Characterizing the effect of mutation, recombination and reassortment on the genetic diversity of African horse sickness virus genomes

Harry Gay Ngoveni

Characterizing the effect of mutation, recombination and reassortment on the genetic diversity of African horse sickness virus genomes

By

Harry Gay Ngoveni

A thesis submitted in partial fulfilment of the requirements for the degree Doctor of Philosophy Veterinary Science Veterinary Tropical Diseases, in the Faculty of Veterinary Science at the University of Pretoria, Republic of South Africa

Submission Date: 2019 March

DECLARATION

I, Ngoveni Harry Gay declare that the thesis, which I hereby submit for the degree Doctor of Philosophy Veterinary Science Veterinary Tropical Diseases, in the Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, is my own work and has not been previously submitted by me for degree purposes at another tertiary institution.

_____ 2019 March 25

Ngoveni Harry Gay Date

DEDICATION

I dedicate this thesis to my late father Rhiyamoya Ngoveni, my late sister Thandi Ngoveni and my late daughter Nhlonhletelo Ngoveni.

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation and gratitude to the Agricultural Research Council Onderstepoort Veterinary Institute for affording me an opportunity to further my studies. Thanks for the financial support of the Economic Competitiveness Support Program in 2013-2015 and the PDP in 2016.

A special thanks to my supervisor Dr Otto Koekemoer for seeing a potential and believing in me, his mentoring and scientific contribution in making sure this study becomes a success.

A great word of appreciation to my co-supervisor Dr. Antoinette van Schalkwyk whom I regard as the great brains behind the success of this study. Thanks for your great contributions and encouragements.

A special recognition to my fellow colleagues in AHSV vaccine project Dr. Maryke Ferreira, Dr. Sonja Maree, Dr. Mathilde Schade-Weskott and Mr. Nick Mokotoane.

A word of appreciation to my colleagues at the ARC-OVR Department of Virology, especially Mrs. Thabi Tshabalala for making sure I learn and understand beyond the scope of my research project.

A special thanks to my lovely wife, Mrs. Tendani Ngoveni for her support and encouragement. Thanks to my two beautiful daughters, Nsuku and Vukona for making sure I continue playing my role as a father while busy writing this thesis.

Thanks to mhani N'wa-Joseph Shiviti-Ngoveni, Malume BG Shiviti and to Kokwani N'wa-Sochangaan Rikhotso-Shiviti for their support all those years.

Last, but not least, a special word of praise to my God for giving me strength to pass through all challenges. Amen.

TABLE OF CONTENTS

THESIS SUMMARY

Characterizing the effect of mutation, recombination and reassortment on the genetic diversity of African horse sickness virus genomes

By

H.G Ngoveni

Promoter : Dr O. Koekemoer
Co-promoter : Dr A van Scalkwyk
Department : Veterinary Tropical Diseases of the University of Pretoria
Degree : PhD in Veterinary Science Veterinary Tropical Diseases

African horse sickness virus (AHSV) is a member of the *Orbivirus* genus within the *Reoviridae* family and the aetiological agent for African horse sickness (AHS). It is an arthropod-borne virus transmitted by biting midges belonging to various *Culicoides* spp. Nine antigenically distinct AHSV (AHSV 1-9) serotypes have been identified and all of them are endemic to South Africa. The AHSV genome comprises of ten double stranded RNA segments and is replicated by an RNA-dependent-RNA polymerase which lacks proof-reading abilities and is therefore prone to mutations. The virus could potentially undergo genetic recombination and reassortment during the co-infection of the same host cells by at least two different virus strains of the same species. Since it is an arbovirus, we proposed that the virus would evolve slower than other RNA viruses that require a single host for replication, due to limitations imposed by its ability to successfully infect and replicate in both insect vector and vertebrate hosts. The focus of this study was to investigate the evolutionary dynamics of AHSV over a period of more than 60 years. The effects of selection, substitution rate and recombination over time were determined. African horse sickness virus isolates from the 1960's to 2014, archived at the OIE World Reference Centre for AHSV and BTV at the ARC-OVI were propagated in cell cultures, double stranded RNA (dsRNA) was extracted and reverse transcribed to cDNA. The amplified cDNA was submitted for Illumina Next Generation Sequencing and full length genome consensus sequences were generated. The assembled genomes were deposited in GenBank and used in the subsequent bioinformatics analyses.

These included phylogenetic analyses, determining substitution rates, selection pressure, intragenic recombination, and reassortment.

All ten segments were predicted to be under purifying selection pressure (dN/dS < 1), while four of the segments included selected sites under positive selection pressure (0.1 significance level). The Bayesian coalescent estimates of mean substitution rates for segments-2, -6, -7, and -9 were the highest of the AHSV genome segments, between 1.75 $x10^{-4}$ and 6.4 x 10^{-4} substitutions/ site/ year. The recorded substitutions rates were similar to that (0.52 x 10^{-4} and 6.9 x 10^{-4} substitutions/ site/ year) observed in BTV. Estimated substitution rates ranged between 1.5 – 6.4 $x10^{-4}$ substitutions/ site/ year over the whole genome. The calculated time to most recent common ancestor, differs significantly among the genome sequences, indicating the influence of reassortment or recombination on the genomes. Using RDP4, Bootscan and Simplot programs, intragenic recombination events were predicted in seg-1, seg-6, seg-7 and seg-10. These included both single and double cross-over events. Genetic recombination appears to be less frequent than in bluetongue virus (BTV). Widespread reassortment events were detected, including between vaccine strains and wild type viruses. Similar to BTV, all ten segments were predicted to evolve under strong purifying selection with selected sites under positive selection. The high percentage sequence identity within serotypes reflects the strong selective constraint imposed on arboviruses by the necessity to replicate in both vertebrate host and vector species. This study also reports on a widespread genomic reassortment of African horse sickness virus, including between wild-type and vaccine viruses and provides the first evidence of intragenic recombination.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

^{o}C	Degree Celsius
µl	microliter
%	percent or percentage
ω	substitution rate
3’	three-end prime
5’	five-end prime
AHS	African horse sickness
AHSV	*African Horse sickness virus*
AGID	Agar gel immunodiffusion
ARC-OVI	Agricultural Research Council – Onderstepoort Veterinary Institute
BEAST	Bayesian Evolutionary Analysis by Sampling Tree
BEUti	Bayesian Evolutionary Analysis Utility
BHK_{21}	Baby hamster kidney cells
BLAST	Basic Local Alignment Search Tool
bp	base pair
BSA	bovine serum albumin
BSR	a derivative of the hamster kidney cell line
BTV	*Bluetongue virus*
CFT	complement fixation test
$ChCl_4$	Chloroform
CO_2	carbon dioxide
CPE	cytopathic effect
CF	Complement fixation
cDNA	complimentary DNA
DISC	Disabled infectious single cycle
DMEM	Dulbecco's Modified Eagle's Medium
d_N	nonsynonymous
d_S	synonymous substitution
ds-cDNA	double stranded complimentary DNA
dsRNA	double stranded ribonucleic acid
EEV	*equine encephalosis virus*
EHDV	*epizootic hemorrhagic disease virus*
ELISA	Enzyme linked immunosorbent assay

EtBr	Ethidium bromide
FBS	fetal bovine serum
GIV	Great Island virus
GTR	General Time Reversal
HyPhy	hypothesis testing using phylogenies
HCl	hydrochloric acid
HPLC	high performance liquid chromatography
ICTV	International Committee for the Taxonomy of Viruses
I-ELISA	indirect ELISA
IFA	Indirect fluorescent antibody
IgG	immunoglobulin gamma
IPTG	isopropyl-beta-D-thiogalactopyranoside
ICTV	International Committee for the Taxonomy of Viruses
KC	Culicoides cells
LiCl	lithium chloride
LAV	Live attenuated vaccine
Maxchi	Maximum Chi-squared Test
MCMC	Markov chain Monte Carlo
MCS	multiple cloning site
MRCA	most recent common ancestor
ml	millilitres
ML	Maximum-likelihood
Min	Minutes
MMOH	Methyl Mercury hydroxide
mRNA	messenger RNA
MTT	3-(4,5-Dimethylthiazol-2-Yl)-2,5-Diphenyltetrazolium Bromide
NCBI	National Centre for Biotechnology Information
NaOH	sodium hydroxide
Ng	nanogram
NGS	Next generation sequencing
NS	Non-structural protein
OBP	Onderstepoort Biological Products
OVI	Onderstepoort Veterinary institute
OIE	World Organisation for Animal Health

PBS	phosphate buffered saline
PCR	Polymerase chain reaction
pH	degree of acidity and alkalinity
PAGE	Polyacrylamide gel electrophoresis
PEG	poly-etheleneglycol
PI	time interval
PSA	Penicillin-Streptomycin
RDP	Recombination Detection Program
r.p.m	revolutions per minute
SDS	sodium dodecyl sulphate
Sec	Seconds
Seg	segment
SCRV	Saint Croix River virus
ssRNA	single stranded RNA
TAE	Tris-acetate-EDTA
TBE	Tris-Borate-EDTA
$TCID_{50}$	50% Tissue Culture Infective Dose
TMRCA	Time to Most Recent Common Ancestor
TRI	guanidinium isothiocyanate
UPGMA	Unweighted Pair Method with Arithmetic Mean
UTR	untranslated region
Vero	African green monkey kidney
VN	Virus neutralization
VIBs	Viral inclusion bodies
VP	Viral protein
VLP	Virus like particles

CHAPTER 1 LITERATURE REVIEW

1.1 INTRODUCTION

African horse sickness virus (AHSV) is a member of the *Orbivirus* genus within the *Reoviridae* family (Attoui *et al.*, 2012). African horse sickness virus causes a non-contagious, but highly infectious arthropod-borne disease of equids known as African horse sickness (AHS). Naïve equids such as horses, mules and donkeys are susceptible to AHSV, with mortality rates in horses that can exceed 90% (Theiler, 1921), while zebras are considered to be vertebrate reservoirs (Mellor and Hamblin, 2004). African horse sickness is a World Organization for Animal Health (OIE) listed disease due to its devastating economic effects and the potential to spread from endemic to non-endemic regions (Mellor and Hamblin, 2004).

In contrast to bluetongue virus (BTV), the prototype virus of the genus Orbivirus, insufficient data concerning the evolution of AHSV are available. Nine antigenically distinct AHSV (AHSV 1-9) serotypes have been identified so far (McIntosh, 1958; Howell, 1963), in contrast to the 27 serotype of BTV (Niedbalski, 2013). Both AHSV and BTV contain double-stranded RNA genomes of 10 segments (Bremer, 1976) and similar to other RNA viruses, their RNA polymerase lacks a proof-reading mechanism to correct errors occurring during replication (Fleischmann, 1996; Manrubia and Lázaro, 2006; Makadiya, 2007). Similar to BTV, AHSV is expected to be able to undergo intragenic recombination during concurrent infection of the same host cells (He *et al*., 2010). Apart from intragenic recombination, co-infection can lead to a genome reassortment during virus packaging (Mellor and Hamblin, 2004; von Teichman and Smit, 2008). Intragenic recombination and reassortment could lead to the formation of new strains or phenotype variants, but the influence of this on AHSV evolution should be determined (Manrubia and Lázaro, 2006, Boulila, 2011).

1.2 AFRICAN HORSE SICKNESS (AHS)

1.2.1 African horse sickness historical perspective of outbreaks

The first historical reference to the disease that resembles AHS was published in an Arabic document ‘Le Kitậb El-Akouậ El-Kafiah Wa El Chafiậh’, which describes an outbreak of scourge recorded in Yemen as early as A.D 1327-8 (Henning, 1956

taken from Moulé, 1890). African horse sickness disease is believed to have originated in Africa (Mellor and Hamblin, 2004), with the earliest reference of the disease documented by Father Monclaro in 1569 on his journey from Portugal, exploring central and east Africa (Theiler, 1921). In 1719 an outbreak resulting in the death of 1700 horses was documented in the Cape of Good Hope (Theiler, 1921). According to the records, the most severe outbreak of AHS occurred in 1854-55 in the Cape Colony with the estimated number of horse fatalities reaching nearly 70 000 (Theiler, 1921). Following the outbreak of 1854-55, AHS re-emerged in 1913-14 in Uitenhage and Humansdorp extending to George and Uniondale (Theiler, 1921).

African horse sickness has been reported in non-endemic countries surrounding the Mediterranean Sea, the Middle East and in the Indian subcontinent (Henning, 1956; Rafyi, 1961; Sellers *et al.,* 1977; Losos, 1986). In 1930, 1944 and 1958 AHS was reported in Yemen, the Middle East and Egypt, respectively (McIntosh, 1958; Mirchamsy and Hazrati, 1973). Between 1959 and 1961 AHS was reported in Persian Gulf regions, Afghanistan, Pakistan, India and the Middle East, causing deaths of an estimated 300 000 horses (Mirchamsy and Hazrati, 1973; Losos, 1986). In 1965 and 1966 AHS was reported in Morocco, Libya, Algeria, Tunisia and Spain (Mirchamsy and Hazrati, 1973; Losos, 1986). African horse sickness re-emerged in Spain and also spread to parts of Portugal in 1987-1990 after the introduction of infected zebras to Spain (Mellor and Hamblin, 2004; Roy, 2004).

1.2.2 AHSV transmission

The transmission of AHSV from one vertebrate host to another requires an insect vector. More than 50 arboviruses, including both the AHSV and BTV, are transmitted to their vertebrate hosts by female *Culicoides* midges feeding on blood as a source of protein prior to laying of eggs (Du Toit, 1944). There are more than 1500 *Culicoides* species identified to date and approximately 30 species are capable of transmitting AHSV, with *C. imicola* identified as the principal vector in South Africa (Du Toit, 1944; Wilson *et al*., 2009; Mellor and Hamblin, 2004). In the chilly areas of South Africa, *C. bolitinos* is the predominant species associated with AHSV transmission (Meiswinkel and Paweska, 2003). *Culicoides* species capable of transmitting arboviruses have been observed on every continent except Antarctica

(Mellor *et al.* 2000). The wide geographical distribution, large population numbers and *Culicoides'* ability to consume multiple blood meals within a female's lifetime, increases the risk of enzootic outbreaks (Mellor *et al.,* 2000). Despite the increased risk in disease distribution due to insect transmission, the evolution of arboviruses is slower than single-host RNA viruses, due to the purifying selective pressure associated with dual vector and host replication cycles (Novella *et al.,* 1999). Even though the rate of genetic drift is lower in arboviruses, the insects are capable of acquiring and amplifying minor variants from a population, thus contributing to the founder effect of different or new viruses in the quasispecies population (Bonneau *et al.,* 2001).

1.2.3 Epidemiology

African horse sickness disease occurs regularly throughout the African continent, but is endemic in central, eastern and southern Africa (Henning, 1956; Hamblin *et al*. 1990). Outbreaks outside the African continent were predominantly caused by serotype 9, except for the Spanish-Portugal outbreaks in 1987-1990, which were due to imported zebra infected with serotype 4 (AHSV-4) (Mellor and Hamblin, 2004). African horse sickness virus is confined within the endemic regions during winter, but can spread to non-endemic regions during the wet summer months, due to the extension of favourable climatic conditions for the breeding of *Culicoides* species (Hening, 1956). In South Africa, both bluetongue and African horse sickness are prevalent during the summer months, due to high numbers of infected insects (Du Toit, 1944). Even though the survival rate of *Culicoides imicola* is likely to decrease at high temperatures, the infection rate of AHSV could increase and the transmission rate could be faster (Mellor and Hamblin, 2004). All nine serotypes have been reported in South Africa (Mellor and Hamblin, 2004). Different serotypes of AHSV are usually present within an area during an outbreak and this frequently leads to one serotype dominating during one season and a different serotype in the next season (Bremer *et al.,* 1998). The presence of multiple serotypes within an endemic area could increase the probability of antigenic shift or reassortment of genome segments between serotypes (Weyer *et al.,* 2016).

1.2.4 Pathogenesis and disease forms

AHSV infects different equids species such as horses, mules, donkeys as well as zebras (Theiler, 1921). Clinical symptoms of AHS are less severe or absent in zebras and they are considered natural vertebrate hosts and reservoirs (Mellor and Hamblin, 2004). Horses are more susceptible to AHS and the mortality rates can reach up to 95% (Theiler, 1921; Mellor and Hamblin, 2004). The virus initially multiplies in the local draining lymph nodes followed by primary viremia and infection of the lungs (Coetzer and Guthrie, 2004). African horse sickness manifests in one of four different forms that include mild or fever, peracute or pulmonary, subacute or oedematous and/or mixed forms (Theiler, 1921; Coetzer and Guthrie, 2004; Mellor and Hamblin, 2004). The four different forms of AHS disease depend on the virus strain and the vertebrate host species (Sailleau *et al.,* 2000). The mild form occurs in donkeys and zebras, which are resistant to disease progression or in horses that are immune to the infecting serotype or that are cross-protected by one or more other serotypes (Coetzer and Guthrie, 2004). The mild form manifest as a slight fever (39-40^{o}C) lasting between one and six days before the body temperature returns to normal and the animal recovers (Theiler, 1921; Coetzer and Guthrie, 2004; Roy, 2004). The pulmonary form of the disease is common in horses that are not immune to AHS (Coetzer and Guthrie, 2004; Roy, 2004). The pulmonary form is characterized by a fever followed by severe dyspnoea, paroxysms of coughing, secretion of serofibrinous fluids from the nostrils that can continue after death (Coetzer and Guthrie, 2004; Mellor and Hamblin, 2004). The pulmonary form of AHS infection has been observed in dogs (Theiler, 1921; Coetzer and Guthrie, 2004). The subacute or oedematous form of AHS is characterized by fever for several weeks and other symptoms that may include swelling of the head, chest and neck, but also the supraorbital fossae (Coetzer and Guthrie, 2004; Mellor and Hamblin, 2004; Roy, 2004). The mixed form is the most common observed form of the AHS disease during which the horses show signs of respiratory distress followed by swelling or initially shows the subacute form before developing respiratory distress and possibly death (Mellor and Hamblin, 2004).

1.2.5 Prevention and control of disease

There is no known available treatment for AHS, but vaccination is compulsory in South Africa and good husbandry practices are advised (Coetzer and Guthrie, 2004; Mellor and Hamblin, 2004). Animals can be protected against AHS if they are immunised annually with all nine serotypes or cross protecting serotypes (du Plessis *et al*., 1998). Cross-protection between serotypes 1 and 2, 3 and 7, 5 and 8, and 6 and 9 has been observed (Erasmus, 1978). Polyvalent attenuated live vaccines (ALV) are currently commercially available from Onderstepoort Biological Products (OBP), South Africa. An estimated 300 000 doses of the AHS ALV vaccine are sold annually, but unfortunately this is not representative of all the equines in the country (Coetzer and Guthrie, 2004; Weyer *et al*., 2016). These vaccines contain live attenuated strains in a trivalent form (AHSV serotype 1, 3 and 4) and quadrivalent form (AHSV serotype 2, 6, 7 and 8) administered 2-3 weeks apart (von Teichman *et al.,* 2010). Serotype 9 was not included in the vaccine while serotype 5 was omitted following reports of foals dying or suffering severe post-vaccination reactions and because serotypes 8 and 6 provide cross-protection against serotypes 5 and 9 respectively (Mellor and Hamblin, 2004; von Teichman *et al*., 2010). Vaccination of horses should be repeated annually, particularly in the endemic regions (Coetzer and Guthrie, 2004; OIE, 2009). Although the live attenuated vaccines have been shown to be effective in endemic areas, there are some questions and worries about their use, which includes the following: 1) The AHSV vaccine manufactured in South Africa is not licenced for use in Europe, 2) the live vaccine may be teratogenic and should not be used in early stages of animal pregnancy, 3) AHSV has a segmented genome and reassortment between the live vaccine viruses and wild-type viruses could take place within host cells, and 4) after reassortment with wild-type-strains, the attenuated virus could revert to a virulent phenotype (Mellor and Hamblin, 2004; Weyer *et al*., 2016).

1.3 AFRICAN HORSE SICKNESS VIRUS (AHSV)

1.3.1 Taxonomic classification of AHSV

The genomes of viruses within the family *Reoviridae* are comprised of several (10, 11, or 12) segments of linear double-stranded RNA (Attoui *et al.,* 2005; Maan *et al.,*

2013). Based on the differences in morphological and physiological properties, the *Reoviridae* family is divided into 15 genera that include; *Aquareovirus, Cardoreovirus, Coltivirus, Cypovirus, Dinovirnavirus, Fijivirus, Idnovirus, Mimoreovirus, Mycoreovirus, Orbivirus, Orthoreovirus, Oryzavirus, Phytoreovirus, Rotavirus and Seadornavirus,* (Attoui *et al.,* 2012). The *Orbivirus* genus is the largest, comprising of 22 species acknowledged by the International Committee for the Taxonomy of Viruses (ICTV) and 15 tentative species (Maan *et al.,* 2013). Orbiviruses require both arthropod vectors (flies, ticks, mosquitos and/or *Culicoides)* and vertebrate hosts (domestic and wild animals, birds and/or humans) for their transmission and life cycle (Hassan *et al.,* 2001; Attoui *et al*., 2012; Maan *et al.,* 2013). Phylogenetic analysis indicated a linear relationship between the evolutionary split of the insect vector species and the genetic distances between the different insect-borne orbivirus groups, providing evidence of co-evolution of the orbiviruses and their insect host (Mohd Jafaar *et al.,* 2014). The co-evolution between host and arbovirus implies that AHSV is genetically closely related to other *Culicoides*-borne Orbiviruses including bluetongue virus (BTV), epizootic haemorrhagic disease virus (EHDV) and equine encephalosis virus (EEV) than other orbiviruses such as the equine infecting Peruvian horse sickness virus (PHSV) transmitted by mosquitoes (Mohd Jaafar *et al.,* 2014). Due to these similarities, the structure and functions of the prototype virus, BTV, are frequently used as the basis for understanding AHSV and other Culicoides-borne orbiviruses (Coetzer and Guthrie, 2004).

1.3.2 The AHSV Viron

The virion consists of a double capsid layer of proteins known as the outer and the inner capsids (Bremer, 1976). The outer capsid consists of VP2 (Huismans and Van Dijk, 1990) and VP5 (Mertens *et al*., 1987) and surrounds the inner capsid proteins comprising of VP7 (outer core) and VP3 (inner core) together with three inner minor core proteins VP1, VP4 and VP6 (Huismans *et al*., 1987). The segmented dsRNA genome is located within the inner core (Prasad *et al.,* 1992) (Figure 1.1). The surface layer of the core particle has a diameter of 73nm that has icosahedral symmetry due to the 260 trimers of VP7 (Prasad *et al*., 1992; Roy, 2004). The inner core proteins are linked with the RNA segments and play a collective role in forming the viral RNA-directed RNA polymerase complex responsible for initiation, extension

and methylation of messenger RNA (mRNA), and the formation of negative-strand RNA during replication and assembly (Basak *et al*., 1997).

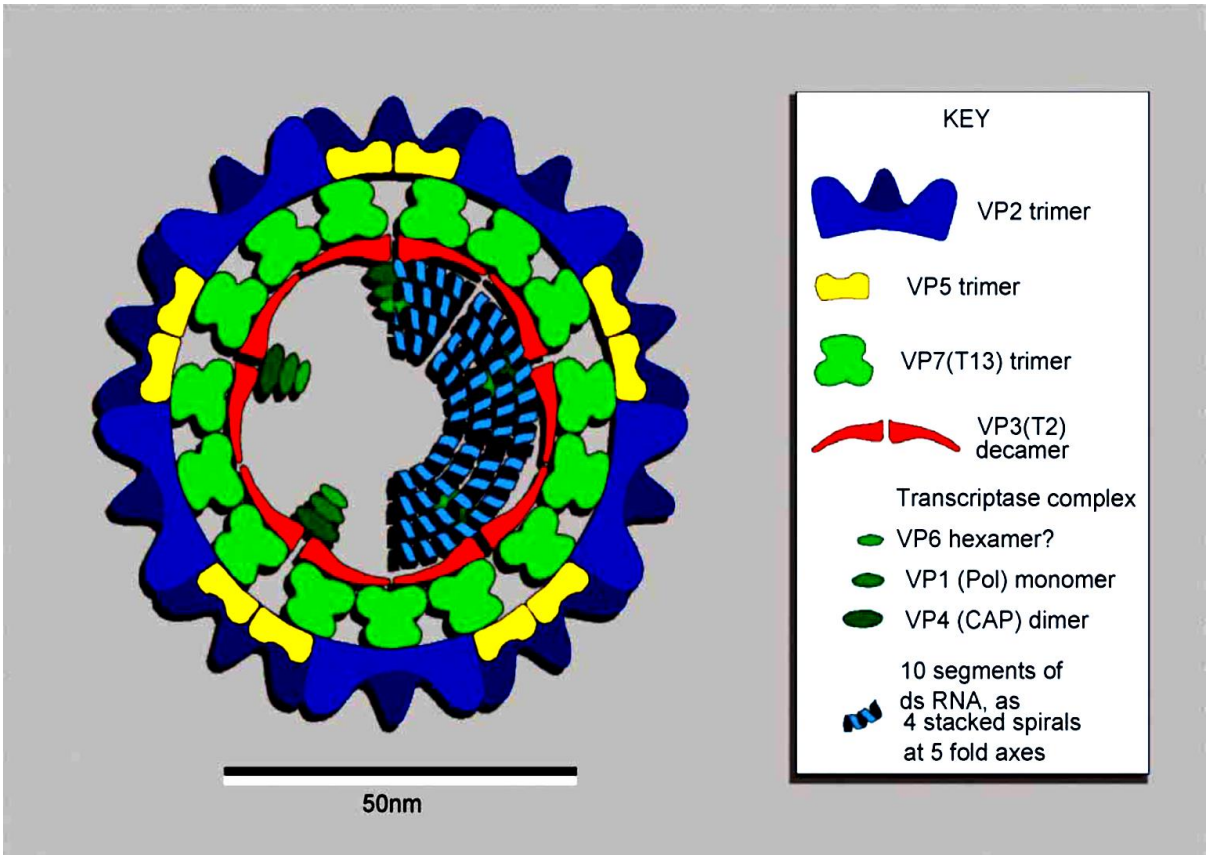

Figure 1.1 Schematic diagram of the structure of bluetongue virus (BTV) particle (Mertens, 2000).

1.3.3 Viral genome

Similar to other species of the *Orbivirus* genus, AHSV contains a linear double stranded-RNA (dsRNA) genome of ten segments (Verwoerd *et al*., 1970; Bremer, 1976; Attoui *et al.,* 2005; Maan *et al.,* 2013). The dsRNA segments are grouped according to their sizes i.e. large (segment 1 to 3), medium (segment 4 to 6) and small (segment 7 to 10) and are separated based on their unique size using polyacrylamide gel electrophoresis (PAGE) (Bremer, 1976; Bremer *et al,* 1990) (Table 1.1). The genome segments range from 3965 bp for segment 1 to 756 bp for segment 10. Of the ten segments, the nucleic acid sequences of segments 1, 3, 5, 7 and 8 are highly conserved between serotypes. Segments 6 and 10 have shown some degree of conservation, while segment 2 which codes for VP2, is highly variable between nine distinct serotypes (Bremer *et al.,* 1990). The genome

segments encode seven structural (VP1-7) and five non-structural proteins (NS1, NS2, NS3, NS3a and NS4) (Bremer *et al*., 1990; Ratinier *et al*., 2011; Zwart *et al*., 2015) (Table 1.1).

1.3.4 The AHSV Proteins

AHSV consists of seven viral structural (VP) and five non-structural proteins (Bremer *et al*., 1990; Ratinier *et al*., 2011; Zwart *et al*., 2015) that are described briefly in this section.

1.3.4.1 The outer capsid proteins

The VP2 and VP5 form the outer capsid proteins and are both involved in host-cell attachment and penetration (Huismans *et al*., 1987; Mertens *et al.,* 1987; Mertens *et al*., 1995). During the infection of the host cells, the outer capsid proteins are removed when the virus passes through the cell membrane to yield a transcriptionally active core (470S) particle (Basak *et al*., 1996). The VP2 is a highly variable protein and the major serotype-specific antigen, resulting in nine distinct VP2 phylogenetic clusters, each representing a different serotype (Mertens *et al.,* 1987; Huismans and Van Dijk, 1990). The VP2 induces the host's neutralisation-specific immune response (Huismans *et al*., 2004). The VP5 forms the inner layer of the outer capsid protein encoded by genome segment 6 (Mertens *et al.,* 1987) and is involved in membrane destabilization prior to virus entry into the cells (Hassan *et al*., 2001).

1.3.4.2 Major core proteins

The major core proteins consist of VP3 and VP7 that interact to form an icosahedral structure (Prasad *et al*., 1992; Roy, 2004). In the absence of VP3, VP7 forms trimers incapable of assembling into icosahedral particles (Roy, 2004). The VP3 is encoded by genome segment 3, while VP7 is encoded by genome segment 7 (Roy *et al*., 1994). Both these proteins are highly conserved across all AHSV serotypes (Bremer *et al*., 1990; Roy *et al*., 1994; Mertens *et al*., 2006). Viral protein 7 is an orbivirus immunodominant group specific-antigen and can be used to draw the distinction between different species of the orbivirus genus, since AHSV and BTV share only

45% sequence identity between their VP7 amino acid sequences (Oldfield *et al*., 1990; Huismans and Van Dijk, 1990; Monastyrskaya *et al*., 1997).

1.3.4.3 Minor core proteins

The inner core viral proteins (VP1, VP4 and VP6) participate in an AHSV genome replication in both the insect and the mammalian hosts (Wilson *et al*., 2009). All three minor proteins combine to form the transcription complex that produces single stranded RNA (ssRNA) from the dsRNA genome (Prasad *et al*., 1992). The ssRNA serves as mRNA for viral protein synthesis as well as templates for the production of a new dsRNA genome (Boyce *et al*., 2007). Encoded by genome segment 1, VP1 serves as an RNA-dependent polymerase and functions at optimal temperatures of between 27^{o}C and 42^{o}C (Roy *et al*., 1988; Wilson *et al*., 2009). Viral Protein 4, is encoded by genome segment 4 and functions as a capping enzyme (functions divided into guanyltransferase, methyltransferases 1 and 2, RNA 5' triphosphatase, inorganic pyrophosphatase and NTPase) (Roy, 2004). Viral protein 6 encoded by genome segment 9, is a helicase and responsible for the unwinding of the viral genomic dsRNA for transcription (Stauber *et al*., 1997). Both VP1 and VP4 are highly conserved among the different AHSV serotypes (inter-serotypical) (Roy *et al*., 1994).

1.3.4.4 Non-structural proteins

Five non-structural proteins (NS1, NS2, NS3, NS3a and NS4) are produced inside infected cells (Ratinier *et al*., 2011). Non-structural protein 1 (NS1), coded by genome segment 5, forms tubular structures within the cytoplasm of the host cell, while NS2, which is encoded by segment 8, forms a major component of granular viral inclusion bodies also in the cytoplasm (Thomas *et al.,* 1990). The NS3 is encoded by genome segment 10, and it is the second most variable protein (Van Niekerk *et al.,* 2001a). The NS3 is a cytotoxic protein synthesized in small quantities within infected cells and participates in alteration of the cell membrane, virus exit and cell death (Martin *et al*., 1998; van Niekerk *et al*., 2001a). The NS4 was first predicted in BTV by Ratinier and co-workers (2011) and subsequently described in AHSV by Zwart and colleagues (2015). The NS4 is encoded in the +1 reading frame of segment 9, which also encodes VP6 in the +3 reading frame (Ratinier *et al*., 2011; Zwart *et al*., 2015). Localized in the nucleoli of infected cells, the NS4 is assumed to

participate in virus-host interaction by acting against antiviral responses of the host (Ratinier *et al.,* 2011; Zwart *et al*., 2015).

Table 1.1 African horse sickness virus genome segments and encoded proteins (Adopted from Martens *et al*., 2006).

†Genome segment(Size: bp)	ORFs Bp	Encoded protein	Protein Location	Number of Amino acids	Properties and Functions for the encoded proteins
Seg 1 (3965)	14-3931	VP1	sub-core	1305	Minor protein, RNA dependent RNA polymerase [§: Pol].
Seg 2 (3205)	13-3174	VP2	outer capsid	1053	Outer layer of outer capsid, Serotype specific antigen.
Seg 3 (2792)	27-2744	VP3	sub-core capsid layer	905	Innermost protein capsid, interacts with sub core minor proteins.
Seg 4 (1978)	12-1940	VP4	sub-core	642	Dimers, capping enzyme (§: Cap),
Seg 5 (1748)	36-1682	NS1	Cytoplasm	548	Forms tubular structures.
Seg 6 (1566)	20-1537	VP5	outer capsid	505	Inner layer of the outer capsid, membrane destabilisation prior to cell entry.
Seg 7 (1169)	18-1067	VP7	Outer core	349	Trimer (§: T13), major serogroup specific antigen.
Seg 8 (1167)	23 – 1220	NS2	Cytoplasm and VIB	365	Viral inclusion body matrix protein, ssRNA binding.
Seg 9 (1166)	18-1127	VP6 and NS4	Sub core	369	1. VP6 ssRNA and dsRNA binding, helicase ATPase. 2. NS4 viral host interaction, act against antiviral responses
Seg 10 (756)	18-672 62-672	NS3 and NS3a	Cell Membranes	217 203	Glycoproteins, membrane proteins, involved in cell exit, and may determine virulence. Cytotoxic protein.

1.3.5 Replication overview

Orbiviruses obligatory replicate in both arthropod vectors and vertebrate hosts, implicating numerous challenges presented by the host humoral response (Mertens *et al*., 2004; Patel and Roy, 2014). The main features of orbivirus replication include: 1) attachment and penetration, 2) removal of viral capsid and formation of replicative complexes, 3) transcription of ssRNA to be used as both mRNA and template for synthesising new genomic segments, 4) generation of virus tubules (VIBs) and inclusion bodies and 5) the movement and the release of virus from the cell (Gould and Hyatt, 1994; Patel and Roy, 2014) (Figure 1.2). Attachment and penetration of the virus are mediated by VP2 binding to cell surface receptors such as sialoglycoprotein or glycoporin A (Roy, 2004).

Viral Protein 5 mediates penetration by destabilising the endosomal membrane prior to virus entry into the cytoplasm (Mertens *et al.,* 1987; Hassan *et al*., 2001; Roy, 2004). Virus attachment to the host cell leads to receptor-mediated endocytosis of virions in clathrin-coated vesicles (Roy, 2004; Patel and Roy, 2014). The low pH in the endosomes causes rearrangement and conformational changes in VP5, which has been shown experimentally to act as a fusion protein (Patel and Roy, 2014). The conformational change results in the interaction between the viral outer capsid proteins and the core proteins to dissociate, allowing the transcriptionally active core particles to be released into the endosomal vesicles (Mertens *et al*., 2004; Roy, 2004; Attoui *et al*., 2012). The majority of dsRNA viruses avoid exposing their genomes to the host cell cytoplasm and subsequent activation of host immune defences (Roy, 2004). The host immune defence activation is achieved by retaining viral genomes and its mRNA synthesis enzymes, within a firm closed-protein capsid (Roy, 2004). The viral core forms the infectious unit of AHSV which is taken into the host's cytoplasm prior to the virus replication (Mertens *et al*., 2004). During mRNA transcription, 10 single-stranded RNA (ssRNA) molecules are synthesized inside the viral core, corresponding to each of the genome segments. The synthesized ssRNA molecules vary in size between 3944 base pairs (bp) to 833 bp (Attoui *et al*., 2012; Patel and Roy, 2014). These ssRNA transcripts encode for the production of viral proteins translated by host cell machinery (Patel and Roy, 2014). Of the ten segments, eight encode a single protein each, while two segments (Segment 9 and

10) encode two proteins each (VP6 and NS4, and NS3 and NS3A, respectively) (Retinier *et al*., 2011). Viral inclusion bodies (VIBs) are formed by NS2 and are considered to be a site for virus morphogenesis or assembly of transcriptionally active virus cores that contains dsRNA genome (Attoui *et al*., 2012). After virus assembly, the viral particles move within the infected cells through interaction with the cellular cytoskeleton and subsequently released as mature virions from the cell via lysis (Attoui *et al*., 2012, Patel and Roy, 2014). As a cytotoxic protein, NS3 is involved in the budding of the host cells, resulting in the release of the virus and host cell death (Martin *et al*., 1998; van Niekerk *et al*., 2001a). (Figure 1.2).

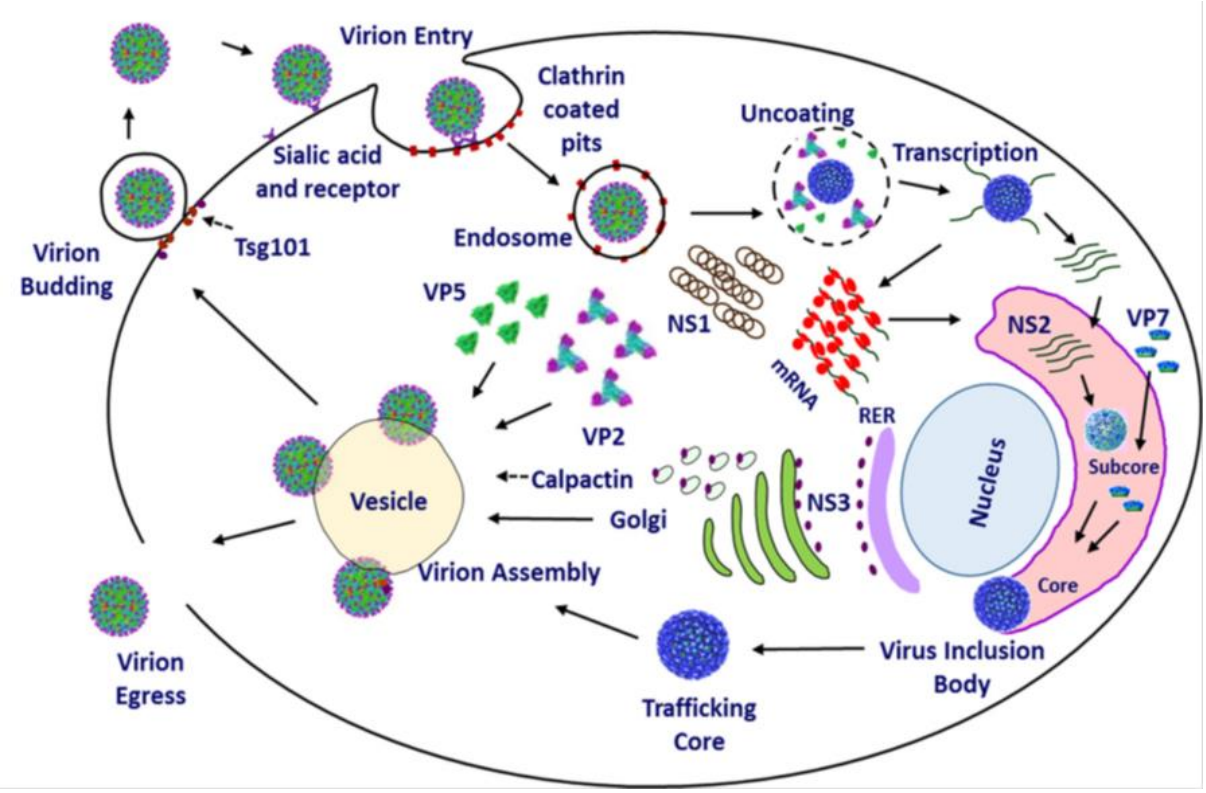

Figure 1.2. Schematic diagram representing the replication cycle of bluetongue virus (Patel and Roy, 2014).

1.3.6 Diagnosis and Laboratory confirmation

African horse sickness should be suspected and subsequently confirmed by laboratory test if an animal develops clinical symptoms as described in section 1.2.4. (*www.cfsph.iastate.edu/Factsheets/pdfs/african_horse_sickness*[CFSPH], last updated in 2015). Mild or the fever form of the disease is often misdiagnosed and a number of differential diagnosis like equine echephalosis, equine infectious anaemia, equine

piroplasmosis, Hendra virus infection, purpura haemorrhagica, and viral arteritis should be considered (CFSFP, 2006; OIE, 2009). Since AHS is an OIE listed disease, it is important that an outbreak be confirmed in an accredited laboratory, using either approved immunoassays, molecular detection of virus nucleic acids or virus isolation (CFSPH, 2015; CFSPH, 2006; OIE, 2009). A variety of mammalian (BHK21 and Vero cells) and insect (mosquito cells (C6/36) and *Culicoides* (KC cells)) cell lines as well as intravascular inoculation of chicken embryo bud or intracerebral inoculation of 2 to 3-day old mice can be used to isolate the virus (Mellor and Hamblin, 2004; CFSPH, 2006; OIE, 2009). Group-specific tests such as agar gel immune-diffusion (AGID), complement fixation test (CFT), ELISA, indirect fluorescent antibody (IFA) direct ELISA and or virus neutralization (VN), can be used for serological diagnosis of AHSV (House *et al.*, 1990). Recombinant VP7 protein is used as antigen in a validated indirect ELISA (I-ELISA) for the diagnosis of group-specific IgG antibodies in horse sera, since the inner core protein is highly conserved between AHSVs (Maree and Paweska, 2005; OIE, 2009).

It is important to rapidly and reliably detect the virus during outbreaks, and since virus isolation is time consuming, molecular assays are preferred for use during the outbreaks (Sailleau *et al.,* 2000). Various tests for the molecular identification of virus nucleic acids have been described and approved, each targeting a conserved region of either the highly conserved segment 7 or 8 (Bremer *et al*., 1998; Sailleau *et al.,* 2000; Agüero *et al.,* 2008). In contrast to serological tests such as VN, IFA and AGID that detect antibodies against the virus, molecular assays target the virus nucleic acids and confirm if the virus is still present in the animal.

1.4 VIRUS EVOLUTION

Evolution is considered to be a change in the genetic composition of an organism from one generation to the next (Makadiya, 2007). Evolution is essential for all living organisms because it provides resources for natural selection and adaptation of the population to environmental changes (Walker and Cowley, 1999; Greenbaum and Ghedin, 2015). When viruses are transmitted from one host to another, they face continuous environmental changes, such as defensive or immunological responses and have to survive or adapt to a new environment (Walker and Cowley, 1999; Makadiya, 2007). Genetic variation plays a critical role in the fitness and adaptability

of the virus to the changing environments, allowing it to overcome multiple selection pressures (Domingo and Holland, 1997; Greenbaum and Ghedin, 2015). Adaptation can be seen as an ability of viruses to rise to its adaptive peaks as a result of high mutation rates, high population turnover, as well as rapid replication cycles (Domingo and Holland, 1997). When viruses are introduced into a new environment, populations will evolve from low fitness to its adaptive peaks (Elena and Sanjuán, 2007; Makadiya, 2007).

According to MacLachlan and Guthrie (2010), viral genetic variation can lead to the differences in disease severity patterns and other biological properties among different field strains of BTV and AHSV. Genetic variation also plays a role in the existence of differences in the organism's reproductive success and fitness (Manrubia and Lázaro, 2006). In a virus population, fitness is regarded as a relative ability to produce infectious progeny viruses that are stable under a given environmental conditions (Domingo and Holland, 1997). Environmental conditions are crucial in the fitness of an organism and therefore small external changes can completely change the quantitative fitness value of an individual or of the population (Manrubia and Lázaro, 2006). During viral passaging, viruses can undergo a large fitness change, with some passage regimes resulting in fitness loss and others to gain (Domingo and Holland, 1997).

When virus cross species barrier and infect a new host, it is initially unable to replicate effectively in the new cell type and the infection can be a dead end for the virus (Manrubia and Lázaro, 2006). The new host may present constraints during virus cell entry, virus replication, or transmission, however, viruses have the adaptive capacity, enabling them to infect different organisms (Manrubia and Lázaro, 2006; Wargo and Kurath, 2012). Changes in the structural properties of the virus can allow the virus to penetrate and replicate in the new host species (Manrubia and Lázaro, 2006).

Viruses' use of cellular metabolism frequently causes injury to the host cells, leading to disease of the host and can result in the death of the infected organism (Eigen, 1993; Manrubia and Lázaro, 2006). The virus challenge to the host sometimes leads to the co-evolution of the virus and the host cells (Mohd Jaafar *et al.,* 2014). The co-evolution implies that the host cells will develop defence mechanisms that are

expressed in the immune system of vertebrates, for example, lymphocytes of the host immune system can acquire antigen-specific receptors in order to provide host immune surveillance against invading pathogens and resist infectious agents (Manrubia and Lázaro, 2006; Sun and Lanier, 2009). As a result of the host immune clearance and the limit in the lifespan of the host, the virus must be transmitted to another host for its survival (Wargo and Kurath, 2012). Factors that contribute to, or affect the virus' evolutionary dynamics may include the type of infection caused by the virus (acute or persistent) to the host cells, the duration of immunity elicited by the viral infectious agent, the capacity of the viral pathogen to acquire immune-escape mutations, as well as the mode of transmission (Manrubia and Lázaro, 2006). According to Manole *et al* (2012), host-driven AHSV evolution can be understood by studying the viral structure and how VP2 interact with host cells.

Certain viruses like measles and smallpox specialize in multiplying inside single host cells, while others like influenza virus are able to infect different host species (Manrubia and Lázaro, 2006). Viruses from the herpes virus family have large DNA genomes, which contain hundreds of genes (Lucas *et al*., 2001). Herpes viruses use a mechanism described as 'camouflage' and 'sabotage', possessing highly evolved molecules that disrupt host immune defence mechanisms (Lucas *et al*., 2001; Sun and Lanier, 2009). Large numbers of evolved genes hinder the recognition of the infected cells by the host's immune defences (Manrubia and Lázaro, 2006; Sun and Lanier, 2009). In contrast to DNA viruses, RNA viruses have smaller genomes and display a higher evolutionary rate (Lucas *et al*., 2001). Virulent viruses have a short generation time, which usually causes serious damage or even death to their hosts (Manrubia and Lázaro, 2006). On the other hand, attenuated viruses limit their persistence inside the host cells without provoking serious disease symptoms (Manrubia and Lázaro, 2006). Many of these viruses can stay inside the host unnoticed ("camouflaged to the host") and are exposed when the virus is transmitted to a new host that lacks an effective defence mechanism (Lucas *et al*., 2001; Manrubia and Lázaro, 2006).

Viruses evolve as a direct result of the incorporation of errors during replication and or packaging of the viral genome (Manrubia and Lázaro, 2006). High error rates allow viruses to adapt to environmental changes and result in RNA virus populations

consisting of different mutants, instead of genomes with identical nucleotide sequences (Eigen, 1993; Novella *et al*., 1999). The virus population composed of different mutants is called quasispecies and are important sources of viral adaptation since they form the continuous changing repositories of genotypic and phenotypic viral variants (Domingo *et al.,* 2012). Quasispecies can evolve as a result of mutation, molecular recombination and genome segment reassortment (Bonneau *et al*., 2001; Domingo *et al.,* 2012). Quasispecies evolution can play an essential role in determining the biological properties as well virulence of AHSV individual strains (MacLachlan and Guthrie, 2010). Quasispecies come about as equilibrium between evolution and natural selection that generates populations with variable genomes (Eigen, 1993; Holmes and Moya, 2002). There are two types of selections, namely: negative selection, in which selection occurs against particular traits and positive selection wherein the favourable traits are inherited across generations (Domingo and Holland, 1997; Makadiya, 2007). Even though individual sequences are targeted by mutations, the effect of selection is only detected in the population (Eigen, 1993).

The fitness value and fitness effects of genetic variation depend on the sequence context of a viral genome as well as physical and biological environments (Eigen, 1993; Domingo, 2010). When the population size is very low, competition and selection of the best-adapted viral phenotypes take place in a reduced number of genomes (Manrubia and Lázaro, 2006). Since most genetic mutations will have deleterious effects, this will result in a decreased fitness of such low size virus population (Domingo and Holland, 1997; Manrubia and Lázaro, 2006). On the other hand, large virus populations encourage a competition between large numbers of genomes and have an advantageous effect on the population, resulting in the increase in fitness (Manrubia and Lázaro, 2006; Elena and Sanjuán, 2007). Once mutation, competition and fitness selection have proceeded for a lengthy duration in an unchanged environment, the stationary phase is reached and the virus population will then be maintained in evolutionary stasis (Manrubia and Lázaro, 2006).

The first study to look at AHSV viral evolution was conducted by Quan *et al*., (2008) who investigated segment 10 (S10) of all nine serotypes using 145 viral isolates, and discovered that AHSV, BTV and EEV formed monophyletic groups, but BTV was more comparable to AHSV on the basis of their evolutionary distance. In the same

year Von Teichman (2008) investigated the genome reassortment between AHSV vaccine strain and wild-type strains. The experimental co-infection of two unrelated AHSV strains was first carried out by Meiring *et al* (2009) focusing their attention on the NS3. Later Weyer *et al*., (2016) reported on AHSV genome reassortment among field and vaccine strains collected between 2004 and 2014. This study analysed AHSV evolutionary dynamics using isolates across all nine serotypes collected from 1933 until 2014. All three mechanisms of RNA genetic variation that contributes to virus evolution were investigated. These mechanisms are described in the next section.

1.4.1 Mechanism of RNA genetic variation

Viruses can undergo minor genetic changes due to mutation, and major genetic changes as a result of genetic recombination and/or reassortment (Fleischmann, 1996; Domingo and Holland, 1997; Bonneau *et al*., 2001). Mutation is also known as vertical evolution because genes are inherited directly from parental offspring (Gogarten and Townsend, 2005). Genomic recombination and reassortment are known as horizontal evolution because genetic information moves across normal mating barriers, between more or less distantly related organism (Keeling and Palmer, 2008). The minor and major genetic changes are also known as genetic drift and genetic shift respectively, and they likely contribute to the genetic diversity of both BTV and AHSV (MacLachlan and Guthrie, 2010). Mutation can produce viruses with new antigenic determinants resulting in the formation of novel viruses (Fleischmann, 1996). The formation of antigenically altered viruses is known as genetic drift and could cause diseases in previously resistant or immune hosts (Fleischmann, 1996; Manrubia and Lázaro, 2006). The development of viruses with new antigenic determinants through genetic reassortment or recombination is called a genetic shift (Fleischmann, 1996; Manrubia and Lázaro, 2006). Both the genomic recombination and reassortment depend on more than one different virus strains of the same species to simultaneously infect the same host cell (Novella *et al*., 2011; Simon-Loriere and Holmes, 2011). Mutation is universal and affects all viruses at different rates, while recombination varies greatly in frequency among different viruses (Domingo, 2010). Genomic reassortment only occurs among the segmented viruses such as ASHV, and provides a better way of genetic variation and virus

evolution (Wilson *et al*., 2009; Domingo, 2010). The three evolutionary mechanisms that can lead to DNA/RNA evolution are discussed below.

1.4.1.1 Genetic mutations

Mutation is regarded as the major source of virus genetic variation (Manrubia and Lázaro, 2006). The genetic mutation occurs as a result of an error incorporated into the viral genome sequence during replication (Fleischmann, 1996; Manrubia and Lázaro, 2006). A single nucleotide mutation could be lethal, neutral or advantageous to the virus depending on the position and subsequent amino acid change (Eigen, 1993; Walker and Cowley, 1999; Makadia, 2007; Novella *et al*., 2011). The position of a nucleotide is important because ribosomes interpret the nucleotide codons during amino acid synthesis and in most cases, the first two positions are important in specifying the amino acid to be incorporated into a protein sequence, while the third codon is more tolerant of changes, i.e. changes may in most cases not affect the amino acid sequence (Eigen, 1993). Lethal mutation can cause distortion or truncation of the protein, resulting in a non-functional protein (Walker and Cowley, 1999). Many mutations are advantageous and can result in the formation of viable genomes that increase virus fitness (Walker and Cowley, 1999). RNA viruses have the highest mutation rates, followed by ssDNA and lastly dsDNA viruses (Domingo and Holland, 1997; Makadiya, 2007). The difference in the mutation rates of viruses is due to the fidelity of enzymes involved in the replication of their nucleic acids (Fleischmann, 1996). DNA viruses use high-fidelity DNA polymerase enzyme that has proof-reading activities to ensure high copying accuracy during replication (Walker and Cowley, 1999; Manrubia and Lázaro, 2006). Unlike DNA viruses, RNA viruses such as HIV, rotaviruses, BTV and AHSV use enzymes that lack proof-reading activities such as RNA-dependant RNA polymerase or reverse transcriptase (Domingo and Holland, 1997; Fleischmann, 1996). Therefore, RNA viruses are prone to errors during replication.

The mutation rates as a result of nucleotide misincorporation in RNA viruses are between 10^{-3} to 10^{-4} per round of replication, which is at least 1000 times higher as compared to the mutation rates in the eukaryotes (Walker and Cowley, 1999). High levels of mutation ensure that various mutants are present in each infected cell, tissue or even the whole organism (Doming *et al*., 1993). Mutation can be of three

different forms, namely, nucleotide substitution (exchange of single base for the other), insertion (an addition of one or more extra bases in the gene sequence) or deletion (the removal of one or more base(s) from the gene sequence) (Saitou and Ueda, 1994; Makadiya, 2007). Nucleotide substitution can lead to either a change in an amino acid sequence known as nonsynonymous (d_N) substitution or amino acid stays the same, which is known as synonymous (d_S) substitution (Gojobori *et al*., 1994). Nonsynonymous mutation is visible to natural selection while synonymous mutations can occur under strong selective pressure (Gojobori *et al*., 1994; Yang *et al*., 2000). The genetic firmness of the viruses can be credited to negative or purifying selection in maintaining the functions of viral proteins (Boulila, 2011). The extent of negative selection in the genes or functional limitation for maintaining the encoded protein sequences can be evaluated by the ratio between nonsynonymous and synonymous (d_N/d_S) substitutions (Yang *et al.,* 2000; Boulila, 2011).

Rates of d_N and d_S are referred to as the amount of nonsynonymous and synonymous substitution per site, and the ratio ω is used to estimate the selection pressure at the viral protein level (Yang *et al.,* 2000). A ω>1 can imply that adaptive or positive selection is driving genetic diversity and nonsynonymous mutations offer fitness advantages to individual proteins (Nielsen and Yang, 1998; Yang *et al*., 2000; Boulila, 2011). However, ω<1 can imply that negative or purifying selection is limiting the genetic diversity and synonymous mutations have higher probabilities than nonsynonymous mutation (Yang *et al.,* 2000; Boulila, 2011). The “neutral” model assumes that two sets of sites exist, namely; conserved sites in which nonsynonymous mutations are eliminated by selection, and only synonymous substitutions are possible (ω=0), and complete neutral sites in which the ratio ω=1 (Nielsen and Yang, 1998; Yang *et al*., 2000). Quan *et al*., (2008) used 145 African horse sickness viruses comprising of nine different AHSV serotypes to determine selection pressure associated with segment 10. Their finding revealed that AHSV segment 10 undergoes a strong purifying selection. However, mutation in AHSV whole genome segments is yet to be reported, although it has been determined in BTV. Nucleotide substitution can play a crucial role in generating new mutants in BTV evolution (Niedbalski, 2013). However, due to the necessity of BTV to infect and replicate in both the insect and vertebrate host, substitution rates are lower compared to single host RNA viruses (mean rates of 0,5-7 $x10^{-4}$ nucleotide

substitutions/ site/ year) and undergo strong purifying selection (Carpi *et al*., 2010; Boyle *et al.,* 2014). Viruses such as BTV and AHSV are vector-borne transmitted and less subjected to diversifying selection as revealed by Woelk and Holmes (2002). Prior to the commencement of this study, it was hypothesized that AHSV will follow similar patterns of BTV substitution rates and selection, because they are both transmitted to their vertebrate hosts by the same insect vectors.

While various studies have documented substitution mutation, the mutation caused by insertion and deletion of nucleotides are not well studied because they are considered to occur infrequently in coding regions due to their strong deleterious effects (Saitou and Ueda, 1994). Insertion mutations occur when one or more extra nucleotides are added into the RNA/DNA sequence (Makadiya, 2007; Saitou, 2013). Insertion or deletion event, could cause changes in the reading frame and result in the premature termination of a protein (Makadiya, 2007). Insertion of extra nucleotides can generate a shift in the reading frame, resulting in changes in the gene product (Walker and Cowley, 1999). On the other hand, deletion mutations cause the removal of one or more nucleotides from the RNA/DNA sequence, which may also result in alteration of the protein product (Taylor *et al*., 2004). While the length of the RNA/DNA sequence does not change when a substitution mutation occurs, both deletion and insertion result in changes in virus genome size (Saitou, 2013). Deletion mutations result in a reduced length of the viral genome and could improve the viral replication speed by polymerase enzymes (Kirkwood and Bangham, 1994). Insertion and deletion of nucleotides are responsible for the gaps when genetic sequences are aligned, however, they are usually ignored when nucleotide substitutions are estimated (Saitou and Ueda, 1994). Insertion and deletion mutations will either result in deleterious effect or strong advantageous effects (Chong *et al*., 2013). So far, there are no studies that have reported mutation as a result of substitution, insertion and/or deletion associated with AHSV field strains.

1.4.1.2 Genetic recombination

Genetic recombination is a process of exchanging genetic information between two or more DNA or RNA molecules (Novella *et al*., 2011; Boulila, 2011). Genetic recombination takes place in both the DNA and RNA viruses, but occurs at very low

rates in negative-stranded RNA viruses (Chare *et al*., 2003). If the recombination event occurs between different mutant genes is called intergenic recombination, and if it occurs between two or more alleles of the same gene it is called intragenic recombination (Posada *et al*., 2002). Recombination generates new genetic combinations from the cross-over between different parental molecules (Naggy and Simon, 1997). The resultant offspring could have increased virulence, an extended host range, the ability to evade the host immunity or resistance to antivirals (Simon-Loriere and Holmes, 2011). It is essential for two or more viruses to simultaneously infect the same host cells, in order for them to exchange their genetic material. The frequency of exchange between different viral genomes depends on the frequency at which simultaneous infection of a same host cell occurs as well as the composition of its genome (Froissart *et al*., 2005). The lowest recombination frequency was observed among the negative-sense RNA viruses (Chare *et al*., 2003; Domingo, 2010). The probability of co-infection is dependent on co-circulation of different strains at a specific time and geographical area, the prevalence of the population in the host as well as the co-infection rate (Perez-Losada *et al*., 2015). Two mechanisms are involved in the exchange of genetic material; the first is limited to segmented genomes and involve the shuffling of complete genome segments known as reassortment. The second creates mosaic sequences following the inter-molecular exchange of nucleotide sequences at a specific cross-over site, which in viruses is normally a non-reciprocal exchange from the beneficiary genome (Perez-Losada *et al*., 2015). The process is divided into homologous and non-homologous recombination based on the structure of the resulting cross-over site (Lai, 1992).

Negative-sense RNA viruses have a genome which is complementary to the messenger RNA and cannot be directly translated making it difficult for genetic recombination to occur (Elena and Sanjuán, 2007). Several factors contribute to the low rates of genetic recombination. These factors include 1). The viral RNA of negative sense viruses undergoes nucleoprotein condensation forming helical nucleocapsid, called ribonucleoprotein complex that serves as a template for replication. The ribonucleoprotein complex of negative sense RNA never disassembles and there are no structural changes of the RNA-N protein template during the RNA synthesis; 2). Since recombination depends on at least two viruses to simultaneously infect a single host cell, a mixed infection at low frequency could

result in low recombination events or no recombination. Negative-sense RNA viruses usually cause acute infection and short recovery periods; 3). Superinfection exclusion, in which the primary virus provides immunity against secondary infection by same or related viruses; and 4). The virus fitness of the recombinants could be another limiting factor of virus recombination associated with negative sense RNA viruses (Han and Worobey, 2011).

Genetic recombination has no restrictions and can take place in segmented and non-segmented viruses, including the orbiviruses such as BTV (Worobey and Holmes, 1999; He *et al*., 2010; Simon-Loriere and Holmes, 2011). Intragenic recombination occurs when a donor nucleotide sequence is presented in a single, adjacent accepter DNA/RNA sequence leading to the generation of a new DNA/RNA that possesses genetic information from different parental sources (Worobey and Holmes, 1999; Makadiya, 2007). This involves the rearrangement of genetic materials joining previous unassociated fragments and is often accomplished by breaking and joining of the DNA/RNA strands (Lewis-Rogers *et al*., 2004). As a result, two or more genetic fragments that belong to different parental molecules are linked together to form one genetic molecule (Manrubia and Lázaro, 2006). Intragenic recombination may arise due to the involvement of cis-acting (intramolecular) versus trans-acting (intermolecular) elements that are involved in the virus infectious cycle, as well as the action of viral replicase enzymes (Domingo and Holland, 1997). Recombination can produce highly fit viral genomes to rescue low fit parental genomes from selection pressures (Domingo, 2010).

Recombination can arise from the genetic molecules separated on the same parental sequences or that are present on different parental sequences (Naggy and Simon, 1997). The outcome of RNA sequences generated as a result of genetic recombination can show minor or major mutational changes. Recombination is therefore divided into two types namely; homologous and non-homologous recombination (Lai, 1992). Homologous recombination simply requires two or more closely associated DNA/RNA molecules, with crossover events occurring at the sites that match between the two DNA/RNAs resulting in the recombinant DNAs/RNAs retaining the same sequence and structural arrangement of the parental molecules (Lai, 1992; Lewis-Rogers *et al*., 2004). Homologous recombination can also occur

between similar sequences, but can be recognized when it affects unrelated though nearby sites on each parental DNA/RNA molecule, resulting in sequence duplicates or deletion or even insertion of unknown origin (Lai, 1992). The type of recombination that result in sequence duplicate, deletion or insertion is known as aberrant homologous recombination and in the majority of cases there may be no sequence homology around the crossover sites on the two DNA/RNA molecules (Lai, 1992). Non-homologous recombination involves sequences of two or more unrelated viral genomes with sequence homology (Lai, 1992; Naggy and Simon, 1997). Since non-homologous recombination takes place in regions with very little or no homology, it often results in deleterious types, and are therefore less observed than homologous recombination (Simon-Loriere and Holmes, 2011). Intragenic recombination events have been reported in bluetongue virus (He *et al*., 2010), however, it is yet to be reported among the African horse sickness viruses.

1.4.1.3 Genetic reassortment

Genome reassortment requires more than one virus particle of the same or related species to simultaneously infect a single host cell, which is followed by genome segment replication and the viral protein synthesis within the host cell (Vijaykrishna *et al*., 2015). Genome reassortment occurs during the packaging of the new virus containing segmented RNA genome such as BTV and AHSV, and because it is flexible, it can involve any viral segment (Fleischmann, 1996; Manrubia and Lázaro, 2006; Shaw *et al*., 2013). Reassortment can result in the formation of a new virus with new properties increasing virus diversification (Gorman *et al*., 1978; Manrubia and Lázaro, 2006). Such reassortment event increases the pressure placed on the host immune system to recognize the newly formed viruses with new antigenic determinants and may result in an increased virulence (Mellor and Hamblin, 2004; Manrubia and Lázaro, 2006; Greenbaum *et al*., 2012). Apart from an increase in virulence, reassortment can cause vaccine failure and pandemic emergence (Greenbaum and Ghedin, 2015), increased transmissibility, resistance to antiviral treatments or escape antibody recognition (Vijaykrishna *et al.,* 2015). Genomic reassortment between a live attenuated vaccine strain and field viruses may lead to the formation of a virulent virus capable of causing disease in previously immune hosts (Greenbaum *et al*., 2012; Shaw *et al*., 2013). Reassortment between AHSV

field and vaccine strains has been reported (Von Teichman and Smit, 2008; Weyer *et al*., 2016) and seem to drive AHS viral evolution (MacLachlan and Guthrie, 2010).

Through the genomic reassortment, the rate of virus acquiring genetic markers that are fit to overpower adaptive host immune barriers can be accelerated faster than mutation alone (Vijaykrishna *et al*., 2015). For the reassortment to take place two or more virus particles must attach and penetrate a single host cell and thereafter followed by the synthesis of genome segments within the host cell (Vijaykrishna *et al*., 2015). When different, but related segmented viruses simultaneously infect the same host cell, there are many possible reassortants that can be formed (Simon-Loriere and Holmes, 2011). For instance, the eight segmented influenza virus has 256 possible reassortants viruses when two strains infect one host cell, however, the strain type and the host cellular environment may influence reassortment outcomes (Greenbaum *et al*., 2012). Reassortment can cause the virus to infect new host species as per the observations in influenza viruses (Shaw *et al*., 2013).

1.4.2 Arbovirus evolution and their host factors

Host factors can influence the rate of virus evolution (Combe and Sajuán, 2014). There is a distinction in evolutionary rates between viruses infecting a single host compared to viruses that infect different host species (Manrubia and Lázaro, 2006). Viruses that infect different hosts have a complex life-cycle in which one host act as a vector, probably an insect and the other as vertebrate host (Manrubia and Lázaro, 2006; Ciota and Kramer, 2010). Viruses that are transmitted between vertebrate hosts and arthropod vectors are known as arboviruses. Arboviruses multiply in the insect tissues and are generally transmitted to the vertebrate host through a bite, transmitting fluids-containing infectious viruses (Mellor, 2000). The majority of arbovirus families are RNA viruses (*Bunyaviridae, Flaviviridae, Orthomyxoviridae, Reoviridae, Rhabdoviridae and Togaviridae)*; with only one DNA virus family of *Asfarviridae* also included (Ciota and Kramer, 2010). It is worth noting that not all members of these virus families are arboviruses. For instance, rotavirus is an important member of *Reoviridae* family, and yet, unlike other members of this family, it does not require an insect host for transmission, while *Thogotovirus* is the only arbovirus included within the family *Orthomyxoviridae.* Arboviruses can be transmitted naturally between susceptible hosts and arthropod vectors usually

through insect bites during blood meals on the vertebrate host (Cooper and Scott, 2001; Novella *et al*., 2011).

Viruses like myxoma virus, capripox virus, avian pox virus as well as equine infectious anaemia virus are transmitted to the vertebrate host by insects that act as mechanical vectors and do not multiply inside the vector hence they are not considered as arboviruses (Mellor, 2000). Arthropod-borne viruses such as BTV and AHSV evolve slower than viruses with a single host because of the constraints associated with replicating in both the insect and mammalian host (Novella *et al*., 1999; Combe and Sajuán, 2014; Novella *et al.,* 2011). Arthropods can act as effective vectors if the virus does not cause pathogenesis that triggers an immune response by the insect and also retains its ability to replicate in the vertebrate host (Wilson *et al*., 2009). Mammals and insects differ in their immune responses to viral infections, and as a result, genetic or genomic changes that encourage immune tolerance or escape in one host cells may have adverse effects on the viral biology of the other host (Novella *et al*., 1999; Woelk and Holmes, 2002).

1.4.3 Orbiviruses evolution

As a result of genetic evolution, BTV have 27 different serotypes and strain variations within each serotype (Niedbalski, 2013). Boyle *et al.* (2014) studied the evolution of BTV serotype 1 collected in northern Australia over the period of 30 years and found that all the genome segments evolve under strong purifying (negative) selection except for three sites that were under positive selection. The ability to infect two different hosts imposes constraints on the viral adaptation as a result of fitness trade-off or different fitness landscapes for replication in both hosts (de Mattos *et al*., 1996; Novella *et al.,* 1999; Anbalagan *et a*l., 2014; Boyle *et al*, 2014). Using the Bayesian coalescent estimate, Boyle *et al* (2014) reported the mean substitution rate that ranged between 3.5 to 5.3 X 10^{-4} substitutions/ site/ year over 33 years using genomic sequence data from 27 BTV genomes. Comparing their findings in BTV with non-vector borne members of *Reoviridae* it was observed that the BTV evolve less rapidly than rotavirus group B, ranging between 1.36 X 10^{-3} to 4.78 X 10^{-4} substitution/ site/ year (Cited by Boyle *et al*., 2014 from Lahon *et al.,* 2012).

The second most variable protein, NS3 of AHSV seems to evolve under positive selection pressure (Quan *et al.,* 2008). The NS3 participates in the alteration of the cell membrane and virus exit (Martin *et al*., 1998; van Niekerk *et al*., 2001a) and because of its membrane-association; it may be under immunological pressure (Quan *et al.,* 2008). Viral proteins (VP1, VP2, VP5, VP7, NS1 and NS3) play an important role in viral adaptation to the host cells (Wilson *et al.,* 2009). Any major changes in the amino acid sequences of the genes associated with VP1, VP2, VP5, VP7, NS1 and NS3 may have an impact on cell attachment, environmental conditions required for replication as well as the transmission and pathogenesis of the virus (Wilson *et al*., 2009).

Genetic recombination has been reported in orbiviruses, such as BTV (He *et al.,* 2010), however, information regarding AHSV genetic recombination is still lacking. Similarly, a number of studies have reported on reassortment between different BTV strains (Oberst *et al.,* 1987; Allison *et al.,* 2012; Shafiq *et al.,* 2013; Shaw *et al.,* 2013; Nomikou *et al*., 2015) and other orbiviruses such as Wallal and Mudjinbarry viruses (Gorman *et al.,* 1978), and epizootic hemorrhagic disease virus (Allison *et al*., 2012). However, limited information regarding genetic reassortment and mutation of AHSV is available. The study conducted by Von Teichman and Smit (2008) reported on the genomic reassortment between the AHSV vaccine strain and wild-type strains. Following the study by Von Teichman and Smit (2008), Weyer *et al*., (2016) reported reassortment between wild-type strains and vaccine strains between 2004 and 2014 using 39 field isolates. However, a comprehensive study looking at more serotypes of AHSV over a longer evolutionary time period is required.

1.4.4 Implication of AHSV genetic variation

African horse sickness virus genetic variation and its contribution to viral evolution have not been well investigated as compared to BTV. However, several implications of AHSV genetic variations and the contribution to the viral evolution have been observed and also theoretically supported based on changes related to BTV. The three main mechanisms of virus genetic variation (mutation, recombination and reassortment) may play a significant role in producing viruses with new antigenic determinants and formation of novel viruses (Fleischmann, 1996; Manrubia and Lázaro, 2006). Furthermore, the three main mechanisms of genetic variation can

contribute to AHSV and BTV evolution, disease severity and changes in virus fitness and adaptation to new hosts or vectors (Oberst *et al.,* 1987; von Teichman and Smith, 2008; Meiring *et al.,* 2009; Niedbalski, 2013). Genetic variation can cause species that are closely related to possess different biological properties such as disease severity, immune evasion, tissue tropism, geographic and seasonal distribution as well as host range (Walker and Cowley, 1999; Makadiya, 2007; Domingo, 2010).

Genetic variation in both AHSV and BTV as a result of mutation, reassortment and recombination can create a condition whereby the most environmentally adapted combinations emerge, leading to an increase in viral survival against host defences (Walker and Cowley, 1999). Genetic mutation, recombination and reassortment among orbiviruses such as BTV and AHSV can generate quasispecies populations within the host cells, which may then encourage the variants with the most favourable or best fitness to be selected under diverse environmental conditions (Niedbalski, 2013). Genome reassortment events are uniquely restricted to segmented viruses such as AHSV and can lead to the formation of new viruses with selective advantages and increase virus diversification (Gorman *et al*., 1978; Manrubia and Lázaro, 2006; Wilson *et al*., 2009). Genome reassortment and/or recombination between field isolates and attenuated vaccine virus can cause the virus to revert to virulence or to produce virulent variants as previously observed by Weyer and co-workers (2016).

Virus genetic variations can have an adverse effect on molecular and serological diagnosis of disease pathogens. Changes in viral proteins could have an adverse effect on the binding of diagnostic reagents such as monoclonal antibodies (Walker and Cowley, 1999), while a nucleotide change could prevent primers from binding to the target sequences (Kwok *et al*., 1990). The VP1, VP3, VP4 and VP6 form the AHSV subcore and play an essential role in the transcription; capping replication as well as packaging of the RNAs (Wilson *et al*., 2009) and therefore changes in their amino acids may affect their functions. For instance, VP1 of the AHSV, which is an RNA-dependent RNA polymerase, depends on temperature, thus if the changes in the genome segment that encodes for VP1 result in structural differences of the coded proteins, the geographic as well as seasonal distribution of AHSV

transmission could be affected (Wilson *et al*., 2009). Climate and evolution can interact to extend transmission rates as well as increasing the development rates of AHSV within the insect vector (CEH Horse Report, 2009; Backer and Nodelijk, 2011). Genetic changes, particularly the changes that affect the VP2 coding region where specific primers or probes can be designed, can cause difficulties in the detection of genetic variants within each AHSV serotype (OIE Terrestrial Manual, 2017). In a study by Maan *et al*., (2013), an AHSV-7 (SEN2007/06) isolate from Senegal, derived from horse that died in 2007, failed to amplify with type 7 specific primers, while the same primers successfully amplified type 7 reference strain suggesting genetic changes in the AHSV-7 sequence. Van Niekerk *et al* (2001b) investigated the protein sequences of NS3 using AHSV field isolates, reference strains as well as vaccine strains and found variation as much as 36.3% across serotypes and 27.6% within serotypes. Based on their findings, they suggested that the sequence variation of NS3 can be used to draw a distinction between the field viruses and live attenuated vaccine strains, and can also be useful during outbreaks of AHS disease. Koekemoer (2008) developed a real-time PCR method for the detection of genome segment 2 in all nine distinct AHSV serotypes. Koekemoer (2008) then concluded that genetic variation that has appeared over time makes the design and the utilisation of a single set of real-time PCR detection probes for the identification of all nine serotypes impossible. And therefore, periodic updating of the primer sequences will be required.

1.5 AIMS AND OBJECTIVES

The aim of this study was to comprehensively investigate the evolutionary dynamics of AHSV using historical archival isolates and recent isolates collected between 1933 and 2014. In this study, the complete genome of 101 AHSVs has been sequenced and analysed comprehensively with various bioinformatics programs in order to determine the evolutionary dynamics in all nine AHSV serotypes. The main objective of this study was to characterise the effect of the genetic mutation, recombination and reassortment on the evolutionary dynamics of African horse sickness virus genomes. To achieve this objective, the following aspects were investigated: The aim of this study was to comprehensively investigate the evolutionary dynamics of AHSV using historical archival isolates and recent isolates

collected between 1933 and 2014. In this study, the complete genome of 101 AHSVs has been sequenced and analysed comprehensively with various bioinformatics programs in order to determine the evolutionary dynamics in all nine AHSV serotypes. The main objective of this study was to characterise the effect of the genetic mutation, recombination and reassortment on the evolutionary dynamics of African horse sickness virus genomes. To achieve this objective, the following aspects were investigated: (1) the genetic diversity of African horse sickness virus genome and determine the substitution rates and selection pressure, (2) intragenic recombination in AHSV, and (3) genomic reassortment between different AHSV field isolates collected over the years, as well as experimentally investigating genomic reassortment between two different AHSV (AHSV 4 and 9) serotypes whose sequences have been elucidated.

1.6 STUDY PROBLEM

The African horse sickness virus features and characteristics make it difficult for the virus to be eradicated. Given the seriousness of the disease and its devastating mortality rate of more than 90% particularly in horses, AHS is a serious economic burden (Theiler, 1921; Mellor and Hamblin, 2004). The high mortality rates and measures imposed to limit AHSV spread impact on the international horse trade, which results in major economic losses (Zientara *et al*., 2015), particularly by the countries within the endemic regions. AHSV is an RNA virus and therefore lacks proof-reading enzyme activities to correct misincorporation during replication, subjecting it to genetic mutations (Fleischmann, 1996; Manrubia and Lázaro, 2006; Makadiya, 2007). RNA viruses mutate faster than DNA viruses. Genetic recombination has been reported among RNA viruses and in orbiviruses, such as BTV (He *et al.,* 2010). However, information regarding AHSV genetic recombination is still lacking. Since AHSV contains segmented dsRNA genome, it can undergo genomic reassortment (Wilson *et al*., 2009). There are different studies that have reported reassortment in orbiviruses (Gorman *et al.,* 1978; Oberst *et al.,* 1987; Allison *et al.,* 2012; Shafiq *et al.,* 2013; Shaw *et al.,* 2013). However, the effect of genomic reassortment on AHSV was not well determined, particularly among the field viruses over a long period of time.

African horse sickness virus may have over the years been subjected to genetic variation due to the occurrence of genetic mutation, genome segment reassortment as well as recombination events. The study conducted by von Teichman and Smit (2008) have reported on the genetic reassortment between the AHSV vaccine strain and the wild-type, however, prior to the commencement of this study, there was very little information regarding the evolutionary dynamics of AHSV. It is important to know how the AHSV evolution take place and what effect does the evolution have on the genome of the virus. Such information about the AHSV evolutionary dynamics will assist in the future vaccine development and other AHSV investigations.

During the planning of this study, no study has reported on the analysis of a large group of AHSV collected over a time-span of decades covering all the serotype in the species. Many countries in the AHSV free regions may not to vaccinate their horses due to fears associated with viral reassortment as well as reversion to virulence (Robin *et al*., 2016; Mellor and Hamblin, 2004). Due to fears regarding the reversion to virulence particularly due to genomic recombination and the reassortment events between the attenuated vaccine and field strains as previously reported (Weyer, *et al*., 2016), it is important to investigate if there is evidence of this in AHSV.

CHAPTER 2 DETERMINING THE SUBSTITUTION RATES AND SELECTION PRESSURE USING COMPLETE GENOME SEQUENCES FROM REPRESENTATIVES OF ALL SEROTYPES OF AFRICAN HORSE SICKNESS VIRUS COLLECTED BETWEEN 1960 AND 2014

2.1 BACKGROUND

African horse sickness virus (AHSV) is a member of the *Orbivirus* genus within the *Reoviridae* family (Attoui *et al*., 2012). It causes a non-contagious, but highly infectious arthropod-borne disease of equids known as African horse sickness (AHS). Virus particles are non-enveloped, icosahedral and consists of a double capsid layer containing the genome comprised of ten linear, double stranded RNA segments (Polson and Deeks, 1963; Bremer, 1976). The genome segments are identified based on their molecular weights ranging in size between 3965 and 756 bp and are known as segment-1 to segment-10 based on their decreasing molecular weight (Bremer, 1976). All 10 AHSV ds-cDNA segments sequences can be identified by utilizing the sequence–independent whole genome amplification followed by ultra-deep sequencing (Potgieter *et al*., 2009).

Phylogenetics has become an important tool to study genetic relationships based on molecular sequence identity or similarity (Drummond *et al*., 2006). The basic method to study virus evolution is through sequencing of the RNA or cDNA produced from total viral RNA by RT-PCR (Isakov *et al*., 2015). Sequencing of dsRNA requires conversion to cDNA. The conversion can be done by the use of sequence–independent whole genome amplification, which is sensitive and specific for dsRNA viruses and is used to amplify cDNA (Potgieter *et al*., 2009). The amplified cDNA can then be subjected to sequencing. Sanger sequencing has been used over the years as a sequencing tool (Sanger *et al*., 1977). The limitations of Sanger sequencing include low resolution; i.e. its ability to detect only high abundant variants (Isakov *et al*., 2015). Unlike Sanger sequencing, next-generation sequencing (NGS) produces thousands to billions of sequence fragment reads, and provides an ideal framework for characterization of highly variable pathogens, with deep sequencing resolution able to capture minority variants (Prosperi and Salemi, 2012). The genome sequences can be analysed by creating alignments using ClustalW in the MEGA v6 software (Tamura *et al*., 2013). ClustalW is used to align multiple nucleotides or protein sequences (Tamura *et al*., 2013). The most similar sequences are aligned

first, followed by the least similar sequences until the whole alignment is reached (Tamura *et al*., 2013). Maximum-likelihood model selection program, also incorporated in the MEGA v6 software can be employed to determine the model of nucleotide substitution (Tamura *et al*., 2013). See section 4.1 for more details on substitution models.

Phylogenetic softwares can also be used to estimates the evolutionary history of an organism, by comparing DNA, RNA or protein sequences of different organisms in that species (Baldauf, 2003). Molecular sequence data could be used directly to assess ancestral-descendant character transitions and can be achieved by applying computational evolutionary biology, statistical phylogenetics as well as coalescent based population genetics (Drummond *et al*., 2012). The BEAUti and the BEAST v1.8.1 software package that implements a family of Markov chain Monte Carlo (MCMC) algorithms for Bayesian phylogenetic inference, divergence time dating, coalescent analysis, phylogeography as well as the related molecular evolutionary analyses is widely used (Drummond *et al*., 2012). Included in the package is the program known as the Bayesian Evolutionary Analysis Utility (BEUti), which allows for the access to advanced models for molecular genetic sequence and provides tools for visualizing and summarizing multiple species coalescent as well as phylogeographic analyses (Drummond *et al*., 2012). By utilizing Markov Chain Monte Carlo (MCMC) implemented in the BEAUti and BEAST software package v1.8.1 the rate of molecular evolution or divergence between a group of organisms (substitution per site per year) and time to the most recent common ancestor (TMRCA) can be estimated for each genome (Drummond *et al*., 2012). The gene and site-specific selection pressure can be calculated and analysed using single-likelihood ancestor counting (SLAC) method and Hasegawa-Kishino-Yano, 85 (HKY85) nucleotide substitution model that are both incorporated in the Datamonkey server (www.datamonkey.org). Datamonkey is a web-based phylogenetic analysis tool used for analysing among others recombination, evolutionary gene fingerprinting, codon model selection, and co-evolution (Delport *et al*., 2010).

This chapter investigated the evolutionary dynamics of AHSV at the hand of complete genome sequences generated from isolates collected between 1933 and 2014 in Sub-Saharan Africa. The phylogenetic relationship among the nine different

AHSV serotypes was determined by analysing genome segments and predicted amino acid sequence identities between and/or within serotypes (Segment 2) of all the 87 virus isolates. Evolutionary rate was determined by the rate of nucleotide substitution, Time to Most Recent Common Ancestor (TMRCA) and selection pressure acting on each of the ten genomic segments were estimated. In comparison to BTV, little is known about AHSV genome evolution even though they are transmitted by the same insect vector and are biologically similar (Du Toit, 1944; Coetzer and Guthrie, 2004).

2.2 AIMS OF THE STUDY

1. Determine the genetic relationship among the nine different AHSV serotypes from 87 virus Isolates.
2. Calculate the rate of nucleotide substitution, Time to Most Recent Common Ancestor (TMRCA) and selection pressure acting on each of the ten AHSV genomic segments across nine serotypes.

2.3 MATERIAL AND METHODS

This section discusses the experimental approach and the material used to determine the genetic relationship between and among the AHSV serotypes as well as selection pressure acting on the AHSV serotypes.

2.3.1 Viruses

The Agricultural Research Council of South Africa – Onderstepoort Veterinary Institute is one of the OIE Reference Centre for AHSV and BTV. Equine tissues and blood samples submitted for AHS diagnoses were tested using either RT-PCR or virus neutralization, and the positive samples were submitted for virus isolation. In order to study the phylogenetic relationship of AHSV, eighty-seven viruses representing all nine serotypes, spanning from 1961 to 2014 were included in this study (Appendix 1, Table 1). To investigate the rate of nucleotide substitution, TMRCA and selection pressure, all (ten) the vaccine derived isolates were removed and from the list of 87 isolates. The vertical evolution is concerned with a genetic inheritance from parental offspring, and therefore isolates that were identical to vaccine strains were removed from the analysis, but retained during the horizontal

evolutionary analysis. The vaccine derived isolates included: AHSV-1_22_01, AHSV-1_107_09, AHSV-3_DG 25324_14, AHSV-3_DG 25327_14, AHSV-3_DG 25423_14, AHSV-4_90_96, AHSV-4_64_99, AHSV-4_97_99, AHSV-5_2_96 and AHSV-5_42_01. These sequences had a significant homology (99%) to the AHS attenuated live virus (ALV) vaccine isolates included AHSV-1_29_62, AHSV-3_13_63, and AHSV-4_32_62 (KT030330 – KT030359) manufactured by Onderstepoort Biological Products (OBP) Ltd as per their individual serotypes. Serotype AHSV-8 _29_00 was included in these analyses, but also downloaded from GenBank (KP009691 - KP009700) and both were used in the horizontal analyses, but only one was omitted from vertical analysis. A total of 14 AHSV genome sequences

Serotypes 1, 2 and 7 were underrepresented (5.8%, 6.9% and 8.0%) in comparison with other serotypes. The majority of the samples were either submitted from South Africa (64) or neighbouring countries (Namibia = 2; Mozambique = 1; Zimbabwe = 3), while two of the 1961/62 Reference strains were obtained from Chad and Kenya. No information concerning the countries of origin was available for 15 samples. They were included in the study since the aim of this study was neither to infer the demographic history nor determine the extent of geographic structure of AHSV population (Table 1, Appendix 1).

2.3.2 Tissue culture and virus infection

Monolayers of baby hamster kidney (BHK_{21}) cells were grown in a $75cm^2$ flask containing Dulbecco's Modified Essential Medium (DMEM) (Gibco) supplemented with 5% (v/v) foetal bovine serum (FBS), 5ml L-glutamine (200mM) and 25µg/ml Penicillin-Streptomycin-Amphotericin B (PSA) (Gibco). The flasks were incubated at 37^oC supplied with 5% CO_2 until the cells show 70-80% confluence. The BHK_{21} cells were then coated with fresh DMEM without serum. Freeze-dried AHSVs were reconstituted in fresh DMEM supplemented with L-glutamine and 25µg/ml Penicillin-Streptomycin-Amphotericin B before inoculation in BHK_{21} cells. The BHK_{21} cells inoculation was followed by incubation at 37^oC with 5% CO_2 until all the cells were showing the cytopathic effect (CPE) of 4+ or 90% cell lysis.

2.3.3 Total dsRNA extraction

Total dsRNA extraction was carried-out as described by Potgieter *et al*., 2009. Briefly; infected cells of 4+ CPE were scraped from the flask and centrifuged at 3000xg for 5 minutes. The centrifugation was followed by RNA extraction using a commercial guanidinium isothiocyanate reagent (TRI-Reagent) (Ambion). The aqueous phase was mixed with $CHCl_3$ prior to centrifugation at 12 000 x g for 40 minutes at 4^{o}C. The supernatants were placed into a clean tube followed by the addition of an equal volume of isopropanol and centrifuged at 12 000 x g for 45 minutes at 4^{o}C to precipitate the RNA. The RNA was dissolved in 90 µl elution buffer. Single-stranded RNA (ssRNA) was precipitated by adding LiCl to a final concentration of 2.6M and incubation at 4^{o}C overnight. The ssRNA was pelleted by centrifugation for 40 minutes at 4^{o}C. The centrifugation step was followed by dsRNA purification from collecting supernatants using a MiniElute gel extraction kit (Qiagen), according the manufacturer's specifications). After extraction, the dsRNA genome was visualised on 1% agarose gel.

2.3.4 Oligo ligation

The 'anchor primer', PC3-T7 loop (5'-p-GGATCCCGGGAATTCG-GTAATACGACTCACTATATTTTTATAGTGAGTCGTATTA–OH-3') synthesized by Tib Molbiol was ligated to the dsRNA as previously outlined by Potgieter *et al* (2009). Briefly; The anchor primer, PC3-T7 loop, was added to the reaction containing 0.5 - 200 ng RNA, 3 µl DMSO (Sigma), PC3-T7 loop primer (300 ng/µl), 1 Weiss unit T4 RNA ligase (Thermo Scientific) in a 10X ligation buffer consisting of 10 µl of 60% PEG 6000; 0.5M HEPES, pH8.0; 0.18M $MgCl_2$; 10mM ATP; 30.6mM DTT with BSA 0.1 % (Takara) for a final volume of 30 µl and incubated at 37°C overnight.

2.3.5 Oligo-Ligation products purification

The dsRNA was purified using the MinElute Gel extraction kit as per manufacturer specifications (Qiagen). Briefly; the volume of the oligo-ligation reaction mixture was increased to 100 µl with Ultra-pure water. Three times buffer QC (Qiagen) and an equal volume of isopropanol was added. The solution was centrifuged through a column for 1 minute at 12 000 x g. The column was washed with PE buffer (Qiagen) and centrifuged at 12 000 x g, for 1 min. The flow-through was discarded and the

columns placed in a clean 1.5 ml collection tube. The ligated dsRNA was eluted in 12 µl elution buffer (Qiagen) and stored at 4^{o}C.

2.3.6 Sequence–independent cDNA synthesis

The dsRNA was denatured using 300 mM methyl mercury hydroxide (MMOH) and reverse transcribed into complementary DNA (cDNA) as described previously by Potgieter *et al* (2009). The dsRNA was denatured using 300 mM mercury hydroxide (MMOH) and incubated at room temperature for 30 minutes. After denaturation of dsRNA, a cDNA mixture was prepared by adding 2.5mM dNTP mix (Takara), cDNA buffer consisting of 300 mM MMOH, 1M Tris-Cl pH 8.3, 3M KCL, 1M β-M ethanol, 15U/µl Cloned Avian Myeloblastosis Virus (AMV) Reverse Transcriptase (Invitrogen) and denatured RNA into 0.5 ml tube. The mixture was incubated at 42^{o}C for 55 minutes, at 55°C for about 15 minutes and allowed to cool to 20°C before adding 1N NaOH. After adding 1N NaOH, the content was incubation at 65°C for 30 minutes and cooling at 20 °C to remove RNA strands. After the removal of RNA, Tris buffer (pH 8.3) was added to the 0.1M final concentration and followed by adding HCl to the 0.1M final concentration and incubated at 65°C for 60 minutes followed by cooling at 20°C.

2.3.7 PCR amplification and cDNA purification

The Phusion® High-Fidelity PCR kit (NEB) was used to amplify the cDNA as described by Potgieter *et al* (2009). A PCR mixture was prepared by adding a 5x Phusion buffer, 10 mM dNTPs (Phusion HF PCR kit- NEB), 50 pmol/µl PC2 primer (5' PO_4-CCGAATTCCCGGGATCC-OH 3'), H_2O, cDNA and Phusion Polymerase (2U/µl) making a final volume of 50 µl. The reaction had an initial denaturation at 98°C for 1.5 min, followed by 20 cycles of 94°C for 15 sec, 65^{o}C for 30 seconds, 72°C for 4 minutes and a final elongation step at 72°C for 5 min. The amplified cDNA was visualized and evaluated on a 1.2 % agarose gel that was stained with ethidium bromide. The cDNA products were subsequently purified using the GeneJET Gel Extraction kit (Thermo Scientific) as per manufacturer specifications.

2.3.8 AHSV ds-cDNA sequencing

Purified PCR products were commercially sequenced using an Illumina MiSeq instrument (Illumina, Inc, CA, USA) at Inqaba Biotec (Pretoria, South Africa).

Libraries were put together using the Nextera DNA Sample Preparation Kit (Illumina, Inc., San Diego, CA, USA) using sequencing adapters and barcodes with the Illumina MiSeq sequencing reagent v3. The purified AHSV dsRNA PCR products were sequenced on a single run. Approximately 75Mb of data was obtained per virus consisting of 300-bp paired-end reads.

2.3.9 AHSV sequence assembly

The forward and reverse reads of each sample were imported into CLC Genomics Workbench v7.0 (CLC bio, Aarhus, Denmark). Pair-end sequence reads were trimmed at the 5' and 3' end in order to remove the sequencing adapters. All the reads below 50 and above 1000 bp were discarded. The trimmed sequences were assembled using *de novo* assembly and best coverages contigs were selected and the consensus sequences extracted. The minimum contig length of 500 bp was set in order to accommodate the smaller genome segment of about 756 to 763 bp. The reads were subsequently mapped to reference virus sequences obtained from NCBI. Consensus sequences obtained via *de novo* assembly and mapping to reference genomes were assembled and new contigs obtained for each of the 10 genome segments. The contigs were trimmed to include the orbivirus start position (GTT) and the end position (TAC). The conflicts were resolved to obtain the final consensus of each genome segment for all the AHSV isolates. The final consensus sequence of each African horse sickness virus genome segment was submitted to GenBank and accession numbers assigned (KP939368- KP940236).

2.3.10 Phylogenetic analysis

Alignment of sequences from each individual genome segment was created using ClustalW in MEGA v6 software (Tamura *et al*., 2013). The model of nucleotide substitution that best fits each data set was determined using the maximum likelihood model selection in MEGA v6 software (Tamura *et al*., 2013). MEGA v6 was used to build the neighbour-joining trees that were then used to construct Maximum-likelihood trees (ML-Trees). The construction of ML-Trees was done using bootstrapping in order to assess the robustness of the tree with 1 000 replicates.

2.3.11 Estimation of evolutionary dynamics

Rates of molecular evolution (substitutions per site per year) and time to the most recent common ancestor (TMRCA in years) were estimated for each genome segment using Bayesian Markov chain Monte Carlo (MCMC) in the BEAUTi and BEAST software package v1.8.1 (Drummond *et al*., 2012). Each AHSV segment across all nine distinct serotypes from 87 isolate was analysed using the General Time Reversible (GTR) model with a gamma distribution (T4) of site variation (Tamura *et al*., 2013). A relaxed molecular clock and general Bayesian skyline coalescent prior was used under an adequate number of generations to ensure convergence of all parameters. The posterior distribution and evolutionary parameters were assessed using the program Tracer v1.6 (available at http://tree.bio.ed.ac.uk/software/tracer). The target tree of each segment was analysed using a 100 burn-in number in TreeAnnotator v1.8.1, available as part of the BEAST package (available at http://tree.bio.ed.ac.uk/software/BEAST) and the maximum clade credibility tree with posterior estimates of TMRCA was visualized in FigTree v1.4.2 (available at http://tree.bio.ed.ac.uk/software/figtree). Gene and site-specific selection pressure on individual codons were calculated and analysed using the single likelihood ancestor counting (SLAC) method and Hasegawa-Kishino-Yano, 85 (HKY85) nucleotide substitution model incorporated in the Datamonkey server (www.datamonkey.org). The positive selection and negative selection were determined by ω=dN/dS (non-synonymous substitutions/ synonymous substitutions). When the ratio was less than one (ω > 1), sites were considered to be under positive selection. For the ratio greater than one (ω < 1), sites were predicted to be under negative selection.

2.4 RESULTS

A total of 77 AHSV isolates chosen across nine distinct serotypes were analysed to determine the phylogenetic relationship between them as well as determining the substitution rates, Time to Most Recent Common Ancestor (TMRCA) and selection pressure. These together with other findings are reported in this section.

2.4.1 The dsRNA extractions

The dsRNA extraction was carried out from monolayers of BHK cells infected with all nine AHSV serotypes. The dsRNA was visualized on 1% agarose gel and uniform migration patterns were observed across all nine distinctive AHSV serotypes. Eight distinct bands were observed representing the 10 individual genome segments (Figure 2.1).

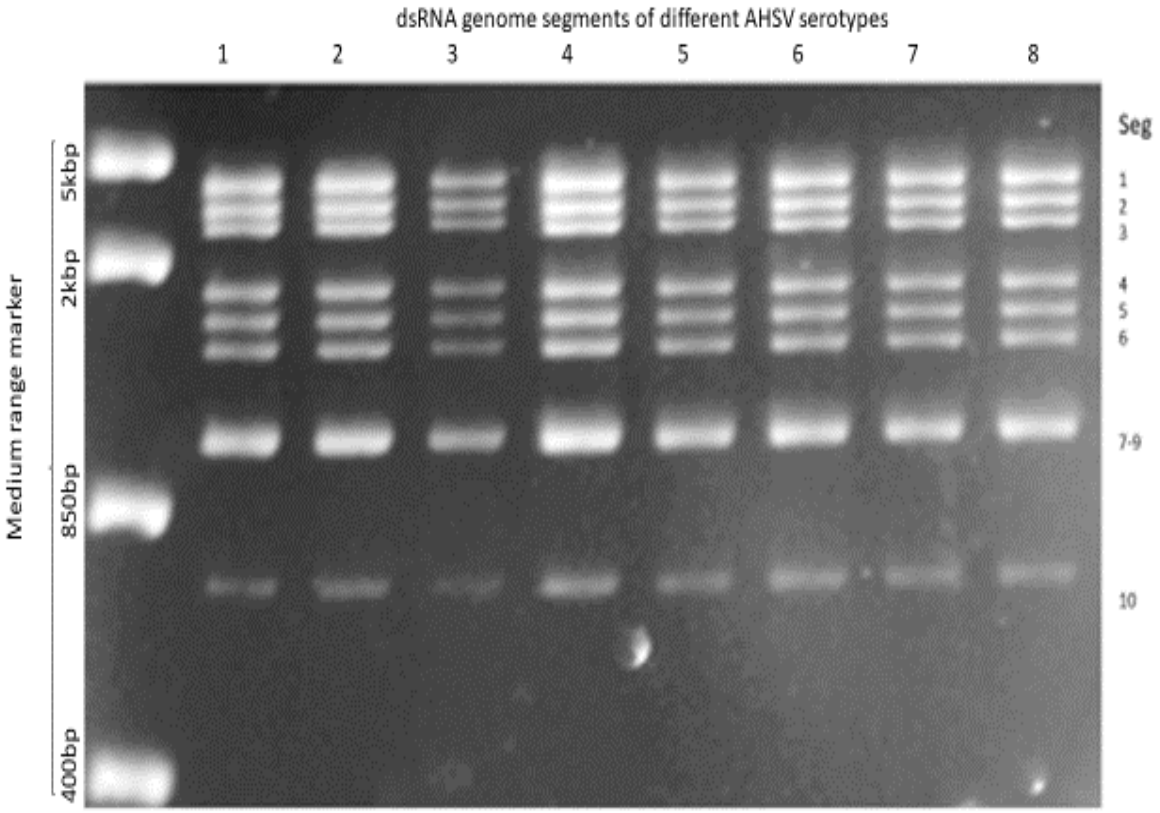

Figure 2.1: A 1% agarose gel showing dsRNA segments of AHSV serotypes 1 to 8 extracted from infected BHK cells using TRI Reagent and stained with EtBr. Only eight out of nine are shown on the gel, but serotype 9 had similar electropherotypic patterns as those in display. Segment 1-6 migrates as individual segments while no size distinction between 7, 8 and 9 could be observed on 1% agarose gel. Segment 10 also migrate as individual fragment.

2.4.2 Synthesis and amplification of cDNA

The PC3-T7 primer was ligated to both ends of the dsRNA from each of the serotypes (AHSV 1-9) using 30 U T4 RNA ligase (Thermo Scientific) and reverse-transcribed using cloned AMV (Invitrogen). The cDNA was subsequently amplified using a single primer (PC2) complementary to the ligated adaptor PC3-T7. The method ensures un-bias amplification of all the genome segments. The amplified

cDNA products were analysed on a 1% agarose gel stained with EtBr. Similar to the dsRNA genome segments, eight amplicons were observed with the first six amplicons representing single segments (segments 1 to 6) while fragment seven contained three segments (S7, 8 and 9). Segment 10 was the eighth and a bottom fragment (Figure 2.2).

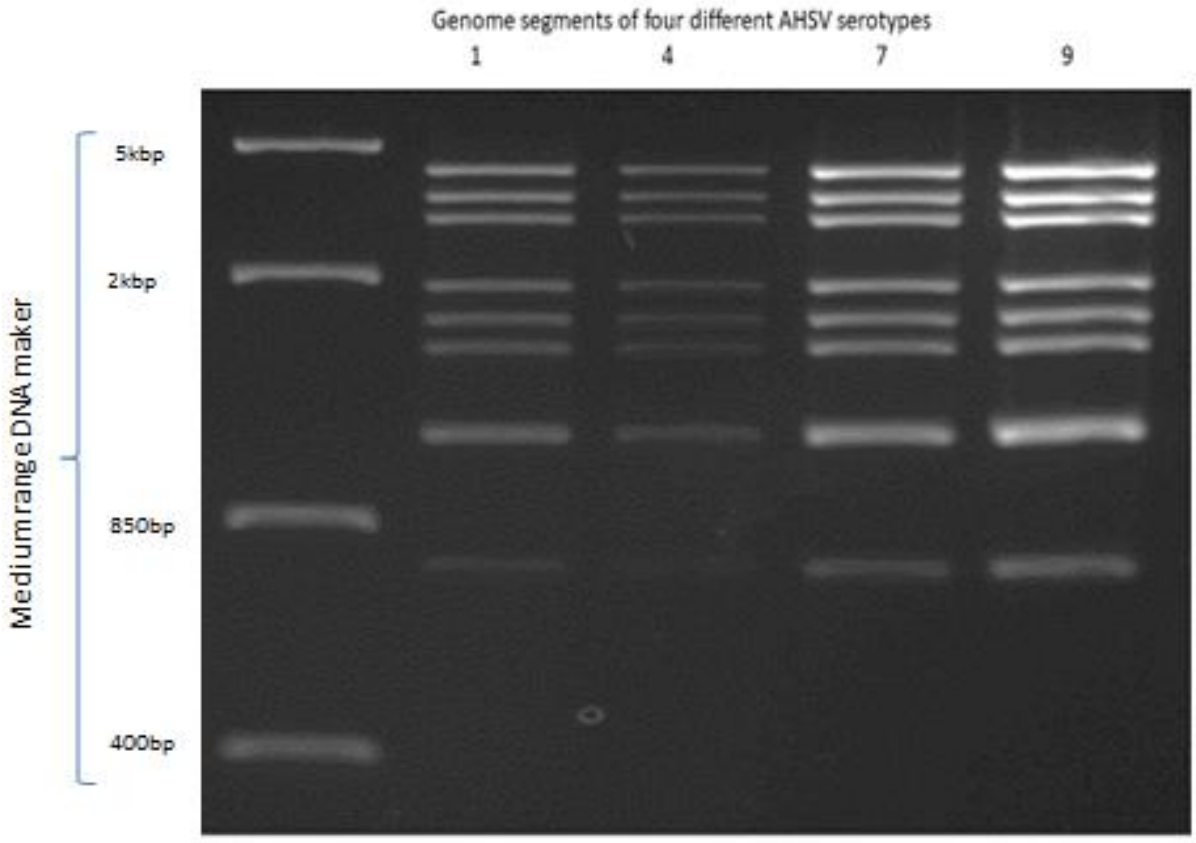

Figure 2.2. A 1% agarose gel showing eight ds-cDNA segments containing 10 cDNA segments of four different AHSV serotypes. The first six segments migrate individually, while the 7th fragment contains three segments (S7, 8 and 9) that are co-migrating. Segment 10 is migrating at the bottom of the gel.

2.4.3 Whole genome sequence analysis

The cDNA of 87 viruses was subjected to Illumina Next Generation sequencing. An average of 250 000 reads of about 250 bp in length were obtained per sample. Consensus sequences of all ten genome segments were obtained by *de novo* assembly and serotype-specific alignments of each individual virus. An average of 96% of the trimmed reads was mapped against the new consensus sequences and variant detection of 10% was performed. These variants were annotated on the consensus sequences since they could represent either mixed infections or

quasispecies. The total genome size of each virus was ~19.5 kb. The sizes of each genome segment (Seg-1 to 10), as well as the length of the predicted proteins, are summarized in Table 2.2. An 11 and 12 nucleotide deletions were observed in the 5'-UTR region of Seg-7 (VP7) and Seg-8 (NS2) respectively (See figure 2.3 and 2.4 respectively). Deletions were observed for both Seg-7 and Seg-8 in isolates AHSV-4_19_97 and AHSV-4_91_00, while AHSV-4_31_00 showed only a deletion in Seg-7. Since no information regarding the virulence or growth capabilities of these isolates are known, therefore no conclusion as to the reason for the deletion could be provided.

The percentage sequence identity, number of differences and gaps between each of the genome segments from an isolate were compared against each other and the reference (1961 - 1975) strains. Sequences of AHSV-1_22_01, AHSV-1_107 _09, AHSV-3 _DG25324_14, AHSV-3_DG_25327_14, AHSV-3 DG 25423_14, AHSV-4_90_96, AHSV-4_64_99 and AHSV-4_97_99 were each 99% identical to the AHS attenuated live virus (ALV) vaccine including AHSV-1_29_62, AHSV-3_13_63, and AHSV-4_32_62 (GenBank accession numbers KT030330 – KT030359) manufactured by Onderstepoort Biological Products (OBP) Ltd as per their individual serotypes (ALV sequences obtained from Guthrie *et al.,* 2015). These isolates were removed from the vertical evolutionary analysis (Genetic mutation analysis) because they were not part of the clock, but since horizontal evolution has been described in the ALV sequences, they were retained during recombination and reassortment analysis. Similar to BTV and EHDV, AHSV has evolved two homologues of Seg-9 encoding viral helicase (VP6), type I and type II. Unlike the clear separation based on geographic origins displayed by VP6 of EHDV, no geographic, serotype specific or chronologic link could be established between the two types of AHSV Seg-9 described here. Sequences of types I and II differ in length (1160 and 1169 bp) as well as predicted protein sizes (361 and 371 aa). Percentage sequence identities between type I and II were as low as 71 % for nucleotides and 62% based on amino acid sequence, but > 93% and > 96% sequence identity respectively within individual serotypes.

The percentage sequence identities in Seg-1 at the nucleotide level ranged between 88.70 to 100% and 97.62 to 100 % at the protein level (Table 2.1). There were no

significant changes, particularly in the amino acid sequences encoded by Seg-1. Within a serotype, Seg-2 has shown a high percentage of sequence identity for both nucleotide as well as amino acid (VP2). However, as expected the percentage sequence identity between the nine AHSV serotypes varied significantly and it was as low as 54.74% for nucleotide sequence and 49.06% for amino acid sequences across nine serotypes (inter-serotype), but highly identical and conserved within individual serotypes (intra-serotype) (Table 2.1, figure 4.3 and Appendix 2 figure 9). Similar to Seg-1, other AHSV segments such as Seg-3, Seg-4, Seg-5, Seg-7 and Seg-8 had a high percentage sequence identity at the nucleotide level and the amino acid sequences they encode. In the smallest genome segment (Seg-10) that encodes NS3 and NS3A protein, the percentage sequence identities between isolates were as low as 66.49% on the nucleotide level and 63.18 % for the amino acids they encode. The analysis of Seg-10 nucleotide and protein sequence were made across three distinct clades resulting in high protein sequence variation compared to nucleotide sequence variation. Although these clades were not defined in this analysis, the ML phylogenetic tree for Seg-10 has shown three main branches confirming the existence of these clades (Appendix 2, Figure 3).

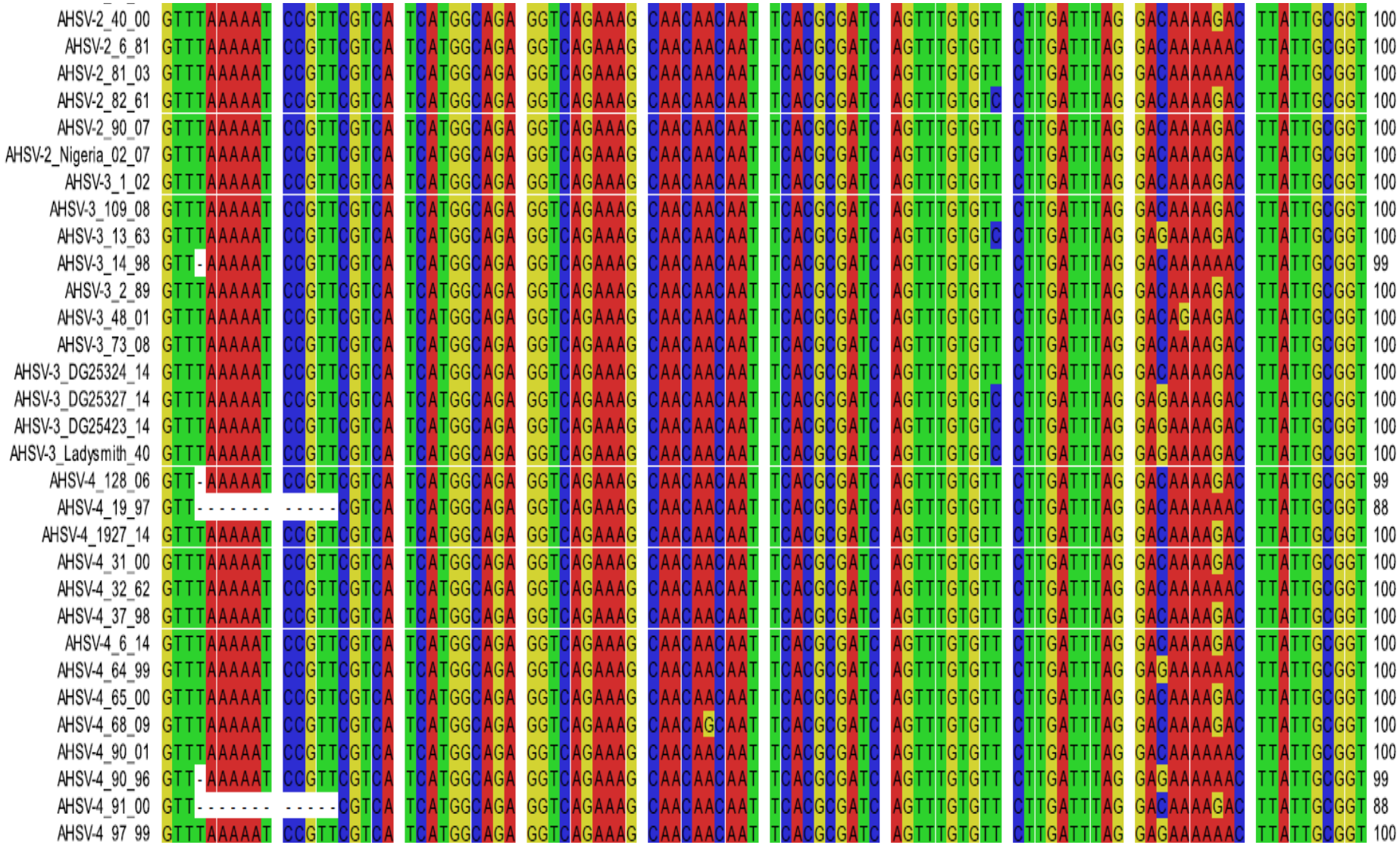

Figure 2.3. Selected section of nucleotide sequence alignment between serotypes of AHSV-2, AHSV-3 and AHSV-4, showing 12 nucleotide deletions in the 5'-UTR region of segment 8 (Seg-8) of isolates AHSV-4_19_97 and AHSV-4_91_00.

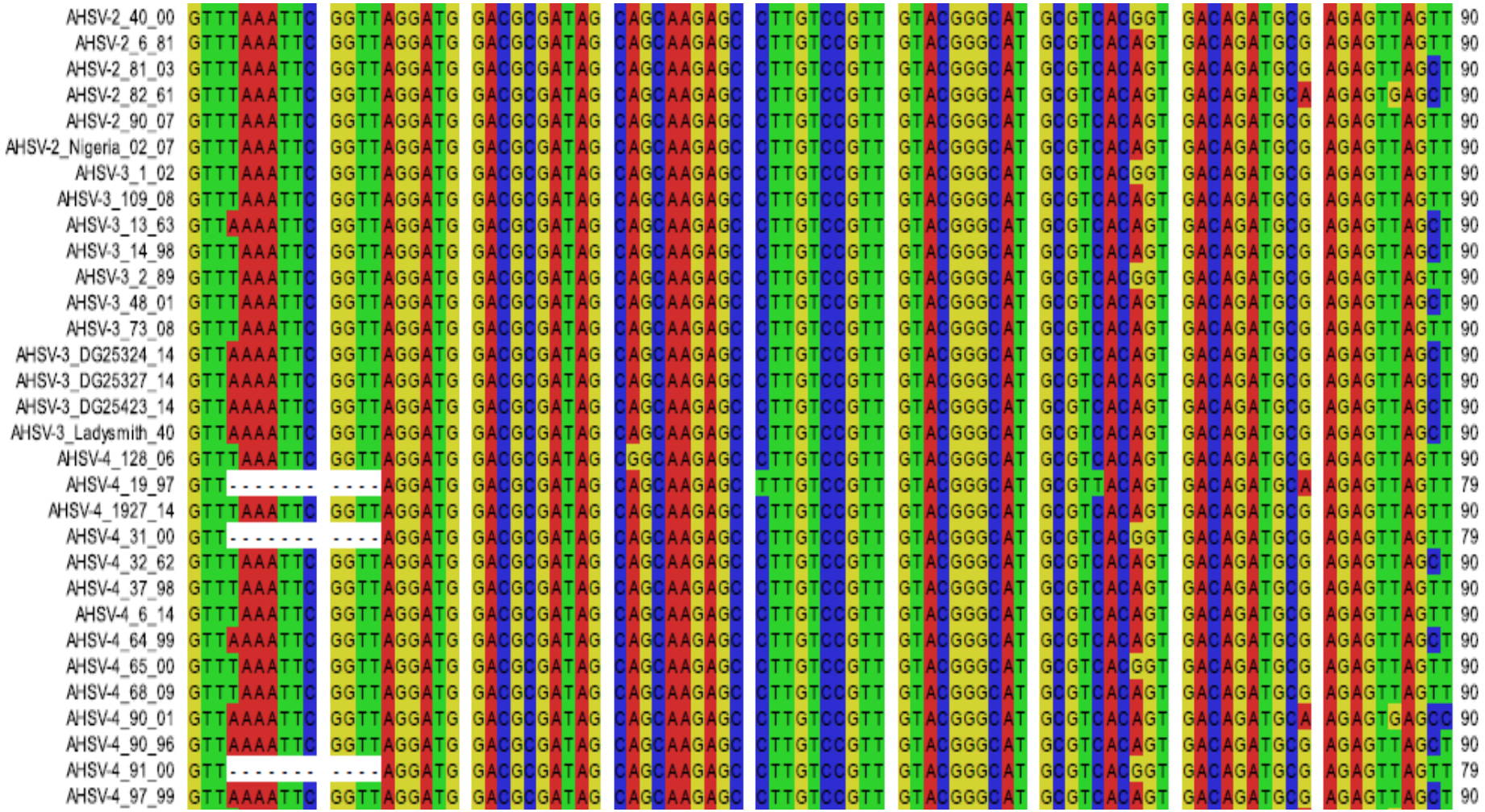

Figure 2.4. Selected section of nucleotide sequence alignment between serotypes of AHSV-2, AHSV-3 and AHSV-4, showing 11 nucleotide deletions in the 5'-UTR region of segment 7 (Seg-7) of isolates AHSV-4_19_97, AHSV-4_91_00 and AHSV-4_31_00.

Table 2.1. Characteristics of dsRNA genome segments and predicted amino acid of sequence identities between or within serotypes (Seg-2) determined from 87 AHSV isolates

Genome segment (Size in bp)	ORFs bp (including stop codon)	Protein nomenclature (Number of Amino acids)	Percentage sequence identity. Nucleotide. (Overall mean distance)	Percentage sequence identity. Protein. (Overall mean distance)
Seg-1 (3965)	14 – 3931	**VP1** (1305)	88.70 – 100 (0.061)	97.62 – 100 (0.010)
Seg-2 (3202-3229) (Inter-serotype)	13 – 3168 / 3195	**VP2** (1051 – 1060)	54.74 – 100 (1.278)	49.06 – 100 (0.551)
Seg-2: AHSV-1 (3218)	13 – 3183	**VP2: AHSV-1** (1056)	96.99 - 100	98.30 – 100
Seg-2: AHSV-2 (3218)	13 – 3183	**VP2: AHSV-2** (1056)	97.08 – 99.81	98.20 – 100
Seg-2: AHSV-3 (3221)	13 – 3186	**VP2: AHSV-3** (1057)	96.46 - 100	97.26 - 100
Seg-2: AHSV-4 (3229)	13 – 3195	**VP2: AHSV-4** (1060)	96.90 - 100	97.36 – 100
Seg-2: AHSV-5 (3217)	13 – 3186	**VP2: AHSV-5** (1057)	96.98 - 100	97.54 – 100
Seg-2: AHSV-6 (3202)	13 – 3168	**VP2: AHSV-6** (1051)	97.22 - 100	98.30 – 100
Seg-2: AHSV-7 (3221)	13 – 3168	**VP2: AHSV-7** (1057)	94.88 - 100	96.22 – 100
Seg-2: AHSV-8 (3217)	13 – 3168	**VP2: AHSV-8** (1057)	97.05 - 100	98.58 - 100
Seg-2: AHSV-9 (3202)	13 – 3171	**VP2: AHSV-9** (1052)	89.82 - 100	97.62 - 100
Seg-3 (2792)	27 – 2744	**VP3** (905)	93.95 – 100 (0.029)	99.34 – 100 (0.002)
Seg-4 (1978)	12 – 1940	**VP4** (642)	92.06 – 100 (0.049)	95.64 – 100 (0.021)
Seg-5 (1748)	36 – 1682	**NS1** (548)	94.05 – 100 (0.026)	97.08 – 100 (0.010)
Seg-6 (1566)	20 – 1537	**VP5** (505)	70.31 – 100 (0.282)	75.64 – 100 (0.133)
Seg-7 (1156 - 1167)	7 / 18 – 1056 / 1067	**VP7** (349)	87.15 – 100 (0.047)	97.71 – 100 (0.007)
Seg-8 (1154 – 1166)	11 / 22 – 1108 /1119	**NS2** (365)	94.51 – 100 (0.026)	97.53 – 100 (0.009)
Seg-9 (1160 – 1169)	18 – 1118 / 1127 193 / 214 – 600 / 657	**VP6** (361 – 371) **NS4** (63 – 169)	71.10 -100 (0.217)	62.03 – 100 (0.209) 19.54 / 37.01 – 100 (0.393)
Seg-10 (756 – 763)	19 – 672 /675 49 - 672 /675	**NS3** (217 – 218) **NS3A** (207 - 208)	66.49 – 100 (0.239)	63.18 – 100 (0.187)

2.4.4 Evolutionary rate and selection pressure on AHSV genome

The Bayesian coalescent estimates of mean substitution rates for segments-2, -6, -7, and -9 were the highest of the AHSV genome segments, between 1.75 $x10^{-4}$ and 6.4 x 10^{-4} substitutions/ site/ year. Estimated substitution rates ranged between 1.5 – 6.4 $x10^{-4}$ substitutions/site/year over the whole genome (Table 2.2). The average age of the most recent common ancestor (MRCA) of Seg-2, Seg-6 and Seg-10 was 4.4 thousand years, yet for the remainder of the genome, the average age of their MRCA was 440 years.

The rate of nonsynonymous (d_N) to synonymous (d_S) substitutions was determined for each genome segment, in order to evaluate the influence that the estimated substitution rate might have on the biological functionality of the protein. All ten segments were predicted to be under purifying selection pressure ($d_N/d_S < 1$), while four of the segments included selected sites under positive selection pressure (0.1 significance level). Segment 4, encoding the viral capping enzyme had three codon sites estimated to be under positive selection. Codon 278 represents a Q > R (0.625: 0.375), codon 295 had T > A > S > R (0.66 : 0.23 : 0.07 : 0.05) and 630 had an R > G (0.77 : 0.23) substitution. Additional to the three sites under positive selection, Seg-4 had a high mean d_N/d_S rate of 0.11. Yet in order to conserve the functionality of the protein, 0.57 rate of overall mean distance of the nonsynonymous substitutions (Table 2.3). Similarly, Seg-9 encoding the viral helicase had a mean d_N/d_S rate of 0.36 and four sites under positive selection pressure (32, 41, 125 and 130). These might be misleading since the comparison included both the homologues of AHSV Seg-9. The two homologues were analysed individually and even though they had similar d_N/d_S rates (Type I, 0.39 and Type II, 0.37) no sites under positive selection were identified. The two most variable segments Seg-2 (VP2) and Seg-10 (NS3) both had sites under positive selection. Segment 10 had two positively selected sites, within its variable domain (139 and 143). Codon 139 was D>N>E>T>K>S (0.43 : 0.28 : 0.15 : 0.09 : 0.05 : 0.01) and 143 was E>K>G>R>D>A>Q (0.33 : 0.27 : 0.14 : 0.12 : 0.09 : 0.02 : 0.02) while Seg-2 had one site (449) when all sequences were compared. Segment 2 (Seg-2) had a high substitution rate (6.4 x 10^{-4}), high mean d_N/d_S rate (0.262) and high rate of substitution of position one and two in a codon (0.52; 0.89; 1.59). Genome segment 10 (Seg-10), encoding NS3/NS3A, the protein required for virus release from

infected cells, had a high d_N/d_S (0.173), but a low substitution rate (1.7×10^{-4}). Significant differences between the substitution rate and selection pressure of individual genomic segments were observed. The differences prompted an investigation into the influence of recombination and reassortment (See Chapter 4 and 5).

Table 2.2 Bayesian estimate of substitution rate and TMRCA (excluding ten vaccine derived isolates).

AHSV segment	Sequence length of coding region (bp)	Substitution model	Substitution rate (substitutions / site / year) (95% HPD)	TMRCA (year) (95% HPD)	d_N/d_S (95% CI)	Relative rates by codon position (1^{st}, 2^{nd}, 3^{rd})	Conservative amino acid substitution	Substitution of basic / acidic amino acids
Seg-1 (VP1)	3915	TN93 + G	2.17E-04 [1.28 - 3.11E-4]	443.964 [186.9 - 815.6]	0.047 (0.034 – 0.095)	0.27; 0.12; 2.61	0.65	0.07
Seg-2 (VP2)	3180	GTR + G + I	2.08E-04 [7.3E-5 - 3.36E-4]	6117.1 (542.7-13481)	0.262 (0.255 – 0.265)	0.52; 0.89; 1.59	0.33 – 0.59	0.02 – 0.16
Seg-3 (VP3)	2715	GTR + G + I	2.25E-04 [1.62 - 2.89E-4]	165.45 [103 - 263.6]	0.017 (0.005 – 0.029)	0.26; 3 $x10^{-2}$; 2.7	0.74	-
Seg-4 (VP4)	1926	TN93 + G	2.55E-04 [1.7 - 2.38E-4]	195.26 [113.1 - 319.8]	0.11 (0.097 – 0.125)	0.49; 0.28; 2.28	0.57	0.01
Seg-5 (NS1)	1644	HKY + G	2.28E-04 [1.7 - 2.85E-4]	156.275 [92.8 - 254.9]	0.084 (0.064 – 0.099)	0.37; 0.2; 2.4	0.71	-
Seg-6 (VP5)	1515	TN93 + G + I	2.36E-04 [9.3E-5 - 3.85E-4]	4182.522 [1273 - 8626.6]	0.067 (0.057 – 0.076)	0.27; 9 $x10^{-2}$; 2.6	0.56	0.08
Seg-7 (VP7)	1047	HKY + G	2.98E-04 [1.79 - 4.17E-4]	442.078 [186.8 - 830.1]	0.021 (0.005 – 0.047)	0.18; 6.9 x 10^{-2}; 2.8	0.76	-
Seg-8 (NS2)	1095	GTR + G + I	2.33E-04 [1.65 - 3.08E-4]	169.227 [102.5 - 272.9]	0.067 (0.038 - 0.095)	0.34; 0.15; 2.5	0.67	-
Seg-9 (VP6 & NS4)	1125	TN93 + G	3.54E-04 [1.88 - 5.19E-4]	850.8 [315.7 - 1604.4]	0.36 (0.34 – 0.38)	0.96; 0.49; 1.54	0.23 - 0.31	0.05 - 0.07
Seg-10 (NS3/NS3A)	654	HKY + G	2.17E-04 [9.43E-5 - 3.55E-4]	2737.016 [835.1 - 5541.6]	0.173 (0.156 – 0.191)	0.6; 0.29; 2.13	0.42	0.06

2.5 DISCUSSION

Virus isolation from samples submitted to the OIE reference centre over the last 60 years in addition to sequence data obtained from NCBI have enabled the investigation into AHSV evolutionary dynamics. Although AHSV shares many similarities with BTV, the total genome size of AHSV was slightly bigger totalling ~19.5 kbp for each genome, while the bluetongue virus genome size is around 19.2kbp (Attoui *et al*., 2012). The deletion observed in Seg-7 (AHSV-4_19_97, AHSV-4_91_00 and AHSV-4_31_00) and Seg-8 (AHSV-4_19_97 and AHSV-4_91_00) in the 5'-UTR region seems to be associated with the highly diverse serotype 4. The impact of those deletions on virus packaging, assembly and growth is unknown, but necessitates additional investigation based on the importance of the untranslated regions in viral packaging (Sung and Roy, 2014). The changes in Seg-8 may affect the NS2 phosphorylation, which will then have consequences in the development of VIB's (Modrof *et al*., 2005). Non-structural protein 2 binds ssRNA that serves as mRNA, which is required for the production of viral proteins (Matsuo *et al*., 2010).

Comparisons of percentage sequence identity and overall mean distance of both nucleotide and amino acid sequences supports previous observations that Seg-2 (VP2) is the most variable (Table 2, figure 4.3 and Appendix 2 figure 9) followed by Seg-10 encoding NS3/NS3A and Seg-6 that encodes VP5 (Martin *et al*., 1998). The nucleotide percentage identity of Seg-2 across different serotypes range between 54.74 – 100%, while the percentage identity was low, ranging between 49.06 – 100 %. However, the comparison of intra-serotypes has shown that both the nucleotide and protein sequences were more highly conserved. VP2 is highly variable protein and act as a major serotype-specific antigen (Mertens *et al.,* 1987; Huismans and Van Dijk, 1990). The study also observed a lower percentage sequence identity on the NS3/NS3A protein sequences across the 87 AHSV isolates (63.18 – 100). The findings are in agreement with the study by van Niekerk and colleagues (2001b), who observed NS3 variation of 36.3% across AHSV serotypes. Three different phylogenetic clades associated with S10 of the AHSV have been Identified namely α, β and γ (Sailleau *et al*., 1997; Martin *et al*.1998; van Niekerk *et al*., 2001b). This study did not investigate in detail how these three clades were distributed across

nine distinct serotypes and how they affected the nucleotide as well as protein variability. However, these clades seem to be responsible for high variability of the NS3/NS3A protein sequence, which was slightly higher compared to its coding nucleotides (Seg-10). To further confirm the presence of the three distinct clades, the ML phylogenetic tree had three main branches.

Even though the percentage sequence identity of AHSV on nucleotide level indicated substantial genetic variation, for example, Seg-1 and Seg-7 at 88 and 87%, respectively, the predicted proteins were conserved at >97% identity. High percentage (>97% predicted protein) sequence identity of Seg-1, Seg-3, Seg-4, Seg-5, Seg-7, Seg-8 as well as Seg-2 (intra serotype), reflects the strong selective constraint imposed on AHSV and other orbiviruses by the necessity to infect and replicate in both the vertebrate host and insect vector species (de Mattos *et al*., 1996; Novella *et al.,* 1999; Anbalagan *et a*l., 2014). In addition, it was common to isolate vaccine derived viruses (ALV) from samples submitted for routine diagnostic assays. Vaccine derived viruses were identified based on comparisons of Seg-2 with published sequence data of the OBP produced ALVs (Guthrie *et al.,* 2015). Viruses were identified from serotype 1 (AHSV-1_22_01 and AHSV-1_107_09), serotype 3 (AHSV-3_DG 25324_14, AHSV-3_DG 25327_14 and AHSV-3_DG 25423_14) and serotype 4 (AHSV-4_90_96, AHSV-4_64_99 and AHSV-4_97_99) with >99% sequence identity to the polyvalent AHS attenuated live virus (ALV) vaccine produced by Onderstepoort Biological Products (OBP) Ltd (Guthrie *et al*., 2015).

The Bayesian coalescent analysis estimated the individual genomic segments of AHSV to evolve at between 1.5 - 6.4 $x10^{-4}$ substitution /site /year. The substitution rate is similar to bluetongue virus (0.52 – 6.9 $x10^{-4}$ substitutions / site / year) and comparable to other vector-borne ssRNA viruses (Iwata *et al*. 1992; Carpi *et al*., 2010; Boyle *et al*., 2014). The estimates of mean substitution rates for segments-2, -6, -7, and -9 were the highest of the AHSV genome segments, between 1.75 $x10^{-4}$ and 4.6 x 10^{-4} substitutions/ site/ year. In comparison with other viruses within the *Reoviridae* family, the AHSV had low substitution rates to human rotaviruses and reoviruses, estimated at 0.5 $x10^{-3}$ and 0.3 x 10^{-3} substitutions/site/year, respectively (Jenkins *et al*., 2002). Both the rotaviruses and reoviruses do not require insect vectors for their transmission from one vertebrate host to another. The substitution

rates observed for AHSV were lower compared to the rates in human immunodeficiency virus (2.5 $x10^{-3}$ substitutions / site / year), which belongs to the *Retroviridae* family.

The RNA viruses that do not require insect vector for their transmissions have fewer evolutionary constraints and therefore have high mutation rates. The Eastern equine encephalomyelitis virus (EEEV) is a togavirus and requires an insect for transmission to vertebrate hosts (Suzuki *et al*., 2000). The EEEV substitution rate is <1.4 $x10^{-4}$ substitutions/ site/ year (Suzuki *et al*., 2000), which is comparable with AHSV substitution rates. However, equine infectious anaemia virus, which belongs to the *Retroviridae* family, has a high substitution rate of 0.2- 2 $x10^{-1}$ substitutions/ site/ year (Suzuki *et al*., 2000). Similar to BTV and EHDV, AHSV Segment 9, which encodes a viral helicase (VP6), has evolved two different homologues known as type I and type II (Zwart *et al*., 2015). Unlike the clear separation based on geographic origins displayed by VP6 of EHDV (Anthony *et al*. 2010), no geographic, serotype-specific or chronologic link could be established between the two types of AHSV Seg-9 described here. In contrast to the observation of Seg-10 in BTV (low ratio of d_N/d_S at 0.060 and the rapid substitution rate of 6.94 x 10^{-4}) by Carpi *et al*., 2010, Seg-10 encoding NS3/NS3A of AHSV had the high ratio of d_N/d_S (0.173) and a low substitution rates (1.7 x 10^{-4}). Our observation might be influenced by higher relative rate of substitution for position one and two in a codon (0.6; 0.29; 2.11), implying that the number of substitutions wasn't the major influence on Seg-10 evolution, but rather the position of the substitution relating to the encoded protein. Similar to BTV, all ten AHSV segments were predicted to evolve under strong purifying selection pressure with selected sites under positive selection (Boyles *et al*., 2014). The high percentage sequence identity between some AHSV genome segments reflect the strong selective constraints imposed on arboviruses by the necessity to infect and replicate in both the host and vector species. Due to these constraints, the AHSV's overall substitution rate favours synonymous substitution and the virus undergoes negative selection. Therefore, as expected and as observed, AHSV evolves very slowly compared to RNA viruses that do not require an insect vector for its transmission.

CHAPTER 3 INVESTIGATING RECOMBINATION IN AHSV GENOMES

3.1 BACKGROUND

African horse sickness virus, a member of *Orbivirus* genus in the family *Reoviridae* is a double stranded segmented RNA virus (Attoui *et al*., 2012) that causes infectious but non contagious disease in equids. As a double stranded RNA virus, AHSV can undergo homologous or non-homologous genetic recombination (Simon-Loriere and Holmes, 2011). Although there are no reports about the influence of genetic recombination in an AHSV virus yet, intragenic recombination has been reported in BTV, the orbivirus prototype. Hybrid genes generated through intragenic recombination have been identified in BTV (He *et al*., 2010). Viruses containing recombinant genome segments became the prevailing strains, demonstrating the influence of these genes on the genetic diversity and epidemiology of BTV (He *et al*., 2010). Except for the study by He *et al* (2010), there are no other known studies that have reported on the influence of genetic recombination to the evolution of orbiviruses among the field isolates.

The use of standard phylogenetic methods such as maximum parsimony, minimum evolution and maximum likelihood are unable to detect the occurrence of genetic recombination. These standard phylogenetic methods, grant that a single evolutionary history underlies the data of samples under study, while ignoring the occurrence of genetic evolution due to cross over events (Posada *et al*., 2002). Genomic recombination events cannot be confirmed or described by a single phylogenetic tree, which is a problem when using standard phylogenetic approaches (Martin *et al*, 2005). The RDP software package can be used for characterizing genomic recombination events in the sequence alignment and is capable of differentiating recombination from reassortment events (Martin *et al*., 2005; Heath *et al*., 2006). The ability of the RDP v4 software program to simultaneously use different methods to detect recombination gives it an upper edge (Martin *et al.*, 2010). The RDP software permits the faster, automated analysis of large sequence alignments of up to 300 containing up to 13 000 sites (Martin *et al*, 2005). Included in the RDP v4 software package are methods such as the original RDP method, bootscan/rescan recombination test (Martin *et al*. 2005), CHIMAERA (Posada and Crandall, 2001), GENECONV (Padidam *et al*., 1999), MAXCHI (Maynard Smith,

1992), and the SISCAN method (Gibbs *et al*. 2000). The original RDP method uses a three-step procedure to screen multiple sequence alignments for the evidence of recombination (Martin *et al*., 2010). The bootscan/rescan recombination test can be used to screen nucleotide sequence alignments for evidence of genomic recombination events without identification of non-recombinant reference sequences (Martin *et al.,* 2005). The MAXCHI (The maximum chi-squared) test is used to identify recombination events between two sequences by comparing the distribution of polymorphic sites along the sequences with those that occur by chance (Maynard Smith, 1992). The GENECONV is a statistical method used for detecting gene conversion events between pairs in the sequence alignment (Padidam *et al.*, 1999). The GENECONV detects regions in the sequence alignment where sequence pairs are similar and could have arisen as a result of recombination (Padidam *et al*., 1999). The CHIMAERA is a method modified from the MAXCHI method, which provides information on the positions of breakpoints, but is unable to offer information on the extent scale of recombination regions (Posada and Crandall, 2001). The SISCAN (Sister scanning) look for recombination signals in the nucleotide sequence data. The SISCAN is done by examining each possible triplet in a sequence alignment for the evidence of genomic recombination (Gibbs *et al*. 2000). Different types of phylogenetic trees can be created using the RDP software, and these include UPGMA, neighbor joining (NJ), Fast neighbour-joining (FatNJ) or approximate least squares (LS), maximum likelihood (ML) or Bayesian trees (Martin *et al*., 2010). These different methods within RDP v4 can be set to analyse and interpret the genomic sequences as being linear or circular, and the program can analyse linear sequences as if they are circular (Martin *et al*., 2010). Each recombinant sequence is split into two pieces and dataset repeatedly scanned until there are no further recombination signals (Martin *et al*., 2015) Another program that can be used in analysing recombination is SimPlot. The SimPlot was created by Lole and co-workers (1999) to identify sequence similarities by calculating and plotting the percentage identity of the query sequence to reference sequences.

In order to determine the influence of homologous recombination in AHSV evolution, each individual segment of 101 complete AHSV genome sequences was analysed using Recombination Detection Program (RDP) (Heath *et al*., 2006).

3.2 AIM OF STUDY

To investigate intragenic recombination in AHSV from the sequences of the viruses isolated between 1933 and 2014 as well as the Live Attenuated Viruses (ALV) using the RDP 4 software program.

3.3 MATERIAL AND METHODS

The following section describes the approach used to investigate the intragenic recombination events predicted after analysing 101 AHSV genomic sequences that included the sequences removed during the vertical analysis.

3.3.1 Datasets

Complete genome sequences of 101 AHSV isolates were used to analyse genomic recombination. The majority of these viruses were isolated between 1933 and 2014 from Sub-Saharan Africa and were stored at ARC-Onderstepoort Veterinary Institute, South Africa (KP939368 - KP940236 and KP009621 - KP009790). Live attenuated viruses (ALV's) from the commercial OBR vaccine (KT030330 – KT030359 and KT715601 – KT715640) were included in genomic recombination analysis. See Table 1, Appendix 1 for more details.

3.3.2 Genomic recombination analysis

Sequences from each individual genome segment were aligned using Clustal W and subsequently used for genomic recombination analysis. In order to detect evidence of intragenic recombination, the individual segment alignments were analysed using different methods described in the Recombination Detection Program (RDP) v4.39 package (Heath *et al*., 2006). The RDP v4.39 was set to default, and a large window sizes between 100 and 200 nucleotides were used. The methods included a bootscan/rescan recombination test (Martin *et al*. 2005), MAXCHI (Maynard Smith, 1992), GENECONV (Padidam *et al*., 1999), CHIMAERA (Posada and Crandall, 2001) and the SISCAN method (Gibbs *et al*. 2000). These different recombination tests were performed under the assumption that the data sequences could be used in a circular form (Different methods within the RDP v4 program can be set such that

it interprets the genome sequences as being linear or circular) and each method had a Bonferroni correction p-Value of 0.05. Bonferroni correction adjusts the P-values when there are several dependent or independent statistical tests that are carried out concurrently on the same data set. Potential recombinants were further described using SimPlot.

3.4 RESULTS

The following chapter investigated the intragenic recombination of AHSV from the complete genome sequence data of viruses that were isolated from 1933 to 2014 using different programs within the RDP 4.39 software package (bootscan/rescan, MAXCHI, GENECONV, CHIMAERA and SISCAN). Individual alignments of each of the ten AHSV genome segments were used to generate maximum likelihood phylogenetic trees (Chapter 4). Incongruent tree topologies were detected between the ten genome segments, suggesting that recombination might be present in AHSV evolution. Four genome segments displayed mosaic regions originating from different viral groups, providing evidence of intragenic recombination in AHSV. Potential intragenic recombination events were predicted in seg-1, seg-6, seg-7 and seg-10. These included both single and double cross-over events and will be discussed individually.

3.4.1 Segment -1 (VP1).

Segment-1 is the largest of the AHSV genome segments consisting of 3965 base pairs. It is known to codes for VP1, one of the enzymatic minor core proteins responsible for the RNA-dependent-RNA-Polymerase activity. A phylogenetic tree of all 101 AHSV sequences, divides the sequences into two major groups. The small group consists of 17 viruses and includes the reference strains of AHSV-1, 2, 3, 4 and 7, their respective attenuated strains (ALV's) as well as field isolates resembling reference strains (Small Group in Figure 3.1.A). The remainder of the segment 1 sequences all clustered into the larger major group comprising 15 viruses (Large Group in Figure 3.1.A. showing selected isolates). Using the RDP v4.39 package, Seg-1 of AHSV presented a double cross-over event (involving positions: 1524 and 1864) in the isolate AHSV-3_2_89. The Maximum-likelihood phylogenetic tree

analysis inferred that virus AHSV-3_2_89 grouped within the larger group, while the minor parental sequences were from the small group (Figure 3.1A).

By utilising the RDP package, the recombination event had statistical support with average P-values for RDP = 6.47 $x10^{-20}$, GENECONV = 3.37 $x10^{-15}$, BootScan = 2.903 $x10^{-21}$, MaxChi = 1.45 $x10^{-8}$, SiScan = 1.83 $x10^{-8}$ and 3Seq = 5.64 $x10^{-11}$ (Figure 3.1B). A similarity plot (SimPlot) was constructed by using two reference strains, AHSV-7_67_99 as the major parent and AHSV-3_13_63 as the minor parent. Since AHSV-3_13_63 was originally used in the generation of the live attenuated vaccine of serotype 3 (Guthrie *et al*., 2015), the sequences of VP1 from LAV-AHSV-1 and LAV-AHSV-3 were included in the similarity plot The AHSV-3_2_89 isolate shared a high percentage sequence identity to AHSV-7_67_99 before (98.9%) and after (99.8%) the two predicted breakpoints than within the predicted breakpoints (89.5%). In contrast, AHSV-3_13_63 had a higher percentage sequence identity between the breakpoints (97%), than before (89.6%) and after (89.2%) to AHSV-3_2_98 (Figure 3.1C).

Distance matrix phylogenetic analysis of regions 1510 – 1850 produced incongruent phylogenetic trees. Analysis of the concatenated regions before and after the breakpoints (1850 – 1510), indicated that AHSV-3_2_89 will cluster in the same large group as the major parental sequence (Figure 3.1D). When the region between the breakpoints was analysed, AHSV-3_2_89 clustered with the minor parental sequences in the small group (Figure 3.1E). Recombination within region 1524 to 1864 of AHSV-3_2_89 with AHSV-3_13_63 resulted in 34 synonymous substitutions compared to AHSV-7_67_99.

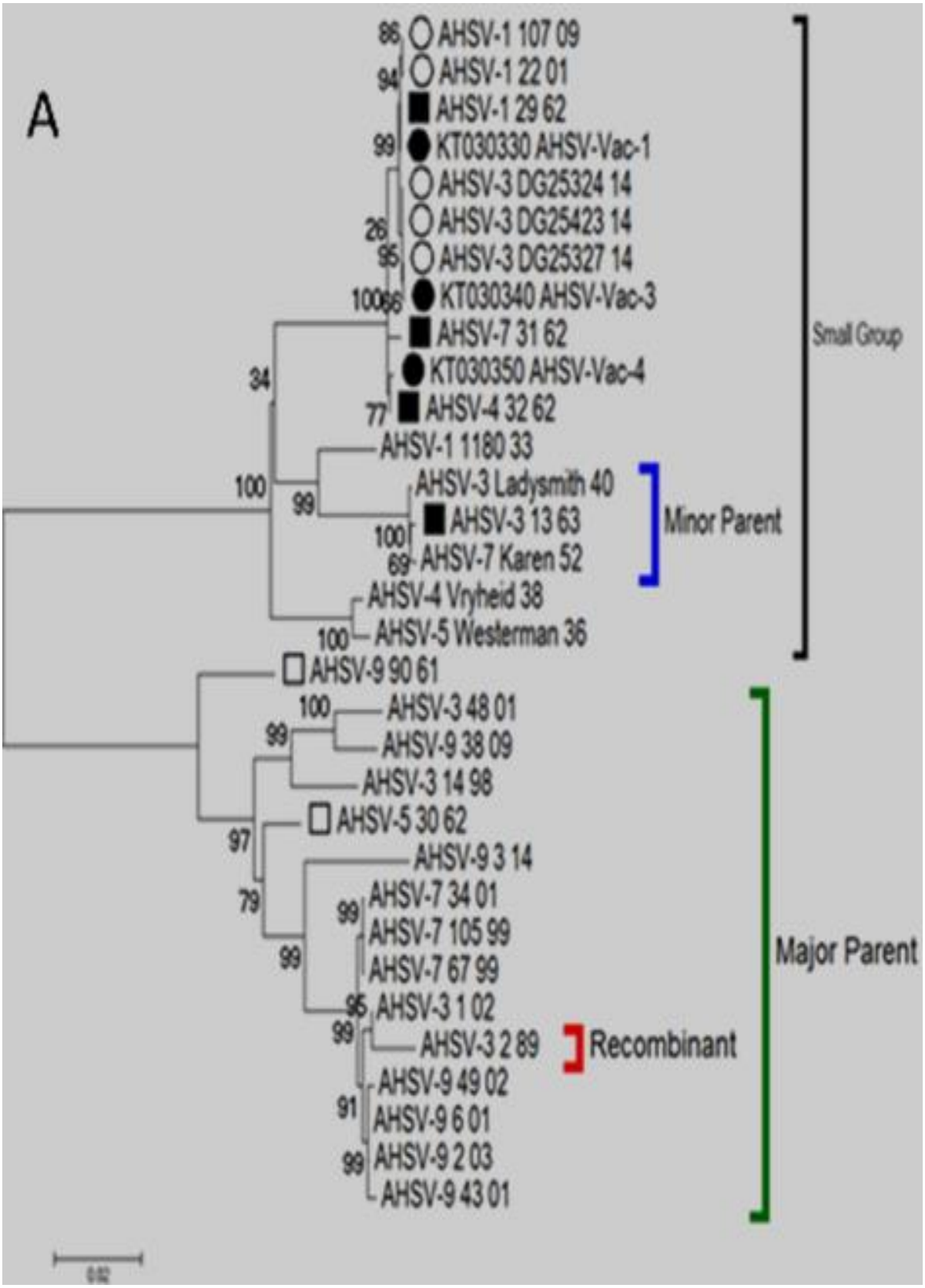

Figure 3.1 (A). Evidence of recombination in Segment-1 of AHSV-3_2_89. Reference isolates are indicated with a square, while attenuated sequences (●) as well as vaccine isolated from the field (○) are marked with a circle. ML phylogenetic tree showing the relationship of segment 1 in selected AHSV isolates.

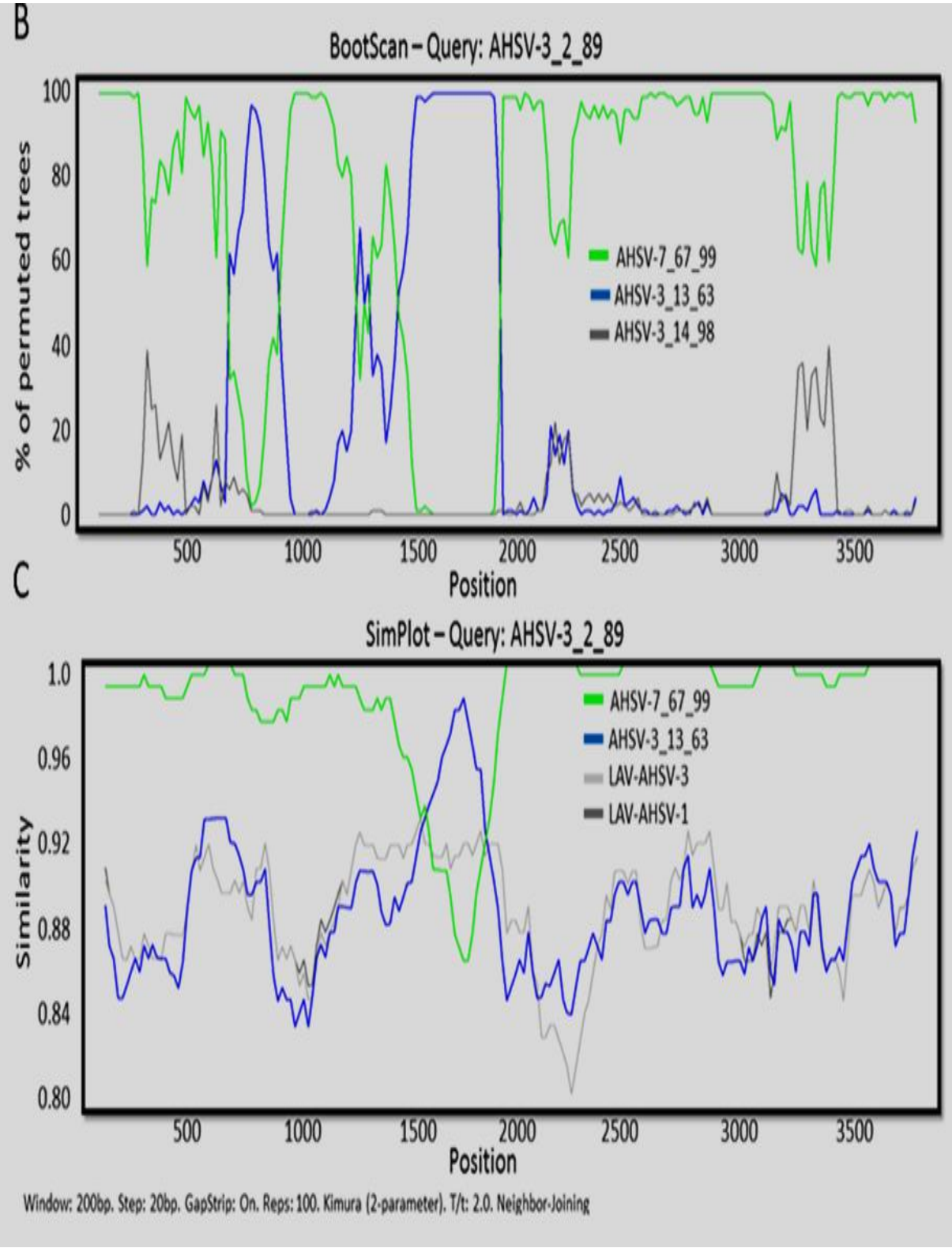

Figure 3.1 (B and C). Evidence of recombination in Segment-1 of AHSV-3_2_89. (B) BootScan results of AHSV-3_2_89 with its parents AHSV-7_67_99 and AHSV-3_13_62 and an outgroup AHSV-3_14_98. (C) Comparison of the sequence similarity between AHSV-3_2_89 and isolates AHSV-7_69_99, AHSV-3_13_62, LAV-AHSV-1 and LAV-AHSV-3 using SimPlot.

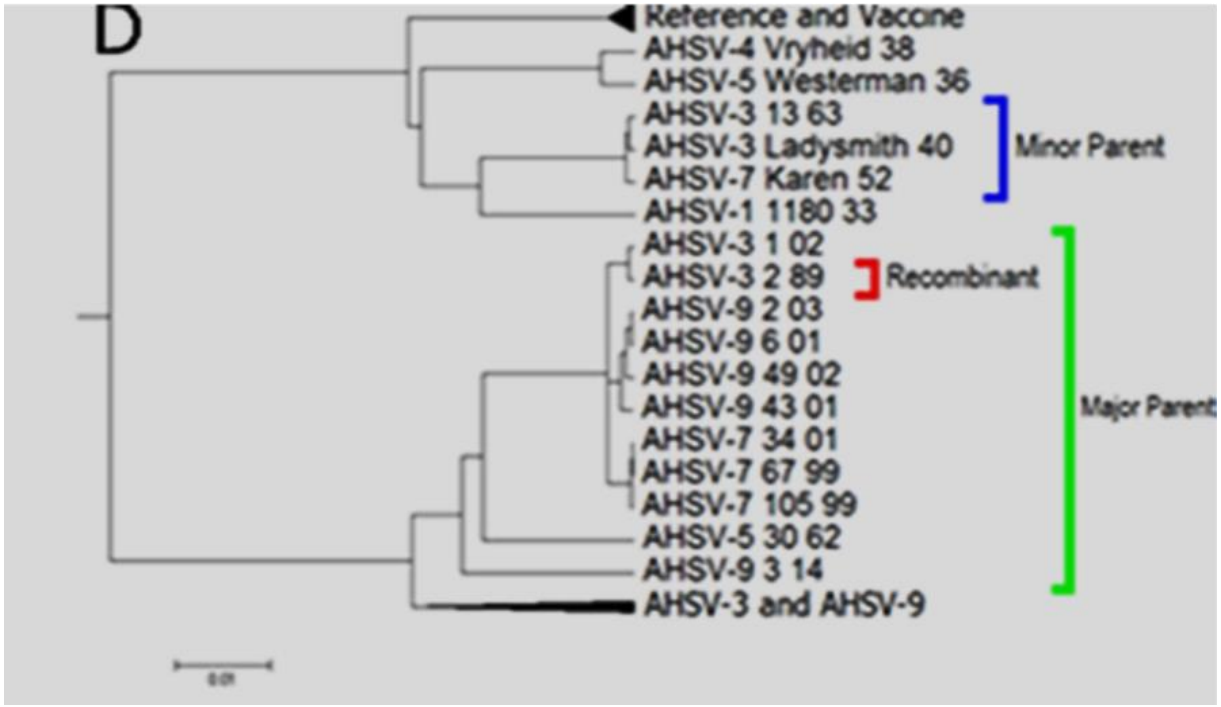

Figure 3.1 (D). Evidence of recombination in Segment-1 of AHSV-3_2_89. (D) Phylogenetic tree of the concatenated segment 1 sequence of region 1865 – 1523 of selected AHSV isolates.

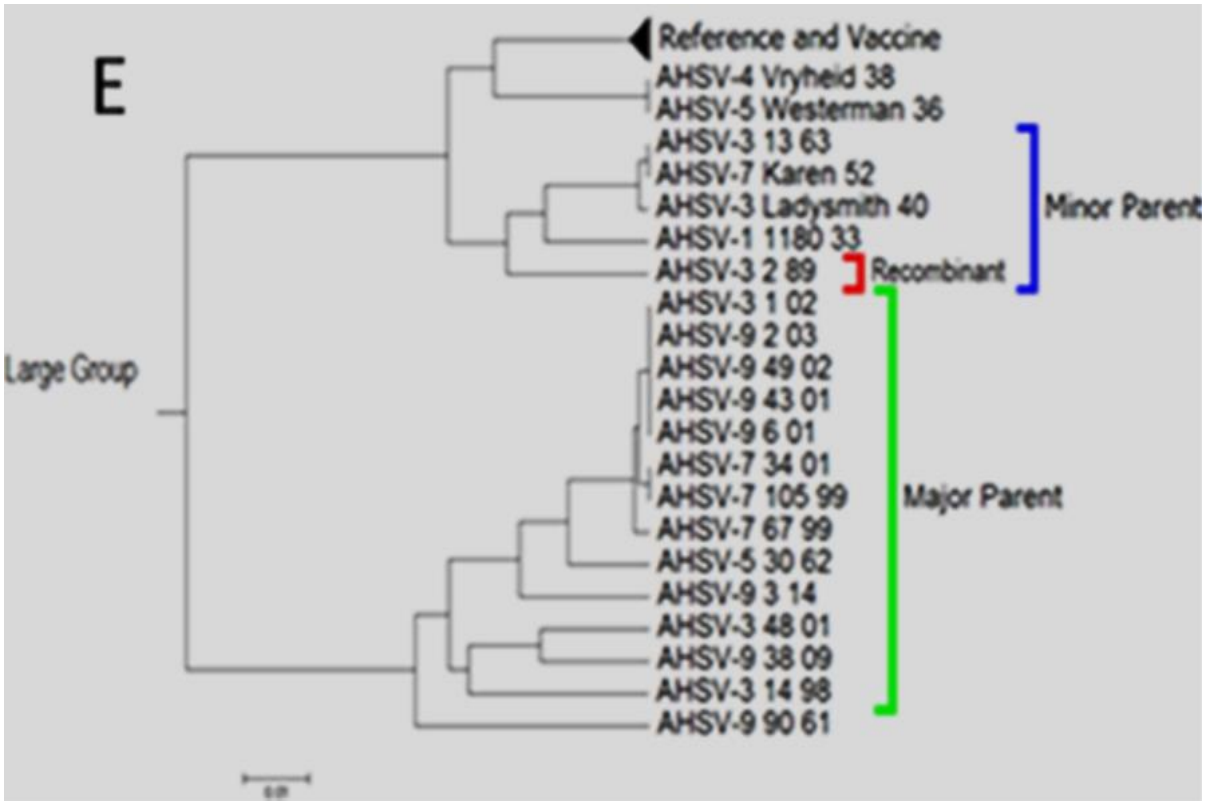

Figure 3.1 (E). Evidence of recombination in Segment-1 of AHSV-3_2_89. (E) Phylogenetic tree of region 1524 – 1864 from segment 1 of selected AHSV isolates**.**

3.4.2 Segment-6 (VP5).

Segment-6 encodes the outer capsid protein VP5. The latter associates with the major serotype determining outer capsid protein VP2. It is therefore not surprising that phylogenetic analysis of all the seg-6 sequences reveals that sequences of individual serotypes cluster together (Figure 3.2A). Four major groups were identified, which were subdivided into seven minor groups. The first major group comprises of two minor groups containing sequences of serotype 1 and serotype 2 respectively. The second major group consist of only serotype 4 sequences, while the third major group is again divided into two minor groups containing serotype 7 and another of serotype 8. The last major group consists of two closely related minor groups. The first minor group contains only serotype 5, while the second minor group consist of a combination from serotypes 3, 6 and 9, 7 and 3 (Figure 3.2.A).

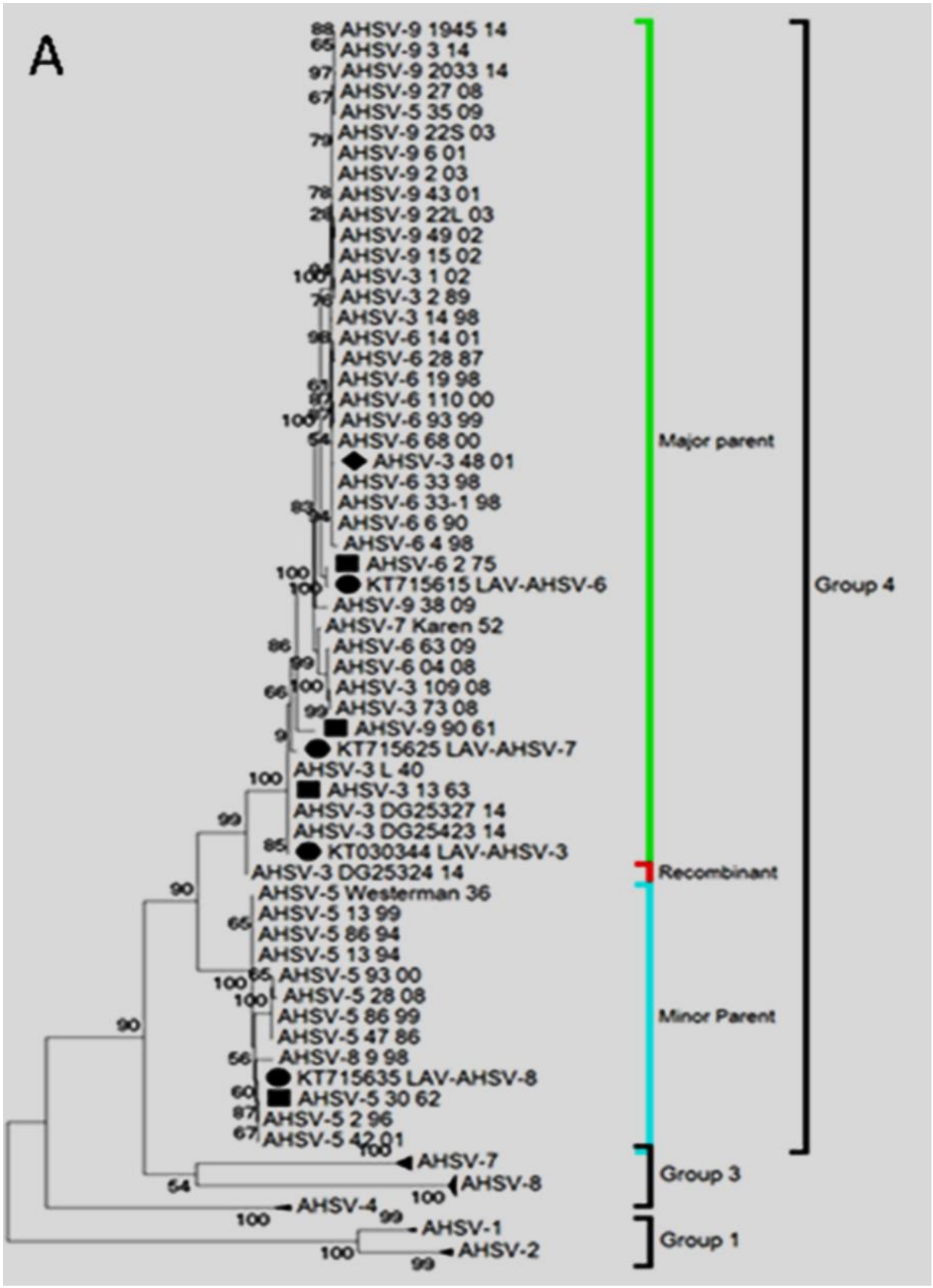

Figure 3.2 (A). Evidence of recombination in Segment-6 of AHSV-3_DG25321_14. Reference isolates are indicated with a square (▪), while LAV sequences are marked with a circle (●). **(**A) ML phylogenetic tree showing the relationship of segment 6 using selected AHSV isolates.

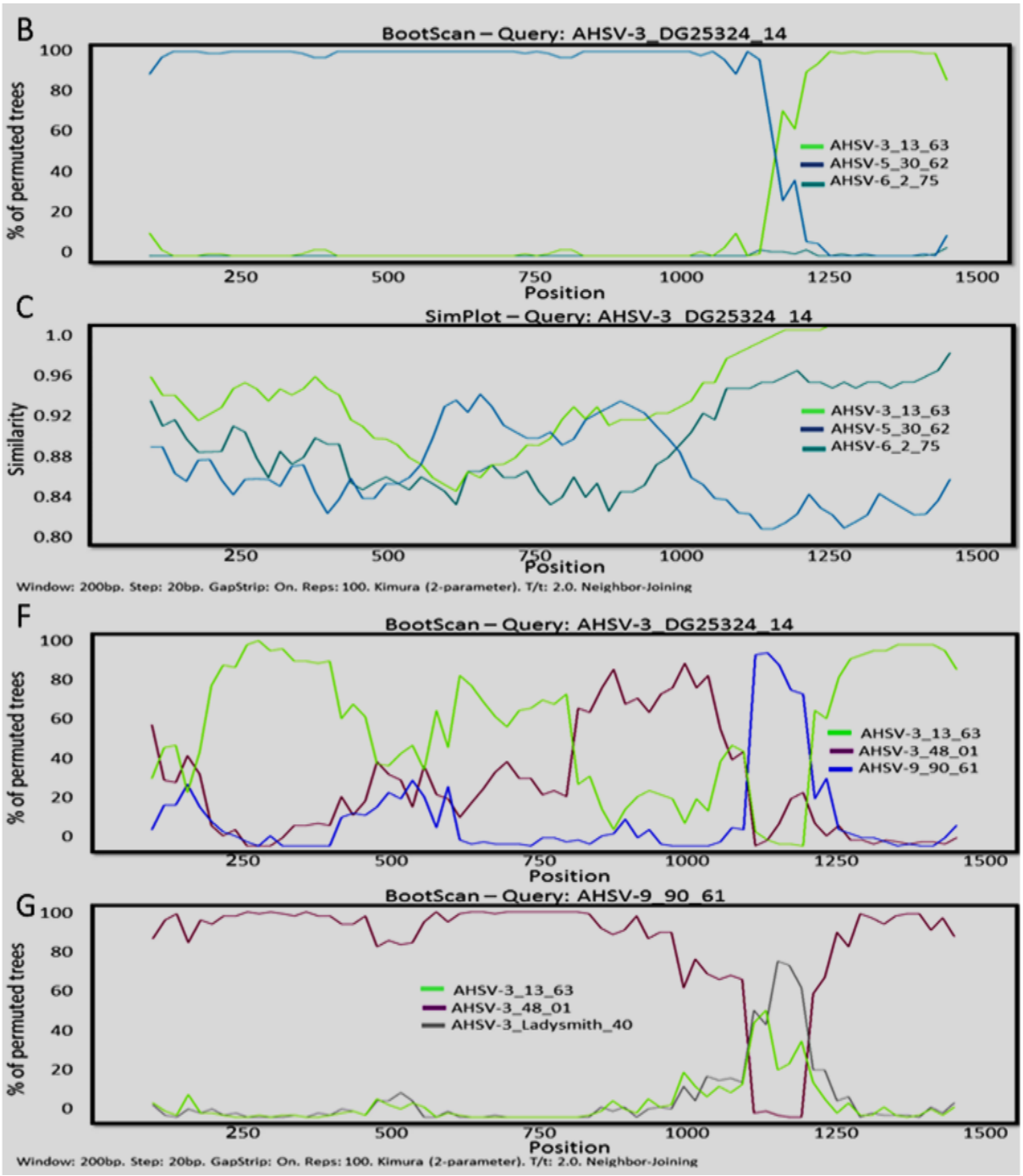

Figure 3.2 (B, C, F, and G). Evidence of recombination in Segment-6 of AHSV-3_DG25321_14. (B) BootScan results of AHSV-3_DG25321_14 with parental isolates AHSV-3_13_63 and AHSV-5_30_62, as well as an outgroup AHSV-6_2_75. (C) Comparison of the sequence similarity between AHSV-3_DG25324_14 and Reference strain AHSV-3_13_63, AHSV-5_30_62 and AHSV-6_2_75 using SimPlot. (F) BootScan results of possible recombination events within the serotype 3, 6 and 9 cluster, using AHSV-3_DG25321_14 as query sequence and reference isolates AHSV-3_13_63 and AHSV-9_90_61 as well as isolate AHSV-3_48_01. (F) Indicating recombination events within the serotype 3, 6 and 9 cluster. (G) BootScan analysis using reference AHSV-9_90_61 as query and isolates AHSV-3_13_63, AHSV-3_Ladysmith_40 and AHSV-4_48_01 for comparison.

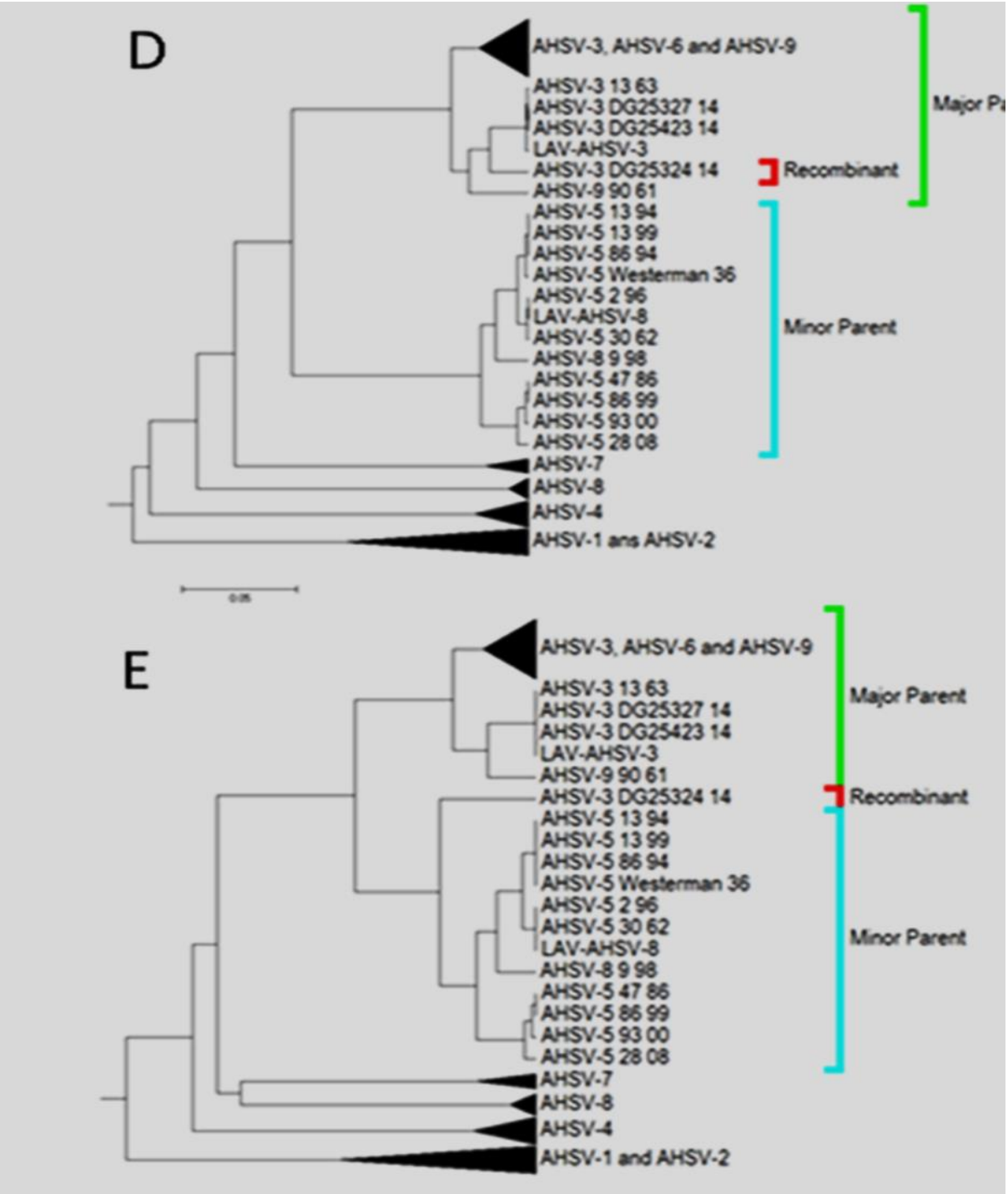

Figure 3.2 (D and E). Evidence of recombination in Segment-6 of AHSV-3_DG25321_14. (D) Phylogenetic tree of the concatenated segment 6 sequence of region 963 – 495 of selected AHSV isolates. (E) Phylogenetic tree of region 496 – 962 from segment 6 of selected AHSV isolates.

Two possible double-crossover events were identified relative to isolate AHSV-3_DG25324_14. The most notable double-crossovers were observed at positions 496 and 962. The AHSV-3_DG25324_14 was isolated in the Gauteng province of South Africa in 2014 and had a Segment 2 identical to AHSV-3 vaccine strain (LAV-

AHSV-3) (Appendix 2, Segment 6). The ML analysis of AHSV seg-6 sequences indicated that AHSV-3_DG25324_14 clustered within Major group 4, but on a separate branch from the minor groups representing serotype 5 and the combined minor group of serotypes 3, 6 and 9 (Figure 3.2A).

From the Simplot / bootscan analysis, the major parental sub-group was composed of sequences from AHSV serotype 3, 6 and 9, while the minor parental sub-group represents the minor group of serotype 5. The latter minor group has two reassortment events from serotype 8 (LAV-AHSV-8 and AHSV-8_9_98). Recombination predictions in RDP suggested that any of the AHSV-5 minor group sequences was the parent between base pairs 496 to 962, with reference strain AHSV-5_30_62 sharing 92.2% sequence identity within this region comparing to 87.9% and 86.1% for the flanking regions. The remainder of the sequence (963 – 1566) was similar to Reference strain AHSV-3_13_63 with percentage sequence identity of 98.5%, in (Figure 3.2B-C). Statistical support of the events between 496 and 962bp had a p-value using RDP = 6.47 x10-20, GENECONV = 3.37 x10-15, Maxchi of P = 3.85 $x10^{-05}$, Chimaera (P = 5.49 $x10^{-05}$) and SiSscan predicted p = 2.30 $x10^{-09}$ (Figure 3.2B). The UPGMA analysis of the concatenated region 963 - 495 had AHSV-3_DG25324_14 clustering in the same minor group as the other serotype 3 ALV/reference strains (Figure 3.2D), while analysis of the region 496 - 962 clustered AHSV-3_DG25324_14 with the sub-group of serotype 5 reference strain (Figure 3.2E). The second recombination event involving AHSV-3_DG25324_14 had less statistical support and could identify only one parent. No breakpoints were defined, but statistical support of MaxChi P = $3.18x10^{-1}$ and Siscan P= $1.4x10^{-4}$ were provided around positions 36 bp and 1047 bp. Both the Bootscan and RDP analysis on the AHSV-3_13_63, AHSV-3_48_01 and AHSV-9_90_61 in relation to AHSV-3_DG25324_14 confirm possible events in these areas, but suggested that AHSV-3_DG25324_14 wasn't the actual recombinant (Figure 3.2F). The analysis was repeated excluding AHSV-3_DG25324_14, but using reference strain AHSV-9_90_61 as query (Figure 3.2G). Based on ML phylogenetic characterization of AHSV segment 6, reference AHSV-9_90_61 grouped between reference/vaccine strains of serotype 3 and a cluster consisting of field isolates of serotype 3, 6 and 9 with sequence identity higher than 98%. Based on BootScan

data, it seems as if multiple recombination events occurred within this lineage, but due to the high sequence similarity it is difficult to predict individual events.

3.4.3 Segment-7 (VP7).

Segment-7 is relatively conserved, 1167 bp in size and codes for the 349 amino acid major core protein VP7 (Basak *et al*., 1996). Recombination analysis of 101 AHSV genomes predicted a single cross-over event at position 642 in isolate AHSV-5_86_94. The only relevant information available for this isolate is that it was isolated from a lung sample in 1994. Maximum-likelihood phylogenetic analysis indicated that AHSV-5_86_94 places separate from two major groups (Figure 3.3A). The recombination event had statistical support with p-Values of GENECONV P = 0.011455, Maxchi P = 4.25 $x10^{-06}$ and Chimaera predicted P = 1.71 $x10^{-06}$. BootScan analyses were performed with Reference strain AHSV-5_30_62 representing the major parent group and AHSV-5_86_99 from the minor parental group (Figure 3.3B). Comparisons of sequence similarity between AHSV-5_86_94 with the two representatives are indicated in Figure 3.3C. A high degree of sequence similarity among all three isolates was observed before position 642, while after the recombination event isolate AHSV-5_86_99 had a significantly higher sequence identity to AHSV-5_86_94 than the reference strain AHSV-5_30_62. Using UPGMA analysis of region 1-625 and region 626 -1167 respectively, at the N-terminal region sample AHSV-5_86_94 cluster with reference isolates of serotype 2 and 5 (Figure 3.3D), while the C-terminal region cluster in the other major group consisting of the remaining serotype 5 sequences (Figure 3.3E).

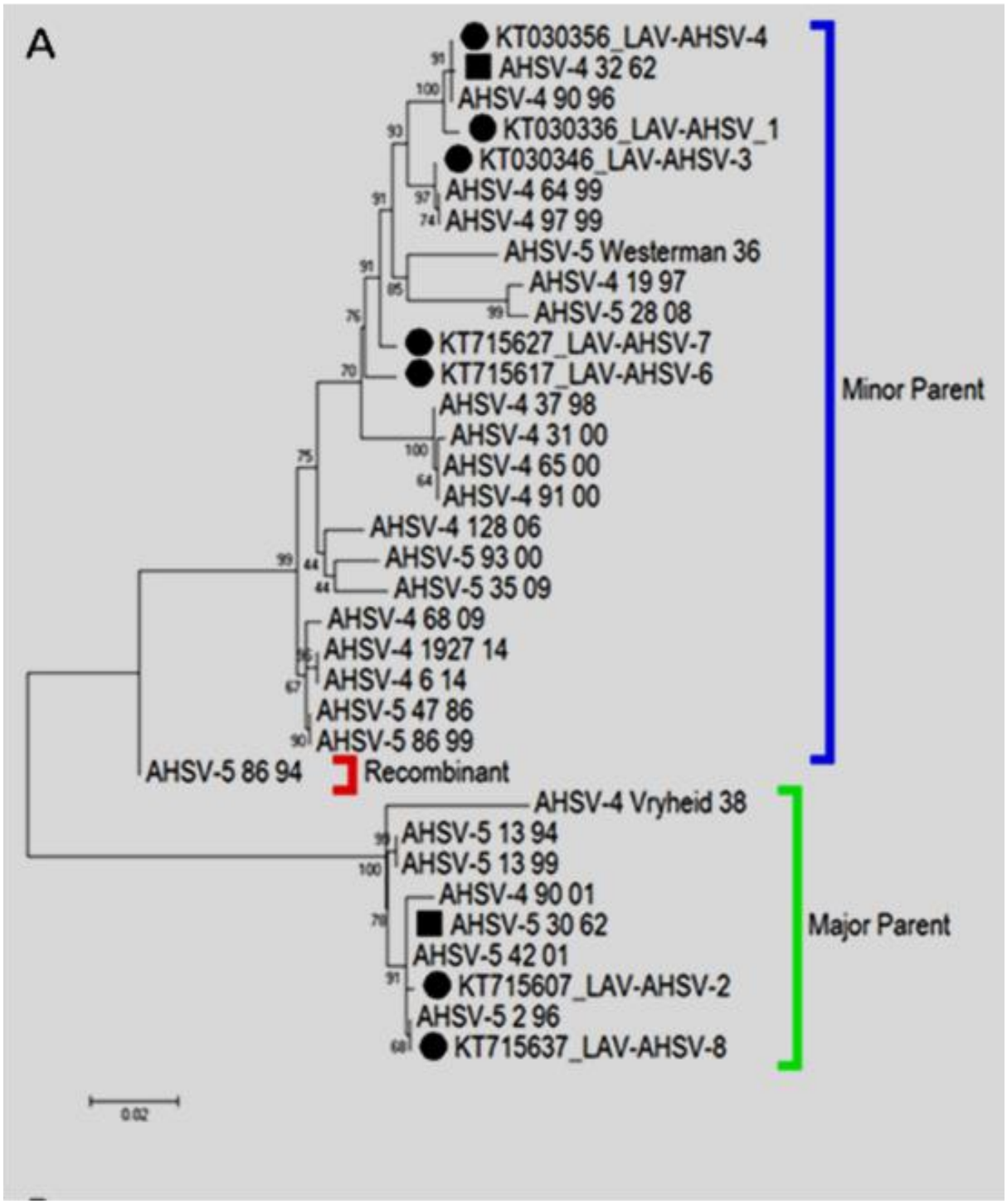

Figure 3.3 (A). Evidence of recombination in Segment-7 of AHSV-5_86_94. Reference isolates are indicated with a square (▪), while LAV sequences are marked with a circle (●). ML phylogenetic tree showing the relationship of segment 7 using AHSV isolates of serotype 4 and 5 as well as all the Live Attenuated vaccine strains. Reference isolates are indicated with a square, while LAV sequences are displayed with a circle.

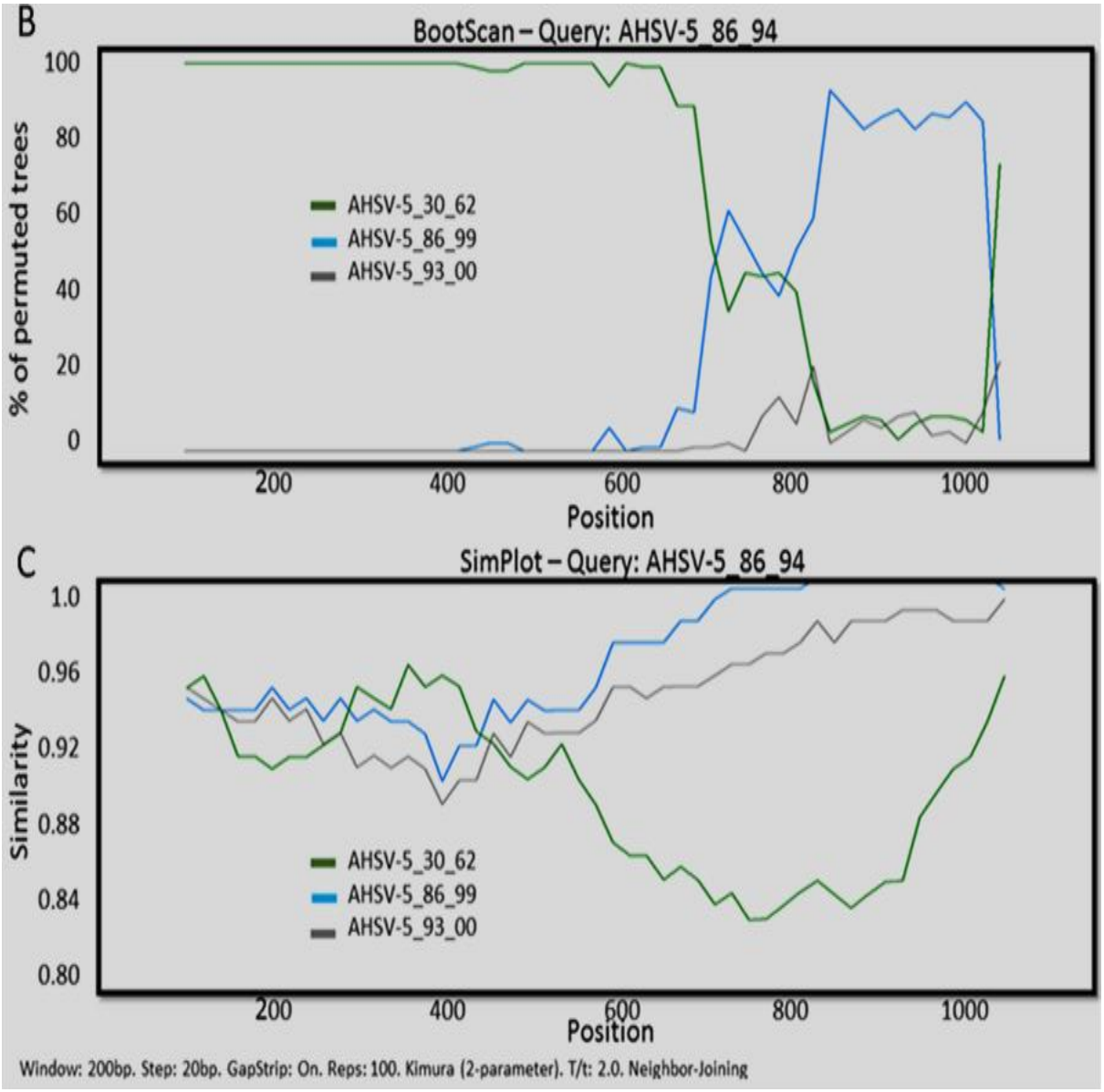

Figure 3.3. Evidence of recombination in Segment-7 of AHSV-5_86_94. (B and C) BootScan results of AHSV-5_86_94 with possible parents AHSV-5_30_62 and AHSV-5_86_99 and an outgroup AHSV-5_93_00. (C) Comparison of the sequence similarity between AHSV-5_86_94 and isolates AHSV-5_30_62, AHSV-5_86_99 and AHSV-5_93_00 using SimPlot.

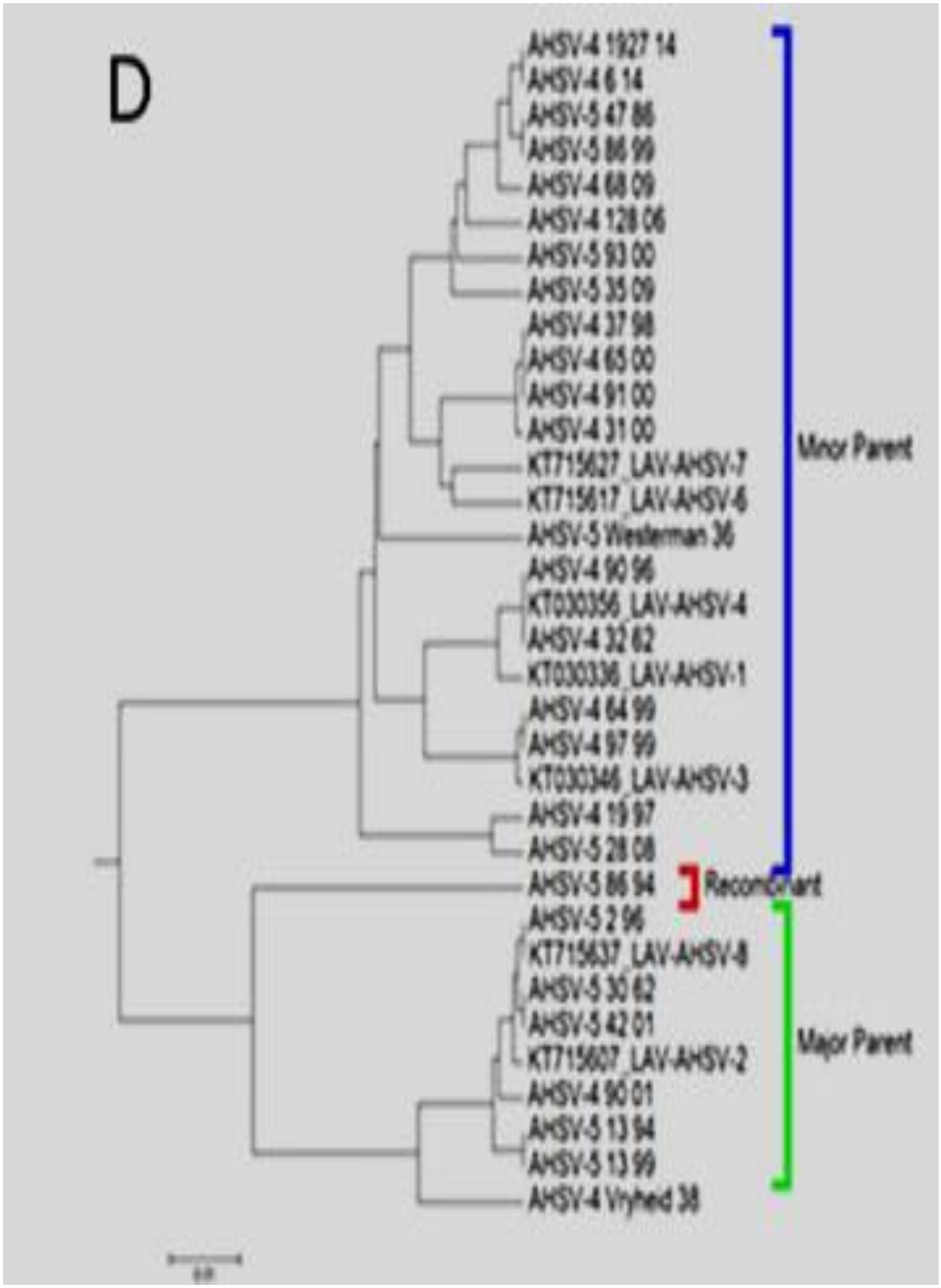

Figure 3.3 (D). Evidence of recombination in Segment-7 of AHSV-5_86_94. (D) Phylogenetic tree of segment 7 sequence of region 1 – 642 of selected AHSV isolates.

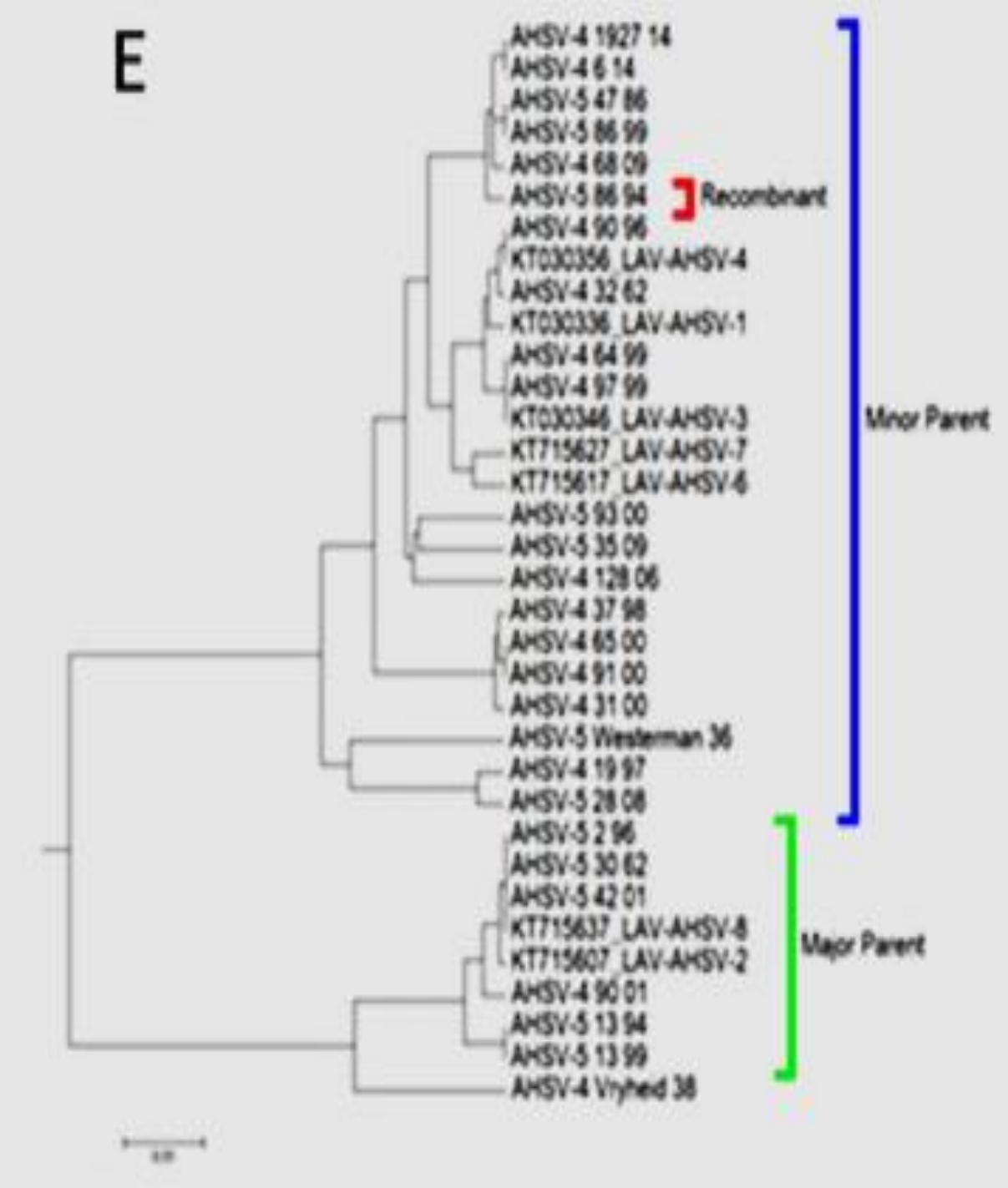

Figure 3.3 (E). Evidence of recombination in Segment-7 of AHSV-5_86_94. (E) Phylogenetic tree of region 642 – 1105 of segment 7 from selected AHSV isolates.

3.4.4 Segment-10 (NS3).

Segment-10 is the smallest and second most variable segment, coding for two in-frame non-structural proteins involved in virus release, NS3 and NS3A (Sailleau *et al*., 1997). ML phylogenetic trees of segment 10 showed 4 major groups A, B, C and

D, while group B is sub-divided into 4 sub-groups. RDP analysis of segment 10 predicted a single cross-over event at position 192 of three isolates, reference strain AHSV-5_30_62 and field isolates AHSV-5_42_01 and AHSV-5_2_96. These three isolates and ref AHSV 9_90_61 formed a single sub-group B1 (Figure 3.4A). The parental sequences predicted by RDP belong to sub-groups B2 and B4. The recombination event had the statistical support of p-values using Siscan of P = $1.3x10^{-15}$ and 3Seq of P = $4.1x10^{-3}$. BootScan analysis using four reference strains of AHSV-5_30_62, AHSV-8_10_62, AHSV-4_32_62 and AHSV-6_2_75 predicted a recombination signal around positions 150 – 250, but it is not clear which of the sequences were the recombinant or parental isolates (Figure 3.4B). Each of these reference strains shares significant sequence similarity with the other isolates in their sub-groups (B1, B2 and B4), resulting in similar Bootscan results when other representatives from each sub-group were analysed (results not shown). Distance matrix analysis of segment 10 sequences belonging to group B, indicated that sub-group B1, B2 and B3 form part of a separate lineage to sub-group B4 when only region 193-757 are considered (Figure 3.4D). In contrast, when region 1 – 192 was analysed sub-group B3, B1 and B4 formed a separate lineage to sub-group B2 (Figure 3.4E).

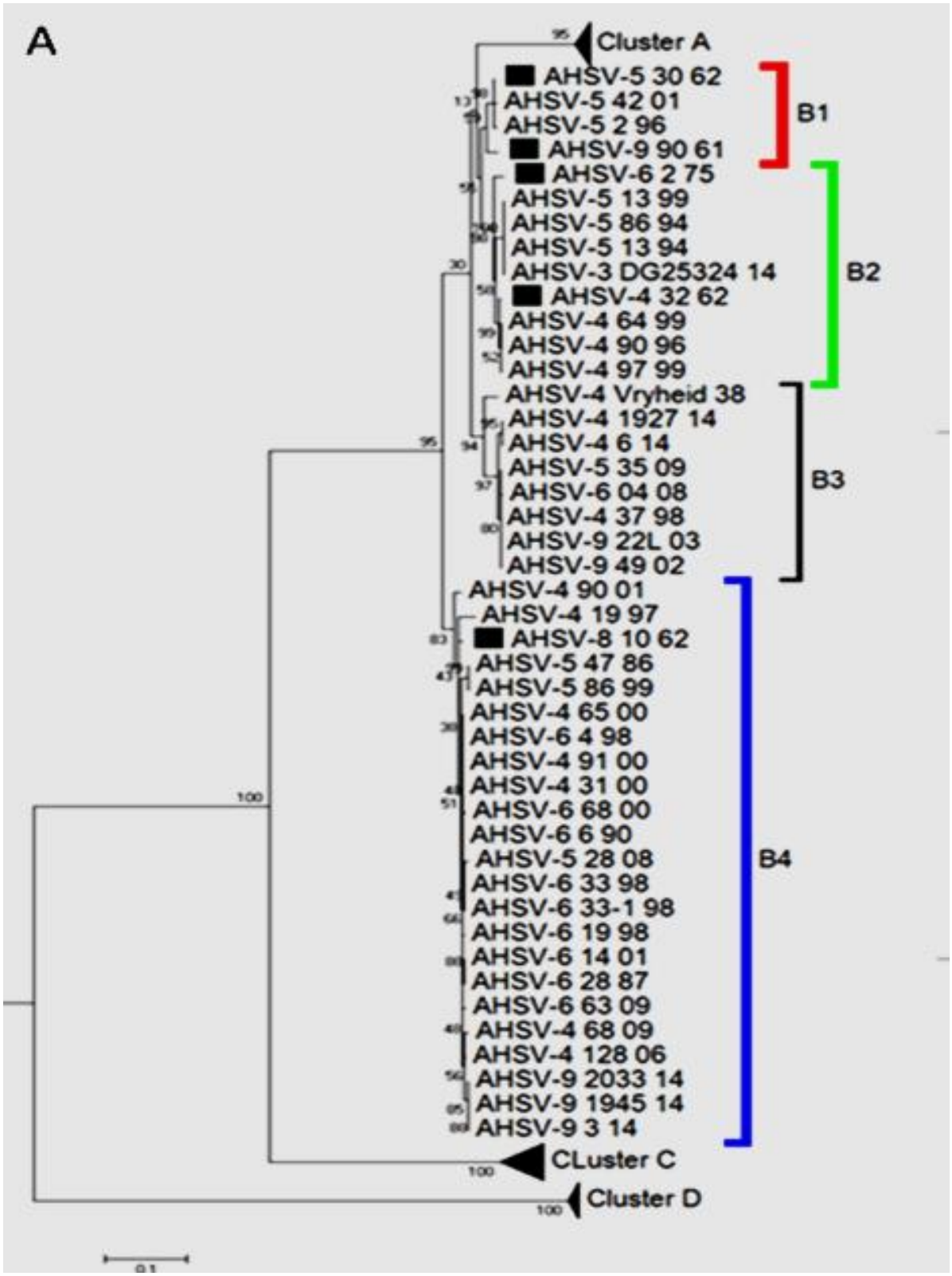

Figure 3.4 (A). Predicted recombination in Segment-10 of AHSV-5_30_62. Reference isolates are indicated with a square. ML phylogenetic tree showing the relationship of segment 10 using 101 AHSV isolates.

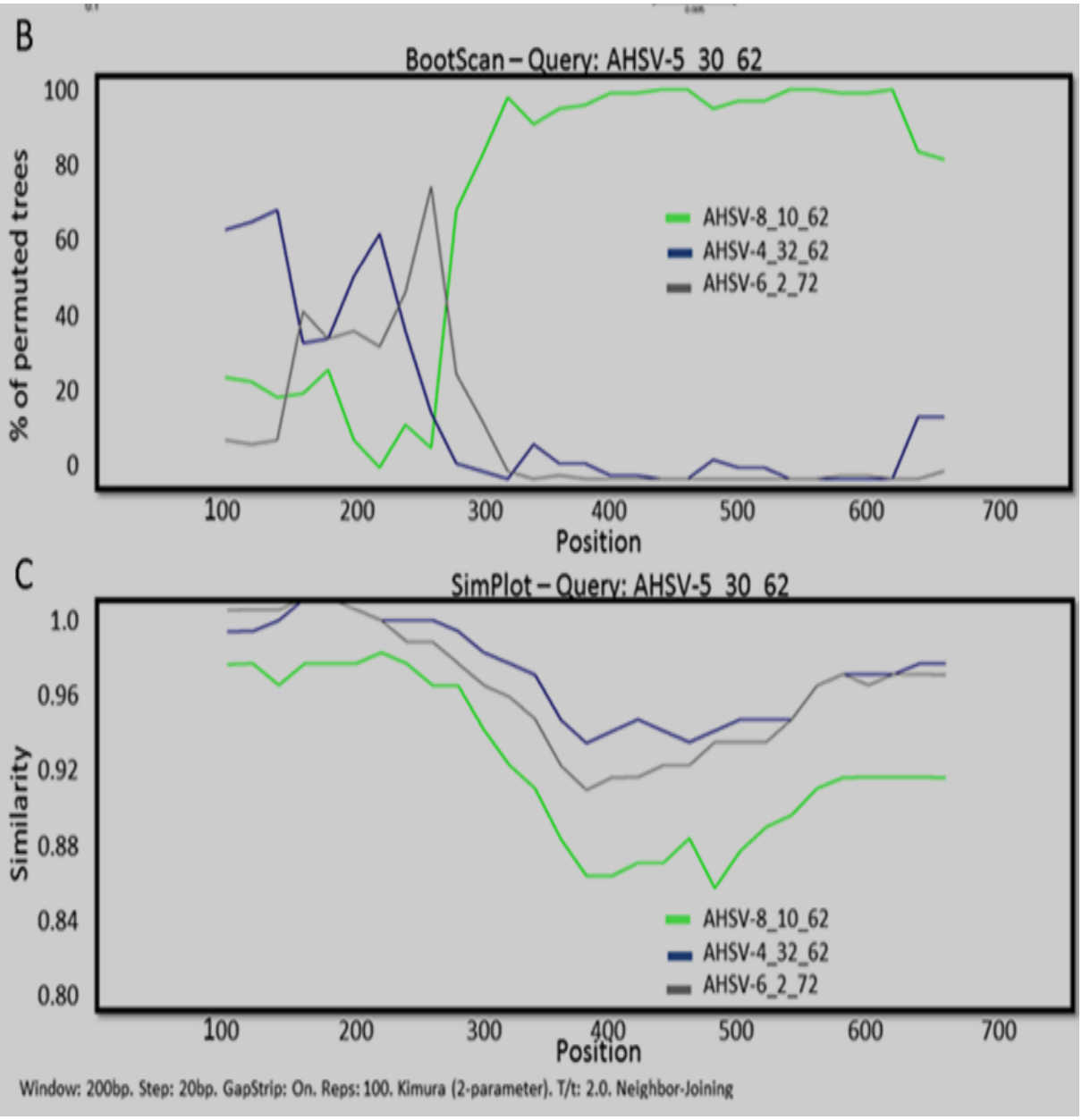

Figure 3.4 (B and C). Predicted recombination in Segment-10 of AHSV-5_30_62. (B) BootScan results of AHSV-5_30_62 with possible parents AHSV-8_10_62 and AHSV-4_32_62 and an outgroup AHSV-6_2_72. (C) Comparison of the sequence similarity between AHSV-5_30_62 and isolates AHSV-8_10_62, AHSV-4_32_62 and AHSV-6_2_72 using SimPlot.

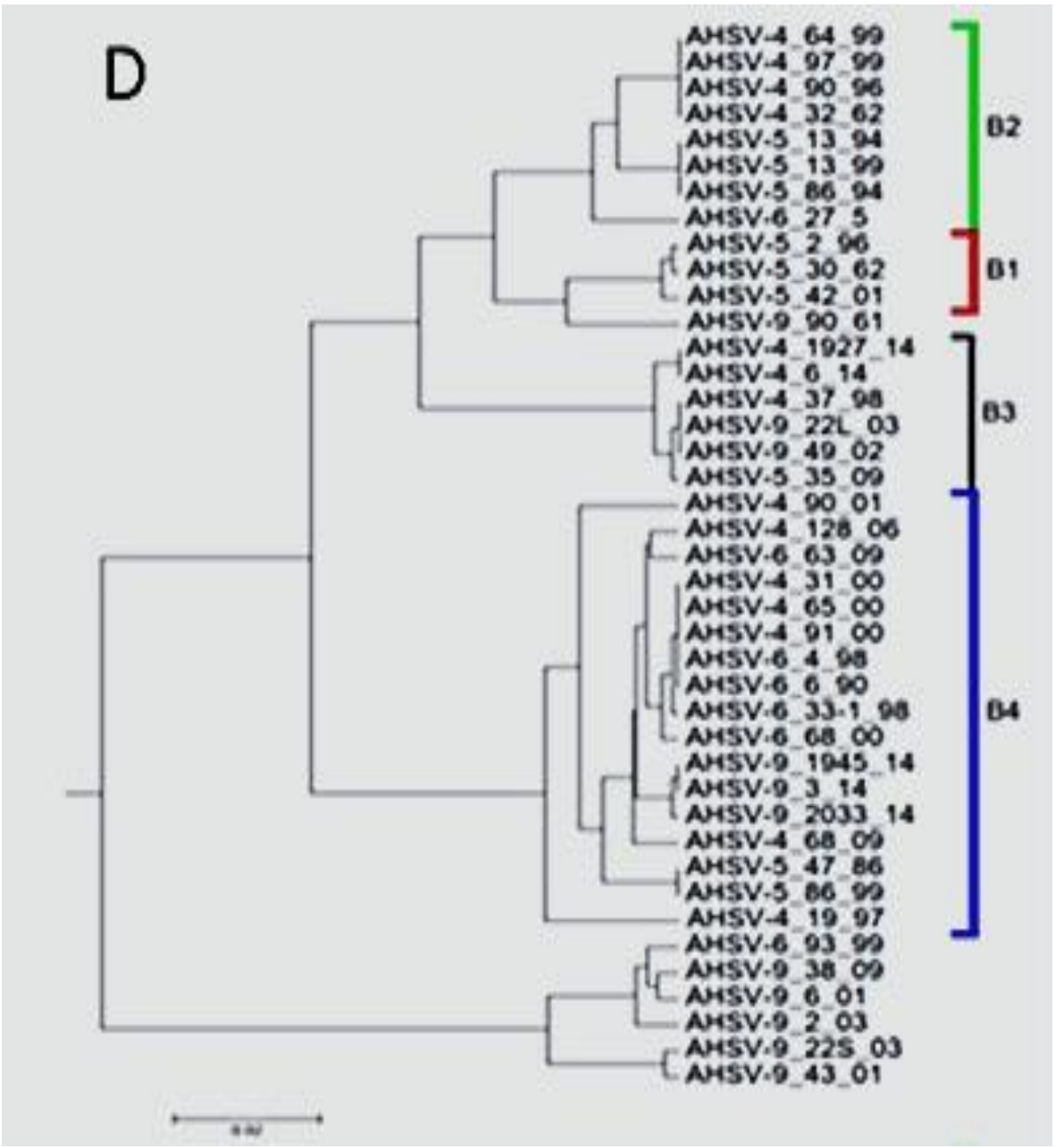

Figure 3.4 (D). Predicted recombination in Segment-10 of AHSV-5_30_62. (D) Phylogenetic tree of segment 10 sequence of region 193 – 757 of selected AHSV isolates.

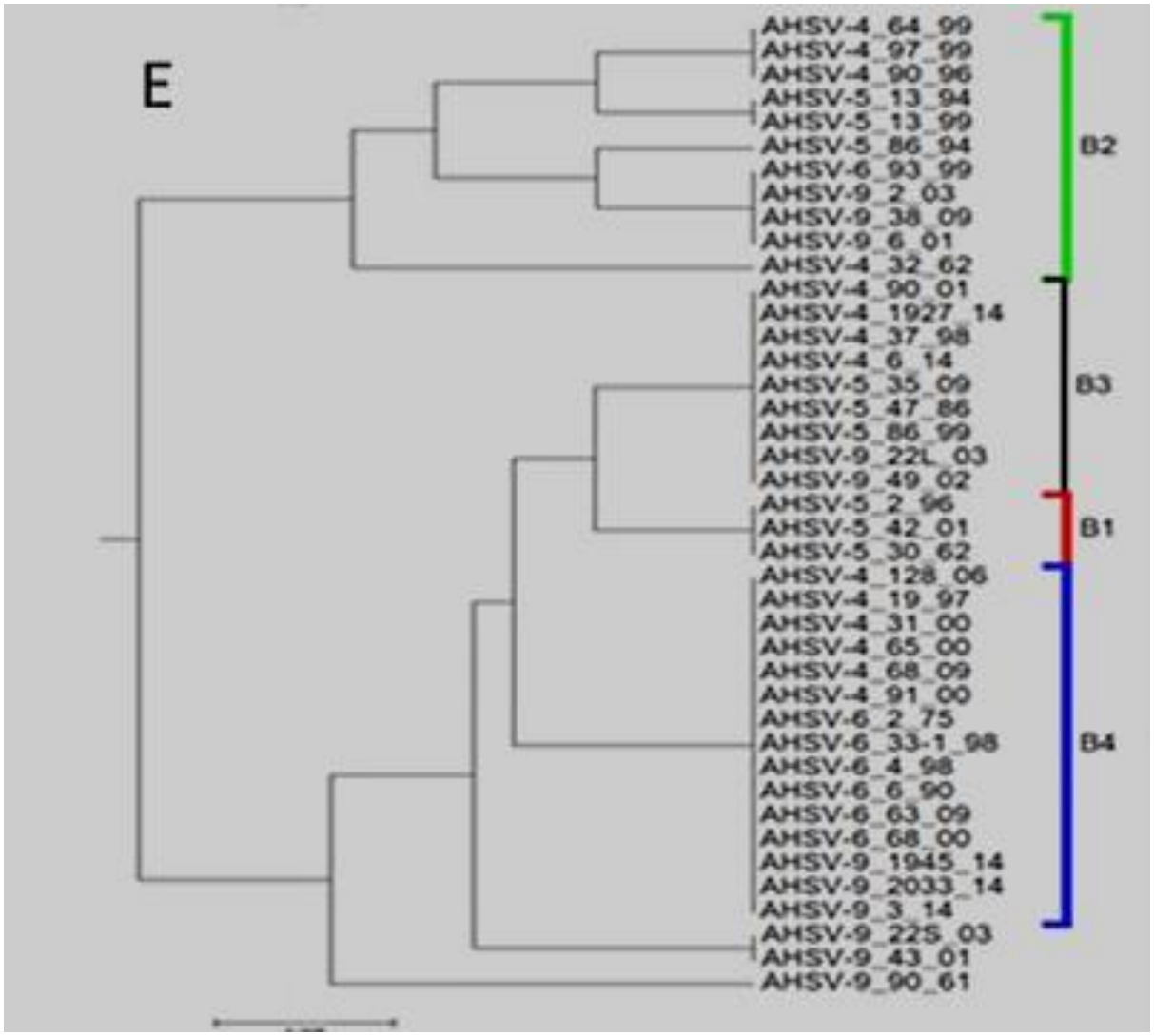

Figure 3.4 (E). Predicted recombination in Segment-10 of AHSV-5_30_62. Phylogenetic tree of region 1 – 192 of segment 10 from selected AHSV isolates.

3.5 DISCUSSION

Recombination has been described in all the major categories of RNA viruses at various frequencies and efficiencies (Perez-Losada *et al*., 2015). The rate of recombination is under the influence of the selection pressure exerted on the virus population. It can contribute to virus evolution if a viable phenotype is selected for by changing environmental conditions, or it rescues viruses by repairing fatal mutations in essential genes (Lai, 1992). In order for the genomic recombination events to occur, it is essential that a single host or vector cell is co-infected by two or more viruses of the same species (Novella *et al*., 2011; Simon-Loriere and Holmes, 2011).

This implies that multiple feedings on the viraemic host by a single vector, or that the host fed upon by multiple vectors infected with different strains of the same virus. Different methods of recombination exist, but in this chapter, analysis was only performed to detect possible homologous intragenic recombination. Intragenic recombination has been described in bluetongue virus in segments 1, 2, 3, 4, 7, 8 and 10 (He *et al*., 2010). According to our knowledge, this is the first study to report on genomic recombination events in an AHSV. These recombination events were less frequent compared to what has been reported previously in BTV (He *et al*., 2010).

The alignments of the individual genomic segments were generated using 101 AHSV sequences. These sequences included field isolates obtained between the year 1933 and 2014, reference strains and live attenuated vaccine (LAV) viruses. The alignments were used to identify possible recombination events in four genome segments and provided evidence of intragenic recombination. Segment-1 predicted a single isolate AHSV-3_2_89, containing a double cross-over mosaic gene. The AHSV-3_2_89 was isolated in 1989 from the blood of a vaccinated horse in KwaZulu-Natal, South Africa. Based on phylogenetic analysis, isolate AHSV-3_2_89 grouped within the major parental group, yet predictions were that the region between 1524 and 1623bp have recombined with viruses clustering with reference strain AHSV-3_13_63. The latter was used in the generation of the Live Attenuated Vaccine (Erasmus, 1978). It was documented that the horse that AHSV-3_2_89 was isolated from, was vaccinated, but no additional information exists pertaining the clinical signs of that horse. It is possible that isolate AHSV-3_2_89 is a result of recombination between the LAV serotype 3 and a field virus. Recent complete genome sequencing of LAV virus strains indicated that LAV-AHSV-3 has exchanged segment 1 with LAV-AHSV-1 (Guthrie *et al*., 2015). Guthrie *et al*. (2015) isolated individual serotypes from the bottle I LAV AHSV vaccine of Onderstepoort Biological Products using plaque selection in the presence of heterologous antibodies to the other serotypes, prior to full genome sequencing. Based on this information, recombination detected in segment 1 of AHSV-3_2_89 could be a result of LAV viruses recombining with field viruses and subsequently explaining why this event was only detected in a single isolate.

The second genome segment to present evidence of genomic recombination is Segment-6 encoding VP5. Clustering of segment 6 sequences according to serotypes was expected, since VP5 interacts with VP2 in the outer capsid, and VP2 as the major serotype-specific determinant (Martınez-Torrecuadrada *et al*., 1994). The inner layer of the AHSV's outer capsid comprises of a VP5 protein encoded by genome segment 6. VP5 has been shown to play a role in membrane destabilization prior to virus entry into the cells (Hassan *et al*., 2001). The genome segment is 1566 bp in size and no recombination was detected in BTV (He *et al*., 2010). The recombination events were predicted in the major parental group, consisting of serotypes 5, 3, 6 and 9. Recombination events with strong statistical support were predicted in isolate AHSV-3_DG25324_14 in relation to reference strains AHSV-3_13_63 and AHSV-5_30_62. The latter sequence was identified in LAV-AHSV-8, where reassortment of segment 6 between serotypes 8 and 5 was observed (Guthrie *et al*., 2015). Since isolate AHSV-3_DG25324_14 possesses a segment 2 sequence identical to LAV-AHSV-3 (Appendix 2, segment 6), it could be assumed that the single isolate recombination event was due to two vaccine viruses recombining. This might be true for the region 496 – 962, supported by SimPlot data and differential clustering of AHSV-3_DG25324_14 in region-specific phylogenetic trees. Evidence of a second recombination event was predicted in an AHSV-3_DG25324_14. Statistical support for this event is lower, no clear breakpoints were identified and the reference strain AHSV-9_90_61 was suggested as a parent along with multiple isolates including AHSV-3_48_01. Based on the data of the region 36bp - 1047bp, AHSV-3_DG25324_14 might not be the actual recombinant, but rather AHSV-9_90_61. Subsequent analysis excluding AHSV-3_DG25324_14 was performed and a strong signal for recombination was detected in an AHSV-9_90_61 with relation to field isolates of serotype 3, 6 and 9. Since sequences in this cluster had a genetic diversity lower than 5%, it was difficult to determine the amount of recombination with sufficient statistical support (Posada *et al*., 2002).

Genomic segment 7 had a double cross-over, predicted at positions 460 and 623 in BTV (He *et al*., 2010). Maximum likelihood analysis of all 101 AHSV segment 7 sequences, divide the isolates into a small group that included reference strains AHSV-5_30_62 and AHSV_2_82_61, their respective LAV isolates as well as field isolates derived from vaccine viruses. The remainder of the sequences, except for

isolate AHSV-5_86_84, were grouped in a major group. AHSV-5_86_84 was clustered independently of the two aforementioned groups and the only isolate to provide evidence of recombination. A single cross-over event around position 642 was predicted with reference isolate AHSV-5_30_62 parent of the region before the breakpoint and field isolates from the major group, AHSV-5_86_99 parent for the remainder of the sequence. Reference isolate AHSV-5_30_62 was originally used in the generation of the live attenuated vaccine of serotype 5, which was commercially available until 1995 from Onderstepoort Biological Products (Erasmus, 1978). Since this was a single isolated event, it is again plausible that recombination between a live attenuated vaccine virus and field isolates resulted in segment 7 of AHSV-5_86_99.

The last recombination event involves multiple isolates of segment-10 coding for the non-structural proteins NS3/NS3A. Phylogenetic analysis of segment-10 produces four independent major groups, with smaller clusters within each group. A single cross-over event around position 192 was predicted in multiple isolates, all of which were associated with the same group B1. Similarly, the parental sequences were multiple isolates belonging to groups B2 and B4 respectively. This point to an ancestral recombination event, which has since manifested into separate lineages.

These results provide evidence of intragenic recombination in African horse sickness virus, albeit less abundant than in bluetongue virus. In comparison, BTV has more antigenic diversity than AHSV with an ever-growing number of serotypes described, 27 and 9 respectively (Jenckel *et al*., 2015). The distribution of BTV includes every continent except Antarctica, while AHSV is endemic to sub-Saharan Africa with sporadic escapes to North Africa, the Middle East and Mediterranean countries (MacLachlan and Guthrie, 2010). Accounting for antigenic diversity, wide distribution, as well as the broader host range BTV has in comparison to AHSV, it is evident that BTV is under lower selection pressure and thus has a greater opportunity for genetic diversification. Intragenic recombination was predominantly observed in BTV field strains and subsequently contributing to genetic diversity and influencing its epidemiology (He *et al*., 2010). In contrast, the majority of intragenic recombination events observed in AHSV isolates were as due to exchange between live attenuated vaccine viruses and field isolates. The LAV viruses have a high prevalence in the

natural population since vaccination of horses using the trivalent and tetravalent vaccine from OBP occurs annually (von Teichman *et al*., 2010). It is possible that the high incidence of recombination detected between LAV and field viruses is an artefact of the recombination detection method or virus isolation method employed. Significant sequence diversity is required to detect recombination with sufficient statistical support, yet seg-3, seg-4, seg-5 and seg-8 of AHSV have more than 90% sequence identity between all 101 sequences compared. Since the majority of the viruses were isolated from South Africa or neighbouring countries, the small geographic distribution of isolates might contribute to the high percentage nucleotide identity observed. Phylogenetic analysis of individual genome segments of AHSV demonstrated that the majority of sequences grouped in temporal distribution rather than serotype-specific, with the exceptions of seg-2 (VP2) and seg-5 (VP5). This implies that co-circulating viruses are similar to one another, inhibiting the detection of recombination events that might have occurred between field isolates. Therefore, the only recombination detected by the methods used in this study is between genetically diverse older and recent samples, where the older samples were used in LAV development. The sequences used in this study were from the complete genomes of viruses isolated from cell culture, it is plausible that the recombination events involving LAV viruses provided the hybrid viruses with a selection advantage to cell culturing, rather than an increase in virulence or viral fitness.

Unfortunately, the intragenic recombination described here cannot account for the significant incongruence observed between AHSV segments. Such observation necessitates subsequent investigations into the role of reassortment on AHSV evolution.

CHAPTER 4 INVESTIGATING GENOMIC REASSORTMENT IN AHSV

4.1 BACKGROUND

The AHSV genome is segmented, which makes the virus prone to genome reassortment which is likely to play a major role in its diversity and evolution (MacLachlan and Guthrie, 2010). Reassortment is a common event among the orbiviruses and has been identified in a number of studies involving orbiviruses as contributing mechanism to their evolution (Gorman *et al.,* 1978; Oberst *et al.,* 1987; von Teichman and Smith, 2008; Allison *et al.,* 2012; Shafiq *et al.,* 2013; Shaw *et al.,* 2013; Weyer *et al*., 2016).

Genomic reassortment events can be analysed from phylogenetic trees by firstly creating sequence alignment (Nomikou *et al*., 2015), and can be done using ClustalW in the MEGA v6 software. The completion of a multiple sequence alignment is usually followed by estimation of the genetic distance between different sequences defined by number of substitutions in sequences that are divergent from the common ancestor (Liò and Goldman, 1998). Substitutions may not be observable particularly in the sequences that have many substitutions and these require the use of substitution models for analysis (Liò and Goldman, 1998). Substitution models are set of assumption about the nucleotide substitution that are used in phylogenetic analysis to describe the rate of change in fixed mutation between sequences that form the basis of evolutionary analysis at the molecular level (Posada and Crandall, 2001; Arenas, 2015).

Model selection implemented in the MEGA v6 can be used to select best fit substitution model to analyse divergence in the multiple sequence alignment (Tamura *et al*., 2013). Depending on the substitution type there are four models (Felstein 1981, Hasegawa-Kishino-Yano, Tamura-Nei and General Time Reversible) available for phylogenetic analysis of multiple sequence alignment in the MEGA v6 software and these models are supported by HyPhy implemented in MEGA software package (Tamura *et al*., 2013). Once a best fit substitution model is selected, an ML phylogenetic tree can be constructed using the Maximum Likelihood method that can be utilised for both the nucleotide and protein sequence data (Nei and Kumar, 2000). Bootstrapping method is applied to evaluate the accuracy of the estimated ML

phylogenetic trees (Efron *et al*., 1996; Ren *et al*., 2013). To achieve a high level of accuracy, the MEGA v6 provides Bootstrap test which uses bootstrap resampling and desired number of replicates can be entered into the program (Tamura *et al*., 2013). The constructed tree will display the highest log likelihood and the percentage of trees whereby the linked taxa grouped together in the branches (Tamura *et al*., 2013). The initial trees for the heuristic search are acquired by using Neighbor-Joining method in pairwise distances that are estimated using the Maximum Composite Likelihood (MCL) (Tamura *et al*., 2013). A Gamma distribution is utilised to model evolutionary rate variations among sites (4 categories). After the ML trees are constructed, Cluster picker v1.2, which is a Java program, can be used to identify clusters in the phylogenetic tree based on support for genetic grouping (bootstrap or posterior probability) as well as within cluster genetic distance (Ragonnet-Cronin *et al*., 2013). The settings in the Cluster Picker software include initial threshold, main support threshold for clusters, genetic distance threshold and large cluster threshold (Ragonnet-Cronin *et al*., 2013). A FigTree v1.4.2 file where sequences are coloured according to their clusters can be used to visually confirm the viral genomic reassortment events.

Apart from investigating natural occurring events, genome reassortment can be induced in the laboratory by co-infecting same cells with two different virus strains of the same or related species. Genomic reassortment has been induced through co-infection to investigate the effect of exchanges between AHSV vaccine and field isolates (Teichman and Smit, 2008). Such investigation was later followed by co-infection of cell line with different AHSV serotypes to analyse the effect of genomic reassortment in NS3 by Meiring et al (2009). Later, induced genomic reassortment events were observed among serological unrelated bluetongue viruses by Shaw *et al*., (2013). Before mixed infection begins, viral dilution factors and multiplicity of infection are determined by propagating and serially diluting two different viruses separately into the monolayer cells (Meiring *et al.,* 2009). Serial dilution is done in order to ensure that viruses grow at an equal rate during co-infection (Meiring *et al.,* 2009). After determining dilution factor and multiplicity of infection, viruses are co-infected, passaged several times and allowed to form plaques (Shaw *et al*., 2013). Individual plaques are then picked through agar overlaying and extracted. Genome

reassortants can be characterised from individual plaque extract using TBE-PAGE and visualised under UV-light.

Prior to the initiation of this study, there was a need to investigate genomic reassortment because the only known studies on reassortment were the ones described by von Teichman and Smit, (2008) and later by Meiring and colleagues in 2009. However, Weyer and colleagues (2016) later reported reassortment between field isolates and vaccine strains using samples collected between 2004 and 2014 and reported genomic reassortment as contributing to AHSV outbreaks. An in-depth investigation into the effect of reassortment on AHSV population structure was necessary. This study comprehensively used historic archival as well as recent field isolates collected between 1933 and 2014 to investigate the role of genome reassortment in AHSV evolution. The study also investigated the induced genomic reassortment by experimentally propagating two different AHSV serotypes into the same cell line.

4.2 AIMS OF STUDY

This work focused on AHSV genomic reassortment by investigating the following: (1). Natural occurring reassortment among 101 AHSV whole genome sequences from all nine AHSV distinct serotypes over 60 years, and (2). Experimentally co-infecting BSR cells with two different AHSV (AHSV 4 and 9) serotypes and detecting reassortant viruses through genome sequencing of purified progeny viruses.

4.3 MATERIALS AND METHODS

This study comprised of two sections that analysed genomic reassortment and included; 1), naturally occurring genomic reassortment using field AHSV isolates and 2) Inducing genomic reassortment using two different AHSV serotypes. The material and methods used for the two sections are described below.

4.3.1 Determining the effect of genomic reassortment on AHSV evolution by analysing 101 viral genomes isolated over 60 years.

4.3.1.1 Data sets

For the completion of the natural occurring genomic reassortment analysis, a total of 101 AHSV complete genome sequences (Table 1; Appendix 1) were used as in section 3.3.1.

4.3.1.2 Genetic reassortment analysis of field isolates using whole genome sequences that were deposited into GenBank.

Genome sequences were aligned using the complete genome nucleotide sequences of each isolate (101 AHSV full-length genome sequences). Alignments of each individual complete genome segment were created using ClustalW implemented in the MEGA v6 software (Tamura *et al*., 2013). The model of nucleotide substitution that best fits each multiple alignment was determined using model selection also implemented in the MEGA v6 software (Tamura *et al*., 2013). The substitution model was determined in order to ensure accuracy during the phylogenetic tree construction. Consequently, data were analysed using the appropriate model of nucleotide substitution to construct phylogenetic trees using Maximum-Likelihood.

Bootstrapping was utilised to evaluate the robustness of the phylogenetic tree with 1,000 replicates. Gamma distribution was utilized to model the evolutionary divergence rates between sites (4 categories) (Tamura *et al*., 2013). The maximum-likelihood trees were analysed using a Cluster Picker software to determine monophyletic clades with posterior estimates of 0.9 initial support threshold, 0.9 main support threshold, and 0.045 genetic distance threshold (Ragonnet-Cronin *et al*., 2013). By using Figtree file, reassortment events were visually confirmed on the ML phylogenetic trees based on grouping of sequences according to different colour clustering.

4.3.2 Experimentally investigating reassortment by co-infecting BSR cells with two different AHSV serotypes

4.3.2.1 Data sets

Two AHSV serotypes, a vaccine strain (Vac-4) AHSV 4 (AHSV-4_32_62) from RSA and a wild-type (WT-4) AHSV 9 (AHSV-9_90_61) isolate from Chad were selected based on differences in their electropherotypic patterns using PAGE analysis.

4.3.2.2 Monolayer Preparations and agar overlay

Similar method as described in section 2.3.2 was employed in the preparation of monolayer cells and agar overlay.

4.3.2.3 Determining viral dilution factors

Before the co-infection procedures commenced, viral dilution and multiplicity of infection were determined in order to ensure that both the viruses grow at an equal rate when incubated at the same conditions. In brief: A 10^{-1} – 10^{-10}serial dilution of both AHSV-4 and AHSV-9 isolates in suspension as described in section 2.3.2 were prepared in separate tubes. Monolayer cells were therefore infected with 330 µl of each of the two viruses per dilution in duplicates. The plates were incubated for 4 hours with regular agitation at 37^{o}C, followed by removal of the inoculum and subsequent overlaying of each well with 2 ml low melting agarose in DMEM medium containing 5% FBS and 25µg/ml PSA. The plates were incubated at 37^{o}C with 5% CO_2 for 3-5 days. Each well was stained with 5mg/ml 3-(4,5-Dimethylthiazol-2-Yl)-2,5-Diphenyltetrazolium Bromide (MTT) for 2-3 hours at 37^{o}C in order to fix the cells and to promote visualization of plaques. The plaque forming unit (PFU) count was determined by multiplying average plaques x dilution factor x ml of virus dilution added to cells. The multiplicity of infection (MOI) was determined by multiplying the PFU with virus used for infection/ number of cells. The desired volume of virus stock was defined as desired MOI (PFU/cell) x C_T (Cells). After determining the dilution factor and the desired volume of virus stock required for inducing genomic reassortment of two unrelated AHSV serotype, the investigations commenced.

4.3.2.4 Virus co-infection assay

A method previously described by Shaw *et al*., (2012) was employed with some modifications. In brief: BSR cells were grown in 12-well plates and co-infected in

duplicates with two different AHSV serotypes at an equal MOI of 1 and incubated at 37 ^{o}C for 2 hours. The inoculum was washed away and fresh media was added. Plates were incubated at 37 ^{o}C until CPE was observed. A volume of 5 µl of the co-infected cell culture material was passaged five times in the BSR cells, each time using the 5 µl of previous passage from harvesting the supernatants and cells at 48h PI. After the 4th passage, supernatants were then collected and used for the virus infection of the BSR cells in 6-well plates, which was then followed by an agar overlay. Single separate plaques were selected through the agar overlay, re-suspended in DMEM, and stored at 4^{o}C. This was then followed by viral RNA extraction from suspended plaques using QIAamp viral RNA mini kit (Qiagen) according to manufacturer's recommendations.

4.3.2.5 Characterization of individual plaques: Identification of genome segment reassortants

The extracted viral dsRNA was subjected to TBE-PAGE employing a ready-to-use 5% Mini-PROTEAN® TBE Gel, 10 well Precast Gels (Bio-Rad Laboratories, Inc. USA, www.bio-rad.com) in order to separate the AHSV genome segments (Laemmli, 1970). The FastRuler High Range and Middle Range DNA ladders (ThermoFisher-Scientific Inc, Waltham, MA USA), dsRNA from AHSV-4 and AHSV-9 and selected plaques were loaded and allowed to migrate at 100 volts for 16 hours. The gels were stained with EtBr in order to visualize the different genome segments under UV-light using a Syngen G.Box Gel Documentation System (GeneTools, Version 4.03.01.0).

4.4 RESULTS

The results in this chapter cover two sections, the investigation of natural occurring genomic reassortment and experimentally induced genomic reassortment.

4.4.1 Detecting naturally occurring genome reassortment

Reassortment events were identified firstly by aligning sequences of individual segments of 101 AHSV isolates using ClustalW on MEGA v6 software package. The evolutionary history of AHSV serotypes were inferred using Maximum Likelihood (ML) with the General Time Reversible model. The ML phylogenetic trees with highest log likelihood are shown. The ML tree analysis included codons of

$1^{st}+2^{nd}+3^{rd}$+non-coding. Consequently, a Gamma distribution was utilised for modelling the evolutionary rate differences between sites and the trees were drawn to scale, with branch lengths that were measured in numbers of substitutions/ site (Tamura *et al*., 2013). The complete genome sequences used for this analysis are the same as the ones used for genomic recombination as stated in section 3.3.1. These complete genome sequences represent viruses collected from 1933 until 2014. The reference strains as well as the ALV sequences from Onderstepoort Biological Products (OBP) South Africa were included. Cluster assignment of each genome segment was identified from the ML trees using Cluster Picker v1.2 software in order to group the sequences into their ancestral taxon. FigTree v1.4.2 was used to visually confirm genomic reassortment based on the monophyletic clustering of genome sequences on the ML trees. Identical sequences clustered together were characterised by same colour clusters and numbers.

4.4.1.1 AHSV Genome clustering and genomic reassortment overview

Based on the construction of Maximum Likelihood (ML) tree for the entire 101 AHSV, the cluster of each genome segment was assigned according to the phylogenetic lineages that were determined using Cluster Picker v1.2. Seg-2 and Seg-6 showed the highest number of clusters, forming 12 clusters each, as well as 7 and 17 numbers of tips not in the cluster, respectively (Table 4.1). Seg-3 and Seg-4 had the lowest number of clusters, each containing five clusters and three number of tips not in clusters that were shared among nine distinct serotypes across 101 isolates. Other segments such as Seg-1, Seg-3 and Seg-4 had a low number of tips that were not in clusters (2, 3 and 3, respectively). The highest number of tips not in clusters was recorded in Seg-7, Seg-9, and Seg-5 (33, 32, and 30 respectively), while Seg-6 and Seg-10 had relatively high number of tips not in clusters (17 and 23, respectively) across all nine serotypes. Seg-2 of AHSV-1, AHSV-2, AHSV-3, AHSV-5, AHSV-6, and AHSV-8 had one cluster each and none of those serotypes had tips that were not in clusters.

Since Seg-2 encodes serotype-specific protein, the VP2, clusters were unique across nine AHSV serotype. Seg-2 of AHSV-9 was grouped into cluster 9, however, there was one isolate (AHSV-9_90_61 reference strain) which could not be grouped into a cluster. Segment 2 (Seg-2) of AHSV-7 was one of the major contributing

serotypes to the Seg-2 high number of clusters with two clusters and three number of tips not in the cluster. The finding makes Seg-2 one of the highly diverse genome segments of AHSV. Base on the analysis of segment 2 (Seg-2), AHSV-4 was the only serotype that did not form any monophyletic branch with other serotypes. The AHSV-4 serotype was highly diverse, showing three clusters and two tips not in the clusters. On the other hand, Seg-3, which is highly conserved, only showed five clusters and three number of sequences (tips) not in clusters across the nine distinct AHSV serotypes. Of the five clusters that were associated with Seg-3, the majority of the AHSV serotypes were grouped into cluster 3. The AHSV-5 serotypes contributed greatly to variations associated with Seg-3, showing three different clusters, while AHSV-3 and 4 were grouped into 2 clusters, each. The AHSV-2 serotypes were dominated by cluster 3 of Seg-3, but two other clusters were picked including cluster 1 which was similar to Seg-3 of AHSV-1 1180 33.

Table 4.1 Cluster Picker showing the number of tips per cluster across 10 segments of 101 AHSV isolates.

								Number of tips per cluster and cluster numbers												
Genome segment	**Protein**	**Initial support threshold**	**Support threshold**	**Genetic distance threshold**	**Large cluster threshold**	**Tree tips**	**Number of clusters found**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Number of tips not in clusters**
Seg-1	VP1	0.9	0.9	0.045	10	101	7	4	2	40	4	14	10	25						2
Seg-2	VP2	0.9	0.9	0.045	10	101	12	8	7	11	14	12	12	5	2	12	4	4	3	7
Seg-3	VP3	0.9	0.9	0.045	10	101	5	6	9	74	3	6								3
Seg-4	VP4	0.9	0.9	0.045	10	101	5	8	3	31	41	15								3
Seg-5	NS1	0.9	0.9	0.045	10	101	9	5	8	2	8	17	7	6	6	12				30
Seg-6	VP5	0.9	0.9	0.045	10	101	12	4	11	11	3	4	4	4	5	3	3	5	27	17
Seg-7	VP7	0.9	0.9	0.045	10	101	11	8	3	9	14	2	2	3	3	8	9	7		33
Seg-8	NS2	0.9	0.9	0.045	10	101	9	6	8	41	27	3	2	4	2	2				6
Seg-9	VP6	0.9	0.9	0.045	10	100	10	37	6	3	3	7	2	4	2	2	2			32
Seg-10	NS3	0.9	0.9	0.045	10	101	8	18	9	9	8	20	9	3	2					23

Although the AHSV isolates were not selected equally across nine distinct serotypes, the findings in this study were significant. Reassortment events were seen to be serotype dependent instead of depending on the number of virus isolates used. For instance, of the seven AHSV-1 isolates used for this analysis, reassortment events were predicted in four isolates, with some isolates showing multiple reassortment events. On the other hand, AHSV-2 only showed evidence of reassortment on one isolate affecting one segment despite using eight isolates for the reassortment analysis. Of the 101 isolates, reassortment events were recorded in at least 39 (38.6%) AHSV isolates across nine different serotypes (See Table 4.2).

Table 4.2 Number of reassorted isolates per AHSV serotype

AHSV Serotypes	**Number of reassorted isolates/total and %**
ASHV-1	5/7 (71.4%)
AHSV-2	1/8 (12.5%)
AHSV-3	7/11 (63.6%)
AHSV-4	6/15 (40.0%)
AHSV-5	4/12 (33.2%)
AHSV-6	4/14 (28.6)
AHSV-7	2/8 (25.0%)
AHSV-8	3/13 (23.1%)
AHSV-9	7/13 (53.8%)
Total predicted reassorted isolates	**39/101 (38.6%)**

Based on the study findings, AHSV-1 had a higher number of reassortants followed by AHSV-3, while AHSV-2 had the least number of reassortants followed by AHSV-8 and AHSV-7 (See Figure 4.1). Reassortment events took place in all nine serotypes despite the variation in their frequencies. In three difference serotypes (AHSV-1, AHSV-3 and AHSV-4), reassortment events were also observed in the vaccine strains. Furthermore, two AHSV-5 isolates were seen to be identical to AHSV-5 reference isolate and these were isolated in 1999 and 2001, but there were no genomic reassortment events predicted in any of their segments. The study could not pick vaccine strains reassortants associated with AHSV-2, AHSV-6, AHSV-7 and AHSV-8, while AHSV-9 is excluded in the vaccine.

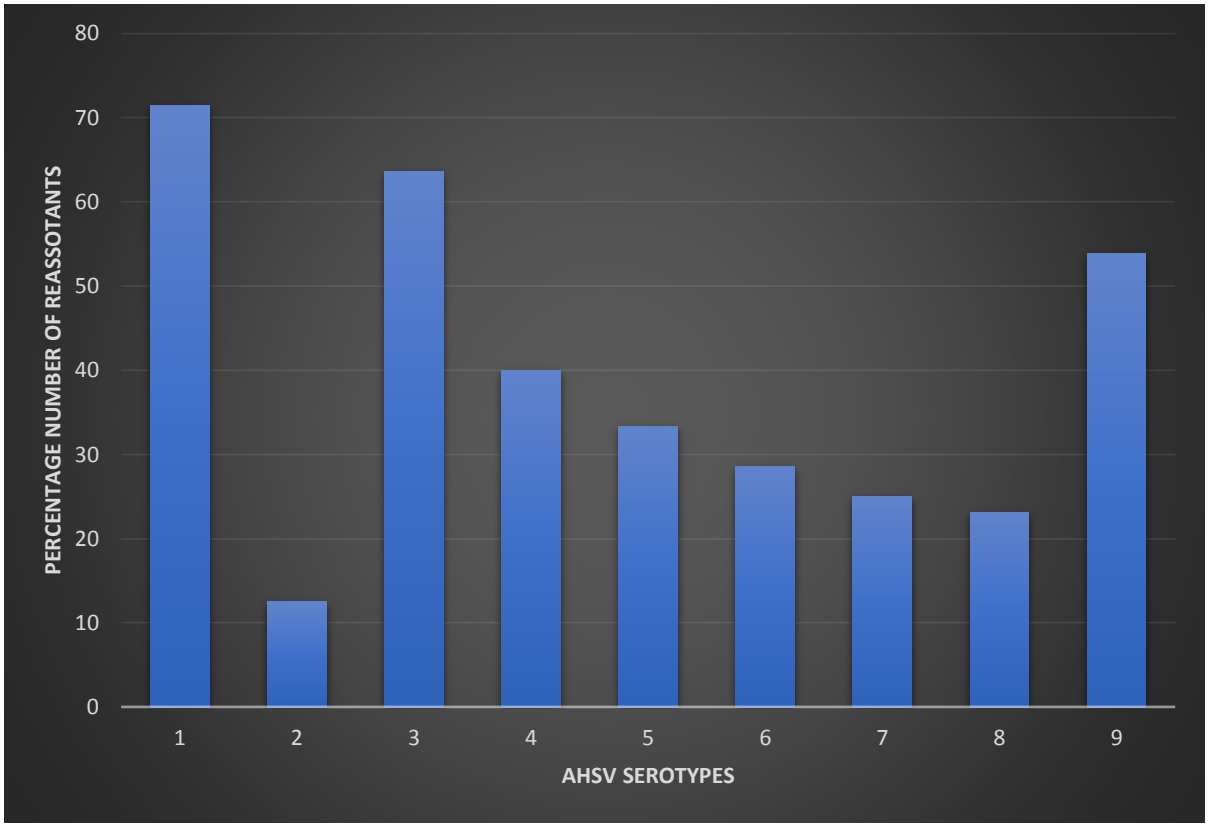

Figure 4.1 Percentage number of reassortants across nine distinct AHSV serotypes.

The analysis of individual genome segments across all nine distinct AHSV serotypes have shown a high number of predicted reassortment events associated with Segment 1, 5 and 7, recording 12 reassortment events each. Segment 6 and 8 had the smallest number of reassortants. As expected, Seg-2 despite having 12 clusters and 7 numbers of tips not in the clusters could not show any reassortment event, and this was also observed in segment 3, with only five clusters (see Figure 4.2).

The three most dominant reassortant AHSV segments were well distributed across the serotypes, with Seg-1 reassortment events predicted in six serotypes, Seg-5 in seven serotypes and Seg-7 in five serotypes. Both the Seg-1, 5 and 7 seem to be more prone to genomic reassortment events when compared to other AHSV genomic segments.

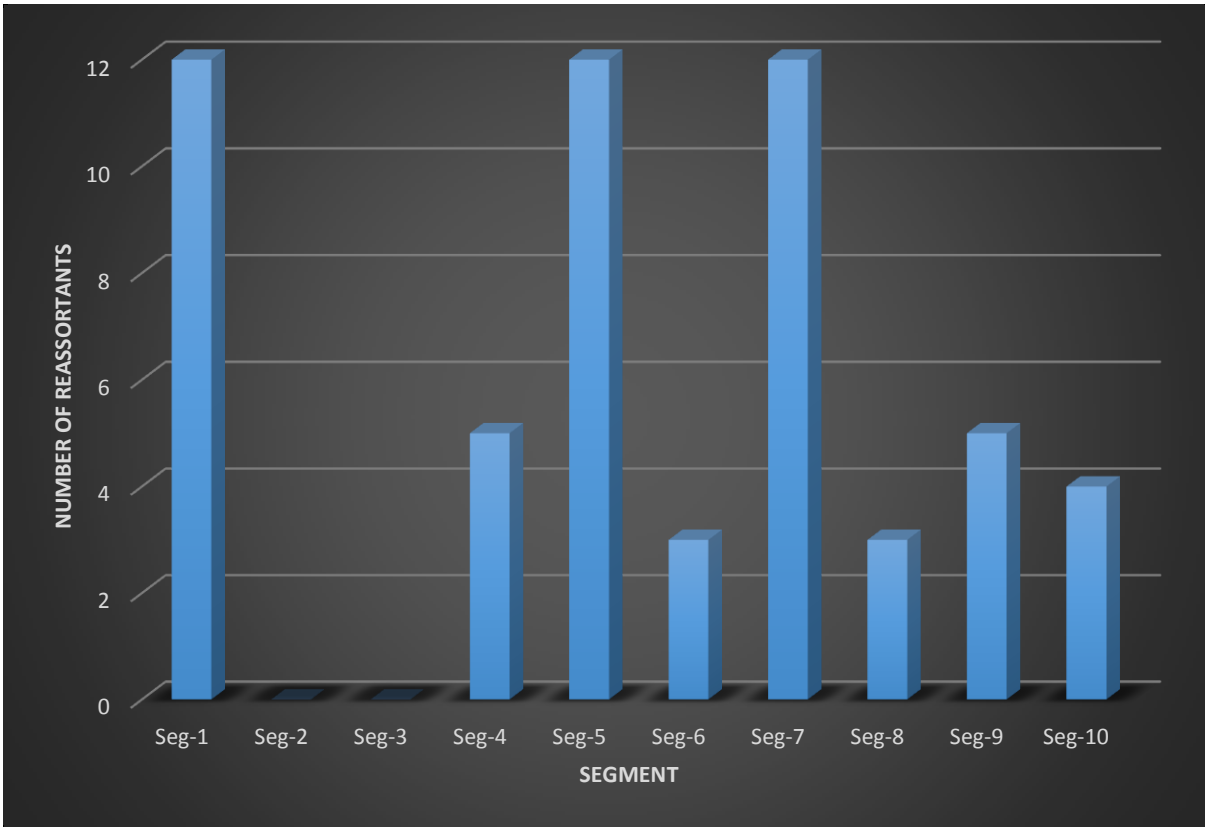

Figure 4.2. Number of inferred reassortment event per AHSV segment across nine distinct serotypes.

4.4.1.2 Evidence of wide-spread genomic reassortment across nine AHSV serotypes.

Cluster assignments of each AHSV genome segment were identified according to their phylogenetic lineages using Cluster Picker software and the sequences of the same genome segment that grouped together to form a cluster were given the same colour and numbers (See Figure 4.3). After alignment of all the 101 complete genome sequences from across nine distinct serotypes in one major ML phylogenetic, each serotype was assigned its phylogenetic tree based on Segment 2 using Maximum Likelihood (ML) implemented in the Mega v6 software program (See Figure 4.4 -4.12). Based on the phylogeny of AHSV seg-2 (Figure 4.3 and also refer to figure 9 in Appendix 2), the ML tree had three main branches, the first comprising of AHSV serotype 5, 6, 8 and 9; the second branch comprising only serotype 4 and the third branch comprising segment 1, 2, 3 and 7. The AHSV serotype 1 and 2 formed a monophyletic taxon separated into two smaller branches based on the variation on serotype level. Other monophyletic taxa separated according to their

seg-2 variations into small branches were observed between serotype 3 and 7, 5 and 8, as well as 6 and 9 (refer to Figure 9 in Appendix 2). However, serotype 4 did not form any monophyletic clustering with other eight serotypes. The evidence of genomic reassortment was observed across all nine AHSV serotypes. Genomic reassortment events are analysed as per individual AHSV serotypes below.

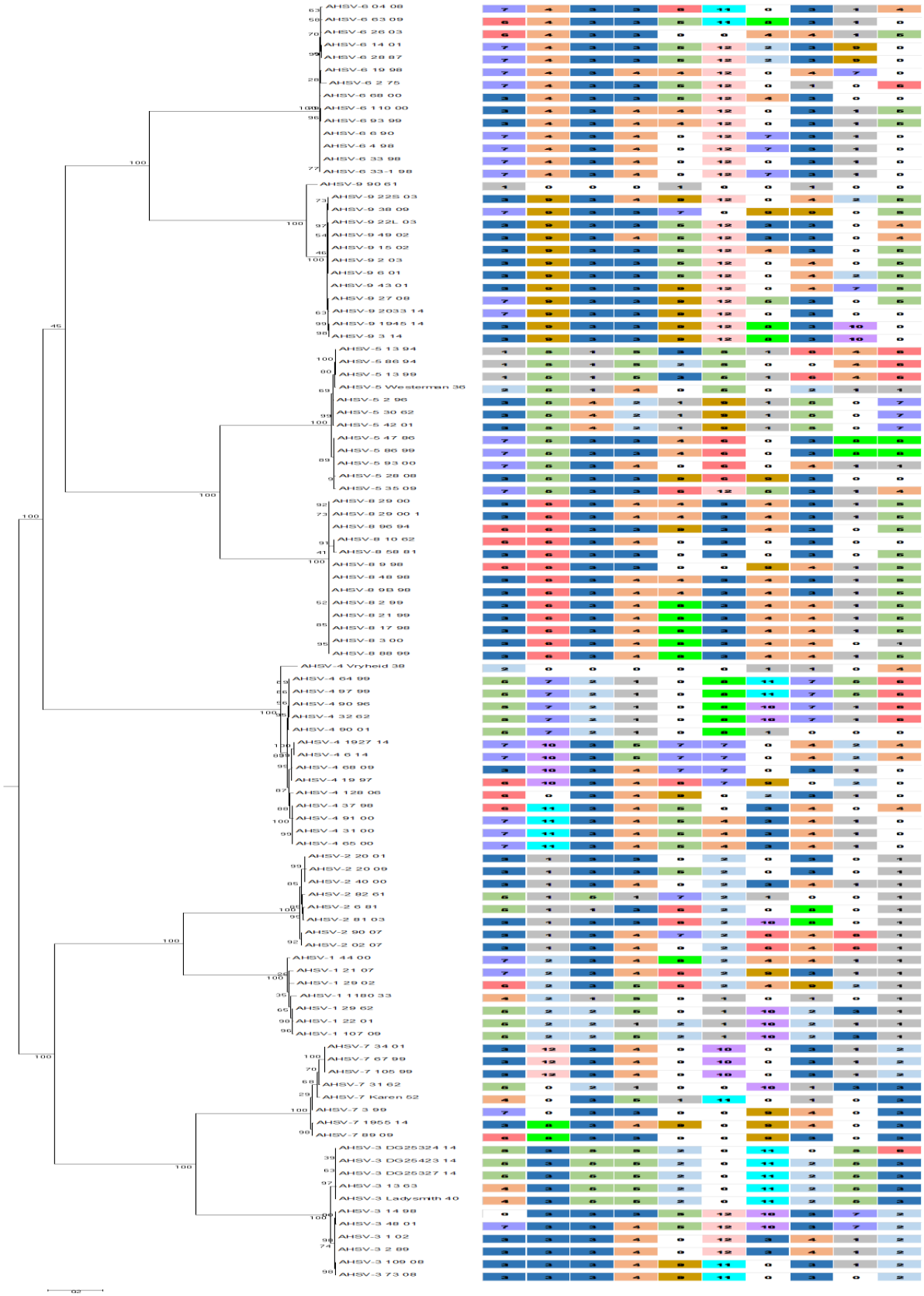

Figure 4.3. Phylogeny reconstruction of AHSV Seg-2 genetic reassortment patterns. The maximum likelihood tree of nine different serotypes making up to 101 AHSV isolates were included for genomic reassortment analysis. (The cluster assignment of each genome segment was assigned based on phylogenetic lineages that were identified using Cluster Picker Software). Sequences of same cluster indicated by same colour per genome segment. Number 0 represent number tips not in clusters. (Refer to Appendix 2, Figure 9 for VP2 ML- phylogenetic tree, and figure 4.4 to 4.12 for ML-tree of individual AHSV serotypes).

4.4.1.2.1 Genetic reassortment events predicted for AHSV-1

The evolutionary history of AHSV-1 was inferred using the ML method based on a GTR model. The analysis of reassortment was carried out on seven AHSV-1 complete genome sequences. These sequences included the reference isolate AHSV-1_29_62 used in making the AHSV live attenuated vaccine, as well as an isolate (AHSV-1_1180_33) from 1933 (See Figure 4.4 below). In this Maximum Likelihood (ML) tree, three branches were observed (See Figure 4.4 below). The first main branch clustered AHSV-1_107_09, AHSV-1_22_01 and AHSV-1_29_62 together. The second branch formed an outer group comprising only AHSV-1 _1180_33, and the third branch clustered AHSV-1_44_00, AHSV-1_21_07 and AHSV-1_29_02 in one group. Isolates AHSV-1 107 09 and AHSV-1_22_01 were vaccine derived strains that have reassorted with other field strains. Segment 5 of both AHSV-1_107_09 and AHSV-1_22_01 isolates showed evidence of reassortment that may have been acquired after mixed infection with AHSV-3 vaccine strain or other AHSV-3 strains.

Other reassortment events were observed in isolate AHSV-1_22_01 Seg-4 (Cluster 1) and Seg-9 (Cluster 1). Seg-4 of AHSV-1_22_01 reassortment event may have resulted from mixed infection with vaccine strain AHSV-2, AHSV-4, AHSV-5 or AHSV-7. On the other hand, Seg-9 of AHSV-1_22_01 reassortment event may have been driven by co-infection with AHSV-3, AHSV-4, AHSV-6, AHSV-7 or AHSV-8. The second evidence of genomic reassortment was observed in AHSV-1_44_00 in Seg-5 (Cluster 8), which could have been reassorted with AHSV-8. Reassortment event was also observed in AHSV-1_29_02 involving three segments (Seg-1, 4 and 9) which were grouped into cluster 6, 5 and 2, respectively. The Seg-1 of AHSV-1_29_02 may have reassorted after mixed infection with AHSV-6 or AHSV-8. The reassortment in Seg-4 of the AHSV-1_29_02 may have resulted from co-infection with other AHSV-1 or AHSV-3. Seg-9 of AHSV-1_29_02 could have be a result of reassortment after simultaneous infection between AHSV-1 and AHSV-4.

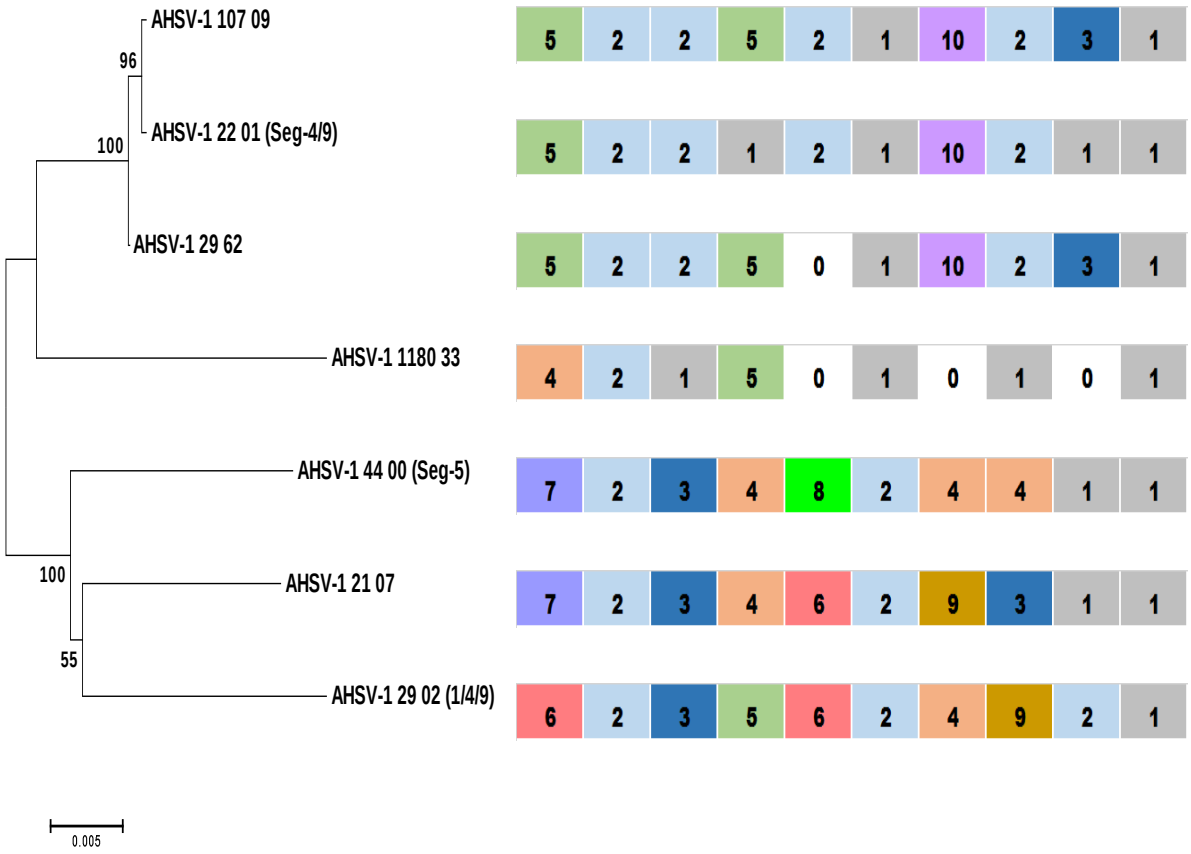

Figure 4.4. Molecular phylogenetic analysis of AHSV 1 isolates by Maximum Likelihood method. The evolutionary history was inferred using ML method based on GTR nucleotide substitution model. The tree with highest log likelihood is displayed. The percentage of trees in which the associated taxa grouped together is shown next to the branches. The maximum Likelihood tree estimated in Mega v6 software program with branch lengths scaled in years. Nucleotide segments are represented by a number and a unique colour that represent a particular cluster, whereas number 0 is assigned for tips not in the cluster.

4.4.1.2.2 Genetic reassortment in AHSV-2

The Maximum likelihood method based on the General Time reversible (GTR) nucleotide substitution model was used to infer the evolutionary history of AHSV-2 serotype. The evolutionary analyses were conducted in the MEGA v6. Eight AHSV sequences have been used for the analysis of AHSV-2 reassortment events (Figure 4.5). The AHSV-2 tree formed three branches. The first branch comprised of AHSV-2_20_01, AHSV-2_40_00 and AHSV-2_20_ 9, the second branch comprised of AHSV-2_82_61, AHSV-2-6_81 and AHSV-2_81_03 while the third branch comprised of AHSV-2_90_07 and AHSV-2_02_07. The evidence of genomic reassortment was observed in AHSV-2_90_07 involving segment 5 (Seg-5) which was grouped into cluster 7. The reassortment of Seg-5 of AHSV-2_90_07 could have resulted from

mixed infection with AHSV-2 vaccine strain. It was not clear whether Seg-3 of AHSV-2_6_81, which grouped into cluster 1 occurred as a result of genomic reassortment. However, it was identical to Seg-3 of AHSV-1_1180_33.

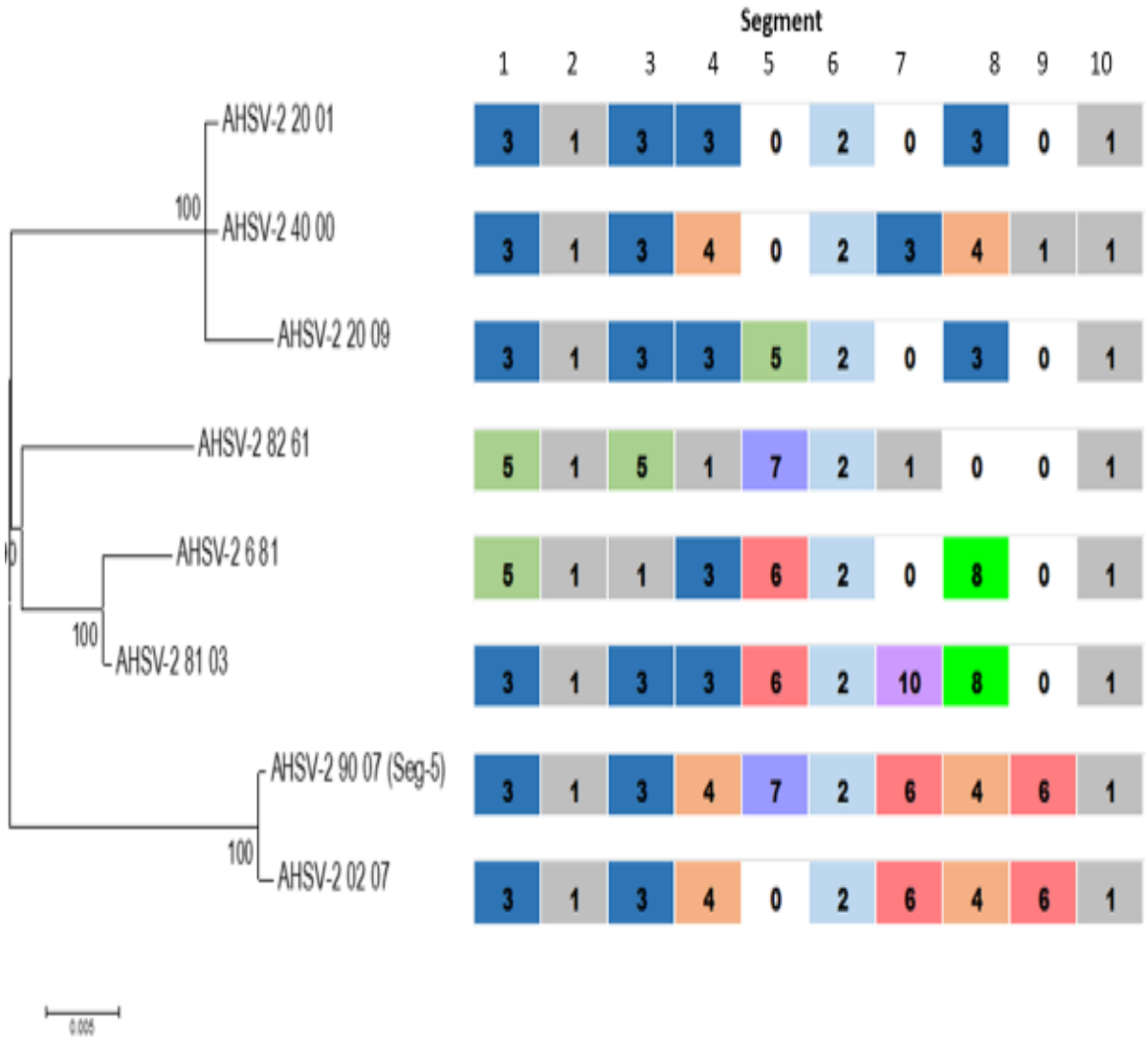

Figure 4.5. Molecular phylogenetic analysis of AHSV-2 isolates by Maximum Likelihood method. See Figure 4.4.

4.4.1.2.3 Genetic reassortment in AHSV-3

The evolutionary history of the AHSV-3 serotype was inferred using the Maximum Likelihood method based on the GTR nucleotide substitution model. Eleven AHSV-3 sequences were analysed for genomic reassortment and these included the OIE reference strain AHSV-3 13 63 that was also used to manufacture the ALV and an old isolate, AHSV-3 Ladysmith 40, retrieved from GenBank (See Figure 4.6 below). The ML tree for AHSV-3 formed two main branches. The first branch comprised of

AHSV-3_1_02, AHSV-3_2_89, AHSV-3_109_08, AHSV-3_73_08, AHSV-3_14_98 and AHSV-3_48_01, while the second branch included AHSV-3 Ladysmith 40, AHSV-3_13_63, AHSV-3_DG25324_14, AHSV-3_DG25327_14 and AHSV-3 DG25423 14. Reassortment events were recorded in isolates that includes AHSV-3_109_08 and AHSV-3_73_08 Seg-6 both belonging to cluster 11. However, there were no clear links between Seg-6's cluster 11 with any of the serotypes except for AHSV-7_Karen_52 which was obtained from GenBank. A possibility that the two isolates could have reassorted with AHSV-7 was evident. The second reassortment event in this branch involved AHSV-3_14_98 Seg-4 grouped into cluster 3 and could have reassorted with AHSV-6. The last reassortment event in this branch was recorded in AHSV-3_48_01 Seg-1 that falls within cluster 7 and could have reassorted with AHSV-5, AHSV-6 or AHSV-9. On the second branch of this ML tree, reassortment events were observed in AHSV-3_DG25324_14 Seg-1 and 10, AHSV-3_DG25327_14 Seg-1 and AHSV-3_ DG25423_14 Seg-1. These three isolates (AHSV-3_DG25324_14, AHSV-3_DG25327_14 and AHSV-3_DG25423_14) were vaccine derived strains that were excluded from the vertical analysis and could have reassorted with AHSV-1 or AHSV-4 vaccine strain. The Seg-1 from both the three serotypes (AHSV-3_DG25324_14, AHSV-3_DG25327_14 and AHSV-3_DG25327_14) was grouped into cluster 5, similar to AHSV-1 and AHSV-4 vaccine strains. Seg-10 of AHSV-3_DG25324_14 may have resulted from a reassortment event after mixed infection with AHSV-4 or AHSV-5.

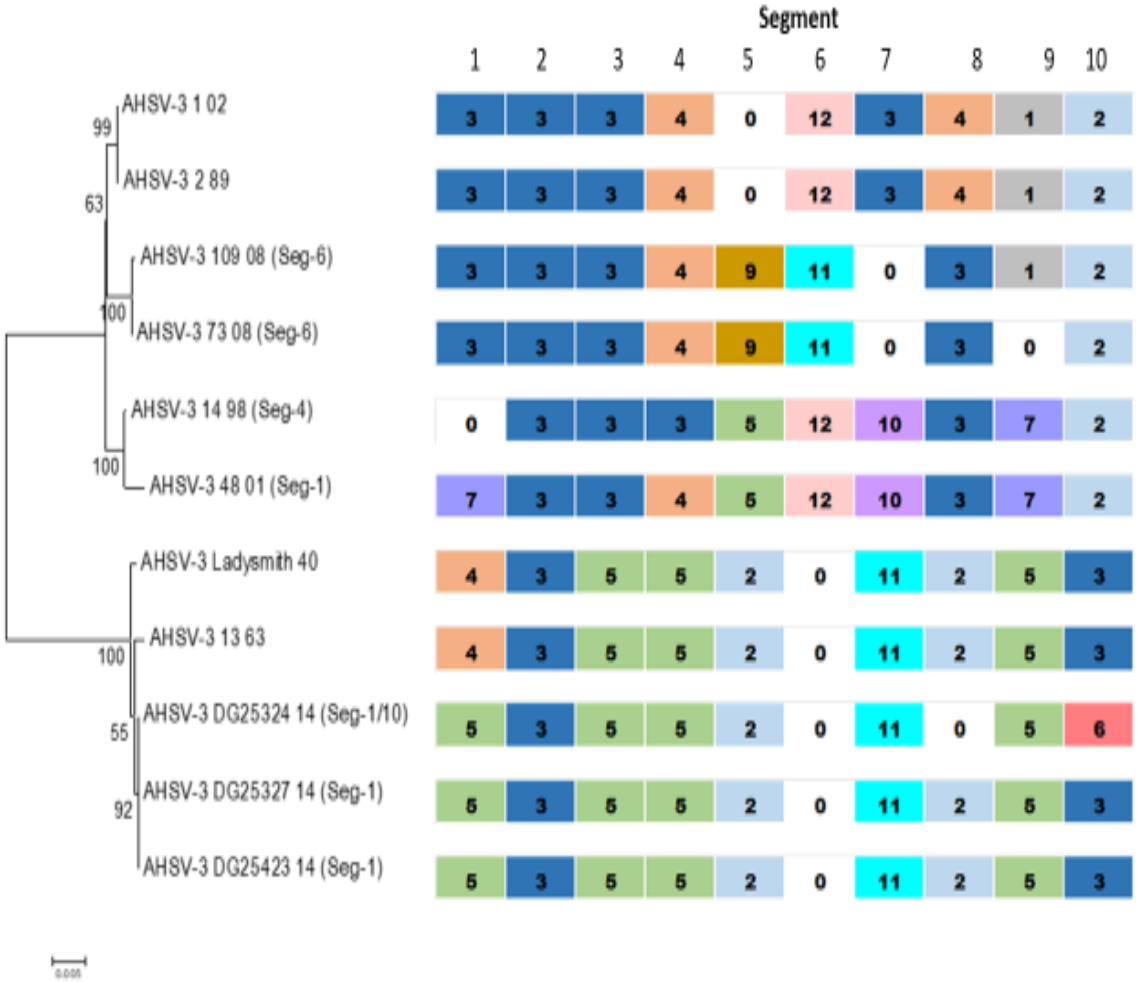

Figure **4.6. Molecular phylogenetic analysis of AHSV -3 isolates analysed using Maximum Likelihood method.** See Figure 4.4.

4.4.1.2.4 Genetic reassortment in AHSV-4

The evolutionary history of AHSV-4 inferred by applying the Maximum Likelihood method based on the GTR nucleotide substitution model. The evolutionary analyses conducted in the MEGA v6 software. To analyse reassortment events in the AHSV-4 isolates, a total number of 15 sequences were used. These included the OIE reference isolate AHSV-4_32_62 and an old isolate, AHSV-4 Vryheid_38 that was retrieved from GenBank (See Figure 4.7 below). The ML three has shown two main branches and one outer branch. The first main branch included AHSV-4_65_00, AHSV-4-91-00, AHSV-4_31_00, AHSV-4_37_98, AHSV-4_128_06, AHSV-4_19_97, AHSV-4_68_09, AHSV-4_1927_14 and AHSV-4_6_14. The second branch comprised of AHSV-4_90_01, AHSV-4 32_62, AHSV-4_90_96, AHSV-4_64_99 and AHSV-4_97_99. The outer group comprised of AHSV-4_Vryheid_38. The sequences of the second branch were more identical to the OIE reference sequences AHSV-

4_32_62, which was also used to manufacture ALV, although Seg-7 of AHSV-4_90_01 (cluster 1), AHSV-4_64_99 and AHSV-4_97_99 (both cluster 11) were different from the reference sequence. The four AHSV-4 isolates (AHSV-4_90_01, AHSV-4_90_96, AHSV-4_ 64_99 and AHSV-4_97_99) were identified as vaccine derived strains that were initially removed from vertical analysis study. While isolate AHSV-4_90_96 was similar to the vaccine strain in all its segments, Seg-7 of AHSV-4_64_99 and AHSV-4_97_99 were grouped into cluster 11 and may have reassorted after co-infection with AHSV-3 vaccine strain. Apart from vaccine strains, reassortment events were observed in the isolate AHSV-4_128_06 involving Seg-5 and Seg-7. The Seg-5 of AHSV-4_128_06 may have reassorted after co-infection of the host cells with AHSV-5, AHSV-7 or AHSV-9, grouped within cluster 9. On the other hand, Seg-2 of AHSV-4_128_06 grouped into cluster 2, may have reassorted after co-infection with AHSV-6. Another reassortment event was observed in Seg-5 (Cluster 6) and Seg-7 (Cluster 9) of the AHSV-4_19_97 isolate. Segment-5 of AHSV-4_19_97 could have resulted from reassortment after co-infection with AHSV-2 while Seg-7 may have emerged from the reassortment event as a result of mixed infection with AHSV-7. Another evidence of a genomic reassortment event was observed in Seg-1 (Cluster 3) of AHSV-4_68_09, which may have reassorted after co-infection with AHSV-2, AHSV-3, AHSV-5, AHSV-6, AHSV-7, AHSV-8 or AHSV-9, taking into consideration that it is the only Seg-1 within AHSV-4 that falls within Cluster 3. The AHSV-4_Vryheid_38, which formed an outer group, was collected in 1938 and its Seg-1 formed cluster 2 with AHSV-5_Westerman_36 that was also collected earlier in 1936. The last reassortment event observed in Seg-7 of AHSV-4_90_01 grouped into cluster 1 and could have reassorted with AHSV-2 or AHSV-5 after mixed infection.

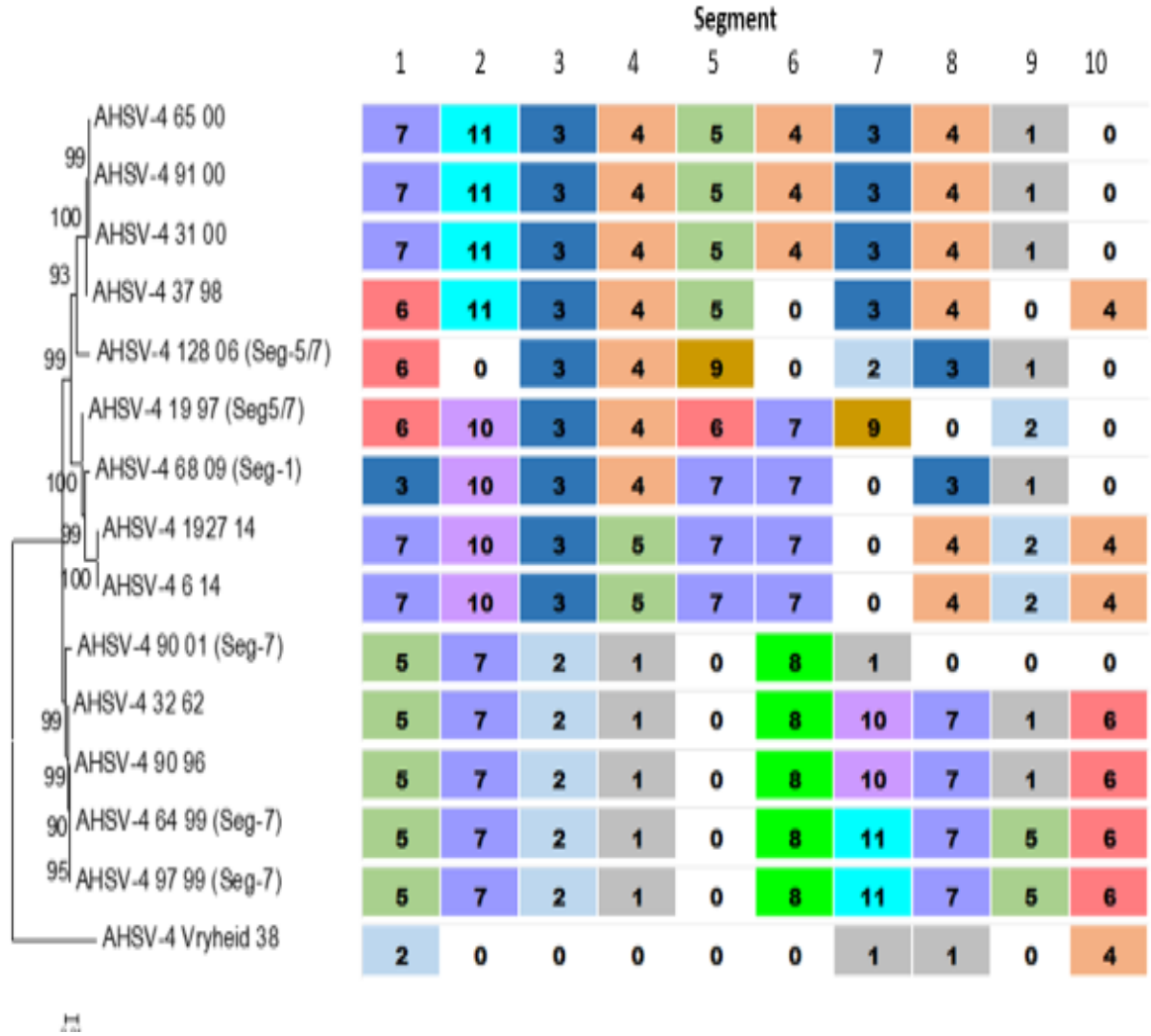

Figure 4.7. Molecular phylogenetic analysis of AHSV-4 isolates analysed using Maximum Likelihood method. See Figure 4.4.

4.4.1.2.5 Genetic reassortment in AHSV-5

The Maximum Likelihood method based on General Time Reversible nucleotide substitution model was utilised to infer the evolutionary history of AHSV-5. The analysis involved 12 AHSV-5 isolates, including the OIE reference strain AHSV-5_30_62 collected in 1962, as well as an old isolate AHSV-5_Wasterman_36 collected in 1936 (Figure 4.8). The AHSV-5 formed two main branches. The first main branch composed of two sub-branches. One sub-branch consisted of AHSV-5_13_99, AHSV-5_86_94, AHSV-5_13_94 and an outer group AHSV-5_Westerman_36. Included on the second sub-branch were isolates AHSV-5_42_01, AHSV-5_2_96 and an OIE reference isolate AHSV-5_30_62. The AHSV-5_42_01 and AHSV-5_2_96 were vaccine derived isolates identical to vaccine strain withdrawn in 1994 and did not undergo genomic reassortment. These two vaccine

derived isolates were initially excluded from vertical analysis study (Chapter 2) because they were not part of the clock. The second main branch included AHSV-5_47_86, AHSV-5_86_99, AHSV-5_28_08, AHSV-5_35_09 and AHSV-5_93_00. The reassortment event was observed in Seg-5 of AHSV-5_86_94, which fell into cluster 2, a cluster common in serotype 3. Another co-infection may have occurred between AHSV-3 with AHSV-5_28_08 resulting in reassortment of Seg-5. Within the same isolate (AHSV-5_28_08), Seg-7 was recorded as a reassortant segment grouped into cluster 9, which could have resulted from mixed infection with AHSV-4, AHSV-1, AHSV-7, AHSV-8, or AHSV-9. The AHSV-5_35_09 had four segments (Seg-5 cluster 6, Seg-6 cluster 12, Seg-7 cluster 5 and Seg-9 cluster 4) that were showing evidence of genomic reassortment. Genomic reassortment in Seg-5 of AHSV-5_35_09 could have resulted from mixed infection with AHSV-1 or 2. The Seg-6 of AHSV-5_35_09 could have reassorted as a result of mixed infection with AHSV-6. In the same isolate (AHSV-5 35 09) Seg-7 reassortment could not be linked with any isolate that was isolated earlier, however, it was similar to AHSV-9_27_08. Seg-7 of AHSV-9_27_08 may have risen due to other evolutionary events instead of reassortment, but later able to reassort with other AHSV serotypes. The last reassortant segment within isolate AHSV-9_27_08 was Seg-10, grouped into cluster 4 and may have reassorted after co-infection with AHSV-4 or AHSV-9. The last reassortment event in AHSV-5 isolates was observed in an isolate AHSV-5_93_00 involving Seg-8 grouped into cluster 4. The Seg-8 of AHSV-5_93_00 may have been acquired after the host cells were co-infected with AHSV-1, AHSV-2, AHSV-3, AHSV-4, AHSV-7, AHSV-8 or AHSV-9 serotype.

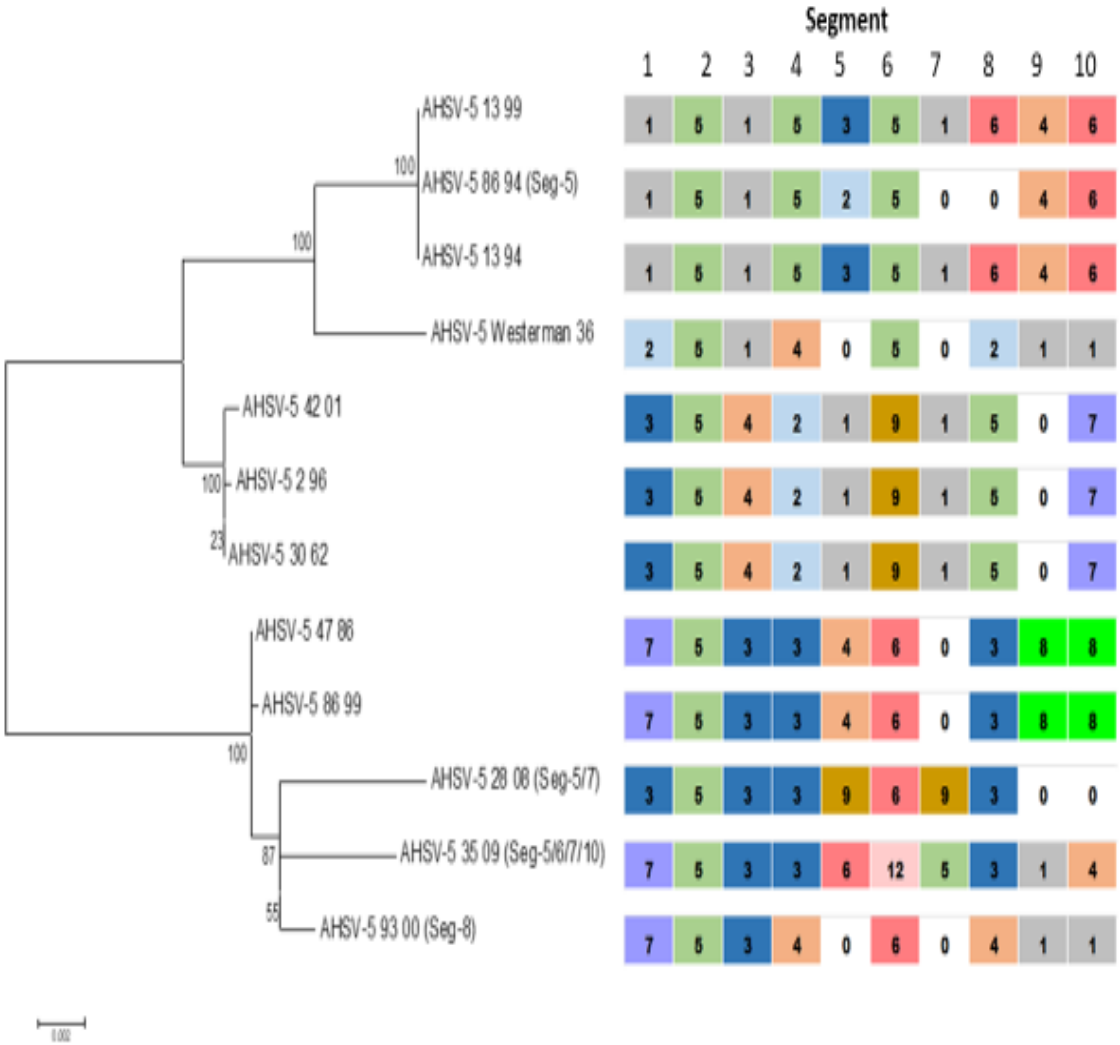

Figure **4.8. Molecular phylogenetic analysis of AHSV -5 isolates analysed using Maximum Likelihood method.** See Figure 4.4.

4.4.1.2.6 Genetic reassortment in AHSV-6

The ML method based on General Time Reversible (GTR) nucleotide substitution model was used to infer the evolutionary history of AHSV-6. Fourteen isolates were used to construct the ML tree for the AHSV-6 serotype. The ML tree was estimated in the MEGA v6 using GRT model for the 5' and 3' UTRs nucleotide substitution type. The 14 isolates included the OIE reference strain AHSV-6_2_75, which formed an outer group while others clustered together to form one main branch (See Figure 4.9 below). Reassortment events were observed in Seg-1 of AHSV-6_26_03 (cluster 6), which may have occurred after co-infection with AHSV-4 or AHSV-8. The AHSV-4_04_08 had two segments (Seg-5 and Seg-10) that emerged due to genomic reassortment and grouped into cluster 6 and 4, respectively. Cluster 6 of Seg-5 was commonly associated with AHSV-5; however, Seg-5 of AHSV-4_04_08 could have

emerged after co-infection of the same host cells with AHSV-1 or AHSV-2 serotypes. On the other hand, Seg-10 of AHSV-4_04_08 was grouped into cluster 4 and may have resulted from co-infection with AHSV-4. Another reassortment event was observed in the isolate AHSV-6 63 09 involving Seg-1 (cluster 6) and Seg-7 (cluster 8), which could have occurred after co-infection with AHSV-5 or AHSV-8 (Seg-1) and AHSV-2 (Seg-7). Segments 8 and 9 of AHSV-6_19_98 grouped into cluster 4 and 7, respectively, were observed as reassortants and could have occurred after a mixed infection with AHSV-4, AHSV-8 or AHSV-9 (Seg-8) and AHSV-4 for Seg-9.

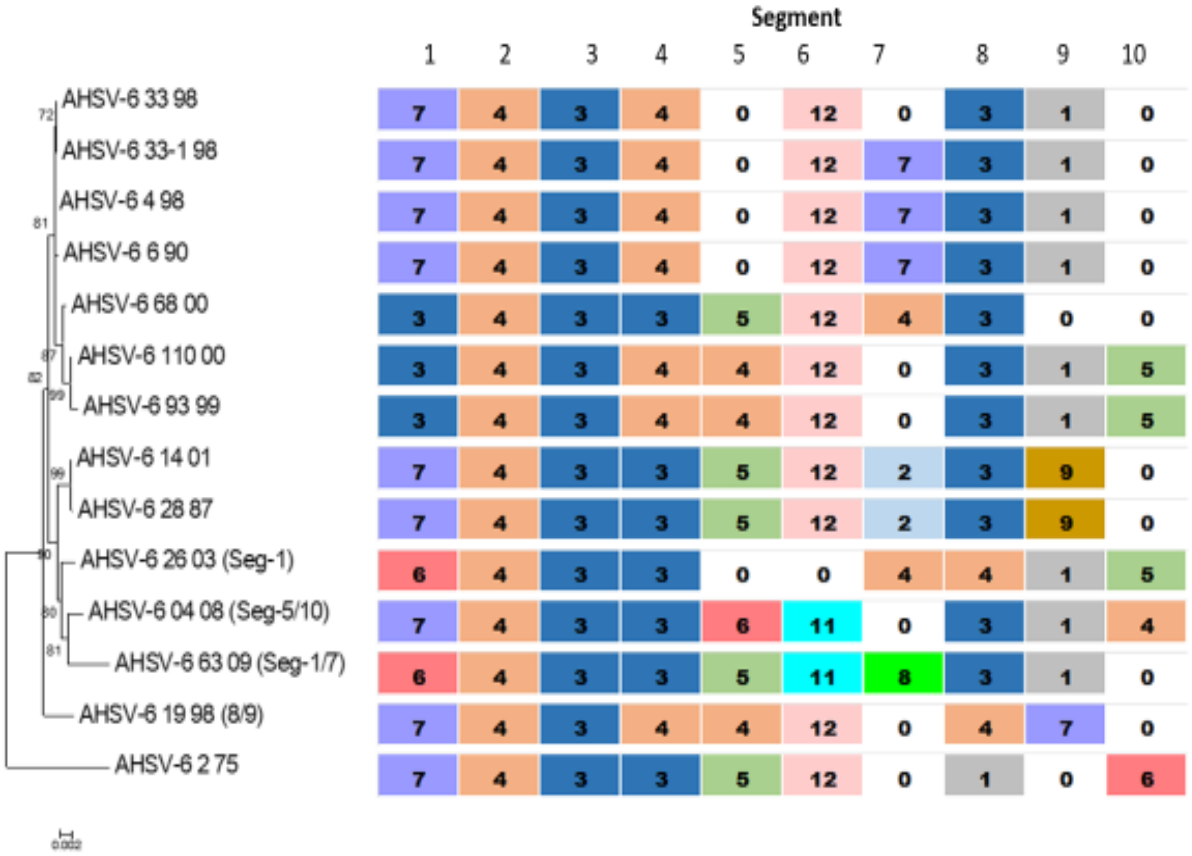

Figure 4.9. Molecular phylogenetic analysis of AHSV -6 isolates analysed using Maximum Likelihood method. See Figure 4.4.

4.4.1.2.7 Genetic reassortment in AHSV-7

The ML method based on the GTR nucleotide substitution model had been used to infer the phylogenetic history of eight AHSV -7 genome sequences. These eight AHSV-7 genome sequences had two main branches and two outer groups (See Figure 4.10 below). The first main branch grouped AHSV-7_105_99, AHSV-7_34_01 and AHSV-7_67_99 and had genome segments that were identical, though Seg-5 and Seg-7 sequences of each isolate could not be grouped into clusters. The second

branch consisted of AHSV-7_3_99, AHSV-7_1955_14 and AHSV-7_89_09. The two outer groups included an OIE reference strain AHSV-7_31_62 and AHSV-7_Karen_52. Reassortment events were observed in the second main branch involving two isolates. The first reassortment event was observed in Seg-1 of AHSV-7_3_99, grouped into cluster 7. The reassortment event may have been due to co-infection with AHSV-5 or AHSV-6. The second reassortment event was observed in Seg-1 of AHSV-7_89_09, which was grouped into cluster 6 and could have reassorted after co-infection with AHSV-4, AHSV-6 or AHSV-8. There were some genomic segments that could not be grouped into clusters and these may have been emerged because of other evolutionary mechanisms such as mutation.

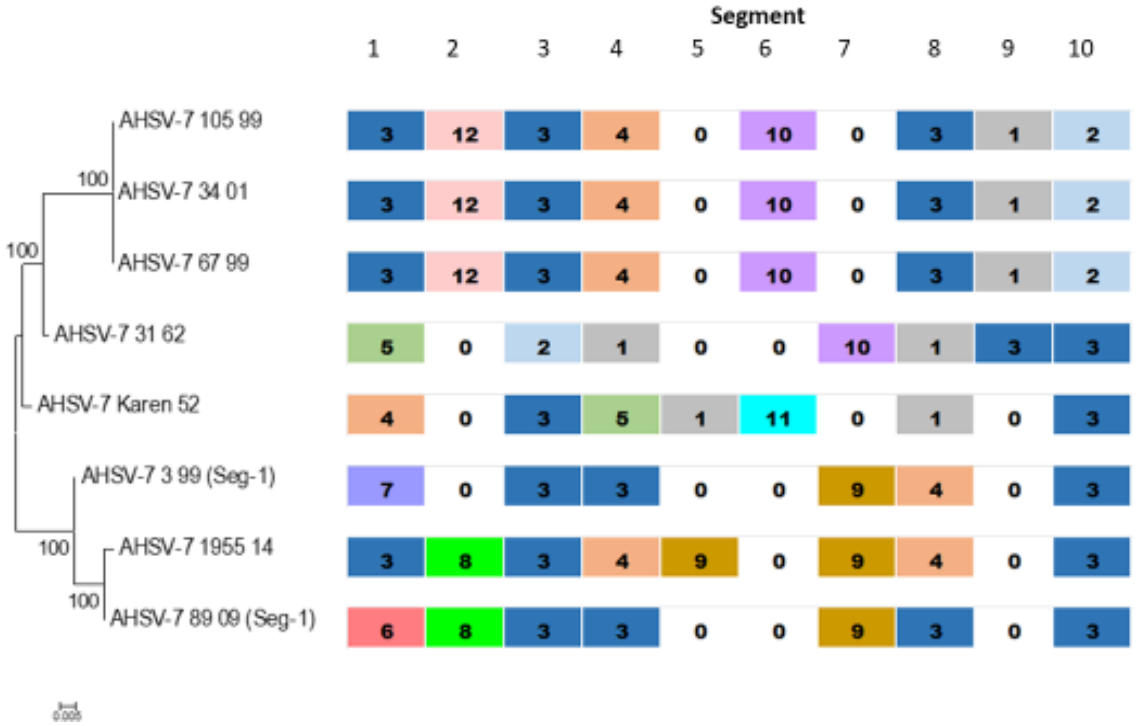

Figure 4.10. Molecular phylogenetic analysis of AHSV-7 isolates analysed using Maximum Likelihood method. See Figure 4.4.

4.4.1.2.8 Genetic reassortment in AHSV-8

The ML method based on the GTR nucleotide substitution model was used to infer the evolutionary history on 13 of AHSV-8 serotype. Subsequently, the phylogenetic tree for the AHSV-8 sequences has shown one big branch and a small branch that formed an outer group with two isolates that included an OIE reference strain (AHSV-8 10 62). The first big branch that separated into two sub-branches with the first branch comprising of 10 isolates collected between 1998 and the year 2001.

The smaller branch comprised of one isolate (AHSV-8_9_98). The first evidence of genomic reassortment was observed in Seg-10 of the AHSV-8_3_00 which grouped under cluster 1 instead of cluster 5 common in other AHSV-5 isolates. The reassortment event may have been driven from co-infection with either AHSV serotype 1 or serotype 2. The second reassortment event was recorded in Seg-5 of AHSV-8_96_94, which belonged to cluster 9 instead of 8 or 4 as shared by other isolates within the main branch. The genomic reassortment event in Seg-5 of AHSV-8_96_94 seems to have occurred after the mixed infection with AHSV-5, or AHSV-9. Segment 7 (Seg-7) of the AHSV-8_9_98 was grouped into cluster 9 instead of cluster 4 shared by other isolates within the main branch and may have reassorted after co-infection with AHSV-7 (See Figure 4.11 below).

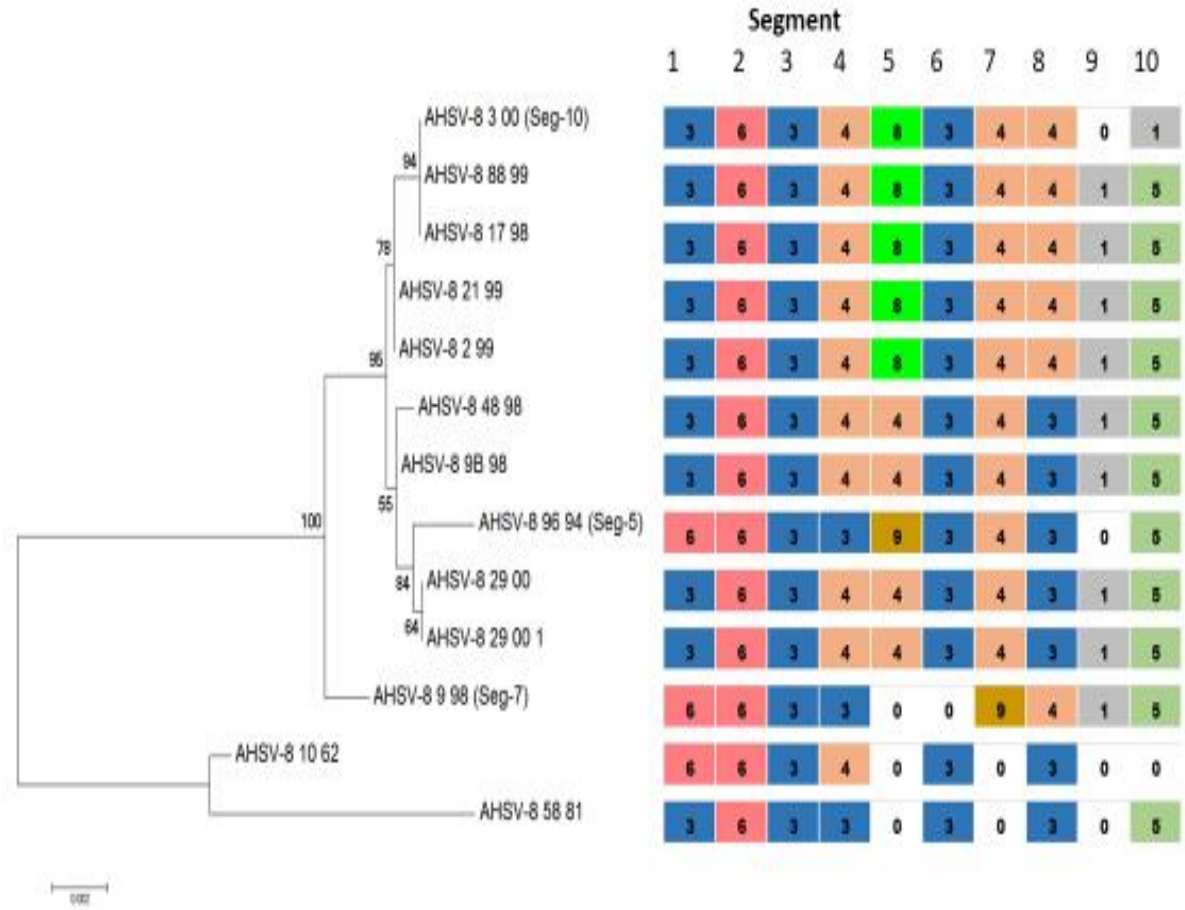

Figure 4.11. Molecular phylogenetic analysis of AHSV-8 isolates analysed using Maximum Likelihood method. See Figure 4.4.

4.4.1.2.9 Genetic reassortment in AHSV-9

The evolutionary analyses conducted in MEGA v6 using the ML method based on GTR nucleotide substitution model. Thirteen isolates, including OIE reference strain

AHSV-9_90_61, were used to analyse reassortment events involving AHSV-9 serotype (See Figure 4.12 below). The ML tree has shown one main branch comprising of 12 isolates and one outer group comprising of the OIE reference isolate. The Majority of the AHSV-9 isolates collected between the year 2001 and 2014 and shared nothing in common with the OIE reference strain in terms of clustering. The OIE reference isolate was collected in 1961 from Chad (Refer to Table 1, Appendix 1). Other isolates were collected within South Africa, except for AHSV-9_2033_14, which came from Mozambique. The geographic distance between the SA and other isolates used could have played a role in the reference isolates forming an outer branch. Two isolates were collected from the same animal at different body organs, one from the lung (AHSV-9_22L_03) and the other from the spleen (AHSV-9_22 S_03). Some of the genome segments associated with the two isolates were not similar. Reassortment events were recorded in Seg-4 and 9 of isolate AHSV-9_22S_03. Segment 4 of AHSV-9_22S_03 grouped into cluster 4 which was similar to AHSV-9_49_02 and could have exchanged segments after mixed infection with AHSV-8, AHSV-7, AHSV-6, AHSV-4, or AHSV-3. Seg-9 of AHSV-9_22S_03 was grouped into cluster 2, sequence identical to Seg-9 of AHSV-9_6_01. The two isolates may have acquired their Seg-9 after co-infection with AHSV-1 or AHSV-4. Among the nine distinct AHSV serotypes, AHSV-9 recorded a higher number of reassortment events totalling 12 involving seven isolates. AHSV-9_38-09 recorded four segments that may have undergone reassortment including Seg-1, 5, 7 and 8. Seg-1 of AHSV-9_38_09 belongs to cluster 7 and could have reassorted after mixed infection with AHSV-1, AHSV-4, AHSV-5, AHSV-6 or AHSV-7. Seg-5 of AHSV-9_38_09 could have been acquired after mixed infection with AHSV-2 or AHSV-4. Segment 7 (Seg-7), grouped into cluster 9 was also observed in AHSV-1, AHSV-4 and AHSV-AHSV-5 and the reassortment event associated with Seg-7 could have caused by co-infection of host cells with one of those serotypes listed above. Reassortment in Seg-1 was also observed in AHSV-9_27_08 and AHSV-9_2033_14 both grouped into cluster 7 and could have resulted from mixed infection with AHSV-1, AHSV-4, AHSV-5, AHSV-6 or AHSV-7. Segment 7 (Seg-7) of AHSV-9_27_08 grouped into cluster 5 and although the sequence clustered together with AHSV-5_35_09, there was no enough evidence to link its originality. Segment 7 of AHSV-9_27_08 may have emerged due to other evolutionary mechanisms or reassortment after co-infection with other AHSV isolates not included in this analysis.

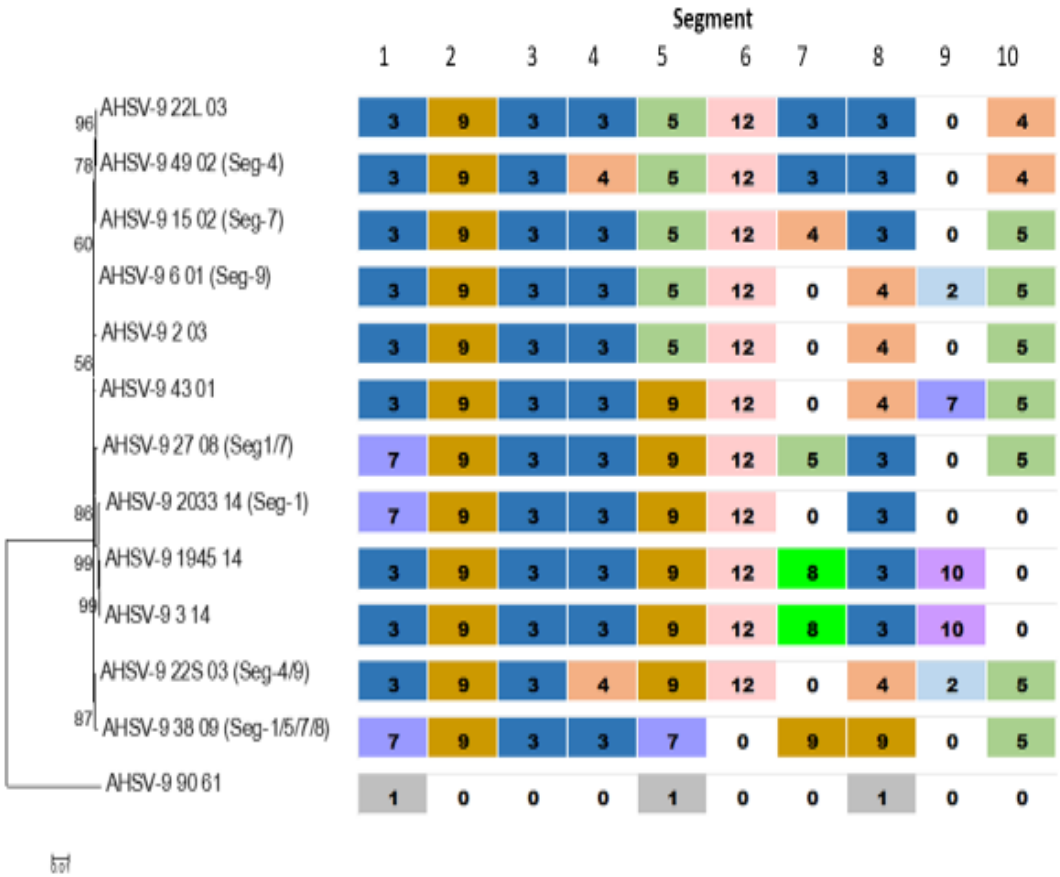

Figure 4.12 Molecular phylogenetic analysis of AHSV -9 isolates analysed using Maximum Likelihood method. See Figure 4.4.

4.4.2 Induced reassortment

The experimentally induced genome reassortment was carried out using two distinct AHSV serotypes (AHSV-4 and AHSV-9). The two serotypes allowed to co-infect the BSR cells and were later subjected to TBE-PAGE analysis to determine if genome reassortment has taken place.

4.4.2.1 Determining the dilution factors of each virus

The dilution factor of each virus required for co-infection was determined experimentally. The dilution factor was determined in order to prevent one serotype from overgrowing the other serotype. The lowest dilution factor for AHSV-4 was at 1 x 10^{-5} while AHSV-9 had the low plaque count at 1x 10^{3} dilution factor (See Figure 4.13). AHSV-4 seems to grow faster than AHSV-9 serotype.

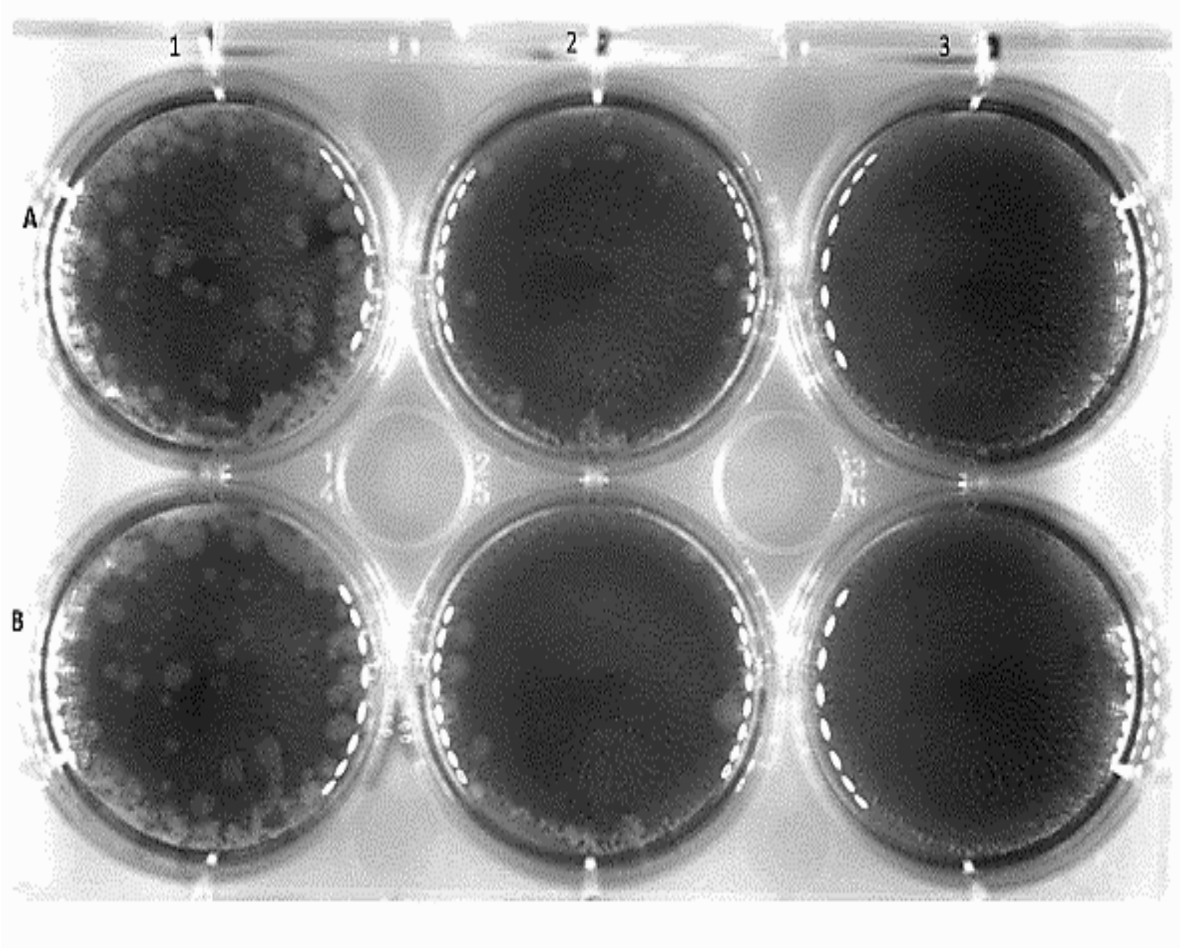

Figure 4.13. A 6-well plate showing AHSV-4 plaques at different dilution factors. The 6-well plate shows three different dilution factors starting from 1x 10^{-1} **to** 1x 10^{-3}, which were done in duplicates (A and B number 1 to 3). Virus plaques are presented as white spots on the wells.

4.4.2.2 Co-infection of two serologically unrelated AHSV serotypes into cell culture

Two different AHSV serotypes were chosen based on variations on their genomic electropherotypic profiles. AHSV-4 (Vac-4/ AHSV-4_32_62) was a South African strain used, while AHSV-9 (AHSV-9_90_61) was collected from Chad. After determining the multiplicity of infection for the two viruses, they were co-infected and passaged four times on BSR cells. The supernatants of the fourth passage were used to infect BSR cells in 6-well plates in order to pick individual plaques through agar overlay and re-infect BSR cells longer flasks (See Figure 4.14).

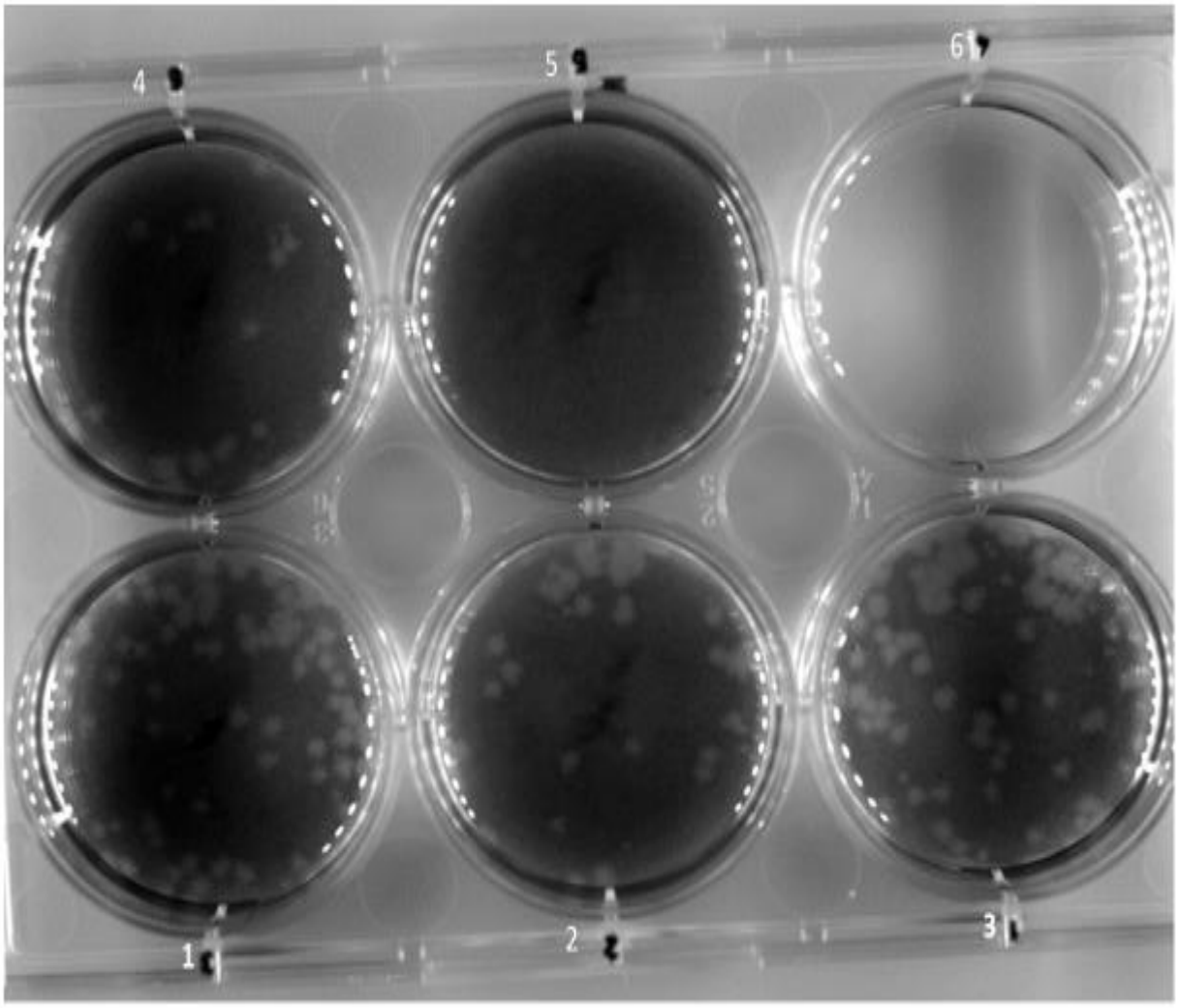

Figure 4.14. An illustration of individual plaques on the agar medium after co-infection of BSR cells with AHSV-4 and AHSV-9 serotype. Number of plaques formed decrease from well 1 to well 4, while well 5 shows no growth. Cells were not cultured on well 6. Virus plaques are presented as white spots on the wells.

4.4.2.3 Genome segment identification by polyacrylamide gel electrophoresis (TBE-PAGE).

The individual plaques obtained from co-infecting both isolates, were re-cultured, RNA extracted and latter subjected to TBE-PAGE. The gels stained with EtBr and analysis indicated that all segments from the plaques resembled the electrophoretic patterns of AHSV-9 (AHSV-9_90_61) (See Figure 4.15). The co-infection experiments were repeated and in contrast, all the genomic segments of the individual plaques had electrophoretic patterns resembling AHSV-4 (Vac-4/ AHSV-4_32_62) (Figure 4.16).

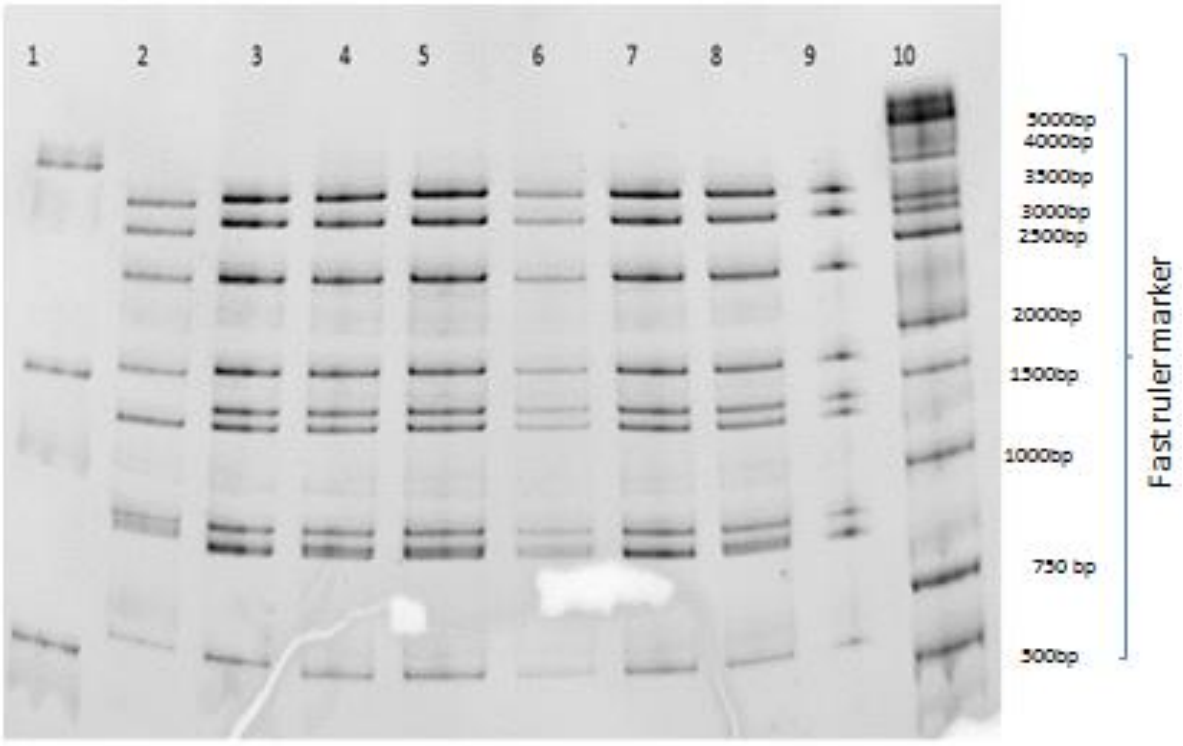

Figure 4.15. TBE-Polyacrylamide gel showing the electropherotypic patterns of AHSV-4 and AHSV-7 after co-infection of BSR cells. (Lane 1, Medium range DNA ladder; lane 2, + control AHSV-4; lane 3, + control AHSV-9; lane 4- 9 segments of AHSV-4 and AHSV9 co-infection; lane 10, Fast Ruler 1kb DNA ladder (Invitrogen).

Further observations of TBE-PAGE gels have shown some of the virus segments with an additional faint fragment co-migrating with segment 10. It is not clear what caused this extra fragment, because the NS3 genome segments for both the AHSV-4 and AHSV-9 were identical in length (See figure 4.16). No further analyses were done to investigate and identify the two different segments that were both associated with Seg-10. And therefore, it is unclear which evolutionary mechanism has caused co-migration in Seg-10.

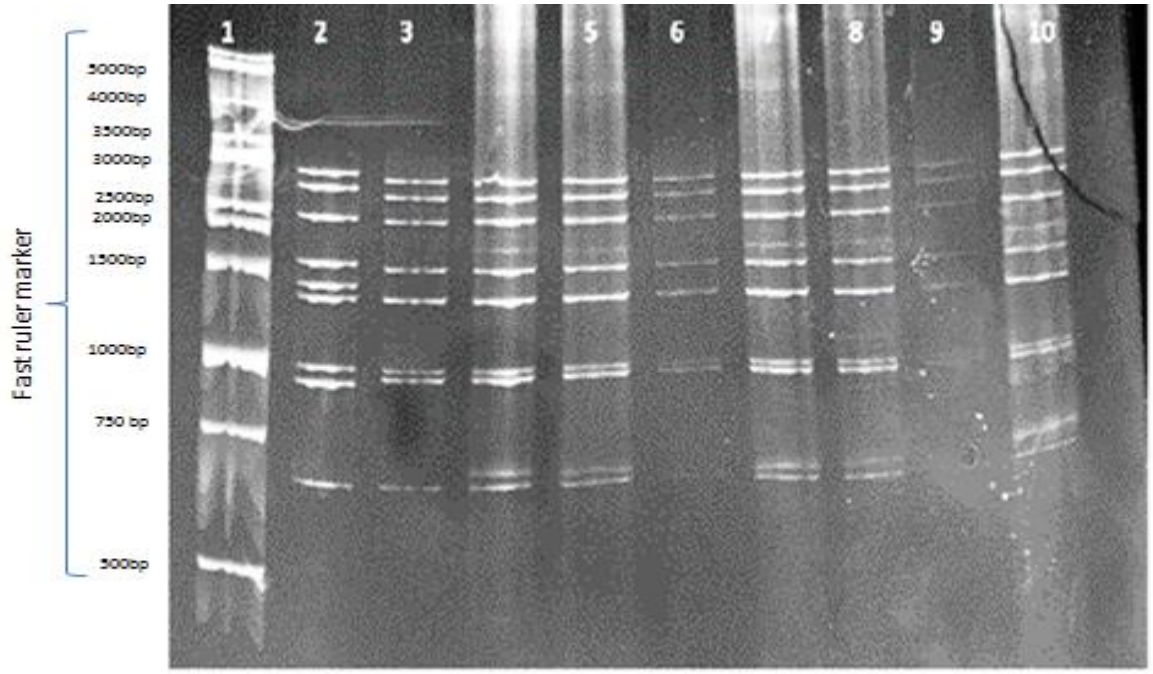

Figure 4.16 Electropherotypic patterns of AHSV-4 and AHSV-9 on TBE-PAGE gel. (Lane 1, 1kbp fast-ruler marker (Invitrogen); lane 2, +control AHSV-9; lane 3 + control AHSV-4; lane 4-10 dsRNA segments resembling AHSV-4 after co-infection with AHSV-4 and AHSV-9).

Based on the electropherotypic arrangements, it was difficult to conclude if reassortment events have taken place since the first co-infection experiments have shown the electropherotypic patterns of AHSV-9 while the second experiments resembled electropherotypic patterns of AHSV-4. The failure to observe clear genomic reassortment between segments of AHSV-4 and AHSV-9 dsRNA led to testing the identification of the VP2 on the individual plaques using the VP2 specific primers. The individual plaques of co-infected cells were extracted and subjected to RT-PCR using VP2 specific primers for AHSV-4 and AHSV-9. The RT-PCR analysed on an agarose gel and several plaques contained the VP2 segments for both AHSV-4 and AHSV-9 (See Figure 4.17). Although no reassortants were identified after co-infection of AHSV-4 and AHSV-9, the two virus isolates did co-infect the BSR cells and were identified in single plaques. Since the determination of reassortment were observed by analysing the migration pattern of the different genome segments, a more conclusive, yet the expensive method would have been performing serotype specific RT-PCR of each individual genome segment.

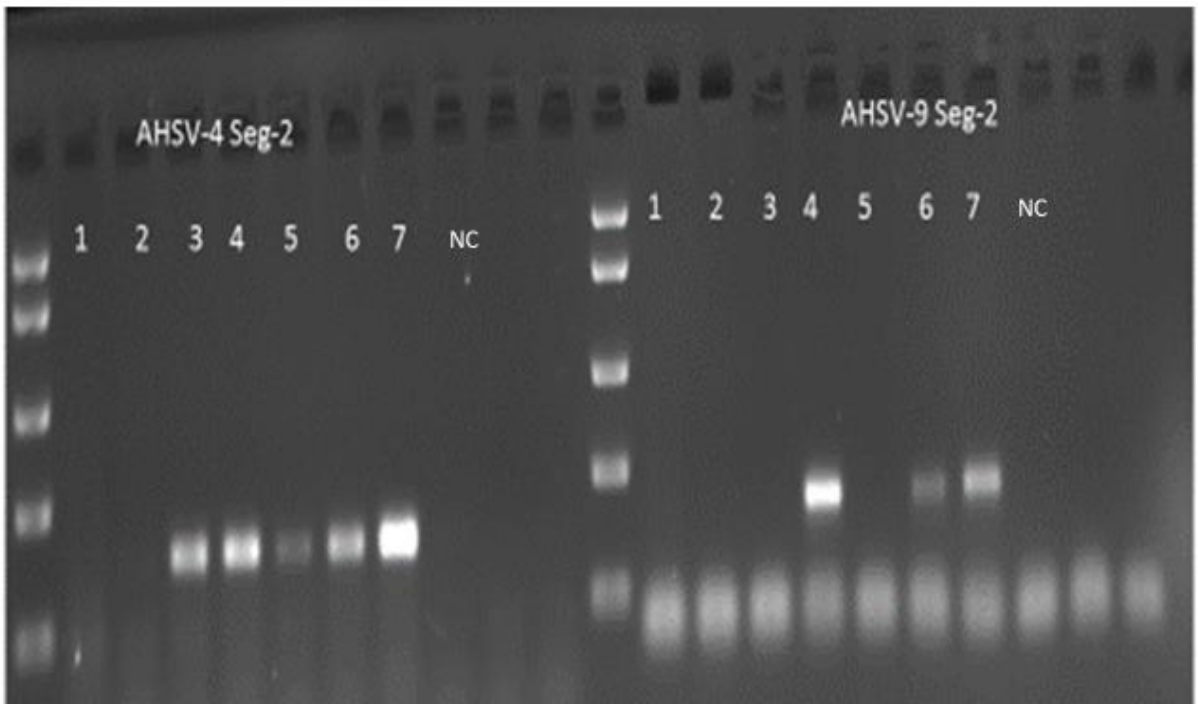

Figure 4.17. Agarose gel showing VP 2 segments of A AHSV-4 and B AHSV-9 separated by Medium Range DNA ladder (Invitrogen). (PCR products for each reassortant were run concurrently, NC represents negative control).

4.5 DISCUSSION

The analysis of naturally occurring genomic reassortment utilised complete sequence data of 101 AHSV isolates representing all the nine distinct serotypes collected from 1933 until 2014. Reassortment of AHSV genome segments has been observed previously (von Teichman and Smit, 2008; Weyer *et al*., 2016). The reassortment events recorded in 39/101 (38.6%) is an indication that genomic reassortment is common among AHSV serotypes and is a major evolutionary drive among AHSV serotypes. Similar to the study by Nomikou and co-workers (2015) on the evolution of BTV, genomic reassortment has shown in this study to be a widespread phenomenon in AHSV. Based on the ML phylogenetic tree, the sequences with less than 5% clustered together. It therefore follows that genetic sequences associated with the most variable proteins such as VP2 and VP5 (Seg-2 and Seg-6, respectively) had many clusters; however, Seg-2 had less number of tips not in clusters. Although Seg-2 had 12 clusters, there was no evidence of a formation of new serotypes and, probably, these clusters formed due to other evolutionary mechanisms such as mutations. This chapter did not look at variations

at the protein level; however, the variations associated with each segment are recorded in Chapter 2 Table 2.1. Furthermore, this study did not identify any reassortment event associated with Seg-2 and Seg-3. As expected, clusters that were picked from Seg-2 were only associated with particular serotypes and could not be shared among different serotypes. The finding was important in confirming the VP2, encoded by Seg-2, as serotype specific protein (Mertens *et al.,* 1987; Huismans and Van Dijk, 1990). Segment 2 (Seg-2) of seven AHSV serotypes had one cluster each. In contrast to the seven AHSV serotypes, Seg-2 of AHSV-4 and AHSV-7 had three and two clusters, respectively. The impacts of having two or more clusters in Seg-2 of AHSV-4 and AHSV-7 could not be defined in this study; however, different clusters did not seem to cause major changes in protein level (Table 2.1, chapter 2). Although there was no reassortment event associated with Seg-3, the OIE reference sequences of at least three serotypes (AHSV-1, 4 and 7) collected in 1962 that were used as vaccine strains fell into cluster 2. Such clustering was also shared by isolates identified as vaccine derived strains such as AHSV-1_107_09, AHSV-1 22_01, AHSV-4_90_01 and the other four isolates that formed second main branch within the AHSV-4 ML tree.

The ML phylogenetic tree has revealed a close association between AHSV-1 and 2, 3 and 7, 6 and 9, and 5 and 8 serotypes respectively. According to Coetzer and Guthrie (2004), there is antibody cross-reactivity between the AHSV-1 and 2, 3 and 7, 6 and 9, and 5 and 8, while AHSV-4 antibodies do not cross-react with other serotypes. Although this study did not include immunological analysis, the observations based on the AHSV ML tree clustered the serotypes that are known to cross-react in the same sub-branches, but separate them into two small branches according to their serotype-specific antigen VP2 genomic segment 2 (Seg-2) variations. The cross-reaction between the AHSV-1 and 2, 3 and 7, 6 and 9, and 5 and 8 could be the main reason why each of those serotypes had a single cluster, except for the AHSV-7 which had two different clusters. It is not known if all of these different clusters associated with AHSV-7 can cross-react with serotype 3 and the role that can be played by the Seg-2 tips that were not in the clusters. The AHSV-4 formed a ‘stand-alone’ branch, however, there were three clusters associated with its VP2 genome segment. These three different Seg-2 clusters of the AHSV-4 can serve as an evidence of high evolutionary rates associated with AHSV-4. Although

the Seg-2 of AHSV-4 had a higher number of clusters, that did not translate into the formation of new serotypes. The lack of cross-reactivity associated with AHSV-4 seems to be the reason why the VP2 genome segment (Seg-2) has a higher number of clusters. The effects of having more than one cluster in AHSV-4 and AHSV-7 on vaccine effectiveness as well as virulence were not investigated in this study but require further investigations. However, Burrage *et al* (1999) used three different AHSV-4 monoclonal antibodies (SA6, OH3, and ME11) and discovered that they strongly neutralized both the homologous and heterologous serotype 4 virus isolates but failed to cross-protect mice against AHSV-1-3 and 5-9 serotypes. However, these monoclonal antibodies varied significantly in the level of their protection. Genome reassortment events associated with AHSV-4 were relatively low at 40%. Instead, AHSV-1 had a very high number of reassortants, followed by AHSV-3 and AHSV-9 (71.4%, 63.6% and 53.8%, respectively). The AHSV-4, genomic evolution could, therefore, be attributed to different evolutionary mechanisms. The different clusters associated with Seg-2 of the AHSV-4 could have emerged as a result of mutation.

Within the putative vaccine isolates of AHSV-1_22_01, AHSV-1_107_09, AHSV-3_DG 25324_14, AHSV-3_DG 25327_14, AHSV-3_DG 25423_14, AHSV-4_90_96, AHSV-4_64_99 and AHSV-4_97_99, various reassortment events were observed. These findings are in agreement with the observations made by Weyer and co-workers (2016) who found that virulent revertants of AHSV-1 live attenuated vaccine as well as genomic reassortment derived from AHSV-1, 3 and 4 were responsible for African horse sickness outbreaks. The genomic segments of AHSV-1_22_01 and AHSV-1_107_09 have >99% sequence identity with AHSV-1 29 62, except for Seg-4, Seg-5 and Seg-9 suggesting they are vaccine derived isolates with reassorted segments. Segments 4 and 9 of AHSV-1_22_01 clustered with AHSV-4_32_62, while Seg-5 of both AHSV-1_22_01 and AHSV-1_107_09 clustered with both AHSV-2_82_61 and AHSV-3 13_63. Similarly, AHSV-3_DG 25324 14, AHSV-3_DG 25327_14 and AHSV- DG 25423 14 had >99% sequence identity to the reference strain AHSV-3_13_63. The Seg-1 of AHSV-3_DG 25324 14, AHSV-3_DG 25327_14 and AHSV- DG 25423 14 were similar to the group of AHSV-1_29_62, AHSV-2_82_61, AHSV-4_32_62 and AHSV-7_31_62.

Although genomic reassortment is flexible and can take place in any of the viral RNA segments (Shaw *et al*., 2013), there was no evidence of genome reassortment involving Seg-2 and Seg-3 in all 101 isolates. However, this does not rule out the fact that genome reassortment can occur between different AHSV serotypes and can affect any segment. Genomic segment 3 (Seg-3) and its viral protein (VP3) are highly conserved between the different *Orbivirus* serogroups (Roy *et al*., 1994). The high conservation of Seg-3 can be the main reason why there were only five clusters across the nine AHSV serotypes used. The analysis of Seg-3 sequence identity at protein level has shown a 99.34- 100% as well as 93.95-100 % at the nucleotide level, further confirming its protein as the most conserved. Although Seg-7, which encodes VP7, is highly conserved across the nine AHSV serotypes, at least 11 clusters were observed, including 33 numbers of tips that were not in the clusters. In chapter 2, genome segment 7 (Seg-7) was found to be among the segments with high substitution rates. Many clusters observed, could have been resulted from the high substitution rate in Seg-7, which translated into synonymous substitution during mutation. This further supported by a low mean d_N/d_S (95% CI) rate of 0.021 (0.005-0.047) which favoured the negative selection. Evidently, the sequence identity at the nucleotide level for Seg-7 was 87.5-100 %, while the sequence identity at protein level was 97.71-100 %. The high percentage identity at the protein level is an indication of high substitution rates that have translated into synonymous substitutions. The synonymous substitution is the main reason why there were many clusters associated with Seg-7, despite the highly conserved viral proteins.

High numbers of reassortment events were recorded in AHSV-1, 3 and 9, while AHSV-4 and 5 were moderately high. There was a low rate of genomic reassortment associated with AHSV-2 followed by AHSV-6, 7 and 8. Despite the different rate of reassortment events, reassortment was recorded across all AHSV serotypes. Collectively, genomic reassortment was seen as one of the main mechanism for AHSV evolution since it was widespread and slightly higher compared to what was reported in BTV (Nomikou *et al*., 2015). One of the important observations made during this analysis was on the isolates AHSV-9_22L_03 and AHSV-9_22S_03, isolated from different body organs (lung and spleen) of the same horse. Despite both being AHSV-9 serotypes, the isolate collected from the spleen (AHSV-9_22S_03) had two reassortment events observed, involving segment 4 and 9, while

its segment 8 did not belong to any cluster. It is possible that there were other AHSV serotypes that have reassorted with serotype 9 after co-infection or the two serotypes co-infected the host without reassorting with any serotype. This can also be assumed to be a direct evidence of genomic reassortment after mixed infection of the vertebrate host. The pathogenic nature of all the isolates that were used for reassortment analysis was not provided, and therefore it is not known whether the vaccine isolates that have shown evidence of genomic reassortment were pathogenic or not. However, a similar analysis of vaccine isolates by von Teichman and Smit (2008) did not find any indication of pathogenicity, death, virulence or reversion to virulence among the vaccine reassortants. Contrary to von Teichman and Smit (2008), a large-scale study involving 39 field isolates by Weyer and colleagues (2016) using virus isolates collected between 2004 and 2014, found that genomic reassortment associated with vaccine derived strains was the main contributor to some of the AHS outbreaks. Although the AHSV isolates used in this study did not have information regarding the pathogenesis, some of the field isolates were noted as vaccine derived isolates, the findings that are concerning and such require further investigations.

Apart from naturally occurring genomic reassortment, an experimental induced genomic reassortment was conducted. Differences were observed in plaque formation in relation to dilution factors between the two viruses (Vaccine AHSV-4 and wild-type AHSV-9). The AHSV-4 seems to grow faster as indicated by the low plaque formation at 10^{-5} compared to the wild-type AHSV-9, which had low plaque formations at 10^{-3}. The induced reassortment experiment has failed to pick any visible reassortants, although VP2 genome segments for both AHSV-4 and AHSV-9 co infected the BSR cells as confirmed by RT-PCR. It is possible that the two viruses (AHSV-4 and AHSV-9) co-infected same cells, but failed to reassort during packaging or may have entered same cells but replication took place at different time intervals. After using specific VP2 primers the detection of AHSV-4 and AHSV-7 serotypes, respectively, some of the plaques had PCR products for both AHSV-4 and AHSV-9. It is possible that there was a co-circulation of the two viruses, but one virus was able to suppress the other. For instance, during the first experiments, AHSV-9 was detected by TBE-PAGE without detecting AHSV-4, except for the positive control. During the second experiment, only AHSV-4 was detected (Except

for positive controls for both AHSV-4 and AHSV-9). Genomic reassortment may have occurred on segments that are equal in length and could not be easily identified as reassortment based on electropherotypic arrangements. As such further, confirmation by sequencing was required. In their induced reassortment study, von Teichman and Smit (2008) inoculated horses with vaccine strains and successfully detected genome reassortants. In contrast to von Teichman and Smit (2008), this study only used BSR cell lines to induce reassortment and could not observe any reassortment event. Shaw *et al*., (2013) used reverse transcription PCR as well as real-time RT-PCR to characterise BTV reassortants between BTV1 and BTV8. Based on Shaw *et al*., (2013)'s findings, it is clear that this study could have picked reassortment if the plaques were not just subjected to TBE-PAGE analysis but also sequencing. While Shaw *et al*., (2013) have used CPT-Tert cells and could pick VP2 segments for BTV1 and BTV8, this study only utilised BSR cells, and however, the choice of a cell line may not be used as a determining factor for the occurrence genome reassortment occurred. Due to the fact that after co-infection of BSR cells with the vaccine AHSV-4 and wild-type AHSV-9 isolates, the RT-PCR using specific primers managed to detect both the VP2 genome segments in some of the plaques. The two distinct serotypes managed to co-enter the same cells, but could not show evidence of genomic reassortment when subjected to TBE-PAGE analysis.

CHAPTER 5 CONCLUDING REMARKS

The main objective of this study was to investigate the evolutionary dynamics of African horse sickness virus in order to determine how the virus has evolved over the years. To achieve this goal, isolates collected between 1933 and the year 2014 were used as the basis for the evolutionary analysis of AHSV. These included mainly the isolates collected in the early 1960s at the time AHSV live attenuated vaccine was introduced up to the year 2014. Where it was required, historical AHSV strains used as vaccine strains or reference strains were either utilised or omitted. The utilisation and/or omission also included vaccine derived isolates. In order to successfully investigate AHSV evolutionary dynamics, this study focused attention on mechanisms of genetic variation that are due to mutations, specifically nucleotide substitutions as the first objective. The second objective was to investigate the role of intragenic recombination in the evolution of AHSV. The last objective of the study was to investigate the role of genomic reassortment on the AHSV evolution, which focused mainly on how the virus may have reassorted in the field. To a less extent, this study also investigated genomic reassortment as a result of mixed infection of cell cultures with two distinct AHSV serotypes.

Analyses of different AHSV serotypes have confirmed the VP2 as the most variable protein, 49.06 – 100 % between nine different serotypes (Mertens *et al*., 1987; Huismans and Van Dijk, 1990). The NS3, which is a cytotoxic protein playing a role in cell membrane alteration and viral exist was confirmed as the second most variable protein (Mertens *et al*., 1987; Huismans and Van Dijk, 1990; Martin *et al*., 1998; van Niekerk *et al*., 2001a). Although the VP2 protein sequences have shown a major variation between different serotypes, comparison within individual serotypes has shown high similarities and most conserved sequences. On the other hand, VP3 and VP7 were confirmed as highly conserved proteins in all distinct AHSV serotypes (Bremer *et al*., 1990; Roy *et al*., 1994; Mertens *et al*., 2006). At the nucleotide level, segment 7 was slightly variable (87.15 – 100%) and this was also confirmed by the Cluster Picker software that has detected 11 clusters, however, the differences did not translate into high sequence variation at the protein level. The differences in clusters can be due to high synonymous substitution rates associated with genome segment 7 (Seg-7). Other genome segments that had high substitution rates includes segment 2, 6 and 9. These three segments (Seg-2, 6 and 9) also recorded

high numbers of clusters and a number of tips not in the clusters. Unlike Seg-7, the proteins associated with these three segments (Seg-2, 6 and 9) were variable. While genome segments such as Segment 1 and 7 have indicated substantial genetic variations, their proteins were highly conserved. The high percentage identity on protein level in segments such as Seg-1, Seg-3, Seg-4, Seg-5, Seg-7, Seg-8 as well as Seg-2 (VP2) within a serotype is a reflection of strong selective constraints imposed on AHSV and other arboviruses (de Mattos *et al*., 1996; Novella *et al.,* 1999; Anbalagan *et al*., 2014). Like other arboviruses, *Orbivirus* genus requires an insect host for transmission between vertebrate hosts (Jenkins, 2002). As a result of these constraints, the AHSV evolve very slowly (Combe and Sajuán, 2014) and therefore, the overall substitution rates of AHSV favours synonymous over nonsynonymous substitutions and therefore AHSV undergo negative selection.

Apart from genetic mutations observed in this study, intragenic recombination events were also observed, although was less abundant than in BTV (He *et al*., 2010). The majority of intragenic recombination events observed in an AHSV isolates occurred due to genetic exchange between live attenuated vaccine viruses and field isolates. Apart from high substitution rates that favour synonymous substitution, genome reassortment seems to be the major drive in AHSV evolution. Genome reassortment events were recorded across all nine AHSV serotypes, but more abundant in serotype 1, 3 and 9. The AHSV -4 was more diverse than the other eight serotypes; however had less number of reassortant isolates when compared with serotype 1, 3 and 9. While it is believed that genome reassortment can affect any segment (Shaw *et al*., 2013), there was no evidence of reassortment involving AHSV segment 2 and 3. The AHSV segment 2 is the most variable gene, coding for the most variable VP2, while segment 3 is the most conserved gene coding for the most conserved VP3.

In conclusion, AHSV evolves slowly due to constraints imposed by its ability to successfully replicate in both the insect vector and vertebrate host. All the AHSV genome segments have shown to undergo a negative selection. Like other RNA viruses, AHSV lack proof-reading mechanism and therefore undergo a genetic mutation that does not translate into major changes in the protein level due to host constraints. Intragenic recombination due occurs, but less abundant than in BTV. A widespread genome reassortment affecting all the serotypes and involving nearly all

the genomic segments has been observed and necessitates a further investigation into its role in the pathogenesis of the virus and vaccine effectiveness.

Congress contributions

H.G. Ngoveni, O.J.J. Koekemoer and A. van Schalkwyk. 2017. **African horse sickness virus** evolutionary **dynamics.** 9th SA Veterinary and Paraveterinary and SASVEPM Congress that will be held at the Birchwood Hotel and OR Tambo Conference Centre from 24-27 July 2017.

Mokotoane, R., Koekemoer, O., Ngoveni, H., van Schalkwyk, A. and Maree, S., **Complete genome comparison of nine recent and historic field isolates of African horse sickness virus.** 12th International dsRNA Virus Symposium 2015, 6-10 October, Goa, India.

CHAPTER 6 REFERENCES

Agüero, M., Gomez-Tejedor, C., Angeles Cubillo, M., Rubio, C., Romero, E. and Jimenez-Clavero, A. (2008). Real-time fluorogenic reverse transcription polymerase chain reaction assay for detection of African horse sickness virus. *J. Vet. Diagn*. *Invest*., **20**: 325–328.

Allison, A.B., Holmes, E.C., Potgieter, A.C., Wright. I.M., Sailleau, C., Breard, M.G. and Stallknecht., D.E. (2012). Segmental configuration and putative origin of the reassortants orbivirus, epizootic hemorrhagic disease virus serotype 6, strain Indiana. *Virol.*, **424**: 67-75.

Anbalagan, S., Cooper E., Klumper P., Simonson RR., Hause BM. (2014). Whole genome analysis of epizootic hemorrhagic disease virus identified limited genome constellations and preferential reassortment. *J. Gen. Virol.*, **95**: 434–441.

Anthony SJ, Darpel KE, Maan S, Sutton G, Attoui H, Mertens PPC. (2010). The evolution of two homologues of the core protein VP6 of epizootic haemorrhagic disease virus (EHDV), which correspond to the geographical origin of the virus. *Virus Genes*, **40**:67–75.

Arenas, M. (2015). Trends in substitution models of molecular evolution. *Front. Genet.* **6**(319):1-9.

Attoui, H., Mohd Jaafar, F., Belhouchet, M., Aldrovandi, N., Tao, S., Chen, B., Liang, G., Tesh, R.B., de Micco, P. and de Lamballerie, X. (2005). *Yunnan Orbivirus*, a new orbivirus species isolated from *Culex tritaeniorhynchus* mosquitoes in China. *J. Gen. Virol.*, **86**: 3409-3417.

Attoui, H., Mertens, P.P.C., Becnel, J., Belaganahilli, S., Bergio, M., Brussaard, C.P., Chappell, J.D., Ciarlet, M., del Vas, M. Dermody, T.S., Dormitzer, P.R., Duncan, R., Fang, Q., Graham, R., Guglielmi, K.M., Harding, R.M., Hillman, B., Makkay, A., Marzachì, C., Matthijnssens, J., Milne, R.G., Mohd Jaafar, F., Mori, H., Noordeloos, A.A., Omura, T., Patton, J.T., Rao, S., Maan, M., Stoltz, D., Suzuki, N., Upadhyaya, N.M., Wei, C. and Zhou, H. (2012). Family Reoviridae. In Virus Taxonomy: Ninth report of the International Committee on Taxonomy of viruses, pp, Ed: King, A,N.Q.,

Adams, M.J., Carstens, E.B. and Lefkowitz, E.J. Amsterdam: Elsevier Academic Press. Pp 541-637.

Backer, J.A. and Nodelijk, G. (2011). Transmission and control of African horse sickness in The Netherlands: A model analysis. *PloS One*, **6**(8):1-12.

Baldauf, S.L. (2003). Phylogeny for the faint heart: a tutorial. *Trends Genet*., **19**: 345-351.

Barnard, B.J.H. and Paweska, J.T. (1993). Prevalence of antibodies against some equine viruses in zebra (*Equus burchellli*) in the Kruger National Park, 1991-1992. *Onderstepoort J. Vet. Res.*, **60**: 175-179.

Basak, A. K., Gouet, P., Grimes, J., Roy and P., Stuart, D. (1996). Crystal structure of the top domain of African horse sickness virus VP7: comparisons with bluetongue virus VP7. *J. Virol.*, ***70***(6): 3797-3806.

Basak, A K., Grimes, J. M., Gouet, P., Roy, P., and Stuart, D. I. (1997). Structure of orbivirus VP7: implications for the role of this protein in the viral life cycle. *Structure*, **5**(7), 871–83.

Bonneau, K.R., Mullens, B.A. and MacLachlan, N.J. (2001). Occurrence of genetic drift and founder effect during quasispecies evolution of VP2 and NS3/NS3A genes of bluetongue virus upon passage between sheep, cattle, and Culicoides sonorensis. *J. Virol*., **75**(17): 8298-8305.

Boulila, M. (2011). Positive selection, molecular recombination structure and phylogenetic reconstruction of members of family Tombusviridae: Implication in virus taxonomy. Gene. *Mol. Biol.*, **34**(4): 647-660.

Boyle, D.B., Amos-Ritchie, R., Broz, I., Walker, P.J., Melville, L., Flanagan, D., Davis, S., Hunt, N. and Weir, R. (2014). Evolution of bluetongue virus serotype 1 in Northern Australia over 30 years. *J. Virol*., **88**(24): 13981-13989.

Boyce, M. and Roy, P. (2007). Recovery of infectious bluetongue virus from RNA. *J. Virol*., **81**(5): 2179-2186.

Bremer, C. W. (1976). A gel electrophoretic study of the protein and nucleic acid components of African horsesickness virus. *Onderstepoort J. Vet. Res*., **43**(4): 193-199.

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

Bremer, C.W., Dungu-Kimbenga, B. and Viljoen, G.J. (1998). Detection of African horsesickness virus in Zebra by RT-PCR and the development of different methods for confirming AHSV specificity of RT-PCR products. Proceedings of the Eighth International Conference on Equine Infectious Diseases, Dubai, 23–26 March 1998. R and W Publications (Newmarket) Ltd, Newmarket, UK.

Burrage, T.G., Trevejo, R., Stone-Marschat, M. and Laegreid, W.W. (1993). Neutralizing epitopes of African horsesickness virus serotye 4 are located on VP2. Virol., **196**(2): 799-803.

Carpi G, Holmes EC, Kitchen A. (2010). The evolutionary dynamics of Bluetongue virus. *J. Mol. Evol*., **70**: 583-592.

CEH Horse Report. (2009). African horse sickness in North America? Lessons learned from recent viral epidemics. *Centre of Equine Health. University of California*, **27**(3):1-12.

CFSPH (Center for Food Security and Public Health). African horse sickness. *www.cfsph.iastate.edu/Factsheets/pdfs/african_horse_sickness*. last updated: November, 2006.

Chare, E.R., Gould, E.A. and Holmes, E.C. (2003). Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. *J. Gen Virol*., **84**: 2691-2703.

Chong, C., Zhai, W., Li. C., Gao, M., Gong, Q., Ruan, J., Li, J., Jiang, L., Lv, X., Hungate, E. and Wu, C. (2013). The evolution of small insertions and deletions in the coding genes of Drosophila melanogaster. *Mol. Biol. Evol*., **30** (12):2699-2708.

Ciota, A.T. and Kramer, L.D. (2010), Insights into arbovirus evolution and adaptation from experimental studies. *Viruses*., **2**: 2594-2617.

Coetzer, J.A. and Guthrie, A.J, 2004. 'African Horse sickness'. In Coetzer J.A.W and Tustin R.C (Eds). Infectious diseases of livestock, 2nd ed., Oxford University Press Southern Africa, Cape Town, pp.1231-1246.

Combe, M. and Sanjuán, R. (2014). Variation in RNA virus mutation rates across host cells. *PLos Pathog*., **10**(1): 1-7.

Cooper, L.A. and Scott, T.W. (2001). Differential evolution of eastern equine encephalitis virus populations in response to host cell type. *Genet*., **157**: 1403-1412.

Delport, W., Poon, A.F.Y., Frost, S.D.W., and Pond, S.L.K (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*. **26**(19):2455-2457.

de Mattos, C.C., de Mattos, C.A., MacLachlan, N.J., Giavedoni, L.D., Yilma, T, Osburn, B.I. (1996). Phylogenetic comparison of the S3 gene of United States prototype strains of bluetongue virus with that of field isolates from California. *J. Virol*., **70**: 5735-5739.

Domingo, E. (2010). Mechanism of viral emergence. *Vet. Res*., **41**: 1-14.

Domingo, E. and Holland, J.J. (1997). RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol*., **51**:151-178.

Domingo, E., Sheldon, J. and Perales, C. (2012). Viral quasispecies evolution. *Microbiol. Mol. Biol. Rev*., **76**(2): 159-216.

Drummond, A.J., Suchard, M.A., Xie, D. and Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Soc. Mol. Biol. Evol*., **29**(8):1969-1973.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J. and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *Plos Biol.,* **4**(5): 699-710.

Du Plessis, M., Cloete, M., Aitchison, H., and Van Dijk, A.A. (1998). Protein aggregation complicates the development of baculovirus-expressed African

horsesickness virus serotype 5 VP2 subunit vaccines. *Onderstepoort. J. Vet. Res*., **65**(4): 321–332.

Du Toit, R.M. (1944). The transmission of blue-togue and horse-sickness by Culicoides. *Onderstepoort J. Vet. Sci. Anim. Ind*., **10**: 7-16.

Efron, B., Halloran, E. and Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sc*., **93**: 13429-13434.

Eigen, M. (1993). Viral quasispecies. *Scie. Ame*., **1**: 42-49.

Elena, S.F. and Sanjuán, R. (2007). Virus evolution: Insights from an experimental approach. *Annu. Rev. Ecol. Evol. Syst*., **38**: 27-52.

Erasmus, B.J. (1978). The pathogenesis of African horsesickness, In Bryans JT, Gerber H (ed), Proceedings of the Third International Conference on Equine Infectious Diseases. *Equine infect. dis*., vol III. S. Karger, Basel, Switzerland, p 1–11.

Fleischmann, R.W Jr. (1996). Viral genetics. Chapter 43. 4th ed, Baron. S (ed) *Med. Microbio*l., pp 1-12.

Froissart, R., Roze, D., Uzest, M., Galibert, L., Blanc, S. and Michalakis, Y. (2005). Recombination every day: abundant recombination in a virus during a single multi-cellular host infection. *PLOS Biol*., **3**: 389-395.

Gibbs, M.J., Armstrong, J.S., Gibbs A.J. (2000). Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics*, **16**:573-582.

Gogarten, J.P. and Townsend, J.P. (2005). Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol*., **3**:679- 687.

Gojobori, T., Yamaguchi, Y., Ikeo, K. and Mizokami, M. (1994). Evolution of pathogenic viruses with special reference to the rates of synonymous and nonsynonymous substitutions. *Jpn. J. Genet*., **69**: 481-488.

Gorman, B.M., Taylor, J., Walker, P.J. and Young, P.R. (1978). The isolation of recombinants between related orbiviruses. *J. Gen. Virol.*, **41**: 333-342.

Gould, A.R. and Hyatt, A.D. (1994). The Orbivirus genus. Diversity, structure, replication and phylogenetic relationships. *Comp. Immunol. Microbiol. Infect. Dis*., **17**: 163-188.

Greenbaum, B.D. and Ghedin, E. (2015). Viral evolution: beyond drift and shift. C*urrent Opin. Microbiol*., **26**: 109-115.

Greenbaum, B.D., Li, O.T.W., Poon, L.L.M., Levine, A.J. and Rabadan, R. (2012). Viral reassortment as an information exchange between viral segments. *Proc. Natl. Acad. Sci. U.S.A.,* **109**:3341–3346.

Guthrie, A.J. and Quan, M. (2009). African horse sickness. In: Infectious Diseases of the Horse, 4th edn., Eds: T.S. Mair and R.E. Hutchinson, Equine Vet. J.Ltd., Cambridgeshire. pp 72-82.

Guthrie, A.J., Coetzee, P., Martin, D.P., Lourens, C.W., Venter, E.H., Weyer, C.T., Joone, C., le Grange, M., Harper, C.K., Howell, P.G., MacLachlan, N.J. (2015). Complete genome sequences of the three African horse sickness virus strains from a commercial trivalent live attenuated vaccine. *Genome Announc*., **3**(4):e00814-15

Hamblin, C., Graham, S. D., Anderson, E. C., and Crowther, J. R. (1990). A competitive ELISA for the detection of group-specific antibodies to African horse sickness virus. *Epidemiol. Infect*., **104:** 303–312.

Han, G. and Worobey, M. (2011). Homologous recombination in negative sense RNA viruses. *Viruses*, **3**(8): 1358-1373.

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

He C.Q., Ding, N-Z., He, M., Li S-N., Wang, X-M., He, H-B., Liu X-F., and Guo H-S. (2010). Intragenic recombination as a mechanism of genetic diversity in bluetongue virus. *J. Virol*., **84**:11487-11495.

Heath, L., van der Walt, E., Varsani, A. and Martin, D.P.(2006). Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *J. Virol.*, **80**:11827-11832.

Henning, M.W. (1956). Animal diseases in South Africa.3rd ed. Cent. News Agen. Ltd. South Africa. pp 785-808.

Holmes, E.C. and Moya, A. (2002). Is the quasispecies concept relevant to RNA viruses? *J. Virol*., **76**:460-462.

House, C., Mikiciuk, P. E. and Berninger, M. L. (1990). Laboratory Diagnosis of African Horse Sickness: Comparison of Serological Techniques and Evaluation of Storage Methods of Samples for Virus Isolation. *J. Vet. Diag. Invest*., **2**: 44–50.

Howell, P.G. (1963). The isolation and identification of further antigenic types of African horsesickness. *Onderstepoort J. Vet. Res*., **29**: 139149.

Huismans, H., van Dijk, A.A., and Bauskin, A.R. (1987). In vitro phosphorylation and purification of a nonstructural protein of bluetongue virus with affinity for single-stranded RNA. *J. Virol*., **61**: 3589-3595.

Huismans, H., van Staden, V., Fick, W. C., van Niekerk, M., and Meiring, T. L. (2004). A comparison of different orbivirus proteins that could affect virulence and pathogenesis. *Vet. Ital*., **40**: 417–425.

Huismans, H. and van Dijk, A. A. (1990) Bluetongue virus structural components. *Current Topics in Microbiology and Immunology*, **162**: 21-42.

Isakov, O., Bordería, A., Golan, D., Hamenahem, A., Celniker, G., Yoffe, L.,Blanc, H., Vignuzzi, M. and Shomron, N. (2015). Deep sequencing analysis of viral infection and evolution allows rapid and detailed characterization of viral mutant spectrum. *Bioinformatics*, **31**:2141-2150.

Iwata, H., Yamagawa, M. and Roy, P. (1992). Evolutionary relationships among the gnat-transmitted orbiviruses that cause African horse sickness, bluetongue, and epizootic hemorrhagic disease as evidenced by their capsid protein sequences. *Virol*., **191**, 251-261.

Jenckel. M., Bréard, E., Schulz, C., Sailleau, C., Viarouge, C., Hoffmann, B., Höper, D., Beer M. and Zientara, S. (2015). Complete coding genome sequence of putative novel bluetongue virus serotype 27. *Genom. Announ*., **3**: 2.

Jenkins, G.M., Rambaut, A., Pybus, O.G. and Holmes, E.C. (2002). Rates of molecular evolution in RNA viruses: a qualitative phylogenetic analysis. J. Mol. Evol., **54**: 156-165.

Keeling, P.J. and Palmer, J.D. (2008). Horizontal gene transfer eukaryotic evolution. *Nat. Rev. Genet*., **9**:605-618.

Kirkwood, T.B and Bangham, C.R. (1994). Cycles, chaos, and evolution in virus cultures: a model of defective interfering particles. *Proc. Natl. Acad. Sci. U.S.A*, **91**(18): 8685-8689.

Koekemoer, J.J.O. (2008). Serotype -specific detection of African horsesickness virus by real-time PCR and the influence of genetic variations. *J. Virol. Methods*, **154**: 104-110.

Kwok, S., Kellogg, D.E., McKinney, N., Spasic, D., Goda, L., Levenson, C. and Sninsky, J.J. (1990). Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nucleic Acids Res*., **18**(4): 999-1005.

Lai, M.M.C. (1992). RNA recombination in animal and plant viruses. *Microbiol. Rev*., **56**: 61-79.

Lahon, A., Walimbe, A.M. and Chitambar, S.D.(2012). Full genome analysis of group B rotaviruses from western India: genetic relatedness and evolution. *J. Gen. Virol*., **93**:2252–2266.

Lewis-Rogers, N., Crandall, K.A. and Posada, D. (2004). Evolutionary analyses of genetic recombination. Dynam. *Genet*., pp 49-78.

Liò, P. and Goldman, N. (1998). Models of molecular evolution and phylogeny. *Genome Res*., **8**:1233-1244.

Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari. D., Kulkarni, S.S., Novak, N.G.,Ingersoll, R., Sheppard, H.W. and Ray, S.C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.*, **73**: 152-160.

Losos, G.j., (1986). African horsesickness. In: Infectious tropical diseases of domestic animals. Longman Scientific and Technical, Essex England. Churchhil, Livingstone Inc New York. pp. 349-375.

Lucas, M., Karrer, U., Lucas, A. and Klenerman, P. (2001). Viral escape mechanisms-escapology taught by viruses. *Int. J. Exp. Pathol.*, **82**: 269-286.

Maan, N.S, Belaganahalli, M.N. and Mertens, P.P.S. (2013). Whole genome sequencing strategies and development of Orbivirus sequence database: Implications for novel dsRNA virus detection, in Salar, R.K., Gahlawat, S.K., Siwach, P. and Duhan, J.S. Biotechnology: Prospects and application (eds), Springer Science and Business Media. Scienc., pp. 237-255.

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

Makadiya, N. (2007). Cosequences of genetic variation and selection in viruses. Student perspective on contemporary virology. Consequen. Gen. Evol., pp 1-18.

Manole, V., Laurinmäki, P., Van Wyngaardt, W., Potgieter, C. a, Wright, I. M., Venter, G. J., van Dijk, A.A., Sewell, B.T. and Butcher, S. J. (2012). Structural insight into African horsesickness virus infection. *J. Virol*., **86**(15), 7858–7866.

Manrubia,S.C. and Lázaro, E. (2006). Viral evolution. *Physics. Life Rev*., **3**: 65-92.

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

Martin, D,P., Lemey, P, Lott, M., Moulton, V., Posada, D. and Lefeuvre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*, **26**:2462-2463.

Martin, P.D., Murrell, B., Golden, M., Khoosal, A. and Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol*., **1**(1): 1-5.

Martin, L.-A., Meyer, A. J., O'Hara, R. S., Fu, H., Knowles, N. J. and Mertens, P. P. C. (1998). Phylogenetic analysis of African horse sickness virus genome segment 10: sequence variation and virulence characteristics. *Arch. Virol. suppl*., **14**:281-293.

Martin, D.P., Posada, D., Crandall, K.A. and Williamson, C. (2005). A modified BOOTSCAN algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS. Res. Hum. Retroviruses*, **21**: 98-102.

Martin, D.P., Williamson, C. and Posada, D. (2005). rdp2: recombination deletion and analysis from sequence alignments. *Bioinformatics*., **21**(2): 260-262.

Martinez-Torrecuandrada, J.L.,Iwata, H., Venteo, A., Casal, J.I. and Roy, P. (1994). Expression and characterization of the two outer capsid proteins of African horse sickness virus: the role of VP2 in virus neutralization. *Virol.*, **202**: 348-359.

Matsuo, E., Celma, C. C. P. and Roy, P. (2010). A reverse genetics system of African horse sickness virus reveals existence of primary replication. *FEBS letters*, **584**(15), 3386–3391.

Maynard Smith J. 1992. Analyzing the mosaic structure of genes. *J. Mol. Evol*., **34**:126-129.

McIntosh, B.M. (1958). Immunological types of horsesickness virus and their significance in immunization. *Onderstepoort J. Vet. Res.*, **27**(4): 465-538.

Meiring, T.L., Huismans, H. and van Staden, V. (2009). Genome segment reassortment identifies non-structural protein NS3 as a key protein in African horsesickness virus release and alteration of membrane permeability. *Arch. Virol.*, **154**:263–271.

Meiswinkel, R. and Paweska, J.T. (2003). Evidence for a new field Culicoides vector of African horse sickness in South Africa. *Prev. Vet. Med*., **60**: 243-253.

Mellor, P.S. (2000). Replication of arboviruses in insect vectors. *J. Comp. Path.*, **123**: 231-247.

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

Mellor P.S., Boorman, J. and Baylis, M. (2000). Culicoides biting midges: their role as arbovirus vectors. *Annu. Rev. Entomol*., **45**:307–340.

Mellor, P.S., Boorman, J. and Jennings, J. (1975). The multiplication of African horse sickness virus in two species of Culicoides (Diptera, Ceratopogonidae). *Arch. Virol*., **47**(4):351-356.

Mertens, P.P.C. (2000). Orbiviruses and coltiviruses—general features. In:Webster R.G., Granoff A. (Eds.), Encyclopedia of Virology. Academic Press, London.

Mertens, P.P.C., Attoui, H. and Bamforf, D.H. (2006). The RNAs and proteins of dsRNA viruses: African horse sickness virus. www.reoviridae.org/dsrna_virus_proteins/AHSV. HTPM. Last updated on the 20th May 2006.

Mertens, P. P. C., Burroughs, J. N. and Anderson, J. (1987). Purification and properties of virus particles, infectious subviral particles, and cores of bluetongue virus serotype 1 and 4. *Virol*., **157**: 375-386.

Mertens, P. P. C., Burroughs, J. N., Walton, A., Welby, M. P., Fu, H., O'Hara, R. S., Brookes, S. M. and Mellor, P. S. (1995). Enhanced infectivities of modified bluetongue virus for two insect cell lines for two Culicoides vector species. *Virol*., **217**: 582-593.

Mertens, P.P.C., Diprose, J., Maan, S., Singh, K.P., Attoui, H. and Samuel, A.R. (2004). Bluetongue virus replication, molecular and structural biology. *Vet. Ital*., **40**(4): 426-437.

Mertens, P. P. C., Pedley, S., Cowley, J. and Burroughs, J. N. (1987). A comparison of six different bluetongue virus isolates by crosshybridisation of the dsRNA genome segments. *Virol*., **161**: 438-447.

Mertens, P. P. C., Pedley, S., Cowley, J., Burroughs, J. N., Corteyn, A. H., Jeggo, M. H., Jennings, A. M. and Gorman, B. M.(1989). Analysis of the roles of bluetongue virus outer capsid proteins VP2 and VP5 in determination of virus serotype. *Virol*., **170**: 561-565.

Mirchamsy, H. and Hazrati, A. (1973). A review on aetiology and pathogeny of African horsesickness. *Arc. Inst. Razi*., **25**: 23-46.

Modrof, J., Lymperopoulos, K. and Roy, P. (2005). Phosphorylation of bluetongue virus nonstructural protein 2 is essential for formation of viral inclusion bodies. *J. Virol*., **79**(15): 10024-10031.

Mohd Jaafar, F., Belhouchet, M., Belaganahalli, M., Tesh, R.B., Mertens, P.P.C.and Attoui, H. (2014). Full-genome characterisation of Orungo, Lebombo and Changuinola viruses provides evidence for co-evolution of orbiviruses with their arthropod vectors. *PLoS ONE* **9**(1): e86392.

Monastyrskaya, K., Staeuber, N., Sutton, G., and Roy, P. (1997). Effects of domain-switching and site-directed mutagenesis on the properties and functions of the VP7 proteins of two orbiviruses. *Virol*., **237**(2): 217–227.

Moulé, L. (1890). Histoire de la Médecine Vétérinaire. Maulde, Paris, p 38.

Naggy, P.D. and Simon, A.E. (1997). New insights into the mechanisms of RNA recombination. *Virol*., **235**: 1-9.

Nei, M. and Kumar, S. (2000). Molecular evolution and phylogenetics. Oxford University Press, New York.

Niedbalski, W. (2013). The evolution of bluetongue virus: genetic and phenotypic diversity of field strains. *Pol. J. Vet. Sci*., **16**(3): 611-616.

Nielsen, R. and Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to HIV-1 envelope gene. *Genetic. Soc. Am*., **148**: 929-936.

Nomikou, K., Hughes, J., Wash, R., Kellam, P., Breard, E., Zientara, S., Palmarini, M., Biek, R. and Mertens, P. (2015). Widespread reassortment shapes the evolution

and epidemiology of bluetongue virus following European invasion. *PLOS Pathog*., pp 1-23.

Novella, I.S., Presloid, J.B., Smith. S.D. and Wilke, C.O. (2011). Specific and nonspecific host adaptation during arboviral experimental evolution. *J. Mol. Microbiol. Biotechnol*., **21**: 71-81.

Novella, I.S., Hershey, C.L., Escarmis, C., Domingo, E. and Holland, J.J. (1999). Lack of evolutionary stasis during alternating replication of an arbovirus in insect and mammalian cells. *J. Mol. Biol*., **287**:459-465.

Oberst, R.D., Stott, J.L., Blanchard-Channell, M. and Osburn B.I. (1987). Genetic reassortment of bluetongue virus serotype 11 strains in the bovine. *Vet. Microbiol*., **15**:11–8.

OIE. (2009). AFfrican horse sickness. OIE Sci. Tec. Dept., last updated: October 2009.

OIE. (2012). AFfrican horse sickness. OIE Terrestrial manual. Section 2.5. Chapt 2.5.1. Pp 1-12.

OIE (2017). African horse sickness (Infection with African horse sickness virus). OIE Terrestrial manual. Section 2.5. Chapter 2.5.1.Pp 1-16.

Oldfield, S., Adachi, a, Urakawa, T., Hirasawa, T., and Roy, P. (1990). Purification and characterization of the major group-specific core antigen VP7 of bluetongue virus synthesized by a recombinant baculovirus. *J. Gen. Virol*., **71**:2649–2656.

Padidam, M., Sawyer, S. and Fauquet, C.M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virol*., **265**: 218-225.

Patel, A. and Roy, P. (2014). The molecular biology of bluetongue virus replication. *Virus Res*., **182**: 5-20.

Perez-Losada, M., Arenas, M., Galan, J.C., Palero, F. and Gonzalez-Candelas, F. (2015). Recombination in viruses: Mechanisms, methods of study and evolutionary consequences. *Infect. Genet. Evol.*, **30**: 296-307.

Polson, A. and Deeks, D. (1963). Electron microscopy of neurotropic African horse-sickness virus. *J. Hyg.Camb*., **61**:149-153.

Posada, D. and Crandall, K.A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc. Natl. Acad. Sci*., **98**:13757-13762.

Posada. D., Crandall, K.A. and Holmes, E.C. (2002). Recombination in evolutionary genomics. *Annu. Rev. Genet*., **23**:75-97.

Potgieter, A.C., Page , N.A., Liebenberg, J., Wright, I.M., Landt, O. and van Dijk, A.A. (2009). Improved strategies for sequence-independent amplification and sequencing of viral double-stranded RNA genomes. *J. Gen. Virol*., **90**: 1423-1432.

Prasad, B. V. V., Yamaguichi, S. and Roy, P. (1992). Three-dimensional structure of single-shelled BTV. *J. Virol*., **66**: 2135-2142.

Prosperi, M.C.F. and Salemi, M. (2012). QURE: Software for viral quasispecies reconstruction from next generation sequence data. *Bioiformatics*., **28**(1): 132-133.

Quan, M., van Vuuren, M., Howell, D.G., Groenewald, D. and Guthrie, A.J. (2008). Molecular Epidemiology of African horse sickness virus S10 gene. *J. Gen. Virol*., **89**: 1159-1168.

Ratinier, M., Caporale, M., Golder, M., Franzoni, G., Allan, K., Nunes, S.F., Armezzani, A., Bayoumy, A., Rixon, F., Shaw, A. and Palmarini, M. (2011). Identification and characterization of a novel non-structural protein of bluetongue virus. *Plos. Pathog*., 7(12) e1002477. doi: 10.1371/journal.ppat.1002477.

Rafyi, A. (1961). Horsesickness. *Bull. Off. Int. Epizoot*., **56**: 216-250.

Ragonnet-Cronin, M., Hodcroft, E., Hué, s., Fearnhill, E., Delpech, V., Brown, A.J.L. and Lycett, S. (2013). Automated analysis of phylogenetic clusters. *BMC Bioinformatics*, **14**: 317-325.

Ren, A., Ishida, T. and Akiyama, Y. (2013). Assessing statistical reliability of phylogenetic tree via a speedy double bootstrap method. *Mol. Phylogenet. Evol*., **67**(2):429-35

Robin. M., Page. P., Archer. D. and Baylis. M. (2016). African horse sickness: The potential for an outbreak in disease-free regions and current disease control and elimination techniques. *Equi. Vet. J*., **48**:659-669.48

Roy, P., 1996. Orbivirus structure and assembly. *Virol*., **216**: 1–11.

Roy, P. (2004). Orbiviruses. Fields Virology. 5th Ed. Knipe DM., Howley, P.M., Griffin, D.E., Lamb, R.A., Marin, M.A., Roizman, B. and Straus, S.E (Ed). **2**:1975-1996.

Roy, P., Hirasawa, T., Fernandez, M., Blinov, V. M., and Sanchez-Vixcain Rodrique, J. M. (1991). The complete sequence of the group-specific antigen, VP7, of African horsesickness disease virus serotype 4 reveals a close relationship to bluetongue virus. *J. Gen. Virol*., **72**(6): 1237– 1241.

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

Sailleau, C., Hamblin, C., Paweska, J. T. and Zientara, S. (2000). Identification and differentiation of the nine African horse sickness virus serotypes by RT-PCR amplification of the serotype-specific genome segment 2. *J. Gen. Virol.*, **81**(3): 831–837.

Sailleau, C., Moulay, S. and Zientara, S. (1997). Nucleotide sequence comparison of the segments S10 of the nine African horse sickness virus serotypes. *Arch. Virol.*, **142**: 965-978.

Saitou, N. and Ueda, S. (1994). Evolutionary rates of insertion and deletion in noncoding nucleotide sequences of primates. *Mol. Biol. Evol*., **11**(3):504-512.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of National Academy of Science of the United States of America*, **74**:5463-5467.

Sellers, R.F., Pedgley, D.E. and Tucker, M.R. (1977). Possible spread of African horse sickeness on the wind. *J. Hyg*., **79**: 279-298.

Shafiq, M., Minakshi, P., Bhateja, A., Ranjan, K. and Prasad, G. (2013). Evidence of genetic reassortment between Indian isolate of bluetongue virus serotype 21 (BTV-21) and bluetongue virus serotype 16 (BTV-16). *Virus Res*., **173**(2):336-343.

Shaw, A.E., Ratinier, M., Nunes, S.F., Nomikou, K., Caporale, M., Golder, M., Allan, K., Hamers, C., Hudelet, Zientara, S., Breard, E., Mertens, P. AND Palmarini, M. (2013). Reassortment between two serologically unrelated bluetongue virus strains is flexible and can involve any genome segment. *J. Virol*., **87**(1): 543-557.

Simon-Loriere, E. and Holmes, E.C. (2011). Why do RNA viruses recombine? *Nat. Rev. Microbiol*., **9**: 617-626.

Stauber, N., Martinez-Costas, J., Sutton, G., Monastyrskaya, K. and Roy, P. (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.

Sun, J.C. and Lanier, L.L. (2009). The natural selection of herpesviruses and virus-specific NK cell receptors. *Rev. Viruses*, **1**: 362-382.

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

Suzuki, Y., Yamaguchi-Kabata, Y. and Gojobori, T. (2000). Nucleotide substitution rates of HIV-1. *AIDS. Rev*., **2**: 39-47.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol*., **30**: 2725-2729.

Taylor, M.S., Ponting, C.P. and Copley, R.R. (2004). Occurrence and consequences of coding sequence insertions and deletions in mammalian genomes. *Genome Res*., **14**: 555-566.

Theiler, A. (1921). African horse sickness (pestis equorum). *Sci. Bull*., **19**: 1-29.

Thomas, C. P. Booth, T. F., and Roy, P. (1990). Synthesis of bluetongue virus encoded phosphoprotein , and formation of inclusion bodies by recombinant

baculovirus in insect cells : it binds the single stranded RNA species. *J. Gen. Virol*., **71**: 2073-2083.

Urakawa, T., Ritter, D.G. and Roy, P. (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 acid Res*., **17**: 7220-7226.

van Niekerk, M., Smit, C. C., Fick, W. C., van Staden, V. and Huismans, H. (2001a). Membrane association of African horsesickness virus nonstructural protein NS3 determines its cytotoxicity. *Virol*., **279**(2): 499–508.

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

Verwoerd, D.W., Louw, H. and Oellermann, R.A. (1970). Characterization of bluetongue virus ribonucleic acid. *J. Virol*., **5**(1): 1-7.

Vijaykrishna, D., MukerJi, R. and Smith, G.J.D. (2015). RNA virus reassortment: An evolutionary mechanism for host jumps and immune evasion. *PLOS Pathog*, DOI:10.1371/Journal.ppat.1004902:1-6.

Von Teichman, B. F., Dungu, B., and Smit, T. K. (2010). In vivo cross-protection to African horse sickness Serotypes 5 and 9 after vaccination with Serotypes 8 and 6. *Vaccine*, **28**(39): 6505–6517.

Von Teichman, B. F. and Smit, T. K. (2008). Evaluation of the pathogenicity of African Horsesickness (AHS) isolates in vaccinated animals. *Vaccine*, **26**(39): 5014–5021.

Wargo, A.R. and Kurath, G. (2012). Viral fitness: definitions, measurement, and current insights. *Cur. Opi. Virol*., **2**: 538-545.

Walker P.J. and Cowley J.A. (1999). Viral genetic variation: Implications for disease diagnosis and detection of shrimp pathogens. Co-operative Research Centre for Aquaculture, *CSIRO Tropical Agric*., p 54-59.

Weyer, C.T., Grewar, J.D., Burger, P., Rossouw, E., Lourens, C., Joone, C., le Grange, M., Coetzee, P., Venter, E., Martin, D.P., MacLachlan, N.J. and Guthrie, A.J. (2016). African horse sickness caused by genome reassortment and reversion to virulence of live, attenuated vaccine viruses, South Africa, 2004-2014. *Emerg. Infect. Dis*., **22**(12):2087-2096.

Weyer, C.T., Joone, C., Lourens, C.W., Monyai, M.S., Koekemoer, O.J.J., Grewar, J.D., van Schalkwyk, A., Majiwa, P.O.A., MacLachlan, N.J. and Guthrie, A.J. (2015). Development of three triplex real-time reverse transcription PCR assays for the qualitative molecular typing of the nine serotypes of African horse sickness virus. *J.Virol. Meth*., **223**: 69-74.

Wilson, A., Mellor, P. S., Szmaragd, C., and Mertens, P. P. (2009). Adaptive strategies of African horse sickness virus to facilitate vector transmission. *Vet. Res*., **40**: 1–25.

Woelk, C.H. and Holmes, E.C. (2002). Reduced positive selection in vector-borne RNA viruses. *Mol. Biol. Evol*., **19**: 2333-2336.

Worobey, M. and Holmes, E.C. (1999). Evolutionary aspects of recommbination in RNA viruses. *J. Gen. Virol*., **80**: 2535-2543.

Yang, Z., Nielsen, R., Goldman, N. and Pedersen, A-M.K. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genet.*, **155**: 431-449.

Zwart, L., Potgieter, C.A, Clift, S.J. and van Staden, V. (2015). Characterising non-structural protein NS4 of African horse sickness virus. *PloS One*, **10**(4):1-18.

Zientara, S., Weyer, C.T. and Lecollinet, S. (2015). African horse sickness. *Rev. Sci. Tech. Off. Int. Epiz*., **34**(2):315-327.

APPENDIX 1

Table 1. Serotype number, origin, passage history and date of collection for the isolates used in this study.

Serotype	Name	Passage history	Date (Year)	Country of origin
1	AHSV-1_107_09	2V	2009	RSA
1	AHSV-1_29_62	1H,1A	1962	RSA
1	AHSV-1_22_01	Spl,1S,2BHK	2001	RSA
1	AHSV-1_29_02	Spl, 1S, 1V, 2 BHK	2002	RSA
1	AHSV-1_44_00	Spl,	2000	RSA
2	AHSV-2_20_09	2V,1V	2009	RSA
2	AHSV-2_82_61	1H,2S	1961	RSA
2	AHSV-2_40_00	Bld,2S,2BHK	2000	RSA
2	AHSV-2_20_01	Spl,1S	2001	RSA
2	AHSV-2_6_81	Spl,	1981	Zim
2	AHSV-2_81_03		2003	
3	AHSV-3_109_08	1V	2008	RSA
3	AHSV-3_13_63	3S	1963	RSA
3	AHSV-3_DG 25324_14	Bld	2014	RSA
3	AHSV-3_DG 25327_14	Bld	2014	RSA
3	AHSV-3_DG 25423_14	Bld	2014	RSA
3	AHSV-3_14_98	Spl, 3V	1998	RSA
3	AHSV-3_48_01	Spl,	2001	RSA
3	AHSV-3_1_02	Spl, 2S, 1V, 1BHK	2002	RSA
3	AHSV-3_2_89	1S	1989	RSA
4	AHSV-4_68_09	2V	2009	RSA
4	AHSV-4_32_62	Spl	1962	Zim
4	AHSV-4_31_00	Bld,1S,3BHK	2000	RSA
4	AHSV-4_91_00	Spl,1S,4BHK	2000	RSA
4	AHSV-4_19_97	2S,1V	1997	
4	AHSV-4_37_98	3V	1998	
4	AHSV-4_65_00	Bld,1S,2BHK	2000	RSA
4	AHSV-4_97_99	1S,2V	1999	RSA
4	AHSV-4_6_14	Lung	2014	Nam
4	AHSV-4_1927_14	Lung	2014	Nam
4	AHSV-4_64_99	1S, 2V	1999	RSA
4	AHSV-4_90_96	Spl, 1V	1996	RSA
4	AHSV-4_90_01	3S	2001	RSA
5	AHSV-5_30_62	2H,1mb	1962	RSA
5	AHSV-5_35_09	3V	2009	RSA

5	AHSV-5_13_94	1V	1994	
5	AHSV-5_93_00	Bld, 2S, 2 BHK	2000	RSA
5	AHSV-5_42_01	Spl, 2S	2001	RSA
5	AHSV-5_13_99	1S, 2V	1999	RSA
5	AHSV-5_86_99	Liver, 1S, 1V	1999	RSA
5	AHSV-5_86_94	Lung, 2V	1994	RSA
5	AHSV-5_47_86	Bld, 4BHK	1986	RSA
5	AHSV-5_2_96	1V	1996	RSA
6	AHSV-6_63_09	3V	2009	RSA
6	AHSV-6_4_98	2V	1998	RSA
6	AHSV-6_2_75	Bld, 2S, 4 BHK	1975	RSA
6	AHSV-6_6_90	Lung, 3V	1990	RSA
6	AHSV-6_68_00	Spl, 1S, 1 BHK	2000	RSA
6	AHSV-6_19_98	Spl, 3V	1998	RSA
6	AHSV-6_33-1_98	Spl, 2V	1998	RSA
6	AHSV-6_110_00	Spl, 1S, 1V, 1 BHK	2000	RSA
6	AHSV-6_14_01	Spl, 1S, 2 BHK	2001	RSA
6	AHSV-6_28_87		1987	
6	AHSV-6_26_03	Bld, 1S, 1 BHK	2003	RSA
6	AHSV-6_93_99	1S, 2V	1999	RSA
6	AHSV-6_33_98	Spl, 2V	1998	RSA
7	AHSV-7_89_09	2V	2009	RSA
7	AHSV-7_31_62	1H,1mb	1962	RSA
7	AHSV-7_1955_14		2014	RSA
7	AHSV-7_67_99	1S, 2V	1999	RSA
7	AHSV-7_34_01	Spl, 2S, 1V, 1BHK	2001	RSA
7	AHSV-7_3_99	1S, 1v	1999	RSA
7	AHSV-7_105_99	1S, 2V,	1999	RSA
8	AHSV-8_3_00		2000	
8	AHSV-8_29_00	1S,4BHK	2000	RSA
8	AHSV-8_10_62	1A	1962	Kenya
8	AHSV-8_96_94	Lung, 1V	1994	RSA
8	AHSV-8_21_99	1S, 2V	1999	RSA
8	AHSV-8_2_99	1S, 2V	1999	RSA
8	AHSV-8_48_98	Spl, 3V	1998	RSA
8	AHSV-8_88_99	1S, 2V	1999	RSA
8	AHSV-8_9_98	Bld, 3V	1998	RSA
8	AH**S**V-8_9B_98	Bld, 3V	1998	RSA
8	AHSV-8_17_98	Spl, 3V	1998	RSA
8	AHSV-8_58_81	Lung, 1S	1981	RSA
9	AHSV-9_90_61	3S	1961	Chad
9	AHSV-9_38_09	2V	2009	RSA
9	AHSV-9_1945_14		2014	RSA
9	AHSV-9_2033_14		2014	Moz
9	AHSV-9_3_14		2014	RSA

9	AHSV-9_43_01	Spl,2S,1V,1BHK	2001	RSA
9	AHSV-9_2_03	Spl, 1S, 1V, 1 BHK	2003	RSA
9	AHSV-9_49_02	Spl, 1S, 2 BHK	2002	RSA
9	AHSV-9_22_03	Spl, 3V	2003	RSA
9	AHSV-9_15_02	Spl, 1S, 1V, 1 BHK	2002	RSA
9	AHSV-9_22_03	Lung, 1S, 1 BHK	2003	RSA
9	AHSV-9_6_01	2 BHK	2001	RSA
Sequences from GenBank (www.ncbi.nlm.nih.gov)				
1	AHSV-1_21_07	KP009621 - KP009630	2007	RSA
1	AHSV-1_1180_33	KP009711 - KP009720	1933	RSA
2	AHSV-2_90_07	KP009631 - KP009640	2007	RSA
2	AHSV-2_02_07	FJ196584 - FJ196593	2007	Nigeria
3	AHSV-3_73_08	KP009641 - KP009650	2008	RSA
3	AHSV-3_Ladysmith_40	KP009761 - KP009770	1940	RSA
4	AHSV-4_128_06	KP009651 - KP009660	2006	RSA
4	AHSV-4_Vryheid_38	KP009771 - KP009780	1938	RSA
5	AHSV-5_28_08	KP009661 - KP009670	2008	RSA
5	AHSV-5_Westerman_36	KP009781 - KP009790	1936	RSA
6	AHSV-6_04_08	KP009671 - KP009680	2008	RSA
7	AHSV-7_Karen_52	KP009751 - KP009760	1952	Kenya
8	AHSV-8_29_00	KP009691 - KP009700	2000	RSA
9	AHSV-9_27_08	KP009701 - KP009710	2008	RSA

Spln, Equine spleen; BLd, Equine blood; Lung, Equine lung; A, adult mice; S, suckling mice; BHK, baby hamster kidney (BHK)-21; V, vero cells

Table 2. Oligo Ligation Reagents

60% PEG 6000		15 g in 25 ml H_2O	Heat 37^oC on shaker to dissolve, filter	
1m HEPES, pH8.0	238.3	11.91 g in 50ml H_2O	pH 8.0 with saturated NaOH, filter	Sigma H0891 100G
1M $MgCl_2$	203.30	10.16 g in 50ml H_2O	Filter	$MgCl_2$.6 H_2O
100Mm ATP		0.0605 g/1ml H_2O	No loves= static, rinse spatula with H_2O	(Roche) Disodium Saly=t. Mr= 605.2
300Mm DTT		0.0462 g/1m H_2O	Fill up to 1ml-water replacement of nearly 200 µl	Roche
Primer PC3T7loop		300 ng/µl		Tib Molbiol
DMSO				
T4 RNA Ligase with BSA 0.1%	99.8%			Takara

Table 3. Oligo Ligation buffer

Volume for 490 µl	Volume for 980 µl	Reagents
250 µl	500 µl	1M Hepes pH 8.0
90 µl	180 µl	$MgCl_2$
50 µl	100 µl	BSA
50 µl	100 µl	DTT
50 µl	100 µl	ATP
Total volume 490 µl	Total volume 980 µl	Store at -20 ^oC for +/- 2 weeks

APPENDIX 2. Phylogenies

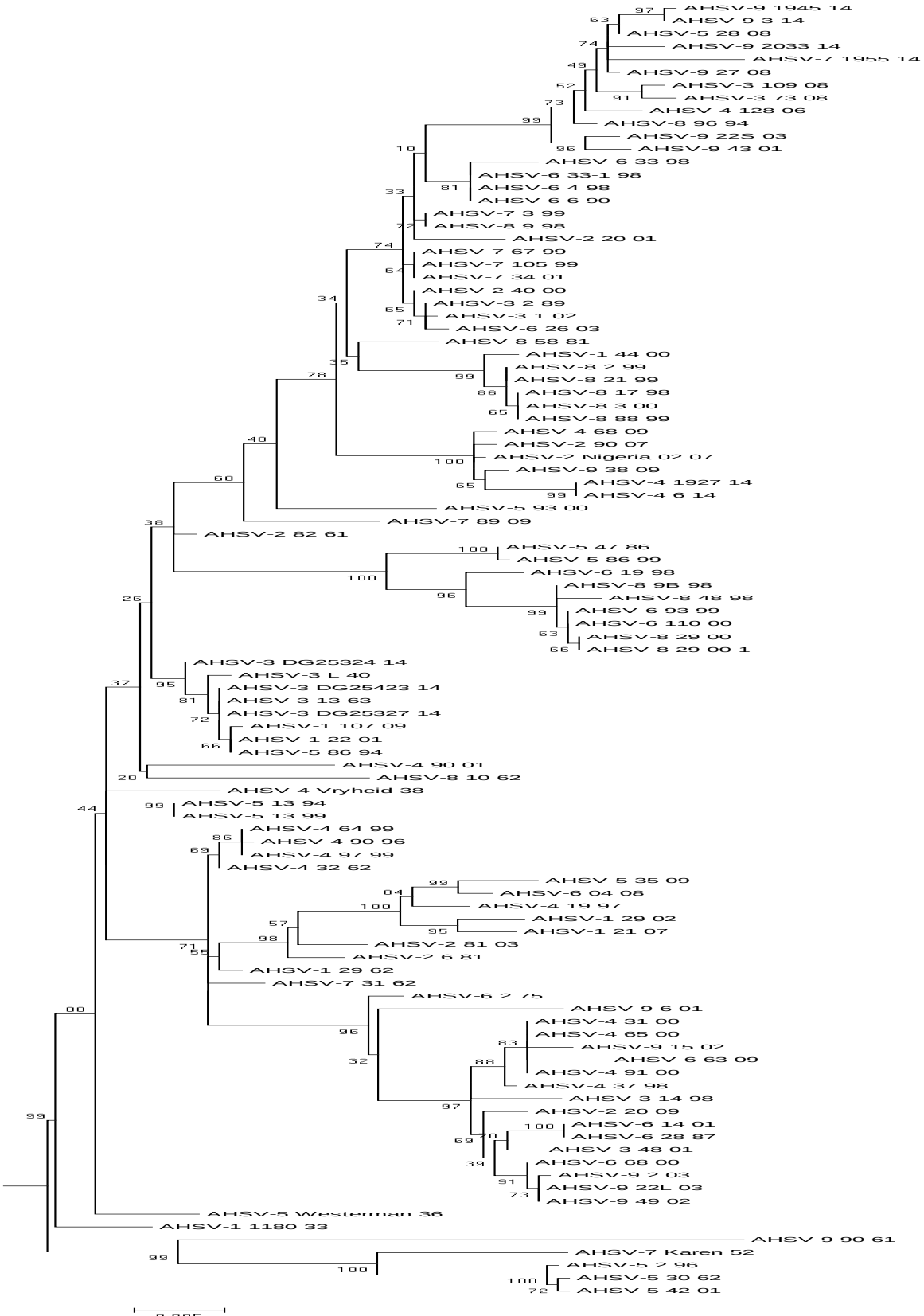

Figure 1. Seg-5 of different AHSV serotypes

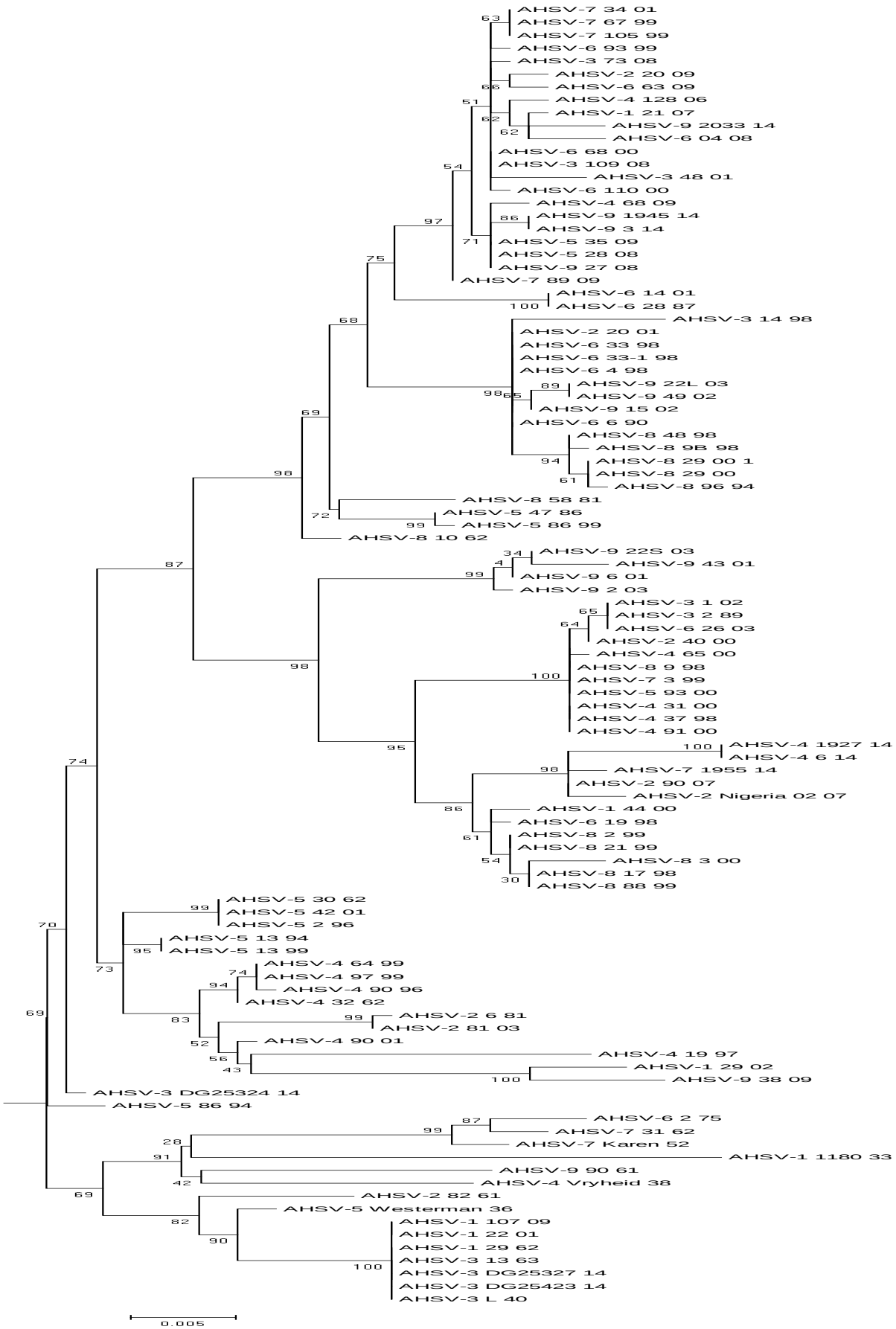

Figure 2. Seg-8 of all nine AHSV serotypes

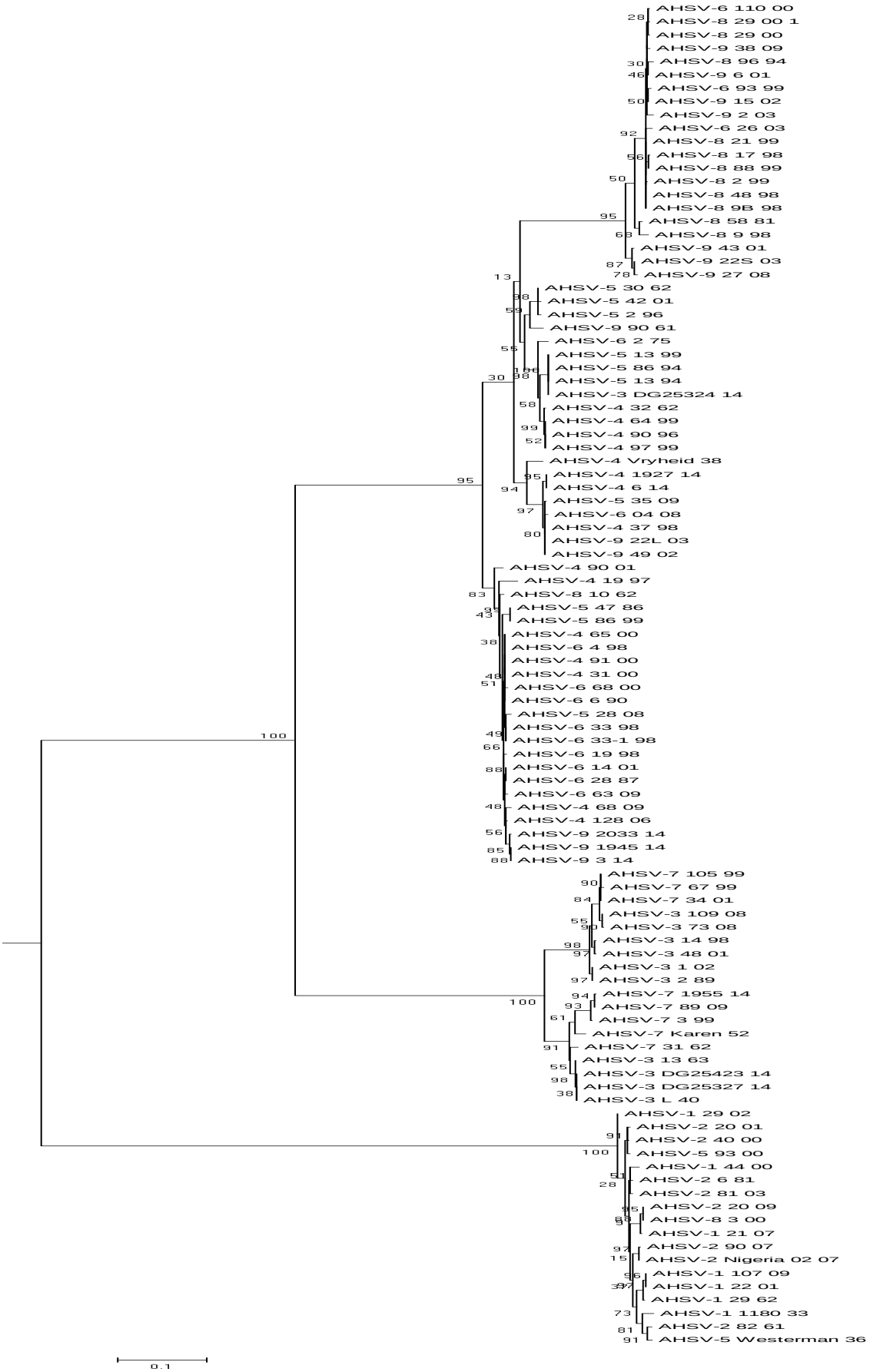

Figure 3. Seg-10 of all nine AHSV serotypes

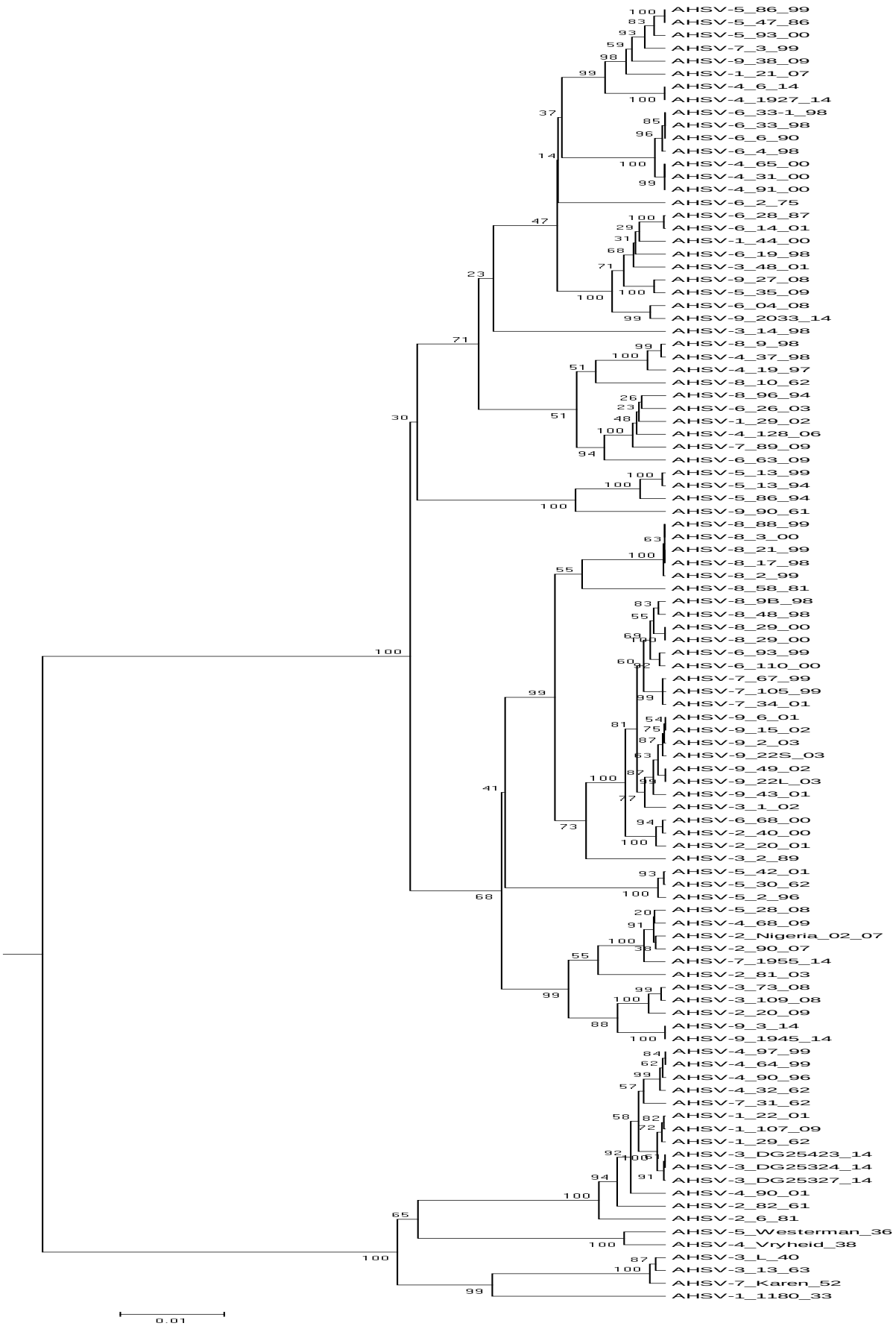

Figure 4. Seg-1 of all nine AHSV serotypes

Figure 5. Seg-3 of all nine AHSV serotypes

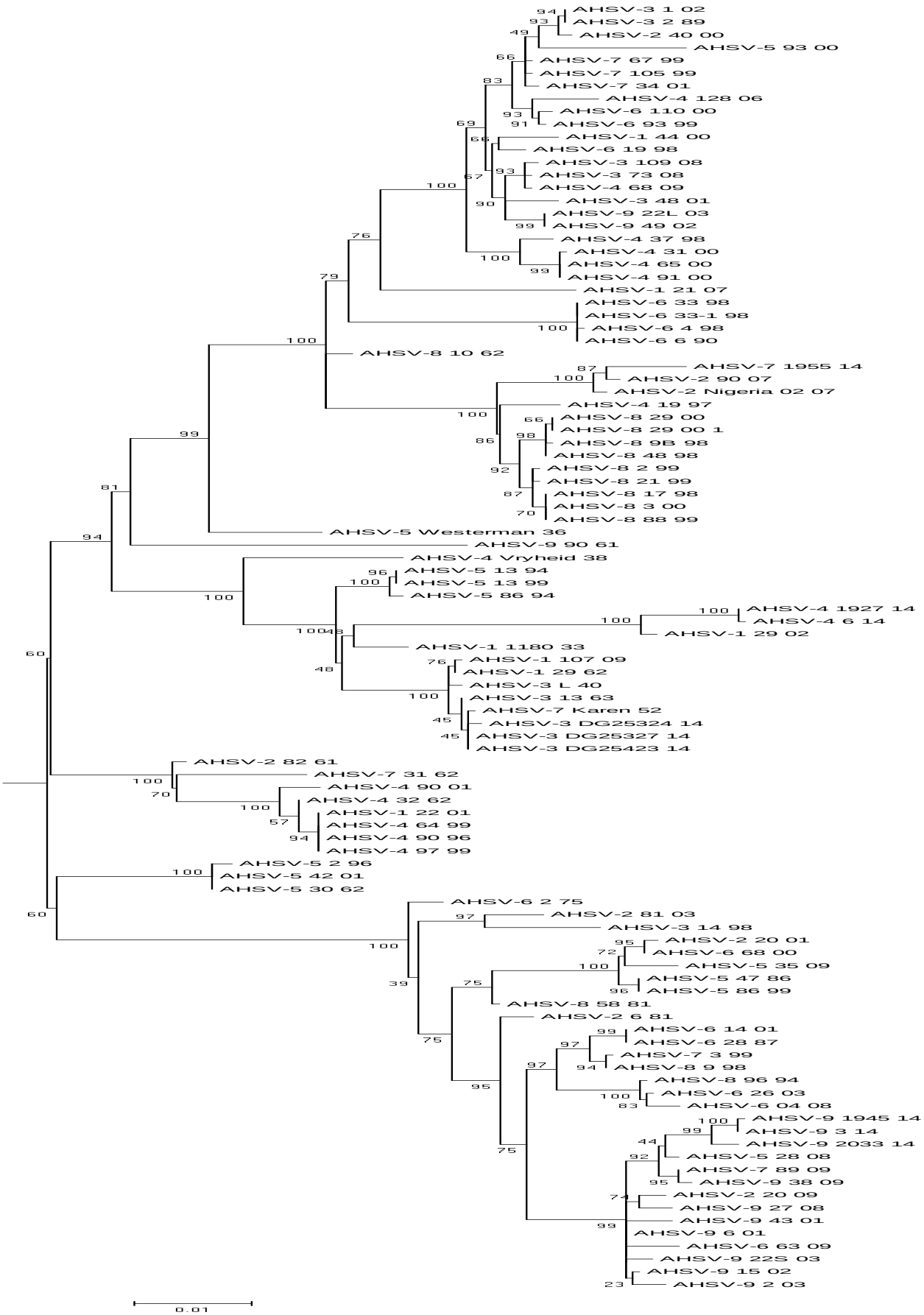

Figure 6. Seg-4 of all nine AHSV serotypes

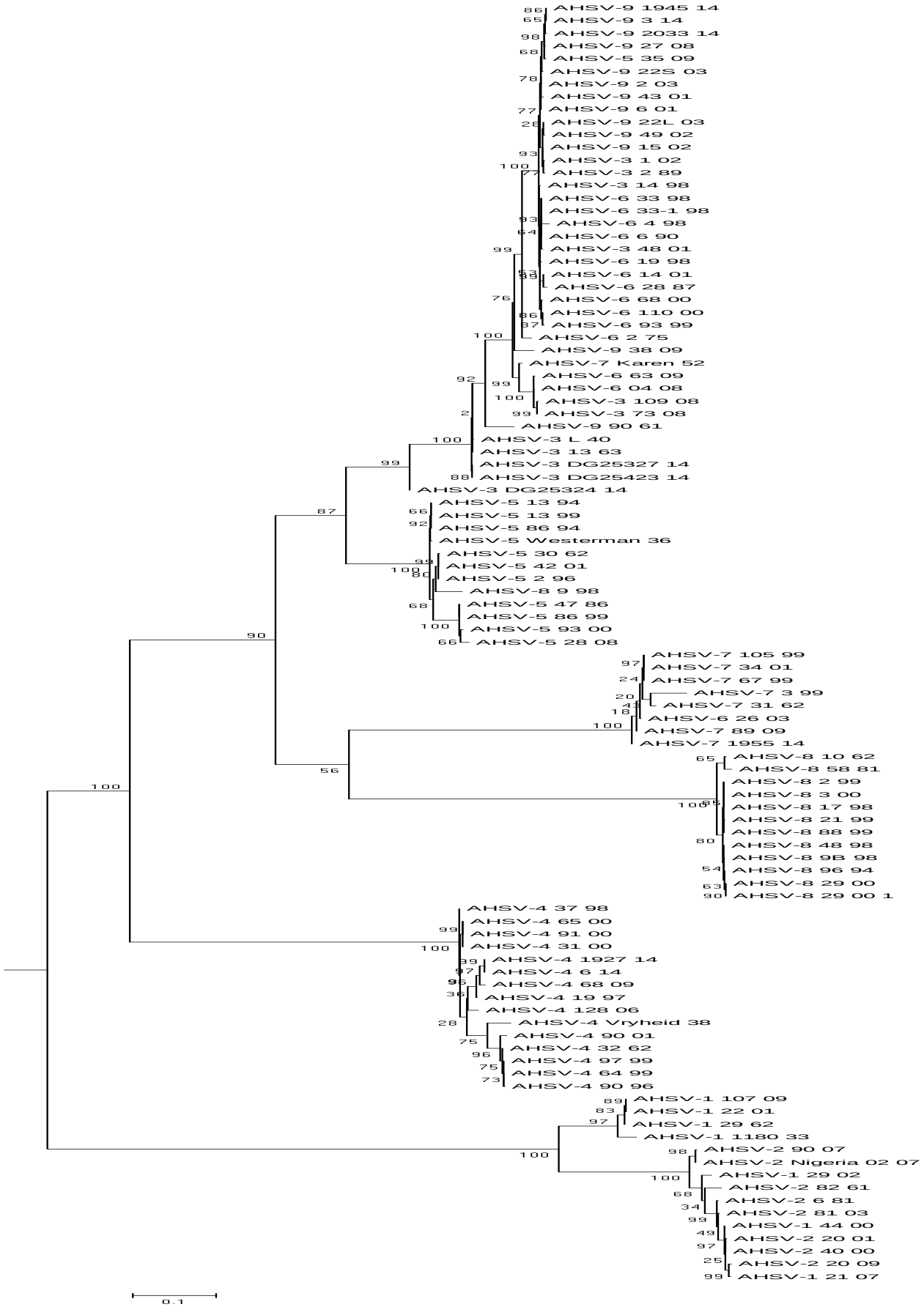

Figure 7. Seg-6 of all nine AHSV serotypes

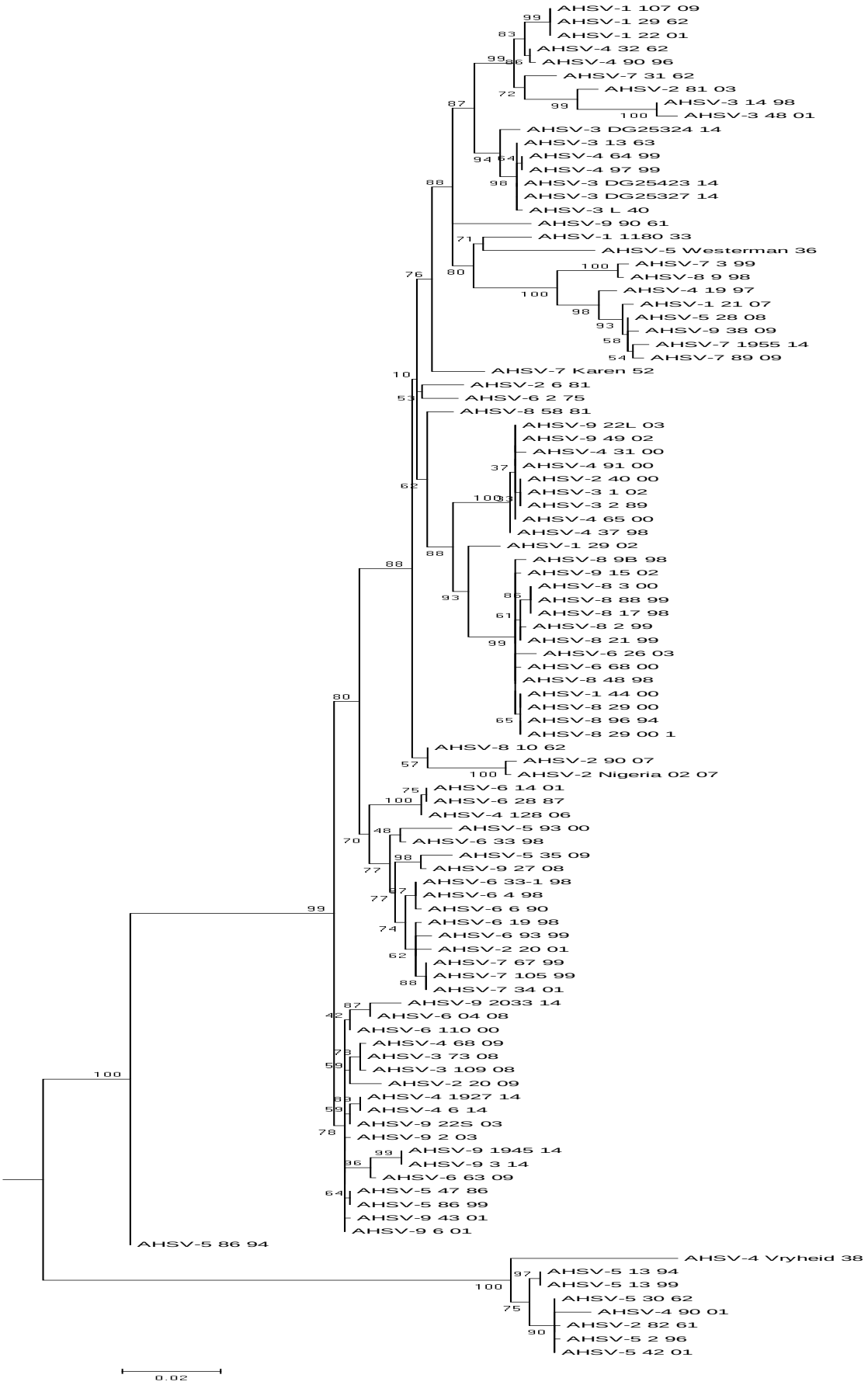

Figure 8. Seg-7 of all nine AHSV serotypes

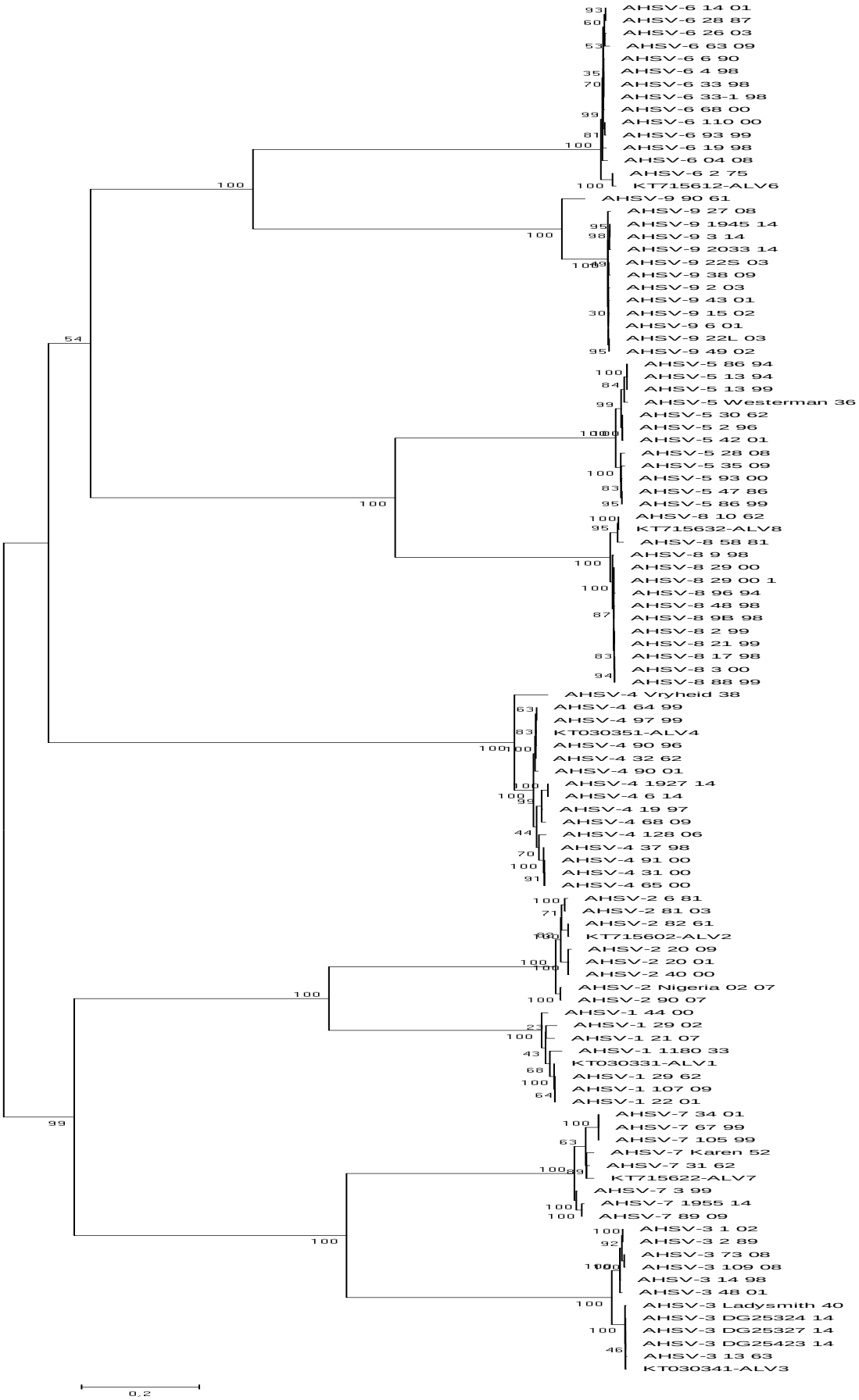

Figure 9. Seg-2 of all nine AHSV serotypes

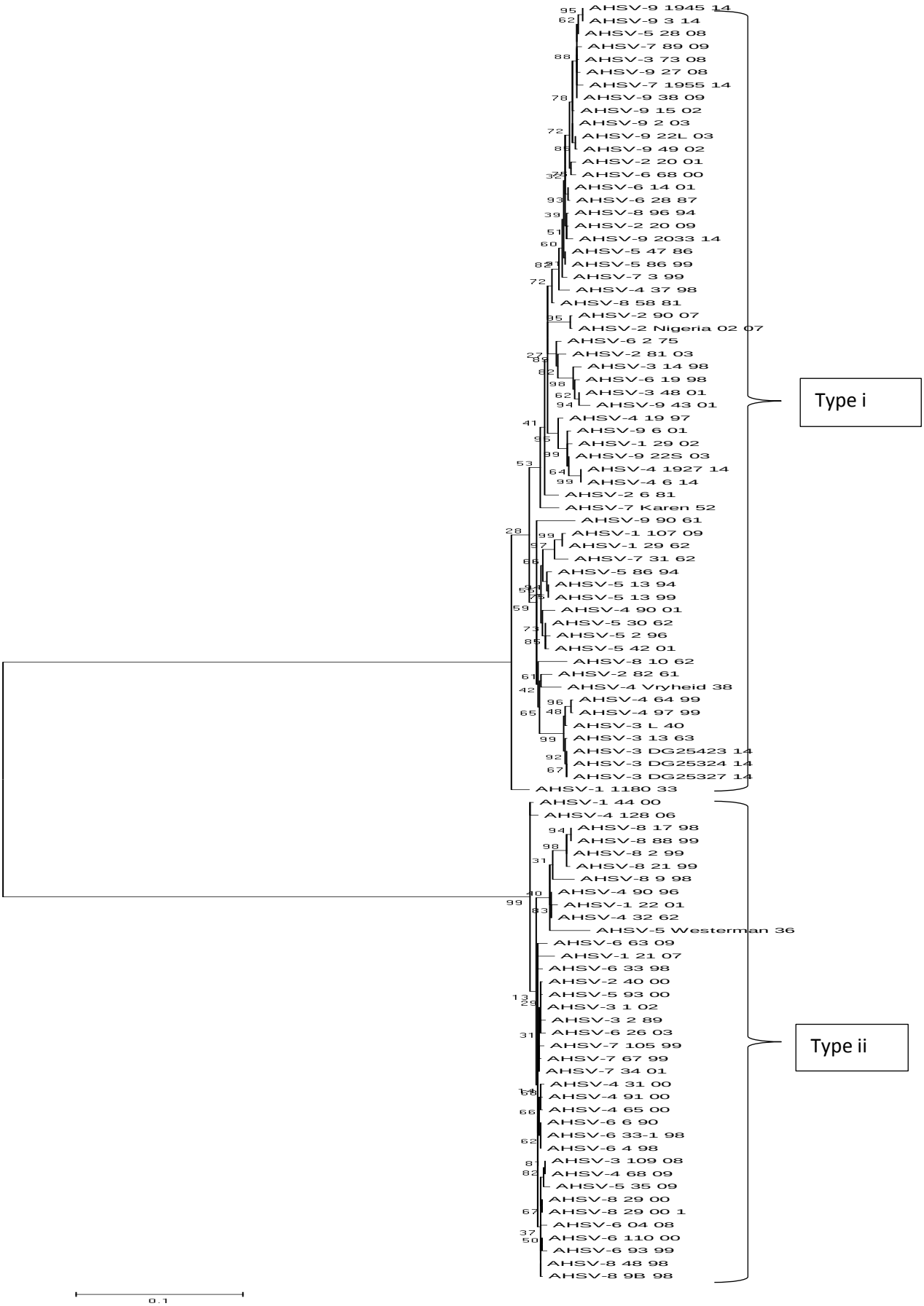

Figure 10. Seg-9 of all nine AHSV serotypes, also showing two homologies (Type i and ii)