**Introduction**

The current distribution of human immunodeficiency virus type 1 (HIV-1) group M subtypes and circulating recombinant forms (CRFs) around the world resulted from the chance exportation of different viral strains out of Central Africa into new geographic regions were these initiated secondary epidemics [1]. A recent study suggests that spatial accessibility (human migrations and movements through transportation link availability and quality) has played a significant role in HIV-1 spread across sub-Saharan Africa and may explain the heterogeneous distribution of HIV-1 subtypes and CRFs in the different African regions [2].

West Africa is one of the most strongly connected regions in the continent [2] and also appears as an area of intense intra-regional migration [3]. This coincides with an overall dominance of the CRF02_AG variant, that accounts for about 50% of all HIV-1 infections in West Africa [4]. A closer inspection of the HIV-1 molecular epidemiological profile in this African region, however, reveals an important intra-regional heterogeneity in the distribution of other viral clades, including subtype G and CRF06_cpx. Subtype G is the second most prevalent HIV-1 clade in West Africa accounting for nearly 30% of infections in the region [4]. Its prevalence greatly varies within and between countries, comprising 30–50% of HIV-1 infections across different regions from Nigeria [5,6,7,8,9,10,11,12], 5–15% in Benin, Niger and Togo [13,14,15,16,17], and ≤4% in other western African countries [14,18,19,20,21,22,23,24,25,26,27,28,29,30]. Similarly, the occurrence of the CRF06_cpx clade ranges from 40–50% of HIV-1 infections in Burkina Faso [18,19,20], to 5–15% in Benin, Ghana, Mali, Niger, Nigeria, Senegal and Togo [5,6,7,8,9,10,11,12,13,14,15,16,17,21,22,23,24,28,29], and <3% in other western African countries [14,26,27,30].

The highly heterogeneous distribution of subtype G and CRF06_cpx across the well-connected western African countries, suggests that spatial accessibility is not enough to fully explain the spatial distribution of those HIV-1 clades in this African region. A recent study conducted by our group suggests that Burkina Faso was the most important epicenter of dissemination of the HIV-1 CRF06_cpx strain at regional level and that CRF06_cpx prevalence decreases exponentially as we move away from the epicenter [31]. Our study also estimated that the CRF06_cpx clade started to spread in West Africa around the late 1970s [31], almost 10 years later than the estimated origin of the CRF02_AG clade in West Central Africa [32]. We postulated that the relatively late introduction of the CRF06_cpx clade into western Africa combined with the stabilization of the HIV epidemic in several
countries from the region since the early/middle 1990s may have resulted in a more limited dissemination away from the epicenter and a more heterogeneous regional distribution of CRF06_cpx when compared with CRF02_AG.

It is unclear whether this hypothesis could also explain the complex distribution of subtype G in West Africa. The objective of this study was to reconstruct the onset date, dissemination routes and demographic history of the HIV-1 subtype G clade in the African continent. To this end, we used a Bayesian coalescent-based framework to analyze 305 HIV-1 subtype G pol sequences isolated from 11 different countries from West (Benin, Ghana, Nigeria, Senegal and Togo), West Central (Cameroon, Equatorial Guinea and Gabon), and Central Africa (Angola, Democratic Republic of Congo and Republic of Congo) over a period of 20 years (1992 to 2011).

Materials and Methods

Sequence Dataset

All HIV-1 subtype G pol sequences from West and Central African countries that covered the entire protease and partial reverse transcriptase (PR/RT) regions (nt 2253–3272 relative to HXB2 clone) and for which the sampling year was known, were downloaded from the Los Alamos HIV Sequence Database (www.hiv.lanl.gov) by August 2013. The subtype assignment of all sequences was confirmed by: REGA HIV subtyping tool v.2 [33], Maximum Likelihood (ML) phylogenetic analysis, and boot-scanning analysis. A ML phylogeny with HIV-1 group M subtype reference sequences was constructed with the PhyML 3.0 program [34] using an online web server [35]. The ML tree was inferred under the GTR+I+G nucleotide substitution model recommended by the jModeltest program [36]. The heuristic tree search was performed using the SPR branch-swapping algorithm and branch support was calculated with the approximate likelihood-ratio (aLRT) SH-like test [37]. In bootscanning analysis, supporting branching of query sequences with HIV-1 group M subtypes reference sequences was determined in Neighbor-Joining trees constructed with the Kimura two-parameter model, within a 250 bp window moving in steps of 10 bases, using Simplot software v.3.5.1 [38]. We detected that 4.7% of the subtype G pol sequences available in database had incorrect subtype classification, consistent with previous estimations [39]. Sequences with incorrect classification, multiple sequences from the same individual and sequences from countries poorly represented (n<4 sequences) were removed, resulting in a final data set of 305 HIV-1 subtype G pol African sequences (Table 1). All codon positions known to be associated with major antiretroviral drug resistance were maintained in the final alignment because ML trees constructed on alignments with or without such positions resulted in the same overall topology (data not shown). Final sequence alignment is available from the authors upon request.

Analysis of Spatiotemporal Dispersion Pattern and Demographic History

The evolutionary rate (μ, nucleotide substitutions per site per year, subst./site/year), the age of the most recent common ancestor (TMRCA, years), the ancestral geographic movements, and the mode and rate (r, years-1) of population growth of HIV-1 subtype G clades circulating in Africa were jointly estimated using the Bayesian Markov Chain Monte Carlo (MCMC) approach as implemented in BEAST v.1.8 [40,41] with BEAGLE to improve run-time [42]. Analyses were performed under a GTR+I+G nucleotide substitution model. The temporal scale of evolutionary process was estimated from the sampling dates of the sequences using a relaxed uncorrelated lognormal molecular clock model and a uniform prior on clock rate (1.0–4.0×10^-3 subst/site/year) [43]. Migration events throughout the phylogenetic history were inferred using a reversible discrete Bayesian phylogeographic model [44], in which all possible reversible exchange rates between locations were equally likely, and a CTMC rate reference prior [45]. To quantify the dissemination process, we estimated the number of viral migrations among locations using ‘Markov Jump’ counts [46] of location-state transitions along the posterior tree distribution as previously described [47,48]. Changes in effective population size through time were initially estimated using a flexible Bayesian Skyline coalescent model [49] that does not require strong prior assumptions of demographic history. Estimates of the population growth rate were subsequently obtained using the parametric model (logistic, exponential or expansion) that provided the best fit to the demographic signal contained in datasets. Comparison between demographic models was performed using the log marginal likelihood (ML) estimation based on path sampling (PS) and stepping-stone sampling (SS) methods [50]. MCMC chains were run for 50–500×10^6 generations. Adequate chain mixing and uncertainty in parameter estimates were assessed by calculating the Effective Sample Size (ESS) and the 95% Highest Probability Density (HPD) values, respectively, using the TRACER v1.6 program [51]. Maximum clade credibility (MCC) trees were summarized from the posterior distribution of trees with TreeAnnotator and visualized with FigTree v.1.4.0 [52]. Migratory events across time were summarized using the cross-platform SPREAD application [53].

Results

Origin of the HIV-1 Subtype G and Identification of Major African Clades

We analyzed 305 HIV-1 subtype G pol sequences isolated from 11 African countries between 1992 and 2011 that were sampled across seven different location states (Table 1). Neighboring countries from West (Togo/Ghana), West Central (Gabon/Equatorial Guinea) and Central (Angola/Democratic Republic of Congo/Republic of Congo) Africa comprising few samples (n<15) were grouped into the same location (Table 1). According to the Bayesian MCMC analysis, the median evolutionary rate of the HIV-1 subtype G lineage at pol gene was estimated at 2.3×10^-3 (95% HPD: 1.8×10^-3–2.8×10^-3) subst./site/year. The estimated coefficient of rate variation in our dataset was 0.28 (95% HPD: 0.24–0.32), thus supporting a significant variation of substitution rate among branches and the use of a relaxed molecular clock model. The most probable root location of the subtype G clade was placed in Central Africa (posterior state probability, PSp=0.88), and the onset date of this clade was estimated to be 1968 (95% HPD: 1956–1976) (Fig. 1).

The Bayesian MCC (Fig. 1) and ML (Fig. S1) trees point to a clear phylogeographic subdivision of subtype G strains from West and Central Africa. Sequences from western Africa branched mostly in two large monophyletic clades (GWA1 and GWA11) that were nested among the most basal clades from Central and West Central Africa (GCA). Distribution of HIV-1 subtype G clades greatly varies across countries within each region (Fig. 2). The GWA1 clade was the predominant subtype G lineage detected in Nigeria (80%) and the GWA11 clade predominates in Togo/Ghana (86%). The subtype G epidemic in Benin is dominated by both GWA1 (47%) and GWA11 (40%) clades, whereas GWA1 (50%) and GCA (42%) clades prevail among subtype G infections in Senegal. Basal GCA clades predominate in countries from both central (100%) and west central (50–71%) regions.
Spatiotemporal Dispersal Pattern of the HIV-1 Subtype G African’s Epidemic

Reconstruction of viral migrations across time revealed the occurrence of multiple introductions of HIV-1 subtype G strains from Central into West Africa since the middle 1970s (Fig. 3). The earliest viral migrations led to the origin of the GWA-I and GWA-II lineages. The GWA-I clade most probably emerged in Nigeria (PSP = 1) around 1974 (95% HPD: 1966–1981) and from this country was later disseminated to Benin, Cameroon, Equatorial Guinea, Ghana, and Senegal. The GWA-II clade most probably emerged in Togo/Ghana (PSP = 0.68) around 1979 (95% HPD: 1973–1984) and was disseminated to Nigeria in 1981 (95% HPD: 1976–1986), where it further spread locally. In the following years, the GWA-II clade was disseminated from both Togo/Ghana and Nigeria to Benin, Cameroon, Gabon, and Senegal. Our phylogeographic analysis also detected several independent introductions of subtype G variants from Central Africa into Cameroon (Figs 1 and 3). The earliest introductions occurred between the late 1970s and the middle 1980s and gave rise to at least three local Cameroonian clades; one of which was further disseminated to Gabon, Equatorial Guinea, Senegal and Angola.

We next quantified the viral flux between locations using Markov jump counts (Fig. 4 and Table S1). Nigeria (16.4), central African countries (14.0), and Togo/Ghana (8.3) displayed positive net viral migration rates (efflux minus influx), whereas Benin (−14.2), Cameroon (−10.1), Gabon/Equatorial Guinea (−8.4), and Senegal (−6.8) displayed negative net viral migration fluxes. The highest numbers of viral transitions were from Nigeria to Benin (8.1), Togo/Ghana (5.5) and Cameroon (4.4), from Central Africa to Cameroon (6.1) and Senegal (4.2), and from Togo/Ghana to Benin (5.7), Nigeria (4.1) and Cameroon (3.9). The estimated viral flux to Gabon/Equatorial Guinea from Cameroon (2.6), Central Africa (2.3), Nigeria (2.3) and Togo/Ghana (2.3) was very similar.

Demographic History of HIV-1 Subtype G African’s Epidemic

Estimations of effective population size (Ne) changes over time were initially obtained using a Bayesian skyline plot (BSP) coalescent model. The BSP analysis of the complete dataset suggests that the subtype G African epidemic experienced a fast exponential growth during the 1970s and 1980s, followed by a more recent stabilization since the early 1990s (Fig. 5A). This overall growth pattern, however, represents the combined population dynamics of the different African subtype G clades that are being disseminated within different countries and regions. In order to better understand the regional differences in the demographic histories of HIV-1 subtype G African epidemics, the GCA, GWA-I and GWA-II clades were analyzed separately (Table S2).

The BSP analyses suggest that all African subtype G clades displayed a similar population growth pattern characterized by an initial phase of exponential growth followed by a decline in growth rate since the early/middle 1990s (Figs. 5B, D and F). To estimate the mean epidemic growth rate of the major subtype G African clades, log ML for the logistic, exponential and expansion growth models were calculated using both PS and SS methods. The best-fit demographic model for all subtype G clades was the logistic one (log BF > 5) (Table S3) that was then used to estimate the initial epidemic growth rate. The overall time-scale and demographic pattern obtained from both BSP (Figs. 5B, D and F) and logistic growth coalescent tree priors (Figs. 5C, E and G) were very similar and important differences in the epidemic growth rate were detected across subtype G clades from West and Central Africa. According to the logistic growth coalescent model, the mean growth rate of clades GWA-I (0.75 year⁻¹) and GWA-II (0.95 year⁻¹) was about two times higher than that estimated for the clade GCA (0.47 year⁻¹) (Fig. 5).

Discussion

This study indicates that the HIV-1 subtype G likely originated in Central Africa around the late 1960s. The root position of the subtype G clade is fully consistent with the most accepted model that traces the origin of all HIV-1 group M subtypes to the DRC [54,55,56,57] and is also resistant to the problem of sampling bias because sequences from Central Africa represent a minor fraction (9.2%) of the total subtype G sequences included in our study. The TMRCA of subtype G clade here estimated (1968: 1956–1976) is also fully consistent to that previously estimated for this subtype (1970: 1960–1978) [58]. This onset date is comparable to that estimated for subtype F (1967: 1956–1976) [59]; but more recent than that of subtypes A1 (1954: 1940–1968), C (1955: 1934–1972), D (1947: 1938–1953) [58].

After emerging in Central Africa around the late 1960s, the HIV-1 subtype G was disseminated to West and West Central

### Table 1. HIV-1 subtype G pol dataset.

<table>
<thead>
<tr>
<th>Region</th>
<th>Country</th>
<th>Location*</th>
<th>N</th>
<th>Sampling interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>West Africa</td>
<td>Benin</td>
<td>BJ</td>
<td>15</td>
<td>2004–2009</td>
</tr>
<tr>
<td></td>
<td>Ghana</td>
<td>TG/GH</td>
<td>8</td>
<td>2002–2009</td>
</tr>
<tr>
<td></td>
<td>Nigeria</td>
<td>NG</td>
<td>183</td>
<td>1992–2010</td>
</tr>
<tr>
<td></td>
<td>Senegal</td>
<td>SN</td>
<td>12</td>
<td>1998–2010</td>
</tr>
<tr>
<td></td>
<td>Togo</td>
<td>TG/GH</td>
<td>13</td>
<td>2006–2008</td>
</tr>
<tr>
<td>West Central Africa</td>
<td>Cameroon</td>
<td>CM</td>
<td>31</td>
<td>1997–2011</td>
</tr>
<tr>
<td>Central Africa</td>
<td>Angola</td>
<td>AO/CD/CN</td>
<td>13</td>
<td>1997–2010</td>
</tr>
<tr>
<td></td>
<td>DRC</td>
<td>AO/CD/CN</td>
<td>12</td>
<td>1993–2007</td>
</tr>
<tr>
<td></td>
<td>Republic of Congo</td>
<td>AO/CD/CN</td>
<td>8</td>
<td>2003</td>
</tr>
</tbody>
</table>

*Location assigned in the Bayesian phylogeographic analysis. DRC: Democratic Republic of Congo.

doi:10.1371/journal.pone.0098908.t001
Africa a few years later (1975–1980). Our phylogeographic analysis supports the occurrence of multiple introductions of HIV-1 subtype G strains from central into the western and west central African regions. Some of the viral strains disseminated during the 1970s fueled secondary outbreaks that led to the origin of specific subtype G clades. The major subtype G clades detected in our study were the GWA-I that most probably emerged in Nigeria around the middle 1970s, and the GWA-II that most probably emerged in Togo or Ghana around the late 1970s. Although we grouped sequences from Togo and Ghana into one single location, the much higher prevalence of subtype G in Togo (9%) [16,17] compared with Ghana (<1%) [23,24] suggests that the GWA-II clade probably arose in Togo. We also detected three minor subtype G clades that resulted of independent introductions of viral strains from central Africa into Cameroon between the late 1970s and the middle 1980s.

Nigeria and Togo/Nigeria were inferred as the most important epicenters of dissemination of the GWA-I and GWA-II clades at regional level, respectively. The GWA-I clade, which corresponds to the clade previously designated G’ [5,8], was the predominant subtype G lineage in Nigeria (80%), Senegal (50%), and Benin (47%), and also comprises a significant fraction of subtype G infections in Gabon/Equatorial Guinea (20%), Cameroon (13%) and Togo/Ghana (9%). The GWA-II clade predominates in Togo/
Ghana (86%) and is responsible for a significant fraction of subtype G infections in Benin (40%), Gabon/Equatorial Guinea (30%), Nigeria (20%), Cameroon (16%) and Senegal (8%). The subtype G clades introduced into Cameroon were mainly disseminated to the neighboring countries in the central west region (Gabon and Equatorial Guinea), although a few disseminations to Senegal were also detected. These results indicate that founder subtype G strains introduced into Nigeria and Togo have been much more efficiently disseminated at regional level than those introduced into Cameroon.

Our demographic reconstructions also revealed another important difference between African subtype G clades mainly disseminated in the western region (GWA-I and GWA-II) and those mainly disseminated in the west central and central regions (GCA). Although all African subtype G clades displayed a similar population growth pattern characterized by an initial phase of exponential growth followed by a decline in growth rate since the early/middle 1990s; the mean epidemic growth rate of GWA-I (0.75 year⁻¹) and GWA-II (0.95 year⁻¹) clades was about two times higher than that estimated for GCA (0.47 year⁻¹) clades. This suggests that subtype G clades introduced into Nigeria and Togo during the 1970s probably encountered more favorable conditions for local and regional expansion than those disseminated within central and west-central African countries around the same time. The median growth rates of the GWA-I and GWA-II clades were comparable to that estimated for the CRF06_cpx in western

Figure 2. Prevalence of GWA-I, GWA-II and GCA clades among subtype G infected individuals from different African countries, estimated from phylogenetic analyses presented in Figs. 1 and S1. The total number of subtype G sequences analyzed in each locality is indicated. Each clade is represented by a color as indicated at the legend. doi:10.1371/journal.pone.0098908.g002
Africa (0.82 year\(^{-1}\)) [31]; whereas the median growth rate of the GCA clades was roughly similar to that estimated for subtype G in Cuba (0.54 year\(^{-1}\)) [60] and higher than that estimated for HIV-1 group M in Democratic Republic of Congo (0.17 year\(^{-1}\)) [61].

The faster epidemic growth and the broader geographic dissemination of subtype G strains introduced into West Africa compared with those circulating in the central west and central African regions could be associated to clade-specific or regional-specific differences in viral transmissibility. It has been suggested that accessibility between locations have played a major role in the spatial spread of HIV-1 in sub-Saharan Africa [2]. Notably, West Africa is one of the most strongly connected regions in the continent [2] and also displays an intra-regional migration rate (3%) above the African average (2%) [3]. Others factors including urbanization [56,62], iatrogenic interventions [63,64], and forced migration [62,65] might have also played a role in the emergence and spread of HIV in Africa. Such alternative scenarios can now be tested in a Bayesian framework [66] to find the hypothesis that best explain the variability in the rate of HIV spread across African regions.

Despite the strong regional accessibility, the prevalence of subtype G and CRF06_cpx clades greatly vary across western African countries. The clades CRF06_cpx, GWA-I, and GWA-II seem to have experienced very similar dissemination dynamics; although their origin was traced to different western African countries (Burkina Faso, Nigeria and Togo, respectively) [31]. The three HIV-1 clades probably started to spread in West Africa around the same time (1975–1980), expanded during the 1980s with similar epidemic growth rates (0.75–0.95 year\(^{-1}\)), started to stabilize around the early/middle 1990s, and their prevalence is

Figure 3. Spatiotemporal dynamics of HIV-1 subtype G clade dissemination in West and Central Africa. Snapshots of viral migration events occurring at different time intervals between 1970 and 2012 are shown. Lines between locations represent branches in the Bayesian MCC tree along which location transitions occur. Each location is represented by a color as indicated at the legend. SN: Senegal; TG/GH: Togo/Ghana; BJ: Benin; NG: Nigeria; CM: Cameroon; GA/GQ: Gabon/Equatorial Guinea; AO/CD/CG: Angola/DRC/Republic of Congo.
Figure 4. Viral migration rates among locations as measured using ‘Markov jump’ counts. Each panel represents the estimated viral exchanges from and to Angola/DRC/Republic of Congo (A), Cameroon (B), Gabon/Equatorial Guinea (C), Benin (D), Togo/Ghana (E), Nigeria (F), and Senegal (G). The width of the arrows is proportional to the corresponding mean estimated number of viral transitions between locations according to the following scale: thin arrows = 1.0–2.9 transitions, medium arrows = 3.0–5.9 transitions, thick arrows = 6.0–8.9 transitions. No arrows were displayed when the mean estimated number of transitions was below one.

doi:10.1371/journal.pone.0098908.g004
Figure 5. Demographic history of the HIV-1 subtype G and the clades GCA, GWA-I and GWA-II circulating in Central and West Africa. Effective number of infections (y-axis; log10 scale) through time (x-axis; calendar years) estimated using Bayesian skyline (A, B, D, F) and logistic (C, E, G) growth coalescent model. Median estimates of the effective number of infections (solid line) and 95% HPD intervals of the estimates (dashed lines) are shown in each graphic. The median growth rate (with the corresponding 95% credibility interval in parenthesis) of each clade estimated under the logistic growth model is indicated in the upper left corner.

doi:10.1371/journal.pone.0098908.g005

greatly reduced as we moved away from the corresponding epicenters [31]. The relatively late spread of subtype G and CRF06_cpx clades in West Africa combined with: 1) stabilization of the HIV epidemic in several western African countries since the early/middle 1990s, and/or 2) depletion of the susceptible populations most at risk by the firstly introduced CRF02_AG lineage, may have limited the dissemination of these viral clades
far from the epicenter, thus generating a heterogeneous spatial distribution.

The most important limitation of our study was the small sampling size of many African countries. Only Nigeria (n = 103) and Cameroon (n = 31) were represented by a high or relatively high number of sequences. Other western (Benin, Niger, and Togo) and central (Central African Republic, Chad, Equatorial Guinea, and Gabon) African countries with circulation of subtype G at significant levels (~5% of all HIV-1 infections) [13,14,15,16,17,67,68,69,70,71,72] were represented by a small number of sequences (n≤15) that may not fully reflect the country’s subtype G diversity, or were not represented at all in our study (Fig. S2). Thus, a more comprehensive and balanced sampling from countries poorly or not represented here would certainly provide more precise estimates of the relative prevalence and migration routes of clades GWA-I, GWA-II, and GCA across different African regions, and may also result in the identification of new regional viral clades not detected in this study.

It will be also interesting to trace the origins and global dispersal pathways of those subtype G lineages found in countries outside sub-Saharan Africa, particularly in Cuba [73,74,75], Portugal [76,77,78], and Russia [79] where this subtype has been disseminated among local populations. It has been showed that the spread of HIV-2 outbreaks Africa mirrors socio historical ties [80] and a previous study conducted by our group showed that the spread of HIV-2 outwards Africa mirrors socio historical ties [80] and a previous study conducted by our group showed that most subtype G Cuban lineages are nested among basal sequences [80].

The most important limitation of our study was the small sampling size of many African countries. Only Nigeria (n = 103) and Cameroon (n = 31) were represented by a high or relatively high number of sequences. Other western (Benin, Niger, and Togo) and central (Central African Republic, Chad, Equatorial Guinea, and Gabon) African countries with circulation of subtype G at significant levels (~5% of all HIV-1 infections) [13,14,15,16,17,67,68,69,70,71,72] were represented by a small number of sequences (n≤15) that may not fully reflect the country’s subtype G diversity, or were not represented at all in our study (Fig. S2). Thus, a more comprehensive and balanced sampling from countries poorly or not represented here would certainly provide more precise estimates of the relative prevalence and migration routes of clades GWA-I, GWA-II, and GCA across different African regions, and may also result in the identification of new regional viral clades not detected in this study.

Acknowledgments

We thank Dr. Vera Bongertz for critical review of the manuscript and corrections on English language usage. We also thank the article reviewers for their helpful comments.

Author Contributions

Conceived and designed the experiments: GB ED. Performed the experiments: ED DM GB. Analyzed the data: ED GB. Contributed to the writing of the manuscript: ED DM GB.


