HIV-1 transmission between MSM and heterosexuals, and increasing proportions of circulating recombinant forms in the Nordic Countries

Joakim Esbjörnsson,[1,2,3,*] Mattias Mild,[4] Anne Audelin,[5] Jannik Fonager,[5]
Helena Skar,[6] Louise Bruun Jørgensen,[5] Kirsi Liitsola,[7] Per Björkman,[8]
Göran Bratt,[9] Magnus Gisslén,[10] Anders Sönnerborg,[11,12,13] Claus Nielsen,[5]
SPREAD/ESAR Programme, Patrik Medstrand,[14] and Jan Albert[1,11]

[1]Department of Microbiology Tumor and Cell Biology, Karolinska Institute, Stockholm, Sweden, [2]Nuffield Department of Medicine, University of Oxford, Oxford, UK, [3]REGA Institute, Katholieke Universiteit, Leuven, Belgium, [4]Department of Microbiology, Public Health Agency of Sweden, Stockholm, Sweden, [5]Department of Microbiological Diagnostics and Virology, Statens Serum Institut, Copenhagen, Denmark, [6]Department of Science and Technology, Linköping University, Campus Norrköping, Norrköping, Sweden, [7]Department of Infectious Diseases, National Institute for Health and Welfare, Helsinki, Finland, [8]Department of Clinical Sciences Malmö, Lund University, Malmö, Sweden, [9]Department of Clinical Science and Education, Venhälsan, Stockholm South General Hospital, Stockholm, Sweden, [10]Department of Infectious Diseases, Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden, [11]Department of Clinical Microbiology, Karolinska University Hospital, Stockholm, Sweden, [12]Division of Clinical Microbiology, Karolinska Institute, Stockholm, Sweden, [13]Department of Infectious Diseases, Karolinska University Hospital, Stockholm, Sweden and [14]Department of Translational Medicine, Lund University, Malmö, Sweden

*Corresponding author: E-mail: joakim.esbjornsson@ki.se

Abstract

Increased knowledge about HIV-1 transmission dynamics in different transmission groups and geographical regions is fundamental for assessing and designing prevention efforts against HIV-1 spread. Since the first reported cases of HIV infection during the early 1980s, the HIV-1 epidemic in the Nordic countries has been dominated by HIV-1 subtype B and MSM transmission. HIV-1 *pol* sequences and clinical data of 51 per cent of all newly diagnosed HIV-1 infections in Sweden, Denmark, and Finland in the period 2000–2012 ($N = 3,802$) were analysed together with a large reference sequence dataset ($N = 4,537$) by trend analysis and phylogenetics. Analysis of the eight dominating subtypes and CRFs in the Nordic countries (A, B, C, D, G, CRF01_AE, CRF02_AG, and CRF06_cpx) showed that the subtype B proportion decreased while the CRF proportion increased over the study period. A majority (57 per cent) of the Nordic sequences formed transmission clusters, with evidence of mixing both geographically and between transmission groups. Detailed analyses showed multiple occasions of transmissions from MSM to heterosexuals and that active transmission clusters more often involved single than multiple Nordic countries. The strongest geographical link was between Denmark and Sweden. Finally, Denmark had a larger proportion of heterosexual domestic spread of HIV-1 subtype B (75 per cent) compared with Sweden (49 per cent) and Finland

(57 per cent). We describe different HIV-1 transmission patterns between countries and transmission groups in a large geographical region. Our results may have implications for public health interventions in targeting HIV-1 transmission networks and identifying where to introduce such interventions.

Key words: HIV-1; subtype; transmission; molecular epidemiology; phylogeny.

1. Introduction

The HIV-1 epidemic in the three Nordic countries Sweden, Denmark, and Finland has been relatively stable over the last decade with low prevalence numbers of <0.2 per cent. The total number of newly diagnosed HIV-1 infections have been ~440, 270, and 160 cases per year for Sweden, Denmark, and Finland, respectively (EpiNorth; UNAIDS). The most prevalent group of HIV-1 is the main (M) group which has been divided into subtypes (A–D, F–H, J–K), sub-subtypes (A1–A4, F1–F2), and seventy circulating recombinant forms (CRFs), distinguished at both genetic level and geographic location (Los Alamos Sequence Database). The early HIV-1 epidemic in the Nordic countries, such as in North America and the rest of Western Europe, was dominated by subtype B infections among men who have sex with men (MSM) (Lukashov et al. 1996; Sonnerborg et al. 1997; Liitsola et al. 2000; Bezemer et al. 2010a; Karlsson et al. 2012; Abecasis et al. 2013; Sweden PHAo). However, less is known about changes in prevalence and dynamics of non-subtype B in the Nordic countries. This may be of particular importance since increasing evidence suggests that there could be differences in infectivity and pathogenicity between different strains and recombinants of HIV-1 (Renjifo et al. 2004; Arien et al. 2005; Esbjornsson et al. 2010; Kiwanuka et al. 2010; Morrison et al. 2010; Palm et al. 2013). The impact of these findings on HIV-1 transmission and spread on the population scale is largely unknown. Phylogenetic analysis has successfully been used to identify and dissect HIV-1 transmission clusters, and when combined with detailed epidemiological and clinical data the results may be of considerable public health relevance, e.g. identifying mixing across transmission, demographic, and behavioural subgroups (Fisher et al. 2010; Aldous et al. 2012; Brenner, Wainberg, and Roger 2013; Grabowski et al. 2014; Wertheim et al. 2014; Frost and Pillay 2015; Poon et al. 2015). The objective of this study was to identify HIV-1 introductions and transmission links within and between three Nordic countries using a large sequence dataset representing half of all newly diagnosed HIV-1 infections in these countries in the period of 2000–2012.

2. Methods

2.1. Dataset

Overall, 3,802 HIV-1 *pol* sequences (~1,000 bp) from Sweden, Denmark, and Finland were included in the analyses representing 51 per cent of all newly diagnosed HIV-1 infections over the study period [Sweden, 2002–2010, 1,538 sequences (44 per cent); Denmark, 2000–2012, 1,795 sequences (57 per cent); Finland, 2003–2009, 469 sequences (54 per cent)]. The sequences were collected as part of the EU project SPREAD/ESAR (http://www.esar-society.eu) from individuals no longer than 6 months after HIV-1 diagnosis (SPREAD programme 2008; Vercauteren et al. 2009; Karlsson et al. 2012). To increase the resolution of phylogenetic reconstructions, sequences from an additional 201 MSM diagnosed with HIV-1 1992–2002 in Sweden were included in the phylogenetic analysis (three subtype A, 194 subtype B, two subtype C, and two CRF01_AE) (Lindstrom et al. 2006). Information about sample collection date, sampling country, self-reported country of infection, gender, and route of transmission [heterosexual transmission (HET); MSM; intravenous drug use (IDU); unknown] were used in subsequent analyses (SPREAD programme 2008). If more than one possible route of transmission was reported, the transmission route was determined according to the following hierarchy: IDU > MSM > HET > Other. All recruited patients were antiretroviral drug naïve at the time of sampling.

2.2. Subtype determination and cluster identification

Details on these analyses can be found in Supplementary Data. In brief, the subtype was determined by maximum-likelihood phylogenies using the HIV-1 subtype reference dataset from Los Alamos Sequence Database. To identify HIV-1 transmission clusters among the most prevalent subtypes and CRFs in the Nordic countries, we used a BLAST-approach to construct a reference sequence dataset from both Genbank and a pan-European dataset kindly provided by SPREAD/ESAR (Altschul et al. 1997). In contrast to the NCBI dataset, sequences from the SPREAD/ESAR dataset had epidemiological information about collection date, sampling country, country of infection, and route of transmission. Each subtype/CRF was analysed separately. A Nordic transmission cluster was defined based on the fulfilment of two criteria: (1) The most basally located and statistically supported cluster node (aLRT-SH >0.90) (Anisimova et al. 2011) in each subtype/CRF-specific phylogeny was identified to maximize the probability of including large transmission networks and identifying links between transmission groups and Nordic countries. (2) If a supported cluster contained ≥80 per cent Nordic sequences, it was defined as a Nordic transmission cluster. If the cluster contained <80 per cent of Nordic sequences it was *not* considered to be a Nordic transmission cluster and the cluster search continued towards the tips of the tree. Nordic clusters with ≥80 per cent of sequences from a particular Nordic country, e.g. Sweden, or of a particular transmission route, e.g. heterosexual, were defined as Swedish and heterosexual clusters, respectively. Clusters with ≥80 per cent Nordic sequences, but <80 per cent from one specific Nordic country and transmission group, were defined as Nordic clusters with mixed transmission route. Clusters of two sequences were defined as dyads, 3–14 sequences as networks, and >14 sequences as large clusters (Aldous et al. 2012). Clusters containing at least one sequence collected ≤2 years before the end of the study period were defined as active transmission clusters. The index sequence was defined as the first collected sequence in a cluster.

2.3. Statistics

The linear-by-linear (LBL) association test, with each year treated independently, was used to analyse trends over time. Bonferroni correction for multiple comparisons was performed

when necessary. Proportions were compared using the 2-tailed Fisher's exact test (FET) and continuous parameters between groups were compared using the Mann–Whitney *U* test (M-W). Statistical analysis was performed using IBM SPSS Statistics for Windows, Version 21.0. (IBM Corp., Armonk, NY).

2.4. Ethics

Informed written or oral consent was obtained from all study participants. The research was approved by the Medical Ethics Committee at the Karolinska Insitutet, Stockholm, Sweden (Dnr 02-367, 04-797 and 2007/1533). Danish data were collected, stored, and analysed as approved by the Danish data protection agency (J. nr. 2015-41-3744). Finnish sequences from the National Microbe Strain Collection was collected based on the Communicable Diseases Act and the Communicable Diseases Decree.

2.5. Nucleotide sequence accession numbers

For similar scientific and ethical reasons as explained in Alizon et al. (2010) and Kouyos et al. (2010), only a proportion of the anonymized sequences is accessible via GenBank (accession numbers, JQ698667–JQ698874). In brief, the sequences analysed in this study constitute a dataset that represents a large proportion of HIV-1 patients in the Nordic region and thereby, in principle, allow for the reconstruction of direct and indirect transmission links. Inappropriate use of the data could thereby endanger the privacy of the patients, which is especially problematic because HIV-1 sequences frequently have been used in court cases. Furthermore, and from a scientific point of view, the consequences of open and uncontrolled access to such densely sampled sequences could jeopardize the future publication (and, thus, the investigation) of similar dense datasets and could thereby be counterproductive even from an "open-access" perspective (Kouyos et al. 2010). However, the entire dataset can be used for well-defined projects that have passed ethical clearance, are in accordance with the guidelines of the cohorts, and is approved by the scientific board of SPREAD/ESAR.

3. Results

3.1. HIV-1 subtype B decreased over time in the Nordic countries

Phylogenetic analyses showed that at least thirty-two subtypes, sub-subtypes, and CRFs were present in the Nordic countries in the period 2000–2012. We also identified forty-seven unique recombinant forms (URFs) and twenty-four sequences that could not be determined. Eight subtypes and CRFs dominated the Nordic epidemic and represented 96 per cent of the dataset (subtypes A, B, C, D, and G, and the CRFs 01_AE, 02_AG, and 06_cpx, Table 1). Most of the remaining sequences (77 per cent) were different recombinant forms and were analysed as one group. Subtype B was the most common variant with an overall proportion of 54 per cent. Trend analysis showed a decrease in the proportion of subtype B in the Nordic countries from 58 to 48 per cent 2003–2009 ($P < 0.001$, LBL) (Supplementary Fig. 1B). Stratified analysis by country confirmed this trend in Sweden (58–34 per cent, 2002–2010, $P < 0.001$, LBL) and Finland (56–49 per cent, 2003–2008, $P = 0.018$, LBL), but not in Denmark (61–57 per cent, 2001–2011, $P = 0.22$, LBL). To study if these differences could be related to differences in country of origin, we stratified

Table 1. Proportions of subtype/CRF and transmission route of the sequences in the Nordic dataset of newly diagnosed individuals in the period 2000–2012

Category	Number of sequences							
	Sweden		Denmark		Finland		Total	
Subtype/CRF								
A	136	9%	108	6%	9	2%	253	7%
B	639	42%	1179	66%	250	53%	2068	54%
C	238	15%	145	8%	33	7%	416	11%
D	26	2%	36	2%	8	2%	70	2%
G	19	1%	26	1%	8	2%	53	1%
CRF01_AE	287	19%	126	7%	103	22%	516	14%
CRF02_AG	117	8%	85	5%	20	4%	222	6%
CRF06_cpx	10	1%	11	1%	29	6%	50	1%
Other	66	4%	79	4%	9	2%	154	4%
Transmission route								
HET	788	51%	665	37%	190	41%	1643	43%
MSM	564	37%	863	48%	201	43%	1628	43%
IDU	145	9%	108	6%	25	5%	278	7%
Unknown	41	3%	159	9%	53	11%	253	7%
Total per category	1538		1795		469		3802	

CRF, circulating recombinant form; HET, Heterosexual; MSM, Men who have sex with men; IDU, Intravenous drug use.

the Nordic dataset by sequences obtained from patients reported to originate from sub-Saharan Africa vs. non-sub-Saharan Africa (since 71 per cent of all global HIV infections are found in sub-Saharan Africa) (see UNAIDS). The proportion of sequences from patients with sub-Saharan African origin was 26 per cent in Sweden, 16 per cent Denmark, and 8 per cent in Finland. No significant time trends in the proportion of sequences from patients with sub-Saharan African origin were found in the different Nordic countries.

3.2. Heterosexual transmission dominated among non-subtype B-infected individuals

In agreement with national surveillance data (EpiNorth), heterosexual (43 per cent) and MSM (43 per cent) transmission were the most common transmission routes (Table 1). However, analysis by country showed that HET was more common than MSM transmission in Sweden (51 vs. 37 per cent), whereas the opposite was observed for Denmark (heterosexual 37 per cent vs. MSM 48 per cent) (Table 1). The percentages of heterosexual and MSM transmission were similar in Finland (41 vs. 43 per cent). The overall proportion of IDU transmission was 7 per cent (range 5–9 per cent). No significant time trends of different transmission routes were found for the complete Nordic dataset, and country-specific trends were in agreement with national surveillance data (see Supplementary Data). Stratified analysis of transmission route dynamics for different subtypes and CRFs showed stable trends in general (Fig. 1 and Supplementary Data). However, closer inspection of the statistically supported changes (as highlighted by arrows in Fig. 1) showed that the significant decreasing proportions among heterosexuals was observed only among different CRFs (only stable trends was observed among the analysed subtypes). In the MSM group, exclusively increasing or stable proportions was observed among both subtypes and CRFs. In contrast, IDUs

Figure 1. Summary of time trends in the major subtypes/CRFs and transmission groups in the analyzed countries. Time trends (as determined by the LBL test for trend) shown for the major subtypes/CRFs (green, increasing trend; red, decreasing trend) and transmission groups (↑, increasing trend; ↓, decreasing trend). Only statistically significant trends are highlighted in the figure, and white areas or no arrows indicate stable or non-significant trends. All minor variants (mainly consisting of different CRFs and URFs) in the Nordic countries (i.e. Sweden, Denmark, or Finland) were analyzed as one group. Details on these analyses and results are presented in the Supplementary data.

showed exclusively decreasing proportions for subtypes and increasing proportions for CRFs.

The subtype-specific analysis showed that MSM was the major transmission route for subtype B (74 per cent whereas HET was more common for the other major subtypes and CRFs in the Nordic countries (subtype A, 82 per cent; C, 83 per cent; D, 86 per cent; G, 81 per cent; CRF01_AE, 67 per cent; CRF02_AG, 77 per cent; and CRF06_cpx, 58 per cent). When subtype B and MSM were dissected by country, Sweden and Finland had larger proportions of MSM with subtype B infections (76 and 76 per cent) than Denmark (68 per cent , $P < 0.001$ and $P = 0.01$, FET). Analysis of country of infection among heterosexuals showed a larger extent of domestically vs. internationally acquired subtype B infection in Denmark (75 per cent domestically acquired subtype B infection) compared with Sweden (49 per cent, $P < 0.001$, FET) and Finland (57 per cent, $P = 0.03$, FET). The male/female ratios among heterosexuals were close to 1:1 in both Denmark and Sweden. In contrast, Finland showed a much more unbalanced male/female ratio with 88 per cent males among the subtype B heterosexuals with domestically acquired HIV-1 infection. No time trends were observed in the proportion of domestically vs. internationally acquired subtype B infection among heterosexuals in any of the Nordic countries ($P > 0.05$ for all comparisons, LBL).

The reverse comparison (i.e. the proportion subtype B among MSM) showed that Denmark and Finland had larger proportions of subtype B among MSM (93 and 95 per cent) compared with Sweden (86 per cent, $P < 0.001$ and $P < 0.001$, FET). Time trend analysis showed a decrease in the proportion of subtype B infections among MSM (from 95 to 87 per cent, 2003–2009, $P = 0.001$, LBL), with the largest decrease in Sweden (from 100 to 75 per cent, 2002–2010, $P < 0.001$, LBL). In agreement, the proportion of domestically acquired HIV-1 non-subtype B infections was larger among MSM in Sweden (13 per cent) compared with Denmark (5 per cent, $P < 0.001$, LBL) and Finland (5 per cent, $P = 0.01$, LBL).

3.3. Sequences clustered according to country and transmission group

Detailed cluster analysis was done for the eight major subtypes/CRFs in the Nordic countries. The Nordic sequences were analysed with 4,537 reference sequences collected during 1983–

Table 2. Number of clusters in different stratifications

Stratification	Dyads[a]	Networks[b]	Large clusters[c]	Total
Sweden (%)	116 (68)	49 (29)	5 (3)	170
Denmark (%)	88 (59)	49 (33)	13 (9)	150
Finland (%)	29 (64)	14 (31)	2 (4)	45
Nordic (%)	24 (39)	29 (48)	8 (13)	61
HET (%)	149 (81)	35 (19)	0 (0)	184
MSM (%)	81 (49)	67 (40)	18 (11)	166
IDU (%)	4 (44)	2 (22)	3 (33)	9
Mixed (%)	20 (31)	37 (58)	7 (11)	64
Unknown (%)	3 (100)	0 (0)	0 (0)	3
Total	257	141	28	426

HET, Heterosexual transmission; MSM, Men who have sex with men; IDU, Intravenous drug use.
[a]Dyads: Clusters of 2 sequences.
[b]Networks: Clusters of 3-14 sequences.
[c]Large clusters: Clusters of >14 sequences.

2012 in 105 countries worldwide. A majority of the Nordic sequences (57 per cent) fell into inferred clusters: 257 dyads, 141 networks (3–14 sequences), and twenty-eight large clusters (fifteen to 121 sequences) (Table 2). The remaining 1,659 sequences likely represented introductions with no or limited spread within the Nordic countries. As expected, the cluster sizes were right-tailed distributed. A majority of the clusters were country-specific (86 per cent), and dyads were the most common cluster form in all three countries (59–68 per cent) and transmission groups (44–81 per cent).

Analyses by transmission route showed that IDU sequences were most likely to form clusters (88 per cent of the sequences clustered), followed by MSM (69 per cent) and heterosexual sequences (41 per cent) ($P < 0.001$, for all pairwise comparisons, FET). Sixty-four (15 per cent) of the clusters were defined as mixed transmission clusters since they were not dominated by a single transmission group. Within those clusters, MSM was generally the most common transmission route (overall proportion 45 per cent), followed by heterosexual (31 per cent) and IDU (17 per cent) transmission. Networks and large clusters were more frequent than dyads among MSM (40 per cent networks and 11 per cent large clusters) and IDU (22 and 33 per cent) clusters as compared with HET clusters (19 and 0 per cent) (Table 2). Furthermore, only 31 per cent of the clustering sequences of HET were found in networks/large clusters, whereas the corresponding number was 86 per cent for the MSM sequences ($P < 0.001$ vs. heterosexual, FET) and 95 per cent for the IDU ($P < 0.001$ vs. heterosexual, and $P < 0.001$ vs. MSM, FET).

HET networks and large clusters were generally smaller and more geographically mixed than MSM clusters that often contained country-specific subclusters (Fig. 2, $P = 0.03$, FET). All large IDU clusters were country-specific. The median number of sequences in HET networks/large clusters was 3 (IQR: 3–5), compared with 5 for MSM (IQR: 4–13.5), 32 for IDU (IQR: 3.5–70), and 4 for mixed transmission clusters (IQR: 3–9). Statistical analyses showed that HET clusters generally contained fewer sequences compared with the other transmission routes ($P < 0.05$ for all comparisons, M-W). Finally, eight and seven of the twenty-eight large clusters (29 and 25 per cent) were defined as Nordic and mixed transmission clusters, respectively, indicating frequent spread of HIV-1 both between Nordic countries and between transmission groups (Table 2).

Next, we analysed the gender distribution among the clustering sequences that were reported as being *heterosexually*
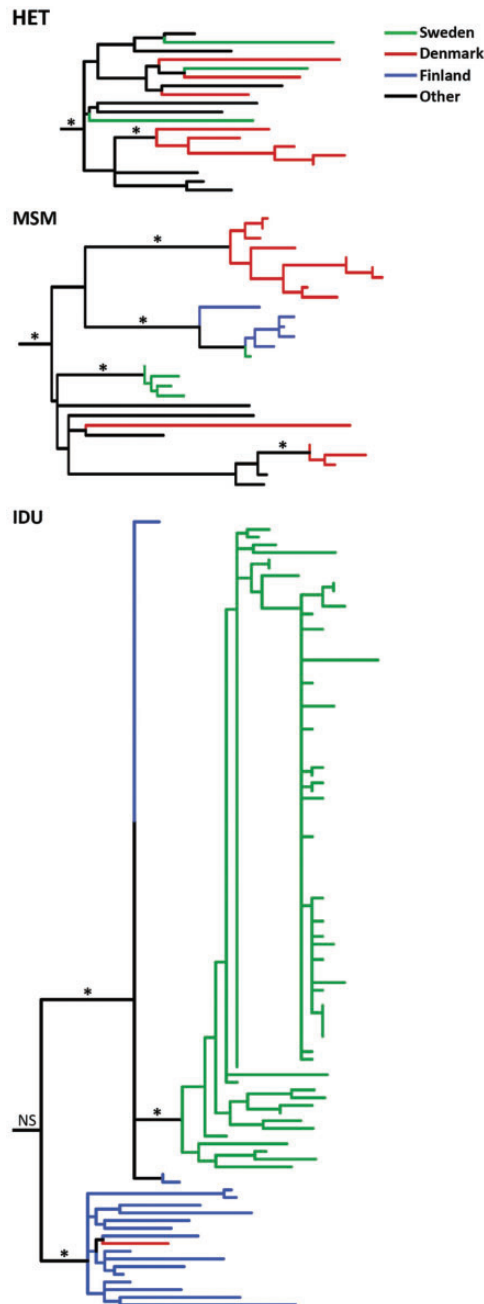
Figure 2. Differences in clustering patterns between transmission groups. Subclusters from maximum likelihood phylogenies showing typical clustering patterns of different transmission groups (HET, heterosexual transmission; MSM, men who have sex with men; IDU, intravenous drug users). HET clusters were typically smaller and more geographically mixed as compared with MSM clusters (where country-specific subclusters often were found) and IDU clusters. The presented IDU cluster contained two subclusters representing a Finnish IDU outbreak between 1998 and 1999 (blue) and a Swedish IDU outbreak between 2005 and 2007 (green). Both outbreaks have been described in detail in the literature and have a well-established link indicating dispersal from the Finnish IDU community to the Swedish IDU community (Skar et al. 2011). The longer branches (higher diversity) seen in the Finnish subcluster likely reflects that these individuals where identified several years after having being infected by HIV-1 (most likely, the majority of those individuals were infected during the outbreak 1998–1999, but the sampling period of the Finnish dataset started first 2003. The colour of the branches represents the country where the sequence was collected (Green, Sweden; Red, Denmark; Blue, Finland; Black, Other countries). Asterisks represent relevant supported branches (aLRT-SH >0.90). NS, non-significant.

transmitted in heterosexual compared with MSM/Mixed transmission clusters (i.e. individuals reported as being, e.g. MSM or IDU were not included in the analysis). The ratio was close to 1:1 in the HET clusters (thirty-six males vs. thirty-one females), and less well balanced in the MSM/Mixed cluster group (~2:1; 108 males vs. sixty-five females, $P = 0.24$, FET). The country-specific analysis indicated similar trends with larger proportions of males reported with HET in MSM/Mixed clusters compared with HET clusters (Sweden: seven males: nine females vs. 13:5, $P = 0.16$; Denmark: 28:21 vs. 59:11, $P = 0.002$; Finland: 1:1 vs. 25:1, $P = 0.14$, FET). Taken together, this suggests a larger extent of misreporting among MSM compared with heterosexuals.

3.4. Fewer active Nordic than country-specific clusters

To further disentangle the HIV-1 transmission patterns in the Nordic countries we performed detailed analyses of larger Nordic, Swedish, Finnish, and Danish clusters (i.e. clusters containing ≥5 sequences, $N = 81$, Fig. 3), and the remaining Result sections present results from these analyses. Some of these clusters, particularly MSM and Mixed clusters, spanned time periods of >15 years. Clusters containing sequences collected ≤2 years before the last enrolment date were defined as active transmission clusters, resulting in 65 per cent (53/81) active clusters. Only 33 per cent of the Nordic clusters (6/18) were active, compared with 85 per cent of the Swedish clusters (17/20, $P = 0.002$, FET), 64 per cent of the Danish clusters (21/33, $P = 0.046$, FET), and 90 per cent of the Finnish clusters (9/10, $P = 0.006$, FET). Moreover, active Nordic clusters were significantly larger (median cluster size 24.5 sequences) than active Swedish (median eleven sequences, $P = 0.045$, M-W) and Finnish clusters (median eleven sequences, $P = 0.045$, M-W), but not significantly larger than active Danish clusters (median sixteen sequences, $P = 0.38$, M-W).

Analyses of active clusters by transmission route indicated a larger proportion of active clusters among MSM (38/50, 76 per cent) compared with mixed clusters (10/21, 48 per cent) ($P = 0.027$, FET). The proportion of active heterosexual and IDU clusters was 43 per cent (3/7) and 67 per cent (2/3), respectively. No other statistically significant differences in proportions of active clusters between transmission groups were found. Forty-five (85 per cent) of the active clusters were subtype B clusters. No significant differences were found in the proportion of active clusters between subtypes and/or CRFs.

3.5. Frequent transmission of HIV-1 from MSM to heterosexuals

To study the directional flow of HIV-1 between countries or transmission groups, we dissected the eighteen Nordic and twenty-one mixed clusters identified in section 3.4 (Supplementary Fig. 2). The analysis suggested a close association between Sweden and Denmark, sharing 61 per cent (11/18) of the Nordic clusters. However, the index sequence or the distribution of sequences over time did not indicate any dominating directional flow between the Nordic countries. The majority of Nordic clusters were MSM clusters (11/18, 61 per cent) and of subtype B (11/18, 61 per cent).

The corresponding analysis of clusters with mixed transmission route showed that most of these clusters were shared by heterosexual and MSM transmission (13/21, 62 per cent). The index sequence was collected from MSM in 85 per cent (11/13) of these clusters. Moreover, 64 per cent (7/11) of these mixed clusters contained sequences collected from heterosexually
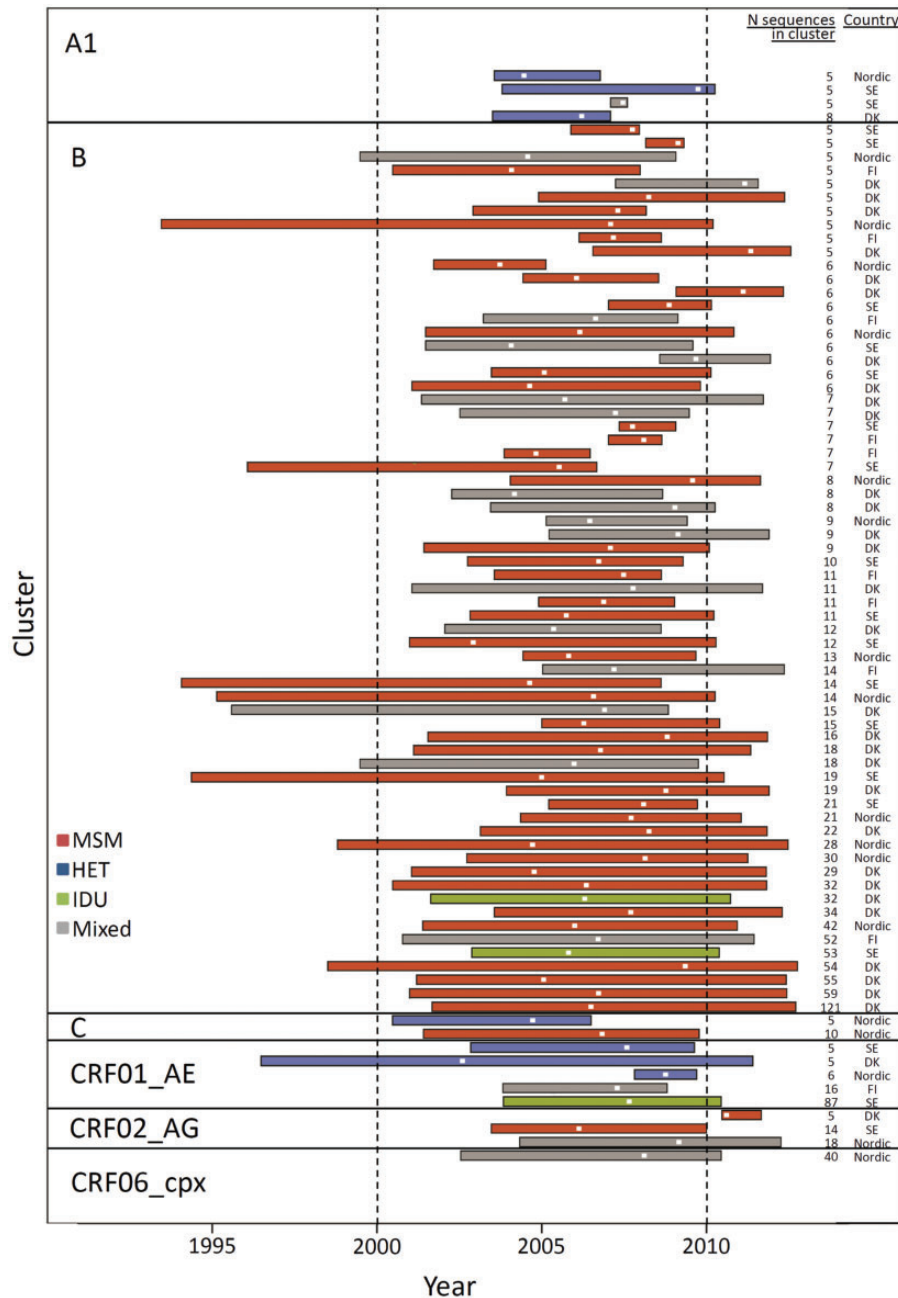
Figure 3. Distribution of sequences over time in the analysed large networks and clusters (N ≥5 sequences). Median (white squares) and range (bars) of the sample collection time periods in years for each cluster. The colour of the bars represents the transmission route: Blue, HET; Red, MSM; Green, IDU; Grey, clusters with sequences of mixed transmission routes. The subtype/CRF is presented on the left, and the number of sequences in each cluster together with each cluster's geographical classification (Nordic; SE, Sweden; DK, Denmark; FI, Finland) are presented on the right side of the figure. Dotted lines have been added to visually clarify the range of the sample collection periods.

infected female patients. In addition, 91 per cent (10/11) of these mixed clusters contained sequences from men reported as heterosexually infected. Notably, all of the three identified heterosexual/IDU transmission clusters had an index sequence of IDU origin. The majority of the clusters with mixed transmission routes were Danish (57 per cent) and of subtype B (81 per cent).

4. Discussion

In this study, we have investigated the molecular epidemiology in three Nordic countries, Denmark, Finland, and Sweden. The study is unique in that it covers three neighbouring countries with a very dense sampling over a long time period. Temporal differences in HIV-1 subtype/CRF dynamics indicated an increasing proportion of CRFs in this geographical region. However, cluster analyses showed that 85 per cent of the active networks and large clusters were subtype B clusters, indicating that subtype B still has the largest influence on the domestic spread of HIV-1 in the Nordic countries.

The HIV-1 epidemic in the Nordic countries, like in many other countries in Western Europe, has been dominated by subtype B and MSM transmission (Public Health Agency of Sweden;

Sonnerborg et al. 1997; Karlsson et al. 2012; Abecasis et al. 2013). Interestingly, we found a larger proportion of MSM among the subtype B infections in both Sweden and Finland compared with Denmark suggesting a larger extent of subtype B spread among heterosexuals and IDUs in Denmark, or alternatively a larger proportion of misreported MSM in Denmark (reported as heterosexuals or IDUs instead of MSM). The larger extent of domestically acquired subtype B among heterosexuals in Denmark supports this. In addition, we found that the proportion of subtype B among MSM decreased over time in the Nordic countries, with the largest decrease seen in Sweden, suggesting less import of non-B subtypes to the MSM communities in Denmark and Finland. Altogether, these results suggests that Finland may have a more closed MSM and subtype B community as compared with Sweden and Denmark. This was further supported by the finding that most Nordic networks and large clusters were MSM clusters shared between Sweden and Denmark.

Recently, a decreasing trend of subtype B infections was shown for Sweden (Neogi et al. 2014). We show that this was true also for Finland and all Nordic countries when analysed together, where the most recent estimates suggest that subtype B now constitutes less than half of all newly diagnosed infections. This also means that the proportion of non-subtype B infections increased during the study period. To determine if differences in migration patterns between Nordic countries could be linked to this trend, we used data from the United Nations Population Division to plot time trends of the net number of migrants per capita (Supplementary Fig. 3) (see United Nations Department of Economic and Social Affairs, Population Division). Sweden, which had the largest increase of non-B subtypes (from 42 to 66 per cent), had the largest number of net number of migrants per capita (0.57 per cent) compared with Denmark (0.32 per cent) and Finland (0.27 per cent) in 2007. Although a stratified analysis indicated that the relatively high overall non-subtype B prevalence (58 per cent) could be related to sub-Saharan African country of origin in Sweden (26 per cent of the Swedish sequences were obtained from individuals with a sub-Saharan African origin), this was not as clear for Denmark or Finland. Moreover, the trend analysis by sub-Saharan Africa vs. non-sub-Saharan Africa did not support the hypothesis that the increase in non-subtype B could be related to changes in proportions of newly discovered infections among individuals with a sub-Saharan origin. Interestingly, all HIV-1 variants with decreasing trends were subtypes whereas all variants with increasing trends were different recombinant forms. A similar trend was recently reported in Sweden when analysing all recombinants as one group (Neogi et al. 2014). Previous reports have suggested that different recombinant forms of HIV-1 may be more pathogenic and with higher replicative capacity compared with the parental strains indicating the importance of further studies of these variants and also their possible epidemiological impact (Konings et al. 2006; Njai et al. 2006; Tebit et al. 2007; Esbjornsson et al. 2010; Palm et al. 2013).

Overall, the Nordic sequences showed a high degree of clustering, similar to what have been seen in other large surveillance cohorts (Kouyos et al. 2010; Leigh Brown et al. 2011). Finland contributed with fewer sequences compared with Sweden and Denmark. However, and more importantly, the coverage of analysed sequences from the different countries were similar (ranging from 44 to 57 per cent), and should therefore reflect the relative impact of the different country-specific HIV-1 epidemics on the overall Nordic HIV-1 epidemic. We also found that clustering patterns differed between transmission groups indicating closer transmission behaviour among IDUs compared with MSM and heterosexuals. This is in line with the report by Kouyos et al., where the majority of the Swiss IDU sequences were found in domestic transmission clusters. The closer transmission pattern among IDUs was further supported by the larger size of IDU clusters compared with MSM and HET clusters. It is possible that the small median size of HET clusters could be linked to generally lower risk behaviour and a larger extent of imported transmission in this group compared with other transmission groups (i.e. higher probability that a large part of the transmission chain was not sampled). Although occasional intentional or unintentional misreporting of transmission route cannot be excluded, only 17 per cent of the sequences in the Nordic clusters with mixed transmission route were IDU sequences, suggesting a larger degree of heterosexual ↔ MSM transmission compared with heterosexual ↔ IDU or MSM ↔ IDU transmission. Moreover, dissection of large networks and clusters ($N \geq 5$ sequences) showed that the majority of clusters with mixed transmission route were heterosexual/MSM transmission clusters, similar to what has been reported from HIV-1 subtype B transmission in Switzerland (Kouyos et al. 2010). Interestingly, 85 per cent of these clusters had an index sequence originating from MSM transmission, suggesting transmission of HIV-1 from MSM to heterosexuals. One limitation of this analysis is that late HIV-1 diagnoses are relatively common and late presenters could still be the source of a transmission chain (Mocroft et al. 2013; Brannstrom et al. 2015). Moreover, it is possible that a proportional overrepresentation of early diagnoses relative to late diagnoses in a phylogenetic cluster could skew the interpretation of directionality, particularly in more recently introduced transmission chains (i.e. higher likelihood of undiagnosed late presenters in recently introduced transmission chains). Across Europe, late HIV diagnosis has been suggested to be most common among heterosexually infected African immigrants and to a lesser extent among MSM (Mocroft et al. 2013). Although inferring transmission directionality from a phylogenetic tree is difficult, extensive analysis using a coalescence-based discrete trait approach could potentially shed some additional light to this issue (Faria et al. 2011).

The presence of sequences from females in the clusters with mixed heterosexual and MSM transmission routes provides solid proof for HET within these clusters, but misreporting cannot be ruled out for men reported as heterosexually infected. Instead the over-representation of men among the heterosexuals in these clusters suggests some misreporting of MSM as being heterosexually infected. Similar evidence for misclassifications of transmission route have been reported from the UK (Hue et al. 2014). In line with our findings of transmission from MSM to heterosexuals, Bruhn et al. (2014) recently showed that HIV-1 was first introduced in Greenland by MSM transmission, which then through bisexual transmission reached and spread within the heterosexual community.

The transmission routes of the study population were similar to those of all diagnosed patients during the study period with a small over-representation of MSM (Supplementary Data). Moreover, the number of undiagnosed cases in the Nordic countries has been estimated to be relatively low (range 5–15 per cent, personal communication and Guerra and Ekstrom 2012). Based on the large number of analysed sequences, the low number of undiagnosed cases, and no systematic loss among transmission groups, we are confident that our results represent the overall trends in the Nordic HIV-1 epidemic during the study period (Novitsky et al. 2014).

Cluster definition is an important but difficult task. Whereas some studies have used pure distance-based criteria, either calculated directly from the sequences or from phylogenetic trees, others have used pure phylogenetic branch support or combinations between distance similarities and branch support (Hue et al. 2004; Hughes et al. 2009; Bezemer et al. 2010b; Esbjornsson et al. 2011; Leigh Brown et al. 2011; Stadler et al. 2012). The selected approach for identifying clusters should depend on the purpose of the study. In this study, we aimed to identify and dissect Nordic HIV-1 transmission clusters. Considering that our dataset had very high coverage, spanned many years and included sequences from more than one country, we wanted to avoid a cluster definition that only detects recently established clusters with short branch lengths (see Fig. 2 and Discussion in Supplementary Data for further discussion on this). Thus, we determined transmission clusters based on phylogenetic branch support and proportion of Nordic sequences. Importantly, the addition of the 80 per cent proportion of Nordic sequences in our cluster definition enabled us to identify transmission links between Nordic countries. Since various types of distance-based thresholds have been used in the literature (Kaye, Chibo, and Birch 2008; Hughes et al. 2009; Bezemer et al. 2010b), we performed a comparative analysis between our cluster definition and a corresponding definition with the addition of a distance threshold (maximum pair-wise distance of 0.045 substitutions/site). This analysis showed that the number of identified clusters was similar between the definitions, but that some of the large clusters were reduced to smaller subclusters. Further dissection of these large clusters indicated that they most likely represented long-lived Nordic transmission clusters in which the genetic distances between the sampled virus strains had diverged apart more than 4.5 per cent (0.045 substitutions/site). These results together with a detailed discussion are available in Supplementary Data.

In this study, we analysed large multinational datasets covering more than 50 per cent of all newly discovered HIV-1 infections for more than a decade in a large geographic area of countries with close historical and cultural inheritance. The majority of previous reports with similar coverage have focused on subtype B-infected individuals and been based on national or local cohorts (Kouyos et al. 2010; Leigh Brown et al. 2011; Aldous et al. 2012). Although further studies are needed to better understand the impact of subtype/CRF on transmission dynamics and spread of HIV-1 in different geographical regions and transmission groups, our observations may have implications for public health interventions in targeting HIV-1 transmission networks and identifying where such interventions should be introduced.

Funding

Conflicts of interest: None declared.

Meeting presentations

Results of this study have been presented in part at The Epidemiology of HIV-1, HCV, and HBV in the Baltic Region Meeting, Stockholm, Sweden (2015); The 11th Annual Conference of the Baltic Network Against Life-threatening Viral Infections. Vilnius, Lithuania (2014); The 20th International HIV Dynamics and Evolution, Utrecht, The Netherlands (2013); The 11th European Meeting on HIV & Hepatitis—Treatment Strategies & Antiviral Drug Resistance, Rome, Italy (2013); The European Society for Antiviral Resistance Meeting, Rome, Italy (2013).

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Acknowledgements

References

Abecasis, A. B., et al. (2013) 'HIV-1 Subtype Distribution and Its Demographic Determinants in Newly Diagnosed Patients in Europe Suggest Highly Compartmentalized Epidemics', *Retrovirology*, 10: 7.

Aldous, J. L., et al. (2012) 'Characterizing HIV Transmission Networks Across the United States', *Clinical Infectious Diseases*, 55: 1135–43.

Alizon, S., et al. (2010) 'Phylogenetic Approach Reveals That Virus Genotype Largely Determines HIV Set-Point Viral Load', *PLoS Pathogens*, 6: e1001123.

Altschul, S. F., et al. (1997) 'Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs', *Nucleic Acids Research*, 25: 3389–402.

Anisimova, M., et al. (2011) 'Survey of Branch Support Methods Demonstrates Accuracy, Power, and Robustness of Fast Likelihood-Based Approximation Schemes', *Systematic Biology*, 60: 685–99.

Arien, K. K., et al. (2005) 'The Replicative Fitness of Primary Human Immunodeficiency Virus Type 1 (HIV-1) Group M, HIV-1 Group O, and HIV-2 Isolates', *Journal of Virology*, 79: 8979–90.

Bezemer, D., et al. (2010a) '27 Years of the HIV Epidemic Amongst Men Having Sex With Men in the Netherlands: An In Depth Mathematical Model-Based Analysis', *Epidemics*, 2: 66–79.

——, et al. (2010b) 'Transmission Networks of HIV-1 Among Men Having Sex With Men in The Netherlands', *AIDS*, 24: 271–82.

Brannstrom, J., et al. (2015) 'Deficiencies in the Health Care System Contribute to a High Rate of Late HIV Diagnosis in Sweden', *HIV Medicine* hiv.12321.

Brenner, B., Wainberg, M. A., and Roger, M. (2013) 'Phylogenetic Inferences on HIV-1 Transmission: Implications for the Design of Prevention and Treatment Interventions', *AIDS*, 27: 1045–57.

Bruhn, C. A., et al. (2014) 'The Origin and Emergence of an HIV-1 Epidemic: From Introduction to Endemicity', *AIDS*, 28: 1031–40.

EpiNorth. Available at: http://www.epinorth.org/.

Esbjornsson, J., et al. (2010) 'Frequent CXCR4 Tropism of HIV-1 Subtype A and CRF02_AG During Late-Stage Disease—Indication of an Evolving Epidemic in West Africa', *Retrovirology*, 7: 23.

——, et al. (2011) 'HIV-1 Molecular Epidemiology in Guinea-Bissau, West Africa: Origin, Demography and Migrations', *PLoS One*, 6: e17025.

Faria, N. R., et al. (2011) 'Toward a Quantitative Understanding of Viral Phylogeography', *Current Opinion in Virology*, 1: 423–9.

Fisher, M., et al. (2010) 'Determinants of HIV-1 Transmission in Men Who Have Sex With Men: A Combined Clinical, Epidemiological and Phylogenetic Approach', *AIDS*, 24: 1739–47.

Frost, S. D., and Pillay, D. (2015) 'Understanding Drivers of Phylogenetic Clustering in Molecular Epidemiological Studies of HIV', *Journal of Infectious Diseases*, 211: 856–8.

Grabowski, M. K., et al. (2014) 'The Role of Viral Introductions in Sustaining Community-Based HIV Epidemics in Rural Uganda: Evidence From Spatial Clustering, Phylogenetics, and Egocentric Transmission Models', *PLoS Medicine*, 11: e1001610.

Guerra, M. V., and Ekstrom, A. M. (2012) 'Adolescents Denied HIV Testing at Adolescent Health Centers. Lack of Knowledge and Interest According to a Study With Simulated Client Study' *Lakartidningen*, 109: 625–8.

Hue, S., et al. (2014) 'Phylogenetic Analyses Reveal HIV-1 Infections Between Men Misclassified as Heterosexual Transmissions', *AIDS*, 28: 1967–75.

——, et al. (2004) 'HIV-1 Pol Gene Variation Is Sufficient for Reconstruction of Transmissions in the Era of Antiretroviral Therapy', *AIDS*, 18: 719–28.

Hughes, G. J., et al. (2009) 'Molecular Phylodynamics of the Heterosexual HIV Epidemic in the United Kingdom', *PLoS Pathogens*, 5: e1000590.

Karlsson, A., et al. (2012) 'Low Prevalence of Transmitted Drug Resistance in Patients Newly Diagnosed With HIV-1 Infection in Sweden 2003–2010', *PLoS One*, 7: e33484.

Kaye, M., Chibo, D., and Birch, C. (2008) 'Phylogenetic Investigation of Transmission Pathways of Drug-Resistant HIV-1 Utilizing Pol Sequences Derived From Resistance Genotyping', *Journal of Acquired Immune Deficiency Syndromes*, 49: 9–16.

Kiwanuka, N., et al. (2010) 'HIV-1 Viral Subtype Differences in the Rate of CD4+ T-Cell Decline Among HIV Seroincident Antiretroviral Naive Persons in Rakai District, Uganda', *Journal of Acquired Immune Deficency Syndromes*, 54: 180–4.

Konings, F. A., et al. (2006) 'Human Immunodeficiency Virus Type 1 (HIV-1) Circulating Recombinant Form 02_AG (CRF02_AG) Has a Higher In Vitro Replicative Capacity Than Its Parental Subtypes A and G', *Journal of Medical Virology*, 78: 523–34.

Kouyos, R. D., et al. (2010) 'Molecular Epidemiology Reveals Long-Term Changes in HIV Type 1 Subtype B Transmission in Switzerland', *Journal of Infectious Diseases*, 201: 1488–97.

Leigh Brown, A. J., et al. (2011) 'Transmission Network Parameters Estimated From HIV Sequences for A Nationwide Epidemic', *Journal of Infectious Diseases*, 204: 1463–9.

Liitsola, K., et al. (2000) 'Analysis of HIV-1 Genetic Subtypes in Finland Reveals Good Correlation Between Molecular and Epidemiological Data', *Scandinavian Journal of Infectious Diseases*, 32: 475–80.

Lindstrom, A., et al. (2006) HIV-1 Transmission Cluster With M41L 'Singleton' Mutation and Decreased Transmission of Resistance in Newly Diagnosed Swedish Homosexual Men', *Antiviral Therapy*, 11: 1031–9.

Los Alamos Sequence Database. <www.hiv.lanl.gov> accessed 5 Sep 2015.

Lukashov, V. V., et al. (1996) 'Evidence for HIV Type 1 Strains of U.S. Intravenous Drug Users as Founders of AIDS Epidemic

Among Intravenous Drug Users in Northern Europe', *AIDS Research and Human Retroviruses*, 12: 1179–83.

Mocroft, A., et al. (2013) 'Risk Factors and Outcomes for Late Presentation for HIV-Positive Persons in Europe: Results From the Collaboration of Observational HIV Epidemiological Research Europe Study (COHERE)', *PLoS Medicine*,10: e1001510.

Morrison, C. S., et al. (2010) 'Plasma and Cervical Viral Loads Among Ugandan and Zimbabwean Women During Acute and Early HIV-1 Infection', *AIDS*, 24: 573–82.

Neogi, U., et al. (2014) 'Temporal Trends in the Swedish HIV-1 Epidemic: Increase in Non-B Subtypes and Recombinant Forms Over Three Decades', *PLoS One*, 9: e99390.

Njai, H. F., et al. (2006) 'The Predominance of Human Immunodeficiency Virus Type 1 (HIV-1) Circulating Recombinant Form 02 (CRF02_AG) in West Central Africa May Be Related to Its Replicative Fitness', *Retrovirology* 3: 40.

Novitsky, V., et al. (2014) 'Impact of Sampling Density on the Extent of HIV Clustering', *AIDS Research and Human Retroviruses*, 30: 1226–35.

Palm, A. A., et al. (2013) 'Faster Progression to AIDS and AIDS-Related Death Among Seroincident Individuals Infected With Recombinant HIV-1 A3/CRF02_AG Compared With Sub-Subtype A3', *Journal of Infectious Diseases*, 209: 721–8.

Poon, A. F., et al. (2015) 'The Impact of Clinical, Demographic and Risk Factors on Rates of HIV Transmission: A Population-Based Phylogenetic Analysis in British Columbia, Canada', *Journal of Infectious Diseases*, 211: 926–35.

Renjifo, B., et al. (2004) 'Preferential In-Utero Transmission of HIV-1 Subtype C As Compared to HIV-1 Subtype A or D', *Aids*, 18: 1629–36.

Skar, H., et al. (2011) 'Dynamics of Two Separate But Linked HIV-1 CRF01_AE Outbreaks Among Injection Drug Users in Stockholm, Sweden, and Helsinki, Finland', *Journal of Virology*, 85: 510–8.

Sonnerborg, A., et al. (1997) 'Dynamics of the HIV-1 Subtype Distribution in the Swedish HIV-1 Epidemic During the Period 1980 to 1993', *AIDS Research and Human Retroviruses*, 13: 343–5.

SPREAD programme (2008) 'Transmission of Drug-Resistant HIV-1 in Europe Remains Limited to Single Classes', *AIDS*, 22: 625–35.

Stadler, T., et al. (2012) 'Estimating the Basic Reproductive Number From Viral Sequence Data', *Molecular Biology and Evolution*, 29: 347–57.

Public Health Agency of Sweden. Hivinfektion. <http://www.folk halsomyndigheten.se/amnesomraden/statistik-och-underso kningar/sjukdomsstatistik/hivinfektion/?t=com&p=6099#stat istics-nav> accessed 9 Sep 2015.

Tebit, D. M., et al. (2007) 'HIV Diversity, Recombination and Disease Progression: How Does Fitness "Fit" Into the Puzzle?' *AIDS Reviews*, 9: 75–87.

UNAIDS. Epidemiological Fact Sheet on HIV and AIDS. <http:// www.unaids.org/en/default.asp>.

United Nations Department of Economic and Social Affairs, Population Division (2013) 'World population prospects: the 2012 revision, DVD Edition', <http://esa.un.org/unpd/wpp/ Excel-Data/migration.htm> UNAIDS accessed 3 Nov 2015.

Vercauteren, J., et al. (2009) 'Transmission of Drug-Resistant HIV-1 Is Stabilizing in Europe', *Journal of Infectious Diseases*, 200: 1503–8.

Wertheim, J. O., et al. (2014) 'The Global Transmission Network of HIV-1', *Journal of Infectious Diseases*, 209: 304–13.