



## Original Article

# CRF19\_cpx is an Evolutionary fit HIV-1 Variant Strongly Associated With Rapid Progression to AIDS in Cuba



Vivian Kouri <sup>a,1</sup>, Ricardo Khouri <sup>b,c,\*</sup>, Yoan Alemán <sup>a</sup>, Yeissel Abrahantes <sup>a</sup>, Jurgen Vercauteren <sup>b</sup>, Andrea-Clemencia Pineda-Peña <sup>b,d</sup>, Kristof Theys <sup>b</sup>, Sarah Megens <sup>b</sup>, Michel Moutschen <sup>e</sup>, Nico Pfeifer <sup>f</sup>, Johan Van Weyenbergh <sup>b</sup>, Ana B. Pérez <sup>a</sup>, Jorge Pérez <sup>a</sup>, Lissette Pérez <sup>a</sup>, Kristel Van Laethem <sup>b</sup>, Anne-Mieke Vandamme <sup>b,g,\*\*</sup>

<sup>a</sup> Virology Department, Institute of Tropical Medicine Pedro Kourí, Autopista Novia del Mediodía Km 6, Marianao 13, Havana City, Cuba

<sup>b</sup> KU Leuven, Department of Microbiology and Immunology, Rega Institute for Medical Research, Laboratory of Clinical and Epidemiological Virology, B-3000 Leuven, Belgium

<sup>c</sup> LIM1-LIP, Centro de Pesquisa Gonçalo Moniz, FIOCRUZ, Salvador-Bahia, Brazil

<sup>d</sup> Clinical and Molecular Infectious Diseases Group, Faculty of Sciences and Mathematics, Universidad del Rosario, Bogotá, Colombia

<sup>e</sup> AIDS Reference Center, Centre Hospitalier Universitaire de Liège, Liège, Belgium

<sup>f</sup> Department of Computational Biology and Applied Algorithmics, Max Planck Institute for Informatics, Campus E1 4, 66123 Saarbrücken, Germany

<sup>g</sup> Centro de Malária e outras Doenças Tropicais and Unidade de Microbiologia, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisbon, Portugal

## ARTICLE INFO

## Article history:

Received 17 October 2014

Received in revised form 22 January 2015

Accepted 26 January 2015

Available online 28 January 2015

## Keywords:

CRF19

HIV-1

Variant

Progression to AIDS

Cuba

## ABSTRACT

**Background:** Clinicians reported an increasing trend of rapid progression (RP) (AIDS within 3 years of infection) in Cuba.

**Methods:** Recently infected patients were prospectively sampled, 52 RP at AIDS diagnosis (AIDS-RP) and 21 without AIDS in the same time frame (non-AIDS). 22 patients were sampled at AIDS diagnosis (chronic-AIDS) retrospectively assessed as >3 years infected. Clinical, demographic, virological, epidemiological and immunological data were collected. *Pol* and *env* sequences were used for subtyping, transmission cluster analysis, and prediction of resistance, co-receptor use and evolutionary fitness. Host, immunological and viral predictors of RP were explored through data mining.

**Findings:** Subtyping revealed 26 subtype B strains, 6 C, 6 CRF18\_cpx, 9 CRF19\_cpx, 29 BG-recombinants and other subtypes/URFs. All patients infected with CRF19 belonged to the AIDS-RP group. Data mining identified CRF19, oral candidiasis and RANTES levels as the strongest predictors of AIDS-RP. CRF19 was more frequently predicted to use the CXCR4 co-receptor, had higher fitness scores in the protease region, and patients had higher viral load at diagnosis.

**Interpretation:** CRF19 is a recombinant of subtype D (C-part of Gag, PR, RT and nef), subtype A (N-part of Gag, Integrase, Env) and subtype G (Vif, Vpr, Vpu and C-part of Env). Since subtypes D and A have been associated with respectively faster and slower disease progression, our findings might indicate a fit PR driving high viral load, which in combination with co-infections may boost RANTES levels and thus CXCR4 use, potentially explaining the fast progression. We propose that CRF19 is evolutionary very fit and causing rapid progression to AIDS in many newly infected patients in Cuba.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Despite a low HIV-1 prevalence, diverse clades are circulating in Cuba (Thomson and Nájera, 2005) mainly subtypes B, C, G and H, BG recombinant forms and complex circulating recombinant forms (CRFs)

CRF18\_cpx and CRF19\_cpx (Thomson and Nájera, 2005; Casado et al., 2005; Sierra et al., 2007; Delatorre and Bello, 2013). Whereas the BG recombinants have a Cuban origin, CRF18\_cpx and CRF19\_cpx seem African (Thomson and Nájera, 2005; Casado et al., 2005; Sierra et al., 2007; Delatorre and Bello, 2013), but became epidemic only in Cuba. This diversity is in contrast with the almost exclusive subtype B epidemic (94%) in the rest of the Caribbean (Vaughan et al., 2003). From 2008 to 2011, 1200 to 1800 individuals have been newly diagnosed with HIV-1 in Cuba each year, of whom 13 to 16% had AIDS at diagnosis, more than in the previous decade (Abrahantes Rodríguez et al., 2010).

Rapid progression to AIDS (within 3 years after seroconversion (Casado et al., 2010)) has been associated to virus genetics (Pushker

\* Corresponding author at: KU Leuven, Department of Microbiology and Immunology, Rega Institute for Medical Research, Laboratory of Clinical and Epidemiological Virology, B-3000 Leuven, Belgium.

\*\* Corresponding author.

E-mail address: [annemie.vandamme@uzleuven.be](mailto:annemie.vandamme@uzleuven.be) (A.-M. Vandamme).

<sup>1</sup> The first two authors contributed equally to this study.

et al., 2010; Lemey et al., 2007; Ng et al., 2014), epidemiological factors (Helleberg et al., 2013; Murphy et al., 2013), co-receptor use (Kaleebu et al., 2007), co-infection (Secor, 2012; Pawlowski et al., 2012; Alemu et al., 2013; Hernandez and Sherman, 2011), and immune activation (Liovat et al., 2012; Giorgi et al., 1999; Roberts et al., 2011). Rapid progression is associated with poorer cellular immune response (Demarest et al., 2001) and, occasionally, also to poorer HIV-specific humoral immune response (Sapsutthipas et al., 2013; Bártolo et al., 2009). In addition, higher expression of immunosuppression markers (Lajoie et al., 2009) and specific host genetic variants (Casado et al., 2010; Le Clerc et al., 2009; Limou et al., 2009) is also associated with AIDS rapid progression. Compared to subtype A, subtype D is associated with a higher risk of progressing to AIDS (Baeten et al., 2007; Kaleebu et al., 2002). Infection with A/D recombinant viruses resulted in significantly faster progression to AIDS than did infection with pure subtype A (Kiwunuka et al., 2008). Phenotypic switching from CCR5 to CXCR4-using variants is well-known to be associated with rapid progression in subtype B (Esbjörnsson et al., 2010), while recent data report similar fast progression in patients infected with CXCR4-using variants of other clades (Li et al., 2014). However, no subtype or CRF has been exclusively associated with rapid progression.

To investigate this potentially increased proportion of patients with rapid progression to AIDS over the last decade in Cuba, we set up an exploratory study evaluating the association of rapid progression with epidemiological, clinical, viral and immunological parameters, comparing three groups of patients. The first two groups were newly diagnosed patients with an HIV negative test less than three years earlier recruited prospectively who within 3 years after seroconversion either were already diagnosed with AIDS (AIDS-RP) or were still AIDS-free (non-AIDS) and chronically infected HIV-1 patients with AIDS recruited retrospectively (chronic-AIDS).

## 2. Patients and Methods

### 2.1. Study Population

From November 2007 onwards, adult HIV-1 patients attending the Institute for Tropical Medicine “Pedro Kourí” (IPK) for medical care were consecutively recruited when 1) having a negative HIV test

less than 3 years before HIV-1 diagnosis and 2) being therapy-naive (Table 1). HIV seroconversion date was the midpoint between last negative and first positive HIV test. AIDS was defined if CD4 count <200 cells/μl, CD4% <14% or if at least 1 of any 26 enlisted opportunistic infections and/or cancers were documented (Rutledge and Boyd, 2008). Patients were recruited into the group of a) rapid progressors (AIDS-RP) when diagnosed with AIDS within 3 years after seroconversion and had at least opportunistic infection or previous consecutive (with at least 1 month interval) CD4 counts <350 cells/μl to exclude potential false-positive AIDS diagnosis due to transient low CD4 counts during acute HIV-1 infection (Casado et al., 2010); b) non-AIDS patients when still AIDS-free 3 years after seroconversion (non-AIDS); or c) as chronically infected AIDS patients (chronic-AIDS) when diagnosed with AIDS during the study period more than 3 years (judged retrospectively) after seroconversion. Median year of HIV-1 diagnosis for chronic-AIDS patients was 2001 (IQR = 1997–2002) and of AIDS definition it was 2008 (IQR = 2008–2008) (Supplementary figure 1A). Recruitment into the study did not interfere with standard medical care, and patients were treated according to the prevailing Cuban/WHO guidelines. The study was in accordance with the Helsinki Declaration and approved by the ethics committee of IPK. All participants provided written informed consent.

### 2.2. Measurement of Immunological Markers and Viral Load

CD4 cell counts were determined by FACScan (Becton Dickinson, USA), plasma HIV-1 viral loads by NucliSENS (Biomérieux, France), cytokine and chemokine plasma levels by FlowCytomix (eBioscience, USA): human IL-1b, IL-2, IL-4, IL-5, IL-6, IL-8, IL-10, IL-12p70, G-CSF, TNF-α, TNF-β, IFN-α and IFN-γ, MIP-1α, MIP-1β, IP-10, MIG, MCP-1 and RANTES. Neopterin (IBL Inc., Germany), β2-microglobulin (Immundiagnostik AG, Germany), FasL (IBL Inc., Germany) and sFas (IBL Inc., Germany) plasma levels were quantified using ELISA kits.

### 2.3. HIV Genetic Analyses

Population-based bi-directional Sanger sequencing of *pol* and *env* fragments resulted in on average 1300 NT for *pol* (HXB2 NT position 2253–3554) and 2078 NT for *env* (HXB2 NT position 6417–8497) (Pérez et al., 2013; Van Laethem et al., 2005) (Supplementary Table 1).

**Table 1**

Patient characteristics at HIV diagnosis for the non-AIDS, chronic-AIDS and AIDS-RP groups. Data are expressed as median values with interquartile ranges, or as proportion (%) with number of patients between brackets. Disease progression statistical differences were tested using Kruskal–Wallis or Chi-square test for trend. Statistical tests were considered significant at p-value <0.05 (displayed in bold). N = number of patients. NS = not significant.

	Non-AIDS (N = 21)	Chronic-AIDS (N = 22)	AIDS-RP (N = 52)	p value
Last negative HIV test before HIV diagnosis (months)	22.0 (10.5–25.0) (N = 21)	22.0 (12.0–24.0) (N = 22)	14.0 (5.3–23.5) (N = 52)	0.17
Log VL at HIV diagnosis (RNA copies/ml plasma)	<1.7 (<1.7–3.5) (N = 12)	3.2 (<1.7–3.9) (N = 8)	4.8 (4.1–5.4) (N = 41)	<b>&lt;0.0001</b>
CD4% at HIV diagnosis	27 (19–35) (N = 20)	24 (19–30) (N = 21)	13 (10–19) (N = 49)	<b>&lt;0.0001</b>
CD4 count at HIV diagnosis (cells per μl blood)	577 (441–815) (N = 20)	522 (374–782) (N = 22)	276 (152–386) (N = 49)	<b>&lt;0.0001</b>
Age at HIV diagnosis (years)	29 (22–37) (N = 20)	31 (24–35) (N = 22)	34 (26–41) (N = 51)	0.34
Number of sexual partners in the year before diagnosis	2.0 (1.0–4.8) (N = 20)	2.0 (1.0–5.3) (N = 22)	2.0 (1.0–3.0) (N = 51)	0.54
Acute retroviral syndrome at HIV diagnosis	65.0% (13/20)	50.0% (11/22)	45.1% (23/51)	0.15
Gender (proportion of male)	85.0% (17/20)	86.4% (19/22)	74.5% (38/51)	0.24
Skin color				
White	60.0% (12/20)	45.5% (10/22)	64.7% (33/51)	0.49
Black	15.0% (3/20)	18.2% (4/22)	23.5% (12/51)	0.40
Mulatto	25.0% (5/20)	36.4% (8/22)	11.8% (6/51)	0.09
Anal sex (as reported by both males and females of various sexual orientation)	80.0% (16/20)	54.6% (12/22)	50.0% (25/50)	<b>0.031</b>
Always condom use after diagnosis	45.0% (9/20)	45.5% (10/22)	19.6% (10/51)	<b>0.016</b>
Sexual orientation				
Heterosexual	20.0% (4/20)	36.4% (8/22)	49.0% (25/51)	<b>0.023</b>
Homosexual	65.0% (13/20)	45.5% (10/22)	41.2% (21/51)	0.087
Bisexual	15.0% (3/20)	18.2% (4/22)	9.8% (5/51)	0.44
Residence in Havana	66.7% (14/21)	71.4% (15/21)	68.0% (34/50)	0.97
Sexual contact with known HIV seropositive	55.0% (11/20)	81.8% (18/22)	54.9% (28/51)	0.60
AIDS definition criteria CD4 low (<200 cells/mm <sup>3</sup> or <14%)	–	95.46% (21/22)	98.08% (51/52)	0.52
Waste Syndrome	–	9.09% (2/22)	9.62% (5/52)	0.94
Opportunistic infections	–	13.64% (3/22)	25.00% (13/52)	0.28
<i>Pneumocystis jirovecii</i>	–	0.00% (0/22)	13.46% (7/52)	0.071

Initial subtype classification was using COMET version 2 (<http://comet.retrovirology.lu>) (Struck et al., 2010) and REGA version 3 (<http://regatools.med.kuleuven.be/typing/v3/hiv/typingtool/>) (De Oliveira et al., 2005; Pineda-Peña et al., 2013), confirmed with manual phylogenetic analysis (Pérez et al., 2006). Because of similar breakpoints in the *pol* region and lack of breakpoints in the *env* region, *pol* and *env* sequences initially assigned to CRF20\_BG, CRF23\_BG or CRF24\_BG were aligned with the *pol* and *env* regions from all full genome sequences of these three CRFs and the full genome subtype reference sequences available in Los Alamos database (accessed August 2013), then a Maximum Likelihood (ML) tree was constructed with RaxML (GTR +  $\Gamma$  model and 1000 bootstrap replicates) (Stamatakis, 2006). All potential CRF20\_BG, CRF23\_BG and CRF24\_BG study sequences clustered mainly outside the cluster of the respective CRF reference sequences precluding a reliable CRF assignment, but together they were monophyletic and were therefore called CRF20–CRF23–CRF24\_BG. Similarly, CRF19\_cpx does not have breakpoints in *pol* (subtype D) and *env* (subtype A), however, CRF19 identification was consistent among subtyping tools, while clustering inside the clade of the reference sequences, consequently CRF19\_cpx assignment was considered reliable.

Transmitted drug resistance was predicted using the 2009 WHO list (Bennett et al., 2009) (<http://cpr.stanford.edu/cpr.cgi>).

Co-receptor use prediction was with `geno2pheno[co-receptor]`: sequences with a false-positive rate (FPR) <5% are mainly CXCR4-using (X4) and  $\geq 20\%$  are mainly CCR5-using (R5) variants. Therefore, we classified V3 loop sequences with FPR  $\geq 20\%$  as R5 viruses, with FPR  $\geq 5\%$  and <20% as dual-tropic viruses and with FPR <5% as X4 viruses. Our motivation for this classification is as follows. Standard methods for co-receptor prediction only classify HIV-1 variants into two categories, namely CCR5 and CXCR4-using viruses (Lengauer et al., 2007). The group of CXCR4-using viruses comprises dual-tropic viruses and X4 viruses. Dual-tropic viruses can use both the CCR5 and CXCR4 co-receptors for cell entry, while X4 viruses can use CXCR4 but not CCR5. The term dual/mixed is used by `geno2pheno[co-receptor]`, referring to either dual tropic viruses or to an unresolved mixture of viruses that can use one or the other co-receptor, `geno2pheno[co-receptor]` cannot discriminate between dual and mixed tropic viruses. If the V3 loop of clinical samples are determined by single population-based Sanger sequences, the European guidelines on the clinical management of HIV-1 tropism testing state that an FPR of `geno2pheno[co-receptor]`  $\geq 20\%$  can be considered as evidence for an R5 virus, while an FPR <20% would indicate a CXCR4-using virus (Vandekerckhove et al., 2011). As in this study we were interested in comparing X4 viruses and R5 viruses, we evaluated whether it is possible to differentiate between X4 and dual-tropic viruses given the `geno2pheno[co-receptor]` FPR. Previously, it has been shown for the PSSM method that dual-tropic viruses mainly had a score between the scores of X4 viruses and R5 viruses (Jensen et al., 2003). We show here that the same holds for the FPR generated by the `geno2pheno[co-receptor]` prediction tool. Supplementary Fig. 2A shows the FPR distribution according to phenotypically confirmed R5 only, X4 only, or dual/mixed strains, for sequences that were published at the LANL HIV sequence database between 2009 and July 2013 (LANL\_recent), while Supplementary Fig. 2B shows all sequences with phenotype information available in July 2013 (LANL\_all). Note that some of the sequences of the “LANL\_all” data set were also used for training `geno2pheno[co-receptor]`. It can be seen that the FPRs of dual/mixed viruses are between the ones from R5 and X4 viruses for both data sets. The 25th percentile for R5 viruses was 20.7% and 17.8% for LANL\_recent and LANL\_all, respectively. This, in combination with the fact that the European guidelines state that the HIV sequences with a `geno2pheno[co-receptor]` FPR  $\geq 20\%$  are considered R5 viruses, motivated us to utilize this threshold also in our study. Since the median FPR for dual/mixed viruses was 3.4% and 5.3% for the two data sets and the median for X4 viruses was 0.3% and 0.5%, respectively, we chose 5% as the FPR threshold for deciding whether a virus is X4 or not. This is also supported by the German and Austrian guidelines for viral tropism

determination stating that patients harboring HIV variants with `geno2pheno[co-receptor]` FPRs below 5% should not be treated with Maraviroc, while patients with an FPR  $\geq 5\%$  might still benefit even if the FPR is not above the R5 threshold (Eberle et al., 2014).

Evolutionary fitness of *pol* sequences (PR and RT regions) was estimated through a quantitative genotypic fitness landscape for PR and RT separately (Theys et al., 2012).

For transmission cluster analysis, *pol* sequences were pooled with a) all other available Cuban sequences at IPK; b) the 30 most similar sequences to each of the study sequences retrieved from NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>); and c) all sequences retrieved from the Los Alamos database (<http://www.hiv.lanl.gov> in December 2013) by using the search term “Cuba” and duplicates were removed. Analysis was done per subtype or CRF separately including also the respective subtype reference sequences (from the Los Alamos database), outgroup for subtype B was subtype D, and for all other subtypes and CRFs it was subtype B. Sequences were aligned with MUSCLE, minimally edited with Mega 5 (Edgar, 2004; Tamura et al., 2011), drug resistance related positions (Bennett et al., 2009) were removed to avoid the effect of convergent evolution (Lemey et al., 2005). An ML phylogenetic tree was constructed using FastTree (Price et al., 2010) on 837 sequences for subtype B, 288 sequences for CRF19\_cpx, and 204 for CRF20–CRF23–CRF24\_BG (average of 918 nucleotides) and transmission clusters were identified with Cluster Picker (Ragonnet-Cronin et al., 2013) using as cut-off genetic distances of 0.03, 0.045 and 0.06 with a bootstrap support of 98% (Frentz et al., 2013; Leigh Brown et al., 2011).

#### 2.4. Statistical Analysis

Parametric and non-parametric tests were used according to Kolmogorov–Smirnov test for normality: Mann Whitney test, Spearman correlation, Pearson correlation, Chi-square test for trend, Chi-square test, One-way ANOVA with Holm–Sidak’s multiple comparison post-test and Kruskal–Wallis test with Dunn’s multiple comparison post-test. These were all two-tailed with a significance level of 5% (GraphPad Prism 5.0 software).

#### 2.5. Data Mining

Bayesian network (BN) learning (using B-course adapted by Deforche et al. (2006)) was used to describe and visualize conditional dependencies between the multiple variables found significant in univariate analysis, discretizing continuous variables into quartiles. Dependencies are represented qualitatively by a directed acyclic graph where each node corresponds to a variable, and a direct arc between nodes represents a direct influence. Robustness of the arcs was scored with a non-parametric bootstrap test (100 $\times$  replicates) (Friedman and WA, 1999) and only arcs with more than 35% support were depicted. Included variables were (Tables 1 and 2): AIDS-RP (AIDS-RP versus chronic-AIDS + non-AIDS), non-AIDS (non-AIDS versus AIDS-RP + chronic-AIDS), RANTES, IP-10, MCP-1,  $\beta 2$ -microglobulin, neopterin, viral load at sampling, heterosexual, anal contact, not always using condom, oral candidiasis, subtype CRF19\_cpx, and not subtype B. Because of their strict association with AIDS definition, collinearity and redundancy, CD4 levels at HIV diagnosis, viral load at HIV diagnosis, CD4 levels at sampling and opportunistic infections were excluded to clarify other dependencies that we were interested in.

### 3. Results

#### 3.1. Descriptive Characteristics of Study Population

We recruited 95 treatment-naïve HIV-1-infected patients, up to a number of 52 in the AIDS-RP group, up to 21 in the non-AIDS group, and up to 22 in the chronic-AIDS group. Study plasma samples and clinical, epidemiological and laboratory data were collected (Tables 1 and 2).

**Table 2**

Clinical, immunological and virological markers at sampling for the three studied groups. Data are expressed as median values with interquartile ranges, or as proportion (%) with number of patients between brackets. Statistical differences between groups were tested using Kruskal–Wallis or Chi-square test for trend. Results were considered significant at p-value <0.05 (displayed in bold). N = number of patients. TDR (transmitted drug resistance): NRTI resistance mutations M41L, F116Y/T215S/K219Q, M41L/T215D and D67N/T215S/K219Q; NNRTI resistance mutations twice K103N and PI resistance mutations N88DN and M46L.

\*MIP-1 $\beta$  and IL-8 statistical graphs are depicted in Supplementary Fig. 6A and B, respectively.

	Non-AIDS (N = 21)	Chronic-AIDS (N = 22)	AIDS-RP (N = 52)	p value
Years elapsed from HIV seroconversion to sampling	1.75 (1.25–2.33) (21)	9.82 (7.97–12.06) (22)	1.40 (0.79–2.16) (52)	–
Log VL at sampling (RNA copies/ml plasma)	3.7 (3.0–4.3) (N = 20)	4.7 (4.2–5.4) (N = 22)	4.7 (4.0–5.4) (N = 47)	<b>0.0012</b>
CD4 count at sampling (cells per $\mu$ l blood)	501 (408–641) (N = 17)	212 (113–266) (N = 17)	189 (91–225) (N = 51)	<b>&lt;0.0001</b>
CD4% at sampling	23 (18–32) (N = 17)	12 (9–13) (N = 17)	11 (8–13) (N = 51)	<b>&lt;0.0001</b>
RANTES/CCL5 (pg/ml)	944 (719–1184) (N = 18)	1291 (932–1430) (N = 21)	1398 (1103–1630) (N = 47)	<b>0.00060</b>
MCP-1 (pg/ml)	0.0 (0.0–128.3) (N = 18)	127.1 (0.0–218.9) (N = 21)	122.70 (0.0–289.0) (N = 47)	<b>0.046</b>
MIP-1 $\beta$ (pg/ml)	0.0 (0.0–0.0) (N = 18)	0.0 (0.0–0.0) (N = 21)	0.0 (0.0–0.0) (N = 47)	*
IP-10/CXCL10 (pg/ml)	0.0 (0.0–129.5) (N = 18)	0.0 (0.0–97.2) (N = 21)	79.43 (0.0–229.3) (N = 47)	<b>0.047</b>
MIG/CXCL9 (pg/ml)	0.0 (0.0–138.8) (N = 18)	139.1 (0.0–212.6) (N = 21)	125.8 (0.0–209.9) (N = 47)	0.15
IL-8 (pg/ml)	0.0 (0.0–0.0) (N = 18)	0.0 (0.0–0.0) (N = 21)	0.0 (0.0–0.0) (N = 47)	*
FAS/APO (pg/ml)	750 (600–1197) (N = 18)	1094 (795–1154) (N = 21)	1030 (886–1381) (N = 48)	0.11
Neopterin (ng/ml)	1.0 (0.7–2.6) (N = 18)	2.1 (1.5–4.1) (N = 21)	2.7 (1.4–7.4) (N = 48)	<b>0.020</b>
$\beta$ 2-microglobulin (mg/ml)	1.1 (0.8–1.7) (N = 18)	1.91 (1.47–2.52) (N = 21)	1.92 (1.2–3.6) (N = 48)	<b>0.0012</b>
Hemoglobin (g/l)	131.5 (123.0–140.0) (N = 18)	125.0 (120.0–139.0) (N = 21)	130.0 (112.8–140.0) (N = 46)	0.95
Co-infections (proportion)				
Oral candidiasis	4.8% (1/21)	0.0% (0/22)	23.1% (12/52)	<b>0.013</b>
Gonorrhoea	28.6% (6/21)	9.1% (2/22)	17.3% (9/52)	0.40
HCV(1)/HBV(5)	9.5% (2/21)	9.1% (2/22)	3.9% (2/52)	0.96
Herpes simplex	23.8% (5/21)	45.5% (10/22)	15.4% (8/52)	0.18
Condyloma	19.1% (4/21)	13.6% (3/22)	7.7% (4/52)	0.16
Herpes zoster	4.8% (1/21)	22.7% (5/22)	13.5% (7/52)	0.52
Syphilis	4.8% (1/21)	18.2% (4/22)	7.7% (4/52)	0.98
Other co-morbidities (chronic non-infectious diseases)				
Allergy	33.3% (7/21)	27.3% (6/22)	23.1% (12/52)	0.42
Hypertension	4.8% (1/21)	9.1% (2/22)	11.5% (6/52)	0.35
HIV subtypes (proportion)				
Subtype B	42.1% (8/19)	42.9% (9/21)	17.3% (9/52)	<b>0.016</b>
Subtype C	0.0% (0/19)	4.8% (1/21)	9.6% (5/52)	0.13
CRF18_cpx	10.5% (2/19)	9.5% (2/21)	3.8% (2/52)	0.28
CRF19_cpx	0.0% (0/19)	0.0% (0/21)	17.3% (9/52)	<b>0.012</b>
CRF20,23,24_BG	36.8% (7/19)	23.8% (5/21)	32.7% (17/52)	0.91
TDR	10.5% (2/19)	4.8% (1/21)	9.6% (5/52)	0.95
Prediction of co-receptor use				
X4 (FPR < 5%)	0.0% (0/13)	18.8% (3/16)	16.2% (6/37)	0.21
R5X4 (5% $\leq$ FPR < 20%)	30.8% (3/13)	31.3% (5/16)	37.8% (14/37)	0.59
R5 (FPR $\geq$ 20%)	70.6% (10/13)	50.0% (8/16)	45.95% (17/37)	0.07

The patients were predominantly white young males from Havana (median 33.0 years old at HIV diagnosis, 80% male) with no significant difference in age, gender and skin color distribution among groups (Table 1).

For all groups, the median time between last negative and first positive HIV test was similar: 16.0 months (IQR = 9.3–24.0 months), its midpoint being the estimated seroconversion date. Median time between seroconversion and AIDS was 1.4 years in the AIDS-RP group and 9.82 years in the chronic-AIDS group (Supplementary Fig. 1B). Sampling was a median of 1.75 years after seroconversion for non-AIDS and AIDS-RP groups (Table 2). At AIDS diagnosis, 16 patients were identified with opportunistic infections, 13 in the AIDS-RP group (25%) and 3 in the chronic-AIDS group (14%). The most common opportunistic infection was *Pneumocystis jirovecii* (Table 1).

For 92 patients, at least partial *pol* and for 70 patients, both partial *pol* and partial *env* regions sequences were available. Twenty six patients were infected with subtype B, 6 with subtype C, 6 with CRF18\_cpx, 9 with CRF19\_cpx and 29 with CRF20–23–24\_BG (Fig. 1A). The others were infected with various other subtypes and URFs (Supplementary Fig. 3).

### 3.2. Variables Associated With AIDS-RP

In univariate analysis with multiple comparison posttest correction, several significant differences were observed (Tables 1–3 and Supplementary Fig. 4A–I). As expected, at HIV diagnosis AIDS-RP patients had a significantly higher median log viral load (4.8 log c/ml)

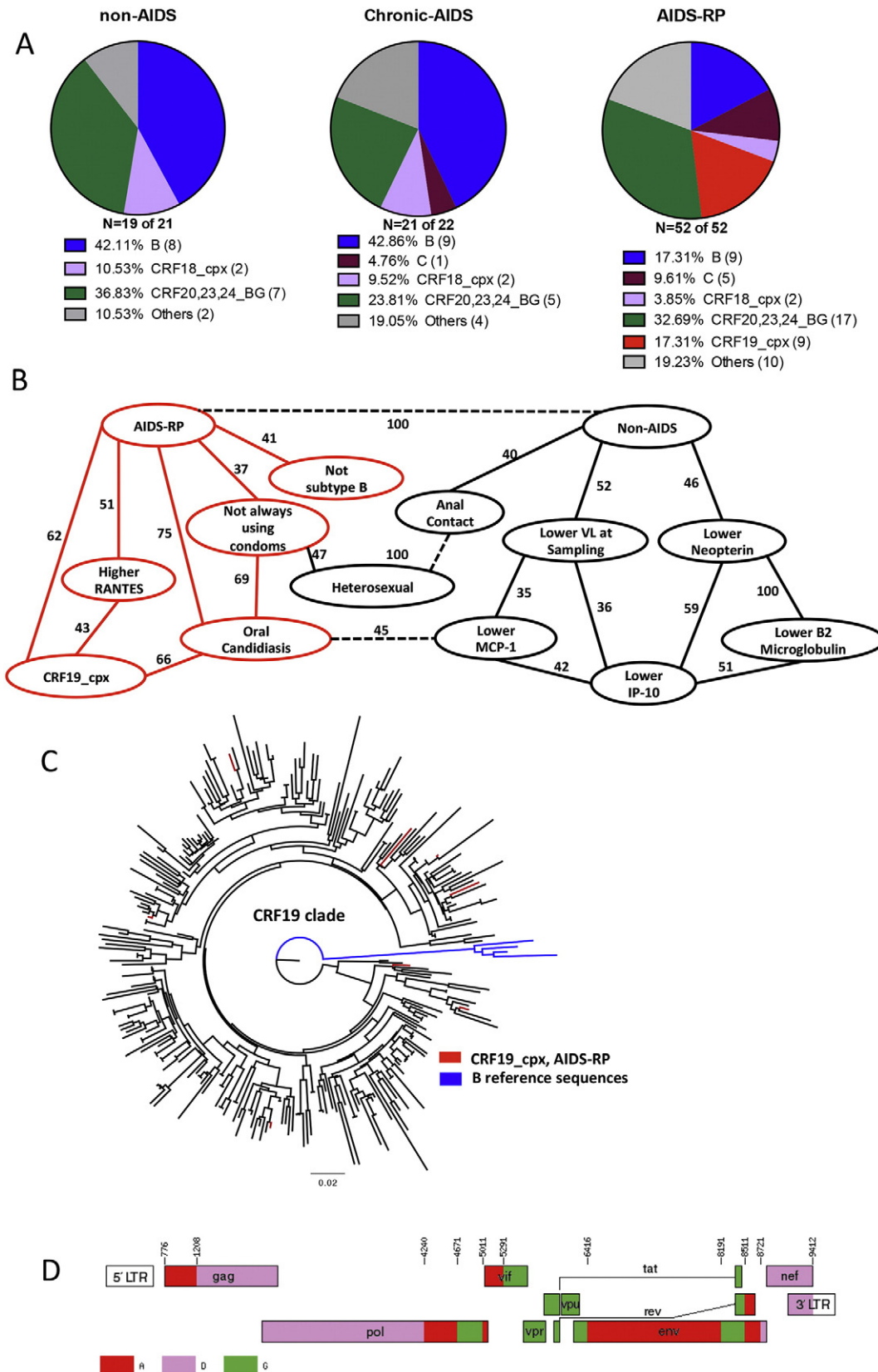
(non-AIDS <1.7 log c/ml, chronic-AIDS 3.2 log c/ml); and significantly lower CD4 levels (276 cells/mm<sup>3</sup>) (non-AIDS 577 cells/ $\mu$ l, chronic-AIDS 522 cells/ $\mu$ l) (Tables 1 & 3). At sampling, levels of RANTES/CCL5, MCP-1,  $\beta$ 2-microglobulin, neopterin, and viral load were significantly higher and CD4 count significantly lower in AIDS-RP group than in the non-AIDS group (Tables 2 & 3 and Supplementary Fig. 4A–I); while AIDS-RP was not associated with a different proportion of transmitted drug resistance (Table 2) or a higher proportion of X4-predicted viruses, irrespective of the cut-off used (<5% or  $\leq$ 20% FPR) (Table 2).

However, we found some surprising associations. Compared to the combined non-AIDS and chronic-AIDS group, AIDS-RP were significantly more often heterosexual (49%), less prone to protect themselves from superinfection or STIs (only 19.6% were always using condoms after diagnosis), presented more often with oral candidiasis (23.1%), were less frequently infected with subtype B (17.3%) and more frequently with CRF19\_cpx (17.3%) (Fig. 1A, Tables 1–3). Strikingly, all 9 CRF19\_cpx infected patients belonged to the AIDS-RP group (Fig. 1A and Table 2). Another unusual association was non-AIDS with anal sex. While we cannot exclude underreporting of injecting drug use in our cohort, according to the Cuban epidemiological information center, there are no reports of IDU infected with HIV in the last years in Cuba.

Because of a complete overlap between some variables (e.g., all CRF19\_cpx were AIDS-RP), the assumptions of the classical regression model were violated by multicollinearity, prohibiting us to calculate odds ratios using logistic regression. We therefore used a

multivariate Bayesian Network analysis (BN) and compiled all patient information significantly associated with disease progression groups in the univariate analysis. The AIDS-RP and non-AIDS groups were displayed explicitly in the graph (Fig. 1B), by coding the groups' binary (only two of the three groups are displayed, such that

chronic-AIDS is coded implicitly as not AIDS-RP and not non-AIDS). This display allowed us to capture better the difference between AIDS-RP and the other two groups, however we sacrifice clarity for associations with chronic-AIDS. As expected, the BN approach revealed that a lower viral load at sampling was associated with non-



**Table 3**

Multiple comparison posttest for the attributes significantly associated with disease progression. Data are expressed as median values with interquartile ranges, or as proportion (%) with number of patients between brackets. Statistical differences between AIDS-RP vs. non-AIDS and AIDS-RP vs. chronic-AIDS groups were tested using Dunn's multiple comparison test or Chi-square test. Results were considered significant at p-value < 0.05 (displayed in bold). N = number of patients.

	Non-AIDS (N = 21)	Chronic-AIDS (N = 22)	AIDS-RP (N = 52)	Multiple comparison posttest AIDS-RP vs. non-AIDS	Multiple comparison posttest AIDS-RP vs. chronic-AIDS	Multiple comparison posttest Chronic-AIDS vs. non-AIDS
Log VL at HIV diagnosis (RNA copies/ml plasma)	<1.7 (<1.7–3.5) (N = 12)	3.2 (<1.7–3.9) (N = 8)	4.8 (4.1–5.4) (N = 41)	<b>&lt;0.0001</b>	<b>0.0054</b>	>0.99
CD4% at HIV diagnosis (%)	27 (19–35) (N = 20)	24 (19–30) (N = 21)	13 (10–19) (N = 49)	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	>0.99
CD4 count at HIV diagnosis (cells per µl blood)	577 (441–815) (N = 20)	522 (374–782) (N = 22)	276 (152–386) (N = 49)	<b>&lt;0.0001</b>	<b>&lt;0.0001</b>	>0.99
Log VL at sampling (RNA copies/ml plasma)	3.7 (3.0–4.3) (N = 20)	4.7 (4.2–5.4) (N = 22)	4.7 (4.0–5.4) (N = 47)	<b>0.0028</b>	>0.99	<b>0.0033</b>
CD4 count at sampling (cells per µl blood)	501 (408–641) (N = 17)	212 (113–266) (N = 17)	189 (91–225) (N = 51)	<b>&lt;0.0001</b>	>0.99	<b>0.0001</b>
CD4% at sampling (%)	23 (18–32) (N = 17)	12 (9–13) (N = 17)	11 (8–13) (N = 51)	<b>&lt;0.0001</b>	>0.99	<b>&lt;0.0001</b>
RANTES/CCL5 (pg/ml)	944 (719–1184) (N = 18)	1291 (932–1430) (N = 21)	1398 (1103–1630) (N = 47)	<b>0.0004</b>	0.41	0.11
MCP-1 (pg/ml)	0.0 (0.0–128.3) (N = 18)	127.1 (0.0–218.9) (N = 21)	122.70 (0.0–289.0) (N = 47)	<b>0.047</b>	0.66	0.83
IP-10/CXCL10 (pg/ml)	0.0 (0.0–129.5) (N = 18)	0.0 (0.0–97.2) (N = 21)	79.43 (0.0–229.3) (N = 47)	0.22	0.093	>0.99
Neopterin (ng/ml)	1.0 (0.7–2.6) (N = 18)	2.1 (1.5–4.1) (N = 21)	2.7 (1.4–7.4) (N = 48)	<b>0.016</b>	0.94	0.34
β2-microglobulin (mg/ml)	1.1 (0.8–1.7) (N = 18)	1.91 (1.47–2.52) (N = 21)	1.92 (1.2–3.6) (N = 48)	<b>0.0016</b>	>0.99	<b>0.0063</b>
Anal sex (as reported by both males and females of various sexual orientation)	80.0% (16/20)	54.6% (12/22)	50.0% (25/50)	<b>0.021</b>	0.72	0.081
Always condom use after diagnosis	45.0% (9/20)	45.5% (10/22)	19.6% (10/51)	<b>0.030</b>	<b>0.023</b>	0.98
Sexual orientation						
Heterosexual	20.0% (4/20)	36.4% (8/22)	49.0% (25/51)	<b>0.025</b>	0.32	0.24
Co-infections (proportion)						
Oral candidiasis	4.8% (1/21)	0.0% (0/22)	23.1% (12/52)	0.064	<b>0.014</b>	0.30
HIV subtypes (proportion)						
Subtype B	38.1% (8/21)	40.9% (9/22)	17.3% (9/52)	0.057	<b>0.031</b>	0.85
CRF19_cpx	0.0% (0/21)	0.0% (0/22)	17.3% (9/52)	<b>0.042</b>	<b>0.037</b>	–

AIDS, whereas infection with CRF19\_cpx, oral candidiasis, increased expression of RANTES, “not always using condoms” and not infected with subtype B were associated with AIDS-RP (red in Fig. 1B). The most robust association was with oral candidiasis (75% of bootstraps) and with CRF19\_cpx (62% of bootstraps). Moreover, the BN approach demonstrated that CRF19\_cpx, RANTES and oral candidiasis were also robustly interconnected (Fig. 1B).

### 3.3. Rapid Progression to AIDS is Associated With CRF19\_cpx Subtype

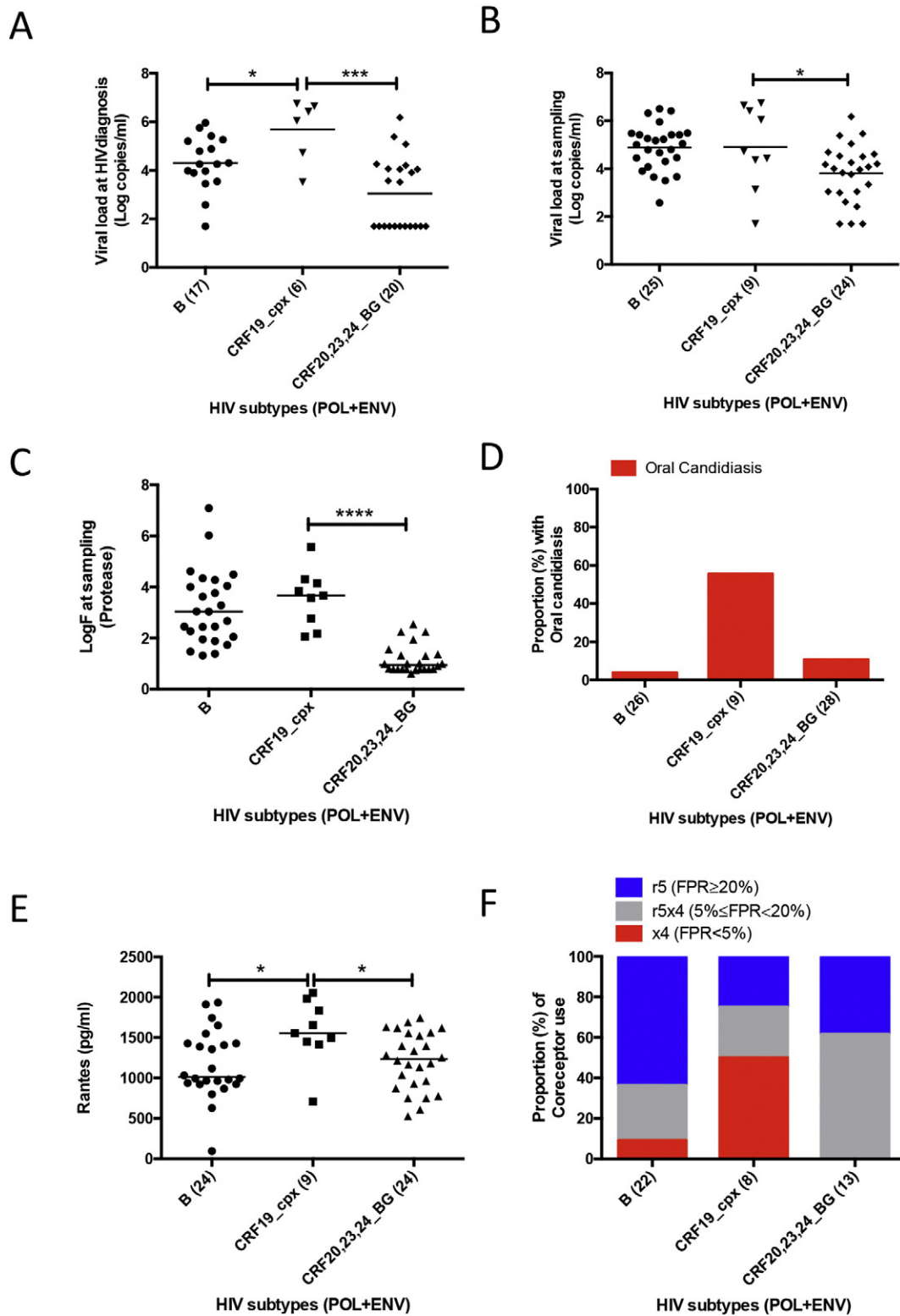
One of the surprising findings from the BN analysis was the robust association of CRF19\_cpx with rapid progression. Our AIDS-RP patients were infected with divergent CRF19\_cpx strains and had thus not transmitted a particular virulent variant among each other: none of them clustered in a transmission cluster (Fig. 1C), while we did find some

transmission clusters among the other clades (Supplementary Fig. 3). There was not sufficient information for the other CRF19\_cpx patients included in the transmission analysis to judge whether or not they were fast progressors. CRF19\_cpx is a complex recombinant HIV-1 virus consisting of subtype D sequence fragments (C-part of Gag, PR, RT and nef), subtype A sequence fragments (N-part of Gag, Integrase, Env) and subtype G fragments (Vif, Vpr, Vpu and C-terminal part of Env) (Casado et al., 2005) (Fig. 1D).

### 3.4. CRF19\_cpx-infected Patients Have High Viral Replication, High Immune Activation and Frequent Co-infection

For all three groups, compared to subtype B and CRF20-23-24\_BG as the only other clades with a sufficient number of sequences, CRF19 infected patients had significantly higher viral load at HIV diagnosis,

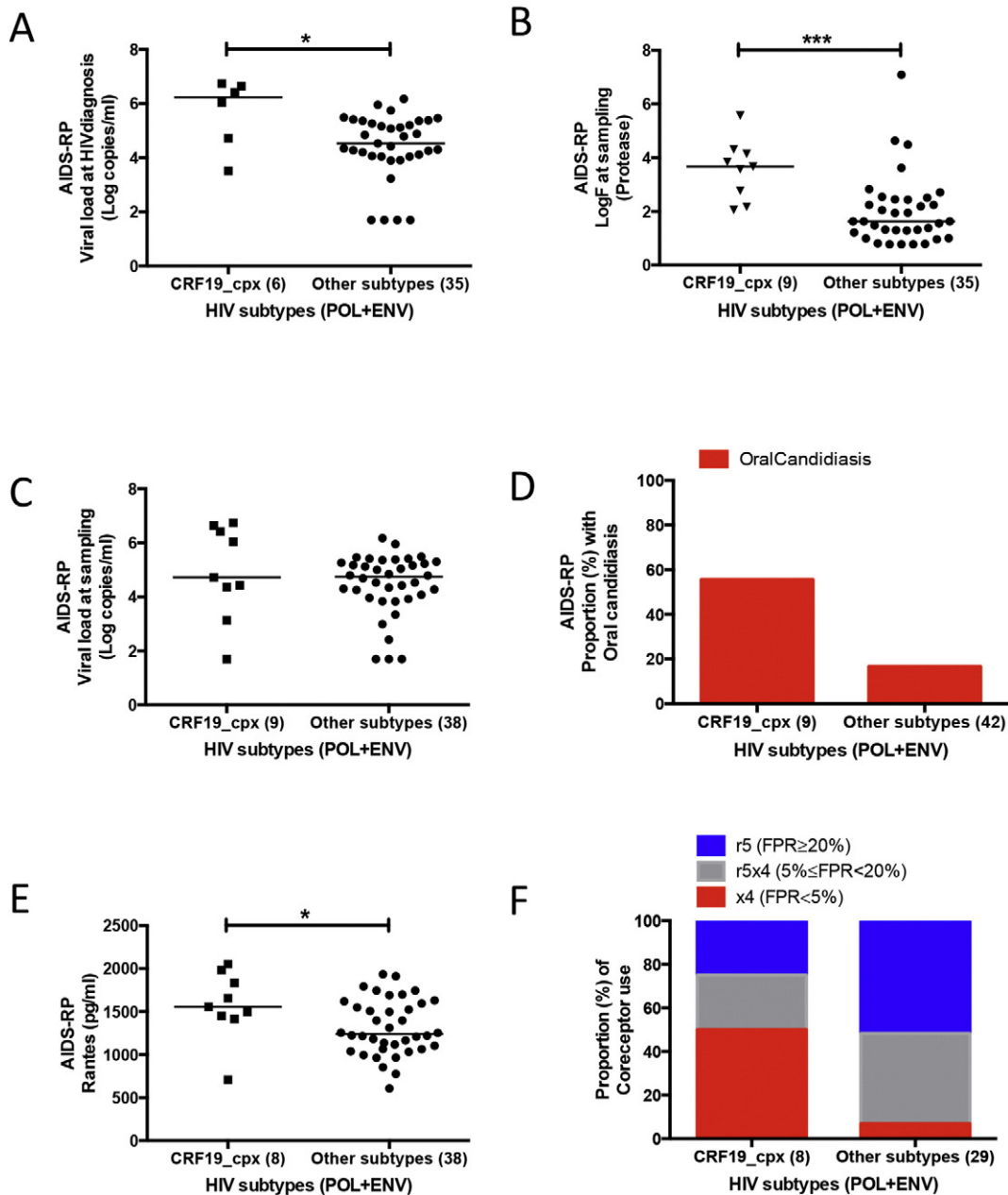
**Fig. 1.** CRF19\_cpx is associated with rapid progression to AIDS. A) HIV subtype distribution among the three studied groups: non-AIDS, chronic-AIDS and AIDS-RP, classified according to the disease progression rates (see Patients and methods). B) Variables found to be significantly associated with AIDS-RP (AIDS-RP versus non-AIDS + chronic-AIDS) or non-AIDS (non-AIDS versus chronic-AIDS + AIDS-RP) patients in univariate analysis were included in the BN model. A central role for CRF19\_cpx in rapid progression to AIDS was demonstrated, showing direct influences (arcs) between AIDS-RP, CRF19\_cpx, oral candidiasis, higher RANTES, not always using condoms and not subtype B (red contours and arcs). The stability of the dependency was assessed with a non-parametric bootstrap (100 × replicates). All arcs with bootstrap over 35% are depicted in the network. All arcs with dashed lines represent antagonistic dependencies. C) The ML phylogenetic tree for subtype CRF19\_cpx (pol region) was constructed using FastTree. CRF19\_cpx-infected AIDS-RP (red), CRF19\_cpx reference sequences and all other CRF19\_cpx sequences retrieved from Los Alamos database (black), all other available CRF19\_cpx sequences from IPK (black), and subtype B Los Alamos reference sequences (blue). One of the nine CRF19\_cpx sequences was incomplete for pol region and not included in the ML phylogenetic tree analysis. D) Mosaic structure of the CRF19\_cpx genome as taken from the Los Alamos HIV Sequence Database (<http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/breakpoints.html#CRF19>).



**Fig. 2.** Comparison of viral load, RANTES, estimated evolutionary fitness, oral candidiasis and co-receptor use prediction among the most frequent HIV subtypes in our study (non-AIDS, chronic-AIDS and AIDS-RP groups). A) Viral load (log copies/ml) at HIV diagnosis (One-way ANOVA,  $p = 0.0002$ , with Holm–Sidak’s multiple comparison test,  $^*p < 0.05$  and  $^{***}p < 0.001$ ). B) Viral load (log copies/ml) at sampling (One-way ANOVA,  $p = 0.0043$ , with Holm–Sidak’s multiple comparison test,  $^*p < 0.05$ ). C) Evolutionary fitness estimated from protease sequence (log of fitness value, logF) at sampling (Kruskal–Wallis test,  $p < 0.0001$ , with Dunn’s multiple comparison test,  $^{****}p < 0.0001$ ). D) Proportion with oral candidiasis at sampling (Chi-square test,  $p = 0.0005$ ). E) RANTES levels (pg/ml) at sampling (Kruskal–Wallis test,  $p = 0.038$ , with Dunn’s multiple comparison test,  $^*p < 0.05$ ). F) Proportion of co-receptor use at sampling (Chi-square test,  $p = 0.0051$ ).

whereas at sampling the difference was only significant compared to CRF20–23–24\_BG (Fig. 2A and B). CRF19 *pol* sequences at sampling had a significantly higher estimated in vivo fitness score for protease

(logF = 3.67) as compared to CRF20–23–24\_BG (logF = 0.95) (Fig. 2C), but not for reverse transcriptase (data not shown), and this score was positively associated with log viral load at HIV diagnosis



**Fig. 3.** Comparison of viral load, RANTES, estimated evolutionary fitness, oral candidiasis and co-receptor use between CRF19\_cpx and all other HIV subtypes among AIDS-RP. A) Viral load (log copies/ml) at HIV diagnosis (Mann Whitney test,  $*p < 0.05$ ). B) Evolutionary fitness estimated from protease sequence (log of fitness landscape, logF) at sampling (Mann Whitney test,  $***p < 0.001$ ). C) Viral load (log copies/ml) at sampling (Mann Whitney test,  $p = 0.33$ ). D) Proportion with oral candidiasis at sampling (Chi-square test,  $p = 0.013$ ). E) RANTES levels (pg/ml) at sampling (Mann Whitney test,  $*p < 0.05$ ). F) Proportion of co-receptor use at sampling (Chi-square test,  $p = 0.014$ ).

and at sampling (Supplementary Fig. 5A and B). CRF19 infected patients had more frequently oral candidiasis (Fig. 2D) and higher RANTES levels than subtype B or CRF20–23–24\_BG infected patients (Fig. 2E). Consistent with RANTES being a CCR5 ligand/competitor, CRF19 strains were more often X4 viruses ( $< 5\%$  FPR cut-off) (Fig. 2F).

Strikingly, when considering only the AIDS-RP group and compared to the other clades, CRF19\_cpx infected patients had a significantly higher viral load at HIV diagnosis and in vivo estimated fitness at sampling ( $\log VL = 6.23$ , and  $\log F = 3.67$ , Fig. 3A and B), although viral load at sampling was not significantly different (Fig. 3C). In addition, CRF19\_cpx infected patients had more often oral candidiasis and had higher RANTES levels (Fig. 3D and E), and CRF19 strains were also

more often predicted as X4 (Fig. 3F). Together, these results suggest a high evolutionary fitness of CRF19\_cpx.

#### 4. Discussion

It has been speculated that inter-subtype recombination may give rise to more pathogenic strains if genomic fragments from different subtypes join together in a better replicating virus, but no direct evidence for this scenario has so far been found. CRF19\_cpx is a genetic (Casado et al., 2005) form hitherto only reported in Cuba, but with evidenced central African ancestry (Thomson and Nájera, 2005; Casado et al., 2005; Delatorre and Bello, 2013). Recently, clinicians in Cuba



noted an increasing trend of rapid progression to AIDS (<3 years since seroconversion) (Abrahantes et al., 2010). Here, we propose that this increase is at least in part due to the spread of CRF19\_cpx, which we found exclusively associated with rapid progression to AIDS (within 3 years of infection), whereas for other subtypes and CRFs, the disease progression was distributed as is generally seen in other parts of the world.

We set up an exploratory study in Cuba to investigate the association of rapid progression with epidemiological, clinical, viral and immunological parameters, comparing three groups of patients. The first two groups were recruited prospectively who at sampling within 3 years after seroconversion either were already diagnosed with AIDS (AIDS-RP) or who were still AIDS-free (non-AIDS) and chronically infected HIV-1 patients with AIDS recruited retrospectively and sampled at AIDS diagnosis (chronic-AIDS). Bayesian network analysis showed that rapid progression to AIDS in Cuba, being infected with CRF19\_cpx, having oral candidiasis, and having higher levels of RANTES are strongly linked. CRF19\_cpx was exclusively found among rapid progressors. Compared to other patients in Cuba, CRF19\_cpx infected patients have a higher viral load at comparable times since infection, higher levels of RANTES and are more frequently co-infected with oral candidiasis. Additionally, the virus at sampling had higher estimated protease fitness and was predominantly using the CXCR4 co-receptor, potentially explaining the rapid disease progression. Surprisingly however, even when restricting the analysis to the 52 rapid progressors only, CRF19 infection was significantly associated with oral candidiasis, CXCR4 co-receptor use prediction, higher RANTES levels, higher protease fitness at sampling and higher viral load at HIV diagnosis, whereas viral load at sampling was not. Together the higher viral load and higher estimated in vivo fitness of the viral protease suggest that CRF19 is an evolutionary fit virus. However, confirmation by a phenotypic assay is still warranted.

CRF19\_cpx is composed of subtype D fragments (C-part of Gag, PR, RT and nef), subtype A fragments (N-part of Gag, Integrase, Env) and subtype G fragments (Vif, Vpr, Vpu and C-terminal part of Env) (Casado et al., 2005). Several studies have reported the association of HIV subtype D with faster disease progression and subtype A with slower disease progression (Baeten et al., 2007; Kaleebu et al., 2002; Kiwanuka et al., 2008). The probability of having an CXCR4-using virus is higher in subtype D than in subtype A infections in non-AIDS clinical status (Kaleebu et al., 2007), suggesting an earlier switch to CXCR4 since dual tropism is not common in subtype D (Tscherning et al., 1998). These findings cannot explain our observations since in CRF19\_cpx, the *env* region responsible for CXCR4 tropism is derived from subtype A. The high viral load, the higher production of RANTES (known to block CCR5 receptors) perhaps resulting from the high viral load (Annunziato et al., 2000) and/or the co-infections (Huang and Levitz, 2000), might be the reason for the high prevalence of CXCR4 use in this CRF19\_cpx, and this may be the cause of the rapid disease progression. This high viral load may be related to and potentially caused by the high protease fitness we observed in CRF19\_cpx, and thus could suggest a more important role of *pol* (subtype D portion), and especially of protease, in disease progression, as compared to the *env* region (subtype A portion) of CRF19\_cpx.

Nevertheless, other subtype D genes in CRF19\_cpx (*gag* and *nef*) not investigated in this study might also be associated with increased viral replication and in vivo pathogenesis. In HIV-1, Gag and Nef have been documented to contain the highest CD8 epitope density and are more frequently recognized by the immune system (Frahm et al., 2004). Notably, amino acid substitution T242N in the Gag p24 TW10 epitope is highly associated with viral escape from the immune system (Leslie et al., 2004). Subtype D-infected patients more often, and subtype A infected patient less often display this T242N (Li et al., 2013). However, T242N in subtype D reduces the viral replication and fitness of the virus (Martinez-Picado et al., 2006), while CRF19-infected patients present a very high viral load. Another important protein is Nef, which is responsible for the down-regulation of CD4 and HLA expression and

up-regulation of immature HLA class II expression (Wang, 2013). Furthermore, infection by HIV-1 with a truncated Nef contributes to low-level viral replication and non-pathogenicity (Deacon et al., 1995). However, subtype B Nef was found to be a better HLA class I down-regulator in normal progressors than subtype D Nef. Yet, subtypes D and A were similar in Nef-mediated CD4 and HLA class I down-regulation (Mann et al., 2013).

Our study has several limitations. Even though the statistical link between fast progression and being infected with CRF19\_cpx is very strong, we only presented data on 9 such patients. An in depth study on progression in patients infected with CRF19\_cpx is needed. We used two different genomic regions and a stringent method of subtyping to increase our confidence in subtyping, however, since we did not have access to the full genome, we cannot exclude recombination in non-sequenced regions. We did not have sufficiently representative data to investigate the trend of CRF19\_cpx incidence over the time period that the increase in rapid progression was noted. However, CRF19\_cpx prevalence at HIV diagnosis at IPK, Cuba, increased over the investigated time period (2007–2013) (data not shown). We were also not able to investigate the host genotype, however we did confirm that there was no significant difference in ethnic composition between the three groups. Future studies should evaluate HLA and others host genetic markers (CCR5) in new CRF19\_cpx-infected patients. It would also be interesting to test humoral and cellular immune responses in new blood samples and to sequence more full-length CRF19\_cpx genomes.

We propose that an evolutionary very fit CRF19\_cpx together with co-infections is linked to the increase of rapid progression to AIDS in newly infected patients in Cuba. The robust and significant associations with a fitter protease, more circulating virus, higher immune-activation and CXCR4 co-receptor use suggest that CRF19\_cpx may be a more pathogenic virus.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.ebiom.2015.01.015>.

## Authors' Contributions

Conceived and designed the experiments: Vivian Kouri, Ricardo Khouri, Michel Moutschen, Johan Van Weyenbergh, Nico Pfeifer, Kristel Van Laethem and Anne-Mieke Vandamme. Performed the experiments: Ricardo Khouri, Yoan Alemán and Yeissel Abrahantes. Analyzed the data: Ricardo Khouri, Andrea-Clemencia Pineda-Peña, Jurgen Vercauteren, Kristof Theys, Nico Pfeifer and Sarah Mergens. Wrote the paper: Ricardo Khouri, Andrea-Clemencia Pineda-Peña, Kristel Van Laethem and Anne-Mieke Vandamme. Contributed data/reagents/materials/analysis tools: Vivian Kouri, Ana B. Pérez, Jorge Pérez, Lissette Pérez and Nico Pfeifer.

## Funding

This work was supported by grants from the Vlaamse Interuniversitaire Raad (VLIR) (ZEIN2008PR358), a collaborative fund between Wallonia and Cuba (CF/JF/PD/Cu-2009/28422), the Global Fund to Fight AIDS, Tuberculosis and Malaria, the University of Leuven (PF/10/018) and the Fonds voor Wetenschappelijk Onderzoek Vlaanderen (1.5.236.11N and G.0692.14). The computational resources and services used in this work were provided by the Hercules Foundation and the Flemish Government – department EWI-FWO Krediet aan Navorsers (Theys, KAN2012 1.5.249.12); AMV was supported by "Programa Ciências sem fronteiras" Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) of Brazil. RK was supported by an IRO scholarship-KU Leuven and KT was supported by the Research Foundation – Flanders (FWO). We declare that funding sources had no role in the study design, data collection, data interpretation or writing of the report.

## Conflict of Interest Disclosure

The Max Planck Institute for Informatics collaborates with a company in the area of medical diagnostics on the topic of viral tropism of HIV. In its context a version controlled implementation of geno2pheno[co-receptor] with secure access is maintained by Max Planck Institute for Informatics under a paid service agreement.

## Acknowledgments

We would like to thank Yoeri Schrooten for his technical assistance.

## References

- Abrahantes Rodríguez, Y., Pérez Avila, J., Kouri Cardellá, V., Fonseca Gómez, C., Baly Gil, A., Tápanes Fernández, T., 2010. Factors linked to rapid progression to AIDS in Cuban subjects. *Rev. Cubana Med. Trop.* 62, 146–153.
- Alemu, A., Shiferaw, Y., Addis, Z., Mathewos, B., Birhan, W., 2013. Effect of malaria on HIV/AIDS transmission and progression. *Parasites Vectors* 6, 18.
- Annunziato, F., Galli, G., Nappi, F., et al., 2000. Limited expression of R5-tropic HIV-1 in CCR5-positive type 1-polarized T cells explained by their ability to produce RANTES, MIP-1alpha, and MIP-1beta. *Blood* 95, 1167–1174.
- Eberle, J., Noah, C., Wolf, E., Stürmer, M., Braun, P., Korn, K., Däumer, M., Thielen, A., Berg, T., Obermeier, M., Ehret, R., Walther, H., Lengauer, T., Verheyen, J., Sierra, S., Kaiser, R., 2014. Empfehlungen zur Bestimmung des HIV-1-Korezeptor-Gebrauchs, (DAIG Recommendations for the HIV-1 Tropism Testing). [http://www.daignet.de/site-content/hiv-therapie/leitlinien-1/Empfehlungen%20zur%20Bestimmung%20des%20HIV-1\\_Korezeptor\\_Gebrauchs.pdf](http://www.daignet.de/site-content/hiv-therapie/leitlinien-1/Empfehlungen%20zur%20Bestimmung%20des%20HIV-1_Korezeptor_Gebrauchs.pdf) (last accessed 2/3/2015).
- Baeten, J.M., Chohan, B., Lavreys, L., et al., 2007. HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. *J. Infect. Dis.* 195, 1177–1180.
- Bártolo, I., Camacho, R., Barroso, H., Bezerra, V., Taveira, N., 2009. Rapid clinical progression to AIDS and death in a persistently seronegative HIV-1 infected heterosexual young man. *AIDS* 23, 2359–2362.
- Bennett, D.E., Camacho, R.J., Otelea, D., et al., 2009. Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. *PLoS ONE* 4, e4724.
- Casado, G., Thomson, M.M., Sierra, M., Nájera, R., 2005. Identification of a novel HIV-1 circulating ADG intersubtype recombinant form (CRF19\_cpx) in Cuba. *J. Acquir. Immune Defic. Syndr.* 40, 532–537.
- Casado, C., Colombo, S., Rauch, A., et al., 2010. Host and viral genetic correlates of clinical definitions of HIV-1 disease progression. *PLoS ONE* 5, e11079.
- De Oliveira, T., Deforche, K., Cassol, S., et al., 2005. An automated genotyping system for analysis of HIV-1 and other microbial sequences. *Bioinformatics* 21, 3797–3800.
- Deacon, N.J., Tsykin, A., Solomon, A., et al., 1995. Genomic structure of an attenuated quasi species of HIV-1 from a blood transfusion donor and recipients. *Science* 270, 988–991.
- Deforche, K., Silander, T., Camacho, R., et al., 2006. Analysis of HIV-1 pol sequences using Bayesian Networks: implications for drug resistance. *Bioinformatics* 22, 2975–2979.
- Delatorre, E., Bello, G., 2013. Phylodynamics of the HIV-1 epidemic in Cuba. *PLoS ONE* 8, e72448.
- Demarest, J.F., Jack, N., Cleghorn, F.R., et al., 2001. Immunologic and virologic analyses of an acutely HIV type 1-infected patient with extremely rapid disease progression. *AIDS Res. Hum. Retroviruses* 17, 1333–1344.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- Esbjörnsson, J., Månsson, F., Martínez-arias, W., et al., 2010. Frequent CXCR4 Tropism of HIV-1 Subtype A and CRF02\_AG During Late-stage Disease – Indication of an Evolving Epidemic in West Africa. pp. 1–13.
- Frahm, N., Korber, B.T., Adams, C.M., et al., 2004. Consistent cytotoxic-T-lymphocyte targeting of immunodominant regions in human immunodeficiency virus across multiple ethnicities. *J. Virol.* 78, 2187–2200.
- Frentz, D., Wensing, A.M.J., Albert, J., et al., 2013. Limited cross-border infections in patients newly diagnosed with HIV in Europe. *Retrovirology* 10, 36.
- Friedman, N.G.M., WA, 1999. Data analysis with bayesian networks: a bootstrap approach. *UAI'99 Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pp. 196–205.
- Giorgi, J.V., Hultin, L.E., McKeating, J.A., et al., 1999. Shorter survival in advanced human immunodeficiency virus type 1 infection is more closely associated with T lymphocyte activation than with plasma virus burden or virus chemokine co-receptor usage. *J. Infect. Dis.* 179, 859–870.
- Helleberg, M., Afzal, S., Kronborg, G., et al., 2013. Mortality attributable to smoking among HIV-1-infected individuals: a nationwide, population-based cohort study. *Clin. Infect. Dis.* 56, 727–734.
- Hernandez, M.D., Sherman, K.E., 2011. HIV/hepatitis C coinfection natural history and disease progression. *Curr. Opin. HIV AIDS* 6, 478–482.
- Huang, C., Levitz, S.M., 2000. Stimulation of macrophage inflammatory protein-1alpha, macrophage inflammatory protein-1beta, and RANTES by *Candida albicans* and *Cryptococcus neoformans* in peripheral blood mononuclear cells from persons with and without human immunodeficiency virus infec. *J. Infect. Dis.* 181, 791–794.
- Jensen, M.A., Li, F.-S., van 't Wout, A.B., et al., 2003. Improved co-receptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. *J. Virol.* 77, 13376–13388.
- Kaleebu, P., French, N., Mahe, C., et al., 2002. Effect of human immunodeficiency virus (HIV) type 1 envelope subtypes A and D on disease progression in a large cohort of HIV-1-positive persons in Uganda. *J. Infect. Dis.* 185, 1244–1250.
- Kaleebu, P., Nankya, I.L., Yirrell, D.L., et al., 2007. Relation between chemokine receptor use, disease stage, and HIV-1 subtypes A and D: results from a rural Ugandan cohort. *J. Acquir. Immune Defic. Syndr.* 45, 28–33.
- Kiwanuka, N., Laeyendecker, O., Robb, M., et al., 2008. Effect of human immunodeficiency virus Type 1 (HIV-1) subtype on disease progression in persons from Rakai, Uganda, with incident HIV-1 infection. *J. Infect. Dis.* 197, 707–713.
- Lajoie, J., Fontaine, J., Tremblay, C., Routy, J.-P., Poudrier, J., Roger, M., 2009. Persistence of high levels of blood soluble human leukocyte antigen-G is associated with rapid progression of HIV infection. *AIDS* 23, 1437–1440.
- Le Clerc, S., Limou, S., Coulonges, C., et al., 2009. Genomewide association study of a rapid progression cohort identifies new susceptibility alleles for AIDS (ANRS Genomewide Association Study 03). *J. Infect. Dis.* 200, 1194–1201.
- Leigh Brown, A.J., Lycett, S.J., Weinert, L., Hughes, G.J., Fearnhill, E., Dunn, D.T., 2011. Transmission network parameters estimated from HIV sequences for a nationwide epidemic. *J. Infect. Dis.* 204, 1463–1469.
- Lemey, P., Derdelinckx, I., Rambaut, A., et al., 2005. Molecular footprint of drug-selective pressure in a human immunodeficiency virus transmission chain. *J. Virol.* 79, 11981–11989.
- Lemey, P., Kosakovsky Pond, S.L., Drummond, A.J., et al., 2007. Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. *PLoS Comput. Biol.* 3, e29.
- Lengauer, T., Sander, O., Sierra, S., Thielen, A., Kaiser, R., 2007. Bioinformatics prediction of HIV co-receptor usage. *Nat. Biotechnol.* 25, 1407–1410.
- Leslie, A.J., Pfafferott, K.J., Chetty, P., et al., 2004. HIV evolution: CTL escape mutation and reversion after transmission. *Nat. Med.* 10, 282–289.
- Li, G., Verheyen, J., Rhee, S.-Y., Voet, A., Vandamme, A.-M., Theys, K., 2013. Functional conservation of HIV-1 Gag: implications for rational drug design. *Retrovirology* 10, 126.
- Li, X., Xue, Y., Zhou, L., et al., 2014. Evidence that HIV-1 CRF01\_AE is associated with low CD4 + T cell count and CXCR4 co-receptor usage in recently infected young men who have sex with men (MSM) in Shanghai, China. *PLoS ONE* 9, e89462.
- Limou, S., Le Clerc, S., Coulonges, C., et al., 2009. Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). *J. Infect. Dis.* 199, 419–426.
- Liovat, A.-S., Rey-Cuillé, M.-A., Lécuroux, C., et al., 2012. Acute plasma biomarkers of T cell activation set-point levels and of disease progression in HIV-1 infection. *PLoS ONE* 7, e46143.
- Mann, J.K., Byakwaga, H., Kuang, X.T., et al., 2013. Ability of HIV-1 Nef to downregulate CD4 and HLA class I differs among viral subtypes. *Retrovirology* 10, 100.
- Martínez-Picado, J., Prado, J.G., Fry, E.E., et al., 2006. Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. *J. Virol.* 80, 3617–3623.
- Murphy, K., Hoover, D.R., Shi, Q., et al., 2013. Association of self-reported race with AIDS death in continuous HAART users in a cohort of HIV-infected women in the United States. *AIDS* 27, 2413–2423.
- Ng, O.T., Laeyendecker, O., Redd, A.D., et al., 2014. HIV type 1 polymerase gene polymorphisms are associated with phenotypic differences in replication capacity and disease progression. *J. Infect. Dis.* 209, 66–73.
- Pawłowski, A., Jansson, M., Sköld, M., Rottenberg, M.E., Källénus, G., 2012. Tuberculosis and HIV co-infection. *PLoS Pathog.* 8, e1002464.
- Pérez, L., Thomson, M.M., Bleda, M.J., et al., 2006. HIV type 1 molecular epidemiology in Cuba: high genetic diversity, frequent mosaicism, and recent expansion of BG intersubtype recombinant forms. *AIDS Res. Hum. Retroviruses* 22, 724–733.
- Pérez, L., Kourí, V., Alemán, Y., et al., 2013. Antiretroviral drug resistance in HIV-1 therapy-naïve patients in Cuba. *Infect. Genet. Evol.* 16, 144–150.
- Pineda-Peña, A.-C., Faria, N.R., Imbrechts, S., et al., 2013. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: performance evaluation of the new REGA version 3 and seven other tools. *Infect. Genet. Evol.* 19, 337–348.
- Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5, e9490.
- Pushker, R., Jacqué, J.-M., Shields, D.C., 2010. Meta-analysis to test the association of HIV-1 nef amino acid differences and deletions with disease progression. *J. Virol.* 84, 3644–3653.
- Ragonnet-Cronin, M., Hodcroft, E., Hué, S., et al., 2013. Automated analysis of phylogenetic clusters. *BMC Bioinforma.* 14, 317.
- Roberts, L., Passmore, J.S., Williamson, C., et al., 2011. Disease progression. 24, 819–831.
- Rutledge, T.F., Boyd, M.F., 2008. Morbidity and Mortality Weekly Report Revised Surveillance Case Definitions for HIV Infection Among Adults, Adolescents, and Children Aged <18 Months and for HIV Infection and AIDS Among Children Aged 18 Months to <13 Years — United States. p. 57 (depa. 2008).
- Sapsuthiphas, S., Tsuchiya, N., Pathipavanich, P., et al., 2013. CRF01\_AE-specific neutralizing activity observed in plasma derived from HIV-1-infected Thai patients residing in northern Thailand: comparison of neutralizing breadth and potency between plasma derived from rapid and slow progressors. *PLoS ONE* 8, e53920.
- Secor, W.E., 2012. The effects of schistosomiasis on HIV/AIDS infection, progression and transmission. *Curr. Opin. HIV AIDS* 7, 254–259.
- Sierra, M., Thomson, M.M., Posada, D., et al., 2007. Identification of 3 phylogenetically related HIV-1 BG intersubtype circulating recombinant forms in Cuba. *J. Acquir. Immune Defic. Syndr.* 45, 151–160.

- Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690.
- Struck, D., Perez-Bercoff, D., Devaux, C.S.J., 2010. COMET: a novel approach to HIV-1 subtype prediction. 8th European HIV Drug Resistance Workshop.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739.
- Theys, K., Deforche, K., Vercauteren, J., et al., 2012. Treatment-associated polymorphisms in protease are significantly associated with higher viral load and lower CD4 count in newly diagnosed drug-naïve HIV-1 infected patients. *Retrovirology* 9, 81.
- Thomson, M.M., Nájera, R., 2005. Molecular epidemiology of HIV-1 variants in the global AIDS pandemic: an update. *AIDS Rev.* 7, 210–224.
- Tscherning, C., Alaeus, A., Fredriksson, R., et al., 1998. Differences in chemokine co-receptor usage between genetic subtypes of HIV-1. *Virology* 241, 181–188.
- Van Laethem, K., Schrooten, Y., Lemey, P., et al., 2005. A genotypic resistance assay for the detection of drug resistance in the human immunodeficiency virus type 1 envelope gene. *J. Virol. Methods* 123, 25–34.
- Vandekerckhove, L.P.R., Wensing, A.M.J., Kaiser, R., et al., 2011. European guidelines on the clinical management of HIV-1 tropism testing. *Lancet Infect. Dis.* 11, 394–407.
- Vaughan, H.E., Cane, P., Pillay, D., Tedder, R.S., 2003. Characterization of HIV type 1 clades in the Caribbean using pol gene sequences. *AIDS Res. Hum. Retroviruses* 19, 929–932.
- Wang, B., 2013. Viral factors in non-progression. *Front. Immunol.* 4, 355.