View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by UCL Discovery

Genetic Analysis Reveals the Complex Structure of HIV-1 Transmission within Defined Risk Groups

Stéphane Hué [*†], Deenan Pillay [*†§], Jonathan P. Clewley [†], and Oliver G. Pybus [‡]

*Centre for Virology, Division of Infection & Immunity, University College London, 46 Cleveland Street, London W1T 4JF, United Kingdom; [†]Centre for Infections, Health Protection Agency, 61 Colindale Avenue, London NW9 5HT, United Kingdom; [‡]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, United Kingdom

[§] To whom correspondence should be addressed. Telephone: + 44 (0) 20 7679 9490. Fax: + 44 (0) 20 7480 5896. E-mail: d.pillay@ucl.ac.uk

Major Category: Medical Sciences
Minor Category: Evolution

Text pages: 16
Figures: 1 double columns figure (Fig.2), 2 single column figures (Fig.1 & 3)
Tables: 1 double column-table (Table 1)

World count in abstract: 140
Total character count: 46,579

Sequences' Genbank Accession numbers: AY669865 to AY670087

Summary

We explored the epidemic history of HIV-1 subtype B in the United Kingdom using statistical methods that infer the population history of pathogens from sampled gene sequence data. Phylogenetic analysis of HIV-1 *pol* gene sequences from Britain showed at least six large transmission chains, indicating a genetically variable, but epidemiologically homogeneous, epidemic among men having sex with men. Through coalescent-based analysis we showed that these chains arose through separate introductions of subtype B strains into the United Kingdom in the early-to-mid 1980s. After an initial period of exponential growth, the rate of spread generally slowed in the early 1990s, which is more likely to correlate with behaviour change than with reduced infectiousness resulting from highly active antiretroviral therapy. Our results provide new insights into the complexity of HIV-1 epidemics that must be considered when developing HIV monitoring and prevention initiatives.

More than 57,700 people have been infected with human immunodeficiency virus type 1 (HIV-1) in the United Kingdom (UK) since the first identification of AIDS in 1982 (*http://www.hpa.org.uk/)*. Despite a recent increase in heterosexually acquired infections within the UK, predominantly originating in sub-Saharan Africa, the most prevalent clade of virus within the country remains subtype B, from the main group (M) of HIV-1, which is mainly transmitted through sex between men (1). To date, very little is known about how subtype B successfully invaded the British population, and more importantly, how the virus has subsequently spread and evolved.

Phylogenies reconstructed from sampled viral gene sequences hold valuable and unique information about the past structure of a population and can be used to understand the course of a viral epidemic over time (2, 3). Hence the history of a pathogen population can be inferred from the genealogy of randomly sampled strains (as represented by a phylogenetic tree) using the coalescent theory of population genetics (4, 5). By this means, one can reconstruct the changing number of infected individuals through time and estimate the demographic parameters that shape the epidemic, such as the rate of growth in the number of infections and the date of introduction of a lineage into a host population (6). Molecular data on HIV-1 within the UK has become increasingly available since the introduction of routine HIV-1 gene sequencing for drug-resistance monitoring. The genetic variability of the envelope (*env)* gene has previously made it attractive for evolutionary studies. However, we have recently demonstrated that the polymerase *(pol)* gene encodes sufficient variation to reconstruct transmission events, despite the potential bias conferred by emergence of drug resistance-associated mutations (7). Moreover, while the coalescent framework assumes neutral evolution, the HIV-1 *pol* gene is known to be under positive and negative selection (8-11). However, selection on HIV genes within infected individuals does not appear to generate non-neutral genealogies at the epidemiological (among-individual) level (12) and therefore should not significantly bias coalescent estimates. Importantly, previous coalescent analyses have yielded similar demographic estimates from different HIV-1 genes which are under considerably different selection pressures (13).

Using a new statistical framework, we reconstructed the history of the HIV-1 subtype B epidemic in the UK from a large dataset of contemporary *pol* gene sequences. For the first time, we characterised separate sub-epidemics of HIV-1 within a defined risk group, dating the introduction of epidemiologically significant viral lineages and estimating their rates of spread. Our analysis, using UK data, illustrates the complexity of HIV-1 epidemics that is applicable to other transmission groups and geographic regions.

Methods

Study population

HIV-1 subtype B *pol* gene sequences were generated from plasma samples collected in the UK by the Health Protection Agency's Antiviral Susceptibility Reference Unit. The samples were submitted for routine genotypic drug resistance testing between 1999 and 2003, and included samples from acute infections, chronic but drug-naïve infections, and from patients at the time of therapy failure. The sequences were 952 base pairs (bp) long, including the full protease gene as well as the first 218 codons of the reverse transcriptase (RT) gene. Around 85% these sequences were from men who had sex with men (MSM).

Phylogenetic reconstruction

To identify HIV-1 lineages derived from single independent introductions of the virus into the UK population, a neighbor-joining (NJ) phylogenetic tree was constructed from 3429 HIV-1 subtype B *pol* gene sequences (1645 UK isolates plus 1784 subtype B reference sequences from throughout the world) (14). The tree was estimated under the Hasegawa-Kishino-Yano model of nucleotide substitution (15). The non-UK sequences used for the study were extracted from GenBank (*http://www.ncbi.nlm.nih.gov/*) and the Los Alamos HIV Sequence Database (http://www.hiv.lanl.gov/). The size of the sequence alignment, as well as the computational power required, prevented the use of a more complex evolutionary model.

After identification of UK transmission clusters, sequences of non-UK origin were removed and the phylogenies of the clusters were re-estimated with the program Paup*, using a maximum likelihood approach (16). The trees were constructed under the General Time Reversible model of nucleotide substitution (17), with proportion of invariable sites and substitution rate heterogeneity, since this was the optimal model selected by the program Modeltest (18). Each UK cluster was rooted using a subtype D *pol* sequence from our database. The statistical robustness of the ML topologies was assessed by bootstrapping, using 1000 replicates (19). The sequences in the transmission clusters are deposited in GenBank under the accession numbers AY669865 to AY670087.

Estimation of HIV-1 subtype B *pol* gene rate of nucleotide substitution

In order to work within a calendar timescale (i.e. years), the genealogies were rescaled by applying a constant rate of nucleotide substitution μ (units are nucleotide substitutions/ site/year) to the branches of the phylogenies. Preliminary analyses demonstrated that the time span covered by our UK samples (i.e. five years) was not sufficient to reliably estimate μ. The rate of nucleotide substitution was therefore estimated from an independent dataset of 106 subtype B *pol* gene sequences. The sequences used to estimate μ were sampled between 1983 and 2000 from men having sex with men and injecting drug users (IDUs) participating in cohort studies at the Academic Medical Centre of Amsterdam (20). The sequences were 804 bp long, including the entire protease gene (294 bp) and the first 510 bp of the RT gene. GenBank accession numbers for these sequences are available in the original publication. A posterior distribution for substitution rate was estimated by Bayesian Markov Chain Monte Carlo (MCMC) inference (21) using a MCMC chain of 10,000,000 states sampled every 100th generation, as implemented in the program Beast (http://evolve.zoo.ox.ac.uk/beast). The estimated posterior distribution was subsequently used as an empirical prior distribution in the coalescent analyses that follow.

Estimation of demographic history and population dynamics

The investigation of the epidemic history of the six UK clusters involved two steps. Firstly, several different models of demographic history, each of which illustrate effective numbers of infections through time, were compared in order to select the model that best describes the epidemiological history of the UK transmission clusters. The demographic models were evaluated by likelihood ratio test (LRT), from likelihoods calculated by the program Genie (22). The five models tested in the present study were constant population size, exponential growth, piecewise logistic (exponential growth followed by constant population size), piecewise expansion (constant population size followed by exponential growth), and piecewise con-exp-con (constant growth periods flanking an exponential growth phase). See reference 16 for more details of these models. In order to fit a constant molecular clock framework, as required for coalescent analyses, the program TipDate (23) was used to rescale each transmission tree under the Single Rate Dated Tip (SRDT) model.

Secondly, the demographic and evolutionary parameters of the epidemic, together with their confidence intervals, were estimated by Bayesian MCMC inference using a chain of 10,000,000 states sampled every 100th generation, as implemented in the program Beast. The estimated parameters include the date of the most recent common ancestor (MCRA) of the cluster, the effective number of infections at the most recent time of sampling *Ne* (i.e. the effective number of prevalent infections), and the growth rate during the exponential phase *r*. The Bayesian MCMC results were used to calculate a marginal posterior distribution of the demographic model for each cluster, a graphical representation of the effective number of infections through time, generated using the program Tracer (http://evolve.zoo.ox.ac.uk/tracer/*).*

Results

Introduction of HIV-1 subtype B into the UK

The initial NJ phylogenetic tree constructed from 3429 UK and worldwide subtype B pol sequences is too large to display here (see supplementary information). A schematic representation of the clustering patterns seen within the phylogeny is presented in Fig.1. Three clustering patterns were distinguished, namely, sporadic UK sequences, non-UK transmission clusters, and UK transmission clusters. Sporadic UK sequences (i.e. those that do not group with other UK lineages in the tree) probably represent single, independent introductions of the virus without subsequent spread. Transmission clusters were identified as clades of sequences from a particular location that descend from a common ancestor, indicating spread of the virus in that region. UK transmission clusters were differentiated from non-UK clusters on the basis of the size of the clade and the proportion of UK sequences within it: UK transmission clusters were defined as those clades with more than 25 sequences, 90% or more of which were of UK origin. A minimum clade size of 25 was used because smaller sample sizes are unlikely to give reliable coalescent estimates under complex demographic models. A minimum fraction of 90% UK sequences was chosen to ensure that the clusters that were identified represent chains of transmission that have overwhelmingly occurred in the UK. However, we note that this methodology probably underestimates the number of transmission chains identified.

Most of the UK sequences represented sporadic lineages (86%), scattered among sequences from other geographical areas, suggesting much geographical mixing and migration of subtype B strains on a worldwide scale. Nonetheless, six UK transmission clusters were identified, involving 45, 62, 29, 26, 27 and 34 sequences. These transmission chains were distinct (i.e. reciprocally monophyletic), indicating that at least six independent introductions of subtype B HIV-1 have succeeded in sustaining onward transmission within the UK over time, and until the present. Each transmission chain contained sequences from a variety of locations within the UK and no obvious geographic correlations were observed. The robustness of the clusters within the overall tree could not be statistically evaluated due to the huge size of the dataset. Nonetheless, the branching patterns of the six UK lineages showed statistical robustness when compared to subsets of worldwide control sequences using bootstrap analyses (neighbor-joining method with 1000 replicates, as implemented in the program Paup*; data not shown). To further explore the history of these six successful viral lineages, sequences of non-UK origin were removed from the six clusters and the phylogenetic histories of the UK sequences were rigorously re-estimated using a maximum likelihood approach. The ML trees are available from the authors on request.

Estimation of the rate of evolution for the HIV-1 subtype B *pol* gene

The rate of evolution for the subtype B HIV-1 *pol* gene was calculated using an independent dataset of 106 sequences, sampled between 1983 and 2000 in Amsterdam (20). Using a Bayesian MCMC method, this rate was estimated to be 2.55x10^{-3} substitutions per nucleotide site per year (95% confidence intervals: 1.74x10^{-3} to 3.51x10^{-3}). In comparison, previous estimates of HIV-1 evolution rates have typically relied on partial *env* gene sequences and have ranged from 2.4 x 10^{-3} to 6.7 x 10^{-3} subst./site/year (24-26). Our estimate is consistent with the order of magnitude of 10^{-3} expected for an HIV-1 gene. The phylogenetic trees in Fig.2 are thus shown on a timescale of years.

Epidemic history and parameter estimation

For each of the six clusters, a model of logistic population growth best fitted the demographic information contained in the tree topologies (likelihoods shown in supplementary information). Under the logistic model, the effective number of infections *Ne* grows exponentially at rate *r* from time t_a (time of the most recent common ancestor of the cluster) then decreases in growth rate towards the present. A schematic representation of the logistic model is given in Fig.3. Note that *Ne* reflects the number of infections contributing to new infections, rather than the total number of prevalent infections within the transmission cluster.

The demographic parameters that determine the shape of the logistic growth curve were estimated by Bayesian MCMC inference (Table 1) and the epidemic histories of the six clusters were reconstructed, with appropriate confidence limits (Fig. 2). Our estimates suggest that three of the six genealogies originated in the early 1980s (1981 for cluster 2, 1983 for clusters 1 and 3), whereas the remaining clusters were introduced later in the same decade (1986 for clusters 4 and 6, 1987 for cluster 5). While the initial exponential growth phase clearly ended in the early 1990s for clusters 1 to 5 (see fig. 2a to 2e), the growth rate decrease is more tentative for cluster 6 and is only apparent very recently (see Fig. 2f), such that cluster 6 appears to also fit a model of exponential growth. To explore this issue further, we estimated the epidemic doubling time of each transmission cluster at the most recent sampling time, year 2003 (by rearrangement of the model in Pybus et al., 2001)(27). This ‘current’ epidemic doubling time is considerably shorter for cluster 6 than for clusters 1-5; specifically, the current doubling time for cluster 6 is significantly more likely to be <20 years (equal to an exponential growth rate >0.035 years^{-1}) in comparison to the other clusters (data not shown). In marked contrast, the exponential growth rates at the time of initiation of each cluster (*r*) are very similar, with an average of 0.80 years^{-1}. Finally, the current effective number of infections *Ne* varied from cluster to cluster, ranging from 94 (cluster 5) to 1350 (cluster 6) effective infections.

Discussion

Our estimates suggest that the HIV-1 subtype B epidemic currently circulating within the UK is comprised of at least six established chains of transmission, introduced in the early and mid 1980s. This demonstrates the existence of distinct, possibly non-overlapping sexual networks within the predominant MSM risk group and argues against the hypothesis that one initial entry of HIV-1 was responsible for the spread of the subtype B epidemic. It also emphasises the role of migration in the HIV-1 epidemic in Britain, as illustrated by the overwhelming prevalence of sporadic lineages (86% of the total UK samples) in the genealogy, representing viruses arising from outside the UK that have failed to establish a large outbreak.

The transmission clusters we characterised had similar epidemic curves and geographic distributions within the UK, indicating a concurrent spread under similar demographic pressures, at least during the early stages of the epidemic. The introduction of the earlier viruses in the early 1980s (i.e. clusters 1-3) seems to coincide with the explosion of new infections reported by epidemiological data at the time (*http://www.hpa.org.uk/)*. The coupling of HIV strain ‘immigration’ with epidemiological changes is likely to have favoured the emergence and persistence of the transmission chains presently circulating amongst MSM. However, the first UK cases of AIDS were reported in 1982 (*http://www.who.int/emc-hiv/fact_sheets/*), and these individuals were probably infected within a window of 10 years prior to that time, hence the currently circulating strains may not represent the first HIV-1 lineages identified within the UK. If earlier strains existed they may have been unsuccessful in sustaining transmission chains until the present, and may no longer be of epidemiological significance. However, the absence of older strains could also reflect a sampling bias.

For all six transmission clusters, the exponential growth phase coincides with a reported augmentation of newly-acquired HIV-1 infections within MSM and IDU in the UK (*http://www.hpa.org.uk/)*. The average growth rate during the initial exponential phase was estimated to be 0.80 years^{-1} (ranging from 0.47 to 1.38), approximating a doubling time of 1 year. This value is similar to that estimated for the US subtype B epidemic (0.83 years^{-1}, 0.72 to 0.94), suggesting that the two epidemics follow similar trends at the macro-epidemiological scale (26). This idea is supported by the effective number of infections estimated for the two epidemics. Despite a wide variation in *Ne* across the six UK transmission clusters, the average effective number of infections among the six UK clusters is 445, which is approximately 2.5% of the infected population. This is remarkably similar to the values for the US epidemic, where the effective number of infections and prevalence in 1995 reached 5000 and 200,000 infections, respectively. *Ne* represents the number of infections contributing to onward transmission, rather than the larger number of actual infections. Importantly, we observe that the population represented by cluster 6 exhibits a faster doubling time in 2003 than the other five clusters, suggesting a difference in current growth rate among clusters. Current surveillance data for the UK reports a very recent increase in infections in MSM (*http://www.hpa.org.uk/)* and it is reasonable to suppose that the lineage we have identified as cluster 6 has contributed to this recent upturn in infection.

Since 1990, there have been important changes in Britain's demographic structure, social attitude and awareness of HIV-1/AIDS (28). Despite a very recent increase in high-risk behaviour among men having sex with men (such as the number of sexual partners or concurrent partnerships), a significant increase in consistent condom use has been reported since 1990. Such a change in sexual health, coupled to large-scale educational campaigns over the past decade, could explain the equilibrium reached by the effective number of prevalent infections. The effect of antiretroviral therapy on past epidemic dynamics should also be considered: although such therapy is instituted primarily to reduce progression of disease, it may also impact on transmission through reduction of infectivity. If so, we would expect evidence of growth rate decrease in the late (rather than early) 1990s – the time that highly active antiretroviral therapy became widely used. In fact, Health Protection Agency data suggests no significant changes in the incidence of HIV-1 within gay men since the late 1980s, and an actual increase over the past 3 years (29). We therefore suggest that antiviral therapy has not had a significant impact on the growth of the epidemic; indeed, some studies suggest that the epidemic is driven by transmissions in primary infection (30-32), before therapy is usually initiated. The current increase in new infections is too recent to be reflected in the growth dynamics of any of the six populations identified by our analysis. On-going analyses of the type undertaken here will clarify whether the recent increase in new subtype B infections derive from longstanding viral lineages, or newly introduced viruses.

In conclusion, we show that currently circulating HIV-1 subtype B strains entered the UK in the mid 1980s and that the rate of spread of these lineages slowed in the early 1990s. It is often assumed that the HIV-1 epidemic within the UK is composed of smaller, independent epidemics defined by risk group. We demonstrate here the existence of multiple sub-epidemics (at least six) within MSM that obey similar demographic constraints during their early stages, yet exhibit differences in their more recent rates of spread. The identification of these multiple lineages within the predominant risk group of the HIV-1 epidemic in the UK suggests the existence of sub-epidemics within groups of MSM, and it is reasonable to assume that this structure exists in comparable risk groups in other countries. Such heterogeneity must therefore be considered when developing HIV monitoring prevention and treatment initiatives.

Acknowledgements

We thank Dr Patricia Cane, from the Health Protection Agency Antiretroviral Susceptibility Reference Unit, and Dr Vladimir Lukashov, from the Academic Medical Center of Amsterdam, for kindly providing the sequences used in this study. We thank Dr Andy Rambaut for his support in the phylogenetic analyses. We also thank Prof. Robin Weiss and Dr Paul Kellam for their useful comments on the manuscript. This work was funded by the Health Protection Agency, UK. Dr Oliver Pybus was funded by the Welcome Trust and The Royal Society.

References

1. Murphy, G., Charlett, A., Jordan, L. F., Osner, N., Gill, O. N. & Parry, J. V. (2004) *Aids* **18,** 265-72.

2. Holmes, E. C., Nee, S., Rambaut, A., Garnett, G. P. & Harvey, P. H. (1995) *Philos Trans R Soc Lond B Biol Sci* **349,** 33-40.
3. Nee, S., Holmes, E. C., Rambaut, A. & Harvey, P. H. (1995) *Philos Trans R Soc Lond B Biol Sci* **349,** 25-31.
4. Kingman, J. F. (2000) *Genetics* **156,** 1461-3.
5. Griffiths, R. C. & Tavare, S. (1994) *Philos Trans R Soc Lond B Biol Sci* **344,** 403-10.
6. Kuhner, M. K., Yamato, J. & Felsenstein, J. (1995) *Genetics* **140,** 1421-30.
7. Hué, S., Clewley, J. P., P.A., C. & Pillay, D. (2004) *AIDS* **18,** 719-728.
8. Leal, E. d. S., Holmes, E. C. & Zanotto, P. M. (2004) *Virology* **325,** 181-91.
9. Richman, D. D., Havlir, D., Corbeil, J., Looney, D., Ignacio, C., Spector, S. A., Sullivan, J., Cheeseman, S., Barringer, K., Pauletti, D. & et al. (1994) *J Virol* **68,** 1660-6.
10. Frost, S. D., Nijhuis, M., Schuurman, R., Boucher, C. A. & Brown, A. J. (2000) *J Virol* **74,** 6262-8.
11. Rouzine, I. M. & Coffin, J. M. (1999) *J Virol* **73,** 8167-78.
12. Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L., Daly, J. M., Mumford, J. A. & Holmes, E. C. (2004) *Science* **303,** 327-32.
13. Lemey, P., Pybus, O. G., Wang, B., Saksena, N. K., Salemi, M. & Vandamme, A. M. (2003) *Proc Natl Acad Sci U S A* **100,** 6588-92.
14. Saitou, N. & Nei, M. (1987) *Mol Biol Evol* **4,** 406-25.
15. Hasegawa, M., Kishino, H. & Yano, T. (1985) *J Mol Evol* **22,** 160-74.
16. Felsenstein, J. (1973) *Am J Hum Genet* **25,** 471-92.
17. Yang, Z. (1994) *J Mol Evol* **39,** 105-11.
18. Posada, D. & Crandall, K. A. (1998) *Bioinformatics* **14,** 817-8.
19. Felsenstein, J. (1985) *Evolution* **39,** 783-791.
20. Lukashov, V. V. & Goudsmit, J. (2002) *J Mol Evol* **54,** 680-91.
21. Drummond, A. J., Nicholls, G. K., Rodrigo, A. G. & Solomon, W. (2002) *Genetics* **161,** 1307-20.
22. Pybus, O. G. & Rambaut, A. (2002) *Bioinformatics* **18,** 1404-5.
23. Rambaut, A. (2000) *Bioinformatics* **16,** 395-9.
24. Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B. H., Wolinsky, S. & Bhattacharya, T. (2000) *Science* **288,** 1789-96.
25. Leitner, T., Escanilla, D., Franzen, C., Uhlen, M. & Albert, J. (1996) *Proc Natl Acad Sci U S A* **93,** 10864-9.
26. Robbins, K. E., Lemey, P., Pybus, O. G., Jaffe, H. W., Youngpairoj, A. S., Brown, T. M., Salemi, M., Vandamme, A. M. & Kalish, M. L. (2003) *J Virol* **77,** 6359-66.
27. Pybus, O. G., Charleston, M. A., Gupta, S., Rambaut, A., Holmes, E. C. & Harvey, P. H. (2001) *Science* **292,** 2323-5.
28. Johnson, A. M., Mercer, C. H., Erens, B., Copas, A. J., McManus, S., Wellings, K., Fenton, K. A., Korovessis, C., Macdowall, W., Nanchahal, K., Purdon, S. & Field, J. (2001) *Lancet* **358,** 1835-42.
29. Brown, A. E., Sadler, K. E., Tomkins, S. E., McGarrigle, C. A., LaMontagne, D. S., Goldberg, D., Tookey, P. A., Smyth, B., Thomas, D., Murphy, G., Parry, J. V.,

Evans, B. G., Gill, O. N., Ncube, F. & Fenton, K. A. (2004) *Sex Transm Infect* **80,** 159-66.

30. Koopman, J. S., Jacquez, J. A., Welch, G. W., Simon, C. P., Foxman, B., Pollock, S. M., Barth-Jones, D., Adams, A. L. & Lange, K. (1997) *J Acquir Immune Defic Syndr Hum Retrovirol* **14,** 249-58.
31. Jacquez, J. A., Koopman, J. S., Simon, C. P. & Longini, I. M., Jr. (1994) *J Acquir Immune Defic Syndr* **7,** 1169-84.
32. Yerly, S., Vora, S., Rizzardi, P., Chave, J. P., Vernazza, P. L., Flepp, M., Telenti, A., Battegay, M., Veuthey, A. L., Bru, J. P., Rickenbach, M., Hirschel, B. & Perrin, L. (2001) *Aids* **15,** 2287-92.

Figure legends

Fig.1

Schematic representation of the phylogeny generated from 3429 UK and worldwide HIV-1 subtype B *pol* sequences. Filled circles represent sequences from the UK, while open squares represent non-UK sequences. Three branching patterns were distinguished: (a) non-UK transmission clusters, (b) sporadic UK infections, and (c) UK transmission clusters. Transmission clusters are clades of sequences from a particular location that descend from a common ancestor, indicating a successful spread of the virus in that location. UK transmission clusters are defined as those clades that include at least 25 sequences, 90% or more of which are of UK origin.

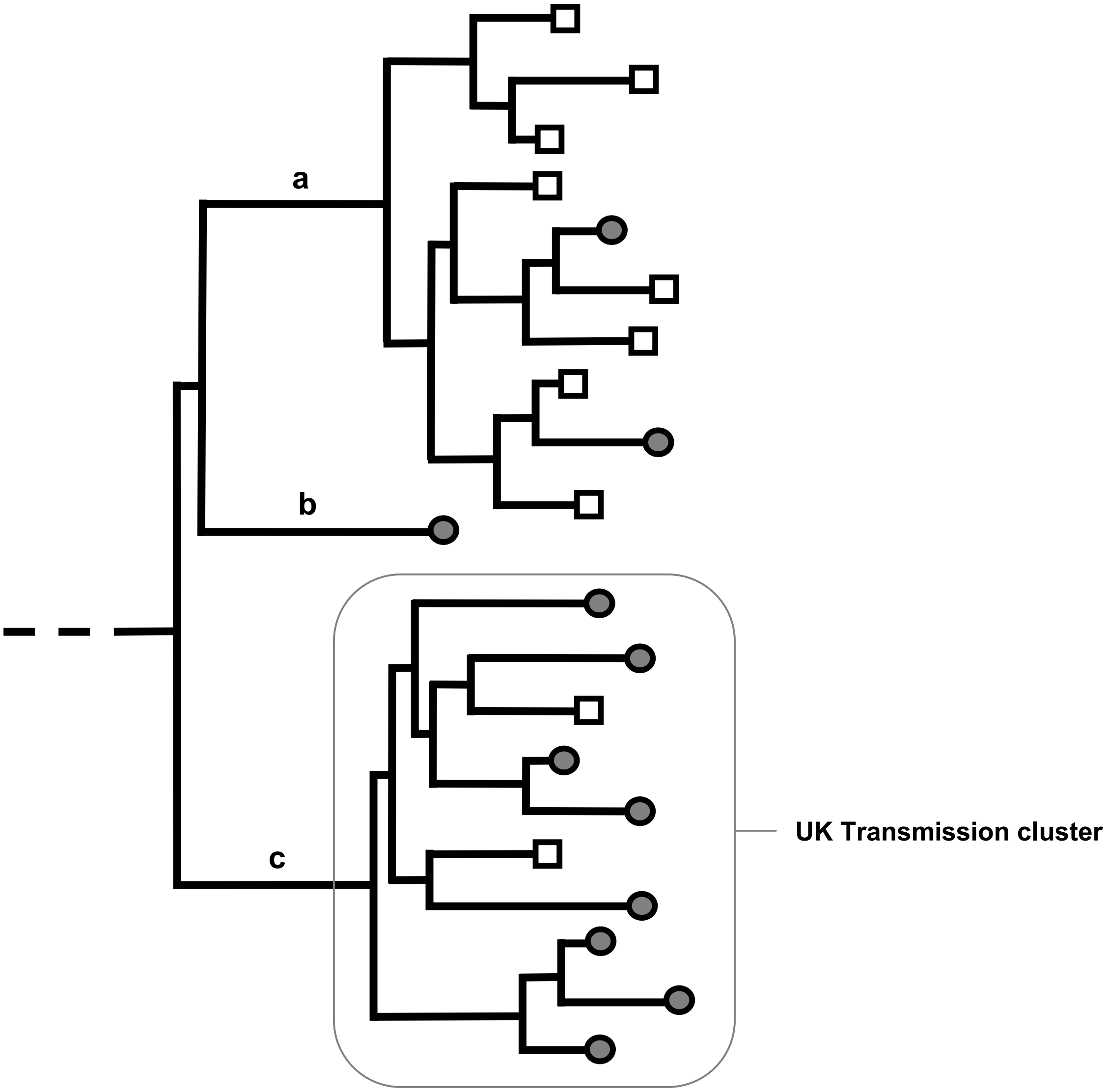
Fig. 2

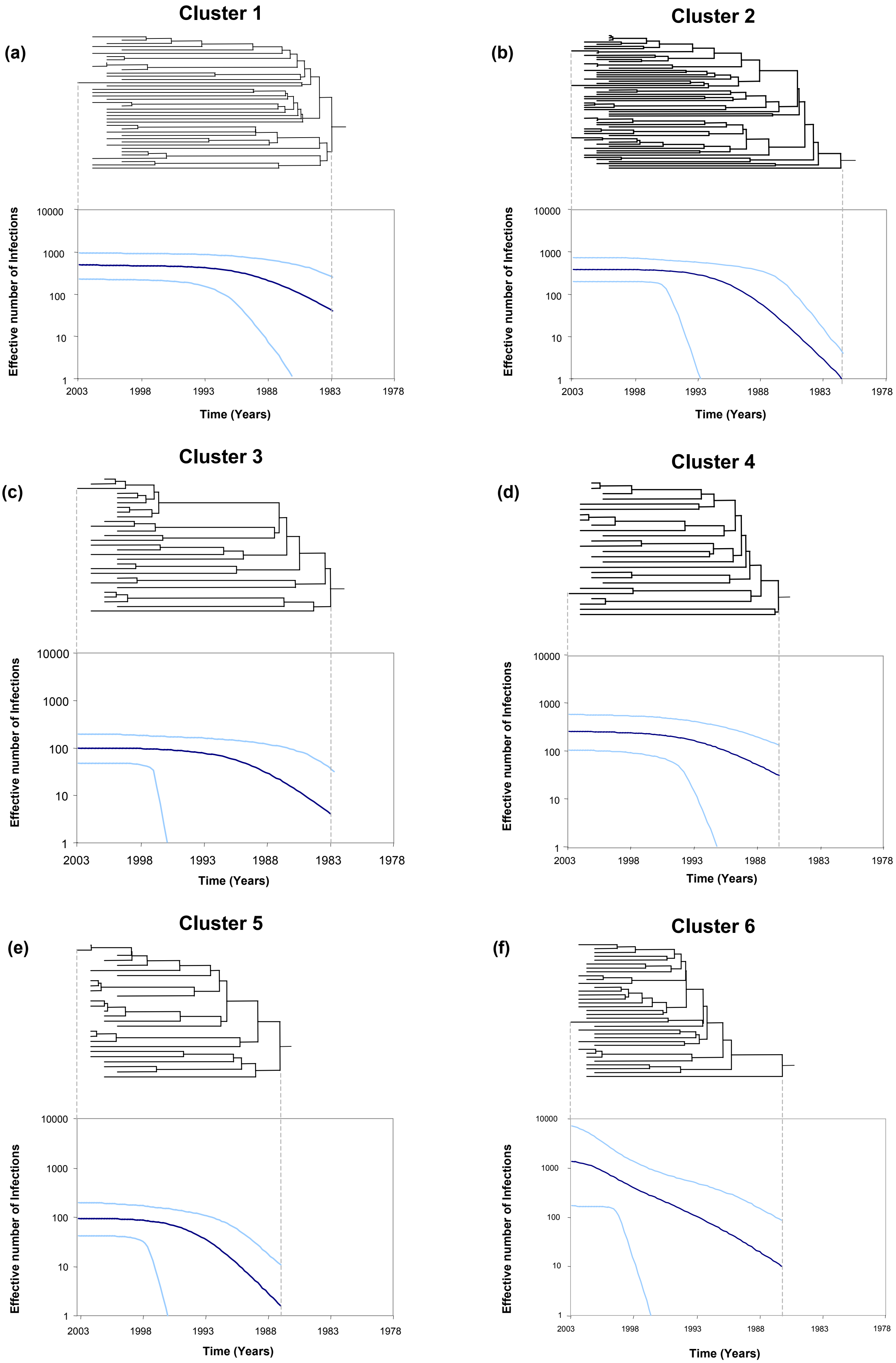
Phylogenetic trees of the six UK transmission clusters and their corresponding estimated epidemic histories (all shown on the same timescale). The trees represent the ancestral relationships of sequences belonging to each cluster. (a) cluster 1, (b) cluster 2, (c) cluster 3, (d) cluster 4, (e) cluster 5, (f) cluster 6. The demographic histories were estimated by Bayesian MCMC inference using a model of logistic growth (see text for details) and show change in the effective number of infections through time (timescale in calendar years). The dark line shows the median estimate of the effective number of infections, whereas the light lines show the 95% confidence limits of the estimate.

Fig. 3

Schematic representation of the logistic model of population growth.

According to this model, the number of infections population grows exponentially at rate *r* from time t_a (time of the most recent common ancestor of the sampled sequences). The growth rate slows as time moves towards the present, such that *Ne* represents the effective number of infections at the present. *Ne* can be thought of as the number of infections contributing to new infections, rather than the total number of prevalent infections within the cluster.


a
b
c
UK Transmission cluster


Cluster 1
(a)
Effective number of Infections
10000
1000
100
10
1
2003
1998
1993
1988
1983
1978
Time (Years)
Cluster 2
(b)
Effective number of Infections
10000
1000
100
10
1
2003
1998
1993
1988
1983
1978
Time (Years)
Cluster 3
(c)
Effective number of Infections
10000
1000
100
10
1
2003
1998
1993
1988
1983
1978
Time (Years)
Cluster 4
(d)
Effective number of Infections
10000
1000
100
10
1
2003
1998
1993
1988
1983
1978
Time (Years)
Cluster 5
(e)
Effective number of Infections
10000
1000
100
10
1
2003
1998
1993
1988
1983
1978
Time (Years)
Cluster 6
(f)
Effective number of Infections
10000
1000
100
10
1
2003
1998
1993
1988
1983
1978
Time (Years)

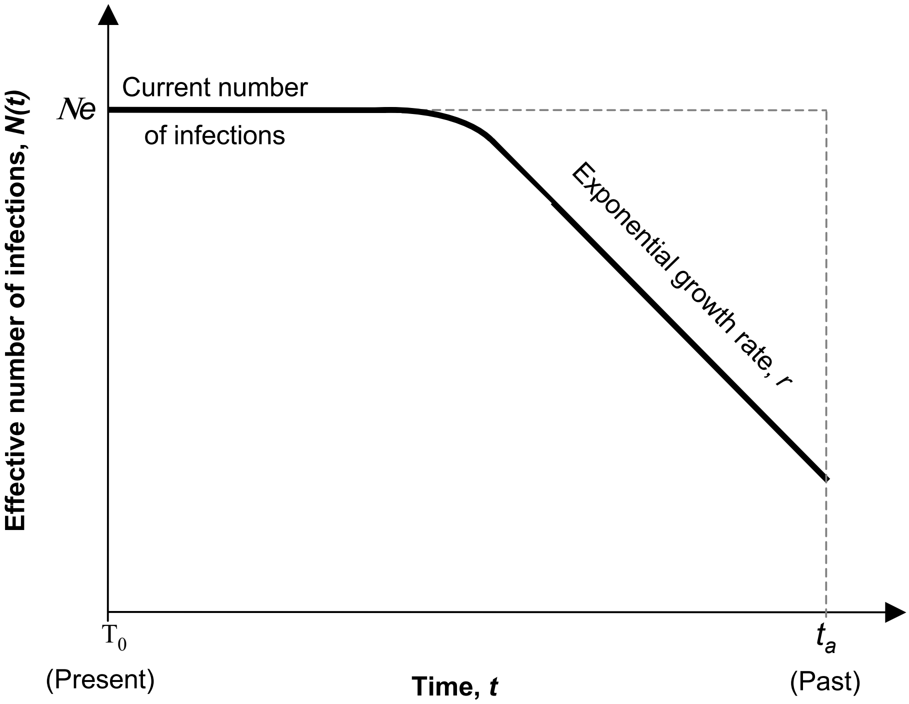

Effective number of infections, N(t)
Ne
Current number
of infections
Exponential growth rate, r
T0
(Present)
Time, t
ta
(Past)

Table 1. Parametric estimates (with 95% confidence intervals) under the logistic growth demographic modelfor the six lineages

Cluster	μ [a]	*Ne* [b]	*r* [c]	Origin of the tree (*yrs)*
Cluster 1	2.55 x 10^{-3} (0.0017, 0.0035)	493 (201, 833)	1.08 (0.66, 2.56)	1983 (1978, 1988)
Cluster 2	2.55 x 10^{-3} (0.0017, 0.0035)	386 (190, 655)	0.47 (0.30, 0.95)	1981 (1976, 1987)
Cluster 3	2.55 x 10^{-3} (0.0017, 0.0035)	98 (42 , 171)	0.50 (0.19, 4.62)	1983 (1977, 1988)
Cluster 4	2.55 x 10^{-3} (0.0017, 0.0035)	250 (88, 483)	1.38 (0.63, 2.50)	1986 (1982, 1991)
Cluster 5	2.55 x 10^{-3} (0.0017, 0.0035)	94 (36, 85)	0.68 (0.35, 2.10)	1987 (1983, 1991)
Cluster 6	2.55 x 10^{-3} (0.0017, 0.0035)	1350 (109, 5489)	0.67 (0.37, 3.85)	1986 (1981, 1991)
US cluster [d]	6.7 x 10^{-3} (n/a, n/a)	4 830 (1995, 26 750)	0.834 (0.72 / 0.945)	1968 (1966, 1970)

[a] Rate of nucleotide substitution, in substitutions per site per year, estimated from an independent dataset of subtype B *pol* sequences

[b] Effective number of infections

[c] Rate of exponential growth, in *years* $^{-1}$

[d] from Robbins *et al.* , 2003