Characterization of Complete HIV Type 1 Genomes from Non-B Subtype Infections in U.S. Military Personnel

SODSAI TOVANABUTRA,1 STEPHANIE K. BRODINE,2,3 JOHN R. MASCOLA,4,5 JEAN-LOUIS SANKALE6 ERIC SANDERS-BUELL,1 BOHYE KIM,1 DEBORAH L. BIRX,4 and FRANCINE E. MCCUTCCHAN1

ABSTRACT
Infections with non-B HIV-1 subtypes are rare in the United States, but comprise a significant percentage of infections among U.S. military personnel. Risk behavior while on overseas deployment correlates with non-B infection in this population. Extensive genetic characterization will be required to define HIV-1 diversity, and to effectively evaluate requirements for HIV-1 vaccines and other prevention strategies in this group. From 1997 to 2000, 520 recent seroconverters, identified through routine HIV-1 testing in the U.S. active military force, volunteered for a prospective study. V3 loop serology or partial genome sequencing identified 28 non-B subtype infections; 14 were studied by full genome sequencing and phylogenetic analysis. Five strains were CRF01_AE. Four of these clustered with CM240 from Thailand, and one clustered with African CRF01_AE. Four strains were CRF02_AG, prevalent in West and West Central Africa. Two strains were subtype C. One strain was a unique recombinant between CRF01_AE and subtype B, and another was a complex unique recombinant between subtype A and D. The final strain was a member of a complex circulating recombinant first identified in Senegal, CRF09_cpx, incorporating subtypes A, F, G, and an unclassified genome. This diversity of non-B subtype HIV-1 strains, encompassing three globally prevalent non-B strains and including rare or even possibly unique strains, illustrates the breadth of U.S. military exposure while deployed and sets the bar higher for breadth of cross-subtype protection to be afforded by an HIV-1 vaccine.

The HIV-1 pandemic is genetically diverse. Nine subtypes (A-D, F-H, I, and K) and 21 circulating recombinant forms (CRF) have been identified to date. Detailed molecular epidemiological studies of HIV-1 have revealed that each subtype is endemic in different parts of the world. CRF01_AE, or former subtype E, is predominant in Southeast Asia where subtype B is also circulating, along with CRF15_01B and unique recombinant forms (URF) between CRF01_AE and subtype B. In South Asia, the epidemic is subtype C, and again, subtype B cocirculates. Subtypes B, C, CRF01_AE, and two CRF_BC are circulating in China. HIV-1 subtypes in Africa are highly diverse: CRF02_AG (BNEG) is epidemic in West and West Central Africa, subtypes A, C, and D are cocirculating in East Africa, and C is the most prevalent subtype in the South and Southeast of this continent. Subtypes A, B, and CRF03_AB have been identified in Eastern Europe and Central Asia. The epidemics in Australia, Western Europe, and North America are mainly subtype B, while subtype B, BF recombinants, and C are cocirculating in South America.3,5

In the United States, the AIDS epidemic was first recognized
in the late 1970s and the early 1980s among gay men in San Francisco and New York City.4,5 The identification of HIV-1 first occurred with these patients4,6 later classified as HIV-1 subtype B. This virus strain also spread in intravenous drug users, heterosexual, and other populations over time and became the most prevalent strain in North America.7 Although different U.S. populations may have different risk behaviors, the most common infecting strain is subtype B. Nonetheless, it has occasionally been reported that other HIV-1 subtypes are present, albeit in very low numbers, in the United States.8–12

U.S. military personnel are a unique population, with virtually worldwide potential for deployments and duty stations. Since 1985, periodic screening for HIV has allowed early identification of HIV infections in U.S. military personnel, with counseling and state of the art clinical management, and HIV-1-positive U.S. military personnel are not deployed.

The vaccine requirements for military personnel often differ from the general U.S. population, and HIV-1 vaccine may follow this paradigm. While a subtype B HIV-1 vaccine would be most genetically matched to most military and civilian U.S. HIV-1 infections, the non-B infections in the military constitute a significant percentage that could expand with increasing international engagement on the ground. To anticipate this need, the best available strategy is to study the breadth of strains acquired through current deployments and to study the active force prospectively. Here we report full characterization of 14 non-B strains, the first detailed analysis of the non-B risk to U.S. forces.

Between February 1997 and May 2000, 520 out of 701 recent seroconverters, identified through routine HIV-1 testing in U.S. military personnel, volunteered for a longitudinal evaluative study. The study has been described in detail elsewhere.13 A competitive-binding enzyme immunosassay (EIA) with V3 loop peptides was employed to screen for non-B subtype.4,13 Further confirmation was performed either by sequence analysis of envelope C2–V5 from peripheral blood mononuclear cell (PBMC) DNA or by a gene chip assay from extracted serum viral RNA. Of 520, 28 (5.4%) were non-B subtype: 2 (RF, MN, and RL42), subtype C (ETH2220, OBR025, 98TZ013, 90ZM651, 93SN905, and 94IN12146), subtype D (ZZZ6, ELI, NDK, and 94UG114), CRF01_AE (CM240, 93TH253, 97CNGX_11F, and 90CF602), CRF02_AG (BNG and DA263), subtype F (93V1850 and F93N363), subtype G (SE1665 and HB8793), subtype H (90CF756 and V1991), subtype J (SE1973J and SE92980.9), and CRF09_cps (95SN7705, 95SN7808, and 96GH2911). Phylogenetic trees were constructed and the consistency was evaluated using SEQBOOT, DNADIST (Kimura 2-parameter with a transition/transversion ratio 2.0), NEIGHBOR, CONSENSE, and DNAPARS modules of the Phylib Package (V1.52i)16 and TREESTOL.16 Recombinant strains were identified and breakpoints were mapped with maximum parsimony bootstrap using 300-nt windows overlapping by 50 nt.16 Breakpoints were confirmed by Recombinant Breakpoint Analysis software.20 The recombinant genome was broken into segments at the breakpoints and a phylogenetic analysis of each segment was performed with a neighbor-joining tree with the bootstrap. Phytophoot values of 0.70% in separate analyses confirmed the subtype of individual segments.

The HIV-1 strains were obtained from U.S. military personnel who seroconverted between 1996 and 2000 (Table 1). They were primarily men (n = 13), with a mean age of 35.8 ± 8.5 (SD) years (range 25–57) and of African-American (n = 7), white (n = 5), and Hispanic (n = 1) ethnicity. Only one female was in this group, a 23-years-old African-American. They gave a history of having heterosexual exposures, with 10 out of 12 (83%) reporting sex outside the United States. Their foreign exposures included several distinct HIV-1 epidemic areas, ranging from the Far East, Southeast Asia, the Middle East, Western Europe, and West Africa. Absolute CD4 lymphocyte counts of these subjects were 450 ± 62 (mean ± SE), with a range of 150–874 cells/µL.

A neighbor-joining tree is depicted in Fig. 1A. Strains MSC1100, 1120, 1164, 2008, and 3012 clustered with CRF01_AE reference sequences (bootstrap value 95%). There were two monophyletic groups among CRF01_AE.
<table>
<thead>
<tr>
<th>Case</th>
<th>Sample ID</th>
<th>Sample Test dates (neg/pos)</th>
<th>CD4 count (cells/μL)</th>
<th>Subtype</th>
<th>Foreign country exposed</th>
<th>Ethnicity (neg/pos)</th>
<th>Sex</th>
<th>Age</th>
<th>Sample</th>
<th>Country Exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 MSC1100</td>
<td>4/98</td>
<td>1/96–2/97</td>
<td>307</td>
<td>CRF01_AE</td>
<td>Thailand, Indonesia, Philippines, Singapore</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 MSC1120</td>
<td>9/98</td>
<td>10/95–7/98</td>
<td>704</td>
<td>CRF01_AE</td>
<td>Thailand, Japan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 MSC1164</td>
<td>3/00</td>
<td>5/97–2/00</td>
<td>256</td>
<td>CRF01_AE</td>
<td>Thailand, Japan, Philippines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 MSC3012</td>
<td>9/98</td>
<td>1/94–8/96</td>
<td>337</td>
<td>CRF01_AE</td>
<td>Philippines, Hong Kong</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 MSC3043</td>
<td>6/98</td>
<td>1/94–8/96</td>
<td>481</td>
<td>BE recombinant</td>
<td>Thailand, Singapore, Guam</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 MSC3134</td>
<td>4/99</td>
<td>5/98–4/99</td>
<td>150</td>
<td>CRF02_AG</td>
<td>Cameroon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 MSC3083</td>
<td>6/00</td>
<td>3/99–4/00</td>
<td>261</td>
<td>CRF02_AG</td>
<td>United Arab Emirates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22 MSC4041</td>
<td>12/98</td>
<td>7/92–4/97</td>
<td>647</td>
<td>CRF02_AG</td>
<td>Ghana, Nigeria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 MSC3007</td>
<td>9/98</td>
<td>8/95–3/96</td>
<td>874</td>
<td>CRF02_AG</td>
<td>Liberia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 MSC3057</td>
<td>5/99</td>
<td>1/98–1/99</td>
<td>269</td>
<td>CRF09-cpx</td>
<td>Germany</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26 MSC3018</td>
<td>2/98</td>
<td>9/96–6/97</td>
<td>817</td>
<td>C</td>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 MSC3016</td>
<td>3/98</td>
<td>10/92–7/97</td>
<td>468</td>
<td>C</td>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28 MSC4068</td>
<td>11/99</td>
<td>9/97–9/99</td>
<td>471</td>
<td>AD recombinant</td>
<td>Germany</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Reported in Brodine et al. (2003).11

Eight of these individuals also reported possible HIV-1 exposure in the United States.

This isolate is clustered most closely with African CRF01_AE.
formed a separate group with CF402, an African strain (100% bootstrap), while the others clustered (100% bootstrap) with Thai strains (CM240 and TH253) and with CNGX_11F from China. The intersubject nucleotide distance of the four Thai CRF01_AE strains was 4.3% over the full genome, and gag, pol, and envelope gene diversity were 4.1%, 3.0%, and 5.7%, respectively. Two subtype C strains, MSC3018 and 5016, clustered with the subtype C reference sequences (100% bootstrap). The new subtype C sequences, together with some reference sequences from Africa and India, formed a monophyletic group separate from the Brazilian and Ethiopian C strains. MSC1134, 3083, 4041, and 5007 were assigned to CRF02_AG with a bootstrap value of 80%. These CRF02_AG strains were more diverse than the CRF01_AE strains, with intersubject nucleotide distances of 8.4% for the whole genome, 6.9% in gag, 6.2% in pol, and 11.6% in envelope. The remaining three strains were recombinants.

The first recombinant, MSC4057, clustered with three CRF09_cpx strains from West Africa, which has a very complex structure containing subtypes A, F, G, and some unclassified regions. There were at least 12 shifts in subtype across the genome of MSC4057 and the other CRF09_cpx, located in gag, pol, env, vpr, and tat/rex coding regions, respectively. These four strains were recently classified as CRF09_cpx.47 The two other recombinants combined subtypes A and D, and CRF01_AE and B, respectively. Neither of them corresponded to any known CRF, hence they are classified as URFs. By bootscan analysis using two parental subtypes identified in Uganda (99UGA07072 as subtype A and 94UG1141 as subtype D) it was determined that MSC4068 was a unique recombinant form between subtypes A and D, with five crossover points (Fig. 1B, top panel). The first fragment of the genome belonged to subtype D (100% bootstrap value) starting from p17 in gag and encompassing p24, p2, p7, p1, p6, protease, RT, RNaseH, and part of integrase. The second fragment clustered with subtype A, with a bootstrap value of 69%. It encoded part of integrase, vif, and part of vpr. The genome again shifted to subtype D, forming a monophyletic group with strain 99UG1141 at a bootstrap value of 82%, in the region encoding part of vif, vpr, and tat1. The fourth fragment was too small to assign to either parent subtype, about 204 nt in length. A visual inspection of the alignment showed mostly subtype A genetic material. The next fragment was subtype D (bootstrap 80%), encompassing vpu and the beginning of envelope. The last fragment consisted of most of env, partial nef, and part of 3’ LTR, and was of subtype A (100% bootstrap). The last strain identified was MSC5043, a recombinant containing CRF01_AE and subtype B. It had three crossover points, two in pol and another in vif, breaking the genome into four fragments (Fig. 1B, bottom panel). The first fragment, from the beginning of gag up to near the end of RNaseH, was subtype B (bootstrap 100%). The genome shifted to CRF01_AE through approximately 80% of integrase. This second fragment clustered with CM240 and TH253 with the bootstrap value of 100%. Subtype B was as-

FIG. 1. (A) A neighbor-joining tree of 14 nearly full-genome sequences with the HIV-1 reference subtypes and circulating recombinant forms. A bootstrap value indicates a significant cluster of sequences with reference subtypes and circulating recombinant forms. The sequences reported here are in bold characters. The scale bar represents 10% genetic distance. (B) Bootscan analysis showing recombinant structure of unique HIV-1 recombinant forms, MSC4068 (A/D recombinant) and MSC5043 (CRF01/B recombinant) identified among U.S. military HIV-1 infections. A diagram at the top of each bootscanning is a genome structure corresponding to the landmarks of the HIV-1 genome at the bottom. Each subgenome region is presented as a box containing a bootstrap value of the node joining between a subgenome region of reference sequences and a sequence of interest. The subtype of origin for each region is shown.
signed in the next region (bootstrap 99%). This region contained the rest of integrase and more than half of vif. The last fragment was CRF01_AE (100% bootstrap) encoding 36 amino acids at the 3’ end of vif. vpr, tat/12, rev/12, vpu, gp160, partial nef, and partial 3’ LTR.

Demographic characteristics in this study included a large age range, broad ethnicity, and geographic risk exposure. Not surprisingly, only one was female, consistent with the 8–12% of females in U.S. military personnel. African-Americans were overrepresented, similar to data reported in the overall HIV-1 incidence in military personnel, 1985–1999.22

The complete genome characterization of these strains is almost a miniature of the global HIV-1 pandemic, including the globally prevalent subtype C, CRF01_AE, and CRF02_AG, alongside rare CRF09_cpx and even UR/F AD and 01B. The U.S. military personnel from whom HIV-1 sequences were derived had broad geographic exposures. Of note, 83% of the infections with international exposure were consistent with the predominant subtypes circulating in the area where the exposure occurred: four Thai CRF01_AE and one UR/F 01B in Southeast Asia and four CRF02_AG in West Africa. However, some strains, such as the African CRF_01AE and CRF09_cpx that are relatively rare and apparently endemic only in certain areas, were also identified. The study provides the first evidence suggesting the wider spread of African CRF01_AE outside Central Africa where it was found exclusively.23,24 CRF09_cpx has only been found previously in Senegal,23 but was in the U.S. military person with foreign exposure in Germany. UR/F AD should be identified in the area where subtypes A and D are circulating, like East Africa.25 The AD recombinant found in Germany, MSC 4068, reflects an East African origin, but was most likely obtained in Germany. This diversity of HIV-1 is similar to that recently reported in an immigrant population in New York City. The diversity of non-B subtypes in this civilian population was also broad, and the strain in each individual usually was a predominant subtype in the exposure area.26

The concordant findings between these two unique populations, immigrants and deployed U.S. military personnel, suggest a continuing influx of non-B strains into the United States, and, collectively, the incoming viruses can be as genetically diverse, if not more so, than in any single part of the world. These results reemphasize the challenges that must be faced to develop an effective HIV-1 vaccine for populations, such as the U.S. military with worldwide exposure. Of note, the most frequently identified non-B strains in this study were Thai CRF01_AE and West African CRF02_AG strains. The ongoing phase III vaccine efficacy trial in Thailand and phase IIb trials in Africa, respectively, will address these strains directly, and will be of clear benefit to the U.S. military vaccine development effort. The GenBank accession number of MSC 4068 is AY09360. The other sequences are available under GenBank accession numbers AJ444799 to AJ444801 and AJ444803 to AJ444812.

ACKNOWLEDGMENTS

The U.S. Military HIV Research Program, USMHRP, is jointly planned and funded by the Department of Defense, and NIAID/NIH interagency agreement. Additional support was from the Cooperative Agreement DAMD17-93-V5004, between the U.S. Army Medical Research and Material Command, Fort Detrick, Maryland, and the Henry M. Jackson Foundation for the Advancement of the Military Medicine, Rockville, Maryland. The views and opinions expressed herein do not necessarily reflect those of the U.S. Army or the Department of Defense.

REFERENCES


Address reprint requests to: Francine E. McCutchan Global Molecular Epidemiology Program Henry M. Jackson Foundation for the Advancement of Military Medicine 16000 East Gude Drive Rockville, Maryland 20850 E-mail: fmccutchan@hivresearch.org

**NON-B SUBTYPE HIV-1 IN U.S. MILITARY PERSONNEL**

429