THE UNIVERSITY
of EDINBURGH

The Interaction of Host and Viral MicroRNAs with Infectious Laryngotracheitis Virus Transcripts

Jack Ferguson

THE UNIVERSITY *of* EDINBURGH

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2019

Declaration

I declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work described in this thesis has been executed by myself, all work of other authors is duly acknowledged.

Jack Ferguson

Division of Infection and Immunity
The Roslin Institute
Royal (Dick) School of Veterinary Studies
University of Edinburgh
Easter Bush Campus
Edinburgh
Midlothian
EH25 9RG

Acknowledgements

Firstly, I would like to thank my supervisors Bob Dalziel and Finn Grey. They have provided help, support and guidance throughout the duration of my PhD and listened irrespective of how daft the question/comment or suggestion may have been. The project would not have been a success without their insights and knowledge. I'd also like to thank Bob's whiteboard for the endless diagrams and scrawling's that have helped shaped this project from the get go.

Secondly, I would like to thank Inga Dry and Spring Tan as current members of the Dalziel lab for putting up with me for the last few years. Your advice, friendship and support has been invaluable to me throughout.

In addition, I would like to thank Katie Nightingale and Erin Brady as previous member of the Dalziel lab for their friendship. As well as this, I would also like to thank the wider virology department at the Roslin Institute for their valued friendship and time over the last three years. In particular, I would like to thank both Marlynne Quigg-Nicol and Nikki Smith.

I am also like to thank my peers and other PhD students both in the Division of Infection and Immunity and the wider Institute. They have provided support, an ear, friendship and questionably too many nights in the pub along the way.

This thesis would not be complete without the acknowledgement of cake club. The club has provided a weekly dose of much needed sugar and discussions on a strange mixture of topics. I'd like to thank all members past and present but particular thanks go to emeritus president Dominique McCormick and current incumbent president Elly Gaunt.

Away from the Institute, I'd like to thank the players of Dollys F.C. past and present for the laugher over the duration of my PhD. It was an honour to be able to lead the team for two seasons as captain, something that was thoroughly enjoyable. Up the Dollys!

Without question, I would like to thank my family and friends back in Yorkshire and further afield. The unconditional support and patience throughout my whole education has been priceless. I would also like to thank Chloe Brown and Bruno the dog for the last 3 years of support and adventures away from the lab.

Abstract

Infectious Laryngotracheitis Virus (ILTV) is an Alphaherpesvirus of the domesticated chicken and other economically important fowl such as pheasants, peafowl and turkeys. It causes an upper respiratory disease that is clinically characterised by dyspnoea, rales and expulsion of a thick, sometimes hemorrhagic, tracheal exudate. Incidences of mortality range from 10 – 70 % whilst morbidity ranges from 50 – 100 %. The disease causes significant financial losses to the poultry industry through bird death, stunted growth and a marked decrease in egg production. Due to its economic importance, attenuated live-vaccines have been developed by serial passage of virus either in eggs or tissue culture. These have the ability to protect birds against ILTV however they do not stop latent infection which can result in reactivation of the virus termed 'vaccinal laryngotracheitis'. The molecular biology underlying virus-host interactions for ILTV is poorly understood and there are large gaps in knowledge regarding the pathogenesis of ILTV infection.

MicroRNAs (miRNAs) are short, non-coding RNAs that post-transcriptionally regulate gene expression through targeting of specific mRNAs. Several herpesviruses have been shown to encode miRNAs that have the ability to regulate both viral and cellular gene expression which can impact virus-host interactions. Previous work in the literature has shown that ILTV encodes for 10 miRNAs with sparse data on what they may be regulating.

It was hypothesised that the virus-encoded miRNAs may have an effect upon the pathogenesis of the virus by targeting both cellular and viral mRNAs. To investigate this hypothesis initially, the biochemical technique CLASH (Cross-Linking and Sequencing of Hybrids) was attempted however a lack of suitable reagents such as physiologically relevant cell lines of chicken origin made this technically challenging and this approach was halted. Instead, a bioinformatic approach was developed and split into two avenues of research.

Firstly, it was hypothesised that miRNAs encoded by ILTV would target virally derived transcripts. As the virus genome is poorly annotated, transcripts for all 79 open reading frames (ORFs) were created manually using an arbitrary system of 1000 bp upstream of the ATG start site and 50 bp downstream of the designated PolyA tail. These were then fed into the online algorithm RNA Hybrid alongside sequences for all 10 virus-encoded miRNAs. Results from the bioinformatic predictions were then sorted and filtered using pre-defined conditions. This left a total of 227 predicted interactions. These were then filtered again leaving 28 novel targets that were screened in a reporter based system. Three of the predicted interactions showed a decrease in luciferase-reporter activity compared to the siRNA control (UL24, UL29 and UL46/48), however only the latter two showed statistically

significant decreases in activity of 15 % and 20 % respectively. Mutation of the seed sequences in both UL29 and UL46/48 targets abrogated the effects of the miRNA mimic. Further work on UL29 and its interaction with ILTV-miR-I2 looked at validating this interaction by western blotting however these results were inconclusive. Investigations into the interaction between UL46/48 and ILTV-miR-I6-5p first confirmed by RT-PCR that UL46 was targeted by ILTV-miR-I6-5p. Validation of the interaction between UL46 and ILTV-miR-I6-5p by western blotting was inconclusive. Investigations into the interplay between UL46, UL48 and the ICP4 promoter were also characterised with UL46 able to negatively modulate the effects of UL48 on ICP4 promoter activity in a reporter-based system.

Secondly, the same viral transcripts were then used in conjunction with high confidence chicken miRNAs as per MiRBase (Release 21, Jun 2014). The sorting and filtering of results mirrored that of the viral transcript study giving a final list of 103 predicted targets. From the list, three targets were picked that were all targeted by the cellular miRNA gga-miR-133a-3p and tested using the same reporter system. Two targets, one in UL20 and one in the coding region of ICP4 showed no statistical difference between the miRNA mimic and siRNA control. In contrast, one target, located in the 5'UTR of ICP4 and confirmed by RT-PCR to be within the expressed mRNA transcript was found to cause a 55 % reduction in luciferase activity. This effect was then abrogated upon mutation of the miRNA seed sequence. Further investigations found that this miRNA can cause an apparent reduction in virus titer and a statistically significant decrease in plaque size morphology when virus is harvested from cells transfected with the miRNA mimic and used to infect naïve cells. Moreover, a combination RT-qPCR and sequencing was used to confirm the sequence of gga-miR-133a-3p in several tissues of the chicken including the Dorsal Root Ganglia (DRG) and Harderian gland (HG). These are of importance to ILTV biology as the DRG is a site of latent infection and the HG is a secondary lymphoid organ (SLO) in the bird which monitors the upper respiratory tract, the site of lytic replication/clinical symptoms.

Finally, CRISPR-Cas9 genome editing was used to delete a cluster of five miRNAs from the viral genome. Guide RNAs (sgRNAs) were designed to target the miRNA cluster and shown to efficiently direct cleavage of target DNA in an *in vitro* system. Following transfection/infection of cells, virus was harvested and subsequent sequencing showed that this approach was successful in creating a recombinant ILTV. This was detectable after passage of the virus through naïve cells although a pure population of recombinant virus was not obtained due to a lack of time.

Lay Summary

Infectious Laryngotracheitis (ILT) is a deadly disease of chickens caused by a virus called Infectious Laryngotracheitis Virus (ILTV). The disease is financially important all over the world due to how much chicken is eaten. If birds catch this virus, they have symptoms such as shortness of breath and produce a thick, sometimes bloodied mucus that they cough up. Birds also do not grow properly and egg laying birds do not lay as many eggs as healthy animals would. Many birds recover from the virus but it can establish a life-long infection in the neurons of animals. Because of this, vaccines have been developed to protect animals from developing the symptoms of the virus but they currently do not stop the virus setting up a life-long infection. There is still not much known about how the virus interacts with chickens when they are infected.

All life depends upon the making of new proteins from the DNA blueprint found within cells of the body. The middle step in the process, known as messenger RNA (mRNA) carries the instructions from DNA to make new proteins. Small molecules known as microRNAs (miRNAs) affect how much protein is produced by targeting these mRNAs and binding to them. Viruses can also make miRNAs and they have the ability to target both virus mRNAs and cellular mRNAs. To date, ILTV is known to make 10 miRNAs but very little is known about what they do. The hypothesis is that these virus-encoded miRNAs, as well as cellular miRNAs can alter levels of specific virus mRNAs which will affect how the virus interacts with its host.

The first aim was to predict virus mRNA targets of the virus-encoded miRNAs using computer software. These predictions were analysed and initially 227 possible targets were identified. Further analysis gave a final list of 28 targets that were tested in the laboratory. Three of these targets showed a change in virus mRNA levels but further work to validate these interactions was inconclusive.

The second aim was to predict virus mRNA targets of chicken miRNAs. These were also filtered and gave a total of 103 predicted targets. Three of the targets of the chicken miRNA miR-133a were tested in the lab and one was found to affect an essential virus mRNA called ICP4. Further work showed that this miR-133a is found in chicken tissues that are important to the virus lifecycle in the chicken.

Finally, genome editing was used to make a mutant virus that lacked 5 of the virus-encoded miRNAs. To do this, guide RNAs (sgRNAs) were designed that instruct a protein called Cas9 to cut at specific points in virus DNA. Experiments showed that these sgRNAs were successful in cutting the virus DNA at the specific points and a mutant virus was made.

Table of Contents

List of Figures

List of Tables

Abbreviations

3’ – 3 prime

5’ – 5 prime

~ - approximately

$ - US dollars

% - percent

* - asterisk

°C – degrees Celsius

α - alpha

β – beta

γ - gamma

bp – base pair

Ago2 – Argonaute 2

BAC – Bacterial artificial chromosome

BFV – Bovine foamy virus

BLV – Bovine leukaemia virus

ChHV – Chimpanzee herpesvirus

CHX – Cycloheximide

CKCs – Chicken kidney cells

CLASH – Crosslinking and sequencing of hybrids

CLEC213s – Chicken lung epithelial cells

CRISPR – Clustered regularly interspaced palindromic repeats

CyP – Cyclophilin

DEFRA – Department for environment, food and rural affairs

DF1s – Immortalised chicken embryo fibroblasts

DGCR8 – DiGeorge syndrome chromosomal region 8

DPI – Days post infection

DRG – Dorsal root ganglia

DSB – Double-stranded breaks

dsDNA – double-stranded DNA

dsRNA – double-stranded RNA

EBV – Epstein-Barr virus

eGFP – Enhanced green fluorescent protein

FPV – Fowl pox virus

gDNA – Genomic DNA

HBV – Hepatitis B virus

HCF-1 – Host cell factor 1

HCMV – Human cytomegalovirus

HCPG – Heperan sulphate proteoglycans

HCV – Hepatitis C virus

HD11s – Chicken hematopoietic cells

HDR – Homology directed repair

HHV6A – Human herpesvirus 6A

HIV – Human immunodeficiency virus

HITS-CLIP – high throughput sequencing or RNAs by crosslinking and immunoprecipitation

HPI – Hours post infection

HSV-1 – Herpes simplex virus 1

HSV-2 – Herpes simplex virus 2

HVEM – Herpesvirus entry mediator

HVS – Herpes saimiri virus

HVT – Herpesvirus of turkeys

I – infected

IE – Immediate-early

ICP – Infected cell polypeptide

IFN - Interferon

ILT – Infectious laryngotracheitis

ILTV – Infectious laryngotracheitis virus

IP – immunoprecipitation

I_R – Inverted internal repeat

ITC – Its Coming Home

JNK – c-Jun N-terminal kinase

KSHV – Kaposi's sarcoma associated virus

K/O – Knock out

L – Ladder

LANA – Latency associated nuclear antigen

LAT – Latency associated transcript

LMHs – Leghorn male hepatocellular carcinoma cells

lncRNAs – long non-coding RNAs

LTR – left terminal repeat

M – Mock

MCMV – Murine cytomegalovirus

MCF – Malignant catarrhal fever

MDV-1 – Marek's disease virus 1

MDV-2 – Marek's disease virus 2

mRNA – messenger RNA

miRNA – MicroRNA

MTOC – microtubule organising centre

NGF – nerve growth factor

NHEJ – Non-homologous end joining

nM - nanomolar

nt – nucleotide

NTC – No template control

nts – nucleotides

ORF – Open reading frame

OvHV-2 – Ovine herpesvirus 2

PAR-CLIP – Photoactivable ribonucleoside enhanced CLIP

PAZ – piwi-argonaute-zwille domain

PCR – Polymerase chain reaction

PFV – Primate foamy virus

PGCs – Primordial germ cells

PK – Protein kinase

Pre-miRNA – Precursor miRNA

Pri-miRNA – Primary miRNA

QM7s – Quail muscle cells

qPCR – Quantitative PCR

QT-35s – Quail fibroblast cell line

RLC – RISC loading complex

RISC – RNA induced silencing complex

R_L – Repeat long

R_S – Repeat short

RTR – right terminal repeat

-RT – minus reverse transcriptase

SILAC – Stable isotope labelling by amino acids in cell culture

ssDNA – single-stranded DNA

sgRNA – Guide RNA

ssRNA – single-stranded RNA

SuHV-1 – Suid herpesvirus 1

TALENs – Transcription activator like effector nucleases

TGN – trans-Golgi-network

TK – Thymidine kinase

tracRNA – transactivating RNA

T_R – Terminal repeat

µg – Microgram

U_L – Unique long

µl – Microliter

µM - Micromolar

U_S – Unique short

UTR – Untranslated region

UV – ultra violated

vDNA – Viral DNA

vmiRNAs – Virus encoded miRNAs

V/O – Vector only

VZV – Varicella zoster virus

ZFN – Zinc finger nucleases

Chapter 1: Introduction

1.1 – Herpesviruses

1.2 – Herpesvirus Classification

1.3 – Herpesvirus Lifecycle

1.4 – Infectious Laryngotracheitis

1.5 – Infectious Laryngotracheitis Virus

1.6 – MicroRNAs

1.7 – Virus-encoded miRNAs

1.8 – Genome Editing

1.9 - Aims

1.1 – Herpesviruses

Herpesviruses are large, double-stranded DNA (dsDNA) viruses that are found throughout the animal kingdom. Most animal species are thought to be infected with at least one herpesvirus and it is generally considered that most vertebrate species are infected with several herpesviruses (Pellet and Roizmann, 2007, Fenner et al., 2011). There are now over 130 characterised viruses contained within the herpesvirus family with many more likely to be identified in species that are generally not as well studied (Brown and Newcomb, 2011). Herpesviruses are considered to have co-evolved with their main, natural host over several millennia. In the case of herpes simplex viruses this can be traced back to the common ancestor of old and new world apes around 44.2 million years ago (Wertheim et al., 2014). At the genomic level, this can be seen in the divergence but conserved functionality of open reading frames (ORFs) found across different species and families of herpesviruses (Fossum et al., 2009)

Herpesviruses cause a diverse range of disease manifestations from an asymptomatic infection through to cancer and even death. The latter are typically associated with immunocompromised hosts, neonates and non-natural hosts. Some members of the family have also been linked with life limiting illnesses such as Alzheimer's disease however these findings were and still are controversial (Readhead et al., 2018).

There are four common biological properties that all herpesviruses share. Firstly, all encode a number of enzymes that are involved with DNA synthesis, nucleic acid synthesis and metabolism. Secondly, viral genome replication and capsid assembly occurs in the host nucleus. Thirdly, the production of infectious progeny leads to the destruction of the infected host cell and finally, herpesviruses have the ability to set up a lifelong, latent infection in their natural host. In this latent state, the viral genome exists as an episome with little to no viral gene expression occurring. Latent virus is also able to reactivate and allow for the production of infectious progeny (Roizmann et al., 1992).

1.1.1 – Herpesvirus Structure

Structurally, herpesvirus virions are extremely similar. At the core of the virion is a linear, non-segmented dsDNA genome that is packed into an icosahedral shaped capsid of T = 16 symmetry. The size of the capsid is estimated to be between 100 nm and 110 nm in diameter. Surrounding the capsid is a loosely arranged protein layer called the tegument. It is now well established that tegument proteins play important roles in initiating lytic gene expression, regulating viral gene expression and virus assembly (Owen et al., 2015). The tegument links the capsid to the envelope that is a lipid bilayer derived from the infected cell.

It is studded with a number of viral glycoproteins that play a number of important roles, particularly in virus entry and attachment (reviewed in section 1.3.1). The overall size of the virion can vary between 120 nm to 300 nm due to changes in the thickness of the tegument (Pellet and Roizmann, 2007). A typical herpesvirus virion is shown alongside a diagrammatic representation of a virion (Figure – 1.1).

1.1.2 – Herpesvirus Genomes

Herpesvirus genomes are split into 6 categories, labelled A-F depending upon the presence, number and the genomic location, of repeat regions within the genome (Roizmann et al., 1992). The six categories are shown in Figure 1.2. 'A' viruses have a long region that is flanked by repeat regions that are termed left terminal repeat (LTR) and right terminal repeat (RTR). Category B viruses have a long region that is flanked by repeat regions at both ends that vary in number. Likewise, category C viruses have repeat regions at the ends of the genome however these repeat regions are often smaller than that of category B viruses. Within the category C viruses, there are also unrelated repeat regions over 100 nt in length that subdivide the genome. Category D viral genomes are split into two unique regions that are the Unique Long (U_L) and Unique Short (U_S) regions. These regions are separated by an inverted repeat region (I_R) that is also found as a terminal repeat (T_R) region at the end of the U_S genome segment. Likewise, category E viruses also have U_L and U_S genomic segments however they are both flanked by their own inverted repeat regions (termed R_L and R_S). Finally the genomes of category F viruses have been found to contain no repeat regions (Roizmann et al., 1992).

The actual length of herpesvirus genomes vary from around 125 kb up to 240 kb with a GC content ranging from 32 to 75 % dependent upon the virus species (Davidson, 2007). The minimum number of ORFs encoded by the viruses is around 70 with the upper limit not well defined (McGeoch et al., 2006). The genome of Human Cytomegalovirus (HCMV) is thought to encode around 169 ORFs though some studies have suggested that there could be an additional 604 ORFs (Van Damme and Van Loock, 2014). Between the three subfamilies, there are a collection of genes that are defined as 'core genes' and in total there are 43 genes that are shared between the *Herpesviridae* family (Davidson, 2007). These genes tend to encode for proteins that are necessary for DNA replication, and structural elements of the virion such as the capsid and tegument proteins. Subfamilies (alpha, beta and gamma) may also have conserved genes between them that are unique to that subfamily. In addition, individual genes may be restricted to a genera or individual species. As well as having ORFs that are protein coding, many herpesviruses genomes also contain non-coding

RNAs. These include long, non-coding RNAs (lncRNAs), specific examples include the four major HCMV lncRNAs named RNA2.7, RNA1.2, RNA4.9, and RNA5.0 respectively (Hancock, 2017). Moreover, a number of herpesviruses have also been shown to encode for MicroRNAs (miRNAs) which are discussed in detail in section 1.7 (Pfeffer et al., 2004a, Kincaid and Sullivan, 2012).

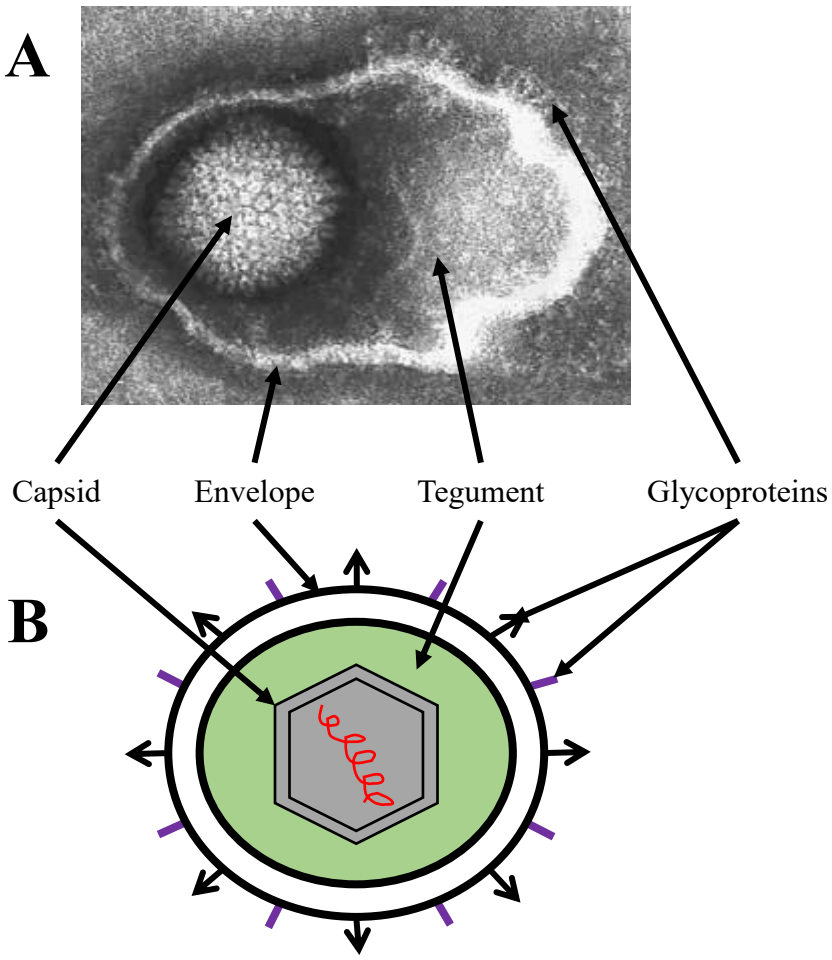


Figure – 1.1. A Typical Herpesvirus Virion
A – An electron micrograph of a Herpes Simplex 1 virion. Image taken from http://web.stanford.edu/group/virus/herpes/2000/herpes2000.html (accessed: 15/11/2017)
B – A diagrammatic representation of a Herpes virion showing major structural features: Capsid, Envelope, Tegument and Glycoproteins.

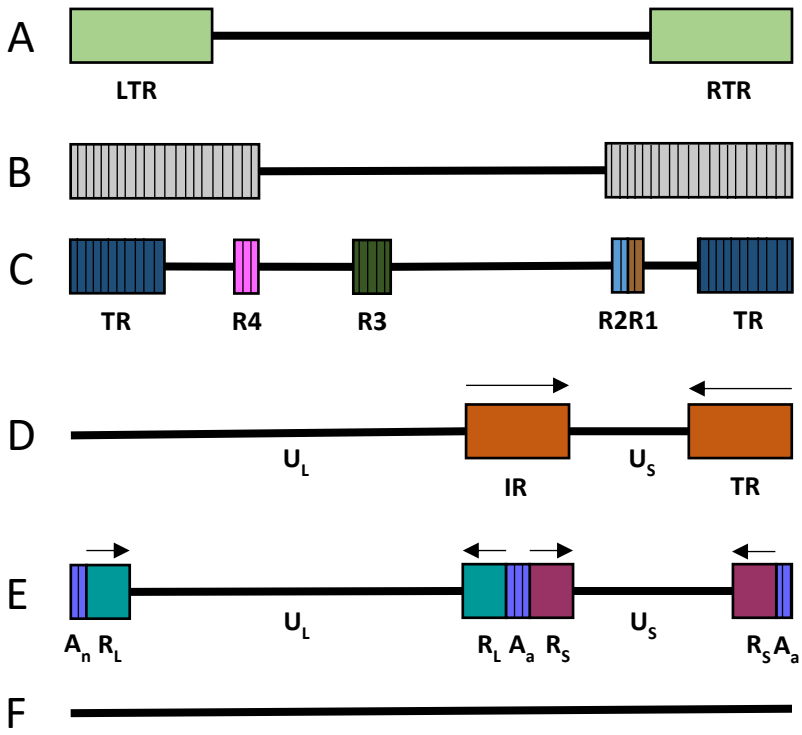


Figure – 1.2. Diagrammatic representation of the genomes found in the family *Herpesviridae*
Schematic drawings of the six different types of genomes found within the *Herpesviridae* family. Solid black horizontal lines represent genomes and rectangles depict repeat regions. Arrows indicate the relative orientation of the repeat regions. Genomes of group A (e.g. *Ictalurid herpesvirus 1*) viruses have a large terminal repeat region at both the left and right end of the genome. Group B viruses (e.g. *Saimirine herpesvirus 2*) have terminal sequences repeated numerous times at both ends of the genome. Group C viruses (e.g. *Human gammaherpesvirus 4*) harbour both terminal repeats and also 4 unrelated repeat regions that subdivide the genome. The genomes of group D (e.g. *Gallid alphaherpesvirus 1*) viruses are divided into a long and short segment by an inverted repeat of the terminal region of the short (U_S) region. Group E viruses (e.g. *Human alphaherpesvirus 1*) are similar to Group D viruses in terms of genome segmentation and inversion of the repeat regions flanking both the long and short regions. No terminal or inverted regions have been described for group F viruses (e.g. *Tupaiid betaherpesvirus 1*).
Image adapted from (Roizmann et al., 1992)

1.2 – Herpesvirus Classification

The *Herpesvirales* order is split into three distinct families; *Alloherpesviridae*, *Herpesviridae* and *Malacoherpesviridae*. Within the *Alloherpesviridae* family there are 12 species split across 4 genera. Viruses found in this family typically infect fish and amphibians (Waltzek et al., 2009). Likewise, the 2 species split into 2 separate genera in the *Malacoherpesviridae* family infect molluscs such as oysters (Davison et al., 2009). Finally, the *Herpesviridae* family is the largest and split into three subfamilies; *Alphaherpesvirinae*, *Betaherpesvirinae* and *Gammaherpesvirinae*. In total, 89 species are found across the 3 subfamilies split into a total of 12 genera (King et al., 2018). Classification of herpesviruses was originally based upon rudimental conditions including host range, cell tropism and site of latency however this was changed in favour of classification based upon genome arrangement and sequence homology (Roizmann et al., 1992). This change in classification conditions lead to the movement of some viruses from one subfamily to another.

1.2.1 – Alphaherpesviruses

The *Alphaherpesvirinae* subfamily is the largest of the 3 subfamilies containing 38 virus species across 6 genera. The *Iltovirus* genus contains 2 species and includes *Gallid alphaherpesvirus 1*, commonly referred to as Infectious Laryngotracheitis Virus (ILTV), this is discussed in detail in sections 1.4 and 1.5. The *Mardivirus* genus also contains species that infect birds with *Gallid alphaherpesvirus 2*, commonly known as Marek's disease virus 1 (MDV1) the type species of the genus. Only 1 species is found in the *Scutairus* genus and that is *Chelonid alphaherpesvirus 5*. There are 12 species in the *Simplexvirus* genus with *Human alphaherpesvirus 1*, also referred Herpes Simplex 1 (HSV-1) the type species for the genus. Finally there are 17 species in the *Varicellovirus* genus which contains *Human alphaherpesvirus 3*, commonly known as Varicella zoster virus (VZV) whilst there is one virus species in the unassigned genus which is *Chelonid alphaherpesvirus 6* (King et al., 2018).

The first two human herpesviruses identified were HSV-1 and HSV-2. As evidenced by their naming they are highly similar in terms of pathogenesis and clinical symptoms however phylogenetic analysis suggests HSV-2 jumped species and is more closely related to Chimpanzee Herpesvirus (ChHV) than HSV-1 (Wertheim et al., 2014, Pimenoff et al., 2018). Infection by HSV-1 and HSV-2 typically occurs at the oral and genital regions and initial infection is typically asymptomatic. Latency is established in the ganglia which innervates the initial site of infection. Outcome of infection varies from small localised lesions to invasion of the brain resulting in encephalitis and in some cases, even death.

Reactivation of HSV-1 and HSV-2 from latency causes disease. Symptoms tend to include blisters on epithelial surfaces such as the lips which develop into ulcerated lesions that pustulate (commonly referred to as cold sores). In addition, clinical manifestations include conjunctivitis, eczema and as previously mentioned, encephalitis. Congenital infection is common and could be considered vertical transmission and there have been reports of this in the literature (Conrady et al., 2010).

Aujeszky's disease, commonly called pseudorabies, is caused by the virus *Suid alphaherpesvirus 1* (SuHV-1) of the *Varicellovirus* genus. It is universally considered to be one of the most important swine pathogens and it is highly contagious possessing the ability to infect several other species such as cattle, cats, dogs and sheep. In pigs, SuHV-1 causes a variety of clinical symptoms including coughing, sneezing, fever and depression with more severe signs such as abortions, seizures and death also seen. Mortality of animals is near 100 % in piglets under 4 weeks of age yet is as low as 10 % in older piglets. Infection can cause pregnant sows to abort litters or can cause stillbirths and mummifications. Non-natural hosts that can be infected also show a large range of clinical signs and almost always die from infection (Fenner et al., 2011).

Marek's disease virus 1 (MDV1) is caused by *Gallid alphaherpesvirus 2* and is a deadly disease of poultry. It is highly virulent and can cause widespread death in a chicken shed. Symptoms include paralysis of the legs, neck and wings, weight loss, depression and lumps on the skin. The latter of these is caused by numerous T cell lymphomas that are induced by the virus infecting lymphocytes (Swayne et al., 2013). Due to the oncogenic potential and the preferential infection of lymphocytes by MDV1, it was originally classed as a gammaherpesvirus however the genome organisation is more similar to the alphaherpesvirus subfamily and was reclassified (Roizmann et al., 1992). Related viruses, such as Marek's disease virus 2 (MDV2) or *Gallid alphaherpesvirus 3* is clinically benign and has been used extensively as a vaccine against MDV1.

1.2.2 – Betaherpesviruses

There are 5 genera in the *Betaherpesvirinae* subfamily totalling 18 species, making it the smallest of the subfamilies. The *Cytomegalovirus* genus is the largest containing 8 species including *Human betaherpesvirus 5*, also known as Human cytomegalovirus (HCMV). Similarly, the *Muromegalovirus* genus contains the murine equivalent of HCMV, *Murid betaherpesvirus 1*, commonly called Murine cytomegalovirus (MCMV). Within the *Proboscivirus* genus is 1 species, *Elephantid betaherpesvirus 1*. 3 virus species are contained in the *Roseolovirus* genus including *Human betaherpesvirus 6A* (HHV6A).

Finally, there are 3 species that are classed as unassigned such as *Suid betaherpesvirus 2* (King et al., 2018).

HCMV is ubiquitous throughout the human population with estimates on infection levels somewhere between 60 – 70 % for industrialised nations and nearer 100 % for developing countries. *De novo* infection is typically asymptomatic however a very small proportion of individuals can develop infectious mononucleosis (commonly referred to as glandular fever) as a result of HCMV infection (Pellet and Roizmann, 2007). HCMV complications occur during pregnancy and it is the leading cause of congenital birth defects. It is estimated that 1 in 100-500 babies born worldwide have congenital HCMV, 1 in 300 babies present with congenital symptoms and 1 in 7000 will die as a result (Butler, 2016). Babies born with congenital HCMV that survive have a number of symptoms including liver, lung or spleen damage, mental retardation as well as hearing and vision loss. In addition to congenital birth defects, HCMV is a problem in the immunocompromised individual. This group of people include recipients of solid organ transplants and individuals infected with human immunodeficiency virus (HIV). Associated diseases in the immunocompromised include increased risk of death, HCMV associated hepatitis, severe colitis and pneumonia (Sager et al., 2015, Navarro, 2016).

1.2.3 – Gammaherpesviruses

The final subfamily found within the *Herpesviridae* family is *Gammaherpesvirinae*. It is the second largest subfamily containing 33 viruses across 5 genera. Contained within the *Lymphocryptovirus* genus is *Human gammaherpesvirus 4*, more commonly called Epstein-Barr virus (EBV). The *Macavirus* genus contains viruses such as *Ovine gammaherpesvirus 2* (OvHV-2) whilst the *Percovirus* genus has 3 species including *Equid gammaherpesvirus 2*. There are a total of 8 species in the *Rhadinovirus* genus which includes *Saimiriine gammaherpesvirus 2* (also known as Herpesvirus Saimiri, HVS) and *Human gammaherpesvirus 8* which is also referred to as Kaposi's sarcoma associated herpesvirus (KSHV). 3 viruses are currently unassigned and they include *Equid gammaherpesvirus 7* (King et al., 2018).

Epstein-Barr virus (EBV) is one of two human Gammaherpesviruses that has oncogenic potential. It is highly prevalent in the human population and similar to HCMV, can cause infectious mononucleosis (glandular fever) and is responsible for around 90 % of cases reported. Infection with EBV is commonly asymptomatic if it happens during childhood but primary infection in adolescents can lead to glandular fever. EBV is heavily associated with several cancers such as Hodgkin's lymphoma, Burkitt's lymphoma, and gastric cancers

(Maeda et al., 2009). This is because EBV latency occurs in B lymphocytes. These diseases and cancers are linked more with the immunocompromised individual but can also happen in the immunocompetent. As well as cancers, EBV is linked with autoimmune diseases too including systemic lupus erythematosus (Lupus) and rheumatoid arthritis (Toussirot and Roudier, 2008).

KSHV (HHV-8) is the other human gammaherpesvirus with oncogenic potential. Its prevalence varies between continents with approximately 10 % of people in Mediterranean countries infected compared to 40 % in sub-Saharan countries (Fu et al., 2009). Malignancies associated with KSHV include Kaposi's sarcoma, primary effusion lymphoma and Castleman's disease (Goncalves et al., 2017). Similar to EBV, immunocompromised patients are at a higher risk of developing these malignancies compared to a healthy individual. Moreover, AIDS patients in Sub-Saharan Africa are at the highest risk of having a Kaposi's sarcoma diagnosis (Boshoff and Weiss, 2002). Primary infection of KSHV is thought to occur through the oral/salivary routes with latency establishment in B lymphocytes similar to other Gammaherpesviruses.

Both *Alcelaphine gammaherpesvirus 1* and *Ovine gammaherpesvirus 2* are causative agents of the fatal lymphoproliferative disease malignant catarrhal fever (MCF) in their non-natural host. In their natural reservoir host, wildebeest and sheep respectively, no disease is apparent (Russell et al., 2009). Animals susceptible to MCF include cattle, deer, buffalo and other ungulates. Clinically the disease manifests with symptoms including fever, depression, opacity of the cornea and a thick mucopurulent discharge from the nose. The outcome of infection in these non-natural hosts is typically death but some animals have been shown to recover from the infection (O'Toole and Li, 2014). Infected animals are dead-end hosts with no known transmission between cattle and other MCF susceptible animals.

1.3 – Herpesvirus Lifecycle

The lifecycle of herpesviruses can be split into two distinctly separate stages. Lytic infection with herpesviruses is associated with the temporal expression of the majority of virus-encoded ORFs and subsequent release of infectious, daughter progeny. In stark contrast, the latent infection of herpesviruses is highly regulated with tight controls on gene expression with the viral genome existing in the nucleus of infected cells as an episome. The transition from latency to lytic replication is known as reactivation and it is generally associated with a stimuli that results in cellular stress. Upon reactivation, the full temporal cascade of gene

expression is seen along with any associated clinical symptoms. The end result is the assembly and egress of infectious virus particles.

1.3.1 – Cellular Attachment and Entry

The process of herpesvirus cellular attachment and entry are distinctly separate complex mechanisms that require both cellular and viral factors. As the lifecycle can be split into two stages also (lytic and latent replication) in addition to the large breadth of cellular tropism across the family, it is not surprising to learn that this process is also complex giving the viruses the ability to enter this large variety of cell types. First and foremost, the herpesvirus genome encodes for a number of glycoproteins that allow for attachment to a cell utilising a number of different receptors found upon the cell surface. There are 10 glycoproteins encoded for by HSV-1 however this number can vary depending on the virus species however it is widely reported that only 4 glycoproteins, gB, gD, gH and gL, are essential for HSV-1 entry into host cells (Agelidis and Shukla, 2015).

To begin the process of attachment, the glycoproteins gB and/or gC interact with heperan sulphate proteoglycans (HSPG) which are found on the cell surface. The latter is not essential as shown by recombinant viruses lacking gC but there is an overall reduction in binding without gC (Herold et al., 1994, Shukla and Spear, 2001). Following attachment, fusion occurs. This involves several glycoproteins including gB, gD, gH and gL. As well as viral factors, cellular factors are also necessary to facilitate fusion and these are nectin-1, herpes virus entry mediator (HVEM) or 3-O-sulfated HS (3-OS HS) (Shukla and Spear, 2001). When gD binds to one of these cellular receptors it undergoes a conformational change recruiting gB, gH and gL to form a fusion complex causing the merging of the lipid bilayers. Once this has happed, the viral capsid along with the tegument proteins are released into the cytoplasm of the cell (Agelidis and Shukla, 2015).

Following entry of the viral capsid and the tegument into the cytoplasm, directed transport to the nucleus is required due to the fact that molecules larger than 500 kDa have restricted diffusion movement in the cytoplasm, therefore a transport system is essential. This is especially true for cells like neurons which have very long axons linking the site of infection to the cell nucleus meaning virions have to travel long distances (Enquist et al., 1998, Campadelli-Fiume and Menotti, 2007). Microtubules found within the cell act as the "molecular motorways" on which the capsid travels to the nucleus from the cell periphery. They are polarised structures with the positive end located towards the cell membrane and the negative end found near the nucleus at the microtubule organizing centre (MTOC) (Dohner and Sodeik, 2005). Once the viral capsids coated with the inner tegument are in the

proximity of the nucleus, filaments protruding from the nuclear pores are used for binding. This binding is thought to enable the release of the viral DNA and allow for translocation into the nucleus (Campadelli-Fiume and Menotti, 2007).

1.3.2 – Lytic Replication

During lytic replication, gene expression occurs in a highly controlled temporal manner. The initial stage of this begins with immediate-early genes (also commonly referred to as alpha or α genes). The second stage in the cascade are early genes (also known as beta or β genes), these are typically enzymes and the viral replication machinery. Once these early genes have been transcribed, viral DNA (vDNA) replication occurs. By definition, anything transcribed following viral DNA replication is referred to as a late gene (gamma or γ genes). This tends to include structural genes such as the capsid. A simplified diagrammatic representation of the lytic (and latent) lifecycle is shown in Figure 1.3.

Translocation of viral DNA is the beginning of the lytic replication cycle. In the case of HSV-1 replication, there are 6 immediate-early genes (ICP4, ICP0, ICP22, ICP27, US1.5, and ICP47) that do not require *de novo* protein synthesis for their expression (Pellet and Roizmann, 2007). Instead, all have a response element termed the 'TAATGARAT' sequence upstream of their respective coding regions. A cellular protein, OCT-1 binds to this sequence and functions as a transcription factor. This however is not enough to induce IE gene synthesis. To initiate expression, the viral tegument protein VP16 (UL48 gene) binds to Host Cell Factor 1 (HCF-1) and this complex is then carried into the nucleus of the infected cell (LaBoissière and O'Hare, 2000). The VP16/HCF-1 complex then binds to OCT-1 forming a complex that can initiate transcription. This binding causes high levels of IE gene transcription kick starting the temporal cascade of gene expression (Thomas et al., 1998).

Once the temporal cascade of gene expression has begun, it is irreversible unless inhibited artificially. Immediate-early gene expression, in particular that of ICP4 is absolutely required for progression beyond this stage, other IE genes have been shown to be dispensable for infection however an extremely high multiplicity of infection (MOI) is necessary. One such example of an IE gene that can be removed is ICP0. Studies have shown that is not necessary for replication (with a high MOI) however it is multifunctional and one such function is to decrease the total amount of histone H3 associated with viral promoters suggesting it helps early gene expression by counteracting epigenetic silencing of the genes (Leib et al., 1989, Cliffe and Knipe, 2008). Early genes do not contain a specific response element akin to the TAATGARAT sequence found within the IE-genes but there is

evidence suggesting some have binding sites for transcription factors such as SP1 (Pellet and Roizmann, 2007). Proteins encoded for by the early genes include the viral DNA replication machinery thus preparing the infected cell for vDNA replication.

The full mechanisms of vDNA replication has not been fully elucidated however it is widely accepted that UL9 which encodes for the origin binding protein initiates replication by binding to an origin of replication. HSV-1 contains 3 origins of replication, two copies of OriS in the IRS/TRS and one copy in the U_L region (Summers and Leib, 2002). Replication follows a rolling circle model which produces long head to tail concatamers of vDNA (Boehmer and Lehman, 1997).

In the first instance, replication is thought to initiate at one of the origin of replication sites with both UL29 and UL9 been shown to be present in a complex surround an OriS site which is negatively supercoiled (Makhov et al., 2003). This complex is thought to facilitate unwinding of the DNA as well help form the pre-replicative and mature replication sites (Darwish et al., 2016). These replication sites are punctate within the nucleus and take up significant space once vDNA synthesis is underway. The viral proteins then undergo a hypothesised conformational changes within the UL29/UL9 complex to form the DNA hairpin and breaking of the dsDNA to form ssDNA whereby UL29 acts to stop reannealing of the two complementary strands (Weller and Coen, 2012). Folowing this, the helicase/primase (H/P) complex consisting of UL5, UL8 and UL52 is recruited for several functions. Firstly, it unwinds duplex DNA and subsequently synthesise short RNA primers to allow for the initiation of DNA replication. In an *in vitro* system, Chen et al. (2011), showed that a minimum of 6 bp of ssDNA is necessary for the H/P complex to unwind duplex DNA. The last of the essential viral proteins recruited to the replication fork are the proteins that make up the viral polymerase; UL30 and UL42. For the recruitment of polymerase to the replication site, an active form of the primase must be present. Several mechanisms for this have been suggested including a conformational change in the H/P complex, changes in the RNA primer and also the possibility of UL29 interacting with the polymerase directly (Weller and Coen, 2012). In total, 7 viral proteins along with as yet unidentified cellular factors are required for vDNA replication and they are; UL5, UL8, UL9, UL29, UL30, UL42 and UL52 (Pellet and Roizmann, 2007).

Finally, the last set of genes are transcribed and these are designated as late genes. They on the whole encode for genes necessary for virion assembly and egress from the cell. The first stages of capsid assembly happen in the cytoplasm though the final stages of capsid assembly including the insertion of vDNA happen in the nucleus of the cell (Pellet and

Roizmann, 2007). Once capsids contain vDNA, the first stage of egress occurs. This happens through the inner nuclear membrane with capsids budding with the assistance of UL31 and UL34 proteins into the perinuclear space (Mettenleiter, 2002). Subsequent steps of virion maturation have been hotly disputed in the literature with differing models proposed. The model with the most data, and generally the most accepted, is the envelopment–de-envelopment–re-envelopment model. To move out of the perinuclear space, de-envelopment occurs by the capsid envelope fusing with the outer nuclear membrane facilitating the release of the capsid and tegument proteins into the cytoplasm. The tegument of the herpesvirus capsid is typically described as a loose structure however there is evidence suggesting at least the inner tegument follows the same icosahedral symmetry; most likely due to the large tegument protein interacting with major capsid protein (Zhou et al., 1999). With this scaffold, tegumentation can occur with both the inner and outer tegument occurring in the cytoplasm.

The final stage in assembly and egress is the addition of a final envelope which is studded with glycoproteins which are necessary to facilate entry into cells. This is thought to occur through the trans-golgi vesicles whereby assembled glycoproteins are in such an orientation that their cytoplasmic tails interact with tegument proteins. Interactions between both glycoprotein E (gE) and glycoprotein M (gM) and the tegument protein UL49 has been shown in SuHV-1 (Fuchs et al., 2002b). These interactions are thought to drive the final envelopment as well as unidentified mechanisms as all three of these proteins have been shown to be dispensible for replication (Mettenleiter, 2002). Release of virions from the cell is carried out by exocytosis (Pellet and Roizmann, 2007).

1.3.3 – Latency

Latency is the second stage of the herpesvirus lifecycle and it is the one by which they are able to establish lifelong infections in their host. The viral genome in latent cells exists as an episome and no viral replication is seen with viral gene expression tightly regulated. The level of gene expression varies between species and in this section, HSV-1 latency will be reviewed with mention of other species from other subfamilies.

The site of latency for Alphaherpesviruses is the sensory neurons that innervate the initial site of infection. Following lytic replication, virions track up the axon of the neuron by retrograde transport. Upon reaching the neuronal cell body, the herpesvirus genome forms a closed circle or episome. No known protein-encoding genes are transcribed during this period. In contrast however, a long viral RNA referred to as the latency associated transcript (LAT) is heavily expressed. It is encoded antisense to the IE-gene ICP0, a transactivator of

viral gene expression (discussed in section 1.3.2) (Pellet and Roizmann, 2007). The function of the LAT is to repress lytic gene expression thus blocking reactivation of the virus (Nicoll et al., 2016). No proteins are produced from the LAT however the 8.3 kb primary transcript is spliced into two major LAT introns of 1.5 and 2 kb size respectively along with a 6.3 kb minor LAT exon that is further processed into 8 mature microRNAs (Pellet and Roizmann, 2007). Numerous studies have been undertaken elucidating the role of the LAT and its associated miRNAs. Umbach et al. (2008b), showed that the LAT encoded miRNAs were able to suppress expression of the viral genes ICP0 and ICP4. As well as repressing viral gene expression the LAT can also block apoptosis. Two small RNA products from the 1.5 kb LAT intron were shown to be able to block cold shock induced apoptosis in latently infected neuroblastoma cells (Shen et al., 2009).

Moreover, it is not only the LAT that works to help maintain latency. Whilst the TAATGARAT response element can kick start lytic replication in permissive cells, in neuronal cells, the opposite effect is seen (Lillycrop and Latchman, 1992). OCT-2, the neuronal version of OCT-1 can bind to these response elements in IE-genes and repress lytic gene expression. In addition, cellular miRNAs can also target essential IE-genes thus blocking reactivation (discussed more in section 1.7.3) (Pan et al., 2014).

Numerous other mechanisms have also been reported to help maintain latency across the virus family. Latency in the gammaherpesvirus EBV is much more complex involving 3 stages defined by the viral transcripts that can be detected (Rowe et al., 1992). Recent reports investigating KSHV latency implicated that chromatin remodelling controls reactivation (Hopcraft et al., 2018). Epigenetic regulation has also been implicated in controlling latency of cytomegaloviruses found in the Betaherpesvirus subfamily (Liu et al., 2013).

1.3.4 – Reactivation

Similar to latency and other life stages of the virus, reactivation is tightly controlled. The causes of reactivation are still mysterious and wide ranging but the molecular control of reactivation has been elucidated for several virus species. The outcome of reactivation is dependent upon the virus subfamily, for example Alphaherpesvirus reactivation leads to clinical symptoms such as cold sores in the case of HSV-1 whilst reactivation of Betaherpesviruses and Gammaherpesviruses is more complex especially owing to the oncogenic nature of the latter.

Reactivation requires the virus to overcome all of the mechanisms that are in place to maintain latency. As discussed in the previous section (1.3.3), both cellular and viral factors

contribute to the maintenance of latency. The end goal of reactivation is the production of infectious daughter progeny in a similar fashion to *de novo* lytic infection however there are two major differences. Firstly, the genome exists as an episome that is epigenetically modified with chromatin and other factors. Secondly, VP16, necessary to facilitate IE-gene transcription is most likely not present however there is little evidence surrounding this (Cliffe and Wilson, 2017).

The process of reactivation is broken down into two separate stages. In the first instance, a burst of viral transcription happens following a reactivation stimulus typically defined as a 'stressor'. These stressors have been found to include heat shock, mental tension, fatigue, and exposure to UV light amongst others (Grinde, 2013). Studies observed the transcription of VP16 upon inhibition of some cell signalling pathways such as those downstream of nerve growth factor (NGF) (Wilcox and Johnson, 1987, Kim et al., 2012). Even on the production of viral transcripts, these then have to overcome the RNAi pathway including the LAT associated miRNAs that are known to target ICP4 and ICP0 suggesting there is some redundancy in this mechanism. Not all latently infected cells exposed to a stressor will undergo reactivation but rather a subset. One theory surrounding this is a threshold of lytic protein production that must be reached in phase I in order to trigger full blown reactivation (Cliffe and Wilson, 2017). If this threshold is surpassed, phase II reactivation which results in the synthesis of infectious particles occurs.

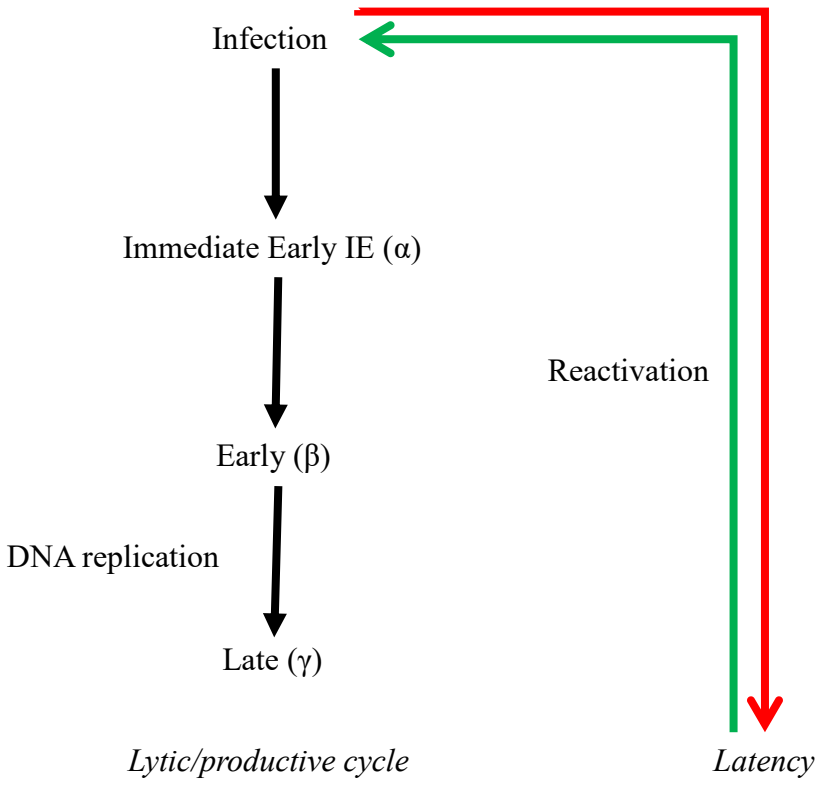


Figure – 1.3. Diagrammatic representation of the Herpesvirus gene expression cascade
Following *de novo* infection, gene expression follows a temporal pattern with immediate-early genes (IE) (also known as alpha or α genes) expressed initially. Following IE gene expression, early genes (E) (also referred to as beta or β genes) are activated. Following early gene expression, viral DNA replication typically occurs. Finally, late genes (L) (also called gamma or γ genes) are expressed prior to virion assembly and nuclear egress. Completing the lytic or productive viral cycle. Conversely, herpesviruses also have a latent cycle whereby no infectious progeny are produced. Upon certain triggers, reactivation can occur resulting in a lytic or productive cycle

1.4 – Infectious Laryngotracheitis

Infectious Laryngotracheitis (ILT) is a fatal respiratory disease of poultry. It is caused by the Alphaherpesvirus *Gallid alphaherpesvirus 1* (Infectious Laryngotracheitis Virus [ILTV]). Endemic worldwide, it can be a significant burden on poultry producers both in the broiler and laying industries. The virus and its properties are discussed in greater detail in section 1.5.

1.4.1 – Occurrence and Significance

The first reports of ILT came from Canada in 1925 with the USA reporting it a year later. By the early 1960's, over 40 countries had reported the disease (Menendez et al., 2014). Even in the earliest reports, it was described as a 'major problem' for the poultry industry with researchers also reporting of 'chronic carrier' birds that survived the initial infection and possessed the ability to pass the virus onto other birds (Brandly and Bushnell, 1934). Today, it is still a major economic problem for the global poultry industry though it is confined to certain geographic regions/areas. This tends to be where the highest population density of poultry is seen (Bagust et al., 2000).

Vaccination is commonly used to protect birds but this prevents the clinical symptoms as opposed to preventing infection. The first vaccine for ILT was developed in 1934 by inoculating birds with infected tissue in the cloaca (Brandly and Bushnell, 1934). This is generally considered as the first effective vaccine against a major avian pathogen (Swayne et al., 2013). More recently, the USDA labels licensed vaccines as '*Prevention of Laryngotracheitis'* instead of *'Prevention of infection'* due to the fact that a latent infection is still established (Koski et al., 2015). The majority of vaccines used are live-attenuated through serial passage of virus either in tissue culture or chicken embryos (Coppo et al., 2013b). However, attenuated vaccine viruses have the ability to recombine and the emergence of vaccine derived virulent strains has become more common (Lee et al., 2012). A study carried out in 2008 found that 94 % of the strains circulating in Europe over the last 35 years were related to the vaccine strain (Neff et al., 2008). Recently, there have been efforts to create vectored vaccines using Herpesvirus of Turkeys (HVT) and Fowl pox virus (FPV) expressing ILTV glycoproteins to confer protection. As well as this, recombinant vaccines are also been developed which lack the genes for some viral glycoproteins (Devlin et al., 2007, Schneiders et al., 2018).

1.4.2 – Clinical Signs

Clinically, there are two forms of ILT recognised: the more severe epizootic form and a milder enzootic form. The former is characterised by dyspnoea, gasping, coughing and

expulsion of thick, sometimes haemorrhagic, tracheal exudate. The milder form presents with marked decreases in egg production, watery eye, conjunctivitis, swelling of the infraorbital regions and nasal discharge (Bagust et al., 2000). Some of the symptoms are shown in Figure – 1.4. There have been some reports of other clinical symptoms such as esophagitis and pharyngitis in some backyard flocks but this is not typically common and has only been reported in a handful of cases (Sary et al., 2017). Symptoms present in the animal after a 6 – 12 day incubation period or after 2 – 4 days incubation in birds experimentally inoculated with the virus (Fuchs et al., 2007). After the onset of clinical signs, most chickens tend to recover in around 10 – 14 days (Bagust et al., 2000). Typical outbreaks see morbidity reach around 90 – 100 % whilst mortality ranges greatly, usually between 5 – 70 % although mortality is generally around 10 - 20 % (Devlin et al., 2011, Coppo et al., 2013b).

1.4.3 – Gross Pathology

Post mortem examinations reveal extensive pathological findings. Lesions can be observed in the conjunctiva and throughout the respiratory tract of infected animals, commonly located in the larynx and trachea of the bird (Bagust et al., 2000). Swelling and inflammation of the respiratory tract coupled with necrosis and haemorrhage can also be seen (Ou and Giambrone, 2012) (Figure – 1.5A). Thick tracheal exudate and caseous plugs also line the lumen of the trachea and can cause complete occlusion of the lumen as shown in Figure – 1.5B. Virus can also disseminate throughout the body of the bird with a number of internal organs including throat, trachea, lung, cecum, kidney, pancreas, thymus and oesophagus, testing positive for the virus via quantitative PCR analysis (qPCR) (Wang et al., 2013). Work carried out by another lab group in the same year detected ILTV in a larger number of organs with a more widespread dissemination throughout the body. Zhao et al. (2013), found ILTV in 17 different locations including the heart, liver, spleen, lung, kidney, larynx, tongue, thymus, glandular stomach, duodenum, pancreatic gland, small intestine, large intestine, cecum, cecal tonsil, bursa of Fabricius, and brain of chickens. This pattern of detection was seen across both the directly infected and contact exposed group of birds up to 28 days post infection (DPI).

At the histopathological level, intranuclear inclusion bodies are typical, though only present in the early stages of infection. Infected epithelial cells start to form large, multinucleated cells (syncytia). The loss of goblet cells and invasion of the mucosa by inflammatory cells is also seen. In later stages, the loss of the epithelial layer through cell death can expose the blood capillaries leading to rupture and haemorrhage (Bagust et al., 2000).

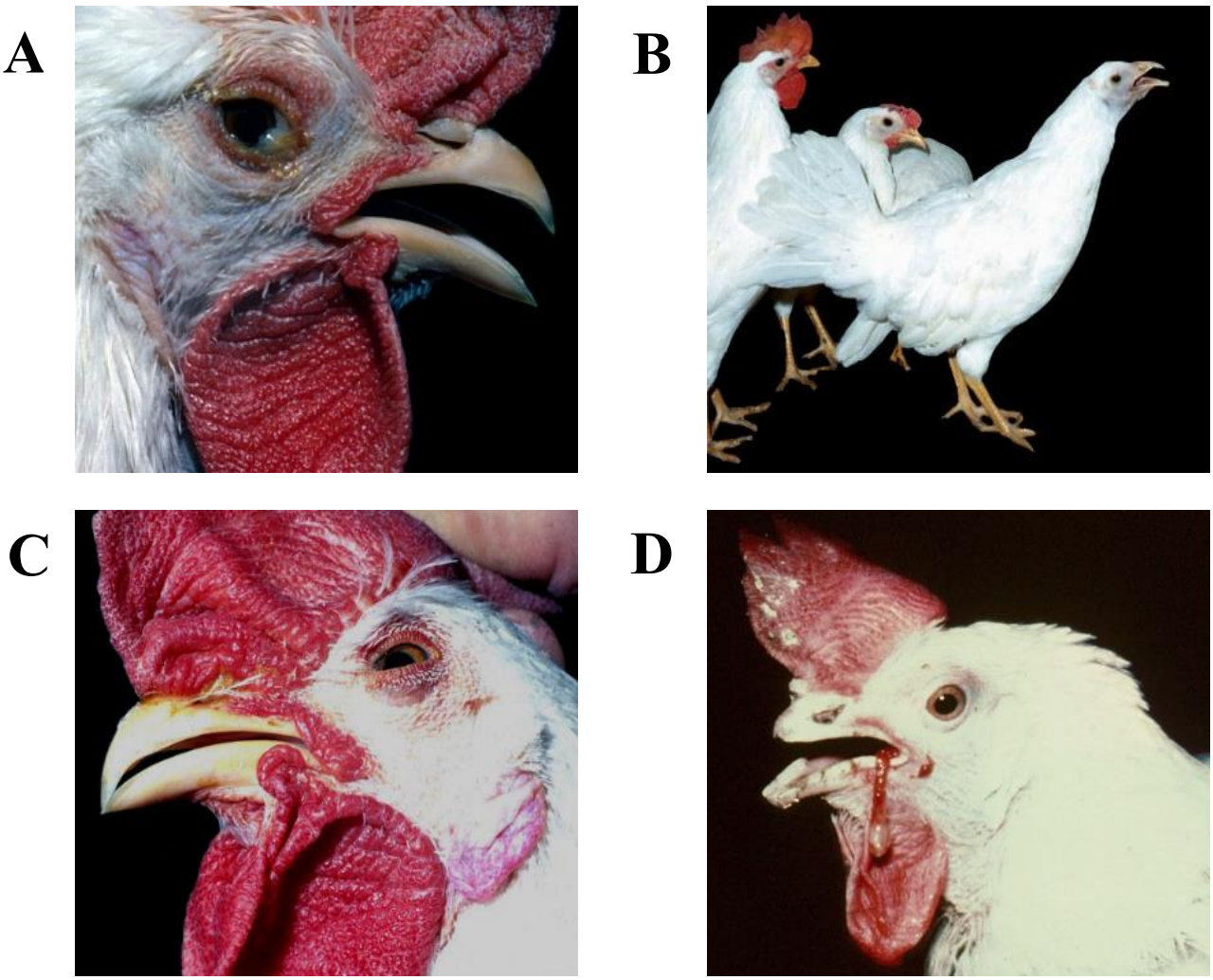


Figure – 1.4. Clinical Signs of Infectious Laryngotracheitis Virus infection
A – Animal is showing conjunctivitis and open-mouthed breathing

B – Central bird is showing open-mouthed breathing and craning of the neck

C – Animal is displaying signs of conjunctivitis and nasal discharge

D – The animal has exuded bloodied, tracheal exudate coming from the mouth

Images adapted from Buckles (2009).

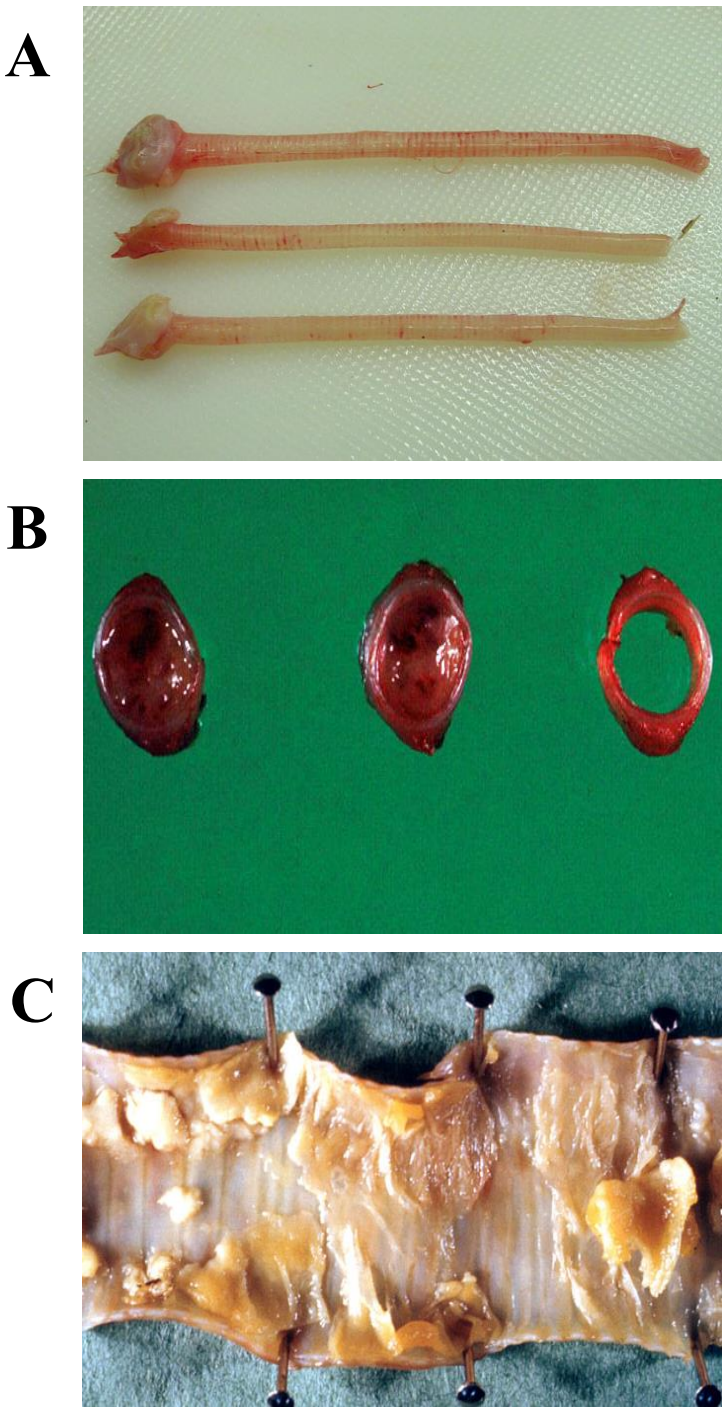


Figure – 1.5. Gross Pathology of Chickens infected with ILTV
A – Image depicts the trachea of infected animals with the bottom showing the fewest lesions and haemorrhage and the top displaying the more severe haemorrhage and lesions
B – Cross sections of trachea from birds infected with ILTV showing complete occlusion of the lumen with tracheal exudate. On the right a healthy trachea is shown for comparison
C – Tracheal exudate lining the lumen of the trachea from an infected animal. In some cases this can also be haemorrhagic

Images adapted from Buckles (2009).

1.4.4 – Poultry as a worldwide food source

At a global level, just over 118 thousand tonnes of poultry meat was eaten in 2017. The United States of America accounted for around 18 thousand tonnes of this total last year (OECD et al., 2017). When broken down into amount of poultry eaten per capita, Israel eats the most with 56.9 Kg of poultry eaten per head. This is over double the amount the combined EU28 eat per person which is at 24.2 kg per head (OECD et al., 2017).

Chickens and other birds are a staple source of protein in the average British household with 91 % of consumers reporting that they ate chicken and of these, 70 % eat chicken at least once per week (Mintel, 2017). Economically, chicken in particular is valuable to the British economy with figures from 2016 showing unprocessed poultry sales equated to £2,894 million rising to an estimated £3,405 million by 2022 (Mintel, 2017). In addition, government figures show that approximately 112 million hens (including spent laying hens and surplus breeder stock) were slaughtered in the UK during the month of July 2018 (DEFRA, 2018). These figures underline the importance of chicken and other poultry as a valuable food source and how food security lapses and pathogen outbreaks can have serious financial consequences at a governmental level right down to the individual farmer.

1.5 – Infectious Laryngotracheitis Virus

Infectious Laryngotracheitis Virus (ILTV) is a member of the *Iltovirus* genus in the *Alphaherpesvirinae* subfamily. It is also sometimes referred to as *Gallid alphaherpesvirus 1*. It is the etiological agent of Infectious Laryngotracheitis (ILT).

1.5.1 – Genome Structure

The genome of ILTV is consistent with a category D genomic structure with a U_L region and a U_S region that is flanked by both an internal and terminal inverted repeat regions (I_R/T_R) (as outlined in Figure – 1.2.). The genome is ~150 kb in length and it encodes for a predicted 80 ORFs. Eight of these ORFs are unique to the *Iltovirus* genera. Sixty five of these ORFs are found within the U_L with a further 9 encoded for in the U_S. Three ORFs (ICP4, US10 and sORF4/3) are found within the internal inverted and terminal repeat regions (I_R/T_R) flanking the U_S region (Piccirillo et al., 2016). Of the predicted 80 ORFs, 63 are homologous to the HSV-1 genome (McGeoch et al., 1988). Within the U_L region there is a lack of UL16 homologue and large inversion of the genomic arrangement from UL22 to UL44 genes (Ziemann et al., 1998a). Moreover, the UL47 gene, termed sORF1 in ILTV is translocated from the U_L to the U_S region however its function is generally considered to be the same (Helferich et al., 2007). Flanking the ends of the U_L region are the eight unique genes to the *iltovirus* genera. Six are found at the left hand end of the region and are designated ORF F

followed by ORF A through E in that genomic order (displayed on Figure – 1.7) (Veits et al., 2003b, Garcia et al., 2013). Previous work stated that ORFs A-E were dispensable for replication *in vitro* but more recent work has found that deletant ORF C recombinant viruses are heavily attenuated *in vivo* compared to wild type and the deletant virus is now a possible vaccine candidate (Veits et al., 2003b, Garcia et al., 2016, Schneiders et al., 2018). At the other end of the U_L region are another two unique ORFs termed UL0 and UL-1 (Thureen and Keeler, 2006, Ziemann et al., 1998b). The former of these, UL0 has been shown to be dispensable for replication *in vitro* however recombinant viruses lacking UL0 have been observed to be attenuated *in vivo* (Veits et al., 2003a). In comparison, UL-1 has recently been shown to be essential for ILTV replication suggesting it plays an important role in the ILTV lifecycle (Nadimpalli et al., 2017).

1.5.2 – Transmission

Natural infection of birds with ILTV primarily occurs through the nasal, oral or ocular route (Bagust et al., 2000). The source of this infection can come from other infected animals, contaminated dust, litter, drinking water and fomites (Ou and Giambrone, 2012). Other indirect sources of transmission have also been reported including cats, crows and dogs (Kingsbury and Jungherr, 1958). Infected birds that are actively shedding virus can do so through a number of routes. This includes both respiratory and ocular routes as described above but also through faeces with high levels of virus detected between 2 and 7 DPI and lower but sustained presence of virus until 28 DPI (Roy et al., 2015). The virus itself is relatively hardy with a series of experiments showing that virions could survive on wooden surfaces, away from light, for up to three months. In the same experiments, it was shown virus could survive in deep litter for 20 days and that carcasses of birds that had succumbed to the virus harboured infective virions up to three weeks after death (Bagust et al., 2000). Due to the large number of potential transmission routes, biosecurity controls are considered to be paramount in helping to stem the spread of the virus alongside vaccination efforts.

1.5.3 – Replication

The site of initial replication for ILTV is widely believed to be the epithelial cells lining both the trachea in the upper respiratory tract and the conjunctival mucosal surfaces (Reddy et al., 2014). Different strains of ILTV have been shown to also have differing virulence and tropisms to the trachea or conjunctival surfaces (Kirkpatrick et al., 2006). The highest level of viral titres in experimentally infected animals is seen in the trachea and conjunctiva around 4 – 6 days post infection (DPI) (Oldoni et al., 2009). Following initial replication, it is believed that ILTV then invades the basement membrane to disseminate into the lamina

propria. This mechanism is not well studied for ILTV however it is well characterised in other Alphaherpesviruses and so is inferred (Reddy et al., 2014). Viral DNA is detected in a large number of an infected animals organs suggesting systemic spread of the virus through an unknown mechanism (Wang et al., 2013, Zhao et al., 2013). Infection of leucocytes, in particular macrophages *in vitro* has been suggested as the possible mechanism of systemic spread in the bird (Calnek et al., 1986). This is also purported as a possible mechanism of resistance and susceptibility to ILTV in different bird lines (Loudovaris et al., 1991a, Loudovaris et al., 1991b).

During the initial, acute infection phase, invasion of the neurons and latency establishment occurs. Viral DNA is detectable in the trigeminal ganglia (TG) from three to six days post onset of acute infection from both field and vaccine strains (Bagust, 1986). The tracheal ganglion has been shown to be the main site of latency however dorsal root ganglia (DRG) have also been shown to harbour ILTV DNA (*this study*) (Bagust et al., 2000). Reactivation of virus in latently infected birds has been put down to a number of different factors. It is generally reported that stress factors such as the mixing of unfamiliar birds and even the onset of lay can trigger reactivation of the virus (Coppo et al., 2012, Coppo et al., 2013a).

The kinetics of ILTV replication have previously been studied *in vitro* to determine the expression pattern of the viral transcripts (Mahmoudian et al., 2012). Many of the viral transcripts with homologs in other Alphaherpesviruses were classified in a similar pattern, for example ICP4 as a designated immediate-early gene however they were some exceptions such as US4 (Glycoprotein G [gG]), which is typically classed as a late gene however it was only partially inhibited with the addition of Cycloheximide (CHX).

1.5.4 – Pathogenesis

Pathogenesis of ILTV has been studied using a number of different techniques. The initial attempts to measure pathogenicity revolved around looking at the microscopic pathology induced in the trachea following infection (Guy et al., 1990). This was then replaced with a more comprehensive approach taking into account clinical signs, body weight gain and the presence of vDNA in the trachea. This was considered to be a more precise measure of pathogenicity compared to the trachea pathology index (Kirkpatrick et al., 2006). As well as this, *in vitro* growth characteristics have been evaluated in a number of studies however these do not link to pathogenicity (Mahmoudian et al., 2012, Oldoni et al., 2009).

The pathology and clinical symptoms associated with ILTV have been well documented in the literature and until recently, the underlying pathogenesis was not well characterised but in recent years a number of papers have looked at varying aspects of the immune response to

ILTV infection. The initial barrier to ILTV infection is the mucus that lines both the respiratory and conjunctival tracts. This innate barrier is the first hurdle for ILTV to overcome prior to infection. Within the mucus, there are a number of innate immune mediators including interferons, collectins, IgA, defensins and lactoferins though there has been no research into the effects of these on ILTV (Coppo et al., 2013a). Following infection of epithelial cells 789 genes were differentially expressed compared to uninfected chicken embryo lung cells *in vitro* with 54 of these genes linked to inflammatory response pathways (Lee et al., 2010). More recently, cytokine profiling in a number of different tissues in infected animals revealed peak pro-inflammatory and anti-inflammatory cytokine transcription 5 DPI (Vagnozzi et al., 2018). In the same study, it was shown that interferon-beta (IFN-β), is not induced in the trachea of birds suggesting ILTV mediates and blocks the type 1 interferon response.

At the initial infection sites, a small infiltration of what is presumed to be Heterophils is observed in the lamina propria. This infiltration persists and between 3 to 5 DPI, more infiltrate is seen with macrophage and other lymphocytes observed (Purcell, 1971). $CD4^+$ and $CD8^+$ cells are seen throughout the mucosa as revealed by immunohistochemistry (Devlin et al., 2010). It was also reported that the types of inflammatory cells present at this stage of infection influences the outcome of infection as well as the balance of the adaptive immune response (Devlin et al., 2010).

Virus neutralising antibodies against ILTV are detectable in the serum of infected animals around 5 – 7 DPI and continually raise until a peak at around 21 DPI. These antibodies levels then wane over time and are detectable at low levels for several months (Bagust et al., 2000). Glycoproteins expressed by the virus are considered to be the most immunogenic and can elicit both a humoral and cell mediated response (Coppo et al., 2013a). Early work showed vertical transmission of passive immunity to offspring in vaccinated birds though this was found to offer no greater protection than parents who were non-vaccinated when the animals were challenged (Hayles et al., 1976, Coppo et al., 2013a). This work suggested that cell-mediate immunity was the most important response to infection and outcome of infection compared to humoral immunity (Coppo et al., 2013a).

A great deal of work has been carried out on glycoprotein G (US4) and its role in pathogenesis. Original studies showed that gG is a broad spectrum viral chemokine binding protein (Bryant et al., 2003). The first gG deficient knockout of ILTV was made in 2006 and it has subsequently undergone a number of different studies (Devlin et al., 2006). Studies looking at the *in vitro* growth characteristics found no significant differences when compared

to wild-type virus however *in vivo* work showed extensive attenuation of clinical signs and weight changes in challenged birds (Devlin et al., 2007, Devlin et al., 2010). From this work it was deduced that gG played a significant role in ILTV pathogenesis. Recently, gG has been shown to alter the transcription of key inflammatory markers such as CXCL8 and IL-18 both *in vitro* and *in vivo* (Coppo et al., 2018).

1.6 – MicroRNAs

MicroRNAs (miRNAs) are short, non-coding RNAs around ~22 nt in length. They have been identified across the animal kingdom in a large number of species ranging from algae to humans (Berezikov, 2011). Their main function is the post-transcriptional regulation of genes by guiding the RNA-induced silencing complex (RISC) to target mRNAs halting the translation of the target through a number of means (discussed in section 1.6.3). The effects of miRNAs are seen in nearly every biological process from cell proliferation to disease development (Winter et al., 2009).

MiRNAs were first identified around 25 years ago by two groups at the same time. Researching the genomes of *Caenorhabditis elegans* (*C. elegans*), the *lin-4* gene was shown to encode for a small RNA as opposed to a protein. It is encoded in the antisense orientation to the protein coding gene *lin-14* and was shown to result in the repression of the *lin-14* protein (Lee et al., 1993, Wightman et al., 1993). A further 7 years passed until a second, small RNA was identified again in *C. elegans* named *let-7* (Reinhart et al., 2000). This ignited the field of small RNA research and the coining of the term microRNA. Since then, 38589 miRNA entries have been added to the miRNA database MiRBase (release 22, March 2018) (Griffiths-Jones et al., 2006). Within the human genome, miRNAs are predicted to account for 1 – 5 % of the total genome and it is suggested that they can regulate at least 30 % of the protein coding genes (Macfarlane and Murphy, 2010).

As mentioned above, miRNAs can play a role in disease development. This can be through the direct interaction between miRNAs and their targets but also through the levels of certain miRNAs changing. Changes in miRNAs and disease development have been shown in a wide variety of complications including cancer and cardiovascular disease (Ha, 2011). Within cancer, miRNA expression is reduced or eliminated through deletions of miRNA coding regions in some cases. This is the case with chronic lymphocytic leukemia (CLL) whereby a common deletion affects the levels of miR-15 and miR-16 and in this deletion is frequently the only abnormality seen in patients suggesting the loss of these miRNAs is the direct cause of CLL (Schetter et al., 2010). On a similar note, miR-155 over expression can drive b-cell tumours and this strategy is utilised by MDV (discussed in section 1.7)

(Costinean et al., 2006). Finally, following acute myocardial infarction (AMI) in both humans and mice, levels of several miRNAs are altered including miR-1, miR-133a and miR-133b withing the blood plasma. This finding suggests that circulating miRNA levels can be used as possible diagnostic markers for prognosis of diease in patients presenting with cardiovascular symptoms (Ha, 2011).

MicroRNAs are not the only non-coding RNAs (ncRNAs) that are found in nature. There are several other species of ncRNAs that have a variety of different functions including small interfering RNAs (siRNAs), piwi interacting RNAs (piRNAs), long non-coding RNAs (lncRNAs) and circular RNAs (circRNA). The latter two will be dicussed below.

Firstly, there are long non-coding RNAs (lncRNAs) which are defined as transcripts longer than 200 nucleotides that are not protein coding (Perkel, 2013). Around ~28,000 lncRNA transcripts have been identified in the human genome highlighting their importance in vital cellular processes (Hon et al., 2017). Functionally, lncRNAs can affect the regulation of gene transcription right through to post-transcriptional regulation. One such example of of a lncRNA affecting mRNA turnover is gadd7. In response to damaging radiation, gadd7 levels increase with the consequence of increased binding to TDP-43 at the detrement of cdk6 mRNA. The result of the change in binding affinity of TDP-43 is a reduction of cdk6 levels thus leading to cell cycle arrest (Liu et al., 2012).

Secondly, a more recent discovery is the identification of circular RNAs (circRNAs). Instead of a typical RNA which is in a linear conformation, the 5'and 3' of the RNA are joined together forming a circle. Functionally, they have many possible functions as described in the literature including acting as miRNA sponges, transporting miRNAs in the cell and also binding to RNA-binding proteins (Wilusz and Sharp, 2013, Memczak et al., 2013). One such example of a circRNA acting as a miRNA sponge is circular RNA sponge for miR-7 (ciRS-7). The circRNA has been shown to have over 60 seed sequence sites (discussed in section 1.6.3) aciting as anchor sites for miR-7 binding thus reducing the amount of 'free' miR-7 that can be used to target mRNAs (Hansen et al., 2013).

This feature of circRNAs adds another layer of gene regulation as they can modulate the levels of other ncRNAs however circRNAs are not the only ncRNA species that effectively regulate the expression of other ncRNAs. The lncRNA PTENP1 has been shown to act as a molecular decoy for miRNAs targeting the tumour suppressor protein PTEN. The 3'UTR of PTENP1 was found to bind the same miRNAs as the 3'UTR of PTEN mRNA (Wang and Chang, 2011)

1.6.1 – MicroRNA Biogenesis

MicroRNA biogenesis is a complicated process. It involves several steps that starts at the genome (A diagrammatic flow chart of miRNA biogenesis is shown in Figure – 1.6). The majority of miRNAs genes are transcribed by RNA polymerase II (RNA pol II) but in some cases, RNA polymerase III (RNA pol III) is also the starting point of transcription (Lee et al., 2002). In the case of human miRNAs, transcription factors such as p53, MYC ZEB1 and ZEB2 can positively or negatively regulate the transcription of miRNAs. Likewise, DNA methylation and histone modification can also contribute to their regulation similar to protein encoding genes (Ha and Kim, 2014). Once transcribed, these transcripts, known as primary miRNAs (pri-miRNAs) form a self-hairpin structure within which the mature miRNA sequence is embedded. The pri-miRNAs can be over 1 kb in length and possess both a 5' cap and a 3' polyA tail. They can encode for a single mature miRNA (monocistronic transcript) or several mature miRNAs (polycistronic transcript) either through a single promoter or through several promoters for each of the miRNAs in the pri-miRNA (Cai et al., 2004). Pri-miRNAs are spread throughout the genome, many are found in intergenic regions or in an antisense orientation to protein coding genes but in some cases, they are also found in introns and exons (Kim, 2005).

Following transcription, the pri-miRNA undergoes the first stage of processing. The RNase III enzyme Drosha and its co-factor DiGeorge syndrome chromosomal region 8 (DGCR8) form what is known as the microprocessor complex. This complex cleaves the stem-loop structure from the pri-miRNA to form a ~65 nt precursor miRNA (pre-miRNA). The models currently suggested in the literature propose that DGCR8 recognises the pri-miRNA at the ssRNA-dsRNA junction and directs Drosha to a precise cleavage point around 11 nt away from this single-stranded-double-stranded RNA junction (Macfarlane and Murphy, 2010). The Drosha and DCGR8 proteins are essential for normal development and function. Deficiency in Drosha is embryonically lethal whilst DCGR8 knockout mice have early arrest in development. Lack of DGCR8 in humans is associated with DiGeorge syndrome which can cause a variety of congenital problems (Ha and Kim, 2014).

Once the pre-miRNA is formed, it requires translocation to the cytoplasm for further processing. For this to happen, the pre-miRNA requires 2 important structural features. At the 5' end, a phosphate group is required whilst at the 3' end, a 2 nt overhang must be present (Zeng and Cullen, 2005). This is to allow recognition by the Exportin/Ran complex that is made up of Exportin 5 and Ran-GTP co-factors that enables the export of the pre-miRNA from the nucleus to the cytoplasm. Following translocation to the cytoplasm, the

Ran-GTP is hydrolysed to Ran-GDP subsequently causing the complex to dissemble releasing the pre-miRNA into the cytoplasm (Wang et al., 2011).

The second stage of miRNA processing revolves around the RNase III endonuclease Dicer. Dicer along with its co-factors TAR RNA-binding protein (TRBP), protein activator of PKR (PACT) and argonaute-2 protein (Ago-2) form the RISC loading complex (RLC). Several of these proteins do not play a role in the processing of the pre-miRNAs but help to stabilise the complex as well as recruit the pre-miRNA and help form the RLC (Lee et al., 2006). The pre-miRNA is recognised by several domains of the Dicer protein including the piwi-argonaute-zwille (PAZ) domain, RNase IIIa and RNase IIIb allowing it to cleave the pre-miRNA into a dsRNA duplex that is approximately ~22nt in length that still has a 5' phosphate group and a 3' 2 nt overhang that has a hydroxyl group. The latter of which is necessary for Dicer to bind to the pre-miRNA (Ha and Kim, 2014, Macfarlane and Murphy, 2010).

This dsRNA duplex then interacts with Ago-2 to eventually form the RISC that has the ability to target specific mRNAs for silencing. The PAZ domain of Ago-2 interacts with the 2 nt overhang at the 3' of the dsRNA duplex whilst the 5' interacts with the middle domain (MID) of Ago-2. These interactions help to anchor the duplex to Ago-2 (Gregory et al., 2005, Kim, 2005). The final stage is the unwinding of the dsRNA duplex by helicases. One strand, termed the guide strand or mature miRNA remains in the RISC and goes on to target mRNAs. The second strand or passenger strand is released from the complex and can follow two fates. It is either degraded or can be incorporated into a separate RISC where it also can target mRNAs and act as a mature miRNA (Ha and Kim, 2014, Kim et al., 2009). As both strands have the possibility of becoming the mature miRNA, a naming system was devised based upon which strand of the duplex the miRNA came from. This feature is seen on miRNA naming system denoted by either a 5p or 3p at the end of the miRNA name (Griffiths-Jones et al., 2006).

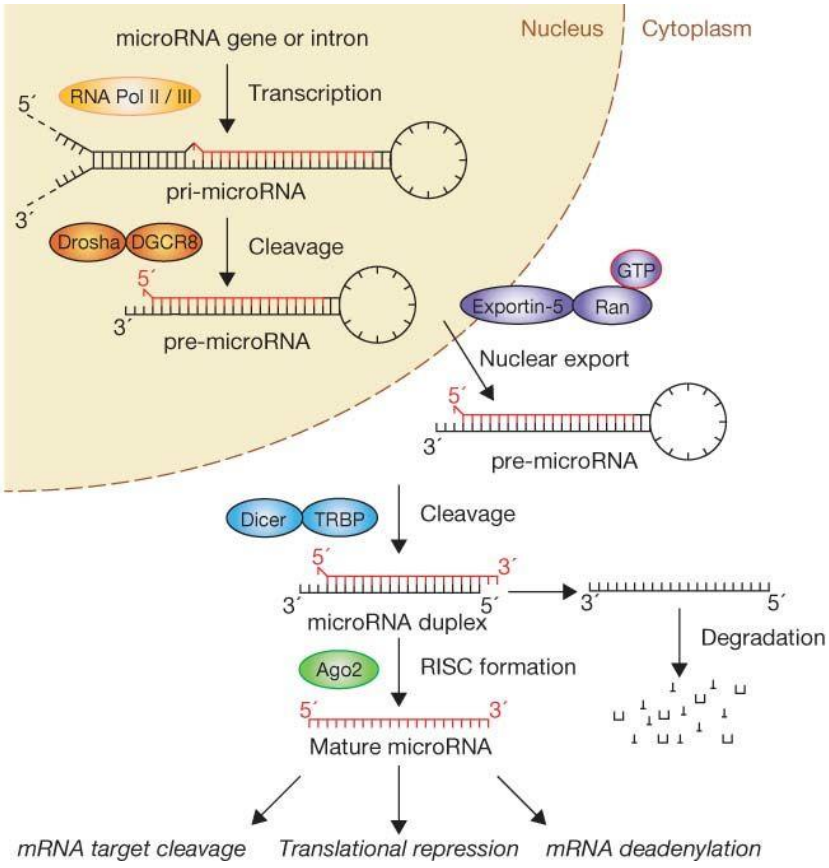


Figure – 1.6. MicroRNA Biogenesis pathway
Section 1.6.1 discusses the miRNA biogenesis pathway.
Imaged adapted from Winter et al. (2009)

1.6.2 – MicroRNA Target Recognition

Target recognition of miRNAs relies upon sequence complementarity between the miRNA and the mRNA target in a Watson-Crick manner. In plants, perfect (or near perfect) complementarity is required along the full ~22 nt miRNA sequence for target recognition resulting the repression of the mRNA by degradation of the target (Bartel, 2009). In comparison, animals as well as viruses do not require near perfect or perfect complementarity. The nucleotide sequence at positions 2 – 8 nt at the 5' end of the miRNA which is termed the seed sequence is considered to be the most important for target recognition (Gottwein et al., 2007). The targeting by, and binding with the seed sequence is known as canonical binding and miRNAs that are conserved across different species use the same seed sequence (Lewis et al., 2005). Non-canonical binding between a miRNA and mRNA can also occur. This is where the seed sequence contains a mismatch, the miRNA only matches at the 3' end or at the centre of the miRNA (Helwak et al., 2013). Mismatching of sequences can happen through G:U binding known as G:U wobble. This was first thought to have a negative effect upon the miRNA targeting but binding can still happen though the specificity and activity of the miRNA can change (Macfarlane and Murphy, 2010). Initial observations dating back the first identified miRNAs in *C. elegans* suggested that miRNAs only target the 3'UTR of mRNA transcripts however more recently they have been shown to target mRNAs both in the 5'UTR and the coding region in addition to the 3'UTR (Lee et al., 1993, Grey et al., 2010, Lytle et al., 2007)

1.6.3 – Modes of Action of MicroRNAs

There is general consensus in the literature that miRNAs negatively regulate gene expression. A single miRNA can have multiple targets and conversely, several miRNAs can target a single mRNA transcript (Macfarlane and Murphy, 2010). Reports in the literature suggest that there are two main methods of gene silencing by miRNAs; mRNA cleavage and translational repression. As well as this, other methods have been reported such as mRNA cap inhibition and ribosomal drop-off causing premature termination of translation (Morozova et al., 2012). A recent study in mice observed 48 % of mRNAs targeted by miRNAs are silenced by translational repression with only 29 % by mRNA degradation. The rest (23 %) have a combination of both (Jin and Xiao, 2015).

Cleavage of the target mRNA is catalysed by Ago-2 and tends to require a large percentage of complementarity between the miRNA and its target however there is exceptions to this (Mallory and Vaucheret, 2004). At the beginning of the process, the RISC complex recruits the GW182 family of proteins. This provide a scaffold on which subsequent effector

proteins can be recruited (Behm-Ansmant et al., 2006). Following miRNA-mRNA interactions, deadenylation of the poly(A) tail is carried out by the deadenylase enzymes PAN2/PAN3 and CCR4-NOT with the former initiating the process and the latter completing it (O'Brien et al., 2018). Decapping of a target mRNA can follow this, primarily with decapping protein 2 (DCP2) alongside other related proteins (Valencia-Sanchez et al., 2006). For effective mRNA degradation, both the CCR4-NOT complex and DCP2 enzymes are necessary to facilitate deadenylation and decapping of the mRNA (Behm-Ansmant et al., 2006)

As well causing mRNA decay, miRNA target recognition can also employ a number of other post-transcriptional methods to effect protein expression. These all broadly fall under the category of translational repression however the miRNA target mRNA can have different fates (Morozova et al., 2012). Translation occurs in three stages; initiation, elongation and termination and all three have been shown to be targeted in different aspects. Inhibition of translation initiation can be caused by disassociation of eIF4A from the initiation complex (Fukaya et al., 2014). During elongation, inhibition can occur through a decrease in the number of associated ribosomes thus lowering protein production potential (Morozova et al., 2012). Premature termination of of translation can also occur through ribosome drop off (Petersen et al., 2006). The decision on which stage is targeting through this mechanism is also unclear but some research suggests that the promoter used to transcribe the mRNA influences the mechanism of silencing (Kong et al., 2008).

In addition to repressing translation by the methods outlined above, target mRNAs can be sequestered into Processing bodies (P bodies) within the cytoplasm. These lack translational machinery and so protein production is halted for the targeted mRNA. P bodies have been shown to arise as a consequence of RNAi mediated gene silencing (Eulalio et al., 2007). mRNAs that are sequestered to P bodies are not always decayed and are simply repressed, thus can act as a storage of targeted mRNAs (Hubstenberger et al., 2017). Stored mRNAs within the P bodies can then either be decayed or can by recycled and sent to the polysomes for protein synthesis (Yang and Bloch, 2007).

1.6.4 – Approaches to Identify MicroRNA Targets

Following the identification of miRNAs, an entirely new field in gene regulation opened up looking at miRNA target recognition and analysis of their targets. MicroRNAs are small and consist of RNA making elucidation of targets tricky due to their size and stability. Despite this, several approaches have been successfully used to identify their targets either through *in silico* approaches or more traditional *in vitro* laboratory techniques. With the advent of next

generation sequencing and tumbling costs of this technology, a combination of bioinformatics and laboratory techniques is now more frequently used.

Bioinformatics can be used to predict miRNA targets and a number of computer algorithms and programmes have been developed by independent groups. These programmes include; miRTar (Hsu et al., 2011), miRanda (John et al., 2004), miRDB (Wong and Wang, 2015) and RNA Hybrid (Rehmsmeier et al., 2004). They all work in slightly different manners to achieve the same end goal of target prediction. Differences include sequence complementarity, thermo dynamic stability and data input methods. Whilst bioinformatic prediction is a powerful tool, it does have several drawbacks. Firstly, predictions generated can contain false positives which affects the confidence in the process and secondly, any predictions made *in silico* have to be validated experimentally *in vitro*.

There are several *in vitro* approaches that can be taken to identify miRNA targets. They can be split into three distinct research strategies; transcriptomics, biochemical and proteomic analysis. Transcriptomics measures changes in gene expression in the presence of a stimuli which can be either the presence or loss of a mature miRNA. Microarrays are the most common approach for this method and they have the ability to test a few hundred to several thousand transcripts at once allowing for multiplexing and the testing of several potential targets on the same chip (Thomson et al., 2007). RNA sequencing (RNA-seq) can also be applied in the same way using specific miRNAs but both techniques have limitations. Primarily, any specific changes in gene expression levels seen cannot be attributed to a specific miRNA and so further validation steps are required to match the gene target and the miRNA.

There are a number of different biochemical approaches that have been developed since miRNAs were first discovered. They are more sensitive than transcriptomic approaches as they revolve around maintaining miRNA:mRNA interactions so as to identify specific targets through immunoprecipitation (IP) methods (Doelken et al., 2010). IPs typically use the Ago-2 protein of the RISC complex as the main target as it is this protein that is anchored to the miRNA (RISC-IP). These pull downs can then be processed further through microarray analysis or by next generation sequencing to determine targets. If a known target is been pulled down, RT-qPCR can be used to determine the level of enrichment of a specific mRNA against a whole cell lysate sample (Pavelin et al., 2013). Similar to transcriptomic approaches, RISC-IP does have some drawbacks. One particular problem is the pulling down of indirect targets as the approach is not very stringent.

RISC-IP has been improved upon several times; each becoming more sensitive than its predecessor. The development of high throughput sequencing of RNAs isolated by crosslinking and immunoprecipitation (HITS-CLIP) improved RISC-IPs by utilising UV crosslinking. This crosslinked the RISC to the mRNA target at the miRNA:mRNA interaction. By using this technique, the sensitivity increased which in turn made the approach more powerful (Chi et al., 2009). HITS-CLIP has been used several times in a number of different herpesviruses to identify miRNA:mRNA interactions (Guo et al., 2015, Riley et al., 2012).

Photoactivable-ribonucleoside-enhanced CLIP (PAR-CLIP) became the successor to HITS-CLIP through improvements and efficiencies in the technique. Incorporation of ribonucleoside analogues that are photoreactive, such as 4-thiouridine (4-SU) and 6-thioguanosine (6-SG) into RNA transcripts allows for improved crosslinking efficiency when UV irradiated thus increasing the amount of mRNA pulled down compared to HITS-CLIP (Hafner et al., 2010). PAR-CLIP has been used numerous times to also identify miRNA:mRNA interactions (Pan et al., 2014, Benhalevy et al., 2017). Whilst both HITS-CLIP and PAR-CLIP are more stringent than RISC-IP, they still do not directly map out specific miRNA:mRNA interactions and downstream work is required to match miRNA sequences to their mRNA targets.

The latest and most sensitive technique is known as cross linking and sequencing of hybrids (CLASH) (Helwak et al., 2013, Helwak and Tollervey, 2014). It builds upon HITS-CLIP by adding an extra step whereby the miRNA and mRNA bound by UV crosslinking are ligated together to create a chimera. These ligated miRNA:mRNA chimeras can then be identified in the sequencing stage. Most recently, a quicker version of this approach has been developed called qCLASH that is essentially the same however a lower input quantity can be used (Gay et al., 2018). Coupled with this approach is a bioinformatic pipeline that was specifically developed to partner CLASH to identify these interactions (Travis et al., 2014).

In addition to the development of CLASH/qCLASH as described, the use of locked nucleic acids (LNAs) as bait to pull down specific miRNA interactions (Xu et al., 2017). The experimental procedure referred to as Target-Link relies upon a LNA with a sequence antisense to the miRNA of interest. This LNA can have a protein attached such as biotin which can be used to purify the miRNA:mRNA complexes of interest that have been UV crosslinked together. Due to the high affinity binding between the LNA and miRNA, purification steps can be more stringent reducing the number of false positives that are seen

with some other methods. This approach maybe of use when looking at specific miRNA:mRNA interactions that have been investigated by other experimental methods.

Highlighting this combination of using both bioinformatics and the wet-lab based techniques described above are the number of new computer algorithms that take raw data reads generated from next generation sequencing of CLIP/CLASH experiments to identify miRNA:mRNA interactions. These programmes include miRTarCLIP (Chou et al., 2013) and MiRTarget (Wang, 2016).

The final method of identifying miRNA:mRNA interactions is using a proteomic approach. Stable isotope labelling by amino acids in cell culture (SILAC) can track the synthesis and degradation of proteins in a cell. Addition of a miRNA to a sample will change the levels of specific proteins and this can be analysed by SILAC methods (Gallaher et al., 2013). Whilst this approach can be used, it does have its downsides. The major flaw in using proteomic approaches is that only one miRNA can be investigated at once which is much narrower than the global approaches of CLASH for example. Secondly, protein deregulation may not necessarily be due to the addition of a particular miRNA and may be an indirect effect.

1.7 – Virus-encoded MicroRNAs

Viruses are obligate intracellular pathogens and it is well known that they modulate the cellular environment to their own advantage. It is therefore no surprise that viruses were found to encode miRNAs considering that they are non-immunogenic and take up relatively little space in already compact genomes (Pfeffer et al., 2004a, Sarnow et al., 2006).

The first virus-encoded miRNAs were identified in cells latently infected with EBV. It was found that these miRNAs were differentially expressed depending upon the latent stage of the virus life cycle (Pfeffer et al., 2004a). The latest release of MiRBase (Release 22, March 2018) lists 530 mature miRNAs from 35 viruses (shown in Table – 1.1). Most virus-encoded miRNAs discovered thus far have been identified in large DNA viruses with the majority identified in herpesviruses (Grey, 2015). Some such as poxviruses however have not been shown to encode for miRNAs and this most likely due to the location of replication. Herpesviruses and adenoviruses both replicate in the nucleus of infected cells whereas poxviruses replicate in the cytoplasm so lack access to critical components of the miRNA processing machinery (Cullen, 2010).

Furthermore, RNA viruses have also been shown to encode for miRNAs although there has been some controversy regarding whether or not they are real. This was the case for HIV-1 which is still listed to encode for 4 mature miRNAs from 3 precursors on MiRBase

(Bennasser et al., 2004, Pfeffer et al., 2005). The retroviruses bovine foamy virus (BFV) and simian foamy virus (SFV) have both been shown to encode for 4 and 13 mature miRNAs respectively (Whisnant et al., 2014, Kincaid et al., 2014). Both of these studies observed that these miRNAs were generated via non-canonical pathways as the primary miRNA (pri-miRNA) transcripts are transcribed by RNA pol III as opposed to RNA pol II. Another retrovirus virus that has been shown to encode for miRNAs is bovine leukaemia virus (BLV) which encodes for 10 mature miRNAs from 5 precursors (Rosewick et al., 2013).

Similar to their hosts, viruses can also encode for other types of ncRNA in addition to miRNAs. There are several instances of viruses encoding for lncRNAs and there is an emerging field documenting the identification of circRNAs encoded for within viruses. Possibly the most well studied virally encoded lncRNA is the HSV-1 encoded latency associated transcript (LAT). It has been strongly implicated in the epigenetic regulation of HSV-1 gene regulation during latency and directly influences reactivation of the virus (Nicoll et al., 2016). HCMV has also been shown to encode for at least 4 lncRNAs, one of which, lncRNA 2.7, is the most abundant transcript found during the early stages of HCMV infection and it has numerous roles including binding to retinoid/interferon-induced mortality-19 subunits of the mitochondrial enzyme complex 1 to prevent stress induced apoptosis and allowing continued ATP production (Reeves et al., 2007). lncRNAs can also be used to subvert the host cell antiviral response. KSHV encodes for the lncRNA PAN which during the lytic phase of infection can reduce the expression of IFN-α, IFN-γ and ISGs (Qiu et al., 2018).

In addition to lncRNAs, viruses have been recently shown to also encode for circRNAs. Both KSHV and EBV have been shown to have circRNAs (Tagawa et al., 2018, Huang et al., 2019). As this is a very emerging field, information regarding the function of the circRNAs is still sparse however the study by Huang et al. (2019), explored the regulation of host miRNAs by the virus-encoded circRNAs. The study found several host miRNAs including miR-28-5p, miR-151a-5p and miR-1248 were down regulated by EBV-circRPMS1. As previously mentioned in section 1.6, circRNAs have the ability to regulate other ncRNAs. This adds another dimension and layer of complexity when one is investgating the role of non-coding RNAs during virus infection and no doubt this emerging field will rapidily expand with the identification of circRNAs in other virus families also.

Table – 1.1. List of Virus-encoded MicroRNAs adapted from MiRBase

Family	Subfamily	Virus	No. of mature miRNAs
Herpesvirus	α-herpesvirus	Bovine Herpesvirus 1	12
		Bovine Herpesvirus 5	5
		Duck enteritis virus	33
		Herpes B virus	15
		Herpesvirus of Turkeys	28
		HSV-1	27
		HSV-2	24
		ILTV	10
		MDV-1	26
		MDV-2	36
		Pseudorabies virus	18
	β-herpesvirus	HCMV	26
		Murine CMV	29
	γ-herpesvirus	EBV	44
		HSV	6
		KSHV	25
		Rhesus lymphocryptovirus	70
Papillomavirus		BPCV1	1
		BPCV2	1
Polyomavirus		Simian virus 40	2
		Raccoon polyomavirus	2
		Gorilla polyomavirus 1	2
Retrovirus		Bovine Leukaemia virus	10
		Bovine Foamy virus	4

Adapted from MiRBase (release 22, March 2018). Some have been omitted from the full list.

1.7.1 – Cellular Targets of Virus-encoded miRNAs

It is no surprise that many of the characterised virus-encoded miRNAs found in herpesviruses regulate cellular pathways that are key to survival including apoptosis, immune evasion and cell cycle. All of these are crucial for the maintenance of latency but by regulating these pathways, viral pathogenesis and in the case of some herpesviruses, oncogenic development can be controlled.

Firstly, there a number of reported viral miRNAs that target genes associated with apoptosis. By blocking this pathway, the virus promotes cell survival which is naturally beneficial to the invading pathogen. EBV encodes several miRNAs that target several parts of this pathway. A report observed that 12 EBV miRNAs tested *in silico* against the 3'UTR of caspase 3 showed one or more binding sites. When tested in in a reporter based system, 9 of these showed a statistically significant reduction in caspase 3 expression (Harold et al., 2016). Caspase 3 is known to be the mediator of apoptosis and by targeting this, EBV can help maintain a persistent infection. EBV miRNAs have also previously been shown to target genes upstream of caspase 3. EBV-miR-BART5 has been shown to target the pro-apoptotic gene P53 upregulated modulator of apoptosis (PUMA). Deletion of this miRNA from the EBV genome caused increased levels of PUMA mediated cell death suggesting that this miRNA helped to protect EBV infected cells from apoptosis (Cullen, 2009). Taken together, the literature surrounding viral miRNAs and apoptosis avoidance suggests that several genes in the apoptotic pathway are targets of these viral miRNAs and by reducing their expression it promotes cell survival and thus survival of the virus.

Viral miRNAs can also target genes associated with the triggering of the immune system thus promoting the chance of immune evasion. HCMV miRNA HCMV-miR-UL112 has been shown to target the major histocompatibility complex class I–related chain B (MICB) gene, firstly through bioinformatic prediction and then with more traditional laboratory methods. The consequence of this targeting is a reduced number of activated natural killer cells (NK) as MICB is the ligand for natural killer cell activating receptor (NKG2D) which causes NK cell activation (Stern-Ginossar et al., 2007). This targeting of MICB expression by herpesviral miRNAs is somewhat conserved with both EBV and KSHV encoding miRNAs that also target this gene indicating that this pathway of NK cell activation shutdown is crucial in the herpesviral lifecycle (Nachmani et al., 2009).

Finally, the cell cycle can be targeted by virus-encoded miRNAs. One such report observed that HCMV encodes the miRNA HCMV-miR-US25-1 that was found to target several genes associated with cell cycle control including cyclin E2, BRCC3, EID1, MAPRE2, and CD147

and deletion of this miRNA caused an over expression of cyclin E2 with regards to viral infection (Grey et al., 2010). Similarly, a miRNA encoded by herpesvirus saimiri (HVS), HVS-miR-HSUR5-3p targets WEE1 which is a negative regulator of the cell cycle. This downregulation leads to the reduced phosphorylation of one of its substrates, cyclin dependent kinase 1 (CDK1) (Guo et al., 2015).

Overall, targeting of cellular transcripts with virus-encoded miRNAs helps viruses to survive in a hostile environment. It allows for the regulation of specific targets and pathways in a non-immunogenic way thus assisting in the completion of the viral life cycle.

1.7.2 – Viral Targets of Virus-encoded miRNAs

As well as targeting cellular transcripts, virus-encoded miRNAs have the ability to target other viral transcripts. As previously stated, the majority of virus-encoded miRNAs identified to date have been in the herpesvirus family. The ability to set up a life-long infection requires precise control of gene expression and miRNAs are a method of regulation during these latent periods (Goodrum et al., 2012). There is a large body of evidence in the literature that herpesviruses use miRNAs as a means to maintain and promote latency. This is achieved by targeting viral genes that are typically necessary for lytic replication and/or reactivation. For example, KSHV mature miRNAs are produced from 12 stem-loops and are all encoded for in the latency associated locus. This complex region allows for the expression of the miRNAs and the latency associated nuclear antigen (LANA) in a co-ordinated manner (Gottwein, 2012). Both KSHV-miR-K12-7 and KSHV-miR-K12-9 target ORF50 (RTA), a protein that is required for the induction of lytic replication (Qin et al., 2017). Similarly, miRNAs found in HSV-1, HCMV and OvHV-2 target essential immediate-early gene transcripts (Duan et al., 2012, Grey et al., 2007, Riaz et al., 2014).

Control of gene expression is a major part of the viral life cycle and by utilising virus-encoded miRNAs to regulate viral gene expression as well as endogenous targets, it helps the virus gain a competitive edge against the cellular immune system. This in turn can drive pathogenesis of the virus or as is the case with herpesviruses, help to maintain a life-long infection in the host.

1.7.3 – Viral Targets of Cellular-encoded miRNAs

As the targeting of transcripts is dictated by the miRNA sequence, any transcript containing the corresponding sequence can be silenced. As previously explored in section 1.7.2, viral targets can be silenced by virus-encoded miRNAs however there is a growing body of evidence that endogenous miRNAs are also able to target viral transcripts thus affecting vital viral processes. The first reports of endogenous miRNAs targeting viral transcripts came out

just over a decade ago and they have been found to both promote and inhibit virus replication (Cullen, 2011). The liver specific miRNA miR-122 was reported to bind to two sites in the 5'UTR of Hepatitis C virus (HCV). It was observed that this miRNA is essential for HCV replication and also enhanced HCV mRNA translation (Jopling, 2008, Jopling et al., 2005). Introduction of a locked nucleic acid (LNA) antisense to the miRNA caused a reduction of HCV viral burden *in vivo* (Cullen, 2011).

Conversely, endogenous miRNAs also have the ability to inhibit virus replication. This has been reported for several viruses with the introduction of artificial miRNA target sites into the viral genome but there is no reason why cellular miRNAs cannot natively target viral transcripts. An early report showed that the miRNA miR-32 was able to inhibit the replication of primate foamy virus (PFV) in the 293T human kidney cell line however the physiological relevance of this was trivial considering PFV does not replicate in the kidneys *in vivo* (Lecellier et al., 2005). One such study that showed physiological relevance as well as the inhibition of viral transcripts was carried out with HSV-1. The report by Pan et al. (2014), found the neuronal specific miRNA miR-138 was able to target ICP0 and repress its expression. The ICP0 gene functions as a transactivator of lytic gene expression and suppression by miR-138 promoted host survival and latency in mouse models. Similar reports of miRNA targeting of essential genes for reactivation from latency have been reported in other herpesviruses.

These interactions add another layer of complexity in understanding virus-host interactions and the regulation of viral gene expression with regards to viral pathogenesis. As approaches to identify miRNA interactions become more sensitive and powerful (reviewed in section 1.6.4) the number of reports describing cellular miRNAs targeting virus transcripts will undoubtedly increase.

1.7.4 – ILTV-encoded miRNAs

Initially, 8 miRNAs were identified in the ILTV genome by deep-sequencing with four of these then validated by northern blot analysis (Waidner et al., 2009). In the same year, a second lab group, also using sequencing, identified the previously published 8 miRNAs plus an extra 2 novel miRNAs. They went onto confirm the expression of all 10 miRNAs by PCR using small-RNA enriched libraries (Rachamadugu et al., 2009). To date, these are the only identified and validated miRNAs within the ILTV genome. Figure – 1.7 shows the position of the miRNAs relative to the ILTV genome in a representative diagrammatic format.

Little work has been carried out regarding the role of these ILTV-encoded miRNAs with only one paper to date identifying one target of the miRNAs ILTV-miR-I5 and ILTV-miR-I6. Both of these miRNAs are encoded anti-sense to the major immediate-early gene ICP4 and previous work identified ICP4 as a target of both ILTV-miR-I5 and ILTV-miR-I6. Using a luciferase based reporter system, the group found the addition of miR-I5 caused a statistically significant 60 % reduction in luciferase activity whilst the addition of miR-I6 caused a reduction of 23 % that was not significant. The effects observed were alleviated upon addition of an antagonist to the respective miRNAs or by mutating the miRNA seed sequence. Further work went on to identify that miR-I5 reduced ICP4 transcript levels by 50 % and it did this by cleavage of the mRNA at the miRNA binding site as validated by modified RACE analysis (Waidner et al., 2011).

Table – 1.2. ILTV-encoded miRNAs

MicroRNA Name (Mature Sequence)	Sequence	Accession Number
ILTV-miR-I1-5p	AGACUGAUUGGGGAAUGAUUGG	MIMAT0012723
ILTV-miR-I1-3p	AAUUCCAUUCCUCUUUCUGUCUCC	MIMAT0012724
ILTV-miR-I2	GGAAGGCUGUGCGAUAGGAGCCGA	MIMAT0012725
ILTV-miR-I3	UCUUGUCUCUGGGUGGGUUCGGA	MIMAT0012726
ILTV-miR-I4	AUGUAUAGCGAGCAAUGACCGUGU	MIMAT0012727
ILTV-miR-I5-5p	CUUCUCGUCCCCGUCUUCUUCAGA	MIMAT0012728
ILTV-miR-I5-3p	UGAAGAAGACGACGACGAGGAGCAU	MIMAT0012729
ILTV-miR-I6-5p	GUCUCCUGUACCCUCAUCGUCG	MIMAT0012730
ILTV-miR-I6-3p	ACGCUGAGGGGCCAUGAGACAGU	MIMAT0012731
ILTV-miR-I7	UUUUAAUACUGAGGUGCGAAUG	MIMAT0012861

Adapted from MiRBase (Release 22, March 2018)

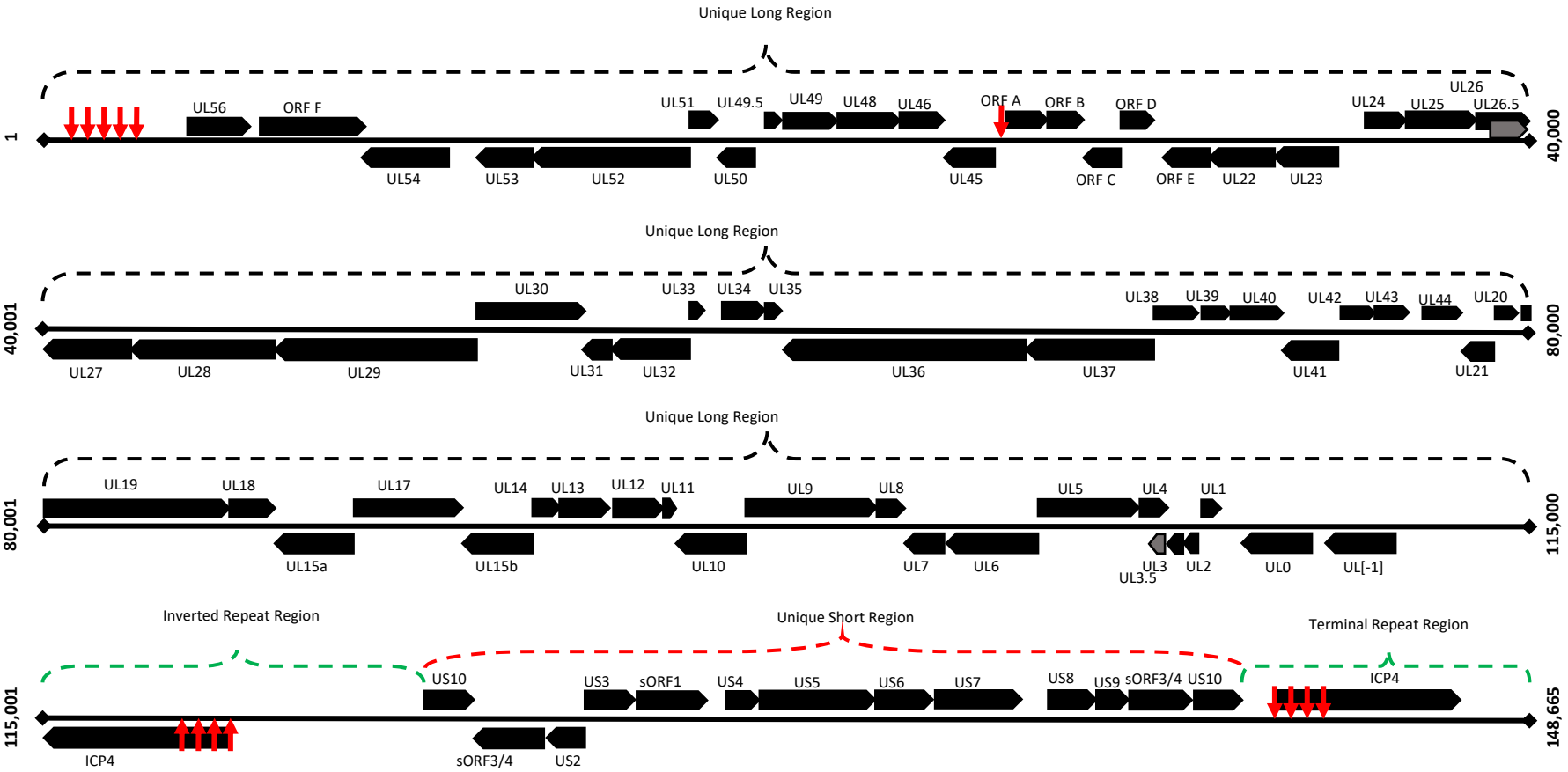


Figure – 1.7. Genome Layout of ILTV

Genomic layout of ILTV showing each of the 79 ORFs in the genome (solid black arrows). Direction of arrows depicts orientation of the ORF coding in either a left to right or right to left orientation. Red arrows show locations of the virus-encoded miRNAs. Adapted from Thureen and Keeler (2006).

1.8 – Genome Editing

Genome editing is a natural occurrence in nature through spontaneous mutations to DNA causing deletions, substitutions and insertions in the genome. Early work on genome editing showed that these mutations could be increased with the use of radiation or chemical treatment (Muller, 1927, Auerbach et al., 1947). It took a further 30 years or so for genome editing to be targeted using a process of homologous recombination; something which is still in use today (Smithies et al., 1985). This approach however, is highly precise but very inefficient and so it takes a large amount of time and patience to generate the desired mutations. Currently, there are three separate approaches that all revolve the ability to make double stranded breaks in DNA at a precise location to induce a programmable change in the genome either through non-homologous end joining (NHEJ) or homology directed repair (HDR) which relies upon the cell to repair DNA damage (Carroll, 2017). Discussed below are three systems, zinc-finger nucleases (ZFNs), transcription activator like effector nucleases (TALENs) and clustered interspaced short palindromic repeats (CRISPR). All three are in use today however the latter is now the most dominant. The systems have all been employed in a number of different species across the natural kingdom for a wide variety of different applications. In future, these systems have the ability to affect the natural world profoundly. This includes applying to system to agriculture, germline editing and the control of invasive species through gene drives (Carroll, 2017).

1.8.1 – Zinc Finger Nucleases (ZFN)

Genome editing using Zinc finger nucleases (ZFNs) relies upon a fusion between a zinc finger DNA binding domain and a DNA cleavage domain. By modifying the former of these domains, it allows for the targeting of specific sequences in a genome (Carroll, 2011). Each zinc finger (ZF) comprises around 30 amino acids and can bind to 3 nucleotides (nts) of DNA in a modular fashion. Subsequently, it was shown that by altering these ZFs, different sequences of DNA could be targeted (Pabo et al., 2001).

The DNA cleavage domain of ZFN is a type 2 restriction endonuclease called *Fok1*. Early work on the proteolytic fragments of *Fok1* showed that it contained both a DNA-binding domain and a non-specific DNA-cleavage domain (Kim et al., 1996). By combining the ZFs with the *Fok1* restriction endonuclease, ZFN were synthesised.

Following the synthesis and successful testing of ZFNs *in vitro* it was soon applied to a variety of different species with different targets using both NHEJ and HDR. Activity in animals was first shown in *Drosophila melanogaster* with higher order animals such as

zebrafish following (Carroll, 2011). ZFNs have since been succeeded by more specific and easier systems that offer more precision and greater flexibility in targeting of genomes.

1.8.2 – Transcription Activator-Like Effector Nucleases (TALENs)

The successor to ZFNs, transcription activator like effector nucleases (TALENs) are an advancement on ZFNs as they still rely upon the Fok1 restriction enzyme (nuclease domain) to cut the DNA however they use transcription activator like effectors (TALEs) which can be programmed to target any DNA sequence allowing for more flexibility. By combining a TALE with a nuclease domain, precise genome editing can be carried out (Boch, 2011). TALE proteins were first identified in pathogenic bacteria that infect plants. They were shown to contain a repeated ~34 conserved amino acid sequence with differing amino acids at positions 12 and 13 known as the repeat variable diresidue (RVD) (Boch and Bonas, 2010). TALEs work on the basis of amino acid sequence recognising DNA sequence and by altering this ~34 amino acid sequence with appropriate RVDs allows for the targeting of specific DNA (Boch, 2011).

TALENs have been applied to a large number of different organisms to efficiently modify their genomes including plants of economic importance and to generate knockout animals used in research (Tesson et al., 2011, Haun et al., 2014). Their usage has been far reaching due to the ability to target individual nucleotides when compared to ZNF nucleotide triplets however with development of CRISPR-Cas systems as outlined below that are cheaper and offer greater flexibility in targeting, the usage of TALENs has in declined in recent years.

1.8.3 – Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)

The most recent advancement in precise genome editing is using the CRISPR system. Although its application and usage has exploded in recent years, CRISPR has been known about for over 30 years with the first observations of clustered repeats made in 1987 (Ishino et al., 1987). A further 20 years passed before these clustered repeats were identified as derivatives of bacteriophage DNA and other viruses that had tried to attack the cells (Pourcel et al., 2005). Previous to this however was the discovery of a set of genes termed CRISPR associated systems genes (Cas genes) that encoded proteins with both helicase and nuclease motifs which were suggested to have a functional relationship with the CRISPRs (Jansen et al., 2002). It took until 2010 for the first publications to show the targeting and cleavage ability of the CRISPR-Cas system against fragments of bacteriophage and plasmid DNA in the bacteria *Streptococcus thermophilus* (Garneau et al., 2010). Research into Cas systems rapidly increased and so far 93 individual *cas* genes split into 35 families have been discovered (Makarova et al., 2015). To date, *cas* genes have been identified that can cleave

single stranded DNA (ssDNA), double-stranded DNA (dsDNA) and also even RNA (Sinkunas et al., 2011, Gasiunas et al., 2012, Cox et al., 2017). The simplistic system relies upon a guide RNA (sgRNA) that is homologous to the region targeted for cleavage and a small transacting RNA (tracrRNA) that helps to process the sgRNA and cleavage of the target when coupled with a single *cas* protein that carries out the cleavage of desired target (Carroll, 2017).

The most utilised CRISPR system for genome editing is the CRISPR-Cas9 system which was shown to be programmable to cleave specific sequences in the human genome simultaneously by two independent research labs (Hsu et al., 2014, Mali et al., 2013). It has now been applied to a large variety of organisms to make precise genomic changes as well as allowing researchers to activate and repress specific genes (Larson et al., 2013).

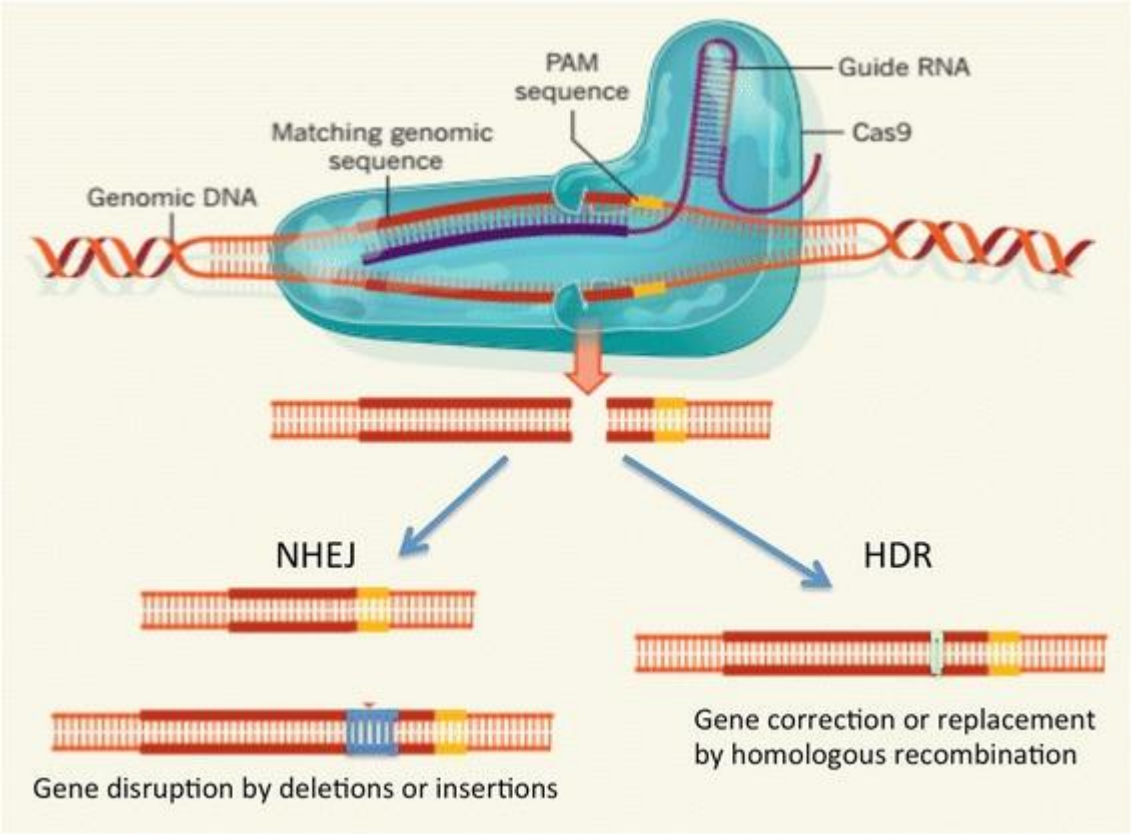


Figure – 1.8. Genome Editing using the CRISPR-Cas9 System
The CRISPR-Cas9 system relies upon a target sequence of 20 bp within a sequence of DNA (or RNA depending upon the Cas protein) that has a Photospacer Adjacent Motif (PAM) that is essential for Cas9 binding and cleavage. Once DNA is cleaved by the Cas9 protein, host cell machinery repair the double stranded break (DSB) either through Non-Homologous End Joining (NHEJ) or Homology Directed Repair (HDR). Figure adapted from Tu et al. (2015)

1.8.4 – Creation of Recombinant Viruses

The first reports of targeted editing of viruses using homologous recombination came out some 40 years ago (Mocarski et al., 1980). Since this report, a vast number of recombinant viruses have been created using this approach to aid investigations into viral gene function and inadvertently cell signalling amongst other things. As stated previously homologous recombination is very precise but highly inefficient even on large animal genomes. Anecdotal reports suggest successful homologous recombination in viruses is as low as 1 in a million virions produced. This approach was updated in 1996 when it was shown that inducing double stranded breaks (DSBs) into the viral DNA greatly improved the efficiency of targeted recombination (Ryan and Shankly, 1996). The downside to this was that it relied upon a unique restriction endonuclease site within the Suid Herpesvirus-1 genome, thus greatly reducing the targeting ability of this approach with improved efficiency.

Aside from homologous recombination, the development of the Bacterial Artificial Chromosome (BAC) system allowed for even more control over the production of mutant viruses. First reported in 1997, the BAC system works by cloning the entire viral genome into an artificial chromosome of stable bacteria such as *Escherichia coli* (Messerle et al., 1997). This allows for mutations to be made in the viral DNA as if it were a plasmid however the drawback is that in the early uses, residual bacterial sequences were incorporated into the virion. Another drawback to this approach was that not all viruses could have BACs created. In particular, several attempts have been made to produce an ILTV BAC but to no avail. This is thought to be down to the palindromic sequences found within the viral genome that are toxic to the bacteria. More recently however, a cosmid system using yeast has been reported for ILTV using three plasmids to encode for the full genome (Spatz. et al., 2018).

The advancement of precision genome editing has also affected the methods of creating recombinant viruses. Reports of ZFNs being effectively used to create recombinant viruses are sparse however the technology was explored as a possible gene therapy to chronic viral infections as well as targeting cellular genes to confer resistance against HIV-1. ZFNs targeting the genomes of HSV-2 and HIV-1 have been reported both with the conclusions that the approach could be used for gene therapy (Wayengera, 2011, Wayengera, 2012). Moreover, targeting of the cellular genome to confer protection against viral infection has also been reported using ZFNs. Targeting of the CCR5 gene in CD4+ T cells conferred protection against HIV-1 infection both *in vitro* and *in vivo* (Perez et al., 2008).

TALENs have been used in some capacity to modify viral genomes. Bi et al. (2014), reported the use of both TALENs and CRISPR-Cas9 systems to modify adenoviruses. For proof of principle, a recombinant adenovirus was used that carried the enhanced green fluorescent protein (eGFP) with TALENs and sgRNAs designed against it. Results showed that both systems were able to competently cleave the target DNA however the CRISPR-Cas system was reported to be much more efficient at cutting the sequence (Bi et al., 2014). The TALENs system has also been used to target the genome of Hepatitis B virus (HBV) genome. Four sets of TALEs were designed against the episomal covalently closed circular HBV DNA (cccDNA). Around 35 % of target cccDNA molecules had some form of mutation and replication was impaired (Bloom et al., 2013). The findings also suggested that genome editing technologies could not only be used to create recombinant viruses but also to efficiently clear chronic viral infections.

With the advent of the CRISPR-Cas system, it was soon applied to viruses. In the first instance, sgRNAs were designed to target the LTR of HIV-1. Loss of the LTR caused a reduction in HIV-1 gene expression (Ebina et al., 2013). The following year it was applied to herpesviruses and since then a number of papers have come out detailing recombinant viruses created using the CRISPR-Cas system (Russell et al., 2015). Since the first papers were published detailing that this system could be used to modify herpesviruses, a large number of recombinant viruses have been created and in the case of HSV-1, these modifications are detailed succinctly by Wang et al. (2018). Across the reports there is a general consensus that for efficient editing of herpesviruses, two plasmids are transfected into cells prior to infection with virus. These plasmids consist of the donor template containing the homologous repair arms with the desired mutations (gene knock out/fusion gene ETC.) and a second plasmid encoding for the sgRNA and cas9 protein. Effieciency improvements vary but one report suggests a 10,000 fold increase in efficiency in HDR using CRISPR-Cas9 compared with homologous recombination (Lin et al., 2016). Selection of viruses containing the desired mutations is variable and dependent upon the marker used for selection. Typically, fluorescent markers are used such as eGFP/RFP as this allows for FACs processing but other methods have also been used such as plaque assays. No doubt with these systems becoming cheaper and more accessible, the continuation of modifying viruses via targeted genome editing will continue.

1.9 – Aims

Elucidating the roles of ILTV-encoded miRNAs is an important step forwards in the understanding of ILTV biology in the broader area of virus-host interactions. Previous work in the literature has already identified and validated these miRNAs however there is little knowledge on their targets both in the viral genome as well as the host genome.

The major aims of this thesis were to:

- Characterise the expression of viral genes and viral miRNAs in a temporal manner and investigate the tropism of ILTV in a number of different avian cell lines *in vitro*
- Employ the use of the experimental procedure CLASH (outlined in section 1.6.4) to identify miRNA:mRNA interactions during ILTV infection
- Use bioinformatics to predict viral targets of virus-encoded miRNAs and test a panel of these in a laboratory setting by a number of methods
- Identify viral targets of high-confidence cellular miRNAs by bioinformatic analysis and investigate these interactions and validate them using laboratory methods
- Explore the potential use of genome editing technology, specifically CRISPR-Cas9 and whether it is a viable option to create recombinant viruses. Efforts were focused upon the deletion of five miRNAs found at the left hand end of the genome and replacing them with a reporter cassette

Chapter 2: Methodology

2.1 – Tissue Culture

2.2 – PCR Methods

2.3 – Cloning Methods

2.4 – Western Blotting

2.5 – Production of a Stably Expressing Cell Line

2.6 – Other Methods

2.1 – Tissue Culture

2.1.1 – Cell lines used

HEK293T Cells (Graham et al., 1977) (A gift from Dr. N Smith, The Roslin Institute) were grown in T75 flasks with Dulbecco's Modified Eagles Medium (DMEM) containing 8 % foetal bovine serum and 1 % penicillin-streptomycin. Cells were incubated at 37 °C with 5 % CO_2.

HEK293FT cells (Javanbakht et al., 2003) (A gift from Dr. N Smith, The Roslin Institute) were grown in T75 flasks with Dulbecco's Modified Eagles Medium containing high glucose and high pyruvate. Media was supplemented with 10 % foetal bovine serum, 1 % penicillin-streptomycin and 500 µg/ml Geneticin™ selective antibiotic (Invitrogen). Cells were incubated at 37 °C with 5 % CO_2.

LMH cells (Kawaguchi et al., 1987) (A gift from Dr K. Russell, Roslin Institute) were grown in a T75 flask coated with 0.1 % gelatin in Waymouth's media (Invitrogen, Paisley, UK [All reagents were obtained from Invitrogen unless otherwise stated]) supplemented with 10 % foetal bovine serum, 1 % chicken serum, 1 % sodium bicarbonate and 1 % penicillin-streptomycin and incubated at 37 °C, 5 % CO_2.

CLEC213 cells (Esnault et al., 2011) (A gift from Dr P. Quéré, INRA, France) were grown in T75 flasks with Dulbecco's Modified Eagles Medium containing F12 nutrient mix. This was supplemented further with 8 % foetal bovine serum and 1 % penicillin-streptomycin. Cells were incubated at 37 °C at 5 % CO_2.

Primary Chicken kidney cells (CKC's) were prepared fresh (as described previously by Barrow and Lovell (1989)) and maintained in Minimal Essential Medium (MEM). Media was supplemented with 10 % foetal bovine serum and 1 % penicillin-streptomycin. Cells were grown in T175 flasks and incubated at 38 °C, 5 % CO_2.

DF-1 cells (Himly et al., 1998) were grown in T75 flasks with Dulbecco's Modified Eagle's Medium supplemented with 10 % foetal bovine serum and 1 % penicillin-streptomycin. Cells were incubated at 37 °C at 5 % CO_2.

QT-35 cells (Cowen and Braune, 1988) (A gift from Miss C Conceicao, Roslin Institute) were grown in T75 flasks with Dulbecco's Modified Eagle's Medium (DMEM) which was supplemented with 10 % foetal bovine serum and 1 % penicillin-streptomycin. Cells were kept in a 37 °C at 5 % CO_2.

QM-7 (Rong et al., 2014) cells were grown in T75 flasks with Dulbecco's Modified Eagles Medium containing F12 nutrient mix. This was supplemented further with 8 % foetal bovine serum and 1 % penicillin-streptomycin. Cells were incubated at 37 °C at 5 % CO_2.

HD-11 cells (Beug et al., 1979) (A gift from Miss K Miedzinska, Roslin Institute) were grown in a T75 flask with Dulbecco's Modified Eagles Medium containing; 36 ml tryptose phosphate broth (TSB), 10 % foetal bovine serum and 1 % penicillin-streptomycin-glutamate. Cells were incubated at 38 °C, 5 % CO_2.

Primordial Germ Cells (PGCs) (Glover and McGrew, 2012) (A gift from Dr M McGrew, Roslin Institute) were grown in 24-well plates with serum free media. They were maintained at 38 °C, 5 % CO_2.

2.1.2 – Propagation of adherent cells

Cells were passaged every 2-3 days depending upon the cell line. Media was removed from cells and cells washed with sterile PBS. Once washed, cells (except HD11) were detached from the plastic using trypsin-EDTA (Sigma Aldrich, Dorset, UK). They were then seeded out into fresh flasks with appropriate media.

For growth of LMH cells, flasks coated with gelatin were prepared prior to cell passage. 5 mL of 0.1 % gelatin (10 mL for T175) was pipetted into T75 flasks and shaken to coat the bottom of the flask evenly. Prior to the addition of cells, excess gelatin was removed.

HD-11 cells were detached using a sterile scraper. Cells were transferred to a falcon tube and centrifuged at 435 x g for 5 minutes. Supernatant was removed and fresh media containing the appropriate supplements was added. Cells were resuspended, counted and seeded out in fresh T75 flasks.

2.1.3 – Propagation of suspension cells

PGCs were monitored every day. Media was changed every 48 hours by removing half of the media (~ 250 µl) using a pipette. Fresh media was then added to the cells. Cells were passaged by transferring well contents to a falcon tube and centrifuging cells for 5 minutes at 435 x g. Supernatant was replaced with fresh media and cells were counted before been seeded out into a clean 24-well plate.

2.1.4 – Preparation of cells for long term storage

Cells were resuspended to 5 x 10^6 cells/ml and centrifuged at 500 x g for 5 minutes. Cell supernatant was removed and 1 ml of freezing medium (90 % FBS, 10 % [v/v] dimethyl sulphoxide) was used to resuspend the cell pellet. Samples were then transferred to cryovials

and placed in a Mr. Frosty® (Thermo Scientific, Loughborough, UK) which was stored at -80 °C overnight. Samples were put into liquid nitrogen the following day for long term storage.

2.1.5 – Growth of cells from long term storage

Vials of cells were removed from liquid nitrogen and placed in a 37 °C water bath to thaw quickly. Cells were then added to a 25 cm^2 tissue culture flask (T25) and an appropriate quantity of pre-warmed media was added. The flasks were then transferred to the applicable incubator (temperature dependent) and left overnight. The following day, the growth medium was replaced.

2.1.6 – Transfection of cells using Lipofectamine 2000®

HEK293T cells were seeded at a density of $2x10^5$ cells per well in a 12 well plate and left for 24 hours prior to transfection. 100ng of plasmid DNA and 100 nanomolar (nM) miRNA mimic were added to Opti-MEM to a final volume of 50 µl and left to incubate for 5 minutes at room temperature. 5 µl of Lipofectamine 2000 was added to 45 µl of Opti-MEM and also left to incubate for 5 minutes at room temperature. After incubation, the Lipofectamine/Opti-MEM mix was added to the plasmid DNA/miRNA mimic mix and reactions were left to incubate for 15 minutes at room temperature. The total 100 µl Lipofectamine/DNA mixture was added to the cells which were subsequently left to incubate for 48 hours at 37°C with 5% CO_2 before harvesting.

2.1.7 – Transfection of cells using Xfect™ Polymer

The day before transfection, HEK293T cells were seeded at a density of $2x10^5$ cells per well in a 12 well plate. 100 ng of Plasmid DNA and 100 nM of miRNA mimic were mixed to a final volume of 50 µl of Xfect reaction buffer and gently mixed. 0.3 µl of Xfect polymer was added to the mixture and the reaction was mixed gently before been incubated at room temperature for 20 minutes. Following incubation, the total 50.3 µl mixture was added to the cells which were incubated for 48 hours at 37°C with 5% CO_2 before harvesting.

LMH cells were seeded out at density of $1x10^5$ cells per well in a 24 well plate and left to adhere overnight. 100 nM of miRNA mimic was then added to Xfect reaction buffer to a final total volume of 25 µl and mixed gently. 0.3 µl Xfect polymer was added to the reaction and mixed once again before a 20 minute incubation period at room temperature. The total 25.3 µl reaction was then added to the cells which were then placed in a 37°C with 5% CO_2 incubator for 12 hours before infection with ILTV virus.

24 hours before transfection, LMH cells were seeded out at a density of $5x10^5$ cells per well in a 6-well plate and left to adhere. 2.5 µg of plasmid DNA was added to a final volume of 100 µl of Xfect reaction buffer and gently mixed. 0.75 µl of Xfect polymer was added to the reaction and mixed again and subsequently incubated at room temperature for 20 minutes. All of the reaction mixture was then added to the cells, which were then incubated for 48 hours at 37°C with 5% CO_2 before harvesting.

For CRISPR-Cas9 transfections, LMH cells were seeded out at density of $5x10^5$ cells per well in a 6 well plate and left for 24 hours to adhere. Once adhered, 5 µg of repair template DNA and 5 µg of cas9 plasmid DNA which also encoded for the specific sgRNAs were mixed with Xfect reaction buffer to a final volume of 100 µl. Reactions were mixed and then 3 µl of Xfect polymer was added to the samples and mixed again before a 20 minute incubation period at room temperature. The total reaction volume was then added to the cells which were incubated for 5-6 hours at 37°C with 5% CO_2 before cells were infected with ILTV.

2.2 – PCR Methods

2.2.1 – DNA isolation

Total DNA was isolated from cell lines at a maximum density of 5×10^6 cells per sample using the DNeasy Blood and Tissue Kit (Qiagen) as per the manufacturer's protocol. Extracted DNA concentrations were determined using a NanoDrop ND-100 Spectrophotometer (Thermo Scientific) and samples stored at – 20 °C.

2.2.2 – RNA isolation

Total RNA was isolated from cell lines at a maximum density of 1×10^7 cells per sample using the RNeasy Mini Kit (Qiagen) as per the manufacturer's protocol. RNA concentration was then determined using a NanoDrop ND-100 Spectrophotometer and samples stored at – 80 °C.

Total small RNA was isolated from tissues using the miRNeasy Mini Kit (Qiagen). Before isolation of total RNA, tissue was disrupted using stainless steel beads (Qiagen Cat: 69989) in a TissueLyser II machine (Qiagen, Cat: 85300). Samples were shaken for 2 minutes at 27.5 shakes per second and then rotated on the machine and the process repeated. Once tissue was disrupted, the miRNeasy Mini Kit was used as per the manufacturer's protocol. Purified RNA was then determined using a NanoDrop ND-100 Spectrophotometer and samples stored at – 80 °C until used.

2.2.3 – DNase treatment of RNA

DNase treatment of RNA was carried out using the TURBO DNA-free™ Kit (Life Technologies). 1 µg of sample was treated with 2 µl TURBO DNase and 5 µl 10 x TURBO buffer into a final reaction volume of 50 µl. Reactions were placed in a 37 °C water bath for 30 minutes. 6 µl DNase STOP was then added and left to stand at room temperature for 5 minutes.

2.2.4 – RNase treatment of DNA

2 µl of RNAsecure (Thermo Scientific Fisher AM7005) was added to the extracted DNA and samples were vortexed. They were then incubated at 60 °C for 20 minutes in a heat block. Samples were stored at – 20 °C until further use.

2.2.5 – Reverse transcription of RNA

DNase treated RNA was primed with either Oligo (DT) primer or random primers (Both Promega, Southampton, UK). 10 µl of DNase treated RNA was added to 1µl of either Oligo (DT) primer or random primers and incubated at 70 °C for 5 minutes. Samples were then placed on ice. 5 µl of supplied buffer, 30 U Avian Myeloblastosis Virus Reverse Transcriptase (AMV RT) and 2.5 µl dNTPs (10nM) (all Promega) were added to the DNase-treated RNA – Primer mix to a total volume of 50 µl and incubated at 37 °C for 1 hour. cDNA concentrations were determined using a NanoDrop ND-100 Spectrophotometer and samples stored at – 20 °C.

2.2.6 – PCR

All Polymerase Chain Reactions (PCR) were carried out in a Veriti 97-well thermal cycler (Life Technologies). Primer3 (http://primer3.ut.ee/) was used for primer design. Sample reactions contained 1 U of Taq DNA Polymerase (Qiagen), 2 µl 10x reaction buffer (supplied with the Taq DNA Polymerase), 0.4 µl dNTPs (10nM) (Promega), 8 pmols forward and reverse primer in a final volume of 20 µl. Cycling conditions were as follows: 95 °C for 15 minutes, 35 cycles of 95 °C for 30 seconds, 55 – 60 °C for 1 minute (temperature dependent upon the T_m of Primers), 72 °C for 1 – 5 minutes (depending upon length of product amplifying) followed by a final cycle of 72 °C for 7 minutes.

2.2.7 – Colony PCR

Colonies were picked and individually spread over a fresh Luria-Bertani (LB)/agar plates containing appropriate antibiotic (100 µg/ml Carbenicillin or 50 µg/ml Kanamycin) that was divided and numbered with a grid. The spreader was then agitated in 20 µl of dH_2O in a tube labelled with the same number. Numbered plates were incubated at 37 °C overnight.

Samples agitated in water were boiled at 95 °C for 10 minutes. A conventional PCR (section 2.2.7) was used to screen for positive colonies. PCR products were analysed on a 2 % Agarose gel as described in section 2.2.12. One positive colony was then selected and grown in 5 ml LB broth containing appropriate antibiotic, plasmid DNA purified (section 2.3.9) and sent for sequencing (see section 2.3.7). Large cultures were then prepared of positive colonies that contained the correct insert (see section 2.3.10.)

2.2.8 – Mutagenesis PCR

Mutagenesis PCR was carried out using the Quik-Change Lightning Site-Directed Mutagenesis Kit (Agilent Technologies, Manchester, UK). Primers for mutagenesis were designed using the QuikChange Primer Design Program (http://www.genomics.agilent.com/primerDesignProgram.jsp) and primers can be found in appendix 2. Site-Directed Mutagenesis reactions, PCR cycling conditions and *Dpn*1 digestion were carried out as per the manufacturer's protocol.

2.2.9 – Quantitative PCR

Quantitative PCR (both RT-qPCR and qPCR) were carried out in a Rotor Gene RG-3000 (Qiagen). SensiFAST SYBR Hi-ROX Kit (Bioline) was used for reactions that were set up follows: 10 µl SensiFAST SYBR Hi-ROX reaction buffer, 8 pmols of both forward and reverse primer, 2 µl DNA (or 500 ng cDNA template for RT-qPCR) template topped up to 20 µl using dH_2O. Primers were designed using primer3 (http://primer3.ut.ee/) and amplified no more than 200 bp. Cycling conditions were as per the 3-step cycle outlined in the manufacturer's protocol.

2.2.10 – Synthesis of RNA from a DNA template

To synthesise RNA, the NEB Quick High Yield Kit (New England Bio labs, E2050s) was used as per the manufacturer's protocol. A total of 100 ng DNA template was used in a final reaction volume of 30 µl. Reactions were set up and incubated at 37 °C overnight. Once incubated, samples were treated with 2 µl of DNase as per the instructions in the manufacturer's protocol.

2.2.11 – Agarose gel electrophoresis

DNA was analysed by electrophoresis using 0.5 – 3 % agarose gels containing 1 x SYBR® Safe DNA Gel stain (Life Technologies) in Tris-acetate (TAE) buffer. Samples were mixed with 6 x loading dye (New England Biolabs, #B7024S). Electrophoresis was carried out in tanks with gels submerged in TAE buffer at 80 – 100 V for an appropriate length of time for band separation. Estimation of DNA size was carried out using 1 kb, 100 bp or 50 bp

GeneRuler DNA Ladders (Thermo Scientific). Gel images were captured using an alpha UV imager.

DNase treated guide RNA (sgRNA) was then extracted as per section 2.2.2. RNA was eluted using 50 µl of nuclease free water and stored at – 80 °C until needed. Two aliquots of RNA (each 2 µl in volume) were taken and left at room temperature for 1 hour to check for degradation. After the incubation period, samples were tested on a NanoDrop ND-100 Spectrophotometer for RNA quantification. In addition to this, an aliquot of RNA was analysed on a 3 % agarose gel to qualify any RNA degradation and to ensure correct size of RNA. Equipment used for gel electrophoresis was treated with 0.5 M NaOH for 1 hour prior to use to remove any RNases.

2.2.12 – DNA extraction from agarose gels

DNA was extracted from agarose gels using the QIAquick Gel extraction Kit (Qiagen) according to the manufacturer's protocol. Extracted DNA concentration was determined using a NanoDrop ND-100 Spectrophotometer and samples stored at – 20 °C.

2.2.13 – PCR Purification

PCR products were extracted from reactions using the QIAquick PCR purification kit (Qiagen) according to the manufacturer's protocol. Extracted DNA concentration was determined using a NanoDrop ND-100 Spectrophotometer and samples stored at – 20 °C.

2.3 – Cloning Methods

2.3.1 – PCR purification

PCR products were purified using the QIAquick Purification kit (Qiagen) as per the manufacturer's protocol.

2.3.2 – Annealing of oligonucleotides

Oligonucleotides up to 120 bases in length were manufactured by Sigma-Aldrich (Sigma-Aldrich) along with the equivalent reverse complement counterpart. Oligonucleotides were then resuspended in dH2O to a concentration of 100 µM. 1 µl of each pair was mixed with 5 µl of 10 x Buffer 2 (New England Biolabs) to a final volume of 50 µl. Samples were heated to 95 °C for 5 minutes in a heat block and allowed to cool slowly to room temperature. 5 µl of annealed oligonucleotides were then digested with appropriate restriction endonucleases before been ligated into a vector cut with the same endonucleases.

2.3.3 – DNA ligation

When PCR products required sequencing, the TOPO® TA Cloning® Kit (Life Technologies) was used for cloning. 4 µl of PCR product mixed with 1 µl of salt solution and 1 µl of TOPO 4 vector and left to stand at room temperature for 10 minutes.

For cloning DNA/annealed oligonucleotides cut with restriction endonucleases into expression vectors the LigaFast™ Rapid DNA Ligation System (Promega) was used. 100 ngs of vector was used to 2 µl of DNA insert along with 3 U of T4 DNA ligase (supplied in the LigaFast™ kit) in a final volume of 15 µl. Ligations were placed on ice overnight to gently warm to room temperature.

2.3.4 – Transformation of chemically competent cell lines

5 µl of ligated product was added to 25 µl (or 30 µl for STBL3 Cells) of TOP10 chemically competent cells (Life Technologies) and placed on ice for 30 minutes. Samples were then heat shocked at 42 °C for exactly 45 seconds before been placed back on ice for a further 5 minutes. 150 µl of pre-warmed SOC media was then added to each ligation. Samples were then placed in a 37 °C shaking (200 RPM) incubator for 1 hour. After incubation, samples were spread on to LB plates containing appropriate antibiotic (100 µg/mL Ampicillin or 50 µg/mL Kanamycin) and incubated overnight at 37 °C.

2.3.5 – Transformation of Ultracompetent cell lines

2 µl of β-Mercaptoethanol was added to 45 µl of XL-10 Gold Ultracompetent Cells (supplied with Quik-Change Lightning Site-Directed Mutagenesis Kit, Agilent) and incubated on ice for 10 minutes with gentle, intermittent mixing. 2 µl of site-directed mutagenesis PCR product was added to the cells and incubated on ice for 30 minutes. Samples were then heat shocked at 42 °C for precisely 30 seconds before been placed back on ice for a further 2 minutes. 200 µl of pre-warmed SOC medium was added to the samples and placed in a 37 °C shaking incubator for 1 hour at 200 rpm. Following incubation, samples were spread on LB plates and were left in a 37 °C incubator overnight.

2.3.6 – Restriction digests

DNA and annealed oligonucleotides were digested with restriction endonucleases (Afl2, Bbs1, BamH1, Bgl2, EcoR1, Kpn1, Not1, Pme1, Xho1, NEB) according to the manufacturer's instructions using the supplied buffers. Restriction digests were incubated at 37 °C for 1 hour in a total volume of 30 µl for double digests and 40 µl total volume for sequential digests. Oligonucleotides containing endonuclease sites were digested and purified using PCR purification as described in section (section 2.3.1). Expression vectors

were treated with FastAP Thermosensitive Alkaline Phosphatase (Thermo Scientific) according to the manufacturer's protocol before samples were analysed on a 1 % agarose gel. Bands were excised using a clean scalpel and purified as described in section 2.2.13. The DNA concentration of cut expression vectors was then determined using a photospectrometer.

2.3.7 – Sequencing of samples

Plasmid DNA was sent to GATC biotech (GATC, London, UK) as per their instructions.

2.3.8 – Sequencing analysis

DNA sequences were analysed using NCBI nucleotide BLAST (http://blast.ncbi.nlm.nih.gov/Blast). NEBcutter (http://tools.neb.com/NEBcutter2/) was used for restriction digest mapping.

2.3.9 – Plasmid DNA isolation from bacteria (small scale)

A single bacterial colony was isolated from LB plates and placed into 5 ml LB broth containing appropriate antibiotic. Cultures were incubated overnight in a 37 °C shaking incubator set at 200 RPM. Plasmid DNA was extracted using the QIAprep Spin Miniprep Kit (Qiagen) as per the manufacturer's protocol. Eluted plasmid DNA concentration was determined using a NanoDrop ND-1000 spectrophotometer. Samples were stored at – 20 °C.

2.3.10 – Plasmid DNA isolation from bacteria (large scale)

200 ml of LB broth containing appropriate antibiotic was inoculated with 1 ml of culture as set up in section 2.3.9. Cultures were shaken overnight at 37 °C at 200 rpm. Plasmid DNA was then extracted using the Qiagen EndoFree Plasmid Maxi Kit (Qiagen) as per the manufacturer's protocol. Eluted plasmid DNA concentration was determined using a NanoDrop ND-1000 spectrophotometer. Samples were stored at – 20 °C.

2.3.11 – Preparation of Bacterial stocks for long term storage

Colonies that were confirmed to have the correct insert by sequencing were used to set up long term glycerol stocks. 500 µl of culture (a single bacterial colony grown in 5 ml LB broth containing appropriate antibiotic(100 µg/ml Carbenicillin or 50 µg/ml Kanamycin) was mixed with 500 µl 80 % glycerol in H_2O and stored at – 80 °C.

2.4 – Detection of Proteins by Chemiluminescence and Western Blotting

2.4.1 – Sample preparation

Laemmli buffer (Bio-Rad, Hemel Hempstead, UK) was prepared by adding 50 µl of β-Mercaptoethanol to 950 µl Laemmli sample buffer. Cells were harvested as per section 2.1.2

and transferred into a suitable Falcon tube and centrifuged at 435 x g before been resuspended in one volume of sterile PBS. Once re-suspended, an equal volume of Laemmli buffer was added. Samples were mixed by vortexing and placed in a dry hot plate at 95 °C for 10 minutes. After boiling, samples were loaded on to a pre-cast SDS-PAGE gel (Bio-Rad, UK).

2.4.2 – SDS-Polyacrylamide gel electrophoresis

Precast 4-20 % Mini-PROTEAN® TGX™ Gels (Bio-Rad) were used for SDS-PAGE. Samples were loaded into the gel along with Odyssey® Protein Molecular Weight Marker (LI-COR). The gel was run in running buffer (Tris/Glycine/SDS buffer (Bio-Rad) for 90 minutes at 100 volts.

2.4.3 – Transfer of protein to nitrocellulose membrane

For transfer of protein to nitrocellulose membrane (Bio-Rad nitrocellulose mini membranes), the Trans-Blot® Turbo™ Blotting System (Bio-Rad) was used. The Trans-Blot Turbo Transfer Buffer was prepared as per the manufacturer's protocol and then the transfer stacks and nitrocellulose membrane were soaked in the transfer buffer for 2 – 3 minutes. Assembly of the transfer stack is as follows from bottom to top: bottom cassette electrode, bottom reservoir stack, blotting (nitrocellulose) membrane, gel, top reservoir stack and top cassette electrode. To transfer the protein to the nitrocellulose membrane, the manufacturer's protocol for Mini-PROTEAN TGX gels was followed.

2.4.5 – Blocking of nitrocellulose membrane

A blocking solution was made using 4 % w/v ECL Prime Blocking Agent (GE Healthcare Sciences) in PBS-T (PBS containing 0.1 % v/v Tween 20 [Sigma-Aldrich]). The membrane was incubated at room temperature in a dark box for 1 hour on a rocker using 50 mL of blocking buffer.

2.4.6 – Probing nitrocellulose membranes with antibodies

Primary antibodies were added at a dilution of 1:1,000 into the blocking solution and left overnight on a rocker at 4 °C. The membrane was then washed x 3 using PBS-T. Secondary antibodies were added at a dilution of 1:10,000 in 4 % blocking solution. Membranes were incubated for 1 hour before been washed x 3 using PBS-T. Membranes were then analysed on the Li-Cor image analyser (LI-COR). A list of antibodies used during this project can be found in table 2.1.

Table – 2.1. Antibodies and their appropriate dilutions

Primary Antibodies	Dilution	Use	Manufacturer
Mouse anti-Actin	1:5000	WB	Abcam (Cambridge, UK) (Cat #AB8226)
Rabbit anti-Argonaute 2	1:1000	WB	A gift from Dr. finn Grey (The Roslin Institute)
Rat anti-HA	1:1000	WB	Roche (Burgess hill, UK) (Cat #11-867-423-001)
6xHis	1:1000	WB	Roche (Cat #11-922-416-001)
tubulin	1:5000	WB	Abcam (Cat #AB18251)
Peroxidase Anti-Peroxidase Soluble Complex antibody	1:1000	CHEM	Sigma-Aldrich (Cat #P1291)
Secondary Antibodies	**Dilution**	**Use**	**Manufacturer**
IRDye 680RD donkey anti-mouse	1:10000	WB	Li-Cor (Cat #926-68072)
IRDye 800CW donkey anti-rabbit	1:10000	WB	Li-Cor (Cat #926-32213)
IRDye 680RD donkey anti-rat	1:10000	WB	Li-Cor (Cat #926-68076)

CHEM = Chemiluminescence, WB = western blotting.

2.4.7 – Detection of Protein by Chemiluminescence

Detection using chemiluminescence used a peroxidase anti-peroxidase antibody (listed in Table - 2.1) following the running, transferring and blocking of a nitrocellulose membrane (as laid out in sections 2.4.2 – 2.4.5). For detection, the Pierce™ ECL Western Blotting Substrate (ThermoFisher Scientific) was used as per the manufacturer's protocol. Blots were then analysed on a G-box imaging system (Syngene, Cambridge, UK).

2.5 – Production of a cell line stably expressing tagged Ago2

2.5.1 – Determining antibiotic resistance of cells

LMH cells were seeded at a density of $1x10^4$ per well on a 24 well plate and incubated overnight (section 2.1.2). Puromycin (Sigma-Aldrich) was added to the cells with dilutions ranging from 0 µg/mL to 10 µg/mL. The medium was replaced every 48 hours containing the appropriate concentration of antibiotic. Cell viability was checked every 48 hours using trypan blue staining and determining the percentage of dead cells, this was carried out for 11 days to determine the minimum concentration of antibiotic that caused complete cell death in 4-6 days.

2.5.2 – Production of Lentivirus particles

HEK293FT cells were seeded out for 70-90 % confluency in 100 mm plates using appropriate media containing no antibiotics or Geneticin. 24 hours after seeding out cells, media was replaced with 10 ml pre-warmed media. Lentiviral media contained 150 µl Optimem (Invitrogen), 4 µg Vector plasmid, 2 µg Envelope plasmid and 6 µg Packaging plasmid. Once mixed, it was left to stand for 5 minutes at room temperature. Transfection media was also made containing 150 µl Optimem (Invitrogen) and 36 µl Lipofectamine 2000 (Thermo Scientific) and left to stand for 5 minutes. After standing, both lentiviral and transfection media were mixed together and left to stand for 20 minutes at room temperature. The mixture was then added to the seeded out plates. Cells were then placed back into a 37 °C, 5 % CO_2 incubator. 16 hours post transfection, media was replaced with pre-warmed media containing appropriate supplements. Supernatant was harvested from cells 3 Days Post Infection (D.P.I) by scrapping cells from the plate and centrifuging them at 435 x g for 5 minutes. The resulting supernatant was then filtered using a 0.45 µm filter. At this point, supernatant was either aliquoted into 15 mL Falcon tubes and frozen at -80 °C or further concentrated.

2.5.3 – Concentration of Lentivirus particles by ultra-centrifugation

14 mL of supernatant was added to each 14 mL ultra-centrifuge tube (Beckmann Coulter, USA) (or sterile PBS for balancing). Tubes were weighed and balanced to within +/- 0.1 g before been placed into corresponding buckets (Beckmann SW40 Swing-out Rotor). The samples were loaded and then centrifuged for 2 hours at 19,500 RPM at 16 °C. Once finished, supernatant was removed and the pellet was resuspended in 200 µl sterile PBS.

2.5.4 – Transduction of LMH cells with Lentivirus particles

Cells were seeded out on a 24 well plate at a density $1x10^5$ per well and left overnight to adhere. Polybrene (Millipore) was added to concentrated virus at a concentration of 8 µg/mL and mixed by inversion. Virus supernatant was then added to cells at 100 µl supernatant per well. Cells were incubated at 38 °C for 30 minutes. Cells were then centrifuged for 2 hours at 625 x g at 30 °C. 18 hours after centrifugation, media was changed on the cells.

2.6 – Other Methods

2.6.1 – Infection of cells with ILTV

All cells listed in section 2.1.1 with the exception of HEK293T cells were seeded out at the appropriate density on a 24-well plate. Plates were placed back in the incubator for 24 hours to allow the cells to adhere.

Virus was retrieved from LN_2 and thawed rapidly using a 37 °C water bath. Cells were washed with sterile PBS before a virus suspension with an appropriate multiplicity of infection (MOI) was added. Plates were rocked manually every 15 minutes to ensure complete coverage of the monolayer. After 1 hour, the virus suspension was removed from cells and replaced with pre-warmed complete medium with 2 % FBS concentration.

2.6.2 – Plaque assays

LMH cells were seeded out at a density of $1x10^6$ per well in a six-well dish. Cells were then placed in the incubator and allowed to adhere prior to virus infection. Virus was recovered from LN_2 and thawed quickly in a 37 °C water bath. A 10 fold dilution series of virus in appropriate medium was made, media was removed from cell monolayers and 1 ml of virus solution was added per well. Plates were placed back into the incubator and rocked every 15 minutes for 1 hour. A 0.5 % avicell solution (FMC Biopolymer, Girvan, Scotland, UK) was made up using stock 2.4 % avicell solution and 2 % virus media with appropriate supplementary reagents (section 2.1.1 for media recipes). This was kept warm at 37 °C until needed. After the initial one hour incubation the inoculum was removed from cells. 2 ml of avicell overlay was added per well and plates were put back in a 37 °C incubator at 5 % CO_2. After 48 hours, plates were fixed by adding 2 ml of 10 % buffered formalin per well and placed in a fume cupboard overnight. Fixed plates had both formalin and avicell overlay removed and they were subsequently stained with 2 ml of 0.1 % toluidine blue. Plates were left for approximately 4 hours before the toluidine blue solution was removed. Plates were then washed using H_2O and allowed to dry.

2.6.3 – Luciferase assays

Luciferase assays were carried out using the Dual Luciferase Reporter Assay System (Promega). Cells were transfected as in section 2.1.6 and harvested using the 5x passive lysis buffer contained within the kit diluted to 1 x concentration. Lysed cells were left to rock for 30 minutes. After 30 minutes, cell lysates were transferred to a clean 1.5 ml eppendorf tube and centrifuged at 1800 RPM for 5 minutes. 20 µl of the sample supernatant was then added in triplicate to a white, opaque 96-well plate. Luciferase Assay Reagent (LAR II) and Stop & Glo reagent were then prepared according to the manufacturers protocol but they were used at a dilution of 1:10. A GloMax 96 luminometer (Promega) was used to measure the luminescence readings of both the Renilla and firefly luciferase. Measurements of luciferase used 50 µl of each reagent instead of the suggested 100 µl in the manufacturer's protocol.

2.6.4 – RNA Synthesis

Synthesis of RNA was carried out using the HiScribe™ T7 Quick High Yield RNA Synthesis Kit (New England Biolabs) according to the manufacturer's protocol. RNA Synthesis used 100 ng of template DNA and reactions were incubated for 24 hours. Synthesised RNA was then extracted and stored at – 80 °C.

2.6.5 – *In Vitro* digestion of DNA using Cas9 protein and sgRNAs

To synthesise both the template DNA and sgRNA DNA templates, conventional PCR was used as laid out in section 2.2.7. Both sgRNA DNA and template DNA was then purify using a PCR purification kit as per the manufacturer's instructions (Qiagen) (section 2.3.14). The DNA was then stored at – 20 °C until reactions were assembled.

sgRNA template DNA was used for RNA synthesis as per the manufacturer's protocol. Synthesised RNA was purified using an RNeasy mini kit (Qiagen). Extracted RNA concentrations were determined using a NanoDrop ND-100 Spectrophotometer (Thermo Scientific) and samples stored at – 80 °C.

Reactions were set up following the NEB protocol (M0386). Molar ratios of 10:10:1 were set up (sgRNA: Cas9 protein: Template DNA) in a final volume of 30 µl using nuclease free water. Samples were left for either 1 hour or 24 hours in a 37 °C incubator. Following incubation, samples were treated with RNase A and proteinase K to denature both the cas9 protein and sgRNA. Samples were then analysed on a 3 % agarose gel.

2.6.6 – CRISPR-Cas9 transfection and infections

LMH cells were seeded in 6-well plates at a density of $1x10^5$ cells per well and left for 24 hours to adhere in a 37 °C, 5 % CO_2 incubator. Cells were then transfected using Xfect polymer (Takara) according to the manufacturer's instructions. 5 µg of sgRNA (or if two were used 2.5 µg of each) was transfected alongside 5 µg of homology repair cassette totalling 10 µg per well. Transfection reactions were placed back in a 37 °C, 5 % CO_2 incubator and left for 12 hours.

After 12 hours incubation, cells were infected with wild type ILTV at an MOI of 0.001 as per section 2.6.1. Following incubation with virus, cells were washed with PBS and complete media was added to the cells (described in section 2.1.1) except with 2 % FBS. Phosphonoacetic acid (PAA) was added to the media at a concentration of 100 µg/ml. Cells were then placed back in 37 °C, 5 % CO_2 incubator for a further 12 hours.

Media was replaced again with complete media with 2 % FBS concentration and cells were placed in a 37 °C, 5 % CO_2 incubator for 48-72 hours and monitored for cell death and fluorescence using a fluorescent microscope.

Samples were harvested by scraping wells with a sterile scraper and well contents transferred to a clean 1.5 ml cryovial tube. Samples were then stored at – 80 °C.

2.6.7 – Statistical analysis of data

Statistical analysis was carried out using Minitab 17 software. Graphs were made using Graphpad Prism 6 software. All errors bars on figures represent the standard error of the mean unless otherwise stated. Specific statistical analysis for individual experiments is as stated in each corresponding section of this thesis.

Commonly Used Solutions

PBS

50 x TAE buffer

LB medium

LB/Agar

All solutions provided by the Central Services Unit at the Roslin Institute.

Chapter Three – Cross Linking and Sequencing of Hybrids (CLASH)

3.1 – Introduction

3.2 – Aims

3.3 – Detection of viral genes using cDNA

3.4 – Detection of viral miRNAs

3.5 – Production of a stably transduced cell line

3.6 – Assessment of other avian cell lines for susceptibility to ILTV

3.7 – Discussion

3.1 – Introduction

There have been several different types of experimental procedures developed to look at the interaction of miRNAs with their targets; each advancing on its predecessor (reviewed in section 1.6.4). Cross Linking and Sequencing of Hybrids (CLASH) is the latest of these biochemical techniques allowing for direct miRNA:mRNA interactions to be identified by ligating the two pieces of RNA (Helwak and Tollervey, 2014). The procedure is reviewed in section 1.6.4 and an outline of the procedure is shown in Figure 3.1. With the use of these techniques, targets of virus-encoded miRNAs can be elucidated.

The role of virus-encoded miRNAs in the pathogenesis of disease is well documented in the literature; especially in the case of herpesviruses (reviewed in section 1.7). Marek's Disease Virus (MDV) encodes for the miRNA MDV1-miR-M4 which is a functional homologue of the cellular miRNA miR-155. Abolition of the miRNA from the virus abrogates the formation of lymphomas during lytic infection (Zhao et al., 2011). Similarly, Kaposi's sarcoma associated Herpesvirus (KSHV) has been shown to encode a miR-155 homologue that is implicated in the induction of B cell tumours (Gottwein et al., 2007). Human Cytomegalovirus (HCMV) encodes the HCMV-miR-US25.1 that has been shown to target the cellular gene ATP6V0C, a gene essential for viral replication. This targeting of a gene reducing pathogenesis is different to what is seen in other viruses but it suggests that ATP6V0C regulation may be important for immune invasion or maintenance of latency (Pavelin et al., 2013). Epstein-Barr virus (EBV) is known to encode for 44 mature miRNAs found in three distinct clusters in the genome. One virus-encoded miRNA, EBV-miR-BART6-5p has been shown to target Dicer, an essential component in the biogenesis of miRNAs. It also has the negative consequence that it down-regulates the effect of EBV-encoded miRNAs so may play a critical role in regulating the expression of miRNAs (Lizasa et al., 2010).

ILTV has been shown to encode for 10 miRNAs (Rachamadugu et al., 2009, Waidner et al., 2009). One of these, ILTV-miR-I6 was found to target ICP4. A second miRNA, ILTV-miR-I5 was also suggested to target ICP4 but this was not statistically significant (Waidner et al., 2011). As there is little information on the remaining miRNAs and no information on possible cellular targets of ILTV-encoded miRNAs, the use of biochemical techniques to identify targets of ILTV-encoded miRNAs would advance the understanding of pathogenesis for this virus. It is important to elucidate these targets as a deeper understanding of pathobiology of the virus will inform vaccine strategy and/or breeding of birds for resistance to ILTV.

As well as using biochemical approaches to identify targets for virus-encoded miRNAs, the use of alignment software can be employed to assess homology between ILTV-miRNAs and any cellular miRNAs. This approach was previously used by the Dalziel group whereby the OvHV-2-encoded miRNA OvHV2-miR-73.1 was found to be homologous to the cellular miRNA miR-216a (Levy, 2011).

Finally, as discussed in chapters 4 and 5, bioinformatical approaches can be used to identify targets of both viral and cellular encoded miRNAs which can be then tested using more traditional laboratory techniques.

3.2 – Aims

The aims of this part of the project were to create a cell line that stably expressed a tagged Argonaute 2 transgene for subsequent use in CLASH experiments. In addition to this, infection studies with LMH cells were carried out to elucidate an optimal time point for when to harvest infected cells for CLASH by assessing the expression of both viral genes and viral miRNAs across a time course of infection. Other avian cell lines were also screened for their permissiveness to infection by ILTV.

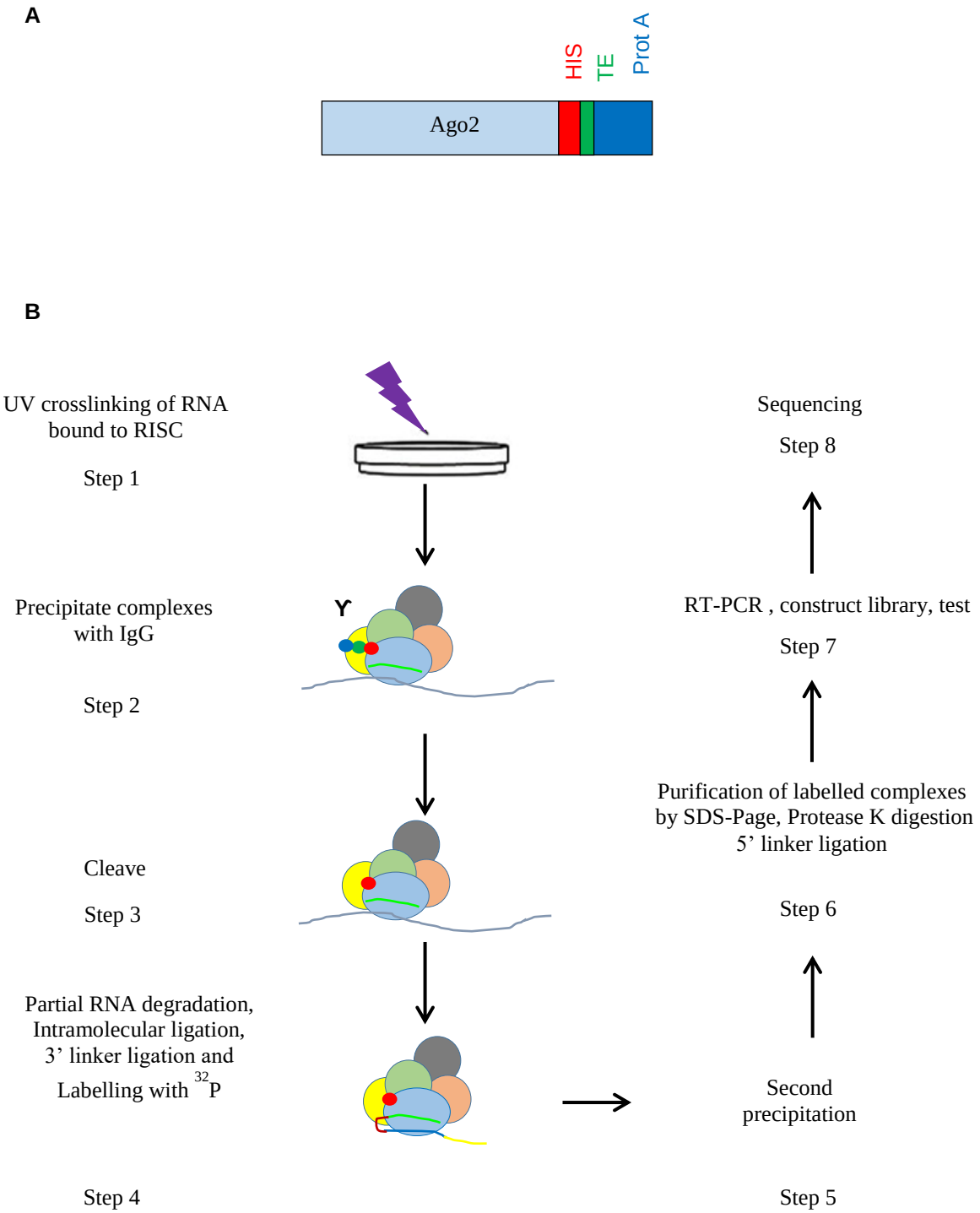


Figure 3.1 - An Overview of the CLASH Technique

A – Schematic diagram showing the His6x, Tobacco Etch Virus (TEV) and Protein A (HTP) fused to Argonaute 2 (Ago2) protein which is used for immunoprecipitation of the Ago2.
B – Brief outline of steps carried out during the CLASH experimental protocol

3.3 – Detection of viral gene expression by Reverse –Transcriptase PCR

Viral gene expression was determined as part of the optimisation for the CLASH experiment, essentially to assist in picking an optimal time to harvest samples for the CLASH protocol. LMH cells were either infected or mock-infected as described in section 2.6.1. Cells were harvested and RNA extracted. RNA was DNase treated and was subsequently used for cDNA synthesis as described in section 2.2.5. Conventional PCR (see section 2.2.7 for details) was then used to amplify specific viral genes and PCR products analysed by gel electrophoresis (as per section 2.2.12). PCR products that were detected in the infected cDNA samples and not the mock infected or –RTs were taken as a positive result. Genes could be detected from across the temporal gene expression profile. One immediate-early designated gene, ICP4 was detected in a time dependent manner (Figure – 3.2A). Two early genes, Protein kinase (PK) and Thymidine Kinase (TK) were detected also (Figure – 3.2B & C). Furthermore, one early/late gene, UL41 was detected (Figure – 3.2D). Finally, a late gene, glycoprotein E (gE) could be detected (Figure – 3.2E).

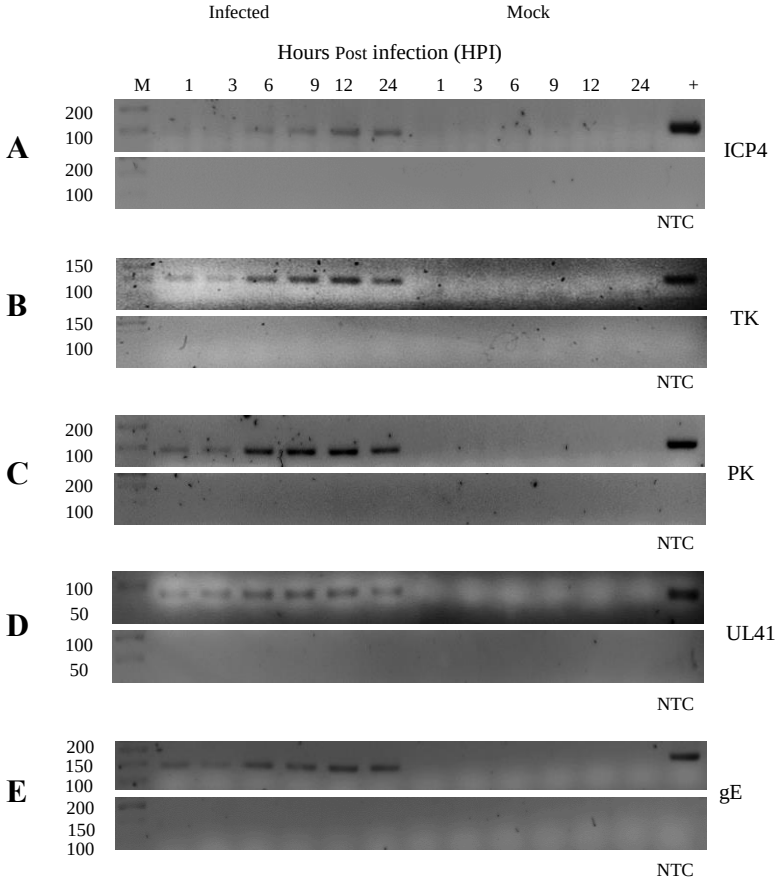


Figure – 3.2. Detection of expression of viral genes in a time dependent manner using RT-PCR

LMH Cells were infected with ILTV at an MOI of 0.1. Samples were harvested at 1, 3, 6, 9, 12 and 24 hours post infection (HPI). RNA was extracted and DNase treated. cDNA was synthesised and used for detection for the detection of viral genes. A gDNA positive control was used as well as minus reverse transcriptase (-RT) controls. Top panel in each case displays cDNA results alongside a gDNA control whilst lower panel shows –RT controls and a NTC. A – ICP4, an immediate-early gene B – Thymidine Kinase (TK), an early gene C – Protein Kinase (US3/PK), an early gene D – UL41, a delayed-early gene E – Glycoprotein E (gE), a late gene

3.4 – Detection of viral miRNAs

The expression of virus-encoded miRNAs was also investigated using the same RNA used for the detection of viral genes. This was again done as an optimisation step in preparation for the CLASH protocol to maximise the number of miRNAs expressed thus garnering the maximum amount of data possible for one time point. RNA was DNase treated and the MiScript II kit was used to amplify mature miRNAs. The SYBR green real-time PCR kit (Qiagen) was then used for mature miRNA detection along with a forward primer specific to each ILTV-encoded miRNA (as shown in Figure 3.3). Samples were then analysed by agarose gel electrophoresis. Bands detected were then cloned and sequenced. Due to the process of cDNA synthesis with the MiScript II kit, primarily the polyadenylation step and use of an oligo-DT prime with a 3' universal tag, genomic DNA is not detected during PCR and so –RTs are not applicable. Therefore an uninfected sample was used as a control for each time point for each miRNA in addition to the detection of a cellular miRNA. According to the MiScript PCR handbook, mature miRNA PCR products are 85-87 nts in length.

In total, 7 of the 10 virally encoded miRNAs could be detected with this method (Figure 3.4). ILTV-miR-I1-3p could be detected at 6, 9, 12 and 24 HPI but not before whilst it's complimentary strand, ILTV-miR-I1-5p could be found at 1, 6, 9, 12 and 24 HPI. Only at one time point could ILTV-miR-I4 be detected which was 1 HPI. ILTV-miR-I5-3p has the same expression profile as ILTV-miR-I1-5p, detected at all times except 3 HPI. Finally, three miRNAs could be detected at all time points and they were ILTV-miR-I5-5p, ILTV-miR-I6-3p and ILTV-miR-I6-5p. In addition, gga-miR92a-3p could be found at all time points including the mock controls. Some other bands were detected at around 50 bp. These are most likely off target effects of the primers which are constrained due to the length of miRNA thus restricting primer design.

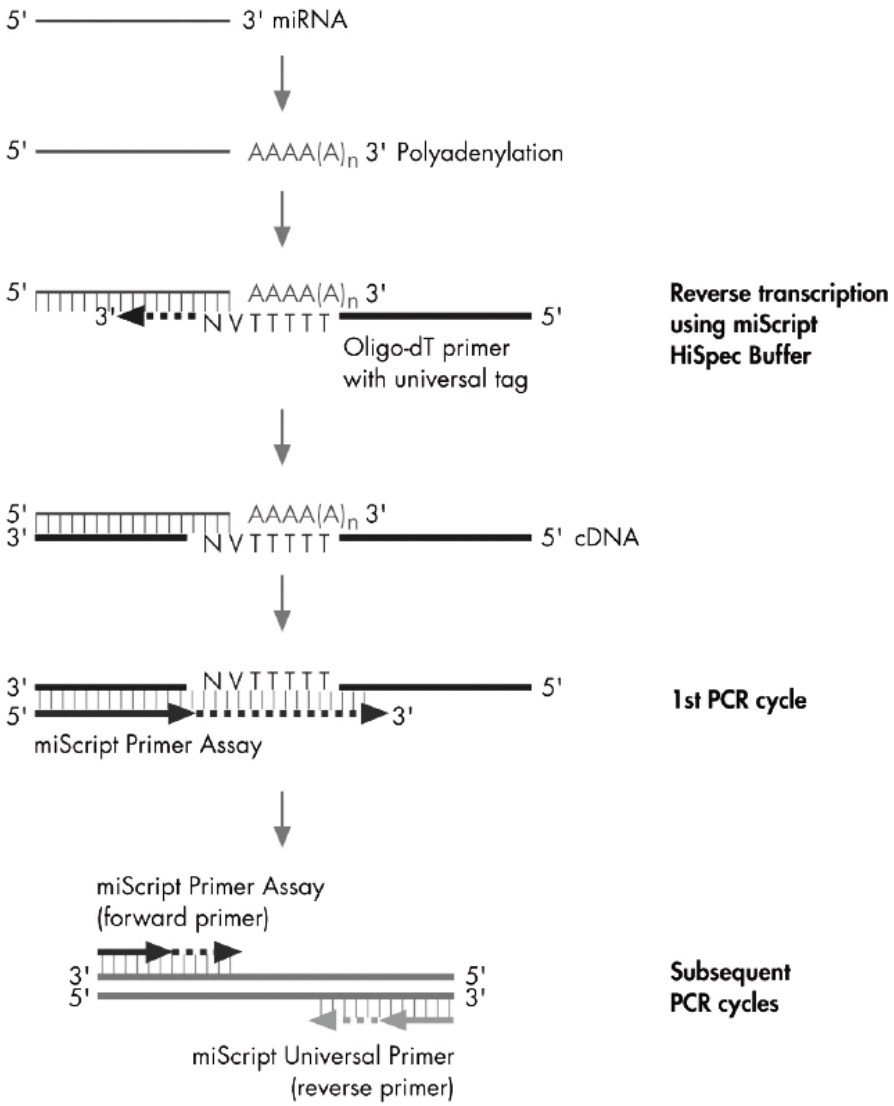


Figure – 3.3. Reverse Transcription and amplification of miRNAs
For detection of miRNAs, cDNA was synthesised using the miScript II RT Kit (Qiagen). This was used in conjunction with the miScript SYBR® Green PCR Kit (also Qiagen) to detect specific miRNAs using a specific sense primer lacking the last 6 nts. Sequencing was then used to confirm the miRNA sequence.

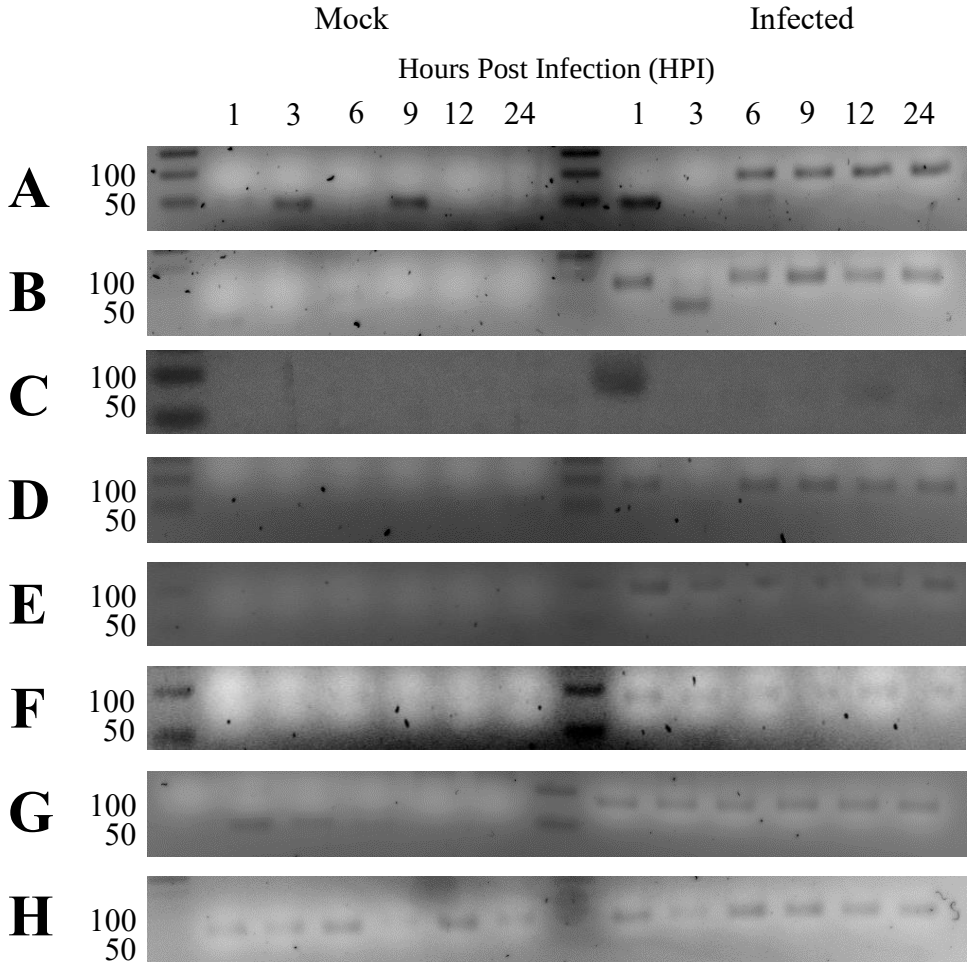


Figure – 3.4. Time course of Expression for Viral miRNAs
LMH cells were seeded out at 1x104 per well in a 24-well plate and left to adhere overnight. Cells were then infected with an MOI of 0.1 and subsequently harvested at 1, 3, 6, 9, 12 and 24 hours post infection (HPI) by removing the media, scrapping the cells and suspending them in 350 µl of RLT buffer. From this, RNA was extracted and DNase treated. cDNA was synthesised using the MiScript II kit and subsequently used in a Real-time PCR for detection of miRNAs with the MiScript SYBR Green PCR Kit (Both Qiagen). 7/10 miRNAs have so far been detected using this method as well as the cellular miRNA gga-miR-92-3p. The miRNAs detected are listed below:

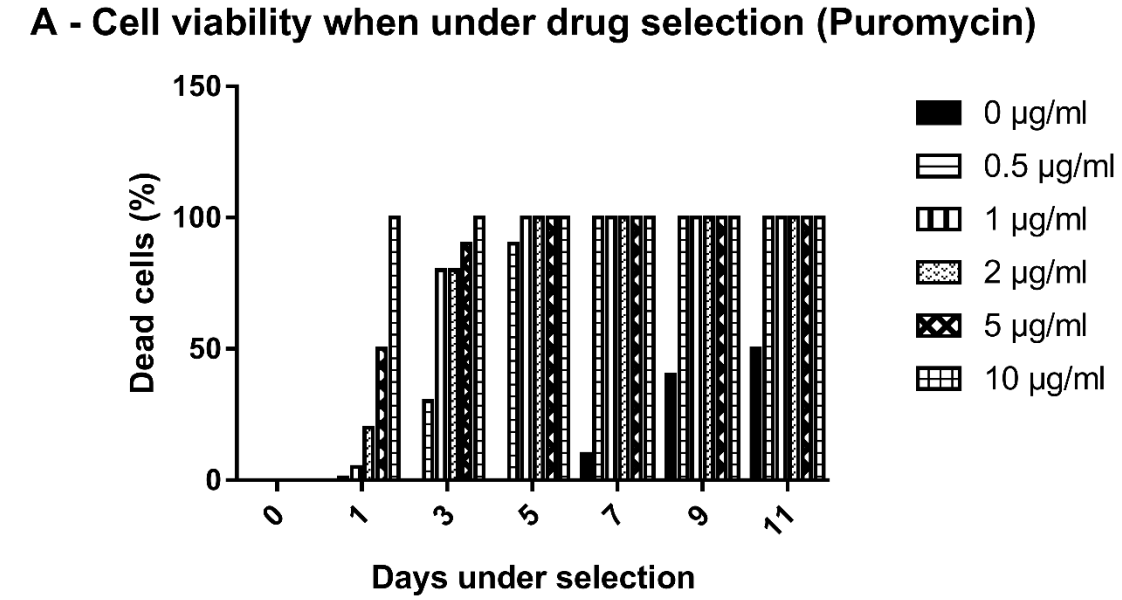
A. ILTV-miR-I1-3p
B. ILTV-miR-I1-5p
C. ILTV-miR-I4
D. ILTV-miR-I5-3p
E. ILTV-miR-I5-5p
F. ILTV-miR-I6-3p
G. ILTV-miR-I6-5p
H. gga-miR-92-3p

3.5 – Transduction of LMH cells with lentivirus particles

For the CLASH experiment (shown in Fig – 3.1.), purification of RISC complexes is required. To do this, a tagged Ago2 transgene is used. For delivery of the tagged transgene, lentiviruses are used which harbour the desired transgene and can intergrate with the host genome allowing for the creation of a stably transduced cell line. To help create a pure population of transduced cells, a selection marker is used to remove any untransduced cells from the population.

3.5.1 – LMH susceptibility to the drug Puromycin

To help create a clonal population of transduced cells expressing a transgene, a selection marker, in this case, Puromycin was used to kill any untransduced cells following transduction. To determine the death rate of LMH cells in the presence of the drug Puromycin, cells were seeded out at a density of $1x10^4$ cells per well in a 24-well plate. In total, 3 wells were seeded per drug concentration and the average cell death was calculated for each condition. After 24 hours, cells were placed under antibiotic selection pressure at differing concentrations to find an optimum concentration where complete cell death was observed in 3-5 days. Cell death was assessed by trypan blue staining and estimating the percentage of dead cells. Experiment one used a broad range of antibiotic concentrations (0 – 10 µg/ml) and found an optimal range somewhere between 0.5 and 1 µg/ml (Figure – 3.5A). Subsequently, a second experiment was set up with a concentration range between 0 and 1 µg/Ml. The result of this second experiment found that 0.5 µg/ml Puromycin was optimal for complete LMH cell death within the 3 – 5 day range (Figure – 3.5B).



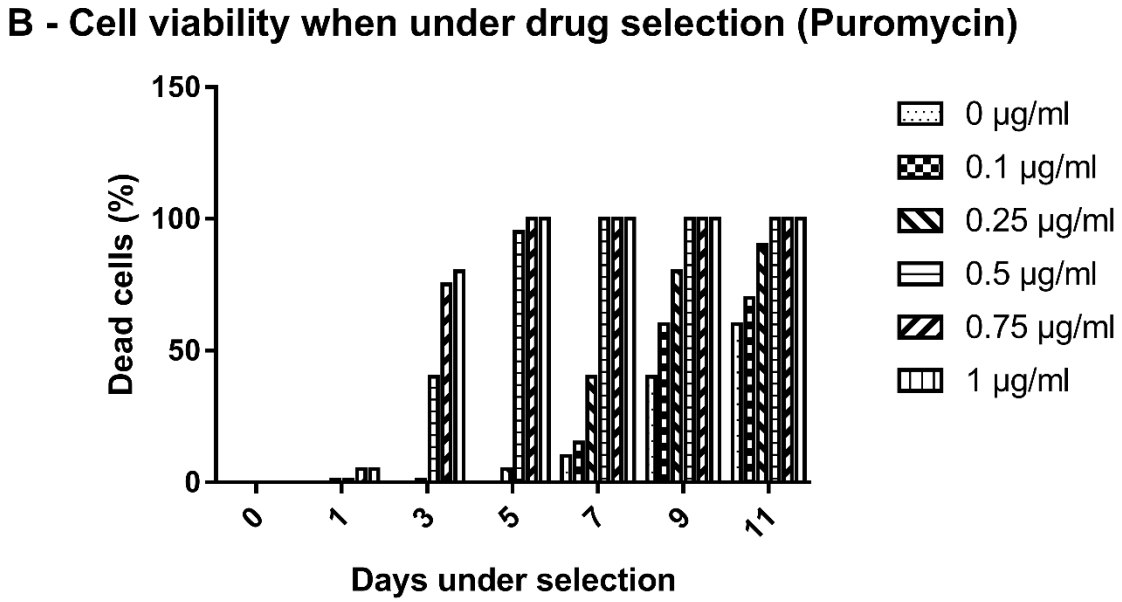


Figure – 3.5. LMH cell viability whilst under selection pressure using Puromycin
LMH cells were seeded out in 24 wells plates and left to adhere. Cells were then placed under drug selection pressure with Puromycin. Dead cells were counted using trypan blue staining which gave an estimated percentage of dead cells (%). A – A large concentration range of Puromycin (0 – 10 µg/ml) used. B – A narrow drug concentration range (0 – 1 µg/ml)

3.5.2 – Transduction of LMH cells with lentivirus particles

Lentivirus particles were produced (diagrammatic flowchart in Figure – 3.6) via transfection of three plasmids into HEK293FT cells (A gift from Dr. N Smith, The Roslin Institute) and harvested as per section 2.5.2. Lentivirus plasmids (A gift from Dr. F Grey, The Roslin Institute) were part of the 3rd generation of lentivirus generation plasmids. A transfer plasmid containing a tagged Ago2 fusion gene was used in conjunction with the packaging plasmid psPAX2 and envelop plasmid pMD2.G. Following transfection of cells, the supernatant was subsequently concentrated by ultracentrifugation before the virus pellet was resuspended in 200 µl PBS. The resulting concentrated viral supernatant was spun on to cells (See section 2.5.4). After approximately 18 hours, transduced LMH cells were placed under antibiotic selection pressure with Puromycin as determined in 3.5.1 and cells were allowed to proliferate. After three passages under selection pressure, a portion of the cells were harvested for testing by chemiluminescence to detect the transgene (tagged Ago2 protein). Results from this were inconclusive. Initial chemiluminescent blots showed faint banding at the correct predicted size for the tagged transgene (~100 kDa) as indicated by the red asterisks (Figure – 3.7A). Upon successive passage of cells and re-testing of the cells for the transgene, no band could be detected at the correct size despite them still been under antibiotic selection pressure (Figure – 3.7B & C).

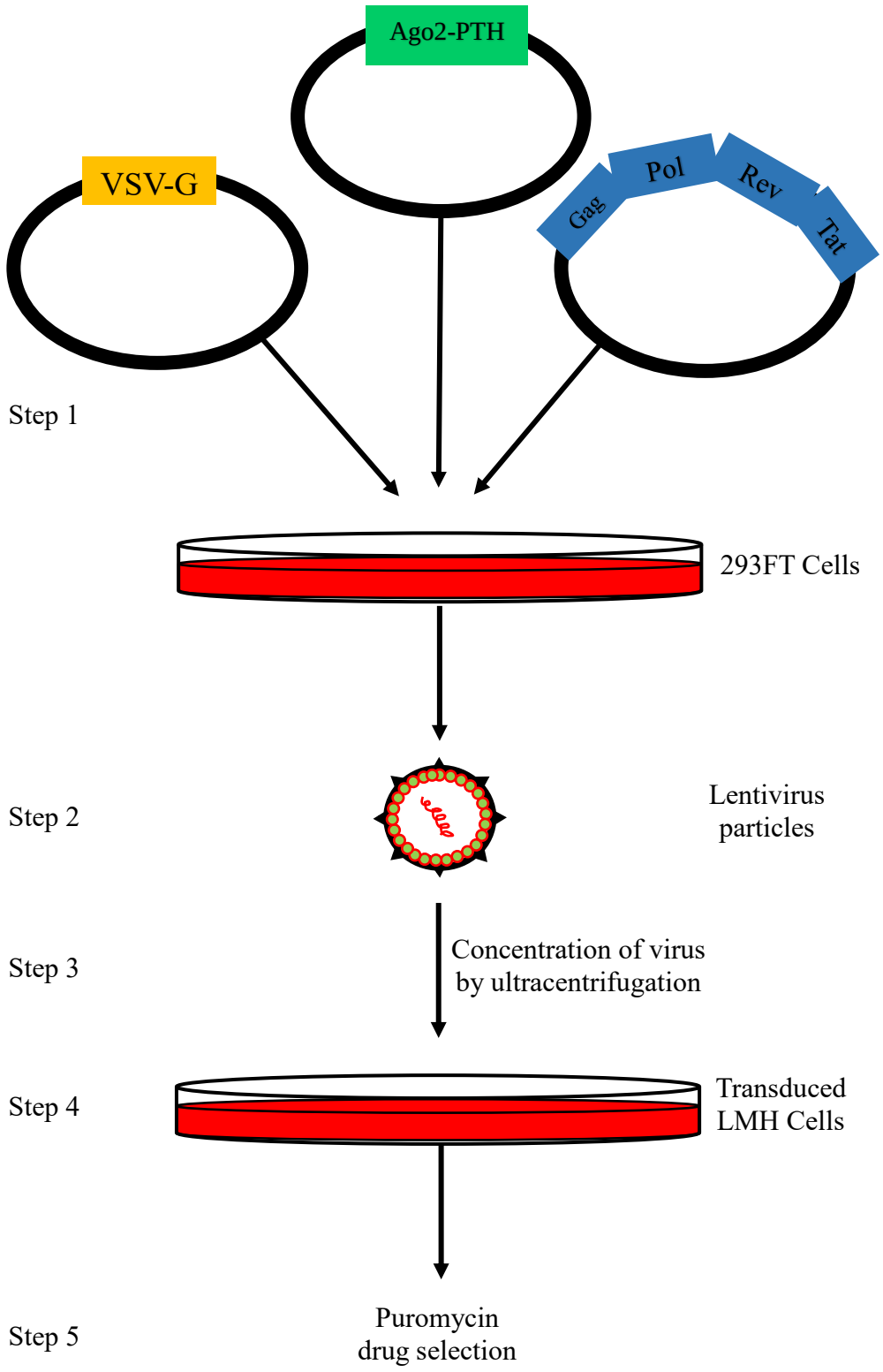


Figure – 3.6 The production of Lentivirus particles for delivery of an Ago 2 Transgene
PLVX-tight-puro lenti vector (pLVX-puro) which contained HTP tagged human Ago2 (a gift from Dr F. Grey, The Roslin Institute) was transfected into HEK293FT cells (a gift from Dr. N Smith, The Roslin Institute). Resulting Ago2 lentivirus was harvested and concentrated by ultracentrifugation (19500 rpm for 2 hours at 16 °C in a Beckman SW40 rotor). Virus pellets were resuspended in an appropriate amount of PBS. LMH cells were then transduced via spinning virus onto cells are per section 2.5.4. After 24 hours, cells were put under selection pressure to obtain cells carrying the transgene.

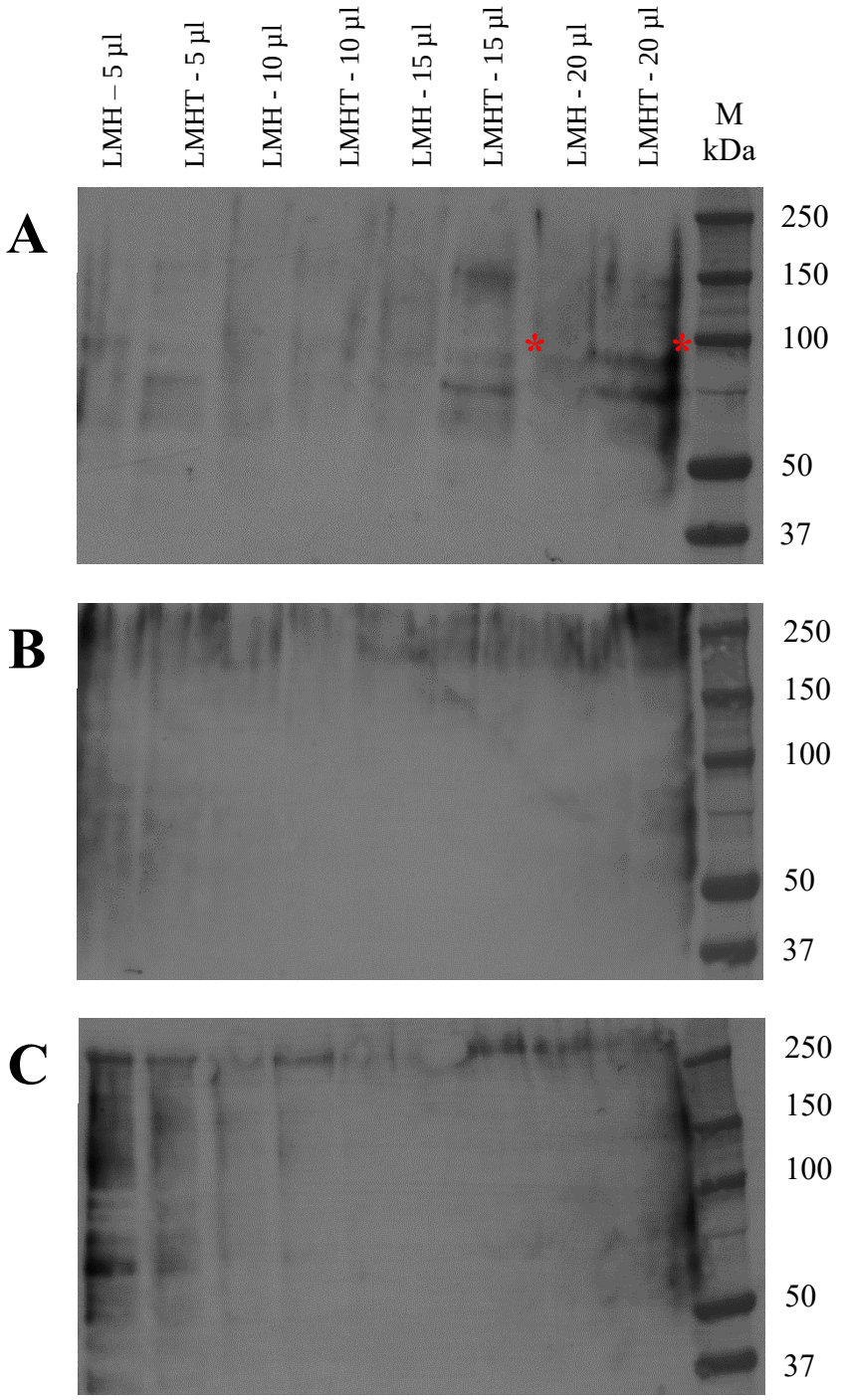


Figure – 3.7. Detection of tagged Ago2 protein in transduced LMH cells
LMH cells were transduced with concentrated lentivirus particles carrying a tagged Ago2 transgene. After transduction, cells were placed under antibiotic selection pressure with Puromycin. After 3 days, cells were tested for detection of the Ago2 protein by chemiluminescence. A – Cells tested 3 days after antibiotic selection pressure. B – Cells passaged and kept under selection pressure. C – A second passage of cells still under selection pressure. Red asterisks – Possible detection of expression of tagged Ago2.

3.6 – Assessment of other avian cells lines for susceptibility to ILTV

Whilst LMHs are able to support virus replication and cytopathic effect can be observed (CPE), there was interest in looking at other avian cell lines as LMHs are difficult to work with (discussed in section 3.7). Chicken cell lines were chosen as ILTV is predominatly a chicken virus however it does infect other species. A more suitable chicken cell line was sought and the panel test below have previously been used for other virsues. In total, four immortal chicken cell lines (LMH, DF1, HD11 and CLEC213 cells) and two primary chicken cell lines (chicken kidney cells and primordial chicken germ cells) were tested for their permissiveness to infection with ILTV, two immortalised quail cell lines (QT-35 and QM7 cells) were also used. Initial infections were set up as described in 2.6.1 with an MOI of 0.1 and cells were harvested for DNA at 1, 6, 24, 48 & 72 hours post infection. Samples were then used for real-time quantitative PCR with primers designed against ILTV genomic DNA as previously published (Mahmoudian et al., 2012). To calculate the fold change, the expression of ILT gDNA was normalised to the cellular gene cyclophilin (CyP) using the 2ΔΔCT protocol (Livak and Schmittgen, 2001b). All samples were carried out in biological triplicate (n = 3). One-way ANOVA was used to compare fold change over time. Tukey's comparison test was then used to determine if the means of each time point were significantly different from one and other. Means that are significantly different from one and other have different letters.

3.6.1 – Chicken cell lines

LMH cells were found to have an increasing viral burden as time progressed with a peak fold change of ~6,400 compared to CyP at 72 hours post infection (HPI). However this was not statistically significant from the input virus or at any other time point during the experiment ($p => 0.05$) (Figure – 3.8A). Similar to published work, CPE was observed when infected cells were assessed using a light microscope.

Primary chicken kidney cells displayed a slight dip in viral burden 6 HPI (fold change = 752) compared with the input virus (fold change = 1087) however this was not statistically significant ($p => 0.05$). At 24 HPI viral burden peaked (fold change = 7987) before decreasing once again at 48 HPI (fold change = 3960). There was no statistical significance between the final two time points ($p => 0.05$) though there was between 24 HPI and the earlier two time points ($p =< 0.01$) (Figure – 3.8B). CPE was seen when cells were assessed visually before harvesting.

DF1 cells showed a decrease in viral burden over time with input virus at 1 HPI having the highest fold change of 191.39 compared with 13.84 at 48 HPI Statistically, each time point

was significant from the other when using the Tukey's comparison test (P =<0.05) (Figure – 3.8C). No CPE could be seen cells were visually checked prior to harvest at each time point.

CLEC213 cells (A gift from Dr. Pascale Quéré, INRA, France) also showed a decrease in viral load over time. No significant difference was found between 1 and 6 HPI (p =>0.05) but both time points were significantly different from the remaining three time points (p =<0.01) (Figure -3.7D). No CPE could be seen when cells were assessed using a light microscope.

The final chicken cell line, HD11 cells also showed a decrease in viral burden. At 1 HPI the highest fold change was observed (fold change = 383.8) which was statistically significant from both 24 and 72 HPI which had fold changes of 101.9 and 117.5 respectively. From 6 HPI onwards however, there was no significant difference observed between the time points (p =>0.05) (Figure – 3.8E). No CPE was seen when cells were observed using a light microscope prior to harvesting.

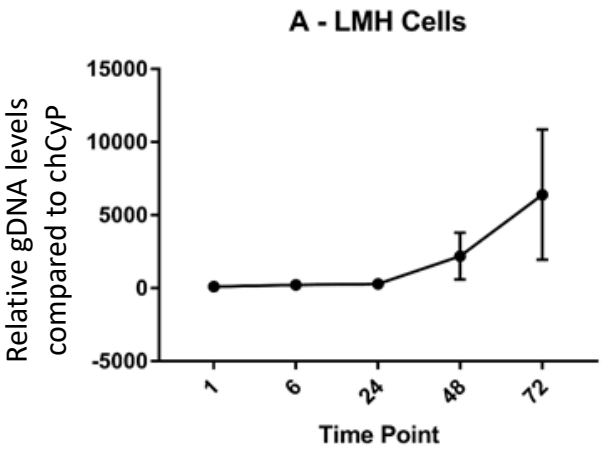


Table 3.1A – Fold change of ILTV gDNA in LMH cells

Time point	N	Mean	Tukey's Grouping
1	3	104	A
6	3	217.32	A
24	3	293	A
48	3	2193	A
72	3	6396	A

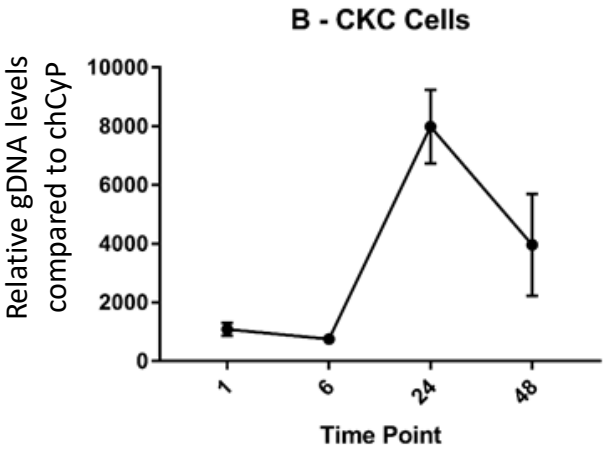


Table 3.1B – Fold change of ILTV gDNA in Primary CK cells

Time point	N	Mean	Tukey's Grouping
1	3	1087	A
6	3	752	A
24	3	7987	B
48	3	3960	A B

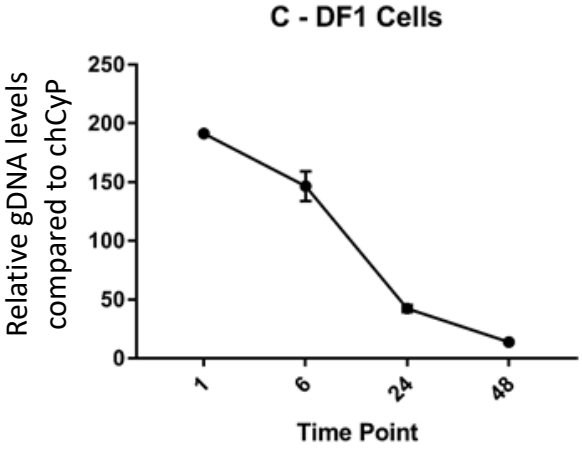


Table 3.1C – Fold change of ILTV gDNA in DF1 cells

Time point	N	Mean	Tukey's Grouping
1	3	191.39	A
6	3	146.5	B
24	3	42.39	C
48	3	13.84	D

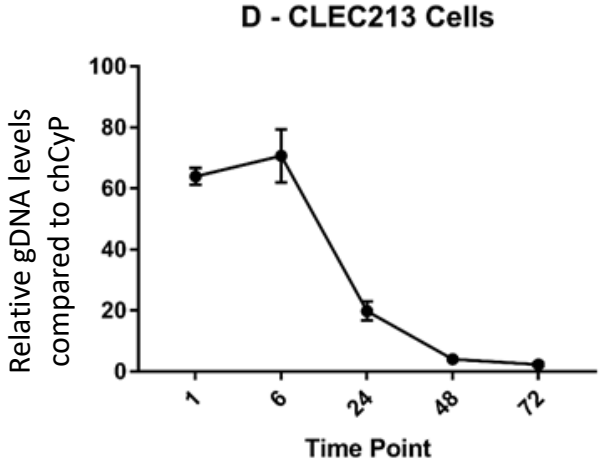


Table 3.1D - Fold change of ILTV gDNA in CLEC213 cells

Time point	N	Mean	Tukey's Grouping
1	3	63.97	A
6	3	70.68	A
24	3	19.81	B
48	3	4.065	B
72	3	2.309	B

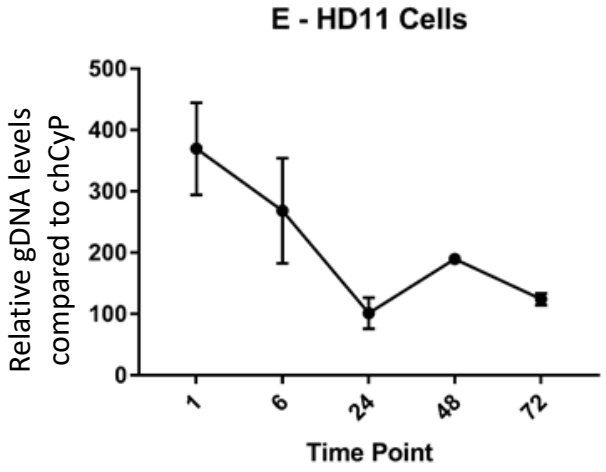


Table 3.1E - Fold change of ILTV gDNA in HD11 cells

Time point	N	Mean	Tukey's Grouping
1	3	383.8	A
6	3	261	A B
24	3	101.9	B
48	3	190.76	A B
72	3	117.5	B

Figure – 3.8. Relative ILTV genomic DNA (gDNA) levels in different Chicken cell lines
Cells were seeded out as per their individual requirements in a 24 well plates. They were left to adhere overnight before infection with ILTV at an MOI = 0.1. Samples were then taken at 1, 6, 24, 48 and 72 hours post infection (HPI). DNA was extracted and subsequently used for qPCR to determine the relative viral load. Data was analysed using the 2ΔΔCT method and normalised to Chicken CyP (chCyP). Data displayed shows the relative fold change of gDNA compared to Chicken Cyp. Error bars represent the standard error of the mean. A – LMH Cells B – Primary Chicken Kidney Cells (CKCs) C – DF-1 cells D – CLEC213 Cells E – HD11 Cells

3.6.2 – Quail cell lines

In addition to 5 chicken cell lines, 2 quail cell lines were also tested for their ability to support ILTV productive infection. Infection studies were set as previously described in section 3.6.

QT-35 cells showed a decrease in viral burden over time with the highest fold change seen at 1 HPI (fold change = 218.1) which was significantly different from every other time point ($p =<0.01$). At 6, 24 and 48 HPI there was no significant difference in fold change even though a decrease was observed ($p =>0.05$). Similarly, there was no significant difference between 24, 48 and 72 HPI ($p =>0.05$) but there was significance between 6 and 72 HPI ($p =<0.05$) (Figure – 3.9A). No CPE was observed when cells were assessed prior to harvesting.

The second quail cell line, QM7s also displayed a decrease in ILTV viral load over time. Statistical analysis revealed no significant difference between 1 and 6 HPI ($p =>0.05$). Likewise, there was no significant difference seen between 24, 48 and 72 HPI ($p =>0.05$). There was however a significant difference between the two Tukey's test groupings ($p =<0.01$) (Figure - 3.9B). Similar to QT-35's, no CPE was seen when cells were assessed using a light microscope.

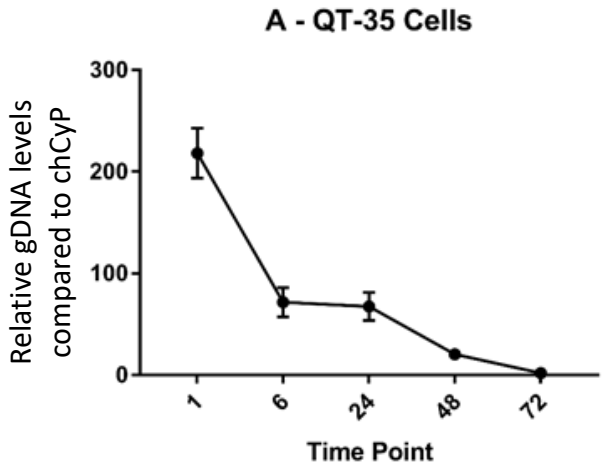


Table 3.2A - Fold change of ILTV gDNA in QT-35 cells

Time point	N	Mean	Tukey's Grouping
1	3	218.1	A
6	3	71.5	B
24	3	67.5	B C
48	3	20.43	B C
72	3	1.752	C

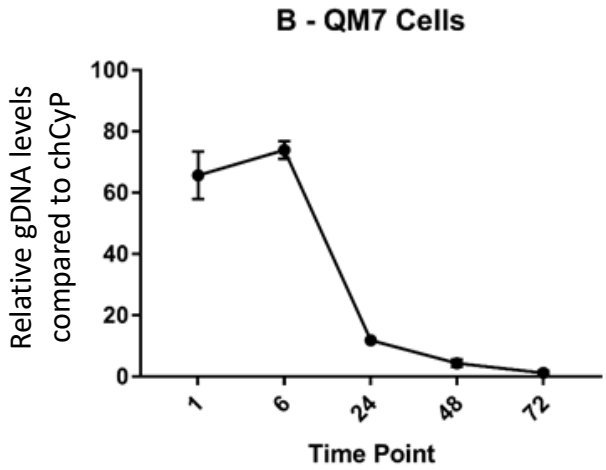


Table 3.2B - Fold change of ILTV gDNA in QM7 cells

Time point	N	Mean	Tukey's Grouping
1	3	65.68	A
6	3	73.97	A
24	3	11.851	B
48	3	4.41	B
72	3	1.218	B

Figure – 3.9. Relative ILTV genomic DNA (gDNA) levels in two Quail cell lines
Quail cell lines were seeded out as per their requirements. After 24 hours incubation, cells were infected with ILTV at an MOI = 0.1. Samples were then taken at 1, 6, 24, 48 and 72 hours post infection (HPI). DNA was extracted and used for qPCR to assess viral burden in the cells. Data was analysed using the 2ΔΔCT method against the endogenous gene Chicken CyP. Figures display the average relative fold change of ILTV gDNA. Error bars represent the standard error of the mean. A – QT-35 cells B – QM-7 cells

3.6.3 – Primordial Germ Cells (PGCs) viral burden

Primordial germ cells (a gift from Dr. Mike McGrew, The Roslin Institute) were also assessed for their susceptibility to ILTV infection. Experiments were set up as described in section 6.7 except virus was spun onto cells via centrifugation as described in section 2.6.1. Viral burden was seen to fluctuate with an initial decrease from 1 to 24 HPI which was statistically significant ($p =<0.01$). There was no significant difference between 1 and 6 HPI as well as between 6 and 24 HPI ($p =>0.05$). However, at 48 HPI an increase in viral load was observed compared to 24 HPI which was statistically significant ($p =<0.01$). At the final time point (96 HPI), a modest increase in fold change was seen compared to 48 HPI (fold change = 2488.6 compared to fold change = 2067) but this was not significant ($p =>0.05$) (Figure – 3.10). As PGCs are a suspension cell line, CPE observations were not apparent.

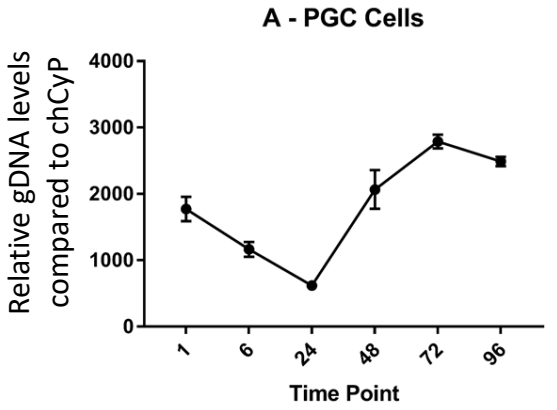


Table 3.3 - Fold change of ILTV gDNA in PG cells

Time point	N	Mean	Tukey's Grouping
1	3	1770	A B
6	3	1163	B C
24	3	618.89	C
48	3	2067	A D
72	3	2788	D
96	3	2488.6	A D

Figure – 3.10. Relative ILTV genomic DNA (gDNA) levels in chicken Primordial Germ Cells (PGCs)

Primordial Germ Cells (PGCs) (A gift from Dr. Mike McGrew, The Roslin Institute), were counted and seeded into 24 well plates with appropriate media. Cells were then spinoculated with ILTV virus at an MOI = 0.1 (details on spinoculation can be found in section 2.6.1 of this thesis). Samples were then harvested at 1, 6, 24, 48, 72 and 96 hours post infection (HPI). DNA was extracted and used for qPCR. Relative gDNA levels of ILTV were calculated using the 2ΔΔCT method of analysis with Chicken CyP gene to compare against. Figure displays the average relative fold change of ILTV gDNA. Error bars represent the standard error of the mean.

3.6.4 – Viral gene expression in primordial germ cells (PGCs)

In addition to assessing the susceptibility of PGCs to ILTV, viral gene expression was also assessed. Cells were infected as previously described (See Section 2.6.1) and at set time points, RNA was harvested. Following RNA extraction and DNase treatment, cDNA was prepared and used in a conventional PCR (See sections 2.2.5 & 2.2.7 for protocols). Bands detected in the cDNA 'infected' samples and not the mock cDNA or –RT samples were taken as positives. Three classes of temporal gene expression were looked at. Firstly, ICP4, representing an immediate-early gene was detected as early as 1 HPI and through until 96 HPI but the bands were not as strong as earlier time points (Figure - 3.11A). Protein Kinase (US3/PK), an early gene was also detected at every time point (Figure – 3.11B). Finally, glycoprotein E (gE), a late gene was also found to be expressed in PGCs (Figure – 3.11C).

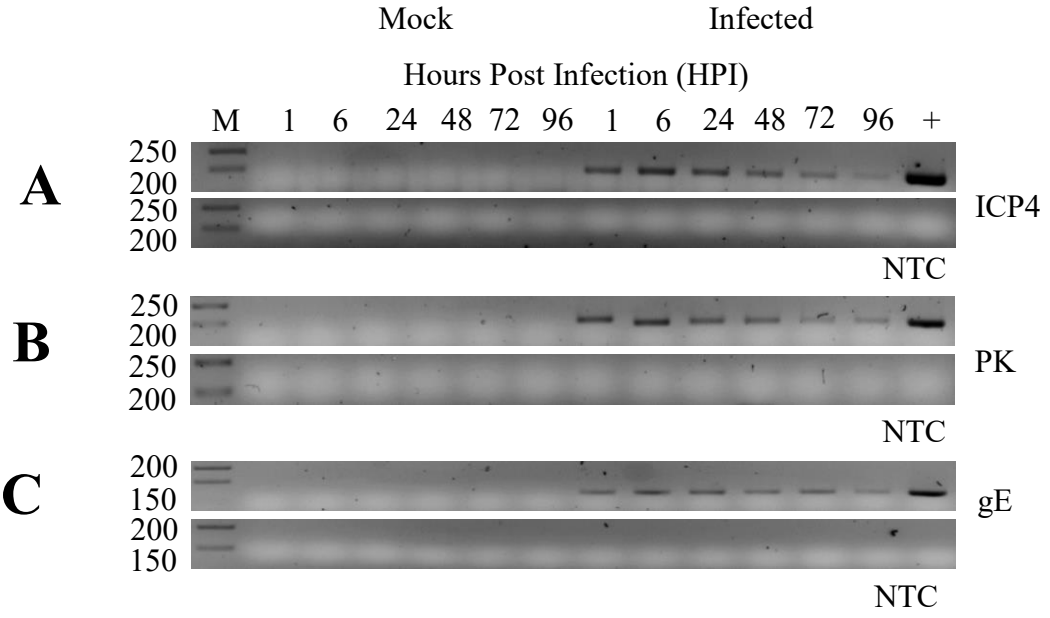


Figure – 3.11. Viral gene expression of ILTV in primordial germ cells (PGCs)
PGCs (A gift from Dr. Mike McGrew, The Roslin Institute) were infected with ILTV at an MOI of 0.1 as well as a parallel set of mock controls. At set time points (1, 6, 24, 48, 72 & 96 HPI), cell lysate was harvested for RNA. RNA was extracted using a RNeasy kit (Qiagen) and subsequently DNase treated. RNA was used to synthesise cDNA with RNA primed using an Oligo DT primer. PCR reactions were carried out on the resulting cDNA. Reactions included a gDNA loading control alongside –RT samples a NTC. PCR products were then visualised on a 2.5 % agarose gel. Bands detected in the infected samples and not in the mock controls or –RT samples were taken as positive results. In total, three genes across the temporal gene expression cascade were tested for and they were; ICP4 (A), protein kinase (B) and glycoprotein E (C).

3.7 – Discussion

The original aims of this part of the project were to use CLASH (as reviewed in Chapter one) to identify targets of miRNAs encoded by ILTV. However, due to unforeseen circumstances, this project was shelved and reasons for this are outlined in this section. Work leading up to CLASH showed the expression of five viral genes that spanned the temporal cascade of gene expression across a range of time points. As well as identifying viral genes, seven out of a possible ten virally-encoded miRNAs were identified. Moreover, attempts were made to engineer a stably expressing cell line using a lentiviral delivery system. Infection work in another 7 cell lines in addition to LMH cells was also carried out to identify a cell line that supported lytic infection that would be suitable for the CLASH experiment.

CLASH is a technically difficult but extremely powerful biochemical approach to identifying miRNA-mRNA interactions in the cell. It builds upon previous forms of identifying miRNA interactions within a cell such as cross-linking and immunoprecipitation (CLIP) and also cross-linking and analysis of cDNAs (CRAC) (Kudla et al., 2011). The exploitation of the technique could have potentially yielded a large quantity of data from which downstream validation work could have been carried out. Due to technical challenges in the optimisation of this experiment, it was however deemed to be not viable in the time frame allotted as well as the hurdles imposed by the lack of both reagents and suitable cells lines available. The lack of a cell line that was both permissive to ILTV infection as well as expressing a tagged transgene was the main stumbling block in this project. Whilst there was most likely detection of tagged Ago2 within transduced LMHs during early passages (Figure – 3.6), the banding at the predicted molecular weight was not detected in subsequent passages of the cells whilst still under selection marker pressure. This suggested that the cells were expressing Puromycin resistance genes as they were actively dividing and multiplying however the likelihood is that the cells had switched off the expression of the tagged Ago2. Whilst not directly comparable due to the difference in species, the switching off of lentiviral delivered transgenes has been reported in murine embryonic carcinoma cell lines (He et al., 2005). Another approach was suggested, by way of using a commercially available Ago2 antibody and carrying out immunoprecipitation against this under the experimental term HITS-CLASH. This is a blend of CLASH and its predecessor HITS-CLIP. This approach has been successfully used in the Grey lab (*personal communication*) however this does not guarantee success in chickens. Firstly, the issue of a workable cell line still exists and then there is the issue of finding suitable antibodies that react in the chicken. Given the time constraints of a PhD project it was deemed to be too much of a risk to try and get this project

off the ground and so therefore work was postponed until more reagents became available to make the project viable.

In parallel with this work, the time course of expression of viral genes and viral miRNAs was investigated in LMH cells. This was in the first instance to look at the most optimal time of when to harvest samples for CLASH. In an ideal situation, one would want the maximum number of virally-encoded miRNAs to be expressed. As well as this, detection of viral genes was necessary to ensure that the full temporal cascade of gene expression was occurring in the LMH cell line. In total, five viral genes were detected spanning the full temporal genome (Figure – 3.2). From this data, it was decided that 24 HPI would be the optimal time for harvesting samples as viral genes from all classes of expression kinetics were detected.

Coupled with this was the investigation of the expression of miRNAs. In total seven of the described ten miRNAs were detected in this project (Figure – 3.4). MicroRNAs are highly tissue specific and so the remaining three miRNAs not detected may only be expressed in certain tissues or their expression may be tightly regulated (Guo et al., 2014). Moreover, similar issues of detection have been seen with the previous studies looking at ILTV-encoded miRNAs. The initial publication showing ILTV encoded for miRNAs found a total of eight miRNAs and in a second study by an independent lab group, a further two were identified. However, in the latter study, some of the miRNAs identified in the first study could not be detected (Rachamadugu et al., 2009, Waidner et al., 2009). Furthermore, one miRNA, ILTV-miR-I7 only had 3 reads detected from a total of ~9100 unique small RNA reads. This was one of the miRNAs not detected in this study and it is possible that it was not detected due to sensitivity issues caused by the low number of copies of this particular miRNA. Similar to the detection of viral genes, the conclusions from the detection of viral-miRNAs was to use a time point of 24 HPI where six of the ten miRNAs were detected. As mentioned above, using CLASH may have identified more of the miRNAs as it uses sequencing whereas the methods used for this experiment were not as sensitive.

As it became increasingly clear that the LMH cell line, whilst permissive to ILTV, was not the most suitable cell line for *in vitro* studies, a number of other avian cell lines of differing origins were investigated to see if they were permissive to ILTV infection. One may question why only lytic replication was considered for the experiement considering the majority of herpesvirus miRNAs are expressed during latency (Grey, 2015). The ideal scenario would be to carry out CLASH in a latent model of infection however this is very difficult to achieve even for the most studied of herpesviruses such as HSV-1. Latency is a

very complex process and difficult to compact into a 2D culture model system without the use of pharmacueticals and/or replication defective viruses to help achieve a 'latent' state (Thellman and Triezenberg, 2017). Whilst not ideal, lytic infection has been used previously for CLASH and data generated has proved significant in previous studies. Other cells lines that were chosen to be tested were picked as they have been shown to permissive to infection with other viruses, including other herpesviruses such as MDV-1. The cell lines that were tested were primary chicken kidney cells (CKCs), CLEC213 cells, DF-1 cells, HD-11 cells and primordial germ cells. The former of these were very much permissive to ILTV infection and in agreement with numerous previous studies (Chang et al., 1960, Meulemans and Halen, 1978, Hughes and Jones, 1988). Primary CKCs were not chosen for the experimental work however due to several reasons. Firstly, fresh cells have to be made from kidneys each time an experiment needs to be carried out and so this becomes problematic in large scale experiments. Secondly, primary cells cannot be passaged indefinitely unlike an immortalised cell line and as the introduction of a tagged transgene via lentiviral delivery followed by extended passage as required for CLASH such cells would not be useful. Finally, one may question the moral implication of sacrificing birds purely for their kidneys, something that does not help the 3 R's of animal experimentation. Similarly, chicken embryo fibroblasts (CEFs) were not used as whilst they are permissive to infection, they cannot be passaged indefinitely and so introducing a tagged transgene would face the same hurdles as primary CKCs.

CLEC213 cells (A gift from Dr. Pascale Quéré, INRA, France) are a spontaneously immortalised cell line derived from chicken lung epithelial cells and are somewhat biologically relevant to ILTV pathogenesis and its invasion of the respiratory tract. Previous work using CLEC213 cells had shown they were permissive to Avian Influenza virus infection (Esnault et al., 2011). In stark contrast though, the cells were not permissive to ILTV infection with the Illinois strain and so no follow up work was carried out regarding these cells. A different result may have been observed using the CSW-1 strain which is more virulent however by the time this strain was available, this work had been halted. A cell line of this nature that is permissive to infection would be a useful tool however as it would be more biologically relevant than the permissive cell lines currently available.

DF-1 cells are derived from CEFs through infection with avian leucosis virus (ALV) (Himly et al., 1998). But when tested, viral load was found to decrease over time suggesting they were not permissive. Whilst studies were not carried out to find out the source of the

replication defect, it maybe of future interest as it may identify candidate genes involved in ILTV replication when compared to CEFs.

HD-11 cells are a macrophage like cell line derived by the transformation of haematopoietic cells by ALV (Beug et al., 1979). Macrophages have previously reported to be permissive to infection and support viral replication and therefore HD-11s were looked at as a possible alternative (Von Buelow and Klasen, 1983, Loudovaris et al., 1991a). The results were not as observed in the literature however with a decrease in viral load seen over the time course of the experiment (Figure – 3.7E). Whilst there was an increase in viral burden at 48 HPI, the general trend was still a decrease. There is room for some investigation in to this area of research however as the original experiments suggesting that macrophages were susceptible to ILTV infection were carried out some 25+ years ago. This was based upon the MHC class of the animals with some been resistant to infection whilst others were susceptible (Loudovaris et al., 1991a). It could be quite simply that the birds HD-11 cells were derived from were of a line of birds known to be resistant to ILTV infection.

In addition to investigating chicken cell lines, two quail cell lines were also look at for their permissiveness to ILTV infection. Firstly, QT-35s were investigated. The results showed a decrease in viral load over time (Figure – 3.9A). This is in contrast to the literature that suggests that they are permissive to infection with ILT however this previous experiment involved serial passage of infected cells with uninfected monolayers (Schnitzlein and Tripathy, 1995). Serial passage of viruses is known to cause point mutations within the virus and thus may have interfered with any downstream experiments (Hildebrandt et al., 2014). This would therefore not be useful in the long term as any data generated would not be truly representative of the wild type virus. In addition to QT-35 cells been investigated, QM-7 cells, another quail cell line, were used. The cell line supports replication of both MDV-1 and HVT but no information was available regarding ILTV (Rong et al., 2014). Upon analysis, it was found that QM-7 cells do not support ILTV replication with a decrease in viral load over time (Figure – 3.9B). Whilst the investigation as to whether quail cell lines are permissive to infection, quail are not known to be affected by the disease and so it may not be a representative model for infection. Furthermore, if any *in vivo* experiments are carried out, the main host of the virus, the chicken can be used, as it is a well-defined model in biomedical science.

The final cell line to be investigated was primordial germ cells (PGCs). Isolated from early chicken embryos, they are the progenitors to the germ cells found within adult birds. It was conveyed that they support replication of influenza virus however any data surrounding this

is yet to be published (*Personal communication*, Dr. Nikki Smith, The Roslin Institute). Data presented in this study suggests that PGCs are supportive of ILTV replication with a modest increase in viral burden at latter time points in the experiment. Furthermore, due to the nature of the cells and the potential they have for differentiation, investigations into viral gene expression were carried out similar to LMH cells. Three genes were analysed spanning the temporal cascade of gene expression and they were ICP4, PK and gE. All three gene transcripts could be detected using cDNA at all time points in the experiment (Figure – 3.11A, B & C). Even with these findings, it was decided not to use the cell lines for further downstream experiments as they are a suspension cell line and so some common experiments such as conventional plaque assays would become problematic. More recently however, the development of an adherent PGC cell line may open up the possibility of using this cell line in future (*Personal communication*, Dr. Mike McGrew, The Roslin Institute). There is also the possibility of differentiating these cells into a specific cell line lineage. By doing this, one can produce a number of cell lines of differing linages. This may include epithelial cells to mimic the upper respiratory tract of the bird to model a site of lytic replication however on the other hand one could produce a neuronal cell line capable of mimicking a latent infection. Whilst both would be advantageous to studies investigating the virus biology, the latter maybe of more use when looking at the role of miRNAs considering the majority of herpesviral miRNAs are involved with latency in some capacity (Grey, 2015).

To conclude, efforts were made to try and use CLASH to identify miRNA targets during ILTV infection however experiments necessary for the success of this biochemical approach suggested that it was not possible due to a number of reasons; primarily the lack of a workable cell line that allows for the expression of a stable transgene. Whilst other approaches such as a combination of CLASH and HITs-CLIP could be used (HITS-CLASH), this was deemed to be unsuitable as there was still no guarantee that this approach would work and yield meaningful results. On the other however, several other advancements were identified such as the potential use of PGCs for use in studies, especially considering they have the potential to be differentiated into a number of cell linages. Whilst no conclusive proof of replication can be shown in this study, the detection of all temporal classes of gene at late time points does warrant further investigation.

Chapter 4 – Screening the Infectious Laryngotracheitis Virus genome for viral microRNA target sites

4.1 – Introduction

A number of different cellular pathways can have a profound effect upon the viral lifecycle. Included in these is the RNA interference (RNAi) pathway, which can be manipulated by the invading pathogen to help facilitate completion of the viral lifecycle. MicroRNAs have been known to be encoded by viruses for over a decade now and since their first identification, novel as well as homologous, miRNAs have been identified in a large range of differing virus families (Grey, 2015). Their discovery has opened up a new field under the broader umbrella of virus-host interactions.

Virus encoded miRNAs have the ability to silence gene expression in a similar fashion to endogenous miRNAs using the same host cell machinery (outlined in Section – 1.6.1). This allows them to target both viral and cellular transcripts thus manipulating the host environment to their own advantage. To date, a large proportion of the identified viral miRNAs are encoded for by herpesviruses with adenoviruses, polyomaviruses and retroviruses also having been shown to encode for them (Grey, 2015). As they are non-immunogenic and can have multiple targets, they are particularly useful for viruses to manipulate the host cell environment (Goodrum et al., 2012). Furthermore, they take up very little space in viral genomes which are already compact.

Herpes simplex virus 1 (HSV-1) encodes for 27 mature viral miRNAs, with a number of these miRNAs implicated in maintaining and promoting latency. Four miRNAs have been shown to target viral transcripts that are necessary for reactivation of HSV-1 from latency (Bernier and Sagan, 2018). HSV1-miR-H6 has been shown to target the major immediate-early protein ICP4 whilst HSV1-miR-H2-3p is able to reduce the expression of ICP0, another immediate-early protein (Duan et al., 2012, Umbach et al., 2008a). The final two miRNAs, HSV-miR-H3 and HSV-miR-H4 respectively are encoded antisense to the neurovirulence factor ICP34.5 with experimental data in HSV-2 showing that these miRNAs are able to downregulate the expression of this protein (Tang et al., 2008).

Similar occurrences of virally-encoded miRNAs regulating viral transcripts are present across the herpesvirus subfamilies. Viral miRNAs of KSHV are expressed during latent infection and are positioned alongside latency associated genes in the viral genome. Of these expressed viral miRNAs, KSHV-miR-K12-7 and KSHV-miR-K12-9 both directly target RTA (ORF50), a protein essential for the induction of lytic replication (Qin et al., 2017). In addition to directly targeting RTA, KSHV encoded miRNAs can also indirectly cause a reduction in the expression of RTA. Targeting of Nuclear Factor I/B (NFIB) by KSHV-

miR-K12-3 also causes a reduction in RTA expression as the former can activate the RTA promoter (Lu et al., 2010).

Taken together, the majority of herpesvirus encoded miRNAs so far investigated play integral roles in the maintenance of latency and whilst targets of herpesvirus encoded miRNAs are still being elucidated, it is clear from the targets already validated that these viral miRNAs can and do play a major role in manipulating both the host and virus gene expression profile to ensure persistence. This is perhaps reflective of the long evolutionary history between herpesviruses and their natural hosts.

As more viral miRNAs are identified and characterised using new, more sensitive, laboratory techniques complimented by the ever-decreasing cost of next-generation sequencing and bioinformatical approaches, the diverse mechanisms and interactions between virally-encoded miRNAs and their targets will only improve. These continuing advancements will no doubt aid in the understanding of virus-host interactions at the small RNA level.

4.2 – Aims

The aims for this part of the project were to employ bio-informatics to predict viral targets of virus-encoded miRNAs. To carry this out, viral gene transcripts needed to be mined from the genome and coupled with the miRNA sequences. Following target predictions, top hits were to be investigated in a reporter based screen. Any targets showing knockdown were then investigated further.

4.3 – RNA Hybrid

4.3.1 – Creation of Viral transcripts for Bioinformatic Use

Every predicted viral transcript was screened for potential viral miRNA target sites. To create this list, a set of parameters was laid out. As the ILTV genome is not extensively annotated, an arbitrary 1000 bp upstream of the ATG start codon was included to ensure full coverage of the 5'UTR region of transcripts. Likewise, a minimum of 50 bp downstream of the designated polyA site was used to ensure full coverage of the predicted mRNA transcript. For genes encoded in a right to left orientation, the sequences were then reverse complemented to put all genes in a left to right orientation. In total, 79 ORFs were included in the data covering every predicted gene of ILTV. All sequences used in the predictions were taken from the ILTV genome (NCBI reference sequence: NC_006623.1).

4.3.2 – ILTV miRNA sequences

Sequences for the 10 mature miRNAs were taken from MiRBase (release 21, [http://www.mirbase.org/]). These were then appropriately named and added to a separate file and this was then used in conjunction with the viral transcripts files created in 4.3.1.

4.3.3 – RNA Hybrid Parameters

RNA Hybrid was used for target prediction (https://bibiserv2.cebitec.uni-bielefeld.de/rnahybrid) (Rehmsmeier et al., 2004). To ensure consistency throughout the prediction process, a general set of rules were employed. Firstly, to ensure the seed sequence was the core binding site, helix constraints were set from nt 2 to nt 8. No minimum free energy (MFE) threshold was used and no G:U base pairing was allowed for the seed sequence. Finally, the number of hits per target was set to 10 and no approximate p-values were calculated. These parameters were used for the entirety of the following project laid out in this chapter.

4.3.3 – Filtering of RNA Hybrid predictions

Once the initial screen was carried out, results were then filtered and sorted. In the first instance, targets were sorted on two main criteria, 1) the MFE had to be greater than or equal to -15 kcal/mol and 2), results were filtered to only include viral genes that were designated to be an immediate-early or early gene class in reference to the temporal regulation of gene expression (reviewed in section 1.3.2). The former was chosen as a cut off as the Dalziel lab group has previously identified and validated targets with this MFE (Riaz, 2014). With respect to the second criterium, this was done as any effects upon IE or E genes would have more of an effect upon the virus and they were also likely to have more interactions with the host. This left a total of 227 predictions in the data set. Further filtering on this dataset look at the prescribed functions of the genes by reviewing the literature on homologues, primarily in HSV-1, but also in other Alphaherpesviruses. When these filters were applied a total of 28 candidates were taken forward for experimental testing with one previously validated target site within ICP4 (miR-I6-5p) also taken forward to act as a control for the initial phase of laboratory experiments (Waidner et al., 2011). Table 4.1 details the targets taken forward in the screen.

Table 4.1 – List of sorted and filtered viral targets of ILTV-encoded miRNAs. Targets listed were tested in the initial reporter assay screen

Viral Gene	Target number	miRNA	Target Site	Region?	MFE (Kcal/mol)	Gene Function?
ICP4	1	I6-5p	4521	CDS	-50.6	Absolutely required for viral growth and initiation of lytic replication
UL54	1	I6-3p	1543	CDS	-33.7	ICP27 – several roles in virus life cycle and highly conserved
UL48	1	I5-3p	3376	3'UTR	-18.4	VP16 – forms a complex with Oct1 & HCF to transcriptionally activate IE-genes
	2	I6-5p	3716	3'UTR	-21.5	
UL46	1	I5-3p	2126	CDS	-18.4	VP11/12 – Interacts with UL48 and modulates its activity
	2	I6-5p	2466	CDS	-21.5	
UL30	1	I6-5p	645	5'UTR	-25.8	Encodes for DNA polymerase and is responsible for viral DNA replication
	2	I5-3p	2162	CDS	-25.2	
	3	I3	3747	CDS	-29.7	
UL28	1	I5-5p	5696	3'UTR	-29.3	ICP18.5 – Responsible for processing and packaging of viral DNA
	2	I5-5p	2389	CDS	-30.5	
	3	I2	57	5'UTR	-30.3	
UL24	1	I6-5p	2162	3'UTR	-27.7	Conserved across the herpesviridae family but function is not fully established
	2	I1-5p	3665	3'UTR	-22.9	
UL9	1	I6-3p	972	5'UTR	-24.0	

	2	I6-3p	1068	CDS	-26.3	Facilitates the docking of viral replication machinery to viral DNA origins
UL2	1	I2	1423	CDS	-24.5	Uracil DNA Glycosylase which is responsible for removing of uracil effectively proof-reading synthesised genomes
	2	I6-5p	747	5'UTR	-23.9	
	3	I2	164	5'UTR	-31.7	
UL-1	1	I3	2539	CDS	-30.1	I/E protein with unknown function but it is absolutely required *in vitro*
	2	I5-5p	2290	CDS	-33.9	
	3	I6-5p	2284	CDS	-29.6	
UL47 (sORF1)	1	I5-5p	74	5'UTR	-20.4	VP13/14 – an I/E protein which interacts with VP16 and VHS
US3	1	I2	773	5'UTR	-28.5	Conserved amongst Alphaherpesviruses. Multifunction protein kinase involved with optimal virus replication
	2	I4	1108	CDS	-29.5	
	3	I6-5p	1489	CDS	-27.5	
ORF F	1	I6-5p	601	5'UTR	-30.3	Unknown protein/function but contains 6xCTD domains which are normally found in RNA polymerase II. Possible polymerase activity
	2	I6-3p	677	5'UTR	-33.0	
	3	I6-3p	2458	CDS	-33.3	

Target site refers to the first nucleotide position where the miRNA binds to the mRNA.
Region refers to where in the mRNA transcript the miRNA binds. 5'UTR = 5'End, CDS = Protein coding region and 3'UTR = 3' End
Gene function is taken from existing data in the literature or inferred from HSV-1 homologues

4.4 – Screening of RNA Hybrid Predictions by Luciferase Assay

The 29 targets taken forward for testing were all cloned into the dual luciferase reporter plasmid Psi-Check 2 (Appendix 1). To clone the targets, sense and antisense long oligonucleotides ~110 bp were designed spanning the target site region. These were annealed and cloned into the vector using restriction endonucleases. To confirm positive clones, colony PCR was used with primers spanning the multiple cloning site. From this, positive clones were picked, amplified and sent for sequencing to ensure the target sites were intact with no mutations.

To test the constructs *in vitro*, Psi-Check 2 plasmids with and without the cloned target site were transfected into HEK293T cells alongside a mature miRNA mimic or a scrambled siRNA control. Samples were harvested after 48 hours incubation and used for luciferase assays as described in 2.6.3.

4.4.1 – Testing the system using a known target

To ensure the system was working, a validated target of ILTV-miR-I6-5p in ICP4, (termed ICP4 T1 in this project (Table 4.1)) (Waidner et al., 2011) was tested first. RNA Hybrid predictions carried out as part of this study showed perfect Watson-crick binding between the miRNA and target along the full length of the miRNA that was reflected in the MFE result of -50.6 kcal/mol (Figure – 4.1A). When the luciferase vector containing ICP4 T1 was co-transfected with a miR-16-5p mimic a 35 % reduction in luciferase expression was observed (Figure - 4.1B). Upon statistical analysis, this was found to be significant from the scrambled siRNA ($p =<0.001$) with n = 8 biological replicates (n = 24 technical repeats).

To confirm this reduction was due to the action of the miRNA, the seed sequence was mutated. This was carried out by redesigning the long oligonucleotides with changes to the seed sequence only and repeating the cloning process followed by sequencing of the cloned product. When the mutated target site was tested (termed Mut-ICP4 T1), the effects of ILTV-miR-I6-5p were abrogated with a reduction of luciferase expression of only 10 % (Figure – 4.1B). This was not significantly different from the scrambled siRNA control but was statistically different from the wild type ($p =<0.001$).

A

```
Target: ICP4
MiRNA: ILTV-miR-I6-5p
Length: 22
mfe: -50.6 kcal/mol
Position:  4521
Target 5' A                      A 3'
          CGACGAUGAGGGUACAGGAGAC
          GCUGCUACUCCCAUGUCCUCUG
MiRNA  3'                         5'
```

B

```
Wild Type - CGACGATGAGGGTACAGGAGACA
Mutant    - CGACGATGAGGGTCCCAACAACA
```

C

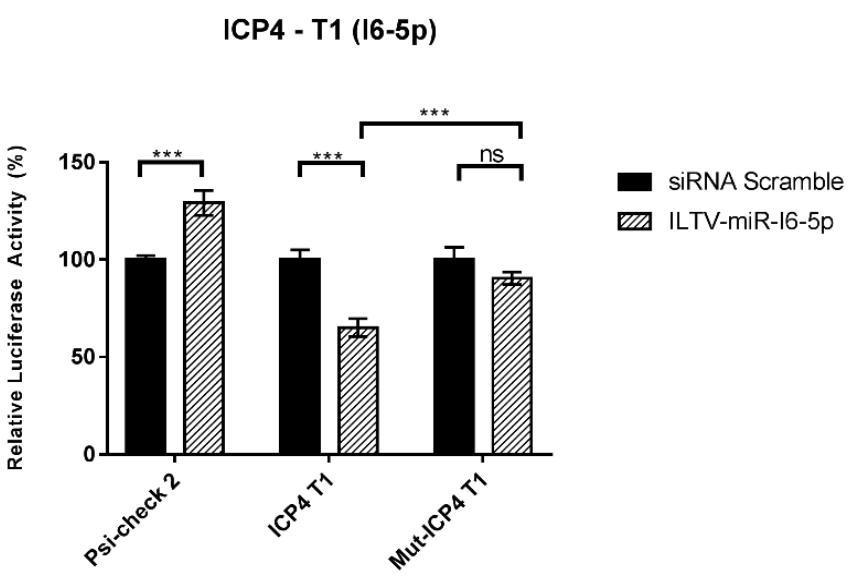


Figure – 4.1. Relative luciferase expression levels of ICP4-T1 and Mut-ICP4-T1 sequences in Psi-check 2 with ILTV-miR-I6-5p compared to a scrambled siRNA control

RNA Hybrid was used to predict viral targets of virally encoded miRNAs (A). HEK293T cells were transfected with a Psi-Check 2 reporter plasmid harbouring a 110 bp sequence containing an ILTV-miR-I6-5p sequence or a mutated sequence. Wild type and mutated seed sequence of the mRNA in bold and underlined (B). ILTV-miR-I6-5p or a scrambled siRNA were also transfected into cells. After 48 hours incubation, *Renilla* luciferase levels were measured and normalised to firefly luciferase levels with expression levels compared between control and test miRNA conditions (C). N = 8 biological replicates with n = 24 technical repeats. Error bars display Standard error of the mean (SEM)

P-values - * = <0.05, ** = <0.01, *** = <0.001

4.5 – Testing of targets using a Luciferase based system

4.5.1 – Screening of predicted targets

In total, 28 novel miRNA targets were tested in this screen. A summary of the results can be found below in Table 4.2. From the constructs tested (excluding the positive control ICP4 target), three targets showed decreases in luciferase expression however upon statistical analysis, only two of these were found to be statistically significant. Experiments listed below were carried out in duplicate with n = 4 biological replicates and n = 12 technical replicates per experiment totalling n = 8 biological and n = 24 technical repeats.

4.5.2 – Targeting of UL46 & UL48 by ILTV-miR-I6-5p

Using RNA Hybrid, a target for ILTV-miR-I6-5p contained within the 3'UTR of UL48 by was identified (Detailed in Figure - 4.2A). The target site is approximately 1520 nt downstream of the stop codon for UL48 and around 200 nt upstream of the predicted PolyA site. Due to its location in the genome, this same target was also found within the coding region of UL46 by RNA Hybrid (diagrammatic representation shown in Figure – 4.6A). Within UL46, the predicted site lies around 150 nt upstream of the stop codon of UL46 and 200 nt upstream of the same PolyA site (Figure - 4.2B). The target sequence for ILTV-miR-I6-5p was cloned into Psi-Check 2 and tested using a mature miRNA mimic. Upon analysis, luciferase expression was reduced by 20 % compared to the scrambled siRNA ($p =<0.001$) (Figure - 4.2C). Mutation of the seed sequence abrogated this effect with no significant difference between the mutant construct and empty construct plus mimic. There was a statistically significant difference between the wild type sequence and the mutant sequence ($p =<0.001$). Further validation work was carried out and detailed in section 4.7.

4.5.3 – Targeting of UL29 by ILTV-miR-I2

RNA Hybrid analysis predicted a potential target site for ILTV-miR-I2 in the 5'UTR of UL28 (Target 1, Table 4.2) however this was around 940 nt upstream of the ATG start site. This same site was also situated in the coding region of UL29 with the same mean free energy of -30.3 kcal/mol (Figures 4.3A&B). When cloned and tested in the reporter construct, a reduction of 15 % was observed when compared to the scramble siRNA control (Figure – 4.3C). Upon statistical analysis, this was found to be significantly different ($p =0.004$). Upon mutation of the seed sequence, the effects of the miRNA mimic were abrogated and this was statistically different from the wild type sequence ($p =<0.001$). In addition to this, there was no significant difference found between the mutant sequence and empty vector control plus mature miRNA mimic (Figure – 4.3C). Further validation work is laid out in section 4.6.

4.5.4 – Targeting of UL24 by ILTV-miR-I6-5p

UL24 was predicted to be targeted by ILTV-miR-I6-5p in the 3'UTR approximately 300 nt downstream of the stop codon. The predicted PolyA site is around 1500 nt further downstream. RNA Hybrid gave a mean free energy prediction of -27.7 kcal/mol (Figure – 4.4A). The target site was cloned as previously described into Psi-Check 2 and tested using a mature miRNA mimic in conjunction with a siRNA scramble control. Upon analysis, luciferase expression was reduced by 5 % when compared to the siRNA scramble. Statistical analysis revealed that this interaction was not significant (Figure – 4.4B). For this reason, no mutant construct was tested and this target was not taken further.

Table 4.2 – Summary of results from testing of viral targets in a reporter system

Viral Gene	Target number	miRNA	MFE (Kcal/mol)	Decrease in luciferase expression?	Statistically significant?	Relative luciferase activity (%) in the presence of the miRNA	Standard Error of the Mean (%)
ICP4	1	I6-5p	-50.6	Yes	Yes	65.22 %	4.63 %
UL54	1	I6-3p	-33.7	No	No	109.64 %	2.31 %
UL48	1	I5-3p	-18.4	No	No	136.33 %	1.87 %
	2	I6-5p	-21.5	Yes	Yes	79.78 %	4.01 %
UL46	1	I5-3p	-18.4	No	No	136.33 %	1.87 %
	2	I6-5p	-21.5	Yes	Yes	79.78 %	4.01 %
UL30	1	I6-5p	-25.8	No	No	107.34 %	1.78 %
	2	I5-3p	-25.2	No	No	105.87 %	2.32 %
	3	I3	-29.7	No	No	99.64 %	2.46 %
UL28*	1	I5-5p	-29.3	No	No	109.38 %	3.62 %
	2	I5-5p	-30.5	No	No	108.57 %	5.30 %
	3*	I2	-30.3	Yes	Yes	85.87 %	11.96 %
UL24	1	I6-5p	-27.7	Yes	No	94.64 %	1.48 %
	2	I1-5p	-22.9	No	No	102.18 %	3.34 %
UL9	1	I6-3p	-24.0	No	No	124.34 %	2.74 %
	2	I6-3p	-26.3	No	No	106.34 %	2.33 %
UL2	1	I2	-24.5	No	No	134.32 %	2.72 %
	2	I6-5p	-23.9	No	No	111.39 %	0.56 %
	3	I2	-31.7	No	No	138.04 %	1.79 %
UL-1	1	I3	-30.1	No	No	127.46 %	2.73 %
	2	I5-5p	-33.9	No	No	113.13 %	1.30 %
	3	I6-5p	-29.6	No	No	117.84 %	1.64 %
UL47 (sORF1)	1	I5-5p	-20.4	No	No	119.64 %	1.64 %
US3	1	I2	-28.5	No	No	142.44 %	1.40 %
	2	I4	-29.5	No	No	113.96 %	2.23 %
	3	I6-5p	-27.5	No	No	125.12 %	3.64 %
ORF F	1	I6-5p	-30.3	No	No	112.34 %	1.84 %
	2	I6-3p	-33.0	No	No	117.93 %	2.21 %
	3	I6-3p	-33.3	No	No	105.24 %	2.37 %

*** = Refers to UL29 once subsequent investigations were carried out regarding the location of the target site.**

A

```
Target: UL48
MiRNA: ILTV-miR-I6-5p
Length: 22
mfe: -21.5 kcal/mol
Position 3716
Target 5'              A          G 3'
                        ACAGGAGAC
                        UGUCCUCUG
MiRNA  3' GCUGCUACUCCCA            5'
```

B

```
Target: UL46
Length: 2730
MiRNA: ILTV-miR-I6-5p
Length: 22
mfe: -21.5 kcal/mol
Position 2466
Target 5'              A          G 3'
                        ACAGGAGAC
                        UGUCCUCUG
MiRNA  3' GCUGCUACUCCCA            5'
```

C

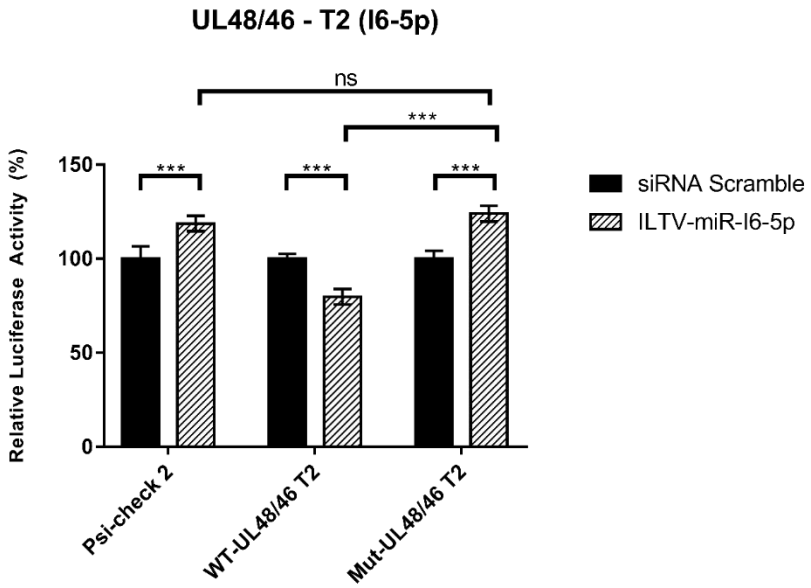


Figure – 4.2. Predicted targets of ILTV-miR-I6-5p and relative luciferase expression levels of UL48/46-T2 and Mut-UL48/46-T2 sequences in Psi-check 2 with ILTV-miR-I6-5p compared to a scrambled siRNA control

RNA Hybrid was used to predict viral targets of virally encoded miRNAs (A&B). HEK293T cells were transfected with a dual-luciferase reporter plasmid containing either a wild-type or a mutated sequence containing a predicted miRNA target. A mature miRNA mimic or a scrambled siRNA control were co-transfected with the reporter plasmid and after 48 hours incubation, cells were harvested. *Renilla* luciferase levels were measured and normalised to firefly luciferase levels with expression then compared between control and test miRNA conditions (C). Error bars display Standard error of the mean (SEM)

P-values - * = <0.05, ** = <0.01, *** = <0.001, ns = Not significant

A

```
Target: UL28
MiRNA : ILTV-miR-I2
Length: 24
mfe: -30.3 kcal/mol
Position  57
Target 5' C         AGAU         A 3'
           CGGC CCUA     CAGCCUUU
           GCCG GGAU     GUCGGAAG
MiRNA  3' A    A    AGCGU        G 5'
```

B

```
Target: UL29
MiRNA : ILTV-miR-I2
Length: 24
mfe: -30.3 kcal/mol
Position  3134
Target 5' C         AGAU         A 3'
           CGGC CCUA     CAGCCUUU
           GCCG GGAU     GUCGGAAG
MiRNA  3' A    A    AGCGU        G 5'
```

C

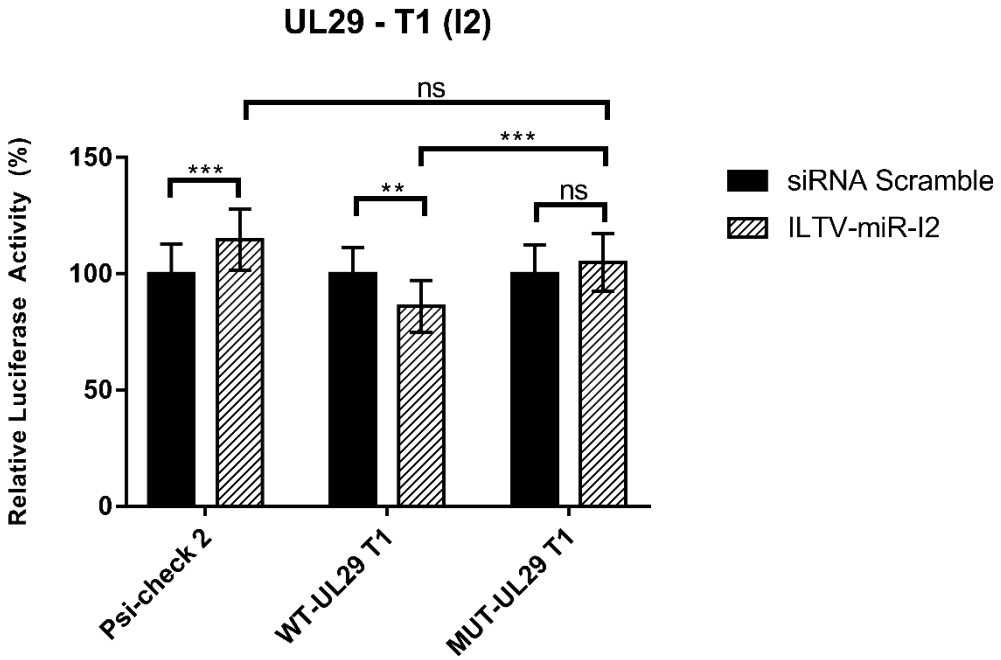


Figure – 4.3. Relative luciferase expression levels of Psi-check 2 plasmid containing either a wild type UL29-T1 sequence or mutant UL29-T1 co-transfected with ILTV-miR-I2 compared to a scrambled siRNA control

RNA Hybrid was used to predict viral targets of virally encoded miRNAs and ILTV-miR-I2 was predicted to target the same sequence in both UL28 and UL29 (A&B). HEK293T cells were transfected with a Psi-Check 2 reporter plasmid harbouring a 110 bp sequence predicted to be target by the viral miRNA ILTV-miR-I2. ILTV-miR-I2 or a scrambled siRNA were also transfected into cells. After 48 hours incubation, *Renilla* luciferase levels were measured and normalised to firefly luciferase levels with expression levels compared between control and test miRNA conditions (C). Error bars display Standard error of the mean (SEM)

P-values - * = <0.05, ** = <0.01, *** = <0.001, ns = Not significant

A

```
Target: UL24
MiRNA : ILTV-miR-I6-5p
Length: 22
mfe: -27.7 kcal/mol
Position  2162
Target 5' G      UUAGCCCAGC   CC          A 3'
           GAUGAU          GGG  GUGCAGGAGA
           CUGCUA          CUC  CAUGUCCUCU
MiRNA  3' G                 C           G 5'
```

B

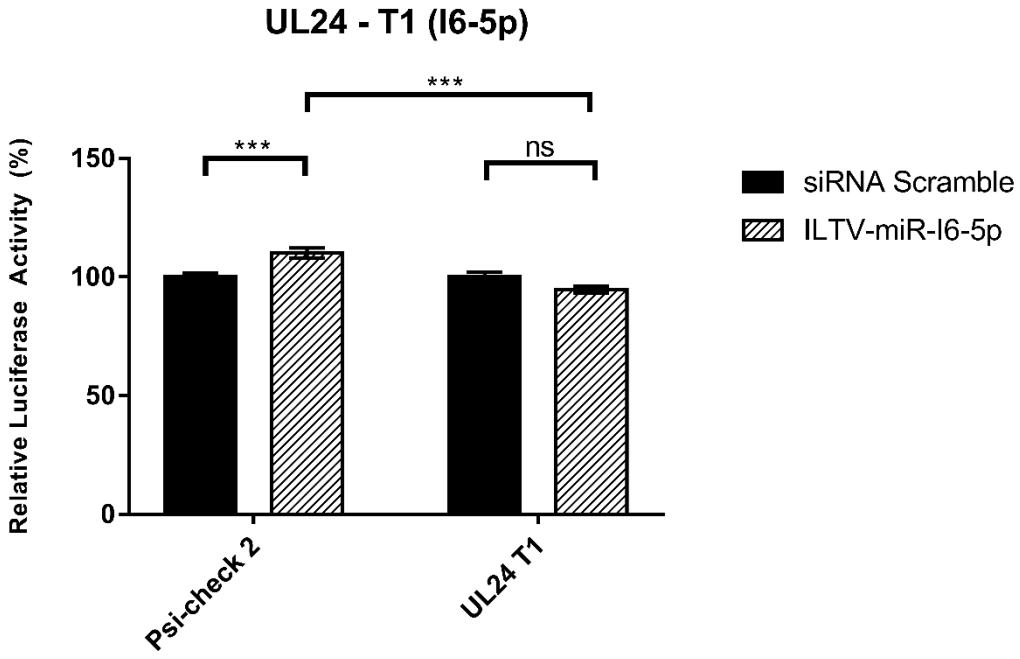


Figure – 4.4. Relative luciferase expression levels of Psi-check 2 plasmid containing UL24-T1 sequence co-transfected with ILTV-miR-I6-5p compared to a scrambled siRNA control
RNA Hybrid was used to predict viral targets of virally encoded miRNAs (A). HEK293T cells were transfected with a Psi-Check 2 reporter plasmid harbouring a 110 bp sequence predicted to be target by the viral miRNA ILTV-miR-I6-5p. ILTV-miR-I6-5p or a scrambled siRNA were also transfected into cells. After 48 hours incubation, *Renilla* luciferase levels were measured and normalised to firefly luciferase levels with expression levels compared between control and test miRNA conditions (B). Error bars display Standard error of the mean (SEM)
P-values - * = <0.05, ** = <0.01, *** = <0.001, ns = Not significant

4.6 – Investigating UL29 as a target of ILTV-miR-I2

4.6.1 – Introduction

UL29, sometimes referred to as ICP8, encodes for an early gene product (Mahmoudian et al., 2012). The protein produced is commonly referred to as the major DNA-binding protein and plays several key roles during virus infection to such an extent that a homologue is encoded for by all known herpesviruses (Weller and Coen, 2012). It is believed to play central roles in viral DNA synthesis, control of viral gene expression and the formation of both pre-replicative and replication compartments (Weller and Coen, 2012).

More recent work has shown the importance of UL29 in the formation of filaments that are essential for the assembly of replication compartments. These filaments are proposed to form a scaffold to which other proteins are recruited to aide in the replication of viral DNA (Darwish et al., 2016). Taken together, these interactions suggest a key role for UL29 (ICP8) in the herpesvirus replication cycle making it a prime target for suppression by both the virus and cellular response pathways.

4.6.2 – Cloning of a 6xHis-UL29 construct

Attempts to validate the findings of the luciferase based reporter screen were carried out using a plasmid expressing UL29 tagged with 6xHis to allow for detection using conventional antibodies. To carry this out, primers were designed to amplify the full coding region of UL29 and included a 6xHis tag on the sense primer (forward) downstream of the ATG start site to ensure the tag was incorporated into the expressed protein.

PCR was used to amplify up the DNA region of interest and PCR products visualised on an agarose gel. Bands of the approximate molecular weight were excised, DNA was extracted, cloned into the TOPO4 vector and *E. coli* transformed as outlined in 2.3.4. Colonies were picked and amplified in liquid LB broth overnight. DNA was extracted and sent for sequencing to confirm the success of the PCR *(Data Not Shown)*. Restriction endonuclease digestion was then used to insert the PCR product into the expression plasmid PcDNA3.1+. Using Kpn1 and Not*1*, the verified UL29 fragment was excised from TOPO4 and inserted into pcDNA3.1+ digested with the same restriction endonucleases. Following ligation and transformation, small cultures were prepared and DNA prepped for a diagnostic digest. Digestion with Kpn1 and Not*1* confirmed the presence of 6xHis-UL29 in pcDNA3.1+ using the sequenced TOPO 4 plasmid as a control *(Data Not Shown)*.

4.6.3 – Mutagenesis of UL29-6xHis using mutagenesis PCR

Once UL29-6xHis was cloned, this construct was used to make a mutant plasmid whereby the miRNA target seed sequence was mutated. This was done using mutagenesis primers designed using the Agilent QuikChange Primer Design (https://www.genomics.agilent.com/primerDesignProgram.jsp). These primers were then used for mutagenesis PCR as per laid out in Section 2.2.9 of this thesis *(Data Not Shown)*. The plasmid made through mutagenesis PCR was named mut-UL29-6xHis.

4.6.4 – Validation attempts by Western Blot of ILTV-miR-I2 knockdown of UL29-6xHis

Using both the wild-type and mutant UL29 plasmid, validation of the miRNA interaction was carried out. HEK293T cells were seeded at an appropriate density in 12-well plates and transfected with either the wild-type or mutant UL29-6xHis construct on its own, with the mature miRNA mimic or a siRNA scramble control. Cells were left for 48 hours and then harvested for protein quantification using Laemmli buffer. 20 µl of sample were loaded per lane in a SDS—PAGE gel (Bio-Rad Industries) and gels were run for 100 minutes at 80 v. Following transfer of the proteins onto nitrocellulose membrane (Bio-Rad) using the TransBlot Turbo system (Bio-Rad) the membrane was probed with antibodies against 6His and Alpha-tubulin at appropriate dilutions laid out in Table 2.1 of section 2.4.6. Appropriate secondary Li-Cor antibodies were then used (also laid out in Table 2.1 of section 2.4.6) and membranes were imaged using the Li-Cor system.

Alpha tubulin bands were detected at approximately 50 kDa and are shown in green on Figure – 4.6. No banding for 6His (UL29) could be detected at the estimated molecular weight of 110 kDa (marked with white asterisks on Figure – 4.6). Multiple biological repeats (n=3) failed to detect either the wild-type or mutant UL29-6xHis proteins on an immunoblot whilst alpha-tubulin could readily be detected on all.

Due to this lack of detection, validation of the interaction between UL29 and ILTV-miR-I2 could not be achieved. Time restrictions and resources meant that other methods of validation were not attempted for this particular interaction.

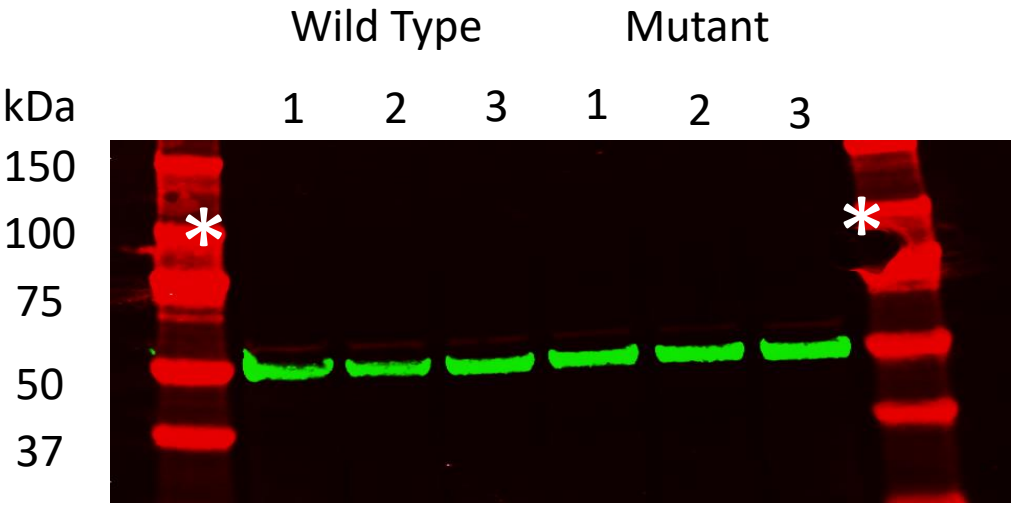


Figure – 4.5. UL29-6xHis Western blot validation attempt
UL29 was cloned into pcDNA3.1+ with a 6xHis tag to allow for detection. Once cloned, a mutant UL29-6xHis with an altered miRNA seed sequence was created using mutagenesis primers as laid out in Sectio2.2.9 of this thesis. These constructs were transfected into HEK293T cells either on their own (1), with the mature miRNA mimic (2) or with a scramble siRNA control (3). Cells were then harvested and lysates prepared for protein quantification using Laemmli buffer. Lysates were used in a western blot as described (section 2.4) and probed with appropriate antibodies. Green banding shows alpha tubulin. White asterisks indicate the size of the band where UL29-6xHis and Mut-UL29-6xHis should be.

4.7 – Further Investigation into UL48 or UL46 as a target of ILTV-miR-I6-5p

4.7.1 – Introduction

Most of the information regarding the interactions between UL46, UL48 and ICP4 promoter (ICP4p) is based upon research from other Alphaherpesviruses, particularly HSV-1 and HSV-2 (Kato et al., 2000). UL46, also referred to as VP11/12 is an ILTV early/late (Mahmoudian et al., 2012). It encodes a tegument protein that is unique to the Alphaherpesvirus family and has several roles in both virion assembly as well as interactions with UL48, the major trans inducing factor (α-TIF) (Kato et al., 2000).

During virion assembly, UL46 is thought to play a role alongside UL47, UL48 and UL49 in organising the tegument by forming interactions with both the inner and outer tegument proteins as well as viral glycoproteins (Owen et al., 2015). The gene is however dispensable for virus replication *in vitro* with only a decrease in plaque size observed (Kopp et al., 2002).

Moreover, UL46 known as the major trans inducing factor (α-TIF) has been shown to interact with UL48. Infection studies have shown that it co-localises with UL48 in infected cells and in reporter based assays can increase the expression from the ICP4 promoter when co-expressed with UL48. In stark contrast however, it inhibits ICP4 promoter driven gene expression when expressed on its own (Kato et al., 2000).

UL48 (VP16) is a key activator of lytic replication and interacts with the cellular proteins HCF and OCT-1, forming a transcriptional regulatory VP16-induced complex in HSV-1 that binds to the sequence 'TAATGARAT' (Thomas et al., 1998). This induces the expression of all sequences with this sequence and includes the major immediate-early gene ICP4. In addition to this, UL48 also plays roles in tegument formation and virion maturation.

Work carried out using pseudorabies virus (SuHV-1) UL48 null viruses demonstrated that virus replication kinetics were greatly delayed and these viruses have significantly reduced viral titres and plaque sizes (Fuchs et al., 2003). This group also demonstrated that UL48 was necessary to link the capsid to future envelope-associated tegument proteins during virion formation (Fuchs et al., 2002a) suggesting that UL48 has two important roles in the virus life cycle at different stages.

4.7.2 – Investigating whether UL46 or UL48 is the primary target of ILTV-miR-I6-5p

Findings from the luciferase based screen (section 4.5.1 of this chapter) showed knockdown of either UL46 or UL48 in the presence of a ILTV-miR-I6-5p mimic however it was not known which viral transcript(s) was targeted by the miRNA. To investigate this, a series of

RT-PCRs were carried out to determine if UL46 or UL48 or both transcripts contained the predicted miRNA target sequence. The rationale behind this experiment was to try and elucidate which transcript is targeted as there is a lack of information regarding transcriptional start and termination sites for ILTV. It was thought that by using cDNA, it would help distinguish somewhat if the site is situated

Primers were designed using a sense (forward) primer situated upstream of the UL48 stop codon and then two anti-sense (reverse) primers. The first of these was situated down steam of the UL48 stop codon but crucially upstream of the UL46 start codon whilst the second primer was situated downstream of miRNA target site (displayed in Figure – 4.6A).

A second set of primers were designed to target the UL46 coding region with a sense primer downstream of the UL46 AUG start codon and the same reverse primer downstream of the miRNA target site.

RT-PCRs were then carried out on DNase treated RNA with cDNA synthesised from either randomly primed RNA or Oligo (DT) primed RNA from both infected and mock infected samples. In addition to this, appropriate minus reverse transcriptase (-RT) and DNA positive controls were included. PCR products were then visualised using agarose gel electrophoresis and imaged.

Primer set one for UL48 showed bands produced in both randomly primed and oligo (DT) primed infected cDNA at the same size as the DNA positive control (Figure – 4.6B). No bands could be detected using primer set 2 apart from the DNA positive control (Figure – 4.6C).

In contrast, primer set 3 which targeted the UL46 coding region showed strong banding for both randomly primed and oligo (DT) primed infected cDNA alongside the DNA positive control (Figure – 4.6D). These results show the miRNA target site to be in the UL46 coding region and not in the UL48 3’UTR.

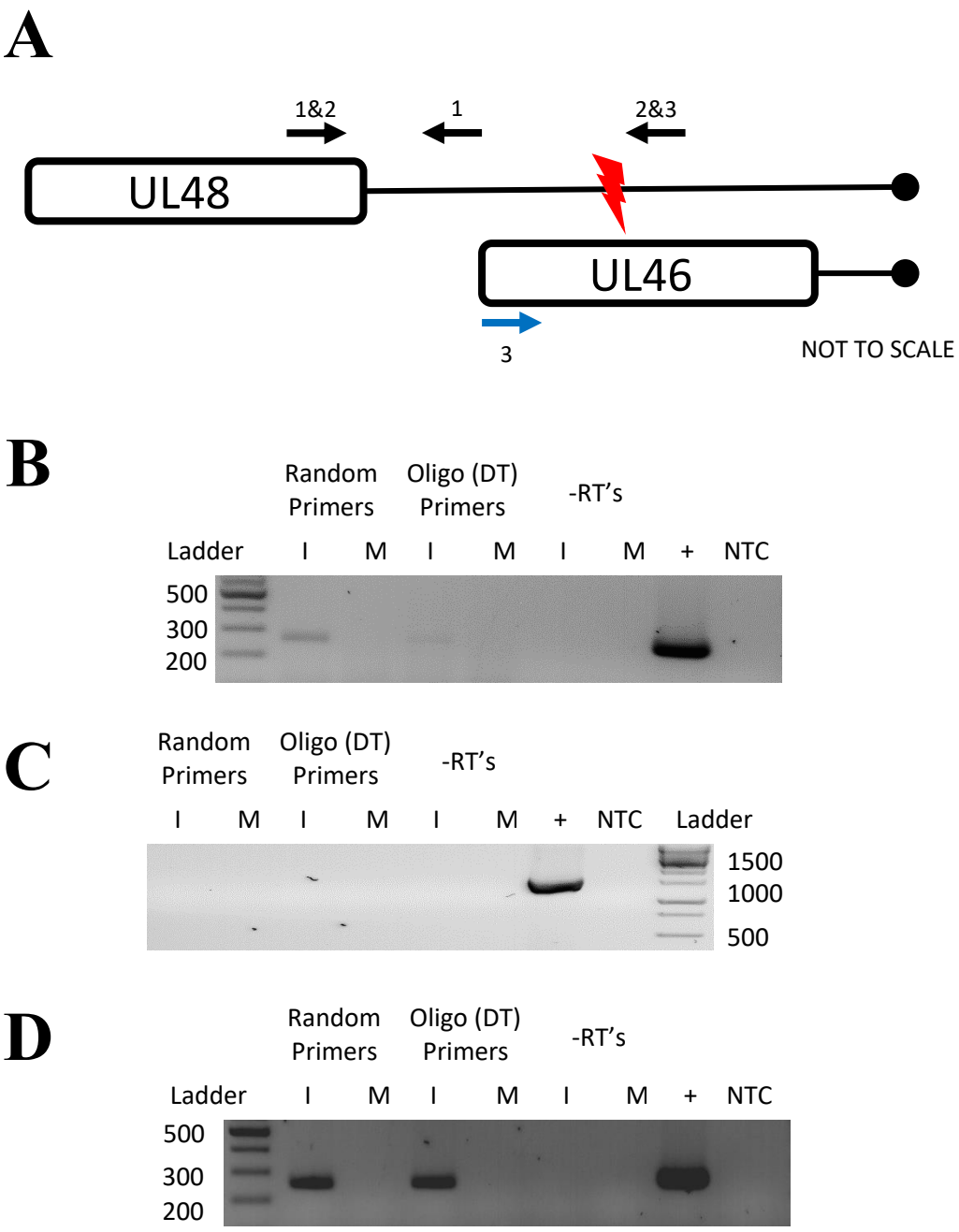


Figure – 4.6. RT-PCR investigations into which ILTV transcript is targeted by ILTV-miR-I6-5p
RNA Hybrid predicted ILTV-miR-I6-5p to target both the 3'UTR of UL48 and the coding region of UL46. Schematic diagram of the genome arrangement surrounding UL48 and UL46 (A). RT-PCR analysis was carried out using a forward primer upstream of the UL48 stop codon along with a reverse primer upstream of UL46 (primer set 1, B) or downstream of the predicted miRNA target site (primer set 2, C). Finally, a primer situated downstream of the UL46 start codon was used in conjunction with the reverse primer downstream of the miRNA target site (primer set 3, D). Arrows depict primer direction and numbers refer to primer set. Boxes in (A) represent the coding region of ORFs whilst solid black line represents the 3'UTR of the respected ORFs.

4.7.3 – Cloning of a UL46-HA construct

As the miRNA target was shown to be within the coding region of UL46, primers were designed to amplify the coding region of UL46 and included a small tag to allow for detection. Initial attempts to PCR amplify up the UL46 region proved difficult *(Data Not Shown)* and so a set further set of primers termed outer UL46 primers were designed to amplify up a larger region than just the specific coding region. From this, the HA-tagged primers were then used to amplify up the region of interest with the tag at either the N or C-terminal.

PCR products were visualised on an agarose gel with bands excised and purified prior to ligation and transformation into TOPO 4 vector. Sequencing was then used to confirm the sequence as UL46 with the HA tag at either the N or C-terminal *(Data Not Shown)*. Tagged UL46 HA was then cloned into the plasmid pcDNA3.1+. Following ligation and transformation, DNA was extracted from small cultures and diagnostic digests were carried out to look for insertion of the product. Upon digest, it was found only C-terminal tagged UL46; termed UL46-HA, had been successfully cloned into pcDNA3.1+. In the interests of time, the decision not to carry on with cloning an N-terminal tagged UL46 was taken, therefore only UL46-HA was taken forward for use in downstream experiments.

4.7.4 – Mutagenesis of UL46-HA plasmid to ablate the miRNA seed sequence

Upon successful cloning of UL46-HA, mutagenesis primers were designed to mutate the miRNA seed sequence of ILTV-miR-I6-5p using Agilent QuikChange Primer Design as described in Section 4.6.3 of this chapter. Mutagenesis PCRs were carried out as previous and products were transformed into ultra-competent cells as per the manufacturer's instructions. The mutant construct was termed mut-UL46-HA.

4.7.5 – Validation attempts of ILTV-miR-I6-5p targeting UL46

Using both wild-type and mutant UL46-HA constructs, HEK293T cells were transfected with the plasmids either on their own, with the mature ILTV-miR-I6-5p mimic or with a scramble siRNA control. Following incubation, cells were harvested for protein quantification with Laemmli buffer. SDS-PAGE gels were loaded with 20 µl of sample per lane. Following separation, proteins were transferred using the TransBlot Turbo system (Bio-Rad) and immunoblots were probed with HA and alpha-tubulin antibodies at appropriate dilutions (See Table 2.1 in Section 2.4.6). Secondary Li-Cor antibodies were then used for visualisation (also Table 2.1 for dilution concentrations) before immunoblots were imaged on the Li-Cor system.

Alpha-tubulin was detected at 50 kDa and is shown in green on Figure – 4.7A. Both wild-type and mutant UL46-HA could be detected with an approximate molecular weight of 63 kDa (bands shown in red, Figure – 4.7A). Protein levels were then quantified using Image Studio (Li-Cor). Individual protein levels were calculated for alpha tubulin and UL46 for each lane. UL46 levels were then normalised against the alpha-tubulin reading and this was carried out for both wild-type and mutant constructs across all conditions and biological repeats (n=4). Normalised UL46-HA levels were then used to look at protein level changes by using the protein only lane as control levels. Data analysis found no significant difference between the wild-type and mutant UL46-HA plasmids in the presence of the mature miRNA mimic. Similarly, there was no significant difference between the protein only (both wild-type and mutant) and protein + miRNA mimic across the biological replicates. A small, statistically significant difference was observed between the WT-UL46-HA and mut-UL46-HA in the presence of the siRNA scramble ($p = 0.039$) though the mut-UL46-HA plus scramble was not statistically significant from any other condition.

The Western blots could not determine a statistically significant knockdown of WT-UL46-HA in the presence of ILTV-miR-I6-5p over several biological replicates ($n = 4$) and therefore this target was also not validated.

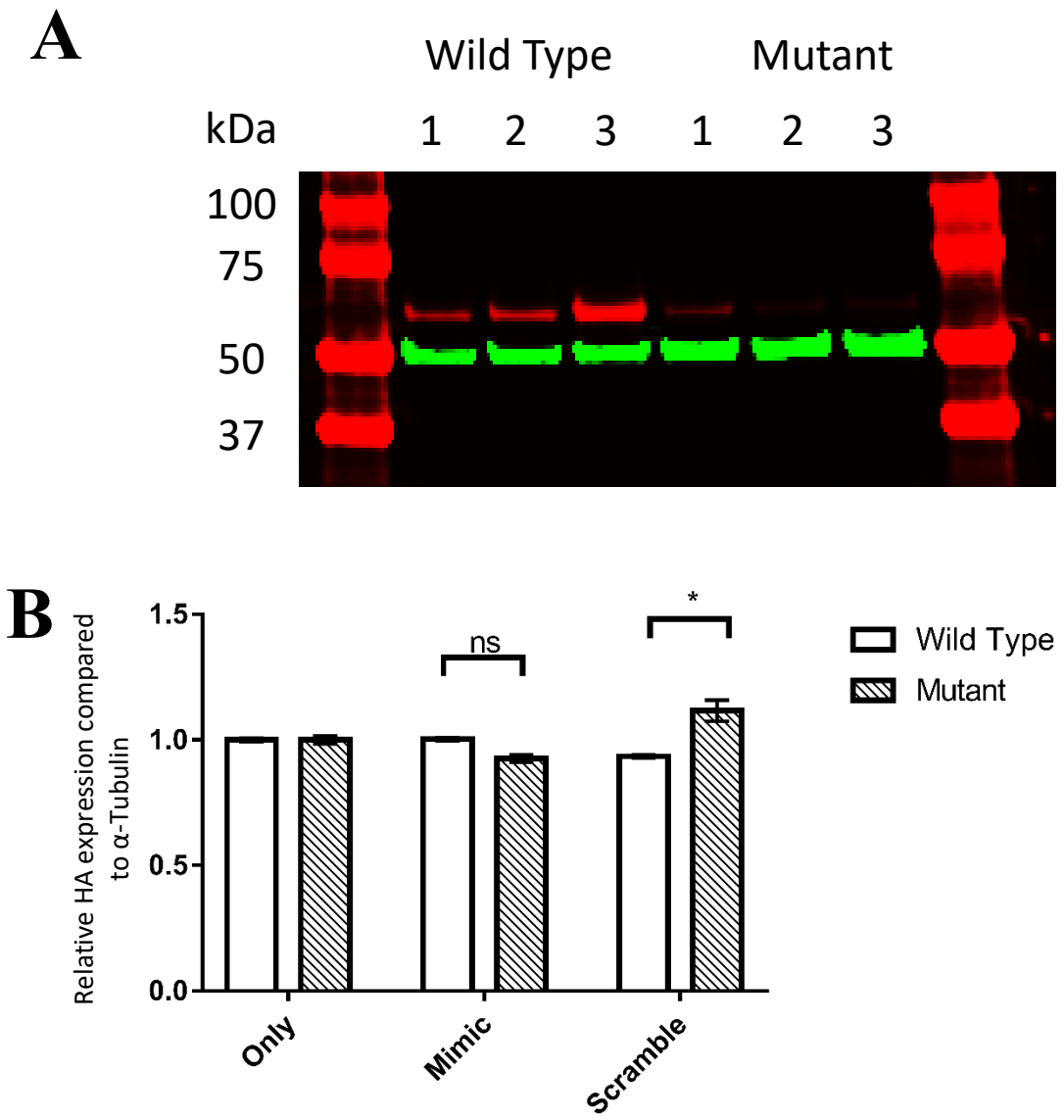


Figure – 4.7. UL46-HA Western blot validation attempts and relative expression of UL46-HA and mut-UL46-HA

UL46 was cloned with a HA tag into pcDNA3.1+. Following successful cloning a mutant UL46-HA (termed mut-UL46-HA) was created by mutagenesis PCR to alter the miRNA seed sequence of the ILTV-miR-I6-5p target site. These constructs were transfected into HEK293T cells either on their own (1), with the mature miRNA mimic (2) or with a scramble siRNA control (3). 48 hours post transfection, cells were harvested for protein quantification using Laemmli buffer. Lysates were boiled and used for conventional western blotting. Immunoblots were probed with a HA antibody (red banding) and an alpha-tubulin antibody (green banding) as a loading control (A). Protein levels were quantified using Image studio software (Li-Cor). HA levels were compared to alpha tubulin levels as a control and then HA levels were compared between conditions using the only lane as the standard to look at relative expression levels of UL46-HA (B).
P-values, * = <0.05, ns = not significant

4.8 – Investigating the Interplay between UL46, UL48 and ICP4 Promoter

4.8.1 – Cloning of ICP4 promoter and UL48-FLAG constructs

As well as cloning the coding region of UL46 with a HA tag into pcDNA3.1+, the coding region of UL48 with a FLAG tag as well as the full promoter region of ICP4 was cloned into pcDNA3.1+ and PGL3b respectively. The same process as used for UL46-HA was used to clone UL48-FLAG into pcDNA3.1+ except there was no need for an extra PCR step using outer primers to amplify across the UL48 region.

To clone the ICP4 promoter, termed ICP4p, primers were designed to amplify up the 5'UTR of ICP4. To ensure the full promoter was covered, a sense (forward) primer was designed approximately 1000 bp upstream of the AUG start codon and an anti-sense (reverse) primer was designed around 50 bp downstream of the AUG start codon. PCR was then used to amplify the region of interest which was then visualised on an agarose gel. Excised bands were purified and ligated into the TOPO 4 vector as previously described. This was then sequenced which confirmed the PCR was successful *(Data Not Shown)*. The product of interest was then excised from TOPO 4 and cloned into PGL3b *(Data Not Shown)*.

4.8.2 – Investigating the interplay between UL46, sORF1 (UL47), UL48 and ICP4 promoter by luciferase assay

Work in HSV has shown that UL46 interacts with UL48 which in turn interacts with the promoter of ICP4 at the TAATGARAT sequence (Thomas et al., 1998, Kato et al., 2000). Therefore investigations into the interactions between UL46, UL48 and ICP4p in ILTV were carried out. This was done using the cloned pUL46-HA, pUL48-FLAG and ICP4p constructs previously mentioned. In addition, pUL47 (sORF1), another protein thought to mediate UL48 was also used. This was cloned by Dr. I Dry (The Roslin Institute) and provided as a kind gift. This included a 6xHis tag at the C-terminal.

Six-well plates were seeded with CLEC213 cells and left to adhere overnight. CLEC213 cells are of chicken lung epithelial origin and are biologically relevant to ILTV infection. Cells were transfected with a total of 2.5 µg of plasmid DNA which consisted of 820 ng ICP4p promoter, 820 ng "protein 1" and 820 ng "protein 2" (either pUL46, pUL47 or pUL48 with empty pcDNA3.1+ or a combination of two of the proteins) plus 40 ng of SV40 plasmid which contains a *Renilla* luciferase and acts as a control. Transfected cells were incubated for 48 hours before harvesting as described (Chapter 2, Section 2.6.3). Luciferase assays were then performed as described and data was then analysed.

Firstly, luciferase expression was analysed to determine if the ICP4 promoter was constitutively active. Compared to an empty vector control, luciferase levels were increased

129-fold in the cells transfected with the plasmid containing the ICP4 promoter. This was statistically significant ($p =<0.0001$).

Luciferase expression in the presence of the pUL46, pUL47 or pUL48 or combinations of these was then analysed using the levels in the presence of only the ICP4p plasmid as baseline. In total, 2.5 µg of DNA was transfected per well consisting of 820 ngs ICP4p, 820 ngs protein 1, 820 ngs protein 2 and 40 ngs of *Renilla* luciferase. Where only a single protein is used, control DNA (empty pcDNA 3.1+) was used to ensure a standard 2.5 µg of DNA was transfected per well each time. When co-transfected with pUL46 only, promoter activity decreased by 50 %. Statistical analysis confirmed this to be significant ($p =<0.001$). ICP4p in the presence of pUL47 also showed a statistically significant decrease in luciferase expression ($p =0.047$) whereas there was no significant difference of expression when co-transfected with pUL48. Dual transfection of both pUL46 and pUL47 with ICP4p also caused a statistically significant decrease in expression ($p =<0.001$). In contrast there was no difference seen between ICP4p alone and co-transfection with UL46 and UL48. Likewise, there was no significant difference between ICP4p alone and UL47 + UL48 co-transfection.

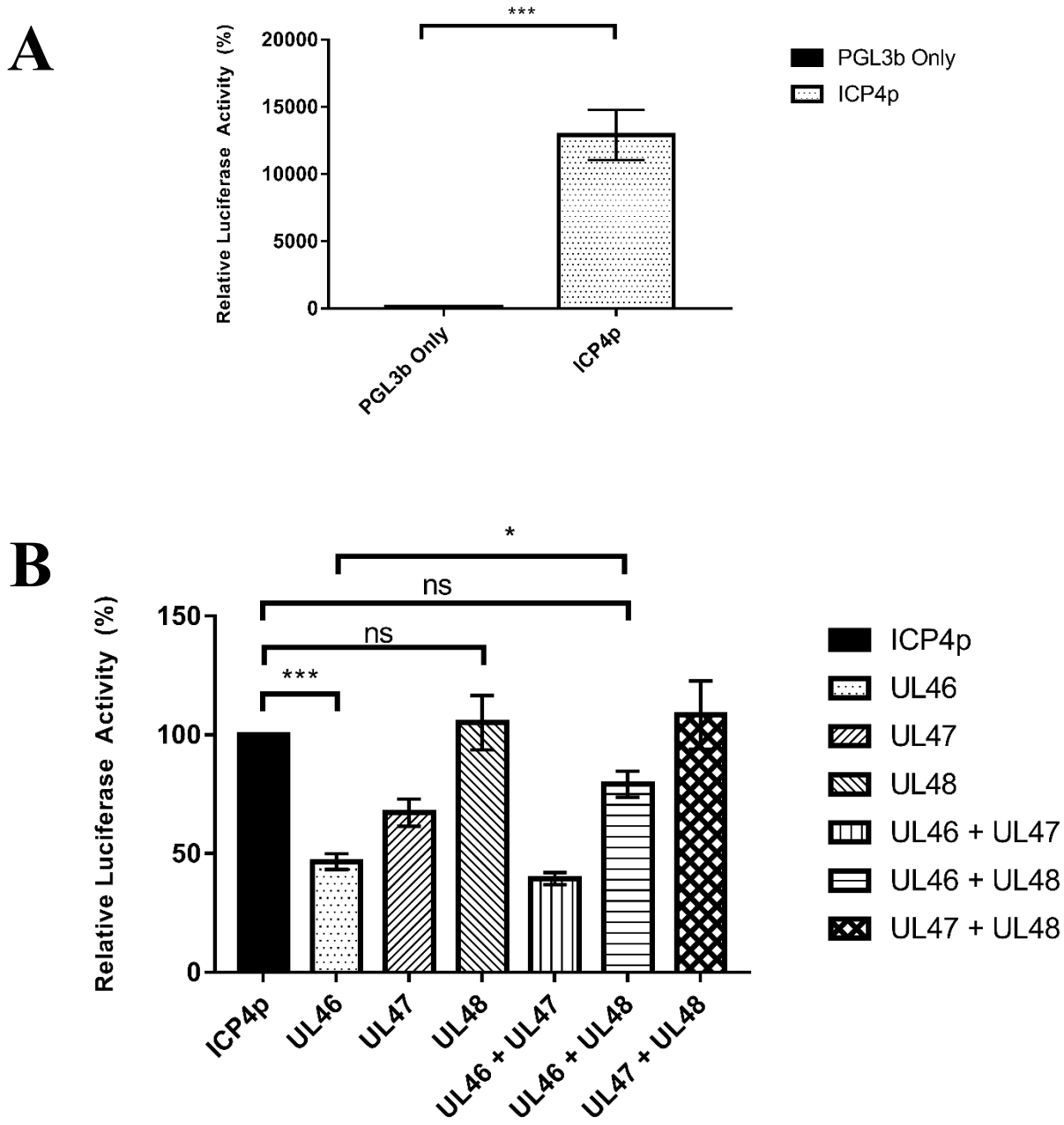


Figure – 4.8. The interplay between UL46, UL47 and UL48 in activating the ICP4 promoter in a reporter based system

UL46, UL47 and UL48 were cloned into the expression plasmid pcDNA3.1+ whilst ICP4 promoter (ICP4p) was cloned into PGL3b. CLEC213 cells were transfected with 2.5 µg total plasmid DNA consisting of 820 ng of ICP4p, 820 ng protein 1, 820 ng protein 2 (where only one [or no protein] was transfected, empty plasmid control pcDNA3.1+ was used) and 40 ng of Renilla luciferase plasmid (B). Activity of the promoter was tested against an empty PGL3b control and normalised to the empty vector control (A). To look at promoter activity in the presence of viral proteins, expression levels were normalised to ICP4p (B). Error bars represent Standard error of the mean.

P-values - * = <0.05, ** = <0.01, *** = <0.001, ns = Not significant

4.9 – Discussion

To date, very little work has been done looking at the role of ILTV-encoded miRNAs during infection and their potential targets. The work described in this chapter investigated the role of virus-encoded miRNAs targeting virally derived transcripts. Using a bio-informatics pipeline, potential viral targets of virally encoded miRNAs were first predicted using RNA Hybrid, an online programme that predicts the binding patterns between miRNAs and mRNA transcripts. The data generated from this bioinformatic screen was sorted and filtered which left a total of 227 predictions of interactions across the viral genome. An additional sort was then carried out to only include viral genes that were either essential to the virus (excluding capsid proteins or glycoproteins) or suppression of which may lead to a change in virus biology, this left a total of 28 potential viral miRNA targets that were taken forward to an initial reporter-based screen. As well as the 28 novel targets, a previously identified ILTV target/miRNA combination was included to ensure that the system chosen was working correctly. Using ILTV-miR-I6-5p and its known target in the virus IE gene ICP4, the system was shown to work with reduction in luciferase levels observed which were then abrogated upon mutation of the seed sequence. The 28 novel targets were subsequently cloned into the dual-luciferase vector Psi-Check 2 using pairs of long oligonucleotides of ~110 bp. Testing of the targets found three that had lower levels of luciferase activity when compared to the scrambled siRNA control however only two of these turned out to be statistically significant.

The first miRNAs were identified in the early 1990's in *C.elegans* with work some ten years later finding miRNAs that were encoded for by Epstein-Barr virus (EBV) and since then a large new field of virus-host interactions has opened up looking at the role of RNAi in virus infection (Lee et al., 1993, Pfeffer et al., 2004b). These interactions are not only looked at from the perspective of how the virus might manipulate the cell to gain a competitive advantage over the immune response organisms mount against invading pathogens but also in terms of how viruses can use small RNAs to regulate their own expression.

Virus encoded miRNAs have now been known about for nearly15 years and since then, upwards of 530 virally encoded miRNAs have been identified across 35 individual viruses according to the latest release of MiRBase (Release 22, March 2018) (Griffiths-Jones et al., 2006). This number relies upon researchers submitting their findings to the repository but with new viruses been discovered coupled with the tumbling cost of next-generation sequencing, new viral miRNAs are discovered much more frequently than they were previously and so this number whilst accurate to MiRBase may be different to the true number of actual virus-encoded miRNAs discovered to date.

RNA Hybrid is an online tool that works on the basis of looking for the 'best fit' between a miRNA and a possible mRNA target. Therefore, the sequences of the 10 known miRNAs of ILTV were taken from MiRBase and used to predict virus targets of virally encoded miRNAs. Due to the poor annotation of the viral genome, an arbitrary 1000 bp upstream of the AUG start site and 50 bp downstream of the designated PolyA signal were used to ensure that the full mRNA transcript was taken into account. Whilst this approach should cover the full transcripts of every open reading frame, it also has the complication that predicted miRNA targets can overlap between two or more viral transcripts. Herpesvirus genomes are compact and have very little space between open reading frames thus making it problematic when looking at binding patterns of miRNAs (Afonso et al., 2001). This was the case for several of the predicted targets and also for some of the targets taken forward into the luciferase based screen. Predicted target sequences in both UL29 and UL46 were also present in the UL28 and UL48 sequences respectively. For the former of these, the predicted site in UL28 was approximately 950 bp upstream of the AUG start site but was within the coding region of UL29. Because of this and the observation in the literature that 5'UTRs of HSV-1 vary between ~148 – 530 bp in length, the decision was taken that the site was in UL29 (Greco et al., 1994).

Conversely, the predicted site of ILTV-miR-I6-5p within UL46 & UL48 was more complex and required RT-PCR to decipher. Primer sets designed to amplify only from the UL48 transcript at the 3'UTR were designed whilst a further primer was designed to amplify only from the UL46 coding region (Figure – 4.7). This approach gave the result that the site was within the UL46 coding region as opposed to the 3'UTR of UL48. This of course does not guarantee that the target site is not within the 3'UTR of UL48 but more sensitive approaches would need to be taken such as 3' and 5' Rapid Amplification of cDNA Ends (RACE) to map the specific RNA transcripts of the respective open reading frames. Whilst this would have given a definitive answer, it was not within the aims of this project to map out gene transcripts.

Filtering and sorting of the data followed strict criteria to minimise the chances of false predictions been taken forward. Firstly, the decision was taken to omit any predictions that were above a minimum threshold of – 15 kcal/mol. This was due to previous findings in the lab which validated targets with minimum free energies of – 15.4 kcal/mol (Riaz, 2014). Duplex formation between a miRNA and mRNA is stronger when the minimum free energy is low thus in turn making it more thermodynamically stable requiring the need for more

energy to disrupt the duplex formation (Huang et al., 2010). Due to this, a large proportion of the predicted targets were immediately omitted and cut the number of predictions down.

These predictions were then filtered again using an approach based on what was known about the potential role of genes in virus biology. The core of this was to identify when in the virus life cycle the gene was expressed i.e. what temporal class of gene expression it was. Targets affecting immediate-early and early genes were considered arbitrarily to be likely of interest as opposed to late genes. As well as this, the function of the target was also investigated. As the genome of ILTV is again poorly annotated, homologues of transcripts were examined primarily in HSV-1 to see if they had any designated function and if they were classed as essential. From this, 28 novel targets were identified out of the 227 predictions following the initial filtering.

In addition to the 28 novel targets, one known viral target was included to ensure the luciferase based reporter system was working. ICP4, the essential major immediate-early protein is known to be targeted by both ILTVmiR-I6-5p and ILTV-miR-I5-5p (Waidner et al., 2011). These miRNAs have perfect Watson-crick binding complementarity to ICP4 as they are encoded anti-sense to the open reading frame. Only ILTV-miR-I6-5p was included in this study and not ILTV-miR-I5-5p. Data presented in this project found a 35 % reduction in luciferase expression when compared to a scrambled siRNA control and this was statistically significant ($p =<0.001$). This knockdown is greater than what was seen in the previous work that found a 23 % reduction in luciferase activity in the presence of ILTV-miR-I6 that was not statistically significant. Similarly, introduction of mutations to the miRNA seed sequence alleviated the effect of the miRNA which suggest a specific interaction. As the system was working with the replication of published data, the novel targets were subsequently tested.

Of the 28 novel targets tested, only three showed decreases in luciferase activity (outlined in Table – 4.2) and these were UL24 T1, UL28/29 T3 and UL46/48 T2. The former of these showed only a minor decrease in luciferase activity of 5 % which when tested statistically showed no significant difference from the scramble siRNA control (Figure – 4.4B). The data from these results meant that this work halted here as whilst an apparent difference was observed, it was not significant and a 5 % reduction is unlikely to have a large effect upon the virus biology. Intriguingly, they were several predicted targets in UL24 and whilst its function is still not fully understood, homologues are found across the herpesvirus family suggesting it does have important functions for the virus. Therefore, the significance of these predicted interactions may become apparent in future.

In comparison, the latter two predicted targets that were tested showed 15 % & 20 % reduction in luciferase activity respectively (Figures – 4.2C and 4.3C). Whilst again this was lower than that of the positive control for the reporter based assay (Figure – 4.1B), one must take into account that the target in ICP4 is perfectly complimentary to the miRNA and so the minimum free energy is much lower. Moreover, previous work in the lab by Dr. Riaz (The Roslin Institute) found knockdowns of around 20 % in a reporter based system (Riaz, 2014). The interactions were found to be specific as upon mutation of the miRNA seed sequence for both UL29 and UL46, an abrogation of the miRNA effect was seen with levels of the mutant sequence similar to that of the siRNA scramble and empty vector controls (Figures – 4.2C and 4.3C). This suggests that the interaction and observed knockdown is due to the action of the miRNA.

Whilst knockdown was seen in the reporter system, validation using alternative methods, specifically western blotting proved to be inconclusive. Firstly, both UL29 and UL46 coding regions were cloned into the expression plasmid pcDNA3.1+ with a small tag to allow for detection using antibodies. Cloning of these genes was successful as laid out in sections 4.6.2 and 4.7.2 yet both had their own separate problems when it came to validating the interaction of these transcripts with the miRNAs.

Firstly, UL29 was shown to be cloned successfully with a 6xHis tag *(Data Not Shown)*. Transfections of the UL29-6xHis plasmid into HEK293T cells alongside the mature miRNA mimic ILTV-miR-I2 or a siRNA scramble control were carried out before cells were harvested for protein by lysing with Laemmli buffer and boiling. These lysates were then used for conventional western blot as outlined in section 2.4 of this thesis. Upon imaging, no banding could be detected for UL29 in any of the lysate samples, including UL29-6xHis only lanes across multiple biological replicates. This could have simply been down to human error however across several biological repeats that is not plausible. Furthermore, mammalian and insect cells have high levels of histidine residues and so the antibody could have simply bound with more affinity to another protein, possibly explaining the banding pattern found at approximately 55 kDa (Figure – 4.5) (Kimple et al., 2013). Another possible cause of this lack of detection is the half-life of UL29. In previous experiments looking into the function of UL29 (ICP8) of HSV-1, detection of ICP8 via western blot was not possible in certain cell lines (Orberg and Schaffer, 1987). This may be the case here as validation work was carried out in HEK293T cells (of human origin) and the results may have differed if using a chicken cell line. To date, no explanation has solved why the protein could be not detected by immunoblot but it is of significance to finish validating this work as

it would shed light on to one of the functions of ILTV-miR-I2. As discussed in Chapter 3 of this thesis, biochemical approaches such as CLASH/HITS-CLIP may be able to aid in validating this target when the appropriate biological reagents are available.

In contrast, UL46-HA could be detected by western blot (Figure – 4.7A) following cloning of the coding region into pcDNA3.1+ *(Data Not Shown)*. Both a wild type UL46-HA and a mutant MUT-UL46-HA were constructed with the latter having the same mutations made as the reporter assay. Analysis of the protein levels by image studio (Li-Cor Biosciences) found there to be no difference in levels of the wild-type protein and the mutant protein in the presence of the miRNA mimic when compared to a protein only lane (Figure – 4.7B). This therefore did not validate the previous findings using the reporter assay and so one cannot categorically state that UL46 is targeted by ILTV-miR-I6-5p. In a similar manner to UL29, other biochemical approaches such as CLASH would be able to validate the findings if the interaction is real during virus infection.

Whilst the validation work for UL46 was ongoing, other viral genes were cloned as well as the ICP4 promoter region to investigate the interplay between UL46, UL48 and ICP4p. As briefly introduced in section 4.7.1 of this chapter, UL46 can modulate the effects of UL48 on ICP4 expression in HSV-2. Therefore, the interplay between these three proteins and ICP4p was investigated using a promoter assay system. As stated previously, ICP4p was cloned into PGL3b to drive the expression of firefly luciferase and following transfection of a chicken cell line a 129-fold increase in luciferase expression with ICP4p was observed when compared to the plasmid alone (Figure – 4.8A). This suggested that the promoter was active and could drive the expression of the reporter protein.

Co-transfection of the reporter plasmid in conjunction with either the plasmids expressing pUL46, pUL47 or pUL48 on their own or in a combination of the two was then carried out as described. In line with previous published data for HSV-2, transfection of pUL46 alongside ICP4p caused a decrease in luciferase activity suggesting an inhibition of ICP4p activity by pUL46. Kato et al. (2000) previously published similar findings on this interaction using HSV-2 gene products in a reporter based system. Transfection of pUL48 on its own with ICP4p showed no statistically significant difference from ICP4p on its own which differs from results seen in HSV-2 and other viruses (Kato et al., 2000). Also, co-transfection of both pUL46 and pUL48 did not result in an increase in ICP4p luciferase expression but in fact a decrease on ICP4p only levels however this was not statistically significant (Figure – 4.8B). pUL46 and pUL48 co-transfection showed an increase in ICP4p activity in HSV-2 in previously published reports (Kato et al., 2000). These data would

suggest that pUL46 of ILTV does modulate the interaction between pUL48 and ICP4p however it has a negative effect in ILTV when compared to the previously published data. As this is the opposite of what is seen in the literature, further investigation is warranted to see if this interaction is indeed correct in a more biologically relevant context as opposed to using a reporter system.

In summary, the use of bioinformatics and *in silico* approaches are complimentary to traditional biochemical approaches that are used to identify small RNA interactions however they also have many drawbacks. As found in this project, predicted miRNA targets that show protein knockdown in reporter based systems do not necessarily translate to real-world targets. In conjunction, predicted targets also don't necessarily have an effect when tested biochemically as found with a large proportion of the novel targets tested in the first part of this project following bioinformatic prediction. The basis of this project was to use bioinformatic approaches to identify viral targets of virus-encoded miRNAs and validate them using available biochemical methods. This was carried out as previous attempts to use biochemical based methods (CLASH) as outlined in chapter 3 of this thesis are technically very challenging with the currently available reagents. Despite this, investigating the RNAi pathway in the context of ILTV infection should still be of priority due to the relative lack of knowledge regarding the virus encoded miRNAs of ILTV. Follow up work investigating the interplay between pUL46, pUL48 and the ICP4 promoter also warrants further investigation as these data goes against what is seen in the published literature.

Chapter 5 – Screening the ILTV genome for Cellular miRNA target sites

5.1 – Introduction

MicroRNAs (miRNAs) have been shown to be effective regulators of gene expression. Endogenous cellular miRNAs are able to target a wide variety of cellular genes and subtly changing the signalling pathways. An increasing body of evidence is beginning to show that they also have the ability to effectively target viral transcripts in a wide range of differing virus families.

Hepatitis C Virus (HCV) infection sets off a cascade of antiviral responses through the interferon pathway resulting in the expression of 8 miRNAs that target the positive sense genome (miR-1, miR-30, miR-128, miR-196, miR-296, miR-351, miR-431 and miR-448) (Pedersen et al., 2007). Introduction into cells of synthetic mimics of these miRNAs replicated the antiviral response seen *in vitro*. Likewise, Coxsackievirus B3 (CVB3) is targeted by the host miRNA miR-342-5p in the 2C region leading to viral RNA degradation thus reducing both RNA and protein levels. The target site was found across CVB types 1-5 suggesting that miR-342-5p may also have an effect on other types of CVB viruses (Wang et al., 2012).

Larger, double stranded DNA viruses are also targets of the host RNAi pathway. The neuronal tissue specific miRNA hsa-miR-138 has been shown to target the immediate-early gene ICP0 of HSV-1. Mice infected with mutant HSV-1 with the miRNA target sites ablated were four times more likely to die from infection compared with mice infected with wild type virus (Pan et al., 2014). Likewise, the immediate-early gene UL112 of HCMV is targeted by the cellular encoded miRNA hsa-miR-200 family. The viral protein is critical in the reactivation of HCMV from latency thus the targeting promotes maintenance of viral latency (O'Connor et al., 2014). One of the other members of this miRNA family has also been implicated in the switch between latency and lytic replication in EBV. Both hsa-miR-200b and hsa-miR-429 have been shown to reduce the expression levels of ZEB1 and ZEB2 in EBV-positive cell lines resulting in a switch to lytic replication. This suggests both of these miRNAs are key regulators in the reactivation of EBV from latency (Ellis-Connell et al., 2010) (Lin et al., 2010).

Effects on virus infection by RNAi are not always necessarily as a direct result of a host miRNA targeting a viral mRNA transcript. Downregulation of cellular genes by host miRNAs can also have a significant effect upon viral infection. Cellular miRNAs hsa-miR-27a/b have been shown to negatively regulate the levels of SNAP25 and TXN2 which consequently causes an 80 % reduction in adenovirus genomic copy number in cells transfected with a synthetic mimic. Closer inspection showed reduced levels of SNAP25

was implicated in adenovirus entry into cells. In addition, reduced levels of TXN2 caused cell cycle arrest, thus blocking adenovirus replication which requires the cell cycle (Machitani et al., 2017). In some cases however, the targets can be both direct and indirect. The chicken miRNA gga-miR-130b has been shown to target the segment A portion of the Infectious Bursal Disease Virus (IBDV) genome causing a direct reduction in IBDV protein expression but in addition to this, expression of miR-130b also caused a downregulation in the levels of suppressors of cytokine signalling 5 (SOCS5) thus increasing the expression levels of IFN-β. Taken together, the results show a crucial role of miR-130b in host defence against IBDV infection (Fu et al., 2017).

The manipulation of the host environment by RNAi can either be advantageous or detrimental to a virus in both a productive infection but also for viruses which possess the ability to establish lifelong latency within infected hosts. These effects can be either direct or indirect to the virus with some highlighted cases using both pathways as a method of host defence. With more powerful laboratory methods being developed continuously in addition to the perpetual advancements in sequencing technologies, no doubt more of these interactions will be identified in the future. HITS-CLIP, PAR-CLIP and CLASH all have the ability to identify miRNA:mRNA interactions which includes cellular miRNAs targeting viral mRNAs. Complementary to this is the use of bioinformatical approaches to identify potential target sites for miRNAs which can then be tested in a traditional laboratory setting.

5.2 – Aims

The aims for this part of the project was to take the existing data set of viral transcripts created in Chapter 4 of this Thesis (Appendix 3) and screen them for miRNA binding sites using a list of high confidence chicken miRNAs taken from MiRBase 21 (released June 2014, [http://www.mirbase.org/]). Any identified target sites would then be investigated further using different biochemical approaches.

5.3 – The Use of RNA Hybrid to screen for candidates

RNA Hybrid parameters were set as per Chapter 3 except the minimum free energy (mfe) had a minimum requirement of -25 kcal/mol or below. All 79 ORFs were tested against a total of 54 mature miRNA sequences as taken from MiRBase (release 21, June, 2014). Similar to Chapter 3, predicted target sites were then ranked according to minimum free energy value. No exclusions were made with regards to gene function or temporal expression profile in this instance. This led to a final list of 103 target predictions of cellular miRNAs targeting viral transcripts. A summary of the predictions can be found in Table 5.1.

Table 5.1 – RNA Hybrid predictions using high confidence chicken miRNAs

Viral Gene	miRNA	Target Site	Predicted Region?	MFE	Gene Function
UL51	128-2-5p	589	5' UTR	-34.9	unknown
UL-1	222b-5p	270	5' UTR	-34.9	Unknown but essential
UL15a	125b-5p	986	5' UTR	-33.2	terminase/DNA Packaging
UL15b	125b-5p	4453	3' UTR	-33.2	terminase/DNA Packaging
UL36	128-2-5p	2091	CDS	-32.6	Major Tegument Protein
UL24	460b-3p	702	5'UTR	-32.2	unknown
UL1	10a-5p	89	5' UTR	-31.7	Glycoprotein gL
UL0	10a-5p	1670	CDS	-31.7	unknown
UL-1	30c-1-3p	1621	CDS	-30.9	unknown but essential
ICP4	133a-3p	584	5' UTR	-30.3	Gene Regulation
UL56	499-3p	415	5' UTR	-30.3	Vesicular Trafficking
UL22	17-3p	2261	CDS	-30.1	Glycoprotein gH
UL23	17-3p	3390	3’UTR	-30.1	thymidine kinase

US6	27b-5p	3288	3’UTR	-29.9	Glycoprotein gD
US7	27b-5p	2048	3’UTR	-29.9	Glycoprotein gI
US8	27b-5p	761	5' UTR	-29.9	Glycoprotein gE
UL42	133a-3p	1929	CDS	-29.8	process factor for DNA pol
UL43	133a-3p	546	5' UTR	-29.8	unknown
ORF A	128-2-5p	317	5' UTR	-29.6	unknown
UL7	222b-5p	484	5' UTR	-29.4	unknown
UL6	222b-5p	2402	CDS	-29.4	minor capsid protein
UL-1	125b-5p	600	CDS	-29.1	unknown but essential
sORF1	222b-5p	2919	CDS	-29.1	modulates UL48
US4	222b-5p	969	5' UTR	-29.1	Glycoprotein gG
UL44	22-3p	1091	CDS	-29.1	Glycoprotein gC
UL36	460b-5p	4173	CDS	-29	Major Tegument Protein
UL27	128-2-5p	510	5' UTR	-28.8	Glycoprotein gB
UL28	128-2-5p	2778	3’UTR	-28.8	ICP18.5
UL39	221-3p	1530	CDS	-28.8	Large-subunit Ribonucleotide Reductase

UL30	140-3p	29	5' UTR	-28.7	DNA POL
UL54	33-5p	2295	5' UTR	-28.6	post-transcriptional regulator of gene expression
ICP4	133a-3p	2563	CDS	-28.5	Gene Regulation
UL24	126-3p	3322	CDS	-28.3	unknown
UL25	126-3p	2482	CDS	-28.3	DNA Packaging Protein
UL26	126-3p	592	5' UTR	-28.3	Capsid Protein p40
UL51	34b-5p	1555	CDS	-28.3	unknown
ORF C	34b-5p	1345	CDS	-28.3	unknown
UL21	460b-3p	859	5' UTR	-28.3	nucleocapsid protein
UL26	128-2-5p	1797	CDS	-27.8	Capsid Protein p40
UL26.5	128-2-5p	927	5' UTR	-27.8	virion scaffold protein
UL15b	187-5p	3132	3’UTR	-27.8	terminase/DNA Packaging
UL1	10b-5p	89	5' UTR	-27.7	Glycoprotein gL
UL0	10b-5p	1670	CDS	-27.7	unknown
UL31	133a-3p	134	5' UTR	-27.7	nuclear phosphoprotein
UL32	133a-3p	1873	CDS	-27.7	envelope glycoprotein

ORF C	29b-1-5p	1197	CDS	-27.7	unknown
UL22	34b-5p	3540	3'UTR	-27.7	Glycoprotein gH
UL29	222b-5p	457	5' UTR	-27.4	Major Single Strand DNA binding protein
ICP4	181b-5p	1968	CDS	-27.3	Gene Regulation
UL15b	221-5p	1430	CDS	-27.3	terminase/DNA Packaging
UL14	460b-3p	773	5' UTR	-27.3	unknown
UL13	460b-3p	322	5' UTR	-27.3	protein kinase
UL7	128-3p	1506	CDS	-27.2	unknown
UL6	128-3p	3424	3'UTR	-27.2	minor capsid protein
UL36	460b-3p	7436	CDS	-27	Major Tegument Protein
UL5	187-3p	1533	CDS	-26.9	Helicase-primase component
ICP4	30c-2-3p	1109	CDS	-26.9	Gene Regulation
UL39	125b-5p	929	5' UTR	-26.7	Large-subunit Ribonucleotide Reductase
UL36	133a-3p	418	5' UTR	-26.7	Major Tegument Protein
UL37	133a-3p	3712	CDS	-26.7	Tegument Protein
ICP4as	181b-5p	1969	CDS	-26.7	Gene Regulation

sORF1	30c-1-3p	1799	CDS	-26.7	modulates UL48
ORF E	34b-5p	1072	CDS	-26.7	unknown
UL23	34b-5p	4669	3’UTR	-26.7	thymidine kinase
UL36	460b-5p	8070	3' UTR	-26.7	Major Tegument Protein
UL52	128-2-5p	3032	CDS	-26.6	DNA Helicase-Primase
US5	128-2-5p	1657	CDS	-26.6	Glycoprotein gJ
UL29	187-3p	2793	CDS	-26.6	Major Single Strand DNA binding protein
UL50	30c-1-3p	1113	CDS	-26.3	Deoxyuridine triphosphate
UL20	133a-3p	2902	CDS	-26.1	membrane protein
UL19	133a-3p	2122	CDS	-26.1	major capsid protein
UL38	30c-1-3p	2026	CDS	-26.1	DNA binding/Capsid Protein
UL39	30c-1-3p	449	5' UTR	-26.1	Large-subunit Ribonucleotide Reductase
UL41	187-3p	1350	CDS	-26	VHS
UL20	221-3p	659	5' UTR	-26	membrane protein
ICP4	30a-5p	3708	CDS	-25.9	Gene Regulation
ICP4	30c-5p	3685	CDS	-25.9	unknown

UL15a	Let-7	1425	CDS	-25.9	teminase/DNA packaging
UL15b	Let-7	4893	CDS	-25.9	terminase/DNA Packaging
UL33	221-3p	596	5' UTR	-25.8	DNA Packaging
UL34	221-3p	161	5' UTR	-25.8	Membrane phosphoprotein
UL38	187-5p	1553	CDS	-25.7	DNA binding/Capsid Protein
UL14	33-3p	786	5' UTR	-25.7	unknown
UL13	33-3p	335	5' UTR	-25.7	protein kinase
UL24	22-3p	2043	CDS	-25.6	unknown
UL25	22-3p	1203	CDS	-25.6	DNA Packaging Protein
ORF C	30c-1-3p	1683	3’UTR	-25.6	unknown
UL41	499-3p	1763	CDS	-25.5	VHS
ICP4	140-3p	4923	CDS	-25.4	Gene Regulation
ICP4	126-3p	3153	CDS	-25.3	Gene Regulation
UL36	22-3p	158	5' UTR	-25.2	Major Tegument Protein
UL37	22-3p	3452	CDS	-25.2	Tegument Protein
UL3.5	10a-5p	1091	CDS	-25.1	Virion Morphogenesis

UL3	10a-5p	1786	CDS	-25.1	unknown
UL3.5	10b-5p	1091	CDS	-25.1	Virion Morphogenesis
UL3	10b-5p	1786	CDS	-25.1	unknown
US5	222b-5p	1878	CDS	-25.1	Glycoprotein gJ
UL44	34b-3p	569	5' UTR	-25.1	Glycoprotein gC
UL42	34b-5p	3852	CDS	-25.1	process factor for DNA pol
UL43	34b-5p	2469	CDS	-25.1	unknown
UL17	10a-3p	207	5' UTR	-25	Tegument Protein
UL20	429-3p	1704	CDS	-25	membrane protein
UL19	429-3p	924	5' UTR	-25	major capsid protein

Target site is the first base pair where the miRNA interacts with the miRNA starting from ‘0’ 1000 bp upstream of the AUG start codon
Predicted region = 5’UTR – 5’ Untranslated region, 3’UTR – 3’ Untranslated region, CDS – Protein coding sequence

5.4 – Screening of RNA Hybrid Predictions by Luciferase Assay

From the sorted list of predicted targets, 3 were chosen that were all predicted to be targeted by the cellular miRNA gga-miR-133a-3p. Two of these were contained within ICP4, a known essential immediate-early gene and one was within UL20, an ORF encoding for a membrane protein. Some viral genes had more than one target for the same miRNA as was the case for ICP4.

Long Oligonucleotide probes were ordered, annealed and cloned into the reporter plasmid Psi-Check-2 using restriction digest (see appendix 1 for vector maps). Positive clones were screened using colony PCR as described and one positive clone was then subsequently picked, amplified and sequenced to confirm the presence of the cloned region containing the predicted miRNA target site (*Data Not Shown*). Luciferase assays were then carried out as described in section 2.6.3. Data shown in Figure -5.1 A, B & C show the combined results of two independent assays with n = 8 biological replicates and n = 24 technical replicates.

Two targets, gga-ICP4-T2 and gga-UL20 showed no statistical difference when compared to an empty vector control in the presence of miR-133a-3p (Figure - 5.1 A & B). However, they were statistically significant from the siRNA scramble control ($p =<0.001$). In contrast, one target, gga-ICP4-T1 showed a statistically significant reduction in luciferase activity when compared to the empty vector control in the presence of gga-miR-133a-3p ($p =<0.01$). There was also a statistically significant difference ($p =<0.001$) between the siRNA scramble control and both the empty vector control and gga-ICP4-T1 plasmid. Upon mutation of the seed sequence contained within the gga-ICP4-T1 plasmid, luciferase levels were restored to similar levels (no significant difference) as the empty vector control (Figure - 5.1 C).

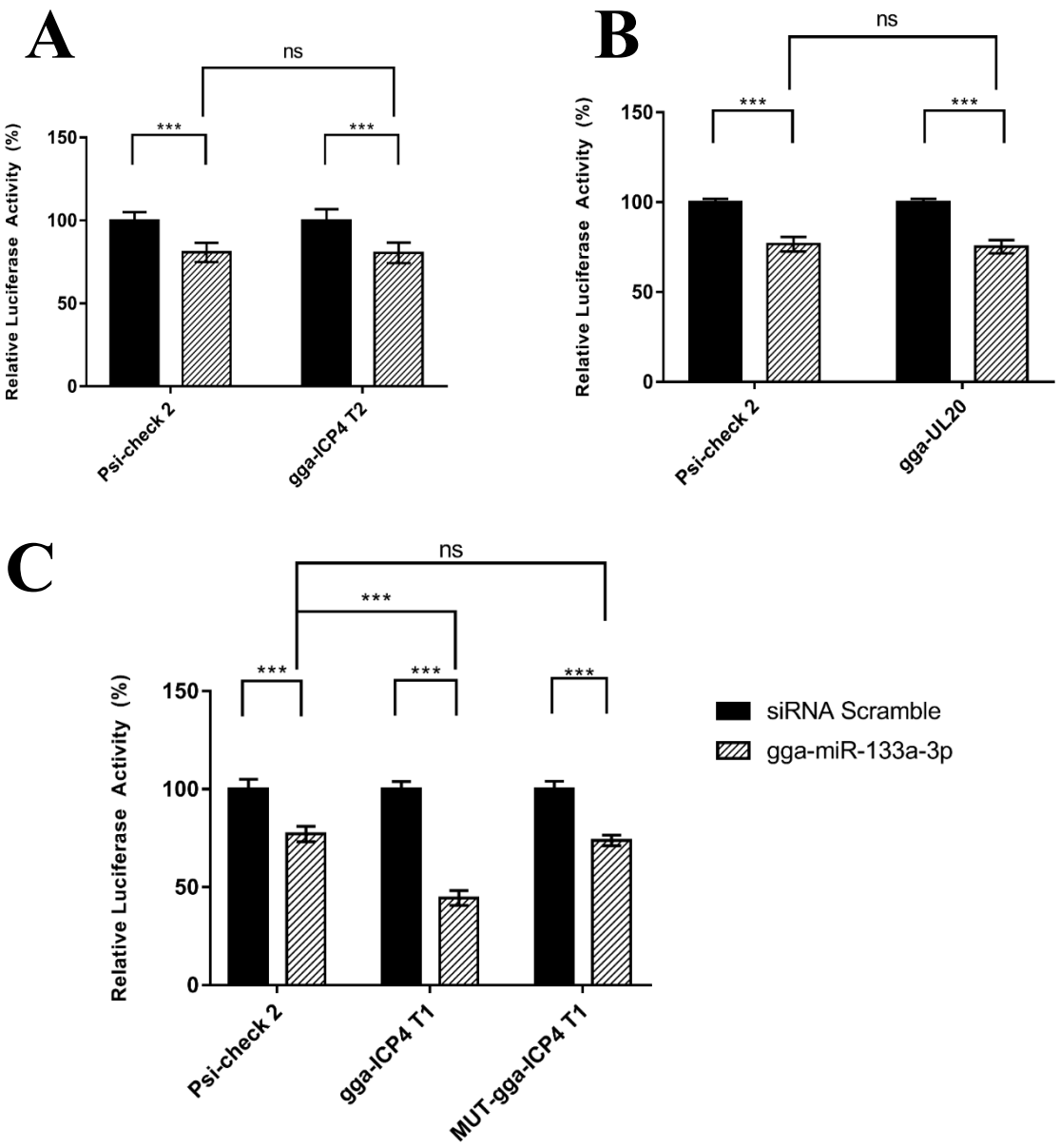


Figure - 5.1 – Relative luciferase expression levels of cloned gga-miR-133a-3p targets compared to a scrambled siRNA

HEK293T cells were transfected with a Psi-Check-2 plasmid containing a ~110 bp region of ILTV genomic DNA harbouring the predicted gga-miR-133a-3p target sites; or with the miRNA site mutated. After 48 hrs, *Renilla* luciferase levels were measured and normalised to firefly luciferase levels and expression in control and test miRNA compared. Three constructs were tested; A – gga-ICP4-T2, B – gga-UL20 and C - gga-ICP4-T1. N = 8 Biological replicates and N = 24 technical repeats. Error bars display standard error of the mean (SEM).
*** = $p<0.0001$, ns = not significant

5.5 – RT-PCR analysis of the 5'UTR of ICP4

The miR-133a-3p target site starts 416 bp upstream of the predicted AUG codon and it was not clear if it would be incorporated into the mRNA transcript of ICP4. Therefore a series of RT-PCRs (see section 2.2.7) were set up with a reverse primer positioned approximately 20 bp downstream of the AUG codon and a panel of forward primers which were progressively further away from start codon. RNA was extracted from LMH cells infected with ILTV and from this cDNA was synthesised using both Oligo (DT) primer and random primers as per section 2.2.5. The resulting cDNA was used for amplification of the region alongside an uninfected control, -RTs, a viral DNA positive control and a no template control. Initially, forward primers were designed to bind 1000, 450 and 100 bp upstream of the AUG start site. Both 1000 and 450 bp primer locations did not produce a band from cDNA synthesised with random primers of oligo DT primers (designated primer set 1 & 2 respectively) when analysed using agarose gel electrophoresis (Figure – 5.2 B & C). However, the DNA positive control produced a band of the predicted size. Conversely, primer set 3 (situated 100 bp upstream) did produce bands in both random primed and oligo DT primed cDNA that were of the same size as the DNA positive control (Figure – 5.2 D).

This prompted a closer inspection of the area surrounding the miRNA target site that identified a potential TATA box region upstream of the target site. Due to this finding, a fourth primer was designed which was approximately 420 bp upstream of the ATG and encompassed some of the predicted binding site for the miRNA but crucially not the seed sequence. Upon analysis, a band was visualised in the random primed infected cDNA but not in the oligo DT primed cDNA. The band visualised was at the same size as the DNA positive control (Figure – 5.2 E).

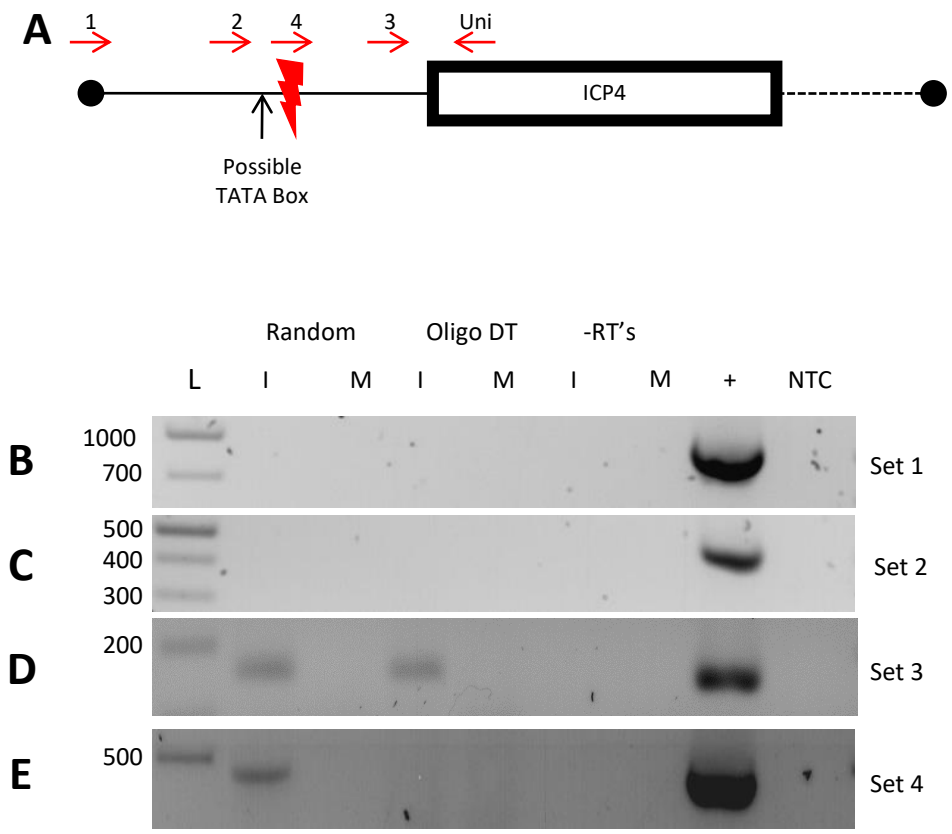


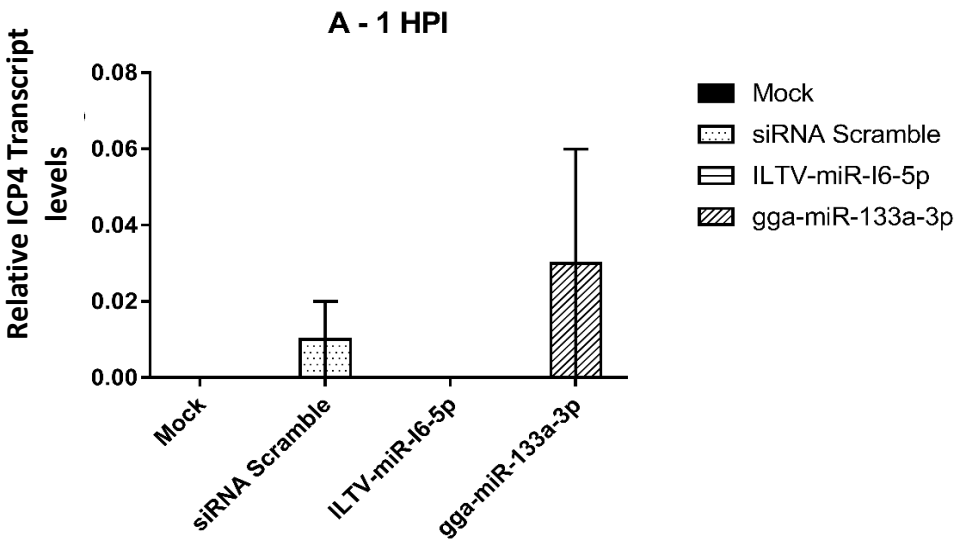
Figure - 5.2 – Detection of 5'UTR PCR products of ICP4 using RT-PCR
Primers were designed at 1000, 450 and 100 bp upstream of the ATG start site with a universal reverse primer situated 50 bp downstream. RNA was extracted from either ILTV infected cells or mock infected cells and DNase treated. cDNA was synthesised using either random primers or Oligo dT. PCRs were run in conjunction with minus reverse transcription (-RT) and viral DNA positive controls. PCR products were visualised on 2 % agarose gel and imaged. Lightning bolt represents the predicted miR-133a target site, red arrows depict primer location and direction. 'Uni' = universal reverse primer.
A – Diagrammatic representation of the viral transcript showing approximate primer locations. B – Primer set one. C – Primer set two. D – Primer set three. E – Primer set four.

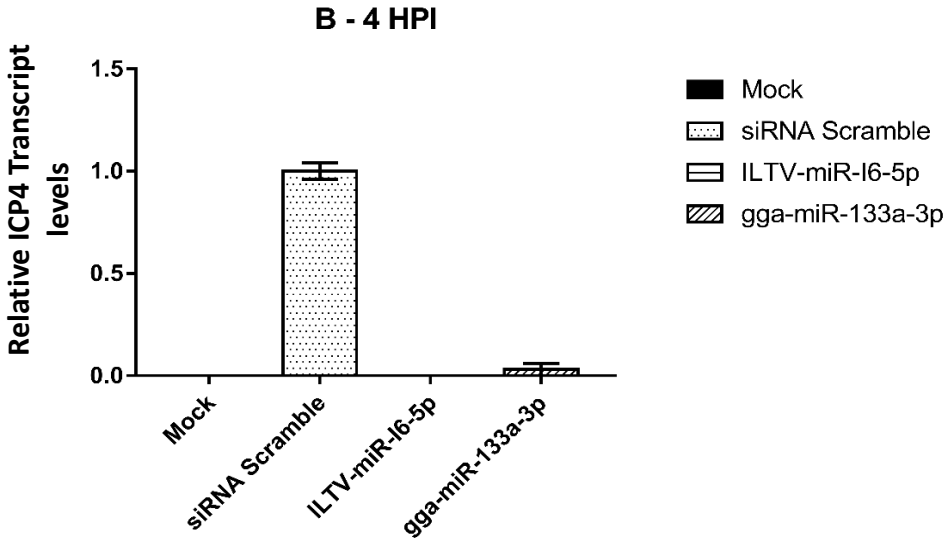
5.6 – The Effect of miR-133a on Virus Replication?

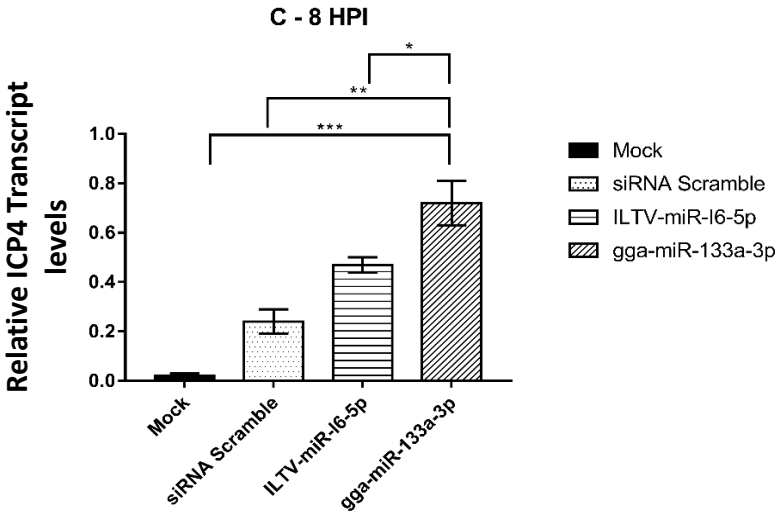
Virus infection studies were conducted to determine if the effects seen in the luciferase reporter assays translated into a more biologically relevant setting. LMH cells were transfected with gga-miR-133a-3p, ILTV-miR-I6-5p or a siRNA scramble alongside a mock control. After 12 hours transfection incubation, cells were infected with cell free virus (CSW-1) at an MOI of 0.1. Samples were taken at 1, 4, 8, 12 and 24 hours post infection. Supernatant taken from wells was then used for conventional plaque assay analysis whilst cells were harvested and processed for RT-qPCR analysis of ICP4 transcript levels.

5.6.1 – Levels of ICP4 mRNA transcript from virus grown in transfected cells

Primers targeting the viral gene ICP4 were designed with the help of Dr. W Tan, The Roslin Institute as previously published primers for ICP4 (Mahmoudian et al., 2012) were found to amplify products in uninfected cells. Extracted RNA was DNase treated and cDNA synthesised using random primers. Synthesised cDNA was then analysed by qPCR to determine the relative levels of ICP4 transcript. The 2ΔΔCt method of analysis [as outlined by (Livak and Schmittgen, 2001a)] was used to determine the relative fold change of ICP4 transcripts compared to the house keeping gene chicken GAPDH. In total, 3 biological replicates (N = 3). Error bars display standard error of the mean (SEM). Generally, apparent RNA levels of ICP4 were unaffected by the introduction of miR-133a compared to the other conditions across the whole experiment apart from at specific time points (data summarised in Table – 5.2.). At 8 HPI, levels of ICP4 transcript were significantly higher compared to both siRNA scramble and ILTV-miRI6-5p ($p = <0.05$). At subsequent time points, no significant difference could be seen between the three conditions however levels of ICP4 mRNA at 24 HPI were apparently higher in miR-133a samples.


A - 1 HPI
Relative ICP4 Transcript levels
0.08
0.06
0.04
0.02
0.00
Mock
siRNA Scramble
ILTV-miR-I6-5p
gga-miR-133a-3p
Mock
siRNA Scramble
ILTV-miR-I6-5p
gga-miR-133a-3p


B - 4 HPI
Relative ICP4 Transcript levels
1.5
1.0
0.5
0.0
Mock
siRNA Scramble
ILTV-miR-I6-5p
gga-miR-133a-3p
Mock
siRNA Scramble
ILTV-miR-I6-5p
gga-miR-133a-3p


C - 8 HPI
Relative ICP4 Transcript levels
*
**
***
1.0
0.8
0.6
0.4
0.2
0.0
Mock
siRNA Scramble
ILTV-miR-I6-5p
gga-miR-133a-3p
Mock
siRNA Scramble
ILTV-miR-I6-5p
gga-miR-133a-3p

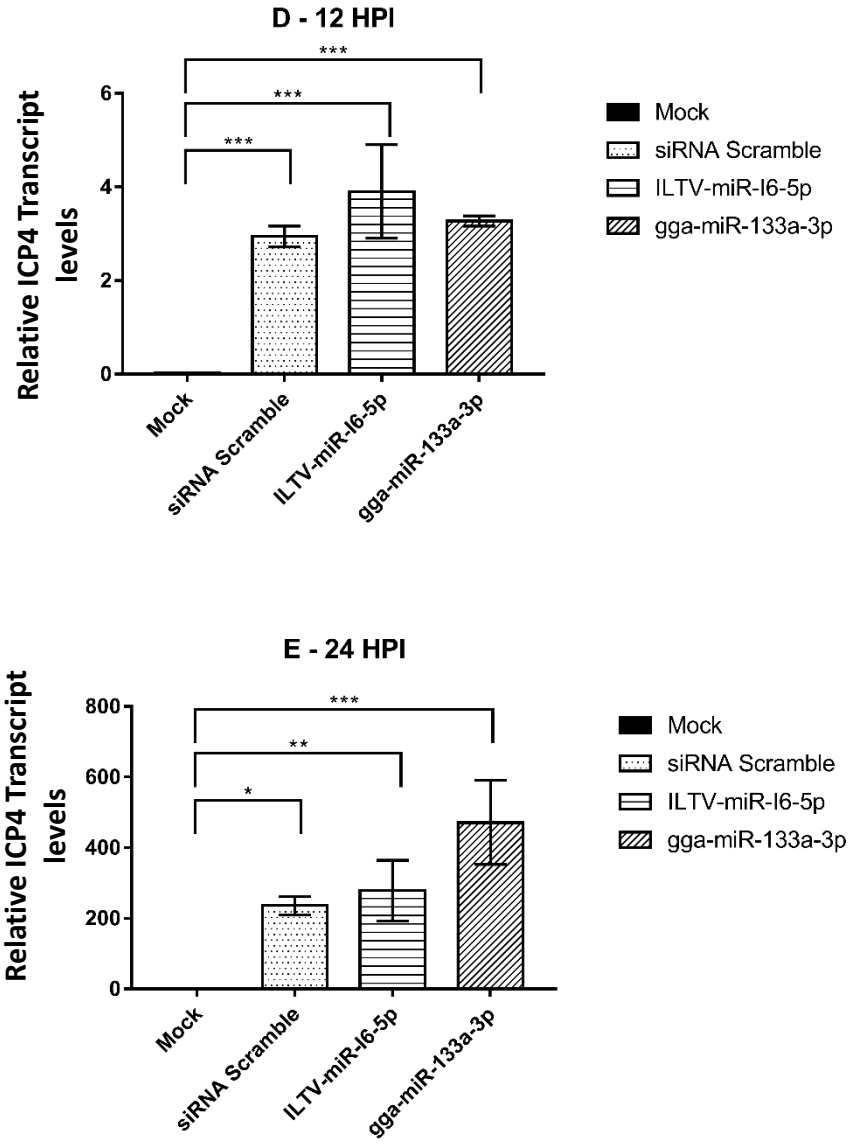


Figure – 5.3. Relative expression of ICP4 transcripts in differing conditions in a time-dependent manner

LMH cells were transfected with either a scrambled siRNA, ILTV-miR-I6-5p or gga-miR-133a-3p alongside a mock control. 12 hours after, transfected cells were infected with live CSW-1 ILTV at an MOI of 0.1. Mock transfected cells were mock infected. Cells were harvested at 1, 4, 8, 12 and 24 HPI. Extracted RNA was used for RT-qPCR analysis of ICP4 transcripts using the 2ΔΔCT method with GAPDH used as control. N = 3 biological replicates. Error bars display standard error of the mean (SEM).

P-values - * = <0.05, ** = <0.01, *** = <0.001

A – 1 HPI. B – 4 HPI. C – 8 HPI. D – 12 HPI. E – 24 HPI.

Table – 5.2. Relative expression of ICP4 transcripts in differing conditions in a time-dependent manner

Time Point (HPI)	Condition	Mean Expression of ICP4 transcript	Standard Error	Post Hoc Tukey's Grouping
1	Mock	0.00	0.00	A
	Scramble	0.01	0.01	A
	miR-I6-5p	0.00	0.00	A
	miR-133a	0.02	0.04	A
4	Mock	0.00	0.00	A
	Scramble	0.10	0.04	A
	miR-I6-5p	0.00	0.00	A
	miR-133a	0.03	0.03	A
8	Mock	0.02	0.01	A
	Scramble	0.24	0.05	A B
	miR-I6-5p	0.47	0.03	B
	miR-133a	0.72	0.09	C
12	Mock	0.02	0.01	A
	Scramble	2.94	0.22	B
	miR-I6-5p	3.90	1.00	B
	miR-133a	3.27	0.11	B
24	Mock	0.09	0.02	A
	Scramble	236.1	25.50	A B
	miR-I6-5p	278.1	86.01	B
	miR-133a	471	119.06	B

Conditions that do not have the same letter are statistically significant from one and other. Experiment was carried out N = 3 Biological repeats

5.6.2 – Titres of ILTV following growth of virus in transfected cells

Harvested supernatant was used for conventional plaque assays as described previously. Following fixing and staining, plaques were counted and the titre calculated (pfu/ml). Experiment was carried out in N = 7 biological repeats except for miR-133a which was N = 6 due to one replicate unable to be counted. Plaques were formed from supernatant harvested at 1 HPI and data analysis found there to be a significant difference between the scrambled siRNA and miR-133a ($p =<0.05$). No significant difference could be determined between the scrambled siRNA or miR-I6-5p and miR-I6-5p and miR-133a (Figure – 5.5A). The reason for this production of plaques harvested from 1 HPI supernatant samples is unclear as virus replication will not have happened at 1 hour. Both 4 and 8 HPI samples showed no statistical difference between the 3 conditions (Figure – 5.5B&C). Similarly, in the 12 HPI there was no statistical evidence for difference between the 3 conditions however there was a trend towards lower virus titres in both I6-5p and miR-133a compared to the scrambled siRNA (Figure – 5.5D). This trend continued at 24 HPI with even clearer difference in absolute pfu/ml. Virus that was incubated in cells transfected with 133a showed a virus titre which was 66 % lower than that of the siRNA control or 42 % lower than that of I6-5p (Figure – 5.5E). Furthermore, I6-5p showed a decreased viral titre compared to the siRNA control of around 25 %. This was not statistically significant when tested using a one-way ANOVA combined with a post-hoc Tukey's comparison test (Table – 5.3E). A summary of the viral titre including Post-Hoc statistical testing can be seen in Table 5.3.

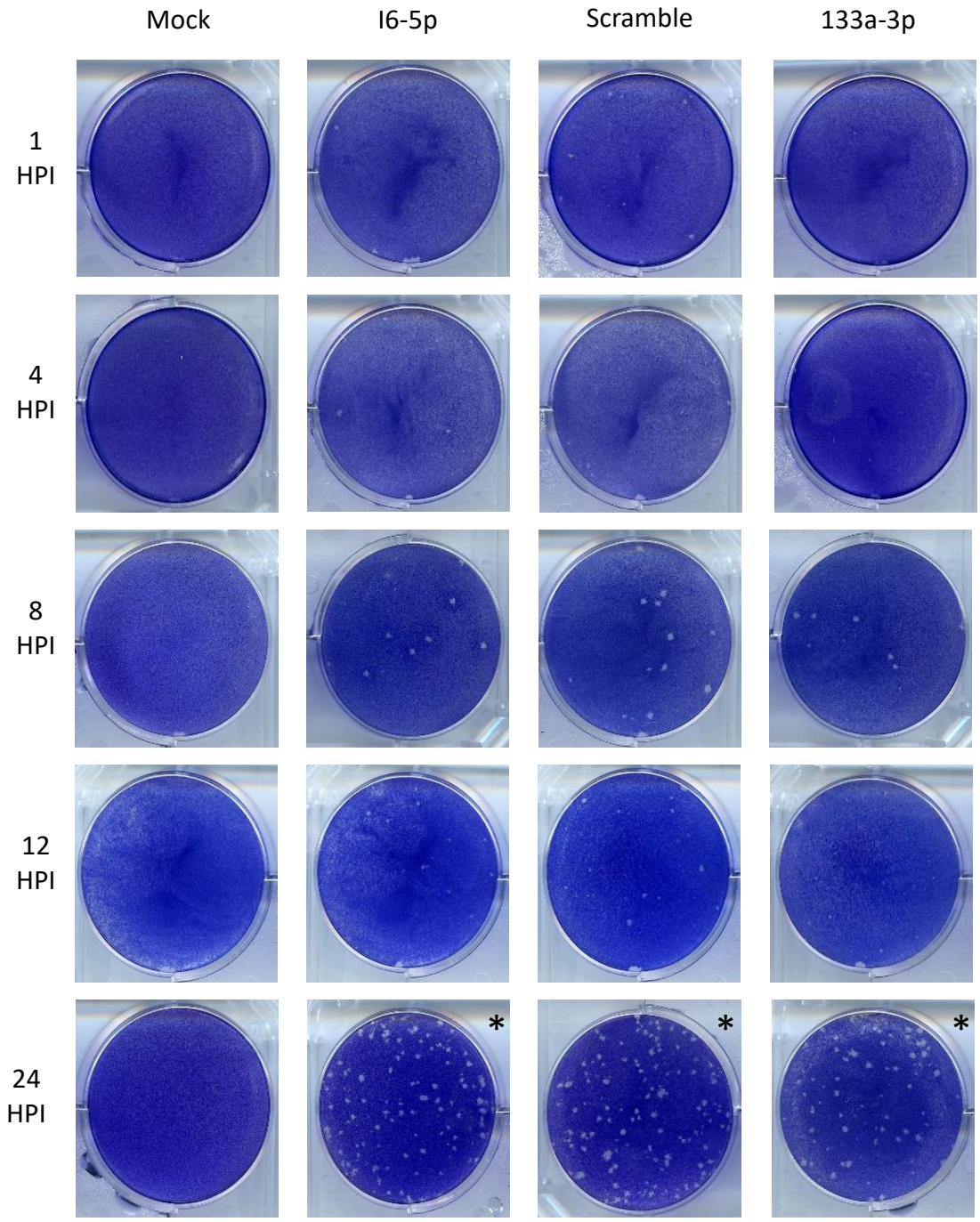
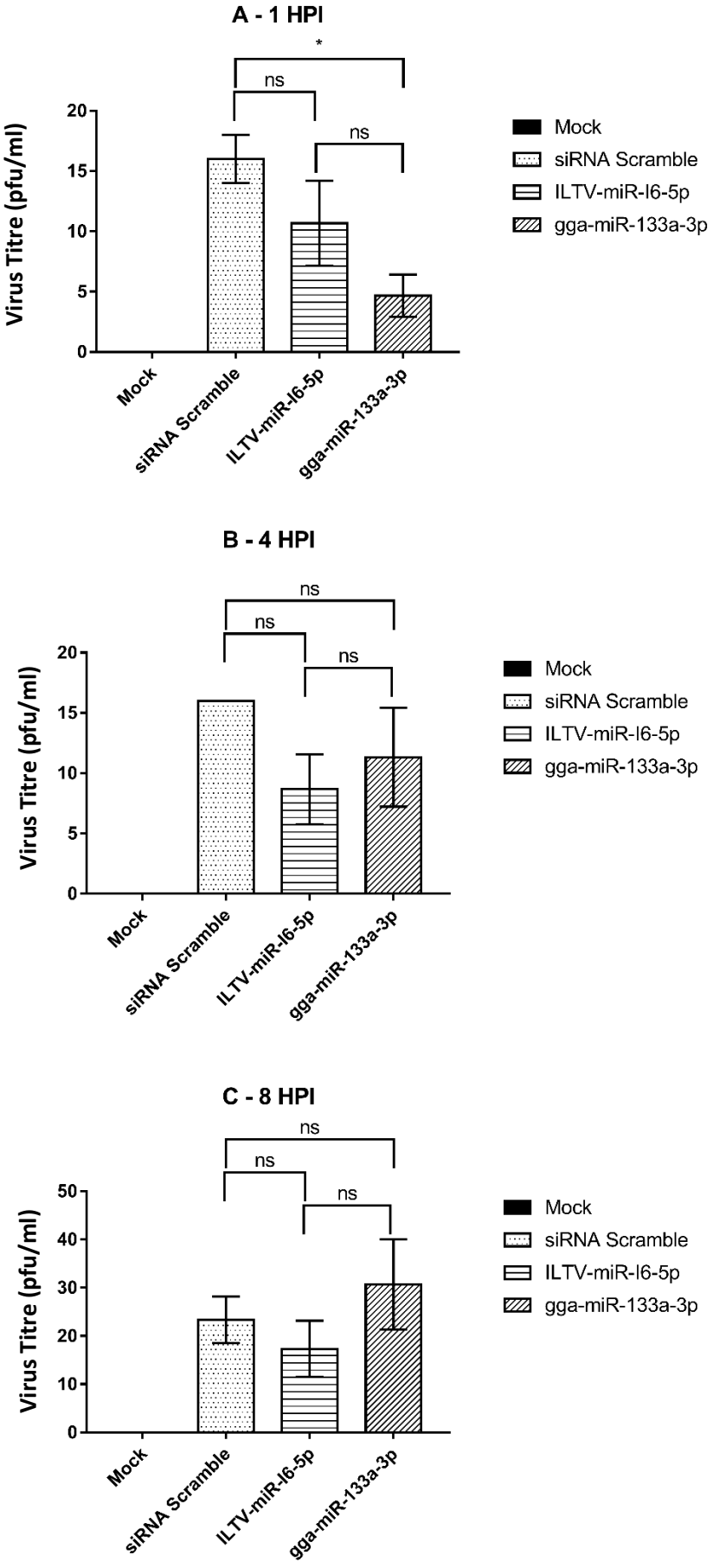


Figure – 5.4. Assessment of virus supernatant infectivity by plaque assay

LMH cells were seeded out at $1x10^{6}$ cells per well and infected with supernatant harvested from transfected/infected cells. Samples were left for 72 hours before been fixed by 10 % non-buffered formalin (NBF) and stained with 0.1 % toluidine blue. Plates shown here are representative of experiment. Images with an asterisk are at 10^{-2} dilution whilst the rest are neat samples.


A - 1 HPI
*
ns
ns
Virus Titre (pfu/ml)
20
15
10
5
0
Mock
siRNA Scramble
ILTV-miR-I6-5p
gga-miR-133a-3p
Mock
siRNA Scramble
ILTV-miR-I6-5p
gga-miR-133a-3p
B - 4 HPI
ns
ns
ns
Virus Titre (pfu/ml)
20
15
10
5
0
Mock
siRNA Scramble
ILTV-miR-I6-5p
gga-miR-133a-3p
Mock
siRNA Scramble
ILTV-miR-I6-5p
gga-miR-133a-3p
C - 8 HPI
ns
ns
ns
Virus Titre (pfu/ml)
50
40
30
20
10
0
Mock
siRNA Scramble
ILTV-miR-I6-5p
gga-miR-133a-3p
Mock
siRNA Scramble
ILTV-miR-I6-5p
gga-miR-133a-3p

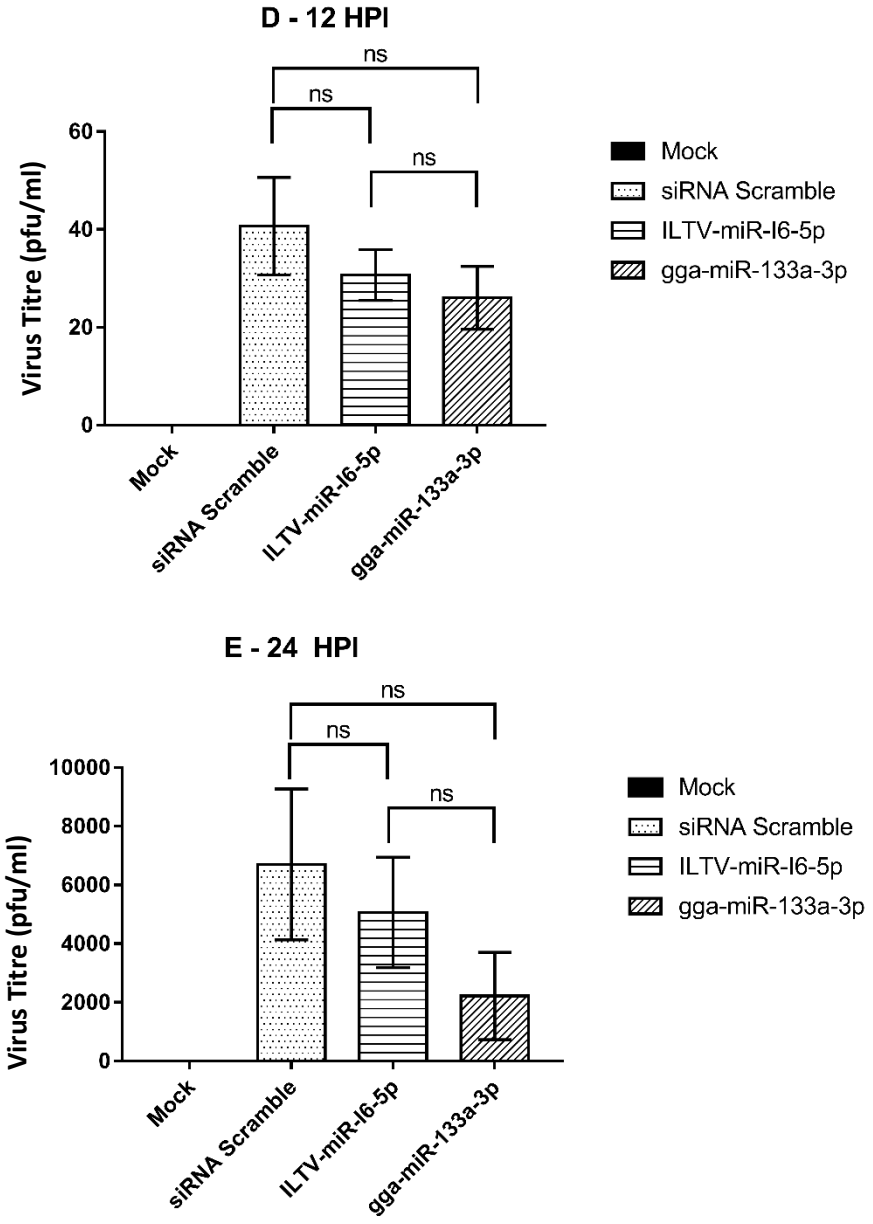


Figure – 5.5. Viral titres of ILTV incubated with a scrambled siRNA, ILTV-miR-I6-5p or gga-miR-133a-3p

Plaque plates as shown in Figure 5.4 were each individually counted. From this, an average PFU/ml was calculated for each condition at every time point. N = 7 biological repeats except for miR-133a which N = 6. Error bars represent standard error of the mean (SEM).

P-values = * - <0.05, ns = Not significant

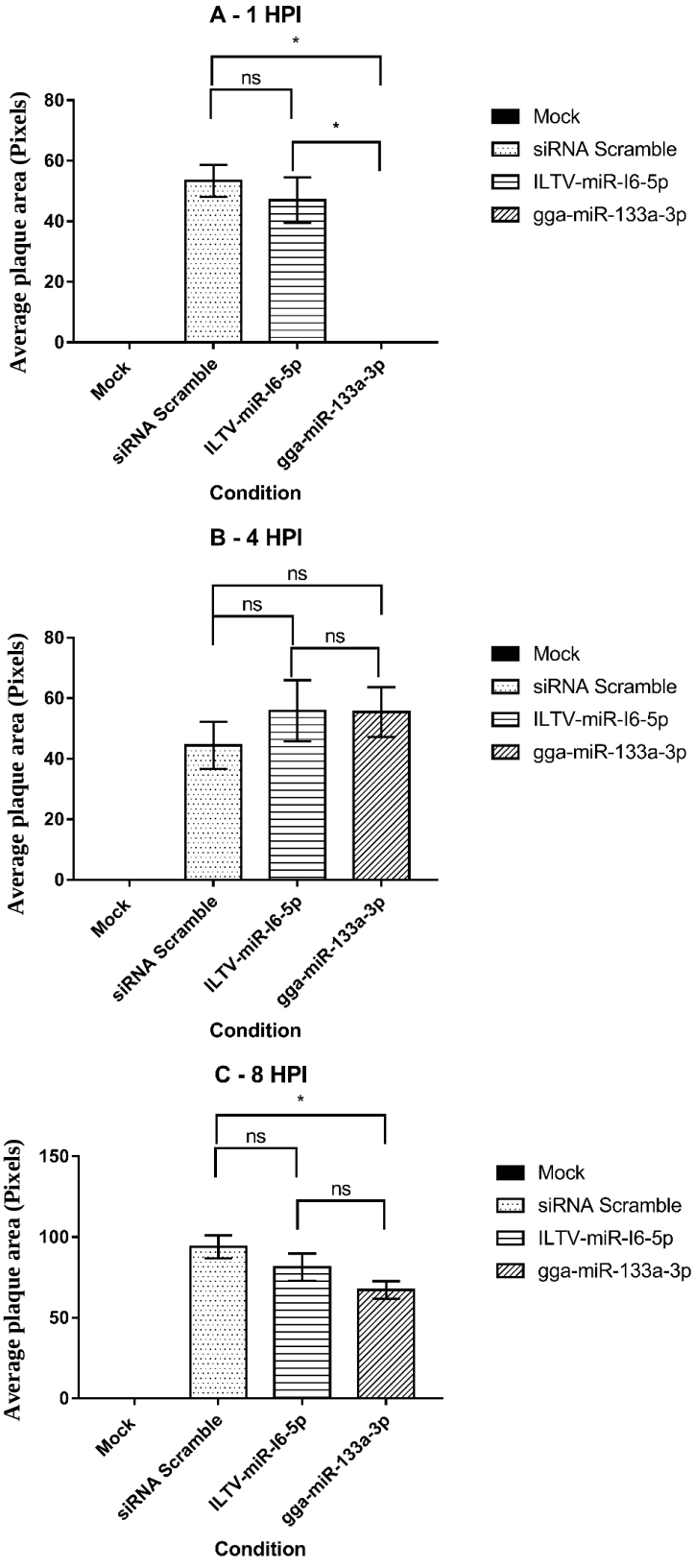
Table – 5.3. Viral titres of ILTV incubated with a scrambled siRNA, ILTV-miR-I6-5p or gga-miR-133a-3p

Time Point (HPI)	Condition	Mean (PFU/ml)	Standard Error	Post Hoc Tukey's Grouping
1	Mock	0	0.00	A
	Scramble	16	2.00	B
	miR-I6-5p	10.67	3.52	B C
	miR-133a	4.67	1.76	C
4	Mock	0	0.00	A
	Scramble	16	0.00	B
	miR-I6-5p	8.67	2.90	A B
	miR-133a	11.33	4.05	B
8	Mock	0	0.00	A
	Scramble	23.33	4.80	A B
	miR-I6-5p	17.33	5.81	A B
	miR-133a	30.67	9.33	B
12	Mock	0	0.00	A
	Scramble	40.67	9.95	B
	miR-I6-5p	30.67	5.20	B
	miR-133a	26	6.42	A B
24	Mock	0	0.00	A
	Scramble	6701	2573.50	A
	miR-I6-5p	5063	1880.26	A
	miR-133a	2217	1483.35	A

Conditions that do not have the same latter are statistically significant from one and other.
Experiment was carried out n=7*
*133A at 24HPI was n=6

5.6.3 – Plaque size and morphology produced by ILTV harvested from transfected cells

Plaque size was measured using Image J software. A minimum of 10 plaques were counted per replicate per condition which were then combined to give an average plaque area per condition. From this data set, statistical analysis was carried out using a one-way ANOVA with post-hoc Tukey's comparisons test. If ten plaques could not be counted per replicate, per condition, an n = 0 was put into the data set. Times stated all refer to when the original virus was harvested (e.g. 1 HPI) and not how long plaque assays were incubated for. For the 1 HPI samples, not enough plaques could be counted for the miR-133a treated cells. There was no significant difference between the siRNA scramble and miR-I6-5p however (Figure – 5.6A). No statistically significant difference was observed with the 4 HPI samples between the three conditions but there was a significant difference between all three conditions and the mock control ($p =<0.001$) (Figure – 5.6B). For the 8 HPI samples however, a significant difference was observed between miR-133a and the siRNA control ($p =<0.05$) yet there was no difference between I6-5p and miR-133a (Figure – 5.6C). At 12 HPI, the same pattern was observed with no difference between 133a and I6-5p but there was a significant difference between 133a and the siRNA scramble ($p =<0.05$) (Figure – 5.6D). Analysis revealed a statistically significant reduction in plaque size ($p =<0.001$) with samples taken at 24 HPI from 133a-transfected cells when compared to either I6-5p or the siRNA control (Figure – 5.6E). A summary of plaque size and Post-Hoc testing is can be seen in Table 5.4.


A - 1 HPI
*
ns
*
Average plaque area (Pixels)
80
60
40
20
0
Mock
siRNA Scramble
ILTV-miR-I6-5p
gga-miR-133a-3p
Condition
Mock
siRNA Scramble
ILTV-miR-I6-5p
gga-miR-133a-3p
B - 4 HPI
ns
ns
ns
Average plaque area (Pixels)
80
60
40
20
0
Mock
siRNA Scramble
ILTV-miR-I6-5p
gga-miR-133a-3p
Condition
Mock
siRNA Scramble
ILTV-miR-I6-5p
gga-miR-133a-3p
C - 8 HPI
*
ns
ns
Average plaque area (Pixels)
150
100
50
0
Mock
siRNA Scramble
ILTV-miR-I6-5p
gga-miR-133a-3p
Condition
Mock
siRNA Scramble
ILTV-miR-I6-5p
gga-miR-133a-3p

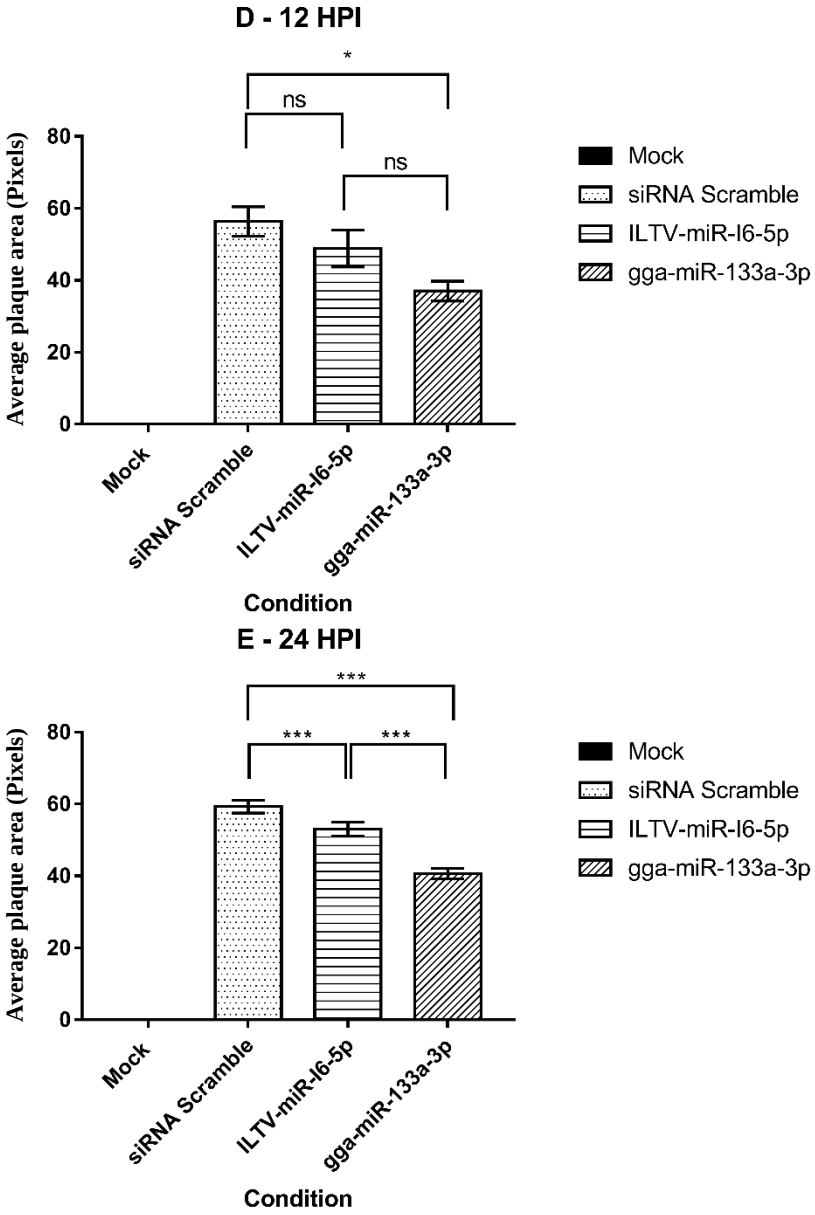


Figure – 5.6. Average plaque area caused by ILTV after incubation with a scrambled siRNA, ILTV-miR-I6-5p or gga-miR-133a-3p

Plaque assays were carried out as previously described. Plates were scanned and images were then used to measure the plaque area of individual plaques using ImageJ software. A minimum of 10 plaques was counted per plate per condition and replicate. Data comprises of two independent experiments with a total of n=7 biological replicates per condition (*n=6 for miR-133a at 24 HPI). Average area was calculated from the combination of replicate plates for each condition. A one-way ANOVA was used for analysis followed by a post-hoc Tukey's test. Error bars display the standard error of the mean (SEM).

P-values - * = <0.05, ** = <0.01, *** = <0.001

Table – 5.4. Average plaque size caused by ILTV virus incubated in different conditions in a time dependent manner

Time Point (HPI)	Condition	N	Mean Plaque Size (Pixel Area)	Standard error	Post Hoc Tukey's Grouping
1	Mock	0	0	0.00	A
	Scramble	21	53.43	5.31	B
	miR-I6-5p	13	47.00	7.52	B
	miR-133a	9	0	0.00	A
4	Mock	0	0	0.00	A
	Scramble	21	44.43	7.82	B
	miR-I6-5p	14	55.86	10.06	B
	miR-133a	16	55.44	8.20	B
8	Mock	0	0	0.00	A
	Scramble	34	93.94	7.10	B
	miR-I6-5p	19	81.36	8.43	B C
	miR-133a	46	67.26	5.44	C
12	Mock	0	0	0.00	A
	Scramble	33	56.36	4.07	B
	miR-I6-5p	29	48.86	5.13	B C
	miR-133a	29	37.00	2.73	C
24	Mock	0	0	0.00	A
	Scramble	297	59.28	2.00	B
	miR-I6-5p	305	53.02	1.25	C
	miR-133a	277	40.65	1.17	D

Conditions within time points that do not have the same latter are statistically significant from one and other.
N refers to the number of individual plaques measured

5.7 – Does miR-133a affect transcription or translation?

Although a decrease in luciferase protein level was observed in the reporter assays, a concurrent decrease in ICP4 transcript levels was not measured in infected cells transfected with miR-133a. The former was carried out using the plasmid Psi-Check 2 with the target sequence cloned into the 3'UTR multiple cloning site. This is not the biological location of the target site and so moving the taget site to the 5'UTR mimics what is seen within ILTV. To investigate whether gga-miR-133a-3p affected the transcription or translation of ICP4. As there is currently no antibody available for ILTV ICP4 the full 5'UTR of ICP4 was cloned upstream of *Firefly* luciferase in the PGL3-Basic plasmid. HEK293T cells were then transfected with this plasmid on its own (mock control) and then with the addition of either miR-133a or a scrambled siRNA. After 48 hours, cells were harvested and RNA extracted. RT-qPCR was performed to determine relative levels of luciferase RNA transcript in the presence of miR-133a. – RT samples were also run as internal controls to ensure only cDNA was investigated and not any residual DNA following RNA harvest *(Data Not Shown)*. Relative RNA levels were normalised using the ΔΔCt method of analysis with GAPDH acting as an internal control (Livak and Schmittgen, 2001a). Once relative expression levels were calculated, they were normalised to the mock control to give a relative expression level as a percentage (%). A one-way ANOVA was performed on the ΔΔCt values, which also included a post-hoc Tukey's test to look at the correlation between differing conditions.

Data analysis revealed that there was a 55 % drop in relative RNA levels (%) compared to the mock control although this difference was not statistically significant (Figure – 5.7). Furthermore, the presence of a scrambled siRNA caused an 85 % drop in relative RNA levels compared to the mock control and a 30 % reduction compared to miR-133a although these differences were not statistically significant (Figure – 5.7).

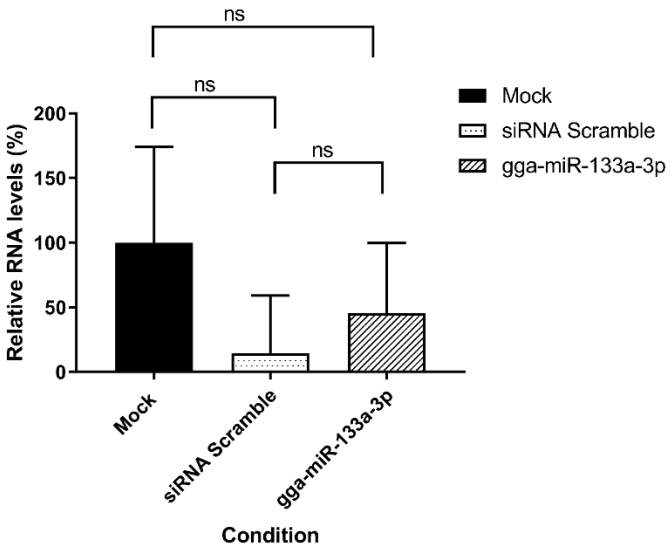


Figure – 5.7. Relative RNA levels (%) of a reporter plasmid containing a miR-133a-3p target sequence in the presence of miR-133a miRNA mimic or a scrambled siRNA control when compared to GAPDH

A reporter plasmid containing the miR-133a-3p target sequence in the 5'UTR was cloned using PGL3b vector. HEK293T cells were transfected with the plasmid on its own (mock) or with a mature miR-133a mimic or siRNA scramble control or were transfected on their own. After 48 hours, RNA was harvested and used for RT-qPCR to look at expression levels of luciferase. The 2ΔΔCT method of analysis was used with GAPDH as a control. Error bars display standard error of the mean (SEM). N = 7 (biological replicates)

5.8 – miR-133a and the Interferon response pathway

With an apparent reduction in virus titre and a statistically significant reduction in plaque size, it was not clear if this was due to a direct effect upon the virus or an indirect effect by carry over of any mimic induced interferon components that could initiate an antiviral response. To test this, LMH cells transfected with either mimics or a scramble siRNA. 12 hours post transfection, media was replaced with fresh complete media containing 2 % FBS and samples were left for a further 12 hours. Concurrently, DF-1 cells in 12 well plates were transfected with a luciferase reporter plasmid driven by the ISG mX promoter (A kind gift from Prof Steve Goodbourn, St George's, University of London) alongside an SV40 driven firefly luciferase as an internal control. 24 hours post transfection of the reporter plasmids, media was transferred from the LMH cells onto the DF-1 cells and left for a period of 24 hours. Samples were then harvested and luciferase assays carried out as previously described. Data shown in Figure – 5.8 is the combined results of two independent assays with n = 4. Data was at first normalised to the internal control to give a fold change level. This was then compared to the mock control as a percentage fold change. Errors indicate the standard error of the mean (SEM). Statistical analysis using a one-way ANOVA was carried out on the luciferase levels after normalisation to the internal control and any p values stated derive from this data.

Data analysis showed that recombinant chicken interferon α/β protein was able to elicit a 50-fold increase in luciferase expression when compared to the mock media control. This was statistically significant from the mock control and all other conditions ($p =<0.001$). A second control, Poly I:C was added directly into the media of transfected DF-1 cells and resulted in a 5-fold increase in luciferase activity compared to the mock control. However, this was only statistically significant from recombinant chicken α/β protein ($p =<0.001$) and no significant difference was observed between Poly I:C and any of the other conditions. Analysis of the control siRNA scramble showed no difference in luciferase activity from the mock condition which was also not statistically significant. In contrast, both gga-miR-133a-3p and ILTV-miR-I6-5p showed decreased luciferase expression of ~73 % and 65 % compared to the mock. Whilst luciferase expression (%) was lower, no significant difference was found upon statistical analysis. Similarly, transfer of LMH mock media onto transfected cells and the introduction of a recombinant mouse IFN γ protein showed a reduction in luciferase expression compared to the mock, this was also found not to be statistically significant.

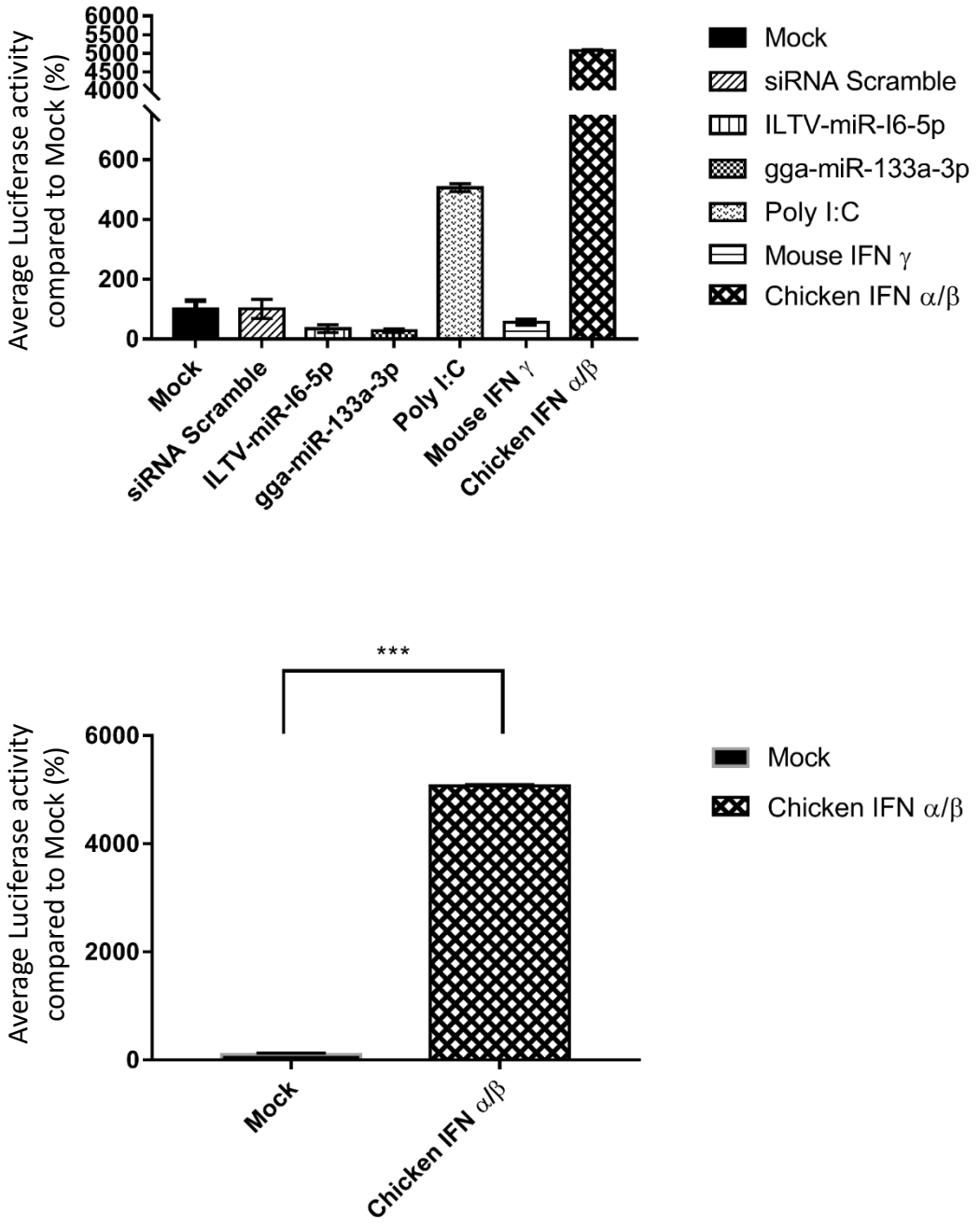


Figure – 5.8. The response of mX promoter, an IFN α/β immune stimulated gene (ISG) to transfection of miR-133a-3p

LMH cells were transfected with a scrambled siRNA, ILTV-I6-5p or gga-miR-133a-3p and left for 12 hours. Media was then replaced on the cells, which were incubated for a further 24 hours. DF-1 cells were transfected with a luciferase reporter plasmid driven by mX promoter. Transfection media was left on for 24 hours and then replaced with media harvested from transfected LMH cells. In addition to this, poly I:C, recombinant chicken IFN α/β protein or recombinant mouse IFN γ was added directly to DF-1 media. After 24 hours incubation, cell lysate was harvested for a luciferase reporter assay. *Renilla* luciferase was measured against *Firefly* luciferase for normalization. Data was then compared to the mock control to give an average fold change percentage (%). Error bars represent the standard error of the mean.

5.9 – miR-133a-3p Expression in the Chicken and its relevance to ILTV

5.9.1 – Confirming the sequence of miR-133a-3p in chickens

To attempt to detect gga-miR-133a-3p in the chicken and confirm the sequence was the same as in other species, muscle tissue (breast and thigh) from two six week old Hyline birds was taken and RNA was extracted using the miRNeasy Kit (Qiagen). RNA was then DNase treated and the MiScript II kit (Qiagen) was used for reverse transcription of RNA. Detection of mature miRNA sequences was then carried out using the MiScript SYBR green PCR kit (Qiagen). Due to the nature of the reverse transcription using polyadenylation and a oligo-dT primer with a 3' universal tag, DNA should not be detected and so minus reverse transcription (-RTs) controls were not necessary (as per section 3.4, Figure - 3.3).

Resulting qPCR products were run on a 3 % agarose gel for visualisation. According to the Qiagen MiScript PCR system handbook, mature miRNA sequences should produce a PCR product size of 85-87 bp.

Breast and thigh muscle were tested using this method as muscle tissue has been shown to highly express miR-133a-3p in other species. All three breast muscle samples and two out of three thigh muscle samples produced bands of between 50 and 100 bp, which fits with the expected product size. In addition, two positive controls were included with cDNA synthesised from the mature miRNA mimic. These also produced a similar product size of between 50 and 100 bp when visualised on a gel (Figure – 5.9A). Following gel analysis, bands visualised were extracted, cloned and sent for sequencing. Using this method, the sequence of miR-133a-3p in the chicken was confirmed to be the same as predicted in MiRBase.

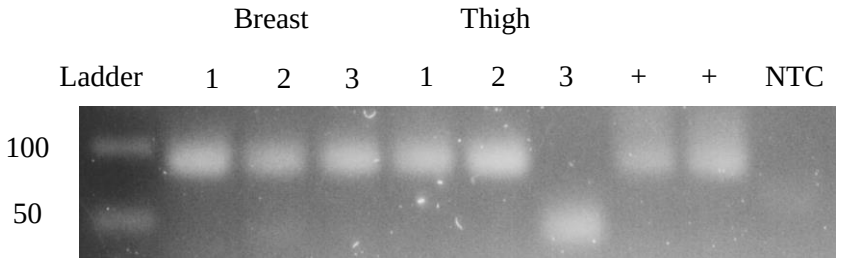


Figure – 5.9. Detection of gga-miR-133a-3p in Chickens
Samples of chicken tissue were taken from birds (1-3) and placed in RNA later prior to RNA extraction using the miRNeasy extraction kit (Qiagen). cDNA was synthesised using the MiScript II kit before detection of gga-miR-133a-3p was carried out using the MiScript SYBR green kit (both Qiagen). qPCR products were visualised on a 3 % agarose gel. Bands detected were then cut out, DNA extracted and the product TOPO cloned for sequencing. A – Detection of miR-133a-3p in muscle tissue.

5.9.2 – Detection of ILTV genomic DNA in Dorsal Root Ganglia

Chickens in the UK are vaccinated against ILTV using live-attenuated strains and so the virus is still able to set up a latent infection in the birds. To investigate the distribution of latent ILTV DNA, 2x 8-month laying birds were obtained from the breeding stock at the National Avian research Facility (NARF) when surplus to requirements. Spinal cord (SpC) and Dorsal Root Ganglia (DRG) were harvested. From these tissues, DNA was extracted and used in a conventional PCR for ILTV genomic DNA (gDNA) and the cellular gene GAPDH.

PCR products were analysed on a 2 % agarose gel and imaged. Bands were visualised at approximately ~350 bp for ILTV gDNA and ~550 bp for GAPDH. ILTV genomic DNA could detected in the DRG in one of the birds and not the other (Figure – 5.9A). DNA could not be detected in either SpC sample. GAPDH could be detected in both DRG samples but not in the spinal cord (Figure – 5.9B).

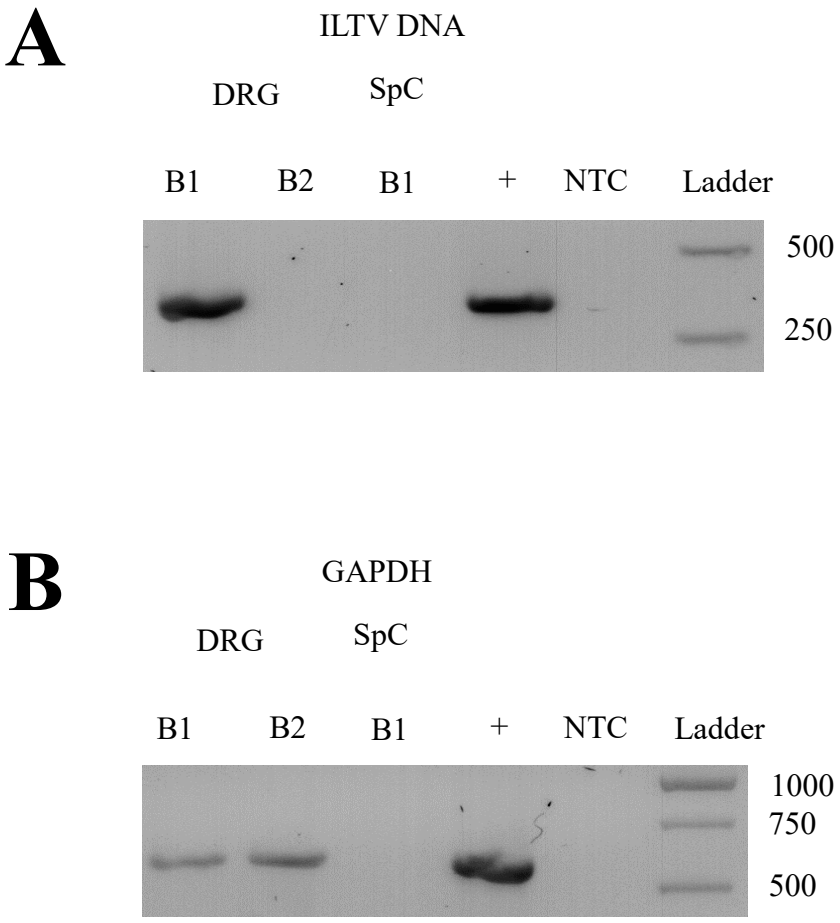


Figure – 5.10. Detection of ILTV DNA in the Dorsal Root Ganglia (DRG) of Vaccinated Chickens
Tissue samples were taken from 2x 8-month old laying hen's surplus to requirements at the National Avian Research Facility (NARF). DRG and Spinal Cord (SpC) were harvested and from this DNA was extracted as per section 2.2.1. DNA was used for conventional PCR with primers for ILTV genomic DNA (gDNA) (A) and GAPDH (B) along with appropriate controls.
B1: Bird 1; B2: Bird 2

5.9.3 – Detection of miR-133a in tissues relevant to ILTV biology

Once it was established that ILTV DNA could be detected in DRG, the expression of miR-133a was investigated. To look at tissues of relevance to the biology of ILTV, 3x 18-month laying birds were obtained from the breeding stock at NARF when surplus to the facilities requirements. Tissues harvested in the first instance were dorsal root ganglia (DRG) and Harderian gland (HG) alongside tissue samples taken from breast muscle (BM) of the birds. Samples were processed in the same manner as described previously and cDNA synthesised using the MiScript II RT Kit. Upon analysis of the qPCR products using gel electrophoresis, bands were visualised at around 100 bp which fits with the predicted mature miRNA product size (Figure – 5.9B). Bands were excised and TOPO cloned before been prepared and sent for sequencing. Sequencing confirmed the presence of the mature miRNA gga-miR-133a-3p in the DRG, HG and BM in addition to the positive control.

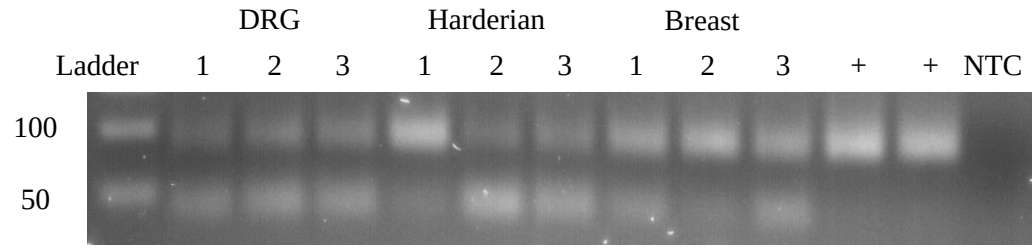


Figure – 5.11. Detection of gga-miR-133a-3p in DRG, Harderian gland and Breast Muscle

Samples of chicken tissue were taken from birds (1-3) and placed in RNA later prior to RNA extraction using the miRNeasy extraction kit (Qiagen). cDNA was synthesised using the MiScript II kit before detection of gga-miR-133a-3p was carried out using the MiScript SYBR green kit (both Qiagen). qPCR products were visualised on a 3 % agarose gel. Bands detected were then cut out, DNA extracted and the product TOPO cloned for sequencing.

5.10 – Discussion

Interactions between high confidence chicken miRNAs and viral transcripts were predicted using RNA Hybrid. The data set was compiled from 79 viral transcripts and 54 mature miRNA sequences taken from MiRBase Release 21 (June, 2014). In total, 103 miRNA target predictions were made using RNA Hybrid. From these predictions, three targets were taken forward, cloned into a reporter plasmid and tested biochemically. This led to the identification of one target showing a statistically significant decrease in luciferase expression when compared to the empty vector and siRNA scramble control. Following on from this, the target site for gga-miR-133a-3p was shown to be contained within the 5'UTR of the expressed mRNA of ICP4. Virus incubated in the presence of a mature miRNA mimic and subsequently used for conventional plaque assay analysis was shown to have defective growth morphology both in terms of resulting virus titre and plaque diameter. Whilst the viral load showed a large decrease between 133a and both I6-5p and siRNA scramble, this was not statistically significant when tested using a one-way ANOVA. Plaque size diameter was however determined to be statistically significant ($p =<0.001$) between 133a and both I6-5p and siRNA scramble. Investigations in this study could not attribute this reduction to the IFN response triggered by the transfection of either mimic or control and this was shown to not be significantly different from mock cells whereas the introduction of recombinant chicken interferon α/β protein caused a statistically significant increase in MX promoter driven luciferase expression ($p =<0.001$). The potential biological relevance of this interaction was then investigated by determining the tissue expression of the miRNA using a select panel of tissues. First, using muscle tissue, the sequence of the mature miRNA was confirmed to be what is predicted in MiRBase. Leading from this, tissues of potential biological relevance i.e. Dorsal root ganglia (DRG) and Harderian gland (HD) alongside a breast muscle tissue sample (BM) were selected and tested. Results from this confirmed the presence of gga-miR-133a-3p within the DRG and HD; both sites important with regards to virus biology.

The miR-133a family was first described in 2002 by cloning tissue specific 21 nt small RNAs from a panel of mouse tissues (Lagos-Quintana et al., 2002). MiR-133, as it was then called, was first identified in heart muscle and the cortex of the mouse brain. Naturally as technology has progressed, the identification of miRNAs by sequencing became the gold standard approach alongside other approaches such as *in situ* hybridisation. Several separate studies identifying chicken miRNAs followed giving some insight into what miRNAs were expressed in the developing bird (Darnell et al., 2006, Xu et al., 2006, Glazov et al., 2008). More recently, miR-133a has been identified in a wide range of species ranging from the

common fruit fly (*Drosophila melanogaster*) through to painted turtles (*Chrysemys picta*). A simple search of MiRBase for '*133a*' reveals a total of 75 miRNAs across the animal kingdom highlighting the conservation of the miRNA. Whilst the sequence of miR-133a was reported in the chicken, this was a predicted sequence and the results presented here are the first to confirm the sequence as the same as predicted in MiRBase.

First reports of miR-133a specific targets were published around a decade ago. Yin et al. (2008) reported the *in vivo* targeting of Mps1 kinase in Zebrafish (*Danio rerio*). Artificial expression of miR-133a caused the abrogation of fin regeneration and conversely, antagonism of miR-133a accelerated tissue regeneration. In the same year, Ivey et al. (2008) showed that miR-133a along with miR-1 were potent repressors of non-muscle genes during the regulation of cell linages from human and mouse embryonic stem cells. Depending on the algorithms used, miR-133a is found to target 441 – 497 gene transcripts in the human genome (https://www.exiqon.com and http://www.targetscan.org, both accessed 20th May, 2018). To date, approximately 43 protein-coding genes have some experimental data in the literature to show that they are targeted by miR-133a-3p (http://mirtarbase.mbc.nctu.edu.tw/php/index.php). Whilst these predictions were made based on human data and the confirmed targets are also based on humans, one can argue that the same targets are seen in the chicken if the gene homolog is encoded for in the bird.

Incidences of miR-133a targeting viruses either directly or indirectly are more recent. Aujeszky's disease is a highly important disease of the domestic pig (*Sus scrofa domesticus*). The aetiological agent is *Suid Herpesvirus 1* (SuHV-1, pseudorabies virus) (Fenner et al., 2011). A study investigating the role of viral and host miRNAs during *in vivo* infection identified miR-133a as a differentially expressed miRNA following infection with a virulent strain of SuHV-1. MiR-133a was up regulated 108 fold compared to the attenuated strain infected group. Further analysis using gene interaction networks predicted a total of 33 viral gene interactions between SuHV-1 and miR-133a which included the regulatory genes; *EP0, IE180, UL41* and *UL48* (Timoneda et al., 2014). IE180 of SuHV-1 is the homolog of ICP4, the major immediate-early gene identified in this study as a target of the miR-133a in the chicken. Timoneda and colleagues suggested that miR-133a may play an active role in combating viral infection and they go on to state the importance of miR-133a due to its potential interaction with many structural and non-structural genes within SuHV-1. In this study, the focus was primarily on the role of miR-133a and its regulation of ICP4 however in the initial bioinformatics screen, several other genes were identified as potential targets of

miR-133a in the ILTV genome and these included; *UL48, UL36, UL20* and *UL19* (a full list can be seen Appendix 7).

A total of 12 viral targets were identified in this study, which whilst lower than that of the SuHv-1 study, suggests an important role for miR-133a in the infection process. Two targets within *ICP4* were tested in addition to the predicted target of *UL20* and only one showed a statistically significant reduction ($p =<0.001$) in luciferase activity when compared to the siRNA scramble control and empty vector control. The effect of this interaction was then abrogated upon mutation of the miRNA seed sequence. First, it was shown that the miRNA target site is within the expressed mRNA of the *ICP4* gene transcript. Unusually, the target site was contained within the 5'UTR, whereas the vast majority of miRNA targets sites are found within the 3'UTR (Lewis et al., 2005). Whilst uncommon, it has been reported that targeting of the 5'UTR is as efficient as targets contained within the 3'UTR (Lytle et al., 2007). Interestingly, miR-133a sites are found within ICP4 homologs of several other *Alphaherpesvirinae* family members including; *Bovine Herpesvirus 1* (BoHV-1), *Equine Herpesvirus 1* (EqHV-1) and HSV-1 (Appendix 8). Equally, all of the species' listed encode for a miR-133a homolog and the sequence is highly conserved (Appendix 8). Whilst it was out of the remit of this project to investigate the function of miRNAs in other species' and other viruses, the identification of the same target site within the same gene for each raises the possibility that this interaction is conserved and may play a role in virus-host interactions.

Once the target was confirmed to be within the 5'UTR of the ICP4 mRNA transcript, the effect of miR-133a expression on virus infection was investigated. Firstly, a time course of infection was set up to investigate the relative expression levels of ICP4 in the presence of miR-133a. For this, two controls were used, one was a scrambled siRNA and the second, ILTV-miR-I6-5p, a miRNA shown previously to target ICP4 (Waidner et al., 2011). Relative RNA levels of ICP4 at both 8 and 24 HPI were significantly higher than that of the scrambled siRNA suggesting possibly that miR-133a was stabilising ICP4 RNA levels. This is in contrast to the translational data shown in Figure – 5.1C that show a knockdown of luciferase protein suggesting that translation is affected. Jopling et al. (2008) showed that the position of the miRNA target site has differing consequences. Positioning of miR-122 target sites within the 5'UTR of Hepatitis C virus causes an increase in viral RNA abundance whereas locating the miRNA site in the 3'UTR of a reporter mRNA resulted in a down regulation of mRNA expression suggesting binding site locations can affect gene regulation. Data presented in this thesis suggests that mRNA levels are indeed higher when the target

site of miR-133a in situated in the 5'UTR of ICP4 in both a reporter assay setting and also in the context of live virus (Figures – 5.7 and 5.3 respectively).

It was hypothesised that by affecting ICP4 expression, miR-133a may help to block reactivation of virus from latency. This is difficult to model in a tissue culture system and there is currently not a latency model available for ILTV and so this is something that could potentially be followed up on if the reagents and systems become available. As there is currently no latency model of infection for ILTV, including reactivation models, the role of this interaction in latency was difficult to establish, however investigations focused on its possible involvement during latency. As reviewed in Chapter 1, latency of ILTV is established within the DRG of the bird, similar to that of other *Alphaherpesvirinae* family members. Furthermore, we know through previous studies that miRNAs have the ability to influence and maintain latency. A study carried out by Pan et al. (2014), showed that miR-138, a neuronal specific miRNA, was able to target ICP0 of HSV-1. The study showed mutant viruses with disrupted miR-138 target sites exhibited increased expression of ICP0 and other lytic proteins in neuronal cells. Moreover, *in vivo*, mice infected with viruses containing the disrupted miR-138 sites displayed increased mortality and encephalitis symptoms alongside increased ICP0 and lytic gene transcripts within the DRG. To determine whether miR-133a could potentially play a role in latency, two approaches were taken. Firstly, what were the effects of the miRNA on the virus if the former was overexpressed in a permissive cell line and secondly, what is the biological relevance of this interaction. The former focused upon what the effects of miR-133a were on virus replication. The study found that virus harvested from cells transfected with miR-133a had a titre 66 % lower than that of virus grown in cells transfected with a siRNA scramble control. The siRNA control was used as a 'mock' in the sense that transfection can have an effect upon the cells and in addition, the transfection efficiency is not 100 % and so a virus only control would not have been an effective control. Similarly, miR-133a transfected cells produced 42 % less virus than cells transfected with the virally encoded ILTV-miR-I6-5p. In addition to the reduced viral titre, a statistically significant difference in plaque size diameter was observed at 8, 12 and 24 HPI between miR-133a and the siRNA control ($p =<0.05$). There was no significant difference between miR-133a plaque size and I6-5p at 8 and 12 HPI however at 24 HPI there was a statistically significant difference ($p =<0.001$). This suggested that the effect miR-133a was having was greater than that of I6-5p, which is encoded antisense to ICP4.

One possible reason for this reduction in virus titre and plaque size area is the interferon response. This pathway is triggered upon detection of invading viral DNA and initiates a signalling cascade to mount an antiviral response. Included in this response is the interferon pathway that has the ability to prime neighbouring cells through secreted proteins and switch on the same anti-viral pathways (Perry et al., 2005). As supernatant was used for plaque assay analysis of virus titre and subsequent plaque size area, it is feasible that if treatment with the dsRNA mimics induced a type I interferon response (IFN α/β) then IFN α/β present in the virus supernatants could induce an antiviral response in the plaque assay cultures. To examine this possibility, a reporter plasmid driven by MX promoter, an IFN α/β induced gene was used. In the presence of miR-133a, MX promoter levels were not induced/increased and there was no significant difference between miR-133a and the mock control. This suggested that IFN was not directly responsible for the reduction in titre or plaque size.

In this study ICP4 has been confirmed to be a target of miR-133a leading to a decrease in protein levels in a luciferase reporter system. Another possible theory is that ICP4 is incorporated into the ILTV virion and the effects of miR-133a result in a decrease in copy number of this potential virion associated ICP4. The composition of ILTV virions has not been reported in the literature but from looking at related viruses contained within the *Alphaherpesvirinae* family, we can infer what maybe contained within the particle. Studies examining the virion composition of HSV-1 determined that ICP4 is contained within the particle and is strongly associated with the capsid of the virion (Loret et al., 2008, Loret and Lippe, 2012). Likewise, ICP4 homologs have been found in other Alphaherpesvirus virions such as BoHV-1 (Barber et al., 2017). It is therefore feasible to postulate that ICP4 is incorporated into the virion of ILTV and a reduced copy number is the cause of the decreased viral tires and small plaque size. Due to the nature of ICP4 and its role in activating the lytic replication cascade, a reduced copy number contained within a particle may result in reduced fitness and thus a delay in initiation of the lytic replication cascade resulting in the smaller plaque size phenotype that is observed.

As well as targeting ICP4, miR-133a has a further 8 predicted targets within ORFs of ILTV (laid out in Appendix – 6). Included in the list are membrane proteins (UL43 and UL20), the large tegument protein (UL36) and also the major capsid protein encoded for by UL19 (Full list can be found in Appendix 6 of this thesis). Manipulation of UL36 alone could have an effect upon virus replication and egress. Studies elucidating the role of UL36 in HSV-1 found that replication was affected when UL36 was deleted from the virus (Desai, 2000).

Likewise, work investigating UL31 HSV-1 mutants observed defective replication kinetics (Chang et al., 1997). Taken together, it would be of interest to follow up on these potential targets which was not carried out during this study. If any are found to be affected by miR-133a, it may help to explain the phenotypes observed.

Another explanation to this is the targeting of cellular genes that are required for lytic gene expression. As previously stated, there are ~450 genes predicted to be targeted by miR-133a within the human genome and one can speculate that a similar number may be targeted in the chicken. Viruses are obligate intracellular parasites that rely heavily upon the host cell to provide essential tools for efficient replication and a reduction of a required protein/pathway will inevitably have an effect on virus replication. Interactions between cellular mRNAs and miR-133a would possibly have been identified using the CLASH procedure but this was not feasible due to the reasoning already outlined in chapter 3. As stated, until a suitable cell line that is permissive to ILTV infection and replication is either identified or made using laboratory methods of cell line immortalisation, this follow up work is technically challenging.

The last part of this study was to begin to investigate the biological relevance of the interaction identified. Many miRNAs families are conserved between species' allowing for the identification of miRNAs in livestock animals that have already been described in other animals (Shi et al., 2012). Chicken embryos have been used for several years as a model for vertebrate developmental biology due the speed at which they develop alongside the well-defined developmental process (Brown et al., 2003, Hamburger and Hamilton, 1992). This has allowed for 'gene-hunters' to identify a large number of cellular miRNAs within the bird. Currently, for the chicken, there are 1238 mature miRNA sequences from 904 precursors according to miRBase (release 22, March 2018). This compares with human, (2693 mature sequences from 1982 precursors), mouse (2013 mature miRNAs from 1303 precursors) and rat (769 mature miRNAs from 510 precursors). With the exception of rat, it suggests that there are many more miRNAs to be identified within chicken that may impact on all aspects of biology. With the existing data set, it was suggested that miR-133a was in fact a miRNA restricted to muscle tissue however it has been shown to target neuronal polypyrimidine-tract binding protein (nPTB) which is expressed predominantly in neurons and testes (Boutz et al., 2007).

To examine the expression of miR-133a, confirmation of the mature miRNA sequence was carried out by using muscle tissue that is known to express it at high levels (Ouyang et al., 2015). Using the commercially available MiScript II RT Kit (Qiagen) and sequencing, the

mature sequence was confirmed to be the same as predicted in MiRBase. Following on from this, two tissues of biological relevance were tested, the DRG and Harderian gland. The same method was applied and found that miR-133a was present in DRG, Harderian gland and Breast muscle tissue. The latter of these is not of biological relevance and was used primarily as a control for the experiment. As discussed earlier, DRG is the site of latency for ILTV and so the identification of miR-133a in this tissue has implications on latency and reactivation. It is possible that the miRNA helps to maintain viral gene expression below the reactivation threshold thus helping to maintain latency. This could be through targeting of just ICP4 or a combination of viral targets in addition to the possible targeting of cellular genes that help facilitate the reactivation of ILTV from latency.

Furthermore, the identification of miR-133a in Harderian gland is also novel. As reviewed in Chapter 1 of this thesis, the Harderian gland is a secondary lymphoid organ (SLO) of the avian immune system and is classed as the conjunctival associated lymphoid tissue (CALT) (Swayne et al., 2013). As this monitors the orbital region of the eye and the upper respiratory tract of the bird, it implicates miR-133a as potentially been involved with lytic replication also as virus can be found in these areas.

Whilst the study of the biological relevance identified miR-133a in tissues where virus is harboured, this study used birds that were 18 months old and so there is no temporal study to see if the miRNA is found in birds of different ages in addition to different breeds and gender. It has been suggested to be a possible marker for the onset of puberty in birds, whereby the levels of miR-133a in serum drop significantly at the onset of puberty (Han et al., 2016) suggesting it is found in younger birds at high levels also. Further work would ideally look at birds at different ages and more relevant to the poultry industry ages of birds.

Overall, this part of the project has characterised the interaction between the 5'UTR of ICP4 in ILTV and the endogenous miRNA gga-miR-133a-3p. It has shown that virus incubated in the presence of miR-133a has a reduced virus titre and a smaller plaque size area, which is not due to the involvement of the interferon system. The sequence of the mature miR-133a as predicted in MiRBase was also confirmed by sequencing. This was found to be in tissues of biological relevance to the virus, primarily the dorsal root ganglia and Harderian gland. The findings were looked at from a latency perspective in relation to the DRG but identifying this miRNA within Harderian gland opens up the possibility that it is expressed in other tissues relevant to the virus biology during lytic replicaiton. 13 targets of miR-133a were predicted in this study and a total of 3 were tested leaving a possible 10 other targets to be looked at. Some of the other predicted targets already defined roles in the virus

replication cycle as well as in addition to been antagonists of the host innate immunity. The likelihood is that as more miRNAs in both animals and viruses are found and characterised, more interactions between miRNAs and target mRNAs will be identified thus expanding the field that is very much in its infancy still.

Chapter 6 – Generation of a Recombinant ILTV lacking a cluster of five microRNAs

6.1 – Introduction

Genome editing is a powerful tool that will shape the future of biomedical research in the following years and decades. A relatively recent advancement, it allows for precise alterations to DNA using a number of methods outlined in Chapter One of this thesis. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) genome editing is the most recent of the approaches used and relies upon a cas9 nuclease protein guided by a small RNA, termed guide RNA (sgRNA) to direct the protein to a specific sequence through Watson-Crick base pair complementarity with target DNA (Ran et al., 2013). Several repair mechanisms can be employed as laid out in Figure 1.8 of Chapter One, however Homology directed repair (HDR) allows for precise changes to be made to DNA through the use of templates that have flanking sequences identical to the sequence you are wishing to change with the desired alterations incorporated into the template (Byrne et al., 2015).

Mutations within large DNA viruses are infrequent, with the error rate calculated to be somewhere between 10^{-8} and 10^{-11} errors per incorporated nucleotide, giving rise to a mutation possibly once in several hundred to some thousands of genome copies (Fleischmann, 1996). With this in mind, the application of homologous recombination has sought to speed up the process of mutations in herpesviruses with the classical approach using a repair template containing the desired changes, for example a deleted gene replaced with a fluorescent marker to aid in the selection process. Numerous studies have used this approach as a means of producing gene-deletant viruses for characterisation of gene function and role in the virus lifecycle (Fuchs et al., 2003, Veits et al., 2003a, Devlin et al., 2007, Pavlova et al., 2010). Whilst these approaches were, and are still, widely used, the number of recombination events resulting in the generation of a virus with the desired changes is low, estimated at less than 10 % of the overall virus population (Ryan and Shankly, 1996).

Inducing double stranded breaks (DSBs) into DNA however can greatly improve the efficiency of targeted homologous recombination to around 75 %. Studies carried out using SuHV-1 as a model genome showed that DSBs caused by a unique restriction enzyme site found within the genome could vastly improve targeted recombination (Ryan and Shankly, 1996). However, this approach relies upon a restriction enzyme site in the target area which limits its' potential whereas CRISPR-Cas9 overcomes this limitation.

Other methodologies have also been deployed such as the use of Bacterial Artificial Chromosomes (BACs) whereby the viral genome is cloned and the DNA construct becomes the artificial chromosome. This approach was first shown using murine cytomegalovirus (MCMV) and since then several other herpesviruses have also had BACs created for them

(Messerle et al., 1997). In contrast though, some herpesviruses including ILTV have not had BACs created due to several palindromic repeats which cause significant problems and result in unwanted mutations to the viral genome. This therefore warrants a different approach which negates these issues.

The first studies which combined genome editing using CRISPR-Cas9 and virology were carried out within the last decade. Ebina et al. (2013) showed a loss in LTR driven expression of HIV-1 genes following stimulation when sgRNAs were designed to target the LTR. Work on large double stranded DNA viruses that do not integrate into the host genome soon followed with work carried out in both adenovirus and HSV-1 (Bi et al., 2014). The findings of this paper suggested that the application of CRISPR-Cas9 increased efficiency of recombination to around 8 % in HSV-1.

Since these first studies were carried out efficient genome editing using CRISPR-Cas9 has been shown in a number of different herpesviruses spanning the subfamilies and includes EBV, HCMV and HSV-1 in addition to several other members of these subfamilies (Russell et al., 2015, Yuen et al., 2015, van Diemen et al., 2016). The future research in virology will almost certainly be shaped by the use of genome editing (Chen et al., 2018, Wang et al., 2018).

6.2 – Aims

The aims of this part of the project were to explore the feasibility of using genome-editing technology to delete a cluster of five miRNAs from the viral genome. This region was targeted as there are 5 miRNAs in a compact area (~1200 nt) and there is only a single copy of each miRNA. To carry this out, sgRNAs targeting the region of interest were designed and tested alongside the creation of homology repair templates containing a fluorescent reporter to aid in the selection of any edited virus particles. Cells were transfected and then infected with virus to induce double-stranded DNA breaks in the hope that upon DNA repair, the homology template containing the reporter gene was incorporated into the viral DNA.

6.3 – Selection of sgRNAs to target the ILTV genome

6.3.1 – Design of sgRNAs around the miRNA cluster

In the first instance, sgRNAs were designed using the online CRISPR Design program (http://crispr.mit.edu). These focused upon the right hand side of the last miRNA approximately ~1780 - 1800 bp from the left hand end of the viral genome following conversations with Dr. S. Lillico (The Roslin Institute). In total, three sgRNAs were selected and taken forward. A schematic diagram of the miRNA cluster and wider viral genome is shown in Figure - 6.1. The program was set to have minimum off target effects in the human (default) due to the limitations of the free software not including the chicken genome. Other online software programs were sought but the same limitation was seen. Chosen sgRNAs are shown in Table 6.1 and highlighted in yellow.

In addition, sgRNAs were also designed in collaboration with Dr W.S. Tan (The Roslin Institute, University of Edinburgh). Here, a custom Python script (written by Dr Tan, unpublished) was used to scan the ILTV genome in the designated areas (512 – 514 nts for the left hand side of the cluster and 1741 – 1818 nts on the right hand side of the miRNA cluster) with the following criteria. 1) Potential sgRNAs must be followed by a PAM (NGG) motif, 2) They do not contain tetramers or above (e.g. AAA, TTT), 3) They do not contain any *Bbs*1 sites and 4) They have a GC content between 20 and 80 % inclusive. Two areas of the ILTV genome flanking the miRNA cluster were chosen for sgRNA design. Once a list of sgRNAs was compiled, the efficiency of the sgRNAs in directing cleavage of the DNA was tested using Azimuth 2.0 (Doench et al., 2016). Once the efficiency was estimated, sgRNAs were ranked high to low. A second package called Cas-OFFinder (http://www.rgenome.net/cas-offinder/) was then used to scan the sgRNAs for any potential off targeting within the ILTV genome (but not the chicken genome) with a maximum of 5 mismatches allowed within the sgRNA sequence to the genome (Bae et al., 2014). Another package called CFD was then used to estimate the efficiency of these off target sites with the lowest off target sgRNAs ranked the highest in this list. Finally, from the two rankings, a list of sgRNAs for the left hand side of the miRNA cluster (designated the miRNA 5', Table – 6.2) and also the right hand side of the cluster (designated the miRNA 3', Table – 6.1). From this list, the top two sgRNAs for the left and right hand side were taken forward into the study.

Two methods of sgRNA were used to maximise the likelihood of designing sgRNAs that were efficient at directing cleavage of the DNA. sgRNAs designed using the online program used the human genome as a threshold for off-targeting limiting the number of sgRNAs that

could direct cleavage in the viral genome. In contrast, the sgRNAs designed in collaboration with Dr W.S. Tan used the viral genome as a threshold for off-targeting disregarding the chicken genome.

Chosen sgRNAs were then ordered as sense/antisense primer sets with the inclusion of a *Bbs*1 endonuclease restriction site to allow for cloning. Primers were annealed together and subsequently cloned into the Cas9 plasmid Px458 using *Bbs*1 endonuclease sites (Primers listed in Appendix 2 and Vector map in Appendix 1). Clones were selected and sent for sequencing to confirm the correct sgRNA insertion. Successfully cloned sgRNAs were then amplified and large DNA stocks were made and stored at – 20 °C until needed.

6.3.2 – Design of sgRNAs around Glycoprotein G (US4)

Glycoprotein G has been successfully deleted from ILTV previously using homologous recombination (Devlin et al., 2006). Therefore to see if the approach designed to delete the miRNAs was feasible, knock out GΔG viruses were constructed in parallel as a positive control. The design of sgRNAs directing cleavage of the gG region was carried out by Dr. I Dry (The Roslin Institute). The chosen sgRNAs were cloned the same as miRNA sgRNAs in 6.3.1.

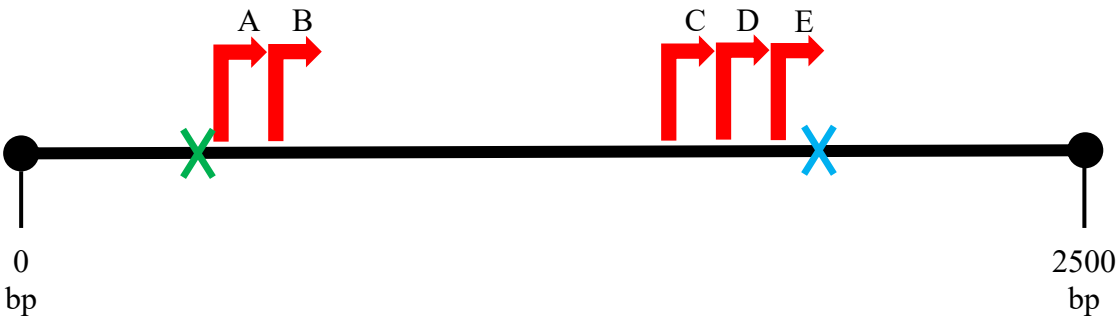


Figure – 6.1. Schematic diagram of the miRNA cluster at the left hand end of the ILTV genome

Diagrammatic representation of the first 2500 bp of the ILTV genome. Displayed are the 5 microRNAs found within the area. Red arrows and letters refer to miRNAs and direction. Green cross is where sgRNAs were picked at the left hand side of the miRNA cluster. Blue cross is where sgRNAs were picked at the right hand side of the cluster. Diagram not to scale.

A – ILTV-miR-I1-5p starts at position 534
B – ILTV-miR-I1-3p starts at position 568
C – ILTV-miR-I2-5p starts at position 1425
D – ILTV-miR-I3-3p starts at position 1634
E – ILTV-miR-I4-5p starts at position 1721

Table – 6.1. List of sgRNAs targeting the right hand side of the miRNA cluster in the ILTV genome

Organism	Efficiency	Area	Start	End	Strand	CRISPR sgRNA	PAM	Cut Position	CRISPR Start	CRISPR End	Chosen?
ILTV	0.774373	miRNA3'	1737	1766	-	TCCGTGATGAGGGAACCACA	CGG	1746	1743	1763	Y
ILTV	0.692173	miRNA3'	1721	1750	+	TATAGCGAGCAATGACCGTG	TGG	1741	1725	1744	Y
ILTV	0.6275	miRNA3'	1735	1764	+	ACCGTGTGGTTCCCTCATCA	CGG	1755	1739	1758	
ILTV	0.614389	miRNA3'	1763	1792	+	TCGTTATGCATAGATGCCTG	CGG	1783	1767	1786	
ILTV	0.612486	miRNA3'	1747	1776	-	TAACGAGCACTCCGTGATGA	GGG	1756	1753	1773	
ILTV	0.584543	miRNA3'	1748	1777	-	ATAACGAGCACTCCGTGATG	AGG	1757	1754	1774	
ILTV	0.542116	miRNA3'	1764	1793	+	CGTTATGCATAGATGCCTGC	GGG	1784	1768	1787	
ILTV	0.540315	miRNA3'	1780	1809	-	GCAAGTGCGACCTAGCCCGC	AGG	1789	1786	1806	Y
ILTV	0.522977	miRNA3'	1789	1818	+	AGGTCGCACTTGCTGGGCAG	CGG	1809	1793	1812	
ILTV	0.4588	miRNA3'	1769	1798	+	TGCATAGATGCCTGCGGGCT	AGG	1789	1773	1792	
ILTV	0.453547	miRNA3'	1783	1812	+	CGGGCTAGGTCGCACTTGCT	GGG	1803	1787	1806	Y
ILTV	0.226674	miRNA3'	1798	1827	+	TTGCTGGGCAGCGGCTAAAC	TGG	1818	1802	1821	
ILTV	0.195193	miRNA3'	1782	1811	+	GCGGGCTAGGTCGCACTTGC	TGG	1802	1786	1805	Y

Efficiency on a scale of 0 – 1 with 1 been the most efficient at cleaving the target.
Start and end refer to the position of the target site in the ILTV genome
Strand refers to the positive or negative strand of the ILTV genomic DNA
Cut position is the specific point where the DSB is induced and the CRISPR start and end points are the sgRNA position
Chosen refers to whether the sgRNA was chosen for the study
Yellow highlighted rows refer to sgRNAs designed using the first method outlined in section 6.3.

Table – 6.2. List of sgRNAs targeting the left hand side of the miRNA cluster in the ILTV genome

Organism	Efficiency	Area	Start	End	Strand	CRISPR	PAM	Cut Position	CRISPR Start	CRISPR End	Chosen?
ILTV	0.697143411	miRNA5'	521	550	+	TCCGCAGAGGAGACTGATTG	GGG	541	525	544	Y
ILTV	0.673674478	miRNA5'	508	537	+	GATTTCGCGAGGCTCCGCAG	AGG	528	512	531	Y
ILTV	0.559910063	miRNA5'	503	532	-	CGGAGCCTCGCGAAATCCAA	CGG	512	509	529	
ILTV	0.433027156	miRNA5'	519	548	+	GCTCCGCAGAGGAGACTGAT	TGG	539	523	542	
ILTV	0.377577048	miRNA5'	520	549	+	CTCCGCAGAGGAGACTGATT	GGG	540	524	543	

Efficiency on a scale of 0 – 1 with 1 been the most efficient at cleaving the target.
Start and end refer to the position of the target site in the ILTV genome
Strand refers to the positive or negative strand of the ILTV genomic DNA
Cut position is the specific point where the DSB is induced and the CRISPR start and end points are the sgRNA position
Chosen refers to whether the sgRNA was chosen for the study

6.4 – Testing of sgRNA cutting ability *In Vitro*

To determine if the sgRNAs had the ability to direct cleavage of the DNA sequence, they were tested *in vitro*. DNA for both the template cutting sequence and the sgRNAs was amplified using conventional PCR (section 2.2.7, primers found in Appendix 2). DNA to be used as a template cutting sequence was then purified and stored until the reactions were assembled. DNA of the sgRNA sequences was purified and used for RNA synthesis as described in section 2.6.4. sgRNA was purified and DNase treated to remove the DNA template. Reactions were then assembled *in vitro* using a commercial recombinant Cas9 protein as per section 2.6.5. Samples were harvested at 1 and 24 hours post assembly and treated with both proteinase K and RNase A to degrade the Cas9 protein and sgRNA. Samples were separated on an agarose gel by electrophoresis and imaged. Figure 6.2. shows the workflow of this experiment.

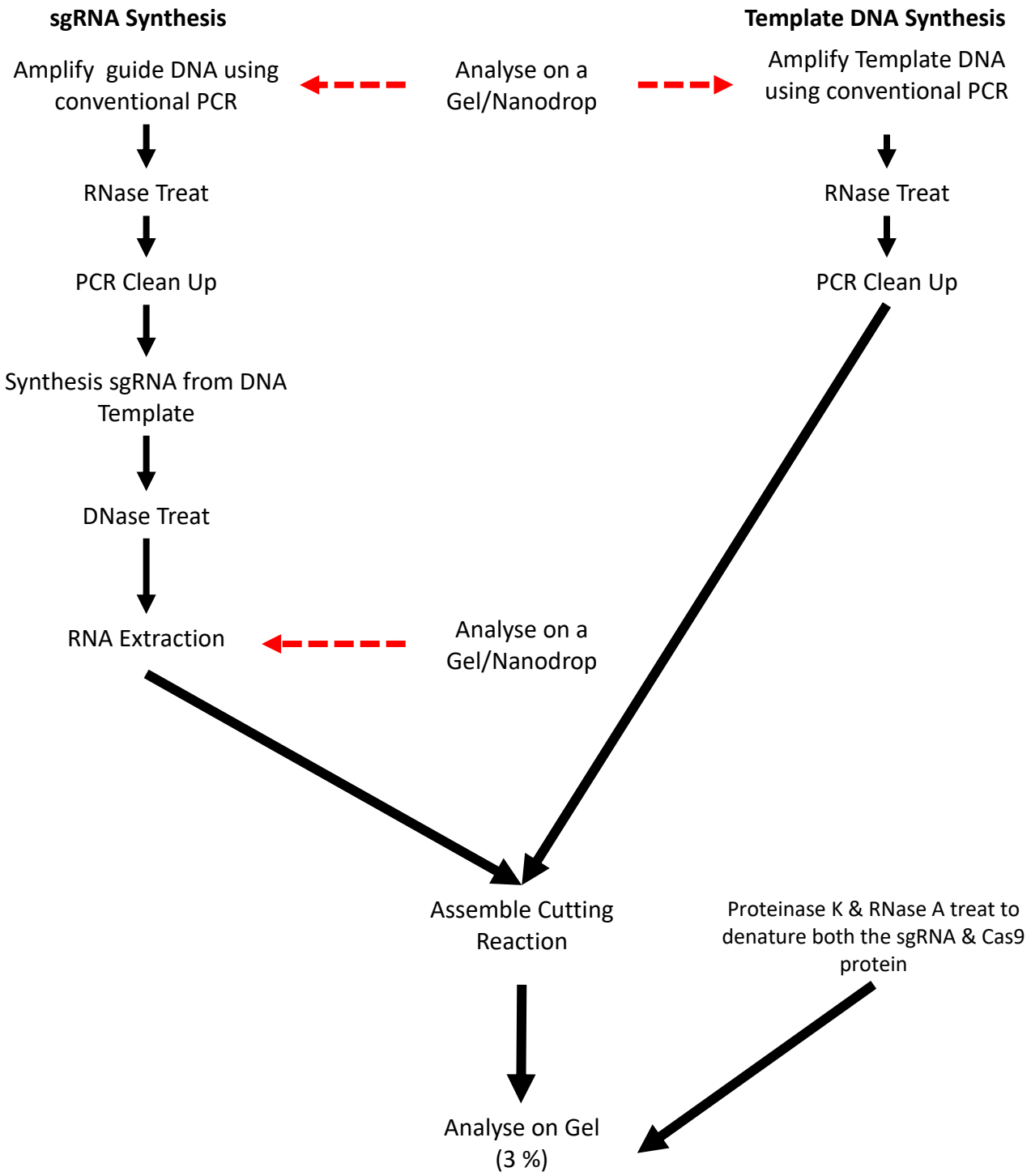


Figure 6.2 – Overview of the testing of sgRNAs cutting ability *in vitro*

In vitro transcription was used to synthesise sgRNAs from a DNA template amplified using conventional PCR. sgRNAs were RNA extracted and DNase treated to remove any DNA contamination and then frozen at – 80 °C until needed. Template DNA that contained the complimentary sgRNA site was amplified also by conventional PCR. PCR products were cleaned up and RNase treated to move any contamination. Template DNA was stored at – 20 °C until needed. Cutting reactions were assembled with a 10:10:1 molar ratio of Cas9:sgRNA:template DNA and left for either 1 hour or 24 hours before analysis. Prior to agarose gel electrophoresis, cas9 protein was degraded with proteinase K and the sgRNAs were degraded with RNase A. Red arrows represent optimisation steps whereas black arrows depict workflow.

6.4.1 – miRNA K/O sgRNAs testing *in vitro* using an online program

Testing of the sgRNAs showed that the original sgRNAs designed were unable to direct cleavage of the template DNA after incubation of the reactions for either 1 hour or 24 hours *in vitro* (Figure 6.3A & B). Due to this finding, the sgRNAs designed using this method were not taken forward.

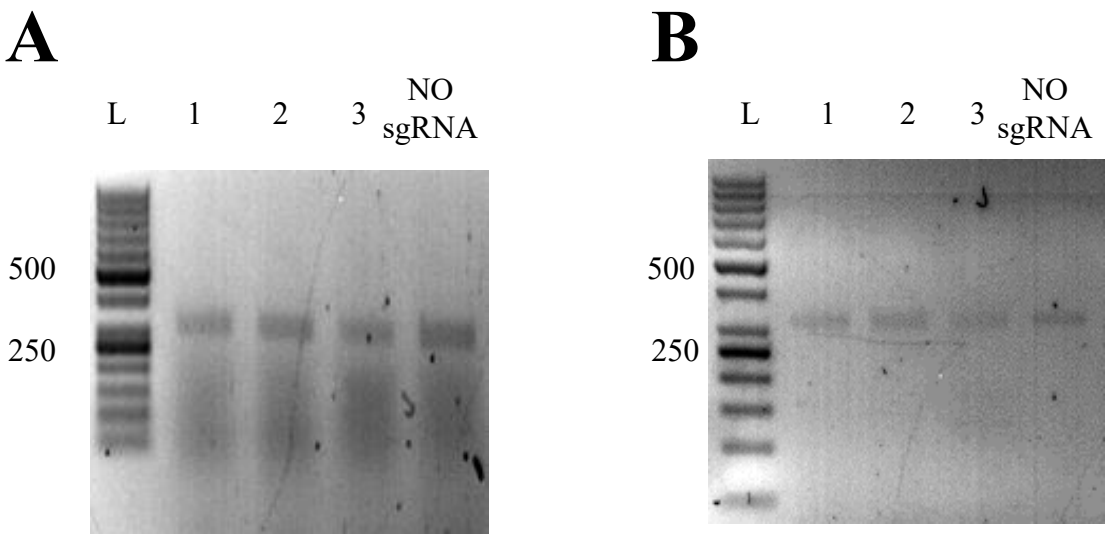


Figure 6.3. – Agarose gels of *in vitro* testing of sgRNAs designed to target the miRNA cluster using an online program
sgRNAs designed using the online CRISPR program (found at: http://crispr.mit.edu) were tested using the methodology laid out in Figure 6.1. L = Ladder (50 bp generuler ladder), 1 = sgRNA 1, 2 = sgRNA 2, 3 = sgRNA 3, no sgRNA = no sgRNA added to sample
A – Samples were left to incubate for 1 hour
B – Samples were left to incubate for 24 hours

6.4.2 – miRNA K/O sgRNAs testing *in vitro* set two

In total, four sgRNAs designed in collaboration with Dr W.S. Tan (The Roslin Institute) were tested which consisted of two sgRNAs from the left hand side and two sgRNAs from the right hand side of the miRNA cluster. Testing of the sgRNAs showed all four were able to direct the Cas9 protein to cleave the DNA template at both time points chosen *in vitro* (Figure 6.4A & B). As these sgRNAs were able to facilitate cleavage of the DNA template, they were taken forward and used in experiments to create a miRNA deletant ILTV virus.

6.4.3 – Testing the cutting ability of sgRNAs against a portion of GΔG DNA sgRNAs *in vitro*

Two sgRNAs were tested *in vitro* for their ability to direct cleavage against a DNA template amplified from the ORF encoding for glycoprotein G (US4) GΔG. One sgRNA was not recovered during the initial stages of the amplification of gDNA synthesis via PCR *(Data Not Shown)*. Of the two sgRNAs that were taken through the process, one showed cutting ability against the template DNA (Figure 6.5C & D). This sgRNA was used in downstream experiments.

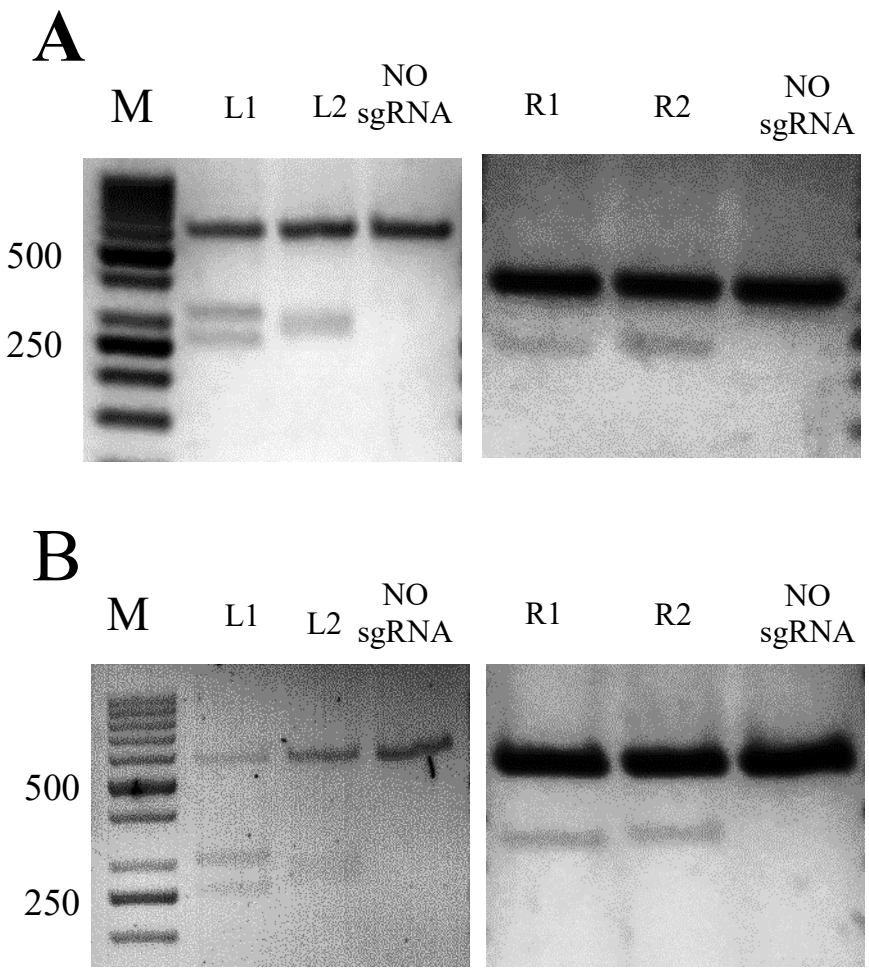


Figure 6.4. – Agarose gels of *in vitro* testing of sgRNAs to direct cleavage at both sides of the miRNA cluster

A second set of sgRNAs were designed in collaboration with Dr W.S. Tan (The Roslin Institute). Guides were designed at both the left and right hand side of the miRNA cluster (designed sgRNAs listed in Tables – 6.1 & 6.2). They used the viral genome as an off-target parameter and ignored the chicken genome. Guides were tested were tested using the methodology laid out in Figure 6.1.

M = DNA marker

Letters above lanes (L&R) refer to left or right hand side of the miRNA cluster whilst numbers refer to the sgRNA

A – Samples were left to incubate for 1 hour

B – Samples were left to incubate for 24 hours

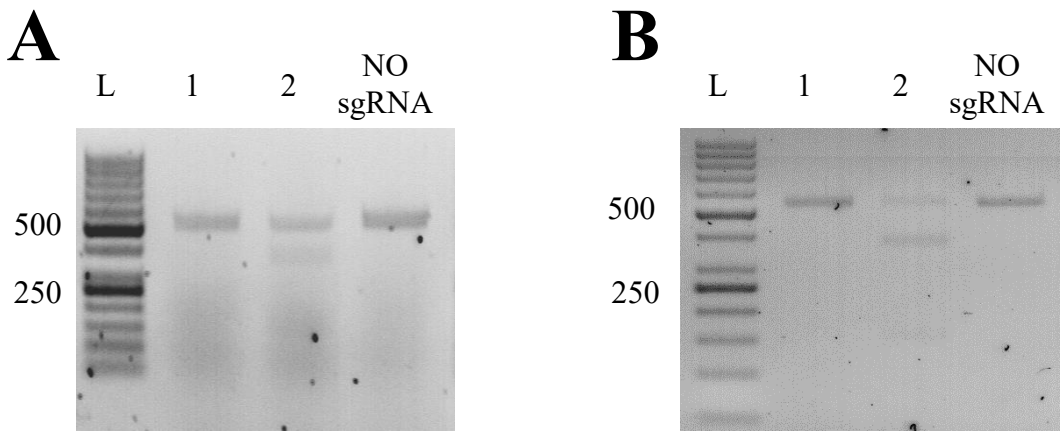


Figure 6.5. – Agarose gels of *in vitro* testing of sgRNAs to direct cleavage at Glycoprotein G (US4)
Glycoprotein G has already been shown to be dispensable for *in vitro* ILTV replication (Devlin et al., 2006). Guides were designed by Dr. I Dry (The Roslin Institute). Guides were tested using the methodology laid out in Figure 6.1.
Numbers above lanes refer to sgRNA
A – Samples were left to incubate for 1 hour
B – Samples were left to incubate for 24 hours

6.5 – Construction of Homology repair templates

To increase the chances of recovering a recombinant virus with the miRNAs deleted, the approach taken used homology directed repair (HDR) coupled with a fluorescent reporter construct to allow for detection via microscopy. This approach required a homology repair cassette that was constructed from the flanking sequences of the miRNA cluster with the reporter construct inserted into the miRNA region instead.

6.5.1 – Creation of a Homology repair template

Homology repair flanks were synthesised using a two-step process. Primers were designed to amplify the left flank and right flank independently. To allow for insertion of a fluorescent reporter construct, a common sequence was included at the end of the left reverse primer and the forward right primer. This common sequence consisted of the restriction endonucleases Kpn1 and BamH1 in addition to a 4 bp consensus sequence. Once both flanks were amplified by conventional PCR, there were analysed on an agarose gel before being extracted *(Data Not Shown)*. Resulting DNA was then used in an overlapping PCR to join the left and right flanks together. This was achieved using the left flank forward primer and the right flank reverse primer. PCR products were again visualised on an agarose gel and bands of the right size were extracted, TOPO cloned and sent for sequencing *(Data Not Shown)*. PCR products were confirmed to be the sequence of the expected flanks. This same process was carried out for creation of the GΔG homology repair template *(Data Not Shown)*. Primers are listed in Appendix 2 for the above. Workflow for the creation of a homology repair template is shown in Figure – 6.6.

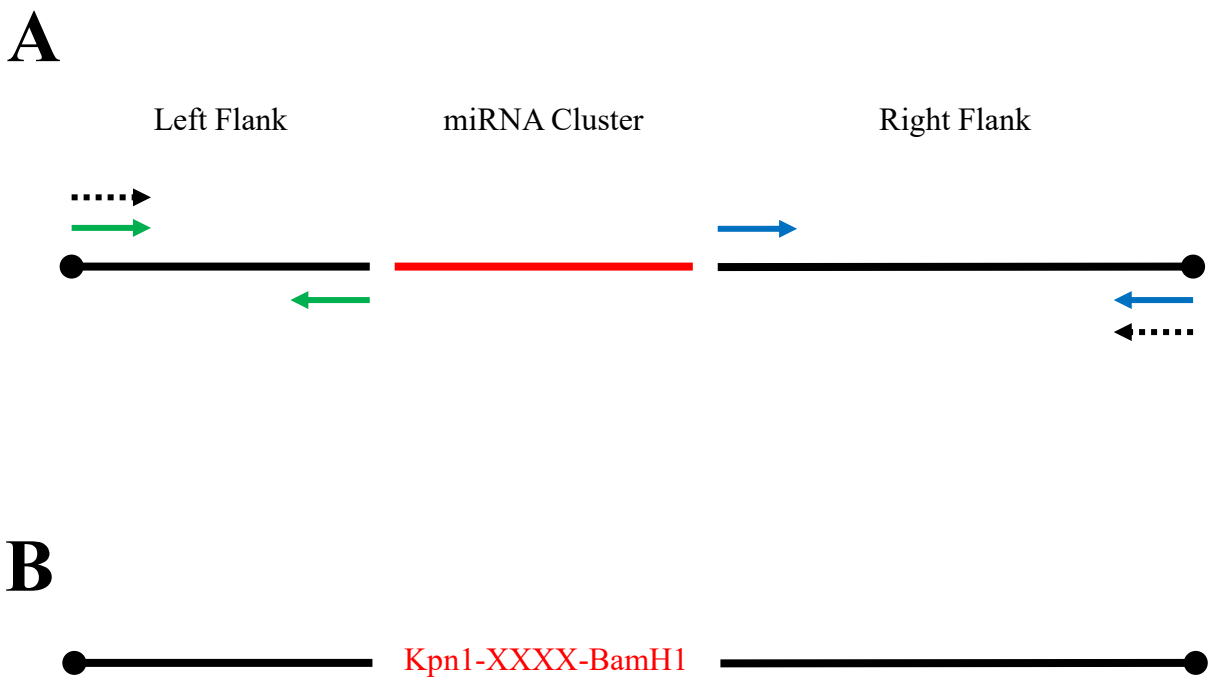


Figure – 6.6. Schematic diagram depicting the construction of the Homology repair template used in the CRISPR experiment
Two-step PCR was used to amplify flanking sections of ILTV genomic DNA around the miRNA cluster. First round PCR amplified the left and right hand flanks separately (A- green and blue arrows). The left reverse primer and right forward primer (marked with asterisks) contained a common sequence shown in (B) to allow for insertion of a reporter construct. Second round PCR used the left forward primer and right reverse primer (A - dotted black arrows) to create a full length homology repair template with a common cloning sequence (B).

6.5.2 – Creating of a fluorescent reporter construct

To create a reporter construct, overlapping primers were designed to amplify the CBh promoter and eGFP. Primers included the Kpn1 and BamH1 restriction endonuclease sites to allow for insertion into the homology repair template created in section 6.5.1. The first round of PCR amplified the CBh promoter and eGFP separately. These products were analysed on an agarose gel and bands were excised and purified for a second round of PCR to create a full fluorescent construct *(Data Not Shown)*. The full construct was then cloned into TOPO4 to allow for sequencing which confirmed the creation of a CBh driven eGFP *(Data Not Shown)*. Primers for this are listed in Appendix 2 and workflow for this is shown in Figure – 6.7.

6.5.3 – Insertion of a reporter construct into the homology repair template

Using the restriction endonucleases as outlined above, Kpn1/BamH1, the CBh-eGFP PCR product was inserted into the homology flanks (Figure – 6.8). To check for its successful insertion, a diagnostic digest was set up with appropriate controls. This confirmed the presence of CBh-eGFP within the homology flanks *(Data Not Shown)*.

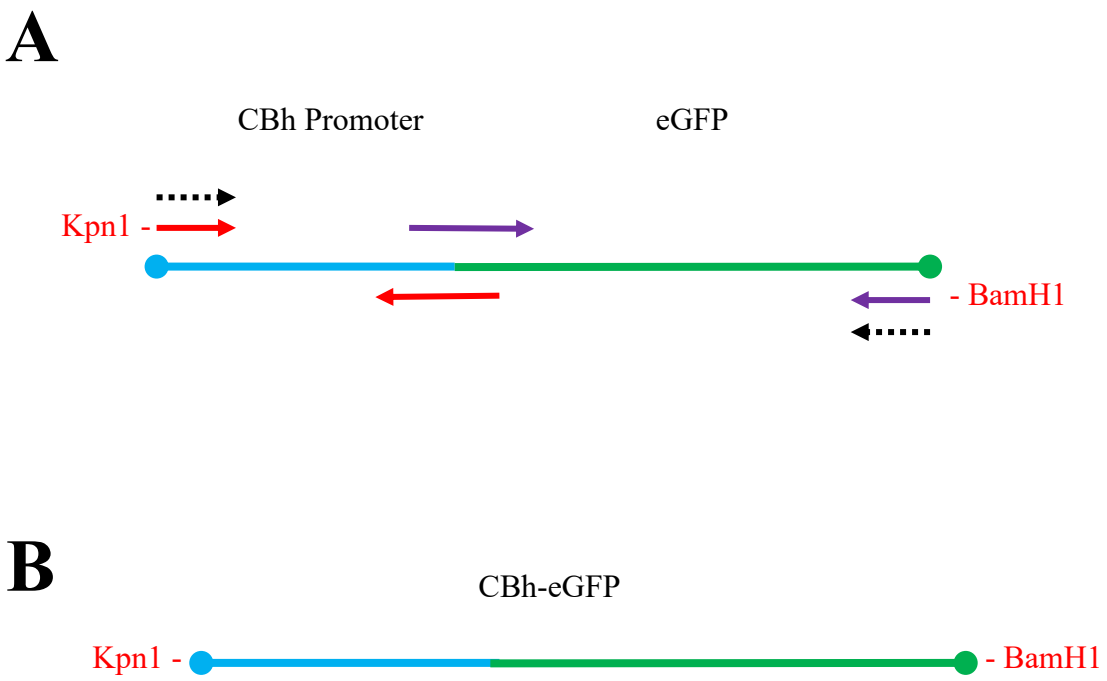


Figure – 6.7. Schematic diagram depicting the construction of the fluorescent reporter construct used in the CRISPR experiment
Two-step PCR was used to amplify a CBh-driven eGFP reporter construct. Primers were designed to amplify CBh promoter and eGFP separately and included a overlapping region on the CBh reverse and eGFP forward primer (A – red and purple primers). PCR products were purified and used as templates for a second PCR using the CBh forward and eGFP reverse primer (A- black dotted arrows) to create a CBh-eGFP with unique restriction enzymes at either end for insertion into a homology repair template (B).

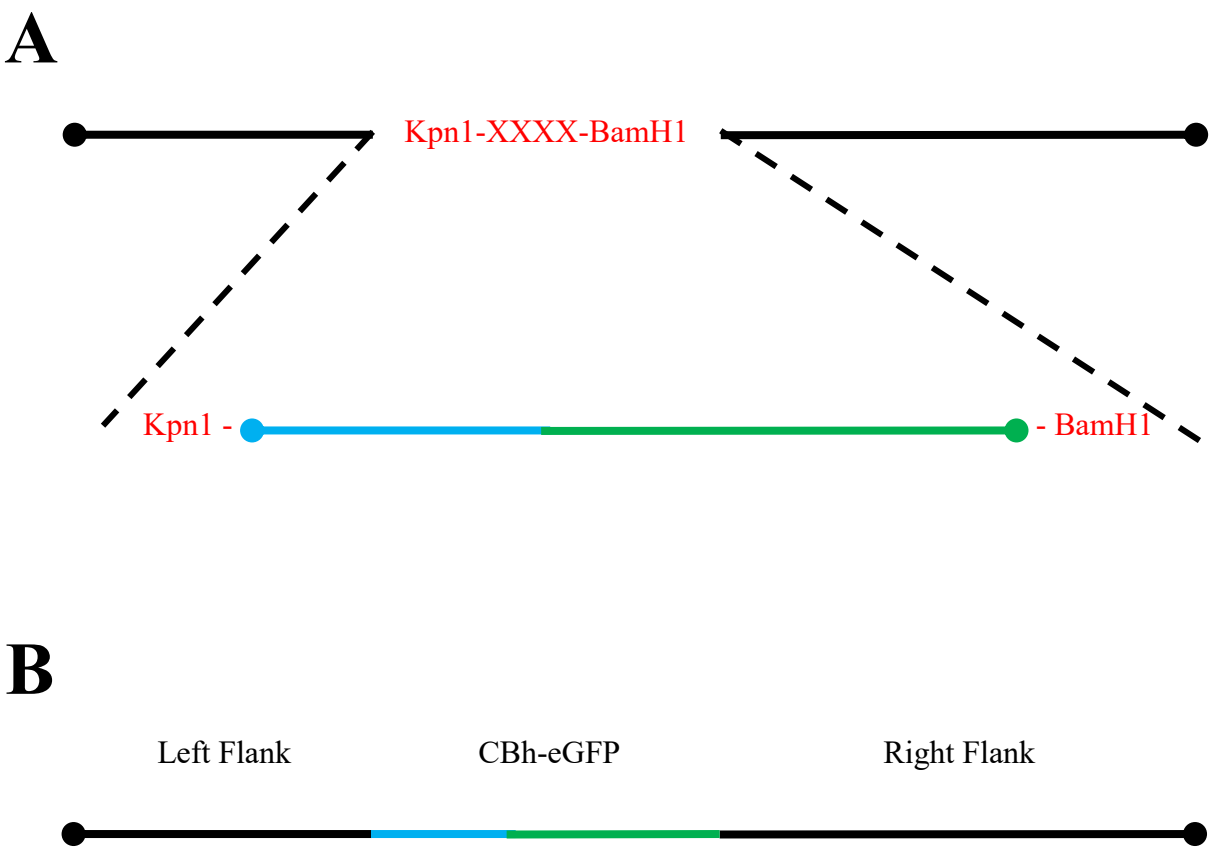


Figure – 6.8. Schematic diagram showing the insertion of the reporter construct into the homology repair template
The restriction endonucleases Kpn1 and BamH1 were used for insertion of the reporter construct into the homology repair template. Both homology repair template and reporter construct were digested with Kpn1 and BamH1 (A). The reporter construct was then ligated into the homology repair template as described in section 2.3.3. This produced a homology repair cassette containing a CBh-eGFP fluorescent reporter (B).

6.6 – Detection of Glycoprotein G recombinant viruses

CRISPR experiments were carried out as described in section 2.6.6 of this thesis. After harvesting of the samples, they were freeze-thawed a total of three times to disrupt any cells that were collected during the harvest process. These samples were then used to infect new cells and were serially diluted. Areas of green were picked using a 10 µl pipette tip and transferred onto fresh, naïve cells and picked virus was allowed to replicate for 48 hours before harvesting. Samples were then harvested and frozen down with a small aliquot removed for DNA extraction and testing.

Primers were designed to look for recombination events and should only detect mutant virus. A forward primer was designed against the eGFP sequence whilst a reverse primer was designed against the virus and sat outside of the flanking sequence. Using extracted DNA from for conventional PCR and using extracted wild type virus and gDNA primers as controls, bands were detected in two of the four samples tested which also tested positive for gDNA. No bands were detected in WT virus except for in the gDNA control PCR (Figure 6.9A & B). Experiments were then halted here to concentrate upon the miRNA deletant virus as a gΔg virus has previously been characterised and studied (Devlin et al., 2006).

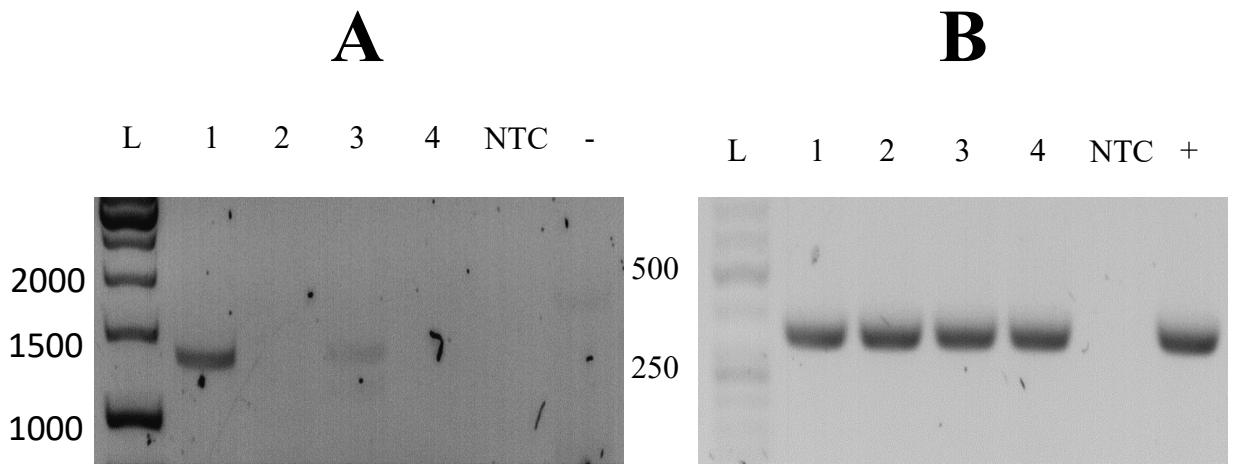


Figure 6.9. – Detection of recombination events using CRISPR-Cas9 against Glycoprotein G of ILTV

Primers were designed to only amplify recombination events with a forward primer situated in the eGFP portion of the reporter construct and a reverse primer located downstream of the right hand flank homology cassette. In addition, primers termed gDNA were used to detect against another part of the viral genome. PCR products were analysed using a suitable % agarose gel and imaged. Numbers refer to PCR reaction. NTC = No template control. '-' refers to a negative control using wild type virus (A) whilst '+' refers to a gDNA positive control (B).

A – Primers designed to detect recombination events

B – gDNA primers to check for virus presence

6.7 – Detection of miRNA K/O recombinant viruses

CRISPR experiments were carried out as described in section 2.6.6. For the detection of recombination events, primers were designed to amplify across the region of interest. A forward primer was situated towards the terminal end of the left hand flank whilst a reverse primer was situated downstream of the right flank and into the viral genome. Virus harbouring recombination events will have a larger PCR product produced compared to the wild type sequence. Bands detected with a larger PCR product were excised, DNA extracted and cloned for sequencing. Using these primer sets, several bands were visible at the correct product size in line with the positive control (Figure 6.10A). Upon sequencing, two of the clones sent for sequencing were confirmed to be from recombinant virus (Figure 6.10A, lanes 2 & 9). Sequencing from lane 8 (Figure – 6.10A) could not be achieved. As whole virus samples were used for these PCR reactions, recovery of these viruses after detection could not be carried out.

However, further testing of CRISPR virus stocks using a portion of the sample (the rest of sample was added to cells for virus infection) for recombination events detected partial recombination events in the virus (Figure – 6.10B, Lane 6). This could have been a partial insertion of the GFP construct or a deletion event in the virus but this was unclear and so was not taken forward. This PCR reaction had the addition of 6 % DMSO to allow for detection of wild type virus also. Due to this, many non-specific bands were also detected upon visualisation on an agarose gel. Bands of the correct size (~3000 bp) were excised, cloned and sent for sequencing. Results from this confirmed the presence of recombination *(Data Not Shown)*.

The insertion of eGFP into the viral genome allowed for fluorescent microscopy imaging and analysis. Samples were assessed at all stages of the process from initial transfection/infection of cells through to plaque purification of the recombinant viruses. Images taken during the initial transfection/infection are shown in Figure – 6.11. The plasmid encoding for the Cas9 protein and the sgRNA also encodes for a red fluorescent protein (RFP) (Figure – 6.11C&D) whilst the homology repair template contains the eGFP (Figure – 6.11E&F).

Whole well lysates were harvested and freeze thawed three times to release virus. These lysates were used to infect naïve LMH cells. Samples were placed under a 0.5 % Avicell overlay to restrict virus spread. Areas of green were selected and added to a second set of naïve LMH cells and numbered. Images of cells under avicell overlay were taken 24 HPI. No images of RFP were taken due to lack of detection. GFP could be seen under the Avicell

overlay in both GFP only and GFP + brightfield phases (Figure – 6.12B&C). Taken together, these results pointed to the creation of a recombinant virus lacking the five miRNAs at the left hand end of the genome (laid out in Figure – 6.1). Time constraints meant the purification of a pure population of recombinant virus could not be achieved through a further round of plaque purification and PCR analysis.

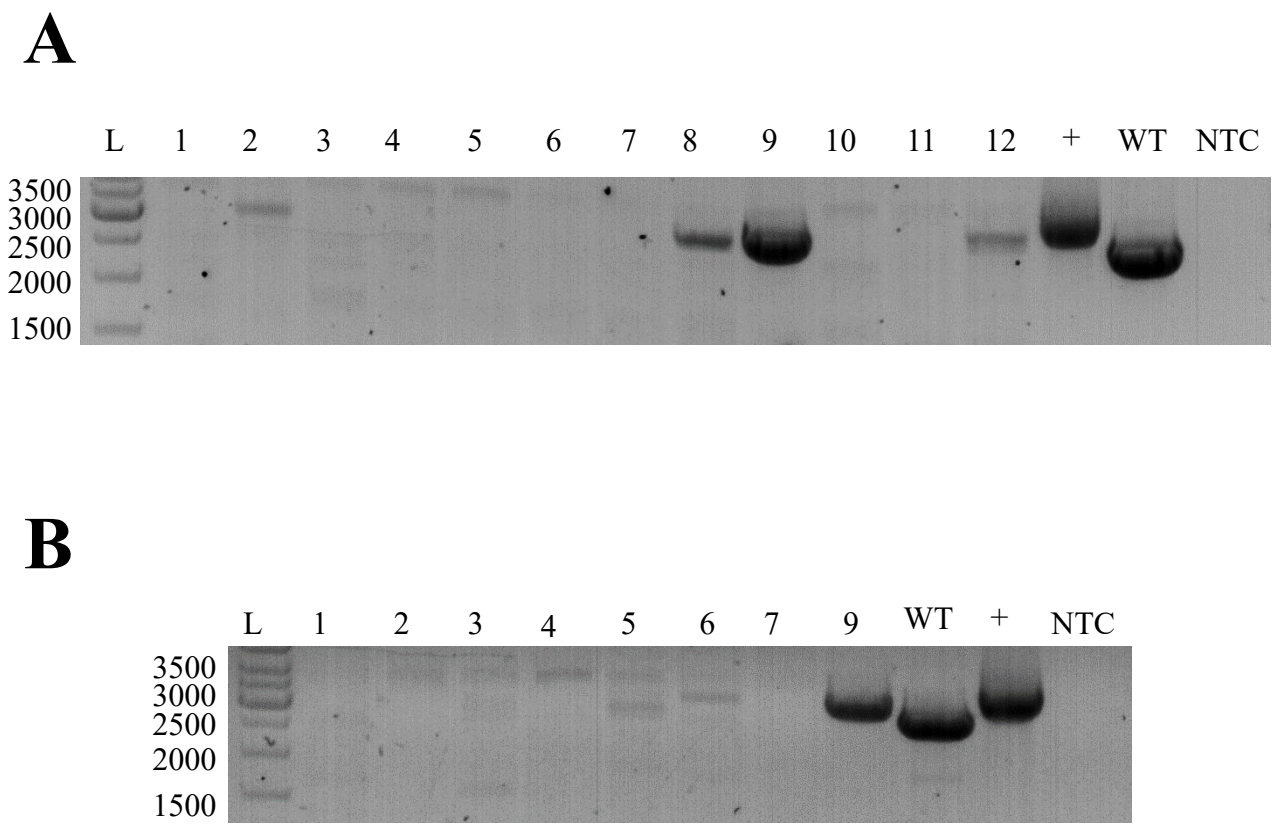


Figure - 6.10. Detection of ILTV recombinants lacking five miRNAs from the genome after passage in cells

Following plaque purification, genomic DNA was isolated from samples for use in PCR. Using primers spanning the miRNA cluster and appropriate controls, PCRs were carried out and visualised on an agarose gel (A). Bands visualised at the correct size (~3000 bp) were excised and cloned for sequencing. A second batch of recombinant viruses were tested (B) using the same method except with the addition of 6 % DMSO to the PCR reaction. Recombination events were detected and bands of the correct size (~3000 bp) were excised, cloned and sent for sequencing.

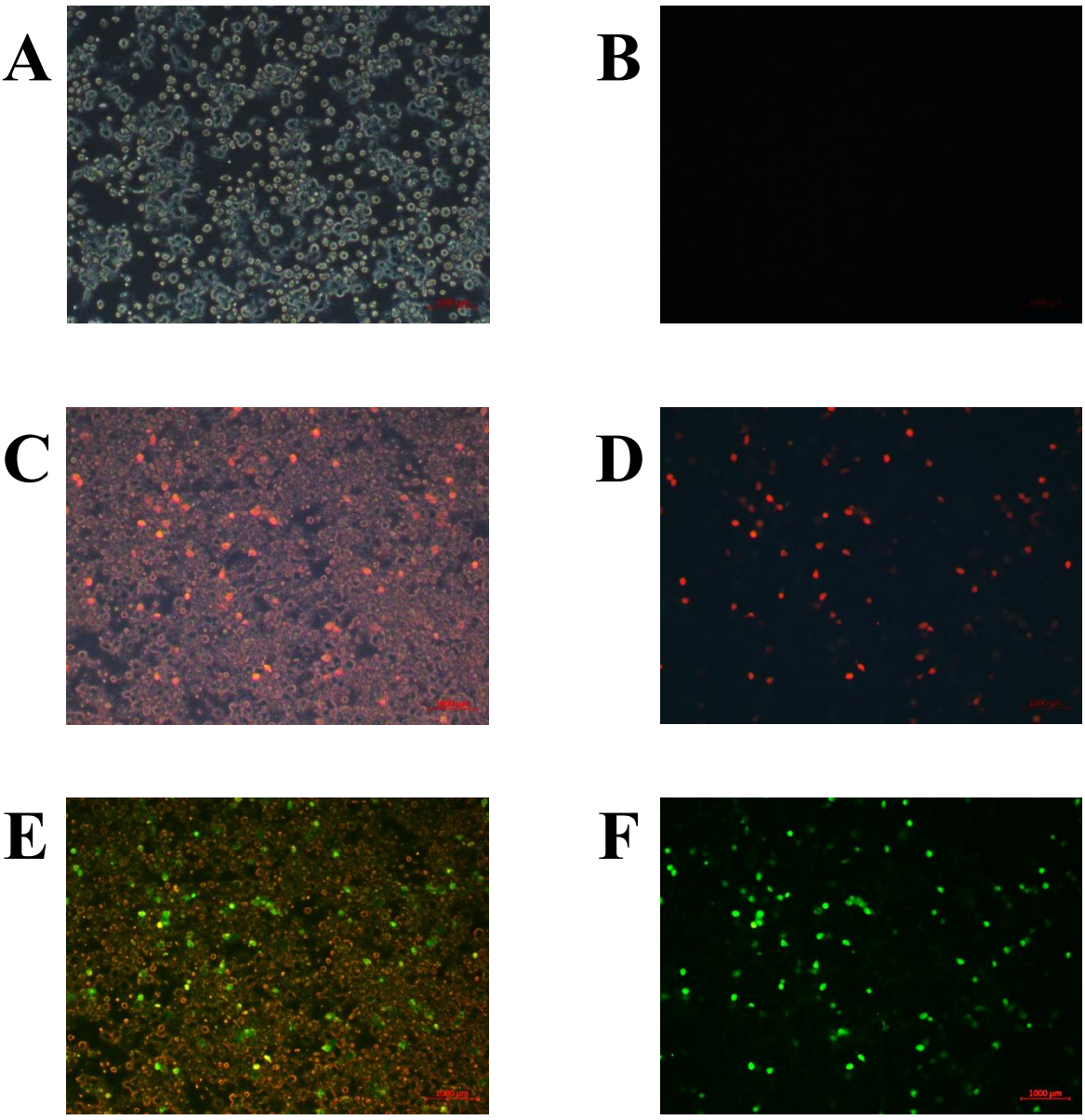


Figure – 6.11. Fluorescent microscope images of transfected and infected LMH cells
LMH cells were seeded out at $1x10^5$ cells per well in a 6-well plate. Cells were then transfected with a plasmid encoding for both the Cas9 protein and a sgRNA against the ILTV cluster and also a second plasmid which encoded for the homology repair cassette encoding the homology directed repair template. After 12 hours, cells were infected with WT ILTV at an MOI = 0.001. Untransfected LMH cells in brightfield phase and GFP phase (A&B). Expression of the sgRNA/Cas9 protein in brightfield + RFP and RFP only phases (C&D). Expression of the homology repair cassette in brightfield + GFP and GFP only phases (E&F). All images are taken 24 hours post infection of cells and are representative.

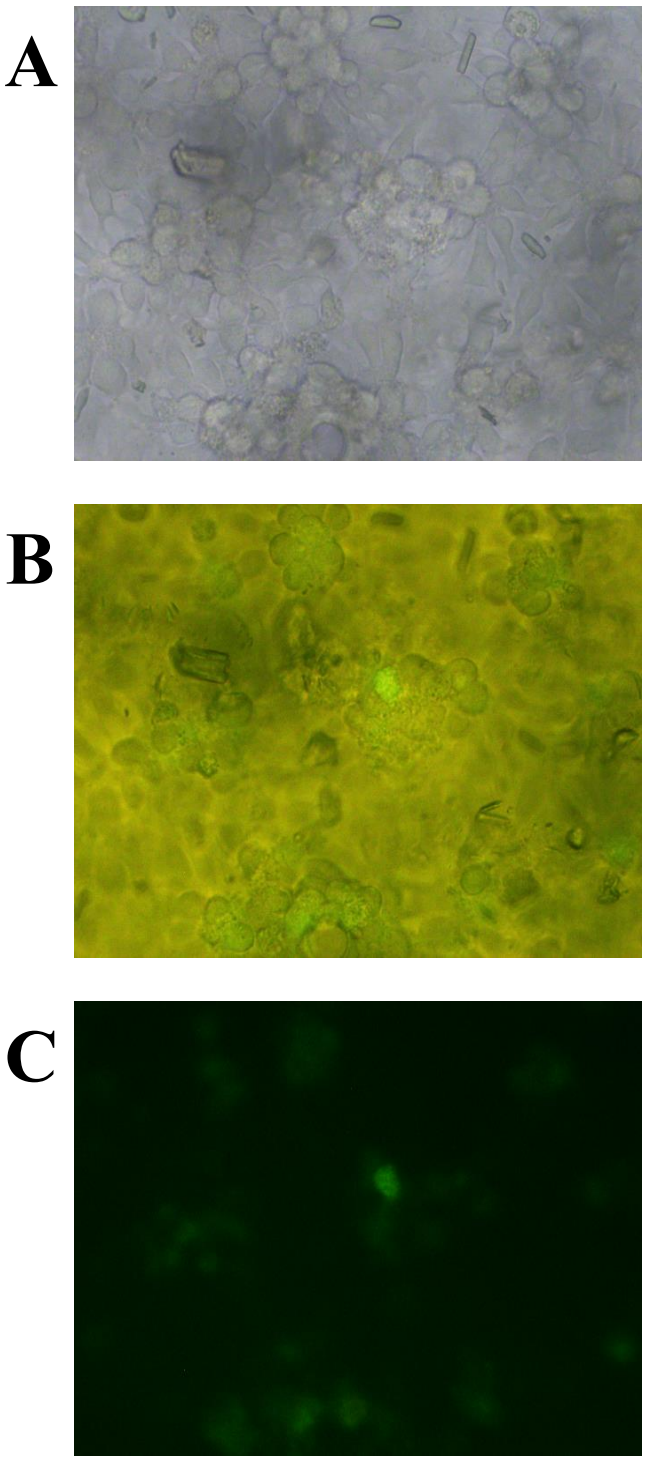


Figure – 6.12. Detection of eGFP following plaque purification of Recombinant ILTV virus

LMH cells were seeded out at $1x10^5$ cells per well in a 6-well plate. Cells were then infected with harvested lysates that had been freeze/thawed 3x and then placed under a 0.5 % Avicell overlay. Images were taken 24 HPI. No RFP images were taken due to lack of detection. A – Brightfield phase only. B – Brightfield + GFP phase. C – GFP only. Images are representative.

6.8 – Discussion

This project aimed to employ the gene editing technology CRISPR-Cas9 to create a recombinant ILTV virus lacking a cluster of five miRNAs found at the left hand end of the genome. To do this, sgRNAs were designed and tested *in vitro* to analyse their ability to direct a recombinant cas9 protein to cut the template DNA sequences. Two attempts, using two different methods, were made to create sgRNAs that were able to direct cutting of the desired sequence. It was found sgRNAs designed using an online algorithm were unable to direct the Cas9 protein to cause double stranded breaks however in contrast, sgRNAs designed in collaboration with Dr. Tan (The Roslin Institute) were able to direct and subsequently cause DNA cutting. Concurrently, a homology repair template which harboured a fluorescent reporter gene was assembled through multiple PCRs and cloning. This was then used alongside the sgRNAs in an experiment to create recombinant virus. Upon testing of the recombinant virus, recombination events were discovered and confirmed through sequencing indicating that the experiment was successful. Due to time limitations, further work to fully purify this virus and then characterise the virus *in vitro* for any possible growth/replication retardation has not been possible thus far.

Recombinant viruses are a very useful tool to decipher the roles specific viral genes play in terms of pathogenesis, replication and overall virus biology. To date, 19 genes have been knocked out of ILTV with two showing they are absolutely required for virus replication *in vitro* (Pavlova et al., 2013, Nadimpalli et al., 2017). The most heavily research gene of ILTV is Glycoprotein G. Deletion of the gene *in vitro* causes little effect upon replication kinetics whereas *in* vivo, significant attenuation is observed making such deleted viruses possible vaccine candidates (Coppo et al., 2011, Devlin et al., 2007, Devlin et al., 2006, Devlin et al., 2010). These recombinant viruses have led to a deeper understanding of not only the role each individual gene plays in terms of virus biology but they have also advanced the knowledge of virus-host interactions. In stark contrast though, there is very little research regarding the role of ILTV-encoded miRNAs. Of the 10 known miRNAs, only one so far, ILTV-miR-I5 has been shown to have a function. Waidner et al. (2011), showed that the miRNA was able to downregulate the protein expression of ICP4 by cleaving the mRNA transcript. The study also observed ILTV-miR-I6 was able to downregulate ICP4 expression however this was not statistically significant. Therefore, investigations into the role these miRNAs play during ILTV infection may uncover novel functions for them.

MicroRNAs can play a significant role in pathogenesis of virus infection. In MDV-1 for example, miRNA MDV1-miR-M4-5p has been shown to be a functional homologue of the cellular encoded miR-155 and they share common targets (Zhao et al., 2009). Further investigation involving the deletion of this miRNA from the viral genome observed the ablation of lymphoma formation as a result of MDV-1 infection *in vivo* (Zhao et al., 2011).

Creation of recombinant viruses has relied mostly upon natural recombination or the use of a BAC to create the desired knockout/fusion protein. Whilst the former has been extensively applied to ILTV, a BAC system has not successfully been created for ILTV. There have been recent reports of a creation of a cosmid system using 3 plasmids that encode for the full ILTV genome that can be propagated in yeast (Spatz. et al., 2018). With these limitations in mind, it was postulated that genome editing technology, in particular CRISPR-Cas9 may be of use to create recombinant viruses.

CRISPR-Cas9 has been used successfully to modify the genomes of other herpesviruses and delete protein encoding genes as well as miRNAs (Bi et al., 2014, Yuen et al., 2015). The study deleting UL23 from HSV-1 encountered issues regarding sgRNA targeting similar to this study. The inherent problem of using a 20 bp sequence as a guide is that there are off target effects. Whilst no mutations were found in the UL23 deletion studies when assessed by deep sequencing, the issue can be problematic (Bi et al., 2014). In this project, two methods of sgRNA were employed to maximise the chances of finding suitable sgRNAs that were able to direct cleavage of the miRNA cluster (shown in Figure – 6.1.).

Following conversations with Dr. S. Lillico (The Roslin Institute), sgRNAs were designed to target the right hand side of the miRNA cluster (shown in Figure – 6.1.). The first system of sgRNA design used a freely available online program along with the region to be targeted. In this instance, the chicken genome could not be selected as a 'target' genome highlighting the issues regarding lack of reagents once again. This was seen across several of the free sgRNA design tools and is problematic. Therefore, suggested sgRNAs had to be chosen upon their supposed efficiency in targeting the specified area minus the off targeting scores calculated by an algorithm against the human genome. From this, three sgRNAs were chosen at the top of the list with predicted efficiencies of 99 %, 99 % and 97 % in targeting the correct sequence. These high scores are most likely due to the differences in the human and chicken genome sequence though this is unclear. When these sgRNAs were tested *in vitro* (laid out in Figure – 6.2.), cleavage of template DNA by a Cas9 protein guided by these sgRNAs was undetectable at 1 and 24 hours post reaction assembly (Figure – 6.3A&B).

A second method of sgRNA design was also carried out in collaboration with Dr. W.S. Tan (The Roslin Institute). As detailed in section 6.3.1, this method used a custom Python script along with several downstream steps to filter the results. As well as designing sgRNAs to the right hand side of the miRNA cluster, sgRNAs were chosen at the left hand side of the cluster (highlighted in Figure – 6.1.). This created two double stranded breaks in the region and hopefully increase the likelihood of a recombination event occurring. The sgRNAs were then ranked on their combined scores of targeting the region (highest cutting efficiency to lowest) and their off targeting scores (lowest off targeting to highest) producing a list of sgRNAs for both the left and right hand side of the cluster (shown in Tables – 6.1 & 6.2). When these sgRNAs were tested *in vitro*, they were able to direct cleavage of the template DNA by the Cas9 protein (Figure – 6.4).

The difference in sgRNA design is not particularly comparable as in the first approach, one is relying upon free software where as in the second approach, custom scripts were utilised in the first instance which were then supplemented by online algorithms. The results of the approaches can however be compared. In a black and white sense, sgRNAs picked from the second approach were able to successfully direct cleavage of a template DNA sequence *in vitro* whilst the former could not. Not surprisingly, the sgRNAs from the first design attempt were identified in the second approach however their efficiencies were much lower than that of the sgRNAs chosen and taken forward (highlighted in Table – 6.1. The first set of sgRNAs are shown in yellow).

As well as having difficulty designing sgRNAs, which was overcome, the location of the miRNA cluster proved problematic. As seen with homologous recombination and also HDR via CRISPR, the flanking homology arms are generally considered to be around 1000 bp or longer to ensure correct targeting and maximise the chances of a recombination event happening (Byrne et al., 2015). This therefore proved a problem as the left hand flank was only 520 bp long in total which may have decreased the efficiency of recombination. New approaches have been developed which may aid future work and also increase the chances of recombination happening. These include using ssDNA donor templates as opposed to double stranded repair templates as they employ dramatically shorter homology repair arms whilst maintaining specificity (Yoshimi et al., 2016). This approach has not been applied to viruses thus far though it may be of use in genomic areas that are constrained by the genome size.

CRISPR-Cas9 modifications were first carried out using the Illinois strain of ILTV though this was soon replaced by the CSW-1 strain for a number of reasons. Firstly, the CSW-1

strain is more virulent and also has a full genomic sequence available for it on NCBI. Secondly, the virus produces more cell-free particles which can be isolated. This was seen as useful in the downstream purification steps to isolate a deletant virus. Finally, the virus has been used extensively for previous deletant viruses both *in vitro* and *in vivo* showing that it can be manipulated (Devlin et al., 2007, Nadimpalli et al., 2017).

The process of the transfection/infection was based upon the protocol from Russell et al. (2015). This was modified though with the addition of Phosphonoacetic acid (PAA). PAA has previously been to inhibit virus replication and can be washed out making it reversible (Elliott et al., 1980). The theory was that by adding PAA to the virus media following infection, virus would not be able to replicate giving the CRISPR-Cas9 system chance to cleave the DNA and allow host cell machinery to carry out homology directed repair (HDR) thus increasing the chances of creating a recombinant virus. This process was used through the project and no comparisons were made with the protocol lacking PAA but if the efficiency was increased, it may be worth noting for future work as it can be applicable to all other herpesviruses.

Detection of recombination events was carried out by PCR. Primers were designed to amplify the region spanning the miRNA cluster with a forward primer situated in the left flank region and a reverse primer downstream of the right homology arm. This was done so that PCR products would not be detected from just the homology repair cassette. The insertion of eGFP into the cluster caused PCR products to be ~300 bp bigger than the wild type sequence and so would be identifiable when separated by agarose gel electrophoresis. Moreover, the identification of a run of ~16 Cytosine's (Cs) contained within the miRNA cluster caused problems during attempts to create a 'revertant' homology repair cassette *(Data Not Shown)*. Whilst unclear, the presence of this sequence possibly caused the amplification of the wild type to be difficult without the use of DMSO. Therefore, the detection of PCR products at the correct size without DMSO was promising (Figure – 6.10A). Sequencing of these PCR products confirmed the presence of recombination events suggesting that a virus lacking the 5 miRNAs was successful. This first PCR analysis used the full virus sample and so recovery of these viruses was not possible but it confirmed that the system was working.

A follow up study using partial virus samples was carried out however this time 6 % DMSO was added to the PCR reaction. This was done to amplify wild type virus also to give a comparison to the recombination events (Figure – 6.10B). Recombination events were also detected in these samples at the correct predicted size which upon sequencing were

confirmed to be recombinant viruses. There were also some other bandings that were unexpected. These were most likely partial recombinants through partial insertion/deletion (indels) however in the interest of time they were not followed up on.

As well as using PCR analysis to look for recombinant viruses, fluorescent microscopy was used. This was carried out at all stages of the process from initial transfection/infection through to plaque purification steps. The Cas9 plasmid which also encoded for the specific sgRNAs also encoded for a mCherry (RFP) marker allowing for detection (Figure – 6.11C&D). At first, a GFP plasmid containing the sgRNA and Cas9 protein was used however this could not be distinguished from the homology repair template as this contained eGFP to replace the miRNAs, this was then changed to allow for a clear difference between the two plasmids.

During the plaque purification process, images were taken of the samples when they were infected with the virus samples. Images were taken whilst cells were under an Avicell overlay and this hampered efforts in obtaining clear images. GFP could still be detected in samples during the plaque purification process (Figure – 6.12) whilst no mCherry could be seen. These results suggested that the eGFP was incorporated into the virus and allowed for passage as detection of the plasmids (Cas9 mCherry) was not observed. The chance of the DNA plasmids surviving a freeze/thaw cycle 3x and then been able to get into naïve cells without a transfection reagent is very small. This gave more confidence that the green that was observed was coming from a recombinant virus.

Due to these difficulties, the ability to test this recombinant virus *in vitro* was not achieved however subsequent work should make this a priority. As the virus was never separated from wild type virus following two rounds of plaque purification, one must also question whether the virus lacking the 5 miRNAs can replicate without a helper wild type virus. The likelihood of this is slim but reports of this are seen in the literature, mainly through the deletion of protein coding genes however. Deletion of UL-1 from ILTV and rescue of recombinant virus could never be fully established suggesting that the gene is essential during virus replication though the function of UL-1 is still to be elucidated (Nadimpalli et al., 2017). Therefore, any future work should look to see if the two viruses can be separated from each other to create a pure population of recombinant virus.

In summary, work presented in this chapter has shown the successful application of CRISPR-Cas9 genome editing technology. This work has produced a virus that is lacking 5 miRNAs as shown by the PCR analysis. Future work must focus on purification of this virus which due to time constraints and technical issues made this not possible during the project.

Chapter 7: Concluding Remarks

At the beginning of this project, there was one major aim, to use a biochemical technique CLASH to investigate miRNA:mRNA interactions during ILTV infection. Following this, top targets were to be validated via a secondary laboratory method. Early on in the project this was deemed to be not feasible due to a number of factors laid out below and so the aims of the project evolved whilst still focusing upon ILTV miRNA:mRNA interactions. The modified aims of the project were to use bio-informatics to predict viral targets of ILTV-encoded miRNAs and to explore the use of CRISPR-Cas9 as a tool to manipulate the ILTV genome by deleting a cluster of five miRNAs from the genome. From this first modified aim, another project aim developed which investigated virus targets of high confidence cellular miRNAs.

CLASH is a technically challenging procedure even with optimal reagents for each individual process in the experiment. The difficulty in working with viruses that infect chickens is the current lack of reagents when compared to the human or mouse. This is not just limited to antibodies but includes cell lines relevant to virus biology and well annotated genomes of both the virus and host. The predominant cell line used for ILTV experiments are LMH cells which come with their own limitations. In terms of getting the CLASH project off the ground, the first step would be to identify a new cell line that was 1 – capable of supporting ILTV replication and 2 – removed the limitations seen with LMH cells such as the use of gelatin for cell adherence. Following this, for CLASH to work, this cell line would be need to be transduced with the same lentivirus expressing the tagged Ago2 to allow for the purification of RISC complexes. However, as explored in Chapter 6 of this thesis, the use of genome editing to 'tag' the endogenous protein maybe be more preferable. This has been shown to be feasible in chickens in a number of experiments (Oishi et al., 2016, Bai et al., 2016). As eluded to in the discussion of Chapter 3, a more suitable basis for CLASH would be look at the virus in a latent state, but as this is already technically challenging in more well defined viruses of the alphaherpesvirus family, carrying this out in the chicken may be some years away though it is something that should be aimed for in the long term as this would be beneficial with respect to virus – host interactions and elucidating the role miRNAs play in the latent state of the virus.

Due to the technical problems and limitations outline in Chapter 3 of this thesis, a new set of project aims were developed. These included the use of bioinformatics to predict viral targets of ILTV-encoded miRNAs. For the bioinformatic predictions, viral transcripts had to be made for each ORF in the viral genome which gave rise to another problem. As mentioned previously, the reference genome for ILTV does not contain all of the details one

would find in another virus such as HSV-1. Therefore, an arbitrary system was employed to ensure full coverage of the viral transcripts. Whilst this worked and predicted transcripts were generated for each ORF, it complicated the downstream analysis. This is highlighted in Chapter 4, section 4.4, with the viral target in UL29 first recognised in UL28. Due to this, if a target showed knockdown, further investigation had to be under taken to examine which transcript was actually targeted.

RNA Hybrid was used to make bioinformatic predictions for both Chapter 4 and Chapter 5 of this thesis. Previous members of the lab group have used this program successfully to predict miRNA targets (Riaz, 2014 , Nightingale, 2016). There are many programs available to predict miRNA:mRNA interactions however some are constraint to specific species' such as human targets making them unfeasible for this project. Whilst this gave RNA Hybrid the advantage as it more flexible in terms of species', it does produce a false positive predictions; something that is evident in Chapter 4. Of the 28 novel targets tested, only 3 showed a reduction in luciferase activity when tested *in vitro*.

The method of testing predicted targets *in vitro* was using a luciferase based reporter plasmid with the predicted target site cloned into the 3'UTR of a *Renilla* luciferase gene. It provided a direct link between the predicted miRNA target and the expression of a reporter gene. Due to the large number of targets tested from the bioinformatic screen, around 110 bp portions were cloned into the multiple cloning site of the plasmid. Using this approach, the context of the predicted target in the viral gene was lost and in some cases, moved the target site from the 5'UTR or coding region into the 3'UTR. Because of this, a secondary method of validation had to be undertaken to see if the target site was indeed real and not artificial.

In the case of the all of the targets that showed statistically significant knockdown in Chapter 4, the targets sites were contained within the coding region of the target gene. As there are a lack of reagents available to detect specific viral genes in ILTV, whole coding regions were cloned into an expression plasmid along with a small tag to allow for detection by western blotting.

For chapter 4, several miRNA targets were investigated further following the initial luciferase screen. These data results were inconclusive however there are several experiments that could be done in future that would answer the lingering questions from this part of the project. In in the first instance, replacing the small 6x His tag on the UL29 plasmid to another one such as HA or FLAG would possibly allow for the detection of the protein allowing for validation. Whilst this would be a small change, it might be of use interest to use Locked Nucleic Acids (LNAs) as outline in chapter one of this thesis as a

possible alternative (Xu et al., 2017). This could be used in conjunction with live virus allowing for RT-qPCR analysis of the interested transcripts. On a similar ilk, the targeting of UL46/UL48 could also be investigated. By using a scramble siRNA alongside the LNA of interest, one could look at the relative expression levels of the transcripts to determine if there was an enrichment in the presence of the LNA compared to the scramble control. In theory, this approach could replace CLASH also as to date, there are only 10 known ILTV-encoded miRNAs and this is unlikely to change.

Moreover, thinking about the investigations into the UL46/UL48 targeting, the data garnered in this project still does not fully answer which transcript is targeted and so this would need to be elucidated also. To do this, the employment of 5' and 3'RACE for the two transcripts would definitively state the transcription start and termination sites. There are several commercial kits available for this and any future work could use utilise these.

The final part of Chapter 4 investigated the interplay between UL46, UL47, UL48 and the ICP4 promoter. This work was very much in its infancy and there is a lot of scope for downstream experimentation. In the first instance, the data needs repeating in another cell line, preferentially one that is permissive to ILTV infection such as LMH cells to see if the data is similar to the results presented in Figure – 4.8. Secondly, one might investigate the sub-cellular localisation of the aforementioned proteins on their own and in the presence of one and other alongside the ICP4 promoter. This could be carried out using the already existing plasmids as each one was cloned with a different small tag to allow for immunofluorescencent detection. This could be done to see if the protein localisations are similar to their homologues from other herpesviruses. One would argue that the importance of this is less than that of the points for chapter 4 above as the interplay between these proteins is well defined in other herpesviruses however the results are interesting in the fact that UL46 possibly acts as a negative regulator though this still needs to be investigated further. Hypothetically, if this system was set up and was reliable, one could introduce the miRNAs that target the coding regions of these proteins to investigate their effects also but again, there are other methods that should be used first.

Once the viral transcripts were identified they were also tested against high confidence cellular miRNAs as listed in MiRBase (release 21, June 2014). This led to the creation of a new avenue of work that was investigated during the project (results outlined in Chapter 5). Reports of cellular miRNAs targeting herpesvirus transcripts are found in the literature but there are to date no reports with respects to ILTV (Pan et al., 2014). The same approach as Chapter 4 was taken which presented the same technical issues such as the arbitrary mRNA

transcript methodology and the filtering of results with some minor adjustments. Initial experiments focused on immediate-early genes as any effect on them is most likely to have an impact upon the virus biology. The results laid out in Chapter 5 focus on ICP4 and the cellular miRNA gga-miR-133a-3p. Similar to above, a secondary method of validation was not achieved and this is of up most importance. The quickest method to validate this interaction would be employ LNAs similar to above and carry out RT-qPCRs to look for ICP4 enrichment. At the time of writing this however, the approach taken was to clone both the 5'UTR and coding region of ICP4 in a plasmid with a small tag (HA) and validate the interaction via western blotting. This line of enquiry was taken for two reasons, 1 – validation of the interaction could be achieved and 2 – the creation of a natural promoter driven ICP4 protein would be advantageous to the lab for other ILTV related experiments.

This project has also been the first to formally identify the mature miRNA sequence of gga-miR-133a-3p in the chicken. One question that does need answering is how strongly expressed is this miRNA in the tissues of biological relevance. This could be done either using the existing method with the Qiagen MiScript II/SYBR green kit to estimate the quantity of the miRNA in the tissue (not carried out in this project) or by carrying out RNA deep sequencing on the tissues of interest to look at number of counts for the miRNA. The former of these approaches is feasible quickly and can be used with archived samples whilst the latter would require fresh samples in addition to expensive sequencing runs.

In a similar fashion, it would be of interest to look at the levels of expression within LMH cells. Coupled with this, the use of a miRNA sponge/anti-miR to remove the levels of miR-133a in the cells and subsequently infecting the cells to look for a difference in viral transcript levels would be of interest. This would probably mimic the levels seen naturally in the chicken as opposed to the overexpression of the miRNA as carried out in the project.

Finally for this part of the project, work carried out during whilst writing this thesis identified this miRNA interaction in several other species of Alphaherpesviruses in ICP4 as well as the miR-133a sequence (details outlined in Appendix 6&7). This opens up the possibility that this interaction is conserved within the alphaherpesvirus family however there is currently no experimental data to corroborate this suggestion. It would be of interest to see if these interactions are indeed real and whether the miRNA is expressed in tissues relevant to the individual virus species biology. Alongside this line of enquiry, work carried out by Dr I. Dry (The Roslin Institute) has shown that there is some differences in the sequence homology surrounding this miRNA site from different genotypes of ILTV

suggesting that this site may have higher incidences of mutation compared with other genomic regions of the virus (*Personal communication, Data Not Shown)*.

The final part of this project used CRISPR-Cas9 to create a mutant virus lacking 5 miRNAs. Results presented in this thesis suggest that a recombinant virus was successfully produced. Polyclonal populations of recombinant and wild type virus were frozen down as a pure population was not achieved by the end of the project. Purification of the virus is of the up most importance to allow for downstream experimentation. This could be done by further plaque purification as outlined in the project to create a monoclonal population.

Following the creation of a monoclonal population, the first stage would be to make a revertant virus whereby the miRNAs are knocked back into the recombinant virus and then confirm this by sequencing. Once the panel of viruses are made, they could then be tested *in vitro* for their growth kinetics. This can be done through two means. An outright look at the viral fitness by plaque assay using a set amount of input virus and measuring the output. This could look at both the absolute value of viral titre but also look at the plaque size phenotype akin to data presented in Chapter 5. Secondly, one could look more in depth at the growth kinetics by RT-qPCR for viral transcript levels at different time points. In theory the project could then go one of two ways dependent upon the data generated. If a phenotypic difference was observed, this virus could then be used *in vivo* to study the effects within the host and measure clinical scores alongside the other parameters typically measured *in vivo* as laid out in Devlin et al. (2006). Downstream of this, it could be a vaccine candidate which would open up a completely new avenue of research. If no phenotypic difference was seen, the virus might be tested *in vivo* anyway as several papers have shown that recombinant viruses *in vitro* behave very different when used *in vivo* (Devlin et al., 2007, Garcia et al., 2016).

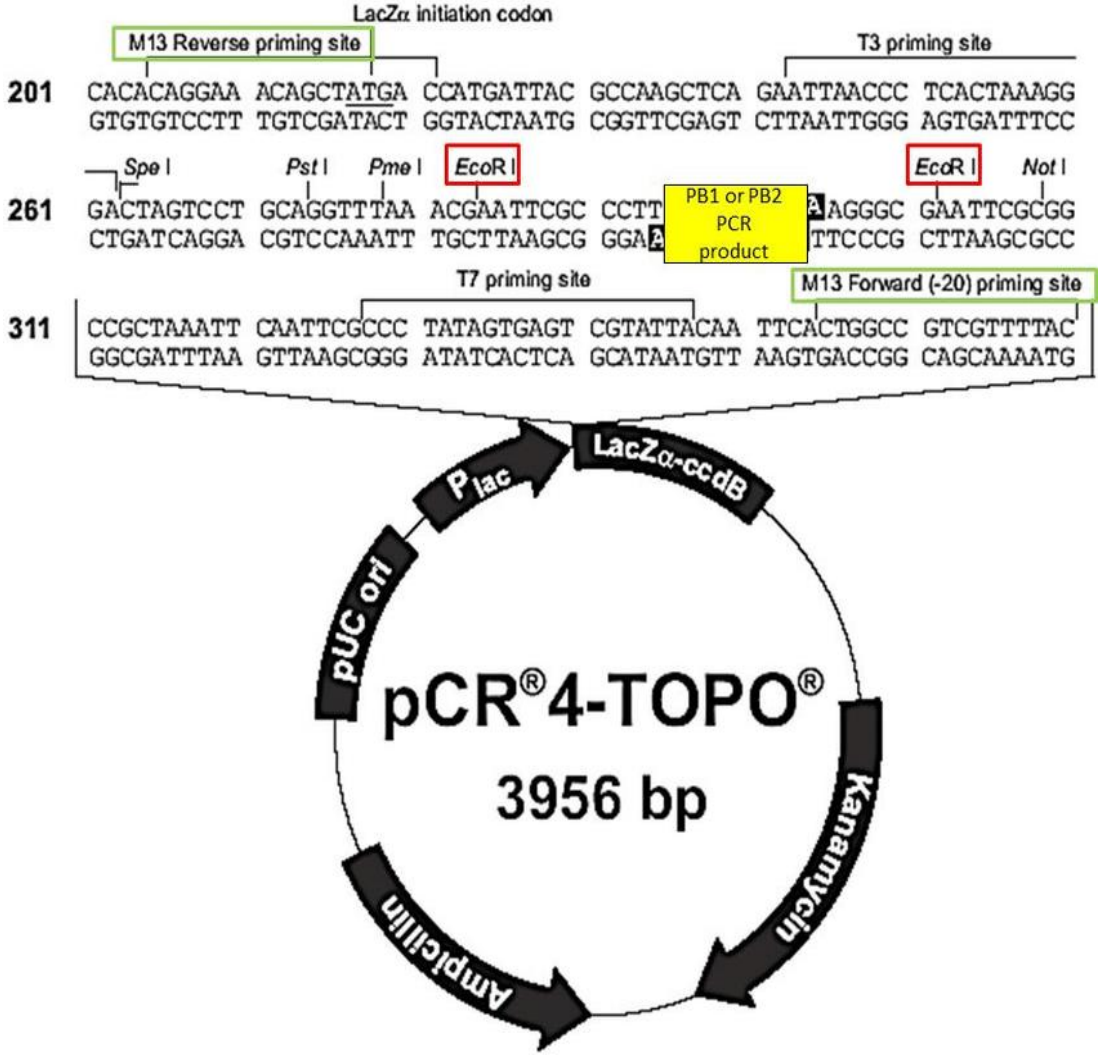
As well as *in vivo* experiments, one might also create a series of single miRNA knockout viruses to look at the individual effects of each miRNA though this is something to only consider as the work laid out above would take precedence.

Results described in this thesis present some novel findings with regards to ILTV biology and microRNAs as well as outlining a new method of manipulating the ILTV genome. Data described goes some way to elucidating mechanisms of viral gene regulation by both ILTV-encoded and cellular miRNAs. Whilst the story is incomplete, mostly due to the technical hurdles outlined above and the time constraints, this thesis has created more questions than it has answered. Future work will be able to build upon these findings and help elucidate the role of miRNAs during ILTV infection.

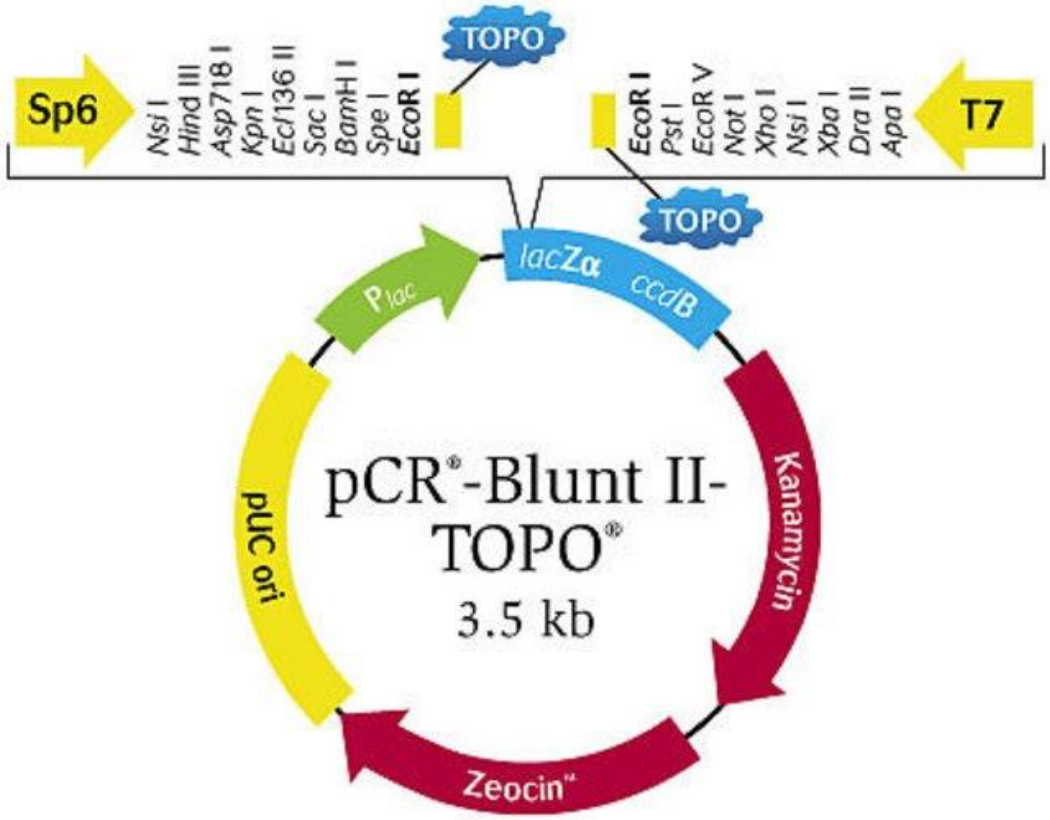
Appendices

Appendix 1: Vectors and Plasmids

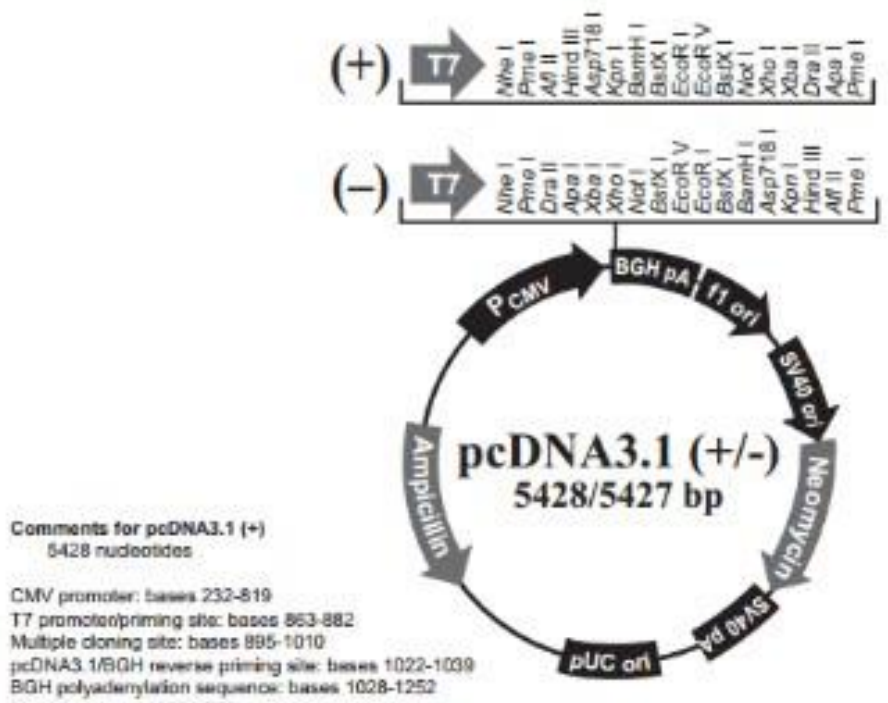
pCR4-TOPO



pCR-Blunt-II-TOPO



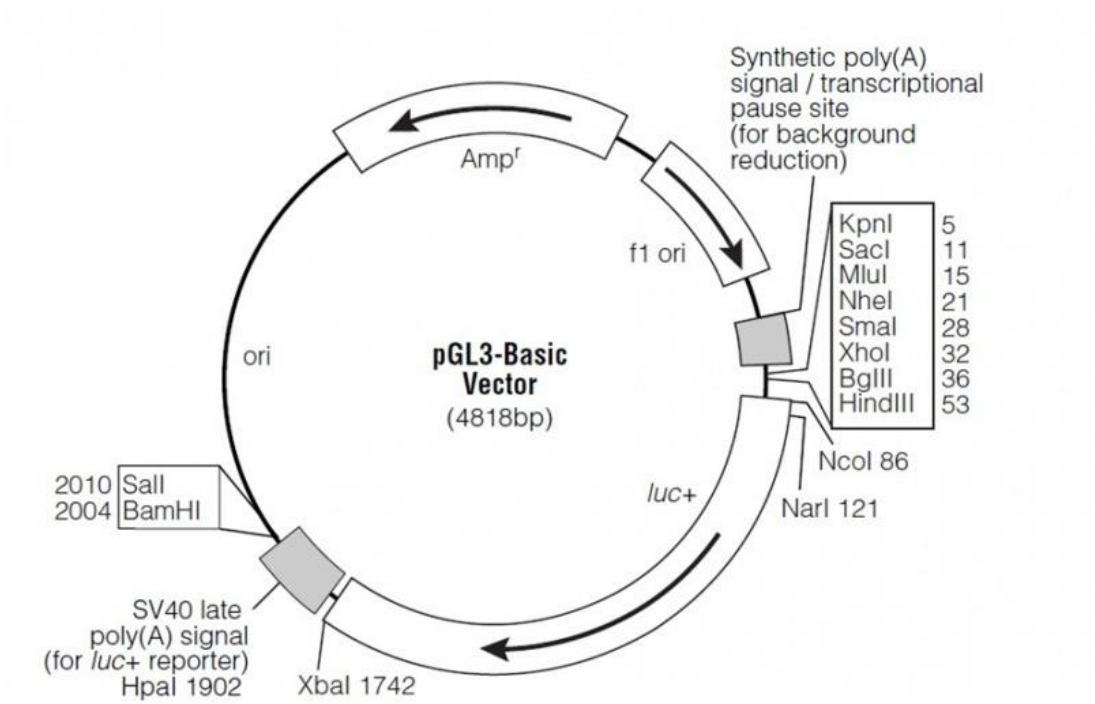
pcDNA3.1(+)



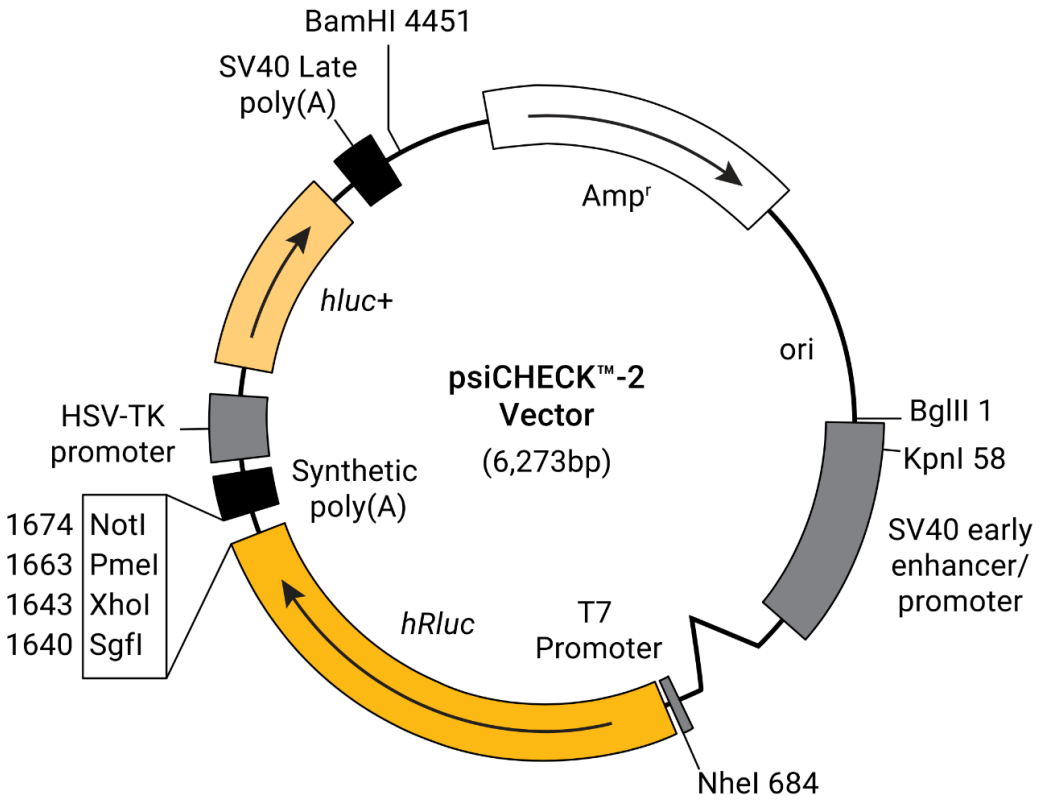
Comments for pcDNA3.1 (+)
5428 nucleotides

CMV promoter: bases 232-819
T7 promoter/priming site: bases 863-882
Multiple cloning site: bases 895-1010
pcDNA3.1/BGH reverse priming site: bases 1022-1039
BGH polyadenylation sequence: bases 1028-1252
f1 origin: bases 1298-1726
SV40 early promoter and origin: bases 1731-2074
Neomycin resistance gene (ORF): bases 2136-2930
SV40 early polyadenylation signal: bases 3104-3234
pUC origin: bases 3617-4287 (complementary strand)
Ampicillin resistance gene (*bla*): bases 4432-5428 (complementary strand)
ORF: bases 4432-5292 (complementary strand)
Ribosome binding site: bases 5300-5304 (complementary strand)
bla promoter (P3): bases 5327-5333 (complementary strand)

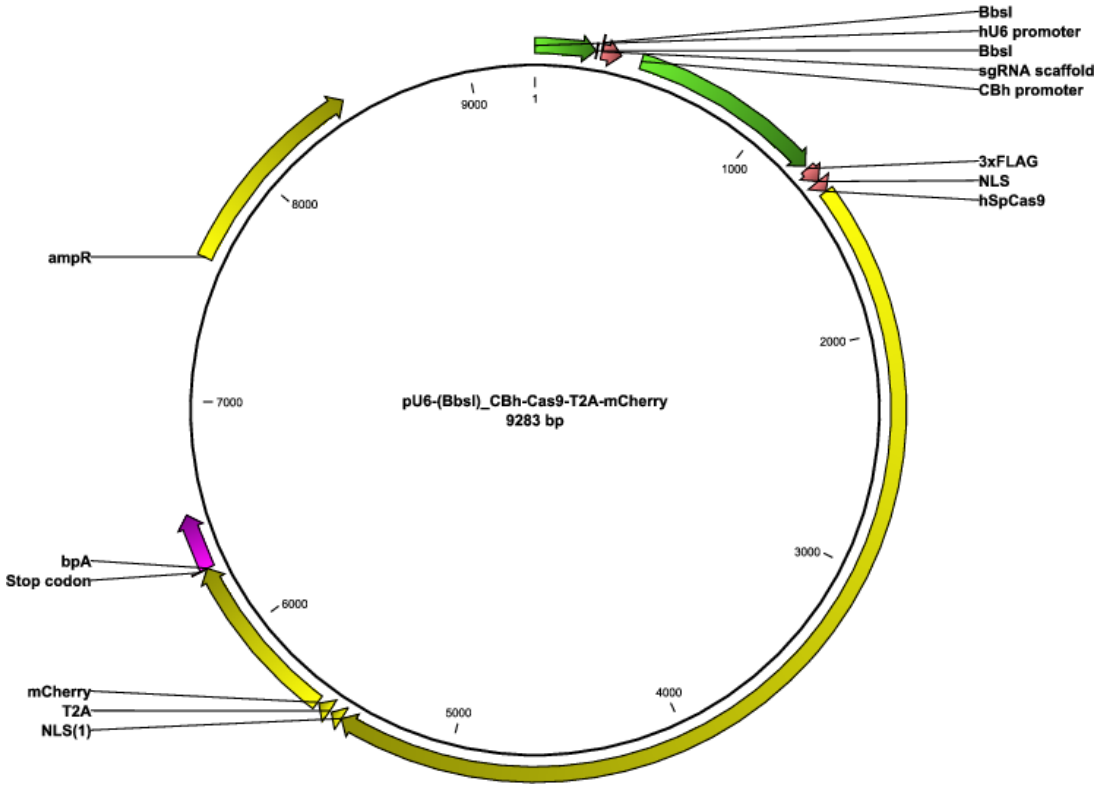
pGL3-BASIC


Synthetic poly(A) signal / transcriptional pause site (for background reduction)
Amp^r
f1 ori
ori
pGL3-Basic Vector
(4818bp)
luc+
KpnI 5
SacI 11
MluI 15
NheI 21
SmaI 28
XhoI 32
BglII 36
HindIII 53
NcoI 86
NarI 121
2010 SalI
2004 BamHI
SV40 late poly(A) signal (for luc+ reporter)
HpaI 1902
XbaI 1742

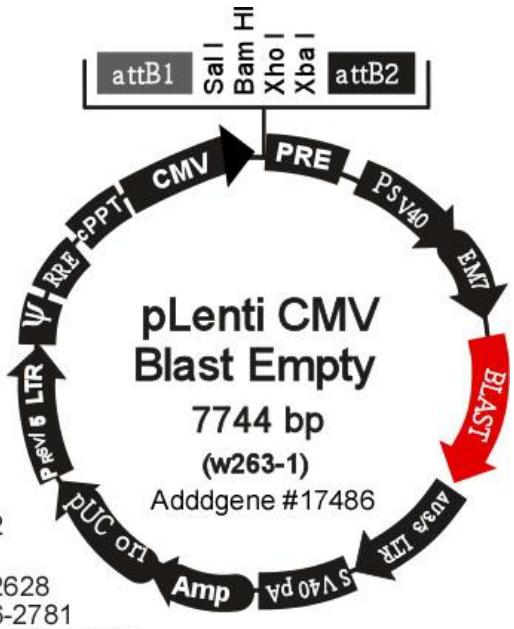
psi-CHECK™-2


BamHI 4451
SV40 Late poly(A)
Amp^r
hluc+
ori
psiCHECK™-2 Vector
(6,273bp)
HSV-TK promoter
Synthetic poly(A)
BglII 1
KpnI 58
SV40 early enhancer/ promoter
1674 NotI
1663 PmeI
1643 XhoI
1640 SgfI
hRluc
T7 Promoter
NheI 684

PX458-mCherry



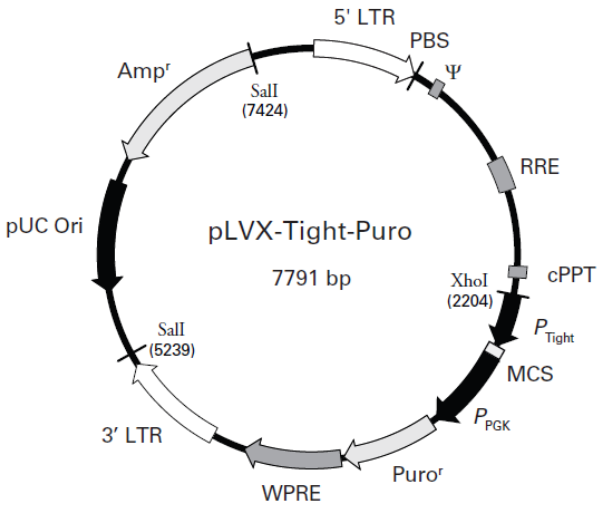
pLenti CMV Blast Empty



comments for pLenti CMV Blast Empty
7744 nucleotides

bla promoter: bases 32-131
ampicillin resistance gene: bases 132-982
pUC origin: bases 1138-1810
RSV/5'LTR hybrid promoter: bases 2218-2628
HIV-1 psi (ψ)packaging signal: bases 2736-2781
HIV-1 Rev response element (RRE): bases 3272-3525
3' splice acceptor: base 3862
3' splice acceptor: base 3901
Central polypurine tract (cPPT): bases 4011-4073
CMV promoter: bases 4145-4730
attB1 site: bases 4764-4788
Multiple cloning site (MCS): bases 4809-4857
*att*B2 site: bases 4857-4881
Woodchuck post-transcriptional element (PRE): bases 4911-5504
SV40 early promoter and origin: bases 5568-5930
EM7 promoter: bases 5934-5999
Blasticidin resistance gene: bases 6000-6398
ΔU3/3'LTR: bases 6486-6718
SV40 polyadenylation signal: bases 6789-6942

PLVX-Tight-Puro Vector



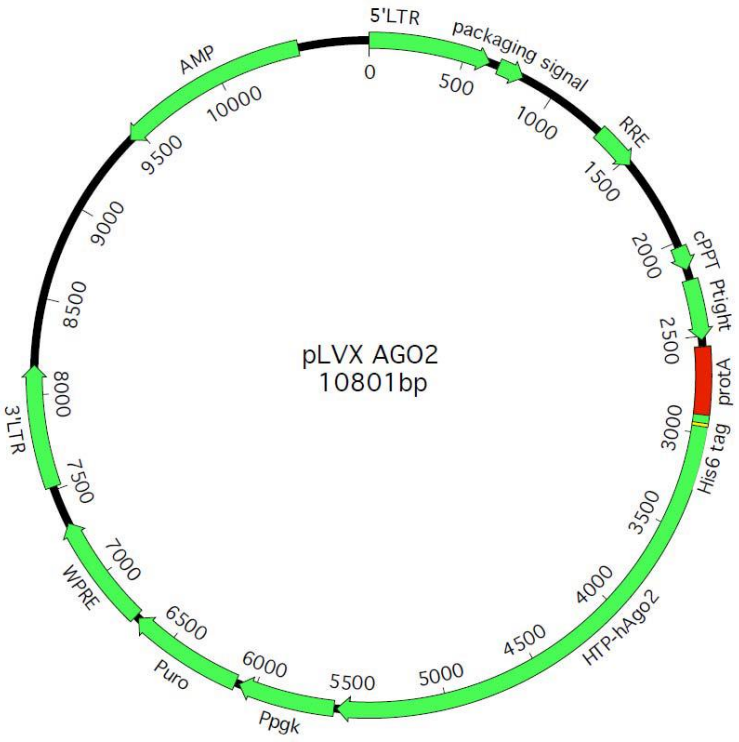
BamHI MluI
NotI XbaI* EcoRI

```
2521  TGGAGAAGGA TCCGCGGCCG CGCCGGCTCT AGATCGCGAA CGCGTGAATT CTACCGGGTA
      ACCTCTTCCT AGGCGCCGGC GCGGCCGAGA TCTAGCGCTT GCGCACTTAA GATGGCCCAT
```

Xba I site (*) is methylated in the DNA provided by Clontech Laboratories, Inc. If you wish to digest the vector with Xba I enzyme, you will need to transform the vector into a dam- host and make fresh DNA.

pLVX-Tight-Puro Vector Map and Multiple Cloning Site (MCS).

PLVX-Ago2 Vector



Appendix 2: Primers and Oligonucleotides

miRNA RT-qPCR Primers

miRNA name	Sequence	Annealing Temperature (°C)
ILTV-miR-I1-5p	GCGGCGAGACTGATTGGGGAAT	74
ILTV-miR-I1-3p	GCGGCGAATTCCATTCCTCTTTCT	72.4
ILTV-miR-I2	GCGGCGGGAAGGCTGTGCGATAGG	81
ILTV-miR-I3	GCGGCGTCTTGTCTCTGGGTGGG	78.3
ILTV-miR-I4	GCGGCGATGTATAGCGAGCAATGA	74.1
ILTV-miR-I5-5p	GCGGCGCTTCTCGTCCCCGTCTTC	80.8
ILTV-miR-I5-3p	GCGGCGTGAAGAAGACGACGACGAG	78.7
ILTV-miR-I6-5p	GCGGCGGTCTCCTGTACCCTCA	75.3
ILTV-miR-I6-3p	GCGGCGACGCTGAGGGGCCATGA	84.4
ILTV-miR-I7	GCGGCGTTTTAATACTGAGGTGC	69.2
gga-miR-92a-3p	GCGGCGTATTGCACTTGTCCC	72.6
gga-miR-133a-3p	GCGGCGTTGGTCCCCTTCAACC	77.8

Sense Viral Target Gene Long Oligonucleotides

Gene		Targeted by	Sequence (**sense**)
ORF F	1	I6-5p	tttggCTCGAGgccaccgtggcggtcgcagccaccctcgtatgatgaggcaatgggggatggacctttactacgactgggggtcgcgacctcgctcacgcGCGGCCGCaggcc
	2	I6-3p	ctgggGTTTAAACggtcgcgacctcgctcacgcaggccggtcgtcgccggtctcgaggtcgatctcggcgacgcaattggtgtgcagcaaagttgtgctcagaGCGGCCGCggca
	3	I6-3p	cacaaCTCGAGgcattggtgagtgcacaagctattagcagcatggtctctggctctgcttcctcagtgggcgtagaagtagactgtgggtacagtcagactGCGGCCGCcatat
UL54	1	I6-3p	ctttcCTCGAGtgtagaatgagatctgcggcgaatacaatgtgcttcgtggcgggattctgacctcagcgtgaacggccatccagggcgatcttttgtccgGCGGCCGCggccg
UL48	1	I5-3p	taataCTCGAGttggcgctgagttccagggaactcctaagccatcttttggctacaaatccttcttcattgagggaagtagccagctgcccgaaaatacttGCGGCCGCttcta
	2	I6-5p	accaaCTCGAGactcaagcgcagcatccctgccccctacacaaggctcaatgcaaacacaaaacaggagacgcgaacgacgcagtaacagtgtcgttatccGCGGCCGCgaaag
UL46	1	I5-3p	Same as UL48
	2	I6-5p	
UL30	1	I6-5p	cgtttCTCGAGggcttgaaatcagactctacagtaagcccccgtatgagaggagcgatggcgcaggagggctggtcttgacttttggcgcacagtatcttcGCGGCCGCcattc
	2	I5-3p	tttccCTCGAGgaagattgtagtgatgggataactgtatgcaaatgtggctctgagttcgagcttcttctgtgctttatgacttttttcaagcaatattcaGCGGCCGCcccga
	3	I3	agccgCTCGAGgatttgaggattttgactgtcttccgagcggacttactaagcttggccggttgattgcagaggcaaggctagctattaccggcaacggacGCGGCCGCtaaac
UL28	1	I2	gcggcCTCGAGaaccagagtcaaaaatgttccgttcttctgccatgccggccctaagatcagcctttaacgggatgctggacaaaggattcctttctggaaGCGGCCGCaatat
	2	I5-5p	ggagaCTCGAGcacaacctgtcacagaaggaactagtacaagcgacctgactgctaacctcagaggaactattgaaagggacgaggattcgatagaagttcGCGGCCGCttcta
	3	I5-5p	cctccCTCGAGacgaccctctagacgttactacaaggatgaggaggaggttgaggaggatagtgatgaggacgacaggatacttgccaccagagttctgaaGCGGCCGCaggcc
UL24	1	I6-5p	aatatCTCGAGcccctagagactgtaattgatgtcttggatgatttagcccagcgggccgtgcaggagaaggacattgttgggtcttataaaacactagacGCGGCCGCatccg
	2	I1-5p	tggatCTCGAGaatgcttagtattttcgcgatcgcgtctctgtttagacaataaaagggttatatctttctgatcagtccgtctgttttgtcagtgtgttGCGGCCGCgatac
UL9	1	I6-3p	atagtCTCGAGagtagccgctcattttctgcgctatgtgaattagcgcctacgtcgctcagcgggaggaacgatatgtcgccaaataattttgctcctggaGCGGCCGCgaaca
	2	I6-3p	ctggaCTCGAGgaacaactgtggttgtcgcccagcgtcggactcgcacgaagactgtacggatgcgatctcagcgatcgcctattatccaatccgacaatgGCGGCCGCagtag
UL2	1	I2	aggttCTCGAGatcatccctcccctcgccatctagacgctcccctccacagctttcgccccaccacccctcgtctaattatgatggcgcagggcaaaatacGCGGCCGCgactg

	2	I6-5p	gccgcCTCGAGaccatgtgatagtaattcgcgagataacggtgcgcgaattttcgggccgttgccatcgggagactactttgaaaatttgactgtgattatGCGGCCGCttaca
	3	I2	atccgCTCGAGcaagatgtgcgtgttattatcgttgggcaagacccgtatcccacggaaggacacgcgcatggcttagctttcagtgtccctaggggtgcGCGGCCGCcgtat
UL-1	1	I6-5p & I5-5p	actctCTCGAGgagtacgacccaccgccggaaatggtggagggttggcatggagggggacagggcgggagggccgagagcccgcagccgctggctgatgtgGCGGCCGCcgtga
	2	I6-5p & I5-5p	SAME AS ABOVE AS THEY OVERLAP
	3	I3	gcgtgCTCGAGaaatattacgcaatgtaggtttggacgatagaagaattgggcctcacaggggcagggatagataccagttaaggtccaggtctaggaaccGCGGCCGCgatcc
US3	1	I2	agatgCTCGAGccgggcgcacgattaccgataatgtactcggacgatcgtaactcgccatagttttcactgcgtgaaccaattctttccatccagaatccgGCGGCCGCagagc
	2	I4	ctcaaCTCGAGaacgcgtttgcgtattggatagtttctcacggacaatgtcattgcgcccctatgcagaaattttgccgaccgcggaaggcgtcgagcgccGCGGCCGCtcgcc
	3	I6-5p	ttatcCTCGAGtgcgacgtccgtcgctacaccgaggaagagcaacgtcgaagaggggttaacagtactaaccaggggaaatcaaaatgtaagcgcctgataGCGGCCGCgctaa
sORF1	1	I5-5p	aatggCTCGAGaacaagggcggcctctcagctggaaaatgaaattttggttctcgggcgcctaaatcacgagaatgttctcaagatccaggaaatccttcgGCGGCCGCgtacc
ICP4	1	I6-5p & I6-3p	agtaaCTCGAGtggacaggcgactcactgtctcatggcccctcagcgttgggaagaccgacaacgacgatgagggtacaggagacagagagaagtgtcctcGCGGCCGCccgtt
	2	I5-5p & I5-3p	tcttcCTCGAGgtcgtcttcctcttcctcatgctcctcgtcgtcgtcttcttcagactccagctcatctgaagaagacggggacgagaagaacgagaaagaGCGGCCGCagatc
Note, ICP4 targets have both strands of corresponding miRNA in them. EG - both I5-3p and I5-5p			
Red = Xho1 and Blue = Not1 and Green = Pme1			

Antisense Viral Target Gene Long Oligonucleotide

Gene		Targeted by	Sequence (**antisense**)
ORF F	1	I6-5p	ggcctGCGGCCGCgcgtgagcgaggtcgcgacccccagtcgtagtaaaaggtccatcccccattgcctcatcatacgagggtggctgcgaccgccacggtggcCTCGAGccaaa
	2	I6-3p	tgccGCGGCCGCtctgagcacaactttgctgcacaccaattgcgtcgccgagatcgacctcgagaccggcgacgaccggcctgcgtgagcgaggtcgcgaccGTTTAAACcccag
	3	I6-3p	atatgGCGGCCGCagtctgactgtacccacagtctacttctacgcccactgaggaagcagagccagagaccatgctgctaatagcttgtgcactcaccaatgcCTCGAGttgtg
UL54	1	I6-3p	cggccGCGGCCGCcggacaaaagatcgccctggatggccgttcacgctgaggtcagaatcccgccacgaagcacattgtattcgccgcagatctcattctacaCTCGAGgaaag

UL48	1	I5-3p	tagaaGCGGCCGCaagtattttcgggcagctggctacttccctcaatgaagaaggatttgtagccaaaagatggcttaggagttccctggaactcagcgccaaCTCGAGtatta
	2	I6-5p	ctttcGCGGCCGCggataacgacactgttactgcgtcgttcgcgtctcctgttttgtgtttgcattgagccttgtgtaggggcagggatgctgcgcttgagtCTCGAGttggt
UL46	1	I5-3p	Same as UL48
	2	I6-5p	
UL30	1	I6-5p	gaatgGCGGCCGCgaagatactgtgcgccaaaagtcaagaccagccctcctgcgccatcgctcctctcatacgggggcttactgtagagtctgatttcaagccCTCGAGaaacg
	2	I5-3p	tcgggGCGGCCGCtgaatattgcttgaaaaaagtcataaagcacagaagaagctcgaactcagagccacatttgcatacagttatcccatcactacaatcttcCTCGAGggaaa
	3	I3	gtttaGCGGCCGCgtccgttgccggtaatagctagccttgcctctgcaatcaaccggccaagcttagtaagtccgctcggaagacagtcaaaatcctcaaatcCTCGAGcggct
UL28	1	I2	atattGCGGCCGCttccagaaaggaatcctttgtccagcatcccgttaaaggctgatcttagggccggcatggcagaagaacggaacatttttgactctggttCTCGAGgccgc
	2	I5-5p	tagaaGCGGCCGCgaacttctatcgaatcctcgtccctttcaatagttcctctgaggttagcagtcaggtcgcttgtactagttccttctgtgacaggttgtgCTCGAGtctcc
	3	I5-5p	ggcctGCGGCCGCttcagaactctggtggcaagtatcctgtcgtcctcatcactatcctcctcaacctcctcctcatccttgtagtaacgtctagagggtcgtCTCGAGggagg
UL24	1	I6-5p	cggatGCGGCCGCgtctagtgttttataagacccaacaatgtccttctcctgcacggcccgctgggctaaatcatccaagacatcaattacagtctctaggggCTCGAGatatt
	2	I1-5p	gtatcGCGGCCGCaacacactgacaaaacagacggactgatcagaaagatataacccttttattgtctaaacagagacgcgatcgcgaaaatactaagcattCTCGAGatcca
UL9	1	I6-3p	tgttcGCGGCCGCtccaggagcaaaattatttggcgacatatcgttcctcccgctgagcgacgtaggcgctaattcacatagcgcagaaaatgagcggctactCTCGAGactat
	2	I6-3p	ctactGCGGCCGCcattgtcggattggataataggcgatcgctgagatcgcatccgtacagtcttcgtgcgagtccgacgctgggcgacaaccacagttgttcCTCGAGtccag
UL2	1	I2	cagtcGCGGCCGCgtattttgccctgcgccatcataattagacgaggggtggtggggcgaaagctgtggaggggagcgtctagatggcgaggggagggatgatCTCGAGaacct
	2	I6-5p	tgtaaGCGGCCGCataatcacagtcaaattttcaaagtagtctcccgatggcaacggcccgaaaattcgcgcaccgttatctcgcgaattactatcacatggtCTCGAGgcggc
	3	I2	atacgGCGGCCGCgcacccccctagggacactgaaagctaagccatgcgcgtgtccttccgtgggatacgggtcttgcccaacgataataacacgcacatcttgCTCGAGcggat
UL-1	1	I6-5p & I5-5p	tcacgGCGGCCGCcacatcagccagcggctgcgggctctcggccctcccgccctgtccccctccatgccaaccctccaccatttccggcggtgggtcgtactcCTCGAGagagt
	2	I6-5p & I5-5p	SAME AS ABOVE AS THEY OVERLAP
	3	I3	ggatcGCGGCCGCggttcctagacctggaccttaactggtatctatccctgcccctgtgaggcccaattcttctatcgtccaaacctacattgcgtaatatttCTCGAGcacgc
US3	1	I2	gctctGCGGCCGCcggattctggatggaaagaattggttcacgcagtgaaaactatggcgagttacgatcgtccgagtacattatcggtaatcgtgcgcccggCTCGAGcatct
	2	I4	ggcgaGCGGCCGCggcgctcgacgccttccgcggtcggcaaaatttctgcataggggcgcaatgacattgtccgtgagaaactatccaatacgcaaacgcgttCTCGAGttgag
	3	I6-5p	ttagcGCGGCCGCtatcaggcgcttacattttgatttcccctggttagtactgttaacccctcttcgacgttgctcttcctcggtgtagcgacggacgtcgcaCTCGAGgataa

sORF1	1	I5-5p	ggtacGCGGCCGCcgaaggatttcctggatcttgagaacattctcgtgatttaggcgcccgagaaccaaaatttcattttccagctgagaggccgcccttgttCTCGAGccatt
ICP4	1	I6-5p & I6-3p	aacggGCGGCCGCgaggacacttctctctgtctcctgtaccctcatcgtcgttgtcggtcttcccaacgctgaggggccatgagacagtgagtcgcctgtccaCTCGAGttact
	2	I5-5p & I5-3p	gatctGCGGCCGCtctttctcgttcttctcgtccccgtcttcttcagatgagctggagtctgaagaagacgacgacgaggagcatgaggaagaggaagacgacCTCGAGgaaga

Mutant Viral Target Gene Long Oligonucleotides

Gene		Targeted by	Orientation	Sequence
UL48	2	I6-5p	Sense	accaaCTCGAGcaagcgcagcatccctgccccctacacaaggctcaatgcaaacacaaacccaacaacgcgaacgacgcagtaacagtgtcgttaGCGGCCGCgaaag
			Antisense	ctttcGCGGCCGCtaacgacactgttactgcgtcgttcgcgttgttgggtttgtgtttgcattgagccttgtgtaggggcagggatgctgcgcttgCTCGAGttggt
ICP4	1	I6-5p & I6-3p	Sense	agtaaCTCGAGacaggcgactcactgtctcatggcccctcagcgttgggaagaccgacaacgacgatgagggtcccaacaacagagagaagtgtcGCGGCCGCccgtt
			Antisense	aacggGCGGCCGCgacacttctctctgttgttgggaccctcatcgtcgttgtcggtcttcccaacgctgaggggccatgagacagtgagtcgcctgtCTCGAGttact
UL29	1	I2	Sense	gcggcCTCGAGaaccagagtcaaaaatgttccgttcttctgccatgccggccctaagaAGCGCGTTCaacgggatgctggacaaaggattcctttctggaaGCGGCCGCaatat
			Antisense	atattGCGGCCGCttccagaaaggaatcctttgtccagcatcccgttGAACGCGCTtcttagggccggcatggcagaagaacggaacatttttgactctggttCTCGAGgccgc
NB - Purple colour shows mutation site				
Red and blue colours refer to the same restriction endonucleases as above				

Viral Gene Targets of Cellular miRNAs (Long Nucleotide)

Gene		Targeted by	Orientation	Sequence
ICP4	1	133a-3p	Sense	aaccaaCTCGAGtattctgttataaattttggggtgggattaggtgggctgtcattatttgaggggaccacgaaatatccttttgcatggggtatcgtgcGCGGCCGCaaaaat
			Antisense	atttttGCGGCCGCgcacgataccccatgcaaaaggatatttcgtggtcccctcaaataatgacagcccacctaatcccaccccaaaatttataacagaataCTCGAGttggtt
	2		Sense	ttgtatCTCGAGttccaggagcggccgtggatcctacggcggcgtctgcttctggggaccaccctgtactcgttggcgccaggcgaatagcgcgtgccacGCGGCCGCgaggga
			Antisense	tccctcGCGGCCGCgtggcacgcgctattcgcctggcgccaacgagtacagggtggtccccagaagcagacgccgccgtaggatccacggccgctcctggaaCTCGAGatacaa
UL20	1		Sense	aataacCTCGAGgcagaccaagttccaacaatgtccgtcaaaatgtccctgctcaacgtgggggaccatcttgtcagcatagaagcactagagcgtgtttGCGGCCGCacacgc
			Antisense	gcgtgtGCGGCCGCaaacacgctctagtgcttctatgctgacaagatggtcccccacgttgagcagggacattttgacggacattgttggaacttggtctgcCTCGAGgttatt

Mutant Viral Gene Targets of Cellular miRNAs (Long Nucleotide)

Gene		Targeted by	Orientation	Sequence
ICP4	1	133a-3p	Sense	aaccaaCTCGAGtctgttataaattttggggtgggattaggtgggctgtcattatttgCAAAACCCCcgaaatatccttttgcatggggtatcgGCGGCCGCaaaaat
			Antisense	atttttGCGGCCGCcgataccccatgcaaaaggatatttcgGGGGTTTTGcaaataatgacagcccacctaatcccaccccaaaatttataacagaCTCGAGttggtt

UL46/48 RT-PCR Primers

Gene	Primer Sequence
UL48 Universal Forward	tagctgattccgcccttgat
UL48 Reverse (Set 1)	cgacttagctgtgttagctg
UL48/46 Reverse (Set 2) and (UL46 Set 1)	cgacgcagtaacagtgtcgt
UL46 Forward	gctcggccagtcctactgaa

ICP4 5'UTR RT-PCR Primers

Gene	Primer Sequence
Universal Reverse	attggagcggccaaattact
ICP4-1000 Forward (Set 1)	cttggcactcccaggaagcg
ICP4-500 Forward (Set 2)	taagtcatgaaaccaatatt
ICP4-100 Forward (Set 3)	ccaccagaaagcttcacgtt
ICP4-475 Forward (Set 4)	attaggtgggctgtcattatttga

Protein and Promoter Cloning Primers

Gene	Primer Sequence
ICP4p Forward	GGTACCtgcttcccggtgtggccaataac
ICP4p Reverse	GGTACCCTCATCAACAATTGGAGCGG
UL46-HA Forward	attgcagcgGATATCgccaccATGgctgaaggcgatggg
UL46-HA Reverse	TACAGTTATCTCGAGACAGGCTTAAGCGTAATCTGGAACATCGTATGGG TAATTCATTACTA
HA-UL46 Forward	CAGCGGATATCGCCACCATGTACCCATACGATGTTCCAGATTACGCTGC CACCATGGCTGAAGGC
HA-UL46 Reverse	AGTTATACACTCGAGGGCTTAATTCATTACTACGTAATG
UL48-FLAG Forward	ggttccaccATGGAAgaagaatcttccactggagcc
UL48-FLAG Reverse	tgtttagggcatCTTATCGTCGTCATCCTTGTAATCaggtgtatc
FLAG-UL48 Forward	ggttccaccATGGAAgattacaaggatgacgacgataaggaagaatcttccactggagcc
FLAG-UL48 Reverse	TTAGGGCATAGGTGTATCAAG
6xHis-UL29 Forward	gagataactcgcGGTACCGCCACCATGgaacatcatcaccatcaccacaagagttcatctggc
6xHis-UL29 Reverse	gcgcagttaGCGGCCGCctgTTAacagaacagaatatcag

UL46 Outer Forward	ggtgggaaagccaacagct
UL46 Outer Reverse	ttattggtccgggagcactt
MUT-UL29 1	tttgtccagcatcccgttgaacgcgcttcttagggccggcatggcag
MUT-UL29 2	ctgccatgccggccctaagaagcgcgttcaacgggatgctggacaaa
MUT-UL46 1	tactgcgtcgttcgcggcgacgattttgtgtttgcattgagccttgtgtagggg
MUT-UL46 2	Cccctacacaaggctcaatgcaaacacaaaatcgtcgccgcgaacgacgcagta

qPCR Primers

Gene	**Primer Sequence**
ICP4 Forward	TGTGGAGGAGTTCATGGTCC
ICP4 Reverse	CAGAGCTAATGACACACGGC
ChCYP Forward	GAGGGAGACAAGCCAAAGTT
ChCYP Reverse	GAGGGAGACAAGCCAAAGTT
Firefly Luciferase Forward	TCCATCTTGCTCCAACACCC
Firefly Luciferase Reverse	TGCGTCGAGTTTTCCGGTAA

CRISPR primers

Gene	Primer Sequence
miR K/O gRNA 1 (Set 1) SENSE	CACCGCGGGCTAGGTCGCACTTGCT
miR K/O gRNA 1 (Set 1) ANTISENSE	CGCCCGATCCAGCGTGAACGACAAA
miR K/O gRNA 2 (Set 1) SENSE	CACCGCGGGCTAGGTCGCACTTGC
miR K/O gRNA 2 (Set 1) ANTISENSE	CGCCCGATCCAGCGTGAACGCAAA
miR K/O gRNA 3 (Set 1) SENSE	CACCGCAAGTGCGACCTAGCCCGC
miR K/O gRNA 3 (Set 1) ANTISENSE	CGTTCACGCTGGATCGGGCGCAAA
miR left side K/O 1 SENSE	CACCTCCGCAGAGGAGACTGATTG
miR left side K/O 1 ANTISENSE	AGGCGTCTCCTCTGACTAACCAAA
miR left side K/O 2 SENSE	CACCGATTTCGCGAGGCTCCGCAG
miR left side K/O ANTISENSE	CTAAAGCGCTCCGAGGCGTCCAAA
miR Right side K/O 1 SENSE	CACCTCCGTGATGAGGGAACCACA
miR Right side K/O 1 ANTISENSE	AGGCACTACTCCCTTGGTGTCAAA
miR Right side K/O 2 SENSE	CACCTATAGCGAGCAATGACCGTG
miR Right side K/O 2 ANTISENSE	ATATCGCTCGTTACTGGCACCAAA
GΔG gRNA 1 SENSE	CACCGACGTACTCGTCCAGCGGAC
GΔG gRNA 1 ANTISENSE	CTGCATGAGCAGGTCGCCTGCAAA
GΔG gRNA 2 SENSE	CACCGGACCAGAGTAGTCGCTCCA
GΔG gRNA 2 ANTISENSE	CCTGGTCTCATCAGCGAGGTCAAA
GΔG gRNA 3 SENSE	CACCGACCCTTGGTAGTTACGTGTC

GΔG gRNA ANTISENSE	CTGGGAACCATCAATGCACAGCAAA
IVT miR K/O gRNA 1 (Set 1) Forward	TTAATACGACTCACTATAGGCGGGCTAGGTCGCACTTGCTGTTTTAGAGCTAGAAAT
IVT miR K/O gRNA 2 (Set 1) Forward	TTAATACGACTCACTATAGGCGGGCTAGGTCGCACTTGCGTTTTAGAGCTAGAAAT
IVT miR K/O gRNA 3 (Set 1) Forward	TTAATACGACTCACTATAGGCAAGTGCGACCTAGCCCGCGTTTTAGAGCTAGAAAT
IVT GΔG gRNA 1 Forward	TTAATACGACTCACTATAGGACGTACTCGTCCAGCGGACGTTTTAGAGCTAGAAAT
IVT GΔG gRNA 2 Forward	TTAATACGACTCACTATAGGGACCAGAGTAGTCGCTCCAGTTTTAGAGCTAGAAAT
IVT GΔG gRNA 3 Forward	TTAATACGACTCACTATAGGACCCTTGGTAGTTACGTGTCGTTTTAGAGCTAGAAAT
IVT Universal Reverse	AAAAGCACCGACTCGGTGCC
IVT miR cut site (right side) Forward	TTCGGAGAGTGCGGGATTTT
IVT miR cut site (right side) Reverse	CCACTGTTAAACTAACCGTTACT
IVT GΔG cut site Forward	GCCACCGTTTCCCTAGTATG
IVT GΔG cut site Reverse	CGTCTAGATAAACAGACCCGGT
IVT miR cut site (left side) Forward	TTTTCCCAGAACCGAGGCG
IVT miR cut site (left side) Reverse	CAATTCAGCCGAGGATTTGG
IVT miR left side K/O 1 Forward	TTAATACGACTCACTATAGGTCCGCAGAGGAGACTGATTGGTTTTAGAGCTAGAAAT
IVT miR left side K/O 2 Forward	TTAATACGACTCACTATAGGATTTCGCGAGGCTCCGCAGGTTTTAGAGCTAGAAAT
IVT miR right side K/O 1 Forward	TTAATACGACTCACTATAGGTCCGTGATGAGGGAACCACAGTTTTAGAGCTAGAAAT
IVT miR right side K/O 2 Forward	TTAATACGACTCACTATAGGTATAGCGAGCAATGACCGTGGTTTTAGAGCTAGAAAT
GΔG Left Flank Forward	CACTCGATATCATGGACGCAGCC
GΔG Left Flank Reverse	GGATCCTCATGGTACCAGCTGAAGTTGTCTCTCTCCCCTC
GΔG Right Flank Forward	GGTACCATGAGGATCCCCACCCGAGAGTGTTTTT

GΔG Right Flank Reverse	CGGAGTCCGCCGGCGAATAATTGG
miR K/O Left Flank Forward	GGATTCGAAACCCCTCGCGGCC
miR K/O Left Flank Reverse	GGATCCTCATGGTACCTCGCGAAATCCAACGGCGGGCGGTCC
miR K/O Right Flank Forward	GGTACCATGAGGATCCCTCATCACGGAGTGCTCGTTATGC
miR K/O Right Flank Reverse	ATAGCACTTGCTCTCGTTCTGTATA
CBH Promoter Forward	GGCTCTAGAGGTACCCGTTACATAACTTACGGTAAATGGC
CBH Promoter Reverse (GFP overlap)	CAGCTCCTCGCCCTTGCTCACCATGGTGGCCCAACCTGAAAAAAAGTG
GFP Forward (CBH overlap)	GTTGGGCCACCATGGTGAGCAAGGGCGAGGAGCTG
GFP Reverse	CGGCCGCGGATCCTCCCCAGCATGCCTGCTATTC
CMV Promoter Forward	GGTACCATAGTAATCAATTACGGGGTCATT
CMV Promoter Reverse (GFP overlap)	CTCGCCCTTGCTCACGCTTATATAGAC
GFP Forward (CMV overlap)	GTGAGCAAGGGCGAG
GFP Reverse (CMV)	GGATCCACTAGAATGCAGTGAAAAAAATGC

Sequencing Primers

Gene	Primer Sequence
psiCHECK-2 Forward	TGCTGAAGAACGAGCAGTAA
psiCHECK-2 Reverse	CGAGGTCCGAAGACTCATTT
PX458 gRNA Forward	GCCTATTTCCCATGATTCCT
GFP Universal Forward (CRISPR)	ACATGGTCCTGCTGGAGTTCGTGA
GΔG-GFP Reverse (CRISPR)	TTCCCTAAAGGCCGTAAACGCGAGGACGC
miRΔ1-5-GFP Forward (CRISPR)	CTTAGGCGCGGTGTTGCTAAG
miRΔ1-5-GFP Reverse (CRISPR)	TTCGTGTCAGTGCAGTTTCGC
M13 Forward	ACTGGCCGTCGTTTTAC
M13 Reverse	GGAAACAGCTATGACCATG

Appendix 3: ILTV Virus Transcripts used for Bioinformatic Analysis

Virus Transcripts

>UL56

AATTCTTTTATTTTTAGCTATATTGTATCTACAGTGGGGTCCGTCATCGTAGAGAAGCCGCAGAGAGGCTGCTATTTCTGCTTTT
AATACACAACATTCAGAACGATGCGGGCTTAAGACGGAACCGCGCGGCACAAGGAGCATAACTGCGGGGTCAGTGTTTGTTGCTG
CGTGATTGTAATTCTCTTTTTGGCACTGCGAGGCTAGGTTTGTTGCCGCGCGATTGCAGGGTTTTTCTTTGTGAGACTGGTGATG
GGGAAAAGCTTGATTGTGCAATAGTAAAAGTGACTTAGGTACTTCATATACTGGGCTGTGGCATGTAGGCACGTCCAAGCTCGGC
TCTGAATTCAAGCTTCTTTGTCGAACCAGTTTTGTTTTCTTTTTGGGGGAGGGTAGCACACTCTGCCCGAGTCTCGGCATTGACT
TAACAGTGATGTGAAACCCGGAAGATCGAGCATGAACTAATAGCATTAAAGAATTGTTATCCGAGGAATAATCGTGGACGCGAAT
TTACTCGACCGCTAAAATCTTTCTTCTACTGAGCTGGATACGTGAAATTTGGTGAGTATAACCTCTCGGGATACATAGCTTTTAA
ATACGGGGCGTGCAATATTAAATTCTGCACTCGGGGCTGCAATGGAGCGCGGAGCTTTATTTGACAAGACCGCCAATTGCAAGGA
CTGGGTCTCGGTCGGAACTACTGTGTGGGCGGATCGATGCAGATGACGGGGACGACTTAGTCTGGGATTATGAAAATAGCCCATA
TCCAAGCATAGTTTCCTCACTATTCCCGGGGGAAGAAACGGACTCGGCAATTTGTAACTCTGTTGTTGCCGCAAACCCCTGTAGC
ATACCTCCTGGGCGGCACGTTTGGCGTGGCCATGCTGCTGTTTTCGTCGGCCAGACAGCTCCTCCGTCCCGCGCGTGGAAGTTAA
TGCTCGCCTTGTTGCCGCGGTTGCACTGATAATTTTCTCATTGCTTGTAGTGATCTGTGTTGCGTCATATTGGGGGTAACATGTC
TTCAGAGGACACATCGGGATTCCTAACGCCCCCCGCAAGTGATGACGACACTGACCCTTCCGAGCCACCACCAAATTTATGGGAT
CCTCACCATGACGATTTTCCGAGGGACGCTGATTCCCCAAACCCACTTTTCTACCCCTGGGATGACTCTGTGAATAATACTGGGG
ATACGGGCAGTAACGAAGATGACTATGTAGATATGGGAGGGGTAGGTGGATCCGAAGACTATGAAGACCTCGGTACGGGCGGGGA
CTCTGACTATGACAATGTATCTACAGCGACCGGCGGGACGTGGTTTCCTTCCCTTACTTCTTGGTCCTCTGATGACCACGGCCCA
ACTTCTCCGGAAAACCCTATGCAACAACTTCAAGTAACAATTCAGCAGGATTCAGATCCACAGCAGGAACCCGATCCCCAGCAAG
TTCCCGGTCTCCAGCAGGAACCTGACCCCCAGCAAGATCCACGAGAGCCTCATGATCCTCCTCCCTATAGTCCGCCCCCAGAGGA
CCCTTTTGGGCTCTCGCCATTTACTAGTGGGATGGGCGGGTTTGGGCCACCGTGGCGGTCGCAGCCACCCTCGTATGATGAGGCA
ATGGGGGATGGACCTTTTACTACGACTGGGGGTCGCGACCTCGCTCACGCAGGCCGGTCGTCGCCGGTCTCGAGGTCGATCTCGG
CGACGCAATTGGTGTGCAGCAAAGTTGTGCTCAGAGGCATGTTTGTGGCATCTCTTTTGGGTTGGAGTGGCCTTAATGTGTTGGT
GGCTCTTGTATCTTATTTTGCGCATTGTCTGGGGACAGACTCCGGGATAAGGAAGGTTGTATCCGCATCCAGTACTCCTCAATAA
AAGCGTGGTGGTGCTACACGATGTCTGTTAATTTTACAACTCCATTTTAC

>ORFF

CATATTGGGGGTAACATGTCTTCAGAGGACACATCGGGATTCCTAACGCCCCCCGCAAGTGATGACGACACTGACCCTTCCGAGC
CACCACCAAATTTATGGGATCCTCACCATGACGATTTTCCGAGGGACGCTGATTCCCCAAACCCACTTTTCTACCCCTGGGATGA
CTCTGTGAATAATACTGGGGATACGGGCAGTAACGAAGATGACTATGTAGATATGGGAGGGGTAGGTGGATCCGAAGACTATGAA
GACCTCGGTACGGGCGGGGACTCTGACTATGACAATGTATCTACAGCGACCGGCGGGACGTGGTTTCCTTCCCTTACTTCTTGGT
CCTCTGATGACCACGGCCCAACTTCTCCGGAAAACCCTATGCAACAACTTCAAGTAACAATTCAGCAGGATTCAGATCCACAGCA
GGAACCCGATCCCCAGCAAGTTCCCGGTCTCCAGCAGGAACCTGACCCCCAGCAAGATCCACGAGAGCCTCATGATCCTCCTCCC
TATAGTCCGCCCCCAGAGGACCCTTTTGGGCTCTCGCCATTTACTAGTGGGATGGGCGGGTTTGGGCCACCGTGGCGGTCGCAGC
CACCCTCGTATGATGAGGCAATGGGGGATGGACCTTTTACTACGACTGGGGGTCGCGACCTCGCTCACGCAGGCCGGTCGTCGCC
GGTCTCGAGGTCGATCTCGGCGACGCAATTGGTGTGCAGCAAAGTTGTGCTCAGAGGCATGTTTGTGGCATCTCTTTTGGGTTGG
AGTGGCCTTAATGTGTTGGTGGCTCTTGTATCTTATTTTGCGCATTGTCTGGGGACAGACTCCGGGATAAGGAAGGTTGTATCCG
CATCCAGTACTCCTCAATAAAAGCGTGGTGGTGCTACACGATGTCTGTTAATTTTACAACTCCATTTTACAGGTGATCTAGAGAG
ACGCTGAGTGGCACTTGTCCCGACGGGACCATGCAGTCGAACAGCAGCGATCGAGGCCCAGTGTGATGATGGTCGAGGAGGGATG
GTCGTCCATAGCTCCAGGTGATGCACTGGATACAGATTTCATTCCAGGGCCTTGTGCCACGTCCATACATGGTATATCCAAGACA
GTTTATTTTTTTCTGTGTGGAGTTAATCTGGAGGAATGTAGTACACTCCCACAGCATGTCCAATCTCACCCATATGGACATCCTG
AGCTGAAATCAGGCAAATGGTACAAGAGGTTTTGCTCCGGGCTAGGCGAAATTGGAGATACAAGCCAGTGTCAGCTGACACGACT
ATGCTGCACTTCCGGAATGCCGGCACAGATTTTTGGGCCTTCGAGATTCAGGTCTCTGCAACAGAAGCCAACCCATATGCGGGCC
CAAGATTTGCTCACTAGGCCTTGCCATATACTAGAGTTCGATGTTGGCGCTGACCTAATCAATCTTTTCTTGTATATGGAACCAT
GTTCAGGGAATCGATATTGCGTACATTTAGGATACCATAAAACTAATGCCATGCGTGTTTTGAGCGGTGGTGGGATTCTATGGGG
CAGACTTCCGTGGAAGGACAACACCGAGGAGCACGGGTACTCGTTGCCTATGCGAGTATTTGGGATCAAATTGCCCCATAAAGTT
TATGTGGCATGTCGCTGCCCTGCAACTCGGACGGAACTATTATTTGGTGAGGGGGGGGTAGGATTCAACGCGGAAAACTTTAAAC
AGTGCGGACGGTTGAAAAAAGAGTGTGAATGTCTGCAGAACCGTTGTTTTACTGCACAAACGGTGTTAGGTGCGGCATGTAAGTT
TACTGTATACTCGGGCAAGGGACGAGGTCAAGAAATTCTGCTATATTCAGGATTCATGAATGCTACAACGGTAATGCCTGTAGTA
CTGGGTATGTTAAGTAGAGAACCCCACAGGTGTGCAGGTACACTCATACTGTCCAGGTCCTCTGGAAATTGCCGTGGATTTCATG
AGACCCAACACGATATTCCCACTAACCCGGGTCTGTATCCTCTGTGTAATCATGAGCACCCTTACTATGTGACAGTTACAGATGT
ATGCGGCAACTGTTGTTCAGGCTTGAGCGGGTTTTTGGGAGAATAGCTGCCCCTGCTGGTCTAAGCTCCGTATCTGTATCCATTA
AAGGCTCCACCCACAGCGGGACTGACGTGACAGAAGAACGTGAAGAGGACCCAGGGACACAGCAAACCTCCCACGACAAATTGCC
GGAGCGCAACCGCATGGGAGATCAAAATTCGAATTTGCGGGGAAGAGATCAATATTGGCCGCCTGCCCCACACCGTAGTCATTGT
CACTCGGATTTTATATTCGATGAACCTGAGCCAGAAAGTGGGGAAGACGTGCATAACATGCATCCCCCACGAGGTGCAGATGAGC
AAACAGCCGCTTCTGTGTCAGCGCTAATGCAAAGTCTAGCACAAGCATTGGTGAGTGCACAAGCTATTAGCAGCATGGTCTCTGG
CTCTGCTTCCTCAGTGGGCGTAGAAGTAGACTGTGGGTACAGTCAGACTCATATTACAGAGGGGCCGGGGAGGGAACAATTCGGT
AGAGTCCCAGAACGAGGGCCAGAGTATCCTCAAGATTACTGTGATATATATGGTCCTGTAAGTAATGGGCCTGCTGGATACAGAG
CAGGACCAGATGCTCCTAGTATACAAGATAGGACCTTCCCATGCGGCAGAAGATGTGACGAAGCATGGCTTGCCTTAGAAGTAGG
GAATATGCCTTGGATTTCTTCTGGTTCACATAGTCCACCTTCTCAGTATCATAACCCTTATGGTTCACATAGTCCACCTTCCCAG
TATCATAACCCTTATGGTACATATAGTCCGCCTTCTCAGTCTCATAACCCTTATGGCTCATATAGTCCACCTTCCCAGTATCATA
ACCCTTATGGTACATATAGTCCGCCTTCTCAGTCTCGTAAGCATGACTATTCACCTCCATATCCGATACTCAAACCAAAGCCTCG
ATTACCCCCAGGCTTTGAAAATACTGCTGGGATGCGGCCTCGATGTCCCCCTGGGTTTGAGGGGCGTCCATACAAATCTGGGGGC
ATGGGTAACTTTCCTGGAAGTGCATGGACGGTAATAGATAGGGGGTCTAACCAATGGCCAGCAGACGTGCGGGGGCCATTCTCAG
ATCAACGATGGGCCCCCACAGAGCATGAAACGCGACGGTTTTTGCGGGTATTACAGCTGAGCTCTCATCATACCCATAACTCCAC
TCATAACCCAAGGCTCATAAATCCATAACTCATAACATAAATTCATACTTTCCGGTCGTCCAGGGCACCACGTTCATCAACCAAG
GATTTGCAGACTAAATAAAAATGCTCCACGTTGTCGGTGTCCGTTGTATTGTGTTCTTTATTATACCTCCGTTA

>UL54

TGCAGACTAAATAAAAATGCTCCACGTTGTCGGTGTCCGTTGTATTGTGTTCTTTATTATACCTCCGTTAATTTTCGAGAGTCGG
GGAACATTCTAAAAATTTTCAACCGTGCTAATACTACAGTGTATTTACAAGGCCGGATTGCAACAGTGAACGCTTACATCATTGA
GCTCGCGGCGCCATCTGCTGACCAGTCACAGAGATGGCAATCTTCAGAAACGTAGGATGGCACCAATTCCAATACAATACCGCCA
TGTGTCGATAGGTGTATAGAACTGTCAAACAAGTCGCAAGAGAAAATTTCCCTACTGTATACTGCGGCTTAGAGCTGCGCACAAA
CCACTCTGCATTCCTCTTTGCGGCACACATTTGCGTGCTGCGCAGAACGAGTGGGATTTTTTTAGAACAGGTCCCAGGATAGTAC
ATGTCCCACAATGTTCTGGCCGGGTCTATTGCTTTATGATTCATGACTATGGCCTCTGATCGCGGATACACAATTCTTGAGAACC
GGTCGAAGAAGGTCAGTAATAAAGTTAAAGGACATTTTGCGCTACTCAGTGATAGCTCCTGAGATCTAGTGGTATCTCTTAGTTG
ACTGCCAATGCTAGAGAGATAACACGGCAGGATTGGCCCCAGATGCATGGCTAGAGATTGACATGCGCAGTAGATGTTAGAGAAG
ATAGGATCGTGGGGGTAAATCCTTTCATCTTCGAACTGATGCCAAAGCATCCATACAAGTGTCTCATCGCATGCAAAAAGTAGCT
CTTCAAATGAGCAGTTCGCCAAATATACAGCTCGTGAAATTTTTGCCAACCTGGCTATATCCGGACGCGATGTCCAGCGGCCTTT
CAGTGAAGCTGCGCGCCCACAAAACTGCTTCCACGAAGTGAATGCAGCATCGGCTGCAAGGTCAGATGATCCCGAAGACAAAAAT
GCTGGAAAGCAGATTCCTCTATCACGATCGATATCATCACAATCATCATCATCCACTGCCCGGTTTACCATGTCTAAAAGACATT
TCTGATTTTCTAATCTTAACTCTTCAGTAATGCACTTTCCGAGACCGCCAAATGCAGTTGCGGCCTTTTCAAAATATTGGGCCGG
TGTTACGTTTCGCAACTCCTTCGTTTCGGTCCGTGATGACGTTGGGCATCGGACAAAGTCTCTCCAAATCGGTCTTCGAAGTTCA
TCCCGATTTCTTTCCCAAGACCTGCGCGAATGCTTCAACGAAACAGTAAAGATAGGCGCCCTATATCGCTTTTCTGGTGTACCTG
CACGGCGCCTGGGTCTAGGGGGGATGCCTCTTCGGACTTGGATATGCGCACGGCCTCTCATAAACTTATTTCGAGGGCTACTGAC
CCGCCTTATGGAAGACCGTACGCTCATACCTGACTGGTTCTCGTAACGGGAACGAGGCCTCGCCATTCGAACAGCGACGGCGCCC
CCTGTCACTTGAATCACGGACTCGTTTGGACGTACCGCGCCCAGATGAACCGGGATCTGGAGCTACTTTCTGTAGAATGAGATCT
GCGGCGAATACAATGTGCTTCGTGGCGGGATTCTGACCTCAGCGTGAACGGCCATCCAGGGCGATCTTTTGTCCGGGCCGTTGCT
GATCTCGCTTCGCTATGAGTACTTGTATTGAAGATGCTCTGAACGTGTCTCACGCTATCTTTATTTCTTCCAGATTTCTCCGGTC
TCCATCGCAGTCAGTGGGTTGATGTACCGCGTGCACGTCAAAAAAAATGAAACCGCATACACAACGGTTGAGACTTCTACGGACT
CAGAACAGGTGTCAAGCTCGGAGCAGGTGCTGAAAGGTAAGCTGACAGTAATCTGGCACGCTGTTTGCGAGCTAATCCACTTGGC
TTTTGAATGGTCTGGGCCACTCCCAGTATACGTCATAACACATACACTGGAACCCACAAACTACAATTGCGATCCAGTAGTTGGT
GCGAAATATTCACGCAGACATAATAATCTGCGAGAACTTCTGCGGATCCGACATGTAACTTAATCCCGTAATGTAGTGCGGCATG
CCGTCTAAACCGCAAACATCCGCTTAGTAGAACACGCCCTAAAATCACCCACGAGTATACTTTGTACATTCTGACCGCCAGATGT
TACTCCTTTCAAACAATGATACTCAGCCGTTAGAACTAGGGCTGTCTTCAAATGGACCAAATTCAGACACAACACCGCACAACGT
GTTTTAACATTTTATTGCCGTTCAAGGCCCGAACAATTTGTTTTGTATCTTCTGTTCGTATTTAAATGCAATTATTACAATGCTC
GCAATCGCAGCCACGCACAATGCACGCAAGACTAAGCTCGAAGCAATATTGGCAAGGCATGAAGTCAAGATACTGGGAGGTTTCG
ATGCGGAGACTTTAGTCTTTGCCGGAGTCGTGTAACCCATC

>UL53
ACCTTATGCAAATGCAAAGAAAAGCTCGGGTTCAGAGTCTGCGTCCCAATCCCATCCCCATATGCTTTGTTTGGTCTAGAGACAG
TAAAAACTATCGCGAAGCTAGCTCAGCACGTAATTTTACTGCAGGAAGAAATCATTGAATGTTTAGACGATGTTATCAGAGACTA
TGACTTTATTGACACTGGAATCTACTCGCCAGGACGAAGTCTCCGGCTACCTTTCTTTGCCAAGGTTTCTGAATCCGGATTCATG
TCTGGGAGACTCCTGCCGTTTATAATTTTCCCCCCTGACTGTGCAGATAAAGTAGCTTTCGCGGCAGCGCACAAAGACCCTAATA
ATTTCCATTTTCATGCTTTCCGTCCAGACAATCCCACCCCAAATATTATCGTCACTCGAATCGCGTGTCCTCCTGACGTTATAGG
GGGAGTTTCAAAACCCTCTAGCAACTTGCCCCAAACGGCTGTGTCTCTTAGCCTCGCGGAAGCATTCAGTCGTGTCAATTTAGCA
CCGTGCGAAGACAGGACAGGAGCCGGGGTGGATACGATCGATGGAGTCTTTGTGCTCACGACATATGTTCTGCCGGCCATAACAA
ATTACATAAAAGAGCATTTTCCATCGTTGGCGCACGAATACAGCGACATTAGTTTTGGAGACGTGCGGGTGCTCAAAACGCGCAT
TACCGCATCACTATTACGAAATCGTCGAGGGTGCGGAGGGCGCGCAACTTTTACCTGCCTTAAACATTCGCATCGAAGTGCCGCT
GCCCAGACCGTCATTACTTCTGTAGCGGTTGCAATTAACTCACAGGGCAATCCCTACGCGGCATTTCAGACCAAATGCTTTGCGA
CTAAGTGCGGAGGCAACACGCTGCAAACTCAGTTTACAGTGCGTCTGACCAAAGAATGCTGCGACCAGAATGTCTAAAATGGGCA
GTAATTTTAACCGGAACGATACATATAGTCTTCTTGGTATGGTATGTATGCTCTAAAATTTCCACTGACTCAAAGGACGACTGCT
TGTACGTACTCGCAAATATTAAATGGCTCCTGCCCGGACAGCAAGAATTCAATCCTGTTCACCAACCCCCCGCTACATTTAATTC
TTCACTAGTTTATGTCCTCACAAAATATACACAGCGTCTAAACAACTACGAGAACCACCAGACGAGATGTGTTAAAGGAGCTTAT
TTTGAAAATGAGACTGTACTCATTTCTCGTCTGATCCCACTGGCCAAAGAACAGTATTCGTCATGGAAGTGGCAGACTGTCTCTT
TACATATGGTTTTCCCAGATCAGAGTTGCATTTCCACGGTTATTGTACATATGTTACTTGCTGACCCGTGCCAGAGGCGAATGTT
CGGCTCTGTCTGCCGCGAGAACGCATTGCGATTGGATGCATATCATCTAAACTACTGGACAGCGTTTACTTCGAGGCTGATATTA
CGGGTGCCATACACAAAGATGCAACGGTTTTTGAGGGAATTTGAACATGTCCGAGATTGCAAAAGCTTGAACTACGTAGCAGACC
CTCTAGGCTTTTGCATCTGTAATCCAGGGGTCTTAGTACTGAAAACACTCGAGATCGGTTTATATTTAGCATCGCTTATTATGTC
CACCATGACATTGCGGATTTGCTATGATCCGTGTGCATATATTTTACATGAACACGTAAAAATTAGTGCTTGGGTATATGTAATT
GTCTCAGCGGTTCTAGAACTCTTATCACTGATGGGTTACACGACTCCGGCAAAGACTAAAGTCTCCGCATCGAAACCTCCCAGTA
TCTTGACTTCATGCCTTGCCAATATTGCTTCGAGCTTAGTCTTGCGTGCATTGTGCGTGGCTGCGATTGCGAGCATTGTAATAAT
TGCATTTAAATACGAACAGAAGATACAAAACAAATTGTTCGGGCCTTGAACGGCAATAAAATGTTAAAACACGTTGTGCGGTGTT
GTGTCTGAATTTGGTCCATTTGAAGACAGCCCTAG

>UL52
GGACTGTTAGTCGACCCGATACATGTGCAGGATAATGATTGGATATCATTTTCGCTAACTAACATTCGAGATGAAGCGGCCAGTA
TTCACGCAGGTGACCGAGTGGCACAGCTAATAGTTACTGAGAAGAGAAGCACCTTTCTCGGAGAGCCAGATGCCCTTCGCTGGAA
AATAGTAAACTCAGTAGTAGACGAACGCAATACGTTGTCGGTTCGGGGTGACGGGGGGTTTGGATCTACCGGCAAATAATGTACT
GACATTGGAAAATTTAAACATACTTTATTAAAGCGCTGAGAAGTCACTTAACTCGCAGCAACCAGTTCTTTACGTGTCTTTGCAG
GTCTAGTACGCTCCATGGGTGGGTTTTCAACTGGCAGTGTCGGTACGGGATCAGGGAACAGTACCTTGTCAGGCAGTTTATCAAT
TACCGGCACGAGCTCGATTGGATCTTTCAGGTCTGAATTTGGTACCAGACCCATGCGCAGTAGTTCCAAAGAGCTATGGTCACGA
TCCATTCCGTCTACGTGCATTGCATGTGTAAGCACGGTCATGTCACCTAGCATCATCGTCCTGTCTGCGGTTTGAGCGGTTACTT
GTGCAAGTAGTCCATCCGTGTCTGCCCCAGGGGAGAGATATAGGTAGGCTAGTACAGAGTACAGAGAACGAAATACCGTCCCATT
GTCCACAATCGTTTGCCGCAAAACCGCCTCTGCTCCTCGCGGGCATCTCATATTCTTAGCAATCTGCATAGCCGCGGCCGATTTT
GCCAGAATGCGGCTAATTTGCCTGGCTTGAGCGAGTTTGTTAAGTTCATCCCTCGAGCCGATGACATCTTCGACGTCCAAGCAGT
TAGGGATAAAGATTTTAATAAGCTTCACAGCCGACTGTACTTTTCGATCATCTACGGGCAGGCACCGCCCAGTTGGAAGGGCTTC
GTATTTACTCTCTGAGTTAGACCGGCCGCAGAACACGCTCTTCAGTAGCTGCCACATTTCCTTGCTATCATGAACTTCGACTGTG
TCCCAGACGCCAAATGGGAGTATGAGGCAGATGACTGGATGCTCGTCGAAATGCTAAATGAGACAGAGACTTCCCAAGCTGGATA
TTGTGAACCCTGCGGCGCTCCCCCTAGTGAGGATGTTTCGTACGAAGAGATTCGCGTGTTGTATGCTACGGATGGGTATCCAGTA
GTATGCTCCCTGCACCTTCTACTCGGCCAAGCCCTGCAGCCTGAAATTTATGTTCTGGCCTTTAGTCATGTTGACGCTCTAGCGC
AAGATTTTTGCAAAGATTCCCCTCTCAAAATCACGTTCTTCCTTCCTGCTAATACCGGAGCGGAGAAGTACAGGCAGAGGTCTAG
ACCAGTGTTTGTGTGTCGTTTTGTCGGTCAAGGGGCCGCAACTATGAGAGAAGCCTTAGCCGAAGGGTTTCCAATCGATGCAAAG

TTATTACTGGAAACCATGGATTTGGAGGAAACATTTGCATTGCACACGGACATTATTGAGGCACTCGGAGTAATTTTCCGAGCAG
AAAATACAACCAACAAAATTCCTTTTTCTGCCATACATGCGACTCTACCCAGAGGAAGACCTGCCCTGGAGGGCCTGATGCTTCA
GCATGAGTCAAGGGTGATTGCCGCATACAGGAGGCTTTACAACTCGAACTTTTCGAACCCCTTTTGGTTCGTATCAAAATTCGGC
CCGTCAGAAATTACCCTTGTTGCAGCCATGAGATATTACTTGTTAGAAGCAACCCCTGACGATGGTATTGCTGGGACGTTTGACC
TTCAAGCAGTAAAGGATCTATGCATTACATTTCAAACGGAGAATAGCAACAACAGGAGTAACCTGACTATAGCGCGCCTAACGTC
CTTCTCTGCGTTTTCTAAATTCTGGTGCAGCAGCTACTATGCCACTCGCGAATCTGCAAAGCTATTCAGAGATCATTTGAGATTA
CGTACTAGCGCCGACAGGGGCTATTCAGAACAATTGGAAGCACTTATCTGCAAATACAGACATGCCCTGCGGCTTCCAGCTCGTG
ACTTCATCTCATTTGTATATTTAGCTTACAGTGAAGGATATGGACGAAACGTGATTAGAGAACACTTGCAATGTCTTTCTGAAAA
TCTCACCAATCCAGTTAAGAGTTCTGCTCGTGCTCAAAGTGATTTTTTTAAAAGCTTAAGGGAATGTTGTAATTTTGCTTCGTAC
TTCAGACAGCATGTGACTATGCAAGTCATGAAATTGCCGGCTGTAGTAGCAGGGAAGTTTGCAGCTATAAAATCTTACGGTGAAA
TACAGGGGCGCAATTTTCTTGGTTTTTGTCGTTCCGCATTTGATGTATCAACGCGTATAGACCGTCTATATTCCGATACTCTCTG
TCGCTACGGGGCACACTTGGCTGACATGCTGGTCAGAAATGTGAGCACTGTATCTAGCAATGAACAGGAAGAAGAGGAAGAAGAG
GAAGAAGAGGAAGGAGATGATGCTTCCGCTCCACACGCCCTTGCACAAAAAAGAGAGCGCGACGGGCAATACTGTAAGACAGGAG
ATATCATTTTTGCTCCCTTAAGTAGGTTACTTTCTCTCGTTTCATCTGATCAAGTTATTTGCGGAAAGCGACCAGTGACACTCTT
ATTCCATTCTCTGAAACCCACAAGCATTCCTCCACCAACAGCTAGGAACGGACCATGTCCCGTTTACAGAGTAACATTGGCTGGA
CACGGACAGGCTTTTGCAGCCATTGCCAGCGATGAGTGGGACTCGATTATAACCGCTGATGTAAGATCTGACACCAGCATACACG
ACTCCTGCTTACTAATGCGCAGTGATGCCCCCCTCGCAGATCAGGCCATGTTGCCAGTAATAGACTTTGCTTGGACTAGTGCAAT
GAACGAAACCGCATGGGGAACTAGTTCGAAAAATAAAATTCGGGCATCTTGCAGCGGTCTGTCACAAATGTACCTGAATAGGAAT
GAAGTCCTTAACGAATCTCTAGCAATTCAATCCCTGATACTGGATATCGACATACCCCTCAAGCTAGATCACGGCCCCATAACGA
TGTTGACCTTACATAAAGCGATGCGGGCCGTGCGCGTGGCATTGATCCAATTAATAGCCTTATTGTTTCCGGAGACATCAATCAG
TCATGACACCTATCCCGTGTACTTTTACAAAAGTCACTGCTCGTCCCCGGAGAGCCATGACGCACAATATTCCTCAGTGGATATT
GATTTTTTGGAGAATACATGTTCTGATGACATAGAAGACGAAGAGAACATGTACCTCTGCGATAGCGGCTGGGTAGAAGAAATGT
CATGGTGTGATGAGCATGCGAGTGAACATGGAGCCATAGTCTCTGTTCAAGACTTACAGGATGCGTTAAATGTATTTCGGGATCG
TGCCAATTCAACCTTATGCAAATGCAAAGAAAAGCTCGGGTTCAGAGTCTGCGTCCCAATCCCATCCCCATATGCTTTGTTTGGT
CTAGAGACAGTAAAAACTATCGCGAAGCTAGCTCAGCACGTAATTTTACTGCAGGAAGAAATCATTGAATGTTTAGACGATGTTA
TCAGAGACTATGACTTTATTGACACTGGAATCTACTCGCCAGGACGAAGTCTCCGGCTACCTTTCTTTGCCAAGGTTTCTGAATC
CGGATTCATGTCTGGGAGACTCCTGCCGTTTATAATTTTCCCCCCTGACTGTGCAGATAAAGTAGCTTTCGCGGCAGCGCACAAA
GACCCTAATAATTTCCATTTTCATGCTTTCCGTCCAGACAATCCCACCCCAAATATTATCGTCACTCGAATCGCGTGTCCTCCTG
ACGTTATAGGGGGAGTTTCAAAACCCTCTAGCAACTTGCCCCAAACGGCTGTGTCTCTTAGCCTCGCGGAAGCATTCAGTCGTGT
CAATTTAGCACCGTGCGAAGACAGGACAGGAGCCGGGGTGGATACGATCGATGGAGTCTTTGTGCTCACGACATATGTTCTGCCG
GCCATAACAAATTACATAAAAGAGCATTTTCCATCGTTGGCGCACGAATACAGCGACATTAGTTTTGGAGACGTGCGGGTGCTCA
AAACGCGCATTACCGCATCACTATTACGAAATCGTCGAGGGTGCGGAGGGCGCGCAACTTTTACCTGCCTTAAACATTCGCATCG
AAGTGCCGCTGCCCAGACCGTCATTACTTCTGTAGCGGTTGCAATTAACTCACAGGGCAATCCCTACGCGGCATTTCAGACCAAA
TGCTTTGCGACTAAGTGCGGAGGCAACACGCTGCAAACTCAGTTTACAGTGCGTCTGACCAAAGAATGCTGCGACCAGAATGTCT
AAAATGGGCAGTAATTTTAACCGGAACGATACATATAGTCTTCTTGGTATGGTATGTATGCTCTAAAATTTCCACTGACTCAAAG
GACGACTGCTTGTACGTACTCGCAAATATTAAATGGCTCCTGCCCGGACAGCAAGAATTCAATCCTGTTCACCAACCCCCCGCTA
CATTTAATTCTTCACTAGTTTATGTCCTCACAAAATATACACAGCGTCTAAACAACTACGAGAACCACCAGACGAGATGTGTTAA
AGGAGCTTATTTTGAAAATGAGACTGTACTCATTTCTCGTCTGATCCCACTGGCCAAAGAACAGTATTCGTCATGGAAGTGGCAG
ACTGTCTCTTTACATATGGTTTTCCCAGATCAGAGTTGCATTTCCACGGTTATTGTACATATGTTACTTGCTGACCCGTGCCAGA
GGCGAATGTTCGGCTCTGTCTGCCGCGAGAACGCATTGCGATTGGATGCATATCATCTAAACTACTGGACAGCGTTTACTTCGAG
GCTGATATTACGGGTGCCATACACAAAGATGCAACGGTTTTTGAGGGAATTTGAACATGTCCGAGATTGCAAAAGCTTGAACTAC
GTAGCAGACCCTCTAGGCTTTTGCATCTGTAATCCAGGGGTCTTAGTACTGAAAACACTCGAGATCGGTTTATATTTAGCATCGC
TTATTATGTCCACCATGACATTGCGGATTTGCTATGATCCGTGTGCATATATTTTACATGAACACGTAAAAATTAGTGCTTGGGT
ATATGTAATTGTCTCAGCGGTTCTAGAACTCTTATCACTGATGGGTTACACGACTCCGGCAAAGACTAAAGTCTCCGCATCGAAA
CCTCCCAGTATCTTGACTTCATGCCTTGCCAATATTGCTTCGAGCTTAGTCTTGCGTGCATTGTGCGTGGCTGCGATTGCGAGCA
TTGTAATAATTGCATTTAAATACGAACAGAAGATACAAAACAAATTGTTCGGGCCTTGAACGGCAATAAAATGTTAAAACACGTT
GTGCGGTGTTGTGTCTGAATTTGGTCCATTTGAAGACAGCCCTAG

>UL51

AAGTGCTTCCAATTGTTCTGAATAGCCCCTGTCGGCGCTAGTACGTAATCTCAAATGATCTCTGAATAGCTTTGCAGATTCGCGA
GTGGCATAGTAGCTGCTGCACCAGAATTTAGAAAACGCAGAGAAGGACGTTAGGCGCGCTATAGTCAGGTTACTCCTGTTGTTGC
TATTCTCCGTTTGAAATGTAATGCATAGATCCTTTACTGCTTGAAGGTCAAACGTCCCAGCAATACCATCGTCAGGGGTTGCTTC
TAACAAGTAATATCTCATGGCTGCAACAAGGGTAATTTCTGACGGGCCGAATTTTGATACGAACCAAAAGGGGTTCGAAAAGTTC
GAGTTGTAAAGCCTCCTGTATGCGGCAATCACCCTTGACTCATGCTGAAGCATCAGGCCCTCCAGGGCAGGTCTTCCTCTGGGTA
GAGTCGCATGTATGGCAGAAAAAGGAATTTTGTTGGTTGTATTTTCTGCTCGGAAAATTACTCCGAGTGCCTCAATAATGTCCGT
GTGCAATGCAAATGTTTCCTCCAAATCCATGGTTTCCAGTAATAACTTTGCATCGATTGGAAACCCTTCGGCTAAGGCTTCTCTC
ATAGTTGCGGCCCCTTGACCGACAAAACGACACACAAACACTGGTCTAGACCTCTGCCTGTACTTCTCCGCTCCGGTATTAGCAG
GAAGGAAGAACGTGATTTTGAGAGGGGAATCTTTGCAAAAATCTTGCGCTAGAGCGTCAACATGACTAAAGGCCAGAACATAAAT
TTCAGGCTGCAGGGCTTGGCCGAGTAGAAGGTGCAGGGAGCATACTACTGGATACCCATCCGTAGCATACAACACGCGAATCTCT
TCGTACGAAACATCCTCACTAGGGGGAGCGCCGCAGGGTTCACAATATCCAGCTTGGGAAGTCTCTGTCTCATTTAGCATTTCGA
CGAGCATCCAGTCATCTGCCTCATACTCCCATTTGGCGTCTGGGACACAGTCGAAGTTCATGATAGCAAGGAAATGTGGCAGCTA
CTGAAGAGCGTGTTCTGCGGCCGGTCTAACTCAGAGAGTAAATACGAAGCCCTTCCAACTGGGCGGTGCCTGCCCGTAGATGATC
GAAAAGTACAGTCGGCTGTGAAGCTTATTAAAATCTTTATCCCTAACTGCTTGGACGTCGAAGATGTCATCGGCTCGAGGGATGA
ACTTAACAAACTCGCTCAAGCCAGGCAAATTAGCCGCATTCTGGCAAAATCGGCCGCGGCTATGCAGATTGCTAAGAATATGAGA
TGCCCGCGAGGAGCAGAGGCGGTTTTGCGGCAAACGATTGTGGACAATGGGACGGTATTTCGTTCTCTGTACTCTGTACTAGCCT
ACCTATATCTCTCCCCTGGGGCAGACACGGATGGACTACTTGCACAAGTAACCGCTCAAACCGCAGACAGGACGATGATGCTAGG
TGACATGACCGTGCTTACACATGCAATGCACGTAGACGGAATGGATCGTGACCATAGCTCTTTGGAACTACTGCGCATGGGTCTG
GTACCAAATTCAGACCTGAAAGATCCAATCGAGCTCGTGCCGGTAATTGATAAACTGCCTGACAAGGTACTGTTCCCTGATCCCG
TACCGACACTGCCAGTTGAAAACCCACCCATGGAGCGTACTAGACCTGCAAAGACACGTAAAGAACTGGTTGCTGCGAGTTAAGT
GACTTCTCAGCGCTTTAATAAAGTATGTTTAAATTTTCCAATGTCAGTACATTATTTGCCGGTAGATCCAAA

>UL50

GCTGGAATAGTAGAAGTATGCCATGGGTCTTCATCCCCTGAGGGTGTAGCACTAGGACCAAATCCCTTAGAGATAGTTAAACGTC
GCATCAGGCCAGCGCAATGATTATCAGGATTCCAGCCCCTTTTACTTTGAGGGGGCTCCTCCTTTGGAGTGGTTTCTTTGCGAAA

ACTTTCCATCGGGGGGCTTCGGCGGCCTTGAGGCGGTGCCGTATGATCTCTCGAATAGTGAGCGCGCGATGGATGGCCGTAATCC
CCTTCTGAAGAATCAGAATCGCTTCCTCTATCCAAATCTAGATATGGAGATGTTCTTCGCCCAGACCAGGGGTGTTCGAATTGGG
GTGGTTTGGGGCGGCGGGTGGTGGTATCTCGCCTCCGTCGAGTATCGTCGTGATACTTCTTCGCCTCCTCCGAGAGATCTTTGTA
GTAAGACATAGTGTTCGAGCTAATATGGTCCCACCACCGATGTTTTCGTCTTTACAGCAGATACTGCGCTACCATCGAGAACTAA
TGACACTGCTTTTAAAGCTGTTGAAACAACCGTTATATAGCACGGCCAGCAGGCCAATCAACGCCACTGCCAGGCTAATGAAAAA
GAAAACGGAAGCAGCTGTTCCGTTTACTAAGGAAAATCCGCGCGACGAGCAGGCACTGTGGTAAAAGTGTTCCATGTTCGCGCTA
TACATCGAATTTTCCGCTCCAAGCATATTTCCAGCGCCAAGACCATAATACCCATGCTCCATTCCGTGTCCTGGTGTCACACCCG
GTTGCGCGAATACTAGTGCCAAATGCAGCAAATACACCACTGCTCCCAACAGCATAAAACTCCTTTCCCACGGCAGCCTCATTGC
TCTTTTGCGAGAGCTCGCCCTGCAGTCTCTTTCACGGGTCCTCTAAGTCGCGCAGTGCGGTACTATTTGTATCCTACCGGTCTAC
TGCCGATTTTAACTACGCGGATTTCCTGAAAGGGTCGGTGTTATTTTGGCGTCTAAGCAACTTGCATGGAGATGGAAATGACAAC
GGAAAAAAATGCATTAATAGAGATTGGATCTGGGTGGGACATTTTTGGCCTCGATTCCTTGCAATGTGTCATTACAAACGTTGAA
GCAACAATAGCGAAAACACGCACCGACTCTCCCGTCCTAAGAATTGATTCTGCGATTAGGACTTCTTTGCCAGTCGGGTATGGAA
TCGTTATCAGTGATGTGGAAGGCCATGCTGCATGTTATCAAATTATTCCAGGATTGGTAGATGCAGACTATACAGGATTATTGGG
AATACTGGTAGTACTTGTAAACGATGGCGGGAATGTTTTGACTGGCACAGAGACAAGAGACGGAAATGTTATATTTCCACCAGGC
ACTGTAAGGGCTCGGCTAAATGTTATAAAGTTGGCTAACGATTCCATGCTTAGCAAGTGTGGGACGGGATACGCGCGCCTACTCT
TAGAAACCGCAGATTCTTTCCAAGGCGAAGATGATTATTTGGGTGAAGGTTTCGAGAAATGTATAAGTTCCCTTTTAAGTATACC
TCCAGATATATTAAGGGCGAAAATAACTATTCCGGGCAACATTGGTTGTATGGGATGTACAACGTTTTACAGAAGAACATTCGAA
GAATTGCAGGAGCAAGATGCAGCCTGCAATGAACGCATCGTTATACTGAATGGAGCGACGCCAGAAAATATCAATCGTTTTTGTG
GCGCATCGAACATCAAGTTTGTAGCCCAGGTGCGACACCATATCATCGTCGGAGCTACACGATCGCTCGTTAAAGATCAGACAGT
CGAGCTGCGATCAGATCCAGATTGCCCTGCCGCACTGTTTCCGTTAACGAAATCTTTGCCCCCAAACGTCCAGGAGATGGAGGC
TATGATATTAGAGCTACTAAAGATGTAACAATTCTTCCAAAGTCGTCTACCCGTATTACCCTTCCCCAAAAACTAGCATGCGGAC
CTTTCTGGCGCGCATTCATACTAGGGAGATCCTCCATGAATCTGAAAGGACTGTTAGTCGACCCGATACATGTGCAGGATAATGA
TTGGATATCATTTTCGCTAACTAACATTCGAGATGAAGCGGCCAGTATTCACGCAGGTGACCGAGTGGCACAGCTAATAGTTACT
GAGAAGAGAAGCACCTTTCTCGGAGAGCCAGATGCCCTTCGCTGGAAAATAGTAAACTCAGTAGTAGACGAACGCAATACGTTGT
CGGTTCGGGGTGACGGGGGGTTTGGATCTACCGGCAAATAATGTACTGACATTGGAAAATTTAAACATACTTTATTAAAGCGCTG
AGAAGTCACTTAACTCGCAGCAACCAGTTCTTTACGT

>UL49.5

TTTGGGGGCAAAGATTTCGTTAAACGGAAACAGTGCGGCAGGGCAATCTGGATCTGATCGCAGCTCGACTGTCTGATCTTTAACG
AGCGATCGTGTAGCTCCGACGATGATATGGTGTCGCACCTGGGCTACAAACTTGATGTTCGATGCGCCACAAAAACGATTGATAT
TTTCTGGCGTCGCTCCATTCAGTATAACGATGCGTTCATTGCAGGCTGCATCTTGCTCCTGCAATTCTTCGAATGTTCTTCTGTA
AAACGTTGTACATCCCATACAACCAATGTTGCCCGGAATAGTTATTTTCGCCCTTAATATATCTGGAGGTATACTTAAAAGGGAA
CTTATACATTTCTCGAAACCTTCACCCAAATAATCATCTTCGCCTTGGAAAGAATCTGCGGTTTCTAAGAGTAGGCGCGCGTATC
CCGTCCCACACTTGCTAAGCATGGAATCGTTAGCCAACTTTATAACATTTAGCCGAGCCCTTACAGTGCCTGGTGGAAATATAAC
ATTTCCGTCTCTTGTCTCTGTGCCAGTCAAAACATTCCCGCCATCGTTTACAAGTACTACCAGTATTCCCAATAATCCTGTATAG
TCTGCATCTACCAATCCTGGAATAATTTGATAACATGCAGCATGGCCTTCCACATCACTGATAACGATTCCATACCCGACTGGCA
AAGAAGTCCTAATCGCAGAATCAATTCTTAGGACGGGAGAGTCGGTGCGTGTTTTCGCTATTGTTGCTTCAACGTTTGTAATGAC
ACATTGCAAGGAATCGAGGCCAAAAATGTCCCACCCAGATCCAATCTCTATTAATGCATTTTTTTCCGTTGTCATTTCCATCTCC
ATGCAAGTTGCTTAGACGCCAAAATAACACCGACCCTTTCAGGAAATCCGCGTAGTTAAAATCGGCAGTAGACCGGTAGGATACA
AATAGTACCGCACTGCGCGACTTAGAGGACCCGTGAAAGAGACTGCAGGGCGAGCTCTCGCAAAAGAGCAATGAGGCTGCCGTGG
GAAAGGAGTTTTATGCTGTTGGGAGCAGTGGTGTATTTGCTGCATTTGGCACTAGTATTCGCGCAACCGGGTGTGACACCAGGAC
ACGGAATGGAGCATGGGTATTATGGTCTTGGCGCTGGAAATATGCTTGGAGCGGAAAATTCGATGTATAGCGCGAACATGGAACA
CTTTTACCACAGTGCCTGCTCGTCGCGCGGATTTTCCTTAGTAAACGGAACAGCTGCTTCCGTTTTCTTTTTCATTAGCCTGGCA
GTGGCGTTGATTGGCCTGCTGGCCGTGCTATATAACGGTTGTTTCAACAGCTTTAAAAGCAGTGTCATTAGTTCTCGATGGTAGC
GCAGTATCTGCTGTAAAGACGAAAACATCGGTGGTGGGACCATATTAGCTCGAACACTATGTCTTACTACAAAGATCTCTCGGAG
GAGGCGAAGAAGTATCACGACGATACTCGACGGAGGCGAGATACCACCACCCGCCGCCCCAAACCACCCCAATTCGAACACCCCT
GGTCTGGGCGAAGAACATCTCCATATCTAGATTTGGATAGAGGAAGCGATTCTGATTCTTCAGAAGGGGATTACGGCCATCCATC
GCGCGCTCACTATTCGAGAGATCATACGGCACCGCCTCAAGGCCGCCGAAGCCCCCCGATGGAAAGTTTTCGCAAAGAAACCACT
CCAAAGGAGGAGCCCCCTCAAAGTAAAAGGGGCTGGAATCCTGATAATCATTGCGCTGGCCTGATGCGACGTTTAACTATCTCTA
AGGGATTTGGTCCTAGTGCTACACCCTCAGGGGATGAAGACCCATGGCATACTTCTACTATTCCAGCAAATCGTTCGGCCTTCGT
GCAAGCCGTCTCTGTGACGGCAATGGCCCAGGCAGAATTGGCAGCGAGAGAGGTGTGGGACGTAACAAAACCACGCACGAATAGG
GAGTTGAGGGATATGGTGAGAGAGCTAGAAATCACTATAATTATCAATCCGGGTGAATCTTTATGGTCTGTTGCCACTTCGGTGG
CAAGAGCAATCAAAGAAGGAACCCCAATAACCCACGAGCTATTGCAGAAAAGACCATCAAAGCCTCCAACCCGCCGCAAAACTGA
AGATGGAGCCAGAAAGAGTTCTTCAAGGCCTTCTCAACCCAAGCCTGAGCATTTCCCCCCTCCAAGAAGAAAGACATCCGAACGA
AAGTACTAGTTTCGAGAGTCTTTGCATTACTGTGTTCCTCGCCATATTTGTAACAATAAAGTATACTTATGATTGGTAACATGCG
TGTATTGTCATGTGTTCATAGACTA

>UL49

TAGGCGCGCGTATCCCGTCCCACACTTGCTAAGCATGGAATCGTTAGCCAACTTTATAACATTTAGCCGAGCCCTTACAGTGCCT
GGTGGAAATATAACATTTCCGTCTCTTGTCTCTGTGCCAGTCAAAACATTCCCGCCATCGTTTACAAGTACTACCAGTATTCCCA
ATAATCCTGTATAGTCTGCATCTACCAATCCTGGAATAATTTGATAACATGCAGCATGGCCTTCCACATCACTGATAACGATTCC
ATACCCGACTGGCAAAGAAGTCCTAATCGCAGAATCAATTCTTAGGACGGGAGAGTCGGTGCGTGTTTTCGCTATTGTTGCTTCA
ACGTTTGTAATGACACATTGCAAGGAATCGAGGCCAAAAATGTCCCACCCAGATCCAATCTCTATTAATGCATTTTTTTCCGTTG
TCATTTCCATCTCCATGCAAGTTGCTTAGACGCCAAAATAACACCGACCCTTTCAGGAAATCCGCGTAGTTAAAATCGGCAGTAG
ACCGGTAGGATACAAATAGTACCGCACTGCGCGACTTAGAGGACCCGTGAAAGAGACTGCAGGGCGAGCTCTCGCAAAAGAGCAA
TGAGGCTGCCGTGGGAAAGGAGTTTTATGCTGTTGGGAGCAGTGGTGTATTTGCTGCATTTGGCACTAGTATTCGCGCAACCGGG
TGTGACACCAGGACACGGAATGGAGCATGGGTATTATGGTCTTGGCGCTGGAAATATGCTTGGAGCGGAAAATTCGATGTATAGC
GCGAACATGGAACACTTTTACCACAGTGCCTGCTCGTCGCGCGGATTTTCCTTAGTAAACGGAACAGCTGCTTCCGTTTTCTTTT
TCATTAGCCTGGCAGTGGCGTTGATTGGCCTGCTGGCCGTGCTATATAACGGTTGTTTCAACAGCTTTAAAAGCAGTGTCATTAG
TTCTCGATGGTAGCGCAGTATCTGCTGTAAAGACGAAAACATCGGTGGTGGGACCATATTAGCTCGAACACTATGTCTTACTACA
AAGATCTCTCGGAGGAGGCGAAGAAGTATCACGACGATACTCGACGGAGGCGAGATACCACCACCCGCCGCCCCAAACCACCCCA
ATTCGAACACCCCTGGTCTGGGCGAAGAACATCTCCATATCTAGATTTGGATAGAGGAAGCGATTCTGATTCTTCAGAAGGGGAT
TACGGCCATCCATCGCGCGCTCACTATTCGAGAGATCATACGGCACCGCCTCAAGGCCGCCGAAGCCCCCCGATGGAAAGTTTTC
GCAAAGAAACCACTCCAAAGGAGGAGCCCCCTCAAAGTAAAAGGGGCTGGAATCCTGATAATCATTGCGCTGGCCTGATGCGACG

TTTAACTATCTCTAAGGGATTTGGTCCTAGTGCTACACCCTCAGGGGATGAAGACCCATGGCATACTTCTACTATTCCAGCAAAT
CGTTCGGCCTTCGTGCAAGCCGTCTCTGTGACGGCAATGGCCCAGGCAGAATTGGCAGCGAGAGAGGTGTGGGACGTAACAAAAC
CACGCACGAATAGGGAGTTGAGGGATATGGTGAGAGAGCTAGAAATCACTATAATTATCAATCCGGGTGAATCTTTATGGTCTGT
TGCCACTTCGGTGGCAAGAGCAATCAAAGAAGGAACCCCAATAACCCACGAGCTATTGCAGAAAAGACCATCAAAGCCTCCAACC
CGCCGCAAAACTGAAGATGGAGCCAGAAAGAGTTCTTCAAGGCCTTCTCAACCCAAGCCTGAGCATTTCCCCCCTCCAAGAAGAA
AGACATCCGAACGAAAGTACTAGTTTCGAGAGTCTTTGCATTACTGTGTTCCTCGCCATATTTGTAACAATAAAGTATACTTATG
ATTGGTAACATGCGTGTATTGTCATGTGTTCATAGACTA

>UL48

GCAGTATCTGCTGTAAAGACGAAAACATCGGTGGTGGGACCATATTAGCTCGAACACTATGTCTTACTACAAAGATCTCTCGGAG
GAGGCGAAGAAGTATCACGACGATACTCGACGGAGGCGAGATACCACCACCCGCCGCCCCAAACCACCCCAATTCGAACACCCCT
GGTCTGGGCGAAGAACATCTCCATATCTAGATTTGGATAGAGGAAGCGATTCTGATTCTTCAGAAGGGGATTACGGCCATCCATC
GCGCGCTCACTATTCGAGAGATCATACGGCACCGCCTCAAGGCCGCCGAAGCCCCCCGATGGAAAGTTTTCGCAAAGAAACCACT
CCAAAGGAGGAGCCCCCTCAAAGTAAAAGGGGCTGGAATCCTGATAATCATTGCGCTGGCCTGATGCGACGTTTAACTATCTCTA
AGGGATTTGGTCCTAGTGCTACACCCTCAGGGGATGAAGACCCATGGCATACTTCTACTATTCCAGCAAATCGTTCGGCCTTCGT
GCAAGCCGTCTCTGTGACGGCAATGGCCCAGGCAGAATTGGCAGCGAGAGAGGTGTGGGACGTAACAAAACCACGCACGAATAGG
GAGTTGAGGGATATGGTGAGAGAGCTAGAAATCACTATAATTATCAATCCGGGTGAATCTTTATGGTCTGTTGCCACTTCGGTGG
CAAGAGCAATCAAAGAAGGAACCCCAATAACCCACGAGCTATTGCAGAAAAGACCATCAAAGCCTCCAACCCGCCGCAAAACTGA
AGATGGAGCCAGAAAGAGTTCTTCAAGGCCTTCTCAACCCAAGCCTGAGCATTTCCCCCCTCCAAGAAGAAAGACATCCGAACGA
AAGTACTAGTTTCGAGAGTCTTTGCATTACTGTGTTCCTCGCCATATTTGTAACAATAAAGTATACTTATGATTGGTAACATGCG
TGTATTGTCATGTGTTCATAGACTATAAAAGCGCCGTGATTTACATTTCTGCGCACACTGCGCAGAACCATGGAAGAAGAATCTT
CCACTGGAGCCTTTGCACTGTATGAAGTCGAAGACTTCATCGAGGAAGTAGAATCTGGGGCTGGGGGGAGGTGCAGCGATTTTGA
GGATGATGCGACTGACCTAATTGAATTAGAAGATCTCTATTTTGATTGGTTACGGAATCCAGGTCCCCCCGCGCTGGTCAATGCG
CCAAAGTACGTTAGGCCTTCAAAATGTTCAGCATCGTTTCTCGCAAAAAACGCGCTAACGGAGCTGCAATTTGCCGACTGGGACA
AAATACAGCACGTGTTGAAGACGATAAACATTGATATTTTCTCGTGTGTACCAGTGGGAGAAAATGGTCTATGCAATATGGATTG
GTTTTCCCTCGATACAGAAAAAATATGGACGTTAATGGATTCAGGAACACCTAATGTTAGCCCAAATTTGCTGATAACACCTGGC
CCGCCACCCGAATGCATTAAAACTCCAGAAGAAGCTCCAGCTTATTTTGAAGCAATCTCGAATTACTTTGAATCCGTGTTAGCCG
CCAGAGAAGCCGAATATGCAAAATTACTGGCGGAGTACGCGAAGTCGCTCGCTAAATACATCAGACATAAATCACAAAAAGTAGG
GCGCGGTCTGCGAGCCCTCGAACTGGACACAGAGGAGGCCAAAAAACGATTTACCGTTCTAATGAAATCCGAATACTACAAAGGG
ATCGAAAAACTTCTAATTATATTTCTGGTTCATTTAGTGCTTAACGCGTGTAGGGATGTTGCTCGGATAATATGGGTCCAACAGA
GGGACTCCTCGTCTGTATTGCAGAAGGTAACGTATAGGTGGGAAAGCCAACAGCTACACTGTATTTTTCAACCATTACTTGTTTC
TCCGGGTCCAGTTTTTTTGGACAACGACCCCTTGCCTTTACAGACTTTGCAGCATGTAAACTACGTACGATATCTTTTAGGAGTT
CCGGTGCTAAGATTCTCCATGATAGAAGAATCCCCCAGTGCACAATATATCCCTTATGACAAAACTGCACCTTCTGCATATGATG
TTTTAGCGACTATTGCACGTACTTGCCTAACTTCCAGCGGCTGTTCGCGTGCTTCTCACGACCATACCTACGCACGCGATAACGA
AAAACCAAACTACTCCTCCTCTCGAATCGTCTTATTAGCTGATTCCGCCCTTGATACACCTATGCCCTAAACATGACCGCCCAGG
CTATAAAACAGATGGACTGGCCAGACAACATCATTGCAGCGATGGCTGAAGGCGATGGGTTTGAAGAGCGTCTCCGCGCTAAAGC
GCGGTGCCATCAGATGAAGAGGACTCCGATTGGACTCCTCCAAGTAATCCTTGACGGGGCAAGGAAAATTGATCCGAAACGGATT
TTAGATAATTCTTCCCCGAGCGACTTAGCTGTGTTAGCTGAGCGAGAGCTACCATCCTCCCTGCTAGGACCGGCGCGCGAAGAAT
ACTTAAAATTTGGGGATAAAAAACTGCCAGACAACTTTATCTCACGCACTGTCTCCAATGACACCACGGATGAGTATAGAAGAAA
TCTTGACGAAGCGTTTAAACTCTTTATTGCACGCAGAGGGGTCACTGAAAAATGCGTTGCCCGAATGCTCAGACATCTCTATGCC
AAGCGGTTGTCCGTGGTTTTGGATCTCGATGTCCCATTTGGTGAAAAAATAAACAAAGCGCTTTCCGCCGATGATCAAGATTTTG
TACTGAAACTGTACTACTGGGTTTTCCCGGACAAGGTCTCTGCGCCTAGACTTAAGGAACCAGAAATGCAAGCTTACCTGGCTTC
AGTAACCTATTTGTCGTGCACCTTTAAGTTCATTGAAGGATACGCTTATTATATGCGCCCCCCTGGGCCAGAACTCCATTCAACA
GAAACGACCGCTCGTCTAGCAGCCCTCCTTGTATATGTGAGAAGCATGTACTGCAAACTGGCGTCCTTATTTGACGCAATCGATG
CCGCTGTGCTGCGCTCTAAATTTAATAGGGGAGCTTCGGCTCAAGCTCTGACTGGAATTTATGGATACACGCGAAACGGGAAGGG
AACTATAATGTCGCGGTATTCTTCGGAGGCCATGAGTGATGAAATCATGTCTGTGTGTCGCACTGGTTCGGACACTTTGATAGAA
TCTTTACGCAAAGCTCTGGGGAGATTGACCTACCTTATTGCTAGGTGGGAGCTCATTTCGGTAGAGAAGAAAAGGGACGCAAAAG
AGGCCGTGACGGCAGCGTTAGCATTAAGTGGACTAGTTAGCGGACATTTGGGATACTTGCTAAACCTAATCTCAATGGCATACGT
TTTTTGGCTCCCTGGGCATTATTTAAATATGCGGTTAATATTGGCGCTGAGTTCCAGGGAACTCCTAAGCCATCTTTTGGCTACA
AATCCTTCTTCATTGAGGGAAGTAGCCAGCTGCCCGAAAATACTTTTCTACCAATCTGTAGAAGAATGGAGAGGACAGGAAAGAG
GAACACAGATTTGGTTGGATTATGTTACTGTAAGGGCATGGTTGGATTATATCGACACTGGGATCCCTACTATTCGCTCCGAGGA
AACAAATGCTTCCATAAATGAAATTTCCGAGTATCTGGAACAGCGACGCTCGGCCAGTCCTACTGAATCCTTTTTATCTTCAGAT
GGACCACCAAACTCAAGCGCAGCATCCCTGCCCCCTACACAAGGCTCAATGCAAACACAAAACAGGAGACGCGAACGACGCAGTA
ACAGTGTCGTTATCCGAAAGCCCTCCCGCTCTGCGCATTCAAAGAGTAATACCATTCCCGTGTTTCAGGAAACGACCATTTCTTA
TTCGGCACCGAATTCCCCGGCGGATCATTACGTAGTAATGAATTAAGCCTGTATAACTGTACAGTAAATGCATCGCCAATTTTTA
TATTTACGCTCTGAATAAATAATGACATTACATTGACGTAAGATGTACATAATTCGTATTTATTTACTTT

>UL46

GAGCTGCAATTTGCCGACTGGGACAAAATACAGCACGTGTTGAAGACGATAAACATTGATATTTTCTCGTGTGTACCAGTGGGAG
AAAATGGTCTATGCAATATGGATTGGTTTTCCCTCGATACAGAAAAAATATGGACGTTAATGGATTCAGGAACACCTAATGTTAG
CCCAAATTTGCTGATAACACCTGGCCCGCCACCCGAATGCATTAAAACTCCAGAAGAAGCTCCAGCTTATTTTGAAGCAATCTCG
AATTACTTTGAATCCGTGTTAGCCGCCAGAGAAGCCGAATATGCAAAATTACTGGCGGAGTACGCGAAGTCGCTCGCTAAATACA
TCAGACATAAATCACAAAAAGTAGGGCGCGGTCTGCGAGCCCTCGAACTGGACACAGAGGAGGCCAAAAAACGATTTACCGTTCT
AATGAAATCCGAATACTACAAAGGGATCGAAAAACTTCTAATTATATTTCTGGTTCATTTAGTGCTTAACGCGTGTAGGGATGTT
GCTCGGATAATATGGGTCCAACAGAGGGACTCCTCGTCTGTATTGCAGAAGGTAACGTATAGGTGGGAAAGCCAACAGCTACACT
GTATTTTTCAACCATTACTTGTTTCTCCGGGTCCAGTTTTTTTGGACAACGACCCCTTGCCTTTACAGACTTTGCAGCATGTAAA
CTACGTACGATATCTTTTAGGAGTTCCGGTGCTAAGATTCTCCATGATAGAAGAATCCCCCAGTGCACAATATATCCCTTATGAC
AAAACTGCACCTTCTGCATATGATGTTTTAGCGACTATTGCACGTACTTGCCTAACTTCCAGCGGCTGTTCGCGTGCTTCTCACG
ACCATACCTACGCACGCGATAACGAAAAACCAAACTACTCCTCCTCTCGAATCGTCTTATTAGCTGATTCCGCCCTTGATACACC
TATGCCCTAAACATGACCGCCCAGGCTATAAAACAGATGGACTGGCCAGACAACATCATTGCAGCGATGGCTGAAGGCGATGGGT
TTGAAGAGCGTCTCCGCGCTAAAGCGCGGTGCCATCAGATGAAGAGGACTCCGATTGGACTCCTCCAAGTAATCCTTGACGGGGC
AAGGAAAATTGATCCGAAACGGATTTTAGATAATTCTTCCCCGAGCGACTTAGCTGTGTTAGCTGAGCGAGAGCTACCATCCTCC
CTGCTAGGACCGGCGCGCGAAGAATACTTAAAATTTGGGGATAAAAAACTGCCAGACAACTTTATCTCACGCACTGTCTCCAATG
ACACCACGGATGAGTATAGAAGAAATCTTGACGAAGCGTTTAAACTCTTTATTGCACGCAGAGGGGTCACTGAAAAATGCGTTGC

CCGAATGCTCAGACATCTCTATGCCAAGCGGTTGTCCGTGGTTTTGGATCTCGATGTCCCATTTGGTGAAAAAATAAACAAAGCG
CTTTCCGCCGATGATCAAGATTTTGTACTGAAACTGTACTACTGGGTTTTCCCGGACAAGGTCTCTGCGCCTAGACTTAAGGAAC
CAGAAATGCAAGCTTACCTGGCTTCAGTAACCTATTTGTCGTGCACCTTTAAGTTCATTGAAGGATACGCTTATTATATGCGCCC
CCCTGGGCCAGAACTCCATTCAACAGAAACGACCGCTCGTCTAGCAGCCCTCCTTGTATATGTGAGAAGCATGTACTGCAAACTG
GCGTCCTTATTTGACGCAATCGATGCCGCTGTGCTGCGCTCTAAATTTAATAGGGGAGCTTCGGCTCAAGCTCTGACTGGAATTT
ATGGATACACGCGAAACGGGAAGGGAACTATAATGTCGCGGTATTCTTCGGAGGCCATGAGTGATGAAATCATGTCTGTGTGTCG
CACTGGTTCGGACACTTTGATAGAATCTTTACGCAAAGCTCTGGGGAGATTGACCTACCTTATTGCTAGGTGGGAGCTCATTTCG
GTAGAGAAGAAAAGGGACGCAAAAGAGGCCGTGACGGCAGCGTTAGCATTAAGTGGACTAGTTAGCGGACATTTGGGATACTTGC
TAAACCTAATCTCAATGGCATACGTTTTTTGGCTCCCTGGGCATTATTTAAATATGCGGTTAATATTGGCGCTGAGTTCCAGGGA
ACTCCTAAGCCATCTTTTGGCTACAAATCCTTCTTCATTGAGGGAAGTAGCCAGCTGCCCGAAAATACTTTTCTACCAATCTGTA
GAAGAATGGAGAGGACAGGAAAGAGGAACACAGATTTGGTTGGATTATGTTACTGTAAGGGCATGGTTGGATTATATCGACACTG
GGATCCCTACTATTCGCTCCGAGGAAACAAATGCTTCCATAAATGAAATTTCCGAGTATCTGGAACAGCGACGCTCGGCCAGTCC
TACTGAATCCTTTTTTATCTTCAGATGGACCACCAAACTCAAGCGCAGCATCCCTGCCCCCTACACAAGGCTCAATGCAAACACAA
AACAGGAGACGCGAACGACGCAGTAACAGTGTCGTTATCCGAAAGCCCTCCCGCTCTGCGCATTCAAAGAGTAATACCATTCCCG
TGTTTCAGGAAACGACCATTTCTTATTCGGCACCGAATTCCCCGGCGGATCATTACGTAGTAATGAATTAAGCCTGTATAACTGT
ACAGTAAATGCATCGCCAATTTTTATATTTACGCTCTGAATAAATAATGACATTACATTGACGTAAGATGTACATAATTCGTATT
TATTTACTTT

>UL45
AATCGCGGTGCGAACGTTACGTATGCAGTTACCGCCCATATCATGTAAATCAGGTCACGGGACTGGATTCGTAACCTATGATTGG
TGCGGGGAGGTCTAGAACAAACATTCGCACCTCAGTATTAAAAATATAGTATATCTATATTATATTTTTAATACTGAGGTGCGAA
TGTTTGTTCTAGACCTCCCCGCACCAATCATAGGTTACGAATCCAGTCCCGTGACCTGATTTACATGATATGGGCGGTAACTGCA
TACGTAACGTTCGCACCGAGTGCTATGAACCAGTTTCTCGCGCCCGCACATTTTTCGCGCAATTAAGCAATAGCGTCACCATGTG
ATGACCTTGACACCACTTTCAATAACATAATACGTGAAATATTGACGTTAACGCCATAACGGGGCCGAAACTATACGCAGAAATG
ACATTTGTTCTTTTTAAAATTCGAGCGTGCAGCGCCCTCGTTACTCGCTTCGTCTAGCGGCCCAAGTGGATTTTCTAGATATGCC
TGAAATAGCAGCAGGCCAGTCCCTTCCAGTAATTTTGCAGCCCTCATCTGAAGATCAGCGAAGCTACCTGAACCCCCCTATAATT
TCACAAACCCCATCTGAAGATCAACGAAGCCATTTGGAAAAAATGATGGCAGAACTTCCCTTCGAAGTCAGGGCTATTTTGGAAG
ATGAGCCAAGCGTTTGTTCCTCTACTTCATCCTCTACGACAAGTACCTATTCCCCCCATCCTTCCGGAGAGCTTCATTCAATGCG
CCCCGAAAATACAGGCGGACAATATTGGAAGGCGTGTAAAGTGATTCTTCTAATCCTGCTTGGCATTTTCCTAGGATTACTGATT
TTTATGTTTGTAATAAGCACTGTCATGGCATTTGCACTGCCGAACAGCACCTGGGAAGGCGGATTGTGTCCGCCTGACTGGATTC
AACATCTGGATTCATGTTATCGGGCCGGGGGTGGTCCGCCTCAAGACACGTATGAAAATGCTAAACTTTCTTGCCGTGCCCTCGG
GGGGAAAATTGCCTCATCTTCAAAAGCAATCGGCTTCCTCACTATTTTAGACCGGTTATTTTATTGGTCCGGGAGCACTTCACCT
CAAGGACTATGGTGGGCGGATGATGGCGGCAAATGCGCCAACGTTTACTTCATAAATTCTATACATGATGACTCCACCGCATCTC
TAAGCACGCAAAAAATGTCCGTGGCCCTAGTCGAGAATCTACAATGTTCTGGAACCGCCGGAGTAATATGCTACATGTCCCCTGC
TATTTCGCCTGCCTTGCGGTATATTAAGGCATTGAGAGTTGCCCTGTCTTTCGGAGTCTACGAGGAAAGTAAATAAATACGAATT
ATGTACATCTTACGTCAATGTAATGTCATTATTTATTCAGAGCGTAAATATAAAAATTGG

>ORFA
GATGGGGGGAATAGGTACTTGTCGTAGAGGATGAAGTAGAGGAACAAACGCTTGGCTCATCTTCCAAAATAGCCCTGACTTCGAA
GGGAAGTTCTGCCATCATTTTTTCCAAATGGCTTCGTTGATCTTCAGATGGGGTTTGTGAAATTATAGGGGGGTTCAGGTAGCTT
CGCTGATCTTCAGATGAGGGCTGCAAAATTACTGGAAGGGACTGGCCTGCTGCTATTTCAGGCATATCTAGAAAATCCACTTGGG
CCGCTAGACGAAGCGAGTAACGAGGGCGCTGCACGCTCGAATTTTAAAAAGAACAAATGTCATTTCTGCGTATAGTTTCGGCCCC
GTTATGGCGTTAACGTCAATATTTCACGTATTATGTTATTGAAAGTGGTGTCAAGGTCATCACATGGTGACGCTATTGCTTAATT
GCGCGAAAAATGTGCGGGCGCGAGAAACTGGTTCATAGCACTCGGTGCGAACGTTACGTATGCAGTTACCGCCCATATCATGTAA
ATCAGGTCACGGGACTGGATTCGTAACCTATGATTGGTGCGGGGAGGTCTAGAACAAACATTCGCACCTCAGTATTAAAAATATA
ATATAGATATACTATATTTTTAATACTGAGGTGCGAATGTTTGTTCTAGACCTCCCCGCACCAATCATAGGTTACGAATCCAGTC
CCGTGACCTGATTTACATGATATGGGCGGTAACTGCATACGTAACGTTCGCACCGCGATTTCGAGTAACCGGGAGCATAAATACC
GCGTGCGTTTGGCACTCCGTGCGCAATATGCGCCCGCCACTGCCGACGTCAGATGCTGTTCTGATCGAAAGGTTACAGGCGCCGG
AATTCTGTACAACCTAACGTAGGGCAGGGTTCCGGCGCCATACGGGTCGCAAGAAACAGGTACAATCCCCTTTCCCCTAATTCAT
TATGTCAATATTTTTTAGGTTGTGGGGCGGCACATCCGGCAGGATGTCAATCGAGATTTTTGAATACAGGTGATGGAGGGAGTAA
CCGACACGGTGCTGGTTTTGACGGAGCGCGACAGGCTCGCCCTATTCGGACCCATTCTGTTTCGAGGAGGGGGACTTACGGCAGC
CCGGTGGGCTGTTTCAGTAGATGCTGTTCTACCGTGGACCGAAAAAACCATCCCCCCGTTAAAGGGGTCCAAGCTCATCATTGCT
GGGGAACCTTACTTGGGGTTAACATCATCTGATGTTCTCTCCAAGTGTGAAGGGAAAACATTTTTTGTGCTCATCGCTGTTCCAG
CAGCCCTCGGACGCTGGATTTGCGCGCCCCCGCCACCGCTTCTAGAGTCCGTTGCAGAAAATGTGCGTTGTAAAATTTCAGCAAC
CCTATCCAAAGAGCATGGAGCGCCCTTTTGCACTGCCGTATACATTTTGGGGTCAACTACACAGGTTCCCCCTGGCTCTAGCTTA
GGTTGGTCTTGTGTCGGGCAACCTGTCCTTCGGGCCCCTTTAAAAGAAGCTACCAAGCCCGACGAATCGTTACAATCTGCAGAGT
GTGTATTTAAGTGTCCGCTTTTTGCGGATGAGTTTTCGGGGAATAACCCGTTTGCAATTGGGTTTTTGTGCGCACTGGACAAGGG
AGTTTTGAAGCACGAATTGTTTGCAGCAATTTATGCAGCAAAAAAGGCAATTTCGAATGCATCCTTAGACGGGACATTTTGTGCT
TCGTTTGGAGATAAAGAGATGAAATTAGAAAGTGTAAGCACCCACTTAGGGCCTGCCGTTGTGGTAACGCGTGGGGCAATTACAG
ACTTCCCTCCAAAGAACGAATTTATGGATCGGCTTACATGCAAGCGAACCTTTTTGCGTAAAGATGGGGACTTTCTCCTGGGATC
TGCGTTCTTAATGTGTAAGCGACAAAATAGAACGTCGCCGGTTTTACGCAGGTGGAAAAGGGCCTGGCTTCGAGATAGGCGAAGG
AAACATTTAATGCGCAGGGCAAGGGTTTCGGTAACAGCTGAGCTGGGTTGTCGGATTTTATCGCATCCAAGCAGAATAGCATCCT
TCGGGCTAGAAGTTACAGAAAGAGATCTGAATTTACAGCGGAGGCTTTCGGCACATAGACCGAGGGCCCGGGGGGTGGTATATGA
AGATTTTGATTAGTTACCTGAAGAGACACGCCCCAAAATCCCACCCACGCTGATGCATGCCTTGGGCATAAAATACTCTGCGATC
TTCGCAGTGGTCAGGCATACTTTATACGCTCCAACTTCAGTTGTCATGGAACAGCGGCGAGACGAAAAAGAAGCGGAGGAACAGA
AAGTATCTGCAATAGCCAGTGCCTTTCTCAATGTGGGACAATTTTTTTCTCGTGGGCAAACGAACAAGTCTCGACCGTGGGTCAC
ATCATCTGAAACAATAAAAGTTTCTATATATGCATGGACTCGGTGCAGAGACTCTCTACTATTTTCAGATTACTTGTTTCGCCGC
GGCACCATGGATAGTGCCCGGCAGCAGTTGCAGGTTGGAGTGCCTAAGAAGGGCGGGGTGTACAGACAGCTATCACGTGATGATG
GAACGGTGTTCGAGGTCAGTTTGGATCCTGCAGTGTGCTTCGAGTTCTCAGTGACGTTAATTTTACCGGGACATGATCTTTTCTG
GCCGGTAGTGCCGCCGCTGCAATTTCTCGAGCTGATAAGTCAGCGCGCGGTTATTCTGGCAGACCAGTTTATTTCGTCCTCTATC
ATGAAGCGAGTGAGCTTATCTCCAATCATGCTATTTCCGAAGAATACATTCATGCCCGGGTACTGGTCCCCGGATCCCCAGCCTG
GTCGGTACCGACCAAAATTTCAGCCTACTCGATCAGAACAACATTTCATGACCGTCGGCAAGCTGGTGCCACCGTTTCAAATTGA
CTTGCATGGGAAAAAGAATAATTTCATGGCTGGAATTGCCGTGGGTTTTCACGGCCCGCCGACCTTGACGGGCACTATTCGCGCC
CTGACTGAACAAGCAATTCACAACGCTGTGGCCGAGGTGGTTCGAACGCTTGAACCGATGACTGTGGTTCCGATCACTCTGAAAA

ATGGAACTGGAGCTCTGATGGGATTGACTTCAGACAACAAGGGTCTTCGGATATTGATAAAACCGGCACTAGGAGAATTGGCCAG
GTCTCCACGAAGACGGCGTTCCGAATCGCGCGGAAGAGAGTCTGTTTTATTATAAATTGAATAAACGCAGACTTTTTATTACTGT
CAATAATTTATTTATTGGTCTTTATTACAC

>ORFB

GGTCCAAGCTCATCATTGCTGGGGAACCTTACTTGGGGTTAACATCATCTGATGTTCTCTCCAAGTGTGAAGGGAAAACATTTTT
TGTGCTCATCGCTGTTCCAGCAGCCCTCGGACGCTGGATTTGCGCGCCCCCGCCACCGCTTCTAGAGTCCGTTGCAGAAAATGTG
CGTTGTAAAATTTCAGCAACCCTATCCAAAGAGCATGGAGCGCCCTTTTGCACTGCCGTATACATTTTGGGGTCAACTACACAGG
TTCCCCCTGGCTCTAGCTTAGGTTGGTCTTGTGTCGGGCAACCTGTCCTTCGGGCCCCTTTAAAAGAAGCTACCAAGCCCGACGA
ATCGTTACAATCTGCAGAGTGTGTATTTAAGTGTCCGCTTTTTGCGGATGAGTTTTCGGGGAATAACCCGTTTGCAATTGGGTTT
TTGTGCGCACTGGACAAGGGAGTTTTGAAGCACGAATTGTTTGCAGCAATTTATGCAGCAAAAAAGGCAATTTCGAATGCATCCT
TAGACGGGACATTTTGTGCTTCGTTTGGAGATAAAGAGATGAAATTAGAAAGTGTAAGCACCCACTTAGGGCCTGCCGTTGTGGT
AACGCGTGGGGCAATTACAGACTTCCCTCCAAAGAACGAATTTATGGATCGGCTTACATGCAAGCGAACCTTTTTGCGTAAAGAT
GGGGACTTTCTCCTGGGATCTGCGTTCTTAATGTGTAAGCGACAAAATAGAACGTCGCCGGTTTTACGCAGGTGGAAAAGGGCCT
GGCTTCGAGATAGGCGAAGGAAACATTTAATGCGCAGGGCAAGGGTTTCGGTAACAGCTGAGCTGGGTTGTCGGATTTTATCGCA
TCCAAGCAGAATAGCATCCTTCGGGCTAGAAGTTACAGAAAGAGATCTGAATTTACAGCGGAGGCTTTCGGCACATAGACCGAGG
GCCCGGGGGGTGGTATATGAAGATTTTGATTAGTTACCTGAAGAGACACGCCCCAAAATCCCACCCACGCTGATGCATGCCTTGG
GCATAAAATACTCTGCGATCTTCGCAGTGGTCAGGCATACTTTATACGCTCCAACTTCAGTTGTCATGGAACAGCGGCGAGACGA
AAAAGAAGCGGAGGAACAGAAAGTATCTGCAATAGCCAGTGCCTTTCTCAATGTGGGACAATTTTTTTCTCGTGGGCAAACGAAC
AAGTCTCGACCGTGGGTCACATCATCTGAAACAATAAAAGTTTCTATATATGCATGGACTCGGTGCAGAGACTCTCTACTATTTT
CAGATTACTTGTTTCGCCGCGGCACCATGGATAGTGCCCGGCAGCAGTTGCAGGTTGGAGTGCCTAAGAAGGGCGGGGTGTACAG
ACAGCTATCACGTGATGATGGAACGGTGTTCGAGGTCAGTTTGGATCCTGCAGTGTGCTTCGAGTTCTCAGTGACGTTAATTTTA
CCGGGACATGATCTTTTCTGGCCGGTAGTGCCGCCGCTGCAATTTCTCGAGCTGATAAGTCAGCGCGCGGTTATTCTGGCAGACC
AGTTTATTTCGTCCTCTATCATGAAGCGAGTGAGCTTATCTCCAATCATGCTATTTCCGAAGAATACATTCATGCCCGGGTACTG
GTCCCCGGATCCCCAGCCTGGTCGGTACCGACCAAAATTTCAGCCTACTCGATCAGAACAACATTTCATGACCGTCGGCAAGCTG
GTGCCACCGTTTCAAATTGACTTGCATGGGAAAAAGAATAATTTCATGGCTGGAATTGCCGTGGGTTTTCACGGCCCGCCGACCT
TGACGGGCACTATTCGCGCCCTGACTGAACAAGCAATTCACAACGCTGTGGCCGAGGTGGTTCGAACGCTTGAACCGATGACTGT
GGTTCCGATCACTCTGAAAAATGGAACTGGAGCTCTGATGGGATTGACTTCAGACAACAAGGGTCTTCGGATATTGATAAAACCG
GCACTAGGAGAATTGGCCAGGTCTCCACGAAGACGGCGTTCCGAATCGCGCGGAAGAGAGTCTGTTTTATTATAAATTGAATAAA
CGCAGACTTTTTATTACTGTCAATAATTTATTTATTGGTCTTTATTACAC

>ORFC

GCTGGAAAACTATGCCAAGCCCAGCTTGAAAGTTATTAGAATCCTTAATACCATCACATCCAAAAAATTGTGGCTGAACAAGTAA
TGAAGCAGGGTCAAACAAAGGGTCATATAGACTTCGCTGCTTCTGTCTCGTGCGAAGGAAAGTTTCTTTGCTCTCTTCTCCAAAA
GTTGGTACGCTAAGATAGTATAGAGGGCAAATGAACCCTTTGGCAAGTTTTCCTGAAACGAGTATCTTGGGGGGAAGATTGTTAA
GCGCGCGCGCCAGGCCCCAAAACAGAGCAGGATTGGGAACACACGGCCACATGCGATCTAACTTAGGAACCACCGGTAAAAAAAA
CATTTCTTCATGTGCCCACACTTCTTCGCCCATCTCCAACACGCCGATGGCGCACGCTCCAAGCTCTCCATCCACATCGACAGTA
TTTTCTGATTGTGGAGCTGGGGACATATCATTTCTGTATGCTAAACGGCTGAACCTTCCACGCTGGCTTGTCCTGGCAGGGTAAA
TGGTCAACCCATGTGCACTCCCAAGACAAGATGGCAAGCCTCCGTTTACGCGATTTAGGATCACTGCATGCTTAGCAACTGCAAA
CGCTGATCCGTCGGGAATACAATCACAGGAAAAAATTGCCTCGTACATCCACGAAGAATATGAAATCTTTCCAGCTCCGACGAAG
ACCATCTCTTCCTCTAATTTTTCCGAATCATTGCAATAATCACTTGCTGCGACAATTTGCTTGGCAGTGCGTTCACATTTTTCGG
GCGAACTATCATTCTTCGCTTTTGGTTTCTGAAAACGAGAGCAGGCCATGGAACTCATGTCTGCGGAAATCGTGCGTGGCGGGCG
TCGAGGCACATGTCTTATATCCACGCCCAGTACACGCCCCATGCAACAAGCACCTAAAAGCCCTCGAACGCGTTCGAGGGTCTGC
TCTGAATCTCAGAACTCTCTAGGCGAGAGCCTGCTCTGAAACTCAGAACTCTCTAGGCGAGACACCATGGGCTGGCTGAGCAAAT
TCATGCCGTTTGGACGCCGCAAGCGCTTCTCTCGATCGGAACCTCAGCTACACCGACTGGAAATATCTGGGCCGATTTCCCCTCC
GCTGAAACCCGCGCCATCATTACCTTCAGTTAAGAACTCCCACTCCGGTCGGAGTGGCTGGATACGCCAAACACTCCGGCGTAAA
AAAGATGATGAACCACAAGAAAATGCTGAAATAACGAACCAGCTTCTCGACAGTATGGTAGACAAGGAACGACACGGAAACTCTA
TACGACAGCAAAGGCCTACTCCTAGTGCTCAGGGCCCTTGTTGGATATTTTCTAACAGATCTTATCATCACATCAGCATACTGCC
TGAGCATCTACCGGAACTCTTTGGCAAGACCAGTGATGAGATACAGGGGATTCAATCGGCTTTTGATCACACAGGCGACCCTTTA
TTTGCACCAGCACTACAGGCAATATTAGTGGAAATATTGGGCCCGCGCTTTTTATACGAAAAGCGCTGCAAGGGAGCATCTTCCC
CAATATCTCTTGACATTCCGATATATTTTTCCGGAACATGTCTAACGCGAAGAGTACGCATTGGAATATTTGACGCCCCAATCGC
CTTCCCAGAAGAGTTTCTCGGAGAGCAATACTTCGTCCCATATCTGAGATTATGGCCTAAAACCTTCACGCGTGCGGAATCCCTG
TCTCTCCCCGATGCGCTCTCCGCAGCAACCTACGCAAACTTGAGATGGGACAGGTCCTATTCGATTCTTGGAAACGTGAGTTTTT
ACGCCGCCGCCATGTTGTCAGAAAGTGGACCAACGCCAGATTTTACAACGGCCGGGGCGCAAGTAATAGATTCCCCGTACAGGTT
TTGCAGTTTTTTGCCAGGCATTTTGGTCGTTTTGCCGCAACAACAAGAACTGTGTATCCCTACTATGCAGAACGCGGTAGTGGTT
GCCAGAGAGGCAATCGAGTCGTGTATAATGTTACCCGAAGGAGCCGTGTAATAAAGACCAATAAATAAATTATTGACAGTAATAA
AAAGTCTGCGTTTATTCAATTTATAATAAAACAGACTCTCTTCCGCGCGATTCGGAACGCCGTCTTCGTG

>ORFD

ACGACCAAAATGCCTGGCAAAAAACTGCAAAACCTGTACGGGGAATCTATTACTTGCGCCCCGGCCGTTGTAAAATCTGGCGTTG
GTCCACTTTCTGACAACATGGCGGCGGCGTAAAAACTCACGTTTCCAAGAATCGAATAGGACCTGTCCCATCTCAAGTTTGCGTA
GGTTGCTGCGGAGAGCGCATCGGGGAGAGACAGGGATTCCGCACGCGTGAAGGTTTTAGGCCATAATCTCAGATATGGGACGAAG
TATTGCTCTCCGAGAAACTCTTCTGGGAAGGCGATTGGGGCGTCAAATATTCCAATGCGTACTCTTCGCGTTAGACATGTTCCGG
AAAAATATATCGGAATGTCAAGAGATATTGGGGAAGATGCTCCCTTGCAGCGCTTTTCGTATAAAAAGCGCGGGCCCAATATTTC
CACTAATATTGCCTGTAGTGCTGGTGCAAATAAAGGGTCGCCTGTGTGATCAAAAGCCGATTGAATCCCCTGTATCTCATCACTG
GTCTTGCCAAAGAGTTCCGGTAGATGCTCAGGCAGTATGCTGATGTGATGATAAGATCTGTTAGAAAATATCCAACAAGGGCCCT
GAGCACTAGGAGTAGGCCTTTGCTGTCGTATAGAGTTTCCGTGTCGTTCCTTGTCTACCATACTGTCGAGAAGCTGGTTCGTTAT
TTCAGCATTTTCTTGTGGTTCATCATCTTTTTTACGCCGGAGTGTTTGGCGTATCCAGCCACTCCGACCGGAGTGGGAGTTCTTA
ACTGAAGGTAATGATGGCGCGGGTTTCAGCGGAGGGGAAATCGGCCCAGATATTTCCAGTCGGTGTAGCTGAGGTTCCGATCGAG
AGAAGCGCTTGCGGCGTCCAAACGGCATGAATTTGCTCAGCCAGCCCATGGTGTCTCGCCTAGAGAGTTCTGAGTTTCAGAGCAG
GCTCTCGCCTAGAGAGTTCTGAGATTCAGAGCAGACCCTCGAACGCGTTCGAGGGCTTTTAGGTGCTTGTTGCATGGGGCGTGTA
CTGGGCGTGGATATAAGACATGTGCCTCGACGCCCGCCACGCACGATTTCCGCAGACATGAGTTCCATGGCCTGCTCTCGTTTTC
AGAAACCAAAAGCGAAGAATGATAGTTCGCCCGAAAAATGTGAACGCACTGCCAAGCAAATTGTCGCAGCAAGTGATTATTGCAA
TGATTCGGAAAAATTAGAGGAAGAGATGGTCTTCGTCGGAGCTGGAAAGATTTCATATTCTTCGTGGATGTACGAGGCAATTTTT

TCCTGTGATTGTATTCCCGACGGATCAGCGTTTGCAGTTGCTAAGCATGCAGTGATCCTAAATCGCGTAAACGGAGGCTTGCCAT
CTTGTCTTGGGAGTGCACATGGGTTGACCATTTACCCTGCCAGGACAAGCCAGCGTGGAAGGTTCAGCCGTTTAGCATACAGAAA
TGATATGTCCCCAGCTCCACAATCAGAAAATACTGTCGATGTGGATGGAGAGCTTGGAGCGTGCGCCATCGGCGTGTTGGAGATG
GGCGAAGAAGTGTGGGCACATGAAGAAATGTTTTTTTTACCGGTGGTTCCTAAGTTAGATCGCATGTGGCCGTGTGTTCCCAATC
CTGCTCTGTTTTGGGGCCTGGCGCGCGCGCTTAACAATCTTCCCCCCAAGATACTCGTTTCAGGAAAACTTGCCAAAGGGTTCAT
TTGCCCTCTATACTATCTTAGCGTACCAACTTTTGGAGAAGAGAGCAAAGAAACTTTCCTTCGCACGAGACAGAAGCAGCGAAGT
CTATATGACCCTTTGTTTGACCCTGCTTCATTACTTGTTCAGCCACAATTTTTTGGATGTGATGGTATTAAGGATTCTAATAACT
TTCAAGCTGGGCTTGGCATAGTTTTCCAGCCTCAGGATGGATGGACGGCAAAAACAAAGCTGCTCGAGTTGGACAGAGTTCGGTC
GGAACTTGTCACACTATTTGCACGCCAGGAGCAGTCAAGGCCTATTACCGTGATAATTACTAGGGGGGCAGTCAGACGCCGGCTT
GTGTTTTCTACTGTTGGAGGGTTTAGGTTTTCATCCTCGTTCGAACCCTCTGGGACACCCCTTAAAACCCACTCGGTGACTTTAC
ATATTTAATGGAACACACATACGAAATAAAAGTTTGTCAGAAATGTATTTGCGTTTATTGATGTTTATGTTTCGGGTAAT

>ORFE

AAGGAACTTCCTCAGGCGCATGACAATTTTATGCTCTGTCGACCGGCTAGCCGATATCAACGAGATTCACTCCCAACCAGACGGG
AGCTTCAAGAAAGGACAGGCATCGAAAATTGAAGAGTGGTTTTCTCCTTGTGCAATTGCGACAAGGTTTGATATTTCCAGAAGTG
CACCATGGAAAATAATTCTGACAGCATCTGCTCCATTTCGATTTTTTTACACCAAAATATATTTATCATCTGAAGGTGAATTGCC
ATCATATAGTGCACCGATACCGCCACAAGCTCTGTTGACCCCTCAGATCCTGGCAACTGCGTTTTTTCCAGAGGTAAAGAAATGT
TTAAACTGGGACCTGGAACACCACATCGTATTTTTCCTCCCTGTTACCGCGCATTCTAGTTGGGTGATGACAAGTAGCCCGCCGA
GCACTGGAATGACCTACAGACTTCCGGATACCTCAACAGCTGCCCCTTTGTTTGTAACGTTTGTGAATTCAAGTTGCAACTCCCT
ACCGCAAATCGCGCGAACAGCAAAAATGGAGGATGGTCATTTGAATGGACGAGCGAAAACTTGCATATATTGCAACCACCTGATA
GTTCAATACGGTCTACGGGGGTTCCAGGATGCTTTCTTTATAGACAGTGCATATGTTCAATCAGAACTGTTAGCCGGTGCCAACT
CATCTATTGCATTTTTTAAGTCTATGAACCCCGAGTTCTATAACTACGCGCTTCTAGGACCACACGGAACTGCTGTCCGAATTTT
AGGACTTGACAGAAGCATAATAGTTTTTGACAAGGCATACATTATCGGTATTGTGGTAGGGGCACTTGTTGGATTGGTTGCTCTA
TGGAGTGCCATCAGGGCTATATATAAATTCTGCCTCAGCTCGGCAACGCAATTTAGACGTGTACGGCTACCGTCTGACGAGGAGA
GCGAACTTTAATCCTTGTTCCTTACAAACCCTTTTCAACTGCACTTGCGGCAGCACGAATAACTGGCTATGGCTCCGTCAAAGAA
GCAGTCTCAGGCTTCCAAAGCACTGTCTACCAGCAATCGCGAATGTCGTTTTGCAGTAACTCACTGCCTCGCAAACTTTGTAGAC
TCGTCTCACATGGAATTAGGAACACATCGTGGAAAACTATGGACAGAAAAGGTAGAATATGGCCTAGCTGTACTGCACGGCACAG
CATCGTTTTTTCTGGATTGCAACAAGGATTTGACTATTCATGTGTTTCAAGACCATGACTCAAAACAGCCCGATAACTCTCAAAT
CCGATTCGTAGCAGCACCATTATTTGAGGATCCGCAAGAATCCCTTTCCAACGACACTGGAATTCCTTCTACTGCAATTATCATG
ACTAGAACTCCCTACACAACTAATTCTAATTTAGAGATTGGCGTCGAGCAGCTCCTTCTTGCCATGAAAAGAGCTTCCTTAAAAA
ATTTCCAAAGATGGCAAGCCGTAGCTGATAGCGTGCATTCAAAAAACCCGATGACTTTTGGAGCCAAGTATTTGCCTTTTAAACT
GTTGTCTGCAGACAAGTGGAAAGATGGAGTCGGATTTCTAATGTTAGGACTTATCTCGGATTTGGAACTTTCTCCGGTTGTTGCT
GCTGTACGCGCATTCAGCGGGGCCTTGTTCTTTCAAAGTCCTCCATCACTTACTCTAAGCGGAGAAATCCTTGATTTAAAGCACG
GAGACAGCCCGCATCACTCGCCCGTCGGACTACTACAAGCGGCGATATTTACTGGGTTACGCGCGTATCGATCACAGATAGCTTC
GCCGATAATTACTTGCCGCGAAGAATGGGCTTTCCGTATCCCGTTTTGTGTTGCGATTGGCGACAGCCAAATACATCCACCGCGC
GCACAGCAGACAGCCTCGATTCCGAAAGGAACTCACATTCTTTGCGCGTTTTTCCCAGTTTCGGGGGTATTGGCATCGGACGAAC
GCTCATATCTAGCGGTAGATACATGGGCATCAAGAGCTACTAGCTGGCTTACTGGGCGAGGCTTTACCTTCTGGCGCGCTCTTTT
AGAAGAAACTCCCATGTGCGGCAAGTTTCTGTGCGCAGCAATTTGCCACGCGTACGATCCACCTAGACTCGGCATTTTATTGCAC
TCTCATTGGACCCCCGATGAGATAGGCGATTTTAAGACGGCTGCTTGCACAGCAGTGTCCGAAGCATGGAAGAACATGGGAATTG
CGACGCGGCCTGATCTGTTTGGATAATTACCCGAAACATAAACATCAATAAACGCAAATACATTTCTGACAAACTTTTATTTCGT
ATGTGTGTTCCATTAAAT

>UL22

CAGTCACTGCAAAGACGTTGATGCAAATGCCAAATTGGAGAGGTTGCCGTCTATACTTAGCGGAACCTATGCAAGCATGGCGCCA
ATGGTTTGGCGGAGCGGATATGATCAAAGAAATTAATGAAATACAAACCCTAAAGGCTTCCGGAAAACTTGAATGTCGGGAGGCG
TCTCCGGTTGCCGTAGCGGAAGTTCAGATGACTATTGCTGCCCCACTAAGAATAATGAACCACGTCATTTATAATTATTTGGGAT
CTGAACGCTGCTACAGCGCAGCTGCATCCGGACCAGATGATGTCTTATTCCTCGTAGATAGGCACCCACTCGCGGCATGTTTGTG
TTTCCCTGTTGCACAATATCTAAGCGGAGCGCTCGAATTTGGAGATTTAATAACTTTATTGTCAGGAATTCCTGACATTCCAACA
CACTCCAACATTGTTTTAATGGATTTGGATATTTGCGAACAGGCACGGCGTATAATACAAAGGGGGCGCCCAGGGGAAACGGTCG
ACTGGACGTATTTGTGTGCATTACGTAACTCGTACATCTGCCTCATGAATACTACCACCTACCTCCAACGTACATCTTATCCAGC
ATTGTTGAAGGAGCAAGAAGCCTTAACAAGTGCCACGCTCTTAAAATTCAAGAGAGAGTGCTTAGAAACTGCTACTGTTCCAGAA
ATCAATCCTTCAATCGACCAGACGCTATTTGCAATATTAGCTTTTGATCAGCAAAATGTTCACGGGGAAAGATTAAAAACTGTAC
TTTCATTTGTGGTTCAAAAACTCGCGACGGTATTGAAAAACTTGTGCATTTTTTACTTACCAGCACATGGCCTCACCCCGGAGGC
ATGTGCACTGAAATGTTTAGAGTTTGCCGAGACGGCAAGTTCTCTTACAACCAAACGAGCGGCGATCGCGAGCTTAATTGACGCA
GTAGAGCGCTACAATGCTGATATGGGTTCGTAATGTTCCGCTTCCATAATCCTTCACAATAAGAGTATGTCCTTTACTCATTTCC
TTGCTTTGTACTCATTCTTACTCGAGAGAGCGTGGCTTCACCAGCAACCCGCGCCGATGGGACACGCGAGAGAAATATTTTTCCT
CATTCTGGCGATATTACTTAGAGCTAAAGGGGATGACAAGGCAAAGGCAAAAATTGTCTTCGTGCCCCCCACCAGTGAAAATTTG
GCCCCTCCGAAAGAATTCATTGCGATAGCGACAGCAGTCGAAAACGGCCACATCACACAACTATGCCTTGTCGATCGAAACTTAA
TCTTTGGGCCTCCTGGAAGCGCTAACAGGAATTTATCGATCGCTCTAGATTTGCTTGGACAACTGTACGTAGCCTATACTTACTC
GAGATGGGGTTCTGCTAGCGAGGCTACCATCCAACCACGAAAATTCAGCAACGATCCAAAGCTTGAGCATCGACTTTTGGTAGGA
TACAAGGGGAATGCACTTACATCTGCGAAACTCGTCAACTTTCCACCAAATAACTCAGTTCATAGGGGCGAGAGCATCATCGCTC
CGATTCGTGTAATTTACAGTGATCCTTTACAAACTGTTTTTCCATTTGTGTTTACACGATATATTAGCCTCGCAGTTACTAGTGA
CATAGATCTTAAGATTAACATCGGACTGAATGTGGCGACCCTTATAATCGCGACCCAGCATTCTCCTCAAATAGAGATGATAATT
ACACCAACCAGCGCGAAGGTTAAAGAAACGTCCCTGAAGTCATGGCCCTATATGCTGAAGGCTGAAGATCAGGGGTTAACTGTGC
GCATCAACGTCGCAGCTGGACCAAATTATCGGCCAAATTTGTTTACAGCAGTAACGACCGCTGCTCAAGATTTCTTAGAGCTGGC
CAGCTTCGACGAACAAATATCCGCGGGACACTTTTTAACTTTGCAGAAGGTCGCAATCGCAGGTTGCCCGAATACCATAGACGAG
TCTTTTTTTGTAAGATTCGCAGCACGGTTTATGGCCGCGTGGTTTCGATTTCTAAAGTTTACGATCGATGCATACGTAACTGATT
CGGAGATCGTCGAGTTTATTGCAGAAACGCAAACATGGGCAGACCAAGTCGCAGTGTGTTTTGGGGGGGCGAATTCAGTAAACTA
TACAACGAATATTCGCTTTATTCAGAATGTCGGTGGTGCTCTTACTTCTATCCCATCCAGTATTCTTAAAATTATGCACGACAGG
CCATGGGAACCCATGATAAAAAATCATAACCTGCTTCTTACTATCATGAATGCGGTGCTTATCACTGCAGAACAAGATCCCCAAA
ATGCATTCCTGGCGAGCATCGATGGAGTGACCCCCAGCCAACTGGCTCTTGAAATAATAGATAACAGTTACAGAGAAATCATTAC
ACAAAAGGTAATGACCCCCGCGTACAGAAAGTTATTGGCAGTTTCTACGTACATTATTTTGCGGGCTCCTCAACTTTCTATACAA
CATGGAAGGAACTTCCTCAGGCGCATGACAATTTTATGCTCTGTCGACCGGCTAGCCGATATCAACGAGATTCACTCCCAACCAG
ACGGGAGCTTCAAGAAAGGACAGGCATCGAAAATTGAAGAGTGGTTTTCTCCTTGTGCAATTGCGACAAGGTTTGATATTTCCAG

AAGTGCACCATGGAAAATAATTCTGACAGCATCTGCTCCATTTCGATTTTTTTACACCAAAATATATTTATCATCTGAAGGTGAA
TTGCCATCATATAGTGCACCGATACCGCCACAAGCTCTGTTGACCCCTCAGATCCTGGCAACTGCGTTTTTTCCAGAGGTAAAGA
AATGTTTAAACTGGGACCTGGAACACCACATCGTATTTTTCCTCCCTGTTACCGCGCATTCTAGTTGGGTGATGACAAGTAGCCC
GCCGAGCACTGGAATGACCTACAGACTTCCGGATACCTCAACAGCTGCCCCTTTGTTTGTAACGTTTGTGAATTCAAGTTGCAAC
TCCCTACCGCAAATCGCGCGAACAGCAAAAATGGAGGATGGTCATTTGAATGGACGAGCGAAAACTTGCATATATTGCAACCACC
TGATAGTTCAATACGGTCTACGGGGGTTCCAGGATGCTTTCTTTATAGACAGTGCATATGTTCAATCAGAACTGTTAGCCGGTGC
CAACTCATCTATTGCATTTTTTAAGTCTATGAACCCCGAGTTCTATAACTACGCGCTTCTAGGACCACACGGAACTGCTGTCCGA
ATTTTAGGACTTGACAGAAGCATAATAGTTTTTGACAAGGCATACATTATCGGTATTGTGGTAGGGGCACTTGTTGGATTGGTTG
CTCTATGGAGTGCCATCAGGGCTATATATAAATTCTGCCTCAGCTCGGCAACGCAATTTAGACGTGTACGGCTACCGTCTGACGA
GGAGAGCGAACTTTAATCCTTGTTCCTTACAAACCCTTTTCAACTGCACTTGCGGCAGCACGAATAACTGGCTATGGCTCCGTCA
AAGAAGCAGTCTCAGGCTTCCAAAGCACTGTCTACCAGCAATCGCGAATGTCGTTTTGCAGTAACTCACTGCCTCGCAAACTTTG
TAGACTCGTCTCACATGGAATTAGGAACACATCGTGGAAAACTATGGACAGAAAAGGTAGAATATGGCCTAGCTGTACTGCACGG
CACAGCATCGTTTTTTCTGGATTGCAACAAGGATTTGACTATTCATGTGTTTCAAGACCATGACTCAAAACAGCCCGATAACTCT
CAAATCCGATTCGTAGCAGCACCATTATTTGAGGATCCGCAAGAATCCCTTTCCAACGACACTGGAATTCCTTCTACTGCAATTA
TCATGACTAGAACTCCCTACACAACTAATTCTAATTTAGAGATTGGCGTCGAGCAGCTCCTTCTTGCCATGAAAAGAGCTTCCTT
AAAAAATTTCCAAAGATGGCAAGCCGTAGCTGATAGCGTGCATTCAAAAAACCCGATGACTTTTGGAGCCAAGTATTTGCCTTTT
AAACTGTTGTCTGCAGACAAGTGGAAAGATGGAGTCGGATTTCTAATGTTAGGACTTATCTCGGATTTGGAACTTTCTCCGGTTG
TTGCTGCTGTACGCGCATTCAGCGGGGCCTTGTTCTTTCAAAGTCCTCCATCACTTACTCTAAGCGGAGAAATCCTTGATTTAAA
GCACGGAGACAGCCCGCATCACTCGCCCGTCGGACTACTACAAGCGGCGATATTTACTGGGTTACGCGCGTATCGATCACAGATA
GCTTCGCCGATAATTACTTGCCGCGAAGAATGGGCTTTCCGTATCCCGTTTTGTGTTGCGATTGGCGACAGCCAAATACATCCAC
CGCGCGCACAGCAGACAGCCTCGATTCCGAAAGGAACTCACATTCTTTGCGCGTTTTTCCCAGTTTCGGGGGTATTGGCATCGGA
CGAACGCTCATATCTAGCGGTAGATACATGGGCATCAAGAGCTACTAGCTGGCTTACTGGGCGAGGCTTTACCTTCTGGCGCGCT
CTTTTAGAAGAAACTCCCATGTGCGGCAAGTTTCTGTGCGCAGCAATTTGCCACGCGTACGATCCACCTAGACTCGGCATTTTAT
TGCACTCTCATTGGACCCCCGATGAGATAGGCGATTTTAAGACGGCTGCTTGCACAGCAGTGTCCGAAGCATGGAAGAACATGGG
AATTGCGACGCGGCCTGATCTGTTTGGATAATTACCCGAAACATAAACATCAATAAACGCAAATACATTTCTGACAAACTTTTAT
TTCGTATGTGTGTTCCATTAAAT

>UL23

CAATGCCGATGAGACATGTACGGGAACAGTAGCGCTCATATTTTCCAGTTGCTCTAAAGCTGCCTTGGCAGCCCTGTTCCGCCTT
TCCAGATCTTCAATCCTTGCACGATTGGTCATGATTTCTGCGTCTATTAGAGTTCCTTCGAAAAGCTTAGATGCATATTTACTCG
AGCCTTCCGCGCAGGAGATAAAGTTTAAACGATCGGCTAAAACCATCAAGGTAGACGGCTTGTTGCTGTCCTTTGGTAAATTCAT
AGGTTCGAATCTGGGTCGAAACATGGTTCGGACAAGTTTTTGGACTCCTCCGTAGAAGTGGAATAGCTCGCGGGTTTGGTTGCGG
ACTCTTTTCCCACGCGCGACGAAAACAATGCAGCTACGGTTTGTAAGGCTGAAGGGGCGGGAGACGCTAGATCCGCATTCTTTGC
ACTTTCTTGTTGATACTTACTGCGCGCATGCGACACGCGTGGCTCTAATTTCGCGCAGAAAGTCTTTAAATACTGGCTTCTGGTA
GCTAGAGTGACAAAGAAAAGCTCAGACGATACATTTATAACCGTTTCCCCAACCGTTTTTACCGCCAAGATGTTCATCCCCCGTT
GAGCAACAAATACCAGTAGAGAAAGAATCTTTACTTTCTCAGTGCCAGGAACAGCCAGATTCTCCAACAATCGTGCAGAATCGCG
AAGTTGGAGAGTGCCAGTATATCCTTGTCTTAATTTACTTTCAGTCATCAAAGATTTGGAAAAACGACATGTTTTCAGTTCAATC
ACAATACATTTCATTTCATGTTGTGTCTCCAGCAAACAAATGCAATCAGGTTTCCGCAACCCTAGGTTCACTTCAAACATGACTA
CAATTTTGCCCCCGGCAGGTTTGCATTGGGGAATTATCGTATAGGCCAGCCTTCCGTCTCCACCCCCTTCAAAGACTTCCTCCAG
TGATCTGACGAGAGCTCGGTAAAAGCGATTATGGCAACGGATTCCGGCATTTAGTCTAGCCCGCAGAGATGGCCGTAGCTGGCGC
CGTGAAAACTTCCGGTGGTGTGCAGTTTTGCTCCGAGTTCGAGAACGATGACTCCGACTTTCGCCGCGTTGTACTTCTTTACGTC
GACGGGCCATTCGGAGTCGGTAAAACAGTCACTGCAAAGACGTTGATGCAAATGCCAAATTGGAGAGGTTGCCGTCTATACTTAG
CGGAACCTATGCAAGCATGGCGCCAATGGTTTGGCGGAGCGGATATGATCAAAGAAATTAATGAAATACAAACCCTAAAGGCTTC
CGGAAAACTTGAATGTCGGGAGGCGTCTCCGGTTGCCGTAGCGGAAGTTCAGATGACTATTGCTGCCCCACTAAGAATAATGAAC
CACGTCATTTATAATTATTTGGGATCTGAACGCTGCTACAGCGCAGCTGCATCCGGACCAGATGATGTCTTATTCCTCGTAGATA
GGCACCCACTCGCGGCATGTTTGTGTTTCCCTGTTGCACAATATCTAAGCGGAGCGCTCGAATTTGGAGATTTAATAACTTTATT
GTCAGGAATTCCTGACATTCCAACACACTCCAACATTGTTTTAATGGATTTGGATATTTGCGAACAGGCACGGCGTATAATACAA
AGGGGGCGCCCAGGGGAAACGGTCGACTGGACGTATTTGTGTGCATTACGTAACTCGTACATCTGCCTCATGAATACTACCACCT
ACCTCCAACGTACATCTTATCCAGCATTGTTGAAGGAGCAAGAAGCCTTAACAAGTGCCACGCTCTTAAAATTCAAGAGAGAGTG
CTTAGAAACTGCTACTGTTCCAGAAATCAATCCTTCAATCGACCAGACGCTATTTGCAATATTAGCTTTTGATCAGCAAAATGTT
CACGGGGAAAGATTAAAAACTGTACTTTCATTTGTGGTTCAAAAACTCGCGACGGTATTGAAAAACTTGTGCATTTTTTACTTAC
CAGCACATGGCCTCACCCCGGAGGCATGTGCACTGAAATGTTTAGAGTTTGCCGAGACGGCAAGTTCTCTTACAACCAAACGAGC
GGCGATCGCGAGCTTAATTGACGCAGTAGAGCGCTACAATGCTGATATGGGTTCGTAATGTTCCGCTTCCATAATCCTTCACAAT
AAGAGTATGTCCTTTACTCATTTCCTTGCTTTGTACTCATTCTTACTCGAGAGAGCGTGGCTTCACCAGCAACCCGCGCCGATGG
GACACGCGAGAGAAATATTTTTCCTCATTCTGGCGATATTACTTAGAGCTAAAGGGGATGACAAGGCAAAGGCAAAAATTGTCTT
CGTGCCCCCCACCAGTGAAAATTTGGCCCCTCCGAAAGAATTCATTGCGATAGCGACAGCAGTCGAAAACGGCCACATCACACAA
CTATGCCTTGTCGATCGAAACTTAATCTTTGGGCCTCCTGGAAGCGCTAACAGGAATTTATCGATCGCTCTAGATTTGCTTGGAC
AACTGTACGTAGCCTATACTTACTCGAGATGGGGTTCTGCTAGCGAGGCTACCATCCAACCACGAAAATTCAGCAACGATCCAAA
GCTTGAGCATCGACTTTTGGTAGGATACAAGGGGAATGCACTTACATCTGCGAAACTCGTCAACTTTCCACCAAATAACTCAGTT
CATAGGGGCGAGAGCATCATCGCTCCGATTCGTGTAATTTACAGTGATCCTTTACAAACTGTTTTTCCATTTGTGTTTACACGAT
ATATTAGCCTCGCAGTTACTAGTGACATAGATCTTAAGATTAACATCGGACTGAATGTGGCGACCCTTATAATCGCGACCCAGCA
TTCTCCTCAAATAGAGATGATAATTACACCAACCAGCGCGAAGGTTAAAGAAACGTCCCTGAAGTCATGGCCCTATATGCTGAAG
GCTGAAGATCAGGGGTTAACTGTGCGCATCAACGTCGCAGCTGGACCAAATTATCGGCCAAATTTGTTTACAGCAGTAACGACCG
CTGCTCAAGATTTCTTAGAGCTGGCCAGCTTCGACGAACAAATATCCGCGGGACACTTTTTAACTTTGCAGAAGGTCGCAATCGC
AGGTTGCCCGAATACCATAGACGAGTCTTTTTTTGTAAGATTCGCAGCACGGTTTATGGCCGCGTGGTTTCGATTTCTAAAGTTT
ACGATCGATGCATACGTAACTGATTCGGAGATCGTCGAGTTTATTGCAGAAACGCAAACATGGGCAGACCAAGTCGCAGTGTGTT
TTGGGGGGGCGAATTCAGTAAACTATACAACGAATATTCGCTTTATTCAGAATGTCGGTGGTGCTCTTACTTCTATCCCATCCAG
TATTCTTAAAATTATGCACGACAGGCCATGGGAACCCATGATAAAAAATCATAACCTGCTTCTTACTATCATGAATGCGGTGCTT
ATCACTGCAGAACAAGATCCCCAAAATGCATTCCTGGCGAGCATCGATGGAGTGACCCCCAGCCAACTGGCTCTTGAAATAATAG
ATAACAGTTACAGAGAAATCATTACACAAAAGGTAATGACCCCCGCGTACAGAAAGTTATTGGCAGTTTCTACGTACATTATTTT
GCGGGCTCCTCAACTTTCTATACAACATGGAAGGAACTTCCTCAGGCGCATGACAATTTTATGCTCTGTCGACCGGCTAGCCGAT
ATCAACGAGATTCACTCCCAACCAGACGGGAGCTTCAAGAAAGGACAGGCATCGAAAATTGAAGAGTGGTTTTCTCCTTGTGCAA
TTGCGACAAGGTTTGATATTTCCAGAAGTGCACCATGGAAAATAATTCTGACAGCATCTGCTCCATTTCGATTTTTTTACACCAA
AATATATTTATCATCTGAAGGTGAATTGCCATCATATAGTGCACCGATACCGCCACAAGCTCTGTTGACCCCTCAGATCCTGGCA

ACTGCGTTTTTTCCAGAGGTAAAGAAATGTTTAAACTGGGACCTGGAACACCACATCGTATTTTTCCTCCCTGTTACCGCGCATT
CTAGTTGGGTGATGACAAGTAGCCCGCCGAGCACTGGAATGACCTACAGACTTCCGGATACCTCAACAGCTGCCCCTTTGTTTGT
AACGTTTGTGAATTCAAGTTGCAACTCCCTACCGCAAATCGCGCGAACAGCAAAAATGGAGGATGGTCATTTGAATGGACGAGCG
AAAACTTGCATATATTGCAACCACCTGATAGTTCAATACGGTCTACGGGGGTTCCAGGATGCTTTCTTTATAGACAGTGCATATG
TTCAATCAGAACTGTTAGCCGGTGCCAACTCATCTATTGCATTTTTTAAGTCTATGAACCCCGAGTTCTATAACTACGCGCTTCT
AGGACCACACGGAACTGCTGTCCGAATTTTAGGACTTGACAGAAGCATAATAGTTTTTGACAAGGCATACATTATCGGTATTGTG
GTAGGGGCACTTGTTGGATTGGTTGCTCTATGGAGTGCCATCAGGGCTATATATAAATTCTGCCTCAGCTCGGCAACGCAATTTA
GACGTGTACGGCTACCGTCTGACGAGGAGAGCGAACTTTAATCCTTGTTCCTTACAAACCCTTTTCAACTGCACTTGCGGCAGCA
CGAATAACTGGCTATGGCTCCGTCAAAGAAGCAGTCTCAGGCTTCCAAAGCACTGTCTACCAGCAATCGCGAATGTCGTTTTGCA
GTAACTCACTGCCTCGCAAACTTTGTAGACTCGTCTCACATGGAATTAGGAACACATCGTGGAAAACTATGGACAGAAAAGGTAG
AATATGGCCTAGCTGTACTGCACGGCACAGCATCGTTTTTTCTGGATTGCAACAAGGATTTGACTATTCATGTGTTTCAAGACCA
TGACTCAAAACAGCCCGATAACTCTCAAATCCGATTCGTAGCAGCACCATTATTTGAGGATCCGCAAGAATCCCTTTCCAACGAC
ACTGGAATTCCTTCTACTGCAATTATCATGACTAGAACTCCCTACACAACTAATTCTAATTTAGAGATTGGCGTCGAGCAGCTCC
TTCTTGCCATGAAAAGAGCTTCCTTAAAAAATTTCCAAAGATGGCAAGCCGTAGCTGATAGCGTGCATTCAAAAAACCCGATGAC
TTTTGGAGCCAAGTATTTGCCTTTTAAACTGTTGTCTGCAGACAAGTGGAAAGATGGAGTCGGATTTCTAATGTTAGGACTTATC
TCGGATTTGGAACTTTCTCCGGTTGTTGCTGCTGTACGCGCATTCAGCGGGGCCTTGTTCTTTCAAAGTCCTCCATCACTTACTC
TAAGCGGAGAAATCCTTGATTTAAAGCACGGAGACAGCCCGCATCACTCGCCCGTCGGACTACTACAAGCGGCGATATTTACTGG
GTTACGCGCGTATCGATCACAGATAGCTTCGCCGATAATTACTTGCCGCGAAGAATGGGCTTTCCGTATCCCGTTTTGTGTTGCG
ATTGGCGACAGCCAAATACATCCACCGCGCGCACAGCAGACAGCCTCGATTCCGAAAGGAACTCACATTCTTTGCGCGTTTTTCC
CAGTTTCGGGGGTATTGGCATCGGACGAACGCTCATATCTAGCGGTAGATACATGGGCATCAAAGAGCTACTAGCTGGCTTACTGG
GCGAGGCTTTACCTTCTGGCGCGCTCTTTTAGAAGAAACTCCCATGTGCGGCAAGTTTCTGTGCGCAGCAATTTGCCACGCGTAC
GATCCACCTAGACTCGGCATTTTATTGCACTCTCATTGGACCCCCGATGAGATAGGCGATTTTAAGACGGCTGCTTGCACAGCAG
TGTCCGAAGCATGGAAGAACATGGGAATTGCGACGCGGCCTGATCTGTTTGGATAATTACCCGAAACATAAACATCAATAAACGC
AAATACATTTCTGACAAACTTTTATTTCGTATGTGTGTTCCATTAAAT

>UL24

GAAGGATTATGGAAGCGGAACATTACGAACCCATATCAGCATTGTAGCGCTCTACTGCGTCAATTAAGCTCGCGATCGCCGCTCG
TTTGGTTGTAAGAGAACTTGCCGTCTCGGCAAACTCTAAACATTTCAGTGCACATGCCTCCGGGGTGAGGCCATGTGCTGGTAAG
TAAAAAATGCACAAGTTTTTCAATACCGTCGCGAGTTTTTGAACCACAAATGAAAGTACAGTTTTTAATCTTTCCCCGTGAACAT
TTTGCTGATCAAAAGCTAATATTGCAAATAGCGTCTGGTCGATTGAAGGATTGATTTCTGGAACAGTAGCAGTTTCTAAGCACTC
TCTCTTGAATTTTAAGAGCGTGGCACTTGTTAAGGCTTCTTGCTCCTTCAACAATGCTGGATAAGATGTACGTTGGAGGTAGGTG
GTAGTATTCATGAGGCAGATGTACGAGTTACGTAATGCACACAAATACGTCCAGTCGACCGTTTCCCCTGGGCGCCCCCTTTGTA
TTATACGCCGTGCCTGTTCGCAAATATCCAAATCCATTAAAACAATGTTGGAGTGTGTTGGAATGTCAGGAATTCCTGACAATAA
AGTTATTAAATCTCCAAATTCGAGCGCTCCGCTTAGATATTGTGCAACAGGGAAACACAAACATGCCGCGAGTGGGTGCCTATCT
ACGAGGAATAAGACATCATCTGGTCCGGATGCAGCTGCGCTGTAGCAGCGTTCAGATCCCAAATAATTATAAATGACGTGGTTCA
TTATTCTTAGTGGGGCAGCAATAGTCATCTGAACTTCCGCTACGGCAACCGGAGACGCCTCCCGACATTCAAGTTTTCCGGAAGC
CTTTAGGGTTTGTATTTCATTAATTTCTTTGATCATATCCGCTCCGCCAAACCATTGGCGCCATGCTTGCATAGGTTCCGCTAAG
TATAGACGGCAACCTCTCCAATTTGGCATTTGCATCAACGTCTTTGCAGTGACTGTTTTACCGACTCCGAATGGCCCGTCGACGT
AAAGAAGTACAACGCGGCGAAAGTCGGAGTCATCGTTCTCGAACTCGGAGCAAAACTGCACACCACCGGAAGTTTTCACGGCGCC
AGCTACGGCCATCTCTGCGGGCTAGACTAAATGCCGGAATCCGTTGCCATAATCGCTTTTACCGAGCTCTCGTCAGATCACTGGA
GGAAGTCTTTGAAGGGGGTGGAGACGGAAGGCTGGCCTATACGATAATTCCCCAATGCAAACCTGCCGGGGGCAAAATTGTAGTC
ATGTTTGAAGTGAACCTAGGGTTGCGGAAACCTGATTGCATTTGTTTGCTGGAGACACAACATGAAATGAAATGTATTGTGATTG
AACTGAAAACATGTCGTTTTTCCAAATCTTTGATGACTGAAAGTAAATTAAGACAAGGATATACTGGCACTCTCCAACTTCGCGA
TTCTGCACGATTGTTGGAGAATCTGGCTGTTCCTGGCACTGAGAAAGTAAAGATTCTTTCTCTACTGGTATTTGTTGCTCAACGG
GGGATGAACATCTTGGCGGTAAAAACGGTTGGGGAAACGGTTATAAATGTATCGTCTGAGCTTTTCTTTGTCACTCTAGCTACCA
GAAGCCAGTATTTAAAGACTTTCTGCGCGAAATTAGAGCCACGCGTGTCGCATGCGCGCAGTAAGTATCAACAAGAAAGTGCAAA
GAATGCGGATCTAGCGTCTCCCGCCCCTTCAGCCTTACAAACCGTAGCTGCATTGTTTTCGTCGCGCGTGGGAAAAGAGTCCGCA
ACCAAACCCGCGAGCTATTCCACTTCTACGGAGGAGTCCAAAAACTTGTCCGAACCATGTTTCGACCCAGATTCGAACCTATGAA
TTTACCAAAGGACAGCAACAAGCCGTCTACCTTGATGGTTTTAGCCGATCGTTTAAACTTTATCTCCTGCGCGGAAGGCTCGAGT
AAATATGCATCTAAGCTTTTCGAAGGAACTCTAATAGACGCAGAAATCATGACCAATCGTGCAAGGATTGAAGATCTGGAAAGGC
GGAACAGGGCTGCCAAGGCAGCTTTAGAGCAACTGGAAAATATGAGCGCTACTGTTCCCGTACATGTCTCATCGGCATTGCAAAC
TATTGAATATCCCCTAGAGACTGTAATTGATGTCTTGGATGATTTAGCCCAGCGGGCCGTGCAGGAGAAGGACATTGTTGGGTCT
TATAAAACACTAGACATCCGTGCGCCTGGCGAGGATGTCCCTGCGAACGTAATTTGGATAGTTAAGAATGGAGAACCATTAACTT
TTAATACAGATTTTCAAGTAGATTTTCTGACAACTTCGTTTGCTATTGCTGGCAATGGCCGGCTGGGATTCGGGTCTTGGTTTCG
CGCGCTACAGACTCAGCTCCTAGATAACAATAAAGCTATCGCTAGGGTTTTGAATGTAATGGGAGACACGCGCATCTCGGGGCGG
TTTATGAAAACGGCTATTCGAGCGCTGCGATCGGCGATGGAAATTTATGCTGGGACTCGACAATATAGCGGATTCGAGGCTACTG
TTTTATGCTTACTGCACTACTCGCGGTCTAGACAATCTGCTTCAAACATTAGACATGGTCTTGATGTTTCTATATTTGAAGATGC
ATTAAGACATGTTCCAACGTATTTAAATTATATGCTGGAAGATATTCGGGCCGAGTGGGGCTCAGTTACATTTTCATTTGACCGG
TCAAAACTCCCCGTAAACTTTTTTTCCCCAATTGACGGCAGAAAATATTCGAATGGAGTGTTCGATCCTCACATAGTGTATCAAT
TGCTAAAGCGTACTGGGACCCTTTCCACCACAGTACGAGATATAACGAAAGAAACTCTGCTTCCAATAGATCCAGATTTTGTTCG
TTTTGATGACCCCATTGCTGCTCTCTCCATTTCATTTTTCCCTTCGAGACGAACGCCGTTAATTCTTCATGAAGATGATCCGCTT
GTGCGAACGGTTATAGACTCGATATCTCTGCTCCTAGTCTTGCAAAAGCTAATGTTTAATAGTAATGTCTATACGAGCACACATC
TCAATCGATTTCAGCCATCTGCGTTCTTTGAATTGCCACTTGGAACCCAATCGGAACAGGAGGCAGCTAAATGGCCCGTCGCCCC
TGGGTCAAGACCTCAAGCCACCGCAAGCACATTTGATGACAATGGACAAGATATGGCCAGTCGCGACAACAACTTGTTCTTTTTA
TTCGAGAAATATGTTGTACCTATGTATAGATATGACAACAGATGTGAAGTTACTGGTTTCTTTCCTGGTTTAGCTGCACTCTGCA
TAACGGGGCGCGTTAAAGGAATACCGACAGCGGTACGGTTGGGGGAGTACTATTCCTCCCTCTGTAATCTAATTGAGCTTGATCT
CAGAAAAACGTCACATGTGGGTAGCGGGGCAGCTGCTGTTTTGGCGGTACACGATTCATTGACAGGTGACGTTGAAGAGGGAGTG
TCCAGGTTACTTGAGGTATTTGATGCAAAAAAAGCATTTGAAAGGAATATTGAGGACATTCAACGTGGAGTCTGATTCTGACCTG
ATCTATTTTATGTGTCTTGGATGCTTGCCACATCACGTGACATATGGATAATGCTTAGTATTTTCGCGATCGCGTCTCTGTTTAG
ACAATAAAAGGGTTATATCTTTCTGATCAGTCCGTCTGTTTTGTCAGTGTGTTGATAC

>UL25

TTCCGGAAGCCTTTAGGGTTTGTATTTCATTAATTTCTTTGATCATATCCGCTCCGCCAAACCATTGGCGCCATGCTTGCATAGG
TTCCGCTAAGTATAGACGGCAACCTCTCCAATTTGGCATTTGCATCAACGTCTTTGCAGTGACTGTTTTACCGACTCCGAATGGC

CCGTCGACGTAAAGAAGTACAACGCGGCGAAAGTCGGAGTCATCGTTCTCGAACTCGGAGCAAAACTGCACACCACCGGAAGTTT
TCACGGCGCCAGCTACGGCCATCTCTGCGGGCTAGACTAAATGCCGGAATCCGTTGCCATAATCGCTTTTACCGAGCTCTCGTCA
GATCACTGGAGGAAGTCTTTGAAGGGGGTGGAGACGGAAGGCTGGCCTATACGATAATTCCCCAATGCAAACCTGCCGGGGGCAA
AATTGTAGTCATGTTTGAAGTGAACCTAGGGTTGCGGAAACCTGATTGCATTTGTTTGCTGGAGACACAACATGAAATGAAATGT
ATTGTGATTGAACTGAAAACATGTCGTTTTTCCAAATCTTTGATGACTGAAAGTAAATTAAGACAAGGATATACTGGCACTCTCC
AACTTCGCGATTCTGCACGATTGTTGGAGAATCTGGCTGTTCCTGGCACTGAGAAAGTAAAGATTCTTTCTCTACTGGTATTTGT
TGCTCAACGGGGGATGAACATCTTGGCGGTAAAAACGGTTGGGGAAACGGTTATAAATGTATCGTCTGAGCTTTTCTTTGTCACT
CTAGCTACCAGAAGCCAGTATTTAAAGACTTTCTGCGCGAAATTAGAGCCACGCGTGTCGCATGCGCGCAGTAAGTATCAACAAG
AAAGTGCAAAGAATGCGGATCTAGCGTCTCCCGCCCCTTCAGCCTTACAAACCGTAGCTGCATTGTTTTCGTCGCGCGTGGGAAA
AGAGTCCGCAACCAAACCCGCGAGCTATTCCACTTCTACGGAGGAGTCCAAAAACTTGTCCGAACCATGTTTCGACCCAGATTCG
AACCTATGAATTTACCAAAGGACAGCAACAAGCCGTCTACCTTGATGGTTTTAGCCGATCGTTTAAACTTTATCTCCTGCGCGGA
AGGCTCGAGTAAATATGCATCTAAGCTTTTCGAAGGAACTCTAATAGACGCAGAAATCATGACCAATCGTGCAAGGATTGAAGAT
CTGGAAAGGCGGAACAGGGCTGCCAAGGCAGCTTTAGAGCAACTGGAAAATATGAGCGCTACTGTTCCCGTACATGTCTCATCGG
CATTGCAAACTATTGAATATCCCCTAGAGACTGTAATTGATGTCTTGGATGATTTAGCCCAGCGGGCCGTGCAGGAGAAGGACAT
TGTTGGGTCTTATAAAACACTAGACATCCGTGCGCCTGGCGAGGATGTCCCTGCGAACGTAATTTGGATAGTTAAGAATGGAGAA
CCATTAACTTTTAATACAGATTTTCAAGTAGATTTTCTGACAACTTCGTTTGCTATTGCTGGCAATGGCCGGCTGGGATTCGGGT
CTTGGTTTCGCGCGCTACAGACTCAGCTCCTAGATAACAATAAAGCTATCGCTAGGGTTTTGAATGTAATGGGAGACACGCGCAT
CTCGGGGCGGTTTATGAAAACGGCTATTCGAGCGCTGCGATCGGCGATGGAAATTTATGCTGGGACTCGACAATATAGCGGATTC
GAGGCTACTGTTTTATGCTTACTGCACTACTCGCGGTCTAGACAATCTGCTTCAAACATTAGACATGGTCTTGATGTTTCTATAT
TTGAAGATGCATTAAGACATGTTCCAACGTATTTAAATTATATGCTGGAAGATATTCGGGCCGAGTGGGGCTCAGTTACATTTTC
ATTTGACCGGTCAAAACTCCCCGTAAACTTTTTTTCCCCAATTGACGGCAGAAAATATTCGAATGGAGTGTTCGATCCTCACATA
GTGTATCAATTGCTAAAGCGTACTGGGACCCTTTCCACCACAGTACGAGATATAACGAAAGAAACTCTGCTTCCAATAGATCCAG
ATTTTGTTCGTTTTGATGACCCCATTGCTGCTCTCTCCATTTCATTTTTCCCTTCGAGACGAACGCCGTTAATTCTTCATGAAGA
TGATCCGCTTGTGCGAACGGTTATAGACTCGATATCTCTGCTCCTAGTCTTGCAAAAGCTAATGTTTAATAGTAATGTCTATACG
AGCACACATCTCAATCGATTTCAGCCATCTGCGTTCTTTGAATTGCCACTTGGAACCCAATCGGAACAGGAGGCAGCTAAATGGC
CCGTCGCCCCTGGGTCAAGACCTCAAGCCACCGCAAGCACATTTGATGACAATGGACAAGATATGGCCAGTCGCGACAACAACTT
GTTCTTTTTATTCGAGAAATATGTTGTACCTATGTATAGATATGACAACAGATGTGAAGTTACTGGTTTCTTTCCTGGTTTAGCT
GCACTCTGCATAACGGGGCGCGTTAAAGGAATACCGACAGCGGTACGGTTGGGGGAGTACTATTCCTCCCTCTGTAATCTAATTG
AGCTTGATCTCAGAAAAACGTCACATGTGGGTAGCGGGGCAGCTGCTGTTTTGGCGGTACACGATTCATTGACAGGTGACGTTGA
AGAGGGAGTGTCCAGGTTACTTGAGGTATTTGATGCAAAAAAAGCATTTGAAAGGAATATTGAGGACATTCAACGTGGAGTCTGA
TTCTGACCTGATCTATTTTATGTGTCTTGGATGCTTGCCACATCACGTGACATATGGATAATGCTTAGTATTTTCGCGATCGCGT
CTCTGTTTAGACAATAAAAGGGTTATATCTTTCTGATCAGTCCGTCTGTTTTGTCAGTGTGTTGATAC

>UL26

CCGTAAACTTTTTTTCCCCAATTGACGGCAGAAAATATTCGAATGGAGTGTTCGATCCTCACATAGTGTATCAATTGCTAAAGCG
TACTGGGACCCTTTCCACCACAGTACGAGATATAACGAAAGAAACTCTGCTTCCAATAGATCCAGATTTTGTTCGTTTTGATGAC
CCCATTGCTGCTCTCTCCATTTCATTTTTCCCTTCGAGACGAACGCCGTTAATTCTTCATGAAGATGATCCGCTTGTGCGAACGG
TTATAGACTCGATATCTCTGCTCCTAGTCTTGCAAAAGCTAATGTTTAATAGTAATGTCTATACGAGCACACATCTCAATCGATT
TCAGCCATCTGCGTTCTTTGAATTGCCACTTGGAACCCAATCGGAACAGGAGGCAGCTAAATGGCCCGTCGCCCCTGGGTCAAGA
CCTCAAGCCACCGCAAGCACATTTGATGACAATGGACAAGATATGGCCAGTCGCGACAACAACTTGTTCTTTTTATTCGAGAAAT
ATGTTGTACCTATGTATAGATATGACAACAGATGTGAAGTTACTGGTTTCTTTCCTGGTTTAGCTGCACTCTGCATAACGGGGCG
CGTTAAAGGAATACCGACAGCGGTACGGTTGGGGGAGTACTATTCCTCCCTCTGTAATCTAATTGAGCTTGATCTCAGAAAAACG
TCACATGTGGGTAGCGGGGCAGCTGCTGTTTTGGCGGTACACGATTCATTGACAGGTGACGTTGAAGAGGGAGTGTCCAGGTTAC
TTGAGGTATTTGATGCAAAAAAAGCATTTGAAAGGAATATTGAGGACATTCAACGTGGAGTCTGATTCTGACCTGATCTATTTTA
TGTGTCTTGGATGCTTGCCACATCACGTGACATATGGATAATGCTTAGTATTTTCGCGATCGCGTCTCTGTTTAGACAATAAAAG
GGTTATATCTTTCTGATCAGTCCGTCTGTTTTGTCAGTGTGTTGATACTTTGCGGGTCGAGTCGGTCTCGCTATGTCGGAGAACG
TCGACATCAAATATATCTTCGTTGCCGGCTATTTGGTAGTATATGACCACCAAGAAAGTGCCGGGCGAGAATATGAGCTTACTCG
CGAGCAAAGCAAATCGGCTTTACCAGTTTTGCCTGGAACAATTCCGATAAACATTGACCACGAGTCATCCTGCGTGGTAGGTACA
GTGCTGACGATCTTAGACCTGCCGAGAGGACTGTTTTGTTTGGGTGTCGTGTCAACTGCTCTTGCACCTATTTTTTTAAGCTACG
TGCAAGATGATGCTCTTTTTGCGAATGCCGAAGAGGGGATGGTTTTGACGGAGACAGAGAAATTTCTCTATCTTCTCAGCAACAT
TCTACCTTCTCTATCGCTCTCTTCAAGGCGGCTTGAAAAGAACGAAGTGCCCGGAAAAGATTTTTTTGCACACGTGGCGCTCTGT
GAGTTGGGTAGACGTGAAGGCACGGTAGCAATTTACGGGGCCACCGCTTCGGAAGCTATTGGAGCATTTGATGATTTGAGCGCAC
CAATTAAAGAACAGCTGTATGAAATAGCTACACGTGAGAAATGTGCAGAAGTCCCTAGAGAGTTGTCGCGACCAGAAATTACGCG
AGTCCTGATGAAGAAATTTATTCACGGGGCGTTTTTGATGGATAGGGGTACCTGTTTGAAAACGCGACGCGAGATGGCAGCTGTT
TACAACCCGAAATATTTACAAGCGAATGAGGTAATAACCATAGGGATTAAGGAACACTCCGAAGAGACTCCTGAGAATGCTATTA
AAGACCGTTCAGTTTCTACGCAAACGGCCCCAAGCTTCGATATAAGCGAGAGTCAGCAACCGTCCGGGCAAACACACGTTCCAGC
CATGGAGTCGGCGACATGCTCCGGCCAGTTCTTGCAAACAAAAAACGGAGCTTCACCGTCCGCATCTCGTGAAGATATGGTCTAT
GTCCCATTCGAGAAGTATGCGAGTCTTCTTGCGGCATCAGCCCGCAGAGATAACGACCGGAGACCCGTGTCGCCATCACGAGAGT
TTTCCCGTAGGTCGCGGGACTCTACTCACGAATGTTCACCAGGACGTGATATTTGGCCGCGTGGATTCGAACGACACCCTCGCCT
AGAATCTTTTATGGGTCCGGGGATGAACCACACTTACAGACCGGCTCTTTACGAAGACCCAAATTTCTGCGGGCGTTTCCCATAT
ATTCCTTATCAAAGCCCCGCATCTACTTACCCCGTGCATCCAAATTATTATAGTTCGAACTTCGGCCAGTTTCCTGGGGCTGGAA
CGTATCCTATACAGTATCCATCATTACATGAGCAGACTGTTGTTTCACGCTTAGATGCCTTGATTTCGGCTCTTGAGAAAAATAA
TAAACGGGATAGCGAATACTCGGAGAATAACCCTCGGAAGAGATCCGCTAGGACCATTTCTGAGAATGACCCGTACTTCCCTGGG
GAAATGGTTCCAGCAAAAAAGATAACTACGGAGCAGCAGCTATGTGAAAAAAAGGAACCCGTTGGTAGCGGAATTAATGATATTT
TACAAGGAATTCTGACCCTTCAAAAAGAGGTTGCCGGTCTTAAATCAGCTTCCAATGCCGATTCGTCTTCAGAGAGAAAAGAGGA
ACTAGAAAATAGTAATCAAGAGTCTGCTCGGGAGACAGTCGATGCTTCTATGCCAAAACGACTGAAAGATGCTCAGACAAAATTG
AAAAGAAAGAAAGAAGCAGCCGCATTTGCGCAAATGATGGCAGACTGAGTCATGTGAATAGTAATGTCATTTTATTAAATAAACT
GCATTTAAACTTATTCGTCTTCGCTTTCTTCTGCCCACGAAGGGTATT

>UL26.5

CATCACGTGACATATGGATAATGCTTAGTATTTTCGCGATCGCGTCTCTGTTTAGACAATAAAAGGGTTATATCTTTCTGATCAG
TCCGTCTGTTTTGTCAGTGTGTTGATACTTTGCGGGTCGAGTCGGTCTCGCTATGTCGGAGAACGTCGACATCAAATATATCTTC
GTTGCCGGCTATTTGGTAGTATATGACCACCAAGAAAGTGCCGGGCGAGAATATGAGCTTACTCGCGAGCAAAGCAAATCGGCTT
TACCAGTTTTGCCTGGAACAATTCCGATAAACATTGACCACGAGTCATCCTGCGTGGTAGGTACAGTGCTGACGATCTTAGACCT

GCCGAGAGGACTGTTTTGTTTGGGTGTCGTGTCAACTGCTCTTGCACCTATTTTTTTAAGCTACGTGCAAGATGATGCTCTTTTT
GCGAATGCCGAAGAGGGGATGGTTTTGACGGAGACAGAGAAATTTCTCTATCTTCTCAGCAACATTCTACCTTCTCTATCGCTCT
CTTCAAGGCGGCTTGAAAAGAACGAAGTGCCCGGAAAAGATTTTTTTGCACACGTGGCGCTCTGTGAGTTGGGTAGACGTGAAGG
CACGGTAGCAATTTACGGGGCCACCGCTTCGGAAGCTATTGGAGCATTTGATGATTTGAGCGCACCAATTAAAGAACAGCTGTAT
GAAATAGCTACACGTGAGAAATGTGCAGAAGTCCCTAGAGAGTTGTCGCGACCAGAAATTACGCGAGTCCTGATGAAGAAATTTA
TTCACGGGGCGTTTTTGATGGATAGGGGTACCTGTTTGAAAACGCGACGCGAGATGGCAGCTGTTTACAACCCGAAATATTTACA
AGCGAATGAGGTAATAACCATAGGGATTAAGGAACACTCCGAAGAGACTCCTGAGAATGCTATTAAAGACCGTTCAGTTTCTACG
CAAACGGCCCCAAGCTTCGATATAAGCGAGAGTCAGCAACCGTCCGGGCAAACACACGTTCCAGCCATGGAGTCGGCGACATGCT
CCGGCCAGTTCTTGCAAACAAAAAACGGAGCTTCACCGTCCGCATCTCGTGAAGATATGGTCTATGTCCCATTCGAGAAGTATGC
GAGTCTTCTTGCGGCATCAGCCCGCAGAGATAACGACCGGAGACCCGTGTCGCCATCACGAGAGTTTTCCCGTAGGTCGCGGGAC
TCTACTCACGAATGTTCACCAGGACGTGATATTTGGCCGCGTGGATTCGAACGACACCCTCGCCTAGAATCTTTTATGGGTCCGG
GGATGAACCACACTTACAGACCGGCTCTTTACGAAGACCCAAATTTCTGCGGGCGTTTCCCATATATTCCTTATCAAAGCCCCGC
ATCTACTTACCCCGTGCATCCAAATTATTATAGTTCGAACTTCGGCCAGTTTCCTGGGGCTGGAACGTATCCTATACAGTATCCA
TCATTACATGAGCAGACTGTTGTTTCACGCTTAGATGCCTTGATTTCGGCTCTTGAGAAAAATAATAAACGGGATAGCGAATACT
CGGAGAATAACCCTCGGAAGAGATCCGCTAGGACCATTTCTGAGAATGACCCGTACTTCCCTGGGGAAATGGTTCCAGCAAAAAA
GATAACTACGGAGCAGCAGCTATGTGAAAAAAAGGAACCCGTTGGTAGCGGAATTAATGATATTTTACAAGGAATTCTGACCCTT
CAAAAAGAGGTTGCCGGTCTTAAATCAGCTTCCAATGCCGATTCGTCTTCAGAGAGAAAAGAGGAACTAGAAAATAGTAATCAAG
AGTCTGCTCGGGAGACAGTCGATGCTTCTATGCCAAAACGACTGAAAGATGCTCAGACAAAATTGAAAAGAAAGAAAGAAGCAGC
CGCATTTGCGCAAATGATGGCAGACTGAGTCATGTGAATAGTAATGTCATTTTATTAAATAAACTGCATTTAAACTTATTCGTCT
TCGCTTTCTTCTGCCCACGAAGGGTATT

>UL27

AGAAACCCAGAGAAAGTAGATACAGATGCACTGGAGGAATTAGTGGAAAAACTGCGTGGCTCAGACAAGGAGACACAACCTGTCA
CAGAAGGAACTAGTACAAGCGACCTGACTGCTAACCTCAGAGGAACTATTGAAAGGGACGAGGATTCGATAGAAGTTCTTCTAAA
AAAGGCAGAGCGAGATCTTGATGTCAGACAAAAGAACTACGTCGAAAAGCTTTCTGCACGATCATTTTCAAACTTGGATCGATGC
ATTAAAATTCAGCGAGCCGAGCTAGAAAAATTAATGCGTGTGAATATTTACGGAGAAACTCTTCCGGTAATGTACGTAACGCTAA
AGAATGCGTTTCTAGCAAGACGCGCATTTCTGGATATTGTTTGCGCTGAAGAGTCTCTACATTCGAAAATTCTACGCATCGATCA
AAATGATGGGCGTGCTTATGATTGGCATCAATACATACGCTCTTCCATCACCAGATACCAAATCGACCAAACCATGTTACCGCGC
CTCACCAATCGGTTTTTTGAAATTTTAAGCGGCCCCCTTTTCAAACATCACAAGGAGAGATTTCCCCAACCCCCAAACACGTCTA
TATACTTTTCTGTAGAGAACGTAGGACTTCTTCCTCATTTGAAAGAAGAGTTGGCCAACTTTACAAAACGTCACAGATCATGTGA
TTGGATGACAAGTGAGTTTCGCCGATTCTATAACTTCGCGGGCATTTCTAATTCAACTGCTATTCAGAGAGCCGCGTGCGGATAT
ATCAGAGAGGCACTCTTTGCTACAACGGCTTTCGAGAAAATTTTCCACAGCGGTAACACTAAGTTACTGCGAGCAGATTGTGTCG
AATTCGACCTCGACGGCCCGATTTTGGAAAACGGGCTATATTTAACATTTGAAGAGGAACATCCATTAGTTGCAGTGTGGGGCGT
CGATGACCGCGGGAAACTGGGTCCAGCGTCTACTATCATAGTAGAGAAAGATTTATACGCGGTCCTGTATGCAATCCTACATCGC
CGTGAACATTGACATGGCTAGCTTGAAAATGCTGATCTGCGTGTGCGTGGCAATCCTGATCCCATCTACCCTATCTCAAGATTCA
CACGGAATTGCTGGAATAATAGACCCTCGTGATACAGCCAGCATGGATGTTGGAAAAATCTCTTTCTCCGAAGCCATTGGGTCGG
GGGCACCGAAAGAACCCCAGATTAGAAACAGAATTTTTGCGTGCTCATCTCCAACTGGCGCCAGTGTTGCGAGGCTTGCCCAGCC
ACGACATTGTCACCGACATGCCGATTCGACTAACATGACTGAAGGAATTGCCGTAGTCTTCAAGCAAAACATTGCCCCGTACGTC
TTTAATGTGACTCTATACTATAAACATATAACCACAGTTACTACGTGGGCATTATTCTCAAGACCCCAAATAACAAATGAGTACG
TGACCAGGGTTCCAATAGACTATCATGAAATTGTCAGGATTGATCGATCGGGAGAATGCTCATCCAAAGCAACGTATCATAAAAA
TTTCATGTTTTTTGAAGCTTACGACAATGATGAAGCAGAAAAAAAATTGCCCCTGGTTCCATCACTGTTAAGATCAACTGTCTCC
AAGGCGTTTCATACAACTAACTTTACTAAGCGACATCAAACCCTGGGATACCGAACGTCTACATCGGTCGACTGTGTTGTGGAAT
ATCTACAGGCTAGATCTGTATACCCGTATGATTACTTTGGAATGGCGACAGGTGATACAGTAGAAATTTCTCCTTTTTATACCAA
AAACACGACCGGACCAAGGCGTCACAGTGTCTACAGAGACTATAGATTTCTCGAAATCGCAAATTATCAAGTCAGGGATTTGGAA
ACCGGACAAATAAGACCCCCTAAAAAAAGAAACTTTCTAACAGATGAACAATTCACTATAGGCTGGGATGCAATGGAAGAAAAGG
AATCTGTATGTACTCTCAGTAAATGGATTGAAGTCCCGGAAGCAGTTCGTGTTTCGTACAAAAACAGTTACCACTTTTCACTTAA
AGATATGACTATGACGTTCTCGTCCGGAAAACAACCTTTTAACATCAGCAGGCTTCATTTGGCTGAATGCGTTCCTACCATAGCC
ACGGAGGCCATAGATGGCATCTTTGCCAGAAAGTATAGTTCGACTCATGTCCGTTCTGGGGACATCGAATACTATCTCGGTAGTG
GCGGATTTCTGATCGCATTTCAGAAACTCATGAGCCATGGCTTGGCTGAAATGTACCTAGAAGAGGCACAAAGACAAAATCATCT
CCCGAGAGGGAGAGAGCGTCGCCAAGCCGCAGGTCGCCGCACGGCGTCGCTGCAGTCTGGACCTCAGGGTGATAGAATTACTACC
CACAGTTCTGCAACATTTGCCATGTTACAATTTGCATACGACAAAATCCAAGCCCATGTTAACGAGCTTATCGGAAATTTGTTGG
AAGCGTGGTGTGAGCTTCAGAACCGCCAACTGATTGTATGGCATGAGATGAAGAAACTAAACCCGAACTCACTGATGACATCTTT
GTTCGGACAACCTGTAAGCGCCAGGCTATTGGGAGACATCGTAGCGGTATCAAAATGTATAGAAATTCCAATCGAAAATATTAGG
ATGCAGGATTCCATGCGCATGCCAGGGGACCCAACCATGTGCTATACCAGACCAGTACTTATTTTCAGGTATTCGTCCTCCCCTG
AGTCACAGTTTTCTGCGAACTCAACAGAAAACCACAATCTTGACATATTAGGCCAACTCGGAGAACATAATGAAATTTTACAAGG
GCGGAATTTGATAGAACCATGCATGATCAATCACAGACGGTACTTTCTGTTGGGAGAAAACTACCTTCTTTACGAAGACTATACA
TTTGTTAGACAAGTAAATGCTTCCGAGATCGAAGAAGTGAGCATATTCATCAACTTGAACGCCACTATACTAGAAGATTTGGACT
TTGTGCCCGTCGAAGTATACACTCGCGAGGAACTCAGAGATACTGGGACTTTAAACTATGATGATGTGGTCAGATATCAAAATAT
TTATAACAAAAGGTTCAGAGACATTGACACTGTAATACGTGGAGATAGGGGAGATGCAATCTTTAGAGCAATAGCAGATTTTTTT
GGCAACACTCTTGGAGAAGTAGGAAAGGCATTGGGAACTGTAGTGATGACAGCCGCGGCAGCAGTAATTTCTACAGTATCTGGCA
TCGCCTCATTTCTTTCTAACCCGTTCGCCGCACTCGGAATTGGGATAGCGGTGGTGGTGAGCATTATTTTAGGACTGCTGGCGTT
CAAATATGTAATGAACCTGAAATCAAACCCAGTTCAGGTTCTGTTCCCAGGCGCAGTTCCCCCGGCCGGAACTCCTCCACGACCC
TCTAGACGTTACTACAAGGATGAGGAGGAGGTTGAGGAGGATAGTGATGAGGACGACAGGATACTTGCCACCAGAGTTCTGAAAG
GCCTTGAGCTTCTACACAAGGATGAACAGAAAGCTCGAAGACAGAAAGCGCGGTTTTCTGCTTTTGCTAAAAATATGAGAAACCT
ATTTCGCAGAAAACCCCGAACCAAGGAAGATGACTACCCCCTGCTCGAATACCCTTCGTGGGCAGAAGAAAGCGAAGACGAATAA
GTTTAAATGCAGTTTATTTAATAAAATGACATTACTATTCACATGACTCAGTCTGCCATCATTTGCGCAAATGCG

>UL28

ATGGAACATTTGATACGCGGCAACCAGAGTCAAAAATGTTCCGTTCTTCTGCCATGCCGGCCCTAAGATCAGCCTTTAACGGGAT
GCTGGACAAAGGATTCCTTTCTGGAAAATATGAGCCATTTTCAATATCAGTTCCGTGTGTTAACGCTCCGGATACAATGCGCCCA
AACACAGAACAGGCAATCGCACAATATGAGTATTCTCTTTCTAGGGGACAAGTCCTGAAACTGAAAGAGTTTAAGGTGAAGAACC
GCGTCGTGTTCAACGGATCCAGTACGTCAAACTCGTGCCGCGTGGTATTGAGTTCTAAACCTATAGTAATGTTAGGGATTATGAT
CAGCAAATATGTAGGACAACAAAACTCGACCACCGTCTTTCAGTCCGGCAATTGGGGTAGCCTTATAGGCAATTCAGGGTCCCAA
TCGGTTAATTCCGCGCTGACGAATGATCCCGTCAGAAAGTTTGCACTTGCTTGTAAAAGGGTAGGAACTATTCTGTCATCTGGGC

AAGTGGCATCAGGGTTTCAAGAAAACACCATTGCCTCTCAAGTGAGAAGCCTAATTGATGCAGGAGGTAGTCCCACTTCCATCTA
CACAGCCGTTCTGAGAGTTCTGGGAGAAGGAATGAAAGACATTACTGCCGAAACATGGATGGCAATTACTGACGACAAGTATCTG
GTAAATATATTAGTAGAGCTCCGTGAGGACATTTCTGGAAGTCAGAATGGGTGGTCCGTGGCTACCGCACAAAGTATGCTCAGTG
AGCTAGAGGGGAAAGATACATGCCCGGAAAATGCAGGAGAGATGTTGGTCTGGGCAGACGAGGATGAATATACTAACACTGGCGA
CATCAGCGAAGATATTGAACTTGGCGCGCCAAGTGCAAAAAAACCTTCGCTATCGGCTGATATTCTGTTCTGTTAACAGTAACTG
CGCGTGCTCAGAATAAACAAATAAATATAAGAAAGTCCTTGTACGTATTTTGTTTCATTCTTACTGGGGAATATGGCTGAAGAAA
CACAGTTACGCCTTATGGCTCTGGCAGGTCAGCTACAGACGCTGTTATTTCAAATAGATATTTTGAAGCGCTGCGATCCAGAGAT
ATTGTTGCAGAAGCGACTGCGGTCGAGAATCAAACACAACGCGCTAATGGCATTGTACATTCATTCGCTCCTTTCTAATGAGCTC
TCTATGCAAGTAAATTATAGGTTAACCCTTCAGACATATTGCGTGTGGCTCTGGCTTCGAAGAGCGTGCTCCGAAATAGCAGCTT
TGTGTAATGAGATAGACAAGTATTCTGTCCTAAAGGATAGAAAATGTTTTTTTGAAGCTACGCTTGGGTACGGGAAAGGATTCAT
CTGTCCTTTGCATGCGGAGGTCCAGCTACAATTATATGGAATGAGTTCAATTATTTCAAAGGAACTCGTTCTGATCAACGATGTA
GAAAACTTTATGAAGCAACTAAACTACTGTCACTGCGTAGTATCTTGTAAAAACGCAATTGAAGCTCTAAATACAATAGATGAGT
TTTTGGTTCTCAGCTCGGGGGGAAGTCTCGTAGCAGCACCGGATACTTACGATCATGGTCAGCCGTGTTATGTGTGTTTAGAAGA
GCTTAGTGTAACCGCTAATCAGGGAGAAACAATTTATCGACGCCTTGGATACAAAATTTGTGATCACTTGACAAGACAGTTTCCG
GTCAATGTATCTACGGATGACCTTAAACGCCATCTTCCATTTTTAAGGGATGTGGATCCCGAAAAATTAAAAAATGTGCTTTCAA
ACGAGACTCGCATCGAAGCTGAAAACCACGTCGAAGCAGGACGCACCTTATCAAGTTCCCCTGGAGAGTCAGAAGCTTCCAAAAT
CTTAAACAAATATGATGTGTTTACGGAACCACCGGGGCCCGTCTATCAACTAAGTGAATTACAATACTGGCTAGCGTCGGGTAAA
AGGACCCCCAGCGGAGTAGAGTCACAAACTACTACTTCACAATCTGTGTTAAAGCAGTTAGACAAAGATCTGAGCGAACTGTTCA
AAAGAGCTGAAGACTTTGAACAGGAATGCATCCGTCTAGAAAGAACACTGTTTGAAAAGGTGCATGCACATTTTCACAGGGTCGC
GCAAGCTGAGTTAATATCCCCACAACATTCTTTAGTGGATACGCTGATCACAGGAGTACTGGCCACTTCACCGGATACCGAAATA
GCTGCCTTGATCAAAGCCTGCTATGACCACCATCTATCCCTCCCATTGTTTCGCAGATTAAGAAACCCAGAGAAAGTAGATACAG
ATGCACTGGAGGAATTAGTGGAAAAACTGCGTGGCTCAGACAAGGAGACACAACCTGTCACAGAAGGAACTAGTACAAGCGACCT
GACTGCTAACCTCAGAGGAACTATTGAAAGGGACGAGGATTCGATAGAAGTTCTTCTAAAAAAGGCAGAGCGAGATCTTGATGTC
AGACAAAAGAACTACGTCGAAAAGCTTTCTGCACGATCATTTTCAAACTTGGATCGATGCATTAAAATTCAGCGAGCCGAGCTAG
AAAAATTAATGCGTGTGAATATTTACGGAGAAACTCTTCCGGTAATGTACGTAACGCTAAAGAATGCGTTTCTAGCAAGACGCGC
ATTTCTGGATATTGTTTGCGCTGAAGAGTCTCTACATTCGAAAATTCTACGCATCGATCAAAATGATGGGCGTGCTTATGATTGG
CATCAATACATACGCTCTTCCATCACCAGATACCAAATCGACCAAACCATGTTACCGCGCCTCACCAATCGGTTTTTTGAAATTT
TAAGCGGCCCCCTTTTCAAACATCACAAGGAGAGATTTCCCCAACCCCCAAACACGTCTATATACTTTTCTGTAGAGAACGTAGG
ACTTCTTCCTCATTTGAAAGAAGAGTTGGCCAACTTTACAAAACGTCACAGATCATGTGATTGGATGACAAGTGAGTTTCGCCGA
TTCTATAACTTCGCGGGCATTTCTAATTCAACTGCTATTCAGAGAGCCGCGTGCGGATATATCAGAGAGGCACTCTTTGCTACAA
CGGCTTTCGAGAAAATTTTCCACAGCGGTAACACTAAGTTACTGCGAGCAGATTGTGTCGAATTCGACCTCGACGGCCCGATTTT
GGAAAACGGGCTATATTTAACATTTGAAGAGGAACATCCATTAGTTGCAGTGTGGGGCGTCGATGACCGCGGGAAACTGGGTCCA
GCGTCTACTATCATAGTAGAGAAAGATTTATACGCGGTCCTGTATGCAATCCTACATCGCCGTGAACATTGACATGGCTAGCTTG
AAAATGCTGATCTGCGTGTGCGTGGCAATCCTGATCCCATCTACCCTATCTCAAGATTCACACGGAATTGCTGGAATAATAGACC
CTCGTGATACAGCCAGCATGGATGTTGGAAAAATCTCTTTCTCCGAAGCCATTGGGTCGGGGGCACCGAAAGAACCCCAGATTAG
AAACAGAATTTTTGCGTGCTCATCTCCAACTGGCGCCAGTGTTGCGAGGCTTGCCCAGCCACGACATTGTCACCGACATGCCGAT
TCGACTAACATGACTGAAGGAATTGCCGTAGTCTTCAAGCAAAACATTGCCCCGTACGTCTTTAATGTGACTCTATACTATAAAC
ATATAACCACAGTTACTACGTGGGCATTATTCTCAAGACCCCAAATAACAAATGAGTACGTGACCAGGGTTCCAATAGACTATCA
TGAAATTGTCAGGATTGATCGATCGGGAGAATGCTCATCCAAAGCAACGTATCATAAAAATTTCATGTTTTTTGAAGCTTACGAC
AATGATGAAGCAGAAAAAAAATTGCCCCTGGTTCCATCACTGTTAAGATCAACTGTCTCCAAGGCGTTTCATACAACTAACTTTA
CTAAGCGACATCAAACCCTGGGATACCGAACGTCTACATCGGTCGACTGTGTTGTGGAATATCTACAGGCTAGATCTGTATACCC
GTATGATTACTTTGGAATGGCGACAGGTGATACAGTAGAAATTTCTCCTTTTTATACCAAAAACACGACCGGACCAAGGCGTCAC
AGTGTCTACAGAGACTATAGATTTCTCGAAATCGCAAATTATCAAGTCAGGGATTTGGAAACCGGACAAATAAGACCCCCTAAAA
AAAGAAACTTTCTAACAGATGAACAATTCACTATAGGCTGGGATGCAATGGAAGAAAAGGAATCTGTATGTACTCTCAGTAAATG
GATTGAAGTCCCGGAAGCAGTTCGTGTTTCGTACAAAAACAGTTACCACTTTTCACTTAAAGATATGACTATGACGTTCTCGTCC
GGAAAACAACCTTTTAACATCAGCAGGCTTCATTTGGCTGAATGCGTTCCTACCATAGCCACGGAGGCCATAGATGGCATCTTTG
CCAGAAAGTATAGTTCGACTCATGTCCGTTCTGGGGACATCGAATACTATCTCGGTAGTGGCGGATTTCTGATCGCATTTCAGAA
ACTCATGAGCCATGGCTTGGCTGAAATGTACCTAGAAGAGGCACAAAGACAAAATCATCTCCCGAGAGGGAGAGAGCGTCGCCAA
GCCGCAGGTCGCCGCACGGCGTCGCTGCAGTCTGGACCTCAGGGTGATAGAATTACTACCCACAGTTCTGCAACATTTGCCATGT
TACAATTTGCATACGACAAAATCCAAGCCCATGTTAACGAGCTTATCGGAAATTTGTTGGAAGCGTGGTGTGAGCTTCAGAACCG
CCAACTGATTGTATGGCATGAGATGAAGAAACTAAACCCGAACTCACTGATGACATCTTTGTTCGGACAACCTGTAAGCGCCAGG
CTATTGGGAGACATCGTAGCGGTATCAAAATGTATAGAAATTCCAATCGAAAATATTAGGATGCAGGATTCCATGCGCATGCCAG
GGGACCCAACCATGTGCTATACCAGACCAGTACTTATTTTCAGGTATTCGTCCTCCCCTGAGTCACAGTTTTCTGCGAACTCAAC
AGAAAACCACAATCTTGACATATTAGGCCAACTCGGAGAACATAATGAAATTTTACAAGGGCGGAATTTGATAGAACCATGCATG
ATCAATCACAGACGGTACTTTCTGTTGGGAGAAAACTACCTTCTTTACGAAGACTATACATTTGTTAGACAAGTAAATGCTTCCG
AGATCGAAGAAGTGAGCATATTCATCAACTTGAACGCCACTATACTAGAAGATTTGGACTTTGTGCCCGTCGAAGTATACACTCG
CGAGGAACTCAGAGATACTGGGACTTTAAACTATGATGATGTGGTCAGATATCAAAATATTTATAACAAAAGGTTCAGAGACATT
GACACTGTAATACGTGGAGATAGGGGAGATGCAATCTTTAGAGCAATAGCAGATTTTTTTGGCAACACTCTTGGAGAAGTAGGAA
AGGCATTGGGAACTGTAGTGATGACAGCCGCGGCAGCAGTAATTTCTACAGTATCTGGCATCGCCTCATTTCTTTCTAACCCGTT
CGCCGCACTCGGAATTGGGATAGCGGTGGTGGTGAGCATTATTTTAGGACTGCTGGCGTTCAAATATGTAATGAACCTGAAATCA
AACCCAGTTCAGGTTCTGTTCCCAGGCGCAGTTCCCCCGGCCGGAACTCCTCCACGACCCTCTAGACGTTACTACAAGGATGAGG
AGGAGGTTGAGGAGGATAGTGATGAGGACGACAGGATACTTGCCACCAGAGTTCTGAAAGGCCTTGAGCTTCTACACAAGGATGA
ACAGAAAGCTCGAAGACAGAAAGCGCGGTTTTCTGCTTTTGCTAAAAATATGAGAAACCTATTTCGCAGAAAACCCCGAACCAAG
GAAGATGACTACCCCCTGCTCGAATACCCTTCGTGGGCAGAAGAAAGCGAAGACGAATAAGTTAAATGCAGTTTATTTAATAAA
ATGACATTACTATTCACATGACTCAGTCTGCCATCATTTGCGCAAATGCGGCTGCTTCTTTCTTTCTTTT

>UL29

CTAAAGCGATACCACCCAAAAGAAGTCAGATTATTGTCGAGGATGAACCTGGTAATAGCGTCAACATTGCTTTCATATTTAGGTA
CGCTCGGGAAAAAATTATCACATATAAATTTCATCACTTTGCCATTACACACTTTTACACGGTAGAAGTCCTCCTCCTCAGATTC
AGAATAATAAATGTCTCTGCGCCGAACAACATCAATGCAAAAACTTTCCGCTGAAACAGCTAGAAAAGGTGAATTTTTTAAAGAT
GTCTTACGCGTCGCTAATGCTAGAGCATGCGCCAACTCTTCCGAGTTCGTGACTTTCGCGGCTTGGTCAATTTCTCGCTTTTTAA
CAAAAAAGTATGGCAAAGATCCATACACATGAACGGCGACGCGCTTTCCGCATGCAGAATTTCCAACTAAGGTTATTACAGTACC
GGTCGGTTTCAGTGTATCAAGAAATCTGGAATGCATCCATCGTACACATGACTGAGCGATCTCAGTGTTTTCAAAAAGGTCATAA
ACGTGAAATTCATAAAACTTTACATCTCCCTTGTTGTGGTGAGTCATTCCAGGACCAGCATTCCAATATTTCATCCTGACGGGCC

ACGCTCCATGTGTGCTTGTCAAGATGTCTAATTGTCGCTCGTTGTGGTAAAAGGTTGGTTTGCTTGTAGCCCTTCCAATCTGCGT
TCCTATCCTTCCGCCCCCATTAATGCACGTTGGACATACATACGGGAATTCTCTCACAGAAGTATAATAAGACGCGCCAGAGTGC
GGCCCGAGGCCAATTATCGGGGCTACACGAGCAGCTCCAGTACCGGCGGAGGAAAAAGTGTCCATAGTGCAGAGTGACAAACTCG
GTGCCCAGCTACTGCAATCATCGATATTTATGAACTCTCTAAAGTATACATAGGTGTGTCCTTTCCGCGGTACTTAAGTCCGTGC
TCAGTGACGGAGACTACTCTACTACGCGGAACTCTGATCTGAGATCATACTGTCTGAGATAACTCGCGCAAATGGAAAAGAGTTC
ATCTGGCACAACTAAGGGCGGAAAGTCGGATGGCGGACCATGGTGTATTGGGCCAATTGGCTACGTATATGCAAGACCCCGCGAA
GAGATTGATGGAAAGGAATGGAAGATACTGTGCGCCAAAAGTCAAGACCAGCCCTCCTGCGCCATCGCTCCTCTCATACGGGGGC
TTACTGTAGAGTCTGATTTCAAGCCAAACGTCGCAGCAGTCATCGGAACTAAGTCATCGGGAGTCGTTGGAGGTAATTGTACGGC
TATTTTATCTCCCTGCCATTTTTCAACAACGGTATATATCTTTCATGGAGGAGAGTGCATCCCGCCTACATCATTCACTCCAAAC
CTCACAAAAATATGCGAGGACGCAAGGGAAAGGTTCGGGTTCAGTTCACTACCCCCAAATGGACCGGTTCCAAATGCCATCGAAA
CTACAGGAGAAGAGATCTGCAAATCCCTGAATATGGACCCTGACAAAACGATGCTCTATCTCGTAGTTGCCGAACCGTTTTGTGA
GGCGGTGTATGTATGCAACACATACATTCACTATGGATCAGTCGAGACGCTTTACATCAATTCTAAGCATGTTACGCGCATACCA
ATCTACCCTGTGCAAATGTACATGCCAGACATCGCGCTGCGGATTTGCCGTAACCCATTTGATTCAAACAGCCGTAACATCGGTG
AAGGCTGCATTTACCCTACCCCATTATTCAACAAAGCTTTGAACAGACTCATTCATGGGGCAACACTCGGAACACAGGGACAAAG
TCTGAGAACCAGAGACCTAGAAGCCGTGGCGCGCGGAGCGGCAATGCTCGCGTTCGATGGAAGCTTCGAAGGCTGTGTATTAACC
TCGGACAAGACCTTTACCGAGCTGGCTCAGAACATACAACAAACAGGACCCCCTAAACAAAATGTAGAAGTAGAACGGCGCGCCG
CCTGTAGCTTGGCTGCTGAACTTGCGCTGGCAACTAGACTCTCTGTCTCCTGCGCTCCATATCCCTTTGATAATGGCTGCGGTCT
GCCGTATGAAGAGTGGCCGTTGTTCTGCGATTGCAAAACTCAAATGGAGCGCGTTCAAGCTCTCGATCGGTTCTCAGCAGAATTG
GCCGGCATTGTAGGAGCAGCAATTTTTTCAAATAACTCCCCCCTTTATGCAAGTGAGGTAGTCGATGGGGGTGCGCCGGAGCCCG
GCGAGCGCGCCTCGAGCGGACTAACCCGTTTTTATCTGACATGTGGATTGCATTTATTGGGATGCCCCCAGGTGGACTACTCAGG
GAAGAGAGTTTGGGAAGGCACATGTCCATCTCCGCTCACACCTCAGAACACATTCGAATATGATCCTGAGCACCTCGCATATGCC
TGTGGTTTTAGCCCCGAACTCCTGTCCCGTATCATATTTTATTTGGAAAAATGTTCCAAGCAACTCAGCACTCGGCCAAACATGA
ATCTGATTTCGATTTCTAACAGTGGAAAAGCGCCTTGTACTGGAGCCGTACCAGAATGCCGGTGGTGCAATGACGAGTCTAGGAA
TCATTGCATCCGTTACACTATGCAGCGACTGAGAACCCGTTTTCCTATTCCACGAACCATGAGAAGAAGCCCTATCGCAGTGATA
GGAGCAGTTGATGGAGACTACACGGACTGCGACATTTTGGGGAATTTTGCCCCGTATTCCCAGTTAAAGAGGGCAGGTGACGGTG
AACCAGCCAAAGCAGTTATGAATGACACATACAGGGGCATAGGCTGGCGAGTCTTTCAATATCTAGTATCAGAAGGGTTGATAAA
CAAAGACACGGGCGAAGACACGCATAATATTTCAAGCCTAGCAGAATTAAAAACTGTATTTGAAAAAATTCAGAACTTCGTTGGA
TCAGAATGCTCTAAGTTCATTTCTGCCCTCTCAGGAGTGAGAGCATATCATTACAAAGAGCACTTATCCTCAATCGTACATACAT
TTGGATCACCTGAACCGTATTCTTCTGCGTACTGTCCAGTACTTAGCTTATTGTGCGCCCAGACTCGCTCAATACTTTTCCAAGA
TTTGATTCTCAGTCAAATCCATGGAACATTTGATACGCGGCAACCAGAGTCAAAAATGTTCCGTTCTTCTGCCATGCCGGCCCTA
AGATCAGCCTTTAACGGGATGCTGGACAAAGGATTCCTTTCTGGAAAATATGAGCCATTTTCAATATCAGTTCCGTGTGTTAACG
CTCCGGATACAATGCGCCCAAACACAGAACAGGCAATCGCACAATATGAGTATTCTCTTTCTAGGGGACAAGTCCTGAAACTGAA
AGAGTTTAAGGTGAAGAACCGCGTCGTGTTCAACGGATCCAGTACGTCAAACTCGTGCCGCGTGGTATTGAGTTCTAAACCTATA
GTAATGTTAGGGATTATGATCAGCAAATATGTAGGACAACAAAACTCGACCACCGTCTTTCAGTCCGGCAATTGGGGTAGCCTTA
TAGGCAATTCAGGGTCCCAATCGGTTAATTCCGCGCTGACGAATGATCCCGTCAGAAAGTTTGCACTTGCTTGTAAAAGGGTAGG
AACTATTCTGTCATCTGGGCAAGTGGCATCAGGGTTTCAAGAAAACACCATTGCCTCTCAAGTGAGAAGCCTAATTGATGCAGGA
GGTAGTCCCACTTCCATCTACACAGCCGTTCTGAGAGTTCTGGGAGAAGGAATGAAAGACATTACTGCCGAAACATGGATGGCAA
TTACTGACGACAAGTATCTGGTAAATATATTAGTAGAGCTCCGTGAGGACATTTCTGGAAGTCAGAATGGGTGGTCCGTGGCTAC
CGCACAAAGTATGCTCAGTGAGCTAGAGGGGAAAGATACATGCCCGGAAAATGCAGGAGAGATGTTGGTCTGGGCAGACGAGGAT
GAATATACTAACACTGGCGACATCAGCGAAGATATTGAACTTGGCGCGCCAAGTGCAAAAAAACCTTCGCTATCGGCTGATATTC
TGTTCTGTTAACAGTAACTGCGCGTGCTCAGAATAAACAAATAAATATAAGAAAGTCCTTGTACGTATTTTGTTTCATTCTTACT
GGGGAATATG

>UL30

ATTGCCGCTCCGCGCGCCACGGCTTCTAGGTCTCTGGTTCTCAGACTTTGTCCCTGTGTTCCGAGTGTTGCCCCATGAATGAGTC
TGTTCAAAGCTTTGTTGAATAATGGGGTAGGGTAAATGCAGCCTTCACCGATGTTACGGCTGTTTGAATCAAATGGGTTACGGCA
AATCCGCAGCGCGATGTCTGGCATGTACATTTGCACAGGGTAGATTGGTATGCGCGTAACATGCTTAGAATTGATGTAAAGCGTC
TCGACTGATCCATAGTGAATGTATGTGTTGCATACATACACCGCCTCACAAAACGGTTCGGCAACTACGAGATAGAGCATCGTTT
TGTCAGGGTCCATATTCAGGGATTTGCAGATCTCTTCTCCTGTAGTTTCGATGGCATTTGGAACCGGTCCATTTGGGGGTAGTGA
ACTGAACCCGAACCTTTCCCTTGCGTCCTCGCATATTTTTGTGAGGTTTGGAGTGAATGATGTAGGCGGGATGCACTCTCCTCCA
TGAAAGATATATACCGTTGTTGAAAAATGGCAGGGAGATAAAATAGCCGTACAATTACCTCCAACGACTCCCGATGACTTAGTTC
CGATGACTGCTGCGACGTTTGGCTTGAAATCAGACTCTACAGTAAGCCCCCGTATGAGAGGAGCGATGGCGCAGGAGGGCTGGTC
TTGACTTTTGGCGCACAGTATCTTCCATTCCTTTCCATCAATCTCTTCGCGGGGTCTTGCATATACGTAGCCAATTGGCCCAATA
CACCATGGTCCGCCATCCGACTTTCCGCCCTTAGTTGTGCCAGATGAACTCTTTTCCATTTGCGCGAGTTATCTCAGACAGTATG
ATCTCAGATCAGAGTTCCGCGTAGTAGAGTAGTCTCCGTCACTGAGCACGGACTTAAGTACCGCGGAAAGGACACACCTATGTAT
ACTTTAGAGAGTTCATAAATATCGATGATTGCAGTAGCTGGGCACCGAGTTTGTCACTCTGCACTATGGACACTTTTTCCTCCGC
CGGTACTGGAGCTGCTCGTGTAGCCCCGATAATTGGCCTCGGGCCGCACTCTGGCGCGTCTTATTATACTTCTGTGAGAGAATTC
CCGTATGTATGTCCAACGTGCATTAATGGGGGCGGAAGGATAGGAACGCAGATTGGAAGGGCTACAAGCAAACCAACCTTTTACC
ACAACGAGCGACAATTAGACATCTTGACAAGCACACATGGAGCGTGGCCCGTCAGGATGAAATATTGGAATGCTGGTCCTGGAAT
GACTCACCACAACAAGGGAGATGTAAAGTTTTATGAATTTCACGTTTATGACCTTTTTGAAAACACTGAGATCGCTCAGTCATGT
GTACGATGGATGCATTCCAGATTTCTTGATACACTGAAACCGACCGGTACTGTAATAACCTTAGTTGGAAATTCTGCATGCGGAA
AGCGCGTCGCCGTTCATGTGTATGGATCTTTGCCATACTTTTTTGTTAAAAAGCGAGAAATTGACCAAGCCGCGAAAGTCACGAA
CTCGGAAGAGTTGGCGCATGCTCTAGCATTAGCGACGCGTAAGACATCTTTAAAAAATTCACCTTTTCTAGCTGTTTCAGCGGAA
AGTTTTTGCATTGATGTTGTTCGGCGCAGAGACATTTATTATTCTGAATCTGAGGAGGAGGACTTCTACCGTGTAAAAGTGTGTA
ATGGCAAAGTGATGAAATTTATATGTGATAATTTTTTCCCGAGCGTACCTAAATATGAAAGCAATGTTGACGCTATTACCAGGTT
CATCCTCGACAATAATCTGACTTCTTTTGGGTGGTATCGCTTTAGAGCTCAAGGTGGCGCACTCCAAATTCGAGACCCAGGACAA
CATGCTACATCAGCTGATGTCGAAATTAACTGCACAGCAAGTAATTTGGAACTCGGAAACAATGTCTCGTGGCCAGACTACAAAT
TACTGTGCTTTGATATAGAATGTAAAGCCGGAGGGGCAAATGAGTTTGCATTTCCTTCTGCTGAGAAGGTAGAAGATCTTGTTAT
CCAAATTTCGGCAGTTACATATTCTTTGTTAACTAAGGAAAAAGAGCAGGAAATCATTTTTTCCCTAGGGACCTGTGAACTTTCC
GAAGATTGTAGTGATGGGATAACTGTATGCAAATGTGGCTCTGAGTTCGAGCTTCTTCTGTGCTTTATGACTTTTTTCAAGCAAT
ATTCACCCGAATTTGTGAGCGGATACAATATTCTAGGGTTCGATTGGGCCTATCTTTCTAACAAGCTGAGCAAGGTGTATGGCAT
GCGTCTCGATGGATATGGGAAAGCAAACTCGTGGGGGACATTCAAACTTCAAGATCCGTCTGCGCGCGGACTTGGAAGATTCAAA
AAAGTCAAAATCAATGGCGTCGTCAACATTGATATGTTTACCATTAGTTATGAAAAGCTAAAGTTACCATCTTATAAGTTAAATG
CAGTAGCAGAGTGCGTTCTTGGAGAACAAAAAATTGATCTGGCTTACAAAGATATCCCAGTCATGTTTGCTGCAGGCCCTAAAGA

AAGGGGGAAAATAGGGGAGTATTGTTTGCAGGATTCTAGACTGTCAGGGAGTTTGTTTTTCAAATTCTCCCCTCATTTAGAAATG
TCCGCTGTCGCTAAATTGGCCTGTATACCCTTGACAAGAGCAATAGGCGATGGACAGCAGCTGCGCGTTTATACATGCTTGCTCC
AACGCTCTACGGCGTCCGGGTTTGTGTTGCCGGACAAAAAGAGTGCGTTTTCCTTCGGTTCTACTCTTGCTTCAGACGCGTCTGA
TGCACCTGCTACAAGAAATGTAGGCTACCAGGGAGCCAAGGTACTAGATCCAGAAATAGGATTCCATGTAACTCCTGTGGTTGTA
TTCGATTTTGCCAGTTTGTATCCGAGTATTATTCAGAGTAACAATCTATGTTACAGTACGTTAACGCACGACCCCGCCGCACTGG
CAGGTCTACAGCCAGAAAAAGATTTCTTCCAGATTGATGTTCAGGGTCGGAAATTTTATTTTGTCAGGGAGCATATTCGCAAAAG
TTTGCTGTCAGACTTGCTAGGTGATTGGTTATCTATGAGGAAAGCCGTGCGCGCAAAAATCAAGACAGCCGAGACTGAAGAAGAG
CGAATCTTATTAGACAAACAGCAAGCTGCCTTAAGTTGTCTGGAACTCTGCTATGGGTTTACTGGGGTGATGCATGGAATGCTTC
CTTGTTTAGAAGTTGCTTCAACAGTTACAGCTATAGGAAGGGACATGCTTTTGCGTACAAAGGCGCACATCGAGAAGGAGTGGAG
AAGTGGAAATCAATTTGCCGAAAAATTTTTGCCCGGGTCCGAACGTATTCAGCTAAACCAATACTCTGTCCGGGTCATTAATGGA
GAAACTGATTCTGCATATTGTGAAGTTTACCGGAGTTTACATTTAAACTTTAACTCAAGTTGGTGGCTGCATGGCCGAATAGATT
TAAACAAGCGCTTTTTAGAGGTCCGGGTAAAACTTGAGTGGAAAAGATTTTCGTGCGCCCTGTTACTCATTGGTAAGGAAAAATA
CATTGGGGTGATGCATGGAGGTAAAATGTTGATGAAGGGCGTAGATTTAGTCAGGAAAACAAACTGTAAATTTGTCAATACCACA
GCCTCCCGATTAGTGAATCTACTTTTCGAAGACGACGAAATAGCGATGGCTGCAGATACAGCAAGAGCCGGATTTGAGGATTTTG
ACTGTCTTCCGAGCGGACTTACTAAGCTTGGCCGGTTGATTGCAGAGGCAAGGCTAGCTATTACCGGCAACGGACTAAACATTAG
GGACTTCATAATGACCGCGGAATTAAGTCGTGCGGTGGATAACTATGCCAGTTTGAAATTGCCTCATCTCACGGTTTATCAAAAG
AAAGCTGCTCCGCGCGAAGAATTACCTCCAATTAAAGAAAGAATCGAATATGTCATCATAGAACCGGGACAATCTGTCCCGAATG
ATCCAGCAACTGAGTCTTTTCCTAACTAAAAGAAACTTCTTAAATTTCTTAGCCTGGCAGAAGATCCAGAATGGGTCGTATCAAA
TGGTCTAAAATTGAATGTAGAATACTACTTTGATTCGCTTATTAGAACATTAAGTGTAACATTCAACGCAATTTTCGGTGATGCA
AAAACAGCAGAGGATGTTTTAAGAAGCTTTATTCCAGAGAAAATACAGTTTTCTGAGAAAGTTGCAGAGGCTCTTGCACGAAACA
CCGCGACATTTGTATCCATCCAGAAGTAATCGGAGGGCTGTAAGTTGCAAATGTTGAATAAAGTTTAGTGTATTCGTGAATTTTT
TCGTCTGTTACTAGGATATCGCAGTATCGTCGATA

>UL31

TGCCAGACCACTCGAGTGTTTTTGCTCCGCGCAGTGTATGTTAAAGAGGAAAAATACAGAGACTGCCTCCTTCGCTTTTCGAACC
TGTGCGCAAAAAGCTGCTTTAGAAGATATTAGCACGGAGTCAAAATATGTGGAAGGGGGACCAAGCGAAACCCTGACAACCCCAA
GTAAATGGGGATTTACCGACCTGACGGCATTATTAATTGCTGGAACAGCCGGTATGCCAAATCGGGAACTCGATGCCAGCTGCAG
CGACGAGCCTGAGTATGATGATACAAATGATAATGGAGAAAATTTCGTTGCCCCTATAATTCGGACAAGAAATACCGTAGTGGAG
AACTTTCACGCAACGCTAATCCAACAGGTCAAAAATACTAGGCTCCCCAAGACAGCGAAGCAGTACATTTGCGACAATACCAAAC
CATCACTTGCCATTGGCCCAATTTTGGCGTCTATTTTGCCTCACCCGACTAGCCGAGGAGATACTGGAGGGGAATGCGTGTTGTG
CAATCTCATGCTTACTCGGGAACACTGGCATGCTCTTCGAAAGCTGAAAAGTAAAGTAGTCGGTTGCAGTAACGTAAATAGCAGC
TTGTTCGATGGGATTGAACCAGCATTAGAAACGTTCGAAGACTATACTGCGTTAAACGATGGGGGGCGAATGTTAACCTTATTGA
GGATGGCCGGGGCAAACGCAATTTACAAACACTTTTTTTGCGATCCATTGTGTGCAGTTAATACGCTTCGCGTAAATCCAAAAGT
TTTGTGGGAAGGCCAACCCAAGGACCCAGAACTGCTGCAGCTTTATAAAGCCGAAATCGCCACGGCTAACATATTTCAAGAACGG
GTTTGCCGCGGGCTGTGGATTCTAGCCTTCACATTTAAAGCTTACCAGCTTTCTCCTCCTCGCCCTACCGCTTTAAATTCTTTCA
TTCGCGGCGCTGAGACCTATTTAAAGAGGCACGGAATCAGCTGCATTGCTTTAGAACACGCACTGACTCGATATGTCTGACTTCC
GTTCTCGCGCCGCTGCTCTGCGCATCCGACGAGAAGGTATTACGGGGGCCATCGAACGCCATCGCTACATCTCATACCTCAGATC
GAAGAGGCGATCGTCTAGGTCAAGGGCTAATCTAGACACACGGAGAACTTTCTTCGATGCTTTTTTCAAGCTAACGGCTTCATCT
CCGGAAGATACTATCTCCCTTCTCAGAACAATGACGATTCCTGTCATACGCCAAGAAAATATCTCCCTTCCTTACAATGTGAATG
CAAGTTTTGCCCCTGGCGACTGTATCTCTCTTTCCGAGATGGGCTACACCGTTGGACCTGGTGGATGCTGCTCACTTTGCTCATA
TGGATGGTCAAATGCCCCACCGTTAGACTTGCCTGCGTTGGAATTGGCCTTTCTCCATCACCTGAGCTCTGTGGTCGAATTCAAA
GAGTTAGTGGCGTCCCTGCGGGTATTTTCAGGTGAATCGGTAGTTGGATCTGGAGCATATGAAAACGAATTCCTACTTGACACTA
TCAAAAATCTACTGGCCCAGGCGACTTTGTTTTATGCATACTATACAGTAAAGGGCGGACTCTCTCAAGATTTTACCGTTCTCTT
GGTGAACGATCCAAGATCTAAGAAGTTTCCATCGGGATACCTTATGTATTTTGTATTTAAACAGAGCACGCCACTTCATCTAGGG
GCAAAGCTTCTGAAACAATTAATATTTAATTGCCCAGGATACACTTGGCACGTGGACGTATTCGCTGGTACGTTCTTATTGATCG
TTTCCAAAGAAGATGGCCCAGGTTCCTCTGAGCATCACCGCGTTAAGCTGACCCCGGATTCAGTGTATCGACGATACTGCGATAT
CCTAGTAACAGACGAAAAAATTCACGAATACACTAAACTTTATTCAACATTTGCAACTTACAGCCCTCCGATTACTTCTGGATGG
ATACAAATGTCGCGGTGTTTCGTGCAAGAGCCTCTGCAACTTTCTCAGAAAACTGTATTTTCTCTGGAATAAAGCTTCTTAAAAC
ATCCTCTGCTGTTTTTGCATCACCGAAAATTGCGTTGAAT

>UL32

GAATCTTCTTGCTCCGCGGAGTCTATAGCCTCTTGCACTCTGCGAGTAATTTTAATCAAATTGGGTTCCGAGTCTGAGCGGTAAA
ATGCGAATCTGATGGTAATATCAGTTCTCGACCCCCTGTAAAAACTAATGTACATCATTGTTGCGAGCGGGGCGTTTTTTAACTT
TTGAAGTCTTTTCATTTCCTTCACCGATGAGGAACTTACCCACAGAGATTTTGCAGCCGCTAAAGATATGGTAGGACTAGAAGCA
GGCCATATTTGGTCTTTGCAATTAGCGGTCTTGGTAAAACATCCGGCAACAATGACAGCCTGGCCCGTATTTTGAATACGCAGCC
ACGACTCTCCAATATTAAGGAAGCTATTCATTGCCTCTAGGATATATTCCGCTGGAATTGGGGGACCGTCAAGTTTAGCAAGAAG
CGTAAAAATATCTCCAGTGGGTGCGAGACGTGCATTATCTCTGGTATACACCAGACTCGCTGTCGTTCCGGCAAGGCTTAATCCA
GCATTCACGAAGCGCACCAATTGACAGTATTTGTCTCCTTGATTTCTGGCGCCTTGGGGTTTCATTGCTTTCTTTATGATCTTCT
GGATCCGTCTTAATACCATACACACTCGCCCTCTTATTTATAATACTACCAGACTAACAAGGTTCTAGCGCCCCTAAAATTTGGT
GAAGATCCAAAAACTTACTTAGAATTTGCGCGAACCGTTCACGACGGCGAGCAATAAGTTCTGCGTGTGCGCACATCCTTTTCCC
ATCTAAATTGCGGTACCGAAGTTCGGCCGCGACATTGGCTGTGTGCGCAAGGTAGTTTAGTTTAGCATCGGTAGTAGGCAGAGCT
ACTTCAATTTCCGGAGGCGATAGGTCATAAAACCATATTTCATATACCGGACTACCTGCAGATCCAATATACTTTCTGAACAGTT
CTGAGAGTGGCCTTTCCGCTAACTCCTCTTCATCTACCAATTCACTAAGACGCCTCGGAACTGTGGATGCCATGACATCCAAACG
ATCAACTGCAGAAACTCGCACTGAACTTTCCAAGGGATGGAAAGCTGGAGTGCTCTCAAAACTATATACTGGCTACGACCCGGCG
ATTCTGACGTTTAATGACAAGTTGTCGGACGAACTATTGTATGGAGCTCACCTAATTGAAATATATCATGAAGAAAGTACAGAAC
AAGCTGAAGAACAAAACAGAATTCCGGACATTGATTCTTTCGTAGAAACCATGGCTGACTTATACACCCTAGATAAATCATGCGC
CGTCTGCAAGATTATTACCAAATATCAAAAAAAATTCCCCGTGACTCCCGAATGGATGGTGGACTACTCTTTATTGTGCTTCAAA
TCTCAAAGCATCCCCCTGTGTGCTACATCTACCTTCATAACTGCTTTTGAGTTCATCTTCATCATGGACAAACACTACTTGCATT
CCCACAGCACCTCTTTAGTTGGCGCTATAAGCCGTCGGGAATTTTCCCTCGGGGATATTCAGAAGCACTTCTTTTTTAACGGCTG
TTTCAAAGCGGTGGAAGGGGGACTGGACTCTAAAATAGATCTAAATAATTATTCGTTCCTAGTGCAGTCGGTAGCACGCTATGCG
TTACTATCTACTAGATATCAAGAAGCACTTAGGGCAAAAATTGCCCCGGTCTCTGGCGCGTCTGCAGATTTCAACTTCCCAAAAA
GTAGCGGAATGGCCGTGGCTTTGTTAAAATGGCGTGAGTATGCCAGACCACTCGAGTGTTTTTGCTCCGCGCAGTGTATGTTAAA
GAGGAAAAATACAGAGACTGCCTCCTTCGCTTTTCGAACCTGTGCGCAAAAAGCTGCTTTAGAAGATATTAGCACGGAGTCAAAA
TATGTGGAAGGGGGACCAAGCGAAACCCTGACAACCCCAAGTAAATGGGGATTTACCGACCTGACGGCATTATTAATTGCTGGAA

CAGCCGGTATGCCAAATCGGGAACTCGATGCCAGCTGCAGCGACGAGCCTGAGTATGATGATACAAATGATAATGGAGAAAATTT
CGTTGCCCCTATAATTCGGACAAGAAATACCGTAGTGGAGAACTTTCACGCAACGCTAATCCAACAGGTCAAAAATACTAGGCTC
CCCAAGACAGCGAAGCAGTACATTTGCGACAATACCAAACCATCACTTGCCATTGGCCCAATTTTGGCGTCTATTTTGCCTCACC
CGACTAGCCGAGGAGATACTGGAGGGGAATGCGTGTTGTGCAATCTCATGCTTACTCGGGAACACTGGCATGCTCTTCGAAAGCT
GAAAAGTAAAGTAGTCGGTTGCAGTAACGTAAATAGCAGCTTGTTCGATGGGATTGAACCAGCATTAGAAACGTTCGAAGACTAT
ACTGCGTTAAACGATGGGGGGCGAATGTTAACCTTATTGAGGATGGCCGGGGCAAACGCAATTTACAAACACTTTTTTTGCGATC
CATTGTGTGCAGTTAATACGCTTCGCGTAAATCCAAAAGTTTTGTGGGAAGGCCAACCCAAGGACCCAGAACTGCTGCAGCTTTA
TAAAGCCGAAATCGCCACGGCTAACATATTTCAAGAACGGGTTTGCCGCGGGCTGTGGATTCTAGCCTTCACATTTAAAGCTTAC
CAGCTTTCTCCTCCTCGCCCTACCGCTTTAAATTCTTTCATTCGCGGCGCTGAGACCTATTTAAAGAGGCACGGAATCAGCTGCA
TTGCTTTAGAACACGCACTGACTCGATATGTCTGACTTCCGTTCTCGCGCCGCTGCTCTGCGCATCCGACGAGAAGGTATTACGG
GGGCCATCGAACGCCATCGCTACATCTCATACCTCAGATCGAAGAGGCGATCGTCTAGGTCAAGGGCTAATCTAGACACACGGAG
AACTTTCTTCGATGCTTTTTTCAAGCTAACGGCTTCATCTCCGGAAGATACTATCTCCCTTCTCAGAACAATGACGATTCCTGTC
ATACGCCAAGAAAATATCTCCCTTCCTTACAATGTGAATGCAAGTTTTGCCCCTGGCGACTGTATCTCTCTTTCCGAGATGGGCT
ACACCGTTGGACCTGGTGGATGCTGCTCACTTTGCTCATATGGATGGTCAAATGCCCCACCGTTAGACTTGCCTGCGTTGGAATT
GGCCTTTCTCCATCACCTGAGCTCTGTGGTCGAATTCAAAGAGTTAGTGGCGTCCCTGCGGGTATTTTCAGGTGAATCGGTAGTT
GGATCTGGAGCATATGAAAACGAATTCCTACTTGACACTATCAAAAATCTACTGGCCCAGGCGACTTTGTTTTATGCATACTATA
CAGTAAAGGGCGGACTCTCTCAAGATTTTACCGTTCTCTTGGTGAACGATCCAAGATCTAAGAAGTTTCCATCGGGATACCTTAT
GTATTTTGTATTTAAACAGAGCACGCCACTTCATCTAGGGGCAAAGCTTCTGAAACAATTAATATTTAATTGCCCAGGATACACT
TGGCACGTGGACGTATTCGCTGGTACGTTCTTATTGATCGTTTCCAAAGAAGATGGCCCAGGTTCCTCTGAGCATCACCGCGTTA
AGCTGACCCCGGATTCAGTGTATCGACGATACTGCGATATCCTAGTAACAGACGAAAAAATTCACGAATACACTAAACTTTATTC
AACATTTGCAACTTACAGCCCTCCGATTACTTCTGGATGGATACAAATGTCGCGGTGTTTCGTGCAAGAGCCTCTGCAACTTTCT
CAGAAAACTGTATTTTCTCTGGAATAAAGCTTCTTAAAACATCCTCTGCTGTTTTTGCATCACCGAAAATTGCGTTGAAT

>UL33

TACTCAGGCTCGTCGCTGCAGCTGGCATCGAGTTCCCGATTTGGCATACCGGCTGTTCCAGCAATTAATAATGCCGTCAGGTCGG
TAAATCCCCATTTACTTGGGGTTGTCAGGGTTTCGCTTGGTCCCCCTTCCACATATTTTGACTCCGTGCTAATATCTTCTAAAGC
AGCTTTTTGCGCACAGGTTCGAAAAGCGAAGGAGGCAGTCTCTGTATTTTTCCTCTTTAACATACACTGCGCGGAGCAAAAACAC
TCGAGTGGTCTGGCATACTCACGCCATTTTAACAAAGCCACGGCCATTCCGCTACTTTTTGGGGAAGTTGAAATCTGCAGACGCGC
CAGAGACCGGGGCAATTTTTGCCCTAAGTGCTTCTTGATATCTAGTAGATAGTAACGCATAGCGTGCTACCGACTGCACTAGGAA
CGAATAATTATTTAGATCTATTTTAGAGTCCAGTCCCCCTTCCACCGCTTTGAAACAGCCGTTAAAAAAGAAGTGCTTCTGAATA
TCCCCGAGGGAAAATTCCCGACGGCTTATAGCGCCAACTAAAGAGGTGCTGTGGGAATGCAAGTAGTGTTTGTCCATGATGAAGA
TGAACTCAAAAGCAGTTATGAAGGTAGATGTAGCACACAGGGGGATGCTTTGAGATTTGAAGCACAATAAAGAGTAGTCCACCAT
CCATTCGGGAGTCACGGGGAATTTTTTTTGATATTTGGTAATAATCTTGCAGACGGCGCATGATTTATCTAGGGTGTATAAGTCA
GCCATGGTTTCTACGAAAGAATCAATGTCCGGAATTCTGTTTTGTTCTTCAGCTTGTTCTGTACTTTCTTCATGATATATTTCAA
TTAGGTGAGCTCCATACAATAGTTCGTCCGACAACTTGTCATTAAACGTCAGAATCGCCGGGTCGTAGCCAGTATATAGTTTTGA
GAGCACTCCAGCTTTCCATCCCTTGGAAAGTTCAGTGCGAGTTTCTGCAGTTGATCGTTTGGATGTCATGGCATCCACAGTTCCG
AGGCGTCTTAGTGAATTGGTAGATGAAGAGGAGTTAGCGGAAAGGCCACTCTCAGAACTGTTCAGAAAGTATATTGGATCTGCAG
GTAGTCCGGTATATGAAATATGGTTTTATGACCTATCGCCTCCGGAAATTGAAGTAGCTCTGCCTACTACCGATGCTAAACTAAA
CTACCTTGCGCACACAGCCAATGTCGCGGCCGAACTTCGGTACCGCAATTTAGATGGGAAAAGGATGTGCGCACACGCAGAACTT
ATTGCTCGCCGTCGTGAACGGTTCGCGCAAATTCTAAGTAAGTTTTTGGATCTTCACCAAATTTTAGGGGCGCTAGAACCTTGTT
AGTCTGGTAGTATTATAAATAAGAGGGCGAGTGTGTATGGTATTAAGACGGATCCAGAAGATCATAAAGAAAGCAATGAAACCCC
AAGGCGCCAGAAATCAAGGAGACAAATACTGTCAATTGGTGCGCTTCGTGAATGCTGGATTAAGCCTTGCCGGAACGACAGCGAG
TCTGGTGTATACCAGAGATAATGCACGTCTCGCACCCACTGGAGATATTTTTACGCTTCTTGCTAAACTTGACGGTCCCCCAATT
CCAGCGGAATATATCCTAGAGGCAATGAATAGCTTCCTTAATATTGGAGAGTCGTGGCTGCGTATTCAAAATACGGGCCAGGCTG
TCATTGTTGCCGGATGTTTTACCAAGACCGCTAATTGCAAAGACCAAATATGGCCTGCTTCTAGTCCTACCATATCTTTAGCGGC
TGCAAAATCTCTGTGGGTAAGTTCCTCATCGGTGAAGGAAATGAAAAGACTTCAAAAGTTAAAAAACGCCCCGCTCGCAACAATG
ATGTACATTAGTTTTTACAGGGGGTCGAGAACTGATATTACCATCAGATTCGCATTTTACCGCTCAGACTCGGAACCCAATTTGA
TTAAAATTACTCGCAGAGTGCAAGAGGCTATAGACTCCGCGGAGCAAGAAGATTCATATAAAACGGCGGCCCTTCTTAAAAAAAC
GTTGATGGAAACTACAAGCGAGATGCCCACAAGTCACTCGGTTGCTGAGTCGAAGCCATCACGTTCTATTTTCGAGCGCGTCACC
ACATATTTTCGAATACATGTAAAATGTTTAAACCCGAGGTCCTTCTCTTTACAGCCACTGATGTGGGTGCTTATGGGTGTGGCAT
GGCCTGCTTTTTTGACACTACTGATCTTCTATTTAATCCGAGTACAGGCGACGAGCTGAAATCATACCACACCTGGAGGCAGATA
TGAGCAGGCAATCTAACTCTCATTCCACGCCCCAACCACAGGACAGAGTCAATGCGGCCGCAGAAATGTTGAAAGCTTTTTTATT
GACCCCTCCGAGAGATCGAGCTCCGGACTACATGCAAACCTTAAACGGAATTACCTTGGATCAACTCACGGAAGTAGCTCGAGCC
TTGAAAAATGAAATCCCGAGAGAAGATACAGTTCGGCAAAATATTCTTAGAAATAATATCACACTCGCGCTAATGGCTTTAAACC
GAGCTCCAATGCTTCGAGATCGGTTAAATATAAGGCCCGTATTTGCTCGCCTTTCAGGCCCTCAAGGATTGTGGACGTTTGGAGT
GAGACAACGCACCAGGCCAAGACCGAGACGCTAAGCAAAACGCAATAAATATACCAATTTAACCCAGTATTATTTTTATTTTGTG
TGTTTCATCCGCTATGCGGGCCTGCTGATCTAAAGCGTAATCTCTCGCGTGTTTTATTCC

>UL34

TTTAGATCTATTTTAGAGTCCAGTCCCCCTTCCACCGCTTTGAAACAGCCGTTAAAAAAGAAGTGCTTCTGAATATCCCCGAGGG
AAAATTCCCGACGGCTTATAGCGCCAACTAAAGAGGTGCTGTGGGAATGCAAGTAGTGTTTGTCCATGATGAAGATGAACTCAAA
AGCAGTTATGAAGGTAGATGTAGCACACAGGGGGATGCTTTGAGATTTGAAGCACAATAAAGAGTAGTCCACCATCCATTCGGGA
GTCACGGGGAATTTTTTTTTGATATTTGGTAATAATCTTGCAGACGGCGCATGATTTATCTAGGGTGTATAAGTCAGCCATGGTTT
CTACGAAAGAATCAATGTCCGGAATTCTGTTTTGTTCTTCAGCTTGTTCTGTACTTTCTTCATGATATATTTCAATTAGGTGAGC
TCCATACAATAGTTCGTCCGACAACTTGTCATTAAACGTCAGAATCGCCGGGTCGTAGCCAGTATATAGTTTTGAGAGCACTCCA
GCTTTCCATCCCTTGGAAAGTTCAGTGCGAGTTTCTGCAGTTGATCGTTTGGATGTCATGGCATCCACAGTTCCGAGGCGTCTTA
GTGAATTGGTAGATGAAGAGGAGTTAGCGGAAAGGCCACTCTCAGAACTGTTCAGAAAGTATATTGGATCTGCAGGTAGTCCGGT
ATATGAAATATGGTTTTATGACCTATCGCCTCCGGAAATTGAAGTAGCTCTGCCTACTACCGATGCTAAACTAAACTACCTTGCG
CACACAGCCAATGTCGCGGCCGAACTTCGGTACCGCAATTTAGATGGGAAAAGGATGTGCGCACACGCAGAACTTATTGCTCGCC
GTCGTGAACGGTTCGCGCAAATTCTAAGTAAGTTTTTGGATCTTCACCAAATTTTAGGGGCGCTAGAACCTTGTTAGTCTGGTAG
TATTATAAATAAGAGGGCGAGTGTGTATGGTATTAAGACGGATCCAGAAGATCATAAAGAAAGCAATGAAACCCCAAGGCGCCAG
AAATCAAGGAGACAAATACTGTCAATTGGTGCGCTTCGTGAATGCTGGATTAAGCCTTGCCGGAACGACAGCGAGTCTGGTGTAT
ACCAGAGATAATGCACGTCTCGCACCCACTGGAGATATTTTTACGCTTCTTGCTAAACTTGACGGTCCCCCAATTCCAGCGGAAT
ATATCCTAGAGGCAATGAATAGCTTCCTTAATATTGGAGAGTCGTGGCTGCGTATTCAAAATACGGGCCAGGCTGTCATTGTTGC

CGGATGTTTTACCAAGACCGCTAATTGCAAAGACCAAATATGGCCTGCTTCTAGTCCTACCATATCTTTAGCGGCTGCAAAATCT
CTGTGGGTAAGTTCCTCATCGGTGAAGGAAATGAAAAGACTTCAAAAGTTAAAAAACGCCCCGCTCGCAACAATGATGTACATTA
GTTTTTACAGGGGGTCGAGAACTGATATTACCATCAGATTCGCATTTTACCGCTCAGACTCGGAACCCAATTTGATTAAAATTAC
TCGCAGAGTGCAAGAGGCTATAGACTCCGCGGAGCAAGAAGATTCATATAAAACGGCGGCCCTTCTTAAAAAAACGTTGATGGAA
ACTACAAGCGAGATGCCCACAAGTCACTCGGTTGCTGAGTCGAAGCCATCACGTTCTATTTTCGAGCGCGTCACCACATATTTTC
GAATACATGTAAAATGTTTAAACCCGAGGTCCTTCTCTTTACAGCCACTGATGTGGGTGCTTATGGGTGTGGCATGGCCTGCTTT
TTTGACACTACTGATCTTCTATTTAATCCGAGTACAGGCGACGAGCTGAAATCATACCACACCTGGAGGCAGATATGAGCAGGCA
ATCTAACTCTCATTCCACGCCCCAACCACAGGACAGAGTCAATGCGGCCGCAGAAATGTTGAAAGCTTTTTTATTGACCCCTCCG
AGAGATCGAGCTCCGGACTACATGCAAACCTTAAACGGAATTACCTTGGATCAACTCACGGAAGTAGCTCGAGCCTTGAAAAATG
AAATCCCGAGAGAAGATACAGTTCGGCAAAATATTCTTAGAAATAATATCACACTCGCGCTAATGGCTTTAAACCGAGCTCCAAT
GCTTCGAGATCGGTTAAATATAAGGCCCGTATTTGCTCGCCTTTCAGGCCCTCAAGGATTGTGGACGTTTGGAGTGAGACAACGC
ACCAGGCCAAGACCGAGACGCTAAGCAAAACGCAATAAATATACCAATTTAACCCAGTATTATTTTTATTTTGTGTGTTTCATCC
GCTATGCGGGCCTGCTGATCTAAAGCGTAATCTCTCGCGTGTTTTATTCC

>UL35
GGTTCGCGCAAATTCTAAGTAAGTTTTTGGATCTTCACCAAATTTTAGGGGCGCTAGAACCTTGTTAGTCTGGTAGTATTATAAA
TAAGAGGGCGAGTGTGTATGGTATTAAGACGGATCCAGAAGATCATAAAGAAAGCAATGAAACCCCAAGGCGCCAGAAATCAAGG
AGACAAATACTGTCAATTGGTGCGCTTCGTGAATGCTGGATTAAGCCTTGCCGGAACGACAGCGAGTCTGGTGTATACCAGAGAT
AATGCACGTCTCGCACCCACTGGAGATATTTTTACGCTTCTTGCTAAACTTGACGGTCCCCCAATTCCAGCGGAATATATCCTAG
AGGCAATGAATAGCTTCCTTAATATTGGAGAGTCGTGGCTGCGTATTCAAAATACGGGCCAGGCTGTCATTGTTGCCGGATGTTT
TACCAAGACCGCTAATTGCAAAGACCAAATATGGCCTGCTTCTAGTCCTACCATATCTTTAGCGGCTGCAAAATCTCTGTGGGTA
AGTTCCTCATCGGTGAAGGAAATGAAAAGACTTCAAAAGTTAAAAAACGCCCCGCTCGCAACAATGATGTACATTAGTTTTTACA
GGGGGTCGAGAACTGATATTACCATCAGATTCGCATTTTACCGCTCAGACTCGGAACCCAATTTGATTAAAATTACTCGCAGAGT
GCAAGAGGCTATAGACTCCGCGGAGCAAGAAGATTCATATAAAACGGCGGCCCTTCTTAAAAAAACGTTGATGGAAACTACAAGC
GAGATGCCCACAAGTCACTCGGTTGCTGAGTCGAAGCCATCACGTTCTATTTTCGAGCGCGTCACCACATATTTTCGAATACATG
TAAAATGTTTAAACCCGAGGTCCTTCTCTTTACAGCCACTGATGTGGGTGCTTATGGGTGTGGCATGGCCTGCTTTTTTGACACT
ACTGATCTTCTATTTAATCCGAGTACAGGCGACGAGCTGAAATCATACCACACCTGGAGGCAGATATGAGCAGGCAATCTAACTC
TCATTCCACGCCCCAACCACAGGACAGAGTCAATGCGGCCGCAGAAATGTTGAAAGCTTTTTTATTGACCCCTCCGAGAGATCGA
GCTCCGGACTACATGCAAACCTTAAACGGAATTACCTTGGATCAACTCACGGAAGTAGCTCGAGCCTTGAAAAATGAAATCCCGA
GAGAAGATACAGTTCGGCAAAATATTCTTAGAAATAATATCACACTCGCGCTAATGGCTTTAAACCGAGCTCCAATGCTTCGAGA
TCGGTTAAATATAAGGCCCGTATTTGCTCGCCTTTCAGGCCCTCAAGGATTGTGGACGTTTGGAGTGAGACAACGCACCAGGCCA
AGACCGAGACGCTAAGCAAAACGCAATAAATATACCAATTTAACCCAGTATTATTTTTATTTTGTGTGTTTCATCCGCTATGCGG
GCCTGCTGATCTAAAGCGTAATCTCTCGCGTGTTTTATTCC

>UL36
TTGCCGTGACAAAAAAGAACTAGGACACTGGAAGGACGATACTGTAATGCAGGCCATATGGAAAGCCGTCAATTTTGCAGAAACT
GCAATCCGACACACAGAAAAAGCTATTCGGGCATTAGATTCCACAAAGGTTAAAACGCATGCACTTGAGACCACATACTGGCAGC
TTGCGGGTTATGCTCAAACACGCTCTCACATTTGTAGGAAAGTAGCTGAAGAATTATCTGGCATAGCGGACGCGAGGAATACTGT
CGTCCAGCGTTCGTGGACATTATGTCGTTCACTCGCAGGCGTCGGGGTAAAAAGGGCCCGCGATGTGGTAAAAGCCTGGAAAACT
TTTTCAGATAATGAAGCTGCTCTCGAGACATCTTAATCCAAAACCCTGAAGATGTAGTTAAAGTTATTGGCGCGGCATCAGCTCT
GCTAGATATAGAGGAGGGAGGGACCACGACCCTTGAAGGAAAATTGGAAGAAATTGAAATTTCTGCCCCTTCTAAAGAAGATACA
GCGTCAGCTCTTACTCTTTTAGGGCCGCCTGATGCCAGGGGATATCAAGATACATCATCAAATCAATTTATTACCGAGGCGTCAG
ACATTACTTCGTGGGATAAAACCAACAAAGCCCCGCTCTATGTTCGACACTCGACAAAAGAAGGGATAGACAATGGGTGTGTAGA
CACACTTCCAGACAAACTAATCTCATCAGAGACTCTTCTGACAGTTTCCGATCGAGTTTTGGAAAGTGTAAATGTAAAAATGTTT
GAATAAAGAAAGACTAATGAGTAATTGAGTGGTGTGTATTTATCGCGTTGTTGTGTGCGGTTGGCTAGACATTTACGAAACTCTT
CCCAAACAAATAAAAGGCCCCAGAAGCAGACTGATGTACAGTTTTGGCAGGTCACGAACATGGCTCAGGATTTCCATCCTCCGAC
AAGATTACCACCGGCATTCGGAGAATTTATTGTAGCCGCGTCGACCAGCCAATTTGCTCGAAAATATGAACCTGTTAGATACTAT
ATGTGTATGGAGACGTCCGCCGCCTTCCTGCGGGGGTGCGAATGCTACGGGATCAGATTCGATGCTCTCTGCGGACAGTTTGGAT
GCGATTTTAGACCAGGGGGCGGAGATTACGAGAACAGGGAAGGCCAAGGTGGAGGACAAGTCAAAAAAATGTTCGTTTCGAATGA
TTGATTGTCACGAACTTCCCAGACTCTGGAGATCATCATGGGATGAACCTGACATGCTCATGATACCTTCTATACCATACGGGAA
TGTGGGTTATGACATGGAAGTCACAATCGAAGGTTTCCCATCTTCCATGAATATACGCGGCATTACCAAAATCATTGCCAAGAAA
ATTAGGAAAGGATTGGGATTCATAATGATCATCGGTGACAGGGGAGTTGCAGCCGTGTGTTCTGGGGGAAAGGACGATCCTGTGT
ACATTTTTGATTCACATGGATGGGAAATGTCAAGTGCTTATGTATGCCAGCTAAGGAGAATTGATGACCTCCCAGAATTTGTTCG
GGCATATGTAAAGAATATGTCTGGAGTTGCAATAGATATGAATTTTCTCTACTGGTTCAATGCGGATGGGTGGGACGAAGAAAAG
GTATCGGAGAAAACGCTCCAGGAAATCATTGCGGCATCAGTAATGGCATCATATGGCGCATCAGATGCTGAATTAAGCAGTGTTA
CGTTAAGCACACAAATAATTCATCCATTTCCGGCGCATGTCTATGCTCTCGGACTAAACCCAGAGTTTTCTCAGAAGAACCTTTT
TGAACTCGCGGACCGTAGAAAGGCTATTGCTCCATGGAGCGTTCCGGTAAGACCAAGATCAAAAAAGAGACAGAAGCCACGGCCA
AAGAATCCTGATCAAATCGGAAGCTCAGAAAAACTGACAAACCTCACACCACGAAAACAAAGAAAACACTCCGTACGTTCAGACC
CAGAAGCTCAAAGCAATACTAAGCCGCATAGCGCCAGGAGGAAGCTATCTTGGAATTTACTTACGGCAGAAGACACAAACCCATC
AGCAACAATTGAATTAGGATCCATAACACCAATAAAAAATGCCTTGGAAGATCCTCTTAAGTGGTTACCGGCCCCACAGGAGACA
TTCGAACCGGCACTGACGAGATTTTTGCGATCTGGGATTTTACGGGAAAAGGTGAGAATTACCTGCAATCGCATATTATTTTTTA
TTATAGAGAACGGGGTCTCCACAAAAAACTTGGAAAGCACGGTGGACCTTCTAATAGACCCATTGGCAACGCTATTGGCTCGCAT
CGGAGAGAAAAATATCAGCAATATATTGACCACTACAAGACTGCGTTATCAGGATCTTATCGAAAAGAAACACTTGCTCCTGGAG
CTATGCCACTCTAACAGTGAAGGGGGCAGTGTTCTTTTAAGAAAAACGCAGCAGGTCAGTTTATCCCTGTTAACGCAGACAGAAA
TATTAATTGCTAAATTGAATCAAATCGCAGACAGTGCCGAGCGTGATCAACTTCCAACTGCGTACGAACAGTTATCCGAAGAACT
GCTGGAATGTTCAAATAAAACGAGTATTCTAATAGAAGATGTTAATTCACACGGACACAAAATTACTGATGTAGTAGAGCGTGCT
GCGAAATCTGTGATAGAAAGTGAAGCACGCTTCGCTGATGCCAAAATTGCAGTACTATCACTTCTCGCCAATACGTGGACAGTCA
TTGATGAAATTTGTAACATAGGAGTAGATGTGATCAATATCCAAGCAACCCAAACTCCGGACATACCACAAGCAATTATACAGGG
TTTGAGGAAATTTACAGGACTCGACGTTCAGGATATTAATCGAAAGCTCGACGACGCAATTATCAGAATTGAGGATGCTGCGGTC
AAATTCCTAAAGAAATATGCGGGTGCCATTGACTATAGCATATCGGCATTATCATCTGGGGTCGGTGGAGTATTGGGAAGGTTTA
GCATCCCTACTCTATTTGTAGAGAATGCTGTCAATTGGCTAAAAAACATAGAGGAGATTGAAAGGCGGTTTCAGTTGATTAGGCA
ACTTGCAGGAACCCCATCGAAGAACTTACAAGTACTGTTCACAAGAAAGCCATTGAGAATTTTCAGAGATCTGGTAATCGATGGG
AAAGACCTTACCACCGATGAAAAACTCAAGAGATGGTCCCTACATCTATTCGAAGCTCATGCAGAGGGACTTGTCCCATCACCTA
CAAAATGGATCTCCGCTATTCATGACATCAACGCTATTGCTCAGAGTTTTGGGCTAACTGAAAATATGCTGGCTGAATTAGAAAA

CGAGCTCGAGGGAGTGGAAGCTATTACAGCGCCCCCCCAAGTACGCCTGCAATATGTCCTTCACACACAGGAATCAATCAAAGTA
GCCAAAGCCGCGAGAACTACCGTGGAACAAAGCCGTCGCCTTGACGCGCTAGAAGACCGCGCTTCACGTTTGATAGATCAATTCT
CCAAAGAGATTAGAAAGGAAGAAACTGCGCACGAAGAAACGAAACGTACCGTGACGCAGATATTGCGCCCGTTAAAAAAATTCGA
AGGACTAAGAAATATAAAAAGCGCGCTCCAGACAAGAAACATTCCCCAAGAGGAAGTTATAAACCTTGCTAGAAATGATGATTTT
ATTGGAAAGCATGTACGGGACGATTTCGAAAACTTGCTGGAAGAGTATGGACGTGCATTTGGATCCCTAAATACCACAGATTTGC
ATGATCTTGTCGACGCTATCCGTTTTACAGCAACGCTTCTTTCCGGAGGAAAAAACAAACATGCACTGAACAGAGCGTTTGAAGT
CTTTCTCGATCATGCAGAACGGCTTCTGTTCGTCATAGAGACGGCGAAAGCAAATGGAGATGCGGATGGGTTCTCTCGCGCCAGA
CAAGAGGTGGAGGCAGTTATTCATGATGTGCAAAATAACAGAGGGGCTTCGCAAATACCTCAGTTTCTATTTAGCATCCGTGATA
ATCTCTACGAGCTGGAGTCTGAAGCTTATCAACGGGCAAGGAAAAAACAGCTATCGGATGTTATTGAAAGACTGACTAAAGACAT
TTTGAACGCCACGGAAAAACTGAATGTCCCAGAATTACGTGATCGACAAATAGCAGACTCGGCTACAGTGCTGGTGGAAAGGGTT
AAGAAAGAGGTGGCACAGCTAGATCTTACCAATGAGGAAAGATTTTCCCCGGAATTACTATCTCTAGAAAAAGCATGTAGCAACC
TTTCCGAGGCTGCGCATAATGCCATACTTGCCATAGAGGAAATCGAGTGGCTGGAAGCAACGAAGGCTGCACTCCTCCAAGATCG
TGGCAGTGCTAAGTTTTCGGCAGACGATTGGGAAACGATATATGATCTCCGGCCACATGTTAATAGTGCACATAAAGCCCACACA
GAACTAAAACATCTAGCTCAAAACGTTGTAAATGACTGGCTCGCAACCGGCACAAAACTTCTTCGATCGTTTTTCTCCTTCAACC
CATATTCTGGTAATGCAGAAGCAGGGCAGGGAAACAAAATAGCCGTAGCTCCTATAGCCGCTTCTACGTGGGTACATTCATTCCC
GTTGGTCAGTAAATACTATGAAACTTTATTTGGAGCAAAGGTAGAAGGCCTGACGACCGTTTGTGCAATTGCAAAGGATATTTTG
GAAACGGCAGAGGTTACCGTAGAAGGGTCTCAAATTGATACCAGGGGCGTAGTTTTCCGACTGTCCGAGCAATTGATGCGCATTC
CTGAATTATCAGAATTTGTTGACTTTTACTTGCAAGCGCATGAACAATTCATCGAATTTTCATCAAAGATTCAAACTATTAAAGA
AAAAGCAATCCTGTTGCAAGATGAGTTAAAGAAGGAGCTTCAATTAGCTGAAGAGAGTGCACAGATAGAGAGAAGTCCAGAAAAT
GCCAAAAAGCGGTTAGACCACGGCGAAGTGATACCCACGGCCCTCATCGCACTCAAGGAGCTGCCAGACAGTCTACACTCCGATG
ATGTCACTGTGTTTGACCATACCGCCTATCAAGATGCTTCTCACAGGGCGTTGCGAGATGTTAAATTATTTTTATCCGCAGAAAT
TAACAAGGCGGAACAACAAAAAAAGAACACAGACGGCAAAATTAAAAACATGCTGGAGCAATTATGGAGTGTAAGAGAAGAGGAA
GTCTTGCGTATGCAGAAGGGCTTAAAAAAGCTGAAAGGGTTACTATTGGGTCGCGTGCCCAAGGAAGTTGCCCCGGGCATCGAAA
GGGCAGAATCACTTGAAGATATAATTATCCTTTTCCTTAATCTCTTTCGCGAAGTGGAAGAGAAAAGCAATTCCACCGGTATAAC
TTCCGTGGACCCCAAATCATTGGAATGGATGTTTTCTTCGGCTCACATTCTCGATACATGTGAACTAGTAAACTCTATCGATGAA
CGCGGACCATTAGAAGATTTTAAAAAGCGATTGGATGAACTTAAACGACTGTATGAAGAGGCTGACGCGGGTCTGAAAGATCTCG
AAGCTTACTATAAGTATTTTTTATCAGCTGCCACTTCCTTTAAGGAAACCTCCGGAAAAGAAAAAGACACCAGCGAAGAGTCATG
GGGTGCATACGTCAATGCAGTTCACAATCTAAAGGAATCACGAGAACGCTTAGCCAGCCGGTTGCAAATGGTAGAGGCTCAAGTT
ACATCAACTCCGGAAATAAAGAAGTATCAAAAAGACATTTCAGAATATACTCCCATATTACAAAAGGCAGAAGAAAACATCATCA
CACTGACTAATGCCATGCGAACATTCGACGAAGTAATCTTTAAAGCAAAGACAGAGATGCATTTCAAAGCTTTGAAATCTCTCCA
TTTAGAATTAACTGCAATTTTTGAGTCTGTTCCAAAGTGGTACCGAGCAAAGTATGAAGATATTAAAAAGTTAATACTGCTAAGA
CTTGAACTTTACCTTGCTTACAAGGCCTTGGGTATAGCTTTTTCAGACGACATCATCCCGAGACTATCTTCTGCGGAGACTCCCT
TTTCTTCCAGAGTCCGGGAACTCGCCAGGGCTTCCAGCTATATGAGACAGAATAAAAAAGCGGTTATGACGGTTACGTTTCACCA
CCTCGGAACAATAGGCTCGGTCATGGTCACAGATAACGGAATGCCCCTGGAATACACTCTTTGTTATCGAACGGCCACTGCAAAA
GCTGCCGCCTTGTGGTCAGCTGCGTTATGCACATCTTCGGACCAAAGTGAGCCTCTCTCCTTGGAACTTGAGAATCCAAAAGACA
ACGAGTTAGTAAAAAAATTATCCGTCAGACAAAATGCAGCCCTCTCATTGGTCGCGGCAACTCTATGGATGAGAGCCAGGACTAG
CGATATAGAAGCGGGCTTCCATTCTTACACCATTTTCTGTTCCACCCGAACCTGGCCAAGCTCGAACAAACAAACAGATACTCCG
ATGGTCGCCGCATGTACAATTTACAGTGCCCTGACTTATCTGACTCTGTCAAAAACTTTAGGGTCAGCGCGGGATGTAATAATAG
ATGAGATGGGCAACTTTATACCCCACACAGACCTGGATAAGATCAATCAGAAAAATCAATTCGACAAAGCAGTTCGCATTTACTC
AATAACCGCTATTGATATTTTAATCCTTTTAACTGCATGCGAACCTGCACACCTAGTTTACTTTGGCAGGCTTAATTATTTAAAG
CAAACTGAATACATAGTAAATACCCTAGATGTTGTTTTATCTAGGGCAGTGCGCGAGAAGGTAGGAATAACATCTCTTGATCCTT
ACCAAAGAAAAACTCAGAATAAAAAAGCACTCTCTTTCCCTCTATCAAAAGGTAGGCAGCATTTCGATCCTTCAGCTGGGGCATG
TTTTGTGATAGACCCTAAAGACTGGCTTTCCCAACCTCCGCCCTCCTCTATTTTGAGATTTTTAGACCCCTGGAAAAGATTTCCC
TCTGTCAGGTCAGGGATATGCATGCTTGAAAAATTTTGTTCCGGAGAAGAATATGAGACAAGTGCCTTTCTAATGATTCTTGCCC
TCAAATGCATCCCGAGGAATATCCTGGAGGCACTATGGGTTTCCATGGGACCCCATGACGATATTGATGGAGAGCCCATCGATAC
CATGATTCATTATATGGTTTCCCGCTCATCGCCAAGCTGCGATTATGACTTTTCCCCAAACAGTGATCTCAGCCGAGTCTTCGAG
GCCCCTGATGACAAATTGTTCGGGACCGTAGATCGTGCTGGGATATATTCCCTTCGGGAGGAGCCCCATAAATCCATATGGGACT
TAACCGCTTTTGATATCGCAATTATTTCACTTCTTTTTGGAGCACCTGTCGTCATCGCATATGAAGCAGGGGAACTAAGCTCAAA
TAACGGTTTGGTTTTGTTTTCTACTATTTTCGATGGACGGAAAACAAATCCATTAGAAAGTGCGCTGTCCTCTACTACTGCCAAG
ACGGCAATTCAATTGGGAGACATTATTGCCGGTGATGCCAACCCCATAGAGAACGCCTGTCTGGCCTCCCAGATAGCAGAGTTAT
CAACCGTACTGTCCAGCAAACCCCTCAAGCATGCTCCGTCGTTACTTTTTGCCATCGACAGGGGAAACAAAATAACTCAGATATT
AGTTTCTCTCGAAACAGCTCCATCGCCATTCGTTTTAAGGCTCCATGGAGACCCCAAATATGAGGACCTCCCAGTAAAATTTTCA
GACACGGGGACGTTCTCGTCTAATCTCGGGGATCCAAATTATGATACCAGTCTATACTCTGTTGCTAGTCCAGTGCAGAGCGGAA
CAGAATCAGAATTATCGGAATTCGAACAAGTACTCTCAGAAACTTGTGATTGGAAAATTCAGAGTCCGTCAATAACTTCTGACCC
GTCTGATAATTTCGCAGTTATTACAGACGAGACTAACCTTGAGCATCCCACAGACCCGGAACCAAAAGATCGAAGCCACCTGGCA
ACAACACGTAAAACTAGCGAACACGAGGGCCCGCCAGAAAGCATAACAAGCCACAGAGATAAACATGACACCGCGGAAGCCAAAC
AGACACATTGCTCAGATAATGAGGATATAGCTCACTTCCCAACATACAGTTACACGTGGGACAAGGAAGTCGATCCCCCGGAAGG
AGAAAGTTCTGACGAGGATTTTACACCGGCATATGTCAAAGGGTCGCACAAACTTCACGAGGAATCTTCTCGCAGAGCCGATGAA
CATAAGAAAGGGCGCATTGAAGAGGGAGAAAGGTTAGTAAGGGATACTCAAGCTTTAGATACTTATTTGAGAAAATTGGACAACG
CATTGCGGGATGGAACGATCACACCCTATCTGAAATCCATTCAGACTCCACCTGATTCTAGTTCGGAAGAGGATGATGCTGTAAA
TGAGATATCTTCCGGAACCTCAATAAATGACAAAGCCCCATCAGAAGACCTTGCAAGCGATGCGGAAGAGTCTAAAGCGGGTGAA
AGAGAAATTCTTAGTCCCTTGCCACCCACCGCGCGCAGTGATATTTATGCATCGCTGTCAGTCACCTGCCGCATGGTGCTTAAAG
GAATAAAACACGCGAGAGATTACGCTTTAGATCAGCAGGCCCGCATAGCGGATGAAACACACAAAATAAAAATAATACTGGGTTA
AATTGGTATATTTATTGCGTTTTGCTTAGCGTCTC

>UL37
TGCTGCTTCGTGAGTCCCTCTGACCAAGAGTCGCATGCACGTTACTGGTTCACTTTCGGTAGCTTTTATCAAGCCCCCACCTAAA
GATAGAAAAAAATGGGGGTACGCTCGTGCCTCAGTTGTTGCCAGAAGCAATGCATCAAAATTGGCAAGCTTTTCGGTAATGGCGG
AATCCTTATAATTCGCGATTACGAACGCGCGTAGTGCCTCTACTCCGCAATTATTTCGGTAGTGAGCCGCGCCGGCCGCAGCTAC
AAAACTCAAAGACACCGCTGAAAGCTCGAAGGATTTGACAGCACTCTCCCCAGTAACCGATAAGCCGCAGGCGCGGTTCAAATTG
ATAAATGCAGTCTCTAAATCTCTCACAATTTGCTCGCGGGCGGGTCGTGATCCCAAAATCTTGAGTAAGTCTTTAACCCCTTTCA
AAAAGAGCATTAGAGACCCAGACTCTTCGACCCCTGGCTGGCAGAGGTCTATGACAGTTATTTGGGTTAAAGAAATTACATCATT
TGTCCTCGTAGATTCAACATCGAGAAGAGGTTTAGATGCGATGGCCCTTGTATCTTCCTTTCCAAGCGTTGCATTATTTGTTTCT
TCGGGGTTACGTGGCCCCGAGCCAATAAAGAATTTCCCTGGCCTCGTTCCAAGACCATGAGAAAAATTCCGGCCAGTTGTTAGGC

GCCCCAGCGCGGCTGATAACAGATTATGCTGGTGAGTTCTTCGAGGAAGTACACTTTGTAAATAAGGATGAGTGCGAGTGTTTGC
GAATCCACTCAGCCCTCTTTCTGGCGCTCGAATTTTATCGTTCGATCCATTGCGAGCAAGTAATCCGGACATGCTTACGATGTAA
CAGCGCGGAAAGGAAGTTCTGACGCTTATTATTAGCCTGGAAACACTACCTCTAATCACACTTCTGAAAAGAACTGTGCTGTGAA
AGCATTCAGAAGCGTGTCTATAAAGCTACCAGCTGGTTCCGTGAGTTGACGCGACTATTGAAACCATGCAGGCCATCACGGATAG
CCTCAAGGCTTGCCTTGAAGCTGTACACACCGACTGCAGTAAAGTTCCGGGTGTTACTTCAGCCCTCACCGGAATTCTAATCTCC
AGAAACAGAATTCCACTTGAAGATTTGGAGAAAGTAGAAGCTAGGAAAAACATGGTTGAAACTATAATGCTCGCATGCACAATGG
CGCCTCCGGCAGTAGCGGACGTTACATTGAAATTATTTGCACGCTCTCTTATTAGCCGCATATCTGTTCCATGGAACTCTGGAGA
AGATTTCAGAATAATTCAATATTTGAGAGAAGCATTTACTGGACTGCAACATGACCTAGAAACAGCTGTACGTAAAGAATTTCCT
GGGTTAAATGATCCAAATGTAGAATATTGGAGTCTAATTAGTGGCTGGTTAACGGAATTTGGGAGTTTGGCAAGATTGGTAGTTG
ACGAGCGACAATTGTTCGAGACATCCGGAGACCGACTCGAGCCAACCAAACTCCTATCTCCTTTAATTGAAAATTATCCACTTTT
ATACGACCACCAAATGGTTCAAGACGGGATACAGTATCTTGGAACAAAGCTACAAGGCTTAGTGGGATACCATGCAATGTTAAAT
TATATCACGGCTTCAACCGGGCTACCCAAAACTAAAGCGCTGATGACCCTCTCTATGGTCAGAGAGTATTTTGAACCTACAGTAA
AACCCCCTTATCGGAAGTTTCCACAGACGGCGCTAGCTGCCTTTGATCTGGACACAGATTCGGCCAAGTCATGGATCGCTAAAGA
TTTAGACTCAAGATACTTTTCTATCTCGCAACTGTACGCCAATGCTGCATACAGTCGAGACCCACTACTCGTGTTAGGAGATCCT
CCGAAGACGCATGTTCCTGGACATGTAGTTTGCTGGCGGCACGACTTAGATAGTCTATTAACTATTCCCATTAAACGCCTCGCCC
TAGATGTCCCAACCATCGTACAGCTTTTCGAGGACGCGCCTAATAACTTGACAGAGGAACAAAAGGAGACCCTAACTGCTTATAC
AACAAGTACCGCCGATATGATGGGGCGTTCGCCGGACTTGGCCGCTACCGGTGGCCCAGAGGAAGTAATTCGTACTTTAGTGCTT
AGGGGGTTTACCAAGAGCAACTGCGAACTGTATATCCAAAAGGCTGAGGCGCTTTGGGGTGAAGATACTGACGGAACAATCATTG
GAGATTTCCTAGGAGCCGTTGTATATACATCACTTATAGGACTCGCTGCGCACTCAATGTATACATATAATCCGCGGACGCTACA
GTATTCAGGGCATTACGCTGGGCTTATATCTGATTGGGGAGACCCATATTCGCATCTTCTCCGGCGGGTAGGGGTGTCTGAAAGC
GAACTACTACTACCACTACGATCATTTGCCCCAGCCCCATCAAAAAATGCTCTAAACATCATTAGGGACACTCTGGCCTCATTTA
TAGATGGGAGCGATGGAGGACCCGAGGCGCTTGGAAATTTGGGGGTGTTCAATTTTGTTCAACACCTCATATCTGACTTAGGAGG
AAAGACCCAGGAAAATTTGGATTGGCTCGAACGCAGGGAAAAGGCCAAAGATAAGGGGTATGCAAGTGCCGACAAACAATCCTAC
ACTTATGGATTTGACCAAACCGTAAAATTAGTCCAGGCCATTGAAGTTGGGGGCGCGGAGCCTGAAGAGATATGTTCGCGTAATG
GAACGGACCCGAAAATCGATATGCAACTCTTTTCCAATCTTTCCGTAGCGACCGTATTGAGAGACATTGTTTACGCAGTATCAAC
TTATCCATTCACCGTTGACGCGGTACAACACGCGGTAAAGCTTTTATCGGGGATAGAGATTAGGGCTCTCAGAAAACCGTCTGAT
ACTCAGAAGTACAGGCGAGCAATTTTGGAATTTCAGTTAGCGGTCTCCCCATTTATTACAAAGACCAATCCGACAGAAAAGATAT
CCATCGAAAGTGCAAGCGCGATAGAAAAGGCATTGCTTAAAATTTCTGACGCATGCGACAGGAGCCTCGACACTCTTCCAGAAAC
ATTTAAGAAAACAGTCCAGCCTGTCCCTGCCGTCGATACCATGAGATCAAAGTTCATTTCCCAGGCATTCGAGCAAGCTGCAACG
AGAGACCTTGAGGAGGTTTCCAAGACTCTAAATAACGCATCTATGCAGATATCTACTGCAGTGCAAATACTACTTTCTAAGACTG
CTAACATAAGACCACTGCTTTCACAGTCAGCAAATGCGGATTTTTTGGGCAATGTTTCTTTCTCAAGTTGCCGTGACAAAAAAGA
ACTAGGACACTGGAAGGACGATACTGTAATGCAGGCCATATGGAAAGCCGTCAATTTTGCAGAAACTGCAATCCGACACACAGAA
AAAGCTATTCGGGCATTAGATTCCACAAAGGTTAAAACGCATGCACTTGAGACCACATACTGGCAGCTTGCGGGTTATGCTCAAA
CACGCTCTCACATTTGTAGGAAAGTAGCTGAAGAATTATCTGGCATAGCGGACGCGAGGAATACTGTCGTCCAGCGTTCGTGGAC
ATTATGTCGTTCACTCGCAGGCGTCGGGGTAAAAAGGGCCCGCGATGTGGTAAAAGCCTGGAAAACTTTTTCAGATAATGAAGCT
GCTCTCGAGACATCTTAATCCAAAACCCTGAAGATGTAGTTAAAGTTATTGGCGCGGCATCAGCTCTGCTAGATATAGAGGAGGG
AGGGACCACGACCCTTGAAGGAAAATTGGAAGAAATTGAAATTTCTGCCCCTTCTAAAGAAGATACAGCGTCAGCTCTTACTCTT
TTAGGGCCGCCTGATGCCAGGGGATATCAAGATACATCATCAAATCAATTTATTACCGAGGCGTCAGACATTACTTCGTGGGATA
AAACCAACAAAGCCCCGCTCTATGTTCGACACTCGACAAAAGAAGGGATAGACAATGGGTGTGTAGACACACTTCCAGACAAACT
AATCTCATCAGAGACTCTTCTGACAGTTTCCGATCGAGTTTTGGAAAGTGTAAATGTAAAAATGTTTGAATAAAGAAAGACTAAT
GAGTAATTGAGTGGTGTGTATTTATCGCGTTGTTGTGTGCG

>UL38

GACTGTATGCAGCATTGGCGTACAGTTGCGAGATAGAAAAGTATCTTGAGTCTAAATCTTTAGCGATCCATGACTTGGCCGAATC
TGTGTCCAGATCAAAGGCAGCTAGCGCCGTCTGTGGAAACTTCCGATAAGGGGGTTTTACTGTAGGTTCAAAATACTCTCTGACC
ATAGAGAGGGTCATCAGCGCTTTAGTTTTGGGTAGCCCGGTTGAAGCCGTGATATAATTTAACATTGCATGGTATCCCACTAAGC
CTTGTAGCTTTGTTCCAAGATACTGTATCCCGTCTTGAACCATTTGGTGGTCGTATAAAAGTGGATAATTTTCAATTAAAGGAGA
TAGGAGTTTGGTTGGCTCGAGTCGGTCTCCGGATGTCTCGAACAATTGTCGCTCGTCAACTACCAATCTTGCCAAACTCCCAAAT
TCCGTTAACCAGCCACTAATTAGACTCCAATATTCTACATTTGGATCATTTAACCCAGGAAATTCTTTACGTACAGCTGTTTCTA
GGTCATGTTGCAGTCCAGTAAATGCTTCTCTCAAATATTGAATTATTCTGAAATCTTCTCCAGAGTTCCATGGAACAGATATGCG
GCTAATAAGAGAGCGTGCAAATAATTTCAATGTAACGTCCGCTACTGCCGGAGGCGCCATTGTGCATGCGAGCATTATAGTTTCA
ACCATGTTTTTCCTAGCTTCTACTTTCTCCAAATCTTCAAGTGGAATTCTGTTTCTGGAGATTAGAATTCCGGTGAGGGCTGAAG
TAACACCCGGAACTTTACTGCAGTCGGTGTGTACAGCTTCAAGGCAAGCCTTGAGGCTATCCGTGATGGCCTGCATGGTTTCAAT
AGTCGCGTCAACTCACGGAACCAGCTGGTAGCTTTATAGACACGCTTCTGAATGCTTTCACAGCACAGTTCTTTTCAGAAGTGTG
ATTAGAGGTAGTGTTTCCAGGCTAATAATAAGCGTCAGAACTTCCTTTCCGCGCTGTTACATCGTAAGCATGTCCGGATTACTTG
CTCGCAATGGATCGAACGATAAAATTCGAGCGCCAGAAAGAGGGCTGAGTGGATTCGCAAACACTCGCACTCATCCTTATTTACA
AAGTGTACTTCCTCGAAGAACTCACCAGCATAATCTGTTATCAGCCGCGCTGGGGCGCCTAACAACTGGCCGGAATTTTTCTCAT
GGTCTTGGAACGAGGCCAGGGAAATTCTTTATTGGCTCGGGGCCACGTAACCCCGAAGAAACAAATAATGCAACGCTTGGAAAGG
AAGATACAAGGGCCATCGCATCTAAACCTCTTCTCGATGTTGAATCTACGAGGACAAATGATGTAATTTCTTTAACCCAAATAAC
TGTCATAGACCTCTGCCAGCCAGGGGTCGAAGAGTCTGGGTCTCTAATGCTCTTTTTGAAAGGGGTTAAAGACTTACTCAAGATT
TTGGGATCACGACCCGCCCGCGAGCAAATTGTGAGAGATTTAGAGACTGCATTTATCAATTTGAACCGCGCCTGCGGCTTATCGG
TTACTGGGGAGAGTGCTGTCAAATCCTTCGAGCTTTCAGCGGTGTCTTTGAGTTTTGTAGCTGCGGCCGGCGCGGCTCACTACCG
AAATAATTGCGGAGTAGAGGCACTACGCGCGTTCGTAATCGCGAATTATAAGGATTCCGCCATTACCGAAAAGCTTGCCAATTTT
GATGCATTGCTTCTGGCAACAACTGAGGCACGAGCGTACCCCCATTTTTTTCTATCTTTAGGTGGGGGCTTGATAAAAGCTACCG
AAAGTGAACCAGTAACGTGCATGCGACTCTTGGTCAGAGGGACTCACGAAGCAGCAGCTGCCGGAGATCTTCTGAGAAAAAATGT
GCCGTACTCCACTTTCCATGCATGTCTGTTACTTGACCTCGATGAAGGTTTTTGCCCACCAAGGACAAAGATCAATGCAGAAGAC
GGTGTTTACTTTCTACATTTATTATTCCTGTACTCTACCGATATGGGACACCCGAGGTACGAACTATATGTGGCAAAAACATCTC
TCCCCGAAAGTTGCATGAGAGACATTTTAAATGAAAGATTTACACGACGCAGAATAAATAATACTATAGCTTCCTTAACACACAT
GCAAGCTCAAGAACGCCAAGCCTTTCCCCCTTTTAATATTGCAGAAGCTGCGGTTGCCGCGCGAATGGATCTGAGACGAGAAAGA
TCAAGACATCCACTGGATCCGTTAAGGTCGTGATCAACGTCTTAGGGATATGCAGATAACGTCCACTGGAATACCAGACCATTCG
GCTCTAGTATTCGCCGCATACTGCATGCTAGGGACGGCGGGCAGTATTGGCGGAGAGCCTTCTTTTTCTACTATCCAGAAAAACG
GGACTAGTCTAAAGTATGTTCACATTCGCGATTTTAACTTTAGGGGCGGCCCATGGAATGTTTGTATGTAACACACCCACTTCTA
AATAAATACAAGTCTGTTCACCCCATGTGTCTATTCAGTTCTCATTCGAGGCAAGG

>UL39

TTGAGTTTTGTAGCTGCGGCCGGCGCGGCTCACTACCGAAATAATTGCGGAGTAGAGGCACTACGCGCGTTCGTAATCGCGAATT
ATAAGGATTCCGCCATTACCGAAAAGCTTGCCAATTTTGATGCATTGCTTCTGGCAACAACTGAGGCACGAGCGTACCCCCATTT
TTTTCTATCTTTAGGTGGGGGCTTGATAAAAGCTACCGAAAGTGAACCAGTAACGTGCATGCGACTCTTGGTCAGAGGGACTCAC
GAAGCAGCAGCTGCCGGAGATCTTCTGAGAAAAAATGTGCCGTACTCCACTTTCCATGCATGTCTGTTACTTGACCTCGATGAAG
GTTTTTGCCCACCAAGGACAAAGATCAATGCAGAAGACGGTGTTTACTTTCTACATTTATTATTCCTGTACTCTACCGATATGGG
ACACCCGAGGTACGAACTATATGTGGCAAAAACATCTCTCCCCGAAAGTTGCATGAGAGACATTTTAAATGAAAGATTTACACGA
CGCAGAATAAATAATACTATAGCTTCCTTAACACACATGCAAGCTCAAGAACGCCAAGCCTTTCCCCCTTTTAATATTGCAGAAG
CTGCGGTTGCCGCGCGAATGGATCTGAGACGAGAAAGATCAAGACATCCACTGGATCCGTTAAGGTCGTGATCAACGTCTTAGGG
ATATGCAGATAACGTCCACTGGAATACCAGACCATTCGGCTCTAGTATTCGCCGCATACTGCATGCTAGGGACGGCGGGCAGTAT
TGGCGGAGAGCCTTCTTTTTCTACTATCCAGAAAAACGGGACTAGTCTAAAGTATGTTCACATTCGCGATTTTAACTTTAGGGGC
GGCCCATGGAATGTTTGTATGTAACACACCCACTTCTAAATAAATACAAGTCTGTTCACCCCATGTGTCTATTCAGTTCTCATTC
GAGGCAAGGGAGCAAGGATCCCTCTCAGGGCAATTATAGGATTATCTCCTGAGCAGCCCGGAGCCATGCTTTCGTTTTTCGATGA
ACTCGAAAGCATTGCCCATCTAGAAGAGTGTCTGCGGGACATCGAAACGCGTATGCGAACGTCTGGGTTTGATCCTAAATCCATC
ATTAGAGACGATTTTGACATAAGGGATGGGCGCGAGGGCCTCGCCAACCGGATCTCGATGCTTGTGGACGGGTTTAAAGTTTCCG
TGCTATTTAACGTAGAACTATACCGACTCCTCGCGGAACTCGTTCATTTAAGAATTCGCACAAAAGCAGTCTCATTTTCGGAGTG
GCTAGATACTAGGGGGCTATCTGCGGACTGTAAACAATTTATTTTGGAAAATGCAGAACATATTTCTGCGGTCGTCAAAGACTTC
TACAACGGAACATACCATCAGCTTGCAAGAGTCGGTCTTCAGTCAGCTCAAAAATATGAATCGCTGTATTTGGGAAAACTTGGCA
ATGGAAAGCTGGAGAGTATGGGTCAATTTTTTACTAGACTCTCTGCTGAGGCTGCCAGAGGAGCTTTTAATATTCCCCAATTTGC
ACAGGCATTGGAAGTTGATGGCCAGGTTACACCGGCAGACGTCTTTACTAGGTTTTTTACGTGTTTGAGTAGCCAACTGATCGTA
CCTCCTACCCCAGTAATGCTTTTCGGGGGTACGTCTCTGGCTGCTTATGCCAGTTGTTTTTTGATAGATTCATCTGGCAGAAATA
CAGAGAGGCGTTCAATGTCGTTGCCCGAAGAAGTAATTCCGATCATGAATAATCGCGGAGGAATTGGAATATCACTGCCGTGCCT
TGGCGCTCCCCAAAATGAGTCGGGGTGCCTGGGATTTCTTAAGGCTCTCGATTCGCTAGTAGCCGCGAGCAACGGCGCCGCAAAG
CGGCCAACAGGGCTTTGTGTTTATTTCGAGCCATGGCATTGCGACACATTAAAGATTCTGAAAATTCGTGGCAGCTCGGCCGGGA
ATGAAGAGTTCAGATGCGATAACATTTTTACTGCCATTTGGATGCCTGATCTTTTTATGAAAAGACTGCAGCAGCCAGGAAGTAA
ATGGACACTGTTCGACCATAGGGGGGAGCATCTGTCTAACTTGTTTGGAGAGGAGTTTGAGAAGGAGTATGAACGTCTTGAAAGG
AAAATGTTGGAGTGGCCACTATCCCTATCACCGAGATTATGTTCAGATAATCAAAAGTGCAGTTCCACCGGCACTCCATTTGTCG
TCTTTAAAGACGCCTTCAACCGTCACTATTTTTACAATATGCAACACCGAGCACTGAAGTGTTCTAACCTCTGTACAGAAATTGC
CCACATGGCAGACTCCGATACTGTTGGAGTATGCAACTTGATTAGCATCAATTTGGCGGCAATGGTCAAAAATGCTCGCGCAGGA
CTTGCTGGGAAACCCGGCAGCTTCTTCGACTATGAACTCTTAAGGGAAACGGCAAGAACTGCAACAATTTTCGCCAATGTCATGA
TATCACTTGGAAACATGCCAAGCCAACGGGCACAGAGCGGGAACCACAGGTTACGATCCCTCGGTGTTGGAGTCCAGGGACTGCA
CACAGCCTGCTTAATGCAAGGTTTTGGGATGACTAGCGTTGAGGGCTTTGAATTTAATGACACTGTGTTTGAGCTACTTGCCTTG
GAAACAACTGGAATCAGCTGTAGATTGTGTGAGCTGGGGTTGCCTCCGTTTGATAAATATAGAGAAAGCTATTATGCCTTAGGAT
GGCTGCATATTGATGGGTGGCCCAATACGAAGCTGAGATACAAAAATGAGTGGGATAGTCTAAGACATAGAATAGATCCATCTGG
TTTGTACAACTGCCAAACAGTTGCTTTGATGCCAACCGCCAGTTCTTCCCAAATTACAGAAGTCAGCGAGGGGTTCCAGCCCGTA
TTCGGGAACATGTTTAGTAAGATCTCTACTACAGGAGAGGAAGTGCGACCCCATTTGGCGCTAATGGATGCCATCGACGAGCTGT
ATTCTGACCAGGGAGAAAAACAAGCATTTCTAGCAAACCTAAAAAAACATCAGTGGTCTGTGCGCGCTGCTCTGGGAAGTGCTTG
GTCAGAGTCTCACGTCCTCGCGAAATTTCAAACCGCATTTGAACTCGATCAGGAAAAGTTACTATACTTAAGCGCACGCCGCGCT
CCATTCATTGATCACTCTCAATCAAACACCCTCTACATCACCGAAGACGTGGACGGTACTTTAAGTGCGTCACGTGTATCCCGCC
TACTGCAAGTTGCCTTTAAATACGGGCTCAAGACTGCCATGTACTACTGCAAGGTGCGCAAGATCACAAACAACGGAGTGTTTGT
CGGGTCTTGCGGAGACAGCCTCATTTGCTCTGCGTGTCAATAATCTTGCTAACATGGCAGAATATCAGTCCACAGACTATTATTA
CTTGCCCCAGTGTGATGACATTCGCGAGCTCCGAGACCTGAGCATAGCAAATAACTGGAACGAATTTGAACTATGCTACAGTCGC
GATGAGAAAGATGTTGATCTACTAACTCCGGAAGAACTAGATTTTTACAAATTGGTGTTTGCATTTCTGGCAGCCGCAGACGATT
TGATAAATTTGGATATCGGGAATCTCATGACTTTGTTCCTTCACAAGGACATTCAACACTATTATGCTGAGCAAATAAGAATAGA
GACGGTCCACTCTAGAACCTACAGCTTGGTGCAAATGATCCTTTTCAAGGGAGACCTTCATGCGCGAGACAGATATGTGACAGAA
GCTATTAAAGATCCCGCCATAAAGAAGAAGATAGACTGGCTCAACCGTGTCCAAGTAGAAACGGATCTGACACTTCCTGAAAAAT
ACATTCTAATGATTCTCCTTGAGGGGATTTTTTTCGTGGCGTCCTTTGCCGCGATTGCGTTTCTTAGGAGGCGGAATATTTTTGT
GGTGATGTGCCAGTCGAACGACCTCATCAGTCGCGATGAAGCGGTGCATACAACAGCATCATGCCGCATTTACAACAACTGGCTC
GGGGATCACAAGAAACCATCGGCTAAACGAATTCACCAACTTTTCAAAGAGGCCATTGAAATTGAGTGTAACTTTTTAGCATCTA
GAGCGCCTCAAGCTTCGAGACTCATAGATCTAGAAGCCATACAGAGTTTCGTGAGATATAGCGGAGATCGACTTTTGACCGCCAT
CGGGGTTCCCACTATTTTTGATGAACCGCCCCCTGATCCGTCGTTTCCCCTTACCCTGATGTCGATCGACAAAAATGTCAATTTT
TTTGAGCACCGGAGCACCGCGTATAATGGCAACGTTATCAATGACTTGTAAAAGAAAGTTCAAACAGTCAAGCATTATATAAATA
AAAACTACTCAGACAGATATTAACATTTTATACTTGTTCTTGTTCATTTTTT

>UL40

AAAATGCTCGCGCAGGACTTGCTGGGAAACCCGGCAGCTTCTTCGACTATGAACTCTTAAGGGAAACGGCAAGAACTGCAACAAT
TTTCGCCAATGTCATGATATCACTTGGAAACATGCCAAGCCAACGGGCACAGAGCGGGAACCACAGGTTACGATCCCTCGGTGTT
GGAGTCCAGGGACTGCACACAGCCTGCTTAATGCAAGGTTTTGGGATGACTAGCGTTGAGGGCTTTGAATTTAATGACACTGTGT
TTGAGCTACTTGCCTTGGAAACAACTGGAATCAGCTGTAGATTGTGTGAGCTGGGGTTGCCTCCGTTTGATAAATATAGAGAAAG
CTATTATGCCTTAGGATGGCTGCATATTGATGGGTGGCCCAATACGAAGCTGAGATACAAAAATGAGTGGGATAGTCTAAGACAT
AGAATAGATCCATCTGGTTTGTACAACTGCCAAACAGTTGCTTTGATGCCAACCGCCAGTTCTTCCCAAATTACAGAAGTCAGCG
AGGGGTTCCAGCCCGTATTCGGGAACATGTTTAGTAAGATCTCTACTACAGGAGAGGAAGTGCGACCCCATTTGGCGCTAATGGA
TGCCATCGACGAGCTGTATTCTGACCAGGGAGAAAAACAAGCATTTCTAGCAAACCTAAAAAAACATCAGTGGTCTGTGCGCGCT
GCTCTGGGAAGTGCTTGGTCAGAGTCTCACGTCCTCGCGAAATTTCAAACCGCATTTGAACTCGATCAGGAAAAGTTACTATACT
TAAGCGCACGCCGCGCTCCATTCATTGATCACTCTCAATCAAACACCCTCTACATCACCGAAGACGTGGACGGTACTTTAAGTGC
GTCACGTGTATCCCGCCTACTGCAAGTTGCCTTTAAATACGGGCTCAAGACTGCCATGTACTACTGCAAGGTGCGCAAGATCACA
AACAACGGAGTGTTTGTCGGGTCTTGCGGAGACAGCCTCATTTGCTCTGCGTGTCAATAATCTTGCTAACATGGCAGAATATCAG
TCCACAGACTATTATTACTTGCCCCAGTGTGATGACATTCGCGAGCTCCGAGACCTGAGCATAGCAAATAACTGGAACGAATTTG
AACTATGCTACAGTCGCGATGAGAAAGATGTTGATCTACTAACTCCGGAAGAACTAGATTTTTACAAATTGGTGTTTGCATTTCT
GGCAGCCGCAGACGATTTGATAAATTTGGATATCGGGAATCTCATGACTTTGTTCCTTCACAAGGACATTCAACACTATTATGCT
GAGCAAATAAGAATAGAGACGGTCCACTCTAGAACCTACAGCTTGGTGCAAATGATCCTTTTCAAGGGAGACCTTCATGCGCGAG
ACAGATATGTGACAGAAGCTATTAAAGATCCCGCCATAAAGAAGAAGATAGACTGGCTCAACCGTGTCCAAGTAGAAACGGATCT
GACACTTCCTGAAAAATACATTCTAATGATTCTCCTTGAGGGGATTTTTTTCGTGGCGTCCTTTGCCGCGATTGCGTTTCTTAGG

AGGCGGAATATTTTTGTGGTGATGTGCCAGTCGAACGACCTCATCAGTCGCGATGAAGCGGTGCATACAACAGCATCATGCCGCA
TTTACAACAACTGGCTCGGGGATCACAAGAAACCATCGGCTAAACGAATTCACCAACTTTTCAAAGAGGCCATTGAAATTGAGTG
TAACTTTTTAGCATCTAGAGCGCCTCAAGCTTCGAGACTCATAGATCTAGAAGCCATACAGAGTTTCGTGAGATATAGCGGAGAT
CGACTTTTGACCGCCATCGGGGTTCCCACTATTTTTGATGAACCGCCCCCTGATCCGTCGTTTCCCCTTACCCTGATGTCGATCG
ACAAAAAATGTCAATTTTTTTGAGCACCGGAGCACCGCGTATAATGGCAACGTTATCAATGACTTGTAAAAGAAAGTTCAAACAGT
CAAGCATTATATAAATAAAAACTACTCAGACAGATATTAACATTTTATACTTGTTCTTGTTCATTTTTT

>UL41

TAGCTGAGACAAAATTGACGCAGACGATACAATACTTGTCTGGTCGCCGAACCATTTCGCGTTAAAGGTTAGGCGTTGCTCGTCT
TCTGCGGGACGAGAACGTGACATACTCTCTTGGGAGTGGGACTGTAATTGAGTATCTTTAGGGATCTTCGAAAGCCACTTTGTAA
TTTCATTGAACGAATACTGTTCCAGCGAGGCCCGACATGCCTTATCGGAACGAGGCAGGTATATAGGAGGTCCATCCGATGAAGA
GTACTTGAATTTCGTGCTTATTTCCGTATTTTGGGGGCACGACGCGCTCTCTTCTATCTCGCGACAGACTTTCAGTCTAAGTACA
GCCCTAGTTTCATCGGATTTGAAACTTTCTGATATAGTAAATGTGGCTAAAGTTGGTATTCCGACAGTTTTTTTCTTAGAGTCCC
ACAGAAAATCCAAGAGGTCTCCAGGGCGCTGGTGTATAGTATTATTGAACATGATAGGTAGTATACTCTCATCGTCTGATAAATA
TTCAAATGAGTCGAACAATTGCCACTTGATTTTTAAGAAAGCCATCCCAAACGGAGCGGAACAGGAAATTACTATGCAGTCATTC
ATAAAAGCCAGCACAGAGTGTTGCAGAATCCCACGCATGTTTTTGGCAATGGGGGAAACCACCGAAATCCCATGCTTTCCTAAAG
TTAGTCGGAGCAAACATTTACTTCTTGTCGCAGTATCCAAATCCATCTTCCAACGAACGTAATTCGTGTGTACGTCCGTTCCCTT
CCCGGTGTCTCTGAACATCTCCTAAACATCCGCGCGGTTTATATACAACTGTAGCCTGTGCATAAAATTTTGGGGCTACCACCAT
GTCACACCTTATGTGTAATGCGATTACGAGGGTGTGGATTGCCCAACCCCTGTACATACAAATACCGCCTTTCAAAGCGCCCTAA
ATATTAGTTTGCTGATACGAGACTTGCGGGCTTCTGTCTTTTCATTTGCTCTCCTCGAAATATACCGATATGGGAATCTTGGGAA
TGCGTCGCTTTATTCGAAACCATGGACTCAGTGTTCATTTATCGTTAAAAATTGAGAAGGGATTTTATCTTCCACTGGCAGTCGA
CACATGGAATGTTCTTTTTCTAATTTTGCGAAGAATTGACCCGGACGGAAATATGGGATGCATGGAGAGAACTTTTAAAAGTTTG
GTGTTCTTATTTTCATTACTCAGTCGGAAATCGTGCTATCCTGTATTTGTTGTTGATGGTGGGCGTCGGCGCGAATACAAGGAAC
CTAAACACCTTAGCCAGGAAGAATCTGTTCCGACTGCACCTGCCGCGCGAAGATGCAGCGATATTAAGGAGCCCAAGGATTGTAT
TTACAAGACACGTCGAAAAATACGTCCTCCGCACTACAAGTTATGTTGGGACTTAATAACTACCGCAGGATTTGATACGGTCTAT
GTGCAAGGAATGGAAGCCGACTATGCATGTGCCAATCTATTCCATACCCGCACGGTGGCGTACGTCCTGTCTAGCGATAGTGACT
TAGTATTCATGGGTTGTGATGTGATTACGGATCTGACCCCAGCATTTCCAGTCGCTATCTTTTATAAAAGTATTATTGAATATTT
GGGCATGTCCGGGGAAGAGTTTGCCAATGCATGTGTGGACTGTCATACTAACATCCATTCTACCGATTCAATATACTCATTTTCT
GCCAGGTTGGCGGAATGGCGTGACCACGAAAAAAATGAAAACGATACCAAAAAGTGTGTTGAGCAGCGCGGGGTCTTGCCTGCGG
ATAGTGATGTTTTTGGAATAAAAACAGATGTGCGTGAACCTTTGCCTGGACGCTATGTATATGTACTGACTCAGTTTCCTCAAGA
ACCTCCCGACGATGGTAACTCGGACATTGATTACCGCCCTCAGGATCCTTCGGTAAATAATGGCGCCAGAAATGAAAAAAATGAC
GAGTTTCTGTCCTACGTTTTGAGTGTCATCGGCCCACGAGAAATAAATGGCAGACACAAACTACTTAAGCGCGTGCCCGTGGTTC
AGGAACAGCTCGACATGATCGCTGTATATGAAGTATTGTGTAAATACTTTTCTAATGATGCCCCGAAATATTTTTCTGAGCTTAT
TAAGAAGGGGCATCTAAAGCCTCTGCCGTCATTACAAGAAGTTATTTGTGGAAAAAATGAACAAGAACAAGTATAAAATGTTAAT
ATCTGTCTGAGTAGTT

>UL42

GCAGGCAAGACCCCGCGCTGCTCAACACACTTTTTGGTATCGTTTTCATTTTTTTCGTGGTCACGCCATTCCGCCAACCTGGCAG
AAAATGAGTATATTGAATCGGTAGAATGGATGTTAGTATGACAGTCCACACATGCATTGGCAAACTCTTCCCCGGACATGCCCAA
ATATTCAATAATACTTTTATAAAAGATAGCGACTGGAAATGCTGGGGTCAGATCCGTAATCACATCACAACCCATGAATACTAAG
TCACTATCGCTAGACAGGACGTACGCCACCGTGCGGGTATGGAATAGATTGGCACATGCATAGTCGGCTTCCATTCCTTGCACAT
AGACCGTATCAAATCCTGCGGTAGTTATTAAGTCCCAACATAACTTGTAGTGCGGAGGACGTATTTTTCGACGTGTCTTGTAAAT
ACAATCCTTGGGCTCCTTAATATCGCTGCATCTTCGCGCGGCAGGTGCAGTCGGAACAGATTCTTCCTGGCTAAGGTGTTTAGGT
TCCTTGTATTCGCGCCGACGCCCACCATCAACAACAAATACAGGATAGCACGATTTCCGACTGAGTAATGAAAATAAGAACACCA
AACTTTTAAAAGTTCTCTCCATGCATCCCATATTTCCGTCCGGGTCAATTCTTCGCAAAATTAGAAAAAGAACATTCCATGTGTC
GACTGCCAGTGGAAGATAAAATCCCTTCTCAATTTTTAACGATAAATGAACACTGAGTCCATGGTTTCGAATAAAGCGACGCATT
CCCAAGATTCCCATATCGGTATATTTCGAGGAGAGCAAATGAAAAGACAGAAGCCCGCAAGTCTCGTATCAGCAAACTAATATTT
AGGGCGCTTTGAAAGGCGGTATTTGTATGTACAGGGGTTGGGCAATCCACACCCTCGTAATCGCATTACACATAAGGTGTGACAT
GGTGGTAGCCCCAAAATTTTATGCACAGGCTACAGTTGTATATAAACCGCGCGGATGTTTAGGAGATGTTCAGAGACACCGGGAA
GGGAACGGACGTACACACGAATTACGTTCGTTGGAAGATGGATTTGGATACTGCGACAAGAAGTAAATGTTTGCTCCGACTAACT
TTAGGAAAGCATGGGATTTCGGTGGTTTCCCCCATTGCCAAAAACATGCGTGGGATTCTGCAACACTCTGTGCTGGCTTTTATGA
ATGACTGCATAGTAATTTCCTGTTCCGCTCCGTTTGGGATGGCTTTCTTAAAAATCAAGTGGCAATTGTTCGACTCATTTGAATA
TTTATCAGACGATGAGAGTATACTACCTATCATGTTCAATAATACTATACACCAGCGCCCTGGAGACCTCTTGGATTTTCTGTGG
GACTCTAAGAAAAAAACTGTCGGAATACCAACTTTAGCCACATTTACTATATCAGAAAGTTTCAAATCCGATGAAACTAGGGCTG
TACTTAGACTGAAAGTCTGTCGCGAGATAGAAGAGAGCGCGTCGTGCCCCCAAAATACGGAAATAAGCACGAAATTCAAGTACTC
TTCATCGGATGGACCTCCTATATACCTGCCTCGTTCCGATAAGGCATGTCGGGCCTCGCTGGAACAGTATTCGTTCAATGAAATT
ACAAAGTGGCTTTCGAAGATCCCTAAAGATACTCAATTACAGTCCCACTCCCAAGAGAGTATGTCACGTTCTCGTCCCGCAGAAG
ACGAGCAACGCCTAACCTTTAACGCGAAATGGTTCGGCGACCAGACAAGTATTGTATCGTCTGCGTCAATTTTGTCTCAGCTATG
TGGCCATGAACCTGCGCCTAAAAAAAGAGAGCTCTCAAGCCGAGCCATCGGTAAGAAACTAAAAGAAGCCCGCATTATCGCTGTT
CATGGGGACAAAGATGCCTGCCCCATTGTAGTATCACTCTCCAGACCCGGATCTCTGAAACAGTCACTAGGGTGGCTGAAAAGCG
GTCCTTGGGGACCACCGTGTCTTACATTTTACAAAGACTCTGTGAATAGTCTTGGAGTAGAACTCTCTGAAAGAGGAGGAGAAGA
ATTAGCTGCTGGGATTTTTTTCCTTTCCGCGTTTTCTGCAGATGGAGCAATCTTTGAGCAATGCCATGATGACTCAGACACAGCC
ATGCACGAATTTTTGGCGGAGGAGGAGCGCCTCATACAACAGACCACGCTATCCCATTCCAACTCAAGTAAGAAGAGGTCGCTCG
AAAATTACGAGGACACCGATATTAGCCCATCCCACCACCCACAAAAGCGGGGGAAATTAAAGAACGGCTCACTTGCACGGAAGAA
CTAAAATCGTCTCCTCGCGTCTACCGGGGAGCTCAGTTTACTGCCAGCGTTTGTGACCCAAGTGTGCCCCGTGATTTCAACTCTA
CCGCAACTATGACGGGGGCTACCATAATTGATCCATTCGCACCGCCCAAGGGTAAATGGTGCCCGTTCAATTTGAACGGGATAGT
TTTTTCCTTGATGATGTTTATTATATTTTTAGCCTGGATACTGTGCATTGACTATGGACTCGCGTTAGCTTACATTACCTGGGCA
AAGCTTTCTACGAAAGAGGCAAGATTCGGATGGATGATCGGACTATTGGTGGCTACGATTACTGCCAGTTTTCTGGATATTCAAT
ACTCGGCTCACAAAACAGTCCGAATTTATTTTCTGGTGATGCTTTCTATGGCGAGCGCAGTAATAATTATATCTCTCATCCATTC
CAACAGCCCCAATGCCGCGATAGTTATGGGGCTATTCTCCGTTCTTTCGGAAGTTTGCTTGATACTGATTTTGGGATTTCAACTC
CGACCGGCCATTTTCTGCAGCATAAACATGACCTGGCTCTTTCTTGAAGCCATGCTCCTAAATTTGACCGTACTTTCTTGGAACT
TGATGCACCTTCGAGTAAACCCTAGATACTTGGAACCGTTGGCCCTTTTTACTATTAACATTGTTGGCATACAATCCCTCTCGTT
TTATCTGCTCAAGAGTGATTTTTTTAAGACCAGCATGATAACTCTGACGGGCAGTATAGAACCATTTTCCGAAGATAACACGATT
TATACACCCCCAAGACAACATAAAGATACTCGCCCTTCACTGAATGACGGACCAACTCGGTGGTGCGGTTACTGTATTCTCGTAT

CTACAACATTGGTTACTGCCGCTTTCGCCTGCACATTATCATTACCGTTCCTGGGCAAAGATTTAGGTACTGTACGCATCGGCAT
GCAAACGAATTTTAAAATCCTCATGGTAGCGTGCGGTTCGGTTTTGGCATTTGGATCTACTTGCATTGGAAAACTATGCAAAATC
CATATCGTTGTATGGTTCGTGATAAGCATACTATTAACCTTCGTGTCTCTGCTATCACTGATTAAGTTGATTGAGGACCCAGCTG
GCATTCCATTTGGTGTCATTCTTGCATCGGTTTCTTGTCTGTTTCAAGTTGGAGCCCTCTTTTTCCGAGAATTAAAAACGGCCAC
CCATACACAAGGATGGATTTCATGCGCCCTTCTTTTCTGCTCCCTTTTCATTCCAATTGCCGCGCCGCTTGTGTGTGAGTACAAG
CTCTGAATTCTTGTCTAAGGGAGACGTGCCAATTCTGACAACGCCCTAAGCCAACACAAATGCCTTCCTCAATTTACGCGCTAGC
TTGAACATTCCAACAAGATGAATGCATCGCTAACATGGCTTGCTTTAACTTTTAAAGTTACCTTGAGTTTCAGCCTGCTCTGAAT
GTTTTCCTCCAAACCTAAGATGTTCTTAGTTGTACGATTTTTGTATTGCGAATACCACATCATCCAGTACCAAAGTACTAATGGT
CGGCAATCGAATAGAAATCAAACATGGCACGCAGGAGACTGTCGCCTCGAAACTGCCTTCGCCGAGCTGCAGAAATTCGGAACGA
GGACATTAGAAACAGATCTACATTCCACCGAGCAACCGTTTACTGGAGAAAATGCTAGAGAATAACCTAACAGGCGGCCTTCAAC
TTTCACTACTTAACAACTGGAACTATATTGTCGAGGATTGTCATAGTCACGAACATGATGCAGTATTGCTTCGGATGCAATCTAT
ATTTTGACATTGTTAATAAAGCACATGTACTATGCAAAGTTTGCATTCGTGTAATTCGTCGAGAGAGAAAGTT

>UL43

AATACCAACTTTAGCCACATTTACTATATCAGAAAGTTTCAAATCCGATGAAACTAGGGCTGTACTTAGACTGAAAGTCTGTCGC
GAGATAGAAGAGAGCGCGTCGTGCCCCCAAAATACGGAAATAAGCACGAAATTCAAGTACTCTTCATCGGATGGACCTCCTATAT
ACCTGCCTCGTTCCGATAAGGCATGTCGGGCCTCGCTGGAACAGTATTCGTTCAATGAAATTACAAAGTGGCTTTCGAAGATCCC
TAAAGATACTCAATTACAGTCCCACTCCCAAGAGAGTATGTCACGTTCTCGTCCCGCAGAAGACGAGCAACGCCTAACCTTTAAC
GCGAAATGGTTCGGCGACCAGACAAGTATTGTATCGTCTGCGTCAATTTTGTCTCAGCTATGTGGCCATGAACCTGCGCCTAAAA
AAAGAGAGCTCTCAAGCCGAGCCATCGGTAAGAAACTAAAAGAAGCCCGCATTATCGCTGTTCATGGGGACAAAGATGCCTGCCC
CATTGTAGTATCACTCTCCAGACCCGGATCTCTGAAACAGTCACTAGGGTGGCTGAAAAGCGGTCCTTGGGGACCACCGTGTCTT
ACATTTTACAAAGACTCTGTGAATAGTCTTGGAGTAGAACTCTCTGAAAGAGGAGGAGAAGAATTAGCTGCTGGGATTTTTTTCC
TTTCCGCGTTTTCTGCAGATGGAGCAATCTTTGAGCAATGCCATGATGACTCAGACACAGCCATGCACGAATTTTTGGCGGAGGA
GGAGCGCCTCATACAACAGACCACGCTATCCCATTCCAACTCAAGTAAGAAGAGGTCGCTCGAAAATTACGAGGACACCGATATT
AGCCCATCCCACCACCCACAAAAGCGGGGGAAATTAAAGAACGGCTCACTTGCACGGAAGAACTAAAATCGTCTCCTCGCGTCTA
CCGGGGAGCTCAGTTTACTGCCAGCGTTTGTGACCCAAGTGTGCCCCGTGATTTCAACTCTACCGCAACTATGACGGGGGCTACC
ATAATTGATCCATTCGCACCGCCCAAGGGTAAATGGTGCCCGTTCAATTTGAACGGGATAGTTTTTTCCTTGATGATGTTTATTA
TATTTTTAGCCTGGATACTGTGCATTGACTATGGACTCGCGTTAGCTTACATTACCTGGGCAAAGCTTTCTACGAAAGAGGCAAG
ATTCGGATGGATGATCGGACTATTGGTGGCTACGATTACTGCCAGTTTTCTGGATATTCAATACTCGGCTCACAAAACAGTCCGA
ATTTATTTTCTGGTGATGCTTTCTATGGCGAGCGCAGTAATAATTATATCTCTCATCCATTCCAACAGCCCCAATGCCGCGATAG
TTATGGGGCTATTCTCCGTTCTTTCGGAAGTTTGCTTGATACTGATTTTGGGATTTCAACTCCGACCGGCCATTTTCTGCAGCAT
AAACATGACCTGGCTCTTTCTTGAAGCCATGCTCCTAAATTTGACCGTACTTTCTTGGAACTTGATGCACCTTCGAGTAAACCCT
AGATACTTGGAACCGTTGGCCCTTTTTACTATTAACATTGTTGGCATACAATCCCTCTCGTTTTATCTGCTCAAGAGTGATTTTT
TTAAGACCAGCATGATAACTCTGACGGGCAGTATAGAACCATTTTCCGAAGATAACACGATTTATACACCCCCAAGACAACATAA
AGATACTCGCCCTTCACTGAATGACGGACCAACTCGGTGGTGCGGTTACTGTATTCTCGTATCTACAACATTGGTTACTGCCGCT
TTCGCCTGCACATTATCATTACCGTTCCTGGGCAAAGATTTAGGTACTGTACGCATCGGCATGCAAACGAATTTTAAAATCCTCA
TGGTAGCGTGCGGTTCGGTTTTGGCATTTGGATCTACTTGCATTGGAAAACTATGCAAAATCCATATCGTTGTATGGTTCGTGAT
AAGCATACTATTAACCTTCGTGTCTCTGCTATCACTGATTAAGTTGATTGAGGACCCAGCTGGCATTCCATTTGGTGTCATTCTT
GCATCGGTTTCTTGTCTGTTTCAAGTTGGAGCCCTCTTTTTCCGAGAATTAAAAACGGCCACCCATACACAAGGATGGATTTCAT
GCGCCCTTCTTTTCTGCTCCCTTTTCATTCCAATTGCCGCGCCGCTTGTGTGTGAGTACAAGCTCTGAATTCTTGTCTAAGGGAG
ACGTGCCAATTCTGACAACGCCCTAAGCCAACACAAATGCCTTCCTCAATTTACGCGCTAGCTTGAACATTCCAACAAGATGAAT
GCATCGCTAACATGGCTTGCTTTAACTTTTAAAGTTACCTTGAGTTTCAGCCTGCTCTGAATGTTTTCCTCCAAACCTAAGATGT
TCTTAGTTGTACGATTTTTGTATTGCGAATACCACATCATCCAGTACCAAAGTACTAATGGTCGGCAATCGAATAGAAATCAAAC
ATGGCACGCAGGAGACTGTCGCCTCGAAACTGCCTTCGCCGAGCTGCAGAAATTCGGAACGAGGACATTAGAAACAGATCTACAT
TCCACCGAGCAACCGTTTACTGGAGAAAATGCTAGAGAATAACCTAACAGGCGGCCTTCAACTTTCACTACTTAACAACTGGAAC
TATATTGTCGAGGATTGTCATAGTCACGAACATGATGCAGTATTGCTTCGGATGCAATCTATATTTTGACATTGTTAATAAAGCA
CATGTACTATGCAAAGTTTGCATTCGTGTAATTCGTCGAGAGAGAAAGT

>UL44

GATCTACTTGCATTGGAAAACTATGCAAAATCCATATCGTTGTATGGTTCGTGATAAGCATACTATTAACCTTCGTGTCTCTGCT
ATCACTGATTAAGTTGATTGAGGACCCAGCTGGCATTCCATTTGGTGTCATTCTTGCATCGGTTTCTTGTCTGTTTCAAGTTGGA
GCCCTCTTTTTCCGAGAATTAAAAACGGCCACCCATACACAAGGATGGATTTCATGCGCCCTTCTTTTCTGCTCCCTTTTCATTC
CAATTGCCGCGCCGCTTGTGTGTGAGTACAAGCTCTGAATTCTTGTCTAAGGGAGACGTGCCAATTCTGACAACGCCCTAAGCCA
ACACAAATGCCTTCCTCAATTTACGCGCTAGCTTGAACATTCCAACAAGATGAATGCATCGCTAACATGGCTTGCTTTAACTTTT
AAAGTTACCTTGAGTTTCAGCCTGCTCTGAATGTTTTCCTCCAAACCTAAGATGTTCTTAGTTGTACGATTTTTGTATTGCGAAT
ACCACATCATCCAGTACCAAAGTACTAATGGTCGGCAATCGAATAGAAATCAAACATGGCACGCAGGAGACTGTCGCCTCGAAAC
TGCCTTCGCCGAGCTGCAGAAATTCGGAACGAGGACATTAGAAACAGATCTACATTCCACCGAGCAACCGTTTACTGGAGAAAAT
GCTAGAGAATAACCTAACAGGCGGCCTTCAACTTTCACTACTTAACAACTGGAACTATATTGTCGAGGATTGTCATAGTCACGAA
CATGATGCAGTATTGCTTCGGATGCAATCTATATTTTGACATTGTTAATAAAGCACATGTACTATGCAAAGTTTGCATTCGTGTA
ATTCGTCGAGAGAGAAAGTTACAAGTTCGATTCTCTCGCGCTAGGAGTGTTTCCACGTGCGAAAACGCAAAAATTTTCATATTAT
TCGGGCGGACTGTGTCCATAGTAGCTAAATTACCGCGATCTGGAGACTAGGGCATTCACGACTCAACATGCAGCATCAGAGTACT
GCGCTAGTTTCGAGTATACTTTTGCTCTTGAGCCTGCAAAGCCTTGCGTTTGAATTTTTCTGTGATCCGCCACACGTTTTTCGAG
GGCAGCTCGGTGACCCCATTCTATTGCAATGCTTCAGCGACAGACCTCTAACCCACGAAGAATCTGTAAAAGTAGAAGTAATTCG
ACACCCAGCCAGCTTAGTTGAAACTGCGCTAAGCGCCTACGGGATCCCCCCTTCGCTAGATCCATGGAGAGCTACTCCAAGAACT
CTCTACACATATGATGCCGCTACTGATTCAATCAAGGACCTAGGATACATTGGTGAAGATGGAATTAACCCACCATATTTGGACG
ACTGTCGTTCAGGTTTTTTCAATGTCTCTATCAAGTCTAGCATGAGATCTCACATGGCGCGTTATCAGTGGACCGCAAGTCGAGG
GTCTACAAAACTAAATAGCTCTTTTATCGACGTCTTTTTGGCAAGACCACCTACAACTGTCCGCATCAAATCAGAAGAACTGTAC
GAAGACTCAGATAAGGCTTCGCACTTAAGTGTTGAAGCGCTTGGCGCTTATCCTCCATCTGCTGCGCTGGGTACATGGATGATAC
ATAATGCATCTCTTGCTGAAAAATACAGTTTAGAAAGAAGAGTTCTTTATGCATCAGGAGAGAATGGATCGGTGGATCAGACATG
GGAACTGGAAATACGTGGAGAAGCCAGCCAGCCCCTCCCTTCCAAAATTCAATTTGTATATCGATGGACCCCTCCTGAGGACTTT
GAAATGCTACGACCTGAAACTCGCTTGTTAAGGTTGACTCCCAGCTGGATTAGCAAGCCCCGCATCACGGTACAATTCGTCCCTC
CTGCCTATGCCCTGTGTAGAGCAGCTAATATTATAGACGGCCGAGGATTTATTGAATGGATCGTAGATAATAGAATTTCGACGAG
CCCACACCAGACCTTTGTTTTGGATGAGCCCGAGGGGAAAAATATCGTTACACTAATGGACGTCATAAAACTACCACCGGAGGAT
ACATTTCAATCTGCCTCTAATTACGTGTGCGTCATAAGAGGCTATGAACATGCATACAGATATCTCAACGCCTCCTTAATGATAG

ATAATCTGCCAATGCGGCAAGGATTCCCCGCAGTCGCTGCGATTTTTATTATAATTAGTATCGCTTTTGTGGGTGGGTTACTAGT
TGCTTGCTTGGGCGCATGGTGCTGGAAGACAACATAAACGCTCATTTAATAAATGACATTACAAACGTGCACTACTGTCTGTCCA
ATTTATTTCATCGGAAA

>UL21

GCCCGTTATAATACTGCAGTCTCTGCCTTTATCGTGTCTTTACCGGGGGTGAGGCTTCAAGGCGGGGTTTATGAGGCCAGTAACG
CGATCCTTCCACCTAACTTTGCGATTTCGTGTAGCAGTAAAACGATTTAGAAGCTTATTGCTGTCCTTAATATTGTCCAGGCCCA
CAAAAGAGGAAGTGTCACGTCAATAACGTCGTATAGAAGAATTATAAAGTACAGACAGGCAGCCCACATCGAAAAATACGGAGAG
AGATAACAATGACCAATTCTCAAGGCTTCCTGTGTGTCTGTCCAGTTTTTCGCAAATATTGAAATAGCAAGGTCATCGTCCGCAA
AAAGGTGCCTTGAAAAGATAATCGCCCCATATATAGCCGTTATCATACATAACACCATTTGAATGAGTTTTTCAGGACCTTTTAA
AGGAAGACGATCAGCCTTCACGTTTCGGTACACAAAAAAGGAAATCAATAGTCCTTCTAGACAGGAGAGGGCAGTGATAATAAGA
GCGTTTACAAATAAATATTTCTCTCGCGTGCAATTCCCGTACCAAAAGAAAGGCAGACAACAGAGTATTTTTAAAAGAAGGGCTA
AACTAGGATAAAAAGTTTGTTTCCTCATTCGCGTATATTGTAGATCAGATGTATCTTCACTTCCCGAAGCTATAAGGCGATCGTT
TAGAAATATGGATCCTTCGTTCAGTGTCGAATAGTCACCATCACGTTCTTTTTCGTAGTCAAATCCAGCCAACAGCTTGTCAATC
TCACAATTTCCCGGTACTTGTACTGAAGAGAGGCGGATTGAATTGGCTTCGTTTCCAGTTGTCATGTTGAAGGAGAGCTCAGGTT
TGGTTGTAGGTGGTGCCAAGACGCAGGCGCTGTATGATCGTCAAAACTTTATGCTACGATAGAATTGGAGCAGTATAAAGGGCGG
ATACTTTTAAAGCACCAATAGCGGTACCGTCACTTGACGCGCAAGCTCTTTGGAGAAGGACTGAAATGGAAGTGGCTTTTCAAGA
AGTCTCCAAGCGCGGATCAAAGTCTCCTGCACGCTACATAGATCACCACAGCGGACGACTCTACATTATCTATGGAGGATGCATC
TACTCGATCTCTACTCGCGGGCCGATAAAAAATATTACCAAATTCGGTTTTAGGATCAAACACCCTACCATGTGTGCGATTATTA
ACACTGAACTGAATCGTTTGAATGGGATGCCTTTTGACCGCCAGGAGGAAGAATACATTTTTACTTCTCATCTTGGTGAAATTTC
AGAATCTCCACATGATGAATTGATTGATAATTTTACCGTGGAGATCAGACGACCGATCTTGTGCAAAATATTTGTCGGCCTGCAG
GTTTCCGAAACGCTAGAAATACGAGTTGGGGAAGTCTCTCTGAGGAATAAAGTACTGAGGTTATTTAGAACCCCGGATCTAATCA
ATACGACGACAGGATTTAGGTATCAAGTAGATAGTAGTGCGCTGGTGCTAACGCAGGCACATATTACTACTCTGCCGAAGTTCAT
AGAAGAAGATTTAGAAGATTTATTCTTGAACTCCCTAACAAGAGAACCGATTCATGAACAATCACAGATTTTTACTCCTATTATT
ATTACGGGAGGTAAAATCATTAGTACGGTCGCAGTAGGAAATGAGAAAAAAGCCCTGCGTGGGTGTAAGAAAAAAACTATCCTCT
CAGATCATGTGCAGGTGAAACACATTTACCCCGACCAAATGTTCGACGAAGATGTAGAAAATTTCTGGGAATGTTGGAAAGTATT
AAAGCGACTCTGGAAGATGTTAATGATCGTGGATGCGGCCACGATTTCACGACATGGACTTCTCGCAACTATCGGACTTTCATGG
GAGACTGATGGAGCAAAACTGAAAAGTACATTGCGCGACGAGGCTGACATTCTGGTCACGCTGCCATACCCTGGCCAACGATGCC
TCAGCAAAGCATCATTCGGGCGAAGCTCAGTGATTCCTTATACTCTGTTTAATTATTTACTTATGCTCGGGCACTATCCTTGTCA
CTTTGACGAAATTACAGTTGTCTGTGAAAGATATATTGAAAAAAGCGGGCTCAAAAAAATCCAGAGCCCCGAAACTCTACAATTA
CAAGAATTGTCGGAAGTTGTTCTCAATATATTGAAAACTTTGCTGGCCACGCACTATGTTTTGCAAGATGCACTAGTCAGCTGTT
GTTTGAATCCCAAAGTGCAACCCAAACCGAGATGTTCGTATCCACCAGAGAGTATCCCGTTAATCCCAGAAACAGCGGTACTCCT
CGATGCGGCAAAAACCCGTTTAATGAAAGATGTAAGTACCGCAGCGTGCCTTCGAACTAAAATTCCACAAGAAATTTTGGTCAAC
ACTGCAAAGCGAACCGCCGCGGCTCTCGTCTGCATTTACTTGTACGAGGAGGCATACGCGGCCAGTGAGGTTTCTGAAATTACTG
GAAATTTTGAGCTTGGGCGCAAAATCGAAGATCTGTGGAAGATGACTAAGGAAGAAATTTCTCCAATCATTTGGGAATATCGGAC
GGATCTGGTAGAAATATTGAAATCTATAATATCGTTTAGGCTTGCTTGAGACTACCTGCCAGTGCAAGGTTTCCGATGAAATAAA
TTGGACAGACAGTAGTGCACGTTTGTAATGTCATTTATTAAATGAGCGTTTATGTT

>UL20

GGGTAAATGTGTTTCACCTGCACATGATCTGAGAGGATAGTTTTTTTCTTACACCCACGCAGGGCTTTTTTCTCATTTCCTACTG
CGACCGTACTAATGATTTTACCTCCCGTAATAATAATAGGAGTAAAAATCTGTGATTGTTCATGAATCGGTTCTCTTGTTAGGGA
TTCAAGAATAAATCTTCTAAATCTTCTTCTATGAACTTCGGCAGAGTAGTAATATGTGCCTGCGTTAGCACCAGCGCACTACTAT
CTACTTGATACCTAAATCCTGTCGTCGTATTGATTAGATCCGGGGTTCTAAATAACCTCAGTACTTTATTCCTCAGAGAGACTTC
CCCAACTCGTATTTCTAGCGTTTCGGAAACCTGCAGGCCGACAAATATTTTGCACAAGATCGGTCGTCTGATCTCCACGGTAAAA
TTATCAATCAATTCATCATGTGGAGATTCTGAAATTTCACCAAGATGAGAAGTAAAAATGTATTCTTCCTCCTGGCGGTCAAAAG
GCATCCCATTCAAACGATTCAGTTCAGTGTTAATAATCGCACACATGGTAGGGTGTTTGATCCTAAAACCGAATTTGGTAATATT
TTTTATCGGCCCGCGAGTAGAGATCGAGTAGATGCATCCTCCATAGATAATGTAGAGTCGTCCGCTGTGGTGATCTATGTAGCGT
GCAGGAGACTTTGATCCGCGCTTGGAGACTTCTTGAAAAGCCACTTCCATTTCAGTCCTTCTCCAAAGAGCTTGCGCGTCAAGTG
ACGGTACCGCTATTGGTGCTTTAAAAGTATCCGCCCTTTATACTGCTCCAATTCTATCGTAGCATAAAGTTTTGACGATCATACA
GCGCCTGCGTCTTGGCACCACCTACAACCAAACCTGAGCTCTCCTTCAACATGACAACTGGAAACGAAGCCAATTCAATCCGCCT
CTCTTCAGTACAAGTACCGGGAAATTGTGAGATTGACAAGCTGTTGGCTGGATTTGACTACGAAAAAGAACGTGATGGTGACTAT
TCGACACTGAACGAAGGATCCATATTTCTAAACGATCGCCTTATAGCTTCGGGAAGTGAAGATACATCTGATCTACAATATACGC
GAATGAGGAAACAAACTTTTTATCCTAGTTTAGCCCTTCTTTTAAAAATACTCTGTTGTCTGCCTTTCTTTTGGTACGGGAATTG
CACGCGAGAGAAATATTTATTTGTAAACGCTCTTATTATCACTGCCCTCTCCTGTCTAGAAGGACTATTGATTTCCTTTTTTGTG
TACCGAAACGTGAAGGCTGATCGTCTTCCTTTAAAAGGTCCTGAAAAACTCATTCAAATGGTGTTATGTATGATAACGGCTATAT
ATGGGGCGATTATCTTTTCAAGGCACCTTTTTGCGGACGATGACCTTGCTATTTCAATATTTGCGAAAAACTGGACAGACACACA
GGAAGCCTTGAGAATTGGTCATTGTTATCTCTCTCCGTATTTTTCGATGTGGGCTGCCTGTCTGTACTTTATAATTCTTCTATAC
GACGTTATTGACGTGACACTTCCTCTTTTGTGGGCCTGGACAATATTAAGGACAGCAATAAGCTTCTAAATCGTTTTACTGCTAC
ACGAAATCGCAAAGTTAGGTGGAAGGATCGCGTTACTGGCCTCATAAACCCCGCCTTGAAGCCTCACCCCCGGTAAAGACACGAT
AAAGGCAGAGACTGCAGTATTATAACGGGCTGACTTCGTAGGATAAAGGCAACCTACAACTCTGTGTGCTCAATATTTAGCCATG
AATCCTGACAACGGGATCCCGCATAACAGTCATCATGATCGCGCAGCATTCCCAAGATCTGCTGCCCCTTTCGTAGCATCTGGGG
AACTGTTAGGAATTCTTCAGAAAATTGCCATGCGCATCTATATGAATGGATAAGCCGCGAAGGGGATTGTTGCTACAGACACAGC
TTTGATATTCTGCTGGGATCTTATTTCAATACACTAACGCTCACCAACTTTCTAGAAACCGGACTTTCAGTTGCATGTATTTGCG
TGAAATTTCCAGAGTTACGCTATGCAGACCGAGGAATAATCCAGTTTGTAGTGGCTAATCCCATGATTGCAAGAAGTGATTGTGA
AGTACCTTCTCGGCCATCATTTACCTACATCAGTAAGAGATGGTCTAGGACGACATTATCCTCATCCCTTGTGATTTGTGCACCA
GCTCTGGGCTTGCTAAGTGGCGAGTCACTTGACGGGACCGAAATATCTGAGTTTTCTAGATTACAGGCGTTAAACCAACTTGCAC
GAAACCTCAAACTAACTCTAGACTCATTTGAAAGAGGAACAATAAATCATGTGCTGAGAATTCTAATCCGAAAAGCTCCACCACT
TCCTCTTCTACGACCCATGATGGCGGCGTTGGAATGTGAACGAGAAATGTCTACGGTTGCCAGAGCAAATATTATCTCCAGTATG
AAGGCAGCCCTATGTGAAGATTTGTTTTTCATAGATAAGGAGAGGGGACACGAGACTCCGGATTTCGCCCGAAAGCTATTAGCCT
TAATTAATTGCACGCTTCCTAGTGTAACCGATGCTCGTGTTACACACATAGGTCCGGATGGTCGACTTATAGAAGGGGTGATTGT
TACAACCGAAGCCGTAAAAGGCCTCATCGCGGCCCGCCTTGGCATCGAAACTGCCCGTGCCAACGTCCCGGCAATGTACAGTGAG
ATGGTATTATCTGGAAATAGCCTTGTAACGGCCTTACTGCTGGGGAAAACCATTCGTAACTTTGATGAAGCCGCCGCAAATTTGT
TAAGCTTTTTGGACGGTGAAAAAAATCTCGCCGATTTTCCAGAAATACCGTCTAATAACGCAGACCAAGTTCCAACAATGTCCGT
CAAAATGTCCCTGCTCAACGTGGGGGACCATCTTGTCAGCATAGAAGCACTAGAGCGTGTTTACACGCGCACTGGAGTTCCCTAC

CCACTCTGTGAAAATGTGGATCTGACATTCTTTTTCCCACTTGGGTTGTTCAAACCTGCAATAGATCGTTATTCCACATCTGAGA
TTGCTCCTGCGGTCGGAGCCCCGGATTATCGCCAGTTTCCCCCTACAGAAATGTACTTTTTCAATAAGGACGGCATGATGGTTAA
ACTCACGTTTGAAGACTCATTGCCAACCGTCGCACATCCCATAGCACATGGAATGCTGGAAGCTTTGGCAGAACTTTGTCAAGAA
CCATGGGTTTCAAATCGCAGACCTGCTCCGATGGGGTTCACAATTCAACGTATCGGACTAAACCCACCTCGCCTGTTAATGATCG
AATTCTTGGAAGCGGTAGCACGTACTGCCCCTGCACCATACCCAGACGCAACTTTAATCAATCGGAAAAATCCAGATCAATTCTC
TAGCCATACCAACCCGTTCCTTCCATTAGAGGTTCACCCGTTCTATGATGTGTACAGAGTTGCGCAGGATCTAACTATTCCCTGT
GATGAACCATTGTTTTTCCCAGCCGAACCAATAACACTGGCCGCCTCTAGGCGACTCTGCAACGGGGATATTCCGTTGCCGCTGT
CTTCAGTCGATTTTAGACTTGCCAGGGGGTACTGTATCGCTGCAGGAAGACACAGACTGCACCCTTCAACGGACGCAGCCATAGA
AACAACACTATCTGATGTGAATTACCCACTGGCTTTCTATGTAATTGAAGCTTGCATCCATGGCGATGAATTGATTTTCATGGAA
TCGCAGAGGCTAGTAGCGCAGTGCATCAATAGCTATTGGCATACATCAAGGGGACTAGCTTTTATAAACAGTTATCCAATGGTAA
CGTACATTTATCACAACATGACTGGAATGATCGATCGAGATTGCCACTCCAGATATGCGGATGTAATGGGCCTGCTTCACGCACT
GAGGGAAACAATCAGAAATTATACTCTCCCTGCGGAACCAATTCTCCTGCGTTCTCATGAGGAATTAAACAATCTCATGACGGAC
CCGGCACTTTTTCCGCCCATGATTTATGACTGTGATTCCATCCTAAGGGCACGCACTGCATGCGCGACTAGAGGGGTTACGATCT
CCACCTTCGGAGAGAGAGCACCTACCGTTTCTGTCAGGGAATATCCAGCCCAAGTGGATTTCACCGTTCTTAGTAACACTCTAAA
TCACGGCCCAGCATATACTGCTGGCGGAAGACATGAAGGAGGAACCCATCATGACTCCGAATGGACAGTACTCAGCAAATTGTTT
TACTATGTTTTTCTACCTGCTGTCTCTAGAGGGAGGTGTTGCAGTGCAGGAGTCGAGTTTGAAATGATCTACGATCTCATTAACA
CTACCAGACTTCCAGATACAGTTGATGAAATGGACAATCAGGCAGGTGCCGTCGACAGAGGTCCTTTGGCGGATGAAAATCTTGC
CCCCGAATCCTTTAACACTTTGTTGGCAAATGGATCCATAAATCTGGTGGATAATGAAGCTCTGGTAGCATTTATCGCTGCCGCA
CGGCGCCGTCAGGCAGTACACACGATACCTCTGAGAGTTAACTATTTACCTGACCCGGGCTTCGAAACTATAGACAGCCCGAGGA
ATTTTTTGGTAGACGGAGTCTTATACAATGGAATAATCATGATGAACTATGCTCAATATGACGCGACAGCTATTCCATCTCGCTA
TTTCTATGCATTGCCTGTGAATGGATTCTTTATGAACCGTACTATCATAGAAGCATCTCACAGGGCTAACGTCAACCTTACTAAT
GTACCTGAGGACTTGCCACTAGTTCCTACATTTTTGGGGAGCGAAGTGTATCGCTCCATAAGAGCACCGTCGTACCTATTGCTTC
GAGAGCTTCAAACTATGCCTCAAACTCGGTCGCGGCATACAGCTTGCTGGCAGGATACTTCAAGACGTCTCCCGTCGCACTTGTC
CATCAGTTGAAATTAAATCTTCATCCAGGGTTTGCTCTCACTGTCGCCCGCCAAGACAGATTTGCGGCGGATCAAATTTTATTTG
CCCGAAGAGCATCGGAATCTTACTATCTAGGATCTCCAGTTGTAACCAACCGCCCAGAGAATGATTCTCTCGTAATTGAGATAGT
CAGCCAAGGGGGCACATAGATATGGGTCTGGGATTTACAGCGAGTCGTGTTCCTGCAAAAATTAATACAGTGGTTACAGACATGG
GCAATCATTGCCAAAATCTCTTCAATGCCCGCTACCCTGGTCAGTTTCGCCATGCGGAAGTTGCGGATTTTATTGCCTCTGAAAT
AACAGACAATGACTCTACAGCCTTGCCTAGAGCGCACCTCCCATTCTGCTGTCTTACGAAAAGGCACCCATCCCCCCTTGCATAG
AGCGGGGACAGCTTGCTACATGTGAGTTTCTTTTAACTCCAGTGACTGCCGATCTTGCATATTTTTATACATCAGCAAATCCCCG
AGGGCGCAGCAGCTGTATCGCGTGCACCAACTGCGAGGATCCATGCGCGAGTGAAACTGAAAAGGCTATGTATGACCATTCTACC
CCAGATGCAGCACACCCGTCTAGGGCTACGAATAATGCTTGGGCATCGCAAAAGTATTCAGTGGGAGACAAAATGTATAATGCGC
GGCGGGGTTTTATTACGGCCAGTGATTTTTACAGTCCTCTGAGCAAGTTTATGACGCCATCTAGAGCTGAAGATAAGAGTCGGTG
CCTGGCTCGTTTGATGAGAGATTCCTCTGCGGCAATTAGTTCAGTTACGGGTGACACTGAGTATCAATTTGTCGCCCCGCCCGGT
GCCAACGAATTAGTAACTGATCCGTGCGCAATGTTTCAAGAAGCGTTCCCACCCCTTTGTAGCAGTGACAAGGTCCTCTTCGCAA
CCTACGAAGGCCCTAATAGAGCATCTGGATCTGGTGCAAGAGAAAACCACTTTGCGCAGTATTTGATTCACGATAAGTCTCCAAT
CGCCAATGCGTTAAAAACCCCATGCAACACGCGTCGCTAAGACGGAATTCGCAAACGATTTCTACCCGTGAAATCACGTCACCGC
CTGGTCTGTTTGGGCGGAGCCCTTTCAAAACTTCTTCATAAGTGTTTCCTGTGCAGATGATTTGCCCACGGAGGGATATATAGCG
CATACCTAACTAAATTTCAGAGGACAGAAGTACAGATCGTATCAGCCAGAACGATGACAAGCGGACAGTACGAGATCCGCGTTGT
TCTGCCAAATGGACTGACACGAGATGAAGAAGATCGTCTCAGAAGCTTGCGTGGGACTATTTTAATGGCGCCAATTTTGCGAAGG
TGTGTTTTTCTGCACGAAATCGATCAAAAATCATTTTTCGCCCATGGAAAAGAACCAGACTATGCAACTCTGCTCACAGCCTACA
GGAGACGCTTTCCCATATTAATAGTTTGTGTCGAAAACCGAGAATTGAGCGCCATCGCGCTATCCATAGGATACCCGCGCGGAAT
CAGCGTGCGCAACACTGGTCCATTTAGTTTGAACAACGGTGATCTCGTGTCTTTACTGCCACCCATTACGAATACGCGTTTTCGC
GTCGACTTGCCATCCTGCGGCTCTGTAATCGAGCCAGCCATGACAATTCCTTTCGAGATCGGAACAGAACTGATGGGAAAGATTT
TTGCAGGGATGGCCTATGACTTCTGCGTTAGAAATCAAATAGCGACAACCCGTCCTCGTGATATTTATGTTGTAACATATAAGAA
CAAGACACTCGATCTATCTACATTGCCCCCGTCTGATGCGGCTGCACTTCAAGATACAATGAAAAGCCTATTTTCATCGGTTCTA
TTTTCGATACACGAGGGGGTCATGTCCGTTTTGTCACTAATGCCTGCTTTATTGGCAGGAGGGGCGAACGATCCATTTCTCAATG
CCATTCTACAAATGCAAAGCATGACCAGACTTTCCGTGCAGCTTTTTAATCCACCCGCGCTAGAATTACCGGAACCTGCTGGCAG
CTCTGGGCGATATCATGTATTTGATGCTTTTGCCGCGTGGCTTTCAATGTCTCACCGACTTGGTGATTTATTCAATTTGAAACCG
GTCCTCAAAGTAGTAATGTTTTACTCGGATGATTCTACCGCAGACGAGGGAGACCTCCTCAACGCCATTGTACCTTAACCCGTCT
GTGCATCTCTATGAAATGACGAGACATCATATCTCTTTGATTTGTGAATAAAGTATTTATTATCGAAATTGAATTTATGCTGGTC
TCTGTCTTCTACCGATCGGT

>UL19

GGTGCTTTAAAAGTATCCGCCCTTTATACTGCTCCAATTCTATCGTAGCATAAAGTTTTGACGATCATACAGCGCCTGCGTCTTG
GCACCACCTACAACCAAACCTGAGCTCTCCTTCAACATGACAACTGGAAACGAAGCCAATTCAATCCGCCTCTCTTCAGTACAAG
TACCGGGAAATTGTGAGATTGACAAGCTGTTGGCTGGATTTGACTACGAAAAAGAACGTGATGGTGACTATTCGACACTGAACGA
AGGATCCATATTTCTAAACGATCGCCTTATAGCTTCGGGAAGTGAAGATACATCTGATCTACAATATACGCGAATGAGGAAACAA
ACTTTTTATCCTAGTTTAGCCCTTCTTTTAAAAATACTCTGTTGTCTGCCTTTCTTTTGGTACGGGAATTGCACGCGAGAGAAAT
ATTTATTTGTAAACGCTCTTATTATCACTGCCCTCTCCTGTCTAGAAGGACTATTGATTTCCTTTTTTGTGTACCGAAACGTGAA
GGCTGATCGTCTTCCTTTAAAAGGTCCTGAAAAACTCATTCAAATGGTGTTATGTATGATAACGGCTATATATGGGGCGATTATC
TTTTCAAGGCACCTTTTTGCGGACGATGACCTTGCTATTTCAATATTTGCGAAAAACTGGACAGACACACAGGAAGCCTTGAGAA
TTGGTCATTGTTATCTCTCTCCGTATTTTTCGATGTGGGCTGCCTGTCTGTACTTTATAATTCTTCTATACGACGTTATTGACGT
GACACTTCCTCTTTTGTGGGCCTGGACAATATTAAGGACAGCAATAAGCTTCTAAATCGTTTTACTGCTACACGAAATCGCAAAG
TTAGGTGGAAGGATCGCGTTACTGGCCTCATAAACCCCGCCTTGAAGCCTCACCCCCGGTAAAGACACGATAAAGGCAGAGACTG
CAGTATTATAACGGGCTGACTTCGTAGGATAAAGGCAACCTACAACTCTGTGTGCTCAATATTTAGCCATGAATCCTGACAACGG
GATCCCGCATAACAGTCATCATGATCGCGCAGCATTCCCAAGATCTGCTGCCCCTTTCGTAGCATCTGGGGAACTGTTAGGAATT
CTTCGAGAAAATTGCCATGCGCATCTATATGAATGGATAAGCCGCGAAGGGGATTGTTGCTACAGACACAGCTTTGATATTCTGC
TGGGATCTTATTTCAATACACTAACGCTCACCAACTTTCTAGAAACCGGACTTTCAGTTGCATGTATTTGCGTGAAATTTCCAGA
GTTACGCTATGCAGACCGAGGAATAATCCAGTTTGTAGTGGCTAATCCCATGATTGCAAGAAGTGATTGTGAAGTACCTTCTCGG
CCATCATTTACCTACATCAGTAAGAGATGGTCTAGGACGACATTATCCTCATCCCTTGTGATTTGTGCACCAGCTCTGGGCTTGC
TAAGTGGCGAGTCACTTGACGGGACCGAAATATCTGAGTTTTCTAGATTACAGGCGTTAAACCAACTTGCACGAAACCTCAAACT
AACTCTAGACTCATTTGAAAGAGGAACAATAAATCATGTGCTGAGAATTCTAATCCGAAAAGCTCCACCACTTCCTCTTCTACGA
CCCATGATGGCGGCGTTGGAATGTGAACGAGAAATGTCTACGGTTGCCAGAGCAAATATTATCTCCAGTATGAAGGCAGCCCTAT
GTGAAGATTTGTTTTTCATAGATAAGGAGAGGGGACACGAGACTCCGGATTTCGCCCGAAAGCTATTAGCCTTAATTAATTGCAC

GCTTCCTAGTGTAACCGATGCTCGTGTTACACACATAGGTCCGGATGGTCGACTTATAGAAGGGGTGATTGTTACAACCGAAGCC
GTAAAAGGCCTCATCGCGGCCCGCCTTGGCATCGAAACTGCCCGTGCCAACGTCCCGGCAATGTACAGTGAGATGGTATTATCTG
GAAATAGCCTTGTAACGGCCTTACTGCTGGGGAAAACCATTCGTAACTTTGATGAAGCCGCCGCAAATTTGTTAAGCTTTTTGGA
CGGTGAAAAAAATCTCGCCGATTTTCCAGAAATACCGTCTAATAACGCAGACCAAGTTCCAACAATGTCCGTCAAAATGTCCCTG
CTCAACGTGGGGGACCATCTTGTCAGCATAGAAGCACTAGAGCGTGTTTACACGCGCACTGGAGTTCCCTACCCACTCTGTGAAA
ATGTGGATCTGACATTCTTTTTCCCACTTGGGTTGTTCAAACCTGCAATAGATCGTTATTCCACATCTGAGATTGCTCCTGCGGT
CGGAGCCCCGGATTATCGCCAGTTTCCCCCTACAGAAATGTACTTTTTCAATAAGGACGGCATGATGGTTAAACTCACGTTTGAA
GACTCATTGCCAACCGTCGCACATCCCATAGCACATGGAATGCTGGAAGCTTTGGCAGAACTTTGTCAAGAACCATGGGTTTCAA
ATCGCAGACCTGCTCCGATGGGGTTCACAATTCAACGTATCGGACTAAACCCACCTCGCCTGTTAATGATCGAATTCTTGGAAGC
GGTAGCACGTACTGCCCCTGCACCATACCCAGACGCAACTTTAATCAATCGGAAAAATCCAGATCAATTCTCTAGCCATACCAAC
CCGTTCCTTCCATTAGAGGTTCACCCGTTCTATGATGTGTACAGAGTTGCGCAGGATCTAACTATTCCCTGTGATGAACCATTGT
TTTTCCCAGCCGAACCAATAACACTGGCCGCCTCTAGGCGACTCTGCAACGGGGATATTCCGTTGCCGCTGTCTTCAGTCGATTT
TAGACTTGCCAGGGGGTACTGTATCGCTGCAGGAAGACACAGACTGCACCCTTCAACGGACGCAGCCATAGAAACAACACTATCT
GATGTGAATTACCCACTGGCTTTCTATGTAATTGAAGCTTGCATCCATGGCGATGAATTGATTTTCATGGAATCGCAGAGGCTAG
TAGCGCAGTGCATCAATAGCTATTGGCATACATCAAGGGGACTAGCTTTTATAAACAGTTATCCAATGGTAACGTACATTTATCA
CAACATGACTGGAATGATCGATCGAGATTGCCACTCCAGATATGCGGATGTAATGGGCCTGCTTCACGCACTGAGGGAAACAATC
AGAAATTATACTCTCCCTGCGGAACCAATTCTCCTGCGTTCTCATGAGGAATTAAACAATCTCATGACGGACCCGGCACTTTTTC
CGCCCATGATTTATGACTGTGATTCCATCCTAAGGGCACGCACTGCATGCGCGACTAGAGGGGTTACGATCTCCACCTTCGGAGA
GAGAGCACCTACCGTTTCTGTCAGGGAATATCCAGCCCAAGTGGATTTCACCGTTCTTAGTAACACTCTAAATCACGGCCCAGCA
TATACTGCTGGCGGAAGACATGAAGGAGGAACCCATCATGACTCCGAATGGACAGTACTCAGCAAATTGTTTTACTATGTTTTTC
TACCTGCTGTCTCTAGAGGGAGGTGTTGCAGTGCAGGAGTCGAGTTTGAAATGATCTACGATCTCATTAACACTACCAGACTTCC
AGATACAGTTGATGAAATGGACAATCAGGCAGGTGCCGTCGACAGAGGTCCTTTGGCGGATGAAAATCTTGCCCCCGAATCCTTT
AACACTTTGTTGGCAAATGGATCCATAAATCTGGTGGATAATGAAGCTCTGGTAGCATTTATCGCTGCCGCACGGCGCCGTCAGG
CAGTACACACGATACCTCTGAGAGTTAACTATTTACCTGACCCGGGCTTCGAAACTATAGACAGCCCGAGGAATTTTTTGGTAGA
CGGAGTCTTATACAATGGAATAATCATGATGAACTATGCTCAATATGACGCGACAGCTATTCCATCTCGCTATTTCTATGCATTG
CCTGTGAATGGATTCTTTATGAACCGTACTATCATAGAAGCATCTCACAGGGCTAACGTCAACCTTACTAATGTACCTGAGGACT
TGCCACTAGTTCCTACATTTTTGGGGAGCGAAGTGTATCGCTCCATAAGAGCACCGTCGTACCTATTTGCTTCGAGAGCTTCAAA
CTATGCCTCAAACTCGGTCGCGGCATACAGCTTGCTGGCAGGATACTTCAAGACGTCTCCCGTCGCACTTGTCCATCAGTTGAAA
TTAAATCTTCATCCAGGGTTTGCTCTCACTGTCGCCCGCCAAGACAGATTTGCGGCGGATCAAATTTTATTTGCCCGAAGAGCAT
CGGAATCTTACTATCTAGGATCTCCAGTTGTAACCAACCGCCCAGAGAATGATTCTCTCGTAATTGAGATAAGTCAGCCAAGGGG
GCACATAGATATGGGTCTGGGATTTACAGCGAGTCGTGTTCCTGCAAAAATTAATACAGTGGTTACAGACATGGGCAATCATTGC
CAAAATCTCTTCAATGCCCGCTACCCTGGTCAGTTTCGCCATGCGGAAGTTGCGGATTTTATTGCCTCTGAAATAACAGACAATG
ACTCTACAGCCTTGCCTAGAGCGCAGCCTCCCATTCTGCTGTCTTACGAAAAGGCACCCATCCCCCCTTGCATAGAGCGGGGACA
GCTTGCTACATGTGAGTTTCTTTTAACTCCAGTGACTGCCGATCTTGCATATTTTTATACATCAGCAAATCCCCGAGGGCGCAGC
AGCTGTATCGCGTGCACCAACTGCGAGGATCCATGCGCGAGTGAAACTGAAAAGGCTATGTATGACCATTCTACCCCAGATGCAG
CACACCCGTCTAGGGCTACGAATAATGCTTGGGCATCGCAAAAGTATTCAGTGGGAGACAAAATGTATAATGCGCGGCGGGGTTT
TATTACGGCCAGTGATTTTTACAGTCCTCTGAGCAAGTTTATGACGCCATCTAGAGCTGAAGATAAGAGTCGGTGCCTGGCTCGT
TTGATGAGAGATTCCTCTGCGGCAATTAGTTCAGTTACGGGTGACACTGAGTATCAATTTGTCGCCCCGCCCGGTGCCAACGAAT
TAGTAACTGATCCGTGCGCAATGTTTCAAGAAGCGTTCCCACCCCTTTGTAGCAGTGACAAGGTCCTCTTCGCAACCTACGAAGG
CCCTAATAGAGCATCTGGATCTGGTGCAAGAGAAAACCACTTTGCGCAGTATTTGATTCACGATAAGTCTCCAATCGCCAATGCG
TTAAAAACCCCATGCAACACGCGTCGCTAAGACGGAATTCGCAAACGATTTCTACCCGTGAAATCACGTCACCGCCTGGTCTGTT
TGGGCGGAGCCCTTTCAAAACTTCTTCATAAGTGTTTCCTGTGCAGATGATTTGCCCACGGAGGGATATATAGCGCATACCTAAC
TAAATTTCAGAGGACAGAAGTACAGATCGTATCAGCCAGAACGATGACAAGCGGACAGTACGAGATCCGCGTTGTTCTGCCAAAT
GGACTGACACGAGATGAAGAAGATCGTCTCAGAAGCTTGCGTGGGACTATTTTAATGGCGCCAATTTTGCGAAGGTGTGTTTTTC
TGCACGAAATCGATCAAAAATCATTTTTCGCCCATGGAAAAGAACCAGACTATGCAACTCTGCTCACAGCCTACAGGAGACGCTT
TCCCATATTAATAGTTTGTGTCGAAAACCGAGAATTGAGCGCCATCGCGCTATCCATAGGATACCCGCGCGGAATCAGCGTGCGC
AACACTGGTCCATTTAGTTTGAACAACGGTGATCTCGTGTCTTTACTGCCACCCATTACGAATACGCGTTTTCGCGTCGACTTGC
CATCCTGCGGCTCTGTAATCGAGCCAGCCATGACAATTCCTTTCGAGATCGGAACAGAACTGATGGGAAAGATTTTTGCAGGGAT
GGCCTATGACTTCTGCGTTAGAAATCAAATAGCGACAACCCGTCCTCGTGATATTTATGTTGTAACATATAAGAACAAGACACTC
GATCTATCTACATTGCCCCCGTCTGATGCGGCTGCACTTCAAGATACAATGAAAAGCCTATTTTCATCGGTTCTATTTTCGATAC
ACGAGGGGGTCATGTCCGTTTTGTCACTAATGCCTGCTTTATTGGCAGGAGGGGCGAACGATCCATTTCTCAATGCCATTCTACA
AATGCAAAGCATGACCAGACTTTCCGTGCAGCTTTTTAATCCACCCGCGCTAGAATTACCGGAACCTGCTGGCAGCTCTGGGCGA
TATCATGTATTTGATGCTTTTGCCGCGTGGCTTTCAATGTCTCACCGACTTGGTGATTTATTCAATTTGAAACCGGTCCTCAAAG
TAGTAATGTTTTACTCGGATGATTCTACCGCAGACGAGGGAGACCTCCTCAACGCCATTGTACCTTAACCCGTCTGTGCATCTCT
ATGAAATGACGAGACATCATATCTCTTTGATTTGTGAATAAAGTATTTATTATCGAAATTGAATTTATGCTGGTCTCTGTCTTCT
ACCGATCGGT

>UL18

ACAGACATGGGCAATCATTGCCAAAATCTCTTCAATGCCCGCTACCCTGGTCAGTTTCGCCATGCGGAAGTTGCGGATTTTATTG
CCTCTGAAATAACAGACAATGACTCTACAGCCTTGCCTAGAGCGCAGCCTCCCATTCTGCTGTCTTACGAAAAGGCACCCATCCC
CCCTTGCATAGAGCGGGGACAGCTTGCTACATGTGAGTTTCTTTTAACTCCAGTGACTGCCGATCTTGCATATTTTTATACATCA
GCAAATCCCCGAGGGCGCAGCAGCTGTATCGCGTGCACCAACTGCGAGGATCCATGCGCGAGTGAAACTGAAAAGGCTATGTATG
ACCATTCTACCCCAGATGCAGCACACCCGTCTAGGGCTACGAATAATGCTTGGGCATCGCAAAAGTATTCAGTGGGAGACAAAAT
GTATAATGCGCGGCGGGGGTTTTATTACGGCCAGTGATTTTTACAGTCCTCTGAGCAAGTTTATGACGCCATCTAGAGCTGAAGAT
AAGAGTCGGTGCCTGGCTCGTTTGATGAGAGATTCCTCTGCGGCAATTAGTTCAGTTACGGGTGACACTGAGTATCAATTTGTCG
CCCCGCCCGGTGCCAACGAATTAGTAACTGATCCGTGCGCAATGTTTCAAGAAGCGTTCCCACCCCTTTGTAGCAGTGACAAGGT
CCTCTTCGCAACCTACGAAGGCCCTAATAGAGCATCTGGATCTGGTGCAAGAGAAAACCACTTTGCGCAGTATTTGATTCACGAT
AAGTCTCCAATCGCCAATGCGTTAAAAACCCCATGCAACACGCGTCGCTAAGACGGAATTCGCAAACGATTTCTACCCGTGAAAT
CACGTCACCGCCTGGTCTGTTTGGGCGGAGCCCTTTCAAAACTTCTTCATAAGTGTTTCCTGTGCAGATGATTTGCCCACGGAGG
GATATATAGCGCATACCTAACTAAATTTCAGAGGACAGAAGTACAGATCGTATCAGCCAGAACGATGACAAGCGGACAGTACGAG
ATCCGCGTTGTTCTGCCAAATGGACTGACACGAGATGAAGAAGATCGTCTCAGAAGCTTGCGTGGGACTATTTTAATGGCGCCAA
TTTTGCGAAGGTGTGTTTTTCTGCACGAAATCGATCAAAAATCATTTTTCGCCCATGGAAAAGAACCAGACTATGCAACTCTGCT
CACAGCCTACAGGAGACGCTTTCCCATATTAATAGTTTGTGTCGAAAACCGAGAATTGAGCGCCATCGCGCTATCCATAGGATAC
CCGCGCGGAATCAGCGTGCGCAACACTGGTCCATTTAGTTTGAACAACGGTGATCTCGTGTCTTTACTGCCACCCATTACGAATA

CGCGTTTTCGCGTCGACTTGCCATCCTGCGGCTCTGTAATCGAGCCAGCCATGACAATTCCTTTCGAGATCGGAACAGAACTGAT
GGGAAAGATTTTTGCAGGGATGGCCTATGACTTCTGCGTTAGAAATCAAATAGCGACAACCCGTCCTCGTGATATTTATGTTGTA
ACATATAAGAACAAGACACTCGATCTATCTACATTGCCCCCGTCTGATGCGGCTGCACTTCAAGATACAATGAAAAGCCTATTTT
CATCGGTTCTATTTTCGATACACGAGGGGGTCATGTCCGTTTTGTCACTAATGCCTGCTTTATTGGCAGGAGGGGCGAACGATCC
ATTTCTCAATGCCATTCTACAAATGCAAAGCATGACCAGACTTTCCGTGCAGCTTTTTAATCCACCCGCGCTAGAATTACCGGAA
CCTGCTGGCAGCTCTGGGCGATATCATGTATTTGATGCTTTTGCCGCGTGGCTTTCAATGTCTCACCGACTTGGTGATTTATTCA
ATTTGAAACCGGTCCTCAAAGTAGTAATGTTTTACTCGGATGATTCTACCGCAGACGAGGGAGACCTCCTCAACGCCATTGTACC
TTAACCCGTCTGTGCATCTCTATGAAATGACGAGACATCATATCTCTTTGATTTGTGAATAAAGTATTTATTATCGAAATTGAAT
TTATGCTGGTCTCTGTCTTCTACCGATCGGT

>UL15A

GCAACCTGGTGGCTCTAATTCAGATGTTGCAGTATTTGATGAAAACCGAGCATCTGCTATCTTTTCAGTGCCTCGGGTCCCTAGC
GTTTTTACGAGAGCGCTCAACCCTTGGGCCACCGCCCTCAAATTTTCACTTCCTTGGGGAACGGGGGTGTACTTCTTCATGGGCA
GCATGGAAGCGCCTGCCACCTGTTTCAAGAGATGTGTAGACTCGTCTATCCTTTCTTTTATAGTAGCCTGTTTGGCTGACGTGTC
CGGTGGCGCATCAAATCGTTGTAGATTGTTTGAGAGATATTCAGCTGCGCTTATTACATATTGCCGGTTTTTGCCGAGTAATTCG
GAGAGCCGTTTTTTAATATCGCGAAAATACTTGGTTTGAGCTTCATTTAAATGTGATGGAGAGTACTCTAATAGCTGGCACGAAA
GCGCCTCTAATCTGGCACTGTCTACATTGTAGCGAGTAGAAGAGTTTACCAGCTCTCCATACGTGAATAGAAGTTCATAATATTC
TTCCAACGCAGCTCCTATAGCTATTCGTATAGCCACCGCAGTAAAATCATCATATGTCATCTTTTCTGCGTCGCAATATACCGGG
ATAGTGATGAACAGTCCTCCCCAGTCATTCCCACCATCATACGTAATTCGAAAGGGGCGCTCTTGACAATTCCCCAACTCAGGTA
ATAGAATTTTGTCGAGTGCCCGGTGTCCGACAAAAGTTGAGAAAACTTGTTTCCACTCTGTGCAGTCAAAACATCCGTGGTGTCG
AATTTGAATTTCTGTATAAAACTTATTGTCACTAAGAAGGTGAACCGGAACTTCAAATCCTTCTAAGCACTGCGTGGACAGGACT
AGGTGACTTAGTATCGGCGCAGGACCTCCCTCCTGTTCACCGCAAGATAATAGCTTCGATGTCGCTGCTAAGACATGTACTTCCA
TTGCGACGCGCTGGTGAATGAATCATGTCCGTTGAGTGTGATCTTTTATTGCTGGTCTGTTTTTGCAGGGTCTCAGGGGGCAGGA
CTTCAATTTCCTGTTCGTCGATGAAGCAAATTTTATAAAACCGGCTGCTCTCCACACGGTGATGGGGTTTTTGAATCAGACTAAC
TGCAAACTTTTCTTCGTATCTTCTACCAACACTTGCCATAGTAACACGAGTCTACTGTACAATCTTAAGGGAAAGACAAACTCTC
TGCTTAATGTTGTTACCTATATATGCGACGAACACATGCCTGAAATCCAGAAAAGAACAGACGTGACTACCTGTTCTTGTTATGT
TTTGCACAAACCAGTCTTTGTTAGTATGGACAGCGAGGTTCGAAATACCGCAGATTTGTTCGTAAAAGATTCGTTTATGCATGAG
ATTGCCGGTGGACGTGCTGGAAAATATGACTCCGATCGTACACTTGTTCCTGTGCGAGCTTTGGATCAGTTTTTGATTTACAGAC
CATCTACCTCAAGTAAACCCAACATCTCTGGGCTCGGGAAAATTCTCACTGTGTATGTAGATCCTGCATTTACCACAAACAGAAG
CGCATCTGGTACTGGCATAGCATTAGTGACTGCGCTGCGAGACTCTATGGTTCTCATGGGAGCTGAACATTTTTACTTGGATGCA
CTGACCGGAGAGGCGGCGCTGGAAATTGCCCAGTGCGTGTATTTGTGTATAGCATACTGTTGTCTTATTCACGCTGGCGCATTTC
GAGAAATTCGTATAGCGGTGGAGGGGAATTCGAGTCAGGATTCTGCTGCTGCCATAGCCGGCAACTTGACAGAGCTTTTGGATTC
TTTGCGTAGGAGGTTAGGCTTTTCTCTGACTTTTGCCCACAGTCGGCAGCCTGGAACTGCAATGGCTCACCCTTTTTATCTTCTA
AACAAACAAAAGAGCAGAGCATTTGACTTGTTTGTCAGCCTATTTAACTCTGGGCGCTTTATGGCTTCCCAGGAACTTGTCTCGA
ACACCCTTGTACTCAGTAAAGATCCATGCGAGTATCTCGTAGACCAAATAAGGAACATTACTGTTACTCATGGCCAAGGGCCCGA
TTCCTTTCGAACTTTTAGCGGTAAGCAGGGCCGCGTACCCGATGACATGTTAGTGGCCGCTGTAATGTCTACATACTTAGCACTA
GAAGGTTCCCCGACTGCCGGATATCACCCCATCGCACCGATCGGTAGAAGACAGAGACCAGCATAAATTCAATTTCGATAATAAA
TACTTTATTCACAAATCAAAGAGATATGATGTCTCGTCATTTCATAGAGA

>UL17

AGCCATAAAGCGCCCAGAGTTAAATAGGCTGACAAACAAGTCAAATGCTCTGCTCTTTTGTTTGTTTAGAAGATAAAAAGGGTGA
GCCATTGCAGTTCCAGGCTGCCGACTGTGGGCAAAAGTCAGAGAAAAGCCTAACCTCCTACGCAAAGAATCCAAAAGCTCTGTCA
AGTTGCCGGCTATGGCAGCAGCAGAATCCTGACTCGAATTCCCCTCCACCGCTATACGAATTTCTCGAAATGCGCCAGCGTGAAT
AAGACAACAGTATGCTATACACAAATACACGCACTGGGCAATTTCCAGCGCCGCCTCTCCGGTCAGTGCATCCAAGTAAAAATGT
TCAGCTCCCATGAGAACCATAGAGTCTCGCAGCGCAGTCACTAATGCTATGCCAGTACCAGATGCGCTTCTGTTTGTGGTAAATG
CAGGATCTACATACACAGTGAGAATTTTCCCGAGCCCAGAGATGTTGGGTTTACTTGAGGTAGATGGTCTGTAAATCAAAAACTG
ATCCAAAGCTCGCACAGGAACAAGTGTACGATCGGAGTCATATTTTCCAGCACGTCCACCGGCAATCTCATGCATAAACGAATCT
TTTACGAACAAATCTGCGGTATTTCGAACCTCGCTGTCCATACTAACAAAGACTGGTTTGTGCAAAACATAACAAGAACAGGTAG
TCACGTCTGTTCTTTTCTGGATTTCAGGCATGTGTTCGTCGCATATATAGGTAACAACATTAAGCAGAGAGTTTGTCTTTCCCTT
AAGATTGTACAGTAGACTCGTGTTACTATGGCAAGTGTTGGTAGAAGATACGAAGAAAAGTTTGCAGTTAGTCTGATTCAAAAAC
CCCATCACCGTGTGGAGAGCAGCCGGTTTTATAAAATTTGCTTCATCGACGAACAGGAAATTGAAGTCCTGCCCCCTGAGACCCT
GCAAAAACAGACCAGCAATAAAAGATCACACTCAACGGACATGATTCATTCACCAGCGCGTCGCAATGGAAGTACATGTCTTAGC
AGCGACATCGAAGCTATTATCTTGCGGTGAACAGGAGGGAGGTCCTGCGCCGATACTAAGTCACCTAGTCCTGTCCACGCAGTGC
TTAGAAGGATTTGAAGTTCCGGTTCACCTTCTTAGTGACAATAAGTTTTATACAGAAATTCAAATTCGACACCACGGATGTTTTG
ACTGCACAGAGTGGAAACAAGTTTTCTCAACTTTTGTCGGACACCGGGCACTCGACAAAATTCTATTACCTGAGTTGGGGAATTG
TCAAGAGCGCCCCTTTCGAATTACGTATGATGGTGGGAATGACTGGGGAGGACTGTTCATCACTATCCCGGTATATTGCGACGCA
GAAAAGATGACATATGATGATTTTACTGCGGTGGCTATACGAATAGCTATAGGAGCTGCGTTGGAAGAATATTATGAACTTCTAT
TCACGTATGGAGAGCTGGTAAACTCTTCTACTCGCTACAATGTAGACAGTGCCAGATTAGAGGCGCTTTCGTGCCAGCTATTAGA
GTACTCTCCATCACATTTAAATGAAGCTCAAACCAAGTATTTTCGCGATATTAAAAAACGGCTCTCCGAATTACTCGGCAAAAAC
CGGCAATATGTAATAAGCGCAGCTGAATATCTCTCAAACAATCTACAACGATTTGATGCGCCACCGGACACGTCAGCCAAACAGG
CTACTATAAAAGAAAGGATAGACGAGTCTACACATCTCTTGAAACAGGTGGCAGGCGCTTCCATGCTGCCCATGAAGAAGTACAC
CCCCGTTCCCCAAGGAAGTGAAAATTTGAGGGCGGTGGCCCAAGGGTTGAGCGCTCTCGTAAAAACGCTAGGGACCCGAGGCACT
GAAAAGATAGCAGATGCTCGGTTTTCATCAAATACTGCAACATCTGAATTAGAGCCACCAGGTTGCAGCAGAGGTTTCGAAAAGG
CGGCACCACCAATGCCTGATATAAGAAATGTTATCATGACAGACAGGCAAAAAACCACAATGTCTATTACATCAAGATCGGTAAC
TTATTATACTATGTATGATTGTCTGGCAGCAGCCTGTGAAATCATGGAGGCAGAAAATAATTCTACCAGAAGATCATGGTCACTC
GGAAAAATCTCAATTGTGTTGATGTCGTGCTATAACTCTGGAGCTCCAATCATGGTTGTAAATTATTCCCAAGACTCGTTAAAGT
TATGCTATCCAAAGGTTAGATCTGGAGCTTCAATACTAGCTACAATAATGAGACCGCTGGGAGAAACCTCAGTACAAAATTTAAC
CGAAGCAGCTCAGCAAGAAATAGTGAAGAATGCCCCATGTTTTCAATGTCCTTTGAAAGCTTTGAAAGATCCAGATTCTCTAACT
AAATTGTTTGCAGTGGACTGTGAACAGTTGCTCCTTTATAGTTTCCAAGCGCGCGTGGCTGCAGCATTCACCGCAGCGGTGTCCG
AGGCCATTGTACGAGCAACTCACACAGAGAGCAATACTACTCGGCTCGGACAGCTAATAAACTACGACATCCCTTTATACCAGCC
ATTTGACAGCTTGGAAGCAAACTGGCCAACACAAATAGCACGTATGGCTTCAGACCTAATGGCAGTTCTCTGTATAACTCTTCAA
AGTGAACGCTTGTCCGTCTTTATGAAAAGTGGAATATGGCGCGCCATCTTGGCAATGTTAATTCGAAAGGAAAATCACCGAAGAC
CATACCTGGCCCCTGTTCATGTGGAGGATGACGTGTACCTATTCGATTATTTCAGATTCGGATCGAGTAATATTGTAAGAATTAC
CACAGAGCCTGTTGTGTTAAAGGCTCGCAAACCCAAAATAGACGGCCTCTACGAACTTGATTTCATCGCTAGTCCCCCCTCTGGC

CTCCACCCGTGGACTAAACAAAAATTCCGCCCAGGAGAATTTCACTCCTATATATGCGTTGGTTTCAATTCGGCCCTGAATGCTC
TTCTCATCTTTCCTGGAGGGTTTGGATTAGAATTTGATTTTGGCGAAGCACTAAAAGAAGTCTGGGAGGATCACCTGGATCACTT
AGTATTGAAACGCTTTAGTAGGCGTGCACCATACTACCCTGAATCTTATCTCCCGAGCTGCGAATTCATGCAAAAGATAAACACA
GCTGGGTAGCTAGAAATTGCGCCATTGGCACGTCTGCAGGCATTCTTGCTTAAATGTAATAGCGTAGGACATAGAAAGCAGTTAA
TACCGAGTAATTAACAATTAGTGGTCAAATGGATTACTGTTAGGACTAGGTAAATGTACAGAATACTTACATTCGTATTTTGACT
GGAGGCGAAGACAATCGCACTTTTGTTTCCATTGCGAAATGTAAATGTGATCGTTTCCCCCTTGACCTGTTCTACCTGTTTTGCG
CCAAACCATTTGTACAGGCGAGTAAAAATTTCAATAAAAACGGGCTCTGTTGCTTTTCTTAAGTGGGCAGTGTATCCAATGCGTA
TTC

>UL15B

CGTACGTTTTTAGGTCATGGCTCAACGGTTCCATTACAAAAGCTCTCTCGATTTTTGAAAAGGCATAGATCTGTAAGAAATGTTT
GGCGATGTCGGAAAGTTCTGGTCTTCTTGCCGCATTGGAAAGAATCAAGGACATTGCTAGTTCCCATCGAAAATGACCATCCGAT
TCAAATATTTTTACTGCCATGTGGTACTTTGGGGCATACTGCACTGATCCATATCCACCATTTCCAATTACCGTTGACCCTTCTA
GGTTAAACCATCGGTAGCTCATTTTTGGATTGACCTTGAGACATATTTTGTTAGGTTTCGGCATTATGTAGATCGCAAGCTGTCT
CGATTCTTTTGTCAAATCTTCTGAGACCGGGGCTATATGCTCAAGGTCGATTCCATCATCGTTGCTTAGTATTCTTCTACTGGGC
ATTCGTCTTCTTGAAAAGTTGGACTGGGAGTCGGTGGAGTTCTTGGCGAGTTCCCGCGACTGACGGGACTTGGTGCATGTGGTGA
TGGCAGTAGCGTCGGAGGATTTACGTTTTCTAGTTGCCATTTTAATAAAAGGTCGTCTTCCTCACAATCTAAGCTGTCTCTGCGA
TCAGGGTCACCTGAAATATACAGAGCAATTTCTTCTTCTGCCCGCTGCAAGTCTAGCTCCGCATCTGCAATGGCTTCTGACTCTT
GAGCGAATGTTTCTAAAAAGTCGTTCCTTAGATATCTTCTGTAGCCATCCAACTCCAGTCTTAATGCTCTCTGACTCTCTGTAGC
TGCTTTTATCCGAGCAACATGACACCTGACAGAGCTGAGGCGCTTATTTGAGCGCACGTTTATCTCTGAAACAGCTTCCTGCTCC
CTAGCTGATGCAAGAAAATAGCGCAGAGCTGGGTGCATTCGAGGAACTTTGAGAGACATCATCTCGATGACCGCAGCTTTATAGG
CTTGCCCTTTAATGTAGTTCATAACTTCGGCCCGAGCCCTGTTTCGGTAGTCTCTCGCAGAAGACATGCTTGGAAAGGAATCGGT
CGAAATTGTCAAAAGATACCGCGATGCTTTGCGTAAGCGAACAATGGAACGAGGACCAGACGACGTGGATGGGCAAGAAATGAGT
GACTCCAATTTTATAACTACTGCAAGTATATGTGATAGGAATGATTCAGCGAGGGATACCATGAATTCACCGGCGAGCAGATTTC
AATTTGCGATTGATGTTCCACAGCGCCATCAAGCGTGCATTGCCCCTATAGGATCTTTTCATAATTGCTGTGCTATTTCGCGTGC
GTTTTCATATATGGCGTCAGAAATCATTTACGAAAATCTGGCCTCCTATTCTACCAAGTACACAGATACCGATGCCGCACTAAAC
GATTTACAGGTATCCCCTAAACGGCAACTTTTCACAGGTGCAGCCGAGGACAGTATCCTTCCGGCACTAAGACAGAAACTCGCAA
ACTTAAATTTTGCCAGGTTTGCTCCAAGCGACTCCCTAATACATGACAAAGCATTCGACGGAATAATGAATGGCTACCGGGGGTT
TGTCAAGTCTGATGAATTTTCCCAACTGAACAACTTTATTTACAGATTTCACACACTGTTAAAAAAGTCATTTTCTGGTCAAGCT
TCAAATGACTACAAACGGGCAAAGTTGGAAAAAACAACTAGCGAACAGAGAGACGGAACACTGGAATTATTTCAAAAAATGATCT
TAATGCATGCTACATACTTTGCGTCAAGTATTTGCTTGGGGGAAGGATCCACGGAACGAAGTAATAGATACTTGTCTACCGTTTT
TAATACCTCATTATTTTCAGAGAATATCATTCAACATTTTAGACAAAGAACTACTGTATTTTTAGTTCCAAGGAGGCACGGGAAA
ACGTGGTTTTTAGTTCCGCTCATTTCCCTACTGGTTTCATCTTTTGAAGGAATACGCATTGGATACACTGCCCACTTAAGAAAAG
CAACAGAGCCCGTTTTTATTGAAATTTTTACTCGCCTGTACAAATGGTTTGGCGCAAAACAGGTAGAACAGGTCAAGGGGGAAAC
GATCACATTTACATTTCGCAATGGAAACAAAAGTGCGATTGTCTTCGCCTCCAGTCAAAATACGAATGTAAGTATTCTGTACATT
TACCTAGTCCTAACAGTAATCCATTTGACCACTAATTGTTAATTACTCGGTATTAACTGCTTTCTATGTCCTACGCTATTACATT
TAAGCAAGAATGCCTGCAGACGTGCCAATGGCGCAATTTCTAGCTACCCAGCTGTGTTTATCTTTTGCATGAATTCGCAGCTCGG
GAGATAAGATTCAGGGTAGTATGGTGCACGCCTACTAAAGCGTTTCAATACTAAGTGATCCAGGTGATCCTCCCAGACTTCTTTT
AGTGCTTCGCCAAAATCAAATTCTAATCCAAACCCTCCAGGAAAGATGAGAAGAGCATTCAGGGCCGAATTGAAACCAACGCATA
TATAGGAGTGAAATTCTCCTGGGCGGAATTTTTGTTTAGTCCACGGGTGGAGGCCAGAGGGGGGACTAGCGATGAAATCAAGTTC
GTAGAGGCCGTCTATTTTGGGTTTGCGAGCCTTTAACACAACAGGCTCTGTGGTAATTCTTACAATATTACTCGATCCGAATCTG
AAATAATCGAATAGGTACACGTCATCCTCCACATGAACAGGGGCCAGGTATGGTCTTCGGTGATTTTCCTTTCGAATTAACATTG
CCAAGATGGCGCGCCATATTCCACTTTTCATAAAGACGGACAAGCGTTCACTTTGAAGAGTTATACAGAGAACTGCCATTAGGTC
TGAAGCCATACGTGCTATTTGTGTTGGCCAGTTTGCTTCCAAGCTGTCAAATGGCTGGTATAAAGGGATGTCGTAGTTTATTAGC
TGTCCGAGCCGAGTAGTATTGCTCTCTGTGTGAGTTGCTCGTACAATGGCCTCGGACACCGCTGCGGTGAATGCTGCAGCCACGC
GCGCTTGGAAACTATAAAGGAGCAACTGTTCACAGTCCACTGCAAACAATTTAGTTAGAGAATCTGGATCTTTCAAAGCTTTCAA
AGGACATTGAAAACATGGGGCATTCTTCACTATTTCTTGCTGAGCTGCTTCGGTTAAATTTTGTACTGAGGTTTCTCCCAGCGGT
CTCATTATTGTAGCTAGTATTGAAGCTCCAGATCTAACCTTTGGATAGCATAACTTTAACGAGTCTTGGGAATAATTTACAACCA
TGATTGGAGCTCCAGAGTTATAGCACGACATCAACACAATTGAGATTTTTCCGAGTGACCATGATCTTCTGGTAGAATTATTTTC
TGCCTCCATGATTTCACAGGCTGCTGCCAGACAATCATACATAGTATAATAAGTTACCGATCTTGATGTAATAGACATTGTGGTT
TTTTGCCTGTCTGTCATGATAACATTTCTTATATCAGGCATTGGTGGTGCCGCCTTTTCGAAACCTCTGCTGCAACCTGGTGGCT
CTAATTCAGATGTTGCAGTATTTGATGAAAACCGAGCATCTGCTATCTTTTCAGTGCCTCGGGTCCCTAGCGTTTTTACGAGAGC
GCTCAACCCTTGGGCCACCGCCCTCAAATTTTCACTTCCTTGGGGAACGGGGGTGTACTTCTTCATGGGCAGCATGGAAGCGCCT
GCCACCTGTTTCAAGAGATGTGTAGACTCGTCTATCCTTTCTTTTATAGTAGCCTGTTTGGCTGACGTGTCCGGTGGCGCATCAA
ATCGTTGTAGATTGTTTGAGAGATATTCAGCTGCGCTTATTACATATTGCCGGTTTTTGCCGAGTAATTCGGAGAGCCGTTTTTT
AATATCGCGAAAATACTTGGTTTGAGCTTCATTTAAATGTGATGGAGAGTACTCTAATAGCTGGCACGAAAGCGCCTCTAATCTG
GCACTGTCTACATTGTAGCGAGTAGAAGAGTTTACCAGCTCTCCATACGTGAATAGAAGTTCATAATATTCTTCCAACGCAGCTC
CTATAGCTATTCGTATAGCCACCGCAGTAAAATCATCATATGTCATCTTTTCTGCGTCGCAATATACCGGGATAGTGATGAACAG
TCCTCCCCAGTCATTCCCACCATCATACGTAATTCGAAAGGGGCGCTCTTGACAATTCCCCAACTCAGGTAATAGAATTTTGTCG
AGTGCCCGGTGTCCGACAAAAGTTGAGAAAACTTGTTTCCACTCTGTGCAGTCAAAACATCCGTGGTGTCGAATTTGAATTTCTG
TATAAAACTTATTGTCACTAAGAAGGTGAACCGGAACTTCAAATCCTTCTAAGCACTGCGTGGACAGGACTAGGTGACTTAGTAT
CGGCGCAGGACCTCCCTCCTGTTCACCGCAAGATAATAGCTTCGATGTCGCTGCTAAGACATGTACTTCCATTGCGACGCGCTGG
TGAATGAATCATGTCCGTTGAGTGTGATCTTTTATTGCTGGTCTGTTTTTGCAGGGTCTCAGGGGGCAGGACTTCAATTTCCTGT
TCGTCGATGAAGCAAATTTTATAAAACCGGCTGCTCTCCACACGGTGATGGGGTTTTGAATCAGACTAACTGCAAACTTTCTT
CGTATCTTCTACCAACACTTGCCATAGTAACACGAGTCTACTGTACAATCTTAAGGGAAAGACAAACTCTCTGCTTAATGTTGTT
ACCTATATATGCGACGAACACATGCCTGAAATCCAGAAAAGAACAGACGTGACTACCTGTTCTTGTTATGTTTTGCACAAACCAG
TCTTTGTTAGTATGGACAGCGAGGTTCGAAATACCGCAGATTTGTTCGTAAAAGATTCGTTTATGCATGAGATTGCCGGTGGACG
TGCTGGAAAATATGACTCCGATCGTACACTTGTTCCTGTGCGAGCTTTGGATCAGTTTTTGATTTACAGACCATCTACCTCAAGT
AAACCCAACATCTCTGGGCTCGGGAAAATTCTCACTGTGTATGTAGATCCTGCATTTACCACAAACAGAAGCGCATCTGGTACTG
GCATAGCATTAGTGACTGCGCTGCGAGACTCTATGGTTCTCATGGGAGCTGAACATTTTTACTTGGATGCACTGACCGGAGAGGC
GGCGCTGGAAATTGCCCAGTGCGTGTATTTGTGTATAGCATACTGTTGTCTTATTCACGCTGGCGCATTTCGAGAAATTCGTATA
GCGGTGGAGGGGAATTCGAGTCAGGATTCTGCTGCTGCCATAGCCGGCAACTTGACAGAGCTTTTGGATTCTTTGCGTAGGAGGT
TAGGCTTTTCTCTGACTTTTGCCCACAGTCGGCAGCCTGGAACTGCAATGGCTCACCCTTTTTATCTTCTAAACAAACAAAAGAG
CAGAGCATTTGACTTGTTTGTCAGCCTATTTAACTCTGGGCGCTTTATGGCTTCCCAGGAACTTGTCTCGAACACCCTTGTACTC

AGTAAAGATCCATGCGAGTATCTCGTAGACCAAATAAGGAACATTACTGTTACTCATGGCCAAGGGCCCGATTCCTTTCGAACTT
TTAGCGGTAAGCAGGGCCGCGTACCCGATGACATGTTAGTGGCCGCTGTAATGTCTACATACTTAGCACTAGAAGGTTCCCCGAC
TGCCGGATATCACCCCATCGCACCGATCGGTAGAAGACAGAGACCAGCATAAATTCAATTTCGATAATAAATACTTTATTCACAA
ATCAAAGAGATATGATGTCTCGTCATTTCATAGAGA

>UL14

CCATTTGTACAGGCGAGTAAAAATTTCAATAAAAACGGGCTCTGTTGCTTTTCTTAAGTGGGCAGTGTATCCAATGCGTATTCCT
TCAAAAGATGAAACCAGTAGGGAAATGAGCGGAACTAAAAACCACGTTTTCCCGTGCCTCCTTGGAACTAAAAATACAGTAGTTC
TTTGTCTAAAATGTTGAATGATATTCTCTGAAAATAATGAGGTATTAAAAACGGTAGACAAGTATCTATTACTTCGTTCCGTGGA
TCCTTCCCCCAAGCAAATACTTGACGCAAAGTATGTAGCATGCATTAAGATCATTTTTTGAAATAATTCCAGTGTTCCGTCTCTC
TGTTCGCTAGTTGTTTTTTCCAACTTTGCCCGTTTGTAGTCATTTGAAGCTTGACCAGAAAATGACTTTTTTAACAGTGTGTGAA
ATCTGTAAATAAAGTTGTTCAGTTGGGAAAATTCATCAGACTTGACAAACCCCCGGTAGCCATTCATTATTCCGTCGAATGCTTT
GTCATGTATTAGGGAGTCGCTTGGAGCAAACCTGGCAAAATTTAAGTTTGCGAGTTTCTGTCTTAGTGCCGGAAGGATACTGTCC
TCGGCTGCACCTGTGAAAAGTTGCCGTTTAGGGGATACCTGTAAATCGTTTAGTGCGGCATCGGTATCTGTGTACTTGGTAGAAT
AGGAGGCCAGATTTTCGTAAATGATTTCTGACGCCATATATGAAAACGCACGCGAAATAGCACAGCAATTATGAAAAGATCCTAT
AGGGGCAATGCACGCTTGATGGCGCTGTGGAACATCAATCGCAAATTGAAATCTGCTCGCCGGTGAATTCATGGTATCCCTCGCT
GAATCATTCCTATCACATATACTTGCAGTAGTTATAAAATTGGAGTCACTCATTTCTTGCCCATCCACGTCGTCTGGTCCTCGTT
CCATTGTTCGCTTACGCAAAGCATCGCGGTATCTTTTGACAATTTCGACCGATTCCTTTCCAAGCATGTCTTCTGCGAGAGACTA
CCGAAACAGGGCTCGGGCCGAAGTTATGAACTACATTAAAGGGCAAGCCTATAAAGCTGCGGTCATCGAGATGATGTCTCTCAAA
GTTCCTCGAATGCACCCAGCTCTGCGCTATTTTCTTGCATCAGCTAGGGAGCAGGAAGCTGTTTCAGAGATAAACGTGCGCTCAA
ATAAGCGCCTCAGCTCTGTCAGGTGTCATGTTGCTCGGATAAAAGCAGCTACAGAGAGTCAGAGAGCATTAAGACTGGAGTTGGA
TGGCTACAGAAGATATCTAAGGAACGACTTTTTAGAAACATTCGCTCAAGAGTCAGAAGCCATTGCAGATGCGGAGCTAGACTTG
CAGCGGGCAGAAGAAGAAATTGCTCTGTATATTTCAGGTGACCCTGATCGCAGAGACAGCTTAGATTGTGAGGAAGACGACCTTT
TATTAAAATGGCAACTAGAAAACGTAAATCCTCCGACGCTACTGCCATCACCACATGCACCAAGTCCCGTCAGTCGCGGGAACTC
GCCAAGAACTCCACCGACTCCCAGTCCAACTTTTCAAGAAGACGAATGCCCAGTAGAAGAATACTAAGCAACGATGATGGAATCG
ACCTTGAGCATATAGCCCCGGTCTCAGAAGATTTGACAAAAGAATCGAGACAGCTTGCGATCTACATAATGCCGAAACCTAACAA
AATATGTCTCAAGGTCAATCCAAAAATGAGCTACCGATGGTTTAACCTAGAAGGGTCAACGGTAATTGGAAATGGTGGATATGGA
TCAGTGCAGTATGCCCCAAAGTACCACATGGCAGTAAAAATATTTGAATCGGATGGTCATTTTCGATGGGAACTAGCAATGTCCT
TGATTCTTTCCAATGCGGCAAGAAGACCAGAACTTTCCGACATCGCCAAACATTTCTTACAGATCTATGCCTTTTCAAAAATCGA
GAGAGCTTTTGTAATGGAACCGTTGAGCCATGACCTAAAAACGTACGCCAAAAGATACAAAGACAATTTTACAATGGAAACATTG
AATACCCTGACATCTGAGTTCAAAGGACTTGCGAAGGCCTTGGCATTTCTGAACATTGATTGCGGATTGGTACACATGGACGTCA
AGAGTAACAACATACTGGTCAAATGTGATGCGAATGGAAAACTTAGTCGACTTGTATTGGCAGACTTTAGTTTGACAGGCCCTAA
TACTAATTCTATTCTCAACCAGAGCATGATGGTGTGTCCTTCAAGAGGAGTAGTGGAAGGGCTAAAAATTATAGATTCAACGACT
GTAAAGAATCACATTCCATCTGACTCTTTTATCATATATAACGGCCACTGCCGAGCTCCAGAAGTGATAATAAATTATTGTAATG
GGAAAAGGTATCGCGACCAGCCAATGGACGCCCTGGAAACACTTGGATTGGATCTATTTTCCCTTGGGCAAGTAGTGCAGGAAAT
CTTATTTGAGGGAATTCTGTGCAGATCGCGCGAGTTTTCGTTTAAACCAAAACCAAACACTATTCATGAAAAATTAAGTCACGAT
TACATGATGAGGGTTTTGGCATATCGAATTGTGCTTTCCGACAATCTAATCTCTAGAGGCTGTGATCTAAGTTTCACCGGTCCAC
TAAGTGGAACAGTAGAATCAGTGAACGCATCCCTATTTCGAGAATTAGACAGAATCCTTTTCCAGAGTCACGTCGAAATGTATGA
GCGCATCGATCTACGAGAAAAACTGCTGAACGTCATCATTCCTCCGGAAACCAGAGGACTTTGTACCCTGGCGGGACTGCTATGT
CACTGGGATGCGGATTTGCGCCGCTCTGCAGTGACCTTCTTTTAGAGATTGTGATCGTTACTTATACCAATAAACTTTTTCAGAG
CTTTGTATTACGTCGTTTTTTTGTGACCGGGGACTCTATCA

>UL13

GAAAATTCATCAGACTTGACAAACCCCCGGTAGCCATTCATTATTCCGTCGAATGCTTTGTCATGTATTAGGGAGTCGCTTGGAG
CAAACCTGGCAAAATTTAAGTTTGCGAGTTTCTGTCTTAGTGCCGGAAGGATACTGTCCTCGGCTGCACCTGTGAAAAGTTGCCG
TTTAGGGGATACCTGTAAATCGTTTAGTGCGGCATCGGTATCTGTGTACTTGGTAGAATAGGAGGCCAGATTTTCGTAAATGATT
TCTGACGCCATATATGAAAACGCACGCGAAATAGCACAGCAATTATGAAAAGATCCTATAGGGGCAATGCACGCTTGATGGCGCT
GTGGAACATCAATCGCAAATTGAAATCTGCTCGCCGGTGAATTCATGGTATCCCTCGCTGAATCATTCCTATCACATATACTTGC
AGTAGTTATAAAATTGGAGTCACTCATTTCTTGCCCATCCACGTCGTCTGGTCCTCGTTCCATTGTTCGCTTACGCAAAGCATCG
CGGTATCTTTTGACAATTTCGACCGATTCCTTTCCAAGCATGTCTTCTGCGAGAGACTACCGAAACAGGGCTCGGGCCGAAGTTA
TGAACTACATTAAAGGGCAAGCCTATAAAGCTGCGGTCATCGAGATGATGTCTCTCAAAGTTCCTCGAATGCACCCAGCTCTGCG
CTATTTTCTTGCATCAGCTAGGGAGCAGGAAGCTGTTTCAGAGATAAACGTGCGCTCAAATAAGCGCCTCAGCTCTGTCAGGTGT
CATGTTGCTCGGATAAAAGCAGCTACAGAGAGTCAGAGAGCATTAAGACTGGAGTTGGATGGCTACAGAAGATATCTAAGGAACG
ACTTTTTAGAAACATTCGCTCAAGAGTCAGAAGCCATTGCAGATGCGGAGCTAGACTTGCAGCGGGCAGAAGAAGAAATTGCTCT
GTATATTTCAGGTGACCCTGATCGCAGAGACAGCTTAGATTGTGAGGAAGACGACCTTTTATTAAAATGGCAACTAGAAAACGTA
AATCCTCCGACGCTACTGCCATCACCACATGCACCAAGTCCCGTCAGTCGCGGGAACTCGCCAAGAACTCCACCGACTCCCAGTC
CAACTTTTCAAGAAGACGAATGCCCAGTAGAAGAATACTAAGCAACGATGATGGAATCGACCTTGAGCATATAGCCCCGGTCTCA
GAAGATTTGACAAAAGAATCGAGACAGCTTGCGATCTACATAATGCCGAAACCTAACAAAATATGTCTCAAGGTCAATCCAAAAA
TGAGCTACCGATGGTTTAACCTAGAAGGGTCAACGGTAATTGGAAATGGTGGATATGGATCAGTGCAGTATGCCCCAAAGTACCA
CATGGCAGTAAAAATATTTGAATCGGATGGTCATTTTCGATGGGAACTAGCAATGTCCTTGATTCTTTCCAATGCGGCAAGAAGA
CCAGAACTTTCCGACATCGCCAAACATTTCTTACAGATCTATGCCTTTTCAAAAATCGAGAGAGCTTTTGTAATGGAACCGTTGA
GCCATGACCTAAAAACGTACGCCAAAAGATACAAAGACAATTTTACAATGGAAACATTGAATACCCTGACATCTGAGTTCAAAGG
ACTTGCGAAGGCCTTGGCATTTCTGAACATTGATTGCGGATTGGTACACATGGACGTCAAGAGTAACAACATACTGGTCAAATGT
GATGCGAATGGAAAACTTAGTCGACTTGTATTGGCAGACTTTAGTTTGACAGGCCCTAATACTAATTCTATTCTCAACCAGAGCA
TGATGGTGTGTCCTTCAAGAGGAGTAGTGGAAGGGCTAAAAATTATAGATTCAACGACTGTAAAGAATCACATTCCATCTGACTC
TTTTATCATATATAACGGCCACTGCCGAGCTCCAGAAGTGATAATAAATTATTGTAATGGGAAAAGGTATCGCGACCAGCCAATG
GACGCCCTGGAAACACTTGGATTGGATCTATTTTCCCTTGGGCAAGTAGTGCAGGAAATCTTATTTGAGGGAATTCTGTGCAGAT
CGCGCGAGTTTTCGTTTAAACCAAAACCAAACACTATTCATGAAAAATTAAGTCACGATTACATGATGAGGGTTTTGGCATATCG
AATTGTGCTTTCCGACAATCTAATCTCTAGAGGCTGTGATCTAAGTTTCACCGGTCCACTAAGTGGAACAGTAGAATCAGTGAAC
GCATCCCTATTTCGAGAATTAGACAGAATCCTTTTCCAGAGTCACGTCGAAATGTATGAGCGCATCGATCTACGAGAAAAACTGC
TGAACGTCATCATTCCTCCGGAAACCAGAGGACTTTGTACCCTGGCGGGACTGCTATGTCACTGGGATGCGGATTTGCGCCGCTC
TGCAGTGACCTTCTTTTAGAGATTGTGATCGTTACTTATACCAATAAACTTTTTCAGAGCTTTGTATTACGTCGTTTTTTTGTGA
CCGGGGACTCTATCA

>UL12

TGGAAACATTGAATACCCTGACATCTGAGTTCAAAGGACTTGCGAAGGCCTTGGCATTTCTGAACATTGATTGCGGATTGGTACA
CATGGACGTCAAGAGTAACAACATACTGGTCAAATGTGATGCGAATGGAAAACTTAGTCGACTTGTATTGGCAGACTTTAGTTTG
ACAGGCCCTAATACTAATTCTATTCTCAACCAGAGCATGATGGTGTGTCCTTCAAGAGGAGTAGTGGAAGGGCTAAAAATTATAG
ATTCAACGACTGTAAAGAATCACATTCCATCTGACTCTTTTATCATATATAACGGCCACTGCCGAGCTCCAGAAGTGATAATAAA
TTATTGTAATGGGAAAAGGTATCGCGACCAGCCAATGGACGCCCTGGAAACACTTGGATTGGATCTATTTTCCCTTGGGCAAGTA
GTGCAGGAAATCTTATTTGAGGGAATTCTGTGCAGATCGCGCGAGTTTTCGTTTAAACCAAAACCAAACACTATTCATGAAAAAT
TAAGTCACGATTACATGATGAGGGTTTTGGCATATCGAATTGTGCTTTCCGACAATCTAATCTCTAGAGGCTGTGATCTAAGTTT
CACCGGTCCACTAAGTGGAACAGTAGAATCAGTGAACGCATCCCTATTTCGAGAATTAGACAGAATCCTTTTCCAGAGTCACGTC
GAAATGTATGAGCGCATCGATCTACGAGAAAAACTGCTGAACGTCATCATTCCTCCGGAAACCAGAGGACTTTGTACCCTGGCGG
GACTGCTATGTCACTGGGATGCGGATTTGCGCCGCTCTGCAGTGACCTTCTTTTAGAGATTGTGATCGTTACTTATACCAATAAA
CTTTTTCAGAGCTTTGTATTACGTCGTTTTTTTGTGACCGGGGACTCTATCAAGGGGCGTGGTTGCGCGATAATGGGGTATATTT
TGGCCTAATGGGTAACAGGCAGCAAGGTGTTCCGTTCACTAGTCTTCATCTCGCAGAGGAGAAACATGACAAGTGTTCAAGACCA
GTCTCCTCTGAACGAAAACATGCAATGCCAGAAGACTTTAGAATTGTCGGTTGGAATGCAGATTCCGAAAGAACTACCAGGAGTG
GCAATGTTTACATTCTACGATTACTTAGCCGACCTGGCACGCGGAGGACACGCGCAAAGTGACATTCTTCAACAGGAACCACTAC
ATCACCGCCTGGCTTACATCAGCCACCTGTTTTCGTGGCTCGAAAAAGAAGGTTTCGCGTATGGCATTTTGAACAAATTTTTTGG
TTCTGAGCGCGCTCTGAGCGCATCAGAAATTAGTAGTAATTTATCGCAACAACAGATCAATGAAATCTTATTTTTCATAGAATCT
GAAACGAAACAACAAGCGTCGTGTGACCTGTGGAAAGTTTTAAGACAATTTTTATTGACGGCTTCAACTTTGAAGTGGATGAAAA
ATAAACCGTGCTCAAAGCCGGAATGGTTTAAGGTACAGGAATTCAAAGGCGGACACCTTGGCTATGCCACACAGTCCATGCCTTT
GATTTTTGGCAATACTAACGAAAGCTGCGCCCGATCCCTTCTTTTGGGATATGTAACTGGAGAGGCATGGAAAACTTCAGAAGAT
CGAGAAGAGTTTTACAAATTTGACGACGGCCATCCACCAGAAGAAGCATTTACATGTGGACTTCTTTTAGATAAACGGAGTGGAA
TGTTGGGAGCATCAATGGATATGGCGATAGTAAAGCGCAGGAAACAGTGCGCTAGAAAAGTGGAGATATACGAAATTAAGTGCAG
AGCAAAGTATGTGTTCTCGGTCGAGAACCAGACGCATCCACTTTCTCAACTGTACGATAAAATGTTGCAACACCCATGTGAAAAT
TCTATTCGCGATTTCTTGCTGGGAATTTCCTCACCGGGGGTAGAATTTGTTGAAGAGAGCGGAATCCCTACAGCATCAGAAGCTC
TACTAACCTGCGACAAAACTTGGAAAACTGACAGGTGGAAGAAAAATTTACGAGAACGTGCTTGCCTAATGGAAAAGCGTCACTT
GTCCCTGAATAGAACCAACAACTCATCTGTGTTTTTGTTTGAAAGCCCCTGTCTGGAAACAAACACGATCAGACCTGTCCAGTGG
CCGGACGGTGAAAACAATATCGAGTTACCAATCTTCATTAACCCAAAACATCAGAATTTCAAACAGATTTTTGTACAGACATACG
TCTTGGCCGAATATTTTGAAACAATTCCTATATCTCCATTTTTAGTTACATTTATAGGTCGCAACAGAAAGACGGTAGAGCGGGG
ACGTGTATTTAAGCTCGAACACACATTAGACGGTATAGAAGAACCGGTGGAATTGAACTGCAAGCACGCGATACCGGTACTACTA
ATAATAACCCCCACAACCATCGACAGAAATCATTTCAGTGATTTGGACTCGCTTGGCAGAGAAGCGTTCGAGTTTTCTGTGAAGG
AAACATGGGCCAAGGTCTCTGTGGACACCTCGGAGAATGTTGCTGCCCCTGCTGTTCAAAGTCCTGCGGGGGTGACACAAGATCC
CGAAAATCTGGAGAGCGCTTGTCCCTATTAGACGACAGCTTTGACGAATTCACCATAACCGAAGACATGCCTATGATTCAAGGAG
GAAAAAAGAAAGAGAAGCGTAGCAAAAAAGAAGCCGAACCAATTGTAGTTGAGCCAGTGGCTGCTAGAAGGCATTGAAACATGTT
ATTTCAAATAAACAGAACTACAACCAAATACATGCGTCAACCTTTATTTGCGTTATGCGTAA

>UL11

CCGGAATGGTTTAAGGTACAGGAATTCAAAGGCGGACACCTTGGCTATGCCACACAGTCCATGCCTTTGATTTTTGGCAATACTA
ACGAAAGCTGCGCCCGATCCCTTCTTTTGGGATATGTAACTGGAGAGGCATGGAAAACTTCAGAAGATCGAGAAGAGTTTTACAA
ATTTGACGACGGCCATCCACCAGAAGAAGCATTTACATGTGGACTTCTTTTAGATAAACGGAGTGGAATGTTGGGAGCATCAATG
GATATGGCGATAGTAAAGCGCAGGAAACAGTGCGCTAGAAAAGTGGAGATATACGAAATTAAGTGCAGAGCAAAGTATGTGTTCT
CGGTCGAGAACCAGACGCATCCACTTTCTCAACTGTACGATAAAATGTTGCAACACCCATGTGAAAATTCTATTCGCGATTTCTT
GCTGGGAATTTCCTCACCGGGGGTAGAATTTGTTGAAGAGAGCGGAATCCCTACAGCATCAGAAGCTCTACTAACCTGCGACAAA
ACTTGGAAAACTGACAGGTGGAAGAAAAATTTACGAGAACGTGCTTGCCTAATGGAAAAGCGTCACTTGTCCCTGAATAGAACCA
ACAACTCATCTGTGTTTTTGTTTGAAAGCCCCTGTCTGGAAACAAACACGATCAGACCTGTCCAGTGGCCGGACGGTGAAAACAA
TATCGAGTTACCAATCTTCATTAACCCAAAACATCAGAATTTCAAACAGATTTTTGTACAGACATACGTCTTGGCCGAATATTTT
GAAACAATTCCTATATCTCCATTTTTAGTTACATTTATAGGTCGCAACAGAAAGACGGTAGAGCGGGGACGTGTATTTAAGCTCG
AACACACATTAGACGGTATAGAAGAACCGGTGGAATTGAACTGCAAGCACGCGATACCGGTACTACTAATAATAACCCCCACAAC
CATCGACAGAAATCATTTCAGTGATTTGGACTCGCTTGGCAGAGAAGCGTTCGAGTTTTCTGTGAAGGAAACATGGGCCAAGGTC
TCTGTGGACACCTCGGAGAATGTTGCTGCCCCTGCTGTTCAAAGTCCTGCGGGGGTGACACAAGATCCCGAAAATCTGGAGAGCG
CTTGTCCCTATTAGACGACAGCTTTGACGAATTCACCATAACCGAAGACATGCCTATGATTCAAGGAGGAAAAAAGAAAGAGAAG
CGTAGCAAAAAAGAAGCCGAACCAATTGTAGTTGAGCCAGTGGCTGCTAGAAGGCATTGAAACATGTTATTTCAAATAAACAGAA
CTACAACCAAATACATGCGTCAACCTTTATTTGCGTTATGCGTAA

>UL10

CAGGACTTAGAATTTCTGAACAAAAGAACGCAACGACGTTTGAAAATGATAGCGTAGAGGAAAAGACACAAATATTTTCTCCACG
CAATAAGCGCGCCTGCATTTCATGAAAAAAGGACCCTTCAAGTAGAGATGGATCCACCTGTGCAGTACTTTTGTTATCCTCGAGT
CCTTTGTCCTCATTTCCTTCTGAGGGCTTTAACAAACTGGCAGGTATTGCCGCGCCAAGCGCGTTAAGCATTGTGCATTGACGAT
TGGCAAACCCGCTCGATACGAAATCATTCACTATCACGTGTATATTTTCTGCCCCTCTCAAATCTGCGAGCATTTCAACAAGAGT
TGCATTTATCGTTGCGTCCATCGCTACGATATACTGACATTTAGATAACAATAATGAAAACAAAGCTTCCGTTTCACGTAGTTTT
CTAACTGTAGGCGAAAAGAATTGATTCATCAAAGACATTACTTCATCAACTACTAGAACATCATAGTGATCCAATAGTACACTAG
AGACGCGTGGCAGACTTTCCAGTTGTACCAATAGCCGTCGGTGTGAAATCTCATCTATTATGTATTTTTCGCAGTCTAGATACGT
AGCAAAACCCACCAAACCATCCCGTTGGAATCGTCGACAAAGACTGTTAGTAAAGCTTTTTCTGCATGACACGACTATAGCGCTT
CTATCCTGGTGGTCTAGAAATCCAGCCAACCATTCAATCAGTGCCGTAGTTTTCCCGGATCCCATCGGAGCGCGCGCAATCAACA
CCGGTCGCGAGCGTGAGGGGGGTGGGAACGTAACCGGATGGCCATGCTGCTGATCTAGGCTCAGCCTACTCATTGTCGGATTGGA
TAATAGGCGATCGCTGAGATCGCATCCGTACAGTCTTCGTGCGAGTCCGACGCTGGGCGACAACCACAGTTGTTCTCCAGGAGCA
AAATTATTTGGCGACATATCGTTCCTCCCGCTGAGCGACGTAGGCGCTAATTCACATAGCGCAGAAAATGAGCGGCTACTACTAT
AGTGGACAGGAATCGCGTTTAGAGAGGATTGCATGGCGCATGTGGCTGATTGAAGTAATCTCTTTTGTGGTTCTGGTGTTGATAT
CACTGGCCTCAGCTTTCTCCGCCCTTAGTCCGCAAGCAGGGTTTCCCTGTTTTTTCGGCAACGTCGCCGGTGGACACCTTAAGCC
AGCAACGCCTTTGATCAATCAACCTGGACTGCTAAGGGCGAACGAAGATGGACTTTTTTTTATTTCGTCGCCAGGAATGATATTT
ACCGTTTACGCTACTATGGCGGTGTGGATACTCTTGGCGATATACGTTACTCTTGGCGCAGTTAGAGCCAAGACTTGTAATCACG
ATCGTTCATACGGAGCGTCCGAATTGACTTCGGCCGTGATCAGCACACCTGCCTTAATGTCTATTATTATCACAGTATGGGCGTG
GCAGTTGTTTGTACTACTCCTTTCGTACAGACAGCTTACGCTATCAGCTGTCACATTTTGCATTATTTTTGTATCCGGCATCGTA
TTTTTGGGTACGTTCGCGTCTTCCGGAAAGACACCTGACCAGTACTCAGTTTTTAACTCACAGCTCAAAAGTGCATGCCCTGAAG
TTCATTGTGTTGTATCACCGTTTAAGGCAGTGGTATTAAACGTCTACGCCTCTATCATGGGAGTTTGGGGGCTTATAATGGCCTT

GTCCGGAGCACTCGTCATGACCGTGAATTTTGGAGTCAGTGTCCCTGCTTCTACACTTGGGGCAATGATTGCATTTCTTATCATA
GGTATCATATACTTGATTGTGACCGAGCTGGTTTTAGCGAAGTATGTGCATGTACTTATTGGTCCGCATTTCGGACTCATCATTG
CTCTTGGAATGCTTGGGACGGCATCCCTGCACTATGCTAATACGTTGGATAGCGGAACTCATCCCAACTGGAAGCCGGCTGCGGC
AGGAATTCTAGGAGTATTTGCCCTCCTTGTTTTAGTTTTGGCTATTCTGCGAGTCGTACGGGCATACAGATATCACCAGTACAGA
CAGACGGATTTCATCCGCAAGGTAACGACCGTTGCCACAGAAGCTTTAGATAAAACCGTGAAACGCAAAACGCGCCGCGAAAAAT
ACAAAAGCCGAAGAGACAAGCACGTTTATGAAACTGTACCCTATGACGACGAACTGTAATTTAGATTACGCATAACGCAAATAAA
GGTTGACGCATGTATTTGGTTGTAGTTCTGTTTATTTGAAATAACATGTT

>UL9

GCAGCCGGCTTCCAGTTGGGATGAGTTCCGCTATCCAACGTATTAGCATAGTGCAGGGATGCCGTCCCAAGCATTCCAAGAGCAA
TGATGAGTCCGAAATGCGGACCAATAAGTACATGCACATACTTCGCTAAAACCAGCTCGGTCACAATCAAGTATATGATACCTAT
GATAAGAAATGCAATCATTGCCCCAAGTGTAGAAGCAGGGACACTGACTCCAAAATTCACGGTCATGACGAGTGCTCCGGACAAG
GCCATTATAAGCCCCCAAACTCCCATGATAGAGGCGTAGACGTTTAATACCACTGCCTTAAACGGTGATACAACACAATGAACTT
CAGGGCATGCACTTTTGAGCTGTGAGTTAAAAACTGAGTACTGGTCAGGTGTCTTTCCGGAAGACGCGAACGTACCCAAAAATAC
GATGCCGGATACAAAAATAATGCAAAATGTGACAGCTGATAGCGTAAGCTGTCTGTACGAAAGGAGTAGTACAAACAACTGCCAC
GCCCATACTGTGATAATAATAGACATTAAGGCAGGTGTGCTGATCACGGCCGAAGTCAATTCGGACGCTCCGTATGAACGATCGT
GATTACAAGTCTTGGCTCTAACTGCGCCAAGAGTAACGTATATCGCCAAGAGTATCCACACCGCCATAGTAGCGTAAACGGTAAA
TATCATTCCTGGCGACGAAATAAAAAAAAGTCCATCTTCGTTCGCCCTTAGCAGTCCAGGTTGATTGATCAAAGGCGTTGCTGGC
TTAAGGTGTCCACCGGCGACGTTGCCGAAAAAACAGGGAAACCCTGCTTGCGGACTAAGGGCGGAGAAAGCTGAGGCCAGTGATA
TCAACACCAGAACCACAAAAGAGATTACTTCAATCAGCCACATGCGCCATGCAATCCTCTCTAAACGCGATTCCTGTCCACTATA
GTAGTAGCCGCTCATTTTCTGCGCTATGTGAATTAGCGCCTACGTCGCTCAGCGGGAGGAACGATATGTCGCCAAATAATTTTGC
TCCTGGAGAACAACTGTGGTTGTCGCCCAGCGTCGGACTCGCACGAAGACTGTACGGATGCGATCTCAGCGATCGCCTATTATCC
AATCCGACAATGAGTAGGCTGAGCCTAGATCAGCAGCATGGCCATCCGGTTACGTTCCCACCCCCCTCACGCTCGCGACCGGTGT
TGATTGCGCGCGCTCCGATGGGATCCGGGAAAACTACGGCACTGATTGAATGGTTGGCTGGATTTCTAGACCACCAGGATAGAAG
CGCTATAGTCGTGTCATGCAGAAAAAGCTTTACTAACAGTCTTTGTCGACGATTCCAACGGGATGGTTTGGTGGGTTTTGCTACG
TATCTAGACTGCGAAAAATACATAATAGATGAGATTTCACACCGACGGCTATTGGTACAACTGGAAAGTCTGCCACGCGTCTCTA
GTGTACTATTGGATCACTATGATGTTCTAGTAGTTGATGAAGTAATGTCTTTGATGAATCAATTCTTTTCGCCTACAGTTAGAAA
ACTACGTGAAACGGAAGCTTTGTTTTCATTATTGTTATCTAAATGTCAGTATATCGTAGCGATGGACGCAACGATAAATGCAACT
CTTGTTGAAATGCTCGCAGATTTGAGAGGGGCAGAAAATATACACGTGATAGTGAATGATTTCGTATCGAGCGGGTTTGCCAATC
GTCAATGCACAATGCTTAACGCGCTTGGCGCGGCAATACCTGCCAGTTTGTTAAAGCCCTCAGAAGGAAATGAGGACAAAGGACT
CGAGGATAACAAAAGTACTGCACAGGTGGATCCATCTCTACTTGAAGGGTCCTTTTTTCATGAAATGCAGGCGCGCTTATTGCGT
GGAGAAAATATTTGTGTCTTTTCCTCTACGCTATCATTTTCAAACGTCGTTGCGTTCTTTTGTTCAGAAATTCTAAGTCCTGGGA
CAGTTCTCTTACTGAACTCCAACTCTCCCCACGTAGATACGACTAACTGGGGGCGATTTCGAGCAGTCATCTATACTACAGTGGT
GACAGTTGGATTGAGTTTTGACTCTTGCCACTTCCACTCAATGTTTGCGTTCGTAAAGCCCACAATCCATGGACCTGACATGATG
GCTGTCTATCAAGCGATGGGTCGGGTAAGACGTCTACTACATGATAAGCTTTTTATATATTTAGATGTTTCTGGGGCATGGGGCG
CACCAATTTTCACTCCCATGATTTTAAACTCGGAGTTTAGCACTACCCCATGGCCAGTAGACATTACAGTTCCAGCAGACGCCAT
GTGTGTAAAATTTAAAAACCGTTGCAGACAGATCAACAATCATCGAGACGGGGTATTTACCCGGTTCAAAAATAAGCATTATGTA
GAAAGATGTACACTAACCAGTGCTAATGACAGTTTTAGTCTCCTACATACTCTTCTAGTCAATAACAAAATAAATGTGGAGATAA
TGAGCGTTGACTCCTCCGATCCGAAACTGGAAAATTTGGGTCAATTCATATCTGGTCTGCGTGCAGATAGTTATCGGAATAGGGC
TCCACTTAAGCATCTGTACTCCGCGTTGGAAAAGTATGAAAAGAGTCAAGATATATTTTGCCTTCCATCAAACCCAACTCCTGAG
GAAACAGCTTTGATGCTCATGGAGAGTGAAAATGTAAACCAGTTCATTGCTCGTTTTTTTGAACTTGACCGATCGCTAGAAAATT
TCGGGGAAGACTTGCAAAGCCTTTCTGCAACAGCTTGGTGACAAAAGCAAAACTATAGATGCACTAATTAATGCGGCTGTTGTGGA
AACAGGATGCACCTGCGATGAAGATGAATGGTTTCAGATCAATACAGAGGCATTGTCCAAAGATCGAGACAGTAAACTCGACTCT
TGGATATCATATTATCTCGATGGTCCATTTGTGTGTCTAAGCAACGGACGTCCGGCCGTGATTGAGGTTTGTAATGAAACCGAAG
TTAAGAATCGTCGAAACCTATTAAGACTATGCATAAACATTGCCAGGACGATAGGCTGGAAACCTTCAGCCGATCCAAACGAGAC
TGAAATTGAACTTGATGTTATCGTGAAGGCAGTAAATGACGCATTGAACGCAGGGATGGGGAATTGTATGCTAGAATATCTTCGG
TTAAACATAACAGAACCCTCATGGTTAACTGGACCTATAAGGAACCTCCAGTGTTATCTTGGAGAAAAAAAAAATGTCTCGCCGCG
AACGGAAGACTTCTGACCCACTAGAGATGGCAGCTGTAAAGATCTTGCGCACTCTATGGGCGGAGTTATTTGATGTACGGGTATT
TAAAAGCCAAAAGACATTTCCTGGTACGGCAAGGGTCAAGAACTTGCGAAAGGAGGAGCTCTGTGCTTTGCTAGACCGCATTCAC
GTTCCCTACGATCGCAAGGAAACGCACAAGCAACTATATGCTAAACTTATGTGTCATAGAGAACAGTTCAAGGGTTCTCGACTAA
GTTTCCGGACTGCTGCGTGGACCCGTTTTTTAAAAGGACGGTGCCTGGAGGACTTCCATTTGATGACTCCAATCTCTTCCCATCA
TGTATGATGAAAGCATACTTTATGGCTACTTATGGGGCATCTCGATATATACGATTGAGACTAAACAAAATGCGTCAGGAAATAT
TACAGACATGTGGGTATTGTATGAATTAGCTGTCCATGATTCTACTAAGGGTGCCCATGAAGTACTTTATGCATCCTTACAGTTT
GAAGAAAGCGACCTCAGTGTGGCCTCGTGGCCTCATCCAGCTCTGAATGCCAGATTACTGACGGACTTCGTTGGGGCTGTAAAGA
ATTTCCATCAGAGGGTGTACAAGCTGAATGCAAAAACAATTAACTTCGGATGGATTCTCTCATGTTCTACTTCTAGCCAGATTGC
TCTCCGCCTGGTGACAGGAAGACTACTTCATGAAATCAGGCGCGCCTTGACTCTGCCAGAATTTTATTCTCCTTCTATTTTTTAT
GTTTGCAAAGACAGTGGATTGATCACTAAGGTCTGCGAGGATAAATCCAAGCCACGTATGGCAACTGCATCTTACGCAGCATTGA
ATGCTTCGTCTAACTTCCAAGCAAATTACATTGAACGCAATCTGCAAGCCGGACATTGTTGTGAACTGAGAAATTTTGGATGGGC
TCGCATACAGGCAATCTCATCTGGAAAAATTAATCCAAGAAGCATGACCGCGGAATGGGTGTGGGCTGGTGGAAAGTGGATTGAT
GGAAGAGGAAATGAAGCATTTTCTGAAACAGGGACTGATCCCACTTCCGACGATGGCCTTGCTGTCCCCTTTGCCTTAACACCCG
CGAAAACAATTTATGGACCGACGTGGTTTGTTCCCAGACGCATGCTTGTAGCGTCATTGGTGCCCAAGGAGTTCGAATATGCCAT
CTATCTCAAGGATGGAAGTACGATAATCTCTTTAGTGTGCGCAATAATAAATCTGTATTGCAGATTTTACCAAGGAAACATTGTT
GCACAATCTACATTTCTCAAACCTATTATCCTTTTTCTATTTCCGACAAGTTCTTCCCGAATGGACGGCTCAGAGAAGGTTTCAT
CTATAAAAGACTCCTACATGTATACACCTGGATTCCCGAGTATAAATTTTGTTCCCATTACGACGCAGAACATGTCAAAGATGGG
AGCCTTGAATGCATGCAGAATAGTCTCTCTTGTGGAGGGGTTATGGCCAGCCTATAATATCAGAGCATTACTCATGCTGGATCGC
GGAGATCGAGCGAAACATGGGTCTCAGATCACAACACTACCCAGAGACCTCGAGAGCACTTTAGAAGTGTACCCAGCCGGTAAAA
TATCCACCATAGCCGATCTCCCGACTGTTATCAGTGGCAGGATTTTAAAAATGGACTTTTCTGCTTTCTTTCCATGTTTGTACAT
GGCATGTGGCGGAGGCAGCCAAGCGCTATGCCGCATTATCGAGGCAAGATTAAACAGGGAGCCGCAGTCAGAGAAGCTTAAAGCT
GCCTTGGTAGCTCTGGTGGGTGGGCTAAAATATACAGATCCATCGAAATATAAATTAGTAATAGCGCTCTGCAACAGCATTGCAT
TGGCTGTCGAAAATGCTGCCAATAGCCTCCAGTTTGGAATAGCTATTTATATGAAGGATGGATTCATAGGAGCTTTTGATAAACA
TTCCTCTACAAGCGCAGAGGAACTGAGAAGCAAATGCGAATGGGCCGCCATTGAAGAATTGCAGAAGATTCTTCTAGAATGCGGC
CAAGCAATTACAGGAATGCCGACTCTGAAACTAAGGCTAGAGGGAGAATTTACAGAGGGCCTGCTTCTTAATTGCAACAAGTATT
GGCTCCACAACAGGAATACCGGAAAGAGTTTCATATGTGGGATACCTGGGCTTCGAGAAGAGAATGGGCTGAGTGTGTTAACAGA
ACGTACAGCTTGTGAACTGCTCGCTGGCATTTACACTGCCAACACCGTAAGCACCGCTACCGAAACACTCACGCGAATTCTAGAT

GGATATGCCTTCTCTGCCTTCGAAGCTAGAGGTGACATAAACTTCTGGCAAGAGACGCTAAAAAGTAGCTTTACCCCATGCATTT
CAGATTCCGCAGCCATCAGATCTGCCAGTTTCATGACACGCCCAGATGAGCTGGAAGGAGAGACTCACTTTGTTTACCTTACTCC
GACAAATTTGTCCTCCCATGGCTCCGCCAACGCTGGCGGTAAAGTTATCTACCCAAGTTCTCTCGCAGAAGAGAATTTCTGCATT
AAAATATGTTATTCGGCTCACCTAATCCCCAAAATGTCAGCAATGGTGGAGCTTGTACAGAATATGGTGTGGTTAAAATTTTATC
ATAATTCTGTGGACAATGATGAGGAGACACGGGGTAAGTTATTGAAAACGTGTGACTACGATTACGAGAGAACATCTTTTTTATT
CTCATAATTAAACAGAACATTTCCAAAGACCATGATACTTGGTACTTCGTCATTTCAAAAAT

>UL8

CTCCGCGTTGGAAAAGTATGAAAAGAGTCAAGATATATTTTGCCTTCCATCAAACCCAACTCCTGAGGAAACAGCTTTGATGCTC
ATGGAGAGTGAAAATGTAAACCAGTTCATTGCTCGTTTTTTTGAACTTGACCGATCGCTAGAAAATTTCGGGGAAGACTTGCAAG
CCTTTCTGCAACAGCTTGGTGACAAAAGCAAAACTATAGATGCACTAATTAATGCGGCTGTTGTGGAAACAGGATGCACCTGCGA
TGAAGATGAATGGTTTCAGATCAATACAGAGGCATTGTCCAAAGATCGAGACAGTAAACTCGACTCTTGGATATCATATTATCTC
GATGGTCCATTTGTGTGTCTAAGCAACGGACGTCCGGCCGTGATTGAGGTTTGTAATGAAACCGAAGTTAAGAATCGTCGAAACC
TATTAAGACTATGCATAAACATTGCCAGGACGATAGGCTGGAAACCTTCAGCCGATCCAAACGAGACTGAAATTGAACTTGATGT
TATCGTGAAGGCAGTAAATGACGCATTGAACGCAGGGATGGGGAATTGTATGCTAGAATATCTTCGGTTAAACATAACAGAACCC
TCATGGTTAACTGGACCTATAAGGAACCTCCAGTGTTATCTTGGAGAAAAAAAAATGTCTCGCCGCGAACGGAAGACTTCTGACC
CACTAGAGATGGCAGCTGTAAAGATCTTGCGCACTCTATGGGCGGAGTTATTTGATGTACGGGTATTTAAAAGCCAAAAGACATT
TCCTGGTACGGCAAGGGTCAAGAACTTGCGAAAGGAGGAGCTCTGTGCTTTGCTAGACCGCATTCACGTTCCCTACGATCGCAAG
GAAACGCACAAGCAACTATATGCTAAACTTATGTGTCATAGAGAACAGTTCAAGGGTTCTCGACTAAGTTTCCGGACTGCTGCGT
GGACCCGTTTTTTAAAAGGACGGTGCCTGGAGGACTTCCATTTGATGACTCCAATCTCTTCCCATCATGTATGATGAAAGCATAC
TTTATGGCTACTTATGGGGCATCTCGATATATACGATTGAGACTAAACAAAATGCGTCAGGAAATATTACAGACATGTGGGTATT
GTATGAATTAGCTGTCCATGATTCTACTAAGGGTGCCCATGAAGTACTTTATGCATCCTTACAGTTTGAAGAAAGCGACCTCAGT
GTGGCCTCGTGGCCTCATCCAGCTCTGAATGCCAGATTACTGACGGACTTCGTTGGGGCTGTAAAGAATTTCCATCAGAGGGTGT
ACAAGCTGAATGCAAAAACAATTAACTTCGGATGGATTCTCTCATGTTCTACTTCTAGCCAGATTGCTCTCCGCCTGGTGACAGG
AAGACTACTTCATGAAATCAGGCGCGCCTTGACTCTGCCAGAATTTTATTCTCCTTCTATTTTTTATGTTTGCAAAGACAGTGGA
TTGATCACTAAGGTCTGCGAGGATAAATCCAAGCCACGTATGGCAACTGCATCTTACGCAGCATTGAATGCTTCGTCTAACTTCC
AAGCAAATTACATTGAACGCAATCTGCAAGCCGGACATTGTTGTGAACTGAGAAATTTTGGATGGGCTCGCATACAGGCAATCTC
ATCTGGAAAAATTAATCCAAGAAGCATGACCGCGGAATGGGTGTGGGCTGGTGGAAAGTGGATTGATGGAAGAGGAAATGAAGCA
TTTTCTGAAACAGGGACTGATCCCACTTCCGACGATGGCCTTGCTGTCCCCTTTGCCTTAACACCCGCGAAAACAATTTATGGAC
CGACGTGGTTTGTTCCCAGACGCATGCTTGTAGCGTCATTGGTGCCCAAGGAGTTCGAATATGCCATCTATCTCAAGGATGGAAG
TACGATAATCTCTTTAGTGTGCGCAATAATAAATCTGTATTGCAGATTTTACCAAGGAAACATTGTTGCACAATCTACATTTCTC
AAACCTATTATCCTTTTTCTATTTCCGACAAGTTCTTCCCGAATGGACGGCTCAGAGAAGGTTTCATCTATAAAAGACTCCTACA
TGTATACACCTGGATTCCCGAGTATAAATTTTGTTCCCATTACGACGCAGAACATGTCAAAGATGGGAGCCTTGAATGCATGCAG
AATAGTCTCTCTTGTGGAGGGGTTATGGCCAGCCTATAATATCAGAGCATTACTCATGCTGGATCGCGGAGATCGAGCGAAACAT
GGGTCTCAGATCACAACACTACCCAGAGACCTCGAGAGCACTTTAGAAGTGTACCCAGCCGGTAAAATATCCACCATAGCCGATC
TCCCGACTGTTATCAGTGGCAGGATTTTAAAAATGGACTTTTCTGCTTTCTTTCCATGTTTGTACATGGCATGTGGCGGAGGCAG
CCAAGCGCTATGCCGCATTATCGAGGCAAGATTAAACAGGGAGCCGCAGTCAGAGAAGCTTAAAGCTGCCTTGGTAGCTCTGGTG
GGTGGGCTAAAATATACAGATCCATCGAAATATAAATTAGTAATAGCGCTCTGCAACAGCATTGCATTGGCTGTCGAAAATGCTG
CCAATAGCCTCCAGTTTGGAATAGCTATTTATATGAAGGATGGATTCATAGGAGCTTTTGATAAACATTCCTCTACAAGCGCAGA
GGAACTGAGAAGCAAATGCGAATGGGCCGCCATTGAAGAATTGCAGAAGATTCTTCTAGAATGCGGCCAAGCAATTACAGGAATG
CCGACTCTGAAACTAAGGCTAGAGGGAGAATTTACAGAGGGCCTGCTTCTTAATTGCAACAAGTATTGGCTCCACAACAGGAATA
CCGGAAAGAGTTTCATATGTGGGATACCTGGGCTTCGAGAAGAGAATGGGCTGAGTGTGTTAACAGAACGTACAGCTTGTGAACT
GCTCGCTGGCATTTACACTGCCAACACCGTAAGCACCGCTACCGAAACACTCACGCGAATTCTAGATGGATATGCCTTCTCTGCC
TTCGAAGCTAGAGGTGACATAAACTTCTGGCAAGAGACGCTAAAAAGTAGCTTTACCCCATGCATTTCAGATTCCGCAGCCATCA
GATCTGCCAGTTTCATGACACGCCCAGATGAGCTGGAAGGAGAGACTCACTTTGTTTACCTTACTCCGACAAATTTGTCCTCCCA
TGGCTCCGCCAACGCTGGCGGTAAAGTTATCTACCCAAGTTCTCTCGCAGAAGAGAATTTCTGCATTAAAATATGTTATTCGGCT
CACCTAATCCCCAAAATGTCAGCAATGGTGGAGCTTGTACAGAATATGGTGTGGTTAAAATTTTATCATAATTCTGTGGACAATG
ATGAGGAGACACGGGGTAAGTTATTGAAAACGTGTGACTACGATTACGAGAGAACATCTTTTTTTATTCTCATAATTAAACAGAAC
ATTTCCAAAGACCATGATACTTGGTACTTCGTCATTTCAAAAAT

>UL7

ATTGTGTCAATTAGTAAATACCGCTCCGATCAAGGTTTTGCTTGGTCTACGAACTGAGAATGCACAGGATCCAGATCATGTAACA
AAAGTGGTGGATAAGGCTCTGGGAGAAGGTGTGGAAGAGGCCGGTACATCTGCAGCAGCTAGGCTGATCAAGTTCATTATTGATA
TTCAAAATATGCGCAAAGTTGGTGATGTGGCAGACGTAGTGGACTCTTACTTGCGTGAAAATGTTTCTAACATCGAGAGACAACT
CGTGCTTGACCCTACCCGTGTTGGGTGTGGCGGAAGAGCGGCAGCGCAGGGGCAGCGGCCCATTGATAATGCAGAAGATCGCATG
GGGAATGCTGTGCGAGCTGCCGTTAATACAAGTGTTGGAAAGGTAGTAAATACTTTGTTCAAGACTGTCTCTGATCTCAAGGAAA
GCAATAAAAACTTGGTCTATAGTAATGGGAAAATGAAAAAAGAATTGGTTGAGGCTAGAATGGAATCTGCCAGGCGGACTGAGCG
CACCCAAGCAAACTATGACCGTTGCCCAGCATGGGAGACCCTTGGAGGTATCGAAGAGGCTTGGGAGTGCAAAGAAACTCTAGAG
TCGCTATCGAAGCTCCCTGAGAATTTATCATATAACCTCGTTGATGTAAGTAAGGACATGAATGACCAACAGTATATTGCTAATA
GTTTTCTGGCGCGGTACATTCCTCCTTATCTGGAGGAGGAGACTCGATTGTCGGCGCTCTGGGAACAAGAATTACTGCGCGTGTT
TAAAATTCAAAGAAATATGAATAATCAAGGCGAGGAAGTATCTTTGTCGTACAGTACTAGCGCAATCACGCTTATTGTCGGTCCA
TTTTTTCACAAAGTTCTGAAGGCGTCACGTCTAGGATTCCTTGTATCGAGCCAGACTACATATAAGTCAGAGGAGGAGTTGTGTC
AAACATTATTTGCCAGCTCGCGGGTAGGTTCGTATTTGAAAGATCTGAACGTACGTTACATAGCAAATGTCAGACGTTCGTGCGC
CAATCGGCAAGATGAACGATGGGATGGAAGAGCTGAAAGAGACAGGAGAGCCTATTCTAGAGGCAGCGACAGAGAATCGGCCGAC
TCGAAAGTTAGTATCATGGGCCGAAGAAGCATTCAACGCGATATGCCGGCATCAGAAAAAAGTAGGACCTATACGATTGGAAGAC
ATCGCAGAAGAACCCGAACACGGCCCTACCCTTGAGGAACTTGAACGTCTGGCCTCCTCTGGAGAATCGACATCACAGTTAATTA
AGGAAATATCGCGGCAGGCTATCCCGAGATTTTTTTTTGAAGTACGATCCACGCCAGGGGCGCCAGCATTGTTTGCGTCTGATTC
CATAACATATATGGACATCCAGATGGATGGGGAAAAAATTATCCTAGAGCTGATTTCTGGAGCAGTAGTGGATATTGAAGAGTAC
TTTTGCGCGTGTGTGCGCCTAGATGATCATGCTGGCTTTATATTTTCAGTAATCACAGCCTCAGAGGACCGCGTTGCCACTGTGC
ATGTTACATGGCCTGATTTTTTAAACAGGATTCACAGAATTATGCCCCAAAAATTTGAAGACTTTGCGCTCTGTTTGATTTGCAT
GTATTTGGAAAACATTGACCCCATTAAGGTTACTCAAGAACACTGCTATAAGTTACGAAGCGCAATTGCCCGTGTCGCGCGAACT
ATTTCAATCAATGATAAGGTGCGCGAGCTATTGGTGAATGGAGGGGAGTGGCTCTTAAATACATTGCTATTAATGCATGATCTTC
CTCCCTTTGATCCCGGGTTGGTATTGCCTCATTATCGAGTGGCGCAGCCCCTCATGCAAATTACCGCTCCAATTGGAGAAGAGCT
AATATCTTCTCTTTATTCTACCCACGCACTTATGATCATGAACCTGCCTTCCAATGAAAATGAACGAGCCATCATCTTAGAGTCA

AGGAATTCAACTTTGAATCGGGTATATGGCAACGGAAACTTTTTTACTCGCGTATTAAGATTGTGGTGGAATGACCGCAAGCTTC
ATCGCAGTGCATATAAGTTATATGATACCTATTTTTGAAATGACGAAGTACCAAGTATCATGGTCTTTGGAAATGTTCTGTTTAA
TTATGAGAATAAAAAAGATGTTCTCTCGTAATCGTAGTCACACGTTTTCAATAACTTACCCCGTG

>UL6

AACCCAAATTCAAATGCCTTTAATACTTCTCCGAAAGCAGGCTCCGAGCATCGTTTATTATTAATAAAGATGGCCCACTGCTTCT
TGATATTCAGGACTGAAAACAGTGTAGGAGTACAAATAAGATTACTTAAAATGTTTATGCTGCTGGATATAAGGTGTCGTTGATT
TCTGTGCTCGAAGCTGCTTTCCATTGCGTCTGTCTGTGTAGGGGAGCCAATGCACACGATCACTGGCTTCTTCCCGTCCTGATAC
ATAGGCGTTTTCCACAATGCATTCATGAGCCACCACGAATATACTATTGCAGTCAATATATGTTTCCCTAAAACTCCAGCCTCAT
CAACAAGGATAATGTTGCTTTTGACAAATGGAGGCATAGAGGAAATTAGGAAAGGTGCAAGGTTTACAAATTTGCGTGAAGTTTT
CTGCTGGAGTGTTTGAAGGACAGACAAAGCTACCGGAGATGCCGTGTCTATGGCGCGTGCAGTAATGTCCTTTATCACGTCCCAA
TAATAATAAATGTCGGCCATTTGATGTTCTGCCAACGAACGTTGTTCGTGAGGTTTTTCAAACTTGAATCGTCCTAGCACAGCCT
GTACGTTGTTTCCTTTGAAGCCAAAGTTTTGAAAAATAGTATGAATGGGACAAGAGGTGTAAGAGGCAGATAGCTTATTGAAGAT
ATTAAGAGCAGCTATGCGCGTTGAGCCAGTAACGATGCAATTCAATGTTTCATTAAGAGTTTGAATGCAAGTACTTTTTCCTGAG
CCGGCGTTACCGGTGATTAGATAAACATTAAATGGTAATTCCGCCAGAGGCAAAGTGGTCGGTTCATCTAAACGTGCCACAGTCT
CAAACCAAGACAGTTGCGGCTTGGAGTCTTCATGAACAGCTTGTTCTGATAGGATTGTAATGTCCGATAAAATCGCCTGAATGCT
CTGCATTGCCGAAAAATTTAAGTAAACAGGAGTGGTGATTTCTATTTCCCGCCTGCTCTTCCCCGAAAATGGACATGCCATTTTC
CCAATTGCGTCAGGTACCTTGTGCCTTTGATATCGACAACCTTATCGCGGTCTTGGAGCCGATTCTAACAGGCGAGCCTGATGAC
TGGATAATAATCCATCCAACAAAGCGCACGATGGGATTCAGGGAAATTCTCCTGGGCGCATCGACGTACACTTCTGGACAAGGCG
TGTATAACTCGGTCCGAGCAACAGAATCTGTTGTTAGACAAATTCAAGTGGCGGTTCTCCGCAATATATTGGATGGCTGTCGATT
TGAGGATGTGGAGCGCGATTGGGACACTCACATCAGAACAAGAGGGTTGACACCGGATGAACTGGCAGCACGTTACTGTCACAAT
TCGGAAGATGGAATTCGAGTTGCAGAGCAGGCGTTCGAGTTGTGGTCCTTTTTATTGCAGACAGTGTTACTTGATTTTGTACGGG
GCGTGTCTGAGTTAAGCTTGCGCGGGTTAAGCGAAGCTGGGAAGGCTCCTTCCTACGTACGTTATATAGATTGGATAACATGTCT
GGGTATGGTTCCTTTGGTACGAGAAAGAAAACCTCAAGAAAAGCGCCCCACTGGCAGTGATCTCGGAAATGAAGACCTGGAAAGT
ACAACGACTGAAATACACCACCACCTACATGTCGCGCCTTCACTTTTTGCCAGAGGGTTCGAAATGCTGAAGTATCTCCGTGCTG
TATTAGGGGCAGTACATATTTTAGAGTATGAGCGAACACTAATAATAATGAACGTTAAAACCAGTGAAATCCGCGCCTATGATGG
AAATACTGGGGAGCTTGGGGAATGTATGGTTGTTTGGGAACCATTGGTAACTGACAGGGGAGTATTATTTGATTCTCCTATGCAA
CGATTGCACGGTCCTGTTCTCAAAGCTCAGGCTATGCGCGAGCATGCAAAATTGTGTCAATTAGTAAATACCGCTCCGATCAAGG
TTTTGCTTGGTCTACGAACTGAGAATGCACAGGATCCAGATCATGTAACAAAAGTGGTGGATAAGGCTCTGGGAGAAGGTGTGGA
AGAGGCCGGTACATCTGCAGCAGCTAGGCTGATCAAGTTCATTATTGATATTCAAAATATGCGCAAAGTTGGTGATGTGGCAGAC
GTAGTGGACTCTTACTTGCGTGAAAATGTTTCTAACATCGAGAGACAACTCGTGCTTGACCCTACCCGTGTTGGGTGTGGCGGAA
GAGCGGCAGCGCAGGGGCAGCGGCCCATTGATAATGCAGAAGATCGCATGGGGAATGCTGTGCGAGCTGCCGTTAATACAAGTGT
TGGAAAGGTAGTAAATACTTTGTTCAAGACTGTCTCTGATCTCAAGGAAAGCAATAAAAACTTGGTCTATAGTAATGGGAAAATG
AAAAAAGAATTGGTTGAGGCTAGAATGGAATCTGCCAGGCGGACTGAGCGCACCCAAGCAAACTATGACCGTTGCCCAGCATGGG
AGACCCTTGGAGGTATCGAAGAGGCTTGGGAGTGCAAAGAAACTCTAGAGTCGCTATCGAAGCTCCCTGAGAATTTATCATATAA
CCTCGTTGATGTAAGTAAGGACATGAATGACCAACAGTATATTGCTAATAGTTTTCTGGCGCGGTACATTCCTCCTTATCTGGAG
GAGGAGACTCGATTGTCGGCGCTCTGGGAACAAGAATTACTGCGCGTGTTTAAAATTCAAAGAAATATGAATAATCAAGGCGAGG
AAGTATCTTTGTCGTACAGTACTAGCGCAATCACGCTTATTGTCGGTCCATTTTTTCACAAAGTTCTGAAGGCGTCACGTCTAGG
ATTCCTTGTATCGAGCCAGACTACATATAAGTCAGAGGAGGAGTTGTGTCAAACATTATTTGCCAGCTCGCGGGTAGGTTCGTAT
TTGAAAGATCTGAACGTACGTTACATAGCAAATGTCAGACGTTCGTGCGCCAATCGGCAAGATGAACGATGGGATGGAAGAGCTG
AAAGAGACAGGAGAGCCTATTCTAGAGGCAGCGACAGAGAATCGGCCGACTCGAAAGTTAGTATCATGGGCCGAAGAAGCATTCA
ACGCGATATGCCGGCATCAGAAAAAAGTAGGACCTATACGATTGGAAGACATCGCAGAAGAACCCGAACACGGCCCTACCCTTGA
GGAACTTGAACGTCTGGCCTCCTCTGGAGAATCGACATCACAGTTAATTAAGGAAATATCGCGGCAGGCTATCCCGAGATTTTTT
TTTGAAGTACGATCCACGCCAGGGGCGCCAGCATTGTTTGCGTCTGATTCCATAACATATATGGACATCCAGATGGATGGGGAAA
AAATTATCCTAGAGCTGATTTCTGGAGCAGTAGTGGATATTGAAGAGTACTTTTGCGCGTGTGTGCGCCTAGATGATCATGCTGG
CTTTATATTTTCAGTAATCACAGCCTCAGAGGACCGCGTTGCCACTGTGCATGTTACATGGCCTGATTTTTTAAACAGGATTCAC
AGAATTATGCCCCAAAAATTTGAAGACTTTGCGCTCTGTTTGATTTGCATGTATTTGGAAAACATTGACCCCATTAAGGTTACTC
AAGAACACTGCTATAAGTTACGAAGCGCAATTGCCCGTGTCGCGCGAACTATTTCAATCAATGATAAGGTGCGCGAGCTATTGGT
GAATGGAGGGGAGTGGCTCTTAAATACATTGCTATTAATGCATGATCTTCCTCCCTTTGATCCCGGGTTGGTATTGCCTCATTAT
CGAGTGGCGCAGCCCCTCATGCAAATTACCGCTCCAATTGGAGAAGAGCTAATATCTTCTCTTTATTCTACCCACGCACTTATGA
TCATGAACCTGCCTTCCAATGAAAATGAACGAGCCATCATCTTAGAGTCAAGGAATTCAACTTTGAATCGGGTATATGGCAACGG
AAACTTTTTTACTCGCGTATTAAGATTGTGGTGGAATGACCGCAAGCTTCATCGCAGTGCATATAAGTTATATGATACCTATTTT
TGAAATGACGAAGTACCAAGTATCATGGTCTTTGGAAATGTTCTGTTTAATTATGAGAATAAAAAAGATGTTCTCTCGTAATCGT
AGTCACACGTTTTCAATAACTTACCCCGTG

>UL5

CTTATCCACCACTTTTGTTACATGATCTGGATCCTGTGCATTCTCAGTTCGTAGACCAAGCAAAACCTTGATCGGAGCGGTATTT
ACTAATTGACACAATTTTGCATGCTCGCGCATAGCCTGAGCTTTGAGAACAGGACCGTGCAATCGTTGCATAGGAGAATCAAATA
ATACTCCCCTGTCAGTTACCAATGGTTCCCAAACAACCATACATTCCCCAAGCTCCCCAGTATTTCCATCATAGGCGCGGATTTC
ACTGGTTTTAACGTTCATTATTATTAGTGTTCGCTCATACTCTAAAATATGTACTGCCCCTAATACAGCACGGAGATACTTCAGC
ATTTCGAACCCTCTGGCAAAAAGTGAAGGCGCGACATGTAGGTGGTGGTGTATTTCAGTCGTTGTACTTTCCAGGTCTTCATTTC
CGAGATCACTGCCAGTGGGGCGCTTTTCTTGAGGTTTTCTTTCTCGTACCAAAGGAACCATACCCAGACATGTTATCCAATCTAT
ATAACGTACGTAGGAAGGAGCCTTCCCAGCTTCGCTTAACCCGCGCAAGCTTAACTCAGACACGCCCCGTACAAAATCAAGTAAC
ACTGTCTGCAATAAAAAGGACCACAACTCGAACGCCTGCTCTGCAACTCGAATTCCATCTTCCGAATTGTGACAGTAACGTGCTG
CCAGTTCATCCGGTGTCAACCCTCTTGTTCTGATGTGAGTGTCCCAATCGCGCTCCACATCCTCAAATCGACAGCCATCCAATAT
ATTGCGGAGAACCGCCACTTGAATTTGTCTAACAACAGATTCTGTTGCTCGGACCGAGTTATACACGCCTTGTCCAGAAGTGTAC
GTCGATGCGCCCAGGAGAATTTCCCTGAATCCCATCGTGCGCTTTGTTGGATGGATTATTATCCAGTCATCAGGCTCGCCTGTTA
GAATCGGCTCCAAGACCGCGATAAGGTTGTCGATATCAAAGGCACAAGGTACCTGACGCAATTGGGAAAATGGCATGTCCATTTT
CGGGGAAGAGCAGGCGGGAAATAGAAATCACCACTCCTGTTTACTTAAATTTTTCGGCAATGCAGAGCATTCAGGCGATTTTATC
GGACATTACAATCCTATCAGAACAAGCTGTTCATGAAGACTCCAAGCCGCAACTGTCTTGGTTTGAGACTGTGGCACGTTTAGAT
GAACCGACCACTTTGCCTCTGGCGGAATTACCATTTAATGTTTATCTAATCACCGGTAACGCCGGCTCAGGAAAAAGTACTTGCA
TTCAAACTCTTAATGAAACATTGAATTGCATCGTTACTGGCTCAACGCGCATAGCTGCTCTTAATATCTTCAATAAGCTATCTGC
CTCTTACACCTCTTGTCCCATTCATACTATTTTTCAAAACTTTGGCTTCAAAGGAAACAACGTACAGGCTGTGCTAGGACGATTC
AAGTTTGAAAAACCTCACGAACAACGTTCGTTGGCAGAACATCAAATGGCCGACATTTATTATTATTGGGACGTGATAAAGGACA

TTACTGCACGCGCCATAGACACGGCATCTCCGGTAGCTTTGTCTGTCCTTCAAACACTCCAGCAGAAAACTTCACGCAAATTTGT
AAACCTTGCACCTTTCCTAATTTCCTCTATGCCTCCATTTGTCAAAAGCAACATTATCCTTGTTGATGAGGCTGGAGTTTTAGGG
AAACATATATTGACTGCAATAGTATATTCGTGGTGGCTCATGAATGCATTGTGGAAAACGCCTATGTATCAGGACGGGAAGAAGC
CAGTGATCGTGTGCATTGGCTCCCCTACACAGACAGACGCAATGGAAAGCAGCTTCGAGCACAGAAATCAACGACACCTTATATC
CAGCAGCATAAACATTTTAAGTAATCTTATTTGTACTCCTACACTGTTTTCAGTCCTGAATATCAAGAAGCAGTGGGCCATCTTT
ATTAATAATAAACGATGCTCGGAGCCTGCTTTCGGAGAAGTATTAAAGGCATTTGAATTTGGGTTACCTTTGACAGAACGCCATG
CCAGATTCTTAGATCAGTTCATTGTATCAGAATCATTCATCAAAGATCCATCAAAGCTTCCTGGATGGACTAGGCTTTTCTCGTC
CCACGAAGAAGTTAAAGAATACGTTTCTAAATTGCACGCAAAGCTTCGTGCACAGAAATCTGAAAAATATAGGGTATTTCTCCTG
CCGATGTATACAATAGTAGATATGGCGGCATTTGAAAAGTATAAATCCCTGACGGGTCAAGAGACACTAAATATTGATCGCTGGC
TGCAAAACAATTCCTCTCGCCTTGGAAACTATTCACAGAGCAGGGATCTAGATGTGACCGCGCCCAGATTTGAATACCATCGAGA
TGAGAGTAATGAATATACACTAGTAACGACTGATGCCTCGCACGTACTCAATAGCCAGATCACGGTAACAAAAAAGGTAAAAAAA
TTGATCTTCGGTTTTGAGGGAACATTTGAAAAATTCGCACATGTATTATCCGAAGATAGCTTTCTCAAGACGTATGGAGAAGACA
AAGTAGAATTTGCCTATCACTTTCTGTCGACCTTGCTATACTATGGGATGATTAAATTTTACGAATTTCTCCGAACAGAGGGCCT
CCCCGAAGATAAATTAACAATGGCGTATGACAGGTTGTACTCCCTAGCAATGCCGGAGCCAGAAACTCAGAACTTTGACATGGGA
AATGGTGAAAAATTTTGTTTTAACGAAGATGAATGTCCGAATGCCGATGTGAATGACAAAGATGATTTATTCGATATTTTTGATA
AGTCGTTAGATCAATTCTATCTAAATTATGAGATTTGCGGGAGCGATGTCCACGGTCAAGAAATATTTAGTTATTTCGAGCAGAT
GAAAAGAATTTATACTTTACGTTACGCGGTGCTGTGTGAGCTCTTTGGGAGTGTATTTACTGCCGCCCCCTTTAGCTCATTCGTC
GGAACGGCTTCTTTCTCAAGTCAAGAAATTTCTATCTCCAGTTTCAAGGGAGCAGTATGTGCATTTGCTGCTCAAACCGACACAT
ATACCTTACGCGGGATTACGCGAGCAAGATTTCCGGGATATGCTGAGGACACCTCAAAAGCCCACGAATGGGCCGAGCCAATTTT
ACAAATGCTAGACTTGCCAAGACTTGTCGTAAGGGATCAAATGGGGTTCGTTTCAGTCCTGTGCCATAACAAAGCAACTTTGTC
GATAATATAGGTGGACAGGAATTAAGGATGGCCATAACAATTGACCACGGAATAAGTTCGAGCCTTGCAATGACGATCACACGAT
CTCAAGGGTTAAGTTTAGACAGGGTAGCAATATGTTTTTCTCACGGAACAATGAAATTAAATACTGCATACGTAGCAATGTCTAG
AGTAACGAGTAGCGAATATTTGCGCATGAACTTAAACCCCCTGCGAACGAAATACGAAGATACTCGACAAGTAAGCCAACACATT
CTTCGCGCGCTCAGATGCAAAGAAACGCGACTAGTATACTAATCCTCAAAAGTAGCCACACGCGTGGCTTTAAAATCAGTTTCCC
ACCGTCGTTTGGAGTGTTGCATACCCTCCTCGTCCCGAGAACATGATTTTCATAAGCTACGTAATGGTAGGAGTCCAAGGCTTGG
GGCATGGCTCGGCCACCGAATACGAACAAGTAGTATACTCATGCGATGGAGGAATGCGGTTTATTTGCATCGGAAATAAAATGTA
TCGCCACCAATTACCACCCGGGAAAGTAATAGTAATTCACAATCCCGTCGCAACAATGATCGCTGTAGACTGTGAAGAAGAATTT
TGCGCATACTGTTTGGAGCGCAATGGTTCTCACCGAGGACCCTCAGGAGAAACATTAGCTTTTCAGTTCTCGGCATGTTGGTTTC
TAGGGCGCGGAGGAACGCGAGAACGGTGGAGCAGCGGCAACATCACCATGATGAATTTTTTGGGCGTCGCTCACTTAACTGTAAC
CATATACGAAACCCCGGAAGATCTAGCATCTAGCGCGCACACCACCCCCACATGTTCCCAATCCTCTCCAGAATCGCTGGAGTGT
CCAGAGGAAAGCGTTCCTCGAGATCTTGTAGAGTTTGCGGCAAGGCATGCCGGACTGTTAGAGGAATAAATAAAAGTAATAAACA
CCAAATTATGAAATACGAGCATTTAATTATTGCGTCGCATTGCGGAAA

>UL4

ACGAATTTCTCCGAACAGAGGGCCTCCCCGAAGATAAATTAACAATGGCGTATGACAGGTTGTACTCCCTAGCAATGCCGGAGCC
AGAAACTCAGAACTTTGACATGGGAAATGGTGAAAAATTTTGTTTTAACGAAGATGAATGTCCGAATGCCGATGTGAATGACAAA
GATGATTTATTCGATATTTTTGATAAGTCGTTAGATCAATTCTATCTAAATTATGAGATTTGCGGGAGCGATGTCCACGGTCAAG
AAATATTTAGTTATTTCGAGCAGATGAAAAGAATTTATACTTTACGTTACGCGGTGCTGTGTGAGCTCTTTGGGAGTGTATTTAC
TGCCGCCCCCTTTAGCTCATTCGTCGGAACGGCTTCTTTCTCAAGTCAAGAAATTTCTATCTCCAGTTTCAAGGGAGCAGTATGT
GCATTTGCTGCTCAAACCGACACATATACCTTACGCGGGATTACGCGAGCAAGATTTCCGGGATATGCTGAGGACACCTCAAAAG
CCCACGAATGGGCCGAGCCAATTTTACAAATGCTAGACTTGCCAAGACTTGTCGTAAGGGATCAAATGGGGTTCGTTTCAGTCCT
GTGCCATAACAAAGCAACTTTGTCGATAAATATAGGTGGACAGGAATTAAGGATGGCCATAACAATTGACCACGGAATAAGTTCG
AGCCTTGCAATGACGATCACACGATCTCAAGGGTTAAGTTTAGACAGGGTAGCAATATGTTTTTCTCACGGAACAATGAAATTAA
ATACTGCATACGTAGCAATGTCTAGAGTAACGAGTAGCGAATATTTGCGCATGAACTTAAACCCCCTGCGAACGAAATACGAAGA
TACTCGACAAGTAAGCCAACACATTCTTCGCGCGCTCAGATGCAAAGAAACGCGACTAGTATACTAATCCTCAAAAGTAGCCACA
CGCGTGGCTTTAAAATCAGTTTCCCACCGTCGTTTGGAGTGTTGCATACCCTCCTCGTCCCGAGAACATGATTTTCATAAGCTAC
GTAATGGTAGGAGTCCAAGGCTTGGGGCATGGCTCGGCCACCGAATACGAACAAGTAGTATACTCATGCGATGGAGGAATGCGGT
TTATTTGCATCGGAAATAAAATGTATCGCCACCAATTACCACCCGGGAAAGTAATAGTAATTCACAATCCCGTCGCAACAATGAT
CGCTGTAGACTGTGAAGAAGAATTTTGCGCATACTGTTTGGAGCGCAATGGTTCTCACCGAGGACCCTCAGGAGAAACATTAGCT
TTTCAGTTCTCGGCATGTTGGTTTCTAGGGCGCGGAGGAACGCGAGAACGGTGGAGCAGCGGCAACATCACCATGATGAATTTTT
TGGGCGTCGCTCACTTAACTGTAACCATATACGAAACCCCGGAAGATCTAGCATCTAGCGCGCACACCACCCCCACATGTTCCCA
ATCCTCTCCAGAATCGCTGGAGTGTCCAGAGGAAAGCGTTCCTCGAGATCTTGTAGAGTTTGCGGCAAGGCATGCCGGACTGTTA
GAGGAATAAATAAAAGTAATAAACACCAAATTATGAAATACGAGCATTTAATTATTGCGTCGCATTGCGGAAA

>UL3.5

ACGGTGTTGATGCTTTGGGGCCAAGAAGCTAGACGATTTATTCCCCGCGTGCCTAAAAATCATTTAAAGTTGGAGTTCATCCATC
CTTCTACCTGCACGCGACGTCCGTTCAACTGTCGCCATTTCGAGGAAGCAAATCAATACTTCGAATCTAAGGGCCTGTCACCGAT
TGACTGGAATTTAGATTAGTATGTTTTTTCCGCCGGTGTTTATTTTGGGGCGGAGCTTGTTGCTGTTAAAATATATGTAAAATTG
CGAGATAAACAATAAATTGAAGACCTTTCTCTGTGGTGGCGTGTGGTCTCTGATATGGGTACCGAGATCCCCTCTGTGACAGAAG
TTCTATATGGAGCCTCGTCTATTGACTGTGTTTCCGGATCGTACGAATGTCATGATACGCATTCTCATGTTACTCCGGACCACGC
AAAAGACGTAGCCGTCCAAGGGCATGTCAAAACGAATAACACCGAAGATGTAGAATCTTTGGACTCGTGTGGCTTTGACAGTGTT
TTCATGATATTTTCATTCGGGGAACTCGGGAGAAGACAGCTTACCGATAACATTCGAAAAGACATTTGTACTTCCCTAGACAGAG
TTCCGATGGCATGTACTAAGACGTCCGCATTTGCAGGTGCAAATAGACATCAGAAAAGTTTGCAGATGTTTCTCTTTTGCAAGAG
AAGACATGCCCCGCAAATAAGGGCTCGCCTAAAAGACATTATTTCGTCAAGAAAGTCAAGAAAATATTTTACGCAGTCCGAGGAT
GGAGAAACTCACCCCGGTGTGCCAGTTTTCTTTCACGAGTTTGTAGCCCATGCCCCGGTATTTATTCCACGCGACAATCTTGCCC
ATGCCTGTCGCAGATTGGCCAGGCATATGACTGGAGGAATGGCGTGTTGACTTACGTGGGCGCGCCTCTGGGTGGAACCCAGCGC
TAGAACATTTATACCTGCCCCATTGCGAAAGTTACTCAGAACGCGAATATTGCACTTCCTGGACTATGAAGTACGGTCAGGCTGC
TATTGCAACTGATATGGACTTTGCCCGCATGTCTCGCCAGCCTCCACGGAAGAACTCCAGGTGCTCTTCGGCACGGACGGCTGCG
TTACAGGGTAATGGATATTGCTTTCTACCGAAGTCGGAAAAGTTGCCTGAGCTACCCTCTAGACATTTCGAGACCCGATTTTCCT
CACTACTACCTCCAGCTGCAGCTAAGTGACAGCAACTATGAAGGTCAACTCATTTCCGCAATGCGACGCAATAATTAAATGCTCG
TATTTCATAATTTGGTGTTTATTACTTTTATTTATTCCTCTAACAGTCCG

>UL3

ACGAACATCAACCACTATCACCCGGACTATGCGAGCAGCCCATAAAAAGTCCCTTCAAGCGTCCGCTGGACATTAGCTCTCCTAC
TGAAAGCAACCAAAGAGATACCGACCCCAACGCTGCTAAAAAACGCCGTACTCTCCCCCCTCCGAGCATGACTTCCGTCCAATTG
GTTATACCATCCGCGAGGAAGCTCGTACCGGAGTGGACCGACGTGGCCATAGAGTTTTCAATACCATCATCATGGGAGTCAATTC
TGAAACCTCTAATTGAAAATGAAGATTTCGGTGAGACGCTGTTACAGTACTCCTACTGTTCTAGATTCCGCAAAATCTTTCCTCC
TAAAGGAGATATTTTTGCATGGACAAGATACTGCAATCCGCAAGATGTGCGTGTTATTATCGTTGGGCAAGACCCGTATCCCACG
GAAGGACACGCGCATGGCTTAGCTTTCAGTGTCCCTAGGGGGTGCCGTATACCGCCCAGCTTGCGCCGCATTTTGGAAGCTCTAC
GCGACTGTTACCCATCTATTCCAAAGGAAACTCACGGGTGTCTAGAATCTTGGGCCAGGAAGGGAGTATTGCTTTTGAACAAATA
CTTGACGGTGGAACAAGGTTTTCCTAGATCGCACTCAACCATCGGCTGGGACAAATTTACAAAGGGAGTGATCTCATCGCTACTA
GAAAAGTTGTCATCTACGGTGTTGATGCTTTGGGGCCAAGAAGCTAGACGATTTATTCCCCGCGTGCCTAAAAATCATTTAAAGT
TGGAGTTCATCCATCCTTCTACCTGCACGCGACGTCCGTTCAACTGTCGCCATTTCGAGGAAGCAAATCAATACTTCGAATCTAA
GGGCCTGTCACCGATTGACTGGAATTTAGATTAGTATGTTTTTTCCGCCGGTGTTTATTTTGGGGCGGAGCTTGTTGCTGTTAAA
ATATATGTAAAATTGCGAGATAAACAATAAATTGAAGACCTTTCTCTGTGGTGGCGTGTGGTCTCTGATATGGGTACCGAGATCC
CCTCTGTGACAGAAGTTCTATATGGAGCCTCGTCTATTGACTGTGTTTCCGGATCGTACGAATGTCATGATACGCATTCTCATGT
TACTCCGGACCACGCAAAAGACGTAGCCGTCCAAGGGCATGTCAAAACGAATAACACCGAAGATGTAGAATCTTTGGACTCGTGT
GGCTTTGACAGTGTTTTCATGATATTTTCATTCGGGGAACTCGGGAGAAGACAGCTTACCGATAACATTCGAAAAGACATTTGTA
CTTCCCTAGACAGAGTTCCGATGGCATGTACTAAGACGTCCGCATTTGCAGGTGCAAATAGACATCAGAAAAGTTTGCAGATGTT
TCTCTTTTGCAAGAGAAGACATGCCCCGCAAATAAGGGCTCGCCTAAAAGACATTATTTCGTCAAGAAAGTCAAGAAAATATTTT
ACGCAGTCCGAGGATGGAGAAACTCACCCCGGTGTGCCAGTTTTCTTTCACGAGTTTGTAGCCCATGCCCCGGTATTTATTCCAC
GCGACAATCTTGCCCATGCCTGTCGCAGATTGGCCAGGCATATGACTGGAGGAATGGCGTGTTGACTTACGTGGGCGCGCCTCTG
GGTGGAACCCAGCGCTAGAACATTTATACCTGCCCCATTGCGAAAGTTACTCAGAACGCGAATATTGCACTTCCTGGACTATGAA
GTACGGTCAGGCTGCTATTGCAACTGATATGGACTTTGCCCGCATGTCTCGCCAGCCTCCACGGAAGAACTCCAGGTGCTCTTCG
GCACGGACGGCTGCGTTACAGGGTAATGGATATTGCTTTCTACCGAAGTCGGAAAAGTTGCCTGAGCTACCCTCTAGACATTTCG
AGACCCGATTTTCCTCACTACTACCTCCAGCTGCAGCTAAGTGACAGCAACTATGAAGGTCAACTCATTTCCGCAATGCGACGCA
ATAATTAAATGCTCGTATTTCATAATTTGGTGTTTATTACTTTTATTTATTCCTCTAACAGTCCG

>UL2

ATATGAGATGTCGTGGGTGCGACGGGTCGTGCGGCATGAAATAGAAAATAGCCCCGCTCATAGTCCACCGCCCGCACGTCCGTCC
TCTATACCCCCTTTATTTTCTGCGCCATCTCCTACGGCCCTAGACCTCAGGTTATCATCCCTCCCCTCGCCATCTAGACGCTCCC
CTCCACAGCTTTCGCCCCACCACCCCTCGTCTAATTATGATGGCGCAGGGCAAAATACGACTGGTGCACAGACTCGTCATGGGGG
GATAAAGAATGAACCCGAATCCCCCGTCGCCGCAGACAGAGGTATTTTAGACCTCGATTCCATGCAGAATCGGAGCCTCCGCATG
CGAAACGATGACGCAGGTGAAGGTCCTTCAGGACTACAGAGCCGAAGTACTCCATCTAGGCCGTCATATTTATATTGGGACGACT
CTGATTCGGACGACGATCCATTCAAAAGTCCGCCACAAAGAGTACCTACTGGTTCCTGTAGACCACTACATTTGTATTCGGACGA
CTCAGATTCTGACGACGATCCATTTAGAAGTCGGGCGCAAAGAGTTCCATCTGCTCGTCGACGTAACACGGATCGTGAACGGAGA
CCTAGGTCTCCATATGATAATGATTTTTAAAATGTTGACTTGTTTTTTTCTAATAGCATCTCTTATGCTATGTTCAAGTGTCGAT
ATACCTGGCATTTTTCGTGATCCTGTTTCTGCCGCGGCCGCACCATGTGATAGTAATTCGCGAGATAACGGTGCGCGAATTTTCG
GGCCGTTGCCATCGGGAGACTACTTTGAAAATTTGACTGTGATTATTTACAGACCTCTGTGCGTAAACACGGAAGTTGTTTTCTA
CTATAAGAATGGGTACGCTGCAAAAGTGAATCCGTACTTGGGGATTAGACTTCTTCAAGACCGGCTGGAAGCTACCGGGCTTGCG
GCTGCTAACGTCTTGGCTCGTGGACTTGTTGATTCGTTGTATAAAAGGACCACCATAAATTTCAGATGGCCGGAAACGAACATCA
ACCACTATCACCCGGACTATGCGAGCAGCCCATAAAAAGTCCCTTCAAGCGTCCGCTGGACATTAGCTCTCCTACTGAAAGCAAC
CAAAGAGATACCGACCCCAACGCTGCTAAAAAACGCCGTACTCTCCCCCCTCCGAGCATGACTTCCGTCCAATTGGTTATACCAT
CCGCGAGGAAGCTCGTACCGGAGTGGACCGACGTGGCCATAGAGTTTTCAATACCATCATCATGGGAGTCAATTCTGAAACCTCT
AATTGAAAATGAAGATTTCGGTGAGACGCTGTTACAGTACTCCTACTGTTCTAGATTCCGCAAAATCTTTCCTCCTAAAGGAGAT
ATTTTTGCATGGACAAGATACTGCAATCCGCAAGATGTGCGTGTTATTATCGTTGGGCAAGACCCGTATCCCACGGAAGGACACG
CGCATGGCTTAGCTTTCAGTGTCCCTAGGGGGTGCCGTATACCGCCCAGCTTGCGCCGCATTTTGGAAGCTCTACGCGACTGTTA
CCCATCTATTCCAAAGGAAACTCACGGGTGTCTAGAATCTTGGGCCAGGAAGGGAGTATTGCTTTTGAACAAATACTTGACGGTG
GAACAAGGTTTTCCTAGATCGCACTCAACCATCGGCTGGGACAAATTTACAAAGGGAGTGATCTCATCGCTACTAGAAAAGTTGT
CATCTACGGTGTTGATGCTTTGGGGCCAAGAAGCTAGACGATTTATTCCCCGCGTGCCTAAAAATCATTTAAAGTTGGAGTTCAT
CCATCCTTCTACCTGCACGCGACGTCCGTTCAACTGTCGCCATTTCGAGGAAGCAAATCAATACTTCGAATCTAAGGGCCTGTCA
CCGATTGACTGGAATTTAGATTAGTATGTTTTTTCCGCCGGTGTTTATTTTGGGGCGGAGCTTGTTGCTGTTAAAATATATGTAA
AATTGCGAGATAAACAATAAATTGAAGACCTTTCTCTGTGGTGGCGTGTGGTCTCTGATATGGGTACCGAGATCC

>UL1

ACCAGGATCAATGGATGGATACGTGAGTTTACGACGAATCACCCTAGCGGATATGATAGCTTTTTGGCGTACGAACAAACCCCGA
CGGTACAGAGTGCATCGGGTCTAGAGTACAGGGTAGGAATGGCCATTCCTGAGACGATACAGCCCGACATGCTGACCAGACCACT
CACGGTGGATTTGAAATATTTTAAAAATTCCGCCGGGATGGTGTCGATCCGCCTTCCTGTCAAAGGCGAGTCTCATGACAGCACG
GACGTCAAGCCAGTTGGGCTCGCACATCCAATGGTTAAAAAGCGGCCGACCACCGTATCCGATACGTTTATGTTACCGGAGACAG
AAGAGGGGGAAGGCAACTATGGCCTTGCGAGTAGAGCAGAAGTGCGAGATGTATATGAGATGTCGTGGGTGCGACGGGTCGTGCG
GCATGAAATAGAAAATAGCCCCGCTCATAGTCCACCGCCCGCACGTCCGTCCTCTATACCCCCTTTATTTTCTGCGCCATCTCCT
ACGGCCCTAGACCTCAGGTTATCATCCCTCCCCTCGCCATCTAGACGCTCCCCTCCACAGCTTTCGCCCCACCACCCCTCGTCTA
ATTATGATGGCGCAGGGCAAAATACGACTGGTGCACAGACTCGTCATGGGGGGATAAAGAATGAACCCGAATCCCCCGTCGCCGC
AGACAGAGGTATTTTAGACCTCGATTCCATGCAGAATCGGAGCCTCCGCATGCGAAACGATGACGCAGGTGAAGGTCCTTCAGGA
CTACAGAGCCGAAGTACTCCATCTAGGCCGTCATATTTATATTGGGACGACTCTGATTCGGACGACGATCCATTCAAAAGTCCGC
CACAAAGAGTACCTACTGGTTCCTGTAGACCACTACATTTGTATTCGGACGACTCAGATTCTGACGACGATCCATTTAGAAGTCG
GGCGCAAAGAGTTCCATCTGCTCGTCGACGTAACACGGATCGTGAACGGAGACCTAGGTCTCCATATGATAATGATTTTTAAAAT
GTTGACTTGTTTTTTTTCTAATAGCATCTCTTATGCTATGTTCAAGTGTCGATATACCTGGCATTTTTCGTGATCCTGTTTCTGCC
GCGGCCGCACCATGTGATAGTAATTCGCGAGATAACGGTGCGCGAATTTTCGGGCCGTTGCCATCGGGAGACTACTTTGAAAATT
TGACTGTGATTATTTACAGACCTCTGTGCGTAAACACGGAAGTTGTTTTCTACTATAAGAATGGGTACGCTGCAAAAGTGAATCC
GTACTTGGGGATTAGACTTCTTCAAGACCGGCTGGAAGCTACCGGGCTTGCGGCTGCTAACGTCTTGGCTCGTGGACTTGTTGAT
TCGTTGTATAAAAGGACCACCATAAATTTCAGATGGCCGGAAACGAACATCAACCACTATCACCCGGACTATGCGAGCAGCCCAT
AAAAAGTCCCTTCAAGCGTCCGCTGGACATTAGCTCTCCTACTGAAAGCAACCAAAGAGATACCGACCCCAACGCTGCTAAAAAA
CGCCGTACTCTCCCCCCTCCGAGCATGACTTCCGTCCAATTGGTTATACCATCCGCGAGGAAGCTCGTACCGGAGTGGACCGACG
TGGCCATAGAGTTTTCAATACCATCATCATGGGAGTCAATTCTGAAACCTCTAATTGAAAATGAAGATTTCGGTGAGACGCTGTT
ACAGTACTCCTACTGTTCTAGATTCCGCAAAATCTTTCCTCCTAAAGGAGATATTTTTGCATGGACAAGATACTGCAATCCGCAA
GATGTGCGTGTTATTATCGTTGGGCAAGACCCGTATCCCACGGAAGGACACGCGCATGGCTTAGCTTTCAGTGTCCCTAGGGGGT

GCCGTATACCGCCCAGCTTGCGCCGCATTTTGGAAGCTCTACGCGACTGTTACCCATCTATTCCAAAGGAAACTCACGGGTGTCT
AGAATCTTGGGCCAGGAAGGGAGTATTGCTTTTGAACAAATACTTGACGGTGGAACAAGGTTTTCCTAGATCGCACTCAACCATC
GGCTGGGACAAATTTACAAAGGGAGTGATCTCATCGCTACTAGAAAAGTTGTCATCTACGGTGTTGATGCTTTGGGGCCAAGAAG
CTAGACGATTTATTCCCCGCGTGCCTAAAAATCATTTAAAGTTGGAGTTCATCCATCCTTCTACCTGCACGCGACGTCCGTTCAA
CTGTCGCCATTTCGAGGAAGCAAATCAATACTTCGAATCTAAGGGCCTGTCACCGATTGACTGGAATTTAGATTAGTATGTTTTT
TCCGCCGGTGTTTATTTTGGGGCGGAGCTTGTTGCTGTTAAAATATATGTAAAATTGCGAGATAAACAATAAATTGAAGACCTTT
CTCTGTGGTGGCGTGTGGTCTCTGATATGGGTACCGAGATCC

>UL0

ACCTGTAGCTCCGGAGCGCTATAAGTCGCGCAGGGAAGCCGCATTTAAGAACTCTGTGGGGGAGGTGACAATTAAGCTGCCTGTC
CGAGATGCTGGGGACCCCCTGGTAAGCTCTGTACGTCCCGTCAACGTGAGTGTGGCTATACCACAGTCAGAGAGCTGGCGTCTCC
GAGCACCTGTAGATAACGAATTTGTGGTCCACTACCGGCACTCCTCCTCCAGTGAGGATGCCCGACTGTCCACAAGTGAAAACGA
TAGCGATGACTCAATGAACGACGCTGCAGTACAAGAAGCGCTGATGGAAGACGCAGACGATGCCCCATTTGGAGCCGAAGATGAG
CAGCTTAGTTCCGAATCAAATGATTCAGACGGCAATTACTCTGAGTACGACCCACCGCCGGAAATGGTGGAGGGTTGGCATGGAG
GGGGACAGGGCGGGAGGGCCGAGAGCCCGCAGCCGCTGGCTGATGTGCGTGAGGTTGAACCCCGCGACCCCCAACGGAGGGAGGA
ACCAATAGAGACTATCGTATTGAGCTCAGAATCTGATTTGAGCGCGGATGAAGGAGAAAACAACAACCCCATGCAACGCCCGACG
GACCCGCGCGTATTTTTAATTCGTGTCCGCAGGGCGCGTGAAATATTACGCAATGTAGGTTTGGACGATAGAAGAATTGGGCCTC
ACAGGGGCAGGGATAGATACCAGTTAAGGTCCAGGTCTAGGAACCGATCCAGATCCATCGACGGACCTGATTTATAGACTGACAA
AGTATCAATAAACGAATGTTTATTATTTTGGCGCGTACGGTGTGTTTATTTGATTAGTAAGTGGGTGAGTGTGCCGCTACCGAAC
TAATCCGAAAACTCGGATATACCGTTACGTCGCGCATGTGTATCCGCGCGGGGGCTAATAAGGACGTGCGTCGGAGCTGCACGGA
TATTCTTTCCAGTGCAACGGCGAGCGCGTTAGTGATCTGGAGAGAGTAAAGAATCGCGCGCAAGAATGACAATGAACGGAGATTT
CAATATAGCATTCGTGGCGGACATCGATGAAGCGCACTATATGGATTTCTGGGTCTTGATGAGGCGTAAGTGGTTTGTGGGGTGG
GATGACGCGTGTAAGATTTTTTGAGGTACGAGGATTAATGGTTTGGTTTCTGTTGTAGCTTATTTTCCGACAACAGTTGAAAGCG
TTAGGAGATTACTTCAAACCAGGCGGCAGGCCGGTACCCCTATTACCGATGGAGAACGCCAAGCCATCCAGACTATTTTGGAACA
GGAAAGACCAGTAACTCGAAGACACCATACCGGAAGGCTATCCCCCATAGATGACGTTGTAGTCGTTTTGAGGGCGATAATGCTT
GAAACGTTTGTAACTCGTCAACTAACGTGGAGAAATTTTGAGGCGCTGTATTCTACTCTACTGAGGGCTAAAAGTGTCGTTTATA
GTATGGCAATCGCCTCAGCTCCAGAAATCTCAAACGCGCGTCATCTGGTGACGCGCGGAGCAAATTCCGCACTACGCGCAGTGGA
ATTGTCATTCCAGGATGAAGAACAGATGCTTGAAGTATTTTACTCGTTTCTAACCAGGATCAATGGATGGATACGTGAGTTTACG
ACGAATCACCCTAGCGGATATGATAGCTTTTTGGCGTACGAACAAACCCCGACGGTACAGAGTGCATCGGGTCTAGAGTACAGGG
TAGGAATGGCCATTCCTGAGACGATACAGCCCGACATGCTGACCAGACCACTCACGGTGGATTTGAAATATTTTAAAAATTCCGC
CGGGATGGTGTCGATCCGCCTTCCTGTCAAAGGCGAGTCTCATGACAGCACGGACGTCAAGCCAGTTGGGCTCGCACATCCAATG
GTTAAAAAGCGGCCGACCACCGTATCCGATACGTTTATGTTACCGGAGACAGAAGAGGGGGAAGGCAACTATGGCCTTGCGAGTA
GAGCAGAAGTGCGAGATGTATATGAGATGTCGTGGGTGCGACGGGTCGTGCGGCATGAAATAGAAAATAGCCCCGCTCATAGTCC
ACCGCCCGCACGTCCGTCCTCTATACCCCCTTTATTTTCTGCGCCATCTCCTACGGCCCTAGACCTCAGGTTATCATCCCTCCCC
TCGCCATCTAGACGCTCCCCTCCACAGCTTTCGCCCCACCACCCCTCGTCTAATTATGATGGCGCAGGGCAAAATACGACTGGTG
CACAGACTCGTCATGGGGGGATAAAGAATGAACCCGAATCCCCCGTCGCCGCAGACAGAGGTATTTTAGACCTCGATTCCATGCA
GAATCGGAGCCTCCGCATGCGAAACGATGACGCAGGTGAAGGTCCTTCAGGACTACAGAGCCGAAGTACTCCATCTAGGCCGTCA
TATTTATATTGGGACGACTCTGATTCGGACGACGATCCATTCAAAAGTCCGCCACAAAGAGTACCTACTGGTTCCTGTAGACCAC
TACATTTGTATTCGGACGACTCAGATTCTGACGACGATCCATTTAGAAGTCGGGCGCAAAGAGTTCCATCTGCTCGTCGACGTAA
CACGGATCGTGAACGGAGACCTAGGTCTCCATATGATAATGATTTTTAAAATGTTGACTTGTTTTTTTCTAATAGCATCTCTTAT
GCTATGTTCAAGTGTCGATATACCTGGCATTTTTCGTGATCCTGTTTCTGCCGCGGCCGCACCATGTGATAGTAATTCGCGAGAT
AACGGTGCGCGAATTTTCGGGCCGTTGCCATCGGGAGACTACTTTGAAAATTTGACTGTGATTATTTACAGACCTCTGTGCGTAA
ACACGGAAGTTGTTTTCTACTATAAGAATGGGTACGCTGCAAAAGTGAATCCGTACTTGGGGATTAGACTTCTTCAAGACCGGCT
GGAAGCTACCGGGCTTGCGGCTGCTAACGTCTTGGCTCGTGGACTTGTTGATTCGTTGTATAAAAGGACCACCATAAATTTCAGA
TGGCCGGAAACGAACATCAACCACTATCACCCGGACTATGCGAGCAGCCCATAAAAAGTCCCTTCAAGCGTCCGCTGGACATTAG
CTCTCCTACTGAAAGCAACCAAAGAGATACCGACCCCAACGCTGCTAAAAAACGCCGTACTCTCCCCCCTCCGAGCATGACTTCC
GTCCAATTGGTTATACCATCCGCGAGGAAGCTCGTACCGGAGTGGACCGACGTGGCCATAGAGTTTTCAATACCATCATCATGGG
AGTCAATTCTGAAACCTCTAATTGAAAATGAAGATTTCGGTGAGACGCTGTTACAGTACTCCTACTGTTCTAGATTCCGCAAAAT
CTTTCCTCCTAAAGGAGATATTTTTGCATGGACAAGATACTGCAATCCGCAAGATGTGCGTGTTATTATCGTTGGGCAAGACCCG
TATCCCACGGAAGGACACGCGCATGGCTTAGCTTTCAGTGTCCCTAGGGGGTGCCGTATACCGCCCAGCTTGCGCCGCATTTTGG
AAGCTCTACGCGACTGTTACCCATCTATTCCAAAGGAAACTCACGGGTGTCTAGAATCTTGGGCCAGGAAGGGAGTATTGCTTTT
GAACAAATACTTGACGGTGGAACAAGGTTTTCCTAGATCGCACTCAACCATCGGCTGGGACAAATTTACAAAGGGAGTGATCTCA
TCGCTACTAGAAAAGTTGTCATCTACGGTGTTGATGCTTTGGGGCCAAGAAGCTAGACGATTTATTCCCCGCGTGCCTAAAAATC
ATTTAAAGTTGGAGTTCATCCATCCTTCTACCTGCACGCGACGTCCGTTCAACTGTCGCCATTTCGAGGAAGCAAATCAATACTT
CGAATCTAAGGGCCTGTCACCGATTGACTGGAATTTAGATTAGTATGTTTTTTCCGCCGGTGTTTATTTTGGGGCGGAGCTTGTT
GCTGTTAAAATATATGTAAAATTGCGAGATAAACAATAAATTGAAGACCTTTCTCTGTGGTGGCGTGTGGTCTCTGATATGGGTA
CCGAGATCC

>UL-1

TGCGGCCTTAGGCCAAGGTGTCGTTACCCGGATAGCTTTCACATCGATAGCGCGTATTACCTTCGGCCTGACAACTGTCAGCTAC
CCAACCGCTCTGTCTTCTTTCAATCCTTGAGTATTATAACTCGCGGTAACCGGCCGTGGCCGCGTAGTGAGGCGGTTATACGGAA
TAAGACCCCCCCCCCGGGGCCTTCACGTTCGTGTCCGTTCTTGCCAGCGAGCGTTGCTGGAAATCGAAAATTCCAGTGTGGCGCT
TATTGTGCAAGAACCACGGGCGTGACCTGTATGACTACTGAGCAAAGCACCTCGCGCTGCCGCGAAACTTACAAACGGGGCGAAC
TGCGTGACGATTCTGTGATCGCGCTTCGAGGTCCAGTCGGAGCCCACCGGTGCCCACCTGAACCTTTTCCGGGGCTGAAAAGGCT
AAACTGGAAAATCTAGTAGAACTAGTTTCACGGCTGCCCGCGCTGACCGTTTGCTCTCCACGGGTGGAAGAACTTTGGGGGGGTT
GGCGGACGCTGGCGTTTGATGTCGGGCCGTCAAGTATCTCGCAACCCCAGAACCTGTCCAGCATCTAATAGTAGGGGTCGGACAT
GAAACCACAAGGGATTGGAAGGGCGGGGAGCACTCAGGGGACTGGGAGGGGACGGGTCGGAGAAAACCGCGGACCACCAGTGCAC
ACACCCTGCATCCTTGGCTGCTCATATATAGCCCCCCCCGCAAAACACCCCCCCCGGGTTTATTGCCGACGACGGGGCTGTGGAG
CTATTCGAAACCCCTCGCGCGCGGCACTGCATGCGCTTCCGCGACGCAATGCAGTACATCAATGGTCACTTTCGCGTGGCGGAAA
AAATGAAAGAAAGGTAGTACAGAGCGGGCAGCGGCGATGGGCAAGTCATTAAGAAGCAGCGCATCGAGCACCTCGTTGCAATAGA
CGGTAGATCGCTGCGACGCGTCGGGAGTTGCTCTTCCCACAGCGCCATCGGGCCCGTTCGTTGAAATGGCAGCGGCCCGGCGGGG
CATGGCGATAAATTTCGCGCGAATGGGTGACGAGGAAAATCTTAGTCGCGTGCAGAAAGATTTGACGGTAGAACTCGAGTTGACG
CCCCAGTTAACGGAACGCATACTGGAGATAGAAGAACGTAAGAGCCGGGCGGGAAGGGAAACTTGGGCGAGTGATGTGGATGGGT
AGTGGGTGCCTGAGTGGACATGGGCAAATGTGAGACACCACCATACTATTGTTTTTGTGTTTCTTTTTGCTAGCGCTCTTTCCTA

CCAGAATGCATGACGTAAAACAGTTGGGGAAGCTAGGGGAGGTCCGAGCGAGACATGAAAATCCGGAAGAGGGGCGTAAGATCTC
TGCATCCGACCACAGTGCAATTCAGCATGTGCTGGACAACGAACGCACCGTCATCAGACGACACCAGTCGGGTCGCATGTCTCCT
GTAGACGAATACCTGATGGTAGTGAAAGCCGCCACGATGGAGATGCTATTGCTGCACAAATTGACGATAAAGCACATTCTGTACA
TACTAAAGGCTTTTGTACGACTTAAGGAGGTAGTATATGATATCATGACACGTACAAATCAGCCCGAGCTCCTAACGGAAAAGGT
GTACGGGGGCGCATACGAAATTTTCCTACGATTTCCTCTCCCAGTACGTGACCCAGAGTCGATGGCAGAGGATTTCCTACAATTC
CTGCGACACTCCATTAGGTGGTGTCGTAAGTTCAAGAGTAGAACTAATCGACGCGGTCACAAAGCTGACATAGGCTATTTGGAGC
ATCGCACAAATATCCAGGAAAATGAGATGAGGTATTACGTCAACCCTAGAAAGATGCCCAAAAAAGACACGCTTGTGAGGGTGCG
ACCTGTAGCTCCGGAGCGCTATAAGTCGCGCAGGGAAGCCGCATTTAAGAACTCTGTGGGGGAGGTGACAATTAAGCTGCCTGTC
CGAGATGCTGGGGACCCCCTGGTAAGCTCTGTACGTCCCGTCAACGTGAGTGTGGCTATACCACAGTCAGAGAGCTGGCGTCTCC
GAGCACCTGTAGATAACGAATTTGTGGTCCACTACCGGCACTCCTCCTCCAGTGAGGATGCCCGACTGTCCACAAGTGAAAACGA
TAGCGATGACTCAATGAACGACGCTGCAGTACAAGAAGCGCTGATGGAAGACGCAGACGATGCCCCATTTGGAGCCGAAGATGAG
CAGCTTAGTTCCGAATCAAATGATTCAGACGGCAATTACTCTGAGTACGACCCACCGCCGGAAATGGTGGAGGGTTGGCATGGAG
GGGGACAGGGCGGGAGGGCCGAGAGCCCGCAGCCGCTGGCTGATGTGCGTGAGGTTGAACCCCGCGACCCCCAACGGAGGGAGGA
ACCAATAGAGACTATCGTATTGAGCTCAGAATCTGATTTGAGCGCGGATGAAGGAGAAAACAACAACCCCATGCAACGCCCGACG
GACCCGCGCGTATTTTTAATTCGTGTCCGCAGGGCGCGTGAAATATTACGCAATGTAGGTTTGGACGATAGAAGAATTGGGCCTC
ACAGGGGCAGGGATAGATACCAGTTAAGGTCCAGGTCTAGGAACCGATCCAGATCCATCGACGGACCTGATTTATAGACTGACAA
AGTATCAATAAACGAATGTTTATTATTTTGGCGCGTACGGTGTGTTTATTTGATTAGTA

>ICP4AS

AACTTCCTTGGCACTCCCAGGAAGCGTGACCCATGTGCTTTCGTTGGCATTCAAAATAAACTTCACTGCTTCCCGGTGTGGCCAA
TAACGTACAAGCGAGCGTGTTTCTCGGGTAATATCAATGCCAACAAAACACACGCGTGACCTTTCTAGAGTAGGCACACACCCCG
GCACCGAATTGCACGTACCTGTGGACACAAGTAATGTATTGTTTTGTTATCTTCGTCATCTGTTCTACAGTAGGCTCAGGTTTGT
TTATTATGTTGGCTCATAAAAAATGCCGGAAGTTCCTGGTACGGTAATGGTATGCTGGGCGCATGGCATGCATTGGCTATGGCAC
ACCGGAACTGGAAATACGCTCGCTAACGATATGCAGGCCGCATGGGCGCCATTGGATCTGAATTATATCATGCCTCAATACAAAT
GCCATACCCAAACCAACTGCACAAGACCGAGGGGTTGACTCTCCAAAGTTCTTTCTTCACAACACCTTTGCTGAAACCTACCCAG
CGTGGAAAGGACCTCTTCGCAAAACGTAAGTCATGAAACCAATATTCTGTTATAAATTTTGGGGTGGGATTAGGTGGGCTGTCAT
TATTTGAGGGGACCACGAAATATCCTTTTGCATGGGGTATCGTGCAAAAATATGACTGTGGGCTTCGAATAATTCCCCCATACAG
TCAAGAGTATGGGCTGCTGTTATGGATGAACATGAAGCGCACAAAACGATAATCCACTTCTTTAGCATGTATTGCTGTTCCAAAA
TTGTGCCTATGGTTGCGGAGCTTTATTGTGCTCTGCATTGCGGGGACGGAAATAAACAGGCCAACGCGAGAAACGGGAGGTTTCG
AGAGTCGCGCATCTTTATTTTTTCTGTTACTACTCCCCACCAGAAAGCTTCACGTTTCTTCCAACTTCTGCACAAAACTTACGTA
ACCTTGTCACTCGGGTCTTGTTCTGCAGGATTCTGGAAATCGCTAGTCCTTCCACTAGCATTCTATCATGTCCTCTGATTCCTCG
ACGCCGAGTAATTTGGCCGCTCCAATTGTTGATGAGTCGGGATTCTCAACTGGCCCGCCCTCAGATCTATTTACCTTTTTGCGCG
CCGATAAGAGTGGGCGCTACGATTCTTTGGGTTCTCCCATCCCGAACCCTCCAAACGCACCTACCTGTGAGACCACTACTGATAG
CTTTTCGTACAGCACGCCACCTCTTTTTTCCAGCGAGGACGAGCTCTTTCACTCTGTCATCGCTGAGAATGCGGCCCAAGACGCG
GCCCAAGACGTGGACGGGCGAGAAATGTCTCTCCGCGACCTTATAGCGATGTTAAACGACATGGATCCCGAAAACATAAACTCTA
TAGAACCGTGTCCGAACGCTACCAGTGATGTCGTGGCACTAGATATTAACGTGGATACCGATCAGCCCGGAGTTTATACGGTCAT
CGACTCGCTCGCGCATCCCTGCGAAACTCCCATGGAGCAGCAGCACGTTGAATGCCCGTGTCCATTACCTGCAGTAGAGAAAACA
ACCGCTGTGAGGCGTCAAGAAGTGGTGAGGAAGTCAGAGAGAGTGGCCAGATCTCGTTCGCGACGGCCGTGGCGCGATCTTTGGA
GCAACCGGCGCCCCGTCCCTTCGCCGCAGAGGACCAGCAAAGACCGTCTCCCCAAGCGAGGGAAGAGGGAATTTTCGAAGAAGAT
GGGCCCGAGCCACCTTACCTCTTCCTCCTCTTCCTCCTCTTCCTCATTCTCGCTGTCAGGGAGGAGGGGACGTCTCGCCCGAAGG
CTGGAAAGCGCAGGAAGTGAAAAATCCACCGAAAGCCCCGACCGGGACTCTGCCCAATTAGTGCTACCTGGAGAATGTCCCGATG
TCGCGCTAGAGCCACTCTGGCGAGTTTTGAGGGAGTGGGTCGAAACGGACGGTACAGACGGGTCGGTTCAGTCAGTAACCGCGCC
TACGGAGCTCCCAGACCTAGGCCGCGGCATTACCCTGAATGTACTCTCACGAGCGTTGGCCGTCGACCTCCATAGTTCCGAGGCT
CGCGTTTTTGTTCCAGAAGCTATTATCCCATCCCCACCCAGTAGAGGACAAGTATATCATGGAAGTGGCTATAGACAGTCTTTGG
TCGATGACCCGCAGGTTCGGGAGGCCGCGGCCGAATTCCATGCCAATCCGAAATCAGCTTCCGTGTATCTCGAAGAGTACGGGAT
GCCAGTGGAACAGTTAGAGCGCCTTGTGGAGGAGTTCATGGTCCATCCCCTGTCGCGCGTTTCGTATGGGTATGGTGGCCCCAAG
CTGAGCCCACGTGACAAACGGCTCTGCGAACTGTGCGCTCTCCGCGGTCCGGGTTCTTCCGAGGGAGGATTGTCTCGTTCTGCGG
TGCCGACGTGCCTCCCGCACGTAGCGCACGCCATGCGTGATGGCCGTGTGTCATTAGCTCTGCCGCACCTTACCGAGCTCGTACG
CGTGTGCCGCCGCTATGACCACCCGCAGAAAACTTACCTGTTGGCGGCTCTTAGGAGAGCGTTTGTCCCGTTTGTATTTCCAGGA
GCGGCCGTGGATCCTACGGCGGCGTCTGCTTCTGGGGACCACCCTGTACTCGTTGGCGCCAGGCGAATAGCGCGTGCCACGAGGG
AGCTGGAAGCCGCAGCGAAAGAGCCACCTGCGGCAACCCCGGAAAAACTTTCCTCTTCGGCGCTATTGCGCCATACTAGAACTTG
CATTGTTGCTGCCGCAGATCTCATGGCGGCGCTTACTGCTATAGACTGCCCACCAGTTACAATAAACCCCGCCAAGCGTCGCCCG
GATCTGTATCAAGCACTCTCGAGGGCGCCCAAAACAGTCTTTACGTGGGAGCGCGCCATGCCGTACTGCGCGGCTTCGCTATGGA
GTCGCCTCGTCATAGGGTCCCTGCCGCTGCCATCTACCGATGATGTCACTCTTCTTGTAGAAGCCATGGAAGCTGCCGTGGGCGT
GCTTGGCTGCTCCGACGAGGGTGACGTGCGGCCTTCAATTGCCTACGCGGCCTCTGGTTGGAGGATGATCGCTACCATGCTAGGA
GATGATTCCGCGCCCGGCAGCGAGCACGTCACGGACGATTGTCCCGCAGATGATGACACCCACCACCATGACAGGCGCAAAAGAC
GAGGCGGCCGGCGGACGACTTGCCGGTACGACCTACCTCCCGGCACTCCCGATCGCGCGGCCCACCTCCTACTCTACGGGAGTCC
TGGCGGCGGCGACGGAGCAGAGCCTTCGGCCGCCCAAGAATCTCCCGCCAACCCCTGGCCCAGAGCTCCACCGTGCGATGAACAG
GAGCCATTATCTGTGTCACCGTACGGGCCCGAACCCGACCGCCCTCCAGATGATGACTTTGAGACCCGTAAAGGTCTGAAAAGAA
AGTCATCTGAAGATCATGCGGACCCGATCCCAGAAGGCAACGCGACTAAAAAAACATGTGGCCTCCAGGGCTTACCAGATTCATT
ACCTCCAGCGGTTCCAGAAACAGATAGAGACAACCCCCTCCTTCCGCCATGCCCCATAACTCCCGAGGGTCCACCATGTCCGCCC
CGAGAAGAACCGCAACAACCGCAAGAACCGCAAGAACCACAATCCCCGTCCTTTCACATTTCCGAGATAGGTGAGGCCCTGTTTC
ATTCTACCCCAGTCAGCCCTACCATCCTGTTCGCACCGGAGGGTTTCATCCCGAACCAAGGGGATGTTTACCGTCCCCATGGCAA
AAACTTGGGCGTTGCGCTCTCCGCAGGGGAGCCATACGTGCCCGCCCATCGCGCGCGTCCTGCTTTGGAGCGGCTCGCGAACTAT
TTGCGCGGGTTGAGCAGCGGTCCGGCGGGAAGCTATTCCGCGACCGGACCCGCAGATTGGCTACCTGCTTTTGCGAGGGACCCAG
TAGCCCTGGGGGAGTTTTGCAAGAGGATTTCACCCGTGAACAGAGGGGGACGCTGCGTCTTAGAGGAGCGGCTTTCCTGGACGTT
GCGACATCCGGCTGACAAAATCTATGACCTCATCATTCTTGTCTCAGAACGTGCCCCGGGCGAATTTTTAGAACGCGCGTACGCC
TACGCGGCCCTCTGCGGACGTGCCTGCATGCCGGGCGGTGCCTGTTGGCCAGAATCGTGGACACGCGGGGAGTATCCCTATATCC
CTCCCGACGAGAAAGACCCGGGCCGCAGCGTGGCCATTCTCTGTACGACAGACCTTGGATACGCCGGGGCGGTAGAATGGATGCT
TGAGCGCGCTCTCAGTACATCTGGGCGCGCATGCGTGGTCGTAGATGCTAGGGATCCCGGGGACGAGGATTCCCGCTCGGGTCCG
CGCATACCTACCGGGCGGAGGGGTGTTGTCTATGCCCGGACTAGCCCCCCGTCGCACGCTTACCTTGTAAAACTGTTTACCGGTC
CGGCGGCCGAACGGTTGGAACGGGGTGGGCTCAATGTAGGTCACGTAGTAATGGACAGGCGACTCACTGTCTCATGGCCCCTCAG
CGTTGGGAAGACCGACAACGACGATGAGGGTACAGGAGACAGAGAGAAGTGTCCTCCCGTTAAGCACCCCTTCCTCACTCACGTA
ACTCTCAGGAGATTACGCGACGTTGCTCGTGCCATGGATGCCGGGTGGAGCCACTCCGAAATCGGAAAAGCTTCAGCCGTCTTGT

ATCCGTACATGGCCCAGCCCCATGTGGGCGCCGGGGCGATGCCGCCATGTCCGGACCTTTCCGAGTCCAGTTCCACTATGCATTC
TTCTTCGTCGTCTTCCTCTTCCTCATGCTCCTCGTCGTCGTCTTCTTCAGACTCCAGCTCATCTGAAGAAGACGGGGACGAGAAG
AACGAGAAAGAAGATCGTGAGCGCGCCGGAGGTGGAAAGCGGCGCGGCCGCCAGCGCCTCCCGATTCGGGACCGCGTGTACCGTG
GCCGGGACTACCGCGTGCGCACTCCCTGTGCTGATATTGTGGGAGGCACTTCCGCTTCCTTTAGGGCCGTCGCTGCGTGGGGGAG
GGACACGGTAATTCCCGCCCTCCCTCTCGCAGCGTCCTGGTCCGATCCCAGCAAAATTCCACAGGAAGTTCTCCGCATCATCTCC
GATTATTACCCGGATGCGCCAGGGGCGCGGGCGCGTCGGAACCCGCTTCACGTGGAAGGACTAGCGCTTATGCGTGCCAGAAATC
CCGCCCCTCTCGCATTGCTACTTGGTGACGACTATTCCCATTATCACACTCCCCGCAATCGGTCCCCAATGTGTTGTTGCTGGAG
TGCTCCGCGGGGAGCAACGCCAGAGTCACTCAGATGCACCGGACACACGCGCACCAGGCGCCCAAGGCTATCCGTCTCCTCAGGG
AGGGGGGAAACTAGAACCCTTGTCCCCAGATTAGATTAGACTCTTCAGAGGATAGGCGGTCGCGTCAACGCCATAGACGCATGAC
TAACTACCCTCTAACGCATCCCACTGTCACTGACCCTTGGTGGTAACCGTGGTTTCCGTTATTAACTTATTAAAACTTTTTGTAC
AATTTACTTCTGGTTGCGTGCTCATTCTTTTGCTTGTCGGAGCTT

>US10AS

CGTGGTCTGGAGGCATCGGCGGGGGGATTTTCGGGTTGCTCATGTGAGAAGACAATGACCCCTCCCATTGTGTCCCCAGTATGAA
ATCTTCTCCCCTCTTCCTGGGGTAATCCGAGCAGGACCCGATGGTCCCGCTATATGCGGGGACAGCGACCCGGATCGTTCTTTTA
TGCAGTGCCCGAAGAGGCAAGTCACAGCTCATTGGCTGGAAGGGGGGATGCAGGACCCTCGGTGCGAACGTGACGTAGGTGGGTG
CCACCCACATCATGTAAATTAGGTCACGGGGTAGGGTTCTGGATCCAAAAAGGATCCCGAAGAGCTCTCCCAGAAGTTTTTCTTT
TCGGACGTATCGGAGGACGAAGAACCGGCACGCGGGAGGAGCTGGAGCGACCCGGAGTCGGAGGAAGAGCAGCCTGGGTGCCGGG
GAGTGGACTTGGGCGAGGAGGACACGGGACACAGCTCCACCGAGTCAGAGCCCACGCAATCTGACTTAGACTTTATTGACGACAG
CTCTCCGGCGCCGCCGCCATTTGCTATCCCCCGCGTCCGTGCGTTATTGCGGTGCGCGGCACCCGCAAAGACCCACGGAAGGCTT
CGGCCGCCAGGGCGGGTAGGCGCACTCTTAAAAGACGGAGGTTGTCATTTTCTTCTTCCTCTGACGAGGAATCCGAGGAGAGAAG
TAAAAAAGAAGAAGCGGCCTCGACCCCTGCACGGCGACGCAAGGCCGAGGCCTCGACGAGCAGATAGAGGAGACGCGGGGCAGAA
CCTCCCCCTCCCTCCCACCCCCCTACTCTGGACATTTATTGCCCGCTCGATCCATTCTCATCCAGAACTTCTTTCCCGCTCAGCC
TTCACGCAGAAGCGGACGCGCGCCCCTTTGCGACCGCCGGACATCCCGCCGCCCCCCCCCCTTCACGCCCGGCGCAATCCGTAGC
CGTCCAACTCGGCCCAGCACAACCGCAGTAGACCGCCCGGACCGCTCTCCTCTAGACACATCCCTAAATGGAAAACATGCTCGAC
GGGTGCTACCCGCTGGCGCTGATGGACAGCGATCACATTACTGCGCACGCGGTACCTCGTGGCGAGCGCAGGCGGCAAGGTGCCG
CTGTCGCCTCGTCGGAGTCGGCCGACTCGGTAGACCCGTGCATTCGGATCGCCTCGCGGCTCTGGCGCGAGTTAGTCGAGATATC
GTCCGAACTCAAGGACGGTTACGGAGAGTTCACGTCAGCGAGAGACCGCCGCAACGCGCTGATTGCTGCCAACGAACGGCTACGT
TCGGCTTTTCTGGGGGCCAGCCGGGCGACGCGCGGCCTAGGTTTGAGGCCGCGGTGGGCGTCGACGGAGAGCGTCGCCAACTCCC
CCACTGACCCGAATAACGGCAACGGGTTGGGAGAATTAGAGGAGGCAATGGAAGGGATCGAGGGCGATTTCTGGCTCGACTCTCT
GGACGGTGACCGCTTCGAGGACGAGAGCCGTACCATGCAGAGCGAGAATATGCGTTTCGTGATCGAGAAAGAACTGTTATCCTGG
CTGTCCCGACACCTGCCGGCCGACCTCGCGTCCGCCGAGCGAGAGACCTCCCGGTCTCTCCTGGCGGCCGGGCACTGGTGCTGCT
TGTGGCACCCTCGGCCGTGCCGCGAAGCGTGTTTGTACGACTCGATTTACGTGCAGAGTCTTTTCTGCGTCGGGACGGGGAGAGT
CCCGCAATCGGAGATGCGCCGTCGCGAATACCTGGCCGCCTTGCGCGCCGGCGCGGCTGCCGCCAACTCTCCCGAAGTGAGCGCC
TCGATCTTTGCGAGGGACGCTGGAATCGCGCTGGCGCTGGCGCGGCGCCGTTGACGGGAGAATGACGCCCTCTAGCGGCTTCCTT
ACCTCCGCGTCCCTGACAACCTCGCGGGTTTTTACACTGTCCTCCGTCCACTCTCCCCCCTCACCCACTCCGCGGCAGCGAAACA
CAACCCCCCCCCCCCCCCCAGAAACGAGCGACACGCGAGCGCTGCGAAATAAATAAAGTAATATTATTGTGTGTTTTTCACGTTGT
TGCAATCGAGAGGCCGTTTGT

>SORF43AS

GCATCTTCTCTAAGCTGTATATATTGGCCTGACTCTGCTAGAATAAGAGCGTTATATTTCTGACATCATAGCGGGGTGGAGTCTG
AGCGGCGATGGCGCCTGTAAAAGTGACTATAGTTTCTGCGGTCGATTCGCACTACAAACTACCTAATTCTAGATTTGAGCTCTCG
GATTCTGGATGGAAAGAATTGGTTCACGCAGTGAAAACTATGGCGAGTTACGATCGTCCGAGTACATTATCGGTAATCGTGCGCC
CGGCATCTCTGTACGAAGTTTCCGGGGAGCTGTTTTCCCTTCCCAGGATGTGCAGACCCGTGATTCGGTTCGGTGAGGGGGGCGA
CCCGCCTGGAGTAAGTCCCGAGTGGAGCGGCTTGGACGCAGGGTTTTACCATTTGTCATCTGGCGCGTATGCCGCAAAAGAGTTC
CATTTGTGGGTGCTGGGTACCGCTGACATATGCATGGCAGCTTTAAACCTCCCTGCGCCAAAAACTTTCCTAATTACCGAAACCG
GAGGTAAAAATTTTGAGAGAGGAGTGGAAATATTTTTGGTAAACGGAGACAAGACAACGCTGTCTCTGAGTCACCCATCAGTCTG
GACAACTCTTGCCCCTTCGAGCCTGAGAACGCCCTGGCCGTACAGCACGGTAAAGTTTTTAAAAGTAAAACCTAACTCGGCCGCA
TACTGTGTTTCCGACTCGGATGATGGCGAACGGCAGCCAAAATTTTTTCTCGGGAGTCTATTTAAGTCGAAGAAACCCCGCTCCC
CGCGGCGCCGACGTTAGCGGCGACGCTCGCCGCGTCGGCGGCCGCTGCCGGAATAGTGGCCTGGGCCGCGGCGAGGGCCTTGCAG
AAACAAGGGCGGTAGCAGTAATAATAACCACACAAATATTGACAATAATAAACGCGTACGCGGATGAGTAAGTGTTATTGTCTCG
CGCGCCATCTTTATAAAAGCCCGCGTTGCGTGGGCCGGCGGGTAGCATTTGGAGGGTTGGCGACCATGTCGAGACCTCCGACGTC
ACATTTGGACTTAGCTTTCTCGGCGGCCTTTAGGGGCACGGACCTGCCCGGAGGGAGATTCTGGCGGGCGTCGCAGAGTTGCGAT
ATTTTCTTTTGGCCCGATCTGGCCGCGGTGATCGTACAGGCCGCCCGCGCGTATTTTGAAGGGAAGGAAAGGCTGGGCAGTCTGC
AGGTCGCCGAAGATATCACGGCGCACGACCCGCGAATAGCGCCCGCGGCTAAGCGCGCCGTCGCAGCGGCGGTAGGACTGTGGAC
CGCGCTGTCGGAGTTAGTTGGGGGGCCGAACGGGGAGTTGGAAAGCAAGGTCTGGGGCAAGCAGATTCCCCGGGCCGCCGCGTGG
GAAATAAGAGACGTGCCCAAAGTTCCAGTCATTGGGCCGGACATTCTTTCTTTTTTCTCCGCCGCCGTCGAGCTGCCGGTGCTCT
ATATCAGAGCCCGGGGAGGGGCGCACTCGCGGTCCGCGCACTGGAATAACCAGAGCAGCGCGCCGGCCGCCGGACTCGCGGCGAT
AAGGATAGGCATGGAGATGGTGCGGAGCCTCCTGGTGATAGCGCTGCCTCTGTCAAACTTCACCCTCCCGGAAGACCTCCCCGAA
GGTTCCCAAAACTCGATCCGCGCGTTCGTGGCCCACCTCATGAACTGTGTAGCTACCGATAAGATCATGTCTCCGGACGTGCGCG
TCCCAGTCGAAGAAAGCTTTTACAGCCACTGTTTAAGGGAAATCATTATGTGCGAGAGAGCTTTTTGTTACCCGTGCAATCCCCC
GCCAAAATGGTGAGCTCAGGGTCCCATTTACCCCCGCAACACCCTCTCGCCGCCAGGGCGCGCGCTCTATCTTTCTCTATGTCCC
GTCGCCACCGGCCTAACCGAACGGTGGAACGGGGCCGCCCGGGAAAGCCTAGCTCCGCACAGACACAGACAGA

>US2

TTCAAAAGGCGTGACAGTTCCAAAGTCGCCCAGCACTGTCTTGCCGTCGCAGTTTAAGAAAATATTTTCGAGTTTGATGTCCCTG
TGAATCAGTTTCTTTGAATGGATATACGAGACCGCTGACATGAGCTGCTTCATGATGCGTCTAGTCTGTTTAAGCATTGGACTGT
CTTTCCAGTCGAACGCTTCATCGTACATGTAGCTGTACAAGTCGAACTGATACCTCTGCGTTAACATGTACGTATTATCCGGGTA
CCGAAGGATTTCCTGGATCTTGAGAACATTCTCGTGATTTAGGCGCCCGAGAACCAAAATTTCATTTTCCAGCTGAGAGGCCGCC
CTTGTTCCATTTTTCACATATTTAGCTATCAGGCGCTTACATTTTGATTTCCCCTGGTTAGTACTGTTAACCCCTCTTCGACGTT
GCTCTTCCTCGGTGTAGCGACGGACGTCGCAGATAAACCCGTAAGCCCCCTCCACCAAATCGTCTAAAACTCGGAAGTGCTGAAG
AAATTCATCCGAATGCTTCAATCTAGTTTGCGCGAGCGGGACAGGGCCGGCGTAGTCGCTCGATGCTTCATCGAAGTCTGTATGC
GAGCCGTCAACTTTTTCAGTAGTCGAGGGCCCATCGCCTGCATATGCGAAGTTCTGCCCGTTCTCGTTAGCCGGGGGGATACTGT
TTACAGCTGTATTCTCTGTCACAGGTTCCGCGCGTTCTGTCATTGTCACACTAACAAGTTCGGCGAGGCGCTCGACGCCTTCCGC

GGTCGGCAAAATTTCTGCATAGGGGCGCAATGACATTGTCCGTGAGAAACTATCCAATACGCAAACGCGTTTTGAGGTAAGCGGA
TTCTCGGTTGTTCTTCTTCGTTTTTCTGCCCTAGAGCGTGAACAGATGCGCCGAAACCGCATCTTCTCTAAGCTGTATATATTGG
CCTGACTCTGCTAGAATAAGAGCGTTATATTTCTGACATCATAGCGGGGTGGAGTCTGAGCGGCGATGGCGCCTGTAAAAGTGAC
TATAGTTTCTGCGGTCGATTCGCACTACAAACTACCTAATTCTAGATTTGAGCTCTCGGATTCTGGATGGAAAGAATTGGTTCAC
GCAGTGAAAACTATGGCGAGTTACGATCGTCCGAGTACATTATCGGTAATCGTGCGCCCGGCATCTCTGTACGAAGTTTCCGGGG
AGCTGTTTTCCCTTCCCAGGATGTGCAGACCCGTGATTCGGTTCGGTGAGGGGGGCGACCCGCCTGGAGTAAGTCCCGAGTGGAG
CGGCTTGGACGCAGGGTTTTACCATTTGTCATCTGGCGCGTATGCCGCAAAAGAGTTCCATTTGTGGGTGCTGGGTACCGCTGAC
ATATGCATGGCAGCTTTAAACCTCCCTGCGCCAAAAACTTTCCTAATTACCGAAACCGGAGGTAAAAATTTTGAGAGAGGAGTGG
AAATATTTTTGGTAAACGGAGACAAGACAACGCTGTCTCTGAGTCACCCATCAGTCTGGACAACTCTTGCCCCTTCGAGCCTGAG
AACGCCCTGGCCGTACAGCACGGTAAAGTTTTTAAAAGTAAAACCTAACTCGGCCGCATACTGTGTTTCCGACTCGGATGATGGC
GAACGGCAGCCAAAATTTTTTCTCGGGAGTCTATTTAAGTCGAAGAAACCCCGCTCCCCGCGGCGCCGACGTTAGCGGCGACGCT
CGCCGCGTCGGCGGCCGCTGCCGGAATAGTGGCCTGGGCCGCGGCGAGGGCCTTGCAGAAACAAGGGCGGTAGCAGTAATAATAA
CCACACAAATATTGACAATAATAAACGCGTACGCGGATGAGTAAGTGTTATTGTCTCGCGCGCCATCTTTATAAAAGCCCGCGTT
GCGTGGGCCGGCGGGTAGCATTTGGAGGGTTGGCGACCATGTCGAGACCTCCGACGTCACATTTGGACTTAGCTTTCTCGGCGGC
CTTTAGGGGCACGGACCTGCCCGGAGGGAGATTCTGGCGGGCGTCGCAGAGTTGCGATATTTTCTTTTGGCCCGATCTGGCCGCG
GTGATCGTACAGGCCGCCCGCGCGTATTTTGAAGGGAAGGAAAGGCTGGGCAGTCTGCAGGTCGCCGAAGATATCACGGCGCACG
ACCCGCGAATAGCGCCCGCGGCTAAGCGCGCCGTCGCAGCGGCGGTAGGACTGTGGACCGCGCTGTCGGAGTTAGTTGGGGGGCC
GAACGGGGAGTTGGAAAGCAAGGTCTGGGGCAAGCAGATTCCCCGGGCCGCCGCGTGGGAAATAAGAGACGTGCCCAAAGTTCCA
GTCATTGGGCCGGACATTCTTTCTTTTTTCTCCGCCGCCGTCGAGCTGCCGGTGCTCTATATCAGAGCCCGGGGAGGGGCGCACT
CGCGGTCCGCGCACTGGAATAACCAGAGCAGCGCCGGCCGCCGGACTCGCGGCGATAAGGATAGGCATGGAGATGGTGCGGAG
CCTCCTGGTGATAGCGCTGCCTCTGTCAAACTTCACCCTCCCGGAAGACCTCCCCGAAGGTTCCCAAAACTCGATCCGCGCGTTC
GTGGCCCACCTCATGAACTGTGTAGCTACCGATAAGATCATGTCTCCGGACGTGCGCGTCCCAGTCGAAGAAAGCTTTTACAGCC
ACTGTTTAAGGGAAATCATTATGTGCGAGAGAGCTTTTTGTTACCCGTGCAATCCCCCGCCAAAATGGTGAGCTCAGGGTCCCAT
TTACCCCCGCAACACCCTCTCGCCGCCAGGGCGCGCGCTCTATCTTTCTCTATGTCCCGTCGCCACCGGCCTAACCGAACGGTGG
AACGGGGCCGCCCGGGAAAGCCTAGCTCCGCACAGACACAGACAGA

>US3

ACATGGTCGCCAACCCTCCAAATGCTACCCGCCGGCCCACGCAACGCGGGCTTTTATAAAGATGGCGCGCGAGACAATAACACTT
ACTCATCCGCGTACGCGTTTATTATTGTCAATATTTGTGTGGTTATTATTACTGCTACCGCCCTTGTTTCTGCAAGGCCCTCGCC
GCGGCCCAGGCCACTATTCCGGCAGCGGCCGCCGACGCGGCGAGCGTCGCCGCTAACGTCGGCGCCGCGGGGAGCGGGGTTTCTT
CGACTTAAATAGACTCCCGAGAAAAAATTTTGGCTGCCGTTCGCCATCATCCGAGTCGGAAACACAGTATGCGGCCGAGTTAGGT
TTTACTTTTAAAAACTTTACCGTGCTGTACGGCCAGGGCGTTCTCAGGCTCGAAGGGGCAAGAGTTGTCCAGACTGATGGGTGAC
TCAGAGACAGCGTTGTCTTGTCTCCGTTTACCAAAAATATTTCCACTCCTCTCTCAAAATTTTTACCTCCGGTTTCGGTAATTAG
GAAAGTTTTTGGCGCAGGGAGGTTTAAAGCTGCCATGCATATGTCAGCGGTACCCAGCACCCACAAATGGAACTCTTTTGCGGCA
TACGCGCCAGATGACAAATGGTAAAACCCTGCGTCCAAGCCGCTCCACTCGGGACTTACTCCAGGCGGGTCGCCCCCCTCACCGA
ACCGAATCACGGGTCTGCACATCCTGGGAAGGGAAAACAGCTCCCCGGAAACTTCGTACAGAGATGCCGGGCGCACGATTACCGA
TAATGTACTCGGACGATCGTAACTCGCCATAGTTTTCACTGCGTGAACCAATTCTTTCCATCCAGAATCCGAGAGCTCAAATCTA
GAATTAGGTAGTTTGTAGTGCGAATCGACCGCAGAAACTATAGTCACTTTTACAGGCGCCATCGCCGCTCAGACTCCACCCCGCT
ATGATGTCAGAAATATAACGCTCTTATTCTAGCAGAGTCAGGCCAATATATACAGCTTAGAGAAGATGCGGTTTCGGCGCATCTG
TTCACGCTCTAGGGCAGAAAAACGAAGAAGAACAACCGAGAATCCGCTTACCTCAAAACGCGTTTGCGTATTGGATAGTTTCTCA
CGGACAATGTCATTGCGCCCCTATGCAGAAATTTTGCCGACCGCGGAAGGCGTCGAGCGCCTCGCCGAACTTGTTAGTGTGACAA
TGACAGAACGCGCGGAACCTGTGACAGAGAATACAGCTGTAAACAGTATCCCCCCGGCTAACGAGAACGGGCAGAACTTCGCATA
TGCAGGCGATGGGCCCTCGACTACTGAAAAAGTTGACGGCTCGCATACAGACTTCGATGAAGCATCGAGCGACTACGCCGGCCCT
GTCCCGCTCGCGCAAACTAGATTGAAGCATTCGGATGAATTTCTTCAGCACTTCCGAGTTTTAGACGATTTGGTGGAGGGGGCTT
ACGGGTTTATCTGCGACGTCCGTCGCTACACCGAGGAAGAGCAACGTCGAAGAGGGGTTAACAGTACTAACCAGGGGAAATCAAA
ATGTAAGCGCCTGATAGCTAAATATGTGAAAAATGGAACAAGGGCGGCCTCTCAGCTGGAAAATGAAATTTTGGTTCTCGGGCGC
CTAAATCACGAGAATGTTCTCAAGATCCAGGAAATCCTTCGGTACCCGGATAATACGTACATGTTAACGCAGAGGTATCAGTTCG
ACTTGTACAGCTACATGTACGATGAAGCGTTCGACTGGAAAGACAGTCCAATGCTTAAACAGACTAGACGCATCATGAAGCAGCT
CATGTCAGCGGTCTCGTATATCCATTCAAAGAAACTGATTCACAGGGACATCAAACTCGAAAATATTTTCTTAAACTGCGACGGC
AAGACAGTGCTGGGCGACTTTGGAACTGTCACGCCTTTTGAAAATGAGCGGGAGCCCTTCGAATATGGATGGGTGGGGACCGTGG
CTACTAACTCTCCCGAGATACTCGCCAGGGATTCGTACTGTGAAATTACAGACATTTGGAGCTGCGGAGTAGTATTGCTGGAAAT
GGTAAGCCATGAATTTTGCCCGATCGGCGATGGCGGGGGAAATCCGCACCAGCAATTGCTGAAAGTTATCGACTCTCTCTCAGTT
TGTGATGAAGAGTTCCCAGACCCCCCGTGTAATCTGTACAATTATTTGCATTATGCGAGCATCGATCGCGCCGGACATACGGTCC
CGTCGCTCATACGGAACCTCCACCTTCCGGCGGATGTGGAATACCCTCTAGTTAAAATGCTTACTTTTGACTGGCGTTTGAGACC
CAGCGCGGCCGAAGTATTGGCAATGCCACTGTTTTCGGCTGAAGAGGAACGGACCATAACAATTATTCATGGAAAACATAAACCC
ATCCGACCCGAAATCCGTGCGCGGGTGCCACGGTCCATGAGTGAAGGTTAATAATAAAGGACGGAGATAGAGAACTGAAGCGTCA
GATTTTTTTAAAAAAATAAATGATCG

>SORF1

CTAACCAGGGGAAATCAAAATGTAAGCGCCTGATAGCTAAATATGTGAAAAATGGAACAAGGGCGGCCTCTCAGCTGGAAAATGA
AATTTTGGTTCTCGGGCGCCTAAATCACGAGAATGTTCTCAAGATCCAGGAAATCCTTCGGTACCCGGATAATACGTACATGTTA
ACGCAGAGGTATCAGTTCGACTTGTACAGCTACATGTACGATGAAGCGTTCGACTGGAAAGACAGTCCAATGCTTAAACAGACTA
GACGCATCATGAAGCAGCTCATGTCAGCGGTCTCGTATATCCATTCAAAGAAACTGATTCACAGGGACATCAAACTCGAAAATAT
TTTCTTAAACTGCGACGGCAAGACAGTGCTGGGCGACTTTGGAACTGTCACGCCTTTTGAAAATGAGCGGGAGCCCTTCGAATAT
GGATGGGTGGGGACCGTGGCTACTAACTCTCCCGAGATACTCGCCAGGGATTCGTACTGTGAAATTACAGACATTTGGAGCTGCG
GAGTAGTATTGCTGGAAATGGTAAGCCATGAATTTTGCCCGATCGGCGATGGCGGGGGAAATCCGCACCAGCAATTGCTGAAAGT
TATCGACTCTCTCTCAGTTTGTGATGAAGAGTTCCCAGACCCCCCGTGTAATCTGTACAATTATTTGCATTATGCGAGCATCGAT
CGCGCCGGACATACGGTCCCGTCGCTCATACGGAACCTCCACCTTCCGGCGGATGTGGAATACCCTCTAGTTAAAATGCTTACTT
TTGACTGGCGTTTGAGACCCAGCGCGGCCGAAGTATTGGCAATGCCACTGTTTTCGGCTGAAGAGGAACGGACCATAACAATTAT
TCATGGAAAACATAAACCCATCCGACCCGAAATCCGTGCGCGGGTGCCACGGTCCATGAGTGAAGGTTAATAATAAAGGACGGAG
ATAGAGAACTGAAGCGTCAGATTTTTTTAAAAAAATAAATGATCGAGAACTTATGATTTGTCTTTCTTGAATGACCTTGCCCCAT
CGATTAACGAAAAGACCTTTCGCGCGTCGATTCTGCTCGGTCTTTGTGATACATTATAGTGAGACTAAACTCGACCGATATAACA
AGACAATGTTACTCTATAGACCGGACTCAACCATGCGGCATAGCGGAGGCGACGCAAATCACAGAGGGATAAGGCCGAGGCGGAA
ATCTATTGGAGCGTTTAGCGCGCGCGAAAAGACTGGAAAACGAAATGCGCTGACGGAAAGCAGCTCCTCCTCCGACATGCTAGAT

CCGTTTTCCACGGATAAGGAATTTGGCGGTAAGTGGACGGTAGACGGACCTGCCGACATTACTGCCGAGGTCCTTTCTCAGGCAT
GGGACGTTCTCCAATTAGTGAAGCATGAAGATGCGGAGGAGGAGAGAGTGACTTATGAGTCCAAACCGACCCCGATACAGCCGTT
CAATGCCTGGCCGGACGGGCCGAGTTGGAACGCGCAGGATTTTACTCGAGCGCCAATAGTTTATCCCTCTGCGGAGGTATTGGAC
GCAGAGGCGTTGAAAGTAGGGGCATTCGTTAGCCGAGTTTTACAATGTGTACCGTTCACGCGATCAAAGAAAAGCGTTACGGTGC
GGGATGCGCAGTCGTTTTTGGGGGACTCGTTCTGGAGAATAATGCAGAACGTTTACACGGTTGTCTTACGACAGCACATAACTCG
ACTCAGGCACCCTTCCAGCAAAAGCATTGTTAACTGCAACGACCCTCTATGGTACGCCTACGCGAATCAATTTCACTGGAGAGGA
ATGCGCGTGCCGTCGCTTAAATTAGCCTCTCCCCCGGAGGAGAATATTCAACACGGCCCAATGGCCGCCGTTTTTAGAAACGCGG
GGGCTGGTCTGTTCCTGTGGCCTGCCATGCGCGCAGCCTTTGAAGAGCGCGACAAGCGACTGTTAAGAGCATGCCTGTCTTCACT
CGATATCATGGACGCAGCCGTCCTCGCGTCGTTTCCATTTTACTGGCGCGGCGTCCAAGACACCTCGCGCTTCGAGCCTGCGCTG
GGCTGTTTGTCAGAGTACTTTGCACTAGTGGTGTTACTGGCCGAGACGGTCTTAGCGACCATGTTCGACCACGCACTGGTATTCA
TGAGGGCGCTGGCAGACGGCAATTTCGATGACTATGACGAAACTAGATATATAGACCCCGTTAAAAACGAGTACCTGAACGGAGC
CGAGGGTACTCTGTTACGGGGCATAGTGGCCTCCAACACCGCTCTGGCGGTGGTTTGCGCAAACACCTATTCGACGATAAGAAAA
CTCCCGTCCGTGGCAACTAGCGCGTGCAATGTTGCCTACAGGACCGAAACGCTGAAAGCGAGGCGCCCTGGCATGAGCGACATAT
ACCGGATATTACAAAAAGAGTTTTTCTTTTACATTGCGTGGCTCCAGAGGGTTGCAACACACGCAAATTTCTGTTTAAACATTCT
GAAGAGAAGCGTGGATACGGGGGCCCCGCCATTTTTGTTCAGGGCCAGCTCGGAGAAGCGGCTGCAGCAGTTAAATAAAATGCTC
TGCCCCCTTCTCGTGCCGATTCAATATGAAGACTTTTCGAAGGCCATGGGGTCTGAGCTCAAGAGGGAAAAGTTAGAGACATTCG
TTAAAGCTATTTCCAGCGACAGGGACCCGAGGGGGTCCTTAAGATTTCTCATTTCGGACCATGCAAGGGAAATTATTGCAGACGG
AGTACGGTTTAAGCCGGTGATAGACGAGCCGGTTCGGGCTTCAGTTGCGCTGAGTACCGCTGCCGCTGGGAAAGTGAAAGCGCGA
CGCTTAACCTCAGTTCGCGCGCCCGTACCGGGCGCAGGCGCCGTTTCCGCGCGCCGGAAATCGGAAATATGATAAAAATGCTTGG
CATTTGCGGGCGAAGAGGCGTGATCTGAAGGGCTCCACAATGACGTAACTGAGCTACGCATCCCTATAAAGTGTACCCGCTGACC
GCTAGCCCATACAGTGTTACAGGAGGGGAGAGAGACAACTTCAGCTCGAAGTCTGAAGAGACATCATGAGCGGCTTCAGTAACAT
AGGATCGATTGCCACCGTTTCCCTAGTATGCTCGCTTTTGTGCGCATCTGTATTAGGGGCGCCGGTACTGGACGGGCTCGAGTCG
AGCCCTTTCCCGTTCGGGGGCAAAATTATAGCCCAGGCGTGCAACCGCACCACGATTGAGGTGACGGTCCCGTGGAGCGACTACT
CTGGTCGCACCGAAGGAGTGTCAGTCGAGGTGAAATGGTTCTACGGGAATAGTAATCCCGAAAGCTTCGTGTTCGGGGTGGATAG
CGAAACGGGCAGTGGACACGAGGACCTGTCTACGTGCTGGGCTCTAATCCATAATCTGAACGCGTCTGTGTGCAGGGCGTCTGAC
GCCGGGATACCTGATTTCGACAAGCAGTGCGAAAAAGTGCAGAGAAGACTGCGCTCCGGGGTGGAACTTGGTAGTTACGTGTCTG
GCAATGGATCCCTGGTGCTGTACCCAGGGATGTACGATGCCGGCATCTACGCCTACCAGCTCTCAGTGGGTGGGAAGGGATATAC
CGGGTCTGTTTATCTAGACGTCGGACCAAACCCCGGATGCCACGACCAGTATGGGTACACCTATTACAGCCTGGCCGACGAGGCG
TCAGACTTATCATCTTATGACGTAGCCTCGCCCGAACTCGACGGTCCTATGGAGGAAGATTATTCCAATTGTCTAGACATGCCCC
CGCTACGCCCATGGACAACCGTTTGTTCGCATGACGTCGAGGAGCAGGAAAACGCCACGGACGAGCTTTACCTATGGGACGAGGA
ATGCGCCGGTCCGCTGGACGAGTACGTCGACGAAAGGTCAGAGACGATGCCCAGGATGGTTGTCTTTTCACCGCCCTCTACGCTC
CAGCAGTAGCCACCCGAGAGTGTTTTTTGTGAGCGCCCACGCAACATACCTAACTGCTTCATTTCTGATCAATTATTGCGTATTG
AATAAATAAACAGTACAAAAGCATCAGGTGTGGTTTGCGTGTCTGTGCTAAACCA

>US4

TCACTCGATATCATGGACGCAGCCGTCCTCGCGTCGTTTCCATTTTACTGGCGCGGCGTCCAAGACACCTCGCGCTTCGAGCCTG
CGCTGGGCTGTTTGTCAGAGTACTTTGCACTAGTGGTGTTACTGGCCGAGACGGTCTTAGCGACCATGTTCGACCACGCACTGGT
ATTCATGAGGGCGCTGGCAGACGGCAATTTCGATGACTATGACGAAACTAGATATATAGACCCCGTTAAAAACGAGTACCTGAAC
GGAGCCGAGGGTACTCTGTTACGGGGCATAGTGGCCTCCAACACCGCTCTGGCGGTGGTTTGCGCAAACACCTATTCGACGATAA
GAAAACTCCCGTCCGTGGCAACTAGCGCGTGCAATGTTGCCTACAGGACCGAAACGCTGAAAGCGAGGCGCCCTGGCATGAGCGA
CATATACCGGATATTACAAAAAGAGTTTTTCTTTTACATTGCGTGGCTCCAGAGGGTTGCAACACACGCAAATTTCTGTTTAAAC
ATTCTGAAGAGAAGCGTGGATACGGGGGCCCCGCCATTTTTGTTCAGGGCCAGCTCGGAGAAGCGGCTGCAGCAGTTAAATAAAA
TGCTCTGCCCCCTTCTCGTGCCGATTCAATATGAAGACTTTTCGAAGGCCATGGGGTCTGAGCTCAAGAGGGAAAAGTTAGAGAC
ATTCGTTAAAGCTATTTCCAGCGACAGGGACCCGAGGGGGTCCTTAAGATTTCTCATTTCGGACCATGCAAGGGAAATTATTGCA
GACGGAGTACGGTTTAAGCCGGTGATAGACGAGCCGGTTCGGGCTTCAGTTGCGCTGAGTACCGCTGCCGCTGGGAAAGTGAAAG
CGCGACGCTTAACCTCAGTTCGCGCGCCCGTACCGGGCGCAGGCGCCGTTTCCGCGCGCCGGAAATCGGAAATATGATAAAAATG
CTTGGCATTTGCGGGCGAAGAGGCGTGATCTGAAGGGCTCCACAATGACGTAACTGAGCTACGCATCCCTATAAAGTGTACCCGC
TGACCGCTAGCCCATACAGTGTTACAGGAGGGGAGAGAGACAACTTCAGCTCGAAGTCTGAAGAGACATCATGAGCGGCTTCAGT
AACATAGGATCGATTGCCACCGTTTCCCTAGTATGCTCGCTTTTGTGCGCATCTGTATTAGGGGCGCCGGTACTGGACGGGCTCG
AGTCGAGCCCTTTCCCGTTCGGGGGCAAAATTATAGCCCAGGCGTGCAACCGCACCACGATTGAGGTGACGGTCCCGTGGAGCGA
CTACTCTGGTCGCACCGAAGGAGTGTCAGTCGAGGTGAAATGGTTCTACGGGAATAGTAATCCCGAAAGCTTCGTGTTCGGGGTG
GATAGCGAAACGGGCAGTGGACACGAGGACCTGTCTACGTGCTGGGCTCTAATCCATAATCTGAACGCGTCTGTGTGCAGGGCGT
CTGACGCCGGGATACCTGATTTCGACAAGCAGTGCGAAAAAGTGCAGAGAAGACTGCGCTCCGGGGTGGAACTTGGTAGTTACGT
GTCTGGCAATGGATCCCTGGTGCTGTACCCAGGGATGTACGATGCCGGCATCTACGCCTACCAGCTCTCAGTGGGTGGGAAGGGA
TATACCGGGTCTGTTTATCTAGACGTCGGACCAAACCCCGGATGCCACGACCAGTATGGGTACACCTATTACAGCCTGGCCGACG
AGGCGTCAGACTTATCATCTTATGACGTAGCCTCGCCCGAACTCGACGGTCCTATGGAGGAAGATTATTCCAATTGTCTAGACAT
GCCCCCGCTACGCCCATGGACAACCGTTTGTTCGCATGACGTCGAGGAGCAGGAAAACGCCACGGACGAGCTTTACCTATGGGAC
GAGGAATGCGCCGGTCCGCTGGACGAGTACGTCGACGAAAGGTCAGAGACGATGCCCAGGATGGTTGTCTTTTCACCGCCCTCTA
CGCTCCAGCAGTAGCCACCCGAGAGTGTTTTTTGTGAGCGCCCACGCAACATACCTAACTGCTTCATTTCTGATCAATTATTGCG
TATTGAATAAATAAACAGTACAAAAGCATCAGGTGTGGTTTGCGTGTCTGTGCTAAACCAT

>US5

TGGACGGGCTCGAGTCGAGCCCTTTCCCGTTCGGGGGCAAAATTATAGCCCAGGCGTGCAACCGCACCACGATTGAGGTGACGGT
CCCGTGGAGCGACTACTCTGGTCGCACCGAAGGAGTGTCAGTCGAGGTGAAATGGTTCTACGGGAATAGTAATCCCGAAAGCTTC
GTGTTCGGGGTGGATAGCGAAACGGGCAGTGGACACGAGGACCTGTCTACGTGCTGGGCTCTAATCCATAATCTGAACGCGTCTG
TGTGCAGGGCGTCTGACGCCGGGATACCTGATTTCGACAAGCAGTGCGAAAAAGTGCAGAGAAGACTGCGCTCCGGGGTGGAACT
TGGTAGTTACGTGTCTGGCAATGGATCCCTGGTGCTGTACCCAGGGATGTACGATGCCGGCATCTACGCCTACCAGCTCTCAGTG
GGTGGGAAGGGATATACCGGGTCTGTTTATCTAGACGTCGGACCAAACCCCGGATGCCACGACCAGTATGGGTACACCTATTACA
GCCTGGCCGACGAGGCGTCAGACTTATCATCTTATGACGTAGCCTCGCCCGAACTCGACGGTCCTATGGAGGAAGATTATTCCAA
TTGTCTAGACATGCCCCCGCTACGCCCATGGACAACCGTTTGTTCGCATGACGTCGAGGAGCAGGAAAACGCCACGGACGAGCTT
TACCTATGGGACGAGGAATGCGCCGGTCCGCTGGACGAGTACGTCGACGAAAGGTCAGAGACGATGCCCAGGATGGTTGTCTTTT
CACCGCCCTCTACGCTCCAGCAGTAGCCACCCGAGAGTGTTTTTTGTGAGCGCCCACGCAACATACCTAACTGCTTCATTTCTGA
TCAATTATTGCGTATTGAATAAATAAACAGTACAAAAGCATCAGGTGTGGTTTGCGTGTCTGTGCTAAACCATGGCGTGTGCGGG
TGAAACCGTAAATTACGTGATAATAAATAGCATAGGAGTTGGCGTGCAGCGTATTTCGCCGAGAGATGGGGACAATGTTAGTGTT

GCGCCTTTTCCTACTTGCAGTAGCGGACGCGGCGTTGCCGACCGGCAGATTCTGCCGAGTTTGGAAGGTGCCTCCGGGAGGAACC
ATCCAAGAGAACCTGGCGGTGCTCGCGGAATCGCCGGTCACGGGACACGCGACATATCCGCCGCCTGAAGGCGCCGTCAGCTTTC
AGATTTTTGCGGACACCCCTACTTTGCGCATTCGCTACGGCGCTACGGAGGACGAACTTGCACTGGAGCGCGGGACGTCCGCCTC
AGACGCGGACAACGTGACATTTTCGCTGTCATATCGCCCGCGCCCAGAAATTCACGGAGCATACTTCACCATAGGGGTATTCGCT
ACTGGCCAGAGCACGGAAAGCAGCTATTCGGTCATCAGTCGGGTCTTAGTTAACGCCTCTCTGGAACGGTCCGTGCGCCTGGAAA
CGCCGTGCGATGAAAATTTTTTGCAGAACGAGCCTACATGGGGCTCGAAGCGTTGGTTAGGCCCCCCGTCGCCTTATGTGCGAGA
TAACGATGTCGCCGTGTTGACAAAAGCGCAGTACATTGGGGAGTGCTACTCCAACTCGGCGGCCCAGACGGGGCTCACGTCTCTC
AACATGACCTTTTTCTATTCGCCTAAAAGAATAGTAAACGTCACGTGGACAACCGGCGGCCCCTCCCCCTCGCGCATAACGGTAT
ACTCGTCGCGGGAGAACGGGCAGCCCGTGTTGAGGAACGTTTCTGACGGGTTCTTGGTTAAGTACACTCCCGACATTGACGGCCG
GGCCATGATAAACGTTATTGCCAATTATTCGCCGGCGGACTCCGGCAGCGTCCTCGCGTTTACGGCCTTTAGGGAAGGAAAACTC
CCATCCGCGATTCAACTGCACCGGATAGATATGTCCGGGACTGAGCCGCCGGGGACTGAAACGACCTTCGACTGTCAAAAAATGA
TAGAAACCCCGTACCGAGCGCTCGGGAGCAATGTTCCCAGGGACGACTCTATCCGTCCGGGGGCCACTCTGCCTCCGTTCGATAC
CGCAGCACCTGATTTCGATACAGGTACTTCCCCGACCCCCACTACCGTGCCAGAGCCAGCCATTACTACACTCATACCGCGCAGC
ACTAGCGATATGGGATTCTTCTCCACGGCACGTGCTACCGGATCAGAAACTCTTTCGGTACCCGTCCAGGAAACGGATAGAACTC
TTTCGACAACTCCTCTTACCCTTCCACTGACTCCCGGTGAGTCAGAAAATACACTGTTTCCTACGACCGCGCCGGGGATTTCTAC
CGAGACCCCGAGCGCGGCACATGAAACTACACAGACCCAGAGTGCAGAAACGGTGGTCTTTACTCAGAGTCCGAGTACCGAGTCG
GAAACCGCGCGGTCCCAGAGTCAGGAACCGTGGTATTTTACTCAGACTCCGAGTACTGAACAGGCGGCTCTTACTCAGACGCAGA
TCGCAGAAACGGAGGCGTTGTTTACTCAGACTCCGAGTGCTGAACAGATGACTTTTACTCAGACTCCGGGTGCAGAAACCGAGGC
ACCTGCCCAGACCCCGAGCACGATACCCGAGATATTTACTCAGTCTCGTAGCACGCCCCCCGAAACCGCTCGCGCTCCGAGCGCG
GCGCCGGAGGTTTTTACACAGAGTTCGAGTACGGTAACGGAGGTGTTTACTCAGACCCCGAGCACGGTACCGAAAACTACTCTGA
GTTCGAGTACTGAACCGGCGATTTTTACTCGGACTCAGAGCGCGGGAACTGAGGCCTTTACTCAGACTTCGAGTGCCGAGCCGGA
CACTATGCGAACTCAGAGTACTGAAACACACTTTTTCACTCAGGCCCCGAGTACGGTACCGAAAGCTACTCAGACTCCGAGTACA
GAGCCGGAGGTGTTGACTCAGAGTCCGAGTACCGAACCTGTGCCTTTCACCCGGACTCTGGGCGCAGAGCCGGAAATTACTCAGA
CCCCGAGCGCGGCACCGGAGGTTTATACTCGGAGTTCGAGTACGATGCCAGAAACTGCACAGAGCACACCCCTGGCCTCGCAAAA
CCCTACCAGTTCGGGAACCGGGACGCATAATACTGAACCGAGGACTTATCCAGTGCAAACGACACCACATACCCAGAAACTCTAC
ACAGAAAATAAGACTTTATCGTTTCCTACTGTTGTTTCAGAATTCCATGAGATGTCGACGGCAGAGTCGCAGACGCCCCTATTGG
ACGTCAAAATTGTAGAGGTGAAGTTTTCAAACGATGGCGAAGTAACGGCGACTTGCGTTTCCACCGTCAAATCTCCCTATAGGGT
AGAAACTAATTGGAAAGTAGACCTCGTAGATGTAATGGATGAAATTTCTGGGGAACAGTCCCGCCGGGGTTTTTAACAGTAATGAG
AAATGGCAGAAACAGCTGTACTACAGAGTAACCGATGGAAGAACATCGGTCCAGCTAATGTGCCTGTCGTGCACGAGCCATTCTC
CGGAACCTTACTGTCTTTTCGACACGTCTCTTATAGCGAGGGAAAAAGATATCGCGCCAGAGTTATACTTTACCTCTGATCCGCA
AACGGCATACTGCACAATAACTCTGCCGTCCGGCGTTGTTCCGAGATTCGAATGGAGCCTTAATAATGTTTCACTGCCGGAATAT
TTGACGGCCACGACCGTTGTTTCGCATACCGCTGGCCAAAGTACAGTGTGGAAGAGCAGCGCGAGAGCAGGCGAGGCGTGGATTT
CTGGCCGGGGAGGCAATATATACGAATGCACCGTCCTCATCTCAGACGGCACTCGCGTTACTACGCGAAAGGAGAGGTGCTTAAC
AAACACATGGATTGCGGTGGAAAACGGTGCTGCTCAGGCGCAGCTGTATTCACTCTTTTCTGGACTTGTGTCAGGATTATGCGGG
AGCATATCTGCTTTGTACGCAACGCTATGGACCGCCATTTATTTTTGAGGAATGCTTTTTGGACTATCGTACTGCTTTCTTCCTT
CGCTAGCCAGAGCACCGCCGCCGTCACGTACGACTACATTTTAGGCCGTCGCGCGCTCGACGCGCTAACCATACCGGCGGTTGGC
CCGTATAACAGATACCTCACTAGGGTATCAAGAGGCTGCGACGTTGTCGAGCTCAACCCGATTTCTAACGTGGACGACATGATAT
CGGCGGCCAAAGAAAAAGAGAAGGGGGGCCCTTTCGAGGCCTCCGTCGTCTGGTTCTACGTGATTAAGGGCGACGACGGCGAGGA
CAAGTACTGTCCAATCTATAGAAAAGAGTACAGGGAATGTGGCGACGTACAACTGCTATCTGAATGCGCCGTTCAATCTGCACAG
ATGTGGGCAGTGGACTATGTTCCTAGCACCCTTGTATCGCGAAATGGCGCGGGACTGACTATATTCTCCCCCACTGCTGCGCTCT
CTGGCCAATACTTGCTGACCCTGAAAATCGGGAGATTTGCGCAAACAGCTCTCGTAACTCTAGAAGTTAACGATCGCTGTTTAAA
GATCGGGTCGCAGCTTAACTTTTACCGTCGAAATGCTGGACAACAGAACAGTATCAGACTGGATTTCAAGGCGAACACCTTTAT
CCGATCGCAGACACCAATACACGACACGCGGACGACGTATATCGGGGATACGAAGATATTCTGCAGCGCTGGAATAATTTGCTGA
GGAAAAAGAATCCTAGCGCGCCAGACCCTCGTCCAGATAGCGTCCCGCAAGAAATTCCCGCTGTAACCAAGAAAGCGGAAGGGCG
CACCCCGGACGCAGAAAGCAGCGAAAAGAAGGCCCCTCCAGAAGACTCGGAG

>US6

CCTTTCACCCGGACTCTGGGCGCAGAGCCGGAAATTACTCAGACCCCGAGCGCGGCACCGGAGGTTTATACTCGGAGTTCGAGTA
CGATGCCAGAAACTGCACAGAGCACACCCCTGGCCTCGCAAAACCCTACCAGTTCGGGAACCGGGACGCATAATACTGAACCGAG
GACTTATCCAGTGCAAACGACACCACATACCCAGAAACTCTACACAGAAAATAAGACTTTATCGTTTCCTACTGTTGTTTCAGAA
TTCCATGAGATGTCGACGGCAGAGTCGCAGACGCCCCTATTGGACGTCAAAATTGTAGAGGTGAAGTTTTCAAACGATGGCGAAG
TAACGGCGACTTGCGTTTCCACCGTCAAATCTCCCTATAGGGTAGAAACTAATTGGAAAGTAGACCTCGTAGATGTAATGGATGA
AATTTCTGGGAACAGTCCCGCCGGGGTTTTTAACAGTAATGAGAAATGGCAGAAACAGCTGTACTACAGAGTAACCGATGGAAGA
ACATCGGTCCAGCTAATGTGCCTGTCGTGCACGAGCCATTCTCCGGAACCTTACTGTCTTTTCGACACGTCTCTTATAGCGAGGG
AAAAAGATATCGCGCCAGAGTTATACTTTACCTCTGATCCGCAAACGGCATACTGCACAATAACTCTGCCGTCCGGCGTTGTTCC
GAGATTCGAATGGAGCCTTAATAATGTTTCACTGCCGGAATATTTGACGGCCACGACCGTTGTTTCGCATACCGCTGGCCAAAGT
ACAGTGTGGAAGAGCAGCGCGAGAGCAGGCGAGGCGTGGATTTCTGGCCGGGGAGGCAATATATACGAATGCACCGTCCTCATCT
CAGACGGCACTCGCGTTACTACGCGAAAGGAGAGGTGCTTAACAAACACATGGATTGCGGTGGAAAACGGTGCTGCTCAGGCGCA
GCTGTATTCACTCTTTTCTGGACTTGTGTCAGGATTATGCGGGAGCATATCTGCTTTGTACGCAACGCTATGGACCGCCATTTAT
TTTTGAGGAATGCTTTTTGGACTATCGTACTGCTTTCTTCCTTCGCTAGCCAGAGCACCGCCGCCGTCACGTACGACTACATTTT
AGGCCGTCGCGCGCTCGACGCGCTAACCATACCGGCGGTTGGCCCGTATAACAGATACCTCACTAGGGTATCAAGAGGCTGCGAC
GTTGTCGAGCTCAACCCGATTTCTAACGTGGACGACATGATATCGGCGGCCAAAGAAAAAGAGAAGGGGGGCCCTTTCGAGGCCT
CCGTCGTCTGGTTCTACGTGATTAAGGGCGACGACGGCGAGGACAAGTACTGTCCAATCTATAGAAAAGAGTACAGGGAATGTGG
CGACGTACAACTGCTATCTGAATGCGCCGTTCAATCTGCACAGATGTGGGCAGTGGACTATGTTCCTAGCACCCTTGTATCGCGA
AATGGCGCGGGACTGACTATATTCTCCCCCACTGCTGCGCTCTCTGGCCAATACTTGCTGACCCTGAAAATCGGGAGATTTGCGC
AAACAGCTCTCGTAACTCTAGAAGTTAACGATCGCTGTTTAAAGATCGGGTCGCAGCTTAACTTTTTACCGTCGAAATGCTGGAC
AACAGAACAGTATCAGACTGGATTTCAAGGCGAACACCTTTATCCGATCGCAGACACCAATACACGACACGCGGACGACGTATAT
CGGGGATACGAAGATATTCTGCAGCGCTGGAATAATTTGCTGAGGAAAAAGAATCCTAGCGCGCCAGACCCTCGTCCAGATAGCG
TCCCGCAAGAAATTCCCGCTGTAACCAAGAAAGCGGAAGGGCGCACCCCGGACGCAGAAAGCAGCGAAAAGAAGGCCCCTCCAGA
AGACTCGGAGGACGACATGCAGGCAGAGGCTTCTGGAGAAAATCCTGCCGCCCTCCCCGAAGACGACGAAGTCCCCGAGGACACC
GAGCACGATGATCCAAACTCGGATCCTGACTATTACAATGACATGCCCGCCGTGATCCCGGTGGAGGAGACTACTAAAAGTTCTA
ATGCCGTCTCCATGCCCATATTCGCGGCGTTCGTAGCCTGCGCGGTCGCGCTCGTGGGGCTACTGGTTTGGAGCATCGTAAAATG
CGCGCGTAGCTAATCGAGCCTAGAATAGGTGGTTTCTTCCTACATGCCACGCCTCACGCTCATAATATAAATCACATGGAATAGC
ATACCAATGCCTATTCATTGGGACGTTCGAAAAGCATGGCATCGCTACTTGGAACTCTGGCTCTCCTTGCCGCGACGCTCGCACC

CTTCGGCGCGATGGGAATCGTGATCACTGGAAATCACGTCTCCGCCAGGATTGACGACGATCACATCGTGATCGTCGCGCCTCGC
CCCGAAGCTACAATTCAACTGCAGCTATTTTTCATGCCTGGCCAGAGACCCCACAAACCCTACTCAGGAACCGTCCGCGTCGCGT
TTCGGTCTGATATAACAAACCAGTGCTACCAGGAACTTAGCGAGGAGCGCTTTGAAAATTGCACTCATCGATCGTCTTCTGTTTT
TGTCGGCTGTAAAGTGACCGAGTACACGTTCTCCGCCTCGAACAGACTAACCGGACCTCCACACCCGTTTAAGCTCACTATACGA
AATCCTCGTCCGAACGACAGCGGGATGTTCTACGTAATTGTTCGGCTAGACGACACCAAAGAACCCATTGACGTCTTCGCGATCC
AACTATCGGTGTATCAATTCGCGAACACCGCCGCGACTCGCGGACTCTATTCCAAGGCTTCGTGTCGCACCTTCGGATTACCTAC
CGTCCAACTTGAGGCCTATCTCAGGACCGAGGAAAGTTGGCGCAACTGGCAAGCGTACGTTGCCACGGAGGCCACGACGACCAGC
GCCGAGGCGACAACCCCGACGCCCGTCACTGCAACCAGCGCCTCCGAACTTGAAGCGGAACACTTTACCTTTCCCTGGCTAGAAA
ATGGCGTGGATCATTACGAACCGACACCCGCAAACGAAAATTCAAACGTTACTGTCCGTCTCGGGACAATGAGCCCTACGCTAAT
TGGGGTAACCGTGGCTGCCGTCGTGAGCGCAACGATCGGCCTCGTCATTGTAATTTCCATCGTCACCAGAAACATGTGCACCCCG
CACCGAAAATTAGACACGGTCTCGCAAGACGACGAAGAACGTTCCCAAACTAGAAGGGAATCGCGAAAATTTGGACCCATGGTTG
CGTGCGAAATAAACAAGGGGGCTGACCAGGATAGTGAACTTGTGGAACTGGTTGCGATTGTTAACCCGTCTGCGCTAAGCTCGCC
CGACTCAATAAAAATGTGATTAAGTCTGAATGTGGCTCTCCAATCATTTCGATTCTCTAATCTCC

>US7

CAAAGAAAAAGAGAAGGGGGGCCCTTTCGAGGCCTCCGTCGTCTGGTTCTACGTGATTAAGGGCGACGACGGCGAGGACAAGTAC
TGTCCAATCTATAGAAAAGAGTACAGGGAATGTGGCGACGTACAACTGCTATCTGAATGCGCCGTTCAATCTGCACAGATGTGGG
CAGTGGACTATGTTCCTAGCACCCTTGTATCGCGAAATGGCGCGGGACTGACTATATTCTCCCCCACTGCTGCGCTCTCTGGCCA
ATACTTGCTGACCCTGAAAATCGGGAGATTTGCGCAAACAGCTCTCGTAACTCTAGAAGTTAACGATCGCTGTTTAAAGATCGGG
TCGCAGCTTAACTTTTTACCGTCGAAATGCTGGACAACAGAACAGTATCAGACTGGATTTCAAGGCGAACACCTTTATCCGATCG
CAGACACCAATACACGACACGCGGACGACGTATATCGGGGATACGAAGATATTCTGCAGCGCTGGAATAATTTGCTGAGGAAAAA
GAATCCTAGCGCGCCAGACCCTCGTCCAGATAGCGTCCCGCAAGAAATTCCCGCTGTAACCAAGAAAGCGGAAGGGCGCACCCCG
GACGCAGAAAGCAGCGAAAAGAAGGCCCCTCCAGAAGACTCGGAGGACGACATGCAGGCAGAGGCTTCTGGAGAAAATCCTGCCG
CCCTCCCCGAAGACGACGAAGTCCCCGAGGACACCGAGCACGATGATCCAAACTCGGATCCTGACTATTACAATGACATGCCCGC
CGTGATCCCGGTGGAGGAGACTACTAAAAGTTCTAATGCCGTCTCCATGCCCATATTCGCGGCGTTCGTAGCCTGCGCGGTCGCG
CTCGTGGGGCTACTGGTTTGGAGCATCGTAAAATGCGCGCGTAGCTAATCGAGCCTAGAATAGGTGGTTTCTTCCTACATGCCAC
GCCTCACGCTCATAATATAAATCACATGGAATAGCATACCAATGCCTATTCATTGGGACGTTCGAAAAGCATGGCATCGCTACTT
GGAACTCTGGCTCTCCTTGCCGCGACGCTCGCACCCTTCGGCGCGATGGGAATCGTGATCACTGGAAATCACGTCTCCGCCAGGA
TTGACGACGATCACATCGTGATCGTCGCGCCTCGCCCCGAAGCTACAATTCAACTGCAGCTATTTTTCATGCCTGGCCAGAGACC
CCACAAACCCTACTCAGGAACCGTCCGCGTCGCGTTTCGGTCTGATATAACAAACCAGTGCTACCAGGAACTTAGCGAGGAGCGC
TTTGAAAATTGCACTCATCGATCGTCTTCTGTTTTTGTCGGCTGTAAAGTGACCGAGTACACGTTCTCCGCCTCGAACAGACTAA
CCGGACCTCCACACCCGTTTAAGCTCACTATACGAAATCCTCGTCCGAACGACAGCGGGATGTTCTACGTAATTGTTCGGCTAGA
CGACACCAAAGAACCCATTGACGTCTTCGCGATCCAACTATCGGTGTATCAATTCGCGAACACCGCCGCGACTCGCGGACTCTAT
TCCAAGGCTTCGTGTCGCACCTTCGGATTACCTACCGTCCAACTTGAGGCCTATCTCAGGACCGAGGAAAGTTGGCGCAACTGGC
AAGCGTACGTTGCCACGGAGGCCACGACGACCAGCGCCGAGGCGACAACCCCGACGCCCGTCACTGCAACCAGCGCCTCCGAACT
TGAAGCGGAACACTTTACCTTTCCCTGGCTAGAAAATGGCGTGGATCATTACGAACCGACACCCGCAAACGAAAATTCAAACGTT
ACTGTCCGTCTCGGGACAATGAGCCCTACGCTAATTGGGGTAACCGTGGCTGCCGTCGTGAGCGCAACGATCGGCCTCGTCATTG
TAATTTCCATCGTCACCAGAAACATGTGCACCCCGCACCGAAAATTAGACACGGTCTCGCAAGACGACGAAGAACGTTCCCAAAC
TAGAAGGGAATCGCGAAAATTTGGACCCATGGTTGCGTGCGAAATAAACAAGGGGGCTGACCAGGATAGTGAACTTGTGGAACTG
GTTGCGATTGTTAACCCGTCTGCGCTAAGCTCGCCCGACTCAATAAAAATGTGATTAAGTCTGAATGTGGCTCTCCAATCATTTC
GATTCTCTAATCTCC

>US8

ACTCATCGATCGTCTTCTGTTTTTGTCGGCTGTAAAGTGACCGAGTACACGTTCTCCGCCTCGAACAGACTAACCGGACCTCCAC
ACCCGTTTAAGCTCACTATACGAAATCCTCGTCCGAACGACAGCGGGATGTTCTACGTAATTGTTCGGCTAGACGACACCAAAGA
ACCCATTGACGTCTTCGCGATCCAACTATCGGTGTATCAATTCGCGAACACCGCCGCGACTCGCGGACTCTATTCCAAGGCTTCG
TGTCGCACCTTCGGATTACCTACCGTCCAACTTGAGGCCTATCTCAGGACCGAGGAAAGTTGGCGCAACTGGCAAGCGTACGTTG
CCACGGAGGCCACGACGACCAGCGCCGAGGCGACAACCCCGACGCCCGTCACTGCAACCAGCGCCTCCGAACTTGAAGCGGAACA
CTTTACCTTTCCCTGGCTAGAAAATGGCGTGGATCATTACGAACCGACACCCGCAAACGAAAATTCAAACGTTACTGTCCGTCTC
GGGACAATGAGCCCTACGCTAATTGGGGTAACCGTGGCTGCCGTCGTGAGCGCAACGATCGGCCTCGTCATTGTAATTTCCATCG
TCACCAGAAACATGTGCACCCCGCACCGAAAATTAGACACGGTCTCGCAAGACGACGAAGAACGTTCCCAAACTAGAAGGGAATC
GCGAAAATTTGGACCCATGGTTGCGTGCGAAATAAACAAGGGGGCTGACCAGGATAGTGAACTTGTGGAACTGGTTGCGATTGTT
AACCCGTCTGCGCTAAGCTCGCCCGACTCAATAAAAATGTGATTAAGTCTGAATGTGGCTCTCCAATCATTTCGATTCTCTAATC
TCCCAATCCTCTCAAAAGGGGCAGTATCGGACACGGACTGGGAGGGGCGTACACGATAGTTATATGGTACAGCAGAGGCCTCTGA
ACACTTAGGAGGAGAATTCAGCCGGGGAGAGCCCCTGTTGAGTAGGCTTGGGAGCATATTGCAGGATGAACATGTTAGTGATAGT
TCTCGCCTCTTGTCTTGCGCGCCTAACTTTTGCGACGCGACACGTCCTCTTTTTGGAAGGCACTCAGGCTGTCCTCGGGGAAGAT
GATCCCAGAAACGTTCCGGAAGGGACTGTAATCAAATGGACAAAAGTCCTGCGGAACGCGTGCAAGATGAAGGCGGCCGATGTCT
GCTCTTCGCCTAACTATTGCTTTCATGATTTAATTTACGACGGAGGAAAGAAAGACTGCCCGCCCGCGGGACCCCTGTCTGCAAA
CCTGGTAATTTTACTAAAGCGCGGCGAAAGCTTCGTCGTGCTGGGTTCTGGGCTACACAACAGCAATATAACTAATATCATGTGG
ACAGAGTACGGAGGCCTGCTCTTTGATCCTGTAACTCGTTCGGACGAGGGAATCTATTTTCGACGGATCTCTCAGCCAGATCTGG
CCATGGAAACTACATCGTACAACGTCAGCGTTCTTTCGCACGTAGACGAGAAGGCTCCAGCACCGCACGAGGTGGAGATAGACAC
CATCAAGCCGTCAGAGGCCCACGCGCACGTGGAATTACAAATGCTGCCGTTTCATGAACTCAACGACAACAGCCCCACCTATGTG
ACCCCTGTTCTTAGAGTCTTCCCACCGACCGAGCACGTAAAATTTAACGTTACGTATTCGTGGTATGGGTTTGATGTCAAAGAGG
AGTGCGAAGAAGTGAAACTGTTCGAGCCGTGCGTATACCATCCTACAGACGGCAAATGTCAGTTTCCCGCAACCAACCAGAGATG
CCTCATAGGATCTGTCTTGATGGCGGAATTCTTGGGCGCGGCCTCTTTGCTGGATTGTTCCCGCGATACTCTAGAAGACTGCCAC
GAAAATCGCGTGCCGAACCTACGGTTCGATTCGCGACTCTCCGAGTCACGCGCAGGCCTGGTGATCAGTCCTCTTATAGCCATCC
CCAAAGTTTTGATTATAGTCGTTTCCGACGGAGACATTTTGGGATGGAGCTACACGGTGCTCGGGAAACGTAACAGTCCGCGCGT
AGTAGTCGAAACGCACATGCCCTCGAAGGTCCCGATGAACAAAGTAGTAATTGGCAGTCCCGGACCAATGGACGAAACGGGTAAC
TATAAAATGTACTTCGTCGTCGCGGGGGTGGCCGCGACGTGCGTAATTCTTACATGCGCTCTGCTTGTGGGGAAAAAGAAGTGCC
CCGCGCACCAAATGGGTACTTTTTCCAAGACCGAACCATTGTACGCGCCGCTCCCCAAAAACGAGTTTGAGGCCGGCGGGCTTAC
GGACGATGAGGAAGTGATTTATGACGAAGTATACGAACCCCTATTTCGCGGCTACTGTAAGCAGGAATTCCGCGAAGATGTGAAT
ACCTTTTTCGGTGCGGTCGTGGAGGGAGAAAGGGCCTTAAACTTTAAATCCGCCATCGCATCAATGGCAGATCGCATCCTGGCAA
ATAAAAGCGGCAGAAGGAATATGGATAGCTATTAGTTGGTCATGCCTTTTAAGACCAGAGGGGCCGAAGACGCGGCCGCGGGCAA
GAACAGGTTTAAGAAATCGAGAAATCGGGAAATCTTACCGACCAGACTGCGTGGCACCGGTAAGAAAACTGCCGGATTGTCCAAT

TATACCCAGCCTATTCCCTGGAACCCTAAATTCTGCAGCGCGCGCGGGGAATCTGACAACCACGCGTGTAAAGACACTTTTTATC
GCAGGACGTGCTGCGCATCGCGCTCTACCGTTTCCAGTCAACCCGATTCCCCCCACACACCCATGCCTACTGAGTATGGGCGCGT
GCCCTCCGCAAAGCGCAAAAAACTATCATCTTCAGACTGCGAGGGCGCGCACCAACCCCTAGTATCCTGTAAACTTCCGGATTCT
CAAGCAGCACCGGCGCGAACCTATAGTTCTGCGCAAAGATATACTGTTGACGAGGTTTCGTCGCCAACTCCGCCAGGCGTCGACG
CTGTTGCGGACTTAGAAACGCGCGCGGAACTTCCTGGCGCTACGACGGAACAAACGGAAAGTAAAAATAAGCTCCCCAACCAACA
ATCGCGCCTGAAGCCGAAACCCACAAACGAGCACGTCGGAGGGGAGCGGTGCCCCTCCGAAGGCACGGTCGAGGCGCCATCGCTC
GGCATCCTCTCGCGCGTCGGGGCAGCGATAGCAAACGAGCTGGCTCGTATGCGGAGGGCGTGTCTTCCGCTCGCCGCGTCGGCGG
CCGCTGCCGGAATAGTGGCCTGGGCCGCGGCGAGGGCCTTGCAGAAACAAGGGCGGTAGCAGTAATAATAACCACACAAATATTG
ACAATAATAAACGCGTACGCGGATGAGTAAGTGTTATTGTCTCGCGCGCCATCTTTATAAAAGCCCGC

>US8A

AGCACCGCACGAGGTGGAGATAGACACCATCAAGCCGTCAGAGGCCCACGCGCACGTGGAATTACAAATGCTGCCGTTTCATGAA
CTCAACGACAACAGCCCCACCTATGTGACCCCTGTTCTTAGAGTCTTCCCACCGACCGAGCACGTAAAATTTAACGTTACGTATT
CGTGGTATGGGTTTGATGTCAAAGAGGAGTGCGAAGAAGTGAAACTGTTCGAGCCGTGCGTATACCATCCTACAGACGGCAAATG
TCAGTTTCCCGCAACCAACCAGAGATGCCTCATAGGATCTGTCTTGATGGCGGAATTCTTGGGCGCGGCCTCTTTGCTGGATTGT
TCCCGCGATACTCTAGAAGACTGCCACGAAAATCGCGTGCCGAACCTACGGTTCGATTCGCGACTCTCCGAGTCACGCGCAGGCC
TGGTGATCAGTCCTCTTATAGCCATCCCCAAAGTTTTGATTATAGTCGTTTCCGACGGAGACATTTTGGGATGGAGCTACACGGT
GCTCGGGAAACGTAACAGTCCGCGCGTAGTAGTCGAAACGCACATGCCCTCGAAGGTCCCGATGAACAAAGTAGTAATTGGCAGT
CCCGGACCAATGGACGAAACGGGTAACTATAAAATGTACTTCGTCGTCGCGGGGGTGGCCGCGACGTGCGTAATTCTTACATGCG
CTCTGCTTGTGGGGAAAAAGAAGTGCCCCGCGCACCAAATGGGTACTTTTTCCAAGACCGAACCATTGTACGCGCCGCTCCCCAA
AAACGAGTTTGAGGCCGGCGGGCTTACGGACGATGAGGAAGTGATTTATGACGAAGTATACGAACCCCTATTTCGCGGCTACTGT
AAGCAGGAATTCCGCGAAGATGTGAATACCTTTTTCGGTGCGGTCGTGGAGGGAGAAAGGGCCTTAAACTTTAAATCCGCCATCG
CATCAATGGCAGATCGCATCCTGGCAAATAAAAGCGGCAGAAGGAATATGGATAGCTATTAGTTGGTCATGCCTTTTAAGACCAG
AGGGGCCGAAGACGCGGCCGCGGGCAAGAACAGGTTTAAGAAATCGAGAAATCGGGAAATCTTACCGACCAGACTGCGTGGCACC
GGTAAGAAAACTGCCGGATTGTCCAATTATACCCAGCCTATTCCCTGGAACCCTAAATTCTGCAGCGCGCGCGGGGAATCTGACA
ACCACGCGTGTAAAGACACTTTTTATCGCAGGACGTGCTGCGCATCGCGCTCTACCGTTTCCAGTCAACCCGATTCCCCCCACAC
ACCCATGCCTACTGAGTATGGGCGCGTGCCCTCCGCAAAGCGCAAAAAACTATCATCTTCAGACTGCGAGGGCGCGCACCAACCC
CTAGTATCCTGTAAACTTCCGGATTCTCAAGCAGCACCGGCGCGAACCTATAGTTCTGCGCAAAGATATACTGTTGACGAGGTTT
CGTCGCCAACTCCGCCAGGCGTCGACGCTGTTGCGGACTTAGAAACGCGCGCGGAACTTCCTGGCGCTACGACGGAACAAACGGA
AAGTAAAAATAAGCTCCCCAACCAACAATCGCGCCTGAAGCCGAAACCCACAAACGAGCACGTCGGAGGGGAGCGGTGCCCCTCC
GAAGGCACGGTCGAGGCGCCATCGCTCGGCATCCTCTCGCGCGTCGGGGCAGCGATAGCAAACGAGCTGGCTCGTATGCGGAGGG
CGTGTCTTCCGCTCGCCGCGTCGGCGGCCGCTGCCGGAATAGTGGCCTGGGCCGCGGCGAGGGCCTTGCAGAAACAAGGGCGGTA
GCAGTAATAATAACCACACAAATATTGACAATAATAAACGCGTACGCGGATGAGTAAGTGTTATTGTCTCGCGCGCCATCTTTAT
AAAAG

>SORF43S

TTAAATCCGCCATCGCATCAATGGCAGATCGCATCCTGGCAAATAAAAGCGGCAGAAGGAATATGGATAGCTATTAGTTGGTCAT
GCCTTTTAAGACCAGAGGGGCCGAAGACGCGGCCGCGGGCAAGAACAGGTTTAAGAAATCGAGAAATCGGGAAATCTTACCGACC
AGACTGCGTGGCACCGGTAAGAAAACTGCCGGATTGTCCAATTATACCCAGCCTATTCCCTGGAACCCTAAATTCTGCAGCGCGC
GCGGGGAATCTGACAACCACGCGTGTAAAGACACTTTTTATCGCAGGACGTGCTGCGCATCGCGCTCTACCGTTTCCAGTCAACC
CGATTCCCCCCACACACCCATGCCTACTGAGTATGGGCGCGTGCCCTCCGCAAAGCGCAAAAAACTATCATCTTCAGACTGCGAG
GGCGCGCACCAACCCCTAGTATCCTGTAAACTTCCGGATTCTCAAGCAGCACCGGCGCGAACCTATAGTTCTGCGCAAAGATATA
CTGTTGACGAGGTTTCGTCGCCAACTCCGCCAGGCGTCGACGCTGTTGCGGACTTAGAAACGCGCGCGGAACTTCCTGGCGCTAC
GACGGAACAAACGGAAAGTAAAAATAAGCTCCCCAACCAACAATCGCGCCTGAAGCCGAAACCCACAAACGAGCACGTCGGAGGG
GAGCGGTGCCCCTCCGAAGGCACGGTCGAGGCGCCATCGCTCGGCATCCTCTCGCGCGTCGGGGCAGCGATAGCAAACGAGCTGG
CTCGTATGCGGAGGGCGTGTCTTCCGCTCGCCGCGTCGGCGGCCGCTGCCGGAATAGTGGCCTGGGCCGCGGCGAGGGCCTTGCA
GAAACAAGGGCGGTAGCAGTAATAATAACCACACAAATATTGACAATAATAAACGCGTACGCGGATGAGTAAGTGTTATTGTCTC
GCGCGCCATCTTTATAAAAGCCCGCGTTGCGTGGGCCGGCGGGTAGCATTTGGAGGGTTGGCGACCATGTCGAGACCTCCGACGT
CACATTTGGACTTAGCTTTCTCGGCGGCCTTTAGGGGCACGGACCTGCCCGGAGGGAGATTCTGGCGGGCGTCGCAGAGTTGCGA
TATTTTCTTTTGGCCCGATCTGGCCGCGGTGATCGTACAGGCCGCCCGCGCGTATTTTGAAGGGAAGGAAAGGCTGGGCAGTCTG
CAGGTCGCCGAAGATATCACGGCGCACGACCCGCGAATAGCGCCCGCGGCTAAGCGCGCCGTCGCAGCGGCGGTAGGACTGTGGA
CCGCGCTGTCGGAGTTAGTTGGGGGGCCGAACGGGGAGTTGGAAAGCAAGGTCTGGGGCAAGCAGATTCCCCGGGCCGCCGCGTG
GGAAATAAGAGACGTGCCCAAAGTTCCAGTCATTGGGCCGGACATTCTTTCTTTTTTCTCCGCCGCCGTCGAGCTGCCGGTGCTC
TATATCAGAGCCCGGGGAGGGGCGCACTCGCGGTCCGCGCACTGGAATAACCAGAGCAGCGCGCCGGCCGCCGGACTCGCGGCGA
TAAGGATAGGCATGGAGATGGTGCGGAGCCTCCTGGTGATAGCGCTGCCTCTGTCAAACTTCACCCTCCCGGAAGACCTCCCCGA
AGGTTCCCAAAACTCGATCCGCGCGTTCGTGGCCCACCTCATGAACTGTGTAGCTACCGATAAGATCATGTCTCCGGACGTGCGC
GTCCCAGTCGAAGAAAGCTTTTACAGCCACTGTTTAAGGGAAATCATTATGTGCGAGAGAGCTTTTTGTTACCCGTGCAATCCCC
CGCCAAAATGGTGAGCTCAGGGTCCCATTTACCCCCGCAACACCCTCTCGCCGCCAGGGCGCGCGCTCTATCTTTCTCTATGTCC
CGTCGCCACCGGCCTAACCGAACGGTGGAACGGGGCCGCCCGGGAAAGCCTAGCTCCGCACAGACACAGACAGACAAACGGCCTC
TCGATTGCAACAACGTGAAAAACACACAATAATATTACTTTATTTATTTCGCAGCGCTCGCGTGTCGCTCGTTTCTGGGGGGGGG
GGGGGGTTGTGTTTCGCTG

>US10AS

TGGTCTGGAGGCATCGGCGGGGGGATTTTCGGGTTGCTCATGTGAGAAGACAATGACCCCTCCCATTGTGTCCCCAGTATGAAAT
CTTCTCCCCTCTTCCTGGGGTAATCCGAGCAGGACCCGATGGTCCCGCTATATGCGGGGACAGCGACCCGGATCGTTCTTTTATG
CAGTGCCCGAAGAGGCAAGTCACAGCTCATTGGCTGGAAGGGGGGATGCAGGACCCTCGGTGCGAACGTGACGTAGGTGGGTGCC
ACCCACATCATGTAAATTAGGTCACGGGGTAGGGTTCTGGATCCTTTTTGGATCCCGAAGAGCTCTCCCAGAAGTTTTTCTTTTC
GGACGTATCGGAGGACGAAGAACCGGCACGCGGGAGGAGCTGGAGCGACCCGGAGTCGGAGGAAGAGCAGCCTGGGTGCCGGGGA
GTGGACTTGGGCGAGGAGGACACGGGACACAGCTCCACCGAGTCAGAGCCCACGCAATCTGACTTAGACTTTATTGACGACAGCT
CTCCGGCGCCGCCGCCATTTGCTATCCCCCGCGTCCGTGCGTTATTGCGGTGCGCGGCACCCGCAAAGACCCACGGAAGGCTTCG
GCCGCCAGGGCGGGTAGGCGCACTCTTAAAAGACGGAGGTTGTCATTTTCTTCTTCCTCTGACGAGGAATCCGAGGAGAGAAGTA
AAAAAGAAGAAGCGGCCTCGACCCCTGCACGGCGACGCAAGGCCGAGGCCTCGACGAGCAGATAGAGGAGACGCGGGGCAGAACC
TCCCCCTCCCTCCCACCCCCCTACTCTGGACATTTATTGCCCGCTCGATCCATTCTCATCCAGAACTTCTTTCCCGCTCAGCCTT
CACGCAGAAGCGGACGCGCGCCCCTTTGCGACCGCCGGACATCCCGCCGCCCCCCCCCCTTCACGCCCGGCGCAATCCGTAGCCG

TCCAACTCGGCCCAGCACAACCGCAGTAGACCGCCCGGACCGCTCTCCTCTAGACACATCCCTAAATGGAAAACATGCTCGACGG
GTGCTACCCGCTGGCGCTGATGGACAGCGATCACATTACTGCGCACGCGGTACCTCGTGGCGAGCGCAGGCGGCAAGGTGCCGCT
GTCGCCTCGTCGGAGTCGGCCGACTCGGTAGACCCGTGCATTCGGATCGCCTCGCGGCTCTGGCGCGAGTTAGTCGAGATATCGT
CCGAACTCAAGGACGGTTACGGAGAGTTCACGTCAGCGAGAGACCGCCGCAACGCGCTGATTGCTGCCAACGAACGGCTACGTTC
GGCTTTTCTGGGGGCCAGCCGGGCGACGCGCGGCCTAGGTTTGAGGCCGCGGTGGGCGTCGACGGAGAGCGTCGCCAACTCCCCC
ACTGACCCGAATAACGGCAACGGGTTGGGAGAATTAGAGGAGGCAATGGAAGGGATCGAGGGCGATTTCTGGCTCGACTCTCTGG
ACGGTGACCGCTTCGAGGACGAGAGCCGTACCATGCAGAGCGAGAATATGCGTTTCGTGATCGAGAAAGAACTGTTATCCTGGCT
GTCCCGACACCTGCCGGCCGACCTCGCGTCCGCCGAGCGAGAGACCTCCCGGTCTCTCCTGGCGGCCGGGCACTGGTGCTGCTTG
TGGCACCCTCGGCCGTGCCGCGAAGCGTGTTTGTACGACTCGATTTACGTGCAGAGTCTTTTCTGCGTCGGGACGGGGAGAGTCC
CGCAATCGGAGATGCGCCGTCGCGAATACCTGGCCGCCTTGCGCGCCGGCGCGGCTGCCGCCAACTCTCCCGAAGTGAGCGCCTC
GATCTTTGCGAGGGACGCTGGAATCGCGCTGGCGCTGGCGCGGCGCCGTTGACGGGAGAATGACGCCCTCTAGCGGCTTCCTTAC
CTCCGCGTCCCTGACAACCTCGCGGGTTTTTACACTGTCCTCCGTCCACTCTCCCCCCTCACCCACTCCGCGGCAGCGAAAC

>ICP4S

CTTCCTTGGCACTCCCAGGAAGCGTGACCCATGTGCTTTCGTTGGCATTCAAAATAAACTTCACTGCTTCCCGGTGTGGCCAATA
ACGTACAAGCGAGCGTGTTTCTCGGGTAATATCAATGCCAACAAAACACACGCGTGACCTTTCTAGAGTAGGCACACACCCCGGC
ACCGAATTGCACGTACCTGTGGACACAAGTAATGTATTGTTTTGTTATCTTCGTCATCTGTTCTACAGTAGGCTCAGGTTTGTTT
ATTATGTTGGCTCATAAAAAATGCCGGAAGTTCCTGGTACGGTAATGGTATGCTGGGCGCATGGCATGCATTGGCTATGGCACAC
CGGAACTGGAAATACGCTCGCTAACGATATGCAGGCCGCATGGGCGCCATTGGATCTGAATTATATCATGCCTCAATACAAATGC
CATACCCAAACCAACTGCACAAGACCGAGGGGTTGACTCTCCAAAGTTCTTTCTTCACAACACCTTTGCTGAAACCTACCCAGCG
TGGAAAGGACCTCTTCGCAAAACGTAAGTCATGAAACCAATATTCTGTTATAAATTTTGGGGTGGGATTAGGTGGGCTGTCATTA
TTTGAGGGGACCACGAAATATCCTTTTGCATGGGGTATCGTGCAAAAATATGACTGTGGGCTTCGAATAATTCCCCCATACAGTC
AAGAGTATGGGCTGCTGTTATGGATGAACATGAAGCGCACAAAACGATAATCCACTTCTTTAGCATGTATTGCTGTTCCAAAATT
GTGCCTATGGTTGCGGAGCTTTATTGTGCTCTGCATTGCGGGGACGGAAATAAACAGGCCAACGCGAGAAACGGGAGGTTTCGAG
AGTCGCGCATCTTTATTTTTTCTGTTACTACTCCCCACCAGAAAGCTTCACGTTTCTTCCAACTTCTGCACAAAACTTACGTAAC
CTTGTCACTCGGGTCTTGTTCTGCAGGATTCTGGAAATCGCTAGTCCTTCCACTAGCATTCTATCATGTCCTCTGATTCCTCGAC
GCCGAGTAATTTGGCCGCTCCAATTGTTGATGAGTCGGGATTCTCAACTGGCCCGCCCTCAGATCTATTTACCTTTTTGCGCGCC
GATAAGAGTGGGCGCTACGATTCTTTGGGTTCTCCCATCCCGAACCCTCCAAACGCACCTACCTGTGAGACCACTACTGATAGCT
TTTCGTACAGCACGCCACCTCTTTTTTCCAGCGAGGACGAGCTCTTTCACTCTGTCATCGCTGAGAATGCGGCCCAAGACGCGGC
CCAAGACGTGGACGGGCGAGAAATGTCTCTCCGCGACCTTATAGCGATGTTAAACGACATGGATCCCGAAAACATAAACTCTATA
GAACCGTGTCCGAACGCTACCAGTGATGTCGTGGCACTAGATATTAACGTGGATACCGATCAGCCCGGAGTTTATACGGTCATCG
ACTCGCTCGCGCATCCCTGCGAAACTCCCATGGAGCAGCAGCACGTTGAATGCCCGTGTCCATTACCTGCAGTAGAGAAAACAAC
CGCTGTGAGGCGTCAAGAAGTGGTGAGGAAGTCAGAGAGAGTGGCCAGATCTCGTTCGCGACGGCCGTGGCGCGATCTTTGGAGC
AACCGGCGCCCCGTCCCTTCGCCGCAGAGGACCAGCAAAGACCGTCTCCCCAAGCGAGGGAAGAGGGAATTTTCGAAGAAGATGG
GCCCGAGCCACCTTACCTCTTCCTCCTCTTCCTCCTCTTCCTCATTCTCGCTGTCAGGGAGGAGGGGACGTCTCGCCCGAAGGCT
GGAAAGCGCAGGAAGTGAAAAATCCACCGAAAGCCCCGACCGGGACTCTGCCCAATTAGTGCTACCTGGAGAATGTCCCGATGTC
GCGCTAGAGCCACTCTGGCGAGTTTTGAGGGAGTGGGTCGAAACGGACGGTACAGACGGGTCGGTTCAGTCAGTAACCGCGCCTA
CGGAGCTCCCAGACCTAGGCCGCGGCATTACCCTGAATGTACTCTCACGAGCGTTGGCCGTCGACCTCCATAGTTCCGAGGCTCG
CGTTTTTGTTCCAGAAGCTATTATCCCATCCCCACCCAGTAGAGGACAAGTATATCATGGAAGTGGCTATAGACAGTCTTTGGTC
GATGACCCGCAGGTTCGGGAGGCCGCGGCCGAATTCCATGCCAATCCGAAATCAGCTTCCGTGTATCTCGAAGAGTACGGGATGC
CAGTGGAACAGTTAGAGCGCCTTGTGGAGGAGTTCATGGTCCATCCCCTGTCGCGCGTTTCGTATGGGTATGGTGGCCCCAAGCT
GAGCCCACGTGACAAACGGCTCTGCGAACTGTGCGCTCTCCGCGGTCCGGGTTCTTCCGAGGGAGGATTGTCTCGTTCTGCGGTG
CCGACGTGCCTCCCGCACGTAGCGCACGCCATGCGTGATGGCCGTGTGTCATTAGCTCTGCCGCACCTTACCGAGCTCGTACGCG
TGTGCCGCCGCTATGACCACCCGCAGAAAACTTACCTGTTGGCGGCTCTTAGGAGAGCGTTTGTCCCGTTTGTATTTCCAGGAGC
GGCCGTGGATCCTACGGCGGCGTCTGCTTCTGGGGACCACCCTGTACTCGTTGGCGCCAGGCGAATAGCGCGTGCCACGAGGGAG
CTGGAAGCCGCAGCGAAAGAGCCACCTGCGGCAACCCCGGAAAAACTTTCCTCTTCGGCGCTATTGCGCCATACTAGAACTTGCA
TTGTTGCTGCCGCAGATCTCATGGCGGCGCTTACTGCTATAGACTGCCCACCAGTTACAATAAACCCCGCCAAGCGTCGCCCGGA
TCTGTATCAAGCACTCTCGAGGGCGCCCAAAACAGTCTTTACGTGGGAGCGCGCCATGCCGTACTGCGCGGCTTCGCTATGGAGT
CGCCTCGTCATAGGGTCCCTGCCGCTGCCATCTACCGATGATGTCACTCTTCTTGTAGAAGCCATGGAAGCTGCCGTGGGCGTGC
TTGGCTGCTCCGACGAGGGTGACGTGCGGCCTTCAATTGCCTACGCGGCCTCTGGTTGGAGGATGATCGCTACCATGCTAGGAGA
TGATTCCGCGCCCGGCAGCGAGCACGTCACGGACGATTGTCCCGCAGATGATGACACCCACCACCATGACAGGCGCAAAAGACGA
GGCGGCCGGCGGACGACTTGCCGGTACGACCTACCTCCCGGCACTCCCGATCGCGCGGCCCACCTCCTACTCTACGGGAGTCCTG
GCGGCGGCGACGGAGCAGAGCCTTCGGCCGCCCAAGAATCTCCCGCCAACCCCTGGCCCAGAGCTCCACCGTGCGATGAACAGGA
GCCATTATCTGTGTCACCGTACGGGCCCGAACCCGACCGCCCTCCAGATGATGACTTTGAGACCCGTAAAGGTCTGAAAAGAAAG
TCATCTGAAGATCATGCGGACCCGATCCCAGAAGGCAACGCGACTAAAAAAACATGTGGCCTCCAGGGCTTACCAGATTCATTAC
CTCCAGCGGTTCCAGAAACAGATAGAGACAACCCCCTCCTTCCGCCATGCCCCATAACTCCCGAGGGTCCACCATGTCCGCCCCG
AGAAGAACCGCAACAACCGCAAGAACCGCAAGAACCACAATCCCCGTCCTTTCACATTTCCGAGATAGGTGAGGCCCTGTTTCAT
TCTACCCCAGTCAGCCCTACCATCCTGTTCGCACCGGAGGGTTTCATCCCGAACCAAGGGGATGTTTACCGTCCCCATGGCAAAA
ACTTGGGCGTTGCGCTCTCCGCAGGGGAGCCATACGTGCCCGCCCATCGCGCGCGTCCTGCTTTGGAGCGGCTCGCGAACTATTT
GCGCGGGTTGAGCAGCGGTCCGGCGGGAAGCTATTCCGCGACCGGACCCGCAGATTGGCTACCTGCTTTTGCGAGGGACCCAGTA
GCCCTGGGGGAGTTTTGCAAGAGGATTTCACCCGTGAACAGAGGGGGACGCTGCGTCTTAGAGGAGCGGCTTTCCTGGACGTTGC
GACATCCGGCTGACAAAATCTATGACCTCATCATTCTTGTCTCAGAACGTGCCCCGGGCGAATTTTTAGAACGCGCGTACGCCTA
CGCGGCCCTCTGCGGACGTGCCTGCATGCCGGGCGGTGCCTGTTGGCCAGAATCGTGGACACGCGGGGAGTATCCCTATATCCCT
CCCGACGAGAAAGACCCGGGCCGCAGCGTGGCCATTCTCTGTACGACAGACCTTGGATACGCCGGGGCGGTAGAATGGATGCTTG
AGCGCGCTCTCAGTACATCTGGGCGCGCATGCGTGGTCGTAGATGCTAGGGATCCCGGGGACGAGGATTCCCGCTCGGGTCCGCG
CATACCTACCGGGCGGAGGGGTGTTGTCTATGCCCGGACTAGCCCCCCGTCGCACGCTTACCTTGTAAAACTGTTTACCGGTCCG
GCGGCCGAACGGTTGGAACGGGGTGGGCTCAATGTAGGTCACGTAGTAATGGACAGGCGACTCACTGTCTCATGGCCCCTCAGCG
TTGGGAAGACCGACAACGACGATGAGGGTACAGGAGACAGAGAGAAGTGTCCTCCCGTTAAGCACCCCTTCCTCACTCACGTAAC
TCTCAGGAGATTACGCGACGTTGCTCGTGCCATGGATGCCGGGTGGAGCCACTCCGAAATCGGAAAAGCTTCAGCCGTCTTGTAT
CCGTACATGGCCCAGCCCCATGTGGGCGCCGGGGCGATGCCGCCATGTCCGGACCTTTCCGAGTCCAGTTCCACTATGCATTCTT
CTTCGTCGTCTTCCTCTTCCTCATGCTCCTCGTCGTCGTCTTCTTCAGACTCCAGCTCATCTGAAGAAGACGGGGACGAGAAGAA
CGAGAAAGAAGATCGTGAGCGCGCCGGAGGTGGAAAGCGGCGCGGCCGCCAGCGCCTCCCGATTCGGGACCGCGTGTACCGTGGC
CGGGACTACCGCGTGCGCACTCCCTGTGCTGATATTGTGGGAGGCACTTCCGCTTCCTTTAGGGCCGTCGCTGCGTGGGGGAGGG
ACACGGTAATTCCCGCCCTCCCTCTCGCAGCGTCCTGGTCCGATCCCAGCAAAATTCCACAGGAAGTTCTCCGCATCATCTCCGA

TTATTACCCGGATGCGCCAGGGGCGCGGGCGCGTCGGAACCCGCTTCACGTGGAAGGACTAGCGCTTATGCGTGCCAGAAATCCC
GCCCCTCTCGCATTGCTACTTGGTGACGACTATTCCCATTATCACACTCCCCGCAATCGGTCCCCAATGTGTTGTTGCTGGAGTG
CTCCGCGGGGAGCAACGCCAGAGTCACTCAGATGCACCGGACACACGCGCACCAGGCGCCCAAGGCTATCCGTCTCCTCAGGGAG
GGGGGAAACTAGAACCCTTGTCCCCAGATTAGATTAGACTCTTCAGAGGATAGGCGGTCGCGTCAACGCCATAGACGCATGACTA
ACTACCCTCTAACGCATCCCACTGTCACTGACCCTTGGTGGTAACCGTGGTTTCCGTTATTAACTTATTAAAACTTTTTGTACAA
TTTACTTCTGGTTGCGTGCTCATTCTTTTGCTTGTCGGA

Appendix 4 – RNA Hybrid analysis of viral targets of ILTV-encoded miRNAs

Gene	miRNA	Position	MFE (Kcal/Mol)
ORF F	ILTV-miR-I1-3p	1497	-16.7
	ILTV-miR-I2	1295	-27.7
	ILTV-miR-I2	2043	-23.9
	ILTV-miR-I3	240	-26.0
	ILTV-miR-I3	812	-25.7
	ILTV-miR-I4	1560	-28.1
	ILTV-miR-I4	1771	-21.9
	ILTV-miR-I5-5p	251	-28.2
	ILTV-miR-I5-5p	184	-23.3
	ILTV-miR-I5-3p	2728	-16.4
	ILTV-miR-I6-5p	601	-30.3
	ILTV-miR-I6-5p	2059	-24.8
	ILTV-miR-I6-3p	2458	-33.3
	ILTV-miR-I6-3p	677	-33.0
	ILTV-miR-I7	1752	-21.2
UL54	ILTV-miR-I6-3p	1543	-33.7
UL52	ILTV-miR-I1-5p	3106	-25.3
	ILTV-miR-I6-5p	2524	-25.7
	ILTV-miR-I6-3p	5091	-21.7
	ILTV-miR-I7	4434	-15.9
UL50	ILTV-miR-I1-3p	1162	-18.6
	ILTV-miR-I3	1310	-30.7
	ILTV-miR-I4	647	-25.9
	ILTV-miR-I4	1243	-20.6
	ILTV-miR-I5-3p	370	-32.0
	ILTV-miR-I6-3p	1852	-18.7
UL48	ILTV-miR-I5-3p	3376	-18.4
	ILTV-miR-I6-5p	3716	-21.5
UL46	ILTV-miR-I5-3p	2126	-18.4
	ILTV-miR-I6-5p	2466	-21.5
UL23	ILTV-miR-I1-3p	4697	-16.0
	ILTV-miR-I2	867	-33.2
	ILTV-miR-I2	134	-28.3
	ILTV-miR-I3	5097	-20.9
	ILTV-miR-I4	3567	-20.1
	ILTV-miR-I4	3227	-17.4
	ILTV-miR-I6-5p	154	-20.8
UL26	ILTV-miR-I1-5p	935	-22.9
	ILTV-miR-I1-5p	991	-18.2
	ILTV-miR-I1-3p	23	-25.1
	ILTV-miR-I1-3p	798	-21.5
	ILTV-miR-I2	2012	-26.7
	ILTV-miR-I2	1947	-23.1
	ILTV-miR-I3	425	-22.0
	ILTV-miR-I3	1063	-23.9
	ILTV-miR-I4	2288	-19.3
	ILTV-miR-I4	2491	-18.4

	ILTV-miR-I5-5p	1071	-21.8
	ILTV-miR-I5-5p	2013	-18.0
	ILTV-miR-I5-3p	2804	-24.2
UL26	ILTV-miR-I5-3p	1246	-15.8
	ILTV-miR-I6-5p	372	-23.4
	ILTV-miR-I6-5p	2654	-21.6
	ILTV-miR-I6-3p	679	-19.2
	ILTV-miR-I7	1488	-17.2
UL26.5	ILTV-miR-I1-5p	65	-22.9
	ILTV-miR-I1-5p	121	-18.2
	ILTV-miR-I2	1142	-26.7
	ILTV-miR-I2	1077	-23.1
	ILTV-miR-I3	193	-23.9
	ILTV-miR-I3	749	-18.0
	ILTV-miR-I4	1418	-19.3
	ILTV-miR-I4	1621	-18.4
	ILTV-miR-I5-5p	201	-21.8
	ILTV-miR-I5-5p	1143	-18.0
	ILTV-miR-I5-3p	1934	-24.2
	ILTV-miR-I5-3p	376	-15.8
	ILTV-miR-I6-5p	1784	-21.6
	ILTV-miR-I7	618	-17.2
UL28	ILTV-miR-I1-5p	409	-27.6
	ILTV-miR-I1-5p	2772	-22.3
	ILTV-miR-I1-3p	1374	-19.2
	ILTV-miR-I1-3p	3926	-17.9
	ILTV-miR-I2	57	-30.3
	ILTV-miR-I2	3040	-26.9
	ILTV-miR-I3	2376	-25.1
	ILTV-miR-I3	5666	-24.1
	ILTV-miR-I4	1582	-23.2
	ILTV-miR-I4	1155	-20.6
	ILTV-miR-I5-5p	2389	-30.5
	ILTV-miR-I5-5p	5696	-29.3
	ILTV-miR-I5-3p	5973	-19.2
	ILTV-miR-I5-3p	2867	-18.4
	ILTV-miR-I6-5p	1522	-22.6
	ILTV-miR-I6-5p	5362	-21.7
	ILTV-miR-I6-3p	733	-28.1
	ILTV-miR-I6-3p	1590	-21.5
	ILTV-miR-I7	287	-23.9
	ILTV-miR-I7	2381	-20.9
UL29	ILTV-miR-I1-5p	3486	-27.6
	ILTV-miR-I1-5p	1549	-23.1
	ILTV-miR-I2	3134	-30.3
	ILTV-miR-I2	2759	-22.2
	ILTV-miR-I3	2668	-24.4
	ILTV-miR-I3	1848	-23.6
	ILTV-miR-I4	2544	-22.2
	ILTV-miR-I4	1814	-19.2
	ILTV-miR-I5-5p	3873	-26.8

	ILTV-miR-I5-5p	2176	-18.4
	ILTV-miR-I5-3p	2096	-16.0
	ILTV-miR-I6-5p	3856	-21.7
UL29	ILTV-miR-I6-5p	3676	-20.6
	ILTV-miR-I6-3p	3810	-28.1
	ILTV-miR-I6-3p	916	-23.4
	ILTV-miR-I7	3364	-23.9
	ILTV-miR-I7	1052	-16.2
UL30	ILTV-miR-I1-5p	2415	-15.0
	ILTV-miR-I1-3p	2107	-20.7
	ILTV-miR-I1-3p	3199	-18.7
	ILTV-miR-I2	107	-24.2
	ILTV-miR-I2	3974	-20.9
	ILTV-miR-I3	3747	-29.7
	ILTV-miR-I3	432	-20.1
	ILTV-miR-I4	2681	-22.1
	ILTV-miR-I4	2637	-17.5
	ILTV-miR-I5-5p	2302	-19.1
	ILTV-miR-I5-5p	630	-16.6
	ILTV-miR-I5-3p	2162	-25.2
	ILTV-miR-I5-3p	1787	-20.0
	ILTV-miR-I6-5p	645	-25.8
	ILTV-miR-I6-5p	2548	-24.1
	ILTV-miR-I6-3p	1600	-24.6
	ILTV-miR-I6-3p	673	-21.7
	ILTV-miR-I7	1473	-16.9
	ILTV-miR-I7	1234	-16.1
UL39	ILTV-miR-I1-5p	3919	-27.0
	ILTV-miR-I1-5p	1374	-20.7
	ILTV-miR-I1-3p	2754	-22.2
	ILTV-miR-I1-3p	845	-19.2
	ILTV-miR-I2	752	-29.1
	ILTV-miR-I2	552	-26.5
	ILTV-miR-I3	609	-26.9
	ILTV-miR-I3	1152	-26.0
	ILTV-miR-I4	2260	-20.1
	ILTV-miR-I4	418	-17.0
	ILTV-miR-I5-5p	1117	-27.6
	ILTV-miR-I5-5p	586	-22.4
	ILTV-miR-I5-3p	2366	-23.7
	ILTV-miR-I5-3p	3839	-22.8
	ILTV-miR-I6-5p	3875	-22.4
	ILTV-miR-I6-5p	2910	-20.0
	ILTV-miR-I6-3p	2511	-24.2
	ILTV-miR-I6-3p	1762	-24.0
	ILTV-miR-I7	2514	-24.2
UL40	ILTV-miR-I1-5p	1556	-27.0
	ILTV-miR-I1-5p	992	-17.3
	ILTV-miR-I1-3p	391	-22.2
	ILTV-miR-I4	1950	-15.3
	ILTV-miR-I5-5p	117	-19.9

	ILTV-miR-I5-5p	499	-18.9
	ILTV-miR-I5-3p	3	-23.7
	ILTV-miR-I5-3p	1476	-22.8
UL40	ILTV-miR-I6-5p	1512	-22.4
	ILTV-miR-I6-5p	547	-20.0
	ILTV-miR-I6-3p	148	-24.2
	ILTV-miR-I6-3p	549	-17.8
	ILTV-miR-I7	151	-24.2
UL12	ILTV-miR-I5-5p	1982	-23.5
	ILTV-miR-I5-5p	688	-16.2
UL9	ILTV-miR-I6-3p	1068	-26.3
	ILTV-miR-I6-3p	972	-24.0
UL5	ILTV-miR-I2	505	-25.1
	ILTV-miR-I6-5p	849	-23.6
	ILTV-miR-I6-5p	3859	-23.0
	ILTV-miR-I7	1973	-18.1
UL4	ILTV-miR-I6-5p	1249	-23.0
UL3.5	ILTV-miR-I2	779	-24.3
	ILTV-miR-I5-3p	776	-19.1
	ILTV-miR-I6-5p	528	-21.9
	ILTV-miR-I6-3p	1268	-16.8
UL2	ILTV-miR-I1-5p	1155	-22.5
	ILTV-miR-I2	164	-31.7
	ILTV-miR-I2	1423	-24.5
	ILTV-miR-I3	1330	-22.7
	ILTV-miR-I3	1379	-22.1
	ILTV-miR-I4	953	-22.5
	ILTV-miR-I4	1021	-14.9
	ILTV-miR-I5-5p	230	-19.2
	ILTV-miR-I5-5p	1176	-14.2
	ILTV-miR-I5-3p	883	-19.7
	ILTV-miR-I6-5p	614	-19.2
	ILTV-miR-I6-5p	747	-23.9
	ILTV-miR-I6-3p	1275	-15.9
	ILTV-miR-I6-3p	1441	-26.4
UL0	ILTV-miR-I5-3p	2855	-19.7
UL-1	ILTV-miR-I2	2	-28.2
	ILTV-miR-I2	157	-23.4
	ILTV-miR-I3	2541	-30.1
	ILTV-miR-I3	2249	-25.5
	ILTV-miR-I4	126	-20.8
	ILTV-miR-I4	1961	-19.6
	ILTV-miR-I5-5p	2292	-33.9
	ILTV-miR-I5-5p	1126	-30.3
	ILTV-miR-I5-3p	80	-23.4
	ILTV-miR-I5-3p	1217	-16.6
	ILTV-miR-I6-5p	616	-29.8
	ILTV-miR-I6-5p	2286	-29.6
	ILTV-miR-I6-3p	2651	-15.5
US10	ILTV-miR-I2	832	-28.7
	ILTV-miR-I5-3p	614	-24.6

sORF3/4	ILTV-miR-I1-5p	578	-24.4
	ILTV-miR-I3	544	-23.8
US2	ILTV-miR-I1-5p	1485	-24.4
US2	ILTV-miR-I3	1451	-23.8
	ILTV-miR-I5-3p	845	-27.5
US3	ILTV-miR-I1-3p	2220	-20.6
	ILTV-miR-I2	773	-28.5
	ILTV-miR-I2	32	-26.9
	ILTV-miR-I3	643	-26.7
	ILTV-miR-I3	383	-25.8
	ILTV-miR-I4	1108	-29.5
	ILTV-miR-I4	1683	-21.7
	ILTV-miR-I5-5p	2054	-20.7
	ILTV-miR-I5-5p	1585	-20.4
	ILTV-miR-I5-3p	213	-24.8
	ILTV-miR-I5-3p	1365	-18.7
	ILTV-miR-I6-5p	1489	-27.5
	ILTV-miR-I6-5p	2064	-26.8
	ILTV-miR-I6-3p	496	-18.3
	ILTV-miR-I6-3p	966	-18.3
	ILTV-miR-I7	1064	-20.5
sORF1 (UL47)	ILTV-miR-I5-5p	74	-20.4

Appendix 5: High Confidence Chicken miRNAs from MiRBase Release 21 (June 2014)

Chicken miRNA Sequences

```
>gga-miR-200b-5p MIMAT0026534
UCUUACUGGGCAGCAUUGGA
>gga-miR-181a-3p MIMAT0001150
ACCAUCGACCGUUGAUUGUACC
>gga-miR-126-3p MIMAT0001169
UCGUACCGUGAGUAAUAAUGCGC
>gga-miR-29b-3p MIMAT0001097
UAGCACCAUUUGAAAUCAGUGUU
>gga-miR-10b-5p MIMAT0001148
UACCCUGUAGAACCGAAUUUGU
>gga-miR-181b-5p MIMAT0001151
AACAUUCAUUGCUGUCGGUGGG
>gga-miR-460b-5p MIMAT0007326
UCCUCAUUGUACAUGCUGUGUG
>gga-miR-429-3p MIMAT0003371
UAAUACUGUCUGGUAAUGCCGU
>gga-miR-30c-1-3p MIMAT0031105
UGGGAGAGGAUUGUUUACGCC
>gga-miR-34b-5p MIMAT0001179
CAGGCAGUGUAGUUAGCUGAUUG
>gga-miR-22-3p MIMAT0007288
AAGCUGCCAGUUGAAGAACUGU
>gga-miR-499-5p MIMAT0003367
UUAAGACUUGUAGUGAUGUUUAG
>gga-miR-30a-5p MIMAT0001135
UGUAAACAUCCUCGACUGGAAG
>gga-miR-29a-3p MIMAT0001096
UAGCACCAUUUGAAAUCGGUU
>gga-miR-22-5p MIMAT0007287
AGUUCUUCAGUGGCAAGCUUUA
>gga-miR-222b-5p MIMAT0025617
UGCUCAGUAGUCAGUGUAGGAUCUGU
>gga-miR-126-5p MIMAT0003723
CAUUAUUACUUUUGGUACGCG
>gga-miR-460b-3p MIMAT0007327
CACAGCGCAUGCAAUGUGGACAU
>gga-miR-125b-3p MIMAT0026493
ACAAGUCAGGCUCUUGGGACCU
>gga-miR-34b-3p MIMAT0026540
AAUCACUAAAUUCACUGCCAUC
>gga-miR-144-3p MIMAT0003776
CUACAGUAUAGAUGAUGUACUC
>gga-miR-140-3p MIMAT0003722
CCACAGGGUAGAACCACGGAC
>gga-let-7i MIMAT0001098
UGAGGUAGUAGUUUGUGCUGU
>gga-miR-187-3p MIMAT0001124
UCGUGUCUUGUGUUGCAGCC
>gga-miR-33-5p MIMAT0001100
GUGCAUUGUAGUUGCAUUGC
>gga-miR-10b-3p MIMAT0026521
AGAUUCGAUUCUAGGGGAAUA
>gga-miR-30a-3p MIMAT0001136
CUUUCAGUCGGAUGUUUGCAGC
```

>gga-miR-128-3p MIMAT0001123
UCACAGUGAACCGGUCUCUUU
>gga-miR-10a-3p MIMAT0007732
AAAUUCGUAUCUAGGGGAAUA
>gga-miR-221-3p MIMAT0001108
AGCUACAUUGUCUGCUGGGUUUC
>gga-miR-200b-3p MIMAT0001172
UAAUACUGCCUGGUAAUGAUGAU
>gga-miR-27b-3p MIMAT0001187
UUCACAGUGGCUAAGUUCUGC
>gga-miR-221-5p MIMAT0026494
AACCUGGCAUACAAUGUAGAUUUCUGU
>gga-miR-17-5p MIMAT0001114
CAAAGUGCUUACAGUGCAGGUAGU
>gga-miR-181a-5p MIMAT0001168
AACAUUCAACGCUGUCGGUGAGU
>gga-miR-222b-3p MIMAT0025618
AGCUACAUCUGAUUACUGGGUCAC
>gga-miR-181b-1-3p MIMAT0026522
UCACUGAACAAUGAAUGCAAC
>gga-miR-17-3p MIMAT0001115
ACUGCAGUGAAGGCACUUGU
>gga-miR-33-3p MIMAT0026490
AAUGUUCCUGCAGUGCAGUA
>gga-miR-133a-3p MIMAT0001126
UUGGUCCCCUUCAACCAGCUGU
>gga-miR-429-5p MIMAT0026635
UGUCUUACCAGGCAAAGUUAGA
>gga-miR-30c-2-3p MIMAT0026515
UGGGAGAAGGCUGUUUACUCU
>gga-miR-125b-5p MIMAT0001105
UCCCUGAGACCCUAACUUGUGA
>gga-miR-187-5p MIMAT0026507
GGCUACAACACAGGACAUGGGA
>gga-miR-29b-1-5p MIMAT0026488
AGCUGGUUUCAUAUGGUGGUUUAGA
>gga-miR-30c-5p MIMAT0001137
UGUAAACAUCCUACACUCUCAGCU
>gga-miR-29a-5p MIMAT0026487
CUGAUUUCUUUUGGUGUUCAGA
>gga-miR-27b-5p MIMAT0026547
AGAGCUUAGCUGAUUGGUGAACA
>gga-miR-133a-5p MIMAT0026509
AGCUGGUAAAAUGGAACCAAAUC
>gga-miR-10a-5p MIMAT0007731
UACCCUGUAGAUCCGAAUUUGU
>gga-miR-128-2-5p MIMAT0026506
GGGGGCCGUUACACUGUAAGAGA
>gga-miR-144-5p MIMAT0026644
GGAUAUCAUCAUAUACUGUAAG
>gga-miR-499-3p MIMAT0026634
AACAUCACUUUAAGUCUGUGCU
>gga-miR-140-5p MIMAT0001159
AGUGGUUUUACCCUAUGGUAG

Appendix 6 – Alternative Targets of gga-miR-133a-3p in the ILTV Genome

Alternative miR-133a-3p Targets

Gene	Target Position	MFE (kcal/mol)	Possible Gene Function? †
UL31	134	-27.7	Nuclear matrix protein
UL32	1873	- 27.7	Envelope glycoprotein
UL36	418	-26.7	Large tegument protein
UL37	3712	-26.7	Capsid assembly
UL42	1929	-29.8	DNA Polymerase processivity factor
UL43	546	-29.8	Membrane protein
UL20	2902	-26.1	Membrane protein
UL19	2122	-26.1	VP5 – Major capsid protein
ICP4as*	586	-30.3	Essential Immediate-Early gene which is absolutely required for lytic gene replication
ICP4as*	2563	-28.5	
ICP4s	584	-30.3	
ICP4	2563	-28.5	

*ICP4 rows are direct copies of each other due to there been two copies of ICP4 in the ILTV genome, one been in the inverted repeat region hence ‘as’ for antisense orientation. These two were tested as laid out in Chapter 5 of this thesis.

†Data taken from HSV-1 genome as it is the most annotated of the Alphaherpesviruses and is therefore inferred on ILTV. Some functions may differ.

Appendix 7: miR-133a-3p Targeting of ICP4 in different herpesvirus species

Alignment of miR-133a-3p from different animal species

```
gga-miR-133a-3p   : 98
hsa-miR-133a-3p   : 98
eca-miR-133a      : 98
ssc-miR-133a-3p   : 97
bta-miR-133a      : 96
cons              : 98

gga-miR-133a-3p   -UUGGUCCCCUUCAACCAGCUGU
hsa-miR-133a-3p   UUUGGUCCCCUUCAACCAGCUG-
eca-miR-133a      UUUGGUCCCCUUCAACCAGCUG-
ssc-miR-133a-3p   -UUGGUCCCCUUCAACCAGCUG-
bta-miR-133a      UUUGGUCCCCUUCAACCAGCUG-

cons              **********************
```

Species included:

gga = Chicken
hsa = Human
eca = Horse
ssc = Pig
bta = Cow

RNA Hybrid analysis of miR-133a-3p Targeting ICP4 in different Herpesvirus species

```
Target: BOHV-1_ICP4_L14320.1
MiRNA: gga-miR-133a-3p
MFE: -29.4 kcal/mol
Position: 4033
Target 5'  A   C  GAC  GG CC       A 3'
          GGC GG   UG  G  GGGACCAG
          UCG CC   AC  C  CCCUGGUU
miRNA  3' UG   A  A    UU          5'
------------------------------------------------------------
Target: EQHV-1_ICP4_NC_001491.2
MiRNA: gga-miR-133a-3p
MFE: -28.7 kcal/mol
Position: 2918
Target 5' C     CGC  CCGCUCC        G 3'
         GCGGC   GG      GGGGACCAG
         UGUCG   CC      CCCCUGGUU
miRNA  3'       A    AACUU          5'
------------------------------------------------------------
```

Target: HSV-1_ICP4_NC_001806.2
MiRNA: gga-miR-133a-3p
MFE: -30.6 kcal/mol
Position: 3724
Target 5' G C GC CGGCCUC C 3'
ACGGC GG GG GGGGACCA
UGUCG CC CU CCCCUGGU
miRNA 3' A AA U U 5'

Target: MDV_ICP4
MiRNA: gga-miR-133a-3p
MFE: -23.6 kcal/mol
Position: 7320
Target 5' C CCCGGCCUUCUC UUUC C 3'
GCUGG GA GGGACCA
CGACC CU CCCUGGU
miRNA 3' UGU AA UC U 5'

Target: ILTV
MiRNA: gga-miR-133a-3p
MFE: -30.3 kcal/mol
Position: 584
Target 5' G UCAUUAU C 3'
GGCUG UUG AGGGGACCA
UCGAC AAC UCCCCUGGU
miRNA 3' UG C U U 5'

Target: ILTV
MiRNA: gga-miR-133a-3p
MFE: -28.5 kcal/mol
Position: 2563
Target 5' U G C C CUUCU C 3'
ACGGC G GU UG GGGGACCA
UGUCG C CA AC CCCCUGGU
miRNA 3' A UU U 5'

Virus Species included in screen:

BOHV-1 = Bovine alphaherpesvirus-1
EQHV-1 = Equine alphaherpesvirus-1
HSV-1 = Human Simplex virus-1
MDV-1 = Marek's Disease Virus-1
ILTV = Infectious Laryngotracheitis virus

References

AFONSO, C. L., TULMAN, E. R., LU, Z., ZSAK, L., ROCK, D. L. & KUTISH, G. F. 2001. The genome of turkey herpesvirus. *J Virol,* 75**,** 971-8.

AGELIDIS, A. M. & SHUKLA, D. 2015. Cell entry mechanisms of HSV: what we have learned in recent years. *Future virology,* 10**,** 1145-1154.

AUERBACH, C., ROBSON, J. M. & CARR, J. G. 1947. The Chemical Production of Mutations. *Science,* 105**,** 243-7.

BAE, S., PARK, J. & KIM, J. S. 2014. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics,* 30**,** 1473-5.

BAGUST, T. J. 1986. LARYNGOTRACHEITIS (GALLID-1) HERPESVIRUS-INFECTION IN THE CHICKEN .4. LATENCY ESTABLISHMENT BY WILD AND VACCINE STRAINS OF ILT VIRUS. *Avian Pathology,* 15**,** 581-595.

BAGUST, T. J., JONES, R. C. & GUY, J. S. 2000. Avian infectious laryngotracheitis. *Revue Scientifique Et Technique De L Office International Des Epizooties,* 19**,** 483-492.

BAI, Y. C., HE, L. J., LI, P. C., XU, K., SHAO, S. M., REN, C. H., LIU, Z. T., WEI, Z. H. & ZHANG, Z. Y. 2016. Efficient Genome Editing in Chicken DF-1 Cells Using the CRISPR/Cas9 System. *G3-Genes Genomes Genetics,* 6**,** 917-923.

BARBER, K. A., DAUGHERTY, H. C., ANDER, S. E., JEFFERSON, V. A., SHACK, L. A., PECHAN, T., NANDURI, B. & MEYER, F. 2017. Protein Composition of the Bovine Herpesvirus 1.1 Virion. *Veterinary Sciences,* 4**,** 11.

BARROW, P. A. & LOVELL, M. A. 1989. Invasion of Vero cells by Salmonella species. *J Med Microbiol,* 28**,** 59-67.

BARTEL, D. P. 2009. MicroRNAs: Target Recognition and Regulatory Functions. *Cell,* 136**,** 215-233.

BEHM-ANSMANT, I., REHWINKEL, J., DOERKS, T., STARK, A., BORK, P. & IZAURRALDE, E. 2006. mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes & development,* 20**,** 1885-1898.

BENHALEVY, D., MCFARLAND, H. L., SARSHAD, A. A. & HAFNER, M. 2017. PAR-CLIP and streamlined small RNA cDNA library preparation protocol for the identification of RNA binding protein target sites. *Methods,* 118-119**,** 41-49.

BENNASSER, Y., LE, S. Y., YEUNG, M. L. & JEANG, K. T. 2004. HIV-1 encoded candidate micro-RNAs and their cellular targets. *Retrovirology,* 1**,** 43.

BEREZIKOV, E. 2011. Evolution of microRNA diversity and regulation in animals. *Nat Rev Genet,* 12**,** 846-60.

BERNIER, A. & SAGAN, S. 2018. The Diverse Roles of microRNAs at the Host–Virus Interface. *Viruses,* 10**,** 440.

BEUG, H., VON KIRCHBACH, A., DÖDERLEIN, G., CONSCIENCE, J.-F. & GRAF, T. 1979. Chicken hematopoietic cells transformed by seven strains of defective avian leukemia viruses display three distinct phenotypes of differentiation. *Cell,* 18**,** 375-390.

BI, Y. W., SUN, L., GAO, D. D., DING, C., LI, Z. H., LI, Y. D., CUN, W. & LI, Q. H. 2014. High-Efficiency Targeted Editing of Large Viral Genomes by RNA-Guided Nucleases. *Plos Pathogens,* 10.

BLOOM, K., ELY, A., MUSSOLINO, C., CATHOMEN, T. & ARBUTHNOT, P. 2013. Inactivation of hepatitis B virus replication in cultured cells and in vivo with engineered transcription activator-like effector nucleases. *Mol Ther,* 21**,** 1889-97.

BOCH, J. 2011. TALEs of genome targeting. *Nat Biotechnol,* 29**,** 135-6.

BOCH, J. & BONAS, U. 2010. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol,* 48**,** 419-36.

References

BOEHMER, P. E. & LEHMAN, I. R. 1997. Herpes simplex virus DNA replication. *Annu Rev Biochem,* 66**,** 347-84.

BOSHOFF, C. & WEISS, R. 2002. AIDS-related malignancies. *Nat Rev Cancer,* 2**,** 373-82.

BOUTZ, P. L., CHAWLA, G., STOILOV, P. & BLACK, D. L. 2007. MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development. *Genes & Development,* 21**,** 71-84.

BRANDLY, C. A. & BUSHNELL, L. D. 1934. A Report of Some Investigations of Infectious Laryngotracheitis. *Poultry Science,* 13**,** 212-217.

BROWN, J. C. & NEWCOMB, W. W. 2011. Herpesvirus capsid assembly: insights from structural analysis. *Curr Opin Virol,* 1**,** 142-9.

BROWN, W. R., HUBBARD, S. J., TICKLE, C. & WILSON, S. A. 2003. The chicken as a model for large-scale analysis of vertebrate gene function. *Nat Rev Genet,* 4**,** 87-98.

BRYANT, N. A., DAVIS-POYNTER, N., VANDERPLASSCHEN, A. & ALCAMI, A. 2003. Glycoprotein G isoforms from some alphaherpesviruses function as broad-spectrum chemokine binding proteins. *The EMBO Journal,* 22**,** 833-846.

BUCKLES, E. 2009. *Atlas of Avian Diseases,* [Ithaca, N.Y.] : T1 - Atlas of avian diseases.

BUTLER, D. 2016. Zika raises profile of more common birth-defect virus. *Nature,* 535**,** 17.

BYRNE, S. M., ORTIZ, L., MALI, P., AACH, J. & CHURCH, G. M. 2015. Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. *Nucleic Acids Res,* 43**,** e21.

CAI, X. Z., HAGEDORN, C. H. & CULLEN, B. R. 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna-a Publication of the Rna Society,* 10**,** 1957-1966.

CALNEK, B. W., FAHEY, K. J. & BAGUST, T. J. 1986. INVITRO INFECTION STUDIES WITH INFECTIOUS LARYNGOTRACHEITIS VIRUS. *Avian Diseases,* 30**,** 327-336.

CAMPADELLI-FIUME, G. & MENOTTI, L. 2007. *Entry of alphaherpesviruses into the cell,* Cambridge, Cambridge University Press.

CARROLL, D. 2011. Genome engineering with zinc-finger nucleases. *Genetics,* 188**,** 773-82.

CARROLL, D. 2017. Genome Editing: Past, Present, and Future. *The Yale Journal of Biology and Medicine,* 90**,** 653-659.

CHANG, P. W., YATES, V. J., DARDIRI, A. H. & FRY, D. E. 1960. Some Observations of the Propagation of Infectious Laryngotracheitis Virus in Tissue Culture. *Avian Diseases,* 4**,** 384-390.

CHANG, Y. E., VAN SANT, C., KRUG, P. W., SEARS, A. E. & ROIZMAN, B. 1997. The null mutant of the U(L)31 gene of herpes simplex virus 1: construction and phenotype in infected cells. *Journal of Virology,* 71**,** 8307-8315.

CHEN, Y., BAI, P., MACKAY, S., KORZA, G., CARSON, J. H., KUCHTA, R. D. & WELLER, S. K. 2011. Herpes Simplex Virus Type 1 Helicase-Primase: DNA Binding and Consequent Protein Oligomerization and Primase Activation. 85**,** 968-978.

CHEN, Y. C., SHENG, J., TRANG, P. & LIU, F. 2018. Potential Application of the CRISPR/Cas9 System against Herpesvirus Infections. *Viruses,* 10.

CHI, S. W., ZANG, J. B., MELE, A. & DARNELL, R. B. 2009. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature,* 460**,** 479-86.

CHOU, C. H., LIN, F. M., CHOU, M. T., HSU, S. D., CHANG, T. H., WENG, S. L., SHRESTHA, S., HSIAO, C. C., HUNG, J. H. & HUANG, H. D. 2013. A computational approach for identifying microRNA-target interactions using high-throughput CLIP and PAR-CLIP sequencing. *BMC Genomics,* 14 Suppl 1**,** S2.

CLIFFE, A. R. & KNIPE, D. M. 2008. Herpes simplex virus ICP0 promotes both histone removal and acetylation on viral DNA during lytic infection. *J Virol,* 82**,** 12030-8.

References

CLIFFE, A. R. & WILSON, A. C. 2017. Restarting Lytic Gene Transcription at the Onset of Herpes Simplex Virus Reactivation. *J Virol,* 91.

CONRADY, C. D., DREVETS, D. A. & CARR, D. J. 2010. Herpes simplex type I (HSV-1) infection of the nervous system: is an immune response a good thing? *J Neuroimmunol,* 220**,** 1-9.

COPPO, M. J. C., DEVLIN, J. M., LEGIONE, A. R., VAZ, P. K., LEE, S. W., QUINTEROS, J. A., GILKERSON, J. R., FICORILLI, N., READING, P. C., NOORMOHAMMADI, A. H. & HARTLEY, C. A. 2018. Infectious Laryngotracheitis Virus Viral Chemokine-Binding Protein Glycoprotein G Alters Transcription of Key Inflammatory Mediators In Vitro and In Vivo. *Journal of Virology,* 92**,** 17.

COPPO, M. J. C., DEVLIN, J. M. & NOORMOHAMMADI, A. H. 2012. Comparison of the replication and transmissibility of an infectious laryngotracheitis virus vaccine delivered via eye-drop or drinking-water. *Avian Pathology,* 41**,** 99-106.

COPPO, M. J. C., HARTLEY, C. A. & DEVLIN, J. M. 2013a. Immune responses to infectious laryngotracheitis virus. *Developmental and Comparative Immunology,* 41**,** 454-462.

COPPO, M. J. C., NOORMOHAMMADI, A. H., BROWNING, G. F. & DEVLIN, J. M. 2013b. Challenges and recent advancements in infectious laryngotracheitis virus vaccines. *Avian Pathology,* 42**,** 195-205.

COPPO, M. J. C., NOORMOHAMMADI, A. H., HARTLEY, C. A., GILKERSON, J. R., BROWNING, G. F. & DEVLIN, J. M. 2011. Comparative in vivo safety and efficacy of a glycoprotein G-deficient candidate vaccine strain of infectious laryngotracheitis virus delivered via eye drop. *Avian Pathology,* 40**,** 411-417.

COSTINEAN, S., ZANESI, N., PEKARSKY, Y., TILI, E., VOLINIA, S., HEEREMA, N. & CROCE, C. M. 2006. Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 7024-7029.

COWEN, B. S. & BRAUNE, M. O. 1988. THE PROPAGATION OF AVIAN VIRUSES IN A CONTINUOUS CELL-LINE (QT35) OF JAPANESE QUAIL ORIGIN. *Avian Diseases,* 32**,** 282-297.

COX, D. B. T., GOOTENBERG, J. S., ABUDAYYEH, O. O., FRANKLIN, B., KELLNER, M. J., JOUNG, J. & ZHANG, F. 2017. RNA editing with CRISPR-Cas13. *Science,* 358**,** 1019-1027.

CULLEN, B. R. 2009. Viral and cellular messenger RNA targets of viral microRNAs. *Nature,* 457**,** 421-425.

CULLEN, B. R. 2010. Five Questions about Viruses and MicroRNAs. *Plos Pathogens,* 6.

CULLEN, B. R. 2011. Viruses and microRNAs: RISCy interactions with serious consequences. *Genes & Development,* 25**,** 1881-1894.

DARNELL, D. K., KAUR, S., STANISLAW, S., KONIECZKA, J. H., YATSKIEVYCH, T. A. & ANTIN, P. B. 2006. MicroRNA expression during chick embryo development. *Dev Dyn,* 235**,** 3156-65.

DARWISH, A. S., GRADY, L. M., BAI, P. & WELLER, S. K. 2016. ICP8 Filament Formation Is Essential for Replication Compartment Formation during Herpes Simplex Virus Infection. *Journal of Virology,* 90**,** 2561-2570.

DAVIDSON, A. 2007. *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis,* Cambridge, Cambridge University Press.

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**,** 171-7.

DEFRA 2018. United Kingdom Poultry and Poultry Meat Statistics – July 2018. London: DEFRA.

References

DESAI, P. J. 2000. A null mutation in the UL36 gene of herpes simplex virus type 1 results in accumulation of unenveloped DNA-filled capsids in the cytoplasm of infected cells. *J Virol,* 74**,** 11608-18.

DEVLIN, J. M., BROWNING, G. F., HARTLEY, C. A. & GILKERSON, J. R. 2007. Glycoprotein G deficient infectious laryngotracheitis virus is a candidate attenuated vaccine. *Vaccine,* 25**,** 3561-3566.

DEVLIN, J. M., BROWNING, G. F., HARTLEY, C. A., KIRKPATRICK, N. C., MAHMOUDIAN, A., NOORMOHAMMADI, A. H. & GILKERSON, J. R. 2006. Glycoprotein G is a virulence factor in infectious laryngotracheitis virus. *Journal of General Virology,* 87**,** 2839-2847.

DEVLIN, J. M., HARTLEY, C. A., GILKERSON, J. R., COPPO, M. J. C., VAZ, P., NOORMOHAMMADI, A. H., WELLS, B., RUBITE, A., DHAND, N. K. & BROWNING, G. F. 2011. Horizontal transmission dynamics of a glycoprotein G deficient candidate vaccine strain of infectious laryngotracheitis virus and the effect of vaccination on transmission of virulent virus. *Vaccine,* 29**,** 5699-5704.

DEVLIN, J. M., VIEJO-BORBOLLA, A., BROWNING, G. F., NOORMOHAMMADI, A. H., GILKERSON, J. R., ALCAMI, A. & HARTLEY, C. A. 2010. Evaluation of immunological responses to a glycoprotein G deficient candidate vaccine strain of infectious laryngotracheitis virus. *Vaccine,* 28**,** 1325-1332.

DOELKEN, L., MALTERER, G., ERHARD, F., KOTHE, S., FRIEDEL, C. C., SUFFERT, G., MARCINOWSKI, L., MOTSCH, N., BARTH, S., BEITZINGER, M., LIEBER, D., BAILER, S. M., HOFFMANN, R., RUZSICS, Z., KREMMER, E., PFEFFER, S., ZIMMER, R., KOSZINOWSKI, U. H., GRASSER, F., MEISTER, G. & HAAS, J. 2010. Systematic Analysis of Viral and Cellular MicroRNA Targets in Cells Latently Infected with Human gamma-Herpesviruses by RISC Immunoprecipitation Assay. *Cell Host & Microbe,* 7**,** 324-334.

DOENCH, J. G., FUSI, N., SULLENDER, M., HEGDE, M., VAIMBERG, E. W., DONOVAN, K. F., SMITH, I., TOTHOVA, Z., WILEN, C., ORCHARD, R., VIRGIN, H. W., LISTGARTEN, J. & ROOT, D. E. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol,* 34**,** 184-191.

DOHNER, K. & SODEIK, B. 2005. The role of the cytoskeleton during viral infection. *Curr Top Microbiol Immunol,* 285**,** 67-108.

DUAN, F., LIAO, J., HUANG, Q., NIE, Y. & WU, K. 2012. HSV-1 miR-H6 Inhibits HSV-1 Replication and IL-6 Expression in Human Corneal Epithelial Cells In Vitro. *Clinical and Developmental Immunology,* 2012**,** 8.

EBINA, H., MISAWA, N., KANEMURA, Y. & KOYANAGI, Y. 2013. Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. *Scientific Reports,* 3**,** 2510.

ELLIOTT, R. M., BATESON, A. & KELLY, D. C. 1980. Phosphonoacetic Acid Inhibition of Frog Virus 3 Replication. *Journal of Virology,* 33**,** 539-542.

ELLIS-CONNELL, A. L., IEMPRIDEE, T., XU, I. & MERTZ, J. E. 2010. Cellular MicroRNAs 200b and 429 Regulate the Epstein-Barr Virus Switch between Latency and Lytic Replication. *Journal of Virology,* 84**,** 10329-10343.

ENQUIST, L. W., HUSAK, P. J., BANFIELD, B. W. & SMITH, G. A. 1998. Infection and spread of alphaherpesviruses in the nervous system. *Adv Virus Res,* 51**,** 237-347.

ESNAULT, E., BONSERGENT, C., LARCHER, T., BED'HOM, B., VAUTHEROT, J.-F., DELALEU, B., GUIGAND, L., SOUBIEUX, D., MARC, D. & QUÉRÉ, P. 2011. A novel chicken lung epithelial cell line: Characterization and response to low pathogenicity avian influenza virus. *Virus Research,* 159**,** 32-42.

EULALIO, A., BEHM-ANSMANT, I., SCHWEIZER, D. & IZAURRALDE, E. 2007. P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Molecular and cellular biology,* 27**,** 3970-3981.

References

FENNER, F. J., GIBBS, E. P. J., MURPHY, F. A., ROTT, R., STUDDERT, M. J. & WHITE, D. O. 2011. *Fenner's Veterinary Virology,* Academic Press, London, Elsevier.

FLEISCHMANN, W. 1996. *Medical Microbiology,* Galveston (TX), University of Texas Medical branch at Galveston.

FOSSUM, E., FRIEDEL, C. C., RAJAGOPALA, S. V., TITZ, B., BAIKER, A., SCHMIDT, T., KRAUS, T., STELLBERGER, T., RUTENBERG, C., SUTHRAM, S., BANDYOPADHYAY, S., ROSE, D., VON BRUNN, A., UHLMANN, M., ZERETZKE, C., DONG, Y. A., BOULET, H., KOEGL, M., BAILER, S. M., KOSZINOWSKI, U., IDEKER, T., UETZ, P., ZIMMER, R. & HAAS, J. 2009. Evolutionarily conserved herpesviral protein interaction networks. *PLoS Pathog,* 5**,** e1000570.

FU, B., SUN, F., LI, B., YANG, L., ZENG, Y., SUN, X., XU, F., RAYNER, S., GUADALUPE, M., GAO, S.-J. & WANG, L. 2009. Seroprevalence of Kaposi's Sarcoma-Associated Herpesvirus and Risk Factors in Xinjiang, China. *Journal of medical virology,* 81**,** 1422-1431.

FU, M., WANG, B., CHEN, X., HE, Z., WANG, Y., LI, X., CAO, H. & ZHENG, S. J. 2017. gga-miR-130b suppresses infectious bursal disease virus replication via targeting the viral genome and cellular SOCS5. *Journal of Virology*.

FUCHS, W., GRANZOW, H., KLUPP, B. G., KOPP, M. & METTENLEITER, T. C. 2002a. The UL48 tegument protein of pseudorabies virus is critical for intracytoplasmic assembly of infectious virions. *Journal of Virology,* 76**,** 6729-6742.

FUCHS, W., GRANZOW, H. & METTENLEITER, T. C. 2003. A pseudorabies virus recombinant simultaneously lacking the major tegument proteins encoded by the UL46, UL47, UL48, an UL49 genes is viable in cultured cells. *Journal of Virology,* 77**,** 12891-12900.

FUCHS, W., KLUPP, B. G., GRANZOW, H., HENGARTNER, C., BRACK, A., MUNDT, A., ENQUIST, L. W. & METTENLEITER, T. C. 2002b. Physical interaction between envelope glycoproteins E and M of pseudorabies virus and the major tegument protein UL49. *J Virol,* 76**,** 8208-17.

FUCHS, W., VEITS, J., HELFERICH, D., GRANZOW, H., TEIFKE, J. P. & METTENLEITER, T. C. 2007. Molecular biology of avian infectious laryngotracheitis virus. *Veterinary Research,* 38**,** 261-279.

FUKAYA, T., IWAKAWA, H.-O. & TOMARI, Y. 2014. MicroRNAs Block Assembly of eIF4F Translation Initiation Complex in Drosophila. *Molecular Cell,* 56**,** 67-78.

GALLAHER, A. M., DAS, S., XIAO, Z., ANDRESSON, T., KIEFFER-KWON, P., HAPPEL, C. & ZIEGELBAUER, J. 2013. Proteomic screening of human targets of viral microRNAs reveals functions associated with immune evasion and angiogenesis. *PLoS Pathog,* 9**,** e1003584.

GARCIA, M., SPATZ, S. J., CHENG, Y., RIBLET, S. M., VOLKENING, J. D. & SCHNEIDERS, G. H. 2016. Attenuation and protection efficacy of ORF C gene-deleted recombinant of infectious laryngotracheitis virus. *Journal of General Virology,* 97**,** 2352-2362.

GARCIA, M., VOLKENING, J., RIBLET, S. & SPATZ, S. 2013. Genomic sequence analysis of the United States infectious laryngotracheitis vaccine strains chicken embryo origin (CEO) and tissue culture origin (TCO). *Virology,* 440**,** 64-74.

GARNEAU, J. E., DUPUIS, M.-È., VILLION, M., ROMERO, D. A., BARRANGOU, R., BOYAVAL, P., FREMAUX, C., HORVATH, P., MAGADÁN, A. H. & MOINEAU, S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature,* 468**,** 67.

GASIUNAS, G., BARRANGOU, R., HORVATH, P. & SIKSNYS, V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A,* 109**,** E2579-86.

References

GAY, L. A., SETHURAMAN, S., THOMAS, M., TURNER, P. C. & RENNE, R. 2018. Modified Cross-Linking, Ligation, and Sequencing of Hybrids (qCLASH) Identifies Kaposi's Sarcoma-Associated Herpesvirus MicroRNA Targets in Endothelial Cells. *J Virol,* 92.

GLAZOV, E. A., COTTEE, P. A., BARRIS, W. C., MOORE, R. J., DALRYMPLE, B. P. & TIZARD, M. L. 2008. A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. *Genome Research,* 18**,** 957-964.

GLOVER, J. D. & MCGREW, M. J. 2012. Primordial Germ Cell Technologies for Avian Germplasm Cryopreservation and Investigating Germ Cell Development. *The Journal of Poultry Science,* 49**,** 155-162.

GONCALVES, P. H., ZIEGELBAUER, J., ULDRICK, T. S. & YARCHOAN, R. 2017. Kaposi sarcoma herpesvirus-associated cancers and related diseases. *Curr Opin HIV AIDS,* 12**,** 47-56.

GOODRUM, F., CAVINESS, K. & ZAGALLO, P. 2012. Human cytomegalovirus persistence. *Cell Microbiol,* 14**,** 644-55.

GOTTWEIN, E. 2012. Kaposi's Sarcoma-Associated Herpesvirus microRNAs. *Front Microbiol,* 3**,** 165.

GOTTWEIN, E., MUKHERJEE, N., SACHSE, C., FRENZEL, C., MAJOROS, W. H., CHI, J. T. A., BRAICH, R., MANOHARAN, M., SOUTSCHEK, J., OHLER, U. & CULLEN, B. R. 2007. A viral microRNA functions as an orthologue of cellular miR-155. *Nature,* 450**,** 1096-U17.

GRAHAM, F. L., SMILEY, J., RUSSELL, W. C. & NAIRN, R. 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol,* 36**,** 59-74.

GRECO, A., SIMONIN, D., DIAZ, J. J., BARJHOUX, L., KINDBEITER, K., MADJAR, J. J. & MASSE, T. 1994. THE DNA-SEQUENCE CODING FOR THE 5' UNTRANSLATED REGION OF HERPES-SIMPLEX VIRUS TYPE-1 ICP22 MESSENGER-RNA MEDIATES A HIGH-LEVEL OF GENE-EXPRESSION. *Journal of General Virology,* 75**,** 1693-1702.

GREGORY, R. I., CHENDRIMADA, T. P., COOCH, N. & SHIEKHATTAR, R. 2005. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell,* 123**,** 631-640.

GREY, F. 2015. Role of microRNAs in herpesvirus latency and persistence. *Journal of General Virology,* 96**,** 739-751.

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 Pathogens,* 3**,** 1593-1602.

GREY, F., TIRABASSI, R., MEYERS, H., WU, G., MCWEENEY, S., HOOK, L. & NELSON, J. A. 2010. A Viral microRNA Down-Regulates Multiple Cell Cycle Genes through mRNA 5 ' UTRs. *Plos Pathogens,* 6.

GRIFFITHS-JONES, S., GROCOCK, R. J., VAN DONGEN, S., BATEMAN, A. & ENRIGHT, A. J. 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research,* 34**,** D140-D144.

GRINDE, B. 2013. Herpesviruses: latency and reactivation – viral strategies and host response. *Journal of Oral Microbiology,* 5**,** 10.3402/jom.v5i0.22766.

GUO, Y. E., OEI, T. & STEITZ, J. A. 2015. Herpesvirus saimiri MicroRNAs Preferentially Target Host Cell Cycle Regulators. *J Virol,* 89**,** 10901-11.

GUO, Z., MAKI, M., DING, R., YANG, Y., ZHANG, B. & XIONG, L. 2014. Genome-wide survey of tissue-specific microRNA and transcription factor regulatory networks in 12 tissues. *Scientific Reports,* 4**,** 5150.

GUY, J. S., BARNES, H. J. & MORGAN, L. M. 1990. VIRULENCE OF INFECTIOUS LARYNGOTRACHEITIS VIRUSES - COMPARISON OF MODIFIED-LIVE

VACCINE VIRUSES AND NORTH-CAROLINA FIELD ISOLATES. *Avian Diseases,* 34**,** 106-113.

HA, M. & KIM, V. N. 2014. Regulation of microRNA biogenesis. *Nature Reviews Molecular Cell Biology,* 15**,** 509-524.

HA, T.-Y. 2011. MicroRNAs in Human Diseases: From Cancer to Cardiovascular Disease. *Immune network,* 11**,** 135-154.

HAFNER, M., LANDTHALER, M., BURGER, L., KHORSHID, M., HAUSSER, J., BERNINGER, P., ROTHBALLER, A., ASCANO, M., JUNGKAMP, A. C., MUNSCHAUER, M., ULRICH, A., WARDLE, G. S., DEWELL, S., ZAVOLAN, M. & TUSCHL, T. 2010. PAR-CliP--a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J Vis Exp*.

HAMBURGER, V. & HAMILTON, H. L. 1992. A series of normal stages in the development of the chick embryo. 1951. *Dev Dyn,* 195**,** 231-72.

HAN, W., ZHU, Y. F., SU, Y. J., LI, G. H., QU, L., ZHANG, H. Y., WANG, K. H., ZOU, J. M. & LIU, H. L. 2016. High-Throughput Sequencing Reveals Circulating miRNAs as Potential Biomarkers for Measuring Puberty Onset in Chicken (Gallus gallus). *Plos One,* 11**,** 13.

HANCOCK, M. H. S., R. L. 2017. *Roles of Non-coding RNAs During Herpesvirus Infection,* Berlin, Springer, Berlin, Heidelberg.

HANSEN, T. B., JENSEN, T. I., CLAUSEN, B. H., BRAMSEN, J. B., FINSEN, B., DAMGAARD, C. K. & KJEMS, J. 2013. Natural RNA circles function as efficient microRNA sponges. *Nature,* 495**,** 384-8.

HAROLD, C., COX, D. & RILEY, K. J. 2016. Epstein-Barr viral microRNAs target caspase 3. *Virology Journal,* 13**,** 145.

HAUN, W., COFFMAN, A., CLASEN, B. M., DEMOREST, Z. L., LOWY, A., RAY, E., RETTERATH, A., STODDARD, T., JUILLERAT, A., CEDRONE, F., MATHIS, L., VOYTAS, D. F. & ZHANG, F. 2014. Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. *Plant Biotechnol J,* 12**,** 934-40.

HAYLES, L. B., HAMILTON, D. & NEWBY, W. C. 1976. Transfer of parental immunity to infectious laryngotracheitis in chicks. *Can J Comp Med,* 40**,** 218-9.

HE, J., YANG, Q. & CHANG, L.-J. 2005. Dynamic DNA Methylation and Histone Modifications Contribute to Lentiviral Transgene Silencing in Murine Embryonic Carcinoma Cells. *Journal of Virology,* 79**,** 13497-13508.

HELFERICH, D., VEITS, J., TEIFKE, J. P., METTENLEITER, T. C. & FUCHS, W. 2007. The UL47 gene of avian infectious laryngotracheitis virus is not essential for in vitro replication but is relevant for virulence in chickens. *Journal of General Virology,* 88**,** 732-742.

HELWAK, A., KUDLA, G., DUDNAKOVA, T. & TOLLERVEY, D. 2013. Mapping the Human miRNA Interactome by CLASH Reveals Frequent Noncanonical Binding. *Cell,* 153**,** 654-665.

HELWAK, A. & TOLLERVEY, D. 2014. Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH). *Nature Protocols,* 9**,** 711-728.

HEROLD, B. C., VISALLI, R. J., SUSMARSKI, N., BRANDT, C. R. & SPEAR, P. G. 1994. Glycoprotein C-independent binding of herpes simplex virus to cells requires cell surface heparan sulphate and glycoprotein B. *J Gen Virol,* 75 (Pt 6)**,** 1211-22.

HILDEBRANDT, E., DUNN, J. R., PERUMBAKKAM, S., NIIKURA, M. & CHENG, H. H. 2014. Characterizing the molecular basis of attenuation of Marek's disease virus via in vitro serial passage identifies de novo mutations in the helicase-primase subunit gene UL5 and other candidates associated with reduced virulence. *J Virol,* 88**,** 6232-42.

HIMLY, M., FOSTER, D. N., BOTTOLI, I., IACOVONI, J. S. & VOGT, P. K. 1998. The DF-1 Chicken Fibroblast Cell Line: Transformation Induced by Diverse Oncogenes

and Cell Death Resulting from Infection by Avian Leukosis Viruses. *Virology,* 248**,** 295-304.

HON, C. C., RAMILOWSKI, J. A., HARSHBARGER, J., BERTIN, N., RACKHAM, O. J., GOUGH, J., DENISENKO, E., SCHMEIER, S., POULSEN, T. M., SEVERIN, J., LIZIO, M., KAWAJI, H., KASUKAWA, T., ITOH, M., BURROUGHS, A. M., NOMA, S., DJEBALI, S., ALAM, T., MEDVEDEVA, Y. A., TESTA, A. C., LIPOVICH, L., YIP, C. W., ABUGESSAISA, I., MENDEZ, M., HASEGAWA, A., TANG, D., LASSMANN, T., HEUTINK, P., BABINA, M., WELLS, C. A., KOJIMA, S., NAKAMURA, Y., SUZUKI, H., DAUB, C. O., DE HOON, M. J., ARNER, E., HAYASHIZAKI, Y., CARNINCI, P. & FORREST, A. R. 2017. An atlas of human long non-coding RNAs with accurate 5' ends. *Nature,* 543**,** 199-204.

HOPCRAFT, S. E., PATTENDEN, S. G., JAMES, L. I., FRYE, S., DITTMER, D. P. & DAMANIA, B. 2018. Chromatin remodeling controls Kaposi's sarcoma-associated herpesvirus reactivation from latency. *PLOS Pathogens,* 14**,** e1007267.

HSU, J. B., CHIU, C. M., HSU, S. D., HUANG, W. Y., CHIEN, C. H., LEE, T. Y. & HUANG, H. D. 2011. miRTar: an integrated system for identifying miRNA-target interactions in human. *BMC Bioinformatics,* 12**,** 300.

HSU, P. D., LANDER, E. S. & ZHANG, F. 2014. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell,* 157**,** 1262-1278.

HUANG, J. T., CHEN, J. N., GONG, L. P., BI, Y. H., LIANG, J., ZHOU, L., HE, D. & SHAO, C. K. 2019. Identification of virus-encoded circular RNA. *Virology,* 529**,** 144-151.

HUANG, Y., ZOU, Q., SONG, H., SONG, F., WANG, L., ZHANG, G. & SHEN, X. 2010. A study of miRNAs targets prediction and experimental validation. *Protein & Cell,* 1**,** 979-986.

HUBSTENBERGER, A., COUREL, M., BÉNARD, M., SOUQUERE, S., ERNOULT-LANGE, M., CHOUAIB, R., YI, Z., MORLOT, J.-B., MUNIER, A., FRADET, M., DAUNESSE, M., BERTRAND, E., PIERRON, G., MOZZICONACCI, J., KRESS, M. & WEIL, D. 2017. P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. *Molecular Cell,* 68**,** 144-157.e5.

HUGHES, C. S. & JONES, R. C. 1988. Comparison of cultural methods for primary isolation of infectious laryngotracheitis virus from field material. *Avian Pathol,* 17**,** 295-303.

ISHINO, Y., SHINAGAWA, H., MAKINO, K., AMEMURA, M. & NAKATA, A. 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol,* 169**,** 5429-33.

IVEY, K. N., MUTH, A., ARNOLD, J., KING, F. W., YEH, R. F., FISH, J. E., HSIAO, E. C., SCHWARTZ, R. J., CONKLIN, B. R., BERNSTEIN, H. S. & SRIVASTAVA, D. 2008. MicroRNA regulation of cell lineages in mouse and human embryonic stem cells. *Cell Stem Cell,* 2**,** 219-29.

JANSEN, R., EMBDEN, J. D., GAASTRA, W. & SCHOULS, L. M. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol,* 43**,** 1565-75.

JAVANBAKHT, H., HALWANI, R., CEN, S., SAADATMAND, J., MUSIER-FORSYTH, K., GOTTLINGER, H. & KLEIMAN, L. 2003. The interaction between HIV-1 Gag and human lysyl-tRNA synthetase during viral assembly. *J Biol Chem,* 278**,** 27644-51.

JIN, H. Y. & XIAO, C. 2015. MicroRNA Mechanisms of Action: What have We Learned from Mice? *Front Genet,* 6**,** 328.

JOHN, B., ENRIGHT, A. J., ARAVIN, A., TUSCHL, T., SANDER, C. & MARKS, D. S. 2004. Human MicroRNA targets. *PLoS Biol,* 2**,** e363.

References

JOPLING, CATHERINE L. 2008. Regulation of hepatitis C virus by microRNA-122. *Biochemical Society Transactions,* 36**,** 1220-1223.

JOPLING, C. L., SCHÜTZ, S. & SARNOW, P. 2008. Position-Dependent Function for a Tandem MicroRNA miR-122-Binding Site Located in the Hepatitis C Virus RNA Genome. *Cell Host & Microbe,* 4**,** 77-85.

JOPLING, C. L., YI, M., LANCASTER, A. M., LEMON, S. M. & SARNOW, P. 2005. Modulation of Hepatitis C Virus RNA Abundance by a Liver-Specific MicroRNA. *Science,* 309**,** 1577-1581.

KATO, K., DAIKOKU, T., GOSHIMA, F., KUME, H., YAMAKI, K. & NISHIYAMA, Y. 2000. Synthesis, subcellular localization and VP16 interaction of the herpes simplex virus type 2 UL46 gene product. *Archives of Virology,* 145**,** 2149-2162.

KAWAGUCHI, T., NOMURA, K., HIRAYAMA, Y. & KITAGAWA, T. 1987. Establishment and characterization of a chicken hepatocellular carcinoma cell line, LMH. *Cancer Res,* 47**,** 4460-4.

KIM, J. Y., MANDARINO, A., CHAO, M. V., MOHR, I. & WILSON, A. C. 2012. Transient reversal of episome silencing precedes VP16-dependent transcription during reactivation of latent HSV-1 in neurons. *PLoS Pathog,* 8**,** e1002540.

KIM, V. N. 2005. MicroRNA biogenesis: Coordinated cropping and dicing. *Nature Reviews Molecular Cell Biology,* 6**,** 376-385.

KIM, V. N., HAN, J. & SIOMI, M. C. 2009. Biogenesis of small RNAs in animals. *Nature Reviews Molecular Cell Biology,* 10**,** 126-139.

KIM, Y. G., CHA, J. & CHANDRASEGARAN, S. 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 1156-1160.

KIMPLE, M. E., BRILL, A. L. & PASKER, R. L. 2013. Overview of affinity tags for protein purification. *Curr Protoc Protein Sci,* 73**,** Unit 9.9.

KINCAID, R. P., CHEN, Y., COX, J. E., RETHWILM, A. & SULLIVAN, C. S. 2014. Noncanonical MicroRNA (miRNA) Biogenesis Gives Rise to Retroviral Mimics of Lymphoproliferative and Immunosuppressive Host miRNAs. *mBio,* 5.

KINCAID, R. P. & SULLIVAN, C. S. 2012. Virus-Encoded microRNAs: An Overview and a Look to the Future. *Plos Pathogens,* 8.

KING, A. M. Q., LEFKOWITZ, E. J., MUSHEGIAN, A. R., ADAMS, M. J., DUTILH, B. E., GORBALENYA, A. E., HARRACH, B., HARRISON, R. L., JUNGLEN, S., KNOWLES, N. J., KROPINSKI, A. M., KRUPOVIC, M., KUHN, J. H., NIBERT, M. L., RUBINO, L., SABANADZOVIC, S., SANFACON, H., SIDDELL, S. G., SIMMONDS, P., VARSANI, A., ZERBINI, F. M. & DAVISON, A. J. 2018. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2018). *Arch Virol*.

KINGSBURY, F. W. & JUNGHERR, E. L. 1958. Indirect Transmission of Infectious Laryngotracheitis in Chickens. *Avian Diseases,* 2**,** 54-63.

KIRKPATRICK, N. C., MAHMOUDIAN, A., COLSON, C. A., DEVLIN, J. M. & NOORMOHAMMADI, A. H. 2006. Relationship between mortality, clinical signs and tracheal pathology in infectious laryngotracheitis. *Avian Pathol,* 35**,** 449-53.

KONG, Y. W., CANNELL, I. G., DE MOOR, C. H., HILL, K., GARSIDE, P. G., HAMILTON, T. L., MEIJER, H. A., DOBBYN, H. C., STONELEY, M., SPRIGGS, K. A., WILLIS, A. E. & BUSHELL, M. 2008. The mechanism of micro-RNA-mediated translation repression is determined by the promoter of the target gene. *Proc Natl Acad Sci U S A,* 105**,** 8866-71.

KOPP, M., KLUPP, B. G., GRANZOW, H., FUCHS, W. & METTENLEITER, T. C. 2002. Identification and characterization of the pseudorabies virus tegument proteins UL46 and UL47: Role for UL47 in virion morphogenesis in the cytoplasm. *Journal of Virology,* 76**,** 8820-8833.

References

KOSKI, D. M., PREDGEN, A. S., TRAMPEL, D. W., CONRAD, S. K., NARWOLD, D. R. & HERMANN, J. R. 2015. Comparison of the pathogenicity of the USDA challenge virus strain to a field strain of infectious laryngotracheitis virus. *Biologicals,* 43**,** 232-237.

KUDLA, G., GRANNEMAN, S., HAHN, D., BEGGS, J. D. & TOLLERVEY, D. 2011. Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 10010-10015.

LABOISSIÈRE, S. & O'HARE, P. 2000. Analysis of HCF, the Cellular Cofactor of VP16, in Herpes Simplex Virus-Infected Cells. *Journal of Virology,* 74**,** 99-109.

LAGOS-QUINTANA, M., RAUHUT, R., YALCIN, A., MEYER, J., LENDECKEL, W. & TUSCHL, T. 2002. Identification of tissue-specific microRNAs from mouse. *Curr Biol,* 12**,** 735-9.

LARSON, M. H., GILBERT, L. A., WANG, X., LIM, W. A., WEISSMAN, J. S. & QI, L. S. 2013. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc,* 8**,** 2180-96.

LECELLIER, C. H., DUNOYER, P., ARAR, K., LEHMANN-CHE, J., EYQUEM, S., HIMBER, C., SAIB, A. & VOINNET, O. 2005. A cellular microRNA mediates antiviral defense in human cells. *Science,* 308**,** 557-60.

LEE, J. Y., SONG, J. J., WOOMING, A., LI, X. Y., ZHOU, H. J., BOTTJE, W. G. & KONG, B. W. 2010. Transcriptional profiling of host gene expression in chicken embryo lung cells infected with laryngotracheitis virus. *Bmc Genomics,* 11**,** 15.

LEE, N. S., DOHJIMA, T., BAUER, G., LI, H. T., LI, M. J., EHSANI, A., SALVATERRA, P. & ROSSI, J. 2002. Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. *Nature Biotechnology,* 20**,** 500-505.

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. W., MARKHAM, P. F., COPPO, M. J. C., LEGIONE, A. R., MARKHAM, J. F., NOORMOHAMMADI, A. H., BROWNING, G. F., FICORILLI, N., HARTLEY, C. A. & DEVLIN, J. M. 2012. Attenuated Vaccines Can Recombine to Form Virulent Field Viruses. *Science,* 337**,** 188-188.

LEE, Y., HUR, I., PARK, S. Y., KIM, Y. K., SUH, M. R. & KIM, V. N. 2006. The role of PACT in the RNA silencing pathway. *Embo Journal,* 25**,** 522-532.

LEIB, D. A., COEN, D. M., BOGARD, C. L., HICKS, K. A., YAGER, D. R., KNIPE, D. M., TYLER, K. L. & SCHAFFER, P. A. 1989. Immediate-early regulatory gene mutants define different stages in the establishment and reactivation of herpes simplex virus latency. *J Virol,* 63**,** 759-68.

LEVY, C. S. 2011. Identification and Characterisation of Ovine Herpesvirus 2 microRNAs. *PhD Thesis University of Edinburgh*.

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.

LILLYCROP, K. A. & LATCHMAN, D. S. 1992. Alternative splicing of the Oct-2 transcription factor RNA is differentially regulated in neuronal cells and B cells and results in protein isoforms with opposite effects on the activity of octamer/TAATGARAT-containing promoters. *J Biol Chem,* 267**,** 24960-5.

LIN, C., LI, H., HAO, M., XIONG, D., LUO, Y., HUANG, C., YUAN, Q., ZHANG, J. & XIA, N. 2016. Increasing the Efficiency of CRISPR/Cas9-mediated Precise Genome Editing of HSV-1 Virus in Human Cells. *Sci Rep,* 6**,** 34531.

LIN, Z., WANG, X., FEWELL, C., CAMERON, J., YIN, Q. & FLEMINGTON, E. K. 2010. Differential expression of the miR-200 family microRNAs in epithelial and B cells

and regulation of Epstein-Barr virus reactivation by the miR-200 family member miR-429. *J Virol,* 84**,** 7892-7.

LIU, X.-F., WANG, X., YAN, S., ZHANG, Z., ABECASSIS, M. & HUMMEL, M. 2013. Epigenetic Control of Cytomegalovirus Latency and Reactivation. *Viruses,* 5**,** 1325-1345.

LIU, X., LI, D., ZHANG, W., GUO, M. & ZHAN, Q. 2012. Long non-coding RNA gadd7 interacts with TDP-43 and regulates Cdk6 mRNA decay. *Embo j,* 31**,** 4415-27.

LIVAK, K. J. & SCHMITTGEN, T. D. 2001a. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods,* 25**,** 402-8.

LIVAK, K. J. & SCHMITTGEN, T. D. 2001b. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods,* 25**,** 402-408.

LIZASA, H., WULFF, B. E., ALLA, N. R., MARAGKAKIS, M., MEGRAW, M., HATZIGEORGIOU, A., IWAKIRI, D., TAKADA, K., WIEDMER, A., SHOWE, L., LIEBERMAN, P. & NISHIKURA, K. 2010. Editing of Epstein-Barr Virus-encoded BART6 MicroRNAs Controls Their Dicer Targeting and Consequently Affects Viral Latency. *Journal of Biological Chemistry,* 285**,** 33358-33370.

LORET, S., GUAY, G. & LIPPE, R. 2008. Comprehensive characterization of extracellular herpes simplex virus type 1 virions. *J Virol,* 82**,** 8605-18.

LORET, S. & LIPPE, R. 2012. Biochemical analysis of infected cell polypeptide (ICP)0, ICP4, UL7 and UL23 incorporated into extracellular herpes simplex virus type 1 virions. *J Gen Virol,* 93**,** 624-34.

LOUDOVARIS, T., CALNEK, B. W., YOO, B. H. & FAHEY, K. J. 1991a. GENETIC SUSCEPTIBILITY OF CHICKEN MACROPHAGES TO INVITRO INFECTION WITH INFECTIOUS LARYNGOTRACHEITIS VIRUS. *Avian Pathology,* 20**,** 291-302.

LOUDOVARIS, T., YOO, B. H. & FAHEY, K. J. 1991b. GENETIC-RESISTANCE TO INFECTIOUS LARYNGOTRACHEITIS IN INBRED LINES OF WHITE LEGHORN CHICKENS. *Avian Pathology,* 20**,** 357-361.

LU, C. C., LI, Z. H., CHU, C. Y., FENG, J. Y., FENG, J., SUN, R. & RANA, T. M. 2010. MicroRNAs encoded by Kaposi's sarcoma-associated herpesvirus regulate viral life cycle. *Embo Reports,* 11**,** 784-790.

LYTLE, J. R., YARIO, T. A. & STEITZ, J. A. 2007. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3′ UTR. *Proceedings of the National Academy of Sciences,* 104**,** 9667-9672.

MACFARLANE, L. A. & MURPHY, P. R. 2010. MicroRNA: Biogenesis, Function and Role in Cancer. *Curr Genomics,* 11**,** 537-61.

MACHITANI, M., SAKURAI, F., WAKABAYASHI, K., NAKATANI, K., TACHIBANA, M. & MIZUGUCHI, H. 2017. MicroRNA miR-27 Inhibits Adenovirus Infection by Suppressing the Expression of SNAP25 and TXN2. *Journal of Virology*.

MAEDA, E., AKAHANE, M., KIRYU, S., KATO, N., YOSHIKAWA, T., HAYASHI, N., AOKI, S., MINAMI, M., UOZAKI, H., FUKAYAMA, M. & OHTOMO, K. 2009. Spectrum of Epstein-Barr virus-related diseases: a pictorial review. *Jpn J Radiol,* 27**,** 4-19.

MAHMOUDIAN, A., MARKHAM, P. F., NOORMOHAMMADI, A. H. & BROWNING, G. F. 2012. Kinetics of transcription of infectious laryngotracheitis virus genes. *Comparative Immunology Microbiology and Infectious Diseases,* 35**,** 103-115.

MAKAROVA, K. S., WOLF, Y. I., ALKHNBASHI, O. S., COSTA, F., SHAH, S. A., SAUNDERS, S. J., BARRANGOU, R., BROUNS, S. J., CHARPENTIER, E., HAFT, D. H., HORVATH, P., MOINEAU, S., MOJICA, F. J., TERNS, R. M., TERNS, M. P., WHITE, M. F., YAKUNIN, A. F., GARRETT, R. A., VAN DER

OOST, J., BACKOFEN, R. & KOONIN, E. V. 2015. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol,* 13**,** 722-36.

MAKHOV, A. M., LEE, S. S.-K., LEHMAN, I. R. & GRIFFITH, J. D. 2003. Origin-specific unwinding of herpes simplex virus 1 DNA by the viral UL9 and ICP8 proteins: Visualization of a specific preunwinding complex. 100**,** 898-903.

MALI, P., YANG, L., ESVELT, K. M., AACH, J., GUELL, M., DICARLO, J. E., NORVILLE, J. E. & CHURCH, G. M. 2013. RNA-Guided Human Genome Engineering via Cas9. *Science (New York, N.Y.),* 339**,** 823-826.

MALLORY, A. C. & VAUCHERET, H. 2004. MicroRNAs: something important between the genes. *Current Opinion in Plant Biology,* 7**,** 120-125.

MCGEOCH, D. J., DALRYMPLE, M. A., DAVISON, A. J., DOLAN, A., FRAME, M. C., MCNAB, D., PERRY, L. J., SCOTT, J. E. & TAYLOR, P. 1988. The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. *J Gen Virol,* 69 (Pt 7)**,** 1531-74.

MCGEOCH, D. J., RIXON, F. J. & DAVISON, A. J. 2006. Topics in herpesvirus genomics and evolution. *Virus Research,* 117**,** 90-104.

MEMCZAK, S., JENS, M., ELEFSINIOTI, A., TORTI, F., KRUEGER, J., RYBAK, A., MAIER, L., MACKOWIAK, S. D., GREGERSEN, L. H., MUNSCHAUER, M., LOEWER, A., ZIEBOLD, U., LANDTHALER, M., KOCKS, C., LE NOBLE, F. & RAJEWSKY, N. 2013. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature,* 495**,** 333-8.

MENENDEZ, K. R., GARCIA, M., SPATZ, S. & TABLANTE, N. L. 2014. Molecular epidemiology of infectious laryngotracheitis: a review. *Avian Pathology,* 43**,** 108-117.

MESSERLE, M., CRNKOVIC, I., HAMMERSCHMIDT, W., ZIEGLER, H. & KOSZINOWSKI, U. H. 1997. Cloning and mutagenesis of a herpesvirus genome as an infectious bacterial artificial chromosome. *Proceedings of the National Academy of Sciences of the USA,* 94**,** 14759-63.

METTENLEITER, T. C. 2002. Herpesvirus assembly and egress. *J Virol,* 76**,** 1537-47.

MEULEMANS, G. & HALEN, P. 1978. A comparison of three methods of diagnosis of infectious laryngotracheitis. *Avian Pathol,* 7**,** 433-6.

MINTEL 2017. Unprocessed Poultry and Red meat - UK. 1/11/17 ed. London: Mintel.

MOCARSKI, E. S., POST, L. E. & ROIZMAN, B. 1980. Molecular engineering of the herpes simplex virus genome: insertion of a second L-S junction into the genome causes additional genome inversions. *Cell,* 22**,** 243-55.

MOROZOVA, N., ZINOVYEV, A., NONNE, N., PRITCHARD, L. L., GORBAN, A. N. & HAREL-BELLAN, A. 2012. Kinetic signatures of microRNA modes of action. *Rna,* 18**,** 1635-55.

MULLER, H. J. 1927. ARTIFICIAL TRANSMUTATION OF THE GENE. *Science,* 66**,** 84-7.

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**,** 376-85.

NADIMPALLI, M., LEE, S. W., DEVLIN, J. M., GILKERSON, J. R. & HARTLEY, C. A. 2017. Impairment of infectious laryngotracheitis virus replication by deletion of the UL[-1] gene. *Arch Virol,* 162**,** 1541-1548.

NAVARRO, D. 2016. Expanding role of cytomegalovirus as a human pathogen. *J Med Virol,* 88**,** 1103-12.

NEFF, C., SUDLER, C. & HOOP, R. K. 2008. Characterization of Western European Field Isolates and Vaccine Strains of Avian Infectious Laryngotracheitis Virus by Restriction Fragment Length Polymorphism and Sequence Analysis. *Avian Diseases,* 52**,** 278-283.

References

NICOLL, M. P., HANN, W., SHIVKUMAR, M., HARMAN, L. E., CONNOR, V., COLEMAN, H. M., PROENCA, J. T. & EFSTATHIOU, S. 2016. The HSV-1 Latency-Associated Transcript Functions to Repress Latent Phase Lytic Gene Expression and Suppress Virus Reactivation from Latently Infected Neurons. *PLoS Pathog,* 12**,** e1005539.

NIGHTINGALE, K. 2016. *Validation and functional analysis of Ovine Herpesvirus 2-encoded microRNAs.* PhD, University of Edinburgh.

O'BRIEN, J., HAYDER, H., ZAYED, Y. & PENG, C. 2018. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Frontiers in endocrinology,* 9**,** 402-402.

O'CONNOR, C. M., VANICEK, J. & MURPHY, E. A. 2014. Host MicroRNA Regulation of Human Cytomegalovirus Immediate Early Protein Translation Promotes Viral Latency. *Journal of Virology,* 88**,** 5524-5532.

O'TOOLE, D. & LI, H. 2014. The Pathology of Malignant Catarrhal Fever, With an Emphasis on Ovine Herpesvirus 2. *Veterinary Pathology,* 51**,** 437-452.

OECD, FOOD & NATIONS, A. O. O. T. U. 2017. OECD-FAO Agricultural Outlook (Edition 2017).

OISHI, I., YOSHII, K., MIYAHARA, D., KAGAMI, H. & TAGAMI, T. 2016. Targeted mutagenesis in chicken using CRISPR/Cas9 system. *Scientific Reports,* 6**,** 23980.

OLDONI, I., RODRIGUEZ-AVILA, A., RIBLET, S. M., ZAVALA, G. & GARCIA, M. 2009. Pathogenicity and growth characteristics of selected infectious laryngotracheitis virus strains from the United States. *Avian Pathology,* 38**,** 47-53.

ORBERG, P. K. & SCHAFFER, P. A. 1987. Expression of herpes simplex virus type 1 major DNA-binding protein, ICP8, in transformed cell lines: complementation of deletion mutants and inhibition of wild-type virus. *Journal of Virology,* 61**,** 1136-1146.

OU, S.-C. & GIAMBRONE, J. J. 2012. Infectious laryngotracheitis virus in chickens. *World journal of virology,* 1**,** 142-9.

OUYANG, H., HE, X., LI, G., XU, H., JIA, X., NIE, Q. & ZHANG, X. 2015. Deep Sequencing Analysis of miRNA Expression in Breast Muscle of Fast-Growing and Slow-Growing Broilers. *Int J Mol Sci,* 16**,** 16242-62.

OWEN, D. J., CRUMP, C. M. & GRAHAM, S. C. 2015. Tegument Assembly and Secondary Envelopment of Alphaherpesviruses. *Viruses-Basel,* 7**,** 5084-5114.

PABO, C. O., PEISACH, E. & GRANT, R. A. 2001. Design and selection of novel Cys2His2 zinc finger proteins. *Annu Rev Biochem,* 70**,** 313-40.

PAN, D. L., FLORES, O., UMBACH, J. L., PESOLA, J. M., BENTLEY, P., ROSATO, P. C., LEIB, D. A., CULLEN, B. R. & COEN, D. M. 2014. A Neuron-Specific Host MicroRNA Targets Herpes Simplex Virus-1 ICP0 Expression and Promotes Latency. *Cell Host & Microbe,* 15**,** 446-456.

PAVELIN, J., REYNOLDS, N., CHIWESHE, S., WU, G. M., TIRIBASSI, R. & GREY, F. 2013. Systematic MicroRNA Analysis Identifies ATP6V0C as an Essential Host Factor for Human Cytomegalovirus Replication. *Plos Pathogens,* 9**,** 13.

PAVLOVA, S., VEITS, J., METTENLEITER, T. C. & FUCHS, W. 2013. Identification and Functional Analysis of Membrane Proteins gD, gE, gI, and pUS9 of Infectious Laryngotracheitis Virus. *Avian Diseases,* 57**,** 416-426.

PAVLOVA, S. P., VEITS, J., BLOHM, U., MARESCH, C., METTENLEITER, T. C. & FUCHS, W. 2010. In vitro and in vivo characterization of glycoprotein C-deleted infectious laryngotracheitis virus. *Journal of General Virology,* 91**,** 847-857.

PEDERSEN, I. M., CHENG, G., WIELAND, S., VOLINIA, S., CROCE, C. M., CHISARI, F. V. & DAVID, M. 2007. Interferon modulation of cellular microRNAs as an antiviral mechanism. *Nature,* 449**,** 919-22.

PELLET, P. & ROIZMANN, B. 2007. *Fields Virology 6th Edition,* Philadelphia, Lipincott Williams & Wilkins.

References

PEREZ, E. E., WANG, J., MILLER, J. C., JOUVENOT, Y., KIM, K. A., LIU, O., WANG, N., LEE, G., BARTSEVICH, V. V., LEE, Y. L., GUSCHIN, D. Y., RUPNIEWSKI, I., WAITE, A. J., CARPENITO, C., CARROLL, R. G., ORANGE, J. S., URNOV, F. D., REBAR, E. J., ANDO, D., GREGORY, P. D., RILEY, J. L., HOLMES, M. C. & JUNE, C. H. 2008. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol,* 26**,** 808-16.

PERKEL, J. M. 2013. Visiting "noncodarnia". *Biotechniques,* 54**,** 301, 303-4.

PERRY, A. K., CHEN, G., ZHENG, D., TANG, H. & CHENG, G. 2005. The host type I interferon response to viral and bacterial infections. *Cell Res,* 15**,** 407-22.

PETERSEN, C. P., BORDELEAU, M.-E., PELLETIER, J. & SHARP, P. A. 2006. Short RNAs Repress Translation after Initiation in Mammalian Cells. *Molecular Cell,* 21**,** 533-542.

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. C., RUSSO, J. J., JU, J. Y., RANDALL, G., LINDENBACH, B. D., RICE, C. M., SIMON, V., HO, D. D., ZAVOLAN, M. & TUSCHL, T. 2005. Identification of microRNAs of the herpesvirus family. *Nature Methods,* 2**,** 269-276.

PFEFFER, S., ZAVOLAN, M., GRÄSSER, F. A., CHIEN, M., RUSSO, J. J., JU, J., JOHN, B., ENRIGHT, A. J., MARKS, D., SANDER, C. & TUSCHL, T. 2004a. Identification of Virus-Encoded MicroRNAs. *Science,* 304**,** 734-736.

PFEFFER, S., ZAVOLAN, M., GRASSER, F. A., CHIEN, M. C., RUSSO, J. J., JU, J. Y., JOHN, B., ENRIGHT, A. J., MARKS, D., SANDER, C. & TUSCHL, T. 2004b. Identification of virus-encoded microRNAs. *Science,* 304**,** 734-736.

PICCIRILLO, A., LAVEZZO, E., NIERO, G., MORENO, A., MASSI, P., FRANCHIN, E., TOPPO, S., SALATA, C. & PALU, G. 2016. Full Genome Sequence-Based Comparative Study of Wild-Type and Vaccine Strains of Infectious Laryngotracheitis Virus from Italy. *Plos One,* 11.

PIMENOFF, V. N., HOULDCROFT, C. J., RIFKIN, R. F. & UNDERDOWN, S. 2018. The Role of aDNA in Understanding the Coevolutionary Patterns of Human Sexually Transmitted Infections. *Genes (Basel),* 9.

POURCEL, C., SALVIGNOL, G. & VERGNAUD, G. 2005. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology,* 151**,** 653-63.

PURCELL, D. A. 1971. The ultrastructural changes produced by infectious laryngotracheitis virus in tracheal epithelium of the fowl. *Res Vet Sci,* 12**,** 455-8.

QIN, J., LI, W., GAO, S. J. & LU, C. 2017. KSHV microRNAs: Tricks of the Devil. *Trends Microbiol,* 25**,** 648-661.

QIU, L., WANG, T., TANG, Q., LI, G., WU, P. & CHEN, K. 2018. Long Non-coding RNAs: Regulators of Viral Infection and the Interferon Antiviral Response. 9.

RACHAMADUGU, R., LEE, J. Y., WOOMING, A. & KONG, B. W. 2009. Identification and expression analysis of infectious laryngotracheitis virus encoding microRNAs. *Virus Genes,* 39**,** 301-308.

RAN, F. A., HSU, P. D., WRIGHT, J., AGARWALA, V., SCOTT, D. A. & ZHANG, F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nature protocols,* 8**,** 2281-2308.

READHEAD, B., HAURE-MIRANDE, J. V., FUNK, C. C., RICHARDS, M. A., SHANNON, P., HAROUTUNIAN, V., SANO, M., LIANG, W. S., BECKMANN, N. D., PRICE, N. D., REIMAN, E. M., SCHADT, E. E., EHRLICH, M. E., GANDY, S. & DUDLEY, J. T. 2018. Multiscale Analysis of Independent Alzheimer's Cohorts Finds Disruption of Molecular, Genetic, and Clinical Networks by Human Herpesvirus. *Neuron,* 99**,** 64-82.e7.

References

REDDY, V. R., STEUKERS, L., LI, Y., FUCHS, W., VANDERPLASSCHEN, A. & NAUWYNCK, H. J. 2014. Replication characteristics of infectious laryngotracheitis virus in the respiratory and conjunctival mucosa. *Avian Pathol,* 43**,** 450-7.

REEVES, M. B., DAVIES, A. A., MCSHARRY, B. P., WILKINSON, G. W. & SINCLAIR, J. H. 2007. Complex I binding by a virally encoded RNA regulates mitochondria-induced cell death. *Science,* 316**,** 1345-8.

REHMSMEIER, M., STEFFEN, P., HOCHSMANN, M. & GIEGERICH, R. 2004. Fast and effective prediction of microRNA/target duplexes. *Rna-a Publication of the Rna Society,* 10**,** 1507-1517.

REINHART, B. J., SLACK, F. J., BASSON, M., PASQUINELLI, A. E., BETTINGER, J. C., ROUGVIE, A. E., HORVITZ, H. R. & RUVKUN, G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature,* 403**,** 901-6.

RIAZ, A. 2014 *Functional analysis of ovine herpesvirus 2 encoded microRNAs.* PhD Thesis, Univeristy of Edinburgh.

RIAZ, A., DRY, I., LEVY, C. S., HOPKINS, J., GREY, F., SHAW, D. J. & DALZIEL, R. G. 2014. Ovine herpesvirus-2-encoded microRNAs target virus genes involved in virus latency. *Journal of General Virology,* 95**,** 472-480.

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 Journal,* 31**,** 2207-2221.

ROIZMANN, B., DESROSIERS, R. C., FLECKENSTEIN, B., LOPEZ, C., MINSON, A. C. & STUDDERT, M. J. 1992. The family Herpesviridae: an update. The Herpesvirus Study Group of the International Committee on Taxonomy of Viruses. *Arch Virol,* 123**,** 425-49.

RONG, S., WHEELER, D. & WEBER, F. 2014. Efficient Marek's disease virus (MDV) and herpesvirus of turkey infection of the QM7 cell line that does not contain latent MDV genome. *Avian Pathol,* 43**,** 414-9.

ROSEWICK, N., MOMONT, M., DURKIN, K., TAKEDA, H., CAIMENT, F., CLEUTER, Y., VERNIN, C., MORTREUX, F., WATTEL, E., BURNY, A., GEORGES, M. & VAN DEN BROEKE, A. 2013. Deep sequencing reveals abundant noncanonical retroviral microRNAs in B-cell leukemia/lymphoma. *Proc Natl Acad Sci U S A,* 110**,** 2306-11.

ROWE, M., LEAR, A. L., CROOM-CARTER, D., DAVIES, A. H. & RICKINSON, A. B. 1992. Three pathways of Epstein-Barr virus gene activation from EBNA1-positive latency in B lymphocytes. *Journal of Virology,* 66**,** 122-131.

ROY, P., ISLAM, A. F. M. F., BURGESS, S. K., HUNT, P. W., MCNALLY, J. & WALKDEN-BROWN, S. W. 2015. Real-time PCR quantification of infectious laryngotracheitis virus in chicken tissues, faeces, isolator-dust and bedding material over 28 days following infection reveals high levels in faeces and dust. *Journal of General Virology,* 96**,** 3338-3347.

RUSSELL, G. C., STEWART, J. P. & HAIG, D. M. 2009. Malignant catarrhal fever: A review. *Veterinary Journal,* 179**,** 324-335.

RUSSELL, T. A., STEFANOVIC, T. & TSCHARKE, D. C. 2015. Engineering herpes simplex viruses by infection-transfection methods including recombination site targeting by CRISPR/Cas9 nucleases. *J Virol Methods,* 213**,** 18-25.

RYAN, P. & SHANKLY, F. L. 1996. A double-strand break in a herpesvirus genome stimulates targeted homologous recombination with exogenous, cloned viral sequences. *Journal of Virological Methods,* 57**,** 95-107.

SAGER, K., ALAM, S., BOND, A., CHINNAPPAN, L. & PROBERT, C. S. 2015. Review article: cytomegalovirus and inflammatory bowel disease. *Aliment Pharmacol Ther,* 41**,** 725-33.

References

SARNOW, P., JOPLING, C. L., NORMAN, K. L., SCHUTZ, S. & WEHNER, K. A. 2006. MicroRNAs: expression, avoidance and subversion by vertebrate viruses. *Nat Rev Microbiol,* 4**,** 651-9.

SARY, K., CHENIER, S., GAGNON, C. A., SHIVAPRASAD, H. L., SYLVESTRE, D. & BOULIANNE, M. 2017. Esophagitis and Pharyngitis Associated with Avian Infectious Laryngotracheitis in Backyard Chickens: Two Cases. *Avian Diseases,* 61**,** 255-260.

SCHETTER, A. J., HEEGAARD, N. H. H. & HARRIS, C. C. 2010. Inflammation and cancer: interweaving microRNA, free radical, cytokine and p53 pathways. *Carcinogenesis,* 31**,** 37-49.

SCHNEIDERS, G. H., RIBLET, S. M. & GARCIA, M. 2018. Attenuation and Protection Efficacy of a Recombinant Infectious Laryngotracheitis Virus (ILTV) Depleted of Open Reading Frame C (DeltaORFC) when Delivered in ovo. *Avian Dis,* 62**,** 143-151.

SCHNITZLEIN, W. M. & TRIPATHY, D. N. 1995. REPLICATION OF INFECTIOUS LARYNGOTRACHEITIS VIRUS IN A QUAIL CELL-LINE, QT-35. *Avian Diseases,* 39**,** 528-531.

SHEN, W., SA E SILVA, M., JABER, T., VITVITSKAIA, O., LI, S., HENDERSON, G. & JONES, C. 2009. Two small RNAs encoded within the first 1.5 kilobases of the herpes simplex virus type 1 latency-associated transcript can inhibit productive infection and cooperate to inhibit apoptosis. *J Virol,* 83**,** 9131-9.

SHI, B., GAO, W. & WANG, J. 2012. Sequence Fingerprints of MicroRNA Conservation. *PLOS ONE,* 7**,** e48256.

SHUKLA, D. & SPEAR, P. G. 2001. Herpesviruses and heparan sulfate: an intimate relationship in aid of viral entry. *Journal of Clinical Investigation,* 108**,** 503-510.

SINKUNAS, T., GASIUNAS, G., FREMAUX, C., BARRANGOU, R., HORVATH, P. & SIKSNYS, V. 2011. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *Embo j,* 30**,** 1335-42.

SMITHIES, O., GREGG, R. G., BOGGS, S. S., KORALEWSKI, M. A. & KUCHERLAPATI, R. S. 1985. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature,* 317**,** 230-4.

SPATZ., S., FUCHS., W., ROSS., T., RIBLET., S., KIM., T., LONCOMAN., C., VOLKENING., J. & GARCIA., M. 2018. Recombineering of the Infectious Laryngotracheitis Genome using Synthetic Genomics Assembly. *In:* SPATZ, S., ed. International Herpesvirus Workshop 2018, 2018 Vancouver, Canada.

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.

SUMMERS, B. C. & LEIB, D. A. 2002. Herpes simplex virus type 1 origins of DNA replication play no role in the regulation of flanking promoters. *J Virol,* 76**,** 7020-9.

SWAYNE, D. E., GLISSON, J. R., MCDOUGALD, L. R., NOLAN, L. K., SUAREZ, D. L. & NAIR, V. L. 2013. *Diseases of Poultry,* Wiley-Blackwell.

TAGAWA, T., GAO, S., KOPARDE, V. N., GONZALEZ, M., SPOUGE, J. L., SERQUINA, A. P., LURAIN, K., RAMASWAMI, R., ULDRICK, T. S., YARCHOAN, R. & ZIEGELBAUER, J. M. 2018. Discovery of Kaposi's sarcoma herpesvirus-encoded circular RNAs and a human antiviral circular RNA. *Proc Natl Acad Sci U S A,* 115**,** 12805-12810.

TANG, S., BERTKE, A. S., PATEL, A., WANG, K., COHEN, J. I. & KRAUSE, P. R. 2008. An acutely and latently expressed herpes simplex virus 2 viral microRNA inhibits expression of ICP34.5, a viral neurovirulence factor. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 10931-10936.

TESSON, L., USAL, C., MENORET, S., LEUNG, E., NILES, B. J., REMY, S., SANTIAGO, Y., VINCENT, A. I., MENG, X., ZHANG, L., GREGORY, P. D., ANEGON, I. & COST, G. J. 2011. Knockout rats generated by embryo microinjection of TALENs. *Nat Biotechnol,* 29**,** 695-6.

THELLMAN, N. M. & TRIEZENBERG, S. J. 2017. Herpes Simplex Virus Establishment, Maintenance, and Reactivation: In Vitro Modeling of Latency. *Pathogens (Basel, Switzerland),* 6**,** 28.

THOMAS, S., COFFIN, R. S., WATTS, P., GOUGH, G. & LATCHMAN, D. S. 1998. The TAATGARAT motif in the herpes simplex virus immediate-early gene promoters can confer both positive and negative responses to cellular octamer-binding proteins when it is located within the viral genome. *J Virol,* 72**,** 3495-500.

THOMSON, J. M., PARKER, J. S. & HAMMOND, S. M. 2007. Microarray analysis of miRNA gene expression. *Methods Enzymol,* 427**,** 107-22.

THUREEN, D. R. & KEELER, C. L. 2006. Psittacid herpesvirus 1 and infectious laryngotracheitis virus: Comparative genome sequence analysis of two avian alphaherpesviruses. *Journal of Virology,* 80**,** 7863-7872.

TIMONEDA, O., NUNEZ-HERNANDEZ, F., BALCELLS, I., MUNOZ, M., CASTELLO, A., VERA, G., PEREZ, L. J., EGEA, R., MIR, G., CORDOBA, S., ROSELL, R., SEGALES, J., TOMAS, A., SANCHEZ, A. & NUNEZ, J. I. 2014. The Role of Viral and Host MicroRNAs in the Aujeszky's Disease Virus during the Infection Process. *Plos One,* 9**,** 11.

TOUSSIROT, E. & ROUDIER, J. 2008. Epstein-Barr virus in autoimmune diseases. *Best Pract Res Clin Rheumatol,* 22**,** 883-96.

TRAVIS, A. J., MOODY, J., HELWAK, A., TOLLERVEY, D. & KUDLA, G. 2014. Hyb: A bioinformatics pipeline for the analysis of CLASH (crosslinking, ligation and sequencing of hybrids) data. *Methods,* 65**,** 263-273.

TU, Z., YANG, W., YAN, S., GUO, X. & LI, X. J. 2015. CRISPR/Cas9: a powerful genetic engineering tool for establishing large animal models of neurodegenerative diseases. *Mol Neurodegener,* 10**,** 35.

UMBACH, J. L., KRAMER, M. F., JURAK, I., KARNOWSKI, H. W., COEN, D. M. & CULLEN, B. R. 2008a. MicroRNAs expressed by herpes simplex virus 1 during latent infection regulate viral mRNAs. *Nature,* 454**,** 780-U108.

UMBACH, J. L., KRAMER, M. F., JURAK, I., KARNOWSKI, H. W., COEN, D. M. & CULLEN, B. R. 2008b. microRNAs expressed by herpes simplex virus 1 during latent infection regulate viral mRNAs. *Nature,* 454**,** 780-783.

VAGNOZZI, A. E., BELTRAN, G., ZAVALA, G., READ, L., SHARIF, S. & GARCIA, M. 2018. Cytokine genes transcription in trachea, Harderian gland, and trigeminal ganglia of chickens inoculated with virulent infectious laryngotracheitis virus (ILTV) strain. *Avian Pathol***,** 1-41.

VALENCIA-SANCHEZ, M. A., LIU, J., HANNON, G. J. & PARKER, R. 2006. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev,* 20**,** 515-24.

VAN DAMME, E. & VAN LOOCK, M. 2014. Functional annotation of human cytomegalovirus gene products: an update. *Front Microbiol,* 5**,** 218.

VAN DIEMEN, F. R., KRUSE, E. M., HOOYKAAS, M. J. G., BRUGGELING, C. E., SCHURCH, A. C., VAN HAM, P. M., IMHOF, S. M., NIJHUIS, M., WIERTZ, E. & LEBBINK, R. J. 2016. CRISPR/Cas9-Mediated Genome Editing of Herpesviruses Limits Productive and Latent Infections. *Plos Pathogens,* 12**,** 29.

VARKONYI-GASIC, E. & HELLENS, R. P. 2011. Quantitative Stem-Loop RT-PCR for Detection of MicroRNAs. *In:* KODAMA, H. & KOMAMINE, A. (eds.) *Rnai and Plant Gene Function Analysis: Methods and Protocols.* Totowa: Humana Press Inc.

VEITS, J., LUSCHOW, D., KINDERMANN, K., WERNER, O., TEIFKE, J. P., METTENLEITER, T. C. & FUCHS, W. 2003a. Deletion of the non-essential UL0

gene of infectious laryngotracheitis (ILT) virus leads to attenuation in chickens, and UL0 mutants expressing influenza virus haemagglutinin (H7) protect against ILT and fowl plague. *J Gen Virol,* 84**,** 3343-52.

VEITS, J., METTENLEITER, T. C. & FUCHS, W. 2003b. Five unique open reading frames of infectious laryngotracheitis virus are expressed during infection but are dispensable for virus replication in cell culture. *Journal of General Virology,* 84**,** 1415-1425.

VON BUELOW, V. & KLASEN, A. 1983. EFFECTS OF AVIAN VIRUSES ON CULTURED CHICKEN BONE MARROW DERIVED MACROPHAGES. *Avian Pathology,* 12**,** 179-198.

WAIDNER, L. A., BURNSIDE, J., ANDERSON, A. S., BERNBERG, E. L., GERMAN, M. A., MEYERS, B. C., GREEN, P. J. & MORGAN, R. W. 2011. A microRNA of infectious laryngotracheitis virus can downregulate and direct cleavage of ICP4 mRNA. *Virology,* 411**,** 25-31.

WAIDNER, L. A., MORGAN, R. W., ANDERSON, A. S., BERNBERG, E. L., KAMBOJ, S., GARCIA, M., RIBLET, S. M., OUYANG, M., ISAACS, G. K., MARKIS, M., MEYERS, B. C., GREEN, P. J. & BURNSIDE, J. 2009. MicroRNAs of Gallid and Meleagrid herpesviruses show generally conserved genomic locations and are virus-specific. *Virology,* 388**,** 128-36.

WALTZEK, T. B., KELLEY, G. O., ALFARO, M. E., KUROBE, T., DAVISON, A. J. & HEDRICK, R. P. 2009. Phylogenetic relationships in the family Alloherpesviridae. *Dis Aquat Organ,* 84**,** 179-94.

WANG, D., WANG, X.-W., PENG, X.-C., XIANG, Y., SONG, S.-B., WANG, Y.-Y., CHEN, L., XIN, V. W., LYU, Y.-N., JI, J., MA, Z.-W., LI, C.-B. & XIN, H.-W. 2018. CRISPR/Cas9 genome editing technology significantly accelerated herpes simplex virus research. *Cancer Gene Therapy*.

WANG, K. C. & CHANG, H. Y. 2011. Molecular mechanisms of long noncoding RNAs. *Molecular cell,* 43**,** 904-914.

WANG, L., QIN, Y., TONG, L., WU, S., WANG, Q., JIAO, Q., GUO, Z., LIN, L., WANG, R., ZHAO, W. & ZHONG, Z. 2012. MiR-342-5p suppresses coxsackievirus B3 biosynthesis by targeting the 2C-coding region. *Antiviral research,* 93**,** 270-279.

WANG, L. G., MA, J., XUE, C. Y., WANG, W., GUO, C., CHEN, F., QIN, J. P., HUANG, N. H., BI, Y. Z. & CAO, Y. C. 2013. Dynamic distribution and tissue tropism of infectious laryngotracheitis virus in experimentally infected chickens. *Archives of Virology,* 158**,** 659-666.

WANG, X. 2016. Improving microRNA target prediction by modeling with unambiguously identified microRNA-target pairs from CLIP-ligation studies. *Bioinformatics,* 32**,** 1316-22.

WANG, X., XU, X., MA, Z., HUO, Y. Q., XIAO, Z. T., LI, Y. & WANG, Y. H. 2011. Dynamic mechanisms for pre-miRNA binding and export by Exportin-5. *Rna-a Publication of the Rna Society,* 17**,** 1511-1528.

WAYENGERA, M. 2011. Proviral HIV-genome-wide and pol-gene specific Zinc Finger Nucleases: Usability for targeted HIV gene therapy. *Theoretical Biology & Medical Modelling,* 8**,** 26-26.

WAYENGERA, M. 2012. Zinc finger arrays binding human papillomavirus types 16 and 18 genomic DNA: precursors of gene-therapeutics for in-situ reversal of associated cervical neoplasia. *Theor Biol Med Model,* 9**,** 30.

WELLER, S. K. & COEN, D. M. 2012. Herpes Simplex Viruses: Mechanisms of DNA Replication. *Cold Spring Harbor perspectives in biology,* 4**,** a013011-a013011.

WERTHEIM, J. O., SMITH, M. D., SMITH, D. M., SCHEFFLER, K. & POND, S. L. K. 2014. Evolutionary Origins of Human Herpes Simplex Viruses 1 and 2. *Molecular Biology and Evolution,* 31**,** 2356-2364.

References

WHISNANT, A. W., KEHL, T., BAO, Q., MATERNIAK, M., KUZMAK, J., LOCHELT, M. & CULLEN, B. R. 2014. Identification of novel, highly expressed retroviral microRNAs in cells infected by bovine foamy virus. *J Virol,* 88**,** 4679-86.

WIGHTMAN, B., HA, I. & RUVKUN, G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell,* 75**,** 855-62.

WILCOX, C. L. & JOHNSON, E. M. 1987. Nerve growth factor deprivation results in the reactivation of latent herpes simplex virus in vitro. *Journal of Virology,* 61**,** 2311-2315.

WILUSZ, J. E. & SHARP, P. A. 2013. Molecular biology. A circuitous route to noncoding RNA. *Science,* 340**,** 440-1.

WINTER, J., JUNG, S., KELLER, S., GREGORY, R. I. & DIEDERICHS, S. 2009. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nature Cell Biology,* 11**,** 228-234.

WONG, N. & WANG, X. 2015. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res,* 43**,** D146-52.

XU, H. T., WANG, X. B., DU, Z. L. & LI, N. 2006. Identification of microRNAs from different tissues of chicken embryo and adult chicken. *Febs Letters,* 580**,** 3610-3616.

XU, Y., CHEN, Y., LI, D., LIU, Q., XUAN, Z. & LI, W.-H. 2017. TargetLink, a new method for identifying the endogenous target set of a specific microRNA in intact living cells. *RNA Biology,* 14**,** 259-274.

YANG, W.-H. & BLOCH, D. B. 2007. Probing the mRNA processing body using protein macroarrays and "autoantigenomics". *RNA (New York, N.Y.),* 13**,** 704-712.

YIN, V. P., THOMSON, J. M., THUMMEL, R., HYDE, D. R., HAMMOND, S. M. & POSS, K. D. 2008. Fgf-dependent depletion of microRNA-133 promotes appendage regeneration in zebrafish. *Genes Dev,* 22**,** 728-33.

YOSHIMI, K., KUNIHIRO, Y., KANEKO, T., NAGAHORA, H., VOIGT, B. & MASHIMO, T. 2016. ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. *Nat Commun,* 7**,** 10431.

YUEN, K. S., CHAN, C. P., WONG, N. H. M., HO, C. H., HO, T. H., LEI, T., DENG, W., TSAO, S. W., CHEN, H. L., KOK, K. H. & JIN, D. Y. 2015. CRISPR/Cas9-mediated genome editing of Epstein Barr virus in human cells. *Journal of General Virology,* 96**,** 626-636.

ZENG, Y. & CULLEN, B. R. 2005. Efficient processing of primary microRNA hairpins by drosha requires flanking nonstructured RNA sequences. *Journal of Biological Chemistry,* 280**,** 27595-27603.

ZHAO, Y., KONG, C. C., CUI, X. L., CUI, H. Y., SHI, X. M., ZHANG, X. M., HU, S. L., HAO, L. W. & WANG, Y. F. 2013. Detection of Infectious Laryngotracheitis Virus by Real-Time PCR in Naturally and Experimentally Infected Chickens. *Plos One,* 8**,** 10.

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 Pathogens,* 7.

ZHAO, Y. G., YAO, Y. X., XU, H. T., LAMBETH, L., SMITH, L. P., KGOSANA, L., WANG, X. W. & NAIR, V. 2009. A Functional MicroRNA-155 Ortholog Encoded by the Oncogenic Marek's Disease Virus. *Journal of Virology,* 83**,** 489-492.

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. 73**,** 3210-3218.

ZIEMANN, K., METTENLEITER, T. C. & FUCHS, W. 1998a. Gene arrangement within the unique long genome region of infectious laryngotracheitis virus is distinct horn that of other alphaherpesviruses. *Journal of Virology,* 72**,** 847-852.

References

ZIEMANN, K., METTENLEITER, T. C. & FUCHS, W. 1998b. Infectious laryngotracheitis herpesvirus expresses a related pair of unique nuclear proteins which are encoded by split genes located at the right end of the U-L genome region. *Journal of Virology,* 72**,** 6867-6874.