in Japan. J. Fish Dis. 30, 59–61.

Hedrick, R.P., Gilad, O., Yun, S., Spangenberg, J.V., Marty, G.D., Nordhausen, R.W., Kebus, M.J., Bercovier, H., Eldar, A., 2000. A Herpesvirus Associated with Mass Mortality of Juvenile and Adult Koi, a Strain of Common Carp. Journal of Aquatic Animal Health 12, 44–57.

Hedrick, R.P., Gilad, O., YUN, S.C., MCDOWELL, T.S., WALTZEK, T.B., KELLEY, G.O., Adkison, M.A., 2005. Initial Isolation and Characterization of a Herpes-like Virus (KHV) from Koi and Common Carp. Bull Fish Res Agen 2, 1.

Hedrick, R.P., Waltzek, T.B., McDowell, T.S., 2006. Susceptibility of Koi Carp, Common Carp, Goldfish, and Goldfish × Common Carp Hybrids to Cyprinid Herpesvirus-2 and Herpesvirus-3. Journal of Aquatic Animal Health 18, 26–34.

Heldwein, E.E., Krummenacher, C., 2008. Entry of herpesviruses into mammalian cells. Cell. Mol. Life Sci. 65, 1653–1668.

Hendrix, D., Levine, M., Shi, W., 2010. miRTRAP, a computational method for the systematic identification of miRNAs from high throughput sequencing data. Genome Biol. 11, R39.

Henkel, C.V., Dirks, R.P., Jansen, H.J., Forlenza, M., Wiegertjes, G.F., Howe, K., Van den Thillart, G.E.E.J.M., Spaink, H.P., 2012. Comparison of the exomes of common carp (Cyprinus carpio) and zebrafish (Danio rerio). Zebrafish 9, 59–67.

Hertel, L., Lacaille, V.G., Strobl, H., Mellins, E.D., Mocarski, E.S., 2003. Susceptibility of immature and mature Langerhans cell-type dendritic cells to infection and immunomodulation by human cytomegalovirus. J. Virol. 77, 7563–7574.

Ho, C.K., Wang, L.K., Lima, C.D., Shuman, S., 2004. Structure and Mechanism of RNA Ligase. Structure 12, 327–339.

Hofmann, H., Sindre, H., Stamminger, T., 2002. Functional Interaction between the pp71 Protein of Human Cytomegalovirus and the PML-Interacting Protein Human Daxx. J. Virol. 76, 5769–5783.

Hon, L.S., Zhang, Z., 2007. The roles of binding site arrangement and combinatorial targeting in microRNA repression of gene expression. Genome Biol 8, R166.

Honjo, M.N., Minamoto, T., Kawabata, Z., 2012. Reservoirs of Cyprinid herpesvirus 3 (CyHV-3) DNA in sediments of natural lakes and ponds. Vet. Microbiol. 155, 183–190.

Hsu, D.H., De Waal Malefyt, R., Fiorentino, D.F., Dang, M.N., Vieira, P., De Vries, J., Spits, H., Mosmann, T.R., Moore, K.W., 1990. Expression of interleukin-10 activity by Epstein-Barr virus protein BCRF1. Science 250, 830–832.

Huang, P.-J., Liu, Y.-C., Lee, C.-C., Lin, W.-C., Gan, R.R.-C., Lyu, P.-C., Tang, P., 2010. DSAP: deep-sequencing small RNA analysis pipeline. Nucleic Acids Res. 38, W385–391.

Huang, S., Hanson, L.A., 1998. Temporal Gene Regulation of the Channel Catfish Virus (Ictalurid Herpesvirus 1). J. Virol. 72, 1910–1917.

Huber, M.T., Compton, T., 1998. The Human Cytomegalovirus UL74 Gene Encodes the Third Component of the Glycoprotein H-Glycoprotein L-Containing Envelope Complex. J Virol 72, 8191–8197.

Huntzinger, E., Izaurralde, E., 2011. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet 12, 99–110.

Hussain, M., Taft, R.J., Asgari, S., 2008. An Insect Virus-Encoded MicroRNA Regulates Viral Replication. J Virol 82, 9164–9170.

Hutoran, M., Ronen, A., Perelberg, A., 2005. Description of an as of Yet Unclassified DNA Virus from Diseased Cyprinus Carpio. Journal of Virology 79, 1383–1991.

ICTV, 2013. ICTV Virus Taxonomy: 2012 Release (current) [WWW Document]. ICTV Virus Taxonomy: 2012 Release (current). URL http://www.ictvonline.org/virusTaxonomy.asp

Illumina.com, 2010. Illumina Sequencing Technology [WWW Document]. Illumina Sequencing Technology. URL http://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf

Ilouze, M., Davidovich, M., Diamant, A., Kotler, M., Dishon, A., 2011. The outbreak of carp disease caused by CyHV-3 as a model for new emerging viral diseases in aquaculture: a review. Ecol Res 26, 885–892.

Ilouze, M., Dishon, A., Kahan, T., Kotler, M., 2006. Cyprinid herpes virus-3 (CyHV-3) bears genes of genetically distant large DNA viruses. FEBS letters 580, 4473–8.

Ilouze, M., Dishon, A., Kotler, M., 2012a. Coordinated and sequential transcription of the cyprinid herpesvirus-3 annotated genes. Virus research 169, 98–106.

Ilouze, M., Dishon, A., Kotler, M., 2012b. Down-regulation of the cyprinid herpesvirus-3 annotated genes in cultured cells maintained at restrictive high temperature. Virus Res. 169, 289–295.

Ito, T., Sano, M., Kurita, J., Yuasa, K., Iida, T., 2007. Carp Larvae Are Not Susceptible to Koi Herpesvirus. Fish Pathology 42, 107–109.

Iwakiri, D., Takada, K., 2010. Chapter 4 - Role of EBERs in the Pathogenesis of EBV Infection, in: George F. Vande Woude and George Klein (Ed.), Advances in Cancer Research. Academic Press, pp. 119–136.

Jacob, R.J., Morse, L.S., Roizman, B., 1979. Anatomy of herpes simplex virus DNA. XII. Accumulation of head-to-tail concatemers in nuclei of infected cells and their role in the generation of the four isomeric arrangements of viral DNA. J. Virol. 29, 448–457.

Jagadeeswaran, G., Zheng, Y., Sumathipala, N., Jiang, H., Arrese, E.L., Soulages, J.L., Zhang, W., Sunkar, R., 2010. Deep sequencing of small RNA libraries reveals dynamic regulation of conserved and novel microRNAs and microRNA-stars during silkworm development. BMC genomics 11, 52.

Jenkins, C., Garcia, W., Godwin, M.J., Spencer, J.V., Stern, J.L., Abendroth, A., Slobedman, B., 2008. Immunomodulatory Properties of a Viral Homolog of Human Interleukin-10 Expressed by Human Cytomegalovirus during the Latent Phase of Infection. J Virol 82, 3736–3750.

Ji, X., 2008. The mechanism of RNase III action: how dicer dices. Curr. Top. Microbiol. Immunol. 320, 99–116.

Jiang, H., Wong, W.H., 2008. SeqMap: mapping massive amount of oligonucleotides to the genome. Bioinformatics (Oxford, England) 24, 2395–6.

Jiang, P., Wu, H., Wang, W., Ma, W., Sun, X., Lu, Z., 2007. MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. Nucleic acids research 35, W339–44.

John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., Marks, D.S., 2004. Human MicroRNA targets. PLoS Biol. 2, e363.

Johnson, D.C., Baines, J.D., 2011. Herpesviruses remodel host membranes for virus egress. Nat Rev Micro 9, 382–394.

Jurak, I., Kramer, M.F., Mellor, J.C., Van Lint, A.L., Roth, F.P., Knipe, D.M., Coen, D.M., 2010. Numerous conserved and divergent microRNAs expressed by herpes simplex viruses 1 and 2. J. Virol. 84, 4659–4672.

Kalla, M., Schmeinck, A., Bergbauer, M., Pich, D., Hammerschmidt, W., 2010. AP-1 homolog BZLF1 of Epstein-Barr virus has two essential functions dependent on the epigenetic state of the viral genome. Proc. Natl. Acad. Sci. U.S.A. 107, 850–855.

Kaufer, B.B., Jarosinski, K.W., Osterrieder, N., 2011. Herpesvirus telomeric repeats facilitate genomic integration into host telomeres and mobilization of viral DNA during reactivation. J. Exp. Med. 208, 605–615.

Kempter, J., Bergmann, S.M., 2007. Detection of koi herpesvirus (KHV) genome in wild and farmed fish from Northern Poland. Aquaculture 272, Supplement 1, S275.

Kempter, J., Sadowski, J., Schuetze, H., Fischer, U., Dauber, M., Fichtner, D., Panicz, R., Bergmann, S.M., 2009. Koi herpes virus: do acipenserid restitution programs pose a threat to carp farms in the disease-free zones? Acta Ichthyologica et Piscatoria v. 39(2) p.119-126.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., Segal, E., 2007. The role of site accessibility in microRNA target recognition. Nature genetics 39, 1278–84.

Khvorova, A., Reynolds, A., Jayasena, S.D., 2003. Functional siRNAs and miRNAs exhibit strand bias. Cell 115, 209–216.

Kielpinski, M., Kempter, J., Panicz, R., Sadowski, J., Schütze, H., Ohlemeyer, S., Bergmann, S.M., 2010. Detection of KHV in Freshwater Mussels and Crustaceans from Ponds with KHV History in Common Carp (Cyprinus carpio).

Kim, V.N., Han, J., Siomi, M.C., 2009. Biogenesis of small RNAs in animals. Nature reviews 10, 126–39.

Kincaid, R.P., Burke, J.M., Sullivan, C.S., 2012. RNA virus microRNA that mimics a B-cell oncomiR. PNAS 109, 3077–3082.

Kincaid, R.P., Sullivan, C.S., 2012. Virus-Encoded microRNAs: An Overview and a Look to the Future. PLoS Pathog 8, e1003018.

Kiriakidou, M., Nelson, P.T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z., Hatzigeorgiou, A., 2004. A combined computational-experimental approach predicts human microRNA targets. Genes Dev 18, 1165–1178.

Knickelbein, J.E., Khanna, K.M., Yee, M.B., Baty, C.J., Kinchington, P.R., Hendricks, R.L., 2008. Noncytotoxic lytic granule-mediated CD8+ T cell inhibition of HSV-1 reactivation from neuronal latency. Science 322, 268–271.

Kongchum, P., Palti, Y., Hallerman, E.M., Hulata, G., David, L., 2010. SNP discovery and development of genetic markers for mapping innate immune response genes in common carp (Cyprinus carpio). Fish & Shellfish Immunology 29, 356–361.

Koscianska, E., Starega-Roslan, J., Krzyzosiak, W.J., 2011. The Role of Dicer Protein Partners in the Processing of MicroRNA Precursors. PLoS ONE 6, e28548.

Kozomara, A., Griffiths-Jones, S., 2011. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 39, D152–157.

Kraus, R.J., Perrigoue, J.G., Mertz, J.E., 2003. ZEB negatively regulates the lytic-switch BZLF1 gene promoter of Epstein-Barr virus. J. Virol. 77, 199–207.

Krek, A., Grün, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., Da Piedade, I., Gunsalus, K.C., Stoffel, M., Rajewsky, N., 2005. Combinatorial microRNA target predictions. Nat. Genet. 37, 495–500.

Ku, W.-C., Lau, W.K., Tseng, Y.-T., Tzeng, C.-M., Chiu, S.-K., 2004. Dextran sulfate provides a quantitative and quick microarray hybridization reaction. Biochem. Biophys. Res. Commun. 315, 30–37.

Kuchenbauer, F., Morin, R.D., Argiropoulos, B., Petriv, O.I., Griffith, M., Heuser, M., Yung, E., Piper, J., Delaney, A., Prabhu, A.-L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., Hansen, C.L., Marra, M.A., Humphries, R.K., 2008. In-depth characterization of the microRNA transcriptome in a leukemia progression model. Genome Res. 18, 1787–1797.

Küppers, R., 2003. B cells under influence: transformation of B cells by Epstein-Barr virus. Nat. Rev. Immunol. 3, 801–812.

La Boissiere, S., Hughes, T., O'Hare, P., 1999. HCF-dependent nuclear import of VP16. EMBO J 18, 480–489.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., Tuschl, T., 2001. Identification of novel genes coding for small expressed RNAs. Science 294, 853–858.

Lai, E.C., Tomancak, P., Williams, R.W., Rubin, G.M., 2003. Computational identification of Drosophila microRNA genes. Genome Biology 4, R42.

Laichalk, L.L., Thorley-Lawson, D.A., 2005. Terminal Differentiation into Plasma Cells Initiates the Replicative Cycle of Epstein-Barr Virus In Vivo. J Virol 79, 1296–1307.

Langenberger, D., Bermudez-Santana, C., Hertel, J., Hoffmann, S., Khaitovich, P., Stadler, P.F., 2009. Evidence for human microRNA-offset RNAs in small RNA sequencing data. Bioinformatics 25, 2298–2301.

Lau, N.C., Lim, L.P., Weinstein, E.G., Bartel, D.P., 2001. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science 294, 858–862.

Le Morvan, C., Deschaux, P., Troutaud, D., 1996. Effects and mechanisms of environmental temperature on carp (Cyprinus carpio) anti-DNP antibody response and non-specific cytotoxic cell activity: a kinetic study. Dev Comp Immunol 20, 331–40.

Le Morvan, C., Troutaud, D., Deschaux, P., 1998. Differential effects of temperature on specific and nonspecific immune defences in fish. J Exp Biol 201, 165–8.

Lee, C.-T., Risom, T., Strauss, W.M., 2007. Evolutionary conservation of microRNA regulatory circuits: an examination of microRNA gene complexity and conserved microRNA-target interactions through metazoan phylogeny. DNA Cell Biol. 26, 209–218.

Lee, L.W., Zhang, S., Etheridge, A., Ma, L., Martin, D., Galas, D., Wang, K., 2010. Complexity of the microRNA repertoire revealed by next-generation sequencing. RNA 16, 2170–2180.

Lee, R.C., Ambros, V., 2001. An extensive class of small RNAs in Caenorhabditis elegans. Science 294, 862–864.

Lee, R.C., Feinbaum, R.L., Ambros, V., 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75, 843–854.

Lee, S.H., Kalejta, R.F., Kerry, J., Semmes, O.J., O'Connor, C.M., Khan, Z., Garcia, B.A., Shenk, T., Murphy, E., 2012. BclAF1 restriction factor is neutralized by proteasomal degradation and microRNA repression during human cytomegalovirus infection. Proc. Natl. Acad. Sci. U.S.A. 109, 9575–9580.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., Kim, V.N., 2003. The nuclear RNase III Drosha initiates microRNA processing. Nature 425, 415–419.

Lee, Y., Jeon, K., Lee, J.-T., Kim, S., Kim, V.N., 2002. MicroRNA maturation: stepwise processing and subcellular localization. EMBO J 21, 4663–4670.

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H., Kim, V.N., 2004. MicroRNA genes are transcribed by RNA polymerase II. EMBO J 23, 4051–4060.

Lerner, M.R., Andrews, N.C., Miller, G., Steitz, J.A., 1981. Two small RNAs encoded by Epstein-Barr virus and complexed with protein are precipitated by antibodies from patients with systemic lupus erythematosus. Proc Natl Acad Sci U S A 78, 805–809.

Lewis, B.P., Burge, C.B., Bartel, D.P., 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120, 15–20.

Li, Y., Zhang, Z., Liu, F., Vongsangnak, W., Jing, Q., Shen, B., 2012. Performance comparison and evaluation of software tools for microRNA deep-sequencing data analysis. Nucl. Acids Res.

Liao, I.C., Chao, N.-H., 2009. Aquaculture and food crisis: opportunities and constraints. Asia Pac J Clin Nutr 18, 564–569.

Ligas, M.W., Johnson, D.C., 1988. A herpes simplex virus mutant in which glycoprotein D sequences are replaced by beta-galactosidase sequences binds to but is unable to penetrate into cells. J Virol 62, 1486–1494.

Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., Bartel, D.P., 2003. The microRNAs of Caenorhabditis elegans. Genes Dev 17, 991–1008.

Lin, H.R., Ganem, D., 2011. Viral microRNA target allows insight into the role of translation in governing microRNA target accessibility. Proceedings of the National Academy of Sciences of the United States of America 108, 5148–53.

Lin, S.-L., Cheng, Y.-H., Wen, C.-M., Chen, S.-N., 2013. Characterization of a novel cell line from the caudal fin of koi carp Cyprinus carpio. Journal of Fish Biology.

Lin, Y.-T., Kincaid, R.P., Arasappan, D., Dowd, S.E., Hunicke-Smith, S.P., Sullivan, C.S., 2010. Small RNA profiling reveals antisense transcription throughout the KSHV genome and novel small RNAs. RNA 16, 1540–1558.

Ling, P.D., Lednicky, J.A., Keitel, W.A., Poston, D.G., White, Z.S., Peng, R., Liu, Z., Mehta, S.K., Pierson, D.L., Rooney, C.M., Vilchez, R.A., Smith, E.O., Butel, J.S., 2003. The Dynamics of Herpesvirus and Polyomavirus Reactivation and Shedding in Healthy Adults: A 14-Month Longitudinal Study. J Infect Dis. 187, 1571–1580.

Linsen, S.E.V., De Wit, E., Janssens, G., Heater, S., Chapman, L., Parkin, R.K., Fritz, B., Wyman, S.K., De Bruijn, E., Voest, E.E., Kuersten, S., Tewari, M.,

Cuppen, E., 2009. Limitations and possibilities of small RNA digital gene expression profiling. Nat. Methods 6, 474–476.
Lio-Po, G.D., 2011. Recent developments in the study and surveillance of koi herpesvirus (KHV) in Asia. Diseases in Asian Aquaculture 7, 13–28.
Liu, C.-G., Calin, G.A., Meloon, B., Gamliel, N., Sevignani, C., Ferracin, M., Dumitru, C.D., Shimizu, M., Zupo, S., Dono, M., Alder, H., Bullrich, F., Negrini, M., Croce, C.M., 2004. An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. Proc. Natl. Acad. Sci. U.S.A. 101, 9740–9744.
Livesey, F.J., 2003. Strategies for microarray analysis of limiting amounts of RNA. Brief Funct Genomic Proteomic 2, 31–36.
Ljungman, P., 1993. Herpes virus infections in immunocompromised patients: problems and therapeutic interventions. Ann. Med. 25, 329–333.
Lo, A.K.F., To, K.F., Lo, K.W., Lung, R.W.M., Hui, J.W.Y., Liao, G., Hayward, S.D., 2007. Modulation of LMP1 protein expression by EBV-encoded microRNAs. Proc Natl Acad Sci U S A 104, 16164–16169.
Lowther, J.A., Henshilwood, K., Lees, D.N., 2008. Determination of Norovirus Contamination in Oysters from Two Commercial Harvesting Areas over an Extended Period, Using Semiquantitative Real-Time Reverse Transcription PCR. Journal of Food Protection® 71, 1427–1433.
Mardis, E.R., 2008. Next-Generation DNA Sequencing Methods. Annual Review of Genomics and Human Genetics 9, 387–402.
Masotti, A., Caputo, V., Da Sacco, L., Pizzuti, A., Dallapiccola, B., Bottazzo, G.F., 2009. Quantification of Small Non-Coding RNAs Allows an Accurate Comparison of miRNA Expression Profiles. Journal of Biomedicine and Biotechnology 2009, 1–9.
Mathelier, A., Carbone, A., 2010. MIReNA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data. Bioinformatics 26, 2226–2234.
McCleary, S.J., Ruane, N.M., Cheslett, D., Hickey, C., Rodger, H.D., Geoghegan, F., Henshilwood, K., 2011. Detection of koi herpesvirus (KHV) in koi carp (Cyprinus carpio L.) imported into Ireland. Bulletin of the European Association of Fish Pathologists 31, 124–128.
Meshesha, M.K., Veksler-Lublinsky, I., Isakov, O., Reichenstein, I., Shomron, N., Kedem, K., Ziv-Ukelson, M., Bentwich, Z., Avni, Y.S., 2012. The microRNA Transcriptome of Human Cytomegalovirus (HCMV). Open Virol J 6, 38–48.
Michel, B., Fournier, G., Lieffrig, F., Costes, B., Vanderplasschen, A., 2010a. Cyprinid Herpesvirus 3. Emerg Infect Dis 16, 1835–1843.
Michel, B., Leroy, B., Stalin Raj, V., Lieffrig, F., Mast, J., Wattiez, R., Vanderplasschen, A.F., Costes, B., 2010b. The genome of cyprinid herpesvirus 3 encodes 40 proteins incorporated in mature virions. J. Gen. Virol. 91, 452–462.
Minamoto, T., Honjo, M.N., Yamanaka, H., Tanaka, N., Itayama, T., Kawabata, Z., 2011. Detection of cyprinid herpesvirus-3 DNA in lake plankton. Res. Vet. Sci. 90, 530–532.
Miranda, K.C., Huynh, T., Tay, Y., Ang, Y.-S., Tam, W.-L., Thomson, A.M., Lim, B., Rigoutsos, I., 2006. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. Cell 126, 1203–1217.

Miska, E.A., Alvarez-Saavedra, E., Townsend, M., Yoshii, A., Sestan, N., Rakic, P., Constantine-Paton, M., Horvitz, H.R., 2004. Microarray analysis of microRNA expression in the developing mammalian brain. Genome Biol. 5, R68.

Miwa, S., Ito, T., Sano, M., 2007. Morphogenesis of koi herpesvirus observed by electron microscopy. Journal of fish diseases 30, 715–22.

Miyazaki, T., Kuzuya, Y., Yasumoto, S., Yasuda, M., Kobayashi, T., 2008a. Histopathological and ultrastructural features of Koi herpesvirus (KHV)-infected carp Cyprinus carpio, and the morphology and morphogenesis of KHV. Dis. Aquat. Org. 80, 1–11.

Miyazaki, T., Yasumoto, S., Kuzuya, Y., Yoshimura, T., 2008b. A primary study on oral vaccination with liposomes entrapping koi herpes virus (KHV) antigens against KHV infection in carp. Diseases in Asian Aquaculture 7, 99–184.

Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.-L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., Eaves, C.J., Marra, M.A., 2008. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res 18, 610–621.

Morissette, G., Flamand, L., 2010. Herpesviruses and Chromosomal Integration. J. Virol. 84, 12100–12109.

Munson, D.J., Burch, A.D., 2012. A novel miRNA produced during lytic HSV-1 infection is important for efficient replication in tissue culture. Arch. Virol. 157, 1677–1688.

Nachmani, D., Stern-Ginossar, N., Sarid, R., Mandelboim, O., 2009. Diverse herpesvirus microRNAs target the stress-induced immune ligand MICB to escape recognition by natural killer cells. Cell Host Microbe 5, 376–385.

Neukirch, M., Böttcher, K., Bunnajirakul, S., 1999. Isolation of a virus from koi with altered gills. Bull Eur Assoc Fish Pathol 19, 221–224.

Ng, K.L.S., Mishra, S.K., 2007. De novo SVM classification of precursor microRNAs from genomic pseudo hairpins using global and intrinsic folding measures. Bioinformatics 23, 1321–1330.

Nuryati, S., ' A., ' S., Soejoedono, R.D., Santika, A., Pasaribu, F.H., Sumantadinata, K., 2012. Construction of a DNA Vaccine Using Glycoprotein Gene and Its Expression Towards Increasing Survival Rate of KHV-Infected Common Carp (Cyprinus carpio). Jurnal Natur Indonesia 13.

Nuwaysir, E.F., Huang, W., Albert, T.J., Singh, J., Nuwaysir, K., Pitas, A., Richmond, T., Gorski, T., Berg, J.P., Ballin, J., McCormick, M., Norton, J., Pollock, T., Sumwalt, T., Butcher, L., Porter, D., Molla, M., Hall, C., Blattner, F., Sussman, M.R., Wallace, R.L., Cerrina, F., Green, R.D., 2002. Gene Expression Analysis Using Oligonucleotide Arrays Produced by Maskless Photolithography. Genome Res 12, 1749–1755.

O'Connor, C.M., Kedes, D.H., 2006. Mass spectrometric analyses of purified rhesus monkey rhadinovirus reveal 33 virion-associated proteins. J. Virol. 80, 1574–1583.

O'Hare, P., Goding, C.R., 1988. Herpes simplex virus regulatory elements and the immunoglobulin octamer domain bind a common factor and are both targets for virion transactivation. Cell 52, 435–445.

Okamura, K., Hagen, J.W., Duan, H., Tyler, D.M., Lai, E.C., 2007. The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell 130, 89–100.

Okamura, K., Phillips, M.D., Tyler, D.M., Duan, H., Chou, Y.T., Lai, E.C., 2008. The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution. Nature structural & molecular biology 15, 354–63.

Ouellet, D.L., Plante, I., Landry, P., Barat, C., Janelle, M.-E., Flamand, L., Tremblay, M.J., Provost, P., 2008. Identification of functional microRNAs released through asymmetrical processing of HIV-1 TAR element. Nucleic Acids Res. 36, 2353–2365.

Oulas, A., Boutla, A., Gkirtzou, K., Reczko, M., Kalantidis, K., Poirazi, P., 2009. Prediction of novel microRNA genes in cancer-associated genomic regions--a combined computational and experimental approach. Nucleic Acids Res. 37, 3276–3287.

Pearson, H., 2004. Carp virus crisis prompts moves to avert global spread. Nature 427, 577.

Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P., Fodor, S.P., 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Natl Acad Sci U S A 91, 5022–5026.

Pedersen, J.S., Bejerano, G., Siepel, A., Rosenbloom, K., Lindblad-Toh, K., Lander, E.S., Kent, J., Miller, W., Haussler, D., 2006. Identification and Classification of Conserved RNA Secondary Structures in the Human Genome. PLoS Comput Biol 2, e33.

Peng, T., Ponce de Leon, M., Novotny, M.J., Jiang, H., Lambris, J.D., Dubin, G., Spear, P.G., Cohen, G.H., Eisenberg, R.J., 1998. Structural and antigenic analysis of a truncated form of the herpes simplex virus glycoprotein gH-gL complex. J. Virol. 72, 6092–6103.

Penkert, R.R., Kalejta, R.F., 2011. Tegument protein control of latent herpesvirus establishment and animation. Herpesviridae 2, 3.

Perelberg, A., Ilouze, M., Kotler, M., Steinitz, M., 2008. Antibody response and resistance of Cyprinus carpio immunized with cyprinid herpes virus 3 (CyHV-3). Vaccine 26, 3750–3756.

Perelberg, A., Ronen, A., Hutoran, M., Smith, Y., Kotler, M., 2005. Protection of cultured Cyprinus carpio against a lethal viral disease by an attenuated virus vaccine. Vaccine 23, 3396–403.

Perelberg, A., Smirnov, M., Hutoran, M., Diamant, A., Bejerano, Y., Kotler, M., 2003. Epidemiological description of a new viral disease afflicting cultured Cyprinus carpio in Israel. Isr J Aquacult-Bamid 55, 5–12.

Perng, G.C., Jones, C., Ciacci-Zanella, J., Stone, M., Henderson, G., Yukht, A., Slanina, S.M., Hofman, F.M., Ghiasi, H., Nesburn, A.B., Wechsler, S.L., 2000. Virus-induced neuronal apoptosis blocked by the herpes simplex virus latency-associated transcript. Science 287, 1500–1503.

Pfeffer, S., Lagos-Quintana, M., Tuschl, T., 2005a. Cloning of small RNA molecules. Curr Protoc Mol Biol Chapter 26, Unit 26.4.

Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grasser, F.A., Van Dyk, L.F., Ho, C.K., Shuman, S., Chien, M., Russo, J.J., Ju, J., Randall, G., Lindenbach, B.D., Rice, C.M., Simon, V., Ho, D.D., Zavolan, M., Tuschl, T., 2005b. Identification of microRNAs of the herpesvirus family. Nature methods 2, 269–76.

Pfeffer, S., Zavolan, M., Grasser, F.A., Chien, M., Russo, J.J., Ju, J., John, B., Enright, A.J., Marks, D., Sander, C., Tuschl, T., 2004. Identification of virus-encoded microRNAs. Science (New York, N.Y 304, 734–6.

Pikarsky, E., Ronen, A., Abramowitz, J., Levavi-Sivan, B., Hutoran, M., Shapira, Y., Steinitz, M., Perelberg, A., Soffer, D., Kotler, M., 2004. Pathogenesis of acute viral disease induced in fish by carp interstitial nephritis and gill necrosis virus. Journal of virology 78, 9544–51.

Pokorova, D., Piackova, V., Cizek, S., Reschova, S., Hulova, J., Vicenova, M., Vesely, T., 2007. Tests for the presence of koi herpesvirus (KHV) in common carp (Cyprinus carpio carpio) and koi carp (Cyprinus carpio koi) in the Czech Republic. Vet Med-Czech 52, 562–568.

Pokorova, D., Vesely, T., Piackova, V., 2005. Current Knowledge on koi Herpesvirus. Veterinarni Medicina 50, 139–147.

Pratt, Z.L., Kuzembayeva, M., Sengupta, S., Sugden, B., 2009. The microRNAs of Epstein-Barr Virus are expressed at dramatically differing levels among cell lines. Virology 386, 387–397.

Quail, M.A., Kozarewa, I., Smith, F., Scally, A., Stephens, P.J., Durbin, R., Swerdlow, H., Turner, D.J., 2008. A large genome center's improvements to the Illumina sequencing system. Nat Methods 5, 1005–10.

Raj, V.S., Fournier, G., Rakus, K., Ronsmans, M., Ouyang, P., Michel, B., Delforges, C., Costes, B., Farnir, F., Leroy, B., Wattiez, R., Melard, C., Mast, J., Lieffrig, F., Vanderplasschen, A., 2011. Skin mucus of Cyprinus carpio inhibits cyprinid herpesvirus 3 binding to epidermal cells. Veterinary Research 42, 92.

Rakus, K.Ł., Wiegertjes, G.F., Adamek, M., Siwicki, A.K., Lepa, A., Irnazarow, I., 2009. Resistance of common carp (Cyprinus carpio L.) to Cyprinid herpesvirus-3 is influenced by major histocompatibility (MH) class II B gene polymorphism. Fish & Shellfish Immunology 26, 737–743.

Reese, T.A., Xia, J., Johnson, L.S., Zhou, X., Zhang, W., Virgin, H.W., 2010. Identification of Novel MicroRNA-Like Molecules Generated from Herpesvirus and Host tRNA Transcripts. J Virol 84, 10344–53.

Reeves, M.B., Sinclair, J.H., 2010. Analysis of latent viral gene expression in natural and experimental latency models of human cytomegalovirus and its correlation with histone modifications at a latent promoter. J. Gen. Virol. 91, 599–604.

Reid, J.G., Nagaraja, A.K., Lynn, F.C., Drabek, R.B., Muzny, D.M., Shaw, C.A., Weiss, M.K., Naghavi, A.O., Khan, M., Zhu, H., Tennakoon, J., Gunaratne, G.H., Corry, D.B., Miller, J., McManus, M.T., German, M.S., Gibbs, R.A., Matzuk, M.M., Gunaratne, P.H., 2008. Mouse let-7 miRNA populations exhibit RNA editing that is constrained in the 5?-seed/ cleavage/anchor regions and stabilize predicted mmu-let-7a:mRNA duplexes. Genome Res 18, 1571–1581.

Riley, K.J., Rabinowitz, G.S., Steitz, J.A., 2010. Comprehensive analysis of Rhesus lymphocryptovirus microRNA expression. J Virol 84, 5148–57.

Riley, K.J., Rabinowitz, G.S., Yario, T.A., Luna, J.M., Darnell, R.B., Steitz, J.A., 2012. EBV and human microRNAs co-target oncogenic and apoptotic viral and human genes during latency. EMBO J. 31, 2207–2221.

Ro, S., Park, C., Young, D., Sanders, K.M., Yan, W., 2007. Tissue-dependent paired expression of miRNAs. Nucleic acids research 35, 5944–53.

Roberts, H.E., 2011. Fundamentals of Ornamental Fish Health. John Wiley & Sons.

Rochat, R.H., Liu, X., Murata, K., Nagayama, K., Rixon, F.J., Chiu, W., 2011. Seeing the Portal in Herpes Simplex Virus Type 1 B Capsids. J Virol 85, 1871–1874.

Roizmann, B., 2001. The family Herpesviridae: A brief introduction, in: Fields Virology. pp. 2381–2397.

Ronen, A., Perelberg, A., Abramowitz, J., Hutoran, M., Tinman, S., Bejerano, I., Steinitz, M., Kotler, M., 2003. Efficient vaccine against the virus causing a lethal disease in cultured Cyprinus carpio. Vaccine 21, 4677–84.

Rosenkranz, D., Klupp, B.G., Teifke, J.P., Granzow, H., Fichtner, D., Mettenleiter, T.C., Fuchs, W., 2008. Identification of envelope protein pORF81 of koi herpesvirus. J Gen Virol 89, 896–900.

Rouha, H., Thurner, C., Mandl, C.W., 2010. Functional microRNA generated from a cytoplasmic RNA virus. Nucleic Acids Res. 38, 8328–8337.

Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., Bartel, D.P., 2006. Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell 127, 1193–1207.

Ruby, J.G., Stark, A., Johnston, W.K., Kellis, M., Bartel, D.P., Lai, E.C., 2007. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. Genome Res 17, 1850–1864.

Ruijter, J.M., Ramakers, C., Hoogaars, W.M., Karlen, Y., Bakker, O., Van den Hoff, M.J., Moorman, A.F., 2009. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic acids research 37, e45.

Sadler, J., Marecaux, E., Goodwin, A.E., 2008. Detection of koi herpes virus (CyHV-3) in goldfish, Carassius auratus (L.), exposed to infected koi. Journal of Fish Diseases 31, 71–72.

Saffert, R.T., Kalejta, R.F., 2006. Inactivating a Cellular Intrinsic Immune Defense Mediated by Daxx Is the Mechanism through Which the Human Cytomegalovirus pp71 Protein Stimulates Viral Immediate-Early Gene Expression. J. Virol. 80, 3863–3871.

Saffert, R.T., Penkert, R.R., Kalejta, R.F., 2010. Cellular and viral control over the initial events of human cytomegalovirus experimental latency in CD34+ cells. J. Virol. 84, 5594–5604.

Sano, M., Ito, T., Kurita, J., Yanai, T., Wantanabe, N., Miwa, S., Iida, T., 2004. First detection of koi herpesvirus in cultured common carp Cyprinus carpio in Japan. Fish Pathology 39, 165–167.

Sano, N., Moriwake, M., Hondo, R., Sano, T., 1993. Herpesvirus cyprini: a search for viral genome in infected fish by infected fish by in situ hybridization. Journal of Fish Diseases 16, 495–499.

Sano, T., Morita, N., Shima, N., Akimoto, M., 1991. Herpesvirus cyprini: lethality and oncogenicity. Journal of Fish Diseases 14, 533–543.

Sawtell, N.M., Poon, D.K., Tansky, C.S., Thompson, R.L., 1998. The latent herpes simplex virus type 1 genome copy number in individual neurons is virus strain specific and correlates with reactivation. J. Virol. 72, 5343–5350.

Schwarz, D.S., Hutvágner, G., Du, T., Xu, Z., Aronin, N., Zamore, P.D., 2003. Asymmetry in the assembly of the RNAi enzyme complex. Cell 115, 199–208.

Scott, E.S., O'Hare, P., 2001. Fate of the Inner Nuclear Membrane Protein Lamin B Receptor and Nuclear Lamins in Herpes Simplex Virus Type 1 Infection. J. Virol. 75, 8818–8830.

Sears, J., Ujihara, M., Wong, S., Ott, C., Middeldorp, J., Aiyar, A., 2004. The amino terminus of Epstein-Barr Virus (EBV) nuclear antigen 1 contains AT hooks that facilitate the replication and partitioning of latent EBV genomes by tethering them to cellular chromosomes. J. Virol. 78, 11487–11505.

Seo, G.J., Chen, C.J., Sullivan, C.S., 2009. Merkel cell polyomavirus encodes a microRNA with the ability to autoregulate viral gene expression. Virology 383, 183–187.

Seo, G.J., Fink, L.H.L., O'Hara, B., Atwood, W.J., Sullivan, C.S., 2008. Evolutionarily Conserved Function of a Viral MicroRNA. J Virol 82, 9823–9828.

Shapira, Y., Magen, Y., Zak, T., Kotler, M., Hulata, G., Levavi-Sivan, B., 2005. Differential resistance to koi herpes virus (KHV)/carp interstitial nephritis and gill necrosis virus (CNGV) among common carp (Cyprinus carpio L.) strains and crossbreds. Aquaculture 245, 1–11.

Shi, W., Hendrix, D., Levine, M., Haley, B., 2009. A distinct class of small RNAs arises from pre-miRNA–proximal regions in a simple chordate. Nat Struct Mol Biol 16, 183–189.

Shimizu, T., Yoshida, N., Kasai, H., Yoshimizu, M., 2006. Survival of koi herpesvirus (KHV) in environmental water. Fish Pathology v. 41(4) p. 153-157.

Shingara, J., KEIGER, K., SHELTON, J., LAOSINCHAI-WOLF, W., POWERS, P., CONRAD, R., BROWN, D., LABOURIER, E., 2005. An optimized isolation and labeling platform for accurate microRNA expression profiling. RNA 11, 1461–1470.

Shukla, D., Spear, P.G., 2001. Herpesviruses and heparan sulfate: an intimate relationship in aid of viral entry. J. Clin. Invest. 108, 503–510.

Sinclair, J., 2010. Chromatin structure regulates human cytomegalovirus gene expression during latency, reactivation and lytic infection. Biochim. Biophys. Acta 1799, 286–295.

Sinha, S., Vasulu, T.S., De, R.K., 2009. Performance and Evaluation of MicroRNA Gene Identification Tools. Journal of Proteomics & Bioinformatics 02, 336–343.

Skalsky, R.L., Cullen, B.R., 2010. Viruses, microRNAs, and Host Interactions. Annual Review of Microbiology 64, 123–141.

Söderberg-Nauclér, C., Streblow, D.N., Fish, K.N., Allan-Yorke, J., Smith, P.P., Nelson, J.A., 2001. Reactivation of latent human cytomegalovirus in CD14(+) monocytes is differentiation dependent. J. Virol. 75, 7543–7554.

Sorefan, K., Pais, H., Hall, A.E., Kozomara, A., Griffiths-Jones, S., Moulton, V., Dalmay, T., 2012. Reducing ligation bias of small RNAs in libraries for next generation sequencing. Silence 3, 4.

sourceforge.net, 2013. MicroRNA Discovery By Deep Sequencing [WWW Document]. MicroRNA Discovery By Deep Sequencing. URL http://sourceforge.net/projects/mireap/

Speck, S.H., Ganem, D., 2010. Viral latency and its regulation: lessons from the gamma-herpesviruses. Cell Host Microbe 8, 100–115.

Starega-Roslan, J., Krol, J., Koscianska, E., Kozlowski, P., Szlachcic, W.J., Sobczak, K., Krzyzosiak, W.J., 2011. Structural basis of microRNA length variety. Nucleic Acids Res 39, 257–268.

Stern, S., Tanaka, M., Herr, W., 1989. The Oct-1 homoeodomain directs formation of a multiprotein-DNA complex with the HSV transactivator VP16. Nature 341, 624–630.

Stern-Ginossar, N., Elefant, N., Zimmermann, A., Wolf, D.G., Saleh, N., Biton, M., Horwitz, E., Prokocimer, Z., Prichard, M., Hahn, G., Goldman-Wohl, D., Greenfield, C., Yagel, S., Hengel, H., Altuvia, Y., Margalit, H., Mandelboim,

O., 2007. Host immune system gene targeting by a viral miRNA. Science 317, 376–381.

St-Hilaire, S., Beevers, N., Way, K., Le Deuff, R.M., Martin, P., Joiner, C., 2005. Reactivation of koi herpesvirus infections in common carp Cyprinus carpio. Diseases of aquatic organisms 67, 15–23.

Stingley, R.L., Gray, W.L., 2000. Transcriptional regulation of the channel catfish virus genome direct repeat region. J Gen Virol 81, 2005–2010.

Stroop, W.G., Rock, D.L., Fraser, N.W., 1984. Localization of herpes simplex virus in the trigeminal and olfactory systems of the mouse central nervous system during acute and latent infections by in situ hybridization. Lab. Invest. 51, 27–38.

Su, H., Trombly, M.I., Chen, J., Wang, X., 2009. Essential and overlapping functions for mammalian Argonautes in microRNA silencing. Genes Dev. 23, 304–317.

Suffert, G., Malterer, G., Hausser, J., Viiliäinen, J., Fender, A., Contrant, M., Ivacevic, T., Benes, V., Gros, F., Voinnet, O., Zavolan, M., Ojala, P.M., Haas, J.G., Pfeffer, S., 2011. Kaposi's sarcoma herpesvirus microRNAs target caspase 3 and regulate apoptosis. PLoS Pathog. 7, e1002405.

Sullivan, C.S., 2008. New roles for large and small viral RNAs in evading host defences. Nat Rev Genet 9, 503–7.

Sullivan, C.S., Grundhoff, A., 2007. Identification of viral microRNAs. Methods Enzymol 427, 3–23.

Sullivan, C.S., Grundhoff, A.T., Tevethia, S., Pipas, J.M., Ganem, D., 2005. SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. Nature 435, 682–6.

Sullivan, C.S., Sung, C.K., Pack, C.D., Grundhoff, A., Lukacher, A.E., Benjamin, T.L., Ganem, D., 2009. Murine Polyomavirus encodes a microRNA that cleaves early RNA transcripts but is not essential for experimental infection. Virology 387, 157–167.

Sunarto, A., Liongue, C., McColl, K.A., Adams, M.M., Bulach, D., Crane, M.S.J., Schat, K.A., Slobedman, B., Barnes, A.C., Ward, A.C., Walker, P.J., 2012. Koi herpesvirus encodes and expresses a functional interleukin-10. J. Virol. 86, 11512–11520.

Syakuri, H., Adamek, M., Brogden, G., Rakus, K.Ł., Matras, M., Irnazarow, I., Steinhagen, D., 2013. Intestinal barrier of carp (Cyprinus carpio L.) during a cyprinid herpesvirus 3-infection: molecular identification and regulation of the mRNA expression of claudin encoding genes. Fish Shellfish Immunol. 34, 305–314.

Szakonyi, G., Klein, M.G., Hannan, J.P., Young, K.A., Ma, R.Z., Asokan, R., Holers, V.M., Chen, X.S., 2006. Structure of the Epstein-Barr virus major envelope glycoprotein. Nat. Struct. Mol. Biol. 13, 996–1001.

Taft, R.J., Simons, C., Nahkuri, S., Oey, H., Korbie, D.J., Mercer, T.R., Holst, J., Ritchie, W., Wong, J.J.-L., Rasko, J.E.J., Rokhsar, D.S., Degnan, B.M., Mattick, J.S., 2010. Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans. Nat. Struct. Mol. Biol. 17, 1030–1034.

Takane, K., Kanai, A., 2011. Vertebrate virus-encoded microRNAs and their sequence conservation. Jpn. J. Infect. Dis. 64, 357–366.

Taylor, N.G.H., Dixon, P.F., Jeffery, K.R., Peeler, E.J., Denham, K.L., Way, K., 2010. Koi herpesvirus: distribution and prospects for control in England and Wales. J. Fish Dis. 33, 221–230.

Tempera, I., Lieberman, P.M., 2010. Chromatin organization of gammaherpesvirus latent genomes. Biochim. Biophys. Acta 1799, 236–245.

Terhune, J.S., Grizzle, J.M., Hayden, K., Mcclenahan, S.., 2004. First report of koi herpesvirus in wild common carp in the Western Hemisphere. Fish Health Newsletter.American Fisheries Society 32, 8–9.

Teune, J.-H., Steger, G., 2010. NOVOMIR: De Novo Prediction of MicroRNA-Coding Regions in a Single Plant-Genome. Journal of Nucleic Acids 2010, 1–10.

Thompson, R.L., Preston, C.M., Sawtell, N.M., 2009. De Novo Synthesis of VP16 Coordinates the Exit from HSV Latency In Vivo. PLoS Pathog 5, e1000352.

Thomson, J.M., Parker, J., Perou, C.M., Hammond, S.M., 2004. A custom microarray platform for analysis of microRNA gene expression. Nat. Methods 1, 47–53.

Toedling, J., Servant, N., Ciaudo, C., Farinelli, L., Voinnet, O., Heard, E., Barillot, E., 2012. Deep-Sequencing Protocols Influence the Results Obtained in Small-RNA Sequencing. PLoS ONE 7, e32724.

Tomari, Y., Du, T., Zamore, P.D., 2007. Sorting of Drosophila small silencing RNAs. Cell 130, 299–308.

Tomé, A.R., Kuś, K., Correia, S., Paulo, L.M., Zacarias, S., De Rosa, M., Figueiredo, D., Parkhouse, R.M.E., Athanasiadis, A., 2013. Crystal structure of a poxvirus-like zalpha domain from cyprinid herpesvirus 3. J. Virol. 87, 3998–4004.

Tronstein E, J.C., 2011. GEnital shedding of herpes simplex virus among symptomatic and asymptomatic persons with hsv-2 infection. JAMA 305, 1441–1449.

Tsai, K., Thikmyanova, N., Wojcechowskyj, J.A., Delecluse, H.-J., Lieberman, P.M., 2011. EBV Tegument Protein BNRF1 Disrupts DAXX-ATRX to Activate Viral Early Gene Transcription. PLoS Pathog 7, e1002376.

Tsutsumi, A., Kawamata, T., Izumi, N., Seitz, H., Tomari, Y., 2011. Recognition of the pre-miRNA structure by Drosophila Dicer-1. Nat Struct Mol Biol 18, 1153–1158.

Tuddenham, L., Jung, J.S., Chane-Woon-Ming, B., Dolken, L., Pfeffer, S., 2012. Small RNA Deep Sequencing Identifies MicroRNAs and Other Small Noncoding RNAs from Human Herpesvirus 6B. J Virol 86, 1638–1649.

Uchii, K., Matsui, K., Iida, T., Kawabata, Z., 2009. Distribution of the introduced cyprinid herpesvirus 3 in a wild population of common carp, Cyprinus carpio L. Journal of Fish Diseases 32, 857–864.

Umbach, J.L., Cullen, B.R., 2010. In-depth analysis of Kaposi's sarcoma-associated herpesvirus microRNA expression provides insights into the mammalian microRNA-processing machinery. J Virol 84, 695–703.

Umbach, J.L., Nagel, M.A., Cohrs, R.J., Gilden, D.H., Cullen, B.R., 2009a. Analysis of Human Alphaherpesvirus MicroRNA Expression in Latently Infected Human Trigeminal Ganglia. J Virol 83, 10677–10683.

Umbach, J.L., Strelow, L.I., Wong, S.W., Cullen, B.R., 2010. Analysis of rhesus rhadinovirus microRNAs expressed in virus-induced tumors from infected rhesus macaques. Virology 405, 592–599.

Umbach, J.L., Wang, K., Tang, S., Krause, P.R., Mont, E.K., Cohen, J.I., Cullen, B.R., 2009b. Identification of viral microRNAs expressed in human sacral ganglia latently infected with herpes simplex virus 2. J Virol 84, 1189–92.

Van Beurden, S.J., Bossers, A., Voorbergen-Laarman, M.H.A., Haenen, O.L.M., Peters, S., Abma-Henkens, M.H.C., Peeters, B.P.H., Rottier, P.J.M., Engelsma, M.Y., 2010. Complete genome sequence and taxonomic position of anguillid herpesvirus 1. J Gen Virol 91, 880–887.

Van Beurden, S.J., Gatherer, D., Kerr, K., Galbraith, J., Herzyk, P., Peeters, B.P.H., Rottier, P.J.M., Engelsma, M.Y., Davison, A.J., 2012. Anguillid Herpesvirus 1 Transcriptome. J Virol 86, 10150–10161.

Van Beurden, S.J., Leroy, B., Wattiez, R., Haenen, O.L., Boeren, S., Vervoort, J.J., Peeters, B.P., Rottier, P.J., Engelsma, M.Y., Vanderplasschen, A.F., 2011. Identification and localization of the structural proteins of anguillid herpesvirus 1. Vet. Res. 42, 105.

Van Beurden, S.J., Peeters, B.P.H., Rottier, P.J.M., Davison, A.J., Engelsma, M.Y., 2013. Genome-wide gene expression analysis of anguillid herpesvirus 1. BMC Genomics 14, 83.

Varble, A., Chua, M.A., Perez, J.T., Manicassamy, B., García-Sastre, A., tenOever, B.R., 2010. Engineered RNA viral synthesis of microRNAs. PNAS.

Varnum, S.M., Streblow, D.N., Monroe, M.E., Smith, P., Auberry, K.J., Pasa-Tolic, L., Wang, D., Camp, D.G., 2nd, Rodland, K., Wiley, S., Britt, W., Shenk, T., Smith, R.D., Nelson, J.A., 2004. Identification of proteins in human cytomegalovirus (HCMV) particles: the HCMV proteome. J. Virol. 78, 10960–10966.

Viollet, S., Fuchs, R.T., Munafo, D.B., Zhuang, F., Robb, G.B., 2011. T4 RNA Ligase 2 truncated active site mutants: improved tools for RNA analysis. BMC Biotechnology 11, 72.

Voinnet, O., 2009. Origin, biogenesis, and activity of plant microRNAs. Cell 136, 669–687.

Walster, C.I., 1999. Clinical observations of severe mortalities in koi carp, Cyprinus carpio, with gill disease. Fish Veterinary Journal 3, 54–58.

Waltzek, T.B., Kelley, G.O., Stone, D.M., Way, K., Hanson, L., Fukuda, H., Hirono, I., Aoki, T., Davison, A.J., Hedrick, R.P., 2005. Koi herpesvirus represents a third cyprinid herpesvirus (CyHV-3) in the family Herpesviridae. J. Gen. Virol. 86, 1659–67.

Waltzek, T.B.T., Kelley, G.O.G., Alfaro, M.E.M., Kurobe, T.T., Davison, A.J.A., Hedrick, R.P.R., 2009. Phylogenetic relationships in the family Alloherpesviridae. Dis Aquat Organ 84, 179.

Walz, N., Christalla, T., Tessmer, U., Grundhoff, A., 2010. A global analysis of evolutionary conservation among known and predicted gammaherpesvirus microRNAs. J Virol 84, 716–28.

Wang, K., Lau, T.Y., Morales, M., Mont, E.K., Straus, S.E., 2005a. Laser-capture microdissection: refining estimates of the quantity and distribution of latent herpes simplex virus 1 and varicella-zoster virus DNA in human trigeminal Ganglia at the single-cell level. J. Virol. 79, 14079–14087.

Wang, L., Liu, H., Li, D., Chen, H., 2011. Identification and Characterization of Maize microRNAs Involved in the Very Early Stage of Seed Germination. BMC genomics 12, 154.

Wang, W.-C., Lin, F.-M., Chang, W.-C., Lin, K.-Y., Huang, H.-D., Lin, N.-S., 2009. miRExpress: Analyzing high-throughput sequencing data for profiling microRNA expression. BMC Bioinformatics 10, 328.

Wang, X., Zhang, J., Li, F., Gu, J., He, T., Zhang, X., Li, Y., 2005b. MicroRNA identification based on sequence and structure alignment. Bioinformatics 21, 3610–3614.

Washietl, S., Hofacker, I.L., Stadler, P.F., 2005. Fast and reliable prediction of noncoding RNAs. PNAS 102, 2454–2459.

Way, K., 2007. Manual of Diagnostic Tests for Aquatic Animals, Chapter 2.1.17, Koi Herpesvirus Disease. OIE.

Way, K., 2008. Koi herpesvirus and goldfishherpes virus: an update of current knowledge and research at Cefas. Fish Veterinary Journal 10, 62–73.

Way, K., 2012. Manual of Diagnostic Tests for Aquatic Animals 2012. Chapter 2.3.6. — Koi herpesvirus disease. pp. 328–344.

Way, K., Beevers, N., Joiner, C., Stone, D.M., St-Hilaire, S., Denham, K.L., Dixon, P.F., 2004a. Koi herpesvirus in the UK: Detection in archive tissue samples and spread of the virus to wild carp. Abstract 6thInternational Symposium on Viruses of LowerVertebrates, Hakodate, Japan, Sept.2004.

Way, K., Le Deuff, R.M., Stone, D.M., Denham, K.L., St-Hilaire, S., 2004b. Koi herpesvirus: Diagnostics and research at CEFAS Weymouth laboratory 2000 – 2003. Report of Int. workshop on Koi Herpesvirus, London, 12-13 Feb 2004 p.9-10.

Wheeler, B.M., Heimberg, A.M., Moy, V.N., Sperling, E.A., Holstein, T.W., Heber, S., Peterson, K.J., 2009. The deep evolution of metazoan microRNAs. Evol. Dev. 11, 50–68.

Whittington, R.J., Chong, R., 2007. Global trade in ornamental fish from an Australian perspective: The case for revised import risk analysis and management strategies. Preventive Veterinary Medicine 81, 92–116.

Wiegant, J.C., Van Gijlswijk, R.P., Heetebrij, R.J., Bezrookove, V., Raap, A.K., Tanke, H.J., 1999. ULS: a versatile method of labeling nucleic acids for FISH based on a monofunctional reaction of cisplatin derivatives with guanine moieties. Cytogenet. Cell Genet. 87, 47–52.

Wildy, P., Watson, D.H., 1962. Electron Microscopic Studies on the Architecture of Animal Viruses. Cold Spring Harb Symp Quant Biol 27, 25–47.

Winter, J., Link, S., Witzigmann, D., Hildenbrand, C., Previti, C., Diederichs, S., 2013. Loop-miRs: active microRNAs generated from single-stranded loop regions. Nucl. Acids Res.

Wolf, K., Darlington, R.W., 1971. Channel catfish virus: a new herpesvirus of ictalurid fish. J. Virol. 8, 525–533.

Wolf, K., Darlington, R.W., Taylor, W.G., Quimby, M.C., Nagabayashi, T., 1978. Herpesvirus salmonis: Characterization of a New Pathogen of Rainbow Trout. J. Virol. 27, 659–666.

Woodhall, D.L., Groves, I.J., Reeves, M.B., Wilkinson, G., Sinclair, J.H., 2006. Human Daxx-mediated repression of human cytomegalovirus gene expression correlates with a repressive chromatin structure around the major immediate early promoter. J. Biol. Chem. 281, 37652–37660.

Workman, C., Krogh, A., 1999. No evidence that mRNAs have lower folding free energies than random sequences with the same dinucleotide distribution. Nucleic Acids Res 27, 4816–4822.

Wu, H., Ye, C., Ramirez, D., Manjunath, N., 2009. Alternative Processing of Primary microRNA Transcripts by Drosha Generates 5′ End Variation of Mature microRNA. PLoS ONE 4, e7566.

Wu, Y.-Q., Chen, D.-J., He, H.-B., Chen, D.-S., Chen, L.-L., Chen, H.-C., Liu, Z.-F., 2012. Pseudorabies Virus Infected Porcine Epithelial Cell Line Generates a Diverse Set of Host MicroRNAs and a Special Cluster of Viral MicroRNAs. PLoS ONE 7, e30988.

Xu, J.-R., Bently, J., Beck, L., Reed, A., Miller-Morgan, T., Heidel, J.R., Kent, M.L., Rockey, D.D., Jin, L., 2013. Analysis of koi herpesvirus latency in wild common carp and ornamental koi in Oregon, USA. J. Virol. Methods 187, 372–379.

Xu, P., Li, J., Li, Y., Cui, R., Wang, J., Wang, J., Zhang, Y., Zhao, Z., Sun, X., 2011. Genomic insight into the common carp (Cyprinus carpio) genome by sequencing analysis of BAC-end sequences. BMC Genomics 12, 188.

Xu, Y., Zhou, X., Zhang, W., 2008. MicroRNA prediction with a novel ranking algorithm based on random walks. Bioinformatics 24, i50–i58.

Xue, C., Li, F., He, T., Liu, G.-P., Li, Y., Zhang, X., 2005. Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. BMC Bioinformatics 6, 310.

Yao, Y., Smith, L.P., Petherbridge, L., Watson, M., Nair, V., 2012. Novel microRNAs encoded by duck enteritis virus. Journal of General Virology 93, 1530–1536.

Yao, Y., Zhao, Y., Xu, H., Smith, L.P., Lawrie, C.H., Sewer, A., Zavolan, M., Nair, V., 2007. Marek's Disease Virus Type 2 (MDV-2)-Encoded MicroRNAs Show No Sequence Conservation with Those Encoded by MDV-1. J. Virol. 81, 7164–7170.

Yasumoto, S., Kuzuya, Y., Yasuda, M., Yoshimura, T., Miyazaki, T., 2006. Oral immunization of common carp with a liposome vaccine fusing koi herpesvirus antigen. Fish Pathology v. 41(4) p. 141-145.

Yi, R., Qin, Y., Macara, I.G., Cullen, B.R., 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. 17, 3011–3016.

Yu, X., Wang, Z., Mertz, J.E., 2007. ZEB1 regulates the latent-lytic switch in infection by Epstein-Barr virus. PLoS Pathog. 3, e194.

Yuasa, K., Motohiko, S., Jun, K., Takafumi, I., Takaji, I., 2005. Improvement of a PCR Method with the Sph I-5 Primer Set for the Detection of Koi Herpesvirus (KHV). Fish Pathology 40, 37–39.

Yuasa, K., Sano, M., 2009. Koi Herpesvirus: Status of Outbreaks, Diagnosis, Surveillance, and Research.

Zak, T., Perelberg, A., Magen, I., Milstein, A., Joseph, D., 2007. Heterosis in the Growth Rate of Hungarian-Israeli Common Carp Crossbreeds and Evaluation of their Sensitivity to Koi Herpes Virus (KHV) Disease.

Zhang, J., Xu, Y., Huan, Q., Chong, K., 2009. Deep sequencing of Brachypodium small RNAs at the global genome level identifies microRNAs involved in cold stress response. BMC Genomics 10, 449.

Zhang, W., Gao, S., Zhou, X., Xia, J., Chellappan, P., Zhou, X., Zhang, X., Jin, H., 2010. Multiple distinct small RNAs originate from the same microRNA precursors. Genome Biology 11, R81.

Zhang, Y., Stupka, E., Henkel, C.V., Jansen, H.J., Spaink, H.P., Verbeek, F.J., 2011a. Identification of common carp innate immune genes with whole-genome sequencing and RNA-Seq data. J Integr Bioinform 8, 169.

Zhang, Y., Yang, Y., Zhang, H., Jiang, X., Xu, B., Xue, Y., Cao, Y., Zhai, Q., Zhai, Y., Xu, M., Cooke, H.J., Shi, Q., 2011b. Prediction of novel pre-microRNAs

with high accuracy through boosting and SVM. Bioinformatics 27, 1436–1437.

Zhao, Y., Xu, H., Yao, Y., Smith, L.P., Kgosana, L., Green, J., Petherbridge, L., Baigent, S.J., Nair, V., 2011. Critical Role of the Virus-Encoded MicroRNA-155 Ortholog in the Induction of Marek's Disease Lymphomas. PLoS Pathog 7.

Zhou, Z.H., Chen, D.H., Jakana, J., Rixon, F.J., Chiu, W., 1999. Visualization of Tegument-Capsid Interactions and DNA in Intact Herpes Simplex Virus Type 1 Virions. J. Virol. 73, 3210–3218.

Zhu, E., Zhao, F., Xu, G., Hou, H., Zhou, L., Li, X., Sun, Z., Wu, J., 2010a. mirTools: microRNA profiling and discovery based on high-throughput sequencing. Nucl. Acids Res. 38, W392–W397.

Zhu, J.Y., Strehle, M., Frohn, A., Kremmer, E., Hofig, K.P., Meister, G., Adler, H., 2010b. Identification and Analysis of Expression of Novel MicroRNAs of Murine Gammaherpesvirus 68. J Virol 84, 10266–75.

Ziegelbauer, J.M., Sullivan, C.S., Ganem, D., 2009. Tandem array-based expression screens identify host mRNA targets of virus-encoded microRNAs. Nat Genet 41, 130–134.