PHYLOGENETIC AND EVOLUTIONARY ANALYSIS OF THE *BORNA DISEASE VIRUS*

By

ELENA BLANK

Submitted in fulfillment of the academic requirements for the degree of Masters in Science in the School of Molecular and Cellular Biosciences in the discipline of Genetics, University of Natal

Pietermaritzburg

2002

PREFACE

The experiment work described in this dissertation was conducted at the University of Natal from January 2000 to December 2002, under the supervision of Professor Annabel Fossey.

The results are the results of my own investigation, except where the work of others is acknowledged in the text; and have not been submitted in any other form to another University.

E. Blank

Elena Blank

November 2002

I certify the above statement is correct.

Professor Annabel Fossey

Supervisor

ACKNOWLEDGEMENTS

I thank Professor Annabel Fossey, my supervisor, for her advice and support.

I thank Ms Colleen LeRoux for her help and support in ensuring that all relevant thesis documents reached my supervisor.

I gratefully acknowledge the financial support I received from the University of Natal.

I am indebted to my brother Nicolas Blank and my mother, Ms Ingrid Blank, for their unwavering love and support throughout my studies.

DEDICATION

This dissertation is dedicated to my son Joshua Alexander, born on the 7th of November 2002, my husband Tim and my mother Ingrid Blank, whose love and support made this work possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The characteristic trait of the *Borna disease virus* is that it is a complicated single negative stranded RNA virus that is capable of infecting a wide array of mammalian species including human beings. It has been implicated in a diverse variety of human neuropsychiatric diseases. The infection capability, mechanism of infection and range of protein action of this virus remain to be identified.

The purpose of the present study was to determine (1) whether the previous *Bornaviridae* family classification is indeed accurate as the action of BDV indicates that it is related to other viruses and (2) to estimate the number of synonymous (nucleotide substitution) and non-synonymous (amino acid change) evolutionary mutation rates of proteins (nucleoprotein, phosphoprotein, glycoprotein, matrix protein) exhibited by various *Borna disease virus* host species and the proteins (nucleoprotein, phosphoprotein, glycoprotein, matrix and X protein) of three *Borna disease virus* strains. The latter study would give an indication as to which proteins are subjected to positive selection.

Phylogenetic methods were used to determine the accuracy of the *Bornaviridae* classification. Phylogenetic trees obtained through an alignment and analysis of the polymerase protein, which displays a uniquely conserved GDN motif, of various RNA negative single stranded viruses using neighbourhood and parsimony methods enabled comparison with other RNA virus families.

A method adapted from Ina, (1995) for estimating the synonymous and non-synonymous evolutionary mutation rate was applied to various BDV proteins in order to provide more information on inter (host virus) and intra (virus) mutation rate. This information in turn was used to create an evolutionary model to clarify the positive and neutral evolutionary trend of the inter- and intra-virus proteins examined, which may help clarify and enhance the lack of current knowledge relating to species infection and the epidemiological nature of the virus.

The results obtained by the polymerase alignment analysis indicates the presence of two newly discovered BDV motifs, v and vi, confirmed by three diverse alignment programmes. An analysis of the alignment of BDV proteins indicated that the BDV nucleoprotein nuclear localization signal aligns the BDV nucleoprotein between motifs IV and vi of the BDV polymerase.

The results obtained by the phylogenetic analysis indicate that the *Rabies* virus and the *Vesicular stomatitis* virus are the most closely related animal viruses to BDV, whereas the *Rice transitory yellowing* (unclassified *Rhabdovirus*) and *Sonchus yellow net* plant virus are closer to BDV than

other animal *Rhabdoviridae* raising intriguing questions on the evolutionary origins of the *Borna disease virus*. The phylogenetic analysis indicates that the *Borna disease virus* does not fall into a separate *Bornaviridae* family classification, and suggests that BDV may be more appropriately placed into a separate subfamily in the family *Rhabdoviridae*.

The results of the evolutionary analysis indicate considerable diversity between BDV host virus (inter-species) and BDV virus (intra-species) protein sequences. In the host virus sequence comparison analysis all of the proteins examined displayed a high pattern of non random evolution, which is in contrast to the intra species comparison in which only three proteins; the BDV glycoprotein, nucleoprotein and X protein; displayed a non random pattern of evolution. The positive selection effect displayed by the inter-species (host) proteins may be attributed to antigenic variation displayed by the inter-species sequences and a super infection hypothesis, which indicates that positive selection on host variants could arise during the course of an infection as a result of specific immune responses.

The positive and neutral selection trend of the proteins displayed by the intra-species (virus) sequences may be a result of a pattern of nucleotide substitution that is physio-chemically conservative. Conservation may be evident in volume, polarity, hydrophilicity, or molecular weight of amino acids of the proteins.

CHAPTER 1

INTRODUCTION

The Borna Disease Virus: A Negative Single Stranded RNA virus

1.1 INTRODUCTION

RNA viruses are a diverse group that infect prokaryotes as well as many eukaryotes, both plants and animals. The large group of RNA viruses includes highly prevalent human pathogens, such as *Respiratory Syncytial virus* (RSV), *Parainfluenza* viruses and influenza viruses, and two of the most deadly human pathogens *(Ebola* and *Marburg* viruses), as well as viruses with a major economic impact on the poultry and cattle industries, such as *Newcastle disease virus* (NDV) and *Rinderpest virus* (RPV).

1.2 CLASSIFICATION OF RNA VIRUSES

In general, most RNA viruses have single-stranded RNA as their genetic material, with the exception of one family the *Reoviridea* (Wickner, 1993), which has double stranded RNA. The study of RNA viruses has been hampered by the extreme virulence of many isolates, the difficulties inherent in direct sequencing of RNA, and the failure of most RNA viruses to undergo ready recombination. The advent of complementary DNA (cDNA) cloning and sequencing of RNA viral genomes as plasmid copies has solved many of the problems and has increased the potential for learning about RNA viruses. The recent development of cDNA clones from which infectious viruses can be recovered has made it possible to engineer specific mutations at will at the DNA level and rescue RNA viruses containing these changes and to make recombinant viruses; which has ushered in a whole new era of RNA virology.

RNA viruses have traditionally been divided into three main groups: the positive stranded viruses; the negative-stranded viruses, and the double stranded RNA viruses. These major groups have, in turn, been subdivided into virus families on the basis of virus structure, hosts and epidemiology. The determination of the nucleotide sequences of many viruses has resulted in detailed comparison

of viruses in terms of both sequence similarities and overall genome organisation and has revealed a number of striking homologies among disparate groups (Goldbach, 1987; Strauss and Strauss, 1983). These comparisons have facilitated compilation of the latest grouping of RNA viruses into a number of superfamilies, which differentiate the RNA viruses into negative single stranded, double strand and positive strand RNA viruses (Table 1.1).

Table 1.1 The current classification of RNA viruses.

SUPERFAMILIES	NUMBER of FAMLIES
Positive single stranded RNA viruses	**35**
Negative single stranded RNA viruses	**8**
Double stranded RNA viruses	**6**

These three superfamilies show different replication strategies:

Positive stranded RNA viruses:

1. Translation of virion RNA as mRNA.
2. Synthesis of (-) sense RNA on the (+) sense template by RNA dependent RNA polymerase (RDRP) which results in the formation of the replicative complex, (RC).
3. Synthesis of (+) sense RNA, mRNA and (-) sense RNA.
4. Translation of (+) sense RNA and mRNA synthesis of structural protein, which biases the replicative complex to produce (+) sense RNA.
5. Assembly of structural protein and (+) sense RNA and maturation of virions.

Negative single stranded RNA viruses:

1. Primary transcription of virion (-) sense RNA by RDRP in the virion core in the cytoplasm. Production of mainly mRNA and (+) sense RNA; formation of the replicative complex (RC).
2. Translation of the mRNAs; accumulation of products.
3. Virion proteins interact with RC and bias it towards production of full-length (+) sense RNA and therefore of genomic (-) sense RNA.
5. Secondary transcription from progeny (-) sense RNA, translation and accumulation of structural proteins.
4. Nucleocapsid assembly and maturation; budding of nucleocapsid through the host membrane which contains viral envelope proteins.

Double stranded RNA viruses:

1. Primary transcription in the virion core in the cytoplasm by viral RDRP, and export of (+) sense RNA to the cytoplasm.
2. Translation of (+) sense RNA, accumulation of viral proteins.
3. Assembly of (+) sense RNA and viral proteins into immature virions.
4. Transcription of (+) sense RNA into double stranded (ds) RNA in virions by viral RDRP.
5. Secondary transcription of double stranded RNA.
6. Final assembly and maturation of virions.

1.3 TAXONOMY OF THE NEGATIVE SINGLE STRANDED RNA VIRUSES

The International Committee on Virus Taxonomy (ICVT) has differentiated the genomes of RNA negative single stranded viruses into monopartite genome viruses, bipartite genome *arenaviruses,* tripartite genome *bunyaviruses,* the multipartite genome *orthomyxoviruses* and *tenuiviruses,* and the helper dependent defective hepatitis delta virus (Pringle and Easton, 1997). The known viruses with monopartite (single stranded) negative-sense RNA genomes are further classified into four families, namely, *Bornaviridae, Filoviridae, Paramyxoviridae and Rhabdoviridae.* These families embrace viruses of diverse biological characteristics, which share certain features of genome organization. In recognition of the common features of genome organization, the four families were grouped together as the order *Mononegavirales* (Pringle, 1991, 1997; Schneemann *et al.*, 1995), which was the first taxon above the family level to be recognized in virus taxonomy (Murphy *et al.*,

1995). The taxonomy of the monopartite, bipartite and multipartite negative single stranded RNA viruses is summarized in Table 1.2.

Table 1.2 Negative Single Stranded RNA viruses of the order *Mononegavirales.*

Order **Family** **Subfamily**	**Genus**	**Species**	**Host**
Mononegavirales			
Bornaviridae	Bornavirus	Borna disease virus	Vertebrates
Bornavirus			
Filoviridae	Ebola like viruses	Ebola virus	Vertebrates
	Marburg like viruses	Marburg virus	Vertebrates
Paramyxoviridae			
Paramyxovirinae		Human parainfluenza virus I	Vertebrates
	Respirovirus	Measles virus	Vertebrates
	Morbillivirus	Mumps virus	Vertebrates
Pneumovirinae	Rubulavirus	Human respiratory syncytial virus	Vertebrates
	Pneumovirus	Turkey rhinotracheitis virus	Vertebrates
	Metapneumovirus		
Rhabdoviridae		Vesicular stomatitis Indiana virus	Vertebrates
	Vesiculovirus	Rabies Virus	Vertebrates
	Lyssavirus	Bovine ephemeral fever virus	Vertebrates
	Ephemerovirus	Infectious haematopoietic necrosis	Vertebrates
	Novirhabdovirus	Lettuce necrotic yellow dwarf virus	Plants
	Cytorhabdovirus	Potato Yellow dwarf virus	Plants
	Nucleorhabdovirus	Sonchus Yellow net virus	Plants
	Unclassified	Viral haemorrhagic septicaemia	Vertebrates
		Rice transitory yellowing virus	Vertebrates
Orthomyxoviridae	InfluenzavirusA	Influenza A virus	Vertebrates
	InfluenzavirusB	Influenza B virus	Vertebrates
	Influenzavirus C	Influenza C virus	Vertebrates
	Thogotovirus	Thogoto virus	Vertebrates
Bunyaviridae	Bunyavirus	Bunyamwera virus	Vertebrates
	Hantavirus	Hantaan virus	Vertebrates
	Nairovirus	Nairobi Sheep disease virus	Vertebrates
	Phlebovirus	Sandfly fever Sicilian virus	Vertebrates
		Rift valley fever virus	Vertebrates
		Uukuniemi virus	Vertebrates
		Toscana virus	Vertebrates
	Tospovirus	Tomato spotted wilt virus	Plants
	Tenuivirus	Rice stripe virus	Plants
		Rice grassy stunt virus	Plants

	Opiovirus	Citrus psorosis virus	Plants
Arenaviridae	Arenavirus	Lymphocytic choriomeningitis	Vertebrates
	Deltavirus	Hepatitis delta virus	Vertebrates

The *Borna Disease virus* (BDV) therefore has been classified as part of the negative single stranded RNA viruses of the order *Mononegavirales*.

1.3.1 Features of the genomes of the order *Mononegavirales*.

The features which characterize the four families of the order *Mononegavirales* can be summarized as follows:

A linear monopartite form of the genome, similar gene order, complementarity of the 3' and 5' termini, a presumptive single 3' - terminal promoter, transcription by sequential interrupted synthesis, replication by synthesis of a complete positive-sense transcript, the virion associated RNA dependent RNA polymerase and maturation by budding.

Features which distinguish the four families are genome size, nucleocapsid structure, cellular site of genome replication and transcription, extent of mRNA processing, virion morphology, host range and several biological properties including tissue specificity and pathogenic potential. Table 1.3 provides a summary of the distinguishing features of the *Mononegavirales*.

Table 1.3 Distinguishing features of the order *Mononegevirales*.

Features	Families			
	Bornaviridae	*Filoviridae*	*Paramyxoviridae*	*Rhabdoviridae*
Absolute genome size	8.9 kb	19.1 kb	15.1-15.9 kb	11-15 kb
Virion Morphology	90 nm diameter spherical particle	Filamentous	Pleomorphic	Bullet-shaped/ Bacilliform
Site of Replication	Nucleus	Cytoplasm	Cytoplasm	Nucleus/Cytoplasm
Mode of transcription	Complex with mRNA splicing and overlapping start/stop signals	Polar with non-overlapping signals and stepwise attenuation	Polar with non-overlapping signals (except RSV)	Polar with non-overlapping signals and stepwise attenuation
Host range	Vertebrates	Primates	Vertebrates	Plants, invertebrates and vertebrates
Pathogenic potential	Immune mediated neurological disease	Haemorrhagic fever	Mainly respiratory disease	Mild febrile to fatal neurological disease

The families *Bornaviridae* and *Filoviridae* are each represented by a single genus, namely, *Bornavirus* and *Filovirus* respectively. The genus *Bornavirus* is defined by a single species, whereas the genus *Filovirus* consists of four species, which have been defined in terms of nucleotide sequence and antigenic divergence and a differential manner of expressing the attachment protein (G). The *Rhabdoviridae* comprises five genera, differentiated on the basis of host range, presence of supplementary genes and the intracellular site of virus replication. The family *Paramyxoviridae* is divided into two subfamilies, the *Paramyxovirinae* with three genera of viruses indicating a greater degree of homology to each other than to the viruses included in the other subfamily, the *Pneumovirinae* (Pringle, 1997).

1.4 *BORNA DISEASE VIRUS*

1.4.1 Pathogenesis

Borna disease in infected horses and sheep is characterized by agitated aggressive behaviour that progresses over weeks to paralysis and inanition (Hatalski *et al.*, 1997). Rats infected as adults exhibit hyperactivity and exaggerated startle responses coincident with viral gene products in limbic system neurons and infiltration of mononuclear cells into the brain (Narayan *et al.*, 1983; Carbone *et al.*, 1987). The inflammation recedes over several weeks, but the virus persists and animals show stereotyped motor behaviour, dyskenesias and dystomas associated with distinct changes in the central nervous dopamine system (Solbrig *et al.*, 1996a; Solbrig *et al.*, 1996b), as well as decreased activity and cachexia (Narayan *et al.*, 1983). In contrast, rats infected as neonates have a disease characterized by stunted growth, hyperactivity, subtle learning preferences and altered taste preferences and do not mount a cellular immune response to the virus (Dittrich *et al.*, 1989; Carbone *et al.*, 1991). Behavioural disturbances have been reported in primates, tree shrews and rhesus monkeys. Infected tree shrews have altered social and asexual behaviour, manifested as abnormal dominance relationships and failure to mate (Hatalski *et al.*, 1997). Infected rhesus monkeys are initially hyperactive and subsequently become apathetic and hypokinetic (Hatalski *et al.*, 1997).

1.4.2 BDV infection

Extensive epizootiologic studies have shown that *Borna disease* (BD) is rare but occurs all over Germany, extending beyond the classic disease-endemic regions (Herzog *et al.*, 1997). Furthermore, BDV-specific antibodies were detected in horses in several European countries, Israel (Herzog *et al.*, 1997; Richt *et al.*, 1994), Japan (Nakamura *et al.*, 1995), Iran (Bahmani *et al.*, 1996) and the United Sates (Kao *et al.*, 1993). In addition to its predominant natural host, the horse, other equidae, sheep, cattle, rabbits, goats, deer, alpaca, llamas, cat and pigmy hippopotamus, sloth, monkeys and ostriches have become naturally infected with BDV (Richt *et al.*, 1992; Rott and Becht, 1995; Becht and Richt, 1996; Schuppel *et al.*, 1995). In sheep flocks, clinical BD can affect large numbers of animals, however, in horse stables, usually only a few animals show clinical signs. It is assumed that the virus is transmitted through salival, nasal or conjunctival secretions, as BDV-specific RNA has been found in these secretions (Becht and Richt, 1996; Herzog *et al.*, 1997; Richt *et al.*, 1994; Richt *et al.*, 1993). Animals become infected by direct contact with these secretions or by exposure to contaminated food or water. A minimum incubation period of 4 weeks

is estimated for horses and sheep with non-specific signs such as hypothermia, anorexia, colic and constipation in the initial phase of the disease. During the acute phase of the disease, neurological signs such as ataxia, depression, circular movement, standing in awkward positions, collapsing, running into obstacles and paralysis result from nonpurulent meningoencephalomyelitis.

In felines BDV has been linked with staggering disease, (Lundgren *et al.*, 1995). However, the clinical signs of BDV in cats differ quite significantly from the symptoms of animals infected with a classic BDV strain (Weissenboeck *et al.*, 1998) suggesting that although the cats were positive for BDV RNA, either a different virus strain is causing the disease or the virus has mutated to such an extent that it is targeting different areas of the brain. Supporting the latter hypothesis is evidence that of several newborn rats inoculated intracerebrally with the cat BDV RNA strain only in one rat did the virus propagate. Several other rats displayed cytopathic changes of neurons, astrocytis and in seven cases degeneration of the hippocamapal denate gyrus, typical of BDV infection (Ludwig *et al.*, 1989). However, no external signs of infection were present, which is the major characteristic of cat BDV infection. When infested with various BDV variants, Lewis rats exhibit clinical manifestations such as behavioural disorders, paralytic disease or obesity in addition to fertility disease (Richt *et al.*, 1994; Rott and Becht, 1995), which indicates that BDV can form virus variants with different biological properties.

1.4.3 BDV and human infection

BDV's broad host and geographic range and its association with behavioural abnormalities in many species such as rhesus macaques (Stitz *et al.*, 1980), tree shrews-*Tupala glis* (Sprankel *et al.*, 1978) and rats (Narayan *et al.*, 1983; Richt *et al.*, 1992) suggest that BDV may be involved in human neuropsychiatric illnesses. As the behavioural disturbances in animals resemble those of affective disorders, particularly bipolar depression and schizophrenia, initial studies in the early 1980s investigated these disorders.

1.4.4 Serology of human BDV infection

The earliest work to suggest a link between BDV and human mental illness came from a serologic survey in 1985 of 285 patients with affective disorders in the Unites States, 694 psychiatric patients in Germany and 200 healthy controls (Rott *et al.*, 1985). An indirect immunofluorescence assay (IFA) was used to detect antibodies reactivity with a BDV infected cell line. Sera from many of these patients were subsequently analysed by a Western immunoblot (WB) assay based on BDV

nucleoprotein (N) and phosphoprotein (P) purified by affinity chromatography from infected rabbit kidney cells. In this study of 138 patients with affective disorders and 117 healthy controls, antibodies to the N protein were found in 53 (38%) patients versus 19 (16%) controls; antibodies to the P protein were found in 16 (12%) patients versus five (4%) controls; antibodies to both proteins were found in nine (6.5%) patients versus one (<1%) control. Further studies (Table 1.4) have indicated a high number of clinically inconspicuous BDV seropositive persons in a BD endemic area, (Richt *et al.*, 1997). Therefore, as is the case in horses, unapparent BDV infections may be very common in humans.

Table 1.4 Prevalence of BDV in Humans (Modified from Gonzalez-Dunia *et al.*, 1997b).

Subject Group	Test	Geographic Area	Subject Specifications	Subject No.	Percentage of Positives	References
Mental disorders	IF	USA	- Major depressive disorder	265	4.5%	Amsterdam *et al.*,1985
			- Healthy volunteers	105	0%	Rott *et al.*, 1985
	IF	Germany	- Inpatients with various Psychiatric disorders	694	0.6%	Rott *et al.*, 1985
			- Healthy disorders	95	0%	
	IF	Southern Germany	- Inpatients with various Psychiatric disorders	1003	6.8%	Bechter *et al.*, 1987,1994
			- Surgical patients	133	3.5%	
	IF	USA	- Major depression	642	2.0%	Bode *et al.*, 1988
		USA,Germany	- Volunteers,blood donors HIV negative patients	540	2.0%	
	IF	USA, Southern Germany, Japan	- Psychiatric and neurological patients	5000	4-7%	Rott *et al.*, 1991
			- Control Patients	1000	1%	
	IF	Southern Germany	Psychiatric inpatients	2377	5.9%	Bechter *et al.*, 1992
			- Surgery Patients	569	3.5%	
	IF/IP	USA	-Major depression (uni and bipolar); surgery patients	550	2.2%	Bode *et al.*, 1992
				365	2.2%	
	IF	Germany	-Acute psychiatric inpatients	71	19.7%	Bode *et al.*, 1993
	WB	USA	-Major depression (uni and bipolar)	138	6.5%	Fu *et al.*, 1993
			-Healthy controls	117	0.85%	

Subject Group	Test	Geographic Area	Subject Specifications	Subject No.	Percentage of Positives	References
	WB	USA	- Schizophrenic outpatients - Normal control subjects	90 20	14.4% 0%	Waltrip *et al.*, 1995
	WB	Japan	- Psychiatric patients	60	30%	Kishi *et al.*, 1995
	WB	Germany	- Patients with various Psychiatric disorders - Surgery patients	416 203	9.6 % 1.4%	Sauder *et al.*, 1996
	Nested RT-PCR	Germany	- Acute and chronic i.p. - Healthy blood donors	6 10	66% 0%	Bode *et al.*, 1995
	Nested RT-PCR	Japan	- Psychiatric inpatients - Healthy blood donors	60 172	37% 4.6%	Kishi *et al.*, 1995 Igata-Yi *et al.*, 1996
	Nested RT-PCR	Japan	- Psychiatric inpatients - Healthy blood donors	55 35	10.9% 0%	
	Nested RT-PCR	Germany	- Psychiatric inpatients - Healthy Volunteers	26 23	50% 0%	Sauder *et al.*, 1996
HIV infection	IF	Germany	- HIV infected individuals - HIV antibody negatives - HIV negative drug abusers	460 125 106	7.8% 1.6% 3.8%	Bode *et al.*, 1988
	IF/IP	Europe	-Asymptomatic HIV infection -HIV infection -HIV negative blood donors	1024 244 118	7.1% 13.9% 2.5%	Bode *et al.*, 1992
	ELISA	Thailand	-Asymptomatic HIV patients -Patients with AIDS -HIV negative blood donors	60 67 103	15% 17.9% 1.9%	Auwanit *et al.*, 1996
Chronic fatigue syndrome	WB	Japan	Symptomatic Patients	25	24%	Nakaya *et al.*, 1996
Parasitic	IF	East Africa	Schistomiasis and Malaria	193	9.8%	Bode *et al.*, 1992

It has become apparent that BDV infection is fairly widespread across a wide range of mammalian species including humans. The research into BDV however is still lacking with the exact mode of inter- and intra-species infection and transmission and the role of individual proteins in the infection cycle. Furthermore, very little research has been done to indicate the change if any of the various proteins across various species as the widespread infection rate of BDV indicates that it has some unique properties. The complex nature of BDV is also clearly indicated by the lack of

clarity of the *Borna disease virus* classification in the phylogenetic analysis of the RNA negative single stranded viruses that has been done to date.

1.5 NEGATIVE SINGLE STRANDED RNA VIRUS REPLICATION AND LIFE CYCLE

The predominant characteristic of the *Borna disease virus* and other negative single stranded RNA viruses is that their templates serve two functions, firstly, they serve as a template for transcription and secondly they serve as a template for replication. The enzymes responsible for initiating transcription are packaged in the virion along with the negative stranded genome. The transcription of the viral genome is the first event after entry of the virus into host cells; the process yields functionally monocistronic mRNAs (+) strands, each specifying a single protein. Replication begins under the direction of newly synthesized viral proteins, a full length (+) strand is made and serves as a template for the synthesis of (-) strand genomic RNAs (Roizman, 1991), summarized in Figure 1.1.

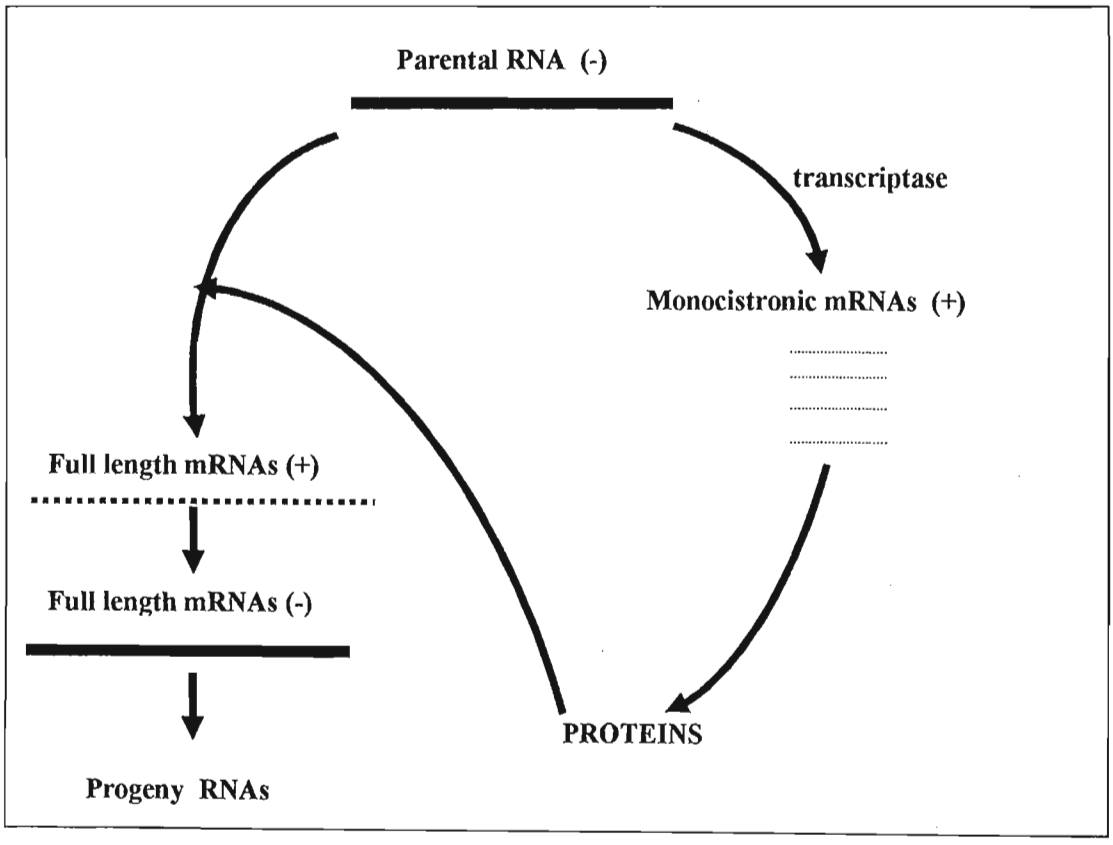


Figure 1.1 Summary of the replication and transcription of the negative single stranded RNA virus.

Central to the replication of the (-) strand viruses is that their genomic RNAs function alternatively as templates for transcription and replication. The consequences are threefold: Firstly, the virus imports into the infected cells the transcriptase enzyme to synthesize mRNAs. Secondly, it therefore follows that naked RNA without transcriptase extracted from virions is not infectious, in contrast to the naked RNA extracted from positive stranded RNA viruses. Thirdly, the mRNAs produced are gene unit lengths; they each specify a single polypeptide, referred to as monocistronic transcription.

RNA splicing signals may result in multiple mRNAs, each specifying a different protein being transcribed from the same region of genomic RNA. Primary transcription of virion (-) sense RNA by RNA-dependent RNA polymerase (RDRP) in the virion core in the cytoplasm, results in the production of (mainly) mRNA and (+) sense RNA, and formation of the replicative complex (RC). The virion proteins interact with the RC and bias it towards production of full-length (+) sense RNA and therefore of genomic (-) sense RNA. Consequently, the (+) transcript, which functions as mRNA, is different from the (+) strand RNA that serves as a template for progeny virus, even though both are synthesized on the genomic RNA. The transcription of multiple mRNAs from the same region, (monocistronic splicing) through splicing of the RNA allows the virus to control the abundance of the individual proteins, (Roizman, 1991), whereas polycistronic splicing (production of various proteins from one mRNA) generates multiple proteins.

1.5.1 Genomic organization of BDV

The recent cloning and the complete sequencing of two BDV isolates have uncovered the genomic organization of BDV (Briese *et al.*, 1994; Cubitt *et al.*, 1994a; de la Torre, 1994; Schneemann *et al.*, 1995). The genome is about 8.9 kb long with complementary 3' and 5' untranslated regions at its termini (Figure 1.3). It is generally believed that the genome contains information for at least six open reading frames (ORF). Similar to other members of the *Mononegavirales*, the genome can be divided into three main blocks (Schneemann *et al.*, 1995) (Table 1.4). The function of the X protein, translated from ORF VI, is not yet clear (Schwemmle *et al.*, 1998). However it is speculated that the X protein may facilitate nuclear export of ribonucleoprotein (RNP) complexes in a manner similar to the NS2 protein of *influenza* virus (O'Neill *et al.*, 1998).

Table 1.5 Main genes and functions of the three gene blocks of BDV.

Gene block	Function
Block 1	Encodes for the nucleoprotein and polymerase cofactors, represented by the p40 (ORF1) and p24 (ORF 2) proteins.
Block 2	Encodes for the matrix and virus envelope proteins, whose likely counterpart in BDV are the gp18 (ORF III) and p57 (ORV IV) proteins.
Block 3	Encodes for the viral polymerase, identified as ORF V. Protein X (ORF VI) – function unclear.

1.5.2 Transcription

In contrast to the vast amount of information known about *Rhabdovirus* and *Paramyxovirus* transcription, studies of BDV transcription are still in the early stages. However, it has been established that BDV has the property, unique among known animal nonsegmented negative single stranded (NNS) RNA viruses, of a nuclear site for replication and transcription of its genome (Cubitt *et al.*, 1994b). A recent report (Pyper *et al.*, 1998) suggested that the nucleolus of the nucleus of the host cell may be involved in BDV transcription and replication. This suggests that BDV appears to be one of a small group of unrelated viruses that are known to have co-opted certain aspects of nucleolar activity.

BDV exhibits a complex transcriptional pattern after infection of host cells. As with other NNS RNA viruses, BDV RNP is infectious upon transfection of susceptible host cells (Cubitt *et al.*, 1994b). Initial mapping of the BDV subgenomic RNAs, done by northern blot hybridisation, (Briese *et al.*, 1994; Cubitt *et al.*, 1994a), revealed a complex pattern of overlapping transcripts that included several polycistronic RNAs (Figure 1.3). Some subgenomic messenger RNAs (mRNAs) encoding the nucleoprotein (p40) and phosphoprotein (p24) are monocistronic, while others encoding the matrix protein (gp18), glycoprotein (p57) and polymerase protein (L) are polycistronic. In the other NNS RNA viruses, these polypeptides are usually encoded by monocistronic mRNAs. In addition, BDV does not exhibit the configuration of transcription termination signal, intergenic region and transcription initiation signal that is characteristically

present at the gene boundaries of NNS RNA viruses (Goldbach and Haan, 1994). Interestingly, BDV utilizes the host splicing machinery to generate some of its mRNAs (Cubitt *et al.,* 1994b; Schneider *et al., 1997*). The transcriptional map of BDV is summarized in Figure 1.2.

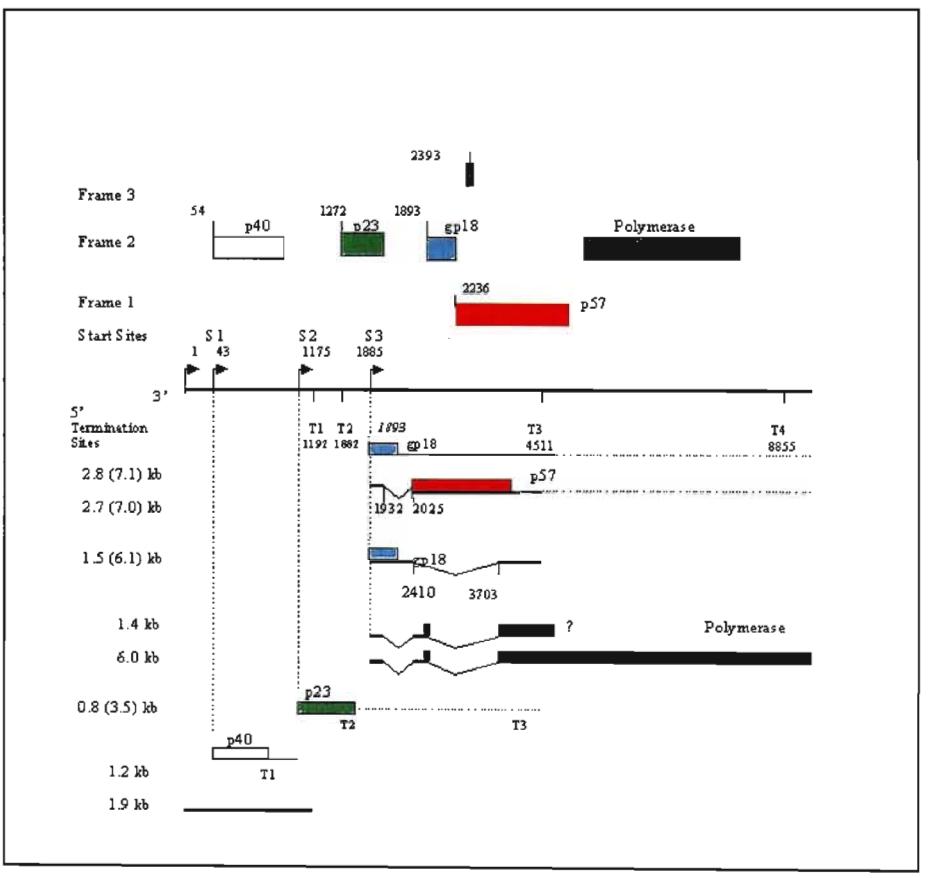


Figure 1.2 Genomic organization and transcriptional map of BDV. (Modified from Schneemann *et al.,* 1995).

1.5.3 BDV proteins and interaction

The products and functions of some of the genes of the three BDV gene blocks have been established.

1. Gene products of block one

The BDV nucleoprotein (p40) is present at high levels in infected cells and tissues. It is likely that this protein is encoded in two forms of 38 and 49 kDa (Haas *et al.,* 1986; Hsu *et al.,* 1994; Pyper *et*

al., 1993), which may be related to the presence of two in-frame initiation codons in the P40 gene sequence.

The p24 phosphoprotein is acidic, with a high serine (Ser) and threonine (Thr) content and is phosphorylated at serine residues (Hsu *et al.*, 1994; Thiedemann *et al.*, 1992; Thierer *et al.*, 1992). The function of the phosphoprotein is not known, although it is assumed that the BDV phosphoprotein associates and co-operates with the polymerase protein to play a vital role in viral transcription and replication, as its acidic features are consistent with the phosphoprotein transcriptional activator found in other NNS RNA viruses. Shoya *et al.*, (1998) established that the BDV phosphoprotein is transported into the nucleus in absence of other viral constituents and that this transportation is accomplished by virtue of BDV phosphoprotein's own nuclear localization signals, which are present in both N and C terminal regions. The nuclear localization signals of the BDV phosphoprotein are unique in that both can function independently and both have several proline residues as key amino acids. The transcription unit encoding the p24 phosphoprotein can also direct the synthesis of a polypeptide of 10kDa (p10 or X protein) as the ORF encoding the p10 protein starts 46 nucleotides upstream from the p24 phosphoprotein and overlaps in a different frame with the 213 first nucleotides of ORF II, which encodes the p24 phosphoprotein. Research done by Schneemann *et al.*, (1995) indicates that the p10 protein is present in infected cells.

2. *Gene products of block two*

The BDV ORF III most likely represents the BDV matrix gp18 protein. In contrast to other NNS RNA viruses, the BDV matrix protein is glycosylated and data suggests that it might be present on the surface of the virion envelope (Kliche *et al.*, 1994). It is predicted that ORF IV encodes for a glycoprotein polypeptide of 56kDa (p57). Sequence features suggest that this protein is a viral surface glycoprotein (p57), (Schneemann *et al.*, 1995; de la Torre *et al.*, 1996).

Recent reports have provided experimental evidence that the p57 glycoprotein is involved in virus entry (Gonzalez-Dunia *et al.*, 1997a; Schneider *et al.*, 1997) and is present as two forms in infected cells (Gonzalez-Dunia *et al.*, 1997a). One form of approximately 84 kDa (GP-84) corresponds to the full-length product encoded by ORF IV and accumulates in the endoplasmic reticulum (ER). The molecular weight of this polypeptide is higher than 5 kDa, due to glycosylation. A shorter product of 43 kDa (Gp-43) corresponds to the C terminus of Gp-84 (Gonzalez-Dunia *et al.*, 1998), and is generated through cleavage by the cellular protease furin (Richt *et al.*, 1998). Furthermore, GP-43 is present at the surface of infected cells, (Gonzalez-Dunia *et al.*, 1998). Both GP-84 and GP-43 are associated with infectious virions. The features indicate a novel maturation pathway for a NNS RNA virus surface glycoprotein (i.e. BDV GP-84 is involved in attachment to the cell

surface receptor whereas its furin-cleaved product, GP-43, is involved in pH-dependent fusion after internalisation of the virion by endocytosis), and hence, for the assembly of BDV particles (Gonzalez-Dunia *et al.*, 1997a; 1998).

3. *Gene products of block three*

ORF V is capable of encoding a polypeptide with a predicted molecular mass of 180 kDa, whose deduced amino acid sequence displays strong homology with the NNS RNA viral polymerases (L protein family) (Cubitt *et al.*, 1994a). This homology is particularly high in the case of the conserved putative catalytic domain.

4. *Protein interaction*

The phosphoproteins of NNS viruses are essential for virus transcription and replication (Schwemmle *et al.*, 1997) and their subsequent phosphorylation by cellular kinases influences the ability of phosphoproteins to form homomultimers, bind other viral proteins and serve as transcriptional activators.

Research by Schwemmle *et al.*, (1998) is geared towards characterizing the interactions of the p10 X protein, p24 phosphoprotein and p40 nucleoprotein in vitro. Their experiments utilizing extracts from infected cells and antibodies directed against the phospho, X and nucleoprotein have indicated interactions between the X and the phosphoprotein, between the phosphoprotein and nucleoprotein and to a lesser extent between the X, phospho and nucleoprotein. These interactions were confirmed through studies in (i) mammalian and yeast two hybrid systems that demonstrated specific binding between the phosphoprotein and the nucleoprotein, between the phosphoprotein and the x protein and between phosphoprotein and phosphoprotein; (ii) noninfected transfected cells where simultaneous expression of the X and phosphoprotein resulted in nuclear localization of the X protein (versus predominantly cytoplasmic localization of the X protein when expression of the phosphoprotein was suppressed); and (iii) infected C6 rat cells where the X, phosphoprotein and nucleoprotein were co-localized in the nucleus.

Analysis of a series of truncation mutants allowed the identification of three nonoverlapping regions important for phosphoprotein oligomerization (amino acids 135-172), binding to X (amino acids 33-115), and binding to nucleoprotein (amino acids 197-201), (Figure 1.3). The amino acids of the phosphoprotein critical for binding to X were mapped to the amino-terminal portion of the protein between amino acids 33 and 115. Deletion of the first 32 amino acids of the phosphoprotein resulted in increasing binding of the X protein, indicating that this region behaves as a negative

regulatory domain. The binding of the X protein and the phosphoprotein was increased 7 fold in assays using a phosphoprotein mutant in which the oligomerization domain was deleted. This suggests that the X protein binds preferentially to the monomeric form of the phosphoprotein and enhances phosphoprotein oligomerization. In other NNS RNA viruses, oligomerization of phosphoproteins correlates with viral transcriptional activity (Schwemmle *et al.*, 1998; Gao and Lenard, 1995). Co-localization of the X and the phosphoprotein in the nucleus of chronically infected cells indicates that the protein is associated with the phosphoprotein at the sites of viral replication and transcription. Taken together, these data suggest the possibility that the X protein may modulate transcriptional activity of BDV via binding to the phosphoprotein.

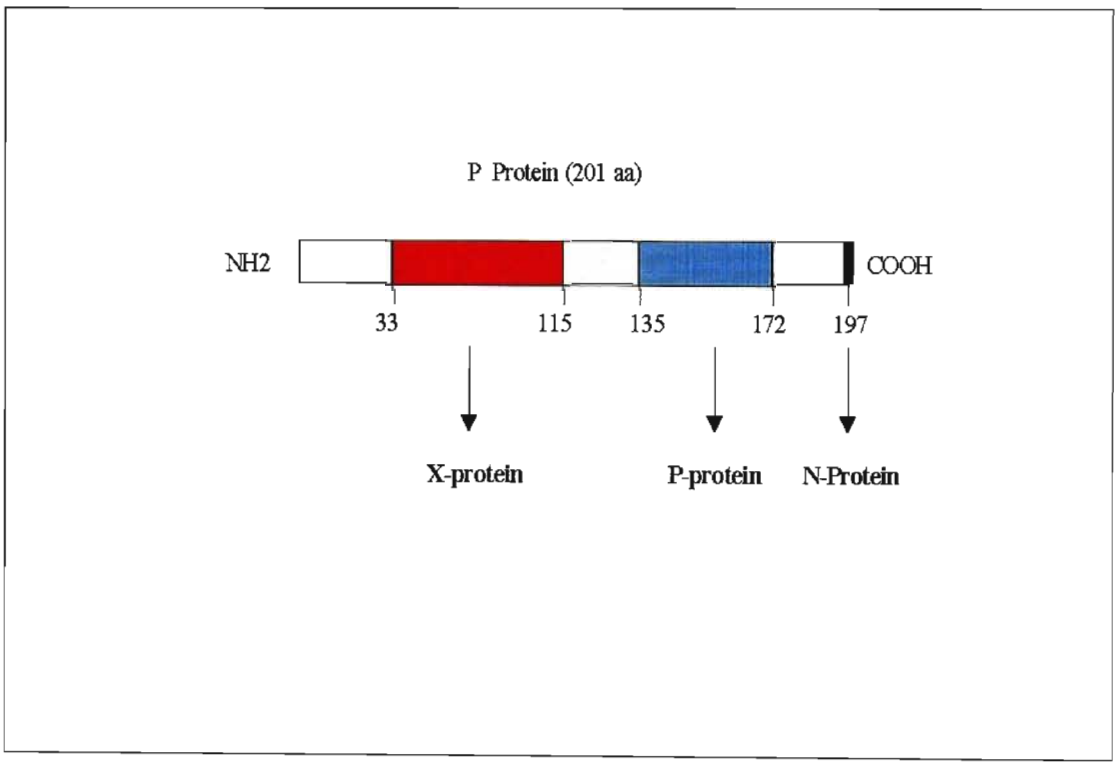


Figure 1.3 Schematic illustrating three distinct regions of the phosphoprotein, which are critical for phosphoprotein oligomerization, binding to X protein and binding to the nucleoprotein. After Schwemmle *et al.*, (1998).

1.6 BDV EVOLUTIONARY RELATIONSHIPS

The *Borna disease virus* belongs to the negative single stranded super group, which forms a rather homogenous group in terms of structure. In principle, the members of this group all have nucleoprotein particles surrounded by a lipid envelope containing one or more glycoproteins. Although only limited homology among the various viral encoded proteins is present (mainly the polymerase proteins), they share several common features of genomic organization and replication strategy, (Goldmann and Haan, 1994). Table 1.5 summarizes some of the features of the genomes of the *Borna disease virus* and other members of the *Mononegavirales*.

Table 1.6 Arrangement of genes of the *Borna disease virus* and other members of the *Mononegavirales* from the 3'- end to 5'-end, (Goldbach and Haan, 1994).

Family	Sub-family	[a]Genus	3'	[b]Gene order							5'
Bornaviridae		**BDV**	le	N	P		M	G		L	tr
Rhabdoviridae		**VSV**	le	N	P		M	G		L	tr
		SYNV	le	N	P	Sc4	M	G		L	tr
		RabiesV	le	N	P		M	G	Ps	L	tr
		IHNV	le	N	P		M	G	NVGns	L	tr
		BEFV	le	N	P		M	G		L	tr
Filoviridae		**Ebola V**	le	N	P		M1	G	? M2	L	tr
Paramyxoviridae	**Paramy-xovirinae**	**Measles V**	le	N	P/C/V		M	F	H	L	tr
		SendaiV	le	N	P/C/V		M	F	H/N	L	tr
		MumpsV	le	N	P/V		M	F	SH	L	tr
	Pneumo-virinae	**TRTV**	le	N	P		M1	F	H/N G	L	tr
		RSV	le	NS1	N P		M1	SH	M2 SH M2	L	tr
				NS2			M1		G F		

[a] BDV = *Borna disease virus*, VSN = *Vesicular stomatits virus*; SYNV = *Sonchus Yellow Net virus*; RabV = R*abies virus*, IHNV = *Infectious hematopoetic necrosis virus*, BEFV = *Bovine ephemeral fever virus*, EbolaV = *Ebola virus* , MeasV = *Measles virus*, SendV = *Sendai virus*, MumV = *Mumps virus*, TRTV = *Turkey Rhinotracheitis virus*, RSV = *Respiratory synctial virus*.

[b] le (leader transcript that is not translated), N (Nucleocapsid protein tightly bound to viral RNA), NS1, NS2 (non-structural protein 1,2), P (phosphoprotein), C (Capsid), M; M1, M2 matrix

proteins embedded in the bilayer membranes), sc4 (an RNA transcripts is observed, but the protein product has not be identified, may be the equivalent of NS protein in *vesicular stomatitis virus*), G (Glycoprotein forming the projections out of the membrane), F (fusion protein, glycoprotein, which binds with other glycoproteins sh and G to form an oligomeric process); H (haemaglutinin protein), L (Core polymerase protein functioning in the viral replication in concert with other proteins), tr (train sequence that is not translated).

1.6.1 RNA polymerase relationships

Kamer and Argos, (1984) were the first researchers to identify sequence relationships between the known *poliovirus* polymerase and those from several positive stranded RNA viruses. One particular conserved positive stranded RNA motif was well preserved and consisted of a central Gly-Asp-Asp triplet (GDD); flanked by pentapeptides composed mostly of hydrophobic residues, hinting at a beta hairpin tertiary structure composed of two hydrogen-bonded anti-parallel beta strands separated by a short exposed loop encompassing the GDD triplet of amino acids, (Kamer and Argos, 1984.) In contrast to the GDD motif found in positive stranded viruses the GDD motif is present as a GDN motif in negative single stranded viruses. Such a motif was found in several different RNA virus species, suggesting that many viral species, across many hosts, and with great variety in structural and infective features and genomic organization, all could be evolutionarily and divergently related (Kamer and Argos, 1984).

Goldbach and Haan, (1994) pioneered the inclusion of RNA negative single stranded viruses in a tentative phylogenetic tree analysis, based partly on a clustal analysis (Figure 1.4) published by Koonin, (1991). The results seem to indicate that: (i) the original super grouping of RNA viruses was appropriate; (ii) the negative single stranded viruses are categorically divided according to their non-segmented and segmented polymerases; and that the most conserved sequences contained the GDD motif.

Very little work has been done to date to confirm the results achieved by Goldbach and Haan, (1994) and there is a considerable lack of research including recently sequenced polymerases. In addition, different methods are being used to investigate the inter- and intra-evolutionary relationships of viral sequences. Although phylogenetic methods including alignment analysis are still one of the most predominant methods to investigate evolutionary relationships a recent shift in the focus of evolutionary studies has taken place. Haydon *et al.*, (1998), and Ina's, (1995) estimations of the extent of synonymous and non-synonymous substitution among viral sequences based on Nei and Gojobori's, (1986) paper, are indicative of such a shift.

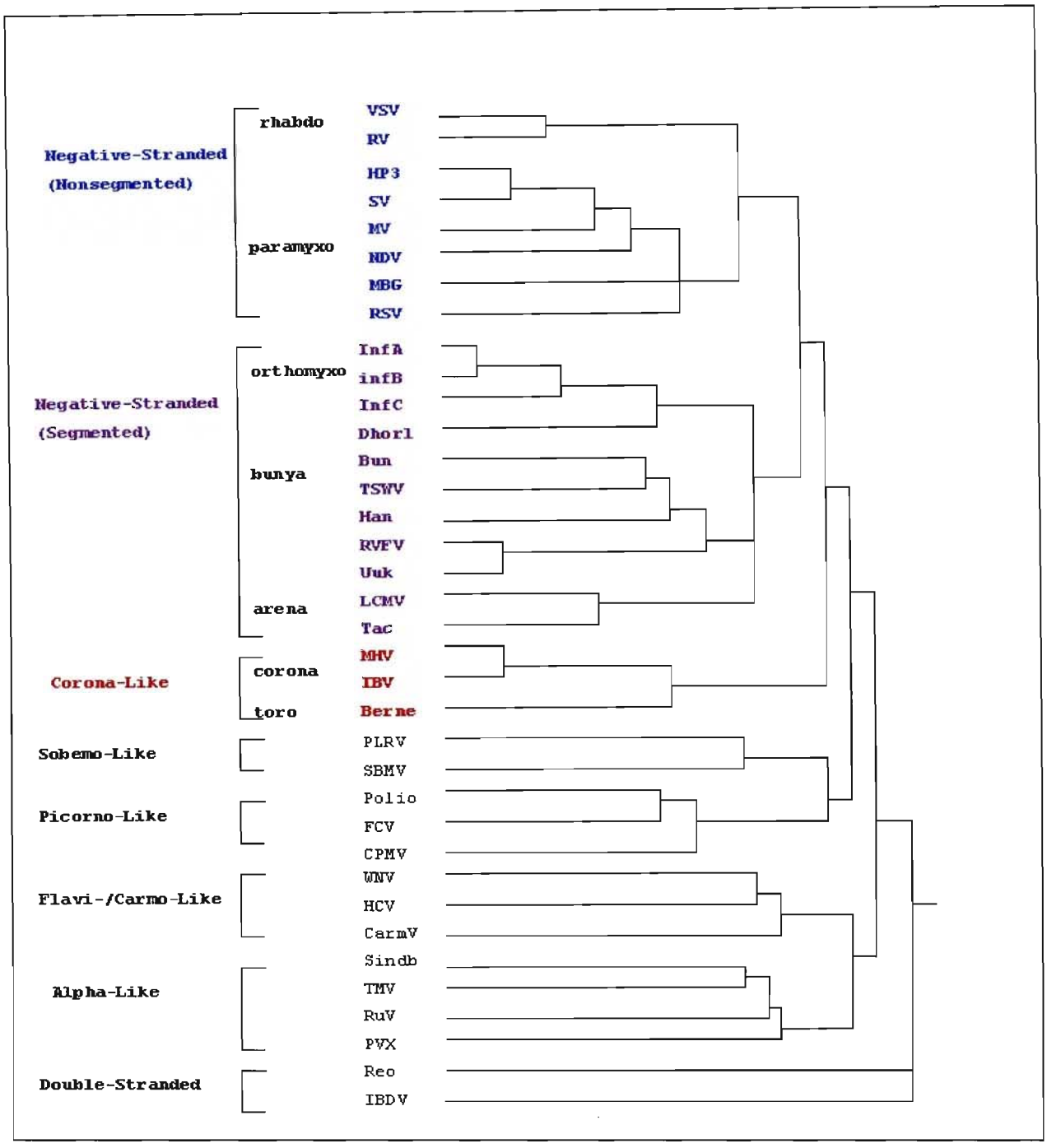


Figure 1.4 Tentative phylogenetic tree for the RNA dependent RNA polymerases of eukaryotic RNA viruses proposed by Goldbach and Haan, (1994). The various supergroups are indicated.

1.7 METHODS FOR INVESTIGATION OF EVOLUTIONARY RELATIONSHIPS

1.7.1 Introduction

Recent advances in whole genome sequencing indicate that vast amounts of sequence data are flooding the DNA and protein databases. Over 40 complete genomes are available, with many more genomes in the sequencing and assembly stage, including five eukaryotic genomes and the recent arrival of a draft of the complete human genome sequence (Lander *et al.*, 2001).

New high throughput technologies in structural proteomics and gene expression analysis have increased the rate in which sequence data is being characterized. In addition, the majority of new sequences have homologues in the new existing sequence databases, so new biological or structural data are now applicable not only in the protein under study but are also propagated through the sequence databases by means of this homology.

The result is a massive explosion of biological information. The sequencing of the human genome and numerous pathogen genomes has resulted in an explosion of potential drug targets, with widespread implications in the pharmaceutical industry (Lenz *et al.*, 2000). The analysis of genomes for example, of extremophile micro organisms, has led to the identification of many enzymes, showing activity and stability at extremes of temperature, pH, pressure and salinity, some of which have potential for industrial and biotechnological applications (Niehaus *et al.*, 1999). It is evident that the systems and methodologies used for analyzing all this information must adapt to the changing perspectives. A major challenge for bionformaticians will be to bring together the wealth of information now available for public access and the vast number of applications that have been developed to process and interpret the biological data into an integrated network.

Multiple alignments and their generation have become increasingly relevant to analysis projects and sequence annotation. Although it is not always apparent, multiple alignments are involved in most of the new computational methods used in genome projects. Recent developments of database search methods have exploited the information to detect more and more distant homologues, (Altschul *et al.*, 1997, Karplus *et al.*, 1998; Neuwald *et al.*, 1997) and many of the new methods referenced above critically depend on accurate multiple alignments. Multiple alignments also play an essential part in the new integrated systems being developed for the analysis and comparison of whole genomes.

The form of viral evolution has been a controversial point in the past and the molecular data now available suggests that virus evolution is based on two mechanisms: divergence from common ancestor and interviral recombination (Goldbach and Haan, 1994).

Multiple alignments of complete sequences play a central role in the integration and analysis of this complex and heterogeneous data. By placing the sequence in the context of the overall family, the multiple alignments permit not only a horizontal analysis of the sequence of its entire length but also a vertical view of the evolution of the protein.

With the accumulation of many RNA stranded viral genome sequences and the enormous variability exhibited by these genome sequences it is has become clear that some point of reference is needed in any alignment and phylogenetic analysis of such sequences in order to present a valid result. In analysing negative single stranded RNA viruses this consensus point takes the form of a polymerase protein, which is particularly well preserved and contains a central Gly-Asp-Asn triplet (GDN) flanked by penta peptides composed mostly of hydrophobic residues, hinting at a Beta hairpin tertiary structure; composed of two hydrogen-bonded anti parallel strands, separated by a short exposed loop encompassing the GDN amino acids. Such a motif was found in several different RNA virus species suggesting that many viral species, across many hosts, and with a great variety in structural and genomic organization, all could be evolutionarily and divergently related (Kamer and Argos, 1984).

1.7.2 Methods of alignment and phylogeny analysis

1. Alignment methods

The comparison or alignment of biological sequences began in the early seventies, with the first dynamic programming algorithm for the global or full full-length alignment of two sequences introduced by Needleman and Wunsch, (1970). The optimal local alignment between a pair of sequences involves a simple modification to the Needleman-Wunsch method, defined by Smith and Waterman, (1981), in which only the highest scoring sub segments of the two sequences are aligned. Sankoff, (1975) who developed his first formal algorithm for multiple sequence alignment extended the pairwise dynamic programming algorithm.

In order to multiple align larger sets of sequences, most methods in use today employ some kind of heuristic approach to reduce the problem to a reasonable size. Traditionally the most popular

approach has been the progressive alignment method (Feng and Doolittle, 1987). A multiple alignment is built up gradually by aligning the closest sequences first and successively adding the more distant ones. A number of alignment programs based on this method exist, for example *Multialign* (Barton and Sternberg, 1987), *Multal* (Taylor, 1988), *Pileup* and *Clustal X* (Thompson *et al* 1997); which provides a graphical interface for *Clustal W* (Thompson, Higgins and Gibson, 1994). They use a global alignment algorithm (Needleman and Wunsch, 1970) to construct an alignment of the entire length of the sequences. They differ mainly in the method used to determine the order of alignment of the sequences. *Multal* uses a sequential branching method to align the closest sequences first and then subsequently align the next close sequences to those already aligned. *Multialign* and *Pileup* construct a guide tree using the UPGMA method (Sneath and Sokal, 1973). A consensus method is then used to align larger and larger trees of sequences according to the branching order of the tree. *Clustal X* uses the alternative neighbour-joining algorithm (Saitou and Nei, 1987) to construct a guide tree incorporating in addition sequence weighting, position specific gap penalties and a choice of residue comparison matrix depending on the degree of identity of the sequences. In contrast to the above global methods, *PIMA* uses a local dynamic programming algorithm (Smith, Waterman and Fitch, 1981) to align only the most conserved motifs. *PIMA* offers two alignments by default using maximum linkage and sequential branching algorithms to decide the order of the alignment, which can be referred to as MLPIMA and SBPIMA respectively.

DIALIGN is a newer program for multiple alignment, (Morgenstern *et al.*, 1996, Morgenstern, 1999). While standard alignment methods rely on comparing single residues and imposing gap penalties, *DIALIGN* combines local and global aspects of sequence alignment by assembling pair wise and multiple alignments from locally conserved gap-free segment pairs, (Morgenstern 1999).

In a recent study by Thompson *et al.*, (1999) a database of benchmark alignments has been used to compare the most widely used programs for multiple sequence alignment under a variety of different conditions. Here, *DIALIGN* was found to be the best method for local multiple alignments. It also performed well on globally related sequence sets, though here *Clustal W* (Thompson *et al.*, 1994) and *SAGA* (Notredam and Higgins, 1996) were superior. A recent paper by Lassmann and Sonnhammer, (2002) provides for a systematic evaluation and comparison of multi alignment programs.

In addition, numerous new alignment programs have recently been developed which provide a fresh approach to the multiple alignment problems. Hidden markov models (HMMS) have come to the forefront of this as they can be used as statistical models of the primary structural consensus of

a sequence family (Baldi *et al.*, 1994). The program *HMMT* (Eddy, 1995) uses a simulated annealing method to maximize the probability that an HMM represents the sequences to be aligned.

As the increasing size of protein sequence databases is straining methods of sequence analysis, even as the increased information offers opportunities for sophisticated analysis of protein structure, function and evolution. A new method *MEME* (Multiple expectation maximization for motif elicitation) (Grundy *et al.*, 1997, Bailey and Gribskov, 1998) was developed to compensate for these weaknesses. This method identifies motif patterns in a protein family, and these motifs are combined into a hidden markov model (HMM), for use as a database searching tool (Figure 1.5). A motif is a sequence pattern that occurs repeatedly in a group of related protein or DNA sequences. *MEME* represents motifs as position-dependent letter-probability matrices which describe the probability of each possible letter at each position in the pattern.

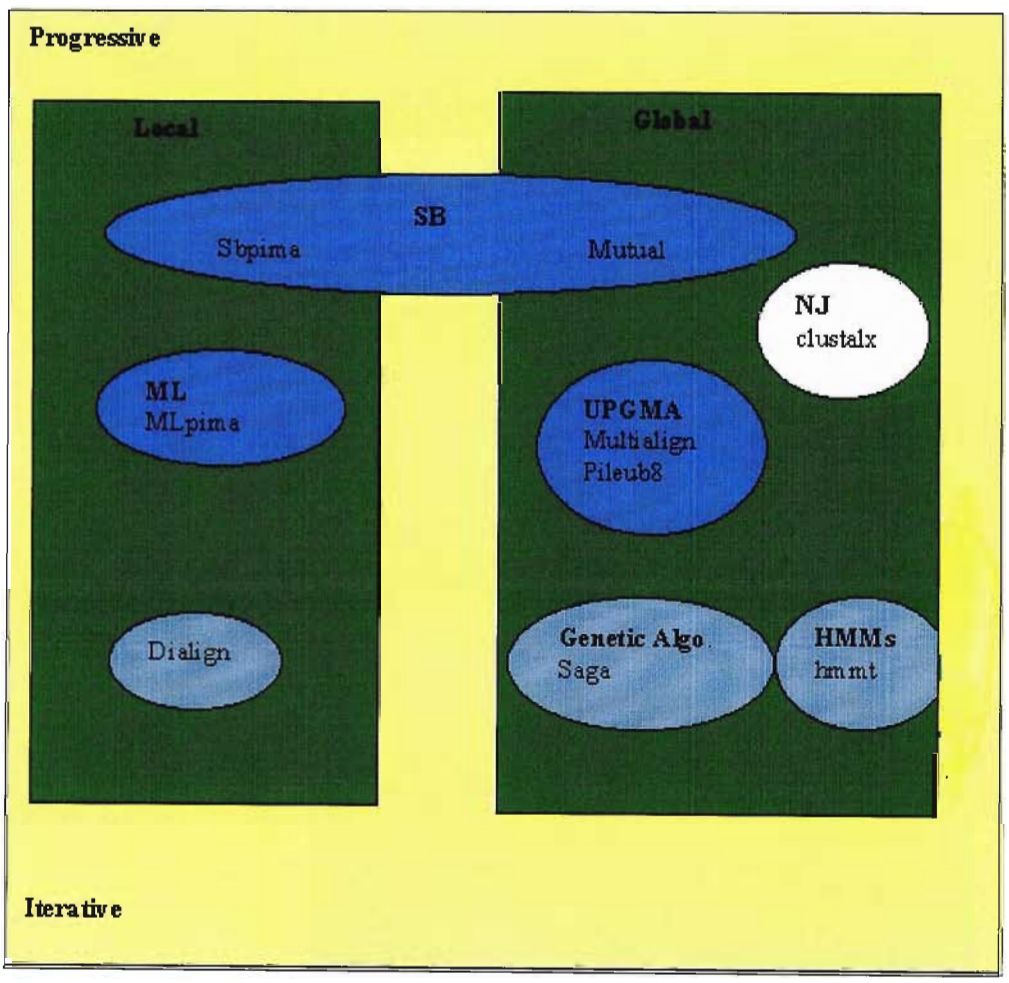


Figure 1.5 Diagram showing the relation between different alignment programmes and algorithms.

Traditional molecular phylogeny studies have relied on multiple alignments to define the phylogenetic relationships between organisms (for a detailed review see Philips *et al.*, 2000). In the post-genomic era, this field of research is experiencing renewed interest since phylogenetic studies can now address the entire proteome of organisms widely scattered across the phylogenetic spectrum. Such evolutionary studies performed at the proteome levels have revealed numerous lineage specific gene losses and the unforeseen importance of lateral transfers in the course of evolution, (Aravind *et al.*, 2000; Eisen 2000, Koonin *et al.*, 2000, Nelson *et al.*, 1999), The construction of the 'phylome 'of an organism, i.e. the complete set of phylogenetic trees derived from the proteome, can be used to assess the variability of single gene phylogenies compared to the organismal phylogeny based on rRNAs and sheds light on the evolution of metabolic pathways.

Numerical methods, for inferring phylogenies from molecular data have existed for over 20 years, but there is still much confusion in the literature about their assumptions and properties. In general there is little coverage of them in textbooks of evolution or of molecular biology; that which exists is usually a brief and mechanical exposition of a particular method familiar to the author. As a result the inference of phylogenies often seems divorced from any connection to other methods of analysis of scientific data.

There are three major methods of inferring phylogenies, which include the parsimony and compatibility methods, the distance methods, and the maximum likelihood methods.

2. Parsimony

This method can be defined simply by counting the minimum number of base substitutions that are required for each proposed tree (leaving aside for the moment the issue of insertion and deletions). The tree requiring the fewest changes is preferred. This is the parsimony criterion. It was first introduced, in the context of estimating phylogenies from gene frequencies, by Cavalli and Edwards, (1967) who called it the "method of minimum net evolution". The word "parsimony" was first associated with it when Camin and Sokal, (1965) published an influential description of this method for discretely coded morphological characters. Eck and Dayhoff, (1968) described the first application to molecular sequences.

Parsimony normally applied to base substitutions counted only base substitutions. Sankoff and Rousseau, (1975) and Sankoff, (1975) describe a method that performs alignment of sequences at the same time as it estimates the phylogenies by minimising the weighted count of substitutions and deletion/insertion events.

Protein sequences are more difficult to analyse. The complexity of mapping from amino acids to codons causes difficulty in computing. Algorithms for counting the number of base substitutions have been given by Moore *et al.*, (1988), Fitch, (1971, 1981). In the *Phylip* package Felsenstein, (1993) counts only those base substitutions that also change the amino acid, under the assumption that synonymous changes are more probable and should thus be demphasised.

3. Compatibility

This is a method closely related to parsimony. It uses a different criterion for resolving conflict among characters. A character in this case is compatible with a phylogeny if its evolution can be explained without assuming that any states may arise more than once. Thus, sites that can show three bases A, C, and T, are compatible with a phylogeny if the observed data could arise with only two nucleotide substitutions. The compatibility method finds the tree on which the maximum numbers of sites are compatible with the user.

The compatibility criterion was first proposed for discrete two state morphological characters by Le Quesne, (1973). Estabrook and Landrum, (1975) and Fitch, (1971) showed how to determine whether two nucleotide sites are compatible with each other, in the sense that there must exists a tree on which they can all evolve with no extra changes. However, Fitch, (1971) also showed that a set of sites that are not pairwise compatible may not be jointly compatible; in that there may not exist one tree on which all can evolve without extra changes. This is in contrast to some classes of multi state morphological characters for which Estabrook *et al.*, (1977a, 1977b) proved that when all characters are pairwise compatible, the must be jointly compatible, and the tree fitting them all can be found very easily. Although the absence of this pairwise compatibility theory for nucleotide sequences, makes it somewhat harder to find the tree with the most sites compatible with it, compatibility method are no harder to use than parsimony methods.

4. Distance matrix methods

Distance methods fit a tree to a matrix of pairwise distances between the species. For nucleotide sequence data the distances might be for example calculated from the fraction of sites different between the two sequences. The phylogeny makes a prediction of the distance for each pair of as the sum of branch lengths in the path form one species to another through the tree. A measure of goodness of the observed distances to the expected sites is used, and the phylogeny is preferred, which minimises the discrepancy between them as evaluated by this measure. There is a widespread misconception that distance assumes a molecular clock, mostly as molecular evolutionists using these methods have also tended to make an assumption and invoke is as the

reason why their methods work. It is possible to either assume or not assume a molecular clock when using distance methods. Fitch and Margolish, (1967) introduced the first distance matrix method, and Cavalli-Sforza and Edwards, (1967) had independently produced another site different between the two sequences.

1.7.3 Measurement of virus mutation rates

Viruses have not generally been considered from the same evolutionary vantage point as other organisms, despite some interest in evolution on the part of virologists. Past efforts to elucidate the concept of viral evolution were problematic with virologists relying on phenotypic characteristics. However, the number, range and resolution of phenotypic characters that could be studied were very limited. The advent of molecular evolution, the field resulting from the introduction of molecular biology into evolution, shifted the evolutionary emphasis away from characteristics based on phenotypic expression towards the viral genome and subsequently to variation in the viral genome, (Morse, 1994).

These variations led to observations of protein polymorphisms (genetic variants in functionally identical proteins) (Zuckerkandl and Pauling, 1962) later extending to gene sequences at the nucleotide level. The neutral evolution theory developed by Kimura, (1968; 1983) expanded by viral evolutionists (Temin, 1989; Temin, 1993) led to a formal expression of these variations on a viral population as fundamental to the development of viral evolutionary theory (Morse, 1994).

RNA viruses in general exhibit a high rate of base mismatch during gene replication; estimated to be between 10^{-4} to 10^{-3} per nucleotide site per replication, (Holland, 1993; Domingo and Holland, 1988). This may be indicative of point mutations forming a major part of virus evolution. Single-base nucleotide substitutions may be classified as synonymous (resulting in no amino acid change) or non-synonymous (resulting in amino acid change). It has been conjectured that base substitutions are likely to be neutral with synonymous point mutations the most probable form of nucleotide substitutions in which neutral evolution may proceed (Haydon *et al.*, 1998), as the rate of accumulation of synonymous substitutions is approximately five fold higher than the rate for non-synonymous substitutions for both mammalian genes (Li *et al.*, 1985) and for influenza virus genes (Hayashida *et al.*, 1985). Furthermore, the rate of synonymous substitutions may be directly related to codon usage. There are several reports showing a preponderance of non-synonymous substitution during RNA virus evolution (Rocha *et al.*, 1991; Dietz *et al.*, 1990), which may be a result of RNA structural functionality, secondary structural constraints and overlapping transcriptional reading frames (Haydon *et al.*, 1998).

One of the more important estimations for molecular evolutionary analysis is the quantification of the number of substitutions per site between nucleotide sequences (Ina, 1995; Muse, 1996). It is imperative to estimate the numbers of synonymous (ds) and non-synonymous (dn) substitutions per site separately, as estimates of ds and dn are used not only for reconstruction of phylogenetic trees but also as a statistical test for a neutral theory of molecular evolution, (Ina, 1995). Various methods exist for estimating synonymous and non-synonymous mutations, however, until recently none of these were based on a valid statistical footing (Muse, 1996). Some of these methods have been utilized for an estimation of the mutation of RNA virus sequences (Saitou, 1987; Haydon *et al.*, 1998), with several methods exhibiting overestimation and underestimation of the number of synonymous sites and non-synonymous sites respectively. New methods have been developed which aim to correct these biases (Ina, 1995).

Genomic analysis of BDV protein isolates has yielded differential estimates of nucleotide sequence conservation with extensive conservation of host BDV nucleoprotein and phosphoprotein isolates being reported (Schneider *et al.*, 1994; Bode *et al.*, 1995; *Sauder et al.*, 1996). Due to their high degree of sequence conservation BDV homologues provide a useful mechanism for the investigation of molecular evolution particularly if the evolution is estimated via point or synonymous mutations.

1.8 AIMS OF THIS INVESTIGATION

There are several areas in the negative single stranded RNA virus field that require investigation, especially information that will assist a more accurate classification of the *Bornaviridae* and provide a better understanding of the evolutionary trends.

In this investigation, the evolutionary trends of *Borna disease virus* proteins were investigated by determining mutation rates, in order to contribute to the body of knowledge addressing these issues. The aims of this study were twofold:

Firstly to investigate the accuracy of the *Bornaviridae* familial classification by an alignment and clustering of the polymerase proteins of several divergent RNA negative stranded viruses to generate an accurate phylogenetic analysis.

Secondly to estimate the number of synonymous (nucleotide substitution) and non-synonymous (amino acid change) mutation rates of proteins (nucleoprotein, phosphoprotein, glycoprotein, matrix protein) exhibited by various *Borna disease virus* host species and the proteins (nucleoprotein, phosphoprotein, glycoprotein, matrix and X protein) of three *Borna disease virus* strains.

The molecular biology of the *Bornaviridae* as reviewed in the above chapter indicates that there are a number of areas involving protein interaction and host virus protein interaction that is puzzling to investigators. The estimation of the protein mutation rate is a fairly unique method used to clarify the evolutionary trend of the proteins examined, as this kind of analysis enables the identification of positive and neutral selection trend of proteins across species, which may help clarify and enhance the lack of current knowledge.

CHAPTER 2

MATERIALS AND METHODS

2.1 INTRODUCTION

Variation in amino acid sequences forms the basis of many evolutionary investigations and phylogenetic analyses. Nucleotide or amino acid sequences are aligned, similarities and differences determined, and the information used to formulate evolutionary pathways in various species. The employment of multiple sequence alignment methods, such as those utilized in this investigation, allows for the detection of common patterns in nucleic acid and protein families, is helpful in suggesting primers for the polymerase chain reaction when amplifying nucleic acids, provides a better understanding of molecular evolution and assists with predicting secondary and tertiary structures of proteins. These alignment investigations form the basis of investigations of conserved regions of DNA, RNA and proteins. It is therefore possible to identify structurally conserved regions between related proteins, even when overall sequence similarity is low. This feature is vital when analysing diverse sequences such as those of the negative stranded RNA viruses.

This chapter presents the materials and methods used to analyse the taxonomic status and evolutionary status, i.e. mutation rate, of the *Borna disease virus*.

The aims of this research were briefly:

1. Investigation of the status of the *Borna disease virus* (BDV) familial classification by an alignment and clustering of the polymerase proteins.
2. An estimation of the mutation rates of various proteins of three *Borna disease virus* strains and a number of BDV hosts.

2.2 INVESTIGATION OF THE TAXONOMIC STATUS OF BDV

2.2.1 Materials

The amino acid polymerase sequences of *Borna disease virus* and a number of other negative stranded RNA viruses were selected to determine the taxonomical status of BDV. These sequences were obtained from the public database of the National Centre for Biotechnology Information (NCBI) http://www.ncbi.nlm.nih.gov/Genbank. The accession numbers and dates of submission are presented in Table 2.1.

Table 2.1 Virus sequences used for the determination of the taxonomic status of BDV.

Virus Strain	Abbreviation used in text	Accession number	Date of submission
Borna disease virus V	BDV V	U04608	28 July 1994
Borna disease virus He	BDV He	L27077	15 August 1994
Marburg virus		Z12132	19 September 1996
Ebola virus		AF086833	28 February 1999
Human Parainfluenza virus III	Human PIV3	V51116	6 March 1997
Measles		AB016162	10 June 1998
Mumps		M19933	2 August 1993
Human respiratory syncytial virus	Human RSV	U3966229	29 March 1991
Rabies virus		AB009663	11 December 1997
Infectious haematopoietic necrosis virus	IHNV	L40883	2 February 1999
Viral haemorrhagic septicemia virus	VHSV	YI8262	16 October 1998
Rice transitory yellowing virus	Rice TYV	AB011257	8 January 1998

Sonchus yellow net virus	SYNV	L32603	10 October 1994
Vesicular stomatitis virus	VSV	J02428	3 August 1993
Influenza A		M38291	2 August 1993
Influenza C		M20861	2 March 1993
Thogoto virus		Y17873	5 August 1998
Bunyamwera virus		X14282	5 July 1991
Hantaan virus		D25531	25 November 1993
Dugbe virus		U15018	22 September 1994
Uukuniemi virus		D10759	21 March 1992
Toscana virus		X68414	16 September 1992
Rift valley fever virus	RVFV	X56494	3 June 1994
Rice Stripe virus	Rice SV	D31879	20 June 1994
Rice grassy stunt virus	Rice GSV	AB009656	10 December 1997
Lymphocytic choriomeningitis virus	LCMV	AF004519	14 May 1997
Tomato spotted wilt virus	TSWV	AF020660	8 May 1998

2.2.2 Alignment procedures

The latest version of the *Clustal X* (version 1.8, 2000) software package described by Thompson, Higgins and Gibson, (1994), was selected for this investigation, as it offers a high degree of accuracy when aligning virus sequences, far superior than other methods (Aiyar, 2001; Jeanmougin *et al.*, 1998,

Thompson *et al.*, 1997, Palacios *et al.*, 2002). This is of great importance when aligning multiple sequences with sequence lengths of more than 2000 amino acid residues, which requires a high degree of accuracy. Motifs of extremely variable RNA sequences may otherwise not be detectable. Two other alignment programmes, *DIALIGN* (version 3.0, 2002), (Morgenstern *et al., 1996),* and *MEME* (version 3.0, 2001), (Grundy *et al.*, 1997, Bailey and Gribskov, 1998), were selected to confirm results obtained by the *Clustal X* alignment programme as recently a paper by Lassmann and Sonnhammer (2002) indicated that *Clustal W* could compete only in strictly global cases with high sequence similarity. This is in contrast to a paper by the creator of *DIALIGN*, Morgenstern *et al.*, (2002), which indicates that *DIALIGN* has been shown to return high scoring fragments that are highly correlated to exons in genomic sequences. The extent of local sequence conservation could not be expected to exactly coincide with protein-coding regions and it was not possible to predict whole gene structures solely based on sequence similarity information. If the evolutionary distance between the compared species is close, even non-functional parts of the sequences may be conserved and it becomes difficult to distinguish functional from non-functional parts of the sequences, (Morgenstern *et al.*, 2002*).* *MEME* was selected as this programme enables the identification of motifs from protein sequences.

The sensitivity of the *Clustal X* software was provided by the employment of a clustering facility based on a Gonnet 350 matrix. The Gonnet matrix was developed by Gonnet, Cohen and Benner, (1992) using exhaustive pairwise alignments of the protein databases as they existed at that time. They utilized classical distance measures to estimate an alignment of the proteins. They then used this data to estimate a new distance matrix, which was used to refine the alignment and estimate a new distance matrix.

The following steps and software parameters were followed for generating the RNA virus sequence alignment:

1. **The sequences to be aligned were placed into a text file in FASTA format and loaded into the *Clustal X* programme**

A sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line is distinguished from the sequence data by a greater- than (>) symbol in the first column (Figure 2.1).

>gi|532319|pir|TVFV2E|TVFV2E envelope protein
ELRLRYCAPAGFALLKCNDADYDGFKTNCSNVSVVHCTNLMNTTVTTGLLLNGSYSENRTQI
WQKHRTSNDSALILLNKHYNLTVTCKRPGNKTVLPVTIMAGLVFHSQKYNLRLRQAWCHFPS
NWKGAWKEVKEEIVNLPKERYRGTNDPKRIFFQRQWGDPETANLWFNCHGEFFYCKMDWF
LNYLNNLTVDADHNECKNTSGTKSGNKRAPGPCVQRTYVACHIRSVIIWLETISKKTYAPPRE
GHLECTSTVTGMTVELNYIPKNRTNVTLSPQIESIWAAELDRYKLVEITPIGFAPTEVRRYTGG
HERQKRVPFVXXXXXXXXXXXXXXXXXXXXXXXVQSQHLLAGILQQQKNLLAAVEAQQQM
LKLTIWGVK

Figure 2.1 Example of a sequence in FASTA format.

2. The following multiple alignment parameters (Figure 2.2) were utilized.

Multiple alignments were carried out in 3 stages:

Stage 1: All sequences were compared to each other (pairwise alignments), followed by a preliminary multiple alignment.

Stage 2: A dendogram, guide tree, (like a phylogenetic tree) was constructed, describing the approximate groupings of the sequences by similarity (stored in a file).

Stage 3: The final multiple alignment was carried out, using the dendogram as a guide.

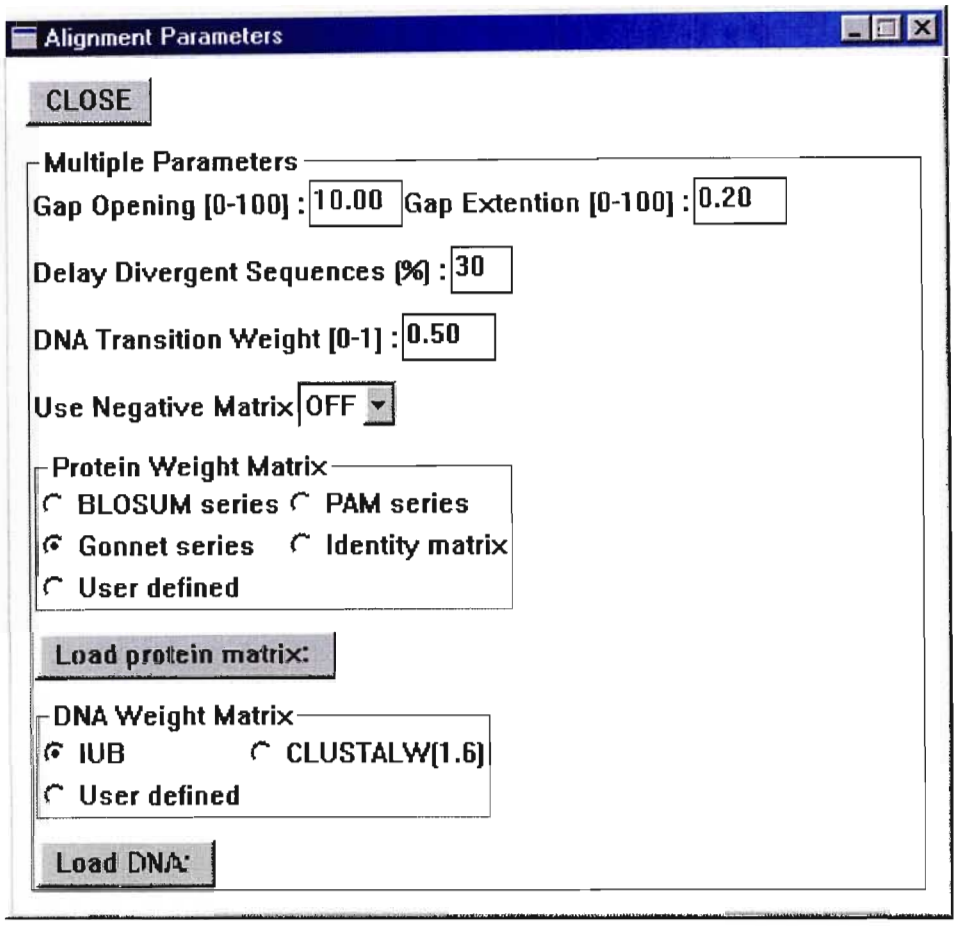


Figure 2.2 Screen of the *Clustal X* programme showing the possible parameters.

The multiple alignment parameters for the RNA virus sequences parameters were chosen in the following order:

1. Gap opening: 10
2. Gap extension: 0.20
3. Delay divergent sequences: 40%
4. Protein Weight matrix: Gonnet

These parameters controlled the final multiple alignment, as this was the core of the *Clustal* X programme.

Each step in the final multiple alignment consisted of aligning two alignments or sequences. This was done progressively, following the branching order in the guide tree or dendogram. The basic

parameters to control this were two gap penalties and the scores for various identical/non-identical residues.

The gap opening and extension penalties were set. These controlled the cost of opening up every new gap and the cost of every item in a gap. Increasing the gap opening penalty made gaps less frequent. Increasing the gap extension penalty made gaps shorter. Terminal gaps were not penalised.

The delay divergent sequence switch delayed the alignment of the most distantly related sequences until after the most closely related sequences were aligned. The parameter setting (Figure 2.2) showed the percent identity level required to delay the addition of a sequence; sequences that were less identical than this level to any other sequences were aligned later.

3. The sequences were aligned and saved in *Phylip* and *Clustal* file format

To ensure that the resulting alignment file could be used for various types of analyses, *Clustal X* allowed the alignment file to be saved in various formats. In this study the alignment file was saved in *Phylip* and *Clustal* format to allow the phylogenetic analysis performed by the *Phylip* compilation of programmes.

The *Clustal* format alignment file generated contained the following characters, which indicated the degree of conservation. Three characters were used: (*) indicated that the codon belonged to a fully conserved group; (:) indicated that the codon belonged to a stronger conserved group; (.) indicated that the codon belonged to a weaker conserved group (Figure 2.3).

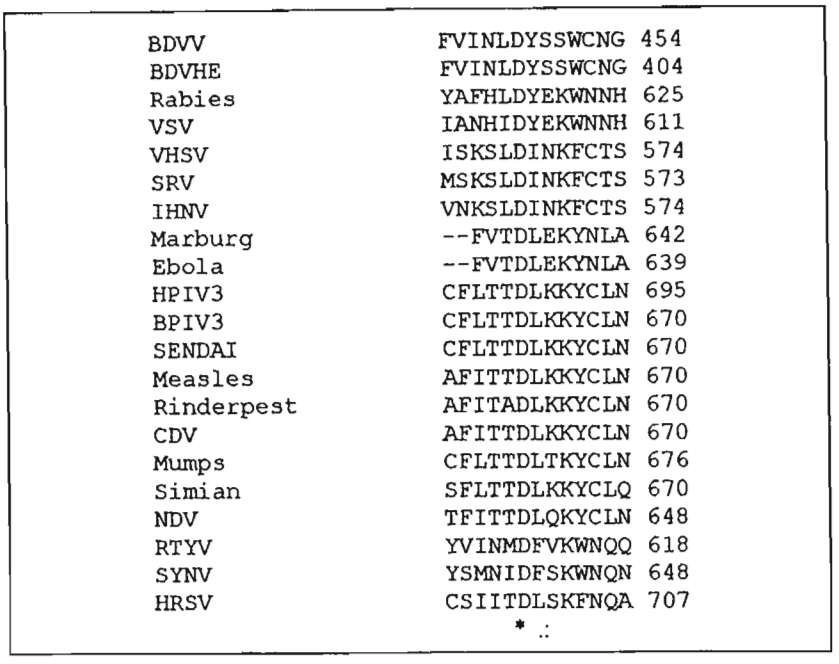


Figure 2.3 Example of a Clustal X multiple alignment showing a fully conserved codon (*), a strongly conserved codon (:) and a weakly conserved codon (.).

In order to validate new results obtained by *Clustal X*, two different alignment programmes, *DIALIGN* and *MEME,* were utilized. *DIALIGN* in contrast to *Clustal X* uses a similarity Blosum (Henikoff and Henikoff, 1992) matrix rather than a distance matrix to create a multiple alignment and is available for academic use at http://www.genomatix.de/cgi-bin/DIALIGN/DIALIGN.pl. The Blosum matrix was created by Henikoff and Henikoff (1992) in order to obtain an increased measure of differences between two proteins, specifically intended for more distantly related proteins.

MEME in contrast to both *DIALIGN* and *Clustal X* utilizes motif based hidden markov modelling of aligned sequences in order to aid in detecting structural motifs and is available at the following server: http://MEME.sdsc.edu/MEME/website/intro.html

The following steps and *DIALIGN* software parameters were followed for generating the RNA virus sequence alignment:

1. **A text file containing the sequences in FASTA format was uploaded to the *genomix* server.**

2. **A threshold value of 5 was chosen.**

The threshold value referred to a number used in order to reduce the number of diagonals under consideration. Diagnols in this case referred to alignments from pairs of similar segments. In contrast to *Clustal X DIALIGN* did not utilize discriminatory gap parameters instead it relied on a threshold value set by the user.

3. **A similarity matrix, Blosum, was implemented.**

The Blosum (Henikoff and Henikoff, 1992) similarity matrix was the default option of the *DIALIGN* programme and was implemented in order to align the sequences.

4. ***DIALIGN* created a file containing the alignment of the input sequences.**

The *DIALIGN* alignment file generated (Figure 2.4) contained the following symbols, which indicated the degree of conservation. Capital letters denoted aligned residues. Lower case letters denoted residues not aligned by DIALIGN. Thus, if a lower case letter appeared above any other letter, this was pure chance. Gaps were denoted by '-'. The number of plus signs below the alignment reflected the degree of local similarity between all the sequences. Regions of maximum similarity between all sequences were represented by 10 plus signs.

```
MoMLV       39   ---DAGT--- --SAQRAELI ALTQA-L--K MAEGKK-L-- NV-YTDSRYA
17.6        31   -SRTLNEHEI NySTIEKELL AIVWA-T--K TF--RH-Y-- LL-GRHFEIS
HBV         29   FSAPLPIHT- ------AELL AACFArS--R SG--AN-I-- -I-GTDNSVV
E.coli      39   YTRTTNNR-- ------MELM AAIVAlealK EH--CE-V-- -IlSTDSQYV
                  ++++++++    ++++++++ +++++ +  + ++  ++ +   ++ +++++++
                  +++++++     ++++++++ +++++      ++     +   ++ +++++++
                   ++++++     ++++++++ +++++                  ++ +++++++
                    ++++      ++++++++ +++++                  ++ +++++++
                     +++          ++++ +++++                   + +++++++
                                  ++++ +++++                   + +++++++
                                  ++++ ++++                    + ++++++
                                   +++ ++++                    + ++++++
                                   +++ +++                        +++++
                                   +++ ++                          ++++
                                   +++ ++                             +
                                   +++ ++
```

Figure 2.4 Example of a *DIALIGN* multiple alignment showing aligned residues (capital letters), non-aligned residues (lower case letters) and gaps (-). Regions of maximum similarity are represented by plus signs.

The following steps and *MEME* software parameters were followed for generating the RNA virus sequence alignment:

1. **A text file containing the sequences in FASTA format was uploaded to the *MEME* server (Figure 2.5).**

An overview of the parameter choices is given in Figure 2.5.

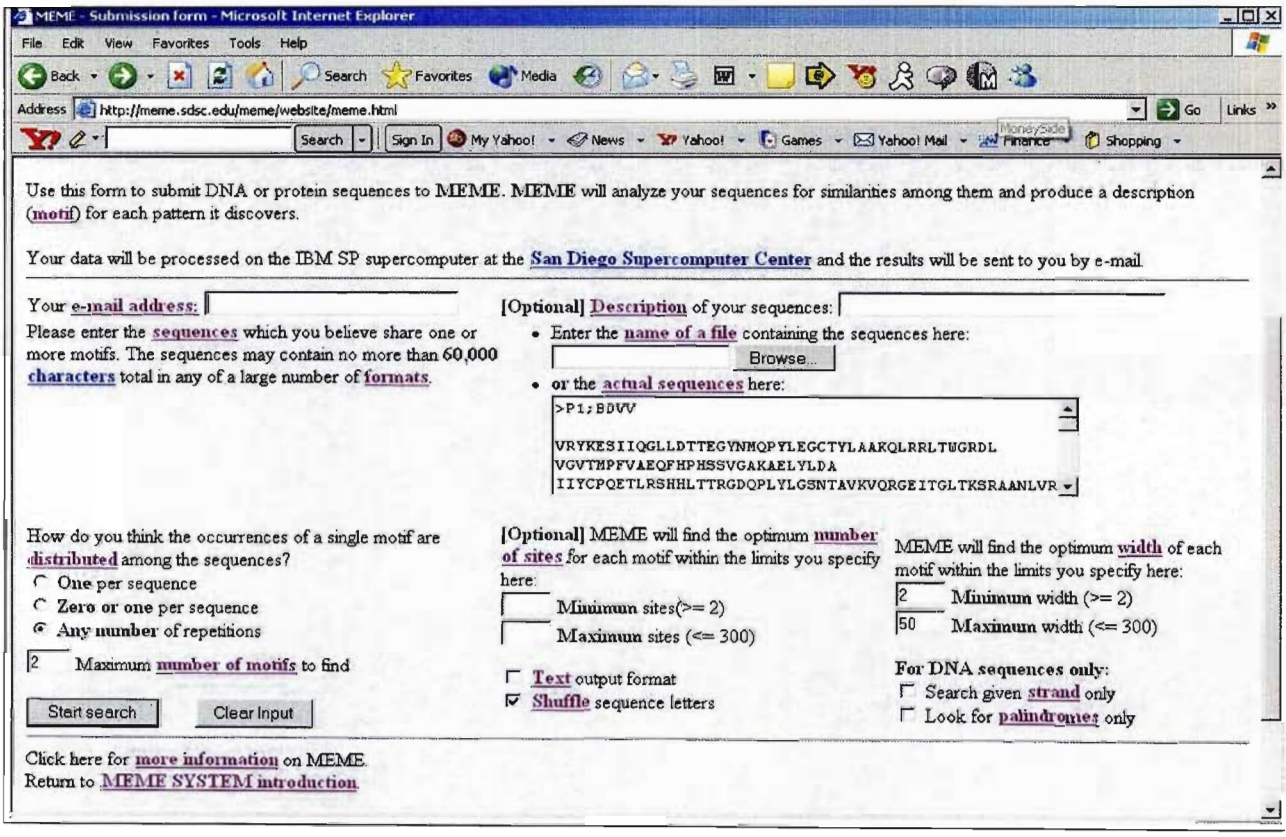


Figure 2.5 Screen of the *MEME* server showing the possible parameters.

2. **The occurrence of the *MEME* motif distribution was selected.**

The selection of the type of distribution improved the sensitivity and quality of the motif search. The motif distribution menu (Figure 2.5) presented the following choices:

1. One per sequence
2. Zero or one per sequence
3. Any number of repetitions

The first option assumed that each sequence in the data set contained exactly one occurrence of a motif. This option was the fastest and most sensitive but the motifs returned by *MEME* may be "blurry" if any of the sequences are missing the motifs.

The second option assumed that each sequence may contain at the most one occurrence of each motif. This option may be useful if it is suspected that some motifs may be missing from the sequence. In that case, the motifs found were more accurate than using the first option.

The third option assumed that each sequence may contain any number of non-overlapping occurrences of each motif. This option was useful when it was suspected that motifs repeat multiple times within a single sequence. In that case, the motifs found were much more accurate than using one of the other options.

Option 1 was selected for the purpose of this study as the multiple alignment of *Clustal X* and *DIALIGN* indicated one occurrence of each motif.

3. The *MEME* output alignment file was analysed.

For each motif that was discovered in the training set, *MEME* provided the following information:

Simplified Position-Specific Probability Matrix

MEME motifs were represented by position-specific probability matrices that specified the probability that each possible letter appeared at each possible position in an occurrence of the motif. The simplified motif showed the letter probabilities multiplied by 10 rounded to the nearest integer.

Multilevel Consensus Sequence

The multilevel consensus sequence corresponding to the motif aided in remembering and understanding the motif was calculated from the motif position-specific probability matrix as follows:

for each column of the motif, the letters in the alphabet were sorted in decreasing order by the probability with which they were expected to occur in that position of motif occurrences. The sorted letters were then printed vertically with the most probable letter on top. Only letters with probabilities of 0.2 or higher at that position in the motif were printed. An example of a multilevel consensus sequence of a motif is given in Figure 2.6.

```
Multilevel  TTATGTGAACGACGTCACACT
consensus   AA  T A G A GA     AA
sequence          T C TT      T
```

Figure 2.6 Example of a *MEME* multilevel consensus sequence.

This multilevel consensus sequence indicated that: firstly, the most likely form of the motif can be read from the top line as TTATGTGAACGACGTCACACT. Secondly, that only letter A had probability more than 0.2 in position 3 of the motif, both T and A had probability greater than 0.2 in position 1. Thirdly, a rough approximation of the motif could be made by converting the multilevel consensus sequence into the Prosite (database of protein families and domains) signature: [TA]-[TA]-A-T-[GT]-[T]-[GA]-A-[AGT]-C-[GAC]-A-[CGT]-[GAT]-T-C-A-C-A-[CAT]-[TA].

Occurrences of the Motif

MEME displayed the occurrences (sites) of the motif in the training set. The sites were shown aligned with each other, and the ten sequence positions preceding and following each site were also shown. Each site was identified by the name of the sequence where it occurred and the position in the sequence where the site began. The sites were listed in order of increasing statistical significance (p-value). The p-value of a site was computed from the match score of the site with the position specific scoring matrix for the motif. The p-value gave the probability of a random string (generated from the background letter frequencies) having the same match score or higher.

2.2.3 Phylogenetic procedures

The phylogenetic status of BDV relative to other negative stranded RNA viruses was assessed by employing a compilation of phylogenetic software programmes, *Phylip* (version 3.5c) (Felsenstein, 1993). *Phylip* was freely available from the following website: http://evolution.genetics.washington.edu/phylip.html and encompassed a diverse package of phylogenetic software. An overview of the phylogenetic analyses steps followed in this investigation are outlined in figure 2.7.

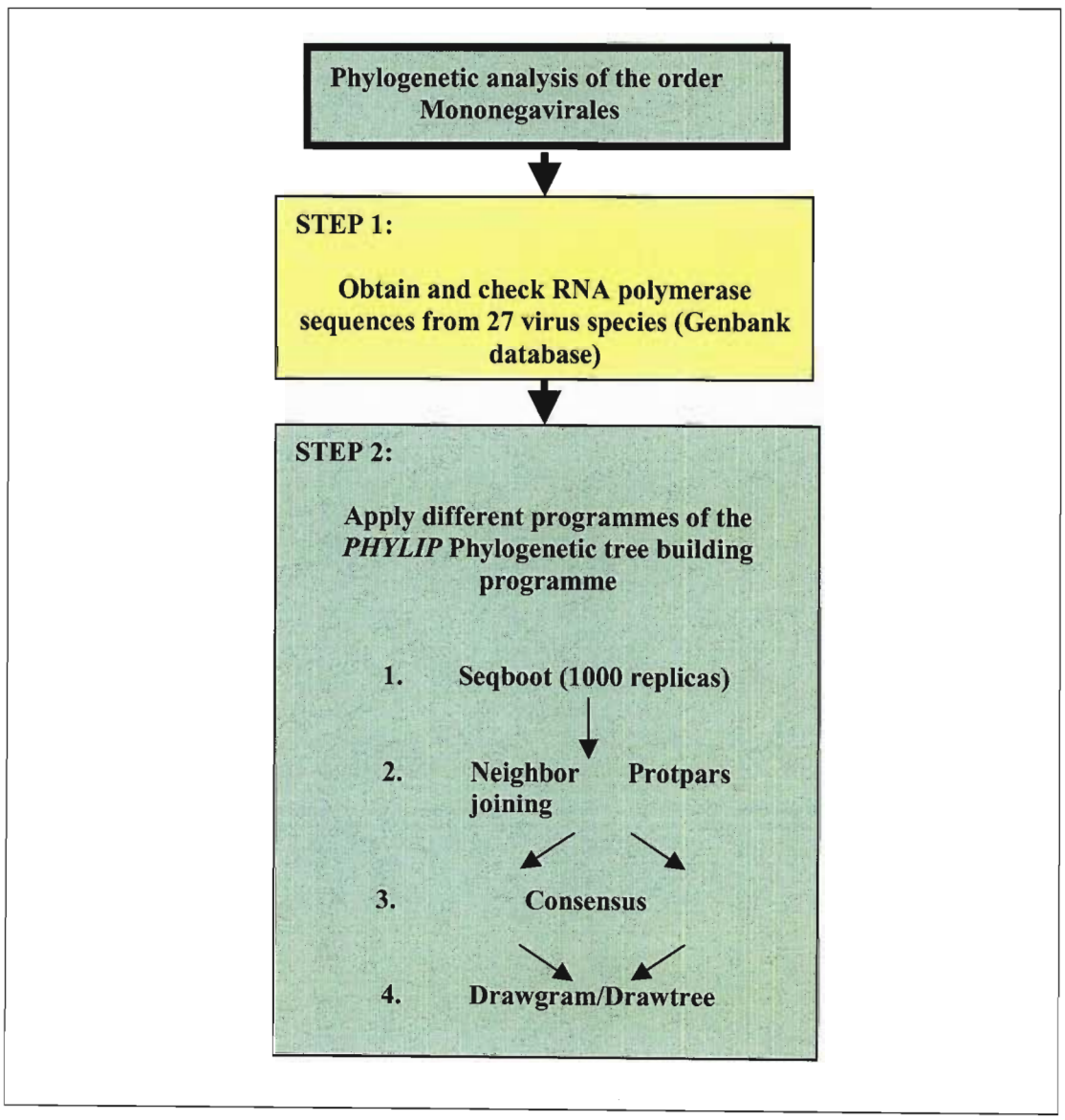


Figure 2.7 Flow diagram of the steps involved in the phylogenetic analysis of RNA negative stranded viruses.

The following steps and software parameters were utilized in the phylogenetic analysis:

1. **The RNA virus sequences were obtained as described in section 2.2.1.**

Alignment of these sequences was carried out by the *Clustal X* programme as detailed in section 2.2.

2. **The *Phylip* bootstrapping programme *Seqboot* was applied to the RNA virus sequences.**

The reliability of different phylogenetic groupings was evaluated by using the bootstrap analysis offered by the *Seqboot* programme, (Figure 2.8), (1000 bootstrap replications). *Seqboot* reads in a data set, and produces multiple data sets from it by bootstrap resampling. Figure 2.8 provides an outline of the programme parameters that were available.

Bootstrapping Sequence Algorithm version 3.573c

Settings for this run:

D	**Sequence, Morph, Rest, Gene freq**	Molecular Sequence
J	**Bootstrap, Jacknife or Permute**	Bootstrap
R	**How many replicates?**	1000
I	**Input sequences interleaved**	Yes
O	**Terminal type (IBM, PC, ANSI)**	PC
1	**Print out the data at the start of the run**	No
2	**Print indications of progress of run**	Yes

Are these settings correct (type Y or letter for change)

Figure 2.8 Programme parameters available for the *Seqboot* programme used for bootstrapping the RNA virus sequences.

The parameters used for the *Seqboot* analysis were selected as follows:

1. Molecular sequences
2. Bootstrap analysis
3. 1000 replicates
4. Input sequences interleaved

3. Two *Phylip* tree construction programmes *Neighbor joining* and *Protpars* were applied to the bootstrap file

The *Neighbor* programme, (Figure 2.9) implemented the neighbour joining method (Saitou and Nei, 1987) which sequentially found the nearest pairs of neighbouring sequences that gave the shortest overall length of the tree. This produced an unrooted tree without the assumption of a clock. The advantage of using *Neighbor* is that it analyses large data sets in a very short amount of time in contrast to *Protpars*.

Neighbor-joining/UPGMA method version 3.5

Settings for this run:

U Search for best tree	**Yes**
J Randomize input order of Seq.	**Yes**
O Outgroup Root?	**Use as outgroup species 1**
T Use threshold parsimony?	**No**
I Input sequences interleaved?	**No**
M Analyze multiple data sets?	**No**
0 Terminal type (IBM, PC, ANSI)	**PC**
1 Print out the data at the start of the run	**No**
2 Print indications of progress of run	**Yes**
3 Print out tree	**Yes**
4 Write out tree file onto tree file	**Yes**

Are these settings correct (type Y or letter for change)

Figure 2.9 Programme parameters available for the *Neighbor* tree building programme.

The parameters for the *Neighbor* programme were selected in the following order:

1. Neighbor-joining tree
2. No outgroup selected
3. Randomize input order of species: Yes
4. Analyze multiple data sets: 1000

The *Protpars* programme, (Figure 2.10), implemented the parsimony tree building method, which was intermediate between Eck and Dayhoff's, (1966) method of allowing transitions between all amino acids and counting those, and Fitch's, (1971) method of counting the number of nucleotide changes that would be needed to evolve the protein sequence. *Protpars* in contrast to *Neighbor* used global rearrangement, which means that sub trees were removed from the tree and put back on in all possible ways so as to have a better chance of finding a more accurate tree. This method however, was very time consuming.

Protein Parsimony algorithm, version 3.573c

Settings for this run:

N Neighbor joining or UPGMA tree?	**Neighbor joining**
O Outgroup Root?	**Use as outgroup species 1**
L Lower triangular data matrix?	**No**
R Upper triangular data matrix?	**No**
T Randmize input order of species?	**No use input order**
M Analyze multiple data sets?	**No**
0 Terminal type (IBM, PC, ANSI)	**PC**
1 Print out the data at the start of the run	**No**
2 Print indications of progress of run	**Yes**
3 Print out tree	**Yes**
4 Write out tree file onto tree file	**Yes**

Are these settings correct (type Y or letter for change)

Figure 2.10 Programme parameters available for the *Protpars* programme.

The parameters were selected in the following order:

1. Search for best tree: Yes
2. Randomize input order of sequences: Yes
3. Outgroup Root: None selected
4. Use ordinary parsimony
5. Analyze multiple data sets: yes 1000 sets
6. Input sequences interleaved: Yes

4. **The reliability of different phylogenetic tree groupings was evaluated by using the *Consenus* programme (Figure 2.11).**

The *Consensus* programme was used as the final step in the bootstrap analyses.

Majority-rule and strict consensus tree programme, version 3.5.7.3c

Settings for this run:

O Outgroup Root?	**Use as outgroup species 1**
R Trees to be treated as Rooted?	**No**
0 Terminal type (IBM, PC, ANSI)	**PC**
1 Print out the data at the start of the run	**No**
2 Print indications of progress of run	**Yes**
3 Print out tree	**Yes**
4 Write out tree file onto tree file	**Yes**

Are these settings correct (type Y or letter for change)

Figure 2.11 Programme Parameters of the *Consensus* programme used in the phylogenetic analysis.

The parameters for the *Consensus* programme were selected in the following order:

1. Outgroup root: None
2. Trees to be treated as rooted:

5. The subsequent trees were visualized with the *Drawgram* and *Drawtree* programmes.

Drawgram plots rooted phylogenies, cladograms, and phenograms in a wide variety of user-controllable formats. The programme was interactive and allowed previewing of the tree on PC graphics screens. *Drawtree* was similar to *Drawgram* however, plotted unrooted phylogenies.

2.3 DETERMINATION OF THE EVOLUTIONARY MUTATION RATE OF BDV

One of the more important estimations for molecular evolutionary analysis is the quantification of the number of substitutions per site between nucleotide sequences (Ina, 1995; Muse, 1996). It is imperative to estimate the numbers of synonymous (ds) (silent mutations) and non-synonymous (dn) (amino acid replacing) substitutions per site separately, as estimates of ds and dn are used not only for reconstruction of phylogenetic trees but also as a statistical test for a neutral theory of molecular evolution, (Ina, 1995). Various methods exist for estimating synonymous and non-synonymous mutations, however until recently none of these were based on a valid statistical footing (Muse, 1996). Some of these methods have been utilized for an estimation of the mutation of RNA virus sequences (Saitou, 1987; Haydon *et al.*, 1998), with several methods exhibiting overestimation and underestimation of the number of synonymous sites and non-synonymous sites respectively, in particular with regard to variable sequence data. New methods have been developed which aim to correct these biases (Ina, 1995). These methods have been used in this study to detect the mode of evolution i.e. neutral or directed.

2.3.1 Materials

The nucleotide and amino sequences of four open reading frames (ORFs) encoding the nucleoprotein (p40), phosphoprotein (p24), matrix (gp18) and glycoprotein (p57) of (i) six different BDV host species (Table 2.2) and (ii) three BDV strains (Table 2.3) were obtained from the public NCBI

Genbank database http://www.ncbi.nlm.nih.gov/Genbank. The BDV strains in table 2.3 refer to two animal prototype viruses, BDV V and BDV He; and one BDV WR98 strain that was isolated from human granulocytes (Planz *et al.*, 1999). In some cases the completely sequenced protein was not available from the genbank database, and thus the partially sequenced protein sequence was used. The *Borna disease virus* strain V (BDV V) was utilized as the reference strain for pairwise comparison of the mutation rates as outlined in section 2.3.2. The BDV human strain used in Table 2.2 represents one of the first human strains isolated from patients suffering from acute depression (Bode *et al.*, 1996). Host isolates (Table 2.2) in this study refers to sequences obtained from different host species. These sequences were only partially sequenced and not classified into the three known BDV strains, BDV V, BDV He and BDV W R98. Inter-species analysis in this study refers to comparison of host isolates with the BDV V strain, whereas intra-species analysis refers to comparison of the three known BDV strains with each other.

Table 2.2 Genbank accession numbers of four *Borna disease virus* proteins from six different host species.

	Protein accession numbers				
BDV host species	**p40**[b]	**p57**	**gp18**	**p24**	**Year submitted**
BDV V[a]	U04608	U04608	U04608	U04608	1994
Feline	U94863	U94866	U94865	U94864	1997
Canine	U94879	U94882	U94881	U94880	1997
Ovine	U95875	U94878	U94877	U94876	1997
Human	U58594	U58597	U58596	U58595	1996
Assine	U94871	U94874	U94873	U94872	1997
Total no. of codons	146	503	142	201	

[a] BDV V (*Borna disease virus* strain V - equine isolate, used as a reference strain for pairwise comparison).
[b] Note that only 146 codons of the P40 protein for all six species (partial cds) were available from the genbank database; polymerase and X protein sequences could not be obtained.

Table 2.3 Genbank accession numbers of five *Borna disease virus* proteins from three BDV strains.

	Protein Accession Numbers					
BDV strain	**p40**	**p10**	**p24**	**gp18**	**p57**	**Year isolated**
BDV RW98	AF158629	AF158631	AF158630	AF158632	AF158633	1999
BDV He	L27077	L27077	L27077	L27077	L27077	1994
BDV V	U04608	U04608	U04608	U04608	U04608	1994

2.3.2 Method for the determination of the evolutionary mutation rate.

Pairwise comparison of BDV protein sequences was performed, by using the BDV V strain as a reference strain, and observed nucleotide differences between two sequences were classified into four categories, (Figure 2.12). When the two compared codons differed at one position the classification was obvious. If they differed in two or three positions, there were two or six pathways along which one codon could change into another. All of these pathways occur when classifying different mutations:

1. Synonymous transitions: The substitution of a purine for a purine or a pyrimidine for a pyrimidine.
2. Synonymous transversions (the substitution of a purine for a pyrimidine or vice versa).
3. Non-synonymous transitions.
4. Non-synonymous transversions.

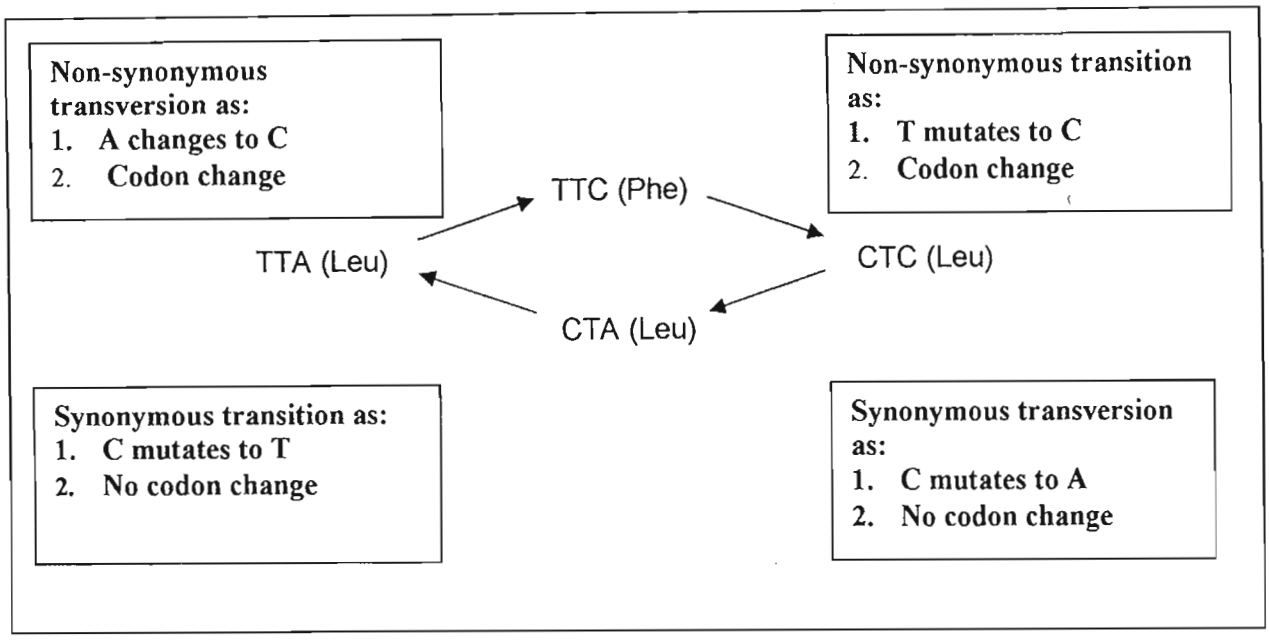


Figure 2.12 Representation illustrating the classification method for differentiating between synonymous and non-synonymous differences between a pair of codons.

The nucleotide substitutions were converted into a percentage by dividing the number of nucleotide substitutions for each protein by the number of codons specific for each protein and multiplying the amount by a 100. This was done in order to enable comparison between the host and virus sequence analysis.

The number of synonymous (ds) substitutions (silent mutations) and non-synonymous (dn) mutations (amino acid replacing mutation) per synonymous and non-synonymous site at the i^{th} nucleotide position was quantified by a site-by-site comparison of sequence pairs (Table 2.2 and Table 2.3) using the method developed by Ina, (1995). The method used to estimate the synonymous (ds) and non-synonymous (dn) ratio is briefly summarized below, followed by a table defining the symbols used in the formula and a detailed description of the formulae used.

1. The Transition/Transversion (TS) ratio (i.e. the number of transitions/ the number of transversions occurring in a pair of codons) using base codon frequencies from the real data through a pairwise sequence comparison was estimated.

2. The number of synonymous and non-synonymous sites respectively was counted using the estimated Transition/Transversion (TS) ratio and the observed base (codon) frequencies.

3. The number of synonymous and non-synonymous differences (both transition and transversion) was counted using the transition/transversion ratio and the codon frequencies.

4. The synonymous and non-synonymous distance ratio was calculated by correcting for multiple hits by using Kimura's (1983) two parameter model.

Table 2.4 Definition of major symbols used in the synonymous and non-synonymous mutation formula.

Symbol	Definition
S	**Number of synonymous sites in a sequence (S=si)**
N	**Number of non-synonymous sites in a sequence (N= 3L-S)**
L	**Number of codons in the sequence**
ds	**Number of synonymous substitutions per synonymous site**
dn	**Number of non-synonymous substitutions per non-synonymous site**
$\hat{S}_{Ts}$	**Synonymous transitional difference between 2 codons**
$\hat{S}_{Tv}$	**Synonymous transversional difference between 2 codons**
N_{Ts}	**Non-synonymous transitional difference between 2 codons**
N_{Tv}	**Non-synonymous transversional difference between 2 codons**
Ps	**Represents the ratio of synonymous transitional differences ($P_s = \hat{S}_{Ts}\backslash\hat{S}$)**
Qs	**Represents the ratio of synonymous transversional differences ($Qs=\hat{S}_{Tv}\backslash\hat{S}$)**

The average transition/transversion ratio (TS) for synonymous base substitutions was estimated from the ratio of observed unambiguously synonymous transitional and transversional differences between codons that differ at only one position from all possible pairwise sequence comparisons.

The total numbers of non-synonymous and synonymous sites for a given nucleotide sequence of **L** codons (number of codons in a given sequence) were estimated by $\hat{S}=\Sigma\hat{s}i$ and $N= 3L-\hat{S}$ respectively,

where Σ represented the summation of $\hat{s}i$ (the number of synonymous sites for codon i) over all codons in the nucleotide sequence. To avoid under- and overestimation of substitution rates the numbers of synonymous and non-synonymous differences were estimated by considering transitional and transversional changes separately (Ina, 1995).

$\hat{S}_{Ts,\,ij}$ and $\hat{S}_{Tv,\,ij}$ denoted the numbers of synonymous transitional and transversional differences between codons i and j, respectively. Furthermore, $n_{TS,ij}$ and $n_{TV,ij}$ represented the numbers of non-synonymous transitional and transversional differences, respectively, between codons i and j. When only one nucleotide difference was observed between a pair of codons, only one synonymous or non-synonymous difference was attributed. This difference was further sub divided into one transversional and transitional difference. When two or three nucleotide differences were observed between a pair of codons, assignment of the differences was complex as two or more possible pathways between the codon pairs were involved. If a stop codon was involved in a pathway, the pathway was eliminated from assignment of nucleotide differences. As substitution of stop codons does not occur such a pathway does not appear in the evolutionary process (Ina, 1995).

The total numbers of synonymous transitional ($\hat{S}_{Ts}$) and transversional ($\hat{S}_{Tv}$) differences between two nucleotide sequences was quantified by the following formulae $\hat{S}_{Ts}= \Sigma\hat{S}_{Ts,ij}$ and $\hat{S}_{Tv}= \Sigma\hat{S}_{Tv,ij,}$ where Σ represented the summation of $\hat{S}_{Ts,ij}$ and $\hat{S}_{Tv,ij}$, over all codon pairs between the two nucleotide sequences. Similarly, the total numbers of non-synonymous transitional (N_{Ts}) and transversional (N_{Tv}) differences between the two nucleotide sequences were estimated by applying $N_{Ts}=\Sigma\ n_{Ts,ij}$ and $N_{Tv}=\Sigma n_{Tv,ij}$, respectively.

The ratio of synonymous transitional (P_s) and transversional (Q_s) differences were estimated by:

$$P_s= \hat{S}_{Ts}\backslash\hat{S};\ Qs=\hat{S}_{Tv}\backslash\hat{S}$$

It was assumed that mutations follow Kimura's (1983) two parameter model ($\lambda_{TC}=\lambda_{CT}=\lambda_{AG}= \lambda_{GA}$ and $\lambda_{TA}=\lambda_{TG}=\lambda_{CA}=\lambda_{CG}=\lambda_{AT}=\lambda_{AC}=\lambda_{GT}=\lambda_{GC}$). Therefore in order to correct multiple substitutions Kimura's (1983) formula was applied. The estimate of d_s was obtained by:

$$d_s= -1\backslash 2\ \ln\left((1-2P_s-Q_s)\sqrt{\ }1-2Q_s\right)$$

To calculate the variance of ds, a derived formula for the sampling variance was used (Kimura, 1983).

$$V(d_s) = (\partial ds/\partial P_s)^2 V(P_s) + (\partial ds\backslash\partial Q_s) V(Q_s) + 2 (\partial d_s\backslash\partial P_s . \partial d_s\backslash\partial Q_s)$$
$$= 1\backslash\hat{S} (a^2P_s + b^2Q_s - (aP_s+bQ_s)^2)$$

where
$$a= 1\backslash(1-2P_s-Q_s)$$
$$b= 0.5 (1\backslash 1-2P_s-Q_s + 1\backslash 1-2Q_s)$$

2.3.3 Investigation of a neutral model of evolution.

To determine whether or not the observed frequency distribution of multiple hit sites was consistent with that expected from a simple Poisson model of synonymous substitution (Haydon *et al.*, 1998), for each data (Table 2.2 and 2.3) set it was estimated how many 1st and 3rd codon position sites received 0, 1, 2, and 3 unambiguously identifiable synonymous substitutions. This information was used to calculate average synonymous hit rates for 1st and 3rd codon positions combined. These per site averages were used to construct the expected frequency distribution of multiple hit sites if the same number of substitution events over the appropriate number of sites according to a Poisson distribution:

Poisson parameter λ = the number of 1^{st} and 3^{rd} codon positions which have received 0, 1, 2 and 3 identifiable synonymous substitutions.

If **X** is a Poisson random variable with parameter λ, then

$$PX(K) = \frac{(\lambda)^K e^{-\lambda}}{K!}$$

Where **k = 0,1,2, ...**

The two site distributions were compared using a Chi-square test. If the expected multiple hit distribution was found to differ from the observed distribution it was concluded that the actual process of synonymous substitution was not homogenous and independent over all sites, as assumed by the Poisson distribution, (Haydon *et al.*, 1998).

CHAPTER 3

RESULTS OF THE *BORNA DISEASE VIRUS* (BDV) ANALYSES

3.1 INTRODUCTION

The results of this investigation will be discussed under the following general major headings:

1. Alignment and phylogenetic analysis of negative stranded RNA viruses with the aim of determining the most suitable BDV taxonomic classification.

2. Estimation of the mutation rate of a number of BDV host species and three different BDV strains, as well as their mode of evolution.

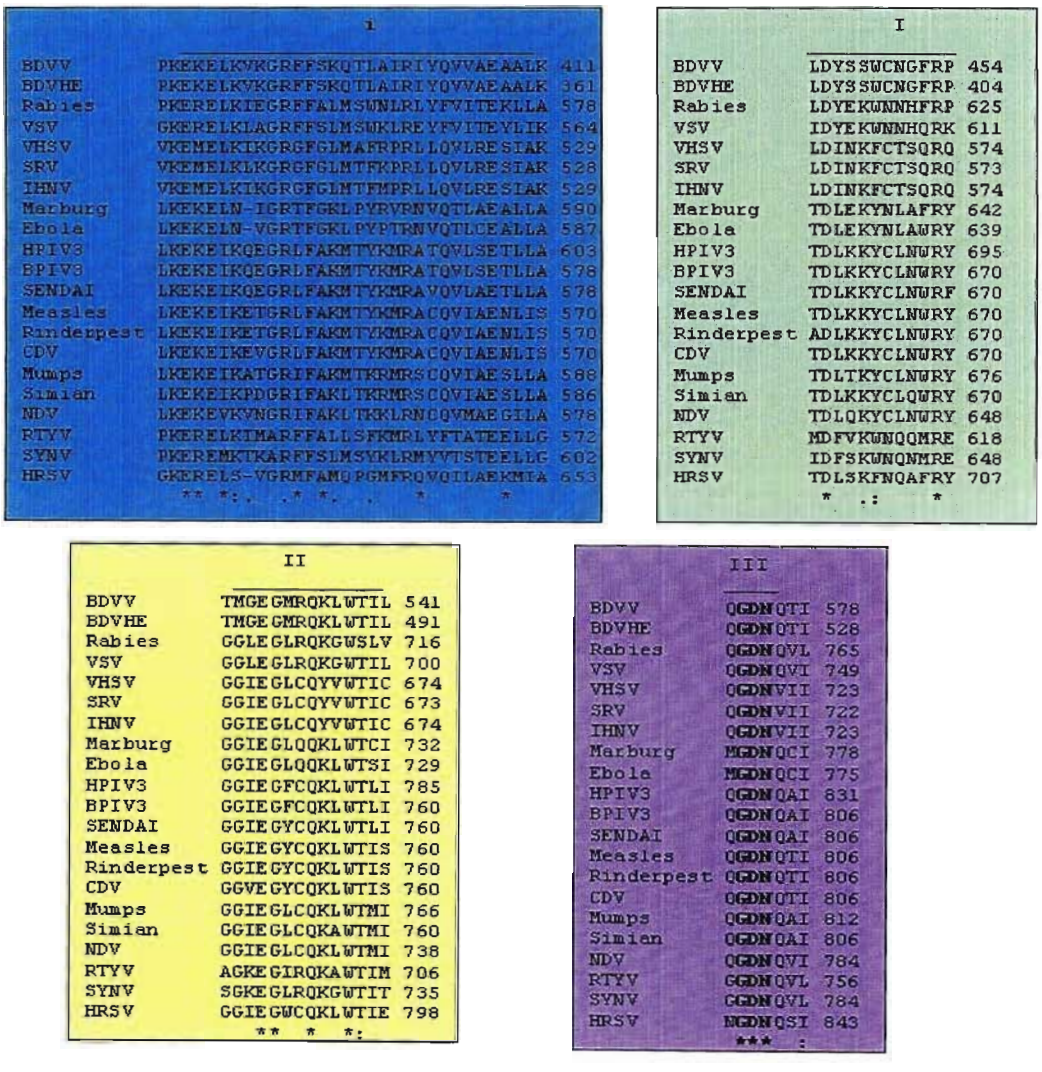
3.2 ALIGNMENT AND PHYLOGENETIC ANALYSIS OF RNA NEGATIVE STRANDED VIRUSES: TAXONOMIC CLASSIFICATION

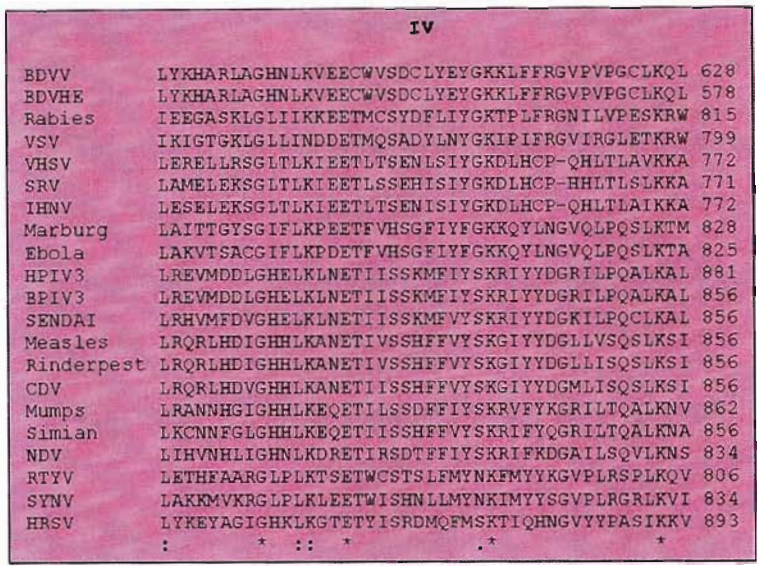
The diverse nature of RNA viruses has created difficulties in identifying sequence relationships between RNA viruses which in turn has complicated taxonomic classification. The first researchers identifying sequence relationships from several positive stranded RNA viruses discovered a particular conserved positive stranded RNA motif consisting of a Gly-Asp-Asp (GDD) triplet, prevalent in the polymerase protein. Further research into the classification of negative stranded viruses has indicated that in contrast to the GDD motif found in positive stranded viruses the GDD motif is present as a Gly-Asp-Asn (GDN) motif in negative stranded viruses. The GDD and GDN motif were found in several different RNA virus species, suggesting that many viral species, across many hosts, and with great variety in structural and infective features and genomic organization are related and provided an accurate basis of classification in a taxonomical classification.

3.2.1 Alignment analysis

The polymerase protein sequences of the *Borna disease virus* and a number of other negative stranded RNA viruses were selected to determine the taxonomic status of BDV. The sequences were aligned with the latest version of the *Clustal X* (version 1.8, 2000) software package (Thompson, Higgins and Gibson, 1994).

Figure 3.1 *Clustal X* alignments of the amino acid sequences of the conserved regions of 21 viral negative stranded RNA dependent RNA polymerase proteins. The consensus patterns of conserved amino acid residues are shown in the lines above the respective blocks of sequences. Asterisks (*) denote positions, which have a single, fully conserved residue; colon (:) denotes positions which have a strongly conserved group and full stop (.) denotes positions, which have a weakly conserved group. Residues in bold, indicate the GDN conserved residues. Coloured blocks indicate conserved polymerase motifs (i,I,II,III,IV,v,vi).





The *Clustal X* multiple alignment consists of seven distinct blocks of amino acid residues (Figure 3.1), which could be considered as extensively conserved motifs. Motifs i-IV have been described previously (designated as A, B, C and D by Briese *et al.*, 1994) whereas motifs v and vi are newly identified. Several amino acid residues were strictly conserved in all polymerase sequences, namely **Lys126, Glu127, Glu129, Arg 129**, **Phe137, Arg136** and **Glu144** in motif i; **Asp198** and **Arg206** in motif I, **Glu271, Gly 272**, **Asp 275** and **Trp 178** in motif II, **Gly 303, Asp 204** and **Asn305** in motif III which form part of the well known GDN motif; **Arg 648**, **Trp 651, Gly 663** and **Pro 667** in motif v, **Gly 710 and Thr713** in motif vi.

In order to confirm the presence of the newly discovered v and vi motifs two different alignment programmes, *DIALIGN* (version 3.0, 2002) and *MEME* (version 3.0, 2001), were utilized to align the negative stranded RNA viruses.

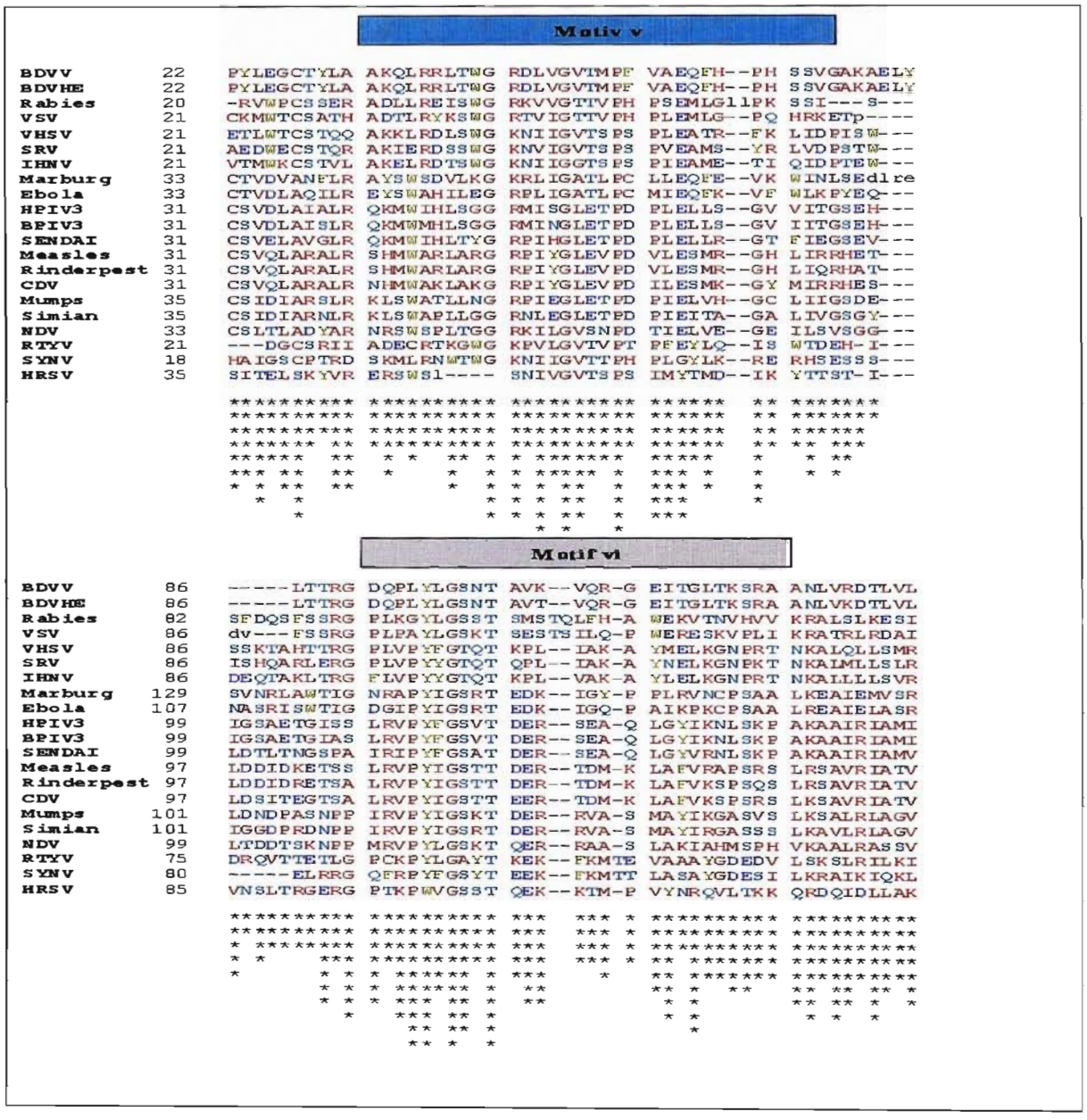


Figure 3.2 *DIALIGN* alignment of the amino acid sequences of the conserved regions of 21 viral negative stranded RNA dependent RNA polymerase proteins. The consensus patterns of conserved amino acid residues are shown in the lines below the respective blocks of sequences. Aligned residues are indicated by capital letters, non-aligned residues by lower case residues) and gaps by (-). Regions of maximum similarity are represented by plus signs. Coloured blocks indicate conserved polymerase motifs (v,vi).

The alignment generated by *DIALIGN* (Figure 3.2) confirms the results from the *Clustal X (*Figure 3.1) alignment. Motifs v and vi exhibit a higher degree of conserved residues when aligned with the *DIALIGN* programme.

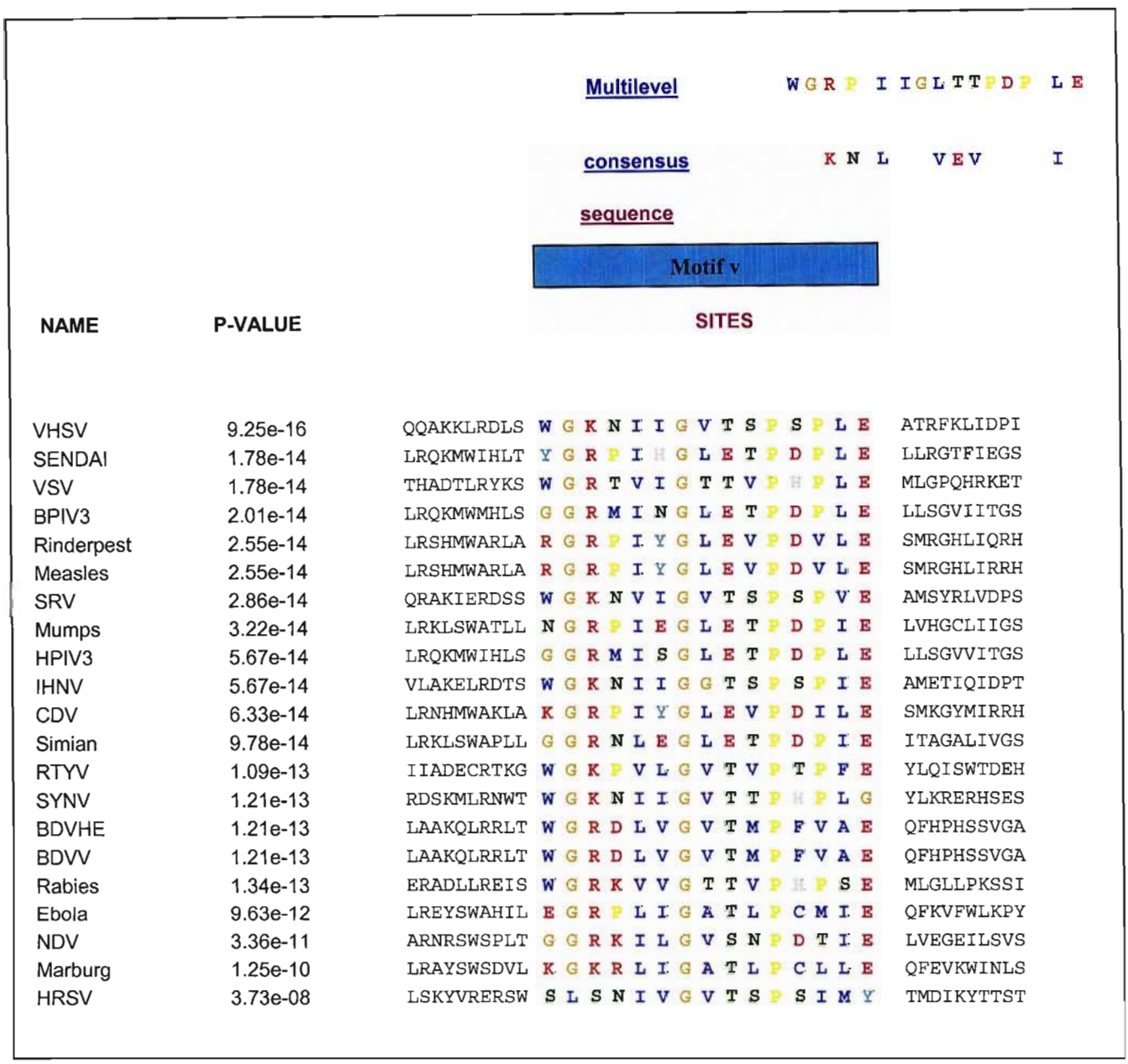


Figure 3.3a *MEME* alignment of the amino acid sequences of the conserved regions of 21 viral negative stranded RNA dependent RNA polymerase proteins. The multilevel consensus sequence indicates that the most likely form of the motif can be read from the top line as WGRPIGLTTPDLE. Aligned residues are indicated in colours and the sites are listed in order of increasing statistical significance (p-value). The coloured block indicates conserved polymerase motif v.

The MEME alignment (Figure 3.3a and 3.3b) confirms the results obtained from the *CLUSTAL X* (Figure 3.1) and *DIALIGN* (Figure 3.2) programme. Motif v and vi are structurally conserved and BDV V, BDV HE and the Rabies polymerase protein have a high probability of containing these motifs.

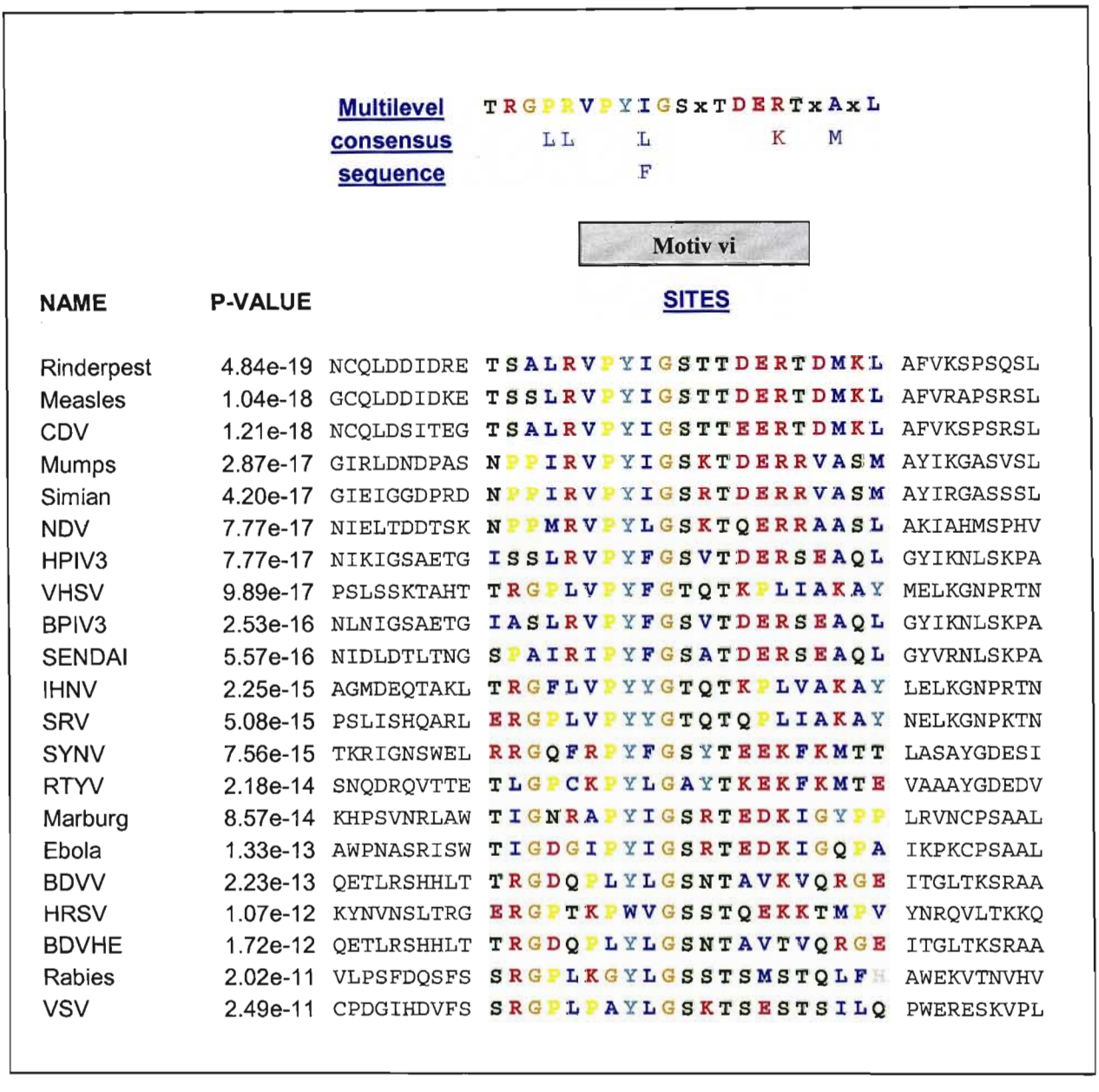


Figure 3.3b *MEME* alignment of the amino acid sequences of the conserved regions of 21 viral negative stranded RNA dependent RNA polymerase proteins. The multilevel consensus sequence indicates that the most likely form of the motif can be read from the top line as TRGPRVPxIGSxTDERTxAxL. Aligned residues are indicated in colours and the sites are listed in order of increasing statistical significance (p-value). The coloured block indicates conserved polymerase motif vi.

The discovery of structural motifs v and vi indicates that these motifs have a particular function. In order to examine the relevance of the newly discovered BDV polymerase protein v and vi motifs in comparison to other BDV proteins, the BDV proteins were aligned with the BDV V and Rabies polymerase protein. It is known that the BDV p40 protein has a nuclear localization signal that interacts with the BDV polymerase protein (Schwemmle *et al.*, 1998). The BDV p40 nucleoprotein was the only protein that aligned between the polymerase protein motifs IV and vi (Figure 3.4).

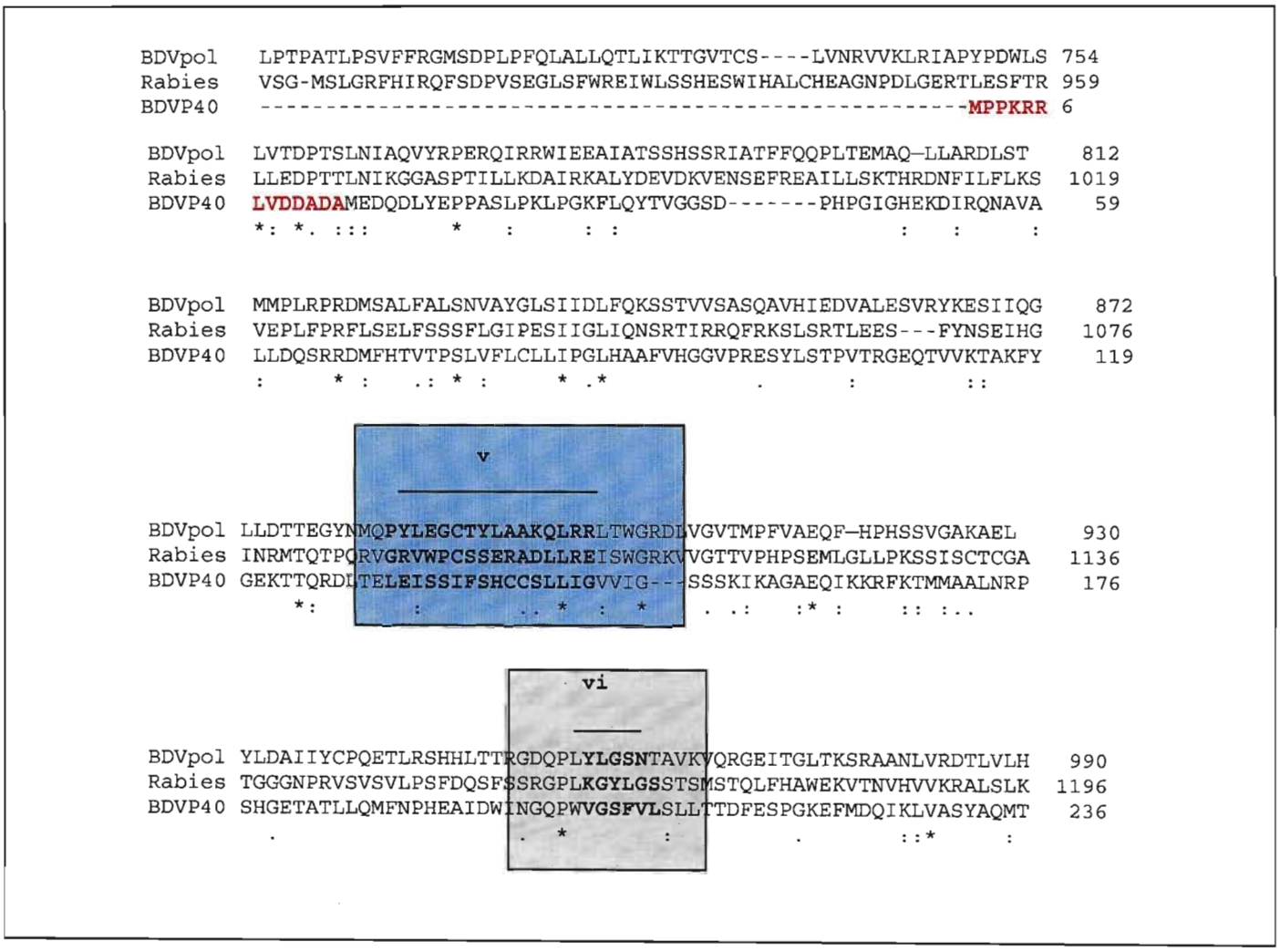


Figure 3.4 *Clustal X* Alignment of the conserved amino acid sequences of the BDV V (BDVpol) and Rabies polymerase protein with the BDV V p40 nucleoprotein. The consensus patterns of conserved amino acid residues are shown in the lines above the respective blocks of sequences. Asterisks (*) denote positions, which have a single, fully conserved residue; colon (:) denotes positions which have a strongly conserved group and full stop denotes positions, which have a weakly conserved group. Coloured blocks indicate conserved polymerase motifs (v,vi). Red residues indicate the nuclear localization signal of the BDV p40 protein.

These alignment results (Figure 3.4) indicate that the BDV p40 protein aligns at its nuclear localization signal (MPPKRRLVDDADA) between motifs IV and vi of the BDV polymerase protein. An overview of the alignment results (Figure 3.1, 3.2, 3.3a, 3.3b and 3.4) with respect to the amino acid sequence of the BDV polymerase protein is provided in Figure 3.5.

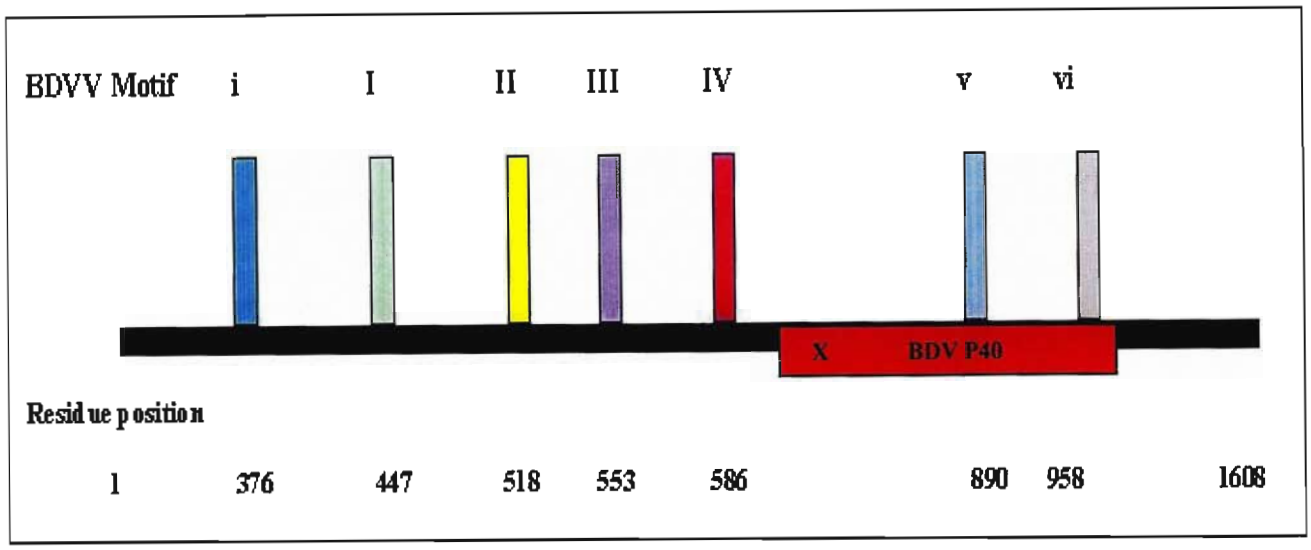


Figure 3.5 Summary of alignment results showing BDV V motif i (light blue), I (green), II (yellow), III (purple), IV (pink) , v (blue) and vi (grey) and the corresponding residue position in the BDV V polymerase protein. The alignment position of the BDV p40 nucleoprotein with the BDV polymerase protein is represented as a red block. X indicates the position of the BDV p40 nuclear localization signal relative to the BDV polymerase protein.

3.2.2 Phylogenetic analysis of the BDV taxonomic status

The phylogenetic status of the BDV virus was assessed by employing programmes available in the *Phylip* compilation of programmes, version 3.5c (Felsenstein, 1993). Two tree construction programmes were implemented: (a) The *Neighbor* programme which utilizes the neighbour joining method (Saitou and Nei, 1987) which sequentially finds the nearest pairs of neighbouring sequences that give the shortest overall length of the tree, (b) the *Protpars* programme which utilizes the parsimony tree building method which is character based and examines each site (character) in a sequence separately (Felsenstein, 1993). The reliability of different phylogenetic groupings was evaluated by using the bootstrap analysis through the *Seqboot* program (1000 bootstrap replications). A consensus of the most probable phylogenetic tree was obtained through

the *Consensus* program and the subsequent trees were visualized with the *Drawgram (produces rooted trees)* and *Drawtree* (produces unrooted trees) programmes. Figure 3.2a and Figure 3.2b present the results of the parsimony analysis with Figure 3.6a indicating a rooted tree and Figure 3.6b indicating an unrooted tree. Figure 3.7 presents the results of the neighbourhood analysis based on a rooted tree.

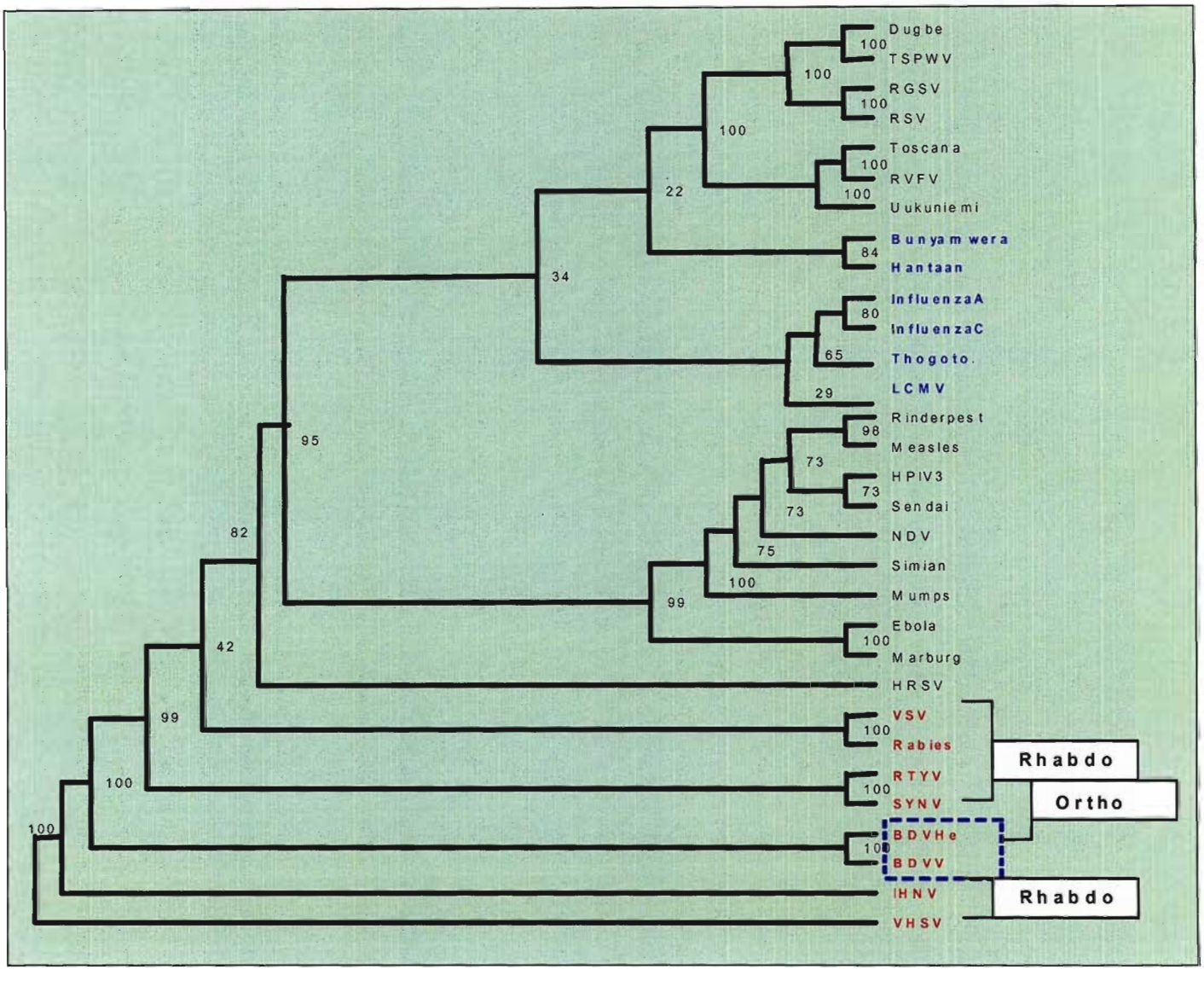


Figure 3.6a Phylogenetic analysis of negative stranded RNA polymerase protein sequences according to a parsimony analysis based on a rooted tree. Segmented viral polymerases are indicated in blue, whereas nonsegmented viral polymerases are indicated in Red. Rhabdo refers to *Rhabdoviridae* and Ortho refers to *Orthomyxoviridae*. The position of BDV is indicated by the dashed block. The black lines indicate tree branches. Percentage bootstrap support is indicated by the numerical values.

The overall phylogenetic scheme of the negative RNA viral polymerase proteins in Figure 3.6a was split into nonsegmented (red) and segmented (blue) viral polymerase proteins. In addition the RNA viruses were divided into their respective families, namely viruses were grouped according to their *Rhabdoviridae* classification (*Rabies virus, Sonchus Yellow Net virus (SYNV), Viral Haemorrhagic Septicaemia (VHS), Vesicular Stomatitis Virus (VSV))* and *Orthomyxoviridae: (Influenza A , B, C, Thogoto virus*) classification.

The above tree strongly (99% -100%) related the *Borna disease virus* strain V and He to the family *Rhabdoviridae* and did not place the *Bornaviridae* in a separate classification, which is supported by the unrooted tree displayed in Figure 3.6b

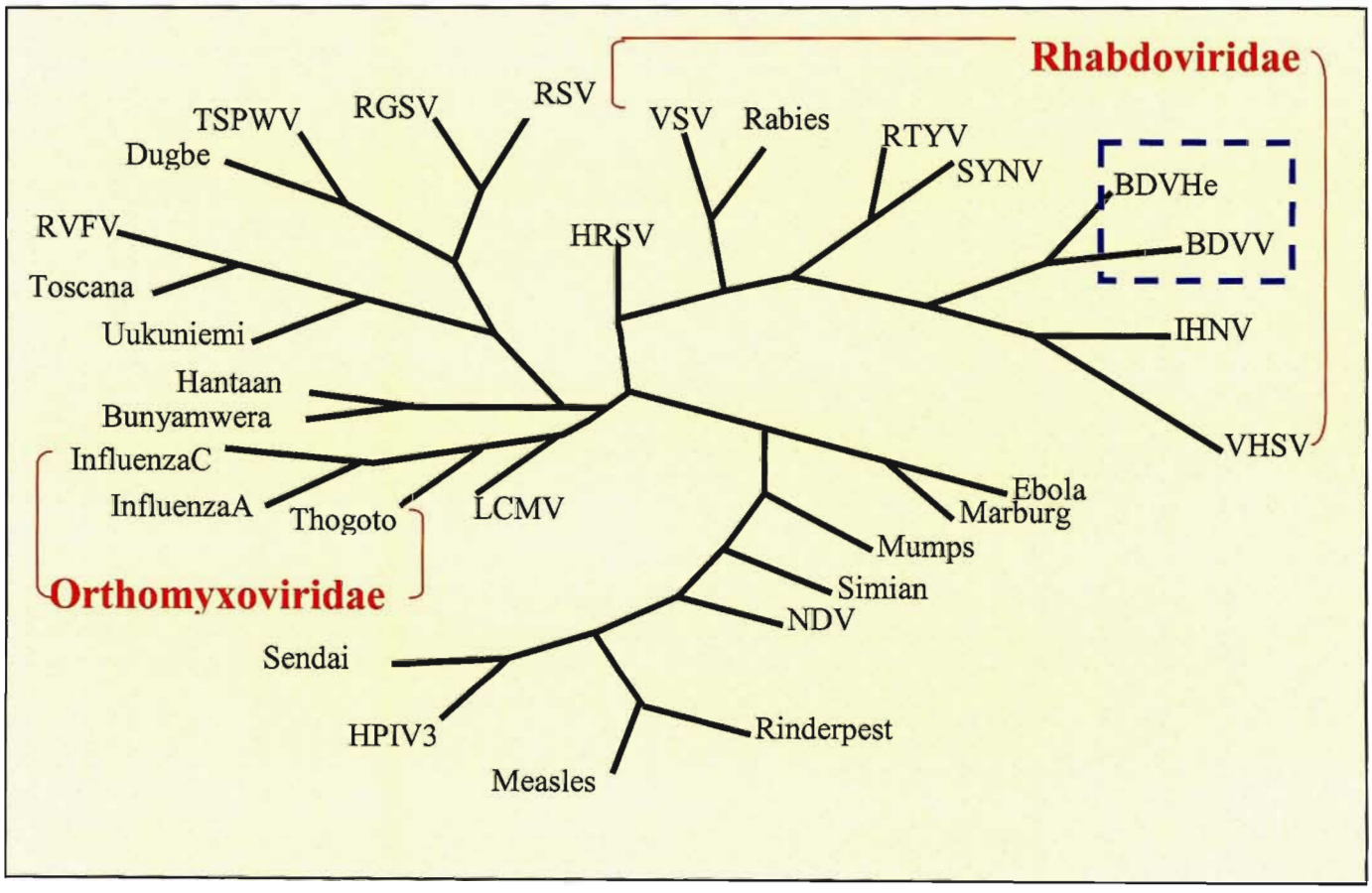


Figure 3.6b Phylogenetic analysis of negative stranded RNA polymerase protein sequences according to a parsimony analysis based on an unrooted tree. The position of BDV is indicated by the dashed block. The black lines indicate tree branches.

Figure 3.6b quite clearly classifies the *Borna disease virus* with the *Rhabdoviridae* family by placing viruses belonging to the *Rhabdoviridae* grouping and *Bornaviridae* grouping onto the same

tree branch or node. As with Figure 3.6a the RNA viruses are grouped into their respective *Orthomyxoviridae* and *Rhabdoviridae* families.

The results of the neighbour joining analysis (Figure 3.7) related the *Borna disease virus* (54 %) with the family *Rhabdoviridae*. Similar to the results displayed in Figures 3.6a and 3.6b the viruses belonging to the *Orthomyxoviridae* and *Rhabdoviridae* families were grouped together.

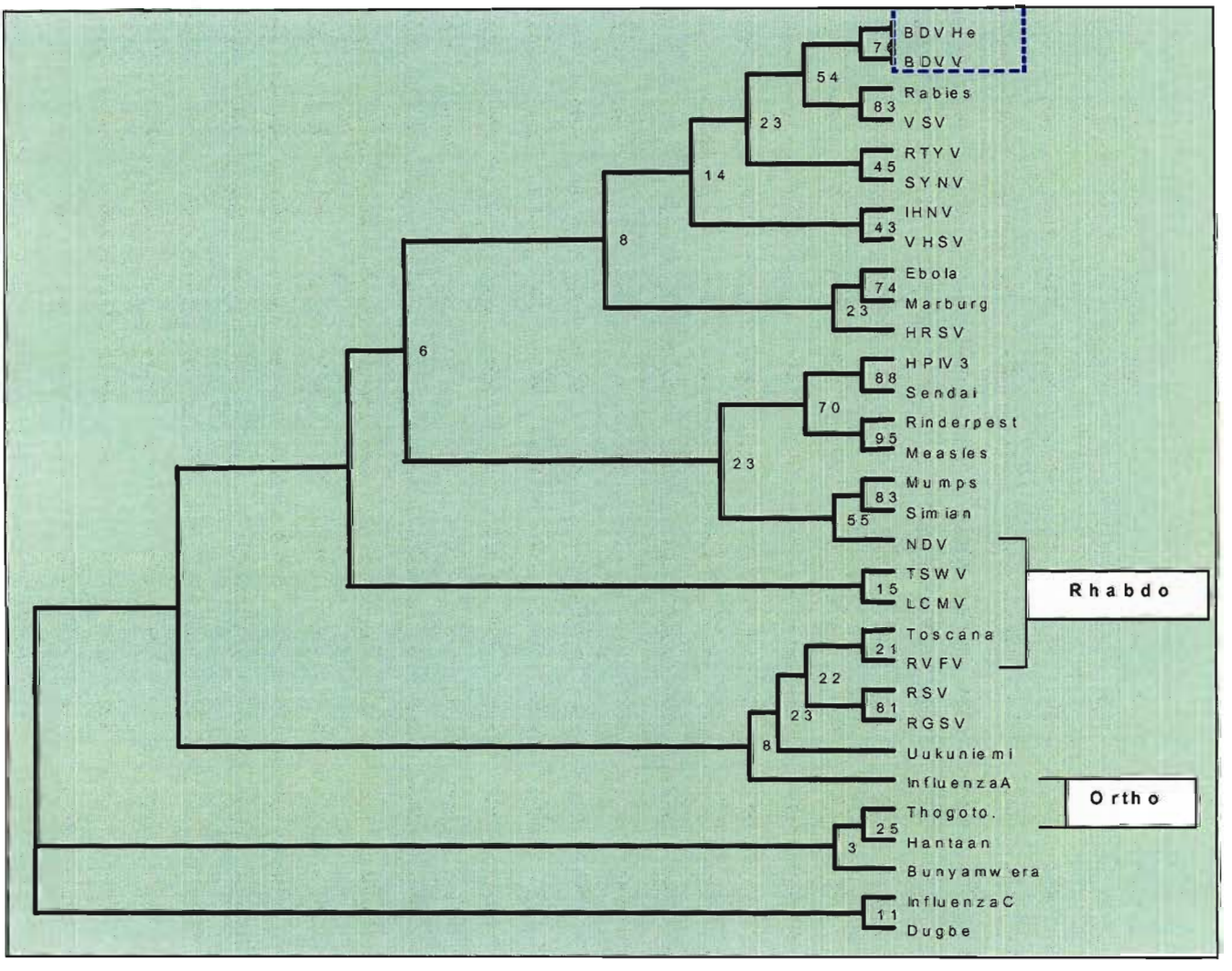


Figure 3.7 Phylogenetic analysis of negative stranded RNA polymerase protein sequences based on a rooted tree. Percentage bootstrap support is indicated by the values at each node. Rhabdo refers to *Rhabdoviridae* and Ortho refers to *Orthomyxoviridae*. The position of BDV is indicated by the dashed block. The black lines indicate tree branches. The scheme is based primarily on the neighbourhood joining analysis of the sequences aligned in Figure 3.1.

3.3 ESTIMATION OF THE EVOLUTIONARY MUTATION RATE OF THE *BORNA DISEASE VIRUS*

In this study the mutation rate of inter-species BDV sequences (Feline, Canine, Ovine, Human and Assine) and intra-species BDV strains (V, He, and RW98) was estimated using a method developed by Ina, (1995). This method was used to shed some light on the evolutionary nature of the *Borna disease virus* (BDV) and to examine whether BDV proteins are under positive or neutral selection. The following BDV proteins were investigated: p40 nucleoprotein, p57 glycoprotein, p24 phosphoprotein, gp18 matrix protein and p10 X protein.

3.3.1 Results of the comparison of BDV inter-species (host) and intra-species (virus) sequences

The results detailed in this section represent the results of a mutation analysis where pairwise comparison of BDV protein DNA sequences was performed, by using the BDV V strain as a reference strain. Pairwise in this case refers to BDV strain comparison as follows:

1. Inter-species host strain comparison with the BDV V strain for example (BDV V and Feline, BDV V and Canine, BDV V and Assine, BDV V and Ovine, BDV V and Human).

2. Intra-species virus strain comparison (BDV V and BDV He, BDV V and BDV RW98).

Observed codon nucleotide differences between two paired sequences were classified into synonymous and non-synonymous substitutions.

The results in this section have been divided into:

1. Synonymous substitution results (Tables 3.1a and 3.1b).
2. Non-synonymous substitution results (Tables 3.2a and 3.2b).
3. Transition/transversion ratio results (Tables 3.3a and 3.3b).
4. Synonymous and non-synonymous mutation rate results (Tables 3.4a and 3.4b).

1. Results of the synonymous substitutions analysis

The percentage of synonymous substitutions and the corresponding nucleotide substitution, transition or transversion, for the pairwise inter-species (host sequences) and intra-species (virus sequences) comparison is displayed in Table 3.1a and Table 3.1b respectively. Transition referred to the frequency of a nucleotide base substitution (a-g or c-t) whereas transversion referred to a nucleotide base substitution that changes the ring structure, for example a-c, a-t, g-c, g-t.

The inter-species pairwise comparison (Table 3.1a) examined the average frequency of synonymous substitutions of various BDV host species (Feline, Canine, Ovine, Human and Assine) and their respective p40 nucleoprotein (146 codons), p57 glycoprotein (503 codons), gp18 matrix protein (142 codons) and p24 phosphoprotein (201 codons). Only 146 codons of the p40 protein across all host species were available and no sequences for the p10 protein could be obtained. This is relevant as normally the p40 nucleoprotein consists of 371 codons. The comparison of synonymous substitutions from three BDV strains: V, He, and RW98 and their respective p40 nucleoprotein (371 codons), p57 glycoprotein (503 codons), gp18 matrix protein (142 codons), p24 phosphoprotein (201 codons) and p10 X protein (89 codons) are displayed in Table 3.1b.

The nucleotide substitutions were converted into a percentage by dividing the number of nucleotide substitutions for each protein by the number of codons specific for each protein and multiplying the amount by a 100. This was done in order to enable comparison between the host and virus sequence analysis.

The inter-species (host) analysis (Table 3.1a) indicated that most of the codons exhibited a transition substitution of a nucleotide base, for example a-g (152.52 in total), or c-t (221.56 in total) (displayed in blue) with very few codons exhibiting a transversion substitution for example a-c (8), a-t (15.5), g-c (0), or g-t (8) (displayed in red). The p57 glycoprotein had the most numerous synonymous mutations across all codons followed by the p40 nucleoprotein and the p24 phosphoprotein. The transversion substitutions were predominantly in the Leucine, Threonine, Isoleucine and Valine codons.

Table 3.1a Comparison of the total number of synonymous substitutions (%) occurring in protein codons of six species. *Borna disease virus* strain V proteins were compared with Feline, Canine, Ovine, Human and Assine protein sequences respectively and the percentage of nucleotide bases that changed for each codon was totalled. The p40 nucleoprotein, p57 glycoprotein, gp18 matrix protein and p24 phosphoprotein were examined. Transition (a-g, c-t) and transversion (a-c, a-t, g-c, or g-t) nucleotide substitutions are indicated in blue and red, respectively.

Codon Name	Codon symbol and nucleotide sequence mutation		Total % of nucleotide base changes per codon			
			p40[a]	p57	gp18	p24
Serine	S	tcc-tct	9.6	2.6	12.2	0
		tcg-tca	11.5	2.6	6.1	4.2
		agc-agt	0	6.5	0	0
Leucine	L	ctg-ttg	7.7	1.96	4.1	8.3
		ctc-ctt	3.8	3.9	6.1	2.1
		cta-ctg	5.8	2.6	8.2	8.3
		cta-tta	0	5.2	0	0
		cta-ttg	0	1.96	0	0
		ctg-ctt	0	0.7	0	0
		cta-ctt	0	1.3	0	4.2
Glycine	G	gga-ggg	9.6	4.6	8.2	4.2
		ggc-ggt	0	3.3	8.2	0
Threonine	T	acg-aca	5.8	2.6	0	0
		acc-act	0	1.3	6.1	4.2
		acc-aca	0	2.6	0	0
		acg-act	0	0.7	0	0
Isoleucine	I	ata-att	5.8	0	0	0
		atc-att	5.8	2.6	6.1	2.1
		atc-ata	0	3.3	0	2.1
Glutamate	E	gag-gaa	5.8	0	2.0	4.2
Proline	P	cca-ccg	5.8	0	0	0
		ccc-cct	0	4.5	2.0	0
Glutamine	Q	cag-caa	7.7	3.9	0	0
Cysteine	C	tgt-tgc	3.8	4.5	0	0
Valine	V	gtt-gtg	1.9	2.6	0	0
		gta-gtg	0	1.96	0	4.2
		gtc-gtt	5.8	3.3	0	0
		gta-gtc	0	0	0	4.2
Lysine	K	aag-aaa	3.8	1.3	4.1	4.2
Phenylalanine	F	ttc-ttt	0	3.9	8.2	0
Alanine	A	gcg-gca	0	8.5	0	0
		gct-gcc	0	2.6	8.2	0
		gct-gcg	0	0	0	2.1
Asparagine	N	aat-aac	0	3.3	6.1	0
Tyrosine	Y	tat-tac	0	4.6	0	0
Histidine	H	cat-cac	0	0.7	0	0
Aspartate	D	gat-gac	0	2.6	2.0	20.8
Arginine	R	agg-aga	0	2.6	2.0	4.2
		cgt-cgc	0	0	0	8.3
	Total codon no.		146	503	142	201

[a]Note that only 146 out of a possible 371 codons of the p40 protein for all six species were available from the genbank database.

The intra-virus species comparison (Table 3.1b) indicated that the p57 glycoprotein had a higher percentage of synonymous substitution (both transition and transversion) than the other proteins, followed by the p40 nucleoprotein and the p24 phosphoprotein. The gp18 matrix protein when compared with Table 3.1a displayed a lower synonymous substitution frequency.

Table 3.1b Comparison of the total number of synonymous substitutions (%) occurring in codons of three BDV strains (V, He, RW98) and their respective p40 nucleoprotein, p57 glycoprotein, gp18 matrix protein and p24 phosphoprotein. Transition (a-g, c-t) and transversion (a-c, a-t, g-c, or g-t) nucleotide substitutions are indicated in blue and red, respectively.

Codon Name	Codon symbol and nucleotide sequence mutation		Total % of nucleotide base changes per codon				
			p40	p57	gp18	p24	p10
Serine	S	tcc-tct	1.9	0.4	1.4	0	0
		tcc-tca	0	0	0	1.0	0
		tcg-tca	0.5	0.4	0	0	0
		agc-agt	0.5	2.6	0	0	2.2
		tcg-tct	0	0.5	0	0	0
Leucine	L	ctg-ttg	0.5	0.4	1.4	1.0	0
		ctc-ctt	0.5	1.2	0	1.5	0
		cta-ctg	0.5	0.8	1.4	2.0	0
		cta-tta	0.5	2.2	0	1.0	2.2
		cta-ttg	0	0	0	0	0
		ctg-ctt	0.5	0	0	0	0
		cta-ctt	0	0.4	0	1.5	0
		tta-ctg	0	0.4	0	0	0
		cta-ctc	0	0	0	0.5	0
Glycine	G	gga-ggg	1.9	1.2	0.7	1.0	0
		ggc-ggt	0	0	1.4	1.0	0
		ggc-gga	0	0.2	0	0	0
		ggc-ggg	0	0.2	0	0	1.1
Threonine	T	acg-aca	1.4	2.0	0	0	0
		acc-act	1.1	1.4	1.4	1.0	0
		acc-aca	0	0.8	0	0	0
		acg-act	0	0	0	0	0
Isoleucine	I	ata-att	0.3	0	0	0	0
		atc-att	1.6	0.8	1.4	2.0	0
		atc-ata	0.5	0.4	0	0	0
Glutamate	E	gag-gaa	2.2	1.1	1.4	1.0	0
Proline	P	cca-ccg	0	0	1.4	0	0
		ccc-cct	1.9	1.0	0	0	0
		ccg-cct	0	0.4	0	0	0
Glutamine	Q	cag-caa	0.5	1.6	0	1.0	0
Cysteine	C	tgt-tgc	0.5	0.4	0	1.0	0
	V	gtt-gtg	0	0		1.0	0

Valine		gta-gtg	0	0.4	1.4	0	0
		gtc-gtt	1.1	0.4	0	1.0	0
		gta-gtc	0	0	0	0	0
		gta-gtt	0	0.4	0	0	0
	K	aag-aaa	2.4	0.4	0	0	0
Lysine	F	ttc-ttt	1.6	2.0	3.5	0	0
Phenylalanine	A	gcg-gca	1.9	2.4	4.2	1.0	0
Alanine		gct-gcc	1.1	0.8	0	0	0
		gct-gcg	0	0	1.4	0	0
	N	aat-aac	0.5	0	0	0	0
Asparagine	Y	tat-tac	0	0.6	1.4	0	0
Tyrosine	H	cat-cac	0.5	0.4	0	0	0
Histidine	D	gat-gac	2.2	2.2	1.4	2.3	0
Aspartate	R	agg-aga	0.3	0.2	0	1.0	0
Arginine		cgt-cgc	0.5	0	1.4	1.0	0
		agg-cgg	0.3	0	0	0	0
		cga-cgg	0.5	0	0	0	0
		agg-cgg	0.3	0	0	0	0
		cga-aga	0	0	0	0	2.2
					0		
	Total no of codons		371	503	142	201	89

2. Results of the non-synonymous substitutions analysis

The frequency of non-synonymous (amino acid change) substitutions and the corresponding codon change from the pairwise inter-species (Feline, Canine, Ovine, Human and Assine) protein comparison with the BDV V strain and the intra-species (BDV V, RW98, BDV He) virus protein sequence comparison is displayed in table 3.2a and table 3.2b respectively.

The intra-species (virus) analysis (Table 3.2a) indicated that in contrast to Table 3.1a more codons exhibited a transversion substitution (a-c, a-t, g-c, or g-t) with a higher occurrence of first and second position substitutions. Examples of the first position substitution were Thr → Ser (**acc-tcc**) and of the second position substitution Asp → Ala (**gac-gcc**). The p57 glycoprotein had the highest number of codon substitutions (43) (Table 3.2a), followed by the p24 phosphoprotein which had 14 substitutions. In addition several codons had mutated either from a hydrophilic to hydrophobic codon or from a hydrophobic codon to a hydrophilic codon. Examples of the former are Cys → Ser (1 codon) and of the latter are Thr → Ala (9 codons), His → Ala (4 codons), Ser → Ala (1 codon), Asp → Ala (1 codon).

Table 3.2a Comparison of the total number of non-synonymous substitutions occurring in protein codons of six species. *Borna disease virus* strain V proteins were compared with Feline, Canine, Ovine, Human and Assine protein sequences respectively. The p40 nucleoprotein, p57 glycoprotein, gp18 matrix protein and p24 phosphoprotein were compared, Transition (a-g, c-t) and transversion (a-c, a-t, g-c, or g-t) nucleotide substitutions are indicated in blue and red, respectively.

Codon Mutation	Nucleotide sequences mutation	Total % of nucleotide base changes per codon			
		p40	p57	gp18	p24
T-A	acg-gcg	0	3	0	1
	aca-gcc	0	1	0	0
	acg-gct	0	1	0	1
	acc-gcc	0	4	0	0
S-G	agc-ggc	0	2	0	0
S-L	tcg-ctg	0	0	0	3
	tcc-ctg	0	0	0	1
P-L	cct-ctt	0	4	0	0
	cca-cta	0	0	0	1
L-E	cga-gaa	0	0	0	1
I-V	att-gtc	0	3	0	0
	atc-gtc	1	0	0	4
C-S	tgc-agc	0	1	0	0
T-S	acc-tcc	0	2	0	0
E-Q	gag-cag	0	1	0	0
E-D	gaa-gca	0	0	1	0
E-V	gag-gtg	0	0	0	1
H-A	cac-cat	0	4	0	0
R-K	agg-aaa	0	1	0	0
	agg-aag	0	1	0	0
S-A	tca-gca	0	1	0	0
C-Y	tgc-tac	0	1	0	0
K-E	aag-gag	0	2	0	0
S-N	agt-aat	0	1	0	0
D-A	gat-gct	0	1	0	0
	gac-gcc	0	0	0	1
D-V	gat-gtc	0	0	0	1
Y-H	tac-cac	0	1	0	1
A-V	gca-gta	0	4	0	2
	gcc-gtc	0	4	0	0
P-D	ccc-gac	0	0	0	1
P-S	tcg-ctg	0	0	0	1

The intra-species analysis (Table 3.2b) indicated that the p57 glycoprotein had the highest number of codon substitutions (32), followed by the p10 X protein (10) and the p24 phosphoprotein (6). In contrast to the inter-species analysis (Table 3.1a) there were considerably fewer non-synonymous mutations among the individual proteins with a lower occurrence of first and second position substitutions.

Table 3.2b Comparison of the total number of non-synonymous substitutions of three BDV virus strains (V, He, and RW98) and their respective p40 nucleoprotein, p57 glycoprotein, gp18 matrix protein and p24 phosphoprotein. Transition (a-g, c-t) and transversion (a-c, a-t, g-c, or g-t) nucleotide substitutions are indicated in blue and red, respectively.

Codon mutation	Nucleotide sequence mutation	Total % of nucleotide base changes per codon				
		p40	p57	gp18	p24	p10
T-A	acg-gcg	0	1	0	2	0
	agt-ggt	0	0	0	0	2
S-F	tct-ttt	0	2	0	0	0
P-L	cct-ctt	0	5	0	0	0
I-V	atc-gtc	0	2	0	2	2
	att-gtt	0	2	0	0	0
I-T	att-ata	1	0	0	0	0
	att-act	0	0	0	0	2
	atc-acc	0	0	0	0	1
E-D	gaa-gat	2	0	0	0	0
	gaa-gac	0	0	2	0	0
	agg-aag	0	1	0	0	0
	aga-aaa	2	4	0	0	0
	aga-agg	0	1	0	0	0
R-Q	cag-cgg	0	2	0	2	0
R-G	cga-gga	0	0	0	0	0
S-G	agc-ggc	0	1	0	0	0
D-G	gat-ggt	0	0	0	0	2
A-V	gca-gta	0	2	0	0	0
	gcc-gtc	0	2	0	0	0
A-G	gcg-ggg	0	0	0	0	1
P-S	tcg-ctg	0	2	0	0	0
V-G	gtt-ggt	0	2	0	0	0
V-M	gtg-atg	0	2	0	0	0
Q-G	ggt-ggc	0	1	0	0	0

3. Transition/Transversion (TS) ratio results

The Transition/Transversion ratio results (Table 3.3a and Table 3.3b) represent a comparative pairwise protein analysis of the BDV V strain with inter-species (Feline, Canine, Sheep, Assine and Human) sequences (Table 3.3a) and intra-species sequences (BDV V, He, and RW98) (Table 3.3b) respectively. The data were obtained from Tables 3.1a, 3.1b, 3.2a and 3.2b respectively. The transition/transversion ratio (TS) was calculated by dividing the number of transitions by the number of transversions. Transition referred to the frequency of a nucleotide base substitution (a-g

or c-t) whereas transversion referred to a nucleotide base substitution that changed the ring structure, for example a-c, a-t, g-c, g-t.

The p57 glycoprotein in general displayed the highest TS ratio for all five host species (Table 3.3a), followed by the p24 phosphoprotein and the p40 nucleoprotein. The human strain exhibited the lowest TS ratio and the gp18 matrix protein displayed the lowest overall TS ratio across all species.

Table 3.3a The Transition/Transversion ratio (TS) as a result of pairwise protein comparison between the BDV V strain p57 glycoprotein, gp18 matrix protein, p24 phosphoprotein and p40 nucleoprotein and the corresponding Feline, Canine, Sheep, Assine and Human protein sequences.

	TS ratio (%) of BDV proteins			
Host Species	**p57**	**gp18**	**p24**	**p40**
Feline	7.33	0	3	0
Canine	4.40	0	4.1	12
Ovine	7.25	0	3.3	13
Assine	8.8	0	9.0	8.5
Human	0	1	8	0

The intra-species comparison (Table 3.3b) displayed a high TS ratio for the BDV HE and BDV RW98 strain p40 protein (18.5%), with the gp18 matrix protein and p10 X protein in general exhibiting the lowest TS ratio for all virus strains.

Table 3.3b The Transition/Transversion ratio (TS) as a result of pairwise BDV p57 glycoprotein, gp18 matrix protein, p24 phosphoprotein, p40 nucleoprotein and p10 virus strain protein comparison.

Strain	TS ratio (%) of BDV proteins				
	p57	p40	p24	gp18	p10
BDV HE-BDV RW98	7.25	18.5	5.67	5.67	2
BDV V - BDV He	11.4	7	7.5	7.5	6
BDV V- BDV RW98	10.8	9.66	5.67	5.67	4

4. Synonymous and non-synonymous mutation rate results

The synonymous and non-synonymous mutation rate results (Table 3.4a and Table 3.4b) reflected the number of synonymous (silent mutations) and non-synonymous (amino acid changes) substitutions obtained through a comparative pairwise analysis of the BDV V strain proteins with inter-species (Feline, Canine, Sheep, Assine and Human) protein sequences and intra-species sequences (BDV V, BDV He, BDV RW98) respectively. The BDV p57 glycoprotein, gp18 matrix protein, p24 phosphoprotein, p40 nucleoprotein and p10 X protein were analysed. The synonymous and non-synonymous substitution mutation rate was calculated according to a formula developed by Ina, (1995), which is detailed in section 2.3.2.

In general the non-synonymous and synonymous mutation rate was very low for all proteins (Table 3.4a). However, the p57 glycoprotein had the highest overall mutation rate followed by the gp18 matrix protein. The human strain had the lowest mutation rate for all proteins.

Table 3.4a Number of synonymous (ds) and non-synonymous (dn) substitutions per site obtained through BDV V strain protein pairwise inter-species (Feline, Canine, Sheep, Assine and Human) comparison, calculated according to Ina's, (1995) method. The p40 nucleoprotein, p57 glycoprotein, gp18 matrix protein and p24 phosphoprotein were compared

Species		p57	gp18	p24	p40
Feline	ds	0.04[a] (8.7×10^{-5})	0.05 (0.0004)	0.04 (0.0002)	0.05 (0.0003)
	dn	0.006 (5.9×10^{-6})	0	0.01 (1.4×10^{-5})	0
Canine	ds	0.09 (2.2×10^{-4})	0.1 (0.0009)	0.01 (0.0006)	0.092 (0.0007)
	dn	0.012 (1.2×10^{-5})	0	0.01 (3.1×10^{-5})	0
Ovine	ds	0.113 (2.8×10^{-4})	0.11 (0.001)	0.045 (0.0003)	0.1 (0.0008)
	dn	0.016 (1.6×10^{-5})	0	0.01 (2.7×10^{-5})	0
Assine	ds	0.115 (2.9×10^{-4})	0.10 (0.0009)	0.085 (0.0005)	0.13 (0.0011)
	dn	0.008 (7.8×10^{-6})	0	0.11 (1.6×10^{-5})	0.003 (1.1×10^{-5})
Human	ds	0.004 (6.9×10^{-8})	0.007 (5.0×10^{-5})	0.118 (9.5×10^{-5})	0
	dn	0.002 (1.9×10^{-6})	0.0009 (7.8×10^{-7})	0.016 (4.0×10^{-5})	0.003 (1.2×10^{-5})

[a]Numbers in brackets refer to standard deviations calculated according to Ina's (1995) method for estimating the number of synonymous and non-synonymous mutations.

The non-synonymous and synonymous mutation rates (Table 3.4b) were in general higher than the mutation rates displayed in Table 3.4a. The p57 glycoprotein displayed the highest ds and dn mutation rates followed by the p40 nucleoprotein.

Table 3.4b Number of synonymous (ds) and non-synonymous (ds) substitutions per site obtained through a BDV strain (V, He, RW98) pairwise comparison (Ina, 1995). The p40 nucleoprotein, p57 glycoprotein, gp18 matrix protein, p10 X protein and p24 phosphoprotein were compared

Protein		Strain		
		BDV He-BDV RW98	**BDV V-He**	**BDV V-BDV RW98**
p57	ds	0.12 [a] (0.00030)	0.114 (0.00003)	0.1106 (0.00028)
	dn	0.012 (1.29×10^{-5})	0.0099 (9.88×10^{-6})	0.0099 (9.37×10^{-5})
gp18	ds	0.0092 (6.5×10^{-5})	0.1177 (0.0011)	0.083 (6.93×10^{-4})
	dn	0	0.0035 (1.26×10^{-5})	0.0035 (1.26×10^{-5})
p24	ds	0.1038 (0.00047)	0.071 (0.00041)	0.104 (6.19×10^{-4})
	dn	0.00494 (1.22×10^{-5})	0.0099 (2.47×10^{-5})	0.0049 (1.23×10^{-5})
p40	ds	0.1078 (3.4×10^{-4})	0.1325 (4.32×10^{-4})	0.0862 (2.59×10^{-4})
	dn	0.0014 (1.9×10^{-6})	0.0028 (3.84×10^{-6})	0.0014 (1.91×10^{-6})
p10	ds	0.0304 (3.16×10^{-4})	0.0307 (3.10×10^{-4})	0.0203 (2.07×10^{-4})
	dn	0.0183 (1.14×10^{-4})	0.0245 (1.50×10^{-4})	0.0183 (1.14×10^{-4})

[a]Numbers in brackets refer to standard deviations calculated according to Ina's (1995) method for estimating the number of synonymous and non-synonymous mutations.

3.3.2 Results of the Analysis of the Mode of evolution of BDV Host proteins (Inter-Species) and BDV Strain (Intra-Species) Proteins

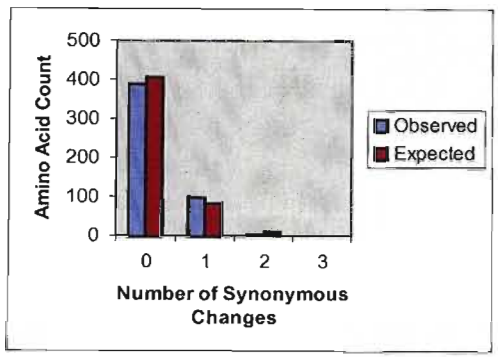
The results detailed in this section represent the results of an evolutionary analysis. A simple Poisson model of synonymous substitution was applied to the BDV V strain glycoprotein, phosphoprotein, matrix and nucleoprotein and various host strains (Feline, Canine, Sheep, Assine and Human), (Figure 3.8). The same model was applied to three BDV strains (BDV V, He and RW98) and their proteins, (Figure 3.9). The raw data was converted from Table 3.1a and Table 3.1b by converting the percentages displayed in Tables 3.3a and 3.3b to the original values. The average synonymous substitutions for 1st (for example **ctg → ttg**) and 3rd (**atg → ata**) codon positions combined are indicated in Figure 3.8 and Figure 3.9 respectively. The site distributions were compared using a Chi-square test and if the expected synonymous distribution was found to differ from the observed distribution it was concluded that the actual process of synonymous substitution was not homogenous and independent over all sites and therefore not neutral in nature.

The inter-species evolutionary analysis (Figure 3.8) indicated that all the proteins displayed a probability higher than 50% that the observed results differed from the expected, for example the pattern of distribution of synonymous substitution was not neutral. The glycoprotein exhibited the highest probability with 97%.

In the intra-species evolutionary analysis (Figure 3.9) the glycoprotein, nucleoprotein and X protein displayed a probability higher than 50% that the observed results differed from the expected, for example the pattern of distribution of synonymous substitution was not neutral, whereas the phosphoprotein and matrix protein displayed a probability less than 50% that the observed results differed from the expected, for example the distribution of synonymous substitution followed a neutral pattern.

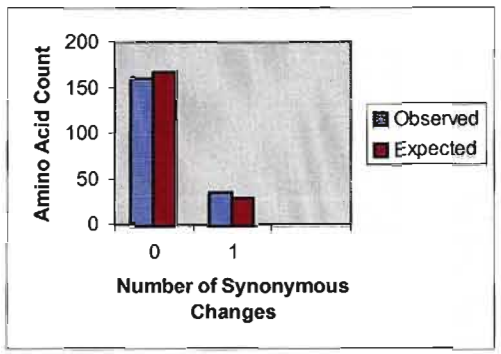
Figure 3.8 Graphical illustration of the Poisson model of synonymous substitution calculated from the BDV glycoprotein, phosphoprotein, matrix protein and nucleoprotein. The *Borna disease virus* strain V proteins were compared with various host (Feline, Canine, Sheep, Assine and Human) strain proteins. The y-axis is the number of codons that have incurred the number of synonymous changes indicated on the x-axis, at 1st and 3rd position sites. Results of chi-squared goodness of fit tests are indicated.

Glycoprotein p57 (Sites 1 and 3)



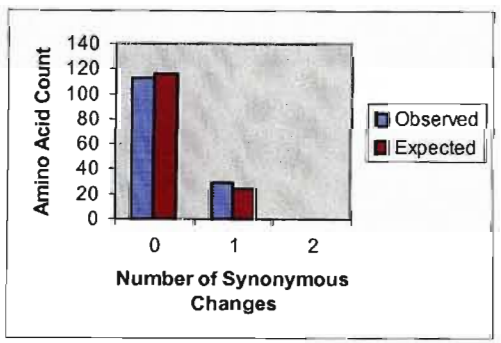
χ^2=8.97
P= 97.01%

Phosphoprotein p24 (Sites 1 and 3)



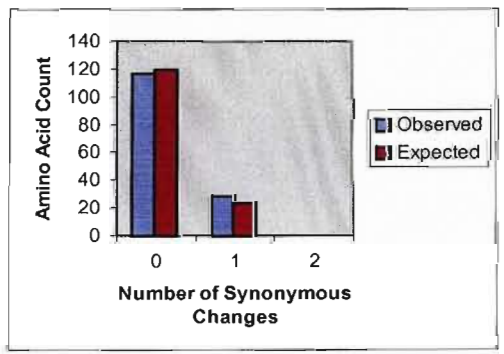
χ^2= 1.477
P= 77.52 %

Matrix protein gp18 (Sites 1 and 3)



χ^2= 1.31
P= 74.8 %

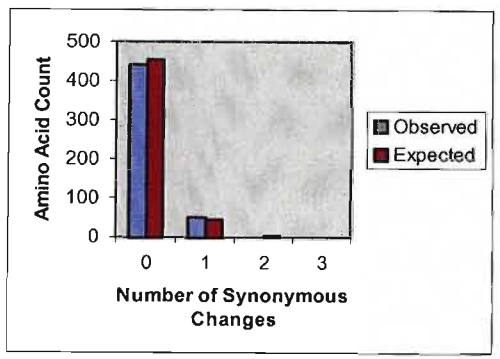
Nucleoprotein p40 (Sites 1 and 3)



χ^2= 1.186
P= 72.38%

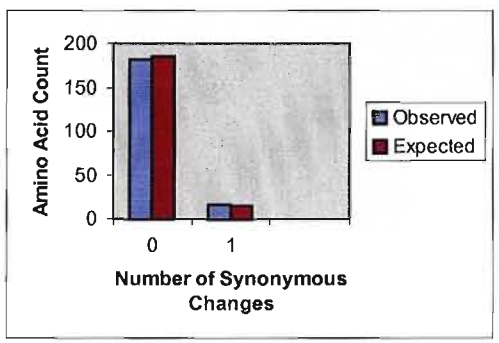
Figure 3.9 Graphical illustration of the Poisson model of synonymous substitution calculated from the glycoprotein, phosphoprotein, matrix protein, nucleoprotein and X protein of three *Borna disease virus* (V, He, RW98) strains. The y-axis is the number of codons that have incurred the number of synonymous changes indicated on the x-axis, at 1st and 3rd position sites. Results of chi-squared goodness of fit tests are indicated.

Glycoprotein p57 (Sites 1 and 3)



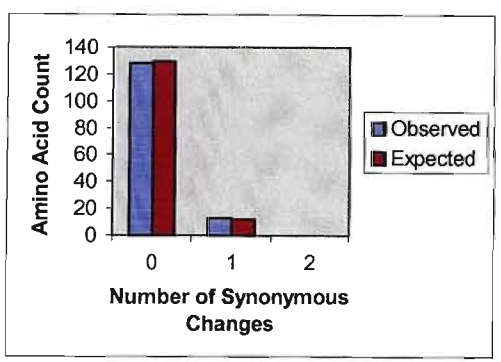
χ^2=1.99
P= 63.065%

Phosphoprotein p24 (Sites 1 and 3)



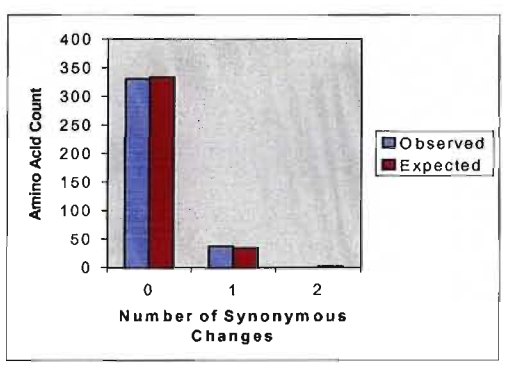
χ^2= 0.172
P= 32.18 %

Matrix protein gp18 (Sites 1 and 3)



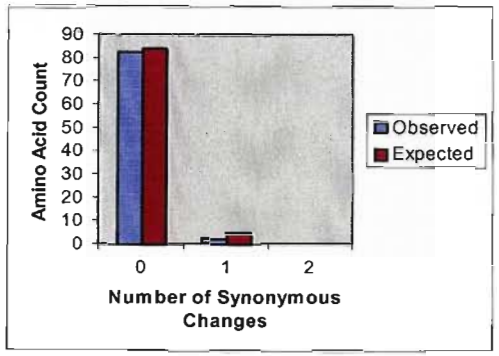
χ^2= 0.116
P= 26.42%

Nucleoprotein p40 (Sites 1 and 3)



χ^2= 1.536
P= 53.61%

X-protein p10 (Sites 1 and 3)



$\chi^2 = 1.02$
P= 68.76 %

CHAPTER 4

DISCUSSION AND CONCLUSION

4.1 INTRODUCTION

RNA viruses are a diverse group that infect prokaryotes as well as many eukaryotes, both plants and animals. Viral fitness, or ability to replicate infectious progeny, can vary a million-fold within short time intervals. Paradoxically, functional and structural studies suggest extreme limitations to virus variation. A hallmark of RNA genomes is the error-prone nature of their replication and retro transcription. The major biochemical basis of the limited replication fidelity is the absence of proofreading/repair and post replicative error correction mechanisms that normally operate during replication of cellular DNA. In spite of this unique feature of RNA replicons, the dynamics of viral populations seem to follow the same basic principles that classical population genetics has established for higher organisms.

The advents of various RNA virus sequencing projects have brought about the possibility of a more accurate classification of these viruses into various orders and families. The negative stranded virus field (order Mononegavirales) is one area in the RNA virus family that still required investigation and a more accurate classification. Therefore, the purpose of this study was to adequately classify the newest member of the Mononegavirales, the *Bornaviridae*, and to evaluate the evolutionary trend in the form of mutation rates of various *Borna disease virus* proteins. The *Bornaviridae* family was chosen as (1) there are a number of areas involving the BDV proteins that is puzzling to investigators in this field, (2) an adequate classification of the *Bornaviridae* would shed some light on the evolutionary relationship of the *Bornaviridae* to other RNA negative stranded viruses, (3) this is the first time that the evolutionary trend of the *Borna disease virus* has been examined.

4.2 BORNAVIRIDAE CLASSIFICATION

4.2.1 Polymerase motifs of the *Mononegavirales*

The *Clustal X* alignment analysis of the polymerase protein of 27 RNA negative stranded viral species indicated seven distinct blocks of amino acid residues. These blocks can be considered as extensively conserved motifs. In the case of the *Borna disease virus* (BDV) polymerase, these motifs were previously named as motifs i-IV (originally designated A, B, C and D) (Briese *et al.*, 1994) and v and vi, identified in this investigation. Several amino acid residues of these motifs were strictly conserved, seven amino acids in the case of motif i, two for motif I, four for II, three for III (which is part of the well known GDN motif), two for IV, four for v and two for vi.

The presence of motif v and vi was confirmed by two different alignment programs *DIALIGN* and *MEME*. The *DIALIGN* alignment indicated that several more residues were conserved in motif v and vi, than in the alignment generated by *Clustal X* which may be due to the fact that *Clustal X* utilized more discriminatory multiple alignment parameters. Furthermore, *DIALIGN* has been shown to return high scoring fragments that are highly correlated to exons in genomic sequences and not to coincide with protein-coding regions that have structural relevance, (Morgenstern *et al.*, 2002). Although the presence of motif v and vi was confirmed with *DIALIGN* it was not possible to distinguish highly conserved structural residues from non-functional conserved residues, as the evolutionary distance between the compared species may have been too close (Morgenstern *et al.*, 2002). The *MEME* alignment, based on a hidden markov model, confirmed the presence of motif v and vi in the BDV V polymerase protein with a high probability.

The functions of the motifs have not been clearly defined. However, the presence of a strongly conserved lysine residue in motif I confirms the speculation that this region could be a template binding site (Briese *et al.*, 1994) as this amino acid is important for DNA phosphate interaction (Heringa and Argos, 1994). The strongly conserved GDN residues in motif III suggest the possibility of a common ancestral fold in the non-segmented RNA viruses, as the aspartate and asparagine residues are involved in catalysis or metal binding and can therefore be considered as the catalytic sub-domain (Heringa and Argos, 1994). The functions of the other motifs are rather speculative. However, the presence of two very strongly conserved lysine residues in motif IV is indicative of a strongly phosphorylative action, which may be involved in phosphorylation of the polymerase protein co-factor p24 phosphoprotein which has been shown to bind to the polymerase protein. (Schwemmle *et al.*, 1998).

Motifs v and vi, identified in this investigation, revealed pronounced conservation, more specifically in motif v, which contradicts the results of Briese *et al.*, (1994). This discrepancy may be due to the fact that (1) the present analysis, unlike the previous analysis, dealt with the full complement of viral polymerases, possibly allowing for more precise identification of the conserved motifs and, (2) more discriminating software programmes were used to generate the alignment. The function of these domains remains speculative. However, alignment results in this investigation have demonstrated that the p40 nucleoprotein nuclear localization signal aligns to a region between motifs IV and vi of the polymerase protein of BDV and Rabies virus, which suggests that p40 may be involved with the co-transport to the nucleus with another protein and therefore acts as an activator for infection and replication. Previous research by McClure *et al.*, (1992) demonstrated that the *Rhabodvirus* and *Paramyxovirus* p40 nucleoprotein had two homologous domains. It is speculated that the BDV p40 nucleoprotein similarly contains two domains, one of which contains the nuclear localization signal acting as a mechanism through which the nucleoprotein moves in the nucleus (Kobayashi *et al.*, 1997; Pyper and Gartner, 1997). The interaction of the BDV p40 nucleoprotein localization signal with the BDV polymerase protein has been confirmed in this investigation.

4.2.2 Taxonomic status of the *Borna disease virus*

A phylogenetic analysis of 27 RNA polymerase virus sequences undertaken in this investigation to obtain a better understanding of the status of the *Bornaviridae* family clearly supports a different BDV classification. The overall phylogenetic scheme conformed to the splitting of all negative stranded RNA viral polymerases into non-segmented and segmented viral polymerases, as proposed by Goldbach and Haan, (1994). A bootstrap analysis related the *Borna disease* virus strain V and He to the family Rhabdoviridae, which was also supported by a neighbour joining analysis. These results further indicated that the *Rabies* virus and the *Vesicular stomatitis* virus are the most closely related animal viruses to BDV, whereas the *Rice transitory yellowing* (RTY) (unclassified *Rhabdovirus*) and *Sonchus yellow net* (SYNV) plant virus are closer to BDV than other animal *Rhabdoviridae*, such as *Viral haemorrhagic septicaemia (VHS)*, raising intriguing questions on the evolutionary origins of the *Borna disease virus*.

The phylogenetic analysis by means of the neighbour joining and the parsimony tree building methods of this investigation has for the most part conformed to the family and subfamily classification of the order Mononegavirales. However, some intriguing deviations have been detected. One of which is that the *Borna* virus does not fall into a separate *Bornaviridae* family but may be more appropriately placed into a separate subfamily in the family *Rhabdoviridae*. The fact

that both the parsimony and the neighbour joining bootstrap analyses, which are reliant on two different algorithms, produced similar results although with a different percentage, indicated the reliability and validity of this result. The slight discrepancy observed in the neighbour joining tree and the parsimony tree building method can be attributed to two very different algorithms. The parsimony method has been described as the more accurate tree building method (Felsenstein, 1993) as it builds the tree at a slower rate allowing the option of global rearrangement, which provides a better method of finding the most accurate tree possible, whereas the neighbour joining method finds the nearest pair of neighbouring sequences in the shortest amount of time, allowing for a "quick" method to confirm results.

Other points relevant to the classification of the *Bornaviridae* as part of the *Rhabdoviridae* family may be that (1) the p40 nucleoprotein of the *Borna disease* virus is homologous to domains of both the *Rabies* virus and the BDV polymerase proteins; (2) a similar mode of dissemination for the *Rabies* virus and BDV along neuronal chains, utilizing natural connections of neurons compatible with a synaptic transfer of these viruses (Gosztonyi *et al.*, 1993); (3) the ability of the *Rabies* virus nucleoprotein to bind to BDV polymerase protein and (4) the multifarious species infection capability of both the *Rabies* virus and the *Borna disease* virus protein.

4.3 EVOLUTIONARY PATTERN OF THE *BORNA DISEASE VIRUS*

Comparison of synonymous (silent) and non-synonymous (amino acid change) substitution rates provides an important method for studying the mechanisms and patterns of DNA sequence evolution (Kimura, 1983; Gillespie, 1991; Ohta, 1993) and enables the establishment of evolutionary models, which indicate whether proteins are under positive or neutral selection. Positive or directed selection on protein variants could arise during the course of an infection as a result of specific immune responses and be transmitted to immunologically naive individuals, as a mechanism for resistance to super-infection

In this first attempt to clarify the evolutionary pattern of the *Borna disease virus* a model of evolutionary analysis was created to identify the evolutionary trend of the BDV proteins that were examined. This in turn served to identify the proteins that are subject to positive and neutral selection in both host (Feline, Canine, Ovine, Assine and Human) and virus (BDV V, He and RW98) protein sequences, which may help clarify and enhance the lack of current knowledge relating to species infection and the epidemiological nature of the virus and whether the virus is subject to super infection interference.

4.3.1 Evolutionary trend of BDV proteins

1. Synonymous nucleotide substitution trend

In this investigation an inter-species (Feline, Canine, Ovine, Assine and Human) and intra-species (BDV V, He and RW98) analysis revealed that the p57 glycoprotein for both analyses displayed a high frequency of synonymous nucleotide substitutions followed by the p24 phosphoprotein and the p40 nucleoprotein. It was expected that the p57 glycoprotein due to its polycistronic transcription (Schneemann *et al.*, 1995; Schwemmle *et al.*, 1998) should display the most diversity of codon substitutions. The nucleoprotein and phosphoprotein, which exhibited monocistronic transcription (Schneemann *et al.*, 1995; Schwemmle *et al.*, 1998), had a high number of synonymous substitutions, which makes the higher percentage of synonymous substitutions exhibited by certain codons in these two proteins more significant.

The varied results may be explained as follows: The host BDV sequences were all obtained from animals that displayed the characteristics of a fatal and lethal BDV infection, whereas the BDV virus strains, with the exception of the BDV V strain have not been shown to be lethal or particularly virulent, allowing more insignificant nucleotide substitutions to take place. Further research and sequence data of both host and virus strains are needed.

2. Non-synonymous nucleotide substitution trend

The non-synonymous nucleotide base substitutions reflect that the p57 glycoprotein in both the inter-species host strain comparison and BDV intra-species virus strain comparison exhibits most of these substitutions, 21 and 15 respectively, indicating a non-random pattern of codon substitution. It can be assumed that the BDV glycoprotein nucleotide substitution observed in this study is not independent, and therefore illustrates an example of directed or positive evolution among different species that may have disparate cell metabolic and physiochemical environments.

Of the 21 codon substitutions several codons have mutated either from a hydrophilic to hydrophobic codon or from a hydrophobic codon to a hydrophilic codon. It is noticeable that a large number of codons were transformed into alanine. Replacements involving alanine, which is the Beta carbon contained in all of the other amino acids in isolation, may be more frequent in the short term rather than the long term. Therefore alanine may act as an intermediate replacement in a stepwise transition between side chains resulting in amino acid mutations (Kyte, 1995). The only other protein that exhibited a high number of non-synonymous mutations is the p24 protein with 14 substitutions in the host sequence analysis and four substitutions in the virus sequence analysis.

The p10 X protein as evidenced in the virus strain protein analysis exhibited a high number of substitutions as well. Unfortunately at the time of analysis no sequences for a p10 protein inter-species comparison were available.

3. Synonymous and non-synonymous mutation rate

The synonymous (ds) and non-synonymous (dn) mutation rate for both the host protein sequence and the virus protein sequence analysis was quite low, which may be a feature caused by overlapping transcription domains (Haydon *et al.*, 1998). The gp18 matrix protein in the host sequence analysis displayed one of the highest synonymous mutation rates and the lowest non-synonymous mutation rate. This is in contrast to the virus sequence analysis, which displayed a lower synonymous mutation rate and a higher non-synonymous mutation rate for the matrix protein.

These results may be interpreted as follows: The gp18 matrix protein exhibits a similar polycistronic transcriptional activity to the p57 glycoprotein and is presumed to have an antigenic function which is ubiquitous across species. Significant antigenic diversification has been shown to arise even as a result of single synonymous point mutations (Mateu *et al.*, 1990, Martinez *et al.*, 1991) and thus in light of the high mutation rate characteristic of the RNA viruses it is likely that a generation of substantial antigenic variability may arise. The adaptive significance of this may be related to a super-infection hypothesis.

Antigenic variation may be of adaptive value for two reasons: Firstly, antigenic variation generated over a course of synonymous or non-synonymous substitution may prolong or intensify a single infection, thereby resulting in greater transmission potential from infected animals. Secondly, sufficiently distinct strains might be capable of more rapid reinfection of hosts with some previous experience of related antigen, thereby effectively increasing the susceptible host population size. Due to the immunogenic nature of virus capsids it is inevitable that point (synonymous) mutations to genes encoding the capsid will result in antigenic variation; this has been repeatedly demonstrated (Klupp *et al.*, 1998). Such a variation seems to be subject to positive selection in BDV, which is supported by the results displayed in comparison of three BDV virus strains. The gp18 matrix protein does not exhibit a directed evolution pattern; rather it conforms to a neutral theory of evolution. This may be a result of lower species infection variability, for example cell culture created strains, which may curtail antigenic variability as explained above.

4. Evolutionary model

To confirm which proteins exhibited a positive or neutral evolutionary trend a model of evolutionary analysis was created. Various protein sequences were investigated by examining whether the observed 1st and 3rd codon position sites of individual BDV host protein and virus strain protein sequences received identifiable synonymous substitutions. This information was used to determine whether the observed frequency distribution was consistent with that expected from a Poisson (neutral evolution) distribution or a positive selection of proteins.

The results of the host protein sequence analysis of BDV proteins, found that all the BDV host proteins investigated displayed a directed and positive evolution pattern. The proteins that exhibited the highest pattern of non-random substitution are the p57 as follows: p57 (97%), p24 (77.5%), gp18 (74.8%), and p40 (72 %). This may be a demonstration that the super infection hypothesis resulted in positive selection of variants.

Super-infection interference was originally described for avian retroviruses (Steck and Rubin, 1966; Temin and Kassner, 1975) and was later also found to affect most noncytolytic (non destruction of infected cells), mammalian retroviruses (Weiss *et al.*, 1985). This phenomenon has also been described for other viruses, including noncytolytic variants of foot-and-mouth disease virus (de la Torre *et al.*, 1985) and measles virus (Fernandez and Celma, 1992). The mechanisms of these restriction phenomena have not yet been elucidated.

A recent study by Formella *et al.*, (2000) gives credence to the super-infection hypothesis as in their study the *Borna disease virus* was exposed to a bottle neck situation in which the virus content of a single persistently infected cell was allowed to spread in uninfected cell cultures. Various virus variants were created as a result. This supports the theoretical results obtained in this study. As BDV is noncytolytic, (Schwemmle *et al.*, 1998) and as it has no growth inhibitory effect on persistently infected cells, resistance to super-infection generates an ideal ecological niche for resident viruses to produce progeny without competition by genetically distinct viruses entering the cell from outside.

The results of the BDV inter-species protein analysis are in considerable contrast to the results displayed by the intra-species (virus) strain protein sequence. In contrast to the inter-species sequence analysis only the p57 glycoprotein, p40 nucleoprotein and p10 X proteins exhibit a pattern of non random evolution. These contrasting results can be interpreted as follows:

The positive selection of the proteins displayed in the intra-species (virus) sequence analysis may be a result of a pattern of nucleotide substitution that is physio-chemically conservative. Conservation may be evident in volume, polarity, hydrophilicity, or molecular weight of amino acids, depending presumably on the exact nature of the structural context of residue positions (Haydon *et al.*, 1998, 2001). Furthermore, as these strains were not obtained from different host species there is no reason for the antigenic variety, which may explain why the gp18 matrix protein in the virus strain protein sequence analysis conforms to a neutral theory of evolution.

In general it is not surprising that the p57 glycoprotein exhibits a positive mode of evolution in both the host species and virus species comparison, which may be directly linked to its glycoprotein function. The glycoprotein in BDV is involved in membrane fusion activity, which allows the entry of enveloped viruses into animal cells, and the polycistronic ability of the glycoprotein enables expression of two products which are 84 kDa and 43 kDa in size (Gonzalez-Dunia *et al.*, 1997a; 1998).

The evolutionary investigation in this study was the first time that such a study had been done involving the *Borna disease virus*. At the time of investigation not enough sequences were available to perform a comparative inter- and intra- species analysis of the polymerase protein and all X proteins, which may shed some light on intriguing strain deviations in transcription and replication; in particular as it has recently been discovered that the p10 X protein may play a role in nuclear localization of BDV proteins (Wolff, 2002). Intriguing results have been detected from this study, which is supported by recent laboratory studies indicating that a theoretical analytical approach has merit and presents a basis for further analysis of RNA negative stranded viruses.

4.4 FUTURE DIRECTIONS IN PHYLOGENETIC AND EVOLUTIONARY ANALYSIS

The present study has demonstrated that there are several differences from the previous *Bornaviridae* classification (Briese *et al.*, 1994). The results indicate that *Bornaviridae* should not be classified as separate family; it should rather be classified as a subfamily of the *Rhabdoviridae*. Furthermore, two additional motifs were discovered which have some relevance to the biological properties of the *Borna disease virus*. Future studies may utilize this discovery to generate three dimensional motifs of the polymerase protein. This in conjunction with the development of three dimensional protein models of other BDV proteins would enable a 3D co-localization of various proteins to confirm interaction patterns of the *Borna disease virus* which have been difficult to

trace with normal in vitro molecular methods, thereby enhancing research efforts in this field and may subsequently provide a mechanism to trace the entry pathway of this disease

A recent study (Nowotny *et al.*, 2000) has indicated that methods previously used to detect the *Borna disease virus* may have only been partially successful, as novel subtypes may exist, which escape detection by currently used RT-PCR protocols. The new findings imply that previous studies, which relied on RT-PCR technology, might have underestimated the true prevalence rates of human and animal BDV infections and might also have missed etiological correlations between BDV infection and certain neuropsychiatric disorders. The isolation of a BDV with a highly variant genome disproves the general opinion of high sequence conservation of all BDV genomes; the possibility that further, yet unidentified, BDV subtypes do exist should be seriously considered Should this be the case, then methods of evaluating the evolutionary nature of BDV as used in the above study may become even more vital as BDV may have a serious epidemiological impact.

In addition, as more RNA negative stranded virus data becomes available, including a complete BDV genome from a human isolate, an evolutionary comparison using the method utilized above to investigate the evolutionary nature of all RNA negative stranded viruses would be possible. Such an investigation would enhance the current phylogenetic analysis performed in this study and although extremely labour intensive as a result of the length of sequence data, would shed light on some puzzling aspects of the overall nature of RNA evolution, and would indicate just how many RNA negative viruses conform to the super-infection hypothesis described above.

In future, additional methods of evolutionary analysis, such as estimation of synonymous and non-synonymous mutation rates, may be utilized to examine the pattern of evolution of RNA negative stranded viruses in general, as this may aid in determining which codons are responsible for some of the significant antigenic effects observed in RNA viruses. As the field of negative stranded viruses encompasses some of the most dangerous viruses known to mankind such an analysis would offer a considerable breakthrough with regard to vaccine development. The investigation of individual proteins and their codons and establishing which proteins are subject to positive selection and are generating antigenic virulent products would open up a fountain of knowledge currently not available.

REFERENCES

Aiyar A. 2000. The use of CLUSTAL W and CLUSTAL X for multiple sequence alignment. Methods Molecular Biology 132: 221-41.

Altschul S.F, Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W. & Lipman D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25(17): 3389-402.

Amsterdam J., Winokur A., Dyson W., Herzog S., Gonzalez F., Rott R. & Koprowski H. 1985. *Borna Disease Virus*: A possible etiologic factor for human affective disorders? Archives of General Psychology 42: 1093-1096.

Aravind L., Watanabe H., Lipman D.J. & Koonin E.V. 2000. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. Proceedings of the National Academy of Sciences U.S.A. 97(21): 11319-24.

Auwanit W., Ayuthaya P., Nakaya T., Fujiwara S., Kurata T., Yamanishi K. & Ikuta K. 1996. Unusually high seroprevalence of *Borna Disease Virus* in clade E human immunodeficiency virus type I infected patients with sexually transmitted diseases in Thailand. Clinical Diagnostic Laboratory of Immunology 3: 590-593.

Bahmani M.K, Nowrouzian I., Nakaya T., Nakamura Y., Hagiwara K., Takahashi H., Rad M.A. & Ikuta K. 1996. Varied prevalence of *Borna Disease Virus* infection in Arabic, thoroughbred and their cross-bred horses in Iran. Virus Research 45(1): 1-13.

Baldi P., Chauvin Y., Hunkapiller T. & McClure M.A. 1994. Hidden Markov models of biological primary sequence information. Proceedings of the National Academy of Sciences U.S.A. 91(3): 1059-63.

Bailey T. & Gribskov M. 1998. Methods and statistics for combining motif match scores. Journal of Computational Biology 5: 211-221.

Barton G.J. & Sternberg M.J. 1987. Evaluation and improvements in the automatic alignment of protein sequences. Protein Engineering 1(2):89-94.

Becht H. & Richt J.A. 1996. *Borna disease*. In: Studdert M.J. (ed). Virus diseases of equines. Elsevier Science Publishers, London. Pp 235-44.

Bechter K., Bauer M., Estler H., Herzog S., Schuettler R. & Rott R. 1994. Expanded nuclear magnetic studies in *Borna Disease Virus* seropositive patients and control probands. Nervenarzt 65: 197-74.

Bechter K., Herzog S., Fleischer B., Schuettler R. & Rott R. 1987. Findings with nuclear magnetic resonance topography in psychiatric patients with and without serum antibodies to the virus of *Borna disease*. Nervenarzt 58: 617-24.

Bechter K., Schuettler R. & Herzog S. 1992. *Borna Disease Virus*: Possible causal agent in psychiatric and neurological disorders in two families. Psychiatry Research 42(3): 291-294.

Bode L., Durrwald R., Rantam F.A., Ferszt R. & Ludwig H. 1996. First isolates of infectious human *Borna disease virus* from patients with mood disorders. Molecular Psychiatry 1(3): 200-12.

Bode L., Ferszt R. & Czech G. 1993. *Borna Disease Virus* infection and affective disorders in man. Archives of Virology 7: 159-167.

Bode L., Riegel S., Lange H., Amsterdam J. & Lange H. 1992. Human infections with *Borna Disease Virus*. Seroprevalence in patients with chronic disease and healthy individuals. Journal of Medical Virology 36: 309-315.

Bode L., Riegel S., Ludwig H., Amsterdam J., Lange W. & Koprowski H. 1988. *Borna Disease Virus* specific antibodies in patients with HIV infection and with mental disorders. Lancet ii: 689.

Bode L., Zimmermann W., Ferszt R., Steinbach F. & Ludwig H. 1995. *Borna Disease Virus* genome transcribed and expressed in psychiatric patients. Nature Medicine 1: 232-236.

Briese T., Schneemann A., Lewis A., Park Y., Kim S., Ludwig H. & Lipkin W. 1994. Genomic Organization of *Borna Disease Virus*. Proceedings of the National Academy of Sciences U.S.A. 91: 4362-4366.

Camin J.H. & Sokal J. 1965. A method for deducing branching sequences in phylogeny. Evolution 19: 311-326.

Carbone K., Duchala C., Griffin J., Kincaid A. & Narayan O. 1987. Pathogenesis of *Borna disease* in rats: evidence that intra-axonal spread is the major route for virus dissemination and the determinant for disease incubation. Journal of Virology 61: 3431-3140.

Carbone K., Park S., Rubin S., Waltrip R. & Vogelsang G. 1991. *Borna disease*: Association with a maturation defect in the cellular immune response. Journal of Virology 65: 6154-64.

Cavalli-Sforza L.L. & Edwards A.W. 1967. Phylogenetic analysis: Models and estimation procedures. Evolution 32: 550-570.

Cubitt B. & de la Torre J. 1994. *Borna Disease Virus (*BDV) a nonsegmented RNA virus replicates in the nuclei of infected cells where infectious nucleoproteins are present. Journal of Virology 68: 1371-1381.

Cubitt B., Oldstone C. & de la Torre J. 1994a. Sequence and genome organization of *Borna Disease Virus.* Journal of Virology 68: 1371-1381.

Cubitt B., Oldstone C., Valcarel J. & de la Torre J. 1994b. RNA splicing contributes to the generation of mature mRNAs of *Borna Disease Virus*, a nonsegmented negative strand RNA virus. Virus Research 34: 69-79.

de la Torre J. 1994. Molecular biology of *Borna Disease Virus*: Prototype of a new group of animal viruses. Journal of Virology 68: 7699-7675.

de la Torre J., Bode L., Duerrwald R., Cubitt B. & Ludwig H. 1996. Sequence characterization of human *Borna Disease Virus.* Virus Research 44: 33-44.

de la Torre J.C., Davila M., Sobrino F., Ortin J. & Domingo, E. 1985. Establishment of cell lines persistently infected with foot-and-mouth disease virus. Virology 145(1): 24-35.

Delwart E.L. & Panganiban A.T. 1989. Role of reticuloendotheliosis virus envelope glycoprotein in superinfection interference. Journal of Virology 63(1): 273-80.

Dietz A., Weisser H.J., Kossel H. & Hausmann R. 1990. The gene for Klebsiella bacteriophage K11 RNA polymerase: sequence and comparison with the homologous genes of phages T7, T3, and SP6. Molecular and General Genetics 221(2): 283-6

Dittrich W., Bode L., Ludwig H., Kao M. & Schneider K. 1989. Learning deficiencies in *Borna Disease Virus* infected but clinically healthy rats. Biological Psychiatry 26: 6254-64.

Domingo E. & Holland J. 1988. *High error rates, population equilibrium, and evolution of RNA replication systems*. In: Domingo E., Holland J. & Ahlquist P. (eds.) RNA genetics, Volume III. CRC Press, Boca Raton. Pp 3-35.

Eck R. & Dayhoff M. 1966. *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation. Silver Spring, Maryland.

Eck R. & Dayhoff M. 1968. Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences. Science 152: 363-366.

Eddy S.R. 1995. Multiple alignment using hidden Markov models. The International Society of Computational Biology 3: 114-20.

Estabrook G.F. & Landrum L. 1975. A simple test for the possible simultaneous evolutionary divergence of two amino acid positions. Taxon 24: 609-613.

Estabrook G.F., Strauch J. & Fialla F. 1977a. An application of compatibility analysis to the Blackith's data on orthopteroid insects. Systematic Zoology 26: 269-276.

Estabrook G.F., Johnson C. & McMorris F. 1977b. When are two quantitative characters compatible? Journal of Mathematical Biology 4: 195-200.

Eisen J.A. 2000. Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. Current Opinion of Genetic Development 6: 606-11.

Felsenstein J. 1993. *PHYLIP (Phylogeny Inference Package) version* 3.5c. Seattle, Department of Genetics, University of Washington.

Feng D.F. & Doolittle R.F. 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. Journal of Molecular Evolution 25(4): 351-60.

Fernandez-Munoz R. & Celma M.L. 1992. Measles virus from a long-term persistently infected human T lymphoblastoid cell line, in contrast to the cytocidal parental virus, establishes an immediate persistence in the original cell line. Journal of General Virology 73: 2195-202.

Fitch W.M. 1971. Toward defining the course of evolution: Minimum change for a specific tree topology. Systematic Zoology 20: 406-416.

Fitch W.M. 1981. A non-sequential method for constructing trees and hierarchical classifications. Journal of Molecular Evolution 18: 30-37.

Fitch W.M. & Margoliash E. 1967. Construction of phylogenetic trees. Science 155: 279-284.

Formella S., Jehle C., Sauder C., Staeheli P. & Schwemmle M. 2000. Sequence variability of *Borna Disease Virus:* resistance to superinfection may contribute to high genome stability in persistently infected cells. Journal of Virology 17: 7878-83.

Frese S., Richt J.A. & Rott R. 1994. Ein Beitrag zur Epizootiologie der Bornaschen Krankheit des Pferdes. Wiener Tierzaertzliche Monatschrift 81: 374-9.

Fu Z., Amsterdam J., Kao M., Shankar V., Koprowski, D. & Dietzchold B. 1993. Detection of *Borna Disease Virus* specific antibodies from patients with affective disorders by western immunoblot technique. Journal of Affective disorders 27: 61-68.

Gao Y. & Lenard J. 1995. Cooperative binding of multimeric phosphoprotein (P) of *Vesicular Stomatitis Virus* to polymerase (L) and template: pathways of assembly. Journal of Virology 69: 7718-7723.

Gillespie J. H. 1991. *The Causes of Molecular Evolution.* Oxford University Press, Oxford.

Goldbach R. 1987. Genome similarities between plant and animal RNA viruses. Microbiological Sciences 4(7): 197-202.

Goldbach R. & de Haan P. 1994. *RNA viral supergroups and the evolution of RNA viruses.* In: Morse S. (ed). The Evolutionary Biology of Viruses. Raven Press, New York.

Gonnet G.H., Cohen M.A. & Benner S.A. 1992. Exhaustive matching of the entire protein sequence database. Science 256: 1443-1445.

Gonzalez-Dunia D., Cubitt B. & de la Torre J. 1998. Mechanism of *Borna Disease Virus* entry into cells. Journal of Virology 72: 783-788.

Gonzalez-Dunia D., Cubitt B., Graesser F. & de la Torre J. 1997a. Characterization of *Borna Disease Virus* p56 protein, a surface glycoprotein involved in virus entry. Journal of Virology 71: 3208-3218.

Gonzalez-Dunia D., Sauder C. & de la Torre J. 1997b. *Borna Disease Virus* and the brain. Brain Research Bulletin 44: 647-664.

Gosztonyi G., Dietzschold B., Kao M., Rupprecht C.E., Ludwig H. & Koprowski H. 1993. *Rabies* and *Borna disease*. A comparative pathogenetic study of two neurovirulent agents. Laboratory Investigation 68(3): 285-95.

Grundy W., Bailey T., Elkan C. & Baker M. 1997. Hidden Markov Model Analysis of Motifs in Steroid Dehydrogenases and their Homologs. Biochemical and Biophysical Research Communications 231: 760-766.

Haas B., Becht H. & Rott R. 1986. Purification and properties of an intranuclear virus-specific antigen from tissue infected with *Borna Disease Virus*. Journal of General Virology 67: 235-241.

Hatalski G., Lewis A. & Lipkin W. 1997. *Borna disease*. Emerging Infectious Diseases 3: 129-135.

Hayashida H., Toh H., Kikuno R. & Miyata T. 1985. Evolution of *influenza* virus genes. Molecular Biological Evolution 2(4):289-303.

Haydon D., Knowles N. & McCauley J. 1998. Methods for the detection of non-random base substitutions in Virus genes: Models of Synonymous nucleotide substitution in *Picornavirus* genes. Virus Genes 16: 252-266.

Haydon D.T. & Woolhouse M.E.J. 1998. Immune avoidance strategies in RNA viruses: Fitness continuums arising from trade-offs between immunogenicity and antigenic variability. Journal of Theoretical Biology 193: 603-612.

Haydon D.T., Bastos A., Samuel A. & Knowles N. 2001. Evidence for positive selection in foot-and-mouth-disease-virus capsid genes from field isolates. Genetics 157: 7-15.

Henikoff S. & Henikoff J. 1992. Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences U.S.A. 89(22): 10915-9.

Heringa J. & Argos P. 1994. *Evolution of viruses as recorded by their polymerase sequences.* In: Morse S. (ed). The Evolutionary Biology of Viruses. Raven Press, New York. Pp 87-103.

Herzog S., Pfeuffer I., Haberzettl K., Feldmann H., Frese K., Bechter K. & Richt J.A. 1997. Molecular characterization of *Borna disease virus* from naturally infected animals and possible links to human disorders. Archives of Virology 13: 183-90.

Holland J. 1993. *Replication error, quasispecies populations and extreme evolution rates of RNA viruses.* In: Morse S. (ed). Emerging Viruses. Oxford University Press, New York, NY.

Hsu T., Cabone K., Rubin S., Vonderfecht S. & Eiden J. 1994. *Borna Disease Virus* p24 and p38/40 synthesized in baculovirus expression system: Virus protein interactions in insects and mammalian cells. Virology 2047: 854-859.

Igata-Yi R., Yamaguchi K., Yoshiki K., Takemoto S. Yamasaki H., Matsuoka M. & Miyakawa T. 1996. *Borna Disease Virus* and the consumption of raw horse meat. Nature Medicine 2: 948-949.

Ina Y. 1995. New methods for estimating the numbers of synonymous and non synonymous substitutions. Journal of Molecular Evolution 40: 190-226.

Jeanmougin F., Thompson J.D., Gouy M., Higgins D.G. & Gibson T.J. 1998. Multiple sequence alignment with Clustal X. Trends in Biochemical Sciences 23(10): 403-5

Kamer G. & Argos P. 1984. Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. Nucleic Acid Research: 12 (7269-7782).

Kao M., Hamir A.N., Rupprecht C.E., Fu Z.F., Shankar V., Koprowski H. & Dietzschold B. 1993. Detection of antibodies against *Borna Disease Virus* in sera and cerebrospinal fluid of horses in the USA. The Veterinary Record 132(10): 241-4.

Karplus K., Barrett C. & Hughey R. 1998. Hidden Markov models for detecting remote protein homologies. Bioinformatics 14(10): 846-56.

Kimura M. 1968. Evolutionary Rate at the Molecular Level. Nature 217: 624-626.

Kimura M. 1983. The Neutral Theory of Molecular Evolution. Cambridge. University Press, Cambridge.

Kishi M., Nakaya T., Nakamura Y., Makaya T., Asahi S., Tobiume M., Yamaguti K., Machii T., Inagi R., Yamanishi K. & Ikuta K. 1995. *Borna Disease Virus* RNA in human peripheral blood mononuclear cells. Federation of European Biochemical Sciences Letters 364: 293-297.

Kliche S., Briese T., Henschen A., Stitz L. & Lipkin W. 1994. Characterization of a *Borna Disease Virus* glycoprotein, gp18. Journal of Virology 68: 6918-6923.

Klupp B., Baumeister J., Dietz P., Granzow H. & Mettenleiter T.C. 1998. Pseudorabies virus glycoprotein gK is a structural component of virions involved in virus release but not required for entry. Journal of Virology 72: 1949-1958.

Kobayashi T., Shoya Y., Koda T., Takashima I., Lai P., Ikuta K., Kakinuma M. and Kishi M. 1998. Nuclear targeting activity associated with the amino terminal region of the *Borna disease virus* nucleoprotein. Virology 243(1): 188-97.

Koonin E. 1991. The phylogeny of RNA dependent polymerases of positive strand viruses. Journal of General Virology 72: 2197-2206.

Koonin E., Aravind L. & Kondrashov A.S. 2000. The impact of comparative genomics on our understanding of evolution. Cell 101(6): 573-6.

Kyte J. 1995. *Structure in Protein Chemistry.* Garland Publishing.

Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., Devon K., Dewar K., Doyle M., FitzHugh W., Funke R., Gage D., Harris K., Heaford A., Howland J., Kann L., Lehoczky J., LeVine R., McEwan P., McKernan K., Meldrim J., Mesirov J.P., Miranda C., Morris W., Naylor J., Raymond C., Rosetti M., Santos R., Sheridan A., Sougnez C., Stange-Thomann N., Stojanovic N., Subramanian A., Wyman D., Rogers J., Sulston J., Ainscough R., Beck S., Bentley D., Burton J., Clee C., Carter N., Coulson A., Deadman R., Deloukas P., Dunham A., Dunham I., Durbin R., French L., Grafham D., Gregory S., Hubbard T., Humphray S., Hunt A., Jones M., Lloyd C., McMurray A., Matthews L., Mercer S., Milne S., Mullikin J.C., Mungall A., Plumb R., Ross M., Shownkeen R., Sims S., Waterston R.H., Wilson R.K., Hillier L.W., McPherson J.D., Marra M.A., Mardis E.R., Fulton L.A., Chinwalla A.T., Pepin K.H., Gish W.R., Chissoe S.L., Wendl M.C., Delehaunty K.D., Miner T.L., Delehaunty A., Kramer J.B., Cook L.L., Fulton R.S., Johnson D.L., Minx P.J., Clifton S.W., Hawkins T., Branscomb E., Predki P., Richardson P., Wenning S., Slezak T., Doggett N., Cheng J.F., Olsen A., Lucas S., Elkin C., Uberbacher E.,

Frazier M., Gibbs R.A., Muzny D.M., Scherer S.E., Bouck J.B., Sodergren E.J., Worley K.C., Rives C.M., Gorrell J.H., Metzker M.L., Naylor S.L., Kucherlapati R.S., Nelson D.L., Weinstock G.M., Sakaki Y., Fujiyama A., Hattori M., Yada T., Toyoda A., Itoh T., Kawagoe C., Watanabe H., Totoki Y., Taylor T., Weissenbach J., Heilig R., Saurin W., Artiguenave F., Brottier P., Bruls T., Pelletier E., Robert C., Wincker P., Smith D.R., Doucette-Stamm L., Rubenfield M., Weinstock K., Lee H.M., Dubois J., Rosenthal A., Platzer M., Nyakatura G., Taudien S., Rump A., Yang H., Yu J., Wang J., Huang G., Gu J., Hood L., Rowen L., Madan A., Qin S., Davis R.W., Federspiel N.A., Abola A.P., Proctor M.J., Myers R.M., Schmutz J., Dickson M., Grimwood J., Cox D.R., Olson M.V., Kaul R., Raymond C., Shimizu N., Kawasaki K., Minoshima S., Evans G.A., Athanasiou M., Schultz R., Roe B.A., Chen F., Pan H., Ramser J., Lehrach H., Reinhardt R., McCombie W.R., de la Bastide M., Dedhia N., Blocker H., Hornischer K., Nordsiek G., Agarwala R., Aravind L., Bailey J.A., Bateman A., Batzoglou S., Birney E., Bork P., Brown D.G., Burge C.B., Cerutti L., Chen H.C., Church D., Clamp M., Copley R.R., Doerks T., Eddy S.R., Eichler E.E., Furey T.S., Galagan J., Gilbert J.G., Harmon C., Hayashizaki Y., Haussler D., Hermjakob H., Hokamp K., Jang W., Johnson L.S., Jones T.A., Kasif S., Kaspryzk A., Kennedy S., Kent W.J., Kitts P., Koonin E.V., Korf I., Kulp D., Lancet D., Lowe T.M., McLysaght A., Mikkelsen T., Moran J.V., Mulder N., Pollara V.J., Ponting C.P., Schuler G., Schultz J., Slater G., Smit A.F., Stupka E. & Szustakowski J. 2001. Initial sequencing and analysis of the human genome. Nature 409(6822): 860-921.

Lassmann T. & Sonnhammer E. 2002. Quality assessment of multiple alignment programs. Federation of European Biochemical Sciences Letters 529: 126-130.

Le Quesne W.J. 1979. Compatibility analysis and the uniquely derived character concept. Systematic Zoology 28: 92-94.

Lenz G.R., Nash H.M. & Jindal S. 2000. Chemical ligands, genomics and drug discovery. Drug Discovery Today 5(4): 145-156.

Li W., Wv C. & Luo C. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Molecular Biological Evolution 2: 150-174.

Ludwig H., Bode L. & Gosztonyi G. 1989. A persistent virus infection of the central nervous system. Progressive medical virology 35: 107-151.

Lundgren A., Zimmermann W., Bode L., Czech G., Gosztonyi G., Lindberg R. & Ludwig H. 1995. Staggering disease in cats. Isolation and characterization of feline *Borna Disease Virus*. Journal of General Virology 76: 2215-222.

Martinez M.A., Carrillo C., Gonzalez-Candelas F., Moya A., Domingo E. & Sobrino F. 1991. Fitness alteration of foot-and-mouth disease virus mutants: measurement of adaptability of viral quasispecies. Journal of Virology 65(7): 3954-7.

Mateu M.G., Martinez M.A., Capucci L., Andreu D., Giralt E., Sobrino F., Brocchi E. & Domingo E. 1990. A single amino acid substitution affects multiple overlapping epitopes in the major antigenic site of foot-and-mouth disease virus of serotype C. Journal of General Virology 71: 629-37.

McClure M.A., Thibault K.J., Hatalski C.G. & Lipkin W.I. 1992.Sequence similarity between *Borna Disease Virus* p40 and a duplicated domain within the *paramyxovirus* and *rhabdovirus* polymerase proteins. Journal of Virology 66(11): 6572-7.

Morgenstern B. 1999. *DIALIGN 2*: improvement of the segment-to-segment approach to multiple sequence alignment. Bioinformatics 15: 211 - 218.

Morgenstern B., Dress A. & Werner T. 1996. Multiple DNA and protein sequence alignment based on segment-to-segment comparison. Proceedings of the national academy of Sciences U.S.A. 93: 12098 – 12103.

Morgenstern B., Rinner O., Abdeddaïm S., Haase D., Mayer K., Dress A. & Mewes H. 2002. Exon Discovery by Genomic Sequence Alignment. Bioinformatics 18: 777-787.

Moore C.W., Hampsey D.M., Ernst J. & Sherman F. 1988. Differential mismatch repair can explain the disproportionalities between physical distances and recombination frequencies of cyc1 mutations in yeast. Genetics 119(1): 21-34.

Morse S. 1994. Prediction and biological evolution. Annals of the New York Academy of Sciences 740: 436-8.

Murphy F., Fauquet C.M., Mayo M.A., Jarvis A., Ghabrial S., Summers M., Martelli G. & Bishop D. 1995. (eds). Sixth Report of the International Committee on Taxonomy of Viruses. Archives of Virology. Springer Verlag, New York.

Muse S. 1996. Estimating synonymous and nonsynonymous substitution rates. Molecular Biological Evolution 13(1): 105-14.

Nakamura Y., Kishi M., Nakaya T., Asahi S., Tanaka H., Sentsui H., Ikeda K. & Ikuta K. 1995. Demonstration of *Borna Disease Virus* RNA in peripheral blood mononuclear cells from healthy horses in Japan. Vaccine 3(12): 1076-9.

Nakaya T., Takahashi H., Nakamura Y., Asahi S., Tobiume M., Kurtsune H., Kitani T., Yaminishi K. & Ikuta. K. 1996. Demonstration of *Borna Disease Virus* RNA in peripheral blood mononuclear cells derived from Japanese patients with chronic fatigue syndrome. Federation of European Biochemical Sciences Letters 378: 145-149.

Narayan O., Hertzog S., Frese K., Scheefers H. & Rott R. 1983. Behavioural disease in rats caused by immunopathological responses to persistent *Borna virus* in the brain. Science 220: 1401-3.

Needleman S.B. & Wunsch C.D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology 48(3): 443-53.

Nei M. & Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biological Evolution 3(5): 418-26.

Nelson K.E. 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature 399(6734): 323-9.

Neuwald A.F., Liu J.S., Lipman D.J. & Lawrence C.E. 1997. Extracting protein alignment models from the sequence database. Nucleic Acids Research 25(9): 1665-77.

Niehaus F., Bertoldo C., Kahler M. & Antranikian G. 1999. Extremophiles as a source of novel enzymes for industrial application. Applied Microbiology and Biotechnology 51(6): 711-29.

Notredame C. & Higgins D. 1996. SAGA: sequence alignment by genetic algorithm. Nucleic Acids Research 24(8): 1515-24.

Nowotny N., Kolodziejek J., Jehle C.O., Suchy A., Staeheli P. & Schwemmle M. 2000. Isolation and characterization of a new subtype of *Borna Disease Virus*. Journal of Virology 74(12): 5655-8.

O'Neill R.E., Talon J. & Palese P. 1998. The *influenza virus* NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. European Molecular Biology Organization Journal 17(1): 288-96.

Ohta T. 1993. An examination of the generation-time effect on molecular evolution. Proceedings of the National Academy of Sciences U.S.A. 90(22): 10676-80.

Palacios G., Casas I., Tenorio A. & Freire C. 2002. Molecular identification of enterovirus by analyzing a partial VP1 genomic region with different methods. Journal of Clinical Microbiology 40(1): 182-92.

Phillips A., Janies D. & Wheeler W. 2000. Multiple sequence alignment in phylogenetic analysis. Molecular Phylogenetics and Evolution 16(3): 317-30.

Pflanz O., Rentzsch C., Batra A., Batra A., Winkler T., Buettner M., Rziha H. & Stitz L. 1999. Pathogenesis of *Borna Disease Virus*: Granulocyte Fractions of Psychiatric Patients Harbor Infectious Virus in the Absence of Antiviral Antibodies. Journal of Virology 73(8): 6251-6256.

Pringle C.R. 1991. The order Mononegavirales. Archives of Virology 117: 137-140.

Pringle C.R. 1997. The order Mononegavirales--current status. Archives of Virology 142(11): 2321-6.

Pringle C.R. & Easton A.J. 1997. Monopartite negative strand RNA genomes. Seminars in Virology 8: 49-57.

Pyper J. & Gartner A. 1997. Molecular basis for the differential subcellular localization of the 38- and 39-kilodalton structural proteins of *Borna disease virus*. Journal of Virology 71(7): 5133-9.

Pyper J., Richt J., Brown L., Rott R., Narayan O. & Clements J. 1993. Genomic organization of a structural protein of Borna Disease Virus revealed by a cDNA clone encoding the 38 kDa protein. Virology 195: 229-338.

Pyper J., Clements J.E. & Zink M.C. 1998. The nucleolus is the site of *Borna Disease Virus* RNA transcription and replication. Journal of Virology 72(9): 7697-702.

Richt J.A., Fuerbringer T., Koch A., Pfeuffer I., Herden C., Bause N. & Garten W. 1998. Processing of the *Borna Disease Virus* glycoprotein gp94 by the subtilisin like endoproteinase furin. Journal of Virology 72: 4528-4533.

Richt J.A., Herzog S., Haberzettl K. & Rott R. 1993. Demonstration of *Borna Disease Virus* specific RNA in secretions of naturally infected horses by the polymerase chain reaction. Medical Microbiology and Immunology 182: 293-304.

Richt J.A., Hertzog S., Schmeed A., Frese K. & Rott R. 1994. *Current knowledge about Borna disease.* In: Nakajima H. & Plowright W. (eds). Equine infectious diseases VII. R and W publications, Newmarket. Pp 55-60.

Richt J.A., Pfeuffer I., Christ M., Frese K., Bechter K. & Herzog S. 1997. *Borna Disease Virus* infection in animals and humans. Emerging Infectious diseases 3: 343-352.

Richt J.A., VandeWoude S., Zink M.C., Clements J.E., Herzog S., Stitz L., Rott R. & Narayan O. 1992. Infection with *Borna Disease Virus*: molecular and immunobiological characterization of the agent. Clinical Infectious Diseases 14(6): 1240-50.

Roizman B. 1991. *Multiplication of viruses: an overview.* In: Fields B. & Knipe D. (eds). Virology. Raven Press, New York. Pp 87-95.

Rocha E., Cox N.J., Black R.A., Harmon M.W., Harrison C.J. & Kendal A.P. 1991. Antigenic and genetic variation in influenza A (H1N1) virus isolates recovered from a persistently infected immunodeficient child. Journal of Virology 65: 2340-2350.

Rott R. & Becht H. 1995. *Natural and experimental Borna disease in animals.* In: Koprowski H. & Lipkin I. (eds). *Borna disease.* Springer Verlag, Berlin. Pp 17-30.

Rott R., Herzog S., Bechter K. & Frese K. 1991. *Borna disease* a possible hazard for man? Archives of Virology 118: 143-149.

Rott R., Herzog S., Fleischer B., Winokur A., Amsterdam J., Dyson W. & Koprowski. H. 1985. Detection of serum antibodies to *Borna Disease Virus* in patients with psychiatric disorders. Science 228: 755-756.

Saitou N. & Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406-425.

Sankoff D. 1975. Minimal mutation trees of sequences. SIAM Journal of Applied Mathematics 21: 35-42.

Sankoff D. & Rousseau P. 1975. Locating the vertices of Steiner tree in an arbitrary metric space. Mathematical Programming 9: 240-246.

Sauder C., Mueller A., Cubitt B., Mayer J., Steinmetz J., Trabert W., Ziegler B., Wanke L., Mueller Lantsch N., de la Torre J. & Graeser F. 1996. Detection of *Borna disease* specific antibodies and BDV RNA in psychiatric patients. Journal of Virology 70: 7713-7724.

Schneemann A., Schneider O., Lamb R. & Lipkin W. 1994. Identification of signal sequences that control transcription of *Borna Disease Virus*, a nonsegmented, negative strand RNA virus. Journal of Virology 68: 5007-5012.

Schneemann A., Schneider P., Lamb R. & Lipkin W. 1995. The remarkable coding strategy of *Borna Disease Virus*. A new member of the nonsegmented negative strand RNA viruses. Virology 210: 1-8.

Schneider P., Hatalski C., Lewis W. & Lipkin W. 1997. Biochemical and functional analysis of *Borna Disease Virus* G protein. Journal of Virology 71: 331-336.

Schneider P., Schneemann A. & Lipkin W. 1994. RNA splicing *in Borna Disease Virus*, a nonsegmented negative strand RNA virus. Journal of Virology 68: 5007-5012.

Schuppel K., Kinne J., Lebelt J. & Reinacher M. 1995. Zwei Faelle von *Bornascher* Krankheit bei Varis (*Memur variegatus*). Tagung der Fachgruppe, Pathologie der DVG, Hamburg.

Schwemmle M., De B., Shi L., Banerjee A., Lipkin W., Bishnu D., Shi L. & Amiya B. 1997. *Borna Disease Virus* P protein is phosphorylated by protein kinase Cepsilon and casein kinase II. Journal of Biological Chemistry 272: 21818-21823.

Schwemmle M., Salvatore M., Shi L., Richt J., Lee C. & Lipkin W. 1998. Interactions of the *Borna Disease Virus* P, N and X proteins and their functional Implications. Journal of Biological Chemistry 273: 9007-9012.

Shoya Y., Kobayashi T., Koda T., Ikuta K., Kakinuma M. & Kishi M. 1998. Two proline-rich nuclear localization signals in the amino- and carboxyl-terminal regions of the *Borna Disease Virus* phosphoprotein. Journal of Virology 72(12): 9755-62.

Smith T.F. & Waterman M.S. 1981. Identification of common molecular subsequences. Journal of Molecular Biology 147(1): 195-7.

Smith T., Waterman M. & Fitch W. 1981. Comparative biosequence metrics. Journal of Molecular Evolution 18: 38-46.

Sneath P. & Sokal R. 1973. *Numerical Taxonomy*. W.H. Freeman and Company, San Francisco. Pp 230-234.

Solbrig M., Koob G., Fallon J., Reid S. & Lipkin W. 1996a. Prefrontal cortex dysfunction in *Borna Disease Virus* infected rats. Biological Psychology 40: 629- 636.

Solbrig M., Koob G., Joyce J. & Lipkin W. 1996b. A neural substrate of hyperactivity in *Borna disease*: Changes in brain dopamine receptors. Virology 2222: 332-338.

Sprankel H., Richartz K., Ludwig H. & Rott R. 1978. Behavioural alterations in tree shrews (*Tupaia glis*, Diard 1820) induced by *Borna Disease Virus*. Medical Microbiology and Immunology 165: 1-18.

Steck F.T. & Rubin H. 1966. The mechanism of interference between an *Avian Leukosis virus* and *Rous Sarcoma virus*. Establishment of interference. Virology 29(4): 628-41.

Stitz L., Krey H. & Ludwig H. 1980. *Borna Disease Virus* in Rhesus monkeys as a model for uveo cerebral symptoms. Journal of Medical Virology 6: 333-340.

Strauss E.G. & Strauss J.H. 1983. Replication strategies of the single stranded RNA viruses of eukaryotes. Current Topics in Microbiological Immunology 105: 1-98.

Taylor W.R. 1988. A flexible method to align large numbers of biological sequences. Journal of Molecular Evolution 28: 161-169.

Temin H. 1989. Retrovirus variation and evolution. Genome 31(1): 17-22.

Temin H. 1993. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. Proceedings of the national academy of sciences U.S.A. 90: 6900-6903.

Temin H. & Kassner V. 1975. Replication of reticuloendotheliosis viruses in cell culture: chronic infection. Journal of General Virology 27(3): 267-74.

Thiedemann N., Presek P., Rott R. & Stitz L. 1992. Antigenic relationship and virus P protein with the L protein or the N protein in cells expressed in recombinant proteins. Virology 208: 821-826.

Thierer J., Riehle H., Grebenstein O., Binz T., Herzog S., Thiedemann N., Stitz L., Rott R., Lottspeich F. & Niemann H. 1992. The 24 K protein of *Borna Disease Virus*. Journal of General Virology 73: 413-416.

Thompson J.D., Higgins D.G. & Gibson T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22(22): 4673-80.

Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F. & Higgins D.G. 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 25(24): 4876-82.

Thompson J.D., Plewniak F. & Poch O. 1999. A comprehensive comparison of multiple sequence alignment programs. Nucleic Acids Research 27(13): 2682-90.

Waltrip R., Buchanan R., Summerfelt A., Breier A., Carpenter W., Bryant N., Rubin S. & Carbone K. 1995. *Borna Disease Virus* and schizophrenia. Psychological Research 56: 33-44.

Weiss A., Hollander H. & Stobo J. 1985. Acquired immunodeficiency syndrome: epidemiology, virology, and immunology. Annual Reviews Medicine 36: 545-62.

Weissenboeck H., Nowotny N., Caplazi P., Kolodziejek J. & Wehrensperger F. 1998. *Borna disease* in a dog with lethal meningoencephalitis. Journal of Clinical Microbiology 36: 2127-2130.

Wickner R.B. 1993. Double-strand RNA virus replication and packaging. Journal of Biological Chemistry 268: 3797-3800.

Wolff T., Unterstab G., Heins G., Richt J.A. & Kann M. 2002. Characterization of an unusual importin alpha binding motif in the *Borna Disease Virus* p10 protein that directs nuclear import. Journal of Biological Chemistry 277(14): 12151-7.

Zuckerkandl E. & Pauling L. 1962. *Molecular disease, evolution, and genetic heterogeneity.* In: Kasha M. & Pullman B. (eds). Horizons in Biochemistry. Academic Press, New York. Pp 189-225.