Impact of Clinical Parameters in the Intrahost Evolution of HIV-1 Subtype B in Pediatric Patients: A Machine Learning Approach

Patricia Rojas Sánchez[1,5], Alberto Cobos[2], Marisa Navaro[3], José Tomas Ramos[4], Israel Pagán[2], and Africa Holguín[1,*]

[1]HIV-1 Molecular Epidemiology Laboratory, Department of Microbiology, Hospital Ramón y Cajal-IRYCIS and CIBER-ESP (for the Madrid Cohort of HIV-1 Infected Children and Adolescents Integrated in the Pediatric Branch of the Spanish National AIDS Network (CoRISPe), Madrid, Spain

[2]Department of Plant-Microbe Interaction, Centro de Biotecnología y Genómica de Plantas (UPM-INIA) and E.T.S.I. Agrónomos, Universidad Politécnica de Madrid, Spain

[3]Department of Infectious Diseases, Hospital General Universitario Gregorio Marañón-CORISPe, Madrid, Spain

[4]Department of Infectious Diseases, Hospital Clínico Universitario and Universidad Complutense-CORISPe, Madrid, Spain

[5]Present address: Transcription-associated genome instability Laboratory, Institute of Cancer and Genomic Sciences, School of Medicine, University of Birmingham, Birmingham, United Kingdom

*Corresponding author: E-mail: africa.holguin@salud.madrid.org

Accepted: September 18, 2017

Abstract

Determining the factors modulating the genetic diversity of HIV-1 populations is essential to understand viral evolution. This study analyzes the relative importance of clinical factors in the intrahost HIV-1 subtype B (HIV-1B) evolution and in the fixation of drug resistance mutations (DRM) during longitudinal pediatric HIV-1 infection. We recovered 162 partial HIV-1B *pol* sequences (from 3 to 24 per patient) from 24 perinatally infected patients from the Madrid Cohort of HIV-1 infected children and adolescents in a time interval ranging from 2.2 to 20.3 years. We applied machine learning classification methods to analyze the relative importance of 28 clinical/epidemiological/virological factors in the HIV-1B evolution to predict HIV-1B genetic diversity (d), nonsynonymous and synonymous mutations (d_N, d_S) and DRM presence. Most of the 24 HIV-1B infected pediatric patients were Spanish (91.7%), diagnosed before 2000 (83.3%), and all were antiretroviral therapy experienced. They had from 0.3 to 18.8 years of HIV-1 exposure at sampling time. Most sequences presented DRM. The best-predictor variables for HIV-1B evolutionary parameters were the age of HIV-1 diagnosis for d, the age at first antiretroviral treatment for d_N and the year of HIV-1 diagnosis for d_s. The year of infection (birth year) and year of sampling seemed to be relevant for fixation of both DRM at large and, considering drug families, to protease inhibitors (PI). This study identifies, for the first time using machine learning, the factors affecting more HIV-1B *pol* evolution and those affecting DRM fixation in HIV-1B infected pediatric patients.

Key words: HIV-1, intrahost evolution, pediatric patients, machine learning.

Importance

This is the first study that analyzes the interactive effects of 28 clinical and virological features in the intrahost evolution of HIV-1 in pediatric patients during a long HIV-1 exposure time. During the HIV infection, clinical, epidemiological, and virological parameters fluctuate over time, presenting differences across patients. These parameters seemed to be relevant in the course of the infection and HIV-1 evolution. Understanding whether variation in these clinical parameters also affect within-host HIV-1 evolution may also help to design more efficient control strategies. However, more related studies are required for a better understanding of HIV evolution in children and adolescents.

Introduction

RNA viruses present high potential for generating large population genetic diversities at the intra and interhost levels (Lemey et al. 2006; Holmes 2009; Gray et al. 2011; Sede et al. 2014). This provides a high capacity for viral adaptation to new environments, which may represent an enormous evolutionary advantage with far-reaching consequences for key viral traits such as disease progression, infectivity, transmissibility, and response to antiviral treatments (Holmes 2009; Santoro and Perno 2013). Thus, understanding the factors that determine the level of intra and interhost genetic diversity in RNA virus populations is required to understand viral disease dynamics (Holmes 2009).

Human immunodeficiency virus type 1 (HIV-1) is an RNA virus presenting high genetic diversity mainly due to the high mutation rates derived from the error-prone nature of its reverse transcriptase (Abecasis et al. 2009; Maldarelli et al. 2013), to its high recombination rate (Zhuang et al. 2002; Moradigaravand et al. 2014), and to the high replication rates and large population sizes (Perelson et al. 1996). Levels of HIV-1 genetic diversity differ between HIV-1 subtypes and recombinants (Abecasis et al. 2009), and may also be affected by clinical factors. A higher within-host HIV-1 genetic diversity in naïve patients has been associated with lower levels of T lymphocyte CD4 count (Markham et al. 1998; Lemey et al. 2007), higher viral load (Mani et al. 2002; Shriner et al. 2006), larger virus exposure time (Maldarelli et al. 2013; Ryland et al. 2010), and age (Carvajal-Rodriguez et al. 2008). In treated patients, antiretroviral therapy (ART) has been demonstrated to have complex effects on HIV genetic diversity, promoting adaptive evolution and fixation of mutations conferring drug resistance (Lorenzo et al. 2004; Pennings 2012) or reducing virus genetic diversity (Gall et al. 2013; Kearney et al. 2014). However, during the course of an infection or an epidemic, HIV-1 populations face changes in viral load (VL), CD4 count, and ART experience that occur simultaneously. Thus, further analysis of the role of these clinical factors in determining the genetic diversity of HIV-1 populations and of their interactive effects is required (Lemey et al. 2006). However, such studies are scant, mainly focused in adult patients and limited to consider the interaction of a maximum of two clinical factors (Carvajal-Rodriguez et al. 2008).

HIV evolution in children has received much less attention than in adults (Carvajal-Rodriguez et al. 2008), even though HIV-1 genetic diversity does not necessarily evolve in the same way in adult and pediatric patients, as clinical features of HIV-1 infection in children and adults are very different. Children present a faster rate of disease progression, substantially higher viremia at early times postinfection and a slower decline after initial infection compared to adult infections (McIntosh et al. 1996). The clinical course of HIV infection in children also varies according to the age of infection and transmission route. Most of the pediatric infections occur in perinatally infected children (Chakraborty et al. 2008) whose immature immune system would exert less selection pressure on the virus than in adult patients, at least in the early stages of infection (Ceballos et al. 2008). Indeed, a large variation in immune responses among pediatric patients has been observed (Becquet et al. 2012). Disease progression has been shown to differ between patients infected through mother-to-child transmission as compared to those infected by other routes (Tobin and Aldrovandi 2013). In a previous study, our group demonstrates that the HIV-1 between-host evolutionary dynamics differs between children and adult populations of Madrid, and identified three clinical factors (age, CD4/mm^3 and antiretroviral experience) as major determinants of HIV-1 population genetic diversity (Pagán et al. 2016). Higher HIV-1 subtype B genetic diversity was observed with increasing child age, decreasing CD4/mm^3 and upon antiretroviral experience (Pagán et al. 2016). However, HIV1 evolutionary dynamics are not the same at the between- and within-host levels (Castro-Nallar et al. 2012). This study analyzes the relative importance of 28 clinical/epidemiological/virological parameters in determining within-host HIV-1 subtype B evolution during longitudinal pediatric HIV-1 infection for a better understanding of HIV evolution in children.

Materials and Methods

Study Population

The Madrid cohort of HIV-infected children and adolescents (established in 2003) registered 561 HIV-1 infected children until March 2016. We selected those perinatally infected patients carrying HIV-1 subtype B (HIV-1B) with three or more available partial *pol* sequences derived from samples collected within a spanning time of at least two years. Following these inclusion criteria, a total of 24 children were enrolled in the study (table 1). The project was approved by the Human Subjects Review Committee at University Hospital Ramón y Cajal (Madrid, Spain), and informed consent of the parents or guardians was obtained.

HIV-1B Sequences

A total of 162 partial HIV-1B *pol* sequences from 24 patients were included in the study. Sequences (1,102 nt) encompassed the complete protease (PR) gene and the nucleotides comprising the first 335 amino acid residues of the reverse transcriptase (RT). For each patient, sequences were obtained at baseline, at least at one intermediate time point, and in the last clinical visit in a time interval ranging from 2.2 to 20.3 years (mean 7.7). From 3 to 24 (mean 7) partial *pol* sequences per patient were included in the analysis (table 1). Sequenced samples were collected during 20 years (from October 1993 to October 2013). Most sequences were generated and previously used by our group for other analyses

Table 1.
Main Features and Available HIV-1 Sequences from the Study Cohort

Features	Patients
HIV-1 vertical transmission	24
HIV-1 subtype B	24
Gender	
Male	15
Female	9
Origin country	
Spain	22
Guatemala	1
Peru	1
Year of infection (birth year)	
1984–1989	7
1990–1999	15
2000–2002	2
Year of HIV-1 diagnosis	
1984–1989	1
1990–1999	19
2000–2004	4
Mean age in years (range)	
At HIV-1 diagnosis[a§]	1.7 (0.1–8.6)
At first ART[b]	3 (0.1–8.5)
At baseline sampled sequence[c]	8.8 (0.3–18.8)
At last sampled sequence[d]	16.8 (5–23.9)
Mean number of HIV-1 sequences per patient (range)	7 (3–21)
Years between the first and the last sequence (range)	7.7 (2.2–20.3)

Unknown data in 6[a], 1[b], 5[c], and 4[d] patients; [§]Thirteen (68%) children were diagnosed before the first year of life, three (15.8%) children were diagnosed between the first and the fourth year of life, and two children were diagnosed between the fourth and the ninth year of life. ART, antiretroviral treatment. HIV-1 sequences, partial *pol* including the complete PR and the first 334 amino acid residues of RT.

(de Mulder et al. 2012; de Mulder et al. 2011; Rojas Sánchez et al. 2015; Rojas Sánchez et al. 2016). For this study, only 7 of the 162 sequences (4.3%) were newly obtained from HIV-1 infected plasma samples provided by the Paediatric HIV BioBank integrated in the Spanish AIDS Research Network (RIS) RD12/0017/0035 and RD12/0017/0037 (García-Merino et al. 2010). Samples were processed following current procedures and frozen immediately after their reception. New HIV-1 sequences were generated as previously reported (de Mulder et al. 2012). Sequence alignments were constructed using MUSCLE 3.7 (Edgar 2004) and adjusted manually according to the amino acid sequences using MEGA 6.0.6 (Tamura et al. 2013). The full list of GenBank accession numbers from the 162 partial *pol* sequences, year of isolation of the sequenced samples and associated relevant clinical parameters is available in supplementary file S1, Supplementary Material online.

Drug Resistance Analysis

Drug resistance mutations (DRM) in pretreated patients were defined following the International AIDS Society—USA (IAS) 2015 list (Wensing et al. 2014). We recorded the DRM to three drug families: nucleoside analogues RT inhibitors (NRTI), non-NRTI (NNRTI) and protease inhibitors (PI). Among drug-naïve patients, transmitted drug-resistance mutations (TDR) were defined according to the mutation list for Transmitted Drug Resistances surveillance as recommended by the WHO (Bennett et al. 2009). Drug susceptibility was predicted using the Stanford HIVdb algorithm (http://sierra2.stanford.edu/sierra/servlet/JSierra), which classifies drug susceptibility in four categories depending on mutation scores: susceptible, low-level, intermediate, and high-level resistance.

Genetic Distances and Selection Pressures

Genetic divergence (*d*) was estimated using the Kimura-2-parameters nucleotide substitution model as implemented in MEGA 6.0.6 (Tamura et al. 2013), which was the best-fitted nucleotide substitution model as determined by jModelTest 2.1.8 (Darriba et al. 2012). Standard errors (SE) of each measure were based on 1,000 bootstrap replicates for permutation tests. Selection pressures were measured as the ratio between the mean number of nonsynonymous (d_N) and synonymous (d_S) nucleotide substitutions per site (d_N/d_S) calculated by the Pamilo–Bianchi–Li method as implemented in MEGA 6.0.6. Individual values of d_N and d_S were also obtained. The d_N/d_S ratio was also estimated at individual codons in the partial *pol* sequence, using different methods implemented in the HYPHY program (SLAC, Single Likelihood Ancestor Counting; FEL, Fixed Effects Likelihood; IFEL, Internal Fixed Effects Likelihood; REL, Random Effects Likelihood; FUBAR, Fast Unbiased Bayesian Approximation) (Kosakovsky and Frost 2005) to determine whether each codon was under negative ($d_N/d_S < 1$), neutral ($d_N/d_S = 1$), or positive ($d_N/d_S > 1$) selection. These analyses were performed after confirmation of the absence of recombinant sequences in our data set by using four different methods available in the RDP4 package: RDP, GENECONV, Bootscan, and Chimaera, and employing the default parameters (Martin et al. 2015). Supplementary file S1, Supplementary Material online includes the observed values for *d*, d_N, d_S, and DRM associated with each viral sequence.

Data on Clinical Factors

We analyzed the influence of 28 clinical/epidemiological/virological factors on HIV-1 subtype B evolution in children. They included five virological parameters (DRM, DRM to NRTI, DRM to NNRTI, DRM to IP, and VL), five clinical parameters (CD4 and CD8 lymphocytes T counts or cells/mm^3 and percent, CD4/CD8 ratio) and 18 epidemiological parameters (children's origin, year of infection (birth year), year of HIV-1 diagnosis, age at HIV-1 diagnosis, coinfection with HBV or HBC, year of sampling sequence, patient's age at HIV-1B sequencing, antiretroviral treatment (ART) exposure (naïve or treated), year of first ART, age at first ART, number of previous ART regimen switches, number of antiretroviral drugs used and drug family experience for NRTI, NNRTI, PI, fusion inhibitor, and integrase inhibitor per patient).

Machine Learning Classification Methods

We applied supervised classification methods to analyze the relative importance of clinical factors in HIV-1B evolution. We constructed multivariate models using HIV-1B d, d_N, d_S, and frequency of DRM as the variables to be predicted, and the previous 28 clinical parameters considered were used as predictors. To build the models, we categorized some of the continuous variables in order to avoid imbalance in the number of instances of each variable. HIV-1B evolutionary parameters were discretized into three categories, according to the three tertiles of the distribution formed by the values of each variable. Thus, HIV-1B evolutionary variables were classified as follows: d (<0.011, 0.011–0.022, >0.022,); d_N (<0.007, 0.007–0.019, >0.019,); and d_S (< 0.002, 0.002–0.075, >0.075). Data related to DRM were discretized into two categories (presence or absence), either considering the different classes of DRM separately (NRTI, NNRTI, and PI) and all classes as a whole. CD4 and CD8 (count and percent), VL, age at sequencing and age at diagnosis were also discretized following the CDC recommendations (Centers for Disease Control 1994). Prior to model selection analyses, we perform Variance Inflation Factor (VIF) analyses to test for predictor collinearity. Given that VIF was smaller than 2 in all predictors, we did not include any variance–covariance matrix or reduced-rank analyses in machine learning methods.

Machine learning methods in Weka (http://www.cs.waikato.ac.nz/ml/weka/) were used to analyze the data. To remove irrelevant variables that could introduce a bias in the predictive models and to evaluate the predictive power of each subset of predictor variables, the Feature Subset Selection (FSS) tool was implemented in the analysis using three different techniques: two univariate methods (1), one multivariate methods (2), and (3) Wrapper method: 1) two univariate algorithms analyze (InfoGainAttributeEval and GainRatioAttributeEval) that explored the importance of each variable separately in the data set; 2) The multivariate algorithm Correlation Feature Selection (CFS) which predicts the value of each individual variable based on the best subset of features using an induction algorithm as a part of the evaluation function; 3) Wrapper methods which uses a search algorithm for predicting the relative usefulness of subsets of variables. To evaluate the contribution of each predictor using Wrapper method, we analyzed the strength of 6 algorithms (classifiers) in order to find the best predictive model: classification tree (J48 tree), Nearest-neighbor (IB-1 and IB-K, where $K = 3$), logistic regression and Bayes algorithm (Naive Bayes and tree-augmented Naive Bayesian Network or TAN) testing the whole data set. To evaluate each algorithm or classifier, a set of measures (confusion matrix, true positive [TP] rate, true negative [TN] rate, precision, accuracy, recall, f-measure, and area under the ROC curve [AUC]) were obtained. For each algorithm, HIV-1B evolutionary parameters were considered as the variables to be predicted, and clinical, epidemiological and virological parameters as the predictor variables. We chose the two best algorithms or classifiers for our data set to evaluate the relative importance of each clinical factor. The clinical factors selected with univariate (InfoGainAttributeEval and GainRatioAttributeEval), multivariate and the best Wrapper algorithms providing the highest values of correctly classified instances and the highest area under the ROC curve (AUC) (were considered the most important variables affecting HIV evolution in the pediatric cohort under study. The ROC curve is a graphical representation of sensitivity versus specificity for a binary classifier system. With a higher area (>0.75), separation of data should be better. Importantly, this representation is independent of the existence of unbalanced sampling).

Results

Clinical, Epidemiological, and Virological Features of the Study Population

A total of 24 patients from the Madrid Cohort of HIV-1 infected children and adolescents were enrolled in the study. Their baseline clinical and epidemiological features are shown in table 1. Most were Spanish (91.7%), male (62.5%), diagnosed before 2000 (83.3%), with a mean age of 3 years at first ART experience and under follow up in pediatric units (table 1). The mean age for collection of the first sequenced viral sample (baseline sequence) was 8.8 years (range 0.3 to 18.8 years) and for the last sampled sequence 16.8 years (range 5–23.9 years). Clinical and virological features of the 24 patients were available in the first clinical report (at collection time of the first sequenced viral sample) and in the last (table 2). In the monitored time interval, we observed an increase in the rate of patients with undetectable VL (≤50 copies/ml) from 8.3% to 17.4%; $\chi^2 = 3.73$; $P = 0.052$. The number of patients with low CD4 counts (<500 CD4 cells/mm^3) also increased from 33.3% to 60.7% ($\chi^2 = 4.52$; $P = 0.033$) (table 2).

We also analyzed the change of viremia and lymphocyte rate over the course of the infection. We calculated the mean values from CD4 and CD8 rates, CD4/CD8 ratio, and VL along the monitored HIV-1B exposure time. We observed an increase in the CD8 rate over time ($r = 0.93$; $P = 0.001$) (fig. 1*A*) from 10% in newborns to 65% at the largest HIV-1 exposure time (age of 18 years). We also noticed a nonsignificant decrease in the CD4 rate ($r = -0.44$; $P = 0.072$) (fig. 1*B*). The CD4/CD8 ratio reached a maximum at early times, and rapidly decreased afterwards, stabilizing its value after 5 years of HIV-1 exposure time. Indeed, we detected a significant logarithmic association between the CD4/CD8 ratio and exposure time ($r = -0.77$;

Table 2.
Clinical Features of the 24 Patients Included in This Study

Features	Number of Patients	
	At First Sampling Time	At Last Clinical Report
Clinical follow-up		
Pediatric Unit	24	11
Adult Unit	0	9
Lost to follow-up	0	1
Exitus	0	1
Unknown	0	2
ART status		
Naive	2	0
Treated	22	24
Regimen switches, mean (range)	3.4 (0–12)	5.9 (2–12)
1–2	9	1
3–6	10	14
7–12	2	6
Unknown	3	3
Previous ARV drugs, mean (range)	5.9 (0–18)	9.7 (5–18)
<3	3	0
3–6	12	4
7–13	5	15
>13	1	4
Unknown	3	1
Number of DRM		
Total	126	147
To NRTI	71	79
To NNRTI	21	34
To PI (major)	34	34
Viral load (HIV-1 RNA-copies/ml)		
≤50	2	4
>50–500	2	2
>500–1,000	2	2
>1,000–10,000	7	8
>10,000–100,000	9	6
>100,000	2	1
Unknown	0	1
CD4+ T rate		
<25%	12	16
25–50%	10	7
>50%	0	0
Unknown	2	1
CD4+ T cells/mm^3		
<350	3	10
350–500	4	4
501–1,000	11	8
1,001–1,500	1	1
>1,500	2	0
Unknown	3	1
CD8+ T cells/mm^3		
<350	1	1
350–500	0	0
501–1,000	3	5
1, 001–1,500	7	5
>1,500	5	6
Unknown	8	7
CD4/CD8 mean ratio (range)	0.8 (0.1–3)[a]	0.4 (0.06–0.9)[b]

Unknown data in 8[a] and 7[b] patients. ART, antiretroviral treatment; ARV, antiretroviral; IQR, interquartile range; DRMs, drug resistance mutations to NRTI, NNRTI and PI (major) according to IAS2015 (Wensing et al., 2014).

$P = 0.006$) (fig. 1*C* and table 2). Finally, VL decreased over time ($r = -0.75$; $P = 0.002$) (fig. 1*C*).

As expected, the 24 patients showed high drug experience and several regimen switches (table 2), and most analyzed viral sequences presented DRMs (supplementary fig. S1, Supplementary Material online), which did not significantly change in number over time (126/1,128 vs. 147/1,128 $\chi^2 = 1.84$, $P = 0.175$) (table 2). The most common DRMs at the last available sequence (last report) were: NRTI mutations T215YF (45.8%), D67N (41.6%), K219QR and L210W (each 37.5%), and M184V (33.3%) at RT; NNRTI mutations Y181C (33.3%), K103NR (29.2%), and G190A (25%) at RT; and PI mutations V82ATS (33.3%), M46I (29.2%), I54V (25%) and L90M (25%) at PR (supplementary fig. S1, Supplementary Material online). Considering each antiretroviral family individually, we observed a similar rate over time of patients carrying DRMs to NRTI (57.3% vs. 53.7%; $\chi^2 = 0.14$; $P = 0.724$) and to PI (27.4% vs. 23.1%; $\chi^2 = 0.49$; $P = 0.476$) but a significant increase of DRM to NNRTI (12.3% vs. 23.1%; $\chi^2 = 3.95$; $P = 0.044$). Despite the presence of DRMs in every viral sequence according to the Stanford algorithm, viruses presented preserved susceptibility to some new antiretrovirals (ARVs) such as darunavir (DRV/r) in six children, tipranavir (TPV/r) in four, and etravirine (ETR) and rilpivirine (RPV) in five patients (fig. 2). During the spanning time between baseline and last analyzed sequence (7.7 years on average, table 1), DRMs to at least one ARV family reverted to wild type (wt) residue in 8 of the 24 children. DRMs to PI, NNRTI, and NRTI reverted to wt in 5, 4, and 2 children, respectively (supplementary table S2, Supplementary Material online). One of the two children with available viral sequence before ARV experience was firstly infected with a resistant virus to PI and NRTI, carrying changes V82A at the PR and L210W and T215S at the RT (supplementary table S2, Supplementary Material online).

Within-Host HIV-1B Genetic Diversity and Selection Pressures

Average within-host HIV-1B genetic diversity in the analyzed virus population was of 0.004 ± 0.001, but varied up to one

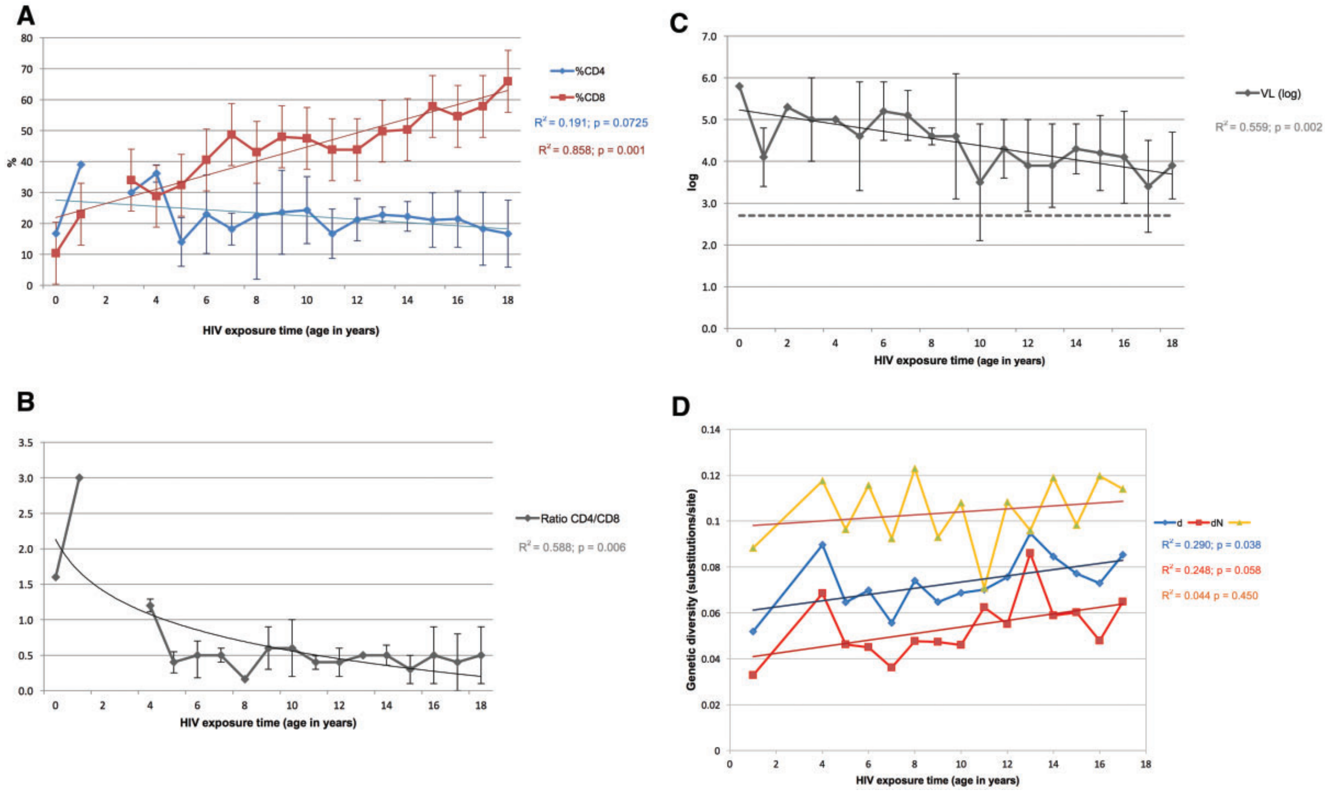

Fig. 1.—Correlations between patient HIV-1B exposure time and T-CD4+ and T-CD8+ rates (*A*), CD4/CD8 ratio (*B*), viral load or VL (*C*) and evolutionary parameters, namely HIV-1B genetic diversity (*D*), mean number of nonsynonymous (d_N) and synonymous (d_S) nucleotide substitutions (*D*). For each correlation the R^2 and the *P* values are shown. Dashed line in panel C refers to undetectable viral load < 2.7 log (<500 HIV-1-RNA copies per milliliter of plasma). *P* values: 0.038 (*d*), 0.058 (d_N), 0.450 (d_S), 0.001 (%CD8), 0.006 (CD4/CD8), and 0.002 (VL).

order of magnitude between patients (*d*: from 0.003 ± 0.000 to 0.040 ± 0.001) (table 3). In addition, averaged nonsynonymous and synonymous diversities were of 0.003 ± 0.001 and 0.008 ± 0.003, respectively, with both evolutionary parameters greatly varying among patients (d_N: from 0.001 ± 0.000 to 0.032 ± 0.001; d_S: from 0.002 ± 0.001 to 0.058 ± 0.003) (table 3). The analyzed *pol* fragment was on average under negative selection (d_N/d_S: 0.565 ± 0.386). Indeed, HIV-*pol* was under negative selection (d_N/d_S significantly smaller than 1) in 14 (58.4%) patients, under positive selection (d_N/d_S values > 1) in 9 children (37.5%) and under neutral evolution in the remaining child (4.1%).

Average evolutionary parameters varied over the course of the infection. Genetic diversity (substitutions per site) increased between the baseline and the last report (from 0.052 ± 0.006 to 0.090 ± 0.014; $r = 0.54$; $P = 0.038$) (fig. 1*D*). This was due to the accumulation of nonsynonymous mutations (from 0.033 ± 0.005 to 0.086 ± 0.004; $r = 0.50$; $P = 0.058$), whereas synonymous mutations remained relatively constant over time (from 0.088 ± 0.009 to 0.123 ± 0.011; $r = 0.21$; $P = 0.450$). When comparing the baseline versus the last collected viral sequences, we also observed a significant increase in the number of sites under neutral evolution (115/334 vs. 293/334 $\chi^2 = 10.32$, $P = 0.002$).

Relative Contribution of Clinical Factors to the Evolution of the Pediatric HIV-1B Population

The relative contribution of each analyzed clinical factor on HIV-1B evolution was obtained after analyzing how each considered clinical factor predicted HIV-1B evolution using six classifiers (J48 tree, IB-1, IB-3, logistic regression, Naive Bayes, and TAN). We observed different accuracy among these classifiers; IB1, IB3 and J48 were the best classifiers. They showed the highest percentage of correctly classified instances (83.1%, 81.4%, and 76.5%, respectively), good precision and high F-Measures (0.83, 0.81, and 0.76) and high AUC (≥0.8) (supplementary file S4, Supplementary Material online). To identify the best predictors of HIV-1B evolution as estimated by IB1, IB3, and J48, we used three different techniques: two univariate algorithms (InfoGainAttributeEval and GainRatioAttributeEval), one multivariate algorithm (CFS) and Wrapper (Wrapper/IB1, Wrapper/IB2, and Wrapper/J48).

Table 4 shows the best predictors of HIV-1B evolutionary parameter and DRM presence, which were more frequently obtained by univariate, multivariate, and wrapped algorithms. Associated with HIV-1B *d*, we found the variable age of HIV-1 diagnosis (values of: 0.105 by GainRatioAttributeEval, 80% by CFS, 100% by Wrapper/IB3 and Wrapper/J48, 70% by

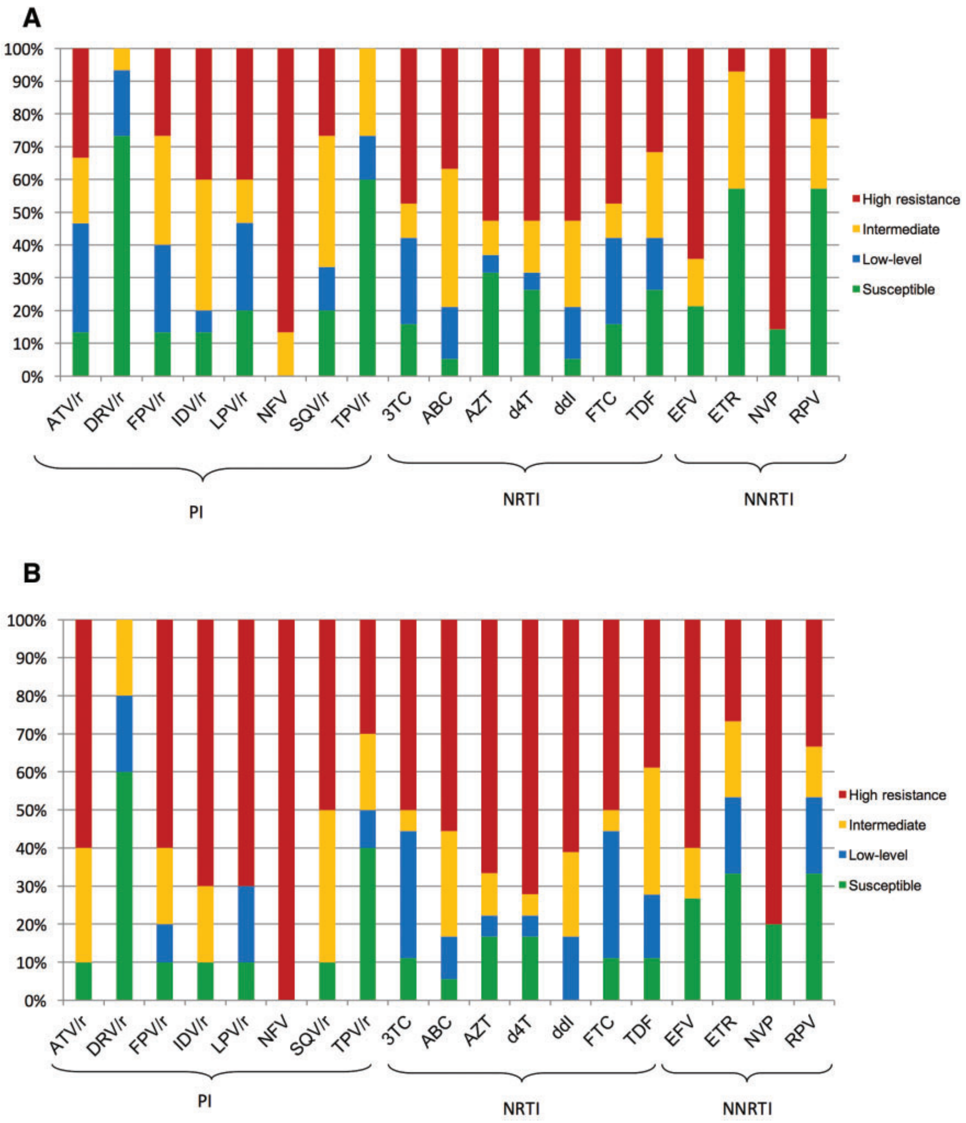

Fig. 2.—Predicted susceptibility according to the Stanford HIVdb Interpretation Algorithm among those virus carrying DRM to PI ($n = 15$), to NRTI ($n = 19$) or to NNRTI ($n = 14$) in the first available partial pol sequence (*A*) and to PI ($n = 10$), to NRTI ($n = 18$) or to NNRTI ($n = 16$) in last sequence (*B*) collected after a mean spanning time of 7.7 years (range 2.2–20.3 years). DRM, drug resistance mutations; NRTI, nucleoside reverse transcriptase inhibitors; NNRTI, non-NRTI; r, ritonavir used for boosting; ATV/r, boosted-atazanavir; DRV/r, boosted-darunavir; FPV/r, boosted-fosamprenavir; IDV/r, boosted-indinavir; LPV/r, boosted-lopinavir; NFV, nelfinavir; SQV/r, boosted-saquinavir; TPV/r, boosted-tipranavir; 3TC, lamivudine; ABC, abacavir; AZT, zidovudine; d4T, estavudine; ddI, didanosine; FTC, emtricitabine; TDF, tenofovir; EFV, efavirenz; ETR, etravirine; NVP, nevirapine; RPV, rilpivirine. The approval year for each drug in Spain was: 1988 (AZT), 1993 (ddI), 1996 (d4T, 3TC, IDV/r, SQV/r), 1998 (NVP, NFV), 1999 (ABC, EFV), 2001 (LPV/r), 2002 (TDF), 2003 (FTC), 2004 (ATV/r, FPV/r), 2005 (TPV/r), 2007 (DRV/r), and 2008 (ETR); data available at Agencia Española de Medicamentos y Productos Sanitarios (https://www.aemps.gob.es/).

Wrapper/IB1). For HIV-1B d_N, the best predictors was age at first ART (values of 0.226 by GainRatioAttributeEval, 60% by CFS and 100% by Wrapper/IB3 and Wrapper/J48); and for HIV1-B d_S, an association with variable year of HIV-1 diagnosis was detected (values of 0.211 by GainRatioAttributeEval, 50% by CFS and 100% by Wrapper/IB1, 90% by Wrapper/IB3, 70% by Wrapper/J48). These three epidemiological variables are interrelated, as most of the children involved in this study had received ART at HIV-1B diagnosis time.

In addition, our analyses indicated that year of infection (birth year) and year of sequencing were relevant for the development of DRMs at large (0.116 by GainRatioAttributeEval, 100% by CFS, for birth year; 0.116 by GainRatioAttributeEval, 100% by CFS and 60% by Wrapper/J48, for year of sequencing), DRMs to PI (0.168 by GainRatioAttributeEval, 100% by CFS and 100% by Wrapper/J48, 90% by Wrapper/IB1 and Wrapper/IB3 and, for birth year; 60% by CFS for year of sequencing), and DRMs to NRTI (0.149 by GainRatioAttributeEval, 60% by

Table 3.

Evolutionary Parameters of the Partial *pol* Sequences Obtained from 24 HIV-1B Population Under Study

Code	HIV Exposure Time (years)	No Sequences	d^1		d_N^2		d_S^3		d_N/d_S^4	
			Mean	SE	Mean	SE	Mean	SE	Mean	SE
P18	2.3	3	0.004	0.001	0.003	0.001	0.008	0.003	0.565	0.386
P11	6	6	0.028	0.008	0.018	0.004	0.050	0.019	1.736	0.508
P14	6.8	5	0.015	0.007	0.008	0.004	0.024	0.012	0.222	0.111
P24	7.3	3	0.028	0.009	0.020	0.005	0.040	0.013	0.628	0.092
P1	9	20	0.017	0.001	0.020	0.002	0.006	0.001	1.919	0.262
P17	10.7	3	0.020	0.004	0.010	0.002	0.044	0.009	0.254	0.022
P16	10.7	5	0.006	0.001	0.004	0.001	0.010	0.003	0.670	0.268
P5	11.7	6	0.006	0.003	0.007	0.004	0.002	0.001	2.444	1.222
P4	12.5	11	0.022	0.004	0.024	0.004	0.013	0.004	1.665	0.416
P2	12.8	20	0.040	0.001	0.030	0.001	0.058	0.003	0.790	0.011
P10	13	4	0.009	0.002	0.007	0.002	0.018	0.003	0.350	0.117
P19	14.3	3	0.018	0.006	0.015	0.006	0.020	0.004	0.657	0.205
P22	14.3	4	0.004	0.001	0.002	0.000	0.007	0.001	0.287	0.038
P3	14.6	18	0.015	0.005	0.008	0.002	0.030	0.009	0.199	0.035
P20	14.7	3	0.008	0.001	0.006	0.003	0.009	0.003	1.215	0.764
P21	14.7	4	0.012	0.005	0.014	0.005	0.005	0.002	1.850	0.513
P9	15.6	5	0.029	0.001	0.026	0.001	0.031	0.002	1.004	0.058
P23	15.7	3	0.007	0.002	0.004	0.001	0.011	0.004	0.235	0.074
P12	15.9	9	0.003	0.000	0.001	0.000	0.006	0.001	0.152	0.054
P13	16	5	0.037	0.011	0.027	0.006	0.053	0.020	2.220	0.818
P12	16	5	0.026	0.001	0.032	0.001	0.008	0.002	1.724	0.145
P8	18.6	5	0.007	0.000	0.006	0.000	0.007	0.000	0.511	0.043
P15	19.8	5	0.020	0.002	0.014	0.001	0.035	0.005	0.359	0.039
P7	25.6	7	0.008	0.002	0.009	0.002	0.007	0.001	1.185	0.168
Total/Average	12	162	0.004	0.001	0.003	0.001	0.008	0.003	0.565	0.386

Patients ordered according to HIV exposure time. *d*, genetic diversity; d_N, frequency of nonsynonymous mutations; d_S, frequency of synonymous mutations; d_N/d_S, selection pressure; SE, standard error.

CFS, 90% by Wrapper/IB1, Wrapper/IB3, and Wrapper J48, for birth year; 90% by CFS and 60% by Wrapper/IB1 for year of sequencing) (supplementary file S4, Supplementary Material online). Other clinical factors had lesser relevance in HIV-1B evolution (both as genetic diversity and DRM fixation) such as year of infection, experience to different ART, number of previous regimen switches, CD8T+ Cell/mm4 and %CD4 and CD4/CD8 ratio (table 4 and supplementary file S4, Supplementary Material online).

To test the robustness of our estimates, we performed FSS analyses, which also indicated that IB1, IB3, and J48 were the supervised classification paradigms that reported the highest values of precision, and identified the same variables as the best predictors of HIV-1B evolution after Infogain, Gain Ratio, and CFS analyses (table 5).

Discussion

This is the first study that analyzes the interactive effects of more than two clinical and virological features in the intrahost evolution of HIV-1 in pediatric patients during a long HIV-1 exposure time. During the HIV infection, clinical, epidemiological, and virological (VL, DRM) parameters fluctuate over time, presenting differences across patients, as we observed in our population study. These parameters may be relevant for the course of the infection and for HIV-1 evolution. Previous results from our group indicate that some of these parameters (children age, decreasing CD4/mm^3, and upon antiretroviral experience) are key determinants of HIV-1 between-host evolution (Pagán et al. 2016). Understanding whether variation in these clinical parameters also affect within-host HIV-1 evolution may help to design more efficient control strategies (Rambaut et al. 2004). Our study showed that the best-predictor variables for HIV-1B evolutionary parameters were the age of HIV-1 diagnosis for *d*, the age at first ART for d_N and the year of HIV-1 diagnosis for d_S. The year of infection (birth year) and the year of sampling seemed to be relevant for fixation of both DRM at large and, considering drug families, to PI.

In this study, we applied machine learning algorithms whose main benefit is the ability to analyze big data sets and construct accurate predictive models, accommodating all types of predictors and response variables (Larrañaga et al. 2006). Importantly, and at odds with most classical methods, these algorithms allow incorporating a stochastic component to

Table 4.
Consensus Best-Predictor Variables for Each Evolutionary Parameter and for Drug Resistance Mutations Presence in the Study Cohort

	Evolutionary parameters			ARV resistance			
	d	d_N	d_S	Major DRMs to PI	DRMs to NRTI	DRMs to NNRTI	DRMs presence
Consensusbest-predictor variables	Age of HIV-1 diagnosis	Age at first ART	Year of infection	Year of infection	Year of infection	NNRTI experience	Sampling year
	Year of infection	NNRTI experience	Year of HIV-1 diagnosis	Patient’s origin	Coinfection with HCV or HBV	No. of ARVs	No. of previous ART regimen switches
	No. of previous ART regimen switches	No. of ARVs	Coinfection with HCV or HBV	Coinfection with HCV or HBV	No. of ARVs	CD4/CD8 ratio	CD4 cel/ mm3switches
	Year at first ART	Year of HIV-1 diagnosis	Year of first ART	Age at first ART	%CD8	%CD4	Year of infection
	Year of HIV-1 diagnosis	CD8 cell counts/mm^3	Age at first ART	No. of previous ART regimen switches	%CD4	CD8 cell counts/mm^3	Year of HIV-1 diagnosis
	NNRTI experience		CD4/CD8 ratio	PI experience	Year of HIV-1 diagnosis	Year of first ART	%CD8
	CD8 cell/m^3		No. of previous ART regimen switches	No. of ARVs	Sampling year	Age at sequencing	
	Iint experience			Sampling year		IP experience	
						CD4 cell/mm^3	
						Age at first ART	
						%CD8	

d, genetic diversity; d_N, frequency of nonsynonymous mutations; d_S, frequency of synonymous mutations; DRMs, drug resistance mutations; PI, protease inhibitor; NRTI, nucleoside reverse transcriptase inhibitor; NNRTI, nonnucleoside reverse transcriptase inhibitor; ART, Antiretroviral treatment; HBV, Hepatitis B viruses; HCV, hepatitis C virus; ARVs, antiretroviral drugs; Iint, Integrase inhibitor. In each column, the variables are sorted by relative importance.

Table 5.
Percentage of Correctly Classified Instances by Each Model Using Wrapper, Gain Ratio, and CFS Data Sets in Our Study Cohort

Model	Data set	Parameter						
			d	d_N	d_S	Major DRM to PI	DRM to NRTI	DRM to NNRTI	DRM presence
J48	Wrapper	Correctly classified instances	47.4%	79.3%	76.4%	91.5%	86.7%	80.6%	93%
	Gain ratio		86.1%	66.8%	72.8%	77.9%	80.6%	79.6%	89.3%
	CFS		68%	68.9%	63.6%	81.2%	85.1%	80.1%	90.6%
IB1	Wrapper		37.1%	69.4%	72.8%	87.7%	82.6%	80.1%	86.8%
	Gain ratio		78.4%	73.6%	75.9%	92.5%	83.7%	86.6%	92.5%
	CFS		76.3%	68.4%	67.7%	85.5%	82.4%	83.3%	93.1%
IB3	Wrapper		47%	68%	67%	88.7%	85%	79%	91.8%
	Gain ratio		74.2%	70.9%	75.4%	85.9%	83.2%	87.1%	93.1%
	CFS		73.7%	68.4%	64.1%	73.7%	89.2%	85.5%	94.9%

Predictive models: J48, Classification tree; IB1 and IB3, Nearest-neighbor; CFS, Correlation Feature Selection; *d*, genetic diversity; d_N, frequency of nonsynonymous mutations; d_S, frequency of synonymous mutations; PI, protease inhibitor; NRTI, nucleoside reverse transcriptase inhibitor; NNRTI, non nucleoside reverse transcriptase inhibitor; DRM, drug resistance mutations.

model construction. We used six supervised classification methods to predict three HIV-1B evolutionary parameters (d, d_N, and d_S) and the fixation of DRM. Supervised classification techniques are algorithms with high predictive power and are designed to optimize the statistical classification procedures (Stephens and Diesing 2014). Among these six methods, IB1, IB3, and J48 generated the best predictor models. Although not all these methods allow building an explicit predictive model, our results suggest that these showed advantages over other methods since they required less preprocessing, had a better performance in the presence of interacting features, generally required less training data to learn good settings, and were more cost-efficient than others. Although multivariate analyses have been extensively used to study others HIV-infected cohorts and clinical studies (Reddy et al. 2016; Auld et al. 2016; Gilbert et al. 2016; Cakır and Demirel 2011; Sahle 2016), this is the first time that machine learning techniques have been applied to understand the importance of clinical parameters in determining within-host HIV-1 subtype B.

Most HIV-infected children under study were born in the 80's–90's (75%), had detectable VL (>50 c/ml) at last sequence (83.3%) and had monotherapy and/or dual therapy experience (34.8%), leading to treatment failure and reducing ART efficacy due to the incomplete virus suppression after DRM selection (Lorenzo et al. 2004; Abrams et al. 1998; Rojas Sánchez and Holguín 2014). In fact, treatment failure in children during ART is frequent, leading to immunological damage (Judd et al. 2016) and the emergence of DRMs is a major obstacle for effective treatment (Rojas Sánchez and Holguín 2014). Clinicians face problems of managing heavily pretreated perinatally infected patients with many resistance mutations, not completely adherent to the treatments or with previous suboptimal regimens, such as in those children born in the mono or biotherapy era or in areas with limited ARV availability. In our study, all 24 patients (except one child) carried viruses with DRMs at first and last available sequence, although they maintained susceptibility to some antiretroviral drugs licensed or under evaluation to control HIV pediatric infection (https://www.aemps.gob.es/). The high resistance rate to most ARVs across resistant viruses reflects the change of treatment choices by clinicians during the last decades depending on the approval time of ARV for pediatric use in Spain (fig. 2).

The effect of ART on plasma HIV-1 diversity has been studied in adults (Lorenzo et al. 2004; Pennings 2012; Gall et al. 2013; Kearney et al. 2014). We previously observed an increase of the mean between-host virus diversity during ART exposure across pediatric patients from the same cohort of the 24 children under study (Pagán et al. 2016), indicating that plasma virus diversity is sustained during each phase of viral decay despite the large decreases in the replicating population size. Our results obtained by machine learning showed that those variables related with ART (as year of first ART, NNRTI experience, number of ARVs and of previous ART regimen switches) had an effect on HIV-1 evolutionary parameters. Therefore, the effect of these variables on within-host HIV-1 subtype B evolution should be analyzed in more detail in future studies. The NNRTI experience seems to have a direct effect on genetic diversity and frequency of nonsynonymous changes at *pol* coding region. This result reinforces the importance of treatment on HIV-1 evolution, since virus replication continues in patients under ART, even in patients with viral suppression and persistent low-level viraemia (Martinez-Picado and Deeks 2016; Vardhanabhuti et al. 2015).

It is known that the evolutionary pathway and HIV evolution at *pol* can also be highly dependent on the viral genetic background, at DRM as well as nonDRM sites (Rath et al. 2013). We observed that the within-host HIV-1B genetic diversity, nonsynonymous and synonymous mutation rates and selection pressures also changed among patients, probably due to the different selection forces at sampling time and different viral genetic background across patients. In the study we observed a similar number of children carrying DRM over time for NRTI and PI. However, we observed an increase of DRM to NNRTI in the last 5 years, maybe due to the reduced ARVs options in children with high virological failure experience and to the approval of two new NNRTI during that period.

Genetic diversity significantly varied over time, increasing when comparing the baseline versus the last collected viral sequences, mainly due to the accumulation of nonsynonymous mutations. This could be favored by clinical changes over time in patients, including decline of CD4/CD8 ratio, incomplete virus suppression, suboptimal therapies and higher experience to ARVs that promote the fixation of DRM (Markham et al. 1998; Castro-Nallar et al. 2012; Rojas Sánchez and Holguín 2014). Note that the number of sites under neutral evolution also increased over time, which would be in apparent contradiction with the observed increase in nonsynonymous mutations. However, if accumulation of nonsynonymous mutations occurred only in a few positions (for instance, DRM-related sites), the rest could have evolved towards neutrality. Indeed, we detected that the number of DRMs increased over time, which would be compatible with this explanation.

According to machine-learning algorithms, other clinical, virological, and epidemiological factors, although lesser, could play a relevant role in DRM fixation and HIV-1B evolution. Thus, VL would not have had an impact on the three HIV-1B evolutionary studied parameters (d, d_N, and d_S) in infected children. However other clinical factors (%T CD4, T CD8 count, and CD4/CD8 ratio) appear to contribute to HIV-1 evolution in children. Twenty-three of the 24 children failed to normalize the CD4/CD8 ratio in the last clinical report, even despite VL suppression or low viraemia after effective ART in most of them. The CD4/CD8 ratio is a surrogate marker of immune activation and immunosenescence (Serrano-Villar et al. 2013; Sainz et al. 2013). In the present study we observed a significantly decreased CD4/CD8 ratio over time. This could be due to the higher immune activation and immunosenescence caused by the long-time HIV infection present in vertically HIV-1-infected patients, whose immune system has developed in the presence of the virus since birth or pregnancy (Sainz et al. 2013). Consequently, all of these interactive clinical factors could modify the HIV-1 replicative environment, affecting the genetic diversity of HIV-1 in children in agreement with previous reports (Carvajal-Rodriguez et al. 2008; Ryland et al. 2010).

Despite the robustness of our predictive models, this study presents limitations. Since all 24 children were perinatally infected, we assumed that HIV-1 was transmitted at delivery time to consider the infant age as HIV-1 exposure time. The exact time of HIV infection is difficult to estimate in infants/children acquiring the HIV infection from their HIV-infected mothers. It will depend on when HIV transmission from mother to child occurred: during 9 months pregnancy, at delivery or during breastfeeding (variable duration after birth). Hence, our estimates of time of infection may have a gap of plus minus a year. However, previous studies indicated little variation in HIV-1 *d*, d_N, and d_S during the first years of infection (Carvajal-Rodriguez et al. 2008). Thus, we do not expect that this uncertainty would have big effects in our results. Another potential caveat is the number of patients included in the study. Although this number and that of analyzed sequences were modest in size, *pol* sequences were sufficient in number to perform an intrahost HIV-1 evolution analysis. Moreover, associated clinical/immunological/virological information was available in all analyzed sequences at sampling time and during the clinical follow up of each patient, and this allowed constructing accurate predictive models. Another limitation of our work is that the monitored time span differed between patients. Using sequences collected during the same time interval would be more a consistent strategy. However, this would have limited our analyses to a time span of 2–3 years, and the analyses shown in figure 1 suggested that few evolutionary changes occur on the HIV-1 genome during such short period of time. Thus, we prioritized an approach that allowed monitoring virus evolution for a longer period. Finally, since only partial *pol* coding region was analyzed, it would be interesting to perform the same analyses using complete *pol* sequences to study the molecular HIV evolution on each three HIV-1 *pol* proteins (PR, RT, and integrase).

In summary, this study identifies for the first time using machine learning and using univariate and multivariate methods, several factors affecting HIV-1B *pol* evolution and those affecting DRM fixation in HIV-1B infected pediatric patients, with high values of precision. More studies are required for a better understanding of HIV evolution across patients and viral genes in children and adolescents with HIV infection.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Disclaimer: The content is solely the responsibility of the authors.

Acknowledgments

This study was supported by Plan Estatal de Investigación Científica y Técnica y de Innovación 2013–2016 and Instituto de Salud Carlos III (PI15/01005 grant) and cofinanced by the European Development Regional Fund "A way to achieve Europe" (ERDF) and CIBER CB06/02/0053. This study is included in the "Subprograma de Inmigración y Salud from CIBERESP (Spain)." IP was supported by a Ramón y Cajal fellowship (RYC-2011-08574).

We want to particularly acknowledge the patients in this study for their participation and the Paediatric HIV BioBank integrated in the Spanish AIDS Research Network for providing two clinical samples used in this work. The Paediatric HIV BioBank, integrated in the Spanish AIDS Research Network, is supported by Instituto de Salud Carlos III, Spanish Health Ministry (Grant no RD06/0006/0035) and Fundación para la investigación y prevención del SIDA en España (FIPSE). CoRISPe cohort is funded by RD12/0017/0035 and RD12/0017/0037 Project, supported by Plan Nacional de I+D+I and cofinanced by the ISCIII-Subdirección General de Evaluación and European Development Regional Fund "A way to achieve Europe" (ERDF). We thank CoRISPE and the professionals responsible for the routine resistance testing in the participant hospitals for providing epidemiological-clinical data and *pol* sequences of enrolled patients when available.

Literature Cited

Abecasis AB, Vandamme AM, Lemey P. 2009. Quantifying differences in the tempo of human immunodeficiency virus type 1 subtype evolution. J Virol. 83(24): 12917–12924.

Abrams EJ, et al. 1998. Association of human immunodeficiency virus (HIV) load early in life with disease progression among HIV-infected infants. J Infect Dis. 178(1): 101–108.

Ammaranond P, Sanguansittianan S. 2012. Mechanism of HIV antiretroviral drugs progress toward drug resistance. Fundam Clin Pharmacol. 26(1): 146–161.

Auld E, et al. 2016. HIV infection is associated with shortened telomere length in ugandans with suspected tuberculosis. PLoS One 11(9): e0163153.

Becquet R, et al. 2012. Children who acquire HIV infection perinatally are at higher risk of early death than those acquiring infection through breastmilk: a meta-analysis. Plos One 7(2): e28510.

Bennett DE, et al. 2009. Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. PLoS One 4(3): e4724.

Cakır A, Demirel B. 2011. A software tool for determination of breast cancer treatment methods using data mining approach. J Med Syst. 35(6): 1503–1511.

Carvajal-Rodriguez A, et al. 2008. Disease progression and evolution of the HIV-1 env gene in 24 infected infants. Infect Genet Evol. 8(2): 110–120.

Castro-Nallar E, Pérez-Losada M, Burton GF, Crandall KA. 2012. The evolution of HIV: inferences using phylogenetics. Mol Phylogenet Evol. 62(2): 777–792.

Ceballos A, et al. 2008. Lack of viral selection in human immunodeficiency virus type 1 mother-to-child transmission with primary infection during late pregnancy and/or breastfeeding. J Gen Virol. 89(Pt 11): 2773–2782.

Centers for Disease Control. 1994. Revised classification system for human immunodeficiency virus infection in children less than 13 years of age. MMWR. Available from: http://www.cdc.gov/MMWr/preview/mmwrhtml/00032890.htm, last accessed September 22, 2017.

Chakraborty R, et al. 2008. HIV-1 drug resistance in HIV-1-infected children in the United Kingdom from 1998 to 2004. Pediatr Infect Dis J. 27(5): 457–459.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9(8): 772.

de Mulder M, et al. 2011. Drug resistance prevalence and HIV-1 variant characterization in the naive and pretreated HIV-1-infected paediatric population in Madrid, Spain. J Antimicrob Chemother. 66: 2362–2371.

de Mulder M, et al. 2012. Trends in drug resistance prevalence in HIV-1-infected children in Madrid: 1993 to 2010 analysis. Pediatr Infect Dis J. 31(11): e213–e221.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5): 1792–1797.

Gall A, et al. 2013. Restriction of V3 region sequence divergence in the HIV-1 envelope gene during antiretroviral treatment in a cohort of recent seroconverters. Retrovirology 10(1): 8.

García-Merino I, et al. 2010. Pediatric HIV BioBank: a new role of the Spanish HIV BioBank in pediatric HIV research. AIDS Res Hum Retroviruses 26(2): 241–424.

Gilbert PB, Huang Y, Janes HE. 2016. Modeling HIV vaccine trials of the future. Curr Opin HIV AIDS 11(6): 620–627.

Gray RR, et al. 2011. The mode and tempo of hepatitis C virus evolution within and among hosts. BMC Evol Biol. 11: 131.

Holmes EC. 2009. The evolution and emergence of RNA viruses. New York, NY: Oxford University Press.

Judd A, et al. 2016. Higher rates of triple-class virological failure in perinatally HIV-infected teenagers compared with heterosexually infected young adults in Europe. HIV Med. doi: 10.1111/hiv.12411.

Kearney MF, et al. 2014. Lack of detectable HIV-1 molecular evolution during suppressive antiretroviral therapy. PLoS Pathog. 10(3): e1004010.

Kosakovsky PSL, Frost SD. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 22(5): 1208–1222.

Larrañaga P, et al. 2006. Machine learning in bioinformatics. Brief Bioinformatics. 7(1):86–112

Lemey P, Rambaut A, Pybus OG. 2006. HIV evolutionary dynamics within and among hosts. AIDS Rev. 8(3): 125–140.

Lemey P, et al. 2007. Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. PLoS Comput Biol. 3(2): e29.

Lorenzo E, et al. 2004. Influence of CD4+ T cell counts on viral evolution in HIV-infected individuals undergoing suppressive HAART. Virology 330(1): 116–126.

Maldarelli F, et al. 2013. HIV populations are large and accumulate high genetic diversity in a nonlinear fashion. J Virol. 87(18): 10313–10323.

Mani I, et al. 2002. Intrapatient diversity and its correlation with viral setpoint in human immunodeficiency virus type 1 CRF02_A/G-IbNG infection. J Virol. 76(21): 10745–10755.

Markham RB, et al. 1998. Patterns of HIV-1 evolution in individuals with differing rates of CD4 T cell decline. Proc Natl Acad Sci U S A. 95(21): 12568–12573.

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 1(1): vev003. 2015,

Martinez-Picado J, Deeks SG. 2016. Persistent HIV-1 replication during antiretroviral therapy. Curr Opin HIV AIDS 11(4): 417–423.

McIntosh K, et al. 1996. Age- and time-related changes in extracellular viral load in children vertically infected by human immunodeficiency virus. Pediatr Infect Dis J. 15(12): 1087–1091.

Moradigaravand D, Kouyos R, Hinkley T, Haddad M, et al. 2014. Recombination accelerates adaptation on a large-scale empirical fitness landscape in HIV-1. PLoS Genet. 10(6): e1004439.

Pagán I, Rojas P, Ramos JT, Holguín A. 2016. Clinical determinants of HIV-1B between-host evolution and their association with drug resistance in pediatric patients. PLoS One 11(12): e0167383.

Pennings PS. 2012. Standing genetic variation and the evolution of drug resistance in HIV. PLoS Comput Biol. 8(6): e1002527.

Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span and viral generation time. Science 271(5255): 1582–1586.

Rambaut A, Posada D, Crandall KA, Holmes EC. 2004. The causes and consequences of HIV evolution. Nat Rev Genet. 5(1): 52–61.

Rath BA, et al. 2013. Evolution in response to triple reverse transcriptase inhibitors & in silico phenotypic analysis. PLoS One 8(4): e61102.

Reddy EA, et al. 2016. Test site predicts HIV care linkage and antiretroviral therapy initiation: a prospective 3.5 year cohort study of HIV-positive testers in northern Tanzania. BMC Infect Dis. 16: 497.

Rojas Sánchez P, et al. 2015. Clinical and virologic follow-up in perinatally HIV-1-infected children and adolescents in Madrid with triple-class antiretroviral drug-resistant viruses. Clin Microbiol Infect. 21(6): 605.e1–609.

Rojas Sánchez P, et al. 2017. Trends in drug resistance prevalence, HIV-1 variants and clinical status in HIV-1- infected paediatric population in Madrid: 1993–2015 Analysis. Pediatr Infect Dis J. (in press).

Rojas Sánchez P, Holguín A. 2014. Drug resistance in the HIV-1-infected paediatric population worldwide: a systematic review. J Antimicrob Chemother. 69(8): 2032–6042.

Ryland EG, Tang Y, Christie CD, Feeney ME. 2010. Sequence evolution of HIV-1 following mother-to-child transmission. J Virol. 84(23): 12437–12444.

Sahle G. 2016. Ethiopic maternal care data mining: discovering the factors that affect postnatal care visit in Ethiopia. Health Inf Sci Syst. 4: 4.

Sainz T, et al. 2013. The CD4/CD8 ratio as a marker T-cell activation, senescence and activation/exhaustion in treated HIV-infected children and young adults. AIDS 27(9): 1513–1519.

Santoro MM, Perno CF. 2013. HIV-1 Genetic variability and clinical implications. ISRN Microbiol. 2013: 481314.

Sede M, Jones LR, Moretti F, Laufer N, Quarleri J. 2014. Inter and intra-host variability of hepatitis C virus genotype 1a hypervariable envelope coding domains followed for a 4-11 year of human immunodeficiency virus coinfection and highly active antiretroviral therapy. Virology 471-473: 19–28.

Serrano-Villar S, Deeks SG. 2015. CD4/CD8 ratio: an emerging biomarker for HIV. Lancet HIV 2(3): e76–e77.

Serrano-Villar S, Gutiérrez C, Vallejo A, Hernández-Novoa B, et al. 2013. The CD4/CD8 ratio in HIV-infected subjects on effective ART could be a surrogate marker of immune activation and possibly of immunosenescence. J Infect. 66: 57–66.

Shriner D, Liu Y, Nickle DC, Mullins JI. 2006. Evolution of intrahost HIV-1 genetic diversity during chronic infection. Evolution 60(6): 1165–1176.

Stephens D, Diesing M 2014. A comparison of supervised classification methods for the prediction of substrate type using multibeam acoustic and legacy grain-size data. PLoS One 9(4): e93950.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 30(12): 2725–2729.

Tobin NH, Aldrovandi GM. 2013. Immunology of pediatric HIV infection. Immunol Rev. 254(1): 43–169.

Vardhanabhuti S, Taiwo B, Kuritzkes DR, Eron JJ, Jr, Bosch RJ. 2015. Phylogenetic evidence of HIV-1 sequence evolution in subjects with persistent low-level viraemia. Antivir Ther. 20(1): 73–76.

Wensing AM, et al. 2014. Update of the drug resistance mutations in HIV-1. Top Antivir Med. 22(3): 642–650.

Zhuang J, et al. 2002. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. J Virol. 76(22): 11273–11282.

Associate editor: Mary O' Connell