Genetic diversity, molecular evolution and classification of the viruses in the genus *Amdoparvovirus* circulating in free-ranging mink in Nova Scotia and detection of novel species

By

Faezeh Kharazyan

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
December 2016

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Aleutian mink disease virus (AMDV) causes Aleutian disease (AD), which results in economic losses to the mink industry globally. Free-ranging mink are reservoirs of AMDV and can transmit the virus to mink farms. To identify the source of infection, entire coding and partial 3′ terminal regions of 25 AMDV isolates in free-ranging mink in Nova Scotia were sequenced. Four groups of *Amdoparvoviruses* in free-ranging mink were identified, of which three were novel species. It was shown that the N-terminus of the NS1 protein plays a significant role in speciation of *Amdoparvoviruses*. Multiple infection of mink with closely related viral isolates was observed and frequent recombination events were detected throughout the viral genomes. Guidelines for *Amdoparvovirus* classification were provided and a phylogenetic marker was developed, which provides great opportunities for farmers to accurately identify the source of infection on their farms with low cost.

LIST OF ABBREVIATIONS USED

aa	Amino acid
AD	Aleutian disease
AMDV	Aleutian mink disease virus
ATP	Adenosine triphosphate
bp	Base pair
CO	Colchester
CPV	Canine parvovirus
CU	Cumberland
dNTP	Deoxynucleotide
FPV	Feline panleukopenia virus
GFAV	Gray fox amdovirus
HA	Halifax
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HVR	Hypervariable region
ICTV	International Committee on the Taxonomy of Viruses
kb	Kilo base
KI	Kings
L-ORF	Left open reading frame
LU	Lunenburg
ML	Maximum Likelihood
mb	Mega base
MVM	Minute virus of mice
NJ	Neighbour Joining
NS1	Non-structural protein 1
NS2	Non-structural protein 2
NS3	Non-structural protein 3
nt	Nucleotide
ORF	Open reading frame
PI	Pictou
R-ORF	Right open reading frame
RFAV	Raccoon dog amdoparvovirus
ssDNA	Single-stranded DNA
UTR	Untranslated region
VP1	Virion protein1
VP2	Virion protein2
YA	Yarmouth

ACKNOWLEDGMENTS

Today is the day to add the finishing touch on my thesis by acknowledging remarkable individuals who so generously contributed to the work presented in this thesis and supported me during my life in graduate school. First and foremost, I offer my sincere gratitude to my supervisor, Dr. Hossain Farid, who provided me the great opportunity to do my Master's degree on the fascinating research field of viral evolution and guided me to grow as a research scientist. I attribute the level of my Master's degree to his consistent guidance, support, encouragement and willingness to give his time so generously whenever I needed it.

My special appreciation goes to my co-supervisor, Dr. Conor Meehan for offering valuable advice on evolutionary analysis during the development of this research work and for his concern about this thesis. Many thanks goes to my advisory committee members Dr. Sarah Stewart-Clark, Dr. Bruce Rathgeber and Dr. Leslie MacLaren for their time in service on my Masters committee. I would like to thank Dr. Robert Beiko and Dr. Darren Martin for teaching me the concepts of recombination analysis.

The collaboration of the Nova Scotia Trappers Association and the trappers who participated in collecting mink samples is appreciated. Financial supports received from the Joint Mink Research Committee and the Research Acceleration Program of the Nova Scotia Department of Agriculture is greatly acknowledged. I would also like to thank the American Society of Virology and the Canadian Society of Animal Science for providing travel awards to present my project in their prestigious conferences. Technical assistance of Priyanka Rupsinghe and Irin Arju is gratefully acknowledged.

To my dear friend Miss Negar Sharifi Mood, thank you for true friendship and caring you provided all the times and for helping me get through the difficult times. My life in graduate school would not be more memorable and blissful without you. I would like to thank Ms. Flora Riyahi for her tremendous kindness and support from the very first day I entered Canada. To my friends Ms. Sheyda Razizade, Ms. Ling Guo, Ms. Maria Caraza Salas and Ms. Zahra Dehghani, thank you for your friendship and always being there for me.

My heartfelt gratitude goes to my beloved parents, brothers and sister-in-law whose unconditional love, support and inspiration helped me to follow my dreams. Without their blessings and encouragement, I would not have been able to finish this work. Mom and dad, I finally became a scientist. Love you both so much.

CHAPTER 1. INTRODUCTION

Nova Scotia is the largest producer of mink pelts in Canada (Statistics Canada, 2014). The high concentration of mink ranches at the western part of Nova Scotia represents great opportunities for the spread of pathogens, most importantly the Aleutian mink disease virus (AMDV) (Farid et al., 2012). The AMDV infection is the major problem for the mink industry in Nova Scotia (Farid et al., 2012) and causes economic losses as a result of increased mortality in newborn kits and adults, decreased reproductive performance (Alexandersen, 1986; Bloom et al., 1994) and the presence of unfavorable white hair fibers on the pelt (Farid and Ferns, 2011). There is no effective treatment for or a vaccine against this disease (Aasted et al., 1998). Since viral eradication has not been successful in Nova Scotia (Farid et al., 2012), identifying sources of repeated reappearance of AMDV on cleaned ranches remains the best way of controlling the AMDV infection (Farid et al., 2012). A high percentage (93.3%) of free-ranging mink in Nova Scotia are infected with AMDV (Farid, 2013), and are a major reservoir of the virus, possibly transmitting the virus to farmed mink. In order to determine if free-ranging mink is the source of contamination of AMDV infection on the mink ranches, a sequence database of the virus in free-ranging mink is needed, on which there is no information available. In addition, AMDV isolates from farmed mink, whose entire coding sequence is available in public databases are limited to the AMDV-G, Utah1, LN1, LN2, LN3 and SL3 (Schuierer et al., 1997; Bloom et al., 1988; Bloom et al., 1990; Li et al., 2012). The AMDV identification and epidemiological studies have been conducted extensively based on partial regions of the AMDV

genome using two pairs of primers published by Oie et al. (1996) and Olofsson et al. (1999), which led to inconsistencies in the literature and to misinterpretation of epidemiological studies. Thus, large number of longer sequences were required to evaluate the validity of the results of analysis of partial regions as well as gaining a better view of the diversity and evolution of *Amdoparvoviruses* in Nova Scotia and globally.

The primary objective of this study was to create the first sequence database of AMDV in free-ranging mink in Nova Scotia by sequencing the entire coding region of 25 AMDV isolates obtained from different counties. The information generated in this research will help farmers to identify whether the source of contamination of their mink is free-ranging mink. Such information would help farmers to implement proper biosecurity standards on their farms to prevent further devastating damages. This sequence database will then be used for several purposes:

1. Exploring genetic characteristics of individual genes and proteins of the local and publically available AMDV isolates.
2. Detecting recombination events and determining the distribution of these event across the AMDV genome.
3. Classifying the genus *Amdoparvoviruses* based on sequence identity, genetic distance and phylogenetic analyses, evaluating the effect of recombination on phylogenetic analysis and the validity of partial genome region analysis as well as developing a phylogenetic marker for molecular epidemiological studies of *Amdoparvoviruses*.

CHAPTER 2. LITERATURE REVIEW

2.1 Permissive and restricted infections of AMDV

Replication of AMDV is not dependent on the presence of other viruses and thus, AMDV is an autonomous parvovirus (Bloom et al., 1994). Autonomous parvoviruses utilize host cellular machinery, such as DNA polymerase, for genome replication. Cellular DNA polymerase is expressed during mitosis and therefore, autonomous parvoviruses replicate in the nuclei of the mitotically active cells, which could be in various tissues, depending on the age of the host (reviewed in Steinel et al., 2001). Parvovirus infection of the fetus and neonates at critical stages of organogenesis produces a severe widespread infection, tissue damage and developmental defects, because there is a high rate of cell division at this stage. Replication of parvoviruses in older animals is usually restricted and produces a subclinical infection in most animals, because the rate of cell division has significantly subsided at this stage. However, continuously dividing cells, such as lymphocytes, are susceptible to infection in older animals (MacLachlann & Dubovi, 2011). Severity of the disease induced by AMDV is related to the age and genotype of the animal, as well as virulence of AMDV strain (Hadlow et al., 1983; Alexandersen et al., 1987; Oie et al., 1996).

The AMDV causes mortality in both newborn and adult mink by permissive and persistent restricted infections, respectively. In newborn kits born of seronegative dams, and therefore, lacking maternal antibodies, virulent AMDV isolates permissively infect the type 2 alveolar cells and replicate in lymph nodes, spleen and kidneys. Permissive replication of the AMDV is characterized by low

levels of anti-AMDV antibody in mink kits and high levels of viral DNA in the infected cells. This infection finally leads to acute interstitial pneumonia, which is a severe and fatal respiratory disease. If mink kits are infected with the highly virulent strains, incidence and mortality rates are more than 90%. If mink kits are infected with low-virulent strains, the incidence of disease and mortality rates are 50-70% and 30-50%, respectively. Survivor mink kits develop the adult form of AD regardless of the virulence level of the infecting strain (Bloom et al., 1994).

In adult mink, the outcome of infection is related to the virus strain and genotype of the mink. Infection of adult Aleutian mink with pathogenic and non-pathogenic AMDV strains leads to AD, which is a chronic disorder of the immune system. In the chronic form of AD, AMDV replication is restricted at low levels and occurs in a small population of macrophages and follicular dendritic cells in lymphoid organs. A large number of these cells, however, sequester AMDV. Persistent AMDV replication leads to production of high titers of non-neutralizing antiviral antibodies, followed by enhancement of virus entry into cells and increased AMDV infectivity, a phenomenon known as antibody-dependent enhancement of infection (ADE). The presence of both the antibodies and the virus results in formation of antigen-antibody (immune) complexes. These circulating infectious immune complexes are deposited in tissues, causing tissue damage, including arteritis and glomerulonephritis. Other clinical signs of AD include lymphadenopathy, splenomegaly, plasmacytosis, progressive hypergammaglobulinemia and anemia. Disease severity in non-Aleutian mink

depends on the virulence level of the AMDV isolate (Bloom et al., 1994; Canuti et al., 2015).

2.2 Pathogenicity of AMDV strains

AMDV strains show a wide range of pathogenicity and have been classified into high, low and nonvirulent strains. Highly virulent strains show severe disease in all mink color types (Canuti et al., 2015). Low virulent strains, such as Pullman, cause severe disease in Aleutian-genotypes, such as sapphire mink, and cause low or no mortality in non-Aleutian genotypes, such as pastel mink (Hadlow et al., 1983; Oie et al., 1996). Infection of non-Aleutian genotypes with the highly virulent Utah-1, Ontario and Montana strains results in severe disease symptoms (Hadlow et al., 1983). Other highly virulent strains include TR (Oie et al., 1996), K (Alexandersen 1986; Gottschalck et al., 1991), United (Gottschalck et al., 1994) and BJ (Xi et al., 2016). It should be mentioned that high and low virulent strains of AMDV are not capable of replicating *in vitro.* The non-virulent AMDV-G is a cell-culture adapted strain, which was derived from the Utah1 strain, but has lost its ability to infect mink and only replicates *in vitro* (Bloom et al., 1994). The SL3 strain is distinguished from those discussed above as it both replicates *in vitro* and causes 50% mortality in the Aleutian-genotype violet mink (Haas et al., 1990). Thus, it has been referred to as an intermediate pathogenic strain (Olofsson et al., 1999). Although pathogenicity of the GFAV is not certain in gray fox (Li et al., 2011), the RFAV isolates have reported to induce clinical signs similar to those of

the AMDV (such as enlargement of lymph nodes, chronic diarrhea, and unkempt fur) and thus are likely to be pathogenic for raccoon dogs (Shao et al., 2014).

2.3 Transmission, host range and epidemiology of AMDV

The AMDV is transmitted both horizontally and vertically. Horizontal transmission of this virus can occur via direct contact through saliva, blood, urine and feces, or indirectly through contaminated food, water and environment, such as farm facilities. The AMDV virion is highly resistant to environmental conditions and this stability facilitates the spread of the virus in the wild and among the mink farms. Vertical transmission of the virus can occur from the infected females to their kits (Canuti et al., 2015). Transmission of AMDV could also occur through disease-carrying vectors, such as mosquitoes (Shen et al., 1973). Resiliency of the virus and its multiple routes of transmission are among the reasons why AMDV eradication is difficult.

In addition to the American mink, AMDV infects several members of the *Mustelidae* family, including weasels, skunks, otters, raccoons and bobcats (Farid, 2013). Raccoons, for example, can be infected with highly virulent strains of AMDV (such as TR and Utah1) and can transmit AMDV to mink (Oie et al., 1996). The GFAV has been detected in gray fox and RFAV has been identified in raccoon dogs (Li et al., 2011; Shao et al., 2014). The existence of so many species that can potentially act as a reservoir of the virus has great implications in the fight against AMDV and other *Amdoparvoviruses*.

Epidemiology of infectious diseases involves studying the prevalence and determinants of infections in populations, including the origins and transmission of the disease agents. The ultimate goal of epidemiological studies is the control and prevention of diseases, by preventing pathogen transmission (reviewed in McCormack & Clewley 2002). The AMDV was first detected in farmed American mink in the USA, but it is uncertain whether the virus first originated in wildlife or on farms (Canuti et al., 2015). Farid (2013) hypothesized that because a large number of parvoviruses infect various wild animal species, some isolates of AMDV had been circulating in wild animals prior to the start of mink farming in North America.

The prevalence of AMDV in free-ranging and farmed mink in Canada and other countries has been recently summarized by Canuti et al. (2015). In Nova Scotia, up to 93.3% (from 2009 to 2011) and 70.7% (in 2003) of the wild and farmed mink, respectively, were infected with AMDV or carried anti-AMDV antibodies. In Ontario, 25% to 38% (from 2005 to 2009) of wild mink, 61.2% (early 1970s) of wild mink, 46.3% (early 1970s) of farmed mink and 14% to 60% (from 1986 to 2006) of farmed mink were positive for AMDV DNA or antibodies against AMDV. The AMDV is also prevalent in free-ranging mink in European countries, including Denmark (up to 45.1%, from 1998 to 2009), Estonia (14.8%, from 2007 to 2010), Finland (54.4%, from 2006 to 2014), France (22.7%, from 1991-2001), Iceland (3.6%, 1980s), Spain (33%, from 1997 to 1999) and Sweden (46.6%, from 2004 to 2009). Evidence of AMDV prevalence has been found in 3% to 60% (from 1980 to 2014), 80% (in 2006) and 5% (in 2001) of the mink farms in Finland, Ireland

and Denmark, respectively. There is limited information on epidemiology of other *Amdoparvoviruses.*

2.4 Genome structure and viral proteins

Members of the family *Parvoviridae* (parvoviruses), are among the smallest viruses. Parvoviruses have a non-enveloped icosahedral capsid covering one linear single-stranded DNA (ssDNA) molecule, approximately 4.0 to 6.3 kb in length. The genome of parvoviruses encompasses two large non-overlapping Open Reading Frames (ORF) and some of them also have two or three minor ORFs in the middle of the genome. The REP (replication) ORF is in the left-end (L-ORF) of the genome and encodes non-structural (NS) proteins, which are required for DNA replication and transcription. The right ORF (R-ORF) encodes structural proteins, which form the viral capsid. The viral genome contains short palindromic sequences at both ends that create hairpin structures by folding back on themselves and are used for replication of viral DNA (King et al., 2011).

The AMDV has a small DNA genome, which is approximately 5 kb in length. Similar to other parvoviruses, AMDV genomes contain two large L- and R-ORFs. The L-ORF encodes three non-structural proteins NS1, NS2 and NS3. The R-ORF encodes structural proteins VP1 and VP2. The NS1, NS2 and NS3 proteins share 59 amino acids at their N-terminus but there are unique amino acids at the C-terminus of the protein. The unique regions of the NS2 and NS3 proteins are encoded by the two short middle ORFs (M-ORF) in the center of the genome, (Alexandersen et al., 1988; Bloom et al., 1988; Qiu et al., 2006). The NS1 protein

is the major non-structural protein of parvoviruses, which is mainly localized in the nucleus of the infected cell (Huang et al., 2014) and possess DNA binding, ATPase and helicase activities, which are required for viral DNA replication (Christensen et al., 1995). In addition, cleavage of the AMDV NS1 by caspases has been shown to be critical to viral replication (Best et al., 2003). The NS2 protein, which also localizes in the nucleus, plays roles in AMDV replication. The NS3 protein is not localized in the nucleus, but is also involved in viral replication (Huang et al., 2014). This is evident, because when nucleotide A(1753)T (amino acid 66) and C(1756)T (amino acid 67) of the NS3 gene of the AMDV-G genome were mutated DNA replicative forms were not produced and this prevented production of infectious virions (Huang et al., 2014). The AMDV virion contains a total of 60 capsid proteins, which have the capability of self-assembly into AMDV empty virions in the nucleus of the AMDV-infected cells. The capsid contains 90% of the major capsid protein, VP2, and 10% of the minor capsid protein, VP1. The VP1 and VP2 proteins have identical sequences, with the exception of 42 unique amino acids at the N-terminal part of the VP1 (Clemens et al., 1992; Christensen et al., 1993). The capsid proteins have short peptide sequences (amino acids 428 to 446 of the VP2) in their flexible loop regions which are targets of immune system (Bloom et al., 2001). In addition, pathogenicity, host range and persistent infection of the AMDV are controlled by interaction of some amino acids in the capsid proteins (Bloom et al., 1993; Bloom et al., 1998; Cheng et al., 2010).

2.4.1 Caspase recognition sites in the NS1 and VP2 proteins

The NS1 is a cytotoxic protein which causes apoptosis in the early stages of AMDV infection. Caspases cleave NS1 at two different sites, following aspartic acid 227 (D227) and 285 (D285). The NS1 cleavage is necessary for translocation of this protein from cytoplasm to the nucleus of infected cells, where NS proteins regulate viral genome replication (Best et al., 2003). While conservation of aspartic acid at both cleavage sites is required for the most efficient viral replication, the presence of aspartic acid at either of the cleavage sites leads to some virus replication. The N-terminal caspase-cleaved fragment (residues 1 to 227) localizes both in cytoplasm and nucleus of the host cell. The C-terminal caspase-cleaved fragment (residues 286 to 641) localizes exclusively in the nucleus and functions as a chaperone of full-length NS1 by binding to it and transferring it to the nucleus; as well, it possibly has roles in viral replication (Best et al., 2003). It is suggested that expression of the AMDV capsid proteins induces production of caspases by the host cell, which in turn, cleave the aspartic acid 420 (D420) of the capsid proteins, thereby decreasing the number of capsid proteins which are required to encapsidate the viral genomes. Low levels of capsid proteins lead to regulating persistent infection of the AMDV (Cheng et al., 2010).

2.4.2 Determinants of *in vitro* and *in vivo* replication in VP2 protein

Replication of the AMDV-G *in vitro* was shown to be regulated by amino acid residues 50 to 226 of the VP2, because exchanging of this region of AMDV-G with that of the Utah1, which is unable to replicate *in vitro*, lead to the resulting chimeric virus to be replication defective for CRFK cells (Bloom et al., 1993; Bloom et al., 1998). Therefore, this region, contained a viral host range determinant for CRFK cells. In this region, amino acid residues H92, Q95 and Y115 are consistently different between AMDV-G and all the pathogenic strains, except for SL3 and TR (Table 2.1) (Bloom et al., 1998). The SL3 strain, which replicates both *in vitro* and in mink, shares the same three amino acids with AMDV-G (Haas et al., 1990; Schuierer et al., 1997). The TR strain is similar to AMDV-G in the first two of these amino acid positions and hence, it has been hypothesized that the TR strain can also replicate *in vitro* (Oie et al., 1996), but this possibility has not been tested.

Replication of AMDV in mink is governed by a short sequence in the VP2 gene, from map unit (mu) 69 to 88, corresponding to nucleotides 3272 to 4176 of the AMDV-G genome. In this region, five amino acids at positions 352, 395, 434, 491 and 534 of the VP2 protein are divergent between AMDV-G and pathogenic strains, such as Utah1 and ZK8 (Bloom et al.,1998) (Table 2.1). In addition, the highly pathogenic TR strain (Oie et al., 1996) was compared with other pathogenic strains and was found to have a high sequence similarity to them in this region, confirming that this region is involved in *in vivo* replication of AMDV.

Table 2.1. Determinants of *in vitro* and *in vivo* replication in VP2 protein

	aa[£] determinants of *in vitro* replication			aa determinants of *in vivo* replication				
Strain	92	94	115	352	395	434	491	534
AMDV-G	H	Q	Y	I	H	N	N	H
SL3	H	Q	Y	I	Q	N	N	H
TR	H	Q	F	V	Q	N	D	D
Utah1	A	K	F	V	Q	H	E	D
Pullman	A	P	F	V	Q	E	E	D
ZK8	A	S	F	V	Q	H	E	D

Adopted from Bloom et al. (1998).
[£]aa: amino acid.

2.4.3 Hypervariable regions in the NS1 and VP2 proteins

A hypervariable region (HVR) is a segment of a genome with a remarkably high number of sequence differences among strains within a species. To date, one and four regions of higher variability have been identified in the NS1 and VP2 proteins, respectively. Sequence comparison of the R-ORF of the AMDV-G, Utah1, TR, ZK8 and SL3 strains has revealed a region of high variability in mu 64 to 65 of the genome, corresponding to nucleotide 3036 to 3196. In this region, there are 11 variant amino acid residues (amino acids 232-VATETLTWDAV-242) in the VP2 protein, of which eight differ between AMDV-G and Utah1 (Bloom et al., 1988; Oie et al., 1996). Primers for amplification of this region were constructed by Oie et al. (1996) and since then, this region has been analyzed extensively for identification of AMDV strains and phylogenetic analysis (Mañas et al., 2001; Jahns et al., 2010; Jensen et al., 2012; Nituch et al., 2012; Leimann et al., 2015). It was reported that the HVR of the VP2 protein is not linked to pathogenicity, because changing this segment of the non-pathogenic AMDV-G genome with that of the pathogenic Utah1 did not induce pathogenicity in the AMDV-G (Bloom et al., 1993). It is also

shown that this region is not responsible for cell culture replication, because no alteration in the replication of AMDV-G in cell culture was observed by replacing this region of the AMDV-G with Utah1 (Bloom et al., 1993). The biological importance of the HVR of the VP2 protein is not clarified yet.

Amino acid sequence comparison of the NS1 proteins has demonstrated a higher variability in the AMDV protein of different strains (AMDV-G, K, Utah1, and United), than this protein from other members of the autonomous parvoviruses, including Minute virus of mice (MVM), Feline panleukopenia virus (FPV) and Canine parvovirus (CPV) (Gottschalck et al.,1994). The high variability in the NS1 protein of AMDV isolates is more frequent in the N- and C-terminus of the protein, than in the middle region (Gottschalck et al.,1994). Sequence comparison of a short region of the NS1 protein, from amino acid 128 to 239 of the AMDV-G, K, SL3, United, Utah-1 and 35 other AMDV isolates, obtained from clinical samples, has revealed four segments of higher variability within this region. These segments are amino acids 149 to 161, 168 to 179, 208 to 215, and 225 to 228 of the NS1 protein (Olofsson et al., 1999).

More AMDV strains have been sequenced since Olofsson et al. (1999) and Bloom et al. (1988) first identified segments of higher variability in the two major AMDV proteins, NS1 and VP2. However, no effort has been made to analyze the entire NS1 and VP2 protein sequences of all available AMDV isolates to determine regions of high variability. The method of choice for identifying HVRs in a protein is by computing entropy or variability at each amino acid position in the sequence alignment. Entropy calculation has been applied by Troesch et al. (2006) for

identifying novel HVRs in the E2 envelope glycoprotein of the hepatitis C virus (HCV).

2.5 Multiple infection and genetic recombination

Populations of all living organisms, such as humans and animals, are exposed to a variety of microorganisms. Simultaneous infections of an individual with multiple microorganisms is widespread in natural conditions (reviewed by Alizon et al., 2013). The presence of several AMDV isolates in naturally infected farmed mink has been reported in one study (Canuti et al., 2016). Canuti et al. (2016) detected multiple isolates in 5 out of the 12 farmed mink, where each were infected with two or three different AMDV strains. The presence of different strains of AMDV in a single farm has been reported in three studies (Olofsson et al., 1999; Jahns et al., 2010; Canuti et al., 2016). Olofsson et al. (1999) detected several AMDV types on a farm in Sweden, which had 16% and 26% nucleotide and amino acid differences, respectively, at the HVR of the NS1 gene (nucleotides 382 to 587). Jahns et al. (2010) detected multiple types of AMDV in 16 mink from an Irish farm, which had a low average litter size, compared to other farms which had only one AMDV type in the tested animals. Hence, this suggested that multiple AMDV infection was a possible factor for increased severity of the disease. Although the presence of multiple infection is widespread in natural populations (Bordes & Morand 2011), AMDV multiple infection has not been reported in wild animals by those who sequenced the virus (Mañas et al., 2001; Jensen et al., 2012; Nituch et al., 2012; Leimann et al., 2015; Knuuttila et al., 2015; Persson et al., 2015). The

high prevalence of viral infections and co-circulation of various viruses in a geographical region could elevate the chances of multiple infection of the same host and consequently, increase the possibility of genetic recombination (reviewed in Martin et al., 2011). In AMDV, a high prevalence of the virus on farms, ability of the virus to establish a chronic persistent infection in the mink and being highly resistant to environmental conditions, have been proposed as factors facilitating multiple infection in the same individual, increasing the chance of genetic recombination, which consequently enhances genetic diversity of AMDV (Canuti et al., 2016).

In viruses, evolution occurs through point mutation and genetic recombination. Mutation introduces new genetic variants by alteration of the nucleotides. Recombination generates viruses by rearrangement of genome regions which contain several existing variants (chimeric genomes) (reviewed in Pérez-Losada et al., 2015). One type of recombination in ssDNA viruses is homologous recombination, which is replacement of segments of one genome with segments from a homologous genome (Martin et al., 2011). Pérez-Losada et al. (2015) proposed the term “gene conversion” in viruses, instead of “genetic recombination,” because recombination in viruses is non-reciprocal, whereas it is reciprocal in diploid, sexually reproducing living organisms. Non-reciprocal recombination means that the recipient of a genome segment does not donate the replaced region to the donor (Pérez-Losada et al., 2015).

If prevalence of recombinant viruses increases in the host population, they will become the circulating recombinant forms. On the other hand, if the newly

generated recombinant virus is incapable of replication and transmission, it will not survive long enough to become an independent viral isolate (Martin et al., 2011). Viability of the generated novel recombinants will be eventually determined by natural selection. If genetic recombination leads to the generation of DNA recombinants that natural selection favors, it could have important biological implications, including enlargement of viral host range, by adapting to new environments and hosts (cross-species transmission), as well as elevation of viral pathogenesis, by generating mechanisms to evade host immunity (Pérez-Losada et al., 2015). Because the presence of extensive recombinations in ssDNA viruses, including *Parvoviruses*, have resulted in the generation of novel genera, species and strains, as well as generation of circulating recombinant forms, Martin et al. (2011) described the type of recombination in ssDNA viruses as advantageous.

2.6 Methods of molecular evolutionary analysis

Evolution is the process of changes in the genetic makeup of biological populations over successive generations. Sequencing genomes of organisms, such as viruses, provides opportunities for unravelling the information stored in their genomes. The sequence information can be used for many purposes. For example, analyzing the relatedness of genera, species or strains of a virus, using sequence identity and phylogenetic analyses, which could be used for virus classification and identifying the sources of infections in outbreaks. It is also possible to measure selection pressures and detect genetic recombination throughout the genome.

To date, full sequences of two AMDV strains, namely AMDV-G (Bloom et al., 1988) and BJ (Xi et al., 2016) and the complete coding sequences of nine other AMDV isolates have been published, namely LN1, LN2, LN3 (Li et al., 2012), SL3 (Schuierer et al., 1997), Utah1 (Bloom et al., 1988) and M228, M173, M195, WM25 (Canuti et al., 2016). Among the above isolates, WM25 is the only one obtained from free-ranging mink, which was isolated in the province of Newfoundland in Canada (Canuti et al., 2016). Most often, sequences of short regions of AMDV in free-ranging mink have been reported, which ranged between 336 and 530 bp. These sequences include 336 bp (HVR) of the NS1 gene (nucleotides 601 to 922) (Jensen et al. 2012; Nituch et al., 2012; Leimann et al., 2015; Persson et al., 2015), 401-435 bp of the NS1 gene (nucleotides 1827 to 2258) (Knuuttila et al., 2015), 530 bp (HVR) of the VP2 gene (nucleotides 2589 to 3275) (Mañas et al., 2001; Nituch et al., 2012; Leimann et al., 2015), and 365 bp of the VP2 gene (nucleotides 3043 to 3406) (Jensen et al., 2012).

2.6.1 Identifying selection pressures

Natural selection determines whether mutation (e.g. substitution, deletion and insertion) and recombination that arise in the viral genome will become fixed in subsequent generations. Positive selection (also known as adaptive or diversifying selection) favors advantageous genetic variants. Negative selection (also known as purifying selection) eliminates disadvantageous genetic variants. If mutations have no effect on fitness, the viral genome does not experience natural selection and thus, is under neutral selection (reviewed in Lam et al., 2010). In

viruses, a specific sequence in the genome may encode several proteins or possess control elements. Therefore, because there is a high mutation rate in viruses, the majority of mutations are harmful and thus, in general, viruses are under a strong negative selection (reviewed in Hungnes et al., 2000). Positive selection in viruses occurs when the virus experiences radical environmental changes, such as infecting a new host, and this requires rapid genetic changes to adapt to the new environment (Hungnes et al., 2000).

In protein-coding sequences of the genome, nucleotide substitutions are either non-synonymous, which alter the amino acids they encode, or synonymous, which preserve the same amino acid. Because non-synonymous nucleotide substitutions probably influence the function of the resulting protein more than synonymous ones, positive and negative selections act more strongly on non-synonymous mutations and more weakly on synonymous mutations. A measure of natural selection is the difference between the patterns of non-synonymous (dN) and synonymous (dS) nucleotide substitutions, and is denoted as dN/dS. If dN/dS < 1 (or dS/dN > 1), then the rate of non-synonymous changes is slower than synonymous ones, indicating that amino acid changes reduce fitness and negative selection acts more strongly on non-synonymous changes, in order to eliminate such substitutions. If dN/dS > 1 (or dS/dN < 1), then the rate of non-synonymous changes is faster than synonymous ones, indicating that amino acid changes are advantageous and positive selection acts more strongly on such changes. If dN/dS = 1, then the rate of the two types of changes are the same and that particular

genome region must be under neutral selection (Lemey et al., 2009; Lam et al., 2010).

Sequence analysis of the two main ORFs of several parvoviruses, including three AMDV sequences (AMDV-G, SL3, Utah1), displayed the prominence of negative selection in shaping evolution of parvoviruses (Lukashov & Goudsmit, 2001). Similarly, sequence analysis of the HVR of the NS1 gene (nucleotides 587 to 922) in 57 AMDV strains provided evidence for negative selection (Knuuttila et al., 2009). Sequence analysis of 15 *Amdoparvoviruses* (AMDV-G, LN1, LN2, LN3, Utah1, SL3, WM25, M173, M228, M195, RFAV-HCR, RFAV-QARF, RFAV-HSR, RFAV-XQJLR and GFAV) showed evidence of negative selection, indicated by dN/dS of the NS1 ORF, ranging between 0.51 and 0.62, dN/dS=1.3 of the unique region of the NS2 protein, dN/dS=3.6 of the unique region of the NS3 protein and dN/dS of the VP2 ORF, ranging between 0.24 and 0.33 (Canuti et al., 2016).

2.6.2 Phylogenetic analysis

In molecular epidemiology of infectious diseases, the evolutionary relationships among taxa are displayed in the form of a phylogenetic tree or phylogeny using nucleotide or amino acid sequences. Phylogenetic analysis of viral sequences could be used for classifying them into specific groups or clades of sequences, for epidemiologic studies and for determining the origin of viruses. Sequences within each clade have a common evolutionary history. Integration of the evolutionary history of viral sequences with other information (geographical distribution for example) aids in understanding the patterns of virus transmission

and the spread of the virus (reviewed in Hungnes et al., 2000 and McCormack & Clewley 2002).

In phylogenetic analysis, homologous sequences, i.e. related sequences that share a common ancestry, should be used (Lam et al., 2010). Choosing a genome region with an appropriate sequence diversity is imperative in phylogenetic analysis. When analyzing closely related strains, segments of genome with more diversity are preferred, while for distantly related viruses, conserved regions of the genome should be analyzed (Hungnes et al., 2000). Similar to real trees, phylogenetic trees contain branches (Figure 2.1). The point at which two branches join is referred to as a node and represents the most recent common ancestor of all taxa originating from that node. At the end of the branches in the tree, there are leaves, which are referred to as tips. The tips represent the sequences that are being compared (known as taxa). Rotation of the branches around each node does not change the interpretation of the relationship between taxa. A group of sequences that are all descendants of a common ancestor is referred to as a clade. In a phylogeny, the branch lengths (i.e. number of substitutions per site) are drawn proportionally to genetic distances between the sequences in the tree and are shown by a scale bar at the bottom of the tree (Figure 2.1). Evolutionary relationships between sequences represent the branching pattern of the tree and is called the tree topology (McCormack & Clewley 2002; Lemey et al., 2009; Lam et al., 2010).

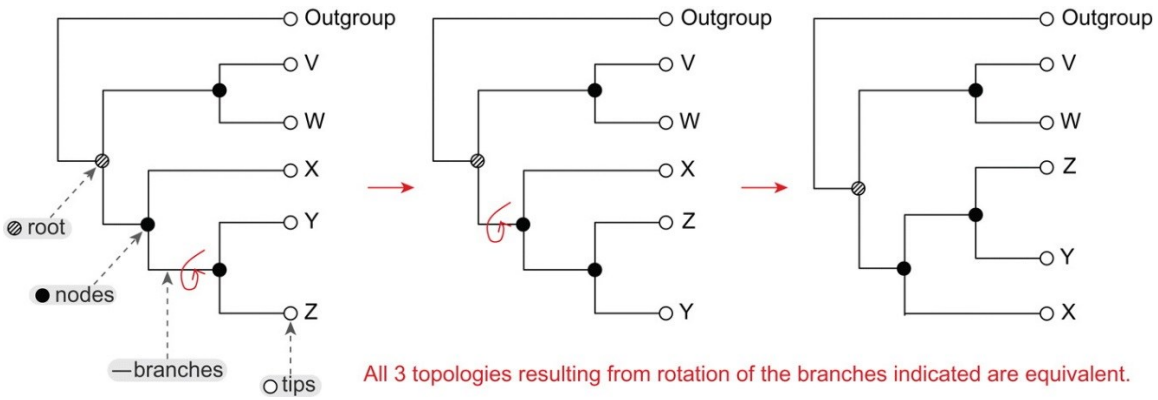


Figure 2.1. The different parts of a rooted phylogeny, representing the root, nodes, branches and tips.
The curved red arrows show interpretation of the relationship among the taxa in the tree remains the same by rotation of the internal single or multiple branches. The root in this tree refers to the root of the ingroup taxa (taxa are V, W, X, Y, and Z) (Lam et al., 2010).

A phylogenetic tree may be either rooted or unrooted. The root of the tree is a node, from which all other nodes descend and hence, is the common ancestor of everything arising from that. Rooted trees can be used for classification of sequences, as well as showing the direction of evolution. Thus, the node closest to the root of the tree is more evolutionarily ancient than the descendant nodes, which are further away. The most widely used method of rooting a tree is to include an outgroup, which is one or more sequences that are distantly related to the sequences in the dataset, and less related to any single sequence in the dataset that they all are to each other (McCormack & Clewley 2002; Lam et al., 2010). Selecting an accurate outgroup is critical because branching patterns of the sequences in the dataset (ingroup) can be altered by the choice of outgroup. It is recommended to include multiple outgroup taxa that have distinct but close

relationships to the ingroup taxa, in order to improve the topological estimate of the ingroup tree (McCormack & Clewley 2002).

No report has evaluated the effect of outgroup selection on tree reconstruction of the entire coding sequence of the genus *Amdoparvovirus*. In a recent review by Simmonds (2015), phylogenetic analysis of a conserved region of the NS1 protein of the family *Parvoviridae*, showed that *Amdoparviviruses,* including AMDV and GFAV, form a clade separate from all other members of the subfamily *Parvovirinae,* i.e *Dependoparvoviruses*, *Erythroparvoviruses*, *Tetraparvoviruses*, *Protoparvoviruses*, *Copiparvoviruses*, *Aveparvoviruses* and *Bocaparvoviruses* (Figure 2.2). Phylogenetic analysis of the complete NS1 protein of the family *Parvoviridae*, reported by Cotmore et al. (2014), however, showed that *Amdoparviviruses,* including AMDV and GFAV, are most closely related to *Protoparvoviruses*, *Aveparvoviruses* and *Bocaparvoviruses* and distantly related to all other members of this family, including *Dependoparvoviruses*, *Erythroparvoviruses*, *Tetraparvoviruses* and *Copiparvoviruses*. Different results is might be because Simmonds (2015) included more sequences and a smaller region in the analysis than Cotmore et al. (2014).

When a proper outgroup is not available, as when all the available taxa are too distantly related to the ingroup and thus the alignment with the outgroup taxa is too ambiguous, an alternate rooting method of mid-point rooting, could be employed. In this approach, the root is located halfway between the two most divergent sequences in the dataset, which leads to a visually balanced phylogeny on either side of the root (Lemey et al., 2009; Lam et al., 2010). Although mid-point

rooted trees show direction of evolution, they assume that all the branches in the phylogeny have similar evolutionary rates. Therefore, they can be used safely when evolutionary rates in different branches are not dramatically different, because in this case a long branch (diverged sequence) might represent faster mutation rates rather than an older branch and thus, misplacing the root at the midpoint of the phylogeny (Lemey et al., 2009). Another method of constructing phylogenetic trees are unrooted trees, which do not indicate the direction of evolution but show clustering of sequences (Lemey et al., 2009; Lam et al., 2010). In the classification of virus families, all three described rooting methods discussed above have been employed, including outgroup-rooting phylogenies (Varsani et al., 2014a; Varsani et al., 2014b) mid-point rooting phylogenies (Varsani et al., 2014b) and unrooted phylogenies (Brown et al., 2015).

Phylogenetic analysis of the complete coding sequence of *Amdoparvoviruses* leads to a more accurate and comprehensive understanding of the evolutionary relationships among them. The majority of phylogenetic analyses have been conducted on partial AMDV fragments (Mañas et al., 2001; Knuuttila et al., 2009; Christensen et al., 2011; Jensen et al., 2012; Li et al., 2012; Nituch et al., 2012; Canuti et al., 2016; Leimann et al., 2015; Knuuttila et al., 2015; Persson et al., 2015). It is not clear, however, whether the phylogenetic trees, based on the applied partial fragments, are compatible with the phylogenies generated from the individual genes or the full AMDV genome sequences. In hepatitis B virus (HBV), for example, phylogenetic analysis of the individual genes (particularly HBV envelope gene) has been considered acceptable, and phylogenetic analysis of the

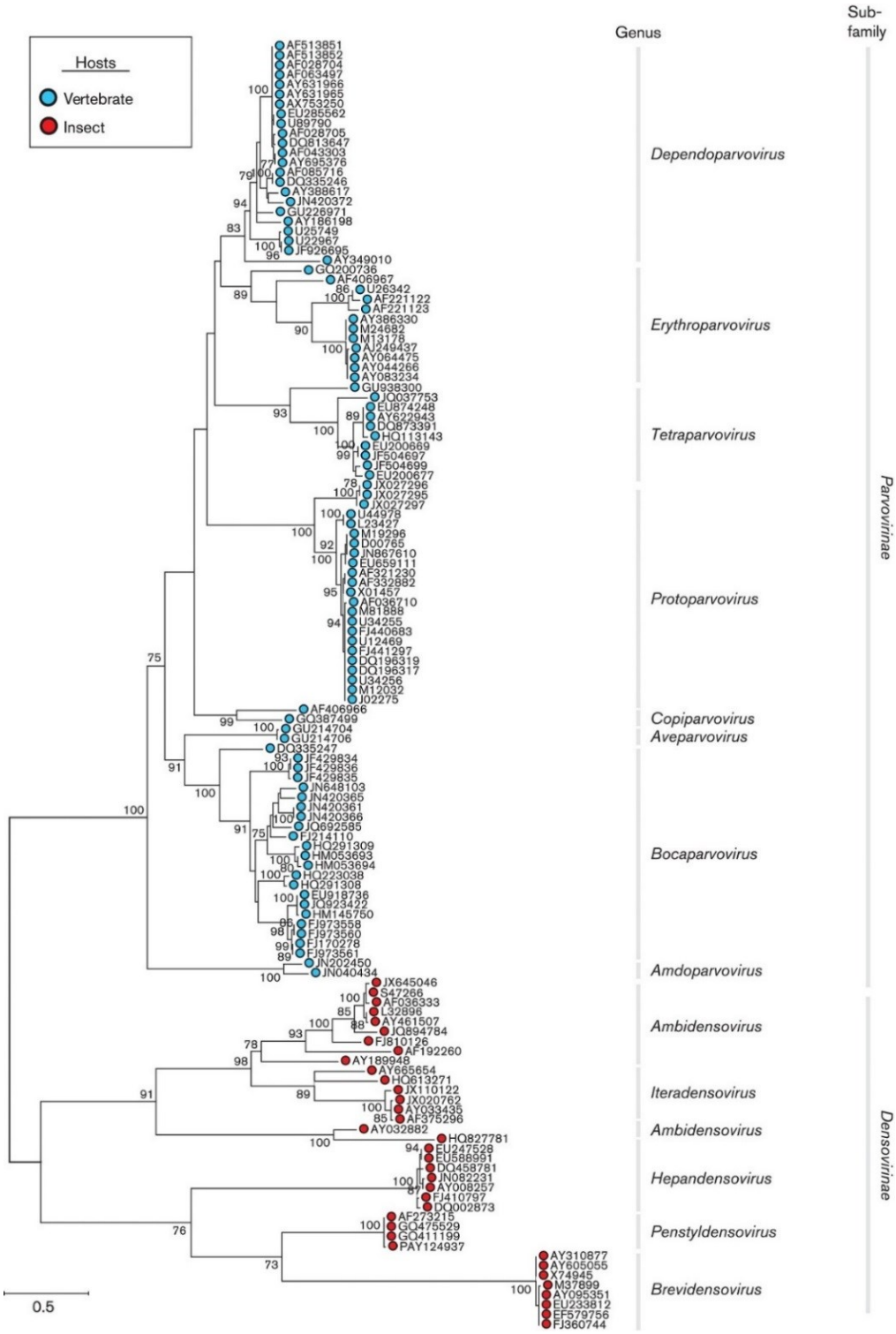


Figure 2.2. Phylogenetic analysis of a conserved region of the NS1 protein of the family *Parvoviridae.*
This phylogeny shows *Amdoparvoviruses* form a clade separate from all other parvoviruses (Simmonds, 2015).

HBV full-genome has been considered the “gold standard” method for HBV genotyping and sub-genotyping (reviewed in Pourkarim et al., 2014). This was because partial genomes were inadequate for classification of a virus like HVB that had a concise genome containing four genes and seven proteins (Pourkarim et al., 2014).

2.6.3 Genetic recombination analysis

Identifying the recombination events in virus genomes can reveal a great deal about their biology and evolution, including molecular epidemiology of recombinant viruses, distribution of recombination breakpoints across the viral genome and the disruptive impact of recombination on the subsequent molecular evolutionary analyses, such as phylogenetic analysis (Martin et al., 2015). Various methods have been developed to detect recombination in the genome of viruses, as well as in other organisms (Martin et al., 2011). The accuracy of recombination detection methods increases when there are several recombination events in the genome and if there is a minimum nucleotide diversity of 5% in the dataset (Pérez-Losada et al., 2015). In the case of datasets containing a low nucleotide diversity, length of the sequences should be increased in order to have a higher number of variable sites for detecting recombination (Martin, 2009).

The Recombination Detection Program (RDP) (http://web.cbio.uct.ac.za/~darren/rdp.html) is a Graphical User Interface (GUI) that accurately detects recombination breakpoints in a dataset, containing up to 2500 sequences with 10 Mb in length. Sequences in a dataset should share higher

than 70% nucleotide identities among each other in order to be analyzed accurately by the RDP software (Martin et al., 2015). The RDP considers every sequence in the dataset as a potential recombinant, without any need for screening the predefined potential recombinants against the predefined potential non-recombinant sequences. Such an approach enables RDP to detect evidence of recombination in all sequences within a dataset and not just in the predefined sequences. In addition, RDP can identify complex recombinants including those that were the result of recombination between parental sequences that were themselves recombinants (Martin et al., 2015). In RDP, the polymorphic sites in every possible combination of three sequences sampled from the dataset are tested in order to identify one recombinant and two parental sequences, as well as the positions of recombination breakpoints positions (Martin et al., 2015).

A few studies have investigated the evidence of recombination in *Amdoparvoviruses* (Knuuttila et al., 2009; Li et al., 2012; Shackelton et al., 2007; Canuti et al., 2016). Knuuttila et al. (2009) used the software SplitsTree4 to construct phylogenetic networks and found no evidence of recombination in the HVR of the NS1 gene (nucleotides 587 to 922) of 54 AMDV isolates. This was possibly because of the short fragment analyzed. Li et al. (2012) conducted phylogenetic analyses on the complete NS1 and VP2 genes of eight AMDV isolates and detected discordant clustering, which means there were differences in the positions of sequences in the two phylogenies related to each other, and is an evidence of genetic recombination. This study, however, did not report the potential recombination breakpoints in the viral genome. Shackelton et al. (2007)

analyzed a short region (1533 bp) of the conserved VP2 gene (nucleotides 2643 to 4175), of 14 AMDV isolates, using RDP, and detected two recombination breakpoints. Canuti et al. (2016) analyzed the entire coding sequence (containing both main ORFs) of 15 *Amdoparvoviruses* (including 10 AMDV, four RFAV and one GFAV), using RDP, and found 11 recombination events. In order to evaluate the results using phylogenetic analysis, the sequence alignment was cut at the approximate nucleotide positions of four of the identified recombination breakpoints. Subsequently, the five sub-alignments were used to construct five phylogenetic trees. Observation of discordant topologies between the NS1 and VP2 ORF phylogenies was suggested as a possible proof of recombination.

2.6.4 Impacts of genetic recombination on phylogenetic estimation

Genetic recombination creates different evolutionary histories throughout the viral genomes. When recombination is present in the dataset, phylogenetic trees based on different segments of genomes, show incongruent topologies because their evolutionary histories are different (Lam et al., 2010). When trees based on different parts of genome are incongruent, a single tree cannot show the true relationships of the viral isolates and thus, is considered unrealistic. It is imperative to analyze recombination in virus genomes prior to phylogenetic analysis in order to decrease the disruptive influence of recombination on the results (Hall & Barlow, 2006). Martin et al. (2015) suggested the exclusion of the recombinant sequences from the dataset prior to phylogenetic analysis or the removal of those regions between any breakpoints from sequences prior to

analysis. Ané (2011), however, suggested that only a portion of the recombination events alter the tree topology and these are more important to detect when constructing species tree.

2.7 Virus classification

2.7.1 Taxonomic classification of the family *Parvoviridae*

According to the current classification of viruses reported by the International Committee on Taxonomy of Viruses (ICTV), the *Parvoviridae* family is classified into two subfamilies: the *Parvovirinae* and the *Densovirinae*, which infect vertebrate and arthropod hosts, respectively (King et al., 2011). Within the *Parvovirinae*, several genera, including *Amdoparvovirus*, *Aveparvovirus*, *Bocaparvovirus*, *Copiparvovirus*, *Dependoparvovirus*, *Erythroparvovirus*, *Protoparvovirus* and *Tetraparvovirus*, have been proposed (Cotmore et al., 2014). The species within the genus *Amdoparvovirus* include *Carnivore amdoparvovirus 1*, which contains AMDV (Bloom et al., 1988; 1990), *Carnivore amdoparvovirus 2*, which contains gray fox amdovirus (GFAV) (Li et al., 2011) and the proposed *Carnivore amdoparvovirus 3*, which contains raccoon dog amdoparvovirus (RFAV) (Shao et al., 2014).

The responsible organization for the classification and nomenclature of viruses at the taxon levels is ICTV (King et al., 2011). Taxon levels that are currently understood by the ICTV are orders, families, subfamilies, genera and species, from the highest to the lowest. In order for a viral agent to be officially assigned to the family *Parvoviridae*, the complete coding sequence, and not the

whole genome, must be identified. This is because sequencing the secondary structures in the viral hairpin telomeres is challenging (Cotmore et al., 2014). For classifying viruses in this family, complete amino acid sequence of the NS1 protein should be available for phylogenetic analysis since the highly conserved domains encoded by the NS1 protein, facilitate a reliable amino acid sequence alignment, which is essential for phylogenetic analysis (Cotmore et al., 2014). Capsid protein analysis generates similar, but less reliable phylogenetic results, compared with the NS1 protein, and thus phylogenies based on capsid protein supported the analysis (Cotmore et al., 2014).

Specific nucleotide identity classification thresholds have been proposed for strain assignment into a species and for species assignment into a genus (Cotmore et al., 2014). As proposed by Cotmore et al. (2014), the NS1 proteins of virus strains within a species should share more than 85% amino acid sequence identity with each other, as determined by pairwise sequence alignments, while diverging by more than 15% from viruses in other species in the same genus. In addition, virus strains within a species should form an independent lineage in the phylogenetic tree (Cotmore et al., 2014). Although divergence threshold of 15% was proposed for species classification threshold in the *Parviviridae* family, it is proposed that there is no uniform species classification for different genera within most virus families, such as *Parvoviridae and Flaviviridae* (Simmonds 2015). Similarly, Brown et al. (2015) suggested that the threshold for species classification differs for each genus in the family *Geminiviridae*. In addition, because pairwise distances and phylogenetic analyses are affected by the genetic recombination, it

is suggested that virus classification should be conducted based on regions of genomes free of recombination (Simmonds 2015), but this criterion was not included in the proposal by Cotmore et al. (2014). For instance, ignoring recombination in phylogenetic analysis of the HBV has been suggested as the main cause of misclassification of this virus (Pourkarim et al., 2014). It is also suggested that inclusion of recombinant HBV sequences can alter the branching pattern of the phylogeny (Shi et al., 2012).

2.7.2 Classification of the viruses in the genus *Amdoparvovirus*

Phylogenetic analysis of the NS1 protein of 18 *Amdoparvoviruses* has shown that these viruses cluster in three clades, containing AMDV, RFAV and GFAV sequences (Canuti et al., 2016). In this analysis, RFAV sequences were more closely related to AMDV and were located in a clade between AMDV and GFAV sequences. AMDV sequences had less than 76.2% and 66.6% amino acid identities with those of RFAV and GFAV sequences, respectively (Canuti et al., 2016), confirming that RFAV and GFAV are different species from AMDV. AMDV sequences were clustered in two clades, of which one contained K strain and the other contained LN1, LN2, LN3, United, Utah1, SL3, G, WM25, M172, M228, M195 and MC42.1 isolates, indicating that K is a distinct isolate. This study also showed that AMDV isolates in Newfoundland (WM25, M172, M228, M195, MC42.1.1) were closely related to some published sequences, such as Utah1. Phylogenetic analysis of the VP2 gene showed a discordant topology with that of the NS1 gene, suggesting the presence of recombination in AMDV genome (Canuti et al., 2016).

CHAPTER 3. MATERIALS AND METHODS

3.1 Animal specimens

Spleen samples from 53 free-ranging American mink tested positive for AMDV DNA by PCR were available for this study. These spleen samples were collected from mink which were trapped in seven counties across Nova Scotia, including Colchester, Cumberland, Halifax, Kings, Lunenburg, Pictou and Yarmouth, between November 2007 and February 2011, and stored at −80 °C (Farid, 2013). Sampling dates and locations of 34 of these mink which were used in this study are presented in Table 3.1.

3.2 Viral nucleic acid extraction

Cell-free homogenates were prepared by cutting 0.25 g of spleen tissue into small pieces with sterile scissors and adding 50 µL of sterile phosphate buffered saline (PBS) (Sigma-Aldrich, St. Louis, MO, USA). The tissue was then homogenized with a battery-operated Kontes grinder (VWR) until it became a uniform paste. Another 700 µL of PBS was added, the mixture was briefly vortexed and centrifuged at 16,000 × g (Eppendorf 5415C, Hamburg, HH, Germany) for 10 min. DNA was extracted from 200 µL of cell-free supernatant, using the Dynabeads Silane viral nucleic acid extraction kit (Invitrogen, Burlington, ON, CAN), according to the manufacturer's protocol. In brief, 50 µL of proteinase K (14-22 mg/mL, Roche, Laval, Quebec) was added to 200 µL of cell-free media and mixed. Lysis buffer, isopropanol and Dynabeads were added to the mixture and Dynabeads were collected by a magnetic rack (DynaMag, Invitrogen). The supernatant was

Table 3.1. Information on 34 mink samples used for AMDV sequencing

Sample ID[£]	County	Township	Sampling date
CO1	Colchester	Brookfield	Feb. 2008
CO2	"	Brule	Nov. 2010
CO3	"	Lower Five Island	Nov. 2010
CO4	"	"	Nov. 2010
CO5[¥]	"	Brookfield	Nov. 2007
CU1	Cumberland	Linden	Feb. 2010
CU2	"	Pugwash Junction	Mar. 2010
CU3	"	Southampton	Nov. 2010
CU4	"	Southampton	Nov. 2010
CU5	"	Conn's Mills	Nov. 2011
CU6	"	"	Nov. 2011
CU7	"	"	Nov. 2011
CU8[¥]	"	Nappan	Nov. 2010
HA1	Halifax	Middle Musquodoboit	Jan. 2008
HA2	"	Musquodabit Harbor	Jan. 2008
HA3[¥]	"	Middle Musquodoboit	Dec. 2007
HA4[¥]	"	"	Nov. 2007
HA5[¥]	"	Musquodabit Harbor	Nov. 2007
HA6[¥]	"	Middle Musquodoboit	Nov. 2007
KI1	Kings	Canning	Nov. 2007
KI2	"	Lake Ville	Feb. 2011
KI3[¥]	"	Forest Home	Jan. 2010
KI4[¥]	"	Canard	Jan. 2010
LU1	Lunenburg	Mahone Bay	Jan. 2010
LU2	"	New Ross	Jan. 2010
LU3	"	Chester Basin	Jan. 2010
LU4	"	Hebville	Jan. 2010
LU5	"	Baker Settlement	Feb. 2011
PI1	Pictou	Saltsprings	Dec. 2010
YA1	Yarmouth	Yarmouth	Feb. 2008
YA2	"	"	Feb. 2008
YA3	"	Woodstock	Mar. 2010
YA4	"	Eel Brook	Mar. 2010
YA5[¥]	"	Yarmouth	Feb. 2008

[£] Sample IDs are based on the first two letters of the county where the animal was trapped.
[¥] Partially sequenced AMDV isolates which are not included in the analysis.

discarded, beads were washed by two washing buffers suspended in 100 µL elution buffer and incubated at 70 ^{o}C for 3 min. Beads were collected by the

magnet and the eluted nucleic acids (100 µL) were transferred to a clean tube and stored at -20 ^{o}C.

3.3 PCR amplification of viral DNA

3.3.1 PCR primer design and optimization

DNA of the 53 samples was initially subjected to polymerase chain reaction (PCR) amplification by the 60F/60R and 70F/70R primer pairs (Table 3.2), which are located on the conserved regions of the VP2 gene. Of the 53 samples, 34 showed strong amplification by at least one of these primer pairs. Subsequently, AMDV DNA of these 34 samples were amplified using six additional primer pairs (140F/40R, 45F/45R, 50F/50R, 55F/55R, 65F/65R and 76F/78R) (Table 3.2). These eight overlapping primer pairs covered the entire AMDV coding region and partial 3' terminal sequence, corresponding to nts 206 to 4615 of the virus (All positions in this study are based on AMDV-G strain, GenBank Accession Number NC_001662). These primers were originally designed based on the AMDV-G genome sequence, using the Oligo Primer Analysis Software version 6 (Molecular Biology Insight, Cascade, CO, USA; http://www.oligo.net/). Sequences of the original primers were later altered when sequence information on viral isolates in Nova Scotia became available. Because of a high degree of nt variation among AMDV isolates circulating on mink farms in Nova Scotia, mixed primers were also designed and optimized when sequences of new isolates became available (Farid, unpublished). These primers were a mixture of two or more primers differing at one or two nts, often at the 5' end, if possible and increased PCR success rate at some

segments of several isolates. In order to amplify as many genomic regions of viral isolates as possible, additional 12 forward and seven reverse primers were designed, as explained above, using sequences of AMDV isolates circulating in Nova Scotia farmed mink and some had to be mixed primers (Table 3.2). Some regions of the virus genome of free-ranging mink were not amplified because of their differences between the primer sequences and the target sites. A combination of many of the primers were tested on these individuals and some were successfully amplified at least one isolate (Table 3.2). In addition, four reverse primers (156R, 157R, 158R and 159R) were designed, based on the divergent regions of AMDV sequences in free-ranging mink (Table 3.3), using Primer3 software version 4 (http://bioinfo.ut.oo/primer3-0.4.0/).

3.3.2 Primer preparation and optimization of PCR conditions

Primers were synthesized by a commercial company (Sigma-Aldrich, Oakville, ON, Canada). The information on molecular weight and the amount of each primer (provided by the vendor) was used to make a 10X stock solution (0.5 mM) in Tris-HCl (pH 7.4), aliquoted and stored at -80 °C. Working solutions (0.01 mM) were prepared by diluting the stock solution with nuclease-free water (Qiagen, Toronto, ON, Canada), when needed, and stored at -20 °C. Because DNA is not very stable at low concentrations, this two-step dilution procedure was used to improve the shelf-life of the primers.

Table 3.2. Primer pairs for PCR amplification of AMDV and sequencing

F primer[£]	R primer[£]	Position nt[¥]	Amplicon size, bp	nt sequence (5'- 3')[†]
145F		176		CAAAGCACAGACCGGTTAC
	142R	929	753	CAAYAATSCCACCGTTACC
	41R	1435	1259	AAGGTATTHTTCATYAAGTCCC
	47R	1856	1680	TGGTTTTCATGCAACGTAT
140F[‡]		206		ATGGCTCAGGCTCARMTTG
	40R[‡]	1418	1212	GTCCCATGTYTTTTATAGTTGC
	41R	1435	1229	See above
	45R	1977	1771	TACCAAYRGCACTTACCT
	50R	2377	2171	ACYGCAGGGTTAGTTTG
	55R	2914	2708	TGYTTGGTAGATGCGTTAC
142F		206		ATGGCTCAGGCTCAAATTGTG
	142R	929	723	See above
36F		724		TCCTGAAGATAGAGCTAAGAAC
	50R	2377	2171	See above
40F		902		TTTRCTGCTGGTAACGGT
	40R	1418	516	See above
48F		1242		AAAGTGACAGATACCTTGAACT
	48R	1862	620	GGTTGATGGTTTTCATGC
45F[‡]		1260		AACTATCTTTAGAACCAAACGG
	45R[‡]	1977	717	See above
	60R	3302	2042	YCCCAAGCAACGTGTACT
46F		1265		TCTTTAGAACCAAACGG
	47R	1856	591	See above
	49R	2145	880	GCAGTTTTCCGTGTTC
49F		1732		TAAAGGCTGTGTGATTGTAA
	50R	2377	645	See above
	55R	2914	1182	See above
50F[‡]		1803		YTCACGCAGARCCACTTAAACA
	50R[‡]	2377	574	See above
52F		2191		AARAGACCTCGGCATGA
	55R	2914	723	See above.
51F		2195		GACCTCGGCATGAGTAR
	55R	2914	719	See above

[£] F = Forward primer; R = Reverse primer.

[¥] NS1 spans from nt 206 to nt 2211; VP1/VP2 span from nt 2204/2406 to nt 4349.

[†] IUPAC ambiguity codes which were used in these primers are underlined including H = A/C/T, M = A/C, R = A/G, S = G/C, W = A/T and Y = C/T.

[‡] Eight original primer pairs.

Table 3.2. Continued.

F Primer	R primer	Position Nt	Amplicon size, bp	nt sequence (5'- 3')
55F‡		2283		TTAGTTCCTCAGCACTATCCTG
	55R‡	2914	631	See above
	78R	4616	2333	GCATACATTWRGCCATAGT
60F‡		2771		GGGTGTATGGATGAGTCCAAA
	60R‡	3302	531	See above
	70R	4209	1438	GCAYGTTACTTGGCTTAGTTTG
	78R	4616	1845	See above
63F		2891		CCAAGGTAACGCATCTAC
	63R	4545	1654	TCCWACATCAGTATATCAAAGC
65F‡		3235		GGCTTRTATGAGTTTAASAGTA
	65R‡	3790	555	CTTCTTCCCAYGAGTCT
66F		3285		AGTACACGTTGCTTGGGGCTAC
	65R	3790	505	See above.
	63R	4545	1260	See above
70F‡		3647		ACGAGGTAGACCTATTAGATGG
	70R ‡	4209	562	See above
72F		3645		AACGAGGTAGACCTATTAGA
	70R	4209	564	See above
76F‡		4029		GAACAACAACGCTCCATTTGTA
	78R‡	4616	587	See above

Table 3.3. Primers designed based on the AMDV sequences in free-ranging mink

Primer name£	Position	nt sequence (5'- 3')
156R¥	2906	AGATGCGTTACCTTGGTTGG
157R	2906	GCTTGCGTTACCTTGGTTGG
158R	2943	GTAACGACGCAGTTAAGTCA
159R	2946	CCTGTAACGACGCAGTTAAGTC

£ These reverse primers worked with 140F, 145F, 40F, 45F, 50F and 55F.
¥ R = Reverse primer.

Optimum conditions for the amplification of each primer set were determined using various annealing temperatures (initially 8 temperature values from 50 to 62 °C), using a gradient thermal cycler (Eppendorf Master Cycler, Hamburg, Germany). DNA of five AMDV-infected animals was mixed for

optimization. Based on the results of this initial step, narrower ranges of temperatures were tested, as necessary, to find the optimal conditions for the PCR amplification. PCR products were tested on a 2.0% agarose gel to assess the amplification success. A single bright band is required for obtaining clean DNA sequences.

3.3.3 PCR conditions

DNA amplification was carried out in 15 µL total volumes, containing final concentrations of 0.1% Tween 20, 0.2 mM each dNTPs (Roche, Mississauga, ON, Canada), 400 nM each primer, 0.75 unit of TaKaRa LA Taq DNA polymerase (Clontech laboratories, Mountain View, CA), 1X LA PCR buffer II (containing Mg_2) and different DNA volumes (see below). The TaKaRa LA Taq enzyme is a high fidelity enzyme with a 3' to 5' exonuclease activity (http://www.clontech.com/US/Products/PCR/Long_PCR/LA_Taq_DNA_Polymerase). This enzyme minimizes incorrect incorporation of nts during PCR amplification. It should be mentioned that this enzyme showed a higher rate of successful amplification and produced a higher concentration of PCR product, compared with Taq polymerase from another company.

Amplification was performed in a Thermal Cycler (C1000, Bio-Rad Laboratories, Hercules, CA, USA), using an initial denaturation at 95 °C for 3 min, followed by 34 cycles of denaturation at 94 °C for 30 sec, annealing at 58 °C for 30 sec, and extension at 72 °C for 2 min for large fragments (>1250 bp) and for 30 sec for short fragments, with a final extension at 72 °C for 6 min. All PCR reactions

included a positive control containing DNA isolated from a known AMDV-infected mink and negative controls containing a blank reaction. The amplified products were run on 1% agarose gels (Agarose 1, Amresco, Solon, OH, USA), stained with ethidium bromide and visualized under the UV light. To prevent contamination during the entire process, sterile filter pipette tips were used and sample preparation, DNA extraction, PCR mixture preparation, PCR amplification and gel electrophoresis were performed in four laboratories with unidirectional sample movement.

3.3.4 DNA preparation for PCR amplification

Initially, three DNA volumes, 1.5, 2.5 and 3.5 µL, were used in 15 µL total PCR volume. This approach was applied because the concentration of extracted DNA was too low to be measured accurately by a spectrophotometer (Farid, 2013). In all three PCR reactions, high levels of smears were observed in PCR products, which could cause poor sequencing results. Subsequently, all DNA samples were two-fold serially diluted five times (1, 1/2, 1/4, 1/8, 1/16) and three volumes (1.5, 2.5, 3.5 µL) from each dilution series were tested by primer pair 60F/60R to determine the optimum DNA volume for amplification of each sample. The optimum DNA volume was 2.5 µL of DNA at 1/4 dilution for amplifying all samples, except for two samples which were amplified using 2.5 µL of 1/8 dilution, due to having high DNA concentrations.

3.3.5 PCR amplification for sequencing

The amounts of PCR product needed by the sequencing facility for fragments less than 1.0 kb, 1.0 to 2.0 Kb and 2.0 to 4.0 kb were 50 ng, 50-100 ng and 100-150 ng, respectively, each in 7μL volume. Two PCR reactions were performed for all fragments less than 1.0 kb and three reactions for the longer ones. Reactions were tested on agarose gels and concentrations of those that produced a single sharp band were measured by a spectrophotometer (Nanodeop-1000, Nano Drop Technology, DE, USA) in triplicate. In cases where the concentrations of PCR products were not sufficient, additional amplifications were performed. In cases where PCR products were faint, the regions were amplified using a different primer pair.

3.4 PCR product purification and sequencing

PCR products were purified using QIAquick PCR Purification kit (Qiagen), according to the manufacturer's protocol and diluted in 30 μL of nuclease-free water. Quality and quantity of purified PCR products were assessed, as described above. Purified products were sequenced by amplification primers (Tables 3.2 and 3.3), as well as by three sequencing primers (Table 3.4) at the Centre for Applied Genomics, The Hospital for Sick Children (Toronto, ON, Canada), which uses an Applied Biosystems 3730xL DNA analyzer. Primers for sequencing were prepared by further diluting the working solutions by 50% in nuclease-free water, as required by the sequencing company. Each amplicon was sequenced in both directions and sequencing was repeated if the results had low quality.

Table 3.4. Sequencing primers

Primer name	Position	nt sequence (5'- 3')
35F[£]	763	AGATGGACCTACTAAGCCTTAC
26R[¥]	647	AACCTAAGGATTTTTGAACAT
30R	931	GTCAACAATGCCACCGTTACCAG

[£] F = Forward primer.
[¥] R = Reverse primer.

3.5 Molecular cloning

Ambiguous positions were detected throughout the genomes of several samples. In order to resolve the ambiguous positions, molecular cloning of viral isolates was initiated in this study. More clone sequences are needed to resolve the sequences and will be conducted in the future. In seven samples which had between 6 to 31 ambiguous codes, 2708 bp of the AMDV genome (from nt 206 to 2914, containing the entire NS1 gene and 513 bp of the VP2 gene), were PCR amplified and cloned. In one sample with 11 ambiguous bases, 1077 bp of the NS1 gene, from nt 902 to 1977, were PCR amplified and cloned. Information about the samples, primers and length of amplicons is shown in Table 3.5.

Amplicons were purified using the QIAquick Gel Extraction Kit (Qiagen), by running 0.8% agarose gel stained with ethidium bromide and quickly cutting the gel with a scalpel blade under UV light. Purified amplicons were eluted in 30 µL elution buffer and their quantity and quality were measured, as previously explained. DNA concentrations of samples lower than 12 ng/µL were increased by reducing the volume in the Speed Vac concentrator (Savant Instruments, Hicksville, New York, USA). Deoxyadenosine was added to the 3' ends of the purified PCR products to increase their chance of insertion into the T-vector. A-

tailing of the PCR products was performed by incubating 20 µL reaction mixture containing 2 µL of 1X PCR buffer, 4 µL of 1 mM dATP (Roche, Mississauga, ON, Canada) and 1 unit of Taq polymerase (Invitrogen, Burlington, ON, CAN), at 72 °C for 20 min.

The gel purified PCR products were cloned into pCR-XL-TOPO vector (Invitrogen), according to the manufacturer's instructions, with 20 min incubation time. Competent One Shot® TOP10 Escherichia coli cells were transformed with the vector. One hundred µL and 150 µL of the cells were spread on two plates and incubated overnight at 37 °C. For each sample, 16 colonies were picked, re-plated and incubated overnight. Colonies were boiled on a hotplate (Hotplate stirrer, Fisher Scientific, ON, CAN) for 10 min followed by centrifugation at 16,000 g (Sorvall Legend Micro 21 centrifuge, Thermo Scientific) for 5 min, the supernatant was discarded and 250 µL of nuclease-free water was added. For each sample, inserts of two to nine clones were PCR amplified by appropriate primers (Table 3.5) and were bi-directionally sequenced. PCR amplifications, quality and quantity measurements, PCR purifications and sequencing were performed, as described previously.

3.6 Nucleotide sequence editing and assembly

The sequences were checked, using the Chromas Lite version 2.1.1 (http://technelysium.com.au) and Sequencher version 5.1 (Gene Codes Corp., Ann Arbor, MI) and were edited manually. First, low quality peaks, particularly at both ends of each read along with the primer annealing sites, were truncated. Next,

the forward and reverse sequences generated by each primer pair were aligned using Sequencher and chromatograms were used to correct inconsistencies and miscalled nts. In case of double peaks, the “Call Secondary Picks” option of the Sequencher was used to identify and confirm ambiguous bases, where the minor peak was at least 25% of the prominent peak on both DNA strands. The following IUPAC ambiguity codes (Cornish-Bowden, 1986) were generated by the Sequencher: D=A/G/T, K=T/G, M=A/C, R=A/G, S=G/C, V=A/C/G, W=A/T and Y=C/T. Finally, the sequences were aligned in the Sequencher, using an 85% minimum gap and a 20 bp minimum overlap, a consensus sequence was then assembled for each sample and exported as a FASTA-formatted file.

3.7 Amino acid sequence editing

The aa sequences were predicted by the ExPASy (Expert Protein Analysis System) program (http://www.expasy.org). This program assigned an ‘X’ for aa where nts were ambiguous. In order to decrease the number of ‘X’s and increase the accuracy of the aa variable sites and substitution analyses, such aa positions were manually converted to the actual aa in cases where the ambiguous positions coded for a synonymous aa.

3.8 Sequence analysis

3.8.1 Dataset construction and multiple sequence alignment

Nine nt sequence datasets were constructed, including the entire coding region (ECR), ECR and a partial 3' terminal region (near-full genome), CDS of individual genes (NS1, NS2, NS3, VP1 and VP2), and NS1-ORF of the 25 local isolates. All nt sequences within each dataset were manually trimmed to the same size. In addition to the local isolates, corresponding *Amdoparvovirus* sequences were retrieved from GenBank (the National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/), trimmed to the appropriate size and were added to the above datasets for additional analyses. Information on the published AMDV sequences retrieved from GenBank (visited September 2015), which were used in this study, are shown in Table 3.6. Analyses were performed on the sequences of the near-full genome, ECR, non-structural and structural genes and partial 3' terminal region of the 25 local isolates (Group 1), with the addition of six available AMDV GenBank sequences (Group 2) and with the addition of four RFAV and one GFAV sequences (Group 5) (Table 3.6). Non-structural genes of the local isolates and 8 corresponding GenBank sequences (Group 3) and VP2 sequences of the local isolates and 16 corresponding GenBank sequences (Group 4) were also analyzed (Table 3.6). It should be noted that VP1 sequences were found on GenBank only in cases where the ECRs were reported and were therefore part of Group 2 sequences (Table 3.6).

Table 3.5. Samples and primers used for cloning of the local AMDV isolates

ID	PCR primers	Amplicon, bp	Sequenced position, nt	Sequenced region, bp
CO1[£]	145F/158R	2708	C1: 211-2887	2669
			C2: 222-2877	2657
			C3: 220-2882	2665
			C4: 220-2884	2714
			C5: 214-2765	2552
CU1[£]	140F/55R	2708	C1: 254-2859	2605
			C2: 217-2815	2598
			C3: 260-2858	2598
			C4: 218-2855	2619
			C5: 221-1373; 1817-2364	1150; 547
			C6: 220-2848	2619
			C7: 219-2502	2283
CU2	140F/55R	2708	C1: 272-841	570
			C2:1324-1888	565
			C3: 251-839	589
			C4: 289-863	575
			C5:1304-1819	516
			C6: 249-813; 953-1506; 1851-2762	565; 553; 885
			C7: 251-813; 954-1509; 1841-2493	563; 470; 653
			C8: 206-604; 958- 1491; 1855- 2859	399; 533; 1002
			C1: 272-784; 814-1536; 1860-2866	513; 722; 1002
CU4[£]	145F/157R	2708	C2: 216- 2849	2609
			C3: 226- 2864	2602
			C4: 226- 2723	2496
			C5: 225-2403	2267
			C6: 225- 2718	2494
			C7: 221-2756	2499
HA1	140F/55R	2708	C1: 217-1146; 1841-2560	930; 720
			C2: 217-1055; 1843-2488	839; 646
KI1[£]	140F/55R	2708	C1: 217- 2864	2648
			C2: 206-2857	2652
			C3: 968-1384	416
			C4: 244-702; 940-1159	459; 220
			C5: 242-485; 941-1169	244; 229
LU1	40F/45R	1075	C1: 913-1965	1053
			C2: 913-1965	1053
			C3: 913-1965	1053
YA4[£]	140F/55R	2708	C1: 219-2928	2654
			C2: 219-2928	2657
			C3: 249-2928	2627

[£] Samples whose NS1 genes were completely sequenced and were used for analysis.
Refer to Tables 3.1 and 3.2 for information on sample IDs and primers sequences, respectively.

Table 3.6. Published *Amdoparvovirus* sequences used for analysis

Isolate ID[£]	Isolate type[¥]	Region†	Groups[‡]	size (bp)	Country of isolation	GenBank accession no.
AMDV-G	AMDV	ECR	2, 3, 4, 5	4596	USA	NC_001662
LN1	"	"	2, 3, 4, 5	4377	China	GU183264
LN2	"	"	2, 3, 4, 5	4400	China	GU183265
LN3	"	"	2, 3, 4, 5	4400	China	GU269892
SL3	"	"	2, 3, 4, 5	4278	Germany	X97629
Utah1	"	"	2, 3, 4, 5	4403	USA	Z18276
HC-R	RFAV	"	5	4218	China	KJ396348
HS-R	"	"	5	4928	China	KJ396347
QA-RF	"	"	5	4416	China	KJ396349
XJLR	"	"	5	4211	China	KJ396350
GFAV	GFAV	"	5	4441	USA	JN202450
K	AMDV	NS1	3	2086	Denmark	X77084
United	"	"	3	2086	USA	X77085
FIN05/C8	"	VP2	4	1944	Finland	GQ336866
Utah1 Kit	"	"	4	1944	USA	U39015
Far-East	"	"	4	1939	Russia	DQ371395
BEL1	"	"	4	1944	Russia	DQ371395
BEL2	"	"	4	1944	Russia	KJ174161
Rus09	"	"	4	1944	Russia	KJ174162
Rus17	"	"	4	1941	Russia	KJ174164
Rus19	"	"	4	1944	Russia	KJ174159
Rus11	"	"	4	1944	Russia	KJ174158
Rus14	"	"	4	1944	Russia	KJ174160

[£] Patented sequences were not used (accession numbers: JB345270 to JB345272).

[¥] AMDV stands for Aleutian mink disease virus; RFAV stands for raccoon dog and fox amdoparvovirus; GFAV stands for Gray fox amdoparvovirus.

† ECR stands for entire coding region, which spans from nt 206 to 4349 and includes non-structural (NS1, NS2, NS3) and structural (VP1, VP2) genes. NS1 ORF spans from nt 206 to 2211. The first 60 aas (nt 206-384) of the non-structural proteins are common between NS1, NS2 and NS3. The non-overlapped regions of the NS1, NS2 and NS3 proteins, from aa 61-247 (nt 384-1961, 2042-2211), 61-114 (nt 2042-2207) and 61-87 (nt 1737-1821), respectively, are in different frames. VP1 ORF spans from nt 2204 to 4349. The predicted aa sequences of the unique region of the VP1 is from aa 1 to 43 (nt 2204-2213, 2287-2406). Overlapped region of the VP1/VP2 is from aa 44 to 690 (nt 2287-4349) of the VP1 and aa 1 to 647 (nt 2406-4349) of the VP2 proteins. Positions are based on the AMDV-G, Reference Sequence: NC_001662.1.

‡ Group 2 = Group 1 (25 local sequences) plus six published AMDV (AMDV-G, LN1, LN2, LN3, SL3 and Utah1), which their ECR was published on GenBank and were used for analysis of the ECR and individual genes. Group 3 = Group 2 plus two AMDV sequences (K and United isolates), which their left ORF was published. Group 4 = Group 2 plus ten AMDV sequences (BEL1, BEL2, Far East, FIN05/C8, RUS09, RUS11, RUS14, RUS17, RUS19, and Utah1 Kit), which their right ORF was published. Group 5 = Group 2 plus four RFAV and one GFAV sequences, which their ECR was published.

The putative aa residues were generated from nt sequences using the ExPASy Translate Tool and were used to create five aa datasets (NS1, NS2, NS3,

VP1 and VP2 proteins). The nt and predicted aa sequences were aligned by the program Multiple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar, 2004), implemented in the graphical user interface (GUI) version of Molecular Evolutionary Genetics Analysis version 6 (MEGA6) (Tamura et al., 2013) (http://www.megasoftware.net), using the default parameters.

3.8.2 Inferring nucleotide and amino acid compositions

Because MEGA6 did not calculate the frequency of ambiguous codes, the Geneious software version 8.1.6 (http://www.geneious.com) was used to calculate nt and GC content of the ECR of Groups 1 and 2 sequences. The MEGA software was used to calculate the number of nt (in ECR, NS1 ORF, CDS of non-structural and structural genes and partial 3' terminal region) and the predicted aa (in non-structural and structural proteins) variable sites of the Groups 1 and 2 sequences (Table 3.6). Variant nt positions containing IUPAC ambiguous codes (R, M, Y etc.), as well as Xs in the aa files, were manually calculated by screening the aligned sequences, because MEGA6 only recognized variations among standard nts or aas, whereas in cases where all elements at a position were the same, but contained any number of ambiguous codes or Xs, they were not counted as variable sites by this software. The aligned predicted aa sequences of the non-structural (Group 3) and structural (VP1: Group 2; VP2: Group 4) proteins were visually inspected to determine the total number of aa substitutions relative to AMDV-G, unique aa substitutions and nt and aa insertions and deletions (indels).

Unique aa substitutions were those aa residues which appeared in only one isolate at each position in the alignment.

3.8.3 Selection pressure analysis

The existence of selective pressures in the non-structural (Groups 1, 2 and 3), VP1 (Groups 1 and 2) and VP2 (Groups 1, 2 and 4) genes was assessed by calculating the ratio of the non-synonymous (d_N) to synonymous (d_S) substitutions (d_N/d_S), for codon-aligned nt sequences. Average of all pairwise comparisons of d_N and d_S estimates were calculated, using the Synonymous Nonsynonymous Analysis Program (SNAP) (www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html). The IUPAC ambiguous codes were manually replaced by 'N,' as required by the software. Ambiguous positions and gaps in these analyses were excluded from the tally of compared codons by the SNAP software.

3.8.4 Entropy measurement

The entropy profile and entropy values at each of the aa positions in the NS1 protein alignment (Group 3) and VP2 protein alignment (Group 4) were obtained using the Entropy-ONE Web tool (http://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy_one.html). Indels, represented by a dash (-), and mixed aas, represented by an X, were considered

variable sites in the calculation. The entropy profile and values at each aa position were visually inspected and those regions with the following criteria were selected to identify the regions with the highest variability: a) had average entropy values greater than that of the entire NS1 protein; b) were four aas or longer in size; c) contained at least two entropy values greater than 0.9; d) the total number of aas with entropy values greater than 0.5 had to be more than those with values lower than 0.5; e) could not have more than seven aas with entropy values <0.5; f) among the overlapped regions which included all the criteria, those with longer aa lengths were selected.

3.8.5 Pairwise sequence identity and genetic distance analyses

Percentage of pairwise nt identities and pairwise identity frequency distribution plots were constructed by the Sequence Demarcation Tool (SDT) version 1.2 (http://web.cbio.uct.ac.za/~brejnev/). The approach that SDT employed for this analysis was as follows: every pair of sequences was aligned using MUSCLE, the identity scores were computed for each of the sequence pairs, a Neighbor Joining (NJ) phylogenetic tree was constructed for clustering closely related sequences based on identity scores and finally a frequency distribution plot of pairwise nt identities was generated (Muhire et al., 2014). SDT ignores alignment positions containing indels (Muhire et al., 2014). Using SDT was recommended by the author of RDP4 software, which was used for genetic recombination analysis, for including sequences with more than 70% nt identity in the recombination analysis (Martin, personal comm. 2015). According to the SDT

software, ambiguous codes were considered as mismatches if different ambiguous letters were used at the same position in either sequence, or as matches if the same ambiguous letter was used in both sequences (Muhire, Personal communication, 2015).

Classification of sequences was conducted based on a method suggested by Muhire et al. (2013), Varsani et al. (2014a) and Brown et al., (2015), which was coupled with phylogenetic support. Based on the classification approach described in the studies mentioned above, the pairwise identity frequency distribution plots of sequences were used in order to find the troughs, which represented the classification thresholds that would result in classification of virus sequences with a low degree of conflict, i.e. the presence of the lowest number of sequences which could be assigned to two or more clusters. The identified thresholds were then applied to classify full genome sequences of the viruses. Based on the conflict-resolution criteria introduced by these studies, each of the clusters consisted of isolates sharing sequence identities higher than the identified classification threshold, to at least one of the sequences in their group, and did not share sequence identities higher than the classification threshold to any member of the other groups. Sequences which had higher identities than the threshold to at least one member of two different clusters were considered outliers and were classified in the group with which it had the highest sequence identity (Muhire et al., 2013; Varsani et al., 2014a; Brown et al., 2015). The pairwise distances were computed for the NS1 aa sequences using the pairwise distance option of the MEGA6. Distances were the proportion of aa sites at which the two sequences were

different, which was obtained by dividing the number of aa differences by the total number of sites compared (MEGA6 manual).

3.9 Recombination analysis

3.9.1 Recombination breakpoint analysis

Detection of potential recombinant sequences, identification of the likely parental sequences and localization of the possible recombination breakpoints were explored with the Recombination Detection Program (RDP) version 4.16 (Martin et al., 2015) (http://web.cbio.uct.ac.za/~darren/rdp.html) as follows: a dataset containing the ECR of the *Amdoparvoviruses* having more than 70% nt identity (nt identity was measured using SDT) with the 25 local sequences were retrieved from GenBank, according to the recommendations provided in the instruction manual of RDP4 (Martin et al., 2015). This dataset included 31 AMDV, 4 RFAV and 1 GFAV sequences (Group 5). Because RDP4 detects recombination signals in a set of aligned nt sequences, Group 5 sequences were subjected to multiple sequence alignment, as described previously, and were used for recombination analysis. All six recombination detection methods available in RDP4, namely RDP (Martin & Rybicki, 2000), Gene Conversion (GeneConv) (Padidam et al., 1999), BootScan (Salminen et al., 1995), Maximum chi-square test (MaxChi) (Smith, 1992), maximum mis-match chi-square (Chimaera) (Posada & Crandall, 2001), sister-scanning (SiScan) (Gibbs et al., 2000) and 3Seq (Boni et al., 2007), were employed. In the above models, ‘linear sequence’ option (rather than ‘circular sequence’) was selected and significance values were set at $P<0.05$,

with the Bonferroni correction. To reduce the chances of obtaining false-positive results while identifying the most detectable recombination events, a recombination event was considered significant if it reached significance level by at least three methods. The software replaced individual ambiguous codes with a gap character (represented by a dash) and was treated as not being present (Martin, personal comm. 2015).

Potential parental sequences referred to sequences within the dataset that were closely related to those contributing to the larger (major parent) and smaller (minor parent) fractions of the recombinant sequence. When the software could not detect any potential parental sequences in the dataset it introduced the sequence which shared the most recent ancestor to the missed parental sequence in a parenthesis after the word 'unknown' in the RDP results file. Only one recombinant and one of the parental sequences were needed to detect a significant recombination event (Martin, personal comm. 2015).

3.9.2 Recombination hotspot test

Once a set of recombination breakpoints were detected, a recombination breakpoint map comprising positions of the positively detectable breakpoints (i.e. excluding those labeled as "unknown") was compiled by RDP4. A breakpoint distribution plot was then constructed from this map by moving a 200-nt window 1 nt at a time along the length of the map. At each window position, all the positively detected breakpoints falling within the window were counted and the breakpoint counts were plotted at the central window position. Significant breakpoint clusters

were identified as those windows within this plot that had more breakpoint positions than the maximum found in more than 95% or 99% of the randomly conducted plots, as explained in instruction manual of RDP4 (http://web.cbio.uct.ac.za/~darren/rdp.html) (Martin et al., 2015). Recombination breakpoint pair matrix was constructed to locate the breakpoint hotspot pairs across the genome.

3.9.3 Proof of recombination using phylogenetic analysis

To assess if recombination events were phylogenetically supported, two NJ phylogenies were constructed for each of the parental regions for each of the detected significant recombination events, with 100 bootstrap replications using RDP4. Each recombination event was considered phylogenetically supported if the recombinant isolate was closely related to the major parent by being present in the same clade in the major parent phylogeny, but distantly related to the major parent in the minor parent phylogeny, according to the recommendations provided in the instruction manual of RDP4 (http://web.cbio.uct.ac.za/~darren/rdp.html) (Martin et al., 2015). The unknown parents (the parental sequences which were missing) were ignored when scanning the trees as they were the closest parents to the missed possible parents in the dataset and not the possible parents (Martin, personal comm., 2015).

3.10 Phylogenetic analyses of the genus *Amdoparvovirus*

3.10.1 Finding the best substitution models

Complete and partial deletion of gaps and missing nts options of MEGA6 were used to find the most appropriate evolutionary model of Maximum Likelihood (ML) nt substitution for phylogenetic analysis of the ECR, of Group 5. The best-fit model of nt substitution for Group 5 sequences was the General Time Reversible Model + Gamma + Invariant Sites (GTR+G+I), with the lowest Bayesian Information Criterion (BIC) value for both complete (46717.08) and partial (53565.42) deletion options, showing that the presence of ambiguous codes in this dataset did not affect the best model. Complete deletion of gaps and missing nts option was used to find the best-fit model of nt substitution for the NS1 and VP1 ORFs of the sequences. For the NS1 and VP1 nt datasets, the best models were GTR+G, with BIC value of 27113.24, and GTR+G+I, with BIC value of 19318.38, respectively. The option of complete deletion of gaps and missing aas was used to find the best-fit model of aa substitution for NS1 and VP1 sequences. For NS1 and VP1 aa datasets, the best models were Jones Thornton Taylor (JTT) +G with BIC value of 15704.69 and retrovirus and reverse transcriptase (rtREV) +G +I with BIC value of 8978.12, respectively (Tamura et al., 2013).

3.10.2 Phylogenetic analyses of the original sequences

To better understand the molecular relationships among the 25 local isolates and between the local isolates and publically available sequences of the genus *Amdoparvovirus*, phylogenetic analyses were performed on the ECR, NS1

and VP1 ORFs as well as aa sequences of the local isolates and corresponding GenBank sequences. Alignments of the analyzed sequences were used for conducting ML analyses using RAxML (Randomized Axelerated Maximum Likelihood) GUI version 1.3.1. (Silvestro & Michalak, 2012; Stamatakis, 2014) (https://sites.google.com/site/raxmlgui/), by the best models explained above. To assess the confidence level of the branching pattern and the robustness of individual nodes in the tree, 1,000 bootstrap replications were performed. The phylogenetic trees were mid-point rooted, which is placing the root of the tree between the most divergent sequences in the dataset (Lam et al., 2010). RAxML handles ambiguous codes, like 'R', as its actual nt composition. For example, in the case of 'R', RAxML assigns equal likelihoods to both 'A' and 'G'. Gaps are treated as Ns, all four nts are considered for each gap (Felsenstein, 2004). A graphical editor software, FigTree version 1.4.2. (http://en.bio-soft.net/tree/figtree.html), was used for mid-point rooting the tree.

3.10.3 Phylogenetic analyses of the modified sequences

Three sets of analyses were conducted to assess the effects of ambiguous codes, presence of outlier sequences and recombination on the bootstrap values and branching patters. First, to assess the possible effect of ambiguous codes, the aligned nt sequences of the ECR of Group 5, prior to the recombination analysis, were exported from RDP4, which replaced ambiguous codes by gap character (-). This file was saved and imported to RAxML for phylogenetic analysis. Second, to assess the effect of outlier isolates, namely CU5, CU6 and YA3, phylogenetic

analyses were performed on the ECR, NS1 and VP1 ORFs and protein sequences, excluding the four outliers. Third, following recombination analysis, aligned sequences of the ECR of the recombination-free segments of Group 5 was exported from RDP4 to RAxML. The NS1 and VP1 nt sections were cut manually from the alignment and used for subsequent phylogenetic analyses. All phylogenetic analyses were performed as explained in Section 3.10.2.

3.10.4 Identifying a phylogenetic marker

In order to determine a phylogenetic marker, six multiple sequence alignments were constructed by moving a 300-aa window 50 aa at a time along the length of the NS1 protein alignment of the Group 5 sequences, excluding the three outlier isolates, CU5, CU6 and YA3. The regions were aa position 1 to 300, 50 to 350, 100 to 400, 150 to 450, 200 to 500 and 250 to 550 of the alignment. For each sequence alignment, phylogenetic trees were conducted as explained in Section 3.10.2.

CHAPTER 4. RESULTS

4.1 Genome characterization

The nt sequences of the entire coding region (ECR), which contained NS1, NS2, NS3, VP1 and VP2 genes, as well as the partial 3' terminal sequences of the 25 AMDV isolates from free-ranging mink in NS (local isolates), were determined in this study (Table 4.1). Six other samples were partially sequenced, but were not included in the analyses, because none of them completely covered an entire gene. The ECR of the 31 local and GenBank sequences (Group 2) had average nt compositions of 37% A, 18.9% C, 19.5% G and 24.3% T (excluding ambiguous codes). The partial 3' terminal regions of the Group 2 had average nt compositions of 35.6% A, 17.3% C, 11.0% G and 36.0% T, which had a lower GC content (28.3%) than that in the coding regions.

4.1.1 Ambiguous codes and mixed amino acids

Ambiguous codes were detected throughout the sequences of 24 of the 25 local isolates. A multiple sequence alignment showing the positions of ambiguous codes is reported in Table A1 and their frequencies are shown in Table A2. Alignment of the near-full genomes of the 25 local isolates, along with the six sequences from GenBank (GenBank sequences did not have ambiguous codes) showed that 228 positions contained ambiguous codes (5.2% of the 4,363 positions in the alignment) (Table A1). Ambiguous positions were ignored in some analyses by some software (MEGA, RDP, SNAP, STD), whereas they were

included in others (RAxML, Geneious) and thus their impact is related to the type of analysis.

Table 4.1. Positions and lengths of the sequences of the 25 local isolates

Isolate ID[¥]	Near-full genome[†] nt position[£]	Size					
		Coding region bp	Partial 3' terminal bp	NS1		VP1	
				bp	aa	bp	aa
CO1	206-4603	4146	262	2008	641	2146	690
CO2	206-4593	4105	268	2006	641	2107	677
CO3	206-4605	4144	257	2006	641	2146	690
CO4	206-4582	4117	234	2006	641	2119	681
CU1	206-4596	4143	247	2005	641	2146	690
CU2	206-4546	4123	247	2006	641	2125	683
CU3	206-4584	4144	235	2006	641	2146	690
CU4	206-4596	4117	247	2006	641	2119	681
CU5	206-4593	4147	265	2006	641	2149	691
CU6	206-4593	4147	266	2006	641	2149	691
CU7	206-4593	4144	269	2009	642	2143	689
HA1	206-4527	4149	188	2008	641	2149	691
HA2	206-4533	4146	192	2008	641	2146	690
KI1	206-4596	4144	255	2006	641	2146	690
KI2	206-4584	4144	243	2006	641	2146	690
LU1	206-4596	4144	247	2006	641	2146	690
LU2	206-4596	4144	255	2006	641	2146	690
LU3	206-4573	4144	232	2006	641	2146	690
LU4	206-4589	4144	248	2006	641	2146	690
LU5	206-4592	4144	251	2006	641	2146	690
PI1	206-4532	4149	214	2008	641	2149	691
YA1	206-4595	4146	257	2008	641	2146	690
YA2	206-4585	4143	246	2008	641	2143	689
YA3	206-4595	4122	256	2008	641	2122	682
YA4	206-4596	4117	255	2006	641	2119	681

[£] Positions are based on the AMDV-G sequence (GenBank accession number NC_001662).
[¥] Refer to Table 3.1 for isolate IDs
[†] Near-full genome sequence includes the coding region (nt 206 to 4349) as well as the partial 3' terminal region.

Multiple sequence alignment of the aa variable sites in the NS1 protein, unique region of the NS2, unique region of the NS3, unique region of the VP1 and overlapped region of the VP1 and VP2 proteins are presented in Tables A3, A4,

A5, A6, and A7, respectively. The aa residues at those codons containing ambiguous codes, which were represented by an 'X' on the ExPASy output, were visually inspected and 109 of the 111 positions in the non-structural proteins and 90 of the 91 positions in the structural proteins were manually replaced. The majority of codons containing ambiguous codes in the non-structural (89.1% of 111) and structural (59.3% of 91) proteins resulted in non-synonymous nt substitutions (Table A8) and the predicted aa residues were manually inserted and separated by "/" in the alignment (such as I/V) and will be referred to as mixed aas. The majority of the mixed aas generated above (66.7% to 100% in non-structural and 0.0 to 59.6% in structural proteins) were a combination of those were found in other AMDV isolates at the same position in the alignment (combination aas) (Table A8).

4.1.2 Nucleotide and amino acid indels

Indels in different parts of genome of the local and GenBank sequences are shown in Table 4.2. The nt indels caused the length of the coding regions to vary between 4,105 and 4,149 bp. The major cause of variability in genome sizes among local isolates was the indels in the Glycine-rich region located in the overlapping segment of the VP1 and VP2 genes (Table 4.3). Amino acid indels in non-structural and structural prtoteins of the 25 local isolates are shown in Table 4.4. The majority of nt and aa indels were detected in various genome regions of the isolates CO1, CO2, CU5, CU6, CU7, HA1, HA2, PI1, YA1, YA2 and YA3.

Table 4.2. Nucleotide indels in local isolates and GenBank sequences

Isolate ID[α]	Position nt[£]	No. nt	nt	Genome region
Insertion				
CU7	568	3	AGG	NS1
HA1/2, PI1, YA1-3	1970[¥]	1	T	NS1 intron
CO1	1970[¥]	1	C	”
CO1, CU7, HA1/2, LN1/2[*], PI1, YA1-3	2003[¥]	1	A	”
CO1/2, HA1/2, KI2, PI1, YA1-3	2363	3	TCT	VP1u[†]
CU5/6	2363	3	GTT	”
LU2-5, KI1, YA3/4,	4358	8	TATGTTAC	3’ terminal
CO1	4358	10	TACATGTTAC	”
HA1/2, YA2, CO2, CU5/6	4358	10	TATATGTTAC	”
CU7	4358	10	TATATATTAG	
PI1	4358	18	TATATGTTACTGTGTTAC	”
PI1	4496	13	TAACATCTAACTA	”
CO4, CU7	4503	1	A	”
PI1	4503	1	C	”
Deletion				
CU1/7, LN1-3	1971[¥]	1	-	NS1 intron
YA3/4, CO4, CU4, LN1[*]	2472	27	-	VP1/2, GT region
CO2, GFAV[*], XQJLR[*]	2469	39	-	”
CU2	2475	21	-	”
CU7, KI2, YA1/2, LN2/3, HCR[*], QARF[*]	2493	3	-	”
CO1/2, HA2, YA2, HSR[*], XQJLR[*]	3099	3	-	VP1/2, HVR[‡]
Far East[*], RUS17[*]	3111	3	-	”
CO1/2, CU5-7, HA1/2, KI2, PI1, YA2	4358	1	-	3’ terminal
HA2	4493	1	-	”
YA1/3	4493	3	-	”
CO2	4500	2	-	”
CO1	4503	1	-	”
YA1/3	4510	7	-	”
CU5/6	4515	3	-	”

[£] Positions are based on the AMDV-G sequence (GenBank accession number NC_001662). GenBank sequences are indicated by an asterisk. Those indels of GFAV and RFAV that were in at least one of the AMDV sequences are indicated.
[¥] Indels which were located in the region from nt 1961 to 2042 are in the NS1 intron.
[†] VP1u: Unique region of the VP1 gene from nt 2204 to 2406 of the AMDV-G.
[‡] HVR: Hypervariable region.
[α] A dash indicates “to” (i.e. CU5-7: CU5, CU6, CU7) and a slash indicates “and” (i.e. HA1/2: HA1, HA2).

Table 4.3. Sequence alignment of the Glycine-rich region

nt[¥]	2469	2472	2475	2478	2481	2484	2487	2490	2493	2496	2499	2502	2505	2508	2511	2514
G	GGT	GGT	GGG	GGG	GGT	GGG	GGT	GGT	GGG	GGT	GGT	GGT	GGT	GGT	GGT	GGG
Gr1[†]	... [£]	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Gr2	...	...	...	...	...	...	..G	...	..A	...	...	...	..A	...	...	...
Gr3	...	...	...	...	...	...	...	...	..T	...	...	...	...	...	...	...
Gr4	...	...	..A	..A	...	...	...	...	...	...	...	...	...	...	...	...
Gr5	...	...	...	...	...	...	...	...	..K	...	...	...	...	...	...	..A
Gr6	...	...	...	..A	A..[‡]	…	...	...	..A	..G	...	...	..G	...	...	...
Gr7	...	...	...	...	...	..A	...	...	-	..G	...	...	..G	...	...	...
Gr8	...	...	...	..T	...	...	...	...	-	..G	...	...	..A	...	...	...
Gr9	...	...	-	-	-	-	-	-	-	...	...	...	...	...	...	...
Gr10	...	-	-	-	-	-	-	-	-	-	...	...	...	...	...	...
Gr11	-[£]	-	-	-	-	-	-	-	-	-	-	-	-	...	...	...

[£] A dash indicates deletion of the triplicate and dots represent similar to the AMDV-G sequence.
[¥] Positions are based on the AMDV-G sequence (GenBank accession number NC_001662), which is represented as G in this Table.
[†] Gr1: LU1, LU4, LU5, CU5, CU6, CO3, CU3 and LU2; Gr2: CO1 and HA2; Gr3: KI1 and CU1; Gr4: HA1; Gr5: LU3; Gr6: PI1; Gr7: YA1 and YA2; Gr8: CU7 and KI2; Gr9: CU2; Gr10: YA3, YA4, CU4 and CO4; Gr11: CO2.
[‡] Amino acid sequence of the AMDV-G and the 25 local isolates from nt 2469 to 2516 consisted of sixteen Glycine, except for PI1, which contains Serine (AGT) in the fifth amino acid of the Glycine-rich region at position 2481 of the AMDV-G genome.

Table 4.4. Position and types of amino acid indels in the 25 local isolates

AMDV isolates	Genome region	Indels	aa position[£]	No. of aa	Aa
CU7	NS1	In.	121	1	R
CO1/2, HA1/2, KI2, PI1, YA1-3	VP1u[¥]	In.	29	1	S
CU5/6	VP1u	In.	29	1	V
CO4, CU4 YA3/4	VP1/VP2	Del.	71/28-79/36	9	-
CO2	VP1/VP2	Del.	67/24-79/36	13	-
CU2	VP1/VP2	Del.	73/30-79/36	7	-
CU7, KI2, YA1/2,	VP1/VP2	Del.	79/36	1	-

[£] Positions are based on the AMDV-G sequence (GenBank accession number NC_001662).
[¥] VP1u: Unique region of the VP1 aa sequence located from aa 1 to 43 of the VP1 protein.

4.1.3 Nucleotide and amino acid variable sites

Multiple sequence alignment of Group 2 showed that 1,132 of the 4,152 nt positions (25.2%) were variable (Table 4.5). Percentages of variant positions were comparable among the NS1, NS2 and NS3 genes (34.2 to 36.2), but they were almost twice as high as those for the VP1 and VP2 genes. Ambiguous codes were not counted in the estimation of variable sites (43 positions). Multiple sequence alignment of the partial 3' terminal region showed that 63 of the 211 nt positions (29.9%) were variable. The number of aa variable sites of the non-structural and structural proteins of the four groups of viruses are presented in Table 4.6. Multiple sequence alignment showed that 44.8%, 50.0% and 47.1% of the aa positions of the NS1, NS2 and NS3 proteins of the Group 2, respectively, were variable. The VP1 and VP2 proteins of Group 2 had 127 (18.4%) and 118 (18.2%) variable aa positions, respectively. The variable positions in the VP2 containing mixed aa were

not counted by MEGA (8 positions). Comparison of the aas and nts of the NS1 and VP2 proteins in Group 2 isolates revealed that the variation at the aa level (44.8%) was greater than that in the nt level (35.9%) in the NS1 protein, whereas they were comparable (18.2% vs. 20.0%) in the VP2 protein (Tables 4.5 and 4.6).

Table 4.5. Variant nucleotide positions in different genome regions

Genome region	No. of positions aligned	No. and % of variant positions[£]		
		Group 1[¥]	Group 2[¥]	Group3[¥]
Coding region	4152	1042 (25.2%)	1132 (27.3%)	-
NS1 ORF	2011	659 (32.8%)	705 (35.1%)	725 (36.0%)
NS1 CDS	1929	646 (33.5%)	691 (35.9%)	709 (36.8%)
NS2 CDS	345	113 (32.8%)	125 (36.2%)	127 (36.8%)
NS3 CDS	263	83 (31.6%)	90 (34.2%)	94 (35.7%)
VP1 CDS	2076	366 (17.6%)	409 (19.7%)	-
VP2 CDS	1944	345 (17.7%)	388 (20.0%)	438 (22.5%)[†]
Partial 3' terminal	211	60 (28.4%)	63 (29.9%)	-

[£] Percentage of variant positions is based on length of the alignment.
[¥] Refer to the footnotes of the Table 3.6 for information on published sequences and sequences within each of the Groups.
[†] Group 4.

Table 4.6. Variant amino acid positions in all proteins

Protein	No. of aligned aa positions	No. and % of variant positions[£]		
		Group 1[¥]	Group 2[¥]	Group3[¥]
NS1	642	273 (42.5%)	287 (44.8%)	292 (45.8%)
NS2	114	53 (46.5%)	57 (50.0%)	57 (50.0%)
NS3	87	39 (44.8%)	41 (47.1%)	42 (48.3%)
VP1	691	114 (16.5%)	127 (18.4%)	-
VP2	647	105 (16.2%)	118 (18.2%)	143 (22.1%)[†]

[£ ¥ †] Refer to the footnotes of the Table 4.5.

4.1.4 Rates of synonymous and non-synonymous substitutions

Selective pressures on Group 2 sequences were assessed using the ratio of non-synonymous (d_N) to synonymous (d_S) substitutions (d_N/d_S). These ratios were less than 1.0 for all the genes, and in non-structural genes were almost twice as much as those in structural genes (Table 4.7). Ambiguous positions and gaps in these analyses were excluded from the tally of compared codons by the SNAP software.

Table 4.7. Averages of pairwise comparisons of non-synonymous and synonymous substitutions in all genes of the local isolates and GenBank sequences

Region	d_N [£]	d_S [¥]	d_N/d_S
NS1	0.07 (0.00-0.12) [†]	0.20 (0.00-0.35)	0.44 (0.09-1.70)
NS2	0.07 (0.00-0.14)	0.18 (0.00-0.43)	0.47 (0.15-3.52)
NS3	0.07 (0.00-0.16)	0.18 (0.00-0.43)	0.47 (0.10-3.71)
VP1	0.02 (0.00-0.04)	0.13 (0.00-0.25)	0.21 (0.08-1.27)
VP2	0.02 (0.00-0.03)	0.13 (0.00-0.26)	0.21 (0.07-2.53)

[£] Non-synonymous substitutions.
[¥] Synonymous substitutions.
[†] Range of estimates are shown in brackets.

4.1.5 Number of total and unique amino acid substitutions

Total aa substitution for each isolate is the number of aa substitutions relative to the AMDV-G. The average number of total aa substitutions in the NS1 (13.4%), NS2 (12.9%) and NS3 (12.8%) proteins of the 25 local isolates was twice as many as that of the VP1 (4.2%) and VP2 (4.1%) proteins (Table A9). Ten of the local isolates (CO1, CO2, CU5, CU6, CU7, HA1, HA2, PI1, YA1 and YA2) had the

greatest number of aa substitutions in the NS1 (98 to 132), NS2 (15 to 26) and NS3 (12 to 23) proteins. Ten isolates (CO1, CO2, CU7, HA1, HA2, KI2, PI1, YA2, YA3 and YA4) had the greatest number of aa substitutions in the VP1 (31 to 51) and VP2 (27 to 47) proteins. Similar to the GenBank isolate K, the local isolate YA1 carried a CAA to TAA substitution at nt 205 of the NS3 coding sequence, which introduced a premature stop codon at position 69 of the NS3 protein and shortened the resulting protein by 18 aa. The local isolates CO1, CO2, CU7, HA1, HA2, PI1, YA1 and YA2 carried a CAA to TAA substitution at nt 214 of the NS3 coding sequence, which introduced a premature stop codon at position 72 of the NS3 protein (Table A5). This stop codon shortened the NS3 protein by 15 aas.

The average number of aas which appeared in only one isolate at each position in the alignment (unique aa substitutions) in the NS1 (1.2%), NS2 (1.5%) and NS3 (1.3%) proteins of the 25 local isolates was twice as many as that of the VP1 (0.4%) and VP2 (0.4%) proteins (Table A10a). Eight isolates (CO1, CO2, CO4, CU5, CU6, CU7, HA1 and PI1) showed the greatest number of unique aa substitutions in the NS1 (9 to 27) and NS2 (3 to 6) proteins (Table A10a). Among the GenBank sequences, the K isolate had the highest number of total and unique aa substitutions (Tables A9 and A7b). Ten of the local isolates (CO1, CO2, CU5, CU6, CU7, HA1, HA2, PI1, YA1 and YA2) contained between 7 and 16 aa residues in the N-terminus of the NS1 protein which did not exist in any other local isolates or in the eight sequences on GenBank, except the K isolate (Table 4.8).

The aa sequence alignment of the VP2 protein of the local isolates, as well as all the complete and partial AMDV, RFAV and GFAV sequences available on

GenBank (144 partial VP2 nt sequences were translated to the predicted aa sequences and used for comparison, mostly from aa 107 to 280), showed that the AMDV (residue D), RFAV (residue A) and GFAV (residue S) viruses could be distinguished by one aa residue at position 498 of the VP2 protein.

4.1.6 Sequence motifs

The aa residues of the left and right caspase recognition sites in the NS1 protein of the 25 local isolates and eight GenBank sequences (Group 3) are shown in Table A11. The aa residue at position 227 (left caspase cleavage site) was highly variable among the local and GenBank isolates, whereas aa residue at position 285 (right caspase cleavage site) was conserved. The aa residues at the caspase recognition site in the VP2 protein of the 25 local isolates and the 16 GenBank sequences are shown in Table A12. The residue at position 420 (caspase cleavage site) was highly conserved among the 41 sequences studied.

Table 4.8. NS1 amino acid residues shared by the most distinct isolates

Isolate	16[£,¥]	19	63	75	76	109	111	120	122	124	143	160	235	255	263	265
CO1	W	F	Y	Q	C	S	Q	S	K	I	H	V	S	L	G	L
CO2	W	F	Y	N	C	R	E	S	K	I	H	V	S	L	.	L
HA1	W	F	Y	D	C	R	E	S	K	I	H	V	S	L	G	L
HA2	W	F	Y	.	C	S	E	S	K	I	H	V	S	L	G	L
PI1	W	F	Y	D	F	S	Q	S	K	I	H	.	.	L	G	L
CU5	W	F	Y	N	C	H	E	S	K	I	H	.	S	L	G	L
CU6	W	F	Y	N	C	H/R	E	S	K	I	H	.	S	L	G	L
CU7	F	V	.	K	F	.	.	A	K	I	.	.	G	.	.	.
YA1	F	F	.	V	C	K	Q	A	K	I	.	.	S	L	G	L
YA2	F	F	.	V	C	K	Q	A	K	I	.	.	S	L	G	L
K	F	F	.	Q	C	.	.	.	.	.	.	.	S	.	G	L
Others[†]	Y	L/V	F	E	H/N/W/Y	K	N	K	Q	F	Q	I	A	M	D	M
Partial[‡]	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	Q/H	I/V/T	A/S/T/D	n/a	n/a	n/a

[£] aa positions are based on the NS1 aa sequence of the AMDV-G sequence.
[¥] A dot indicates the same aa residue as the "Others".
[†] The other 15 local and 7 GenBank AMDV sequences with complete NS1 sequence.
[‡] Partial NS1 sequences on GenBank. n/a refers to the segments which were not available on GenBank.

Variations at three aa residues (92, 94 and 115) of the VP2 protein, which have been identified as those which control *in vitro* replication were compared among 41 local and published sequences (Table A13). The most notable result was that 11 of the local AMDV isolates (CO3, CU2, CU3, CU4, KI1, LU1, LU2, LU3, LU4, LU5 and YA1) had aa residues similar to eight of the AMDV-G at these three positions.

The five aa residues of the VP2 protein which have been associated with the AMDV pathogenicity are presented in Table A14. The most notable observation was that all the five aa residues in the 25 local isolates were different from those in the non-pathogenic AMDV-G, and 12 of them (CO1, CO2, CU2, KI1, LU1, LU3, LU4, LU5, PI1, YA1, YA2 and YA3) had similar aa residues to the highly pathogenic Utah1 isolate.

4.1.7 Hypervariable regions

The entropy profile of the NS1 protein of the Group 3 sequences is shown in Figure 4.1 and the entropy values are depicted in Table A15. A total of nine HVRs were identified in the NS1 protein, ranging from four to 19 aas in length, of which six were located at the N-terminus and were termed N-HVR1 to N-HVR6 (Table 4.9), and those at the C-terminus of the NS1 protein were called N-HVR7 to NHVR9. The entropy profile of aas in the VP2 protein of the Group 4 sequences is shown in Figure 4.2 and the entropy values are presented in Table A16. The entropy profile showed that aa variation in the VP2 protein was much lower than that in the NS1 protein, indicated by three times greater average entropy value of the NS1

compared with the VP2 (0.27 vs 0.09, Table 4.9). In the N-terminus of the VP2 protein, aa variation was concentrated in two segments, V-HVR1 and V-HVR2 which were 5 and 11 aa long (Table 4.9).

4.2 Molecular evolution

4.2.1 Recombination analysis

Details of significant recombination events within the ECR of the Group 2 plus four RFAV and GFAV sequences (total of 36 members of the genus *Amdoparvovirus*), including recombinant sequences, potential major and minor parents, the breakpoint positions, programs which detected the events and the corresponding P-values are listed in Table 4.10. A total of 27 possible recombination events were detected within 21 (67.8%) of the 31 AMDV. sequences, of which 18 events were significant. The NJ phylogenetic analysis confirmed the occurrence of recombination events between the parental sequences in these 18 cases (Figures A1-18). The two recombination events that were detected in YA1 (event 1) and CU5 and CU6 (event 2) by six recombination detection methods, had the strongest statistical support (P= 3.92E-17). Similar recombination events, i.e., the same nt position and parents, were detected in several isolates. Recombination occurred throughout the genome and between both closely related and distantly related parental sequences. Recombination breakpoint distribution plot of the same dataset (Group 5) which showed the distributions of the positively detectable recombination breakpoints (i.e. excluding those labeled as “unknown”), indicated the presence of two statistically significant

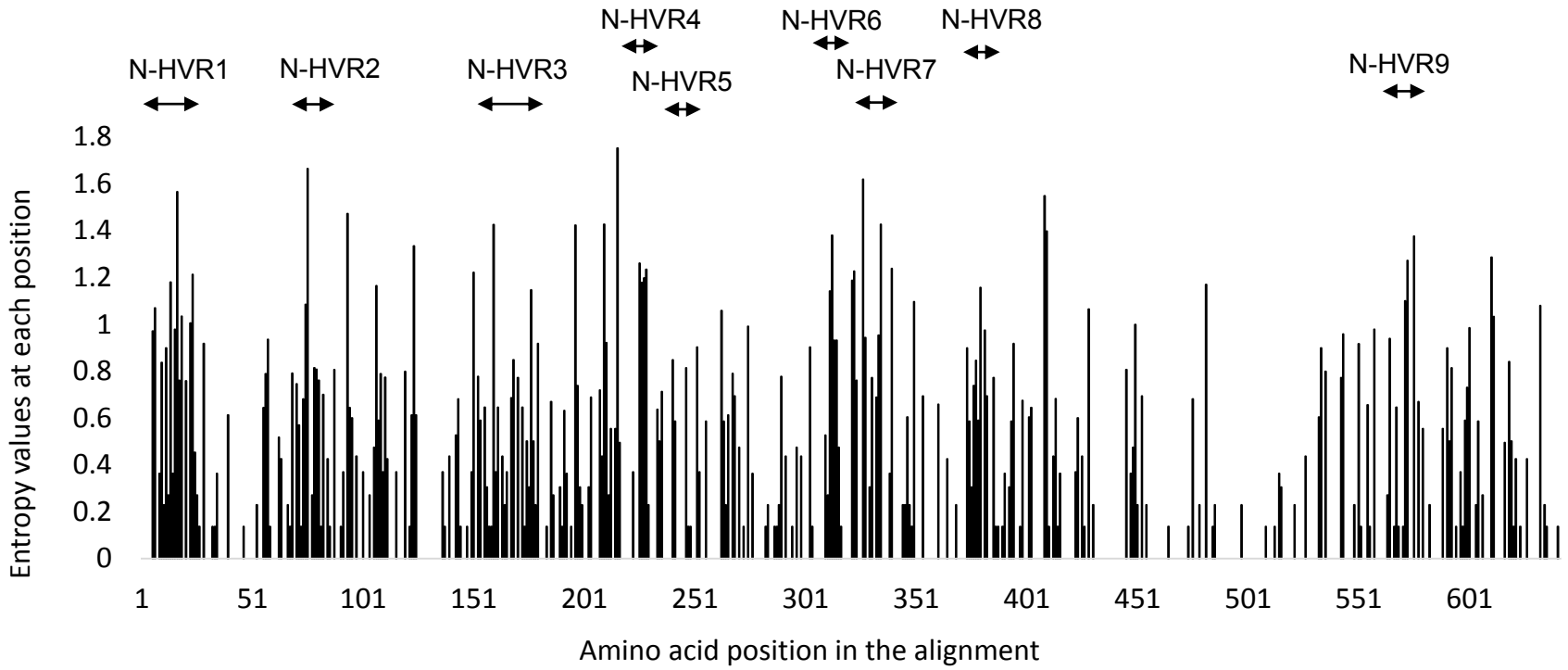


Figure 4.1. Amino acid sequence variability in the NS1 protein of the 25 local isolates and eight AMDV sequences from GenBank using the Entropy-ONE Web tool.
Amino acid positions are based on the AMDV-G sequence (GenBank accession number NC_001662). Approximate positions of hypervariable regions are identified by arrows.

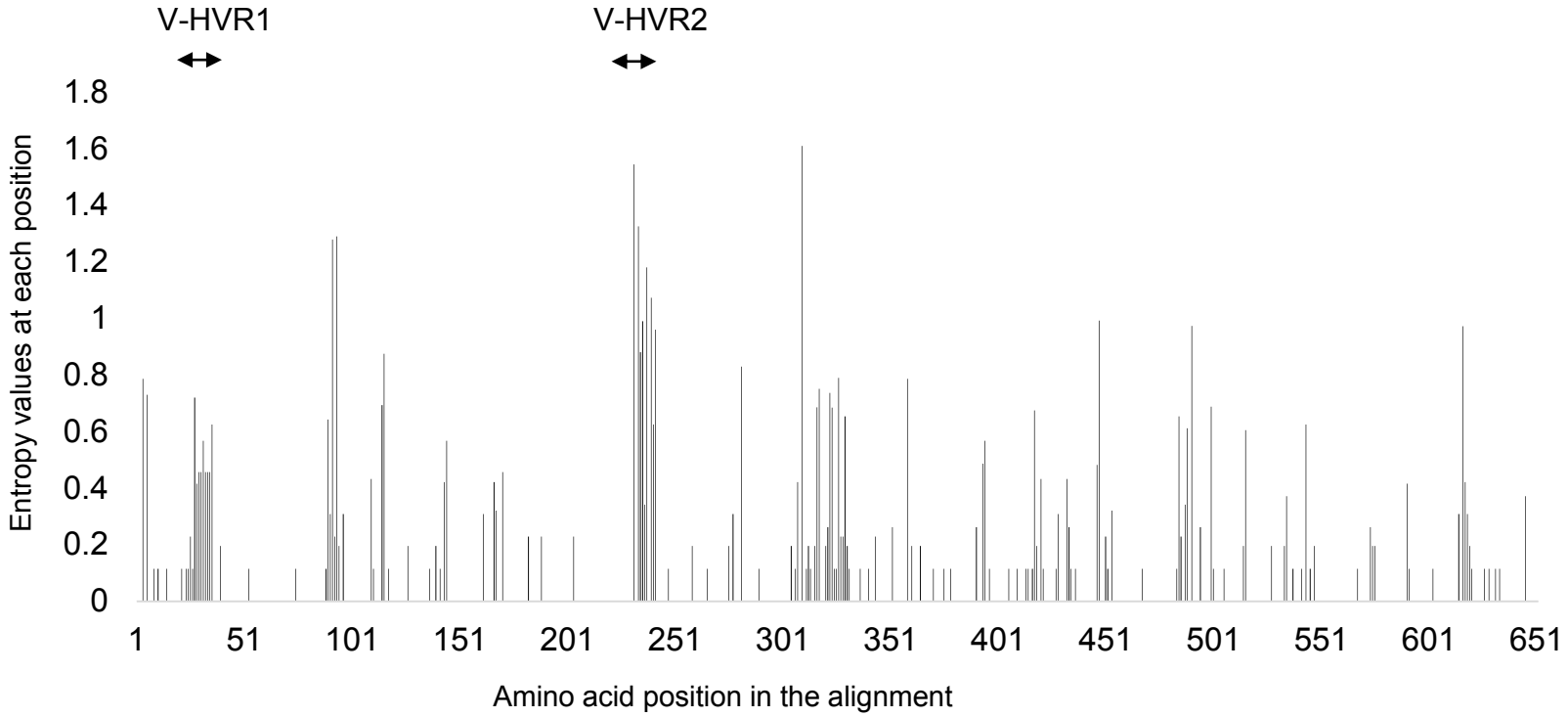


Figure 4.2. Amino acid sequence diversity in the VP2 protein of the 25 local and 16 GenBank sequences using the Entropy-ONE Web tool.
Amino acid positions are based on the AMDV-G sequence (GenBank accession number NC_001662). Approximate positions of hypervariable regions are identified by arrows.

Table 4.9. Hypervariable regions of the NS1 and VP2 proteins

Region identification	aa position[£]	Length aa	Average entropy	Number of aa positions with specified entropy measures				
				>=1.0	>0.9	>0.7	>0.5	<0.5
NS1 protein	1-641	641	0.27	-	-	-	-	-
N-HVR1	6-24	19	0.71	8	8	12	14	7
N-HVR2	69-83	15	0.61	2	2	8	10	6
N-HVR3	167-179	13	0.51	1	2	4	8	5
N-HVR4	207-216	10	0.71	2	3	4	7	3
N-HVR5	225-228	4	1.21	4	4	4	4	0
N-HVR6	309-315	7	0.80	2	4	4	5	2
N-HVR7	321-334	14	0.70	4	6	8	9	5
N-HVR8	373-385	13	0.58	1	2	6	9	4
N-HVR9	571-579	9	0.55	3	3	3	5	4
VP2 protein	1-647	647	0.09	-	-	-	-	-
V-HVR1	90-94	5	0.75	2	2	2	3	2
V-HVR2	232-242	11	0.90	5	7	8	9	2

[£] Amino acid positions are based on the AMDV-G sequence (GenBank accession number NC_001662).

recombination breakpoint hotspots at nucleotide positions around 1000 and 3100 (Figure 4.3). Recombination breakpoint pair matrix of the Group 5 sequences showed a breakpoint hotspot pair at nt position around 2300 and 3100 (Figure 4.4).

Table 4.10. Information on the 18 significant recombination events in the entire coding region of the genus *Amdoparvovirus*

Event No.	Recombination site[£] (programs)[¥]	ORFs[‡]	Recombinant isolates	Major parent × minor parent (similarity)	P-value range
1	2446-4142[†] (**R**GBMCS)	R	YA1	YA2 (98.5%) × SL3 (98.7%)	3.92E-17 to 1.05E-36
2	1[†]-1221 (RGMCS**3**)	L	CU6/CU5	CU2 (97.3%) × Unknown[¤] (KI2)	1.20E-19 to 1.41E-33
3	1871-4138[†] (RMCS**3**)	L & R	KI2	LU5 (97.6%) × YA2 (95.8%)	1.96E-02 to 6.15E-19
4	2217-2972 (**R**GMCS3)	R	HA1	Unknown (Utah1) × G (97.9%)	2.34E-05 to 5.61E-15
5	43[†]-1181 (RGMC**S**3)	L	CU2	CO3 (98.2%) × LU4 (100%)	1.67E-09 to 3.00E-22
6	135[†]-795 (RGBMC**S**)	L	YA3	YA2 (96.5%) × LU3 (96.4%)	2.57E-04 to 2.75E-16
7	920-1454 (RG**M**CS3)	L	LN3	LN1 (97%) × Utah1 (97.8%)	3.68E-06 to 1.73E-13
8	1856-3450[†] (R**M**CS)	L & R	Utah1	G (98.4%) × Unknown (CU1)	4.22E-06 to 1.73E-13
9	849-1160 (**R**GMCS3)	L	CO4/CU1/CU3/CU4/CO3	SL3 (96.9%) × CO2 (95.8%)	1.11E-04 to 6.18E-11
10	814-1750[†] (RGMCS**3**)	L	LN2	LN1 (96.3%) × Utah1 (96.8%)	3.35E-03 to 5.22E-10
11	3545[†]- 3943[†] (**R**MC3)	R	Utah1	G (99.1%) × Unknown (CO2)	4.10E-03 to 2.18E-08
12	3944-806 (RGMC**S**)	L & R	YA4/SL3[P]*	YA2 (94.4%) × LU2 (96.4%)	3.31E-03 to 4.19E-35
13	1927-2965 (MC**S**)	L & R	LN3/LN2[P]/LN1[T]*	CO3 (94.1%) × YA2 (95.9%)	2.11E-05 to 5.50E-08
14	2250-3062 (MC**S**)	R	YA4	YA2 (94.9%) × G (97.9%)	3.76E-05 to 3.51E-08
15	3127-3820[†] (**R**GMCS)	R	CO2	HA2 (93.7%) × CU3 (98.3%)	1.72E-02 to 9.08E-06
16	1730-1937 (R**G**MC)	L	CO2	HA1 (92.8%) × G (99%)	1.38E-03 to 7.65E-06
17	2973-3383[†] (G**M**C)	R	Utah1	Unknown (YA3) × G (99.8%)	7.99E-03 to 3.61E-05
18	3694-134[†] (R**M**C)	L & R	YA2	YA3 (97.7%) × × Unknown (LU3)	4.36E-03 to 9.68E-05

[£] The first and last nt of a recombination fragment, relative to recombinant sequence, detected in the break point analysis.

[¥] Recombination detection method: R=RDP, G=GENECONV, B=Bootscan, M=Maxchi, C=Chimaera, S=SiScan, 3=3Seq. Programs which detected the recombination with the smallest P-values are in **bold**.

[†] The actual breakpoint position is undetermined (it was most likely overprinted by a subsequent recombination event).

[‡] ORFs containing the detected recombination breakpoint: L: Left ORF, R: Right ORF.

[¤] Unknown refers to missing parental sequence.

* [P] and [T] are sequences with partial and trace evidence of the same recombination event, respectively.

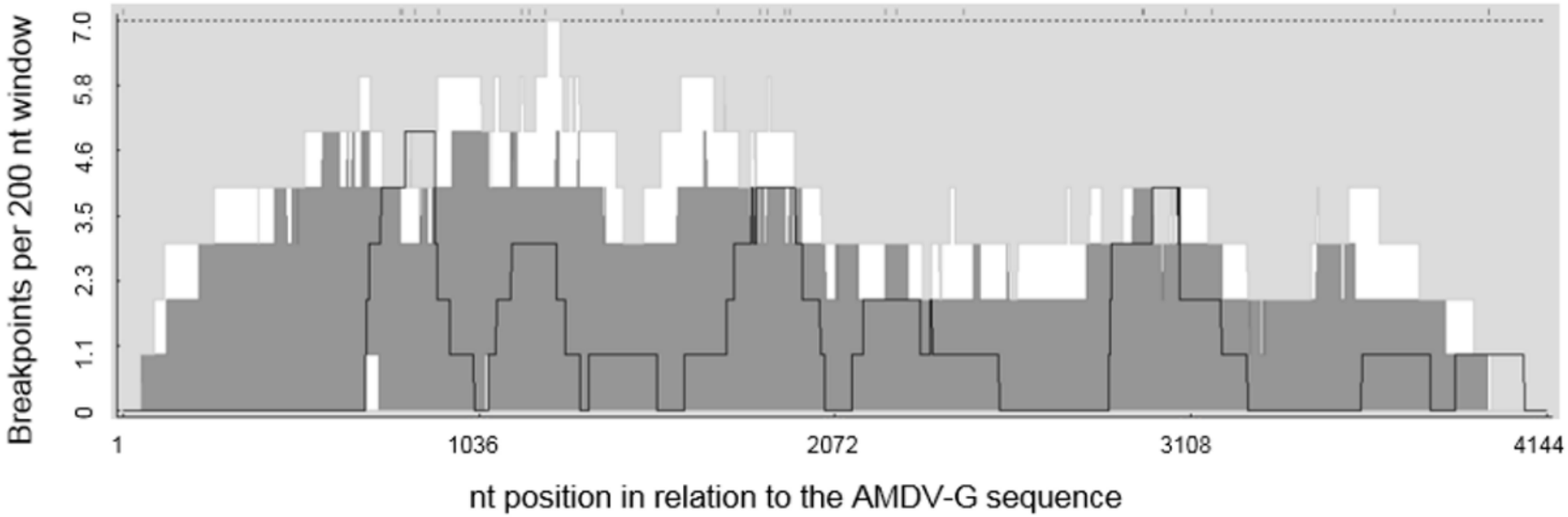


Figure 4.3. Recombination breakpoint distribution plot of the genus *Amdoparvoviruses*, showing the breakpoint hotspots. The small vertical lines at the top of the graph indicate the positively detectable breakpoint positions. Solid line indicates the number of breakpoints detected within a 200-nucleotide window moving along the alignment 1 nucleotide at a time. The breakpoint counts were plotted at the center of the window region. The dark gray and white areas indicate 95% and 99% breakpoint clustering thresholds, respectively. Statistically significant recombination hotspots are at nucleotide positions around 1000 and 3100 of the genome, where the solid line emerged from the dark gray and white shaded areas.

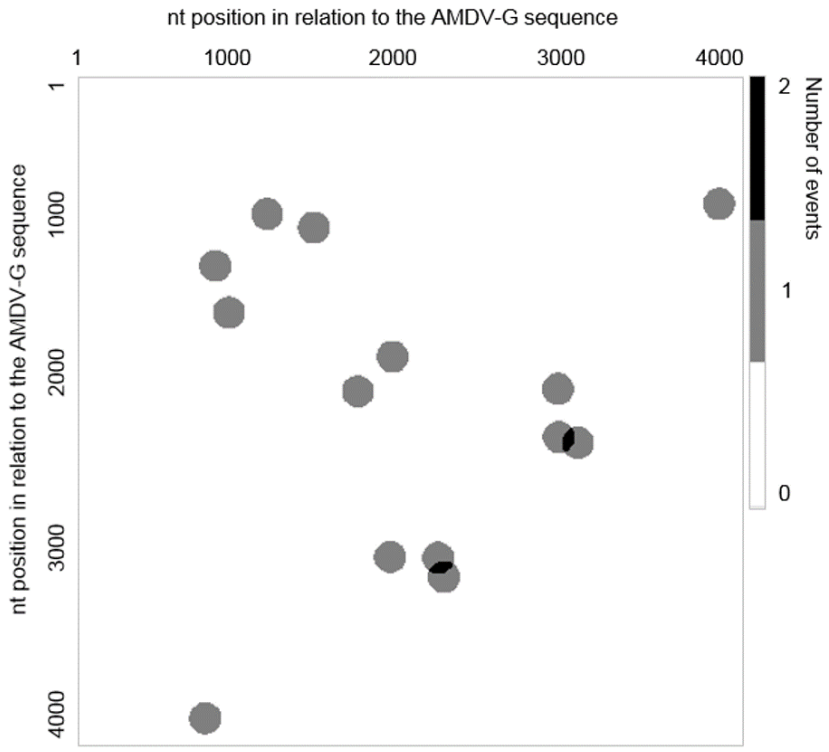


Figure 4.4. Recombination breakpoint pair matrix of the genus *Amdoparvoviruses,* showing the breakpoint hotspot pairs (Note symmetry).
The black spots show a breakpoint hotspot pair at nucleotide position around 2300 and 3100 of the genome.

4.2.2 Pairwise sequence identity and genetic distance analyses

The nt identity analyses of the VP1 and NS1 ORFs of the AMDV, RFAV and GFAV showed higher nt identities over the VP1 compared with the NS1 sequences (Table 4.11). Percentages of pairwise nt identities of the ECR of the Group 5 sequences are presented in Table A17. The AMDV sequences shared between 89.6% and 99.6% nt identities over their ECR, and showed less than 86.5% and 76.8% nt identities with RFAV and GFAV sequences, respectively (Tables 4.12 and A17).

Table 4.11. Nucleotide identity analyses of the VP1 and NS1 ORFs of the *Amdoparvoviruses*

	NS1 ORF		VP1 ORF	
	AMDV	RFAV	AMDV	RFAV
AMDV	85.1-99.6%		92.4-99.7%	
RFAV	80.7-82.7%	96.9-98.4%	88.3-90.5%	97.2-98.9%
GFAV	72.2-74.1%	72.4-73.3%	78.5-80.0%	79.7-79.8%

The pairwise identity frequency distribution plots of the near-full genome (Figure 4.5A) and ECR sequences (Figure 4.5B) of the local and corresponding six AMDV, four RFAV and one GFAV isolates, and the plot for the ECR of the non-recombinant (Section 4.2.2) AMDV sequences (Figure 4.5C) shared three pairwise nt identity troughs and/or semi-troughs (i.e. proportion of nt identities did not reach 0.00) at 77-83%, at 87-88% and at 93%. Using the first trough at 77-83% nt identity threshold, sequences were divided into two groups without any conflict, i.e. all members of each group had identities higher than the 77-83% nt identity threshold with each other and none of the members of each of the two groups had nt identities higher than the threshold (77-83%) with members of the other group. One group included GFAV and the other group included the RFAV and AMDV sequences. Using the 87-88% nt identity threshold, sequences were divided into three groups of AMDV, GFAV and RFAV, without any conflict.

Considering the 93% nt identity threshold, 33 of 36 *Amdoparvoviruses* were classified into six main Clusters (AMDV-1, AMDV-2, AMDV-3, AMDV-4, RFAV and GFAV) (Table 4.12). Members of the six main Clusters had less than 93% nt identities with sequences in other Clusters. Three sequences, CU5, CU6 and YA3, were outliers and were not classified in the aforementioned six main Clusters.

Isolates CU5 and CU6 shared 99.1% nt identities, and were assigned to the conflicting AMDV-RO1/2 Cluster, because they had high but similar nt identities (93.9-94.6%) with several sequences in the AMDV-1 and AMDV-2 Clusters (Tables A17 and 4.12). Isolate YA3 was assigned to the conflicting AMDV-RO1/3 Cluster, because it had high but similar nt identities (93.4-94.6%) with several members of the AMDV-1 and AMDV-3 Clusters. The YA3 isolate was most closely related to the YA4 isolate (94.6%) (Tables A17 and 4.12). Exclusion of the three outlier sequences from nt identity analysis, resulted in the same three thresholds shared among the three pairwise identity frequency distribution plots discussed above (Figure 4.5D).

The percentage of pairwise aa sequence identity and distance matrices of the NS1 protein of Group 5 sequences are presented in Table A18. The 85% identity and 15% distance thresholds based on the NS1 aa revealed the same clustering as the 93% threshold based on the ECR (Table 4.13). Each of the identified six main Clusters consisted of sequences sharing higher than 85% aa identity with at least one of the sequences in their cluster, did not share higher than 85% aa identity with any members of the other clusters, and showed more than 15% divergence with members of the other clusters. Based on the 85% aa identity and 15% aa distance, the same three outliers (CU5, CU6 and YA3) were identified (Tables A18 and 4.13).

Nucleotide identity analysis of a partial region of the NS1 ORF (nt 920 to 1959) of the Group 3 plus RFAV, GFAV and NOVA-OB1 (accession number KM494943) was performed (Data not shown). This analysis revealed that NOVA-

OB1, had higher than 95% nt identities with some members of the AMDV1 Cluster, including LU4, LU5, KI2, CU2, LN2, United, Utah1, SL3 and AMDV-G. The NOVA-OB1 had low nt identities (87.1%-90.4%) with the AMDV-2, AMDV-3 and AMDV-4 sequences. The aa sequence alignment of the HVR of the VP2 at position 232 to 244, showed that there was a high variability among the sequences within each of the main four AMDV Clusters and thus, AMDV Clusters could not be distinguished based on this region (Table A19).

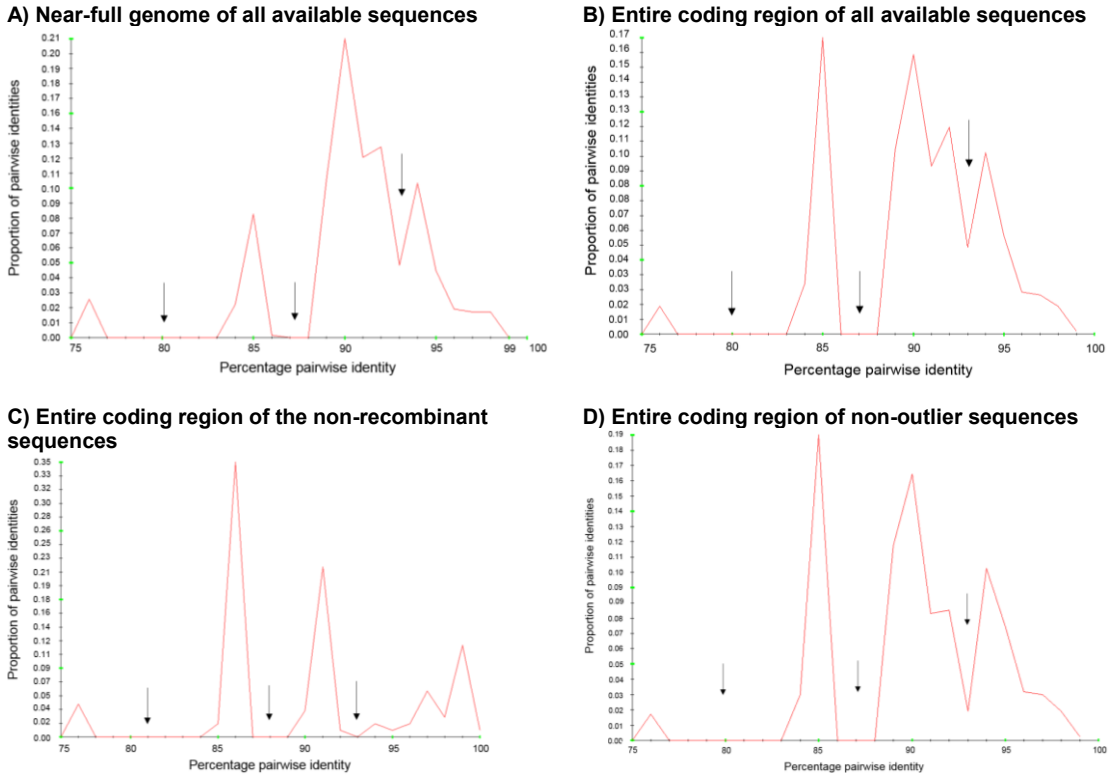


Figure 4.5. Distribution of pairwise nucleotide identities over the entire coding region and near-full genome of *Amdoparvoviruses*.
(**A**) Near-full genome of the 25 local isolates and the corresponding 8 GenBank sequences, including 5 AMDV, 2 RFAV and 1 GFAV. (**B**) Entire coding region of Group 5 sequences. (**C**) Entire coding region of the non-recombinant members of the genus *Amdoparvovirus*, including 11 AMDV (AMDV-G, CO1, CU7, HA2, KI1, LU1, LU2, LU3, LU4, LU5 and PI1), 4 RFAV and 1 GFAV sequences. **D**) Entire coding region of the 25 local isolates and 11 GenBank sequences (Group 5), excluding outlier isolates CU5, CU6 and YA3.

Table 4.12. Classification of the genus *Amdoparvovirus*, based on nucleotide identity analysis of the entire coding region

Clusters	Sub clusters	Sub-Sub-clusters	nt identity % within groups	Max nt identity % with other groups
AMDV-1			91.9-99.6	94.6: AMDV-RO1/2 & 1/3 92.5: AMDV-3
	AMDV-1a		94.9-99.4	95.1: AMDV-1a(2)
		AMDV-1a(1): AMDV-G, SL3, Utah1	97.5-99.4	96.8: AMDV-1a(2)
		AMDV-1a(2): CU2, KI1/2, LU1-5, YA4	94.9-99.4	96.8: AMDV-1a(1)
		AMDV-1a(3): CO3, CU3, CU4	97.3-98.0	96.4: AMDV-1a(1)
		AMDV-1a(4): CO4, CU1	96.8	94.6: AMDV-1a(1)
	AMDV-1b: LN1, LN2, LN3		95.4-96.0	94.8: AMDV-1a(1)
AMDV-2				93.9: AMDV-RO1/2 92.4: AMDV-1a(4)
	AMDV-2a: CO2		-	93.9: AMDV-RO1/2 93.2: AMDV-2c
	AMDV-2b: PI1		-	93.2: AMDV-2c
	AMDV-2c: CO1, HA1, HA2		94.2-95.0	93.2: AMDV-2b
AMDV-3	YA1, YA2		95.9	94.1: AMDV-RO1/3
AMDV-4	CU7		-	92.1: AMDV-1a(1)
AMDV-RO1/2[£]	CU5, CU6		99.1	94.6: AMDV-1a(4) 93.9: AMDV-2a
AMDV-RO1/3[£]	YA3		-	94.6: AMDV-1a(2) 94.1: AMDV-3
RFAV	HCR, HSR, QARF, XQJLR		97.1-98.6	86.5: AMDV-1
GFAV	GFAV		-	76.8: AMDV-1

RO: Recombinant Outliers having higher than the classification threshold with two main Clusters; AMDV-RO1/2: Recombinant Outliers of the Clusters AMDV-1 and AMDV-2; AMDV-RO1/3: Recombinant Outliers of the Clusters AMDV-1 and AMDV-3.

Table 4.13. Percentage of pairwise amino acid identity and genetic distance of the NS1 protein of the *Amdoparvoviruses*

Clusters[£]	AMDV-1	AMDV-2	AMDV-3	AMDV-4	AMDV-RO1/2	AMDV-RO1/3	RFAV
AMDV-1	82.4-99.2[¥] 0-0.17[†]						
AMDV-2	77.4-82.4 0. 11-0.22	82.8-91.6 0.08-0.17					
AMDV-3	77.4-84.9 0.03-0.22	78.8-84.4 0.15-0.20	84.6-96.6 0.03-0.15				
AMDV-4	80.2-82.5 0.17-0.20	78.9-80 0.19-0.21	79.6-80.5 0.20	- -			
AMDV-RO1/2	80.5-85.5 0.12-0.19	81.3-86.1 0.13-0.17	80.5-81.4 0.21-0.22	80.8-81 0.18	98.0 0.00		
AMDV-RO1/3	83.5-91.6 0.07-0.16	79.9-81.7 0.17-0.19	83.6-86.4 0.12-0.16	82.7 0.17	81.6-82.1 0.16-0.17	-	
RFAV	71.9-75.2 0.25-0.28	72.4-74.9 0.25-0.27	72.7-73.9 0.26-0.27	72.5-73.6 0.26-0.28	72.5-74.1 0.25-0.27	71.9-73.3 0.26-0.28	94.4-96.4 0.04-0.06
GFAV	66.3-68.5 0.3-0.34	65.8-66.8 0.17-0.34	64.4-66.3 0.32	66.3 0.34	66.1-66.3 0.33	68.7 0.31	64.4-66.1 0.34-0.35

[£] Refer to Table 4.12 for members of each Cluster. The differences are inclusion of the United strain in AMDV-1b (91.3-94.2% aa identities) and inclusion of K strain in AMDV-3.
[¥] aa identity of the NS1 protein.
[†] Genetic distance

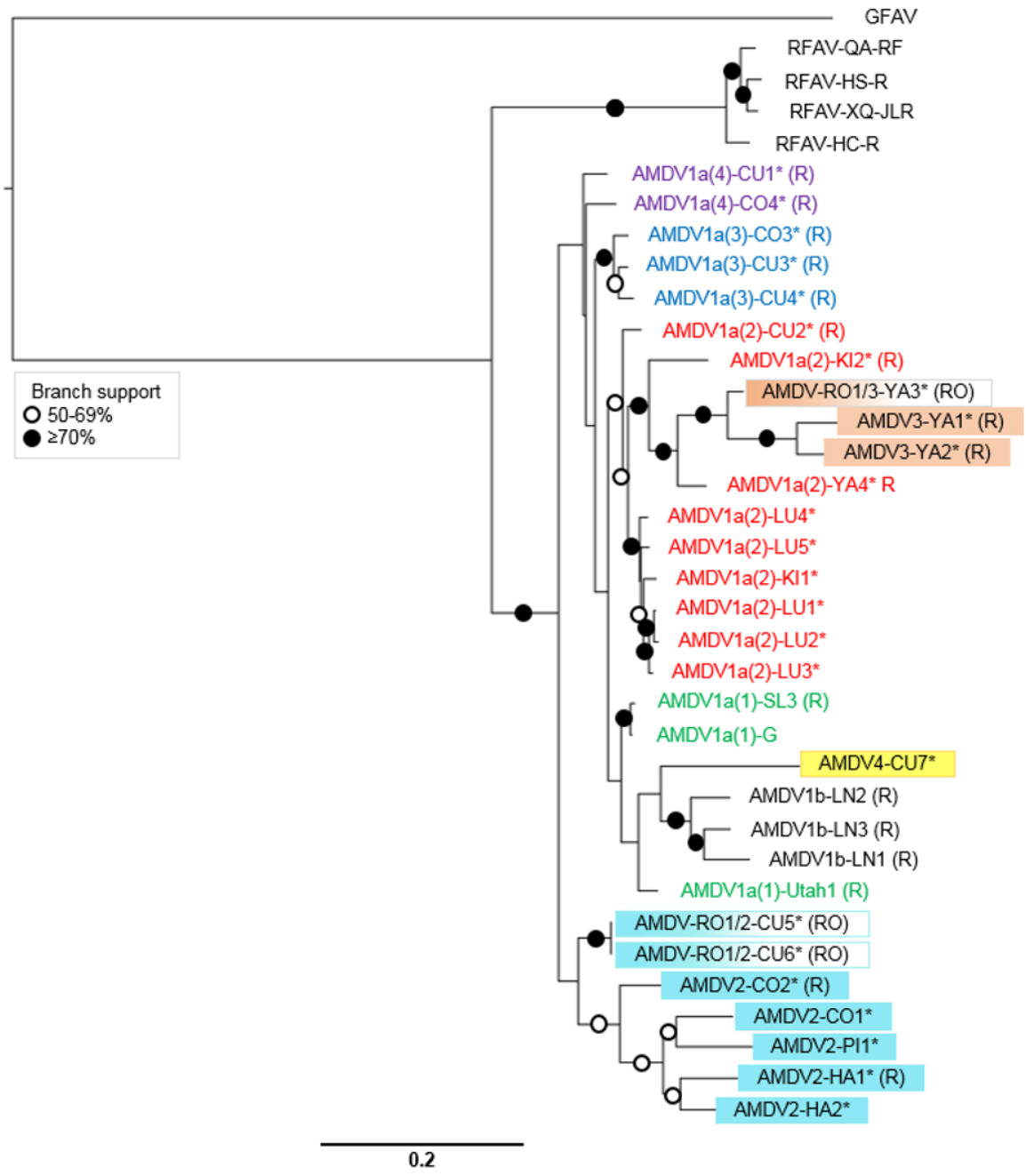
The NS1 ORF of five local isolates which contained ambiguous codes were cloned and sequenced; including CO1 (5 clones), CU1 (6 clones), CU4 (6 clones), KI1 (2 clones) and YA4 (3 clones). The nucleotide identity analysis showed that clones of each of the CO1, CU1, KI1 and YA4 had higher than 99% nucleotide identities with each other and with the original sequences. The clones of CU4 isolate showed more diversity because they had as low as 96% nucleotide identities with each other and with the original sequence.

4.2.3 Phylogenetic analyses of the genus *Amdoparvovirus*

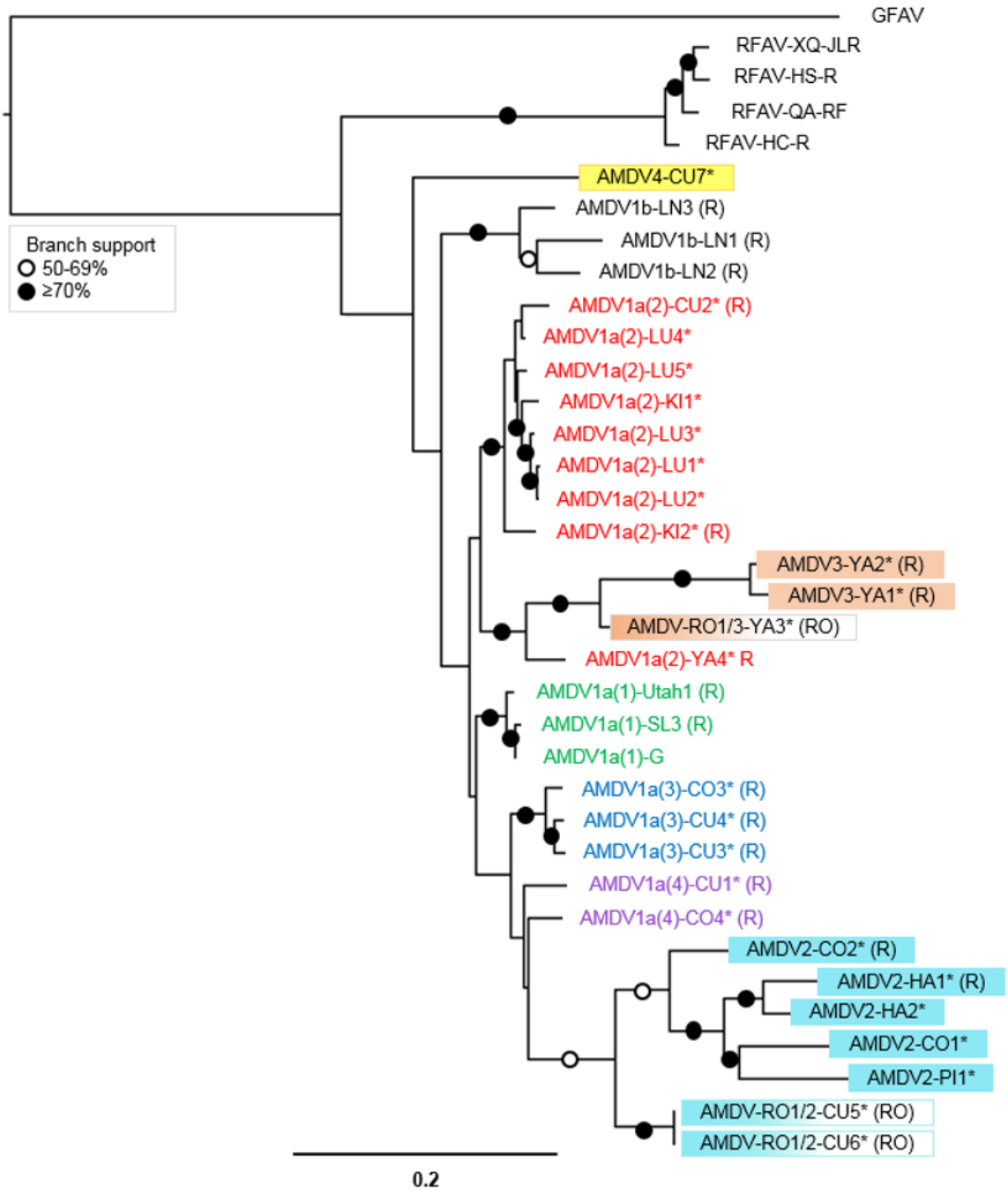
4.2.3.1 Phylogenetic analyses of the original sequences

The phylogenetic analyses of the Group 5 sequences were based on the ECR, NS1 ORF, NS1 protein, VP1 ORF and VP1 protein sequences are shown in Figures 4.6A, B, C, D, and E, respectively. In the phylogenies based on the ECR, NS1 ORF, NS1 protein, VP1 ORF and VP1 protein, only 58%, 55%, 55%, 45% and 18% of the branches had bootstrap supports higher than 70%, respectively. In addition to low bootstrap values throughout the five phylogenies, none revealed a clear clade division into the four main AMDV Clusters (AMDV-1 to AMDV-4) (Table 4.12). The CU5 and CU6 isolates (AMDV-RO1/2) clustered with the AMDV-2 sequences in the ECR and NS1 ORF and protein phylogenies (Figures 4.6A-C), but with AMDV-1 sequences in the VP1 phylogenies (Figures 4.6D-E).

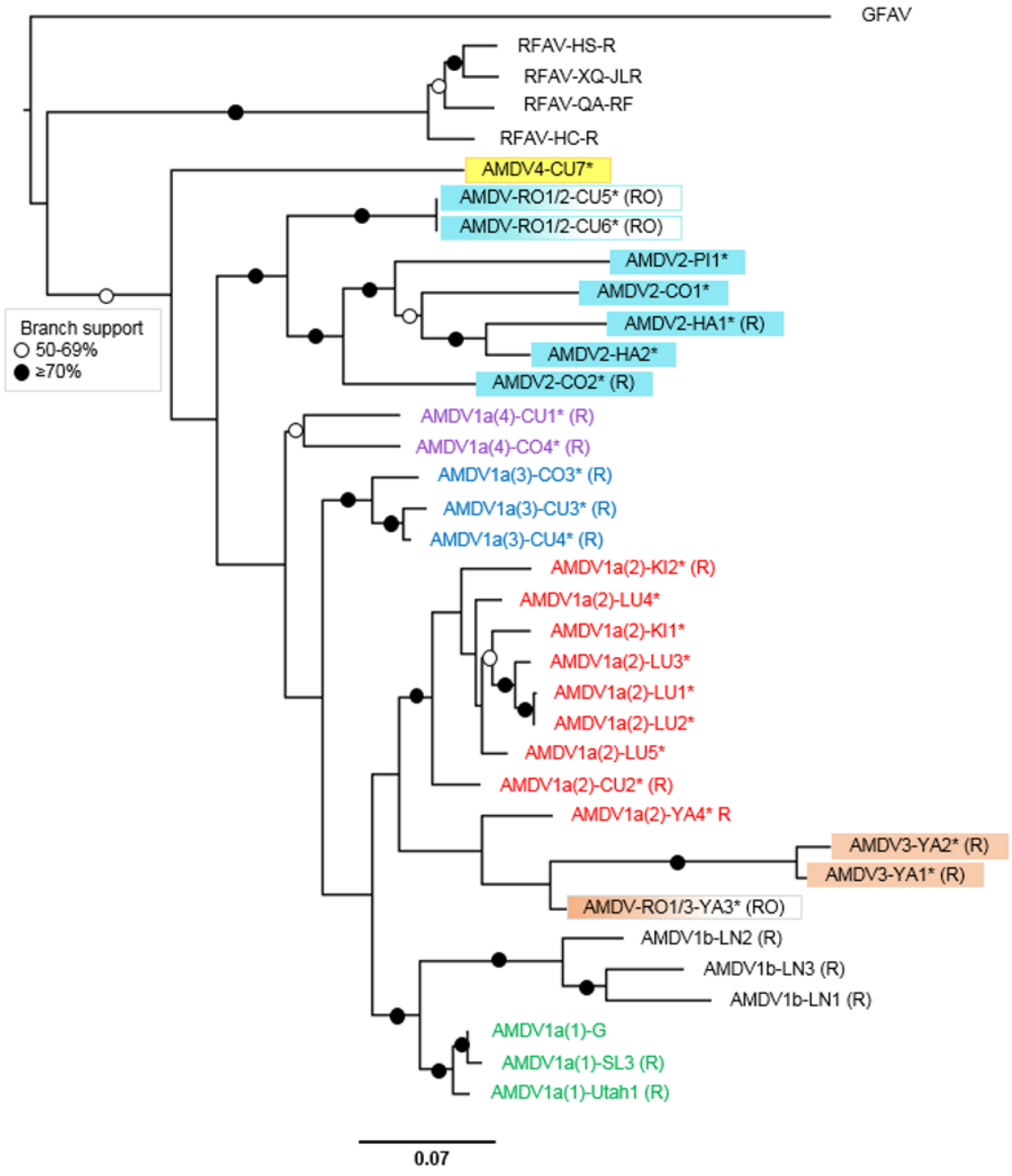
A) Entire coding region



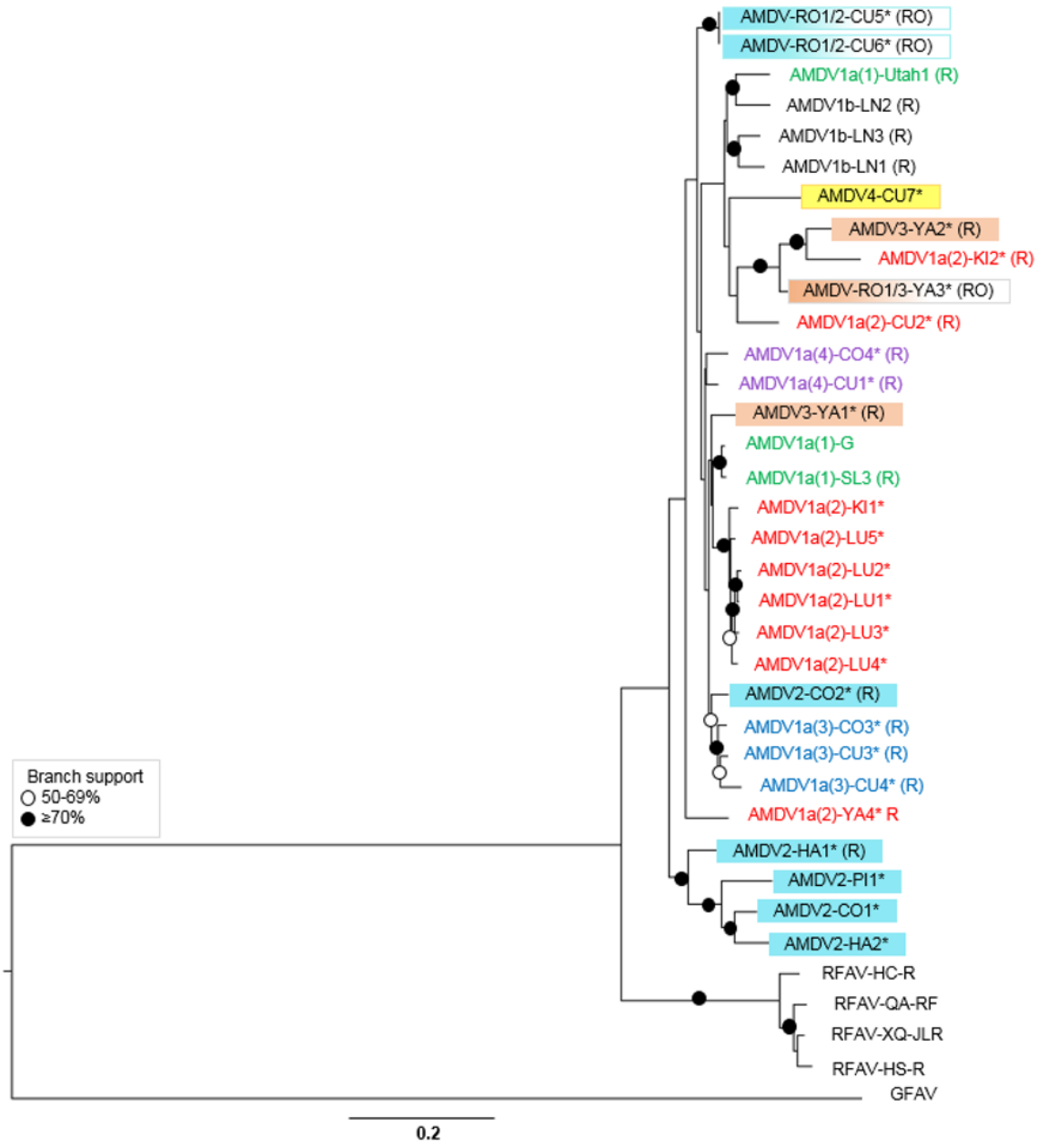
B) NS1 ORF


GFAV
RFAV-XQ-JLR
RFAV-HS-R
RFAV-QA-RF
RFAV-HC-R
AMDV4-CU7*
AMDV1b-LN3 (R)
AMDV1b-LN1 (R)
AMDV1b-LN2 (R)
AMDV1a(2)-CU2* (R)
AMDV1a(2)-LU4*
AMDV1a(2)-LU5*
AMDV1a(2)-KI1*
AMDV1a(2)-LU3*
AMDV1a(2)-LU1*
AMDV1a(2)-LU2*
AMDV1a(2)-KI2* (R)
AMDV3-YA2* (R)
AMDV3-YA1* (R)
AMDV-RO1/3-YA3* (RO)
AMDV1a(2)-YA4* R
AMDV1a(1)-Utah1 (R)
AMDV1a(1)-SL3 (R)
AMDV1a(1)-G
AMDV1a(3)-CO3* (R)
AMDV1a(3)-CU4* (R)
AMDV1a(3)-CU3* (R)
AMDV1a(4)-CU1* (R)
AMDV1a(4)-CO4* (R)
AMDV2-CO2* (R)
AMDV2-HA1* (R)
AMDV2-HA2*
AMDV2-CO1*
AMDV2-PI1*
AMDV-RO1/2-CU5* (RO)
AMDV-RO1/2-CU6* (RO)
Branch support
50-69%
≥70%
0.2

C) NS1 amino acid



D) VP1 ORF


AMDV-RO1/2-CU5* (RO)
AMDV-RO1/2-CU6* (RO)
AMDV1a(1)-Utah1 (R)
AMDV1b-LN2 (R)
AMDV1b-LN3 (R)
AMDV1b-LN1 (R)
AMDV4-CU7*
AMDV3-YA2* (R)
AMDV1a(2)-KI2* (R)
AMDV-RO1/3-YA3* (RO)
AMDV1a(2)-CU2* (R)
AMDV1a(4)-CO4* (R)
AMDV1a(4)-CU1* (R)
AMDV3-YA1* (R)
AMDV1a(1)-G
AMDV1a(1)-SL3 (R)
AMDV1a(2)-KI1*
AMDV1a(2)-LU5*
AMDV1a(2)-LU2*
AMDV1a(2)-LU1*
AMDV1a(2)-LU3*
AMDV1a(2)-LU4*
AMDV2-CO2* (R)
AMDV1a(3)-CO3* (R)
AMDV1a(3)-CU3* (R)
AMDV1a(3)-CU4* (R)
AMDV1a(2)-YA4* R
AMDV2-HA1* (R)
AMDV2-PI1*
AMDV2-CO1*
AMDV2-HA2*
RFAV-HC-R
RFAV-QA-RF
RFAV-XQ-JLR
RFAV-HS-R
GFAV
Branch support
50-69%
≥70%
0.2

E) VP1 protein

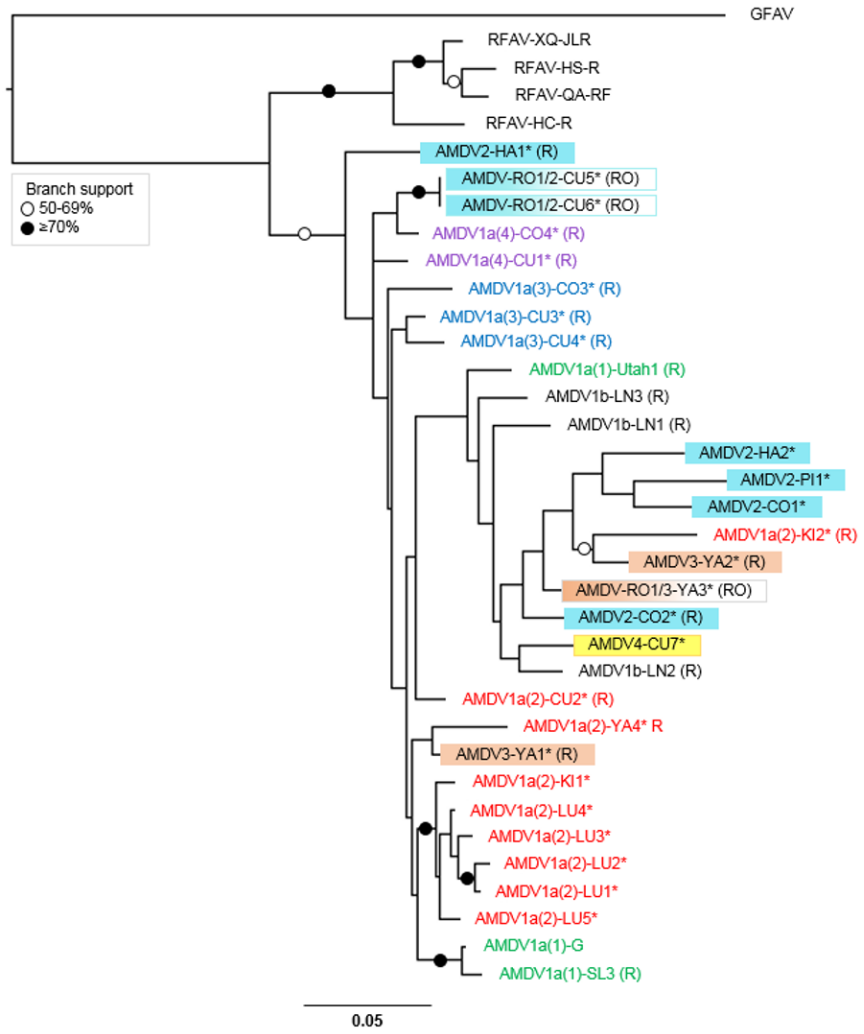


Figure 4.6. ML mid-point rooted phylogenetic analyses of the entire coding region, NS1 and VP1 ORFs and protein sequences of the genus *Amdvoparvovirus*.
A bootstrap analysis of 1000 replicates was used and nodes with ≥50% bootstrap support are shown. The nt substitution model GTR + G + I was used for analyzing the entire coding region and VP1 ORF. The model GTR + G was used for analyzing the NS1 ORF. The models JTT + G and rtREV + G + I were used for analyzing the NS1 and VP1 aa sequences, respectively. The scale bar indicates the number of nt substitutions per site. Acronyms are described in Table 3.1 and sequences of the Groups 1 to 5 are explained in Table 3.6. The local isolates are marked with a star. Recombinant and recombinant outlier sequences are characterized with (R) and (RO), respectively. Sequences belonging to the novel AMDV-2, AMDV-3 and AMDV-4 Clusters are highlighted with blue, orange and yellow, respectively. The outlier isolates CU5 and CU6 (AMDV-RO1/2), are highlighted half in blue. The outlier isolate YA3 (AMDV-RO1/3), is highlighted half in orange. Sequences of the AMDV-1b Sub-cluster are shown in black and sequences of the AMDV-1a Sub-cluster are shown in colors: AMDV-1a(1): green; AMDV-1a(2): red; AMDV-1a(3): blue; AMDV-1a(4): Purple.

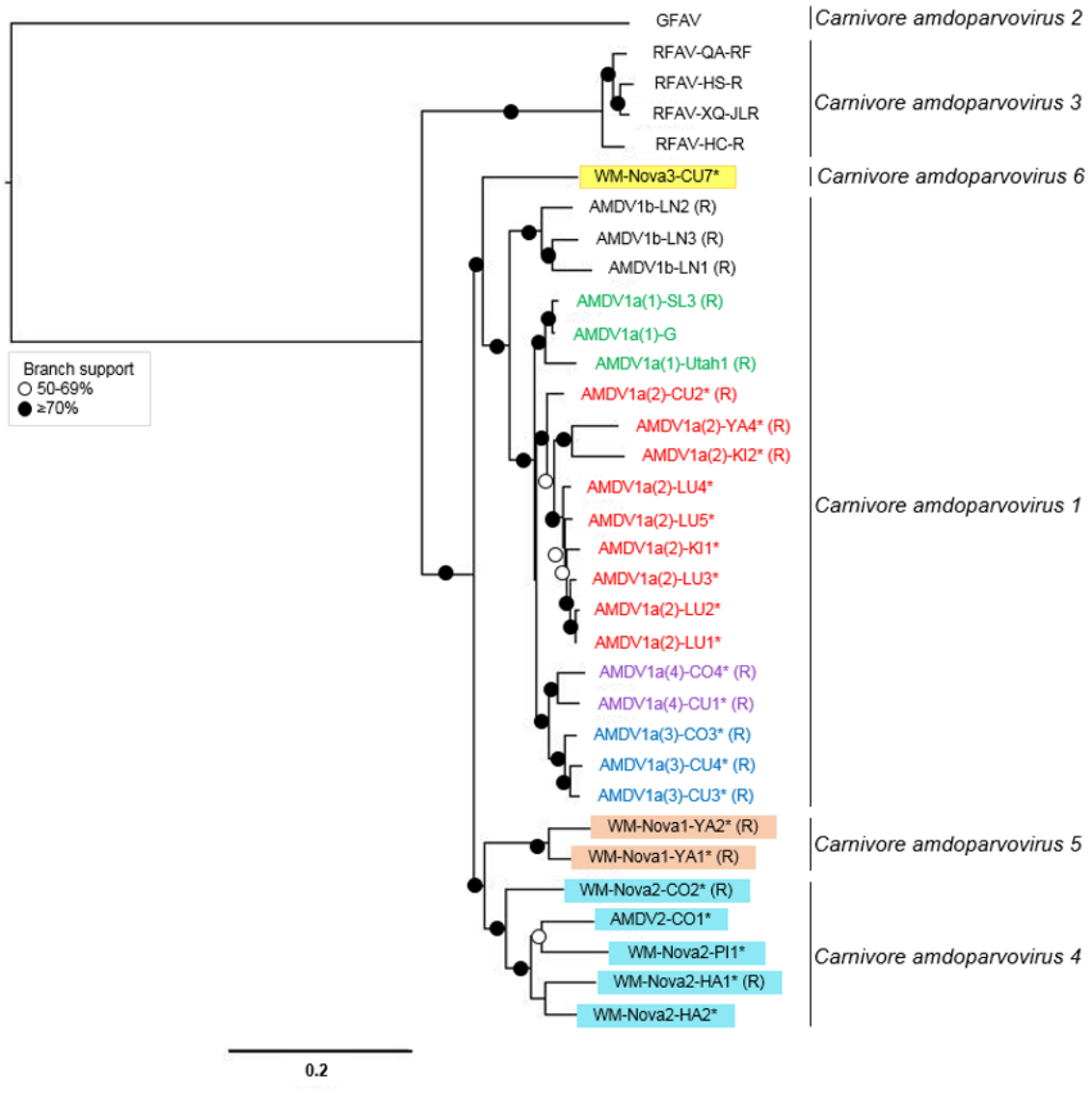
4.2.3.2 Phylogenetic analyses of the modified sequences

To obtain phylogenies with high bootstrap support values which were compatible with clustering based on sequence identity and distance analyses (Tables 4.12 and 4.13), three approaches were taken (1) phylogenetic analysis free of ambiguous codes, (2) phylogenetic analysis free of outlier sequences, and (3) phylogenetic analysis free of the identified recombination regions of the genome. Phylogenetic analysis of the ECR of Group 5 sequences, free of ambiguous codes (Figure A19), revealed 61% of the branches had bootstrap supports higher than 70% and the branching patterns were the same as that of the original phylogeny (Figure 4.6A).

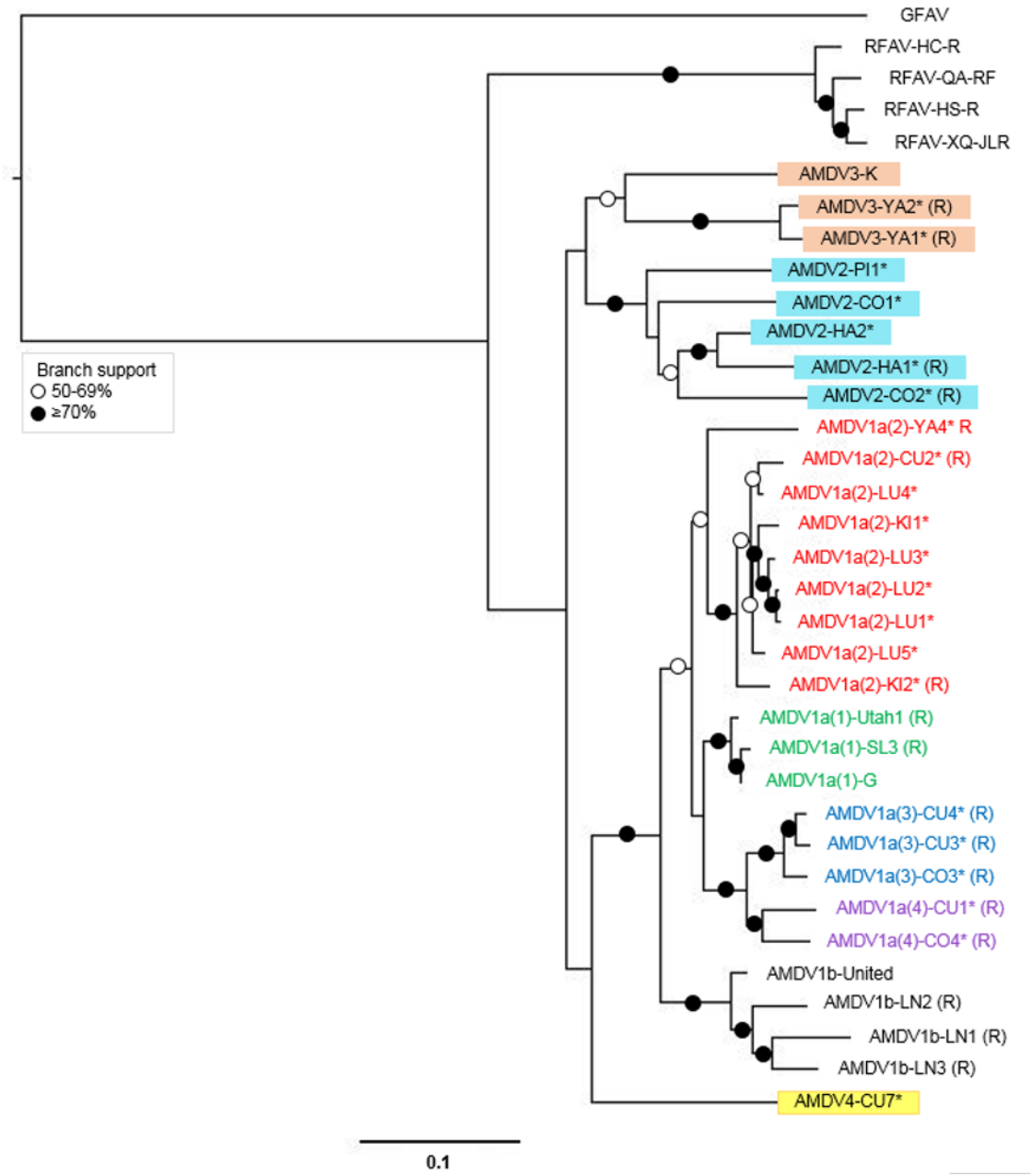
Phylogenetic analyses of the ECR and NS1 ORF and protein sequences, excluding the outliers (CU5, CU6, YA3) are shown in Figures 4.7A-C. In these phylogenies, 77%, 59%, 56% of the branches had bootstrap supports higher than 70%, respectively, and thus, had higher bootstrap supports compared with the corresponding original phylogenies (Figures 4.6A-C). Clustering the AMDV Clusters and Sub-clusters in the ECR, NS1 ORF and protein phylogenies, free of outlier isolates (Figures 4.7A-C), were compatible with clustering based on sequence identity and distance analyses (Tables 4.12 and 4.13). In the VP1 ORF and protein outlier-free phylogenies (Figures 4.7D and 4.7E), only 50% and 23% of the branches had bootstrap supports higher than 70%, respectively. In both phylogenies, members of the four main AMDV Clusters were scattered throughout the clades. The VP1 ORF phylogeny was not fully resolved, i.e., more than two

sequences immediately descended from an internal node and made a polytomy, which could represent uncertainties about branching order (Hall & Barlow, 2006).

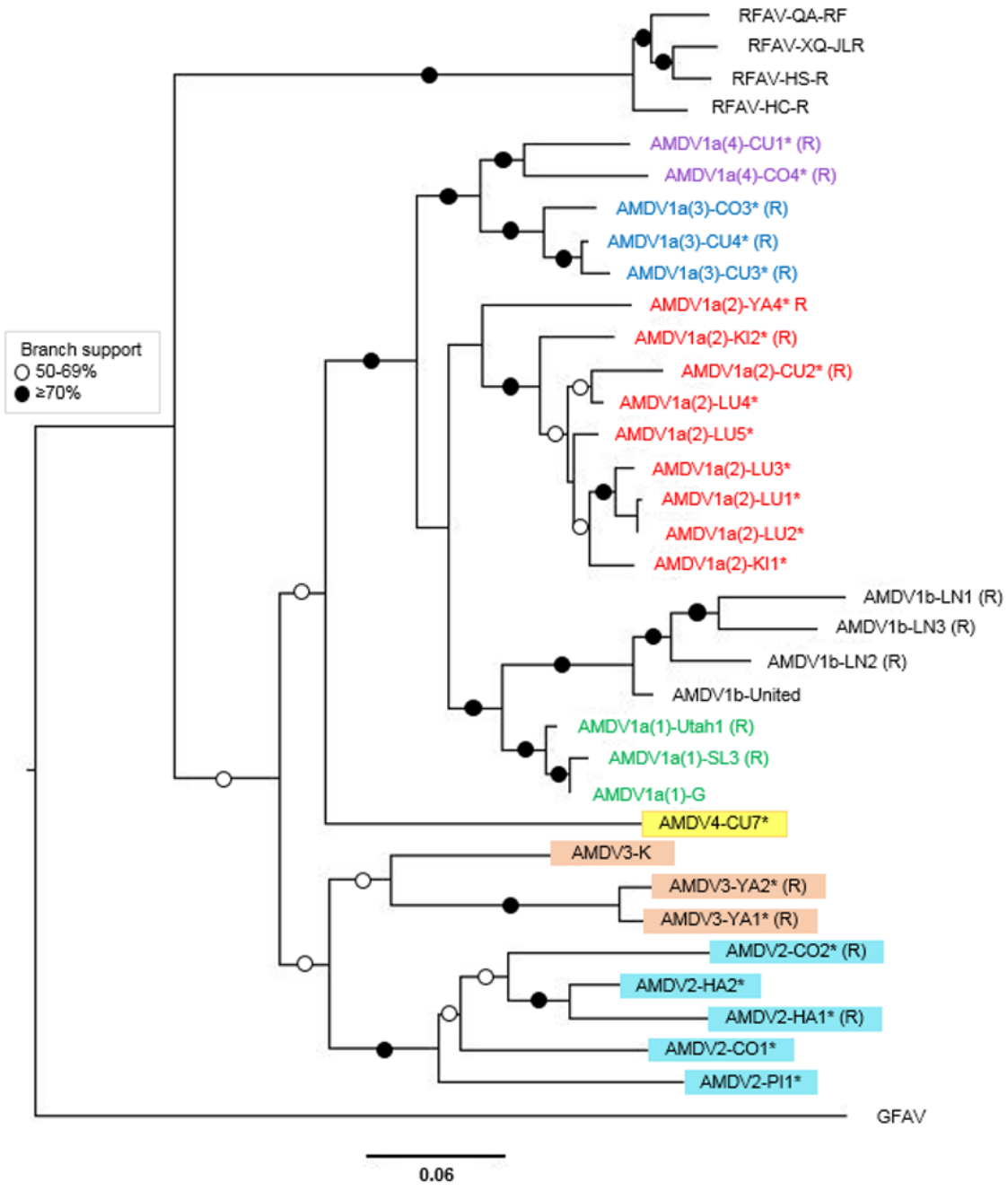
A) Entire coding region



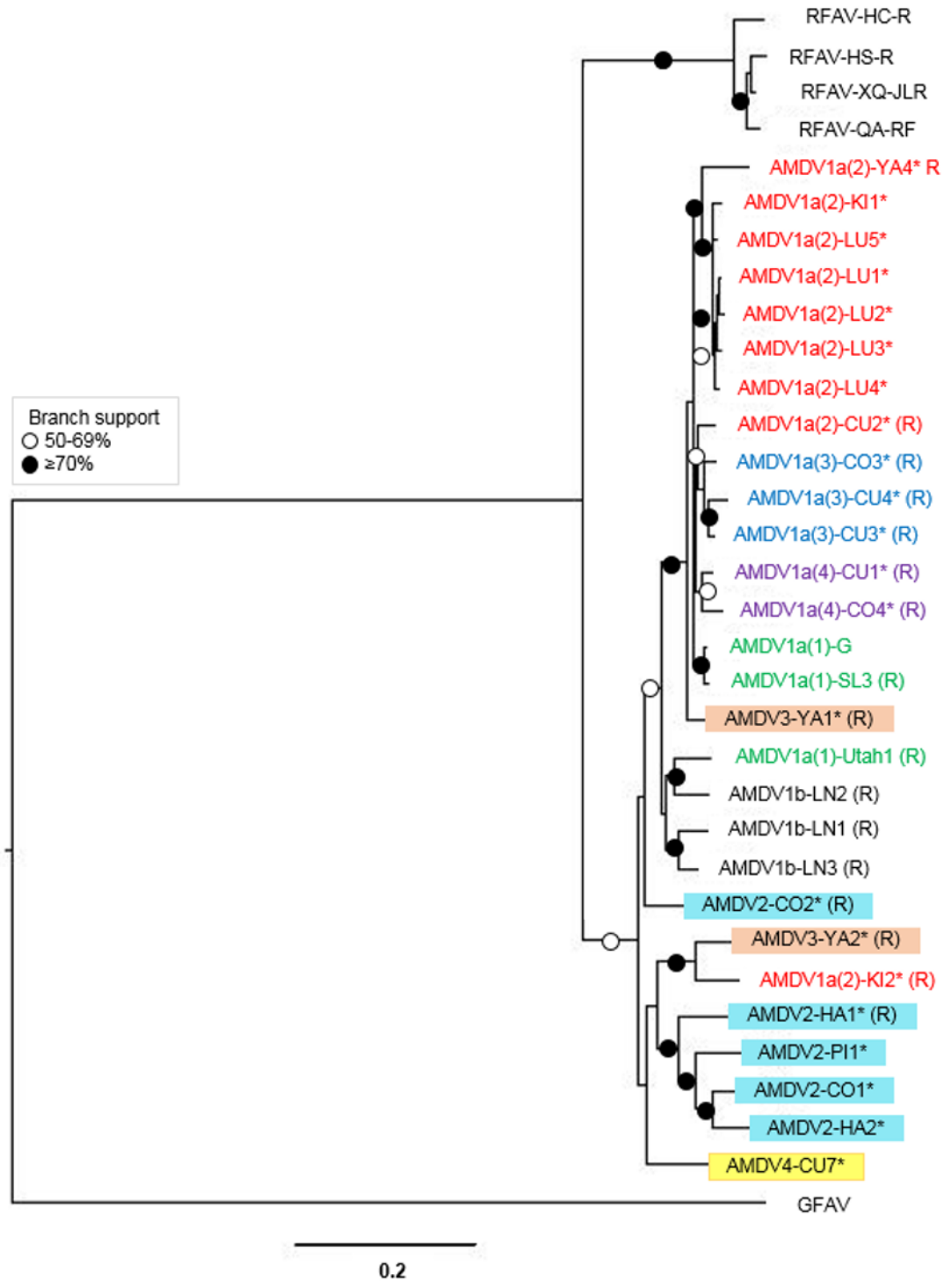
B) NS1 ORF


GFAV
RFAV-HC-R
RFAV-QA-RF
RFAV-HS-R
RFAV-XQ-JLR
AMDV3-K
AMDV3-YA2* (R)
AMDV3-YA1* (R)
AMDV2-PI1*
AMDV2-CO1*
AMDV2-HA2*
AMDV2-HA1* (R)
AMDV2-CO2* (R)
AMDV1a(2)-YA4* R
AMDV1a(2)-CU2* (R)
AMDV1a(2)-LU4*
AMDV1a(2)-KI1*
AMDV1a(2)-LU3*
AMDV1a(2)-LU2*
AMDV1a(2)-LU1*
AMDV1a(2)-LU5*
AMDV1a(2)-KI2* (R)
AMDV1a(1)-Utah1 (R)
AMDV1a(1)-SL3 (R)
AMDV1a(1)-G
AMDV1a(3)-CU4* (R)
AMDV1a(3)-CU3* (R)
AMDV1a(3)-CO3* (R)
AMDV1a(4)-CU1* (R)
AMDV1a(4)-CO4* (R)
AMDV1b-United
AMDV1b-LN2 (R)
AMDV1b-LN1 (R)
AMDV1b-LN3 (R)
AMDV4-CU7*
Branch support
50-69%
≥70%
0.1

C) NS1 protein



D) VP1 ORF


RFAV-HC-R
RFAV-HS-R
RFAV-XQ-JLR
RFAV-QA-RF
AMDV1a(2)-YA4* R
AMDV1a(2)-KI1*
AMDV1a(2)-LU5*
AMDV1a(2)-LU1*
AMDV1a(2)-LU2*
AMDV1a(2)-LU3*
AMDV1a(2)-LU4*
AMDV1a(2)-CU2* (R)
AMDV1a(3)-CO3* (R)
AMDV1a(3)-CU4* (R)
AMDV1a(3)-CU3* (R)
AMDV1a(4)-CU1* (R)
AMDV1a(4)-CO4* (R)
AMDV1a(1)-G
AMDV1a(1)-SL3 (R)
AMDV3-YA1* (R)
AMDV1a(1)-Utah1 (R)
AMDV1b-LN2 (R)
AMDV1b-LN1 (R)
AMDV1b-LN3 (R)
AMDV2-CO2* (R)
AMDV3-YA2* (R)
AMDV1a(2)-KI2* (R)
AMDV2-HA1* (R)
AMDV2-PI1*
AMDV2-CO1*
AMDV2-HA2*
AMDV4-CU7*
GFAV
Branch support
50-69%
≥70%
0.2

E) VP1 protein

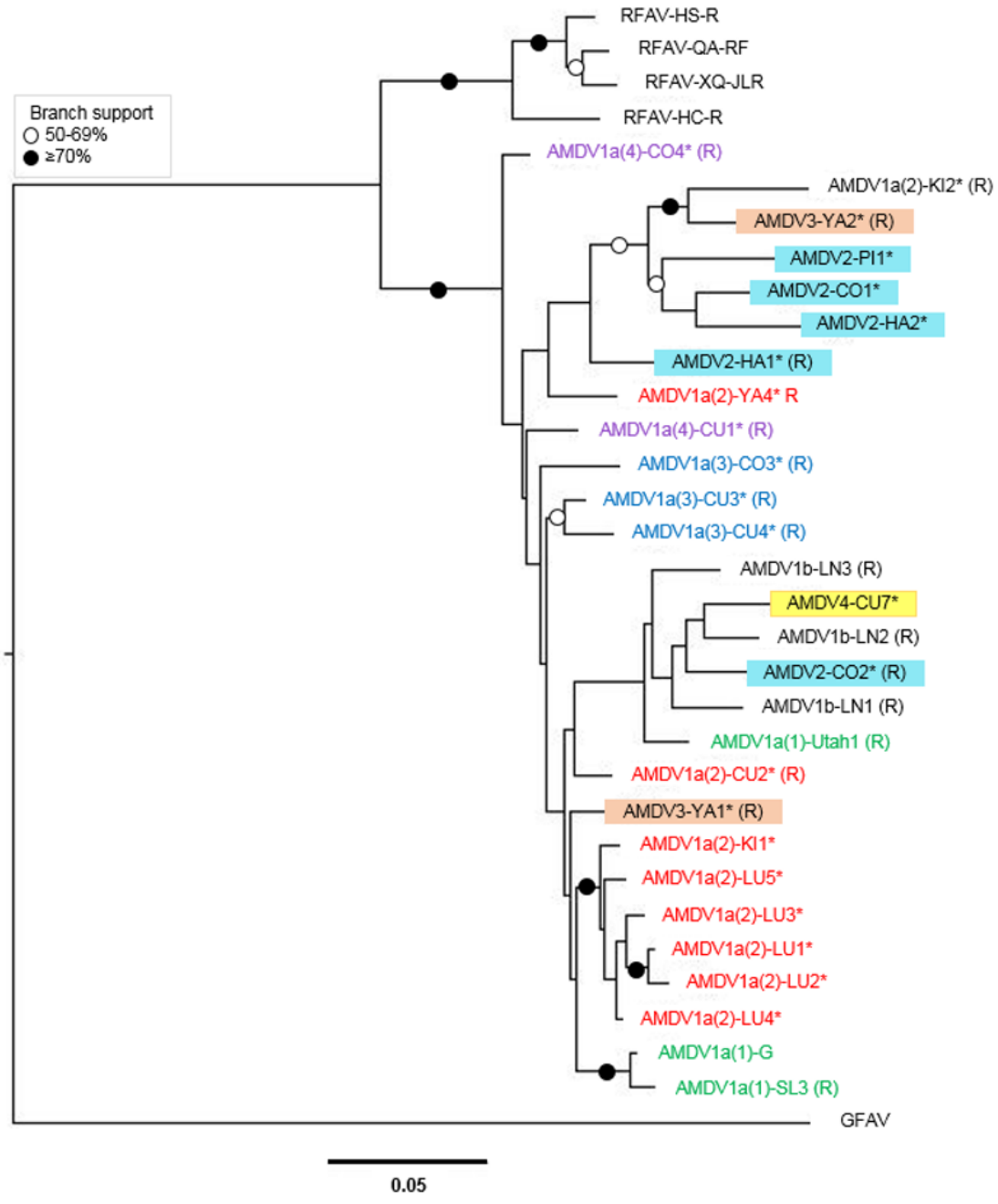
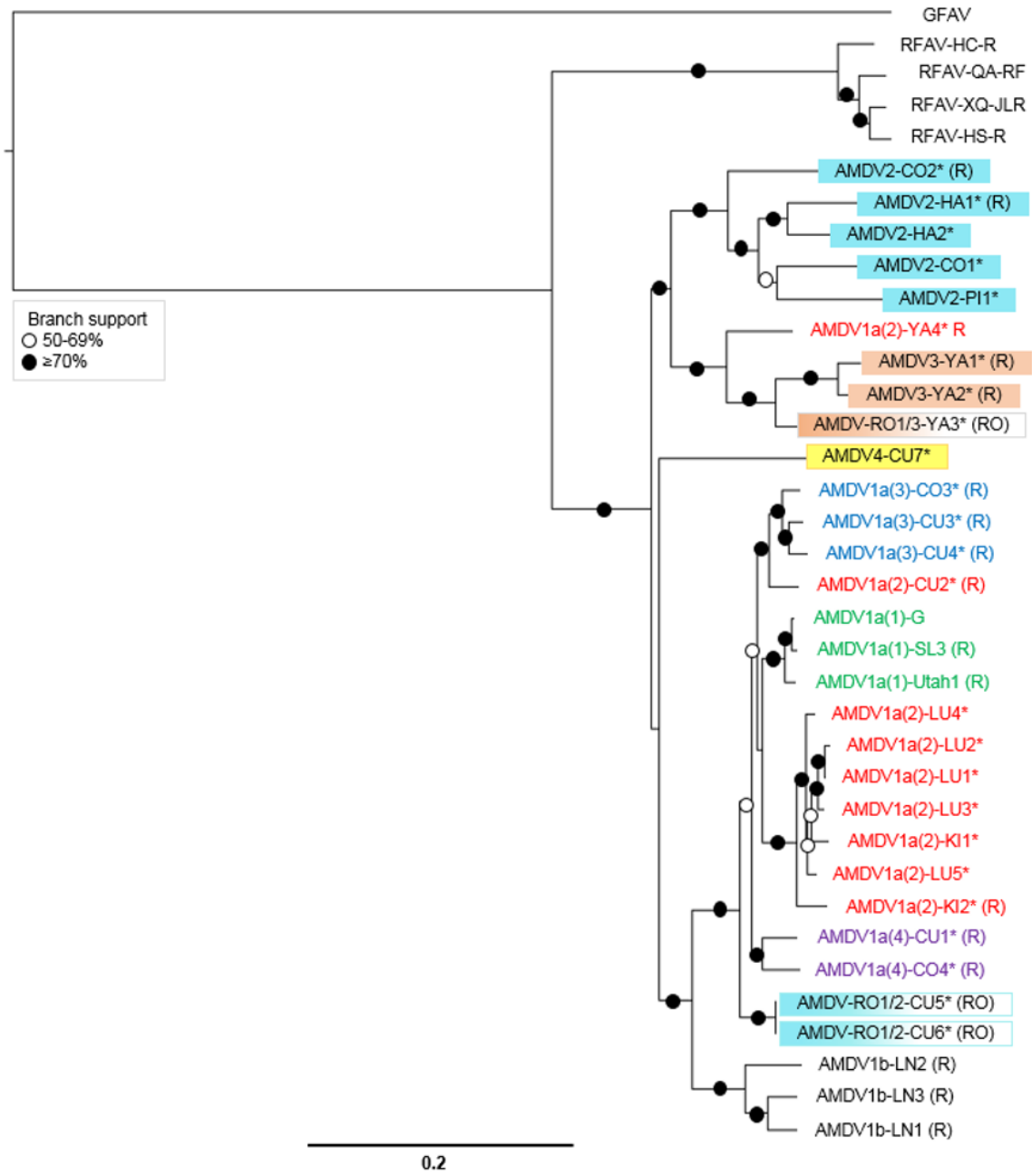


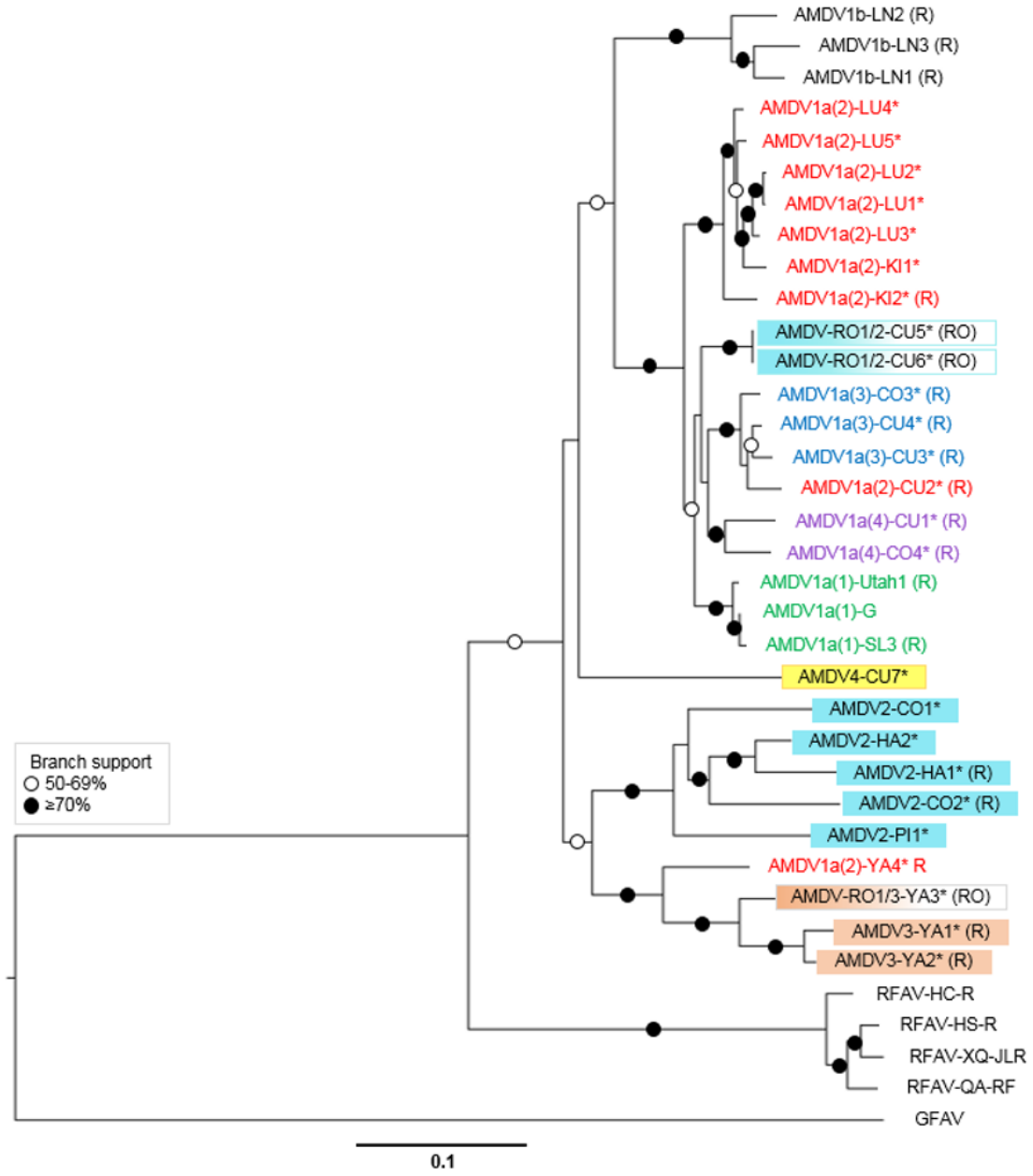
Figure 4.7. ML mid-point rooted phylogenetic analyses of the entire coding region, NS1 ORF and protein, VP1 ORF and protein of the genus *Amdvoparvovirus*, excluding the outlier isolates CU5, CU6 and YA3.
Refer to Figure 4.6 for description of sequences, symbols and colors.

Three ML phylogenetic analyses based on the ECR, NS1 and VP1 ORFs of the non-recombinant genome segments of the Group 5 sequences, along with their bootstrap values, are shown in Figures 4.8A-C. In the phylogenies based on the ECR, NS1 and VP1 ORF (Figures 4.8A-C), 79%, 67% and 40% of the branches had bootstrap supports higher than 70%, respectively, and thus, had similar bootstrap supports to the corresponding outlier-free phylogenies (Figures 4.7A, B and D). In these three phylogenies, similar to the original and outlier-free phylogenies, AMDV, RFAV and GFAV sequences were clearly divided into three main clades. In these three phylogenies, the AMDV-2 and AMDV-4 Clusters each formed clades separate from other Clusters, compatible with clustering based on sequence identity and distance analyses (Tables 4.12 and 4.13). The AMDV-3 sequences formed a separate clade, but together with the AMDV-1-YA4 and the outlier AMDV-RO1/3-YA3 sequences (Figures 4.8A-C). In all three phylogenies, sub-sub-clustering of the AMDV-1 sequences, was not compatible with the sequence identity and distance clustering, because the AMDV-1a(2)-CU2 was clustered with the AMDV-1a(3) sequences (Figures 4.8A-C). In the VP1 ORF phylogeny (Figure 4.8C), most members of the AMDV-1 Cluster were not clustered similar to the clustering based on the sequence identity and distance analysis (Tables 4.13 and 4.13).

A) Entire coding region



B) NS1 ORF


AMDV1b-LN2 (R)
AMDV1b-LN3 (R)
AMDV1b-LN1 (R)
AMDV1a(2)-LU4*
AMDV1a(2)-LU5*
AMDV1a(2)-LU2*
AMDV1a(2)-LU1*
AMDV1a(2)-LU3*
AMDV1a(2)-KI1*
AMDV1a(2)-KI2* (R)
AMDV-RO1/2-CU5* (RO)
AMDV-RO1/2-CU6* (RO)
AMDV1a(3)-CO3* (R)
AMDV1a(3)-CU4* (R)
AMDV1a(3)-CU3* (R)
AMDV1a(2)-CU2* (R)
AMDV1a(4)-CU1* (R)
AMDV1a(4)-CO4* (R)
AMDV1a(1)-Utah1 (R)
AMDV1a(1)-G
AMDV1a(1)-SL3 (R)
AMDV4-CU7*
AMDV2-CO1*
AMDV2-HA2*
AMDV2-HA1* (R)
AMDV2-CO2* (R)
AMDV2-PI1*
AMDV1a(2)-YA4* R
AMDV-RO1/3-YA3* (RO)
AMDV3-YA1* (R)
AMDV3-YA2* (R)
RFAV-HC-R
RFAV-HS-R
RFAV-XQ-JLR
RFAV-QA-RF
GFAV
Branch support
50-69%
≥70%
0.1

C) VP1 ORF

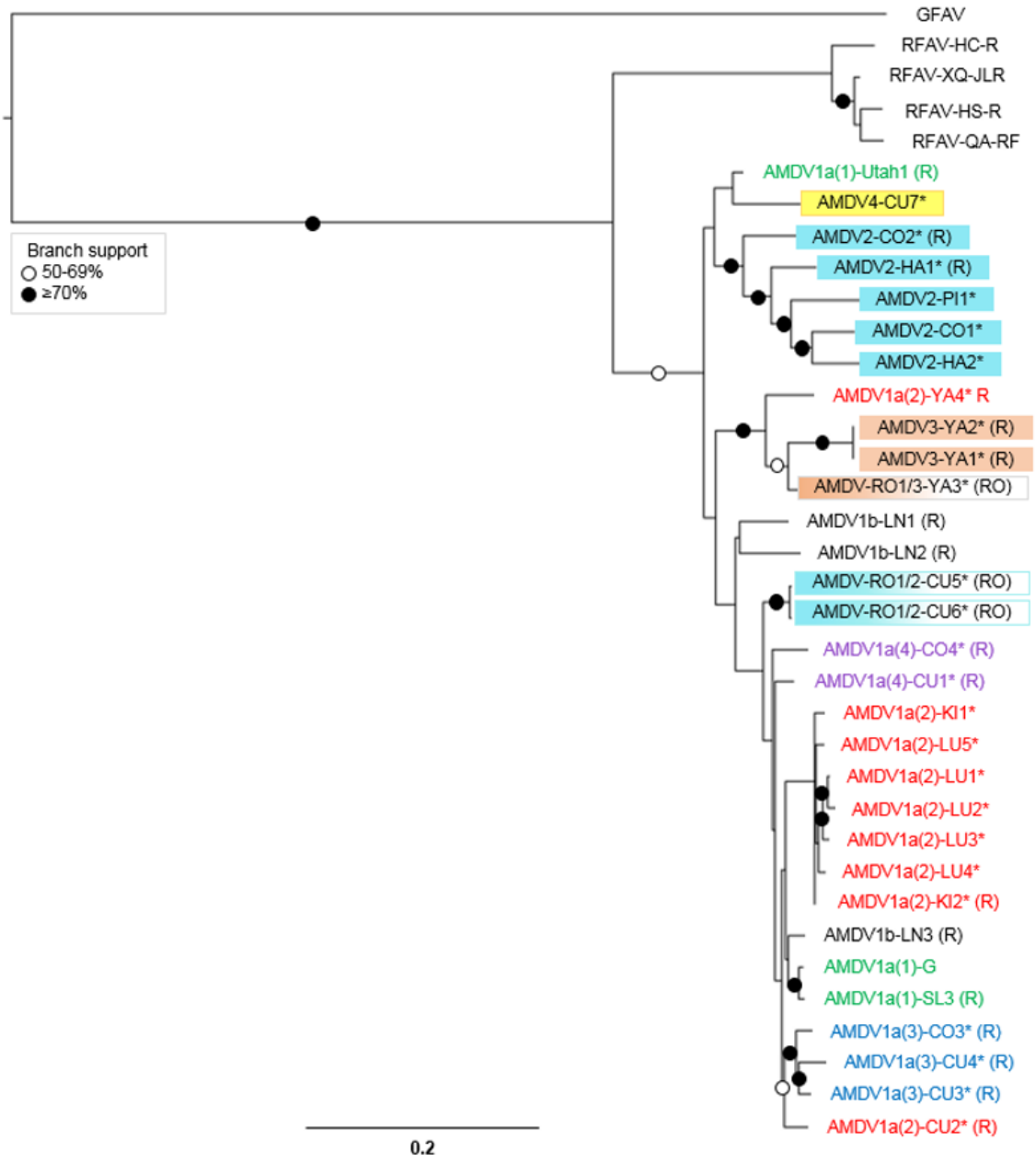


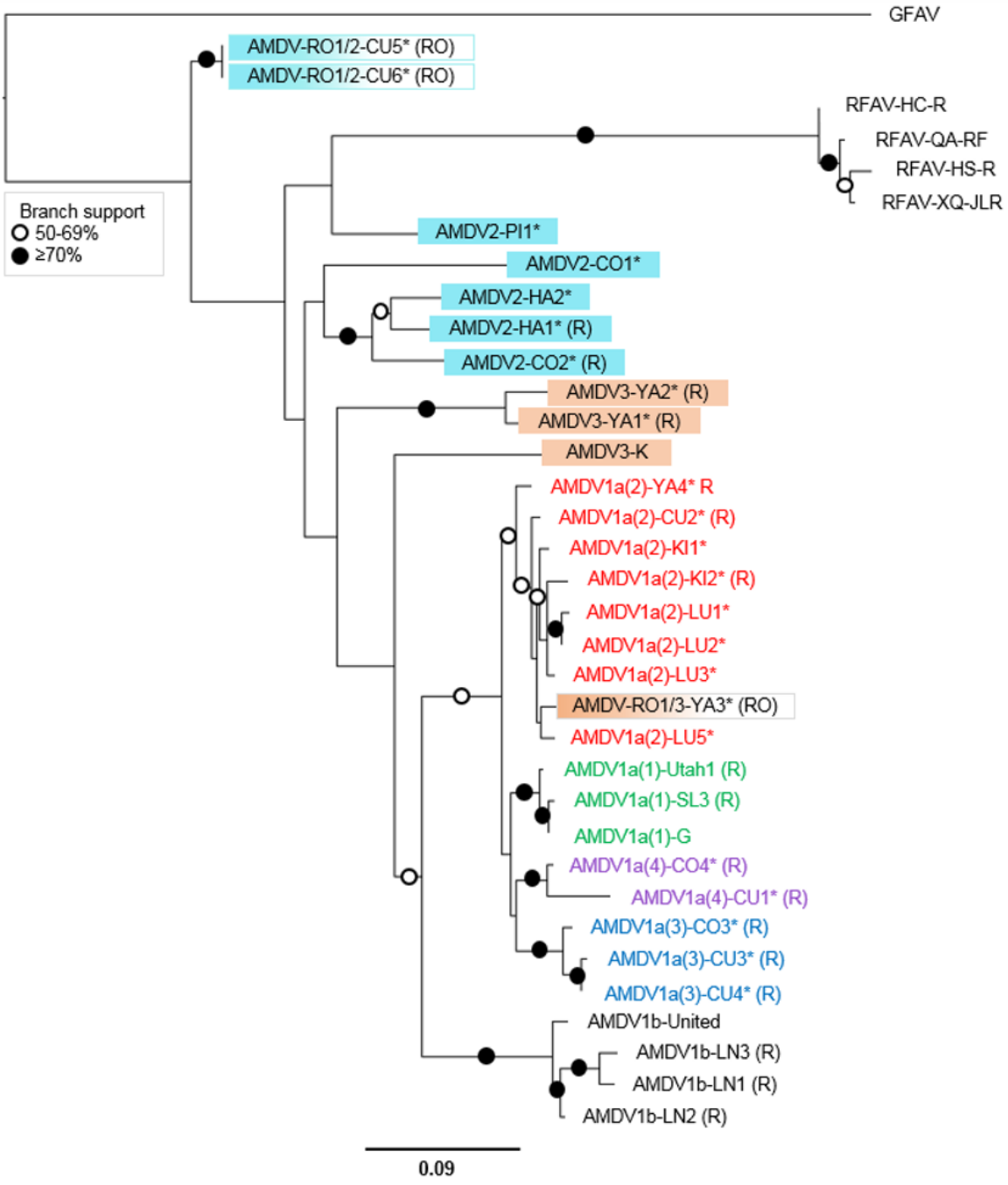
Figure 4.8. ML mid-rooted phylogenetic analyses, based on the recombination-free segments of the entire coding region, NS1 and VP1 ORFs of the genus *Amdvoparvovirus*.
Refer to Figure 4.6 for description of sequences, symbols and colors.

4.2.3.3 Phylogenetic analyses of the partial fragments

Phylogenetic analyses of the HVR of the NS1 gene (nts 587 to 922) (NS1-HVR) of the Group 3 plus RFAV and GFAV sequences, including and excluding the three outlier isolates, are shown in Figures 4.9A and 4.9B, respectively. In the NS1-HVR phylogenies, including and excluding the three outlier sequences, 42% and 55% of the branches, respectively, had bootstrap supports higher than 70%. The AMDV-4-CU7 (the most divergent isolate among the local and GenBank AMDV sequences) had 100% nt identities with AMDV-2-CO2 in this region, and thus was excluded from analyses. Similarly, isolate AMDV-1a(2)-LU4 was excluded from analyses, because of having exactly the same sequence as AMDV-1a(2)-CU2 in this region.

Phylogenetic analyses of the HVR of the VP2 gene (nts 2728 to 3255) (VP2-HVR) of the (Group 5), including and excluding the three outlier isolates, are shown in Figures 4.10A and 4.10B, respectively. In the VP2-HVR phylogenies, including and excluding the three outlier sequences, only 31% and 27% of the branches, respectively, had bootstrap supports higher than 70%. In this region, the AMDV-1-G had 100% nt identities with AMDV-1-SL3, and thus was excluded from the analyses.

A) HVR of the NS1, including the three outlier isolates



B) HVR of the NS1, excluding the three outlier isolates

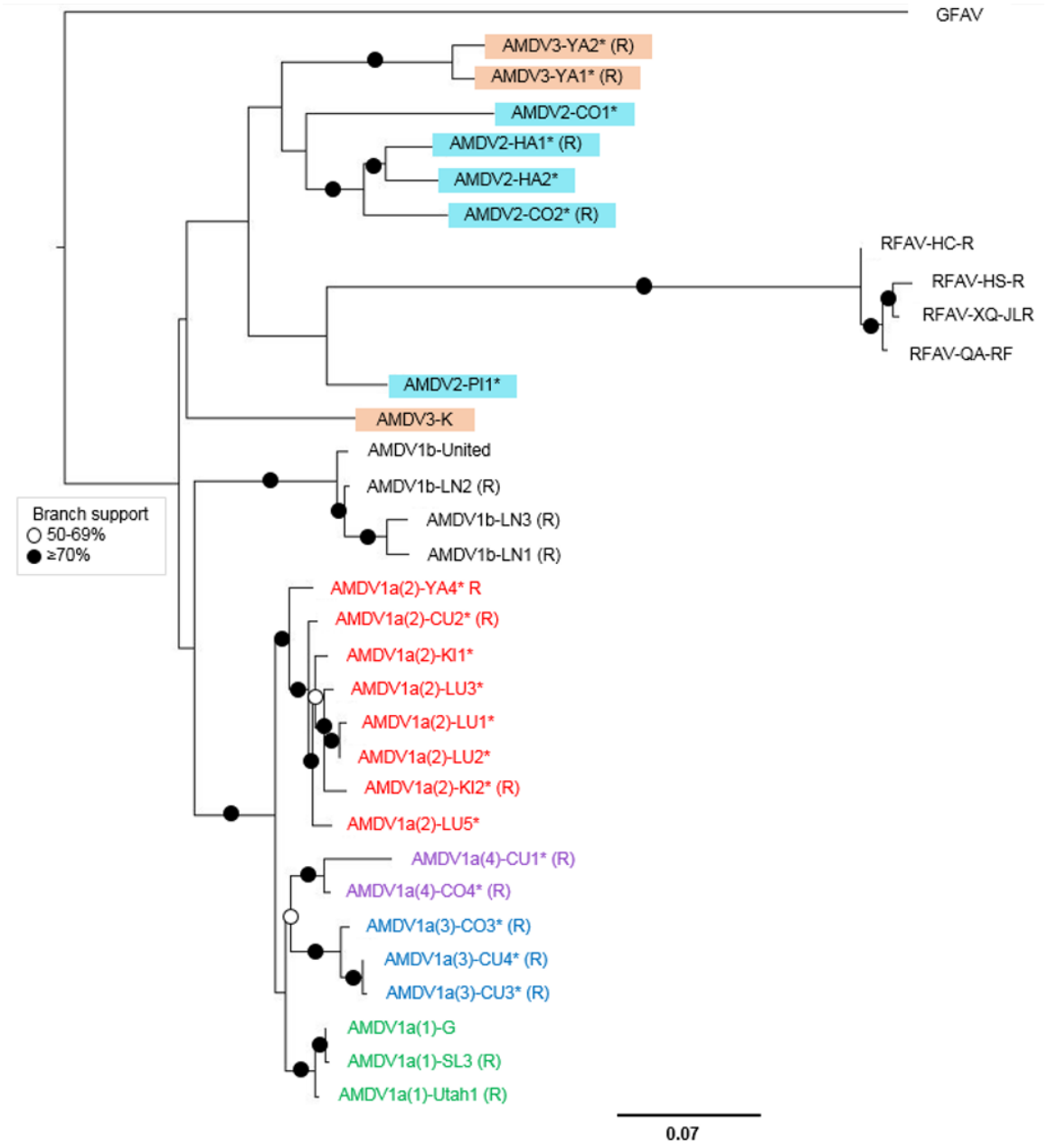
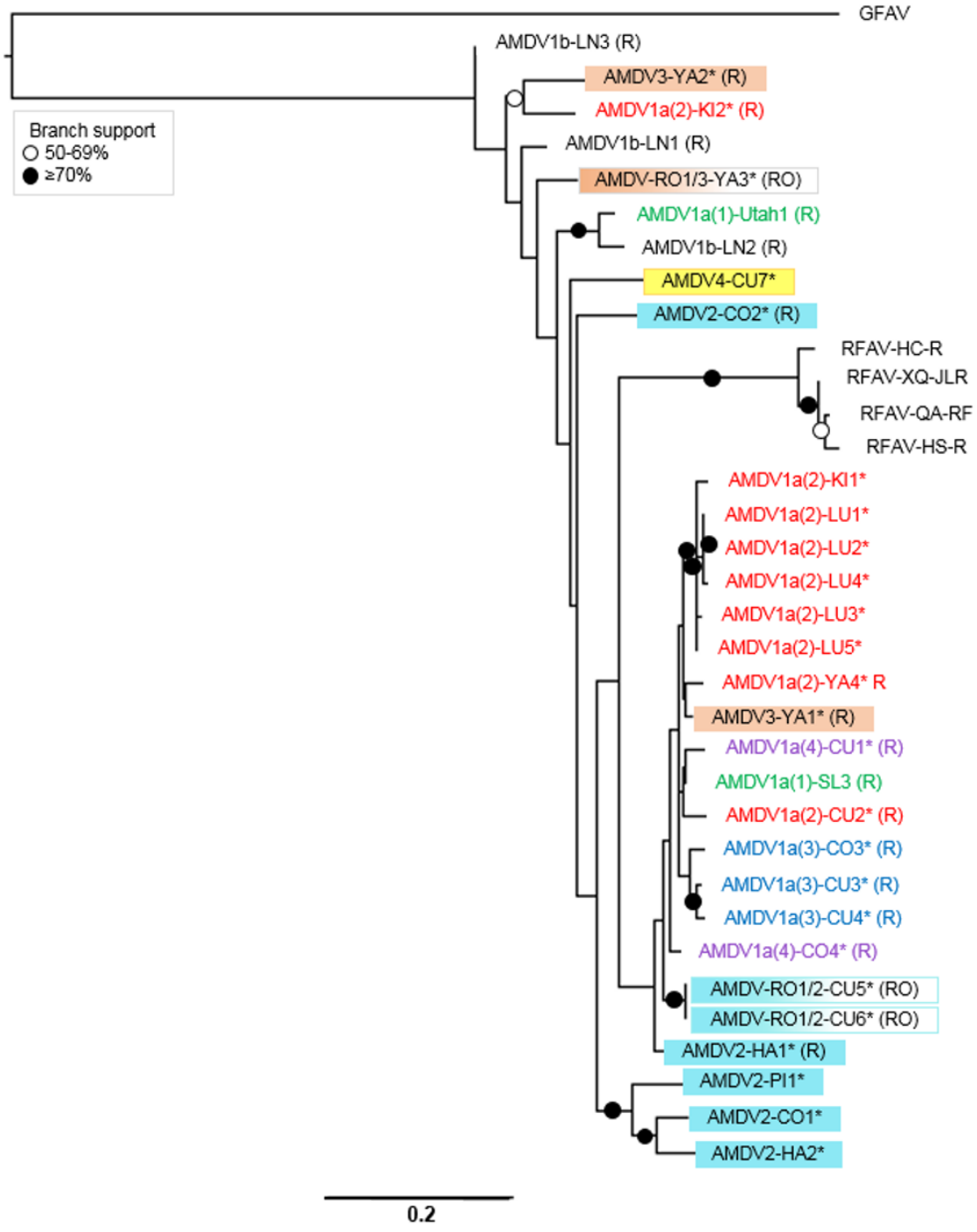


Figure 4.9. ML mid-rooted phylogenetic analyses of the hypervariable region of the NS1 gene in the genus *Amdvoparvovirus.*
These analyses included nucleotide 587 to 922 of the AMDV-G. A) including the outlier isolates CU5, CU6 and YA3, B) excluding the outlier isolates. The AMDV-4-CU7 and AMDV-1a(2)-LU4 isolates had 100% nt identities with AMDV-2-CO2 and AMDV-1a(2)-CU2, respectively, in the HVR of the NS1 gene, and thus were excluded from the analyses. Refer to Figure 4.6 for description of sequences, symbols and colors.

A) HVR of the VP2, including the three outlier isolates



B) HVR of the VP2, excluding the three outlier isolates

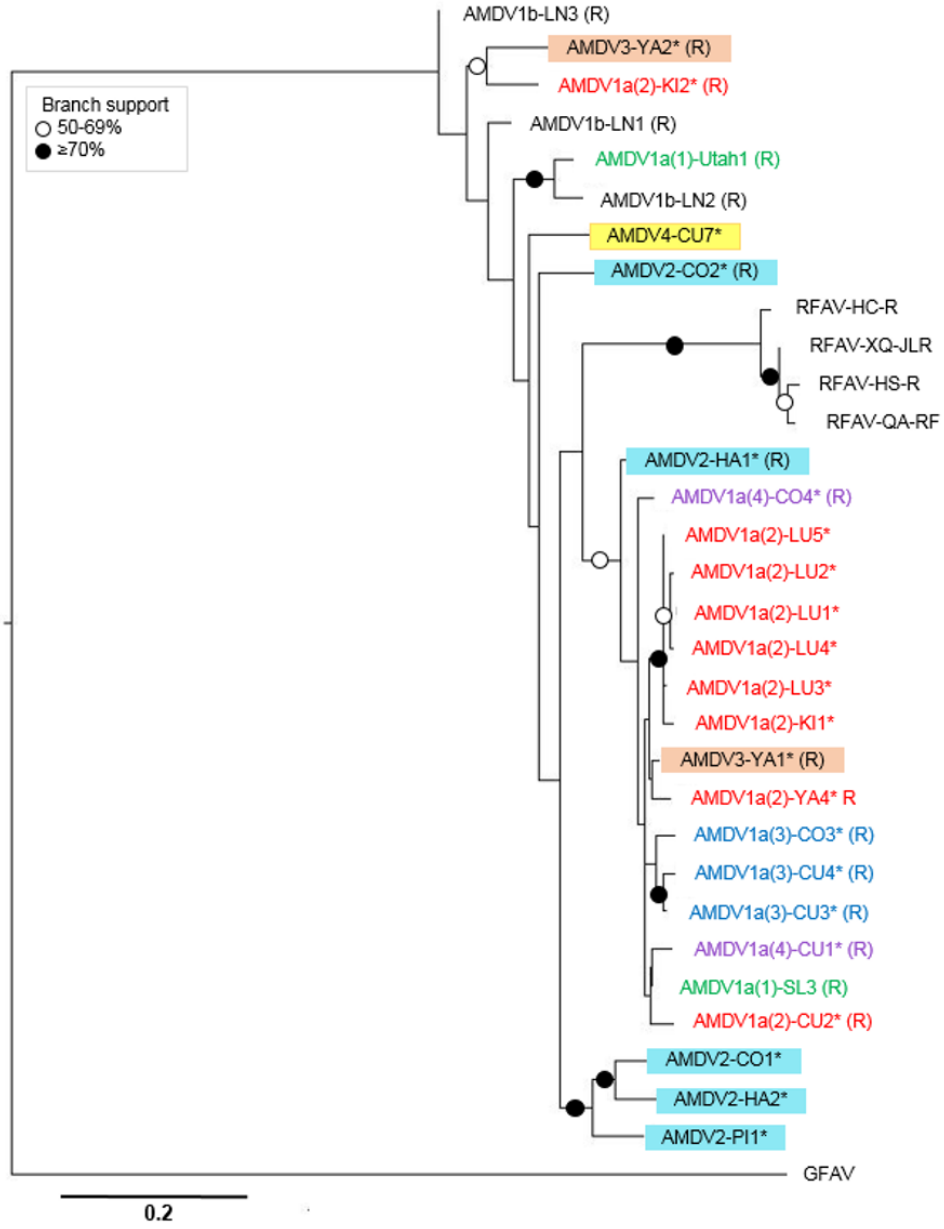


Figure 4.10. ML mid-rooted phylogenetic analyses of the hypervariable region of the VP2 gene in the genus *Amdvoparvovirus.*
These analyses included nucleotide 2728 to 3255 of the AMDV-G. A) including the outlier isolates CU5, CU6 and YA3, B) excluding the outlier isolates. The AMDV-1-G isolate had 100% nt identities with AMDV-1-SL3, in the HVR of the VP1/2, and thus was excluded from analyses. Refer to Figure 4.6 for description of sequences, symbols and colors.

4.2.3.4 Identification of phylogenetic marker

Six phylogenies were constructed by moving a 300-aa window 50 aa at a time along the length of the NS1 protein alignment of the Group 5 sequences, excluding the three outlier isolates, CU5, CU6 and YA3. These five phylogenies were based on aa positions at 1 to 300, 50 to 350, 100 to 400, 150 to 450, 200 to 500 and 250 to 550 of the NS1 protein alignment. Among the six conducted phylogenies, only the phylogeny based on aa position 150 to 450, showed topologies similar to what found based on the ECR of *Amdoparvoviruses* and thus was a phylogenetic marker for this group of viruses.

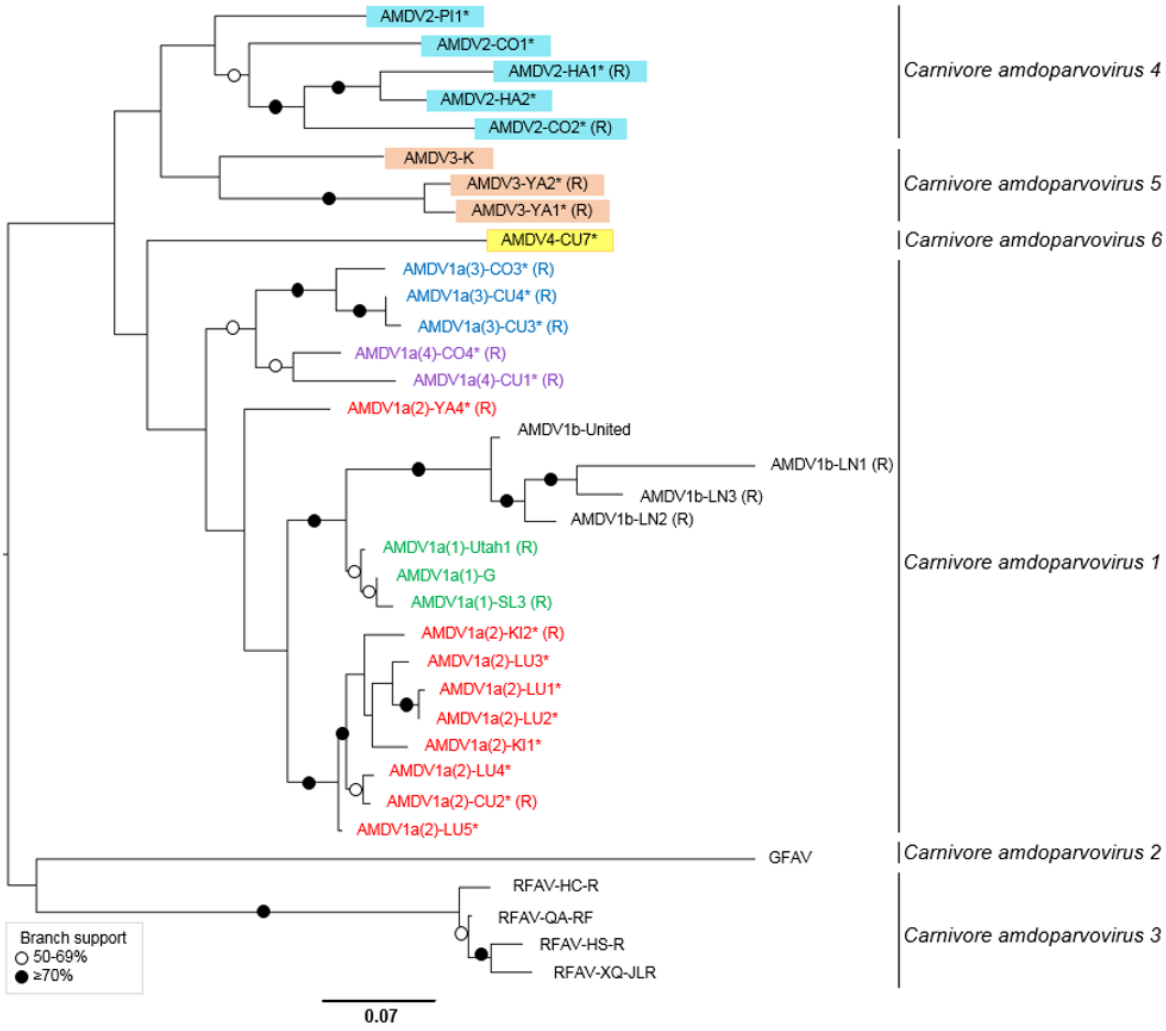


Figure 4.11. ML mid-point rooted phylogenetic analysis of the phylogenetic marker of the genus *Amdoparvoviruses*.
This marker locates at amino acid positions 150 to 450 of the NS1 protein (nt 650 to 1513 of the AMDV-G genome). Refer to Figure 4.6 for description of sequences, symbols and colors.

CHAPTER 5. DISCUSSION

A very high percentage (93.3%) of free-ranging mink in Nova Scotia are infected with AMDV (Farid, 2013), and are thus likely to be major reservoirs of the virus, possibly transmitting the virus to farmed mink (Farid et al., 2012). There is, however, little information about the genome sequence of the AMDV isolates circulating in free-ranging mink in Nova Scotia. In addition, AMDV isolates from farmed mink, whose entire coding sequence is available in public databases are limited to the AMDV-G and Utah1 (Bloom et al., 1988; 1990), LN1, LN2, LN3 (Li et al., 2012) and SL3 (Schuierer et al., 1997). Most published sequences which have been used for AMDV identification and epidemiological studies are shorter than 600 bp (Oie et al., 1996; Olofsson et al., 1999; Mañas et al., 2001; Knuuttila et al., 2009; Jahns et al., 2010; Jensen et al., 2012; Nituch et al., 2012; Sang et al., 2012; Knuuttila et al., 2015; Leimann et al., 2015), and are based on two pairs of primers published by Oie et al. (1996) and Olofsson et al. (1999). These short sequences led to inconsistencies in the literature and to misclassification and misinterpretation of epidemiological studies because it was not clear if partial regions could represent phylogenetic signals of the AMDV genome. Thus, a large number of longer sequences were required to evaluate the validity of the partial region analysis as well as gaining a better view of the diversity and evolution of *Amdoparvoviruses* in Nova Scotia and globally. This study was conducted to explore the genetic characteristics of the entire coding region of AMDV isolates circulating in free-ranging mink in Nova Scotia, and results from this study were compared with published isolates from farmed mink. This study expanded the

number of isolates sequenced, and provided important new findings about genetic characteristics and evolutionary dynamics of AMDV isolates in free-ranging and farmed mink. The genetic diversity of AMDV isolates was also extended by discovering three novel *Amdoparvovirus* species.

5.1 Heterogeneity and taxonomy of the species *Amdoparvovirus*

One of the objectives of this study was to determine the genetic variability of AMDV isolates circulating in free-ranging mink in Nova Scotia, as no published information was available. An AMDV sequence database for free-ranging mink is needed as a reference for monitoring the movement of AMDV between farmed and free-ranging mink, helping mink farmers set up appropriate biosecurity systems to block transmission of the virus from wild animals. Due to the presence of palindromic motifs and secondary structures at the 3′ and 5′ terminal regions, approximately 92% of the genome of 25 local AMDV isolates was sequenced, including the known coding sequences and the partial 3′ terminal region (Bloom et al., 1988). A high degree of heterogeneity was detected among the 25 local isolates circulating in free-ranging mink in Nova Scotia (Table A17). Such a high level of variability can be attributed to single-stranded DNA viruses that show high mutation rates, comparable to RNA viruses (Duffy et al., 2008; Streck et al., 2011) as well as a high incidence of genetic recombination in members of the family *Parvoviridae* (Shackelton et al., 2007). In addition to a high mutation rate, other factors have been proposed as possible factors contributing to the high genetic diversity of this virus, including the high prevalence of AMDV on farms and in the

wild, ability of the virus to establish a chronic persistent infection and a high degree of resistance to environmental conditions such as high temperatures (Canuti et al., 2015) and the long evolutionary history of AMDV (Gottschalck et al, 1994).

One of the most notable findings of this study was discovering three novel *Amdoparvovirus* species, which were designated as *Carnivore amdoparvovirus 4* (including CO1, CO2, HA1, HA2 and PI1), *Carnivore amdoparvovirus 5* (including YA1 and YA2 isolates) and *Carnivore amdoparvovirus 6* (including CU7 isolate) (Figure 4.7A). Identification of three novel species was based on the species demarcation criteria proposed by Cotmore et al. (2014). According to this criterion, the NS1 amino acids of viruses within a species should share more than 85% sequence identity, while diverging by more than 15% from viruses in other species and forming a distinct phylogenetic clade. The novel species also met the species demarcation threshold of 93% over the entire coding region, proposed in this study (Figure 4.5). These novel species contained highly variable non-structural and structural proteins (Tables 4.2, 4.4, A9 and A10) and a short NS3 protein (Table A5). Similarly the NS3 protein of the GFAV and RFAV (*Carnivore amdoparvovirus 2* and *3*, respectively) had shorter lengths that those of the *Carnivore amdoparvovirus 1*. The NS3 protein is reported to be expressed less than the NS2 protein during AMDV infection and has critical roles in viral replication (Huang et al., 2014). In the current study, the highly variable N-terminus of the NS1 protein was found to have an important role in speciation of *Amdoparvoviruses*. This was because genome comparison of the *Amdoparvoviruses* showed five amino acid positions (residues 16-W/F, 76-C/F, 120-S/A, 122-K, 124-I) in this region unique

to the novel *Carnivore amdoparvovirus 4*, *5*, and *6* species as well as the CU5 and CU6 isolates. These residues were not observed in members of the *Carnivore amdoparvovirus 1*, or in the YA3 isolate (residues 16-Y, 76-H/N/W/Y, 120-K, 122-Q, 124-F) (Table 4.8). The three novel *Amdoparvovirus* species were more related to AMDV viruses than to RFAV or GFAV viruses and could be considered AMDV-like viruses. In this study, six other *Amdoparvoviruses* were only partially sequenced due to the high variability of their left ORF, but were not included in the analyses (Table 3.1). The population of the novel *Amdoparvoviruses* in free-ranging mink in Nova Scotia should not be limited to only those whose entire coding region was sequenced and analyzed in this study. The highly pathogenic K isolate, whose left ORF was reported on GenBank (Alexandersen 1986; Gottschalck et al., 1991), was categorized with the *Carnivore amdoparvovirus 5* species. The K isolate cannot be assigned taxonomic status because its entire coding region should be described in order to be assigned as a virus belonging to the family *Parvoviridae*, as stated by the ICTV (Cotmore et al., 2014). The current species within the genus *Amdoparvovirus* include *Carnivore amdoparvovirus 1*, which contains AMDV (Bloom et al., 1988; 1990), *Carnivore amdoparvovirus 2*, which contains GFAV (Li et al., 2011) and the proposed *Carnivore amdoparvovirus 3*, which contains RFAV (Shao et al., 2014). Isolates LU1 to LU5, KI1, KI2, CO3, CO4, CU2 to CU4 and YA4 had high similarities with the previously reported AMDV isolates AMDV-G, Utah1, SL3, LN1 to LN3, and were all classified in the *Carnivore amdoparvovirus 1* species (Figure 4.7A).

5.2 Epidemiology of the *Amdoparvoviruses* in free-ranging mink

Mink genotype analysis was not performed in this study and it is not clear if any of the collected samples belong to free-ranging domestic, wild or domestic-wild hybrid mink. Because the majority of mink ranches are located at the western part of Nova Scotia (Farid et al., 2012), it is possible that samples used in this study, especially those obtained from the western region, contained domestic escapees. Escape of captive mink and their hybridization with wild mink populations have been suggested based on genotyping of free-ranging mink in Ontario (Kidd et al., 2009; Nituch et al., 2012). Although escaped mink are more likely to carry viral isolates circulating on mink ranches, they eventually become co-infected with viral isolates circulating in the wild mink through horizontal transmission. The origin of those mink which were used in the current study may not have a crucial effect on the results.

Most of the viral isolates which were obtained from the same county in Nova Scotia were distributed into different phylogenetic clades, implying different variants of AMDV are circulating in the same county. A small-scale phylogeographical clustering was observed in isolates from Halifax, Lunenburg and Kings counties (Figure 4.7A), in agreement with previous studies in free-ranging mink in Ontario (Nituch et al., 2012). In the current study, the known pathogenic strain Utah1 and intermediate pathogenic isolates SL3, clustered in the same clade with the nonpathogenic isolate AMDV-G, and were separate from the pathogenic isolates K and United, which each clustered in different clades. Pathogenicity of the local isolates were not known, and because phylogeny of the

known viral strains did not reflect their pathogenicity, it was impossible to comment on the pathogenicity of the local strains based on their position in the phylogeny. The observation that most of the AMDV isolates did not cluster according to their pathogenicity, geographical origin or year of collection, was in agreement with those described previously (Schuierer et al., 1997; Olofsson et al., 1999; Knuuttila et al., 2009; Nituch et al., 2012; Sang et al., 2012; Knuuttila et al., 2015; Leimann et al., 2015).

More than half the viral isolates obtained from free-ranging mink in Colchester, Cumberland, Kings and Lunenburg counties in Nova Scotia (members of the *Carnivore amdoparvovirus 1*) were closely related to the known Utah1, United and SL3 isolates obtained from farmed mink, and the laboratory adapted AMDV-G (Figure 4.7A). Several members of the *Carnivore amdoparvovirus 1* isolates in free-ranging mink were more related (95% nucleotide identity) to the NOVA-OB1 isolate obtained from nine infected mink ranches in western Nova Scotia during an outbreak in 2012 and 2013 (Farid and Rupasinghe, 2014). Similarly, this group of AMDV isolates in free-ranging mink seemed to share a common ancestor with the farm D strains in Finland (Knuuttila et al., 2009) because they shared a common ancestor with the Utah1 and United strains. Overall, the similarities of some of the local isolates from free-ranging mink in Nova Scotia with those in farmed mink in Nova Scotia and Finland suggest a continuous movement of viruses between wild and farmed mink populations, similar to reports from Ontario (Nituch et al., 2012) and Newfoundland (Canuti et al., 2016) in Canada, and in Finland (Knuuttila et al., 2015). As such, biosecurity

and AMDV testing of mink farms have to be considered to prevent transmission between farms and with the wildlife. Although the three novel species identified in this study here were highly divergent from those previously described in farmed mink (SL3, United, Utah1) and from the NOVA-OB1, a higher viral diversity may remain undetected on Nova Scotia farms.

In the current study, no information about the physical condition of the trapped mink were available and there is no report on the size of the free-ranging mink populations in Nova Scotia. Trends in the number of pelts from trapped mink have been used to measure changes in the population of free-ranging mink because furbearing animals are difficult to census in the wild. By this measure, the population of free-ranging mink in Canada, including wild and feral ranched-raised mink, has declined during the years 1952 to 2001 (Bowman et al., 2007). Several factors contribute to the decline, including less trapping because of the reduced value of pelt (Wren, 1991, cited in Bowman et al., 2007) and environmental contaminants such as mercury (Yates et al., 2005; Lake et al., 2007).

The introduction of AMDV to wild mink populations by domestic escapees has also been proposed as a factor for the decline in the wild mink populations (Bowman et al., 2007). In Canada, where American mink are native, the effects of AMDV infection on wild mink is limited to one study by Cho & Greenfield, (1978), who observed that 80% (of 120) of the seropositive feral mink from Ontario showed no histological lesions specific to the AD and the infection was unprogressive. Animals with unprogressive infection can live a long time, reproduce, and transmit the infection horizontally and vertically (Cho & Greenfield, 1978). Because severity

of the disease caused by the AMDV depends on the strain of AMDV and the genotype of the mink (Hadlow et al., 1983; Oie et al., 1996), subtle infections of feral American mink in Canada could be indicative of tolerance of this species for certain strains of the AMDV. A limited number of studies, small sample sizes and the possibility of low pathogenicity of the strains circulating in the free-ranging mink in Canada in the above mentioned studies cannot not be excluded.

The three novel *Amdoparvovirus* species discovered in this study were highly divergent. Observation of other isolates with diverged sequences has been reported in free-ranging mink in Ontario (Nituch et al., 2012) and Europe (Mañas et al., 2001; Jensen et al., 2012; Knuuttila et al., 2015; Leimann et al., 2015). Hypothesized by Farid, (2013), it is possible to assume free-ranging mustelids, and particularly mink, in Canada and Europe, carry their own distinct, indigenous *Amdoparvoviruses*, which has been evolving independently long before the start of mink farming in the 1800s. Considering the 40 or 50 million years history of the *Parvoviridae* family (Belyi et al., 2010), a wide range of parvoviruses which infect numerous animal species (Lukashov & Goudsmit, 2001) and the long evolutionary history of the AMDV which is estimated to be 700 evolutionary years old (Gottschalck et al., 1994), it can be hypothesized that AMDV has circulated in wild mustelids in North America before mink farming began in North America (Farid, 2013). It is difficult to believe that a member of this family, AMDV, did not exist in its host in the wild, and appeared in farmed mink spontaneously, or a different member of the *Parvoviridae* family crossed the host species barriers and appeared in farmed mink. If the existence of AMDV in the wild for a long period of time is

true, AMDV isolates circulating in free-ranging mink are valuable resources to use for studying the evolutionary history of AMDV.

5.3 Multiple infection and genetic recombination

A limited number of samples were cloned in the current study and the results suggest either multiple infections with closely related viral isolates resulting from cross-contamination among free-ranging mink or accumulation of mutations. The presence of multiple infection with closely related viral isolates in this study, was not surprising because multiple infection is widespread in natural populations (Bordes & Morand 2011). This study is one of the first to report on the occurrence of multiple infection with *Amdoparvoviruses* in free-ranging mink populations. Canuti et al. (2016) found multiple infections of the farmed mink with different viral strains in Newfoundland, Canada. The results of this study together with that of the Canuti et al. (2016), showed that AMDV infection does not prevent infection by a second virus isolate and frequent recombination should be expected. The presence of polymorphic sites (ambiguous codes) in all but one of the local isolates suggests intra-host mutations or presence of multiple infection with closely related isolates. The presence of polymorphic sites has not been previously reported in free-ranging mink in Canada (Nituch et al., 2012) or Europe (Mañas et al., 2001; Jensen et al., 2012; Leimann et al., 2015; Knuuttila et al., 2015; Persson et al., 2015). Canuti et al. (2016) found polymorphic sites in viruses within the same farmed mink in Newfoundland, and suggested them to result from mutations arising during chronic infection of AMDV.

Frequent inter- and intra-species recombination in the *Amdoparvoviruses* observed in this study (Table 4.10) confirms that recombination represents an important mechanism for generating genetic diversity during the viral lifecycle and plays a significant role in the evolution of *Amdoparvoviruses*. This agrees with other results that parvoviruses are recombination-prone (Shackelton et al., 2007). The majority of recombination identified in this study were inter-species recombination between *Carnivore amdoparvovirus 1* and *4* species and between *Carnivore amdoparvovirus 1* and *5* species. Both resulted in the generation of recombinant isolates belonging to either of the potential parental species. Additionally, there was no evidence of recombination in the most divergent *Amdoparvovirus* species GFAV (*Carnivore amdoparvovirus 2*), RFAV (*Carnivore amdoparvovirus 3*) and CU7 (*Carnivore amdoparvovirus 6*). These novel observations suggest that recombination in *Amdoparvoviruses*, occurs between two species, but do not contribute to generation of a new species and thus, mutation accumulation is the major evolutionary process driving the speciation of *Amdoparvoviruses*.

It was found that recombination can occur throughout the AMDV genome; some regions, however, acted as hotspots of recombination. Recombination hotspots were observed around nucleotide position 1000, located in the overlapped region of the non-structural proteins, and around nucleotide position 3100, located in the overlapped region of the VP1 and VP2 proteins (Alexandersen et al., 1988; Bloom et al., 1988) (Figure 4.3). It was observed that whenever a breakpoint occurred around nucleotide position 2300, which encodes the unique

region of the VP1 protein, there was a strong tendency for the second breakpoint to occur around nucleotide position 3100 (Figure 4.4.). Determining genomic regions that are hotspots for recombination is imperative for understanding the role of recombination in the evolution of AMDV and the emergence of new virus strains in animal populations.

The observation that two groups of closely related AMDV sequences (isolates CU5 and CU6: 99% nucleotide identity; isolates CO3, CO4, CU1, CU3 and CU4: 96% nucleotide identity), each had the same evidence of recombination (Table 4.10) and were obtained from different mink samples, suggests the viability and prevalence of these recombinant isolates in the environment and their eligibility to be considered circulating recombinant forms. Therefore, since the same recombination events were characterized within the genomes of multiple circulating viruses, these two recombination events may have been adaptive enough to have been selectively favored. Interestingly, all these seven circulating recombinant isolates had their 5' recombination breakpoint at a hotspot of recombination (around nucleotide 1000) (Figure 4.3), which suggests that recombination in this region is selectively favorable.

Several isolates from Yarmouth, Colchester, Cumberland and Halifax counties found to be recombinant sequences, suggesting that recombination was not exclusive to a specific geographical region. Isolates from geographically separated counties appeared to have recombined with one another and produced a recombinant isolate obtained from the same or another county (Table 4.10). For example, the KI2 recombinant isolate (from Kings county) contained sequences

closely resembling those found in the adjacent Lunenburg and Yarmouth counties. In recombination analysis, parental isolates are the closest sequences in the dataset to the actual parental isolates. The observation explained above thus, could suggest that similar isolates to those identified in Lunenburg and Yarmouth counties were circulating in Kings county, indicating the diversity of AMDV in free-ranging mink in Nova Scotia is high. Known isolates from China (LN1, LN2 and LN3) had potential parents originated in China, the USA and Canada (Colchester and Yarmouth counties). This finding confirms that Chinese isolates originated from North American mink imported to China. Recombination events were not detected among the AMDV isolates from the same geographical region, and this may be due to the small sample sizes available in each county in Nova Scotia or that recombination between highly similar isolates would not be detected by the recombination methods used in this study. The prediction of six “unknown” major and minor parents for recombination sites indicate other *Amdoparvoviruses* exist and more sequences are required to understand the role of recombination in the evolution of AMDV. Frequent recombination events found in free-ranging mink in Nova Scotia and in farmed mink in Newfoundland (Canuti et al., 2016), highlights the risk of new and distinct AMDV strains arising in mink infected with different AMDV isolates simultaneously.

5.4 Predominance of purifying selection on evolution of the AMDV

The evaluation of the direction of selective forces acting on known AMDV isolates in farmed and free-ranging mink shows a marked predominance of

purifying selection pressure on shaping the evolution of non-structural and structural proteins (Table 4.7). This agrees with previous studies on *Parvoviruses* (Lukashov & Goudsmit, 2001), including AMDV (Knuuttila et al., 2009; Canuti et al., 2016). Predominance of purifying selection in evolutionary history of the left (NS1) and right (VP2) ORFs of the parvoviruses suggests the disadvantageous character of most nonsynonymous substitutions in these viruses (Lukashov & Goudsmit, 2001), i.e., new amino acids often negatively affect the function of the resulting proteins. A stronger purifying selection pressure acting on capsid proteins compared with the non-structural proteins was in agreement with the lower variability of these proteins, brought about by a stronger selection against nonsynonymous substitutions in the capsid proteins of the *Amdoparvoviruses*. It could be hypothesized that a high conservation of capsid proteins is because AMDV does not benefit from evading from immune system, as its pathogenicity is dependent on forming immune-complexes with antibodies.

5.5 Genetic characterization of the AMDV genes and proteins

The higher variability of the non-structural genes and proteins of the AMDV as compared to other parvoviruses (Gottschalck et al.,1994) was supported in the present study by finding that nucleotides and amino acids in the NS1, NS2 and NS3 sequences were twice as variable as those in the VP1 and VP2 sequences (Tables 4.5 and 4.6). In agreement with the previous studies on AMDV (Canuti et al., 2016; Hagberg et al., 2016), a higher degree of nucleotide conservation of the VP1 gene, compared with NS1 was observed within and between AMDV, RFAV

and GFAV (Table 4.11). Since Bloom et al. (1988) and Olofsson et al. (1999), identified a highly variable segment of the NS1 protein, no further efforts have been made to determine other HVR regions of the NS1. In this study, a total of nine HVRs were identified in the NS1 gene. Six were located at the N-terminus of the protein (N-HVR1 to N-HVR6) and three were at the C-terminus (N-HVR7 to N-HVR9) (Figure 4.1. and Table 4.9). Three of the regions (N-HVR3 to N-HVR5) were previously reported as hypervariable segments (Olofsson et al., 1999) but their functions have not been clarified yet. The other six segments (N-HVR1, N-HVR2, N-HVR6 to N-HVR9) were novel HVRs discovered in this study and their functions have to be determined. Gottschalck et al. (1994) found variability in the NS1 gene to be more frequent in the N- and C-terminus of the protein, than in the middle region. In the current study, a similar observation was made with the addition of identifying a larger number of amino acid variation at the N-terminus compared to the C-terminus, where amino acid variations was scattered (Figure 4.1). The N-terminus of the AMDV NS1 protein, which is an overlapped region of the three non-structural proteins (Qiu et al., 2006), therefore, was the most variable region in the AMDV genome. The region at the center of the NS1 protein (amino acid residues 421 to 492) is called the GKRN domain, named after the first four amino acids of this region, and is conserved among parvoviruses (Bloom et al., 1988; Gottschalck et al.,1994), including in the current study.

Only two HVRs were detected in the N-terminus of the VP2 protein (V-HVR1 and V-HVR2) (Figure 4.2. and Table 4.9), reflecting a low degree of variability in the structural proteins, and this agrees with the reports of Canuti et al. (2016) and

Hagberg et al. (2016). The V-HVR1 was a novel HVR and the V-HVR2 was previously identified (Bloom et al., 1988; Oie et al., 1996). The V-HVR2 has been shown to have no role in pathogenicity because changing this segment of the non-pathogenic AMDV-G genome with that of the pathogenic Utah1 did not induce pathogenicity in the AMDV-G (Bloom et al., 1993). The biological importance of the HVR of the VP2 protein has to be determined (Oie et al., 1996).

The results showed that the C-terminal caspase cleavage site at residue D285 of the NS1 protein was fully conserved among the *Amdoparvoviruses* studied (Table A11). This finding confirms the important role of the C-terminal caspase-cleaved fragment of NS1 in AMDV replication because nuclear localization of the full-length NS1 is dependent on cleavage of the C-terminal caspase cleavage site (Best et al., 2003). The results of this study supports the hypothesis that caspase activity and the NS1 cleavage regulate AMDV replication not only in vitro, but also in vivo, regardless of the degree of viral pathogenicity (Best et al., 2003). Although three local AMDV isolates had one or two amino acid variations at the C-terminal caspase recognition site at 282DQTD ↓ S286 region, this variation does not seem to have an important effect on viral replication. This conclusion was drawn because RFAV contains variable amino acids and GFAV contains variable amino acids and a seven amino acid insertion in this region (Table A11), not reported by Li et al. (2011). A high degree of amino acid variation at the left caspase cleavage site (residue D227) and at the caspase recognition site (224INTD ↓ S228), implies that N-terminal caspase-cleaved fragment possibly has a less influential role, if any, in viral replication. This agrees with the results of

Best et al. (2003), who observed no N-terminal caspase-cleaved fragment of NS1 in the nucleus of infected cells. The results of the current study showed that the caspase cleavage site (D420) of the VP2 protein, was conserved among the majority of *Amdoparvoviruses* (Table A12). This finding confirmed the important role of the caspase cleavage of the capsid proteins in persistent infection of the AMDV (Cheng et al. 2010). Cheng et al. (2010), however, reported that mutation of D to glutamic acid (E) at position 420, largely, but not totally, prevented the generation of small capsid cleavage products. The RFAV isolates had a D to a serine mutation in this site, which was not reported by Shao et al., (2014), who observed the same disease manifestations as the AMDV in sick raccoon dogs with high RFAV DNA levels in their blood and tissues. Perhaps the D420S mutation does not completely block the cleavage of the RFAV capsid proteins and further research is required to understand the biological mechanisms of the RFAV infection.

5.6 *Amdoparvovirus* classification accuracy

With no treatment or vaccine available against the AD (Aasted et al., 1998) and failure of viral eradication (Farid et al., 2012), Identifying sources of repeated reappearance of AMDV on cleaned ranches remains the ultimate way of controlling the AMDV infection (Farid et al., 2012). Determining the prevalence and sources of infections requires understanding the relationship of organisms using phylogenetic analysis (McCormack & Clewley 2002). Phylogenetic analysis can show the true relationships only if it has been conducted accurately by paying

special attention to bootstrap values, outgroup selection, recombination and the target genome. In this study, we investigated several factors for accurately classifying members of the *Amdoparvoviruses* species using phylogenetic analysis and thus, proposed a classification method in order to eliminate misinterpretation of epidemiological studies in the future.

5.6.1 Effect of the rooting method on topology of phylogenies

Phylogenetic analyses of the *Amdoparvoviruses* in this study, using several members of each of the genera *Protparvovirus* and *Bocaparvovirus*, as outgroups resulted in phylogenies with different topologies of the ingroup taxa (Figures A20-23). The two genera were found to be the sister taxa of the genus *Amdoparvovirus* by Cotmore et al. (2014). In the current study, several members of each of the mentioned genera had a similar topology with that of the unrooted phylogenetic analysis conducted here (Figure A20 and A22) and several of them had a different topology with that of the unrooted phylogenetic analysis (Figure A21 and A23). Surprisingly, in all the resulting rooted phylogenies, the GFAV and RFAV isolates were clustered within the major AMDV clade, and thus, were appeared to evolve from the AMDV isolates. Among the *Protoparvoviruses* and *Bocaparvoviruses* which were used as outgroup, BuPV1a and CslBoV1, respectively, showed the most compatible topologies with classifications based on sequence identity and genetic distance analyses (Table 4.12). In this study, mid-point rooting was used as an alternate rooting method. Mid-point rooting can be used when evolutionary rates in different branches are not dramatically different (Lemey et al., 2009). The

only study that investigated a molecular clock in the AMDV suggested that molecular evolution of AMDV occurred at a rate that was not constant (Knuuttila et al., 2009). This study was based on the analysis of a partial fragment of the NS1 gene (336 bp), and this result may not be compatible with that of a whole or near-full genome analysis of this virus. In addition, Gottschalck et al. (1994) estimated that AMDV-G and K isolates have separated 700 evolutionary years ago, by calculating nucleotide sequence substitution per year, concluding that heterogeneity of the AMDV is because of the long evolutionary history of the virus and not the high mutation rate. Mid-point rooting was the best available rooting method for phylogenetic analysis in the current study.

5.6.2 Effect of ambiguous codes on phylogenetic analysis

In the current study, a small number of ambiguous codes were detected in 24 of the 25 isolates (Table A2) and sequences containing the ambiguous codes were used for different analyses. This method is frequently used in literature, where cloning of the sequences is not a choice. For example, Carrillo et al. (2005), detected polymorphisms in the chromatograms of the Foot and Mouth Disease Virus, replaced them with ambiguous codes and conducted various analysis, including phylogenetic analysis. In the current study, phylogenetic analyses of viral sequences were conducted using RAxML rather than MEGA, which the latter is more frequently used for this purpose (Knuuttila et al., 2009; Sang et al., 2012; Knuuttila et al., 2015; Canuti et al., 2015; Persson et al., 2015; Canuti et al., 2016; Xi et al., 2016). The MEGA software removes the sequence positions in the

alignment containing gaps and ambiguous codes prior to analysis using a defined site coverage cutoff value by the user. For example, if the site coverage cutoff value is set at 95%, then all the positions having higher than 5% alignment gaps and ambiguous bases are removed (Tamura et al., 2013), which can lead to the loss of informative sites for phylogenetic analyses. In contrast, RAxML treats all ambiguous codes as polymorphic and handles ambiguous codes as a sum of their actual nucleotide compositions. For example, in the case of 'S', RAxML assigns equal likelihoods to both 'C' and 'G'. Gaps are treated as Ns and all four nucleotides are assumed to exist with equal probability for each gap. This procedure can make it more complicated to determine the best phylogenetic tree and reduces bootstrap support values (Felsenstein, 2004). In the current study, phylogenetic analysis of the entire coding region of the *Amdoparvoviruses*, containing ambiguous codes in the local isolates (Figure 4.3A) as well as replacing ambiguous codes gap characters was conducted (Figure A1) and the results showed low bootstrap supported phylogenies in both cases. Therefore, ambiguous codes had minor effect on low bootstrap supports in the original phylogenies. Therefore, ambiguous codes, which only accounted for 0.7% of the genome, had little effect on the low bootstrap supports of the original phylogenies.

5.6.3 Effects of recombination and outlier isolates on phylogenies

Recombination events were identified in the majority of AMDV sequences in the current study (Table 4.10). In order to prevent data loss, the phylogenetic analyses were conducted based on the non-recombinant fragments of the

genomes, as suggested by Martin et al. (2015). Although the resulting phylogenies had high bootstrap supports, clustering of several viral isolates was not consistent with clustering based on sequence identity and distance analyses (Figure 4.8). In this study, three recombinant outlier isolates (CU5, CU6 and YA3) were identified because each had higher than 93% nucleotide and 85% amino acid identity thresholds over the entire coding region and NS1 protein, respectively, to two different species (Tables 4.12 and 4.13). These isolates were evolutionarily related to two different species within the genome segments analyzed (Figure 4.6) and thus, were classified into separate clusters (Tables 4.12 and 4.13). The CU5 and CU6 outlier isolates were intermediates of the *Carnivore amdoparvovirus 1* and *4*, and the outlier isolate YA3 was an intermediate of the *Carnivore amdoparvovirus 1* and *5.* The reason for such conflicts was due to the occurrence of a recombination in the highly variable N-terminus of the NS1 proteins of these isolates, which was found to have important roles in speciation of *Amdoparvoviruses*. Exclusion of three recombinant outlier isolates CU5, CU6 and YA3 from phylogenetic analyses resulted in the formation of well-supported monophyletic clades for new and already identified species (Figure 4.7), which was consistent with the classification based on genetic identity and distance analyses (Tables 4.12 and 4.13). The results showed that although the majority of isolates had evidence of recombination, only those having recombination in the N-terminus of the NS1 protein cause disruptive effects on the bootstrap values and topology of the phylogenies. This was in line with a study by Ané (2011), who suggested only a portion of recombination events alter tree topology and these need to be

detected when constructing a species tree. Identifying recombinant outliers, therefore, is crucial for the accurate classification of species and strains in the genus *Amdoparvovirus*. It should be noted that this concept has not been previously investigated in *Amdoparvoviruses.*

5.6.4 Effect of target genome on phylogenetic analysis

The observation that phylogenetic topology of the NS1 ORF and protein (but not VP1 ORF or protein) sequences was always consistent with the phylogenetic grouping obtained from the analysis of the entire coding region (i.e., supported the same major clades), indicates the NS1 gene evolves in a manner similar to that of the entire-coding region and that the NS1 gene is an appropriate model for studying the evolution of *Amdoparvoviruses*, especially when sequences of the entire coding region are not available. The result of this study confirms the requirement of the NS1 amino acid sequences for the classification of the family *Parvoviridae* (Cotmore et al., 2014). Amino acid sequence comparison of the V-HVR2, showed that AMDV Clusters and Sub-clusters, could not be distinguished on the basis of amino acid sequences in this region (Table A19). The region containing V-HVR2, has been used extensively for identification of the AMDV isolates for determining their origins by using an old type classification (Gottschalck et al., 1991; Bloom et al., 1994; Gottschalck et al., 1994; Oie et al., 1996; Jahns et al., 2010; Li et al., 2012). Based on this classification, AMDV-G is defined as type 1 and highly pathogenic Utah1, K and United isolates are types 2, 3 and 4, respectively (Gottschalck et al., 1991; Bloom et al., 1994; Gottschalck et al., 1994;

Oie et al., 1996). The V-HVR2 of the highly pathogenic TR isolate, however, is exactly the same as that of the non-pathogenic G isolate (Oie et al., 1996). Because it is unknown whether this segment determines the host range or pathogenicity, as proposed by Oie et al. (1996), the V-HVR2 cannot be applied for type classification to estimate the pathogenicity of AMDV strains and this preliminary old type classification should probably be ceased. Furthermore, phylogenetic analysis of the V-HVR2 conducted in this study, showed a miss-classification of the RFAV and AMDV Clusters and Sub-clusters as well as low boot strap values (Figure 4.10). The V-HVR2 has been extensively used for molecular epidemiology of the AMDV in different countries (Mañas et al., 2001; Jahns et al., 2010; Jensen et al., 2012; Nituch et al., 2012; Leimann et al., 2015). Phylogenetic analysis of the known hypervariable region of the NS1 (N-HVR3 to N-HVR5), showed members of the species *Carnivore amdoparvovirus 4* and *5* made a polyphyletic clade (Figure 4.9). In addition, the highly diverged CU7 isolate (*Carnivore amdoparvovirus 6*) was removed from analysis because it had 100% identity with the diverged isolate, CO2, in this region. There was no misclassification in the *Carnivore amdoparvovirus 1* major clade, but analysis of this region of the genome may lead to misinterpretation of epidemiological signals in other species.

5.6.5 Identifying a phylogenetic marker for *Amdoparvoviruses*

Since sequence analysis of the entire coding region might not be practical for any reason, a phylogenetic marker, located in the NS1 protein (Figure 4.11)

was identified. This phylogenetic marker was 300 amino acid long and is located at the positions 150 to 450 of the NS1 protein, containing partial N-terminus region, which was found in this study to have important roles in speciation of *Amdoparvoviruses*. Topology of this phylogenetic marker reflected true clustering of the *Amdoparvovirus* species in phylogenies based on the entire coding region, NS1 ORF and NS1 protein (Figure 4.7A-C). Further research is required to design PCR primers for amplification of this region in *Amdoparvoviruses* based on conserved regions upstream and downstream of the marker. Applying the phylogenetic marker which was proposed in this study, is only recommended if sequencing of the entire coding region or the NS1 protein is not available. The shortcoming of using this phylogenetic marker is the probability of failure in identifying outlier isolates.

To sum up, because partial genome regions result in misclassification of viral isolates, as often reported in the literature, sequence analyses (i.e., identity, divergence, recombination and phylogenetic analyses) of the entire coding region of the *Amdoparvoviruses* was proposed to be the most accurate approach. Careful attention must be paid to all aspects of phylogenetic analysis, including meeting the necessary amino acid thresholds, detecting recombination and removing the recombinant outliers from phylogenetic analysis, outgroup selection and bootstrapping values. These considerations will help harmonize this field, prevents misclassifications and allows for improved epidemiological interpretation of *Amdoparvovirus* sequences.

CHAPTER 6. CONCLUSIONS

The sequence database created in this study is a valuable tool for identifying roots of infection on the mink farms. The discovery of three novel *Amdoparvoviruses* broadened our knowledge on genetic diversity of this virus and widened the list of *Amdoparvoviruses* globally. The obtained near-full genome sequences facilitated the identification of novel hypervariable regions and conserved motifs embedded in the NS1 and VP2 proteins, providing foundational information for future research on protein structure and function of AMDV. The N-terminus of the NS1 protein was founded to have a significant role in speciation of the *Amdoparviruses*. This study was the first to report on the occurrence of multiple infection of the *Amdoparvoviruses* in free-ranging mink populations. Detection of frequent recombination highlighted the risk of distinct AMDV strains arising in farmed mink infected simultaneously with different genotypes of AMDV. Studying the effect of recombination on phylogenetic analysis, marked the urgency to identify and remove the isolates having recombination events in the N-terminus of the NS1 protein, prior to phylogenetic analysis, in order to avoid misclassification. The misclassification of AMDV sequences in the literature as a result of analyzing partial regions of AMDV genome was revealed. Analyzing the entire coding sequences was found to be the foremost prerequisite of classifying *Amdoparvoviruses*. Finally, the proposed phylogenetic marker in this study, will provide great opportunities for mink farmers to identify the source of infection in their farms with a high degree of accuracy as well as low costs, if sequencing the entire coding sequence is not applicable.

APPENDICES

Table A1. Multiple sequence alignment of nucleotides containing ambiguous codes in the entire coding region alignment of *Amdoparvoviruses*.

G¥	224	234	240	241	247	248	252	260	272	275	282	290	302	309	316	370	373	374	379	433	443	454	468	475	485	487	493	498	516	524	531	574	577	620	634	636	653	655	660	669	674
CO1	A	A	C	T	G	C	G	T	G	G	G	**R**	**Y**	A	C	C	T	G	T	C	A	C	C	G	A	T	T	A	G	**R**	G	A	T	T	C	A	C	T	A	G	C
CO2	G	G	.	A	C	.	.	.	.	T	A	A	T	.	.	.	.	C	C	**Y**	.	.	.	.	C	G	.	.	.	A	.	.	.	C	.	.	.	.	.	A	**Y**
CO3	.	G	**M**	A	.	.	T	.	.	.	.	C	T	.	A	.	.	T	.	.	.	.	T	A	G	G	.	.	.	G	A	.	.	.	A	.	A	A	G	A	.
CO4	.	G	.	A	C	.	T	.	.	.	**V**	C	T	.	A	.	.	T	.	.	.	.	T	A	G	G	.	.	.	G	A	.	.	C	A	.	A	A	G	A	.
CU1	G	G	.	A	.	.	.	.	.	T	.	G	T	.	.	.	.	**Y**	C	.	.	.	.	A	C	G	.	.	.	G	.	.	.	**Y**	.	.	A	.	.	.	.	.
CU2	G	G	.	**K**	.	.	A	.	A	.	.	C	T	.	T	.	.	T	G	.	G	.	T	A	G	G	.	.	.	A	A	C	.	C	A	.	.	A	.	**R**	.
CU3	G	G	.	G	.	.	A	.	A	.	.	C	T	.	T	.	.	T	G	.	G	.	T	A	G	G	.	.	.	A	A	C	.	C	A	.	.	A	.	A	.
CU4	G	G	.	G	.	.	A	.	A	.	.	C	T	.	T	.	.	T	**S**	.	G	.	T	A	G	G	.	.	.	A	A	C	.	C	A	**R**	.	A	.	A	.
CU5	G	G	.	G	C	.	A	.	**R**	.	.	C	T	.	T	.	.	T	**K**	.	G	.	T	A	.	G	.	.	.	A	A	C	.	C	A	.	**S**	A	G	A	.
CU6	G	G	.	G	C	.	A	.	A	.	.	C	T	.	**M**	.	.	T	C	.	G	.	**Y**	A	G	G	.	G	.	A	A	T	**Y**	C	A	.	A	A	**R**	A	.
CU7	G	.	.	G	C	G	A	.	.	.	.	C	T	.	T	.	.	T	C	.	G	.	.	A	G	G	.	.	.	A	A	C	.	C	A	.	.	G	G	A	.
HA1	**R**	**R**	.	G	.	.	A	**S**	**R**	.	.	C	T	.	.	**Y**	**W**	A	A	.	G	A	T	.	C	G	.	.	.	A	A	C	.	C	A	.	G	G	G	A	T
HA2	.	G	.	G	**S**	**M**	A	.	**R**	.	.	C	T	C	T	T	.	T	G	.	G	.	T	A	G	G	.	.	.	A	A	**V**	.	C	A	.	.	A	G	A	.
KI1	G	.	G	A	C	T	.	.	.	T	.	**M**	T	.	.	.	C	T	A	T	.	.	.	.	C	G	.	.	.	G	.	.	.	C	.	.	.	.	.	.	.
KI2	G	G	A	A	.	.	.	.	.	.	.	C	T	.	.	.	.	T	A	T	.	.	.	.	**S**	G	**K**	.	.	.	C	.	.	C	.	.	.	.	.	A	.
LU1	.	G	.	G	C	T	A	G	A	.	.	C	T	.	.	.	.	T	C	.	.	A	T	**R**	C	C	.	**R**	.	G	A	C	.	C	A	.	.	A	.	A	.
LU2	**R**	G	.	G	C	T	**W**	G	A	**R**	.	C	T	.	.	.	.	**Y**	C	.	**R**	**M**	.	.	C	**S**	.	.	.	G	A	C	.	C	**M**	.	.	A	.	A	.
LU3	.	G	.	G	C	T	A	G	A	.	.	C	T	.	.	.	.	T	C	.	.	A	T	.	C	G	.	.	.	G	A	C	.	C	A	.	A	G	G	A	.
LU4	.	.	.	G	C	.	A	.	A	.	.	C	T	.	.	.	.	T	A	.	G	.	T	.	C	G	.	.	.	G	A	C	.	C	.	.	.	G	.	A	.
LU5	G	G	.	G	C	G	A	.	**R**	.	.	C	T	C	T	T	.	T	G	.	G	.	T	A	G	G	.	.	.	A	A	**V**	.	C	A	.	.	A	G	A	.
PI1	.	G	.	G	C	.	A	.	**R**	.	.	C	T	**M**	T	.	.	T	A	.	G	.	T	A	G	G	.	.	.	A	A	C	.	C	A	.	.	A	G	A	.
YA1	.	.	.	A	.	.	.	.	.	T	.	G	T	.	.	.	C	T	**M**	T	.	.	.	.	**M**	G	.	G	.	G	A	.	.	C	.	.	A	G	.	A	.
YA2	.	.	.	A	.	.	.	.	.	T	.	**S**	T	.	.	.	C	**Y**	**M**	T	.	.	.	.	C	G	.	G	**R**	G	**R**	.	.	C	.	.	A	**K**	.	A	.
YA3	G	G	.	A	.	.	T	G	.	A	.	C	T	.	.	.	.	T	C	.	.	.	T	.	G	G	.	.	A	G	A	C	.	C	A	.	A	.	.	A	A
YA4	G	G	T	G	C	.	A	.	.	.	.	C	T	.	T	.	.	T	A	.	G	.	T	A	**V**	G	.	.	.	A	A	C	.	C	A	.	.	A	.	A	.
G	G	G	.	G	C	.	A	.	A	A	.	G	T	.	.	.	.	T	C	.	.	.	.	.	C	G	.	.	.	A	A	.	.	C	A	.	A	G	G	A	.
LN1	G	G	.	G	C	.	A	.	.	.	.	C	T	.	.	T	.	T	C	G	G	.	.	.	G	G	.	.	.	G	A	C	.	C	A	.	.	.	.	.	.
LN2	G	.	.	A	C	.	A	.	.	T	.	C	T	.	.	T	.	T	C	G	G	.	.	.	G	G	.	.	.	G	A	C	.	C	A	.	.	.	.	.	.
LN3	G	G	.	G	C	.	A	.	A	A	.	C	T	.	.	T	.	T	C	G	G	.	.	.	G	G	.	.	.	G	A	C	.	C	A	.	.	.	.	.	.
SL3	G	.	.	G	C	.	A	.	A	A	.	G	T	.	.	.	.	T	C	.	.	.	.	.	C	G	.	.	.	A	A	.	.	C	A	.	A	G	G	A	.
Utah1	G	G	.	G	C	.	A	.	A	A	.	G	T	.	.	.	.	T	C	G	.	.	.	.	C	G	.	.	.	A	A	.	.	C	A	.	A	G	G	A	.

Note: Ambiguous codes are represented in bold. CO1 is used as the reference sequence.
¥ Positions are based on the AMDV-G sequence (GenBank accession number NC_001662).

Table A1. Continued

G#	704	714	722	731	528	530	541	570	572	575	585	587	588	589	626	628	636	643	644	645	648	708	711	714	717	720	788	789	821	931	934	938	967	976	980	996	999	1001	1013	1015	1016
CO1	G	A	G	G	C	G	T	T	A	C	C	T	A	A	T	**W**	A	G	T	A	A	T	T	**Y**	**Y**	C	**R**	T	G	G	A	C	T	G	C	T	**K**	T	C	C	C
CO2	A	.	.	.	.	.	.	.	.	.	**Y**	.	T	C	A	A	.	A	.	.	.	.	C	C	C	T	G	.	.	.	**R**	.	.	.	.	.	G	.	.	G	T
CO3	A	**R**	.	.	T	.	.	.	G	.	T	.	T	C	A	A	.	.	C	C	C	.	C	T	G	G	G	.	A	A	G	**Y**	A	A	.	.	A	A	.	G	.
CO4	A	G	.	.	T	**R**	.	.	.	.	T	A	T	C	A	A	.	.	C	.	T	.	C	T	G	G	G	.	A	.	G	.	A	A	.	.	A	A	.	.	.
CU1	A	.	.	.	.	.	.	.	.	.	T	.	**R**	C	A	A	C	.	.	G	G	.	C	.	C	T	G	**Y**	**R**	.	G	.	.	.	.	.	G	.	.	.	.
CU2	.	.	.	.	T	.	C	**W**	.	T	T	.	T	C	A	A	C	A	.	C	.	.	C	T	C	**M**	A	.	.	.	.	.	.	A	**Y**	.	C	**Y**	.	.	A
CU3	.	.	.	.	T	.	C	C	.	T	T	.	T	C	A	A	C	A	.	C	.	.	C	T	C	T	A	.	.	.	.	.	.	A	T	.	C	.	.	.	A
CU4	.	.	.	.	T	.	C	C	C	.	T	.	T	C	A	A	C	A	.	C	.	.	C	T	C	T	A	.	.	.	.	.	.	A	T	.	C	.	.	.	**M**
CU5	.	.	.	.	T	.	C	A	.	.	T	.	T	C	A	A	C	A	.	C	**R**	.	C	T	C	T	A	.	.	.	.	.	.	C	.	**Y**	**Y**	.	.	.	**M**
CU6	.	.	**R**	.	**Y**	.	**Y**	.	G	.	T	**W**	T	C	A	A	C	**W**	**K**	**S**	.	.	C	T	G	A	A	.	.	.	G	.	A	A	.	.	A	A	.	G	.
CU7	.	.	.	.	T	.	C	G	.	.	T	.	T	.	A	T	C	T	.	C	.	.	C	T	C	T	A	.	.	**R**	G	.	A	A	.	.	G	**R**	G	**Y**	.
HA1	.	.	.	.	T	.	C	.	.	.	T	.	T	.	A	A	.	A	.	G	.	.	C	T	C	T	A	.	.	.	.	.	.	.	.	.	G	.	.	.	.
HA2	.	.	.	.	T	.	C	C	G	.	T	.	T	C	A	C	C	A	.	C	.	.	C	T	C	T	A	.	.	.	.	.	.	A	T	.	C	.	.	.	A
KI1	A	.	.	.	.	.	.	.	.	.	T	.	T	C	A	C	**R**	T	.	.	.	.	C	C	C	T	A	.	.	.	G	T	**W**	.	.	.	G	.	.	.	.
KI2	A	.	.	A	.	.	.	.	.	.	T	.	.	.	A	A	G	.	.	.	.	.	C	T	C	.	G	.	.	.	G	.	.	.	.	.	G	.	.	.	.
LU1	.	.	.	.	T	.	C	.	.	.	T	.	T	C	.	C	C	T	.	C	.	.	.	T	C	T	A	.	.	A	G	.	.	.	.	.	G	.	.	.	.
LU2	**R**	.	.	**R**	T	.	C	.	R	.	T	.	T	C	.	C	C	T	.	C	.	**Y**	**Y**	T	C	T	A	.	.	.	G	.	.	.	.	.	G	.	.	.	.
LU3	.	.	.	.	T	.	C	.	.	.	T	.	T	C	.	C	C	T	.	C	.	.	.	T	C	T	A	.	.	.	G	.	.	.	.	.	C	.	.	G	.
LU4	.	.	.	.	T	.	C	.	.	.	T	.	T	C	**M**	A	C	A	.	G	.	.	C	T	C	T	A	.	.	.	G	.	.	.	.	.	G	.	.	.	.
LU5	.	.	.	.	T	.	C	C	G	.	T	.	T	C	A	C	C	A	.	C	.	.	C	T	C	T	A	.	.	.	.	.	.	A	T	.	C	.	.	.	A
PI1	.	.	.	.	T	.	C	C	.	.	T	.	T	**M**	A	A	C	A	.	T	.	.	C	T	G	T	A	.	.	.	.	.	.	A	.	.	C	.	.	.	A
YA1	.	.	.	.	T	.	.	.	.	.	T	A	T	C	A	A	C	.	.	.	.	.	C	C	C	T	G	.	.	.	G	.	.	.	T	.	G	.	G	.	A
YA2	.	.	.	.	T	.	.	.	.	.	T	A	T	C	A	A	C	.	.	.	.	.	C	C	C	T	G	.	.	.	G	.	.	**R**	T	.	G	.	**S**	.	A
YA3	.	.	.	.	T	.	.	.	.	.	.	.	.	.	A	A	.	C	.	.	.	.	.	T	C	A	A	.	.	A	.	A	.	A	G	.	C	.	.	.	A
YA4	.	.	.	.	T	.	C	C	.	**Y**	T	.	T	C	A	A	C	A	.	C	.	.	C	T	C	T	A	.	.	.	.	.	.	A	.	.	C	.	.	.	A
G	.	.	.	.	T	.	C	.	.	.	T	.	T	.	A	T	C	T	.	C	.	.	C	T	C	T	A	.	.	.	.	.	.	A	.	.	C	.	.	.	.
LN1	.	.	.	A	.	A	.	.	.	.	.	A	T	C	C	T	C	A	C	.	.	.	.	T	T	A	A	.	.	.	G	.	A	A	.	.	C	.	.	.	.
LN2	.	.	.	.	.	A	.	.	.	.	.	A	T	C	C	T	C	A	.	.	.	.	.	T	T	A	A	.	.	.	.	.	.	A	.	.	C	.	.	.	.
LN3	.	.	.	.	.	A	.	.	.	.	.	A	T	C	C	T	C	A	.	.	G	.	.	T	T	A	A	.	.	.	.	.	.	A	.	.	C	.	.	.	.
SL3	.	.	.	.	T	.	C	.	.	.	T	.	T	.	A	T	C	T	.	C	.	.	C	T	C	T	A	.	.	.	.	.	.	A	.	.	C	.	.	.	.
Utah1	.	.	.	.	T	.	C	.	.	.	T	.	T	.	A	T	C	T	.	C	.	.	C	T	C	T	A	.	.	.	.	.	.	A	.	.	C	.	.	.	.

Table A1. Continued

G*	1098	1118	1127	1129	1130	1141	1156	1239	1266	1272	1276	1290	1299	1302	1317	1333	1340	1345	1417	1419	1424	1449	1479	1482	1512	1566	1576	1595	1598	1605	1734	1746	1749	1761	1763	1767	1769	1778	1795	1848	1853
CO1	A	A	C	A	T	T	A	A	T	A	G	T	G	A	C	T	A	G	A	C	C	C	A	T	G	T	G	A	T	A	T	A	A	G	A	T	G	C	C	C	C
CO2	.	T	.	.	.	.	.	.	.	.	R	.	A	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	C	.	Y	.	.	.	.	.	G	.	.	S	.	.
CO3	.	G	.	.	.	C	M	G	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	G	.
CO4	.	G	.	.	.	C	.	R	.	R	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	.	.	.	.	G	.
CU1	.	G	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	C	.	.	C	.	.	.	C	.	.	.	.	.
CU2	.	.	G	.	.	.	.	.	.	G	.	C	.	T	.	.	.	A	.	T	.	T	M	M	.	C	.	.	.	.	.	.	G	.	G	.	.	.	.	.	.
CU3	.	.	G	.	.	.	.	.	.	G	.	.	.	T	.	.	.	A	.	T	.	T	.	C	.	C	.	.	.	.	.	.	G	.	G	.	.	.	.	.	.
CU4	.	.	G	.	.	.	.	.	.	G	.	C	.	T	.	A	.	A	.	T	.	T	.	C	R	C	.	.	.	.	.	.	G	.	G	.	.	.	.	.	.
CU5	.	.	G	.	.	.	.	.	.	G	.	C	.	T	.	C	.	.	.	T	.	T	.	C	.	C	.	.	.	.	.	.	G	.	G	.	.	.	.	.	M
CU6	.	G	.	.	.	C	.	G	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	T	.	.	.	.	.	.	G	.
CU7	.	G	.	.	.	.	.	.	.	.	.	C	.	G	T	C	.	.	.	.	.	T	.	C	.	C	.	.	C	.	.	.	G	.	.	.	.	.	.	G	.
HA1	R	G	.	.	.	C	.	.	.	G	.	.	.	.	.	A	.	A	.	T	.	T	.	.	.	C	.	.	C	.	.	Y	G	.	.	.	.	.	.	.	.
HA2	.	.	G	.	.	.	.	.	.	G	.	C	.	C	.	A	.	A	R	T	Y	T	.	.	.	C	.	.	.	.	.	.	G	R	.	.	R	.	.	.	.
KI1	.	G	.	.	.	C	.	.	.	.	.	C	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	G	C	.	.	.	G	.	G	.	.	.	.	.	.
KI2	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	C	.	.	.	T	.	.	.	.	.	.	.	.
LU1	.	R	.	M	M	.	.	.	Y	G	.	C	A	C	.	A	G	A	.	T	.	T	.	.	.	C	.	.	C	.	.	.	G	.	.	.	.	.	.	.	.
LU2	.	.	.	.	C	.	.	.	.	G	.	Y	R	S	M	A	R	A	.	Y	.	Y	.	.	.	C	R	.	Y	.	.	.	G	.	R	K	.	Y	.	.	.
LU3	.	G	G	.	C	.	.	.	.	G	.	C	.	C	.	A	.	A	R	T	.	T	.	.	.	C	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
LU4	.	G	.	.	A	C	.	.	.	G	.	C	.	G	.	C	.	.	.	T	.	T	.	.	.	C	.	.	C	.	.	.	G	.	.	.	.	.	.	M	.
LU5	.	.	G	.	.	.	.	.	.	G	.	C	.	T	.	.	.	A	.	T	.	T	.	C	.	C	.	.	.	.	.	.	G	.	G	.	.	.	.	.	.
PI1	.	.	G	.	.	.	.	.	.	G	.	C	.	T	.	A	G	A	.	T	.	T	.	Y	.	C	.	.	.	.	.	.	G	.	G	.	.	.	.	.	.
YA1	.	G	.	.	.	C	.	.	.	G	.	C	.	G	.	C	.	.	.	T	.	T	.	.	.	M	.	R	C	M	.	.	R	.	R	.	.	.	.	.	.
YA2	.	G	.	.	.	C	.	.	.	G	.	C	.	G	.	M	.	R	.	T	.	T	.	.	.	A	.	G	C	.	.	.	G	.	G	.	.	.	.	.	Y
YA3	G	.	.	.	A	C	.	.	.	.	.	C	A	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.
YA4	.	.	R	.	.	Y	.	.	.	G	.	C	A	G	.	A	.	R	.	T	.	T	.	.	.	C	.	G	C	.	.	.	G	.	G	.	.	.	.	.	.
G	.	.	.	.	.	.	.	.	.	G	.	C	.	G	.	A	.	A	.	T	.	.	.	.	.	C	.	.	C	.	.	.	G	.	G	.	.	.	.	.	.
LN1	.	.	.	C	.	C	.	.	.	G	.	G	A	.	.	A	.	A	.	.	.	.	.	.	.	C	.	G	C	.	.	.	G	.	.	.	.	.	.	.	.
LN2	.	.	.	.	.	.	.	.	.	G	.	C	A	.	.	A	.	.	.	T	.	.	.	.	.	C	.	G	C	.	.	.	G	.	.	.	.	.	.	.	.
LN3	.	.	.	.	.	C	.	.	.	G	.	C	.	G	.	A	.	A	.	T	.	.	.	.	.	C	.	G	C	.	.	.	G	.	.	.	.	.	.	.	.
SL3	.	.	.	.	.	.	.	.	.	G	.	C	.	G	.	A	.	A	.	T	.	.	.	.	.	C	.	.	C	.	.	.	G	.	G	.	.	.	.	.	.
Utah1	.	.	.	.	.	.	.	.	.	G	.	C	.	G	.	A	.	A	.	T	.	.	.	.	.	C	.	.	C	.	.	.	G	.	G	.	.	.	.	.	.

Table A1. Continued

G[≠]	1855	1896	1935	2187	2018	2026	2173	2185	2190	2210	2215	2290	2344	2353	2422	2424	2428	2472	2474	2476	2480	2483	2491	2528	2553	2579	2609	2630	2634	2686	2699	2740	2765	2767	2803	2810	2894	2897	2900	2901	2903
CO1	C	C	A	G	C	A	A	A	C	G	A	A	A	T	C	C	T	C	T	C	G	A	C	G	A	A	C	G	C	G	A	A	T	G	G	G	-	G	A	A	G
CO2	.	.	.	.	T	M	G	.	.	.	.	G	.	.	.	.	.	.	C	T	C	.	.	.	.	.	.	.	.	C	T	G	C	A	.	.	G	.	T	C	.
CO3	.	.	.	.	.	.	G	.	.	.	.	-	.	.	.	.	.	.	C	T	C	.	.	.	.	.	.	.	.	C	T	G	C	.	A	.	G	.	T	C	.
CO4	.	.	.	.	.	.	G	.	.	.	.	-	.	.	.	.	.	.	G	T	A	.	.	.	.	.	.	.	.	C	T	G	.	.	.	.	-	C	.	C	.
CU1	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	G	.	A	.	.	.	.	.	.	.	.	C	.	.	W	.	R	.	-	.	C	.	.
CU2	.	.	.	.	T	.	C	G	.	A	.	G	.	C	.	Y	.	.	C	T	C	.	.	.	.	W	.	.	.	C	T	G	C	.	.	.	G	.	C	.	.
CU3	.	.	.	.	T	.	C	G	.	A	.	G	.	C	.	.	.	.	C	T	C	.	.	.	.	.	.	.	.	C	T	G	C	.	.	.	G	C	C	.	.
CU4	.	.	.	R	T	.	C	G	.	A	.	K	.	Y	.	.	.	M	C	T	C	.	.	.	.	.	.	.	.	C	T	G	C	.	.	.	G	.	T	C	.
CU5	.	.	.	.	T	.	C	G	.	A	.	T	.	C	.	.	.	.	C	T	C	.	.	.	.	.	.	.	.	C	T	G	C	.	.	.	G	.	T	C	.
CU6	.	.	.	.	.	.	G	.	.	.	.	-	.	C	.	.	.	.	G	T	A	.	.	.	.	.	.	.	.	C	T	.	C	.	.	.	G	.	C	.	A
CU7	.	.	.	.	.	.	G	.	.	.	.	-	.	C	.	.	.	.	C	T	A	.	.	.	.	.	.	.	.	C	T	G	C	.	.	.	G	.	T	C	S
HA1	.	.	.	.	.	.	T	G	.	A	.	T	.	C	.	.	.	.	C	T	A	.	.	.	.	.	.	.	G	C	T	G	C	R	.	.	G	.	.	C	.
HA2	T	.	.	.	.	.	T	G	.	A	.	-	.	.	.	.	.	.	C	T	C	.	.	.	.	.	.	.	.	C	T	G	C	.	.	.	G	.	C	.	.
KI1	.	.	R	.	.	.	G	.	.	.	R	-	.	.	.	.	.	.	G	T	A	.	.	.	.	.	.	.	.	C	T	G	A	.	.	.	T	.	C	-	-
KI2	T	.	G	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	T	T	M	.	.	.	.	.	.	.	C	.	.	.	.	.	.	R	.	C	.	T
LU1	T	.	.	.	A	.	T	G	.	A	.	G	.	C	.	.	.	A	C	T	C	.	.	.	.	.	.	.	.	C	T	G	C	.	.	.	G	.	.	C	S
LU2	T	.	.	R	.	.	T	G	.	.	.	-	.	C	.	.	.	.	C	T	C	.	.	.	.	.	.	.	G	C	T	G	C	.	.	.	G	.	.	C	.
LU3	.	.	.	.	A	.	T	M	Y	.	.	G	.	C	.	.	.	.	C	T	C	.	.	.	.	.	M	R	.	S	T	G	C	.	.	.	G	T	T	C	.
LU4	T	.	.	.	.	.	T	G	.	A	.	-	.	C	.	.	.	.	C	T	C	.	.	.	.	.	.	.	.	C	T	G	C	.	.	.	T	.	T	C	.
LU5	.	.	.	.	T	.	C	G	.	A	.	G	.	C	.	.	.	.	C	T	M	.	.	.	.	W	.	.	.	C	T	G	C	.	.	.	G	D	M	V	.
PI1	.	.	.	.	T	.	C	G	.	A	.	G	W	C	.	.	.	.	C	T	C	.	.	.	W	.	.	.	.	C	T	G	C	.	.	R	G	.	T	C	.
YA1	Y	S	.	.	.	.	R	.	.	R	.	G	.	C	Y	.	Y	.	M	R	A	.	M	R	.	.	.	.	M	C	T	R	C	.	.	.	T	.	.	C	.
YA2	T	G	.	.	M	.	.	.	.	A	.	G	.	C	T	.	.	.	M	R	A	.	.	A	.	.	.	.	.	C	T	.	C	.	.	.	T	.	.	C	.
YA3	.	.	G	.	.	.	.	G	.	.	.	-	.	.	.	.	.	A	G	T	A	C	.	.	.	.	.	.	.	C	T	.	.	.	.	.	T	.	C	.	.
YA4	.	.	.	.	.	.	G	.	.	.	.	-	.	.	.	.	.	.	A	T	A	.	.	.	.	.	.	A	.	C	R	G	.	.	.	.	G	.	C	.	.
G	.	.	.	.	.	.	C	G	.	A	.	G	.	C	.	.	C	.	C	T	C	.	.	.	.	.	.	.	.	C	T	G	C	.	.	.	G	.	.	C	.
LN1	.	.	.	.	.	.	G	.	.	A	.	-	.	C	.	.	.	.	G	T	A	.	.	.	.	.	.	.	.	C	T	G	.	.	.	.	T	.	C	.	A
LN2	.	.	.	A	.	.	.	.	.	A	.	-	.	.	.	.	.	.	G	T	A	.	.	.	.	.	.	.	.	C	T	G	C	A	.	A	A	.	C	.	C
LN3	.	.	G	.	.	.	G	.	.	.	.	-	.	.	.	.	C	.	C	T	A	.	.	.	.	.	.	.	.	C	.	G	C	.	.	.	T	.	C	.	A
SL3	.	.	.	.	.	.	C	G	.	A	.	G	.	C	.	.	C	.	C	T	C	.	.	.	.	.	.	.	.	C	T	G	C	.	.	.	G	.	.	C	.
Utah1	.	.	.	.	.	.	G	.	.	.	.	T	.	C	.	.	.	.	G	T	A	.	.	.	.	.	.	.	.	C	T	G	C	.	.	.	A	.	C	.	.

Table A1. Continued

G[≠]	2906	2909	2910	2912	2913	2920	2924	2944	2947	3097	3120	3121	3123	3128	3129	3130	3152	3154	3166	3167	3169	3179	3181	3183	3184	3188	3202	3226	3228	3231	3232	3244	3253	3277	3384	3421	3424	3453	3499	3540	3541
CO1	**R**	**Y**	**M**	G	A	T	A	T	G	T	**R**	**R**	C	A	**R**	A	G	G	T	A	**R**	**S**	A	**R**	**M**	G	G	**R**	G	A	G	A	A	C	A	C	A	C	C	G	T
CO2	A	C	T	A	C	.	G	.	A	.	A	G	**Y**	.	A	.	.	A	C	T	A	C	.	G	A	.	A	A	.	G	.	.	.	.	.	.	G	T	.	.	.
CO3	A	C	T	A	C	.	G	.	.	G	A	G	.	.	C	.	.	A	C	C	A	G	.	G	A	.	A	A	.	.	.	G	.	.	.	.	.	T	T	.	.
CO4	C	A	C	.	.	.	.	**Y**	.	.	A	G	.	**R**	**S**	**D**	.	C	.	.	A	G	.	G	G	.	.	A	.	.	.	.	.	.	.	.	G	T	T	A	C
CU1	A	C	A	.	.	.	.	.	A	.	A	G	T	G	C	.	.	A	C	.	A	C	.	G	A	.	A	A	**S**	.	.	.	.	G	**Y**	.	.	.	.	.	.
CU2	A	C	T	A	C	.	G	.	.	G	A	G	.	**R**	C	.	.	A	.	C	A	C	.	G	A	.	A	A	.	.	.	.	.	.	.	.	.	T	.	.	.
CU3	A	C	T	A	C	.	G	.	.	G	A	G	.	G	T	.	.	A	.	C	A	C	.	G	A	.	**R**	A	.	.	.	.	.	.	.	.	.	A	.	.	.
CU4	A	C	T	A	C	.	G	.	.	G	A	G	.	G	C	.	.	A	.	C	A	C	.	G	A	.	A	A	.	.	.	.	.	.	.	.	.	A	.	.	.
CU5	A	C	T	A	C	.	G	.	A	G	A	G	.	G	C	.	.	A	.	.	A	G	.	G	A	A	A	A	.	.	.	.	.	.	.	.	.	T	.	.	.
CU6	C	T	T	.	.	.	.	C	.	.	A	G	.	**R**	C	.	.	A	.	.	A	G	.	G	G	.	.	A	.	.	.	.	.	.	.	.	.	T	**Y**	**R**	**Y**
CU7	A	C	T	**R**	**M**	.	G	.	.	.	A	G	.	G	C	T	.	A	.	.	A	G	.	G	G	A	.	A	.	.	.	.	.	.	.	.	G	T	T	A	C
HA1	A	C	T	A	C	.	G	.	A	**R**	A	G	.	G	C	.	C	A	C	C	A	G	.	G	A	.	A	A	.	.	.	G	.	.	.	.	.	T	.	.	.
HA2	A	C	T	A	C	.	G	.	.	G	A	G	.	.	C	.	.	A	C	C	A	G	.	G	A	.	A	A	.	.	.	G	.	.	.	.	.	**W**	.	.	.
KI1	C	T	T	.	.	.	.	.	A	.	A	G	.	G	T	.	.	A	C	T	A	G	.	G	A	.	.	A	.	**R**	**K**	G	.	**Y**	.	.	.	T	.	.	.
KI2	C	T	T	.	.	.	.	.	A	.	A	G	.	.	A	.	.	A	.	.	A	G	G	G	A	.	.	A	.	.	.	.	**R**	.	.	**Y**	**R**	T	.	.	.
LU1	A	C	T	A	C	**W**	G	.	.	G	A	G	.	C	A	.	.	A	C	C	A	G	.	G	A	.	A	A	.	.	.	**R**	.	.	.	.	.	T	.	.	.
LU2	A	C	T	A	**Y**	.	G	.	A	.	A	G	.	.	C	.	.	C	C	T	A	G	.	G	A	.	A	A	.	.	.	G	.	.	.	.	.	T	.	.	.
LU3	A	C	T	A	C	.	G	.	.	G	A	G	.	G	C	.	.	A	**R**	C	A	C	.	G	A	.	A	A	.	.	.	G	.	.	.	.	.	T	.	.	.
LU4	A	C	T	A	C	.	G	.	.	G	A	G	.	G	C	.	A	C	C	.	A	G	.	G	A	.	A	A	.	.	.	G	.	.	.	.	.	T	.	.	.
LU5	A	C	T	A	C	.	G	.	.	G	A	G	.	G	C	.	.	A	.	C	A	G	.	G	A	A	A	A	.	.	.	.	.	.	.	.	.	A	.	.	.
PI1	A	C	T	A	C	.	G	.	.	G	A	G	.	G	T	.	.	A	.	C	A	G	.	G	A	A	A	A	.	.	.	.	.	.	.	.	.	T	.	.	.
YA1	A	C	T	A	**S**	.	**R**	.	**R**	G	A	G	.	G	C	.	**R**	**M**	C	**M**	A	G	**R**	G	A	**R**	A	A	.	.	.	**R**	.	.	.	.	.	T	.	.	.
YA2	A	C	T	A	C	.	G	.	.	G	A	G	.	G	C	.	**R**	C	C	.	A	G	**R**	G	A	.	A	A	.	.	.	G	.	.	.	.	.	T	.	.	.
YA3	C	T	T	.	.	G	.	.	.	.	A	G	.	.	G	C	.	.	C	C	A	G	.	G	A	.	A	A	.	.	.	.	.	G	C	.	.	T	.	A	.
YA4	C	T	T	.	.	.	.	C	.	.	A	G	.	G	T	T	.	A	.	.	A	G	.	G	G	.	.	A	.	.	.	.	.	.	C	.	G	T	T	A	C
G	A	C	T	A	C	.	G	.	.	G	A	G	.	.	C	.	.	A	C	C	A	C	.	G	A	.	A	A	.	.	.	G	.	.	.	.	.	T	.	.	.
LN1	C	T	T	.	.	.	.	C	A	G	A	G	.	G	T	G	.	A	C	C	A	C	.	G	A	.	A	A	.	.	.	G	.	.	.	.	.	T	.	.	.
LN2	C	T	T	.	.	.	.	.	.	G	A	G	.	G	T	.	C	.	C	C	A	G	.	G	A	.	A	A	.	.	.	.	.	.	.	.	.	T	.	.	.
LN3	C	T	T	.	.	.	.	.	.	G	A	G	.	.	C	G	.	A	C	C	A	G	.	G	A	.	A	A	.	.	.	G	.	.	.	.	.	T	.	.	.
SL3	A	C	T	A	C	.	G	.	.	G	A	G	.	.	C	.	.	A	C	C	A	C	.	G	A	.	A	A	.	.	.	G	.	.	.	.	.	T	.	.	.
Utah1	C	T	T	.	.	.	.	.	.	G	A	G	.	.	C	.	.	A	C	C	A	C	.	G	A	.	A	A	.	.	.	G	.	.	.	.	.	T	.	.	.

Table A1. Continued

G#	3542	3544	3550	3551	3556	3562	3635	3653	3657	3667	3673	3871	3895	4043	4062	4086	4105
CO1	T	A	A	A	C	A	T	A	**R**	T	G	C	C	**R**	A	A	G
CO2	.	.	.	.	T	.	.	.	G	A	C	.	.	G	.	.	.
CO3	A	.	C	.	T	.	C	.	G	G	A	.	.	G	.	.	.
CO4	.	G	G	.	T	T	C	T	G	A	.	.	.	G	.	.	.
CU1	.	.	.	G	**Y**	.	.	.	G	A	.	.	**M**	G	.	.	.
CU2	.	.	C	.	.	.	C	.	G	G	A	.	.	G	.	.	.
CU3	.	.	C	.	.	.	C	.	G	G	C	.	.	G	.	.	.
CU4	.	.	C	.	.	.	C	.	G	G	A	.	.	G	.	.	.
CU5	.	.	C	.	.	.	C	.	G	G	A	.	.	G	.	.	.
CU6	**W**	**R**	C	.	T	T	C	.	G	**M**	.	**Y**	.	G	.	.	**R**
CU7	.	G	.	.	T	T	C	T	G	G	C	.	T	G	.	.	.
HA1	.	.	C	.	G	.	C	**R**	G	G	C	.	.	G	.	.	.
HA2	.	.	C	.	.	T	C	.	G	G	.	.	.	G	.	.	.
KI1	.	.	C	.	.	.	.	.	G	G	.	.	T	G	.	R	.
KI2	.	.	.	.	.	**W**	.	.	G	.	.	.	T	G	.	.	.
LU1	.	.	C	.	.	.	C	.	G	G	**S**	.	.	G	.	.	.
LU2	**W**	.	**M**	**R**	.	.	C	.	G	G	C	.	.	G	**W**	.	.
LU3	A	.	C	.	.	.	C	.	G	G	C	.	.	G	.	.	.
LU4	A	.	C	.	T	T	C	.	G	G	C	.	.	G	.	.	.
LU5	.	.	C	.	.	.	C	.	G	G	A	.	.	G	.	.	.
PI1	.	.	C	.	.	.	C	.	G	G	A	.	.	.	.	.	.
YA1	**W**	.	C	.	.	.	**Y**	.	G	**R**	C	.	.	G	.	.	A
YA2	A	.	C	.	.	.	C	.	G	G	C	.	.	G	.	.	A
YA3	.	.	C	.	T	.	.	.	G	G	.	A	.	G	.	.	.
YA4	.	G	G	.	T	T	C	T	G	A	A	.	.	G	.	.	.
G	A	.	C	.	.	.	C	.	G	G	C	.	.	G	.	.	.
LN1	A	.	C	.	.	.	C	.	G	A	.	.	.	G	.	.	.
LN2	G	.	C	.	.	.	C	G	G	G	.	.	.	G	.	.	.
LN3	A	.	C	.	.	.	C	.	G	A	.	.	.	G	.	.	.
SL3	A	.	C	.	.	.	C	.	G	G	C	.	.	T	.	.	.
Utah1	.	.	C	.	T	.	C	G	G	A	.	T	A	G	.	.	.

Table A2. The occurrence of ambiguous codes in the near-full genome of the local isolates

IUPAC Codes[£]	D	K	M	R	S	V	W	Y	Number	%[¥]
CO1	-	2	2	13	1	-	1	4	23	0.5
CO2	-	1	1	5	-	-	1	1	9	0.2
CO3	-	-	2	3	1	-	-	1	7	0.2
CO4	-	-	2	-	-	-	-	-	2	0.0
CU1	-	-	-	7	1	-	1	3	12	0.3
CU2	-	-	1	4	1	1	1	1	9	0.2
CU3	-	-	2	5	2	-	1	1	11	0.3
CU4	-	1	4	12	2	-	3	9	31	0.7
CU5	-	-	9	15	2	-	1	4	31	0.7
CU6	-	1	4	7	2	-	-	2	16	0.4
CU7	-	-	-	-	-	-	-	-	0	0.0
HA1	-	-	1	2	1	-	-	5	9	0.2
HA2	-	-	1	4	1	-	1	6	13	0.2
KI1	-	1	2	2	1	-	-	2	8	0.2
KI2	-	-	-	3	-	1	-	2	6	0.1
LU1	-	1	3	2	-	-	2	4	12	0.3
LU2	-	-	-	1	-	-	-	-	1	0.0
LU3	-	1	2	3	1	-	-	1	8	0.2
LU4	1	-	2	1	-	2	1	1	8	0.2
LU5	-	-	2	3	-	-	2	1	8	0.2
PI1	-	1	1	4	1	-	1	1	9	0.2
YA1	-	-	2	1	-	-	-	1	4	0.1
YA2	1	-	-	4	1	1	-	1	8	0.2
YA3	-	1	2	7	1	-	3	7	21	0.5
YA4	-	-	1	3	1	-	-	1	6	0.1
Total	2	10	46	111	20	5	19	59	272	-

[£] IUPAC ambiguity codes: D=A/G/T, K=G/T, M=A/C, R=A/G, S=G/C, V=A/C/G, W=A/T and Y=C/T.

[¥] Number of ambiguous positions divided by the length of each isolate.

Table A3. Variable amino acid positions in the NS1 protein of the *Amdoparvoviruses*.

G†	6	7	9	10	11	12	13	14	15	16	17	18	19	21	23	24	25	26	27	29	33	34	35	40	47
G	I	D	Q	R	R	L	Q	D	L	Y	V	Q	L	K	I	N	D	G	E	V	F	Q	Q	D	K
Utah1	L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SL3	.	.	.	K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	D	.	.	.	.	.
United	L	.	.	K	.	.	.	.	.	.	T	.	.	.	V	A	.	.	.	L	.	.	.	.	.
LN1	L	.	.	.	.	.	.	.	.	.	K	.	.	.	V	A	.	.	.	L	.	.	.	.	.
LN2	L	.	.	K	.	.	K	N	.	.	A	.	.	.	V	S	E	.	.	L	.	.	.	.	.
LN3	L	.	.	.	.	.	.	.	.	.	K	.	.	.	.	T	.	.	.	L	.	.	.	.	.
K	L	N	.	.	.	.	.	E	.	F	E	K	F	T	V	A	.	.	.	L	.	.	.	.	.
LU1	.	.	.	.	.	Q/H	.	E	.	.	A	.	.	.	.	A	.	.	.	L	.	.	.	N	.
LU2	.	.	.	.	.	Q	.	E	.	.	A	.	.	.	.	A	.	.	.	L	.	.	.	N	.
LU3	.	.	.	.	.	Q	.	E	.	.	A	.	.	.	.	A	.	.	.	L	.	K	.	N	.
LU4	.	.	.	.	.	.	.	.	V	.	A	.	.	.	I/V	A	.	.	.	L	.	.	P	N	.
LU5	.	N	.	.	.	Q	.	.	.	.	A	.	.	.	I/V	A	.	.	.	L	.	.	P/Q	N	.
CU1	.	N/D	.	R/K	.	.	.	K	.	.	E	.	L/V	.	I/V	A	.	.	.	L	.	.	.	.	.
CU2	.	N	.	.	.	Q	.	E/D	M/L	.	A	.	.	.	I/V	A	.	.	.	L	.	.	P	N	.
CU3	.	N	.	.	.	.	.	.	.	.	A	.	V	.	.	A	E	.	D	L	.	.	.	.	.
CU4	.	N/D	.	.	.	.	.	.	.	Y/F	A	.	V	.	.	T/A	.	.	.	L	.	.	.	.	.
KI1	.	.	.	.	.	Q	.	.	.	.	A	.	.	.	I/V	A	.	.	.	L	.	.	.	N	.
KI2	.	.	.	.	.	.	.	.	.	.	A	.	.	.	V	A	.	.	.	L	.	.	.	N	.
CO3	.	N	.	.	.	.	.	.	.	.	A	.	V	.	.	A	.	.	.	L	.	.	.	.	.
CO4	V	N	H	K	K	.	.	.	.	.	A	.	.	R	.	G	.	.	.	L	.	.	.	.	.
CU5	V	T	.	K	.	.	.	E	.	W	N	.	F	.	V	S	.	.	.	.	.	.	.	.	.
CU6	V	T	.	K	.	.	.	E	.	W	N	.	F	.	V	S	.	.	.	L/V	.	.	.	.	.
HA1	.	.	.	.	K	.	.	.	.	W	T	K	F	.	V	S	.	N	.	I	.	.	.	.	.
HA2	.	.	.	.	.	.	.	E	.	W	T	K	F	.	V	S	.	.	.	.	.	.	.	.	.
CO1	V	T	.	K	.	.	.	E	.	W	T	.	F	.	V	A	.	.	.	I/V	L/F	.	.	.	.
CO2	L	.	K	K	.	V	.	.	.	W	T	K	F	.	V	S	.	.	.	I/L	.	.	.	.	.
PI1	L	.	.	.	.	T	E	E	.	W	T	R	.	Q	V	A	N	.	.	L	.	.	.	N	.
CU7	.	.	.	S	.	.	.	R	.	F	E	.	V	.	V	T	N	.	.	L	.	.	.	N	N
YA1	V	T	.	.	.	I/L	.	E	.	F	E	K	F	T	V	A	.	.	.	L	.	.	.	.	.
YA2	V	T	.	.	.	.	.	N	.	F	E	K	F	S	V	A	.	.	.	L	.	.	.	.	.
YA3	.	.	.	.	.	.	.	.	.	.	K	R	.	A	.	A	.	.	.	L	.	.	.	.	.
YA4	.	.	K	K	.	.	.	.	V	.	A	R	.	.	V	A	.	.	.	L	.	.	.	.	.

Note: AMDV sequences in free-ranging mink are shown in **bold.** A dash represents deletion and a dot signifies the same aa as the AMDV-G. The possible aas are separated by a slash (mixed aas). Untranslatable aas are represented by a question mark.

† Positions are based on the AMDV-G sequence (GenBank accession number NC_001662), which is represented as G.

Table A3. Continued

G†	53	56	57	58	59	63	64	67	68	69	71	72	73	74	75	76	78	79	80	81	82	83	85	86	88
G	R	S	S	D	L	F	D	E	E	N	T	A	S	N	E	H	T	N	N	E	I	N	C	K	T
Utah1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Q	.	.	.	.	.	.	.	.	.
SL3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
United	.	.	.	.	.	.	.	.	.	T	I	.	.	.	.	W	.	K	D	.	.	.	.	.	.
LN1	.	C	.	.	.	.	.	.	.	T	.	.	.	.	.	W	.	K	D	.	.	.	.	.	.
LN2	.	.	.	.	.	.	.	.	.	T	.	V	.	.	.	W	.	K	D	.	.	.	.	.	.
LN3	.	C	.	.	.	.	.	.	.	T	.	I	.	.	.	W	.	K	D	.	.	.	.	.	.
K	.	.	T	.	.	.	.	.	.	T	K	T	.	.	Q	C	.	.	.	D	.	.	.	.	.
LU1	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.	N	.	.	D	.	.	.	.	.	V
LU2	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.	N	.	.	D	.	.	.	.	.	V
LU3	.	.	.	D/E	.	.	.	.	.	.	.	.	.	.	.	N	.	.	D	.	.	.	.	.	V
LU4	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.	N	.	.	D	.	.	.	.	.	V
LU5	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.	N	.	.	D	.	.	.	.	.	V
CU1	.	.	T	E	.	.	.	.	.	.	.	.	.	T	.	.	.	.	D	Q	.	Q	S	.	V
CU2	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.	N	.	.	D	.	.	.	.	.	V
CU3	.	A	.	.	.	.	.	.	.	.	I	.	.	.	.	.	.	.	.	D	.	Q	.	.	V
CU4	.	A	P/S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N/D	.	.	H/Q	.	.	.
KI1	.	.	.	D/E	.	.	.	.	.	.	.	.	.	.	.	N	.	.	D	.	.	.	S	.	V
KI2	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.	N	.	.	D	.	.	.	S	.	V
CO3	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	D	.	Q	S	.	V
CO4	.	.	.	E	.	.	.	.	D	.	.	T	.	.	.	Y	.	T	D	.	.	.	.	.	V
CU5	.	C	.	D/E	.	Y	.	D	.	T	.	.	.	T	N	C	.	.	.	D	.	.	.	.	.
CU6	.	C	P/S	D/E	.	Y	.	D	.	T	.	.	.	T	N	C	.	.	.	D	.	.	.	.	.
HA1	.	.	P	.	.	Y	E	.	.	T	L	.	.	.	D	C	.	.	.	D	.	.	.	.	.
HA2	.	.	P/S	.	.	Y	E	.	.	.	L	.	.	T	.	C	.	.	.	D	.	.	.	.	.
CO1	.	.	A	.	.	Y	E	.	.	T	L	.	.	T	Q	C	.	K	.	D	.	.	.	.	.
CO2	.	C	.	E	.	Y	.	.	.	T	.	.	.	T	N	C	.	.	.	.	.	.	.	.	.
PI1	.	.	.	E	Q	Y	.	.	.	S	K	.	A	D	D	F	.	E	.	.	R	T	.	V	.
CU7	.	.	.	.	.	.	.	.	.	T	.	.	.	.	K	F	G	G	.	D	.	.	.	.	V
YA1	K	.	.	.	.	.	E	.	.	T	.	.	.	T	V	C	.	.	.	D	.	.	.	.	V
YA2	K	.	.	.	.	.	E	.	.	T	.	.	.	T	V	C	.	.	.	D	.	.	.	.	V
YA3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	S	.	D	.	.	.	S	.	V
YA4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	D	.	.	S	.	.	.

Table A3. Continued

G†	91	92	94	95	96	98	101	104	106	107	108	109	110	111	112	116	120	–	122	123	124	136	137	139	142
G	K	T	L	L	I	K	K	R	D	S	N	K	V	N	L	I	K	-	Q	Q	F	N	E	P	V
Utah1	.	.	.	.	L	.	.	.	.	N	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
SL3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.
United	.	.	V	.	V	.	.	.	.	A	D	.	I	.	F	.	.	-	.	H	.	.	.	.	.
LN1	.	.	V	.	V	.	.	.	.	A	D	.	I	.	F	.	.	-	.	H	.	.	.	.	T
LN2	.	A	V	.	.	.	.	.	.	A	D	.	I	.	F	.	.	-	.	N	.	.	.	.	.
LN3	.	.	V	.	V	.	.	.	.	A	D	.	I	.	F	.	.	-	.	H	.	.	.	.	T
K	.	.	V	.	L	.	R	.	.	A	.	.	.	.	F	V	.	-	.	.	.	D	.	S	.
LU1	.	.	V	V	L	.	.	.	.	.	.	.	.	.	F	.	.	-	.	N	.	.	.	.	.
LU2	.	.	V	V	L	.	.	.	.	.	.	.	.	.	F	.	.	-	.	N	.	.	.	.	.
LU3	.	.	V	V	L	.	.	.	.	.	.	.	.	.	F	.	.	-	.	N	.	.	.	.	.
LU4	.	.	V	V	L	.	.	.	.	.	.	.	.	.	F	.	.	-	.	K	.	.	.	.	.
LU5	.	.	V	.	L	.	.	.	.	.	.	.	.	.	F	.	.	-	.	S	.	.	.	.	.
CU1	.	.	Q	.	L	.	.	.	.	T	.	.	.	.	.	.	.	-	.	H	.	.	G	.	.
CU2	.	.	V	V	L	.	.	.	.	.	.	.	.	.	F	.	.	-	.	K	.	.	.	.	.
CU3	.	.	H	.	L	R/K	.	.	.	A	.	.	.	.	F	.	.	-	.	H	.	.	.	.	.
CU4	.	.	H/Q	.	L	.	.	.	.	A	.	.	.	.	F	.	.	-	.	H	.	.	.	.	.
KI1	.	.	M	.	L	.	.	.	.	.	.	.	.	.	F	.	.	-	.	N	.	.	.	.	.
KI2	.	.	M/L/V	.	L	.	.	.	.	.	.	.	.	.	F	.	.	-	.	N	.	.	.	.	.
CO3	.	.	Q	.	L	.	R	.	.	A	.	.	.	.	F	.	.	-	.	H	.	.	.	.	.
CO4	.	.	Q	.	L	.	.	.	N	A	.	.	.	.	.	.	.	-	.	H	.	.	.	.	.
CU5	.	.	L/M	.	L	R	.	.	.	A	S	H	.	E	F	.	S	-	K	.	L	.	.	.	.
CU6	.	.	.	.	L	R	.	R/K	.	A	S	H/R	.	E	F	.	S	-	K	.	L	.	.	.	.
HA1	.	.	.	.	L	.	.	.	.	T	.	R	.	E	F	V	S	-	K	.	L	.	.	.	.
HA2	.	.	.	.	L	.	.	.	N	A	.	S	.	E	F	V	S	-	K	.	L	.	.	P/S	.
CO1	.	A	I	.	L	.	.	.	N	A/T	.	S	.	Q	F	.	S	-	K	.	L	.	.	S	I
CO2	.	.	.	.	L	.	.	.	.	A	.	R	.	E	F	.	S	-	K	.	L	.	.	.	.
PI1	R	.	L/V	.	L	.	.	.	N	A/T	.	S	.	Q	F	.	S	-	K	.	L	D	.	.	.
CU7	.	.	V	.	L	.	.	K	.	A	.	.	.	.	F	V	A	R	K	H	L	.	.	.	.
YA1	.	A	V	G	L	.	R	.	N	A	.	.	.	Q	F	.	A	-	K	.	L	D	.	S	I
YA2	.	A	V	G	L	.	R	.	N	A	.	.	.	Q	F	.	A	-	K	.	L	D	.	.	I
YA3	.	.	V	.	L	R	.	.	.	.	.	.	.	.	F	.	.	-	.	H	.	.	.	.	.
YA4	.	.	V	.	L	.	.	.	.	.	.	.	.	.	F	.	.	-	.	N	.	.	.	.	.

Table A3. Continued

G†	143	144	147	149	150	142	153	155	156	157	158	159	160	161	163	164	165	167	168	170	172	173	174	175	176
G	Q	K	G	F	M	R	L	K	D	L	A	V	I	Y	N	H	H	D	I	D	K	D	P	E	D
Utah1	.	.	.	.	.	.	.	.	.	.	.	L	.	.	.	.	.	.	.	.	.	.	.	.	.
SL3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
United	.	.	.	L	I	K	.	R	.	.	.	A	.	F	.	.	.	.	Q	.	.	.	.	.	A
LN1	.	.	.	L	L	K	I	R	.	.	.	A	.	F	.	.	Q	.	Q	.	.	.	.	.	T
LN2	.	.	S	L	L	K	.	R	.	.	.	A	.	F	.	.	.	.	Q	.	.	.	.	.	A
LN3	.	.	.	L	L	K	.	R	.	.	.	.	.	F	.	.	Q	.	Q	.	.	.	.	.	A
K	.	.	.	.	I	K	.	.	.	.	.	.	.	W	.	.	.	.	.	G	Q	.	S	D	G
LU1	.	.	.	.	L	K	V	K/R	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.
LU2	.	.	.	.	L	K	V	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.
LU3	.	R/K	.	.	L	K	V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LU4	.	.	.	.	L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LU5	.	.	.	.	L	.	.	.	.	.	.	L	.	.	.	.	.	.	.	.	.	.	.	.	.
CU1	.	.	.	.	V	.	.	.	.	.	.	I	.	.	.	.	.	.	.	.	.	.	.	.	.
CU2	.	.	.	.	L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CU3	.	.	.	.	L	K	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CU4	H/Q	.	.	.	L	K	.	.	.	.	.	.	.	.	.	.	.	N/D	.	.	.	.	.	.	N/D
KI1	.	.	.	.	L/V	.	I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KI2	.	.	.	.	L	K	V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	H	.	.
CO3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CO4	H	.	.	.	L	K	.	.	.	.	.	I	.	.	.	.	.	.	.	.	.	.	.	.	.
CU5	H	.	.	.	.	K	.	.	E	.	.	I	.	.	.	.	.	.	.	.	.	.	.	.	.
CU6	H	.	.	.	M/I	K	.	.	E	.	.	I	.	.	.	.	.	.	.	.	.	.	.	.	.
HA1	H	.	.	.	L	K	.	.	.	.	.	C	V	.	D	Y	.	N	.	N	T	.	.	.	.
HA2	H	.	.	.	I	K	.	R	.	.	.	C	V	.	D	Y	.	N	.	N	T	.	.	.	G
CO1	H	.	.	.	L	K	.	R	.	.	.	Y	V	F	.	.	.	.	V	.	.	.	S	.	G
CO2	H	.	.	.	L	K	.	R	.	.	.	C	V	F	D	.	Q	N	T	N	T	.	.	.	G
PI1	H	.	.	.	L	K	.	.	.	.	G	L	.	.	.	.	Q	T	M	.	T	.	.	.	S
CU7	.	.	.	.	I	K	.	.	E	I	.	.	.	F	.	.	.	.	.	.	T	.	.	.	A
YA1	.	.	.	.	I	.	.	.	.	.	.	C	.	.	.	.	.	N	.	D/G	T	.	S	D	G
YA2	.	.	.	.	I	.	.	.	.	.	.	C	.	.	.	.	.	N	M	G	T	.	S	D	.
YA3	.	.	.	.	I	K/R	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N/D	.	.	.
YA4	.	.	.	.	L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Table A3. Continued

G†	177	178	179	183	185	186	189	190	191	192	194	196	197	198	199	202	203	207	208	209	210	211	212	214	215
G	R	A	K	V	I	E	P	T	K	P	K	F	N	K	Q	Q	D	P	V	H	L	R	D	T	F
Utah1	.	.	.	.	V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SL3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
United	K	.	N	.	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.	T	.	S	.	V	I
LN1	K	.	N	.	.	.	.	.	.	.	.	Y	H	.	.	.	E	S	.	P	.	.	.	V	T
LN2	K	.	N	.	.	.	.	.	.	.	.	Y	H	.	.	.	.	.	.	T	.	.	.	V	I
LN3	K	.	.	.	.	.	.	.	.	.	.	Y	H	.	.	.	E	S	.	P	.	.	.	V	I
K	.	.	D	I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	N	.	.	.	.	L
LU1	.	.	.	.	V	.	.	.	.	I	.	.	H	.	.	.	.	.	.	.	I	.	.	.	I
LU2	.	.	.	.	V	.	.	.	.	I	.	.	H	.	.	.	.	.	.	.	M	.	.	.	I
LU3	.	.	.	.	.	.	.	.	T	.	.	.	H	.	.	.	.	.	.	.	M	.	.	.	I
LU4	.	.	.	.	V	.	.	.	R	.	.	.	H	.	.	.	.	.	.	.	.	.	.	.	I
LU5	.	.	.	.	V	.	.	.	.	.	.	.	H/N	.	.	.	.	.	.	.	M	.	.	.	I
CU1	.	.	.	.	.	.	.	.	.	.	R	L	.	.	.	.	E	.	.	N	M	.	E	.	M
CU2	.	.	.	.	V	.	.	.	R	.	.	.	H	.	.	.	.	.	.	.	.	.	.	.	I
CU3	.	.	.	.	.	.	.	.	.	.	.	.	H	.	.	.	E	.	.	V	.	.	.	V	.
CU4	.	.	.	.	.	.	.	.	R/K	.	.	.	H	.	.	.	E	.	.	V	.	.	.	V	.
KI1	.	.	.	.	V	.	.	.	.	.	.	.	H	.	.	.	.	.	.	.	M	.	.	.	I
KI2	.	.	.	.	V	.	.	.	.	P/L	.	.	H	.	.	.	.	.	.	.	M	.	.	.	I
CO3	.	.	.	.	V	.	.	.	.	.	.	.	H	.	.	.	E	.	I	V	.	.	.	V	.
CO4	.	.	.	.	V	.	.	.	.	.	.	L	H	.	.	.	E	.	.	H/P	M	.	.	.	M
CU5	.	.	R	.	.	.	V	.	.	.	.	H	H	.	.	.	E	Q	.	N	M	.	.	.	V
CU6	.	.	R	.	.	.	V	.	.	.	.	H	H	.	.	.	E	Q	.	N	M	.	.	.	V
HA1	.	.	N	.	V	.	.	.	.	.	.	I	H	.	.	K	E	Q	.	N	M	.	E	.	I
HA2	.	.	N	.	.	.	.	.	.	.	.	I/M	H	.	H	K	E	Q	.	N	M	.	.	.	V
CO1	.	.	N	.	.	.	.	V	.	.	.	L	.	R	.	.	E	.	I	V	M/L	.	E	.	V
CO2	.	.	N	.	.	.	.	.	.	.	.	I	H	.	.	K	E	.	.	N	.	.	E	.	L
PI1	.	.	N	.	.	T	V	.	.	.	.	L	.	.	H	.	E	.	.	N	M	.	E	.	V
CU7	.	.	N	.	.	D	.	.	.	.	.	L	.	R	.	.	E	L	I	.	M	Q	E	V	L
YA1	.	D	.	.	.	.	.	.	R	.	.	.	H	.	.	.	.	.	.	N	M	.	E	.	A
YA2	R/K	D	.	.	.	.	.	.	.	.	.	Y	H	R	.	.	.	.	.	N	M	.	E	.	A
YA3	.	.	.	.	V	.	.	.	R	.	.	Y/F	H	.	.	.	.	.	.	.	M	.	.	.	?
YA4	.	.	.	.	V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Table A3. Continued

G†	272	274	276	282	283	286	287	288	289	291	294	296	298	302	303	309	310	311	312	313	314	315	316	321	322
G	E	G	D	D	Q	S	A	T	T	T	S	I	K	T	K	E	V	A	N	P	V	Q	Q	L	Y
Utah1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SL3	.	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
United	.	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LN1	.	.	N	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	D	.	.	K	.	M	F
LN2	.	.	.	.	.	.	.	.	S	.	.	V	.	.	.	.	.	.	.	.	.	.	.	.	F
LN3	.	.	N	.	.	.	.	.	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
K	.	S	S	.	.	.	.	.	S	.	.	V	.	.	.	.	.	.	I	.	.	K	.	M	T
LU1	.	S	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	V	.	.	A	K	.	.	C
LU2	.	S	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	V	.	.	A	K	.	.	C
LU3	.	S	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	V	.	.	A	K	.	.	C
LU4	.	S	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	V	.	.	.	K	.	.	C
LU5	.	S	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.	K	.	.	C
CU1	.	S	.	N	.	.	.	.	.	.	.	.	N	.	.	.	.	V	T	S	A	K	.	M	C
CU2	.	S	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	V	.	.	.	K	.	.	C
CU3	.	S	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	I	A	S	A	K	.	I	T
CU4	.	S	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	V	A	S	A	K	.	I	T
KI1	D	S	.	.	.	.	.	.	.	.	.	.	.	A	Q	.	.	V	.	.	.	K	P	.	C
KI2	.	S	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	K	.	.	.
CO3	.	S	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	V	A	S	T	K	.	M	T
CO4	.	S	.	.	.	.	.	.	.	.	.	.	N	N	.	D	.	V	A	S	.	K	.	V	C
CU5	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	V	A	S	.	K	.	V	C
CU6	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	V	A	S	.	K	.	V	C
HA1	.	.	.	.	.	.	.	.	.	S	.	V	.	N	.	.	L	V	I/V	.	A	K	.	M	C
HA2	.	D/G	.	.	.	.	.	.	.	S	.	V	.	N	.	.	.	.	A	.	A	K	.	M	C
CO1	.	.	.	.	.	.	.	.	S	L	G	V	.	N	.	.	.	V	T	S	L	K	.	M	T
CO2	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	D	D	V	A	L	I	K	.	V	S
PI1	.	.	.	.	K	.	V	.	A	.	.	V	T	N	.	.	.	.	A	.	.	K	.	V	T
CU7	.	.	.	.	K	L	.	.	.	S	.	.	.	.	.	.	.	S	.	Q	.	K	.	M	C
YA1	.	D	.	.	.	.	.	.	S	.	.	.	.	.	.	N	.	T	D	P/L	.	K	.	M	C
YA2	.	D	.	.	.	.	.	.	S	.	.	.	.	.	.	N	.	.	D	.	.	K	.	M	C
YA3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	D	.	.	K	.	M	C
YA4	.	S	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.	T/A	D	.	.	K	.	M	.

Table A3. Continued

G†	323	326	327	329	330	332	333	334	338	339	344	345	346	347	348	349	353	360	364	368	373	374	375	376	377
G	S	S	T	D	A	F	N	V	T	P	I	K	Q	S	D	K	L	P	N	H	K	T	S	T	M
Utah1	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SL3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
United	.	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LN1	T	.	.	.	D	.	.	F	.	.	.	T	L	G	.	.	M	.	.	.	.	.	.	.	L
LN2	.	N	.	.	.	.	.	.	.	.	.	.	.	.	E	.	.	.	.	.	.	.	A	.	.
LN3	C	N	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	A	.	.
K	.	N	.	.	D	Y	M	F	.	.	L	.	M	.	.	R	M	A	.	.	R	.	.	S	N
LU1	.	N	I/T	.	.	.	.	A/V	.	Q	.	.	L	.	.	T	.	.	.	.	.	V	.	S	.
LU2	.	N	I	.	.	.	.	.	.	Q	.	.	L	.	.	T	.	.	.	.	.	V	.	S	.
LU3	.	N	I	.	.	.	.	.	.	P/Q	.	.	L	.	.	T	.	.	.	.	.	V	.	S	.
LU4	.	N	I	.	.	.	.	.	.	Q	.	.	L	.	.	T	.	.	.	.	.	V	.	S	.
LU5	.	N	.	.	.	.	.	.	.	Q	.	.	L	.	.	T	.	.	.	.	.	V	.	S	.
CU1	.	D	.	.	.	Y	M	L	.	.	.	.	L	.	.	R	.	.	.	.	R	.	A	.	.
CU2	.	N	I	.	.	.	.	.	.	Q	.	.	L	.	.	T	.	.	.	.	.	V	.	S	.
CU3	.	D	.	Q	.	.	M	L	.	.	.	.	L	.	.	R	M	.	.	.	R/K	.	.	.	K/T/Q/P
CU4	.	D	.	Q	.	.	M	L	.	.	.	.	L	.	.	R	M	.	.	.	.	.	.	.	T
KI1	.	H	.	.	.	.	.	.	.	P/Q	.	.	L	.	.	T	.	.	.	.	.	V	.	S	.
KI2	.	N	.	.	.	.	.	.	.	Q	.	.	L	.	.	T	.	.	.	.	.	V	.	N/S	.
CO3	.	D	.	Q	.	.	.	F	.	A	.	.	L	.	.	R	M	.	.	.	R	.	.	S	T
CO4	.	D	A	.	.	Y	M	F	.	.	.	.	L	.	.	R	M	.	.	.	R	.	.	.	N
CU5	.	D	I	.	N	Y	M	F	S	Q	.	.	L	.	.	.	.	.	.	Y	R	.	.	.	.
CU6	.	D/N	I	.	N	Y	M	F	T/S	Q	.	.	L	.	.	.	.	.	.	Y	R	.	.	.	.
HA1	.	G	.	.	.	Y	M	L	.	V	L	.	L	.	.	.	M	A	S	.	M	.	.	.	.
HA2	.	A	.	.	.	Y	M	F	.	.	.	.	L	.	.	R	M	.	.	.	R	.	.	S	.
CO1	.	D	.	.	.	Y	.	L	.	.	.	.	L	.	.	R	M	.	.	.	.	.	.	.	.
CO2	T/S	D	A	.	.	Y	M	L	.	.	.	.	L	.	.	R	M	.	.	.	R	.	.	.	.
PI1	.	D	.	.	.	Y	M	L	.	.	.	.	L	.	.	R	M	.	.	.	.	.	.	.	.
CU7	.	T	C	.	.	Y	.	F	.	Q	.	T	L	.	.	T	M	G	.	.	.	V	.	.	K
YA1	M	T	.	.	D	Y	T	H	.	A	.	.	L	.	.	T	M	T	S	.	R	.	.	.	.
YA2	M	T	.	.	D	Y	T	H	.	.	.	.	L	.	.	T	M	T	S	.	R	.	.	.	.
YA3	M	T	.	.	D	Y	T	H	.	A	.	.	L	.	.	T	M	T	S	.	R	.	.	.	.
YA4	M	N	.	.	E	Y	T	Y/C	S	P/S	.	.	L	.	.	.	M	.	S	.	R	.	.	.	.

Table A3. Continued

G†	378	379	381	382	385	386	387	389	390	392	393	394	397	398	401	402	408	409	410	412	413	415	422	423	425
G	I	T	F	D	I	K	F	E	E	D	D	K	L	A	K	D	Q	Y	L	K	V	C	G	G	R
Utah1	M	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SL3	.	.	.	.	V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
United	M	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LN1	L	N	H	.	.	.	.	.	P	N	.	E	.	.	R	.	.	.	.	.	A	.	S	.	.
LN2	L	S	.	.	V	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.
LN3	M	S	L	.	V	.	.	.	.	.	.	.	.	T	Q	.	.	.	.	.	.	.	.	.	.
K	L	.	.	E	.	.	.	.	.	.	E	.	.	D	.	.	.	H	.	.	.	.	S	.	.
LU1	L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	H	.	.	.	A	.	.	K	.
LU2	L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	H	.	.	.	A	.	.	K	.
LU3	L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	H	.	.	.	A	.	.	R	.
LU4	L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	K	.
LU5	L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.
CU1	L	S	L	E	T	.	.	.	.	.	.	.	.	.	.	.	H	F	.	.	.	.	.	.	.
CU2	L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	I	.	.	.	A	.	.	.	.
CU3	L	S/N	L	E	V	.	.	.	.	.	.	.	.	.	.	.	I	.	.	.	A	.	.	.	.
CU4	L	.	L	E	.	.	.	.	.	.	.	.	.	.	.	.	I	.	.	.	A	.	.	.	.
KI1	L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	A	.	.	R	.
KI2	L	.	F/L	.	M	.	.	.	.	.	.	.	.	.	.	.	E	.	.	.	A	.	S	R	.
CO3	L	D	L	E	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.	A	.	.	.	.
CO4	L	N	L	E	.	.	.	.	.	.	.	.	.	.	.	.	K	.	.	.	.	.	.	.	.
CU5	L	N	L	E	.	.	.	.	.	.	.	T	.	.	.	E	H	F	.	.	A	.	.	.	.
CU6	L	N	L	E	.	.	.	.	.	.	.	T	.	.	.	E	H	F	.	.	A	.	.	.	.
HA1	M	N	.	.	.	.	L	.	.	.	E	Q	.	.	.	N	H	K	.	Q	.	G	.	.	V
HA2	L	N	.	E	.	.	.	.	.	.	E	Q	.	.	.	N	H	K	.	Q	.	G	.	.	V
CO1	L	N	.	E	.	.	.	K	.	.	E	Q	.	.	.	.	G	T	.	.	.	S	.	.	V
CO2	L	N	H	E	M	.	.	.	D	.	E	Q	.	.	.	.	.	T	.	.	.	.	.	.	M
PI1	L	N	.	E	.	.	.	.	.	.	E	Q	.	.	.	.	D	M	.	R	.	.	.	.	.
CU7	L	N	L	E	.	.	.	.	P	.	.	Q	V	.	R	N	N	T	F	Q	A	.	.	.	.
YA1	L	N	L	E	.	Q/K	.	.	.	N	E	Q	.	N	R	.	.	I	.	.	A	.	.	.	.
YA2	L	N	L	E	.	.	.	.	.	N	E	Q	.	D	R	.	.	M	.	.	A	.	.	.	.
YA3	L	N	L	E	.	.	.	.	.	.	E	Q	.	N	R	.	.	V	.	.	A	.	.	.	.
YA4	L	N	.	.	V	.	.	.	.	.	.	.	.	.	R	.	.	.	.	.	A	.	S	.	.

Table A3. Continued

G†	426	428	430	445	447	448	449	450	452	454	464	473	475	478	481	484	485	497	598	512	514	515	521	526	532
G	G	I	F	I	K	A	T	V	Y	M	W	I	A	C	F	W	V	V	S	C	I	V	I	V	N
Utah1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	L	.	.	.	.	.	.	.	.	.	.
SL3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
United	.	.	.	.	.	.	.	.	.	.	.	.	.	.	L	.	.	.	.	.	.	.	.	.	.
LN1	.	.	L	.	.	.	.	K	.	.	.	.	.	.	L	.	.	.	.	.	.	.	.	.	S
LN2	.	L	.	.	.	S	V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S
LN3	.	.	.	.	.	.	.	.	.	.	R	.	.	.	L	.	.	I	.	.	V	I	.	.	S
K	.	V	.	.	.	.	V	.	.	.	.	.	.	.	I	Y	.	.	.	S	.	.	.	.	.
LU1	.	L	.	L	.	.	.	.	F	.	.	.	T	.	L	.	.	.	.	.	.	.	.	.	.
LU2	.	L	.	L	.	.	.	.	F	.	.	.	T	.	L	.	.	.	.	.	.	.	.	.	.
LU3	.	L	.	.	.	.	.	.	.	.	.	.	T	.	L	.	.	.	.	.	.	.	.	.	.
LU4	.	L	.	L	.	.	.	.	.	.	.	.	T	.	L	.	.	.	.	.	.	.	.	.	.
LU5	.	V	.	.	R	.	.	.	.	.	.	.	T	.	L	.	.	.	.	.	.	.	.	.	.
CU1	.	L	.	.	.	.	.	.	F	.	.	.	.	.	.	.	I	.	.	.	.	.	.	.	.
CU2	.	V	.	.	.	.	.	.	F	.	.	.	A/V	.	.	.	.	.	.	.	.	.	.	.	.
CU3	.	V	.	.	R	.	.	.	F	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CU4	.	V	.	.	R/K	.	.	.	F	.	.	.	.	.	.	.	.	.	.	.	.	.	.	I/V	.
KI1	.	L	.	L	.	.	V	.	.	.	.	.	T	.	L	.	.	I	.	.	.	.	.	.	.
KI2	.	L	.	.	.	.	A/T	.	F	.	.	.	.	.	I	.	I	.	.	.	.	.	.	.	S
CO3	.	.	.	.	.	.	.	.	F	.	.	I/V	.	.	.	.	.	.	.	.	.	.	.	.	.
CO4	.	.	.	L	.	S	V	.	F	V	.	.	.	L	I	.	.	.	.	.	.	.	.	.	.
CU5	.	.	.	L	.	.	A	.	F	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	N/S
CU6	.	.	.	I/L	.	.	A/T	.	F	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	S
HA1	S/G	.	.	L	.	.	.	.	F	.	.	.	.	.	I	.	.	.	.	.	.	.	.	.	.
HA2	.	.	.	L	.	.	.	.	F	.	.	.	.	.	I	.	.	.	.	.	.	.	.	.	.
CO1	.	V	.	L	.	.	A	.	F	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CO2	.	.	.	L	.	.	.	.	F	.	.	.	.	.	.	.	.	.	.	.	.	.	.	I	S
PI1	.	.	.	L	.	.	.	.	F	.	.	.	.	.	V	.	.	.	.	.	V	.	.	.	.
CU7	.	V	.	.	.	.	V	K	.	.	.	.	.	L	L	.	.	.	.	.	L	.	.	.	.
YA1	.	V	.	L	.	S	V	.	.	.	.	.	.	.	I	.	.	.	.	.	.	I	V	I	.
YA2	.	V	.	L	.	S	V	.	.	.	.	.	.	.	I	.	.	.	.	.	.	I	V	I	.
YA3	.	V	.	L	.	S	V	.	.	.	.	.	.	.	L	.	.	.	.	.	.	.	.	.	.
YA4	.	.	L	L	.	S	V	.	.	V	.	.	.	.	L	.	.	.	C	.	.	.	.	.	.

Table A3. Continued

G†	533	535	542	543	548	550	551	554	555	557	563	564	566	567	568	570	571	572	575	577	579	582	588	590	591
G	A	A	M	I	M	T	I	K	T	I	K	R	L	N	T	D	R	Q	Q	S	E	E	T	P	N
Utah1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SL3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
United	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LN1	.	.	.	V	.	P	.	.	.	V	.	.	.	.	.	.	G	E	.	N	.	.	.	.	.
LN2	.	.	.	V	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.	.	.
LN3	.	S	.	L	.	.	.	.	.	V	.	T	.	S	S	.	.	.	K	.	.	.	.	.	.
K	V	.	.	V	.	P	V	.	.	V	.	T	I	S	.	.	.	I	K	.	.	.	S	.	.
LU1	V	.	I	L	.	P	.	T	.	V	.	K	.	S	.	.	G	E	P	N	.	.	.	.	.
LU2	V	.	I	L	.	P	.	T	.	V	.	K	.	S	.	.	G	E	P	N	.	.	.	.	.
LU3	V	.	I	L	.	P	.	T	.	V	.	K	.	S	.	.	G	E	P	N	.	.	.	.	.
LU4	V	.	I	L	.	P	.	T	.	V	.	K	.	S	.	.	G	.	P	N	.	.	.	.	.
LU5	V	.	I	L	.	P	.	T	.	V	.	K	.	.	.	.	G	.	P	N	.	.	.	.	.
CU1	.	.	.	V	.	P	.	.	.	.	.	K	.	S	.	.	E	.	.	.	.	D	.	.	.
CU2	I	.	.	L	T	.	.	.	.	V	.	K	.	S	.	.	G	.	.	.	.	.	.	.	.
CU3	.	.	.	L	.	.	.	.	.	V	.	K	.	S	.	.	G	.	.	.	D	.	.	.	.
CU4	A/V	.	.	L	.	.	.	.	.	V	.	K	.	S	.	.	G	.	.	.	.	.	.	.	.
KI1	V	.	I	L	.	P	.	T	.	V	.	K	.	S	.	.	G	K	P	N	.	.	.	.	N/K
KI2	.	V	.	L	.	P	.	T	.	V	.	K	.	S	.	.	G	.	S	N	.	.	.	.	.
CO3	V	.	.	L	.	.	.	.	.	V	.	K	.	S	.	.	G	.	.	.	D	.	.	.	.
CO4	.	.	.	V	.	.	.	.	.	.	.	K	.	S	.	.	Q	E	.	.	.	.	.	Q/K	.
CU5	.	.	.	L	.	P	.	.	.	.	.	K	.	S	.	.	E	.	.	.	.	.	.	.	.
CU6	.	.	.	L	.	P	.	.	.	.	.	K	.	S	.	.	E	.	.	.	.	.	.	.	.
HA1	.	S	.	V	.	P	.	T	.	L	.	K	.	S	.	N	G	V	K	.	D	.	S	R	S
HA2	.	S	.	V	.	P	.	T	.	V	.	.	.	S	.	.	G	T	K	N	.	D	S	R	S
CO1	V	.	I	V	.	P	.	T	.	V	.	K	.	S	.	.	G	V	K	.	D	.	S	R	S
CO2	.	.	.	V	.	P	.	T	.	V	R	K	.	S	.	.	E	I	K	N	.	.	.	.	.
PI1	.	V	T	.	.	P	.	T	.	V	.	K	.	S	.	.	G	I	K	N	.	.	S	R	S
CU7	.	V	L	L	.	P	.	.	.	.	A	K	.	S	.	.	N	.	K	.	.	.	.	A	.
YA1	V	S	.	L	.	Q	.	.	.	L	.	A	.	S	.	.	G	I	T	.	D	.	S	A	.
YA2	V	S	.	L	.	Q	.	.	.	L	.	A	.	S	.	.	G	I	T	.	D	.	.	A	.
YA3	V	S	.	L	T	.	.	.	.	L	.	K	.	R	.	.	G	.	P	.	D	.	S	A	.
YA4	.	S	.	L	.	Q	.	.	S	L	.	K	.	S	.	.	G	.	S	N	D	.	S	A	.

Table A3. Continued

G†	592	594	596	597	598	599	600	603	604	606	610	611	616	618	619	620	621	623	626	632	534	635	640
G	S	A	T	A	T	K	N	N	S	P	K	S	N	E	N	C	D	P	S	A	Q	H	H
Utah1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	P	.	.	.
SL3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
United	.	.	S	.	.	.	T	.	.	.	.	T	S	A	D	.	.	.	G	P	.	.	.
LN1	.	.	S	.	.	R	S	.	.	.	G	G	.	T	.	.	.	.	G	P	.	.	.
LN2	.	.	S	.	.	R	S	.	.	.	G	G	.	T	.	.	.	.	G	P	.	.	.
LN3	.	.	S	.	.	.	I	.	.	.	G	G	.	T	D	.	.	.	G	P	.	.	.
K	.	.	.	.	.	.	S	.	L	.	E	T	.	A	.	.	.	.	G	P	.	.	.
LU1	.	.	.	.	.	E	.	.	.	.	R	T	.	.	.	.	N	.	G	.	.	.	.
LU2	.	.	.	.	.	E	.	.	.	.	R	T	.	.	.	.	N	.	G	.	.	.	.
LU3	.	.	.	.	.	E	T	.	.	.	R	T	.	.	.	.	N	.	G	.	.	.	.
LU4	.	.	.	.	.	.	.	.	.	.	R	T	.	.	.	.	.	.	G	.	.	.	.
LU5	.	.	.	.	.	.	.	.	.	.	R	T	.	.	.	.	N	.	G	.	.	.	.
CU1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	P	.	.	.
CU2	L	.	.	.	.	.	.	.	L	.	.	.	.	.	.	.	.	.	.	P	.	.	.
CU3	L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
CU4	L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	V	.	.	.
KI1	.	.	.	.	.	E	.	.	.	.	R	T	.	.	.	.	N	.	G	.	.	Q	.
KI2	.	V	.	.	.	.	S	.	L	.	E	T	H	.	.	.	.	.	G	V	K	.	.
CO3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
CO4	L	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	P	.	.	.
CU5	S/L	.	.	.	.	.	.	.	.	P/A	.	.	.	.	.	.	.	.	G	V	.	.	.
CU6	L	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	G	V	.	.	.
HA1	.	.	.	.	K	E	S	.	L	.	A	T	.	T	.	S	.	.	G	V	.	.	.
HA2	.	.	.	.	.	.	S	.	.	.	E	T	.	T	.	.	.	.	G	V	.	.	.
CO1	T	.	.	.	.	.	S	S	.	.	E	T	.	T	.	.	.	.	G	V	.	.	.
CO2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	N/D	.	.	.	G	V	K	.	.
PI1	V	.	.	V	K	E	S	.	.	.	.	T	.	T	D	.	.	.	G	V	.	.	Y
CU7	.	.	.	.	.	E	S	S	L	.	.	I	S	T	D	.	.	.	G	P	.	.	.
YA1	.	.	.	.	A	.	S	.	L	.	E	T	D	T	.	.	.	.	G	P	.	.	.
YA2	.	.	.	.	A	.	S	.	L	.	E	T	.	T	.	.	.	L	G	P	.	.	.
YA3	.	.	.	.	A	.	S	.	L	.	.	T	.	.	.	.	.	.	G	P	.	.	.
YA4	.	.	.	.	A	.	S	.	L	.	.	T	.	T	.	.	.	.	G	P	.	.	.

Table A4. Variable amino acid positions in the unique region of the NS2 protein of the *Amdoparvoviruses*.

G	62	64	66	69	70	72	73	74	80	81	84	85	90	92
G	Y	T	L	H	N	H	E	E	T	Y	K	E	E	G
Utah1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SL3	.	.	.	.	.	.	.	.	.	.	.	.	.	.
United	.	.	.	.	E	.	.	D	.	.	.	D	.	.
LN1	.	.	.	.	E	.	.	.	.	.	R	R	.	D
LN2	.	.	.	.	K	.	.	.	.	.	R	R	.	D
LN3	.	.	.	.	E	.	.	D	.	.	R	R	.	N
K	F	.	.	Y	.	.	.	.	.	.	R	D	.	.
LU1	.	.	.	.	.	.	G	.	.	.	.	D	.	.
LU2	.	.	.	.	.	.	G	.	.	.	.	D	.	.
LU3	.	.	.	.	.	.	G	D	.	.	.	D	.	.
LU4	.	.	.	.	.	.	.	.	.	.	.	D	.	.
LU5	.	.	.	.	.	.	.	.	.	.	.	D	.	.
CU1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CU2	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CU3	.	.	.	.	.	.	.	.	.	H	.	.	.	.
CU4	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KI1	.	.	I/L	.	.	.	G	.	.	.	.	D	.	.
KI2	.	.	.	Y	.	.	.	.	.	.	R	D	A	.
CO3	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CO4	.	T/K	.	.	.	.	.	.	.	.	.	.	.	.
CU5	.	.	.	.	.	.	.	.	T/S	.	.	.	.	.
CU6	.	.	.	.	.	.	.	.	S	.	.	.	.	.
HA1	F	.	.	.	.	Q	G	.	.	.	S	N	.	N
HA2	F	.	.	.	.	.	.	.	.	.	R	D	.	N
CO1	F	.	H	.	.	.	.	.	.	.	R	D	.	N
CO2	H	.	.	.	.	.	.	.	.	.	.	.	.	.
PI1	F	.	R	.	.	Q	G	.	.	.	.	D	.	N
CU7	.	S	.	.	.	.	G	.	.	.	.	D	.	N
YA1	F	S	.	.	.	C	.	.	.	.	R	D	R	N
YA2	.	S	.	.	.	C	.	.	.	.	R	D	K	N
YA3	F	S	.	.	.	R	.	.	.	.	.	D	.	.
YA4	F	S	.	.	.	C	K	.	.	.	.	D	.	N

Note: See footnotes of the Table A3.

Table A4. Continued

G	93	95	96	99	100	102	105	106	107	108	109	110	113	114
G	K	R	H	K	E	E	A	C	K	A	A	Q	S	A
Utah1	.	.	.	.	G	.	.	S	E	.	.	.	.	.
SL3	.	.	.	.	.	.	.	.	.	.	.	.	.	.
United	R	.	Y	.	G	.	.	S	E	.	.	.	.	.
LN1	.	.	Y	.	R	.	T	S	.	.	.	.	.	.
LN2	.	.	Y	T	G	K	T	S	E	.	T	.	.	.
LN3	R	.	.	.	G	.	.	S	E	.	.	.	.	.
K	.	.	Y	.	G	.	T	S	.	.	T	.	.	.
LU1	.	Q	.	.	G	.	.	.	.	.	.	.	.	.
LU2	.	Q	.	.	G	.	.	.	.	.	.	.	.	.
LU3	.	Q	.	.	G	.	.	.	.	.	T/A	.	.	.
LU4	.	.	.	.	G	.	.	.	.	.	.	.	.	.
LU5	.	Q	.	.	G	.	.	.	.	.	.	.	.	.
CU1	.	.	.	.	.	.	.	S	.	.	.	.	.	.
CU2	.	.	.	.	.	.	.	S	.	.	.	.	.	.
CU3	.	.	.	.	G	.	.	.	.	.	.	.	.	.
CU4	.	.	.	.	G	.	.	.	E	.	T/A	.	.	.
KI1	.	Q	.	.	G	.	.	.	.	.	.	K	.	.
KI2	.	.	Y	.	G	.	.	.	E	E	.	.	.	.
CO3	.	.	.	.	G	.	.	.	.	.	.	.	.	.
CO4	.	.	.	.	.	.	.	S	.	.	.	.	.	.
CU5	.	.	.	.	G	.	.	.	E	T	.	.	.	.
CU6	.	.	.	.	G	.	.	.	E	T	.	.	.	.
HA1	.	.	Y	.	R	.	.	.	E	.	.	.	T	.
HA2	.	.	Y	.	G	.	.	.	E	.	.	.	.	.
CO1	.	.	Y	.	G	.	.	.	E	.	.	.	.	.
CO2	R/K	.	Y	.	R	.	.	.	E	E	.	.	A	.
PI1	R	.	Y	.	G	.	.	.	E	.	.	.	.	V
CU7	R	.	Y	.	G	.	.	S	E	.	.	.	.	.
YA1	.	.	Y	.	G	.	T	S	E	.	.	.	.	.
YA2	.	.	Y	.	G	.	T	S	E	.	.	.	.	.
YA3	.	.	Y	.	G	.	T	S	E	.	.	.	.	.
YA4	.	.	Y	.	G	.	T	S	E	.	.	.	.	.

Table A5. Variable amino acid positions in the unique region of the NS3 protein of the *Amdoparvoviruses*

G	62	63	64	69	70	71	72	74	75	77	79	81	82	84	85	86
G	C	D	C	Q	H	N	Q	N	C	M	G	K	R	R	R	A
Utah1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SL3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
United	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LN1	.	.	.	.	.	.	.	.	.	.	R	.	C	.	.	.
LN2	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
LN3	.	G	Y	.	.	.	.	.	.	L	.	.	C	L	.	T
K	.	H	.	*	.	Y	*	Y	.	.	.	.	C	.	.	T
LU1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LU2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LU3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LU4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LU5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CU1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CU2	.	N	.	.	.	.	.	.	.	.	.	.	H	.	.	.
CU3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CU4	.	.	.	.	.	.	.	.	./Y	.	.	.	.	.	.	.
KI1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KI2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CO3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CO4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CU5	.	.	.	.	.	.	K/.	.	.	.	.	.	.	.	.	.
CU6	.	.	.	.	.	.	K	.	.	.	.	.	.	.	.	.
HA1	.	H	.	.	.	Y	*	.	.	.	.	.	.	L	.	T
HA2	.	H	.	.	.	Y	*	.	.	.	.	.	.	L	.	T
CO1	.	H	.	.	.	.	*	.	.	.	.	.	.	.	.	T
CO2	.	H	.	.	.	.	*	.	Y	.	.	.	.	.	.	T
PI1	.	G	.	.	.	Y	*	.	.	.	.	.	.	.	.	T
CU7	.	S	.	.	.	Y	*	.	.	.	.	.	.	.	.	T
YA1	.	Y	Y	*	R	Y	*	.	Y	.	.	.	C	L	.	T
YA2	.	Y	Y	.	R	Y	*	.	Y	.	.	.	C	L	.	T
YA3	R	H	.	.	.	.	.	.	.	.	.	.	C	L	.	T
YA4	.	N	.	.	.	.	.	.	.	.	.	.	.	L	.	T

Note: See footnotes of the Table A3.
An asterisk represents the STOP codon.

Table A6. Variable amino acid positions in the unique region of the VP1 protein of the *Amdoparvoviruses*.

G	3	27		35	36	37	38	39	40
G	K	K	-	Y	H	G	E	D	T
Utah1	R	.	-	.	.	.	.	.	.
SL3	.	.	-	.	.	.	.	.	.
LN1	R	.	-	.	.	.	.	.	.
LN2	.	.	-	.	.	.	.	.	.
LN3	.	.	-	.	.	.	.	.	.
LU1	.	.	-	.	.	.	P	.	.
LU2	.	.	-	.	.	.	P	.	.
LU3	.	.	-	.	.	.	P	.	.
LU4	.	.	-	.	.	.	P	.	.
LU5	.	.	-	.	.	.	P	.	.
CU1	R	.	-	.	.	.	.	.	.
CU2	.	.	-	.	.	.	.	.	.
CU3	R	.	-	.	.	.	.	.	.
CU4	R	.	-	.	.	.	.	.	.
KI1	.	.	-	.	.	.	P	.	.
KI2	.	.	S	S	Y	E	.	H	.
CO3	R	.	-	.	.	.	E/D	.	I/T
CO4	R	R	-	.	.	.	.	.	.
CU5	.	R	V	.	.	.	.	.	.
CU6	.	R	V	.	.	.	.	.	.
HA1	R	.	S	S	Y	A	.	.	.
HA2	.	.	S	S	Y	A	.	.	.
CO1	.	.	S	S	Y	A	.	.	.
CO2	R	.	S	S	Y	E	.	.	.
PI1	.	.	S	S	Y	A	.	.	.
CU7	.	.	-	.	.	.	.	.	.
YA1	.	.	S	S	Y	E	.	.	.
YA2	.	.	S	S	Y	E	.	.	.
YA3	.	.	S	S	Y	E	.	.	.
YA4	R	.	-	S	Y	E	.	.	.

Note: See footnotes of the Table A3.

Table A7. Variable amino acid positions in the overlapped region of the VP1 and VP2 proteins of the *Amdoparvoviruses*.

G	47	49	52	54	58	65	69	71	75	83	96	118	132	133	134	135
G	T	A	M	T	T	G	G	G	G	N	V	T	T	K	T	H
Utah1	A	P	.	.	.	.	.	S	.	.	.	.	.	.	.	A
SL3	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.	.
LN1	.	P	V	.	.	.	.	-	-	.	.	.	.	Q	.	A
LN2	.	P	.	.	.	.	.	.	.	.	.	.	.	.	.	A
LN3	A	P	.	.	.	.	.	.	.	.	.	.	.	.	.	P
LU1	.	.	.	.	.	.	.	.	.	.	.	I/T	.	T	.	.
LU2	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
LU3	.	.	.	.	.	.	.	.	.	.	I	.	.	T	I/T	.
LU4	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
LU5	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
CU1	.	P	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CU2	.	P	.	I	.	.	.	.	-	K	.	.	.	.	.	.
CU3	.	P	.	.	.	.	.	.	.	.	.	.	.	Q	K	.
CU4	A	P	.	.	.	.	.	-	-	.	.	.	.	.	.	.
KI1	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KI2	A	P	.	.	.	.	.	.	.	.	.	.	.	.	.	S
CO3	A	P	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CO4	.	P	.	.	.	.	.	-	-	.	.	.	.	.	.	.
CU5	A/T	P	.	.	.	.	.	.	.	.	.	.	.	.	.	Q
CU6	.	P	.	.	.	.	.	.	.	.	.	.	.	.	.	Q
HA1	A	P	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HA2	A	P	.	.	.	.	.	.	.	.	.	.	.	.	.	A
CO1	A	P	.	.	.	.	.	.	.	.	.	.	.	.	.	S
CO2	A	P	.	.	.	.	-	-	-	.	.	.	.	.	.	A
PI1	A	P	.	.	.	.	S	.	.	K	.	.	.	.	.	S
CU7	A	P	.	.	.	.	.	.	.	.	.	.	.	.	K	A
YA1	A	P	.	.	.	.	.	.	.	.	.	.	.	.	.	.
YA2	A	P	.	.	.	.	.	.	.	.	.	.	.	.	.	V
YA3	A	P	.	.	.	.	.	-	-	.	.	.	.	.	.	A
YA4	A	P	.	.	.	.	.	-	-	.	.	.	K	T	.	.
FIN5/C8	A	.	.	.	.	.	.	.	S	.	.	.	.	.	.	.
Utah1 Kit	A	P	.	.	.	.	.	S	.	.	.	.	.	.	.	A
Bel1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bel2	.	P	.	.	.	.	.	.	.	.	.	.	.	.	.	A
Rus09	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Rus19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Rus11	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.
Rus14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Rus17	A	P	.	.	S	S	.	.	.	.	.	.	.	.	.	T
Far East	.	P	.	.	.	.	.	S	.	.	.	.	.	.	.	A

Note: See footnotes of the Table A3.

Table A7. Continued

G	136	137	138	140	153	154	158	159	161	170	180	183	185	187	188	205
G	Q	Q	K	N	V	M	Y	I	D	S	L	E	S	V	T	N
Utah1	.	K	.	.	.	.	F	.	.	.	.	.	.	.	.	.
SL3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LN1	.	T	.	.	.	.	F	L	.	.	.	.	.	.	.	.
LN2	.	K	.	D	.	.	F	L	.	.	.	.	.	.	.	.
LN3	.	S	.	.	.	.	F	.	.	.	.	.	.	.	.	.
LU1	.	.	.	.	.	.	.	L	.	C/S	.	.	.	.	.	.
LU2	.	.	.	.	.	.	.	L	.	.	.	.	.	.	.	.
LU3	.	.	.	.	.	.	.	L	.	.	.	.	.	.	.	.
LU4	.	.	.	.	.	.	.	L	.	C/S	.	.	.	.	.	.
LU5	.	.	.	.	.	.	.	L	D/V	.	.	.	.	.	.	.
CU1	.	K	.	D	.	.	F	M	.	.	.	.	.	.	S	.
CU2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.
CU3	.	.	.	.	.	.	.	L	.	.	.	.	.	.	.	.
CU4	.	.	.	.	.	.	.	L	.	.	.	.	.	.	S	.
KI1	.	.	.	.	.	.	.	L	.	.	.	.	.	.	.	.
KI2	.	S	.	.	.	.	.	L	.	.	.	.	.	I	.	.
CO3	.	.	.	.	.	.	.	.	.	.	L/M	.	.	I/V	S	N/K
CO4	.	.	.	.	.	.	F	L	.	.	.	.	.	.	.	.
CU5	.	T	.	.	I/V	.	F	L	.	.	.	.	.	.	I/T	.
CU6	.	T	.	.	I	.	F	L	.	.	.	.	.	.	.	.
HA1	T	.	.	.	.	.	F	L	.	.	.	.	.	.	.	.
HA2	.	T	.	.	.	.	F	M	.	.	.	.	.	.	.	.
CO1	.	G	.	.	.	.	F	L	.	.	.	.	.	.	.	Q
CO2	.	T	.	.	.	.	F	M	.	.	.	.	.	.	.	.
PI1	.	S	Q	.	.	L	.	L	.	.	.	.	.	.	.	.
CU7	.	K	Q	.	.	.	F	L	.	.	.	.	.	.	.	.
YA1	.	.	.	.	.	.	.	L	.	.	.	.	.	.	.	.
YA2	.	S	.	.	.	.	F	M	.	.	.	.	.	.	.	.
YA3	.	S	.	.	.	.	F	M	.	.	.	.	.	.	.	.
YA4	.	T	.	.	.	.	F	L	.	.	.	.	.	.	.	.
FIN5/C8	.	.	.	K	.	.	.	L	.	.	.	D	N	.	S	.
Utah1Kit	.	K	.	.	.	.	F	.	.	.	.	.	.	.	.	.
Bel1	.	.	.	.	.	.	.	L	.	.	.	.	.	.	.	.
Bel2	.	K	.	.	.	.	F	M	.	.	.	.	.	.	.	.
Rus09	.	.	.	.	.	.	.	L	.	.	.	.	.	.	.	.
Rus19	.	.	.	.	I	.	.	L	.	.	.	.	.	.	.	.
Rus11	R	.	.	.	I	.	.	L	.	.	.	.	.	.	.	.
Rus14	.	.	.	.	I	.	.	L	.	.	.	.	.	L	.	.
Rus17	.	K	.	.	.	.	F	.	.	.	.	.	.	.	S	Q
Far East	.	K	.	.	.	.	F	L	.	.	.	D	.	I	S	.

Table A7. Continued

G	210	211	214	226	232	247	275	276	277	278	279	280	281	283
G	S	T	F	L	L	V	V	A	T	E	T	L	T	D
Utah1	.	.	.	.	.	.	M	G	Q	.	Q	.	E	T
SL3	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LN1	.	.	Y	.	.	.	S	G	Q	S	Q	.	E	T
LN2	.	.	.	.	.	I	T	G	Q	Q	Q	.	E	T
LN3	T	V	Y	.	.	.	S	G	Q	S	Q	.	E	T
LU1	.	.	.	.	.	.	.	.	Q	.	.	.	.	.
LU2	.	.	.	.	.	.	.	P	Q	.	.	.	.	.
LU3	.	.	.	.	.	.	.	.	S	.	.	.	.	.
LU4	.	.	.	.	.	.	.	T/A/S	?	.	.	.	.	.
LU5	.	.	.	.	.	I/V	.	.	S	.	.	.	.	.
CU1	.	.	.	.	.	.	A	.	.	.	.	.	.	.
CU2	.	.	.	.	.	.	.	.	Q	.	.	.	.	.
CU3	.	.	.	.	.	.	.	.	.	E/Q	.	.	.	E/D
CU4	.	.	.	V	.	.	.	.	.	.	.	.	M/T	.
KI1	.	.	.	.	.	.	.	.	S	.	.	.	.	.
KI2	A/T	V	Y	.	.	.	A	G	Q	G	Q	.	E	N
CO3	.	.	.	.	.	.	.	S	S	.	.	.	.	.
CO4	.	.	.	.	.	.	S	.	S	.	.	.	.	.
CU5	.	.	.	.	.	.	S	.	.	.	.	.	T/S	N
CU6	.	.	.	.	.	.	S	.	.	.	.	.	.	N
HA1	.	.	.	.	.	.	A	.	S	.	.	.	.	.
HA2	.	.	Y	.	I/M	.	-	.	Q	G	S	Q	D	T
CO1	.	.	Y	.	.	.	-	.	K	.	S/G	T/Q	D	I
CO2	.	.	.	.	M	.	S	G	Q	-	Q	.	E	T
PI1	.	.	.	.	.	.	I/V	.	Q	S	Q	.	E	T
CU7	.	.	.	.	.	.	S	G	Q	.	Q	.	E	T
YA1	.	.	.	.	.	.	.	.	S	.	.	.	.	.
YA2	.	V	Y	.	.	.	-	R	.	.	Q	T	D	T
YA3	.	.	.	.	.	.	A	.	Q	S	Q	.	E	T
YA4	.	.	.	I	.	.	.	.	S	E/Q	.	.	T/D	.
FIN5/C8	.	.	.	.	.	.	.	S	.	.	.	.	.	.
Utah1Kit	.	.	.	.	.	.	M	G	Q	.	Q	.	E	T
Bel1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bel2	T	V	Y	.	.	.	M	G	Q	.	Q	.	E	T
Rus09	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Rus19	.	.	.	.	.	.	.	.	.	.	.	.	I	.
Rus11	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Rus14	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Rus17	M	.	.	.	.	.	T	G	-	.	Q	.	E	T
Far East	.	.	.	.	.	.	T	G	Q	.	-	.	E	T

Table A7. Continued

G	284	285	291	302	309	319	321	325	333	348	350	351	353	355	356	357
G	A	V	S	I	N	T	L	N	L	K	K	T	T	H	K	V
Utah1	G	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SL3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LN1	G	T	.	.	.	.	.	K	.	.	.	.	V	.	.	.
LN2	G	T	.	V	.	.	.	K	.	.	.	.	V	.	.	.
LN3	G	T	.	.	.	.	.	.	.	.	.	.	.	.	R	.
LU1	.	.	.	.	.	.	.	.	.	.	.	.	T/A	.	.	.
LU2	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.
LU3	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
LU4	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
LU5	.	.	.	.	.	.	.	.	.	.	.	.	V	.	.	.
CU1	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
CU2	.	.	.	.	.	.	V	.	.	.	.	.	.	.	.	.
CU3	.	.	.	.	.	.	.	Q	.	.	.	.	Q	.	R	.
CU4	.	.	.	.	.	.	.	Q	.	.	.	.	.	.	.	.
KI1	.	.	.	.	S	.	.	.	.	.	.	.	A	.	.	.
KI2	.	T	.	.	.	S	I	K	.	.	.	.	V	Y	.	.
CO3	.	.	T	.	.	.	.	K	.	.	.	.	A	.	.	.
CO4	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
CU5	.	I/V	.	.	.	.	.	.	.	.	.	.	A	.	.	.
CU6	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
HA1	.	.	.	.	.	.	.	K	.	T	.	Y/L	K	.	.	.
HA2	G	I	.	.	.	.	.	K	.	T	.	L	A	.	.	.
CO1	.	I	.	.	.	.	.	K	.	.	R/K	.	R/K	.	.	.
CO2	G	T	.	.	.	.	.	K	.	.	.	S	V	.	.	.
PI1	.	I	.	.	.	S	.	K	.	.	.	S	K	.	.	.
CU7	G	T	.	V	.	.	.	K	.	.	.	.	S	.	.	.
YA1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
YA2	G	T	.	.	.	.	.	K	.	.	.	.	?	.	.	.
YA3	G	T	.	.	.	.	I	K	.	.	.	.	T/A	.	.	.
YA4	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
FIN5/C8	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Utah1Kit	G	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bel1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bel2	G	T	.	.	.	.	.	K	.	.	.	.	.	.	.	.
Rus09	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Rus19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Rus11	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Rus14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Rus17	G	T	.	.	.	.	.	K	.	.	.	.	A	.	.	.
Far East	G	T	.	.	.	.	.	K	F	.	.	.	K	.	.	A

Table A7. Continued

G	359	360	361	364	365	366	367	368	369	370	371	372	373	374	375	380
G	S	K	E	A	D	L	I	Y	I	Q	G	Q	D	N	T	F
Utah1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SL3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LN1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LN2	.	L	Q	.	.	.	.	.	.	E	.	.	.	.	.	.
LN3	.	.	.	.	.	.	.	.	.	E	.	.	.	.	.	.
LU1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LU2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LU3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LU4	.	.	.	.	.	.	.	.	.	E	.	.	N	.	.	.
LU5	.	.	.	.	.	.	.	.	.	E	.	.	S	.	.	.
CU1	.	L	Q	.	.	.	.	.	.	E	.	.	.	.	.	.
CU2	.	.	.	.	.	.	.	.	.	E	.	.	.	.	.	.
CU3	.	.	.	.	.	.	.	.	.	E	.	.	.	.	.	.
CU4	N	L	D	G	.	.	.	.	.	E	.	.	.	.	.	.
KI1	.	.	.	.	.	I	.	.	.	E	.	.	N	.	.	.
KI2	.	L	.	.	.	I	T	.	.	E	.	.	.	.	.	.
CO3	.	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.
CO4	.	L	N	.	.	I	.	.	.	E	.	.	.	.	.	.
CU5	.	L	N/D	.	.	I/L	.	.	.	E	.	.	N/D	.	.	.
CU6	.	L	N/D	.	.	I	.	.	.	E	.	.	.	.	.	.
HA1	.	L	.	.	.	.	V	.	.	.	.	.	.	.	.	.
HA2	.	L	.	.	.	I	T	.	.	.	.	G	.	H	V	.
CO1	.	L	.	.	.	I/M	V	.	.	E/Q	D/G	A	.	.	.	.
CO2	.	L	.	.	.	.	L	.	.	E	.	.	.	H	.	.
PI1	.	L	.	G	.	I	V	.	.	E	.	.	.	.	.	.
CU7	.	L	.	.	.	.	.	.	.	E	.	.	.	.	.	.
YA1	.	.	.	.	.	.	.	.	.	E	.	.	.	.	.	Y
YA2	N	L	D	.	.	I	T	.	.	E	.	.	.	.	.	.
YA3	.	L	.	.	.	I	.	.	.	E	.	.	.	.	.	.
YA4	.	L	.	.	.	I	.	.	.	E	.	.	N	.	.	.
FIN5/C8	.	.	.	.	E	.	.	.	.	.	.	.	.	.	.	.
Utah1Kit	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bel1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Bel2	.	L	.	.	.	I	V	.	.	E	.	.	.	.	.	.
Rus09	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Rus19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Rus11	.	.	.	.	.	.	.	.	.	.	.	.	N	.	.	.
Rus14	.	.	D	.	.	.	.	.	.	.	.	.	.	.	.	.
Rus17	.	L	.	.	E	.	.	.	.	.	.	.	.	.	.	.
Far East	.	.	.	.	.	.	.	D	V	.	R	.	R	.	.	.

Table A7. Continued

G	384	387	395	402	404	408	414	419	422	434	437	438	440	449	453	457
G	E	K	I	Y	Y	I	Q	A	P	T	Q	H	Q	Y	H	N
Utah1	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
SL3	.	.	.	.	.	.	.	.	.	.	.	Q	.	.	.	.
LN1	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
LN2	.	.	V	F	.	.	.	.	.	.	.	Q	.	.	.	.
LN3	.	.	V	.	.	.	.	.	.	.	K	Q	.	.	.	.
LU1	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
LU2	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
LU3	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
LU4	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
LU5	.	.	V	.	.	.	.	.	.	S	.	Q	.	.	.	.
CU1	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
CU2	.	.	V	.	.	.	.	.	.	.	.	Q	.	F	.	.
CU3	.	.	V	F	.	.	.	.	.	.	.	Q	.	.	.	.
CU4	.	.	V	F	.	.	.	.	.	S	.	Q	.	.	.	.
KI1	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
KI2	.	.	V	.	.	.	.	.	.	S	H	T	.	.	.	.
CO3	.	.	V	.	F	V	.	.	.	.	.	Q	.	.	.	.
CO4	.	.	V	.	F	V	.	.	.	.	.	Q	.	.	.	.
CU5	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
CU6	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
HA1	.	R	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
HA2	.	.	V	M	.	.	.	.	.	.	M	T/M	.	.	.	.
CO1	.	.	V	F	.	.	.	.	.	.	M	Q	.	.	.	.
CO2	.	K/R	V	F	.	.	.	.	.	.	.	Q	.	.	.	.
PI1	.	.	V	.	.	.	.	.	.	.	M	Q	.	.	.	.
CU7	.	.	V	M	.	.	.	.	.	.	.	A	.	.	.	.
YA1	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
YA2	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
YA3	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
YA4	.	.	V	.	.	.	.	T	.	.	.	Q	.	.	.	.
FIN5/C8	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
Utah1Kit	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
Bel1	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
Bel2	.	.	V	M	.	.	.	.	.	.	.	Q	.	.	.	.
Rus09	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
Rus19	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
Rus11	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
Rus14	.	.	V	.	.	.	.	.	.	.	.	Q	.	.	.	.
Rus17	.	.	V	M	.	.	.	.	.	.	.	Q	.	.	.	.
Far East	Q	.	.	N	.	.	K	.	Q	.	.	N	K	.	N	K

Table A7. Continued

G	458	460	461	462	464	465	471	472	476	477	478	480	490	491	494	495
G	E	D	L	L	G	I	S	N	D	N	E	E	R	T	I	H
Utah1	.	.	.	.	.	.	.	.	N	H	.	.	.	S	.	.
SL3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LN1	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	.
LN2	.	.	.	.	.	.	.	.	.	H	.	.	.	V	.	.
LN3	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	.
LU1	.	.	.	.	.	.	.	.	.	H	.	.	.	S	.	.
LU2	.	.	Q	.	.	.	.	.	.	H	.	.	.	S	.	.
LU3	.	.	Q	I	.	.	.	.	.	H	.	.	.	S	.	.
LU4	.	.	Q	.	.	.	.	.	.	H	.	.	.	S	.	.
LU5	.	.	.	.	.	.	.	.	.	H	.	.	.	S	.	.
CU1	.	.	.	.	.	.	.	.	.	H	.	.	.	S	.	Q
CU2	.	.	L/Q	.	.	.	.	.	.	H	.	.	.	S	.	.
CU3	.	.	.	.	.	.	.	.	.	H	.	.	.	S	.	.
CU4	.	.	.	.	.	.	.	.	.	H	.	.	.	T/S	I/V	.
KI1	.	.	V	.	.	.	.	.	.	H	.	.	.	S	.	.
KI2	.	.	.	.	.	.	N	.	.	H	.	.	H	S	.	.
CO3	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	.
CO4	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	.
CU5	.	.	.	.	.	.	.	T	.	H	.	.	.	T/S	.	.
CU6	.	.	.	.	.	.	.	T	.	H	.	.	.	.	.	.
HA1	.	.	.	.	A	V	.	.	.	H	.	.	.	S	.	.
HA2	.	.	T	.	A	.	.	.	.	H	.	.	.	S	V	.
CO1	.	.	T	.	A	.	.	.	.	H	.	.	.	S	.	.
CO2	.	.	.	.	.	.	.	.	.	H	.	.	.	S	.	.
PI1	.	.	.	I	A	.	.	.	N	H	D	.	.	S	.	.
CU7	.	.	.	.	.	.	.	.	.	H	.	.	H	S	.	.
YA1	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	.
YA2	.	.	.	.	.	.	.	.	.	H	.	.	H	S	.	.
YA3	.	.	.	.	.	.	.	.	.	H	.	.	H/R	T/S	.	.
YA4	.	.	.	.	.	.	.	.	.	H	.	.	H	S	.	.
FIN5/C8	.	.	.	.	.	.	.	.	E	H	.	.	.	.	.	.
Utah1Kit	.	.	.	.	.	.	.	.	N	H	.	.	.	S	.	.
Bel1	.	.	.	.	.	.	.	.	.	H	.	K	.	.	.	.
Bel2	Q	.	.	.	E	.	.	.	.	H	.	.	H	S	.	.
Rus09	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	.
Rus19	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	.
Rus11	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	.
Rus14	.	.	.	.	.	.	.	.	.	H	.	.	.	.	.	.
Rus17	.	.	.	.	.	.	.	.	N	H	.	.	.	.	.	.
Far East	.	N	.	.	.	.	.	K	.	.	.	.	.	.	.	.

Table A7. Continued

G	497	511	527	528	529	531	532	534	538	543	544	549	558	559	571	577
G	S	S	F	S	G	Q	E	N	E	A	A	L	A	Q	I	H
Utah1	.	.	.	G	.	.	.	E	.	.	.	.	.	L	.	D
SL3	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.
LN1	.	L	.	.	.	.	.	E	.	.	.	Q	.	.	.	D
LN2	.	.	.	G	.	.	T	E	.	.	.	.	.	.	.	D
LN3	.	.	.	.	.	L	.	E	.	.	.	.	.	.	.	D
LU1	.	.	.	.	.	.	.	E	.	.	.	.	.	.	.	D
LU2	.	.	.	.	.	.	.	D	.	.	.	.	.	.	.	D
LU3	.	.	.	.	.	.	.	E	.	.	.	.	.	.	.	D
LU4	.	.	.	.	.	.	.	E	.	.	.	.	.	.	.	D
LU5	.	.	.	.	.	.	.	E	.	G	.	.	.	.	.	D
CU1	.	.	.	S/G	.	.	.	D	.	.	.	.	.	.	.	D
CU2	H	.	.	.	.	.	.	E	.	.	.	.	.	.	.	D
CU3	.	.	.	.	.	.	.	E/D	.	G	.	.	.	.	.	D
CU4	.	.	.	.	.	.	.	D	.	G	.	.	.	.	.	D
KI1	.	.	.	.	.	.	.	E	.	.	.	.	.	.	.	D
KI2	H	.	.	.	.	.	V	E	.	.	.	.	G	L	.	D
CO3	.	.	.	.	.	.	.	D	Q	G	.	.	.	.	.	D
CO4	H	.	.	.	.	.	.	D	.	G	.	.	.	.	.	D
CU5	.	.	.	.	.	.	.	D	.	.	.	.	.	.	.	D
CU6	.	.	.	.	.	.	.	D	.	.	.	.	.	.	.	D
HA1	.	.	.	.	.	.	.	D	.	G	.	.	G	L	V	D
HA2	.	.	.	.	.	.	.	E	Q	G	.	.	.	L	.	D
CO1	.	.	.	.	D/G	.	D	E	Q	G	.	.	.	L	V	D
CO2	.	.	.	.	.	.	.	E	.	.	.	.	.	.	.	D
PI1	.	.	.	.	.	H	D	E	.	G	L	.	.	L	.	D
CU7	.	.	.	.	.	.	V	E	.	S	.	.	.	L	.	D
YA1	.	.	.	.	.	.	.	E	.	.	.	.	.	.	.	D
YA2	.	.	.	.	.	.	.	E	.	G	.	.	.	L	.	D
YA3	H	.	.	T	.	.	D/E	E	.	G	.	.	.	.	.	D
YA4	.	.	.	.	S	.	.	D	.	.	.	.	.	.	.	D
FIN5/C8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	D
Utah1Kit	.	.	.	G	.	.	.	E	.	.	.	.	.	L	.	D
Bel1	.	.	.	.	.	.	.	D	.	.	.	.	.	.	.	D
Bel2	.	.	.	.	.	.	.	E	.	.	.	.	.	L	.	D
Rus09	.	.	.	.	.	.	.	D	.	.	.	.	.	.	.	D
Rus19	.	.	.	.	.	.	.	D	.	.	.	.	.	.	.	D
Rus11	.	.	.	.	.	.	.	D	.	.	.	.	.	.	.	D
Rus14	.	.	.	.	.	.	.	D	.	.	.	.	.	.	.	D
Rus17	.	.	.	G	.	.	.	E	.	.	.	.	.	L	.	D
Far East	.	.	L	.	.	T	.	E	.	.	.	.	.	L	.	D

Table A7. Continued

G	578	581	585	587	589	591	611	617	618	619	634	635	646	658	660	661
G	M	K	N	N	P	V	T	N	P	D	V	L	N	V	T	K
Utah1	.	.	.	S	.	.	.	.	.	.	.	.	.	.	S	.
SL3	.	.	.	.	.	.	.	.	.	.	.	.	.	L	.	.
LN1	.	.	D	.	.	.	.	.	.	.	.	.	.	.	S	.
LN2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LN3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LU1	.	.	.	.	.	.	.	.	S	E	I	.	.	.	S	.
LU2	.	.	.	.	.	.	.	.	S	E	I	.	.	.	S	.
LU3	.	.	.	.	.	.	.	.	.	.	I	.	.	.	N	.
LU4	.	.	.	.	.	.	.	.	.	.	I	.	.	.	S	.
LU5	.	.	.	.	.	.	.	.	.	.	I	.	.	I/V	.	.
CU1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.
CU2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.
CU3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.
CU4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.
KI1	.	.	.	.	.	.	.	.	.	.	I	.	.	.	.	.
KI2	V	.	.	S	.	.	.	.	.	.	.	.	.	.	.	.
CO3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
CO4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	S
CU5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.
CU6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.
HA1	V	.	.	S	.	I	.	.	.	.	.	.	.	.	S	.
HA2	.	.	.	S	.	.	.	.	.	.	.	.	.	.	S	R
CO1	V	R	.	S	.	I	.	D	.	.	.	.	.	I/V	S	.
CO2	.	.	.	.	.	.	I	.	.	.	.	.	.	.	.	.
PI1	V	.	.	S	.	.	.	D	.	.	.	.	.	.	S	.
CU7	.	.	.	S	.	.	.	.	.	.	.	.	.	.	S	.
YA1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
YA2	V	.	.	S	.	.	.	.	.	.	.	.	.	.	S	.
YA3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
YA4	.	.	.	S	.	.	.	.	.	.	.	.	.	.	.	.
FIN5/C8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Utah1Kit	.	.	.	S	.	.	.	.	.	.	.	.	.	.	S	.
Bel1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.
Bel2	.	.	.	S	.	.	.	.	.	.	.	.	.	.	G	R
Rus09	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.
Rus19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.
Rus11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.
Rus14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	S	.
Rus17	.	.	.	S	.	.	.	D	.	.	.	.	.	.	N	T
Far East	.	.	.	S	S	.	.	.	.	.	.	I	S	.	.	.

Table A7. Continued

G	662	663	664	670	672	675	677	689
G	D	K	Y	N	K	F	I	I
Utah1	.	N	.	.	.	.	.	.
SL3	.	.	.	H	.	.	.	.
LN1	.	N	.	.	.	.	.	L
LN2	.	N	.	.	.	.	.	.
LN3	.	N	.	.	.	.	.	.
LU1	.	N	.	.	.	.	.	.
LU2	N	N	.	.	.	.	.	.
LU3	.	N	.	.	.	.	.	.
LU4	.	N	.	.	.	.	M	.
LU5	.	N	.	.	.	.	.	.
CU1	.	N	.	.	.	.	.	.
CU2	.	N	.	.	.	.	.	.
CU3	.	N	.	.	.	.	.	.
CU4	.	N	Y/F	.	.	.	.	.
KI1	.	N	.	.	.	.	.	.
KI2	.	N	.	.	.	.	.	L
CO3	.	N	.	.	.	.	.	.
CO4	.	N	.	.	.	.	.	.
CU5	.	N	.	.	.	.	.	.
CU6	.	N	.	.	.	.	.	.
HA1	.	N	.	.	.	.	.	L
HA2	.	N	.	.	.	.	.	.
CO1	.	N	.	.	.	.	.	.
CO2	.	N	.	.	R/K	.	.	.
PI1	.	N	.	.	.	.	.	L
CU7	.	N	.	.	.	.	.	.
YA1	.	N	.	.	.	.	.	.
YA2	.	N	.	.	.	.	.	L
YA3	.	N	.	.	.	.	.	.
YA4	.	N	.	.	.	.	.	.
FIN5/C8	.	N	.	.	.	.	.	.
Utah1 Kit	.	N	.	.	.	.	.	.
Bel1	.	N	.	.	.	.	.	.
Bel2	E	N	.	.	.	.	.	.
Rus09	.	N	.	.	.	.	.	.
Rus19	.	N	.	.	.	.	.	.
Rus11	.	N	.	.	.	.	.	.
Rus14	.	N	.	.	.	.	.	.
Rus17	N	N	.	.	.	.	.	.
Far East	.	N	.	.	.	Y	.	.

Table A8. Classification of codons containing ambiguous codes in the proteins of the 25 local isolates

Genome region[£]	No. of amino acids				
	Translated	Non-translatable	Mixed[¥]	Synonymo. Substitutions	Combination[†]
NS overlapped	27	0	26	1	23 (88.5%)
NS1 unique	69	2	65	4	49 (75.4%)
NS2 unique	8	0	6	2	4 (66.7%)
NS3 unique	5	0	2	3	2 (100%)
Total	109	2	99	10	78 (78.8%)
VP1 unique	3	0	2	1	0 (0.0%)
VP1/VP2	87	1	52	33	31 (59.6%)
Total	90	1	54	34	31 (57.4%)

[£] Overlapped region: aa 1-60; NS1 unique sequence: aa 61-641; NS2 unique sequence: aa 61-114; NS3 unique sequence: aa 61-87; VP1 unique region: aa 1 to 43 and VP1/VP2 overlapped region: aa 44/1 to 690/647.

[¥] Positions containing two or three putative amino acids.

[†] Positions containing two or three putative aas where individual aas were present in other AMDV isolates at the same position. Percentages of total mixed amino acids are shown in brackets.

Table A9. Number and percentage of total amino acid substitutions in the non-structural and structural proteins of *Amdoparvoviruses*

	Non-structural proteins			Structural proteins	
Isolates[£]	NS1	NS2	NS3	VP1	VP2
CO1	123 (19.2%)	20 (17.5%)	15 (17.2%)	51 (7.4%)	47 (7.3%)
CO2	111(17.3%)	20 (17.5%)	17 (19.5%)	38 (5.5%)	33 (5.1%)
CO3	60 (9.4%)	7 (6.1%)	5 (5.7%)	25 (3.6%)	23 (3.6%)
CO4	75 (11.7%)	12 (10.5%)	10 (11.5%)	24 (3.5%)	22 (3.4%)
CU1	67 (10.5%)	11 (9.6%)	10 (11.5%)	22 (3.2%)	21 (3.2%)
CU2	57 (8.9%)	12 (10.5%)	13 (14.9%)	18 (2.6%)	18 (2.8%)
CU3	62 (9.7%)	9 (7.9%)	7 (8.0%)	20 (2.9%)	19 (2.9%)
CU4	61 (9.5%)	10 (8.8%)	8 (9.2%)	25 (3.6%)	24 (3.7%)
CU5	98 (15.3%)	15 (13.2%)	12 (13.8%)	28 (4.1%)	26 (4.0%)
CU6	101 (15.8%)	17 (14.9%)	14 (16.1%)	20 (2.9%)	18 (2.8%)
CU7	120 (18.7%)	19 (16.7%)	15 (17.2%)	34 (4.9%)	34 (5.3%)
HA1	122 (19.0%)	20 (17.5%)	15 (17.2%)	43 (6.2%)	39 (6.0%)
HA2	115 (17.9%)	15 (13.2%)	12 (13.8%)	43 (6.2%)	39 (6.0%)
KI1	68 (10.6%)	12 (10.5%)	6 (6.9%)	18 (2.6%)	17 (1.6%)
KI2	64 (10.0%)	14 (12.3%)	6 (6.9%)	49 (7.1%)	44 (6.8%)
LU1	70 (10.9%)	10 (8.8%)	6 (6.9%)	18 (2.6%)	17 (2.6%)
LU2	56 (8.7%)	13 (11.4%)	9 (10.3%)	19 (2.8%)	18 (2.8%)
LU3	67 (10.5%)	13 (11.4%)	7 (8.0%)	18 (2.6%)	17 (2.6%)
LU4	58 (9.0%)	10 (8.8%)	8 (9.2%)	20 (2.9%)	19 (2.9%)
LU5	53 (8.3%)	3 (2.6%)	0 (0.0%)	18 (2.6%)	17 (2.6%)
PI1	123 (19.2%)	26 (22.8%)	19 (21.8%)	51 (7.4%)	47 (7.3%)
YA1	132 (20.6%)	25 (21.9%)	23 (26.4%)	16 (2.3%)	12 (1.9%)
YA2	126 (19.7%)	23 (20.2%)	21 (24.1%)	42 (6.1%)	38 (5.9%)
YA3	80 (12.5%)	14 (12.3%)	10 (11.5%)	35 (5.1%)	31 (4.8%)
YA4	73 (11.4%)	19 (16.7%)	11 (12.6%)	31 (4.5%)	27 (4.2%)
Total[¥]	85.7 (13.4%)	14.8 (12.9%)	11.2 (12.8%)	29.0 (4.2%)	26.7 (4.1%)
K	102 (15.9%)	20 (17.5%)	17 (19.5%)	n/a	n/a
BEL1	n/a[†]	n/a	n/a	n/a	9 (1.4%)
BEL2	n/a	n/a	n/a	n/a	37 (5.7%)
Far East	n/a	n/a	n/a	n/a	44 (6.8%)
FIN5/C8	n/a	n/a	n/a	n/a	16 (2.5%)
LN1	87 (13.6%)	14 (12.3%)	8 (9.2%)	30 (4.3)	30 (4.6%)
LN2	68 (10.6%)	21 (18.4%)	10 (11.5%)	32 (4.6)	32 (4.9%)
LN3	79 (12.3%)	14 (12.3%)	11 (12.6%)	27 (3.9)	27 (4.2%)
RUS09	n/a	n/a	n/a	n/a	8 (1.2%)
RUS11	n/a	n/a	n/a	n/a	13 (2.0%)
RUS14	n/a	n/a	n/a	n/a	11 (1.7%)
RUS17	n/a	n/a	n/a	n/a	36 (5.6%)
RUS19	n/a	n/a	n/a	n/a	9 (1.4%)
SL3	5 (0.8%)	0 (0.0%)	0 (0.0%)	5 (0.7)	5 (0.8%)
Utah1	11 (1.7%)	4 (3.5%)	1 (1.1%)	27 (3.9)	26 (4.0%)
Utah1 Kit	n/a	n/a	n/a	n/a	26 (4.0%)
United	54 (8.4%)	14 (12.3%)	6 (6.9%)	n/a	n/a

[£] Local AMDV sequences are shown at the top in **bold**.

[¥] Average of total aa substitution.

[†] n/a: segments of genome which were not available on GenBank.

Table A10a. Unique amino acid substitutions in the non-structural and structural proteins of the 25 local isolates

Isolate	Non-structural proteins¥			Structural proteins¥	
	NS1£	NS2	NS3	VP1u	VP2
CO1	L-33†, A-57†, I-94, Y-159, V-168, V-190, L-246, L-291, G-294, L-314, K-389, G-408, S-415, T-592	H-66	-	-	G-94, K-234, G-236, I-240, R-307, R-310, M-323, D/E-328, A-329, D-486, R-538
CO2	V-12†, T-168, L-225, N-226, K-246, D-310, I-314, S-322, D-390, M-425, R-563	H-62, A-113	-	-	L-324, I-568, R-629
CO3	T-314, D-379, V-408, V-473	H-81	-	D-38, I-40	M-137, K-162, T-248
CO4	H-9†, R-21†, G-24†, D-68, Y-76, T-79, K-408, Q-571, Q/K-590	K-64	-	-	S-618
CU1	K-14†, Q-81, G-137, R-194, N-282, T-385	-	-	-	Q-452
CU2	M-15†, V-475, I-533	-	H-82	-	I-11, V-278, F-406
CU3	D-27†, I-311	H-81	-	-	E-240, Q-310
CU4	H-83, N-176	-	-	-	V-183, M-238, F-621
CU5	N-17†, D-67, S-108, H-109, R-179, H-196, D-225, E-262, V-264, A-288, N-330, Y-368, T-394, E-402, A-606	S-80, T-108	K-72	-	Q-92, I-145, S-238
CU6	See CU5	See CU5	See CU5	-	Q-92
CU7	S-10†, R-14†, N-47†, K-75, G-78, G-79, I-157, D-186, L-207, Q-211, G-235, C-246, D-248, T-267, L-286, S-311, Q-313, C-327, G-360, V-397, N-408, F-410, L-514, L-542, A-563, N-571, I-611	-	S-63	-	S-310, A-395, S-500
HA1	N-26†, V-234, L-310, G-326, V-339, M-373, L-387, S-426, N-570, A-610, S-620	S-84, N-85, T-113	-	-	T-93, Y-308 V,-422
HA2	M-196, A-326, T-572	-	-	-	I-189, G-329, V-332, M-395

£ Unique aas are those that are not detected in other AMDV sequences at the same position in the alignment. In each cell, the type of aa substitution is followed by the aa position in the AMDV-G (GenBank accession number NC_001662).

¥ Refer to footnotes of Table 3.6 for description of the overlapped and unique regions of proteins.

† Unique aa substitutions in the overlapped region of the non-structural proteins.

Table A10a. Continued

Isolate	Non-structural proteins¥			Structural proteins¥	
	NS1	NS2	NS3	VP1	VP2
KI1	D-272, Q-303, P-316, H-326, K-572, K-591, Q-635	I-66, K-110	-	-	T-6, S-266, V-418
KI2	H-174, L-192, N-376, E-408, V-594, H-616	A-90	-	H-39	A-167, Y-312, H-394, N-428
PI1	T-12†, E-13†, Q-21†, Q-59†, S-69, A-73, D-74, E-79, R-82, T-83, V-86, R-91, G-158, T-167, S-176, T-186, F-240, V-287, A-289, T-298, D-408, R-412, V-481, T-542, V-592, V-597, Y-640	R-66, V-114	-	-	S-26, L-111, I-232, D-435, L-501
LU1	H-12†, I-210, A-334, R-567	-	-	-	I-75
LU2	-	-	-	-	P-233
LU3	K-34†, R-144, T-191	-	-	-	I-53, I-91
LU4	-	-	-	-	M-634
LU5	-	-	-	-	V-118, S-330
YA1	I-12†, Q-386, I-409, D-616	R-90	-	-	Y-337
YA2	S-21†, D/A-26, L-623	K-90	-	-	V-92, R-233
YA3	A-21†, S-78, N-173, V-409, R-597	R-72	R-62	-	T-485
YA4	S-83, E-330, Y/C-334, S-339, C-598, S-555	K-73	-	-	K-89, I-183, T-376, S-486

Table A10b. Unique amino acid substitutions in the non-structural and structural proteins of the isolates reported on GenBank

Isolate	Non-structural proteins[¥]			Structural proteins[¥]	
	NS1[£]	NS2	NS3	VP1	VP2
K	W-161, Q-172, D-179, I-183, T-208, Q-262, S-276, M-346, H-409, Y-484, S-512, V-551, I-566	-	Y-74	n/a[‡]	n/a
BEL1	n/a	n/a	n/a	n/a	K-437
BEL2	n/a	n/a	n/a	n/a	Q-415, E-421, E-619
Far East	n/a	n/a	n/a	n/a	F-290, A-314, D-325, V-326, R-328, R-330, Q-341, K-371, Q-379, N-395, K-397, N-410, K-414, N-417, K-429, L-484,T-488, S-546, I-592, S-603, Y-632
FIN5/C8	n/a	n/a	n/a	n/a	S-32, K-97, N-142, E-433
G	-	-	-	-	H-395
LN1	T-176, L-216, L-377, E-394	-	R-79	-	V-9, L-468, Q-506, D-542
LN2	K-13, V-72, S-147, E-348	K-70, T-99, K-102	-	-	V-448, T-489
LN3	I-72, N-227, R-247, C-323, Q-401, R-464, S-568, I-600	-	L-77	-	P-92, K-394
RUS09	n/a	n/a	n/a	n/a	-
RUS11	n/a	n/a	n/a	n/a	R-93
RUS14	n/a	n/a	n/a	n/a	L-144
RUS17	n/a	n/a	n/a	n/a	S-15, S-22, T-92, M-167, T-618
RUS19	n/a	n/a	n/a	n/a	I-238
SL3	D-29[†], T-163	-	-	-	R-485, L-615, H-627
Utah1	Q-76, N-107	-	-	-	-
Utah1 Kit	n/a	n/a	n/a	n/a	-
United	S-211	-	-	n/a	n/a

[£,¥,†] Refer to the footnotes for Table 4.11a for the description of symbols.
[‡] n/a: segments of genome which were not available on GenBank.

Table A11. Amino acid variation in the two caspase recognition sites in the NS1 protein of the *Amdoparvoviruses*

AMDV sequences[¥]	Left aa recognition site[£]					Right aa recognition site[£]				
	224	225	226	227[†]	228	282	283	284	285[†]	286
AMDV-G[*]	I	N	T	D	S	D	Q	T	D	S
Utah1[*], SL3[*]	.	.	.	.	.	.	.	.	.	.
LN1[*]	H	K	.	G	.	.	.	.	.	
LN3[*]	H	K	N	G	.	.	.	.	.	
LN2[*], United[*]	.	H	K	E	G	.	.	.	.	.
K[*], HA2	.	T	Q	E	G	.	.	.	.	.
CU2, KI1, KI2, LU1, LU2, LU3, LU4, LU5, YA3, YA4	.	T	.	I	G	.	.	.	.	.
CO1	.	T	E	E	N	.	.	.	.	.
CO2	.	L	N	.	G	.	.	.	.	.
CO3	.	T	.	.	.	.	.	.	.	.
CO4	.	T	.	.	T	.	.	.	.	.
CU1	.	.	.	.	T	N	.	.	.	.
CU3, CU4	.	I	.	.	.	.	.	.	.	.
CU5, CU6	.	D	K	E	N	.	.	.	.	.
CU7	.	T	.	E	T	.	K	.	.	L
HA1	.	T	E	E	G	.	.	.	.	.
PI1	.	T	K	E	G	.	K	.	.	.
YA1, YA2	.	T	E	E	H	.	.	.	.	.
RFAV[*]	.	.	E	E	T	.	.	I	.	.
GFAV[*]	.	T	E	K	G	T	E	T[‡]	.	.

[£] Positions are based on the AMDV-G sequence (GenBank accession number NC_001662).
[¥] GenBank sequences are indicated by an asterisk.
[†] Caspase cleavage sites are highlighted.
[‡] The GVAV had a seven amino acid insertion at this position: ISNVTYV.

Table A12. Amino acid variations in the caspase recognition site in the VP2 protein of the *Amdoparvoviruses*

AMDV sequences		aa recognition sites[£]				
GenBank sequences	Local isolates	417	418	419	420[¥]	421
AMDV-G, BEL1, FIN05/C8, LN1-3, RUS9/11/14/17/19, SL3, Utah1, Utah1 Kit	CO2-4, CU1, CU3-7, KI2, LU1/5	D	L	L	D	G
BEL2	-	.	.	.	.	E
Far East	-	N	.	.	.	.
-	CO1, HA2	.	T	.	.	A
-	CU2	.	L/Q	.	.	.
-	HA1	.	.	.	.	A
-	LU2/4	.	Q	.	.	.
-	LU3	.	Q	I	.	.
-	KI1	.	V	.	.	.
-	PI1	.	.	I	.	A
RFAV	-	.	T	.	S	A
GFAV	-	.	T	I	.	S

[£] Positions are based on the AMDV-G sequence (GenBank accession number NC_001662).
[¥] Caspase cleavage site is highlighted.

Table A13. Variations at the amino acid determinants of in vitro replication in the VP2 gene of *Amdoparvoviruses*

AMDV sequences		aa positions[£]		
GenBank	Local	92	94	115
AMDV-G, SL3, BEL1, FIN05/C8, Rus09/11/14/19	CO3, CU2-4, KI1, LU1-5, YA1	H	Q	Y
TR	CO4, HA1	.	.	F
Utah1, Far East, Utah1 Kit, BEL2, LN2	CU7	A	K	F
ZK8	YA3	A	S	F
PU	-	A	P	F
Rus17	-	T	K	F
LN1	CO2, HA2	A	T	F
LN3	-	P	S	F
-	CO1	S	G	F
-	CU1	.	K	F
-	CU5/6	Q/K	T	F
-	KI2, PI1	S	S	.
-	YA2	V	S	F
-	YA4	.	T	F
RFAV		.	K/T	F
GFAV		P	G	F

[£] Positions are based on the AMDV-G sequence (GenBank accession number NC_001662).

Table A14. Variations at the amino acid determinants of pathogenicity in VP2 gene of *Amdoparvoviruses*

AMDV sequences		aa position[£]				
GenBank	Local	352	395	434	491	534
AMDV- G	-	I	H	N	N	H
SL3	-	.	Q	.	.	.
TR	-	V	Q	.	D	D
Utah1, BEL2, LN1, LN2, LN3, RUS17, Utah1 Kit, ZK8	CO1, CO2, CU2, LU1, LU3, LU4, LU5, KI1, PI1, YA1, YA2, YA3	V	Q	H	E	D
PU		V	Q	E	E	D
Bel1, Rus09, Rus11, Rus14, Rus19	CO3, CO4, CU4, CU5, CU6, HA1, LU2, YA4	V	Q	H	D	D
FIN05/C8	-	.	N	.	E	D
Far East	-	V	Q	H	.	D
-	CU3	V	Q	H	E/D	D
-	CU7	V	A	H	E	D
-	HA2	V	T/M	H	E	D
-	KI2	V	T	H	E	D
RFAV		V	T	S	D	D
GFAV		V	Q	Q	.	D

[£] Positions are based on the AMDV-G sequence (GenBank accession number NC_001662).

Table A15. Entropy values of the aa positions in the NS1 protein

Position[£]	AMDV-G	Entropy[¥]	Position	AMDV-G	Entropy	Position	AMDV-G	Entropy	Position	AMDV-G	Entropy	Position	AMDV-G	Entropy
6	I	0.971	67	E	0.229	112	L	0.425	174	D	0.136	227	T	1.178
7	D	1.069	68	E	0.136	116	I	0.369	175	P	0.501	228	D	1.197
9	Q	0.363	69	N	0.791	120	K	0.799	176	E	0.305	229	S	1.234
10	R	0.837	71	T	0.746	122	-	0.136	177	D	1.147	230	M	0.229
11	R	0.229	72	A	0.569	123	Q	0.613	178	R	0.501	234	F	0.637
12	L	0.899	73	S	0.136	124	Q	1.334	179	A	0.229	235	A	0.501
13	Q	0.271	74	N	0.681	125	F	0.613	180	K	0.918	236	A	0.712
14	D	1.18	75	E	1.085	137	N	0.369	184	V	0.136	241	I	0.848
15	L	0.363	76	H	1.664	138	E	0.136	186	I	0.67	242	V	0.586
16	Y	0.978	78	T	0.271	140	P	0.437	187	E	0.271	247	N	0.814
17	V	1.566	79	N	0.814	143	V	0.527	190	P	0.305	248	K	0.136
18	Q	0.76	80	N	0.808	144	Q	0.681	191	T	0.136	249	E	0.136
19	L	1.034	81	E	0.76	145	K	0.136	192	K	0.632	252	T	0.902
21	K	0.758	82	I	0.136	148	G	0.136	193	P	0.363	253	L	0.369
23	I	1.005	83	N	0.7	150	F	0.369	195	K	0.136	256	M	0.586
24	N	1.213	85	C	0.425	151	M	1.221	197	F	1.423	263	S	1.058
25	D	0.453	86	K	0.136	153	R	0.777	198	N	0.738	264	D	0.586
26	G	0.271	88	T	0.806	154	L	0.59	199	K	0.305	265	I	0.229
27	E	0.136	91	K	0.136	156	K	0.645	200	Q	0.229	266	M	0.613
29	V	0.917	92	T	0.369	157	D	0.305	203	Q	0.305	268	A	0.79
33	F	0.136	94	L	1.472	158	L	0.136	204	D	0.689	269	N	0.693
34	Q	0.136	95	L	0.644	159	A	0.136	208	P	0.719	271	D	0.474
35	Q	0.363	96	I	0.6	160	V	1.425	209	V	0.437	273	E	0.136
40	D	0.613	98	K	0.437	161	I	0.369	210	H	1.427	275	G	0.991
47	K	0.136	101	K	0.369	162	Y	0.645	211	L	0.921	277	D	0.363
53	R	0.229	104	R	0.271	164	N	0.437	212	R	0.271	283	D	0.136
56	S	0.644	106	D	0.474	165	H	0.229	213	D	0.554	284	Q	0.229
57	S	0.788	107	S	1.165	166	H	0.369	215	T	0.554	287	S	0.136
58	D	0.935	108	N	0.59	168	D	0.686	216	F	1.752	288	A	0.136
59	L	0.136	109	K	0.788	169	I	0.848	217	I	0.495	289	T	0.229
63	F	0.517	110	V	0.369	171	D	0.772	223	D	0.369	290	T	0.777
64	D	0.425	111	N	0.773	173	K	0.645	226	N	1.261	292	T	0.437

Note: Analysis was based on local isolates and GenBank sequences AMDV-G, ADV-K, LN1-3, SL3, United and Utah1 (Group 3).

[£] Positions in the alignment of 33 sequences (Group 3), conducting by MEGA6.

[¥] Entropy values of zero indicate no aa variation at the position and are thus not represented.

Table A15. Continued

Position[£]	AMDV-G	Entropy[¥]	Position	AMDV-G	Entropy	Position	AMDV-G	Entropy	Position	AMDV-G	Entropy
295	S	0.136	376	S	0.305	429	I	1.064	572	R	1.1
297	I	0.474	377	T	0.738	431	F	0.229	573	Q	1.272
299	K	0.437	378	M	0.846	446	I	0.806	576	Q	1.376
303	T	0.902	379	I	0.59	448	K	0.363	578	S	0.67
304	K	0.136	380	T	1.157	449	A	0.474	580	E	0.554
310	E	0.527	382	F	0.974	450	T	0.998	583	E	0.229
311	V	0.271	383	D	0.693	451	V	0.229	589	T	0.554
312	A	1.142	386	I	0.772	453	Y	0.693	591	P	0.899
313	N	1.38	387	K	0.136	455	M	0.229	592	N	0.501
314	P	0.932	388	F	0.136	465	W	0.136	593	S	0.814
315	V	0.932	390	E	0.136	474	I	0.136	595	A	0.136
316	Q	0.474	391	E	0.363	476	A	0.681	597	T	0.369
317	Q	0.136	393	D	0.305	479	C	0.229	598	A	0.136
322	L	1.187	394	D	0.586	482	F	1.17	599	T	0.59
323	Y	1.226	395	K	0.918	485	W	0.136	600	K	0.73
324	S	0.761	398	L	0.136	486	V	0.229	601	N	0.985
327	S	1.619	399	A	0.674	498	V	0.229	604	N	0.229
328	T	0.943	402	K	0.604	527	V	0.437	605	S	0.586
331	A	0.772	403	D	0.644	533	N	0.604	607	P	0.271
333	F	0.689	509	S	0.136	534	A	0.898	611	K	1.286
334	N	0.953	513	C	0.136	536	A	0.799	612	S	1.033
335	V	1.427	515	I	0.363	543	M	0.772	617	N	0.495
339	T	0.363	516	V	0.305	544	I	0.958	619	E	0.84
340	P	1.238	522	I	0.229	549	M	0.229	620	N	0.501
345	I	0.229	409	Q	1.548	551	T	0.916	621	C	0.136
346	K	0.229	410	Y	1.397	552	I	0.136	622	D	0.425
347	Q	0.604	411	L	0.136	555	K	0.655	624	P	0.136
348	S	0.229	413	K	0.437	556	T	0.136	627	S	0.425
349	D	0.136	414	V	0.682	558	I	0.978	633	A	1.08
350	K	1.096	415	L	0.136	564	K	0.271	635	Q	0.229
354	L	0.693	416	C	0.363	565	R	0.939	636	H	0.136
361	P	0.658	423	G	0.369	567	L	0.136	641	H	0.136
365	N	0.425	424	G	0.6	568	N	0.645			
374	K	0.898	426	R	0.437	569	T	0.136			
375	T	0.586	427	G	0.136	571	D	0.136			

Table A16. Entropy values of the aa positions in the VP2 protein

Position[£]	AMDV-G	Entropy[¥]	Position	AMDV-G	Entropy	Position	AMDV-G	Entropy	Position	AMDV-G	Entropy	Position	AMDV-G	Entropy
4	T	0.786	115	Y	0.693	290	L	0.115	379	P	0.115	495	E	0.262
6	A	0.73	116	I	0.875	305	K	0.195	391	T	0.262	500	A	0.688
9	M	0.115	118	D	0.115	307	K	0.115	394	Q	0.487	501	A	0.115
11	T	0.115	127	S	0.195	308	T	0.421	395	H	0.567	506	L	0.115
15	T	0.115	137	L	0.115	310	T	1.609	397	Q	0.115	515	A	0.195
22	G	0.115	140	E	0.195	312	H	0.115	406	Y	0.115	516	Q	0.605
24	G	0.115	142	S	0.115	313	K	0.195	410	H	0.115	528	I	0.195
25	G	0.115	144	V	0.421	314	V	0.115	414	N	0.115	534	H	0.195
26	G	0.229	145	T	0.567	316	S	0.195	415	E	0.115	535	M	0.371
27	G	0.115	162	N	0.308	317	K	0.686	417	D	0.115	538	K	0.115
28	G	0.72	167	S	0.421	318	E	0.751	418	L	0.675	542	N	0.115
29	G	0.416	168	T	0.32	321	A	0.195	419	L	0.195	544	N	0.625
30	G	0.457	171	F	0.457	322	D	0.262	421	G	0.432	546	P	0.115
31	G	0.457	183	L	0.229	323	L	0.736	422	I	0.115	548	V	0.195
32	G	0.567	189	L	0.229	324	I	0.684	428	S	0.115	568	T	0.115
33	G	0.457	204	V	0.229	325	Y	0.115	429	N	0.308	574	N	0.262
34	G	0.457	232	V	1.544	326	I	0.115	433	D	0.432	575	P	0.195
35	G	0.457	233	A	1.074	327	Q	0.79	434	N	0.262	576	D	0.195
36	G	0.625	234	T	1.325	328	G	0.229	435	E	0.115	591	V	0.416
40	N	0.195	235	E	0.88	329	Q	0.229	437	E	0.115	592	L	0.115
53	V	0.115	236	T	0.99	330	D	0.654	447	R	0.482	603	N	0.115
75	T	0.115	237	L	0.342	331	N	0.195	448	T	0.992	615	V	0.308
89	T	0.115	238	T	1.18	332	T	0.115	451	I	0.229	617	T	0.972
90	K	0.643	240	D	1.073	337	F	0.115	452	H	0.115	618	K	0.421
91	T	0.308	241	A	0.625	341	E	0.115	454	S	0.32	619	D	0.308
92	H	1.279	242	V	0.96	344	K	0.229	468	S	0.115	620	K	0.195
93	Q	0.229	248	S	0.115	352	I	0.262	484	F	0.115	621	Y	0.115
94	Q	1.29	259	I	0.195	359	Y	0.786	485	S	0.654	627	N	0.115
95	K	0.195	266	N	0.115	361	Y	0.195	486	G	0.229	629	K	0.115
97	N	0.308	276	T	0.195	365	I	0.195	488	Q	0.342	632	F	0.115
110	V	0.432	278	L	0.308	371	Q	0.115	489	E	0.611	634	I	0.115
111	M	0.115	282	N	0.829	376	A	0.115	491	N	0.973	646	I	0.371

Note: Analysis was based on local isolates and GenBank sequences AMDV-G, BEL1/2, Far East, FIN5/C8, LN1-3, RUS09. RUS11, RUS14, RUS17, RUS19, SL3, Utah1 and Utah1 Kit (Group 4).

[£] Positions in the alignment of the 41 sequences (Group 4), conducting by MEGA6.

[¥] Entropy values of zero indicate no aa variation at the position and are thus not represented.

Table A17. Percentage of pairwise nucleotide identities over the entire coding region and partial 3' UTR sequence of the *Amdoparvoviruses*

CU7		88.1	87.2	89.2	87.7	87.7	87.8	87.8	88.3	-	88.3	88.3	87.7	86.6	86.6	86.1	86.1	86.6	86	87.7	88.3	87.7	88.3	88.3	87	87	85.6	85.2	86.7	87.5	87	88.3	-	87.8	-	77.9
YA1	90.2		94.1	96.4	94.7	92.3	92.3	92.9	92.3	-	96.4	94.7	93.5	94.1	94.1	94.1	93.5	94.1	93.5	95.9	95.3	95.9	96.4	94.1	95.2	95.2	91.7	90.5	93.5	91.1	93.5	88	-	88	-	79.9
YA2	90.0	95.9		93.8	93	94.1	91.1	92.2	91.1	-	96.1	93.8	94.9	94.1	94.1	92.5	93.5	93	94.4	95.5	96.1	92.7	94.9	94.9	91.9	91.9	94.1	92	99.5	94.7	96.8	89.2	-	89.3	-	80.4
YA3	91.2	93.3	94.1		96	92.6	95.3	95.9	94	-	97.6	95.2	94.7	95.5	95.5	94.9	94.9	95.5	94.7	97	96.4	97	97.6	95.8	93.8	93.8	91	90.9	93.2	92	96.6	88	-	88	-	79.9
YA4	90.9	92.5	91.8	94.6		95.2	91.7	92.2	92.2	-	98.3	92.7	97.8	97.9	97.9	97.3	97.3	97.9	95.5	97.8	98.3	97.2	97.2	96.1	90.8	90.8	94	91.4	92.5	92.4	93.5	91.2	-	91.2	-	81.6
KI2	91.1	90.0	91.2	93.1	94.3		91.6	91.6	92.1	-	96.1	92.1	96.1	95.2	95.2	94.6	94.6	95.2	93.3	95.5	96.1	94.9	94.9	93.8	90.8	90.8	94	89.3	93.5	89.8	91.4	87.1	-	87.1	-	77.6
LN1	91.5	90.9	90.5	92.3	92.1	91.9		97.8	94.4	-	91.6	96.7	90.5	91.7	91.7	91.1	91.1	91.7	88.8	91.1	90.5	91.6	90.5	91.7	91.5	91.5	94.9	87.8	90.6	89.4	91	87.6	-	86.5	-	81.8
LN3	91.2	91.2	90.5	92.4	92.5	92.6	96.0		94.4	-	92.2	95.6	91.1	92.2	92.2	91.7	91.7	92.2	90.5	91.6	91.1	92.2	91.1	92.2	92.1	92.1	94.9	89	91.7	91.6	90.4	89.8	-	88.8	-	82.4
LN2	91.2	90.8	89.9	91.7	92.0	92.0	95.4	95.8			92.2	96.7	91.1	91.6	91.6	92.2	91.1	91.6	89.4	91.6	92.2	92.2	91.1	92.2	92.6	92.6	96	89.4	90.5	89.9	92.1	87	-	86	-	80.4
SL3	91.4	92.1	89.8	93.1	94.8	93.6	93.4	94.6	94.3		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AMDV-G	91.6	92.3	90.0	93.2	94.8	93.7	93.5	94.8	94.4	99.6		92.7	97.2	97.8	97.8	97.8	97.2	97.8	97.2	99.4	98.9	98.9	98.9	97.2	93.8	93.8	95.5	91.6	95.5	93.3	94.4	89	-	88.7	-	83.7
Utah1	92.1	91.3	90.4	93.4	94.1	94.0	93.7	94.8	94.6	97.5	97.7		91.6	92.2	92.2	92.2	91.6	92.2	89.9	92.2	92.7	92.7	97.1	92.8	91.5	91.5	94.9	88.9	93.2	91	92.7	88.7	-	87.6	-	80.4
LU1	91.3	91.7	89.7	92.8	94.6	94.7	92.5	93.3	92.8	96.1	96.3	95.0		99.4	99.4	98.3	98.9	98.3	94.4	96.6	97.2	96.1	96.1	95	92.6	92.6	93.7	90.5	94.4	92.1	93.2	89.4	-	88.9	-	78.9
LU2	91.2	91.7	89.7	92.8	94.7	94.8	92.5	93.2	92.7	96.1	96.3	94.9	99.4		100	98.4	99.5	98.9	95	97.2	97.8	96.6	96.6	95.5	91.8	91.8	92.9	90.9	93.5	90.9	91.9	91.2	-	90.6	-	79.6
LU3	91.2	91.8	89.7	92.7	94.7	94.7	92.4	93.3	93.1	96.2	96.4	95.0	98.9	99.0		98.4	99.5	98.9	95	97.2	97.8	96.6	96.6	95.5	91.8	91.8	92.9	90.9	93.5	90.9	91.9	91.2	-	90.6	-	79.6
LU4	91.2	91.9	89.9	93.0	95.1	94.9	92.5	93.4	93.0	96.4	96.6	95.3	98.7	98.7	98.7		98.4	98.9	95.5	97.2	97.8	96.6	96.6	96.1	90.8	90.8	93.4	90.4	91.9	92.4	92.9	88.8	-	88.9	-	81
KI1	91.0	91.8	89.7	92.9	94.8	94.6	92.3	93.1	92.8	96.0	96.2	94.8	98.3	98.3	98.5	98.4		99.5	95.5	96.6	97.2	96.1	96.1	96.1	91.8	91.8	92.3	90.9	93	90.9	91.4	90.6	-	90.1	-	79.6
LU5	91.3	92.0	89.9	93.0	95.0	94.9	92.6	93.4	93.3	96.5	96.7	95.3	98.4	98.5	98.5	98.8	98.4		96.1	97.2	97.8	96.6	96.6	96.6	91.3	91.3	94	90.9	92.5	93	93.5	90.9	-	90.9	-	79.6
CU2	91.4	92.1	90.0	93.3	94.6	94.6	92.6	93.7	93.2	96.7	96.8	95.6	97.2	97.2	97.2	97.8	96.8	97.4		97.8	97.2	97.2	96.6	95.5	91.5	91.5	93.2	89.9	93.8	92.1	92.6	87.3	-	87	-	79.1
CU3	91.1	91.7	89.8	92.5	93.9	92.9	92.7	93.2	92.9	96.0	96.2	94.8	95.2	95.2	95.2	95.3	95.0	95.5	96.0		99.4	99.4	98.3	96.6	93.2	93.2	94.9	91.1	94.9	92.7	93.8	88.4	-	88.1	-	83.7
CU4	90.8	91.3	89.6	92.4	93.7	92.7	92.3	92.9	92.7	95.8	95.9	94.7	94.6	94.7	94.7	94.8	94.6	94.9	95.5	97.7		98.9	97.8	96.1	93.8	93.8	95.5	91.1	95.5	93.3	94.4	87.9	-	87.6	-	83.7
CO3	91.1	92.0	89.9	92.7	94.1	92.9	92.8	93.5	92.8	96.2	96.4	95.0	95.4	95.4	95.4	95.6	95.2	95.6	96.2	98.0	97.3		97.8	97.2	93.8	93.8	92.7	91.6	92.2	92.1	95.4	89.2	-	88.7	-	79.7
CU1	91.2	91.9	90.1	93.0	94.0	92.8	92.4	93.0	93.1	96.2	96.4	95.3	95.2	95.3	95.3	95.3	95.2	95.4	95.7	96.1	95.9	96.2		96.1	92.6	92.6	94.4	91.6	94.4	92.7	93.2	90.6	-	87.6	-	79.1
CO4	91.3	91.9	90.1	93.2	94.4	92.9	92.5	92.9	93.0	96.0	96.1	95.0	95.1	95.1	95.0	95.1	95.1	95.3	95.7	96.1	95.8	96.3	96.8		92.6	92.6	94.3	91.7	94.3	91	91	88.1	-	87.6	-	80.4
CU5	90.7	92.0	90.5	91.2	92.2	91.1	91.7	92.1	92.0	93.9	94.1	93.2	93.2	93.2	93.1	93.3	93.1	93.2	93.3	93.8	93.6	93.8	94.4	94.3		100	91.8	90.8	91.4	91.3	94	86.1	-	85.6	-	78.4
CU6	90.8	92.1	90.5	91.2	92.2	91.2	91.8	92.3	92.2	94.1	94.3	93.3	93.4	93.5	93.4	93.5	93.3	93.4	93.5	94.0	93.7	93.9	94.5	94.6	99.1		91.8	90.8	91.4	91.3	94	86.1	-	85.6	-	78.4
CO1	90.5	90.7	90.8	90.1	90.2	90.2	90.2	90.0	90.0	90.4	90.5	90.6	90.4	90.4	90.3	90.3	90.1	90.4	90.1	90.5	90.5	90.7	90.7	90.5	91.3	91.4		92.5	93.6	88.2	93	88	-	89.2	-	81.8
PI1	90.2	89.9	90.2	90.2	90.3	90.7	90.2	90.0	89.7	90.2	90.4	90.6	90.2	90.2	90.1	90.1	90.1	90.2	90.0	90.2	90.5	90.0	90.2	90.1	91.0	91.0	93.2		91.4	89.8	89.2	84.1	-	85.2	-	83.1
HA1	90.3	91.4	90.8	90.3	91.0	90.0	90.5	90.0	90.0	91.1	91.3	90.8	90.8	90.9	90.8	90.7	90.7	90.8	90.5	90.7	90.7	90.8	91.2	90.9	91.7	91.8	92.8	92.4		94.1	96.2	88.6	-	88.7	-	79.7
HA2	90.7	90.9	90.9	90.8	90.7	90.7	90.9	90.6	90.3	90.8	91.1	91.3	90.6	90.7	90.6	90.6	90.5	90.5	90.4	90.6	90.6	90.6	91.0	90.7	91.8	91.9	94.2	93.2	95.0		93.5	86.7	-	87.4	-	79.7
CO2	90.5	91.2	90.5	91.1	91.1	91.0	92.0	91.6	91.5	91.8	92.0	92.0	91.3	91.4	91.2	91.4	91.3	91.5	91.3	91.9	92.2	91.9	92.4	92.2	93.8	93.9	92.0	92.1	92.8	93.2		86.8	-	88	-	81.1
HS-R	85.3	86.1	85.2	85.1	85.7	85.2	85.6	85.1	85.6	86.2	86.2	86.0	85.7	85.8	85.7	85.7	85.6	85.9	85.6	85.7	85.4	86.0	86.3	85.9	86.0	86.1	85.7	85.5	85.4	85.5	85.9		-	98.9	-	82.4
XQ-JLR	85.6	86.2	85.4	85.4	85.9	85.4	85.9	85.6	85.9	86.3	86.4	86.2	85.8	85.9	85.8	85.8	85.7	85.9	85.8	85.8	85.6	86.0	86.5	86.1	86.1	86.2	85.8	85.7	85.6	85.6	85.9	98.6	-	-	-	-
QA-RF	85.5	86.1	85.1	85.1	85.8	85.2	85.8	85.5	85.9	86.4	86.4	86.2	86.0	86.1	86.0	85.9	85.9	86.1	85.8	85.9	85.6	86.1	86.4	86.2	86.1	86.2	85.6	85.6	85.4	85.6	85.7	98.0	98.1		-	83.8
HC-R	85.6	86.3	85.7	85.5	85.9	85.3	86.0	85.5	86.0	86.3	86.3	86.2	85.9	86.1	85.9	86.0	85.7	85.9	85.8	85.8	85.5	86.0	86.5	86.2	86.3	86.5	85.8	85.9	85.8	85.7	85.8	97.1	97.3	97.1		-
GFAV	76.1	76.1	76.5	76.5	76.7	76.7	76.7	76.8	76.6	76.5	76.5	76.5	76.0	76.2	76.2	76.3	76.3	76.2	76.3	76.2	76.1	76.2	76.7	76.6	76.1	76.1	76.4	76.4	75.5	76.3	76.5	76.1	76.2	76.5	76.1	

Note: Values above and below the diagonal shaded frames indicate the percentage of nucleotide sequence identity for the partial 3' UTR sequence and the entire coding region, respectively. The 3' UTR sequence of isolates SL3, HC-R and XQ-JLR were not available on GenBank. See footnotes of the Table 3.6 for sequences in Group 5.

Table A18. Percentage of amino acid identity and divergence in the NS1 protein of the *Amdoparvoviruses*

HC-R		0.06	0.05	0.05	0.26	0.25	0.26	0.26	0.25	0.25	0.25	0.25	0.25	0.25	0.27	0.26	0.25	0.25	0.26	0.26	0.26	0.25	0.25	0.25	0.25	0.26	0.24	0.26	0.25	0.25	0.26	0.25	0.25	0.25	0.26	0.26	0.24	0.35
HS-R	94.4		0.04	0.05	0.28	0.26	0.27	0.27	0.26	0.26	0.26	0.26	0.26	0.26	0.28	0.28	0.26	0.26	0.26	0.27	0.27	0.26	0.26	0.26	0.25	0.27	0.24	0.27	0.27	0.25	0.27	0.25	0.26	0.26	0.26	0.26	0.24	0.35
XQ-JLR	94.9	96.4		0.05	0.27	0.25	0.27	0.27	0.26	0.26	0.25	0.25	0.25	0.26	0.27	0.27	0.25	0.25	0.26	0.26	0.27	0.26	0.25	0.26	0.25	0.27	0.25	0.27	0.26	0.25	0.26	0.25	0.26	0.26	0.26	0.26	0.25	0.35
QA-RF	94.5	95.3	94.9		0.27	0.26	0.27	0.27	0.26	0.27	0.27	0.26	0.26	0.26	0.27	0.28	0.26	0.26	0.26	0.26	0.26	0.27	0.25	0.27	0.26	0.28	0.25	0.27	0.27	0.25	0.26	0.26	0.27	0.26	0.27	0.27	0.25	0.34
CU7	73.6	72.4	72.7	72.5		0.18	0.18	0.18	0.17	0.18	0.19	0.19	0.18	0.19	0.20	0.20	0.17	0.18	0.18	0.18	0.18	0.18	0.17	0.18	0.18	0.17	0.18	0.19	0.21	0.20	0.21	0.21	0.18	0.18	0.20	0.20	0.20	0.34
CU1	74.7	73.9	74.3	73.3	81.1		0.09	0.09	0.08	0.09	0.10	0.10	0.10	0.17	0.17	0.16	0.15	0.13	0.13	0.13	0.13	0.12	0.11	0.11	0.12	0.13	0.14	0.18	0.19	0.17	0.19	0.16	0.14	0.13	0.20	0.19	0.17	0.31
CO4	73.9	73.2	73.2	72.7	81.3	90.2		0.10	0.09	0.10	0.11	0.11	0.11	0.16	0.17	0.15	0.15	0.12	0.13	0.13	0.13	0.12	0.12	0.11	0.13	0.12	0.12	0.18	0.19	0.17	0.18	0.15	0.14	0.13	0.20	0.19	0.17	0.32
CU3	73.2	72.4	72.5	72.4	81.4	89.9	89.1		0.02	0.04	0.09	0.09	0.10	0.15	0.15	0.14	0.13	0.11	0.12	0.12	0.12	0.11	0.10	0.09	0.11	0.12	0.12	0.17	0.19	0.17	0.19	0.16	0.14	0.14	0.19	0.19	0.16	0.32
CU4	73.3	72.2	72.7	72.2	81.1	89.5	89.1	95.6		0.04	0.08	0.08	0.08	0.13	0.13	0.13	0.12	0.10	0.10	0.10	0.11	0.09	0.09	0.07	0.10	0.11	0.11	0.16	0.18	0.16	0.17	0.14	0.13	0.12	0.18	0.18	0.15	0.30
CO3	74.4	74.1	73.6	73.2	82.2	90.2	89.7	94.9	93.8		0.09	0.10	0.10	0.15	0.15	0.16	0.14	0.11	0.11	0.11	0.11	0.10	0.09	0.09	0.11	0.11	0.11	0.17	0.19	0.17	0.20	0.17	0.14	0.14	0.18	0.18	0.15	0.32
AMDV-G	74.9	73.9	74.7	73.5	81.4	89.5	88.3	90.2	90.3	90.6		0.01	0.02	0.14	0.12	0.11	0.08	0.11	0.11	0.10	0.10	0.09	0.08	0.08	0.10	0.12	0.11	0.19	0.19	0.17	0.19	0.17	0.15	0.14	0.20	0.20	0.16	0.32
SL3	75.2	74.3	75	73.8	81	89.5	88.3	90.2	90	90.3	99.2		0.02	0.14	0.13	0.11	0.09	0.11	0.11	0.11	0.10	0.09	0.09	0.08	0.10	0.12	0.11	0.19	0.20	0.18	0.20	0.17	0.15	0.15	0.21	0.20	0.16	0.32
Utah1	75	74.3	74.7	73.9	81.6	89.1	88.6	89.5	89.9	90.3	98.3	97.5		0.13	0.11	0.12	0.08	0.10	0.11	0.10	0.10	0.09	0.08	0.08	0.10	0.11	0.11	0.19	0.19	0.17	0.18	0.17	0.15	0.14	0.20	0.19	0.15	0.32
LN1	74.9	73.8	74.4	73.6	81.4	82.4	83.5	84.7	85	85	86.4	85.8	86.7		0.08	0.09	0.09	0.15	0.16	0.15	0.16	0.15	0.14	0.15	0.14	0.15	0.14	0.20	0.22	0.19	0.20	0.17	0.18	0.17	0.21	0.20	0.17	0.33
LN3	73.3	71.9	72.9	72.5	80.2	82.7	83	84.7	84.9	84.6	87.7	87.4	88.3	91.6		0.08	0.07	0.15	0.16	0.15	0.16	0.14	0.14	0.14	0.15	0.16	0.15	0.22	0.22	0.21	0.21	0.19	0.19	0.18	0.22	0.21	0.18	0.33
LN2	73.6	72.4	73.2	72.4	80.2	83.9	84.4	85	84.9	84.2	89.2	89.2	89.4	90.8	91.9		0.06	0.14	0.15	0.14	0.15	0.14	0.13	0.14	0.13	0.16	0.14	0.19	0.22	0.20	0.21	0.18	0.18	0.17	0.22	0.21	0.17	0.34
United	74.9	73.9	74.9	73.8	82.5	84.9	85	86.1	86.3	86.1	91.6	91.3	92.4	91.3	92.7	94.2		0.13	0.14	0.13	0.14	0.12	0.11	0.12	0.13	0.14	0.13	0.18	0.20	0.19	0.19	0.17	0.17	0.16	0.21	0.20	0.15	0.33
LU1	74.3	73.5	73.9	73.8	81.4	86.3	86.7	87.4	87.5	88.6	88.8	88.5	88.9	84.4	84.2	85.2	86.3		0.00	0.02	0.04	0.03	0.04	0.05	0.06	0.12	0.11	0.18	0.19	0.17	0.18	0.18	0.16	0.16	0.20	0.20	0.17	0.32
LU2	74.4	73.6	74.1	73.9	81.9	86.7	87.2	87.7	87.8	88.9	89.2	88.9	89.4	84.4	84.4	85.3	86.4	99.1		0.02	0.04	0.03	0.04	0.12	0.16	0.12	0.11	0.19	0.19	0.17	0.18	0.11	0.16	0.16	0.20	0.20	0.16	0.32
LU3	73.8	72.9	73.6	73.3	82.1	86.4	86	87.5	87.8	88.5	89.4	89.1	89.2	84.4	84.4	85.3	86.6	96.9	97.8		0.03	0.03	0.04	0.05	0.06	0.12	0.12	0.19	0.19	0.17	0.19	0.19	0.16	0.16	0.20	0.20	0.17	0.32
KI1	73	72.5	72.7	73	81.1	86	85.6	86.3	86.4	88.5	89.1	88.8	89.1	83.6	83.6	84.4	85.5	94.7	95.5	95.8		0.04	0.04	0.06	0.06	0.11	0.11	0.19	0.19	0.18	0.19	0.19	0.17	0.17	0.20	0.20	0.18	0.32
LU4	74.4	73.8	74.1	73.3	82.1	87.4	87.2	88.3	88.6	90	91	90.6	91.1	85	85.5	86.1	87.7	96.3	96.9	96.1	95.5		0.03	0.03	0.05	0.10	0.09	0.19	0.19	0.18	0.19	0.18	0.16	0.16	0.19	0.19	0.17	0.32
LU5	74.7	74.1	74.4	74.4	82.2	87.8	87.2	88.9	88.8	90.2	91.3	91	91.7	85.3	85.3	86.3	88.1	95.2	95.9	95.5	95.3	97.2		0.04	0.06	0.10	0.10	0.18	0.19	0.18	0.18	0.18	0.16	0.15	0.19	0.19	0.16	0.31
CU2	74.1	73.3	73.8	72.4	82.1	88.3	88.5	89.7	90	90.6	91.3	91	91.1	84.1	85	85.8	87.4	93.4	94.2	93.6	92.4	96.3	95.2		0.06	0.10	0.10	0.19	0.20	0.19	0.19	0.18	0.15	0.15	0.19	0.19	0.16	0.31
KI2	74.6	73.9	74.3	73.5	81.7	86.9	86.3	87.5	87.7	88.5	89.4	89.2	89.5	84.9	84.7	86	86.7	93	93.6	93	92.5	93.8	93.3	92.7		0.11	0.09	0.19	0.19	0.17	0.18	0.18	0.16	0.16	0.19	0.19	0.16	0.31
YA3	73.3	72.7	72.5	71.9	82.7	85.5	86.9	86.4	86.1	87.7	87.5	86.9	87.8	84.1	83.5	83.5	84.9	87.1	87.5	87.1	87.4	88.8	88.6	88.5	87.5		0.07	0.19	0.18	0.17	0.19	0.19	0.17	0.16	0.12	0.13	0.16	0.31
YA4	75.7	75.2	75	74.4	81.4	85.3	86.9	87.1	86.4	88	88.6	88.9	89.1	86	84.4	85.8	86.7	88.1	88.3	87.8	88	90	89.4	88.9	89.4	91.6		0.19	0.18	0.18	0.19	0.18	0.17	0.17	0.16	0.16	0.16	0.32
CO1	73.3	72.4	72.4	72.9	80	81.3	81.4	81.9	82.1	82.2	80.7	80.3	80.7	79.7	78	80.3	81.3	80.8	81	80.3	79.6	80.7	80.8	79.9	79.9	79.9	80.3		0.14	0.11	0.15	0.15	0.15	0.15	0.16	0.17	0.16	0.33
HA1	74.6	73.2	74.1	73.2	79.1	80	80	80.2	80	80.2	80.7	80.2	81.1	78.2	77.4	78.2	79.6	80.5	81	80.3	79.7	80.5	80.3	79.7	80.5	81.3	81.3	85.2		0.08	0.17	0.14	0.16	0.16	0.17	0.18	0.18	0.34
HA2	74.9	74.3	74.1	74.4	79.6	81.7	82.1	81.6	81.9	81.9	82.1	81.6	82.1	80.3	78.8	79.9	81	82.1	82.4	81.7	80.8	81.6	81.4	80.3	81.6	81.7	81.3	87.5	91.6		0.14	0.12	0.14	0.14	0.16	0.17	0.15	0.33
PI1	73.9	73	73.3	73.6	79.1	80.3	81.6	80.7	81.1	79.9	80.8	80.2	81.3	79.4	78.3	79.3	81	81.1	81.4	80.7	79.9	81	81.3	80	81.1	80	80.8	84.9	82.8	85.3		0.16	0.17	0.17	0.20	0.20	0.17	0.17
CO2	74.3	74.3	74.4	73.8	78.9	83.2	83.9	82.8	83.5	82.2	82.7	82.7	82.8	82.4	80.8	81.3	82.2	81.4	81.6	80.3	79.7	81.7	81.4	81	81.4	80.2	81	84.7	85.5	87.2	83.6		0.13	0.13	0.19	0.19	0.18	0.33
CU5	74.1	73.3	73	72.7	81	85.2	85.5	84.7	84.9	84.9	84.7	84.2	84.7	81.4	80.7	81.6	82.8	82.8	83.5	82.8	81.9	83.5	83.3	83.8	82.5	82.1	82.2	83.8	82.7	84.7	82.1	86.1		0.00	0.18	0.18	0.18	0.33
CU6	74.1	73.2	72.9	72.5	80.8	84.9	85	84.4	84.7	84.2	84.2	83.9	84.2	81.3	80.5	81.4	82.5	82.4	83	82.5	81.6	83	83	83.6	82.4	81.6	81.6	83.3	82.5	84.6	81.3	86	98		0.18	0.18	0.17	0.33
YA1	73.8	73.3	73.3	72.7	79.6	79.3	79.6	79.9	79.6	81.3	79.4	78.9	79.7	78.6	77.4	78	78.8	79.3	79.4	79.4	78.9	80.5	80.5	80.3	79.6	86.4	83.5	82.7	81.7	82.7	78.8	80.2	81	80.5		0.03	0.14	0.32
YA2	73.9	73.5	73.8	73	79.7	79.9	80.2	80.3	79.7	81.4	80.2	79.7	80.5	79.6	78.3	79.1	79.6	79.3	79.6	79.4	79.3	80.7	80.8	80.3	80	86.3	83.6	82.7	81.9	82.7	79.4	80.2	81	80.5	96.6		0.15	0.32
ADV-K	76.3	75.5	74.9	74.9	80.5	82.4	82.5	83.5	83.3	84.9	84.1	83.8	84.6	82.5	81.6	83.2	84.7	82.4	82.7	82.5	81.6	83.3	83.8	83.3	83.2	83.6	84.1	83.5	81.6	84.4	80.7	81.6	81.4	81.3	85	84.6		0.32
GFAV	65.2	64.4	64.8	66.1	66.3	68.4	67.7	67.7	68.2	68.2	68.2	68.4	68.2	66.8	66.9	66.3	67.4	67.9	68.1	67.7	67.6	68.1	68.5	68.2	68.1	68.7	68.1	66.6	66	66.8	65.8	66.6	66.3	66.1	67.1	67.4	68.2	

Note: Values above and below the diagonal shaded frames indicate the percentage of aa sequence identity and divergence for the NS1 protein, respectively. **See** footnotes of the Table 3.6 for sequences in Group 3.

Table A19. Amino acid sequence alignment of the hypervariable region of the VP2

Clusters	VP2 aa	232	233	234	235	236	237	238	239	240	241	242
AMDV-1a(1)	G	V	A	T	E	T	L	T	W	D	A	V
	SL3	.	.	.	.	.	.	.	.	.	.	.
	Utah1	M	G	Q	.	Q	.	E	.	T	G	T
AMDV-1a(2)	LU1, CU2	.	.	Q	.	.	.	.	.	.	.	.
	LU2	.	P	Q	.	.	.	.	.	.	.	.
	LU3/5, KI1	.	.	S	.	.	.	.	.	.	.	.
	LU4	.	.	X	.	.	.	.	.	.	.	.
	YA4	.	.	S	.	.	.	X	.	.	.	.
	KI2	A	G	Q	G	Q	.	E	.	N	.	T
AMDV-1a(3)	CU3/4	.	.	.	.	.	.	.	.	.	.	.
	CO3	.	S	S	.	.	.	.	.	.	.	.
AMDV-1a(4)	CO4	S	.	S	.	.	.	.	.	.	.	.
	CU1	A	.	.	.	.	.	.	.	.	.	.
AMDV-1b	LN1/3	S	G	Q	S	Q	.	E	.	T	G	T
	LN2	T	G	Q	Q	Q	.	E	.	T	G	T
AMDV-2	CO1	-	.	K	.	S	Q	D	.	I	.	I
	HA1	A	.	S	.	.	.	.	.	.	.	.
	HA2	-	.	Q	G	S	Q	D	.	T	G	I
	CO2	S	G	Q	-	Q	.	E	.	T	G	T
	PI	.	.	Q	S	Q	.	E	.	T	.	I
AMDV-3	YA1	.	.	S	.	.	.	.	.	.	.	.
	YA2	-	R	.	.	Q	T	D	.	T	G	T
AMDV-4	CU7	S	G	Q	.	Q	.	E	.	T	G	T
AMDV-RO1/2	CU5/6	S	.	.	.	.	.	.	.	N	.	.
AMDV-RO1/3	YA3	A	.	Q	S	Q	.	E	.	T	G	T

Note: HVR of the VP2 locates at nucleotide 3036 to 3196 of the AMDV-G genome. A dash indicates deletion of the aa and dots represent similar to the AMDV-G sequence.

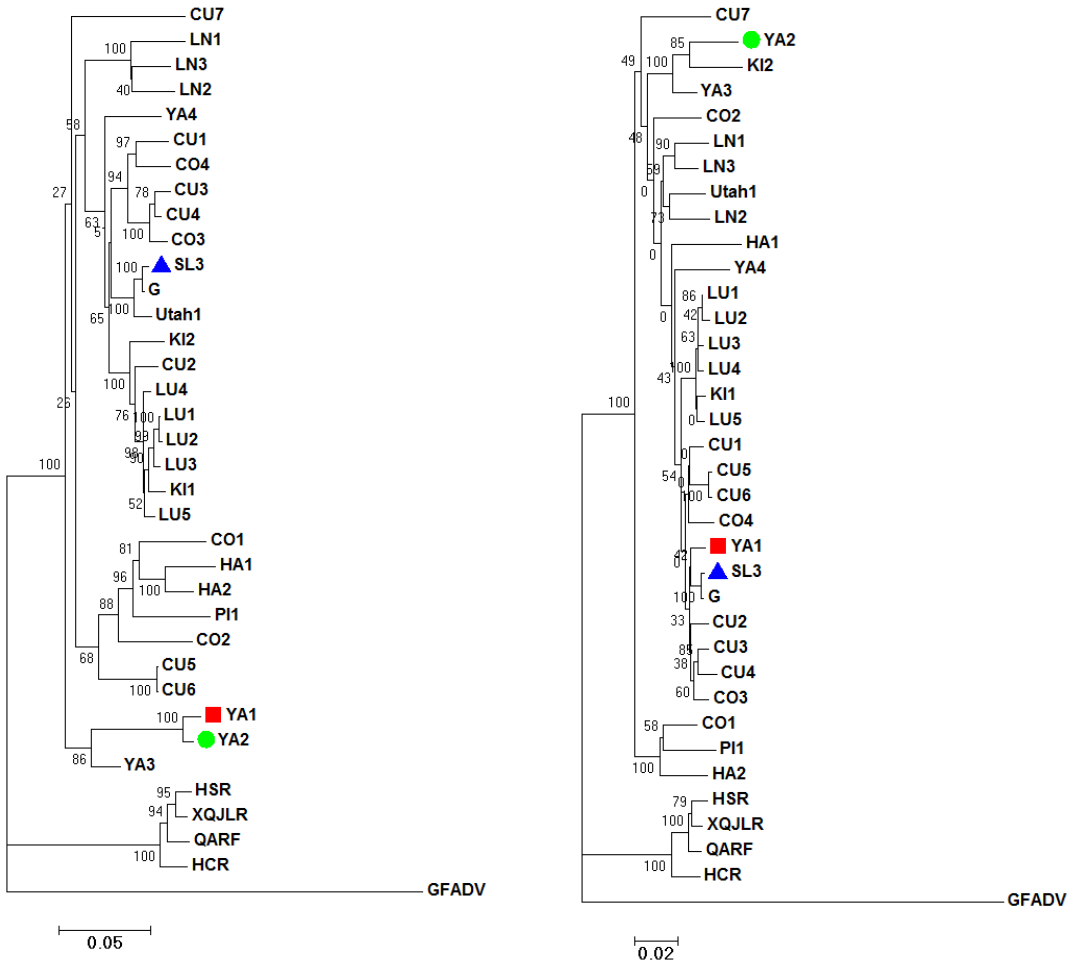


Figure A1. Neighbor Joining tree of the Recombination Event 1.

Neighbor Joining tree of Recombination Event 1 which was constructed, based on the region derived from the portion of alignment between the identified recombination regions (major parent: nt 1 - 2509 and 4215 – 4215 of the alignment). Recombinant regions were removed from alignment. The tree is constructed by RDP4. Recombinant sequence is YA1 (square), major parent is YA2 (circle) and minor parent is SL3 (triangle). Scale below the tree is the number of substitutions per site.

Neighbor Joining tree of the recombination event 1 which was constructed based on the region derived from the identified recombination region in the alignment (minor parent: nt 2510 – 4215 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is YA1 (square), major parent is YA2 (circle) and minor parent is SL3 (triangle). Scale below the tree is substitution per site.

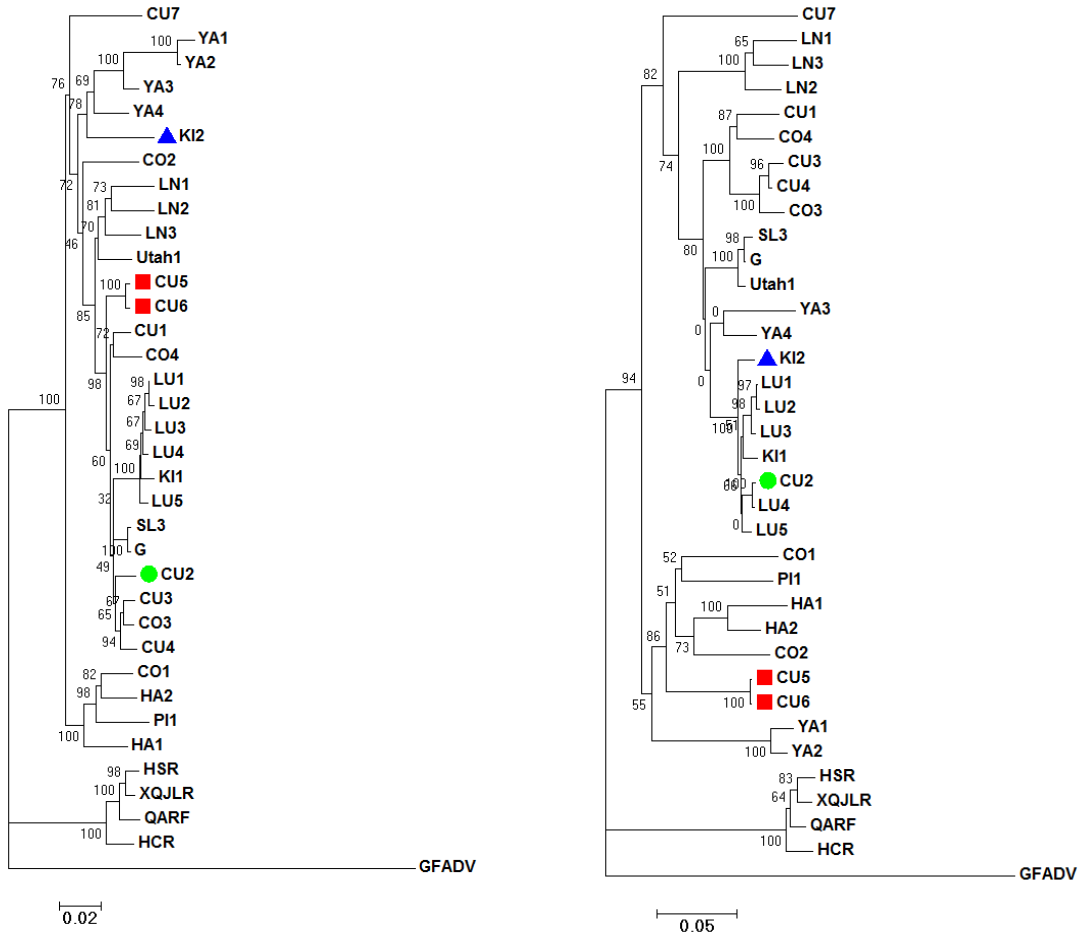


Figure A2. Neighbor Joining tree of the recombination event 2.

Neighbor Joining tree of the recombination event 2 which was constructed based on the region derived from the portion of alignment between the identified recombination region (major parent: nt 1 – 0 and 1269 – 4215 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is CU6 (square). CU5 is a sequence with evidence of the same recombination event (square), major parent is CU2 (circle) and minor parent is unknown (KI2) (triangle). Scale below the tree is substitution per site.

Neighbor Joining tree of the recombination event 2 which was constructed based on the region derived from the identified recombination region in the alignment (minor parent: nt 1 - 1268 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is CU6 (square). CU5 is a sequence with evidence of the same recombination event (square), major parent is CU2 (circle) and minor parent is unknown (KI2) (triangle). Scale below the tree is substitution per site.

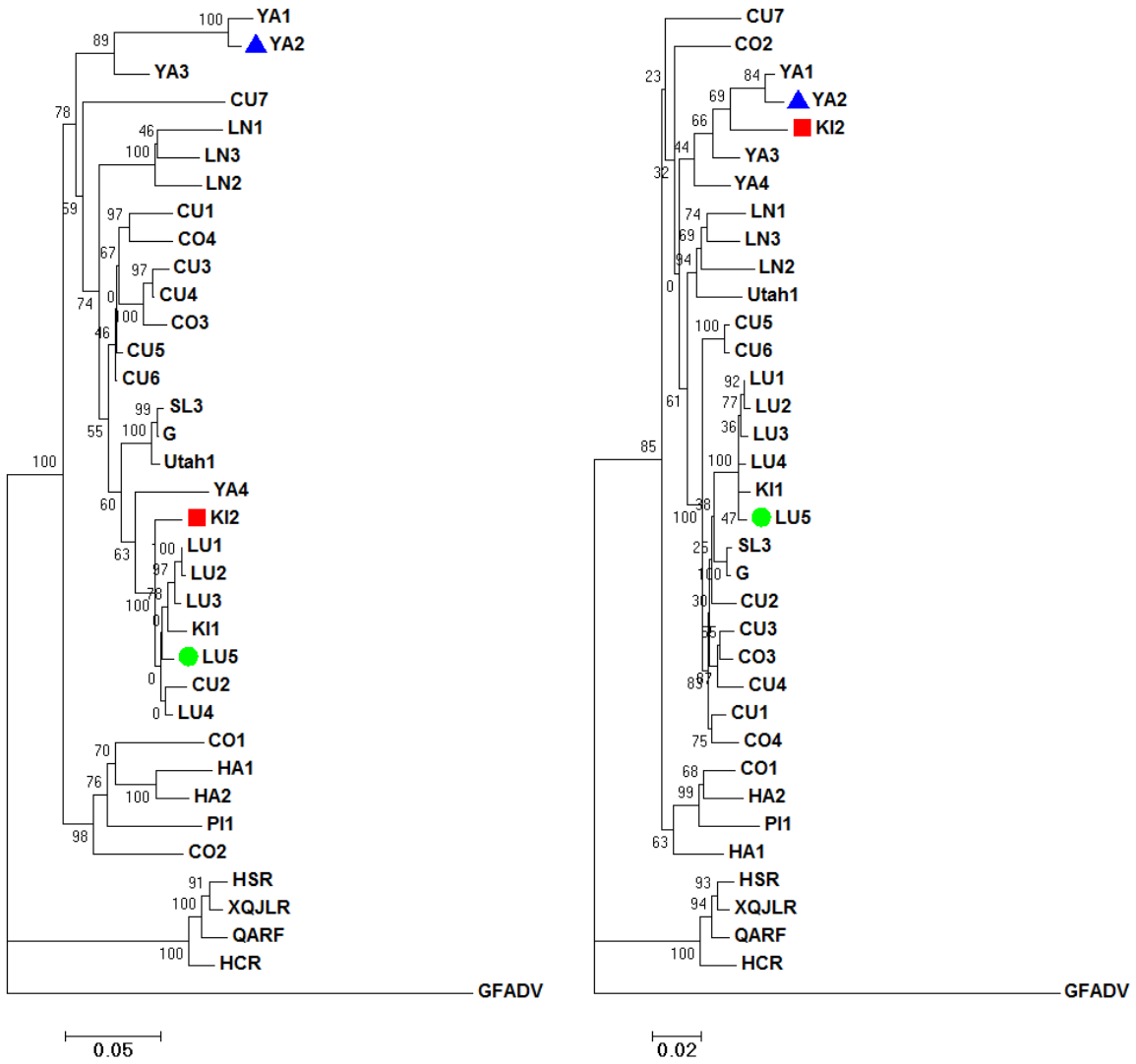


Figure A3. Neighbor Joining tree of the recombination event 3.

Neighbor Joining tree of the recombination event 3 which was constructed based on the region derived from the portion of alignment between the identified recombination region (major parent: nt 1 – 1923 and 4215 – 4215 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is KI2 (square), major parent is LU5 (circle) and minor parent is YA2 (triangle). Scale below the tree is substitution per site.

Neighbor Joining tree of the recombination event 3 which was constructed based on the region derived from the identified recombination region in the alignment (minor parent: nt 1924 – 4215 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is KI2 (square), major parent is LU5 (circle) and minor parent is YA2 (triangle). Scale below the tree is substitution per site.

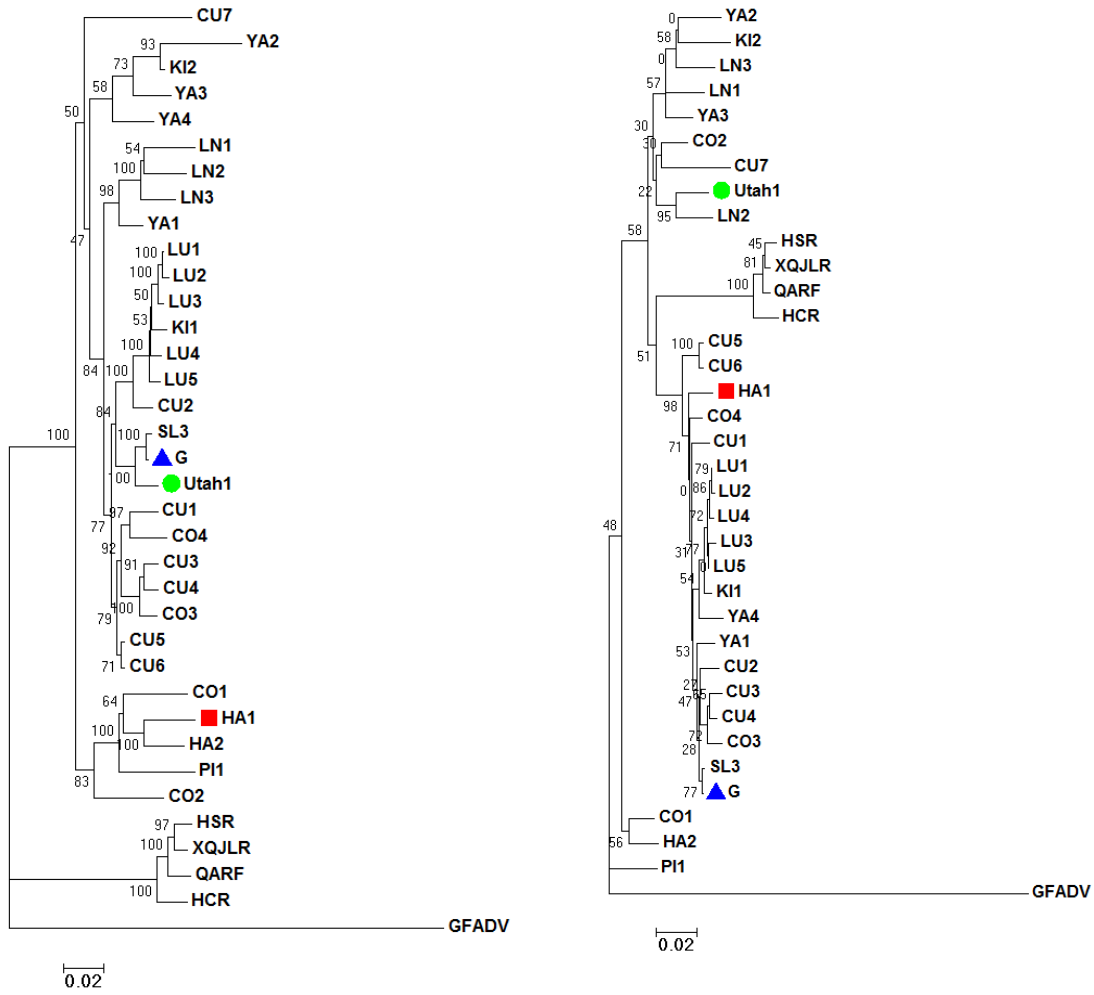


Figure A4. Neighbor Joining tree of the recombination event 4.

Neighbor Joining tree of the recombination event 4 which was constructed based on the region derived from the portion of alignment between the identified recombination region (major parent: nt 1 – 2278 and 3039 - 4215 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is HA1 (square), major parent is Unknown (Utah1) (circle) and minor parent is G (triangle). Scale below the tree is substitution per site.

Neighbor Joining tree of the recombination event 4 which was constructed based on the region derived from the identified recombination region in the alignment (minor parent: nt 2279 - 3038 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is HA1 (square), major parent is Unknown (Utah1) (circle) and minor parent is G (triangle). Scale below the tree is substitution per site.

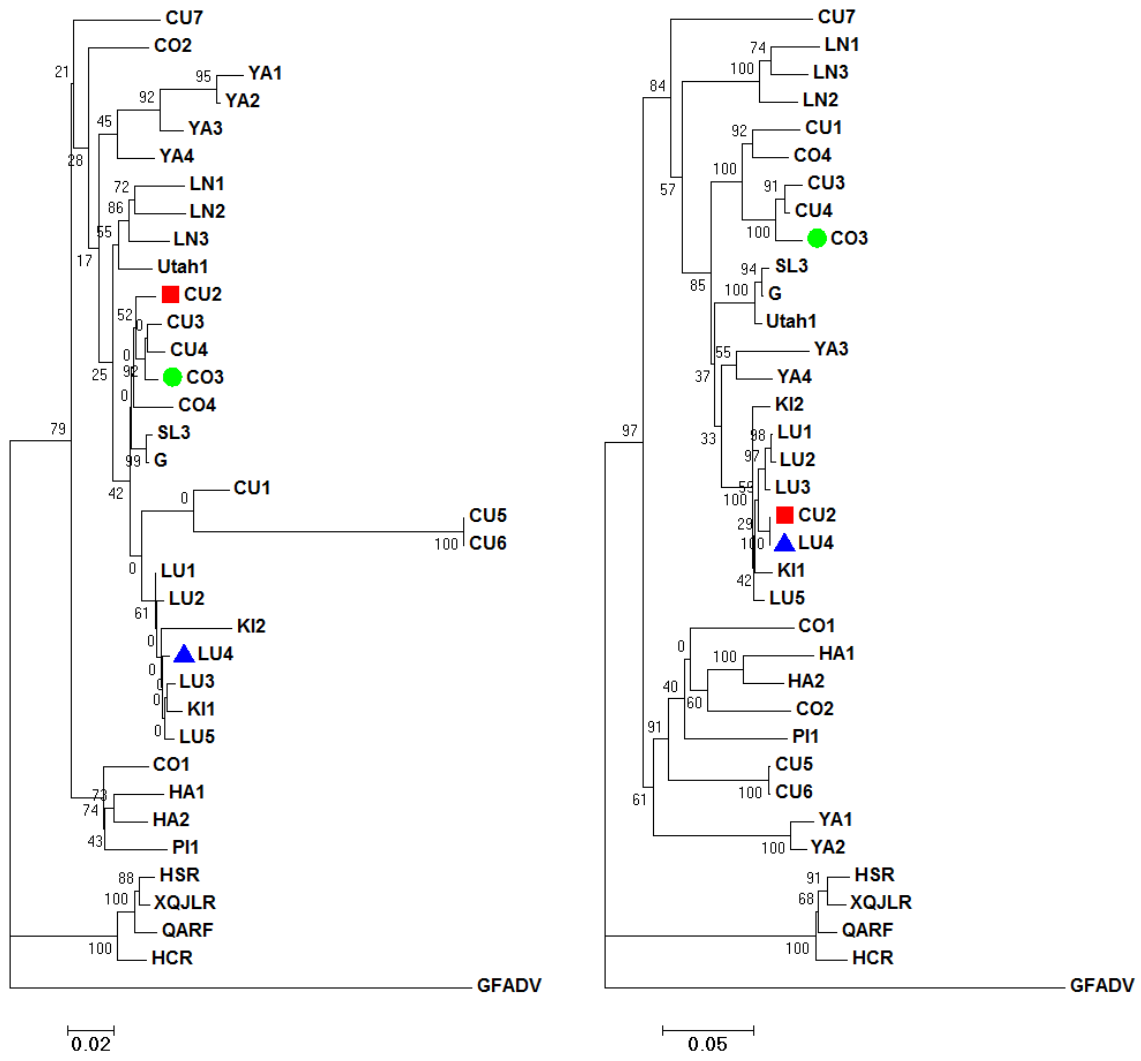


Figure A5. Neighbor Joining tree of the recombination event 5.

Neighbor Joining tree of the recombination event 5 which was constructed based on the region derived from the portion of alignment between the identified recombination region (major parent: nt 1 – 44 and 1225 - 4215 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is CU2 (square), major parent is CO3 (circle) and minor parent is LU4 (triangle). Scale below the tree is substitution per site.

Neighbor Joining tree of the recombination event 5 which was constructed based on the region derived from the identified recombination region in the alignment (minor parent: nt 45 - 1224 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is CU2 (square), major parent is CO3 (circle) and minor parent is LU4 (triangle). Scale below the tree is substitution per site.

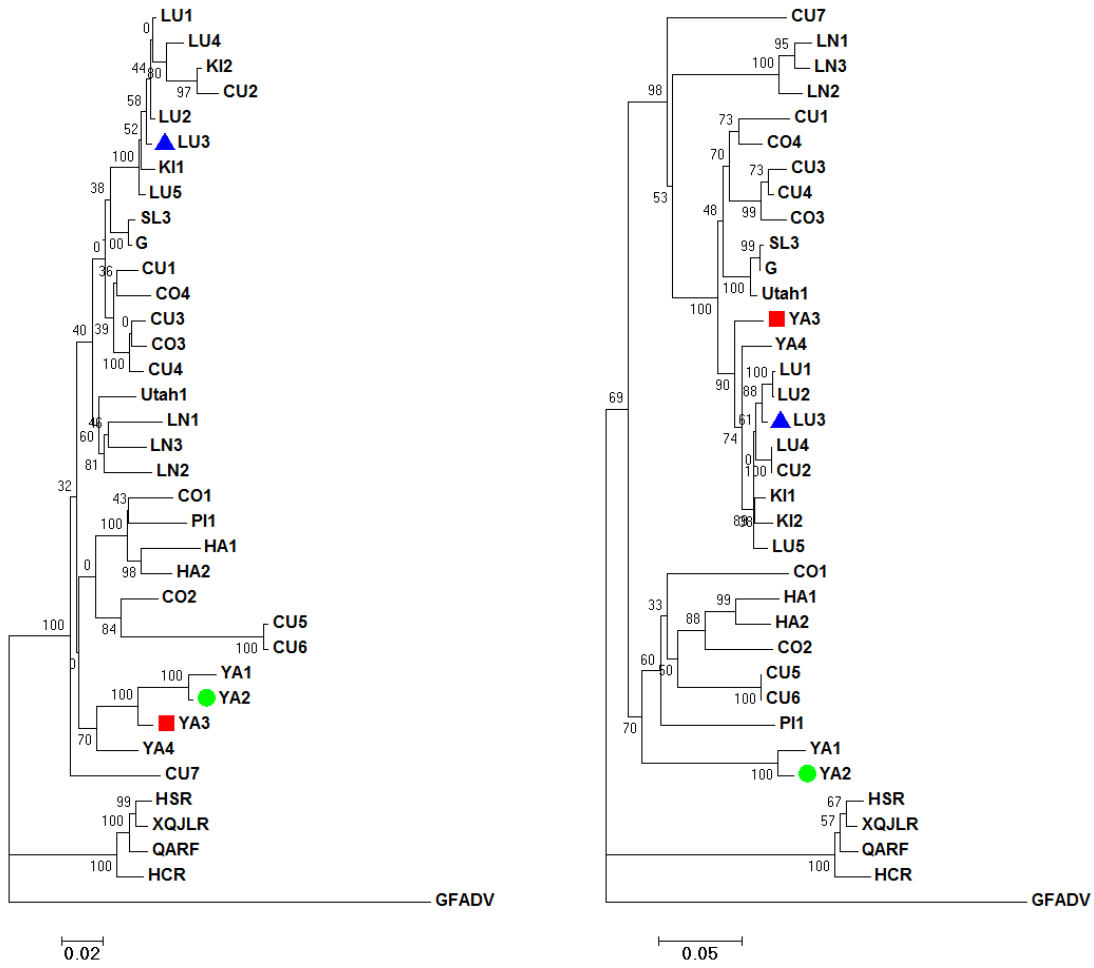


Figure A6. Neighbor Joining tree of the recombination event 6.

Neighbor Joining tree of the recombination event 6 which was constructed based on the region derived from the portion of alignment between the identified recombination region (major parent: nt 1 – 135 and 810 - 4215 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is YA3 (square), major parent is YA2 (circle) and minor parent is LU3 (triangle). Scale below the tree is substitution per site.

Neighbor Joining tree of the recombination event 6 which was constructed based on the region derived from the identified recombination region in the alignment (minor parent: nt 136 - 809 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is YA3 (square), major parent is YA2 (circle) and minor parent is LU3 (triangle). Scale below the tree is substitution per site.

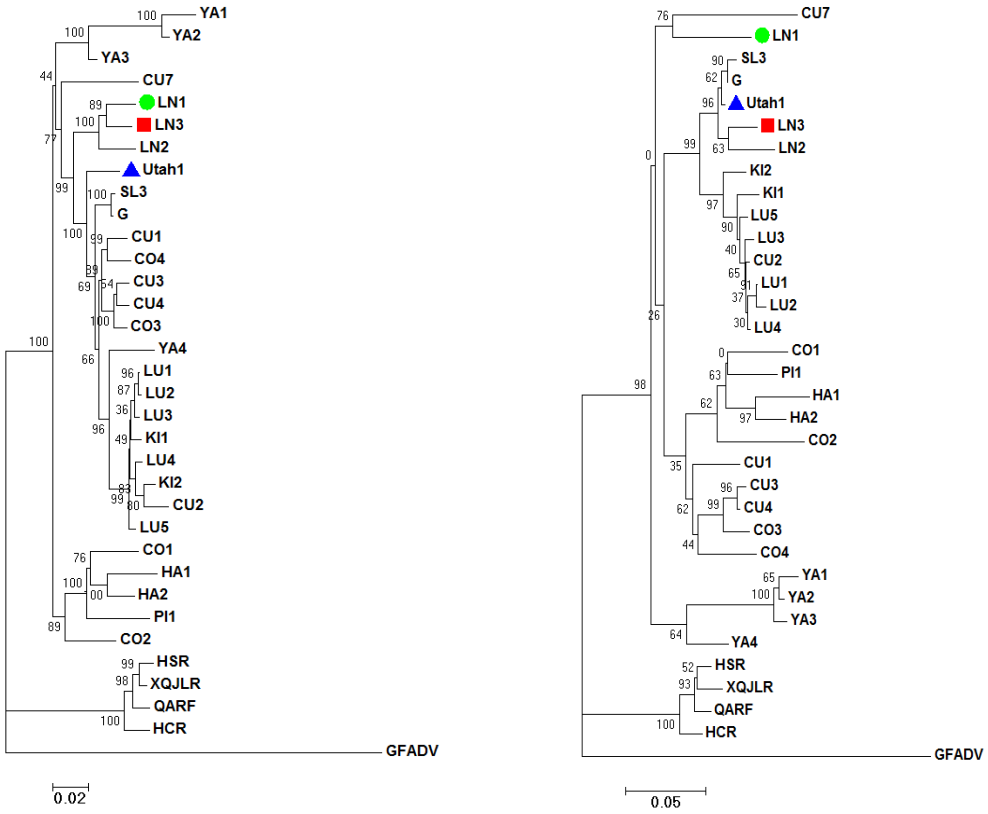


Figure A7. Neighbor Joining tree of the recombination event 7.

Neighbor Joining tree of the recombination event 7 which was constructed based on the region derived from the portion of alignment between the identified recombination region (major parent: nt 1 – 958 and 1494 - 4215 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is LN3 (square), major parent is LN1 (circle) and minor parent is Utah1 (triangle). Scale below the tree is substitution per site.

Neighbor Joining tree of the recombination event 7 which was constructed based on the region derived from the identified recombination region in the alignment (minor parent: nt 959 - 1493 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is LN3 (square), major parent is LN1 (circle) and minor parent is Utah1 (triangle). Scale below the tree is substitution per site.

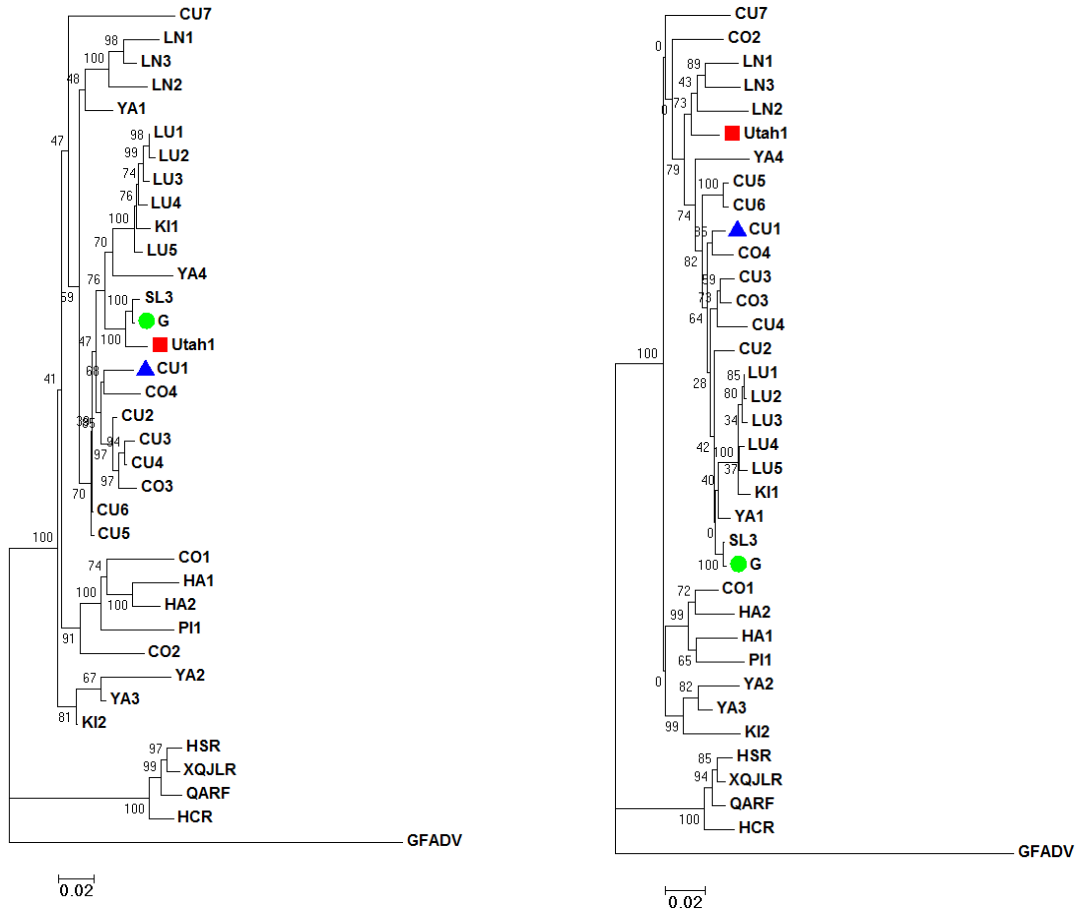


Figure A8. Neighbor Joining tree of the recombination event 8.

Neighbor Joining tree of the recombination event 8 which was constructed based on the region derived from the portion of alignment between the identified recombination region (major parent: nt 1 – 1903 and 3517 - 4215 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is Utah1 (square), major parent is G (circle) and minor parent is Unknown (CU1) (triangle). Scale below the tree is substitution per site.

Neighbor Joining tree of the recombination event 8 which was constructed based on the region derived from the identified recombination region in the alignment (minor parent: nt 1904 - 3516 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is Utah1 (square), major parent is G (circle) and minor parent is Unknown (CU1) (triangle). Scale below the tree is substitution per site.

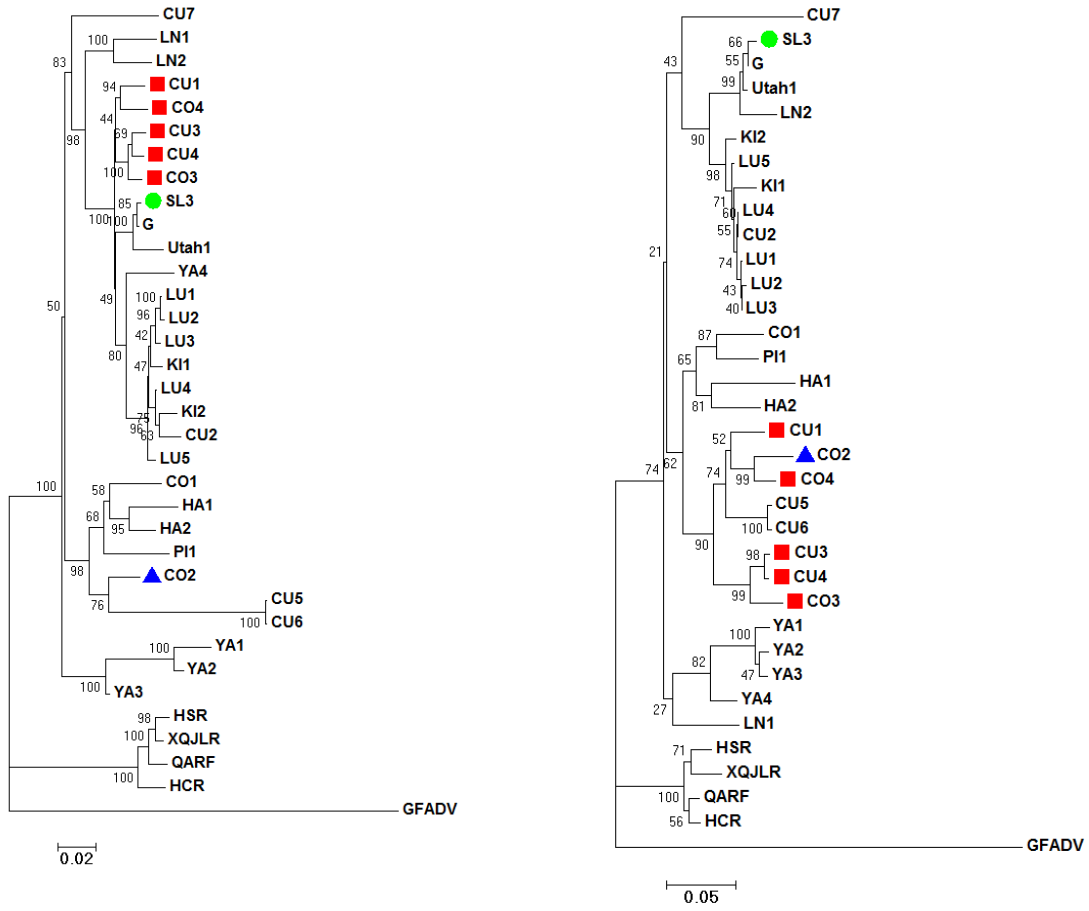


Figure A9. Neighbor Joining tree of the recombination event 9.

Neighbor Joining tree of the recombination event 9 which was constructed based on the region derived from the portion of alignment between the identified recombination region (major parent: nt 1 – 873 AND 1201 – 4215 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is CO4 (square). CU1, CU3, CU4 and CO3 are sequences with evidence of the same recombination (square). Major parent is SL3 (circle) and minor parent is CO2 (triangle). Scale below the tree is substitution per site.

Neighbor Joining tree of the recombination event 9 which was constructed based on the region derived from the identified recombination region in the alignment (minor parent: nt 874 - 1200 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is CO4 (square). CU1, CU3, CU4 and CO3 are sequences with evidence of the same recombination (square). Major parent is SL3 (circle) and minor parent is CO2 (triangle). Scale below the tree is substitution per site.

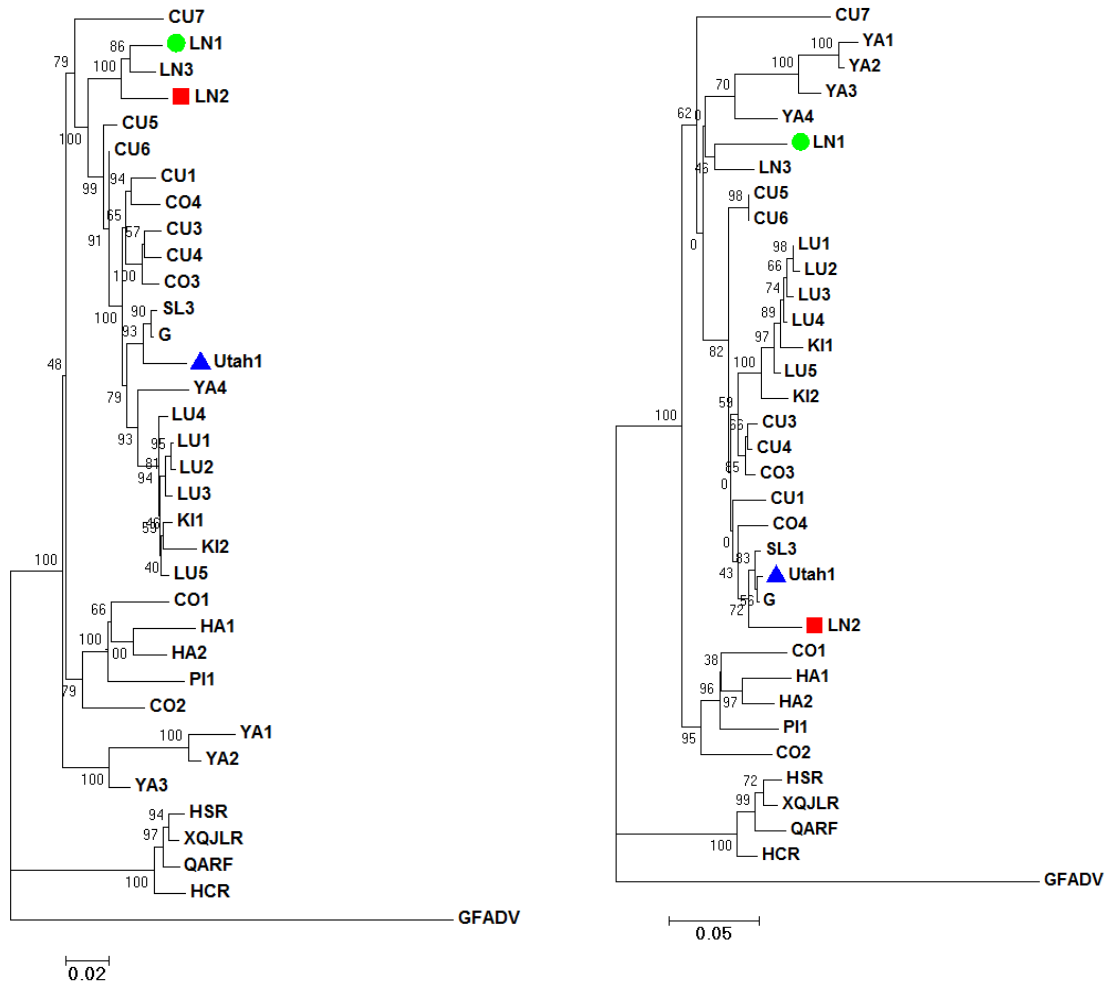


Figure A10. Neighbor Joining tree of the recombination event 10.

Neighbor Joining tree of the recombination event 10 which was constructed based on the region derived from the portion of alignment between the identified recombination region (major parent: nt 1 – 816 AND 1790 - 4215 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is LN2 (square), major parent is LN1 (circle) and minor parent is Utah1 (triangle). Scale below the tree is substitution per site.

Neighbor Joining tree of the recombination event 10 which was constructed based on the region derived from the identified recombination region in the alignment (minor parent: nt 817 - 1789 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is LN2 (square), major parent is LN1 (circle) and minor parent is Utah1 (triangle). Scale below the tree is substitution per site.

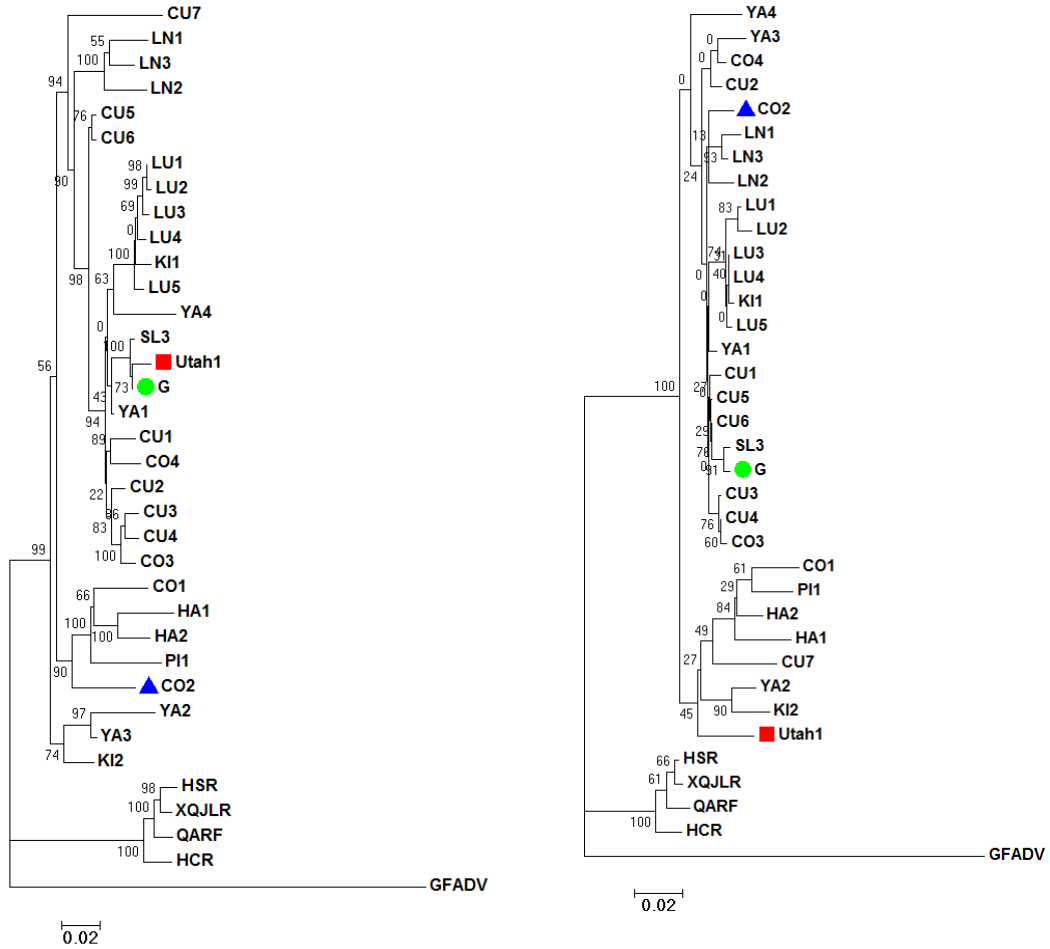


Figure A11. Neighbor Joining tree of the recombination event 11.

Neighbor Joining tree of the recombination event 11 which was constructed based on the region derived from the portion of alignment between the identified recombination region (major parent: nt 1 – 3610 AND 4015 – 4215 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is Utah1 (square), major parent is G (circle) and minor parent is Unknown (CO2) (triangle). Scale below the tree is substitution per site.

Neighbor Joining tree of the recombination event 11 which was constructed based on the region derived from the identified recombination region in the alignment (minor parent: nt 3611 - 4014 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is Utah1 (square), major parent is G (circle) and minor parent is Unknown (CO2) (triangle). Scale below the tree is substitution per site.

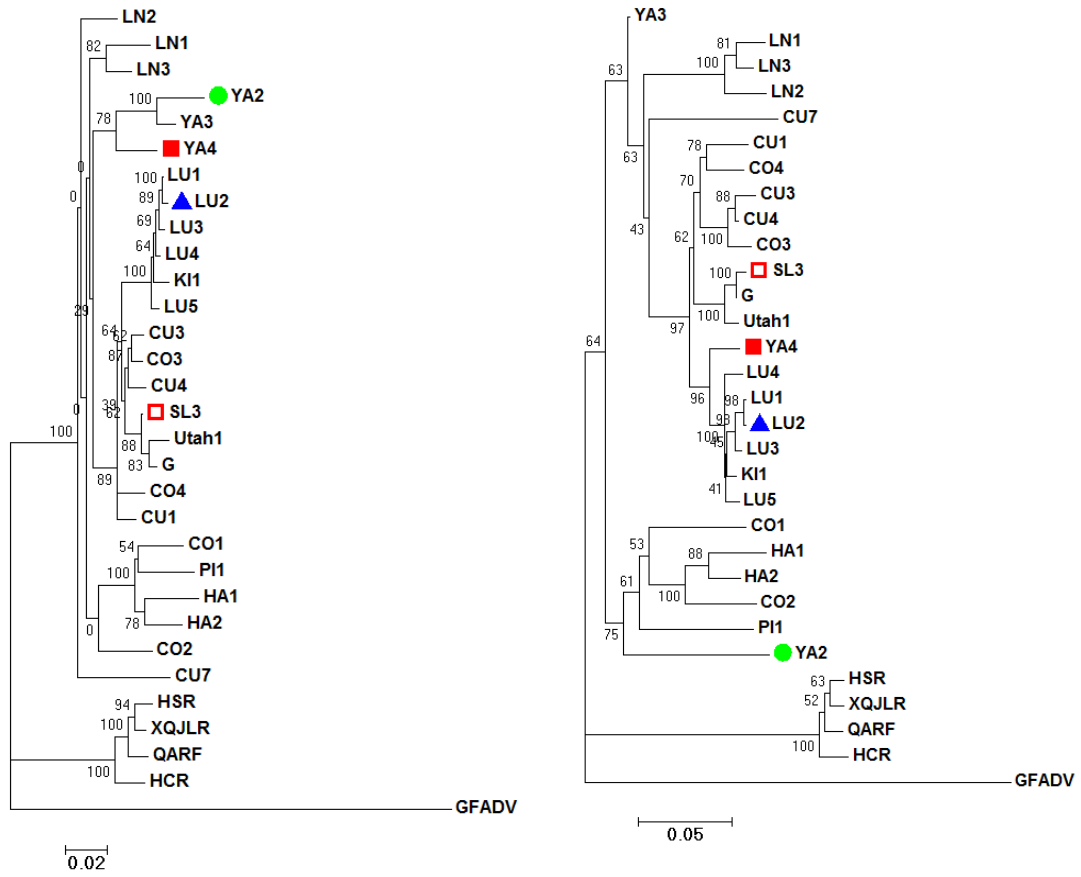


Figure A12. Neighbor Joining tree of the recombination even 12.

Neighbor Joining tree of the recombination even 12 which was constructed based on the region derived from the portion of alignment between the identified recombination region (major parent: nt 810 - 4047 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is YA4 (filled square). SL3 is a sequence with partial evidence of the same recombination event (square). Major parent is YA2 (circle) and minor parent is LU2 (triangle). Scale below the tree is substitution per site.

Neighbor Joining tree of the recombination event 12 which was constructed based on the region derived from the identified recombination region in the alignment (minor parent: nt 1 – 809 and 4048 - 4215 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is YA4 (filled square). SL3 is a sequence with partial evidence of the same recombination event (square). Major parent is YA2 (circle) and minor parent is LU2 (triangle). Scale below the tree is substitution per site.

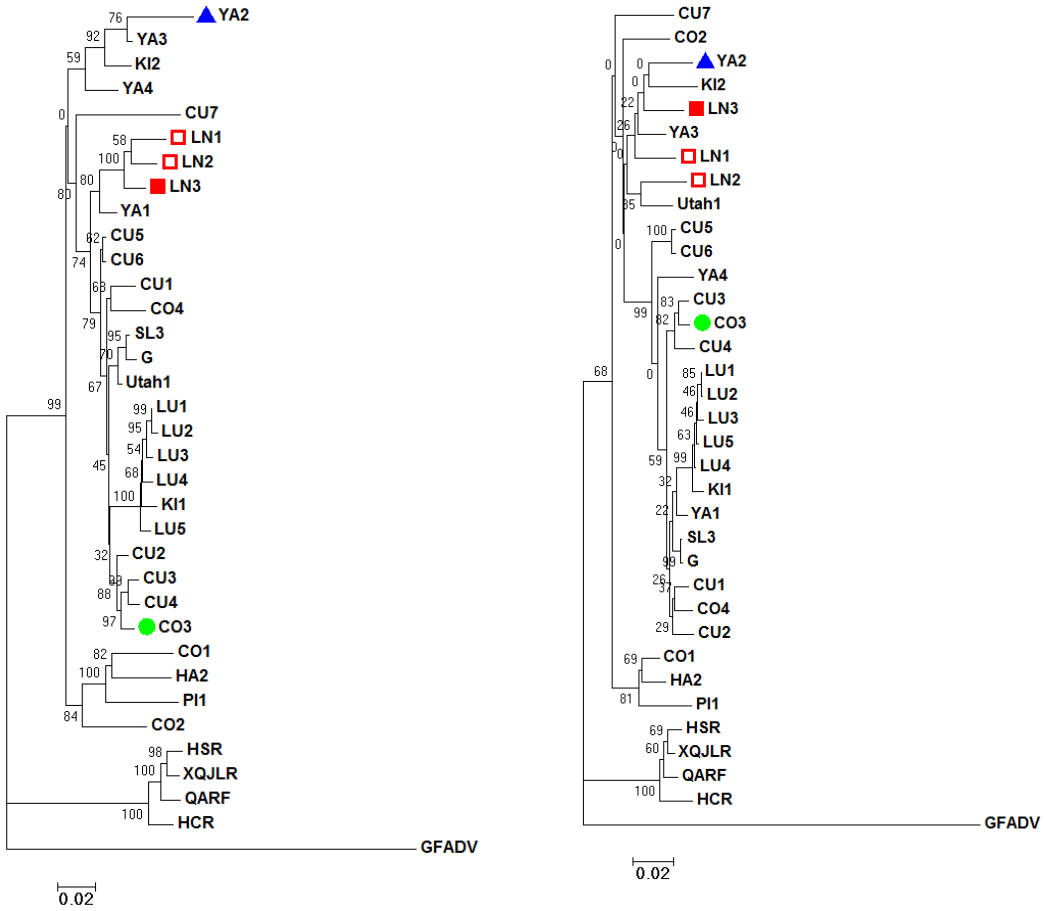


Figure A13. Neighbor Joining tree of the recombination event 13.

Neighbor Joining tree of the recombination event 13 which was constructed based on the region derived from the portion of alignment between the identified recombination region (major parent: nt 1 – 1975 and 3033 – 4215 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is LN3 (square). LN2 and LN1 sequences with partial and trace evidence of the same recombination event, respectively (square). Major parent is CO3 (circle) and minor parent is YA2 (triangle). Scale below the tree is substitution per site.

Neighbor Joining tree of the recombination event 13 which was constructed based on the region derived from the identified recombination region in the alignment (minor parent: nt 1976 - 3032 of the alignment). Tree is constructed by RDP4. Recombinant regions were removed from alignment. Recombinant sequence is LN3 (square). LN2 and LN1 sequences with partial and trace of the same recombination event, respectively (square). Major parent is CO3 (circle) and minor parent is YA2 (triangle). Scale below the tree is substitution per site.

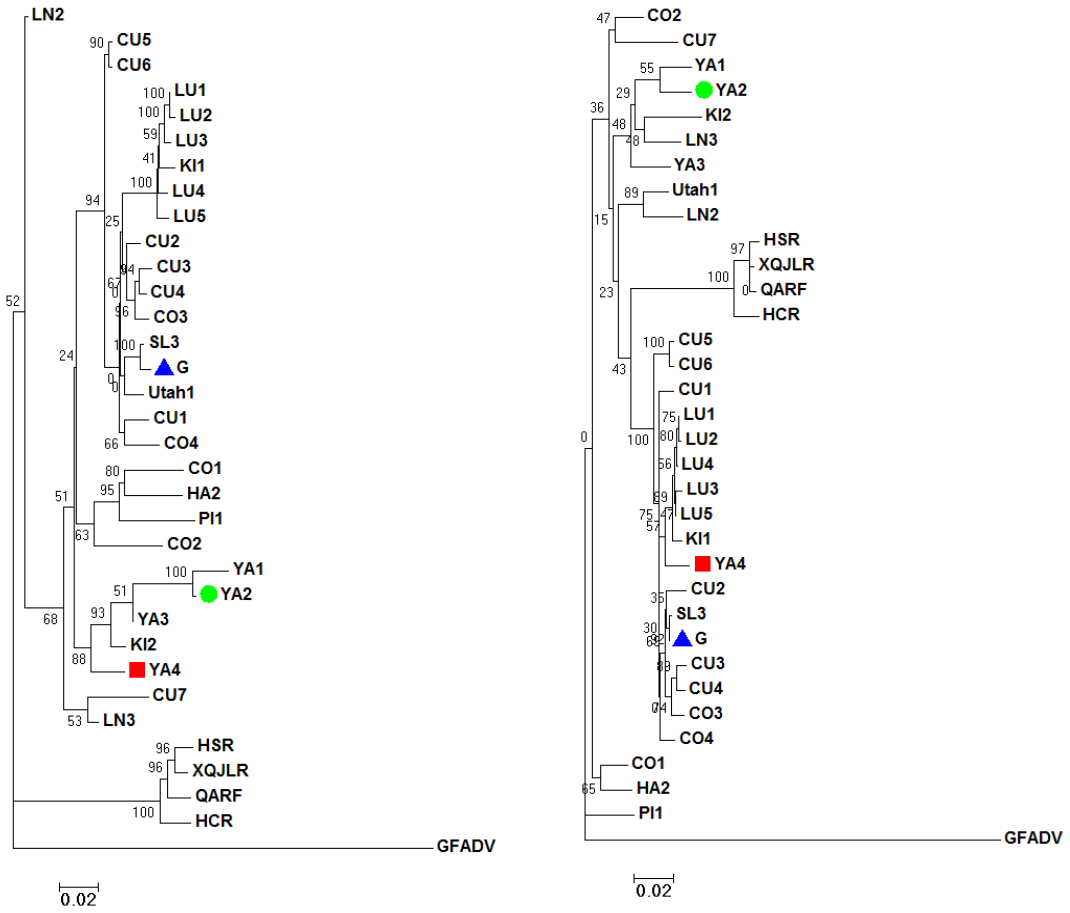


Figure A14. Neighbor Joining tree of the recombination event 14.

Neighbor Joining tree of the recombination event 14 which was constructed based on the region derived from the portion of alignment between the identified recombination region (major parent: nt 1 – 2314 and 3159 - 4215 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is YA4 (square), major parent is YA2 (circle) and minor parent is G (triangle). Scale below the tree is substitution per site.

Neighbor Joining tree of the recombination event 14 which was constructed based on the region derived from the identified recombination region in the alignment (minor parent: nt 2315 - 3158 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is YA4 (square), major parent is YA2 (circle) and minor parent is G (triangle). Scale below the tree is substitution per site.

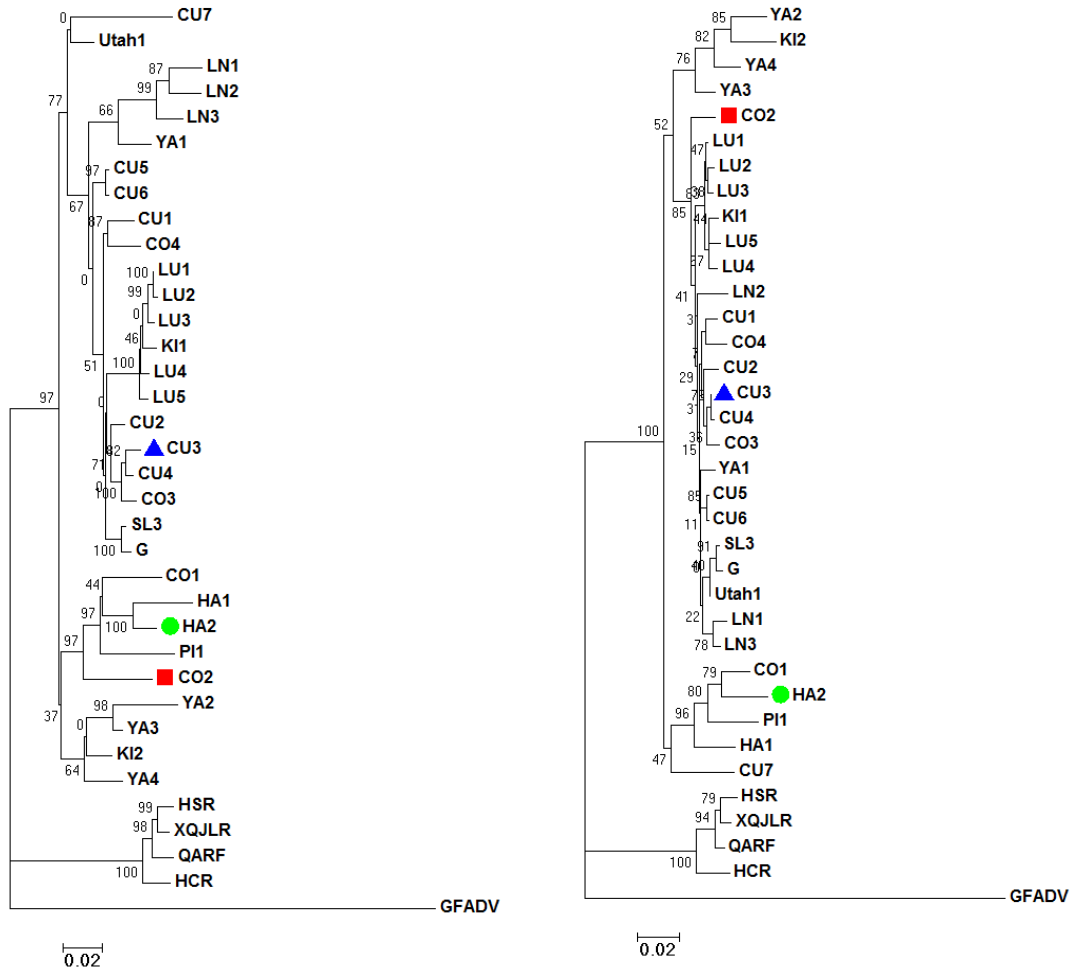


Figure A15. Neighbor Joining tree of the recombination event 15.

Neighbor Joining tree of the recombination event 15 which was constructed based on the region derived from the portion of alignment between the identified recombination region (major parent: nt 1 – 3233 and 3939 - 4215 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is CO2 (square), major parent is HA2 (circle) and minor parent is CU3 (triangle). Scale below the tree is substitution per site.

Neighbor Joining tree of the recombination event 15 which was constructed based on the region derived from the identified recombination region in the alignment (minor parent: nt 3234 - 3938 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is CO2 (square), major parent is HA2 (circle) and minor parent is CU3 (triangle). Scale below the tree is substitution per site.

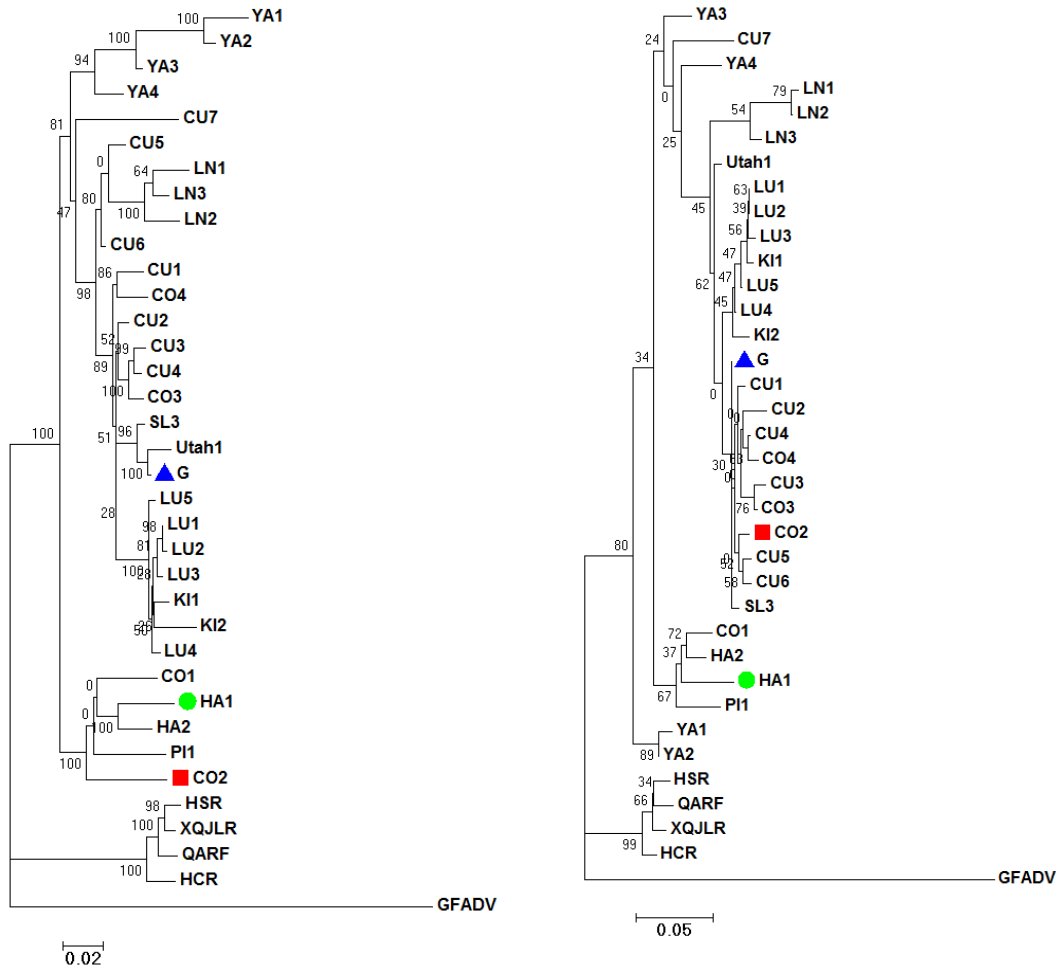


Figure A16. Neighbor Joining tree of the recombination event 16.

Neighbor Joining tree of the recombination event 16 which was constructed based on the region derived from the portion of alignment between the identified recombination region (major parent: nt 1 – 1771 and 1990 - 4215 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is CO2 (square), major parent is HA1 (circle) and minor parent is G (triangle). Scale below the tree is substitution per site.

Neighbor Joining tree of the recombination event 16 which was constructed based on the region derived from the identified recombination region in the alignment (minor parent: nt 1772 - 1989 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is CO2 (square), major parent is HA1 (circle) and minor parent is G (triangle). Scale below the tree is substitution per site.

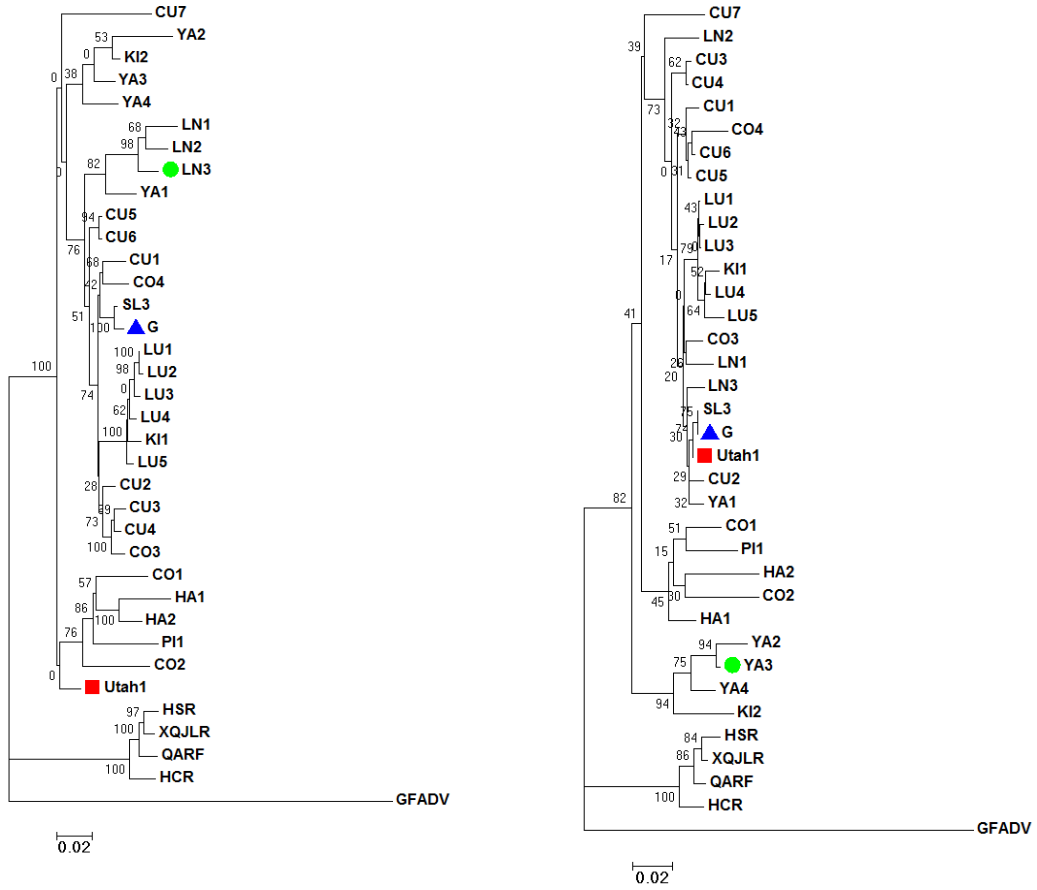


Figure A17. Neighbor Joining tree of the recombination event 17.

Neighbor Joining tree of the recombination event 17 which was constructed based on the region derived from the portion of alignment between the identified recombination region (major parent: nt 1 – 3035 and 3450 - 4215 of the alignment). Recombinant regions were removed from alignment. Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is Utah1 (square), major parent is Unknown (YA3) (circle) and minor parent is G (triangle). Scale below the tree is substitution per site.

Neighbor Joining tree of the recombination event 17 which was constructed based on the region derived from the identified recombination region in the alignment (minor parent: nt 3036 - 3449 of the alignment). Recombinant regions were removed from alignment. Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is Utah1 (square), major parent is Unknown (YA3) (circle) and minor parent is G (triangle). Scale below the tree is substitution per site.

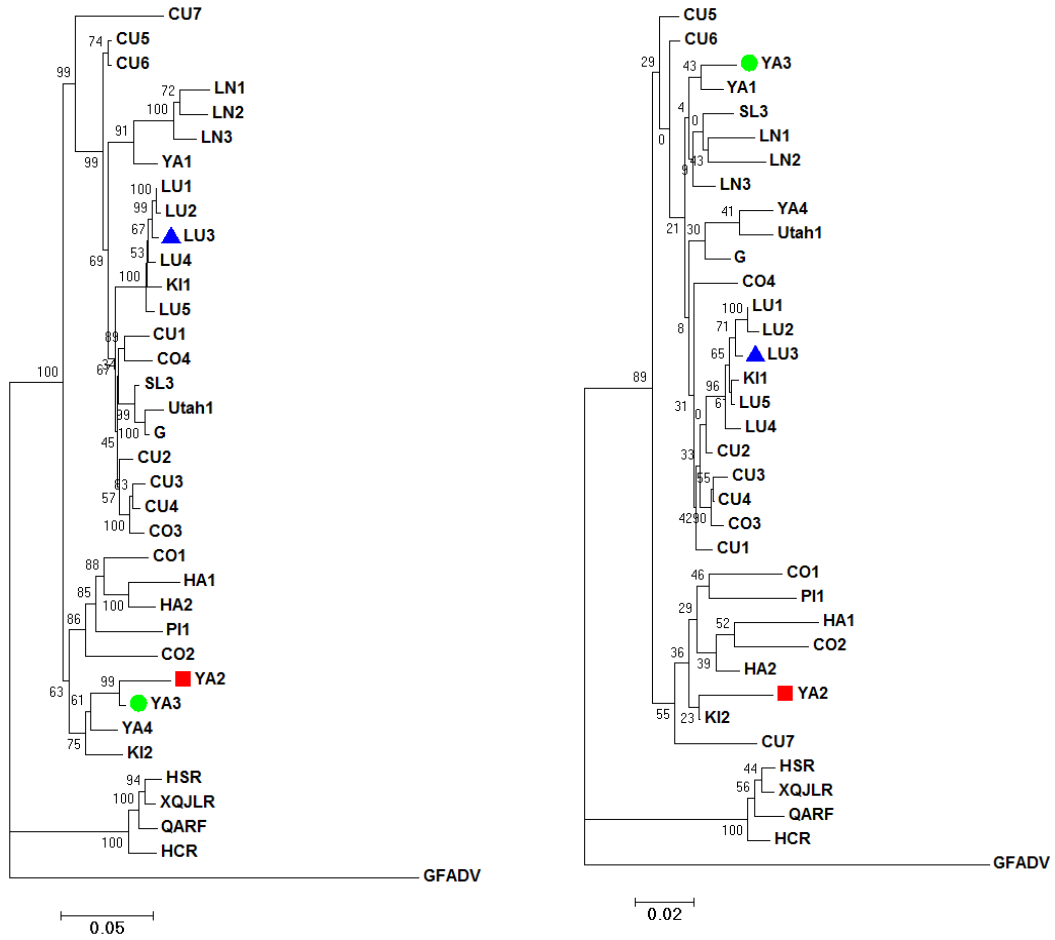


Figure A18. Neighbor Joining tree of the recombination event 18.

Neighbor Joining tree of the recombination event 18 which was constructed based on the region derived from the portion of alignment between the identified recombination region (major parent: nt 136 - 3773 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is YA2 (square), major parent is YA3 (circle) and minor parent is Unknown (LU3) (triangle). Scale below the tree is substitution per site.

Neighbor Joining tree of the recombination event 18 which was constructed based on the region derived from the identified recombination region in the alignment (minor parent: nt 1 – 135 and 3774 - 4215 of the alignment). Recombinant regions were removed from alignment. Tree is constructed by RDP4. Recombinant sequence is YA2 (square), major parent is YA3 (circle) and minor parent is Unknown (LU3) (triangle). Scale below the tree is substitution per site.

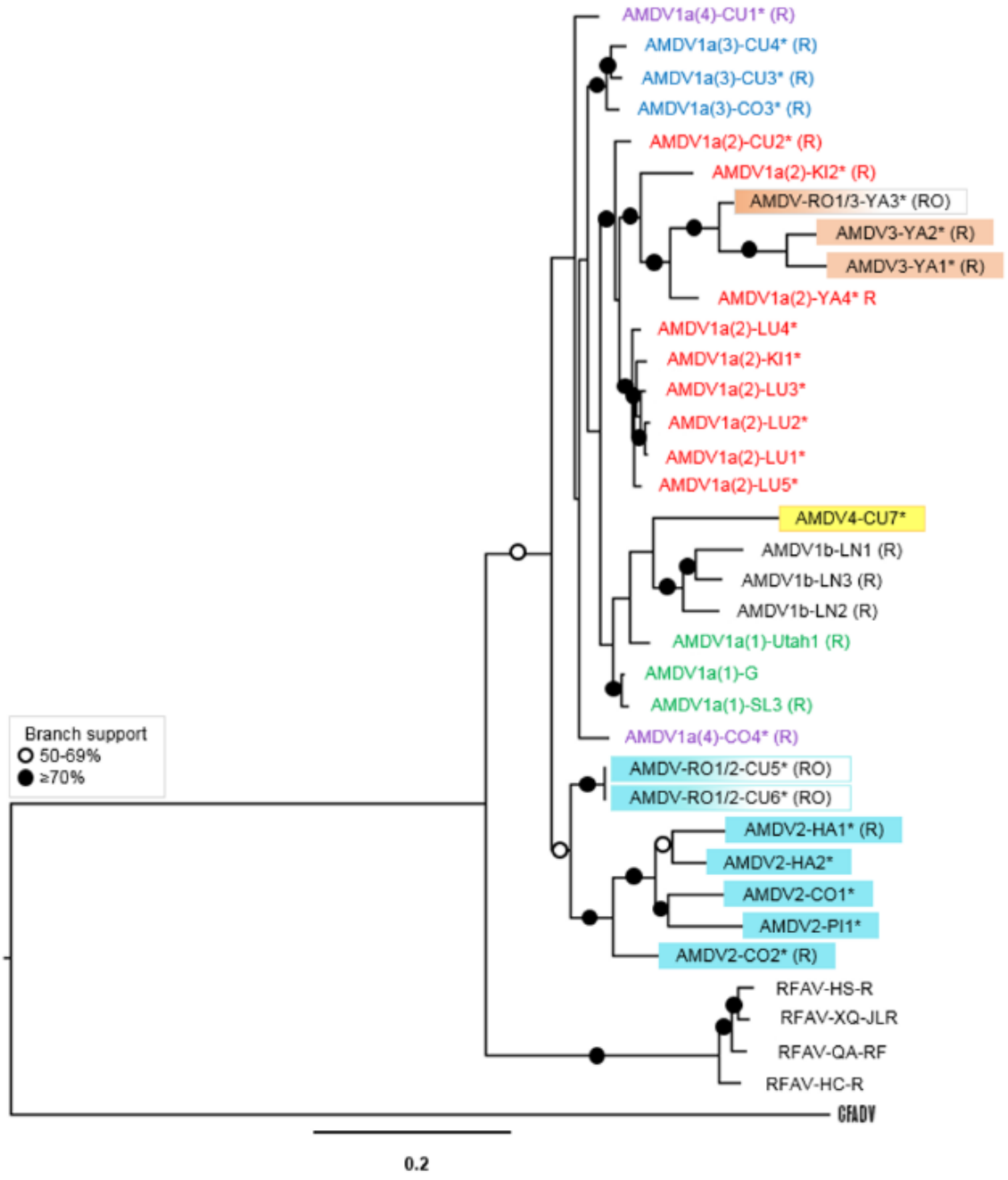


Figure A19. ML mid-point rooted phylogenetic analysis of the entire coding region *Amdoparvoviruses*, with the replacement of ambiguous codes with gap character. Refer to Figure 4.6 for description of sequences, symbols and colors.

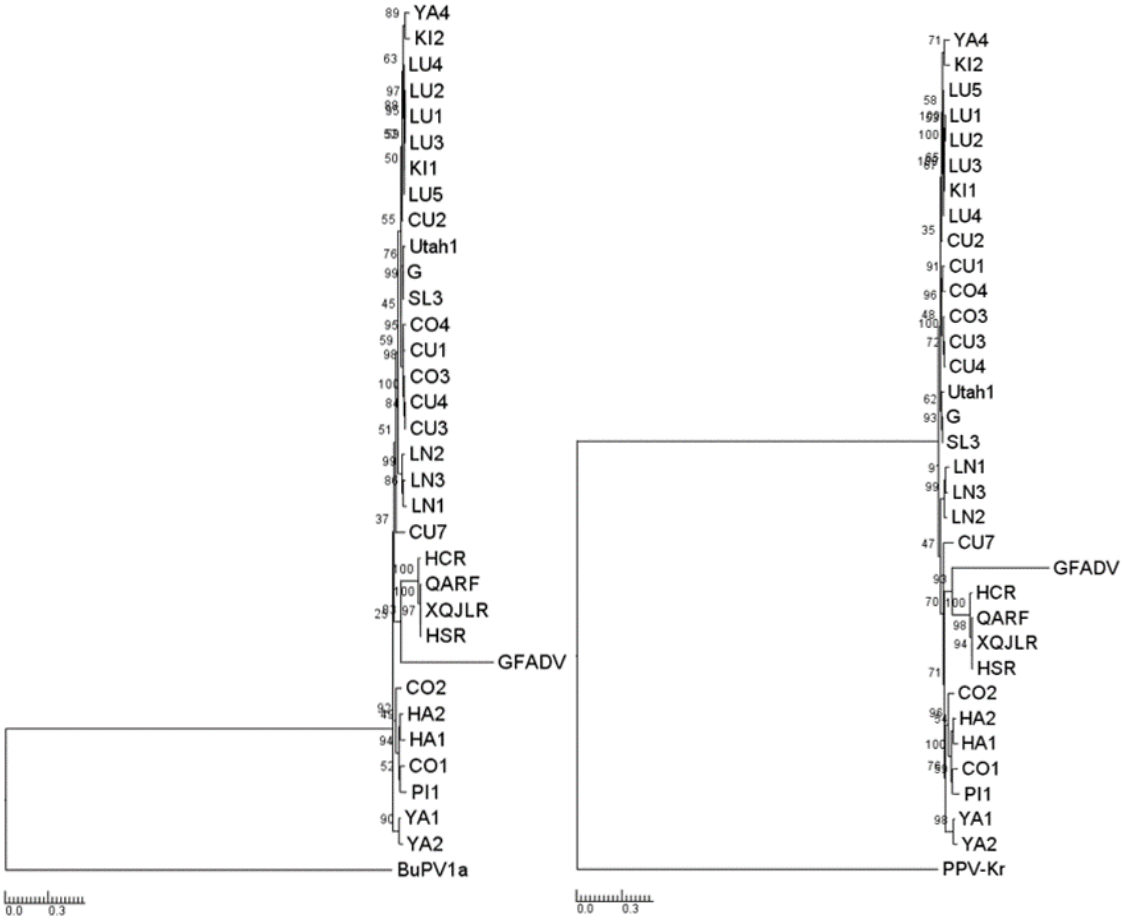


Figure A20. Rooted phylogenetic analysis of the genus *Amdoparvovirus*, using outgroups belonging to the genus *Protoparvovirus,* resulting in a *s*imilar topology with unrooted analysis.
Refer to Figure 4.6 for description on analysis method, sequences, symbols and colors. Outgroups were BuPV1a (accession number JX027296) and PPV-Kr (accession number U44978).

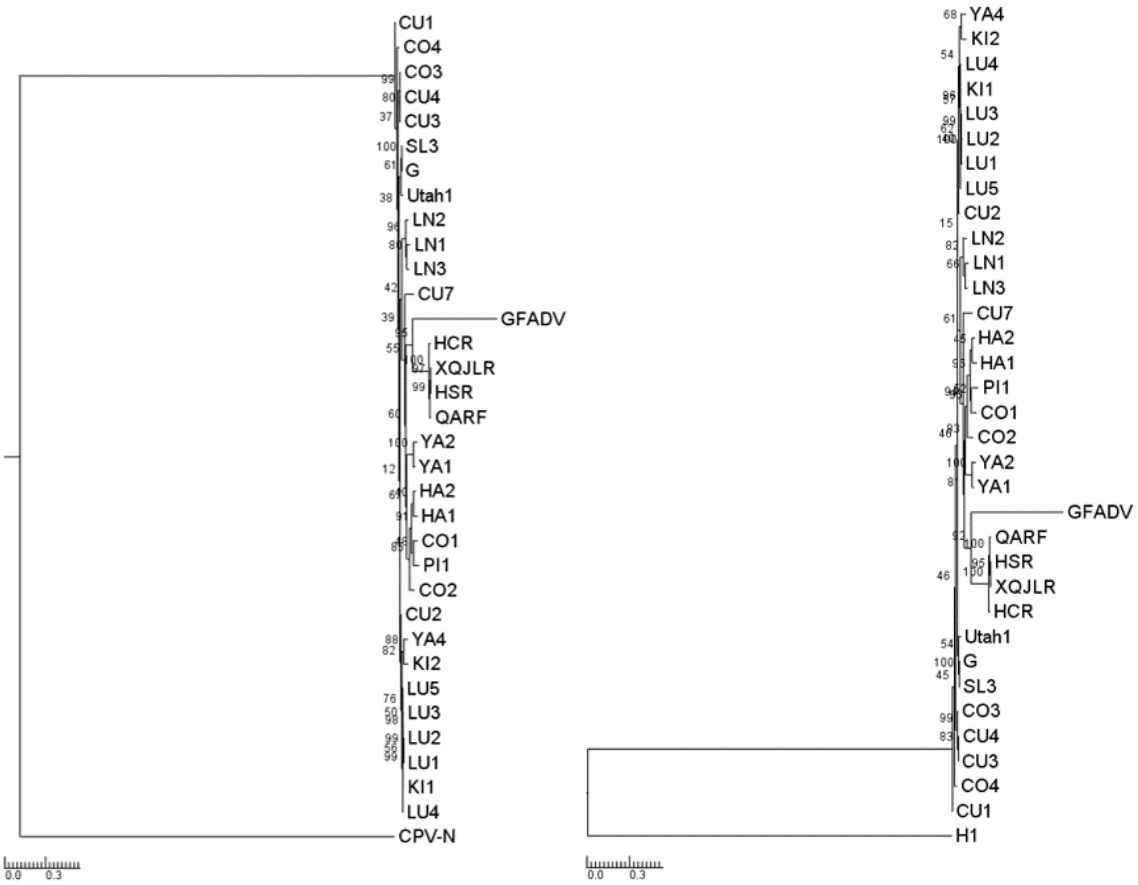


Figure A21. Rooted phylogenetic analysis of the genus *Amdoparvovirus*, using outgroups belonging to the genus *Protoparvovirus,* resulting in *a* different topology with unrooted analysis.
Refer to Figure 4.6 for description on analysis method, sequences, symbols and colors. Outgroups were CPV (accession number M19296) and H1 (accession number X01457).

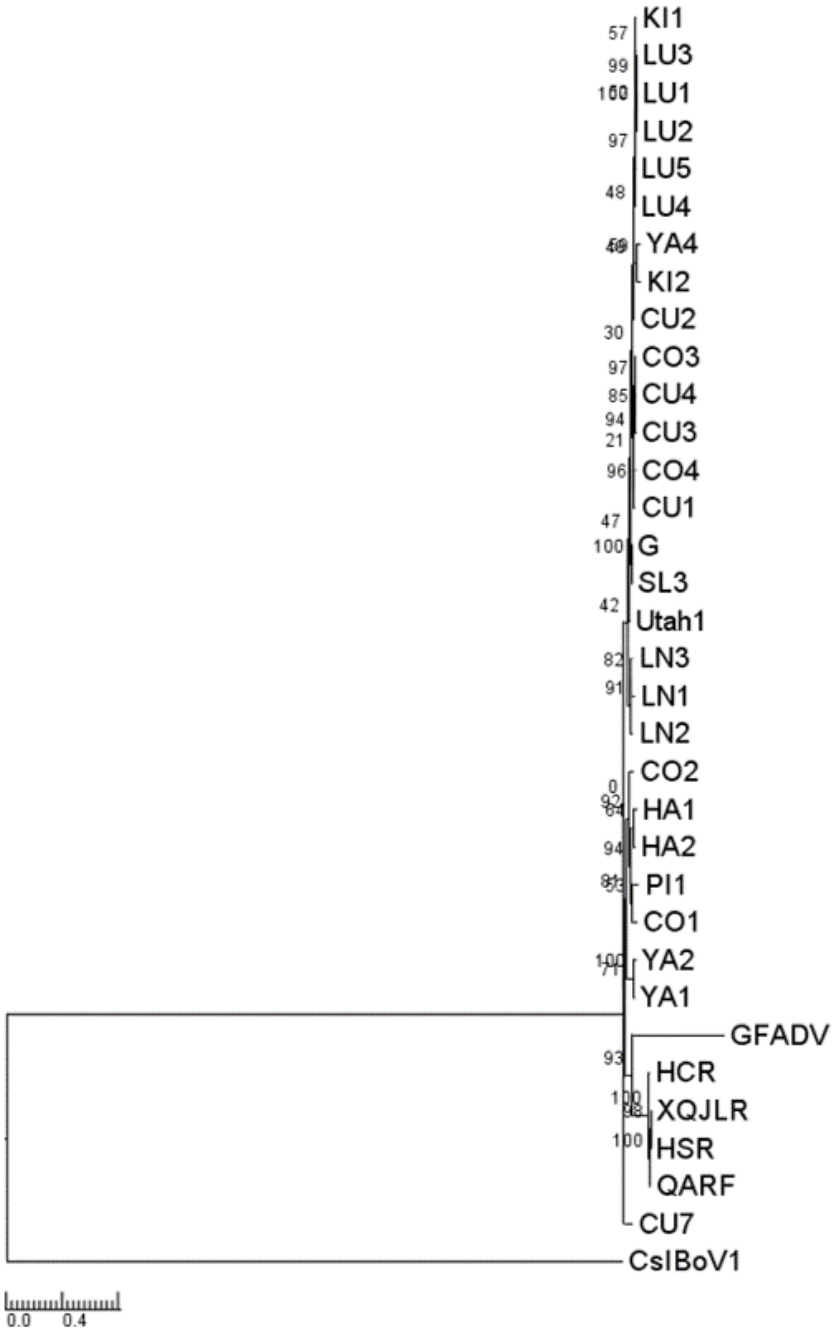


Figure A22. Rooted phylogenetic analysis of the genus *Amdoparvovirus*, using outgroups belonging to the genus *Bocaparvovirus,* resulting in a similar topology with unrooted analysis.
Refer to Figure 4.6 for description on analysis method, sequences, symbols and colors. Outgroups was CslBoV1 (accession number JN420361).

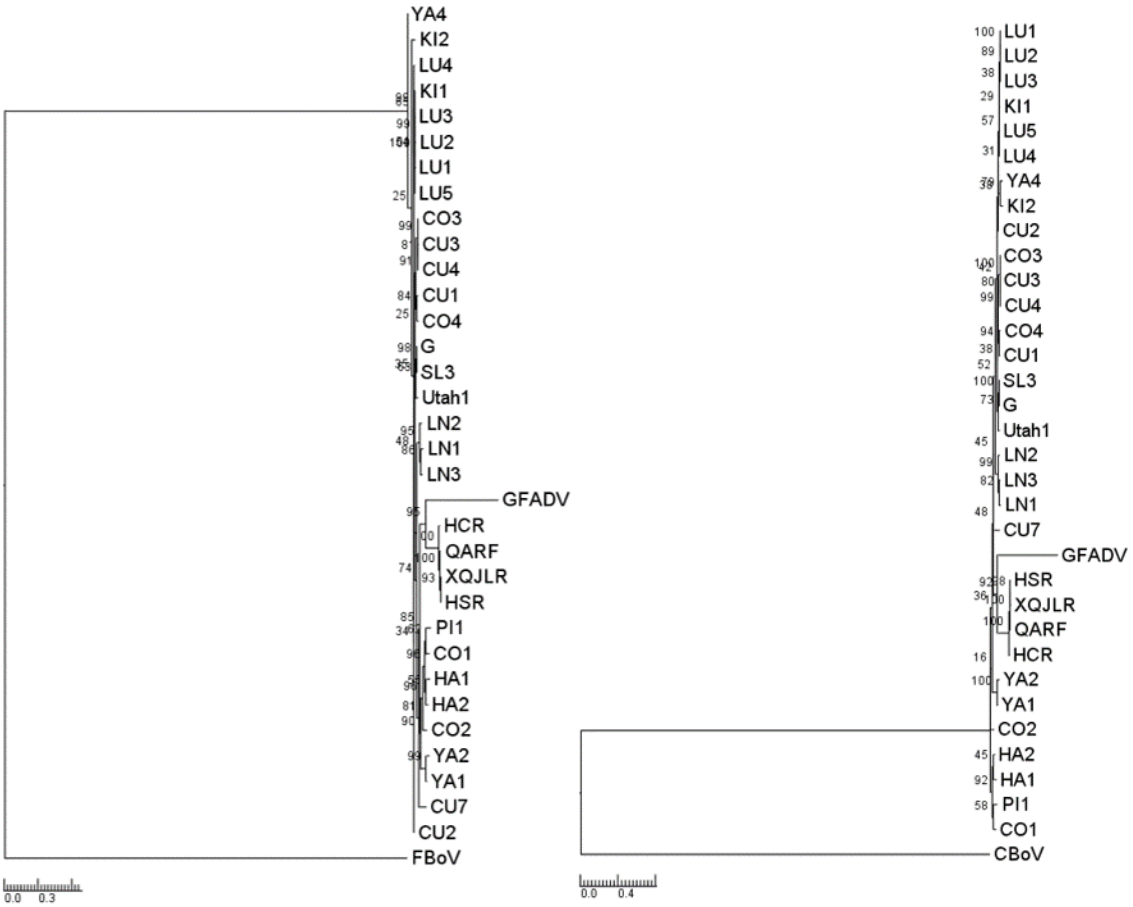


Figure A23. Rooted phylogenetic analysis of the genus *Amdoparvovirus*, using outgroups belonging to the genus *Bocaparvovirus,* resulting in *a* different topology with unrooted analysis.
Refer to Figure 4.6 for description on analysis method, sequences, symbols and colors. Outgroups were FBoV (accession number JQ692585) and CBoV (accession number JN648103).

BIBLIOGRAPHY

Aasted, B., Alexandersen, S., & Christensen, J. (1998). Vaccination with Aleutian mink disease parvovirus (AMDV) capsid proteins enhances disease, while vaccination with the major non-structural AMDV protein causes partial protection from disease. Vaccine, 16(11–12), 1158–1165. https://doi.org/10.1016/S0264-410X(98)80114-X

Alexandersen, S. (1986). Acute interstitial pneumonia in mink kits: experimental reproduction of the disease. Veterinary Pathology, 23(5), 579–88. http://www.ncbi.nlm.nih.gov/pubmed/3022453

Alexandersen, S., Bloom, M. E., & Perryman, S. (1988). Detailed transcription map of Aleutian mink disease parvovirus. Journal of Virology, 62(10), 3684–94. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC253511/

Alizon, S., de Roode, J. C., & Michalakis, Y. (2013). Multiple infections and the evolution of virulence. Ecology Letters, 16(4), 556–567. https://doi.org/10.1111/ele.12076

Ané, C. (2011). Detecting phylogenetic breakpoints and discordance from genome-wide alignments for species tree reconstruction. Genome Biology and Evolution, 3(1), 246–258. https://doi.org/10.1093/gbe/evr013

Belyi, V. A., Levine, A. J., & Skalka, A. M. (2010). Sequences from Ancestral Single-Stranded DNA Viruses in Vertebrate Genomes: the Parvoviridae and Circoviridae Are More than 40 to 50 Million Years Old. Journal of Virology, 84(23), 12458–12462. https://doi.org/10.1128/JVI.01789-10

Best, S. M., Shelton, J. F., Pompey, J. M., Wolfinbarger, J. B., & Bloom, M. E. (2003). Caspase Cleavage of the Nonstructural Protein NS1 Mediates Replication of Aleutian Mink Disease Parvovirus. Journal of Virology, 77(9), 5305–12. https://doi.org/10.1128/JVI.77.9.5305

Bloom, M. E., Alexandersen, S., Garon, C. F., Mori, S., Wei, W., Perryman, S., & Wolfinbarger, J. B. (1990). Nucleotide sequence of the 5'-terminal palindrome of Aleutian mink disease parvovirus and construction of an infectious molecular clone. Journal of Virology, 64(7), 3551–3556. https://doi.org/10.1097/MPG.0b013e318227ad6e

Bloom, M. E., Alexandersen, S., Perryman, S., Lechner, D., & Wolfinbarger, J. B. (1988). Nucleotide sequence and genomic organization of Aleutian mink disease parvovirus (ADV): sequence comparisons between a nonpathogenic and a pathogenic strain of ADV. Journal of Virology, 62(8), 2903–15. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=253728&tool=pmcentrez&rendertype=abstract

Bloom, M. E., Berry, B. D., Wei, W., Perryman, S., & Wolfinbarger, J. B. (1993). Characterization of chimeric full-length molecular clones of Aleutian mink disease parvovirus (ADV): identification of a determinant governing replication of ADV in cell culture. Journal of Virology, 67(10), 5976–88. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=238019&tool=pmcentrez&rendertype=abstract

Bloom, M. E., Best, S. M., Hayes, S. F., Wells, R. D., Wolfinbarger, J. B., McKenna, R., & Agbandje-McKenna, M. (2001). Identification of aleutian mink disease parvovirus capsid sequences mediating antibody-dependent enhancement of infection, virus neutralization, and immune complex formation. Journal of Virology, 75(22), 11116–27. https://doi.org/10.1128/JVI.75.22.11116-11127.2001

Bloom, M. E., Fox, J. M., Berry, B. D., Oie, K. L., & Wolfinbarger, J. B. (1998). Construction of pathogenic molecular clones of Aleutian mink disease parvovirus that replicate both in vivo and in vitro. Virology, 251(2), 288–96. https://doi.org/10.1006/viro.1998.9426

Bloom, M. E., Kanno, H., Mori, S., & Wolfinbarger, J. B. (1994). Aleutian mink disease: puzzles and paradigms. Infectious Agents and Disease, 3(6), 279-301. http://www.ncbi.nlm.nih.gov/pubmed/7889316

Boni, M. F., Posada, D., & Feldman, M. W. (2007). An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics, 176(2), 1035–1047. https://doi.org/10.1534/genetics.106.068874

Bordes, F., & Morand, S. (2011). The impact of multiple infections on wild animal hosts: a review. Infection Ecology & Epidemiology, 1(0), 1–10. https://doi.org/10.3402/iee.v1i0.7346

Bowman, J., Kidd, A, Gorman, R., & Schultehostedde, A. (2007). Assessing the potential for impacts by feral mink on wild mink in Canada. Biological Conservation, 139(1–2), 12–18. https://doi.org/10.1016/j.biocon.2007.05.020

Brown, J. K., Zerbini, F. M., Navas-Castillo, J., Moriones, E., Ramos-Sobrinho, R., Silva, J. C. F., Fiallo-Olivé, E., Briddon, R. W., Hernández-Zepeda, C., Idris, A., Malathi, V. G., Martin, D. P., Rivera-Bustamante, R., Ueda, S., & Varsani, A. (2015). Revision of Begomovirus taxonomy based on pairwise sequence comparisons. Archives of Virology, 160(6), 1593–1619. https://doi.org/10.1007/s00705-015-2398-y

Canuti, M., O'Leary, K. E., Hunter, B. D., Spearman, G., Ojkic, D., Whitney, H. G., & Lang, A. S. (2016). Driving forces behind the evolution of the Aleutian mink disease parvovirus in the context of intensive farming. Virus Evolution, 2(1), vew004. https://doi.org/10.1093/ve/vew004

Canuti, M., Whitney, H. G., & Lang, A. S. (2015). Amdoparvoviruses in small mammals: Expanding our understanding of parvovirus diversity, distribution, and pathology. Frontiers in Microbiology, 6(OCT), 1–9. https://doi.org/10.3389/fmicb.2015.01119

Carrillo, C., Tulman, E. R., Delhon, G., Lu, Z., Carreno, A., Vagnozzi, A., Kutish, G. F., & Rock, D. L. (2005). Comparative Genomics of Foot-and-Mouth Disease Virus. Journal of Virology, 79(10), 6487–6504. https://doi.org/10.1128/JVI.79.10.6487

Cheng, F., Chen, A. Y., Best, S. M., Bloom, M. E., Pintel, D., & Qiu, J. (2010). The capsid proteins of Aleutian mink disease virus activate caspases and are specifically cleaved during infection. Journal of Virology, 84(6), 2687–96. https://doi.org/10.1128/JVI.01917-09

Cho, H. J., & Greenfield, J. (1978). Eradication of aleutian disease of mink by eliminating positive counterimmunoelectrophoresis test reactors. Journal of Clinical Microbiology, 7(1), 18–22. https://www.ncbi.nlm.nih.gov/pubmed/203601

Christensen, J., Pedersen, M., Aasted, B., & Alexandersen, S. (1995). Purification and characterization of the major nonstructural protein (NS-1) of Aleutian mink disease parvovirus. Journal of Virology, 69(3), 1802–9. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=188788&tool=pmcentrez&rendertype=abstract

Christensen, J., Storgaard, T., Bloch, B., Alexandersen, S., & Aasted, B. (1993). Expression of Aleutian mink disease parvovirus proteins in a baculovirus vector system. Journal of Virology, 67(1), 229–38. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=237356&tool=pmcentrez&rendertype=abstract

Christensen, L. S., Gram-Hansen, L., Chriél, M., & Jensen, T. H. (2011). Diversity and stability of Aleutian mink disease virus during bottleneck transitions resulting from eradication in domestic mink in Denmark. Veterinary Microbiology, 149(1–2), 64–71. https://doi.org/10.1016/j.vetmic.2010.10.016

Clemens, D. L., Wolfinbarger, J. B., Mori, S., Berry, B. D., Hayes, S. F., & Bloom, M. E. (1992). Expression of Aleutian mink disease parvovirus capsid proteins by a recombinant vaccinia virus: self-assembly of capsid proteins into particles. Journal of Virology, 66(5), 3077–85. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=241069&tool=pmcentrez&rendertype=abstract

Cornish-Bowden, A. (1986). Nomenclature for incompletely specified bases in nucleic acid sequences, recommendations. Nucleic Acids Research, 14(6), 2707–2720. https://www.ncbi.nlm.nih.gov/pubmed/2582368

Cotmore, S. F., Agbandje-McKenna, M., Chiorini, J. a, Mukha, D. V, Pintel, D. J., Qiu, J., Soderlund-Venermo, M., Tattersall, P., Tijssen, P., Gatherer, D., & Davison, A. J. (2014). The family Parvoviridae. Archives of Virology, 159(5), 1239–47. https://doi.org/10.1007/s00705-013-1914-1

Duffy, S., Shackelton, L. A, & Holmes, E. C. (2008). Rates of evolutionary change in viruses: patterns and determinants. Nature Reviews. Genetics, 9(4), 267–76. https://doi.org/10.1038/nrg2323

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Farid, A. H. (2013). Aleutian mink disease virus in furbearing mammals in Nova Scotia, Canada. Acta Veterinaria Scandinavica, 55(1), 10. https://doi.org/10.1186/1751-0147-55-10

Farid, A. H. & Ferns, L. E. (2011). Aleutian mink disease virus infection may cause hair depigmentation. Scientifur. 35, 55-59.

Farid, A. H., & Rupasinghe, P. P. (2014). Source of the Aleutian mink disease virus outbreak in Nova Scotia. http://www.perennia.ca/wp-content/uploads/2016/09/V1-Outbreak-2013.pdf

Farid, A. H., Zillig, M. L., Finley, G. G., & Smith, G. C. (2012). Prevalence of the Aleutian mink disease virus infection in Nova Scotia, Canada. Preventive Veterinary Medicine, 106(3), 332–8. https://doi.org/10.1016/j.prevetmed.2012.03.010

Felsenstein, J. (2004). Inferring Phylogenies (1st ed.). Sunderland, Mass: Sinauer Associates. https://doi.org/10.1086/383584

Gibbs, M. J., Armstrong, J. S., & Gibbs, A. J. (2000). Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics (Oxford, England), 16(7), 573–582. https://doi.org/10.1093/bioinformatics/16.7.573

Gottschalck, E., Alexandersen, S., Cohn, A., & Poulsen, L. A., Bloom, M. E., & Aasted, B. (1991). Nucleotide sequence analysis of Aleutian mink disease parvovirus shows that multiple virus types are present in infected mink. Journal of virology, 65(8), 4378-4386. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC248877/

Gottschalck, E., Alexandersen, S., Storgaard, T., Bloom, M. E., & Aasted, B. (1994). Sequence comparison of the non-structural genes of four different types of Aleutian mink disease parvovirus indicates an unusual degree of variability. Archives of Virology, 138(3–4), 213–31. https://doi.org/10.1007/BF01379127

Haas, L., Wohlsein, P., Trautwein, G. et al., (1990) Violet mink develop and acute disease after experimental infection with Aleutian disease virus (ADV) isolate ADV SL3. Journal of Veterinary Medicine, 37(1-10), 106-17. https://www.ncbi.nlm.nih.gov/pubmed/2163578

Hadlow, W. J., Race, R. E., & Kennedy, R. C. (1983). Comparative pathogenicity of four strains of Aleutian disease virus for pastel and sapphire Comparative Pathogenicity of Four Strains of Aleutian Disease Virus for Pastel and Sapphire Mink. Infection and Immunity, 41(3), 1016. https://www.ncbi.nlm.nih.gov/pubmed/6193063

Hagberg, E. E., Krarup, A., Fahnøe, U., Larsen, L. E., Dam-Tuxen, R., & Pedersen, A. G. (2016). A fast and robust method for whole genome sequencing of the Aleutian Mink Disease Virus (AMDV) genome. Journal of Virological Methods, 234, 43–51. https://doi.org/10.1016/j.jviromet.2016.03.010

Hall, B. G., & Barlow, M. (2006). Phylogenetic analysis as a tool in molecular epidemiology of infectious diseases. Annals of Epidemiology, 16(3), 157–169. https://doi.org/10.1016/j.annepidem.2005.04.010

Huang, Q., Luo, Y., Cheng, F., Best, S. M., Bloom, M. E., & Qiu, J. (2014). Molecular characterization of the small nonstructural proteins of parvovirus Aleutian mink disease virus (AMDV) during infection. Virology, 452–453, 23–31. https://doi.org/10.1016/j.virol.2014.01.005

Hungnes O., Jonassen, C. M., & Grinde, B. (2000). Molecular epidemiology of viral infections, Apmis, 108(2), 81–97. https://www.ncbi.nlm.nih.gov/pubmed/10737453

Jahns, H., Daly, P., McElroy, M. C., Sammin, D. J., Bassett, H. F., & Callanan, J. J. (2010). Neuropathologic Features of Aleutian Disease in Farmed Mink in Ireland and Molecular Characterization of Aleutian Mink Disease Virus Detected in Brain Tissues. Journal of Veterinary Diagnostic Investigation, 22(1), 101–105. https://doi.org/10.1177/104063871002200120

Jensen, T. H., Christensen, L. S., Chriél, M., Harslund, J., Salomonsen, C. M., & Hammer, A. S. (2012). High prevalence of Aleutian mink disease virus in free-ranging mink on a remote Danish island. Journal of Wildlife Diseases, 48(2), 497–502. https://doi.org/10.7589/0090-3558-48.2.497

Kidd, A. G., Bowman, J., Lesbarrères, D., & Schulte-Hostedde, A. I. (2009). Hybridization between escaped domestic and wild American mink (Neovison vison). Molecular Ecology, 18(6), 1175–86. https://doi.org/10.1111/j.1365-294X.2009.04100.x

King, A. M. Q., Lefkowitz, E. J., Adams, M. J., & Carstens, E. B. (2011). Virus Taxonomy: Classification and Nomenclature of Viruses. Ninth Report of the International Committee on Taxonomy of Viruses, 9, 405–425. https://doi.org/10.1016/B978-0-12-384684-6.00039-2

Knuuttila, A., Aaltonen, K., Virtala, A. M. K., Henttonen, H., Isomursu, M., Leimann, A., Maran, T., Saarma, U., Timonen, P., Vapalahti, O., & Sironen, T. (2015). Aleutian mink disease virus in free-ranging mustelids in Finland - a cross-sectional epidemiological and phylogenetic study. Journal of General Virology, 96(6), 1423–35. https://doi.org/10.1099/vir.0.000081

Knuuttila, A., Uzcátegui, N., Kankkonen, J., Vapalahti, O., & Kinnunen, P. (2009). Molecular epidemiology of Aleutian mink disease virus in Finland. Veterinary Microbiology, 133(3), 229–38. https://doi.org/10.1016/j.vetmic.2008.07.003

Lake, J. L., Ryba, S. A., Serbst, J., Brown, C. F., & Gibson, L. (2007). Mercury and Stable Isotopes of Carbon and Nitrogen in Mink. Environmental Toxicology and Chemistry, 26(12), 2611. https://doi.org/10.1897/06-607.1

Lam, T. T.-Y., Hon, C.-C., & Tang, J. W. (2010). Use of phylogenetics in the molecular epidemiology and evolutionary studies of viral infections. Critical reviews in clinical laboratory sciences, 47(1), 5-49. https://doi.org/10.3109/10408361003633318

Leimann, A., Knuuttila, A., Maran, T., Vapalahti, O., & Saarma, U. (2015). Molecular epidemiology of Aleutian mink disease virus (AMDV) in Estonia, and a global phylogeny of AMDV. Virus Research, 199, 56–61. https://doi.org/10.1016/j.virusres.2015.01.011

Lemey, P., Salemi, M., & Vandamme, A.-M. (2009). The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing (2nd ed.). New York: Cambridge University Press.

Li, L., Pesavento, P. A., Woods, L., Clifford, D. L., Luff, J., Wang, C., & Delwart, E. (2011). Novel amdovirus in gray foxes. Emerging Infectious Diseases, 17(10), 1876–1878. https://doi.org/10.3201/eid1710.110233

Li, Y., Huang, J., Jia, Y., Du, Y., Jiang, P., & Zhang, R. (2012). Genetic characterization of Aleutian mink disease viruses isolated in China. Virus Genes, 45(1), 24–30. https://doi.org/10.1007/s11262-012-0733-x

Lukashov, V. V, & Goudsmit, J. (2001). Evolutionary relationships among parvoviruses: virus-host coevolution among autonomous primate parvoviruses and links between adeno-associated and avian parvoviruses. Journal of Virology, 75(6), 2729–40. https://doi.org/10.1128/JVI.75.6.2729-2740.2001

MacLachlann, N. J., & Dubovi, E. J. (2011). Fenner's Veterinary Virology, Academic press. https://doi.org/DOI:10.1016/B978-0-12-375158-4.00012-2

Mañas, S., Ceña, J. C., Ruiz-Olmo, J., Palazón, S., Domingo, M., Wolfinbarger, J. B., & Bloom, M. E. (2001). Aleutian mink disease parvovirus in wild riparian carnivores in Spain. Journal of Wildlife Diseases, 37(1), 138–44. https://doi.org/10.7589/0090-3558-37.1.138

Martin, D. P. (2009). Recombination Detection and Analysis Using RDP3. Methods in Molecular Biology, 537, 185–205. https://doi.org/10.1007/978-1-59745-251-9

Martin, D. P., Biagini, P., Lefeuvre, P., Golden, M., Roumagnac, P., & Varsani, A. (2011). Recombination in eukaryotic single stranded DNA viruses. Viruses, 3(9), 1699–1738. https://doi.org/10.3390/v3091699

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evolution, 1(1), vev003-vev003. https://doi.org/10.1093/ve/vev003

Martin, D., & Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. Bioinformatics, 16(6), 562–563. https://doi.org/10.1093/bioinformatics/16.6.562

McCormack, G. P., & Clewley, J. P. (2002). The application of molecular phylogenetics to the analysis of viral genome diversity and evolution. Reviews in Medical Virology, 12(4), 221–38. https://doi.org/10.1002/rmv.355

Muhire, B. M., Varsani, A., & Martin, D. P. (2014). SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation. PLoS ONE, 9(9), e108277. https://doi.org/10.1371/journal.pone.0108277

Muhire, B., Martin, D. P., Brown, J. K., Navas-Castillo, J., Moriones, E., Zerbini, F. M., Rivera-Bustamante, R., Malathi, V. G., Briddon, R. W., & Varsani, A. (2013). A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus Mastrevirus (family Geminiviridae). Archives of Virology, 158(6), 1411–1424. https://doi.org/10.1007/s00705-012-1601-7

Nituch, L. A., Bowman, J., Wilson, P., & Schulte-Hostedde, A. I. (2012). Molecular epidemiology of Aleutian disease virus in free-ranging domestic, hybrid, and wild mink. Evolutionary Applications, 5(4), 330–340. https://doi.org/10.1111/j.1752-4571.2011.00224.x

Oie, K. L., Durrant, G., Wolfinbarger, J. B., Martin, D., Costello, F., Perryman, S., Hogan, D., Hadlow, W. J., & Bloom, M. E. (1996). The relationship between capsid protein (VP2) sequence and pathogenicity of Aleutian mink disease parvovirus (ADV): a possible role for raccoons in the transmission of ADV infections. Journal of Virology, 70(2), 852–61. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=189888&tool=pmcentrez&rendertype=abstract

Olofsson, A, Mittelholzer, C., Treiberg Berndtsson, L., Lind, L., Mejerland, T., & Belák, S. (1999). Unusual, high genetic diversity of Aleutian mink disease virus. Journal of Clinical Microbiology, 37(12), 4145–9. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=85904&tool=pmcentrez&rendertype=abstract

Padidam, M., Sawyer, S., & Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. Virology, 265(2), 218–225. https://doi.org/10.1006/viro.1999.0056

Pérez-Losada, M., Arenas, M., Galán, J. C., Palero, F., & González-Candelas, F. (2015). Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. Infection, Genetics and Evolution, 30, 296–307. https://doi.org/10.1016/j.meegid.2014.12.022

Persson, S., Jensen, T. H., Blomström, A.-L., Appelberg, M. T., & Magnusson, U. (2015). Aleutian Mink Disease Virus in Free-Ranging Mink from Sweden. PLOS ONE, 10(3), e0122194. https://doi.org/10.1371/journal.pone.0122194

Posada, D., & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proceedings of the National Academy of Sciences of the United States of America, 98(24), 13757–62. https://doi.org/10.1073/pnas.241370698

Pourkarim, M. R., Amini-Bavil-Olyaee, S., Kurbanov, F., Van Ranst, M., & Tacke, F. (2014). Molecular identification of hepatitis B virus genotypes/ subgenotypes: Revised classification hurdles and updated resolutions. World Journal of Gastroenterology, 20(23), 7152–7168. https://doi.org/10.3748/wjg.v20.i23.7152

Qiu, J., Cheng, F., Burger, L. R., & Pintel, D. (2006). The transcription profile of Aleutian mink disease virus in CRFK cells is generated by alternative processing of pre-mRNAs produced from a single promoter. Journal of Virology, 80(2), 654–62. https://doi.org/10.1128/JVI.80.2.654-662.2006

Salminen, M. O., Carr, J. K., Burke, D. S., & McCutchan, F. E. (1995). Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. AIDS Research and Human Retroviruses, 11(11), 1423–1425. https://doi.org/10.1089/aid.1995.11.1423

Sang, Y., Ma, J., Hou, Z., & Zhang, Y. (2012). Phylogenetic analysis of the VP2 gene of Aleutian mink disease parvoviruses isolated from 2009 to 2011 in China. Virus Genes, 45(1), 31–7. https://doi.org/10.1007/s11262-012-0734-9

Schuierer, S., Bloom, M. E., Kaaden, O. R., Truyen, U., (1997). Sequence analysis of the lymphotropic Aleutian disease parvovirus ADV-SL3. Archives of Virology, 142(1), 157–166. https://www.ncbi.nlm.nih.gov/pubmed/9155880

Shackelton, L. a., Hoelzer, K., Parrish, C. R., & Holmes, E. C. (2007). Comparative analysis reveals frequent recombination in the parvoviruses. Journal of General Virology, 88(12), 3294–3301. https://doi.org/10.1099/vir.0.83255-0

Shao, X.-Q., Wen, Y.-J., Ba, H.-X., Zhang, X.-T., Yue, Z.-G., Wang, K.-J., Li, C.-Y., Qiu, J., & Yang, F.-H. (2014). Novel amdoparvovirus infecting farmed raccoon dogs and arctic foxes. Emerging Infectious Diseases, 20(12), 2085–8. https://doi.org/10.3201/eid2012.140289

Shen, D. T., Gorham, J. R., Harwood, R. F., & Padgett, G. A. (1973). The persistence of Aleutian disease virus in the mosquito Aedes fitchii. Archiv Für Die Gesamte Virusforschung, 40(3), 375–81. http://www.ncbi.nlm.nih.gov/pubmed/4633583

Shi, W., Zhu, C., Zheng, W., Zheng, W., Ling, C., Carr, M. J., Higgins, D. G., & Zhang, Z. (2012). Subgenotyping of Genotype C Hepatitis B Virus: Correcting Misclassifications and Identifying a Novel Subgenotype. PLoS ONE, 7(10), 1–8. https://doi.org/10.1371/journal.pone.0047271

Silvestro, D., & Michalak, I. (2012). RaxmlGUI: A graphical front-end for RAxML. Organisms Diversity and Evolution, 12(4), 335–337. https://doi.org/10.1007/s13127-011-0056-0

Simmonds, P. (2015). Methods for virus classification and the challenge of incorporating metagenomic sequence data. Journal of General Virology, 96(6), 1193–1206. https://doi.org/10.1099/vir.0.000016

Smith, J. M. (1992). Analyzing the mosaic structure of genes. Journal of Molecular Evolution, 34(2), 126–129. https://doi.org/10.1007/BF00182389

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Statistics Canada, (2014). CANSIM, Table number 003-0014. Number and value of mink pelts produced, by color type, annual. 2014. Available: http://www.statcan.gc.ca/pub/23-013-x/2010001/related-connexes-eng.htm

Steinel, A, Parrish, C. R., Bloom, M. E., & Truyen, U. (2001). Parvovirus infections in wild carnivores. Journal of Wildlife Diseases, 37(3), 594–607. https://doi.org/10.7589/0090-3558-37.3.594

Streck, A. F., Bonatto, S. L., Homeier, T., Souza, C. K., Gonçalves, K. R., Gava, D., Canal, C. W., & Truyen, U. (2011). High rate of viral evolution in the capsid protein of porcine parvovirus. Journal of General Virology, 92(11), 2628–36. https://doi.org/10.1099/vir.0.033662-0

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution, 30(12), 2725–2729. https://doi.org/10.1093/molbev/mst197

Troesch, M., Meunier, I., Lapierre, P., Lapointe, N., Alvarez, F., Boucher, M., & Soudeyns, H. (2006). Study of a novel hypervariable region in hepatitis C virus (HCV) E2 envelope glycoprotein. Virology, 352(2), 357–367. https://doi.org/10.1016/j.virol.2006.05.015

Varsani, A., Martin, D. P., Navas-Castillo, J., Moriones, E., Hernández-Zepeda, C., Idris, A., Murilo Zerbini, F., & Brown, J. K. (2014a). Revisiting the classification of curtoviruses based on genome-wide pairwise identity. Archives of Virology, 159(7), 1873–1882. https://doi.org/10.1007/s00705-014-1982-x

Varsani, A., Navas-Castillo, J., Moriones, E., Hernandez-Zepeda, C., Idris, A., Brown, J. K., Murilo Zerbini, F., & Martin, D. P. (2014b). Establishment of three new genera in the family Geminiviridae: Becurtovirus, Eragrovirus and Turncurtovirus. Archives of Virology, 159(8), 2193–2203. https://doi.org/10.1007/s00705-014-2050-2

Xi, J., Wang, J., Yu, Y., Zhang, X., Mao, Y., Hou, Q., & Liu, W. (2016). Genetic characterization of the complete genome of an Aleutian mink disease virus isolated in north China. Virus Genes, (52), 1–11. https://doi.org/10.1007/s11262-016-1320-3

Yates, D. E., Mayack, D. T., Munney, K., Evers, D. C., Major, A., Kaur, T., & Taylor, R. J. (2005). Mercury levels in mink (Mustela vison) and river otter (Lontra canadensis) from northeastern North America. Ecotoxicology, 14(1–2), 263–274. https://doi.org/10.1007/s10646-004-6273-y