Phylogenetic Analysis of the p24-p7 Region of the Human Immunodeficiency Virus Type 1 *gag* Gene To Determine Subtype Distribution among Female Sex Workers in Calcutta, India

Satarupa Sengupta,¹ Smarajit Jana,² Pratim Roy,² Kamalesh Sarkar,¹
Sujit K. Bhattacharya,¹ and Sekhar Chakrabarti^{1*}

HIV/AIDS Laboratory, National Institute of Cholera and Enteric Diseases,¹ and STD/HIV Intervention Project,² Calcutta, India

Received 13 May 2005/Returned for modification 2 August 2005/Accepted 12 August 2005

Human immunodeficiency virus type 1 (HIV-1) subtype C, based on the envelope region, has been reported to be predominant in India. We sequenced the p24-p7 *gag* region from 51 HIV-1 seropositive female sex workers in Calcutta, India, for more-detailed molecular characterization. Subtype C was found to be prevalent, although no strong monophyletic cluster was observed.

Since the first report of human immunodeficiency virus type 1 (HIV-1) infection in 1986, India has been experiencing a rapid spread, primarily through heterosexual transmission (15, 17). The magnitude and complexity of HIV-1 genetic diversity are one of the major challenges for vaccine development (8, 19, 21). Earlier reports from India identified the preponderance of subtype C (10, 18, 20) and small proportions of subtypes A and Thai B of HIV-1 (1, 12). Subsequent reports identifying multiple subtypes and recombinants suggest new introductions and/or their detection due to extended screening (13). The specific region within the envelope gene was used to determine HIV-1 subtypes in India (14, 16, 18). Besides the C2-V3 envelope region (5, 7, 9), the *gag* p24-p7 region has become important to the study of the genetic diversity (4, 9, 11, 22) of HIV-1, since the development of a *gag*-based heteroduplex mobility assay (HMA) (3, 9), followed by sequencing and phylogenetic analysis (6, 22). Based on the C2-V3 and C2-V5 envelope regions, we previously identified subtype C viruses among female sex workers in Calcutta, the most populous city in the eastern part of India (14). HIV-1 subtype determination studies in India based on the *gag* gene are very limited, and so far very few samples have been studied (12). In this study, we report the distribution of HIV-1 subtypes on the basis of the *gag* p24-p7 region.

Three hundred twenty female sex workers (in an age group of 20 to 35 years) in Calcutta were included in an unlinked anonymous study during the years 2003 and 2004 following the ethical guidelines. Blood samples were collected in Na-citrate solution after obtaining informed consent and pre- and post-test counseling. HIV-1 seropositivity was determined by rapid spot test (Immunocomb HIV-1/2 Bi-spot; Organics, Israel), followed by enzyme-linked immunosorbent assay (Immunogenetics, Belgium) and line immunoassay (Inno-LIA HIV-1/

HIV-2). Fifty-one out of 320 samples (16%) were HIV-1 positive. The study population never received any antiretroviral therapy, per their interviews with the research team. (The government of India has started antiretroviral therapy in high-risk states of India in accordance with the WHO 3 by 5 program. As West Bengal [Calcutta is the state capital] is a low-prevalence [in the true sense, concentrated-epidemic] state, antiretroviral therapy has just been initiated.) Peripheral blood mononuclear cells (PBMC) were separated from whole blood by Ficoll-Hypaque gradient centrifugation (2), and the DNA was extracted by using the QIAamp DNA blood mini kit 250 (QIAGEN, Germany) according to the manufacturer's protocol. The HIV-1 DNA fragment comprising a 460-bp *gag* gene fragment corresponding to the region from amino acid 132 of p24 to amino acid 40 of p7 was amplified by nested PCR in a thermal cycler (GeneAmp PCR system 2400; Perkin Elmer). Primers used for the amplification were H1G777, 5'TCACCTAGAAGTTTGAATGCATGGG3' (outer forward); H1P202, 5'CTAATACTGTATCATCTGCTCCTGT3' (outer reverse); H1Gag1584, 5'AAAGATGGATAATCCTGGG3' (inner forward); and G17, 5'TCCACATTTCCAACAGCCCTTTT3' (inner reverse).

PBMC DNA (1 µg) was used as a template for PCR in the presence of 1.5 mM MgCl₂, 0.2 mM deoxynucleoside triphosphates (Perkin Elmer), 10 pmol of each primer, and 2.5 U of *Taq* DNA polymerase (Ampli *Taq* Gold; Perkin Elmer) in a total volume of 50 µl. PCR conditions followed were 94°C for 2 min, 35 cycles consisting of 94°C for 30 s, 50°C for 30 s, and 72°C for 1 min 30 s, with a final extension at 72°C for 7 min in the first round; and 94°C for 2 min, 35 cycles consisting of 94°C for 30 s, 50°C for 30 s, and 72°C for 1 min, with a final extension at 72°C for 7 min in the second round. An HMA was performed as described elsewhere (9). A 460-bp *gag* p24-p7 region was amplified from the reference samples (NIH AIDS Research and Reference Reagent Program, NIH) by using the same sets of primers used to amplify p24-p7 from PBMC. Amplified *gag* DNA fragments from reference strains were mixed separately with the amplicon obtained from the sample

* Corresponding author. Mailing address: HIV/AIDS Laboratory, National Institute of Cholera and Enteric Diseases, P-33, C.I.T. Road, Scheme-XM, Beliaghata, Calcutta-700010, India. Phone: 91-33-2350 0448. Fax: 91-33-2350 5066. E-mail: drsekharchakrabarti@yahoo.co.in.

FIG. 1. Phylogenetic analysis of HIV-1 p24-p7 region of *gag* gene from 51 female sex workers in Calcutta, India. The construction of the tree is described in the text. Samples from Calcutta were designated "cal." Reference isolates of different subtypes used were as follows: subtype A, UG029 and Q23_17; subtype B, Mstd101, HXB2 copy, and NC7; subtype C, mIDU101_3, 96ZM651, 98IS002, KER2010, 93IN101, 93IN999, 95IN21068, 93IN904, 96BW01B03, 97ZA012, 98BR004, AA97202GP, and 97T204; subtype D, MB2059 and 99UGA07412; subtype F1, 93BR020_1 and BZ163; subtype G, 91ZR02 and X558; subtype H, 056 and VI991; subtype J, SE7887 and SE7022; and subtype K, EQTB11C and VI325.

FIG. 2. Phylogenetic study of p24-p7 gene sequences of 51 HIV-1 seropositive strains from Calcutta, the eastern region of India, with 44 other C strains from different regions of India. The symbols used with the isolate names denote the origins of the strains as defined in the figure. Samples from Calcutta are designated "cal." GenBank accession numbers of the Calcutta sequences are given in the text. The different isolates used for the study are as follows: NARI_GAG_1, NARI_GAG_2, NARI_GAG_3, NARI_GAG_4, NARI_GAG_5, NARI_GAG_6, 01IN565_10, 01IN565_11, 01IN565_13, 01IN565_14, 95IN21068, 93IN904, 94IN11246, 93IN101, 93IN905, 93IN9999, 98IN022, 94IN476, 98IN012, 50437, 60161, 60080, 60133, 60615, 50813, 50322, 49670, 50581, 33840, 33842, 49587, 80496, 50542, 70177, 50823, 50950, 50561, MYA1, IMP1, IMP3, IMP4, IMP5, IMP6, and C_GAG_221.

DNA (4.5 μ l each) in the presence of annealing buffer (100 mM NaCl, 2 mM EDTA, 10 mM Tris-HCl, pH 7.8). Denaturation and renaturation of DNAs were done by heating the mixes at 94°C for 2 min, followed by rapid cooling in ice. Heteroduplexes formed were then analyzed in a 5% polyacryl-

amide-20% urea gel in 1 $\times$ Tris-borate-EDTA buffer at 250 V for about 2 h 30 min. A 14- by 16-in vertical gel apparatus (Atto, Japan) was used for the HMA. Heteroduplex molecules that formed between the unknown sample and the most closely related subtype exhibited the fastest mobility. Analysis of het-

eroduplex mobility of 51 samples with respect to the reference strains showed that the prevailing subtype of HIV-1 in Calcutta is subtype C.

PCR amplicons were purified by a QIAGEN PCR purification kit (QIAGEN, Germany), and the purified products were subjected to cycle sequencing reactions in both directions by using fluorescent dye-labeled dideoxy nucleotides in an ABI PRISM 310 automated sequencer following the manufacturer's protocol. Sequences were submitted to GenBank, and the accession numbers were assigned (see below). The p24-p7 sequences were edited by using the BioEdit sequence alignment editor program (version 5.0.6.; Department of Microbiology, North Carolina State University [http://www.mbio.ncsu.edu/BioEdit/BioDoc.pdf]) and were subsequently analyzed with the BASIC BLAST program (http://www.hiv.lanl.gov/content/hiv-db/BASIC_BLAST/basic_blast.html), which revealed a close relatedness to subtype C. At least 5% divergence was shown by each sequence from those in the database, suggesting an absence of sample mix-ups with previously published sequences. All of the 51 sequences were then aligned with a reference panel of reported sequences and/or related sequences of strains isolated from different geographic regions available in the HIV sequence database (http://www.hiv.lanl.gov/content/index) provided by the Los Alamos National Laboratory, operated by the University of California, to generate the nucleotide substitution pattern among them. The reference panel included 30 sequences of different global strains with the same p24-p7 region of HIV-1 but consisting of all subtypes (A to K). At least two to three sequences of each reference subtype were taken for comparison. The multiple alignments were done by the ClustalW program. Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 2.1 (Kumar et al., 2001). Evolutionary distances were measured by a Kimura two-parameter distance matrix method. The mean genetic diversity among the Calcutta samples was found to be 7%, ranging from 2% to 13%. The mean genetic diversity between the strains from Calcutta and other subtype C global strains was also found to be 7%.

A phylogenetic tree was constructed by the neighbor-joining method using the interior branch test of phylogeny, a *t* test that is computed using the bootstrap procedure. Bootstrapping was done for 1,000 replicates, and finally the tree was viewed and edited by the Tree Explorer in MEGA 2.1. Figure 1 shows the phylogenetic tree generated for all 81 strains (Calcutta and others). Only confidence probability values above 50% were given. Sequences from Calcutta belonged to subtype C. However, no rigid monophyletic cluster was observed and all were placed in discrete groups. Some sequences from Calcutta were found to be dispersed among the Indian C strains and some among non-Indian C strains, such as strains from South Africa, Botswana, Brazil, Kenya, Israel, Zambia, and Myanmar. However, in most cases, the Calcutta sequences formed small groups among themselves, characteristic exclusively of the C type from Calcutta.

This result propelled us to study the distribution and phylogenetic pattern of the HIV-1 type C *gag* p24-p7 sequences from Calcutta compared to those of other Indian C strains from different regions. All of the 51 sequences from Calcutta were analyzed, with another 44 sequences from India (mainly western and northern parts) as available in the database (Fig. 2).

Samples from Calcutta represented eastern India. The ClustalW multiple sequence alignment program was used, and a phylogenetic tree was computed using the neighbor-joining method by MEGA version 2.1. The radial pattern of the tree was viewed for better observation. Two main clusters were formed. The majority (35 out of 44) of the Indian strains other than Calcutta strains formed cluster I, while another group of sequences (cluster II) were comprised mostly of the Calcutta strains (46 out of 51). Only five strains from Calcutta were within cluster I, while nine Indian strains other than Calcutta strains resided within cluster II with Calcutta strains. This result showed the distribution pattern of the HIV-1 strains from the eastern part of India to be a little different from that of strains from the rest of the country.

The overall study reveals HIV-1 subtype C to be the most prevalent type in the eastern part of India, based on the *gag* p24-p7 region. This also supports the earlier study done on the subtyping of HIV-1 based on the C2-V3 envelope region from this part of the country (14). A characteristic geographic separation between the Calcutta strains and the other Indian strains was also indicated. These results show that there might be a tendency toward variance persisting among subtype C HIV-1 strains in the eastern part of India. It will be interesting to study further to see whether a changing pattern is evolving in other genomic regions of the HIV-1 strains. Therefore, continuous monitoring and intensive studies for more detailed characterization of HIV-1 with other genes are needed, which might help in understanding the infection pattern and evolutionary lineage in the eastern part of India.

Nucleotide sequence accession numbers. The GenBank accession numbers of the Calcutta sequences are AY651099 through AY651128, DQ003318 through DQ003327, DQ023706 through DQ023712, and DQ054798 through DQ054801.

We thank the staff members of the STD/HIV Intervention Project, Calcutta, for their cooperation in collecting samples from the female sex workers in Calcutta. The *gag* HMA kit provided by the NIH AIDS Research and Reference Reagent Program is thankfully acknowledged.

S.S. was supported by CSIR, Government of India.

REFERENCES

1. Baskar, P. V., S. C. Ray, R. Rao, T. C. Quinn, J. E. Hildreth, and R. C. Bollinger. 1994. Presence in India of HIV type 1 similar to North American strains. *AIDS Res. Hum. Retrovir.* **10**:1039-1041.
2. Boyum, A. 1968. Separation of leukocytes from blood and bone marrow. *Scand. J. Clin. Lab. Investig.* **21**:77-81.
3. Buonaguro, L., M. Tagliamonte, M. L. Tornesello, and F. M. Buonaguro. 2005. Evaluation of a modified version of heteroduplex mobility assay for rapid screening of HIV-1 isolates in epidemics characterized by mono/dual clade predominance. *J. Virol. Methods* **124**:123-134.
4. Cham, F., L. Heyndrickx, W. Janssens, G. Van der Auwera, K. Vereecken, K. De Houwer, S. Coppens, H. Whittle, and G. van der Groen. 2000. Study of HIV type 1 *gag/env* variability in The Gambia, using a multiplex DNA polymerase chain reaction. *AIDS Res. Hum. Retrovir.* **16**:1915-1919.
5. Delwart, E. L., B. Herring, A. G. Rodrigo, and J. I. Mullins. 1995. Genetic subtyping of human immunodeficiency virus using a heteroduplex mobility assay. *PCR Methods Appl.* **4**:S202-S216.
6. Espantaleon, A., S. Kageyama, M. T. Bernardo, T. Nakano, P. S. Leano, P. Alban, R. Abrenica, S. Morimatsu, H. Teraoka, and D. M. Agdamag. 2003. The influence of the expanding HIV genetic diversity on molecular diagnosis in the Philippines. *Int. J. STD AIDS* **14**:125-131.
7. Esteves, A., R. Parreira, T. Venenno, M. Franco, J. Piedade, J. Germano De Sousa, and W. F. Canas-Ferreira. 2002. Molecular epidemiology of HIV type 1 infection in Portugal: high prevalence of non-B subtypes. *AIDS Res. Hum. Retrovir.* **18**:313-325.

8. Garber, D. A., G. Silvestri, and M. B. Feinberg. 2004. Prospects for an AIDS vaccine: three big questions, no easy answers. *Lancet Infect. Dis.* **4**:397–413.
9. Heyndrickx, L., W. Janssens, L. Zekeng, R. Musonda, S. Anagonou, G. Van der Auwera, S. Coppens, K. Vereecken, K. De Witte, R. Van Rempelbergh, M. Kahindo, L. Morison, F. E. McCutchan, J. K. Carr, J. Albert, M. Essex, J. Goudsmit, B. Asjo, M. Salminen, A. Buve, and G. van Der Groen. 2000. Simplified strategy for detection of recombinant human immunodeficiency virus type 1 group M isolates by *gag/env* heteroduplex mobility assay. *J. Virol.* **74**:363–370.
10. Jameel, S., M. Zafrullah, M. Ahmad, G. S. Kapoor, and S. Sehgal. 1995. A genetic analysis of HIV-1 from Punjab, India reveals the presence of multiple variants. *AIDS* **9**:685–690.
11. Kazennova, E. V., A. F. Bobkov, L. M. Selimova, T. A. Khanina, V. V. Pokrovskii, L. Heyndrickx, and G. Van der Groen. 2001. Analysis of gag gene subtypes of HIV-1 variants isolated in Russia by comparative assessment of heteroduplex electrophoretic mobility. *Vopr. Virusol.* **46**:12–16. (In Russian.)
12. Kurle, S., S. Tripathy, S. Jadhav, K. Agnihotri, and R. Paranjape. 2004. Full-length gag sequences of HIV type 1 subtype C recent seroconverters from Pune, India. *AIDS Res. Hum. Retrovir.* **20**:1113–1118.
13. Lole, K. S., R. C. Bollinger, R. S. Paranjape, D. Gadkari, S. S. Kulkarni, N. G. Novak, H. W. Sheppard, and S. C. Ray. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* **73**:152–160.
14. Mandal, D., S. Jana, S. K. Bhattacharya, and S. Chakrabarti. 2002. HIV type 1 subtypes circulating in eastern and northeastern regions of India. *AIDS Res. Hum. Retrovir.* **18**:1219–1227.
15. Mehendale, S. M., J. J. Rodrigues, R. S. Brookmeyer, R. R. Gangakhedkar, A. D. Divekar, M. R. Gokhale, A. R. Risbud, R. S. Paranjape, M. E. Shepherd, A. E. Rompalo, et al. 1995. Incidence and predictors of human immunodeficiency virus type 1 seroconversion in patients attending sexually transmitted disease clinics in India. *J. Infect. Dis.* **172**:1486–1491.
16. Ramalingam, S., R. Kannangai, T. S. Vijayakumar, D. Mathai, O. C. Abraham, S. Subramanian, P. Rupali, M. V. Jesudason, and G. Sridharan. 2005. Subtype & cytokine profiles of HIV infected individuals from south India. *Indian J. Med. Res.* **121**:226–234.
17. Sahni, A. K., V. V. Prasad, and P. Seth. 2002. Genomic diversity of human immunodeficiency virus type-1 in India. *Int. J. STD AIDS* **13**:115–118.
18. Shankarappa, R., R. Chatterjee, G. H. Learn, D. Neogi, M. Ding, P. Roy, A. Ghosh, L. Kingsley, L. Harrison, J. I. Mullins, and P. Gupta. 2001. Human immunodeficiency virus type 1 Env sequences from Calcutta in eastern India: identification of features that distinguish subtype C sequences in India from other subtype C sequences. *J. Virol.* **75**:10479–10487.
19. Stebbing, J., and G. Moyle. 2003. The clades of HIV: their origins and clinical significance. *AIDS Rev.* **5**:205–213.
20. Tripathy, S., B. Renjifo, W. K. Wang, M. F. McLane, R. Bollinger, J. Rodrigues, J. Osterman, S. Tripathy, and M. Essex. 1996. Envelope glycoprotein 120 sequences of primary HIV type 1 isolates from Pune and New Delhi, India. *AIDS Res. Hum. Retrovir.* **12**:1199–1202.
21. van der Groen, G., P. N. Nyambi, E. Beirnaert, D. Davis, K. Fransen, L. Heyndrickx, P. Ondoa, G. Van der Auwera, and W. Janssens. 1998. Genetic variation of HIV type 1: relevance of interclade variation to vaccine development. *AIDS Res. Hum. Retrovir.* **14**:S211–S221.
22. Zhong, P., S. Burda, F. Konings, M. Urbanski, L. Ma, L. Zekeng, L. Ewane, L. Agyingi, M. Agwara, Saa, Z. E. Afane, T. Kinge, S. Zolla-Pazner, and P. Nyambi. 2003. Genetic and biological properties of HIV type 1 isolates prevalent in villagers of the Cameroon equatorial rain forests and grass fields: further evidence of broad HIV type 1 genetic diversity. *AIDS Res. Hum. Retrovir.* **19**:1167–1178.