Sequence analysis of human T cell lymphotropic virus type I strains from southern India: gene amplification and direct sequencing from whole blood blotted onto filter paper

Vivek R. Nerurkar,[1]* P. George Babu,[2] Ki-Joon Song,[1] Rebecca R. Melland,[1] Chandran Gnanamuthu,[3] N. K. Saraswathi,[2] Mammen Chandy,[4] Mark S. Godec,[1] T. Jacob John[2] and Richard Yanagihara[1]

[1] *Laboratory of Central Nervous System Studies, National Institute of Neurological Disorders and Stroke, NIH, Bethesda, Maryland 20892, U.S.A. and* [2] *Departments of Virology,* [3] *Neurology and* [4] *Haematology, Christian Medical College Hospital, Vellore, Tamil Nadu 632004, India*

Human T cell lymphotropic virus type I (HTLV-I) infection in India has been found to be associated with adult T cell leukaemia/lymphoma (ATLL) and HTLV-I-associated myelopathy/tropical spastic paraparesis (HAM/TSP) among life-long residents of southern India. To examine the heterogeneity of HTLV-I strains from southern India and to determine their relationship with the sequence variants of HTLV-I from Melanesia, 1149 nucleotides spanning selected regions of the HTLV-I *gag, pol, env* and *pX* genes were amplified and directly sequenced from DNA extracted from whole blood blotted onto filter paper and from peripheral blood mononuclear cells, obtained from one patient with HAM/TSP, two with ATLL and eight asymptomatic carriers from Andhra Pradesh, Kerala and Tamil Nadu.

Sequence alignments and comparisons indicated that the 11 HTLV-I strains from southern India were 99·2% to 100% identical among themselves and 98·7% to 100% identical to the Japanese prototype HTLV-I ATK. The majority of base substitutions were transitions and silent. No frameshifts, insertions, deletions or possibly disease-specific base changes were found in the regions sequenced. The observed clustering of the Indian HTLV-I strains with those from Japan, as determined by the maximum parsimony method, suggested a common source of HTLV-I infection with subsequent parallel evolution. Amplification of DNA from blood specimens collected on filter paper may be useful for the study of other blood-borne pathogens.

A spastic paraplegic syndrome of possible viral aetiology was described in southern India more than 2 decades ago (Mani *et al.*, 1969), but the failure of recent attempts to demonstrate human T cell lymphotropic virus type I (HTLV-I) infection in such patients has prompted some investigators to conclude that HTLV-I has no aetiological role in the occurrence of spastic myelopathy in this region (Lalkaka *et al.*, 1988; Richardson *et al.*, 1989; Ravi *et al.*, 1992). Recently, however, the controversy over the occurrence of HTLV-I infection and disease in India (Advani *et al.*, 1987; Kelkar *et al.*, 1990) has been settled by finding HTLV-I to be associated with adult T cell leukaemia/lymphoma (ATLL) and HTLV-I-associated myelopathy/tropical spastic paraparesis (HAM/TSP) among inhabitants of southern and western India (Chandy *et al.*, 1991; Babu *et al.*, 1993; Singhal *et al.*, 1993). We have determined the genetic diversity of the Indian strains of HTLV-I and clarified their phylogenetic relationship with the sequence variants of HTLV-I from Melanesia (Yanagihara, 1993). This was

achieved by amplification using PCR and then direct sequencing of selected regions of the HTLV-I *gag, pol, env* and *pX* genes in DNA extracted from whole blood blotted onto filter paper and from uncultured peripheral blood mononuclear cells (PBMC) obtained from 11 virus-infected, life-long residents of Andhra Pradesh, Kerala and Tamil Nadu in southern India, all of whom denied blood transfusion, intravenous drug use or foreign travel. Eight of the HTLV-I-infected individuals were members of three unrelated families (Table 1).

Approximately 5 ml of whole blood, collected after obtaining informed consent, was blotted onto neonatal 903 filter paper (Schleicher & Schuell), allowed to air-dry for 8 to 16 h, then placed between two sheets of cardboard and sealed in individual plastic bags for transshipment at ambient temperature to the NIH. DNA was extracted from these samples within 3 weeks to 15 months after collection. Briefly, randomly selected regions of each filter paper-blotted blood sample were cut into six pieces (measuring approximately 1 mm²) and

Table 1. *Study participants and HTLV-I strains from southern India*

Study subject	Age and sex*	Relation	Birth and residence	Virus	Clinical status
1	52M	Husband	Kerala	HTLV-I CMCH 1	ATLL
2	44F	Wife	Kerala	HTLV-I CMCH 3	Asymptomatic
3	44F	Mother	Tamil Nadu	HTLV-I CMCH 5	Asymptomatic, HIV-positive
4	26F	Daughter	Tamil Nadu	HTLV-I CMCH 6	Asymptomatic
5	52M	Father	Andhra Pradesh	HTLV-I CMCH 7	HAM/TSP
6	48F	Mother	Andhra Pradesh	HTLV-I CMCH 8	Asymptomatic
7	17M	Son	Andhra Pradesh	HTLV-I CMCH 9	Asymptomatic
8	18M	Son	Andhra Pradesh	HTLV-I CMCH 12	Asymptomatic
9	45M	None	Tamil Nadu	HTLV-I CMCH 4	Asymptomatic, HIV-positive
10	35M	None	Tamil Nadu	HTLV-I CMCH 10	Asymptomatic
11	40F	None	Kerala	HTLV-I CMCH 13	ATLL

*M, male; F, female.

placed into 0·5 ml microfuge tubes containing 20 μl of GeneReleaser (Bioventures). Specimens were then overlaid with 40 μl of mineral oil, and DNA was released from the filter paper by denaturation and cooling in a thermocycler, using cycling conditions recommended by the manufacturer. This was then incubated at 80 °C before amplification. Alternatively, six 1 mm² pieces of filter paper-blotted blood, placed into 0·5 ml microfuge tubes containing 200 μl of 5% Chelex solution (Bio-Rad Laboratories) prepared in distilled water, were incubated at 57 °C for 20 min, then boiled for 10 min and centrifuged for 10 min at 15600 g. High M_r DNA was extracted using a non-organic method (Oncor) from Ficoll gradient-separated uncultured PBMC from one of the 11 study participants, a 40-year-old woman with ATLL from Kerala (Table 1, case 11).

Oligonucleotide primers for PCR and direct DNA sequencing were derived from sequences of the Japanese HTLV-I strain ATK (Seiki *et al.*, 1983) for the B cell immunodominant domain on the carboxy-terminal p19-encoding region of the *gag* gene (bases 1081 to 1100, 5′ ACTCATCCAAACCCAAGCCC 3′ and bases 1278 to 1257, 5′ GCCTGTAGGTCTTTCATTTGC 3′); the amino-terminal p24-encoding region of the *gag* gene (bases 1423 to 1444, 5′ CCATCACCAGCAGCTAGA-TAGC 3′ and bases 1560 to 1537, 5′ AGTTGCTGGT-ATTCTCGCCTTAAT 3′); the 3′ end of the *pol* gene (bases 4757 to 4778, 5′ CCCTACAATCCAACCAGC-TCAG 3′ and bases 4942 to 4919, 5′ GTGGTGAAGC-TGCCATCGGGTTTT 3′); the gp46-encoding region of the *env* gene (bases 5228 to 5246, 5′ TTTATTCTTCC-AGTTCTGC 3′ and bases 5596 to 5572, 5′ TAGGGG-CTGGAGACGGCTCCTGTAT 3′); the gp21-encoding region of the *env* gene (bases 6068 to 6087, 5′ TCATA-ACTCCCTCATCCTGC 3′ and bases 6481 to 6462, 5′ CAGCCAGTCAGGACTCGATT 3′); and the orf-II of the *pX* gene (bases 7358 to 7377, 5′ CGGATACCC-AGTCTACGTGT 3′ and bases 7516 to 7496, 5′ GAG-

CCGATAACGCGTCCATCG 3′). Reaction mixes were cycled as described previously (Nerurkar *et al.*, 1992).

Amplicons, purified using Centricon 100 microconcentrators (Amicon), were sequenced directly and in both directions using the *Taq* dye deoxynucleotide terminator cycle sequencing kit (Applied Biosystems) on an automated sequencer (model 373A, Applied Biosystems) (Nerurkar *et al.*, 1993). Sequence ambiguities were resolved by manual sequencing using the Sequenase version 2.0 DNA sequencing kit (U.S. Biochemicals). Sequence analysis was facilitated by using programs available on the VAX computer system, as part of the Genetics Computer Group (Devereux *et al.*, 1984).

Nucleotide and deduced amino acid sequence of the Indian HTLV-I strains were aligned and compared with those of HTLV-I strains ATK (Seiki *et al.*, 1983), H5 (Tsujimoto *et al.*, 1988), TSP-1 (Evangelista *et al.*, 1990) and MT-2 (Gray *et al.*, 1989, 1990) from Japan, HS-35 (Malik *et al.*, 1988) and CH (Paine *et al.*, 1991; Ratner *et al.*, 1991) from the Caribbean basin, pt3 from Brazil (Schulz *et al.*, 1991), ST from Chile (Dekaban *et al.*, 1992), EL from Zaire (Paine *et al.*, 1991; Ratner *et al.*, 1991), BEL 1 from the Polynesian Outlier Bellona (Nerurkar *et al.*, 1993), MEL 1 from Papua New Guinea (Nerurkar *et al.*, 1993) and MEL 5 from the Solomon Islands (Nerurkar *et al.*, 1993), as well as with sequences of simian T cell lymphotropic virus type I (STLV-I) strains PtM3 from a pig-tailed macaque (*Macaca nemestrina*) from Indonesia (Watanabe *et al.*, 1985), MM39-83 from a naturally infected rhesus macaque (*Macaca mulatta*) born in captivity (K.-J. Song, V. R. Nerurkar and R. Yanagihara, unpublished observations) and Tan90 from a tantalus monkey (*Cercopithecus aethiops* var. tantalus) from Central Africa (Saksena *et al.*, 1993), and sequences of human T cell lymphotropic virus type II (HTLV-II) strain Mo from the United States (Shimotohno *et al.*, 1985).

Since the DNA samples employed in PCR were

extracted directly from whole blood blotted onto filter paper and from uncultured PBMC, the proviral sequences presented here did not result from selection of virus populations from *in vitro* manipulation. Alignment and comparison between the Indian HTLV-I strains and virus strains from widely separated geographic regions indicated a high degree of sequence similarity with HTLV-I strains from Japan. Few base substitutions, primarily transitions, were found in each of the gene regions of the Indian HTLV-I strains (Fig. 1 *a* to *f*). The overall nucleotide sequence similarity between the 11 Indian HTLV-I strains and the prototype HTLV-I strain ATK from Japan was 98·7% to 100%, and the intrafamilial nucleotide sequence heterogeneity was nil to 0·1% among the eight Indian HTLV-I strains from three unrelated families. Neither termination codons, frameshifts, insertions nor deletions were identified in the 11 HTLV-I strains from southern India. In addition, no significant differences in nucleotide sequences were found between the virus strains from patients with ATLL or HAM/TSP and from asymptomatic carriers in southern India.

The greatest degree of diversity of the regions sequenced was found in the gp46 external glycoprotein-encoding region of the *env* gene (Fig. 1 *d* and Table 2). At base position 5415, an A→G transition was found in HTLV-I strains CMCH 1, CMCH 3 and CMCH 8, which was shared with HTLV-I strain TSP-1 from Japan and strain CH from the Caribbean and with STLV-I strains PtM3 and MM39-83 (Fig. 1 *d*). HTLV-I strains CMCH 1 and CMCH 3 which belonged to the same family were identical to each other except at position 5265 (T→G) and differed from all other Indian HTLV-I strains at position 5423 (T→C).

All base substitutions in each of the gene regions of the Indian HTLV-I strains were silent and synonymous when compared to the cosmopolitan HTLV-I strain ATK, except in the 108-amino acid gp46-encoding region of the *env* gene, in which HTLV-I strains CMCH 1, CMCH 3, CMCH 8 and CMCH 13 differed by two to four amino acids (1·9% to 3·7%) from HTLV-I ATK, specifically at amino acid positions 19 (Phe→Leu), 22 (Tyr→Asp), 72 (Ser→Gly) and 89 (Thr→Ile). The Phe→Leu change at position 19 and the Ser→Gly replacement at position 72 in the Indian strains of HTLV-I were shared with several HTLV-I strains from widely separated regions.

The amino-terminal neutralizing domain (composed of amino acids 88 to 98, TrpIleLysLysProAsnArgAsn-GlyGlyGly), encoded by bases 5463 to 5495, on the gp46 external envelope glycoprotein was totally conserved between the Indian HTLV-I strains and those HTLV-I strains from other widely separated geographical regions, including Melanesia, except for HTLV-I ATK, in which

the corresponding sequence was TrpThrLysLysProAsn-ArgAsnGlyGlyGly. The considerable sequence and functional conservation of the neutralizing domains on the external envelope glycoprotein among HTLV-I strains from widely separated geographical regions, including Melanesia (Melland, 1992), suggests that synthetic peptide- or recombinant protein-based subunit vaccines may offer protection against HTLV-I world-wide.

To ascertain the reliability of the derived topologies, phylogenetic trees were constructed using the unweighted pair-group method of assortment (UPGMA), which assumes a constant rate of evolutionary change, and the maximum parsimony method, which accommodates variable rates of change. The branch-and-bound search of the PAUP version 3.1 program (Swofford, 1993) was used to ensure that the true number of parsimonious trees was found. Six equally parsimonious trees, re-quiring 637 nucleotide substitutions, were constructed from the 325 bp gp46 and 374 bp gp21-encoding regions of the *env* gene. Bootstrap probabilities (as percentages) for the internal branches of the strict consensus tree, as determined for 1000 resampling using the PAUP pro-gram, were quite high (91% to 100%), demonstrating that the HTLV-I strains from Japan, India and the Caribbean formed a monophyletic group and that the Melanesian HTLV-I strains evolved independently of cosmopolitan strains (Fig. 2). The branch patterns of the UPGMA trees based on the same *env* gene sequences and on 1149 nucleotides spanning the *gag*, *pol*, *env* and *pX* genes were nearly identical to the consensus PAUP tree, using either STLV-I strain PtM3 or HTLV-II strain Mo as the outgroup (data not shown). Congruency of the phylogenetic trees, based on different gene regions by two algorithms, supported the evolutionary relationship between the HTLV-I strains from southern India and other geographic regions, including Melanesia.

Consistent with the data presented here, HTLV-I strains from two patients with HAM/TSP from western India have been shown to exhibit > 98% sequence similarity to the Japanese prototype HTLV-I ATK (Hashimoto *et al.*, 1993). Sequence analysis of HTLV-I strains obtained from Iranian-born Mashhadi Jews in Israel also indicate a possible genetic link with the Indian strains of HTLV-I (V. R. Nerurkar, K.-J. Song, A. Achiron and R. Yanagihara, unpublished). Phylogenetic networks based on a variety of genetic markers, including red cell enzymes and serum proteins, show a close evolutionary relationship between Asiatic Indians and Iranians (Saitou *et al.*, 1992).

Collectively, the data can be interpreted as being consistent with early migrations of human populations from the Middle East to India, as well as to Northeast Asia, dating more than 50000 years before the present

(a)

	1121 ... 1220
Japan	ATK CCACGGCCGCCGTCATCCCCCACCCACGACCCCCGGGATTCTGATCCACAAATCCCCCCTCCCTATGTTGAGCCTACGGCCCCCCAAGTCCTTCCAGTCA
	MT-2
	TSP-1
India	CMCH 1
	CMCH 4
	CMCH 5
	CMCH 13
Caribbean	CH
	HS-35
Zaire	EL
Bellona	BEL 1
Papua New Guinea	MEL 1
Solomon Islands	MEL 5
United States	Mo

(b)

	1445 ... 1536
Japan	ATK CTTATATCAGAGGCCGAAACCCGAGGTATTACAGGTTATAACCCATTAGCCGGTCCCCTCCGTGTCCCAAGCCAACAATCCACAACAACAAGG
	MT-2
	TSP-1
India	CMCH 1
	CMCH 3
	CMCH 5
	CMCH 7
	CMCH 8
	CMCH 9
	CMCH 12
	CMCH 10
	CMCH 13
Caribbean	CH
	HS-35
Chile	ST
Zaire	EL
Bellona	BEL 1
Papua New Guinea	MEL 1
Solomon Islands	MEL 5
Southeast Asia	MM39-83
Central Africa	Tan90
United States	Mo

(c)

	4779 ... 4848
Japan	ATK GACTTGTAGAACCCTCTAATGGCATTCTTAAAACCCTATTATATAAGTACTCTACTGACAAACCCGACCT
	MT-2
	TSP-1
India	CMCH 1
	CMCH 3
	CMCH 5
	CMCH 7
	CMCH 13
Caribbean	CH
	HS-35
Zaire	EL
Bellona	BEL 1
Papua New Guinea	MEL 1
Solomon Islands	MEL 5
Southern Asia	MM39-83
Central Africa	Tan90
United States	Mo
	4849 ... 4918
Japan	ATK ACCCATGGATAATGCTCTATCCATAGCCCTATGGACAATCAACCACCTGAATGTGTTAACCAACTGCCAC
	MT-2
	TSP-1
India	CMCH 1
	CMCH 3
	CMCH 5
	CMCH 7
	CMCH 13
Caribbean	CH
	HS-35
Zaire	EL
Bellona	BEL 1
Papua New Guinea	MEL 1
Solomon Islands	MEL 5
Southern Asia	MM39-83
Central Africa	Tan90
United States	Mo

(d)

	5247 ... 5355
Japan	ATK CCCCTCATCTTCGGTGATTACAGCCCCAGCTGCTGTACTCCCACAATTGGAGTCTCCTCATACCACTCTAAACCCTGCAATCCTGCCCAGCCAGTTTGTTCGTGGACCC
	MT-2
	H5
	TSP-1
India	CMCH 1
	CMCH 3
	CMCH 8
	CMCH 13
Caribbean	CH
	HS-35
Chile	ST
Brazil	pt3
Zaire	EL
Bellona	BEL 1
Papua New Guinea	MEL 1
Solomon Islands	MEL 5
Southern Asia	MM39-83
Indonesia	PtM3
United States	Mo
	5356 ... 5464
Japan	ATK TCGACCTGCTGGCCCCTTCAGCAGATCAGGCCCTACAGCCCCCCGCCCTAACCTAGTAAGTCACTGCAGCTACCATGCCACCTATTCCCCATATCTATTCCCTCATTG
	MT-2
	H5
	TSP-1
India	CMCH 1
	CMCH 3
	CMCH 8
	CMCH 13
Caribbean	CH
	HS-35
Chile	ST
Brazil	pt3
Zaire	EL
Bellona	BEL 1
Papua New Guinea	MEL 1
Solomon Islands	MEL 5
Southeast Asia	MM39-83
Indonesia	PtM3
United States	Mo
	5465 ... 5571
Japan	ATK GACTAAGAAGCCAAACCGAAATGGCGGAGGCTATTATTCAGCCCCTTATCAGACCCTTGCTCCTTAAAGTGCCCATACCTGGGGCGGCAATCATGGACCTGCCCCT
	MT-2
	H5
	TSP-1
India	CMCH 1
	CMCH 3
	CMCH 8
	CMCH 13
Caribbean	CH
	HS-35
Chile	ST
Brazil	pt3
Zaire	EL
Bellona	BEL 1
Papua New Guinea	MEL 1
Solomon Islands	MEL 5
Southeast Asia	MM39-83
Indonesia	PtM3
United States	Mo

(e)

```
                          6088                                                                                                    6212
Japan          ATK   CCCCCTTCTCCTTGTCACCTGTTCCCACCCTAGGATCCCGCTCCCGCCGAGCGGTACCGGTGGCGGTCTGGCTTGTCTCCGCCCTGGCCATGGGAGCCGGAGTGGCTGGCGGGATTACCGGCTCC
MT-2   ...........................................................................................G...........................
H5     ...........................................................................................................A...........
TSP-1  .....................................................................................................................
India          CMCH 1 .....................................................................................................................
CMCH 5 .....................................................................T...............................................
CMCH 13 ....................................................................................................................
Caribbean      CH     ...................................................................................................................
HS-35  ....................G.............A.....................................................A..........................
Brazil         pt3    ....................G..G...........................................................................................
Zaire          EL     .........................C..............T............A...........................A...............................
Bellona        BEL 1  ...T....C....C................................................G.........................
Papua New Guinea MEL 1 ...T....C....G...T..CA....T................A............A...........A...A...A.T.............T......
Solomon Islands MEL 5 .......C....C....GG...C.C....T.............A.........................A...GA.T.......A.............
Indonesia      PtM3   ........TC.........G...C....G....T.......G.A........C....T.........A.......A.....A.........
United States  Mo     ...T.....C.CG.T..C..A..TC.T.CG.CCA.AA.ACG.....T..C..T.AA.A..A..G...............A..GGCC...A.A..TA.C....T..AG.A..A....
```

```
                          6213                                                                                                    6337
Japan          ATK   ATGTCCCTCGCCTCAGGAAAGAGCCTCCTACATGAGGTGGACAAAGATACTTCCCAGTTAACTCAAGCAATAGTCAAAAACCACAAAAATCTACTCAAAATCGCGCAGTCATGCTGCCCAGAACAG
MT-2   ..............................T.....................A.................................................
H5     ..........................................................A.........................................
TSP-1  ..........................................................A.........................................
India          CMCH 1 ..........................................................A.........................................
CMCH 5 ..........................................................A.........................................
CMCH 13 .........................................................A.........................................
Caribbean      CH     .........................................................A.........................................
HS-35  .........................................................A...............................G.......A......
Brazil         pt3    .........................................................A.........................................
Zaire          EL     ..................G............A...........A..............A.........................................
Bellona        BEL 1  ................T.......................A..............A.........................................
Papua New Guinea MEL 1 .......................T.........A.....G........A...........T...............G...A..A.............
Solomon Islands MEL 5 ....................T.........A..A..............A.........T.................G...A..A.............
Indonesia      PtM3   ....................A..T.......A..A..............A..G..C.....................G....A..A......A....
United States  Mo     C.A..T..G..T..GA.T..A....T..C.TC.....T.........C..C.....CC.T..C..G..C..........T..TC...CA.C...CGUG....A.......A....A..T..
```

```
                          6338                                                                                                    6461
Japan          ATK   ACGAGGCCTTGATCTCCTGTTCTGGGAGCAAGGAGGATTATGCAAAGCATTACAAGAACAGTGCTGTTTTCTGAATATTACTAACTCCCATGTCTCAATACTACAAGAAAGACCCCCCCTTGAG
MT-2   ......................................................................A...C.....................
H5     .....................................................................................................
TSP-1  ...................A.................................................................................
India          CMCH 1 .....................................................................................................
CMCH 5 .................................................................C..................................
CMCH 13 ....................................................................................................
Caribbean      CH     .....................................................................................................
HS-35  ..........................................................A................................G...........A
Brazil         pt3    .....................................................................................................
Zaire          EL     G.......................................................G.............A...C..........................
Bellona        BEL 1  ............................................A.............A...C..............C..T..C............A....
Papua New Guinea MEL 1 ....C..................A.................G.GC.............A..C............C..T..C............A....
Solomon Islands MEL 5 ....C..................A.................G.GC.........A..C..............C..T..C............A....
Indonesia      PtM3   .......................T.............C.G....A..G..A..C..C..C..C..C.G..CA.T....A..CG.C..C..G...C.G....T...A
United States  Mo     ......AT.A..C....A......A.....G..T..G....G..CA....G..G..A..T..C..C..C..C..C.G..CA.T....A..CG.C..C..G...C.G....T...A
```

(f)

```
                          7378                                                                                7495
Japan          ATK   TTGGAGACTGTGTACAAGGCGACTGGTGCCCCATCTCTGGGGGACTATGTTCGGCCCGCCTACATCGTCACGCCCTACTGGCCACCTGTCCAGAGCATCAGATCACCTGGGACCCCAT
MT-2   ..A.........................................................................................
TSP-1  ..C.........................................................................................
India          CMCH 1 .......................................G...................................................
CMCH 3 .......................................G...................................................
CMCH 5 ...........................................................................................
CMCH 6 ...........................................................................................
CMCH 8 ...........................................................................................
CMCH 9 ...........................................................................................
CMCH 12 ..........................................................................................
CMCH 4 ...........................................................................................
CMCH 13 ..........................................................................................
Caribbean      CH     ...........................................................................................
HS-35  .......................T...................................................................
Chile          ST     .......................T...................................................................
Zaire          EL     ..............G........................T..................................................
Bellona        BEL 1  ...........................................................................................
Papua New Guinea MEL 1 A.........................................................................................
Solomon Islands MEL 5 ..........................................................................................
Southeast Asia MM39-83 ...........C.G.....................G......A...................................G........C....C
Indonesia      PtM3   ..........G........................C..........G..C..................C..G..............T...C
Central Africa Tan90  .............................................T................................C..G........T...C
United States  Mo     ....C..T........G.C...T....T...G....A..T..T........CA........A..C....C...................C..AC.............
```
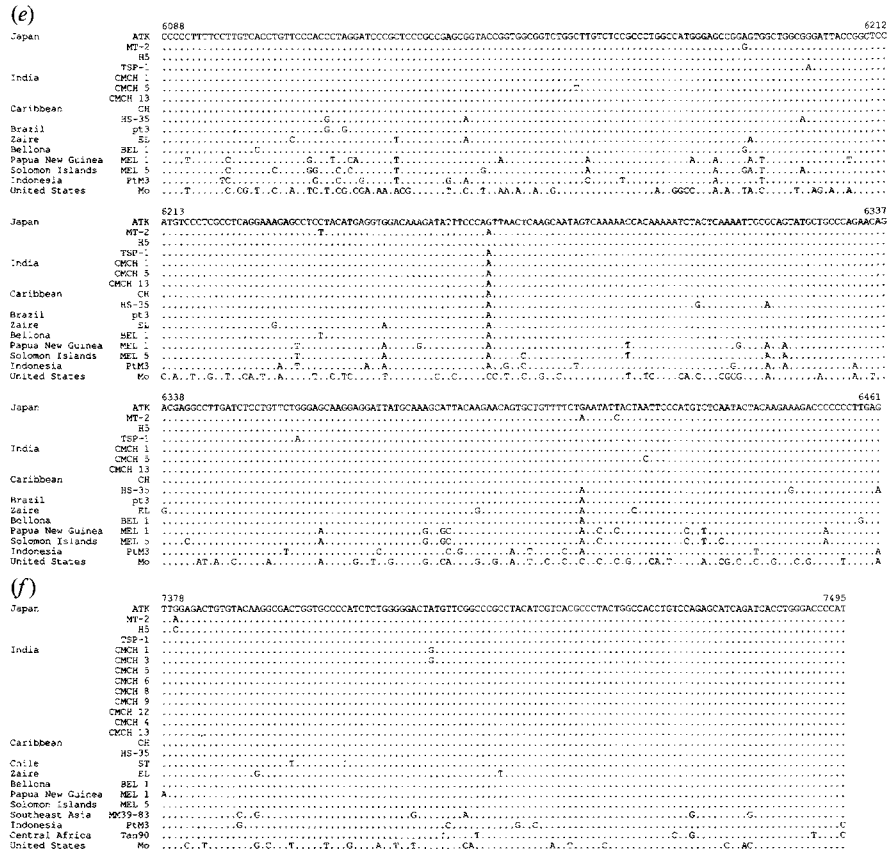
Fig. 1. Alignment and comparison of nucleotide sequences of (*a*) the 100 bp B cell immunodominant epitope-spanning p19-encoding region of the *gag* gene (bases 1121 to 1220), (*b*) the 92 bp p24-encoding region of the *gag* gene (bases 1445 to 1536), (*c*) the 140 bp 3′ end of the *pol* gene (bases 4779 to 4918), (*d*) the 325 bp 5′ end of the gp46-encoding region of the *env* gene (bases 5247 to 5571), (*e*) the 374 bp gp21-encoding region of the *env* gene (bases 6088 to 6463) and (*f*) the 118 bp tax-encoding region of the *pX* gene (bases 7378 to 7495) in HTLV-I strains from southern India. The B cell epitopes on the carboxy-terminal p19 matrix protein-encoding *gag* gene and on the amino-terminal gp46 external envelope glycoprotein-encoding *env* gene are underlined. For comparison, corresponding sequences are shown for HTLV-I strains from Japan (ATK, MT-2, H5, TSP-1), the Caribbean (CH, HS-35), Chile (ST), Brazil (pt3), Zaire (EL), the Polynesian Outlier Bellona (BEL 1), Papua New Guinea (MEL 1) and the Solomon Islands (MEL 5), as well as for Asian and African STLV-I strains (PtM3, MM39-83, Tan90) and for an HTLV-II strain (Mo) from the United States.

Table 2. *Nucleotide sequence comparison of HTLV-I strains from southern India, Japan and Melanesia with a cosmopolitan strain of HTLV-I (ATK) from Japan, based on selected regions of the* gag, pol, env *and* pX *genes*

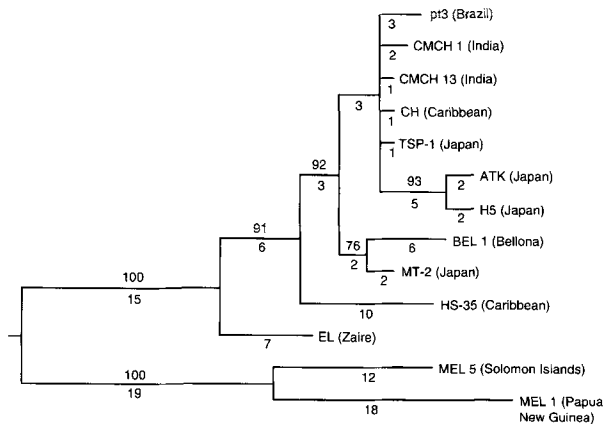| Country of origin	HTLV-I strain	Percentage divergence from HTLV-I ATK					
		p19*gag*	p24*gag*	*pol*	gp46*env*	gp21*env*	*pX*-II
Japan	MT-2	1	2·2	0	1·5	1·3	0·8
	TSP-1	2	1·1	0·7	1·8	0·8	0
India	CMCH 1	1	1·1	0	1·8	0·3	0·8
	CMCH 3	–	1·1	0	2·2	–	0·8
	CMCH 5	0	1·1	0	–	0·8	0
	CMCH 6	–	–	–	–	–	0
	CMCH 7	–	1·1	0·7	–	–	–
	CMCH 8	–	1·1	–	1·8	–	0
	CMCH 9	–	1·1	–	–	–	0
	CMCH 12	–	1·1	–	–	–	0
	CMCH 4	1	–	–	–	–	0
	CMCH 10	–	1·1	–	–	–	–
	CMCH 13	0	1·1	0·7	1·5	0·3	0
Papua New Guinea	MEL 1	3	8·7	9·3	6·8	8·6	0·8
Solomon Islands	MEL 5	4	7·6	7·1	7·7	8·6	0

Fig. 2. Strict consensus phylogenetic tree, constructed by the maximum parsimony method, based on both the 325 bp gp46- and the 374 bp gp21-encoding regions of the *env* gene of HTLV-I strains from southern India (CMCH 1, CMCH 13) and representative HTLV-I strains from other geographical regions, including Japan (ATK, MT-2, TSP-1, H5), the Caribbean basin (CH, HS-35), Brazil (pt3), Zaire (EL), the Polynesian Outlier Bellona (BEL 1), Papua New Guinea (MEL 1) and the Solomon Islands (MEL 5). The tree was rooted by assuming STLV-I strain PtM3 from Indonesia as the outgroup. Branch lengths, drawn in proportion to the number of nucleotide substitutions, are given below each branch, and bootstrap probabilities (%), calculated from 1000 resamplings by PAUP, are given above the internal branches.

(Nei & Roychoudhury, 1993) and with more recent migrations from the Middle East to Kerala in southern India and to Gujarat in western India, approximately 1000 to 1300 years ago (Undevia *et al.*, 1972; Balkrishan, 1978). More recently, visits by Tamils from southern India to southwestern Japan, albeit unverified, are presumed to have occurred during the 16th century (Kantha, 1986). Admittedly the hypothesis that these ancient or more recent population movements and resettlements account for the close genetic similarities between HTLV-I strains in India, Japan and the Middle East is speculative, but not implausible, especially when viewed within the context of the likely spread of HTLV-I to parts of the New World by the African slave trade, beginning in the 16th century (Gessain *et al.*, 1992).

The origin of HTLV-I in India and the extent to which strains of STLV-I contributed to the emergence of HTLV-I in India are not certain. However, the demonstrated marked sequence divergence of more than 10% between the Indian HTLV-I strains and STLV-I strains from rhesus and bonnet macaques from India (K.-J. Song, V. R. Nerurkar, T. Ishida and R. Yanagihara, unpublished) make it less likely that the HTLV-I strains in India evolved recently from STLV-I harboured by Asian macaques. Further molecular genetic studies of HTLV-I strains from elsewhere in India and from the offshore islands in the Indian Ocean, such as Reunion and the Seychelles, and of STLV-I isolates from other

non-human primates in India and Sri Lanka may provide additional insights into the emergence and early dissemination of the primate T cell lymphotropic viruses.

As in other HTLV-I-endemic regions, HTLV-I infection in southern India tended to cluster in families. In one of the Indian families we studied, consisting of a railway guard with HAM/TSP and his asymptomatically infected wife and two sons from Andhra Pradesh, HTLV-I infection was probably transmitted from the patient to his wife by sexual intercourse and the two sons probably acquired infection during infancy via breast feeding (Babu *et al.*, 1992, 1993). The genomic identity of HTLV-I strains from members of the same Indian family is not unexpected and is consistent with the high degree of sequence similarity among HTLV-I strains from infected families in Martinique (Gessain *et al.*, 1992), Zaire (Goubau *et al.*, 1992), Papua New Guinea and the Solomon Islands (Nerurkar *et al.*, 1993).

The logistics of transporting blood samples over long distances from the patients' residences to the laboratory in Vellore and the unavailability of laboratory instrumentation and supplies for adequately processing blood specimens, during the early phases of this study, precluded any possibility of separating PBMC by Ficoll gradient centrifugation for DNA isolation. Thus, we had to resort to extracting DNA from filter paper-blotted whole blood to characterize nearly all of the HTLV-I strains from southern India reported here. However, as evidenced by PCR and subsequent sequencing, the quality of the DNA extracted from filter paper-blotted blood stored at room temperature for 3 weeks to 15 months, using either GeneReleaser or Chelex, was as high as that obtained from Ficoll-gradient-separated uncultured PBMC. Thus, gene amplification from filter paper-blotted blood and subsequent direct sequencing permit rapid genetic analysis of HTLV-I strains from remote populations.

References

ADVANI, S. H., FUJISHITA, M., KITAGAWA, T., TAGUCHI, H. & MIYOSHI, T. (1987). Absence of HTLV-I infection in India – a preliminary report. *Indian Journal of Medical Research* 86, 218–220.

BABU, P. G., SARASWATHI, N. K., JOHN, T. J., ISHIDA, T., IMAI, J., MURPHY, E. & VARNEY, K. (1992). Sexual transmission of HTLV infections in southern India. *Journal of Acquired Immunodeficiency Syndromes* 5, 317.

BABU, P. G., GNANAMUTHU, C., SARASWATHI, N. K., NERURKAR, V. R., YANAGIHARA, R. & JOHN, T. J. (1993). HTLV-I-associated myelopathy in southern India. *AIDS Research and Human Retroviruses* 9, 499–500.

BALKRISHNAN, V. (1978). A preliminary study of genetic distances among some populations of the Indian sub-continent. *Journal of Human Evolution* 7, 67–75.

CHANDY, M., BABU, P. G., SARASWATHI, N. K., ISHIDA, T. & JOHN, T. J. (1991). HTLV-1 infection in patients with leukaemia in southern India. *Lancet* 338, 3810–3811.

DEKABAN, G. A., KING, E. E., WATERS, D. & RICE, G. P. A. (1992). Nucleotide sequence analysis of an HTLV-I isolate from a Chilean patient with HAM/TSP. *AIDS Research and Human Retroviruses* 8, 1201–1207.

DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* 12, 387–395.

EVANGELISTA, A., MAROUSHEK, S., MINNIGAN, H., LARSON, A., RETZEL, E., HAASE, A., GONZALEZ-DUNIA, D., McFARLIN, D., MINGIOLI, E., JACOBSON, S., OSAME, M. & SONODA, S. (1990). Nucleotide sequence analysis of a provirus derived from an individual with tropical spastic paraparesis. *Microbial Pathogenesis* 8, 259–278.

GESSAIN, A., GALLO, R. C. & FRANCHINI, G. (1992). The low degree of HTLV-I genetic drift *in vivo* as a means to follow viral transmission and movement of ancient human populations. *Journal of Virology* 66, 2288–2295.

GOUBAU, P., KAZADI, K., CARTON, H., VANDAMME, A., LIU, H. F. & DESMYTER, J. (1992). The epidemiology of human T-cell lymphotropic viruses in Zaire. (abstract) *American Journal of Tropical Medicine and Hygiene* 47, 258S.

GRAY, G. S., BARTMAN, T. & WHITE, M. (1989). Nucleotide sequence of the core (*gag*) gene from HTLV-I isolate MT-2. *Nucleic Acids Research* 17, 7998.

GRAY, G. S., WHITE, M., BARTMAN, T. & MANN, D. (1990). Envelope gene sequence of HTLV-I isolate MT-2 and its comparison with other HTLV-I isolates. *Virology* 177, 391–395.

HASHIMOTO, K., LALKAKA, J., FUJISAWA, J., SINGHAL, B.S., MACHIGASHIRA, K., KUBOTA, R., SUEHARA, M., OSAME, M. & YOSHIDA, M. (1993). Limited sequence divergence of HTLV-I of Indian HAM/TSP patients from a prototype Japanese isolate. *AIDS Research and Human Retroviruses* 9, 495–498.

KANTHA, S. S. (1986). Portuguese role in spread of HTLV-I virus. *Nature, London* 321, 733.

KELKAR, R., ISHIDA, T., BHARUCHA, Z., ADVANI, S. H. & HAYAMI, M. (1990). A seroepidemiological survey of HTLV-I in blood donors in India. *Indian Journal of Haematology* 8, 11–14.

LALKAKA, J. A., SAVANT, C. V. & SINGHAL, B. S. (1988). HTLV-I antibody study in non-compressive myelopathies. (Abstract) *Abstracts, Neurological Society of India.*

MALIK, K. T. A., EVEN, J. & KARPAS, A. (1988). Molecular cloning and complete nucleotide sequence of an adult T cell leukaemia virus/ human T cell leukaemia virus type I (ATLV/HTLV-I) isolate of Caribbean origin: relationship to other members of the ATLV/ HTLV-I subgroup. *Journal of General Virology* 69, 1695–1710.

MANI, K. S., MANI, A. J. & MONTGOMERY, R. D. (1969). A spastic paraplegic syndrome in southern India. *Journal of Neurological Sciences* 9, 179–199.

MELLAND, R. R. (1992). *Amino acid sequence conservation of three immunodominant gag and env epitopes of human T lymphotropic virus type I from Melanesian families.* M.Sc. dissertation, Pennsylvania State University.

NEI, M. & ROYCHOUDHURY, A. K. (1993). Evolutionary relationships of human populations on a global scale. *Molecular Biology and Evolution* (in press).

NERURKAR, V. R., MILLER, M. A., LEON-MONZON, M. E., AJDU-KIEWICZ, A. B., JENKINS, C. L., SANDERS, R. C., GODEC, M. S., GARRUTO, R. M. & YANAGIHARA, R. (1992). Failure to isolate human T lymphotropic virus type I and to detect variant-specific genomic sequences by polymerase chain reaction in Melanesians with indeterminate Western immunoblot. *Journal of General Virology* 73, 1805–1810.

NERURKAR, V. R., SONG, K.-J., SAITOU, N., MELLAND, R. R. & YANAGIHARA, R. (1993). Interfamilial and intrafamilial genomic diversity and molecular phylogeny of human T-cell lymphotropic virus type I from Papua New Guinea and the Solomon Islands. *Virology* 196, 506–513.

PAINE, E., GARCIA, J., PHILPOTT, T. C., SHAW, G. & RATNER, L. (1991). Limited sequence variation in human T-lymphotropic virus type I isolates from North American and African patients. *Virology* 182, 111–123.

RATNER, L., PHILPOTT, T. & TROWBRIDGE, D. B. (1991). Nucleotide sequence analysis of isolates of human T-lymphotropic virus type I of diverse geographic origins. *AIDS Research and Human Retroviruses* 7, 923–941.

RAVI, V., GOURIE-DEVI, M., VENKATESH, A., SATISHCHANDRA, P., SHENOY, P. K., KALYANARAMAN, V. S. & ROMAN, G. C. (1992). HTLV-I antibody screening in tropical myeloneuropathies – a study from South India. *Abstracts, Fifth International Conference on Human Retrovirology, Kumamoto, Japan.*

RICHARDSON, J. H., NEWELL, A. L., NEWMAN, P. K., MANI, K. S., RANGAN, G. & DALGLEISH, A. G. (1989). HTLV-I and neurological disease in southern India. *Lancet* i, 1079.

SAITOU, N., TOKUNAGA, K. & OMOTO, K. (1992). Genetic affinities of human populations. In *Isolation, Migration and Health*, pp. 118–129. Edited by D. F. Roberts, N. Fujiki & K. Torizuka. Cambridge: Cambridge University Press.

SAKSENA, N. K., HERVÉ, V., SHERMAN, M. P., DURAN, J. P., MATHIOT, C., MÜLLER, M., LOVE, J. L., LEGUENNO, B., BARRÉ SINOUSSI, F., DUBE, D. K. & POIESZ, B. J. (1993). Sequence and phylogenetic analyses of a new STLV-I from a naturally infected tantalus monkey from Central Africa. *Virology* 192, 312–320.

SCHULZ, T. F., CALABRÓ, M.-L., HOAD, J. G., CARRINGTON, C. V. F., MATUTES, E., CATOVSKY, D. & WEISS, R. A. (1991). HTLV-I envelope sequences from Brazil, the Caribbean, and Romania: clustering of sequences according to geographic origin and variability in an antibody epitope. *Virology* 184, 483–491.

SEIKI, M., HATTORI, S., HIRAYAMA, Y. & YOSHIDA, M. (1983). Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. *Proceedings of the National Academy of Sciences, U.S.A.* 80, 3618–3622.

SHIMOTOHNO, K., TAKAHASHI, Y., SHIMIZU, N., GOJOBORI, T., GOLDE, D. W., CHEN, I. S. Y., MIWA, M. & SUGIMURA, T. (1985). Complete nucleotide sequence of an infectious clone of human T-cell leukemia virus type II: an open reading frame for the protease gene. *Proceedings of the National Academy of Sciences, U.S.A.* 82, 3101–3105.

SINGHAL, B., LALKAKA, J. A., SONODA, S., HASHIMOTO, K., NOMOTO, M., KUBOTA, R. & OSAME, M. (1993). Human T-lymphotropic virus type I infections in western India. *AIDS* 7, 138–139.

SWOFFORD, D. L. (1993). *PAUP: Phylogenetic Analysis Using Parsimony*, version 3.1. Computer program distributed by the Illinois Natural History Survey, Champaign, Illinois.

TSUJIMOTO, A., TERUUCHI, T., IMAMURA, J., SHIMOTOHNO, K., MIYOSHI, I. & MIWA, M. (1988). Nucleotide sequence analysis of a provirus derived from HTLV-I-associated myelopathy (HAM). *Molecular Biology and Medicine* 5, 29–42.

UNDEVIA, J. V., BLAKE, N. M., KIRK, R. L. & McDERMID, E. M. (1972). The distribution of some enzyme group systems among Parsis and Iranis in Bombay. *Human Heredity* 22, 274–282.

WATANABE, T., SEIKI, M., TSUJIMOTO, H., MIYOSHI, I., HAYAMI, M. & YOSHIDA, M. (1985). Sequence homology of the simian retrovirus genome with human T-cell leukemia virus type I. *Virology* 144, 59–65.

YANAGIHARA, R. (1993). Geographic-specific genotypes or topotypes of human T-cell lymphotropic virus type I as markers for early and recent migrations of human populations. *Advances in Virus Research* (in press).