REVIEW ARTICLE

The Role of Phylogenetics as a Tool to Predict the Spread of Resistance

Anna Zhukova[1], Teresa Cutino-Moguel[2], Olivier Gascuel[1], and Deenan Pillay[3,4]

1 Unité Bioinformatique Evolutive, C3BI USR 3756 Institut Pasteur et CNRS, 75015 Paris, France
2 Virology Department, UCLH NHS Foundation Trust NW1 3AD London, UK
3 Africa Health Research Institute, 4001 KwaZulu-Natal, South Africa
4 Division of Infection and immunity, University College London, WC1E6BT London, UK

Drug resistance mutations emerge in genetic sequences of HIV through selective pressure during antiretroviral therapy. Drug resistances can be transmitted and reduce the chances of long-lasting successful treatment. Phylogenetic methods have been used to estimate the parameters shaping the emergence of drug resistance and spread of resistant viruses. In this review we discuss the examples of use of phylogenetic methods in studies of drug resistance mechanisms in HIV.

Keywords. HIV, drug resistance, phylogenetics.

The clinical epidemiology of HIV drug resistance

The clear majority of publications on HIV drug resistance emanate from the resource rich world, including those pertaining to the clinical epidemiology of resistance. Whilst there remain important lessons to be drawn to understand the spread of drug resistance in resource limited settings, it is worth comparing these settings in considering drivers of drug resistance (Table 1). We witnessed high levels of resistance in treated and untreated individuals in the 1990's and early 2000's in those settings with access to therapy. To a large extent, this rise was associated with what is now recognised to be suboptimal therapy – limited drug classes, pill burden, toxicities, late initiation of therapy – together with continuing transmission in high risk communities. Since that time, the availability of more than 25 antiretroviral (ARV) drugs across five classes, individualised therapy including the use of resistance testing, and simplified regimens have led to a dramatic reduction in resistance in these settings [1]. Indeed, some predictions at the time of ever increasing levels of resistance [2] have not been bourne out [3]. By contrast, we are observing the opposite phenomenon in resource limited settings, where the burden of infection is greatest [4]. Table 1 identifies some of the drivers of such high levels of resistance.

How can we better understand this phenomenon, and develop tools for predicting future trends? The overall population burden of resistance is contributed to both by the emergence of resistance in treated individuals, as well as by transmission of resistance. It is self-evident that the dynamics of the epidemic itself must be considered in modelling future spread of resistance – in other words, the proportion of infected individuals diagnosed and receiving treatment, as well as the ongoing incidence of infection must be considered. From an overall health burden and policy perspective, there is a big difference between a transmitted drug resistance (TDR) rate of 15% within a setting of population HIV incidence of 2%, compared to a TDR rate of 5% in a population with 6% incidence. This contrast is exemplified in modelling approach undertaken by Phillips and colleagues, which addresses the likely impact of a widespread HIV testing and treatment strategy within the South African epidemic. Based on a 2012 prevalence of TDR of $< 10\%$, their model suggests that over a 20-year period of such a test and treat strategy, overall incidence of infection would be reduced by 50%. Nevertheless, by that time, up to 30% of new infections would be with drug resistant virus [5]. For this reason, programmes on surveillance of drug resistance need to be placed into a wider clinical epidemiology of the epidemic in question.

It is also important to consider the developing use of antiretrovirals for pre-exposure prophylaxis (PrEP). Following the PROUD and IPER-GAY study results [6, 7], there is a strong push for rollout of PrEP within high risk populations in resource

Received 00 00 0000; accepted 00 00 0000.
Potential conflicts of interest: none reported.
Reprints or correspondence: Prof Deenan Pillay Division of Infection and immunity, University College London, WC1E6BT London, UK (dpillay@ahri.org).
Word count: abstract: 69; main text: 1965.

The Journal of Infectious Diseases 0000;00(00):0–3
DOI: 10.1093/infdis/xxxxx

poor settings. The first case of PrEP failure due to resistance has now been reported [8]. Abbas et al. [9] modelled the potential impact of preexposure prophylaxis (PrEP) on HIV transmission and drug resistance in South Africa. They predicted that combined ART + PrEP over 10 years would reduce the number of infections (35Supervie et al. performed two modeling studies on rolling out of PrEP: in San Francisco (i.e., in a resource-rich country) [10] and in Botswana (resource-limited) [11]. They showed that if PrEP is widely used in a "high-risk" community in San Francisco the number of infections as well as the number of transmitted ART resistance is likely to decrease (if risk behavior does not increase significantly). In contrast, the introduction of PrEP interventions in Botswana is likely to lead to increase of transmitted ART (while decreasing the overall number of infections). This occurs because the level of ambient resistance is higher in San Francisco than in Botswana due to a longer treatment history. The differences in the results obtained in the studies by Abbas et al. and by Supervie et al. draw our attention to the importance of taking into account the assumptions that are made, e.g. the initial levels of resistance when the rollout begins.

Several studies have utilised phylogenetics together with detailed clinical and epidemiological data to explore the origin of incident infections. Fisher et al. [12] demonstrated that up to 30% of new infections were from individuals in the highly infectious primary stage of infection. Brenner et al. [13] used phylogenetic clustering analysis of Quebec HIV-infected population to show that early infections may account for a major proportion of onward transmissions. This approach was expanded to the ATHENA cohort in the Netherlands [14] to show that both primary and undiagnosed infections together accounted for the bulk of new infections. By contrast, few transmissions came from those in care and on antiretroviral therapy. However, the incidence of transmissions from treated patients bearing not yet detected resistances due to poor monitoring (a typical situation in developing countries) remain to be estimated.

Against this background, what is the potential role of phylogenetics in enhancing our understanding of emergence and spread of drug resistance? Firstly, who are the main transmitters of drug resistance, and are they receiving antiretroviral therapy or not? Secondly, what is the contribution of transmission during acute infection to spread of drug resistance? Thirdly, what is the persistence of drug resistant strains of virus within the population? Lastly, as PrEP becomes widespread, can we identify emergence and transmission of resistant strains from those infected whilst receiving PrEP?

Table 1. Drivers of high levels of resistance.

Characteristics	Resource rich countries	Resource limited countries	Impact on population drug resistance in resource limited settings
Calendar time of ARV availability	1980's	2000's	↓
Treatment paradigm from time of ARV availability	Mono-, to dual- to triple therapy	Triple therapy	↓
Availability of second and third line regimens	Yes	No	↑
Single dose NVP for PMTCT	No	Yes	↑
VL monitoring availability	Extensive	Limited	↑
Incidence and prevalence	Low	High	↑

ARV, Antiretroviral drugs; NVP, Nevirapine (Viramune); PMTCTC, Prevention of mother-to-child transmission; VL, viral load.

Phylogenetics and drug resistance

HIV viruses rapidly accumulate genetic variation because of short generation times and high mutation rates. Phylogenetic inference methods use these variations for reconstruction of *phylogenies* (phylogenetic trees) from contemporary sequencing data. The root of the tree represents the ancestral lineage, and the tips correspond to the virus sequences at the moment of sampling. Going from the root to the tips corresponds to moving forward in time. When a lineage splits (speciation), it is represented as a branching node of the phylogeny. When the sampling is dense such a split can be interpreted as a virus transmission infecting a new individual, and the whole tree is an approximation of the transmission tree [15].

To access the robustness of the reconstructed tree the *support* values on its branches can be calculated using statistical methods, such as bootstrap [16]. These values tend to decrease when going back in history, from tips to the root. In order to remove the uncertain data from the study, often *genetic clusters* are used instead of the whole tree. Such clusters correspond to the well-supported subtrees that contain sequences closely related to each other and distant from the rest of the tree (see [17] for an overview of genetic clustering methods). A cluster of sequences that also share a common trait values (e.g. geographic location, risk group, presence of a given resistance mutation) is called a *phylotype* [18]. The branch lengths in genetic clusters are typically short, and therefore a cluster can be interpreted as representing a recent outbreak, as for example, in a situation when a virus acquires a DRM under drug selective pressure and the patient starts transmitting the resistant virus. The subtree including this patient, individuals infected by him/her, and those

infected by them, would form a resistance cluster if they are sampled before their virus strains diverge significantly. The root of the cluster would correspond to the first transmission event.

Viral phylodynamics is defined as the study of how epidemiological, immunological, and evolutionary processes act and potentially interact to shape viral phylogenies [19, 20]. Phylodynamics methods have been used to estimate the parameters shaping the emergence of drug resistance and spread of resistant viruses, such as, for example, the persistence time of drug resistance mutations (DRMs) in the untreated population.

Wensing et al. [21] used phylogenetic reconstruction and genetic clustering to study the persistence of DRMs in HIV-infected treatment-naïve patients from 19 countries across Europe. They found a significant difference in the level of baseline resistance between recently infected patients (13.5%) and patients infected for more than one year (8.7%).

The origin of transmitted drug resistance has been addressed by several groups. Yerly et al. [22] reconstructed HIV transmission clusters in Geneva using phylogenetic analysis, and showed that newly diagnosed HIV infections are a significant source of onward transmission, notably of resistant strains. Audelin et al. [23] studied TDR among newly diagnosed HIV-1 individuals in Denmark, and concluded that TDR isolates mostly originate from patients failing therapy. The same conclusion was reached by Lewis et al. [24] using $\approx 2,000$ patients from London, predominantly men who have sex with men (MSM), using a similar transmission-cluster-based approach.

Hué et al. [25], and later Mourad et al. [26] obtained different results while studying HIV-1 transmission in the UK. Hué et al. studied treatment-independent viral clusters with DRMs and demonstrated that sustainable reservoirs of resistance persist in the HIV-1-infected population through continuous transmission of resistant viruses among treatment-naïve individuals. Mourad et al. used a parsimony-based approach [27] to extract phylotypes of sequences, the most recent common ancestor of which was bearing a resistant mutation that is still shared by the majority of the sequences in the phylotype. Once dated and combined with the treatment-naïve/experienced status of those represented by the sequences, these phylotypes were used to zoom on the most readable parts of the phylogeny and compute simple statistics which are immediately accessible from the annotated tree; for example, the number of naïve-to-naïve transmissions of DRMs, or the fraction of extant sequences having lost the ancestral resistance. The simplicity of the method makes it computationally very efficient. It was applied to a large set of $\approx 25,000$ HIV-1 subtype B sequences from the UK, where it showed that around 70% of transmitted drug-resistance had a treatment-naïve source. In this population, the most commonly transmitted mutations were L90M in the protease gene and K103N, T215D and T215S in reverse transcriptase. Moreover, reversion to wild type occurred at a low frequency and drug-independent reservoirs of resistance have persisted for up to 13 years.

These conclusions are very close to those of Drescher et al. [28] who studied the transmission of resistances among MSM in the Swiss HIV Cohort. Their method was different as they did not reconstruct the ancestral resistance status of the sequences; but they also extracted well supported transmission clusters from a large sequence phylogeny, and searched for the potential sources of the resistances observed in these clusters.

The discrepancy between the results obtained by Mourad et al. [26] and Drescher et al. [28], and those obtained by Audelin et al. [23] and Lewis et al. [24], is most likely attributable to the size of the data sets, from $\approx 2,000$ in [24] published in 2008, to $\approx 25,000$ in [26] published in 2015. Moreover, the sampling density is of prime importance ($> 50\%$ in [26] and [28]), because to demonstrate naïve-to-naïve TDR relatively large resistance clusters with no or little missing data are needed. When the ratio of missing data is high, it is not possible to conclude on the origin of the transmission for isolated drug-naïve patients harbouring DRMs.

Conclusions

In summary, we argue for building phylogenetics into a more detailed epidemiological surveillance of HIV drug resistance. With an ever reducing cost of genetic sequencing, there is a move to generate full length HIV sequences [29]. This has the capacity to increase the phylogenetic resolution due to a longer sequence length. Through a large simulated dataset, we have shown that the accuracy of trees was nearly proportional to the length of sequences,with gag-pol-env datasets showing best performance compared to the partial pol sequences commonly created through drug resistance testing [30]. An added advantage of extended sequencing is the ability to capture integrase inhibitor resistance. Care must be taken in the sampling frame in the context of HIV prevalence, to produce realistic estimates. This will facilitate a better understanding of the drivers of resistance spread, the source of transmitted resistance, and how this is changing over time in the face of antiretroviral rollout.

Funding

D.P. was supported by Wellcome Trust and Bill and Melinda Gates Foundation.

References

1. Scherrer AU, von Wyl V, Yang WL, Kouyos RD, Böni J, Yerly S, Klimkait T, Aubert V, Cavassini M, Battegay M, Furrer H, Calmy A, Vernazza P, Bernasconi E, Günthard HF, Swiss HIV Cohort Study, Swiss HIV Cohort Study. Emergence of Acquired HIV-1 Drug Resistance Almost Stopped in Switzerland: A 15-Year Prospective Cohort Analysis. Clinical Infectious Diseases **2016**; 62(10):1310–1317.
2. Smith RJ, Okano JT, Kahn JS, Bodine EN, Blower S. Evolutionary dynamics of complex networks of HIV drug-resistant strains: the case of san francisco. Science **2010**; 327(5966):697–701.
3. Jain V, Liegler T, Vittinghoff E, Hartogensis W, Bacchetti P, Poole L, Loeb L, Pilcher CD, Grant RM, Deeks SG, Hecht FM. Transmitted

drug resistance in persons with Acute/Early HIV-1 in san francisco, 2002-2009. PLoS One **2010**; 5(12):e15510.

4. Villabona-Arenas CJ, Vidal N, Guichet E, Serrano L, Delaporte E, Gascuel O, Peeters M. In-depth analysis of HIV-1 drug resistance mutations in HIV-infected individuals failing first-line regimens in West and Central Africa. AIDS **2016**; :1.
5. Cambiano V, Bertagnolio S, Jordan MR, Pillay D, Perriëns JH, Venter F, Lundgren J, Phillips A. Predicted levels of HIV drug resistance. Aids **2014**; 28(Suppl 1):S15–S23.
6. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, Sullivan AK, Clarke A, Reeves I, Schembri G, Mackie N, Bowman C, Lacey CJ, Apea V, Brady M, Fox J, Taylor S, Antonucci S, Khoo SH, Rooney J, Nardone A, Fisher M, McOwan A, Phillips AN, Johnson AM, Gazzard B, Gill ON. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. The Lancet **2016**; 387(10013):53–60.
7. Molina JM, Capitant C, Spire B, Pialoux G, Cotte L, Charreau I, Tremblay C, Le Gall JM, Cua E, Pasquet A, Raffi F, Pintado C, Chidiac C, Chas J, Charbonneau P, Delaugerre C, Suzan-Monti M, Loze B, Fonsart J, Peytavin G, Cheret A, Timsit J, Girard G, Lorente N, Préau M, Rooney JF, Wainberg MA, Thompson D, Rozenbaum W, Doré V, Marchand L, Simon MC, Etien N, Aboulker JP, Meyer L, Delfraissy JF. On-Demand Preexposure Prophylaxis in Men at High Risk for HIV-1 Infection. The New England journal of medicine **2015**; 373(23):2237–46.
8. Knox DC, Anderson PL, Harrigan PR, Tan DH. Multidrug-Resistant HIV-1 Infection despite Preexposure Prophylaxis. New England Journal of Medicine **2017**; 376(5):501–502.
9. Abbas UL, Glaubius R, Mubayi A, Hood G, Mellors JW. Antiretroviral therapy and pre-exposure prophylaxis: combined impact on HIV transmission and drug resistance in south africa. J Infect Dis **2013**; 208(2):224–234.
10. Supervie V, Garcia-Lerma JG, Heneine W, Blower S. HIV, transmitted drug resistance, and the paradox of preexposure prophylaxis. Proceedings of the National Academy of Sciences **2010**; 107(27):12381–12386.
11. Supervie V, Barrett M, Kahn JS, Musuka G, Moeti TL, Busang L, Busang L, Blower S. Modeling dynamic interactions between pre-exposure prophylaxis interventions & treatment programs: predicting HIV transmission & resistance. Sci Rep **2011**; 1:185.
12. Fisher M, Pao D, Brown AE, Sudarshi D, Gill ON, Cane P, Bucktom AJ, Parry JV, Johnson AM, Sabin C, Pillay D. Determinants of HIV-1 transmission in men who have sex with men: a combined clinical, epidemiological and phylogenetic approach. AIDS **2010**; 24(11):1739–1747.
13. Brenner BG, Roger M, Routy JP, Moisi D, Ntemgwa M, Matte C, Baril JG, Thomas R, Rouleau D, Bruneau J, Leblanc R, Legault M, Tremblay C, Charest H, Wainberg MA, Quebec Primary HIV Infection Study Group. High rates of forward transmission events after acute/early HIV-1 infection. J Infect Dis **2007**; 195(7):951–959.
14. Ratmann O, van Sighem A, Bezemer D, Gavryushkina A, Jurriaans S, Wensing A, de Wolf F, Reiss P, Fraser C, observational cohort A. Sources of HIV infection among men having sex with men and implications for prevention. Science Translational Medicine **2016**; 8(320):320ra2–320ra2.
15. Leitner T, Escanilla D, Franzén C, Uhlén M, Albert J. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. Proc Natl Acad Sci U S A **1996**; 93(20):10864–10869.
16. Felsenstein J. CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP. Evolution; International Journal of Organic Evolution **1985**; 39(4):783–791.
17. Poon AFY. Impacts and shortcomings of genetic clustering methods for infectious disease outbreaks. Virus Evol **2016**; 2(2):vew031.
18. Chevenet F, Jung M, Peeters M, de Oliveira T, Gascuel O. Searching for virus phylotypes. Bioinformatics (Oxford, England) **2013**; 29(5):561–570.
19. Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, Holmes EC. Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. Science **2004**; 303(5656):327–332.
20. Volz EM, Koelle K, Bedford T, Bhattacharya T, Delaporte E. Viral Phylodynamics. PLoS Computational Biology **2013**; 9(3):e1002947.
21. Wensing AMJ, van de Vijver DA, Angarano G, Asjö B, Balotta C, Boeri E, Camacho R, Chaix ML, Costagliola D, De Luca A, Derdelinckx I, Grossman Z, Hamouda O, Hatzakis A, Hemmer R, Hoepelman A, Horban A, Korn K, Kücherer C, Leitner T, Loveday C, MacRae E, Maljkovic I, de Mendoza C, Meyer L, Nielsen C, Op de Coul EL, Ormaasen V, Paraskevis D, Perrin L, Puchhammer-Stöckl E, Ruiz L, Salminen M, Schmit JC, Schneider F, Schuurman R, Soriano V, Stanczak G, Stanojevic M, Vandamme AM, Van Laethem K, Violin M, Wilbe K, Yerly S, Zazzi M, Boucher CA, SPREAD Programme. Prevalence of drug-resistant HIV-1 variants in untreated individuals in europe: implications for clinical management. J Infect Dis **2005**; 192(6):958–966.
22. Yerly S, Junier T, Gayet-Ageron A, Amari EBE, von Wyl V, Günthard HF, Hirschel B, Zdobnov E, Kaiser L, Swiss HIV Cohort Study. The impact of transmission clusters on primary drug resistance in newly diagnosed HIV-1 infection. AIDS **2009**; 23(11):1415–1423.
23. Audelin AM, Lohse N, Obel N, Gerstoft J, Jørgensen LB. The incidence rate of HIV type-1 drug resistance in patients on antiretroviral therapy: a nationwide population-based danish cohort study 1999-2005. Antivir Ther **2009**; 14(7):995–1000.
24. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ. Episodic sexual transmission of HIV revealed by molecular phylodynamics. PLoS Med **2008**; 5(3):e50.
25. Hué S, Gifford RJ, Dunn D, Fernhill E, Pillay D, UK Collaborative Group on HIV Drug Resistance. Demonstration of sustained drug-resistant human immunodeficiency virus type 1 lineages circulating among treatment-naïve individuals. J Virol **2009**; 83(6):2645–2654.
26. Mourad R, Chevennet F, Dunn DT, Fearnhill E, Delpech V, Asboe D, Gascuel O, Hue S, UK HIV Drug Resistance Database & the Collaborative HIV, Anti-HIV Drug Resistance Network. A phylotype-based analysis highlights the role of drug-naive HIV-positive individuals in the transmission of antiretroviral resistance in the UK. AIDS **2015**; 29(15):1917–1925.
27. Fitch WM. Toward defining the course of evolution: Minimum change for a specific tree topology. Syst Biol **1971**; 20(4):406–416.
28. Drescher SM, von Wyl V, Yang WL, Böni J, Yerly S, Shah C, Aubert V, Klimkait T, Taffé P, Furrer H, Battegay M, Ambrosioni J, Cavassini M, Bernasconi E, Vernazza PL, Ledergerber B, Günthard HF, Kouyos RD, Swiss HIV Cohort Study. Treatment-naive individuals are the major source of transmitted HIV-1 drug resistance in men who have sex with men in the swiss HIV cohort study. Clin Infect Dis **2014**; 58(2):285–294.
29. Pillay D, Herbeck J, Cohen MS, de Oliveira T, Fraser C, Ratmann O, Brown AL, Kellam P. PANGEA-HIV: Phylogenetics for generalised epidemics in Africa. The Lancet Infectious Diseases **2015**; 15(3):259–261.
30. Yebra G, Hodcroft EB, Ragonnet-Cronin ML, Pillay D, Brown AJL, Consortium P, Fraser C, Kellam P, de Oliveira T, Dennis A, Hoppe A, Kityo C, Frampton D, Ssemwanga D, Tanser F, Keshani J, Lingappa J, Herbeck J, Wawer M, Essex M, Cohen MS, Paton N, Ratmann O, Kaleebu P, Hayes R, Fidler S, Quinn T, Novitsky V, Project I, Haywards A, Nastouli E, Morris S, Clark D, Kozlakidis Z. Using nearly full-genome HIV sequence data improves phylogeny reconstruction in a simulated epidemic. Scientific Reports **2016**; 6:srep39489.