Ethical Considerations in Global HIV Phylogenetic Research

Cordelia E. M. Coltart[*1], Anne Hoppe[*2], Michael Parker[3], Liza Dawson[4], Joseph J. Amon[5], Musonda Simwinga[6], Gail Geller[7,8], Gail Henderson[9], Oliver Laeyendecker[10], Joseph D. Tucker[11], Patrick Eba[12,13], Vladimir Novitsky[14], Anne-Mieke Vandamme[15,16], Janet Seeley[17,18], Gina Dallabetta[19], Guy Harling[1,17], M. Kate Grabowski[20,21], Peter Godfrey-Faussett[22,23], Christophe Fraser[24], Myron S. Cohen[#11,25], Deenan Pillay[#2,17]

On behalf of the Ethics in HIV Phylogenetics Working Group¶

[1] Institute for Global Health, University College London, London, UK; [2]Division of Infection and Immunity, University College London, London, UK; [3]The Wellcome Centre for Ethics and Humanities (Ethox), Nuffield Department of Population Health, University of Oxford, Oxford, UK; [4]Division of AIDS, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, US; [5]Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ, US; [6]Zambart, Lusaka, ZM; [7]Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD, US; [8]School of Medicine, Johns Hopkins University, Baltimore, MD, US; [9]Center for Genomics and Society, Department of Social Medicine, University of North Carolina, Chapel Hill, NC, US; [10]National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, US; [11]Institute for Global Health and Infectious Diseases, University of North Carolina, Chapel Hill, NC, US; [12]Community Support, Social Justice and Inclusion Department, Joint United Nations Programme on HIV/AIDS, Geneva, CH; [13]School of Law, University of KwaZulu-Natal, Pietermaritzburg, ZA; [14]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, US; [15]KU Leuven - University of Leuven, Department of Microbiology and Immunology, Rega Institute for Medical Research, Clinical and Epidemiological Virology, Leuven, BE; [16]Center for Global Health and Tropical Medicine, Unidade de Microbiologia, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisbon, PT; [17]Africa Health Research Institute, KwaZulu-Natal,

ZA; [18]Department of Global Health and Development, London School of Hygiene & Tropical Medicine, London, UK; [19]Bill & Melinda Gates Foundation, Washington, DC, US; [20]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; [21]Department of Rakai Community Cohort Study, Rakai Health Sciences Program, Kalisizo, Uganda; [22]Joint United Nations Programme on HIV/AIDS, Geneva, CH; [23]Department of Clinical Research, London School of Hygiene & Tropical Medicine, London, UK; [24]Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK; [25]Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, US

*Joint first authorship

Joint senior authorship

¶Members of the Ethics of HIV Phylogenetics Working Group are listed in the Supplementary Materials

Abstract: 150 words

Manuscript: 5001 words

References: 75

Boxes: 5

Figures: 2

Corresponding Author

Dr Anne Hoppe, Division of Infection & Immunity, UCL, Cruciform Building, 90 Gower Street, London, WC1E 6BT, United Kingdom;

Tel: +44 7824700204; Email: a.hoppe@ucl.ac.uk

Abstract

Phylogenetics applied to pathogen genetics is an increasingly powerful tool to help reduce the spread of epidemics, including HIV. As a result, phylogenetic approaches are becoming embedded in public health and research programmes, as well as outbreak responses. This presents unique ethical, legal and social issues which are not addressed adequately by the existing bioethics literature.

A multidisciplinary working group was formed to explore the ethical issues arising from the design, conduct and use of results from HIV phylogenetic studies, and to propose recommendations to minimise the associated risks both to individuals and groups.

We identified eight critical ethical domains within which we highlighted factors that make HIV phylogenetic research unique. We also endeavoured to provide a framework to assist researchers, public health practitioners, and funding institutions to ensure that HIV phylogenetic studies are designed, conducted and disseminated in an ethical manner. Our findings have broader relevance for pathogen phylogenetics.

Introduction

Understanding the transmission dynamics of infectious agents is critical to developing effective public health interventions. Historically, this was done using epidemiological tracking of the evolution of epidemics through time, place and person, based primarily on observation and self-reports of exposure and risk behaviours. However, despite advances in HIV prevention, HIV incidence remains high, notably in sub-Saharan Africa, which accounts for 75% of all new HIV infections worldwide (1). In this context, understanding who is most likely to infect whom remains an important consideration in the development of targeted prevention strategies.

In phylogenetic analysis historical relationships between individuals or groups are deduced by comparing pathogen genomes, establishing how closely related viruses from two individuals are. In combination with traditional epidemiological data, these viral genetic sequence data can be thus help to infer transmissions patterns. The combination of phylogenetics and traditional epidemiology has demonstrated the potential to answer critical questions that are not easily addressed by traditional or molecular approaches alone (2-4).

In recent years, the focus of HIV phylogenetic studies has extended from concentrated to generalised epidemics, and increasingly involves sizeable datasets (5). Funding bodies such as the Wellcome Trust, the Bill & Melinda Gates Foundation, and the National Institutes of Health are committed to sharing such datasets to maximise the benefit of HIV research. Additionally, sequence data analysed for publications in scientific journals are usually required to be submitted to open public sequence repositories. The collection, storage, sharing and research use of such data raises important ethical, legal and social challenges.

International ethical guidelines for research with human participants, such as the Helsinki Declaration and the Council for International Organizations of Medical Sciences (CIOMS) guidelines

address a number of these issues, including the need for informed consent, community engagement, risk minimization, and consideration of the risks and benefits of research for groups and communities (6, 7). In addition, a large and diverse academic and policy literature exists addressing the ethical, legal and social implications (ELSI) of the research and clinical uses of human genomics in high-income countries (HIC)(8). In recent years, this has been accompanied by a growing bioethics and social science literature on the implications of genomic research in low- and middle-income countries (LMIC)(9-10), and by stakeholder engagement initiatives in these settings (11-13). The key documents, position statements and initiatives are summarized in Supplementary Table 1.

HIV phylogenetic research presents complex ethical issues, including two specific challenges. Firstly, (like contact tracing data) phylogenetic analyses are fundamentally relational: analysis of data from one person may impact other people, for example by identifying them as potential sources of infection. Secondly, as sequence data become richer (through next generation sequencing (NGS)), true anonymization of viral sequence data becomes difficult, since virus isolated from another time-point in another study could be used to re-identify an individual. In recognition of this, and of the need for the development of models of good ethical practice in this area, we held a multidisciplinary (scientists, bioethicists, lawyers, human rights advocates, HIV activists and community engagement members from Africa) workshop in London in May 2017 (15). This meeting focused both on identifying the critical issues arising from designing, conducting, or using the results of HIV phylogenetic studies, and on making recommendations regarding the public release and publishing of data obtained from HIV phylogenetic studies in an ethical manner.

This review article summarises the findings and recommendations both from the workshop and follow-up discussions, and sets out a framework for both researchers and funding bodies undertaking HIV genetic studies. This also has relevance for the increasing use of phylogenetics for non-HIV pathogens, including within outbreak response situations.

Phylogenetics and its role in HIV research

Over time and successive generations, mutations occur in the genetic code of species. Phylogenetic inference exploits these changes to determine the genetic similarity of two organisms, assuming the more similar their genetic sequence, the closer in time they are to having a common ancestor. Applying this approach to the HIV detected in blood samples from infected human populations helps to understand HIV transmission patterns.

HIV lends itself to phylogenetic analysis as it is highly genetically variable (16). The transmission of HIV involves two individuals (a couple) and the variability of HIV is used to infer linkages forming phylogenetic clusters between couples and groups of people (17). In addition, it is possible in many cases to infer, with a degree of uncertainty, the direction of transmission within clusters, either by using additional epidemiological data (18, 19), or by using data with sequences from multiple viruses sampled from each individual (20, 21).

There are different techniques to generate the sequence data required for phylogenetic analysis. NGS increases both the potential power and potential risks of phylogenetic approaches compared to conventional Sanger sequencing methods (22, 23), as it provides information on intra-host variation, and returns richer sequence data on multiple viral particles per sample. Additionally, there are many methods and associated assumptions in phylogenetic analyses and a number of methodological considerations for identifying clusters of genetically related viruses through phylogenetic analyses (Supplementary Box 1). Phylogenetic clusters are generally thought to represent groups of infected persons closer together in a transmission chain and can be identified via different methods (3, 24-28).

Caution is required when interpreting phylogenetic clusters for epidemiological purposes. This is because clusters are typically inferred from partially sampled transmission chains, i.e. some infected

individuals were not sampled. Unsampled cases may act as either a common source of infection or as an intermediary in a transmission chain for hosts infected with genetically similar pathogens. Whilst historically it has been difficult clearly to prove transmission between two individuals, the use of NGS together with improved interpretation algorithms makes such inferences more likely in future.

Phylogenetic analyses can be used widely in HIV epidemiology (see Figure 1 for details). For example, they can be used to study viral linkage and risk factors for epidemic spread (molecular epidemiology) (4), the growth/decline of the HIV epidemic (phylodynamics) (3, 29, 30), or the impact of migration on HIV spread and the identification of hubs of transmission (phylogeography)(31).

HIV phylogenetics has been most applied in HIC with greater scientific infrastructure, where HIV is characterised by smaller epidemics, focused on specific risk groups ("concentrated epidemics"). In many HIC, sequencing of the HIV *pol* gene is used to monitor both transmitted drug resistance at time of diagnosis and emerging drug resistance on antiretroviral therapy (ART)(32). This has led to the growth of national HIV genetic databases, such as in the UK (33) and Switzerland (34). If such datasets are linked to epidemiological surveillance and clinical cohort data, inferences may be made with regard to patterns of exposure and risk factors for onwards transmission amongst infected individuals (18, 29, 35). Unlike standard epidemiological data, molecular data can also allow inferences to be made from the time of transmission relative to the time of sample collection. Furthermore, the data obtained through phylogenetic analyses can be used to validate self-reported epidemiological data in relation to sexual and other behaviours. Combined with traditional epidemiological tools, phylogenetic research provides more detailed and precise understanding of epidemic characteristics, thereby enabling improved public health policies, including more effective and better targeted programs for prevention and treatment (36, 37).

Many African epidemics are much larger than those in the US and Europe. The sequencing of virus in Africa is currently not routinely undertaken outside of targeted programmes such as the World Health Organisation's HIV drug resistance surveillance (38) and research projects. Whilst declining cost and progressively easier sequencing will increase the proportion of infected individuals represented within sequence databases, such as PANGEA-HIV (5), Africa does not yet have the extensive and comprehensive datasets seen in HIC. In addition, community and patient mobilisation around HIV takes very different forms compared to that in HIC, and the social, political and economic context is significantly different and varies among African countries. Ethical analysis of phylogenetic work will therefore need to take account of international variation in both epidemic characteristics and local economic, legal and social contexts.

Key ethical issues arising in phylogenetic studies of HIV transmission

Some of the ethical issues raised by HIV phylogenetic research are similar to those in traditional epidemiology studies. These include the potential for stigmatization and risk of social harm to individuals or groups, and concerns about privacy, confidentiality and security of data. However, there are risks that are particularly salient in phylogenetic research, which we discuss over eight critical ethical domains.

(i) Risk and benefit assessments

The harms and benefits of phylogenetic research will vary depending on whether they are assessed at the individual, group or societal level (Figure 2). Information obtained through phylogenetic analyses should be used to advance socially valuable goals, such as reducing the spread of HIV, whilst at the same time minimising the risks to individuals, groups, and populations. Of particular concern with phylogenetics is that complex social and sexual relationships may be deduced by adding minimal clinical and demographic information. In contrast, traditional epidemiological studies

would require far more information in order to draw inferences about transmission of HIV between individuals, particularly with respect to directionality of infection.

Risks to individuals principally arise either from inadvertent or intentional disclosure of HIV status or transmission events, or from demands for these data for judicial or extra-judicial targeting of individuals or groups. In a number of countries, phylogenetic evidence is being used in criminal cases of alleged HIV transmission (39-42). Breaches of confidentiality could occur through inadequate anonymization or deductive disclosure, through misinterpretation, miscommunication, or misuse of the analytic results, or through legal action.

These risks will increase if more data is generated and made publicly available, a requirement of many funding agencies and publishers. While publication might maximize the scientific research value of a dataset, it raises concerns about how the data are used, appropriate consent for such use, confidentiality, and stigma. Furthermore, contrary to epidemiological studies where individuals can choose what information they disclose to investigators, inferences made from viral genetic sequences are not controlled by participants.

Anonymization may provide some protection to individuals. However, even with anonymization, deductive disclosure of identities from HIV sequence and other corresponding data remains theoretically possible. Through the use of rich NGS data applied at successive time-points, a virus sequenced from an individual could in principle be used to re-link that individual to an earlier study with high reliability. Furthermore, small fragments of human DNA sequence contained in NGS data could be accidentally released. In addition, human leukocyte antigen of the infected individual is imprinted on the virus due to immune selection, which may assist individual identification in the future (43). The probability of re-linking individuals to anonymised data can be minimised by

processing sequence data prior to release, for example by only including consensus sequences, which suffice for many phylogenetic methods.

Notwithstanding these risks, maintaining a link to individuals' identities may allow for direct benefits to individuals. Sequence information can provide clinical guidance, for example, by allowing treatment optimization following detection of drug resistance mutations. Indeed, most HIV phylogenetic studies to date have used data obtained for clinical drug resistance testing, from resistance surveillance programmes, or as part of broader research studies.

At a population level, phylogenetic analysis can allow individuals' data to be linked to one-another in a network, enabling inference about the characteristics of networks and identification of risk groups. This information could be used to focus public health interventions towards specific groups at high risk of both acquiring and transmitting the infection.

The choice of meta-data variables used in phylogenetic analysis is an important ethical decision. Phylogenetic analyses are often based upon individual-level demographic, behavioural or clinical variables, ignoring structural and environmental factors. Focusing only on these variables may reinforce the perception that certain groups (for example key populations such as men who have sex with men (MSM) or people who inject drugs (PWID)) are responsible for infecting others and sustaining the HIV epidemic. In contrast, other structural factors, such as those highlighted in the case study of migration in Botswana (see Box 1), as well as sexual violence, lack of access to prevention and treatment, and having experienced discrimination, may play a significant role in HIV transmission (44). Studying these factors and their effect on HIV transmission risk can decrease the "blaming mentality" and create alternative understanding of how to reduce HIV transmission, and which individuals or groups are most at risk and why.

Plans for addressing risks to individuals and to groups should be developed in the planning stages of research projects. For protection of individuals, particularly the risk of criminal prosecution or other targeting based on either HIV status or HIV transmission events, anonymization of data provides considerable protection. While it is theoretically possible that individuals could be identified through re-analyzing and re-linking anonymized data from different sources, it would be difficult and require specialized expertise. In contrast, datasets with individual identifiers still linked could be subpoenaed or obtained through unauthorized means, putting individuals at risk.

Researchers, therefore, need to assess carefully the potential of identifying specific individuals or groups of people from their data, whether this identification could provide benefit in informing targeted treatment or interventions, and whether these benefits outweigh the risks to individuals and groups by being identified. Preference should be given to other approaches that achieve the same research objective, but involve less risk. Finally, an ongoing monitoring of anticipated and unanticipated risks should be built into HIV phylogenetic research, and mitigation strategies identified as early as possible.

(ii) Protection of the rights and interests of study participants while in pursuit of scientific progress and improvements to public health

Effective phylogenetic work often takes place at the interface between research and public health practice: the same data can be used for both purposes. Researchers are typically viewed as obliged, so far as possible, to protect individuals who enrol in a study from risk of harm while pursuing valuable knowledge. In contrast, public health agencies have the mission of protecting the health of the public, which sometimes involve overruling individuals' privacy interests to use data for public health decision making. Where research also has implications for specific population groups, further considerations relating to group harm are important (50-52). Some research and clinical challenges in phylogenetic analyses are highlighted in Box 2.

The obligation of researchers to communicate results to study participants needs to be evaluated for each phylogenetic study. Where clinical action is required, there is an obligation to make results available. In general, this goal (and benefit) is currently theoretical because phylogenetic results are produced with a significant delay from sampling, and therefore any result would likely no longer be timely in informing clinical care. However, with the evolution of real time phylogenetics, reporting of drug resistance data to study participants may result in changes to clinical management. A second potential issue is the source of HIV acquisition in discordant couples. As outlined in Box 2, it may be critical for interpreting the efficacy of prevention strategies to establish whether HIV acquisition events are linked to the known infected partner. However, disclosure of these results to study participants may result in adverse consequences for the individuals involved (53).

(iii) Local social and legal context, including human rights violations

Understanding the local social and legal context is crucial to understanding the risks and benefits associated with phylogenetic research. For example, whilst certificates of confidentiality are legally binding tools to protect both research participants and researchers from being compelled to reveal personal data in the US, similar protocols do not exist in many other HIC or in LMIC (54). Knowledge of local legal proceedings is therefore essential for ensuring that research data is unavailable to subpoena, and thus dramatically reduce the level of individual risk. Furthermore, changes to the social and legal environment need to be regularly monitored to ensure that new risks are not introduced during the course of the study. Finally, researchers need to be clear with policy-makers how proposed laws or policies could negatively impact HIV research efforts and interventions.

The global human right to health (6, 55) encompasses prevention and treatment, and a right to privacy, consent, freedom from discrimination and violence, yet persecution of key populations in Africa remains widespread (56-58). Research methods have been used to violate individuals' rights,

including the use of key population mapping by police to arrest and harass sex workers and MSM in Nigeria in 2014 (59), impose travel bans on foreigners, enforce restrictions on access to housing, schooling and employment, and trigger violent attacks, including murder. Box 3 highlights key legal and human rights pre-requisites for the use of phylogenetic tools for public health. These are based on norms provided under international human rights treaties as well as national constitutions and legislation.

Globally, 72 countries (a third of them in Africa) have laws specifically allowing for HIV criminalisation (60). Box 4 reviews phylogenetic analysis used in criminal convictions. Government officials or other actors may misinterpret, or wilfully misconstrue, the results of phylogenetic research in support of political agendas or criminal convictions, putting individuals at risk of criminal prosecution for HIV transmission or broader human rights abuses. A realisation from these communities of the possible consequences for privacy and prosecution may lead to a reluctance to test, failure to disclose contacts and/or refusal of resistance testing. There is evidence that these effects have already occurred (61-63). The likelihood of misuse and abuse of these data is high, particularly for stigmatised populations. Researchers should alert ethical review committees and suspend research when risks to study participants increase.

Misuse of phylogenetic data, including seizing and subpoena of such data by police and in criminal proceeding or perceptions by people living with HIV (PLHIV) and members of key populations that phylogenetic data might be misused against them, can undermine trust in research projects and health care systems, thus putting HIV prevention and treatment programmes at risk. Research conducted in countries where privileged information between medical practitioners and their patients can be seized in HIV-related criminal trials showed that PLHIV were more reluctant to speak openly with their practitioners about their sexual partners and practices (63). These risks can be

mitigated in phylogenetic studies by ensuring awareness of and addressing social and legal issues at the planning stages and monitoring these throughout the project.

(iv) Risk mitigation strategies to protect individual and group identities

Many of the risks relating to identification of individuals from phylogenetic information in environments with oppressive laws and policies can be reduced through use of anonymization. Therefore, one default presumption is that if scientific objectives can be accomplished with anonymized data, this is preferable. This default presumes there is no overriding interest in individuals receiving research results at individual level. If the data are not relevant for clinical care, given, for example, significant time delay between sample collection and generation of sequence information, then there is little rationale for returning data to health care workers. Where sequence analysis is timely, resistance data should be returned to clinics before performing phylogenetic analyses on anonymised sequencing data.

If anonymization is significantly detrimental to the scientific objectives or public health, further ethical analysis must be undertaken and specific steps taken to protect the data from use in harmful proceedings. These steps might be technical (storage linkage to identifiers in coded, separate databases with controlled access) and/or legal, for example legal agreements that data will remain protected from disclosure for the duration of the study.

The risk that individuals are de-anonymised by use of later samples, for example infected blood collected at a different time point as part of a linked unrelated study, or as part of a criminal investigation, can be mitigated by restricting the amount of data that is shared. For example, by restricting to one virus sequence per individual person, and keeping raw NGS data under managed access, or destroying raw data. However, this reduces the potential scientific benefits of the study

since it limits the ability to apply newly develop bioinformatic algorithms to infer the direction of transmission, for example.

Risks to groups cannot be addressed through anonymization of individuals. Groups can be placed at risk through characterization as high risk or likely to transmit virus, and these can include geographically defined groups, as well as sexual or gender minorities, those defined by ethnicity, nationality, or migration status. Mitigation plans to address these risks need to include consultation with community representatives, consideration of the public health value of the findings and development of communication plans in formats and venues that are least damaging to vulnerable groups. In some cases, detailed findings might need to be communicated confidentially rather than publicly; and some group descriptors may need to be masked in research publications and press releases. Risk mitigation strategies must also provide for redress mechanisms in cases of abuse or misuse of phylogenetic data. These may require the establishment of ties with local legal services organizations working to protect PLHIV and criminalized or stigmatized populations to ensure that they have access to the means to protect their rights.

Finally, training those involved in phylogenetic research on the potential of harm to communities and individuals is an important risk mitigation strategy. Such training should aim to ensure that research staff are sensitive to the risk of harm and understand key issues of anonymity, confidentiality, informed consent and protection of research participants and communities.

(v) Valid informed consent and other safeguards

The formal requirements for the achievement of valid consent are well-established in literature and guidelines (66). The issues arising in relation to consent for phylogenetic studies are likely to be multifaceted. Obtaining community assent (via community leaders) and individual informed consent is particularly challenging for complex scientific studies such as phylogenetic research, which involve

concepts that are hard to both explain and understand, and have multiple possible risks and benefits.

Due to the complex concepts involved, phylogenetic research may raise fears about the aims of the work and the implications of participation among research participants, frontline research staff, healthcare professionals and ethics committee members. Communication tools which increase the understanding of phylogenetic studies need to be designed and evaluated. These must emphasise the potential harms, thoughtful mitigation of harms to risk groups, processes for monitoring risk, and clear protection procedures to minimise risks. Nevertheless, with ever-advancing technologies, a comprehensive "one size fits all" consent model will be hard to design.

Study participants and patients whose samples are being used for phylogenetic analysis should ideally have consented to such use. However, sequence data generated from drug resistance testing and other surveillance data typically does not include explicit consent to participate in large-scale phylogenetics analyses. Data from previous research studies often entails broad consent for HIV-related research, but rarely involves specific consent for phylogenetics. In such situations, a waiver of specific consent may be obtainable from an ethics committee. Waivers of specific consent are allowable when samples are no longer linked to identifiers, or where consent was given for sample collection for research and storage in future studies, without specific consent for the current research.

Independent review of protocols for phylogenetic is also essential for the protection of research participants. The role of local ethics committees is essential for providing local, independent, representation for research participants and others affected by the research, as well as ensuring that the local context in which researchers and participants are situated is taken into account.

(vi) Community engagement

Community engagement should occur early on in the research design process, ensuring that the phylogenetic research is relevant to participating communities and local perspectives are included in the design and overall conduct of research studies (67, 68). Meaningful community engagement is particularly challenging in research-naïve and low-income communities, and in criminalized or socially marginalized populations. Lack of authentic representative structures, low literacy levels and/or poverty place these communities at risk of being exploited, especially where research involves highly technical elements, such as viral genomics (69, 70).

Nevertheless, this should not limit attempts to maximise engagement. The phylogenetics study team of the PopART study (71) in Zambia has performed extensive community engagement in those communities in which the study takes place. The process involved obtaining community input in the design stages, as well as ongoing consultation and the development of a feedback protocol. Community representatives were consulted on the benefits and risks of informing and sharing results with entire communities, and on measures of how to avoid stigmatization of or within communities.

(vii) Communication

Scientific inferences are based on probabilities. Understanding and communicating uncertainty is key to understanding phylogenetic results: technical complexity or lack of familiarity with methods may easily generate a false sense of accuracy and precision. Researchers performing phylogenetic analysis must ensure that caveats, such as the fact that inferences are always based on probabilities and that methods are based on assumptions, are clearly highlighted in any dissemination, including interviews, publications, oral presentations and posters. It is important to note that probabilities vary; an assignment of 50% to a transmission event is very different of an assignment of 99%; in both cases the analyst will report uncertainty, but the conclusions drawn by most observers will be

different. An ethical framework in an area of rapid technological development should prepare for the possibility that, in some cases at least, probabilistic assignments will likely improve over time.

Mass media campaigns, as well as reporting on social media, television, radio and in newspapers has been a powerful tool, raising awareness about HIV, treatments and prevention, and facilitating public health campaigns aiming to change attitudes and behaviours. However, the way the media frames HIV and reports study outcomes can affect both the long-term and short-term success of any campaigns and may generate unintentional consequences, including a lack of trust in healthcare services (72, 73). Any ambiguous or misleading reporting of phylogenetic studies may reduce HIV testing rates, increase scepticism about participating in studies, and make risk groups less likely to access healthcare. Therefore, it is essential to educate the media, local health care personnel and the community about these studies.

Care must be taken especially in reporting findings relevant to specific population subgroups, including identifiable geographic areas, population groups that may be stigmatized or targeted by government, police, others in the community, or subject to criminal charges. Researchers will need to consider the potential social harms and political impact of findings before deciding exactly what information should be publicly shared or published.

(viii) Equitable data sharing

Largely as a result of the funders' requirements, many anonymised HIV sequences are being made publicly available via GenBank and LosAlamos. This is advantageous for some research studies, such as vaccine development. However, there is a real risk of lack of awareness that every sequence is associated with a patient or study participant. Care must be taken to ensure human DNA sequences are not inadvertently released with NGS data. Routinely publishing only limited information (such as the year of sampling and the country of where the sample are collected) with each sequence would

help minimize risk. Any other anonymised information should be provided via a controlled access protocol which ensures that the research proposed is scientifically valid, does not pose any risks to study participants, and is in line with the informed consent obtained. This would require the development of a clear governance plan.

Finally, different participant information sheets and consent forms may allow for different levels of data sharing, and laws may differ as to how data may be re-used. Any phylogenetic researcher must abide to the levels of sharing outlined in the forms, even if this impacts on the quality of the research conducted.

Conclusions and recommendations:

Phylogenetic analyses, either alone or in combination with linked epidemiological data, is a powerful tool with the potential to help reduce the spread of the HIV epidemic. However, an effective and sustainable model of good ethical practice in phylogenetic research is required to help minimise the risks to individuals/groups participating in studies while optimising the scientific benefits. Whilst a one-model approach to address any ethical issues is impractical given the vast variations in studies and contexts, this review article highlights several themes we believe are essential to consider in order to undertake phylogenetic studies in an ethically responsible manner.

We have clustered the critical issues into eight domains, which provide a framework through which to consider them: (i) risk and benefit assessments; (ii) protection of the rights and interests of study participants; (iii) local social and legal context, including human rights violations; (iv) risk mitigation strategies to protect individual and group identities; (v) valid informed consent and other safeguards; (vi) community engagement; (vii) communication and (viii) equitable data sharing. Our recommendations for each domain, based on our review of the literature and input from experts, are summarized in Box 5.

So far, viral sequencing has been effectively restricted to large, highly regulated laboratories. Emerging DNA sequencing technologies are more powerful in terms of their applications to epidemiology (21), more portable, robust to field conditions, and with lower capital costs (74). HIV phylogenetics (and increasingly for other pathogens) may therefore become more widely distributed across geographic areas and laboratory types, with widely differing regulatory frameworks. Furthermore, as phylogenetics is inherently relational, data are likely to be shared amongst wider and looser networks of investigators. While these developments have the potential to promote rapid scientific advances, they also pose new challenges for governance, enhancing the utility of disseminating clear ethical frameworks and promoting positive social norms.

Any researcher conducting phylogenetic analysis should be aware of the risks such analyses pose and undertake steps to mitigate these risks. This is particularly pertinent in LMIC, which often have weak governance structures and limited laws to protect vulnerable populations. These issues are likely to become more problematic as sequence costs decrease and data become more routinely available. Looking forward, real-time phylogenetics may be used more frequently to direct public health responses and increasingly form the basis for surveillance programmes. Whatever the scenario, the fundamental principle of protecting participating individuals and groups must be central to any study design, implementation and reporting of results.

Authors' contributions

CEMC drafted the original manuscript. Significant contributions to the manuscript were made by AH, CF, DP, GG, GH, LD, MP and MSC. OL produced the content for Figure 2. JDT and team produced the infographics. The content of the manuscript was based on discussions at the Ethics of HIV Phylogenetics meeting and excerpts of meeting abstracts (AMV, FM, JW, JJA, MS, PE, VN) have been incorporated into the manuscript. Members of the Ethics of Phylogenetics work group (AH, AR, CEMC, DG, DP, GG, GD, JS, JDT, MSC, OL, PGF, ZK) organised the meeting. AH led the workgroup with support from DP and MSC, under the auspices of the PANGEA-HIV Consortium (Principal Investigator DP). All authors and working group members reviewed and approved the final version of the manuscript.

Declaration of interests

This work was funded by the Bill & Melinda Gates Foundation and carried out with support from the HIV Prevention Trials Network and the PANGEA-HIV project (Grant code: OPP1084362). CEMC is funded by the Wellcome Trust (Grant code: 106551/Z/14/A). OL received additional support from the Division of Intramural Research NIAID. AMV received funding from Fonds voor Wetenschappelijk Onderzoek – Flanders (FWO) grant G.0692.14, and the VIROGENESIS project receives funding from the European Union's Horizon 2020 Research and Innovation Programme (under Grant Agreement no. 634650). GD is employed by the Bill & Melinda Gates Foundation. LD is employed by the US National Institutes of Health. DP is funded by the Wellcome Trust. No other funding sources or representatives thereof had any role in the writing of the manuscript or the decision to submit it for publication. None of the authors has been paid to write this article by a pharmaceutical company or other agency. The views expressed in the article do not represent any policy or position of the US Department of Health and Human Services or any of its components.

Acknowledgements

We would like to thank Shufang Wei for producing the infographics of this paper, Habiba Cooper Diallo and Cait Breeden for their assistance in organising the Ethics of HIV Phylogenetics meeting, Anne Johnson for helpful discussions and commenting on the manuscript, and the PANGEA-HIV Steering Committee (Myron Cohen, Tulio De Oliveira, Ann Dennis, Max Essex, Sarah Fidler, Dan Frampton, Christophe Fraser, Richard Hayes, Joshua Herbeck, Anne Hoppe, Pontiano Kaleebu, Paul Kellam, Cissy Kityo, Andrew Leigh-Brown, Jairam Lingappa, Vladimir Novitsky, Nicholas Paton, Thomas Quinn, Deenan Pillay, Oliver Ratmann, Deogratius Ssemwanga, Frank Tanser and Maria Wawer) for their support.

References

1. UNAIDS. AIDS by numbers 2015. Geneva: UNAIDS, 2015.
2. Dennis AM, Herbeck JT, Leigh Brown A, Kellam P, de Oliveira T, Pillay D, et al. Phylogenetic studies of transmission dynamics in generalized HIV epidemics: An essential tool where the burden is greatest? J Acquir Immune Defic Syndr. 2014;67(2):181-95.
3. Grabowski MK, Lessler J, Redd AD, Kagaayi J, Laeyendecker O, Ndyanabo A, et al. The role of viral introductions in sustaining community-based HIV epidemics in rural Uganda: evidence from spatial clustering, phylogenetics, and egocentric transmission models. PLoS Med. 2014;11(3):e1001610.
4. De Oliveira T, Kharsany AB, Graf T, Cawood C, Khanyile D, Grobler A, et al. Transmission networks and risk of HIV infection in KwaZulu-Natal, South Africa: a community-wide phylogenetic study. Lancet HIV. 2017;4(1):e41-e50.
5. Pillay D, Herbeck JT, Cohen MS, de Oliveira T, Fraser C, Leigh Brown A, et al. PANGEA-HIV: phylogenetics for generalised epidemics in Africa. Lancet Infect Dis. 2015; 15(3):259-61.
6. Association WM. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. 2013. Available from: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ (accessed Mar 28, 2018)
7. WHO CfIOoMSCa. International Ethical Guidelines for health related research involving humans. Geneva: 2016. Available from: https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf (accessed Mar 28, 2018)
8. Geller G, Dvoskin R, Thio CL, Duggal P, Lewis MH, Bailey TC, et al. Genomics and infectious disease: a call to identify the ethical, legal and social implications for public health and clinical practice. Genome Med. 2014;6(11):106.
9. Parker M, Kwiatkowski DP. The ethics of sustainable genomic research in Africa. Genome Biol. 2016;17:44.

10. De Vries J, Munung SN, Matimba A, McCurdy S, Ouwe Missi Oukem-Boyer O, Staunton C, et al. Regulation of genomic and biobanking research in Africa: a content analysis of ethics guidelines, policies and procedures from 22 African countries. BMC Med Ethics. 2017;18(1):8.
11. Wellcome_Trust. Ethical sharing of health research data in low- and middle-income countries. 2014. Available from: https://wellcome.ac.uk/sites/default/files/ethical-sharing-of-health-research-data-in-low-and-middle-income-countries-phrdf-2014.pdf.
12. H3_Africa. [cited 2017 July]. Available from: https://h3africa.org/about/ethics-and-governance.
13. De Vries J, Tindana P, Littler K, Ramsay M, Rotimi C, Abayomi A, et al. The H3Africa policy framework: negotiating fairness in genomics. Trends Genet. 2015;31(3):117-9.
14. Kaye J, Meslin EM, Knoppers BM, Juengst ET, Deschenes M, Cambon-Thomsen A, et al. Research priorities. ELSI 2.0 for genomics and society. Science (New York, NY). 2012;336(6082):673-4.
15. Cohen J. Pinpointing HIV spread in Africa poses risks. Science (New York, NY). 2017;356(6338):568-9.
16. Batorsky R, Sergeev RA, Rouzine IM. The route of HIV escape from immune response targeting multiple sites is determined by the cost-benefit trade off of escape mutations. PLoS Comput Biol. 2014;10(10):e1003878.
17. Grabowski MK, Redd AD. Molecular tools for studying HIV transmission in sexual networks. Curr Opin HIV AIDS. 2014;9(2):126-33.
18. Ratmann O, van Sighem A, Bezemer D, Gavryushkina A, Jurriaans S, Wensing A, et al. Sources of HIV infection among men having sex with men and implications for prevention. Science translational medicine. 2016;8(320):320ra2.
19. Volz EM, Frost SD. Inferring the source of transmission with phylogenetic data. PLoS Comput Biol. 2013;9(12):e1003397.
20. Romero-Severson EO, Bulla I, Leitner T. Phylogenetically resolving epidemiologic linkage. Proc Natl Acad Sci USA. 2016;113(10):2690-5.

21. Wymant C, Hall M, Ratmann O, Bonsall D, Golubchik T, de Cesare M, et al. PHYLOSCANNER: Inferring Transmission from Within- and Between-Host Pathogen Genetic Diversity. Mol Biol Evol. 2017. [Epub ahead of print]
22. Noguera-Julian M, Edgil D, Harrigan PR, Sandstrom P, Godfrey C, Paredes R. Next-Generation Human Immunodeficiency Virus Sequencing for Patient Management and Drug Resistance Surveillance. J Infect Dis. 2017.16 (suppl_9):S829-S833
23. Moscona R, Ram D, Wax M, Bucris E, Levy I, Mendelson E, et al. Comparison between next-generation and Sanger-based sequencing for the detection of transmitted drug-resistance mutations among recently infected HIV-1 patients in Israel, 2000-2014. J Int AIDS Soc. 2017;20(1):21846.
24. Hue S, Clewley JP, Cane PA, Pillay D. Investigation of HIV-1 transmission events by phylogenetic methods: requirement for scientific rigour. AIDS. 2005;19(4):449-50.
25. Brown AE, Gifford RJ, Clewley JP, Kucherer C, Masquelier B, Porter K, et al. Phylogenetic reconstruction of transmission events from individuals with acute HIV infection: toward more-rigorous epidemiological definitions. J Infect Dis. 2009;199(3):427-31
26. Vandamme A. Basic concets of molecular evolution. Lemey P, M S, A V, editors. Cambridge, UK: Cambridge University Press; 2009.
27. Efron B, Halloran E, al. e. Bootstrap confidence levels for phylogenetic trees. Proc Natl Acad Sci USA. 1996;93(14):7085-90.
28. Ratmann O, Hodcroft EB, Pickles M, Cori A, Hall M, Lycett S, et al. Phylogenetic Tools for Generalized HIV-1 Epidemics: Findings from the PANGEA-HIV Methods Comparison. Mol Biol Evol. 2017;34(1):185-203.
29. Volz EM, Ionides E, Romero-Severson EO, Brandt MG, Mokotoff E, Koopman JS. HIV-1 transmission during early infection in men who have sex with men: a phylodynamic analysis. PLoS Med. 2013;10(12):e1001568;

30. Stadler T, Kuhnert D, Bonhoeffer S, Drummond AJ. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proc Natl Acad Sci USA. 2013;110(1):228-33.
31. Faria NR, Azevedo Rdo S, Kraemer MU, Souza R, Cunha MS, Hill SC, et al. Zika virus in the Americas: Early epidemiological and genetic findings. Science (New York, NY). 2016;352(6283):345-9.
32. Pozniak A, Gazzard B, Anderson J, Babiker A, Churchill D, Collins S, et al. British HIV Association (BHIVA) guidelines for the treatment of HIV-infected adults with antiretroviral therapy. HIV Med. 2003;4 Suppl 1:1-41.
33. Dunn D, Pillay D. UK HIV drug resistance database: background and recent outputs. J HIV Ther. 2007;12(4): 97-8.
34. Schoeni-Affolter F, Ledergerber B, Rickenbach M, Rudin C, Gunthard HF, Telenti A, et al. Cohort profile: the Swiss HIV Cohort study. Int J Epidemiol. 2010;39(5):1179-89.
35. Hughes GJ, Fearnhill E, Dunn D, Lycett SJ, Rambaut A, Leigh Brown AJ. Molecular phylodynamics of the heterosexual HIV epidemic in the United Kingdom. PLoS Pathog. 2009;5(9):e1000590.
36. Poon AF, Gustafson R, Daly P, Zerr L, Demlow SE, Wong J, et al. Near real-time monitoring of HIV transmission hotspots from routine HIV genotyping: an implementation case study. Lancet HIV. 2016;3(5):e231-8.
37. Wertheim JO, Kosakovsky Pond SL, Forgione LA, Mehta SR, Murrell B, Shah S, et al. Social and Genetic Networks of HIV-1 Transmission in New York City. PLoS Pathog. 2017;13(1):e1006000.
38. World Health Organization USCfDCaP, The Global Fund to Fight AIDS, Tuberculosis and Malaria. HIV drug resistance report 2017. Available from: http://www.who.int/hiv/pub/drugresistance/hivdr-report-2017/en/ (accessed Mar 28, 2018)
39. Ou CY, Ciesielski CA, Myers G, al. e. Molecular epidemiology of HIV transmission in a dental practice. Science (New York, NY). 1992;256:1165-71.

40. Lemey P, Van Dooren S, Van Laethem K, Schrooten Y, Derdelinckx I, Goubau P, et al. Molecular testing of multiple HIV-1 transmissions in a criminal case. AIDS. 2005;19(15):1649-58.
41. Bernard EJ, Azad Y, Vandamme AM, Weait M, Geretti AM. HIV forensics: pitfalls and acceptable standards in the use of phylogenetic analysis as evidence in criminal investigations of HIV transmission. HIV Med. 2007;8(6):382-7.
42. Abecasis AB, Pingarilho M, Vandamme AM. Phylogenetic analysis as a forensic tool in HIV transmission investigations. AIDS. 2018;32(5):543-54.
43. Carlson JM, Du VY, Pfeifer N, Bansal A, Tan VY, Power K, et al. Impact of pre-adapted HIV transmission. Nat Med. 2016;22(6):606-13.
44. Amon JJ, Kasambala T. Structural barriers and human rights related to HIV prevention and treatment in Zimbabwe. Global Public Health. 2009;4(6):528-45.
45. McGrath N, Eaton JW, Newell M-L, Hosegood V. Migration, sexual behaviour, and HIV risk: a general population cohort in rural South Africa. Lancet HIV. 2015;2(6):e252-e9.
46. Dobra A, Barnighausen T, Vandormael A, Tanser F. Space-time migration patterns and risk of HIV acquisition in rural South Africa. AIDS. 2017;31(1):137-45.
47. SASPEN. Migration and health care: The case of HIV and AIDS in Botswana. 2015. Available from: http://www.saspen.org/brief/SASPEN-brief_2015-6_Refilwe-Sinkamba_Migration-and-Healthcare-HIV-Botswana.pdf (accessed Mar 28, 2018)
48. Statistics_Botswana. Population and Housing Census - Analytical Report. Gaborone, Botswana: 2014. Available form: https://www.statsbots.org.bw/sites/default/files/publications/Population%20%26%20Housing%20Census%20Dissemination%20analytical%20report%20.pdf (accessed Mar 28, 2018)
49. Jacques G, Mmatli TO. Addressing Ethical Non-Sequiturs in Botswana's HIV and AIDS Policies: Harmonising the Halo Effect. Ethics and Social Welfare. 2013;7(4):342-58.
50. Munthe C. The Goals of Public Health: An Integrated, Multidimensional Model. Public Health Ethics. 2008;1(1):39-52.

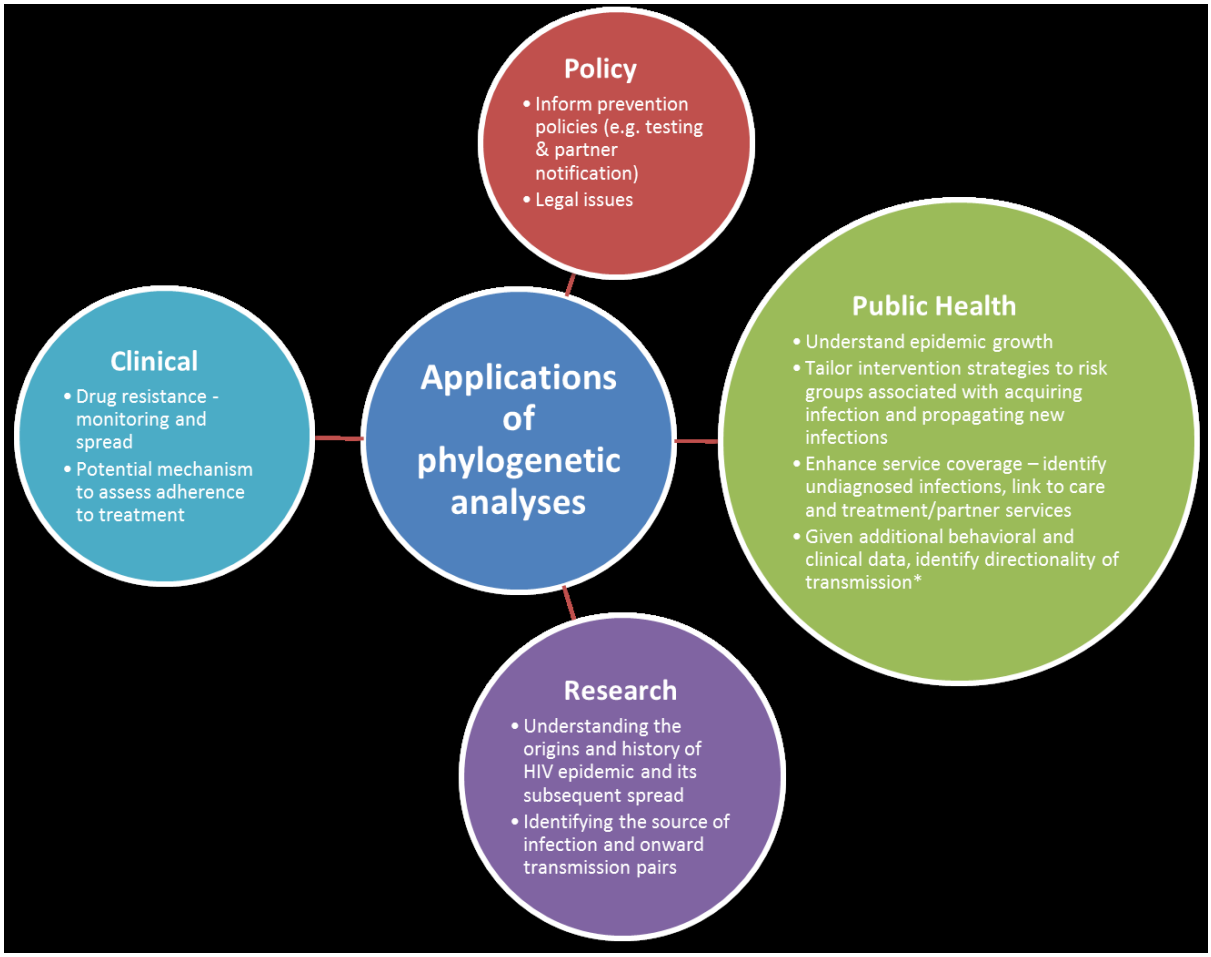
51. Sox HC. Resolving the tension between population health and individual health care. JAMA. 2013;310(18):1933-4.
52. Gostin LO, Bayer R, Fairchild AL. Ethical and legal challenges posed by severe acute respiratory syndrome: implications for the control of severe infectious disease threats. JAMA. 2003;290(24):3229-37.
53. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, et al. Antiretroviral Therapy for the Prevention of HIV-1 Transmission. N Engl J Med. 2016;375(9):830-9.
54. Cohen MS, McCauley M, Sugarman J. The Ethical Odyssey in Testing HIV Treatment as Prevention. Clinical Trials. 2012;9(3):340-7.
55. WHO. Health and human rights 2015 [cited 2017 23 August]. Available from: http://www.who.int/mediacentre/factsheets/fs323/en/ (accessed Mar 28, 2018)
56. Earley CL, Strong LC. Certificates of confidentiality: a valuable tool for protecting genetic data. Am J Hum Genet. 1995;57(3):727-731
57. Shannon K, Strathdee SA, Goldenberg SM, Duff P, Mwangi P, Rusakova M, et al. Global epidemiology of HIV among female sex workers: influence of structural determinants. Lancet. 2015;385(9962):55-71.
58. De Oliveira T, Pybus OG, Rambaut A, Salemi M, Cassol S, Ciccozzi M, et al. Molecular epidemiology: HIV-1 and HCV sequences from Libyan outbreak. Nature. 2006;444(7121):836-7.
59. Human Rights Watch. Tell me where I can be safe? The impact of Nigeria's same sex marriage (prohibition) act 2016 [cited 2017 25 May]. Available from: https://www.hrw.org/sites/default/files/report_pdf/nigeria1016_web.pdf. (accessed Mar 28, 2018)
60. Amon J. Personal communication with UNAIDS. 2017.

61. Bernard E, Cameron S. Advancing HIV Justice and Advancing HIV Justice 2: Building momentum in global advocacy against HIV criminalisation. Brighton/Amsterdam: HIV Justice Network and GNP+. 2016.
62. Galletly CL, Pinkerton SD. Conflicting messages: how criminal HIV disclosure laws undermine public health efforts to control the spread of HIV. AIDS and Behavior. 2006;10(5):451-61.
63. O'Byrne P, Bryan A, Woodyatt C. Nondisclosure prosecutions and HIV prevention: results from an Ottawa-based gay men's sex survey. JANAC. 2013;24(1):81-7.
64. O'Byrne P, Willmore J, Bryan A, Friedman DS, Hendriks A, Horvath C, et al. Nondisclosure prosecutions and population health outcomes: examining HIV testing, HIV diagnoses, and the attitudes of men who have sex with men following nondisclosure prosecution media releases in Ottawa, Canada. BMC Public Health. 2013;13:94.
65. Abecasis AB, Geretti AM, Albert J, Power L, Weait M, Vandamme AM. Science in court: the myth of HIV fingerprinting. Lancet Infect Dis. 2011;11(2):78-9.
66. Leitner T. Guidelines for HIV in court cases. Nature. 2011;473(7347):284.
67. Department of Health, Education, and Welfare; National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report. Ethical principles and guidelines for the protection of human subjects of research. The Journal of the American College of Dentists. 2014;81(3):4-13.
68. UNAIDS. Good participatory practice: Guidelines for HIV biomedical prevention trials. Geneva: UNAIDS, 2011. Available from: http://www.unaids.org/sites/default/files/media_asset/JC1853_GPP_Guidelines_2011_en_0.pdf (accessed Mar 28, 2018)
69. WHO. Ethical considerations in biomedical HIV prevention trials: UNAIDS-WHO guidance document. Geneva, Switzerland: UNAIDS & WHO, 2012. Available from: http://www.unaids.org/en/resources/documents/2012/20120701_jc1399_ethical_considerations (accessed Mar 28, 2018)

70. Singh JA, Mills EJ. The abandoned trials of pre-exposure prophylaxis for HIV: what went wrong? PLoS Med. 2005;2(9):e234.
71. Vermund SH, Blevins M, Moon TD, Jose E, Moiane L, Tique JA, et al. Poor clinical outcomes for HIV infected children on antiretroviral therapy in rural Mozambique: need for program quality improvement and community engagement. PLoS One. 2014;9(10):e110116.
72. PopART. PopART website 2017. Available from: https://www.lshtm.ac.uk/popart.
73. Boyson AR, Zimmerman RS, Shoemaker S. Exemplification of HAART and HIV/AIDS: A News Experiment. Health Commun. 2015;30(9):901-10.
74. Zimmerman RS, Kirschbaum AL. News of Biomedical Advances in HIV: Relationship to Treatment Optimism and Expected Risk Behavior in US MSM. AIDS and Behavior. 2017; 22(2):367-378. doi: 10.1007/s10461-017-1744-3.
75. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530(7589):228-32.

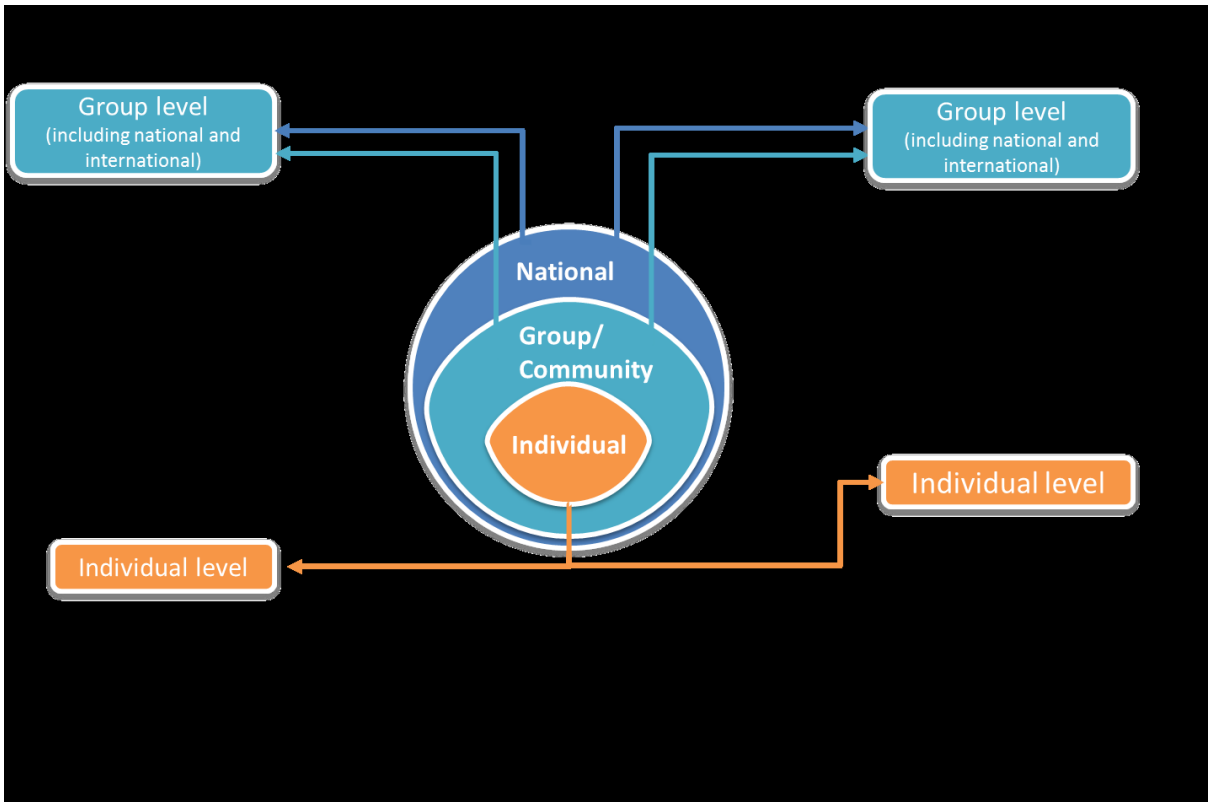
Figures

Figure 1: Applications of phylogenetic analyses



**Denotes a potential future use of phylogenetic analyses*

Figure 2: Potential benefits (left) and harms (right) associated with HIV phylogenetic analysis



Box 1: Migration in Botswana

Studying migrants is often fraught with both logistical and ethical problems. Migration has been identified as a key risk factor for the spread of HIV possibly because of the lack of access to culturally and linguistically appropriate prevention information and clinical care, disruption of established social relationships and the potential for increased risky sexual practices when people are away from home (45, 46). However, a more nuanced approach to migration and the link to HIV is needed in order to take differences in migration flows, risk environments and characteristics of the areas between which migrants move into account.

Migratory populations in Botswana (documented and undocumented, skilled and unskilled) face challenges in accessing health care services and are prohibited from receiving government provided free ART drugs (47). According to the 2014 national census (48) migrants accounted for approximately 14% of employed and 9% of unemployed populations in Francistown, Botswana. A significant proportion of HIV-infected migrants in Botswana are unaware of their positive HIV status and are not on ART, and may therefore disproportionally contribute to new HIV transmissions. Phylogenetic studies are able to assess the level of integration of migrants in the generalized HIV epidemic in Botswana. However, there are significant ethical issues and risks associated with the participation of migrants in phylogenetic research. Migrants may face additional stigmatisation and marginalization. They may also become subject to deportation, imprisonment or extortion in exchange for short-term visas (49). Obtaining consent can inhibit participation due to privacy and identification issues related to their undocumented status. Therefore, it is ethically preferable to enrol migrants anonymously to avoid these social harms.

Box 2: Examples of research and clinical challenges that are prevalent in phylogenetic analyses

Serodiscordant couples: HIV transmission often occurs within discordant couples in which one person is HIV-positive and the other is not. Phylogenetic analysis allows identification of linked transmission between the members of the pair. In one HIV study enrolling discordant couples, 30% of the HIV acquisition events were not linked to the known infected partner (53). While this information was critical for interpreting the efficacy of the prevention strategy tested in this particular trial (the interventions to reduce transmission was focused on the HIV positive partner), these results were not provided to study participants for fear of adverse consequences for the individuals involved (54). For example, domestic violence, loss of trust in relationships, and relationship break ups might all result from disclosure of partnerships. It may not be possible to mitigate these risks with counselling or follow-up.

Detection of transmission events in non-treatment-compliant patients: If ART does not suppress viral replication, some new HIV transmissions from those receiving treatment could occur. Linked to epidemiological data, phylogenetics has the potential to differentiate between transmitted drug resistance and poor adherence to ART thereby allowing healthcare professionals to initiate an appropriate intervention (for example treatment switch versus adherence counselling). However, if non-adherence results in transmission of HIV, then additional ethical issues, particularly in relation to transmission laws, may arise.

Box 3: Key legal and human rights pre-requisites for the use of phylogenetic tools for public health research

- Informed consent for collection and dissemination of phylogenetic data and information
- Confidentiality, safety and prevention of un-authorised use of phylogenetic data and information
- Non-stigmatisation and non-discrimination in collection and publication of phylogenetic data
- Attention to criminalisation and other potential negative consequences relating to collection and dissemination of phylogenetic data
- Specific gender consideration and attention to the particular risks and concerns faced by women and key populations due to coercive social and legal environments
- Awareness, that in some countries, collection and publication of phylogenetic data may require legislative or policy change
- Community participation and accountability for collection and use of data
- Legal redress in case of misuse of phylogenetic data
- Phylogenetic experts need to be consistent in their statements that source attribution cannot be definitively determined from phylogenetics alone.

Legal associated risks of misuse of phylogenetic data:

- Even with anonymization minimal information is required for self-identification which may provide information about individuals in the same network, leading to attribution of blame for infections which may increase prosecution episodes.
- Research data may be subject to subpoena due to laws on accessing public health data. This may result in misuse by governments and police to target vulnerable populations.

Box 4: Use of Phylogenetic Analysis in Criminal Convictions

Since the "Florida dentist case" in the beginning of the 1990s (39), phylogenetic analyses started to be used in court cases as a forensic tool in HIV transmission investigations, for example cases where one or more complainants allege that a defendant has unlawfully infected them with HIV (42). Cases can be criminal (in countries where transmission of HIV infection is specifically criminalized) or civil (in the context of general civil laws, for example, by applying physical or sexual assault laws to HIV-related cases). Most HIV-specific laws are overly broad and/or vague, and as such do not require proof of transmission for conviction; prosecution is often based on potential or perceived exposure with allegations of non-disclosure. However, when general criminal laws (such as those relating to bodily harm) are applied to allegations of HIV transmission proof of causality is often required.

Phylogenetic evidence cannot stand alone in court and should be used in the context of other evidence, such as full epidemiological investigation and contact tracing (41, 64, 65). Experts have worked with the Crown Prosecution Service for England and Wales to highlight the limitations and challenges of phylogenetics in prosecution cases including:

- Phylogenetic information based on Sanger sequencing alone cannot prove transmission beyond reasonable doubt – although an indirect link can never be ruled out. In contrast, significantly separated clustering can be used as evidence against direct transmission, provided the samples have been drawn close enough to the timing of transmission and do not get phylogenetically separated by onward transmission events.
- There are challenges in communicating results, such as the lack of certainly, to non-experts.
- Identification of a source of a transmission is not possible, as it would require for all strains of all patients ever infected with HIV to be available as "controls", and for phylogenetic trees to flawlessly reconstruct a true epidemic history. Both assumptions are unrealistic.

Box 5: Eight considerations for ethically responsible implementation of phylogenetic analyses

(i) **A careful risk-benefit assessment** should be conducted prior to designing, conducting and reporting phylogenetic analysis. Risk assessment should address risks to individuals and to groups that may be identified in the research.

(ii) **Protection of the rights and interests of study participants:** Individuals who participate in studies as well as the social and geographic groups that may be identified in phylogenetic networks need to be protected. Clinically relevant results should be returned to the patient and/or care provider.

(iii) **Social and legal context**: An awareness of the social environment, legal environment, human rights violations and other potential negative consequences is essential. This includes knowledge both of when and how data are subject to subpoena, and of precedent criminal cases. These challenges are context-specific and the legal, political and social environments are subject to change. Furthermore, considerations of gender specific risks and concerns faced by women and key populations should be evaluated due to coercive social and legal environments

(iv) **Risk mitigation strategies** should address risks to individuals and groups, and should take into account the potential for anonymization and masking of individual and group identifiers as protective strategies, as well as accounting for scientific needs of the project and its value in informing public health strategies. The technical nature of sequence data collected needs to be considered in terms of potential for re-linkage or other harms, and data may be pre-processed to reduce this risk in line with the needs of the study. Training should be conducted for research staff on the risks as well as the importance of anonymity, confidentiality, informed consent and protection of research participants and communities. Monitoring and redress mechanisms should be established to accompany and respond to misuse of phylogenetic data.

(v) **Informed consent and other safeguards**: Study participants and patients whose samples are being used for phylogenetic analysis should have consented to such use. In the absence of such consent, waivers of consent must have been obtained from the appropriate ethics committees. Researchers must ensure that specific populations are protected against non-stigmatisation and non-discrimination.

(vi) **Community engagement:** The engagement process should be started during the research design process, thereby ensuring that the research is relevant to participating communities and local perspectives are included in the design and overall conduct of the research studies, including risk assessment, risk mitigation, informed consent, and communication.

(vii) **Communication**: Phylogenetic results require expertise to undertake and to interpret. The results are usually ambiguous and it is crucial that the uncertainty associated with these methods is communicated appropriately during dissemination to the wider scientific community, government bodies, media, and participating communities. Specific efforts are needed to sensitise public health officials, the police and communities on the use of phylogenetic analysis in the context of public health including its benefits and limitations.

(viii) **Equitable data sharing:** Accountability of phylogenetic and sequencing data should be ensured and we recommend that a governance plan is created to address confidentiality, safety and potential un-authorised use of phylogenetic data. Information and protocols for data sharing, including controlled access, must be addressed in the governance plan. Only limited information should be routinely published with each sequence and care is needed to ensure human DNA sequences are not inadvertently released with NGS data.

SUPPLEMENTARY MATERIALS

Ethics of HIV Phylogenetics Working Group Members:

Africa Health Research Institute, KwaZulu-Natal, ZA (Guy Harling, ScD; Prof Deenan Pillay, PhD; Prof Janet Seeley, PhD);

Bill & Melinda Gates Foundation, Washington, DC, US (Gina Dallabetta, MD);

Centers for Disease Control and Prevention, Atlanta, GA, US (Irene Hall, PhD);

European AIDS Treatment Group, Brussels, BE (Nikos Dedes, BA)

Harvard T.H. Chan School of Public Health, Boston, MA, US (Guy Harling, ScD; Vladimir Novitsky, PhD);

HIV Justice Network, Brighton, UK (Edwin J. Bernard, BA (Hons));

Johns Hopkins University, Baltimore, MD, US (Prof Gail Geller, ScD; Danielle German, PhD; M. Kate Grabrowski, PhD);

Joint United Nations Programme on HIV/AIDS, Geneva, CH (Patrick M. Eba, PhD; Prof Peter Godfrey-Faussett, FRPC);

London School of Hygiene & Tropical Medicine, London, UK (Prof Peter Godfrey-Faussett, FRCP; Prof Janet Seeley, PhD);

National Institutes of Health, Bethesda, MD (Prof David Burns, MD; Liza Dawson, PhD; Oliver Laeyendecker, PhD);

Princeton University, Princeton, NJ, US (Dr Joseph J. Amon, PhD)

Public Health England, London, UK (Valerie Delpech; FPHM)

Treatment Advocacy and Literacy Campaign, Lusaka, ZM (Felix Mwanza, BA);

Universidade Nova de Lisboa, Lisbon, PT (Prof Anne-Mieke Vandamme, PhD);

University California, San Diego, La Jolla, CA, US (Joel O. Wertheim, PhD);

University College London, London, UK (Cordelia C. Coltart, PhD; Guy Harling, ScD; Anne Hoppe, PhD; Zisis Kozlakidis, PhD; Prof Deenan Pillay, PhD)

University of KwaZulu-Natal, Pietermaritzburg, ZA (Patrick M. Eba, PhD);

University of Leuven, Leuven, BE (Prof Anne-Mieke Vandamme, PhD);

University of Missouri, St Louis, MO, US (Prof Rick Zimmerman, PhD);

University of North Carolina, Chapel Hill, NC, US (Prof Myron S. Cohen, MD; Prof Gail Henderson; PhD; Joseph D. Tucker, PhD);

University of Oxford, Oxford, UK (Prof Christophe Fraser, PhD; Prof Michael Parker, PhD)

World Health Organization, Geneva, CH (Rachel Baggaley, MBBS; Andreas Reis, MD);

Zambart, Lusaka, ZM (Musonda Simwinga, PhD)

Supplementary Table 1: Summary of key documents, position statements and initiatives relevant to ethical issues of HIV phylogenetics and referred to within the document

Document/ Position statement/ Initiative	**Description**
The Declaration of Helsinki (first published in 1964; amended most recently in 2013)(6)	The World Medical Association developed the Helsinki declaration as a statement of ethical principles for medical research involving human subjects, including research on identifiable human material and data. This declaration states that the interest and well-being of the individual takes precedence over the science and well-being of communities and populations. Many of the principles are relevant to performing HIV phylogenetic studies.
The Council for International Organizations of Medical Sciences (CIOMS) International Ethical Guidelines (7)	CIOMS is an international nongovernmental organization in official relationship with WHO, founded in 1949. The guidelines aim to provide internationally vetted ethical principles and detailed commentary on how universal ethical principles should be applied, with particular attention to conducting research in lower-income countries (LIC). There have been four revisions of the guidelines since they were first published (1982) to take into account scientific developments and bring the guidelines into line with current thinking on ethics and human rights.
Wellcome Trust report on Ethical	This report aims to provide evidence to inform the

sharing of health research data in LMIC: views of stakeholders (11)	development, implementation and evaluation of data-sharing models and identify further research priorities. It is based on a multi-site collaborative study of stakeholder experiences and views in LMIC of best practices in sharing individual-level data from clinical and public health research.
The Human Heredity and Health in Africa (H3Africa) Initiative (12, 13)	H3Africa Initiative aims to facilitate a contemporary research approach to the study of genomics and environmental determinants of common diseases with the goal of improving the health of African populations. To accomplish this, the H3Africa Initiative aims to contribute to the development of the necessary expertise among African scientists, and to establish networks of African investigators.
The ELSI (Ethical, Legal, and Social Implications) Program (8)	The ELSI Program is a multi-disciplinary program funded by the National Human Genome Research Institute at NIH. It focuses on exploring ELSI of human genomics, and developing policy options to address these implications, although the scope has broadened over years in response to rapidly evolving genomic technologies, legal and commercial developments, and translation to clinical applications. Many of the issues from human genomics also apply to viral genomics: Biobank governance is a particular focus of the “ELSI 2.0” initiative (14).

Supplementary Box 1: Key methodological concepts to be considered when inferring HIV molecular transmission clusters

- Phylogenetic support methods use bootstrap or posterior probability to identify groups more closely related to each other than to the rest of the population being analysed:
    - Bootstrapping: a statistical resampling method of random sampling of nucleotide sites with replacement. This process is repeated multiple times and the frequency of identical branch reproduction gives a bootstrap value indicating the robustness of the cluster assignment.
    - Posterior probability (PP): combines the prior probability of a tree with the likelihood of the given data to indicate the probability of the cluster assignment to be correct.
- Phylogenetic distance methods identify groups whose mean/median/maximum genetic distance suggests a common ancestor in recent time. Phylogenetic support and distance methods are often combined.
- Molecular clock methods indicate the timing of the most recent common ancestor, which can contribute to understanding the timing of infection.