22nd International BioInformatics Workshop on Virus Evolution and Molecular Epidemiology

A1 Pairwise diversity and tMRCA as potential markers of HIV infection recency

Sikhulile Moyo,[1,2] Eduan Wilkinson,[3] Alain Vandormael,[3] Rui Wang,[4] Jia Weng,[4] Kenanao P. Kotokwe,[2] Simani Gaseitsiwe,[2,4] Rosemary Musonda,[2,4] Joseph Makhema,[2,4] Max Essex,[2,4] Susan Engelbrecht,[1,5] Tulio de Oliveira,[3,6,7] and Vladimir Novitsky[2,4]

[1]Division of Medical Virology, Stellenbosch University, Tygerberg, South Africa, [2]Botswana-Harvard AIDS Institute Partnership, Gaborone, Botswana, [3]Wellcome Trust Africa Centre for Health and Population Studies, Doris Duke Medical Research Centre, Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa, [4]Harvard T.H. Chan School of Public Health, Boston, MA, USA, [5]National Health Laboratory Services (NHLS), Tygerberg Coastal, South Africa, [6]Research Department of Infection, University College London, London, UK and [7]College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa

Intra-host HIV-1 diversity increases linearly over time. We assessed the extent to which mean pairwise distances and the time to the most recent common ancestor (tMRCA) inferred from intra-host HIV-1 C env sequences were associated with the estimated time of HIV infection. Data from a primary HIV-1 C infection study in Botswana were used for this analysis ($N = 42$). A total of 2,540 HIV-1 C env gp120 V1C5 viral sequences were generated by single genome amplification and sequencing, with an average of sixty-one viral sequences per participant and eleven sequences per time point per participant. Raw pairwise distances were calculated for each time point and participant using the ape package in R software. The tMRCA was estimated using phylogenetic inference implemented in BEAST v1.8.2. Pairwise distances and tMRCA were significantly associated with the estimated time since HIV infection (both $P < 0.001$). Taking into account multiplicity of HIV infection strengthened these associations. HIV-1 C env-based pairwise distances and tMRCA can be used as potential markers for HIV recency. However, the tMRCA estimates demonstrated no advantage over the pairwise distances estimates.

A2 Optimization of the results generated by large-scale sequencing for the study of drug resistance in HIV infection: A systematic review

J. A. Fernández-Caballero,[1] N. Chueca,[1] E. Poveda,[2] and F. García[1]

[1]Department of Clinical Microbiology, Instituto de Investigación Biosanitaria (IBIS)-Complejo Hospitalario Universitario de Granada, Granada, Spain and [2]Division of Clinical Virology, Instituto de Investigación Biomédica de A Coruña (INIBIC)-Complexo Hospitalario Universitario de A Coruña (CHUAC), Sergas, Universidade da Coruña (UDC), Spain

Next-generation sequencing (NGS) approaches are now used in many clinical diagnostic laboratories for the routine diagnosis of resistance to antiretrovirals approved for the treatment of HIV infection. As some of NGS platforms may be a source of sequencing error, it is necessary to improve currently available protocols and implement bioinformatics tools that may help to correctly identify the presence of resistance mutations with clinical impact. In this study, we reviewed all studies dealing with software or methods aiming to decrease these errors, published during the period 2006–2016. We considered, as bioinformatic strategies, software aiming to delete or detect sequencing errors, and as protocol improvements, those changes in PCR temperature profiles and/or reagent concentration aiming to minimize sequencing errors. We used a combination of non-MeSH and MeSH terms related to error correction and NGS sequence filtering. All abstracts of papers available through January 2006 and June 2016 were reviewed. Our search identified 611 studies, we finally selected seven papers that met all the eligibility criteria, three of which dealt with protocol modifications and four with bioinformatics aiming to eliminate errors. Some studies are mainly focused on improving protocols for decreasing the magnitude of errors during the polymerase change reaction. Other studies propose specific bioinformatics tools, able to reach both a 93–98 per cent reduction of indels (insertions/deletions) and a sensitivity and specificity close to 100 per cent in single nucleotide polymorphism variant calling. Moreover, error rates decreased from a median value (95% CI) of 0.2 per cent (0.008–0.4) before processing to 0.06 per cent (0.05–0.08) after using a bioinformatic tool. All the software did not incur in a high loss in the number of reads. New protocols and bioinformatics tools that improve the accuracy of NGS results must be considered for correct analysis of HIV resistance mutations. We recommend using bioinformatic software to filter short and low-quality sequences, and using high fidelity polymerases.

A3 Brazilian network for HIV drug resistance surveillance: An investigation of pre-treatment drug resistance transmission chains

N. M. C. Véras,[1] A. F. Pires,[1,2] A. Tanuri,[3] and A. S. Benzaken[1]

[1]Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Vigilância, Prevenção e Controle das Infecções Sexuais Transmissíveis, HIV/Aids e Hepatites Virais, Setor Administrativo Federal Sul (SAF Sul), Trecho 2, Bloco F, Ed. Premium, Torre I, Auditório, Brasília-DF, CEP 70.070-600, Brazil, [2]Programa de Pós-Graduação em Saúde Coletiva, Faculdade de Medicina, Faculdade de Ciências de Saúde, Universidade de Brasília (UnB), Brasília-DF, Brazil and [3]Laboratório de Virologia Molecular, Departamento de Genética, Instituto de Biologia,

Universidade Federal do Rio de Janeiro (UFRJ), Av. Carlos Chagas Filho, 373, Edifício do Centro de Ciências da Saúde, Bloco A, sala A2-121, Cidade Universitária, Rio de Janeiro-RJ, CEP 21.941-902, Brazil

In Brazil, the transmission of HIV drug resistance strains has increased from 6.6 per cent in 2001 to 9.5 per cent in 2015. Besides being associated with virological failure in first-line antiretroviral therapy, drug resistance can also compromise pre-exposure prophylaxis, mother-to-child transmission prevention, and post-exposure prophylaxis. In order to monitor HIV primary resistance in Brazil, we aimed to identify pre-treatment drug resistance transmission chains in both national and regional levels. Sampling strategy was based on the HIV Threshold Survey methodology (HIV-THS, WHO). Subjects were selected from fifty-one highly populated cities present in all five Brazilian macro-regions. HIV pol subtype was determined by Rega HIV Subtyping Tool and maximum likelihood phylogenetics. The presence of pre-treatment drug resistance transmission clades was verified by maximum likelihood tree (PhyML 3.0) for subtypes B, C, and F, separately, in both national and regional levels. Phylogenetic trees were edited using FigTree v1.4.3. We analyzed 1,566 HIV pol sequences from antiretroviral naïve individuals with recent HIV infection. The presence of surveillance drug resistance mutations (SDRM) was previously characterized by Stanford HIVdb Program, with a nationwide prevalence of 9.5 per cent. Overall, subtype B (66 per cent) was the most prevalent, followed by subtypes C (13 per cent), F (11 per cent) and recombinant forms (10 per cent). Subtypes A, D, and CRF 02_AG were identified in < 1 per cent. The distribution of HIV subtypes was slightly different among the different geographic regions, especially in the South, where subtype C represented 51 per cent of analyzed sequences. Sequences presenting SDRM appeared dispersed on all phylogenetic trees, showing no specific pre-treatment drug resistance transmission clade, when considering both the national level or the five Brazilian geographic regions, separately. The HIV subtype distribution is in accordance with previous reports, emphasizing the high prevalence of subtype C in the Southeast Brazil and the introduction of CRF 02_AG in the Northeast. Phylogenetic analyses showed no specific SDRM transmission clade. Hence, HIV SDRM transmission in Brazil does not seem to occur in a particular population group or geographic region. These results further illustrate the important contribution of phylogenetic studies in predicting future trends in SDRM transmission.

A4 Development of HIV drug resistance in HIV-infected patients failing second line regimen in Zimbabwe: A public health concern

V. Kouamou,[1] D. Katzenstein,[2] T. Shamu,[3] and C. Chimbetete[3]

[1]Infectious Disease Research Laboratory, Department of Medicine, College of Health Science, University of Zimbabwe, Zimbabwe, [2]Division of Infectious Disease, Stanford University Medical Center, USA and [3]Newlands Clinic, Hospital, 56 Enterprise, Harare, Zimbabwe

The countries in southern Africa are not only the most heavily impacted in the world by HIV/AIDS but are now home to the world's largest HIV/AIDS treatment programs, providing care and treatment to millions of people. As antiretroviral therapy (ART) programs in sub-Saharan Africa continue to expand, individuals on ART should be closely monitor to ensure favorable treatment outcomes and to minimize the development and transmission of HIV drug resistance, given the limited antiretroviral drug regimen choices available in these settings. We sought to investigate the frequency and determinants of virologic failure of acquired drug resistance-associated mutations (DRMs) in patients failing protease inhibitor-based antiretroviral treatment in Zimbabwe, using a prospective cohort study with cross-sectional analysis. All participants attended the HIV Clinic at Newlands Hospital-Harare and were on ART for at least twelve months. Participants with virologic failure (VL > 1,000copies/ml) were tested for HIV-DRM. Demographic and clinical data were abstracted from medical records. DRMs were defined according to the Stanford HIV database guidelines. A total number of 187 plasma samples were genotyped, out of the 187 participants. Only 114 participants were on second line regimen. From the 114 participants, ART-associated PI major DRM were identified in forty-five patients (39.47 per cent) with multiclass resistance being the combination of M46I+ I50L+ V82A (2.63 per cent). The most common mutations were M46I (24.77 per cent), V82A (21.43 per cent), I50L (17.70 per cent), I84A (7.08 per cent), and L90M (4.42 per cent). No PI mutations were observed in adolescents with the mean age of < 20 years old and yet were prevalent in adults with mean age of >35 years old. Virologic failure rates in adults were high with the majority of ART-failing adults harboring HIV-DRM, yet in adolescents and young adults this remains an adherence problem. Viral load monitoring and drug resistance testing are urgently needed to maintain future treatment options for the millions of African living with HIV.

A5 Phylogenetic characterization of HIV transmission in Belgium between 2013 and 2015

Kenny Dauwe on behalf of all Belgian AIDS Reference Laboratories

AIDS Reference Laboratory, Department of Clinical Chemistry, Microbiology and Immunology, Ghent University, Ghent, Belgium

A good insight in regional HIV transmission dynamics is important for policy decisions and improved prevention and testing strategies. We present the phylogenetic analysis of a well characterized and dense sampled population of most newly diagnosed HIV patients in Belgium. The patient population consisted of 1,665 individuals newly diagnosed with HIV-1 in 2013, 2014, and 2015 for whom baseline resistance testing has been performed, which is a standard procedure in Belgium. Protease-RT sequences as well as clinical and epidemiological data were collected. Maximum likelihood phylogenetic analysis was performed on concatenated protease-RT sequences trimmed to a total length of 876 nucleotides. Clusters of at least two sequences with a bootstrap value >0.97 and a mean pairwise distance <0.015 were considered as transmission pairs or clusters (n= 177). Since the major aim of the work was a better insight on ongoing HIV transmission, occasional sequences with a genetic distance to the closest relative in the cluster of > 0.030 were removed and considered as individual sequences. Based on tree topology, bootstrap and distance analysis, 873 individual sequences, 93 pairs, 67 small clusters (three to seven individuals), and 17 large clusters (more than seven individuals) were identified. The median cluster size for small and large cluster was respectively 4 (IQR: 3–6) and 10 (IQR: 9–19). Comparison of the characteristics of patients on individual branches and patients in clusters (less than two individuals) revealed that clustered patients are more frequently male (92.6 vs. 65.8 per cent), MSM (77.2 vs. 41.3 per cent), of Belgian origin (68.2 vs. 32.8 per cent), self-reporting infection in Belgium (95.1 vs. 47.4 per cent), infected with subtype B or F (respectively 69.0 vs. 40.5 per cent and 18.8 vs. 2.8 per cent), diagnosed with an infection of <6 months (55.4 vs. 28.8 per cent). They have higher CD4 counts (mean 487 vs. mean 376) and higher viral load (mean log