J Antimicrob Chemother 2018; **73**: 1054–1059
doi:10.1093/jac/dkx508 Advance Access publication 23 January 2018

Journal of Antimicrobial Chemotherapy

Impact of antiretroviral resistance and virological failure on HIV-1 informational entropy

Elidamar Nunes de Carvalho Lima[1,2], José Roberto Castilho Piqueira[2], Michelle Camargo[1], Juliana Galinskas[1], Maria Cecilia Sucupira[1] and Ricardo Sobhie Diaz[1]*

[1]*Division of Infectious Diseases, Paulista School of Medicine, Federal University of São Paulo-UNIFESP, São Paulo, SP, Brazil;*
[2]*Telecommunication and Control Engineering Department, Engineering School, University of São Paulo, São Paulo, SP, Brazil*

*Corresponding author. Tel: +55-(11)-5576-4834; Fax: +55-(11)-4192-3176; E-mail: rsdiaz@catg.com.br

Received 3 August 2017; returned 25 October 2017; revised 4 December 2017; accepted 7 December 2017

Objectives: The present study investigated the relationship between genomic variability and resistance of HIV-1 sequences in protease (PR) and reverse transcriptase (RT) regions of the *pol* gene. In addition, we analysed the resistance among 651 individuals presenting antiretroviral virological failure, from 2009 to 2011, in the state of São Paulo, Brazil.

Methods: The genomic variability was quantified by using informational entropy methods and the relationship between resistance and replicative fitness, as inferred by the residual viral load and CD4+ T cell count.

Results: The number of antiretroviral schemes is related to the number of resistance mutations in the HIV-1 PR ($\alpha = 0.2511$, $P = 0.0003$, $R^2 = 0.8672$) and the RT ($\alpha = 0.7892$, $P = 0.0001$, $R^2 = 0.9141$). Increased informational entropy rate is related to lower levels of HIV-1 viral loads ($\alpha = -0.0121$, $P = 0.0471$, $R^2 = 0.7923$), lower levels of CD4+ T cell counts ($\alpha = -0.0120$, $P = 0.0335$, $R^2 = 0.8221$) and a higher number of antiretroviral resistance-related mutations.

Conclusions: Less organized HIV genomes as inferred by higher levels of informational entropy relate to less competent host immune systems, lower levels of HIV replication and HIV genetic evolution as a consequence of antiretroviral resistance.

Introduction

The selective pressure exerted by any antiretroviral can select mutants with several levels of viral resistance.[1–5] During the evolutionary process, the genome accumulates and stores information about its environment. This information, already fixed in the genome and transmitted to future generations, may be adaptive or not, depending on the information context previously accumulated in a particular environment (the environment is heterogeneous and varies over time). The acquired genetic information is not easily lost and co-evolves with the environment.[6–8] Several studies indicate that when changes in the environment occur, viruses increase the information content in their genome, leading to a genetic evolution, which can result in a decrease in replicative fitness.[9–12]

For some HIV genomic regions, such as protease (PR), selected primary mutations are initially followed by a dramatic decrease in viral fitness to allow viral replication in the presence of antiretrovirals. Additional selected resistance mutations increase the level of resistance, thereby restoring replicative fitness.[13–15]

PR is a structural protein and an enzyme responsible for the selective cleavage of viral polyproteins—a crucial step during the late phase of the viral cycle. PIs are designed based on the structural characteristics of the PR molecule. However, resistance mutations induce disorders in the protein and inhibit long-term binding to the cleavage site of this molecule.[16–18] Reverse transcriptase (RT) is an enzyme that reads the sequence of viral RNA nucleic acids that enter the host cell and transcribes them into complementary DNA sequences.[19,20]

To comprehend the relationship between HIV-1 genetic diversity and the presence of antiretroviral resistance mutations, the genomic sequences encoding PR and RT of HIV-1 were analysed among individuals presenting virological antiretroviral failure. This study aims to quantify the informational entropy and correlate these values with laboratory markers related to HIV-1 disease progression (CD4+ T cell count and viral load), the number of ART schemes followed and the number/profile of resistance related to resistance mutation. Entropy is a measure of the average uncertainty of symbols and is used mainly to quantify uncertainty within a system; it is convenient in many aspects, and can, for example, serve as a measure of variability in a discrete variable.[21–24]

Methods

This survey includes 651 sequences of PR and RT HIV-1 among individuals presenting antiretroviral virological failure to NRTIs with NNRTIs and/or PIs. The samples were evaluated as part of the Brazilian Network for HIV-1 Genotyping (RENAGENO).[25] Genotypes were determined at the Retrovirology Laboratory of the Federal University of São Paulo in Brazil, between 2009 and 2011. The sequences were aligned by using ClustalW[26] and MatLab[27] software, and subtyped on COMET[28] and jpHMM;[29] all selected sequences belong to HIV-1 subtype B. The length of the PR genome was 96 amino acids from residue 4 to 99 and the RT length was 210 amino acids from residue 38 to 247.

We analysed the relationship between the antiretroviral classes received and the frequency of resistance-associated mutations. The sequences were separated into groups according to the RT and PR regions of the *pol* gene. To understand how the informational entropy rate can affect replicative fitness, as inferred by the residual viral load and the CD4+ T cell count, an informational measure of diversity was employed to quantify the number of amino acid variations in each position of the HIV-1 PR and the RT, calculated from Equation (1). The entropy is a standard metric to evaluate protein variability, and considers the number of possible amino acids replaced and their frequency.[30,31] Informational entropy H is commonly used to quantify the uncertainty of information about the amino acid or protein at a given position and their fixation over time. The classical Shannon formula for the mean entropy, or information content, per position of amino acid sequence is written as:

$$H(\mathrm{X}) = -\sum p(x)\ \ln\ [p(x)] \quad (1)$$

where $p(x)$ is the probability (frequency) of the base (A, T, C, G) in the given sequence, x is the total number of sequences compared and ln represents the natural logarithm. A value of 0 for H indicates that all sequences are identical, whereas a value of 1 indicates that in a given position all amino acids differ from the others.[30–34] The entropy analysis was performed using MatLab software[27] and the relation between groups was performed for each genome position, by pairwise correlation, as well as based on the corresponding P value. We analysed the correlation between: (i) the antiretroviral class administered and the resistance-related mutations; (ii) the number of ART schemes and the number of resistance mutations; and (iii) the informational entropy and viral loads and CD4+ T cell counts. In these methods, entropy increase represents a decrease in the informational content of this protein. To analyse correlation between the entropy and markers of disease progression, the sequences of the PR and RT of HIV-1 from patients under virological failure were sequentially divided into five groups according to the intervals of viral load in $\log_{10}$ (0.4 $\log_{10}$ intervals), where group 1 comprised individuals with viral loads <3.7 $\log_{10}$ copies/mL ($n = 96$), group 2 comprised individuals with viral loads between 3.8 and 4.2 $\log_{10}$ copies/mL ($n = 94$), group 3 comprised individuals with viral loads between 4.3 and 4.7 $\log_{10}$ copies/mL ($n = 98$), group 4 comprised individuals with viral loads between 4.8 and 5.2 $\log_{10}$ copies/mL ($n = 94$) and group 5 comprised individuals with viral loads >5.3 $\log_{10}$ copies/mL ($n = 92$). For these correlations, the CD4+ T cell counts were sequentially divided into group 1 comprising individuals with CD4+ T cell counts <100 cells/mm^3 ($n = 93$), group 2 comprising individuals with CD4+ T cell counts ranging from 101 to 265 cells/mm^3 ($n = 96$), group 3 comprising individuals with CD4+ T cell counts ranging from 266 to 372 cells/mm^3 ($n = 98$), group 4 comprising individuals with CD4+ T cell counts ranging from 373 to 540 cells/mm^3 ($n = 93$) and group 5 comprising individuals with CD4+ T cell counts >540 cells/mm^3 ($n = 94$).

Data validation and graphs

The main objective was to quantify and understand the relationship between the ART scheme used and the resistance mutation number. The data were grouped according to previously used antiretroviral schemes (one to nine), and the number of antiretroviral resistance mutations in PR and RT. The number of resistance mutations was determined according to the International Antiviral Society-USA (IAS-USA) mutation list.[35] To analyse the relationship between resistance-related mutations and the number of antiretrovirals used, two distinct analyses were performed evaluating the mutation number within and between groups. As the number of mutations increases, we calculated the relationship of the resistance mutation number and the ART received using Pearson's coefficient, according to the R^2 and the P value of each metric. To analyse the relationship between the groups, linear correlation was performed in addition to the R^2 to verify the metric adherence. The statistical analysis was performed using Excel®, MatLab and GraphPadPrism software.[27,36] We considered a P value <0.05 to be statistically significant.

Results

Number of received ART schemes and the number of resistance-related mutations

Figure 1 shows the relationship between the number of antiretroviral schemes and the number of resistance mutations in the PR and RT regions of the HIV-1 *pol* gene. For patients who received more than six antiretroviral schemes, the number of resistance mutations increased compared with patients who received one antiretroviral scheme, which was statistically significant for PR ($\alpha = 0.2511$, $P = 0.0003$, $R^2 = 0.8672$) and RT ($\alpha = 0.7892$, $P = 0.0001$, $R^2 = 0.9141$).

Relationship between the administered antiretroviral class and the selected resistance-related mutations

The manner in which the frequency of the administered therapeutic class can affect the relationship with the numbers of resistance mutations was analysed. To assess this, a PR and RT HIV-1 sequence was employed with antiretroviral resistance-associated mutations as previously defined. The number of antiretroviral exposures is clearly related to the number of selected antiretroviral resistance mutations (Figure 2). Among the individuals failing the first treatment scheme with NRTI and NNRTI, the prevalence of NRTI mutations was 26.7% and the prevalence of NNRTI mutations was 9.7%. For the NRTI and PI combinations evaluated, the prevalence of NRTI mutations was 6.9%, whereas the prevalence of PI mutations was 2.6%. For NRTI, NNRTI and PI combination schemes, the prevalence of mutations was 37.9%, 5.5% and 10.7%, respectively.

Correlation between informational entropy and the laboratorial markers of HIV-1 disease progression

To understand how the informational entropy affects the inferred replicative fitness and the residual immunity in HIV-1, the informational entropy and the relationships with the viral load and the CD4+ T cell count were analysed. The viral load mean and median were 4.5 and 4.6 $\log_{10}$ copies/mL, respectively, with a range from below detection limits to 6.1 $\log_{10}$ copies/mL, whereas the CD4+ T cell count mean and median were 383.9 and 319 cells/mm^3, respectively, ranging from 10 to 2318 cells/mm^3. We observed a correlation between the informational entropy and the viral loads and the CD4+ T cell counts. As seen in Figure 3(a and b), the

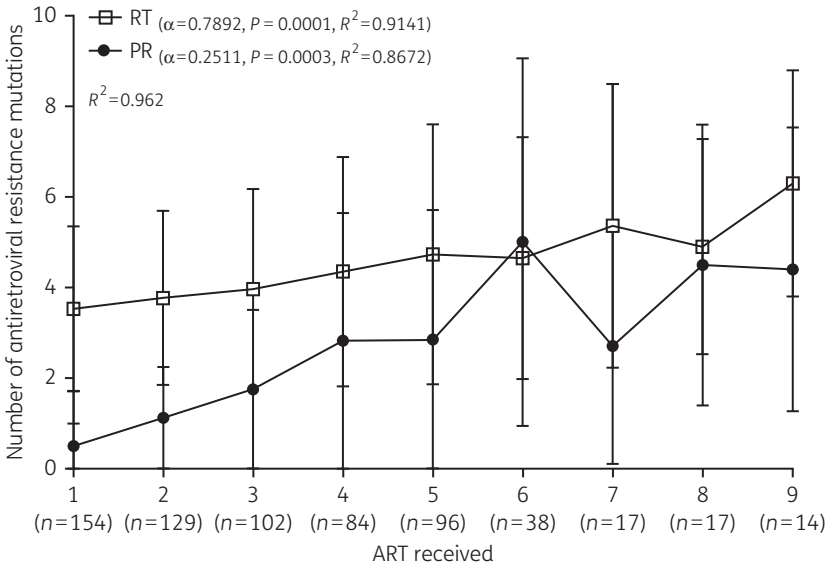

Figure 1. Relationship between the number of received ART schemes and the mean number of detected resistance-associated mutations in the RT and PR regions of the *pol* gene. The numbers of patients are shown in parentheses.

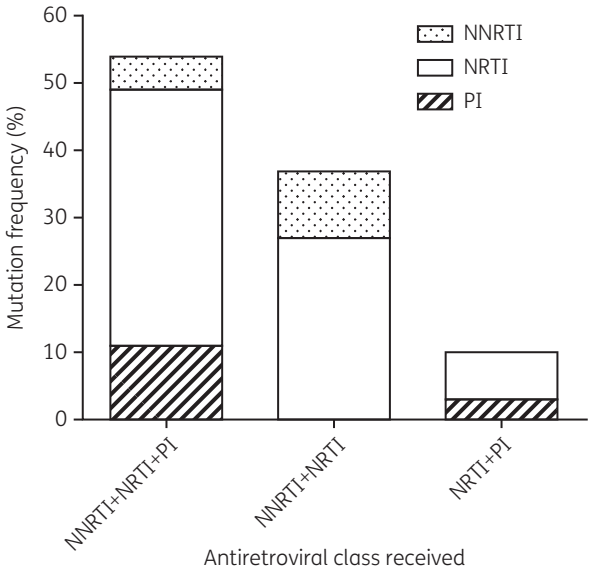

Figure 2. Relationship between the antiretroviral class received and the prevalence of resistance-associated mutations.

informational entropy correlates to the residual HIV replication, as inferred by the viral loads among patients presenting virological failure, and the residual host immunity, as reflected by the CD4+ T cell counts at the time of virological failure ($\alpha = -0.0121$, $P = 0.0471$, $R^2 = 0.7923$ for viral loads; $\alpha = -0.0120$, $P = 0.0335$, $R^2 = 0.8221$ for CD4+ T cell counts). In summary, when the entropy is high, the viral load and the CD4+ T cell count are low, indicating a negative correlation among these quantitative variables.

Discussion

In this study we sought to find the relationships between a number of variables determined in HIV-infected individuals experiencing ART failure. These variables included the resistance mutation profiles from the PR and RT regions of the *pol* gene according to the data generated from Sanger sequencing, the number of previous ARTs in which these individuals have failed, CD4+ T cell counts and HIV viral loads at the moment of virological failure; all of this according to the informational entropy estimation. First of all, and completely intuitively, we have been able to find a positive correlation between the number of previous virological failures and the number of resistance-related mutations. Some studies have pointed out that the magnitude of previous antiretroviral exposure has higher negative predictive values than the genotypic resistance profile in the outcome of salvage therapy antiretroviral schemes.[37,38]

It is well known that the results of genotypic resistance tests for antiretroviral drugs most likely reflect the selective pressure exerted by antiretrovirals used at that moment, and sometimes do not reflect the mutations that have been selected in the past by other antiretrovirals. In this context, genotypic resistance tests present a high positive predictive value and a low negative predictive value. One may argue, in contrast, that the high number of antiretroviral schemes used in the past most likely reflects lack of adherence and could be related to the lower prevalence of resistance-related mutations.

We believe that our results, as well as results reported elsewhere, provide an interesting and unique insight into the relationship between informational entropy and the interplay of HIV and human hosts. The variability patterns in genomic sequences can be quantified by applying informational entropy methods, a reliable tool for measuring probabilities and determining the

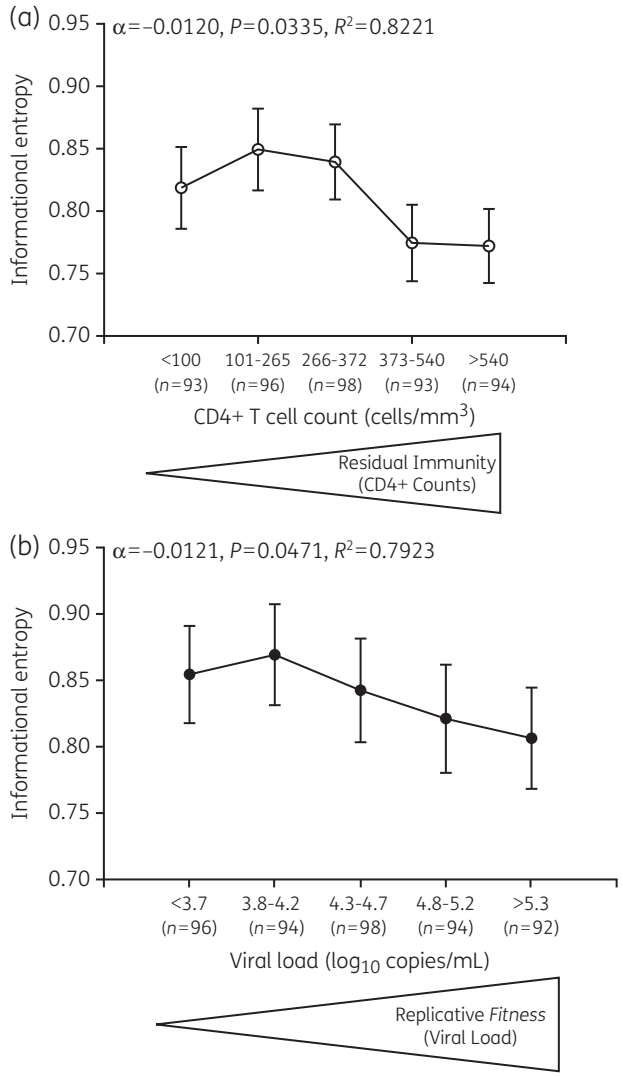

Figure 3. Correlation between the informational entropy and laboratory markers associated with HIV-1 disease progression. The CD4+ T cell count and viral load of patients under virological failure were classified into five groups according to the CD4+ T cell count (a) or viral load (b). The numbers of included sequences from patients are shown in parentheses.

information in sequences of aligned genomes.[7,9–11,39] High informational entropy equates to disorganization of a given environment and to a high level of uncertainty. In this context, a very organized and strong immune system from a host in the context of HIV would be accompanied by lower levels of informational entropy observed in genomic sequences from HIV at a given moment. On the other hand, high informational entropy would reflect a less adapted HIV in a given moment.

We were able to detect that among individuals on ART and experiencing virological failure, the levels of HIV informational entropy were inversely correlated with CD4+ T cell counts, which, as suggested above, suggests that more competent or stronger immune systems from a host may be related to a more genetically stable HIV. In agreement with the above, other studies demonstrate that an increase in the informational entropy of the HIV genome correlates with the efficacy of cytotoxic T-lymphocytes in the recognition and affinity of HIV epitopes.[40–42] Furthermore, we were also able to find a correlation of higher HIV viral loads and lower levels of HIV informational entropy. Higher HIV viral loads in a given human host may reflect higher replication capacity of HIV, which also may correlate with so-called viral fitness. Mutations in PR and RT that cause resistance to antiretrovirals are often associated with the structure and function of HIV enzymes, leading to a decrease in the viral fitness.[43–47] In this context, we can also speculate that a decrease in surrogate markers of HIV fitness, such as viral loads, may be accompanied by a loss of genomic organization. Indeed, it has been previously reported that individuals recently infected with HIV, whose viral loads are higher prior to the viral set point, harbour strains with lower informational entropy as compared with individuals with established HIV infection.[48]

Significantly, we also describe here a relationship between the number of resistance-related mutations and entropy, where a lower number of mutations is accompanied by lower levels of informational entropy. This finding also confirms that selection of HIV resistant to antiretrovirals strongly decreases HIV organization.

We recognize that HIV fitness has not been measured directly in this study. We also recognize that the inclusion of a larger sample size in the current analysis would strengthen the inferences about the relationship between HIV-1 fitness and information content. However, as discussed above, the results presented here, taken together with results from other studies, suggest that originally transmitted HIV soon after HIV infection, when the host immune system is still preserved, has a more organized genomic structure as inferred by the lower levels of informational entropy. As HIV evolves due to the presence of escape mutants and disease progression or due to the selective pressure imposed by antiretrovirals, a less organized genome will emerge, which in part contrasts to original Darwinian concepts.

Acknowledgements

We thank the Brazilian Network for HIV-1 Genotyping (RENAGENO), sponsored by the Brazilian Ministry of Health, for providing access to patient information.

Funding

This work was financially supported by Fundação de Amparo a Pesquisa do Estado de Sao Paulo (FAPESP), research grant # 2011/12156-0 to R. S. D. and a PhD scholarship from Cooedenação de Aperfeicoamento de Pessoal de Nível Superior (CAPES) to E. N. d. C. L. The funders had no role in study design, data collection and analysis, decision to publish or the preparation of the manuscript.

Transparency declarations

None to declare.

Author contributions

Design of research: E. N. d. C. L., R. S. D. and J. R. C. P. DNA sequence generation from plasma samples: M. C. and J. G. Analysis of antiretroviral resistance: M. C. S. and E. N. d. C. L. Analysis and interpretation of the

experiments: E. N. d. C. L. Writing of the manuscript: E. N. d. C. L., R. S. D. and J. R. C. P. All authors read and approved the final manuscript.

References

1 Günthard HF, Leigh-Brown AJ, D'Aquila RT *et al*. Higher selection pressure from antiretroviral drugs *in vivo* results in increased evolutionary distance in HIV-1 *pol*. *Virology* 1999; **259**: 154–65.

2 Santoro MM, Perno CF. HIV-1 genetic variability and clinical implications. *ISRN Microbiol* 2013; **2013**: 481314.

3 Vazquez-Guillen JM, Palacios-Saucedo GC, Rivera-Morales LG *et al*. Mutations related to antiretroviral resistance identified by ultra-deep sequencing in HIV-1 infected children under structured interruptions of HAART. *PLoS One* 2016; **11**: e0147591.

4 Bahia F, Pedroso C, Netto EM *et al*. Evaluation of the genotypic pattern of HIV-1 resistance in AIDS patients failing antiretroviral therapy. *Braz J Infect Dis* 2004; **8**: 281–9.

5 Banke S, Lillemark MR, Gerstoft J *et al*. Positive selection pressure introduces secondary mutations at Gag cleavage sites in human immunodeficiency virus type 1 harboring major protease resistance mutations. *J Virol* 2009; **83**: 8916–24.

6 Østman B, Hintze A, Adami C. Impact of epistasis and pleiotropy on evolutionary adaptation. *Proc Biol Sci* 2011; **279**: 247–56.

7 Schneider TD. Evolution of biological information. *Nucleic Acids Res* 2000; **28**: 2794–9.

8 El-Diasty F. Origin of order: emergence and evolution of biological organization as a problem in thermal physics. *Adv Life Sci* 2011; **1**: 30–9.

9 Gupta A, Adami C. Strong selection significantly increases epistatic interactions in the long-term evolution of a protein. *PLoS Genet* 2016; **12**: e1005960.

10 Adami C. Information theory in molecular biology. *Phys Life Rev* 2004; **1**: 3–22.

11 Donaldson-Matasci MC, Bergstrom CT, Lachmann M. The fitness value of information. *Oikos* 2010; **119**: 219–30.

12 Adami C. The use of information theory in evolutionary biology. *Ann N Y Acad Sci* 2012; **1256**: 49–65.

13 Collins JA, Thompson MG, Paintsil E *et al*. Competitive fitness of nevirapine-resistant human immunodeficiency virus type 1 mutants. *J Virol* 2004; **78**: 603–11.

14 Dapp MJ, Heineman RH, Mansky LM. Interrelationship between HIV-1 fitness and mutation rate. *J Mol Biol* 2013; **425**: 41–53.

15 Hu Z, Kuritzkes DR. Altered viral fitness and drug susceptibility in HIV-1 carrying mutations that confer resistance to nonnucleoside reverse transcriptase and integrase strand transfer inhibitors. *J Virol* 2014; **88**: 9268–76.

16 Mata-Munguía C, Escoto-Delgadillo M, Torres-Mendoza B *et al*. Natural polymorphisms and unusual mutations in HIV-1 protease with potential antiretroviral resistance: a bioinformatic analysis. *BMC Bioinformatics* 2014; **15**: 72.

17 Mager PP. The active site of HIV-1 protease. *Med Res Rev* 2001; **21**: 348–53.

18 Torbeev VY, Raghuraman H, Hamelberg D *et al*. Protein conformational dynamics in the mechanism of HIV-1 protease catalysis. *Proc Natl Acad Sci USA* 2011; **108**: 20982–7.

19 Sarafianos SG, Marchand B, Das K *et al*. Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J Mol Biol* 2009; **385**: 693–713.

20 Hu W-S, Hughes SH. HIV-1 reverse transcription. *Cold Spring Harb Perspect Med* 2012; **22**: a006882.

21 Shannon CE. A mathematical theory of communication. *Bell Syst Tech J* 1948; **27**: 379–423.

22 Tenreiro MacHado JA. Shannon entropy analysis of the genome code. *Math Probl Eng* 2012; **2012**: 132625.

23 Pan K, Deem MW. Quantifying selection and diversity in viruses by entropy methods, with application to the hemagglutinin of H3N2 influenza. *J R Soc Interface* 2011; **8**: 1644–53.

24 Nair C, Prabhakar B, Shah D. On entropy for mixtures of discrete and continuous variables. *arXiv* 2007; cs/0607075.

25 Souza DC, Sucupira MCA, Brindeiro RM *et al*. The Brazilian network for HIV-1 genotyping external quality control assurance programme. *J Int AIDS Soc* 2011; **14**: 45.

26 Larkin MA, Blackshields G, Brown NP *et al*. Clustal W and Clustal X version 2.0. *Bioinformatics* 2007; **23**: 2947–8.

27 MathWorks. *Bioinformatics Toolbox™: User's Guide (R2017b)*. https://www.mathworks.com/help/pdf_doc/bioinfo/bioinfo_ug.pdf.

28 Struck D, Lawyer G, Ternes AM *et al*. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. *Nucleic Acids Res* 2014; **42**: e144.

29 Schultz AK, Zhang M, Bulla I *et al*. jpHMM: improving the reliability of recombination prediction in HIV-1. *Nucleic Acids Res* 2009; **37**: 647–51.

30 Funkhouser S. The entropy of a discrete real variable. *Entropy* 2012; **14**: 1522–38.

31 Gatenby RA, Frieden BR. Information theory in living systems, methods, applications, and challenges. *Bull Math Biol* 2007; **69**: 635–57.

32 Xu W, Miranker DP. A metric model of amino acid substitution. *Bioinformatics* 2004; **20**: 1214–21.

33 Pandey A. Entropy and codon bias in HIV-1. *bioRxiv* 2016; doi:10.1101/052274.

34 Smith JM. The concept of information in biology. *Phil Sci* 2000; **67**: 177–94.

35 Wensing AM, Calvez V, Gunthard F *et al*. Update of the drug resistance mutations in HIV-1. *Top Antivir Med* 2017; **24**: 132–41.

36 Motulsky HJ. *GraphPad Statistics Guide*. http://www.graphpad.com/guides/prism/7/statistics/index.htm.

37 Lazzarin A, Clotet B, Cooper D *et al*. Efficacy of enfuvirtide in patients infected with drug-resistant HIV-1 in Europe and Australia. *N Engl J Med* 2003; **348**: 2186–95.

38 Imaz A, Falcó V, Ribera E. Antiretroviral salvage therapy for multiclass drug-resistant HIV-1-infected patients: from clinical trials to daily clinical practice. *AIDS Rev* 2011; **13**: 180–93.

39 Abidi SH, Kalish ML, Abbas F *et al*. HIV-1 subtype A gag variability and epitope evolution. *PLoS One* 2014; **9**: e93415.

40 Liu Y, McNevin J, Zhao H *et al*. Evolution of human immunodeficiency virus type 1 cytotoxic T-lymphocyte epitopes: fitness-balanced escape. *J Virol* 2007; **81**: 12179–88.

41 Yusim K, Kesmir C, Gaschen B *et al*. Clustering patterns of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. *J Virol* 2002; **76** Suppl: 8757–68.

42 Boutwell CL, Carlson JM, Lin TH *et al*. Frequent and variable cytotoxic-T-lymphocyte escape-associated fitness costs in the human immunodeficiency virus type 1 subtype B Gag proteins. *J Virol* 2013; **87**: 3952–65.

43 Mammano F, Petit C, Clavel F. Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients. *J Virol* 1998; **72**: 7632–7.

44 Hinkley T, Martins J, Chappey C *et al*. A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. *Nat Genet* 2011; **43**: 487–9.

45 Arts EJ, Hazuda DJ. HIV-1 antiretroviral drug therapy. *Cold Spring Harb Perspect Med* 2012; **2**: a007161.

46 Menéndez-Arias L, Martínez MA, Quiñones-Mateu ME *et al*. Fitness variations and their impact on the evolution of antiretroviral drug resistance. *Curr Drug Targets Infect Disord* 2003; **3**: 355–71.

47 Buckheit RWJ. Understanding HIV resistance, fitness, replication capacity and compensation: targeting viral fitness as a therapeutic strategy. *Expert Opin Investig Drugs* 2004; **13**: 933–58.

48 Wu JW, Patterson-Lomba O, Novitsky V *et al*. A generalized entropy measure of within-host viral diversity for identifying recent HIV-1 infections. *Medicine (Baltimore)* 2015; **94**: e1865.