AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 15, Number 18, 1999, pp. 1707–1712
Mary Ann Liebert, Inc.

Sequence Note

The Identification of a Complex A/G/I/J Recombinant HIV Type 1 Virus in Various West African Countries

CELINE MONTAVON,[1] FREDERIC BIBOLLET-RUCHE,[1,2] DAVID ROBERTSON,[3] BRAHIM KOUMARE,[4] CLAIRE MULANGA,[1] EKA ESU-WILLIAMS,[5] COUMBA TOURE,[6] SOULEYMANE MBOUP,[6] ERIC SAMAN,[7] ERIC DELAPORTE,[1] and MARTINE PEETERS[1]

ABSTRACT

In this sequence note we describe the full-length genome sequence of an HIV-1 isolate originating from the west African country of Mali. The phylogenetic tree analysis from the near full-length genome shows that the 95ML84 strain forms a separate cluster, supported by 100% of the bootstrap values, with the previously described A/G/J/? mosaic virus BFP90 from Burkina Faso. Additional analysis showed that throughout the genome the lowest diversity was seen between the 95ML84 and the BFP90 viruses, and bootscan analysis showed a similar complex genomic structure. In addition to the initial report describing the BFP90 virus as an A/G/J/? recombinant, our data show that for the BFP90 and 95ML84 strains the unclassified region corresponds to subtype I. The A/G/I/J BFP90 and 95ML84 strains represent the fifth and most complex circulating recombinant form of HIV-1 detected so far, and our data show its presence in various West African countries. Subtype I and J sequences, initially considered rare, seem to have broadened their geographical spread by way of these recombinant forms.

One of the major characteristics of human immunodeficiency virus type 1 (HIV-1) is its remarkable genetic variation. Phylogenetic analysis of many isolates from various parts in the world allows the subdivision of HIV-1 into three groups: M, N, and O.[1] Group M comprises the majority of HIV-1 strains responsible for the AIDS global epidemic and can be subdivided into at least 10 different *env* subtypes, A to J, approximately equidistantly related to each other.[2] The geographic distribution of HIV-1 group M subtypes is heterogeneous, but owing to the increas in international travel and intermixing of populations, viral strains are becoming more geographically dispersed and the simultaneous presence of multiple subtypes in a region has become more and more common.[3–5] Subtype designations have been powerful molecular epidemiological markers for tracking the course of the HIV global pandemic. In Africa all of the known subtypes have already long cocirculated.[3,6] In regions where multiple distinct subtypes of HIV-1 have cocirculated in the human population, it is likely that past coinfection of individuals with more than one subtype has led to the evolution of recombinant or mosaic forms of the virus. Since the first report on recombinant HIV-1 viruses, increasing numbers of recombinant HIV-1 genomes have been recognized and recombinant forms of epidemiologic importance have been identified. Indeed, up to 20% of the genomes that have been completely sequenced revealed a mosaic structure comprising fragments from two or more genetic subtypes.[7] All representa-

[1]Laboratoire Rétrovirus, Institut pour le Recherche en Developement (IRD), 34032 Montpellier, France.
[2]Present address: Department of Medicine and Microbiology, University of Alabama at Birmingham, Birmingham, Alabama.
[3]Laboratory of Structural and Genetic Information, CNRS-UMR 1889, 13402 Marseille, France.
[4]Institut National de Santé Publique (INSP), Bamako, Mali.
[5]University of Calabar, Calabar, Nigeria.
[6]Hôpital le Dantec, Dakar, Senegal.
[7]Innogenetics, 9052 Ghent, Belgium.

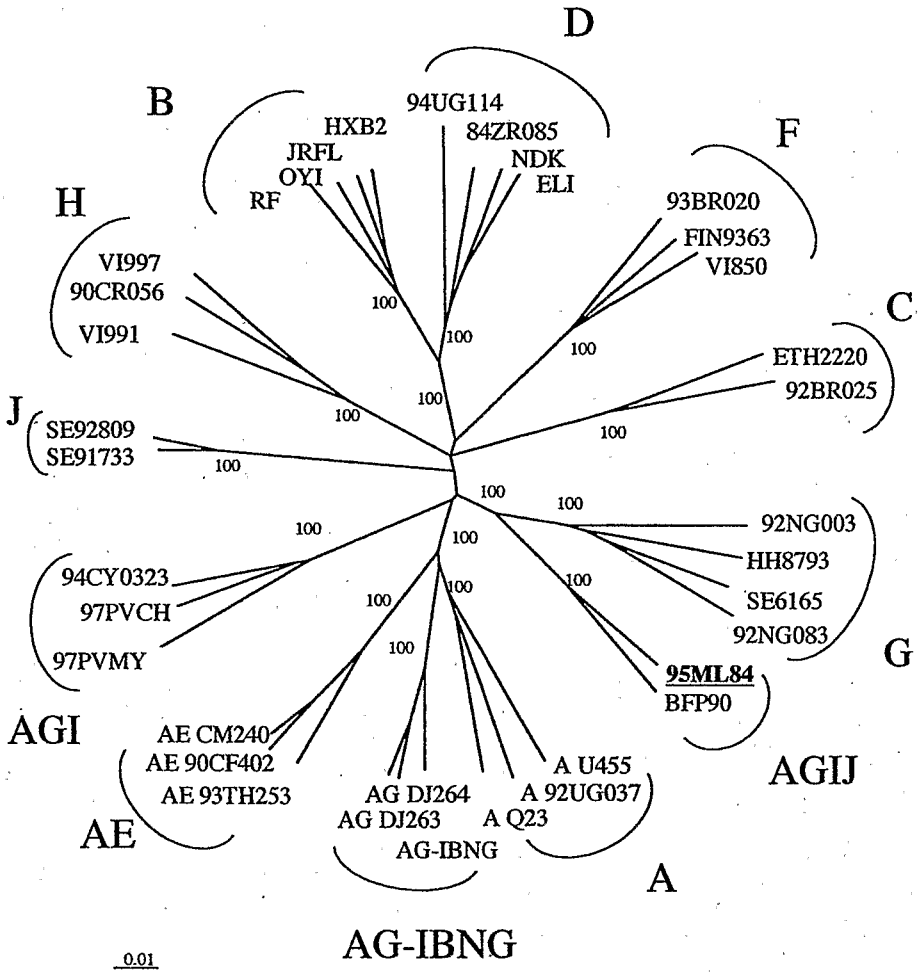
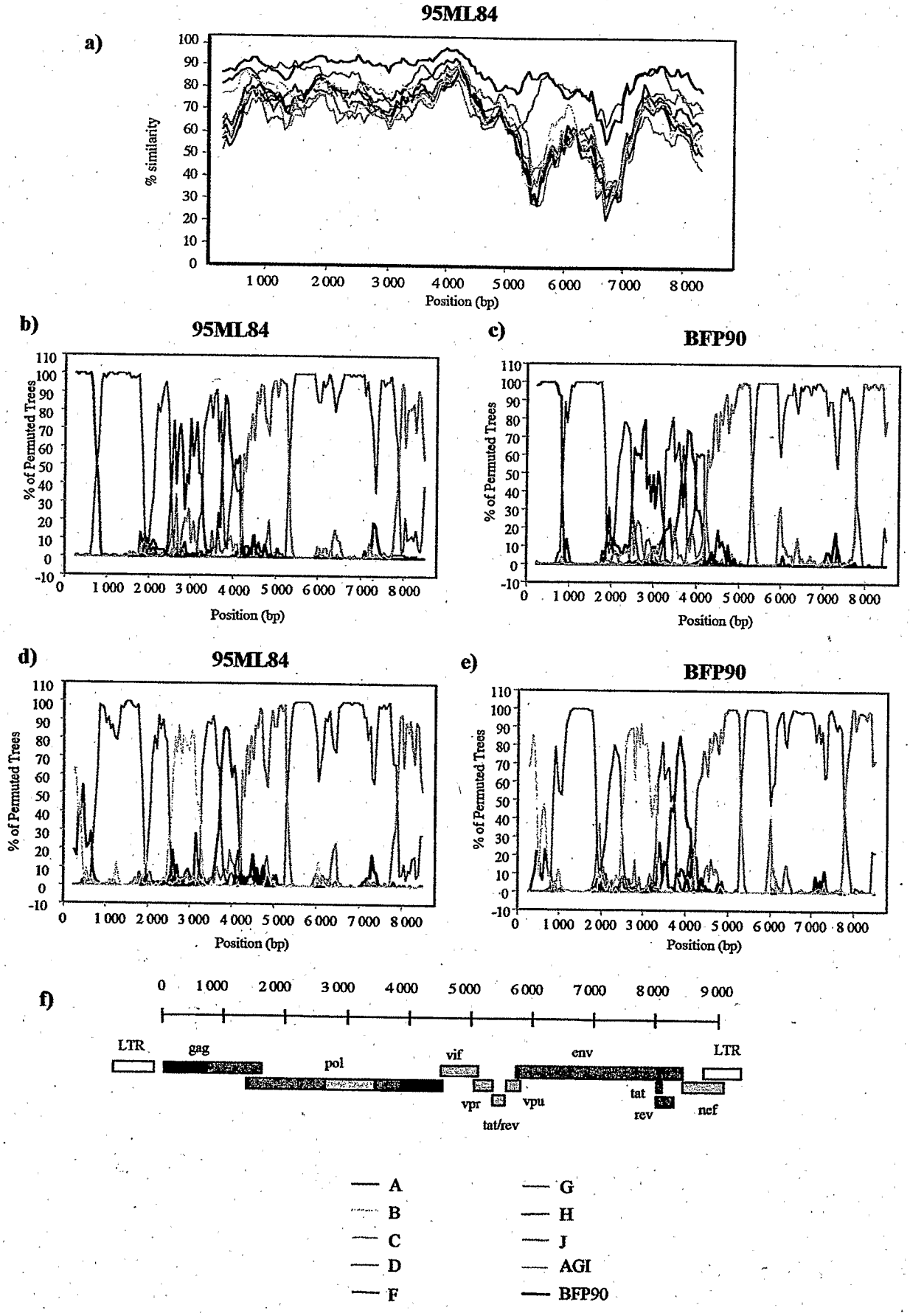


FIG. 1. The phylogenetic relationship of the newly derived virus was estimated from sequence comparisons with previously reported representatives of group M.[17] Nucleotide and amino acid sequences of the near full-length genome were aligned by CLUSTAL W,[18] with minor manual adjustments, bearing in mind the protein sequences. Sites where there was a gap in any of the sequences, as well as areas of uncertain alignement, were excluded from all sequence comparisons. Phylogenetic trees of the enitre genome and various regions of the genome were constructed by the neighbor-joining method and the reliability of the branching patterns was assessed by the bootstrap approach implemented by CLUSTAL W.

tives of subtypes E and G, sequenced to date, represent mosaic genomes, with parts of the genome clustering with subtype A viruses and other parts forming clearly distinguishable clades designated as E and G.[8–10] In certain populations and regions where multiple HIV-1 subtypes cocirculate many combinations of intersubtype recombinant viruses have been documented.[7,8,11,12] Recombination plays a significant role in global HIV evolution, creating novel viral genotypes with in populations. Recombination can be missed if large portions or various regions of the viral genome are not examined.

In this sequence note we describe the full-length genome sequence of an HIV-1 isolate originating from the west African

FIG. 2. Similarity plots, calculated with Simplot software, show that 95ML84 is the most similar to BFP90 (**a**). For bootstrap plots, Simplot software[15] performed bootscanning on neighbor-joining trees by using SEQBOOT, DNADIST (with the Kimura two-parameter method and a transition/transversion ratio of 2.0), NEIGHBOR, and CONSENSE from the Phylip package for a 300-pb window moving along the alignment in increments of 20 bp. We evaluated 100 replicates for each phylogeny. The bootstrap values for the sequences studied were plotted at the midpoint of each window. The same multiple genome alignment generated for the phylogenetic analysis of the nearly full-length genomes shown in Fig. 1 was used to calculate the consensus reference sequences (50% threshold) for the eight groups of nonrecombinant subtypes corresponding to subtypes A, B, C, D, F, G, H, and J. (**b**) and (**c**) show the bootstrap plots of the 95ML84 and BFP90 strains, respectively, against the nonrecombinant subtype references. (**d**) and (**e**) show the bootstrap plots of the 95ML84 and BFP90 strains, respectively, against the nonrecombinant subtypes and the AGI references from Cyprus and Greece. (**f**) shows the overall genomic structure of the 95ML84 and BFP90 viruses.


95ML84
a)
% similarity
Position (bp)
b)
95ML84
% of Permuted Trees
Position (bp)
c)
BFP90
% of Permuted Trees
Position (bp)
d)
95ML84
% of Permuted Trees
Position (bp)
e)
BFP90
% of Permuted Trees
Position (bp)
f)
LTR
gag
pol
vif
env
LTR
vpr
vpu
tat
rev
nef
tat/rev
A
B
C
D
F
G
H
J
AGI
BFP90

country of Mali, and partial sequences from samples obtained in Nigeria and Senegal. The sample from Mali was collected in 1995 during a study performed to identify the prevalence of various genetic subtypes of HIV-1 among female sex workers (FSWs) in Bamako, the capital city.[13] The sample was obtained from a 35-year-old FSW, working in Bamako but of Ghanian nationality. She was in good health and showed no clinical signs of AIDS-like disease. A 10-ml whole blood sample was collected in an EDTA tube and peripheral mononuclear blood cells were separated from the plasma by Ficoll gradient centrifugation. Plasma and cell pellets were stored at −20°C. The Innolia HIV-1/HIV-2 test (Innogenetics, Ghent, Belgium) showed the simultaneous presence of antibodies to HIV-1 and HIV-2 and the preliminary genetic characterization of this sample, previously described,[13] identified HIV-1 and HIV-2 sequences in this patient. On the basis of a partial sequence of the V3C4 region, the sample was classified as HIV-1 subtype G.[13] Sequence analysis of the accessory gene region revealed that in this region the subtype was not G but more A-like,[8] and the sequence did not cluster with any of the sequences known at that time. It was wrongly assumed that 95ML84 was subtype G, as subtype J had not yet been defined.[14]

To elucidate the genomic structure of this virus, the near full-length sequence was derived directly from primary peripheral blood mononuclear cells from the patient. Provirus was amplified by a series of overlapping nested polymerase chain reactions (PCRs) amplifying fragments between 2000 and 5000 bp in length. We used the Boehringer Mannheim (Indianapolis, IN) Long Template Expand DNA polymerase, according to the instructions of the manufacturer. The overlapping PCR fragments from each patient were either cloned and sequenced or the PCR product was directly sequenced. Sequencing was performed by cycle sequencing and dye terminator technology on an automated DNA sequencer (ABI 373A Stretch; Applied Biosystems, Foster City, CA). The sequenced fragments were reanalyzed and assembled into contiguous sequences by using the Seqed program (Applied Biosystems).

Examination of potential coding regions revealed the expected reading frames for *gag*, *pol*, *vif*, *vpr*, *tat*, *rev*, *vpu*, *env*, and *nef*. Long terminal repeat (LTR) extremities were not sequenced. None of the genes contained inframe stop codons, major deletions, insertions or rearrangements.

The phylogenetic tree analysis of the near full-length genome shows that the 95ML84 strain from Mali forms a separate cluster, supported by 100% of the bootstrap values, with the previously described A/G/J/? mosaic virus BFP90 from Burkina Faso[12] (Fig. 1). To find out whether the 95ML84 and BFP90 HIV-1 viruses have the same intersubtype recombinant structure additional analysis was done. Similarity plots, produced by Simplot 2.5 software,[15] determined the percent similarity between selected pairs of sequences by moving a window of 400 bp along the genome alignment in 10-bp increments. The genome alignment used was the same as that from the phylogenetic tree analysis, including nonmosaic reference sequences for subtypes A through J as well as selected intersubtype recombinants. From this analysis it is clear that over all of the genome the lowest diversity was seen between the 95ML84 and BFP90 viruses (Fig. 2a). To identify whether the 95ML84 virus has a similar structure, involving subtypes A, G, J, and some unknown fragments, supplementary similarity plots were done for the 95ML84 and the BFP90 viruses. Careful inspection of the graphs, including some recombinant strains, revealed several small areas of higher sequence similarity involving sequences from subtypes A, G, I, and J and small fragments with unknown relatedness. The graphs for the ML84 and BFP90 viruses were identical (data not shown). Moreover, the Simplot 2.5 software was used to calculate bootstrap plots. The results of the bootscan analysis are shown in Fig. 2b–e. The complexity of the new strain is readily apparent by bootscan analysis, using the nonrecombinant reference subtypes, and shows alternating segments clustering with subtypes A, G, and J, and with unclassified areas, similar to the complexity observed for the BFP90 prototype strain (Fig. 2b and c). However, when the recombinant AGI isolates from Cyprus[14] and Greece are included in the bootscan analysis, the unclassified part of the *pol* region could be classified as subtype I (Fig. 2d and e). The overall genomic structure of the 95ML84 and BFP90 viruses is shown in Fig. 2f. The *gag* gene is subtype A over the segment encoding p17 and part of p24, the remaining portion being subtype G. The *pol* gene is subtype G in the protease region and half of the reverse transcriptase (RT) region, whereas the 3′ end of the RT gene is classified as subtype I; the intergrase region is subtype G at the 5′ end and subtype A at the 3′ end. The accessory gene region, including *vif*, *vpr*, *tat*, and *vpu*, is subtype J, the envelope gene is entirely subtype G, and the *nef* gene is subtype J. All of these different subtype designations have been confirmed by phylogenetic tree analysis of the corresponding fragments. In addition to the initial report describing the BFP90 virus as an A/G/J/? recombinant, our data show that for the BFP90 strain the unclassified regions correspond to subtype I.

Phylogenetic tree analysis of smaller fragments dispersed throughout the genome showed that the 95ML84 and BFP90 strains always formed separate and well-supported clusters within the subtype A and G sequences, depending on the genomic region studied. Genetic characterization of HIV-1 samples from Senegal and Nigeria showed that some subtype A strains in *gag* formed a separate cluster with 95ML84 and BFP90; in addition, in the envelope gene some subtype G strains clustered with these viruses (Fig. 3a and b). To find out whether these strains have a similar complex genomic structure, we sequenced the *pol* gene, in which multiple breakpoints are documented, from the Nigerian sample MAOP11. Indeed, phylogenetic tree analysis of the *pol* gene showed that the MAOP11 virus clusters together with ML84 and BFP90 (Fig. 3a), and bootscan analysis confirmed for the MAOP11 strain the same complex G/I/G/A structure in the *pol* gene (Fig. 3b).

Recombinant viruses are already contributing substantially to the global pandemic; in this study we have identified a new circulating recombinant HIV-1 virus, and four such types are currently documented: A/E-CM240, A/G-IBNG, AGI-CY032, and A/B-Kal153.[16] The A/G/I/J recombinant is the fifth and most complex so far, and circulates in various west African countries: the initial virus, BFP90, was isolated from a patient in Burkina Faso; the 95ML84 strain was from a female commercial sex worker in Mali; and the MAOP11 strain from an HIV-infected person in the northern part of Nigeria (preliminary sequence data from other regions suggest the presence of this recombinant form also in Senegal). Subtype I and J sequences, initially considered rare, seem to have a broader geographical spread by way of these recombinant forms.

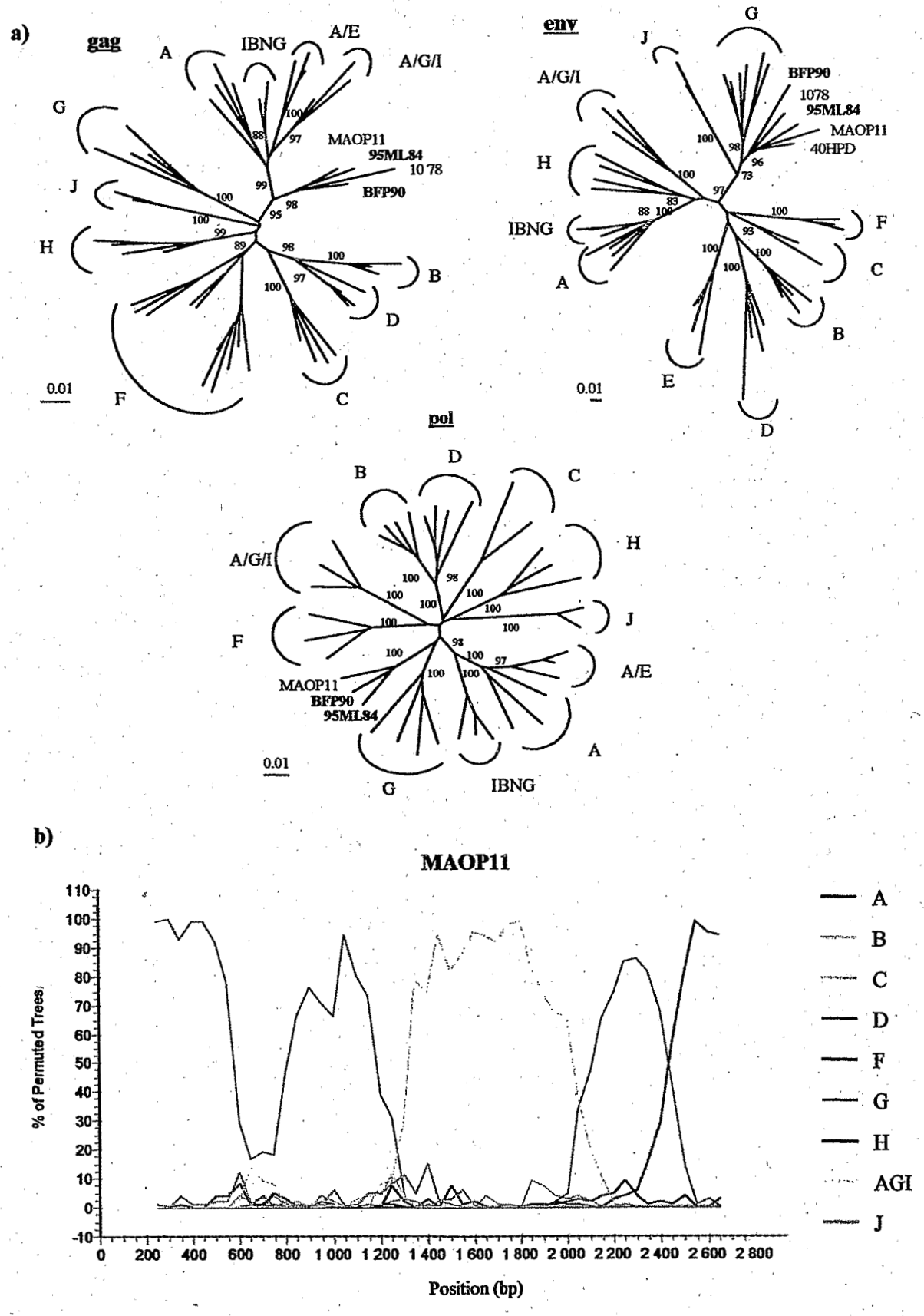


FIG. 3. Phylogenetic relationship of *gag* (p24), *env* (V3–V5 region), and *pol* sequences from strains from Senegal and Nigeria compared with representatives of group M and the A/G/I/J recombinant HIV-1 strains (BFP90 and 95ML84) (**a**). Bootscan analysis of the *pol* gene from the Nigerian MAOP11 strain (**b**).

ACKNOWLEDGMENTS

This work was cosponsored by grants from the Agence Nationale de Recherches sur le SIDA (ANRS, Projet Sidak) and by the Joint United Nations Programme on HIV/AIDS (UNAIDS) (Contract Ref. Moleculo07). Celine Montavon has a doctoral fellowship from the ANRS. The Simplot version 2.5 software was kindly provided by Stuart Ray.

SEQUENCE DATA

The new sequences have been deposited in the GenBank Data Library under the following accession numbers: 95ML84 complete genome, AJ245481; other partial sequences, AJ245482 to AJ245487.

REFERENCES

1. Simon F, Mauclere P, Roques, *et al.*: Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nature Med 1998;4:1032–1037.
2. Janssens W, Buve A, and Nkengasong JN: The puzzle of HIV-1 subtypes in Africa. AIDS 1997;11:705–712.
3. European Commission and the Joint United Nations Programme on HIV/AIDS: HIV-1 subtypes: Implications for epidemiology, pathogenicity, vaccines and diagnostics. AIDS 1997;11:UNAIDS17–UNAIDS36. [Workshop report]
4. Lasky M, Perret JL, Peeters M, *et al.*: Presence of multiple non-B subtypes and divergent subtype B strains of HIV-1 in individuals infected after overseas deployment. AIDS 1997;11:43–51.
5. McCutchan FE, Salminen MO, Carr JK, and Burke DS: HIV-1 genetic diversity. AIDS 1996;10(Suppl. 3):S13–S20.
6. Kanki PJ, Peeters M, and Gueye-Ndiaye A: Virology of HIV-1 and HIV-2: Implications for Africa. AIDS 1997;11(Suppl. B):S33–S42.
7. Robertson DL, Gao F, Hahn BH, and Sharp PM: Intersubtype recombinant HIV-1 sequences. In: *Human Retroviruses and AIDS.* Los Alamos National Laboratory, Los Alamos, New Mexico, 1997, Part III, pp. 25–30.
8. Gao F, Robertson DL, Carruthers CD, *et al.*: A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. J Virol 1998;72:5680–5698.
9. Gao F, Robertson DL, Morrison SG, *et al.*: The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. J Virol 1996;70:7013–7029.
10. Carr JK, Salminen MO, Koch C, *et al.*: Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. J Virol 1996;70:5935–5943.
11. Leitner T, Escanilla D, Marquina S, *et al.*: Biological molecular characterization of subtype D, G, and A/D recombinant HIV-1 transmissions in Sweden. Virology 1995;209:136–146.
12. Oelrichs RB, Workman C, Laukkanen T, McCutchan FE, and Deacon NJ: A novel subtype A/G/J recombinant full-length HIV type 1 genome from Burkina Faso. AIDS Res Hum Retroviruses 1998;14:1495–1500.
13. Peeters M, Koumare B, Mulanga C, *et al.*: Genetic subtypes of HIV-1 and HIV-2 strains in commercial sex workers from Bamako, the capital city of the west African country Mali. AIDS Res Hum Retroviruses 1998;14:51–58.
14. Gao F, Robertson D, Carruthers C, *et al.*: An isolate of human immunodeficiency virus type 1 originally classified as "subtype I" represents a complex mosaic genome comprised of 3 different group M subtypes (A/G/I). J Virol 1998;72:10234–10241.
15. Ray SC: *Simplot for Windows*, version 2.5. [Distributed by the author; http://www.med.jhu.edu/deptmed/sray/download/]
16. Carr J, Foley B, Leitner T, Korber B, and McCutchan F: Reference sequences representing the principal genetic diversity of HIV-1 in the pandemic. In: *Human Retroviruses and AIDS.* Los Alamos National Laboratory, Los Alamos, New Mexico, 1998, Part III, pp. 10–19.
17. Leitner T, Korber B, Robertson D, Gao F, and Hahn B: Update proposal of reference sequences of HIV-1 genetic subtype. In: *Human Retrovirus and AIDS.* Los Alamos National Laboratory, Los Alamos, New Mexico, 1997, Part III, pp. 19–24.
18. Thompson JD, Higgins DG, and Gibson TJ: CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680.

Address reprint requests to:
Martine Peeters
Laboratoire Rétrovirus, IRD
911 Avenue Agropolis
B.P. 5045
34032 Montpellier Cedex 1, France

E-mail: martine.peeters@mpl.ird.fr

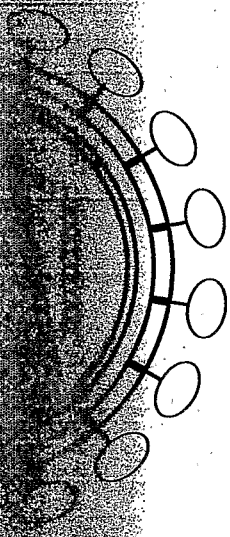
lume 15 Number 18 December 10, 1999

ISSN: 0889-2229

Conference Summary
ARGETING HIV RESERVOIRS AND
ECONSTITUTING THE IMMUNE SYSTEM

AIDS
CH
AN
RE

Edgar [illegible]GNESI, Ph.D.



Mary Ann Liebert, Inc.