Heliyon

Genetic characterization of human immunodeficiency virus type 1 transmission in the Middle East and North Africa

Malik Sallam [a,*], Gülşen Özkaya Şahin [a,b], Mikael Ingman [a], Anders Widell [a],
Joakim Esbjörnsson [c], Patrik Medstrand [a]

[a] Lund University, Faculty of Medicine, Department of Translational Medicine, Malmö, Sweden

[b] Laboratory Medicine Skåne, Lund, Sweden

[c] Lund University, Faculty of Medicine, Department of Laboratory Medicine, Lund, Sweden

* Corresponding author. Faculty of Medicine, Lund University, Department of Translational Medicine, BMC B13,
Sölvegatan 19, 223 62 Lund, Sweden.

E-mail address: malik.sallam@med.lu.se (M. Sallam).

Abstract

Background: The HIV-1 spread in the Middle East and North Africa (MENA) has
not been previously characterised using the phylogenetic approach. The aim of the
current study was to investigate the genetic diversity and domestic transmission of
HIV-1 in the MENA.

Methods: A total of 2036 HIV-1 sequences available in Genbank and collected in
the MENA during 1988–2016 were used together with 715 HIV-1 reference
sequences that were retrieved from Genbank based on genetic similarity with the
MENA sequences. The REGA and COMET tools were used to determine HIV-1
subtypes and circulating recombinant forms. Maximum Likelihood and Bayesian
phylogenetic analyses were used to identify and date HIV-1 transmission clusters.

Results: At least 21 HIV-1 subtypes and recombinant forms were prevalent in the
MENA. Subtype B was the most common variant (39%), followed by CRF35_AD
(19%) and CRF02_AG (14%). The most common genetic region was *pol*, and 675
partial *pol* sequences (average of 1005 bp) were eligible for detailed phylogenetic
analysis. Fifty-four percent of the MENA sequences formed HIV-1 transmission
clusters. Whereas numerous clusters were country-specific, some clusters indicated

transmission links between countries for subtypes B, C and CRF02_AG. This was more common in North Africa compared with the Middle East (p < 0.001). Recombinant forms had a larger proportion of clustering compared to pure subtypes (p < 0.001). The largest MENA clusters dated back to 1991 (an Algerian CRF06_cpx cluster of 43 sequences) and 2002 (a Tunisian CRF02_AG cluster of 48 sequences).

Conclusions: We found an extensive HIV-1 diversity in the MENA and a high proportion of sequences in transmission clusters. This study highlights the need for preventive measures in the MENA to limit HIV-1 spread in this region.

Keywords: Health profession, Medicine, Infectious disease, Public health, Evolution, Genetics, Virology

1. Introduction

An important aspect of studying the epidemiology of infectious diseases is the characterisation of spread of infectious agents. This can be approached through phylogenetic analysis that allows inference of epidemiologic links and depiction of different variables associated with pathogen transmission (Grenfell et al., 2004). Moreover, phylogenetic analysis can be applied to gain in-depth knowledge of the virus spread and to guide efficient preventive strategies (Brenner et al., 2013).

HIV-1 infection has been reported in the Middle East and North Africa (MENA) from mid-1980s (UNAIDS, 2016). Based on the UNAIDS regional classification, the MENA region includes the following countries: Algeria, Bahrain, Djibouti, Egypt, Iran, Iraq, Jordan, Kuwait, Lebanon, Libya, Morocco, Occupied Palestinian Territories, Oman, Qatar, Saudi Arabia, Somalia, Sudan, Syria, Tunisia, United Arab Emirates, and Yemen (UNAIDS, 2016). Despite the low prevalence of HIV-1 in most MENA countries (less than 0.1%), the trends of new HIV-1 infections and AIDS-related mortality have increased since 2001 (Gokengin et al., 2016; UNAIDS, 2011, 2016). In addition, the MENA has the lowest coverage of antiretroviral therapy (ART) in the world (UNAIDS, 2011, 2016).

The epidemiologic features of HIV/AIDS epidemic in the MENA have been reviewed by (Mumtaz et al., 2014). Briefly, HIV-1 infection in the MENA can be described as a diverse and heterogeneous epidemic with the majority of countries afflicted by concentrated epidemics among high risk marginalized populations as injection drug users (IDUs), female sex workers (FSW) and men who have sex with men (MSM). Previous reports have indicated variability in the risk groups affected per country, e.g. IDUs in Iran and Libya, MSM in Lebanon and FSW in Djibouti (Khajehkazemi et al., 2013; Mahfoud et al., 2010; Mirzoyan et al., 2013; Mumtaz et al., 2014). Moreover, within-country variation was noticed in relation to affected risk groups in Morocco, with IDUs concentrated in the North in comparison to MSM in the South (Kouyoumjian et al., 2013). The HIV-1 epidemic

in the MENA has been reported to be genetically diverse on the country-specific level with subtype B being reported as one of the most common variants (Bakhouch et al., 2009; Ben Halima et al., 2001; Bouzeghoub et al., 2006; Mumtaz et al., 2011; Saad et al., 2005). Notably in Iran, the epidemic was dominated by CRF35_AD, particularly among IDUs (Jahanbakhsh et al., 2013).

The local spread of HIV-1 in the MENA region viewed as a single geographic unit has not been elucidated yet. The objectives of the current study were to determine the genetic diversity and spread of HIV-1 within and between countries in the MENA region.

2. Materials and methods

2.1. HIV-1 MENA dataset

All HIV-1 sequences from the MENA countries, were downloaded from the Los Alamos HIV Database (LADB, http://www.hiv.lanl.gov/) with retrieval of the background sequence information, as of 31st December 2016, using the Geography search tool. Based on information about patient code and patient ID, we identified unique sequences (the earliest single sequence from each individual) which were used to investigate the distribution of subtypes and circulating recombinant forms (CRFs) in the region. Multi-locus sequences collected in the same study and from the same patient were concatenated to improve the phylogenetic signal in the subtype/CRF classification and transmission cluster analysis. For multiple clonal sequences from the same patient, a single sequence was randomly selected. We used the sequence locator tool in LADB to identify the HIV-1 sub-genomic region with the largest number of sequences for subsequent analyses (Supplementary File 1). The complete list of the publications with submitted sequences to LADB that were used in our study is provided in (Supplementary File 1).

2.2. Subtype/CRF assignment and selection of *pol* dataset

As it has been shown that the original subtype/CRF assignment in LADB is not necessarily accurate, particularly for recombinant forms (Zhang et al., 2010), the classification of the MENA HIV-1 sequences was checked using REGA v3 and COMET HIV-1 online subtyping tools (Pineda-Pena et al., 2013; Struck et al., 2014). For discordant results, COMET assignments were used to make the final subtype/CRF call as COMET has been shown to be superior to REGAv3 for short sequences (Pineda-Pena et al., 2013). For the MENA sequences that were used to investigate domestic transmission, subtype/CRF assignment was confirmed by maximum likelihood (ML) phylogenetic analysis with LADB HIV-1 GENOME Subtype Reference Alignment from 2010 (the most recent reference alignment). A total of 712 partial *pol* sequences were available for analysis, (average length of 1005 base pairs; nucleotide positions 2259–3264 of HXB2, GenBank accession

number K03455). A slight variation in sequence alignment length was noticed depending on the subtype/CRF (Supplementary File 1).

As the ART-status of the patients at time of sequence collection was unknown, we eliminated resistance codon positions in the protease (PR) region (30, 32, 46, 47, 48, 50, 54, 82, 84, 88 and 90) and in the reverse transcriptase (RT) region (41, 62, 65, 67, 70, 74, 75, 77, 100, 101, 115, 116, 151, 181, 184, 188, 190, 215 and 219) to reduce bias of the ART selective pressure (Wensing et al., 2015). We deleted the initial RT segment from the final sequence alignment for the several subtype/CRF datasets, since a majority of *pol* sequences lacked this genome region, especially for the sequences collected in Algeria, Kuwait and Morocco (Supplementary File 1). Sequence editing and alignment was performed using MEGA6 and ClustalW respectively (Larkin et al., 2007; Tamura et al., 2013).

2.3. Maximum likelihood phylogenetic analyses

For ML analyses, the PR and RT regions of *pol* gene were found to be have the largest number of sequences in LADB (Supplementary File 1), nevertheless, a considerable fraction of these sequences were either only PR or RT short sequences, presumably lacking a sufficient phylogenetic signal for informative HIV-1 cluster analysis (Novitsky et al., 2015). Based on subtype/CRF assignment, eight datasets were created for the most common genetic variants found (A1, B, C, D, CRF01_AE, CRF02_AG, CRF06_cpx and CRF35_AD). For each MENA dataset, a search for similar sequences was performed using NCBI GenBank BLAST tool (Benson et al., 2006), and the ten best sequence hits for every query sequence from the MENA were retained for analysis after removing identical target sequences. To reduce the likelihood of including multiple intra-patient sequences, the reference sequences were further reduced using Skipredundant from the EMBOSS package with a similarity cut-off of 99.0% (Rice et al., 2000). To investigate domestic transmission and lineage mixing of the prevalent subtypes/ CRFs in the MENA region, ML phylogenetic trees were constructed for the final datasets that comprised the MENA sequences together with GenBank reference sequences (Supplementary Files 1 and 2). ML analysis was conducted in GARLI v2.0 with five search replicates using the GTR + I + Γ nucleotide substitution model (Bazinet et al., 2014), and the ML tree with highest likelihood score was retained for subsequent analyses. The measure of statistical support for nodes in the ML tree was determined using the approximate Likelihood Ratio Test Shimodaira-Hasegawa like (aLRT-SH) implemented in PhyML v3.1 (Guindon et al., 2010) and an aLRT-SH value ≥ 0.90 was considered significant (Anisimova et al., 2011). The MENA transmission clusters were identified by inspection of the final ML trees from root to tips and defined as the monophyletic clades with aLRT-SH like support values ≥ 0.90 and that contained at least 80% sequences from the MENA region (Esbjornsson et al., 2016; Kouyos et al., 2010). The phylogenetically

identified transmission clusters were classified into dyads (two sequences), networks (3–14 sequences) and large clusters (15 or more sequences) depending on the number of sequences within each cluster (Aldous et al., 2012; Esbjornsson et al., 2016).

2.4. Bayesian evolutionary analyses

The time to most recent common ancestors (tMRCAs) of the MENA clusters were determined in BEAST v1.8.4 (Drummond et al., 2012), using the following settings: HKY nucleotide substitution model with discrete gamma-distributed rate heterogeneity, uncorrelated relaxed clock model with an uninformative uniform rate prior (lower = 0, upper = 100, initial value of 0.001), and a Bayesian skyline tree density model. The results of five independent runs, chain length 100 million, were combined using LogCombiner v1.8.2 after discarding 20% burn-in and were checked for convergence using Tracer v1.6.0. (http://tree.bio.ed.ac.uk/software/tracer/).

2.5. Sequence accession numbers

A complete list of the nucleotide accession numbers for the MENA sequences analysed in our study is provided in (Supplementary File 2).

2.6. Statistical analysis

All statistical analyses were conducted through IBM SPSS Statistics 21.0 for Windows. P values were calculated using the exact two-sided Fisher's test (FET) and Mann–Whitney U test (M-W) with $p < 0.050$ as the significance level.

3. Results

3.1. Complex genetic diversity of HIV-1 in the MENA

A total of 3319 HIV-1 sequences was retrieved from LADB. Of these sequences, we identified a total of 2036 unique sequences collected between 1988 and 2016 from 13 different MENA countries (Algeria, Djibouti, Egypt, Iran, Kuwait, Lebanon, Libya, Morocco, Saudi Arabia, Somalia, Sudan, Tunisia and Yemen). The concordance rate between the original LADB and our subtype/CRF assignment was 90% (n = 1836 out of 2036). A high proportion of discordant results was noticed for recombinant forms (n = 163, 82%). Based on the short length of a considerable number of sequences, 25% of the MENA sequences (n = 513) could not be assigned to a specific subtype/CRF using the REGA subtyping tool, compared with 4% (n = 74) for COMET. The majority of these sequences were relatively short (median 671 bp, IQR: 299–776) compared with the overall median sequence length of 864 bp (IQR: 638–1040). Hence we decided to use

COMET results for final subtype/CRF assignment besides the previous conclusion that COMET is considered one of the best methods for HIV-1 subtyping (Pineda-Pena et al., 2013). Out of 712 partial *pol* sequences that were subtyped using ML phylogenetic analysis with reference sequences, 708 (99%) had an identical COMET assignment (Supplementary File 1), and were used to conduct transmission cluster analysis.

Of the 2036 unique MENA HIV-1 sequences, 1960 (96%) sequences were assigned to 21 specific subtypes/CRFs. Of these, 98% (n = 1927) belonged to nine subtypes/CRFs that predominated the MENA HIV-1 epidemic (Table 1), with subtype B being the most prevalent variant (n = 804, 39%). Among the recombinant forms, CRF35_AD was the most common (n = 396, 19%) followed by CRF02_AG (n = 293, 14%). Pure subtype assignment was found in a slightly higher number than CRF assignment (1091 [56%] vs. 869 [44%]). Six percent (n = 124) of the sequences represented complex recombinant forms. Subtype B was the most common in six countries while subtype C was the most common in three countries. Only two countries (Libya and Somalia) lacked subtype B sequences, whereas CRF35_AD was found only in Iran (Fig. 1, Table 1).

3.2. High proportion of domestic HIV-1 spread in the MENA

To investigate the proportion of HIV-1 introductions that have led to domestic spread, 708 partial *pol* sequences collected in the MENA were stratified by subtype/CRF (A1, B, C, D CRF01_AE, CRF02_AG, CRF06_cpx and CRF35_AD) and analysed together with 715 GenBank reference sequences selected by similarity (Supplementary File 2). The MENA sequences had been collected in Algeria, Djibouti, Iran, Kuwait, Libya, Morocco, Somalia, Sudan, Tunisia, and Yemen during 1989–2016. The Libyan sequences (n = 33) belonged to a single monophyletic clade that caused a nosocomial outbreak and were thus excluded from transmission cluster analysis as the inclusion of such sequences might result in a spuriously high proportion of clustering (Visco-Comandini et al., 2002). Overall, the proportion of phylogenetic clustering was 54% (363 of the 675 MENA sequences were found in clusters). Tunisia had the highest proportion of clustering (n = 107, 61%), whereas the lowest proportion was found in Sudan (n = 6, 21%). The proportion of clustering per subtype/CRF was highest for CRF06_cpx (n = 52, 98%), whereas the lowest proportion was observed for subtype C (n = 16, 38%, Fig. 2). Risk group information was available for a small fraction of sequences (n = 65, 10%) which precluded its inclusion in subsequent analyses (CRF35_AD = 44, subtype B = 13, subtype C = 6, subtype A1 = 1 and subtype D = 1).

Table 1. HIV-1 subtype/CRF distribution per country in the Middle East and North Africa (MENA).

Country Clade	Algeria N[1] (%)	Djibouti N (%)	Egypt N (%)	Iran N (%)	Kuwait N (%)	Lebanon N (%)	Libya N (%)	Morocco N (%)	Saudi Arabia N (%)	Somalia N (%)	Sudan N (%)	Tunisia N (%)	Yemen N (%)	Total N (%)
B	132 (39)	1 (2)	21 (95)	48 (9)	15 (16)	11 (38)	0	443 (76)	10 (18)	0	1 (3)	113 (52)	9 (47)	804 (39)
CRF35_AD[2]	0	0	0	396 (77)	0	0	0	0	0	0	0	0	0	396 (19)
CRF02_AG	54 (16)	8 (16)	0	2 (0.4)	6	6 (21)	75 (100)	60 (10)	2 (4)	0	0	80 (37)	0	293 (14)
C	1 (0.3)	34 (68)	0	5 (1)	24 (26)	1 (3)	0	11 (2)	22 (39)	4 (80)	11 (37)	3 (1)	6 (32)	122 (6)
CRF06_cpx	99 (29)	0	0	0	1 (1)	0	0	0	0	0	0	2 (0.9)	0	102 (5)
A1	5 (1)	2 (4)	0	57 (11)	5 (5)	5 (17)	0	17 (3)	1 (2)	0	2 (7)	3 (1)	1 (5)	98 (5)
CRF01_AE	1 (0.3)	2 (4)	1 (5)	8 (2)	33 (35)	0	0	5 (0.9)	0	0	0	0	0	50 (2)
G	6 (2)	0	0	0	0	2 (7)	0	21 (4)	6 (11)	0	1 (3)	3 (1)	0	39 (2)
D	2 (0.6)	0	0	0	0	2 (7)	0	0	1 (2)	0	15 (50)	2 (0.9)	1 (5)	23 (1)
Others[3]	40 (12)	3 (6)	0	0	9 (10)	2 (7)	0	27 (5)	14 (25)	1 (20)	0	11 (5)	2 (11)	109 (5)
Total	340	50	22	516	93	29	75	584	56	5	30	217	19	2036

[1] N: Number.

[2] CRF: Circulating recombinant form.

[3] Others: Include the following: sub-subtypes F1 and F2, CRF09_cpx, CRF11_cpx, CRF13_cpx, CRF18_cpx, CRF19_cpx, CRF25_cpx, CRF43_02G, CRF45_cpx, CRF49_cpx, CRF71_BF1 and other unassigned recombinant forms.
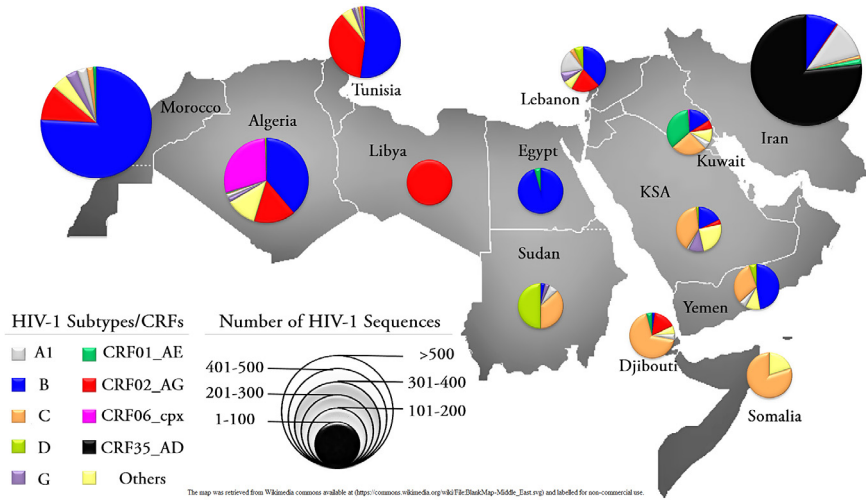
Fig. 1. HIV-1 subtype/circulating recombinant form (CRF) distribution in the Middle East and North Africa (MENA) countries using Los Alamos HIV database sequences collected in the region. The size of each pie chart is proportional to the number of HIV-1 sequences analyzed. The map was retrieved from Wikimedia commons available at (https://commons.wikimedia.org/wiki/File:BlankMap-Middle_East.svg) and labelled for non-commercial use.
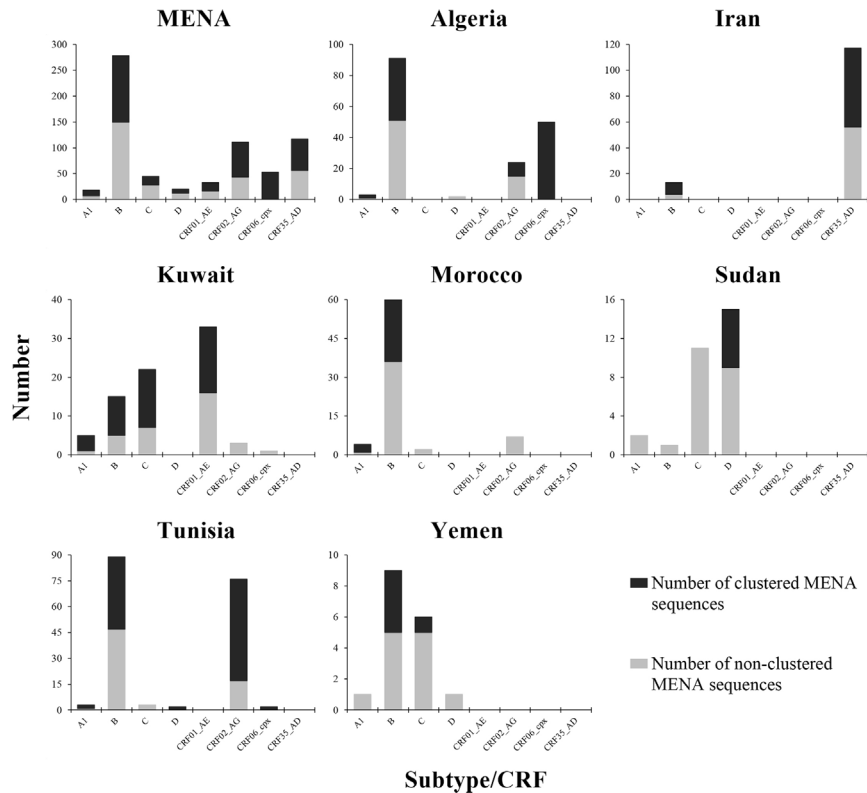

Fig. 2. The total number of Middle East/North African (MENA) clustering vs. non-clustering HIV-1 sequences stratified by country and subtype/CRF.

http://dx.doi.org/10.1016/j.heliyon.2017.e00352

3.3. Phylogenetic evidence of inter-country movement of HIV-1 in the MENA

We identified a total of 73 phylogenetic clusters, of which 36 were dyads (two sequences), 33 were networks (3–14 sequences) and four were large clusters (more than 14 sequences). The majority of the clusters were country-specific (n = 68, 93%). However, five clusters had sequences collected in more than one country as follows: For subtype B, one network and one large cluster having sequences from Algeria and Tunisia, and one network with sequences collected in Morocco and Tunisia. For CRF02_AG, one dyad having sequences from Algeria and Tunisia. For subtype C, one dyad having sequences collected in Yemen and Somalia (Supplementary File 1). Lineage mixing was more common in North Africa (NA) compared with the Middle East (ME, 12% of the NA sequences were found in mixed clusters [n = 30/242] vs. 1% of the ME sequences [1/121], p < 0.001, FET). The total number of sequences included in dyads for all subtypes/CRFs was 72 sequences (20%), compared to 168 sequences (46%) in networks and 123 sequences (34%) in large clusters. To analyse lineage mixing for MENA sequences from a global perspective, we examined the number of reference sequences that were part of all statistically supported phylogenetic clusters (both MENA-specific with ≥80% MENA sequences and non-MENA clusters with <80% MENA sequences). Asian sequences were found to be the most common for CRF01_AE and CRF35_AD while Western European/North American sequences were the most common for subtype B and CRF02_AG. West-Central African sequences were the most common for CRF06_cpx and were found frequently in CRF02_AG clusters (Fig. 3).
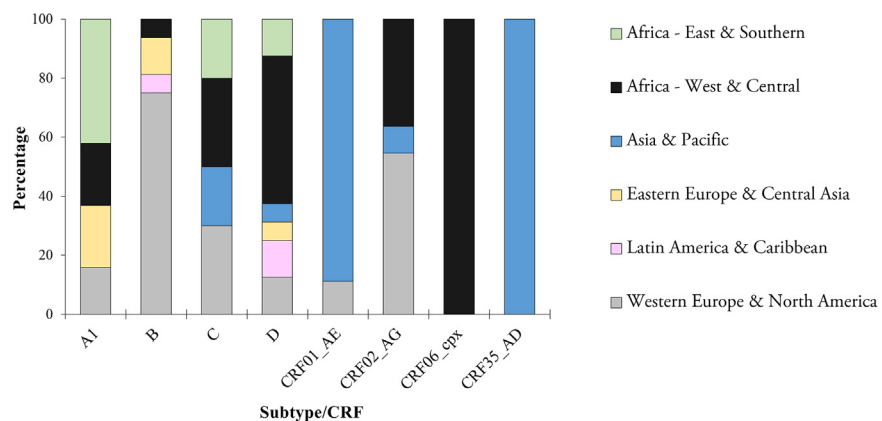

Fig. 3. The percentage of the GenBank reference HIV-1 sequences that were found in the statistically supported phylogenetic clusters with Middle East/North African (MENA) sequences. The stratification was based on UNAIDS regional classification of countries.

3.4. CRFs clustered to a larger extent than pure subtypes in the MENA

The MENA transmission cluster analysis was conducted on four pure subtypes (A1, B, C and D) and on four circulating recombinant forms (CRF01_AE, CRF02_AG, CRF06_cpx and CRF35_AD). Of the identified 73 phylogenetic clusters, 46 (63%) were pure subtype clusters that included 165 sequences (45%), with 24 dyads, 21 networks and one large cluster. Networks included the majority of clustering pure subtypes sequences (n = 100, 61%) followed by dyads (n = 48, 29%) and large clusters (n = 17, 10%). For the CRFs, 27 clusters were identified, that included 198 sequences (55%) and were classified into 12 dyads, 12 networks and three large clusters. Large clusters included the majority of clustering CRF sequences (n = 106, 54%) followed by networks (n = 68, 34%) and dyads (n = 24, 12%). The median size of the pure subtype clusters was two sequences/cluster (range: 2–17, IQR: 2–4) compared with three sequences/cluster for CRFs (range: 2–48, IQR: 2–7) [p = 0.288, M-W]. Clustering CRF sequences had a higher probability to be found in larger transmission clusters compared with pure subtype sequences (88% of the CRF sequences were found in networks or large clusters [n = 174/198] vs. 71% of the pure subtype sequences [117/165], p < 0.001, FET). CRFs had also a higher percentage of clustering compared to pure subtypes (p < 0.001, FET). The details of subtype/CRF clustering stratified per country is illustrated in (Fig. 2).

3.5. Molecular clock analyses of the largest MENA transmission clusters

The Tunisian CRF02_AG cluster of 48 sequences (collected during 2012–2015), and the Algerian CRF06_cpx cluster with 43 sequences (collected during 2001–2013), constituted the two largest transmission clusters in the MENA and were subjected to detailed evolutionary analysis. For evolutionary rate analysis, the median coefficient of variation was 0.62 for the CRF02_AG cluster and 0.42 for the CRF06_cpx cluster, suggesting a large degree of rate variation among branches and that a relaxed clock model would be appropriate. The median evolutionary rate was 2.80×10^{-3} substitutions/site/year for CRF02_AG (95% Highest Posterior Density interval [HPD]: 1.62×10^{-3}-3.74×10^{-3}) and 2.89×10^{-3} substitutions/site/year for CRF06_cpx (95% HPD: 1.74×10^{-3}-4.00×10^{-3}). The tMRCA of the large CRF02_AG cluster in Tunisia dated back to 2002 (95% HPD: 1995–2004), and the tMRCA of the large CRF06_cpx cluster in Algeria dated back to 1991 (95% HPD: 1983–1996). The gap between the median tMRCA and the last date of sequence collection of each cluster was 13 years and 22 years for CRF02_AG and CRF06_cpx, respectively (Fig. 4).
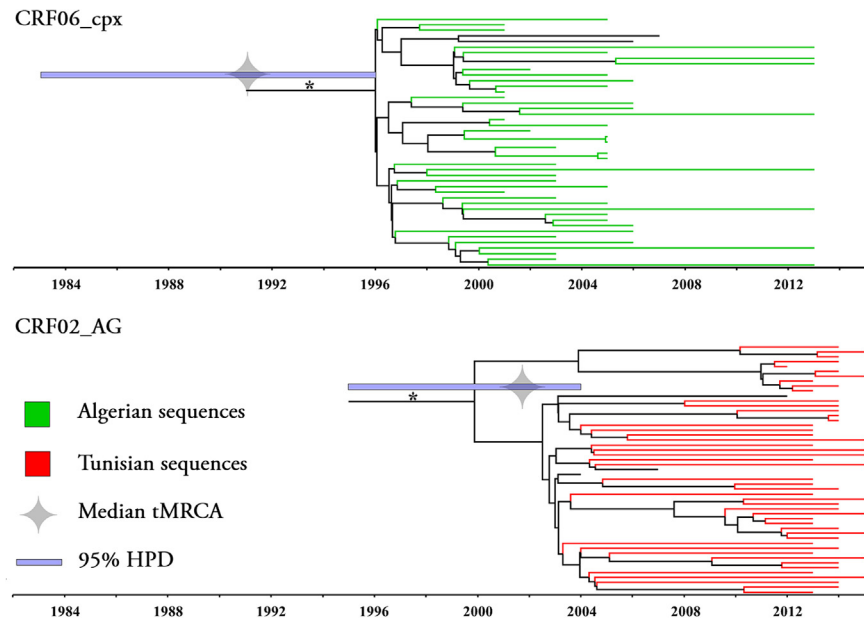
Fig. 4. Maximum clade credibility (MCC) trees for the two largest Middle East and North Africa (MENA) HIV-1 clusters shown on the same timescale. The upper section represents the Algerian CRF06_cpx phylogenetic cluster identified through maximum likelihood (ML) analysis with approximate Likelihood Ratio Test Shimodaira-Hasegawa like (aLRT-SH) value of 0.98 which contained 43 Algerian sequences and two Chadian sequences. The Algerian sequences were collected during 2001–2013. Bayesian analysis revealed that median time to the most recent common ancestor (tMRCA) of the cluster dated back to 1991 (95% highest posterior density interval [HPD]: 1983–1996). The lower section represents the Tunisian CRF02_AG phylogenetic cluster identified through ML analysis with aLRT-SH value of 0.94 which contained 48 Tunisian sequences and one Danish, one Swedish and one Nigerian sequences. The Tunisian sequences were collected during 2012–2015. Bayesian analysis revealed that tMRCA of the cluster dated back to 2002 (95% HPD: 1995–2004). The phylogenetic trees were edited in the software FigTree available freely at (http://tree.bio.ed.ac.uk/software/figtree/).

4. Discussion

The phylogenetic investigation of HIV-1 transmission in the MENA as a single region has not been performed previously and is important for several reasons. First, about 6% of the global population reside in the MENA (approximately 460 million people, United Nations World Population Prospects: The 2015 Revision) and HIV-1 research in this region has been limited. In addition, previous studies have focused on a single country or part of a country and have been limited by low sequence numbers. Furthermore, the majority of countries in the MENA share a sociocultural perspective that might indirectly influence individual behaviour (Alkaiyat and Weiss, 2013), and studying the HIV-1 epidemic in the region as a whole might reveal patterns that can help to guide preventive measures. Finally, the political instability in the MENA, with civil wars ongoing in several countries necessitates increased awareness and surveillance as it has been shown that such

instabilities can result in increased transmission and spread of infectious diseases, including HIV-1 infection (Afonso et al., 2012; Esbjornsson et al., 2011; Mansson et al., 2009). Based on the aforementioned reasons we aimed to characterise the HIV-1 epidemic in the MENA utilizing the powerful tool of phylogenetic inference, as genetic analysis of HIV-1 *pol* sequences has been shown to hold the potential of resolving epidemiologic linkages (Hassan et al., 2017; Hue et al., 2004; Pybus and Rambaut, 2009).

The major result of our study is the demonstration of a high proportion of phylogenetic clustering (54%). A high proportion of phylogenetic clustering is not unique to the MENA, and similar proportions have been previously reported in diverse cohorts from different geographic locations (Aldous et al., 2012; Esbjornsson et al., 2016; Kouyos et al., 2010). Since the majority of the phylogenetically identified transmission clusters were confined by country boundaries, this gives an indication that the sequences within these clusters, were genuinely linked epidemiologically. This is in contrast to the regional perception that HIV-1 transmission is mainly associated with exogenous imports. This view was previously contradicted by (McFarland et al., 2010), and is further negated by the results of our study. The identification of inter-country mixing of different HIV-1 lineages across borders is not a unique finding to the MENA region and similar results were recently reported in Central-West Africa and Europe (Esbjornsson et al., 2016; Faria et al., 2012). This highlights the need of coordinated regional preventive efforts to halt or limit the spread of HIV-1 within and between countries in the MENA.

Despite the considerable heterogeneity of the MENA HIV-1 epidemic, with several established and locally spread subtypes and CRFs, certain patterns could be deduced. First, the epidemic in the North African countries (Algeria, Morocco and Tunisia) was dominated by subtype B and CRF02_AG, in addition to CRF06_cpx in Algeria. It is possible that these countries represent a transitional region where parts of the African and European HIV-1 epidemics are mixing. The facts that subtype B is the most common variant in Europe and that West Africa represents the epicentre of CRF02_AG and CRF06_cpx support this hypothesis (Beloukas et al., 2016; Faria et al., 2012; Yebra et al., 2016). In addition, the evolutionary analysis of the largest CRF clusters in North Africa indicates that local spread has been established from early 1990s for CRF06_cpx and early 2000s for CRF02_AG with a possibility of forward transmission of these variants. The majority of sequences available from Libya and Egypt were originally used to investigate nosocomial spread of CRF02_AG and subtype B in the mid and early 1990s and may not fully represent the national epidemics in these countries (de Oliveira et al., 2006; El Sayed et al., 2000). Hence, further studies are needed to draw firm conclusions about HIV-1 transmission in Libya and Egypt, similar to the recent report by (Daw et al., 2017). The most common HIV-1 variants in countries of

horn of Africa (Djibouti and Somalia) and Sudan were subtypes C and D. This likely reflects the close geographic proximity to major transportation routes to Ethiopia, Kenya and Uganda where these subtypes prevail (Lihana et al., 2012; Tully and Wood, 2010). The picture of the HIV-1 epidemic in the Middle East is more obscure with the exception of Iran, as the majority of the Middle East countries lacked sequences in LADB (Bahrain, Iraq, Jordan, Occupied Palestinian Territories, Oman, Qatar, Syria, and United Arab Emirates). In Iran, CRF35_AD dominated the HIV-1 epidemic, particularly among IDUs as has been shown previously (Eybpoosh et al., 2016). The overall complex genetic diversity of HIV-1 in the MENA has been previously reported by Rolland and Modjarrad, and our results are in line with their conclusions (Rolland and Modjarrad, 2015).

One observation that warrants meticulous study from other regions was the finding of a higher proportion of clustering among recombinant forms and the tendency of these clusters to have a larger size compared to the pure subtypes. Previous studies examined the HIV-1 genetic effects on factors such as disease progression and pathogenicity with conclusions of provisional links among these variables (Abraha et al., 2009; Esbjornsson et al., 2010; Njai et al., 2006; Palm et al., 2014; Vasan et al., 2006). Based on that, the viral genetic effects on epidemiologic features of the virus infection might not be an unforeseen result. However, another study failed to show such an effect (Rubio et al., 2014), which makes our observation pending for further assessment to disentangle the role of virus recombination in HIV-1 epidemiology.

Several shortcomings of our study should be addressed. First, eight of the 21 MENA countries in the Middle East lacked HIV-1 sequences in LADB. Second, only seven of the 13 countries with sequences available for analysis had more than 50 unique sequences submitted to the LADB. Third, many sequences were part of cross sectional studies, sometimes with non-representative and scattered sampling. Fourth, the sampling density was relatively low (< 1%) for the analysed countries resulting in an incomplete picture of the MENA epidemic (Novitsky et al., 2014) (Supplementary File 1). It is important to emphasize that phylogenetic clustering only represents indirect evidence of epidemiologic linkage and may not fully represent the true transmission networks, particularly in datasets with low sequence coverage of the studied epidemic (Romero-Severson et al., 2014). Finally, the results of HIV-1 subtyping varied to some degree depending on the method of subtyping and could not be affirmed particularly for the short sequences and those with potential existence of recombination.

5. Conclusions

In summary we investigated the transmission of HIV-1 in the MENA as a single region using the phylogenetic analysis approach for the first time to our

knowledge. We demonstrate a high proportion of established transmission networks in the MENA. Phylogenetic analysis is an important tool to determine HIV-1 transmission networks and links between most-at-risk-populations and countries. Targeting these groups with awareness and prevention programs as well as screening for undiagnosed infections is fundamental in reverting the increase of HIV-1 infections both globally and in the MENA region.

Declarations

Author contribution statement

Malik Sallam: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Gülşen Özkaya Şahin, Mikael Ingman, Anders Widell: Analyzed and interpreted the data.

Joakim Esbjörnsson: Conceived and designed the experiments; Analyzed and interpreted the data.

Patrik Medstrand: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Funding statement

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at http://dx.doi.org/10.1016/j.heliyon.2017.e00352.

References

Abraha, A., Nankya, I.L., Gibson, R., Demers, K., Tebit, D.M., Johnston, E., et al., 2009. CCR5- and CXCR4-tropic subtype C human immunodeficiency virus type 1 isolates have a lower level of pathogenic fitness than other dominant group M subtypes: implications for the epidemic. J. Virol. 83, 5592–5605.

Afonso, J.M., Morgado, M.G., Bello, G., 2012. Evidence of multiple introductions of HIV-1 subtype C in Angola. Infect. Genet. Evol. 12, 1458–1465.

Aldous, J.L., Pond, S.K., Poon, A., Jain, S., Qin, H., Kahn, J.S., et al., 2012. Characterizing HIV transmission networks across the United States. Clin. Infect. Dis. 55, 1135–1143.

Alkaiyat, A., Weiss, M.G., 2013. HIV in the Middle East and North Africa: priority, culture, and control. Int. J. Public Health 58, 927–937.

Anisimova, M., Gil, M., Dufayard, J.F., Dessimoz, C., Gascuel, O., 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst. Biol. 60, 685–699.

Bakhouch, K., Oulad-Lahcen, A., Bensghir, R., Blaghen, M., Elfilali, K.M., Ezzikouri, S., et al., 2009. The prevalence of resistance-associated mutations to protease and reverse transcriptase inhibitors in treatment-naive (HIV1)-infected individuals in Casablanca, Morocco. J. Infect. Dev. Ctries. 3, 380–391.

Bazinet, A.L., Zwickl, D.J., Cummings, M.P., 2014. A gateway for phylogenetic analysis powered by grid computing featuring GARLI 2.0. Syst. Biol. 63, 812–818.

Beloukas, A., Psarris, A., Giannelou, P., Kostaki, E., Hatzakis, A., Paraskevis, D., 2016. Molecular epidemiology of HIV-1 infection in Europe: An overview. Infect. Genet. Evol. 46, 180–189.

Ben Halima, M., Pasquier, C., Slim, A., Ben Chaabane, T., Arrouji, Z., Puel, J., et al., 2001. First molecular characterization of HIV-1 Tunisian strains. J. Acquir. Immune Defic. Syndr. 28, 94–96.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L., 2006. GenBank. Nucleic Acids Res. 34, D16–20.

Bouzeghoub, S., Jauvin, V., Recordon-Pinson, P., Garrigue, I., Amrane, A., et al., 2006. High diversity of HIV type 1 in Algeria. AIDS Res. Hum. Retroviruses 22, 367–372.

Brenner, B., Wainberg, M.A., Roger, M., 2013. Phylogenetic inferences on HIV-1 transmission: implications for the design of prevention and treatment interventions. AIDS 27, 1045–1057.

Daw, M.A., El-Bouzedi, A., Ahmed, M.O., Dau, A.A., In association with the Libyan Study Group of Hepatitis & HIV, 2017. Molecular and epidemiological characterization of HIV-1 subtypes among Libyan patients. BMC Res. Notes 10, 170.

de Oliveira, T., Pybus, O.G., Rambaut, A., Salemi, M., Cassol, S., Ciccozzi, M., et al., 2006. Molecular epidemiology: HIV-1 and HCV sequences from Libyan outbreak. Nature 444, 836–837.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.

El Sayed, N.M., Gomatos, P.J., Beck-Sague, C.M., Dietrich, U., von Briesen, H., Osmanov, S., et al., 2000. Epidemic transmission of human immunodeficiency virus in renal dialysis centers in Egypt. J. Infect. Dis. 181, 91–97.

Esbjornsson, J., Mansson, F., Martinez-Arias, W., Vincic, E., Biague, A.J., da Silva, Z.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.

Esbjornsson, J., Mild, M., Audelin, A., Fonager, J., Skar, H., Bruun Jorgensen, L., et al., 2016. HIV-1 transmission between MSM and heterosexuals, and increasing proportions of circulating recombinant forms in the Nordic Countries. Virus Evol. 2.

Esbjornsson, J., Mild, M., Mansson, F., Norrgren, H., Medstrand, P., 2011. HIV-1 molecular epidemiology in Guinea-Bissa. West Africa: origin, demography and migrations. PLoS One 6, e17025.

Eybpoosh, S., Bahrampour, A., Karamouzian, M., Azadmanesh, K., Jahanbakhsh, F., Mostafavi, E., et al., 2016. Spatio-Temporal History of HIV-1CRF35_AD in Afghanistan and Iran. PLoS One 11, e0156499.

Faria, N.R., Suchard, M.A., Abecasis, A., Sousa, J.D., Ndembi, N., Bonfim, I., et al., 2012. Phylodynamics of the HIV-1CRF02_AG clade in Cameroon. Infect. Genet. Evol. 12, 453–460.

Gokengin, D., Doroudi, F., Tohme, J., Collins, B., Madani, N., 2016. HIV/AIDS: trends in the Middle East and North Africa region. Int. J. Infect. Dis. 44, 66–73.

Grenfell, B.T., Pybus, O.G., Gog, J.R., Wood, J.L., Daly, J.M., Mumford, J.A., et al., 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303, 327–332.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

Hassan, A.S., Pybus, O.G., Sanders, E.J., Albert, J., Esbjornsson, J., 2017. Defining HIV-1 transmission clusters based on sequence data. AIDS 31, 1211–1222.

Hue, S., Clewley, J.P., Cane, P.A., Pillay, D., 2004. HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. AIDS 18, 719–728.

Jahanbakhsh, F., Ibe, S., Hattori, J., Monavari, S.H., Matsuda, M., Maejima, M., et al., 2013. Molecular epidemiology of HIV type 1 infection in Iran: genomic evidence of CRF35_AD predominance and CRF01_AE infection among individuals associated with injection drug use. AIDS Res. Hum. Retroviruses 29, 198–203.

Khajehkazemi, R., Osooli, M., Sajadi, L., Karamouzian, M., Sedaghat, A., Fahimfar, N., et al., 2013. HIV prevalence and risk behaviours among people who inject drugs in Iran: the 2010 National Surveillance Survey. Sex Transm. Infect. 89 (Suppl 3), iii29–32.

Kouyos, R.D., von Wyl, V., Yerly, S., Boni, J., Taffe, P., Shah, C., et al., 2010. Molecular epidemiology reveals long-term changes in HIV type 1 subtype B transmission in Switzerland. J. Infect. Dis. 201, 1488–1497.

Kouyoumjian, S.P., Mumtaz, G.R., Hilmi, N., Zidouh, A., El Rhilani, H., Alami, K., et al., 2013. The epidemiology of HIV infection in Morocco: systematic review and data synthesis. Int. J. STD AIDS 24, 507–516.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., et al., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.

Lihana, R.W., Ssemwanga, D., Abimiku, A., Ndembi, N., 2012. Update on HIV-1 diversity in Africa: a decade in review. AIDS Rev. 14, 83–100.

Mahfoud, Z., Afifi, R., Ramia, S., El Khoury, D., Kassak, K., El Barbir, F., et al., 2010. HIV/AIDS among female sex workers, injecting drug users and men who have sex with men in Lebanon: results of the first biobehavioral surveys. AIDS 24 (Suppl 2), S45–54.

Mansson, F., Biague, A., da Silva, Z.J., Dias, F., Nilsson, L.A., Andersson, S., et al., 2009. Prevalence and incidence of HIV-1 and HIV-2 before, during and after a civil war in an occupational cohort in Guinea-Bissau, West Africa. AIDS 23, 1575–1582.

McFarland, W., Abu-Raddad, L.J., Mahfoud, Z., DeJong, J., Riedner, G., Forsyth, A., et al., 2010. HIV/AIDS in the Middle East and North Africa: new study methods, results, and implications for prevention and care. AIDS 24 (Suppl 2), S1–4.

Mirzoyan, L., Berendes, S., Jeffery, C., Thomson, J., Ben Othman, H., Danon, L., et al., 2013. New evidence on the HIV epidemic in Libya: why countries must

implement prevention programs among people who inject drugs. J. Acquir. Immune Defic. Syndr. 62, 577–583.

Mumtaz, G., Hilmi, N., Akala, F.A., Semini, I., Riedner, G., Wilson, D., et al., 2011. HIV-1 molecular epidemiology evidence and transmission patterns in the Middle East and North Africa. Sex Transm. Infect. 87, 101–106.

Mumtaz, G.R., Riedner, G., Abu-Raddad, L.J., 2014. The emerging face of the HIV epidemic in the Middle East and North Africa. Curr. Opin. HIV AIDS 9, 183–191.

Njai, H.F., Gali, Y., Vanham, G., Clybergh, C., Jennes, W., Vidal, N., 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., Moyo, S., Lei, Q., DeGruttola, V., Essex, M., 2014. Impact of sampling density on the extent of HIV clustering. AIDS Res. Hum. Retroviruses 30, 1226–1235.

Novitsky, V., Moyo, S., Lei, Q., DeGruttola, V., Essex, M., 2015. Importance of Viral Sequence Length and Number of Variable and Informative Sites in Analysis of HIV Clustering. AIDS Res. Hum. Retroviruses 31, 531–542.

Palm, A.A., Esbjornsson, J., Mansson, F., Kvist, A., Isberg, P.E., Biague, A., et al., 2014. Faster progression to AIDS and AIDS-related death among seroincident individuals infected with recombinant HIV-1 A3/CRF02_AG compared with sub-subtype A3. J. Infect. Dis. 209, 721–728.

Pineda-Pena, A.C., Faria, N.R., Imbrechts, S., Libin, P., Abecasis, A.B., Deforche, K., et al., 2013. 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. 19, 337–348.

Pybus, O.G., Rambaut, A., 2009. Evolutionary analysis of the dynamics of viral infectious disease. Nat. Rev. Genet. 10, 540–550.

Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277.

Rolland, M., Modjarrad, K., 2015. Multiple co-circulating HIV-1 subtypes in the Middle East and North Africa. AIDS 29, 1417–1419.

Romero-Severson, E., Skar, H., Bulla, I., Albert, J., Leitner, T., 2014. Timing and order of transmission events is not directly reflected in a pathogen phylogeny. Mol. Biol. Evol. 31, 2472–2482.

Rubio, A.E., Abraha, A., Carpenter, C.A., Troyer, R.M., Reyes-Rodriguez, A.L., Salomon, H., et al., 2014. Similar replicative fitness is shared by the subtype B and unique BF recombinant HIV-1 isolates that dominate the epidemic in Argentina. PLoS One 9, e92084.

Saad, M.D., Al-Jaufy, A., Grahan, R.R., Nadai, Y., Earhart, K.C., Sanchez, J.L., et al., 2005. HIV type 1 strains common in Europ. Africa, and Asia cocirculate in Yemen. AIDS Res. Hum. Retroviruses 21, 644–648.

Struck, D., Lawyer, G., Ternes, A.M., Schmit, J.C., Bercoff, D.P., 2014. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. Nucleic Acids Res. 42, e144.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.

Tully, D.C., Wood, C., 2010. Chronology and evolution of the HIV-1 subtype C epidemic in Ethiopia. AIDS 24, 1577–1582.

UNAIDS, 2011. UNAIDS Middle East and North Africa regional report on AIDS.

UNAIDS, 2016. The Joint United Nations Programme on HIV/AIDS (UNAIDS) website. Accessed on 31 Aug 2016, cited on 15 Dec 2016 www.unaids.org.

Vasan, A., Renjifo, B., Hertzmark, E., Chaplin, B., Msamanga, G., Essex, M., et al., 2006. Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype. Clin. Infect. Dis. 42, 843–852.

Visco-Comandini, U., Cappiello, G., Liuzzi, G., Tozzi, V., Anzidei, G., Abbate, I., et al., 2002. Monophyletic HIV type 1CRF02-AG in a nosocomial outbreak in Benghazi, Libya. AIDS Res. Hum. Retroviruses 18, 727–732.

Wensing, A.M., Calvez, V., Gunthard, H.F., Johnson, V.A., Paredes, R., Pillay, D., et al., 2015. 2015 Update of the Drug Resistance Mutations in HIV-1. Top Antivir. Med. 23, 132–141.

Yebra, G., Kalish, M.L., Leigh Brown, A.J., 2016. Reconstructing the HIV-1CRF02_AG and CRF06_cpx epidemics in Burkina Faso and West Africa using early samples. Infect. Genet. Evol. 46, 209–218.

Zhang, M., Foley, B., Schultz, A.K., Macke, J.P., Bulla, I., Stanke, M., et al., 2010. The role of recombination in the emergence of a complex and dynamic HIV epidemic. Retrovirology 7, 25.