Journal of General Virology (1998), **79**, 1055–1068. Printed in Great Britain

Analysis of human immunodeficiency virus type 1 *env* and *gag* sequence variants derived from a mother and two vertically infected children provides evidence for the transmission of multiple sequence variants

Christopher M. Wade, Denis Lobidel† and Andrew J. Leigh Brown

Centre for HIV Research, Institute of Cell, Animal and Population Biology, Division of Biological Sciences, The University of Edinburgh, Edinburgh EH9 3JN, UK

In order to investigate the transmission of human immunodeficiency virus type 1 (HIV-1) from mother-to-child we have examined serial plasma RNA samples obtained from a mother over an eight year period spanning four pregnancies. Child 1 and 2 (born January 1987 and June 1990) were uninfected whilst child 3 and 4 (born July 1992 and February 1994) were HIV positive. Genetic variation was examined within the viral population of the mother and her two infected children for both the V3 loop and flanking regions of the *env* gene and the p17 region of the *gag* gene. In one child (child 4) a highly homogeneous virus population was observed within both *env* and *gag* in contrast to the more heterogeneous virus population observed within the mother. Viral sequences of child 4 clustered within a single branch within the reconstructed phylogenetic tree. This is consistent with the transmission of a single maternal variant to the child in this case, which may indicate a selective process. By contrast, child 3 showed substantial genetic heterogeneity even within the first samples obtained shortly after birth. Sequences of child 3 clustered in two distinct groups within the phylogenetic tree and were separated by sequences of the mother. These results are not consistent with the selective transmission of a single maternal variant to the child in this case and we therefore propose that the infection within child 3 is the result of the transmission of multiple sequence variants to the child. All transmitted sequence variants were predicted to be of the macrophage-tropic, non-syncytium-inducing (NSI) phenotype.

Introduction

Mother-to-child transmission of human immunodeficiency virus type 1 (HIV-1) is estimated to occur at a rate of 15–20% of children born to HIV-1 infected mothers in Europe, and accounts for the majority of paediatric AIDS cases (Newell, 1995). Transmission may occur either *in utero* (prepartum) (Courgnaud *et al.*, 1991; Soeiro *et al.*, 1992), at delivery (intrapartum) (Ehrnst *et al.*, 1991; De Rossi *et al.*, 1992) or postnatally through breastfeeding (postpartum) (Ziegler *et al.*, 1985; Lepage *et al.*, 1987). Evidence for *in utero* infection has been provided by the detection of viral nucleic acids in foetal tissues (Joviasis *et al.*, 1985; Sprecher *et al.*, 1986; Courgnaud *et al.*, 1991; Soeiro *et al.*, 1992) and clinical studies have interpreted the early detection of HIV infection in newborns (within the first few days following birth) as indicative of transmission occurring early in pregnancy. Nevertheless, markers for HIV infection are absent at birth in a high proportion of perinatally infected children (Borkowsky *et al.*, 1992; Burgard *et al.*, 1992; Krivine *et al.*, 1992) and this has been interpreted as evidence for transmission either late in pregnancy or at the time of delivery (Ehrnst *et al.*, 1991; De Rossi *et al.*, 1992). Epidemiological data have indeed shown increased risks of transmission associated with vaginal delivery, with a significantly lower rate of perinatal transmission observed in children delivered by Caesarean section (European Collaborative Study, 1994). Studies of vertical transmission in

Author for correspondence: Christopher M. Wade. Present address: Department of Genetics, University of Nottingham, Queens Medical Centre, Clifton Boulevard, Nottingham NG7 2UH, UK. Fax +44 115 970 9906. e-mail pdzcmw@evol.nott.ac.uk

0001-5228

Table 1. Times of sampling, clinical data, $CD4^+$ cell counts and the number of molecules sequenced per time-point for the patients studied

Patient	Sample name	Sample date	Time course* Child 3	Time course* Child 4	Population	$CD4^+$ cell counts (cells/mm^3)	Clinical information	CDC classification	*env* sequences	*gag* sequences
Mother	1R	15.06.86			Plasma	ND		2	0	0
	2R	04.12.87			Plasma	563		2	0	0
	3R	15.02.89			Plasma	343	Asymptomatic	2	0	0
	4R	26.10.89			Plasma	250		2	0	0
	5R	30.01.90	−2·5 yrs	−4 yrs	Plasma	260		2	0	4
	6R	24.04.90			Plasma	330		2	0	0
	7R	15.06.90			Plasma	729		3	0	0
	8R	19.07.90			Plasma	300		3	0	0
	9R	15.10.90			Plasma	290		3	0	0
	10R	22.11.91			Plasma	ND		3	0	0
	11R	20.01.92			Plasma	212		3	0	0
	12R	28.02.92	−5 mths	−24 mths	Plasma	349	Persistent	3	1	1
	13R	28.04.92	−3 mths	−22 mths	Plasma	210	generalized	3	0	3
	14R	08.06.92	−2 mths	−20 mths	Plasma	215	lymphadenopathy	3	3	0
	15R	13.07.92	−18 days	−19 mths	Plasma	650		3	0	3
	16R	03.08.92			Plasma	342		3	0	0
	17R	25.06.93			Plasma	380		3	0	0
	18R	03.09.93	+13 mths	−6 mths	Plasma	330		3	5	12
	19R	21.12.93	+17 mths	−2 mths	Plasma	270		3	4	11
Child 3	1D	25.08.92	1 mth		PBMC	3580		N1	9	10
	1R	25.08.92	1 mth		Plasma	3580	Well	N1	1	2
	2R	30.09.92	2 mths		Plasma	2340		N1	2	11
	3R	10.11.92	4 mths		Plasma	2810		A1	3	8
	4D	18.12.92	5 mths		PBMC	2510	Lymphadenopathy, respiratory infection	A1	9	11
	4R	18.12.92	5 mths		Plasma	2510		A1	0	11
	5D	23.06.93	11 mths		PBMC	960	Respiratory infections, nappy	A2	11	11
	5R	23.06.93	11 mths		Plasma	960	rash, hepatomegaly	A2	2	6
	6D	01.10.93	15 mths		PBMC	1750	Respiratory infection, acute	A2	7	9
	6R	01.10.93	15 mths		Plasma	1750	parotitis	A2	0	11
	7D	03.05.94	22 mths		PBMC	1340	Recurrent upper	A2	11	12
	7R	03.05.94	22 mths		Plasma	1340	respiratory infections	A2	11	5
	8D	17.08.94	25 mths		PBMC	1820	Lymphadenopathy	A2	0	0
	8R	17.08.94	25 mths		Plasma	1820		A2	5	8
Child 4	1R	03.03.94		6 days	Plasma	ND	ND	ND	0	0
	2D	22.03.94		1 mth	PBMC	3550		N1	9	12
	2R	22.03.94		1 mth	Plasma	3550	Well	N1	8	11
	3D	10.06.94		4 mths	PBMC	3210		N1	11	11
	3R	10.06.94		4 mths	Plasma	3210		N1	1	4
	4D	05.08.94		6 mths	PBMC	3590	Hepatomegaly	A1	10	10
	4R	05.08.94		6 mths	Plasma	3590		A1	1	1

* Time-course provides approximate details of the sampling time-points relative to the dates of birth of child 3 (31.07.92) and child 4 (26.02.94). For the mother, the time is given relative to the delivery date of each child. − indicates sampling dates before delivery; + indicates sampling dates following delivery. Time-course details are only provided for maternal samples for which either *env* or *gag* molecules were successfully sequenced. For the children, the age of the child at the time of sampling is given.
ND, Not determined.

twins (Goedert *et al.*, 1991) also provide evidence for transmission at delivery; it has been proposed that the higher rate of transmission observed in first-born twins is due to the exposure of the first twin to infectious material within the birth canal for a greater time than the subsequent twin (Goedert *et al.*, 1991). Such data would suggest the occurrence of transmission as the child encounters the cervix and birth canal.

Vertical transmission of HIV-1 appears to be dependent upon the mother's immunological and virological status. Recent infection, high viral load, p24 antigenaemia, low $CD4^+$ T cell counts, low anti-gp120 antibodies in the mother's serum, absence of antibodies against specific domains of gp120 and gp41, and advanced clinical stage of the mother appear to be associated with an increased risk of vertical transmission (Newell, 1995). Viral factors may also play a role in influencing vertical transmission.

Molecular analyses of HIV-1 sequence variants following sexual or parenteral transmission in adults have revealed that a highly homogeneous sequence population is observed within *env* in the recipient immediately following transmission (McNearney *et al.*, 1990; Wolfs *et al.*, 1992; Zhang *et al.*, 1993; Zhu *et al.*, 1993). This is in contrast to the heterogeneous sequence population typically observed within *env* in the long-term infected patient. It has therefore been postulated that a sequence bottleneck occurs upon infection, with infection initiated by a limited number of variants or even one particular variant. p17 *gag* gene sequences were not however found to show such a restricted level of sequence diversity within the recipient upon seroconversion (Zhang *et al.*, 1993; Zhu *et al.*, 1993). This suggests that the homogeneity observed in *env* is due to strong selection for specific *env* sequences either upon transmission or in the interval between exposure and seroconversion (Zhang *et al.*, 1993).

Molecular studies examining mother-to-child transmission of HIV-1 have suggested that infection within the child is initiated by a single maternal variant (Wike *et al.*, 1992; Wolinsky *et al.*, 1992; Mulder-Kampinga *et al.*, 1993; Scarlatti *et al.*, 1993; Ahmad *et al.*, 1995; Mulder-Kampinga *et al.*, 1995), a situation analogous to that observed for sexual or parenteral transmission in adults. A highly homogeneous sequence population is typically observed within the child which is considerably less diverse than that of the mother. The transmitted variant has been reported to represent a minor maternal form in the majority of mother to-child transmission cases studied (Wike *et al.*, 1992; Wolinsky *et al.*, 1992; Mulder-Kampinga *et al.*, 1993, 1995; Scarlatti *et al.*, 1993; Ahmad *et al.*, 1995) and this has been interpreted as evidence that selection is playing a role in vertical transmission. A number of studies have however reported that the transmitted variant can represent either a major or a minor maternal form (Wike *et al.*, 1992; Wolinsky *et al.*, 1992; Scarlatti *et al.*, 1993). The transmission of multiple HIV-1 genotypes from mother-to-child has nevertheless also been reported for a number of vertical transmission cases (Lamers *et al.*, 1994; Van't Wout *et al.*, 1994; Briant *et al.*, 1995). Multiple HIV-1 genotypes have also been observed to be transmitted to twins during a single pregnancy (Weiser *et al.*, 1993).

In order to examine aspects of the transmission of HIV-1 from mother-to-child we have examined serial plasma RNA samples obtained from a mother across an 8 year period spanning four pregnancies (child 1 and 2 uninfected, child 3 and 4 infected) and compared the viral population of the mother with plasma RNA and proviral DNA sequences obtained from her two infected children. We provide evidence that different transmission processes were in operation for the infection of the two infected children.

Methods

■ **Patients.** Sequential plasma and peripheral blood mononuclear cell (PBMC) samples were obtained over a period of 9 years (1986–1994) from a transmission set comprising a mother and her two HIV-1 infected children. During the studied time-period, the mother gave birth to four children of whom child 1 (born 1.87) and child 2 (born 1.6.90) were uninfected whereas child 3 (born 31.7.92) and child 4 (born 26.2.94) were infected. The mother first tested HIV positive on 17.10.85 with the seroconversion date estimated at May 1984. Her risk factor for HIV infection was intravenous drug use.

The times of sampling, clinical status and $CD4^+$ cell counts of the patients within the mother–child transmission set are presented in Table 1. [In this paper dates are given in the standard British format; thus 1.6.90 is 1 June 1990.]

■ **Viral nucleic acid extraction, PCR amplification and sequencing.** DNA and RNA extraction, PCR amplification and automated DNA sequencing were performed essentially as described previously (Simmonds *et al.*, 1990; Zhang *et al.*, 1991; Leigh Brown & Simmonds, 1995).

An approximately 436 bp fragment spanning the V3 loop and flanking regions of the *env* gene (positions 7029–7464 in the HIV-HXB2 genome (GenBank accession no. K03445) and an approximately 390 bp fragment of the p17 coding region of the *gag* gene (positions 857–1246 in the HIV-HXB2 genome) were amplified by limiting dilution nested PCR (Simmonds *et al.*, 1990; Leigh Brown & Simmonds, 1995), with 30 cycles in both the first and second rounds of amplification. The following primers were used:

env

1 (outer, +): 5′ TACAATGTACACATGGAATT 3′ (nucleotide positions 6957–6976 in the HIV-HXB2 genome)

2 (outer, −): 5′ GGAGGGGCATACATTGC 3′ (7520–7537 HXB2)

3 (inner, +): 5′ TGGCAGTCTAGCAGAAGAAG 3′ (7009–7028 HXB2)

4 (inner, −): 5′ ATTCTGCATGGGAGTGTG 3′ (7465–7482 HXB2)

gag

1 (outer, +): 5′ GCGAGAGCGTCAGTATTAAGCGG 3′ (795–817 HXB2)

2 (outer, −): 5′ TCTGATAATGCTGAAAACATGGG 3′ (1296–1318 HXB2)

3 (inner, +): 5′ GGGAAAAAATTCGGTTAAGGCC 3′ (833–856 HXB2)

4 (inner, −): 5′ CTTCTACTACTTTTACCCATGC 3′ (1247–1270 HXB2)

Both sense and antisense strands were sequenced using a direct solid phase automated sequencing approach using T7 dye terminator sequencing chemistry (detailed in Leigh Brown & Simmonds, 1995). Sequencing products were run on an Applied Biosystems 373A DNA sequencer.

■ **Sequence analysis.** Raw nucleotide sequences (sense and antisense strands) were assembled using the STADEN package (Staden, 1993). Sequences were then aligned using the CLUSTAL V algorithm (Higgins *et al.*, 1992), as implemented in version 2.2 of the Genetic Data Environment (GDE) package (Smith *et al.*, 1994) and the alignment improved manually. Phylogenetic analyses were performed using programs taken from version 3.52c of the Phylogeny Inference Package (PHYLIP; Felsenstein, 1993). Nucleotide distances were estimated using the generalized two-parameter (maximum likelihood) model (Kishino & Hasegawa, 1989) (program DNADIST) and phylogenetic trees were reconstructed using the neighbour-joining method (Saitou & Nei, 1987) (program NEIGHBOR). Bootstrap resampling (Felsenstein, 1985) (programs SEQBOOT and CONSENSE; 1000 replications) was employed to assign support to the inferred phylogenies and alternative phylogenetic hypotheses were evaluated statistically by likelihood ratio tests (Kishino & Hasegawa, 1989). The number of synonymous substitutions per synonymous site (d_s) and the number of non-synonymous substitutions per nonsynonymous site (d_n) were calculated using the Jukes–Cantor one-parameter model (Jukes & Cantor, 1969) as implemented in the Molecular Evolutionary Genetics Analysis program version 1.01 (MEGA; Kumar *et al.*, 1993).

The sequence dataset was screened for the presence of potential contaminants by comparing the patient sequences with sequences of equivalent regions of all clones and other patients examined within the laboratory.

■ **Nucleotide sequence accession numbers.** Nucleotide sequences reported in this study have been assigned the GenBank accession numbers AF050769–AF051090.

Results

We have obtained 124 sequences spanning the V3 loop and flanking regions of the *env* gene and 198 sequences of the p17 region of the *gag* gene from an infected mother and her two HIV-1 infected children. Nineteen plasma samples were obtained from the mother over a period of 8 years (1986–1993) spanning four pregnancies. Plasma RNA viral sequences were however only obtained from a limited number of time-points (Table 1). Sequences were obtained from both plasma RNA and proviral DNA samples from the two infected children. Child 3 was followed for a 2 year period and child 4 was followed for a 5 month period (Table 1).

Sequence variability in the mother and her children

The intra-patient genetic diversities (Table 2, on diagonal), inter-patient genetic distances (Table 2) and intra-sample genetic diversities (Table 3) were calculated for each pairwise comparison between sequences of the mother-to-child transmission set. Relatively high intra-patient sequence diversity was observed in child 3 within the first 2 years of life in both *env* (2·8%) and *gag* (1·2%) (Table 2). Diversity was relatively high even within the earliest samples obtained. The two sequences obtained in *env* from RNA sample 2R (approximately 2 months) differed at 1% of their nucleotide sites with a similar diversity observed in *gag* on a substantial sequence sample ($n = 11$) (Table 3). Within the first proviral DNA sample, 1D (approximately 1 month), 2·2% ($n = 9$) diversity was observed in *env* with 1·1% ($n = 10$) diversity observed in *gag* (Table 3). In contrast, child 4 showed much less variability, with an overall diversity of 0·6% observed in *env* and 0·3% observed in *gag* (Table 2). The earliest sample obtained for child 4, sample 2 (approximately 1 month), showed levels of sequence diversity of approximately 0·1% in both *env* and *gag* within both the plasma RNA and proviral DNA populations (Table 3). Sequences from the infected mother showed diversity levels of 1·7% ($n = 13$) in *env* and 2·1% ($n = 34$) in *gag* (Table 2). The lower levels of intra-patient diversity observed within the mother in *env* than in *gag* reflect the limited number of *env* sequences obtained from this patient.

Table 2. Inter-patient distances and intra-patient diversities for the infected mother and children within the V3 loop and flanking regions of *env* (*a*) and the p17 region of *gag* (*b*)

Distances were calculated using the generalized two-parameter (maximum likelihood) distance estimate (program DNADIST; Felsenstein 1993). Intra-patient distances are shown in bold on the diagonal. *n*, No. of sequences within each sample.

	n	Mother	Child 3	Child 4
(*a*) *env*				
Mother	13	**0·0167**		
Child 3	71	0·0284	**0·0277**	
Child 4	40	0·0218	0·0385	**0·0056**
(*b*) *gag* p17				
Mother	34	**0·0212**		
Child 3	115	0·0207	**0·0125**	
Child 4	49	0·0208	0·0270	**0·0027**

The number of synonymous substitutions per synonymous site (d_s) and the number of nonsynonymous substitutions per nonsynonymous site (d_n) are presented for each sampling time-point in Table 3. Synonymous and nonsynonymous distances were relatively high for all time-points in both *env* and *gag* in child 3 in contrast with low synonymous and nonsynonymous distances in both *env* and *gag* in child 4. Within child 3, the synonymous distances were typically greater than the nonsynonymous distances in both the p17 region of *gag* and within the V3 loop and flanking regions of the *env* gene. This resulted in d_s/d_n ratios typically greater than 1 for all time-points in both *env* and *gag*, although d_s/d_n ratios were considerably greater in *gag* than *env*. Within child

Table 3. Nucleotide sequence diversity within the V3 loop and flanking regions of *env* and the p17 region of *gag* for the mother and her infected children

ML, generalized two-parameter (maximum likelihood) distance estimate (program DNADIST; Felsenstein 1993); d_s, proportion of synonymous nucleotide substitutions per synonymous site; d_n, proportion of nonsynonymous nucleotide substitutions per nonsynonymous site. d_s and d_n values were calculated using the Jukes–Cantor one-parameter model as implemented in MEGA version 1.01 (Kumar *et al.*, 1993). *n*, No. of sequences within each sample.

Patient	Sample name	V3 loop and flanking regions of *env*					p17 region of *gag*				
		n	ML	d_s	d_n	d_s/d_n	*n*	ML	d_s	d_n	d_s/d_n
Mother	5R	0	–	–	–	–	4	0·0160	0·0244	0·0138	1·7681
	12R	1	0·0000	0·0000	0·0000	–	1	0·0000	0·0000	0·0000	–
	13R	0	–	–	–	–	3	0·0199	0·0326	0·0165	1·9758
	14R	3	0·0102	0·0178	0·0082	2·1707	0	–	–	–	–
	15R	0	–	–	–	–	3	0·0203	0·0333	0·0169	1·9704
	18R	5	0·0182	0·0077	0·0209	0·3684	12	0·0166	0·0342	0·0119	2·8739
	19R	4	0·0202	0·0086	0·0232	0·3707	11	0·0181	0·0285	0·0154	1·8506
Child 3	1D	9	0·0221	0·0225	0·0219	1·0274	10	0·0113	0·0246	0·0079	3·1139
	1R	1	0·0000	0·0000	0·0000	–	2	0·0103	0·0243	0·0066	3·6818
	2R	2	0·0108	0·0128	0·0103	1·2428	11	0·0120	0·0208	0·0085	2·4470
	3R	3	0·0183	0·0213	0·0174	1·2241	8	0·0097	0·0289	0·0046	6·2826
	4D	9	0·0232	0·0216	0·0240	0·9000	11	0·0100	0·0306	0·0046	6·6522
	4R	0	–	–	–	–	11	0·0083	0·0260	0·0037	7·0270
	5D	11	0·0178	0·0234	0·0165	1·4182	11	0·0153	0·0373	0·0096	3·8854
	5R	2	0·0168	0·0265	0·0143	1·8531	6	0·0208	0·0448	0·0112	4·0000
	6D	7	0·0145	0·0260	0·0114	2·2807	9	0·0085	0·0334	0·0020	16·700
	6R	0	–	–	–	–	11	0·0077	0·0303	0·0015	20·200
	7D	11	0·0217	0·0236	0·0212	1·1132	12	0·0091	0·0337	0·0026	12·961
	7R	11	0·0213	0·0275	0·0197	1·3959	5	0·0093	0·0398	0·0013	30·615
	8R	5	0·0163	0·0257	0·0139	1·8489	8	0·0164	0·0428	0·0095	4·5053
Child 4	2D	9	0·0006	0·0000	0·0008	0·0000	12	0·0009	0·0000	0·0011	0·0000
	2R	8	0·0012	0·0056	0·0000	–	11	0·0014	0·0022	0·0012	1·8333
	3D	11	0·0042	0·0000	0·0053	0·0000	11	0·0022	0·0084	0·0006	14·000
	3R	1	0·0000	0·0000	0·0000	–	4	0·0000	0·0000	0·0000	–
	4D	10	0·0070	0·0049	0·0084	0·5833	10	0·0065	0·0164	0·0039	4·2051
	4R	1	0·0000	0·0000	0·0000	–	1	0·0000	0·0000	0·0000	–

4, d_s/d_n ratios were typically less than 1 in *env* but greater than 1 in *gag*.

Evolutionary relationships between the mother and her two HIV-1 infected children

Neighbour-joining phylogenetic trees inferred from analyses of all *env* and *gag* sequences derived from the mother and her infected children are presented in Fig. 1. Sequences of the two children were observed to cluster distinctly from one another within both the *env* and *gag* trees (Fig. 1*a*, *b*) with specific sequences of the mother showing an association with sequences derived from the two children. When sequences of the two children were considered independently of those of the mother, 86% (*env*) and 91% (*gag*) bootstrap support was provided for the branch separating the two children. Likelihood tests of alternative phylogenetic hypotheses in which sequences of one child were clustered with sequences of the other child confirmed the significance of the separation of sequences from the two children for both the *env* ($P < 0·02$) and *gag* ($P < 0·02$) analyses.

(*a*) Child 3. The high levels of sequence diversity observed within child 3 immediately following birth were reflected by high degrees of sequence divergence within the inferred evolutionary trees. In *env*, two clearly distinct viral sequence lineages were identified within child 3 (Fig. 1*a*, groups A and B). The distinction of these two groups was supported in 89% of bootstrap replicates when sequences of child 3 were considered on their own. Bootstrap support within the overall tree was reduced by the placement of maternal sequences on the branch to child 3 group A (Fig. 1*a*, group A). Furthermore, likelihood tests of alternative phylogenetic hypotheses in which sequences of one child group were clustered with those of the other confirmed the significance of the two groups within the *env* tree ($P < 0·03$).

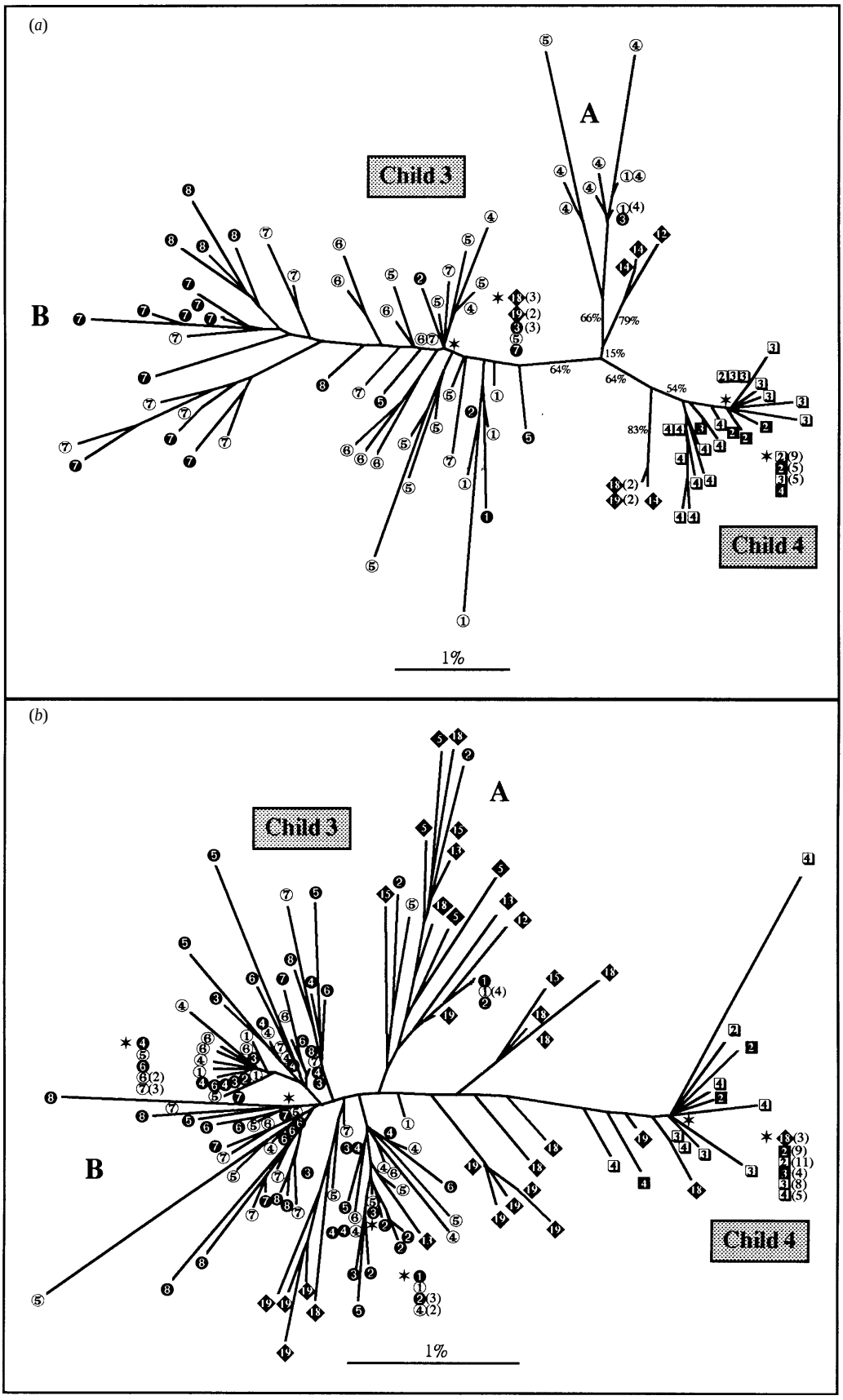


Fig. 1. For legend see facing page.

Child 3 group A formed the minor group of sequence variants circulating within child 3, with a single sequence obtained from the viral RNA population of sample 3R (approximately 4 months following birth) and 12 sequences obtained from the proviral DNA population in samples 1D (5), 4D (6) and 5D (1) (approximately 1, 5 and 11 months after birth respectively). The child 3 group A lineage did not persist within the child subsequent to these time-points and was not observed in the most recent RNA (8R) or DNA (7D) samples in *env* which were taken approximately 2 years after birth. Group A variants appeared to persist within the proviral DNA population marginally longer than in the plasma RNA population. Child 3 group A sequences were observed to cluster, albeit loosely, with maternal sequences obtained from samples taken across pregnancy (samples 14R and 12R taken approximately 2 and 5 months respectively before the birth of child 3). Nonetheless, this branch was not supported in bootstrap analyses and likelihood tests revealed that the maternal group could be associated alternatively with either child 3 lineage B or with child 4.

The major child group (Fig. 1*a*, group B) showed a very strong association with a number of maternal sequences identified in RNA samples 18R and 19R which were obtained just over 1 year following the birth of child 3. Indeed, the maternal sequences were identical at the nucleotide level to a number of child 3 sequences identified in RNA samples 3R (approximately 4 months) and 7R (approximately 22 months) and DNA sample 5D (approximately 11 months) of the child. The mother–child 3 group B cluster was supported in 59% of bootstrap replicates in the overall tree although when all sequences of child 3 (groups A and B) were considered with the group B maternal sequences alone the cluster was supported in 89% of bootstrap replicates (also observed when child 3 was considered independently). Likelihood tests of alternative phylogenetic hypotheses in which the maternal sequences associated with child 3 lineage B were clustered alternatively with lineage A confirmed the significance of the association of these maternal sequences with child 3 group B ($P < 0{\cdot}01$). The poor storage conditions of plasma samples obtained from the mother made the generation of viral RNA sequences from maternal samples difficult. However, in the four *env* sequences that were obtained from samples taken across pregnancy (samples 14R and 12R, taken approximately 2 and 5 months respectively before the birth of child 3) the group B lineage was not identified. Child 3 lineage B persisted until the most recent *env* RNA (8R) and DNA (7D) samples obtained approximately 2 years following birth. During this period, viral sequences of the major child 3 lineage showed a clear increase in genetic distance from the transmitted maternal group B variant with time.

Two lineages were also observed within the reconstructed *gag* gene phylogeny for child 3 (Fig. 1*b*, groups A and B). Bootstrap support for particular lineages within the *gag* tree was however low. A greater number of sequences were obtained from plasma samples of the mother in *gag* than within *env* and thus consequently the relationship between maternal sequences and sequences of the child becomes somewhat clearer. As in *env*, child 3 group A (Fig. 1*b*) consisted of sequences circulating within the plasma RNA and proviral DNA populations of the child earlier in the infection and did not persist until the most recent samples (8R and 8D, approximately 2 years following birth). The group consisted of four RNA sequences from time-points 1R and 2R (approximately 1 month and 2 months after birth respectively) and five DNA sequences from samples 1D and 5D (approximately 1 month and 11 months after birth respectively). Child 3 group A sequences persisted within the proviral DNA population longer than within the plasma RNA population and clustered with a number of maternal sequences from all sequenced maternal time-points (samples 5R, 12R, 13R, 15R taken approximately 2·5 years, 5 months, 3 months, 2 months, 2 weeks, respectively before the birth of child 3 and samples 18R and 19R taken just over 1 year after the birth of child 3). Group A formed the major variant present within the mother, particularly across the child 3 pregnancy.

The major group of child 3 sequences (Fig. 1*b*, group B) contained representatives from all sequenced time-points in *gag* (RNA and DNA samples 1 (approximately 1 month following birth) through 8 (approximately 2 years following birth). A number of maternal sequences obtained predominantly from samples taken a year after pregnancy (18R and 19R) clustered within the child group although the group B sequence population was represented in the mother in sample 13R (approximately 3 months before birth). Child 3 group B sequences in *gag* did not show the clear increase in diversity observed in *env* over the 2 year study period in the child.

(*b*) Child 4. Child 4 sequences, obtained from five samples taken between 1 and 6 months of age, clustered tightly within both the inferred *env* and *gag* phylogenies (Fig. 1*a*, *b*) as indicated by the intra-patient and intra-sample genetic diversities. In both the *env* and *gag* phylogenies, sequences of child 4 were associated with maternal sequences from time-points 18R and 19R, approximately 6 and 2 months respectively before the birth of child 4 (by comparison

Fig. 1. Unrooted neighbour-joining phylogenetic trees reconstructed from 124 *env* (*a*) and 198 p17 *gag* (*b*) gene sequences for the mother–child transmission set. Maternal sequences are represented by diamonds, sequences of child 3 by circles and sequences of child 4 by squares. Numbers within each symbol represent the sampling time-point. Filled symbols denote plasma RNA sequences; open symbols denote proviral DNA sequences. The scale bar corresponds to 1% nucleotide sequence divergence. Bootstrap values, based on 1000 bootstrap replications, are expressed as a percentage.

approximately 1 year after the birth of child 3). In *env*, the maternal sequences associated with child 4 clustered outside the child group. In *gag*, the child sequences fell on a maternal lineage consisting of sequences from time-points 18R and 19R and 3 *gag* sequences of time-point 18R were identical at the nucleotide level to the most common variant observed within the child.

(*c*) Mother. Maternal sequences thus appear to fall within three groups within both the *env* and *gag* phylogenies (Fig. 1*a*, *b*). Early sequences of the mother, including sequences obtained from samples taken during the child 3 pregnancy, clustered with a minor group of sequence variants within child 3 (child 3 lineage A, Fig. 1). Later sequences of the mother, obtained a year after the birth of child 3 and during the child 4 pregnancy, fell into two groups. One group was associated with the major group of sequence variants within child 3 (child 3 group B, Fig. 1). The final group of maternal sequences was associated with viral sequences of child 4 (child 4, Fig. 1). Only a small number of maternal sequences were obtained from a limited number of sampling time-points with sample sizes particularly small in *env*.

Amino acid sequence heterogeneity between mothers and children

Deduced *env* and *gag* amino acid sequences for the mother–child transmission set are presented in Fig. 2.

(*a*) *env* V3 loop and flanking regions. Three distinct amino acid sequence variants were identified within the deduced *env* amino acid sequences (Fig. 2*a*) which corresponded to the three lineages (child 3 lineages A and B, and the child 4 lineage) observed within the phylogenetic tree (Fig. 1*a*).

Child 3 lineage A could be differentiated from lineage B by the presence of specific amino acids at positions 66 and 67. An alanine (A) was observed at both these positions in child 3 lineage A with a glutamic acid (E) at position 66 and typically a glutamine (Q) at position 67 in child 3 lineage B. [Histidine (H) was observed at position 67 in a single child 3 lineage B sequence.]

Further, an alanine (A) and a lysine (K) were commonly observed within child 3 lineage A at positions 117 and 118 although these substitutions were not observed in all lineage A sequences. Arginine (R) was also fairly commonly observed within child 3 lineage B at position 84 but uncommon in child 3 lineage A (observed in one of nine sequences). Within the later child 3 lineage B samples, 7D, 7R and 8R (approximately 2 years post-infection), amino acid substitutions were commonly observed at positions 39 (asparagine, N), 40 (methionine, M) and 91 (lysine, K).

Child 4 was characterized by amino acid substitutions at positions 39 (proline, P), 73 (lysine, K) and 117 (alanine, A) which were observed in all child 4 sequences. These amino acids were fairly uncommon within the deduced amino acid sequences of child 3 and in no cases were all three amino acids observed within a single child 3 sequence. Threonine (T) at position 59 and aspartic acid (D) at position 121 were also commonly observed within the amino acid sequences of the earlier (1 to 4 months) child 4 samples, 2D, 2R and 3D.

Within the maternal *env* amino acid sequence population the earliest sequences obtained were closest to those observed within child 3 lineage A. Alanine (A) at position 66, characteristic of the child 3 lineage A sequence population, was observed within maternal sequences 12Ra and 14Ra, although the alanine (A) at position 67 observed in all child 3 lineage A sequences was not observed within these maternal sequences. The child 3 lineage B amino acid sequence was clearly identified in amino acid sequences 18Ra and 19Ra. These maternal amino acid sequence variants were identical to sequences 3Ra (2 sequences), 4Dc, 5Dd and 7Rg observed within child 3. Amino acid sequence patterns characteristic of child 4 were clearly identified in maternal sequences 14Rb, 18Rb and 19Rb. The maternal sequences did however contain an asparagine (N) at position 6 and a methionine (M) at position 9 which were not observed in the amino acid sequences of child 4. The GPGRAF motif at the crown of the V3 loop was conserved in all mother and child sequences.

Examination of putative *N*-linked glycosylation sites between the mother and her children revealed no pattern of selective loss of *N*-linked glycosylation sites between the mother and child 3 with the 10 glycosylation sites observed within the mother conserved between mother and child. Child 4 showed the loss of the glycosylation site at position 115 in all sequences obtained. This site was also absent in the maternal amino acid sequences of the variant associated with child 4. The *N*-linked glycosylation site at position 121 was also absent within the earlier (1 to 4 months) child 4 samples, 2D, 2R and 3D. Interestingly, the potential *N*-linked glycosylation site at position 7 was absent in the maternal sequences which were associated with child 4 but this site was present within all sequences obtained from the child.

The potential phenotype of the amino acid sequence variants was predicted on the basis of the global net charge of the V3 loop and the degree of sequence divergence from the LaRosa subtype B consensus (Milich *et al*., 1993; Donaldson *et al*., 1994). The majority of sequences from the mother and her children were predicted to be of the macrophage-tropic, non-syncytium-inducing (NSI) phenotype. However, a small number of sequences obtained from child 3 between 15 months and 2 years (6D, 7D and 7R) were predicted to be of the T-cell-tropic, syncytium-inducing (SI) phenotype. These variants were observed to show a positively charged lysine (K) at position 51 as opposed to a negatively charged aspartic acid (D) or glutamic acid (E), which are more commonly observed at this position.

(*b*) p17 *gag* region. The two lineages observed within child 3 were less clearly discernible within the deduced amino acid sequences of the p17 region of *gag* than within *env*. In general,

a threonine (T) was observed at position 107 in child 3 lineage B with either an alanine (A) or a serine (S) observed at this position in lineage A. There were however exceptions to this with the child 3 lineage B 1Dd sequence showing an alanine at this position.

Child 3 amino acid sequences could however be clearly differentiated from those of child 4 at positions 58 [threonine (T) child 3, alanine (A) child 4], 100 [glycine (G) child 3, glutamic acid (E) child 4] and 102 [serine (S) child 3, asparagine (N) child 4]. Position 107 was characterized typically by an alanine (A) or serine (S) in child 3 lineage A, an alanine (A) in child 4 and by contrast a threonine (T) at this position in child 3 lineage B.

Genetic relationships with other Edinburgh HIV-1 sequences

A phylogenetic analysis of sequences from the mother–child transmission set with representative sequences of identified HIV-1 risk groups circulating in Edinburgh (Holmes *et al.*, 1995; Leigh Brown *et al.*, 1997) revealed that the patients fell within the previously identified intravenous drug user (IDU) clade (data not shown). This is consistent with the available clinical information for the mother. An analysis of p17 *gag* gene sequences from the mother and her two children with all available Edinburgh IDU sequences and B subtype reference isolates (Fig. 3) revealed that all sequences from the mother and her children fell within a single group within the reconstructed phylogenetic tree and were not separated by sequences obtained from any other patient. The branch separating the sequences of the transmission set from other Edinburgh IDU patients was supported in 76% of bootstrap replicates. Sequences were not available from these Edinburgh IDU patients for comparison within *env*.

Discussion

We have analysed sequences spanning the V3 loop and flanking regions of the *env* gene and the p17 region of the *gag* gene from sequential plasma RNA populations obtained from a mother, and compared the viral characteristics of the maternal samples with both plasma RNA and proviral DNA populations obtained from her two vertically infected children.

Sequence variability and implications for transmission

(*a*) Evidence for the transmission of a single maternal variant. Substantial differences were observed in the level of genetic heterogeneity within the viral populations of the two infected children. Child 4 showed a highly homogeneous sequence population within both the sequenced *env* and *gag* regions, with sequences of the child clustering tightly within a single group in the reconstructed phylogenetic trees. Such viral homogeneity upon transmission is consistent with the infection of child 4 being initiated by a single maternal variant. Similar levels of sequence diversity have been reported in the newly infected child immediately following birth in the majority of mother-to-child transmission cases studied (Wike *et al.*, 1992; Wolinsky *et al.*, 1992; Mulder-Kampinga *et al.*, 1993; Scarlatti *et al.*, 1993; Ahmad *et al.*, 1995; Mulder-Kampinga *et al.*, 1995), a situation analogous to the restricted levels of sequence diversity observed in recently infected adults upon sero-conversion (McNearney *et al.*, 1990; Wolfs *et al.*, 1992; Zhang *et al.*, 1993; Zhu *et al.*, 1993). This has been interpreted as evidence for selection for particular viral variants from the heterogeneous pool present within the long-term infected transmitter. Although on their own such observations do not distinguish selection from a founder effect, evidence that mother-to-child transmission can be a selective process is provided by the fact that the transmitted variant has been observed to represent a minor form within the maternal sequence population upon transmission in the majority of studied mother-to-child transmission cases (Wike *et al.*, 1992; Wolinsky *et al.*, 1992; Mulder-Kampinga *et al.*, 1993, 1995; Ahmad *et al.*, 1995) although the non-selective transmission of major maternal forms has also been reported in some cases (Wike *et al.*, 1992; Wolinsky *et al.*, 1992; Scarlatti *et al.*, 1993). Based on the extremely limited number of sequences available, the maternal variant transmitted to child 4 appeared to be present in approximately 50% of the maternal sequence population during the child 4 pregnancy. The limited number of maternal sequences obtained do not allow us to make a reliable estimate of the timing of the transmission event on the basis of sequence data alone. Virus was not however detected in child 4 until 1 month following birth (a sample taken at 6 days was negative), which is perhaps more consistent with infection occurring either late in pregnancy or at delivery (Ehrnst *et al.*, 1991; De Rossi *et al.*, 1992).

(*b*) Evidence for the transmission of multiple sequence variants to the child. In contrast to child 4, substantial genetic heterogeneity was detected in child 3 within the first samples obtained following birth. The level of genetic diversity observed is considerably greater than diversity levels typically reported in patients, including vertically infected children, immediately following infection (McNearney *et al.*, 1990; Wike *et al.*, 1992; Wolfs *et al.*, 1992; Wolinsky *et al.*, 1992; Mulder-Kampinga *et al.*, 1993, 1995; Scarlatti *et al.*, 1993; Zhang *et al.*, 1993; Zhu *et al.*, 1993; Ahmad *et al.*, 1995). Phylogenetic analysis revealed the existence of two groups of sequence variants within child 3 which were separated by sequences obtained from the mother. The main child group (group B) showed a clear association with maternal sequences in both *env* and *gag*, with a number of maternal sequence variants obtained in *env* identical to sequences obtained from the child. Again maternal sequences were associated with the minor child 3 group (group A) in both *env* and *gag*, although the association was not as clearly defined within *env*. The

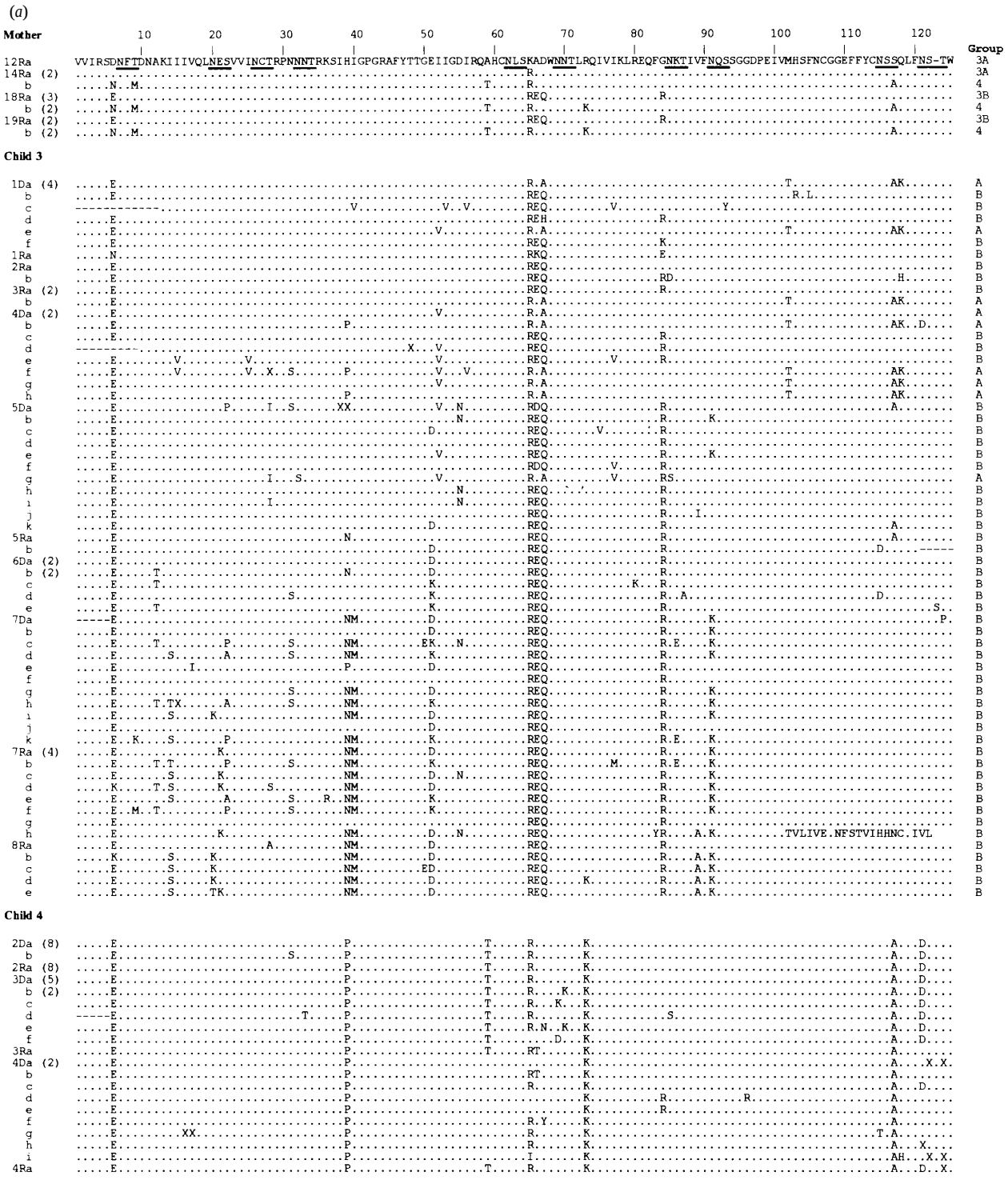
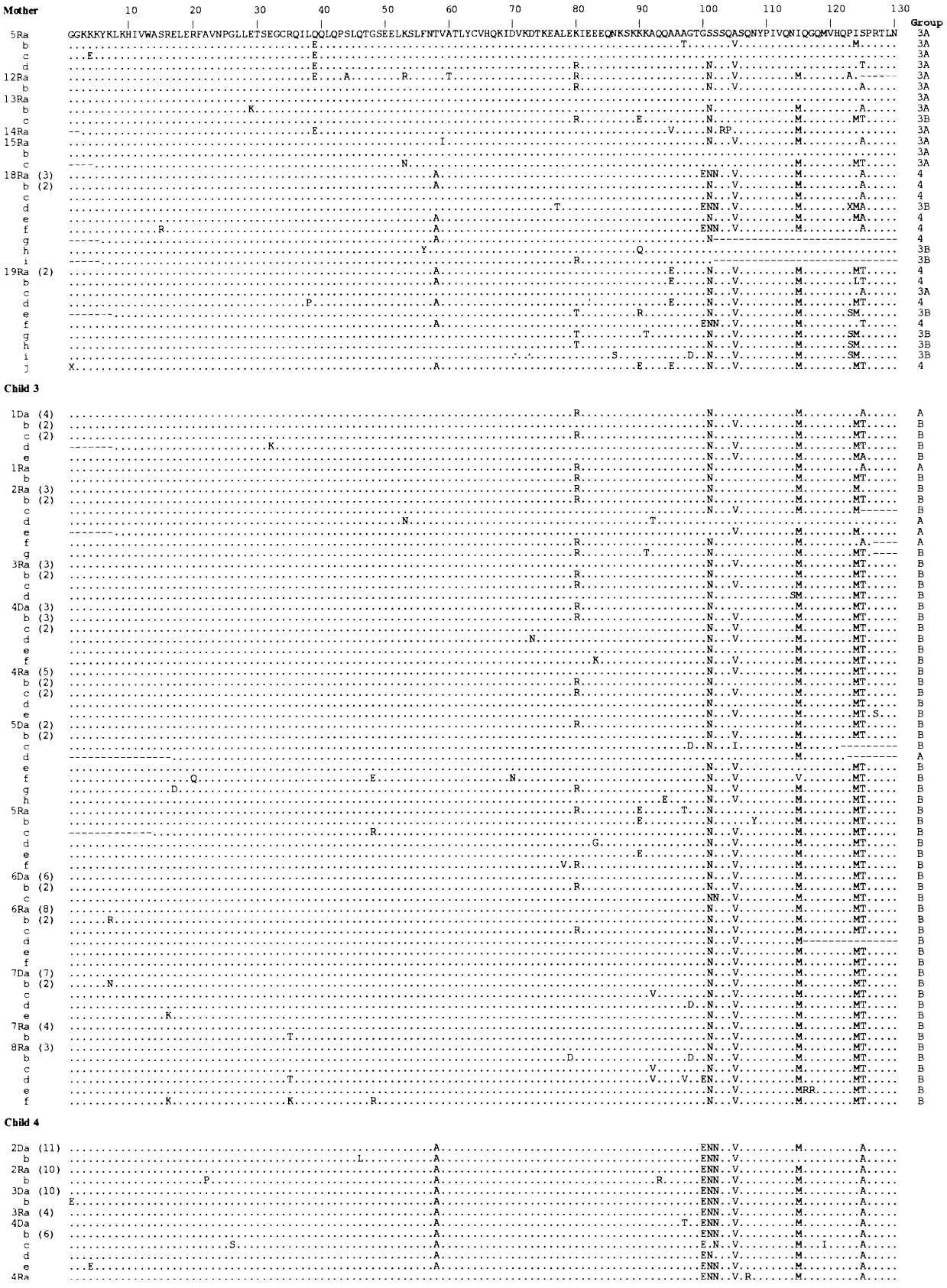


Fig. 2. Multiple alignments of deduced amino acid sequences of the V3 loop and flanking regions of the *env* gene (*a*) and the p17 region of the *gag* gene (*b*) for the mother–child transmission set. Sequences are presented in order of the time of sampling and are aligned relative to the most common variant present within the first sample of the mother. Amino acids identical to the first sequence are replaced by a dot, dashes represent alignment gaps and stop codons are represented by asterisks. X indicates unreadable bases in the nucleotide sequence at this position. Sequences are labelled according to sampling time-point. R denotes plasma RNA sequences; D denotes proviral DNA sequences. The number of identical amino acid sequences for a time-point are indicated in parentheses. Putative *N*-linked glycosylation sites are underlined. Sequences of child 3 were classified according to their phylogenetic grouping (lineages A and B, Fig. 1*a* and 1*b*). Sequences of the mother were classified according to their phylogenetic association with a particular child and a particular child lineage (Fig. 1*a* and 1*b*).

(*b*)



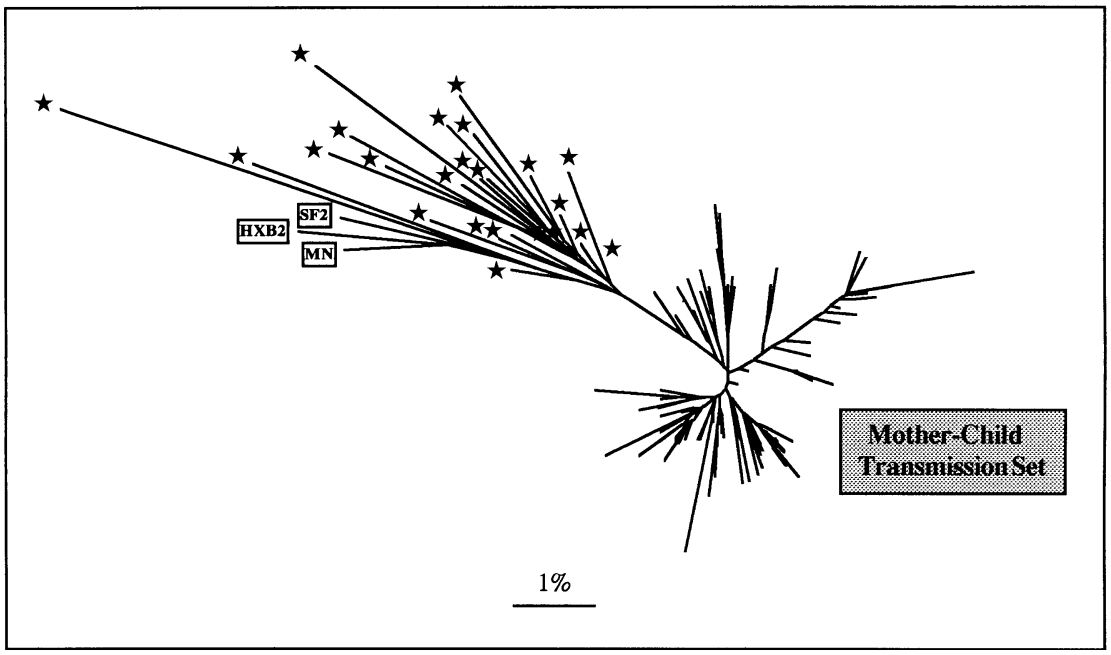


Fig. 3. Unrooted neighbour-joining p17 *gag* gene phylogeny showing the relationships between the 198 sequences obtained from the mother–child transmission set and sequences of 23 epidemiologically unrelated HIV-1 infected IDU patients from the Edinburgh cohort and 3 HIV-1 subtype B reference isolates. Sequences of the subtype B reference isolates are labelled directly on the tree; stars denote individual Edinburgh IDU sequences. Sequences obtained from the mother–child transmission set are not labelled individually. The scale bar corresponds to 1% nucleotide sequence divergence. Bootstrap values, based on 1000 bootstrap replications, are expressed as a percentage.

presence of both these groups within the first samples of the child, and their association with specific maternal sequence variants is indicative of the transmission of two distinct maternal variants to the child. Furthermore, the association of the two child lineages with predominantly early (across pregnancy, group A) or late (after pregnancy, group B) maternal variants is perhaps suggestive of transmission occurring at different times during pregnancy. Child 3 lineage A could possibly be the result of an early *in utero* transmission event with lineage B perhaps the result of transmission at delivery. The limited number of maternal sequences obtained for the transmission set do not however permit reliable inferences to be made regarding the time of transmission and it is possible that both lineages were present simultaneously, albeit with one lineage possibly present as a minor variant. Thus the results of the phylogenetic analysis for child 3 require either infection with multiple variants or transmission of virus on more than one occasion, possibly during pregnancy and at delivery. The transmission of more than one variant appears to be rare, although transmission of multiple variants has been reported in a small number of cases for both adults [one patient of the Florida dentist case (Ou *et al.*, 1992; Korber & Myers, 1992); the victim in the Swedish rape case (Albert *et al.*, 1993); evidence for coinfection with multiple HIV-1 strains within an Australian homosexual male (Zhu *et al.*, 1995); and with heterogeneous virus populations also reported shortly after seroconversion within cervical secretions and/or peripheral blood in five women of a cohort of six Kenyan female sex workers (Poss *et al.*, 1995)] and some vertically infected children (Lamers *et al.*, 1994; Van't Wout *et al.*, 1994; Briant *et al.*, 1995).

In conclusion, our analysis has provided evidence that mother-to-child transmission of HIV-1 is a complex process with different transmission processes occurring even within children born to a single mother.

We would like to thank Dr J. Mok (Regional Infectious Diseases Unit, City Hospital, Edinburgh) and Dr F. Johnstone (Department of Obstetrics and Gynaecology, Edinburgh Royal Infirmary, Edinburgh) for the provision of samples and clinical information. We would also like to thank Dr David Yirrell and Marian Aldhous for identifying patients and samples. This work was supported by the Medical Research Council AIDS Directed Programme.

References

Ahmad, N., Baroudy, B. M., Baker, R. C. & Chappey, C. (1995). Genetic analysis of human immunodeficiency virus type 1 V3 region isolates from mothers and infants after perinatal transmission. *Journal of Virology* **69**, 1001–1012.

Albert, J., Wahlberg, J. & Uhlén, M. (1993). Forensic evidence by DNA sequencing. *Science* **361**, 595–596.

Borkowsky, W., Krasinski, K., Pollack, H., Hoover, W., Kaul, A. & Ilmet-Moore, T. (1992). Early diagnosis of human immunodeficiency virus infection in children < 6 months of age: comparison of polymerase chain reaction, culture, and plasma antigen capture techniques. *Journal of Infectious Diseases* **166**, 616–619.

Briant, L., Wade, C. M., Puel, J., Leigh Brown, A. J. & Guyader, M. (1995). Analysis of envelope sequence variants suggests multiple mechanisms of mother-to-child transmission of human immunodeficiency virus type 1. *Journal of Virology* **69**, 3778–3788.

Burgard, M., Mayaux, M.-J., Blanche, S., Ferroni, A., Guihard-Moscato, M.-L., Allemon, M.-C., Ciraru-Vigneron, N., Firtion, G., Floch, C., Guilht, F., Lachassine, E., Vial, M., Griscelli, C. & Rouzioux, C. (1992). The use of viral culture and p24 antigen testing to diagnose human immunodeficiency virus infection in neonates. *New England Journal of Medicine* **327**, 1192–1197.

Courgnaud, V., Lauré, F., Brossard, A., Bignozzi, C., Goudeau, A., Barin, F. & Bréchot, C. (1991). Frequent and early in utero HIV-1 infection. *AIDS Research and Human Retroviruses* **7**, 83–88.

De Rossi, A., Ometto, L., Mammano, F., Zanotto, C., Giaquininto, C. & Chiecho-Bianchi, L. (1992). Vertical transmission of HIV-1: lack of detectable virus in peripheral blood cells of infected children at birth. *AIDS* **6**, 1117–1120.

Donaldson, Y. K., Bell, J. E., Holmes, E. C., Hughes, E. S., Brown, H. K. & Simmonds, P. (1994). *In vivo* distribution and cytopathology of variants of human immunodeficiency virus type 1 showing restricted sequence variability in the V3 loop. *Journal of Virology* **68**, 5991–6005.

Ehrnst, A., Lindgren, S., Dictor, M., Johansson, B., Sönnerborg, A., Czajkowski, J., Sundin, G. & Bohlin, A.-B. (1991). HIV in pregnant women and their offspring: evidence for late transmission. *Lancet* **338**, 203–207.

European Collaborative Study (1994). Caesarean section and risk of vertical transmission of HIV-1 infection. *Lancet* **343**, 1464–1467.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Felsenstein, J. (1993). PHYLIP manual version 3.52c. Berkeley University Herbarium, University of California, Berkeley.

Goedert, J. J., Duliège, A.-M., Amos, C. I., Felton, S., Biggar, R. J. & The International Registry of HIV-Exposed Twins (1991). High risk of HIV-1 infection for first-born twins. *Lancet* **338**, 1471–1475.

Higgins, D. G., Bleasby, A. J. & Fuchs, R. (1992). CLUSTAL V: improved software for multiple sequence alignment. *Computer Applications in the Biosciences* **8**, 189–191.

Holmes, E. C., Zhang, L. Q., Robertson, P., Cleland, A., Harvey, E., Simmonds, P. & Leigh Brown, A. J. (1995). The molecular epidemiology of human immunodeficiency virus type 1 in Edinburgh. *Journal of Infectious Diseases* **171**, 45–53.

Joviasis, E., Koch, M. A., Schäfer, A., Strauber, M. & Löwenthal, D. (1985). LAV/HTLV-III in 20-week fetus. *Lancet* **ii**, 1129.

Jukes, T. H. & Cantor, C. R. (1969). Evolution of protein molecules. In *Mammalian Protein Metabolism*, chapter 11, pp. 21–132. Edited by H. N. Munro. New York: Academic Press.

Kishino, H. & Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data and the branching order of the Hominoidea. *Journal of Molecular Evolution* **4**, 406–425.

Korber, B. & Myers, G. (1992). Signature pattern analysis: a method for assessing viral sequence relatedness. *AIDS Research and Human Retroviruses* **8**, 1549–1560.

Krivine, A., Firtion, G., Cao, L., Francoual, C., Henrion, R. & Lebon, P. (1992). HIV replication during the first weeks of life. *Lancet* **339**, 1187–1189.

Kumar, S., Tamura, K. & Nei, M. (1993). MEGA: Molecular Evolutionary Genetics Analysis, version 1.0. The Pennsylvania State University, University Park, PA 16802, USA.

Lamers, S. L., Sleasman, J. W., She, J. X., Barrie, K. A., Pomeroy, S. M., Barrett, D. J. & Goodenow, M. M. (1994). Persistence of multiple maternal genotypes of human immunodeficiency virus type 1 in infants by vertical transmission. *Journal of Clinical Investigation* **93**, 380–390.

Leigh Brown, A. J. & Simmonds, P. (1995). Analysis of HIV sequence variation. In *HIV – A Practical Approach*, Chapter 11, pp. 161–188. Edited by J. Karn. Oxford: Oxford University Press.

Leigh Brown, A. J., Lobidel, D., Wade, C. M., Rebus, S., Phillips, A. N., Brettle, R. P., France, A. J., Leen, C. S., McMenamim, J., McMillan, A., Maw, R. D., Mulcahy, F., Robertson, J. R., Sankar, K. N., Scott, G., Wyld, R. & Peutherer, J. F. (1997). The molecular epidemiology of human immunodeficiency virus type 1 in six cities in Britain and Ireland. *Virology* **235**, 166–177.

Lepage, P., Vande Perre, P., Carael, M., Nsengumuremyi, F., Nkurunziza, J. J., Butler, J.-P. & Sprecher, S. (1987). Postnatal transmission of HIV from mother to child. *Lancet* ii, 400.

McNearney, T., Westervelt, P., Thielan, B. J., Trowbridge, D. B., Garcia, J., Whittier, R. & Ratner, L. (1990). Limited sequence heterogeneity among biologically distinct human immunodeficiency virus type 1 isolates from individuals involved in a clustered infectious outbreak. *Proceedings of the National Academy of Sciences, USA* **87**, 1917–1921.

Milich, L., Margolin, B. & Swanstrom, R. (1993). V3 loop of the human immunodeficiency virus type 1 *env* protein: interpreting sequence variability. *Journal of Virology* **67**, 5623–5634.

Mulder-Kampinga, G. A., Kuiken, C., Scherpbier, H. J., Boer, K. & Goudsmit, J. (1993). Genomic human immunodeficiency virus type 1 RNA variation in mother and child following intra-uterine virus transmission. *Journal of General Virology* **74**, 1747–1756.

Mulder-Kampinga, G. A., Simonon, A., Kuiken, C. L., Dekker, J., Scherpbier, H. J., Van De Perre, P., Boer, K. & Goudsmit, J. (1995). Similarity in *env* and *gag* genes between genomic RNAs of human immunodeficiency virus type 1 (HIV-1) from mother and infant is unrelated to the time of HIV-1 RNA positivity in the child. *Journal of Virology* **69**, 2285–2296.

Newell, M.-L. (1995). Children and HIV infection. In *HIV Infection in Children. A Guide to Practical Management*, chapter 1, pp. 1–20. Edited by J. Y. Q. Mok & M.-L. Newell. Cambridge: Cambridge University Press.

Ou, C.-Y., Ciesieleski, C. A., Myers, G., Bandea, C. I., Luo, C.-C., Korber, B. T. M., Mullins, J. I., Schochetman, G., Berkelman, R. L., Economou, A. N., Witte, J. J., Furman, L. J., Satten, G. A., MacInnes, K. A., Curran, J. W., Jaffe, H. W., the Laboratory Investigation Group & the Epidemiological Investigation Group (1992). Molecular epidemiology of HIV transmission in a dental practice. *Science* **256**, 1165–1171.

Poss, M., Martin, H. L., Kreiss, J. K., Granville, L., Chohan, B., Nyange, P., Mandaliya, K. & Overbaugh, J. (1995). Diversity in virus populations from genital secretions and peripheral blood from women recently infected with human immunodeficiency virus type 1. *Journal of Virology* **69**, 8118–8122.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing evolutionary trees. *Molecular Biology and Evolution* **4**, 406–425.

Scarlatti, G., Leitner, T., Halapi, E., Wahlberg, J., Marchisio, P., Clerici-Schoeller, M. A., Wigzell, H., Fenyö, E. M., Albert, J., Uhlén, M. & Rossi, P. (1993). Comparison of variable region 3 sequences of human immunodeficiency virus type 1 from infected children with the RNA and DNA sequences of the virus populations of their mothers. *Proceedings of the National Academy of Sciences, USA* **90**, 1721–1725.

Simmonds, P., Balfe, P., Peutherer, J. F., Ludlam, C. A., Bishop, J. O. &

Leigh Brown, A. J. (1990). Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. *Journal of Virology* **64**, 864–872.

Smith, S. W., Overbeek, R., Woese, C. R., Gilbert, W. & Gillevet, P. M. (1994). The genetic data environment; an expandable GUI for multiple sequence analysis. *Computer Applications in the Biosciences* **10**, 671–675.

Soeiro, R., Rubinstein, A., Rashbaum, W. K. & Lyman, W. D. (1992). Materno–foetal transmission of AIDS: frequency of human immunodeficiency virus type 1 nucleic acids in sequences in human foetal DNA. *Journal of Infectious Diseases* **166**, 699–703.

Sprecher, S., Soumenkoff, G., Puissant, F. & Degueldre, M. (1986). Vertical transmission of HIV in 15-week fetus. *Lancet* ii, 228–289.

Staden, R. (1993). Staden package update. *Genome News* **13**, 12–13.

Van't Wout, A. B., Kootstra, N. A., Mulder-Kampinga, G. A., Albrecht van Lent, N. A., Scherpbier, H. J., Veenstra, J., Boer, K., Coutinho, R. A., Miedema, F. & Schuitemaker, H. (1994). Macrophage-tropic variants initiate human immunodeficiency virus type 1 infection after sexual, parenteral and vertical transmission. *Journal of Clinical Investigation* **94**, 2060–2067.

Weiser, B., Burger, H., Nachman, S., Hsu, Y. J. & Gibbs, R. (1993). Use of serial HIV-1 sequences from a pregnant woman and her twins to study timing of vertical transmission. In *Vaccines 1993, Modern Approaches to New Vaccines Including Prevention of AIDS*, pp. 161–188. Edited by H. S. Ginsberg, F. Brown, R. M. Chanock & R. A. Lerner. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Wike, C. M., Korber, B. T. M., Daniels, M. R., Hutto, C., Muñoz, J., Furtado, M., Parks, W., Saah, A., Bulterys, M., Kurawige, J.-B. & Wolinsky, S. M. (1992). HIV-1 sequence variation between isolates from mother-infant transmission pairs. *AIDS Research and Human Retroviruses* **8**, 1297–1300.

Wolfs, T. F. W., Zwart, G., Bakker, M. & Goudsmit, J. (1992). HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. *Virology* **189**, 103–110.

Wolinsky, S. M., Wike, C. M., Korber, B. T. M., Hutto, C., Parks, W. P., Rosenblum, L. L., Kunstman, K. J., Furtado, M. R. & Muñoz, J. L. (1992). Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. *Science* **255**, 1134–1137.

Zhang, L. Q., Simmonds, P., Ludlam, C. A. & Leigh Brown, A. J. (1991). Detection, quantification and sequencing of HIV-1 from the plasma of seropositive individuals and from factor VIII concentrates. *AIDS* **5**, 675–681.

Zhang, L. Q., MacKenzie, P., Cleland, A., Holmes, E. C., Leigh Brown, A. J. & Simmonds, P. (1993). Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. *Journal of Virology* **67**, 3345–3356.

Zhu, T., Mo., H., Wang, N., Nam, D. S., Cao, Y., Koup, R. A. & Ho, D. D. (1993). Genotypic and phenotypic characterization of HIV-1 in patients with primary infection. *Science* **261**, 1179–1181.

Zhu, T., Wang, N., Carr, A., Wolinsky, S. & Ho, D. D. (1995). Evidence for coinfection by multiple strains of human immunodeficiency virus type 1 subtype B in an acute seroconvertor. *Journal of Virology* **69**, 1324–1327.

Ziegler, J. B., Cooper, D. A., Johnson, R. & Gold, G. (1985). Postnatal transmission of AIDS associated retrovirus from mother to infant. *Lancet* i, 896–897.

Received 24 September 1997; Accepted 15 January 1998

† **Present Address:** Institute of Virology, University of Glasgow, Church Street, Glasgow G11 5JR, UK.