Original research article

Superimposing incident sexually transmitted infections on HIV phylogram to investigate possible misclassification of men who have sex with men as heterosexuals in a cohort in Antwerp, Belgium

Kara K Osbak[1], Conor J Meehan[2], Sergio G Ribas[1],
Leo Heyndrickx[1,2], Kevin K Ariën[2], Achilleas Tsoumanis[1],
Eric Florence[1], Marjan Van Esbroeck[1], Katrien Fransen[1] and
Chris R Kenyon[1,3]

Abstract
In this study, we assessed if the superimposition of incident sexually transmitted infections (STIs) on HIV phylogenetic analyses could reveal possible sexual behaviour misclassifications in our HIV-infected population. HIV-1 sequences collected between 1997 and 2014 from 1169 individuals attending a HIV clinic in Antwerp, Belgium were analysed to infer a partial HIV transmission network. Individual demographic, clinical and laboratory data collected during routine HIV follow-up were used to compare clustered and non-clustered individuals using logistic regression analyses. In total, 438 (37.5%) individuals were identified in 136 clusters, including 76 transmission pairs and 60 clusters consisting of three or more individuals. Individuals in a cluster were more likely to have a history of syphilis, Chlamydia and/or gonorrhoea ($P < 0.05$); however, when analyses were stratified by HIV transmission risk groups (heterosexual and men who have sex with men [MSM]), this association only remained significant for heterosexuals with syphilis ($P = 0.001$). Under closer scrutiny, this association was driven by six heterosexual men who were located in six almost exclusively MSM clusters. A parsimonious conclusion is that these six individuals were potentially misclassified as heterosexual. Improving the accuracy of sexual behaviour reporting could improve care.

Introduction

Phylogenetic analysis of HIV sequences passively accumulated through routine drug resistance testing (DRT) can provide useful insights into local HIV transmission network characteristics when coupled with demographic, clinical and sexually transmitted infection (STI) data. These approaches have proven useful in highlighting HIV transmission networks at various geographical levels,[1–3] thereby allowing the identification of hotspots within (sub)populations such as men who have sex with men (MSM).[4] A recent Belgian study by Chalmet et al.[1]

[1]Department of Clinical Sciences, Institute of Tropical Medicine, Antwerp, Belgium
[2]Department of Biomedical Sciences, Institute of Tropical Medicine, Antwerp, Belgium
[3]Division of Infectious Diseases and HIV Medicine, University of Cape Town, Cape Town, South Africa

The first two authors contributed equally to this work.

Corresponding author:
Kara K Osbak, Department of Clinical Sciences, Institute of Tropical Medicine, Antwerp, Belgium; Institute of Tropical Medicine, Nationale straat 155, Antwerp 2000, Belgium.
Email: kosbak@hotmail.com

demonstrated a correlation between syphilis and clustering. Recently, studies employing phylogenetic analyses of HIV clinical cohorts have highlighted probable occurrences of sexual behaviour misclassification among males,[5,6] although it is still uncertain if these inferences about transmission linkages and sexual behaviours should be used to guide clinical follow-up or services at the individual level.[7]

With the advent of increasingly user-friendly and robust bioinformatic tools,[8] it is becoming easier for clinicians and epidemiologists to apply these approaches to various cohorts and it is expected that phylogenetic analyses will be increasingly employed to inform and improve public health efforts.[9]

HIV-infected individuals play a disproportionately large role in the current STI epidemics in Europe and elsewhere.[10–12] HIV-infected MSM, in particular, contribute up to 49.8% of those diagnosed with certain STIs such as syphilis.[13] In 2015, an estimated 18,758 individuals in Belgium were HIV-positive (1.7 per 1000 people); of the individuals with a known HIV transmission route in 2016, MSM represented 52% of new HIV infections.[14] A resurgence of several STI epidemics including syphilis, hepatitis C and lymphogranuloma venereum has been observed in our HIV/STI clinics during the last two decades (15–17; V Maes, personal communication, 30/09/2016). For example, during a syphilis epidemic observed at our clinic between 1992 and 2012, over half of infections occurred in patients with a previous diagnosis of syphilis and almost all were HIV-infected MSM.[18] From 2000 to 2015, only 14/895 (1.6%) of syphilis episodes reported at our clinic occurred in females.

In this study, we aimed to ascertain if mapping STI diagnoses onto our local HIV phylogenetic tree could help us to better understand the sexual networks underpinning local STI transmission by investigating the possible role of sexual behaviour misclassification.

Methods

Study participants

A retrospective review of clinical data from HIV-1 infected individuals attending the Institute of Tropical Medicine's (ITM) HIV clinic who had at least one plasma sample submitted for HIV sequencing for DRT from 31 December 1997 to 17 February 2014 was performed. The ITM has a medium-sized HIV-outpatient clinic with 3371 HIV-infected individuals in regular care during the study period. From 2008 onwards, DRT in the form of reverse transcriptase (RT) and protease (PR) gene sequencing was incorporated into the HIV diagnostic workup; prior to 2008 HIV sequencing was performed in the case of treatment failure.

Demographic and clinical data were extracted from our clinical database and pseudonymized. Self-reported HIV transmission risk was recorded by the HIV care physician at the time of diagnosis. Individuals were grouped into categories: 'heterosexual', 'MSM', 'homo/bisexual (females only)' and 'other', which included intravenous drug user (IVDU), mother-to-child transmission, blood transfusion and occupational risk transmissions. The sexual behaviour characteristics of the IVDUs were not specified in the clinical database, thus were regarded as unknown. Sexual preference classification was extrapolated from the transmission risk categories, with the exception of the 'other' classification for which no sexual behaviour group was assigned. MSM were defined as males ever reporting sexual contact with someone of the same sex. All episodes of gonorrhoea, Chlamydia and syphilis that were diagnosed at the ITM laboratory from August 1993 to June 2014 were extracted from the laboratory database for each individual.

STI laboratory testing

STI testing is typically performed at the ITM every 3–12 months for patients in routine HIV follow-up care depending on their risk profile.[19] During follow-up visits, MSM are tested for syphilis, Chlamydia and gonorrhoea, even if they are asymptomatic. *Treponema pallidum* ssp. *pallidum* infection was dual tested with a non-treponemal Macro-Vue rapid plasma reagin (RPR) card test (Becton, Dickinson and Company, United States [US]) and treponemal antibody detection was performed via the Cellognost Syphilis H Combipack *T. pallidum* haemagglutination assay (Dade Behring, Germany) up until November 2002, thereafter testing was performed with the SERODIA *T. pallidum* particle agglutination (TPPA) assay (Fujirebio, Inc., Japan). Syphilis episodes were defined as previously described[18]; briefly, initial syphilis was defined as an RPR titre $\geq 1/8$ and a positive TPPA test on serum. Repeat episodes of syphilis, including syphilis reinfections and reactivations, were defined as an episode in a person who had a $\geq$ 4-fold increase in RPR titre compared to a previous RPR titre.

Neisseria gonorrhoeae (NG) and *Chlamydia trachomatis* (CT) presence in first void urine and/or (anal/pharyngeal/vaginal) swabs were assessed with a nucleic acid amplification test from 1999 onwards for CT and 2010 onwards for NG following manufacturer's instructions. Before 2010 NG presence was tested by culture.

HIV RNA extraction, amplification and sequencing

Viral RNA was extracted from plasma using the QIAamp Viral RNA Mini Kit (QIAGEN, Germany)

following the manufacturer's instructions. The resistance genotyping was performed with the Trugene HIV-1 genotyping kit (Siemens Healthcare Diagnostics, US). The PCR primers covered a 1.3 Kb fragment of the *pol* gene. Full length PR and partial RT sequences were sequenced as previously described.[20] Sequencing products were analysed using the OpenGene DNA Sequencing System (Visible Genetics, Inc., Canada) following manufacturer's instructions. Sequences were stored using the Integrated Database Network System software package (SmartGene, Switzerland).

Sequence analyses

HIV subtyping was performed using REGA version 3.0[21] based on the RT gene since the short PR sequence was found to have insufficient signal for subtyping. Sequence alignments were performed for both the PR and RT genes separately. Sequences were placed in-frame using Mesquite version 3.10[22] and then aligned using the codon model of MUSCLE version 3.7[23] implemented in MEGA7.[24] Trailing sequence ends were trimmed using MEGA7 and then the two gene alignments were concatenated using Mesquite. A maximum likelihood phylogenetic tree of all samples was created using a GTR+Γ model of evolution with each gene as a separate partition and 1000 bootstrap replicates as implemented RAxML version 8.2.9.[25] Patients for whom their multiple samples did not cluster on the phylogenetic tree (patristic branch length distance >0.1) were identified using a custom python script implementing DendroPy.[26] Such large distances are likely the result of either multiple HIV-1 infections or a sample switch error. The above alignment and tree building steps were redone using only the first sample for all individuals. Trees were visualized using FigTree version 1.4.2. Cluster Picker[27] version 1.2.3 was used to identify clusters of all individuals. Clusters were defined as those with >90% bootstrap support for the clade and <4.5% genetic difference between all patient samples. These settings have been found to be optimal for similar data previously.[27] Mixed transmission clusters were defined as clusters consisting of two or more individuals, whereby at least one individual was a heterosexual male who clustered with at least one MSM individual and regardless of the presence of female(s) in the cluster. Potentially misclassified HIV sexual transmission cases were defined as heterosexual males with a history of syphilis who were present in a mixed transmission cluster. We chose a history of syphilis as a predictive factor for possible misclassification of sexual behaviour because syphilis almost exclusively occurs in MSM in our cohort.[18]

Statistical analyses

Continuous variables are expressed in median and interquartile range (IQR). Clustered and non-clustered individuals were compared using Chi square (χ2) test or Fisher's exact test for categorical variables. Bivariate logistic regression analyses using maximum likelihood ratio for the whole cohort (Stage 1 analysis) were performed. Penalized Multinomial Logistic Regression[28,29] analyses were performed using the STATA module 'firthlogit'[30] when stratifying STI (syphilis/Chlamydia/gonorrhoea) diagnosis by HIV transmission risk groups (heterosexual versus MSM) (Stage 2 analysis). Multivariate analyses were not performed due to the small number of individuals in multiple categories. A P-value of ≤0.05 was regarded as significant. All analyses were performed in STATA version 13 (StataCorp LP, College Station, TX, US).

Results

General description of study cohort

In total, 1173 HIV-infected individuals who submitted at least one plasma sample for HIV sequencing out of 3371 patients in follow-up over the study period were initially included in the study. Of the 183 individuals that submitted two samples for DRT, four were excluded from the study because the two samples did not cluster on the phylogram. A total of 1169 individuals were retained in the study. General cohort characteristics are presented in Table 1.

High incidence of STIs among HIV-infected MSM

Syphilis was the most common STI, with a total of 298 episodes reported in 198 individuals. Most (133/198; 67.2%) only had one episode and 65/198 (33.9%) individuals had two or more episodes. Almost all syphilis episodes were reported in MSM (290/298; 97.3%). Chlamydia was the second most common STI with a total of 173 episodes in 131 individuals, mostly occurring in MSM (116/131; 88.5%). With regards to gonorrhoea, 173 episodes were reported in 104 individuals – most were MSM (98/104; 94.2%). When considering all study individuals, 324/1169 (29.3%) had at least one STI episode diagnosed during their HIV follow-up and this increased to 42.8% (299/699) when only considering the MSM population.

HIV transmission clustering analysis reveals a high number of mixed transmission clusters

The HIV phylogram constructed placed a total of 438/1169 (37.5%) study participants into 136 transmission clusters, including 60 larger clusters defined as

Table 1. Characteristics of study cohort of HIV-infected individuals (n = 1169).

Characteristic	n = (%)
Male	914 (78.2)
Female	255 (21.8)
HIV sampling for sequencing performed at time of HIV diagnosis	830 (71.0)
On ART at time of sequencing	924 (79.4)
Birth region	
Europe, North America or Australia	788 (67.4)
Belgium	694 (59.4)
Asia	48 (4.1)
Sub-Saharan Africa	286 (24.5)
North Africa	15 (1.3)
Central and South America	32 (2.7)
HIV subtype	
Subtype B	676 (57.8)
Non-B	493 (42.2)
Self-reported HIV transmission risk category	
Heterosexual total	452 (38.7)
Male	204 (17.5)
Female	248 (21.2)
Homo/bisexual total	703 (60.1)
Male (MSM)	699 (59.8)
Female (homo/bisexual)	4 (0.3)
Other[a]	14 (1.2)
Male	11 (0.01)
Female	3 (0.003)
Other	
Characteristic	Median (interquartile range)
Age at time of sample for HIV sequencing (years)	35.9 (29.4–44.2)
Age at time of HIV diagnosis (years)	34.3 (28.3–41.8)
Years of clinical follow-up[b]	4.9 (3.0–8.7)
Months from HIV diagnosis to ART start	8.1 (1.4–27.6)
CD4+ T-cell count nadir	295 (186–414)
CD4+ T-cell count at HIV diagnosis	419 (266–610)
Years of follow-up heterosexual risk group (all participants)	5.7 (3.2–10.0)
Years of follow-up homo/bisexual risk group (all participants)	4.7 (2.9–7.9)
Years of follow-up heterosexual risk group (male only)	5.4 (2.5–10.2)
Years of follow-up MSM	4.7 (2.9–7.8)

ART: antiretroviral therapy; MSM: men who have sex with men.

[a]'Other' transmission risk category includes: n = 8 intravenous drug users, n = 3 mother-to-child transmission, n = 1 occupational and n = 2 blood transfusion HIV transmissions.

[b]Calculated from date of initial HIV diagnosis to end of study inclusion date.

three or more individuals and 76 transmission pairs (also defined as clusters) (Table 2; Figure 1). Larger clusters contained a median of four individuals (IQR 3–6) with a maximum of 17 individuals (n = 1 cluster; MSM only). Details of the clusters defined based on the HIV transmission risk category of the individuals are reported in Table 2, including 24 clusters (containing 94 individuals) that were mixed transmission clusters.

Table 3 outlines various bivariate associations with HIV transmission clustering. There was a strong positive association between male gender (odds ratio [OR] 4.96 [CI 3.4–7.3]; P < 0.001), HIV B-subtype (OR 4.7 [CI 3.6–6.1]; P < 0.001), HIV sequencing performed at

diagnosis (OR 2.3 [CI 1.7–3.0]; P < 0.001) and MSM transmission (OR 3.9 [CI 3.0–5.2]; P < 0.001) with transmission clustering.

Those diagnosed with syphilis were more likely to be in a cluster (overall association; OR 2.3 [CI 1.7–3.2]; P < 0.001). No syphilis episodes were reported in heterosexual only clusters (Table 2). When the analysis was stratified by sexual transmission categories ('MSM' and 'heterosexual'), clustering of syphilis only remained significant in the heterosexual group (Penalized Multinomial Logistic Regression OR 18.5 [CI 3.1–111.1]; P < 0.001). Individuals with a diagnosis of Chlamydia and gonorrhoea were also more likely to

Table 2. HIV transmission clusters identified by Cluster Picker and STI history of individuals in each cluster category.

	Total n = clusters	Total n = individuals	MSM only clusters	Heterosexual only clusters	MSM/heterosexual mixed transmission clusters[a]	'Other' mix clusters
Clusters[b]	136	438	83 (283)	27 (57)	24 (94)	2[c]
Cluster pairs	76	152	42 (84)	24 (48)	8 (16)	2 (4)
Clusters with ≥3 individuals	60	286	41 (199)	3 (9)	16 (78)	NA
STI[d]	Total n = clusters	Total n = individuals	Total n = individuals	Total n = individuals	Total n = individuals	Total n = individuals
Syphilis	NA	198	86	0	22	0
Chlamydia	NA	131	55	0	6	0
Gonorrhoea	NA	104	45	1	4	0

NA: not applicable; STI: sexually transmitted infection.

All data presented in n = clusters (n = individuals) unless otherwise noted.

[a]MSM/heterosexual mix transmission clusters defined as clusters consisting of two or more individuals, whereby at least one individual was a heterosexual male who clustered with at least one MSM individual and regardless of the presence of female(s) in the cluster.

[b]Clusters defined as those with >90% bootstrap support for the clade and <4.5% genetic difference between all individual samples as determined by Cluster Picker version 1.2.3.

[c]Two transmission pairs: (1) male intravenous drug user (IVDU) of unknown sexual behaviour + heterosexual female; (2) Male IVDU of unknown sexual behaviour + heterosexual male.

[d]Individuals ever diagnosed with an episode of Chlamydia, gonorrhoea or syphilis at the ITM laboratory from 1993 to 2014.
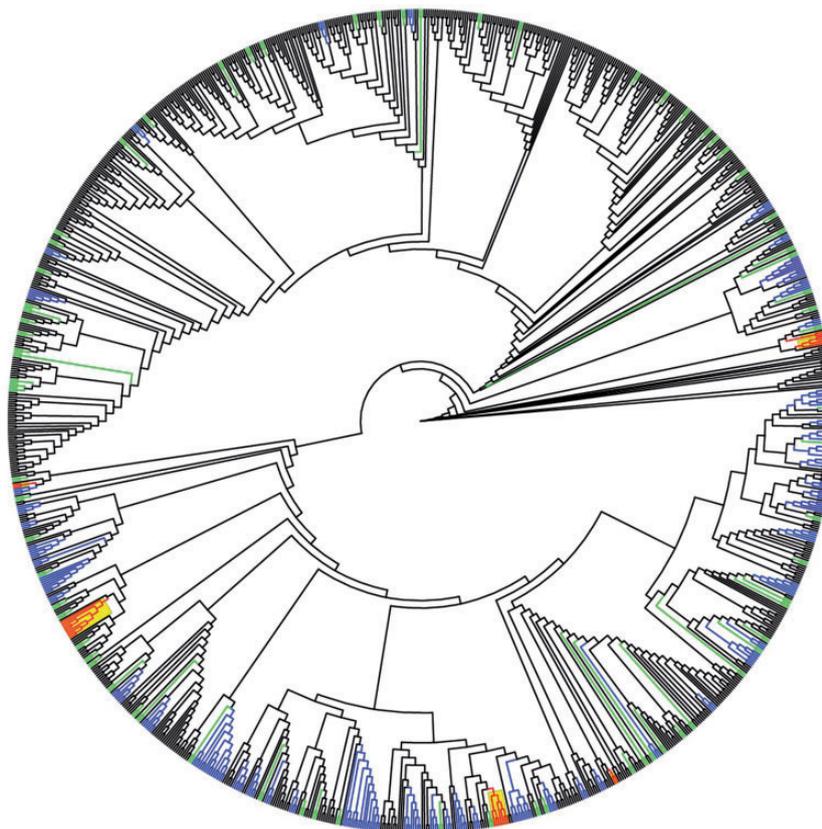

Figure 1. HIV transmission cladogram of 1169 individuals in this study. Clusters (n = 136) are depicted in colour: blue represents clusters containing ≥3 individuals, green represents cluster pairs and red/yellow depicts mixed transmission clusters with possible misclassified HIV sexual transmission among heterosexual males

Table 3. Characteristics of HIV-infected individuals who were identified in HIV transmission clusters compared to non-clustering individuals.

	Total n = (%)	In transmission cluster[a] n = (%)	Not in transmission cluster[a] n = (%)	Odds ratio (95% CI)[b]
Whole study cohort	1169	438 (37.5)	731 (62.5)	NA
Gender				
Female	255 (21.8)	35 (13.7)	220 (86.3)	Ref
Male	914 (78.2)	403 (44.1)	511 (55.9)	4.7 (3.4–7.3)**
HIV subtype				
Non-subtype B	493 (42.2)	91 (18.5)	402 (81.5)	Ref
Subtype B	676 (57.8)	347 (51.3)	329 (48.7)	4.7 (3.6–6.1)**
Belgian origin				
No	475 (40.6)	92 (19.4)	383 (80.6)	Ref
Yes	694 (59.4)	346 (49.9)	348 (50.1)	4.1 (3.2–5.4)**
Sub-Saharan African origin				
No	883 (75.5)	397 (45.0)	486 (55.0)	Ref
Yes	286 (24.5)	41 (14.3)	245 (85.7)	0.2 (0.1–0.3)**
Other region origin				
No	980 (93.8)	387 (39.5)	593 (60.5)	Ref
Yes	189 (16.2)	51 (27.0)	138 (73.0)	0.6 (0.4–0.8)**
Age at time of sampling for sequencing[c]				
< 25 years	112 (9.6)	55 (49.1)	57 (50.9)	Ref
≥ 25 years	1057 (90.4)	383 (36.2)	674 (63.8)	0.5 (0.4–0.8)*
HIV sampling for sequencing performed at time of diagnosis				
No	339 (29.0)	84 (24.8)	255 (75.2)	Ref
Yes	830 (71.0)	354 (42.7)	476 (57.3)	2.3 (1.7–3.0)**
Sexual behaviour[c]				
Heterosexual	452 (39.2)	90 (19.9)	362 (80.1)	Ref
MSM	699 (77.4)	346 (49.5)	353 (50.5)	3.9 (3.0–5.2)**
Homo/bisexual female	4 (0.3)	0 (0)	4 (100)	NA[d]
Syphilis overall association[c,e]				
No	971 (83.1)	330 (34.0)	641 (66.0)	Ref
Yes	198 (16.9)	108 (54.5)	90 (45.5)	2.3 (1.7–3.2)**
Syphilis MSM only[c,e]				
No	509 (72.8)	244 (47.9)	265 (52.1)	Ref
Yes	190 (27.2)	102 (53.7)	88 (46.3)	1.3 (0.9–1.8)
Syphilis heterosexual only[c,e]				
No	455 (98.5)	84 (18.9)	361 (81.1)	Ref
Yes	7 (1.5)	6 (85.7)	1 (14.3)	18.5 (3.1–111.1)**
Chlamydia overall association[c,e]				
No	1038 (88.8)	337 (36.3)	661 (63.7)	Ref
Yes	131 (11.2)	61 (46.6)	70 (53.4)	1.6 (1.1–2.2)*
Chlamydia MSM only[c,e]				
No	583 (83.4)	286 (49.1)	297 (50.1)	Ref
Yes	116 (16.6)	60 (51.7)	56 (48.3)	1.1 (0.7–1.6)
Chlamydia heterosexual only[c,e]				
No	439 (97.1)	89 (20.3)	350 (79.7)	Ref
Yes	13 (2.9)	1 (7.7)	12 (92.3)	0.5 (0.1–2.5)
Gonorrhoea overall association[c,e]				
No	1065 (91.1)	388 (36.4)	677 (63.6)	Ref
Yes	104 (8.9)	50 (48.1)	54 (51.9)	1.7 (1.1–2.5)*
Gonorrhoea MSM only[c,e]				
No	601 (86.0)	298 (49.6)	303 (50.4)	Ref
Yes	98 (14.0)	48 (49.0)	50 (51.0)	1.0 (0.6–1.5)
Gonorrhoea heterosexual only[c,e]				
No	452 (99.1)	88 (19.6)	360 (80.4)	
Yes	4 (0.09)	2 (50.0)	2 (50.0)	NA[d]

CI: confidence interval; MSM: men who have sex with men; NA: not applicable; P-value and odds ratio values are for comparison between clustered and non-clustered groups (clustered code = '1' in bivariate analysis; non-clustered = '0'); Ref: reference comparison group for bivariate analysis; Column 1 is presented as column percentages and column 2 and 3 are presented as row percentages.

[a]Clusters defined as those with >90% bootstrap support for the clade and <4.5% genetic difference between all individual samples as determined by Cluster Picker version 1.2.3.

[b]* P < 0.05, ** P < 0.005.

[c]Calculated excluding the n = 14 'other' (non-sexual) transmission category individuals and n = 4 homo/bisexual females.

[d]Not calculated due to low case number.

[e]Individuals ever diagnosed at the ITM with an episode of syphilis, gonorrhoea or Chlamydia during 1993–2014.

cluster; however, when stratified by transmission category, this relationship was no longer significant.

Six possible cases of HIV sexual transmission misclassification

Twenty-five heterosexual males were present in mixed transmission clusters, of which 6/25 had a diagnosis of syphilis and 5/6 clustered exclusively with other males. According to our definition, these males were categorized as possible HIV sexual transmission misclassifications. We estimate the misclassification rate in our cohort to be 0.7% of total males (6/914) and 3% of total heterosexual males (6/204). When only considering males in clusters, which are the only individuals assessed for misclassification, this rate increased to 1.5% (6/403) for all males and 11% for heterosexual males (6/55). Figure 2 is a graphical depiction of these six possible misclassified heterosexual males.

Discussion

This study exemplifies the utility of coupling phylogenetic analyses of HIV collected through DRT with clinical data to gain insights into local HIV transmission network characteristics and the degree of possible sexual behaviour misclassification of MSM.

Simplified bioinformatic analyses revealed 136 transmission clusters encompassing 37.5% of the study cohort. STIs were highly prevalent among MSM, with almost half (42.5%) having a diagnosis of at least one STI episode. Individuals with a STI diagnosis were more likely to cluster in the close genetic proximity of one another. However, when analyses were stratified per risk transmission category, this relationship was attenuated and only remained significant for heterosexual males with a history of syphilis. This finding was unexpected since almost all syphilis infections in our cohort occur in MSM.[18] The phylogenetic analyses provided a possible explanation for this finding, the fact that five out of six of these men were in clusters that only contained MSM suggested that they might have been misclassified as non-MSM.

A recent HIV phylogenetic study in the United Kingdom estimated that 18.6% (249/1341) of all clustered heterosexual men were potential non-disclosed MSM.[6] This estimation is higher than the 11% suspected misclassifications in our cohort, a discrepancy that could be attributed to the larger national population study cohort that would likely cover sexual networks more extensively than our single-centre analysis. Further sources of misclassification could be that doctors in our clinic are not taking accurate sexual histories and clinical databases are incorrectly filled in
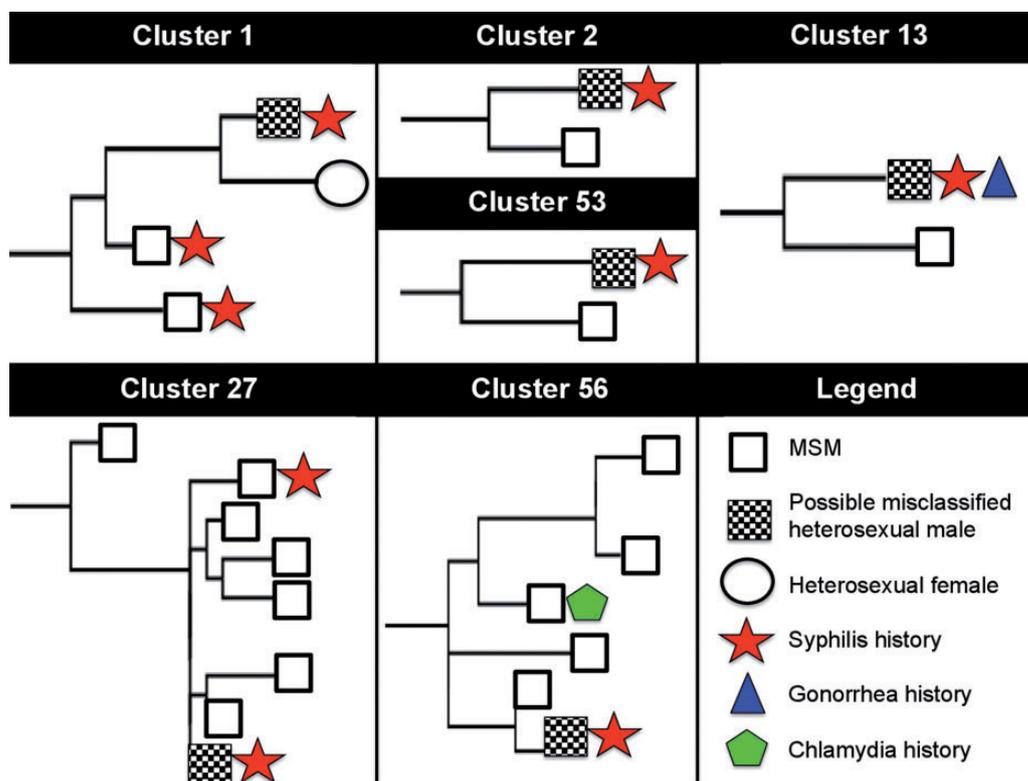

Figure 2. Six HIV transmission clusters containing possible misclassified HIV sexual transmission heterosexual males with a history of syphilis. MSM: men who have sex with men.

and/or not regularly updated if new patient information is obtained.

There are a number of caveats to this analysis, for example the single institution setting and retrospective nature of the study. Only 34.6% of HIV-infected persons in our cohort were included in this study. Nonetheless, approximately a third of study participants still clustered with other individuals, indicating a moderate coverage of the local HIV transmission network. One must also take into account the high diversity of the cohort population. A quarter of the participants originated from sub-Saharan Africa; they were less likely to cluster (P < 0.001) with others, which could be indicative of transmission events outside the cohort or even the region. Moreover, the long study period spanning 17 years means that individuals diagnosed and sequenced early in the study period who then went on to transmit HIV years later within the same cohort may not be genetically linked due to the continued viral evolution that would slowly abolish phylogenetic linkage.[31] This could explain why individuals who were sequenced shortly after diagnosis were more likely to cluster in close proximity to one another compared to those who were sequenced later (OR 2.3 [1.7–3.0]). Participants with more recent HIV infections and shorter clinical follow-up had less STI episode information available (data not shown) and this could have contributed to a lower misclassification rate for those without (reported) STI episodes. Moreover, the STI episode rate for the MSM participants in this study could be underestimated since the mean follow-up time of the MSM group was significantly shorter (P < 0.001) than the heterosexual transmission group.

Our phylogenetic tree is only based on HIV transmission. Since compared to other STIs, HIV is relatively inefficiently transmitted sexually,[32] our tree likely underestimates the sexual network connectivity of individuals therein. Phylogenetic analyses are limited by the fact that they often cannot assign directionality, thus it cannot be ruled out that the one female clustered with the possible misclassified transmission individuals could have been the transmission link instead of a MSM in the same cluster. The number of individuals defined as possibly misclassified was very small; a Penalized Multinomial Logistic Regression analysis was performed in an attempt to compensate for rare event biases[28,29]; however, the sample size remains small and multivariate analytical approaches were not feasible.

Increased use of phylogenetic analysis brings with it important ethical, security and privacy considerations that need to be addressed.[9] With the further refinement of accessible bioinformatics tools, individual smaller centres will be able to use these methods to define and characterize their cohorts. Furthermore, patient care would benefit from standardized and optimized approaches to sexual behaviour reporting. A recent study by Haider et al.[33] investigating patient-centred approaches to sexual behaviour data collection found that patients were willing to be asked about their sexual behaviour, were likely not to take offense at such questioning and both patients and clinicians prefer nonverbal data collection methods.

Conclusions

In our setting the superimposition of STIs on the HIV phylogram provided insights into how sexual behaviour could have been misclassified for some individuals. This has important implications for patient care and understanding local STI transmission characteristics since these individuals could operate as a bridge population into otherwise low-risk groups. With regards to patient care, it is crucial to know if individuals are MSM as they require different anal neoplasia and STI screening[19] than other populations.

Authors' contributions

Study design: CRK, CJM, AT, KKO; Cohort study and sample collection EF, ME; Sequence analysis CJM; Statistical analysis: KKO, CRK, AT; Draft manuscript preparation: KKO, CJM; CRK; Final manuscript edit and approval: CRK, CJM, EF, ME, KKO, LH, KKA, KF, SGR, AT.

Availability of data and material

The datasets generated and analysed during the current study are not publicly available due to them containing information that could compromise research participant privacy but are available from the corresponding author (CRK) on reasonable request.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID iD

Kara K Osbak http://orcid.org/0000-0002-8612-8413

References

1. Chalmet K, Staelens D, Blot S, et al. Epidemiological study of phylogenetic transmission clusters in a local HIV-1 epidemic reveals distinct differences between subtype B and non-B infections. *BMC Infect Dis* 2010; 10: 262.
2. Mehta SR, Wertheim JO, Brouwer KC, et al. HIV transmission networks in the San Diego–Tijuana border region. *EBioMedicine* 2015; 2: 1456–1463.
3. Wertheim JO, Leigh Brown AJ, Hepler NL, et al. The global transmission network of HIV-1. *J Infect Dis* 2014; 209: 304–313.
4. Bezemer D, Cori A, Ratmann O, et al. Dispersion of the HIV-1 epidemic in men who have sex with men in the Netherlands: a combined mathematical model and phylogenetic analysis. *PLoS Med* 2015; 12: e1001898.
5. Hué S, Brown AE, Ragonnet-Cronin M, et al. Phylogenetic analyses reveal HIV-1 infections between men misclassified as heterosexual transmissions. *AIDS* 2014; 28: 1967–1975.
6. Ragonnet-Cronin M, Hué S, Hodcroft EB, et al. Non-disclosed men who have sex with men in UK HIV transmission networks: phylogenetic analysis of surveillance data. *Lancet HIV* 2018; 5: e309–e316.
7. Sullivan PS and Oster AM. Phylogenetic analyses of local HIV transmission. *Lancet HIV* 2018; 5: e270–e271.
8. Lam TT-Y, Hon C-C and Tang JW. Use of phylogenetics in the molecular epidemiology and evolutionary studies of viral infections. *Crit Rev Clin Lab Sci* 2010; 47: 5–49.
9. German D, Grabowski MK and Beyrer C. Enhanced use of phylogenetic data to inform public health approaches to HIV among men who have sex with men. *Sex Health* 2016; 14: 89.
10. Dougan S, Evans BG and Elford J. Sexually transmitted infections in Western Europe among HIV-positive men who have sex with men. *Sexually Transmitted Diseases* 34 (10): 783–790, OCT 2007.
11. Roberts CP and Klausner JD. Global challenges in human immunodeficiency virus and syphilis coinfection among men who have sex with men. *Expert Rev Anti Infect Ther* 2016; 14: 1037–1046.
12. Abara WE, Hess KL, Neblett Fanfair R, et al. Syphilis trends among men who have sex with men in the United States and Western Europe: a systematic review of trend studies published between 2004 and 2015. *PLoS One* 2016; 11: e0159309.
13. Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance 2015. Atlanta: U.S. Department of Health and Human Services; 2016.
14. Belgian Institute of Public Health (IPH). Epidemiology of AIDS and HIV-infection in Belgium, 2016.
15. Bottieau E, Apers L, Van Esbroeck M, et al. Hepatitis C virus infection in HIV-infected men who have sex with men: sustained rising incidence in Antwerp, Belgium, 2001–2009. *Euro Surveill* 2010; 15: 19673.
16. Apers L, Koole O, Bottieau E, et al. Incidence of HCV and sexually transmitted diseases among HIV positive MSM in Antwerp, Belgium, 2001–2011. *Acta Clin Belg* 2013; 68: 421–426.
17. Vandenbruaene M, Ostyn B, Crucitti T, et al. Lymphogranuloma venereum outbreak in men who have sex with men (MSM) in Belgium, January 2004 to July 2005. *Euro Surveill* 2005; 10: E050929.3.
18. Kenyon C, Lynen L, Florence E, et al. Syphilis reinfections pose problems for syphilis diagnosis in Antwerp, Belgium – 1992 to 2012. *Euro Surveill* 2014; 19: 20958.
19. Workowski KA and Bolan GA. Centers for disease control and prevention. Sexually transmitted diseases treatment guidelines, 2015. *MMWR Recomm Rep* 2015; 64: 1–137.
20. Ribas SG, Heyndrickx L, Ondoa P, et al. Performance evaluation of the two protease sequencing primers of the Trugene HIV-1 genotyping kit. *J Virol Methods* 2006; 135: 137–142.
21. Pineda-Peña A-C, Faria NR, Imbrechts S, et al. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: performance evaluation of the new REGA version 3 and seven other tools. *Infect Genet Evol* 2013; 19: 337–348.
22. Maddison, W. P. and D.R. Maddison. 2018. Mesquite: a modular system for evolutionary analysis. http://www.mesquiteproject.org.
23. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004; 32: 1792–1797.
24. Kumar S, Stecher G and Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016; 33: 1870–1874.
25. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014; 30: 1312–1313.
26. Sukumaran J and Holder MT. DendroPy: a python library for phylogenetic computing. *Bioinformatics* 2010; 26: 1569–1571.
27. Ragonnet-Cronin M, Hodcroft E, Hué S, et al. Automated analysis of phylogenetic clusters. *BMC Bioinformatics* 2013; 14: 317.

28. Firth D. Bias reduction of maximum likelihood esti-
    mates. *Biometrika* 1993; 80: 27–38.
29. Heinze G and Schemper M. A solution to the problem of
    separation in logistic regression. *Stat Med* 2002;
    21: 2409–2419.
30. Coveney J. FIRTHLOGIT: Stata module to calculate
    bias reduction in logistic regression. https://econpapers.
    repec.org/software/bocbocode/s456948.htm   (Accessed
    April 15, 2016).
31. Castro-Nallar E, Pérez-Losada M, Burton GF, et al. The
    evolution of HIV: inferences using phylogenetics. *Mol
    Phylogenet Evol* 2012; 62: 777–792.
32. Boily M-C, Baggaley RF, Wang L, et al. Heterosexual
    risk of HIV-1 infection per sexual act: systematic review
    and meta-analysis of observational studies. *Lancet Infect
    Dis* 2009; 9: 118–129.
33. Haider AH, Schneider EB, Kodadek LM, et al.
    Emergency department query for patient-centered
    approaches to sexual orientation and gender identity.
    JAMA Intern Med 2017; 177: 819.