Characterisation and expression of ORF2 in genome segment 10 of African horse sickness virus

By

Carmen Coertzen
17088900

Supervisor: Prof Vida van Staden

Co-supervisors: Prof Christiaan Potgieter & Prof Eshchar Mizrachi

Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae

In the Faculty of Natural and Agricultural Sciences

University of Pretoria

November 2022

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

Declaration

I, **Carmen Coertzen,** declare that the thesis/dissertation, which I hereby submit for the degree *Magister Scientiae* at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE:  DATE: 14 November 2022

Acknowledgments

I would like to show sincere appreciation to the following persons and organisations for supporting me throughout this study:

My supervisor, Prof Vida van Staden, I am grateful for your infinite patience, assistance, and support. You have helped me grow as a researcher and provided me with the necessary tools to succeed.

My co-supervisor, Prof Christiaan Potgieter, thank you for your experimental and overall aid. You are a stream of knowledge when it comes to research.

My co-supervisor, Prof Eshchar Mizrachi, thank you for taking the time to assist me with my bioinformatic analyses.

The University of Pretoria and the Poliomyelitis Research Foundation for their financial support.

The University of Pretoria for the utilisation of their facilities.

Deltamune for their antibody production services.

The Laboratory for Microscopy and Microanalysis (UP) for the use of their facilities and technical assistance.

The sequencing facility (Gladys Shabangu and Renate Zipfel) for the use of their facilities.

Arrie Klopper and Sandra van Wyngaardt for technical assistance.

My fellow past and present peers, Dr Gayle Wall and Dr Shareen Boughan as well as Ariel Buyens and Desré Pinard for all your advice, motivational words and always being willing to lend a helping hand.

Research outputs

1. National Conferences

Coertzen, C., van Staden, V., Potgieter A.C. and Mizrachi, E. 2022. Characterisation and expression of ORF2 in genome segment 10 (seg-10) of AHSV. The South African Genetics Society and the South African Society for Bioinformatics: BIO2022 conference. Online (poster presentation)

Coertzen, C., van Staden, V., Potgieter A.C. and Mizrachi, E. 2022. African horse sickness virus (AHSV) Seg-10 ORF2 potentially encodes a novel non-structural protein. The South African Society of Biochemistry and Molecular Biology (SASBMB) conference. Online (poster presentation)

2. International conferences

Coertzen, C., van Staden, V., Mizrachi, E. and Potgieter A.C. 2022. ORF2 in African horse sickness virus segment 10 potentially encodes a novel non-structural protein. 14th International dsRNA Virus Symposium. Banff, Canada (poster presentation).

Table of contents

Summary

Characterisation and expression of ORF2 in genome segment 10 of African horse sickness virus

by

Carmen Coertzen

Supervisor: Prof V van Staden
Department of Biochemistry, Genetics and Microbiology
University of Pretoria, Gauteng, South Africa

Co-supervisors: Prof A C Potgieter
Deltamune (Pty) Ltd and North-West University, Potchefstroom
&
Prof Eschar Mizrachi
Department of Biochemistry, Genetics and Microbiology
University of Pretoria, Gauteng, South Africa

For the degree *Magister Scientiae*

African horse sickness virus (AHSV) is responsible for the often fatal disease African horse sickness in horses. As viruses have small genomes, they are required to maximise the amount of coding and regulatory information within them by making use of overlapping functional elements. The smallest genome segment in the segmented double-stranded RNA genome of AHSV, segment 10 (Seg-10), encodes the non-structural protein NS3. Recent bioinformatic analyses revealed an additional open reading frame (ORF) on Seg-10, in the +1 frame relative to NS3.

The aim of this study was to completely characterise the additional open reading frame (ORF2) present in AHSV Seg-10, and to identify whether it functions as a structural RNA element or by encoding a protein with a role in the AHSV replication cycle. To accomplish this, the phylogeny, variants, conservation, and selection pressures on AHSV Seg-10 ORF2 were investigated, and related to its possible relevance as an RNA element or potential protein coding sequence. FRESCo analysis revealed high selection pressures on regions of Seg-10 overlapping with ORF2. The ORF2 and NS3 coding regions on Seg-10 were under conflicting selection pressures, characteristic of overlapping reading frames. The presence of ORF2 was maintained throughout all analysed sequences, ranging in size from 60 to 83 codons. RNA analysis did not provide support for functional RNA elements present in the ORF2 region. The high amino acid similarity scores and internal

conservation of the putative ORF2 protein across all serotypes lead us to conclude ORF2 rather encodes a functional protein.

The protein encoded by AHSV Seg-10 ORF2 was subsequently expressed in insect cells, mammalian cells and bacteria to better understand its effect on different host cells, and to be utilised as antigen source to produce antibodies against ORF2. The subcellular localisation and trafficking of the ORF2 protein was also investigated, in order to deduce its possible function. Native and tagged ORF2 proteins were successfully recombinantly expressed in Sf9 insect cells via the baculovirus system. Here ORF2 showed specific intracellular punctate cytoplasmic and nuclear localisation profiles. During transient mammalian expression, recombinant ORF2-eGFP initially localised to the cytoplasm and later moved to the nucleus, and ultimately caused cell nuclei to shrivel showing a cytotoxic effect. Expression of ORF2 in *E. coli* following induction resulted in low yields and influenced cell proliferation. Thus, overall recombinant expression of the ORF2 protein resulted in low expression and/or cytotoxicity.

Lastly the expression of the ORF2 protein during AHSV infection was explored, to identify whether it functions in the AHSV replication cycle. ORF2 was detected as spherical structures in the cytoplasm following infection with AHSV-1 or AHSV-4, and additionally in the nucleus during AHSV-8 infection. The cytoplasmic structures were similar in morphology to viral inclusion bodies, which requires future verification.

This study represents the first report of the potential identification of a novel non-structural protein encoded by ORF2 in AHSV Seg-10, which will then become AHSV NS5. The identification and characterisation of new elements/proteins from additional open reading frames paves the way for improved future understanding of the AHSV replication cycle and viral pathogenesis.

List of abbreviations

aa	amino acid/s
AF	alexa Fluor-conjugated secondary antibody
AHS	African horse sickness
AHSV	African horse sickness virus
APS	ammonium persulphate
ATA	aurintricarboxylic acid
bp	base pairs
BSR-T7	baby hamster kidney cells with T7 promoter
BTV	bluetongue virus
CLSM	confocal laser scanning microscopy
CPE	cytopathic effect
DAPI	4'6'-diamidino-2-phenylindole
ddH_2O	double distilled water
dH_2O	distilled water
DISA	disabled Infectious Single Animal
DMEM	Dulbecco's Modified Eagle Medium
DNA	deoxyribonucleic acid
ds	double stranded
dsRNA	double stranded ribonucleic acid
E. coli	*Escherichia coli*
e.g.	*exempli gratia*; for example
ECRA	entry Competent Replication-Abortive
EDTA	ethylenediaminetetraacetic acid
EEV	equine encephalosis virus
eGFP	enhanced green fluorescent protein
EHDV	epizootic haemorrhagic disease of deer virus
eIF	eukaryotic translation initiation factor
EIP	extrinsic incubation period
ELISA	enzyme-linked immunosorbent assay
et al.	*et alia*; and others
EtBr	ethidium bromide
EtOH	ethanol
FCS	foetal calf serum
Fig.	figure
FRESCo	finding regions of excess synonymous constraint
g	gravitational force
HAV	hepatitis A virus
hpi	hours post infection
hpt	hours post transfection
i.e.	*id est;* that is
IFNAR	interferon-alpha/beta receptor alpha chain
IPTG	isopropyl-β-D-thiogalactoside
IRES	internal ribosome entry site
kb	kilobase
kDa	kilodalton
LAVs	live-attenuated vaccines
LB	Luria-Bertani
LF	lipofectamine

MAFFT	multiple alignment using fast Fourier transform
MEM	Minimum Essential Medium
min	minute/s
mM	millimolar
mRNA	messenger RNA
MVA	modified vaccinia Ankara
NEAA	non-essential amino acids
ng	nanogram
NLS	nuclear localisation signal
NLSs	nuclear localisation signals
nm	nanometre
NoLS	nucleolar localisation signals
NS	non-structural
nt	nucleotide
ºC	degrees Celsius
OD	optical density
OIE	Office International des Epizooties
ORFs	open reading frame/s
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pen/strep	penicillin and streptomycin
PFA	paraformaldehyde
PML-NBs	promyelocytic leukaemia nuclear bodies
pmol	picomol
PSB	protein solvent buffer
RdRp	RNA-dependent RNA polymerase
RDV	rice dward virus
RE	restriction endonuclease
RER	rough endoplasmic reticulum
RG	reverse genetics
RIPA	radioimmunoprecipitation assay
RNA	ribonucleic acid
rpm	revolutions per minute
RT	room temperature
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SBV	Schmallenberg virus
SDS-PAGE	sodium dodecyl polyacrylamide
HYPHY	hypothesis Testing using Phylogenies
sec	second/s
Seg	segment
Sf9	*Spodoptera frugiperda*
SLAC	single-Likelihood Ancestor Counting
SOC	super optimal broth with catabolite repression
ss	single stranded
TAE	tris-acetate-EDTA buffer
TEMED	tetramethylethylenediamine
TGS	tris-Glycine-SDS
TM	transmembrane
Tris	Tris hydroxymethyl aminomethane

U	units
UTR	untranslated region
V	volt
v/v	volume per volume
VIBs	viral inclusion bodies
VLPs	virus-like particles
VP	virus protein
WT	wild type
w/v	weight per volume
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
µg	microgram
µl	microlitre
µM	micromolar

List of buffers

Coomassie:	0.125% Coomassie blue 50% methanol (v/v) 10% acetic acid (v/v)
LB broth:	1% tryptone (w/v) 0.5% yeast extract (w/v) 1% NaCl (w/v) pH 7.4
1X PBS:	137 mM NaCl 2.4 mM KCl 4.3 mM Na2HPO4.2H2O 1.4 mM KH2PO4 pH 7.4
6X PSB:	15% 2-mercaptoethanol (v/v) 40% glycerol (v/v) 12% SDS (v/v) 0.375 M Tris pH 6.6
RIPA:	150 mM sodium chloride 1% Triton X-100 (v/v) 0.5% sodium deoxycholate (w/v) 0.1% SDS (w/v) 50 mM Tris pH 8
SOC medium:	0.5 % Yeast Extract (w/v) 2 % Tryptone (w/v) 10 mM NaCl 2.5 mM KCl 10 mM MgCl2 10 mM MgSO4 20 mM Glucose
1X TAE:	40 mM Tris-HCl 20 mM Na-Acetate 1 mM EDTA
TE buffer:	10mM Tris 1mM EDTA pH 8
1X TGS:	25 mM Tris-HCl, pH 8.3 192 mM glycine 0.1% SDS (w/v)
Transfer buffer:	25 mM Tris 192 mM glycine 20% methanol (v/v) pH 8.3

CHAPTER 1

LITERATURE REVIEW

1.1 Introduction

African horse sickness (AHS) is a disease of horses and other equids, brought about by the etiological agent African horse sickness virus (AHSV). AHSV is an arthropod-borne virus in the genus *Orbivirus* and the family *Reoviridae* (ATTOUI *et al.* 2009). Biting midges, specifically of the genus *Culicoides* are responsible for its transmission. AHS has a mortality rate of more than 90% in susceptible naïve horses and African donkeys. AHSV can also replicate in infected zebras and Asian donkeys, but they rarely show clinical signs (COETZER AND GUTHRIE 2005). The World Organisation for Animal Health (OIE) has classed AHS as a notifiable disease, due to its rapid spread and high mortality numbers in horses. AHS is indigenous to sub-Saharan Africa, but has spread beyond its geographical limitations into North Africa and over the Mediterranean sea into Southern Europe (COETZER *et al.* 1994; OIE 2019). An outbreak of AHS in Thailand in 2020 was confirmed to be from South African virus strains which raises great concern for its spread into non-endemic areas (KING *et al.* 2020; TOH *et al.* 2022).

Bluetongue virus (BTV), the type species also in the genus *Orbivirus,* primarily infects ruminants like cows and sheep. BTV shares a lot of structural as well as functional resemblances with AHSV (ATTOUI *et al.* 2009). All orbivirus particles are non-enveloped, triple-layered icosahedral particles consisting of a subcore, a core layer, and an outer capsid. The core contains the segmented double-stranded (ds) RNA genome (ROY *et al.* 1994; MERTENS AND DIPROSE 2004). This genome encodes 11 viral proteins. These include seven structural (VP1-7) and four non-structural (NS1-4) proteins. Understanding the structure and function of the different viral genes and their encoded proteins in the virus replication cycle is central to understanding the viral pathogenesis.

Because RNA viruses need to express multiple structural and non-structural proteins from compact genomes, the utilisation of overlapping open reading frames (ORFs) is often required (FIRTH AND BRIERLEY 2012). In both AHSV and BTV the two smallest genome segments, namely Seg-9 and Seg-10, encode major proteins VP6 and NS3 respectively. Bioinformatic analyses uncovered that Seg-9 and Seg-10 each contain an additional small open reading frame (ORF) in the +1-reading frame (FIRTH 2008; SEALFON *et al.* 2015). The product of this ORF of Seg-9 was characterised as NS4, a protein shown to potentially be an important virulence factor that interferes with host innate immunity upon AHSV infection (WALL *et al.* 2021). STEWART *et al.* (2015) characterised the protein encoded by the second ORF from Seg-10 in BTV, designated ORF2. This second ORF also exists in Seg-10 of AHSV, but its conservation, and the size, nature, expression, and potential subcellular localisation of the protein encoded by AHSV ORF2 has not yet been determined. This will form the basis for the investigations of this study.

The scope of this literature review will be to briefly highlight the epidemiology and economic impact of AHS, how it is transmitted, as well as its pathogenicity and prevention/treatment. The AHSV molecular biology and replication cycle will be described next. Following this will be background on

overlapping open reading frames and their role in virus genomes, concluding with the recent discovery of ORF2 in Seg-10. Lastly a short description of the baculovirus expression system and its applications to this study will be given. This will allow formulation of the main aim and objectives for this study, and strategies needed to address them.

1.2 African Horse Sickness

1.2.1 Transmission

Biological transmission of AHSV is primarily by *Culicoides* midges, which are haematophagous arthropods. There are a 1 500 *Culicoides* species situated all over the world excluding New Zealand and Antarctica. They thrive in warm humid places such as streams, swamps or shorelines. Orbiviruses like BTV and AHSV can replicate in 30 species when fed infected blood in a laboratory setting and could most probably transmit the virus (WILSON *et al.* 2009). *C. imicola* is the most prevalent vector in warm regions, and is less common in cooler highland regions. *C. bolitinos* has been identified in colder, drier regions of sub-Saharan Africa as another primary vector for transmission (MELLOR *et al.* 2009). Research revealed that *C. obsoletus* and *C. pulicaris* found in Europe are also competent, and might therefore be important as vectors for AHSV (MEISWINKEL *et al.* 2000; CARPENTER *et al.* 2008). In Switzerland, *C. obsoletus* and *C. scoticus* also showed full viral dissemination of AHSV during vector competency experiments making transmission possible in this area, specifically on the Swiss Plateau where the abundance of these vectors is higher (MAURER *et al.* 2022).

In the UK, *Culicoides* species were also found to opportunistically feed on exotic zoo animals (ENGLAND *et al.* 2020; NELSON *et al.* 2022). Zoo animals are at risk of *Culicoides*-borne arboviruses like BTV, Schmallenberg virus (SBV) and AHSV. These native or exotic animals have a high value both in economic terms and reproduction purposes for conservation of endangered species. ENGLAND *et al.* (2020) analysed *Culicoides* blood-meals in two zoos found in the UK. They found that *Culicoides* species were feeding on a wide range of animals including camels, rhinoceros, and Asian elephants, and subsequently six putative vector species were identified. There is a strong need for vaccination as well as mitigating measures for susceptible zoo animals in the event of an outbreak to protect endangered species (ENGLAND *et al.* 2020). The accurate surveillance of *Culicoides* populations in London zoo's is essential for identifying any variations in vector abundance and subsequently implement the necessary prevention measures (NELSON *et al.* 2022).

The bites of adult female *Culicoides* midges are the primary reason for transmission of AHSV. Female midges require blood proteins for egg production. The virus is ingested by the midge upon taking a blood meal, if AHSV is present in the peripheral vascular system or in the skin tissues of the equid host. The orbivirus then localises to the lumen of the hind part of the midgut where it replicates in the gut cells. The newly formed viruses migrate through the basolateral membrane to the haemocoel, and infect secondary organs and eventually progress towards the salivary glands. In

order to successfully be transmitted by the vector, the virus needs to replicate sufficiently without adversely affecting the midge before transmission to a vertebrate host. The extrinsic incubation period (EIP) is defined as the time between ingestion and the ability for consequent transmission of the virus to another vertebrate host. This time depends on the strain of AHSV, the *Culicoides* species and the temperature experienced by the midge vector. Usually there is an initial drop in virus concentration and then a rise until an equilibrium is reached in the vector. It is suggested that this equilibrium is maintained by constant viral replication with inactivation by the insect host. Consequently the insect vector can be continually infected with AHSV without any pathological effect on the vector which would hinder the midge's ability to feed (WILSON *et al.* 2009). Upon the next bloodmeal the virus is subsequently transmitted back to the next vertebrate host upon biting for the next feeding (WILSON *et al.* 2009).

The midge saliva contains proteases and trypsin which are responsible for viral protein activation upon feeding of the midge on the mammalian host and also prevents the phagocytic ability of host macrophages at the biting site. This activation enables virus attachment through cellular receptor binding to equid host cells (DARPEL *et al.* 2011). Once the virus is introduced into the skin of the mammalian host there is a quick rise in viral replication as well as infection of both white blood cells and the microvascular epithelial tissue of skin capillaries resulting in primary viraemia. It then targets circulatory and respiratory systems specifically, with preferential replication in the endothelial cells of the capillaries of the heart, lungs and spleen which is responsible for secondary viraemia (CLIFT AND PENRITH 2010).

1.2.2 Epidemiology

AHS is indigenous to sub-Saharan Africa (CRAIG *et al.* 2019; AHMED *et al.* 2020; HOFFMANN *et al.* 2022; NDEBÉ *et al.* 2022). Travellers from Portugal were the first to report cases of horse mortality due to AHSV in 1569 (MELLOR 1993). During 1652, Dutch colonists imported donkeys and horses to from India to the Cspe of Good Hope (MACLACHLAN AND GUTHRIE 2010). The first documented AHS outbreak in South Africa was in 1719. This outbreak was caused due to this importation of horses by hunters to areas where zebras resided (BARNARD 1998). Asymptomatic infection of zebras and donkeys serve as reservoirs for AHS, especially during trade by land, sea or air. Donkeys can be used to estimate the prevalence and distribution of AHSV since they produce high levels of antibodies. In a study conducted by MOLINI *et al.* (2020) all AHSV serotypes, except serotype 8, were found to be circulating in Namibian donkeys. This widespread distribution of different AHSV serotypes highlights the importance of vaccination of horses to prevent outbreaks. Furthermore, bat cell cultures are commonly used to isolate vector-borne viruses for transmission assessment purposes. Strains of diploid lung cells of the dwarf bat are permissive to the orbiviruses BTV, AHSV and also epizootic hemorrhagic disease of deer virus (EHDV). Thus, like zebras and donkeys, these bats act as natural reservoir and carriers of these vector-borne diseases and can be used to study the prevalence and circulation of AHSV (POVOLYAEVA *et al.* 2022).

AHS was not known to occur in Northern regions of Africa (Tunisia, Morocco and Algeria) due to the Sahara Desert acting as a barrier. Outbreaks have however happened in these regions, presumably due to migrants travelling with infected donkeys through the Sahara (MAURICE AND PROVOST 1967; MELLOR AND HAMBLIN 2004). Recently, AHS was confirmed in horses in Thailand again due to imports of infected zebras and subsequently seven outbreaks occurred in different regions killing 231 horses which was controlled by the live attenuated vaccine utilised in South Africa. AHS remains a threat in Asia due to the high number of unvaccinated horses and the presence of *Culicoides imicola* (KING *et al.* 2020; LU *et al.* 2020). The virus found in Thailand belonged to serotype 1 and was closely related to strains from South Africa (TOH *et al.* 2022). This was the first incidence of serotype 1 outside of Africa. This spread into non-endemic areas raises concern for future AHSV outbreaks (RAKSAKOON AND POTIWAT 2021).

Climatic conditions that favour *Culicoides* survival and breeding influences disease prevalence by increasing vector ranges. With increased night-time temperatures, temperature is no longer a barrier preventing the survival of midges in northerly latitudes. Strong winds also lead to midge distributions over large geographical areas. AHS has shown a high prevalence in horses in Sudan due to the increasing numbers in *Culicoides* midges (AHMED *et al.* 2020). Barriers inhibiting infection of *Culicoides* by AHSV include the midgut infection barrier, the midgut escape barrier and the diffusion barrier (PURSE AND ROGERS 2009). Increased temperatures can also lead to increased viral replication (decrease in generation time) of AHSV inside even more competent *Culicoides*, and also increases blood-feeding by the midges (MELLOR *et al.* 1998). Recent work from India confirmed that temperature, adequate rainfall and moist weather play a crucial role in development of the vector *Culicoides* (MUKHOPADHYAY et al. 2020). Furthermore, a study which analysed the habitat connectivity of *Culicoides* from the first sighting of AHSV in Thailand to China based on temperature variables, land coverage and altitude revealed a strong habitat connectivity and thus a high risk for AHS introduction into China (GAO *et al.* 2022).

Lastly, an ecological niche model was used to determine the global habitat suitability for AHSV. As expected, the sub-Sahara African continent was identified as the most suitable area for the virus. Some disease-free countries have been found to also be suitable habitats for the virus. These include Australia, Brazil, Paraguay, and Bolivia. Thus, correct control measures and surveillance to these identified risk areas are essential (ASSEFA *et al.* 2022).

1.2.3 Economic Impact

The first main AHS epidemic occurring in southern Africa in 1719, resulted in the death of 1 700 animals. During the largest AHS outbreak 70 000 horses died in southern Africa in 1855 (MELLOR AND HAMBLIN 2004). Through subsequent northwards spread, epidemics also occurred in Lebanon, Iraq, Cyprus, Syria, and Jordan as well as in India, Pakistan, Turkey and Afghanistan between 1959 and 1961. About 300 000 horses were lost during these outbreaks. In Southern Africa, specifically

the Western Cape, legislatively defined AHS controlled areas were implemented in 1997, where movement (import and export) of horses from and to the EU is strongly regulated. The legislation guidelines state that if an outbreak occurs in AHS free zones a two-year ban on importation and exportation will be placed. There have subsequently been six outbreaks in surveillance zones in the Western Cape province in 1999, 2004, 2006, 2011, 2013, and 2014 (WEYER *et al.* 2016). This has a significant impact on the South African economy (BOSMAN *et al.* 1995; MELLOR AND BOORMAN 1995).

The outbreak of AHS in 2011 (70 equid deaths) caused the stop of horse trade directly from South Africa to the EU and other countries. This led to major costs in several commercial industries like Racing South Africa, the Cape Breeders' Association and the Acorn Group of Companies. R850 000 was the minimum estimated cost to these companies spent on bulk laboratory testing and monitoring of equines and their samples. More than 200 horses were exported from South Africa per year before this outbreak. This revenue loss due to suspension of exportation lead to costs of up to R20 million per year for industry stakeholders. Poorer communities which relied on the equid trade money suffered the most economically (GREWAR *et al.* 2012).

The surveillance of AHSV-free territories is crucial for controlling AHSV within endemic regions. Diagnostic tests need to be extremely sensitive and precise in order to detect AHSV in nonendemic areas. This allows for confirmation that after the execution of an eradication programme a previously infected population is now free of the disease (DURÁN-FERRER *et al.* 2019). MAREE AND PAWESKA (2005) developed an indirect ELISA detecting group-specific recombinant protein VP7. The commercially available VP7 Blocking ELISA (INGEZIM AHSV COMPAC PLUS 2.0), can detect specific antibodies against the conserved VP7 protein. This is one of the serological tests enforced for the control of importations to the Europe. SERGEANT *et al.* (2016) proposed a stochastic risk model which describes the probability of exporting an AHS-infected horse. This model also describes management measures such as PCR tests before export and after arrival at destination. The additional PCR tests post-arrival decreased the median number of infected horses exported by at least 12 times. Thus, the chances of exporting undetected AHS infected horses can be diminished by suitable risk control methods.

A recent economic assessment study on AHS revealed that its economic cost was about $95 million annually for domestic equine industries and companies involved in international trade. The authors evaluated a cost-benefit model, and determined that the investment required to bring a new AHS vaccine to market was minute in comparison to the benefits (REDMOND *et al.* 2022).

1.2.4 Pathogenicity

Pathogenesis characterises a disease developmental state, while virulence relates to the ability of the virus to cause this state (HUISMANS *et al.* 2004). Virulent strains of AHSV representing all nine serotypes result in disease in horses. The cellular pathogenesis of AHS includes the formation of

microlesions in infected cells, due to severe damage to endothelial cells and macrophage activation. Microvascular endothelial cells are damaged upon infection, and this forms the lesions. The lesions cause capillary walls to have increased permeability. Other molecules found in infected cells include polysomes and lysosome-like bodies (GOMEZ-VILLAMANDOS *et al.* 1999). Apoptotic signals are also observed by condensation of chromatin, membrane blebbing and apoptotic bodies (STASSEN *et al.* 2011). Aggregations of red blood cells, monocytes and platelets formed by the damage leads to fluid build-up.

Clinical manifestations of AHS can be classified based on clinical signs (organs affected) and severity, and there exists four distinct clinical manifestations of AHS (THEILER 1921; MELLOR AND HAMBLIN 2004; CLIFT AND PENRITH 2010). The least severe form is the fever form. The sub-acute cardiac form is categorised by fever and sub-cutaneous dropsy of the brain, neck and legs. The mortality rate for this form is about 50%. The pulmonary form is categorised by pulmonary dropsy (fluid in the chest), resulting in respiratory failure, acute sweating and foaming from the nose. This form has the highest mortality rate of up to 95%. Death usually occurs within a week following infection and is ultimately caused by respiratory distress, oedema, and pleural effusion. The fourth form, which is a combination of the cardiac and pulmonary forms, is the most prevalent form with mortality rates of around 70% (COETZER AND GUTHRIE 2005; CLIFT AND PENRITH 2010).

1.2.5 Prevention

The economic implications, impacts on animal husbandry as well as horse racing/shows, and animal safety and welfare considerations led to the need for vaccines against BTV and AHSV for prevention of disease in non-endemic areas. There are two main vaccination strategies against AHSV/BTV, utilising either live-attenuated or inactivated vaccines. Live-attenuated vaccines (LAVs) contain avirulent strains of AHSV/BTV, and in South Africa the polyvalent LAV against AHSV is produced by Onderstepoort Biological Products (VON TEICHMAN *et al.* 2010). Limitations and concerns regarding LAVs include disease development in immunocompromised horses, the spread of the vaccine strains by vectors, reversion to virulence in the host and generation of reassortant strains (WEYER *et al.* 2016; WEYER *et al.* 2017). Inactivated vaccines are better than LAVs in terms of safety regarding reversion to virulence, uncontrolled spread and reassortment. The problem with inactivated vaccines is their low efficacy. Antibody responses depend on the adjuvant, amount of antigen and vaccination route (VAN RIJN *et al.* 2020). WERNERY *et al.* (2020) evaluated the efficacy of vaccinations produced by the Central Veterinary Research Laboratory in Dubai. Inactivated vaccines containing all 9 serotypes did indeed provide protection against AHS through the production of neutralizing antibodies (WERNERY *et al.* 2021). However, 12% of horses still developed African horse sickness fever, which is the mild subclinical form of AHSV often seen in donkeys and zebras. To address the need for safe and more effective vaccines, several new technologies and approaches have recently been evaluated.

Reverse genetics (RG) is an important system for generation of recombinant virus strains. These strains can provide insight into the replication and the pathogenicity of viruses and more specifically the functions of certain virus proteins. It provides a way for directly investigating certain viral proteins and their significance in virus replication. In the *Reoviridae* family, the first ground breaking study was the development of a reverse genetics system for reovirus in which the introduction and essentially the substitution of a specific gene segment from cDNA allowed for the recovery of an infectious recombinant virus (KOMOTO *et al.* 2006). This study paved the way for the development of reverse genetics systems for other members of the *Reoviridae* family. In essence reverse genetics is the manipulation of the virus replication cycle to produce recombinant viruses. Cells are given synthetic transcripts that appear native to the virus. These are then incorporated into the viral infection cycle, translated, replicated and packaged (BOYCE AND ROY 2007).

Reverse genetics have paved a way of overcoming the limitations of inactivated vaccines. Strategies that have proven effective for RG vaccine development are Entry Competent Replication-Abortive (ECRA) viruses and Disabled Infectious Single Animal (DISA) (VAN RIJN *et al.* 2018; CALVO-PINILLA *et al.* 2020; SULLIVAN *et al.* 2021). For AHSV, ECRA-AHSV viruses have been generated. LULLA *et al.* (2016) disrupted the open reading frames (ORF) of AHSV Seg-9, encoding VP6 and NS4, by introduction of multiple stop codons. They confirmed that the virus containing this mutated Seg-9 did not reveal any changes when grown in a complement cell line, concluding that the mutated ORF is improbable to revert to the virulent form. This complementation system was firstly developed for BTV and then for AHSV (ROY 2020). Furthermore, the protective effect of a vaccine containing this mutant Seg-9 revealed generation of neutralizing antibodies and consequently zero viremia in newly born foals (LULLA *et al.* 2017). Also, a ECRA vaccine comprising all nine vaccine strains was injected into ponies and over 6 months developed neutralising antibody responses against all 9 serotypes, showing a good replacement option for LAVs (SULLIVAN *et al.* 2021). For DISA, viruses deficient for the functional gene of NS3/NS3a were generated which inhibited virus egress. This strain did not show inhibited viral replication in cell culture, but the virus was not present in blood of vaccinated horses reducing the risk of spread or transmission by midge vectors during feeding (VAN DE WATER *et al.* 2015). Recombinant AHSV strains rescued by RG which lack expression of NS4 were recently shown to be fully attenuated in experimentally infected horses (WALL *et al.* 2021)

Transient plant-based expression systems are a rapid and highly scalable means of producing AHSV proteins for vaccine purposes (KUSHNIR *et al.* 2012). Virus-like particles (VLPs) are also deemed as a safe and effective alternative to LAVs for various diseases. These consist of viral structural proteins lacking the dsRNA viral genome. They display the same size and morphology as a normal virus but cannot replicate. RUTKOWSKA *et al.* (2019) co-expressed the four capsid proteins (VP2, VP5, VP3 and VP7) of AHSV in plant cells. Chimeric VLPs, containing capsid proteins produced from different AHSV serotypes, were injected into foals and successfully stimulated the formation of serotype-specific neutralising antibodies (RUTKOWSKA *et al.* 2019). Most recently FEARON *et al.* (2021) showed

that AHSV-5 VP7 quasi-crystals also produced in plant cells induce both humoral and cell-mediated responses in guinea pigs, thus signifying the activation of adaptive immunity. However, experimentation is needed to evaluate these VP7 quasi-crystals as an effective vaccine candidate. This could also be used for diagnostic purposes, as plant-produced AHSV-5 VP7 accurately detected AHSV-specific antibodies (FEARON *et al.* 2022). Additionally, O'KENNEDY *et al.* (2022) produced AHSV serotype 5 chimeric VLPs as well as soluble VP2 in *Nicotiana benthamiana*. When the adjuvanted VLPs or VP2's were injected into IFNAR $^{-/-}$mice it resulted in high titres of antibodies, signifying a safe effective plant-produced vaccine.

Another effective vaccine candidate is based on a modified vaccinia Ankara (MVA) virus (CASTILLO-OLIVARES *et al.* 2011). These are non-replicative live viral vectors designed to express and present viral antigens to the host immune system. Studies have shown that vaccination with these modified viruses expressing one or more AHSV outer capsid proteins like VP2, elicits an immune response in both mice and horses (ALBERCA *et al.* 2014; AKSULAR *et al.* 2018). MARÍN-LÓPEZ *et al.* (2020) also produced recombinant modified vaccinia Ankara (MVA) virus expressing NS1. This provided complete protection against the infection of AHSV-4 when injected into mice. Recombinant expression of only the capsid proteins, have also been explored extensively for BTV and other orbiviruses (FRENCH AND ROY 1990). Baculoviral expressed VP2 and VP5 was used for vaccination of horses and induced neutralizing antibody production and conferred protection against AHSV (ROY *et al.* 1996). Although they give good protection against challenge with homologous serotypes, problems associated with their large scale purification and stability have made commercialisation of these vaccines problematic (ALPAR *et al.* 2009)

There are a few other antiviral strategies still under investigation. Compounds inhibiting viral replication (virostatic drugs) have also been studied that seem to impede BTV cytopathogenic impacts/ cytopathic effect (CPE) on host cells (GU *et al.* 2012). The polyanionic aromatic compound aurintricarboxylic acid (ATA) has recently been shown as a potential antiviral compound against orbivirus infections. *In vitro* ATA effectively inhibited viral replication in mammalian and insect cells (ALONSO *et al.* 2020). However, ATA did not prevent BTV or AHSV infection when used as prophylactic measure in mice. Speculations are that immune system regulation influences the *in vivo* activity of ATA.

In conclusion, many vaccine candidates and antivirals are available for the defence against AHSV (JIMÉNEZ-CABELLO *et al.* 2022). Unfortunately, each of these come with their own restrictions if it be lack of efficacy, safety, or production challenges.

1.3 AHSV molecular biology

1.3.1 Virus genome

The AHSV genome consists of 10 linear double-stranded RNA (dsRNA) segments. The segments are named segment 1 (Seg-1) to segment 10 (Seg-10) in order of decreasing size, with Seg-1 the largest at around 4000 bp, and the smallest Seg-10 only 750 bp in size (BREMER *et al.* 1990; ROY *et al.* 1990; VAN STADEN AND HUISMANS 1991). Each genome segment contains one gene, which encodes one protein, except segments 9 and 10 that both contain more than one open reading frame (ORF). The encoded proteins are divided into two groups, namely the seven structural viral proteins (VP1-VP7) and the four non-structural proteins (NS1-NS4) (BREMER *et al.* 1990; BELHOUCHET *et al.* 2011; ZWART *et al.* 2015). The dsRNA genome is encased in a non-enveloped triple-layered icosahedral particle. The outer capsid is formed by VP2 and VP5, VP7 forms the outer core layer, and VP3 the sub-core. VP1, VP4 and VP6 form the transcription complex contained within the core. The non-structural proteins are only detected during viral replication and do not form part of the virion structure.

There are 9 serotypes of AHSV, AHSV-1 to AHSV-9, based on the neutralising antibody responses generated against the outer capsid protein VP2 (MCINTOSH 1958; HOWELL 1962). There is large VP2 gene sequence diversity among different strains of AHSV and other orbivirus species. (ROY *et al.* 1994). The high mutation rates of orbiviruses inside vertebrate hosts as well as founder effects in vectors led to evolution into quasispecies. In segmented viruses these refer to the variation in segments between different strains of BTV and AHSV, respectively. This is due to each segment evolving independently from one another (BONNEAU *et al.* 2001). Variation between strains are also frequently observed due to whole segment reassortment where one virus will have genome segments coming from different strains (WEYER *et al.* 2016). Intragenic recombination increases genetic diversity in viruses, resulting in increase in possible hosts, or adaptability to a changing environment driven by selection pressures such as vaccines or antivirals (PÉREZ-LOSADA *et al.* 2015). Seg-1, Seg-6, Seg-7 and Seg-10 showed evidence of intragenic recombination in AHSV (NGOVENI *et al.* 2019).

The untranslated regions (UTRs) at both 5’ and 3’ ends of all genome segments are well conserved in each genome segment. These regions appear to interact with each other, by forming hairpin-loops. It has been shown that the function of these UTRs facilitates segment recognition and packaging during the virus replication cycle (VAN STADEN *et al.* 1991; BOYCE *et al.* 2012; BURKHARDT *et al.* 2014; BORODAVKA *et al.* 2018). Orbiviruses enable co-packaging of their segmented genome through the use of trans-acting RNA-RNA interactions (tRRIs). A copy of the complete viral genome is directed for packaging by tRRIs that occur between different viral genome segments (NEWBURN AND WHITE 2019). In BTV, the smaller mRNA segments (segments 7 to 10) play a crucial role in RNA complex formation and thus packaging (ROY 2017). Segment assortment signals (SASs) is the term

used to describe the tRRIs multiple specific sites, which are dispersed across each segment (ALSHAIKHAHMED *et al.* 2018).

Open reading frames (ORF) are under selection pressures. Production of synthetic BTV with deleted/silenced regions in Seg-10 showed that the mutant viruses were not genetically stable (FEENSTRA *et al.* 2014b; FEENSTRA *et al.* 2016). Cis-complementation or random insertions occurred from other segments in genome to compensate for these mutations, but the virus was not able to produce NS3. Seg-10 RNA elements was conserved among orbiviruses as there were exchanges, shown by BTV chimeras containing NS3A ORFs of other orbivirus species.

1.3.2 Viral structural proteins

1.3.2.1 AHSV outer capsid

The outer capsid of AHSV is necessary for host cell recognition, adherence and entry. The viral structural proteins VP2 and VP5 form the outermost layer of the capsid, and directly connect and associate with host cells. VP2 arranges on top of VP5. Both these proteins contribute to the anti-AHSV immune response, due to their external localisation and hence ability to be recognised by the host's immune system (HUISMANS AND ERASMUS 1981).

Seg-2 encodes VP2 (1055 amino acids (aa) and MM = 120 kDa) (POTGIETER *et al.* 2003). This gene segment is extremely variable and has hardly any sequence conservation between different serotypes in AHSV (ROY *et al.* 1994). Thus, VP2 is the main determinant of viral serotype. VP2 is a hydrophilic protein that contains many charged amino acids, and forms trimers through disulphide bonds (HASSAN AND ROY 1999). Immune responses are triggered by neutralising epitopes present on VP2 (BURRAGE AND LAEGREID 1994). For both AHSV and BTV the VP2 multimers form a tip and hub domain. The tip domains project outwards and are responsible for the triskelion appearance of VP2 (MANOLE *et al.* 2012). In BTV, these tips have perfect positioning to recognise and bind to host cell receptors, specifically cellular glycophorin A, a membrane spanning protein containing many sialic acid residues (HASSAN AND ROY 1999; ZHANG *et al.* 2010). The hub domain is necessary for binding to sialic acids, mediating infection into the host cell. This might also be true for AHSV, as VP2 exchange has been shown to affect the growth kinetics of the virus (VAN DE WATER *et al.* 2015). However, there are some structural differences between AHSV and BTV core particles. AHSV contain spikes in the central hub, as well as a protein plugging the 5-fold vertices, whereas BTV does not (MANOLE *et al.* 2012).

Seg-6 encodes VP5 (504 aa and MM = 57 kDa). This protein is moderately variable, with some sequence conservation between different serotypes in orbiviruses (HASSAN *et al.* 2001). VP5 contains many non-polar residues, and like VP2 forms trimers (DU PLESSIS AND NEL 1997). VP5 consist of many α-helices and is characterised as a globular multimeric protein (STASSEN *et al.* 2011). Importantly, helices at the N-terminal end (amphipathic fusion peptide) are responsible for

membrane permeabilizing activity (HASSAN *et al.* 2001). In BTV, the VP5 monomers consist of three domains namely the dagger, unfurling and anchoring domains (ZHANG *et al.* 2016).

1.3.2.2 AHSV core

The core is responsible for protecting the viral genome from host immune responses, and for replication of the virus genome. The AHSV core is formed by two major structural proteins (VP3 and VP7) and three minor structural proteins (VP1, VP4 and VP6), representing the transcription centre with the dsRNA trapped inside (MERTENS AND DIPROSE 2004). The dsRNA genomes occupy only one to two thirds of the space available in core particles of viruses in the *Reoviridae* family (DESSELBERGER 2020).

Seg-3 encodes VP3 (905 aa and MM = 103 kDa). This segment is highly conserved between different orbivirus serotypes (MAREE *et al.* 1998a). VP3 contains many hydrophobic amino acid residues. Seg-7 encodes VP7 (349 aa and MM = 38 kDa) and is also well conserved between serotypes. VP3 assembles into the scaffold or sub-core layer, and associates by automatic hydrophobic bonding with VP7 to form core particles (BASAK et al. 1997). Knockdown studies show that when VP7 is absent, VP3 moves to the proteasome and gets degraded (BASAK *et al.* 1997). VP7 at the outermost layer of core particles then associates with VP5 proteins to facilitate assembly of the outer capsid. The fact that core particles are infectious in arthropod but not mammalian cells, suggests their function in insects for cell attachment and entry (MERTENS *et al.* 1987). AHSV VP7 always, irrespective of the presence of VP3, also forms insoluble crystal hexagons in the cell cytoplasm (BURROUGHS *et al.* 1994). The VP7 crystalline particles were however shown to play a role in AHSV yield and release (BEKKER *et al.* 2022) These insoluble crystals sequester a large proportion of the available VP7 protein, preventing it from being integrated into newly formed cores, which might impact negatively on virus replication (BEKKER *et al.* 2014). When a seven amino acid region of VP7 was targeted by mutagenesis to produce only soluble VP7, VP7 was still capable of interacting with VP3 however no VP7 self-assembly into crystalline particles occurred (BEKKER *et al.* 2022).

Inside the virus core the minor structural proteins VP1, VP4 and VP6 integrate to form a flower-like transcription complex, which is in contact with the VP3 scaffold. All three proteins have enzymatic activity (MERTENS AND DIPROSE 2004).

Seg-9 encodes VP6 (369 aa and MM = 43 kDa). This protein is hydrophilic and contains many positively charged residues capable of association with nucleic acids (DE WAAL AND HUISMANS 2005). VP6 forms hexameric complexes. These are distinctive of helicases and STAUBER *et al.* (1997) suggested that VP6 acts as a viral helicase when single or double stranded RNA is present. RNA-binding, ATP-binding, ATPase and helicase activities are found in interlinked structural motifs of the protein (KAR AND ROY 2003; MATSUO *et al.* 2014). VP6 also appears to bind genomic viral RNA to

help facilitate genome packing, and genomic viral RNA seems to in turn help with VP6 encapsidation (SUNG *et al.* 2019).

Seg-1 encodes VP1 (1305 aa and MM = 150 kDa), the largest of the virus proteins. VP1 is an RNA-dependent RNA polymerase (RdRp), the viral polymerase, and is necessary for RNA strand synthesis by usage of a single-stranded RNA template (ROY *et al.* 1994; BOYCE *et al.* 2004). This RdRp firstly forms plus strand ssRNA transcripts from every genome segment. These have the same function as mRNAs. In the cytoplasm they are used for translation of viral proteins. Secondly, minus strand ssRNA are generated from the plus strand ssRNA to achieve dsRNA through complementation inside new viral cores. The amino and carboxyl terminal regions of VP1 function to stabilize the protein to guarantee efficiency of enzymatic domain.

Seg-4 encodes VP4 (642 aa and MM = 76 kDa). This protein forms dimers through a leucine-zipper conserved at the carboxyl end and is important for capping of the viral mRNAs (RAMADEVI *et al.* 1998). Enzymatic functions of VP4 are as follows: RNA triphosphatase activity, guanylyltransferase activity, methyltransferase activity and pyrophosphatase activity. RNA triphosphatase produces 5' diphosphate ends on mRNA, guanylyltransferase generates a 5'-guanidylic acid, and methylation of the 5' cap and second 5' base occurs via guanine-7-methyltransferase and 2'-O-methyltransferase. The pyrophosphatase activity prevents inactivation of the polymerase activity due to by-products (MARTINEZ-COSTAS *et al.* 1998; ROY 2008).

1.3.3 Viral non-structural proteins

1.3.3.1 NS1, NS2 and NS4

Non-structural proteins are produced inside the host cell by using the host's translational machinery but are not found inside the virus particle itself. In AHSV, the non-structural proteins are important for different aspects of the infection cycle, including virus release and pathogenesis. Although they do not directly play a role in viral replication, they do give increased replicative efficacy (OWENS *et al.* 2004).

Seg-5 encodes NS1 (548 aa and MM = 63 kDa). This segment is highly conserved between different serotypes of AHSV (MAREE AND HUISMANS 1997). NS1 is an extremely hydrophobic protein, which naturally forms tubular structures sequestered in the cytoplasm near or around the nucleus (EATON *et al.* 1988; VAN STADEN *et al.* 1998). The tubules are fragile, fretted with smooth edges and have a diameter of about 23 nm and lengths of up to 4 nm. NS1 contains many cysteine residues which can form disulphide bonds contributing to the ordered structure of the tubules. MONASTYRSKAYA *et al.* (1994) observed that deletion or loss of cysteine amino acids (positions 337 and 340) is detrimental to functioning of NS1. The result was non-functional structures aggregating to one another. Deletions at the N-and C-terminals showed corresponding effects (MONASTYRSKAYA *et al.* 1994; VAN STADEN *et al.* 1998). In BTV, NS1 plays a major role as a positive regulator of viral protein synthesis as shown by BOYCE *et al.* (2012), where the NS1 expression was the highest of all the viral proteins in

mammalian cells. NS1 particularly causes the upregulation of viral mRNAs. This leads to the virus outcompeting the host's mRNA quantities. The host cell's translational machinery is reserved by the virus, inhibiting the host's protein production. Observations of this predominating viral protein synthesis was observed by BOYCE *et al.* (2012). In contrast, this does not happen in insect cells. Translation of insect proteins can still occur, and viral mRNAs do not interfere or outcompete insect mRNAs. Insect cells subsequently stay viable and capable of their own translation. NS1 also integrates with centrosomes. This hinders the normal cell cycle, inhibiting host DNA and RNA synthesis and subsequent death of cell (SHAW *et al.* 2013). NS1 is also needed for release of virus from host cells (OWENS *et al.* 2004). In BTV specifically, KERVIEL *et al.* (2019) showed that an NS1 monomer contains two metal binding, zinc-finger-like motifs and a long C-terminal arm and appears in two forms, tubular or non-tubular. Deletion of this carboxyl terminal prevents tubule formation but not viral replication. The non-tubular form of NS1 upregulates BTV mRNA translation. Disruption of the zinc-fingers diminishes viral mRNA translation, tubule formation and virus replication. This shows functional importance of zinc-fingers in the non-tubular NS1. From personal communications no one has been able to rescue mutant AHSV lacking NS1 or solely lacking the conserved region close to the C-terminal end.

Seg-8 encodes NS2 (365 aa and MM = 46 kDa). Intermediate Seg-8 sequence conservation is observed between different serogroups, where the N-termini are most conserved and the C-termini varying the most (VAN STADEN *et al.* 1991; UITENWEERDE *et al.* 1995). NS2 is the next most highly expressed protein after NS1. NS2 forms large, globular structures called viral inclusion bodies (VIBs) in the cytoplasm around the nucleus (perinuclear) (UITENWEERDE *et al.* 1995). Assembly of new viruses occurs within these VIBs (BROOKES *et al.* 1993). Association with VP3 and minor structural proteins (VP1, VP4 and VP6) confers NS2 with the ability to sequester viral ssRNAs and viral elements for packaging and encasing (KAR *et al.* 2007). The ssRNA binding affinity is proposed to be due to RNA secondary structures interacting with NS2 (LYMPEROPOULOS *et al.* 2006). NS2 is a phosphorylated viral protein. The phosphorylation of two serine amino acids at the carboxyl terminal confers stability to the VIBs (MODROF *et al.* 2005).

Seg-9 also encodes NS4 (143 aa and MM = 17 kDa) not just VP6. Prediction of a second ORF on Seg-9, other than that of VP6, led to its discovery (FIRTH 2008; ZWART *et al.* 2015). This overlapping gene shows strong internal synonymous constraint in VP6. The 5' and 3' termini of Seg-9 also has high synonymous constraint (SEALFON *et al.* 2015). These constraint regions at the ends of the segment might be important for genome packaging and replication, or translation as seen in other viruses of the *Reoviridae* family like rotavirus (LI *et al.* 2010). NS4 has variable conservation between serotypes. AHSV strains express either NS4-I or a somewhat longer NS4-II. AHSV NS4 has nucleic acid binding abilities, and localises to both the cytoplasm and the nucleus (BOUGHAN *et al.* 2020). A subtype of NS4-II containing a nuclear localization signal (NLS), named NLS-NS4-II, displays distinct punctate foci in the nucleus. Due to its small size NS4 can passively diffuse into the nucleus or be

actively transported through use of the NLS. BOUGHAN *et al.* (2020) also showed that NS4 co-localises with promyelocytic leukaemia nuclear bodies (PML-NBs). These PML-NBs regulates antiviral gene expression and thus NS4 potentially affects the host antiviral response. While the precise functioning of NS4 is still unclear in AHSV, and it seems to not be necessary for replication, it is an important virulence factor and is able to disrupt the JAK-STAT signalling pathway involved in host innate immunity (WALL *et al.* 2021). In BTV, RATINIER *et al.* (2016) observed that the expression of NS4 gives the virus a replicative advantage in the presence of interferon, and modulates the host interferon response.

1.3.4 NS3 and Seg-10

Seg-10 encodes NS3 (217 aa and MM = 24 kDa) (VAN STADEN AND HUISMANS 1991). This segment is the second most variable amongst different serotypes in AHSV, with Seg-2 encoding VP2 being the most variable (VAN NIEKERK *et al.* 2001b). As described under the virus genome section, this segment is also involved in tRRIs for correct packaging of the BTV segmented genome (NEWBURN AND WHITE 2019). Seg-10 has two in-frame start codons, encoding NS3 and a truncated version, named NS3A (206 aa and MM = 23 kDa), which lacks the 11 N-terminal amino acids of NS3. NS3/A are integral membrane proteins, with their membrane association mediated by two hydrophobic transmembrane regions (VAN STADEN *et al.* 1998). These proteins are translated on the rough endoplasmic reticulum (RER) from where they are trafficked to the cell membrane. Non-enveloped viruses like BTV and AHSV are released from host cells via lytic and non-lytic mechanisms (VENTER 2014). NS3 is found at virus exit sites and is required for the non-lytic release of new viruses from the host cell (HYATT *et al.* 1991; STOLTZ *et al.* 1996). Both transmembrane domains are essential for NS3's correct functioning. Their proper conformations are essential for trafficking virus particle to sites of release. If these domains are disrupted, aggregates of virions form due to reduced virus release from host cells (FERREIRA-VENTER *et al.* 2019).

It is proposed that NS3 creates hydrophilic pores in the membrane bilayer, which has negative effects on normal cellular transport and gradients (VAN STADEN *et al.* 1995; VAN NIEKERK *et al.* 2001a). FEENSTRA *et al.* (2014a) showed through mutational knockouts of BTV Seg-10 that NS3 is not essential for viral replication. Knockout of NS3 reduced pathogenesis by decreasing replication and release of virus, but not completely stopping it (VAN GENNIP *et al.* 2014). NS3 contains two polybasic motifs (PMB1/PMB2), conserved in orbiviruses. These motifs act as trafficking signals facilitating the retainment of NS3 in the endoplasmic reticulum, as well as exporting of NS3 from the Golgi apparatus to the plasma membrane (LABADIE *et al.* 2019). Recently BTV NS3 was specifically shown to bind serine/threonine-protein kinase B-Raf (BRAF) and activate the mitogen-activated protein kinase/extracellular signal-regulated kinase (MAPK/ERK) pathway (KUNDLACZ *et al.* 2019). In doing this, BTV manipulates the MAPK/ERK pathway to support its own replication. This binding motif to BRAF appears to be only conserved in BTV compared to other members of the *Orbivirus* genus. AHSV NS3 was found to stabilize membrane interactions through a conserved region near

the N-terminal containing a myristylation motif together with an area of basic amino acid residues (VAN NIEKERK *et al.* 2001b). The NS3 protein of AHSV only shares 30% sequence homology with BTV NS3, which may explain some of the differences. Furthermore, NS3 of BTV is also much more conserved across serotypes, having about 93% similarity, whereas AHSV NS3 has about 64% similarity (HWANG *et al.* 1992; SAILLEAU *et al.* 1997; MARTIN *et al.* 1998; VAN NIEKERK *et al.* 2001b).

When MARTIN *et al.* (1998) compared the sequence homology in AHSV Seg-10 from isolates of different serotypes, they grouped into 3 different clades. Serotypes 4, 5, 6 and 9 belong to group α, serotypes 3 and 7 to group β and serotypes 1, 2, and 8 to group γ. VAN NIEKERK *et al.* (2001b) further confirmed this clustering of AHSV NS3 into the different clades. They compared NS3 protein sequences of South African AHSV field isolates, vaccine strains and reference strains. NS3's variation was determined to be 36% across serotypes and 27% within serotypes. Variation is often caused by the reassortment of AHSV genome segments. Multiple AHSV serotypes simultaneously present in hosts or vectors may facilitate Seg-10 reassortment between virus populations of different serotypes. The serotypes in the same clade appear to exchange the NS3 gene more readily. QUAN *et al.* (2008) found the two hydrophobic NS3 regions, corresponding to the transmembrane helices (VAN STADEN *et al.* 1995), to have decreased variation and be under strong purifying (negative) selection. Between the two hydrophobic regions there is a small hydrophilic region, more variable in nature, showing signs of positive selection. NS3 was continuously shown to be the second most variable AHSV protein despite the two highly conserved transmembrane regions. This variability of NS3 may be explained by immunological pressures acting on it.

Sealfon et al. (2015) identified an additional overlapping +1 ORF in BTV Seg-10 by using bioinformatics tools. STEWART *et al.* (2015) showed that the additional ORF expresses a small protein (59 aa) called S10-ORF2, that shows a nucleolar localization when expressed from transfected plasmids. This will be described in detail under the section on additional overlapping open reading frames. This additional overlapping +1 ORF is also present in AHSV Seg-10 (Stewart et al. 2015), but has not yet been investigated nor characterised.

1.4 Virus replication cycle

The orbivirus replication cycle has a few essential steps. First adsorption and entry must occur, followed by gene expression and genome replication. Assembly of new viruses, viral trafficking and release from host cells occur last.

As mentioned previously, VP2 mediates recognition and binding to the host cell receptors, via binding of cell membrane surface glycoproteins. AHSV attaches and enters the host cell via clathrin-mediated endocytosis and macropinocytosis (HASSAN AND ROY 1999; FORZAN *et al.* 2004; NOAD AND ROY 2009; VERMAAK *et al.* 2016). Once inside, an early endosome gets produced (ZHANG *et al.* 2010). The characteristic low pH of this endosome causes VP2 degradation and activation of VP5.

The amphipathic α-helices integrate with the late endosomal membrane, disrupting it. This is characteristic of the permeabilisation capabilities of VP5 (HASSAN *et al.* 2001). This consequently releases the virus core into the cytoplasm of the host cell (ROSS-SMITH *et al.* 2009). In BTV the specific mechanism of action of the VP5 protein was investigated. The low pH triggers three VP5 actions. The dagger domain is projected, the VP5 unfurling domains fold into a six-helix stalk and a surface loop is converted to a β-hairpin and inserts into a gap formed between two adjacent VP7 trimers to anchor VP5 to the BTV core. This gap through the endosomal membrane enables the viral core to enter the cytosol (XIA *et al.* 2021).

The replication cycle of orbiviruses happens in two stages, primary and secondary replication cycles (MATSUO AND ROY 2009). The primary replication cycle constitutes the initial transcription of 10 RNA segments inside the core (MATSUO AND ROY 2013). The core particles prevent dsRNA from entering the host cytoplasm, which would consequently activate host immune responses (DIPROSE *et al.* 2002; STUART AND GRIMES 2006). With VP2 and VP5 (outer capsid proteins) gone, the core actively transcribes all 10 dsRNA segments from the flower-shaped transcription complex (VP1, VP4 and VP6) (NOAD AND ROY 2009). VP6 (viral helicase) unwinds the dsRNA, and the individual minus sense ssRNA molecules can then act as templates. This requires ATP for energy to unwind strands as soon as VP6 is bound to nucleic acids (KAR AND ROY 2003; DE WAAL AND HUISMANS 2005). VP1 (viral polymerase) synthesises plus strand sense mRNA from minus sense template strands (URAKAWA *et al.* 1989; BOYCE *et al.* 2004). VP4 (capping enzyme) modifies the mRNA to increase its stability and inhibit its degradation by the host cell, when released from the core, by addition of a 5' cap. This also ensures the mRNA's efficient recognition by the translational machinery of the host cell (MARTINEZ-COSTAS *et al.* 1998; RAMADEVI *et al.* 1998). Translation of these plus sense ssRNA (mRNA) molecules results in the expression of the seven viral structural and four non-structural proteins.

The secondary replication cycles constitutes the replication and assembly of progeny virions in the viral inclusion bodies (VIBs) formed by phosphorylated NS2 (ROSS-SMITH *et al.* 2009). The VIBs recruit all viral structural proteins as well as the 10 ssRNA segments (KAR *et al.* 2007). Secondary structures formed by RNA-RNA interaction produce an RNA complex which can now be packaged (SUNG AND ROY 2014).

Secondary replication is induced by association of VP1 and VP4 with ssRNA and VP6 (LOURENCO AND ROY 2011). VP3 multimers integrate with the transcription complex to form the transcriptionally active sub-core. Following encapsidation by VP3, the ssRNA templates are converted to dsRNA by VP1. VP7 subsequently assembles on top of the VP3 scaffold, stabilizing the cores (MATSUO AND ROY 2013). The 5-fold vertices in the VP3-VP7 core are necessary for persistent transcription. Substrates and metal ions enter, and waste products and new transcripts get released (BASAK *et al.* 1997). Thus, new cores either continually transcribe to maintain production of viral proteins or get surrounded by an outer capsid which stops transcription. The outer capsid is added when progeny

virions depart from the VIBs. They are then transported to the cell membrane for subsequent release (KAR *et al.* 2007). NS3 is the mediator of this release. Although the precise cell trafficking mechanisms are not known, it is proposed that the cell cytoskeleton might play an important role, specifically interactions with vimentin-rich intermediate filaments (PLOUBIDOU AND WAY 2001; SMITH AND ENQUIST 2002).

NS3/NS3A mediates the release in both insect and mammalian cells but differ in the mechanism (HYATT *et al.* 1989; STOLTZ *et al.* 1996; FERREIRA-VENTER *et al.* 2019). In BTV the carboxyl terminal of NS3 has binding motifs for VP2, and the amino terminal interacts with secretory proteins, subsequently recruiting the virus particles to the cell membrane (BEATON *et al.* 2002; WIRBLICH *et al.* 2006). In mammalian cells primarily lytic release occurs, causing disruption of the plasma membrane, this might contribute to cytotoxicity and host pathogenesis. In insect cells non-lytic release (like budding) occurs, virus particles project through the cell membrane causing no damage to cell membrane leading to persistent infection of the virus and facilitate optimal spread (VENTER 2014). In insect cells, both NS3 and NS3A are required for non-lytic release in the case of BTV. In mammalian cells only NS3 expression is needed for virus release. Thus, NS3A may have a role in facilitating non-lytic release (CELMA AND ROY 2011; VAN GENNIP *et al.* 2014). Different levels of NS3 expression are proposed to be the reason for this difference in release mechanisms. OWENS *et al.* (2004) showed that the NS1:NS3 ratio is a major determinant, in insect cells both are highly expressed throughout infection, but low NS3 levels in mammalian cells corresponds to lytic release. Mammalian cells have their own inherit exocytosis. This may also lead to differences in lytic and non-lytic release mechanisms (CELMA AND ROY 2009).

1.5 Overlapping open reading frames

1.5.1 Viral ORF translation and RNA structure

Viruses do not encode their own ribosomes and are entirely dependent on host translational machinery. Normal canonical eukaryotic translation synthesises one protein from one mRNA. Viruses often need to express multiple proteins from one or a few RNA molecules and have evolved different non-canonical mechanisms to optimise the translation of their ORFs. These include for example ribosome shunting, internal ribosome entry, non-AUG initiation, leaky scanning, ribosomal frameshifting, stop-codon readthrough and re-initiation (BELLI AND SAMUEL 1993; RACINE *et al.* 2007; FIRTH AND BRIERLEY 2012). These different translational mechanics are typically modulated by RNA secondary structures, hence these RNA elements/regions are highly conserved because of their essential functionality.

A viral genome contains RNA structures that regulate different viral processes (GUPTA AND BANSAL 2020). Studies on dengue virus show the complex formation of dynamic structures of genomic RNAs during different stages of their life cycle (BROSNAN AND VOINNET 2009; PONTING *et al.* 2009; IGLESIAS AND GAMARNIK 2011).

In eukaryotes, mRNAs possess a 5' cap and initiation of translation can be described by a scanning model and/or cap-dependent translation initiation. The ribosome consists of a 4OS and 6OS ribosomal subunit. First, a 4OS pre-initiation complex associates with mRNA at the 5′-cap. This complex scans along the mRNA searching for an AUG codon in the right context. Once found, the 60S ribosomal subunit binds and translation begins (CONTE AND PENNELL 2013). When ribosomes fail to start translation from the first AUG, they continue scanning downstream until reaching the second open reading frame AUG codon, allowing expression of multiple proteins from a single mRNA. This is called leaky scanning translation (ERNST AND SHATKIN 1985). Non-AUG initiation occurs when similar codons to that of a start codon is recognized by Met–tRNA, initiating translation at other reading frames (KOZAK 1990b).

Ribosomes can sometimes shunt over parts of the 5' untranslated region accessing downstream ORFs, resulting in 5' independent translation, called ribosome shunting (FIRTH AND BRIERLEY 2012). RNA leader sequences present on mRNA form hairpin loops, inhibiting ribosome scanning. The ribosome then bypasses a large portion translation ORFs further down the mRNA (POOGGIN *et al.* 2006). During translation, re-initiation can also occur when the 40S subunit from the ribosome remains associated with mRNA, further scanning, and reinitiating translation at a downstream AUG codon (JACKSON *et al.* 2012). The presence of RNA structures stimulate this re-initiation (KOZAK 1990a). Ribosomal frameshifting occurs when a ribosome slips by one base in either direction during translation, resulting in translation in a different reading frame (JACKS AND VARMUS 1985).

Internal Ribosome Entry Sites (IRES) recruit ribosomes directly to internal positions on mRNAs, facilitating efficient translation of an mRNA in a 5' and cap-independent manner (JANG *et al.* 1988; BALVAY *et al.* 2009). An IRES element is an RNA sequence present in 5' untranslated regions. These sequences are proposed to fold to adopt highly organized, complex RNA structures that are capable of recruiting ribosomes to internal AUG initiation codons. Examples of IRES that result in translation of additional ORFs have been identified in several virus families (BALVAY *et al.* 2009).

Stop codon readthrough occurs when the ribosome miscodes at a stop codon and translation continues to the next termination codon (NAMY AND ROUSSET 2010; NAPTHINE *et al.* 2012). As seen in ribosomal frameshifting, codon read-through is also modulated by features located in the 5'or 3' end of the mRNA. These suppress the stop codon recognition and result in continuing translation of further ORFs (NAMY *et al.* 2006).

In the *Reoviridae* family, several viruses have been characterised as having one or two genome segments which are polycistronic, containing more than one ORF per mRNA. In the genus *Fijivirus*, Mal de Rı´o Cuarto virus (MRCV) has two non-overlapping open reading frames found in both genomic segment 7 and segment 9 (GUZMÁN *et al.* 2007). In the genus *Orthoreovirus,* segment 1 of avian reovirus encodes for three different proteins from overlapping ORFs, each having a different function (RACINE AND DUNCAN 2010). The first two overlapping ORFs of avian reovirus segment 1

are translated by leaky scanning translation. The third overlapping ORF is translated via ribosome shunting. In the genus *Phytoreovirus,* rice dwarf virus (RDV) segment 12 and wound tumour virus (WTV) segment 9 both have an additional smaller ORF embedded inside the larger main ORF. In these cases, a leaky scanning mechanism was identified for production of two proteins from each mRNA (SUZUKI *et al.* 1996). In some rotavirus strains, the smallest genome segment (Seg-11) contains 2 overlapping reading frames encoding major protein NSP5 and an additional +1 alternative ORF. The +1 ORF was shown to express a small protein, NSP6, which localises to mitochondria (MATTION *et al.* 1991; RAINSFORD AND MCCRAE 2007).

Until recently all genome segments of viruses in the genus *Orbivirus,* including AHSV and BTV, were believed to be monocistronic. However as described in other genera of the *Reoviridae* family, segments containing more than one ORF were recently identified for several orbiviruses. Novel proteins have also been identified to be translated from these additional open reading frames. The first to be identified and characterized was BTV NS4, encoded in a +1-frame relative to the main VP6 ORF on Seg-9 (FIRTH 2008; BELHOUCHET *et al.* 2011; RATINIER *et al.* 2011). The mechanism whereby the NS4 AUG is recognised for translation initiation has however not been investigated.

1.5.2 Computational predictions

Advancements in sequencing technologies permit the identification of many different isolates of viruses, and the annotation of functional elements within viral genomes. This increased availability of the amount of sequence data can in turn be used to find genomic regions under evolutionary pressure. Viruses typically have small, compressed genomes and must maximise the amount of coding information within them. Many different types of overlapping elements have been discovered within viral genes. These include RNA elements such as microRNAs, transcription factor binding sites, packaging signals, RNA secondary structures, RNA editing sites, and overlapping reading frames that encode proteins (SIEGRIST *et al.* 1993; STEWARD *et al.* 1993; GRUNDHOFF AND SULLIVAN 2011; KIM *et al.* 2011; SEALFON *et al.* 2015).

The recently developed framework FRESCo (Finding Regions of Excess Synonymous Constraint) uses a high resolution principled statistical model comparison to pinpoint synonymous constraint elements (SCEs) in the genomes of diverse viruses (SEALFON *et al.* 2015). By systematically scanning genomic coding sequences for regions of excess synonymous constraint, it is possible to identify overlapping functional elements (Fig. 1.1). Synonymous mutations within an ORF are typically evolutionary neutral since they do not change the amino acid sequence. However, in the presence of overlapping open reading frames or other functional elements, synonymous substitutions would disrupt the additional element and be disadvantageous. So, by scanning for these sections of constraint based on rates of synonymous substitutions; a decreased substitution rate suggests a selective pressure inhibiting mutation. As seen in Fig. 1.1 A, a synonymous substitution which does not affect one ORF, will affect the second ORF in the +1 frame relative to

the first ORF, causing an amino acid change. Thus, synonymous constraint in this region ensures both proteins remain correct for translation. Fig. 1.1 B shows how synonymous constraint can also occur due to important RNA elements, e.g. transcription factor binding sites. Lastly Fig. 1.1 C shows

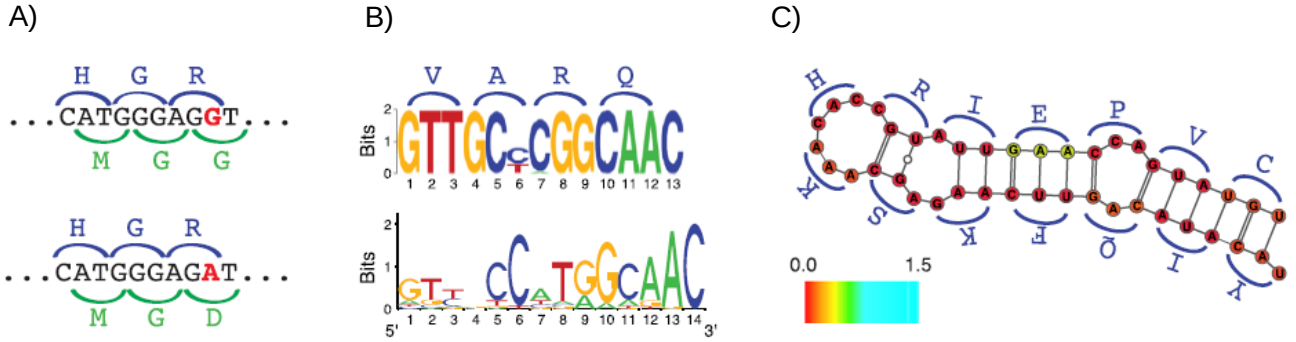


Fig. 1.1: **Examples of synonymous constraint regions and their importance**. **A)** This Hepatitis B virus (HBV) sequence fragment shows overlaps of the HBV polymerase and the HbsAg genes. The mutation in red illustrates a synonymous mutation in the polymerase protein but a nonsynonymous in the overlapping HbsAg protein. **B)** This region encodes a HBV polymerase protein section which overlaps with the transcription factor RFX1 binding site. **C)** The poliovirus genome contains an element within the ORF that has a highly conserved secondary structure. Adapted from SEALFON *et al.* (2015).

how synonymous substitution would affect secondary RNA structure, being disadvantageous. In conclusion synonymous substitution constraint suggests the presence of either another protein-coding region, or the presence of important RNA elements or sequences that are important for formation of RNA secondary structures.

By purely looking at synonymous site conservation, the differences between overlapping coding sequences and overlapping non-coding functional elements like RNA structural elements cannot be distinguished (FIRTH 2014). FIRTH (2014) suggested looking for conserved potential translation mechanisms such as sequence motifs associated with ribosomal frameshifting, or by using RNA folding software such as pknotsRG or ViennaRNA to observe stable RNA structures. This strategy was also adopted in the FRESCo analysis, where RNA structural elements, splicing sites and RNA editing sites were identified in various viruses. These RNA structural elements were found by examining the SCEs for evidence of conserved, stable RNA structures using the RNAz software. As these are only computational predictions of RNA structural elements within SCEs, they will still require experimental validation (SEALFON *et al.* 2015). Overlapping genes are also enriched in high-degeneracy amino acids, these are amino acids translated from many different nucleotide codon combinations (PAVESI *et al.* 2018). This high degeneracy could be present to relieve some of the evolutionary pressure on the overlapping regions.

FRESCo was able to identify known overlapping functional elements very accurately in viral genes. The analyses included viruses with double- and single-stranded DNA and RNA genomes, plus and minus sense RNA genomes, segmented and unsegmented genomes, and plant, insect, and animal hosts (SEALFON *et al.* 2015). In dsRNA viruses SEALFON *et al.* (2015) identified likely packaging

signals in rotavirus, and known overlapping genes in rotavirus and BTV. For BTV the known overlapping NS4 gene was recovered as a strong region of internal synonymous constraint in the +1 reading frame of the Seg-9 VP6 gene (FIRTH 2008). Additionally, FRESCo analysis revealed a strong signal of internal synonymous constraint on BTV Seg-10. This region corresponded to an alternative 50 – 59 codon ORF on Seg-10 in the +1-position relative to the NS3 ORF. However, as this region contained many nonsynonymous substitutions and few synonymous substitutions which is uncharacteristic for a protein-coding gene, but also has a weak signal for RNA structure, it could also possibly encode an RNA structural element (SEALFON *et al.* 2015).

Other bioinformatic tools also exist for identification of overlapping open reading frames with high accuracy. OLGenie utilises a measure of natural selection that is specific to protein coding genes, making it possible to directly compare functional constraint between ORFs and non-ORFs and achieves a false-positive rate of 0% (NELSON *et al.* 2020). FINKEL *et al.* (2018) utilised sequence predictions together with ribosome profiling in DNA viruses to demonstrate functionality of short open reading frames (sORFs). They found that viral sORFs are enriched for specific functional features, suggesting that some of these translation products may act at the protein level. cRegions is a new web-based tool used for identifying functional elements like splice sites, stem-loops, overlapping reading frames, internal promoters and/or ribosome frameshifting signals (PUUSTUSMAA AND ABROI 2019). Datamonkey is a web-server for analysing evolutionary signatures in sequence data (WEAVER *et al.* 2018). One such analysis that forms part of Datamonkey is the Single likelihood ancestor counting (SLAC). SLAC analysis involves a likelihood-based branch lengths, nucleotide and codon substitution.

Many other computational methods exist to characterise regions under evolutionary pressures that might indicate an important RNA structural element or overlapping open reading frame. For RNA structures these include tools such as RNA 3D structures predictions directly from sequences. Various programmes are available for these types of analyses, e.g. RNAComposer, Mfold, 3dRNA and SimRNAweb (ZUKER 2003b; ANTCZAK *et al.* 2016; MAGNUS *et al.* 2016). RNA-RNA interaction (necessary for segment recognition) prediction tools are also widely available. These include Infernal (INFERence of RNA Alignment) and CMfinder that are based on sequence and RNA secondary structure conservation (YAO *et al.* 2006; NAWROCKI AND EDDY 2013). For protein analysis various databases exist which contain information on all known proteins and their secondary structures and/or sequences. These include Uniprot, pfam, SMART and InterPro. These can be used to determine homology between proteins based on either sequence or secondary structure or both. Furthermore, protein prediction tools have also been developed to characterise unknown proteins and their structures such as Jpred, IntFOLD, RaptorX, Biskit and many others (GRÜNBERG *et al.* 2007; DROZDETSKIY *et al.* 2015; WANG *et al.* 2016; MCGUFFIN *et al.* 2019). Recently a new protein structure prediction tool AlphaFold was developed, which incorporates physical and biological knowledge about protein structures. AlphaFold is a novel machine learning approach, which utilised

multi-sequence alignments in the design of a deep learning algorithm (JUMPER *et al.* 2021). AlphaFold has offered great advances for computational predictions, and produces highly accurate protein structure predictions (VARADI *et al.* 2022). By using this machine learning-based approach during the modelling of protein-ligand interactions, it can also be used for better drug discovery (WONG *et al.* 2022).

1.5.3 ORF2 in Seg-10 of BTV and AHSV

STEWART *et al.* (2015) further investigated the alternative ORF in BTV Seg-10 that overlaps with the NS3 ORF in the +1 frame, and named it S10-ORF2. The putative protein encoded from it (50–59 residues) seems to be under strong positive selection. The function of S10-ORF2 is unknown, but it was shown to have nucleolar localisation when transiently expressed from a plasmid in transfected cells as seen in Fig. 1.2. The protein also contains a nucleolar localisation signal (NoLS) which explains its localisation pattern, when the NoLS was mutated it inhibited S10-ORF2 localising to the nucleolus. Because of this localisation, it was proposed that maybe S10-ORF2 associates with BTV NS4, which also localises to the nucleolus (BELHOUCHET *et al.* 2011; RATINIER *et al.* 2011). Through transient transfection reporter assays it was observed that S10-ORF2 inhibited gene expression, but had no effect on translation. No expression of S10-ORF2 could however be detected in BTV-infected cells, using antiserum raised against a recombinant S10-ORF2. Viruses engineered to be S10-ORF2

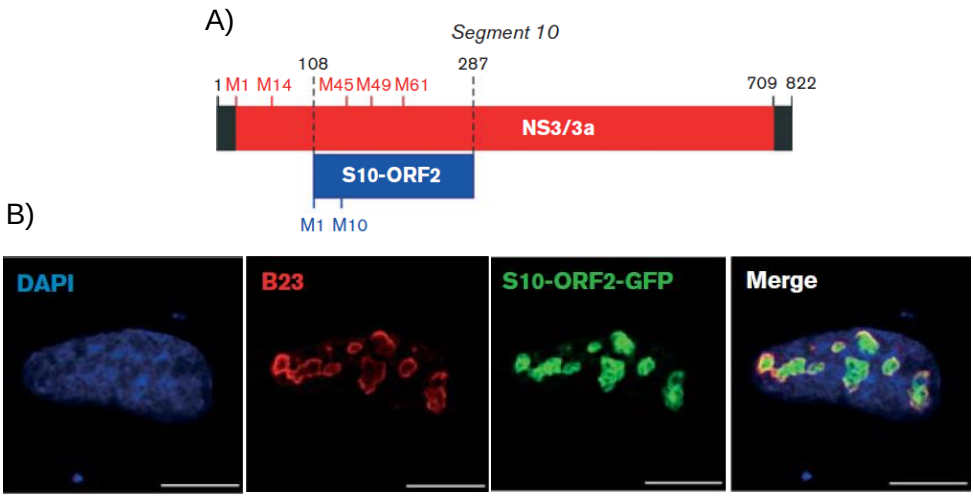


Fig. 1.2: BTV S10-ORF2. A) Representation of the internal overlapping open reading frame found in BTV NS3. **B)** Confocal microscopy images of transfected cells with GFP tagged S10-ORF2 (green). Nuclei were stained with DAPI (blue). Antibodies against the nucleolar protein B23 (red). Adapted from (STEWART et al. 2015).

deletion mutants showed similar replication kinetics to wild-type BTV. When mice were infected with these mutant virus strains, viruses were still pathogenic and caused disease. As a result of this, it is still not clear whether BTV S10-ORF2 plays a role in viral replication or pathogenicity.

In AHSV, Seg-10 ORF2 is similarly maintained in over 200 sequences of AHSV. Because this ORF is conserved in different phylogenetically related viruses, it can be an indication of possibly being evolutionary advantageous for some orbiviruses (STEWART *et al.* 2015). This emphasizes the importance of identifying new proteins from different open reading frames in virology studies. In a recent study a similar method of looking at enhanced conservation at synonymous sites was used to identify a new protein, also encoded from and additional open reading frame, in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (FIRTH 2020). By identifying new functional elements/proteins from additional open reading frames, the number of targets for drug and vaccine design potentially increases (SEALFON *et al.* 2015).

In conclusion, the focus of this research will be further investigations into AHSV Seg-10 ORF2. By utilising different bioinformatic tools combined with expression analyses a better hypothesis of the RNA structure, function and/or protein product encoded by AHSV Seg-10 ORF2 can be formulated.

1.6 Baculovirus expression system

The baculovirus expression vector system is a common eukaryotic gene expression system used to produce a large quantity of proteins of scientific interest, for example biochemical and structural studies. It contributes to the industrial manufacturing of human and veterinary vaccines globally (VAN OERS *et al.* 2015). SMITH *et al.* (1983) first described the production of recombinant proteins inside insect cells by use of the baculovirus *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV). PENNOCK *et al.* (1984) expressed β-galactosidase by use of this baculoviral vector. This was the origin of baculovirus expression vector systems, which expanded over the next 30 years.

Baculoviruses are big dsDNA viruses capable of infecting insect cells. They contain strong promoters used to produce a lot of a specific protein while inhibiting host cell protein production (MEHALKO AND ESPOSITO 2016). Bac-to-Bac is a transposition technology for baculovirus production. It was first described by LUCKOW *et al.* (1993) in *E. coli*. This is one of the most efficient approaches for constructing recombinant baculoviruses, as it is possible to produce completely pure recombinant bacmids without the need for trans-complementation or homologous recombination (MEHALKO AND ESPOSITO 2016).

A donor plasmid is produced which has a strong viral promoter and gene of interest inserted by use of recognition sites. This is then in turn transposed into a bacmid, a large single-copy DNA molecule. This bacmid contains the entire *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV) genome with an origin of replication to ensure replication in bacteria. Transposition of the gene of interest from the donor plasmid into the bacmid allows for selection and screening of recombinant bacmids in bacteria. The purified bacmids are then transfected into insect cells, for the production of recombinant baculoviruses. The insect cells produce large amount of the recombinant protein expressed from the gene of interest, and are able to achieve post-translational modifications like

phosphorylation and glycosylation (JARVIS 2003; HARRISON AND JARVIS 2006). The expressed proteins thus have the same activities as the authentic protein. By adding fluorescent markers like a GFP tag, the location of the protein inside the insect cell can be observed.

Baculoviral expression systems is and have been widely used in research, these systems are used in e.g. vaccine production and to help determine protein functions, interactions and localisations. Baculovirus expression has been used extensively for the expression and characterisation of BTV and AHSV proteins. For example, the baculovirus expressed VP2 and VP5 used for vaccination of horses (ROY *et al.* 1996). MAREE *et al.* (1998b) observed through baculovirus co-expression of VP3 and VP7 that together they form core like particles (CLP). The intracellular distribution and trafficking of AHSV VP7 was observed by measuring of VP7-eGFP fusion protein fluorescence at different time points (BEKKER *et al.* 2014). AHSV NS3 expressed from recombinant baculovirus associated with the plasma membrane, suggesting its function in viral release (STOLTZ *et al.* 1996). Recently, baculovirus expression of AHSV NS4 confirmed that the intracellular localisation of NS4 in Sf9 insect cells represented the native distribution of the protein in AHSV-infected mammalian cells (BOUGHAN *et al.* 2020). The baculoviral expression system will therefore be an ideal system for investigating the expression and intracellular localisation of the AHSV Seg-10 ORF2 viral protein.

1.7 Concluding remarks

Advancements in sequencing technologies permit the identification of many different isolates of viruses and the annotation of functional elements within viral genomes. This increased availability of the amount of sequence data can be used to find genomic regions under evolutionary pressure. Viruses have small genomes and must maximise the amount of coding information within them. In AHSV, an additional overlapping +1 ORF is found in Seg-10. This ORF is conserved in all AHSV and BTV Seg-10 genes, as well in other phylogenetically related viruses, indicating a possible evolutionary advantage for some orbiviruses. However, its specific function in orbivirus biology remains to be elucidated. While identifications of overlapping coding sequences using bioinformatics can provide computational predictions of RNA structural elements or possible protein encoding regions, experimental proof is required of whether these potential ORFs are actually translated. A combination of bioinformatic analysis and experimental expression of the AHSV Seg-10 ORF2 will allow characterisation of its expression and subcellular localisation patterns, which will in turn give an indication of its function. By identifying new functional elements/proteins from additional open reading frames, the number of targets for drug and vaccine design potentially increases. The hypothesis which forms the basis of this study is that the additional ORF in AHSV Seg-10 may encode a novel protein important during AHSV infection, or form important secondary RNA structures which contribute to advancing the AHSV replication cycle.

1.8 Aim and Objectives

The aim of this study was to fully characterise the additional open reading frame (ORF2) present in AHSV Seg-10, and to identify whether it functions either as an RNA structural element or by encoding a protein product expressed during virus infection with a role in the AHSV replication cycle. To address this aim, the following objectives were identified:

- To investigate the phylogeny, variants, conservation, and selection pressures working on AHSV Seg-10 ORF2, and relate those to its possible relevance as an RNA element or potential protein coding sequence.
- To express the protein encoded by AHSV Seg-10 ORF2 in insect cells, mammalian cells and bacteria to better understand its effect on different host cells and utilise it as an antigen source to produce an antibody against ORF2.
- To investigate the subcellular localisation and trafficking of the ORF2 protein in insect and mammalian cells in order to deduce its possible function.
- To explore whether the ORF2 protein is expressed during AHSV infection and identify where it functions in the AHSV replication cycle.

CHAPTER 2

BIOINFORMATIC ANALYSIS OF EVOLUTIONARY PRESSURES, RNA AND PREDICTED PROTEIN STRUCTURES FOR ORF2

2.1 Introduction

Viruses do not encode their own ribosomes and are entirely dependent on host translational machinery. Normal canonical eukaryotic translation synthesises one protein from one mRNA. Moreover, viruses typically have small genomes and must maximise the amount of coding information within them. Thus, they have evolved non-canonical translation initiation methods to optimise translation of multiple proteins from a single mRNA. In the *Reoviridae* family, several viruses have been characterised as having one or two genome segments which are polycistronic, containing more than one open reading frame (ORF) per mRNA.

Historically all genome segments of viruses in the genus *Orbivirus,* including AHSV and BTV, were believed to be monocistronic. However, recently novel orbivirus proteins have been identified which result from translation from additional open reading frames. BTV shares a lot of structural as well as functional resemblances with AHSV (ATTOUI *et al.* 2009). In both AHSV and BTV, Segment 9 (Seg-9) encodes major protein VP6. Seg-9 contains an additional small ORF in the +1-reading frame (FIRTH AND ATKINS 2008; SEALFON *et al.* 2015). The product of this ORF of Seg-9 was characterised as NS4 in both BTV and AHSV (FIRTH 2008; BELHOUCHET *et al.* 2011; RATINIER *et al.* 2011).

The recently developed framework FRESCo (Finding Regions of Excess Synonymous Constraint) was able to identify known overlapping functional elements in viral genes, such as BTV NS4 on Seg-9. FRESCo also revealed a strong signal of internal synonymous constraint on BTV Seg-10 which is known to encode NS3. This region corresponds to an alternative 50-59 codon ORF in the +1-position relative to the NS3 ORF. STEWART *et al.* (2015) characterised the protein encoded by the second ORF from Seg-10 in BTV, which we will refer to as BTV ORF2. In AHSV, Seg-10 ORF2 (AHSV ORF2) was similarly conserved across all serotypes and sequences investigated thus far. However, as is the case for other orbiviruses, its degree of conservation, size, RNA and putative protein structure, and possible selection pressures, have not yet been determined.

The aim of this chapter was to do a comprehensive sequence analysis of AHSV Seg-10 and ORF2. We focussed specifically on selection pressures present on Seg-10, the conservation and characterisation of the putative ORF2 protein as well as RNA structure analysis.

2.2 Material and methods

2.2.1 Sequence filtering

A total of 417 AHSV and 878 BTV Seg-10 nucleotide sequences were bulk downloaded from Genbank (18 January 2022) (BENSON *et al.* 2013). The BTV S10 ORF2 sequence (MN495931.1) (STEWART *et al.* 2015) was also downloaded from Genbank (19 May 2022). The Biopython package (COCK 2009) in Python (VAN ROSSUM 2009; THOMAS KLUYVER 2016) was used to filter downloaded sequences by removing partial or incomplete sequences. Geneious Prime® 2021.2.2 was used for initial MUSCLE alignment (EDGAR 2004) of AHSV and BTV sequences. The “Find ORF’s” function

was used to find open reading frames in respective AHSV and BTV alignments. All translations were done in CLC Main Workbench 21 using the universal code. An algorithm was written in Python to remove any identical sequences and keeping only one of each, *i.e.* only unique sequences. Geneious Prime® 2021.2.2 was also used to determine the pairwise percent identity of all alignments as well as the amino acid percent similarity score for protein alignments. SnapGene® software (from Insightful Science; available at snapgene.com) was used for visualisations. The most prevalent Seg-10 sequences representing each of the ORF2 size variants were extracted. These were designated representatives of each most common ORF2 sequence and used for certain analyses.

2.2.2 Conservation and phylogeny

CLC Main Workbench 21 was used to produce all further sequence alignments. The alignments produce a conservation percentage at each site. The average of these percentages was calculated to obtain overall conservation. Maximum likelihood phylogenetic trees were produced using RAxML 8 (STAMATAKIS 2014) in Geneious Prime® 2021.2.2. Bootstrap value was equal to 500. Phylogenies were annotated according to serotype by using the description of each sequence.

2.2.3 Selection pressure analysis

FRESCo code was obtained from the supplementary data provided in SEALFON *et al.* (2015) and was run using HYPHY 2.5.1(MPI) for Linux on x86_64 using a 10-codon sliding window. A nucleotide sequence alignment and RAxML best tree was used as input for FRESCo. A substitution model selection was performed in Datamonkey (WEAVER *et al.* 2018). The mean number of non-synonymous and synonymous substitutions per site for nucleotide sequence alignments (AHSV NS3 and ORF2, BTV NS3 and ORF2) were determined using the SLAC algorithm (KOSAKOVSKY POND AND FROST 2005). OLGenie code was obtained from https://github.com/chasewnelson/OLGenie and run with Perl 5.32.0 for the NS3 nucleotide sequence alignment with specified frame of ORF2 (ss12).

2.2.4 RNA analysis

Rfam was used to find matching RNA families or secondary RNA structures within the six representative Seg-10 RNA sequences. The Mfold web server (ZUKER 2003a) was used to predict secondary RNA structures of these Seg-10 RNA sequences. Internal ribosome entry sites (IRESs) were predicted by IRESpy (WANG AND GRIBSKOV 2019) and IRESbase (ZHAO *et al.* 2020).

2.2.5 Protein analysis

Representative sequences of AHSV ORF2 proteins and BTV ORF2 (MN495931.1) were used as input for homology searches in Uniprot, pfam, SMART and InterPr. Jpred (DROZDETSKIY *et al.* 2015) was used for initial protein structure prediction. Alphafold (JUMPER *et al.* 2021) was run using Google Colab (ColabFold) (MIRDITA *et al.* 2022). Nuclear localisation signals (NLSs) were predicted with NucPred (BRAMEIER *et al.* 2007) and NLStradamus (NGUYEN BA *et al.* 2009).

2.3 Results

2.3.1 ORF2 is present in an area of high selective pressure on AHSV Seg-10

In order to comprehensively analyse all available AHSV Seg-10 data, 417 AHSV Seg-10 nucleotide sequences were bulk-downloaded from Genbank. Sequences from other AHSV segments that were wrongfully present in the bulk-download were removed by looking at the sequence description as well as sequence lengths. Also, sequences shorter than NS3A (621 bp) were deemed partial or incomplete and were also removed. After filtering, 387 full-length AHSV Seg-10 sequences remained. A large number of the sequences represented samples isolated from AHSV positive field samples submitted from differents parts of South Africa to the Faculty of Veterinary Science, University of Pretoria (QUAN *et al.* 2008). Other sequences represented field isolates from South Africa, Namibia, Japan, and parts of the USA. The Office International des Epizooties (OIE) AHSV reference strains, attenuated vaccine viruses and laboratory strains were also included in the data. The sequences were grouped according to serotype (Table 2.1). All serotypes were represented in the 387 downloaded sequences, with the majority of Seg-10 sequences coming from serotypes 1 and 7. Serotype 2 sequences were the least prevalent. Overall there were 18 to 98 isolates per serotype.

Table 2.1: Number of AHSV Seg-10 sequences per serotype

Serotype	Nr of sequences N=387
AHSV-1	98 (25.3%)
AHSV-2	18 (4.7%)
AHSV-3	23 (5.9%)
AHSV-4	37 (9.6%)
AHSV-5	41 (10.6%)
AHSV-6	32 (8.3%)
AHSV-7	79 (20.4%)
AHSV-8	26 (6.7%)
AHSV-9	33 (8.5%)

Seg-10 sequences were aligned and searched for all possible ORFs in the sense strand. ORF's were additionally filtered for 50 amino acids and up to find the already predicted ORF2 found by STEWART *et al.* (2015). All Seg-10 sequences contained a major ORF, with a start codon at position 20 and a second in-frame methionine codon at position 53. These were annotated as NS3 and NS3A respectively. There was only one additional ORF present in all Seg-10 sequences in the +1 reading frame relative to NS3. This ORF started at position 60 of Seg-10, and position 40 relative to the NS3 start site, across all sequences. This was designated as ORF2. The size of ORF2 varied between

60 and 83 codons across different Seg-10 sequences. An example of a final annotated Seg-10 sequence is shown in Fig. 2.1.

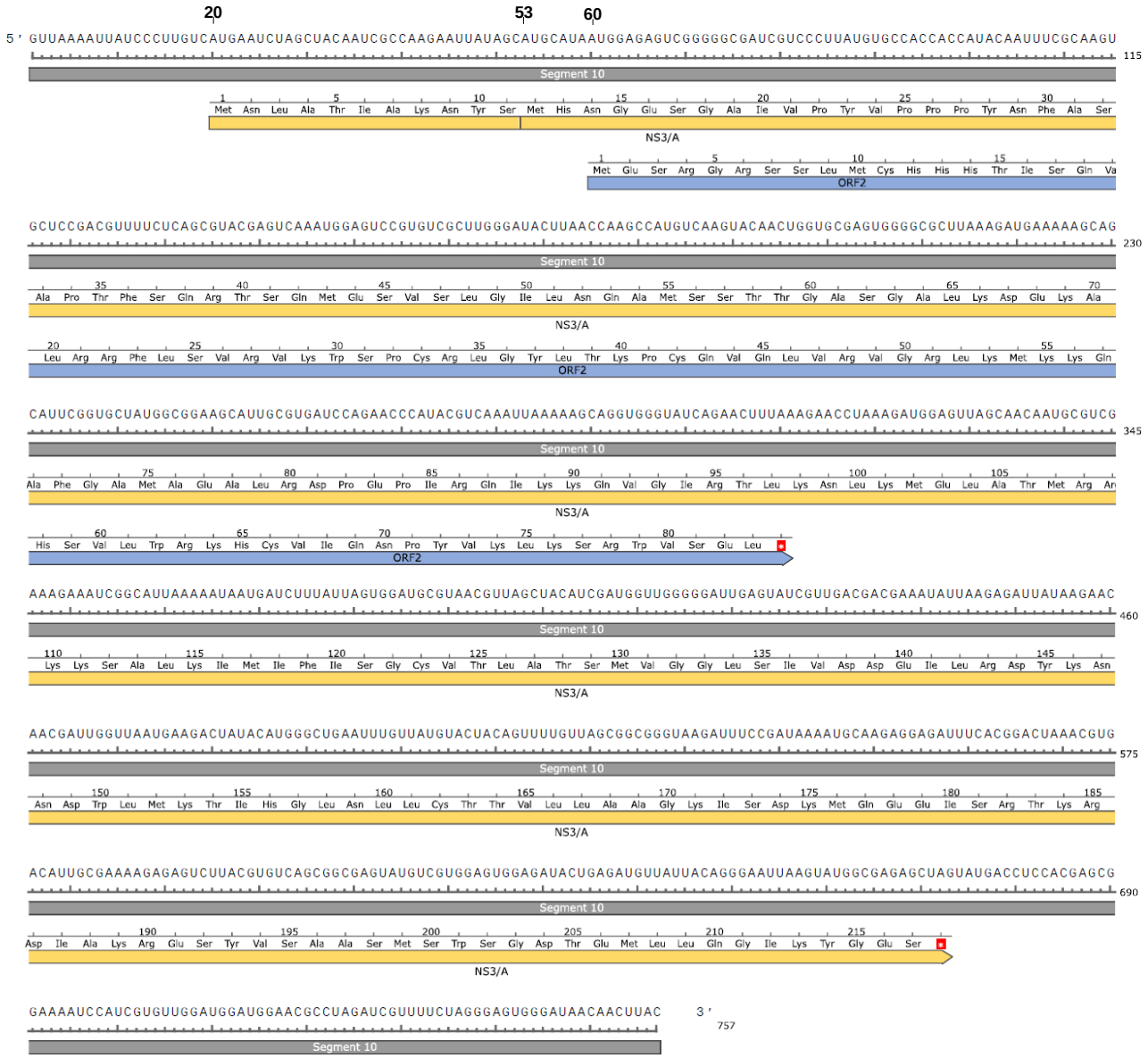


Fig. 2.1: AHSV-5 Seg-10 RNA sequence with NS3/A and ORF2 annotated. The Seg-10 nucleotide sequence is indicated in grey; NS3/A is shown in yellow with the line representing the in-frame start codon for NS3A; ORF2 is shown in blue. Amino acids are listed as 3 letter codes. Image was produced using SnapGene.

The recently developed FRESCo model-comparison framework calculates the rates of nucleotide evolution across an alignment using a maximum likelihood model. FRESCo was the application used to identify an additional open reading frame in BTV Seg-10. They identified a strong signal of internal synonymous constraint in the BTV NS3 gene, which corresponded to a 50-59 codon ORF in the +1 reading frame that was conserved across all aligned isolates (SEALFON *et al.* 2015).

In order to determine if there are any similar synonymous constraint elements (SCEs) present in AHSV Seg-10, FRESCo analysis was performed. The 387 Seg-10 nucleotide sequences were aligned and used to construct a RAxML maximum likelihood tree. The alignment and tree were used

as input for the FRESCo algorithm. FRESCo then returned the overall synonymous substitution rate found across all sequences in the alignment, where a synonymous substitution rate below 1 indicates constraint elements. When analysing the synonymous substitution rate at a 10-codon resolution in AHSV Seg-10, regions of excess synonymous constraint or SCEs were clearly observed (Fig. 2.2). FRESCo identified three regions of excess synonymous constraint in the NS3 nucleotide alignment, with positions 38 -115 for the first region, 123-144 for the second region and 176-227 for the last region as indicated by the red bars in Fig. 2.2. These three SCEs could theoretically represent regions of codon usage in a different reading frame, or important overlapping RNA structural elements such as packaging or any other regulatory signals. The smaller second and third SCEs overlapped with the two transmembrane domains of the NS3 protein. There were no alternative ORFs present in the +1 or +2 frame at these positions, so these SCEs likely have other functions which underlie their evolutionary constraint. The first and largest SCE however overlapped with the region corresponding to the identified ORF2 position, and hence formed the focus of our further investigations.

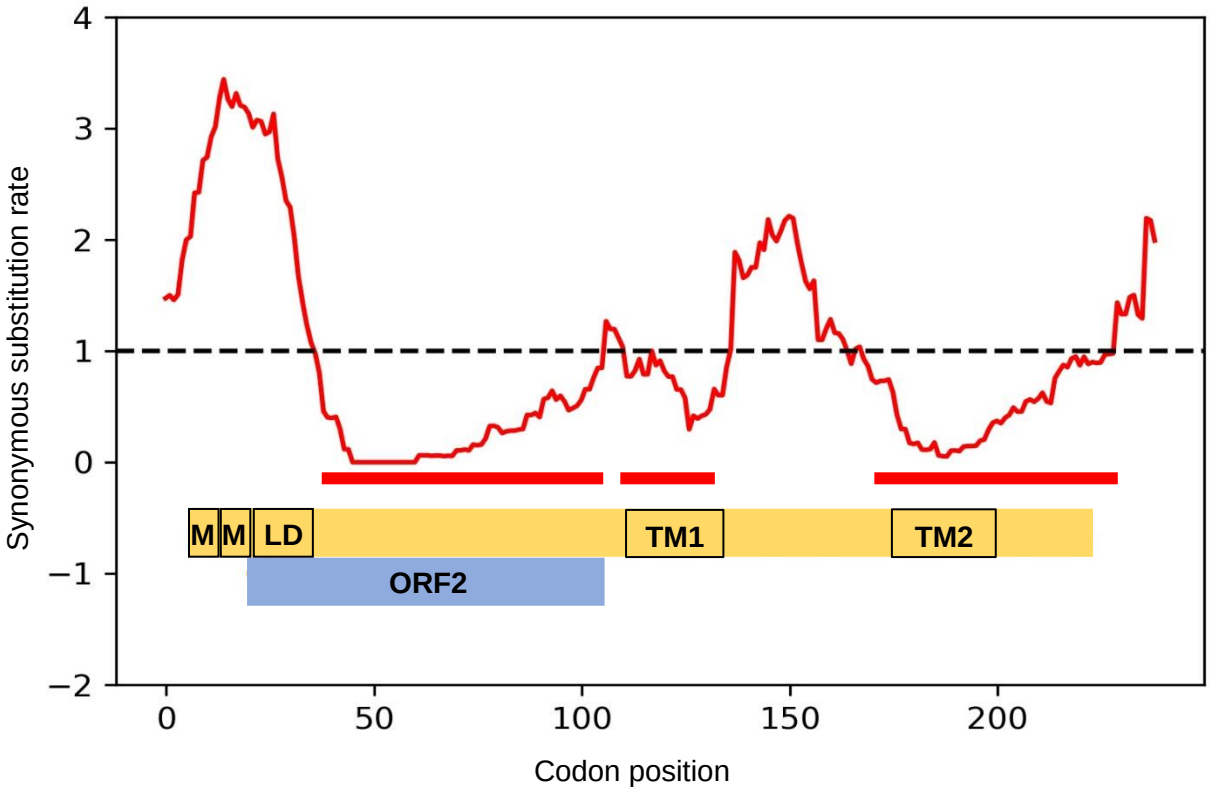


Fig. 2.2: Regions of excess synonymous constraint in the Seg-10 sequence alignment. Red bars indicate positions of synonymous constraint elements (SCEs) at 10-codon resolution, relative to NS3 (yellow) and ORF2 (blue). The NS3 bar shows the start positions (M) of NS3 and NS3/A, as well as other functional domains of NS3 namely the late domain (LD) and the two transmembrane domains (TM1 and TM2).

2.3.2 AHSV Seg-10 ORF2 has specific phylogenetic clustering and high pairwise similarity

Due to the fact that there is high evolutionary constraint in the region of ORF2, we wanted to focus specifcally on ORF2 in terms of its phylogeny in comparison to Seg-10 and NS3 and also its conservationary significance. All Seg-10 nucleotide, NS3 amino acid and ORF2 amino acid sequences were used to create maximum likelihood phylogenetic trees, and branches were labelled to indicate the predominant serotypes present (Fig. 2.3). Seg-10 nucleotide, as well as NS3 protein and ORF2 protein sequences all clustered into three distinct clades, as observed in the radial phylogenetic trees. These were previously described as alpha (α), beta (β) and gamma (γ) (SAILLEAU *et al.* 1997; MARTIN *et al.* 1998; VAN NIEKERK *et al.* 2001b; QUAN *et al.* 2008). According to the results in Fig. 2.3, certain serotypes clustered together in the Seg-10 nucleotide, NS3 and ORF2 protein trees. Serotypes 3 and 7 clustered together in the β clade, serotypes 6, 8 and 9 clustered together in the α clade and serotype 2 clustered in the γ clade. Serotype 1 clustered into both the β and γ clades and serotype 4 and 5 clustered together into both α and γ. The placement of a serotype within more than one clade could be due to segment recombination, as VP2 and not NS3 is the serotype determinant. Overall, the ORF2 protein clustered into the same three clades as the Seg-10 nucleotide and NS3 protein sequences. Lastly, we wanted to see if the ORF2 region was the reason for the specific clustering of Seg-10. The ORF2 region was removed from Seg-10 and subject to maximum likelihood tree building (results not shown). The clustering remained the same with or without ORF2, and thus the entire Seg-10 sequence is responsible for the different clade formation.

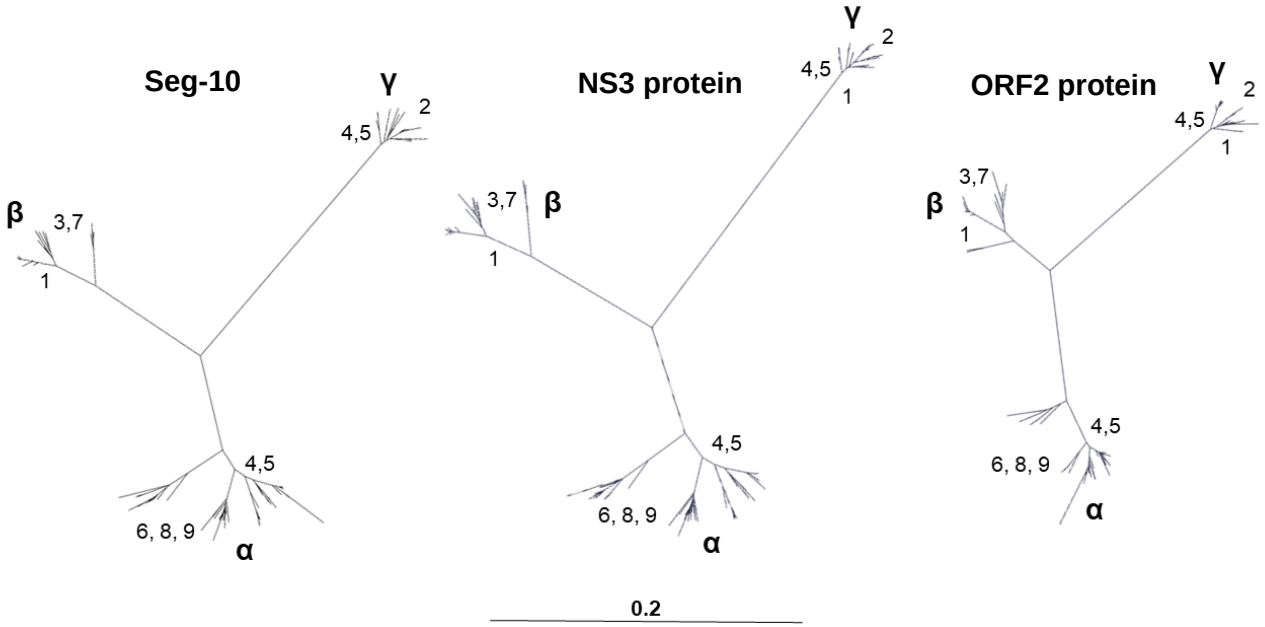


Fig. 2.3: Phylogenetic analysis of Seg-10, NS3 and ORF2 in AHSV. Unrooted maximum likelihood tree of AHSV Seg-10 (left), NS3 protein (middle) and ORF2 protein (right). Scale bar represents the phylogenetic distance of 0.2 nucleotide substitutions per site for Seg-10 and 0.2 amino acid substitutions per site for NS3 and ORF2. Bootstrap value=500.

Next we wanted to obtain a clear understanding on how the putative ORF2 protein is conserved across all AHSV Seg-10 sequences, in terms of the amino acid identity and similarity scores of its encoded protein. The 387 sequences were parsed and all duplicate sequences removed. The remaining unique Seg-10 nucleotide, NS3/A and ORF2 nucleotide and protein sequences were aligned, and their overall pairwise sequence identity, pairwise similarity and degree of conservation per site determined. Table 2.2 details the number of unique sequences found to be present amongst all 387 sequences, with pairwise sequence identities across all unique sequences indicated as a percentage. The overall pairwise sequence identity of Seg-10 ranged between 62.3% and 100% between clades, and when compared within clades it was slightly higher ranging from 64 to 100%. From a total of 387 sequences, there were 101 unique ORF2 coding region nucleotide sequences and 90 unique ORF2 protein sequences. The pairwise nucleotide sequence identity for the ORF2 coding region ranged from 54% to 100%, with protein identity scores from 43% to 100%. When considering amino acid similarity in the ORF2 pairwise amino acid sequence alignments, all alignments had a score of 82% and above between and within clades. This indicated that any amino acid substitutions between aligned sequences were mostly of structurally similar amino acids. This could be to preserve the overall functioning of the putative ORF2 protein. The NS3 pairwise similarity scores were slightly higher than those of ORF2, at over 90% within and between clades.

Table 2.2: Sequence conservation comparisons for Seg-10, NS3 and ORF2 across all AHSV strains. The number of unique sequences of AHSV Seg-10, NS3 and ORF2 are listed with their respective average and range percent pairwise identities for nucleotide and protein sequences, and percent average pairwise similarity for protein sequences.

Sequence	**Size**	**Nr of unique seq.**	**Nucleotide/amino acid pairwise identity average [range]**			
			Between clades	**Within α clade**	**Within β clade**	**Within γ clade**
Seg-10 nucleotide	755-763 nt	217	78.3% [62.3-100]	87.6% [64.2-100]	80.6% [64.8-100]	81.3% [64.5-100]
NS3 coding region	654-657 nt	183	78.5% [64.5-100]	87.3% [65.2-100]	80.5% [64.8-100]	81.0% [64.3-100]
NS3 amino acid	216-217 aa	150	80.3% [61.6-100]	89.6% [61.8-100]	80.8% [63.0-100]	81.1% [62.1-100]
ORF2 coding region	183-252 nt	101	78.9% [54.0-100]	89.7% [56.7-100]	78.8% [55.1-100]	77.2% [55.3-100]
ORF2 amino acid	60-83 aa	90	71.7% [43.0-100]	84.1%[44.6-100]	72.0% [41.5-100]	73.0% [43.0-100]

Sequence	**Amino acid pairwise similarity average**			
	Between clades	**Within α clade**	**Within β clade**	**Within γ clade**
NS3 protein	94.3%	97.1%	94.8%	93.9%
ORF2 protein	82.9%	91.4%	83.4%	84.0%

The conservation at each site was determined from aligned sequences, and plotted for Seg-10 nucleotide, NS3 protein and ORF2 protein sites (Fig. 2.4). The center of the ORF2 protein spans a region of high conservation in Seg-10. Thus, both at a nucleotide and protein sequence level, the ORF2 sequence is relatively highly conserved. The 5' and 3' UTRs of Seg-10 showed very high conservation. These regions could be important for other non-coding functions such as RNA elements or packaging signals. The NS3 protein follows roughly the same conservation pattern as Seg-10.

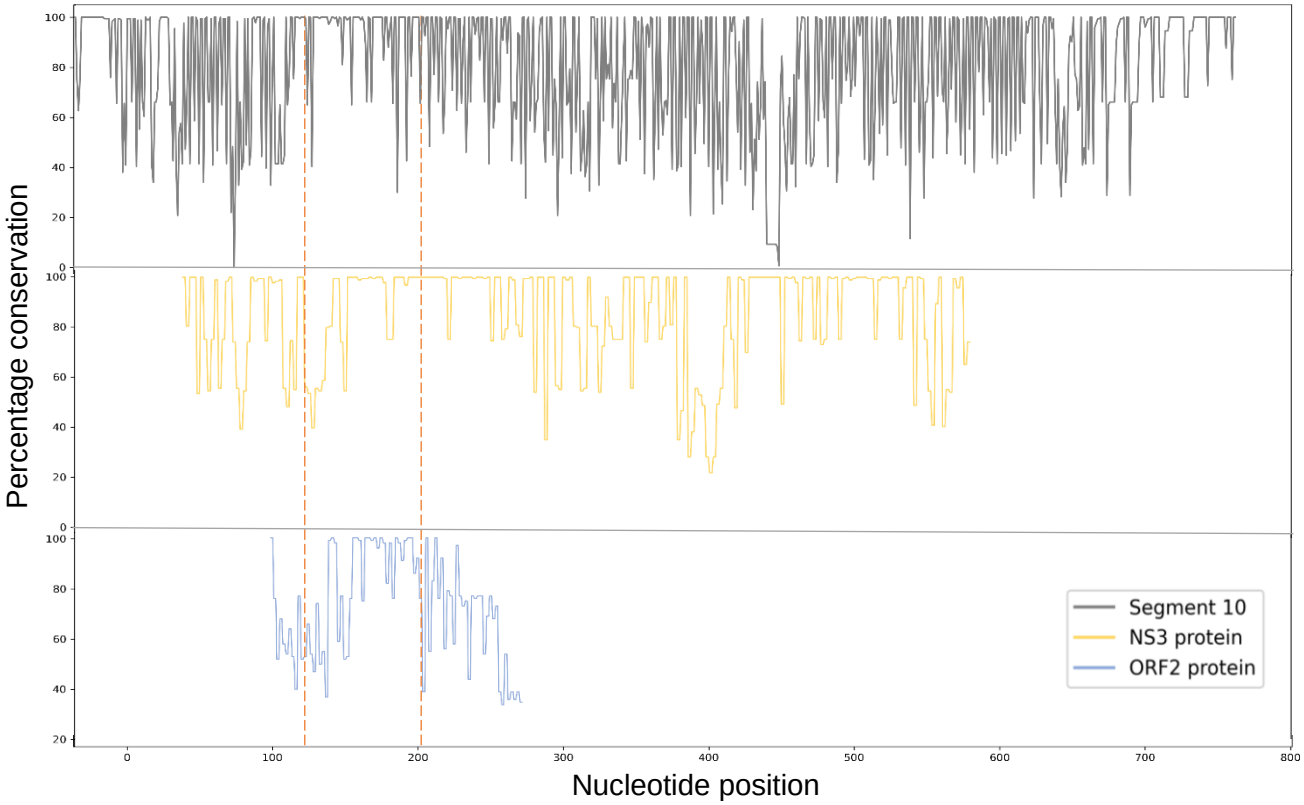


Fig. 2.4: Conservation percentage per site. The nucleotide conservation of Seg-10 (grey), and amino acid conservation of NS3 (yellow) and ORF2 (blue). Dashed lines indicate a region of high conservation.

In conclsuion, the high pairwise similarity of the ORF2 amino acid alignment as well as its high conservation in the Seg-10 nucleotide alignment could suggest a functionally important putative protein.

2.3.3 AHSV Seg-10 ORF2 protein exists as six size variants

2.3.3.1 ORF2 has high conservation across all size variants

Next, we wanted to focus specifically on ORF2 in terms of its size, and conservation within and between size variants. In the initial annotation of ORFs on Seg-10, we had already identified that ORF2 was present in all sequences, starting at nucleotide 60 and terminating at different downstream positions. In total, there were six different size variants of ORF2 with predicted sizes of 60, 64, 74, 78, 80 and 83 aa respectively. For each size category of ORF2, the unique nucleotide

and protein sequences were grouped together and aligned to determine their conservation. Table 2.3 summarizes the prevalence of each size variant in the sample, (total number and categorised based on serotype) as well as the degree of nucleotide and amino acid identity across unique sequences within that size category.

Table 2.3: Analysis of prevalence of six different size variants of ORF2. For each size variant the total sample number, the number of unique sequences, the nucleotide and amino acid conservation, and the breakdown per serotype is indicated.

ORF2 variant size	**60 aa (183 nt)**	**64 aa (195 nt)**	**74 aa (225 nt)**	**78 aa (237 nt)**	**80 aa (243 nt)**	**83 aa (252 nt)**
Nr of sequences (N=387)	10 (2.6%)	73 (18.9%)	19 (4.9%)	38 (9.8%)	88 (22.7%)	159 (41.1%)
Nr of unique sequences	2	18	4	13	16	37
ORF2 coding region pairwise identity	99.3%	97.7% [95.4-99.5]	96.9% [84.3-99.6]	96.4% [89.4-99.6]	95.2% [91.7-99.6]	96.3% [86.9-99.6]
ORF2 amino acid pairwise identity	98.3%	93.4% [87.5-98.4]	91.2% [83.7-98.6]	91.2% [78.2-98.7]	88.7% [78.7-98.8]	90.2% [68.7-98.8]
ORF2 protein similarity	100.0%	96.2%	94.6%	96.4%	95.5%	95.3%
AHSV-1 ORF2 distribution (N = 98)	5.%	44.9%	3.1%	0%	1.0%	45.9%
AHSV-2 ORF2 distribution (N = 18)	11.1%	88.9%	0%	0%	0%	0%
AHSV-3 ORF2 distribution (N = 23)	0%	0%	69.6%	0%	26.1%	4.3%
AHSV-4 ORF2 distribution (N = 37)	0%	5.4%	0%	21.6%	0%	73.0%
AHSV-5 ORF2 distribution (N = 41)	4.9%	26.8%	0%	2.4%	2.4%	63.4%
AHSV-6 ORF2 distribution (N = 32)	0%	0%	0%	9.4%	0%	90.6%
AHSV-7 ORF2 distribution (N = 79)	0%	0%	0%	0%	100%	0%
AHSV-8 ORF2 distribution (N = 26)	3.8%	0%	0%	46.2%	3.8%	46.2%
AHSV-9 ORF2 distribution (N = 33)	0%	0%	0%	42.4%	0%	57.6%

The longest size variant of 83 aa (252 nt) occurred most frequently and was present in 41% of all samples, while the shortest size variant of 60 aa (183 nt) was present the least frequently in 3% of cases. Within each size variant, there were only a few completely unique sequences present. When analyzing the pairwise identity within each size variant alignment of ORF2, the nucleotide sequences for each size class had a high similarity of 95% or higher. The protein sequences of each size class had a slightly lower pairwise identity score, but had extremely high pairwise similarity scores. These pairwise similarity scores were much higher for sequences compared within a specifc size variant as compared to the overall pairwise similarity from Table 2.2. This again infers the maintenace of similar amino acids at most sites in order to preserve the possible functioning of the ORF2 putative protein. The different ORF2 size variants showed some serotype specificity, and each serotype contained one to five different ORF2 size classes with some sizes being more prevalent than the other. For example, the 64 aa and 83 aa size variants represented over 90% of AHSV-1 sequences whereas sequences from AHSV-7 only contained the 80 aa ORF2 variant. Thus, this serotype-ORF2 size specificity could match the phylogeny, where a certain size variant clusters into one of the three lineages.

The 90 unique ORF2 protein sequences were aligned, and the alignment organised according to similarity (Fig. 2.5). Overall, all ORF2 sequences showed a high conservation from amino acid 20 to 62 (red block), with more variability in the terminal regions.

Fig. 2.5: Alignment of all unique ORF2 sequences. ORF2 sequences were aligned using CLC Main Workbench. Sequences are listed according to similarity. Conservation is shown as bars at the bottom of alignment in percentages.

Lastly, the most commonly occurring sequence of each AHSV size variant was aligned with BTV S10-ORF2 (MN495931.1) (STEWART *et al.* 2015) (Fig.2.6). Even though there was low overall homology between BTV and AHSV ORF2, a number of residues (mostly hydrophobic or charged) were identical across the serogroups (red blocks).

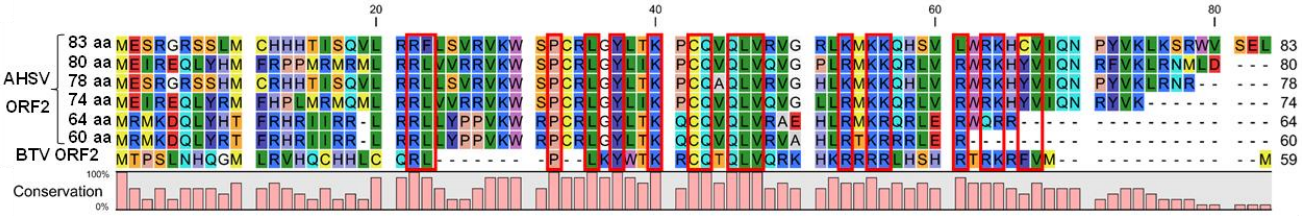


Fig. 2.6: AHSV ORF2 protein aligned with BTV ORF2. Representative AHSV ORF2 sequences of each size variant and BTV ORF2 were aligned in CLC Main workbench. Conservation is shown as bars at the bottom of alignment in percentages

2.3.3.2 NS3 and ORF2 proteins are under conflicting selection pressures

Additional to the FRESCo analysis we wanted to determine the specific selection pressures on AHSV NS3 and its putative overlapping protein ORF2, because overlapping proteins are often under conflicting selection pressures (HUGHES AND HUGHES 2005). We performed a substitution model selection using a Single-Likelihood Ancestor Counting (SLAC) algorithm to infer nonsynonymous (dN) and synonymous (dS) substitution rates on a per-site basis for a given coding alignment.

The 878 BTV Seg-10 sequences downloaded from Genbank were filtered in the same way as for AHSV. The 33 most prevalent BTV Seg-10 sequences, as well as a sequence from each of the new serotypes 25 and 26, were annotated for NS3 and ORF2. Nucleotide sequence alignments of each of BTV NS3, BTV ORF2, AHSV NS3 and ASHV ORF2 were subject to SLAC analysis. Table 2.4 shows the number and sizes of aligned sequences analysed, as well as the output with respect to dN/dS, and positively and negatively selected site values. The BTV results were consistent with what was found by STEWART *et al.* (2015) and used as a control. Overall AHSV followed the same trend as BTV. NS3 in both AHSV and BTV were identified as being under strong purifying/negative selection, with an average dN/dS ratio below one. Also, the majority of codon sites in NS3 show negative selection. This indicates that both BTV NS3 and AHSV NS3 are under constraint, which correlates with the FRESCo results. BTV ORF2 and AHSV ORF2 had more sites under positive selection than sites under negative selection. They also showed an average dN/dS ratio of above one. This is completely the opposite of what was observed in NS3.

Table 2.4: SLAC analysis summary results for non-synonymous (dN) and synonymous (dS) substitution rates for BTV and AHSV NS3 and ORF2

Protein	No. of seq. analysed	Size	dN/dS	Positively selected	Negatively selected
BTV NS3	35	230	0.042	0	165
BTV S10-ORF2	35	59	6.670	16	1
AHSV NS3	387	216-217	0.207	28	159
AHSV S10-ORF2	387	60-83	1.870	29	16

The synonymous (dS) and non-synonymous (dN) substitution rate per codon calculated by SLAC was plotted (Fig. 2.7). NS3 of both AHSV and BTV have high synonymous substitution rates (primarily orange) at each amino acid site. AHSV and BTV ORF2 had a high non-synonymous substitution rate (primarily blue) at each amino acid site. Again, the results of BTV S10-ORF2 matched those of STEWART *et al.* (2015). The results of both BTV and AHSV indicated that NS3 and ORF2 are under conflicting selection pressures, where NS3 is under negative selection and ORF2 is under positive selection.

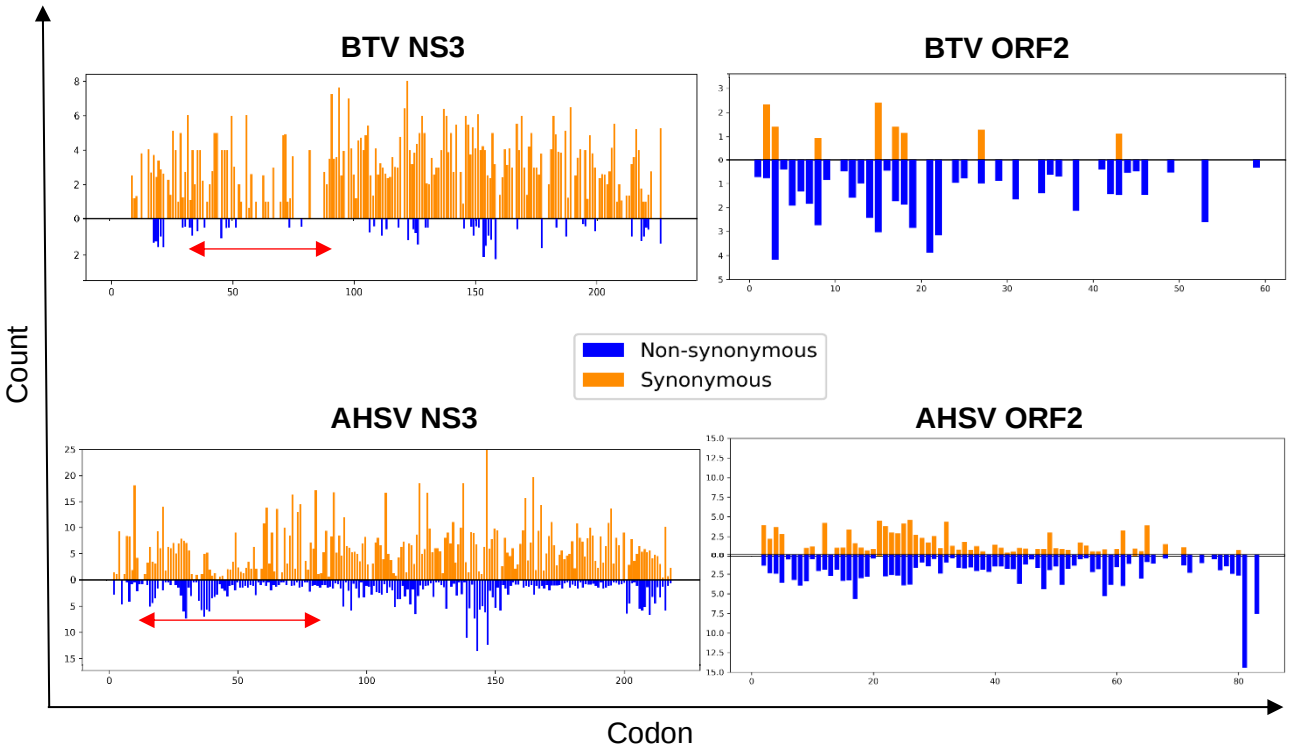


Fig. 2.7: Frequency of synonymous vs non-synonymous changes per codon in BTV and AHSV NS3 and ORF2. Observed synonymous (orange) and non-synonymous (blue) changes determined by the SLAC method for NS3 and S10-ORF2 ORFs. The red double headed red arrow represents the position of ORF2 relative to NS3.

The dN/dS rates for NS3 and ORF2 were also estimated in OLGenie, using modifications of another approach (WEI AND ZHANG 2014), which allows testing whether a specified alternate reading frame contains an overlapping gene that is in fact under selection. This method demonstrates low false-positive rates and good discriminatory ability in differentiating true overlapping genes from non-overlapping genes (NELSON *et al.* 2020). The NS3 nucleotide sequence alignment was provided with the frame of choice relative to NS3 in which the predicted ORF2 lies (+1 frame) as input. Given these values, dN/dS may be estimated for the reference gene as dNN/dSN or for the alternate gene as dNN/dNS. For example, if nonsynonymous changes observed in the reference gene are disproportionately synonymous in the alternate gene, the result will be dN/dS < 1.0, and purifying selection on the alternate gene can be inferred. The OLGenie value calculated for NS3 dN/dS was

0.67, and for the ORF2 was 2.61. These values again sustained the previous observation using SLAC analysis of conflicting selection pressure on the two ORFs on Seg-10.

2.3.3.3 ORF2 putative protein showed no homology to known proteins

This overlapping second open reading frame on Seg-10 of AHSV could theoretically encode an additional protein for expression during the AHSV replication cycle. To identify any structural protein domains in ORF2, or homology to other known proteins, the amino acid sequences of each of the most common ORF2 size variants was subjected to searches against the Uniprot, pfam, SMART and InterPro databases. No homology of ORF2 was detected to any proteins in these databases.

Next the secondary structures of the AHSV ORF2 size variants as well as BTV S10-ORF2 were predicted by Jpred (DROZDETSKIY *et al.* 2015), which uses a database searching method, alignment algorithm and scoring scheme (Fig. 2.8). Secondary structure prediction is an important tool for the analysis of proteins which have no sequence similarity to proteins of known structure. Fig. 2.8 shows the Jpred results, where all protein structures were predicted with high confidence levels throughout. All the AHSV and the BTV ORF2 proteins were predicted to have a relatively long helix in the N-terminal third of the protein, and most sequences also showed a second helix domain predicted in the C-terminal half of the protein.

Fig. 2.8: AHSV ORF2 and BTV ORF2 protein structure predictions. ORF2 sequences of each size variants and BTV ORF2 were subject to Jpred. Red indicates helices and green indicates beta sheets.

AlphaFold is currently the top-ranked method for prediction of protein secondary structures, and is a novel machine learning approach (JUMPER *et al.* 2021). The AlphaFold system uses the input amino acid sequence to query several protein sequence databases and constructs a multiple sequence alignment (MSA) for secondary structure predictions. Input sequences are compared across large

databases including MGnify, UniProt and wwPDB. AHSV S10-ORF2 variants and BTV S10-ORF2 were used as query sequences in AlphaFold. AlphaFold did not find any taxonomic data in MGnify matching the AHSV ORF2 sequences. It did however identify that the BTV sequence matched BTV NS5 (STEWART *et al.* 2015), which was its own sequence. AlphaFold did not find any homology in Uniref for either AHSV or BTV ORF2.

Table 2.5. Protein database hits for six AHSV ORF2 size variants and BTV ORF2 as determined by AlphaFold.

Name	P-value	Organism	Class
AHSV ORF2 83 aa			
Chain A and B, site-determining protein	0.00028	Bacteria	Structural
Chain A, primosomal replication protein	0.00085	Bacteria	DNA binding protein
Chain A, Myosin binding protein C, cardiac-type	0.0011	Human	Cytoskeleton
26s proteasome non-ATPase regulatory subunit 4	0.0022	Human	Proteasome
Chain A, nitrate/nitrite sensor histidine kinase narQ	0.0029	Bacteria	Transferase
Oe1.3 alkylated	0.0039	Archaeon	Hydrolase
Chain C, zinc finger domain-containing protein	0.0041	Fungi	Splicing
Chain A, alpha-(1,3)-fucosyltransferase	0.0042	Bacteria	Transferase
Chain A, ataxin-7-like protein 3	0.0043	Human	Transcription
AHSV ORF2 80 aa			
Chain B, Tni2	0.0003	Bacteria	Antimicrobial protein
Chain A, ABC transporter ATP-binding protein	0.0015	Bacteria	Hydrolase
Chain A, pyridinium-3,5-bisthiocarboxylic acid mononucleotide nickel insertion protein	0.0017	Bacteria	Metal binding protein
Chain A, U4/U6 small nuclear ribonucleoprotein PRP3	0.0017	Yeast	RNA binding protein
Chain A, polycomb protein EED	0.0022	Fungi	Transferase
Chain c, neurogenic locus notch homolog protein 1	0.0023	Metazoa	Protein binding
Chain B, HoBa	0.0029	Bacteria	DNA binding protein
AHSV ORF2 78 aa			
Chain B, Tni2	0.00023	Bacteria	Antimicrobial protein
Chain A, B and F protein SeqA	0.0005	Bacteria	Replication inhibitor
Chain E, 39s ribosomal protein l46, mitochondrial	0.00083	Metazoa	Ribosome
Chain A, Pleckstrin homology domain-containing family A member 3	0.0014	Metazoa	Membrane protein
Chain A, histidine-containing phosphotransfer protein 2	0.0015	Plant	Signalling protein
Chain A, UNC-112-related protein 2	0.0018	Human	Signalling protein
Chain D, 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2	0.0019	Human	Signalling protein
Chain D, DNA-binding protein SATB1	0.0019	Metazoa	DNA binding protein
AHSV ORF2 74 aa			
Chain B, Tni2	0.00039	Bacteria	Antimicrobial protein
Chain D, hypothetical UPF0247 protein YYDA	0.0013	Bacteria	Unknown function
Chain A, U4/U6 small nuclear ribonucleoprotein PRP3	0.0016	Yeast	RNA binding protein
Chain A, hypothetical protein YBEA	0.0016	Bacteria	Unknown function
Chain A, polycomb protein EED	0.0018	Fungi	Transferase
Chain D, 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2	0.0018	Human	Signalling protein
AHSV ORF2 64 aa			
Chain A, Attractin	0.0005	Metazoa	Attractin
Chain A, Apolipoprotein C-II	0.00054	Metazoa	Signalling protein
Chain BN, mL80	0.0012	Parasite	Ribosome
Chain B, Dynein assembly factor with WDR repeat domains 1	0.0015	Algi	Motor protein
Chain B, glycosyl hydrolase family 62 protein	0.0018	Fungi	Hydrolase
Chain A, HCF N-terminal chain 1	0.0023	Metazoa	Protein binding
Chain A ATP synthase	0.0023	Bacteria	Hydrolase
AHSV ORF2 60 aa			
Chain A, Attractin	0.00041	Metazoa	Attractin
Chain J, inward rectifier potassium channel 2	0.0012	Metazoa	Metal transport
Chain B, Tni2	0.0012	Bacteria	Antimicrobial protein
Chain B, dynein assembly factor with WDR repeat domains 1	0.0013	Algi	Motor protein
Chain A ATP synthase	0.0023	Bacteria	Hydrolase
Chain K, ATP synthase subunit K	0.0025	Algi	Membrane protein
Chain A, neurogenic locus notch homolog protein 1	0.0028	Metazoa	Receptor
Chain B, Dr hemagglutinin structural subunit	0.0029	Bacteria	Cell adhesion
BTV ORF2 59 aa			
Chain A, ribosome hibernation promoting factor	0.0025	Bacteria	Ribosomal protein
Chain A, low-density lipoprotein receptor	0.0028	Metazoa	Lipid transport
Chain A and B, Ectatomin	0.0031	Metazoa	Toxin
Chain A, eukaryotic translation initiation factor 3 subunit G	0.0044	Metazoa	Metal binding protein
Chain D, putative 3-epimerase in D-allose pathway	0.0056	Bacteria	Unkown function
Chain A, low-density lipoprotein receptor-related protein 6	0.0059	Human	Protein binding antagonist
Chain B, telomere repeats-binding bouquet formation protein 1	0.0063	Human	DNA binding protein
Chain F, Melanophilin	0.007	Mouse	Protein transport

This matched the lack of homology observed during the previous database searches. Lastly, AlphFold identified some homology secondary structures in the wwPDB database matching AHSV ORF2 or BTV ORF2 with a P-value smaller than 0.05 indicative of significant results. These are listed in Table 2.5. All homology identified was from sources other than viruses. There was also no single protein homologous in common to all size variants of AHSV ORF2, only chain b from the bacterial protein tni2 (highlighted in the table) matched four of the six ORF2 size variants.

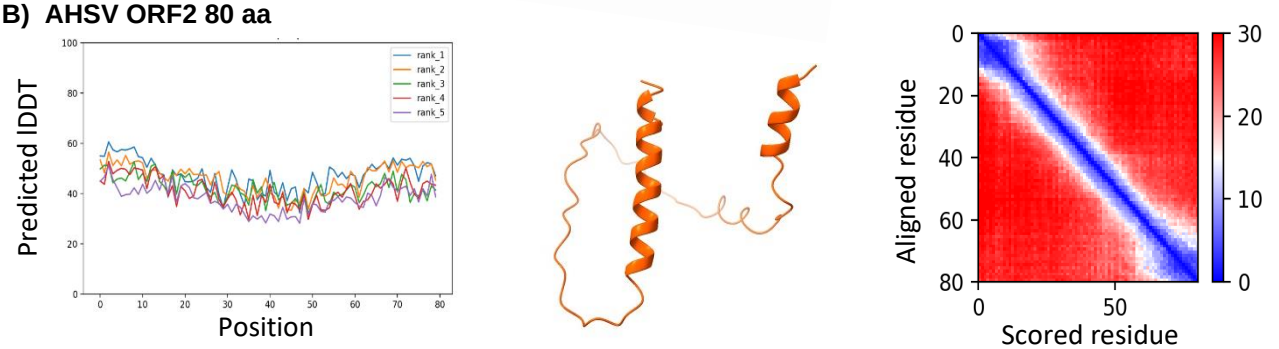
Next AlphaFold provided secondary structure predictions for the ORF2 sequences (Fig. 2.9). Alphafold makes five model predictions. The confidence measure for the structures is called pIDDT and corresponds to each model's predicted score. Regions with pIDDT > 90 are expected to be modelled to high accuracy, regions with pIDDT between 70 and 90 are expected to be modelled well and regions with pIDDT between 50 and 70 are low confidence and may not reflect accurate results. AlphaFold also produces an output called "Predicted Aligned Error" (PAE). These values report AlphaFold's expected position error at residue x, when the predicted and true structures are aligned on residue y. If the PAE is generally low for residue pairs x, y (blue) it indicates that AlphaFold predicts well-defined relative positions and orientations for them. If the PAE is generally high for residue pairs x, y (red) then the relative positions and/or orientations of these domains in the 3D structure are uncertain and should not be interpreted. For each AHSV ORF2 size variant and BTV ORF2 a plDDT plot of all five predicted models with the confidence level shown per site is shown (left), with the best predicted model's secondary structure (middle) and related PAE plot (right) (Fig. 2.9 A-G). The pIDDT levels of AHSV ORF2 83 aa, 80 aa and 78 aa variants showed low confidence with pIDDT scores below 70. This means that AlphaFold was not very confident in the protein secondary structure predicted, and that it might be a poor representation of the actual structure. The pIDDT levels of AHSV ORF2 74 aa, 64 aa, 60 aa and BTV ORF2 showed slightly higher confidence levels ranging from low 50's to around 80. This is well modelled and may represent actual secondary structures.

When considering the secondary structures themselves, all ORF2 sequences contained two alpha helices with short to longer linear regions. The 83 aa and 80 aa size variants had a similar structure showing two alpha helices. The other four size variants showed very similar linear structures with two long alpha helices near the terminal ends. BTV ORF2 was predicted to form a single C-terminal alpha helix. All PAE plots showed a similar distribution with overall high PAE (red) indicative of uncertain relative positions of protein domains (TUNYASUVUNAKOOL *et al.* 2021). The alpha helices of the 83 aa, 80 aa and 78 aa variants were predicted with low confidence and moderate confidence

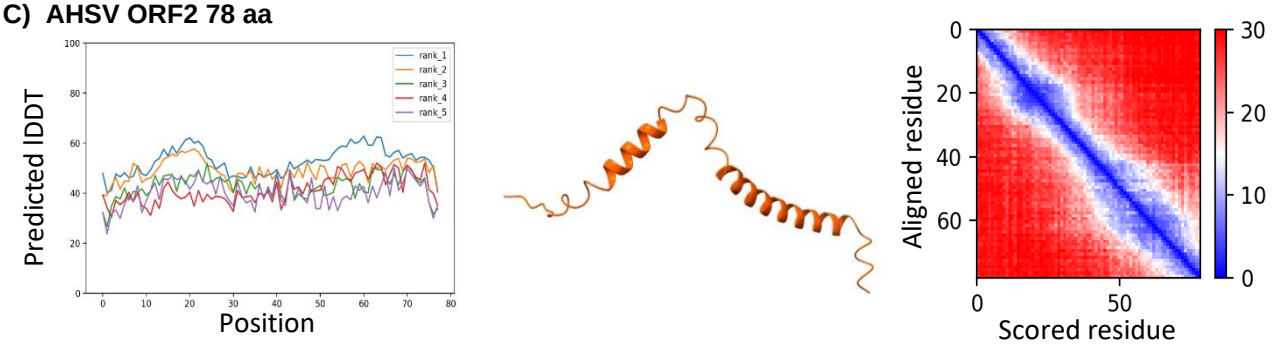
for the 74 aa, 64 aa and 60 aa AHSV ORF2 variants. The linear region of the BTV ORF2 sequence was predicted with high confidence but lower confidence for the alpha helix.

A) AHSV ORF2 83 aa

B) AHSV ORF2 80 aa

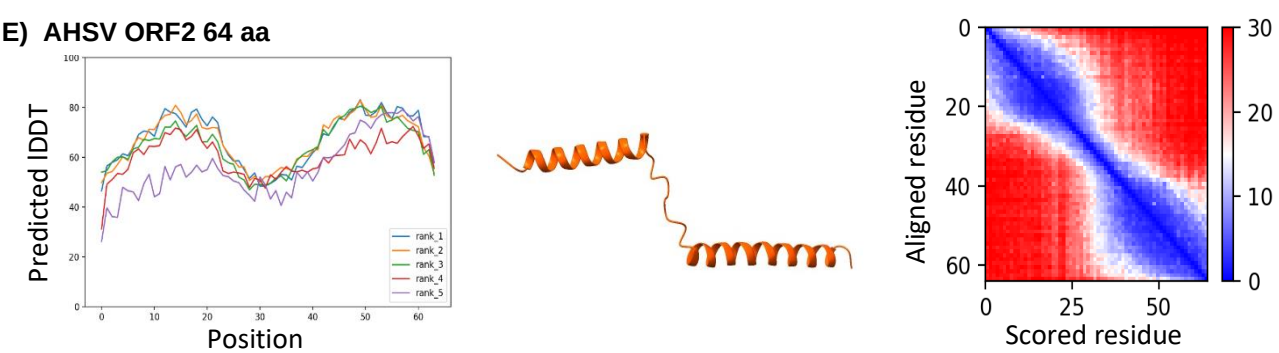


C) AHSV ORF2 78 aa

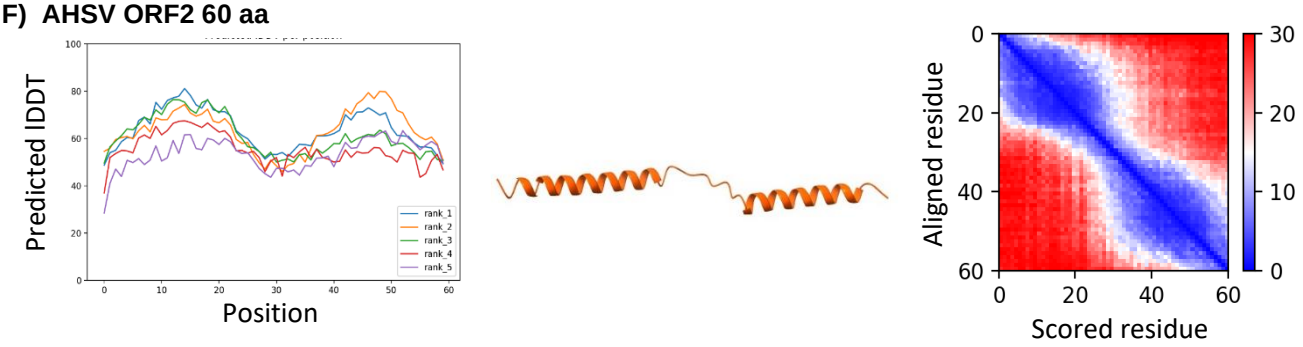


D) AHSV ORF2 74 aa

E) AHSV ORF2 64 aa



F) AHSV ORF2 60 aa



G) BTV ORF2

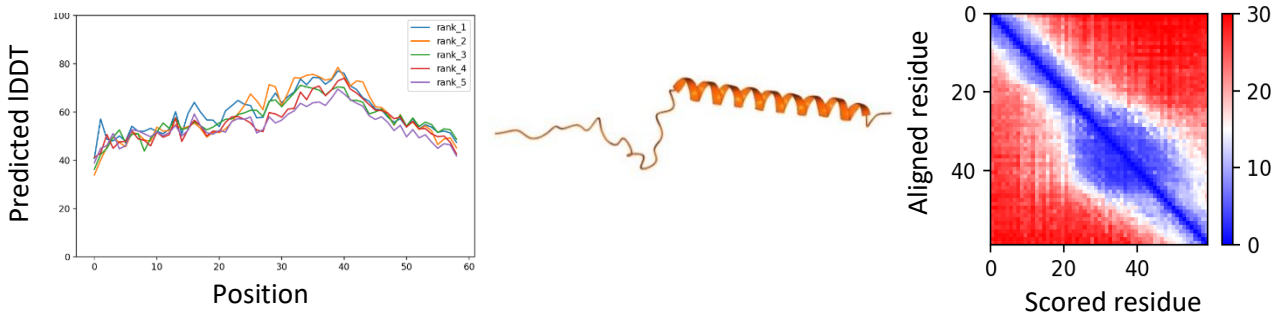


Fig. 2.9: AHSV ORF2 (A-F) and BTV ORF2 (G) AlphaFold protein structure predictions. Representative ORF2 protein sequences were subject to Alphafold. Each prediction has a **p**lot representing the confidence measure (pIDDT) corresponding to the five model's predicted score (rank 1-5) on the lDDT-Cα metric. The secondary structure of the best model is shown in orange with N-terminal on the left and C-terminal on the right. The PAE heat map shows the predicted error (in Angstroms) between all pairs of residues.

Lastly, protein sequences were also analysed by other signal detection software. ORF2 protein sequences were subject to NucPred and NLStradamus for nuclear localisation signal detection. These two programs did not identify any nuclear localisation signals in any of the AHSV ORF2 size variants, but did predict the known NoLS in BTV-ORF2 (STEWART *et al.* 2015) (Fig. 2.10). The smallest ORF2 size variant did show a nuclear localisation signal that was just below the threshold probability of 0.8 and therefore not deemed significant.

In conclusion, AHSV ORF2 showed no homology to any other known sequences and did not produce an overwhelmingly confident protein structure. Thus, it cannot be predicted with a high level of

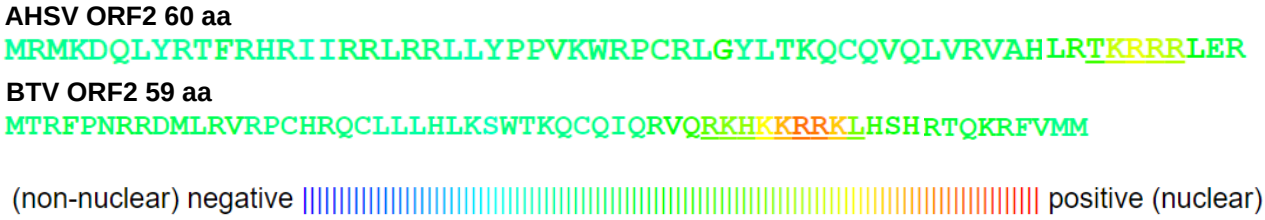
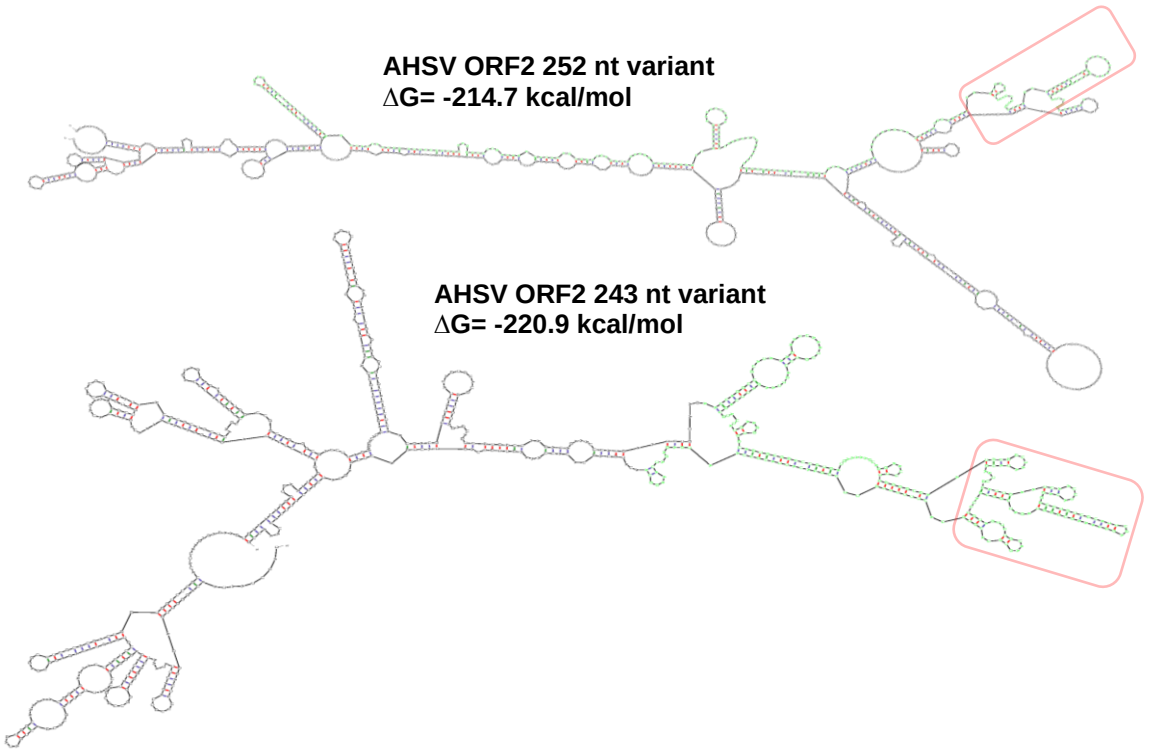


Fig. 2.10: Nucelar localisation prediction for AHSV ORF2 60 aa and BTV ORF2. Nucpred was used to predict NLSs. Red is indicative of a high likelihood of a nuclear localisation signal.

confidence whether AHSV S10 ORF2 actually encodes a protein which will be expressed and functional during the AHSV replication cycle. Experimental work will be required to determine this.

2.3.4. AHSV Seg-10 RNA had no homology to known RNA structures

A synonymous constraint region need not only point towards the utilisation of an alternative reading frame, but could also be indicative of RNA structural elements or segment packaging signals. To determine if ORF2 might be an important RNA element, Seg-10 RNA was subject to several RNA prediction software. The Rfam database did not produce any results to known RNA families or RNA secondary structures in its database. Thermodynamic models is widely used for the prediction of RNA secondary structures, such as Turner's nearest-neighbor model (TURNER AND MATHEWS 2010). Mfold was used to fold Seg-10 RNA sequences which contained one of each of the ORF2 size variants, to see if there were any important RNA structures in the region where ORF2 is found (ZUKER 2003b) (Fig. 2.11). The region spanned by ORF2 (from position 59) is shown in green. The minimum free energy of each RNA secondary structure was also provided (ΔG). The lower the ΔG value, the more thermodynamically stable and most likely a stable accurately predicted RNA secondary structure is. All structures had a very low ΔG indicating stable RNA structures. The stability of an RNA secondary structure is determined mainly by energetically favourable helical regions. Base pair stacking provides stabilizing energy contributions (YAKOVCHUK *et al.* 2006; WANG *et al.* 2019). The entire Seg-10 mRNA structure contained long helical structures with some hairpin loops, multi-branched loops and internal loops. The red boxes on the figure show hairpin structures inside the ORF2 region. Hairpin structures are important RNA structures that might have essential functions during genome packaging.



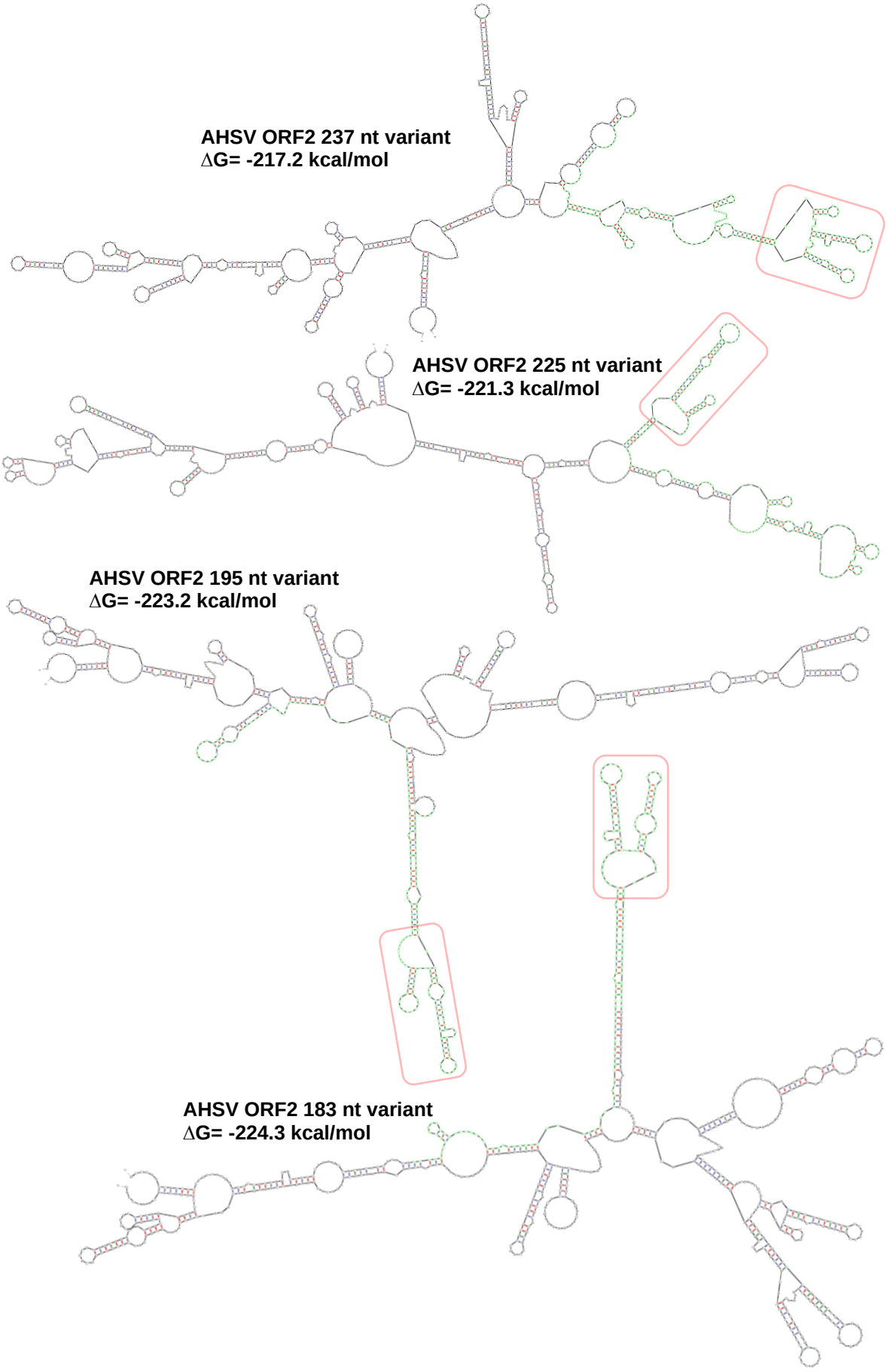


Fig. 2.11: AHSV Seg-10 RNA secondary structures. Seg-10 RNA sequences representing each of the ORF2 size variants were analysed by Mfold. The minimum free energy (ΔG) of each structure is given, with the ORF2 region indicated in green. Light red blocks show hairpins within the ORF2 region.

Furthermore, the Seg-10 RNA sequences were used to search for important consensus sequences such as internal ribosome binding sites. Internal ribosome entry sites (IRESs) are structured RNA elements can recruit ribosomes independently from a 5' cap. Viral IRESs are highly structured (or in some cases unstructured), highly diverse RNA sections with regard to nucleotide length, secondary/tertiary structure and primary sequence (LOPEZ-ULLOA *et al.* 2022). IRESpy and IRESbase were used to search for any known internal ribosome entry site signals, but did not find any indication of IRES near the Seg-10 5' regions. However, when only the 60 nucleotide region preceding ORF2 was subject to IRESite, a few sequences produced alignments. The one present in all six Seg-10 representative sequences was the HAV-HM175 IRES (Fig. 2.12). Nucleotide 14-27 (5' UTR of ORF2) in the Seg-10 sequence was nearly identical to the HAV-HM175 IRES. ORF2 translation could also still occur through other mechanisms such as strand slippage, leaky scanning or ribosome shunting. These are all used by viruses for translation of polycistronic mRNA. Experimental analysis will be required to further assess these different possibilities.

```
Seg-10      14  cttgtcatgaatct 27
                ||| ||||||||||
HAV-HM175   212 ctt-tcatgaatct 224
```

Fig. 2.12: IRES signal identity with AHSV Seg-10. Homology with HAV_HM175 IRES detected by IRESite.

2.4 Discussion

Virus genome segments are under strong selection pressures. A viral genome contains RNA structures that regulate different viral processes (GUPTA AND BANSAL 2020). Many viral genomic RNAs are reported to assume complex dynamic structures during different stages of their life cycle. Viruses also often make use of overlapping reading frames to maximise the amount of coding efficiency within them. Recent bioinformatic analysis revealed overlapping reading frames in genome segments of viruses in the genus *Orbivirus.*

In AHSV and BTV the two smallest genome segments, namely Seg-9 and Seg-10, encode major proteins VP6 and NS3 respectively. Bioinformatics analyses uncovered that Seg-9 and Seg-10 each contain an additional ORF in the +1-reading frame relative to the main ORF (FIRTH 2008). In Seg-9 the product of this +1 ORF was characterised as NS4, a protein shown to potentially be an important virulence factor that interferes with host innate immunity upon AHSV infection (BOUGHAN *et al.* 2020; WALL *et al.* 2021). STEWART *et al.* (2015) characterised the protein encoded by the second ORF from Seg-10 in BTV. BTV ORF2 expresses a small protein (59 aa) in transfected cells. This second ORF is also predicted to exist in Seg-10 of AHSV, but its conservation, size, nature, phylogeny, and selection pressures have not been investigated until now and formed the focus of this study.

We found AHSV ORF2 to be present in all Seg-10 sequences (>380) analysed. This corresponds to results found by STEWART *et al.* (2015), who described ORF2 as conserved in over 300 BTV

sequences and found it to be maintained in 230 AHSV sequences as well as in EHDV, Equine encephalosis virus (EEV) and a few other but not all orbiviruses. This might be an indication of evolutionary pressure to maintain this ORF in the AHSV genome as well as in closely related viruses. As viruses have a high mutation rate, sequenced isolates of the same virus are often substantially different at the nucleotide level. If there are overlapping elements present within viral genes, this can result in regions with unusual evolutionary constraint. Synonymous substitutions would disrupt the additional element. Our FRESCo analysis revealed 3 synonymous constraint regions in the Seg-10 nucleotide alignment. The first SCE overlapped with the late domain of NS3 and the identified ORF2. This SCE did not contain any other known functional regions for the NS3 protein or Seg-10 nucleotide sequence. Synonymous substitutions in the NS3 frame would not change the NS3 amino acid sequence due to the wobble effect, but will change the ORF2 amino acid sequence. Thus, inferring that ORF2 could be the reason for this region's synonymous constraint and is under a selective pressure to be maintained in the +1 reading frame relative to NS3. The second and third synonymous constraint regions overlapped with known functional domains of NS3. These are the two NS3 transmembrane domains of the NS3 integral protein that are essential for its correct functioning (VAN STADEN *et al.* 1998). Their proper conformations are essential for trafficking virus particle to sites of release (FERREIRA-VENTER *et al.* 2019). Transmembrane domains are frequently under more stringent structural constraints and evolve more slowly (JONES *et al.* 1994). The rate of synonymous substitutions in transmembrane domains are typically two times slower than in the rest of the gene. Amino acid content is thus not the reason for the constraint of synonymous substitutions (TOURASSE AND LI 2000). Furthermore, during our SLAC and OLGenie analyses, which considered NS3 and ORF2 protein-specific selection pressures, we found conflicting selection pressures between the two. Overlapping proteins often show conflicting selection pressures and have previously been detected in retroviruses and papillomaviruses (HUGHES AND HUGHES 2005). This is characterised by positive selection in one gene and negative selection in the other. The purifying selection in AHSV NS3 could possibly be driving the positive selection on AHSV ORF2 or, vice versa, positive selection on the AHSV ORF2 gene may favour purifying selection on the overlapping region of NS3.

A viral genome also contains important RNA structures that regulate different viral processes. Orbiviruses enable co-packaging of their segmented genome through the use of trans-acting RNA-RNA interactions (tRRIs). A copy of the complete viral genome is directed for packaging by tRRIs that occur between different viral genome segments. The smaller mRNA segments (Seg-7 to 10) have a crucial role in BTV genome packaging. Seg-10 is has been shown to be involved in tRRI's for correct packaging of the BTV segmented genome (NEWBURN AND WHITE 2019). TRRIs between segments occur at several specific sites, termed segment assortment signals (SASs) (ALSHAIKHAHMED *et al.* 2018). For BTV these are known but they remain undetermined in AHSV. Production of synthetic BT viruses with deleted/silenced regions in Seg-10 showed that the mutant viruses were not genetically stable (FEENSTRA *et al.* 2014b; FEENSTRA *et al.* 2016). Cis-

complementation or random insertions occurred from other segments in the genome to compensate for these mutations. In AHSV, a synthetic virulent AHSV-5 derived by reverse genetics with an in-frame deletion of 77 amino acids codons in genome Seg-10 lead to the inability to produce NS3 and was completely safe for ponies and is a promising vaccine candidate (VAN RIJN *et al.* 2018). The synonymous constraint regions identified in AHSV Seg-10 could also be an indication of important cis-acting RNA packaging signals that are essential for the correct packaging of the AHSV genome into new virions. However, there is no current evidence on the role of Seg-10 in AHSV genome packaging. Further mutational analysis of these synonymous constraint regions could provide insight about their functional importance.

We showed that ORF2 was maintained in all AHSV sequences, implying evolutionary pressures to be maintained. As ORF2 could possibly encode a novel protein, we analysed the predicted protein's size, conservation and phylogeny. ORF2 had stop codons at different positions in different sequences, which resulted in six distinct size classes. The two longest ORF2 proteins (80 and 83 aa) were the most prevalent, being present in over 60% of all sequences analysed. The ORF2 amino acid sequence was highly conserved, with great similarities within and between the different sizes. This high conservation indicates a selection pressure to maintain the protein sequence. Viruses have high mutational rates and this inhibition of amino acid changes suggests some functionality of the ORF2 protein. There was also a connection between the size variant being expressed and the clade which in it groups. When considering the phylogeny of ORF2, we observed its clustering into the previously described three clades identical to Seg-10 and NS3 (SAILLEAU *et al.* 1997; MARTIN *et al.* 1998; VAN NIEKERK *et al.* 2001b; QUAN *et al.* 2008). When excluding the ORF2 region from the sequences, the resulting phylogenetic tree still clustered the same way. Thus, ORF2 is not responsible for the specific clustering but rather the differences over the entire Seg-10 sequence between serotypes.

Viruses do not encode their own ribosomes and are entirely dependent on host translational machinery. Viruses have evolved different non-canonical mechanisms to optimise the translation of their ORFs. IRESs recruit ribosomes to internal positions on mRNAs facilitating efficient translation of an mRNA in a 5' and cap-independent manner (JANG *et al.* 1988; BALVAY *et al.* 2009). These sequences are proposed to fold to adopt highly organized, complex RNA structures that were capable of recruiting ribosomes to AUG initiation codons internally on mRNAs translating additional ORFs identified in several virus families (BALVAY *et al.* 2009). ORF2 is present as an internal additional open reading inside the NS3 open reading frame. Its translation could be initiated from an internal ribosome entry site (IRES) or from leaky scanning translation. IRESs are often used to initiate translation while permitting packaging signals to be accommodated within the 5′ UTR (FIRTH AND BRIERLEY 2012). The only IRES identified during our bioinformatic analysis in AHSV Seg-10 was an IRES that is present in Hepatitis A virus (HAV). This IRES recruits the eukaryotic initiation factor (eIF) 4G, a component of the eIF4F cap-binding protein complex, to achieve cap-independent

translation (BORMAN AND KEAN 1997). This IRES is 450 bp long and thus requires a long 5' UTR to fold into its secondary complexes (BROWN *et al.* 1991). Since the 5' UTR for ORF2 is only 60 nucleotides long it is highly unlikely that this IRES could be responsible for ORF2 translation. During leaky scanning translation, scanning ribosomes that are not successful in initiating translation at the first or second AUG (NS3 and NS3A start codons) will continue scanning until they reach an alternative initiation codon (ORF2 codon) further downstream (KOZAK 2002). Neither the NS3 nor the NS3 start codons are within an optimal Kozak consensus sequence (VAN STADEN AND HUISMANS 1991), and hence would more readily allow leaky scanning. The ORF2 start codon meets the minimum Kozak requirements ($^{A}_{G}$NNAUGG), which would generally increase the efficiency could of its protein translation (KOZAK 1991). This provides support for the hypothesis that ORF2 is translated into a novel protein during AHSV infections, initiated via leaky scanning.

A viral genome contains RNA structures that regulate different viral processes (GUPTA AND BANSAL 2020). Many viral genomic RNAs are reported to assume complex dynamic structures during different stages of their life cycle, as for example illustrated from studies on dengue virus (BROSNAN AND VOINNET 2009; PONTING *et al.* 2009; IGLESIAS AND GAMARNIK 2011). RNA structures are also responsible in modulating the non-canonical translation mechanisms (BELLI AND SAMUEL 1993; RACINE *et al.* 2007; FIRTH AND BRIERLEY 2012). Thus, these RNA elements/regions are highly conserved because of their essential functionality. mRNA–rRNA interactions facilitate ribosome shunting (WALSH AND MOHR 2011). The RNA structures of AHSV Seg-10 show stable structures with hairpin loops lying within the synonymous constraint region. These mutationally constrained hairpin loops could possibly form part of functionally important RNA regions. However, without experimental action it cannot be determined what the function or importance of these hairpins entail.

In the *Reoviridae* family, several viruses have been characterised as having one or two genome segments which are polycistronic, containing more than one ORF per mRNA. In the genus *Fijivirus*, Mal de Rı´o Cuarto virus (MRCV) has two non-overlapping open reading frames found in both genomic segment 7 and segment 9 (GUZMÁN *et al.* 2007). In the genus *Orthoreovirus,* segment 1 of avian reovirus encodes for three different proteins from overlapping ORFs, each having a different function (RACINE AND DUNCAN 2010). The smallest genome segment (Seg-11) in rotavirus contains two overlapping reading frames encoding major protein NSP5 and an additional small 12 kDA protein, NSP6. NSP6 shows some similarities to AHSV ORF2 in that it is about the same size (98 aa), it is expressed from a +1 alternative ORF in the smallest genome segment and its coding sequence lies entirely within that of the major protein, NSP5 (MATTION *et al.* 1991; RAINSFORD AND MCCRAE 2007). When specifically looking at protein similarities between the two, AHSV has a predicted N-terminal helix similar to the N- terminal non amphipathic-helix in NSP6 (HOLLOWAY *et al.* 2015). This helix in NSP6 is necessary for mitochondrial localisation and contains a mitochondrial localisation signal (L**R*L*L**LL). ORF2 contained a similar sequence (R**RRLRRLL). However protein domain detection software did not identify any localisation signals in AHSV ORF2, and thus

this sequence cannot be concluded to be responsible for any putative mitochondrial localisation in a similar way an in NSP6.

In terms of structural predictions, AlphaFold provided similar predicted protein structures of AHSV ORF2 as Jpred but had low confidence for predictions. The lack of homology to other sequences as seen during domain searched in other protein databases underlie this low confidence. Between size variants the predicted structures looked similar for the four smallest variants, while the two larger ORF2 variants looked similar. The low prediction confidence does not necessarily mean that ORF2 is not an expressed protein. During AlphaFold VP8 analysis in rotavirus, it was found that VP8 in serogroup B lacks sequence similarity with VP8 of serogroup A and C resulting in novel folding of VP8 in serogroup B (HU *et al.* 2022). Thus, ORF2 could have novel folding which is completely different from any other viral proteins in the databases. So due to the lack of any homology to know proteins or any protein signals/domains in protein databases, only experimental validation will be able to shed light on ORF2 expression and function during the AHSV replication cycle.

As ORF2 is present in different phylogenetically related viruses, it can be an indication of possibly being evolutionary advantageous for some orbiviruses (STEWART *et al.* 2015). Identifying and understanding the structure and function of new functional elements/proteins from additional open reading frames is critical to understanding viral pathogenesis. Additionally, the number of targets for drug and vaccine design potentially increases. In conclusion the data presented here showed that ORF2 is maintained in all AHSV strains, and that there are significant selection pressures on the nucleotide and protein sequences of ORF2. Due to its high amino acid pairwise similarity scores and conserved Kozak sequence we conclude that ORF2 more likely encodes a protein product, rather than functioning as an RNA structural element. This bioinformatic analysis forms the basis of our current knowledge about ORF2, and experimental efforts are required to further validate these results.

CHAPTER 3

EXPRESSION AND LOCALISTION OF AHSV SEG-10 ORF2 IN DIFFERENT CELL SYSTEMS

3.1 Introduction

AHSV and BTV have 10 linear double stranded RNA segments. This genome encodes for structural and non-structural proteins. The seven structural viral proteins (VP1-VP7) form the virus particle while the four non-structural proteins (NS1-NS4) are expressed during infection in the cytoplasm, promoting the virus life cycle. Thus, these viruses need to express more than 10 proteins from their 10 genome segments. They subsequently make use of overlapping reading frames which encode for two different functional proteins. This was first observed from the two overlapping reading frames on Seg-9 which encodes for VP6 and NS4. BTV shares a lot of structural as well as functional resemblances with AHSV (ATTOUI *et al.* 2009), and more extensive research completed on BTV has guided AHSV studies.

Recent bioinformatics revealed additional open reading frames in BTV and AHSV Seg-10. STEWART *et al.* (2015) characterised the protein encoded by the second ORF from Seg-10 in BTV. The S10-ORF2 encoded protein was predicted to be 50 – 59 aa in size, and when transiently expressed from plasmids it localised to the nucleoli of transfected cells (STEWART *et al.* 2015). During our bioinformatic analysis in Chapter 2 we observed this ORF2 to also be present in over 400 AHSV sequences. Its putative gene product also shows high amino acid conservation, indicative of a possible functional protein. Experimental validation is needed to determine the expression pattern, subcellular localisation profiles and possible function of the protein encoded by AHSV ORF2. Understanding the structure and function of the different viral genes and their encoded proteins in the virus replication cycle, is core to understanding the viral pathogenesis. This will form the basis for the investigations of this chapter.

The aim of this chapter was to express the ORF2 protein in different cell systems. Firstly, expression was done in insect cells as the virus is insect-vectored. Next expression was investigates in mammalian cells which represent the primary host of the virus, followed by expression in bacterial cells in order to purify high amounts for antibody production. Methods used include the baculoviral expression system, transient mammalian expression and bacterial induction respectively. Lastly, we wanted detect whether the ORF2 protein is expressed during AHSV infections.

3.2 Material and methods

3.2.1 Cells

BSR-T7 cells (BSR cells which constitutively express T7 polymerase) (BUCHHOLZ *et al.* 1999), and umbilical cord cells (obtained from Prof Christiaan Potgieter, Deltamune (Pty) Ltd) were maintained as monolayers in Dulbecco's Modified Eagle's Medium (DMEM, Lonza) supplemenyed with 5% fetal calf serum (FCS) (Thermo Scientific),1% antibiotics (streptomycin and penicillin, Lonza) and 1.2% fungizone (Sigma) in a 37°C incubator with 5% CO_2 and 90% humidity. Geneticin (20 µl/ml [v/v]) (Invitrogen) was added to BSR-T7 cells every second passage to maintain the T7 RNA dependent

RNA polymerase (RdRp). The DMEM used for umbilical cord cells were additionally supplemented with 1% non-essential amino acids (NEAA, Lonza).

Spodoptera frugiperda (Sf9) cells (ATCC CRL-1711) were maintained as spinner cultures with shaking at 120 rpm in TC100 medium (Sigma) containing 1% antibiotics (streptomycin and penicillin, Lonza), 1.2% fungizone (Sigma), 10% FCS and 1% pluronic acid and incubated with shaking at 28°C.

3.2.2 AHSV infections

All AHSV strains used in this study are listed in Table 3.1. Recent South African field isolates representing each AHSV serotype were obtained from the Deltamune collection. Virus stocks were obtained by inoculating confluent BSR-T7 monolayers in a T75 flask with virus, and harvested when full cytopathic effect (CPE) was observed and stored at 4°C. BSR-T7 or umbilical cord cells were grown as monolayers on coverslips in a 24-well plate. When at 80% confluency, cells were infected with 100 uL AHSV inoculum and incubated at 37°C for 30 hours.

Table 3.1: AHSV strains used in this study

AHSV strain	Isolate name
AHSV-1 Field	DM21/11
AHSV-4 Field	DM10/14
AHSV-8 Field	DM5/15

3.2.3 Plasmids

All plasmids utilised and generated during this study are listed in Table 3.2.

Table 3.2: plasmid names and description

Plasmid name	Description	Source
pFB	pFastBac™1	ThermoFisher Scientific
pFB-HTA	pFastBac™HTA	ThermoFisher Scientific
pFB-NS4-eGFP	Full length NS4 cloned into pFastBac™1	Shareen Boughan (Department of Biochemistry, Genetics and Microbiology, University of Pretoria)
pFB-ORF2*	Full length ORF2 cloned into pFastBac™1	N/A
pFB-HTA-ORF2*	ORF2 lacking its start codon cloned into pFastBac™-HTA	N/A
pFB-ORF2-eGFP*	ORF2 lacking its stop codon cloned upstream of eGFP gene already present in pFastBac™1	N/A
pcDNA3.1-ORF2	Full length ORF2 cloned into pcDNA3.1	Genscript
pcDNA3.1-ORF2-TST	ORF2 lacking its stop codon linked to a Twin-Strep-Tag cloned into pcDNA3.1	Genscript
phCMV-NS4-eGFP	NS4-eGFP cloned into in phCMV	Shareen Boughan (Department of

		Biochemistry, Genetics and Microbiology, University of Pretoria)
phCMV-ORF2-eGFP*	ORF2 lacking its stop codon cloned upstream of eGFP gene already present in phCMV	N/A
pCMV-eGFP	eGFP gene cloned into pCMV	Litia Yssel (Department of Biochemistry, Genetics and Microbiology, University of Pretoria)
pCMV-T7-eGFP	T7-eGFP gene cloned into pCMV	Shareen Boughan (Department of Biochemistry, Genetics and Microbiology, University of Pretoria)
pUC57-ORF2-S	ORF2 lacking its stop codon linked to a Strep-tag II cloned into pUC	Genscript
pUC57-ORF2-TST	ORF2 lacking its stop codon linked to a Twin-Strep-Tag cloned into pUC	Genscript
pStaby-IB	pStaby containing and Ib gene in its insertion site	Prof Potgieter (Deltamune (Pty) Ltd, Pretoria, South Africa)
pStaby-ORF2-S*	ORF2 lacking its stop codon linked to a Strep-tag II® cloned into pStaby	N/A
pStaby-ORF2-TST*	ORF2 lacking its stop codon linked to a Twin-Strep-Tag® cloned into pStaby	N/A
pSP72-S10	AHSV-4 Seg-10 cloned into pSP72	Dr Liesel Stassen (Department of Biochemistry, Genetics and Microbiology, University of Pretoria)

** Generated during this study*

3.2.4 PCR amplification

The CloneAmp HiFi PCR Premix (Clontech) was used for PCR amplification of all FB-inserts, pFB-eGFP vector and recombinant pFB vectors prepared for baculovirus expression. A reaction of 12.5 µl of CloneAmp HiFi PCR Premix, 7.5 pmol of each forward and reverse primer and less than 100 ng template DNA was prepared. The reaction was filled with dH_2O to a final volume of 25 µl. The reaction was placed in an Applied Biosystems 2720 Thermal cycle with the following cycling conditions: 30 cycles of 98°C for 10 sec, 55°C for 15 sec and 72°C for 5 sec/kb (5 sec or 28 sec).

DreamTaq Green PCR Master Mix (2X) (ThermoFisher Scientific) was used for any PCR amplification of insert or vectors prepared for transient mammalian expression and bacterial expression. A volume of 12.5 µl DreamTaq Green PCR Master Mix (2X) was added to 0.2 µM of each forward and reverse primer and 5 ng template DNA. The reaction was filled with dH_2O to a final volume of 25 µl. The reaction was placed in Applied Biosystems 2720 Thermal cycle with the following cycling conditions: 1 cycle of 95°C for 3 min followed by 25 cycles of 95°C for 30 sec, 55°C for 30 sec and 72°C for 4 min and a final elongation of 72°C for 5 min.

All primers used during this study are listed in Table 3.3. Primers were produced by Integrated DNA Technologies (IDT).

Table 3.3: primers used for PCR during baculoviral, mammalian and bacterial expression

Baculovirus expression system		
Primer Name	**Sequence 5' – 3' ***	**Tm (°C)**
pFB-ORF2 F	**ATCGGGCGCGGATCCATG**GAGAGTCGGGGGCGAT	61
pFB-ORF2 R	**TTCGGACCGGGATCC**TTAAAGTTCTGATACCCACCTGC TTTTTAATTTGA	58
pFB-HT-ORF2 F	**TTTCAGGGCGCCATG**GAGAGTCGGGGGCGATCG	60
pFB-HT-ORF2 R	**TTGAATTCCGGATCC**TTAAAGTTCTGATACCCACCTGCT TTTTAATTTG	57
pFB-ORF2-eGFP insert F	**AAGCGCGCGGAATTC**ATGGAGAGTCGGGGGCGAT	61
pFB-ORF2-eGFP insert R	**CTCGCCCTTGCTCAC**AAGTTCTGATACCCACCTGCTTTT TAATTTGAC	59
pFB-ORF2-eGFP vector F	**GTGAGCAAGGGCGAGGAGCTGTT**	63
pFB-ORF2-eGFP vector R	**GAATTCCGCGCGCTTCGGA**	61
M13 Forward	TGTAAAACGACGGCCAGT	54
M13 Reverse	GAGGAAACAGCTATGACC	50
Transient mammalian expression		
Primer Name	**Sequence 5' – 3' ***	**Tm (°C)**
phCMV-eGFP-ORF2 insert F	**AACCATGGGAGACG**ATGGAGAGTCGGGGGC	69
phCMV-eGFP-ORF2 insert R	**AGAACCTCCTCCTCC**AAGTTCTGATACCCACCTGCTT	66
phCMV-ORF2-eGFP vector F	**GGAGGAGGAGGTTCTGGAGG**	69
phCMV-ORF2-eGFP vector R	**CGTCTCCCATGGTTTTGGTACC**	68
Bacterial expression		
Primer Name	**Sequence 5' - 3' ***	**Tm (°C)**
pStaby-ORF2-S/TST insert F	**TAAGAAGGAGATATACA**TATGGAATCCCGCGGTC	69
pStaby-ORF2-S insert R	**GGTGGTGGTGGTGGTGCTCGA**GTTATTATTTCTCAAAC TGCGGAT	66
pStaby-ORF2-TST insert R	**GGTGGTGGTGGTGGTGCTCGA**GTTATTACTTTTCGAAC TGAGGGTG	69

*Nucleotides in bold indicate overhangs specific to vector, and underlined nucleotides are complementary to ORF2 insert

3.2.5 Agarose gel electrophoresis

Samples were analysed by agarose gel electrophoresis in a 1% or 2% agarose gel prepared using agarose powder (Sigma) dissolved in 1 x TAE (40 mM Tris-HCl, 20 mM Na-Acetate, 1 mM EDTA) and 0.5 µg/µl ethidium bromide. DNA molecular weight markers (Quick-Load® 1 kb Plus DNA Ladder, New England Biolabs) were used to determine the sizes of each DNA fragment. Samples were electrophoresed at 90 V for 30 – 60 min in 1 x TAE. Agarose gels were visualised using the GelDoc™ XR+ Imaging System (Bio-Rad).

3.2.6 DNA purification

The required linear vectors and inserts were excised from a 1% agarose gel. The NucleoSpin® Gel and PCR clean up kit (Macherey-Nagel) was utilised to purify the DNA from the gel slice. The

manufacturer's protocol was followed, this involved column purification using wash buffers included in the kit. The purified DNA was eluted with ddH_2O.

3.2.7 In-Fusion® HD cloning procedure

The In-Fusion® HD Cloning Kit (ClonTech) was utilised for construction of pFastBac recombinants. The reaction contained 2 µl of 5X In-Fusion HD Enzyme Premix, a linearised vector and insert. The reaction was brought to a final volume of 10 µl with dH_2O and incubated for 15 min at 50°C. Afterwards the reaction was put on ice. The recombinant vectors were transformed into competent Stellar cells. The transformation procedure was done by adding 2.5 µl of the In-Fusion mixture to 50 µl thawed Stellar cells and incubated on ice for 30 min. After incubation, cells were heat shocked at 42°C for 45 sec and then placed on ice immediately for 2 min. Next, 450 µl of pre-warmed super optimal broth with catabolic expression (SOC) medium was added to the transformation reaction and incubated at 37°C for 1 h with shaking at 230 rpm. Dilutions ranging from 1:10 to 1:100 were then plated on Luria Broth (LB) agar plates (1% tryptone; 0.5% yeast extract, 1% NaCl; pH 7.4 and 1.2% bacteriological agar) containing 100 µg/mL ampicillin and incubated at 37°C overnight.

3.2.8 Plasmid isolation

All plasmids were isolated from bacterial cultures with the High pure plasmid isolation kit (Roche). The manufacturer's protocol was followed. Briefly, this involved pelleting of bacterial cells from a liquid culture and column purification using wash buffers included in the kit, the purified plasmid was eluted with ddH_2O.

3.2.9 Sequencing and sequencing alignment

Recombinant pFB-ORF2, pFB-HT-ORF2 and pFB-ORF2-eGFP as well as phCMV-ORF2-eGFP, pStaby-ORF2-S and pStaby-ORF2-TST vectors were verified by means of dye-terminator cycle sequencing. Volumes of 1 µl of BigDye and 1 µl of 5 X sequencing buffer were added to 3.2 pmol of forward or reverse primer (Table 3) and 50 ng of the DNA sample to be sequenced. The sequencing reaction was filled with dH_2O to a final volume of 10 µl. The reaction was placed in Applied Biosystems 2720 Thermal cycle with the following cycling conditions: 1 cycle of 96°C for 3 min followed by 30 cycles of 96°C for 10 sec, 50°C for 5 sec and 60°C for 4 min. A volume of 3 µl of 3M NaOAc; pH 4.6 and 62.5 µl of absolute EtOH was added to the sequencing reaction and centrifuged for 30 min at 16 000 xg. The supernatant was discarded and 300 µl of freshly prepared 70% EtOH was immediately added. The mixture was centrifuged for 5 min at 16 000 xg and the supernatant was discarded, this step was then repeated. The supernatant was again discarded, and the pellet was allowed to air dry on a heating block at 60°C for ±10 min. These were analysed by the DNA Sequencing Laboratory at the University of Pretoria using the ABI Prism® 310 Genetic analyser. The putative phCMV-ORF2-eGFP was sequenced with phCMV-eGFP-ORF2 insert F and phCMV-eGFP-ORF2 insert R. The putative pStaby-ORF2-S and pStaby-ORF2-TST were sequenced with pStaby-ORF2-S/TST insert F and pStaby-ORF2-S insert R or pStaby-ORF2-TST insert R, respectively. All

pFB constructs were sequenced with the primers in Table 3.4.

Table 3.4: primers used in sequencing reactions

Primer Name	Description	Sequence 5' - 3'
pFB polyhedrin forward	Binds polyhedrin promoter	TTCCGGATTATTCATACC
pFB1 reverse	Binds downstream of pFB insertion site	GAGGATCATAATCAGCCATACCAC

All sequencing results were aligned to their respective reference sequences using MAFFT. These MAFFT alignment results were imported into SnapGene for visualisation purposes.

3.2.10 Restriction endonuclease digestion

pFB and pFB-HTA were linearised using restriction enzymes (RE) for cloning purposes (30 µl reaction). The restriction enzyme used was BamHI (Buffer B) (Roche). The reactions contained 500 or 300 ng DNA, 3 or 2 µl of 10X buffer and 1.5 or 1 µl of BamHI for the 30 µl or 20 µl respectively and filled to the final volumes with ddH_2O and incubated at 37°C for 1.5 – 3 h. After cloning, all pFB putative recombinants were also confirmed using RE digestion (20 µl reaction). Reactions were analysed by agarose gel electrophoresis.

A recombinant pStaby plasmid containing a 1.3 kb insert was kindly provided by Prof Christiaan Potgieter (Deltamune (Pty) Ltd, Pretoria, South Africa). pStaby is a commercial vector (Delphi Genetics) utilised for bacterial expression using a T7 bacteriophage promoter on the vector and T7 RNA polymerase in *Escherichia coli* (*E. coli*). Vectors were prepared by linearising pStaby with FastDigest NdeI and XhoI enzymes (ThermoFisher Scientific). This resulted in excision of the random insert. Each reaction contained 500 ng of DNA, 2 µl 1 x FastDigest Buffer, 1 µl XhoI FastDigest enzyme, 1 µl NdeI FastDigest enzyme and dH2O to a total volume of 50 µl. Reactions were incubated for 30 min at 37°C and confirmed by agarose gel electrophoresis. This digestion was also used to confirm recombinant pStaby-ORF2-S and pStaby-ORF2-TST.

3.2.11 Transformation into DH10Bac cells

Recombinant pFastBac plasmids were transformed into MAX Efficiency® DH10Bac competent cells (Invitrogen™). A volume of a 100 µl thawed competent cells were transferred to a pre-chilled 15-mL round-bottom polypropylene tube and 20 ng recombinant vector added to the cells. After incubation on ice for 30 min, cells were heat shocked at 42°C for 45 sec and then moved to ice immediately for 2 min. A volume of 900 µl LB broth (1% tryptone; 0.5% yeast extract, 1% NaCl; pH 7.4) was added to the transformation mixture and incubated for 4 hours at 225 rpm. Dilutions ranging from 1:50 to 1:500 were then plated on LB agar plates containing 7 µg/ml gentamycin, 10 µg/ml tetracycline, 50 µg/ml kanamycin, 40 µg/ml Isopropyl β-D-1-thiogalactopyranoside (IPTG) and 100 µg/ml 5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside (X-gal) and incubated at 37°C. Blue and white colonies were selected, grown in a liquid culture containing the appropriate antibiotics and used for bacmid DNA isolation.

3.2.12 Transformation into BL21(DE3)pLysS or Cys21 cells

Recombinant pStaby-ORF2-S and pStaby-ORF2-TST plasmids were transformed into BL21(DE3)pLysS cells (Promega). A 10 ng quantity of plasmid DNA was added to 50 uL of competent BL21 cells and incubated on ice for 30 min. Then the cells were heat-shocked for 30 seconds at 42°C. The volume of the transformation reaction was increased with prewarmed LB broth to a final volume of 500 µl. The transformation reaction was incubated for 1 h at 200 rpm at 37°C. After 1 h the transformation reaction was plated on 100 µg/ml ampicillin-containing agar plates at multiple dilutions (undiluted, 1:5, 1:50; 1:100) and incubated at 37°C overnight.

Recombinant pStaby-ORF2-S and pStaby-ORF2-TST plasmids were transformed into Cys21 cells by Prof Christiaan Potgieter (Deltamune (Pty) Ltd, Pretoria, South Africa).

3.2.13 Bacmid DNA isolation

Blue or white colonies were grown in liquid culture and used for bacmid isolation. A 1 mL culture volume was centrifuged at max speed (16 000 xg) for 1 min and the supernatant discarded. The pellet was resuspended in 300 µl of solution I (15 mM Tris-HCl pH 8, 10 mM EDTA and 100 µg/ml RNase A) and 300 µl of Solution II (0.2 N NaOH, 1% SDS) was added, mixed by inversion and incubated for 5 min at RT. A volume of 300 µl of solution III (3 M potassium acetate pH 5.5) was added incubated on ice for 10 min. The samples were centrifuged at 16 000 xg for 10 min. The supernatant was transferred to a separate tube containing 0.8 ml of isopropanol and kept on ice for 10 min, followed by centrifugation at 16 000 xg speed for 15 min. The supernatant was discarded, and the pellet was washed twice with 0.5 ml of 70% EtOH and centrifuged for 5 min at 16 000 xg speed for each wash. The supernatant was subsequently removed, the pellet was air dried and resuspended in 40 µl of 1 × TE buffer (10 mM Tris pH 8, 1 mM EDTA).

3.2.14 Transfection into Sf9 cells

After isolating and confirming recombinant bacmid DNA, Sf9 insect cells were used for transfection. The Bac-to-Bac® Baculovirus Expression System protocol manual was followed. Briefly, Sf9 cells were grown as a monolayer in a 6-well plate. When confluency was reached, the medium was replaced with antibiotic-free TC-100 medium. Cellfectin® II (8 µl diluted in 100 µl of TC-100 medium) was used as transfection reagent. The bacmid DNA (1 µg diluted in 100 µl of TC-100 medium) was combined with Cellfectin® II dilution and incubated at room temperature (RT) for 30 min. The entire transfection mixture was dropped directly onto Sf9 cells and incubated for 4 hours at 28°C. The medium was then replaced with TC-100 containing antibiotics and 10% FCS and incubated for approximately 7 days at 28°C until CPE was visible.

3.2.15 Amplification of recombinant baculovirus stocks

The P0 recombinant baculoviral stocks were isolated from transfections via scraping the cells and centrifugation at 500 × g for 5 min to remove cell debris. The supernatant was stored as the P0 viral

stock for further amplification. All recombinant baculovirus stocks were amplified by infecting 25 cm^2 or 75 cm^2 flasks containing monolayers of Sf9 cells. The cells were incubated at 28°C for 3-5 days until 100% CPE was observed, thereafter the cells were collected by scraping and centrifuged at 500 × g for 5 min to remove cell debris. The supernatant was stored (P1 virus stock) for later infections needed in expression analysis and confocal microscopy assays. A recombinant baculovirus and a wild type (WT) baculovirus was available and utilised as controls in this study. Sf9 cells grown as monolayers in 6-well plates or on coverslips in 24-well plates were infected with 500 µl or 100 µl baculovirus stocks, respectively. All baculoviruses used during this study are listed in Table 3.5.

Table 3.5: Baculoviral stocks used in this study

Virus Stock	Description
Bac-eGFP	Recombinant baculovirus expressing eGFP
Bac-WT	Baculovirus wildtype
Bac-ORF2*	Recombinant baculovirus expressing AHSV Seg-10 ORF2
Bac-HTA-ORF2*	Recombinant baculovirus expressing ORF2 with an N-terminal His-tag
Bac-ORF2-eGFP*	Recombinant baculovirus expressing an ORF2-eGFP fusion protein

* Generated during this study

3.2.16 Cell Harvest

Sf9 cells and BSR-T7 cells were scraped and collected by centrifugation at 1 550 xg (Eppendorf miniSpin centrifuge) for 10 min, the supernatant was then discarded. The pellet was washed twice with 1 x phosphate buffered saline (PBS) (137 mM NaCl, 2.4 mM KCl, 4.3 mM $Na_2HPO_4.2H_2O$, 1.4 mM KH_2PO_4; pH 7.4) by centrifugation at 900 xg for 5 min for each wash. After discarding the supernatant, the pelleted cells were resuspended in RIPA buffer (1% Triton-X, 0.01 M Tris pH 8, 0.1% SDS, 0.15 M NaCl, 0.5% sodium deoxycholate and 1 protease inhibitor tablet), incubated for 30 min at 4°C, and the lysate was subsequently used for SDS-PAGE or Western blot, or stored at -20°C.

Bacterial cells were harvested using BugBuster®, the cells were pelleted by centrifuging the overnight liquid culture at 10 000 x g for 10 min. A volume of 5 ml/g pellet of BugBuster® was added to the pellet along with 10 uL/g pellet lysonase and incubated on a shaker for 20 min at RT. Extra benzonase was also added in some cases. The soluble (supernatant) and insoluble (pellet) fractions were separated by centrifuging at 16 000 x g for 20 min at 4°C. The insoluble fractions were resuspended in 1x PBS in a volume the same as the supernatant. Additionally 1x Halt™ Protease Inhibitor Cocktail, EDTA-free (ThermoFisher Scientific) was added to all samples.

3.2.17 Transfection of BSR-T7 cells

BSR-T7 cells were grown as monolayers on coverslips and transfected when at about 60% confluency. Prior to transfection, the medium was replaced with 800 µl of opti-MEM® medium (Gibco®) and incubated for 1 h at 37°C, the medium was then replaced with 200 µl of opti-MEM® to just cover the cells. A concentration of 200 – 1000 ng of DNA was added to 150 µl of opti-MEM®. Lipofectamine® 2000 reagent (LF2000) (ThermoFisher Scientific) was added to a separate tube containing 150 µl of opti-MEM®. Both mixtures were incubated for 5 min in the laminar flow at RT. The DNA- opti-MEM® mixture was added to the LF2000-opti-MEM® mixture. The mixture was tapped and slightly shaken to mix and incubated at RT for 20 min. A volume of 40 µl of mixture (equal to 400 ng plasmid DNA) was added to each well directly to the cells dropwise and mixed by tilting. The cells were incubated for 4 hours at 37°C and the medium was replaced with DMEM containing FCS and the appropriate antibiotics. For transient mammalian expression, codon-optimised pCMV-eGFP expressing enhanced green fluorescent protein (eGFP), phCMV-ORF2-eGFP, pcDNA3.1-ORF2 and pcDNA3.1-ORF2-TST expressing ORF2-eGFP, ORF2 and ORF2-TST respectively, were utilised. Cells were then incubated for 16, 24, 36 or 48 h and prepared for CLSM.

3.2.18 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

For SDS-PAGE analysis, protein samples from the cell lysates were resuspended in the appropriate amount of 6 x protein solvent buffer (PSB) (15% mercaptoethanol, 40% glycerol, 12% SDS, 0.375 M Tris; pH 6.6), followed by denaturisation at 90°C for 10 min. A sodium dodecyl sulphate polyacrylamide gel consisted of two layers, a top stacking gel layer (5% polyacrylamide, 0.125 M Tris-HCl pH 6.8, 0.01% TEMED and 0.01% ammonium persulphate) and a bottom separating gel (15% or 20% of 30% polyacrylamide, 0.375 M Tris-HCl pH 8.8, 0.1% SDS, 0.01% TEMED and 0.01% ammonium persulphate) for concentrating and resolving the protein samples respectively. Electrophoresis was carried out in 1 x Tris-glycine-SDS (TGS) (25 mM Tris-HCl pH 8.3, 192 mM glycine, 0.1% SDS) at 130 V for 2-3 hours. Protein sizes were determined using a 10 – 170 kDa PageRuler™ prestained protein molecular weight marker (Thermo Fisher Scientific). Proteins were visualised by staining in Coomassie brilliant blue solution (0.125% Coomassie blue, 50% methanol, 10% acetic acid) for 2 h and subsequently destained in destain solution (5% methanol, 5% acetic acid) RT overnight. Gels were visualised with the GelDoc™ XR+ Imaging System (Bio-Rad).

3.2.19 Western blot

Resolved proteins (from SDS-PAGE) were transferred to a nitrocellulose membrane (Hybond™-C⁺, Amersham Biosciences) in transfer buffer (0.025 M Tris pH 8.8, 192 mM glycine, 20% methanol) at 100 V for 1 h. The nitrocellulose membrane was incubated in 1% blocking solution (1% [w/v] milk powder in 1 x PBS) for 30 min at RT to prevent non-specific binding. The blocking solution was

replaced with the primary antibody diluted in 1% blocking solution (1:100 or 1:500) and incubated with shaking at RT overnight, followed by three consecutive washes with wash buffer (0.05% [v/v] Tween-20 in 1 x PBS) for 5 min. The membrane was then incubated in Protein A peroxidase conjugate (1:1000) for 1 h with shaking at RT. The membrane was subsequently washed three times with wash buffer and one time with 1 x PBS for 5 min for each wash. To detect positive antibody binding, the substrate was prepared by mixing enzyme substrate A (60 mg 4-chloro-1-napthol in 20 ml methanol) with enzyme substrate B (60 µl hydrogen peroxide in 100 ml 1 × PBS) before use and incubated with the membrane. After sufficient band visibility was observed, the reaction was stopped by rinsing with distilled water. Bands were visualised with the GelDoc™ XR+ Imaging System (Bio-Rad).

3.2.20 Protein purification by immobilized metal ion affinity chromatography (IMAC)

Proteins were purified using the Protino® Ni-TED 1000 Packed columns (Macherey-Nagel) as described in the instruction manual (all steps were performed on ice). For purification under native conditions the column was firstly equilibrated by adding 2 ml of 1 x LEW buffer to the column which was allowed to drain by gravity. Next the soluble fraction was added to the pre-equilibrated column for binding. The column was then washed with 1 x LEW buffer twice. The histidine tagged proteins were eluted by adding 1 x Elution buffer to the column which was allowed to drain by gravity into new collection tubes three times.

3.2.21 Microscopy

3.2.21.1 Light and fluorescence microscopy

All transfections and infections of BSR and Sf9 cells, were visualised with the Olympus CKX41 light microscope. Cells could be counted or assessed for the intactness of cell membranes, or to assay transfection efficiency.

3.2.21.2 Confocal laser scanning microscopy (CLSM)

Sf9 infected cells growing on sterile cover slips were incubated and fixed at 30 or 48 hours post infection (hpi). Transfected BSR-T7 cells growing on sterile cover slips were incubated and fixed at 16, 24, 36 or 48 hours post transfection (hpt). AHSV infected BSR-T7 cells were fixed at 30 hpi. The medium was replaced with 4% paraformaldehyde (PFA) dissolved in 1 x PBS and incubated at RT for 30 min. The 4% PFA was removed and replaced with 0.2% TX-100 in 1 x PBS and incubated for 10 min to permeabilise the cell membranes. Cells were either directly labelled with nuclear stain, 6-diamidino-2-phenylindole (DAPI) (Roche Applied Science) or labelled with primary antibody. For DAPI labelling, DAPI (1:1000) diluted in 1% blocking solution was added to cells and incubated for 10 min at RT. The cells were washed once with 1 x PBS for 5 min. Lastly, the coverslip was placed on a microscope glass slide containing 2 µl of VECTASHIELD Mounting Medium (Vector laboratories) and sealed with glue (RavX glue tube). For primary labelling cells were incubated for

30 min at RT in 5% blocking solution, the solution was discarded, and the primary antibody diluted in 1% blocking solution (1:50 or 1:100 or 1:500) was added and incubated overnight at 4°C in a container containing wet tissue paper and sealed with parafilm. Cells were washed three times with wash buffer (0.05% Tween-20 in 1 x PBS) for 5 min for each wash. The secondary antibody (1:250 or 1:500) was then added and incubated for 1 h at RT in a container containing wet tissue paper sealed with parafilm. The cells were washed three times in wash buffer for 5 min for each wash and three times in 1 x PBS for 5 min before nuclear staining with DAPI as above. During transfection of BSR-T7 cells with pcDNA3.1-ORF2-TST, cells were also labelled with Strep-Tactin®XT DY-549 conjugate (IBA Lifesciences). The conjugate was diluted 1:300 in the DAPI blocking solution and incubated for 10 min at RT before washing. Immunofluorescence was visualized using a Zeiss LSM 880 Confocal Laser Scanning Microscope (CLSM) coupled to an Airyscan detector for super resolution microscopy. Cells were viewed using the 10x or 63x objective with or without microscope immersion oil depending on the objective.

3.2.22 Antibodies and conjugates

The antibodies and/or conjugates used in this study are listed in Table 3.6.

Table 3.6: Antibodies and conjugates used for labelling in this study

	Species Reactivity	Dilution	Use	Manufacturer
Primary labelling				
Anti-fibrillarin	Mouse	1:100	CLSM	Abcam; ab5821
Anti-eGFP	Mouse	1:100	Western blot	Sigma-Aldrich
Anti-ORF2	rabbit	1:50	Western blot/CLSM	Deltamune
Anti-NS1	Mouse	1:500	CLSM	Eurofins Technologies Ingenasa
Secondary labelling				
Anti-Alexa Fluor® 594	mouse	1:250	CLSM	ThermoFisher Scientific; A-11005
Anti-Alexa Fluor® 488	rabbit	1:500	CLSM	ThermoFisher Scientific; A-11034
Anti-Alexa Fluor® 633	rabbit	1:250	CLSM	ThermoFisher Scientific; A-21070
Anti-Alexa Fluor® 555	rabbit	1:250	CLSM	ThermoFisher Scientific; A32732
Other				
Strep-Tactin®XT DY-549 conjugate	N/A	1:300	CLSM	iba-lifesciences; 2-1565-050
Protein A peroxidase conjugate	N/A	1:1000	Western blot	Calbiochem

3.2.23 Bacterial expression

For bacterial expression in BL21(DE3)pLysS or Cys21 cells, a recombinant colony was picked and inoculated in 10 ml of LB broth. The culture was incubated overnight at 37°C with shaking at ~120

rpm. The next day, a 50 ml large culture was prepared for induction by adding 1% of overnight culture in LB broth. The large liquid culture was incubated at 37°C with shaking at 120 rpm until an OD600 of 0.4-0.8 was reached. When the correct OD was reached the large culture was induced by adding 0.1 mM, 0.5 mM or 1 mM IPTG and shaking vigorously at 250 rpm for 4h or overnight at 37°C, 16°C or 25°C.

3.2.24 Strep-Tactin®XT High-Capacity Purification

In order to purify the bacterial expressed ORF2-S and ORF2-TST the Strep-Tactin®XT High-Capacity column was used. The column bed volume corresponds to the amount of resin. The columns used were specified to have a column bed volume of 1. Induced bacterial cultures were lysed using BugBuster® as previously described. The column was equilibrated with BugBuster® reagent and left to drain by gravity. The soluble fractions obtained after lysis were syringe filtered. The total filtered soluble fraction was added to the column and left to drain by gravity. This flowthrough was collected and saved. Next, the column was washed with BugBuster® reagent. The wash was also collected and saved. The S or TST tagged proteins were eluted 4 times with 500 µl 1X Buffer BXT (elution buffer). Lysate, flowthrough and wash sample volumes of 5 µl and eluate volumes of 10 µl were analysed via SDS-PAGE.

3.2.25 Preparation of pre-absorbed serum

Sf9 cells or BSR-T7 cells were seeded as monolayers in a 75 cm^2 flask and incubated until 100% confluency was reached. Sf9 cells were infected with Bac-WT and BSR-T7 cells were mock-infected and incubated for 30 h. Cells were scraped and centrifuged for 700 x g for 5 min before removing the supernatant and resuspending the pellet in 1 ml of 1% (w/v) blocking solution (skim milk powder in 1 x PBS) The cells were lysed by multiple passages through a 22G syringe needle, followed by the addition of 10 µl of anti-ORF2 serum to give a 1:100 dilution. The solution was incubated at RT with shaking for 3 hours followed by centrifugation at 1200 x g for 10 min. The supernatant was stored at -20°C until use. The pre-absorbed anti-ORF2 serum was used for labelling during CLSM.

3.2.26 Acrylamide gel purification

To purify proteins from an acrylamide gel, samples were subjected to 15% SDS-PAGE. Duplicates of each sample were loaded adjacent to molecular weight markers. The first duplicate and the molecular weight marker were stained with Coomassie blue stain, while the second duplicate with the molecular weight marker was kept unstained. The stained gel was used to determine the region of excision on the unstained gel. The band representing ORF2 (12 kDa) was excised from the unstained gel and placed in an Eppendorf tube before being crushed. A volume of 150 µl of 1 x PBS was added to just cover the gel slice. The gel solution was incubated at 30°C overnight with shaking. Following incubation, the solution was centrifuged at 10 000 x g for 10 min at 4°C. The supernatant was stored at -20°C, and the presence of the gel purified protein confirmed by SDS-PAGE and Coomassie blue staining.

3.2.27 Hydrophobicity

CLC Main workbench 21 was used to determine the hydrophobicity at each site of NS3, ORF2, HT-ORF2, ORF2-S and ORF2-TST proteins.

3.3 Results

3.3.1 Baculoviral expression of recombinant AHSV Seg-10 ORF2

To investigate the expression and localisation of the AHSV Seg-10 ORF2 protein in insect cells, the Bac-to-Bac® Baculovirus Expression System was used. This system allows the expression/production of the protein of interest in high amounts under the control of a strong polyhedrin promoter. This method uses donor plasmids pFastBac™1 (pFB), or pFastBac™HTA (pFB-HT). These donor plasmids contain a mini-Tn7 element and an ampicillin resistance gene. The gene of interest is cloned into a pFB donor plasmids. The recombinant pFB plasmids are transformed into DH10Bac™ *E. coli* cells containing a bacmid (a very large plasmid containing the baculovirus genome). The bacmid has a kanamycin resistance gene, a DNA segment containing a lacZ gene and a mini-attTn7 site. A helper plasmid containing a tetracycline resistance gene is also present in these specialised *E.coli* cells. Inside transformed DH10Bac™ *E. coli* cells, transposition occurs between the mini-Tn7 element on the pFB vector and the mini-attTn7 attachment site on the bacmid DNA with the aid of proteins from the helper plasmid. This results in recombinant bacmids containing the insert. The recombinant bacmid DNA is purified and transfected into Sf9 cells, which allows formation of recombinant baculoviruses expressing the specific protein encoded by the insert. After amplification of the recombinant baculoviruses, subsequent infections of Sf9 cells with these viruses allows the characterisation of the protein of interest by SDS-PAGE and Western blot analysis, localisation through confocal microscopy as well as purification.

For this study three recombinant baculoviruses were generated, which respectively expressed the native ORF2 protein, a histidine-tagged HT-ORF2 fusion protein, or an eGFP tagged ORF2-eGFP fusion protein.

3.3.2 Production of recombinant pFB plasmids

We aimed to construct three different recombinant pFB vectors (Fig. 3.1). For construction of the pFB-ORF2 vector, the full length ORF2 was cloned into the multiple cloning site of pFastBac™1 (Fig. 3.1 A & D) to enable the expression of the native form of the ORF2 protein from its own start to stop codons. Secondly, we aimed to clone ORF2 into pFastBac™HTA which encodes an N-terminal 6X histidine tag (pFB-HT-ORF2) (Fig. 3.1 B & D) to allow expression of an HT-ORF2 fusion protein that can be purified via histidine-binding columns. Thirdly we designed an ORF2-eGFP fusion protein, which will express an autofluorescing protein that can be detected microscopically to determine the intracellular localisation of the ORF2 protein in insect cells. For this, ORF2 was cloned in frame with and upstream of an eGFP gene previously inserted into a pFastBac™1 vector (Fig. 3.1 C & D). The

three ORF2 inserts used for these cloning procedures will now be referred to as FB-ORF2 insert, FB-HT-ORF2 insert and FB-ORF2-eGFP insert respectively.

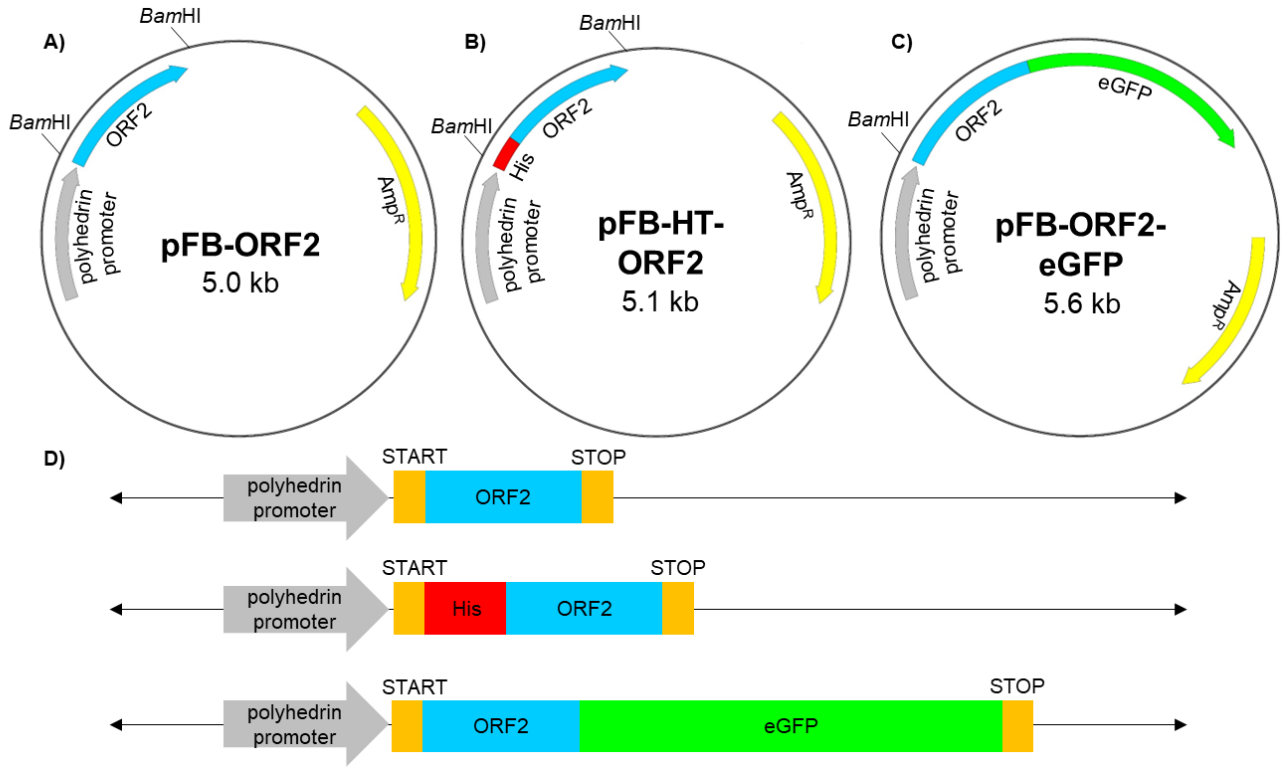


Fig. 3.1: A schematic representation of the three recombinant pFastBac™1 plasmid maps and cloning strategy for the baculoviral expression system. A) Recombinant pFastBac™1 vector, pFB-ORF2. **B)** Recombinant pFastBac™-HTA vector, pFB-HT-ORF2. **C)** Recombinant pFastBac™1 vector, pFB-ORF2-eGFP. Each plasmid contains the polyhedrin promoter and an ampicillin resistance gene (Amp^R). The position of the *Bam*HI restriction site is shown, and the size of each recombinant plasmid is given in kilobases (kb). **D)** Position of translation start and stop sites are indicated for native and fusion proteins.

3.3.2.1 Linearisation and purification of pFB vectors

pFB vector stocks first needed to be linearised. The pFB and pFB-HTA plasmids were linearised using restriction enzyme digestion with *Bam*HI. The success of restriction enzyme digestion was confirmed with an agarose gel (results not shown). A recombinant previously constructed, pFB-NS4-eGFP, was used as a template for substituting NS4 with ORF2. Primers were designed for the linearisation of pFB-NS4-eGFP to ensure the complete removal of the NS4 gene (Table 3.3; pFB-ORF2-eGFP vector F and pFB-ORF2-eGFP vector R). The result of PCR amplification was thus the amplification of the full plasmid pFB-eGFP excluding NS4 and creating complementary overhangs to FB-ORF2-eGFP insert. The success of PCR amplification was confirmed with an agarose gel (results not shown). Gel purification was done of all linear pFB vectors. The agarose gel in Fig. 3.2 shows the undigested pFB plasmid stocks used, and each purified linear pFB vector. All linear vectors correlated with their expected sizes; with purified linear pFB seen at a position correlating to

about 5.0 kb, linear pFB-HT at around 5.1 kb and linear pFB-eGFP at 5.6 kb. These vectors were now ready for use in In-Fusion® cloning.

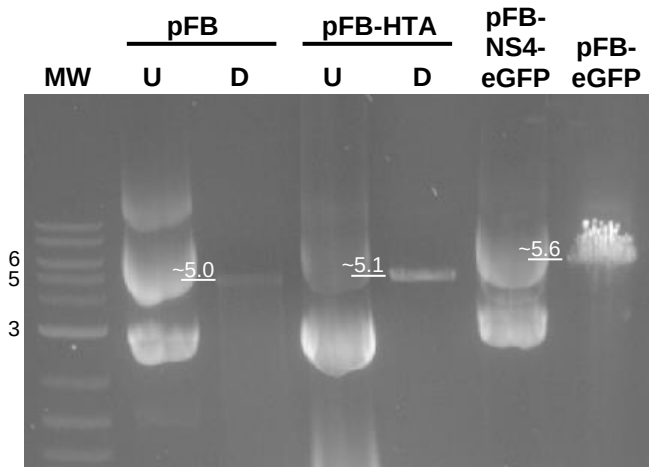


Fig. 3.2: A 1% agarose gel representing undigested and linearised pFB, pFB-HTA and pFB-eGFP plasmids. Digested (D) linear pFB and pFB-HTA plasmids and the purified PCR amplicon of pFB-eGFP were analysed in comparison to their respective undigested (U) controls or the pFB-NS4-eGFP plasmid. A molecular weight marker (MW) was included with the sizes (kb) as indicated on the left.

3.3.2.2 PCR amplification and purification of each ORF2 insert

A plasmid vector pSP72-S10 containing the full-length AHSV-4 Seg-10 gene was used for PCR amplification of ORF2 inserts. Different primers specific for each ORF2 insert were used for amplification (Table 3.3). The FB-ORF2 insert primers (pFB-ORF2 F and pFB-ORF2 R) were designed to amplify the full length ORF2 from segment 10 and produce complementary overhangs to pFB. The FB-HT-ORF2 insert primers (pFB-HT-ORF2 F and pFB-HT-ORF2 R) were designed to amplify ORF2 lacking its start codon and produce complementary overhangs to the vector regions flanking the pFB-HT insertion site. Lastly, FB-ORF2-eGFP insert primers (pFB-ORF2-eGFP insert F and pFB-ORF2-eGFP insert R) were designed to amplify ORF2 lacking its stop codon and produce complementary overhangs to vector sequences and the eGFP gene respectively in pFB-eGFP. This ensured in frame linkage of the eGFP protein at the C-terminal end of ORF2. Successful amplification of each insert was confirmed with agarose gel electrophoresis (results not shown). All PCR amplicons were column purified, and their recovery verified (Fig.3.3). All inserts were of expected size of ~250 bp, pure and of high concentration. The ORF2 inserts were now ready for In-Fusion® cloning into purified, linear pFB vectors.

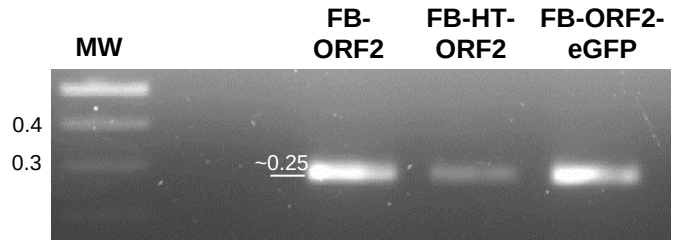


Fig. 3.3: A 2% agarose gel representing purified FB insert DNA. FB-ORF2, FB-HT-ORF2 and FB-ORF2-eGFP specific primers were used for amplification from segment 10 DNA and column purified. A molecular weight marker (MW) was included (kb).

3.3.2.3 Cloning of ORF2 inserts into donor plasmids

In order to clone the three ORF2 inserts into their respective linear pFB donor plasmids, the In-Fusion® cloning procedure was used. FB-ORF2 was cloned into pFB, FB-HT-ORF2 into pFB-HT and FB-ORF2-eGFP into pFB-eGFP. The In-Fusion® reaction mixture was directly transformed into competent Stellar cells. The next day white colonies were observed on all plates except the negative control (non-transformed Stellar cells). Single colonies were used for inoculation and plasmids were isolated. A restriction enzyme digestion was performed with *Bam*HI to confirm the presence of the ORF2 insert in putative recombinants pFB-ORF2, pFB-HTA-ORF2 and pFB-ORF2-eGFP and analysed on a 1% agarose gel (results not shown). *Bam*HI linearised all putative recombinants, but did not excise any insert from pFB-ORF2 and pFB-HT-ORF2 as expected.

Sanger sequencing was performed to obtain better answers on putative recombinants. The pFB polyhedrin forward (anneals to the polyhedrin promoter) and pFB1 reverse (anneals downstream of insertion site) primers were used in the BigDye reaction (Table 3.4). Sanger sequencing results were aligned to the theoretical recombinant pFB plasmid sequences using MAFFT. Full sequences are provided in the Appendix. Recombinants of each pFB-ORF2, pFB-HTA-ORF2 and pFB-ORF2-eGFP were identified, which contained the correct ORF2 insert (Fig. 3.4). For pFB-HTA-ORF2 the start codon was successfully removed and ORF2 was cloned directly downstream of and in frame with the 6x his tag. For pFB-ORF2-eGFP the stop codon of ORF2 successfully removed and ORF2 was in frame with eGFP. Terminal sequences were of low quality and showed mismatches, as is expected in Sanger sequencing reactions. In conclusion, ORF2 was amplified without any PCR-induced errors and correctly cloned for each pFB recombinant and ready for transformation into DH10Bac™ *E. coli*.

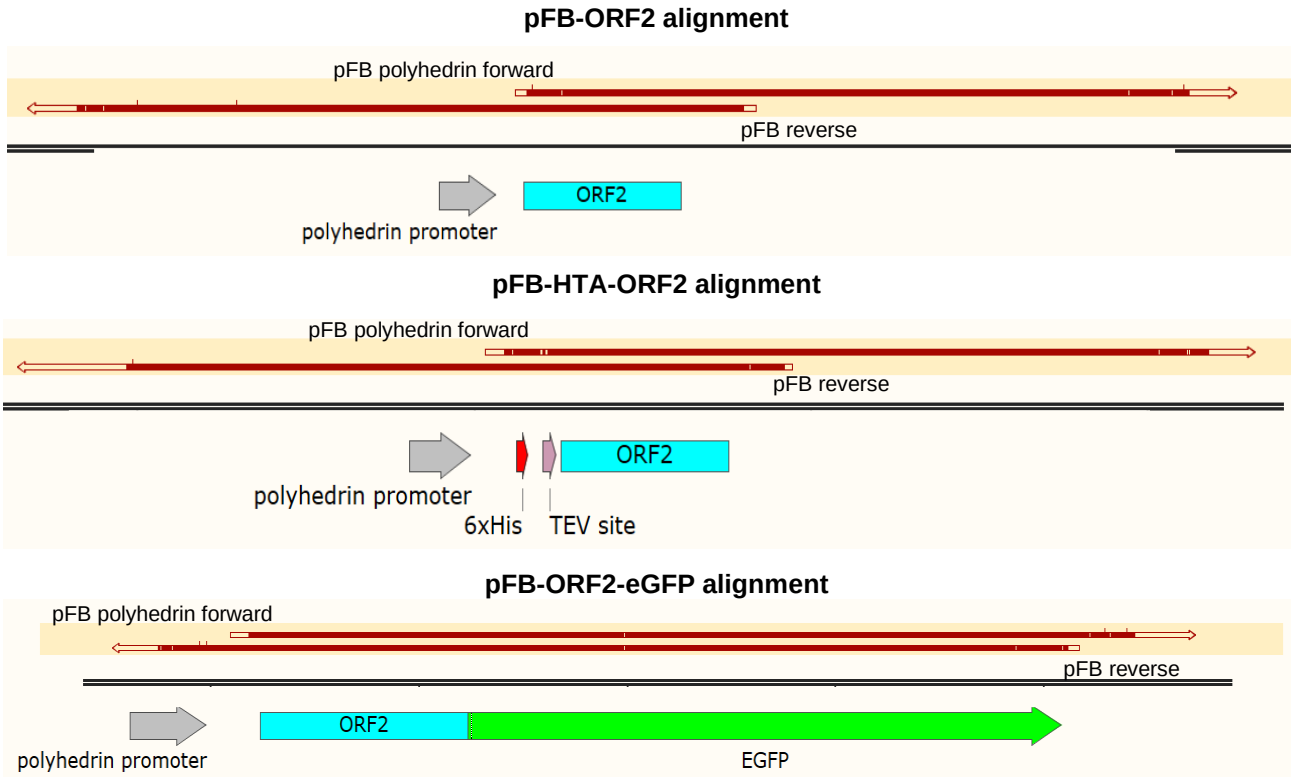


Fig. 3.4: Schematic representation of regions sequenced for recombinant pFB plasmids, MAFFT sequence alignments were imported and visualised with SnapGene. Red bars represent the extent of sequences obtained from Sanger sequencing reactions from pFB polyhedrin forward and pFB reverse primers. White spaces inside red bars represent mismatches in the aligned sequences.

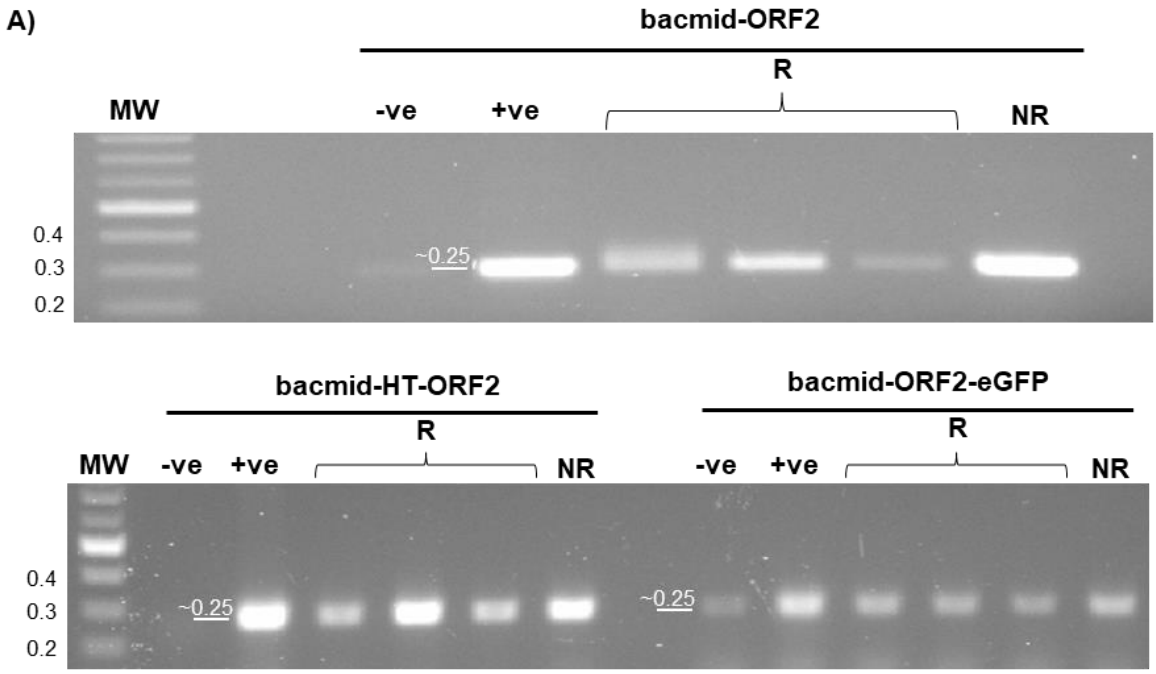
3.3.3 Production of recombinant bacmid

The next step was to generate recombinant bacmid DNA via transposition from the recombinant pFB plasmids. The pFB-ORF2, pFB-HT-ORF2 and pFB-ORF2-eGFP recombinant vectors were now transformed into DH10Bac *E. coli* cells which contain the baculovirus genome (bacmid DNA) and grown under selective conditions. Blue and white colonies were selected for inoculation in a liquid culture, representing clones containing non-recombinant and recombinant bacmid genomes respectively.

3.3.3.1 Isolation and confirmation of recombinant bacmid DNA

Bacmid DNA was isolated, and the presence of the insert was confirmed by PCR. The first PCR reactions were performed with the ORF2 gene-specific primers used to produce amplicons of the respective FB-ORF2 inserts (Fig. 3.5 A). Transposition into the bacmid DNA had now occurred and each pFB-ORF2 insert would now be referred to as bacmid-ORF2, bacmid-HT-ORF2 and bacmid-ORF2-eGFP. The positive control contained the Seg-10 plasmid DNA used for initial production of the inserts. The negative control contained no DNA to ensure contamination was absent. Bacmid DNA from both non-recombinant (blue) and recombinant (white) colonies resulted in an amplification

product with ORF2 gene specific primers. This could be due to mixed colonies resulting in both non-recombinant and recombinant bacmid DNA being isolated together. The ORF2 amplicons for bacmid-ORF2, bacmid-HT-ORF2 and bacmid-ORF2-eGFP were of the expected size of about 250 bp (Fig. 3.5 A). The second set of PCR reactions were performed with M13 forward and reverse primers which anneal to the sites flanking the mini-attTn7 site. Following successful transposition, amplification of a region of the transposed pFB plus the ORF2 insert would occur. All the bacmid DNA samples from white colonies were recombinant and yielded amplicons of the expected size (according to Bac-to-Bac™ manual) of ~2.4 kb + insert size. In the case of recombinant bacmids containing pFB-ORF2 and pFB-HT-ORF2, amplicons were a final size ~2.7 kb (2.4 kb + 250 bp ORF2 insert) (Fig. 3.5 B). Similarly, recombinant bacmids containing pFB-ORF2-eGFP resulted in larger amplicons due to the presence of eGFP, with a band size of ~3.3 kb (2.4 kb + 250 bp ORF2 insert + 710 bp eGFP) (Fig. 3.5 B). In the non-recombinant lanes a smaller fragment of the expected size for bacmid only (without pFB insert) of ~300 bp was obtained. A lighter large band representing amplification from recombinant bacmid DNA was also present in these lanes, further confirming that the blue colonies were mixed. Three recombinant bacmids, namely bacmid-ORF2, bacmid-HT-ORF2 and bacmid-ORF2-eGFP were now ready for transfection into Sf9 cells.



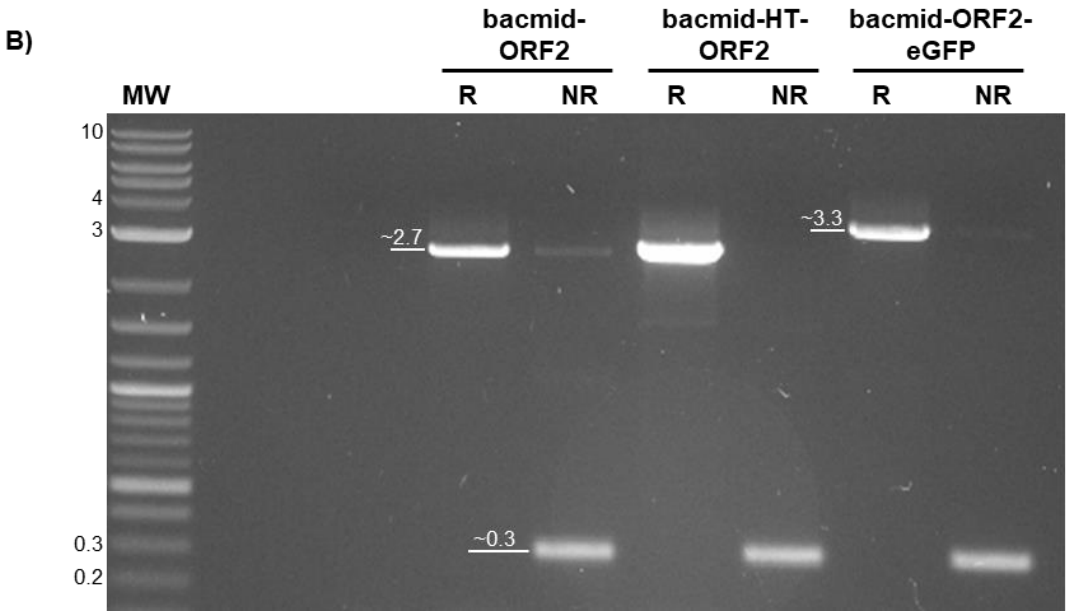


Fig. 3.5: 1% agarose gels representing amplicons of bacmid DNA purified from putative recombinant (R) and non-recombinant (NR) colonies. A) ORF2-specific primers were used for amplification from bacmid DNA. **B)** M13 primers were used for amplification from bacmid DNA. Negative (-ve) and positive (+ve) controls are indicated. A molecular weight marker (MW) was included (kb).

3.3.4 Successful expression of ORF2 from recombinant baculoviruses

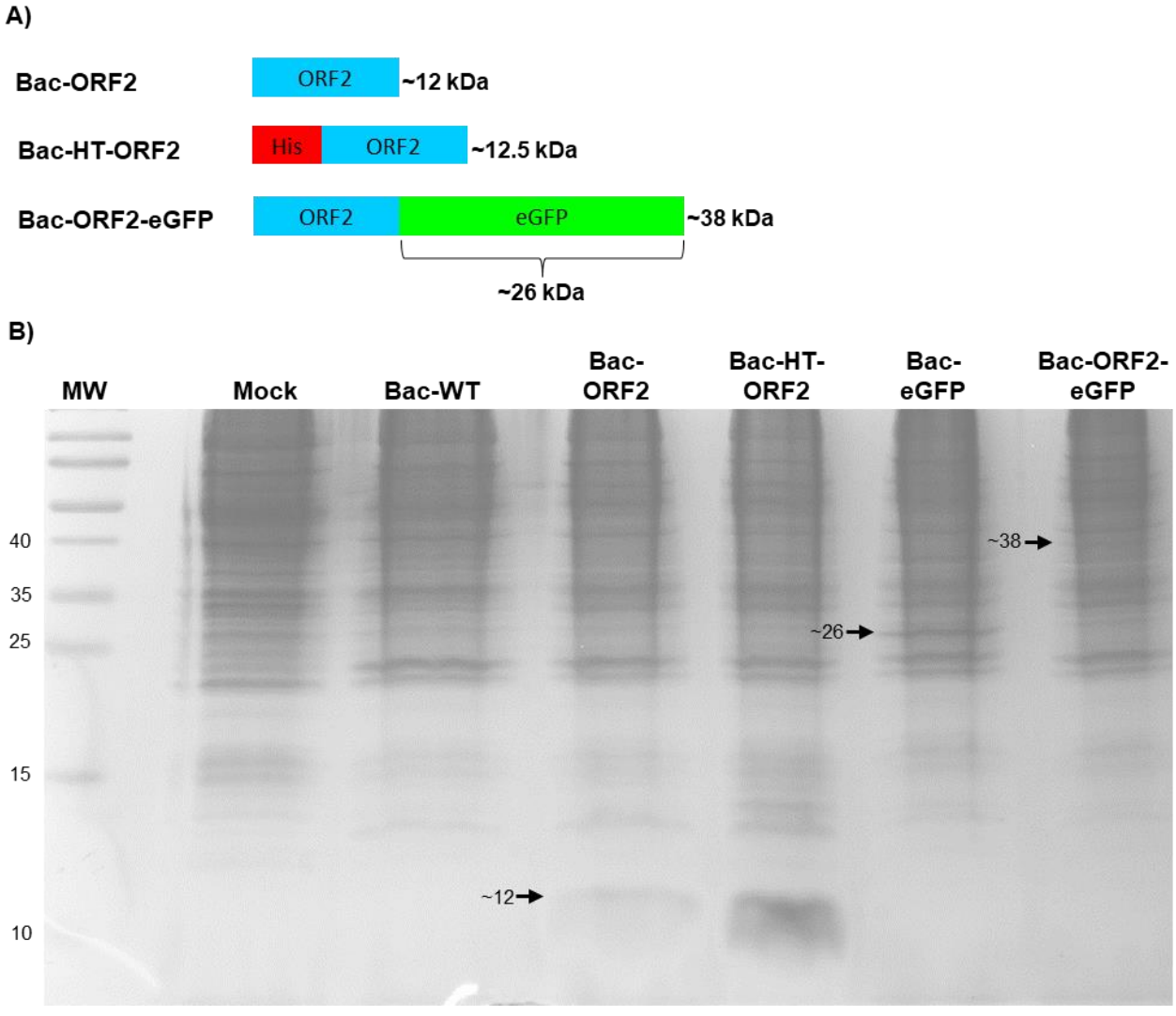
Selected recombinant bacmid DNA was used to transfect Sf9 cells and incubated for 7 days (P1). During incubation transfected Sf9 cells appeared enlarged after two days with an increasing number of unattached cells, compared to mock transfected cells which appeared normal in size and remained attached to the dish. Each viral stock was amplified twice to yield P2 and P3 viral stocks, and the resulting recombinant baculovirus stocks were named Bac-ORF2, Bac-HT-ORF2 and Bac-ORF2-eGFP respectively.

The native full length ORF2 produced by Bac-ORF2 has a predicted size of about 12 kDa. The HT-ORF2 fusion protein produced by Bac-HT-ORF2 has a slightly larger size of about 12.5 kDa. As the eGFP protein alone has a size of about 26 kDa, ORF2-eGFP produced by Bac-ORF2-eGFP has a final size of about 38 kDa (Fig. 3.6 A). Sf9 cells were infected with Bac-ORF2, Bac-HTA-ORF2 or Bac-GFP-ORF2, baculovirus wild-type (Bac-WT) or a baculovirus expressing native eGFP (Bac-eGFP). Mock infected and infected cells were harvested at 48 hpi for analysis. Lysates were first analysed on an SDS-PAGE gel by Coomassie staining (Fig. 3.6 B). The mock infected sample showed no sign of infection, and the Bac-WT did not show any band of the expected size of the ORF2 protein. A clear band with a size of around 12 kDa was observed in Bac-ORF2 and Bac-HTA-ORF2 infected lysates, representing the ORF2 and HT-ORF2 proteins respectively. Native eGFP (~26 kDa) was expressed by Bac-eGFP, and the ORF2-eGFP fusion protein of ~38 kDa can faintly be observed in Bac-GFP-ORF2 lysates.

The proteins from mock infected, Bac-WT, Bac-eGFP and Bac-GFP-ORF2 infected cells were labelled with an anti-eGFP antibody during a Western blot (Fig. 3.6 C). No signal was observed for Bac-WT or mock infected cells, while the eGFP protein of size 26 kDa expressed by Bac-eGFP and the ORF2-eGFP fusion protein of size 38 kDa expressed by Bac-GFP-ORF2 could be clearly distinguished. In conclusion, the baculoviral expression system was successfully utilised to produced recombinant baculoviruses expressing the ORF2 protein of interest.

3.3.5 ORF2 was purified and used for antibody production

The recombinant baculovirus strains, expressing adequate amounts of the ORF2 proteins, could now be utilised to isolate and purify ORF2 for use in antibody production. Anti-ORF2 serum could subsequently be used to detect the presence of native ORF2 protein following AHSV infection. To achieve this, large quantities of pure ORF2 protein was required for injection into rabbits. Bac-HT-ORF2 made this possible, as the presence of the histidine tag from HT-ORF2 would allow column purification. ORF2 expressed by Bac-ORF2 could also be extracted directly from an unstained SDS-PAGE.



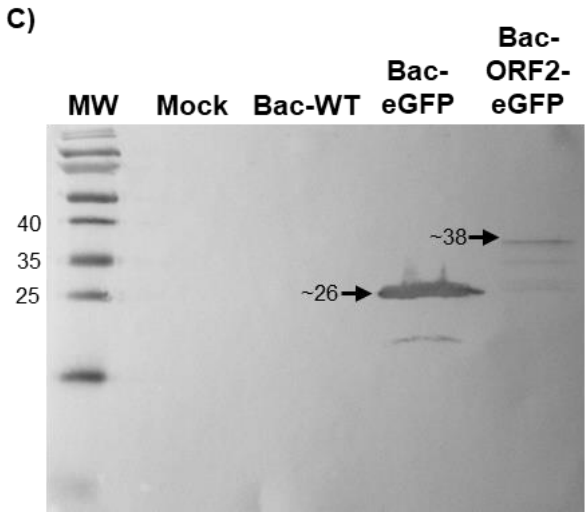


Fig. 3.6: Baculovirus (Bac) infected Sf9 cells expressing the respective recombinant proteins. A) Schematic representation of ORF2 protein, HT-ORF2 and ORF2-eGFP fusion protein sizes expressed by recombinant baculoviruses. **B)** Cell lysates from mock, Bac-WT and recombinant baculoviral infected Sf9 cells were separated by 15% SDS PAGE and stained with Coomassie Blue. **C)** Western Blot of cell lysates from Bac-eGFP and Bac-ORF2-eGFP labelled with anti-eGFP. A molecular weight marker (MW) was included with sizes indicated on the left in kDa.

3.3.5.1 Solubility assay

Firstly, the solubility of HT-ORF2 and normal ORF2 was determined. Both column purification and antibody production require soluble protein to achieve binding of His-tag, and antibody formation in rabbits, respectively. Sf9 cells were infected with Bac-HT-ORF2 or Bac-ORF2, harvested at 48 hpi and lysed. To separate the soluble and insoluble fractions, the lysates were centrifuged at either 3000 x g for 10 min or 16 000 x g for 30 min. The supernatant of each was saved as the soluble fraction, and the pellets resuspended and saved as the insoluble fraction. All samples were run on an SDS-PAGE and stained with Coomassie to visualise solubility (Fig. 3.7). The original cell lysate from Bac-ORF2 expressed the ORF2 protein of the expected 12 kDa size. The mock and Bac-WT lanes did not contain any ORF2 sized bands. Native ORF2 protein was completely soluble, shown by bands only being present in the soluble fractions (supernatant) after centrifugation. Bac-HT-ORF2 expressed the HT-ORF2 protein of the expected size (Fig. 3.7). HT-ORF2 was however mostly insoluble, shown by bright bands in the insoluble fractions (pellet) after centrifugation. At 16 000 x g there was almost no protein left in supernatant. Thus, when ORF is tagged with HT it produces a very insoluble product. This could be due to misfolding of the tagged protein. This is not favourable for column purification and caused problems due to the low amounts of soluble protein.

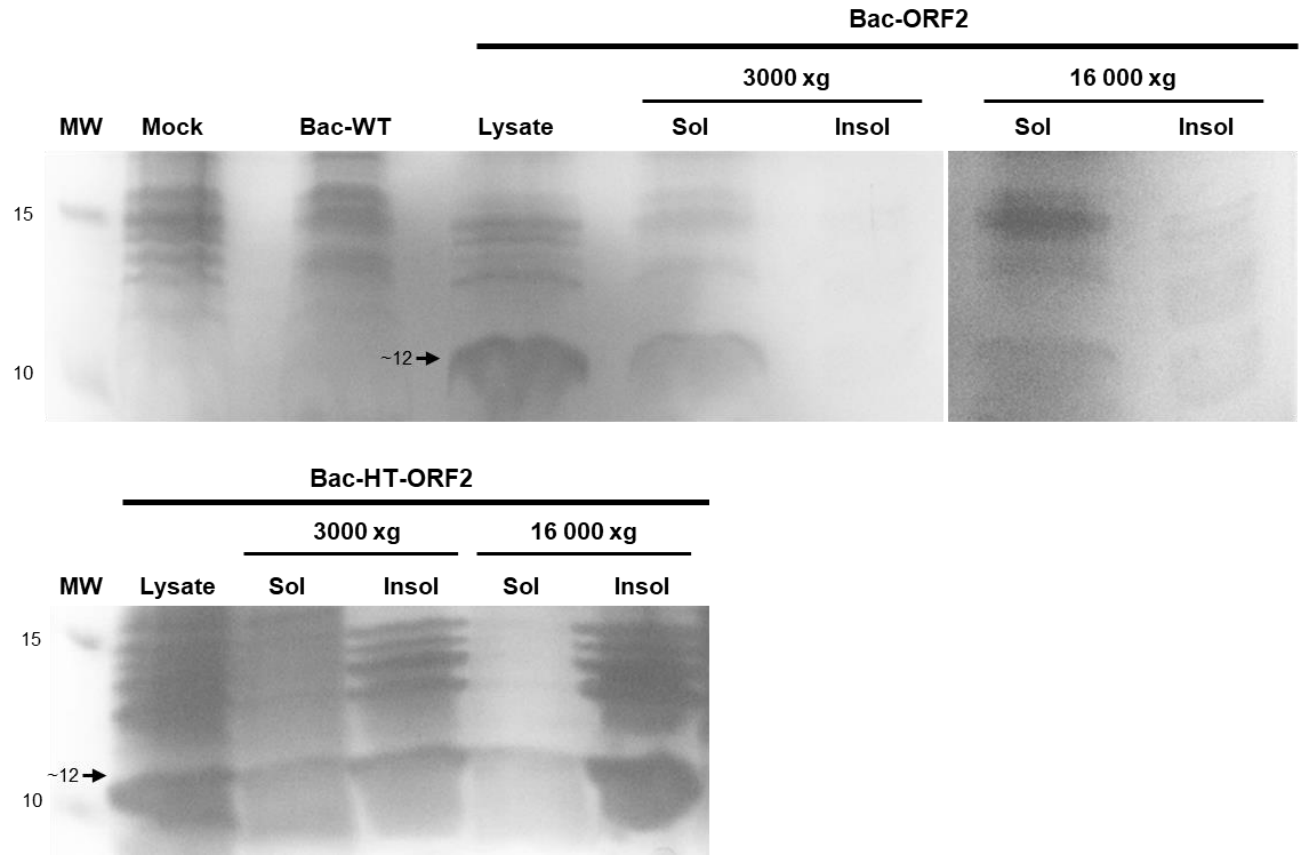


Fig. 3.7: Solubility assay of baculoviral expressed ORF2 and HT-ORF2. Cell lysates from mock, Bac-WT, Bac-ORF2 and Bac-HT-ORF2 infected Sf9 cells were separated by 15% SDS PAGE and stained with Coomassie Blue. The soluble (Sol) and insoluble (Insol) fractions obtained via centrifugation were also analysed. The sizes of molecular weight markers (MW) are indicated on the left in kDa.

3.3.5.2 Column purification of HT-ORF2

The Protino® Ni-TED packed columns allow quick purification of recombinant HT proteins by immobilized metal ion affinity chromatography (IMAC). The 6X his-tag binds to immobilized Ni^{2+} ions in the column. Sf9 cells were infected with Bac-HT-ORF2. After 48 hours the cells were lysed by freezing at -80°C overnight or lysed via dounce homogenisation in LEW buffer. Due to the insolubility of HT-ORF2, a set of samples were additionally treated with 8 M urea (known to increase solubility of proteins). After lysis and centrifugation, the supernatant was run through the column. Higher imidazole concentrations were also used to increase elution from the column. A mock infected cell sample was used as control. An SDS-PAGE was loaded with samples of both lysates, the pellet and supernatant after centrifugation, flow through, washes, and elutions and stained with Coomassie to visualise (results not shown). The HT-ORF2 protein of the expected size was present in lysates and pellets of Bac-HT-ORF2. The supernatants obtained from centrifugation of cell lysates showed a very faint band for the HT-ORF2 protein. Supernatants from cell lysates treated with urea also showed a barely visible HT-ORF2 protein. Thus, urea treatment did not increase solubility of HT-ORF2. After passage through the column there was no HT-ORF2 visible in the flowthrough or washes. Thus, indicating putative attachment of HT-ORF2 protein to the column. HT-ORF2 was also

absent in eluted samples. Possible explanations could be that HT-ORF2 remained attached to the column with or without additional imidazole added to the elution buffer or the already very faint (low concentration) HT-ORF2 from the supernatant was to dilute to be seen during SDS-PAGE analysis. In conclusion the addition of urea did not have any effect on the solubility of HT-ORF2. Higher imidazole concentrations did also not seem to make any difference during elution. Unfortunately, despite multiple attempts, we were unable to purify the HT-ORF2 protein using a column. This could be due to the insolubility of the tagged ORF2 protein, resulting in aggregation. Other explanations could be either that the protein sample was of too low concentration, or that the histidine tag was not exposed and unable to bind freely to the immobilized Ni^{2+} ions in the column resulting in flow through of the protein.

3.3.5.3 Gel purification and antibody production

An alternative approach was therefore followed, whereby the native ORF2 protein was purified. Bac-ORF2 was used to infect Sf9 cells and after 48 hours cells were lysed. Lysates were gel separated, and the ORF2 band excised from an unstained SDS-PAGE. Following elution of the ORF2 protein from gel slices, the purified protein was evaluated by SDS-PAGE and Coomassie staining (Fig. 3.8). The results showed the presence of a single protein band of the expected size of 12 kDa for the gel excised and eluted purified ORF2 (lane 4), corresponding to ORF2 in the Bac-ORF2 lysate (lane 5) and. As seen in the mock and Bac-WT lysates, there were however also native Sf9 and baculoviral proteins of similar or overlapping sizes that could be excised with ORF2. None the less, large scale purification resulted in about 2 mg of protein. The purified ORF2 protein was utilised as antigen for injection into rabbits by Prof Christiaan Potgieter at Deltamune to produce anti-ORF2 serum.

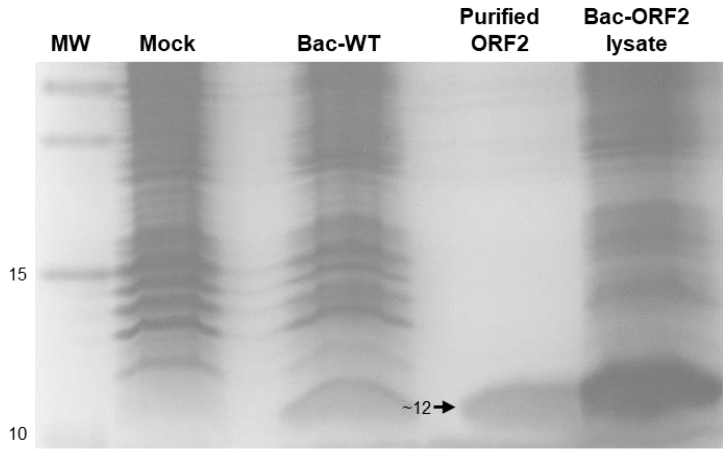


Fig. 3.8: Gel purification of ORF2 from Bac-ORF2 infected Sf9 cells. Cell lysates from mock, Bac-WT, gel eluted purified ORF2 protein and Bac-ORF2 infected Sf9 cells were separated by 15% SDS PAGE and stained with Coomassie Blue. The sizes of molecular weight markers (MW) are indicated on the left in kDa.

To test the produced sera, Sf9 cells were infected with either a recombinant baculovirus (Bac-ORF2, Bac-HTA-ORF2 or Bac-GFP-ORF2), or wild-type baculovirus (Bac-WT). Cells were harvested at 48 hpi and lysed. Cell lysates were gel separated and labelled with anti-ORF2 serum during a Western blot. (Fig 3.9). Fig. 3.9 A and B show Western blots labelled with serum obtained after the first and booster immunisations respectively. Mock, Bac-WT and Bac-ORF2-eGFP lysates showed little to no reaction to the first immunisation anti-ORF2 serum, while very faint bands were detected in Bac-ORF2 and Bac-HT-ORF2 lysates corresponding to the size of ORF2 protein (Fig 3.9 A). When labelled with anti-ORF2 serum from the second bleed following the booster, much more background was detected (Fig 3.9 B). As expected, antibodies in this serum bound to some Sf9 and specifically to a baculoviral protein of similar size as ORF2. This was likely the result of other co-migrating proteins being excised from the gel along with ORF2 during gel purification, providing a multi-protein antigen source. However bands were detected in Bac-ORF2 and Bac-HT-ORF2 lysates which corresponded to the size of the ORF2 protein. The Bac-ORF2-eGFP lane also contained a band of size 12 kDa, which could be explained by overflow when loading the gel. Furthermore, there was a lot of non-specific background signal. However, since there was specific labelling of the ORF2 protein when using the anti-ORF2 booster serum, this serum was subsequently used for detection of ORF2 during localisation studies in Sf9 and mammalian cells via confocal microscopy.

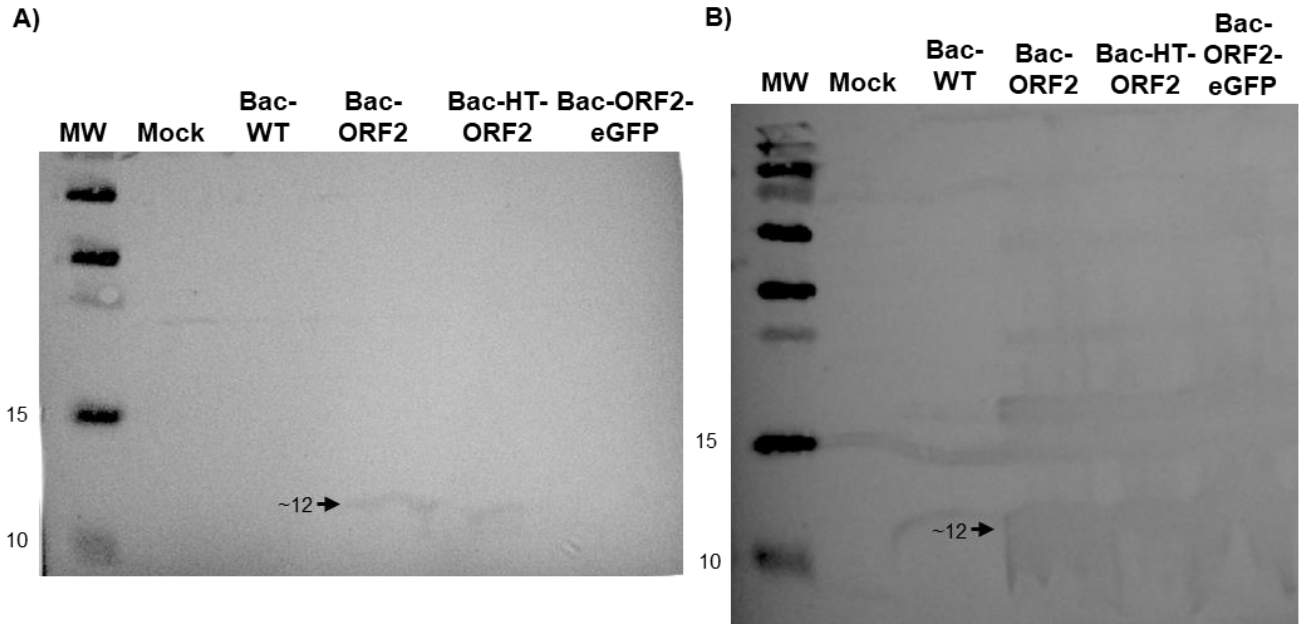


Fig. 3.9: Western Blot to test anti-ORF2 serum: Cell lysates from mock infections, Bac-WT, Bac-ORF2, Bac-HTA-ORF2, Bac-eGFP and Bac-ORF2-eGFP infected Sf9 cells were separated by 15% SDS-PAGE, transferred to a membrane and reacted with **A)** single immunisation anti-ORF2 serum or **B)** booster immunisation anti-ORF2 serum. The sizes of molecular weight markers (MW) are indicated on the left of the blot in kDa.

3.3.6 ORF2 shows punctate localisation in infected Sf9 cells

Following the verification of the successful expression of native ORF2, HT-ORF2 and ORF2-eGFP fusion proteins by the respective recombinant baculoviruses, the intracellular localisation of each was analysed via confocal microscopy. ORF2-eGFP autofluoresces in infected Sf9 cells, while ORF2 and HT-ORF2 were labelled with anti-ORF2 serum for detection. Furthermore, co-labelling of ORF2-eGFP was done with anti-fibrillarin or anti-ORF2 serum.

3.3.6.1 Detection of autofluorescent ORF2-eGFP

Sf9 cells were grown on coverslips and either mock infected or infected with Bac-WT, Bac-eGFP or Bac-ORF2-eGFP. One set of infected cells was fixed at 30 hpi and the other at 48 hpi. The nuclei were stained with DAPI (blue) and cells visualised by confocal laser scanning microscopy (CLSM) (Fig. 3.10). The time points of 30 hpi and 48 hpi were chosen as the genes encoding the proteins of interest are expressed under control of the polyhedrin promoter. This promotor serves as a very late promoter and expression only occurs during the late stages of the viral life cycle from 18 hpi or later (CHEN *et al.* 2013). At 30 hpi and 48 hpi mock infected cells were small as compared to all virus infected cells. Bac-WT infection resulted in enlarged nuclei and cells, as expected from a virus that replicates in the nucleus. Both mock and Bac-WT infections displayed low to no background fluorescence. Sf9 cells infected with Bac-eGFP expressed homogenous green fluorescence over the entire cell (cytoplasm plus nucleus) with no specific localisation. eGFP is a soluble protein, and as a result of its small size it also enters the nucleus via passive diffusion through the nuclear pores. Sf9 cells infected with Bac-ORF2-eGFP had clear specific fluorescence patterns. ORF2-eGFP seemed to aggregate, forming small foci of bright fluorescence, rather than being spread evenly over the entire cell. At 30 hpi the ORF2-eGFP protein showed punctate perinuclear localisation but was also present inside the nucleus in significant quantities. At 48 hpi ORF2-eGFP seemed to be more concentrated in the nucleus. This could be because of the enlargement of the nucleus causing the cytoplasm to shrink, obscuring some ORF2-eGFP cytoplasmic localisation. The ORF2-eGFP fusion protein is still below the 60 kDa size limit, which can theoretically allow passive diffusion through the nuclear membrane (WANG AND BRATTAIN 2007).

In the case of BTV, the segment-10 ORF2 protein localised specifically to the nucleoli within the nucleus (STEWART *et al.* 2015). To investigate potential nucleolar localisation of AHSV ORF2, another set of infected Bac-ORF2-eGFP and controls Sf9 cells were fixed at 30 hpi or 48 hpi. This time nucleoli were labelled with a mouse anti-fibrillarin primary antibody, and anti-mouse AF594 secondary antibody which fluoresces red. The nuclei were stained with DAPI (blue) prior to analysis by CLSM (Fig. 3.11). The same characteristic localisation of ORF2-eGFP was observed as per the previous figure. The red fluorescence from anti-fibrillarin labelled nucleoli did not overlap specifically with any green fluorescence from ORF2-eGFP, as can be clearly seen in the fluorescence intensity graphs. In conclusion, the AHSV ORF2 protein did not colocalise with nucleoli.

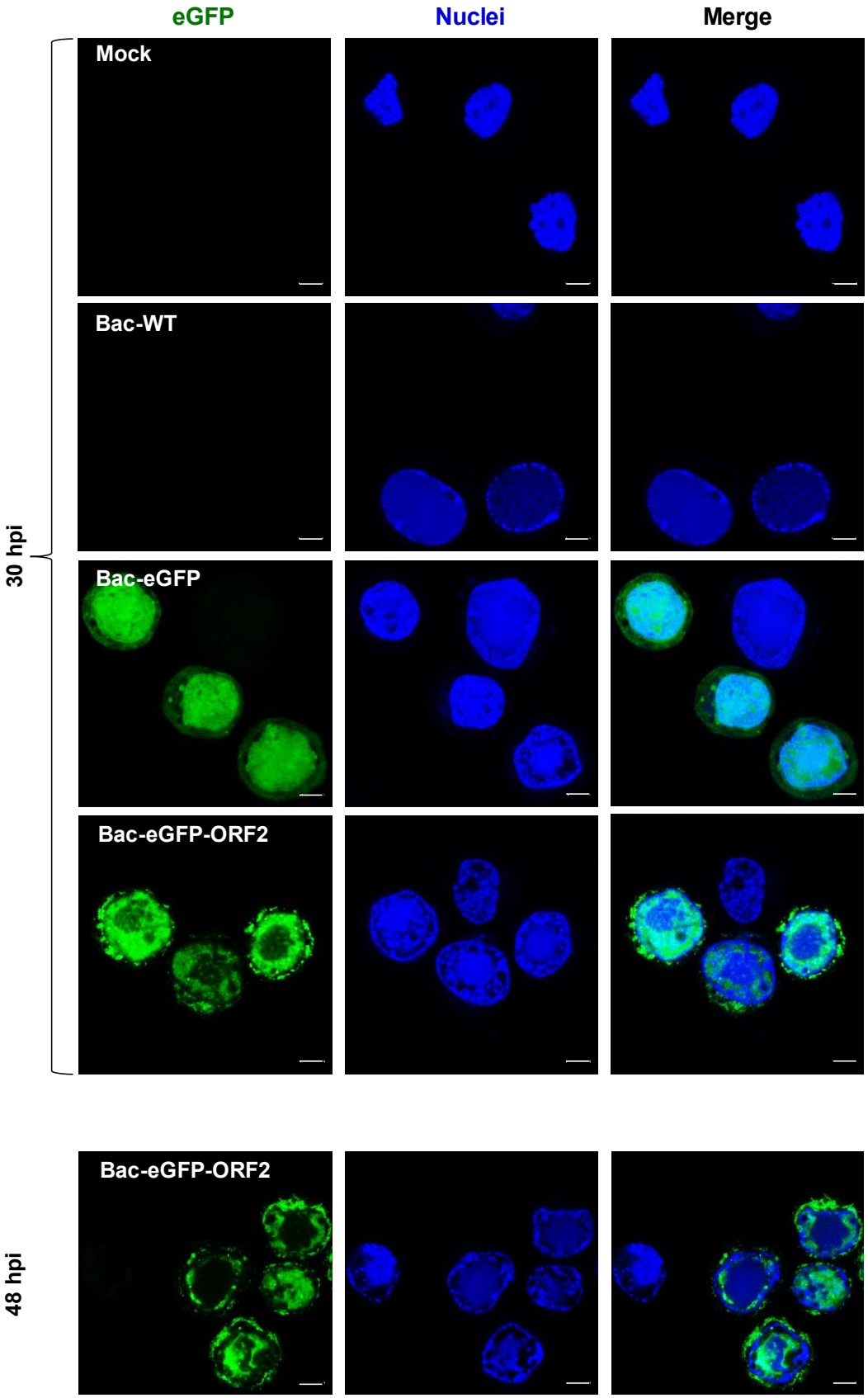


Fig. 3.10: Intracellular localisation of ORF2-eGFP. Sf9 cells were mock-infected or infected with Bac-WT, Bac-eGFP or Bac-ORF2-eGFP. Cells were fixed at 30 hpi or 48 hpi and autofluorescence of eGFP or ORF2-eGFP (green) viewed by CLSM. Nuclei were stained with DAPI (blue). Size bars represent 5µm.

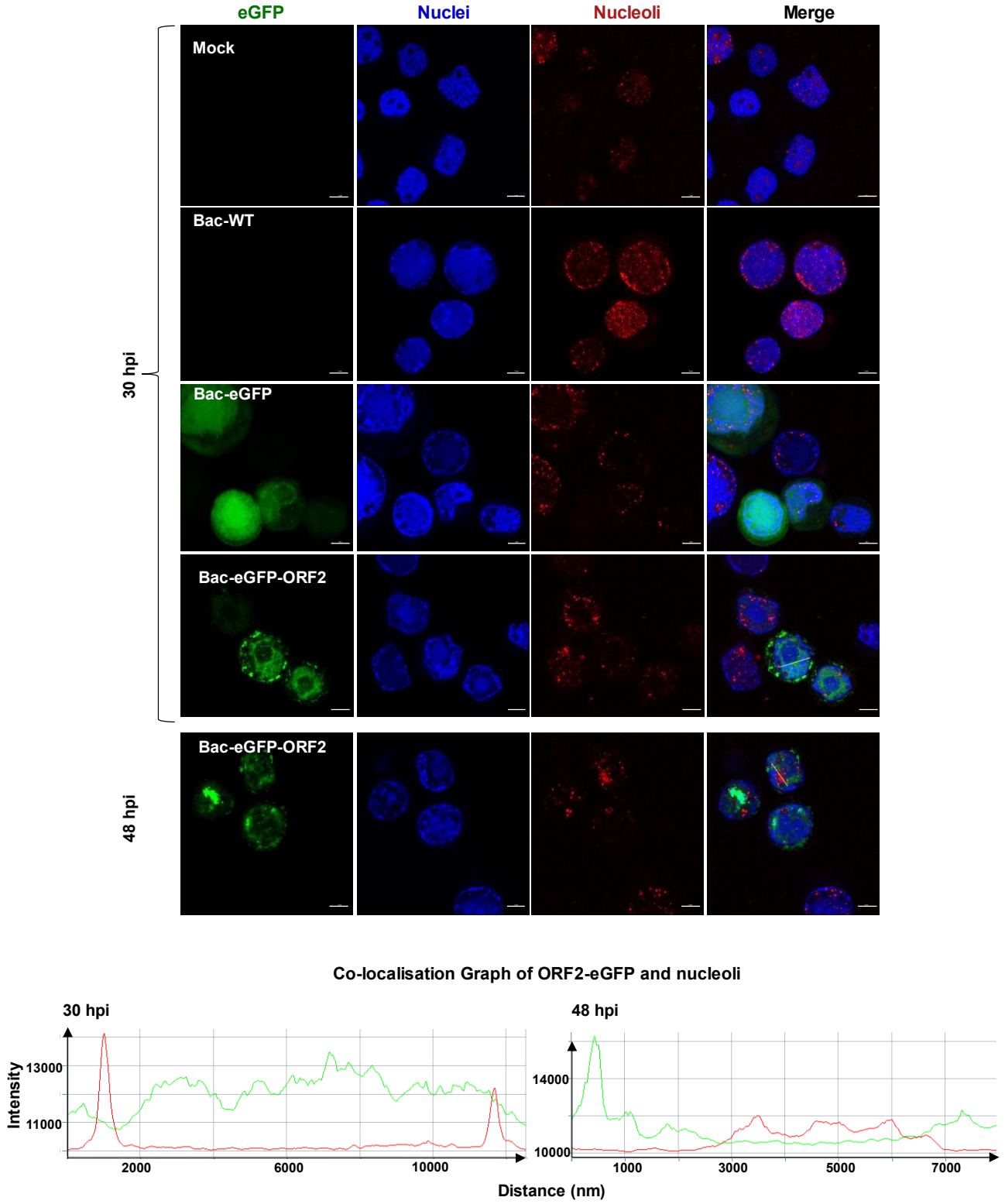


Fig. 3.11: Intracellular co-localisation of ORF2-eGFP and fibrillarin. Sf9 cells were mock-infected or infected with Bac-WT, Bac-eGFP or Bac-ORF2-eGFP. Cells were fixed at 30 hpi or 48 hpi and autofluorescence of eGFP of ORF2-eGFP (green) viewed by CLSM. Nucleoli were labelled with mouse anti-fibrillarin primary antibody, and anti-mouse AF594 secondary antibody (red). Nuclei were stained with DAPI (blue). Histograms show intensity of green (eGFP) and red (fibrillarin) fluorescence measured along the faded white lines in the merged images for Bac-ORF2-eGFP. Size bars represent 5µm.

Lastly, for Bac-ORF2-eGFP a third set of infected Bac-ORF2-eGFP and control Sf9 cells were fixed at 30 hpi or 48 hpi. This time cells were labelled with the anti-ORF2 serum produced in rabbits as primary antibody and detected with an anti-rabbit AF633-conjugated secondary antibody (Fig. 3.12). There was red fluorescence present in mock, Bac-WT and Bac-eGFP infected Sf9 cells. This was expected, as the anti-ORF2 serum also interacted with some Sf9 and baculoviral proteins as seen in the Western blot (Fig. 3.9). The anti-ORF2 serum showed a feint red signal in the cytoplasm in mock, Bac-WT and Bac-eGFP infected Sf9 cells. The red fluorescence was brighter in Bac-ORF2-eGFP infected cells, and additionally some red fluorescent foci were seen in the nucleus. There was unfortunately no clear overlap or any correlation between the green autofluorescence from ORF2-eGFP and the red anti-ORF2 serum labelling.

To get rid of some background binding, pre-absorption was done by incubating a dilution of 1:100 anti-ORF2 serum with a Bac-WT infected Sf9 cell lysate. Cells were lysed by passaging through a 22 G needle. Antibodies in the serum that bound to cytoplasmic cellular or baculoviral proteins were collected by centrifugation, theoretically resulting in a supernatant containing only unbound antibody. This pre-absorbed serum was used to label another set of Bac-ORF2-eGFP infected Sf9 cells and visualised with CLSM (results not shown). There was no major difference in the labelling compared to untreated anti-ORF2 serum. Some of the cytoplasmic red fluorescence was not visible anymore but nuclear red labelling remained the same. This could be since the needle was not able to lyse the nuclear membrane and so nuclear proteins were not able to bind and be removed during pre-absorption of the serum. It was still not possible to observe any green ORF2-eGFP and red anti-ORF2 pre-absorbed serum signal colocalization. The red labelling was also very feint and showed a lot of homogenous red background when overexposed, making localisation observations difficult and not trustworthy.

It could be possible that the specific epitopes which are the targets of the anti-ORF2 serum were unavailable or inaccessible for binding. The eGFP protein is much larger than the ORF2 protein and might have influenced the ability of the anti-ORF2 antibody to bind to ORF2 epitopes. Furthermore, the serum was produced against denatured/linear ORF2 protein excised from a gel, with the result that antibodies directed against linear epitopes of ORF2 might not recognise or efficiently bind their targets in the native and properly folded three-dimensional conformation of ORF2 within the cellular environment.

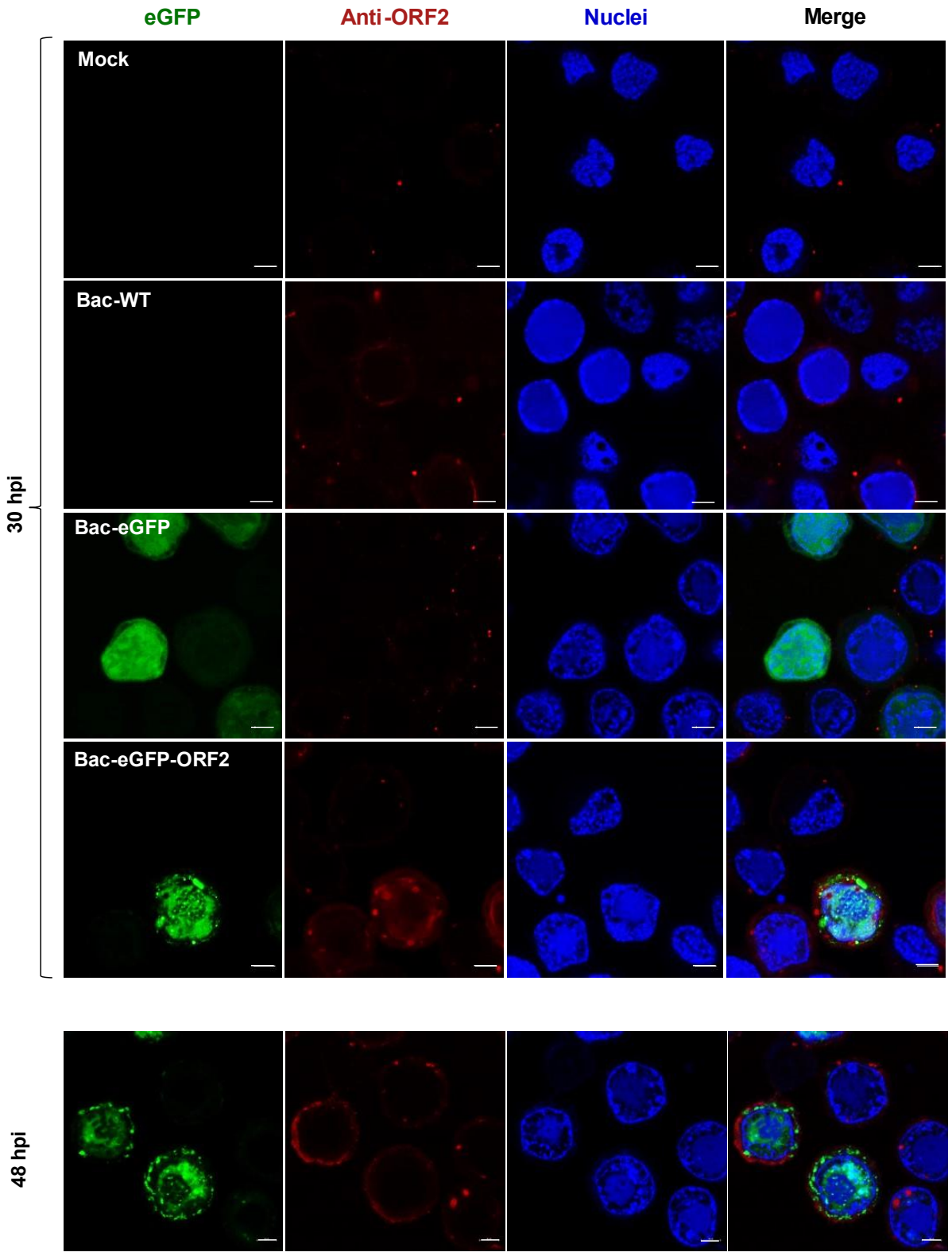


Fig. 3.12: Intracellular localisation of ORF2-eGFP combined with labelling with anti-ORF2 serum. Sf9 cells were mock-infected or infected with Bac-WT, Bac-eGFP or Bac-ORF2-eGFP. Cells were fixed at 30 hpi or 48 hpi and autofluorescence of eGFP or ORF2-eGFP (green) viewed by CLSM. Cells labelled with primary anti-ORF2 serum, and anti-rabbit AF633 secondary antibody (red). Nuclei were stained with DAPI (blue). Size bars represent 5µm.

3.3.6.2 Detection of ORF2 and HT-ORF2 with anti-ORF2 serum

Sf9 cells were grown on coverslips and infected with Bac-ORF2 or Bac-HT-ORF2. The negative controls used were mock infected or wild type baculovirus (Bac-WT) infected Sf9 cells. Cells were fixed at 30 hpi or 48 hpi, labelled with primary anti-ORF2 serum and anti-rabbit AF488 secondary antibody and nuclei stained with DAPI before analysis by CLSM (Fig. 3.13).

Once again, the anti-ORF2 serum showed some background labelling of Sf9 and baculovirus proteins. There were no distinct differences between results from 30 hpi and 48 hpi. Bac-ORF2 and Bac-HT-ORF2 however showed much more fluorescence and much less background than Bac-WT or mock infected cells (Fig. 3.13). HT-ORF2 also displayed a higher intensity of fluorescence than ORF2, which correlates with results in Fig. 3.6 C where HT-ORF2 showed higher expression on an SDS-PAGE than ORF2. Some bright cytoplasmic foci were present in Bac-ORF2 and Bac-HT-ORF2 which were not present in Bac-WT infections, and could indicate the ORF2 protein. However no definite conclusions could be made, as we were hesitant to assume that the anti-ORF2 serum contained any ORF2 specific antibodies as ORF2-eGFP autofluorescence did not correlate with the anti-ORF2 serum labelling as described in the previous section.

To attempt to resolve the problem of non-specific binding to Sf9 and baculoviral proteins, another batch of pre-absorbed anti-ORF2 sera was prepared as described above and used for labelling (Fig 3.14). Bac-WT and mock infected Sf9 cells showed much less background binding with some minor nuclear background fluorescence, and little to no cytoplasmic fluorescence. This suggested that any other cytoplasmic fluorescence would be due to binding of the anti-ORF2 serum to ORF2 specifically. Bac-ORF2 and Bac-HT-ORF2 displayed large foci in the cytoplasm of Sf9 cells. These could indicate the expressed ORF2 protein. When comparing the localisation pattern of autofluorescing eGFP-tagged ORF2, versus anti-ORF2 pre-absorbed serum labelled ORF2 and HT-ORF2, there are some similarities and differences (Fig 3.15). Cytoplasmic fluorescence of all three showed aggregation of the proteins into small to larger foci, with minor nuclear presence. There was much less nuclear fluorescence in Bac-ORF2 and Bac-HT-ORF2 than in Bac-ORF2-eGFP infections.

In conclusion, the recombinant baculoviruses were generated to produce large amounts of the AHSV ORF2 protein and to obtain an initial detection and localisation pattern of ORF2 in insect cells, a potential mimic of the AHSV *Culicoides* insect vector. The protein was also purified from Bac-ORF2 infected cells for antibody production for specific detection of ORF2 in other systems. Unfortunately, background binding made it difficult to accurately determine the antiserum's specificity for ORF2. This serum did show putative binding to the ORF2 protein when considering the profiles obtained with pre-absorbed anti-ORF2 serum, and therefore potentially provides a way to detect ORF2 proteins in other cell types. Thus, the anti-ORF2 serum can be used during transient mammalian expression of ORF2 and AHSV infections in BSR-T7 cells. These should lack Sf9 cellular proteins as well as baculoviral proteins, thus less or no background binding is expected to occur.

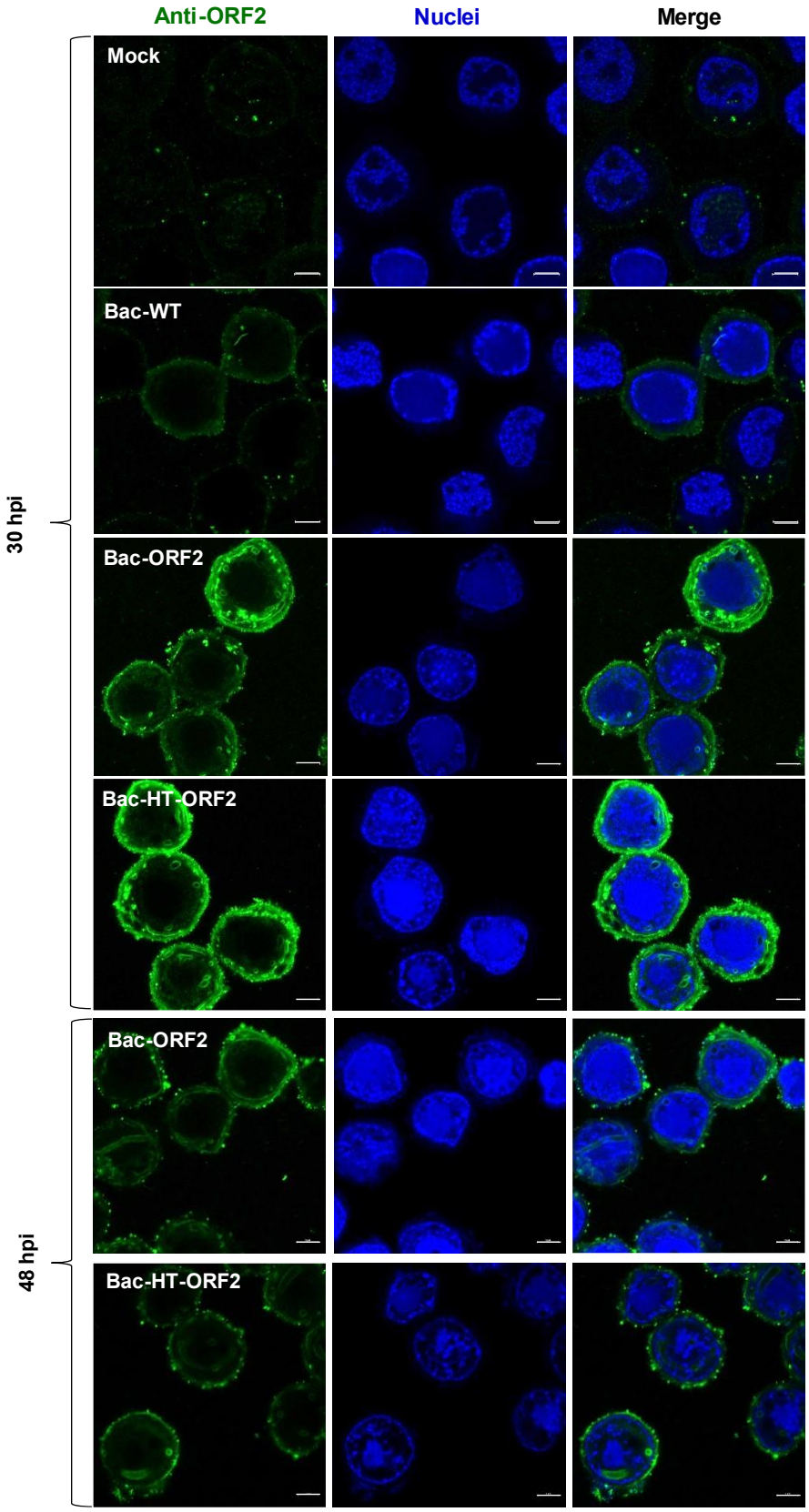


Fig. 3.13: Intracellular localisation of ORF2 and HT-ORF2 labelled with anti-ORF2 serum. Sf9 cells were mock-infected or infected with Bac-WT, Bac-ORF2 or Bac-HT-ORF2. Cells were fixed at 30 hpi or 48 hpi. Cells labelled with primary anti-ORF2 serum, and anti-rabbit AF488 secondary antibody (green). Nuclei were stained with DAPI (blue). Size bars represent 5µm.

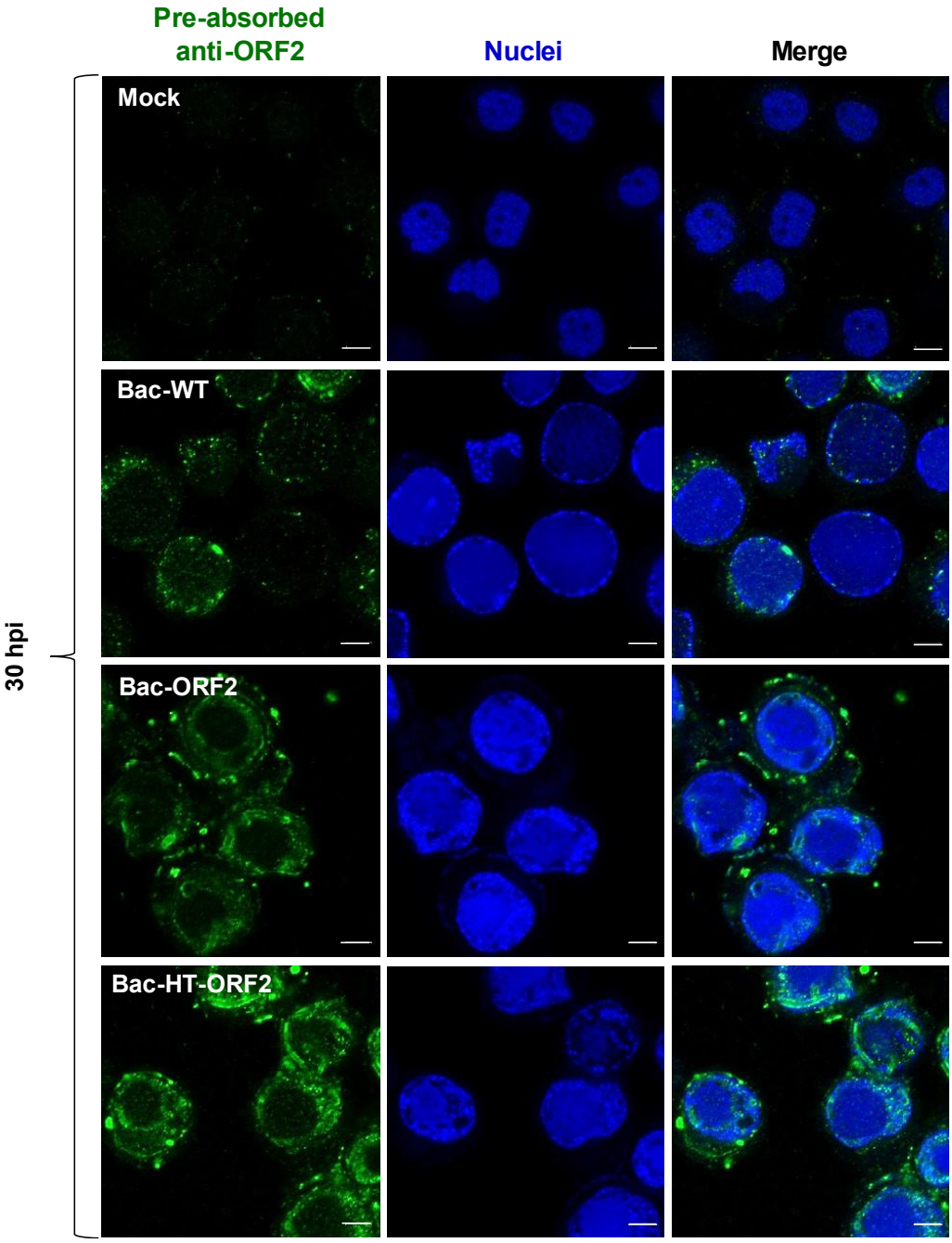


Fig. 3.14: Intracellular localisation of ORF2 and HT-ORF2 labelled with pre-absorbed anti-ORF2 serum. Sf9 cells were mock-infected or infected with Bac-WT, Bac-ORF2 or Bac-HT-ORF2. Cells were fixed at 30 hpi and labelled with primary pre-absorbed anti-ORF2 serum, and anti-rabbit AF488 secondary antibody (green). Nuclei were stained with DAPI (blue). Size bars represent 5µm.

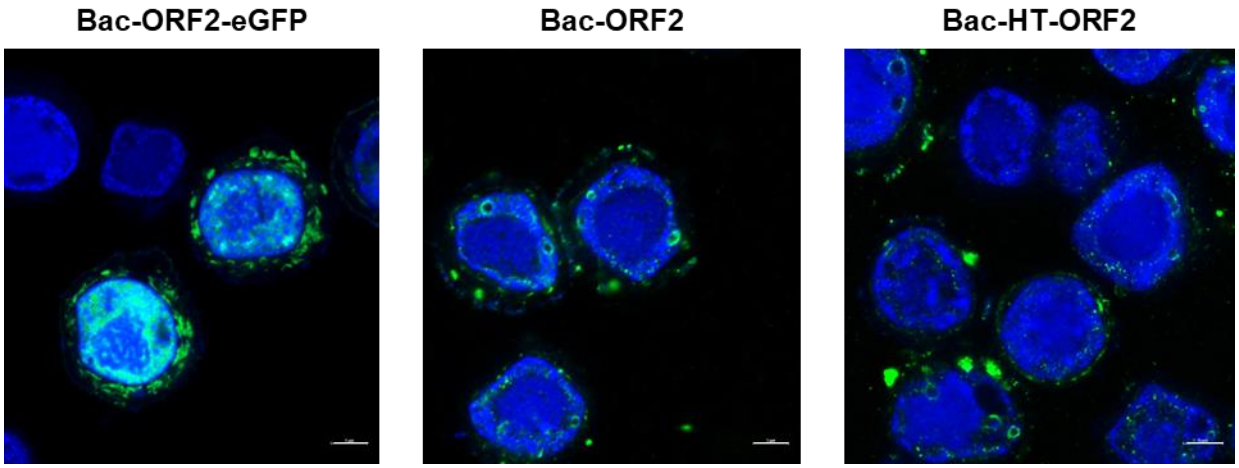


Fig. 3.15: Intracellular localisation comparison between ORF2-eGFP, ORF2 and HT-ORF2. Sf9 cells were infected with Bac-ORF2-eGFP, Bac-ORF2 or Bac-HT-ORF2. Cells were fixed at 30 hpi. Bac-ORF2-eGFP autofluorescence in green. Bac-ORF2 and Bac-HT-ORF2 labelled with primary pre-absorbed anti-ORF2 serum, and anti-rabbit AF488 secondary antibody also in green. Nuclei were stained with DAPI (blue). Size bars represent 5µm.

3.3.7 Transient mammalian expression of AHSV Seg-10 ORF2

To investigate the expression and localisation of the AHSV Seg-10 ORF2 protein in mammalian cells, to mimic the AHSV equine host, a transient mammalian expression system was used. Recombinant plasmids containing a CMV promoter would express ORF2 or fusion proteins when transfected into mammalian cells.

3.3.7.1 Recombinant plasmids for transfection

The expression vectors pcDNA3.1 and phCMV-Dream were used, both containing an ampicillin resistant gene and a CMV promoter for gene expression in mammalian cells (Fig. 3.16). The pcDNA3.1 plasmid additionally contains a T7 promoter, which is beneficial when transfecting into BSR-T7 cells which constitutively express T7 polymerase. The advantage of T7 RNA polymerase is the synthesis of large quantities of RNA, as it transcribes much faster than normal RNA polymerase. Thus, when the ORF2 protein is cloned downstream of a T7 promoter and transfected into BSR-T7 cells, large quantities of ORF2 mRNA will theoretically be transcribed quickly leading to higher ORF2 protein yields.

We aimed to create three recombinant expression plasmids. The plasmids pcDNA3.1-ORF2 and pcDNA3.1-ORF2-TST were produced by Genscript™, and following transfections would express native full-length ORF2 and ORF2 tagged with a C-terminal Twin-Strep-tag® (TST) respectively (Fig. 3.16 A & B). The Twin-Strep-tag® is a sequential arrangement of two Strep-tag®II (8 aa) sequences with an internal Ser-Gly linker region. A two amino acid Ser-Ala spacer between the ORF2 protein and the tag promotes the accessibility of the tag. The size of the total TST is 30 aa. For ORF2 to be tagged with TST, the stop codon from ORF2 was removed (Fig. 3.16 D). This construct will allow detection of ORF2-TST during confocal microscopy via a biotin-binding protein (streptavidin)

covalently attached to a fluorescent label. Streptavidin binds with high affinity to the peptide sequence of the Twin-Strep-tag®. This ORF2-TST protein could thus be detected microscopically to determine its intracellular localisation pattern. Lastly, we designed an ORF2-eGFP fusion protein which would express an autofluorescing eGFP-tagged ORF2. A Ser-Gly linker was included between ORF2 and eGFP to minimise the risk of interactions occurring between the two proteins (Fig. 3.16 C & D). The In-Fusion® cloning kit was utilised for generating this recombinant phCMV-ORF2-eGFP.

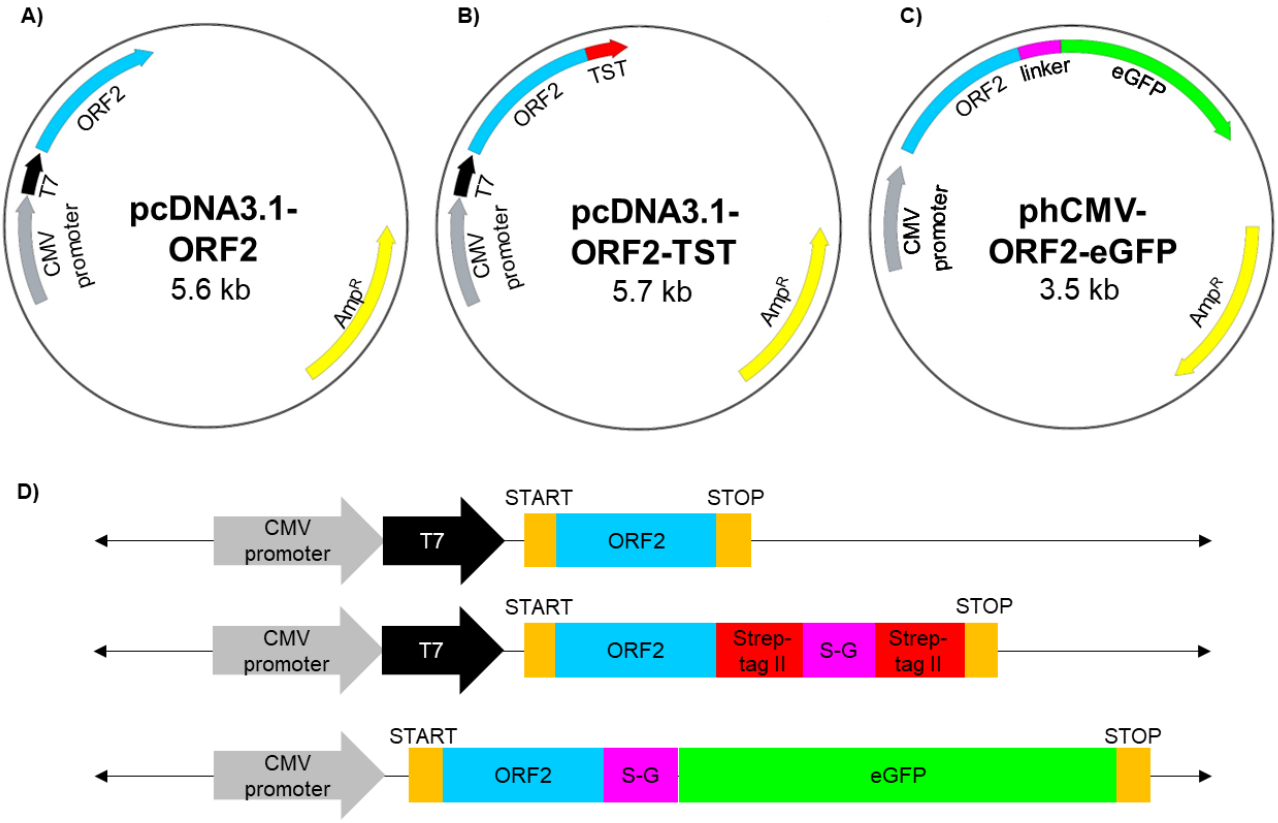


Fig. 3.16: The recombinant plasmid maps and cloning strategy for transient mammalian expression of AHSV ORF2. A) Full-length ORF2 cloned into pcDNA3.1 **B)** ORF2 lacking a stop codon cloned into pcDNA3.1-TST (Twin-strep-tag) downstream of the CMV and T7 promoters. **C)** ORF2 lacking a stop codon cloned into phCMV-eGFP. Each plasmid contains an ampicillin resistance gene (Amp^R). The size of each recombinant plasmid is given in kilobases (kb) **D)** Position of translation start and stop sites as well as position of any linkers and tags on recombinant proteins.

3.3.7.1.1 Generation of recombinant phCMV-ORF2-eGFP

The ORF2 gene was cloned into the phCMV plasmid downstream of a strong CMV promoter, and upstream of and in frame with eGFP to create an expression construct containing the gene of interest. Following transfections, this plasmid would express an autofluorescing ORF2-eGFP fusion protein that could be detected microscopically.

The first step was to linearise phCMV-NS4-eGFP (previously synthesised by Genscript™) through PCR amplification with primers designed to remove the NS4 gene directly upstream of the linker and create complementary overhangs to ORF2 excluding its stop codon (Table 3.3). The plasmid vector pSP72-S10 containing Seg-10 of AHSV-4 was used for PCR amplification of the ORF2 insert. The primers for insert amplification were designed to create overhangs complementary to the termini of the linearised phCMV-eGFP (Table 3.3) to allow joining of the ORF2 gene at the 5' end of eGFP. An agarose gel was used to confirm PCR amplification of both the phCMV-eGFP vector and ORF2 insert with necessary controls (results not shown). Gel purification was done for both vector and insert PCR amplicons, and their recovery was verified (Fig.3.17). The phCMV-eGFP vector without any insert was of the expected size of about 3.3 kb, and the ORF2 insert was also of the expected size of about 250 bp. The linearised phCMV-eGFP vector and ORF2 insert were now ready for In-fusion® cloning.

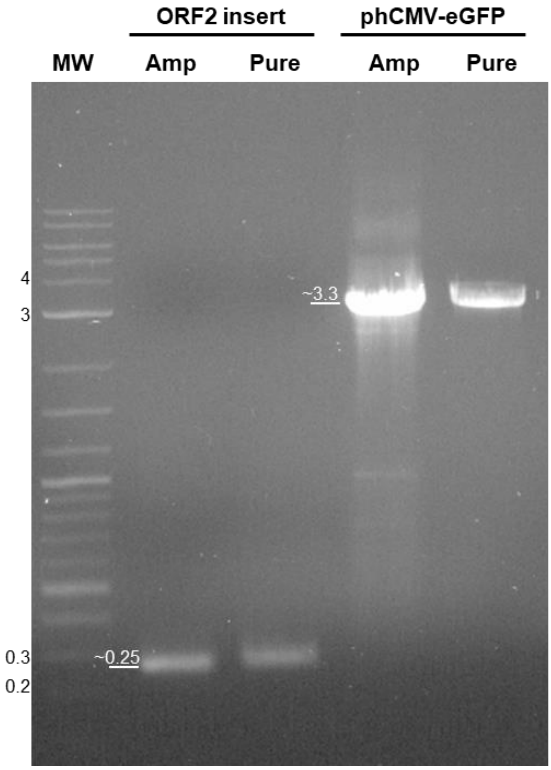


Fig. 3.17: A 1% agarose gel representing amplified (Amp) and purified (Pure) ORF2 insert and phCMV-eGFP. A molecular weight marker (MW) was included (kb).

An In-Fusion® cloning reaction was set up for phCMV-eGFP and ORF2 insert, and the cloning reaction used directly to transform Stellar™ competent cells and spread on ampicillin containing plates. The next day white colonies were observed on all plates except the negative control (non-transformed Stellar cells). Single colonies were picked and inoculated into a liquid culture and grown overnight at 37°C. Plasmids were isolated and PCR amplified with the ORF2 insert primers to confirm presence of ORF2 insert. Plasmid isolation and PCR was analysed by agarose gel electrophoresis (Fig. 3.18). Isolated putative recombinant phCMV-ORF2-eGFP plasmid DNA was of the expected size of about 3.5 kb, which is only slightly higher than non-recombinant phCMV-eGFP plasmid DNA

(Fig. 3.18 A). The PCR reaction confirmed the presence of the ORF2 insert. The negative control contained no DNA template, and the positive control contained the original pSP72-S10 DNA template. An ORF2 amplicon of the right size was successfully amplified from each putative recombinant and confirmed the presence of the ORF2 insert (Fig. 3.18 B).

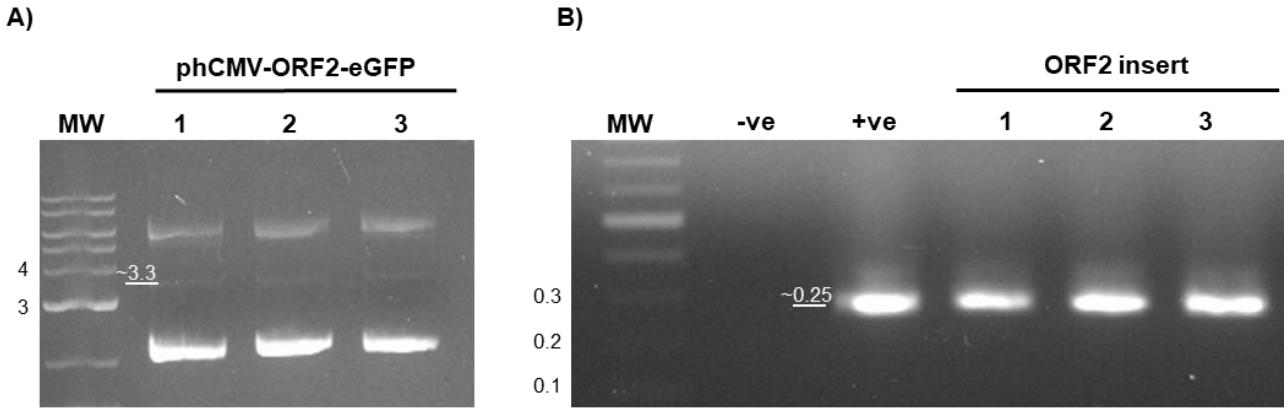


Fig. 3.18: 1% agarose gels representing representing recombinant phCMV-ORF2-eGFP plasmid DNA and PCR amplification of the cloned insert. A) phCMV-ORF2-eGFP plasmids were isolated from putative recombinant Stellar bacterial colonies (1, 2 and 3). **B)** ORF2 inserts were amplified from recombinant phCMV-ORF2-eGFP plasmid DNA samples (1,2 and 3) with ORF2 insert specific primers, with negative (-ve) and positive (+ve) controls. A molecular weight marker (MW) was included (kb).

Sanger sequencing was performed to further confirm the putative recombinants and ensure that ORF2 was inserted in frame with eGFP. The phCMV-ORF2-eGFP insert primers were used in the BigDye reaction (Table 3.3). Sanger sequencing results were aligned to the theoretical phCMV-ORF2-eGFP plasmid sequence using MAFFT. Full sequences are provided in the Appendix. All putative recombinants were confirmed to contain the ORF2 insert with the stop codon successfully removed and ORF2 cloned directly upstream of the linker. An example of Sanger sequencing results alignment is shown in Fig. 3.19. In conclusion, ORF2 was correctly cloned to produce a phCMV-ORF2-eGFP plasmid that could be used to produce recombinant ORF2-eGFP protein during transfection into mammalian cells.

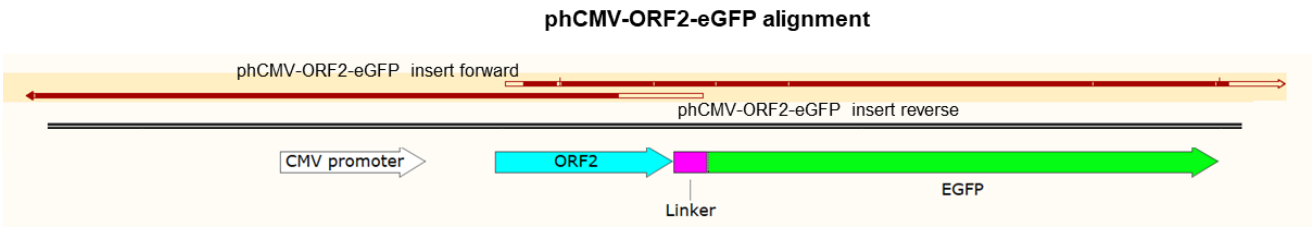


Fig. 3.19: Schematic representation of regions sequenced of recombinant phCMV-ORF2-eGFP plasmids, MAFFT sequence alignment was imported and visualised with SnapGene. Red bars represent the extent of sequences obtained from Sanger sequencing reactions. White spaces inside red bars represent mismatches in the sequences.

3.3.7.2 ORF2 shows low expression during transfections

A transfection protocol for BSR-T7 cells was previously investigated and optimised for transient expression from recombinant plasmids (BOUGHAN 2016). BSR-T7 were grown on coverslips were transfected with 400 ng of plasmid DNA. The plasmids used for transfection were phCMV-ORF2-eGFP, pcDNA3.1-ORF2 and pcDNA3.1-ORF2-TST. Control transfection were either mock transfected with no plasmid DNA or transfected with pCMV-eGFP or pCMV-T7-eGFP. Both express the autofluorescing eGFP protein, but pCMV-T7-eGFP expresses the eGFP gene under control of a T7 promoter. This provided an expression level control for pcDNA3.1 plasmids which also contain the T7 promoter.

In order to observe the expression in mammalian cells, BSR-T7 cells were transfected with each of the above and incubated for 24 hours before cells were fixed, labelled if required, and visualised. Proteins expressed from pCMV-eGFP, pCMV-T7-eGFP and phCMV-ORF2-eGFP would autofluoresce green. The mock, pcDNA3.1-ORF2 and pcDNA3.1-ORF2-TST transfected cells were labelled with primary anti-ORF2 serum and secondary anti-rabbit AF488 antibody. All cells were stained with DAPI. Cells were viewed by CLSM at 10X magnification to determine how many cells were successfully transfected (Fig. 3.20). Mock transfected control cells showed no autofluorescence or labelling (results not shown). The first interesting observation was that when cells were transfected with a plasmid containing the T7 promoter, a higher number of cells expressed fluorescence. This was seen by comparing pCMV-eGFP with pCMV-T7-eGFP, and comparing pcDNA3.1-ORF2/pcDNA3.1-ORF2-TST with phCMV-ORF2-eGFP. When the T7 promoter was present, the BSR-T7 cells expressed ORF2 protein in more cells than when T7 was absent. An explanation could be that when the T7 promoter is not present the expression level of ORF2 is too low to be observed, thus overall, less cells will fluoresce. However, transfections often vary due to many other reasons, including the presence or absence of a Kozak sequence as well as cell generation and viability. Secondly, when comparing transfections of plasmids encoding eGFP to those encoding ORF2 (native or tagged), the number of transfected cells expressing the target protein were much higher for eGFP than for ORF2. This was interesting since the same amount of plasmid DNA was used for each transfection, and all plasmids contained a CMV promoter. This suggested that the ORF2 protein or RNA transcript itself must thus be the reason for the low expression. A possible explanation might include a toxic effect of ORF2 expression on the cells, preventing protein production.

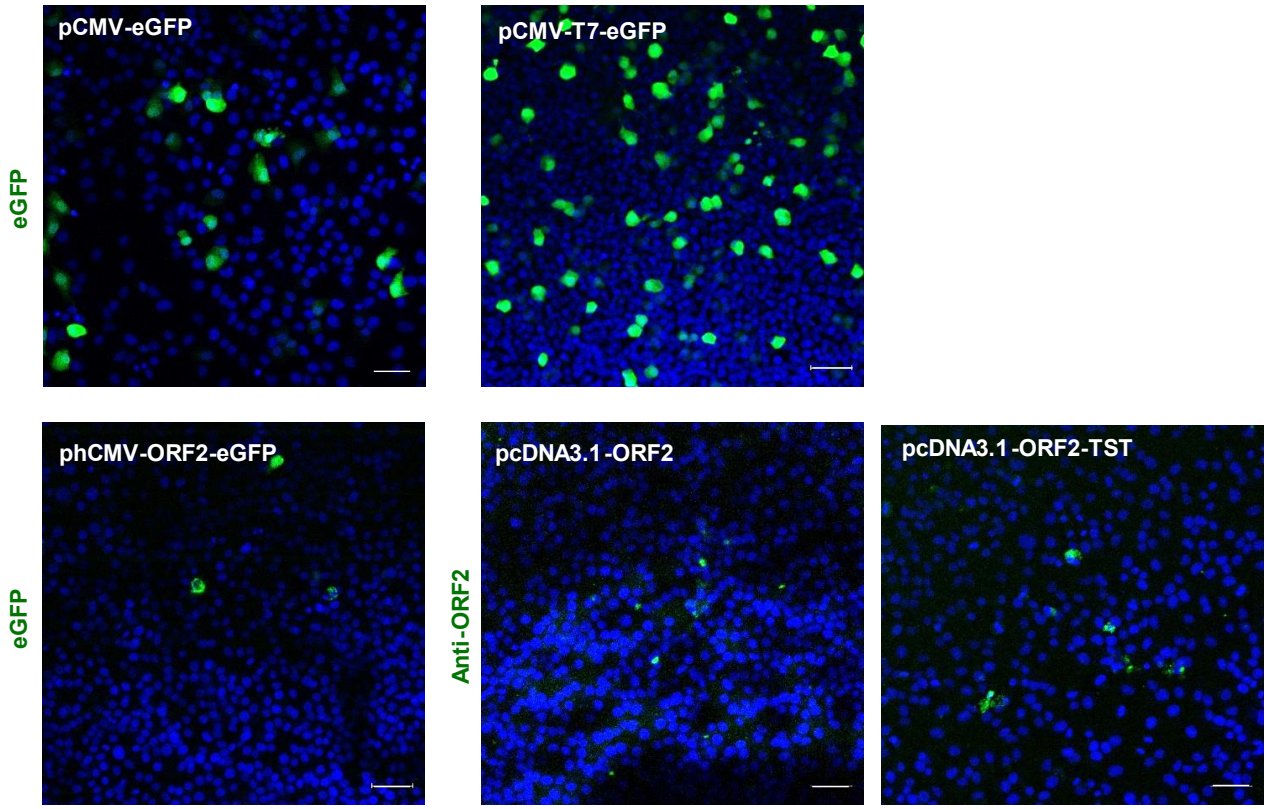


Fig. 3.20: Confocal microscopy images illustrating transfection efficiency. BSR-T7 cells were transfected with pCMV-T7-eGFP, pCMV-eGFP, pcDNA3.1-ORF2, pcDNA3.1-ORF2-TST or phCMV-ORF2-eGFP and fixed at 24 hpt. All eGFP proteins were detected by autofluorescence (green). Cells transfected with pcDNA3.1-ORF2 or pcDNA3.1-ORF2-TST were labelled with anti-ORF2 serum and anti-rabbit AF488 secondary antibody (green). Nuclei were stained with DAPI (blue). All images shown represent merged images. Size bars represent 50µm.

3.3.7.3 Detection of ORF2-eGFP shows a cytotoxic effect on transfected cells

BSR-T7 cells grown on coverslips were either mock transfected or transfected with pCMV-eGFP or phCMV-ORF2-eGFP, and cells fixed at different time points post transfection (Fig. 3.21). Mock transfected and pCMV-eGFP transfected cells were round and healthy at all time points. Mock cells transfected with no plasmid DNA did not show any background fluorescence. Cells expressing native eGFP from pCMV-eGFP plasmids showed green fluorescence homogenously distributed over the entire cell at all the time points. When ORF2 was fused to eGFP, a specific localisation pattern that changed over time was observed. At 16 hpt small round foci were detected in the cytoplasm, and cell nuclei still appear round and overall the cell appeared viable. At 24 hpt the fluorescence was also present in the nucleus, now showing foci in the cytoplasm and nucleus. The movement into the nucleus could be due to passive diffusion of the small ORF2 fusion protein. At 36 hpt transfected cells started to shrink and appeared less viable, with nuclei also appearing more compact. Lastly at 48 hpt cells containing any fluorescence were completely shrivelled. Due to the shrivelling, the exact localisation of ORF2 could not be determined. Cells surrounding the ORF2-eGFP expressing cells

that did not contain any fluorescence still appeared round and healthy. This suggests a specific cytotoxic effect resulting from ORF2 expression.

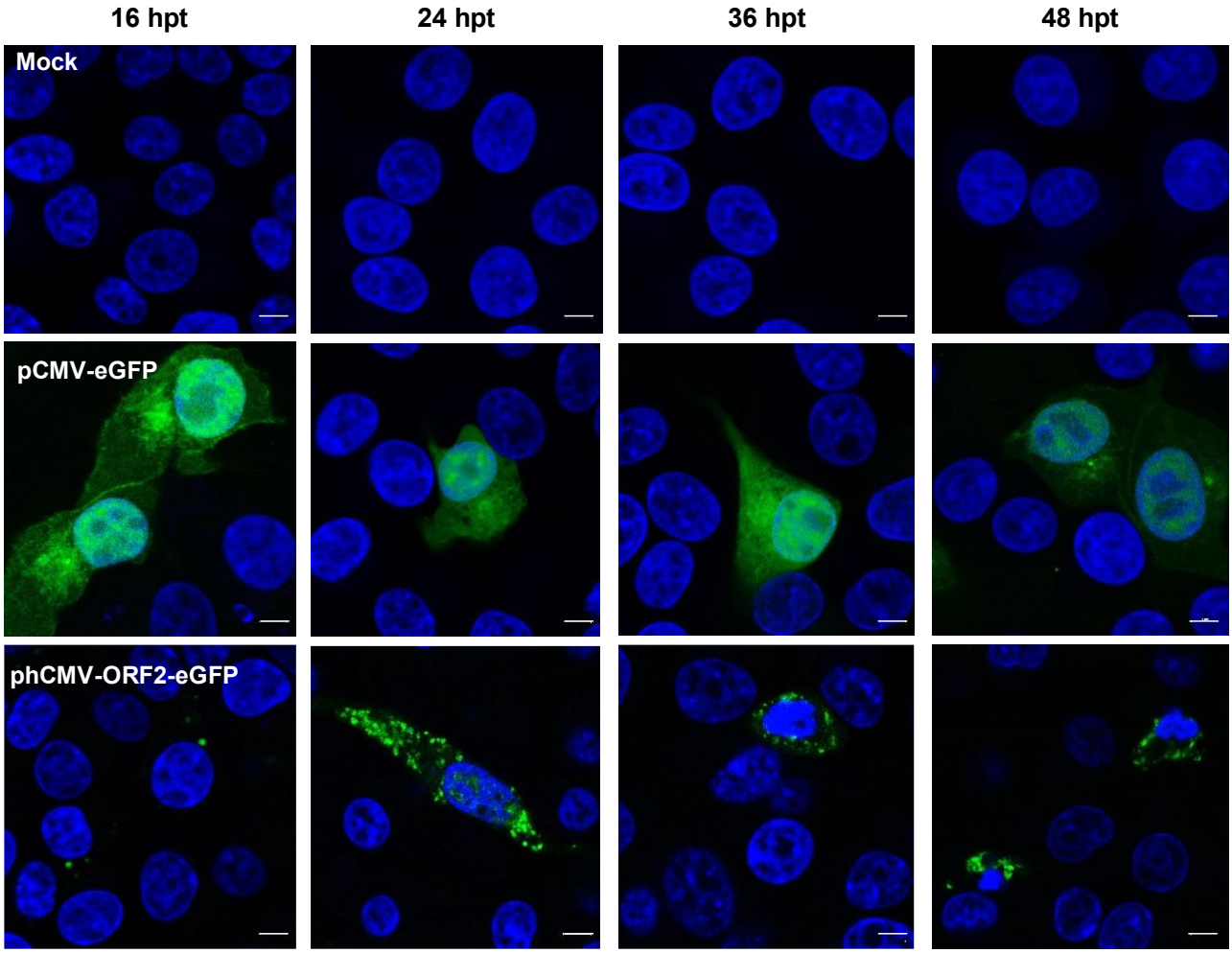


Fig. 3.21: Intracellular localisation of autofluorescing eGFP and ORF2-eGFP at different time points. Cells were mock transfected, or transfected with pCMV-eGFP or phCMV-ORF2-eGFP and fixed at 16, 24, 36 or 48 hpt. Autofluorescence of eGFP (green) was visualised by CLSM. Nuclei were stained with DAPI (blue). All images shown represent merged images. Size bars represent 5µm.

To again test our anti-ORF2 serum, BSR-T7 cells grown on coverslips were either mock transfected or transfected with pCMV-eGFP or phCMV-ORF2-eGFP. Cells were fixed at 24 hpt and labelled with primary anti-ORF2 serum and anti-rabbit AF555 secondary antibody (Fig. 3.22). Mock and pCMV-eGFP transfected cells were again round and viable, in both there was some red background over the entire slide which appeared as small red dots. Cells expressing ORF2-eGFP were shrivelled and appeared to be dying. Red fluoresence from anti-ORF2 labelling was observed in some cells containing ORF2-eGFP, but not all. Once again as seen previously in insect cells there was no clear correlation between green fluorescence from the ORF2eGFP and the red fluorescence representing anti-ORF2 antibody binding. Since the anti-ORF2 serum antibody might contain limited antibodies specific for ORF2, the low red fluorescence might not represent all ORF2 protein. The dying shrivelled cells could also cause proteins to aggregate, impacting on the binding of anti-ORF2 serum

to its respective epitopes. This would explain why the red fluorescence was not always present in cells containing green ORF2-eGFP. Also, the red fluorescnce was extremely dim compared to the bright green from eGFP which made accurate detection without overexposure (which causes background) very tricky. In conclusion, due to low transfection efficiency or other reasons for low ORF2-eGFP expression, it could still not be concluded confidently that the anti-ORF2 serum showed specific/representative binding to the putative ORF2 protein.

3.3.7.4 Detection of transiently expressed ORF2 and ORF2-TST with anti-ORF2 serum

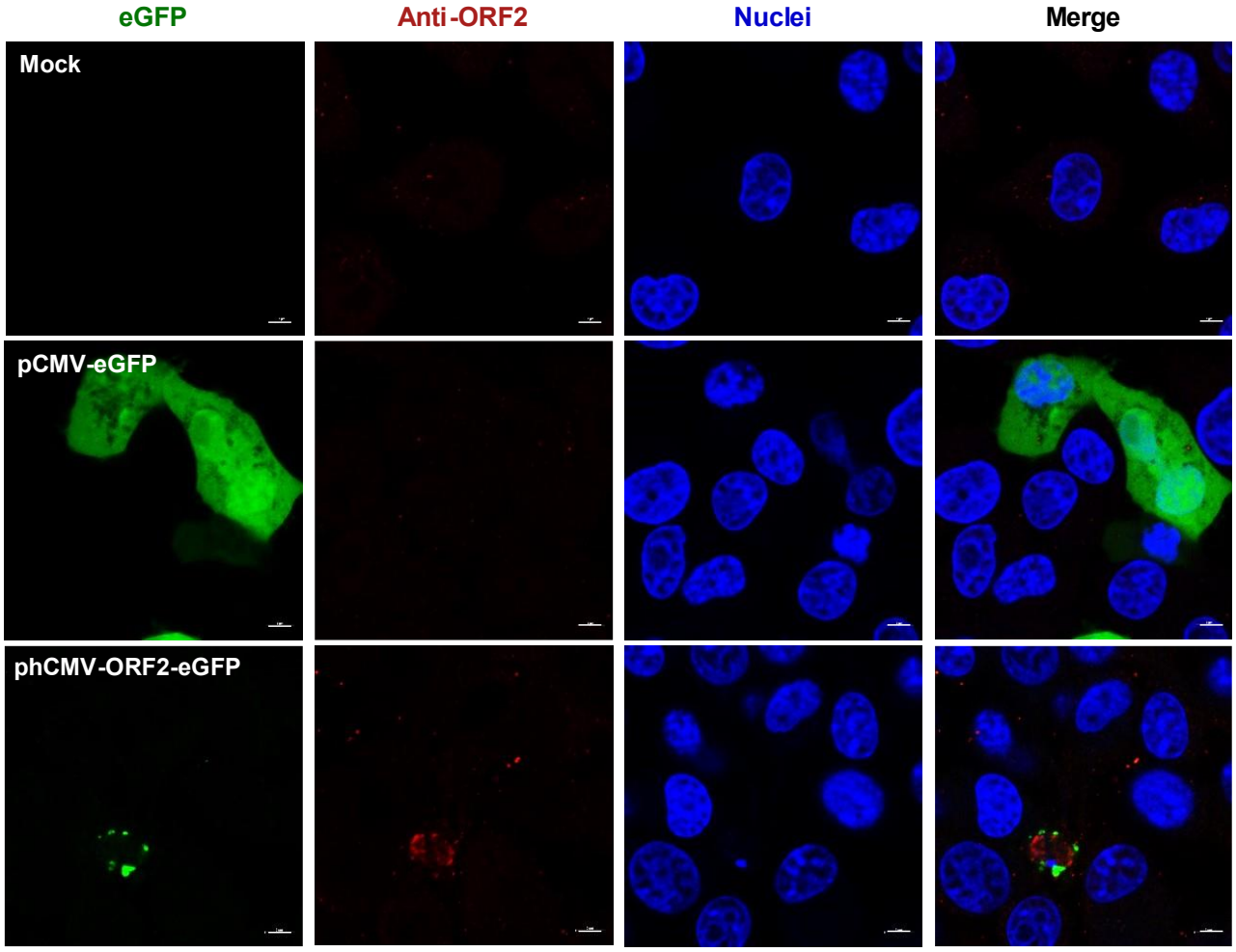


Fig. 3.22: Intracellular localisation of ORF2-eGFP also labelled with anti-ORF2 serum. Cells were mock transfected, or transfected with pCMV-eGFP or phCMV-ORF2-eGFP and fixed at 24 hpt. Autofluorescence of eGFP (green) viewed by CLSM. Cells were labelled with anti-ORF2 serum and anti-rabbit AF555 secondary antibody (red). Nuclei were stained with DAPI (blue). Size bars represent 5µm.

To visualise the localisation of native or TST-tagged ORF2 in mammalian cells, BSR-T7 cells grown on coverslips were either mock transfected or transfected with pcDNA3.1-ORF2 or pcDNA3.1-ORF2-TST. Cells were fixed and labelled with primary anti-ORF2 serum and anti-rabbit AF488 secondary antibody as well as DAPI stained different times post transfection (Fig. 3.23). Mock transfected cells were round and healthy at all time points. All samples showed very low levels of background noise of diffuse light green speckles. The labelling visualised in cells transfected with both pcDNA3.1

recombinant plasmids showed similar localisation patterns to that of phCMV-ORF2-eGFP. The presence of both ORF2 and ORF2-TST could be detected as small foci in the cytoplasm at 16 hpt. Cells assumed to be expressing these proteins (i.e with green labelling) started to shrivel as early as 24 hpt, whilst ORF2-eGFP only caused shrivelling from 36 hpt. Thus, the cytotoxicity of native ORF2 seemed to be higher than that of ORF2-eGFP. Due to the nuclear condensation, potential nuclear fluorescence of ORF2 could not be observed. At 48 hpt it was almost impossible to observe the shrivelled-up nucleus, probably due to its possible degradation.

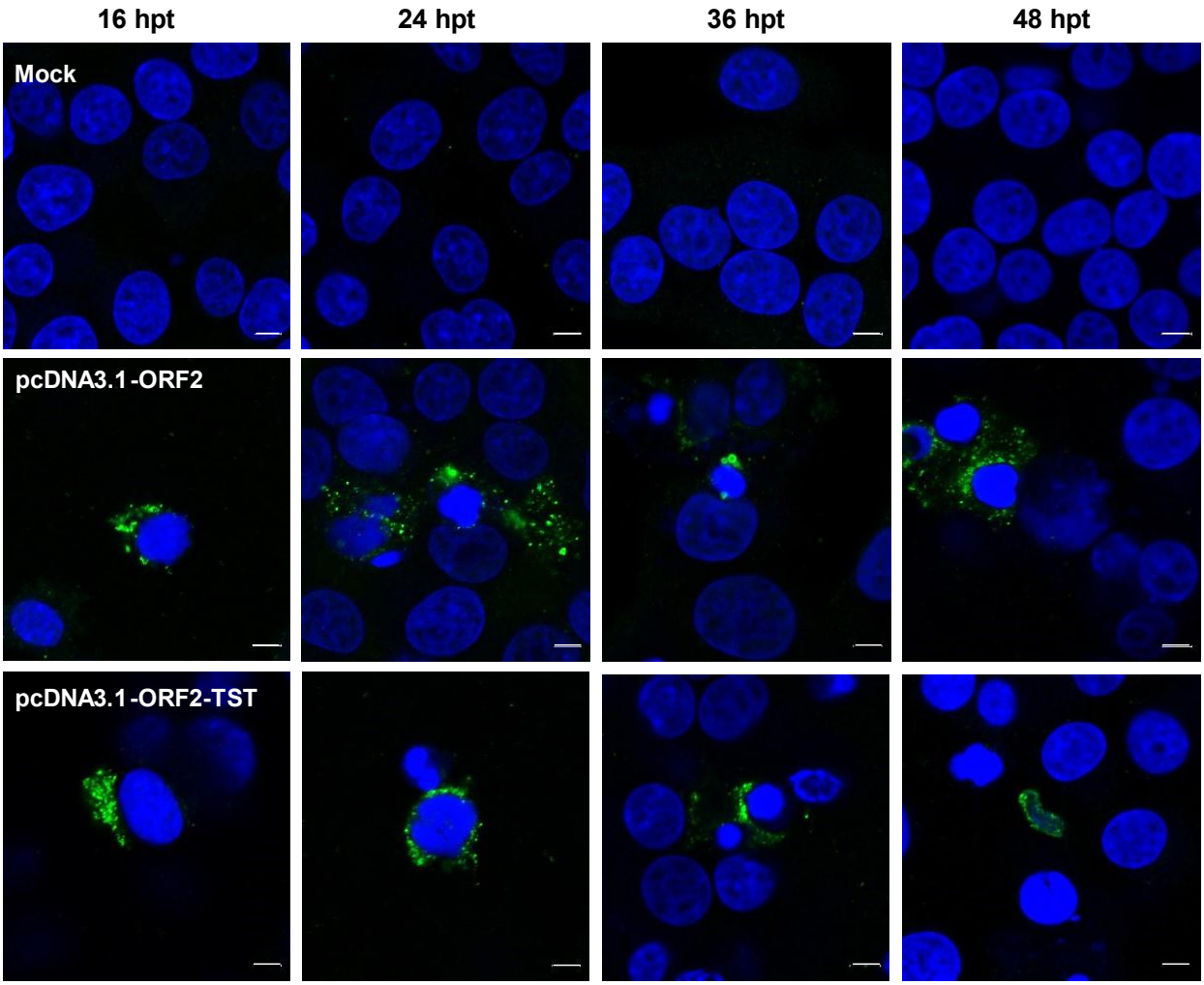


Fig. 3.23: Intracellular localisation of ORF2 and ORF2-TST at different time points detected by anti-ORF2 labelling. Cells were mock transfected, or transfected with pcDNA3.1-ORF2 or pcDNA3.1-ORF2-TSTand fixed at 16, 24, 36 or 48 hpt. Cells were labelled with anti-ORF2 serum and anti-rabbit AF488 secondary antibody (green). Nuclei were stained with DAPI (blue). Images shown are merged images. Size bars represent 5µm.

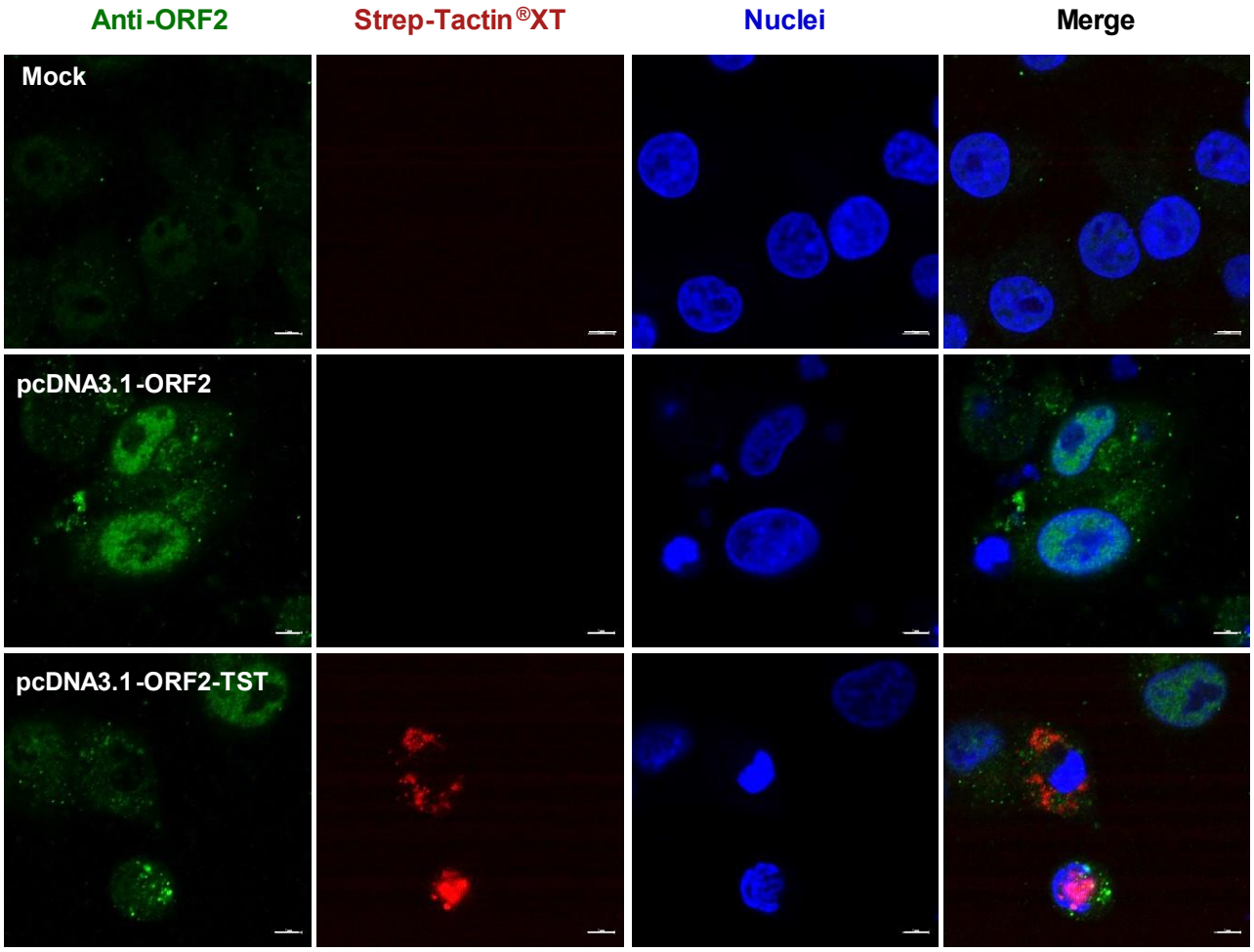


Fig. 3.24: Intracellular localisation of ORF2 and ORF2-TST detected by dual labelling. Cells were mock transfected, or transfected with pcDNA3.1-ORF2 or pcDNA3.1-ORF2-TST and fixed at 24 hpt. Cells were dual labelled with anti-ORF2 serum and anti-rabbit AF488 secondary antibody (green) and Strep-Tactin®XT (red). Nuclei were stained with DAPI (blue). Size bars represent 5µm.

Due to the lack of clear overlap of either baculovirus expressed or transiently expressed ORF2-eGFP green fluorescence with labelling in red with anti-ORF2 serum, an additional experiment was performed to test the specificity of the serum. For this, cells harbouring pcDNA3.1-ORF2-TST were co-labelled with Strep-Tactin®XT. BSR-T7 cells grown on coverslips were either mock transfected or transfected with pcDNA3.1-ORF2-TST and fixed at 24 hpt. Cells were labelled with primary anti-ORF2 serum and anti-rabbit AF488 secondary antibody, and co-labelled with a TST specific label, Strep-Tactin®XT DY-549 conjugate. The nucleus was stained with DAPI (Fig. 3.24). The green fluorescence indicative of the putative expression of ORF2 was present in both the pcDNA3.1-ORF2 control and pcDNA3.1-ORF2-TST transfections. The mock did not contain any red or green fluorescence, and all labelled samples still had low green background noise. There was indeed red fluorescence visible in pcDNA3.1-ORF2-TST cells, and as expected this red fluorescence was not detected in the untagged pcDNA3.1-ORF2 control. This means that any red fluorescence must be the binding of the TST in the expressed ORF2-TST to the Strep-Tactin®XT label. As seen in Fig.

3.24 the pcDNA3.1-ORF2-TST transfected cell contained the green fluorescence indicative of ORF2 also contained the red labelling indicative of a TST tagged ORF2. Unfortunately, there was still no clear overlap of the two. Also, some cells were detected during confocal microscopy to have red labelling without the presence of green fluorescence. This could be due the epitope specific for the anti-ORF2 serum not being exposed during transient mammalian expression. TST will still be accurately labelled, if the ORF2 protein conformation has no effect on the tag's ability to bind to the Strep-Tactin®XT DY-549 conjugate.

In conclusion, transient mammalian expression of the ORF2 native protein and fusion proteins provide an initial putative localisation pattern of ORF2 in mammalian cells. ORF2 also seemed to have a toxic effect when expressed, causing cells to shrivel. Only a few cells expressed the ORF2 protein due to low transfection efficiency or other reasons for low expression. Thus, the small sample size of only a few fluorescent cells made it difficult to conclude the specificity of the anti-ORF2 serum to the ORF2 protein.

3.3.8 Bacterial expression of AHSV Seg-10 ORF2

Due to the challenges encountered with the anti-ORF2 serum produced from the baculoviral expressed antigen, it was decided to attempt ORF2 protein synthesis via bacterial induction. The aim was to express the protein for purification and subsequent new antibody production. The StabyExpress™ expression system was used. This is a T7 expression system, consisting of a pStaby vector containing a T7 bacteriophage promoter region upstream of a cloning site. The gene encoding the T7 RNA polymerase is inserted into the chromosome of the expression bacteria, which in this case was either BL21(DE3)pLysS (BL21) *E. coli* or Cys21 *E. coli*. This T7 RNA polymerase gene is under the control of a lacUV5 promoter region, and works analogous to the lac operon. During normal conditions the gene is repressed by the lac repressor which is encoded by both the bacterial chromosome and the pStaby vector. For induction of expression, medium is supplemented with isopropyl-β-D-thiogalactoside (IPTG) which removes the repressor from the lac operator sequence and allows the expression of T7 RNA polymerase. T7 RNA polymerase binds to the T7 promoter on the pStaby plasmid and produces large amounts of target mRNA and subsequently lots of target protein. This allow for high protein yields, to be utilised in downstream applications.

3.3.8.1 Production of recombinant of pStaby plasmids

In order to express the ORF2 protein for purification purposes, recombinant pUC57 plasmids were produced by Genscript which contained a bacterial codon-optimised AHSV Seg-10 ORF2 gene linked in frame at the 3' end to either a Strep-tag®II (S) or Twin-Strep-tag® (TST) (Appendix). Strep-tag®II is a short peptide tag with negligible effect on the fusion protein. It is a chemically balanced peptide with an additional serine and alanine spacer (10 aa). The spacer promotes the accessibility of the tag. The Twin-Strep-tag® is a sequential arrangement of two Strep-tag®II sequences with an internal linker region (30 aa). These tags make it possible for purification on a Strep-Tactin®XT

column. The ORF2 linked to either tag was cloned into the pStaby insertion site through In-Fusion® cloning for bacterial expression (Fig. 3.25).

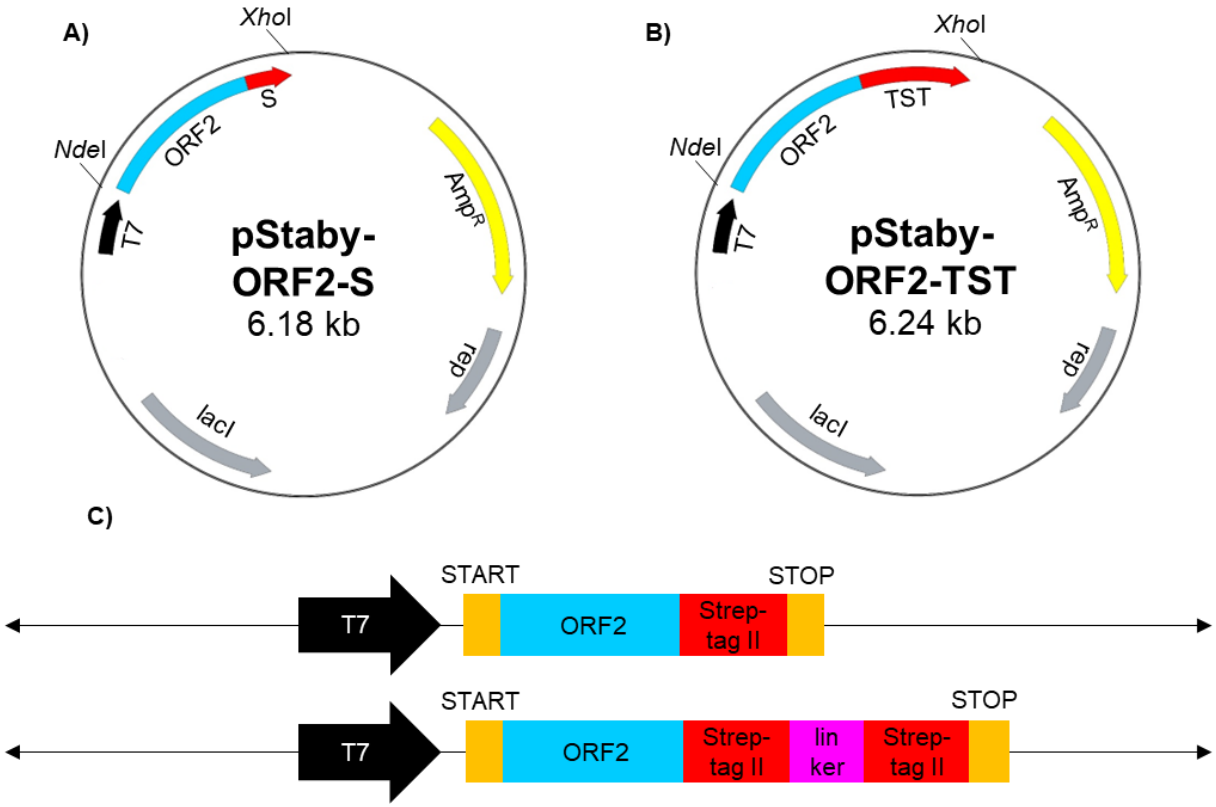


Fig. 3.25: The pStaby expression plasmids and cloning strategy for bacterial expression of AHSV ORF2. ORF2 tagged A) Strep-tag II (S) or **B)** Twin-strep-tag (TST) cloned into the pStaby plasmid, respectively. Each plasmid contains a T7 promoter, ampicillin resistance gene (Amp^R), lacI and repressor (rep) genes as well as a *Nde*I and *Xho*I restriction enzyme site. The size of each recombinant plasmid is given in kilobases (kb). **C)** Position of translation start and stop sites as well as position of any linkers and tags of the gene construct encoding the recombinant protein.

3.3.8.1.1 Linearisation of pStaby and amplification of ORF2 inserts

Plasmid stocks of pStaby containing an insert were double restriction enzyme digested. *Nde*I and *Xho*I flank the insertion site of the pStaby vector making it easy to remove any inserted gene. Linear pStaby without any inset is roughly 5.9 kb. Double digestion of a pStaby plasmid stock with *Nde*I and *Xho*I gave linear pStaby of the expected size and an excised insert of about 1.2 kb (Fig. 3.26 A). The final result was linear pStaby with the insertion site available.

Next, the pUC plasmids containing codon optimised ORF2 tagged to either S or TST were used as template for PCR amplification with gene specific primers which flank the start of ORF2 and the end of either tag respectively (Table 3.3). The inserts will now be referred to as ORF2-S and ORF2-TST. These primers were also designed to produce complementary overhangs to the linear pStaby termini, while reconstructing the *Xho*I and *Nde*I restriction enzyme sites. The size of ORF2-S and ORF2-TST is about 280 bp and 340 bp respectively. PCR amplification of pUC-ORF2-S and pUC-ORF2-TST with these primers resulted in amplicons of the expected size. The amplicons were a bit

larger (300 bp and 360 bp) than the actual gene, probably due to the fact that the primers include a region specific to the pStaby vector which is roughly 20 bp. This provides an extension to each gene, hence the larger size of the amplicon (Fig. 3.26 B). Negative controls containing no DNA did not amplify, but primer dimers were observed at the bottom in each lane. Amplification thus resulted in the target inserts required for cloning. Both restriction enzyme digested pStaby and PCR amplified tagged ORF2 inserts were column purified and purification verified (Fig. 3.26 C). Pure linear pStaby vector and ORF2-S and ORF2-TST inserts were now ready for In-Fusion® cloning.

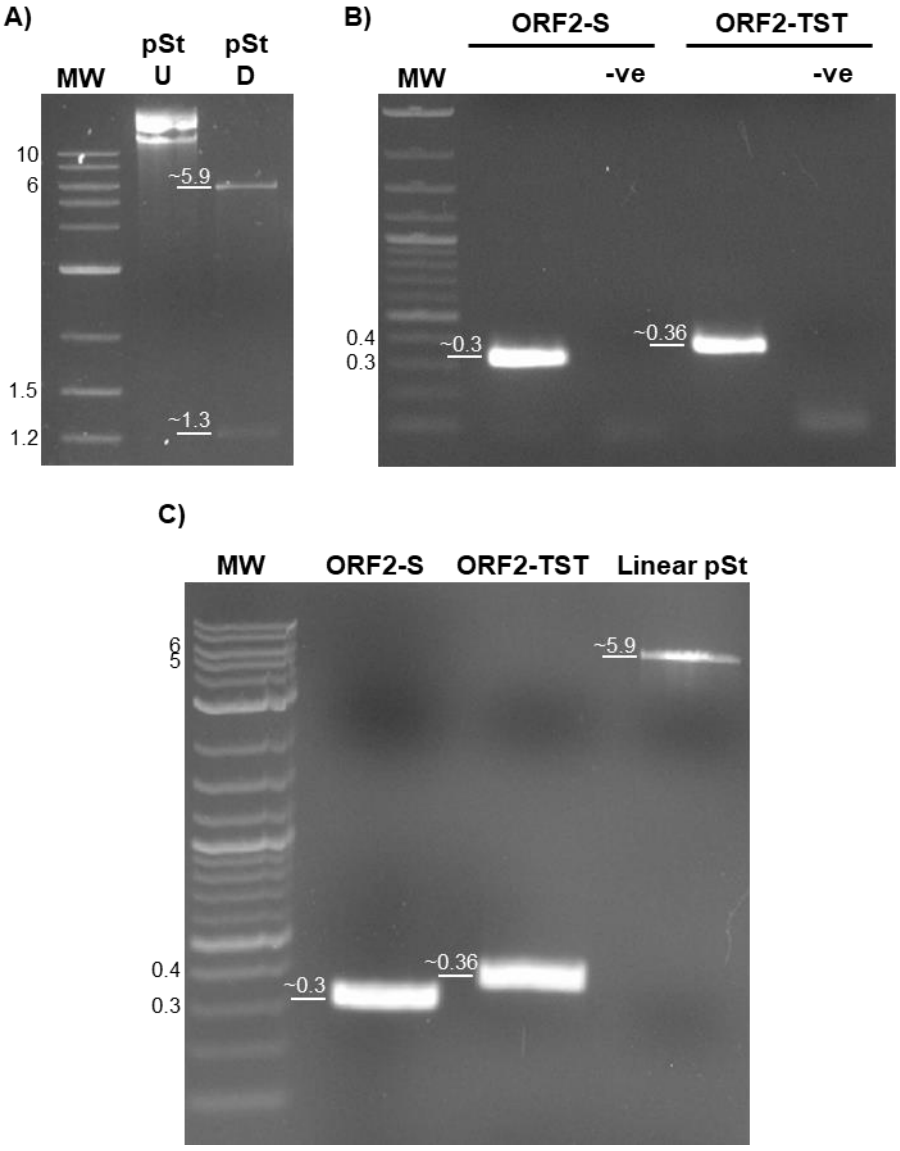


Fig. 3.26: 1% Agarose gels representing RE digested pStaby and insert amplicon DNA as well as their purified products. A) pStaby (pSt U) was double digested (pSt D). **B)** PCR amplification of ORF2-S and ORF2-TST. **C)** Column purified ORF2 inserts and linear pStaby. A molecular weight marker (MW) was included (kb).

3.3.8.1.2 In-Fusion cloning of ORF2-S/TST into pStaby

Pure linear pStaby and either pure ORF2-S or pure ORF2-TST were added together to an In-Fusion® cloning mixture. The mixtures were directly transformed into competent Stellar cells and spread over LB plates containing ampicillin. The next day white colonies were observed on all plates except the negative control (non-transformed Stellar cells). Single colonies were picked and inoculated into a liquid culture and grown overnight at 37°C. Putative recombinant pStaby plasmids were isolated from four colonies each.

3.3.8.1.3 Confirmation of recombinant pStaby-ORF2-S and TST

To confirm that the colonies were indeed recombinant, a double restriction enzyme digestion was performed with *Xho*I and *Nde*I (Fig. 3.27 A). Digestion of a normal wild-type pStaby plasmid stock (which does not contain any gene in its insertion site) resulted in only one band. corresponding to the size of linear pStaby (about 5.9 kb). All pStaby-ORF2-S putative recombinants excised a band of the correct size for ORF2-S, 300 bp. The first recombinant also contained some undigested supercoiled or circular plasmid DNA. The digestion of the pStaby-ORF2-TST recombinants also resulted in 2 bands, with the top one corresponding to the 5.9 kb linear pStaby and the bottom one to the ORF2-TST insert, 340 bp. In conclusion the ORF2-S and ORF2-TST inserts were present in all putative recombinant colonies respectively (Fig. 3.27 A). To further confirm recombinants, plasmid

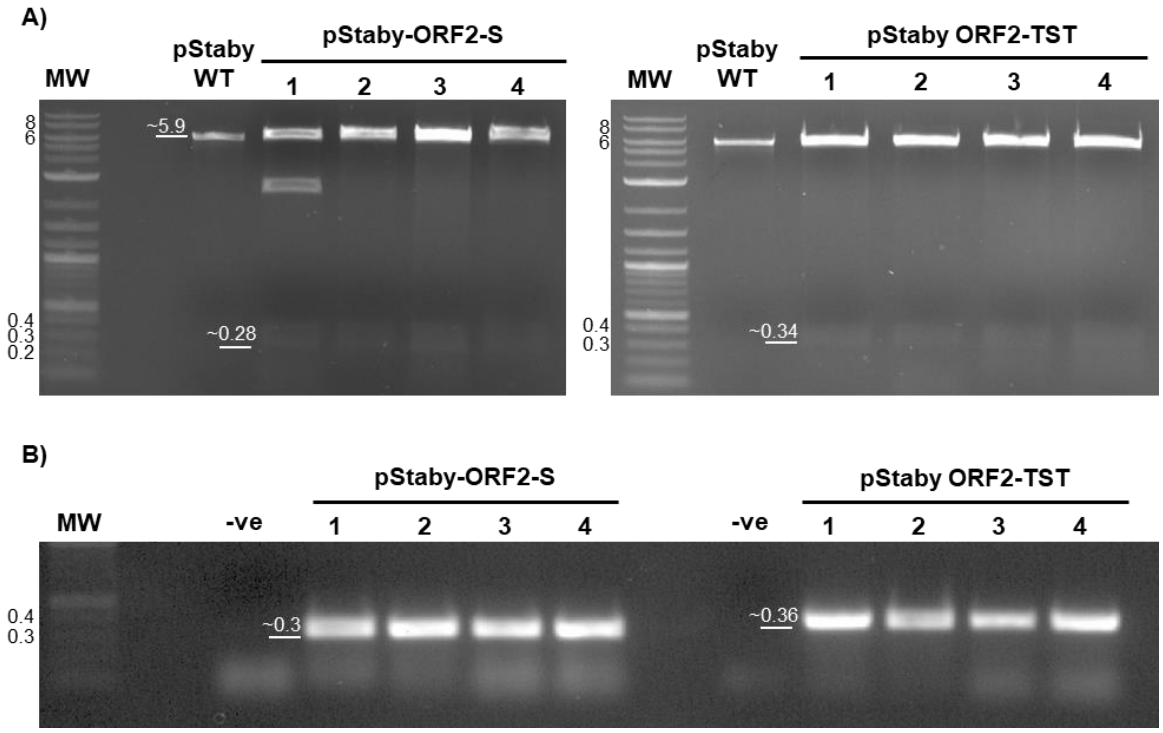


Fig. 3.27: 1% agarose gels representing RE digestions and PCR amplification of recombinant pStaby plasmids. A) pStaby wild-type and putative recombinants (1 - 4) of pStaby-ORF2-S and pStaby-ORF2-TST were double digested with *Xho*I and *Nde*I. **B)** PCR amplification of pStaby-ORF2-S and pStaby-ORF2-TST putative recombinants (1 - 4) and negative controls (-ve). A molecular weight marker (MW) was included (kb).

DNA was also PCR amplified with the previously used ORF2-S and ORF2-TST gene specific primers (Fig. 3.27 B). All recombinant plasmids of both pStaby-ORF2-S and pStaby-ORF2-TST resulted in amplicons of the expected sizes., with some primer dimers present.

Sanger sequencing was performed to further verity the recombinants and ensure the ORF2-S and ORF2-TST were inserted into pStaby correctly. The pStaby-ORF2-S/TST forward and reverse insert primers were used in the BigDye reaction (Table 3.3). Sanger sequencing results were aligned using MAFFT to the theoretical pStaby-ORF2-S and pSTaby-ORF2-TST plasmid sequences. Full sequences are provided in the Appendix. All recombinants were confirmed to contain the correct ORF2-S or ORF2-TST inserts, respectively. An example of Sanger sequencing results alignment is shown in Fig. 3.28. In conclusion, pStaby-ORF2-S and pStaby-ORF2-TST could now be used to transform BL21 or Cys21 *E. coli* for bacterial induction and expression of tagged ORF2.

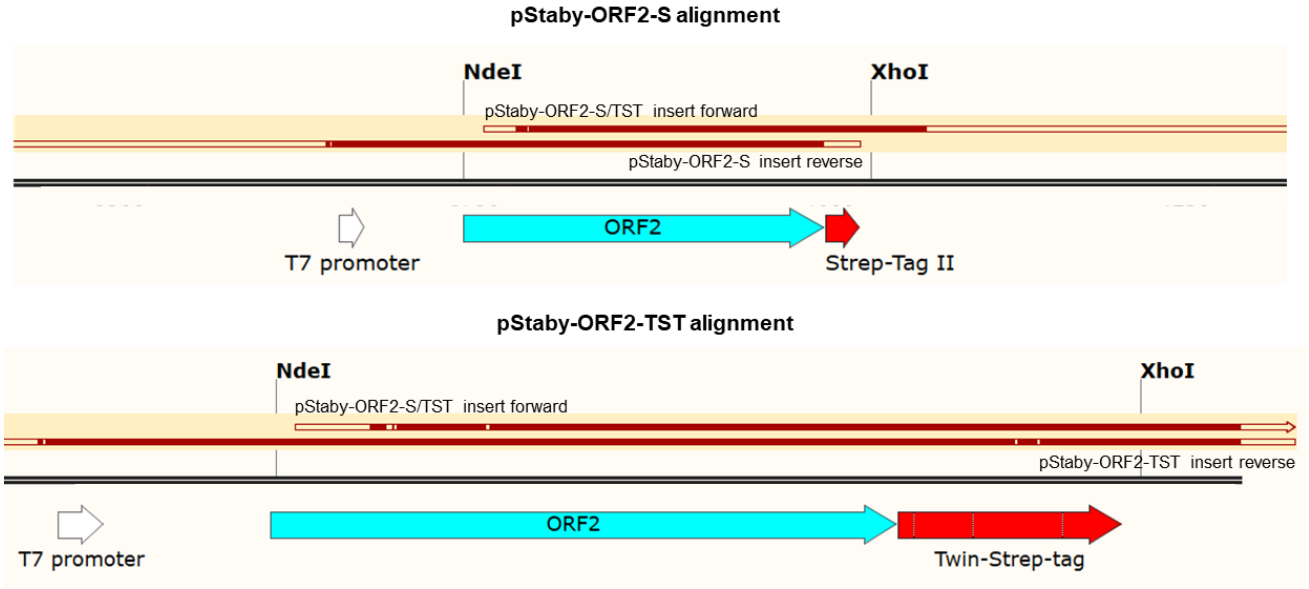


Fig. 3.28: Schematic representation of regions sequenced of pStaby-ORF2-S and pStaby-ORF2-TST plasmids, MAFFT sequence alignments were imported and visualised with SnapGene. Red bars represent the extent of sequences obtained from Sanger sequencing reactions. White spaces inside red bars represent mismatches in the sequences.

3.3.8.2 Expression of ORF2 seems toxic to induced bacterial cells

The relative sizes of the two fusion proteins is presented in Fig. 3.29 A. The sizes predicted by SnapGene® Viewer for ORF2-S is about 12 kDa and for ORF2-TST is 14 kDa. To express ORF2-S and ORF2-TST, purified recombinant plasmids pStaby-ORF2-S and pStaby-ORF2-TST were transformed into competent BL21 cells and protein expression was induced for 2 h using 1 mM IPTG at 37°C. Whole cell lysates were prepared using BugBuster® and lysonase, and the soluble and insoluble fractions were separated by centrifugation. Equal amounts of the uninduced and induced samples for soluble and insoluble fractions were compared using SDS-PAGE followed by Coomassie staining (Fig. 3.29 B). No unique bands of the expected sizes for ORF2-S or ORF2-TST could be detected in soluble samples following induction. However, ORF2-S and ORF2-TST were

clearly visible in the insoluble fractions of induced samples. Absence of bands at the corresponding positions in the uninduced fractions confirmed that both target proteins were expressed. If the soluble induced fractions contained small amounts of ORF2-S and ORF2-TST proteins these were at barely detectable levels, which would not be sufficient for purification purposes.

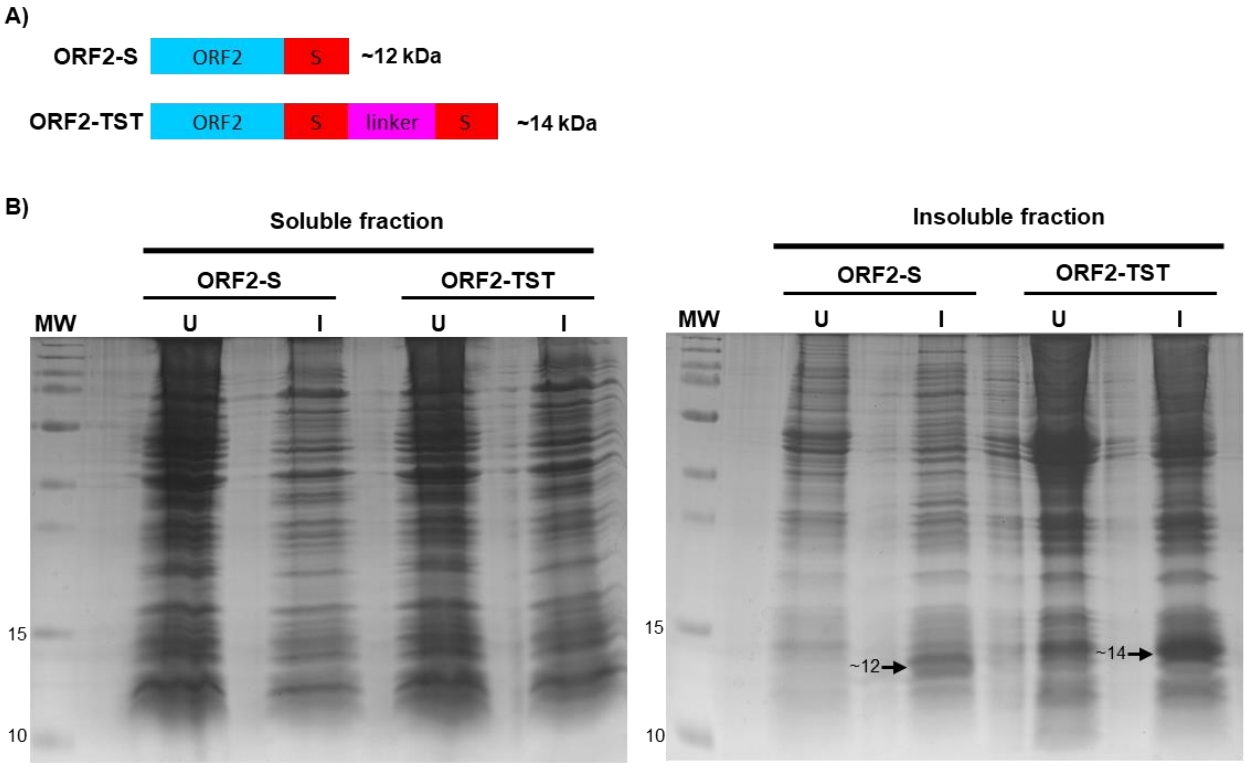


Fig. 3.29: Bacterial expression of ORF2 by induction of BL21 cells with 1mM IPTG at 37°C for 2 hours. **A)** Schematic representation of ORF2-S and ORF2-TST fusion protein sizes. **B)** Soluble (left) and insoluble (right) fractions of uninduced (U) and induced (I) pStaby-ORF2-S or pStaby-ORF2-TST transformed BL21 cells. Fractions were separated by 15% SDS PAGE and stained with Coomassie Blue. The sizes of molecular weight markers (MW) are indicated on the left of the gel in kDa.

The formation of insoluble aggregates of target protein can be the result of the induction at 37°C, when bacterial cells are fast growing and producing lots of protein. In an attempt to prevent this, induction was done with 1 mM IPTG at 16°C overnight. Induction at lower temperatures when bacteria are growing slower, could result in more gradual production of proteins. Unfortunately, no ORF2-TST or ORF2-S could be detected in any fraction of the bacterial lysates following overnight induction (results not shown) and it was concluded that 16°C was too cold for ORF2 protein expression. The optimal temperature for T7 RNA polymerase is between 37°C and 56°C. Furthermore, high IPTG concentrations could be toxic to the bacterial cells.

A last induction experiment was attempted to achieve expression without aggregation. Protein expression was performed by inducing BL21 or Cys21 cells overnight using either 0.1 mM or 0.5 mM IPTG at 25°C. Whole cell lysates were separated by centrifugation into soluble and insoluble fractions, and induced and uninduced samples compared by SDS-PAGE and Coomassie staining (Fig. 3.30). The first observation that could be made was that the BL21 bacterial soluble cell fractions

contained much more protein and also many more proteins of different sizes, when compared to the Cys21 bacterial soluble fractions. This made it very difficult to detect any band specific for the ORF2 protein in the soluble fractions of induced BL21 lysates. The insoluble fraction from both BL21 and Cys21 cells looked similar. The StabyExpress™ system was designed to be used with Cys21 cells, which could explain this difference between BL21 and Cys21cultures. The second observation made during harvesting of the bacterial cell pellets was that the pellets from uninduced bacterial cultures had a much higher wet pellet weight than induced bacterial cells pellets. It appeared that when the ORF2 protein was produced it had a negative effect on the growth of the bacteria, resulting in a lower number of bacterial cells in the induced cell pellets. The ORF2-TST and ORF2-S proteins were visible in the induced lysates of Cys21, but in reduced quantities compared to when previously induced at 37°C for 2 hours. This further sustained the fact that when the ORF2 protein was present in bacterial cells for more than 2 hours, it influenced the growth of the cells consequently leading to less cells being able to express any ORF2. The insoluble fractions did not contain any ORF2-S or ORF2-TST in amounts that could be easily detected. There might be some still present but in very low amounts. In conclusion the induction at 25°C overnight did produce soluble protein, but produced less protein overall than when induced at 37°C.

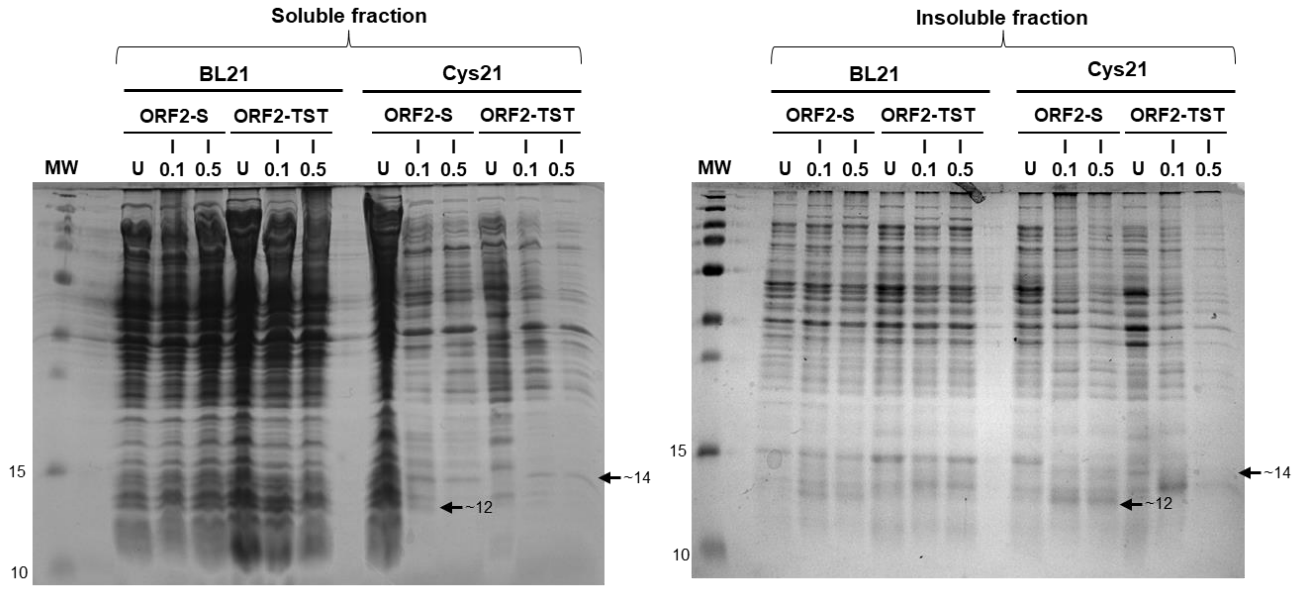


Fig. 3.30: Bacterial expression of ORF2 by induction of BL21 and Cys21 cells with 0.1 mM or 0.5 mM IPTG at 25°C overnight. Soluble (left) and insoluble (right) fractions of uninduced (U) and induced (I) pStaby-ORF2-S or pStaby-ORF2-TST transformed BL21 or Cys21 cells. Fractions were separated by 15% SDS PAGE and stained with Coomassie Blue. The sizes of molecular weight markers (MW) are indicated on the left of the gel in kDa.

3.3.8.3 ORF2-TST/S could not be purified with the Strep-Tactin®XT column

To purify the soluble ORF2 protein, an induction of Cys21 cells was again done at 25°C and grown overnight. The cells were lysed with Bugbuster®, lysonase and extra benzonase was added. This cell lysate was separated by centrifugation and the supernatant (soluble fraction) was run through a

Strep-Tactin®XT column (Fig. 3.31). Both the Strep-tag®II and Twin-Strep-tag® bind with high affinity to streptavidin in the column. This enables the purification of higher protein yields compared to other purification systems, such as the previously used histidine-tag. Moreover, the neutral pI of Strep-Tactin®XT minimizes non-specific protein or nucleic acid binding. The first clear observation that could be made after purification was that if the ORF2 protein was present, it was eluting together with many other proteins, mainly in fraction E3. Also, because of the low expression of the ORF2 protein due to the previously observed toxicity when induced cells were grown overnight, it was impossible to determine whether pure ORF2-S or ORF2-TST was indeed present amongst all the eluted proteins. The arrows in the figure represent the size of band expected for ORF2-S and ORF2-TST. There were indeed proteins of these respective sizes, but still too little and impure. The induction protocol was followed exactly as for the samples depicted in the previous figure, and thus should contain the ORF2-S and ORF2-TST proteins respectively in the lysates. When the eluates were read on a spectrophotometer, they showed extremely high absorbance in the 260 nm range which is characteristic of nucleic acids. This was of concern, since lysonase and extra benzonase were added to the cell lysates to degrade any nucleic acids. We hypothesised that the ORF2-TST/S protein might bind to nucleic acids, protecting them from degradation, and in turn other proteins that also bind nucleic acids were bound to ORF2-TST/S and co-purified and eluted together. In conclusion the purification was unsuccessful. To obtain sufficient quantities of bacterially expressed ORF2 for immunisation, the best option would be to express the protein at 37°C for 2 hours and excise it from a gel as was done for the baculovirus expressed ORF2.

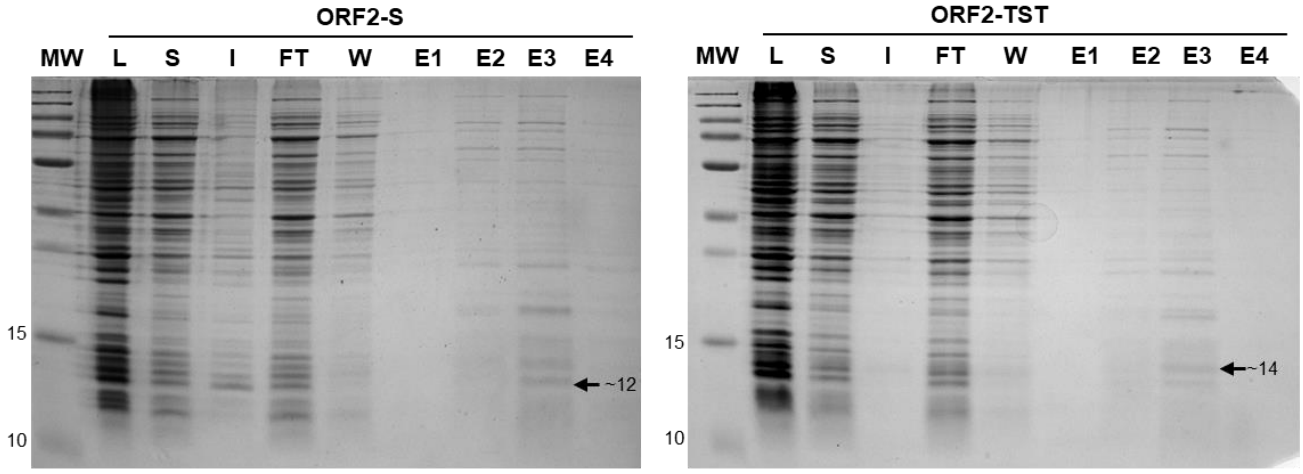


Fig. 3.31: Column purification of soluble ORF2-S and ORF2-TST from induced Cys21 cells with 0.1 mM IPTG at 25°C overnight. The total lysate (L), the soluble (S) and insoluble (I) fractions of the induced cultures were collected. The soluble fraction was run through a Strep-Tactin®XT column and the flowthrough (FT) and wash (W) samples were collected respectively. Bound protein was eluted 4 times (E1, E2, E3 and E4). All samples were separated by 15% SDS PAGE and stained with Coomassie Blue. The sizes of molecular weight markers (MW) are indicated on the left of the gel in kDa.

3.3.9 The ORF2 protein is mainly hydrophilic

Due to the difficulties with insolubility of tagged ORF2 proteins during baculoviral expression of HT-ORF2 and bacterial expressed ORF2-S and ORF2-TST, the hydrophobicity of all peptides were determined. CLC Main Workbench was used to determine the hydrophobicity value at each amino acid site of AHSV-4 ORF2, HT-ORF2, ORF2-S and ORF2-TST (Fig. 3.32). The ORF2 protein (blue) did not show any difference in hydrophobicity when tagged and so overlapped exactly, thus only one ORF2 line is shown in the figure. Hydrophobicity scales are values that define the relative hydrophobicity or hydrophilicity of all amino acid residues. The more positive the value, the more hydrophobic are the amino acids located in that region of the protein. Overall ORF2 was not very hydrophobic but rather largely hydrophilic. As the tags did not have any effect on its theoretical hydrophobicity, there must be a different explanation for its insolubility during expression of tagged ORF2.

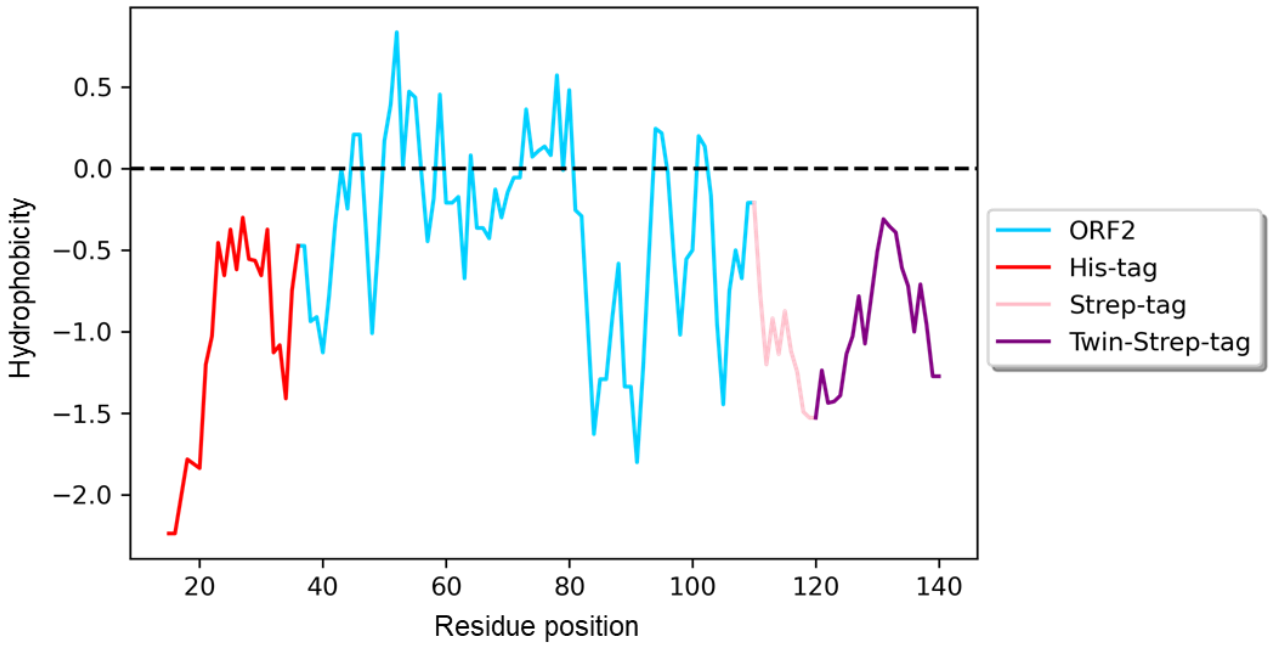


Fig. 3.32: Hydrophobicity plot of ORF2 protein (blue) as well as the tags used during this study (HT-red, S-pink, TST-purple). The hydrophobicity was determined using CLC Main workbench and the plot was created using matplotlib.

3.3.10 Seg-10 ORF2 expression was detected during AHSV infections

Due to time constraints, gel purification of bacterially expressed ORF2 and subsequent immunisation and bleeding of rabbits to obtain new ORF2 specific serum was not done. Instead the existing anti-ORF2 serum, produced from the baculoviral expressed ORF2, was used for labelling of cells following AHSV infection for detection of putative ORF2 expression.

BSR-T7 cells grown on coverslips in a 24-well were mock infected or infected with AHSV-1, AHSV-4 or AHSV-8 and cells were fixed 30 hpi. Some primary and secondary antibody optimisations were required to achieve the best results with low background noise. Different antibody dilutions, as well

as pre-absorption of serum with BSR-T7 cell lysates, were tested. Pre-absorbed serum did not have any effect on the background noise. Cells were finally labelled with dilutions of 1:50 for primary anti-ORF2 serum and 1:500 for secondary anti-rabbit AF 488 and nuclei stained with DAPI prior to visualisation by CLSM (Fig. 3.33). Even after optimisations, there was still some low intensity green background signal across all samples which could not be eliminated. The mock infected cell nuclei were round and overall cells were healthy and did not show any specific localised green fluorescence. All AHSV infected cell nuclei appeared deformed and cells less viable as compared to mock infected cells. As indicated with white arrows, small green foci or larger spherical structures were detected by anti-ORF2 labelling in all AHSV infected samples. These structures showed a more regular spherical shape than the irregular foci detected during transient mammalian or baculovirus expression. In AHSV-1 infected cells, small to larger spherical structures were detected in the cytoplasm. The same was observed for AHSV-4, but there were fewer labelled foci compared to AHSV-1. Interestingly, during AHSV-8 infections the labelled foci were not only present in the cytoplasm but also appeared in the nucleus. The AHSV non-structural protein NS2 forms large, globular perinuclear cytoplasmic structures called viral inclusion bodies (VIBs) in infected cells. The morphology of these structures detected by anti-ORF2 labelling appeared similar to VIBs (UITENWEERDE *et al.* 1995). This could unfortunately not be verified, as the only anti-NS2 primary antibody at our disposal was also produced in rabbits like the anti-ORF2 serum. This precluded simultaneous dual labelling of NS2 and ORF2, as differentially conjugated secondary antibodies could not be used.

To further confirm that the protein detected by the antibody was indeed present only in AHSV infected cells, co-labelling was done with anti-NS1. BSR-T7 cells were either mock infected or infected with AHSV-1, AHSV-4 or AHSV-8 and fixed 30 hpi. Cells were labelled with 1:50 primary anti-ORF2 serum and 1:500 secondary anti-rabbit AF 488 and co-labelled with primary mouse anti-NS1 and anti-mouse secondary AF 594, and nuclei stained with DAPI. NS1 is known to be a highly expressed AHSV protein and is perfect as a positive control to confirm infected cells. As seen in Fig. 3.34, mock infected cells again were healthy and showed no red or green fluorescence. The same spherical structures were detected with anti-ORF2 serum as in the previous experiment. The localisation pattern was also the same throughout for the three strains. Additionally, NS1 was successfully detected as red tubular structures sequestered in the cytoplasm near or around the nucleus. Thus, any BSR cells containing red fluorescence could be concluded to be infected, and any cells that did not have red fluorescence were not infected. With this additional labelling it could clearly be seen that the green anti-ORF2 labelled structures were only present in infected cells also expressing NS1. Uninfected cells did not contain any green labelling.

Lastly, a primary cell line of equine umbilical cord cells were obtained from Prof Christiaan Potgieter at Deltamune. These cells were either mock infected, or infected with either AHSV-1 or AHSV-4. Cells were fixed at 30 hpi and labelled as for Fig. 3.34. The spherical green ORF2 structures were

again observed only in confirmed infected cells expressing NS1. Fewer cells were infected than in BSR-T7 cell cultures (Fig. 3.35). In conclusion, this was the first putative detection of the novel ORF2 protein being expressed during AHSV infections.

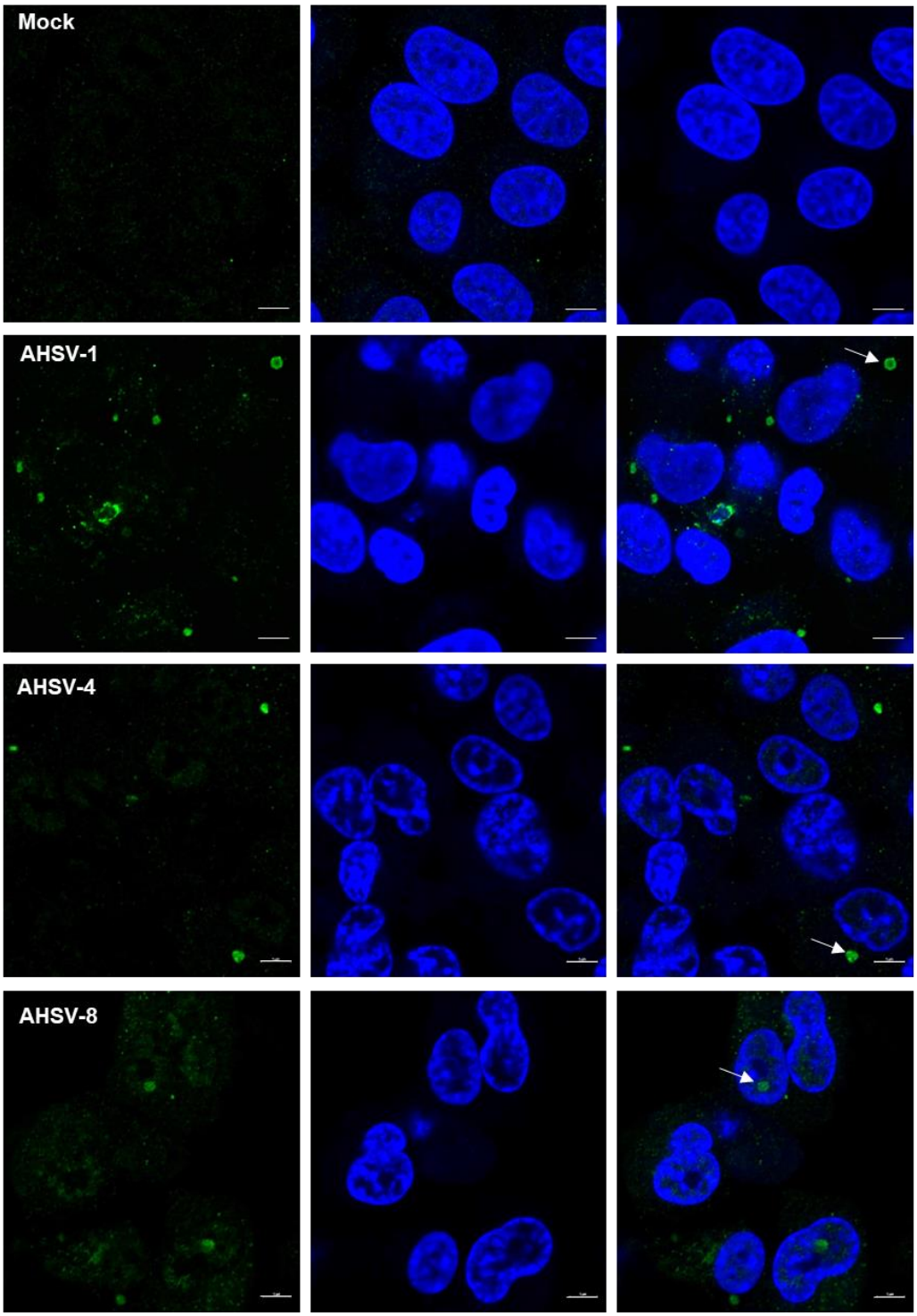


Fig. 3.33: Intracellular localisation of putative ORF2 expressed during AHSV infections of BSR-T7 cells labelled with anti-ORF2 serum. Cells were mock infected, or infected with AHSV-1, AHSV-4 or AHSV-8. Cells were fixed at 30 hpi and labelled with anti-ORF2 serum and anti-rabbit AF488 secondary antibody (green). Nuclei were stained with DAPI (blue). Arrows indicate examples of labelling of ORF2. Size bars represent 5µm.

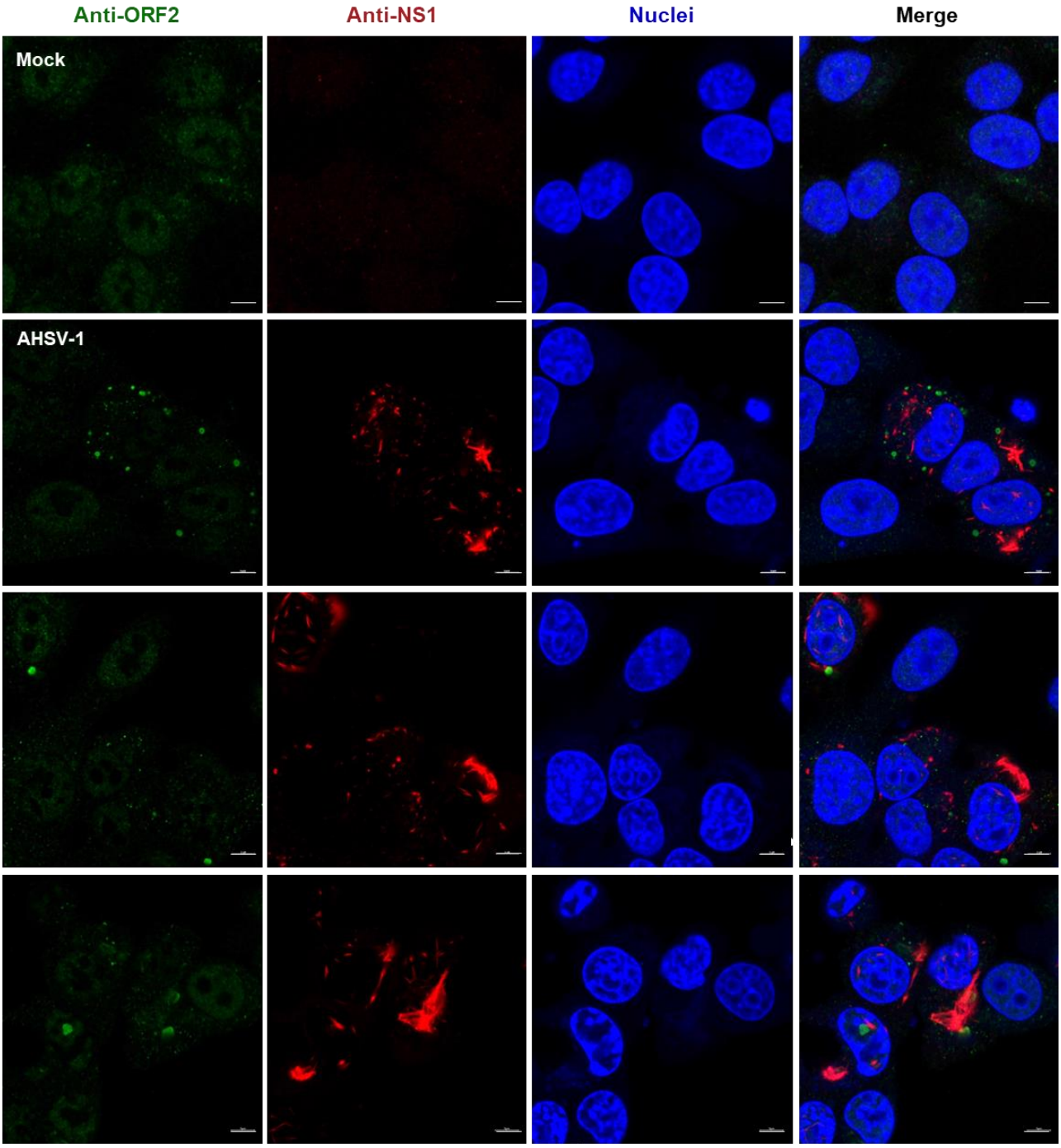


Fig. 3.34: Intracellular localisation of putative ORF2 during AHSV infections of BSR-T7 co-labelled with anti-ORF2 and anti-NS1 serums. Cells were mock infected, or infected with either AHSV-1, AHSV-4 or AHSV-8. Cells were fixed at 30 hpi, dual labelled with anti-ORF2 primary and anti-rabbit AF 488 secondary antibodies (green) plus mouse anti-NS1 primary and anti-mouse AF 594 secondary antibodies (red) and viewed by CLSM. Nuclei were stained with DAPI (blue). Size bars represent 5µm.

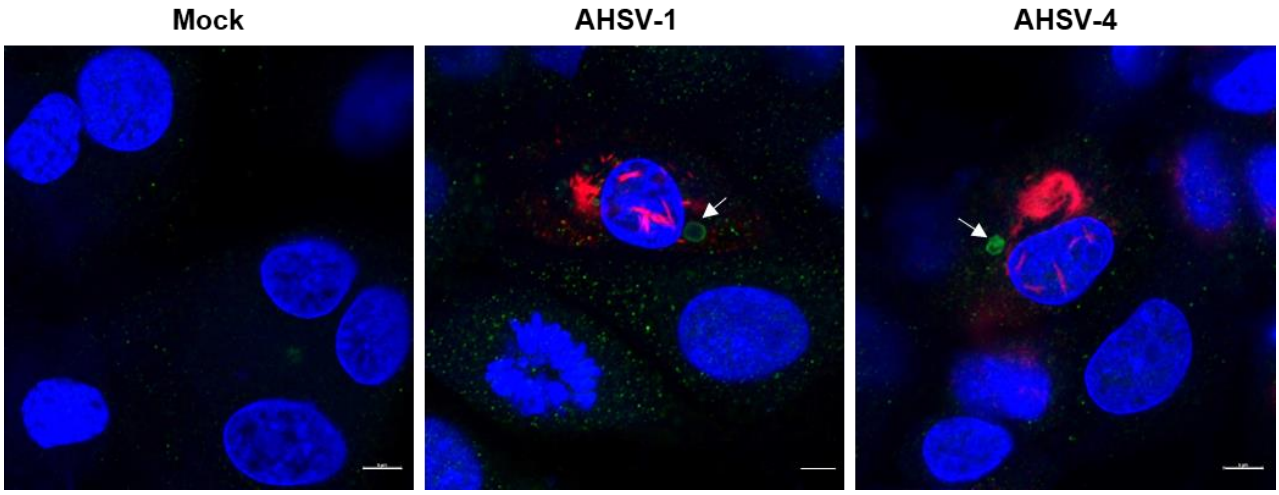


Fig. 3.35: Intracellular localisation of putative ORF2 during AHSV infections of umbilical cord cells co-labelled with anti-ORF2 and anti-NS1 serums. Cells were mock infected, or infected with AHSV-1 or AHSV-4. Cells were fixed at 30 hpi, dual labelled with anti-ORF2 primary and anti-rabbit AF 488 secondary antibodies (green) plus mouse anti-NS1 primary and anti-mouse AF 594 secondary antibodies (red) and viewed by CLSM. Merged images are shown. Arrows indicate the labelling of ORF2. Nuclei were stained with DAPI (blue). Size bars represent 5µm.

3.4 Discussion

The aim of this chapter was to express the protein theoretically encoded by AHSV Seg-10 ORF2, and to characterise its intracellular localisation patterns. This was done in insect and mammalian cells, and ORF2 was additionally expressed in bacteria. Lastly, the potential expression of ORF2 during AHS viral infection was investigated.

In the *Reoviridae* family, several viruses have been characterised as having one or two genome segments which contain more than one ORF per mRNA. In the genus *Orthoreovirus,* segment 1 of avian reovirus encodes three different proteins from overlapping ORFs, each having a different function (RACINE AND DUNCAN 2010). In rotavirus, segment 11 (the smallest genome segment) encodes major protein NSP5 as well as the 12 kDa non-structural protein NSP6 from an internal +1 additional open reading frame (MATTION *et al.* 1991). In AHSV and BTV, Seg-9 encodes major protein VP6, and an additional smaller protein from a +1 open reading frame characterised as NS4 (RATINIER *et al.* 2016; BOUGHAN *et al.* 2020; WALL *et al.* 2021). Seg-10 in BTV and AHSV also contains two overlapping reading frames, encoding major protein NS3 and an internal ORF2. In BTV, the Seg-10 ORF2 encodes a predicted protein of 50-59 aa in size, which when transiently expressed from plasmids localised to the nucleoli of transfected cells. Its expression during BTV infection could however not be confirmed (STEWART *et al.* 2015). In this study we characterised the previously unresearched AHSV ORF2. As this ORF is conserved in different phylogenetically related viruses, it can be an indication of possibly being evolutionary advantageous for some orbiviruses (STEWART *et al.* 2015).

Two systems were investigated which could provide high yields of a suitable product for affinity purification of ORF2, and for its subsequent use as an immunogen in production of anti-ORF2 antibodies. The expression of a polyhistidine tagged ORF2 (HT-ORF2) was done utilising the baculovirus expression system. Unfortunately, problems arose with the insolubility of the HT-ORF2 protein. ORF2 is mainly hydrophilic, and predicted to soluble. Theoretically, the histidine tag should not change the hydrophobicity or interfere with the function of ORF2 due to its small size (~2.5 kDa). Polyhistidine tags are of the most widely used methods for recombinant protein purification and sometimes even provide aid in protein expression, folding and solubility (GRÄSLUND *et al.* 2008). However, there is growing evidence showing that this assumption may be false and polyhistidine tags have also been shown to produce undesirable confirmational changes and/or affect the function of the protein it is tagged to (LEDENT *et al.* 1997; MAJOREK *et al.* 2014; KÖPPL *et al.* 2022). The untagged ORF2 protein was completely soluble during baculovirus expression, however when histidine-tagged it was almost completely insoluble. The polyhistidine tag could have caused a conformational change in ORF2 resulting in an insoluble product.

A Strep-tagged ORF2 was therefore expressed in bacteria in an additional attempt to produce sufficient yields of a soluble protein for affinity purification. Unfortunately, when induced at higher temperatures of 37 °C the tagged ORF2 produced moderate protein yields but was once again insoluble. There was no evidence suggesting that the Strep-tag would affect the confirmation of ORF2, and insolubility of strep-tagged ORF2 was likely due to protein aggregation. This is a common problem during recombinant protein expression in *E. coli* (ROSANO AND CECCARELLI 2014). At high temperatures bacterial growth, and thus translation of proteins, occur at relatively high rates which can lead to improper folding of proteins and thus insoluble aggregates. Lower temperatures can overcome this by slowing the growth rate of bacteria and providing time for newly translated recombinant proteins to fold properly (BALDWIN 1986). However, when working at lower temperatures, slower growth and reduced synthesis rates resulted in lower yields of the Strep-tagged ORF2 protein. Even when cultures were induced for longer times at low temperatures, ORF2 yields remained low. Additionally, ORF2 RNA or expression of the ORF2 protein following induction appeared to have a toxic effect on cells, as deduced from the observed slower growth rate and low final cell density when compared to uninduced cell cultures. Other reasons for low expression to keep in mind are protein turnover rates. A comparable small protein, namely rotavirus NSP6, was observed to have a turnover rate of only 2 hours after which it was completely degraded (RAINSFORD AND MCCRAE 2007). High turnover rates for AHSV ORF2 would also explain the difficulties in obtaining high yields of soluble protein during longer periods of bacterial induction. The low yields and insolubility of Strep-tagged ORF2 protein resulted in failure of attempted affinity column purification. As a result, we had to resort to eluting soluble untagged ORF2 from a polyacrylamide gel slice. The anti-ORF2 serum produced from this less effective gel purification method did contain

antibodies which reacted with a putative ORF2 protein, but additionally showed a lot of background binding to other proteins that were also present in the gel slice.

The baculovirus expression system is a common eukaryotic gene expression system used to produce a large quantity of proteins of scientific interest. ORF2 and tagged ORF2 protein were expressed in insect cells using this system, as AHSV has an insect vector namely *Culicoides* midges. We found that AHSV Seg-10 ORF2 could be recombinantly expressed using the baculovirus system. The ORF2 and HT-ORF2 proteins had a size of approximately 12 kDa. The ORF2-eGFP fusion protein showed cytoplasmic as well as nuclear localisation in the Sf9 insect cells. During confocal microscopy analysis, native eGFP was homogenously diffused over the entire cell and showed no specific localisation to any cellular structures. The moment ORF2 was fused to eGFP the localisation profile changed into discrete irregular foci, and hence it was concluded that the ORF2 was the reason for ORF2-eGFP to now have a different localisation pattern as opposed to native eGFP. Nuclear trafficking of cellular proteins is critical for the regulation of cellular processes, including DNA replication, transcriptional regulation, gene expression, signal transduction and apoptosis. Some viral proteins of certain RNA viruses localise to the nucleus at specific times in the virus life cycle, primarily early in infection, to enhance successful viral replication by interfering with or inhibiting the host antiviral response (AHMED *et al.* 2003; WULAN *et al.* 2015; LI *et al.* 2019), thereby impacting pathogenesis and disease outcome. AHSV replicates exclusively in the cytoplasm, however it was recently shown that its smallest non-structural protein NS4 is a nucleocytoplasmic protein that interacts with specific nuclear components and is a potential virulence determinant (BOUGHAN *et al.* 2020; WALL *et al.* 2021). Nuclear proteins can enter the nucleus via active transport, or through passive diffusion through the nuclear membrane if their size is less than 60kDa (WANG AND BRATTAIN 2007). As ORF2-eGFP is small (38 kDa), it could reach the nucleus through either of these mechanisms, and for the moment any function for it inside the nucleus cannot yet be inferred. In the case of BTV, the Seg-10 ORF2 protein localised within the nucleoli (STEWART *et al.* 2015). However this was not the case for AHSV, and from a co-staining of the nucleolar protein fibrillarin, it was concluded that the ORF2-eGFP protein seemed to not colocalise with the nucleoli. The native ORF2 and HT-ORF2 proteins showed similar localisation patterns to ORF2-eGFP when labelled with anti-ORF2 serum, but the high levels of background binding to other insect and baculoviral proteins made it difficult to interpret the results unequivocally.

Native and tagged ORF2 proteins were also expressed in mammalian cells, since the primary host of AHSV is horses. During transient mammalian expression in BSR-T7 cells, ORF2, ORF2-TST and ORF2-eGFP initially localised to the cytoplasm and later moved to the nucleus, and ultimately caused cells to shrivel showing a cytotoxic effect. Again, nuclear diffusion might simply have resulted from its small size, however it did not show the homogenous distribution expected from a small soluble protein. When plasmids containing only eGFP were used for transfections, many cells expressed green fluorescence. However, when ORF2-eGFP was used for transfections only a few

cells express the green fluorescence. The same low level of expression was seen for ORF2 and ORF2-TST transfections. Different genes exhibit different rate limiting steps for efficient expression. The level of protein expression from genes introduced into mammalian cells depends on various aspects. These can firstly include the number of DNA copies, mRNA transcription, processing, transport, and stability and secondly translation efficiency, protein processing, transport, and stability (KAUFMAN 2000; KHAN 2013). The same quantity of plasmid DNA was used for all transfections, and transcription was done under control of the same promoter for phCMV-eGFP and phCMV-ORF2-eGFP and similarly pCMV-T7-eGFP and pcDNA3.1-ORF2/ORF2-TST experiments. Thus, the difference between the expression levels between eGFP and ORF2 must result from differences from their RNA processing onwards. Furthermore, it was clear that during native eGFP expression even after 48 hours the cell nuclei of cells containing fluorescence remained round and appeared viable. This was not the case when ORF2 was expressed. Cell nuclei started to shrivel and become deformed from as early as 24 hours. The cytotoxic effect of ORF2 might again have influenced normal cell proliferation and thus inhibited its ability for transcription and translation leading to the low expression levels. When ORF2 and ORF2-TST were labelled with anti-ORF2 serum they showed the same localisation pattern as ORF2-eGFP. However, co-labelling of ORF2-eGFP with anti-ORF2 serum and ORF2-TST with both a Strep-Tactin®XT and anti-ORF2 serum did not provide any specific overlap. The binding site of the ant-ORF2 antibodies might not have been always available for binding due to malforming of ORF2 or even aggregations of ORF2 from cytotoxicity to the cells. Due to this fact we could not confidently conclude that the serum contained ORF2-specific antibodies.

The anti-ORF2 serum was lastly used to investigate whether AHSV potentially expresses ORF2 during its replication cycle. We could detect ORF2 in AHSV infected cells as spherical structures, similar in morphology to viral inclusion bodies. These structures were identified in the cytoplasm upon infection with AHSV-1 or AHSV-4, and additionally in the nucleus during AHSV-8 infection. As we were not able to co-label for NS2 we could not confirm that these represented VIBs, which they could potentially represent based purely on morphology. In AHSV, NS2 is a single-stranded RNA binding protein that forms perinuclear cytoplasmic VIBs (UITENWEERDE *et al.* 1995). Furthermore, due to the limitations of the anti-ORF2 serum we were not able to do more detailed investigations of the time or levels of expression of ORF2, or its trafficking or interactions with other cellular organelles, compartments or other viral proteins. Nonetheless, these experiments provide the first evidence that ASHV potentially encodes 12 proteins, with ORF2 representing a novel and previously undescribed non-structural protein which will then be NS5.

From the literature, the only protein which we could identify as a potential functional homolog for ORF2 from any non-orbivirus member of the *Reoviridae* is the rotavirus non-structural protein NSP6. It is also encoded by the smallest genome segment, from a +1 reading frame internal to that of another non-structural protein. Due to its strong start codon motif, conservation in nearly all rotavirus

strains, and evidence that its 12 kDA product was expressed in rotavirus infected cells (MATTION *et al.* 1991). It was postulated that there was selection pressure to maintain this ORF and that the NSP6 protein played a role in the rotavirus replication cycle. Further characterisation of NSP6 showed that it was expressed at a low but constant rate throughput the rotavirus infection cycle, however it had a high turnover rate and was completely degraded within 2 h of synthesis. Transiently expressed GFP-NSP6 fusion proteins appeared as an amorphous perinuclear cytoplasmic mass, which redistributed into the spherical viroplasms and co-localised with NSP5 upon rotavirus infection (RAINSFORD AND MCCRAE 2007). Rotavirus viroplasms, the equivalent of AHSV VIBs, are formed by both NSP2 and NSP5 and are the sites for viral genome replication and packaging. NSP6 was shown to have a sequence-independent broad nucleic acid binding ability (RAINSFORD AND MCCRAE 2007), which would be consistent with its localisation to viroplasms. Further studies revealed that transiently expressed tagged NSP6 of diverse rotavirus strains localised to the outer mitochondrial membrane, and that this was mediated by leucine and arginine residues within the first 50 N-terminal amino acid region of NSP6 predicted to form a non-amphipathic α-helix (HOLLOWAY et al. 2015). In recent studies, reverse genetics was used to generate recombinant NSP6-deficient rotavirus strains. These showed that NSP6 was dispensable for viral replication in cell culture (KOMOTO *et al.* 2017) and the NSP6 knockout strain had slightly lower pathogenicity in suckling mice than the parental strain (FUKUDA *et al.* 2022). The authors concluded that whilst NSP6 expression is not necessarily strictly required for replication and pathogenicity *in vitro* and *in vivo*, it contributes to optimal rotavirus replication and pathogenesis through still to be elucidated mechanisms (FUKUDA *et al.* 2022).

In conclusion we expressed a novel protein for AHSV, which showed specific localisation in the cytoplasm as well as the nucleus in transfected cells. This protein's function is not yet clear and requires further investigation.

CHAPTER 4

CONCLUDING REMARKS

In AHSV genome segment 10 (Seg-10) an additional open reading frame is predicted to exist in the +1 frame relative to the main NS3 ORF. As high selection pressures were detected on its spanning region, this suggests its evolutionary importance. Furthermore, due to the high amino acid conservation of the encoded protein across all AHSV strains, it could indicate that this ORF encodes a previously undescribed AHSV protein. We were successful in putatively detecting this ORF2 protein during AHSV infection in tissue culture, present in structures which morphologically resemble viral inclusion bodies (VIBs). Future work will be critical to confirming its time and levels of expression, localisation and subsequent function during the AHSV replication cycle.

Firstly, the growing world of machine learning methods make it possible to do deep sequence analyses. Bioinformatic analysis remains an ongoing process, as better tools are developed more accurate predictions and results become possible. This includes protein predictions, functional RNA predictions as well as possible genome packaging signals. Alphafold can additionally be used to predict protein-protein binding. ORF2 co-predictions with other non-structural proteins such as NS2 or NS4 can provide insight into its possible interactions with other viral proteins and/or its function. Lastly, the amount of scientific data is ever-growing. Everyday more sequences, protein structures, and known functional domains are loaded and available on NCBI, UniProt and many other databases. This will make it possible to in future possibly find better homology to known proteins and deduce possible functions of ORF2.

The genomes of viruses have strong selecting pressures working on them. AHSV ORF2 initiates with a methionine codon at a fully conserved position across all Seg-10 sequences. The ORF2 start codon lies within the Kozak consensus sequence ($^{A}_{G}$NNAUGG), which is known to increase translation efficiency (KOZAK 1991). Both the NS3 and NS3A start codons upstream from the ORF2 start codon on Seg-10 are in a suboptimal Kozak context (VAN STADEN AND HUISMANS 1991). This could indicate a leaky scanning translation mechanism, in which the suboptimal start codons of NS3 and NS3A are skipped over in order to initiate translation at the ORF2 start codon. Furthermore, if ORF2 does not encode a protein we would not have expected such a high amino acid similarity score and would also expect random early stop codon mutations in its frame. This further sustains the idea of it encoding a functional conserved protein.

However, viral genomes also comprise important RNA structures that control various viral processes. Because the mechanism of AHSV trans-acting RNA–RNA interactions (tRRIs) to enable co-packaging of their segmented genome is still unclear, we cannot definitely conclude that this region of high selection pressure within which ORF2 lies is not responsible for genome packaging. The segment assortment signals (SASs) on Seg-10 for correct packaging of its segmented genome is known for BTV (NEWBURN AND WHITE 2019), but not for AHSV. Production of synthetic AHS viruses with deleted/silenced regions in Seg-10, specifically in the ORF2 region and other high selection

pressure regions identified, could provide insight into the function of this region in genome packaging and/or genetic stability of recombinant viruses.

ORF2 and tagged ORF2 proteins were successfully recombinantly expressed in insect cells, as a model for the AHSV vector *Culicoides* midges, where the protein localised to cytoplasmic and nuclear foci. Whether ORF2 enters the nucleus by passive diffusion versus active trafficking could be verified by tagging ORF2 to a larger protein. A large molecule would not be able to passively diffuse into the nucleus. Next, mammalian expression provided insight into its expression in a system analogous to the AHSV host, various equid species. Lastly, bacterial expression of ORF2 was attempted in *E. coli* cells. In all cell types assayed there was low expression of the ORF2 protein and it detrimentally affected cell proliferation. It has been observed that toxic proteins might interfere with normal proliferation of bacterial cells by slowing translation and lowering overall protein expression (DONG *et al.* 1995; ROSANO AND CECCARELLI 2014). Some microbial genes have been shown to express products that are toxic to *E. coli*. Toxicity assays revealed novel toxins and restriction enzymes, and new classes of small, non-coding toxic RNAs that reproducibly inhibit *E. coli* growth (KIMELMAN *et al.* 2012). Translation-independent RNA toxicity is also a previously unrecognized obstacle in bacterial gene expression (MITTAL *et al.* 2018). The ORF2 protein or RNA could have acted in one of these ways to reduce cell growth and/or cause cell death in all systems. In order to determine if the mRNA is responsible for the toxicity, we could redo the bacterial expression but delete the start codon of ORF2 so that no protein would form. If the same toxicity persists, we know that mRNA plays a vital role in ORF2's cytotoxicity in bacterial cells.

The anti-ORF2 serum was able to detect ORF2 in AHSV infected cells as spherical structures, similar in morphology to VIBs. NS2 is responsible for forming the large, globular cytoplasmic VIBs which act as viral factories for assembly of progeny virions. Co-labelling with NS2 can confirm the co-localisation of ORF2 with viral inclusion bodies. Co-labelling with the other non-structural proteins, NS3 and NS4, would also provide further information on the co-localisation patterns of the ORF2 protein during AHSV infections. Another possible avenue to pursue would be to transfect cells with a plasmid construct that expresses ORF2-eGFP, followed by AHSV infection, to monitor the effect of the viral replication cycle on the distribution of ORF2. The structures detected by the anti-ORF2 serum could also be lipid droplets, since they have the same round shape. A lipid droplet stain like Lipi-Blue would be suitable to investigate this. Lipid droplets have been shown to form complexes with viral inclusion bodies and are crucial for rotavirus replication (CRAWFORD AND DESSELBERGER 2016). There the lipid droplets surround the viral inclusion body, protecting it, and function in signal transduction, membrane trafficking and modulation of immune and inflammatory responses in rotavirus(CRIGLAR *et al.* 2022; MARTINEZ *et al.* 2022). If these structures turn out to be lipid droplets which localise with NS2 it could be the first identification of a similar pathogenicity mechanism in AHSV. Lastly, there are organelle stains such as MitoTracker™ which is a mitochondrial stain that can be used to paint a better picture of the ORF2 protein localisation and provide insight into its

function.

Due to limitations of the anti-ORF2 serum we were not able to do more detailed investigations of the time or levels of expression of ORF2, or its trafficking or interactions with other cellular organelles, compartments or other viral proteins leading to its function still being unclear. However, these experiments do unequivocally provide the first evidence that ASHV potentially encodes 12 proteins, with ORF2 representing a novel and previously undescribed non-structural protein which will then be NS5.

Reverse genetics is an important system for generation of recombinant virus strains containing targeted mutations, which can provide information about the drivers of replication and pathogenicity of viruses and more specifically the functions of certain virus proteins. Reverse genetics can be used to make point mutations, base pair insertions and/or deletions. This in turn can produce modified functional domains, gene knock-outs or replacement of whole or partial genome segments. Reverse genetics could be utilised to rescue a recombinant AHSV strain, in which Seg-10 ORF2 has been knocked out. This can be achieved relatively easily by mutating both the ORF2 ATG start codon, and the ATG codon at amino acid 9 of ORF2, to TAC. This will not affect the NS3 sequence in the alternative reading frame, but will abolish ORF2 expression. This recombinant ORF2-knockout AHSV can be compared to wildtype AHSV for insights into ORF2's function with respect to aspect like viral replication kinetics, viral yield, cytopathogenicity and virus virulence in different systems.

References

Ahmed, A., E. Bakri, M. Hussien, M. Taseen, A. Ahmed *et al.*, 2020 Molecular Detection and Risk Factors of African Horse Sickness Virus (AHSV) in Different Governorates of Sudan. Journal of Animal Health and Production 8.

Ahmed, M., M. O. McKenzie, S. Puckett, M. Hojnacki, L. Poliquin *et al.*, 2003 Ability of the matrix protein of vesicular stomatitis virus to suppress beta interferon gene expression is genetically correlated with the inhibition of host RNA and protein synthesis. J Virol 77**:** 4646-4657.

Aksular, M., E. Calvo-Pinilla, A. Marín-López, J. Ortego, A. C. Chambers *et al.*, 2018 A single dose of African horse sickness virus (AHSV) VP2 based vaccines provides complete clinical protection in a mouse model. Vaccine 36**:** 7003-7010.

Alberca, B., K. Bachanek-Bankowska, M. Cabana, E. Calvo-Pinilla, E. Viaplana *et al.*, 2014 Vaccination of horses with a recombinant modified vaccinia Ankara virus (MVA) expressing African horse sickness (AHS) virus major capsid protein VP2 provides complete clinical protection against challenge. Vaccine 32**:** 3670-3674.

Alonso, C., S. Utrilla-Trigo, E. Calvo-Pinilla, L. Jiménez-Cabello, J. Ortego *et al.*, 2020 Inhibition of Orbivirus Replication by Aurintricarboxylic Acid. International Journal of Molecular Sciences 21**:** 7294.

Alpar, H. O., V. W. Bramwell, E. Veronesi, K. E. Darpel, P.-P. Pastoret *et al.*, 2009 Chapter 18 - Bluetongue virus vaccines past and present, pp. 397-428 in *Bluetongue*, edited by P. S. Mellor, M. Baylis and P. P. C. Mertens. Academic Press, London.

AlShaikhahmed, K., G. Leonov, P.-Y. Sung, R. J. Bingham, R. Twarock *et al.*, 2018 Dynamic network approach for the modelling of genomic sub-complexes in multi-segmented viruses. Nucleic Acids Research 46**:** 12087-12098.

Antczak, M., M. Popenda, T. Zok, J. Sarzynska, T. Ratajczak *et al.*, 2016 New functionality of RNAComposer: an application to shape the axis of miR160 precursor structure. Acta Biochim Pol 63**:** 737-744.

Assefa, A., A. Tibebu, A. Bihon, A. Dagnachew and Y. Muktar, 2022 Ecological niche modeling predicting the potential distribution of African horse sickness virus from 2020 to 2060. Sci Rep 12**:** 1748.

Attoui, H., S. Maan, S. J. Anthony and P. Mertens, 2009 Chapter 3 Bluetongue virus, other orbiviruses and other reoviruses: Their relationships and taxonomy, pp. 23-52 in *Bluetongue*, edited by R. Noad and P. Roy.

Baldwin, R. L., 1986 Temperature dependence of the hydrophobic interaction in protein folding. Proceedings of the National Academy of Sciences 83**:** 8069-8072.

Balvay, L., R. Soto Rifo, E. P. Ricci, D. Decimo and T. Ohlmann, 2009 Structural and functional diversity of viral IRESes. Biochim Biophys Acta 1789**:** 542-557.

Barnard, B. J., 1998 Epidemiology of African horse sickness and the role of the zebra in South Africa. Arch Virol Suppl 14**:** 13-19.

Basak, A. K., J. M. Grimes, P. Gouet, P. Roy and D. I. Stuart, 1997 Structures of orbivirus VP7: implications for the role of this protein in the viral life cycle. Structure 5**:** 871-883.

Beaton, A. R., J. Rodriguez, Y. K. Reddy and P. Roy, 2002 The membrane trafficking protein calpactin forms a complex with bluetongue virus protein NS3 and mediates virus release. Proc Natl Acad Sci U S A 99**:** 13154-13159.

Bekker, S., H. Huismans and V. van Staden, 2014 Factors that affect the intracellular localization and trafficking of African horse sickness virus core protein, VP7. Virology 456-457**:** 279-291.

Bekker, S., H. Huismans and V. van Staden, 2022 Generation of a Soluble African Horse Sickness Virus VP7 Protein Capable of Forming Core-like Particles. Viruses 14.

Belhouchet, M., F. Mohd Jaafar, A. E. Firth, J. M. Grimes, P. P. Mertens *et al.*, 2011 Detection of a fourth orbivirus non-structural protein. PLoS ONE 6**:** e25697.

Belli, B. A., and C. E. Samuel, 1993 Biosynthesis of Reovirus-Specified Polypeptides: Identification of Regions of the Bicistronic Reovirus S1 mRNA That Affect the Efficiency of Translation in Animal Cells. Virology 193**:** 16-27.

Benson, D. A., M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D. J. Lipman *et al.*, 2013 GenBank. Nucleic Acids Res 41**:** D36-42.

Bonneau, K. R., B. A. Mullens and N. J. MacLachlan, 2001 Occurrence of genetic drift and founder effect during quasispecies evolution of the VP2 and NS3/NS3A genes of bluetongue virus upon passage between sheep, cattle, and Culicoides sonorensis. J Virol 75**:** 8298-8305.

Borman, A. M., and K. M. Kean, 1997 Intact eukaryotic initiation factor 4G is required for hepatitis A virus internal initiation of translation. Virology 237**:** 129-136.

Borodavka, A., U. Desselberger and J. T. Patton, 2018 Genome packaging in multi-segmented dsRNA viruses: distinct mechanisms with similar outcomes. Current Opinion in Virology 33**:** 106-112.

Bosman, P., G. K. Brückner and A. Faul, 1995 African horse sickness surveillance systems and regionalisation/zoning: the case of South Africa. Rev Sci Tech 14**:** 645-653.

Boughan, S., 2016 Expression and intracellular localisation of African horse sickness virus non-structural protein NS4, pp. in *In the Faculty of Natural & Agricultural Sciences*. University of Pretoria, University of Pretoria.

Boughan, S., A. Christiaan Potgieter and V. van Staden, 2020 African horse sickness virus NS4 is a nucleocytoplasmic protein that localizes to PML nuclear bodies. Journal of General Virology 101**:** 366-384.

Boyce, M., C. C. Celma and P. Roy, 2012 Bluetongue virus non-structural protein 1 is a positive regulator of viral protein synthesis. Virol J 9**:** 178.

Boyce, M., and P. Roy, 2007 Recovery of Infectious Bluetongue Virus from RNA. Journal of Virology 81**:** 2179-2186.

Boyce, M., J. Wehrfritz, R. Noad and P. Roy, 2004 Purified recombinant bluetongue virus VP1 exhibits RNA replicase activity. J Virol 78**:** 3994-4002.

Brameier, M., A. Krings and R. M. MacCallum, 2007 NucPred--predicting nuclear localization of proteins. Bioinformatics 23**:** 1159-1160.

Bremer, C. W., H. Huismans and A. A. Van Dijk, 1990 Characterization and cloning of the African horsesickness virus genome. J Gen Virol 71 (Pt 4)**:** 793-799.

Brookes, S. M., A. D. Hyatt and B. T. Eaton, 1993 Characterization of virus inclusion bodies in bluetongue virus-infected cells. J Gen Virol 74 (Pt 3)**:** 525-530.

Brosnan, C. A., and O. Voinnet, 2009 The long and the short of noncoding RNAs. Current Opinion in Cell Biology 21**:** 416-425.

Brown, E. A., S. P. Day, R. W. Jansen and S. M. Lemon, 1991 The 5' nontranslated region of hepatitis A virus RNA: secondary structure and elements required for translation in vitro. J Virol 65**:** 5828-5838.

Buchholz, U. J., S. Finke and K. K. Conzelmann, 1999 Generation of bovine respiratory syncytial virus (BRSV) from cDNA: BRSV NS2 is not essential for virus replication in tissue culture, and the human RSV leader region acts as a functional BRSV genome promoter. J Virol 73**:** 251-259.

Burkhardt, C., P. Y. Sung, C. C. Celma and P. Roy, 2014 Structural constraints in the packaging of bluetongue virus genomic segments. J Gen Virol 95**:** 2240-2250.

Burrage, T., and W. Laegreid, 1994 African horsesickness: pathogenesis and immunity. Comparative immunology, microbiology and infectious diseases 17**:** 275-285.

Burroughs, J. N., R. S. O'Hara, C. J. Smale, C. Hamblin, A. Walton *et al.*, 1994 Purification and properties of virus particles, infectious subviral particles, cores and VP7 crystals of African horsesickness virus serotype 9. J Gen Virol 75 (Pt 8)**:** 1849-1857.

Calvo-Pinilla, E., A. Marín-López, S. Utrilla-Trigo, L. Jiménez-Cabello and J. Ortego, 2020 Reverse genetics approaches: a novel strategy for African horse sickness virus vaccine design. Current Opinion in Virology 44**:** 49-56.

Carpenter, S., P. Mellor and S. Torr, 2008 Control techniques for Culicoides biting midges and their application in the UK and northwestern Palaearctic. Medical and veterinary entomology 22**:** 175-187.

Castillo-Olivares, J., E. Calvo-Pinilla, I. Casanova, K. Bachanek-Bankowska, R. Chiam *et al.*, 2011 A Modified Vaccinia Ankara Virus (MVA) Vaccine Expressing African Horse Sickness Virus (AHSV) VP2 Protects Against AHSV Challenge in an IFNAR −/− Mouse Model. PLOS ONE 6**:** e16503.

Celma, C. C., and P. Roy, 2009 A viral nonstructural protein regulates bluetongue virus trafficking and release. J Virol 83**:** 6806-6816.

Celma, C. C. P., and P. Roy, 2011 Interaction of Calpactin Light Chain (S100A10/p11) and a Viral NS Protein Is Essential for Intracellular Trafficking of Nonenveloped Bluetongue Virus. J. Virol. 85**:** 4783-4791.

Chen, Y.-R., S. Zhong, Z. Fei, Y. Hashimoto, J. Z. Xiang *et al.*, 2013 The Transcriptome of the Baculovirus Autographa californica Multiple Nucleopolyhedrovirus in Trichoplusia ni Cells. Journal of Virology 87**:** 6391-6405.

Clift, S. J., and M. L. Penrith, 2010 Tissue and cell tropism of African horse sickness virus demonstrated by immunoperoxidase labeling in natural and experimental infection in horses in South Africa. Vet Pathol 47**:** 690-697.

Cock, P. J. a. A., Tiago and Chang, Jeffrey T and Chapman, Brad A and Cox, Cymon J and Dalke, Andrew and Friedberg, Iddo and Hamelryck, Thomas and Kauff, Frank and Wilczynski, Bartek, 2009 Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25**:** 1422-1423.

Coetzer, J., and A. J. Guthrie, 2005 *African horsesickness. In: Infectious diseases of Livestock*. Oxford University Press, Cape Town, South Africa.

Coetzer, J. A. W., G. R. Thomson and R. C. Tustin, 1994 *Infectious diseases of livestock with special reference to Southern Africa*. Oxford University Press, Cape Town; New York.

Conte, M. R., and S. Pennell, 2013 Translation Initiation: A Eukaryotic Perspective, pp. 2644-2650 in *Encyclopedia of Biophysics*, edited by G. C. K. Roberts. Springer Berlin Heidelberg, Berlin, Heidelberg.

Craig, A. F., G. C. Packer, A. J. Guthrie and E. H. Venter, 2019 Evaluating african horse sickness virus in horses and field-caught culicoides biting midges on the east rand, gauteng province, south africa. Veterinaria Italiana 55**:** 91-94.

Crawford, S. E., and U. Desselberger, 2016 Lipid droplets form complexes with viroplasms and are crucial for rotavirus replication. Curr Opin Virol 19**:** 11-15.

Criglar, J. M., M. K. Estes and S. E. Crawford, 2022 Rotavirus-Induced Lipid Droplet Biogenesis Is Critical for Virus Replication. Front Physiol 13**:** 836870.

Darpel, K. E., K. F. Langner, M. Nimtz, S. J. Anthony, J. Brownlie *et al.*, 2011 Saliva proteins of vector Culicoides modify structure and infectivity of bluetongue virus particles. PLoS ONE 6**:** e17545.

de Waal, P. J., and H. Huismans, 2005 Characterization of the nucleic acid binding activity of inner core protein VP6 of African horse sickness virus. Arch Virol 150**:** 2037-2050.

Desselberger, U., 2020 What are the limits of the packaging capacity for genomic RNA in the cores of rotaviruses and of other members of the Reoviridae? Virus Research 276**:** 197822.

Diprose, J., J. Grimes, G. Sutton, J. Burroughs, A. Meyer *et al.*, 2002 The core of bluetongue virus binds double-stranded RNA. Journal of virology 76**:** 9533-9536.

Dong, H., L. Nilsson and C. G. Kurland, 1995 Gratuitous overexpression of genes in Escherichia coli leads to growth inhibition and ribosome destruction. Journal of Bacteriology 177**:** 1497-1504.

Drozdetskiy, A., C. Cole, J. Procter and G. J. Barton, 2015 JPred4: a protein secondary structure prediction server. Nucleic Acids Res 43**:** W389-394.

Du Plessis, M., and L. H. Nel, 1997 Comparative sequence analysis and expression of the M6 gene, encoding the outer capsid protein VP5, of African horsesickness virus serotype nine. Virus Research 47**:** 41-49.

Durán-Ferrer, M., M. Agüero, S. Zientara, C. Beck, S. Lecollinet *et al.*, 2019 Assessment of reproducibility of a VP7 Blocking ELISA diagnostic test for African horse sickness. Transboundary and Emerging Diseases 66**:** 83-90.

Eaton, B. T., A. D. Hyatt and J. R. White, 1988 Localization of the nonstructural protein NS1 in bluetongue virus-infected cells and its presence in virus particles. Virology 163**:** 527-537.

Edgar, R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32**:** 1792-1797.

England, M. E., P. Pearce-Kelly, V. A. Brugman, S. King, S. Gubbins *et al.*, 2020 Culicoides species composition and molecular identification of host blood meals at two zoos in the UK. Parasites & Vectors 13**:** 139.

Ernst, H., and A. J. Shatkin, 1985 Reovirus hemagglutinin mRNA codes for two polypeptides in overlapping reading frames. Proc Natl Acad Sci U S A 82**:** 48-52.

Fearon, S. H., S. J. Dennis, Hitzeroth, II, E. P. Rybicki and A. E. Meyers, 2022 Plant expression systems as an economical alternative for the production of iELISA coating antigen AHSV VP7. N Biotechnol 68**:** 48-56.

Fearon, S. H., S. J. Dennis, I. I. Hitzeroth, E. P. Rybicki and A. E. Meyers, 2021 Humoral and cell-mediated immune responses to plant-produced African horse sickness virus VP7 quasi-crystals. Virus Research 294: 198284.

Feenstra, F., R. G. van Gennip, M. Maris-Veldhuis, E. Verheij and P. A. van Rijn, 2014a Bluetongue virus without NS3/NS3a expression is not virulent and protects against virulent bluetongue virus challenge. J Gen Virol 95: 2019-2029.

Feenstra, F., R. G. van Gennip, M. Schreuder and P. A. van Rijn, 2016 Balance of RNA sequence requirement and NS3/NS3a expression of segment 10 of orbiviruses. J Gen Virol 97: 411-421.

Feenstra, F., R. G. van Gennip, S. G. van de Water and P. A. van Rijn, 2014b RNA elements in open reading frames of the bluetongue virus genome are essential for virus replication. PLoS ONE 9: e92377.

Ferreira-Venter, L., E. Venter, J. Theron and V. Van Staden, 2019 Targeted mutational analysis to unravel the complexity of African horse sickness virus NS3 function in mammalian cells. Virology 531: 149-161.

Finkel, Y., N. Stern-Ginossar and M. Schwartz, 2018 Viral Short ORFs and Their Possible Functions. Proteomics 18: e1700255.

Firth, A. E., 2008 Bioinformatic analysis suggests that the Orbivirus VP6 cistron encodes an overlapping gene. Virol J 5: 48.

Firth, A. E., 2014 Mapping overlapping functional elements embedded within the protein-coding regions of RNA viruses. Nucleic Acids Res 42: 12425-12439.

Firth, A. E., 2020 A putative new SARS-CoV protein, 3c, encoded in an ORF overlapping ORF3a. Journal of General Virology 101: 1085-1089.

Firth, A. E., and J. F. Atkins, 2008 Bioinformatic analysis suggests that a conserved ORF in the waikaviruses encodes an overlapping gene. Arch Virol 153: 1379-1383.

Firth, A. E., and I. Brierley, 2012 Non-canonical translation in RNA viruses. The Journal of general virology 93: 1385-1409.

Forzan, M., C. Wirblich and P. Roy, 2004 A capsid protein of nonenveloped Bluetongue virus exhibits membrane fusion activity. Proc Natl Acad Sci U S A 101: 2100-2105.

French, T. J., and P. Roy, 1990 Synthesis of bluetongue virus (BTV) corelike particles by a recombinant baculovirus expressing the two major structural core proteins of BTV. J Virol 64: 1530-1536.

Fukuda, S., M. Kugita, Y. Higashimoto, K. Shiogama, H. Tsujikawa *et al.*, 2022 Rotavirus incapable of NSP6 expression can cause diarrhea in suckling mice. J Gen Virol 103.

Gao, S., Z. Zeng, H. Wang, F. Chen, L. Huang *et al.*, 2022 Predicting the possibility of African horse sickness (AHS) introduction into China using spatial risk analysis and habitat connectivity of Culicoides. Sci Rep 12: 3910.

Gomez-Villamandos, J., C. Sanchez, L. Carrasco, M. Laviada, M. Bautista *et al.*, 1999 Pathogenesis of African horse sickness: ultrastructural study of the capillaries in experimental infection. Journal of comparative pathology 121: 101-116.

Gräslund, S., P. Nordlund, J. Weigelt, B. M. Hallberg, J. Bray *et al.*, 2008 Protein production and purification. Nat Methods 5: 135-146.

Grewar, J., C. Weyer, A. Guthrie, P. Koen, S. Davey *et al.*, 2012 The 2011 outbreak of African horse sickness in the African horse sickness controlled area in South Africa. Journal of the South African Veterinary Association 84: 1-7.

Grünberg, R., M. Nilges and J. Leckner, 2007 Biskit—A software platform for structural bioinformatics. Bioinformatics 23: 769-770.

Grundhoff, A., and C. S. Sullivan, 2011 Virus-encoded microRNAs. Virology 411: 325-343.

Gu, L., V. Musiienko, Z. Bai, A. Qin, S. W. Schneller *et al.*, 2012 Novel virostatic agents against bluetongue virus. PloS one 7: e43341-e43341.

Gupta, A., and M. Bansal, 2020 RNA-mediated translation regulation in viral genomes: computational advances in the recognition of sequences and structures. Briefings in Bioinformatics 21: 1151-1163.

Guzmán, F., A. Distéfano, J. Arneodo, H. Hopp, S. Lenardon *et al.*, 2007 Sequencing of the bicistronic genome segments S7 and S9 of Mal de Río Cuarto virus (Fijivirus, Reoviridae) completes the genome of this virus. Archives of virology 152: 565-573.

Harrison, R. L., and D. L. Jarvis, 2006 Protein N-Glycosylation in the Baculovirus–Insect Cell Expression System and Engineering of Insect Cells to Produce "Mammalianized" Recombinant Glycoproteins, pp. 159-191 in *Advances in Virus Research*. Academic Press.

Hassan, S. H., C. Wirblich, M. Forzan and P. Roy, 2001 Expression and functional characterization of bluetongue virus VP5 protein: role in cellular permeabilization. J Virol 75**:** 8356-8367.

Hassan, S. S., and P. Roy, 1999 Expression and functional characterization of bluetongue virus VP2 protein: role in cell entry. J Virol 73**:** 9832-9842.

Hoffmann, B., S. Joseph, N. A. G. Patteril, M. R. Caveney, S. K. Elizabeth *et al.*, 2022 Comparative genome analysis of all nine African horse sickness serotypes isolated from equine fatalities in Kenya and South Africa. J Equine Vet Sci**:** 104137.

Holloway, G., R. I. Johnson, Y. Kang, V. T. Dang, D. Stojanovski *et al.*, 2015 Rotavirus NSP6 localizes to mitochondria via a predicted N-terminal a-helix. J Gen Virol 96**:** 3519-3524.

Howell, P., 1962 The isolation and identification of further antigenic types of African horsesickness virus. Onderstepoort J Vet Res 29**:** 139–149.

Hu, L., W. Salmen, B. Sankaran, Y. Lasanajak, D. F. Smith *et al.*, 2022 Novel fold of rotavirus glycan-binding domain predicted by AlphaFold2 and determined by X-ray crystallography. Communications Biology 5**:** 419.

Hughes, A. L., and M. A. K. Hughes, 2005 Patterns of nucleotide difference in overlapping and non-overlapping reading frames of papillomavirus genomes. Virus Research 113**:** 81-88.

Huismans, H., and B. J. Erasmus, 1981 Identification of the serotype-specific and group-specific antigens of bluetongue virus. Onderstepoort J Vet Res 48**:** 51-58.

Huismans, H. V., V. Van Staden, W. C. Fick and M. M. van Niekerk, 2004 A comparison of different orbivirus proteins that could affect virulence and pathogenesis. Veterinaria Italiana 40**:** 417-425.

Hwang, G.-Y., Y. Yi-Yuan, C. Jwo-Farn and K. K. L. Joseph, 1992 Sequence conservation among the cognate nonstructural NS3/3A protein genes of six bluetongue viruses. Virus Research 23**:** 151-161.

Hyatt, A. D., B. T. Eaton and S. M. Brookes, 1989 The release of bluetongue virus from infected cells and their superinfection by progeny virus. Virology 173**:** 21-34.

Hyatt, A. D., A. R. Gould, B. Coupar and B. T. Eaton, 1991 Localization of the non-structural protein NS3 in bluetongue virus-infected cells. J Gen Virol 72 (Pt 9)**:** 2263-2267.

Iglesias, N. G., and A. V. Gamarnik, 2011 Dynamic RNA structures in the dengue virus genome. RNA Biol 8**:** 249-257.

Jacks, T., and H. E. Varmus, 1985 Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. Science 230**:** 1237-1242.

Jackson, R. J., C. U. Hellen and T. V. Pestova, 2012 Termination and post-termination events in eukaryotic translation. Adv Protein Chem Struct Biol 86**:** 45-93.

Jang, S. K., H. G. Kräusslich, M. J. Nicklin, G. M. Duke, A. C. Palmenberg *et al.*, 1988 A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. Journal of virology 62**:** 2636-2643.

Jarvis, D. L., 2003 Developing baculovirus-insect cell expression systems for humanized recombinant glycoprotein production. Virology 310**:** 1-7.

Jiménez-Cabello, L., S. Utrilla-Trigo, N. Barreiro-Piñeiro, T. Pose-Boirazian, J. Martínez-Costas *et al.*, 2022 Nanoparticle- and Microparticle-Based Vaccines against Orbiviruses of Veterinary Importance. Vaccines (Basel) 10.

Jones, D. T., W. R. Taylor and J. M. Thornton, 1994 A mutation data matrix for transmembrane proteins. FEBS Lett 339**:** 269-275.

Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov *et al.*, 2021 Highly accurate protein structure prediction with AlphaFold. Nature 596**:** 583-589.

Kar, A. K., B. Bhattacharya and P. Roy, 2007 Bluetongue virus RNA binding protein NS2 is a modulator of viral replication and assembly. BMC Mol Biol 8**:** 4.

Kar, A. K., and P. Roy, 2003 Defining the structure-function relationships of bluetongue virus helicase protein VP6. J Virol 77**:** 11347-11356.

Kaufman, R. J., 2000 Overview of vector design for mammalian gene expression. Mol Biotechnol 16**:** 151-160.

Kerviel, A., P. Ge, M. Lai, J. Jih, M. Boyce *et al.*, 2019 Atomic structure of the translation regulatory protein NS1 of bluetongue virus. Nature Microbiology 4**:** 837-845.

Khan, K. H., 2013 Gene expression in Mammalian cells and its applications. Adv Pharm Bull 3**:** 257-263.

Kim, D. Y., A. E. Firth, S. Atasheva, E. I. Frolova and I. Frolov, 2011 Conservation of a packaging signal and the viral genome RNA packaging mechanism in alphavirus evolution. J Virol 85**:** 8022-8036.

Kimelman, A., A. Levy, H. Sberro, S. Kidron, A. Leavitt *et al.*, 2012 A vast collection of microbial genes that are toxic to bacteria. Genome Res 22**:** 802-809.

King, S., P. Rajko-Nenow, M. Ashby, L. Frost, S. Carpenter *et al.*, 2020 Outbreak of African horse sickness in Thailand, 2020. Transboundary and Emerging Diseases 67**:** 1764-1767.

Komoto, S., Y. Kanai, S. Fukuda, M. Kugita, T. Kawagishi *et al.*, 2017 Reverse Genetics System Demonstrates that Rotavirus Nonstructural Protein NSP6 Is Not Essential for Viral Replication in Cell Culture. J Virol 91.

Komoto, S., J. Sasaki and K. Taniguchi, 2006 Reverse genetics system for introduction of site-specific mutations into the double-stranded RNA genome of infectious rotavirus. Proceedings of the National Academy of Sciences of the United States of America 103**:** 4646-4651.

Köppl, C., N. Lingg, A. Fischer, C. Kröß, J. Loibl *et al.*, 2022 Fusion Tag Design Influences Soluble Recombinant Protein Production in Escherichia coli. Int J Mol Sci 23.

Kosakovsky Pond, S. L., and S. D. W. Frost, 2005 Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. Molecular Biology and Evolution 22**:** 1208-1222.

Kozak, M., 1990a Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proc Natl Acad Sci U S A 87**:** 8301-8305.

Kozak, M., 1990b Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. Proceedings of the National Academy of Sciences of the United States of America 87**:** 8301-8305.

Kozak, M., 1991 Structural features in eukaryotic mRNAs that modulate the initiation of translation. J Biol Chem 266**:** 19867-19870.

Kozak, M., 2002 Pushing the limits of the scanning mechanism for initiation of translation. Gene 299**:** 1-34.

Kundlacz, C., M. Pourcelot, A. Fablet, R. Amaral Da Silva Moraes, T. Léger *et al.*, 2019 Novel Function of Bluetongue Virus NS3 Protein in Regulation of the MAPK/ERK Signaling Pathway. Journal of virology 93**:** e00336-00319.

Kushnir, N., S. J. Streatfield and V. Yusibov, 2012 Virus-like particles as a highly efficient vaccine platform: diversity of targets and production systems and advances in clinical development. Vaccine 31**:** 58-83.

Labadie, T., S. Jegouic and P. Roy, 2019 Bluetongue Virus Nonstructural Protein 3 Orchestrates Virus Maturation and Drives Non-Lytic Egress via Two Polybasic Motifs. Viruses 11**:** 1107.

Ledent, P., C. Duez, M. Vanhove, A. Lejeune, E. Fonzé *et al.*, 1997 Unexpected influence of a C-terminal-fused His-tag on the processing of an enzyme and on the kinetic and folding parameters. FEBS Lett 413**:** 194-196.

Li, G., X. Qi, Z. Hu and Q. Tang, 2019 Mechanisms Mediating Nuclear Trafficking Involved in Viral Propagation by DNA Viruses. Viruses 11.

Li, W., E. Manktelow, J. C. von Kirchbach, J. R. Gog, U. Desselberger *et al.*, 2010 Genomic analysis of codon, sequence and structural conservation with selective biochemical-structure mapping reveals highly conserved and dynamic structures in rotavirus RNAs with potential cis-acting functions. Nucleic Acids Res 38**:** 7718-7735.

Lopez-Ulloa, B., Y. Fuentes, M. S. Pizarro-Ortega and M. Lopez-Lastra, 2022 RNA-Binding Proteins as Regulators of Internal Initiation of Viral mRNA Translation. Viruses 14.

Lourenco, S., and P. Roy, 2011 In vitro reconstitution of Bluetongue virus infectious cores. Proc Natl Acad Sci U S A 108**:** 13746-13751.

Lu, G., J. Pan, J. Ou, R. Shao, X. Hu *et al.*, 2020 African horse sickness: Its emergence in Thailand and potential threat to other Asian countries. Transboundary and Emerging Diseases 67**:** 1751-1753.

Luckow, V. A., S. C. Lee, G. F. Barry and P. O. Olins, 1993 Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in Escherichia coli. Journal of Virology 67**:** 4566-4579.

Lulla, V., A. Losada, S. Lecollinet, A. Kerviel, T. Lilin *et al.*, 2017 Protective efficacy of multivalent replication-abortive vaccine strains in horses against African horse sickness virus challenge. Vaccine 35: 4262-4269.

Lulla, V., A. Lulla, K. Wernike, A. Aebischer, M. Beer *et al.*, 2016 Assembly of replication-incompetent African horse sickness virus particles: rational design of vaccines for all serotypes. Journal of virology 90: 7405-7414.

Lymperopoulos, K., R. Noad, S. Tosi, S. Nethisinghe, I. Brierley *et al.*, 2006 Specific binding of Bluetongue virus NS2 to different viral plus-strand RNAs. Virology 353: 17-26.

Maclachlan, N. J., and A. J. Guthrie, 2010 Re-emergence of bluetongue, African horse sickness, and other orbivirus diseases. Vet Res 41: 35.

Magnus, M., M. J. Boniecki, W. Dawson and J. M. Bujnicki, 2016 SimRNAweb: a web server for RNA 3D structure modeling with optional restraints. Nucleic Acids Res 44: W315-319.

Majorek, K. A., M. L. Kuhn, M. Chruszcz, W. F. Anderson and W. Minor, 2014 Double trouble-Buffer selection and His-tag presence may be responsible for nonreproducibility of biomedical experiments. Protein Sci 23: 1359-1368.

Manole, V., P. Laurinmaki, W. Van Wyngaardt, C. A. Potgieter, I. M. Wright *et al.*, 2012 Structural Insight into African Horsesickness Virus Infection. J. Virol. 86: 7858-7866.

Maree, F. F., and H. Huismans, 1997 Characterization of tubular structures composed of nonstructural protein NS1 of African horsesickness virus expressed in insect cells. J Gen Virol 78 (Pt 5): 1077-1082.

Maree, S., S. Durbach and H. Huismans, 1998a Intracellular production of African horsesickness virus core-like particles by expression of the two major core proteins, VP3 and VP7, in insect cells. J Gen Virol 79 (Pt 2): 333-337.

Maree, S., S. Durbach, F. F. Maree, F. Vreede and H. Huismans, 1998b Expression of the major core structural proteins VP3 and VP7 of African horse sickness virus, and production of core-like particles. Arch Virol Suppl 14: 203-209.

Maree, S., and J. T. Paweska, 2005 Preparation of recombinant African horse sickness virus VP7 antigen via a simple method and validation of a VP7-based indirect ELISA for the detection of group-specific IgG antibodies in horse sera. Journal of Virological Methods 125: 55-65.

Marín-López, A., N. Barreiro-Piñeiro, S. Utrilla-Trigo, D. Barriales, J. Benavente *et al.*, 2020 Cross-protective immune responses against African horse sickness virus after vaccination with protein NS1 delivered by avian reovirus muNS microspheres and modified vaccinia virus Ankara. Vaccine 38: 882-889.

Martin, L. A., A. J. Meyer, R. S. O'Hara, H. Fu, P. S. Mellor *et al.*, 1998 Phylogenetic analysis of African horse sickness virus segment 10: sequence variation, virulence characteristics and cell exit. Arch Virol Suppl 14: 281-293.

Martinez-Costas, J., G. Sutton, N. Ramadevi and P. Roy, 1998 Guanylyltransferase and RNA 5'-triphosphatase activities of the purified expressed VP4 protein of bluetongue virus. J Mol Biol 280: 859-866.

Martinez, J. L., C. Eichwald, E. M. Schraner, S. Lopez and C. F. Arias, 2022 Lipid metabolism is involved in the association of rotavirus viroplasms with endoplasmic reticulum membranes. Virology 569: 29-36.

Matsuo, E., E. Leon, S. J. Matthews and P. Roy, 2014 Structure based modification of Bluetongue virus helicase protein VP6 to produce a viable VP6-truncated BTV. Biochem Biophys Res Commun 451: 603-608.

Matsuo, E., and P. Roy, 2009 Bluetongue virus VP6 acts early in the replication cycle and can form the basis of chimeric virus formation. J Virol 83: 8842-8848.

Matsuo, E., and P. Roy, 2013 Minimum Requirements for Bluetongue Virus Primary Replication In Vivo. J. Virol. 87: 882-889.

Mattion, N. M., D. B. Mitchell, G. W. Both and M. K. Estes, 1991 Expression of rotavirus proteins encoded by alternative open reading frames of genome segment 11. Virology 181: 295-304.

Maurer, L. M., A. Paslaru, P. R. Torgerson, E. Veronesi and A. Mathis, 2022 Vector competence of Culicoides biting midges from Switzerland for African horse sickness virus and epizootic haemorrhagic disease virus. Schweiz Arch Tierheilkd 164: 66-70.

Maurice, Y., and A. Provost, 1967 [Horse sickness caused by type 9 virus in Central Africa: serological survey]. Rev Elev Med Vet Pays Trop 20: 21-25.

McGuffin, L. J., R. Adiyaman, A. H. A. Maghrabi, A. N. Shuid, D. A. Brackenridge *et al.*, 2019 IntFOLD: an integrated web resource for high performance protein structure and function prediction. Nucleic Acids Research 47**:** W408-W413.

McIntosh, B. M., 1958 Immunological types of horsesickness virus and their significance in immunization. Onderstepoort J Vet Res 27**:** 465-538.

Mehalko, J. L., and D. Esposito, 2016 Engineering the transposition-based baculovirus expression vector system for higher efficiency protein production from insect cells. Journal of biotechnology 238**:** 1-8.

Meiswinkel, R., M. Baylis and K. Labuschagne, 2000 Stabling and the protection of horses from Culicoides bolitinos (Diptera: Ceratopogonidae), a recently identified vector of African horse sickness. Bulletin of entomological research 90**:** 509-515.

Mellor, P., and J. Boorman, 1995 The transmission and geographical spread of African horse sickness and bluetongue viruses. Annals of Tropical Medicine & Parasitology 89**:** 1-15.

Mellor, P. S., 1993 African horse sickness: transmission and epidemiology. Veterinary research 24 2**:** 199-212.

Mellor, P. S., S. Carpenter and D. M. White, 2009 Chapter 14 - Bluetongue virus in the insect host, pp. 295-320 in *Bluetongue*. Academic Press, London.

Mellor, P. S., and C. Hamblin, 2004 African horse sickness. Vet Res 35**:** 445-466.

Mellor, P. S., P. Rawlings, M. Baylis and M. P. Wellby, 1998 Effect of temperature on African horse sickness virus infection in Culicoides. Arch Virol Suppl 14**:** 155-163.

Mertens, P. P., J. N. Burroughs and J. Anderson, 1987 Purification and properties of virus particles, infectious subviral particles, and cores of bluetongue virus serotypes 1 and 4. Virology 157**:** 375-386.

Mertens, P. P., and J. Diprose, 2004 The bluetongue virus core: a nano-scale transcription machine. Virus Res 101**:** 29-43.

Mirdita, M., K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov *et al.*, 2022 ColabFold - Making protein folding accessible to all. bioRxiv**:** 2021.2008.2015.456425.

Mittal, P., J. Brindle, J. Stephen, J. B. Plotkin and G. Kudla, 2018 Codon usage influences fitness through RNA toxicity. Proc Natl Acad Sci U S A 115**:** 8639-8644.

Modrof, J., K. Lymperopoulos and P. Roy, 2005 Phosphorylation of bluetongue virus nonstructural protein 2 is essential for formation of viral inclusion bodies. J Virol 79**:** 10023-10031.

Molini, U., G. Zaccaria, E. Kandiwa, B. Mushonga, S. Khaiseb *et al.*, 2020 Seroprevalence of African horse sickness in selected donkey populations in Namibia. Veterinary world 13**:** 1005-1009.

Monastyrskaya, K., T. Booth, L. Nel and P. Roy, 1994 Mutation of either of two cysteine residues or deletion of the amino or carboxy terminus of nonstructural protein NS1 of bluetongue virus abrogates virus-specified tubule formation in insect cells. J Virol 68**:** 2169-2178.

Mukhopadhyay, E., S. Hazra and D. Banerjee, 2020 Effect of Environment and Elevation on Seasonal Prevalence of Culicoides in West Bengal. Indian Journal of Animal Research 54**:** 1125-1135.

Namy, O., S. J. Moran, D. I. Stuart, R. J. C. Gilbert and I. Brierley, 2006 A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. Nature 441**:** 244-247.

Namy, O., and J.-P. Rousset, 2010 Specification of Standard Amino Acids by Stop Codons, pp. 79-100 in *Recoding: Expansion of Decoding Rules Enriches Gene Expression*, edited by J. F. Atkins and R. F. Gesteland. Springer New York, New York, NY.

Napthine, S., C. Yek, M. L. Powell, T. D. Brown and I. Brierley, 2012 Characterization of the stop codon readthrough signal of Colorado tick fever virus segment 9 RNA. Rna 18**:** 241-252.

Nawrocki, E. P., and S. R. Eddy, 2013 Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29**:** 2933-2935.

Ndebé, M. M. F., M. M. M. Mouiche, F. Moffo, R. N. S. Poueme and J. Awah-Ndukum, 2022 Seroprevalence and Risk Factors of African Horse Sickness in Three Agroecological Zones of Cameroon. Vet Med Int 2022**:** 2457772.

Nelson, C. W., Z. Ardern and X. Wei, 2020 OLGenie: Estimating Natural Selection to Predict Functional Overlapping Genes. Molecular biology and evolution 37**:** 2440-2449.

Nelson, E., W. Thurston, P. Pearce-Kelly, H. Jenkins, M. Cameron *et al.*, 2022 A Qualitative Risk Assessment for Bluetongue Disease and African Horse Sickness: The Risk of Entry and Exposure at a UK Zoo. Viruses 14.

Newburn, L. R., and K. A. White, 2019 Trans-Acting RNA–RNA Interactions in Segmented RNA Viruses. Viruses 11: 751.

Ngoveni, H. G., A. v. Schalkwyk and J. J. O. Koekemoer, 2019 Evidence of intragenic recombination in African horse sickness virus. Viruses 11.

Nguyen Ba, A. N., A. Pogoutse, N. Provart and A. M. Moses, 2009 NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. BMC Bioinformatics 10: 202.

Noad, R., and P. Roy, 2009 Chapter 4 Bluetongue virus replication and assembly, pp. 53-76 in *Bluetongue*, edited by R. Noad and P. Roy.

O'Kennedy, M. M., P. Coetzee, O. Koekemoer, L. du Plessis, C. W. Lourens *et al.*, 2022 Protective immunity of plant-produced African horse sickness virus serotype 5 chimaeric virus-like particles (VLPs) and viral protein 2 (VP2) vaccines in IFNAR(-/-) mice. Vaccine 40: 5160-5169.

OIE, 2019 OIE-Listed diseases. Office International des Epizooties. World Organization for Animal Health. http://www.oie.int.

Owens, R. J., C. Limn and P. Roy, 2004 Role of an arbovirus nonstructural protein in cellular pathogenesis and virus release. J Virol 78: 6649-6656.

Pavesi, A., A. Vianelli, N. Chirico, Y. Bao, O. Blinkova *et al.*, 2018 Overlapping genes and the proteins they encode differ significantly in their sequence composition from non-overlapping genes. PloS one 13: e0202513-e0202513.

Pennock, G. D., C. Shoemaker and L. K. Miller, 1984 Strong and regulated expression of Escherichia coli beta-galactosidase in insect cells with a baculovirus vector. Molecular and cellular biology 4: 399-406.

Pérez-Losada, M., M. Arenas, J. C. Galán, F. Palero and F. González-Candelas, 2015 Recombination in viruses: mechanisms, methods of study, and evolutionary consequences. Infect Genet Evol 30: 296-307.

Ploubidou, A., and M. Way, 2001 Viral transport and the cytoskeleton. Curr Opin Cell Biol 13: 97-105.

Ponting, C. P., P. L. Oliver and W. Reik, 2009 Evolution and functions of long noncoding RNAs. Cell 136: 629-641.

Pooggin, M. M., L. A. Ryabova, X. He, J. Fütterer and T. Hohn, 2006 Mechanism of ribosome shunting in Rice tungro bacilliform pararetrovirus. RNA (New York, N.Y.) 12: 841-850.

Potgieter, A. C., M. Cloete, P. J. Pretorius and A. A. van Dijk, 2003 A first full outer capsid protein sequence data-set in the Orbivirus genus (family Reoviridae): cloning, sequencing, expression and analysis of a complete set of full-length outer capsid VP2 genes of the nine African horsesickness virus serotypes. J Gen Virol 84: 1317-1326.

Povolyaeva, O. S., A. A. Chadaeva, A. V. Lunitsin and S. G. Yurkov, 2022 [Dwarf bat's (Pipistrellus pipistrellus) lung diploid cell strains and their permissivity to orbiviruses (Reoviridae: Orbivirus) - pathogens of vector-borne animal diseases]. Vopr Virusol 67: 227-236.

Purse, B. V., and D. J. Rogers, 2009 Chapter 16 Bluetongue virus and climate change, pp. 343-364 in *Bluetongue*, edited by R. Noad and P. Roy.

Puustusmaa, M., and A. Abroi, 2019 cRegions-a tool for detecting conserved cis-elements in multiple sequence alignment of diverged coding sequences. PeerJ 6: e6176-e6176.

Quan, M., M. van Vuuren, P. Howell, D. Groenewald and A. Guthrie, 2008 Molecular epidemiology of the African horse sickness virus S10 gene. The Journal of general virology 89: 1159-1168.

Racine, T., C. Barry, K. Roy, S. J. Dawe, M. Shmulevitz *et al.*, 2007 Leaky scanning and scanning-independent ribosome migration on the tricistronic S1 mRNA of avian reovirus. J Biol Chem 282: 25613-25622.

Racine, T., and R. Duncan, 2010 Facilitated leaky scanning and atypical ribosome shunting direct downstream translation initiation on the tricistronic S1 mRNA of avian reovirus. Nucleic acids research 38: 7260-7272.

Rainsford, E. W., and M. A. McCrae, 2007 Characterization of the NSP6 protein product of rotavirus gene 11. Virus Res 130: 193-201.

Raksakoon, C., and R. Potiwat, 2021 Current arboviral threats and their potential vectors in Thailand. Pathogens 10: 1-14.

Ramadevi, N., N. J. Burroughs, P. P. Mertens, I. M. Jones and P. Roy, 1998 Capping and methylation of mRNA by purified recombinant VP4 protein of bluetongue virus. Proc Natl Acad Sci U S A 95: 13537-13542.

Ratinier, M., M. Caporale, M. Golder, G. Franzoni, K. Allan *et al.*, 2011 Identification and characterization of a novel non-structural protein of bluetongue virus. PLoS Pathog 7: e1002477.

Ratinier, M., A. E. Shaw, G. Barry, Q. Gu, L. Di Gialleonardo *et al.*, 2016 Bluetongue Virus NS4 Protein Is an Interferon Antagonist and a Determinant of Virus Virulence. J Virol 90**:** 5427-5439.

Redmond, E. F., D. Jones and J. Rushton, 2022 Economic assessment of African horse sickness vaccine impact. Equine Vet J 54**:** 368-378.

Rosano, G. L., and E. A. Ceccarelli, 2014 Recombinant protein expression in Escherichia coli: advances and challenges. Front Microbiol 5**:** 172.

Ross-Smith, N., K. E. Darpel, P. Monaghan and P. Mertens, 2009 Chapter 5 Bluetongue virus: cell biology, pp. 77-99 in *Bluetongue*, edited by R. Noad and P. Roy.

Roy, P., 2008 Bluetongue virus: dissection of the polymerase complex. J Gen Virol 89**:** 1789-1804.

Roy, P., 2017 Bluetongue virus structure and assembly. Current Opinion in Virology 24**:** 115-123.

Roy, P., 2020 Highly efficient vaccines for Bluetongue virus and a related Orbivirus based on reverse genetics. Current Opinion in Virology 44**:** 35-41.

Roy, P., D. H. Bishop, S. Howard, H. Aitchison and B. Erasmus, 1996 Recombinant baculovirus-synthesized African horsesickness virus (AHSV) outer-capsid protein VP2 provides protection against virulent AHSV challenge. J Gen Virol 77 (Pt 9)**:** 2053-2057.

Roy, P., J. Marshall and T. French, 1990 Structure of the bluetongue virus genome and its encoded proteins, pp. 43-87 in *Bluetongue viruses*. Springer.

Roy, P., P. P. Mertens and I. Casal, 1994 African horse sickness virus structure. Comp Immunol Microbiol Infect Dis 17**:** 243-273.

Rutkowska, D. A., N. B. Mokoena, T. L. Tsekoa, V. S. Dibakwane and M. M. O'Kennedy, 2019 Plant-produced chimeric virus-like particles - A new generation vaccine against African horse sickness. BMC Veterinary Research 15.

Sailleau, C., S. Moulay and S. Zientara, 1997 Nucleotide sequence comparison of the segments S10 of the nine African horsesickness virus serotypes. Archives of Virology 142**:** 965-978.

Sealfon, R. S., M. F. Lin, I. Jungreis, M. Y. Wolf, M. Kellis *et al.*, 2015 FRESCo: finding regions of excess synonymous constraint in diverse viruses. Genome Biol 16**:** 38.

Sergeant, E. S., J. D. Grewar, C. T. Weyer and A. J. Guthrie, 2016 Quantitative Risk Assessment for African Horse Sickness in Live Horses Exported from South Africa. PLoS ONE 11**:** e0151757.

Shaw, A. E., A. Bruning-Richardson, E. E. Morrison, J. Bond, J. Simpson *et al.*, 2013 Bluetongue virus infection induces aberrant mitosis in mammalian cells. Virol J 10**:** 319.

Siegrist, C. A., B. Durand, P. Emery, E. David, P. Hearing *et al.*, 1993 RFX1 is identical to enhancer factor C and functions as a transactivator of the hepatitis B virus enhancer. Mol Cell Biol 13**:** 6375-6384.

Smith, G. A., and L. W. Enquist, 2002 Break ins and break outs: viral interactions with the cytoskeleton of Mammalian cells. Annu Rev Cell Dev Biol 18**:** 135-161.

Smith, G. E., M. D. Summers and M. J. Fraser, 1983 Production of human beta interferon in insect cells infected with a baculovirus expression vector. Mol Cell Biol 3**:** 2156-2165.

Stamatakis, A., 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30**:** 1312-1313.

Stassen, L., H. Huismans and J. Theron, 2011 Membrane permeabilization of the African horse sickness virus VP5 protein is mediated by two N-terminal amphipathic alpha-helices. Arch Virol 156**:** 711-715.

Stauber, N., J. Martinez-Costas, G. Sutton, K. Monastyrskaya and P. Roy, 1997 Bluetongue virus VP6 protein binds ATP and exhibits an RNA-dependent ATPase function and a helicase activity that catalyze the unwinding of double-stranded RNA substrates. J Virol 71**:** 7220-7226.

Steward, M., I. B. Vipond, N. S. Millar and P. T. Emmerson, 1993 RNA editing in Newcastle disease virus. J Gen Virol 74 (Pt 12)**:** 2539-2547.

Stewart, M., A. Hardy, G. Barry, R. M. Pinto, M. Caporale *et al.*, 2015 Characterization of a second open reading frame in genome segment 10 of bluetongue virus. J Gen Virol 96**:** 3280-3293.

Stoltz, M. A., C. F. van der Merwe, J. Coetzee and H. Huismans, 1996 Subcellular localization of the nonstructural protein NS3 of African horsesickness virus. Onderstepoort J Vet Res 63**:** 57-61.

Stuart, D. I., and J. M. Grimes, 2006 Structural studies on orbivirus proteins and particles. Curr Top Microbiol Immunol 309**:** 221-244.

Sullivan, E., S. Lecollinet, A. Kerviel, E. Hue, S. Pronost *et al.*, 2021 Entry-competent-replication-abortive African horse sickness virus strains elicit robust immunity in ponies against all serotypes. Vaccine 39**:** 3161-3168.

Sung, P.-Y., R. Vaughan, S. K. Rahman, G. Yi, A. Kerviel *et al.*, 2019 The Interaction of Bluetongue Virus VP6 and Genomic RNA Is Essential for Genome Packaging. Journal of Virology 93**:** e02023-02018.

Sung, P. Y., and P. Roy, 2014 Sequential packaging of RNA genomic segments during the assembly of Bluetongue virus. Nucleic Acids Res 42**:** 13824-13838.

Suzuki, N., M. Sugawara, D. L. Nuss and Y. Matsuura, 1996 Polycistronic (tri- or bicistronic) phytoreoviral segments translatable in both plant and insect cells. J Virol 70**:** 8155-8159.

Theiler, A., 1921 African Horse Sickness (Pestiis equorum). Union S. Africa Dept. Agric., Pretoria, Sci. Bull.

Thomas Kluyver, B. R.-K., Fernando Perezand Brian Granger, Matthias Bussonnier , Jonathan Frederic , Kyle Kelley , Jessica Hamrick,Jason Grout , Sylvain Corlay , Paul Ivanov , Damian Avila , Safia Abdalla , Carol Willing, 2016 *Jupyter Notebooks -- a publishing format for reproducible computational workflows*. IOS Press.

Toh, X., Y. Wang, M. P. Rajapakse, B. Lee, T. Songkasupa *et al.*, 2022 Use of nanopore sequencing to characterize african horse sickness virus (AHSV) from the African horse sickness outbreak in thailand in 2020. Transbound Emerg Dis 69**:** 1010-1019.

Tourasse, N. J., and W.-H. Li, 2000 Selective Constraints, Amino Acid Composition, and the Rate of Protein Evolution. Molecular Biology and Evolution 17**:** 656-664.

Tunyasuvunakool, K., J. Adler, Z. Wu, T. Green, M. Zielinski *et al.*, 2021 Highly accurate protein structure prediction for the human proteome. Nature 596**:** 590-596.

Turner, D. H., and D. H. Mathews, 2010 NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. Nucleic Acids Res 38**:** D280-282.

Uitenweerde, J. M., J. Theron, M. A. Stoltz and H. Huismans, 1995 The Multimeric Nonstructural NS2 Proteins of Bluetongue Virus, African Horsesickness Virus, and Epizootic Hemorrhagic Disease Virus Differ in Their Single-Stranded RNA-Binding Ability. Virology 209**:** 624-632.

Urakawa, T., D. G. Ritter and P. Roy, 1989 Expression of largest RNA segment and synthesis of VP1 protein of bluetongue virus in insect cells by recombinant baculovirus: association of VP1 protein with RNA polymerase activity. Nucleic Acids Res 17**:** 7395-7401.

van de Water, S. G., R. G. van Gennip, C. A. Potgieter, I. M. Wright and P. A. van Rijn, 2015 VP2 Exchange and NS3/NS3a Deletion in African Horse Sickness Virus (AHSV) in Development of Disabled Infectious Single Animal Vaccine Candidates for AHSV. J Virol 89**:** 8764-8772.

Van Gennip, R. G. P., S. G. Van de Water and P. A. Van Rijn, 2014 Bluetongue Virus Nonstructural Protein NS3/NS3a Is Not Essential for Virus Replication. PLoS ONE 9.

Van Niekerk, M., C. C. Smit, W. C. Fick, V. Van Staden and H. Huismans, 2001a Membrane association of African horsesickness virus nonstructural protein NS3 determines its cytotoxicity. Virology 279**:** 499-508.

van Niekerk, M., V. van Staden, A. A. van Dijk and H. Huismans, 2001b Variation of African horsesickness virus nonstructural protein NS3 in southern Africa. J Gen Virol 82**:** 149-158.

van Oers, M. M., G. P. Pijlman and J. M. Vlak, 2015 Thirty years of baculovirus-insect cell protein expression: from dark horse to mainstream technology. J Gen Virol 96**:** 6-23.

van Rijn, P. A., M. A. Maris-Veldhuis, M. Grobler, I. M. Wright, B. J. Erasmus *et al.*, 2020 Safety and efficacy of inactivated African horse sickness (AHS) vaccine formulated with different adjuvants. Vaccine 38**:** 7108-7117.

Van Rijn, P. A., M. A. Maris-Veldhuis, C. A. Potgieter and R. G. P. Van Gennip, 2018 African horse sickness virus (AHSV) with a deletion of 77 amino acids in NS3/NS3a protein is not virulent and a safe promising AHS Disabled Infectious Single Animal (DISA) vaccine platform. Vaccine 36**:** 1925-1933.

Van Rossum, G. a. D., Fred L, 2009 *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.

Van Staden, V., and H. Huismans, 1991 A comparison of the genes which encode non-structural protein NS3 of different orbiviruses. J Gen Virol 72 (Pt 5)**:** 1073-1079.

Van Staden, V., C. C. Smit, M. A. Stoltz, F. F. Maree and H. Huismans, 1998 Characterization of two African horse sickness virus nonstructural proteins, NS1 and NS3. Arch Virol Suppl 14**:** 251-258.

Van Staden, V., M. A. Stoltz and H. Huismans, 1995 Expression of nonstructural protein NS3 of African horsesickness virus (AHSV): evidence for a cytotoxic effect of NS3 in insect cells, and characterization of the gene products in AHSV infected Vero cells. Arch Virol 140**:** 289-306.

Van Staden, V., J. Theron, B. J. Greyling, H. Huismans and L. H. Nel, 1991 A comparison of the nucleotide sequences of cognate NS2 genes of three different orbiviruses. Virology 185**:** 500-504.

Varadi, M., S. Anyango, M. Deshpande, S. Nair, C. Natassia *et al.*, 2022 AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res 50**:** D439-d444.

Venter, E., 2014 Factors contributing to AHSV pathogenesis in infected mammalian and insect cells, pp. in *Department of Genetics*. University of Pretoria.

Vermaak, E., A. M. Conradie, F. F. Maree and J. Theron, 2016 African horse sickness virus infects BSR cells through macropinocytosis. Virology 497**:** 217-232.

von Teichman, B., B. Dungu and T. Smit, 2010 In vivo cross-protection to African horse sickness Serotypes 5 and 9 after vaccination with Serotypes 8 and 6. Vaccine 28**:** 6505-6517.

Wall, G. V., I. M. Wright, C. Barnardo, B. J. Erasmus, V. van Staden *et al.*, 2021 African horse sickness virus NS4 protein is an important virulence factor and interferes with JAK-STAT signaling during viral infection. Virus Research 298**:** 198407.

Walsh, D., and I. Mohr, 2011 Viral subversion of the host protein synthesis machinery. Nature Reviews Microbiology 9**:** 860-875.

Wang, F., L.-Z. Sun, T. Sun, S. Chang and X. Xu, 2019 Helix-Based RNA Landscape Partition and Alternative Secondary Structure Determination. ACS Omega 4**:** 15407-15413.

Wang, J., and M. Gribskov, 2019 IRESpy: an XGBoost model for prediction of internal ribosome entry sites. BMC Bioinformatics 20**:** 409.

Wang, R., and M. G. Brattain, 2007 The maximal size of protein to diffuse through the nuclear pore is larger than 60kDa. FEBS letters 581**:** 3164-3170.

Wang, S., W. Li, S. Liu and J. Xu, 2016 RaptorX-Property: a web server for protein structure property prediction. Nucleic Acids Res 44**:** W430-435.

Weaver, S., S. D. Shank, S. J. Spielman, M. Li, S. V. Muse *et al.*, 2018 Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. Molecular Biology and Evolution 35**:** 773-777.

Wei, X., and J. Zhang, 2014 A simple method for estimating the strength of natural selection on overlapping genes. Genome Biol Evol 7**:** 381-390.

Wernery, U., S. Joseph, R. Raghavan, B. Dyer and S. Spendrup, 2020 African Horse Sickness Fever in Vaccinated Horses: Short Communication. Journal of Equine Veterinary Science 88**:** 102967.

Wernery, U., M. Rodriguez, R. Raghavan, G. Syriac, M. S. Miriam Thomas *et al.*, 2021 Humoral antibody response of 10 horses after vaccination against African horse sickness with an inactivated vaccine containing all 9 serotypes in one injection. Equine Vet J 53**:** 826-833.

Weyer, C., J. Grewar, P. Burger, C. Joone, C. Lourens *et al.*, 2017 Dynamics of African horse sickness virus nucleic acid and antibody in horses following immunization with a commercial polyvalent live attenuated vaccine. Vaccine 35**:** 2504-2510.

Weyer, C. T., J. D. Grewar, P. Burger, E. Rossouw, C. Lourens *et al.*, 2016 African horse sickness caused by genome reassortment and reversion to virulence of live, attenuated vaccine viruses, South Africa, 2004–2014. Emerging infectious diseases 22**:** 2087.

Wilson, A., P. S. Mellor, C. Szmaragd and P. P. C. Mertens, 2009 Adaptive strategies of African horse sickness virus to facilitate vector transmission. Veterinary research 40**:** 16-16.

Wirblich, C., B. Bhattacharya and P. Roy, 2006 Nonstructural protein 3 of bluetongue virus assists virus release by recruiting ESCRT-I protein Tsg101. J Virol 80**:** 460-473.

Wong, F., A. Krishnan, E. J. Zheng, H. Stärk, A. L. Manson *et al.*, 2022 Benchmarking AlphaFold-enabled molecular docking predictions for antibiotic discovery. Molecular Systems Biology 18**:** e11081.

Wulan, W. N., D. Heydet, E. J. Walker, M. E. Gahan and R. Ghildyal, 2015 Nucleocytoplasmic transport of nucleocapsid proteins of enveloped RNA viruses. Front Microbiol 6**:** 553.

Xia, X., W. Wu, Y. Cui, P. Roy and Z. H. Zhou, 2021 Bluetongue virus capsid protein VP5 perforates membranes at low endosomal pH during viral entry. Nat Microbiol 6**:** 1424-1432.

Yakovchuk, P., E. Protozanova and M. D. Frank-Kamenetskii, 2006 Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. Nucleic Acids Res 34**:** 564-574.

Yao, Z., Z. Weinberg and W. L. Ruzzo, 2006 CMfinder--a covariance model based RNA motif finding algorithm. Bioinformatics 22**:** 445-452.

Zhang, X., M. Boyce, B. Bhattacharya, X. Zhang, S. Schein *et al.*, 2010 Bluetongue virus coat protein VP2 contains sialic acid-binding domains, and VP5 resembles enveloped virus fusion proteins. Proc Natl Acad Sci U S A 107**:** 6292-6297.

Zhang, X., A. Patel, C. C. Celma, X. Yu, P. Roy *et al.*, 2016 Atomic model of a nonenveloped virus reveals pH sensors for a coordinated process of cell entry. Nat Struct Mol Biol 23**:** 74-80.

Zhao, J., Y. Li, C. Wang, H. Zhang, H. Zhang *et al.*, 2020 IRESbase: A Comprehensive Database of Experimentally Validated Internal Ribosome Entry Sites. Genomics, Proteomics & Bioinformatics 18**:** 129-139.

Zuker, M., 2003a Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31**:** 3406-3415.

Zuker, M., 2003b Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research 31**:** 3406-3415.

Zwart, L., C. A. Potgieter, S. J. Clift and V. van Staden, 2015 Characterising Non-Structural Protein NS4 of African Horse Sickness Virus. PLoS ONE 10**:** e0124281.

Appendix

Alignment of pFB recombinants to their reference sequence

A) pFB-ORF2

```
pFB-ORF2        gtttgttcgcccaggactctagctatagttctagtggttggctacgtatactccggaata
polyhedron-Forw ------------------------------------------------------------
pFB-Reverse     gtttgttcgcccaggactctagctatagttctagtggttggctacgtatactccggaata


pFB-ORF2        ttaatagatcatggagataattaaaatgataaccatctcgcaaataaataagtattttac
polyhedron-Forw ------------------------------------------------------------
pFB-Reverse     ttaatagatcatggagataattaaaatgataaccatctcgcaaataaataagtattttac


pFB-ORF2        tgttttcgtaacagttttgtaataaaaaaacctataaatattccggattattcataccgt
polyhedron-Forw ------------------------------------------------------------
pFB-Reverse     tgttttcgtaacagttttgtaataaaaaaacctataaatattccggattattcataccgt


pFB-ORF2        cccaccatcgggcgcggatccatggagagtcgggggcgatcgtcccttatgtgccaccac
polyhedron-Forw ----------------------------gggggcgatcgtcccttatgtgccaccac
pFB-Reverse     cccaccatcgggcgcggatccatggagagtcgggggcgatcgtcccttatgtgccaccac
                                            ****************************


pFB-ORF2        catacaatttcgcaagtgctccgacgttttctcagcgtacgagtcaaatggagtccgtgt
polyhedron-Forw catacaatttcgcaagtgctccsmcgttttctcagcgtacgagtcaaatggagtccgtgt
pFB-Reverse     catacaatttcgcaagtgctccgacgttttctcagcgtacgagtcaaatggagtccgtgt
                ***********************  ************************************


pFB-ORF2        cgcttgggatacttaaccaagccatgtcaagtacaactggtgcgagtggggcgcttaaag
polyhedron-Forw cgcttgggatacttaaccaagccatgtcaagtacaactggtgcgagtggggcgcttaaag
pFB-Reverse     cgcttgggatacttaaccaagccatgtcaagtacaactggtgcgagtggggcgcttaaag
                ************************************************************


pFB-ORF2        atgaaaaagcagcattcggtgctatggcggaagcattgcgtgatccagaacccatacgtc
polyhedron-Forw atgaaaaagcagcattcggtgctatggcggaagcattgcgtgatccagaacccatacgtc
pFB-Reverse     atgaaaaagcagcattcggtgctatggcggaagcattgcgtgatccagaacccatacgtc
                ************************************************************


pFB-ORF2        aaattaaaaagcaggtgggtatcagaacttaaggatcccggtccgaagcgcgcggaatt
polyhedron-Forw aaattaaaaagcaggtgggtatcagaacttaaggatcccggtccgaagcgcgcggaatt
pFB-Reverse     aaattaaaaagcaggtgggtatcagaacttaaggatcccggtccgaagcgcgcggaatt
                ************************************************************


pFB-ORF2        caaaggcctacgtcgacgagctcactagtcgcggccgctttcgaatctagagcctgcagt
polyhedron-Forw caaaggcctacgtcgacgagctcactagtcgcggccgctttcgaatctagagcctgcagt
pFB-Reverse     caaaggcctacgtcgacgagctcactagtcgcggccgctttcgaatctagagcctgcagt
                ************************************************************


pFB-ORF2        ctcgaggcatgcggtaccaagcttgtcgagaagtactagaggatcataatcagccatacc
polyhedron-Forw ctcgaggcatgcggtaccaagcttgtcgagaagtactagaggatcataatcagccatacc
pFB-Reverse     ctcgaggcatgcggkaccca-------------------------------------------
                ************** *** *


pFB-ORF2        acatttgtagaggttttacttgctttaaaaaacctcccacacctcccccgaacctgaaa
polyhedron-Forw acatttgtagaggttttacttgctttaaaaaacctcccacacctcccccgaacctgaaa
pFB-Reverse     ------------------------------------------------------------


pFB-ORF2        cataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttacaaa
polyhedron-Forw cataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttacaaa
pFB-Reverse     ------------------------------------------------------------
```

B) pFB-HT-ORF2

```
pFB-ORF2-HTA     ccggtggtgctgaccccggatgaagtggttcgcatcctcggttttctggaaggcgagcat
polyhedron-Forw  ------------------------------------------------------------
pFB-Reverse      ccggtggtgctgaccccggatgaagtggttcgcatcctcggttttctggaaggcgagcat


pFB-ORF2-HTA     cgtttgttcgcccaggactctagctatagttctagtggttggctacgtatactccggaat
polyhedron-Forw  ------------------------------------------------------------
pFB-Reverse      cgtttgttcgcccaggactctagctatagttctagtggttggctacgtatactccggaat


pFB-ORF2-HTA     attaatagatcatggagataattaaaatgataaccatctcgcaaataaataagtatttta
polyhedron-Forw  ------------------------------------------------------------
pFB-Reverse      attaatagatcatggagataattaaaatgataaccatctcgcaaataaataagtatttta


pFB-ORF2-HTA     ctgttttcgtaacagttttgtaataaaaaaacctataaatattccggattattcataccg
polyhedron-Forw  ------------------------------------------------------------
pFB-Reverse      ctgttttcgtaacagttttgtaataaaaaaacctataaatattccggattattcataccg


pFB-ORF2-HTA     tcccaccatcgggcgcggatctcggtccgaaaccatgtcgtactaccatcaccatcacca
polyhedron-Forw  ----------ggamkcgsggayyggtc--gaaccatgtcgwactaccatcaccatcacca
pFB-Reverse      tcccaccatcgggcgcggatctcggtccgaaaccatgtcgtactaccatcaccatcacca
                           **.  ** .    ****  .********** ******************

pFB-ORF2-HTA     tcacgattacgatatcccaacgaccgaaaacctgtattttcagggcgccatggagagtcg
polyhedron-Forw  tcacgattacgatatcccaacga-tgaaaammkgtattttcagggcgccatggagagtcg
pFB-Reverse      tcacgattacgatatcccaacgaccgaaaacctgtattttcagggcgccatggagagtcg
                 ********************** .*****   **************************

pFB-ORF2-HTA     ggggcgatcgtcccttatgtgccaccaccatacaatttcgcaagtgctccgacgttttct
polyhedron-Forw  ggggcgatcgtcccttatgtgccaccaccatacaatttcgcaagtgctccgacgttttct
pFB-Reverse      ggggcgatcgtcccttatgtgccaccaccatacaatttcgcaagtgctccgacgttttct
                 ************************************************************

pFB-ORF2-HTA     cagcgtacgagtcaaatggagtccgtgtcgcttgggatacttaaccaagccatgtcaagt
polyhedron-Forw  cagcgtacgagtcaaatggagtccgtgtcgcttgggatacttaaccaagccatgtcaagt
pFB-Reverse      cagcgtacgagtcaaatggagtccgtgtcgcttgggatacttaaccaagccatgtcaagt
                 ************************************************************

pFB-ORF2-HTA     acaactggtgcgagtggggcgcttaaagatgaaaaagcagcattcggtgctatggcggaa
polyhedron-Forw  acaactggtgcgagtggggcgcttaaagatgaaaaagcagcattcggtgctatggcggaa
pFB-Reverse      acaactggtgcgagtggggcgcttaaagatgaaaaagcagcattcggtgctatggcggaa
                 ************************************************************

pFB-ORF2-HTA     gcattgcgtgatccagaacccatacgtcaaattaaaaagcaggtgggtatcagaacttta
polyhedron-Forw  gcattgcgtgatccagaacccatacgtcaaattaaaaagcaggtgggtatcagaacttta
pFB-Reverse      gcattgcgtgatccagaacccatacgtcaaattaaaaagcaggtgggtatcagaacttta
                 ************************************************************

pFB-ORF2-HTA     aggatccggaattcaaaggcctacgtcgacgagctcaactagtgcggccgctttcgaatc
polyhedron-Forw  aggatccggaattcaaaggcctacgtcgacgagctcaactagtgcggccgctttcgaatc
pFB-Reverse      aggatccggaattcaaaggcctacgtcgacgagstcaactagtgcggccgctttcgaatc
                 ********************************* ************************

pFB-ORF2-HTA     tagagcctgcagtctcgaggcatgcggtaccaagcttgtcgagaagtactagaggatcat
polyhedron-Forw  tagagcctgcagtctcgaggcatgcggtaccaagcttgtcgagaagtactagaggatcat
pFB-Reverse      tagagcctgcagtctcgaggcatgcgktwccaagytk-----------------------
                 ************************** * ***** *

pFB-ORF2-HTA     aatcagccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctccc
polyhedron-Forw  aatcagccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctccc
pFB-Reverse      ------------------------------------------------------------
```

C) pFB-ORF2-eGFP

```
pFB-ORF2-eGFP   gttcgcccaggactctagctatagttctagtggttggctacgtatactccggaatattaa
pFB-Reverse     ------------------------------------------------------------ttaa
polyhedron-Forw ------------------------------------------------------------


pFB-ORF2-eGFP   tagatcatggagat-aattaaaatgataaccatctcgcaaataaataagtattttactgt
pFB-Reverse     tagatcatggagataaattaaaatgata--catctggcaataaaataagta-tttactgt
polyhedron-Forw ------------------------------------------------------------


pFB-ORF2-eGFP   tttcgtaacagttttgtaat--aaaaaaacctat-aaatattccggattattcataccgt
pFB-Reverse     tttcgtaacagttttgtaataaaaaaaaacctataaaatattccggattattcataccgt
polyhedron-Forw ------------------------------------------------------------


pFB-ORF2-eGFP   cccaccatcgggcgcggatcccggtccgaagcgcgcggaattcatggagagtcgggggcg
pFB-Reverse     cccaccatcgggcgcggatcccggtccgaagcgcgcggaattcatggagagtcgggggcg
polyhedron-Forw -------csssssgggaayccgggtcgaagcgcgcggaattcatggagagtcgggggcg
                       .       ***  **** .**********************************

pFB-ORF2-eGFP   atcgtcccttatgtgccaccaccatacaatttcgcaagtgctccgacgttttctcagcgt
pFB-Reverse     atcgtcccttatgtgccaccaccatacaatttcgcaagtgctccgacgttttctcagcgt
polyhedron-Forw atcgtcccttatgtgccaccaccatacaatttcgcaagtgctccgacgttttctcagcgt
                ************************************************************

pFB-ORF2-eGFP   acgagtcaaatggagtccgtgtcgcttgggatacttaaccaagccatgtcaagtacaact
pFB-Reverse     acgagtcaaatggagtccgtgtcgcttgggatacttaaccaagccatgtcaagtacaact
polyhedron-Forw acgagtcaaatggagtccgtgtcgcttgggatacttaaccaagccatgtcaagtacaact
                ************************************************************

pFB-ORF2-eGFP   ggtgcgagtggggcgcttaaagatgaaaaagcagcattcggtgctatggcggaagcattg
pFB-Reverse     ggtgcgagtggggcgcttaaagatgaaaaagcagcattcggtgctatggcggaagcattg
polyhedron-Forw ggtgcgagtggggcgcttaaagatgaaaaagcagcattcggtgctatggcggaagcattg
                ************************************************************

pFB-ORF2-eGFP   cgtgatccagaacccatacgtcaaattaaaaagcaggtgggtatcagaacttgtgagcaa
pFB-Reverse     cgtgatccagaacccatacgtcaaattaaaaagcaggtgggtatcagaacttgtgagcaa
polyhedron-Forw cgtgatccagaacccatacgtcaaattaaaaagcaggtgggtatcagaacttgtgagcaa
                ************************************************************

pFB-ORF2-eGFP   gggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaa
pFB-Reverse     gggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaa
polyhedron-Forw gggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaa
                ************************************************************

pFB-ORF2-eGFP   cggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgac
pFB-Reverse     cggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgac
polyhedron-Forw cggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgac
                ************************************************************

pFB-ORF2-eGFP   cctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccac
pFB-Reverse     cctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccac
polyhedron-Forw cctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccac
                ************************************************************

pFB-ORF2-eGFP   cctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgactt
pFB-Reverse     actgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgactt
polyhedron-Forw actgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgactt
                 ***********************************************************

pFB-ORF2-eGFP   cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacga
pFB-Reverse     cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacga
polyhedron-Forw cttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacga
                ************************************************************

pFB-ORF2-eGFP   cggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcat
pFB-Reverse     cggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcat
polyhedron-Forw cggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcat
```

```
                 ************************************************************

pFB-ORF2-eGFP    cgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagta
pFB-Reverse      cgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagta
polyhedron-Forw cgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagta
                 ************************************************************

pFB-ORF2-eGFP    caactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggt
pFB-Reverse      caactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggt
polyhedron-Forw caactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggt
                 ************************************************************

pFB-ORF2-eGFP    gaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactacca
pFB-Reverse      gaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactacca
polyhedron-Forw gaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactacca
                 ************************************************************

pFB-ORF2-eGFP    gcagaacaccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcac
pFB-Reverse      gcagaacaccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcac
polyhedron-Forw gcagaacaccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcac
                 ************************************************************

pFB-ORF2-eGFP    ccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagtt
pFB-Reverse      ccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcckgctggagtt
polyhedron-Forw ccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagtt
                 ************************************************* *********

pFB-ORF2-eGFP    cgtgaccgccgccgggatcactctcggcatggacgagctgtacaagtaaaagcttgtcga
pFB-Reverse      cgtgaccgccgccgggatcactctcggcatggacgagctgtacaag-aaaagcykycsgr
polyhedron-Forw cgtgaccgccgccgggatcactctcggcatggacgagctgtacaagtaaaagcttgtcga
                 ********************************************** ******   . *

pFB-ORF2-eGFP    gaagtactagaggatcataatcagccataccacatttgtagaggttttacttgcttt-aa
pFB-Reverse      rwrrc-------------------------------------------------------
polyhedron-Forw gaagtactagaggatcat-atcagccataccacattttgtagagttttacttgctttaaa


pFB-ORF2-eGFP    aaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaattgttgttgtta
pFB-Reverse      ------------------------------------------------------------
polyhedron-Forw aaaacctcccacactcccccct--gatctgaaacat--aatgaatgcagt-----------
```

Fig. 1: MAFFT analysis of DNA sequences of recombinant vectors to be used in the baculoviral expression system. A) pFB-ORF2 alignment. **B)** pFB-HT-ORF2 alignment. **C)** pFB-ORF2-eGFP alignment. Plasmids were sequenced with the pFB reverse primer and polyhedrin forward primer respectively. ORF2 nucleotide sequences are in blue, His-tag sequences in red and eGFP sequences in green. * Indicate matches.

Alignment of phCMV-ORF2-eGFP recombinant to the reference sequence

```
phCMV-ORF2-eGFP gcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgg
ORF2_insert_for ------------------------------------------------------------
ORF2_insert_rev gcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgg


phCMV-ORF2-eGFP gagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgcccc
ORF2_insert_for ------------------------------------------------------------
ORF2_insert_rev gagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgcccc


phCMV-ORF2-eGFP attgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctggttt
ORF2_insert_for ------------------------------------------------------------
ORF2_insert_rev attgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctggttt


phCMV-ORF2-eGFP agtgaaccgtcagatcgcctgctcacgccatccacgctgttttgaacctccatagaagac
```

```
ORF2_insert_for ------------------------------------------------------------
ORF2_insert_rev agtgaaccgtcagatcgcctgctcacgccatccacgctgttttgaacctccatagaagac


phCMV-ORF2-eGFP accgggacggtaccaaaaccatgggagacgatggagagtcgggggcgatcgtcccttatg
ORF2_insert_for ------------------------------------------------------------
ORF2_insert_rev accgggacggtaccaaaaccatgggagacgatggagagtcgggggcgatcgtcccttatg


phCMV-ORF2-eGFP tgccaccaccatacaatttcgcaagtgctccgacgttttctcagcgtacgagtcaaatg
ORF2_insert_for ------------------------------------------------------------
ORF2_insert_rev tgccaccaccatacaatttcgcaagtgctccgacgttttctcagcgtacgagtcaaatg


phCMV-ORF2-eGFP gagtccgtgtcgcttgggatacttaaccaagccatgtcaagtacaactggtgcgagtggg
ORF2_insert_for ------------------------------------------------------------
ORF2_insert_rev gagtccgtgtcgcttgggatacttaaccnannnntgtcaagtacaactggtgcgagtggg


phCMV-ORF2-eGFP gcgcttaaagatgaaaaagcagcattcggtgctatggcggaagcattgcgtgatccagaa
ORF2_insert_for ------------------------------------------------------------
ORF2_insert_rev gcgnnnnnnnnnnnnnnagcagcattcggtgctatggcggaagcattgcgtgatccagaa


phCMV-ORF2-eGFP cccatacgtcaaattaaaaagcaggtgggtatcagaacttggaggaggaggttctggagg
ORF2_insert_for ------------------------------------------------------------
ORF2_insert_rev ccnnnnnnnnnnnnnnncnnnnnnn------------------------------------


phCMV-ORF2-eGFP aggaggtagtggcggaggggttctgtgagtaaaggtgaagagctgttcaccggcgtggt
ORF2_insert_for ------------------------------------------------------------
ORF2_insert_rev ------------------------------------------------------------


phCMV-ORF2-eGFP ccctatcctggtggagctggacggagatgtcaatgggcataagttttccgtgagcgggga
ORF2_insert_for ------------------------------------------------------------
ORF2_insert_rev ------------------------------------------------------------


phCMV-ORF2-eGFP aggcgagggcgatgctacatatggaaagctgactctgaaattcatttgcaccacagggaa
ORF2_insert_for ------------------------------------------------------------
ORF2_insert_rev ------------------------------------------------------------


phCMV-ORF2-eGFP gctgcccgtgccttggccaaccctggtcactaccctgacatacggcgtgcagtgttttag
ORF2_insert_for ------------------------------------------------------------
ORF2_insert_rev ------------------------------------------------------------


phCMV-ORF2-eGFP ccggtatcctgaccacatgaagcagcatgatttctttaaatctgcaatgccagaaggata
ORF2_insert_for ------------------------------------------------------------
ORF2_insert_rev ------------------------------------------------------------


phCMV-ORF2-eGFP cgtgcaggagagaacaatcttctttaaggacgatggcaactataaaaactcgcgccgaagt
ORF2_insert_for ------------------------------------------------------------
ORF2_insert_rev ------------------------------------------------------------


phCMV-ORF2-eGFP gaagttcgagggtgacaccctggtcaaccgaatcgagctgaaggggatcgacttcaagga
ORF2_insert_for ------------------------------------------------------------
ORF2_insert_rev ------------------------------------------------------------


phCMV-ORF2-eGFP agatggtaacatcctgggccacaagctggagtacaactataatagtcataacgtgtacat
ORF2_insert_for ------------------------------------------------------------
ORF2_insert_rev ------------------------------------------------------------


phCMV-ORF2-eGFP catggctgataagcagaaaaacggaattaaggtcaacttcaagatccgccacaatattga
```

```
ORF2_insert_for catggctgataagcagaaaaacggaattaaggtcaacttcaagatccgccacaatattga
ORF2_insert_rev ------------------------------------------------------------


phCMV-ORF2-eGFP ggacggctctgtgcagctggcagatcattaccagcagaacacaccaatcggcgacggccc
ORF2_insert_for ggacggctctgtgcagctggcanatcattaccagcagaacacaccaatcggcgacggccc
ORF2_insert_rev ------------------------------------------------------------


phCMV-ORF2-eGFP cgtgctgctgcctgataatcactatctgtcaactcagtccgccctgagcaaggaccctaa
ORF2_insert_for cgtgctgctgcctgataatcactatctgtcaactcagtccgccctgagcaaggaccctaa
ORF2_insert_rev ------------------------------------------------------------


phCMV-ORF2-eGFP cgagaaacgagatcatatggtgctgctggaatttgtcaccgccgctggtatcaccctggg
ORF2_insert_for cgagaaacgagatcatatggtgctgctggaatttgtcaccgccgctggtatcaccctggg
ORF2_insert_rev ------------------------------------------------------------


phCMV-ORF2-eGFP gatggatgaactgtataaatgacgtctcatagtaaataaagcaatagcatc-acaaattt
ORF2_insert_for gatggatgaactgnataaatgacgtctcatantaaannaagcaanagcatcnnnaaattt
ORF2_insert_rev ------------------------------------------------------------


phCMV-ORF2-eGFP cacaaataaagcatttttttcacgaattc-------------------------------
ORF2_insert_for cacaaataaagcatttttttcncnantnnnagcntngatatcattcnnnngnnnnnnnnn
ORF2_insert_rev ------------------------------------------------------------
```

Fig. 2: MAFFT analysis of DNA sequences of recombinant phCMV-ORF2-eGFP plasmid used in the transient mammalian expression system. Plasmids were sequenced with the phCMV-ORF2-eGFP insert primers. ORF2 nucleotide sequences are in blue, the linker sequences in pink and the eGFP sequences in green. * Indicate matches.

Codon-optimised AHSV Seg-10 ORF2 gene linked to:

A) Strep-tag®II

```
1       catatggaat cccgcggtcg ctcctccctg atgtgccacc accataccat tagccaagtt
61      ctgcgtcgtt ttctgtccgt tcgtgtcaag tggagcccgt gccgtctggg ttacctgacc
121     aagccgtgcc aggtgcagct ggtgcgtgtt ggccgtctga agatgaagaa acagcacagc
181     gttctgtggc gtaaacactg cgtgatccag aacccgtacg ttaagctgaa aagccgtagc
241     gtgagcgagc acagcgcttg gtctcaccct cagttcgaaa agtaataact cgag
```

B) Twin-Strep-tag®

```
1       catatggaat cccgcggtcg ctcctccctg atgtgccacc accataccat tagccaagtt
61      ctgcgtcgtt ttctgtccgt tcgtgtcaag tggagcccgt gccgtctggg ttacctgacc
121     aagccgtgcc aggtgcagct ggtgcgtgtt ggccgtctga agatgaagaa acagcacagc
181     gttctgtggc gtaaacactg cgtgatccag aacccgtacg ttaagctgaa aagccgtagc
241     gtgagcgagc acagcgcttg gagccacccg cagttcgaaa aaggtggcgg ttccggtggt
301     ggcagcggtg gcagcgcctg gtcccatccg cagtttgaga aataataact cgag
```

Fig. 3: Codon-optimised AHSV Seg-10 ORF2 (blue) gene linked in frame at the 3' end to either a Strep-tag®II (red) or Twin-Strep-tag® (Orange). These were cloned into pUC57 by Genscript.

Alignment of pStaby recombinants to their reference sequence

A) pStaby-ORF2-S

```
pStaby-ORF2-S   taatacgactcactataggggaattgtgagcggataacaattcccctctagaaataatttt
R               taatacgactcactataggggaattgtgagcggataacaattcccctctagaaataatttt
F               ------------------------------------------------------------


pStaby-ORF2-S   tgtttaactttaagaaggagatatacatatggaatcccgcggtcgctcctccctgatgtg
R               tgtttaactttaagaaggagatatacatatggaatcccgcggtcgctcctccctgatgtg
F               ---------------------------------------nnnnnnnnnnnnnnngng
                                                                    * *

pStaby-ORF2-S   ccaccaccataccattagccaagttctgcgtcgttttctgtccgttcgtgtcaagtggag
R               ccaccaccataccattagccaagttctgcgtcgttttctgtccgttcgtgtcaagtggag
F               nnnnnnccataccnttagccaagttctgcgtcgttttctgtccgttcgtgtcaagtggag
                      ******* **************************************************

pStaby-ORF2-S   cccgtgccgtctgggttacctgaccaagccgtgccaggtgcagctggtgcgtgttggccg
R               cccgtgccgtctgggttacctgaccaagccgtgccaggtgcagctggtgcgtgttggccg
F               cccgtgccgtctgggttacctgaccaagccgtgccaggtgcagctggtgcgtgttggccg
                ************************************************************

pStaby-ORF2-S   tctgaagatgaagaaacagcacagcgttctgtggcgtaaacactgcgtgatccagaaccc
R               tctgaagatgaagaaacagcacagcgttctgtggcgtaaacactgcgtgatccagaaccc
F               tctgaagatgaagaaacagcacagcgttctgtggcgtaaacactgcgtgatccagaaccc
                ************************************************************

pStaby-ORF2-S   gtacgttaagctgaaaagccgtagcgtgagcgagcacagcgcttggtctcaccctcagtt
R               gtacgttaagctgaaaagccgtagcgtgagcgagcacagcgcnnnnnnnnnnnnnnnnnn
F               gtacgttaagctgaaaagccgtagcgtgagcgagcacagcgcttggtctcaccctcagtt
                ****************************************

pStaby-ORF2-S   cgaaaagtaataactcgagcaccaccaccaccaccactgagatccggctgctaa------
R               nnnnnnnnn-------------------------------------------------
F               cgaaaagtaataactcgagcaccaccaccaccaccactgagatccggctgctaacaaagc
```

B) pStaby-ORF2-TST

```
pStaby-ORF2-TST ataattttgtttaactttaagaaggagatatacatatggaatcccgcggtcgctcctccc
R               ataattttgtttaactttaagaaggagatatacatatggaatcccgcggtcgctcctccc
F               ---------------------------------------nnnnnnnnnnnnnnnnnn

pStaby-ORF2-TST tgatgtgccaccaccataccattagccaagttctgcgtcgttttctgtccgttcgtgtca
R               tgatgtgccaccaccataccattagccaagttctgcgtcgttttctgtccgttcgtgtca
F               nnnnnnnnnnnnnccatacnntnagccaagttctgcgtcgttttctgtccgttcgtgtcn
                      ******  * ************************************

pStaby-ORF2-TST agtggagcccgtgccgtctgggttacctgaccaagccgtgccaggtgcagctggtgcgtg
R               agtggagcccgtgccgtctgggttacctgaccaagccgtgccaggtgcagctggtgcgtg
F               agtggagcccgtgccgtctgggttacctgaccaagccgtgccaggtgcagctggtgcgtg
                ************************************************************

pStaby-ORF2-TST ttggccgtctgaagatgaagaaacagcacagcgttctgtggcgtaaacactgcgtgatcc
R               ttggccgtctgaagatgaagaaacagcacagcgttctgtggcgtaaacactgcgtgatcc
F               ttggccgtctgaagatgaagaaacagcacagcgttctgtggcgtaaacactgcgtgatcc
                ************************************************************

pStaby-ORF2-TST agaacccgtacgttaagctgaaaagccgtagcgtgagcgagcacagcgcttggagccacc
R               agaacccgtacgttaagctgaaaagccgtagcgtgagcgagcacagcgcttggagccacc
F               agaacccgtacgttaagctgaaaagccgtagcgtgagcgagcacagcgcttggagccacc
                ************************************************************
```

```
pStaby-ORF2-TST cgcagttcgaaaaaggtggcggttccggtggtggcagcggtggcagcgcctggtcccatc
R               cgcagttcgaaaaaggtggcggttccggtgggnggcagcggnggcagcgcnnnnnnnnnnn
F               cgcagttcgaaaaaggtggcggttccggtggtggcagcggtggcagcgcctggtcccatc
                ****************************** ******** ********

pStaby-ORF2-TST cgcagtttgagaaataataactcgagcaccaccaccaccaccactgagatccggctgcta
R               nnnnnnnnn--------------------------------------------------
F               cgcagtttgagaaataataactcgagcaccaccaccaccaccactgagatccggctgcta
```

Fig. 4: MAFFT analysis of DNA sequences of recombinant pStaby vectors used in bacterial expression system. Plasmids were sequenced with the pStaby-ORF2-S and pStaby-ORF2-TST insert primers, respectively. ORF2 nucleotide sequences are in blue, and the Strep-tag II and Twin-strep-tag sequences are in red. * Indicate matches.