Hotspots of Transmission Driving the Local Human Immunodeficiency Virus Epidemic in the Cologne-Bonn Region, Germany

Melanie Stecher,1,2 Martin Hoenigl,1,4 Anna Maria Eis-Hübinger,5,6 Clara Lehmann,1,2 Gerd Fätkenheuer,1 Jan-Christian Wasmuth,2,6 Elena Knops,7 Jörg Janne Vehreschild,1,8 Sanjay Mehta,3,8,b and Antoine Chaillon3,b

1Department I of Internal Medicine, University Hospital of Cologne, and 2German Center for Infection Research (DZIF), partner site Bonn-Cologne, Germany; 3Division of Infectious Diseases, University of California San Diego; 4Division of Pulmonology and Section of Infectious Diseases, Medical University of Graz, Austria; 5Institute of Virology, University of Bonn Medical Center, Department for Internal Medicine I, University Hospital of Bonn, and 6Institute of Virology, University Hospital of Cologne, Germany; and 7Department of Medicine, San Diego VA Medical Center, California

Background. Geographical allocation of interventions focusing on hotspots of human immunodeficiency virus (HIV) transmission has the potential to improve efficiency. We used phylogeographic analyses to identify hotspots of the HIV transmission in Cologne-Bonn, Germany.

Methods. We included 714 HIV-1 infected individuals, followed up at the University Hospitals Cologne and Bonn. Distance-based molecular network analyses were performed to infer putative relationships. Characteristics of genetically linked individuals and assortativity (shared characteristics) were analyzed. Geospatial diffusion (ie, viral gene flow) was evaluated using a Slatkin-Maddison approach. Geospatial dispersal was determined by calculating the average distance between the residences of linked individuals (centroids of 3-digit zip code).

Results. In sum, 217/714 (30.4%) sequences had a putative genetic linkage, forming 77 clusters (size range: 2–8). Linked individuals were more likely to live in areas surrounding the city center (P = .043), <30 years of age (P = .009), and infected with HIV-1 subtype B (P = .002). Clustering individuals were nonassortative by area of residency (−.0026, P = .046). Geospatial analyses revealed a median distance between genetically linked individuals of 23.4 kilometers (km), lower than expected (P < .001). Slatkin-Maddison analyses revealed increased gene flow from central Cologne toward the surrounding areas (P < .001).

Conclusion. Phylogeographic analysis suggests that central Cologne may be a significant driver of the regional epidemic. Although clustering individuals lived closer than unlinked individuals, they were less likely to be linked to others from their same zip code. These results could help public health entities better understand transmission dynamics, facilitating allocation of resources to areas of greatest need.

Keywords. HIV transmission; geospatial dispersal; phylogeographic analyses; public health.

With an estimated 3700 people newly diagnosed with human immunodeficiency virus type 1 (HIV-1) in 2016, the annual number of new diagnoses remains stable in Germany [1]. The metropolitan region Cologne-Bonn has one of the highest rates of new HIV diagnoses in Western Europe (13.7 per 100 000 habitants) [2, 3]. Men having sex with men (MSM) bear the major burden, with 67.4% of all new HIV diagnoses in Germany occurring among MSM in 2017 [2].

This high incidence among MSM in the Cologne-Bonn region may be related to several factors [2]. There is a well-known MSM party scene in central Cologne [4] with an increased prevalence of high-risk behaviors associated with chemsex, which involves the use of methamphetamine, mephedrone, poppers, or cocaine [5, 6]. The use of geosocial networking apps, especially among MSM, also may play a key role in increasing sexual contacts and high sexual risk behavior between individuals living, working, and interacting in a reasonably close proximity [7, 8]. Simultaneously, the number of new HIV diagnoses among people who inject drugs (PWID) is also high in Cologne (4.55/100 000) [2], which may additionally contribute to HIV transmission through needle sharing and high-risk sexual behavior [5, 9]. Besides individual-level risk characteristics, there is a lack of findings on neighborhood-level characteristics and geospatial factors to explain different incidence rates and regional burden of HIV transmission in high-risk populations [10]. Identifying these factors could be critical for geographical allocation of interventions to improve efficiency of targeted prevention efforts [11].
Here we sought to determine the dynamics and spread of HIV transmission throughout the area of Cologne-Bonn using molecular epidemiologic techniques. We then used these data to identify hotspots of HIV transmission and individual/population factors driving these transmission dynamics to guide future public health responses.

METHODS

Ethical Considerations

Our study was conducted within the framework of the Translational Platform HIV (TP-HIV) of the German Center for Infection Research (DZIF) (NCT02149004) and was approved by the ethics committee and scientific steering committee of all partner sites. The study was designed and completed in close cooperation with the University of California San Diego.

Study Population

This analysis includes data from 714 HIV-1 infected antiretroviral therapy (ART) naive individuals, followed at the German University Hospitals Cologne (n = 558) and Bonn (n = 156) between 2001 and 2016. The included data (714/1766; 40.4%) was selected from a pool of 1766 individuals (Cologne, n = 1507 and Bonn, n = 259), who were retrospectively confirmed (ie, via chart review and patient records) to be ART naive at the time of inclusion. By the ethics committee and scientific steering committee of all partner sites. The study was designed and completed in close cooperation with the University of California San Diego.

The available partial HIV-1 pol sequence data (HXB2 position 2550–3556) were obtained from blood plasma, and sequenced between 2001 and 2014 using Sanger sequencing on an ABI 3130xl Genetic Analyzer (Applied Biosystems, Carlsbad, CA, USA) [13]. In 2015–2016, consensus pol sequences were obtained using next-generation sequencing (NGS) (Illumina MiSeq, CA, USA) [14], using a reference cutoff of 10% consistent with Sanger sequencing sensitivity [15, 16]. The Subtype Classification using Evolutionary Algorithms (SCUEAL) program was used to subtype all sequences [17]. To construct the genetic transmission network, we used HIV-TRACE [18], following a procedure described previously [19]. Briefly, we calculated the Tamura-Nei 93 (TN93) [20], genetic distance (GD) between all sequences and then inferred a putative linkage when the GD was <1.5% [21]. Multiple links were then resolved into clusters.

Geospatial Spread of the Epidemic

Geospatial dispersal of the clusters was determined by calculating the average spatial distances between reported residences (centroids of 3-digit zip code) of genetically linked individuals using R packages maptools [22] and mapdata [23].

Epidemic Dynamics and Viral Migration

To evaluate the role of central Cologne as a geographical hotspot in the spread of the local HIV epidemic, we applied the Slatkin-Maddison approach [24] implemented in HyPhy [25] to infer the number of viral migration events between the individuals living in central Cologne and the surrounding areas. To accommodate for the differences in sample size (available number of sequences per location, zip code 506, n = 95 and surrounding areas, n = 619), we repeated these analyses on 1000 iterations of random subsets of equal sequence numbers per location as previously described [26]. We assessed whether nodes in the network were more likely to be linked to individuals living in a neighboring area. We therefore compared the geospatial distance between genetically linked individuals (ie, putative transmission pairs) based on the centroid of the 3-digit zip code of residency to a random distribution. We repeated these simulations with 1000 iterations and evaluated the significance using nonparametric t-test.

Mixing Assessment—Assortativity

We next evaluated the assortativity of the study population, which is the mixing pattern of the population with respect to a given classification [27–29]. Given that our population was predominantly male and MSM, we considered mixing (ie, tendency for nodes that share attributes to link to each other) according to discrete geospatial characteristics (ie, zip code of residency). We computed Newman’s assortativity coefficients [27] to describe the mixing patterns in our dataset using R package igraph [30].

Statistical Analysis

We calculated and reported population characteristics as absolute numbers and percentages, medians plus interquartile ranges (IQR) or means plus 95% confidence intervals (95% CI), as appropriate. All P-values of <.05 were considered significant. We performed a uni- and multivariable logistic
regression model. Our multivariable model was adjusted for age, sex, HIV-1 subtype, risk group, residential area, and region of origin, using stepwise backward elimination to reach the simplest model that explained the data. Odds ratio (OR) and 95% confidence intervals were reported to show the direction and strength of the association. Analyses were conducted using Stata (Stata Statistical Software: Release 14. College Station, TX, USA).

RESULTS

Population Characteristics

A total of 714 individuals (129 females and 585 males) sampled between 2001 and 2016 were included in this study. Of 585 males, 408 (70%) reported having sex with men as their main risk factors, whereas 14 (2.4%) reported injection drug use. HIV-1 subtype B was predominant (n = 539, 75.5%) in our study population. MSM were significantly (P < .001) more likely to be infected with subtype B (n = 383, 71.1%) than individuals reporting heterosexual sex (HTS) (n = 95, 17.6%) and PWID (n = 14, 2.6%). Subtype A was significantly more prevalent in HTS (n = 27, 57.4%) compared to people with other risk factors (P < .001).

A significantly higher number of study participants were treated at the University Hospital of Cologne (n = 558; 78.2%) compared to the University Hospital in Bonn (n = 156; 21.8%) (P < .001). Figure 1A illustrates the geospatial distribution of the participants based on the 3-digits zip code of residency. Most sequences were obtained from individuals living in central Cologne (zip code 506), its surrounding districts, and central Bonn (zip code 531). The population characteristics are displayed in Table 1.

Transmission Network Analysis

In total, 217/714 (30.4%) sequences had a putative linkage with at least one other sequence, forming 77 transmission clusters ranging in size from two to eight sequences. Figure 1B shows the spatial distribution of all clustering individuals.

The HIV transmission network annotated by risk factor is displayed in Figure 2A. Most clustering individuals (195/217; 90.3%) were men. Of these, 146 (74.9%) reported sex with men and 25 (12.8%) sex with women as primary risk factor. Among clustering males reporting HTS as their risk factor, 17 (8%) clustered with MSM. A map with the HIV transmission network colored by area is displayed in Figure 2B. Clustering individuals were primarily living in the surrounding areas 192 (88.5%) versus central Cologne 25 (11.5%) (Figure 2B). Interestingly, the largest cluster, composed of eight individuals who were predominantly MSM (n = 6) and residents of the surrounding areas (n = 7), also included 1 individual from the city center of Cologne who also had a central position within the cluster (6 degrees).

Univariate comparison between clustering and not clustering nodes revealed that linked individuals were significantly more likely to be younger than 30 years of age (P = .050), male (P < .001), and infected with HIV-1 subtype B (P < .001). Individuals reporting MSM contact were more likely to cluster compared to HTS (P = .002) (Table 1). Similar trends were observed in PWID and people living in surrounding areas of Cologne, even though these findings were not significant.

This multivariable model revealed that clustering individuals were more commonly younger than 30 years of age (P = .009), infected with a non-B subtype (P = .002) and living in the surrounding areas of Cologne (P = .043). The association between

---

**Figure 1.** Geospatial distribution of (A) number of sequences in the study population (n = 714), and (B) number of clustering individuals based on the 3-digits zip code of residency (n = 217).
risk group and sex did not remain significant in the multivariable model (Table 1).

**Spatial Diffusion of the Local Transmission Network**

To evaluate the spatial connectivity of the local transmission network and the contribution of geographical clusters of transmission on the local epidemic, we first determined the distance between residences of linked individuals (i.e., distances between centroids of zip of residence). Figure 3 shows the inferred putative links and numbers of clustering individuals (blue circles) based on the centroid of zip code of residency. While 25/217 clustering individuals were identified in the central district of Cologne and 22/217 in central Bonn, the transmission network revealed a high degree of putative linkage (i.e., edges, gold lines) between individuals living in these areas and in the surrounding regions (i.e., south of Bonn, the east and western surrounding areas of Cologne). Among the 82 putative linkages involving individuals living in central Cologne, 76 were links with individuals living in the surrounding areas suggesting the spread of the local transmission network.

Of the 77 distinct clusters, 19 (24.7%) included sequences from 25 (11.5%) individuals living in central Cologne. Clustering individuals from central Cologne were more highly linked than individuals from the surrounding areas (1.04 compared to 0.88 per clustering sequence; \( P < .001 \)).

Overall, the median distance between genetically linked individuals was 23.4 kilometers (IQR 11.3–34.6). At the cluster level, the median distance between linked individuals was 23.0 kilometers (IQR 5.24–31.0), which was significantly lower than the median distance of any random sub-sampled population (median 39.68 kilometers (IQR 23.79–62.59), \( P < .001 \)), (Figure 4A).
Figure 2. Transmission network of HIV in the Cologne-Bonn region. Sex of the individuals (nodes) is indicated by shape. In (A) the color indicates the reported risk group. In (B) the color indicates the residence location, i.e., living in the city center (orange) or surrounding areas (blue) of Cologne-Bonn. All edges represent a genetic distance of ≤1.5%. Abbreviations: HTS, heterosexual sex; MSM, men having sex with men; NA, not applicable; PWID, people with injection drug use.

Figure 3. Geographic inferred transmission network by the location of residence overlaid on the maps of the Cologne and Bonn area. The size of the dots defining the number of clustering individuals in each specific area. The golden lines representing the links between different areas. Most of the larger clusters (>20 individuals) are located in the central parts of Cologne and Bonn (n = 217).
Finally, we sought to determine the source/directionality of HIV gene flow between the central regions of Cologne/Bonn and surrounding areas. We performed migration analyses using a Slatkin-Maddison approach to infer prior and ongoing viral gene flow between central Cologne and its surrounding areas. We found significant viral gene flow out from central Cologne, which was significantly higher compared to the gene flow originating from the surrounding areas ($P < .001$), illustrating the potential role of central Cologne as a geographical hotspot in the spread of the local epidemic (Figure 4B). To evaluate geospatial mixing between genetically linked individuals (i.e., putative transmission pairs), we also computed assortativity coefficients based on individual’s location of residence. Individuals were more likely to be genetically linked to a partner living in a different zip code (assortativity coefficient for concurrent area of residency $A_{geo} = -.0026, P = .046$) (Figure 4C).

DISCUSSION

In this study, we explored the dynamics of HIV transmission in therapy-naive HIV-1 infected patients in the metropolitan Cologne-Bonn in Germany, a region with one of the highest rates of new HIV diagnoses in Western Europe [2, 3].

We evaluated the spatial connectivity of the local transmission network and the contribution of geospatial clusters and identified an overall median distance of 23.4 kilometers between linked individuals and a median diffusion/spread of 23.0 kilometers at the cluster level. Interestingly, we also found that individuals were less likely to be linked with individuals from the same neighborhood (same zip-code area). These findings may suggest that individuals were more engaged in sexual contact close to their home community but at the same time not within their immediate vicinity, or that individuals took greater risks with individuals outside their home community. These results are consistent with a previous report suggesting that safer sex behavior was more prevalent among MSM living in the same neighborhood [31]. The widespread use of geosocial-networking apps among MSM may contribute to interconnectivity and influence HIV transmission in more geographically distant individuals [8]. Although further work needs to be performed to understand this phenomenon, we hypothesize that geosocial networking apps and venues that facilitate contacts between individuals in different social circles (i.e., outside the neighborhood) may lead to this pattern of HIV transmission [32, 33]. Importantly, our migration analyses demonstrate an ongoing gene flow originating from the city center of Cologne, well known for its large MSM community and MSM clubs, toward the neighboring areas. Thus, the Cologne city center may be playing a prominent role in driving transmission of HIV across the region. Due to an extensive public transportation network (i.e., bus, train, and tram routes), the city center of Cologne and its surrounding areas are highly interconnected, which may contribute to the HIV gene flow originating from the city center of Cologne to its surrounding areas.

Although previous findings from the United States [34] and South Africa [35] indicate that HIV transmission was identified along main trucking routes, association between HIV transmission and human migration has to date not been described for Germany. In fact, Cologne is one of the key hubs for national and international transportation routes, connecting north-south and east-west, but it remains speculative if trucking routes may be driving interregional HIV transmissions in Germany.

Based on data of the European Centre for Disease Prevention and Control (ECDC), the overall HIV prevalence among sex workers in Germany is <1%, but the prevalence is considerably higher among male sex workers (20%) compared to female sex workers (0.2%) [36]. Although the data from our cohort did not collect information about sex workers, there is a need to continue monitoring of HIV prevalence among all risk groups and to elucidate if other risk groups such as men sex workers may drive the local HIV epidemic in the Cologne-Bonn region.

There are several limitations of our study. Most importantly, our data set is limited to individuals who were diagnosed and had sequences obtained before ART initiation at the University hospital of Cologne or Bonn between 2001 and 2016. However, based on the surveillance report of the Robert-Koch-Institute between 2001 and 2016, an estimated number of 2401 persons were newly diagnosed with HIV in Cologne and Bonn [37]. Our
sample of 714 ART naive patients represented approximately 30% of overall newly diagnosed patient from the Cologne-Bonn region during this time. While not reaching the 50% sampling for phylogenetic studies suggested by Novitsky et al [38], our 30% sampling is comparable to other similar studies [26, 39]. The substantial mixing of MSM and HTS clusters may also be due to missing individuals in our sampled population or due to an incorrect self-reported risk behavior or multiple risk factors [40]. Also, more detailed information on sex (ie, transsexuals) or other risk behaviors (eg, bisexual, having sex with male/ female sex workers) was not collected in our data set, consequently their role in the transmission network could not have been determined.

CONCLUSION

Using this viral sequence data, we inferred significant viral gene flow from the city center of Cologne to the surrounding areas, highlighting the role of this area in the transmission dynamics of the regional epidemic. We also found that transmission links were more likely between individuals residing closer together but less likely if they lived in the same neighborhood. Our findings highlight the importance of phylogenetic analysis to get a better understanding of local transmission patterns. Knowledge of these patterns can improve the efficiency of prevention and intervention efforts and can help public health entities to allocate resources and to ensure that strategies are optimal for priority risk groups and areas of greatest need.

Supplementary Data

Supplementary materials are available at Clinical Infectious Diseases online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyrighted and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Authors contributions. A. C., M. S., M. H., S. M., and J. V. designed the study; A. C., M. S., M. H., and S. M. analyzed and interpreted the data. J. V., M. S., A. E., C. L., G. F., J. W., and E. K. provided the data and contributed critically important ideas on how to interpret the data. M. S., A. C., and M. H. drafted the primary draft of the manuscript. S. M., J. V., A. E., C. L., G. F., J. W., and E. K. revised the manuscript critically for important intellectual content. All authors revised and approved the final version of the manuscript.

Availability of data and materials. The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate. The study was approved by the local Ethics Committees of the University Hospitals of Bonn (reference number 279-14) and Cologne (reference number 13-364) and all individuals gave written informed consent.

Funding. This work was supported by the German Center for Infection Research (DZIF) [NCT02149004] and supported by the National Institutes of Health grants [R25 MH081482, P30 AI036214, P50DA026306, MH113477, AI106039, and MH100974].

Potential conflicts of interest. J. J. V. reports personal fees from Merck / MSD, Gilead, Pfizer, Astellas Pharma, Basilea, Deutsches Zentrum für Infektionsforschung, Uniklinik Freiburg / Kongress und Kommunikation, Akademie für Infektionsmedizin, Universität Manchester, Deutsche Gesellschaft für Infektiologie, Ärztekammer Nordrhein, Uniklinik Aachen, Back Bay Strategies, Deutsche Gesellschaft für Innere Medizin and grants from Merck / MSD, Gilead, Pfizer, Astellas Pharma, Basilea, Deutsches Zentrum für Infektionsforschung, Bundesministerium für Bildung und Forschung. G. F. reports grants from German Center of Infection Research (DZIF), during the conduct of the study; personal fees from MSD, personal fees from Janssen, personal fees from Gilead, grants from Janssen, grants from AbbVie, outside the submitted work. C. L. reports grants from German Center of Infection Research (DZIF), personal fees from ViViV, personal fees from Gilead, and personal fees from MSD, outside the submitted work. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References


