Phylogenetic analysis of pathogens is an increasingly powerful way to 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, presenting unique ethical, legal, and social issues that are not adequately addressed by existing bioethics literature. We formed a multidisciplinary working group to explore the ethical issues arising from the design of, conduct in, and use of results from HIV phylogenetic studies, and to propose recommendations to minimise the associated risks to both individuals and groups. We identified eight key ethical domains, within which we highlighted factors that make HIV phylogenetic research unique. In this Review, we endeavoured to provide a framework to assist researchers, public health practitioners, and funding institutions to ensure that HIV phylogenetic studies are designed, done, and disseminated in an ethical manner. Our conclusions also have broader relevance for pathogen phylogenetics.

Introduction
Understanding of the transmission dynamics of infectious agents is essential for development of effective public health interventions. Historically, transmission dynamics were investigated with epidemiology: tracking of the evolution of epidemics through time, place, and person, primarily with observation and self-reports of exposure and risk behaviours. However, despite advances in HIV prevention, incidence remains high, notably in sub-Saharan Africa, which accounts for 75% of all new HIV infections worldwide.1 In this context, an understanding of who is most likely to infect whom remains important to the development of targeted prevention strategies.

In phylogenetic analyses, historical relationships between individuals or groups are deduced by comparison of pathogen genomes to establish how closely related viruses from two individuals are. In combination with traditional epidemiological data, viral genetic sequence data can help to infer transmission patterns. This combination has the potential to answer key questions that are not easily addressed by traditional or molecular approaches alone.2,4

The focus of HIV phylogenetic studies has extended from concentrated to generalised epidemics, and increasingly involves large 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 guidelines, address several of these issues, including the need for informed consent, community engagement, risk minimisation, and consideration of the risks and benefits of research for groups and communities.6,7 Additionally, a large and diverse collection of academic and policy literature exists that addresses the ethical, legal, and social implications of the research and clinical uses of human genomics in high-income countries.8,9 This literature has been accompanied by a growing assortment of bioethics and social science literature on the implications of genomic research in low-income and middle-income countries (LMICs),10,11 and by stakeholder engagement initiatives in these settings (appendix p 3–4).12–14

HIV phylogenetic research presents complex ethical issues, including two specific challenges. First, (like contact-tracing data) phylogenetic analyses are fundamentally relational: analysis of data from one person might affect other people (eg, by identifying them as potential sources of infection). Second, as next-generation sequencing (NGS) produces richer sequence data, true anonymisation of viral sequence data becomes difficult because virus isolated from another timepoint in another study could be used to reidentify an individual.

In recognition of these issues and the need for the development of good ethical practice models in this area, we held a multidisciplinary (scientists, bioethicists, lawyers, human rights advocates, HIV activists, and community engagement members from Africa) workshop in London (UK) in May, 2017.15 This meeting focused on identifying the key issues arising from study design and conduct or use of 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.

In this Review, we summarise the findings and recommendations from the workshop and follow-up discussions, and we set out a framework for researchers and funding bodies supporting HIV genetic studies. This Review 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 that the more similar their genetic sequence, the closer in time they are to having a common ancestor. Application of this approach to HIV detected in blood samples from infected human populations helps in the understanding of HIV transmission patterns.

HIV is suitable for phylogenetic analysis because it is highly genetically variable. Given that transmission of HIV involves two individuals (a couple), the variability of HIV can be used to infer linkages by forming phylogenetic clusters between couples and groups of people. Furthermore, in many cases, inferred with a degree of uncertainty, the direction of transmission within clusters is possible by use of either additional epidemiological data or data with sequences from multiple viruses sampled from each individual.

Different techniques can be used to generate sequence data required for phylogenetic analysis. NGS increases both the potential power and potential risks of phylogenetic approaches compared with conventional Sanger sequencing methods, because it provides information on intrahost variation and returns richer sequence data for multiple viral particles per sample. Additionally, many methods and associated assumptions in phylogenetic analyses exist for identifying clusters of genetically related viruses (appendix p 5). Phylogenetic clusters are generally thought to represent groups of infected individuals who are closer together in a transmission chain and can be identified with various methods.

Caution is required when interpreting phylogenetic clusters for epidemiological purposes, because clusters are typically inferred from partially sampled transmission chains (ie, some infected individuals were not sampled). Unsampled cases can be either a common source of infection or an intermediary in a transmission chain for people with genetically similar viruses. Although historically proving transmission between two individuals has been difficult, the use of NGS with improved interpretation algorithms makes such inferences more likely to happen in future.

Phylogenetic analyses can be used widely in HIV epidemiology (figure 1), for example, to study viral linkage and risk factors for epidemic spread (molecular epidemiology), to examine the growth and decline of HIV epidemics (phylogeography), or to investigate the impact of migration on HIV spread and to identify hubs of transmission (phylogeography). HIV phylogenetics has been most applied in high-income countries with well-developed scientific infrastructure, where HIV is characterised by smaller epidemics focused in specific risk groups (concentrated epidemics). In many high-income countries, 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. This sequencing has led to the growth of national HIV genetic databases, such as those in the UK and Switzerland. If these datasets are linked to epidemiological surveillance and clinical cohort data, inferences can be made with regard to patterns of exposure and risk factors for onward transmission among infected individuals. Unlike standard epidemiological data, molecular data can allow inferences to be made regarding the time of transmission relative to the time of sample collection. Furthermore, 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 methods, phylogenetic research provides a more detailed and precise understanding of epidemic characteristics, thereby enabling improved public health policies, including more effective and better targeted programmes for prevention and treatment.

Many African epidemics are much larger than those in the USA and Europe. Viral sequencing in Africa is not routinely done outside targeted programmes, such as WHO’s HIV drug resistance surveillance, and research projects. Although declining costs and progressively easier sequencing will increase the proportion of infected individuals represented within sequence databases, such as

---

**Figure 1: Applications of phylogenetic analyses**

*Denotes a potential future use of phylogenetic analyses.*
PANGEA-HIV: Africa does not yet have the extensive and comprehensive datasets seen in high-income countries. Additionally, community and patient mobilisation around HIV takes very different forms compared with that in high-income countries, and the social, political, and economic context is considerably different and varies among African countries. Ethical analysis of phylogenetic work will 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 issues include the potential for stigmatisation and risk of social harm to individuals or groups, and concerns about privacy, confidentiality, and security of data. However, certain risks are particularly salient in phylogenetic research, which we discuss over eight key ethical domains.

**Risk and benefit assessments**

The harms and benefits of phylogenetic research will vary depending on whether they are assessed at an 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, while at the same time minimising the risks to individuals, groups, and populations. Of particular concern to phylogenetics is the possible deduction of complex social and sexual relationships, should phylogenetic data be combined with minimal clinical and demographic information. Conversely, traditional epidemiological studies require far more information to infer transmission of HIV between individuals, particularly with respect to directionality of infection.

Risks to individuals principally arise from inadvertent or intentional disclosure of HIV status or transmission events, or from demands for these data for judicial or extrajudicial targeting of individuals or groups. In several countries, phylogenetic evidence is being used in criminal cases of alleged HIV transmission. Breaches of confidentiality could occur through inadequate anonymisation or deductive disclosure, through misinterpretation, miscommunication, or misuse of the analytic results, or through legal action.

These risks will increase if more data are generated and made publicly available, a requirement of many funding agencies and publishers. Although publication might maximise 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, in which individuals can choose what information they disclose to investigators, inferences made from viral genetic sequences are not controlled by participants.

Anonymisation can provide some protection to individuals. However, even with anonymisation, deductive disclosure of identities from HIV sequences and other corresponding data remains theoretically possible. Through the use of rich NGS data applied at successive timepoints, a virus sequenced from an individual could in principle be used to relink that individual to an earlier study with high reliability. Small fragments of human
DNA sequence contained in NGS data could also be accidentally released. Furthermore, HLA of the infected individual is imprinted on the virus because of immune selection, which might assist identification of individuals in the future.44 The probability of relinking individuals to anonymised data can be minimised by processing sequence data before release (eg, by only including consensus sequences, which suffice for many phylogenetic methods).

Despite these risks, maintaining links to individuals’ identities might allow for direct benefits to individuals. Sequence information can provide clinical guidance (eg, by allowing treatment optimisation following detection of drug resistance mutations). 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 the population level, phylogenetic analysis can allow individuals’ data to be linked 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 metadata variables used in phylogenetic analysis is an important ethical decision. Phylogenetic analyses are often based on individual-level demographic, behavioural, or clinical variables, ignoring structural and environmental factors. The perception that certain groups (eg, key populations, such as men who have sex with men [MSM] or people who inject drugs) are responsible for infecting others and sustaining the HIV epidemic might be reinforced by only focusing on these variables. Conversely, other structural factors, such as those highlighted in the case study of migration in Botswana (panel 1), as well as sexual violence, lack of access to prevention and treatment, and having experienced discrimination, can have a prominent role in HIV transmission.44 Studying these factors and their effect on HIV transmission risk can decrease the blame mentality and create an alternative understanding of how to reduce HIV transmission and which individuals or groups are most at risk and why.

Plans to address the risks to individuals and 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, anonymisation of data provides substantial protection. Although individuals could theoretically be identified through reanalysing and relinking anonymised data from different sources, it would be difficult and require specialised expertise. By contrast, datasets linked with individual identifiers could be subpoenaed or obtained through unauthorised means, putting individuals at risk.

Researchers need to assess carefully the possibility of specific individuals or groups of people being identified 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. An ongoing monitoring of anticipated and unanticipated risks should be built into HIV phylogenetic research and mitigation strategies identified as early as possible.

Protection of the rights and interests of study participants

Effective phylogenetic work often occurs at the interface between research and public health practice because the same data can be used for both purposes. Researchers are typically viewed as obliged to protect individuals who enrol in a study from risk of harm, as far as possible, while pursuing valuable knowledge. Conversely, public health agencies have the mission of protecting the health of the public, which sometimes involves overriding individuals’ privacy interests to use data for public-health decision making. When research also has implications for specific population groups, further considerations relating to group harm are important (panel 2).41-46

The obligation of researchers to communicate results to study participants needs to be evaluated for each phylogenetic study. When clinical action is required, there is an obligation to make results available. In general, this goal (and benefit) is theoretical because phylogenetic results are produced with a substantial delay from sampling; therefore, any result is unlikely to be timely in informing clinical care. However, with the evolution of
real-time phylogenetics, reporting of drug resistance data to study participants could result in changes to clinical management. A second potential issue is the source of HIV acquisition in discordant couples. Whether HIV acquisition events are linked to the known infected partner might be crucial for interpretation of the efficacy of prevention strategies (panel 2). However, disclosure of these results to study participants could lead to adverse consequences for individuals involved.39

**Panel 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 couple. In one HIV study enrolling discordant couples, 30% of the HIV acquisition events were not linked to the known infected partner.38 Although this information was essential for interpreting the efficacy of the prevention strategy tested in this particular trial (the interventions to reduce transmission were focused on the HIV-positive partner), these results were not provided to study participants for fear of adverse consequences for the individuals involved.39 For example, domestic violence, loss of trust in relationships, and relationship break-ups might all result from disclosure of partnerships. It might not be possible to mitigate these risks with counselling or follow-up.

**Detection of transmission events in non-treatment-compliant patients**

If antiretroviral therapy does not suppress viral replication, some new HIV transmissions from those patients receiving treatment could occur. Linked to epidemiological data, phylogenetics have the potential to differentiate between transmitted drug resistance and poor adherence to antiretroviral therapy, thereby allowing health-care professionals to initiate an appropriate intervention (eg, treatment switch vs adherence counselling). However, if non-adherence results in transmission of HIV, then additional ethical issues can arise, particularly in relation to transmission laws.

**Panel 3: Key legal and human rights prerequisites for the use of phylogenetic methods for public health research**

- Informed consent for collection and dissemination of phylogenetic data and information.
- Confidentiality, safety, and prevention of unauthorised use of phylogenetic data and information.
- Non-stigmatisation and non-discrimination in collection and publication of phylogenetic data.
- Attention to criminalisation and other potentially 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 might 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**

- Minimal information is required for self-identification, even with anonymisation, which might provide information about individuals in the same network, leading to attribution of blame for infections, which might increase prosecution episodes.
- Research data can be subject to subpoena because of laws on accessing public health data, which might result in misuse by governments and police to target vulnerable populations.
Panel 4: Use of phylogenetic analysis in criminal convictions

Since the Florida dentist case in the beginning of the 1990s, phylogenetic analyses started to be used in court cases as a forensic tool in HIV transmission investigations (eg, cases in which one or more complainers allege that a defendant has unlawfully infected them with HIV). Cases can be criminal (in countries where transmission of HIV infection is specifically criminalised) 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 or vague, and 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. 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. By contrast, substantially 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.
- Communication of results to non-experts has challenges, such as the lack of certainty.
- Identification of the 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.

populations that data might be misused against them can undermine trust in research projects and health-care systems, putting HIV prevention and treatment programmes at risk. Research done in countries where privileged information between medical practitioners and their patients can be seized in HIV-related criminal trials showed that people with HIV were more reluctant to speak openly with their practitioners about their sexual partners and practices. These risks can be mitigated in phylogenetic studies through awareness of social and legal issues and ensuring they are addressed at the planning stages and monitored throughout the project.

Risk mitigation strategies to protect individual and group identities

Many of the risks associated with identification of individuals from phylogenetic information in environments with oppressive laws and policies can be reduced through use of anonymisation. Therefore, one default position is that anonymisation of data is preferred, if the scientific objectives can be accomplished. This position presumes no overriding interest in individuals receiving research results at the individual level. If data are not relevant for clinical care (eg, because of a substantial time delay between sample collection and generation of sequence information) then little rationale exists for returning data to health-care workers. When sequence analysis is timely, resistance data should be returned to clinics before doing phylogenetic analyses on anonymised sequencing data.

If anonymisation is detrimental to the scientific objectives or public health, further ethical analysis must be done and specific steps taken to protect the data from use in harmful proceedings. These steps might be technical (eg, storage linkage to identifiers in coded, separate databases with controlled access) or legal (eg, legal agreements protecting data from disclosure for the duration of the study).

The deanonymisation risk for individuals by use of later samples, such as infected blood collected at a different timepoint, as part of a linked or unrelated study or as part of a criminal investigation, can be mitigated by restricting the amount of data shared. Such restriction could include limitation to one virus sequence per individual, storage of raw NGS data under managed access, or destruction of raw data. However, these actions could reduce the potential scientific benefits of the study (eg, because they limit the ability to apply newly developed bioinformatic algorithms to infer the direction of transmission).

Risks to groups cannot be addressed through anonymisation of individuals. Certain groups can be placed at risk through characterisation as high risk or likely to transmit virus, including geographically defined groups, sexual or gender minorities, or 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 might 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 strategies might require the establishment of ties with local legal services, organisations working to protect people with HIV, and criminalised or stigmatised populations, to ensure that they have access to the means to protect their rights.

Training researchers and health-care professionals involved in phylogenetic investigations 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.

Valid informed consent and other safeguards

The formal requirements for the achievement of valid consent are well established in literature and...
guidelines.\(^6\)^\(^7\)^\(^8\)^\(^9\) The issues arising in relation to consent for phylogenetic studies are likely to be multifaceted. The procurement of 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 both to explain and to understand, and have multiple possible risks and benefits.

The complexity of concepts involved in phylogenetic research can raise fears about the aims of the work and the implications of participation among research participants, front-line research staff, health-care professionals, and ethics committee members. Communication methods that increase the understanding of phylogenetic studies need to be designed and evaluated. These must emphasise 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 consent model suitable for all circumstances 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 do not include explicit consent to participate in large-scale phylogenetics analyses. When using data from previous studies, researchers must be aware that only broad consent for HIV-related research might have been obtained. In such situations, a waiver of specific consent might be obtainable from an ethics committee. Waivers of specific consent are allowable when samples are no longer linked to identifiers, or when 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 studies 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.

Community engagement

Community engagement should occur early in the research design process, ensuring that phylogenetic research is relevant to participating communities and that local perspectives are included in the design and overall conduct of research studies.\(^6\)^\(^7\)^\(^8\)^\(^9\) Meaningful community engagement is particularly challenging in research-naive and low-income communities, and in criminalised or socially marginalised populations. Lack of authentic representation structures, poor literacy, or poverty place these communities at risk of being exploited,\(^6\)^\(^7\)^\(^8\) especially when research involves highly technical elements, such as viral genomics.

Nevertheless, these challenges should not limit attempts to maximise engagement. The phylogenetics study team of the PopART study in Zambia has performed extensive community engagement in 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 stigmatisation of or within communities.

Communication

Scientific inferences are based on probabilities. Comprehension and communication of uncertainty is key to understanding phylogenetic results; technical complexity or lack of familiarity with methods might easily generate a false sense of accuracy and precision. Researchers doing phylogenetic analyses 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 to 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 probably improve over time.

Mass media campaigns, as well as reporting on social media, television, radio, and in newspapers have been powerful ways to raise awareness about HIV, treatments, and prevention, and to facilitate 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 can generate unintentional consequences, including a lack of trust in health-care services.\(^5\)^\(^7\)^\(^2\)^\(^3\) Any ambiguous or misleading reporting of phylogenetic studies might reduce frequency of HIV testing, increase scepticism about participating in studies, and make risk groups less likely to access health care. Therefore, education of the media, local health-care personnel, and the community about these studies is essential.

Care must be taken when reporting findings relevant to specific population subgroups, including identifiable geographic areas or population groups that might be stigmatised or targetted 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.
Panel 5: Eight considerations for ethically responsible implementation of phylogenetic analyses

1. Careful risk–benefit assessment
Done before designing, conducting, and reporting phylogenetic analysis. Risk assessment should address risks to individuals and to groups that might be identified in the research.

2. Protection of the rights and interests of study participants
Individuals who participate in studies and the social and geographic groups that might be identified in phylogenetic networks need to be protected. Clinically relevant results should be returned to the patient or care provider.

3. Social and legal context
An awareness of the social environment, legal environment, human rights violations, and other potential negative consequences is essential. This understanding includes knowledge both of when and how data are subject to subpoena, and of precedent criminal cases. These challenges are specific to the context 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, owing to coercive social and legal environments.

4. Risk mitigation strategies
Study designs should address risks to individuals and groups and take into account the potential for anonymisation 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 relinkage or other harms, and data can be preprocessed to reduce this risk in line with the needs of the study. Research staff should be trained 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.

5. 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.

6. 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.

7. Communication
Expertise is needed to do and to interpret phylogenetic results. The results are usually ambiguous and therefore the uncertainty associated with these methods must be 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.

8. 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 unauthorised use of phylogenetic data. Information and protocols for data sharing, including controlled access, must be addressed in the governance plan. Only certain information should be routinely published with each sequence, and care is needed to ensure human DNA sequences are not inadvertently released with next-generation sequencing data.

Equitable data sharing
Largely as a result of funders’ requirements, many anonymised HIV sequences are being made publicly available on GenBank and LosAlamos. This availability is advantageous for some research studies, such as vaccine development. However, the lack of awareness that every sequence is associated with a patient or study participant is a real risk. Care must be taken to ensure human DNA sequences are not inadvertently released with NGS data. Routinely publishing only certain information (eg, the year of sampling and the country where the samples are collected) with each sequence would help to minimise risk. Any other anonymised information should be provided by use of 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 protocol would require the development of a clear governance plan.

Finally, different participant information sheets and consent forms can allow for different amounts of data sharing, and laws can differ as to how data can be reused. Any phylogenetic researcher must abide by the amount of sharing outlined in the forms, even if this impacts on the quality of the research conducted.

Conclusions and recommendations
Phylogenetic analysis, either alone or in combination with linked epidemiological data, is a powerful method 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 or groups participating in studies while optimising the scientific benefits. Although a one-model approach to address any ethical issues is impractical, given the vast variations in studies and contexts, this Review highlights...
Search strategy and selection criteria

An outline document was created before the workshop, which included the PubMed search terms “HIV transmission”, “HIV and viral linkage”, “HIV and genetic linkage”, “HIV & phylogenetic”, “sequencing technologies”, “Stigma and HIV”, “Stigma and MSM”, “HIV criminalisation”, “MSM, HIV and criminalisation”, “HIV transmission laws/criminalisation”, and “HIV and female sex workers”.

Additional information was obtained by reviewing resources available from UNAIDS and from human rights associations such as the International Lesbian Gay Bisexual Trans and Intersex Association. The outline document was used to draft an agenda, identify a list of experts to be invited to the workshop and to debrief all workshop attendees. Workshop attendees included social scientists, bioethicists, phylogeneticists, clinicians, virologists, lawyers, human rights advocates, HIV activists, and community engagement members. All speakers were asked to provide an abstract and give a presentation on the key issues that, in line with their field of expertise, needed to be considered for HIV phylogenetic analysis. The issues were subsequently debated and a consensus was formed. This Review was drafted based on the summary of meeting minutes, abstracts submitted beforehand, the meeting, and follow-on email discussions. Reference suggestions from experts were included in this Review.

Any researcher doing phylogenetic analysis should be aware of the risks such analyses pose and take steps to mitigate these risks. These steps are particularly pertinent in LMICs, which often have weak governance structures and few 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 could 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 the design and implementation of any study and the reporting of results.

Contributors
CEMC drafted the original manuscript. AH, MP, LD, GG, GHe, GHa, MKG, CF, MSC, and DP made substantial contributions to the manuscript. 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 in HIV Phylogenetics meeting and excerpts of meeting abstracts (JJ-A, MS, PE, VN, and A-MV) have been incorporated into the manuscript. Members of the Ethics in HIV Phylogenetics Working Group (CEMC, AH, GG, OL, JDT, JS, GD, PG-F, MSC, and DP) organised the meeting. AH led the working group with support from MSC and DP, 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 done with support from the HIV Prevention Trials Network and the PANGEA-HIV project (grant code OPP1088436). CEMC is funded by the Wellcome Trust (grant code 106531/Z/14/A). OL received additional support from the Division of Intramural Research National Institute of Allergy and Infectious Diseases. A-MV received funding from Fonds voor Wetenschappelijk Onderzoek—Flanders (grant G.0692.14), and the VIROGENESIS project receives funding from the European Union’s Horizon 2020 Research and Innovation Programme (under grant agreement 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 have been paid to write this Review by a pharmaceutical company or other agency. The views expressed in this Review do not represent any policy or position of the US Department of Health and Human Services or any of its components. All other authors declare no competing interests.

Acknowledgments
We would like to thank Shufang Wei for producing the infographics for this paper, Habiba Cooper Diallo and Carli Breeden for their assistance in organising the Ethics in HIV Phylogenetics meeting, Anne Johnson for her helpful discussions and commenting on the manuscript, and the PANGEA-HIV Steering Committee for their support.

References
10 Parker M, Kwiatkowski DP. The ethics of sustainable genomic research in Africa. Genome Biol 2016; 17: 44.
Review

53 Sox HC. Resolving the tension between population health and individual health care. JAMA 2013; 310: 1933–34.

© 2018 Elsevier Ltd. All rights reserved.

www.thelancet.com/hiv Published online August 30, 2018 http://dx.doi.org/10.1016/S2352-3018(18)30134-6
Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

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)
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 Grabowski, 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**

<table>
<thead>
<tr>
<th>Document/ Position statement/ Initiative</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Declaration of Helsinki (first published in 1964; amended most recently in 2013)(6)</td>
<td>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.</td>
</tr>
<tr>
<td>The Council for International Organizations of Medical Sciences (CIOMS) International Ethical Guidelines (7)</td>
<td>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.</td>
</tr>
<tr>
<td>Wellcome Trust report on Ethical</td>
<td>This report aims to provide evidence to inform the</td>
</tr>
<tr>
<td>sharing of health research data in LMIC: views of stakeholders (12)</td>
<td>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.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>The Human Heredity and Health in Africa (H3Africa) Initiative (13)</td>
<td>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.</td>
</tr>
<tr>
<td>The ELSI (Ethical, Legal, and Social Implications) Program (8, 9)</td>
<td>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 (9).</td>
</tr>
</tbody>
</table>
**Supplementary Panel 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:
  - Bootstrap: 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.
Supplementary Panel 2: Important social and legal considerations

- Are there particular approaches to handling reporting of results that will reduce risk?

- What role can local and external ethics committees play in addressing risks and handling the potential for political issues that arise in the local context?

- What kinds of discussions with policy makers, government officials or other stakeholders might be helpful in planning the research and communicating findings?

- What on-going monitoring will be conducted to ensure respect for study participants and impact on people with HIV or key populations? What resources are available for advocacy and redress if concerns arise?

- Have groups of people with HIV and key populations been meaningfully consulted? Are their views and concerns taken into account?
Supplementary Panel 3: Five Key questions that need to be addressed as part of responsible and ethical community engagement in phylogenetic research

1. What is the best community engagement strategy for phylogenetic studies and how sustainable is it?

2. How should feedback be provided to communities? Does informing entire communities add value and/or pose risks?

3. How valid is informed consent when a phylogenetic study is nested within an existing study or healthcare setting? Does it become a question of trust?

4. How can we avoid stigmatization when public health interventions are tailored towards specific communities?

5. How can researchers best share results at community level? Is it disrespectful not to do so?