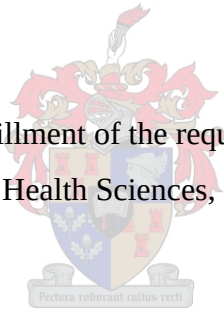
Molecular epidemiology of mother-to-child transmission of HIV-1 in children at Tygerberg Hospital.

Stephen Nicolaas Jacques Korsman



Assignment presented in partial fulfillment of the requirements for the degree of Master of Medicine at the Faculty of Health Sciences, University of Stellenbosch.

December 2006

Promoter: Dr G U van Zyl

Co-promoter: Prof S Engelbrecht

Declaration

I, the undersigned, hereby declare that the work contained in this assignment is my own original work and that it has not previously in its entirety or in part been submitted at any university for a degree.

Signature:.....

Date:.....

Summary

One of the major routes of transmission of human immunodeficiency virus (HIV) in the developing world is vertical transmission from mother to infant – pre-, intra-, or post-partum. In the Western Cape, HIV-1 subtype C is the predominant subtype in the heterosexual population, and this trend was expected to be seen amongst cases of mother-to-child transmission of HIV. The aim of this study was to perform genetic characterisation and phylogenetic analysis of the HIV-1 genome in positive serum/plasma samples obtained from children (age 0 to 18 months) from 2000-2002, and temporally related specimens from their mothers. We obtained 27 suitable pairs of samples taken within 6 months of delivery. From this pool, we obtained 21 infant DNA sequences and 17 maternal sequences, resulting in 16 mother-infant pairs. All patient sequences were identified as HIV-1 subtype C, and, as expected, mother and infant viral sequences clustered together. In some cases where a mother was suspected to have two dominant quasispecies based on the electropherogram, only one sequence was detectable in the infant. Single or multiple amino acid deletions were consistent between mothers and infants, and some pairs showed the same amino acid deletions seen in other pairs.

Opsomming

Een van die belangrikste roetes van verpreiding van menslike immuuniteitsgebreksvirus (MIV) in die ontwikkelende wêreld is verspreiding van moeder na kind – voor, tydens, of na geboorte. In die Wes-Kaap is MIV-1 subtipe C die algemeenste subtipe onder die heteroseksuele bevolking en hierdie tendens is ook verwag by die verspreiding van MIV van moeder na kind. Die doel van hierdie projek is die genetiese karakterisering en filogenetiese analise van die MIV-1 genoom in positiewe serum/plasma monsters van kinders (ouderdom 0 – 18 maande) en hulle moeders. Die monsters is geduurende 2000-2002 versamel en monsters van die moeders is gedurende dieselfde periode geneem. Ons het 27 geskikte pare monsters binne 6 maande van geboorte gekry. Van dié groep, het ons 21 DNS volgordes van babas en 17 volgordes van moeders gekry, wat vir ons 16 moeder/baba pare gegee het. Alle pasiënt volgordes is as MIV-1 subtipe C geïdentifiseer, en moeder en baba virale volgordes het, soos verwag, bymekaar gegroepeer. In sommige gevalle waar die moeder se elektroferogram dalk twee dominante quasispesies getoon het, is net een volgorde by die baba gevind. Enkele of veelvuldige aminosuur delesies het altyd in beide die moeder en baba van ’n paar voorgekom. Sommige pare het dieselfde delesies bevat wat ook in ander pare voorgekom het.

Acknowledgements

I wish to thank the following people for their assistance in this project:

Prof Susan Engelbrecht, for guidance as co-promoter

Prof Estrelita Janse van Rensburg, for guidance during the initial phase of the study as head of department and initial promoter

Dr Gert van Zyl, for guidance during the latter phase of the study as promoter

Mrs. Annette Laten, for performing the sequencing reactions and for assistance with translation

Ms. Brenda Robson, for advice and support with the troubleshooting

I also wish to thank the National Health Laboratory Service and the Polio Research Foundation for funding this project.

Contents

Abbreviations used

A	Adenine
ARV	Antiretroviral drug
ASSA	Actuarial Society of South Africa
AZT	azidothymidine
C	Cytosine
CRF	Circulating recombinant form
CTL	Cytotoxic T-lymphocyte
DC-SIGN	dendritic cell-associated ICAM-grabbing non-integrin
DC-SIGNR	DC-SIGN-related molecule
dNTP	deoxy-nucleotide triphosphate
G	Guanine
HAART	Highly active antiretroviral therapy
HIV	Human immunodeficiency virus
HLA	Human leukocyte antigen
HMA	Heteroduplex mobility assay
HSRC	Human Sciences Research Council
kb	kilobases
LTR	Long terminal repeat
MTCT	Mother to child transmission
NHLS	National Health Laboratory Service
PCP	Pneumocystis pneumonia
PCR	Polymerase chain reaction
PI	Protease inhibitor
RT	Reverse transcriptase
RT-PCR	Reverse transcription polymerase chain reaction
SIV	Simian immunodeficiency virus
T	Thymine
TAE	Tris acetate - EDTA
URF	Unique recombinant form
WHO	World Health Organisation

Chapter 1 - Literature review

1.1 Introduction

One of the major routes of spread of the human immunodeficiency virus (HIV) in the developing world is vertical transmission from mother to infant *in utero*, intrapartum, or post-partum. This means of spread has been studied from different angles, such as mechanism of spread, and factors that can be introduced to decrease this transmission. In regard to decreasing transmission, the role of breastfeeding and replacement feeding, minimising the infant's exposure to maternal secretions, and the use of prophylactic antiretroviral drugs have been the most well-described (Domachowske, 1996; Lyall *et al.*, 1998; McGowan, 2000).

At Tygerberg Hospital, many of the women seen at the antenatal clinic are HIV positive. In South Africa, prevalence studies are largely based on women attending antenatal clinics, and indicate that 29.5% of women at these clinics are infected with HIV (15.4% in the Western Cape) (Department of Health, South Africa, 2005). We have one of the largest epidemics worldwide - 2005 World Health Organisation reports show between 4.3 million and 6 million infected people in South Africa (WHO, 2005b).

Antiretroviral (ARV) drug usage in developed countries is now common, while South Africa is one of the few remaining places in which ARVs were not administered for prevention of mother-to-child transmission (MTCT) prior to 2002. Also making South Africa ideal for a molecular study of the genome of the virus transmitted from mother to child is a high prevalence of HIV, particularly MTCT.

This gave us an opportunity to study the molecular epidemiology of HIV-1 in MTCT in ARV naïve and ARV treated subjects, neither of which has been widely studied in South Africa. Few studies have been found dealing with this aspect of HIV – similar studies of significant size have taken place in Argentina (23 infants) (Carrillo *et al.*, 2002), the Congo (15 infants) (Mokili *et al.*, 2002), the USA (141 infants) (Krogstad *et al.*, 2002), Canada (103 mothers) (Akouamba *et al.*, 2005) and elsewhere (Nicoll *et al.*, 1998; Ho *et al.*, 2001; Verhofstede *et al.*, 2003), with a study in Russia dealing with nosocomial transmission in the late 1980's (22 infants) (Bobkov *et al.*, 1994).

1.2 HIV worldwide

The human immunodeficiency virus, HIV, is believed to have been introduced into the human race in several transmission events from primates harbouring a related virus, Simian Immunodeficiency Virus (SIV). There are two types – HIV-1 and HIV-2, the latter being less pathogenic, and the major impact of HIV-2 is geographically limited to West Africa (Lemey *et al.*, 2003). HIV-1, however, has caused a worldwide pandemic, first noticed in the late 1970s, with the virus itself being identified in 1983 and 1984 by two teams in the USA and France (Karpas, 2004).

The World Health Organisation estimates that there were about 40.3 million people living with HIV in 2005. Of these, 25.8 million live in sub-Saharan Africa, of which 3.2 million were newly infected individuals. An estimated 2.4 million people died of AIDS during the same year (WHO, 2005a).

1.3 HIV in South Africa

HIV in South Africa was first noticed in 1983 in white homosexual males, infected with subtypes B and D (Ras *et al.*, 1983; Loxton *et al.*, 2005). From 1989, a second and separate epidemic was noticed in black heterosexual people. The virus in this population was identified as being subtype C (van Harmelen *et al.*, 1997). Today, subtypes A, D and G, and the circulating recombinant form CRF02_AG, are also seen in our population (Bredell *et al.*, 2002).

Although there is some disagreement as to the actual prevalence of HIV in South Africa, it is clear from the surveys of both antenatal clinics and households that the prevalence is significant.

The Department of Health's National HIV and Syphilis Seroprevalence Survey in 2004 shows that 29.5% of the pregnant women sampled were infected with HIV, compared to less than 1% in 1990. Based on certain assumptions, they have extrapolated the data to the whole population, indicating between 5.7 and 6.7 million individuals infected with HIV in South Africa. Of those, between 2.6 and 3.1 million are male, and between 3.1 and 3.6 million are female. The highest prevalence amongst pregnant women was seen in KwaZulu-Natal, with 40.7% infected, while the lowest prevalence is in the Western Cape, with 15.4% infected (Department of Health, South Africa, 2005).

The Human Sciences Research Council (HSRC) conducted a survey of households in 2002, which indicated a prevalence of 11.4% in the population older than 2 years, with 9.5% of males and 12.8% of females infected. This is similar to the 11.2% estimated through statistical modeling by the Department of Health in 2002. This means that about 55% of HIV-infected adults in Africa are women of childbearing age, compared to 10% in Australia and New Zealand, where intravenous drug use and homosexual spread are the major routes of transmission (Gayle and Hill, 2001; Dabis and Ekpini, 2002; Department of Health, South Africa, 2005). The HIV prevalence in women who reported being pregnant in the previous 12 months was 24%, similar to the 24.8% estimated by the 2001 antenatal clinic data. A discrepancy arose in the Western Cape data between these two studies – the 2002 HSRC study showed a prevalence of 18.5% in women aged 15-49 years, while the antenatal survey of the same year showed a prevalence of 12.4%. The Western Cape was the only province to have a higher estimated prevalence for this group in the HSRC survey than in the antenatal survey (Human Sciences Research Council, 2002; Department of Health, South Africa, 2005).

A number of factors influence transmission in South Africa – poverty, lack of education about HIV, condom use, the status of women in some areas of society, a large migrant workforce, traditions relating to male circumcision, as well as the number and age of sexual partners. In a number of African countries, women have been seen to have partners significantly older than they are, whereas the same does not apply to men (Human Sciences Research Council, 2002).

1.4 HIV in children

In sub-Saharan Africa, according to WHO statistics, there were approximately 1.9 million children under the age of 15 years living with HIV, with 2.2 million worldwide, at the end of 2004. In the same year, approximately 450 000 of the 510 000 AIDS-related deaths worldwide occurred in sub-Saharan Africa. Of the 640 000 new infections in this age group (almost 2000 per day), 560 000 were in sub-Saharan Africa. Clearly, this area has the greatest burden of HIV disease (Laurent and Delaporte, 2001; Gayle and Hill, 2001; WHO, 2005a; WHO, 2005b).

According to the ASSA (Actuarial Society of South Africa) model, 69 000 South African infants were estimated to have been infected with HIV at birth in 2002, with a further 20 000 infected due to breastfeeding during that year. The WHO statistics for 2005 provide similar figures – 61 000 – 89 000 infants infected (Dorrington *et al.*, 2002; WHO, 2005b).

The risk of transmission of HIV to infants from their infected mothers is shown in Table 1.1.

Table 1.1. Risk of HIV transmission to infants. Adapted from De Cock *et al.*, 2000.

Timing	**Transmission**
During pregnancy	5-10%
During labour and delivery	10-20%
During breastfeeding	5-20%
Overall without breastfeeding	15-30%
Overall with breastfeeding till 6 months	25-35%
Overall with breastfeeding till 18-24 months	30-45%

By giving zidovudine (AZT) to mothers and infants, the overall risk decreases to 4-8%, while nevirapine reduces transmission by about 47% (to 10-15%). With the use of three antiretroviral drugs, this can be decreased to 1-2% (Nielsen and Bryson, 2000; McIntyre and Gray, 2002; Moodley *et al.*, 2003; Pancharoen and Thisyakorn, 2003; Coovadia, 2004; Lallemant *et al.*, 2004).

HIV infection of infants can be divided into three groups based on the timing of transmission (Peckham and Gibb, 1995) – *in utero*, intrapartum, and post-partum (mostly due to breastfeeding).

Evidence suggests that most *in utero* infections occur during the third trimester, although HIV can be detected in foetal tissue as early as 10 weeks of gestation (Courgnaud *et al.*, 1991; Ehrnst *et al.*, 1991; Katz *et al.*, 1997). In one study, viruses were detected in only 2% of mid-trimester foetal thymuses (Nielsen and Bryson, 2000). It is also believed that early *in utero* infection causes more foetal losses. Infants infected *in utero* tend to have a quicker disease progression when compared to infants infected intrapartum (Diaz *et al.*, 1998; Dickover *et al.*, 1998).

Two patterns of disease progression are seen with infected infants – about 20% become ill soon after birth, and most of these children die by age 4. It is believed that *in utero* infection may play a significant role in this group. The other 80% tend to show progression in a similar way to adults, taking 5-15 years before reaching AIDS (Auger *et al.*, 1988; Scott *et al.*, 1989; Blanche *et al.*, 1990; Domachowske, 1996; Hermione Lyall, 2002).

Several factors have been associated with a faster progression to severe disease, namely a maternal $CD4^+$ count below 400 cells/μl, a maternal viral load above 10^5 copies/mL, low birth

weight, poor weight gain, a high infant viral load, and a low infant $CD4^+$ count. The infant factors are possibly due to earlier infection, while the maternal factors are thought to potentiate earlier infection. Low birth weight is more controversial, and thought to be more likely due to maternal illness than foetal illness. (Italian Multicentre Study, 1988; European Collaborative Study, 1992; Gabiano *et al.*, 1992; Domachowske, 1996; Bailey *et al.*, 1999; Bongertz, 2001).

When tested for virus at birth, only 20-50% of infants infected *in utero* and intrapartum are identified (Dunn *et al.*, 1995; Kuhn *et al.*, 1996; CDC, 1998; Molina *et al.*, 2004). Several studies found a sensitivity of 22-38% for DNA PCR testing during the first 48 hours of life, 73-93% by 14 days, and 96% by 4 weeks (Dunn *et al.*, 1995; Kuhn *et al.*, 1996). Those detected early are likely to be those infected *in utero*, whereas those infected intrapartum will probably only have detectable replicating virus at a later stage.

The survival tends to be higher in countries with a generally higher socioeconomic status, and this is reflected in Table 1.2, which shows the mortality rate found in a number of studies.

Table 1.2. Infant mortality rate in several countries.

Country	**Mortality rate as % over the years studied**	**Reference**
Europe (6 years)	26%	Blanche *et al.*, 1997
USA (5 years)	25-35%	Barnhart *et al.*, 1996; Langston *et al.*, 2001
Thailand (5 years)	49%	Chearskul *et al.*, 2002
Uganda (2 years)	54%	Berhane *et al.*, 1997
Rwanda (5 years)	62%	Spira *et al.*, 1999

The clinical presentation of HIV-related conditions differs in certain aspects from the presentation in adults – mainly relating to specific infant factors such as growth, and the incidence of opportunistic infections.

Infant growth and development may be affected by HIV infection – many infected infants are underweight for age, and do not grow as rapidly as HIV-negative children (Bailey *et al.*, 1999; European Collaborative Study, 2003). Milestones are also often reached at a later stage, and may later manifest as poor performance at school. Similar to HIV-infected adults, amongst whom increased osteoporosis and osteopenia are seen, up to 50% of HIV-infected children have a lower skeletal age than chronological age, with poorer bone quality, than their

HIV-negative counterparts (Rosso *et al.*, 2005). While these effects have been seen to occur on antiretroviral therapy, especially those regimens containing protease inhibitors, this influence appears to be less significant in infected infants than the natural course of HIV infection itself.

Pneumocystis pneumonia (PCP) is the leading cause of death in children with AIDS. Cytomegalovirus is also a significant contributor to the burden of disease. In infants, these are usually primary infections, whereas in adults, the illness is usually due to reactivation of latent infection. Oral candidiasis is common in infants, but more frequent in HIV-infected infants, and more often unresponsive to treatment.

Serious bacterial infections (such as pneumonia, meningitis, cellulitis, septicaemia) are more common in infants than in adults, while toxoplasmosis and cryptococcal infections tend to occur less frequently in infants. Lymphocytic interstitial pneumonitis (LIP) is rarely seen in adults, but is common in HIV-infected children. Mycobacterial infection is problematic in older infants, although it is now being more readily acknowledged in younger infants. The live attenuated vaccine strain Bacillus Calmette-Guerin also poses a real threat to immunocompromised infants (Domachowske, 1996; Saloojee and Violari, 2001).

1.5 Risk factors for spread from mother to child

A large number of clinical factors have been implicated in the transmission from an infected mother to her infant. These include maternal factors, such as viral load and immunological status, late stage disease as well as new infections, and the presence of symptoms, as well as illicit drug use and nutritional status; intrapartum factors such as prematurity, type of delivery, complications, pre-labour rupture of membranes, invasive procedures, concomitant sexually transmitted infections; and post-partum factors, namely breastfeeding, with duration and sole breastfeeding versus mixed feeding being important (Domachowske, 1996; European Collaborative Study, 1999; Fawzi *et al.*, 2001).

Other factors associated with transmission are infant and viral genotypes, the presence of maternal neutralising antibodies, and infant immune response.

Breastfeeding – The risk of transmission of HIV during breastfeeding in the first 6-24 months of life is 5-20%, considerably lower than that for simian immunodeficiency virus, suggested by limited data (Amedee *et al.*, 2004; Jayaraman *et al.*, 2004) available in macaques. HIV transmission can be successfully reduced by elimination of breastfeeding (Coutsoudis *et al.*, 2001b; McIntyre and Gray, 2002). However, this is a controversial

subject, especially in developing countries, where breastfeeding is of great importance for the protection of the infant, via humoral and cellular factors in the breast milk, until the infant has developed his/her own protective immune responses. Denying breastfeeding during this period to lower the risk of HIV infection may put the infant at a greater risk of disease preventable by breastfeeding (Phadke *et al.*, 2003). Where formula feeding is available, as well as access to adequate medical care, this strategy may be of great value. However, in certain populations, exclusive breastfeeding may pose a lesser risk to the infant when the risk of HIV is weighed up against the risk of other childhood diseases such as rotavirus, which has a significant mortality and morbidity in the developing world (Abdool Karim *et al.*, 2002; Coutsoudis *et al.*, 2003; Ehrnst and Zetterstrom, 2003; Kourtis *et al.*, 2003;). Mixed feeding, i.e. combining breastfeeding with formula feeding, is generally accepted to be detrimental due to irritation of the intestinal mucosa by artificial feeds, enhancing susceptibility to HIV infection (Magoni and Giuliano, 2005). It has also been noted that different components of breast milk may carry a different viral load (Hoffman *et al.*, 2003). The effect of breastfeeding on maternal health is controversial, with one study showing a negative effect, and several others showing no effect (Coutsoudis *et al.*, 2001a; Nduati *et al.*, 2001; Breastfeeding and HIV International Transmission Study Group, 2005; Kuhn *et al.*, 2005). The impact of highly active antiretroviral therapy, or HAART, (with suppressed viral replication) on the risk of transmission by breastfeeding has yet to be formally assessed, and ethical considerations would make assessment of the control arm of such studies difficult. Heat-inactivation (Pretoria Pasteurisation) of breast milk has been suggested as a method for prevention of transmission (Jeffery and Mercer, 2000). Other factors that may limit transmission from breast milk include the Lewis X sugar epitope, which binds to dendritic cell-associated ICAM-grabbing non-integrin (DC-SIGN) and interferes with HIV-1 transfer to $CD4^+$ T-lymphocytes (Naarding *et al.*, 2005); and the anti-HIV $CD8^+$ T-lymphocyte component of breast milk (Sabbaj *et al.*, 2002).

Maternal factors – Early and late stage infections are usually associated with high viral loads, which increase the chance of transmission to the infant during all three time periods – *in utero*, intrapartum, and post partum. Similarly, a weaker immune system, as indicated by the $CD4^+$ count, has similar effects. Illicit drug use and lack of nutrients may result in weakening of the immune system, as well as effects on epithelial integrity, and since the foetus derives its nutrition from the mother, these effects would carry over into the infant (Domachowske, 1996). It has also been seen that certain nutrients are lacking in many HIV-

infected populations, and that low serum vitamin A levels may be associated with increased transmission to infants, although further studies showed that supplementation did not reduce transmission, and it is likely that the low vitamin A levels are a symptom of late stage disease, the actual factor in the higher transmission rates seen (Coutsoudis *et al.*, 1999; Fawzi *et al.*, 2002). The effect of micronutrient supplementation on vertical transmission during breastfeeding is still unknown (Dreyfuss and Fawzi, 2002).

Intrapartum factors – Exposure to maternal body fluids containing virus increases the risk of infection for the uninfected infant. This has resulted in attempts to limit exposure to maternal blood by avoidance of invasive procedures such as probes for foetal monitoring, artificial rupture of membranes, and unnecessary assisted delivery. Caesarean section, while invasive, has been shown to have a limited effect on preventing transmission by limiting exposure of the infant to maternal blood and vaginal secretions during the physical trauma to the mother during vaginal delivery (Italian Multicentre Study, 1988; European Collaborative Study, 1992; Gabiano *et al.*, 1992; Landesman *et al.*, 1996; International Perinatal HIV Group, 1999). Exposure to maternal blood due to abruptio placentae, placenta praevia, and transplacental micro-transfusions during vaginal delivery also increases the risk of transmission (Kwiek *et al.*, 2006). Chorioamnionitis increases the risk of transmission of HIV to the infant. Sexually transmitted infections may increase the risk of infection for the infant – ulcerative conditions expose the infant to detrimental maternal secretions, and *Treponema pallidum* may cause placentitis during active syphilis infection (Domachowske, 1996). Malaria, on the other hand, while known to increase HIV viral loads, and with placental malaria significantly associated with HIV infection, and resulting in an increased presence of HIV-1 presenting cells, as well as more placental tissue damage, does not seem to be associated with a higher risk of HIV transmission to the infant, according to one study (Inion *et al.*, 2003), while another implicates it as a factor in transmission (Brahmbhatt *et al.*, 2003). The placenta itself has been implicated in transmission of HIV – the placenta expresses DC-SIGN and DC-SIGN-related molecule (DC-SIGNR), both mediating infection of other cells by HIV. Furthermore, maternal cells associated with the placenta express DC-SIGN (Soilleux and Coleman, 2003).

Viral genotype – HIV-1 and HIV-2 – related viruses, but significantly different genetically – have different rates of transmission from mother to infant. Furthermore, individual quasispecies may have a greater or lesser replication and infectivity efficiency, and therefore it was considered that within the HIV-1 type, different subtypes may also show a difference in

transmission risk. Early studies of the relative risk for transmission of the various HIV-1 group M subtypes indicated that there was no significant difference between transmission of the different subtypes. Subsequently, it was suggested that the subtype of the long terminal repeat (LTR) was associated with degree of transmission to the infant; however, the study did not have access to viral load data, and the LTR subtype could have influenced the viral load, which in turn influenced transmission, rather than the LTR influencing transmission directly (Blackard *et al.*, 2000; Blackard *et al.*, 2001). Other subsequent studies have produced conflicting results (Murray *et al.*, 2001; Renjifo *et al.*, 2001; Tapia *et al.*, 2003). The possibility that mothers infected with certain HIV-1 subtypes may require more aggressive preventative measures requires further study.

Human genetic factors: – Different HLA types are associated with either better or poorer responses, depending on HIV subtype, and it has been shown that concordant maternal and infant class 1 HLA types increases the risk of HIV transmission to the infant. The reason for this is that cytotoxic T-lymphocyte (CTL) escape mutants that are not detected by the mother, are unlikely to be detected by the infant if they share the same HLA type (Pillay and Phillips, 2005). The HIV co-receptor CCR5 has a variety of polymorphisms associated with HIV progression and infection (Doms and Moore, 1997; Cullen, 2001; Koning *et al.*, 2002). The most well known is a 32 nucleotide deletion, known as CCR5-Δ32. In adults as well as infants, homozygosity for the mutant protects against infection, and both CCR5-Δ32/Δ32 homozygosity and CCR5-Δ32/wild type heterozygosity result in a slower progression of disease in those infected. Several single nucleotide polymorphisms (SNPs) in the CCR5 gene's regulatory region, as well as in the CCR2 gene (CCR2 is a minor co-receptor for HIV) and the SDF1 gene (SDF1 is the natural ligand for the HIV co-receptor CXCR4) have also been associated with the rate of disease progression, although much controversy exists (Brumme *et al.*, 2001; Sei *et al.*, 2001; Ioannidis *et al.*, 2003; Mulherin *et al.*, 2003; Singh *et al.*, 2003). Knowledge of these differences has, in the case of CCR5, led to the development of antiretroviral drugs currently in clinical trials (Barber, 2004; Dorr *et al.*, 2005; Reeves and Piefer, 2005).

Infant immune response – HIV-specific cytotoxic T-cell responses have been observed in uninfected, but not infected, HIV-exposed infants, suggesting that unknown factors may enable certain infants to form an immune response capable of preventing infection by HIV (Wasik *et al.*, 1999).

Maternal neutralising antibodies – Neutralising antibodies are one of the factors that influence the evolution of HIV quasispecies within an individual, and some antibodies have, inconsistently, been associated with an altered risk for transmission (Parekh *et al.*, 1991; Scarlatti *et al.*, 1993).

1.6 Molecular epidemiology of HIV

HIV is a diploid single-stranded RNA virus that incorporates a reverse transcriptase enzyme into its virion, which reverse transcribes the RNA, making DNA that gets incorporated into the host genome by the viral enzyme integrase (Thompson *et al.*, 2002). This reverse transcriptase is one of the major causes of the genetic diversity of HIV today, due to it lacking 3'-exonuclease proof reading ability, and a high mismatch error rate, which averages 1 error per 1700 detectable nucleotide incorporation, but which can be as high as 1 per 70 nucleotides in certain areas of the genome (Roberts *et al.*, 1988). Other factors that contribute to the virus' genetic diversity include the rapid turnover of viruses *in vivo*, selective pressure from immune responses and therapeutic interventions, and the diploid genome, which enables easier recombination events (Thompson *et al.*, 2002; Kandathil *et al.*, 2005).

HIV-2, of less importance than HIV-1, is divided into 8 groups – A-H – and likely originated from SIVsm, a form of simian immunodeficiency virus found in sooty mangabeys, in 8 separate cross-species transmission events. Groups A and B are the predominant groups, with groups C-H represented by only a few sequences (Laurent and Delaporte, 2001; Lemey *et al.*, 2003).

HIV-1 likely originated from SIVcpz, the chimpanzee strain of SIV, in the early 1900's. It is divided into 3 groups – M, N, and O – based on phylogenetic analysis. Each group is believed, based on absence of common branching in phylogenetic trees of HIV-1 and SIVcpz, to be the result of a separate transmission event from chimpanzees to humans. Group M is the main (or major) group, group O the outliers, and group N the Non-M, Non-O, or new, group. Group M is responsible for the majority (99.6%) of HIV infections worldwide, and different subtypes have been found in different areas and population groups, and represent distinct phylogenetic lineages. Currently subtypes A, B, C, D, F, G, H, J, and K are recognised, with subtype A divided into subsubtypes A1, A2 and A3, and subtype F divided into subsubtypes F1 and F2. At least 19 known circulating recombinant forms (CRFs) have been identified, such as CRF01_AE, and CRF02_AG, and many further unique recombinant forms (URFs). HIV is believed to have originated in Africa – most of the identified CRFs were originally

described in Africa, while only 5 are believed to have originated elsewhere (Laurent and Delaporte, 2001; Los Alamos National Laboratory HIV Sequence Database, 2002; Casado *et al.*, 2005; Kandathil *et al.*, 2005; Los Alamos National Laboratory HIV Sequence Database, 2005; Meloni *et al.*, 2004;).

All 3 groups of HIV-1 are found in Africa, but groups N and O are, with few exceptional cases, limited to Central Africa. Group M is found across the whole continent, and, indeed, the whole planet. Subtypes A and D are common in East Africa, with subtype C in the horn of Africa, and subtype A in West Africa, while the predominant form in West and Central Africa is CRF02_AG. In Southern Africa, the most common subtype amongst heterosexuals (and therefore the most commonly seen in vertical transmission) is subtype C, while subtype B predominates in the homosexual community, and is also the predominant form in Western and Central Europe, Australia, South East Asia (with CRF01_AE) and the Americas, although evolutionary analysis suggests that subtype C is starting to out-compete with subtype B in Brazil, after having been present for only a few years. Subtype C is the most common subtype in India, and accounts for 47.2% of all HIV infections worldwide (Laurent and Delaporte, 2001; Puren, 2002; Thompson *et al.*, 2002; Weiss, 2003; Kandathil *et al.*, 2005).

A variety of methods are employed to examine viral genotype, ranging from relatively simple assays such as serotyping to infer genotype, to more complex ones, such as heteroduplex mobility assays, to the gold standard – sequencing, with subsequent phylogenetic analysis (Thompson *et al.*, 2002; Kandathil *et al.*, 2005).

Serotyping is based on antibody binding to peptides from the V3 loop, part of the gp120 protein of the viral envelope, of different subtypes. In areas where there are multiple genotypes present, this method is not as accurate, and in particular is not reliable in differentiating between subtypes A and C (Arens, 1999; Engelbrecht *et al.*, 1999; Thompson *et al.*, 2002).

The heteroduplex mobility assay (HMA) utilises the difference in electrophoretic mobility of a duplex formed between an amplified PCR product and a reference strain. The *gag* gene and the gp120 region of the *env* gene are the preferred regions, and the *env* gp41 region is useful for monitoring subtypes where divergent strains of HIV-1 exist. Multiple regions need to be used if there are multiple subtypes circulating, in order to exclude recombination (Bachman *et al.*, 1994; Thompson *et al.*, 2002).

The gold standard is sequencing of the genome, either partially or fully, with subsequent phylogenetic analysis involving complex mathematical calculations that can take significant amounts of time, even on modern computers. Not only can genotypes be identified, and phylogenetic trees drawn to indicate relationships, but evolutionary rates predicted from the observed diversity in a database of viral sequences, which make modeling of epidemics from a genetic perspective possible. The sequence itself can be utilised for a variety of purposes, ranging from prediction of resistance to antiretroviral drugs, to identification of variants that need to be taken into account in diagnostic assays (Thompson *et al.*, 2002; Kandathil *et al.*, 2005).

1.7 Implications of HIV molecular epidemiology

Knowledge of the genotypes of HIV strains in the region, and in the world, is required in order to understand the epidemic, and to eventually come to a solution for its management. This knowledge is not limited to the subtype itself, within regions, and even within individuals, specific populations may arise that differ from those outside the region, or individual. This understanding impacts on a number of important areas in HIV research, including understanding the biological properties of the virus, such as transmission and pathogenesis, design of sensitive diagnostic assays, predicting influence of genotype on drug resistance patters, and eventual vaccine design.

Transmission – It has been postulated that different HIV subtypes may have different transmission potential, and while phylogenetic analyses may indicate a different epidemic potential, such as seen in Brazil where HIV-1 subtype C is spreading twice as fast as subtype B in Brazil or subtype C in South Africa, no definitive answer has been found regarding genetic influence on transmissibility of virus subtypes (Salemi *et al.*, 2005). There is considerable evidence, however, that viruses containing the subtype C long terminal repeat (LTR) is more likely to transmit than those containing other LTRs. This has also been observed in the infant population. However, other factors may influence transmission, such as a higher viral load, which is also influenced by subtype C LTRs. In an individual, however, different quasispecies exist, some of which may be in some way defective, and less likely to be transmitted. This has been seen in infants and adults where specific quasispecies, often minor quasispecies, have been seen to transmit. Also of significance is the transmission potential of viruses using different co-receptors. Viruses using the CCR5 co-receptor tend to be non-syncytium inducing (NSI) strains, whereas those using the CXCR4 co-receptor tend to

be syncytium inducing (SI) strains (Fenyo *et al.*, 1997). The former are usually seen in earlier infections, while the latter tend to appear as the patient progresses towards AIDS. Subtype D does not appear to have dual-tropic viruses that utilise both co-receptors, while subtype C tends to have a lower frequency of progression to the SI phenotype (Tscherning *et al.*, 1998; Thompson *et al.*, 2002). There has been speculation that, at least for the Indian population infected with subtype C, this is due to a replication advantage caused by a higher natural level of CCR5 expression (Ramalingam *et al.*, 2002). Studies of viral co-receptor usage in combination with human genetic data, as seen in the discussion on risk factors for the transmission from mother to child, have been enlightening.

Pathogenesis – Different subtypes may be associated with different disease progression patterns. As with HIV-2 having a slower progression than HIV-1, different HIV-1 subtypes may cause more rapid or slower progression than others. For instance, subtype A has been linked to slower progression than subtypes C and D (Kanki *et al.*, 1999). Other studies, however, have found no correlation between subtype and disease progression (Alaeus *et al.*, 1999; Amornkul *et al.*, 1999).

Diagnostics – For both serological and molecular diagnostic assays, as well as monitoring assays such as viral loads, it is important to maintain sensitivity, and be able to detect the prevalent HIV subtypes, if not all types and subtypes as in the case of serology, adequately (Nielsen and Bryson, 2000). Different HIV subtypes may be associated with different seroconversion profiles, as seen in subtype B vs. CRF01_AE, where the latter had a mean window period for a sensitive/less sensitive assay, designed to detect recent HIV infection, of 270 days verses the mean window period of subtype B, which was 155 days (Parekh *et al.*, 2001). (The window period here means the time between seroconversion on only one assay to seroconversion on both).

Drug resistance – Since HIV-1 group O and HIV-2 are inherently resistant to non-nucleoside reverse transcriptase inhibitors (NNRTIs), it was suggested that HIV-1 group M subtypes may display differences in inherent sensitivity to antiretroviral drugs (Kandathil *et al.*, 2005). However, most studies have shown similar responses of all group M subtypes to these drugs, with the possible exception of some subtype G strains, which may be less sensitive to protease inhibitors (PIs) (Descamps *et al.*, 1998). Naturally occurring secondary PI resistance mutations have been found to be more prevalent in subtype C than in subtype B, but the clinical significance of this is not yet known (Pieniazek *et al.*, 2000). Until now, studies have indicated that, on a molecular level, subtype B and non-B subtypes studied have no significant

differences in primary mutations, and, on a clinical level, outcome of treatment does not differ (Cane *et al.*, 2001; Perez-Alvarez *et al.*, 2001). More research is needed to ascertain the significance of the different prevalences amongst different subtypes seen for secondary resistance mutations and other polymorphisms seen in the *pol* gene (Kandathil *et al.*, 2005).

Vaccine design – Serotypes defined by neutralising antibody do not appear to correlate with HIV-1genetic subtypes, but subtype-specific neutralising antibodies have been reported, and it has been seen that sensitivity to neutralising antibody is more frequent within a subtype than between subtypes. Broadly neutralising antibodies have only been seen in long-term non-progressors. Correlation of neutralising antibody response with viral sequences, as well as with human genetic factors, remains unclear. Cytotoxic T-lymphocyte (CTL) responses, similarly to neutralising antibody responses, appear to be stronger and more frequent within a subtype than between subtypes. Phylogenetic data has been used to identify sequences that are perhaps more likely to produce a cross-subtype, or even pan-subtype, immune response, as well as investigate ancestral sequences as a possibility for use in a vaccine against multiple subtypes (Wagner *et al.*, 1999; Gordon *et al.*, 2003).

1.8 Molecular epidemiology of HIV in children

One of the aspects of HIV that remains poorly understood, despite more than 2 decades since the discovery of the virus, is the dynamics of mother-to-child transmission. We know little about the effects of pregnancy on the interaction between the virus and the maternal immune system, and the complex interrelationships between the immune system of the infant, the immune system of the mother, the placenta, and the virus itself. Although the cytotoxic T-lymphocyte (CTL) response in pregnant women does not differ from that of non-pregnant women, the CTL response in the placenta has not been well described (Wasik *et al.*, 1999; Pillay and Phillips, 2005). There is still much speculation regarding the transmission potential of different subtypes, the degree to which external factors, such as nutrients, affect transmission, and the factors that cause some infants to become infected, and others not. To some extent, the combined research on different subtypes, the degree of divergence within an individual at various time points after infection, the nature of the transmitted virus, and the variety of immune responses to the virus, has formed a framework for future research on mother-to-child transmission of HIV.

Transmission has been seen to take place in a variety of different forms. Obviously, a replication-competent virus is required in order to infect the infant. However, not all viruses

transmitted are equally efficient (Ahmad *et al.*, 1995; Yedavalli *et al.*, 1998a; Yedavalli *et al.*, 1998b; Blackard *et al.*, 2000; Blackard *et al.*, 2001; Ramakrishnan *et al.*, 2005).

Hypermutation, meaning that there is an excess of purine transitions taking place during viral replication, has been seen shortly after birth in HIV transmitted to infants. G to A transitions in the context of GA or GG dinucleotides are the most common form of hypermutation in lentiviruses, and result in an increased number of stop codons throughout the viral genome. This has been associated with decreased reverse transcription and integration efficiency. The implication is that the virus would be less pathogenic, and this has been supported by associations with slower progression in patients with such viruses. The cause is uncertain, but may reflect a cellular mechanism to protect against viral infections (Koulinska *et al.*, 2003).

In general, however, it appears that the main maternal quasispecies to be transmitted is replication competent. There is considerable data to show that viruses that transmit do not always represent the main quasispecies in circulation in the mother (Ahmad *et al.*, 1995; Yedavalli *et al.*, 1998a; Yedavalli *et al.*, 1998b; Ramakrishnan *et al.*, 2005). While cases of transmitted CXCR4-using viruses have been observed, cases where CCR5-using viruses were transmitted from mothers whose predominant virus utilised CXCR4 as a co-receptor – later, as the infants progressed to AIDS, virus strains using CXCR4 evolved from the original CCR5-using strain (Clevestig *et al.*, 2005). Cases of single quasispecies transmission and multiple quasispecies transmission, with both major and minor maternal variants can occur both *in utero* and intrapartum, although single quasispecies transmission is more likely (Ahmad *et al.*, 1995; Dickover *et al.*, 2001). Transmission can also occur by free virus in the plasma, or by cell-associated proviral DNA (Dickover *et al.*, 2001). There does not seem to be a relationship to either subtype or viral load with regard to the risk of single or multiple quasispecies being transmitted, and data suggests that transmission is likely to be a single transmission event, rather than multiple events. However, transmission of multiple subtypes to infants has been observed, making it likely that, at times, multiple transmission events can and do occur (Renjifo *et al.*, 2003).

Whether the infant is infected with single or multiple quasispecies, their population of virus appears to be less diverse than the mother's virus. Selection of quasispecies transmitted does therefore seem to occur – however, the factors influencing selection are not well understood, nor is the significance of different selection pressures on the nature of the transmitted virus transmitted *in utero*, intrapartum, or postpartum. The interpretation of phylogenetic information is further confounded by an uncertainty about when the transmission event

occurred – multiple quasispecies in a newborn may indicate recent transmission of multiple quasispecies, or it may indicate a single quasispecies transmitted *in utero*, with enough time for divergence within the infant (Ahmad *et al.*, 1995; Yedavalli *et al.*, 1998a; Yedavalli *et al.*, 1998b; Ahmad, 2005; Ramakrishnan *et al.*, 2005).

Although there is conflicting data, it appears that the greater the diversity of viral quasispecies in the mother, the greater the chance of transmission to her infant. In contrast to adults, therefore, where the primary infection is usually due to one quasispecies, and usually in virion form, infants can be infected with multiple quasispecies, in either virion or proviral form (Ahmad, 2005).

Another under-studied aspect of the virus transmitted is the influence of maternal and infant neutralising antibody and CTL response. While maternal antibody and CTL escape mutants have been detected in infants, and phylogenetic data indicate that they were transmitted to the infant from the mother, the influence of the infant's own immune response in selecting for certain quasispecies is not known. However, the fact that mothers and infants share a similar genetic background to their immune systems, immune selection in the mother is likely to be detrimental to the infant in ways not expected in sexual transmission, and perhaps accounts for the more severe disease and more rapid disease progression often seen in infants (Wasik *et al.*, 1999; Pillay and Phillips, 2005).

Few studies related to the comparison between mother and infant virus have been done, and most contain only a few mother-infant pairs (Ahmad *et al.*, 1995; Yedavalli *et al.*, 1998a; Yedavalli *et al.*, 1998b; Biggar *et al.*, 2001; Koulinska *et al.*, 2003; Ramakrishnan *et al.*, 2005). Studies identifying hypermutation and the types and numbers of quasispecies transmitted have contributed significantly to our limited knowledge of vertical HIV transmission. Several large studies on the diversity of vertically transmitted HIV have provided a significant amount of information. Twin studies are extremely interesting, as the infants share a common intrauterine environment, common exposure to maternal factors, and a significant degree of genetic similarity (Hutto *et al.*, 1996; Biggar *et al.*, 2002; Biggar *et al.*, 2003). There has even been a study on an alleged event of child-to-mother transmission, where the study suggests that the infants were infected nosocomially, and then one transmitted HIV to the mother, possibly via breastfeeding (Bobkov *et al.*, 1994).

1.9 This study's aim

The purpose of this study was to contribute to the knowledge of vertical HIV transmission.

The primary aim was to perform genetic characterisation and phylogenetic analysis of the HIV-1 viral genome in positive serum/plasma samples obtained from infants and their mothers during the years 2000, 2001, and 2002.

This was achieved by:

- Genotyping and characterisation of the *env* gene of maternal and infant HIV-1
- Comparison between the genotype and genomic sequence of the mother's virus on the one hand, and the infant's virus on the other
- Comparison of this study's findings with other molecular epidemiology studies and known data in paediatric patients, along with clinical analysis of patient file data (from mothers and infants), in order to assist in the judgment of any findings. Although numbers were too low to be statistically significant, trends were be observable, and a clinical description of the patient sample was be obtained.

Chapter 2: Materials and Methods

2.1 Ethical approval

The study protocol was submitted to the Committee for Human Research of the Faculty of Health Sciences, University of Stellenbosch. The letter approving the study is attached as Appendix A.

2.2 Patient samples

The infant samples forming the preliminary group for further investigation were those for which the HIV PCR was requested during 2000, 2001, or 2002, and the PCR result was positive; the sample had been taken within 6 months of birth, in order to allow minimal divergence from the infecting virus; and sufficient plasma and/or RNA for further investigating remained in storage.

The maternal serum samples were included if the mother was also HIV positive, the sample was taken within 6 months of the infant's specimen's date, in order to allow minimal divergence from the time of transmission to the infant, and sufficient serum for use remained in storage.

All diagnostic HIV PCR results for the period 2000-2002 were obtained from Disa*Lab 204.16.00 (LabSystec, Johannesburg, South Africa), the NHLS database used to store all diagnostic results. The plasma and RNA extracted from the original samples for diagnostic purposes were obtained from storage at –80 °C. Clinical files of these infants were obtained from Tygerberg Hospital, and case notes and laboratory results examined. Mothers were identified by their name and/or hospital number being listed in the clinical file or Disa*Lab's clinical details. For those mothers who had been tested at the department of Medical Virology, their serum specimens' reference numbers were identified using Disa*Lab, and located in the laboratory. Infants and mothers were then assigned numbers in pairs, B01-B27 and M01-M27 respectively, for further identification in the laboratory.

2.3 Clinical information

Clinical information was obtained from the hospital files and from Disa*Lab. The following conditions were specifically noted – parental details; antenatal records; information on the delivery – type of delivery, duration of labour, rupture of membranes, APGAR scores, complications, and the use of antiretroviral drugs for attempted prevention of HIV

transmission; clinical diagnoses and symptoms during hospitalisation or outpatient visits; clinical and laboratory indicators of immune status; treatment given, including antiretroviral drugs; and date of death, if deceased.

2.4 RNA extraction

The QIAamp Viral RNA Mini Kit (QIAGEN, Hilden, Germany) was used for RNA extraction from plasma samples, according the manufacturer's protocol.

All infant plasma specimens had previously undergone RNA extraction by Ms Brenda Robson, one of the NHLS technologists employed in the molecular diagnostic section. However, not all had RNA remaining, and in those cases, RNA was extracted again from remaining plasma. All maternal serum specimens required RNA extraction.

2.5 Primers

Initially the degenerate *env* immunodominant region primers JH38 and JH41 were used for the prenested reverse-transcription polymerase chain reaction (RT-PCR), followed by env 27F and Menv 19R for a nested polymerase chain reaction (PCR) (Swanson *et al.*, 2003).

However, these proved to be problematic, and a smaller region around the V3 loop was chosen, and a well-established PCR used regularly in the department was used. To screen samples for RNA quality, the established in-house diagnostic HIV *gag* RT-PCR use in the department was used with primers GAG A and GAG B for the prenested PCR, and GAG C and GAG D for the nested PCR (Kemp *et al.*, 1989; Engelbrecht and van Rensburg, 1995). For the *env* fragment chosen for analysis, the primers used were ES7x and ES8x for the prenested PCR, and ES7x and ES125 for the heminested PCR (Bachman *et al.*, 1994; Sanders-Buell *et al.*, 1995; Moodley *et al.*, 1998). Table 2.1 shows the primers, their direction on the genome, and their sequences.

Table 2.1. Primers used in this study, their direction, and sequence.

Primer name	Direction	Sequence
JH38	Forward	GGTGARTATCCCTKCCTAAC
JH41	Reverse	CAGCAGGWAGCACKATGGG
env 27F	Forward	CTGGYATAGTGCARCARCA
Menv 19R	Reverse	AARCCTCCTACTATCATTATRA
GAG A	Forward	AGAGAACCAAGGGGAAGTGA
GAG B	Reverse	TCTCTAAAGGGTTCCTTTGG
GAG C	Forward	CATAGCAGGAACTACTAGTA
GAG D	Reverse	TCCTTGTCTTATGTCCAGAA
ES7x	Forward	CTGTTAAATGGCAGTCTAGC
ES8x	Reverse	CACTTCTCCAATTGTCCCTCA
ES125	Reverse	CAATTTCTGGGTCCCCTCCTGAG

2.6 PCR kits and protocols

The one-step RT-PCRs were performed using the Promega Access RT-PCR System (Promega, Madison, WI, USA), according to the manufacturer's protocol and the protocol used for the established in-house diagnostic PCRs.

Nested and heminested PCRs were performed using Promega *Taq* DNA Polymerase in Buffer A (Promega, Madison, WI, USA), according to the protocol used for the established in-house diagnostic PCRs.

The initial PCR methods used with primers JH38, JH41, env 27F and Menv 19R consisted of a 50 μl one-step RT-PCR followed by a 50 μl nested PCR.

The 50 μl prenested PCR reaction mixture consisted of 5 μl of extracted RNA added to a mastermix consisting of 10 μl 5X buffer, dNTPs at a concentration of 200 μM for each dNTP, a concentration of 1000 nM for each primer, 1.0 mM $MgSO_4$, and 1 μl each of Avian myeloblastosis virus reverse transcriptase (RT) and *Thermus flavius* DNA polymerase, with nuclease-free water to bring the final mastermix volume to 45 μl.

Cycling temperatures for the RT-PCR were as follows: 45 minutes at 48 °C for reverse transcription, and 2 minutes at 94 °C for RT inactivation and DNA denaturation, was followed by 40 cycles of 30 seconds at 94 °C for DNA denaturation, 30 seconds at 45 °C for

primer annealing, and 60 seconds at 72 °C for primer extension. A final 10 minutes at 72 °C allowed final extension, and products were stored at 4 °C until collected.

The 50 μl nested PCR reaction mixture consisted of 2 μl of prenested PCR product added to a mastermix consisting of 10 μl 10X buffer, dNTPs at a concentration of 400 μM for each dNTP, a concentration of 800 nM for each primer, 3.0 mM $MgSO_4$, and 0.3 μl of *Thermus aquaticus* DNA polymerase, with nuclease-free water to bring the final mastermix volume to 48 μl.

Cycling temperatures for the RT-PCR were as follows: 2 minutes at 94 °C for DNA denaturation was followed by 40 cycles of 30 seconds at 94 °C for DNA denaturation, 30 seconds at 50 °C for primer annealing, and 60 seconds at 72 °C for primer extension. A final 10 minutes at 72 °C allowed final extension, and products were stored at 4 °C until collected.

The subsequent PCR methods using primers GAG A, GAG B, GAG C, and GAG D for the *gag* PCR, and primers ES7X, ES8X, and ES125 for the *env* PCR, consisted of a 50 μl one-step prenested RT-PCR followed by a 100 μl nested PCR.

The 50 μl prenested PCR reaction mixture consisted of 5 μl of extracted RNA added to a mastermix consisting of 10 μl 5X buffer, dNTPs at a concentration of 200 μM for each dNTP, a concentration of 800 nM for each primer, 1.0 mM $MgSO_4$, and 1 μl each of Avian myeloblastosis virus reverse transcriptase (RT) and *Thermus flavius* DNA polymerase, with nuclease-free water to bring the final mastermix volume to 45 μl.

Cycling temperatures for the RT-PCR were as follows: 45 minutes at 48 °C for reverse transcription, and 2 minutes at 94 °C for RT inactivation and DNA denaturation, was followed by 40 cycles of 60 seconds at 94 °C for DNA denaturation, 60 seconds at 40 °C (for *gag* PCR) or 45 °C (for *env* PCR) for primer annealing, and 90 seconds at 68 °C for primer extension. A final 7 minutes at 68 °C allowed final extension, and products were stored at 4 °C until collected.

The 100 μl nested PCR reaction mixture consisted of 2 μl of prenested PCR product added to a mastermix consisting of 10 μl 10X buffer, dNTPs at a concentration of 800 μM for each dNTP, a concentration of 400 nM for each primer, 1.5 mM $MgSO_4$, and 0.3 μl of *Thermus aquaticus* DNA polymerase, with nuclease-free water to bring the final mastermix volume to 98 μl.

Cycling temperatures for the RT-PCR were as follows: 60 seconds at 94 °C for DNA denaturation was followed by 40 cycles of 60 seconds at 94 °C for DNA denaturation, 60 seconds at 40 °C (for *gag* PCR) or 45 °C (for *env* PCR) for primer annealing, and 90 seconds at 72 °C for primer extension. A final 7 minutes at 72 °C allowed final extension, and products were stored at 4 °C until collected.

Products were identified by agarose gel electrophoresis on a 1% agarose gel containing ethydium bromide, visualised under ultraviolet light. 10 µl of nested PCR product was added to 2 µl loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, 15% Ficoll) and loaded into a slot on the agarose gel submerged in TAE buffer (0.04 M Tris acetate, 0.001 M EDTA), and electrophoresed at a constant current of 5 mA/cm of gel width, resulting in a voltage of approximately 4-5 V/cm. The Promega 1 kb DNA ladder (Promega, Madison, WI, USA) was used as a size marker.

The expected band sizes, based on the reference strain B.FR.83.HXB2 (Korber *et al.*, 1998) were 160 base pairs for the *gag* PCR and 337 base pairs for the *env* PCR.

2.7 Sequencing

Sequencing was performed by Mrs. Annette Laten, a scientist employed in the department of Medical Virology.

Both *gag* and *env* products were sequenced using their respective nested PCR primers, and both strands were sequenced for comparison, to minimise ambiguities.

PCR products were prepared using the PCR Product Pre-sequencing Kit (USB Corporation, Cleveland, USA, according to manufacturer's protocol), which utilises two enzymes, namely exonuclease 1, which removes single stranded DNA (primers and single stranded PCR products), and shrimp alkaline phosphatase, which removes remaining dNTPs from the PCR mixture. Both enzymes are heat inactivated prior to further processing.

The ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit v3.1 (Applied Biosystems, Foster City, USA) was used for sequencing reaction, based on the chain terminator principle, using the manufacturer's protocol, and an annealing temperature of 50 °C.

Purification to remove unincorporated dye terminators was done using the DyeEx 2.0 Spin Kit (QIAGEN, Hilden, Germany) according to the manufacturer's protocol. The products were reconstituted in 25 µl Template Suppression Reagent (Applied Biosystems, Foster City,

USA) and the ABI 310 Genetic Analyzer (Applied Biosystems, Foster City, USA) was used to determine the sequence based on fluorescence of different sized DNA strands moving at different rates in the POP-6 polymer used for electrophoresis.

The data was analysed and converted to electropherograms using the DNA Sequencing Analysis software, version 3.3 (Applied Biosystems, Foster City, USA).

2.8 Reference sequences

Reference HIV sequences were obtained from the Los Alamos National Laboratory HIV Sequence Database, using the HIV reference strain HXB2CG coordinates to obtain the sections of the *gag* and *env* genes required (Korber *et al.*, 1998). The web page at http://www.hiv.lanl.gov/content/hiv-db/SUBTYPE_REF/align.html was accessed, and the 2001 reference set was requested. All reference sequences were obtained, but the following were used in the analyses that followed: All HIV group M reference sequences, in order to show our sequences relative to the rest of group M, and one HIV group O sequence for comparison. The reference sequences used, with their Genbank accession numbers can be seen in Appendix B.

2.9 Sequence analysis

Sequence analyses were performed on the sequences obtained using the software listed in Table 2.2.

Table 2.2. Software used to perform sequence analysis, with source/reference, and purpose.

Software	Source / reference	Used for	Reason for use
FinchTV 1.3.1	Geospiza, Seattle, WA, USA www.finchtv.com	Editing of electropherograms obtained from the ABI 310 Genetic Analyzer.	Identification of readable sequences. Reverse and forward sequences were compared to solve ambiguities.
PrimAlign	Los Alamos National Laboratory, CA, USA http://www.hiv.lanl.gov/content/hiv-db/PRIMALIGN/PRIME.html	Alignment of primers.	To find their coordinate values in the reference HIV sequence HXB2CG.
DNAMan 4.0	Lynnon Biosoft, Quebec, Canada	Alignment of primers and initial sequence alignments.	To obtain the primer alignments in a visual form, and to experiment with simple alignments.
CLUSTAL X 1.81	Thompson *et al.*, 1997	Creating initial sequence alignments. Later used to convert FASTA format files to NEXUS files for use in PAUP*.	Alignments needed for further phylogenetic analysis. NEXUS files required for phylogenetic analysis in PAUP*.

HIV BLAST Search	Los Alamos National Laboratory, CA, USA http://www.hiv.lanl.gov/content/hiv-db/BASIC_BLAST/basic_blast.html	Performing BLAST searches on the sequences obtained.	To identify the sequences as HIV sequences, and see related sequences for the amplified region.
BioEdit 7.0.4.1	Hall, 1999	Performing manual codon alignments on the alignments produced by CLUSTAL X; to create amino acid sequences.	Codon alignments would represent a more natural comparison, and improve phylogenetic analysis.
TREECON 1.3b	Van de Peer and Wachter, 1993; Van de Peer and Wachter, 1994	Performing phylogenetic analyses using the Kimura 2-parameter model for distance calculation, and construction of bootstrapped neighbour joining trees.	Initial phylogenetic analysis.
FindModel	Los Alamos National Laboratory, CA, USA, http://www.hiv.lanl.gov/content/hiv-db/findmodel/findmodel.html, based on ModelTest (Posada and Crandall, 1998; Posada and Buckley, 2004)	Obtaining information on which model of phylogenetic analysis to use in constructing more detailed phylogenetic trees.	To improve phylogenetic analysis.

PAUP* 4.0b10	Swofford, 1993; Swofford, 2003	Performing phylogenetic analysis using the maximum likelihood method.	Obtaining optimised phylogenetic analysis. The PAUP* script used for obtaining PAUP*'s suggested parameters is listed in Appendix D. The PAUP* script used for obtaining FindModel's suggested parameters is listed in Appendix E.
PHYLIP 3.6 (PHYLogeny Inference Package) - Rooted tree plotting program version 3.63	Fenselstein, 2005	To plot the phylogenetic trees produced by PAUP*.	

Chapter 3: Results

3.1 Patient samples

A total of 397 diagnostic RT-PCRs were performed during the period 2000-2002. Of these, 184 were positive on either *gag* or *env* (or both) PCRs. Of those, only 83 fulfilled the criteria of having been taken within 6 months of birth, and having sufficient specimen remaining for either use of stored RNA or RNA extraction from stored plasma.

Clinical files were requested from Tygerberg Hospital for these 83 specimens; however, only 47 files could be found. Twenty seven infants had sufficient information to identify the mother and locate her stored serum specimens in the department. Other infants with identifiable mothers did not have specimen available in the laboratory. The selected infants were assigned numbers from B01 to B27 (for baby), and their mothers were assigned numbers from M01 to M27 (for mother), with B01 being the infant of M01, B02 the infant of M02, and so forth.

3.2 Clinical information

The clinical data obtained from the hospital files was of a limited nature. Information was often not available – for example, information on the delivery of the infant was not always recorded. The only alternative source available for cross-referencing was the laboratory database. Therefore rates and ratios could not be calculated, and only a qualitative description could be obtained.

Furthermore, there were a number of obvious errors in the clinical notes, and the accuracy of the notes is therefore questioned. The three most notable examples are:

a) notes record one infant, at age 3.5 months, as being hypoxic since birth (likely to mean that the infant suffered from hypoxia at birth), and

b) notes for another infant stating that both parents of the infant had died 2 years previously. This infant was 4 months of age at the time of these notes, which means the death of the parents did not fit in with the infant's age and a normal gestation period of 9 months.

c) One infant's file has two women, both designated as the infant's mother, with different first names and surnames signing consent forms for a variety of procedures,

both with different handwriting. Insufficient data was available to determine if one was the grandmother, and which was the biological mother.

Only 26 of the 47 files read contained a detailed obstetric history, with some mention of risk factors for HIV transmission. It is unclear whether such a history was unavailable or simply not taken in the other cases. In total, there were 26 girls, and 21 boys.

Fourteen instances of normal vaginal delivery were recorded, and 2 of breech delivery. Five files mentioned a Caesarean section, none of which was elective, but rather due to complications such as foetal distress (4 infants), halting of progression in labour, and pre-eclampsia. Five infants were born before 34 weeks gestation, one of which was complicated by pre-eclampsia and prelabour rupture of membranes. In total, 3 deliveries were complicated by prelabour rupture of membranes. Only 5 files recorded mothers receiving nevirapine in labour as prophylaxis for HIV transmission, but the prevention of mother-to-child transmission programme in the Western Cape was only introduced in 2002.

The files record that at least 7 infants were eventually placed on antiretroviral therapy – 6 of these survive today on HAART, and 1 received AZT monotherapy in the time before antiretroviral drugs were easily available to those who could not afford them. This infant unfortunately died. Eleven infants were recorded as having been breastfed, and 4 as having been given mixed breast and formula feeding. One of the infants on sole breastfeeding was breastfed while the mother was suffering from mastitis.

Eighteen of the 47 infants were recorded as having died below the age of 6 months. No other deaths were recorded. However, in the Disa*Lab database, some can be observed to have $CD4^+$ counts that drop, and infections that increase, until they no longer get further laboratory investigations. Some of these infants probably died.

The files record 6 cases of developmental abnormalities. Three cases were of known etiology – two with intrapartum hypoxia, and one with foetal alcohol syndrome. One of these infants was delivered by Caesarean section for foetal distress, and the mother suffered from pre-eclampsia. This was the infant that was recorded, at age 3.5 months, as being hypoxic since birth. This infant (included in the study as B24) was the first of two twins, and the twin was healthy, and HIV negative.

Although only 47 files were obtained, the total number of recorded clinical conditions indicative of a compromised immune system, as well as the total number of admissions,

exceeds this. Several infants had several diagnoses of infectious conditions during a single admission.

A number of HIV-related conditions were frequently recorded. Files record that 28 of the infants were classified as marasmic on admission, 21 of whom had gastroenteritis. Several infants had been underfed. Respiratory conditions were common. There were 39 recorded cases of acute bacterial pneumonia, 13 cases of *Pneumocystis carinii* (now *Pneumocystis jirovecii*) pneumonia, 13 cases of pulmonary tuberculosis, and 1 case of BCGosis, disseminated infection caused by Bacillis Calmette-Guerin used in the tuberculosis vaccine. Only one case of otitis media was mentioned – it was a chronic infection. In the cases of pulmonary tuberculosis, most patients needed 5 or more specimens to obtain a positive culture or Zeihl-Neelson stain. Eight cases of meningitis were recorded, one of which was caused by *Mycobacterium tuberculosis*. There were 10 recorded cases of documented septicaemia, and 6 cases where multiple bacteria of clinical significance were isolated. Anaemia was recorded for twenty-nine infants, with haemoglobin levels between 7-9 g/dl.

3.3 PCR results

Table 3.1 shows a summary of the results of the *gag* and *env* PCRs, readable sequences, as well as which sequences were eventually acceptable for inclusion for phylogenetic analysis (discussed in section 3.4).

Using the initial set of primers (Swanson *et al.*, 2003), JH38 and JH41 for the prenested one-step RT-PCR, and env 27F and Menv 19R for the nested PCR, few of the samples could be amplified. Figure 3.1 shows an example of the agarose gel electrophoresis results obtained. The nested products of two PCRs were run together on one gel, with the first PCR on the left of the black bar, and the second on the right. As can be seen, only samples M23, M25, and B10 showed amplification in the first of the two PCRs. Figure 3.2 shows an example of one the checkerboard optimisation attempts for these primers. Different primer and $MgSO_4$ concentrations were used. The most successful appeared to be a $MgSO_4$ concentration of 2.0 mM with a primer concentration of 600 nM, and a $MgSO_4$ concentration of 3.5 mM with a primer concentration of either 200 nM or 400 nM.

Table 3.1 – PCR results, readable sequences, and sequence inclusion for analysis. "POS" in capital letters indicates a positive result, "pos" in small letters indicates a positive result, but with a weak band on agarose gel electrophoresis, and "NEG" indicates a negative result. Asterisks next to "NO" indicate sequences that were not read due to multiple peaks possibly indicative of quasispecies.

	Infants (B01-B27)		Mothers (M01-M27)		Readable sequence		Included in sequence analysis	
Pair	***gag***	***env***	***gag***	***env***	**Infant**	**Mother**	**Infant**	**Mother**
01	POS	pos	POS	NEG	YES	NO	YES	NO
02	POS	POS	POS	pos	YES	YES	NO	YES
03	POS	POS	POS	POS	YES	YES	YES	NO
04	POS	POS	POS	POS	NO*	YES	NO	YES
05	POS	POS	POS	NEG	YES	NO	YES	NO
06	POS	POS	POS	NEG	NO*	NO	NO	NO
07	POS	POS	POS	POS	YES	YES	YES	YES
08	POS	POS	POS	POS	YES	YES	YES	YES
09	POS	POS	POS	NEG	NO*	NO	NO	NO
10	POS	POS	POS	POS	YES	YES	YES	YES
11	POS	POS	POS	POS	YES	YES	YES	YES
12	POS	NEG	POS	NEG	NO	NO	NO	NO
13	POS	POS	POS	POS	YES	YES	YES	NO
14	POS	POS	POS	POS	YES	YES	YES	NO
15	POS	pos	POS	NEG	YES	NO	YES	NO
16	POS	POS	POS	POS	YES	YES	YES	NO
17	POS	pos	POS	NEG	NO*	NO	NO	NO
18	POS	POS	POS	POS	YES	YES	YES	YES
19	POS	POS	POS	POS	YES	YES	YES	YES
20	POS	POS	POS	POS	YES	YES	YES	YES
21	POS	POS	POS	POS	YES	YES	YES	YES
22	POS	NEG	POS	NEG	NO	NO	NO	NO
23	POS	POS	POS	POS	YES	YES	YES	NO
24	POS	POS	NEG	POS	YES	YES	YES	NO
25	POS	POS	POS	POS	YES	YES	YES	NO
26	POS	POS	POS	POS	YES	NO*	YES	NO
27	POS	POS	POS	pos	YES	NO	YES	NO

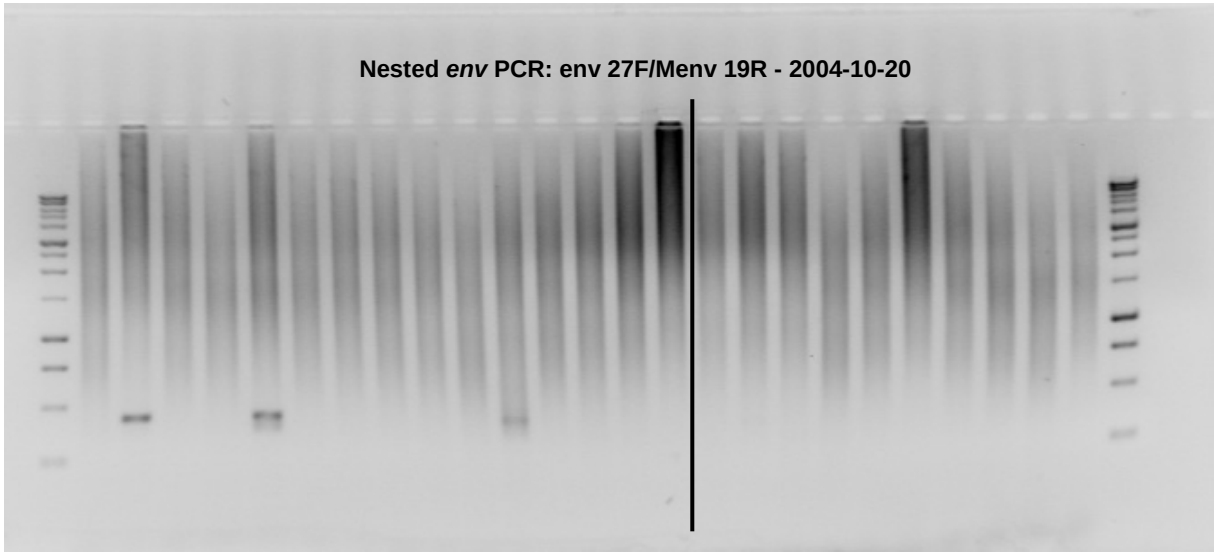


Figure 3.1. 1% agarose gel using nested PCR products. Nested PCR was performed on the prenested products of 2 RT-PCRs. Primers: env 27F and Menv 19R. From left to right: 1kb DNA marker; M18, M23, M12, M14, M25, M24, M13, M15, M26, M16, B10, Pos control 1, Pos control 2, Neg control, prenested PCR's reagent blank; M20a, M20b, M21a, M21b, M19, M22, M27, M17, Pos control 1, nested PCR's reagent blank, 1kb DNA marker.

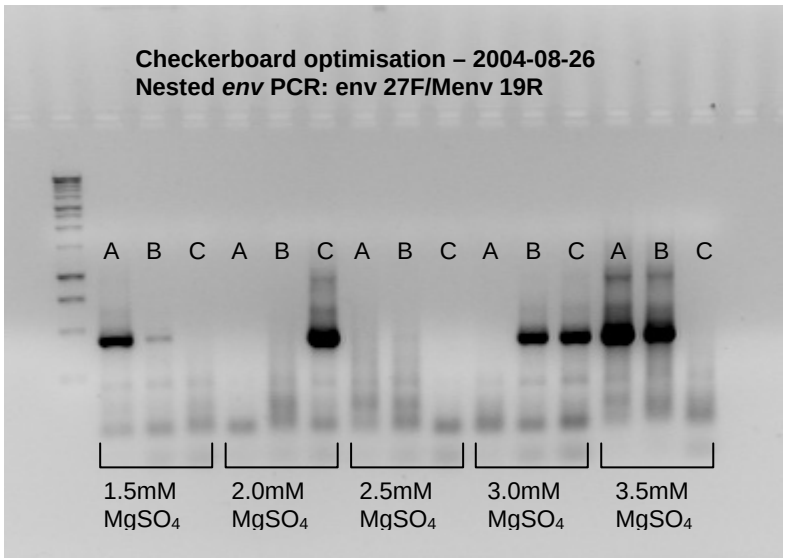


Figure 3.2. Checkerboard optimisation - 1% agarose gel using nested PCR products. Primers: env 27F and Menv 19R. After the 1kb DNA marker, magnesium concentrations are shown in groups of 3, and primer concentrations shown as A, B, and C. A = 200 nM of each primer; B = 400 nM; C = 600 nM. DNA input was uniform as it was added to the mastermix. DNA used was a prenested PCR product (JH38/JH41) from a previous successful nested PCR (B10).

When repeated attempts failed to amplify many of the study samples, it was decided that for this project we would use two PCRs that had already been optimised and proven in the molecular diagnostic section. The *env* PCR amplified a smaller region around the V3 loop of the gp120 protein, and used primers ES7x and ES8x for the prenested one-step RT-PCR, and primers ES7x and ES125 for the heminested PCR (Bachman *et al.*, 1994; Sanders-Buell *et al.*, 1995; Moodley *et al.*, 1998). The *gag* PCR used primers GAG A and GAG B for the prenested one-step RT-PCR, and primers GAG C and GAG D for the nested PCR (Kemp *et al.*, 1989; Engelbrecht and van Rensburg, 1995). Figure 3.3 shows an example of the agarose gel electrophoresis results obtained. All patient samples and the positive control showed amplification.

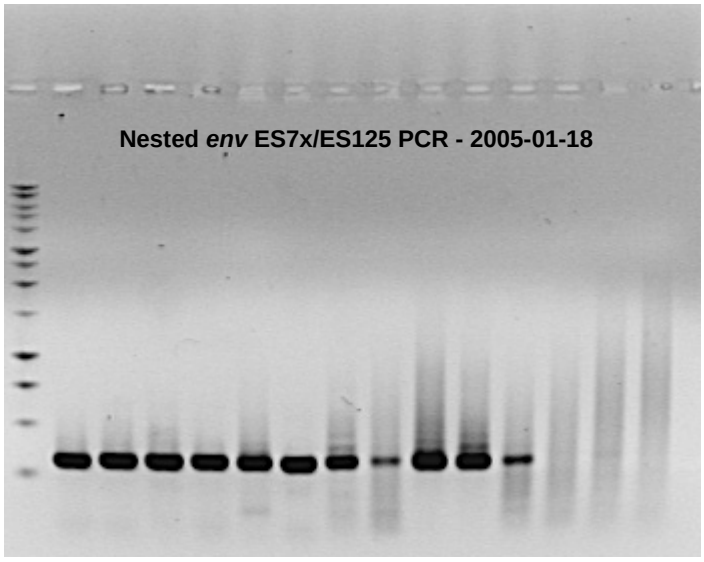


Figure 3.3. 1% agarose gel using nested PCR products. Primers: ES7x and ES125. From left to right: 1kb DNA marker; B6, B2, B7, B3, B4, B5, B1, B9, B8, B10, Pos control, Neg control, reagent blank, reagent blank (repeated due to possible contamination during loading).

The diagnostic *env* PCR was known to be less sensitive than the *gag* PCR; in two pairs, both mother and infant were negative on the *env* PCR. See Table 3.1 for details on the PCR results.

3.4 Sequence alignment

FinchTV 1.3.1 was used to interpret the electropherograms produced by the ABI 310 Genetic Analyzer. Figures 3.4 and 3.5 show the electropherograms of M08 and B08, showing two

areas in M08's sequence that could not be resolved. Letters M and R (see arrows) were used in the mother's sequence to indicate ambiguity. Using the IUPAC Nucleotide Ambiguity Codes, M indicates either A or C, while R indicates either A or G. Arrows on the infant's sequence indicate the corresponding position.

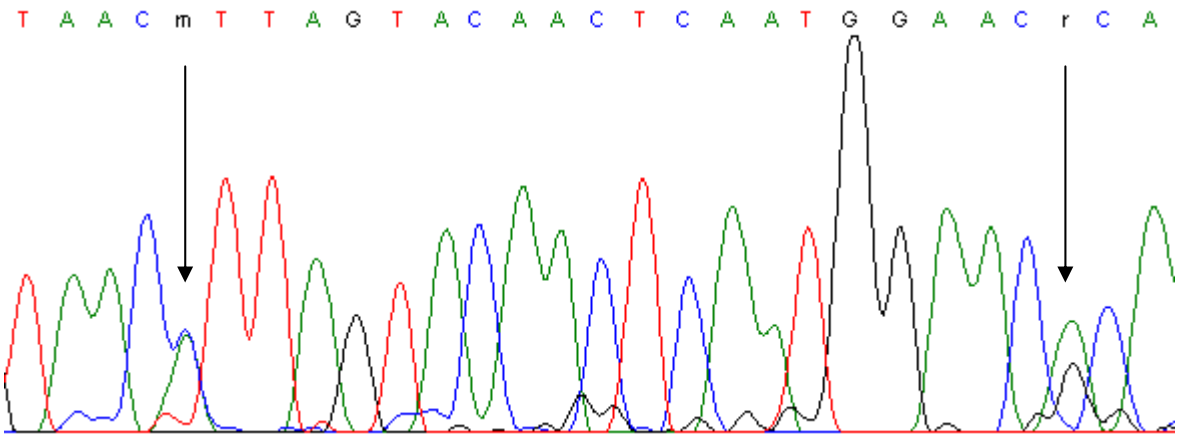


Figure 3.4. Partial electropherograms of M08's sequence sequenced with primer ES7x.

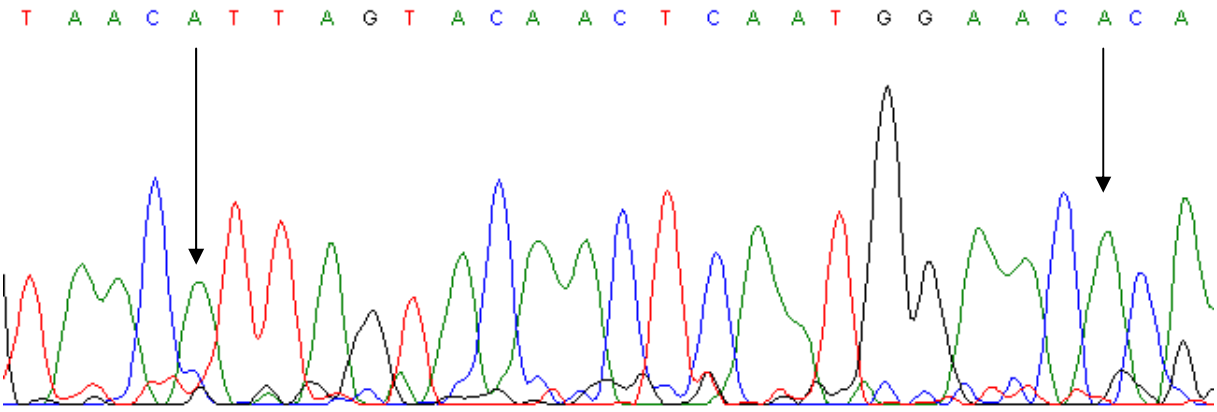


Figure 3.5. Partial electropherograms of B08's sequence sequenced with primer ES7x.

CLUSTAL X was used to create an alignment with minimised gaps using the IUB DNA Weight Matrix and a gap creation penalty of 10, and a gap extension penalty of 0.2. Appendix E shows the alignment created.

Some PCR products did not produce a usable sequence, and were therefore not included in further analysis. See Table 3.1 for the sequences included in further analysis.

BioEdit 7.0.4.1 was used for performing manual codon alignments (Appendix F) on the alignments produced by CLUSTAL X, and to create amino acid sequences (Appendix G).

For the purposes of phylogenetic analysis, short sequences were removed from the data set, and the ends trimmed to leave as few sequences as possible with excessive gaps at either end.

Appendix H shows the codon alignment of the final set of sequences, and Appendix I shows the amino acid alignment of these sequences.

3.5 Phylogenetic analysis

TREECON 1.3b (Van de Peer 1994) was used for performing phylogenetic analyses using the Kimura 2-parameter 1980 model for distance calculation, and construction of bootstrapped neighbour joining trees.

Figure 3.6 shows the neighbour joining tree produced on TREECON. All monophyletic clusters representing HIV subtypes and subsubtypes other than subtype C have been marked as triangles. Names of subtype C reference sequences have been marked in green, mother-infant pairs have been marked in different colours, and other mother or infant sequences are in black. M20a and M20b refer to two sequences obtained from one mother. All paired sequences from mothers and infants cluster together, and all study sequences cluster with the subtype C reference sequences.

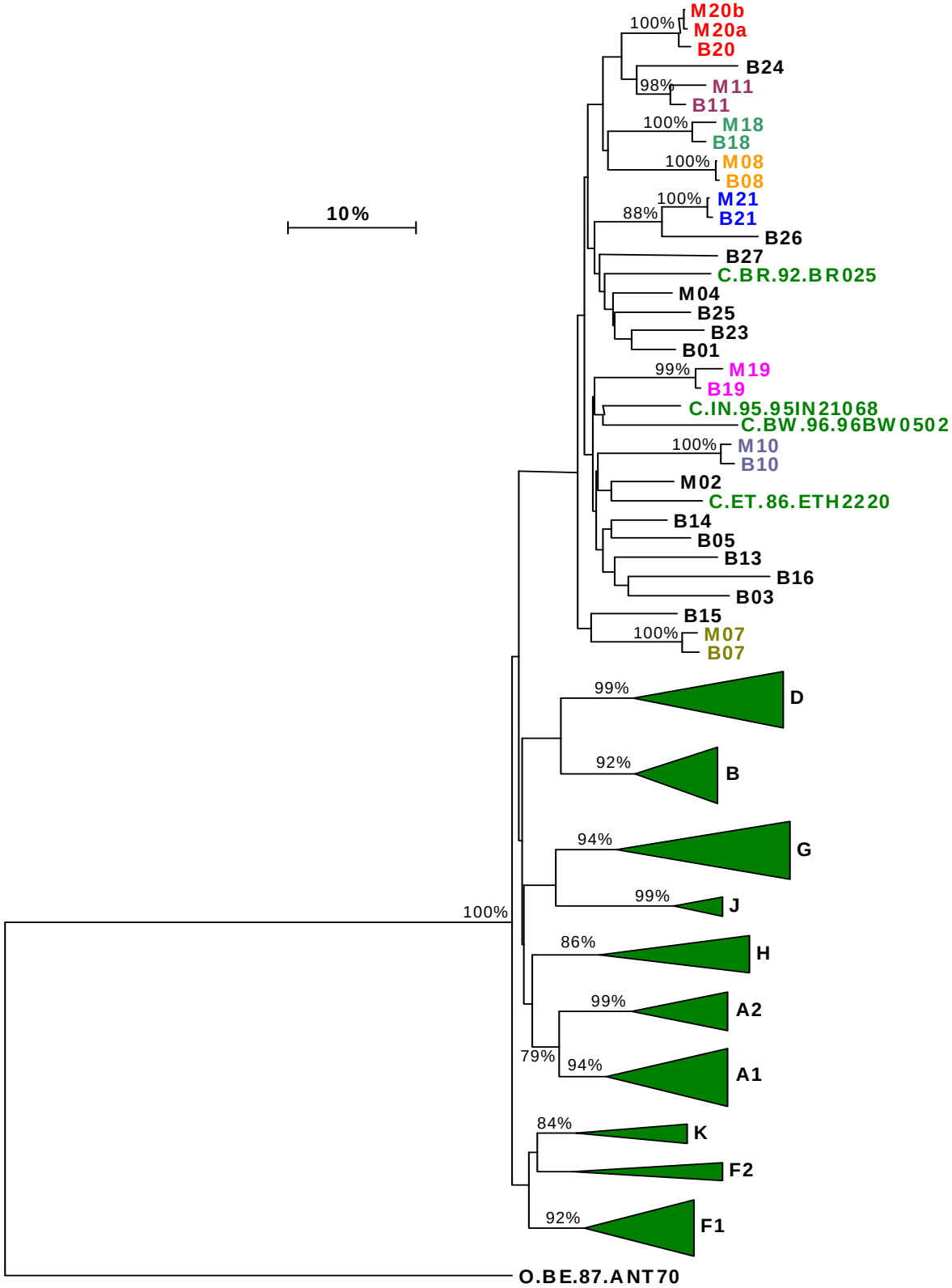


Figure 3.6. Neighbour-joining drawn using the Kimura 2-parameter 1980 model. Bootstraps (1000 bootstraps performed) above 75% are shown. Reference sequences are in green, mother/infant pairs are in colour, and unpaired mother/infant sequences are in black.

The FindModel (Los Alamos National Laboratory, CA, USA) resource at http://www.hiv.lanl.gov/content/hiv-db/findmodel/findmodel.html, based on ModelTest (Posada and Crandall, 1998; Posada and Buckley, 2004) was used to provide information on which model of phylogenetic analysis to use in constructing more detailed phylogenetic trees.

FindModel recommended the General Time Reversible plus Gamma model. The parameters it recommended were base frequencies of 0.2262, 0.1605, 0.4254, and 0.1878 for T, C, A, and G respectively; a transition/transversion ratio of 1.9048, and a shape parameter alpha of 0.65026. The rate matrix Q can be seen in Table 2.2.

Table 3.2. Rate matrix Q recommended by FindModel

	A	C	G	T
A	-0.872477	0.506487	0.233530	0.132460
C	0.713890	-1.270528	0.449971	0.106662
G	0.124173	0.169747	-0.795270	0.501350
T	0.159516	0.091130	1.135467	-1.386112

PAUP* is only able to accept 5 positive values, with the A→C rate equal to the C→A rate, the A→G rate equal to the G→A rate, and so forth. It assumes the T→G / G→T rate is equal to 1. Several methods to adapt this were tried, but no input was accepted that allowed these values to be used. The modified rate matrix used can be seen in Table 2.3.

Table 3.3. Rate matrix Q used in analysis

	A	C	G	T
A	-	0.713890	0.124173	0.159516
C	0.713890	-	0.169747	0.091130
G	0.124173	0.169747	-	1.000000
T	0.159516	0.091130	1.000000	-

After 108 hours, the calculation was only on the 2nd bootstrap, and it was decided not to obtain bootstrap values. The zero branch length test was used instead, with p being the probability of obtaining a likelihood ratio as large or larger than the observed ratio under the null hypothesis that a branch has zero length (PAUP* 4b10 package, 2001).

Figures 3.7a and 3.7b show the maximum likelihood tree produced using these parameters. All monophyletic clusters representing HIV subtypes and subsubtypes other than subtype C have been marked as bars. All paired sequences from mothers and infants cluster together, and all study sequences cluster with the subtype C reference sequences.

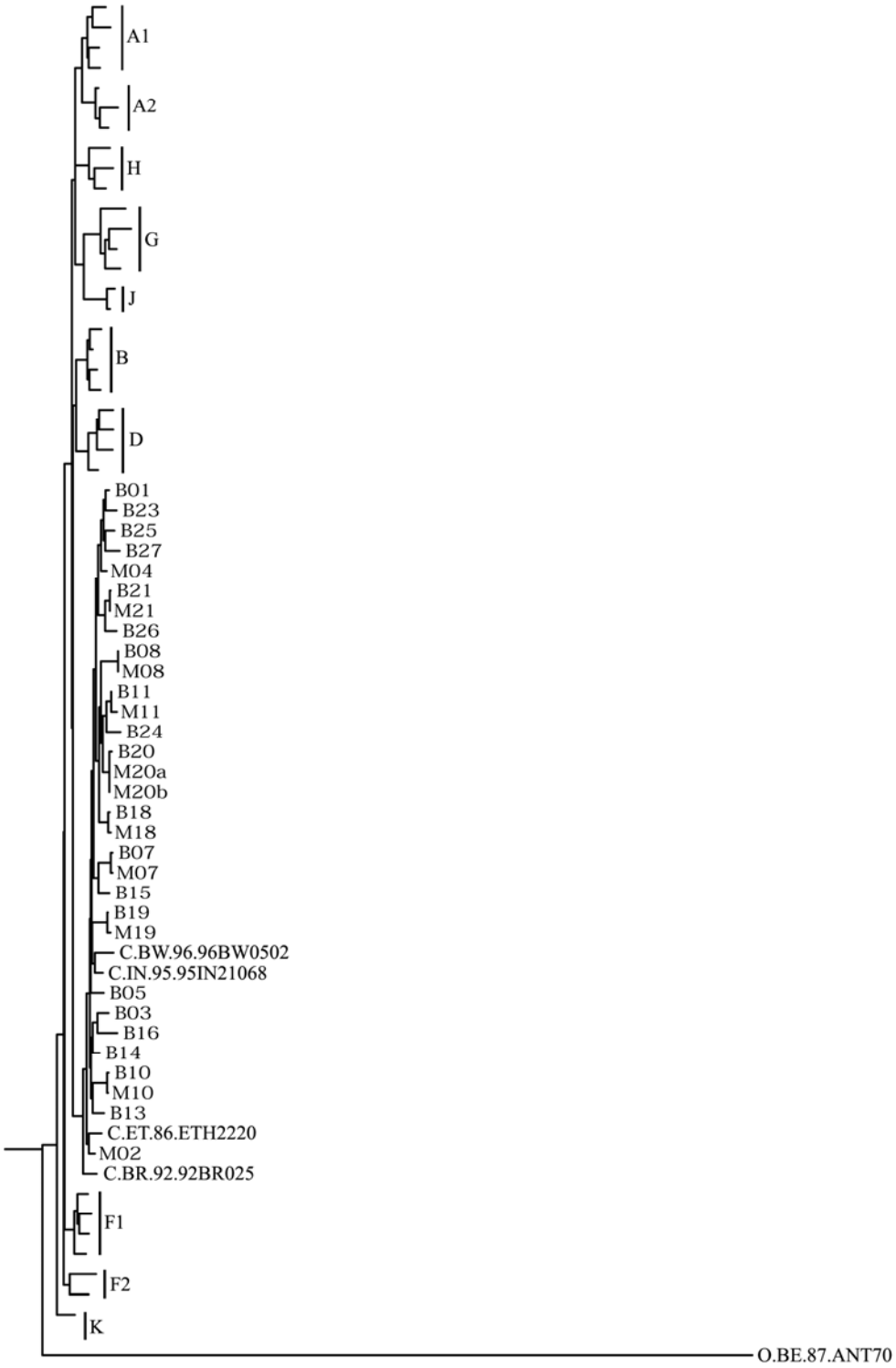


Figure 3.7a. Maximum likelihood tree calculated by PAUP* using FindModel's parameters. PHYLIP was used to draw the tree.

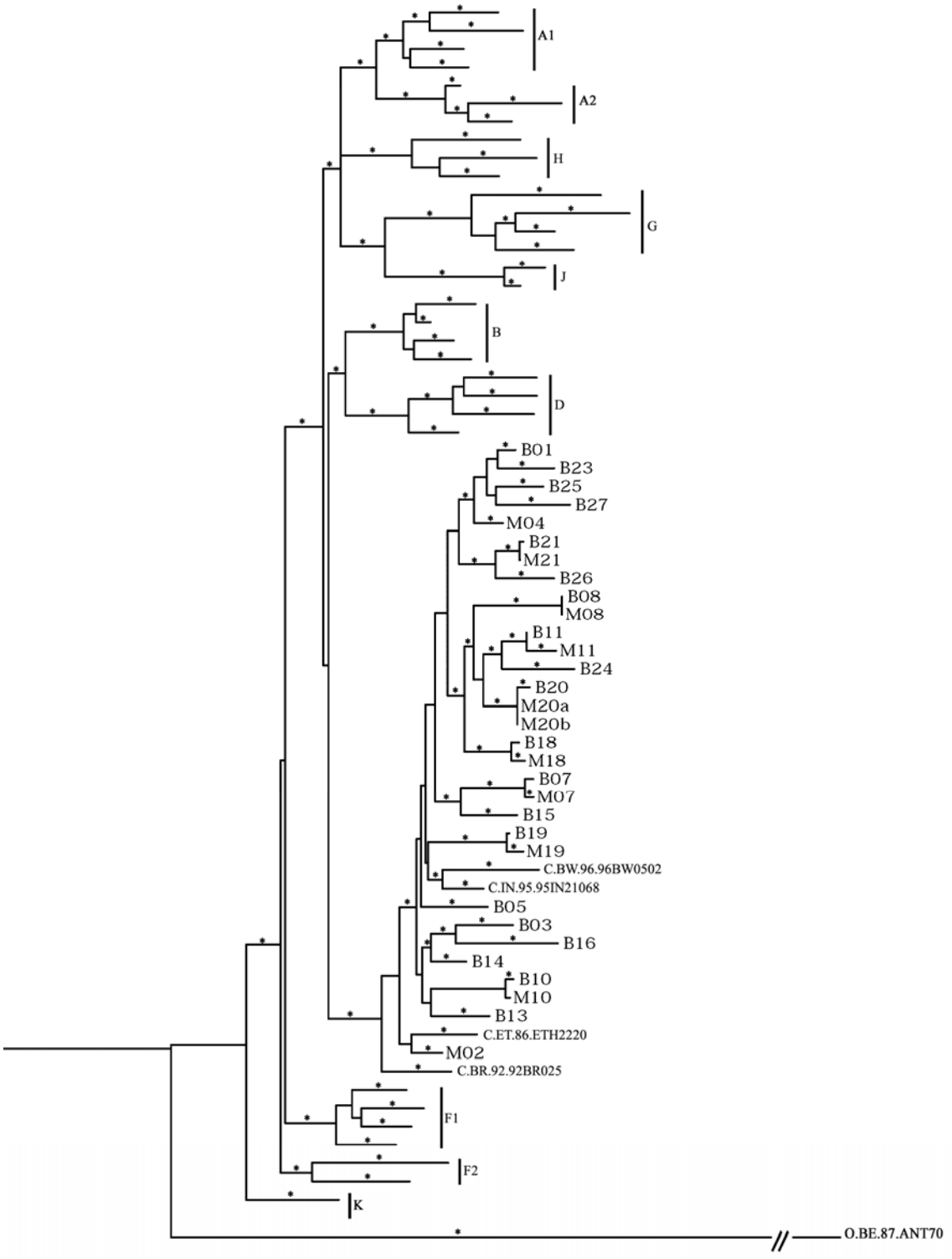


Figure 3.7b. Maximum likelihood tree calculated by PAUP* using FindModel's parameters. The tree has been stretched horizontally, and the long branch of O.BE.87.ANT70 cut off. PHYLIP was used to draw the tree. Asterisks on branches indicate statistical support for the branch (p<0.001, zero branch length test).

PAUP* was allowed to determine its own values for these parameters as well. Those obtained were base frequencies of 0.18843, 0.18539, 0.43483, and 0.19135 for T, C, A, and G respectively; a transition/transversion ratio of 1.694153, and a shape parameter alpha of 0.702772. The rate matrix Q can be seen in Table 2.4.

Table 3.4. Rate matrix Q suggested by PAUP*

	A	C	G	T
A	-	1.400891	3.327623	0.746876
C	1.400891	-	0.695150	3.480019
G	3.327623	0.695150	-	1.000000
T	0.746876	3.480019	1.000000	-

Again, bootstrap values were not calculated, and the statistical support provided by the zero branch length test was used instead.

Figures 3.8a and 3.8b show the maximum likelihood tree produced using these parameters. All monophyletic clusters representing HIV subtypes and subsubtypes other than subtype C have been marked as bars. All paired sequences from mothers and infants cluster together, and all study sequences cluster with the subtype C reference sequences. The inserted text box on the right of the tree in Figure 3.8b indicates the two samples that reflect the least (M08/B08) and greatest (M11/B11) distances between sequences, and the corresponding cumulative time between samples.

When selecting the mother's sample, an error was made, and the dates used were not dates 6 months from the birth, but rather dates 6 months from the date of the infant's specimen. If the dates 6 months from the date of birth were used, there would be a maximum cumulative time of 12 months separating the maternal and infant specimens. Due to the date chosen, a maximum cumulative time of 18 months was possible. (Due to the fact that HIV is constantly evolving, and maternal and infant viruses would continue to diversify between infection and the taking of paired specimens at, for example, 6 months after birth, the virus in the mother's specimen would be separated from the virus infecting the infant by 6 months of evolution, and similarly, the virus in the infant's specimen would be separated from the virus infecting the infant by 6 months of evolution. The cumulative time here would therefore be 12 months).

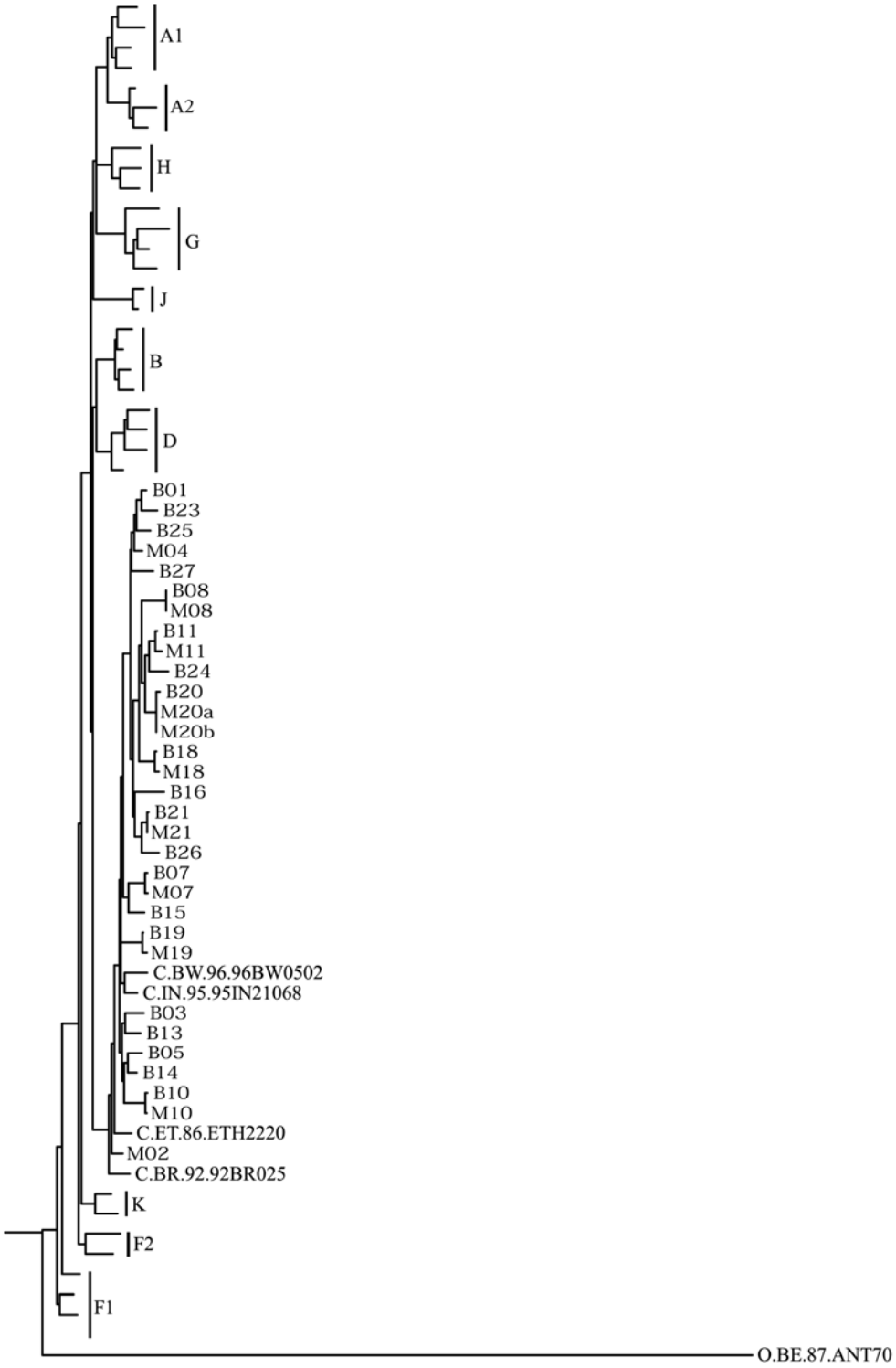


Figure 3.8a. Maximum likelihood tree calculated by PAUP* using PAUP*s calculated parameters. PHYLIP was used to draw the tree.

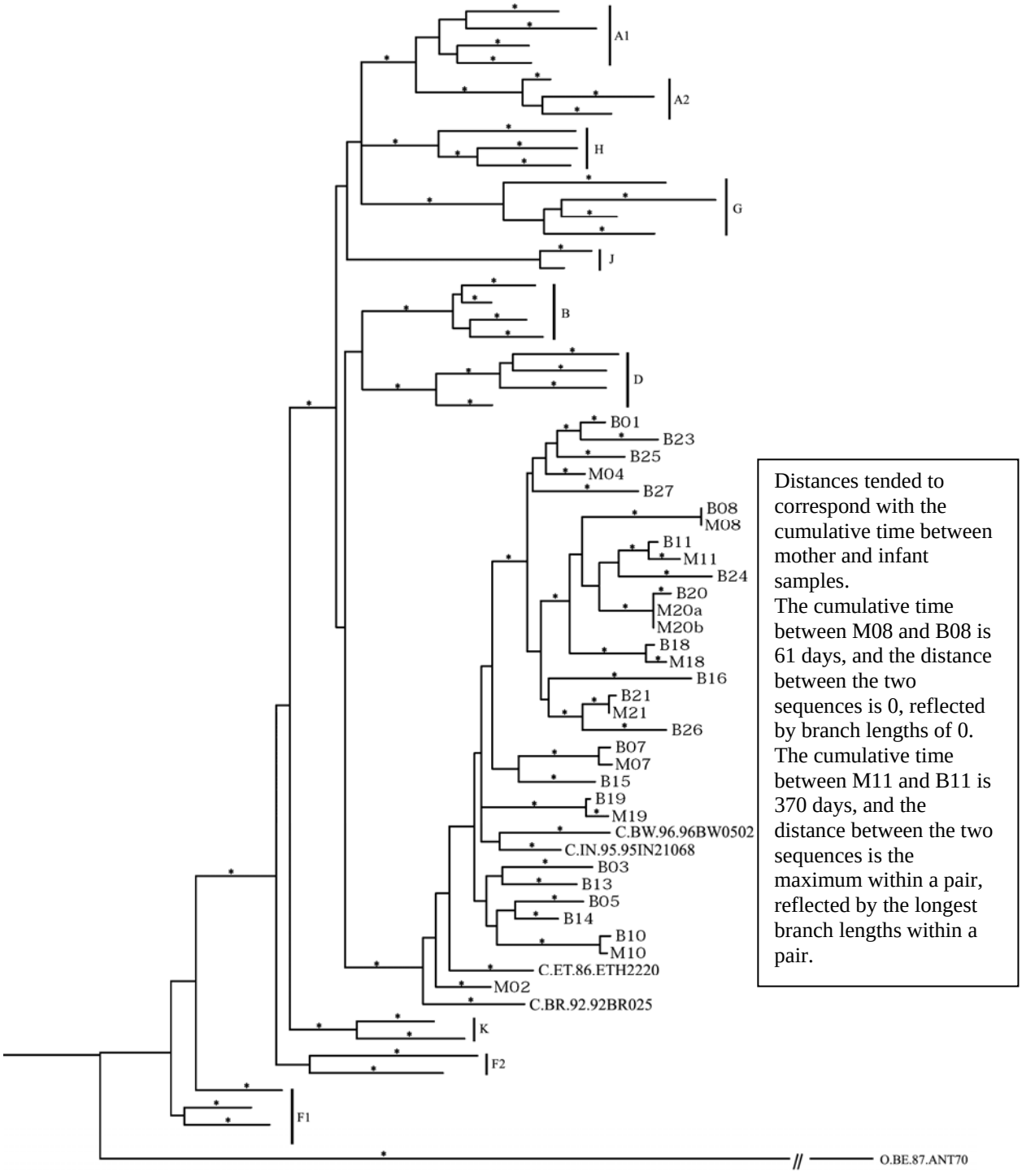


Figure 3.8b. Maximum likelihood tree calculated by PAUP* using PAUP*'s calculated parameters. The tree has been stretched horizontally, and the long branch of O.BE.87.ANT70 cut off. PHYLIP was used to draw the tree. Asterisks on branches indicate statistical support for the branch (p<0.001, zero branch length test).

Chapter 4: Discussion

4.1 Patient samples

Due to the retrospective nature of this study, it was subject to several limitations.

Stored patient samples that had been collected for diagnostic purposes were used for the study, and therefore paired maternal and infant samples were not available. Ideally, such paired samples should be obtained at specified time points relative to parturition, and stored with the aim of the research in mind. The opportunity to estimate timing of infection was therefore lost, and the ability to assess the significance of findings, such as quasispecies diversity, restricted.

4.2 Clinical information

Another disadvantage of a retrospective study is that the clinical notes were often incomplete, and didn't contain all the information relevant to the study. The aim of such clinical notes is to keep record of the management of the patient, for future clinical reference, and so that other clinicians that become involved in the acute management of the patient can be made aware of the situation at that time. Using such data for purposes it was not primarily intended for does result in many uncertainties and potential inaccuracies. However, such retrospective data is valuable in identifying areas for future research.

Errors in clinical notes could be avoided, but after a 24-hour shift, it is understandable that sometimes notes can be ambiguous. Obvious errors are easily detectable, but there may be errors that are not as easily noticed, and the reliability of the data contained in the notes can therefore be questioned.

A large number of risk factors were observed based on the clinical notes, and it became clear that there were many practical problems involved in prevention of vertical HIV transmission. A low level of education and difficult social circumstances in the community, as well as inadequate training and resources for many public health care workers, understandably makes the recommendations for prevention of mother-to-child transmission of HIV difficult to follow completely.

It was therefore not unexpected to find mothers with inadequate knowledge regarding the high risk of mixed feeds with both formula feeds and breast milk, as well as the increased risk of transmission due to mastitis (Semba *et al.*, 1999; Geijtenbeek *et al.*, 2001). Mothers often present at the labour ward for the first time when they realise they are in labour, without

having attended antenatal clinics. The result of this is that caesarean sections are often emergency procedures and not elective ones, and avoidance of breech deliveries and foetal distress is not optimal.

Twin studies are exceptionally valuable in HIV research, and discordant pairs are especially interesting. Studies of twins can lead to a greater understanding of the underlying mechanisms that may protect one infant but allow or cause infection of another. In our patient database, we have one pair of discordant twins, indicated in the clinical notes as being identical twins. Further study of the samples from this infant and mother are warranted.

4.3 PCR results

Initially we experienced a problem with the two sets of primers chosen for *env* PCR. The primers are degenerate primers designed to detect most of group M sequences. Certain samples could be amplified using these primers, while most could not. This finding was therefore unexpected, but it is possible that the long duration of storage of samples, with repeated thawing and re-freezing, may have influenced the quality of the viral RNA. RNA is known to fragment when exposed to such stresses. The problems caused by fragmentation can be minimised by amplifying a short section of the genome, as there is a greater possibility that the short region will be intact in a higher percentage of genomes present in the sample. Another possible cause for the failed amplification is that the development of the primers, which were meant to amplify most HIV-1 type M sequences (Swanson *et al.*, 2003), did not use many subtype C sequences into the analysis, and therefore may not be ideal for amplification of subtype C genomes. It was therefore decided to use a PCR well-established in the department's diagnostic section, which amplifies a shorter fragment (Engelbrecht and van Rensburg, 1995; Moodley *et al.*, 1998*).*

This PCR, a combination of two RT-PCRs and two nested PCRs, one each for *gag* and *env*, produced some discordant results. The *gag* PCR was the most sensitive, as had been seen previously in its use in diagnostics. Only one sample was negative on the *gag* PCR, and that sample was positive on the *env* PCR. A number of samples were negative on the *env* PCR. The infant specimens negative on this PCR were also negative previously when this PCR was used for diagnosis. Of the eight maternal specimens negative on the *env* PCR, two correlated with a negative *env* PCR on the infant's sample, while three were associated with a weak band for the infant's specimen. Further research may be warranted to investigate diversity in the *env* primer binding regions.

4.4 Sequences

In areas of some of the sequences, there were ambiguities that could not be resolved. Fig 3.4 and 3.5 in chapter 3 showed a partial electropherogram from a mother's sequence and one from her infant's sequence. In the mother's sequence, two peaks seemed to coincide in two places, whereas there seemed to be only one peak in the infant's sequence. This could be due to a sequencing problem, but it could represent two different sequences in the mother. Further studies, with the creation of at least 20 clones from each sample that showed this phenomenon, may clarify this issue by identifying multiple quasispecies amongst the clones.

Codon aligned sequences represent a state that is more likely to be the true alignment of the sequences used. A single nucleotide deletion or insertion (indel) will usually cause a nonsense mutation, shifting the reading frame by one nucleotide, which would result in a nonfunctional protein. A virus with such a protein is unlikely to replicate efficiently. The more natural types of indels are sets of 3 nucleotides, representing a single amino acid. Similarly, it is unlikely that a conserved motif will be deleted in one position, and an identical one inserted in another position. Therefore, where such motifs, such as glycosylation sites, exist, an alignment that aligns these motifs is more likely to represent the true alignment.

Observed amino acid deletions were conserved in all sequences from mother and infant sequence pairs. Sequences from infant B05 and pair B14/M14 showed a two amino acid deletion in the same location. Three single amino acid insertions relative to the subtype C reference sequence were observed, one in infant B03's sequence, one in infant B26's sequence, and another in the sequence from infant B27.

Further analysis on these sequences in the future, including a more optimised alignment using glycosylation sites and other conserved motifs, may provide more information.

4.5 Phylogenetic analysis

The neighbour joining method was used initially, to obtain a phylogenetic tree with bootstrap values. Maximum likelihood is a better method for this sort of sequence, but is extremely computationally expensive. Each heuristic search for the best maximum likelihood tree took 8-12 hours. When a bootstrapped tree was attempted, the software was still processing the 2nd bootstrap after approximately 108 hours, and it was decided not to continue. Another statistical approach was used – the zero branch length test. This test provides a statistical assessment for the existence of a particular branch.

On all threes drawn, all the study sequences clustered with subtype C, indicating that for the *env* sequence analysed, these viruses were subtype C viruses. Multiple regions were not analysed, and therefore the presence of recombinant viruses cannot be excluded.

As was expected, all infant and maternal sequences clustered as pairs, indicating that there was not a sample or infant mix-up, and corresponding to known studies regarding diversity of mother/infant sequences.

Some studies have noted that viral sequence divergence, and the formation of a significant number of the infant's own quasispecies, tends, on average, to become prominent at about 7 months after infection. For this reason, we chose to use only infant samples taken within 6 months of birth. However, when selecting the mother's sample, an error was made, and the dates used were not dates 6 months from the birth, but rather dates 6 months from the date of the infant's specimen. If the dates 6 months from the date of birth were used, there would be a maximum cumulative time of 12 months separating the maternal and infant specimens. Due to the date chosen, a maximum cumulative time of 18 months was possible. (Due to the fact that HIV is constantly evolving, and maternal and infant viruses would continue to diversify between infection and the taking of paired specimens at, for example, 6 months after birth, the virus in the mother's specimen would be separated from the virus infecting the infant by 6 months of evolution, and similarly, the virus in the infant's specimen would be separated from the virus infecting the infant by 6 months of evolution. The cumulative time here would therefore be 12 months).

This error was, however, serendipitous, as it permitted us to have extra sequences for a bigger analysis, as well as have one more mother-infant pair included in the final phylogenetic analysis. The potential disadvantage of having an extra 6 months of evolution between mother and infant virus did not realise – although the maximum cumulative time could have been 18 months, in the mother/infant pairs used, the longest time between the specimens was 12 months and 15 days.

The advantage of having the extra pair of sequences was that we could see that the branch distances between the mother and infant sequences tended to correspond to the cumulative time between them – the shortest distance between the pair closest to each other in time (61 days, M08/B08), and the longest distance between the pair furthest from each other (370 days, M11/B11). This finding is in keeping with other studies on viral diversity after infection.

4.6 Conclusions

As these infants were recently infected by their mothers, and some of the mothers have had subsequent children, these sequences represent currently circulating viruses in the population of women in the Western Cape, and, by extension, in the heterosexual community of the Western Cape. Studies on viruses currently circulating are important in observing trends in subtype evolution, and discerning clinical factors that may arise from these trends.

The study supports previous data showing that subtype C is predominant amongst the pregnant women of the Western Cape, and, by extension, their sexual partners.

The study has also identified areas where further research can be directed, both on the samples in the current database, as well as vertically transmitted HIV in the Western Cape.

Since vertically transmitted HIV is an ideal transmitter-recipient model for HIV infection, due to it being common in South Africa, and the timing of infection more easily determined than in adults, it is an important aspect of HIV on which to focus research.

References

Abdool Karim S, Abdool Karim Q, Adhikari M, *et al*. Vertical HIV transmission in South Africa: translating research into policy and practice. *Lancet* 2002; 359: 992-3.

Ahmad N, Baroudy BM, Baker RC, Chappey C. Genetic analysis of human immunodeficiency virus type 1 envelope V3 region isolates from mothers and infants after perinatal transmission. *J Virol* 1995; 69: 1001-12.

Ahmad N. The vertical transmission of human immunodeficiency virus type 1: molecular and biological properties of the virus. *Crit Rev Clin Lab Sci* 2005; 42: 1-34.

Akouamba BS, Viel J, Charest H, *et al*. HIV-1 genetic diversity in antenatal cohort, Canada. *Emerg Infect Dis* 2005; 11: 1230-4.

Alaeus A, Lidman K, Bjorkman A, Giesecke J, Albert J. Similar rate of disease progression among individuals infected with HIV-1 genetic subtypes A-D. *AIDS* 1999; 13: 901-7.

Amedee AM, Rychert J, Lacour N, Fresh L, Ratterree M. Viral and immunological factors associated with breast milk transmission of SIV in rhesus macaques. *Retrovirology* 2004; 1: 17.

Amornkul PN, Tansuphasawadikul S, Limpakarnjanarat K, *et al*. Clinical disease associated with HIV-1 subtype B´ and E infection among 2104 patients in Thailand. *AIDS* 1999; 13: 1963-9.

Arens M. Methods for subtyping and molecular comparison of human viral genomes. *Clin Microbiol Rev* 1999; 12: 612-26.

Auger I, Thomas P, De Gruttola V, *et al*. Incubation periods for paediatric AIDS patients. *Nature* 1988; 336: 575-7.

Bachman MH, Delwart EL, Shpaer EG, *et al*. Rapid genetic characterization of HIV type 1 strains from four WHO-sponsored vaccine evaluation sites using a heteroduplex mobility assay. *AIDS Res Hum Retroviruses* 1994;10:1345-53.

Bailey RC, Kamenga MC, Nsuami MJ, Nieburg P, St Louis ME. Growth of children according to maternal and child HIV, immunological and disease characteristics: a prospective cohort study in Kinshasa, Democratic Republic of Congo. *Int J Epidemiol* 1999; 28: 532-40.

Barber CG. CCR5 antagonists for the treatment of HIV. *Curr Opin Investig Drugs* 2004; 5: 851-61.

Barnhart HX, Caldwell MB, Thomas P, *et al*. Natural history of human immunodeficiency virus disease in perinatally infected children: an analysis from the Pediatric Spectrum of Disease Project. *Pediatrics* 1996; 97: 710-6.

Berhane R, Bagenda D, Marum L, *et al*. Growth failure as a prognostic indicator of mortality in pediatric HIV infection. *Pediatrics* 1997; 100:

Biggar RJ, Broadhead R, Janes M, Kumwenda N, Taha TE, Cassol S. Viral levels in newborn African infants undergoing primary HIV-1 infection. *AIDS* 2001; 15: 1311-3.

Biggar RJ, Cassol S, Kumwenda N, *et al*. The risk of human immunodeficiency virus-1 infection in twin pairs born to infected mothers in Africa. *J Infect Dis* 2003; 188: 850-5. Epub 2003 Sep 9.

Biggar RJ, Janes M, Pilon R, *et al*. Human immunodeficiency virus type 1 infection in twin pairs infected at birth. *J Infect Dis* 2002; 186: 281-5. Epub 2002 Jun 17.

Blackard JT, Renjifo B, Chaplin B, Msamanga G, Fawzi W, Essex M. Diversity of the HIV-1 long terminal repeat following mother-to-child transmission. *Virology* 2000; 274: 402-11.

Blackard JT, Renjifo B, Fawzi W, *et al*. HIV-1 LTR subtype and perinatal transmission. *Virology* 2001; 287: 261-5.

Blanche S, Newell ML, Mayaux MJ, *et al*. Morbidity and mortality in European children vertically infected by HIV-1. The French Pediatric HIV Infection Study Group and European Collaborative Study. *J Acquir Immune Defic Syndr Hum Retrovirol* 1997; 14: 442-50.

Blanche S, Tardieu M, Duliege A, *et al*. Longitudinal study of 94 symptomatic infants with perinatally acquired human immunodeficiency virus infection. Evidence for a bimodal expression of clinical and biological symptoms. *Am J Dis Child* 1990; 144: 1210-5.

Bobkov A, Garaev MM, Rzhaninova A, *et al*. Molecular epidemiology of HIV-1 in the former Soviet Union: analysis of env V3 sequences and their correlation with epidemiologic data. *AIDS* 1994; 8: 619-24.

Bongertz V. Vertical human immunodeficiency virus type 1--HIV-1--transmission--a review. *Mem Inst Oswaldo Cruz* 2001; 96: 1-14.

Brahmbhatt H, Kigozi G, Wabwire-Mangen F, *et al*. The effects of placental malaria on mother-to-child HIV transmission in Rakai, Uganda. *AIDS* 2003; 17: 2539-41.

Breastfeeding and HIV International Transmission Study Group. Mortality among HIV-1-infected women according to children´s feeding modality: an individual patient data meta-analysis. *J Acquir Immune Defic Syndr* 2005; 39: 430-8.

Bredell H, Hunt G, Casteling A, *et al*. HIV-1 Subtype A, D, G, AG and unclassified sequences identified in South Africa. *AIDS Res Hum Retroviruses* 2002; 18: 681-3.

Brumme ZL, Chan KJ, Dong W, *et al*. CCR5Delta32 and promoter polymorphisms are not correlated with initial virological or immunological treatment response. *AIDS* 2001; 15: 2259-66.

Cane PA, de Ruiter A, Rice P, Wiselka M, Fox R, Pillay D. Resistance-associated mutations virus type 1 subtype C protease gene from treated and untreated patients in the United Kingdom. *J Clin Microbiol* 2001; 39: 2652–54.

Carrillo MG, Avila M, Hierholzer J, *et al*. Mother-to-Child HIV Type 1 Transmission in Argentina: BF Recombinants Have Predominated in Infected Children Since the Mid-1980s. *AIDS Res Hum Retroviruses* 2002;18:477–83.

Casado G, Thomson MM, Sierra M, Najera R. Identification of a novel HIV-1 circulating ADG intersubtype recombinant form (CRF19_cpx) in Cuba. *J Acquir Immune Defic Syndr* 2005; 40: 532-7.

Centres for Disease Control (CDC). Guidelines for the use of antiretroviral agents in pediatric HIV infection. *MMWR* 1998; 47: 1–43.

Chearskul S, Chotpitayasunondh T, Simonds RJ, *et al*. Survival, disease manifestations, and early predictors of disease progression among children with perinatal human immunodeficiency virus infection in Thailand. *Pediatrics* 2002; 110:

Clevestig P, Maljkovic I, Casper C, *et al*. The X4 phenotype of HIV type 1 evolves from R5 in two children of mothers, carrying X4, and is not linked to transmission. *AIDS Res Hum Retroviruses* 2005; 21: 371-8.

Coovadia H. Antiretroviral agents--how best to protect infants from HIV and save their mothers from AIDS. *N Engl J Med* 2004; 351: 289-92. Epub 2004 Jul 9.

Courgnaud V, Laure F, Brossard A, *et al*. Frequent and early in utero HIV-1 infection. *AIDS Res Hum Retroviruses* 1991; 7: 337-41.

Coutsoudis A, Coovadia H, Pillay K, Kuhn L. Are HIV-infected women who breastfeed at increased risk of mortality? *AIDS* 2001a; 15: 653-5.

Coutsoudis A, Pillay K, Kuhn L, Spooner E, Tsai WY, Coovadia HM. Method of feeding and transmission of HIV-1 from mothers to children by 15 months of age: prospective cohort study from Durban, South Africa. *AIDS* 2001b; 15: 379-87.

Coutsoudis A, Pillay K, Spooner E, Coovadia HM, Pembrey L, Newell ML. Morbidity in children born to women infected with human immunodeficiency virus in South Africa: does mode of feeding matter? *Acta Paediatr* 2003; 92: 890-5.

Coutsoudis A, Pillay K, Spooner E, Kuhn L, Coovadia HM. Randomized trial testing the effect of vitamin A supplementation on pregnancy outcomes and early mother-to-child HIV-1 transmission in Durban, South Africa. South African Vitamin A Study Group. *AIDS* 1999; 13: 1517-24.

Cullen BR. Species and Tissue Tropisms of HIV-1: Molecular Basis and Phenotype Consequences. In: HIV Sequence Compendium 2001. (Kuiken C, Foley B, Hahn B, Marx P, McCutchan F, Mellors JW, Wolinsky S, Korber B, editors). Los Alamos, New Mexico, USA. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, 2001; 71-82.

Dabis F, Ekpini ER. HIV-1/AIDS and maternal and child health in Africa. *Lancet* 2002; 359: 2097-104.

De Cock KM, Fowler MG, Mercier E, *et al*. Prevention of mother-to-child HIV transmission in resource-poor countries: translating research into policy and practice. *JAMA* 2000; 283: 1175-82.

Department of Health, South Africa. National HIV and syphilis antenatal sero-prevalence survey in South Africa, 2004. Department of Health, Pretoria, South Africa, 2005.

Descamps D, Apetrei C, Collin G, Damond F, Simon F, Brun-Vezinet F. Naturally occurring decreased susceptibility of HIV-1 subtype G to protease inhibitors. *AIDS* 1998; 12: 1109-11.

Diaz C, Hanson C, Cooper ER, *et al*. Disease progression in a cohort of infants with vertically acquired HIV infection observed from birth: the Women and Infants Transmission Study (WITS). *J Acquir Immune Defic Syndr Hum Retrovirol* 1998; 18: 221-8.

Dickover RE, Dillon M, Leung KM, *et al*. Early prognostic indicators in primary perinatal human immunodeficiency virus type 1 infection: importance of viral RNA and the timing of transmission on long-term outcome. *J Infect Dis* 1998; 178: 375-87.

Dickover RE, Garratty EM, Plaeger S, Bryson YJ. Perinatal transmission of major, minor, and multiple maternal human immunodeficiency virus type 1 variants in utero and intrapartum. *J Virol* 2001; 75: 2194-203.

Domachowske JB. Pediatric human immunodeficiency virus infection. *Clin Microbiol Rev* 1996; 9: 448-68.

Doms RW and Moore JP. HIV-1 Coreceptor Use: A Molecular Window into Viral Tropism. In: HIV Molecular Immunology Database 1997. (Korber B, Brander C, Haynes B, Koup R, Moore J, Walker B, editors). Los Alamos, New Mexico, USA. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, 1997; IV-25-36.

Dorr P, Westby M, Dobbs S, *et al*. Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. *Antimicrob Agents Chemother* 2005; 49: 4721-32.

Dorrington RE, Bradshaw D, Budlender D. HIV/AIDS profile of the provinces of South Africa – indicators for 2002. Centre for Actuarial Research, Medical Research Council and the Actuarial Society of South Africa. 2002.

Dreyfuss ML, Fawzi WW. Micronutrients and vertical transmission of HIV-1. *Am J Clin Nutr* 2002; 75: 959-70.

Dunn DT, Brandt CD, Krivine A, *et al*. The sensitivity of HIV-1 DNA polymerase chain reaction in the neonatal period and the relative contributions of intra-uterine and intra-partum transmission. *AIDS* 1995; 9:

Ehrnst A, Lindgren S, Dictor M, *et al*. HIV in pregnant women and their offspring: evidence for late transmission. *Lancet* 1991; 338: 203-7.

Ehrnst A, Zetterstrom R. Vertical transmission of HIV-1 infection and dilemma of infant feeding. *Acta Paediatr* 2003; 92: 990-1.

Engelbrecht S, Smith TL, Kasper P, *et al*. HIV type 1 V3 domain serotyping and genotyping in Gauteng, Mpumalanga, KwaZulu-Natal, and Western Cape Provinces of South Africa. *AIDS Res Hum Retroviruses* 1999; 15: 325-8.

Engelbrecht S, van Rensburg EJ. Detection of southern African human immunodeficiency virus type 1 subtypes by polymerase chain reaction: evaluation of different primer pairs and conditions. *J Virol Methods* 1995; 55: 391-400.

European Collaborative Study. Maternal viral load and vertical transmission of HIV-1: an important factor but not the only one. *AIDS* 1999; 13: 1377-85.

European Collaborative Study. Risk factors for mother-to-child transmission of HIV-1. *Lancet* 1992; 339: 1007-12.

Fawzi W, Msamanga G, Renjifo B, *et al*. Predictors of intrauterine and intrapartum transmission of HIV-1 among Tanzanian women. *AIDS* 2001; 15: 1157-65.

Fawzi WW, Msamanga GI, Hunter D, *et al*. Randomized trial of vitamin supplements in relation to transmission of HIV-1 through breastfeeding and early child mortality. *AIDS* 2002; 16: 1935-44.

Felsenstein, J. PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, USA, 2005.

Fenyo EM, Schuitemaker H, Asjo B, McKeating J. The History of HIV-1 Biological Phenotypes Past, Present, and Future. In: Human Retroviruses and AIDS 1997. (Korber B, Hahn B, Foley B, Mellors JW, Leitner T, Myers G, McCutchan F, Kuiken CL, editors). Los Alamos, New Mexico, USA. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, 1997; III-13-18.

Filho DJ, Sucupira MC, Casiero MM, Sabino EC, Diaz RS, Janini LM. Identification of Two HIV Type 1 Circulating Recombinant Forms in Brazil. *AIDS Res Hum Retroviruses* 2006; 22: 1-13.

Gabiano C, Tovo PA, de Martino M, *et al*. Mother-to-child transmission of human immunodeficiency virus type 1: risk of infection and correlates of transmission. *Pediatrics* 1992; 90: 369-74.

Gayle HD, Hill GL. Global impact of human immunodeficiency virus and AIDS. *Clin Microbiol Rev* 2001; 14: 327-35.

Geijtenbeek TB, van Vliet SJ, van Duijnhoven GC, Figdor CG, van Kooyk Y. DC-SIGN, a dentritic cell-specific HIV-1 receptor present in placenta that infects T cells in trans-a review. *Placenta* 2001; 22: Suppl.

Gordon M, De Oliveira T, Bishop K, *et al*. Molecular characteristics of human immunodeficiency virus type 1 subtype C viruses from KwaZulu-Natal, South Africa: implications for vaccine and antiretroviral control strategies. *J Virol* 2003; 77: 2587-99.

Hall, TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 1999;41:95-8.

Hermione Lyall EG. Paediatric HIV in 2002--a treatable and preventable infection. *J Clin Virol* 2002; 25: 107-19.

Ho KM, Ho KK, Lim WL, Li P, Wong KH. Epidemiology and detection of human immunodeficiency virus among pregnant women in Hong Kong. *Hong Kong Med J* 2001; 7: 335-42.

Hoffman IF, Martinson FE, Stewart PW, *et al*. Human immunodeficiency virus type 1 RNA in breast-milk components. *J Infect Dis* 2003; 188: 1209-12. Epub 2003 Oct 1.

Human Sciences Research Council. Nelson Mandela/HSRC study of HIV/AIDS - South African National HIV Prevalence, Behavioural Risks and Mass Media - Household Survey 2002. Human Sciences Research Council Publishers, Cape Town, South Africa, 2002.

Hutto C, Zhou Y, He J, *et al*. Longitudinal studies of viral sequence, viral phenotype, and immunologic parameters of human immunodeficiency virus type 1 infection in perinatally infected twins with discordant disease courses. *J Virol* 1996; 70: 3589-98.

Inion I, Mwanyumba F, Gaillard P, *et al*. Placental malaria and perinatal transmission of human immunodeficiency virus type 1. *J Infect Dis* 2003; 188: 1675-8. Epub 2003 Nov 10.

International Perinatal HIV Group. The mode of delivery and the risk of vertical transmission of human immunodeficiency virus type 1--a meta-analysis of 15 prospective cohort studies. *N Engl J Med* 1999; 340: 977-87.

Ioannidis JP, Contopoulos-Ioannidis DG, Rosenberg PS, *et al*. Effects of CCR5-delta32 and CCR2-64I alleles on disease progression of perinatally HIV-1-infected children: an international meta-analysis. *AIDS* 2003; 17: 1631-8.

Italian Multicentre Study. Epidemiology, clinical features, and prognostic factors of paediatric HIV infection. *Lancet* 1988; 2: 1043-6.

Jayaraman P, Mohan D, Polacino P, *et al*. Perinatal transmission of SHIV-SF162P3 in Macaca nemestrina. *J Med Primatol* 2004; 33: 243-50.

Jeffery BS, Mercer KG. Pretoria pasteurisation: a potential method for the reduction of postnatal mother to child transmission of the human immunodeficiency virus. *J Trop Pediatr* 2000; 46: 219-23.

Kandathil AJ, Ramalingam S, Kannangai R, David S, Sridharan G. Molecular epidemiology of HIV. *Indian J Med Res* 2005; 121: 333-44.

Kanki PJ, Hamel DJ, Sankale JL, *et al*. Human immunodeficiency virus type 1 subtypes differ in disease progression. *J Infect Dis* 1999; 179: 68-73.

Karpas A. Human retroviruses in leukaemia and AIDS: reflections on their discovery, biology and epidemiology. *Biol Rev Camb Philos Soc* 2004; 79: 911-33.

Katz JM, Fox CH, Eglinton GS, Meyers WA 3rd, Queenan JT. Relationship between human immunodeficiency virus-1 RNA identification in placenta and perinatal transmission. *J Perinatol* 1997; 17: 119-24.

Kemp DJ, Smith DB, Foote SJ, Samaras N, Peterson MG. Colorimetric detection of specific DNA segments amplified by polymerase chain reactions. *Proc Natl Acad Sci* U S A 1989; 86: 2423-7.

Koning F, van Rij R, Schuitemaker H. Biological and Molecular Aspects of HIV-1 Coreceptor Usage. In: HIV Sequence Compendium 2002. (Kuiken C, Foley B, Freed E, Hahn B, Korber B, Marx PA, McCutchan F, Mellors JW, Wolinksy S, editors). Los Alamos, New Mexico, USA. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, 2002; 24-42.

Korber B, Foley BT, Kuiken C, Pillai SK, Sodroski JG. Numbering Positions in HIV Relative to HXB2CG. In: Human Retroviruses and AIDS 1998. (Korber B, Kuiken CL, Foley B, Hahn B, McCutchan F, Mellors JW, and Sodroski J, editors). Los Alamos, New Mexico, USA. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, 1998; III-102-111.

Koulinska IN, Chaplin B, Mwakagile D, Essex M, Renjifo B. Hypermutation of HIV type 1 genomes isolated from infants soon after vertical infection. *AIDS Res Hum Retroviruses* 2003; 19: 1115-23.

Kourtis AP, Butera S, Ibegbu C, Beled L, Duerr A. Breast milk and HIV-1: vector of transmission or vehicle of protection? *Lancet Infect Dis* 2003; 3: 786-93.

Krogstad P, Eshleman SH, Geng Y, *et al*. Mother-to-child transmission in the United States of subtypes D and A/G human immunodeficiency virus type 1. *AIDS Res Hum Retroviruses* 2002; 18: 413-7.

Kuhn L, Abrams EJ, Chinchilla M, Tsai WY, Thea DM. Sensitivity of HIV-1 DNA polymerase chain reaction in the neonatal period. The New York City Perinatal HIV Transmission Collaborative Study Group. *AIDS* 1996; 10: 1181-2.

Kuhn L, Kasonde P, Sinkala M, *et al*. Prolonged breast-feeding and mortality up to two years post-partum among HIV-positive women in Zambia. *AIDS* 2005; 19: 1677-81.

Kwiek JJ, Mwapasa V, Milner DA, *et al*. Maternal-Fetal Microtransfusions and HIV-1 Mother-to-Child Transmission in Malawi. *PLoS Med* 2006; 3:

Lallemant M, Jourdain G, Le Coeur S, *et al*. Single-dose perinatal nevirapine plus standard zidovudine to prevent mother-to-child transmission of HIV-1 in Thailand. *N Engl J Med* 2004; 351: 217-28. Epub 2004 Jul 9.

Landesman SH, Kalish LA, Burns DN, *et al*. Obstetrical factors and the transmission of human immunodeficiency virus type 1 from mother to child. The Women and Infants Transmission Study. *N Engl J Med* 1996; 334: 1617-23.

Langston C, Cooper ER, Goldfarb J, *et al*. Human immunodeficiency virus-related mortality in infants and children: data from the pediatric pulmonary and cardiovascular complications of vertically transmitted HIV (P(2)C(2)) Study. *Pediatrics* 2001; 107: 328-38.

Laurent C, Delaporte E. Epidemiology of HIV Infection in sub-Saharan Africa. *AIDS Rev* 2001;3:59-66.

Lemey P, Pybus OG, Wang B, Saksena NK, Salemi M, Vandamme AM. Tracing the origin and history of the HIV-2 epidemic. *Proc Natl Acad Sci* U S A 2003; 100: 6588-92. Epub 2003 May 12.

Los Alamos National Laboratory HIV Sequence Database. HIV and SIV Nomenclature. Updated 23 September, 2002. (Accessed on 30 December 2005 at http://www.hiv.lanl.gov/content/hiv-db/HelpDocs/subtypes-more.html)

Los Alamos National Laboratory HIV Sequence Database. The Circulating Recombinant Forms (CRFs). Updated 26 January, 2005. (Accessed on 30 December 2005 at http://www.hiv.lanl.gov/content/hiv-db/CRFs/CRFs.html)

Loxton AG, Treurnicht F, Laten A, van Rensburg EJ, Engelbrecht S. Sequence analysis of near full-length HIV type 1 subtype D primary strains isolated in Cape Town, South Africa, from 1984 to 1986. *AIDS Res Hum Retroviruses* 2005; 21: 410-3.

Lyall EG, Stainsby C, Taylor GP, *et al*. Review of uptake of interventions to reduce mother to child transmission of HIV by women aware of their HIV status. *BMJ* 1998; 316: 268-70.

Magoni M, Giuliano M. Authors´ response to ´HIV and infant feeding: a complex issue in resource-limited settings´ by Becquet and Leroy, to the letter to the editors by Coutsoudis *et al*., and to ´Increased risk of infant HIV infection with early mixed feeding´ by Piwoz and Humphrey. *AIDS* 2005; 19: 1720-1.

McGowan JP, Shah SS. Prevention of perinatal HIV transmission during pregnancy. *J Antimicrob Chemother* 2000; 46: 657-68.

McIntyre J, Gray G. What can we do to reduce mother to child transmission of HIV? *BMJ* 2002; 324: 218-21.

Meloni ST, Kim B, Sankale JL, *et al*. Distinct human immunodeficiency virus type 1 subtype A virus circulating in West Africa: sub-subtype A3. *J Virol* 2004; 78: 12438-45.

Mokili JL, Wade CM, Burns SM, *et al*. Genetic heterogeneity of HIV type 1 subtypes in Kimpese, rural Democratic Republic of Congo. *AIDS Res Hum Retroviruses* 1999; 15: 655-64.

Molina RM, Toro AD, Silva MT, Vilela MM, Costa SC. Early diagnosis of HIV-1: infected infants in Brazil using nested-PCR. *J Trop Pediatr* 2004; 50: 107-13.

Moodley D, Moodley J, Coovadia H, *et al*. A multicenter randomized controlled trial of nevirapine versus a combination of zidovudine and lamivudine to reduce intrapartum and early postpartum mother-to-child transmission of human immunodeficiency virus type 1. *J Infect Dis* 2003; 187: 725-35. Epub 2003 Feb 24.

Moodley D, Smith TL, Van Rensburg EJ, Moodley J, Engelbrecht S. HIV type 1 V3 region subtyping in KwaZulu-Natal, a high-seroprevalence South African region. *AIDS Res Hum Retroviruses* 1998; 14: 1015-8.

Mulherin SA, O´Brien TR, Ioannidis JP, *et al*. Effects of CCR5-Delta32 and CCR2-64I alleles on HIV-1 disease progression: the protection varies with duration of infection. *AIDS* 2003; 17: 377-87.

Murray MC, Embree JE, Ramdahin SG, Anzala AO, Njenga S, Plummer FA. Effect of human immunodeficiency virus (HIV) type 1 viral genotype on mother-to-child transmission of HIV-1. *J Infect Dis* 2000; 181: 746-9.

Naarding MA, Ludwig IS, Groot F, *et al*. Lewis X component in human milk binds DC-SIGN and inhibits HIV-1 transfer to CD4+ T lymphocytes. *J Clin Invest* 2005; 115: 3256-64. Epub 2005 Oct 20.

Nduati R, Richardson BA, John G, *et al*. Effect of breastfeeding on mortality among HIV-1 infected women: a randomised trial. *Lancet* 2001; 357: 1651-5.

Newell ML, Borja MC, Peckham C. Height, weight, and growth in children born to mothers with HIV-1 infection in Europe. *Pediatrics* 2003; 111:

Nicoll A, McGarrigle C, Brady AR, *et al*. Epidemiology and detection of HIV-1 among pregnant women in the United Kingdom: results from national surveillance 1988-96. *BMJ* 1998; 316: 253-8.

Nielsen K, Bryson YJ. Diagnosis of HIV infection in children. *Pediatr Clin North Am* 2000; 47: 39-63.

Pancharoen C, Thisyakorn U. Preventive strategies of perinatal HIV-1 transmission: an experience from Thailand. *Expert Opin Pharmacother* 2003; 4: 179-82.

Parekh BS, Hu DJ, Vanichseni S, *et al*. Evaluation of a sensitive/less-sensitive testing algorithm using the 3A11-LS assay for detecting recent HIV seroconversion among individuals with HIV-1 subtype B or E infection in Thailand. *AIDS Res Hum Retroviruses* 2001; 17: 453-8.

Parekh BS, Shaffer N, Pau CP, *et al*. Lack of correlation between maternal antibodies to V3 loop peptides of gp120 and perinatal HIV-1 transmission. The NYC Perinatal HIV Transmission Collaborative Study. *AIDS* 1991; 5: 1179-84.

Peckham C, Gibb D. Mother-to-child transmission of the human immunodeficiency virus. *N Engl J Med* 1995; 333: 298-302.

Perez-Alvarez L, Cuevas MT, Villahermosa ML, *et al*. Prevalence of drug resistance mutations in B, non-B subtypes, and recombinant forms of human immunodeficiency virus type 1 in infected individuals in Spain (Galicia). *J Hum Virol* 2001; 4: 35-8.

Phadke MA, Gadgil B, Bharucha KE, *et al*. Replacement-fed infants born to HIV-infected mothers in India have a high early postpartum rate of hospitalization. *J Nutr* 2003; 133: 3153-7.

Pieniazek D, Rayfield M, Hu DJ, *et al*. Protease sequences from HIV-1 group M subtypes A-H reveal distinct amino acid mutation patterns associated with protease resistance in protease inhibitor-naive individuals worldwide. HIV Variant Working Group. *AIDS* 2000; 14: 1489-95.

Pillay T, Phillips RE. Adaptive evolution in perinatal HIV-1. *Best Pract Res Clin Obstet Gynaecol* 2005; 19: 211-29. Epub 2005 Jan

Posada D, Buckley TR. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Syst Biol* 2004; 53: 793-808.

Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 1998; 14: 817-8.

Puren AJ. The HIV-1 epidemic in South Africa. *Oral Dis* 2002; 8: Suppl 2:27-31

Ramakrishnan R, Hussain M, Holzer A, Mehta R, Sundaravaradan V, Ahmad N. Evaluations of HIV type 1 rev gene diversity and functional domains following perinatal transmission. *AIDS Res Hum Retroviruses* 2005; 21: 1035-45.

Ramalingam S, Kannangai R, Vijayakumar TS, *et al*. Increased number of CCR5+ CD4 T cells among south Indian adults probably associated with the low frequency of X4 phenotype of HIV-1 in India. *Indian J Med Res* 2002; 116: 90-5.

Ras GJ, Simson IW, Anderson R, Prozesky OW, Hamersma T. Acquired immunodeficiency syndrome. A report of 2 South African cases. *S Afr Med J* 1983; 64: 140-2.

Reeves JD, Piefer AJ. Emerging drug targets for antiretroviral therapy. *Drugs* 2005; 65: 1747-66.

Renjifo B, Chung M, Gilbert P, *et al*. In-utero transmission of quasispecies among human immunodeficiency virus type 1 genotypes. *Virology* 2003; 307: 278-82.

Renjifo B, Fawzi W, Mwakagile D, *et al*. Differences in perinatal transmission among human immunodeficiency virus type 1 genotypes. *J Hum Virol* 2001; 4: 16-25.

Roberts JD, Bebenek K, Kunkel TA. The accuracy of reverse transcriptase from HIV-1. *Science* 1988; 242: 1171-3.

Rosso R, Vignolo M, Parodi A, *et al*. Bone quality in perinatally HIV-infected children: role of age, sex, growth, HIV infection, and antiretroviral therapy. *AIDS Res Hum Retroviruses* 2005; 21: 927-32.

Sabbaj S, Edwards BH, Ghosh MK, *et al*. Human immunodeficiency virus-specific CD8(+) T cells in human breast milk. *J Virol* 2002; 76: 7365-73.

Salemi M, de Oliveira T, Soares MA, *et al*. Different epidemic potentials of the HIV-1B and C subtypes. *J Mol Evol* 2005; 60: 598-605.

Saloojee H, Violari A. Regular review: HIV infection in children. *BMJ* 2001; 323: 670-4.

Sanders-Buell E, Salminen M, McCutchan FE. Sequencing primers for HIV-1. In: Human retroviruses and AIDS 1995 (Myers G, Korber B, Hahn BH, Jeang K-T, Mellors JW, McCutchan FE, Henderson LE, Pavlakis GN, editors). Los Alamos, New Mexico, USA. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, 1995; III.

Scarlatti G, Albert J, Rossi P, *et al*. Mother-to-child transmission of human immunodeficiency virus type 1: correlation with neutralizing antibodies against primary isolates. *J Infect Dis* 1993; 168: 207-10.

Scott GB, Hutto C, Makuch RW, *et al*. Survival in children with perinatally acquired human immunodeficiency virus type 1 infection. *N Engl J Med* 1989; 321: 1791-6.

Sei S, Boler AM, Nguyen GT, *et al*. Protective effect of CCR5 delta 32 heterozygosity is restricted by SDF-1 genotype in children with HIV-1 infection. *AIDS* 2001; 15: 1343-52.

Semba RD, Kumwenda N, Hoover DR, *et al*. Human immunodeficiency virus load in breast milk, mastitis, and mother-to-child transmission of human immunodeficiency virus type 1. *J Infect Dis* 1999; 180: 93-8.

Singh KK, Barroga CF, Hughes MD, *et al*. Genetic influence of CCR5, CCR2, and SDF1 variants on human immunodeficiency virus 1 (HIV-1)-related disease progression and neurological impairment, in children with symptomatic HIV-1 infection. *J Infect Dis* 2003; 188: 1461-72. Epub 2003 Nov 18.

Soilleux EJ, Coleman N. Transplacental transmission of HIV: a potential role for HIV binding lectins. *Int J Biochem Cell Biol* 2003; 35: 283-7.

Spira R, Lepage P, Msellati P, *et al*. Natural history of human immunodeficiency virus type 1 infection in children: a five-year prospective study in Rwanda. Mother-to-Child HIV-1 Transmission Study Group. *Pediatrics* 1999; 104:

Swanson P, Devare SG, Hackett J Jr. Molecular characterization of 39 HIV isolates representing group M (subtypes A-G) and group O: sequence analysis of gag p24, pol integrase, and env gp41. *AIDS Res Hum Retroviruses* 2003; 19: 625-9.

Swofford, DL. PAUP*: Phylogenetic analysis using parsimony (* and other methods), version 4.0b 10. Sinauer Associates, Sunderland, Massachusetts, USA, 2003.

Swofford, DL. PAUP: Phylogenetic Analysis Using Parsimony. Illinois Natural History Survey, Champaign, USA, 1993.

Takebe Y, Kusagawa S, Motomura K. Molecular epidemiology of HIV: tracking AIDS pandemic. *Pediatr Int* 2004; 46: 236-44.

Tapia N, Franco S, Puig-Basagoiti F, *et al*. Influence of human immunodeficiency virus type 1 subtype on mother-to-child transmission. *J Gen Virol* 2003; 84: 607-13.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997; 25: 4876-82.

Thomson MM, Perez-Alvarez L, Najera R. Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. *Lancet Infect Dis* 2002; 2: 461-71.

Tscherning C, Alaeus A, Fredriksson R, *et al*. Differences in chemokine coreceptor usage between genetic subtypes of HIV-1. *Virology* 1998; 241: 181-8.

Van de Peer Y, De Wachter R. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput Appl Biosci* 1994; 10: 569-70.

Van de Peer Y, De Wachter R. TREECON: a software package for the construction and drawing of evolutionary trees. *Comput Appl Biosci* 1993; 9: 177-82.

van Harmelen J, Wood R, Lambrick M, Rybicki EP, Williamson AL, Williamson C. An association between HIV-1 subtypes and mode of transmission in Cape Town, South Africa. *AIDS* 1997; 11: 81-7.

Verhofstede C, Demecheleer E, De Cabooter N, *et al*. Diversity of the human immunodeficiency virus type 1 (HIV-1) env sequence after vertical transmission in mother-child pairs infected with HIV-1 subtype A. *J Virol* 2003; 77: 3050-7.

Wagner R, Shao Y, Wolf H. Correlates of protection, antigen delivery and molecular epidemiology: basics for designing an HIV vaccine. *Vaccine* 1999; 17: 1706-10.

Wasik TJ, Bratosiewicz J, Wierzbicki A, *et al*. Protective role of beta-chemokines associated with HIV-specific Th responses against perinatal HIV transmission. *J Immunol* 1999; 162: 4355-64.

Weiss RA. HIV and AIDS: looking ahead. *Nat Med* 2003; 9: 887-91.

World Health Organization (WHO) 2005a. AIDS epidemic update: December 2005. (Accessed on 2 January 2006 at http://www.who.int/hiv/epi-update2005_en.pdf)

World Health Organization (WHO), UNAIDS 2005b. "3 by 5" Progress report - South Africa - Country Profile for HIV/AIDS Treatment Scale-up, June 2005. (Accessed on 21 November 2005 at http://www.who.int/entity/3by5/support/june2005_zaf.pdf)

Yedavalli VR, Chappey C, Ahmad N. Maintenance of an intact human immunodeficiency virus type 1 vpr gene following mother-to-infant transmission. *J Virol* 1998a; 72: 6937-43.

Yedavalli VR, Chappey C, Matala E, Ahmad N. Conservation of an intact vif gene of human immunodeficiency virus type 1 during maternal-fetal transmission. *J Virol* 1998b; 72: 1092-102.

Appendix A

Letter from the Committee for Human Research of the Faculty of Health Sciences, University of Stellenbosch approving this study.

UNIVERSITEIT·STELLENBOSCH·UNIVERSITY
jou kennisvennoot • your knowledge partner

27 January 2004

Dr SNJ Korsman
Dept of Medical Virology

Dear Dr Korsman

RESEARCH PROJECT: "MOLECULAR EPIDEMIOLOGY OF MTCT OF HIV-1 IN CHILDREN AT TYGERBERG HOSPITAL WHO TESTED POSITIVE ON HIV-1 *GAG* PCR IN 2000, 2001 AND 2002"
PROJECT NUMBER : N04/01/002

At a meeting of the Committee for Human Research that was held on 17 November 2003 the above project was approved on condition that further information that was required, be submitted.

This information was supplied and the project was finally approved on 20 January 2004. This project is therefore now registered and you can proceed with the work. Please quote the above-mentioned project number in all further correspondence.

Patients participating in a research project in Tygerberg Hospital will not be treated free of charge as the Provincial Administration of the Western Cape does not support research financially.

Due to heavy workload the nursing corps of the Tygerberg Hospital cannot offer comprehensive nursing care in research projects. It may therefore be expected of a research worker to arrange for private nursing care.

Yours faithfully

CJ VAN TONDER
RESEARCH DEVELOPMENT AND SUPPORT (TYGERBERG)

CJVT/ev

C:\DOCUMENTS AND SETTINGS\EVISAGIE.000\MY DOCUMENTS\KMN\PROJEKTE\2004\N04-01-002-001.DOC

Fakulteit Gesondheidswetenskappe • Faculty of Health Sciences

Verbind tot Optimale Gesondheid • Committed to Optimal Health
Afdeling Navorsingsontwikkeling en -steun • Division of Research Development and Support
Posbus/PO Box 19063 • Tygerberg 7505 • Suid-Afrika/South Africa
Tel: +27 21 938 9207 • Faks/Fax: +27 21 933 6330
E-pos/E-mail: cjvt@sun.ac.za

Appendix B

Reference sequence	Genbank accession number
A1.KE.94.Q23_17	AF004885
A1.SE.94.SE7253	AF069670
A1.UG.85.U455	M62320
A1.UG.92.92UG037	U51190
A2.CD.97.97CDKS10	AF286241
A2.CD.97.97CDKTB48	AF286238
A2.CY.94.94CY017_41	AF286237
B.FR.83.HXB2	K03455
B.US.83.RF	M17451
B.US.86.JRFL	U63632
B.US.90.WEAU160	U21135
C.BR.92.92BR025	U52953
C.BW.96.96BW0502	AF110967
C.ET.86.ETH2220	U46016
C.IN.95.95IN21068	AF067155
D.CD.83.ELI	K03454
D.CD.83.NDK	M27323
D.CD.84.84ZR085	U88822
D.UG.94.94UG114	U88824
F1.BE.93.VI850	AF077336
F1.BR.93.93BR020_1	AF005494
F1.FI.93.FIN9363	AF075703
F1.FR.96.MP411	AJ249238
F2.CM.95.MP255	AJ249236
F2.CM.95.MP257	AJ249237
G.BE.96.DRCBL	AF084936
G.FI.93.HH8793_12_1	AF061641
G.NG.92.92NG083	U88826
G.SE.93.SE6165	AF061642
H.BE.93.VI991	AF190127
H.BE.93.VI997	AF190128
H.CF.90.056	AF005496
J.SE.93.SE7887	AF082394
J.SE.94.SE7022	AF082395
K.CD.97.EQTB11C	AJ249235
K.CM.96.MP535	AJ249239
O.BE.87.ANT70	L20587

Appendix C

The final PAUP* script used for obtaining PAUP*'s suggested values was:

```
begin paup;
outgroup O.BE.87.ANT70;
set autoclose=yes;
hsearch addseq=random rseed=5 multrees=no;
set criterion=distance;
dset distance=f84;
nj;
lset nst=6;
lset rmatrix=estimate basefreq=estimate;
lset shape=estimate showqmatrix=yes rates=gamma;
lscore;
lset shape=previous rmatrix=previous basefreq=previous;
set criterion=distance;
dset distance=gtr;
lset nst=2 tratio=estimate;
lscore;
lset tratio=previous nst=6;
nj;
bootstrap search=nj nreps=1000 brlens=yes;
describe /brlens=yes;
set criterion=likelihood;
hsearch start=nj;
showtrees;
describe /brlens=yes;
lset zerolentest=full;
showtrees;
describe /brlens=yes;
savetree brlens=yes savebootp=both file=7msn.txt format=altnexus;
savetree brlens=yes savebootp=both file=7msp.txt format=phylip;
end;
```

Appendix D

The final PAUP* script for using the suggested values from FindModel was:

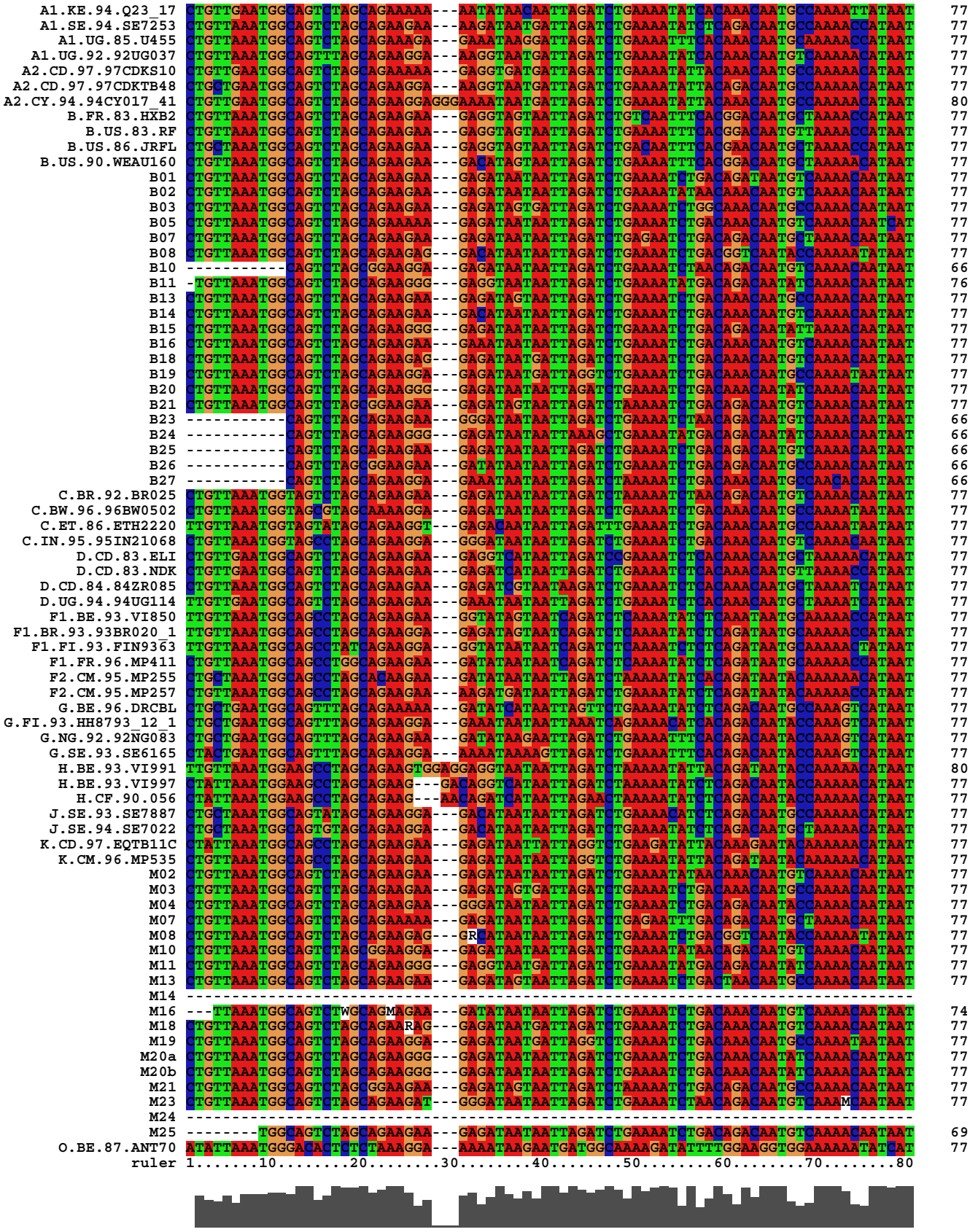
```
begin paup;
outgroup O.BE.87.ANT70;
set autoclose=yes;
hsearch addseq=random rseed=5 multrees=no;
set criterion=likelihood;
dset distance=gtr;
nj;
lset tratio=1.9048;
lset rmatrix=(0.71389 0.124173 0.159516 0.169747 0.091130);
lset basefreq=(0.4254 0.2262 0.1878);
lset shape=0.65026;
lset showqmatrix=yes rates=gamma;
lset nst=6;
nj;
set criterion=distance;
bootstrap search=nj nreps=1000 brlens=yes;
describe /brlens=yes;
set criterion=likelihood;
hsearch start=nj;
showtrees;
describe /brlens=yes;
lset zerolentest=full;
showtrees;
describe /brlens=yes;
savetree brlens=yes savebootp=both file=4finlanln.txt format=altnexus;
savetree brlens=yes savebootp=both file=4finlanlp.txt format=phylip;
end;
```

Appendix E

Nucleotide alignment – original alignment created by CLUSTALX.

CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

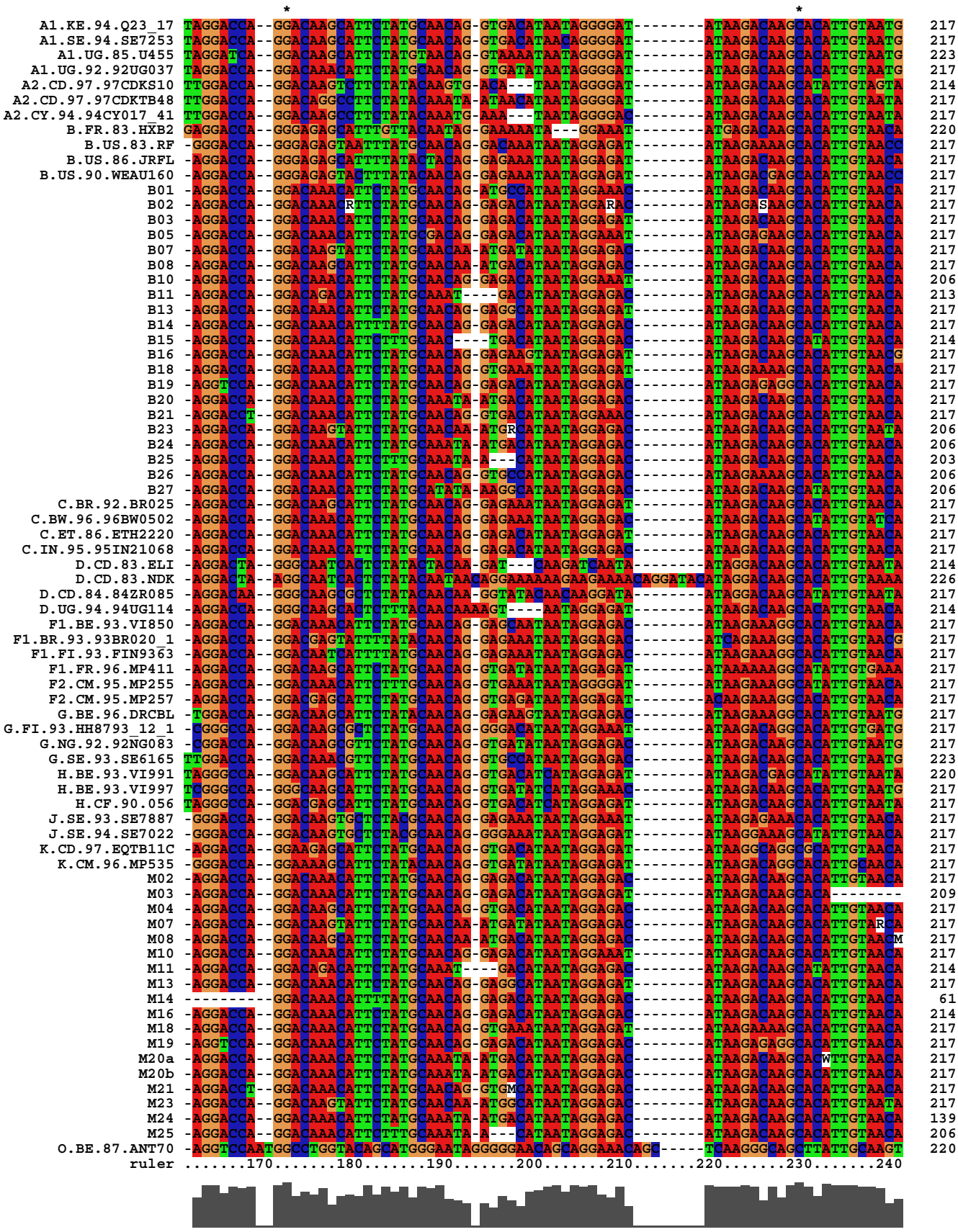
File: Nucleotide alignment in CLUSTAL X



CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

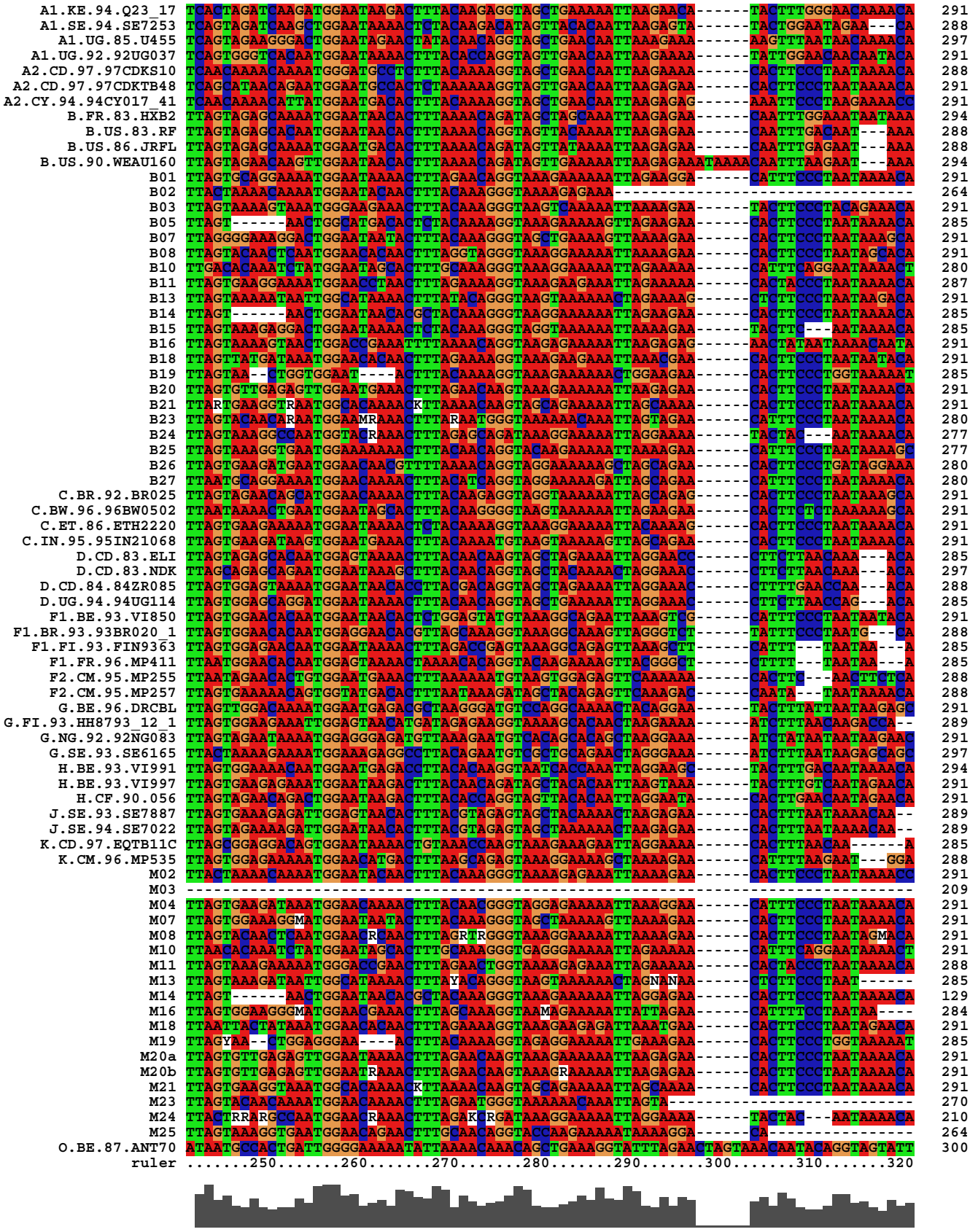
File: Nucleotide alignment in CLUSTAL X


```
     A1.KE.94.Q23_17 AGTACAACTTGTCCAACCTGTGACAATTAAGTGTATCAGACCTAACAACAATACAAGAAAAAGTATACGTA---------   148
     A1.SE.94.SE7253 AGTACAACTTACCGAGCCTGTAACAATTAATTGTACCAGACCTAGCAACAATACAAGAACAAGTATACGTA---------   148
       A1.UG.85.U455 AGTACAGCTTGTCAATCCTGTAAAAATCAATTGTTCCAGACCTTACAA---TACAAGAAAAAATATACGTAGGTATAGTA   154
    A1.UG.92.92UG037 AGTACAACTTAACGAGTCTGTAACAATTAATTGTACCAGACCTAACAATAATACAAGAAGAAGTGTACGTA---------   148
   A2.CD.97.97CDKS10 AGTACAGTTTAATGAATCGGTACCAATTACTTGTATCAGACCCAACAACAATACGAGAAAAGGTATACCTA---------   148
   A2.CD.97.97CDKTB48 AGTACAGTTTAATAAACCTGTACCAATTAATTGTACCAGACCCAACAACAATACAAGAAAAAGTATACGCT---------   148
  A2.CY.94.94CY017_41 AGTTCAGTTTACTAAGCCTGTACTAATTACTTGTATCAGACCCAACAACAATACAAGAAAAAGTATACGCT---------   151
       B.FR.83.HXB2 AGTACAGCTGAACACATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAGAAAAAGAATCCGTATCCAGA---   154
         B.US.83.RF AGTACAGCTGAATGCATCTGTACAAATTAATTGTACAAGACCCAACAACAATACAAGAAAAAGTATAACTAA--------   149
       B.US.86.JRFL AGTACAGCTGAAAGAATCTGTAGAAATTAATTGTACAAGACCCAACAACAATACAAGAAAAAGTATACATAT--------   149
    B.US.90.WEAU160 AGTACAGCTGAATGTATCCATAGAAATTAATTGTACAAGACCCAACAACAATACAAGAAAAAAAATAACTTT--------   149
                B01 AGTACATCTTAATGAATCTGTAGAGATTAATTGTACAAGACCCAACAATAATACAAGGAAAAGTGTAAGGAT--------   149
                B02 AGTACATCTTAATGAATCTGTAGAAATTAATTGTACAAGACCCAACAATAATACAAGAAAAAGTATAAGGAT--------   149
                B03 AGTACATCTCAATGAATCTGTAGAGATAGAGTGTATAAGACCTGGCAATAATACAAGAACAAGTACGAGGAT--------   149
                B05 AGTACACCTCAATGAATCTGTAGAAATTGTGTGTACAAGGCCCAACAATAATACAAGAAAAAGTGTGAGGAT--------   149
                B07 AGTACATCTTAATGAGTCCATAGAAATTGTRTGCACAAGACCTAACAATAATACAAGAAAAAGTATAAGGAT--------   149
                B08 AGTACACCTTAATGACCCTATAGAAATTGTGTGTACAAGACCCAACAATAATACAAGAAGAAGTATAAGGAT--------   149
                B10 AGTACAGCTTAATGATCCTGTAGAGATTATATGTACAAGACCCAACAATAATACAAGAAAAAGTGTGAGAAT--------   138
                B11 AGTACATCTTAATGAATCTGTAGAAATTATGTGTACAAGACCCGGCAATAATACAAGAAAAAGTATAAGGAT--------   148
                B13 AGTGCATCTCAATGAATCTGTAGAGATTAATTGTACAAGACCAAACAATAATACAAGGAAAAGTGTGAGAAT--------   149
                B14 AGTACATCTTAATGAATCTGTAGAGATAATATGTATAAGACCTAACAATAATACAAGAAAAAGTGTAAGGAT--------   149
                B15 AGTACACCTTAATGAATCTATAAACATTGAGTGTACAAGACCCAACAATAATACAAGAAAAAGCGTAAGGAT--------   149
                B16 AGTACATCTTAATAAAGCTGTAAAGATTGTGTGTGTAAGACCTGGCAATAATACAAGGAAAAGTGTAAGGAT--------   149
                B18 AGTACACCTTAATGAGTCTATAGCAATTGTGTGTATAAGACTCGGCAATAACACAAGAAAAAGTATAAGGAT--------   149
                B19 AGTACATCTTAATGAATCTGYAGAAATTGTGTGTGCAAGACCCAACAATAATACAAGAAMAAGTATAAGGAT--------   149
                B20 AGTCCACCTTAATAAATCTGTAGAAATTATGTGTATAAGACCCAACAATAATACAAGAAAAAGTATAAGGAT--------   149
                B21 AGTACACCTTAATGAGTCTGTAGAAATTGTGTGCATAAGACCCGGCAATAATACAAGGAAAAGTGTAAGGAT--------   149
                B23 AGTACATTTTAATRAATCTGTACAGATTACTTGTACAAGACCCAACAACAATACAAGGAAAAGTGTAAGGAT--------   138
                B24 AGTCCATCTTAATGAATCTGTACAAATTGTGTGTTCAAGACCCAACAATAATACAAGAAAAGGTATAAGGAT--------   138
                B25 AGTACATCTTAATGAATCTGTAGAGATTGAGTGTACAAGACCCAACAATAATACAAGAAAAGGTATAGGGAT--------   138
                B26 AGTACACTTTAATGAGTCTGTAGACATTGTGTGCATAAGACCAGGCAATAACACAAGAAAAAGTGTAAGGAT--------   138
                B27 AGTACATCTTAATGAATCTGTAGAAATTATGTGTACAAGACCCAACAATAATACAAGGGAAAGTGTAAGGAT--------   138
       C.BR.92.BR025 AGTACATCTTAACGAATCTGTAGAGATTAATTGTACGAGACCCAACAATAATACAAGAAAAAGTATAAGGAT--------   149
    C.BW.96.96BW0502 AGTACAACTTAATAAACCTGTAAAAATTGTGTGTGTAAGGCCTAACAATAATACAAGAAAAAGTGTAAGGAT--------   149
     C.ET.86.ETH2220 AGTACAGCTTAATGAATCTGTAGAAATTACTTGTACGAGACCCAGCAATAATACAAGAGAAAGTATAAGGAT--------   149
  C.IN.95.95IN21068 AGTACATCTTAATCAACCTGTAGAAATTATGTGTACAAGACCCGACAATAATACAAGAAAAAGTATAAGGAT--------   149
         D.CD.83.ELI AGCACATCTTAATGAATCTGTAAAAATTACCTGTGCAAGGCCCTATCAAAATACAAGACAAAGAACACCTAT--------   149
         D.CD.83.NDK AGTACAGCTTAATGCATCTATAGTAATTAATTGTACAAGGCCCTACAAATATACAAGGCAAAGGACATCGAT--------   149
    D.CD.84.84ZR085 AGTACATCTTAATCAATCTGTAGAAATTAATTGTACAAGGCCCTACAAAAAGGAAAGACAAAGGACACCTAT--------   149
    D.UG.94.94UG114 AGTACAGCTTAATGAGTCTGTACCAATTAATTGCATAAGGCCCTATAACAATACAAGACAAAGTACACGTAT--------   149
      F1.BE.93.VI850 AGTACACCTTAATGAATCTGTACAGATTAATTGTACAAGACCCAACAACAATACAAGGAAAGGTATACATTT--------   149
 F1.BR.93.93BR020_1 AGTGCACCTTAATGAATCTGTACAGATTAATTGTACAAGACCCAACAACAATACAAGAAAAAGAATATCTTT--------   149
   F1.FI.93.FIN9363 AGTACACCTTAATGAATCTGTACAGATCAATTGTACAAGACCCAACAACAATACAAGAAAAAGTATACGTAT--------   149
     F1.FR.96.MP411 AGTACACCTTAATGAAAGTGTACAGATTAATTGTACAAGACCCAACAACAATACAAGAAAAAGTATACATTT--------   149
     F2.CM.95.MP255 AGTACAGTTTAATAGATCTGTAATAATTGATTGTAGAAGACCCAACAACAATACAAGAAAAGGGATACGTAT--------   149
     F2.CM.95.MP257 AGTACAGTTTAAAAATCCTGTAAAAATTAATTGTACAAGACCCAACAACAATACAAGAAGAAGTATACATAT--------   149
      G.BE.96.DRCBL AGTGCACCTTAATAGGTCTGTAGAAATTAATTGTACCAGACCCAACAACAATACAAGAAGAAGTGTAGCAAT--------   149
  G.FI.93.HH8793_12_1 AGTGCAGCTTAATGAAACTGTAGAAATTACGTGTGTCAGACCCAACAACAATACAAGAAAAAGTATACACCT--------   149
    G.NG.92.92NG083 AGTGCAGCTTAATAATAGTATAGAAATTAATTGTATCAGACCCAATAACAATACAAGAAAAAGTATACCAAT--------   149
     G.SE.93.SE6165 AGTACAGCTTAATAAAACTGTAGAAATTAATTGTACCAGACCCAACAACAATACAATGAAAAGGATAAGAATGGGAA---   154
      H.BE.93.VI991 AGTACAGTTAAATGAACCTGTACAAATTAACTGTACCAGAACAGGCAATAATACGAGAAAAAGTATACGTA---------   151
      H.BE.93.VI997 AGTACAGCTTGATAGTCCTATAGAAATTACCTGTACCAGACCTAACAATAATACAAGAAAAGGTATACATT---------   148
       H.CF.90.056 AGTACAGCTTAAGACACCAGTAAACATTACATGTACCAGGCCTAACAATAATACGAGAACAAGTATACATT---------   148
     J.SE.93.SE7887 AGTACAACTTAATAAAACTGTAGAAATTGTGTGTTACAGACCTAATAACAATACAAGGAAAGGTATACACAT--------   149
     J.SE.94.SE7022 AGTACAACTTAATGACACTGTAGAAATTGTGTGTACCAGACCTAATAACAATACAAGAAAAGGTATACACAT--------   149
   K.CD.97.EQTB11C AGTACAGCTTAATGAAGCTGTAGAAATTAATTGTACAAGGCCAAGCAACAATACAAGAAAAAGTATACATAT--------   149
     K.CM.96.MP535 AGTACAGCTTAATGAAACTGTACAAATTAATTGTACAAGGCCAAACAACAATACAAGAAAAAGTATACATAT--------   149
                M02 AGTACATCTTAATGAATCTGTAGAAATTAATTGTACAAGACCCAACAATAATACAAGAAAAGGTATAAGGAT--------   149
                M03 AGTACATCTCAATGAATCTGTAGARATTGAGTGTATAAGACCTGGCAATAATACAAGAACAAGTACGAGGAT--------   149
                M04 AGTACACCTTAATGAATCTGTAGAGATTAAGTGTACAAGACCCAACAATAATACAAGGAAAAGTATAAGGAT--------   149
                M07 AGTACATCTTAATGAGTCCATAGAAATTGTGTGCACAAGACCTAGCAATAATACAAGAAAAAGTATAAGGAT--------   149
                M08 AGTACACCTTAATGACCCTATAGAAATTGTGTGTACAAGACCCAACAATAATACAAGAAGAAGTATAAGGAT--------   149
                M10 AGTACAGCTTAATGATTCTGTAGAGATTATATGTACAAGACCCAACAATAATACAAGAAAAAGTGTGAGAAT--------   149
                M11 AGTACATCTTAATGAATCTGTAGAAATTATGTGTACAAGACCCGGCAATAATACAAGAAAAAGTATAAGGAT--------   149
                M13 AGTGCATCTCAATGAATCTGTAGAGATTAATTGTACAAGACCAAACAATAATACAAGGAAAAGTGTGAGAAT--------   149
                M14 --------------------------------------------------------------------------------
                M16 AGTACATCTTAAARACCCTGTAGAGATTGTGTGTACAAGACCTGGCAATAATACAAGAAAAAGTGTAAGGAT--------   146
                M18 AGTACACCTTAATGAGTCTATAGCAATTGTGTGTACAAGACTCGGCAATAACACAAGAAAAAGTATAAGGAT--------   149
                M19 AGTACATCTTAATGAATCTGCAGRAATTGTGTGTGCAAGACCCAACAATAATACAAGAACAAGTATAAGGAT--------   149
               M20a AGTCCACCTTAATGAATCTGTAGAAATTGTGTGTATAAGACCCAGCAATAAYACAAGAAAAAGTATAAGGAT--------   149
               M20b AGTCCACCTTAATRAATCTGTAGAAATTGTGTGTATAAGACCCARCAATAAYACAAGAAAAAGTATAAGGAT--------   149
                M21 AGTACACCTTAATGAGTCTGTAGAAATTGTGTGCATAAGAMCCGGCAATAATACAAGGRAAAGTGTAAGGAT--------   149
                M23 AGTMCATTTTAATGAATCTGTACAGATTACTTGTACAAGACCCAACAACAATMCAAGGAAAAGTGTAAGGAT--------   149
                M24 -GTCCATCTTAATGAATCTGTACAAATTGTGTGTTCAAGACCCAACAATAATACAAGAAAAGGTATAAGGAT--------    71
                M25 AGTACATCTTAATAAATCTGTAGAGATTGWGTGTACAAGACCCAACAATAATACAAGAAAAGGTATAGGGAT--------   141
       O.BE.87.ANT70 AGTGACCCTAAACTCTACCCTAAACATGACCTGTGAAAGACC--ACAAATAGACATACAAGAG-ATGAGAAT--------   146
              ruler .........90.......100.......110.......120.......130.......140.......150.......160
```


A1.KE.94.Q23_17 217
A1.SE.94.SE7253 217
A1.UG.85.U455 223
A1.UG.92.92UG037 217
A2.CD.97.97CDKS10 214
A2.CD.97.97CDKTB48 217
A2.CY.94.94CY017_41 217
B.FR.83.HXB2 220
B.US.83.RF 217
B.US.86.JRFL 217
B.US.90.WEAU160 217
B01 217
B02 217
B03 217
B05 217
B07 217
B08 217
B10 206
B11 213
B13 217
B14 217
B15 214
B16 217
B18 217
B19 217
B20 217
B21 217
B23 206
B24 206
B25 203
B26 206
B27 206
C.BR.92.BR025 217
C.BW.96.96BW0502 217
C.ET.86.ETH2220 217
C.IN.95.95IN21068 217
D.CD.83.ELI 214
D.CD.83.NDK 226
D.CD.84.84ZR085 217
D.UG.94.94UG114 214
F1.BE.93.VI850 217
F1.BR.93.93BR020_1 217
F1.FI.93.FIN9363 217
F1.FR.96.MP411 217
F2.CM.95.MP255 217
F2.CM.95.MP257 217
G.BE.96.DRCBL 217
G.FI.93.HH8793_12_1 217
G.NG.92.92NG083 217
G.SE.93.SE6165 223
H.BE.93.VI991 220
H.BE.93.VI997 217
H.CF.90.056 217
J.SE.93.SE7887 217
J.SE.94.SE7022 217
K.CD.97.EQTB11C 217
K.CM.96.MP535 217
M02 217
M03 209
M04 217
M07 217
M08 217
M10 217
M11 214
M13 217
M14 61
M16 214
M18 217
M19 217
M20a 217
M20b 217
M21 217
M23 217
M24 139
M25 206
O.BE.87.ANT70 220
ruler .....170.....180.....190.....200.....210.....220.....230.....240

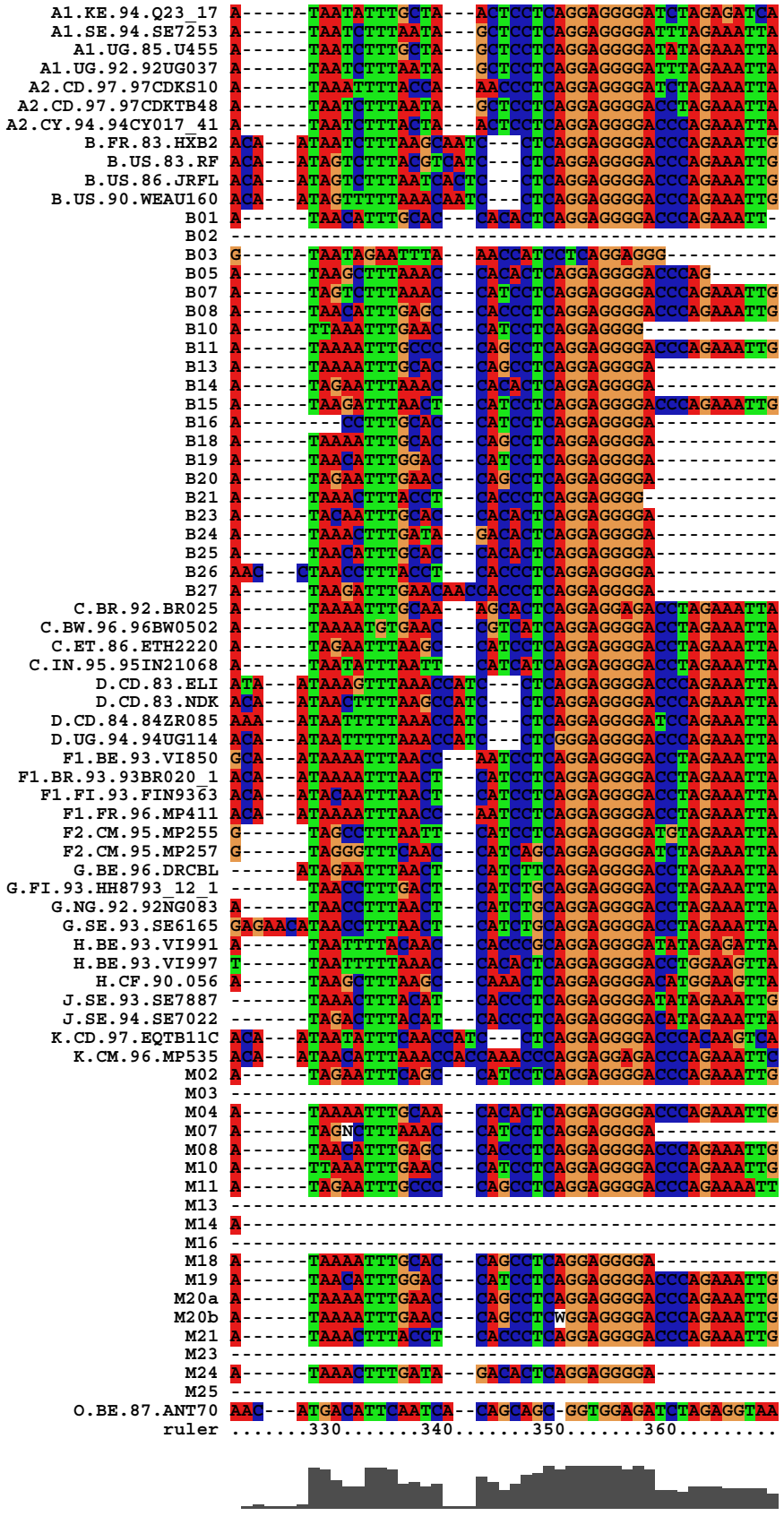
CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

File: Nucleotide alignment in CLUSTAL X



CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

File: Nucleotide alignment in CLUSTAL X

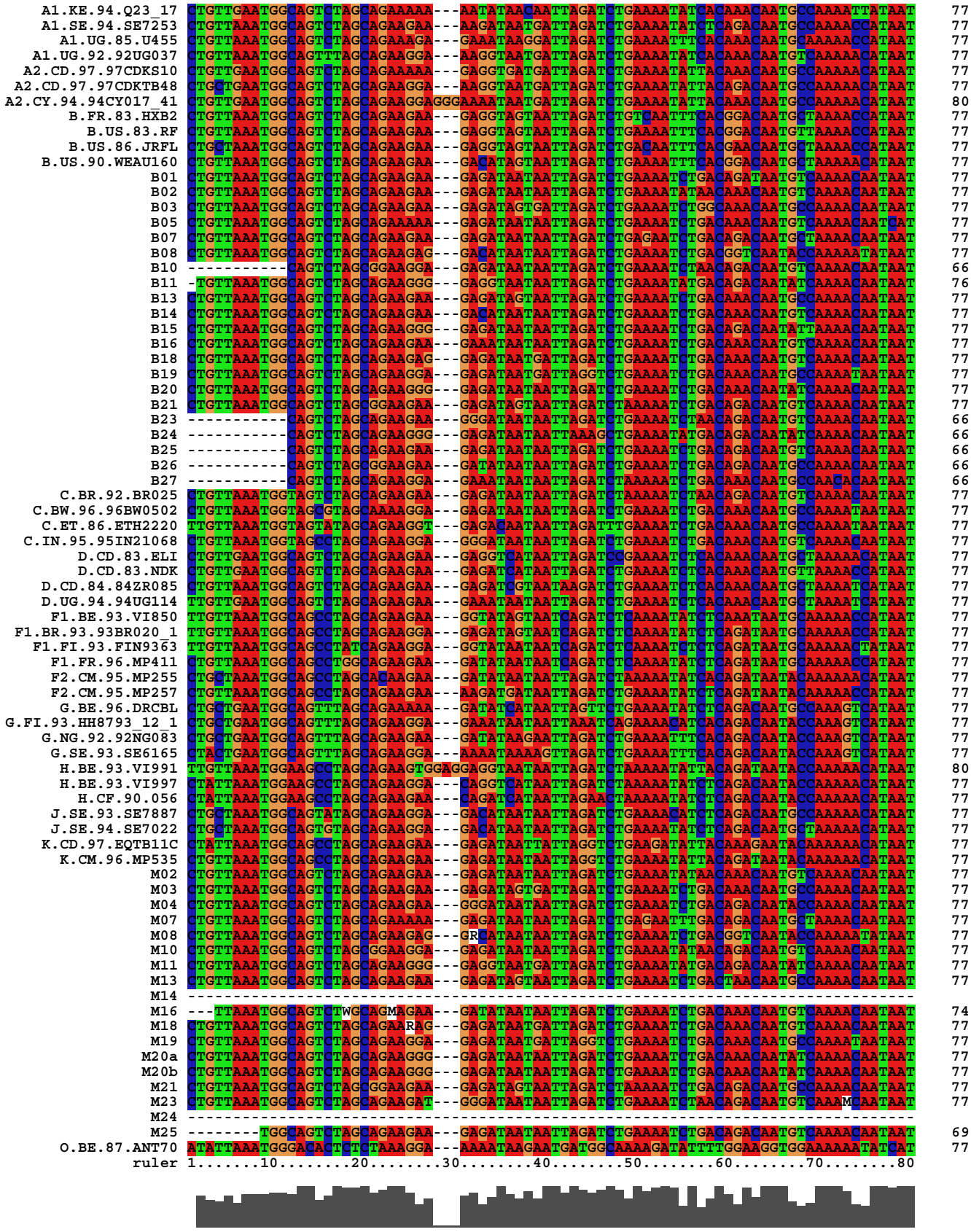


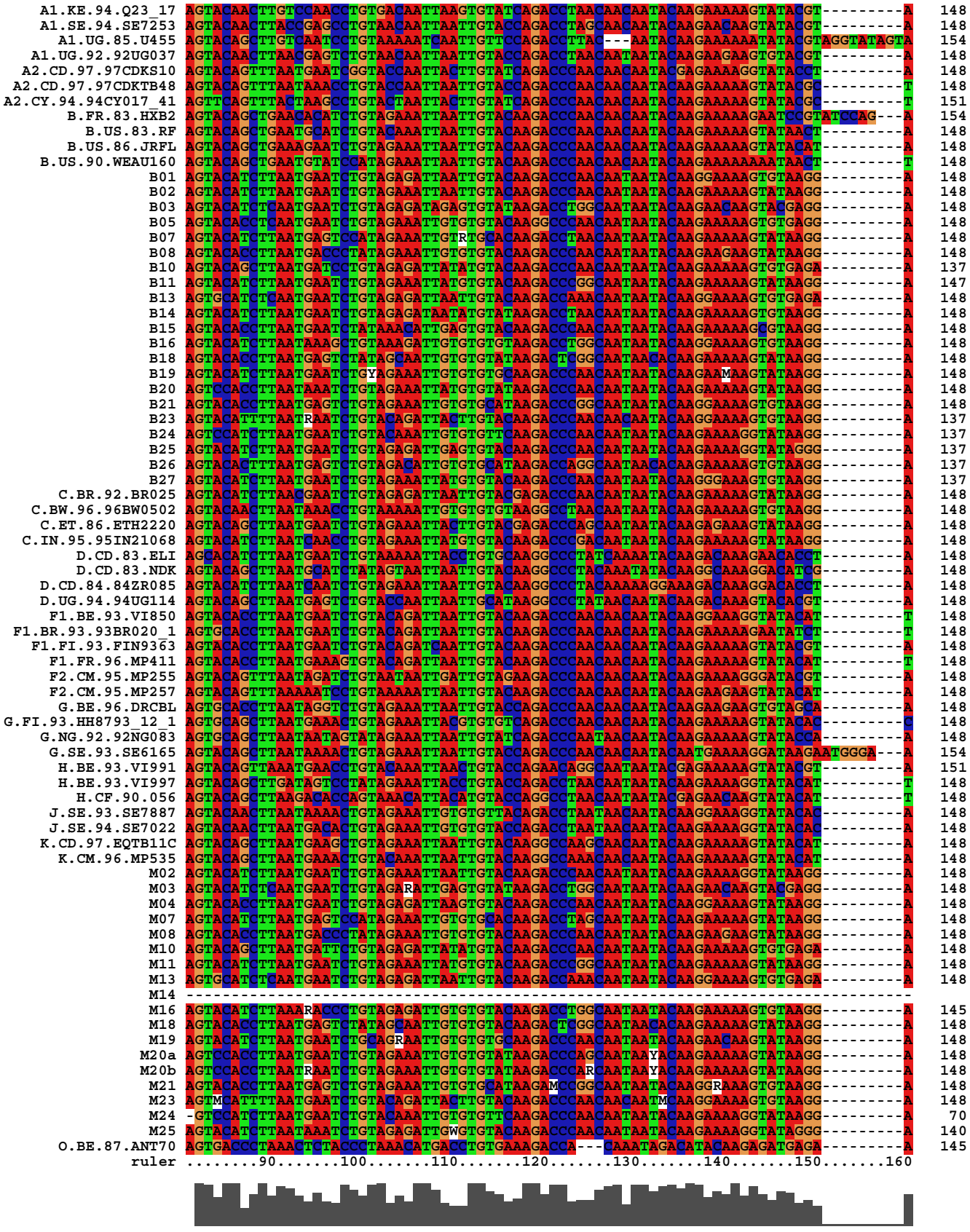
Appendix F

Nucleotide alignment – codon alignment of all sequences obtained.

CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

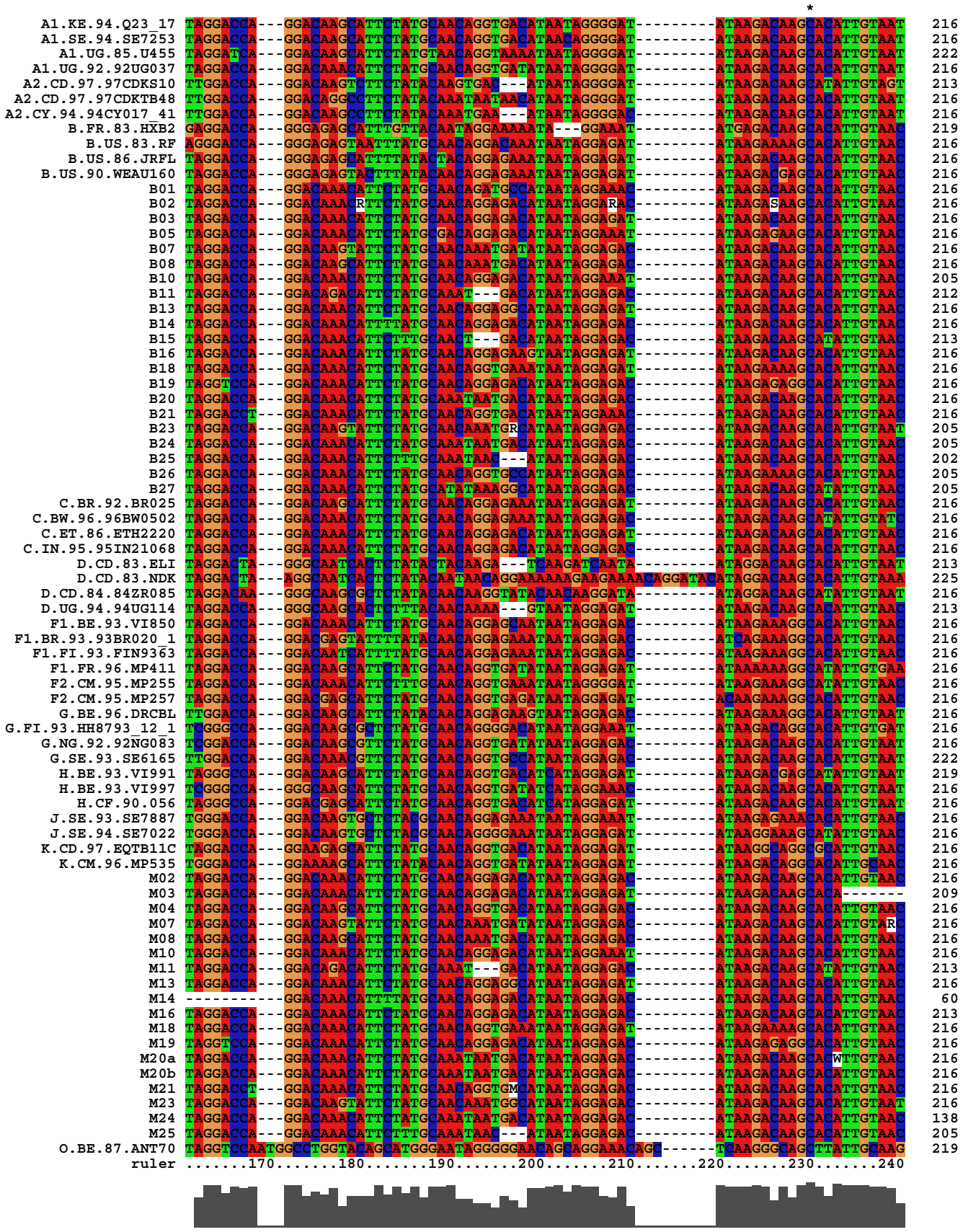
File: Codon alignment - all sequences




A1.KE.94.Q23_17
A1.SE.94.SE7253
A1.UG.85.U455
A1.UG.92.92UG037
A2.CD.97.97CDKS10
A2.CD.97.97CDKTB48
A2.CY.94.94CY017_41
B.FR.83.HXB2
B.US.83.RF
B.US.86.JRFL
B.US.90.WEAU160
B01
B02
B03
B05
B07
B08
B10
B11
B13
B14
B15
B16
B18
B19
B20
B21
B23
B24
B25
B26
B27
C.BR.92.BR025
C.BW.96.96BW0502
C.ET.86.ETH2220
C.IN.95.95IN21068
D.CD.83.ELI
D.CD.83.NDK
D.CD.84.84ZR085
D.UG.94.94UG114
F1.BE.93.VI850
F1.BR.93.93BR020_1
F1.FI.93.FIN9363
F1.FR.96.MP411
F2.CM.95.MP255
F2.CM.95.MP257
G.BE.96.DRCBL
G.FI.93.HH8793_12_1
G.NG.92.92NG083
G.SE.93.SE6165
H.BE.93.VI991
H.BE.93.VI997
H.CF.90.056
J.SE.93.SE7887
J.SE.94.SE7022
K.CD.97.EQTB11C
K.CM.96.MP535
M02
M03
M04
M07
M08
M10
M11
M13
M14
M16
M18
M19
M20a
M20b
M21
M23
M24
M25
O.BE.87.ANT70
ruler
.....90.....100.....110.....120.....130.....140.....150.....160

CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

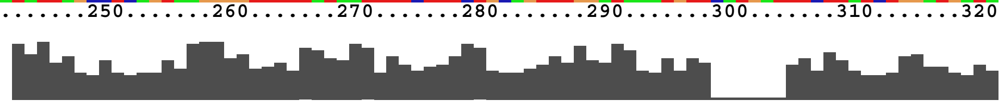
File: Codon alignment - all sequences
Page 3 of 5



CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

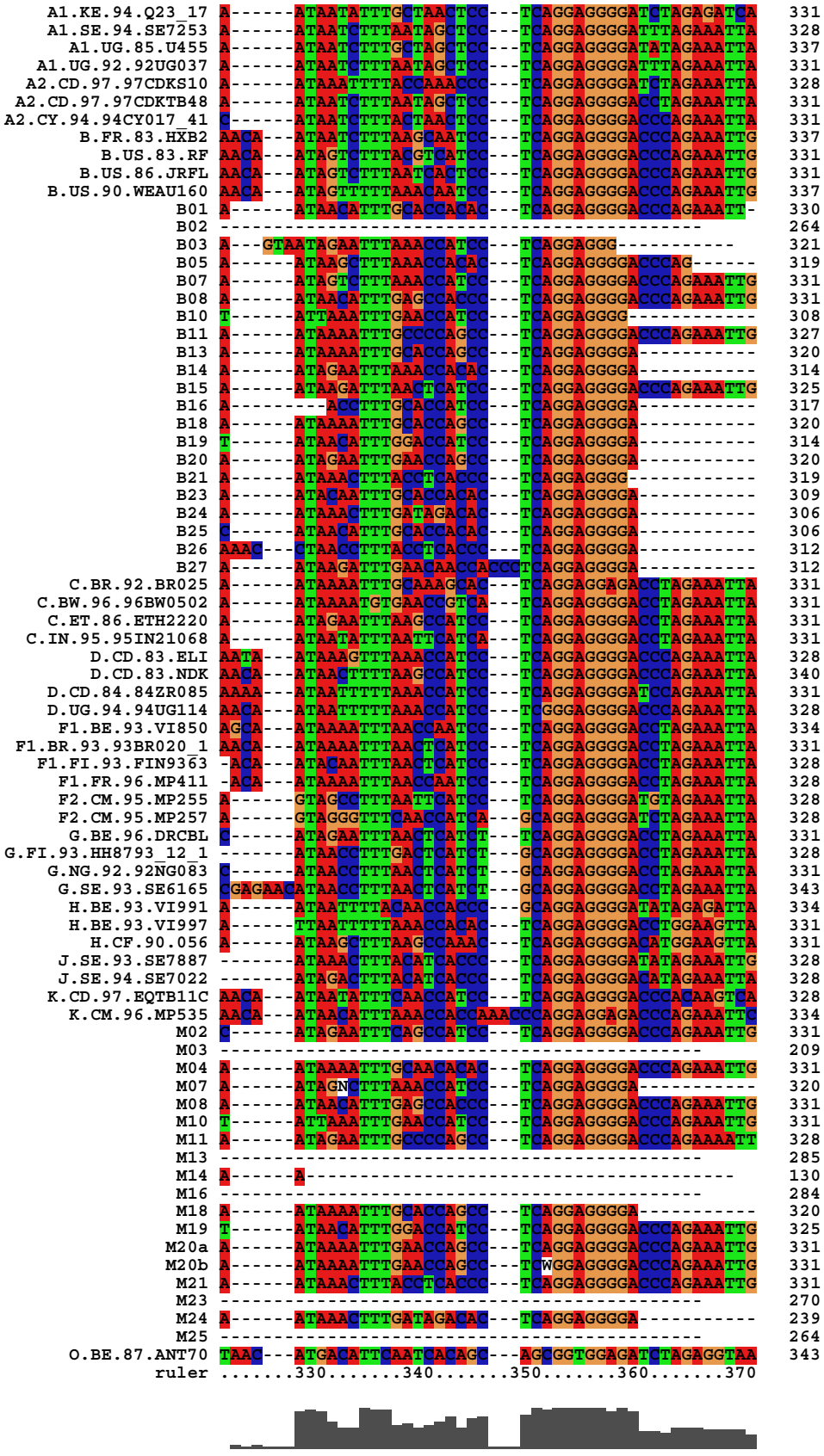
File: Codon alignment - all sequences


```
      A1.KE.94.Q23_17 GTCACTAGATCAAGATGGAATAAGACTTTACAAGAGGTAGCTGAAAAATTAAGAACA------TACTTTGGGAACAAAAC   290
      A1.SE.94.SE7253 GTCAGTAGATCAAGCTGGAATAAAACTCTACAAGACATAGTTACACAATTAAGAGTA------TACTGGAATAGA---AC   287
       A1.UG.85.U455 GTCAGTAGAAGGGACTGGAATAGAACTATACAACAGGTAGCTGAACAATTAAAGAAA------AAGTTTAATAACAAAAC   296
     A1.UG.92.92UG037 GTCAGTGGGTCACAATGGAATAAAACTTTACACCAGGTAGTTGAACAATTAAGAAAA------TATTGGAACAACAATAC   290
    A2.CD.97.97CDKS10 ATCAACAAAACAAAATGGGATGCCTCTTTACAAAAGGTAGCTGAACAATTAAGAAAA------CACTTCCCTAATAAAAC   287
    A2.CD.97.97CDKTB48 ATCAGCATAACAGAATGGAATGCCACTCTAAAAAAGGTAGTTGAACAATTAAGAGAA------CACTTCCCTAATAAAAC   290
  A2.CY.94.94CY017_41 ATCAACAAAACATTATGGAATGACACTTTACAAAAGGTAGCTGAACAATTAAGAGAG------AAATTCCCTAAGAAAAC   290
        B.FR.83.HXB2 ATTAGTAGAGCAAAATGGAATAACACTTTAAAACAGATAGCTAGCAAATTAAGAGAA------CAATTTGGAAATAATAA   293
          B.US.83.RF CTTAGTAGAGCACAATGGAATAACACTTTAAAACAGGTAGTTACAAAATTAAGAGAA------CAATTTGACAAT---AA   287
        B.US.86.JRFL ATTAGTAGAGCAAAATGGAATGACACTTTAAAACAGATAGTTATAAAATTAAGAGAA------CAATTTGAGAAT---AA   287
      B.US.90.WEAU160 CTTAGTAGAACAAGTTGGAATAACACTTTAAAACAGATAGTTGAAAAATTAAGAGAAATAAAACAATTTAAGAAT---AA   293
                  B01 ATTAGTGCAGGAAAATGGAATAAAACTTTAGAACAGGTAAAGAAAAAATTAGAAGGA------CATTTCCCTAATAAAAC   290
                  B02 ATTACTAAAACAAAATGGAATACAACTTTACAAAGGGTAAAAGAGAAA-------------------------------   264
                  B03 ATTAGTAAAAGTAAATGGGAAGAAACTTTACAAAGGGTAAGTCAAAAATTAAAAGAA------TACTTCCCTACAGAAAC   290
                  B05 ATTAGT------AACTGGCATGACACTCTACAAAAGGTAAAGAAAAAGTTAGAAGAA------CACTTCCCTAATAAAAC   284
                  B07 ATTAGGGGAAAGGACTGGAATAATACTTTACAAAGGGTAGCTGAAAAGTTAAAAGAA------CACTTCCCTAATAAAGC   290
                  B08 ATTAGTACAACTCAATGGAACACAACTTTAGGTAGGGTAAAGGAAAAATTAAAAGAA------CACTTCCCTAATAGCAC   290
                  B10 ATTGACACAAATCTATGGAATAGCACTTTGCAAAGGGTAAAGGAAAAATTAGAAAAA------CATTTCAGGAATAAAAC   279
                  B11 ATTAGTGAAGGAAAATGGAACCTAACTTTAGAAAAGGTAAAGAAGAAATTAGAAAAA------CACTACCCTAATAAAAC   286
                  B13 ATTAGTAAAAATAATTGGCATAAAACTTTATACAGGGTAAGTAAAAAACTAGAAAAG------CTCTTCCCTAATAAGAC   290
                  B14 ATTAGT------AACTGGAATAACACGCTACAAAGGGTAAGGAAAAAATTAGAAGAA------CACTTCCCTAATAAAAC   284
                  B15 ATTAGTAAAGAGGACTGGAATAAAACTCTACAAAGGGTAGGTAAAAAATTAAAAGAA------TACTTC---AATAAAAC   284
                  B16 GTTAGTAAAAGTAACTGGACCGAAATTTTAAAACAGGTAAGAGAAAAATTAAGAGAG------AACTATAATAAAACAAT   290
                  B18 ATTAGTTATGATAAATGGAACACAACTTTAGAAAAGGTAAAGAAGAAATTAAACGAA------CACTTCCCTAATAATAC   290
                  B19 ATTAGT---AACTGGTGGAAT---ACTTTACAAAAGGTAAAGAAAAAACTGGAAGAA------CACTTCCCTGGTAAAAA   284
                  B20 ATTAGTGTTGAGAGTTGGAATGAAACTTTAGAACAAGTAAAGAAAAAATTAAGAGAA------CACTTCCCTAATAAAAC   290
                  B21 ATTARTGAAGGTRAATGGCACAAAACKTTAAAACAAGTAGCAGAAAAATTAGCAAAA------CACTTCCCTAATAAAAC   290
                  B23 ATTAGTACAACARAATGGAAMRAAACTTTARAATGGGTAAAAAACAAATTAGTAGAA------CATTTCCCTAATAAAAC   279
                  B24 ATTAGTAAAGGCCAATGGTACRAAACTTTAGAGCAGATAAAGGAAAAATTAGGAAAA------TACTAC---AATAAAAC   276
                  B25 ATTAGTAAAGGTGAATGGAAAAAAACTTTACAACAGGTACAAGAAAAATTAAAAGAA------CATTTCCCTAATAAAAG   276
                  B26 ATTAGTGAAGATGAATGGAACAACGTTTTAAAACAGGTAGGAAAAAAGCTAGCAGAA------CACTTCCCTGATAGGAA   279
                  B27 ATTAATGCAGGAAAATGGAACAAAACTTTACATCAGGTAGGAAAAAGATTAGCAGAA------CATTTCCCTAATAAAAC   279
        C.BR.92.BR025 ATTAGTAGAACAGCATGGAACAAAACTTTACAAGAGGTAGGTAAAAAATTAGCAGAG------CACTTCCCTAATAAAGC   290
     C.BW.96.96BW0502 ATTAATAAAACTGAATGGAATAGCACTTTACAAGGGGTAAGTAAAAAATTAGAAGAA------CACTTCTCTAAAAAAGC   290
      C.ET.86.ETH2220 ATTAGTGAAGAAAAATGGAATAAAACTCTACAAAAGGTAAAGGAAAAATTACAAAAG------CACTTCCCTAATAAAAC   290
    C.IN.95.95IN21068 ATTAGTGAAGATAAGTGGAATGAAACTTTACAAAATGTAAGTAAAAAGTTAGCAGAA------CACTTCCCTAATAAAAC   290
         D.CD.83.ELI ATTAGTAGAGCACAATGGAGTAAAACTTTACAACAAGTAGCTAGAAAATTAGGAACC------CTTCTTAACAAA---AC   284
         D.CD.83.NDK ATTAGCAGAGCAGAATGGAATAAAGCTTTACAACAGGTAGCTACAAAACTAGGAAAC------CTTCTTAACAAA---AC   296
      D.CD.84.84ZR085 ATTAGTGGAGTAAAATGGAATAACACCTTACGACAGGTAGCTAGAAAATTAGGAAAC------CTTTTGAACCAA---AC   287
      D.UG.94.94UG114 ATTAGTGGAGCAGGATGGAATAAAACTTTACAACAGGTAGCTGAAAAATTAGGAAAC------CTTCTTAACCAG---AC   284
       F1.BE.93.VI850 ATTAGTGGAACACAATGGAATAACACTCTGGAGTATGTAAAGGCAGAATTAAAGTCG------CATTTCCCTAATAATAC   290
   F1.BR.93.93BR020_1 GTTAGTGGAACACAATGGAGGAACACGTTAGCAAAGGTAAAGGCAAAGTTAGGGTCT------TATTTCCCTAAT---GC   287
     F1.FI.93.FIN9363 ATTAGTGGAGAACAATGGAATAAAACTTTAGACCGAGTAAAGGCAGAGTTAAAGCTT------CATTTT---AATAAA--   285
       F1.FR.96.MP411 ATTAATGGAACACAATGGAGTAAAACTAAAACACAGGTACAAGAAAAGTTACGGGCT------CTTTTT---AATAAA--   285
       F2.CM.95.MP255 ATTAATAGAACACTGTGGAATGAAACTTTAAAAAATGTAAGTGGAGAGTTCAAAAAA------CACTTC---AACTTCTC   287
       F2.CM.95.MP257 ATTAGTGAAAAACAGTGGTATGACACTTTAATAAAGATAGCTACAGAGTTCAAAGAC------CAATAT---AATAAAAC   287
       G.BE.96.DRCBL GTTAGTTGGACAAAATGGAATGAGACGCTAAGGGATGTCCAGGCAAAACTACAGGAA------TACTTTATTAATAAGAG   290
 G.FI.93.HH8793_12_1 GTTAGTGGAAGAAATTGGAGTAACATGATAGAGAAGGTAAAAGCACAACTAAGAAAA------ATCTTTAACAAGACC--   288
     G.NG.92.92NG083 GTTAGTAGAATAAAATGGAGGGAGATGTTAAAGAATGTCACAGCACAGCTAAGGAAA------ATCTATAATAATAAGAA   290
      G.SE.93.SE6165 GTTACTAAAAGAAAATGGAAAGAGGCCTTACAGAATGTCGCTGCAGAACTAGGGAAA------ATCTTTAATAAGAGCAG   296
       H.BE.93.VI991 ATTAGTGGAAAACAATGGAATGAGACCTTACACAAGGTAATCACCAAATTAGGAAGC------TACTTTGACAATAAAAC   293
       H.BE.93.VI997 GTTAGTGAAGAGAAATGGAATAAGACTTTACAACAGATAGCTACACAATTAAGTAAA------TACTTTGTCAATAGAAC   290
        H.CF.90.056 ATTAGTAGAACAGACTGGAATAAGACTTTACACCAGGTAGTTACACAATTAGGAATA------CACTTGAACAATAGAAC   290
      J.SE.93.SE7887 ATTAGTGAAAGAGATTGGAGTAACACTTTACGTAGAGTAGCTACAAAACTAAGAGAA------CACTTTAATAAAACA--   288
      J.SE.94.SE7022 ATTAGTAGAAAAGATTGGAATAACACTTTACGTAGAGTAGCTAAAAAACTAAGAGAA------CACTTTAATAAAACA--   288
     K.CD.97.EQTB11C ATTAGCGGAGGACAGTGGAATAAAACTGTAAACCAAGTAAAGAAAGAATTAGGAAAA------CACTTTAAC------AA   284
       K.CM.96.MP535 ATTAGTGGAGAAAAATGGAACATGACTTTAAGCAGAGTAAAGGAAAAGCTAAAAGAA------CATTTTAAGAAT---GG   287
                 M02 ATTACTAAAACAAAATGGAATACAACTTTACAAAGGGTAAAAGAGAAATTAAAAGAA------CACTTCCCTAATAAAAC   290
                 M03 --------------------------------------------------------------------------------   209
                 M04 ATTAGTGAAGATAAATGGAACAAAACTTTACAACGGGTAGGAGAAAAATTAAAGGAA------CATTTCCCTAATAAAAC   290
                 M07 ATTAGTGGAAAGGMATGGAATAATACTTTACAAAGGGTAGCTAAAAAGTTAAAAGAA------CACTTCCCTAATAAAAC   290
                 M08 MTTAGTACAACTCAATGGAACRCAACTTTAGRTRGGGTAAAGGAAAAATTAAAAGAA------CACTTCCCTAATAGMAC   290
                 M10 ATTAACACAAATCTATGGAATAGCACTTTGCAAAGGGTGAGGGAAAAATTAGAAAAA------CATTTCAGGAATAAAAC   290
                 M11 ATTAGTAAAGAAAAATGGGACCGAACTTTAGAACTGGTAAAAGAGAAATTAGAAAAA------CACTACCCTAATAAAAC   287
                 M13 ATTAGTAAAGATAATTGGCATAAAACTTTAYACAGGGTAAGTAAAAAACTAGNANAA------CTCTTCCCTAAT-----   285
                 M14 ATTAGT------AACTGGAATAACACGCTACAAAGGGTAAAGAAAAAATTAGGAGAA------CACTTCCCTAATAAAAC   128
                 M16 ATTAGTGGAAGGGMATGGAACGAAACTTTAGCAAAGGTAAMAGAAAAATTATTAGAA------CATTTTCCTAATAA---   284
                 M18 ATTAATTACTATAAATGGAACACAACTTTAGAAAAGGTAAAGAAGAGATTAAATGAA------CACTTCCCTAATAGAAC   290
                 M19 ATTAGY---AACTGGAGGGAA---ACTTTACAAAAGGTAGAGGAAAAATTGAAAGAA------CACTTCCCTGGTAAAAA   284
                M20a ATTAGTGTTGAGAGTTGGAATAAAACTTTAGAACAAGTAAAGAAAAAATTAAGAGAA------CACTTCCCTAATAAAAC   290
                M20b ATTAGTGTTGAGAGTTGGAATRAAACTTTAGAACAAGTAAAGRAAAAATTAAGAGAA------CACTTCCCTAATAAAAC   290
                 M21 ATTAGTGAAGGTAAATGGCACAAAACKTTAAAACAAGTAGCAGAAAAATTAGCAAAA------CACTTCCCTAATAAAAC   290
                 M23 ATTAGTACAACAAAATGGAACAAAACTTTAGAATGGGTAAAAAACAAATTAGTA-------------------------   270
                 M24 ATTACTRRARGCCAATGGAACRAAACTTTAGAKCRGATAAAGGAAAAATTAGGAAAA------TACTAC---AATAAAAC   209
                 M25 ATTAGTAAAGGTGAATGGAACAGAACTTTGCAACAGGTACCAAGAAAAATAAAAGGA------CA--------------   264
        O.BE.87.ANT70 TATAATGCCACTGATTGGGGAAAAATATTAAAACAAACAGCTGAAAGGTATTTAGAACTAGTAAACAATACAGGTAGTAT   299
               ruler .......250.......260.......270.......280.......290.......300.......310.......320
```



CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

File: Codon alignment - all sequences

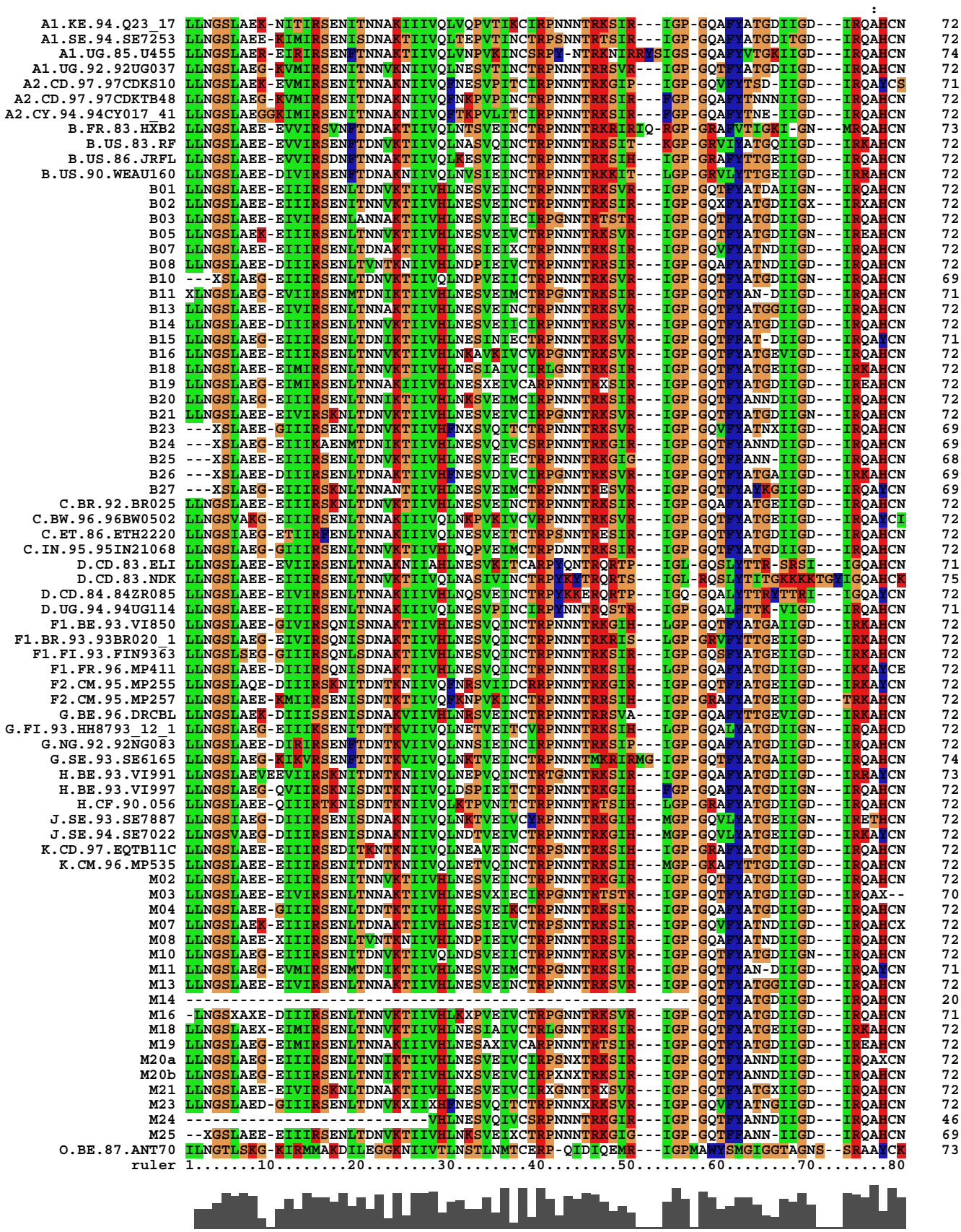


Appendix G

Amino acid alignment – all sequences obtained.

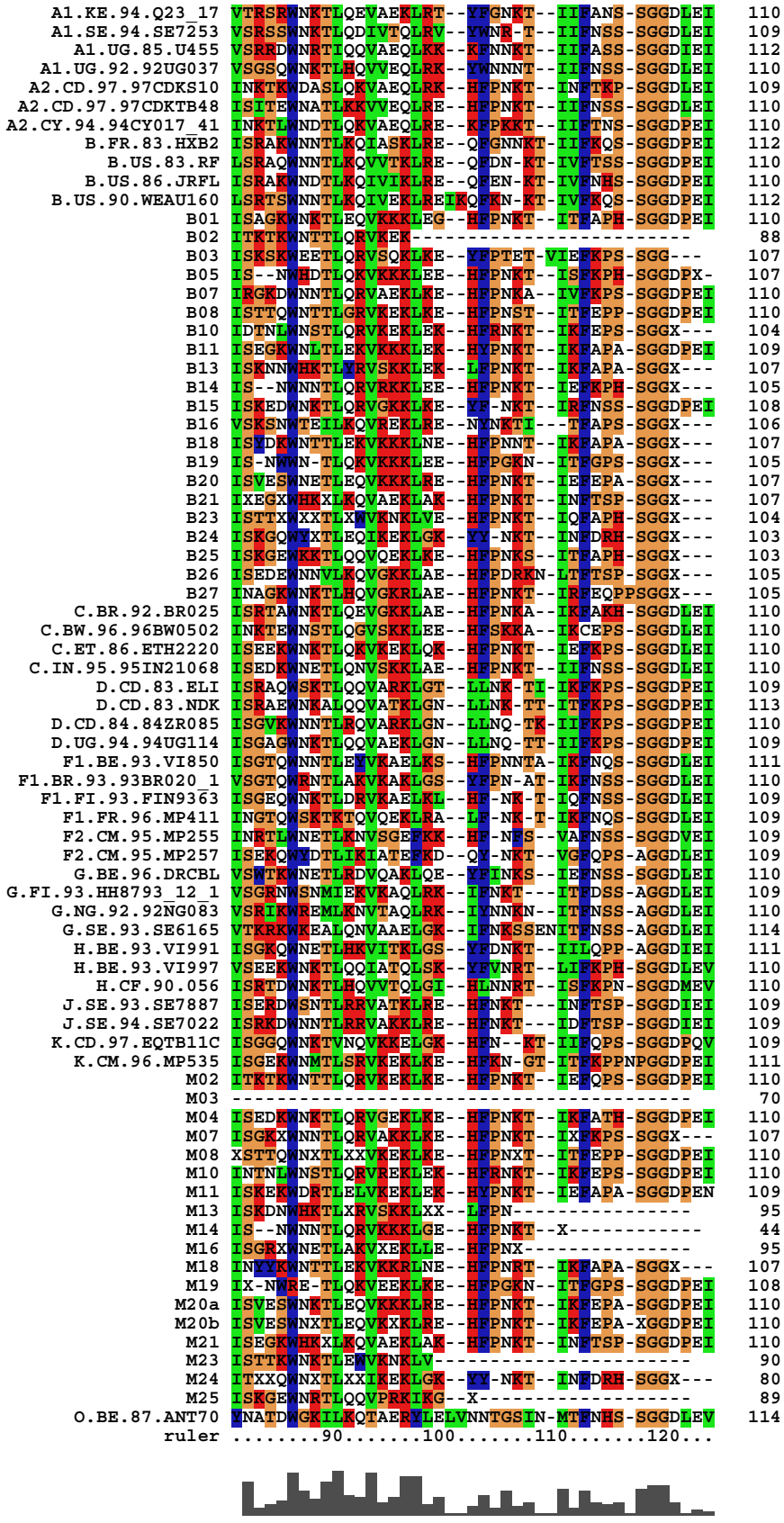
CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

File: Amino acid alignment - all sequences



CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

File: Amino acid alignment - all sequences
Page 2 of 2

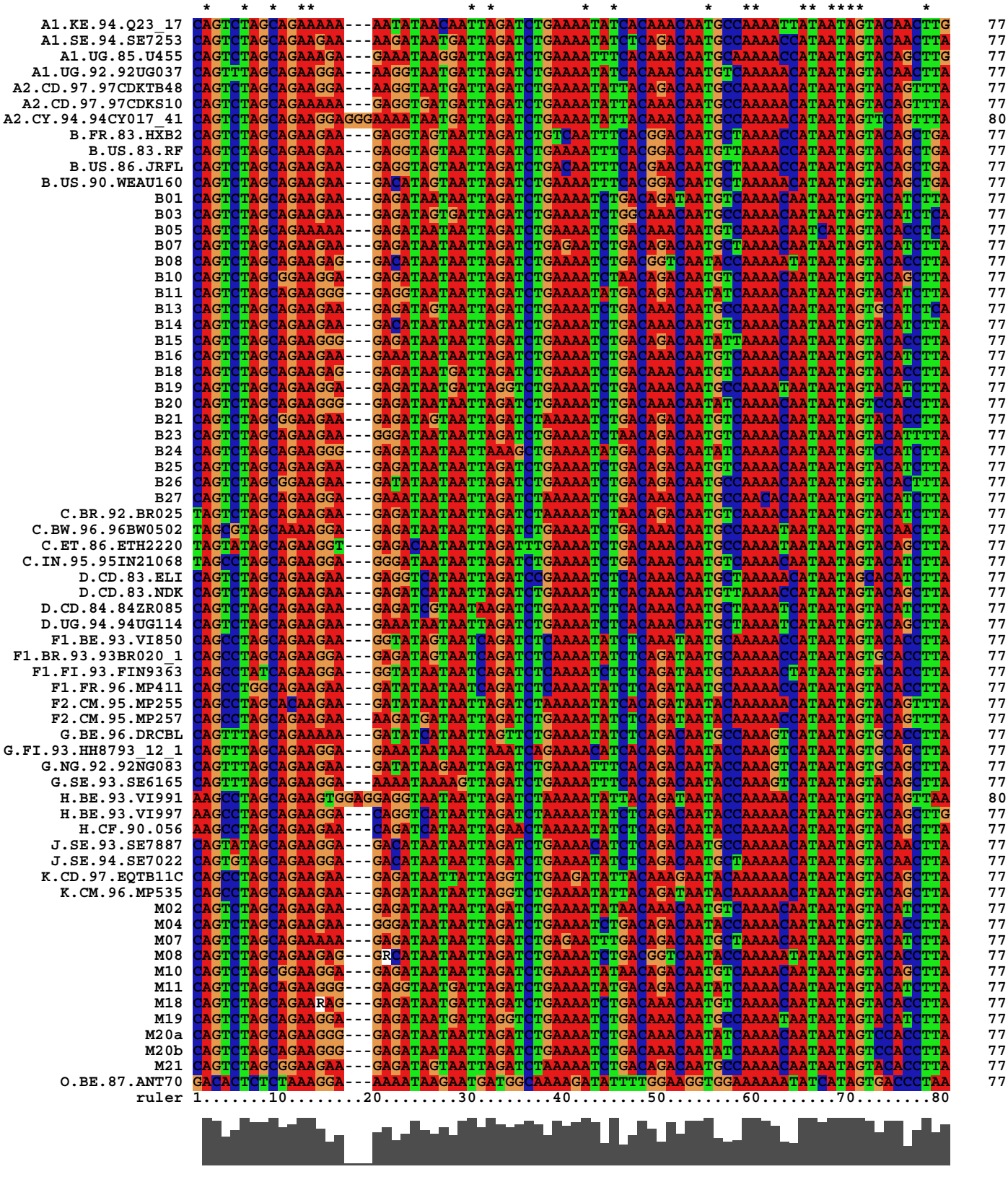


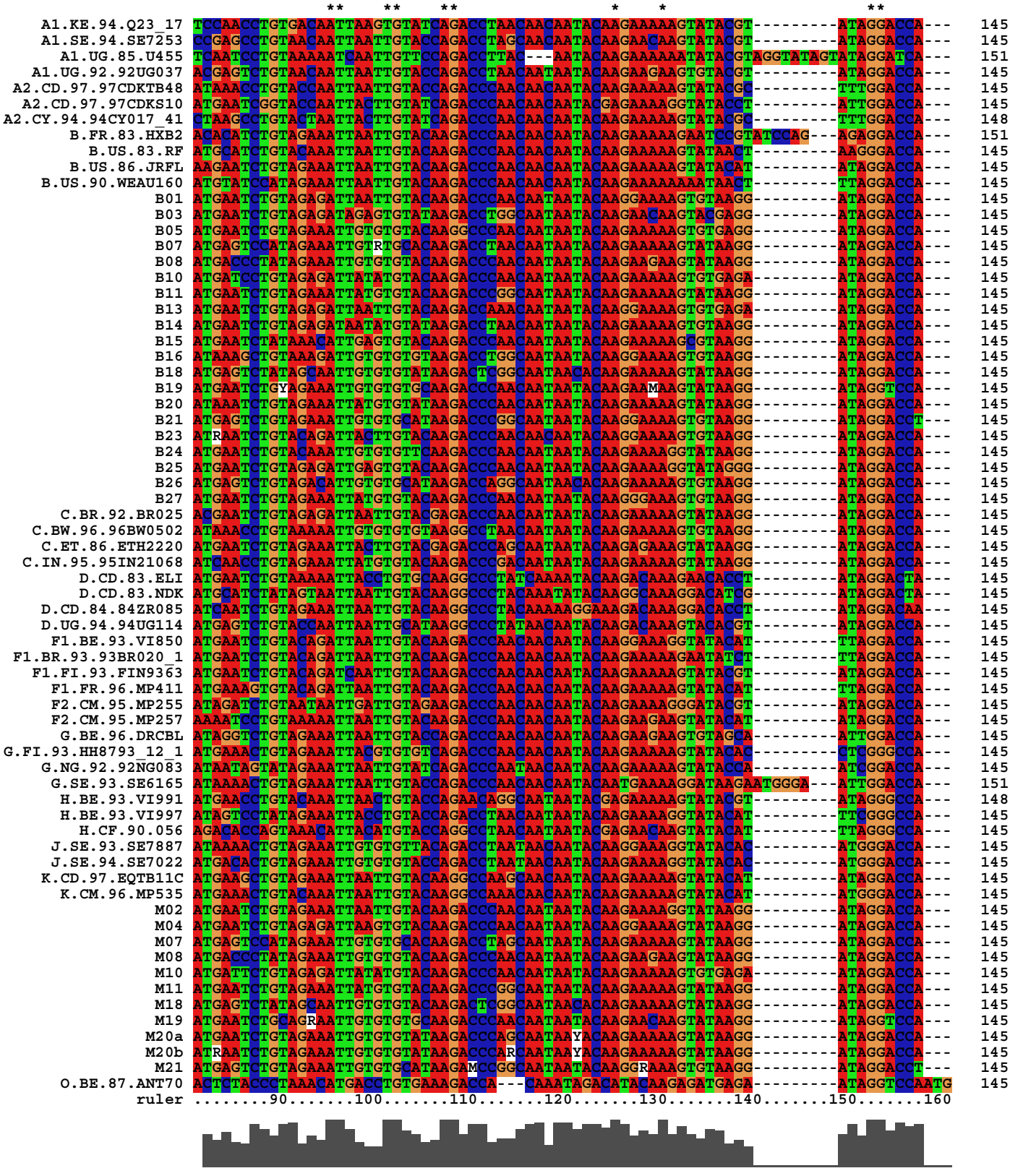
Appendix H

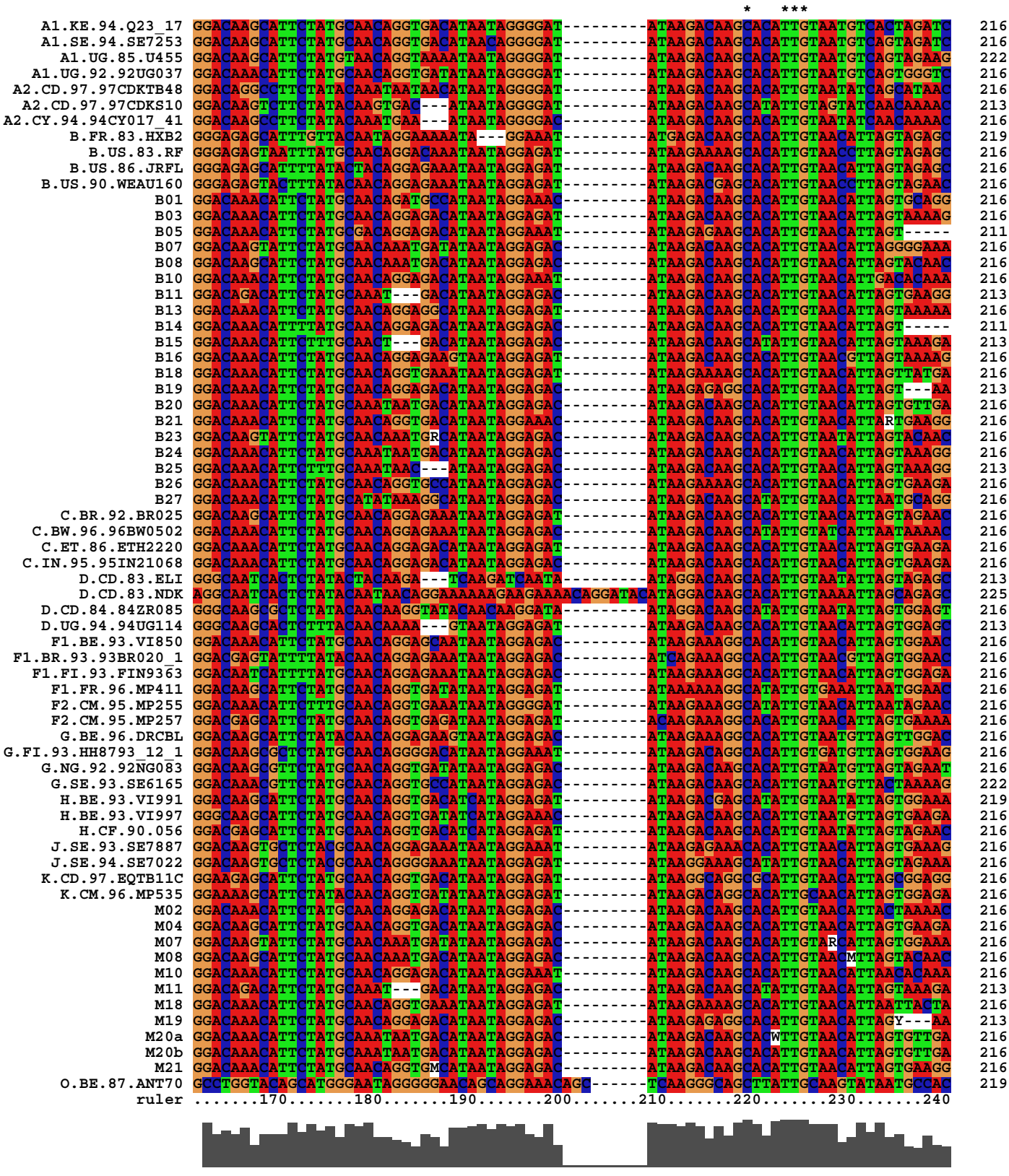
Nucleotide alignment – codon alignment of sequences used for analysis.

CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

File: Nucleotide sequences - codon aligned - sequences selected for phylogenetic analysis

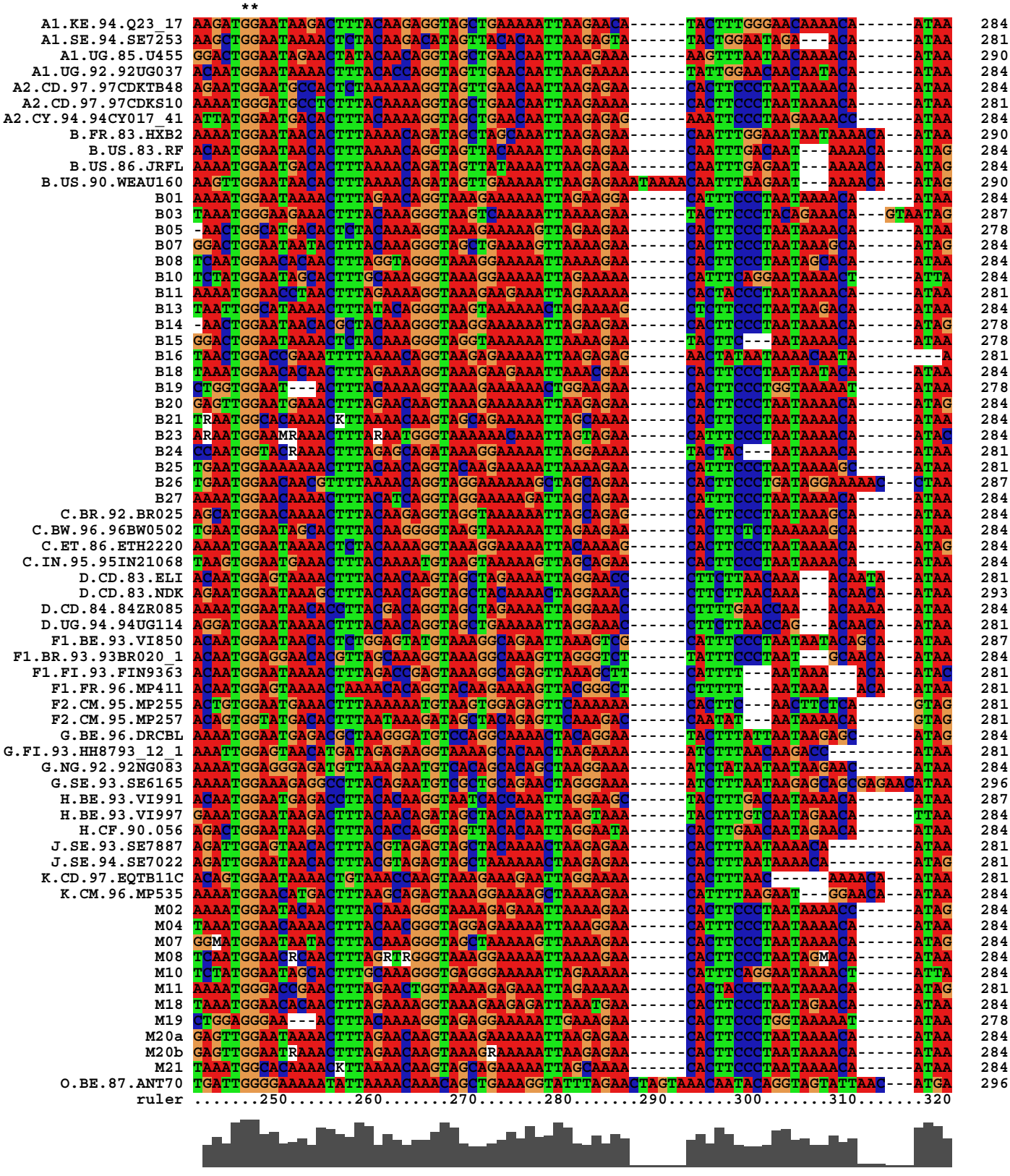


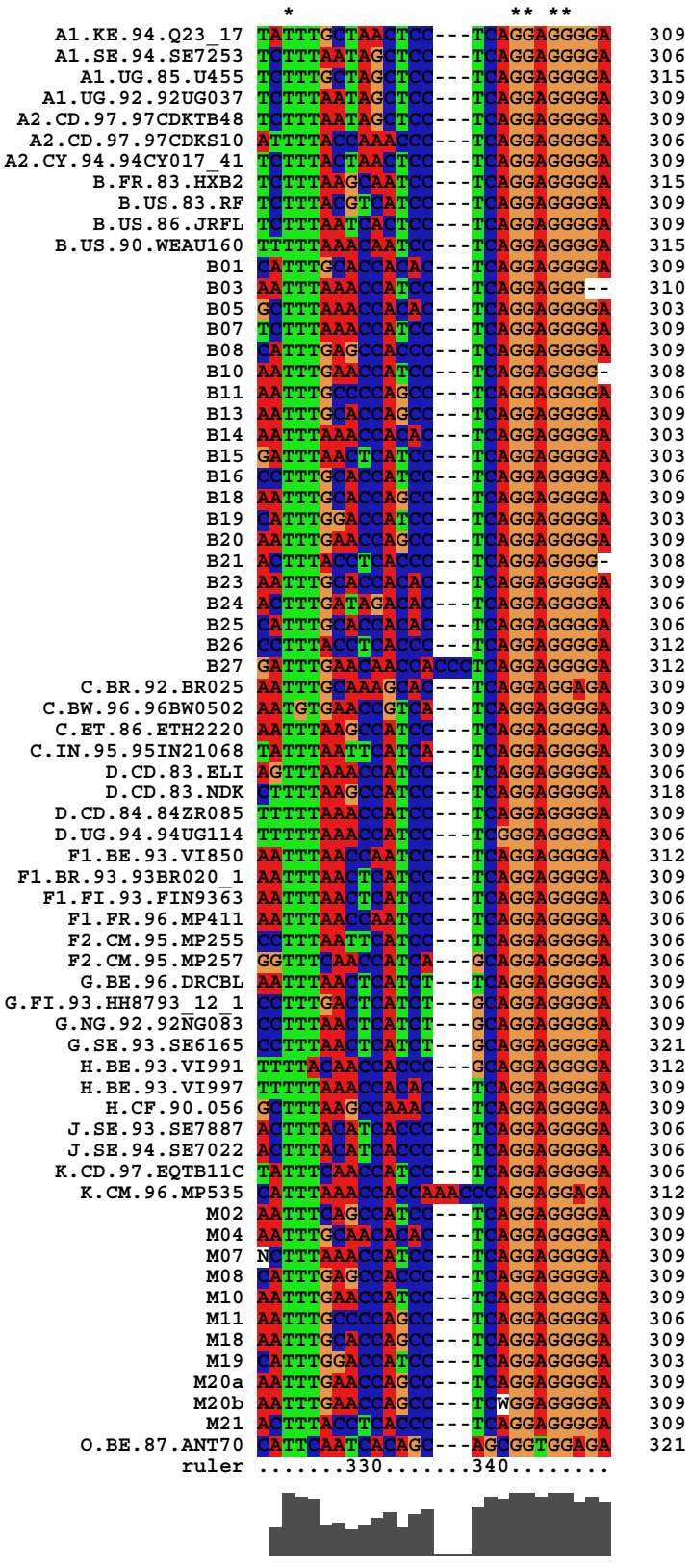

A1.KE.94.Q23_17 145
A1.SE.94.SE7253 145
A1.UG.85.U455 151
A1.UG.92.92UG037 145
A2.CD.97.97CDKTB48 145
A2.CD.97.97CDKS10 145
A2.CY.94.94CY017_41 148
B.FR.83.HXB2 151
B.US.83.RF 145
B.US.86.JRFL 145
B.US.90.WEAU160 145
B01 145
B03 145
B05 145
B07 145
B08 145
B10 145
B11 145
B13 145
B14 145
B15 145
B16 145
B18 145
B19 145
B20 145
B21 145
B23 145
B24 145
B25 145
B26 145
B27 145
C.BR.92.BR025 145
C.BW.96.96BW0502 145
C.ET.86.ETH2220 145
C.IN.95.95IN21068 145
D.CD.83.ELI 145
D.CD.83.NDK 145
D.CD.84.84ZR085 145
D.UG.94.94UG114 145
F1.BE.93.VI850 145
F1.BR.93.93BR020_1 145
F1.FI.93.FIN9363 145
F1.FR.96.MP411 145
F2.CM.95.MP255 145
F2.CM.95.MP257 145
G.BE.96.DRCBL 145
G.FI.93.HH8793_12_1 145
G.NG.92.92NG083 145
G.SE.93.SE6165 151
H.BE.93.VI991 148
H.BE.93.VI997 145
H.CF.90.056 145
J.SE.93.SE7887 145
J.SE.94.SE7022 145
K.CD.97.EQTB11C 145
K.CM.96.MP535 145
M02 145
M04 145
M07 145
M08 145
M10 145
M11 145
M18 145
M19 145
M20a 145
M20b 145
M21 145
O.BE.87.ANT70 145
ruler .....90.....100.....110.....120.....130.....140.....150.....160


A1.KE.94.Q23_17
A1.SE.94.SE7253
A1.UG.85.U455
A1.UG.92.92UG037
A2.CD.97.97CDKTB48
A2.CD.97.97CDKS10
A2.CY.94.94CY017_41
B.FR.83.HXB2
B.US.83.RF
B.US.86.JRFL
B.US.90.WEAU160
B01
B03
B05
B07
B08
B10
B11
B13
B14
B15
B16
B18
B19
B20
B21
B23
B24
B25
B26
B27
C.BR.92.BR025
C.BW.96.96BW0502
C.ET.86.ETH2220
C.IN.95.95IN21068
D.CD.83.ELI
D.CD.83.NDK
D.CD.84.84ZR085
D.UG.94.94UG114
F1.BE.93.VI850
F1.BR.93.93BR020_1
F1.FI.93.FIN9363
F1.FR.96.MP411
F2.CM.95.MP255
F2.CM.95.MP257
G.BE.96.DRCBL
G.FI.93.HH8793_12_1
G.NG.92.92NG083
G.SE.93.SE6165
H.BE.93.VI991
H.BE.93.VI997
H.CF.90.056
J.SE.93.SE7887
J.SE.94.SE7022
K.CD.97.EQTB11C
K.CM.96.MP535
M02
M04
M07
M08
M10
M11
M18
M19
M20a
M20b
M21
O.BE.87.ANT70
ruler
.....170.....180.....190.....200.....210.....220.....230.....240

CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

File: Nucleotide sequences - codon aligned - sequences selected for phylogenetic analysis



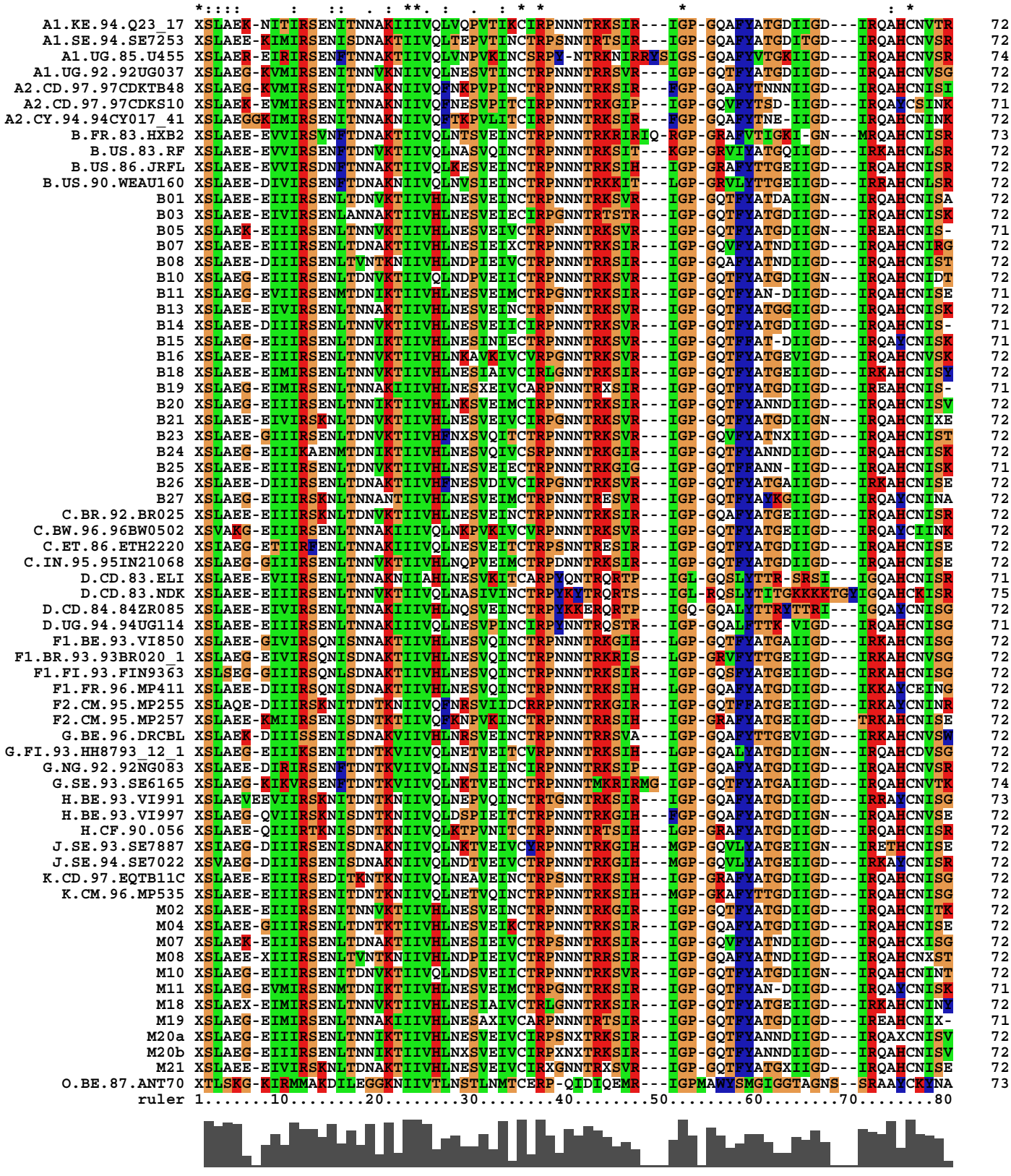

A1.KE.94.Q23_17 309
A1.SE.94.SE7253 306
A1.UG.85.U455 315
A1.UG.92.92UG037 309
A2.CD.97.97CDKTB48 309
A2.CD.97.97CDKS10 306
A2.CY.94.94CY017_41 309
B.FR.83.HXB2 315
B.US.83.RF 309
B.US.86.JRFL 309
B.US.90.WEAU160 315
B01 309
B03 310
B05 303
B07 309
B08 309
B10 308
B11 306
B13 309
B14 303
B15 303
B16 306
B18 309
B19 303
B20 309
B21 308
B23 309
B24 306
B25 306
B26 312
B27 312
C.BR.92.BR025 309
C.BW.96.96BW0502 309
C.ET.86.ETH2220 309
C.IN.95.95IN21068 309
D.CD.83.ELI 306
D.CD.83.NDK 318
D.CD.84.84ZR085 309
D.UG.94.94UG114 306
F1.BE.93.VI850 312
F1.BR.93.93BR020_1 309
F1.FI.93.FIN9363 306
F1.FR.96.MP411 306
F2.CM.95.MP255 306
F2.CM.95.MP257 306
G.BE.96.DRCBL 309
G.FI.93.HH8793_12_1 306
G.NG.92.92NG083 309
G.SE.93.SE6165 321
H.BE.93.VI991 312
H.BE.93.VI997 309
H.CF.90.056 309
J.SE.93.SE7887 306
J.SE.94.SE7022 306
K.CD.97.EQTB11C 306
K.CM.96.MP535 312
M02 309
M04 309
M07 309
M08 309
M10 309
M11 306
M18 309
M19 303
M20a 309
M20b 309
M21 309
O.BE.87.ANT70 321
ruler .....330.....340.....

Appendix I

Amino acid alignment – sequences used for analysis.

CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

File: Amino acid sequences - sequences selected for phylogenetic analysis



CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

File: Amino acid sequences - sequences selected for phylogenetic analysis