Human immunodeficiency virus type 1 (HIV-1) shows great genetic diversity due to its high replication rate, the error-prone reverse transcriptase, and recombination events that may occur during virus replication (57). On the basis of genetic homology, HIV-1 has been classified in four groups: M (main), O (outlier), N (non-M, non-O), and the recently identified group P (36). HIV-1 group M is subdivided into 9 subtypes (A to D, F to H, J, and K), at least 49 circulating recombinant forms (CRFs) (http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html), and multiple unique recombinant forms (URFs). CRFs are defined as intersubtype recombinants for which at least three genetically unrelated variants are monophyletic, sharing an identical genetic structure along their full genomes. URF variants are widely distributed worldwide, with recombination breakpoints different from those found in CRFs. Genetic complexity is not always detected, mainly due to the subtyping of only one genetic region and not of the full genome. Consequently, specimens previously considered “pure” variants may be classified as recombinants when additional viral genes are analyzed. Therefore, the frequency of recombinant variants is underestimated in the pandemic. Recombination, in addition to purifying selection, is involved in the evolution of HIV (42), adaptation to its host, and escape from antiviral treatments (37). Recombination can also increase HIV fitness (50).

HIV-1 subtype B is the prevalent variant in developed areas, such as North America and Western Europe, including Spain (16, 23). However, subtypes other than subtype B and recombinants (HIV-1 non-B variants) are responsible for 90% of the 33 million infections worldwide (20, 46). These variants are increasing in prevalence and heterogeneity in developed countries (3, 15, 23, 45, 49, 54), mainly due to immigration and the movement of populations from areas of endemicity. The coexistence of multiple variants in the same region favors recombination between them after coinfection and/or superinfection events. In Spain, an increase in the frequency of HIV-1 non-B subtypes has been found among native Spaniards and immigrants newly diagnosed with HIV-1 in recent years (23), and the presence of different recombinants has been published (11, 14, 16, 21, 23, 24, 34, 52, 53).

The increasing prevalence of HIV-1 non-B variants could have implications for diagnosis (5), vaccine design (56), and the clinical management of HIV infection (39). HIV-1 non-B variants present clade-specific substitutions in positions related to drug resistance (26, 52). They could accelerate the emergence of drug-resistant viruses, change or induce alternative pathways of resistance (17, 19), influence viral replicative capacity in vitro (25), impair the interpretation of genotypic resistance algorithms (9, 43, 52), reduce the genetic barrier of certain protease inhibitors (47), and affect drug-binding affinity (27). Additionally, patients infected by certain HIV-1 non-B subtypes present accelerated disease pro-
gression (24, 48) and higher cognitive impairment (40). Thus, the proper detection and description of HIV-1 variants in representative cohorts is essential for further studies.

CoRIS, the cohort of the Spanish Research Network of Excellence on HIV/AIDS, has recently reported a prevalence of 15.2% for HIV-1 non-B subtypes (16). To further explore the molecular epidemiology of HIV in Spain, the objective of the present study was to characterize the HIV-1 recombinant variants detected in CoRIS by using phylogenetic analysis (phy), the gold-standard method for typing and discrimination between subtypes and/or CRFs. We also defined the complex mosaic patterns in variants classified as unique recombinants (not assigned to any known subtype or circulating recombinant form), as well as the phylogenetic clusters including such variants.

METHODS

Study population. CoRIS is an open, multicenter, prospective cohort of HIV-positive, antiretroviral (ART)-naive subjects more than 13 years old seen at 31 HIV units of the 18 Autonomous Regions in Spain from January 2004 on. Ethics approval was obtained from participating sites. The study was designed to protect the rights of all subjects involved under the appropriate local regulations. Written informed consent was obtained at the respective sites from every patient included in the study. A detailed description of the cohort has been published previously (7).

Of the 3,351 subjects included from 2004 to 2008, 670 patients provided a FASTA sequence while naïve to antiretroviral treatment (ART) and were included in this study. The alignment including these sequences is available as file S1 in the supplemental material. Overall, 375 patients (56%) were Spanish, 181 (27%) were immigrants (118 Central and South Americans, 17 sub-Saharan Africans, 16 Western Europeans, 11 Eastern Europeans, 10 North Africans, 6 North Americans, and 3 Asians), and 114 were of unknown origin. A pol sequence including the complete protease (codons 1 to 99) and part of the reverse transcriptase (codons 38 to 260 or 408) and/or CRFs. We also defined the complex mosaic patterns in variants classified as unique recombinants (not assigned to any known subtype or circulating recombinant form), as well as the phylogenetic clusters including such variants.

Phylogenetic analysis. HIV-1 subtypes and CRFs were identified by phylogenetic analysis (phy) of the 670 pol sequences. The 2008 version of the subtype reference data set provided by the Los Alamos National Laboratory (http://www.hiv.lanl.gov/content/sequence/NewAlign/align.html) was used. It was updated to include more sequences of CRFs that had been absent or scarcely represented (26 AU, 30_0206, 32_06A1, 34_01B, 38_BF, 41_CD, and 42_BF). Therefore, at least 2 representative sequences of each of the 9 subtypes and 43 CRFs of HIV-1 group M available in GenBank at the moment of the analysis were taken as references. DNA sequences were aligned using the ClustalX program, version 2.0.11. The tree topology was obtained using the neighbor-joining method. The pairwise distance matrix was estimated using the Kimura two-parameter model within the DNAdist program, as implemented in the PHYLIP software package. Bootstrap resampling (1,000 data sets) of the multiple alignments was performed to test the statistical robustness of the tree. We considered the associations of pol sequences showing a bootstrap value higher than 700 in the phylogenetic tree to be “clusters.”

URF characterization. Recombination analyses of sequences not assigned to any known subtype or CRF by phy were performed using several methods: SimPlot, version 3.5.1, the Recombination Detection Program (RDP; version 3alpha44), and the jumping-profile hidden Markov model (jHMM) (http://jphmm.gobics.de/jphmm.html). In the SimPlot analysis, which applies the bootscanning method, windows of 300 nucleotides moving in 10-nucleotide increments were used, as recommended by Zhang et al. (55). Sequences obtained from GenBank with the same geographical origin as the sequences analyzed were used as references when possible. After the different recombination analyses, the definitive breakpoints were those confirmed by constructing phylogenetic trees of the subsequences in order to assign them to the parental subtypes involved in the specific recombination event.

RESULTS

HIV-1 subtypes defined by phy in CoRIS: high frequency of recombinants. A total of 670 available pol sequences from different patients included in CoRIS were collected. As expected, 588 (87.8%) were assigned to subtype B and 82 (12.2%) to HIV-1 non-B variants. The majority (n = 59 [71.9%]) of HIV-1 non-B sequences were, in fact, viruses recombinant at pol (Fig. 1). Forty-eight viruses (58.5% of HIV-1 non-B variants) were 12 different CRFs (2 CRF01_AE, 31 CRF02_AG, 2 CRF03_AB, 1 CRF06_cpx, 1 CRF11_cpx, 1 CRF12_BF, 4 CRF14_BG, 1 CRF15_01B, 2 CRF19_cpx, 1 CRF20_BG, 1 CRF28_BG, and 1 CRF42_BF). Eleven viruses were unique recombinants (URFs), i.e., not assigned to any subtype or CRF. CRFs and URFs caused 7.2% and 1.6% of total infections, respectively. URF sequences did not cluster with any other known HIV-1 subtype or CRF after phylogenetic analyses of pol sequences, and they presented complex mosaic patterns due to recombination events between different subtypes (Fig. 1). In summary, among the 82 HIV-1 non-B pol sequences, 59 (71.9%) were shown by phy to be recombinants (81.4% CRFs and 18.6% URFs). Recombinant form CRF02_AG and URFs accounted for one-third and one-fifth, respectively, of HIV-1 non-B infections. The data showed that HIV-1 recombinants caused 8.8% of HIV-1 infections (this proportion rose to 14% when only the year 2008 was considered) and represented 71.9% of the HIV-1 non-B variants identified by phy.

HIV-1 non-B variants found in native Spanish individuals. Of the 375 native Spaniards included in the study, 31 (8.3%) carried HIV-1 non-B variants. Among these, only 9 harbored pure HIV-1 non-B subtypes; 22 were infected with recombinants (18 CRF and 4 URF). It is remarkable that 1 out of 2 CRF01_AE viruses, the CRF12_BF virus, 3 out of 4 CRF14_BG viruses, 1 out of 2 CRF19_cpx viruses, the CRF20_BG virus, and the CRF42_BF virus were found in Spanish individuals. Although the frequency of these variants among native Spaniards was low, native patients accounted for more than one-third (37.8%) of the total number of patients infected with HIV-1 non-B variants.

Origins of CoRIS patients infected by HIV-1 recombinants. The origins of subjects infected by each HIV-1 recombinant variant are shown in Fig. 2. According to the origins of patients, the prevalences of HIV-1 recombinants were as follows, in descending order: 64.7% (11/17) for sub-Saharan Africans, 60% (6/10) for North Africans, 33.3% (1/3) for Asians, 18.2% (2/11) for Eastern Europeans, 5.9% (7/118) for South and Central Americans, 5.9% (22/375) for Spaniards, and 0% for North Americans (0/6) and Western Europeans other than Spaniards (0/16). Among patients of unknown origin, this prevalence was 8.8% (10/114). Surprisingly, a CRF02_AG variant was isolated from an Asian patient, and CRF03_AB recombinants, first described in Russia, were isolated from South American patients.

Complex variants in CoRIS and epidemiological data. The 11 URFs presented mosaic patterns due to recombination events in the pol region between different subtypes and/or CRFs (Fig. 1). Recombinants presented fragments from 7 different pure subtypes (A, B, C, F, G, J, and K) and CRF02_AG. Some carried B/CRF02_AG (4/11), B/F1 (2/11), and B/A1 (2/11) sequences, and...
one showed a complex pattern including regions of subtypes A1, C, J, and K. Interestingly, 10 of the 11 URFs contained subtype B regions.

Table 1 records the epidemiological features of 11 HIV-1 URF-infected CoRIS patients defined by phy (Fig. 2) and confirmed by bootscanning and other methods (Fig. 1). Seven of these patients came from sub-Saharan Africa (n = 3), Eastern Europe or Russia (n = 2), South or Central America (n = 1), or North Africa (n = 1). Of note, the remaining four were native Spaniards (36.4% of cases), demonstrating the increasing heterogeneity of HIV-1 even in native Western Europeans. The modes of transmission for the URF-infected patients were as follows: 6 and 2 acquired HIV in-
fection through heterosexual or homosexual routes, respectively; 2 were injecting drug users (IDU); and for 1 patient, viral transmission took place through a contaminated-blood transfusion. In other words, URF variants in the CoRIS cohort were transmitted mostly through heterosexual contacts.

Clusters of CoRIS patients carrying complex recombinants. One cluster was found in the phylogenetic analysis of the 11 sequences assigned to HIV-1 URF variants (Fig. 1). This cluster consisted of three samples harboring sequences from HIV-1 subtype B, mainly at the protease, and CRF02_AG sequences at the reverse transcriptase (Fig. 1, URFs 3 to 5), with very similar breakpoints. The three patients were diagnosed in Madrid, Spain, and were infected by heterosexual transmission. Two of them were from sub-Saharan Africa (Table 1), which could explain the presence of CRF02_AG sequences, given the predominance of this strain in their country of origin (Equatorial Guinea). The third was a North African. The remaining eight URFs did not cluster with any other sample in the study population.

**DISCUSSION**

One of 10 HIV-1-infections in Spain could represent recombinant variants. This study was performed within a large and representative Spanish cohort of ART-naïve HIV-infected patients included in the Research Network on HIV/AIDS (CoRIS). CoRIS collects data on patients from different areas of Spain, a country with one of the highest HIV prevalences in the European Union (13, 46). The results presented confirmed that subtype B is still the main HIV-1 variant in Spain, as reported previously for the same sample of subjects from CoRIS (16), and are concordant with data from other Spanish studies (23, 31, 34). What our study adds is the finding that 8.8% of the 670 CoRIS patients (14% of the 164 patients assessed in 2008) were infected with HIV-1 recombinant variants (CRFs and URFs) and the description of the genetic nature of the complex recombinants. The recombinant variants represented almost three-quarters (72%) of the 82 HIV-1 non-B variants found. The highest prevalence of recombinants was found in patients from sub-Saharan Africa, followed by Eastern Europeans. As previously reported in the initial CoRIS description (16) and in other studies (23, 24, 31, 34), recombinant CRF02_AG was the most frequently found HIV-1 non-B variant in our country, accounting for more than one-third of HIV-1 non-B infections. This is due to the high prevalence of this recombinant in Central and West Africa, the most common origin of infected sub-Saharan Africans residing in Spain. Of note, we detected CRF02_AG in an Asian patient, as well as CRF03_AB in two South American patients. To our knowledge, this was the first time that these two variants were reported for patients from those regions in Spain, suggesting that transmission occurred in Spain.

**Introduction of complex HIV-1 variants into the native population.** In recent years, increasing prevalences of HIV-1 non-B subtypes and recombinants have been reported across Western Europe (23, 51, 54). The rising prevalence of HIV-1 non-B variants has traditionally been attributed to the growing number of immigrants from developing regions where these variants are prevalent. In fact, one-third of patients newly diagnosed with HIV in Spain in 2008 were immigrants (8, 38). However, the current heterogeneity of the HIV epidemic in Spain caused by recombinants can be only partially explained by immigration. Among the 23 and 59 subjects infected by HIV-1 non-B pure subtypes or recombinants, 9 (39.1%) and 22 (37.3%), respectively, were native patients. Among the 22 native subjects infected by recombinants, 18 and 4 patients carried CRF and URF variants, respectively. The detection in native Spaniards of 37.3% of the recombinant variants found in this study is evidence of the intro-

---

**TABLE 1** Epidemiological features of the 23 HIV-1-infected patients included in CoRIS carrying URF sequences at pol

<table>
<thead>
<tr>
<th>URF no.</th>
<th>Patient ID</th>
<th>Sex</th>
<th>Origin</th>
<th>Parental strains</th>
<th>Exposure category</th>
<th>City (region) of sampling</th>
<th>Yr of sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21</td>
<td>F</td>
<td>Eastern Europe</td>
<td>B/A1</td>
<td>IDU</td>
<td>Granada (East Andalusia)</td>
<td>2005</td>
</tr>
<tr>
<td>2</td>
<td>41</td>
<td>M</td>
<td>Eastern Europe</td>
<td>A1/B</td>
<td>IDU</td>
<td>Elche (Valencia)</td>
<td>2008</td>
</tr>
<tr>
<td>3</td>
<td>46</td>
<td>M</td>
<td>Equatorial Guinea</td>
<td>B/CRF02</td>
<td>Heterosex.</td>
<td>Madrid</td>
<td>2007</td>
</tr>
<tr>
<td>4</td>
<td>52</td>
<td>M</td>
<td>Equatorial Guinea</td>
<td>B/CRF02</td>
<td>Heterosex.</td>
<td>Madrid</td>
<td>2005</td>
</tr>
<tr>
<td>5</td>
<td>29</td>
<td>F</td>
<td>North Africa</td>
<td>B/CRF02</td>
<td>Heterosex.</td>
<td>Madrid</td>
<td>2005</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>M</td>
<td>Spain</td>
<td>B/CRF02</td>
<td>Homosex.</td>
<td>Malaga (South Andalusia)</td>
<td>2008</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>U</td>
<td>Spain</td>
<td>B/F1</td>
<td>Heterosex.</td>
<td>Seville (West Andalusia)</td>
<td>2008</td>
</tr>
<tr>
<td>8</td>
<td>76</td>
<td>M</td>
<td>Spain</td>
<td>B/F1</td>
<td>Homosex.</td>
<td>Madrid</td>
<td>2006</td>
</tr>
<tr>
<td>9</td>
<td>39</td>
<td>M</td>
<td>South America</td>
<td>CRF02/B</td>
<td>Other</td>
<td>Madrid</td>
<td>2006</td>
</tr>
<tr>
<td>10</td>
<td>34</td>
<td>M</td>
<td>Spain</td>
<td>B/C</td>
<td>Heterosex.</td>
<td>Madrid</td>
<td>2008</td>
</tr>
<tr>
<td>11</td>
<td>66</td>
<td>F</td>
<td>Sub-Saharan Africa</td>
<td>J/C/A1/K</td>
<td>Heterosex.</td>
<td>Terrassa (Catalonia)</td>
<td>2008</td>
</tr>
</tbody>
</table>

*Numbered as in Fig. 1.

Patient number according to the sequence alignment provided as File S1 in the supplemental material.

M, male; F, female; U, unknown.

Heterosex., heterosexual risk behavior; Homosex., homo/bisexual risk behavior; IDU, injecting drug user.
duction of complex HIV-1 variants into the native population of Spain, which also happens in other developed countries. Thus, the attribution of the increasing HIV-1 heterosexual transmission to immigrant populations is a prejudgment that only partially explains this phenomenon.

**Origin and transmission of URFs in Spain.** Interestingly, 13% of infections with HIV-1 non-B variants were caused by URFs. These are very frequent in regions where multiple clades cocirculate, such as sub-Saharan Africa (10), and are increasingly present in developed countries (15). However, whether the complex recombinant variants were transmitted directly from the immigrant community to the native population or whether the recombinant events took place in native Spaniards cannot be ascertained with the current data.

URF sequences from heterosexualy infected sub-Saharan Africans that form a clade (URFs 3 to 5) resulted from very similar recombinant events involving CRF02_AG and subtype B, a phenomenon previously reported in both Spain and France (24, 29). Two of these patients came from Equatorial Guinea, where CRF02_AG is highly prevalent (12), but the third patient was from North Africa, where subtype B is prevalent (1). The different origins of the patients suggest that the infection and subsequent spread occurred in Spain. Of note, URFs 6 and 9 also were B/CRF02_AG recombinants, although these recombinants did not share a common origin with those included in the cluster. In addition, other recombinants, including sequences typically found in Eastern European (URFs 1 and 2, including subsubtype A1) and South American (URFs 7 and 8, including subtypes B and F) countries, were found, reflecting the wide variety of geographical origins of immigrants in Spain.

Despite the pandemic spread of HIV-1 recombinants, their times of origin are not well understood. A recent paper suggests that recombination was common in the early evolutionary history of HIV-1 (44). In this cohort, the oldest URF was sampled in 2005, but complex recombinants have been reported in our country at least since the end of the 1990s (22–24, 30, 33), including the description of the first CRF that originated in Western Europe (CRF14_BG) (11). More epidemiological data about the infection date, risk behavior, and the geographical regions in which these patients have resided or traveled, as well as additional sequencing of longer genetic regions and specific computer programs to study the viruses’ genetic evolution, would be necessary to confirm these findings and to define the origin of complex recombinants circulating in Spain.

**Possible underestimation of the frequency of HIV-1 URFs in molecular epidemiology studies.** This work reveals the circulation and spread of complex HIV-1 variants in a large and representative cohort of HIV-infected persons in Spain. However, the detection of URFs could be even higher if more viral regions were analyzed. For instance, more than one-third of HIV-1 sequences described in the Los Alamos HIV database to date might be found to be recombinant forms if different genes or full-length sequences were analyzed (41), and therefore, complex recombinants could be more frequent than expected in the HIV-1 pandemic (35). Another limitation of our study lies in the representativeness of the data, given the unequal distribution of the patients across the 18 centers and the territory of Spain. Thus, due to the scarcity of data from certain regions and/or hospitals, our findings could not be representative of the HIV-1 epidemic in our country. Nevertheless, this is the largest HIV-1 molecular epidemiology study performed in Spain using phylogeny to determine the distribution of HIV-1 variants.

**Biological consequences of recombination in HIV-1 evolution.** Not only is the frequency of recombinants likely underestimated, but an increasing presence of URFs in developed countries is expected in the coming years due to the movement of populations between countries where different HIV-1 variants are prevalent. In fact, decreasing numbers of pure subtype B viruses and increases in the numbers of unique recombinants including subtype B sequences among HIV-1-seropositive patients have also been reported recently in neighboring countries (15). It has also been suggested that the HIV-1 epidemic could be evolving toward a more complex epidemiological landscape (18). Recombination seems to be very important in the evolution of HIV-1 (42), since it can provide a biological advantage versus parental viruses (28), promoting biological adaptation and enhancing fitness (6). It can also facilitate drug resistance and may allow superinfecting HIV-1 strains to evade preexisting immune responses (32). Thus, the continuous spread of HIV-1 recombinants may have serious implications for efforts to control the AIDS pandemic (including future vaccination trials) and could represent one of the highest barriers to HIV-1 eradication (32). However, despite some cases where URFs are described as highly pathogenic (4), the clinical implications of the presence of URFs for the AIDS pandemic remain to be clarified.

**Acknowledgments**

This study would not have been possible without the collaboration of all the patients, medical and nursery staff, and data managers who have taken part in the project.

The RIS Cohort (CoRIS) is funded by the Instituto de Salud Carlos III through the Red Temática de Investigación Cooperativa en Sida (RIS C03/173). This work was supported in part by grants from the Fondo de Investigaciones Sanitarias (FIS 09/00284). A.H. is supported by the Agenzia Lain Entralcolo. G.Y. is supported by the Consejería de Educación de la Comunidad de Madrid and the Fondo Social Europeo (FSE). Centers and investigators involved in CoRIS. Steering committee: Juan Berenguer, Julia del Amo, Federico García, Félix Gutiérrez, Pablo Labarga, Santiago Moreno, and María Ángeles Muñoz. Data management and statistical analyses: Ana María Caro-Murillo, Paz Sobrino Vegas, Santiago Pérez-Cachafeiro, Victoria Hernando Sebastián, Belén Alejos Ferreras, and Débora Álvarez. BioBank: M Ángeles Muñoz-Fernández, Isabel García-Merino, Coral Gómez Rico, Jorge Gallego de la Fuente, and Almudena García Torre. Centers: Hospital General Universitario de Alicante, Alicante (Joanit Portilla Sogorb, Esperanza Merino de Lucas, Sergio Reus Baus, Vicente Boix Martínez, Lívia Giner Oncina, Carmen Gadea Pastor, Irene Portilla Tamarit, and Patricia Arcaina Toledo), Hospital Universitario de Canarias, Santa Cruz de Tenerife (Juan Luis Gómez Sivrent, Patricia Rodríguez Fortuño, María Remedios Alemán Valls, María del Mar Alonso Sos, Ana María López Liro, María Inmaculada Hernández Hernández, and Felicitas Díaz-Flores), Hospital Carlos III, Madrid (Vicente Soriano, Pablo Labarga, Pablo Barreiro, Carol Castañera, Pablo Rivas, Andrés Ruiz, Francisco Blanco, Pilar Garcia, and Mercedes de Diego), Hospital Universitario Central de Asturias, Oviedo (Vicente Asensi, Eulalia Valle, and José Antonio Martín), Hospital Clinic, Barcelona (José M. Miró, María López-Díez, Christian Manzano, Laura Zamora, Iñaki Pérez, M. Teresa García, Carmen Ligero, José Luis Blanco, Felipe García-Alcaide, Esteban Martín, Josep Mallolas, and José M. Gatell), Hospital Doce de Octubre, Madrid (Rafael Rubio, Federico Pulido, Silvana Fiorante, Jara Llenas, Violeta Rodríguez, and Mariano Matarranz), Hospital Donostia, San Sebastián (José Antonio Iribarren, Julio Arrizabalaga, María José Aramburu, Xabier Camino, Francisco Rodríguez-Arrondo, Miguel Ángel von Wichmann, Lidia Pascual Tomé, and Débora Álvarez).
References
Most HIV-1 Non-B Variants in Spain Are Recombinants

Circulating recombinant form 02_AG (CRF02_AG) has a higher in vitro replicative capacity than its parental subtypes A and G. J. Med. Virol. 78:523–534.


