PHYLOGENETIC CLUSTER ANALYSIS IDENTIFIES VIROLOGICAL AND BEHAVIORAL DRIVERS OF HIV TRANSMISSION IN MSM

Nadine Bachmann[1,2], Katharina Kusejko[1,2], Huyen Nguyen[1,2], Sandra E. Chaudron[1,2], Claus Kadelka[1,2], Teja Turk[1,2], Jürg Böni[2], Matthieu Perreau[3], Thomas Klimkait[4], Sabine Yerly[5], Manuel Battegay[6], Andri Rauch[7], Alban Ramette[7], Pietro Vernazza[8], Enos Bernasconi[9], Matthias Cavassini[10], Huldrych F. Günthard[1,2,*], Roger D. Kouyos[1,2,*] and *the Swiss HIV Cohort Study*

**Authors contributed equally*

[1] Department of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Zurich, Switzerland.

[2] University of Zurich, Institute of Medical Virology, Swiss National Reference Center for Retroviruses, Zurich, Switzerland.

[3] Division of Immunology and Allergy, Centre Hospitalier Universitaire Vaudois, University of Lausanne, Lausanne, Switzerland.

[4] Molecular Virology, Department Biomedicine - Petersplatz, University of Basel, Basel, Switzerland.

[5] Laboratory of Virology, Geneva University Hospital, Geneva, Switzerland.

[6] Department of Infectious Diseases and Hospital Epidemiology, University Hospital Basel and University of Basel, Basel, Switzerland.

[7] Institute for Infectious Diseases, Bern University Hospital and University of Bern, Bern, Switzerland.

[8] Division of Infectious Diseases and Hospital Epidemiology, Cantonal Hospital of St. Gallen, St. Gallen, Switzerland.

[9] Division of Infectious Diseases, Regional Hospital Lugano, Lugano, Switzerland.

[10] Division of Infectious Diseases, Centre Hospitalier Universitaire Vaudois, University of Lausanne, Lausanne, Switzerland.

To whom correspondence should be addressed: Roger D. Kouyos, Department of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Rämistrasse 100, 8091 Zürich; roger.kouyos@usz.ch

Summary: By developing a quantitative integrative approach combining phylogenetic, virological, behavioral, and STI data, we identified changing drivers of ongoing HIV transmission among MSM enrolled in the SHCS over the last decade indicating an increasing importance of undiagnosed individuals.

ABSTRACT

Background. Identifying local outbreaks and their drivers is a key step towards curbing HIV transmission and potentially achieving HIV elimination. Such outbreaks can be identified as transmission clusters extracted from phylogenetic trees constructed of densely sampled viral sequences. In this study, we combined phylogenetic transmission clusters with extensive data on virological suppression and behavioral risk of cluster members to quantify the drivers of ongoing transmission over ten years.

Methods. Using the comprehensive Swiss HIV Cohort Study and its drug-resistance database, we reconstructed phylogenetic trees for each year between 2007-2017. We identified HIV transmission clusters dominated by men who have sex with men(MSM) and determined their annual growth. We used Poisson regression to assess if cluster-growth was associated with a per-cluster-infectivity and behavioral-risk score.

Results. Both infectivity and behavioral risk scores were significantly higher in growing MSM transmission clusters compared to non-growing clusters ($p \leq 0.01$). The fraction of transmission clusters without infectious members acquiring new infections increased significantly over the study period. The infectivity score was significantly associated with per-capita incidence of MSM transmission clusters in eight years, while the behavioral risk score was significantly associated with per-capita incidence of MSM transmission clusters in three years.

Conclusions. We present a phylogenetic method to identify hotspots of ongoing transmission among MSM. Our results demonstrate the effectiveness of treatment as prevention at the population level. However, the significantly increasing number of new infections among transmission clusters without infectious members highlight a relative shift from diagnosed to undiagnosed individuals as drivers of HIV transmission in Swiss MSM.

Keywords: Phylogenetics, HIV transmission clusters, MSM, drivers of transmission

INTRODUCTION

The global epidemic of human immunodeficiency virus (HIV) is still not under control with approximately 1.7 million new HIV infections in 2018 [1]. As the HIV community embraces the UNAIDS 90-90-90 targets for HIV diagnosis, treatment and viral suppression, the focus has largely been on increasing access to antiretroviral treatment (ART) [2]. However, reducing HIV incidence remains a mains challenge in HIV medicine. Thus, optimizing prevention using state-of-the-art methodology and allocating resources for prevention to subpopulations at risk are crucial to curbing the epidemic.

Molecular epidemiology is essential for the improvement of prevention of new HIV infections [3,4]. The possibilities of molecular epidemiology have advanced with increasing computational power and availability of densely sampled HIV sequencing data from genotypic resistance testing (GRT). Representative HIV phylogenies can be reconstructed to subsequently extract phylogenetic clusters. Epidemiologically, these phylogenetic clusters reflect transmission chains that are or were ongoing and can identify local outbreaks within epidemics. Subgroups within the epidemic with a higher risk of leading to local outbreaks may be identified by specific clinical, demographic, or behavioral factors [5–9]. This study presents a new longitudinal method of identifying HIV clusters at risk of transmission, combining molecular epidemiology tools with classical clinical and behavioral risk factors. Our method was applied to men who have sex with men (MSM) enrolled in the Swiss HIV Cohort Study (SHCS) as a case study.

Within the SHCS it has been shown that more than half of Swiss HIV infected individuals were included in transmission clusters and that most of these individuals were infected via the same transmission route [10]. MSM having unprotected anal sex represent the major risk group for ongoing transmission of HIV in Western Europe and North America [11]. Based on numbers of newly diagnosed cases over the last decade, the epidemic in MSM does not seem to be under control [12]. Recent studies in this community indicate reduced condom use in some subgroups of MSM [9,13,14]. Additionally, the numbers of HIV-positive MSM infected with sexually transmitted infections (STI), such as Syphilis, Chlamydia, and Hepatitis C, have been on the rise in the past decade [15–19].

In 2011 global HIV-prevention efforts progressed considerably when the HPTN 052 trial demonstrated that ART reduces the sexual transmission of HIV in HIV-serodiscordant couples by more than 96% [20]. In MSM the risk of HIV transmission through condom-less sex when the HIV viral load is suppressed was recently shown to be effectively zero (upper 95% confidence interval of 0.23 per 100 couple-years of follow-up) [21] thereby confirming the concept of treatment as prevention (TasP) on the individual level for MSM. The impact of TasP at the population level is less clear. The results of the HPTN 071 (PopART) trial revealed that, initiating ART immediately after diagnosis did not significantly reduce

incidence [22]. In Switzerland, TasP was gradually introduced after 2008 based on the Swiss statement [23]. What still remains controversial about TasP is its impact on risk behavior of HIV-positive individuals potentially compensating effects of TasP by increasing risk behavior [16]. Utilizing the comprehensive database of the SHCS we sought to better understand impacts of TasP and the potentially increased risk behavior on HIV transmission in MSM at both individual and population levels.

METHODS

The Swiss HIV Cohort Study

The SHCS is a large, prospective, multi-center study established in 1988 [24]. At each semi-annual follow-up visit clinical and laboratory data of three-monthly visits are obtained from SHCS participants. The SHCS has an estimated coverage of approximately 75% of all HIV-positive individuals on ART [24] in Switzerland. It includes a drug resistance database that contains HIV sequences for approximately 60% of the over 20,000 SHCS participants and 75% of all participants enrolled after 1995, because over 11,000 genotypes were retrospectively sequenced from the SHCS biobank [25,26]. Written informed consent was obtained from each SHCS study participant.

Phylogenetic Tree Construction

For each of the years from 2007 to 2017, we considered all partial *pol* sequences of SHCS participants, which were available prior to the beginning of the respective year and had adequate quality (at least 250 base pairs from protease and at least 500 base pairs from reverse transcriptase). If multiple sequences per patient were available, the first one was considered. Importantly, all sequences included in a certain year were also included in subsequent years. For each year, we blasted non-Swiss Los Alamos background sequences against the described choice of SHCS sequences (targeting the 10 closest sequences with a threshold of 90%). SHCS and Los Alamos sequences were pooled and pairwise aligned to the HXB2 reference genome using MUSCLE [27]. Insertions relative to HXB2 and resistance mutations according to the Stanford (http://hivdb.stanford.edu/) and International Antiviral Society-USA (https://www.iasusa.org/) lists were removed, and positions with many gaps were eliminated by trimAl [28]. Maximum likelihood phylogenetic trees were reconstructed with FastTree using a generalized time-reversible model [29]. We obtained eleven phylogenetic trees reflecting the observed transmission network within the SHCS in the beginning of each year from 2007 to 2017 *(Figure 1)*.

Extraction of MSM transmission clusters

We identified transmission clusters using phylogenetic tree topology including Los Alamos background sequences to split individual groups who were not closely related in the transmission chain into phylogenetic clades consisting of at least 80% SHCS sequences. As we aimed to study MSM transmission clusters, we discarded transmission clusters if they did not consist of at least 50% Swiss MSM. This definition allowed to also capture transmission clusters that did not exclusively contain MSM, which is especially of important considering MSM who did not disclose as MSM and women infected by bisexual men. We did not consider newly emerging clusters since we were interested in the predictive power of the clusters that already existed the year before. Therefore, we defined "preexistent" clusters as clusters that existed in the previous year containing at least 50% of the already infected members appearing in MSM transmission clusters *(Figure 1)*. Transmission clusters were extracted using custom scripts incorporating the R package APE [30].

Identification of newly infected MSM

For each study year, we identified all MSM in the SHCS who were newly infected in that year and reported to have been infected in Switzerland (thus for the new infections considered, we excluded MSM reporting to have been infected abroad. This information is only collected in the SHCS since 2007, thus, we did not consider earlier years). The time period spanned the first to the last day of the year (i.e., newly infected MSM in 2016 were infected between 2016-01-01 and 2016-12-31). MSM were considered newly infected in a certain year, if they had both (i) their first GRT done within the year and (ii) their estimated date of infection (EDI) was within the year (MSM with first GRT and EDI in different years were discarded as newly infected).

We used the following two methods to estimate HIV infection dates as accurately as possible:

(1) Defined HIV primary infection: For individuals with available seroconversion dates (negative and positive HIV tests less than 1 year apart) or with a diagnosis of primary infection as previously described [31], we used the midpoint between the dates of the negative and the positive test, or, if known, the date of the primary infection as best estimate of the HIV infection date.

(2) Defined recent HIV infection: For individuals who (i) exhibited low HIV diversity (less than 0.5% of ambiguous nucleotides) in a genotypic HIV drug resistance test within the first year after diagnosis

and (ii) had >200 CD4+ cells/µl blood at cohort registration [32–34], we used the diagnosis date as best estimate of the HIV infection date.

Individuals for whom neither (1) nor (2) was possible to estimate the date of infection, were viewed as not newly infected within that given year and thus excluded from analysis (i.e. newly diagnosed chronic infections were excluded from our analysis).

Definition of explanatory variables

For each MSM transmission cluster and each considered calendar year, we calculated the following explanatory variables using all individuals within the transmission clusters (not only those who disclosed as MSM).

Annual infectivity score: For each MSM transmission cluster, the infectivity score is defined as the fraction of cluster members who had at least one viral load greater than 1,000 copies in the year of interest. Laboratory values such as the viral load are usually measured during every quarterly to half yearly follow-up visit for individuals participating in the SHCS.

Annual behavioral risk score: For each MSM transmission cluster, the behavioral risk score is defined as the fraction of cluster members who either reported risky behavior and/or had a coinfection with Hepatitis B, C or Syphilis (*Figure 3d*) as a surrogate for risky behavior. Risky behavior was defined as reported condom-less anal or vaginal intercourse with occasional partners or refusal to answer this question (on median 3 participants per year refused an answer, IQR=[2,5]). A Hepatitis B coinfection was present if both the qualitative result of an Anti-HBc-test and the qualitative result of an HBsAg-test were positive. A Hepatitis C coinfection was present if either the qualitative result of an Anti-HCV-test was positive or Hepatitis C RNA viral load from a quantitative test was detectable. A Syphilis coinfection was present if the qualitative result of the screening test and the interpretation of the quantitative result were positive. Coinfections were primarily included as proxies for risk behavior [19] and not because we assumed that they biologically facilitate HIV-transmission.

Statistical Analysis

Regression analysis: In order to evaluate the infectivity and behavioral risk score as predictors for new infections, we performed multivariable Poisson regression analyses using behavioral risk score and infectivity score as predictor variables. For each study year, we identified pre-existing clusters from the beginning of that year and the beginning of the subsequent year and quantified the number of

newly infected MSM within each cluster. We then used the number of new infections within each cluster as an outcome variable and included infectivity and behavioral risk score of each cluster from the previous year as predictor variables. By separating predictor and outcome variables by time we aimed at establishing a causal relationship (i.e. to test that the former cause the latter). Further, we included the logarithm of the number of cluster members as an offset to exclude the risk of observing an effect driven by the higher chance of acquisition of newly infected MSM in larger MSM transmission clusters. Thus, we model the numbers of new infections per cluster as

$$N_{new} \sim Poiss\big(\lambda(N, RiskScore, InfectivityScore)\big)$$

Where

$$\lambda(N, RiskScore, InfectivityScore) = e^{\log(N)+a_0+a_1\mathrm{RiskScore}+a_1\mathrm{InfectivityScore}} = Nr(RiskScore, InfectivityScore)$$

and thus

$$r(RiskScore, InfectivityScore) = e^{a_0+a_1\mathrm{RiskScore}+a_1\mathrm{InfectivityScore}}$$

is the per capita growth rate of the clusters. Accordingly, the exponential of the coefficients derived from the poisson regression analysis can be interpreted as the effect from the risk and the infectivity scores on the per capita cluster growth rate of the considered transmission clusters. In other words, these coefficients can be interpreted as the per capita incidence risk ratios.

P-values for differences and time trends: P-values for time trends over our study period were derived using a linear regression analysis. P-values for differences were derived using the paired Wilcoxon rank sum test.

RESULTS

Study population of MSM within MSM transmission clusters

Table 1 displays characteristics of MSM within MSM transmission clusters and of all non-clustered MSM on the respective phylogeny for the first and last study year (*Table S1 for all study years).* Of the individuals in the MSM transmission clusters at least 77.9% were MSM, while up to 9.5% of individuals were males reporting infection through heterosexual contacts (*Figure S1*).

Incidence of newly infected MSM within MSM transmission cluster during 10-year study period

The annual number of newly infected MSM, identified by incorporating estimated infection dates and excluding new diagnosis of chronic infections *(see Methods),* decreased significantly over our study period *(p_{trend} < 0.001, Figure 2a)*. Between 30.61% (2009) and 57.14% (2014) of the newly infected MSM appeared within preexistent MSM transmission clusters *(p_{trend} = 0.802, Figure 2a)*. The majority of newly infected MSM not in preexisting MSM transmission clusters appeared in newly emerged clusters containing only two individuals including the newly infected MSM (*Figure S2).*

Characteristics of MSM transmission clusters acquiring newly infected MSM

The mean annual infectivity score for MSM transmission clusters significantly decreased over the study period *(Figure 3a)*. This paralleled the observation that the median time between HIV diagnosis to treatment initiation decreased from a maximum of 22 weeks in 2008 to a minimum of 3 weeks in 2016 (*Figure S3*), indicating a decreasing fraction of infectious SHCS participants over our study period. In contrast, the mean annual behavioral risk score for MSM transmission clusters strongly increased over the study period *(Figure 3b).* Both the infectivity and the behavioral risk score were significantly higher in MSM transmission clusters that acquired newly infected MSM, compared to MSM transmission clusters not acquiring newly infected MSM *($p_{difference}$ < 0.001 each, Figure 3a&b)*. The fraction of transmission clusters acquiring new infections, even though not containing any infectious members (i.e., infectivity score of 0), increased significantly over the study period *(p_{trend} = 0.017, Figure 3c)*, reaching levels of up to 50% by 2016.

Per capita cluster growth was determined by annual infectivity and behavioral risk scores

Multivariable Poisson regression analysis revealed that the annual infectivity score was significantly and independently associated with per capita cluster growth of MSM transmission clusters in 8 out of 10 study years (2007-2011, 2013-2014, 2016) (*Figure 4)*. At the same time, the annual behavioral risk score was significantly associated with the per capita MSM transmission clusters growth in 3 out of 10 study years (2011-2012, 2014). Combining the two scores did not improve the observed associations (*Table S2*). These associations highlight the potential of identifying transmission clusters at high risk of transmission using these already available variables. At the same time, our power to predict cluster growth in the subsequent year using the previous regression analysis decreased strongly over our study period (p_{trend} = 0.0002, *Figure S4*).

DISCUSSION

This is the first study quantifying the risk of ongoing transmission in phylogenetically identified clusters by combining data on virological suppression with behavioral risk thereby identifying the changing drivers of the HIV epidemic. Based on the comprehensive SHCS, this analysis was performed for over a decade and thus provides a long-term perspective on challenges underlying HIV elimination. The decreasing infectivity score along with the decreasing total number of newly infected MSM over the study period provide evidence for a beneficial effect of TasP at the population level. Despite the increase of the behavioral risk score in the last years of our study period HIV-infections still decreased, illustrating that TasP also works in the context of risky behavior. Over one decade, we found a significantly increasing fraction of transmission clusters which acquired new infections but did not contain any infectious members. In particular, in the last study years up to one half of MSM transmission clusters acquiring newly infected MSM contained only virally suppressed cluster members. Given the proven effectiveness of TasP at the individual level [20,21,35], this implies that there were unsampled or undiagnosed HIV-positive individuals within the transmission networks corresponding to the MSM transmission clusters we identified. This is additionally supported by our finding of a significantly decreasing power to predict cluster growth over our study period and illustrates the declining role of people living with diagnosed but unsuppressed HIV in driving the more recent epidemic. Furthermore, this highlights that in recent years a relative shift occurred from the diagnosed to the undiagnosed population as the driver of the HIV epidemic in Swiss MSM.

Our study has limitations and strengths. As any study in molecular epidemiology, we cannot achieve perfect sampling. Even though the SHCS is highly representative of the HIV epidemic in Swiss MSM, phylogenetically identified transmission clusters are likely to have unsampled members. Thus, the detected transmission clusters should be interpreted as proxies for the actual transmission networks. However, despite the incomplete sampling, our approach allows for relevant conclusions about the drivers of the epidemic. In particular, in the last study years we observed newly infected individuals in clusters where all members were virally suppressed, implying that there were unsampled/undiagnosed HIV-positive individuals within the transmission networks of the corresponding MSM transmission clusters. Further, the estimated dates of infection sometimes harbor large confidence intervals. Thus, we may not exclude the possibility of incorrectly categorizing individuals as “newly infected”. However, the quality of our estimated infection dates using defined HIV primary or recent infections was not time-dependent (data not shown) and misclassification of “newly infected” individuals was equally likely for individuals of clusters with varying characteristics

over the time period examined. These shortcomings are therefore unlikely to explain any of the observed outcomes. Further, there is no consensus or gold-standard for phylogenetically defining HIV transmission clusters. The frequently used cluster detection methods fall into three categories: distance-based, topology-based or a combination of the two. Suitability of the methods highly depends on the research question of interest. In our case, a distance-based approach was not suitable since we did not want to prioritize new infections within recently established compared to long-circulating transmission clusters – thus, we applied a purely topological approach. The strengths of this study include that the statistical analysis directly yields the impact of risk and infectivity scores on the observed per-capita growth rate of the cluster over a long-term perspective in a representative population.

We demonstrate a new method to track the epidemiology of HIV, thereby identifying hotspots of ongoing transmission among MSM. The identified high-risk clusters could be used to target social networks harboring undiagnosed individuals spreading HIV. In these social networks, intensification of prevention (e.g. testing campaigns and PrEP) is likely to have the largest impact. The significantly increasing number of new infections within transmission clusters without infectious members, indicates a relative shift of epidemiological importance from diagnosed to undiagnosed individuals as drivers of the epidemic. Further, this highlights the need for additional efforts to identify infectious individuals as early as possible and to link them into care. Our method, combining phylogenetics, clinical and behavioral data, sheds light on high risk social networks, where prevention and HIV-testing should be intensified and may thus inform the optimal allocation of resources for ending HIV transmission.

Acknowledgements

We thank the patients for participating in the SHCS, the study nurses and physicians for excellent patient care, A. Scherrer, A. Traytel, and S. Wild for excellent data management and D. Perraudin and M. Amstutz for administrative assistance.

The members of the SHCS are Anagnostopoulos A, Battegay M, Bernasconi E, Böni J, Braun DL, Bucher HC, Calmy A, Cavassini M, Ciuffi A, Dollenmaier G, Egger M, Elzi L, Fehr J, Fellay J, Furrer H, Fux CA, Günthard HF (President of the SHCS), Haerry D (deputy of "Positive Council"), Hasse B, Hirsch HH, Hoffmann M, Hösli I, Huber M, Kahlert CR (Chairman of the Mother & Child Substudy), Kaiser L, Keiser O, Klimkait T, Kouyos RD, Kovari H, Ledergerber B, Martinetti G, Martinez de Tejada B, Marzolini C, Metzner KJ, Müller N, Nicca D, Paioni P, Pantaleo G, Perreau M, Rauch A (Chairman of the Scientific Board), Rudin C, Scherrer AU (Head of Data Centre), Schmid P, Speck R, Stöckle M (Chairman of the Clinical and Laboratory Committee), Tarr P, Trkola A, Vernazza P, Wandeler G, Weber R, Yerly S.

Funding

This work was funded within the framework of the Swiss HIV Cohort Study (SNF grant# 33CS30_177499 to H.F.G). The data were gathered by the five Swiss University Hospitals, two Cantonal Hospitals, 15 affiliated hospitals and 36 private physicians listed in http://www.shcs.ch/180-health-care-providers. The work was furthermore supported by Swiss National Science Foundation grant 324730B_179571 (to H.F.G), SNF grant SNF 310030_141067 (to H.F.G), SNF grants no. PZ00P3-142411 and BSSGI0_155851 to R.D.K., the Yvonne-Jacob Foundation (to H.F.G.), the University of Zurich's Clinical Research Priority Program viral infectious disease, ZPHI (to H.F.G) and an unrestricted research grant from Gilead Sciences to the SHCS research foundation.

Competing interests

M.C. has received research and travel grants for his institution from ViiV and Gilead.

E.B. has received fees for his institution for participation to advisory board from MSD, Gilead Sciences, ViiV Healthcare, Pfizer, Sandoz, and Abbvie.

M.B. has received research or educational grants by Abbvie AG, Gilead Sciences Switzerland Sàrl, Janssen-Cilag AG, MSD Merck Sharp & Dohme AG and ViiV Healthcare GmbH.

The institution of P.V. has received advisory fees for participation of team members in advisory boards from MSD, ViiV Healthcare GmbH, TEVA, and Gilead-Sciences.

T.K. has received honoraria from Gilead Sciences and Roche Diagnostics.

A.Rauch has received honoraria for advisory boards and/or travel grants: MSD, Gilead Sciences, Abbvie, Bristol-Myers Squib, Janssen-Cilag, and Pfizer, and an investigator-initiated trial (IIT) grant from Gilead Sciences. All remuneration went to his home institution and not to AR personally, outside the submitted work.

R.D.K. has received grants from the Swiss National Science Foundation and personal fees from Gilead Sciences, outside the submitted work.

H.F.G. has received unrestricted research grants from Gilead Sciences and Roche; fees for data and safety monitoring board membership from Merck; consulting/advisory board membership fees from Gilead Sciences, ViiV, Sandoz and Mepha.

All other authors have no competing interests.

References

1. UNAIDS. UNAIDS data 2019. 2019. Available at: https://www.unaids.org/sites/default/files/media_asset/2019-UNAIDS-data_en.pdf. Accessed 2 September 2019.

2. Baggaley R, Dalal S, Johnson C, et al. WHO | Beyond the 90-90-90: refocusing HIV prevention as part of the global HIV response. WHO **2018**; Available at: https://www.who.int/hiv/pub/journal_articles/article-beyond-90-90-90/en/. Accessed 31 October 2018.

3. Oster AM, France AM, Mermin J. Molecular Epidemiology and the Transformation of HIV PreventionMolecular Epidemiology and the Transformation of HIV PreventionMolecular Epidemiology and the Transformation of HIV Prevention. JAMA **2018**; 319:1657–1658. Available at: https://doi.org/10.1001/jama.2018.1513.

4. Poon AFY, Gustafson R, Daly P, et al. Near real-time monitoring of HIV transmission hotspots from routine HIV genotyping: an implementation case study. The Lancet HIV **2016**; 3:e231–e238. Available at: https://doi.org/10.1016/S2352-3018(16)00046-1.

5. Grabowski MK, Herbeck JT, Poon AFY. Genetic Cluster Analysis for HIV Prevention. Current HIV/AIDS reports **2018**; 15:182–189. Available at: https://www.ncbi.nlm.nih.gov/pubmed/29460226.

6. Mishra S, Pickles M, Blanchard JF, Moses S, Shubber Z, Boily M-C. Validation of the modes of transmission model as a tool to prioritize HIV prevention targets: a comparative modelling analysis. PloS one **2014**; 9:e101690–e101690. Available at: https://www.ncbi.nlm.nih.gov/pubmed/25014543.

7. Tanser F, de Oliveira T, Maheu-Giroux M, Bärnighausen T. Concentrated HIV subepidemics in generalized epidemic settings. Current opinion in HIV and AIDS **2014**; 9:115–125. Available at: https://www.ncbi.nlm.nih.gov/pubmed/24356328.

8. Frost SDW, Pillay D. Understanding drivers of phylogenetic clustering in molecular epidemiological studies of HIV. The Journal of infectious diseases. 2015; 211:856–858.

9. Salazar-Vizcaya L, Kusejko K, Schmidt AJ, et al. Clusters of sexual behaviour in HIV-positive men who have sex with men reveal highly dissimilar time trends. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2019**;

10. Kouyos RD, von Wyl V, Yerly S, et al. Molecular Epidemiology Reveals Long-Term Changes in HIV Type 1 Subtype B Transmission in Switzerland. The Journal of Infectious Diseases **2010**; 201:1488–1497. Available at: http://dx.doi.org/10.1086/651951.

11. O'Leary D. The syndemic of AIDS and STDS among MSM. The Linacre quarterly **2014**; 81:12–37. Available at: https://www.ncbi.nlm.nih.gov/pubmed/24899736.

12. Beyrer C, Baral SD, van Griensven F, et al. Global epidemiology of HIV infection in men who have sex with men. Lancet (London, England) **2012**; 380:367–377. Available at: https://www.ncbi.nlm.nih.gov/pubmed/22819660.

13. Kouyos RD, Hasse B, Calmy A, et al. Increases in Condomless Sex in the Swiss HIV Cohort Study. Open Forum Infectious Diseases **2015**; 2:ofv077. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4498263/.

14. Paz-Bailey G, Mendoza MCB, Finlayson T, et al. Trends in condom use among MSM in the United States: the role of antiretroviral therapy and seroadaptive strategies. AIDS (London,

England) **2016**; 30:1985–1990.

15. Mohammed H, Mitchell H, Sile B, Duffell S, Nardone A, Hughes G. Increase in Sexually Transmitted Infections among Men Who Have Sex with Men, England, 2014. Emerging infectious diseases **2016**; 22:88–91. Available at: https://www.ncbi.nlm.nih.gov/pubmed/26689861.

16. Shilaih M, Marzel A, Braun DL, et al. Factors associated with syphilis incidence in the HIV-infected in the era of highly active antiretrovirals. Medicine **2017**; 96:e5849–e5849. Available at: https://www.ncbi.nlm.nih.gov/pubmed/28079818.

17. Shilaih M, Marzel A, Scherrer AU, et al. Dually Active HIV/HBV Antiretrovirals as Protection Against Incident Hepatitis B Infections: Potential for Prophylaxis. The Journal of infectious diseases **2016**; 214:599–606.

18. Wandeler G, Gsponer T, Bregenzer A, et al. Hepatitis C virus infections in the Swiss HIV Cohort Study: a rapidly evolving epidemic. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2012**; 55:1408–1416.

19. Braun DL, Marzel A, Steffens D, et al. High Rates of Subsequent Asymptomatic Sexually Transmitted Infections and Risky Sexual Behavior in Patients Initially Presenting With Primary Human Immunodeficiency Virus-1 Infection. Clinical Infectious Diseases **2018**; 66:735–742. Available at: https://doi.org/10.1093/cid/cix873.

20. Cohen MS, Chen YQ, McCauley M, et al. Prevention of HIV-1 Infection with Early Antiretroviral Therapy. New England Journal of Medicine **2011**; 365:493–505. Available at: https://doi.org/10.1056/NEJMoa1105243.

21. Rodger AJ, Cambiano V, Bruun T, et al. Risk of HIV transmission through condomless sex in serodifferent gay couples with the HIV-positive partner taking suppressive antiretroviral therapy (PARTNER): final results of a multicentre, prospective, observational study. The Lancet **2019**; Available at: https://doi.org/10.1016/S0140-6736(19)30418-0.

22. Hayes RJ, Donnell DJ, Floyd S, et al. IMPACT OF UNIVERSAL TESTING AND TREATMENT IN ZAMBIA AND SOUTH AFRICA: HPTN071(POPART) | CROI Conference. 2019. Available at: http://www.croiconference.org/sessions/impact-universal-testing-and-treatment-zambia-and-south-africa-hptn071popart. Accessed 22 April 2019.

23. Vernazza P, Hirschel B, Bernasconi E, Flepp M. HIV-positive individuals not suffering from a ny other STD and adheri ng to an effective antiretroviral treatment do not transmit HIV sexually. Bulletin des médecins suisses 89 (5) **2008**; Available at: http://i-base.info/qa/wp-content/uploads/2008/02/Swiss-Commission-statement_May-2008_translation-EN.pdf. Accessed 2 May 2018.

24. Schoeni-Affolter F, Ledergerber B, Rickenbach M, et al. Cohort profile: the Swiss HIV Cohort study. International journal of epidemiology **2010**; 39:1179–1189.

25. Yang W-L, Kouyos R, Scherrer AU, et al. Assessing the Paradox Between Transmitted and Acquired HIV Type 1 Drug Resistance Mutations in the Swiss HIV Cohort Study From 1998 to 2012. The Journal of infectious diseases **2015**; 212:28–38.

26. Yang W-L, Kouyos RD, Böni J, et al. Persistence of Transmitted HIV-1 Drug Resistance Mutations Associated with Fitness Costs and Viral Genetic Backgrounds. PLOS Pathogens **2015**; 11:e1004722. Available at: https://doi.org/10.1371/journal.ppat.1004722.

27. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **2004**; 32:1792–1797. Available at:

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC390337/.

28. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics (Oxford, England) **2009**; 25:1972–1973.

29. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLOS ONE **2010**; 5:e9490. Available at: https://doi.org/10.1371/journal.pone.0009490.

30. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics **2004**; 20:289–290. Available at: http://dx.doi.org/10.1093/bioinformatics/btg412.

31. Rieder P, Joos B, von Wyl V, et al. HIV-1 transmission after cessation of early antiretroviral therapy among men having sex with men. AIDS (London, England) **2010**; 24:1177–1183.

32. Kouyos RD, von Wyl V, Yerly S, et al. Ambiguous nucleotide calls from population-based sequencing of HIV-1 are a marker for viral diversity and the age of infection. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2011**; 52:532–539.

33. Ragonnet-Cronin M, Aris-Brosou S, Joanisse I, et al. Genetic diversity as a marker for timing infection in HIV-infected patients: evaluation of a 6-month window and comparison with BED. The Journal of infectious diseases **2012**; 206:756–764.

34. Andersson E, Shao W, Bontell I, et al. Evaluation of sequence ambiguities of the HIV-1 pol gene as a method to identify recent HIV-1 infection in transmitted drug resistance surveys. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases **2013**; 18:125–131.

35. Rodger AJ, Cambiano V, Bruun T, et al. Sexual Activity Without Condoms and Risk of HIV Transmission in Serodifferent Couples When the HIV-Positive Partner Is Using Suppressive Antiretroviral Therapy. JAMA **2016**; 316:171–181.

year 2007	MSM in clusters	non-clustered MSM on tree	$p_{cluster/tree}$	MSM not on phylogeny	$p_{tree/SHCS}$
N	1767	1124	-	785	-
median age (IQR)	45 (40,52)	45 (40,54)	**0.091**	49 (43,58)	**<0.001**
white ethnicity (%)	95.59%	92.45%	**<0.001**	93%	0.51
median CD4 cell count at diagnosis (IQR)	367 (196.5,557)	345 (184,537)	0.065	290 (115,487)	**<0.001**
median year of SHCS registration (IQR)	1998 (1994,2003)	1998 (1994,2003)	0.5	1995 (1992,2000)	**<0.001**
year 2017					
N	3429	1544	-	1372	-
median age (IQR)	51 (44,58)	52 (43,60)	**0.056**	53 (45,63)	**0.003**
white ethnicity (%)	92.97%	86.69%	**<0.001**	88.63%	0.104
median CD4 cell count at diagnosis (IQR)	381.5 (222,561)	367.5 (200,546.5)	**0.03**	369.5 (194,577.25)	0.735
median year of SHCS registration (IQR)	2005 (1997,2010)	2004.5 (1996,2010)	0.888	2004.5 (1995,2012)	0.301

Table 1. Individuals' characteristics in the first (2007) and the last (2017) year of the study. *MSM transmission clusters were defined as phylogenetic transmission clusters consisting of at least 50% MSM. The first column "MSM in clusters" refers to MSM cluster members, the second column "non-clustered MSM on tree" refers to all MSM who were on the respective phylogeny (i.e. with an available GRT), while the column "MSM not on phylogeny" includes all MSM registered in the SHCS but not on the phylogeny (i.e., individuals without sequencing data) with follow-up data after the beginning of the year 1995, when the collection of plasma samples was initiated. P-values comparing characteristics of MSM in clusters to non-clustered MSM on the tree and all MSM included in the analysis to all MSM without GRT, respectively, were derived using paired Wilcoxon rank sum test.*

Figure legends

Figure 1. Methods overview. *For all pairs of two consecutive years (year x and year x+1, x being one of the years from 2007 to 2016) phylogenetic trees were built of sequences available prior to the respective years (green tips refer to MSM transmission clusters, blue tips refer to all other). We extracted MSM transmission clusters from these phylogenetic trees and checked which MSM transmission clusters were preexistent in the previous year with at least 50% of the already infected members in MSM transmission clusters. For all preexisting clusters, we calculated the number of newly infected MSM hat were acquired by the respective cluster within the year of interest. Further, we assigned a behavioral risk and an infectivity score to each preexistent cluster of year x. We performed a regression analysis to investigate the association of behavioral risk and infectivity score and the number of newly infected individuals with MSM transmission clusters (indicated by question mark).*

Figure 2. Newly infected MSM within MSM transmission clusters. (a) Number of newly infected MSM with acute or recent HIV-infection (as defined in the Methods section, solid green line) and fraction of all newly infected MSM who appear within preexisting phylogenetic MSM transmission clusters (as defined in the Methods section, dashed pink line). *P-values for time trends were derived using linear regression.* ***(b)*** *Example of 3 MSM transmission clusters acquiring newly infected MSM between 2007-01-01 and 2008-01-01. Olive green color represents individuals who were already in the cluster in the previous year, pink newly infected individuals within the cluster, black Los Alamos background sequences and light green other SHCS participants (SHCS participants who were not in the same cluster in the previous year).*

Figure 3. Characteristics of MSM transmission clusters acquiring newly infected individuals. *The x-axis denotes the study year (first year per tick: year of transmission clusters used to identify cluster characteristics; second year per tick: year of transmission clusters used to identify cluster growth). (a) Mean annual per-cluster infectivity scores of clusters acquiring newly infected individuals (squares) compared to clusters not acquiring newly infected individuals (circles). The fraction of cluster members who were infectious (defined as at least one viral load ≥ 1'000 copies/ml) within the study years displayed on the y-axis. (b) Mean annual per cluster behavioral risk scores of clusters acquiring newly infected individuals (squares) compared to clusters not acquiring newly infected individuals (circles). The fraction of cluster member who either reported risk-behavior and/or had a coinfection (Hepatitis B, Hepatitis C or Syphilis) within the study years displayed on the y-axis. (c) Infectivity and behavioral*

risk scores of zero. The fraction of clusters acquiring newly infected individuals even though harboring an infectivity (green) or behavioral risk score (pink) of zero is displayed on the y-axis. Total number of clusters identified at the beginning of the first year per tick ("clusters") and number of clusters acquiring newly infected MSM during the next year ("growing") are displayed. P-values for time trends were derived using linear regression. P-values for differences were derived using paired Wilcoxon rank sum test.

Figure 4. Association of cluster infectivity and cluster risk behavior with per capita cluster growth. *Results of a Poisson regression with per capita growth as outcome, and infectivity as well as behavioral risk score as dependent variables. The effect of **(a)** the infectivity score and **(b)** the behavioral risk score on the incidence rate ratio is depicted on the y-axis, while the x-axis shows the study years. Coefficients from univariable (green) and multivariable (purple) models are shown. The y-axes depict the incidence rate ratio.*

Figure 1

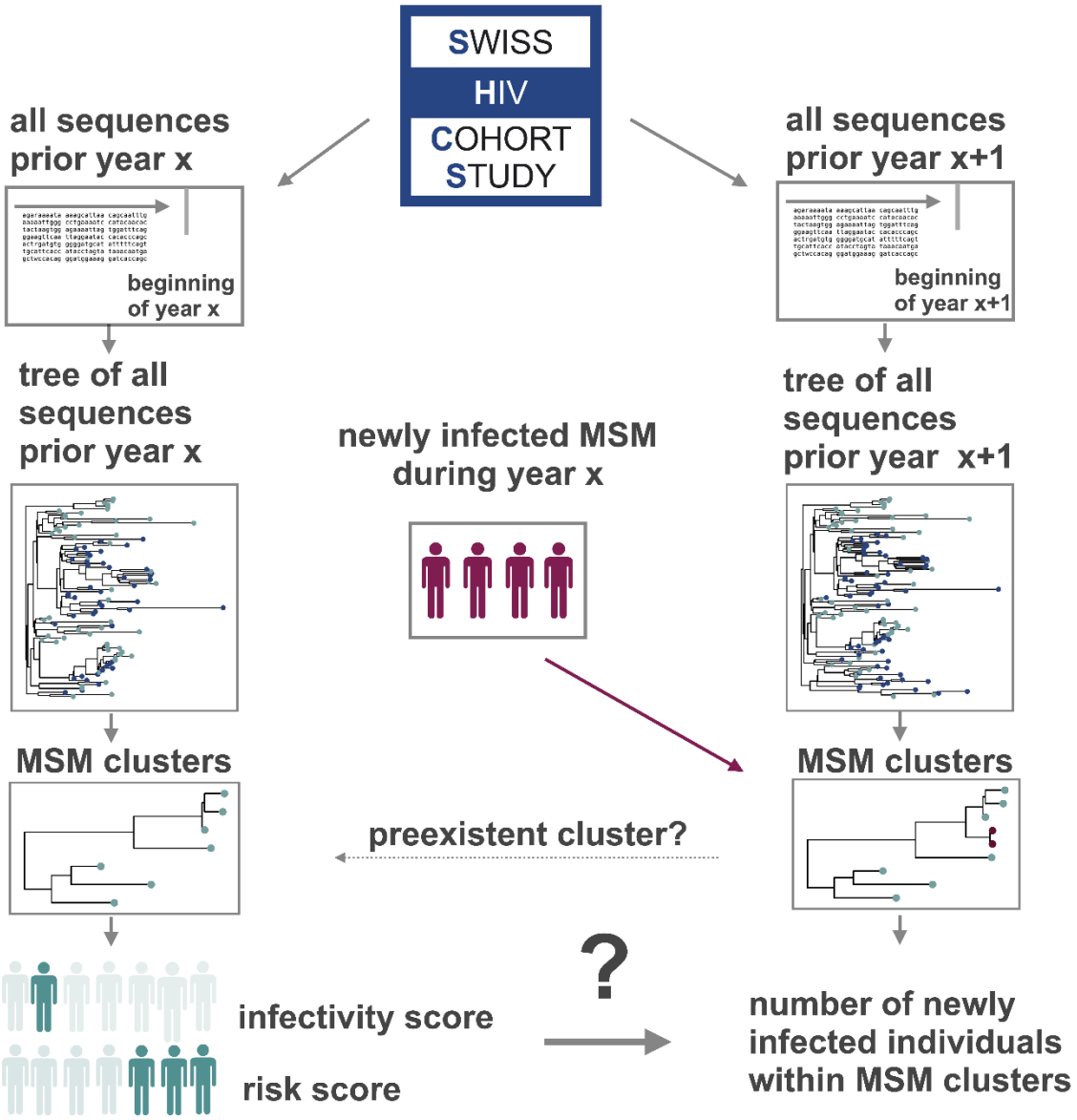

Figure 2

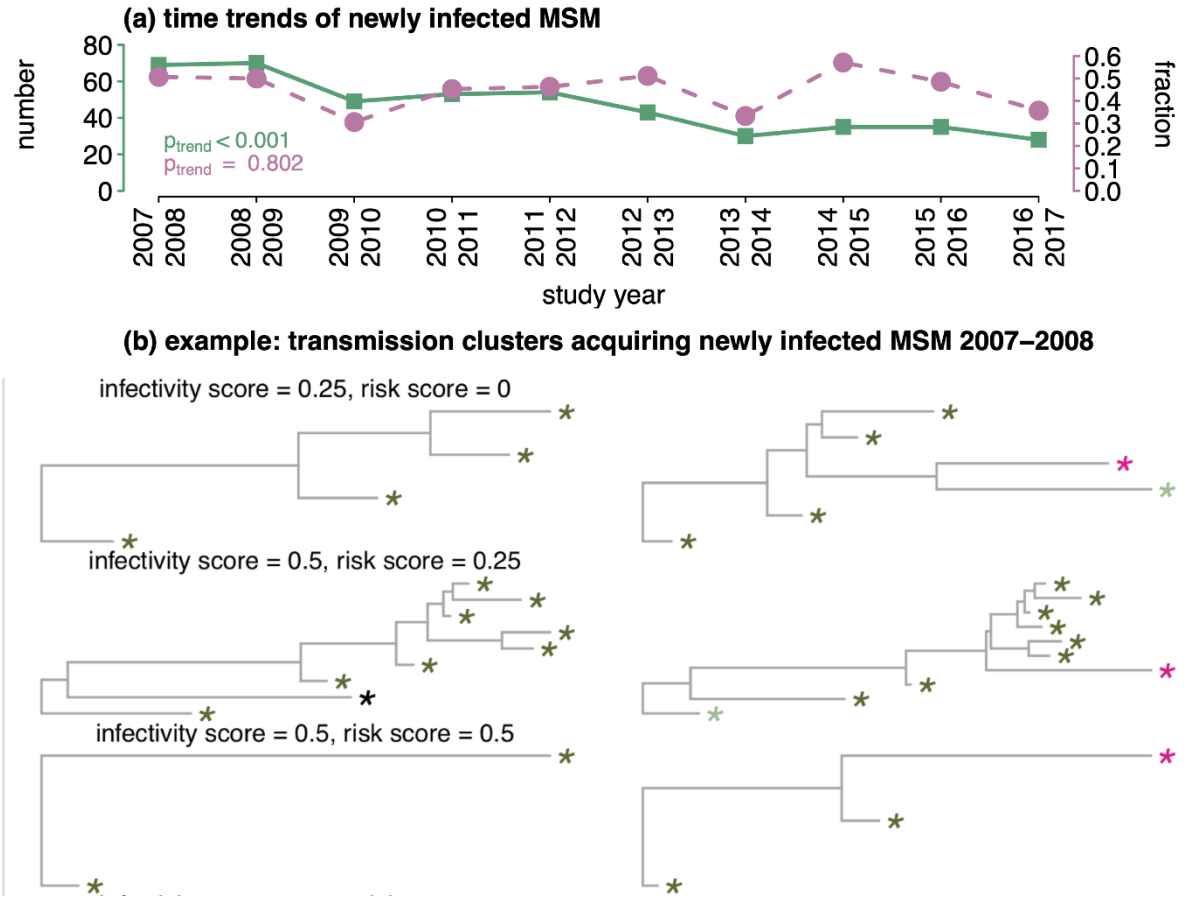

Figure 3

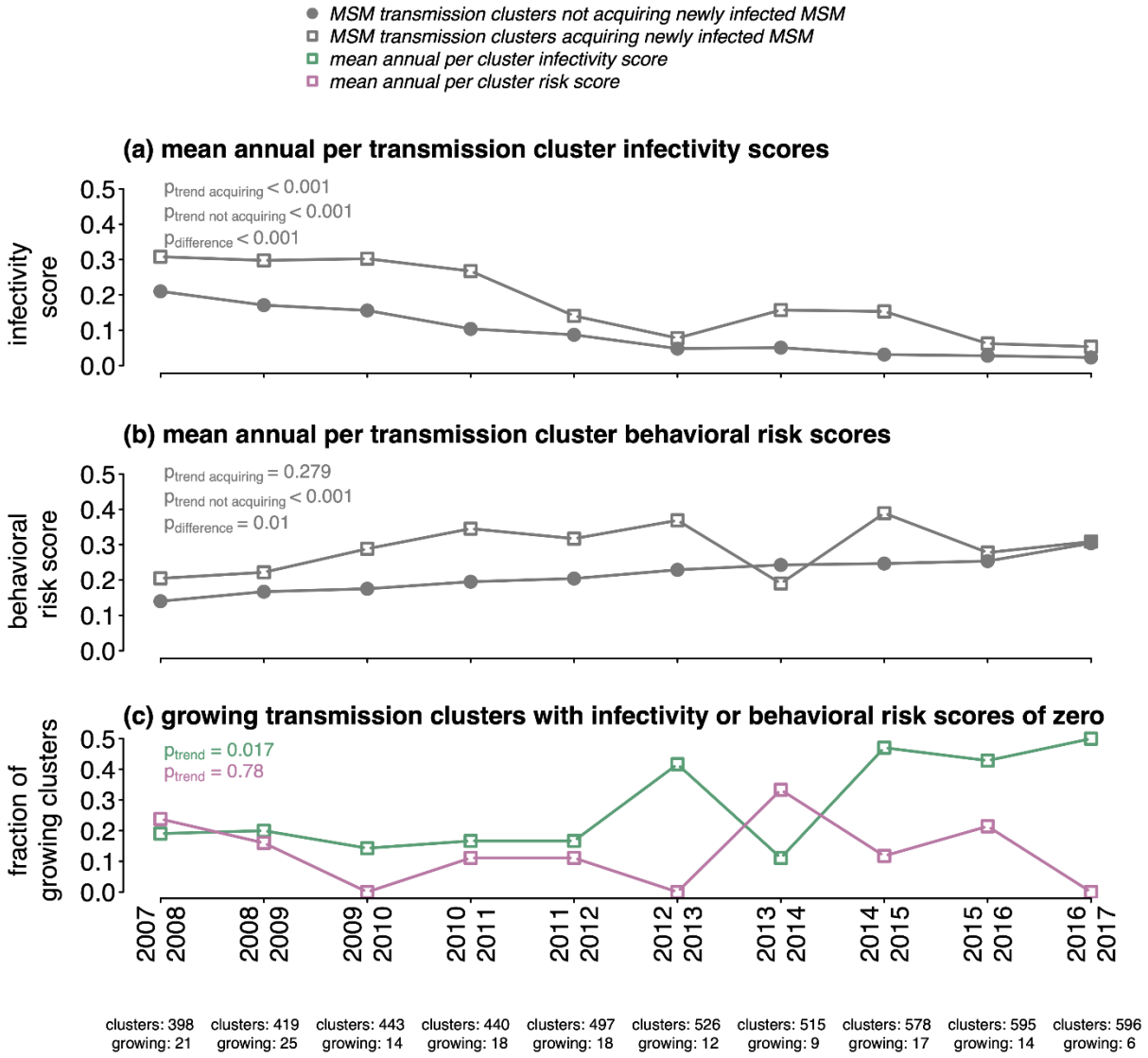

Figure 4