Molecular Epidemiology of Human Immunodeficiency Virus Type 1 (HIV-1) in Southern Africa and Northern California

By

Sudeb C. Dalai

A dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Epidemiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Arthur Reingold
Professor Maya Petersen
Professor Montgomery Slatkin

Summer 2017

Abstract

Molecular Epidemiology of Human Immunodeficiency Virus Type 1 (HIV-1)
in Southern Africa and Northern California

by

Sudeb C. Dalai

Doctor of Philosophy in Epidemiology

University of California, Berkeley

Professor Arthur Reingold, Chair

Human immunodeficiency virus type 1 (HIV-1) exhibits extraordinary genetic diversity that is driven by high rates of mutation and recombination, coupled with elevated rates of viral turnover and the persistent nature of infection. Through these mechanisms, HIV-1 group M, the group largely responsible for the global pandemic, diversified into nine distinct subtypes and additional circulating recombinant forms. Subtype C HIV-1, which accounts for approximately 50% of the estimated 37 million individuals living with HIV/AIDS worldwide, predominates the epidemics in southern Africa and South / Southeast Asia. In contrast, subtype B HIV-1, predominant in the United States and Europe, comprises 12% of the global HIV-1 prevalence.

Current US and, increasingly, international guidelines recommend that HIV genotypic sequencing be performed for newly-identified HIV-positive patients, in order to identify genetic mutations that may confer drug resistance and pose a barrier to effective antiretroviral therapy (ART). This practice has resulted in the generation of enormous amounts of genotypic data often linked with clinical, demographic, and geospatial information. Analyses of these large datasets, utilizing sophisticated statistical and computational methods, has enabled identification of important and previously unrecognized trends in epidemiologic and vertical transmission, persistence of HIV within and among distinct risk groups, and the accumulation and propagation of ART resistance. Results of these analyses, in turn, advance treatment and care for persons living with HIV.

This thesis takes advantage of large, de-identified and anonymized datasets of HIV-1 genotypic and demographic information available through clinical testing and treatment programs in two distinct and epidemiologically important regions in the global HIV pandemic, northern California and southern Africa. Employing innovative and computationally intensive tools which combine experience from molecular biology, evolutionary biology and biostatistics, in-depth analyses were undertaken to explore and define risk factors, underlying epidemiologic features and clinical characteristics relevant to transmission and ART resistance. The final chapter provides a comprehensive literature review of the relevance of molecular epidemiologic principles and innovative biology to the complex and continued phenomenon of mother-to-child transmission of HIV, as a case study exemplifying the need for continued molecular studies grounded in solid epidemiologic principles. The findings presented here are ultimately intended to help guide domestic and global efforts to scale up ART treatment while considering well-known and lesser-known factors that may influence eventual clinical and epidemiologic outcomes.

ACKNOWLEDGEMENTS

To the Dissertation Committee, Dr. Maya Petersen and Dr. Montgomery Slatkin, thank you for your helpful advice, your thoughtful comments and your inspiring example during the process of this important work.

To the Committee Chair, Dr. Arthur Reingold, thank you for your supportive mentorship, patience and helpful guidance throughout the process of my graduate training.

To the Qualifying Committee examiners, Dr. Jack Colford, Dr. Maya Petersen, Dr. Montgomery Slatkin, and Dr. Anand Chokkalingam, thank you for your rigorous evaluation and helpful advising during the process of my qualifying examination.

To my lifelong Mentor Dr. David Katzenstein, thank you for all the tremendous generosity, support, stimulating ideas, advising, devotion and companionship throughout the years of my academic training and career pursuits. Your inspiration for and impact on my academic and personal endeavors is substantial and immeasurable.

Thank you to Drs. Brian Blackburn, Lucy Tompkins, Upinder Singh, and Julie Parsonnet in the Stanford University Division of Infectious Disease for their academic and career support and their unwavering commitment to medical training, physician education and clinical care.

Formal acknowledgements:

This author would like to acknowledge the generous support of the Doris Duke Charitable Foundation, the Infectious Disease Society of America, the Paul and Daisy Soros Fellowship for New Americans, the California HIV Research Program and the National Institutes of Health / National Institute of Allergy and Infectious Diseases T32 Institutional Training Award.

This author would like to acknowledge the following co-authors on the relevant published materials.

Tulio de Oliveira, Gordon Harkins, Seble Kassaye, Jennifer Lint, Justen Manasa, Elizabeth Johnston, David Katzenstein for AIDS (2009).

Philip Bulterys and David Katzenstein for Clinics in Perinatology (2010).

TABLE OF CONTENTS

Chapter 1: Molecular Epidemiology of Subtype C HIV-1 in Southern Africa

INTRODUCTION

HIV-1 subtype C now accounts for approximately 50% of the estimated 33 million people living with HIV/AIDS and half of the 2-3 million new infections annually [1]. While the majority of subtype C infections are in southern Africa, this subtype also dominates the epidemics in India, Ethiopia, and southern China, and has entered East Africa, Brazil, and many European countries. Recombinant viruses including genes derived from subtype C have been increasingly recognized in China, Thailand, and Taiwan, where phylogenetic studies indicate that complex BC subtype recombinants, such as circulating recombinant forms CRF_07 and CRF_08, comprise much of the current epidemics [2-4].

The predominance of a single clade of HIV-1 in the most severely affected countries in sub-Saharan Africa has been ascribed to a founder effect [5]. High levels of continued subtype C transmission are thought to be sustained by sexual-social factors, including low rates of male circumcision, the frequency of concurrent partnerships, increased virus load, and shedding of virus in a context of other sexually transmitted infections [6, 7]. Comparisons with subtype B isolates, which predominate infections in the Americas and western Europe, have identified characteristics of subtype C viruses that may explain differences in infectivity, including enhanced tropism for macrophages and dendritic cells, elevated viral replication rates through transcriptional regulation [8], and significantly higher rates of mutation and emergence of drug resistance in women receiving single-dose nevirapine [9, 10]. Studies of *in vitro* replication rates of subtypes A, B, C and D suggest lower pathogenic fitness but equivalent transmission efficiency of HIV-1 subtype C, suggesting a higher rate of transmission [11, 12]. This may provide a partial, virologic explanation for the disproportionately high rates of HIV-1 infection in southern Africa, where a longer period of persistent infection and transmission preceding symptomatic disease could increase both the population prevalence and the reproductive rate of the epidemic.

The routine, population-based genotyping of circulating viruses, as a surveillance tool for drug resistance, has been exploited for evolutionary and phylogenetic mapping and to explore the origins, molecular epidemiology, and genetic diversity of HIV-1 [13, 14]. Analysis of spatio-temporally sampled sequence data enables the reconstruction of epidemic histories and estimation of demographic parameters, including measures of circulating viral diversity and population-level prevalence over time [15]. We analyzed publicly available subtype C pol sequences over a fifteen-year period between 1991-2006 from successive cohorts of women screening for HIV infection in antenatal clinics in Harare, Zimbabwe. Using a combination of molecular clock analysis, to estimate the timescale of the epidemic, and a Bayesian coalescent-based approach, to infer demographic parameters of virus transmission, we present new information on the origins, timing, and epidemic growth patterns of subtype C HIV-1 during a period of in-migration and political change in Zimbabwe.

MATERIALS AND METHODS

Study Population

Plasma samples were previously obtained in four prior studies of HIV infection and pregnancy from HIV-positive women enrolled at antenatal clinics in Harare, Zimbabwe (1991) and the neighboring suburb of Chitungwiza (1998, 2000 and 2006). Each cohort comprised HIV-infected women enrolled in prevention of mother-to-child-transmission (pMTCT) studies approved by the Medical Research Council of Zimbabwe and Stanford University. Samples were collected at 28-36 weeks of pregnancy before antiretroviral drugs were initiated for the prevention of mother-to-child transmission (**Table 1**).

All sequence data utilized for analysis were publicly available and obtained from the public repository NCBI GenBank (https://www.ncbi.nlm.nih.gov/genbank/) in association with the relevant prior publications, which are cited accordingly.

Sequence Generation

We obtained 177 pol sequences from 4 sequential cohorts of young treatment naïve women presenting to antenatal clinics in Zimbabwe from 1991-2006. All datasets of HIV-1 subtype C pol used in this analysis were previously published (SIDA 1998 [16], HPTN023 2001 [17], WHO 1991 [18], NIH 2006 [19] and obtained from NCBI GenBank. All sequence data were created through bi-directional dideoxynucleotide sequencing of pol genes, followed by direct sequencing of PCR products as previously described.

Sequence Alignment

Sequences were aligned with ClustalW [20] and manually edited using BioEdit [21]. A large dataset of reference HIV-1 subtype C sequences (n=981) was downloaded from the BioAfrica website (http://www.bioafrica.net/subtype/subC/) and used to characterize the relationship between the Zimbabwean sequences and other subtype C sequences worldwide.

Subtype Classification

HIV-1 subtype was characterized for the Zimbabwean sequence datasets using the REGA Subtyping Tool v2.0 [22]. Evidence for inter-subtype recombination was assessed with bootscanning analysis implemented in Simplot v3.5 [23].

Phylogenetic Analysis

A best-fitting nucleotide substitution model for the Zimbabwean HIV-1 subtype C pol sequences was estimated using hierarchical likelihood ratio tests (hLRT's) implemented in the program Modeltest v3.7 [24] and manual examination in PAUP v4.0 [25]. Maximum likelihood (ML) phylogenetic trees were constructed using the inferred model, GTR + I + G, with the program PhyML v2.4. This method employs a neighbor-joining (NJ) tree as a starting tree and implements the tree bisection-reconnection (TBR) branch-swapping algorithm to identify the final ML tree. Support for internal nodes in the trees was obtained via parametric bootstrapping with 1000 replicates.

Evolutionary Rate Estimation and Analysis

A co-estimate of nucleotide substitution model parameters, phylogeny and time to the most recent common ancestor (tMRCA) was obtained using the Bayesian Markov chain Monte Carlo (MCMC) method implemented in BEAST v1.4.8 [26]. The approximate marginal likelihoods were calculated for six coalescent demographic models. These included both parametric (constant population size, exponential growth) and non-parametric models (Bayesian skyline plot) with both a strict and relaxed (uncorrelated LogNormal prior) molecular clock. All analyses were performed using the best fitting model of nucleotide substitution as determined by Modeltest v3.7 [24].

For each demographic model, two independent runs of length 1.0×10^8 steps in the Markov chain were performed using BEAST and checked for convergence using Tracer v1.4 [27]. Samples of trees and parameter estimates were collected every 10,000 steps to build a posterior distribution of parameters. The Estimated Sample Sizes (ESS's) for each run were > 200, indicating sufficient mixing of the Markov chain and parameter sampling. When similar results were produced from the independent runs of the Markov chain, the log files were combined with the program LogCombiner v1.4.7 available in the BEAST package [26].

The program TreeStat v1.1 [28] was used to calculate the proportion of lineages that existed over five-year intervals, between 1980-85, 1985-90, and 1990-95. This methodology utilized the posterior distribution of trees obtained previously in the Bayesian molecular clock analyses (as described by de Oliveira and colleagues) [29].

Model comparison

Model comparison in a Bayesian framework was achieved by calculating a measure known as the Bayes Factor (BF), which is the ratio of the marginal likelihoods of the two models being compared [30, 31]. This flexible method enables the comparison of non-nested models (such as the non-parametric BSP vs. parametric constant or exponential demographic models) that cannot validly be compared using mean log posterior probabilities.

RESULTS

All 177 Zimbabwean HIV-1 pol sequences generated from samples collected between 1991-2006 were identified as subtype C with no evidence for inter-subtype recombination, reflecting the predominance of this subtype in the HIV-1 epidemic in Harare. A maximum likelihood (ML) phylogenetic tree constructed from the Zimbabwe sequences demonstrated a clear relationship between sampling year and phylogenetic distance (branch length). For example, the earliest 1991 sequences were closest to and the recent 2006 sequences were most divergent from the most recent common ancestor (MRCA) node (**Figure 1A**).

To place the Zimbabwean sequences in the broader context of the subtype C pandemic and identify cross-epidemic relationships that could suggest common geographic origins, another ML tree was constructed with the Zimbabwean sequences and an additional 981 publicly available subtype C pol reference sequences isolated from several other HIV-endemic countries (**Figure 2**). Most Zimbabwean sequences (158/177) were highly intermingled with sequences from neighboring African nations of South Africa, Botswana, Zambia, Malawi, and Mozambique;

very few Zimbabwe sequences (19/177) clustered with subtype C sequences isolated from individuals sampled in non-adjacent African countries of Tanzania and Somalia, or from Sweden, Denmark, Yemen, and India.

Evolutionary Rate and Origin of the Zimbabwean Epidemic

To identify the time of the most recent common ancestor (tMRCA) and test hypotheses concerning the initial entry of HIV-1C into the Zimbabwean population, time-resolved phylogenetic trees were constructed under a Bayesian coalescent framework with BEAST (**Figure 3**). BEAST consensus trees showed evidence for multiple independent introductions of subtype C HIV-1 into Zimbabwe between 1979-1981. Notably, since the Zimbabwean sequences were scattered among other sequences from neighboring countries (**Figure 2**), BEAST trees taken in this context support a scenario of multiple cross-border transmissions of HIV-1C followed by subsequent spread.

The estimated nucleotide substitution rates and tMRCA dates for the Zimbabwe pol sequences obtained under the six different evolutionary models of population growth are presented in **Table 2**. All parameter estimates were highly consistent across the different evolutionary models, and replicate runs of the same model produced almost-identical results. The mean rate of 2.33×10^{-3} nucleotide substitutions / site / year (highest posterior density, HPD: 1.91×10^{-3} to 2.75×10^{-3}) produced an average estimate of the date of origin of the Zimbabwean epidemic in the year 1972 (HPD: 1969-1974:). The median estimates of the coefficient of variation parameter were 0.43, 0.24 and 0.25 for the constant, exponential and Bayesian skyline relaxed clock analyses respectively, indicating relatively little variation in evolutionary rates among branches in the tree irrespective of the evolutionary model employed. Estimates of the Bayes Factor (BF; see Methods – Model comparison) for the HIV-1C pol dataset supported models enforcing a relaxed clock over a strict clock and population growth models over a constant population size model. In turn, a relaxed clock exponential growth parametric model was statistically supported over the non-parametric BSP model (Supplementary Tables S1 and S2).

Specification of a Bayesian skyline plot (BSP) coalescent tree prior enables the estimation of effective population size (Ne) through time directly from sequence data. Our reconstruction of the demographic history of HIV-1C in Zimbabwe through the BSP analysis identified three epidemic growth phases: an initial, slow growth phase in 1974-1976, followed by an exponential growth phase in 1979-1984, and an asymptotic phase approaching the present time (**Figure 1C**). The most rapid increase in the curve occurred during 1979-1981, reflecting a logarithmic expansion in Ne, or effective number of infections over this short time period. Estimates indicated an initial median value of 8 effective infections (95% HPD 3 to 18) around 1972 while final estimates in the year 2006 were approximately 20,115 effective infections (95% HPD 8,334 to 61,200). Sequence diversity was estimated within each cohort by calculating average pairwise nucleotide distances using the HKY + G model of nucleotide substitution in Phylip v3.6 implemented in BioEdit. Consistent with the results of the BSP model, mean sequence diversity significantly increased from 1991-1998 ($p < 0.0001$), from 2001-2006 ($p < 0.0001$), and over the entire 15-year period ($p < 0.0001$), calculated using one-way ANOVA (**Figure 1B**).

In addition to estimating changes in Ne through time using a BSP, we also calculated the proportion of current lineages that existed during 1980-85, 1985-90, and 1990-95 (see Methods –

Evolutionary Rate Estimation and Analysis). The results showed that for 4 of the 6 evolutionary models (expo-strict clock, expo-relaxed clock, BSP-strict clock and BSP-relaxed clock) approximately 80% of the lineages were already present in Zimbabwe by 1985 (ranging from 55.4% to 98.3%; **Figure 3**).

DISCUSSION

The persistence and rapid increase of HIV-1C infection in much of southern Africa has been attributed largely to heterosexual transmission. This is certainly true of Zimbabwe, where continued heterosexual transmission underlies a generalized HIV-1 epidemic with relatively high population prevalence, particularly among young women. In the present investigation, phylogenetic analysis of pol sequences from cohorts of young women sampled from 1991-2006 confirms the predominance of subtype C infection in the Zimbabwean HIV-1 epidemic. Our demographic and evolutionary reconstruction of the Zimbabwean epidemic suggests that multiple closely-related subtype C viruses with a common ancestor originating in the early 1970's entered the country in the early 1980's, followed by an explosive growth in effective number of infections over the next decade.

The origin and timing of this epidemic expansion may be partly explained by the political and military history of the region. From 1953-63, migration of populations in Southern Africa was facilitated by a pre-existing colonial infrastructure in which Zimbabwe (Southern Rhodesia), Zambia (Northern Rhodesia), and Malawi (Nyasaland) were politically and economically merged as the Central African Federation. In 1965, in response to the end of colonial rule and the independence of Northern Rhodesia (Zambia), the Southern Rhodesian government issued a unilateral declaration of independence to maintain minority rule and denial of majority rights. This led to a prolonged civil conflict throughout the 1970's between the self-declared Rhodesian government and Black Nationalist liberation groups, during which movement in and out of Rhodesia was severely constrained by international sanctions and government restrictions. The conflict in Zimbabwe ended with the Lancaster House Agreement in December 1979, followed by the return of exiled liberation forces. Thus in 1980, tens of thousands of expatriates and liberation fighters returned from adjoining southern African countries to a newly independent Zimbabwe.

Our phylogenetic and evolutionary analyses of HIV-1C in the region reflect these historical events. The clustering of nearly 90% of Zimbabwean sequences with sequences from adjacent southern African countries (**Figure 2**) supports the regional origin and localized expansion of the subtype C epidemic in Zimbabwe. The subsequent rapid increase of HIV-1C following Zimbabwean independence in the early 1980's is consistent with our retrospective reconstruction and Bayesian estimates of HIV-1 growth, with 98% of the current lineages present in Zimbabwe by 1990 (**Figure 3**). This exponential growth of the Zimbabwean epidemic over a period of less than a decade in the 1980's provides an example of in-migration of a small number of ancestors (founders) and subsequent amplification during a period of heightened political and demographic change.

Epidemic expansion in Zimbabwe in the 1980's, as we have estimated by calculating effective number of infections in a coalescent framework, is supported by three independent sources

estimating historical HIV prevalence: blood donor screening, antenatal surveillance, and back-calculation of incidence from mortality statistics. The first clinical case of AIDS in Zimbabwe was documented in 1985, the same year that diagnostic screening for donated blood units was initiated by the Zimbabwe National Blood Transfusion Service (NBTS). Our estimated increase in HIV-1C prevalence corresponds well with NBTS records, which documented a near doubling in seroprevalence among blood donors each year from 1986-1990 [32]. Our estimates also mirror WHO sentinel surveillance data, which document a rapidly increasing prevalence among antenatal women and the general population through the early 1990's [33]. Notably, our demographic estimates indicate a peak in number of infections between 1989-1991. Epidemiologic modeling studies back-calculating HIV incidence from mortality statistics estimate a likely peak in incidence in Harare during the same period between 1988 and 1990 [34]. These independent approaches paint a remarkably similar picture, each identifying a period of rapid, peaking expansion of HIV-1C in Zimbabwe in the 1980's.

Analysis of population demographics using Bayesian coalescent methods has a number of limitations. The presence of recent (slightly) deleterious mutations, which have not yet been eliminated at the population level by purifying selection, would result in an overestimation of the time to the most recent coalescence event in the tree [35]. The present analyses may also be limited by an exclusive focus on HIV isolated from cohorts of pregnant women, where infection was identified through antenatal screening. In sub-Saharan Africa young women comprise a demographic group with high risk of infection in association with patterns of sequential overlapping partnerships, intergenerational sex, and low condom usage [36]. More recent studies and surveillance data from Zimbabwe have provided encouraging evidence that gradual behavioral changes have reduced the prevalence among 18-24-year-old women from more than 25% to about 20%, suggesting that incidence in this vulnerable group has declined since 1998 [33, 37]. This is consistent with our Bayesian estimates of a steady-state prevalence as the number of effective infections reaches an asymptotic phase approaching the present time (**Figure 1C**). However, it has been noted that all Bayesian skyline plots show some signal of steady-state dynamics approaching the present that may be partly attributed to within-host evolution [35]. Moreover, like recent surveillance studies our findings cannot rule out the possibility that the observed decrease in HIV-1 prevalence may be due, in part, to selective AIDS-induced mortality rates, especially given the disruptions in health infrastructure introduced by political volatility in Zimbabwe.

Large-scale migration in response to political and economic instability is just as evident in Zimbabwe today. The recent cholera epidemic, spiraling inflation and political turmoil are limiting disease prevention services and treatment and escalating out-migration to neighboring countries. The political and economic displacement of millions of individuals from Zimbabwe poses further challenges to regional programs operating in the context of a generalized HIV-1 epidemic in southern Africa. These programs urgently need to expand testing, prevention and treatment services for HIV and other infectious diseases to reach migratory and displaced populations in a rapidly changing political and economic environment.

Table 1. Samples From Women Testing Positive for HIV-1 in Antenatal Clinics in Zimbabwe, 1991-2006.

Cohort	WHO	SIDA	HPTN 023	NIH
Collection Date	1991	1998	2001	2006
N samples	39	56	26	56
ARV Exposure History	None	None	None	None
Study and Sponsor	"Natural History of MTCT" WHO	"Feasibility of SC AZT in pMTCT" SIDA	"HIVNET 023 Phase I study of SD NVP" NIH	"Drug Resistance and Pathogenesis in Subtype C HIV-1" NIH

WHO, World Health Organization; SIDA, Swedish International Development Cooperation Agency; HPTN, HIV Prevention Trials Network; NIH, National Institutes of Health; pMTCT, prevention of mother-to-child transmission; SC AZT, short-course zidovudine; SD NVP, single dose nevirapine.
References for cohorts and genotype data: WHO [18], SIDA [16], HPTN 023 [17], NIH [19].

Table 2. Bayesian estimates of mean time to the most common ancestor (tMRCA), mean nucleotide substitution rates, and percent lineages at selected time periods for HIV-1C pol in Zimbabwe.

	Expo Strict Clock	Expo Relax Clock	BSP Strict Clock	BSP Relax Clock
MRCA	1973.638	1974.078	1971.831	1972.035
95% HPD lower	1977.702	1978.246	1977.487	1978.175
95% HPD upper	1969.578	1969.371	1965.799	1965.062
Mut. Rate	0.002068	0.00206	0.002186	0.002192
95% HPD lower	0.001729	0.001664	0.001834	0.001791
95% HPD upper	0.00241	0.002456	0.002555	0.002597
% Strains in 1975	1.861	1.861	1.406	1.406
95% HPD lower	0.565	0.565	0.565	0.565
95% HPD upper	5.65	5.65	2.26	2.26
% Strains in 1980	20.7	20.7	10.1	10.1
95% HPD lower	1.695	1.695	1.13	1.13
95% HPD upper	47.5	47.5	31.6	31.6
% Strains in 1985	82.9	82.9	81.3	81.3
95% HPD lower	63.8	63.8	55.4	55.4
95% HPD upper	96.6	96.6	98.3	98.3
% Strains in 1990	98.5	98.5	97.8	97.8
95% HPD lower	97.7	97.7	96	96
95% HPD upper	99.4	99.4	98.3	98.3

The table contains the mean time to the most recent common ancestor (tMRCA), the estimated mean mutation rate (substitutions/site/year), and the percentages of current strains estimated to exist in 1975, 1980, 1985 and 1990 for clade HIV-1C in Zimbabwe. 95% lower and upper highest posterior density (HPD) intervals are shown. MRCA and mutation rate were calculated using BEAST under exponential (Expo) and Bayesian skyline plot (BSP) demographic growth models assuming both a strict and relaxed molecular clock. Lineage estimations were calculated from the posterior distribution of Bayesian trees using TreeStat.

Supplementary Table S1. Estimates of the coefficient of variation, mean nucleotide substitution rates, and mean time to the most common ancestor (tMRCA) for the HIV-1 subtype C pol sequences.

Evolutionary Model	Coefficient of variation	*hky.kappa*	*u*	*tMRCA*
Constant (Strict clock)	n/a	13.771 (12.128-15.593)	0.00264 (0.002171-0.00314)	1971 (1976-1964)
Constant (Relaxed clock)	0.442 (0.308-0.589)	13.943 (12.194-15.787)	0.00281 (0.002272-0.00335)	1969 (1978-1958)
Exponential (Strict clock)	n/a	13.948 (12.216-15.684)	0.00207 (0.001729-0.00241)	1973 (1977-1969)
Exponential (Relaxed clock)	0.251 (0.169-0.328)	14.098 (12.335-15.823)	0.00206 (0.001664-0.00246)	1974 (1978-1969)
BSP (Strict clock)	n/a	14.051 (12.339-15.854)	0.00219 (0.001834-0.00256)	1972 (1977-1965)
BSP (Relaxed clock)	0.24 (0.165-0.319)	14.180 (12.383-15.939)	0.00219 (0.001791-0.00260)	1972 (1978-1965)

Mean mutation rate *u* is given as substitutions/site/year. 95% HPD is shown in parenthesis.

Supplementary Table S2. Comparisons of Bayes Factors between different evolutionary models for HIV-1C in Zimbabwe.

Model Comparison	log10 Bayes Factor	Evidence against Ho
Const Strict (H0) vs Relaxed (H1) clock	27.732	Very Strong
Expo Strict (H0) vs Relaxed (H1) clock	19.496	Very Strong
BSP Strict (H0) vs Relaxed (H1) clock	17.736	Very Strong
Const (H0) vs Expo (H1) Relaxed clock	0.86	Positive
Const (H0) vs BSP (H1) Relaxed clock	3.694	Very Strong
BSP (H0) vs Expo (H1) Relaxed clock	4.554	Very Strong

BF = Bayes Factor is the difference (in log space) of the marginal likelihood of the null (H0) and alternative (H1) model. BFs were estimated by comparing the approximate marginal likelihoods of the different models given in Table S1.
Const = constant population size; Expo = exponential population growth; BSP = Bayesian skyline plot; Strict = strict molecular clock; Relaxed = relaxed molecular clock.

FIGURE 1A

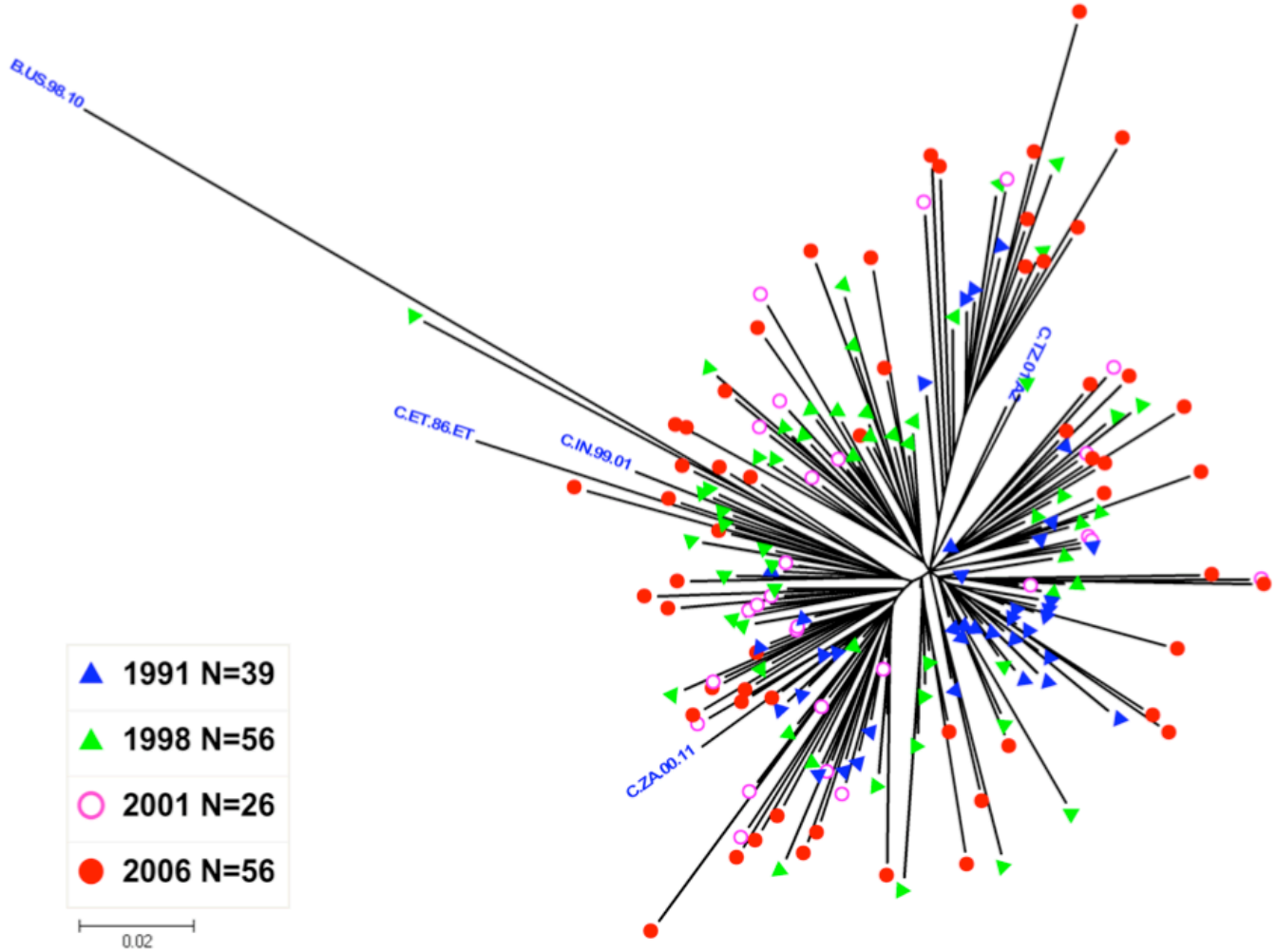


Figure 1A. Midpoint-rooted maximum-likelihood tree of 177 HIV-1 subtype C *pol* sequences sampled from young women in Zimbabwe. The tree was constructed under the GTR + I + G model of evolution using sequences sampled in Harare over a 15-year period between 1991-2006. Branches are color-tagged by sampling year and subtype reference strains from the Los Alamos HIV Sequence Database (www.hiv.lanl.gov) are labeled with blue text

FIGURE 1B

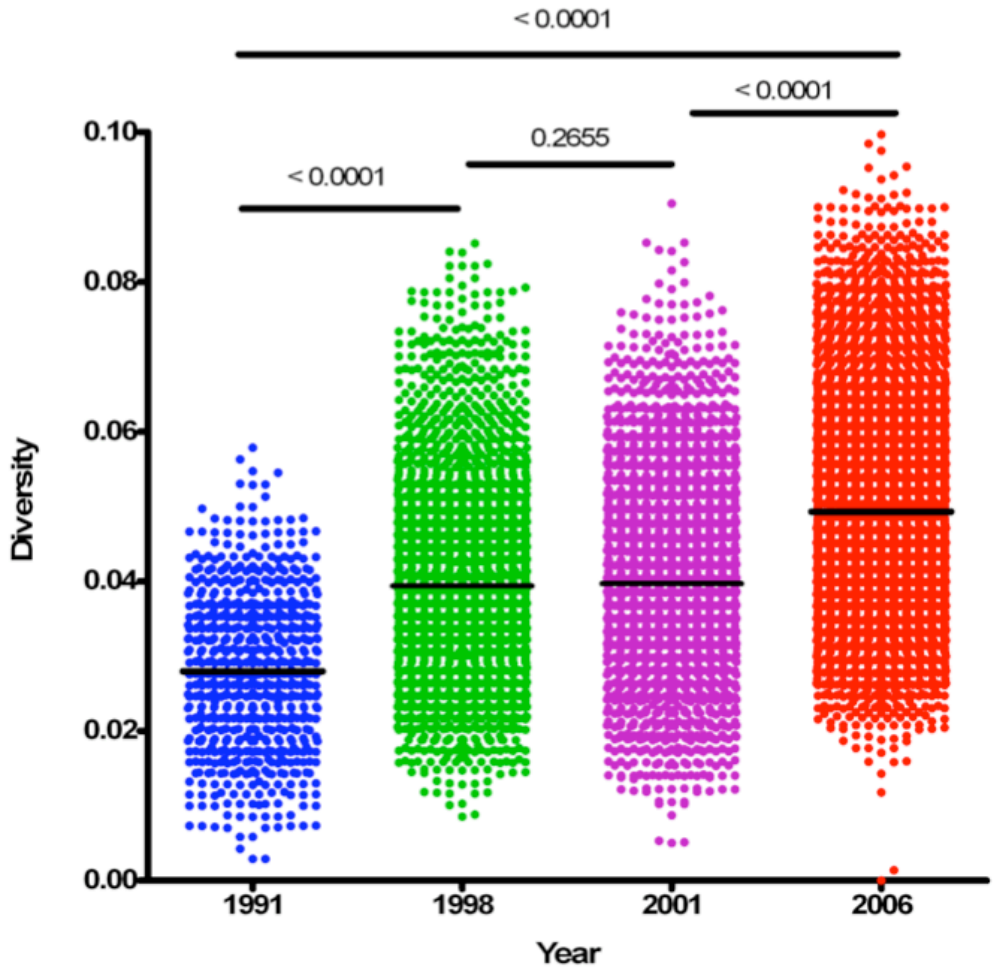


Figure 1B. Scatterplots of sequence diversity for successive cohorts of HIV-1 infected women in Zimbabwe. Sequence diversity within each cohort was calculated from average pairwise nucleotide distances using the HKY + G model of nucleotide substitution. Three time periods demonstrated a significant increase in sequence diversity: 1991-1998, 2001-2006, and 1991-2006 ($p<0.0001$, one-way ANOVA).

FIGURE 1C

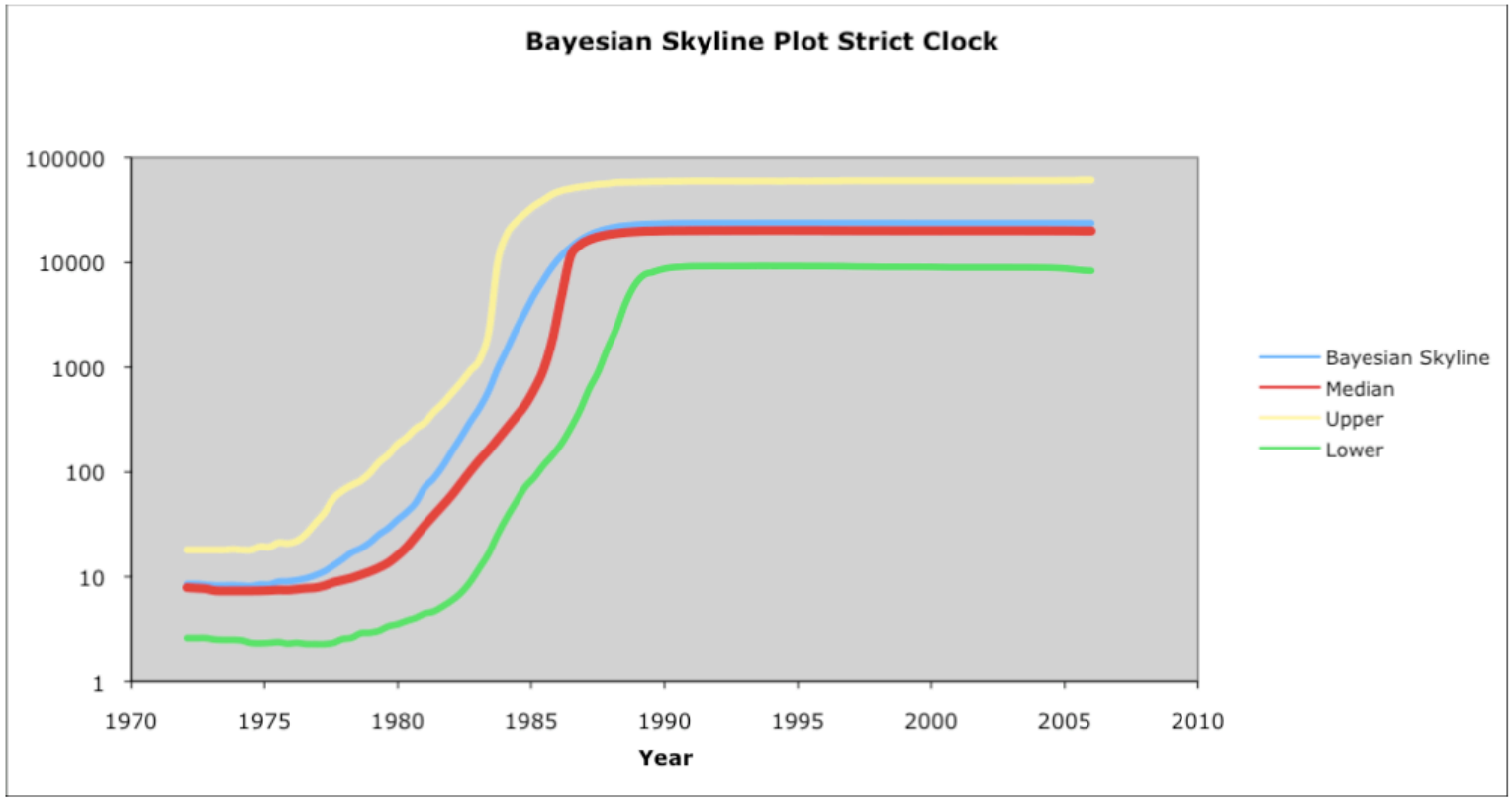


Figure 1C. Bayesian skyline plot of HIV-1C *pol* demographic growth patterns in Zimbabwe. The X-axis represents year and the Y-axis represents HIV-1 effective population size (effective number of infections, Ne; $\log_{10}$ scale). The red line represents the median estimate for Ne, and yellow and green lines represent the upper and lower 95% highest posterior density (HPD) estimates of Ne, respectively.

FIGURE 2

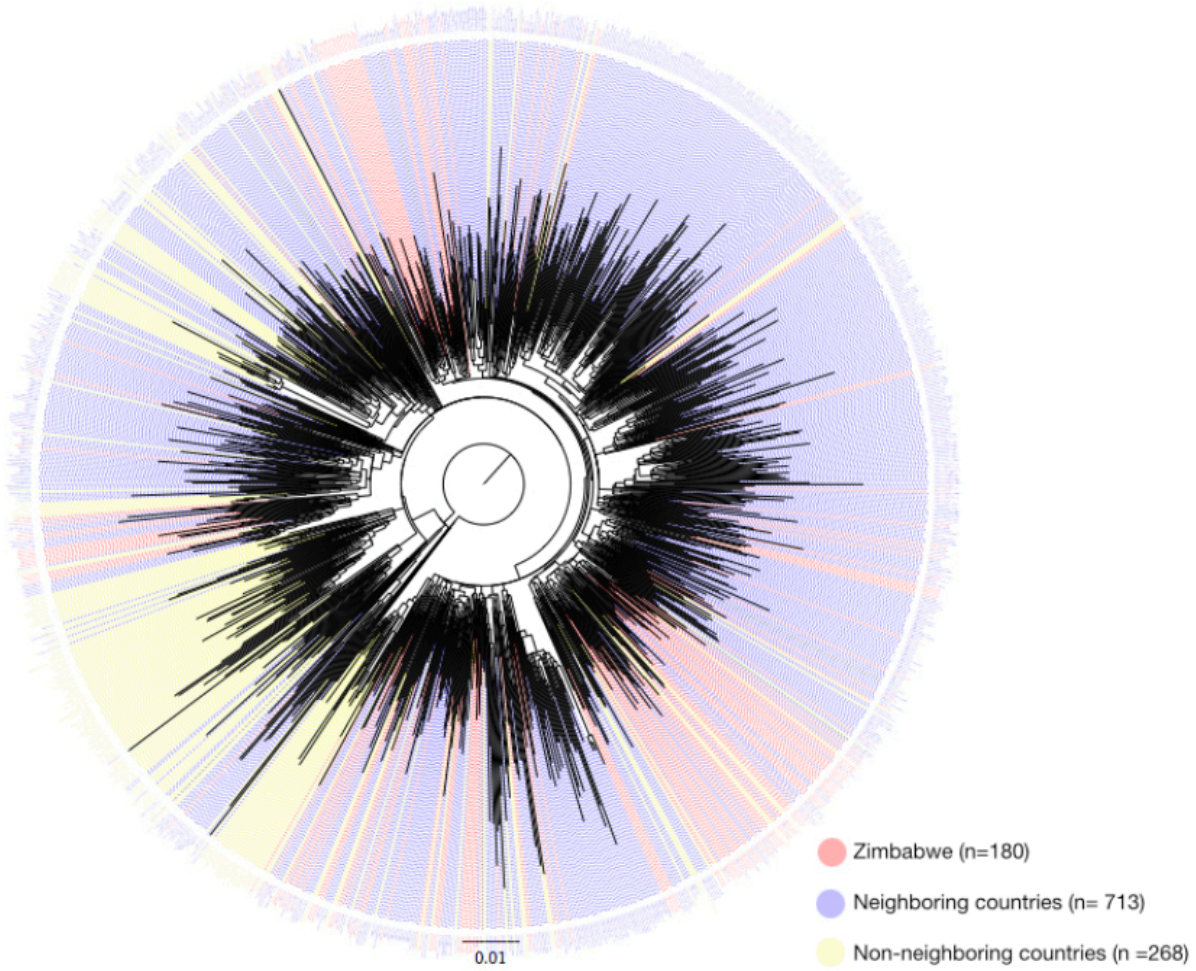


Figure 2. Maximum-likelihood tree of HIV-1 subtype C sequences. The tree contains 1,161 HIV-1C *pol* sequences isolated from Zimbabwe (n=180); Neighboring countries (n=713) including sequences from Botswana, Mozambique, Malawi, South Africa and Zambia; and Non-neighboring countries (n=268) including African, Asian, European and South American sequences. The complete alignment and list of strains is provided in supplementary information. The tree was constructed under the GTR + I + G model of evolution using PhyML.

FIGURE 3

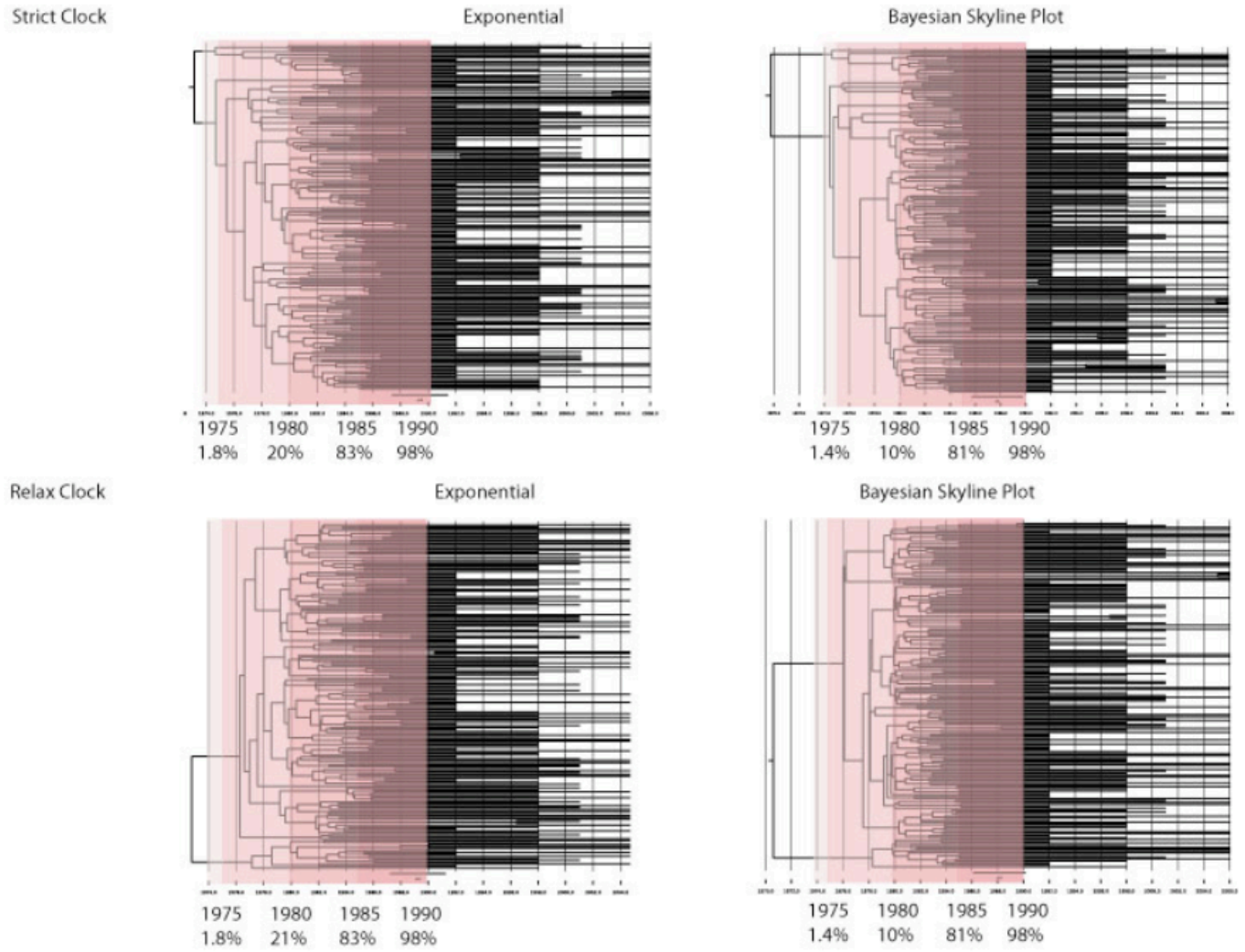


Figure 3. Bayesian maximum clade-credibility trees for HIV-1C *pol* in Zimbabwe. The consensus trees were constructed with BEAST from a posterior distribution of 10.000 Bayesian trees under Exponential and Bayesian Skyline Plot models, enforcing either a strict or relaxed molecular clock. The estimated percentage of strains at 1975, 1980, 1985 and 1990 are indicated in the trees, with red highlighting marking this time interval.

REFERENCES

1. Hemelaar, J., et al., Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS, 2006. 20(16): p. W13-23.
2. Lin, Y.T., et al., Molecular epidemiology of HIV-1 infection and full-length genomic analysis of circulating recombinant form 07_BC strains from injection drug users in Taiwan. J Infect Dis, 2007. 195(9): p. 1283-93.
3. Su, L., et al., Characterization of a virtually full-length human immunodeficiency virus type 1 genome of a prevalent intersubtype (C/B') recombinant strain in China. J Virol, 2000. 74(23): p. 11367-76.
4. Watanaveeradej, V., et al., Subtype C/CRF01_AE recombinant HIV-1 found in Thailand. AIDS, 2003. 17(14): p. 2138-40.
5. Rambaut, A., et al., The causes and consequences of HIV evolution. Nat Rev Genet, 2004. 5(1): p. 52-61.
6. Halperin, D.T. and R.C. Bailey, Male circumcision and HIV infection: 10 years and counting. Lancet, 1999. 354(9192): p. 1813-5.
7. Cohen, M.S., Preventing sexual transmission of HIV. Clin Infect Dis, 2007. 45 Suppl 4: p. S287-92.
8. Montano, M.A., et al., Divergent transcriptional regulation among expanding human immunodeficiency virus type 1 subtypes. J Virol, 1997. 71(11): p. 8657-65.
9. Eshleman, S.H., et al., Nevirapine (NVP) resistance in women with HIV-1 subtype C, compared with subtypes A and D, after the administration of single-dose NVP. J Infect Dis, 2005. 192(1): p. 30-6.
10. Kassaye, S., et al., Drug resistance in plasma and breast milk after single-dose nevirapine in subtype C HIV type 1: population and clonal sequence analysis. AIDS Research and Human Retroviruses, 2007.
11. Ball, S.C., et al., Comparing the ex vivo fitness of CCR5-tropic human immunodeficiency virus type 1 isolates of subtypes B and C. J Virol, 2003. 77(2): p. 1021-38.
12. Abraha, A., et al., CCR5- and CXCR4-tropic subtype C HIV-1 isolates have lower pathogenic fitness as compared to the other dominant group M subtypes: Implications for the epidemic. J Virol, 2009.
13. Gifford, R.J., et al., Phylogenetic surveillance of viral genetic diversity and the evolving molecular epidemiology of human immunodeficiency virus type 1. J Virol, 2007. 81(23): p. 13050-6.
14. Korber, B., et al., Timing the ancestor of the HIV-1 pandemic strains. Science, 2000. 288(5472): p. 1789-96.
15. Salemi, M., et al., High-Resolution Molecular Epidemiology and Evolutionary History of HIV-1 Subtypes in Albania. PLoS ONE, 2008. 3(1): p. e1390.
16. Mateta, P., et al., Feasibility of preventing mother-to-child transmission of HIV in Zimbabwe [B10499]. in XIV International AIDS Conference. 2002: Barcelona, Spain.
17. Shetty, A.K., et al., Safety and trough concentrations of nevirapine prophylaxis given daily, twice weekly, or weekly in breast-feeding infants from birth to 6 months. J Acquir Immune Defic Syndr, 2003. 34(5): p. 482-90.

18. Katzenstein, D.A., et al., Serum level of maternal human immunodeficiency virus (HIV) RNA, infant mortality, and vertical transmission of HIV in Zimbabwe. J Infect Dis, 1999. 179(6): p. 1382-7.
19. Manasa, J., et al., HIV-1 Subtype C Drug Resistance Surveillance Among Young Treatment Naive Pregnant Women in Chitungwiza, Zimbabwe (2006-2007). in XVII International AIDS Conference. 2008: Mexico City, Mexico.
20. Thompson, J.D., D.G. Higgins, and T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res, 1994. 22(22): p. 4673-80.
21. Hall, T., BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser., 1999. 41: p. 95-98.
22. de Oliveira, T., et al., An automated genotyping system for analysis of HIV-1 and other microbial sequences. Bioinformatics, 2005. 21(19): p. 3797-800.
23. Ray, S., Simplot v3.5.1. Available from http://sray.med.som.jhmi.edu/SCRoftware/simplot/.
24. Posada, D. and K.A. Crandall, MODELTEST: testing the model of DNA substitution. Bioinformatics, 1998. 14(9): p. 817-8.
25. Swofford, D.L., PAUP: Phylogenetic Analysis Using Parsimony (*and Other Methods). 1998, Sinauer Associates, Inc. Publishers, Sunderland, MA.
26. Drummond, A.J. and A. Rambaut, BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol, 2007. 7: p. 214.
27. Rambaut, A., Drummond AJ., Tracer v1.4, Available at http://beast.bio.ed.ac.uk/Tracer. 2007.
28. Rambaut, A. and A.J. Drummond, TreeStat v1.1: Tree statistic calculation tool, Available at http://tree.bio.ed.ac.uk/software/treestat/. 2005.
29. De Oliveira, T., et al., Molecular epidemiology: HIV-1 and HCV sequences from Libyan outbreak. Nature, 2006. 444(7121): p. 836-7.
30. Kass, R., Raftery, A., Bayes Factors. Journal of the American Statistical Association, 1995. 90(430): p. 773-795.
31. Suchard, M.A., R.E. Weiss, and J.S. Sinsheimer, Bayesian selection of continuous-time Markov chain evolutionary models. Mol Biol Evol, 2001. 18(6): p. 1001-13.
32. Zimbabwe National Blood Transfusion Service, H., Annual Reports. 1985-1990.
33. UNAIDS, Report on the global AIDS epidemic. 2008.
34. Lopman, B. and S. Gregson, When did HIV incidence peak in Harare, Zimbabwe? Back-calculation from mortality statistics. PLoS ONE, 2008. 3(3): p. e1711.
35. Lemey, P., A. Rambaut, and O.G. Pybus, HIV evolutionary dynamics within and among hosts. AIDS Rev, 2006. 8(3): p. 125-40.
36. Hallett, T.B., et al., Declines in HIV prevalence can be associated with changing sexual behaviour in Uganda, urban Kenya, Zimbabwe, and urban Haiti. Sex Transm Infect, 2006. 82(suppl_1): p. i1-8.
37. Gregson, S., et al., HIV decline associated with behavior change in eastern Zimbabwe. Science, 2006. 311(5761): p. 664-6.

Chapter 2: Molecular Epidemiology and Antiretroviral Drug Resistance in Subtype B HIV-1 in Northern California

INTRODUCTION

Molecular epidemiologic analyses of the HIV-1 epidemics in Africa [3], Asia [4, 5], and North America[6, 7], many employing sophisticated computational and phylogenetic approaches, have provided evidence for distinct epidemic dynamics and patterns of transmission in defined communities. While the southern African and Asian epidemics are thought to be sustained by sexual-social factors, high-risk injecting drug use (IDU), and commercial sex work (CSW), male-male sex has remained the primary mode of transmission in the US, accounting for over half of new infections [8]. In recent years, the HIV-1 epidemic in California has shifted from a primarily white MSM (men who have sex with men) population to a diverse range of overlapping risk groups, where heterosexual women now comprise the fastest growing demographic for new infections and an increasing fraction of new diagnoses are documented among underrepresented minorities [9, 10]. Hispanic monolingual men and women now predominate among new HIV-1 cases in several California counties followed by African American and Asian/Pacific Islander populations [8, 11].

The HIV-1 epidemic in San Mateo County is sustained by multiple ethnic, migratory, and behavioral networks, including MSM; migrant populations from Africa, Asia, and Latin America; and IDU, each having distinct patterns of HIV acquisition and transmission. Within these communities access to care, including ART, has historically been constrained by financial, cultural, linguistic, and community barriers. Recent studies from San Mateo County have shown that immigrants (78.7% Hispanic) attending the publicly-funded County AIDS Program have substantially delayed clinical presentation of HIV, marked by lower CD4 cell counts, greater prevalence of opportunistic infections, and higher hospitalization rates as compared with US-born individuals [12]. Population-based surveys in San Mateo have also recently identified multiple risk behaviors associated with immigrant status, including unstable, overlapping sexual partnerships, CSW contacts, intravenous drug use, and unprotected sex [13]. The implications of prolonged, undiagnosed HIV infection, viremia, delayed treatment, and expanded risk behavior for community transmission networks remain largely uncharacterized.

HIV-1 is constantly changing in response to drug and immune pressures through evolutionary mechanisms, including point mutations, insertions, deletions, reduplications, and recombination [14-17]. These mechanisms operate within the virus quasi-species of individuals [18], between individuals as adaptation to divergent immune pressures [19], and across populations and risk groups where recombination between divergent viruses and viral substrains may occur [20]. ART, the mainstay of patient care, may be compromised by the selection and evolution of drug resistance, the consequence of mutations in critical loci of viral proteins targeted by antiretroviral agents. The presence of certain mutations is associated with diminished virologic response to specific ART drugs and combinations, leaving patients who develop drug resistance with fewer treatment options.

There is increasing evidence that drug-exposed individuals may contribute substantially to new transmission and perpetuation of HIV epidemics. Clinical experience in the US and Europe has shown that up to 50% of newly identified, drug naïve patients fail the initial triple therapies used

in clinical practice, providing evidence of increasing community transmission of multiple drug-resistant viruses [21, 22]. This has been confirmed in more recent studies in California, where a high prevalence of drug resistance among untreated, newly diagnosed patients has demonstrated widespread community-level transmission and cross-border introduction of multi-drug resistant HIV-1 associated with an expansion in migration and risk behaviors [23, 24]. In recent years a focus on addressing health disparities has increased availability of AIDS treatment and care for historically marginalized populations in California. However, there are very limited data on the virologic and clinical consequences of treatment in these communities. These observations underscore the need for broad and systematic surveillance of HIV transmission networks and drug resistance in the context of comprehensive health service delivery systems, particularly among populations who present late and where ARV treatment access is constrained.

Molecular and phylogenetic methods are powerful approaches for characterizing the distribution and transmission of infectious diseases and have great potential for enhancing HIV-1 monitoring, surveillance, and prevention efforts. The extensive genetic variability of HIV-1 enables phylogenetic reconstruction of transmission events, including sexual, IDU-related, and healthcare-related transmission [25-27], based on the unique molecular signatures of HIV sequences within and between individuals. Analysis of HIV sequence data sampled over time and geographic space also allows inference of population-level processes such as viral population size, epidemic growth rate, and migratory behavior [28, 29]. Despite the genetic conservation of the HIV-1 pol gene, a number of studies have demonstrated its suitability for phylogenetic inference of transmission clusters and patterns in the context of epidemiologic investigation [30, 31]. Indeed, the high frequency of HIV-1 pol gene sequencing throughout the world, heralded by the increasingly prominent role of drug resistance monitoring as a standard of clinical care, presents valuable and unique opportunities to characterize HIV transmission clusters based on demographic, behavioral, and geospatial parameters; the origin and introduction of novel HIV subtypes; the transmission and distribution of multi-drug resistance; and the evolutionary selection of drug resistance mutations in treated populations.

Defining the patterns of HIV transmission and resistance in communities is important for regional prevention and treatment programs to develop effective, integrated testing and treatment strategies to reduce transmission and to identify and appropriately treat newly infected individuals. This study combines phylogenetic and molecular virology/epidemiology approaches to characterize HIV viral transmission and ARV drug resistance in the San Mateo epidemic. These studies are undertaken to test hypotheses about localized HIV transmission, as a prelude to more extensive analyses of transmission, risk groups, and migratory populations in northern California.

MATERIALS AND METHODS

Study Population, Data and Sequences

HIV-positive adults receiving antiretroviral treatment and care as part of the publicly funded San Mateo County AIDS Program underwent clinically indicated genotypic antiretroviral resistance testing (GART) from 1996-2010 as part of routine HIV care. Resistance assays were performed at the Stanford University Hospital Diagnostic Virology Laboratory and/or Monogram Biosciences (South San Francisco, CA). This study population included both acute and chronic infections ranging from treatment naïve to multi-drug experienced individuals. Demographic (age, gender, race/ethnicity), epidemiologic (date of diagnosis, mode of transmission, social/risk behavior, partner information, location data), and clinical information (ARV treatment history, HIV clinical stage, HIV viral load, CD4+ cell count, co-infections) was de-identified and extracted from electronic and written medical records at San Mateo Medical Center. 637 HIV-1 pol sequences were obtained from 316 patients tested over the study period.

Mode of HIV transmission was extracted from the medical records and was determined by the physician of record at the time of patient intake. Transmission categories included 1) MSM, men who have sex with men; 2) IDU, injection drug use; 3) MSM + IDU; and 4) heterosexual/other (including participants reporting infection through contaminated blood products).

The use of anonymized, de-identified clinical/demographic and sequence data was reviewed and approved under an exempt protocol by the Institutional Review Boards of the University of California-Berkeley and Mills-Peninsula Health Services on behalf of San Mateo Medical Center.

Sequence Selection and Alignment

Sequences were aligned with ClustalW [32] and manually edited using BioEdit [33]. For individuals who had multiple sequences, the earliest sequence was utilized for transmission/clustering analyses and identification of drug resistance.

Subtype Classification and Drug Resistance Analysis

HIV-1 subtype and evidence for inter-subtype recombination were assessed using the REGA Subtyping Tool v2.0 (available at hivdb.stanford.edu). The HIVseq algorithm implemented in the Stanford HIV Drug Resistance Database [1] was used for genotypic resistance interpretation and to identify known ART drug resistance mutations [2].

Phylogenetic Analysis

A best-fitting nucleotide substitution model for the San Mateo subtype B pol sequences was estimated using hierarchical likelihood ratio tests (hLRT's) implemented in the program Modeltest v3.7 [34] and manual examination in PAUP v4.0 [35]. Maximum likelihood (ML) phylogenetic trees were constructed using the inferred model, GTR + I + G, with the program PhyML v2.4 [36]. This method employs a neighbor-joining (NJ) tree as a starting tree and implements the tree bisection-reconnection (TBR) branch-swapping algorithm to identify the

final ML tree. Support for internal nodes in the trees was obtained via parametric bootstrapping with 1000 replicates. In accordance with empirically-accepted values, bootstrap values above 975 were considered indicative of a statistically significant transmission cluster, and values between 750-974 were considered statistically probable. Medical records were independently reviewed to identify epidemiologic linkages among study participants irrespective of phylogenetic linkages present.

Statistical Analyses

Univariate analyses of risk factors associated with being in a cluster were performed using either chi-squared or Fisher's exact tests for categorical variables, and the Student's t-test for continuous variables. A multiple logistic regression model included variables significant in the univariate analyses and empirical confounders. All standard statistical analyses were performed in the statistical package R (cran.r-project.org).

RESULTS

Between 1996 and 2010, 620 GART samples were identified from 316 distinct outpatients receiving ART in the San Mateo County AIDS program. **Table 1** summarizes baseline demographic characteristics for 241 participants for whom full demographic information was available. Nearly 80% of participants were male, with ages ranging from 19 – 67 years old. The study population consisted mainly of White non-Hispanics (43.1%), Black or African Americans (28.0%), and White Hispanics (19.8%). The most frequent mode of HIV transmission was MSM (40.1%), followed by heterosexual and other modes of transmission (27.7%), IDU (17%), and combined MSM/IDU (15.3%). Three participants reported infection through contaminated blood products.

HIV Epidemiology and Transmission Clustering

Maximum-likelihood phylogenetic trees constructed from 316 individual patient HIV genotypes identified a total of 15 transmission clusters, 10 being statistically significant and 5 statistically probable (**Figures 1A and 1B**). Fourteen of the clusters were composed of 2 members and one of the clusters had 3 members. **Table 2** summarizes the characteristics of clustered patients. A review of medical records indicated that Cluster 1 was a husband/wife couple who were monolingual White Hispanics, Cluster 4 was a heterosexual White non-Hispanic couple with a history of IDU, and Cluster 12 was a White non-Hispanic heterosexual couple. No other seropositive partners of study participants were identified as they either did not attend the San Mateo AIDS Clinic or did not have any genotypic testing performed at Stanford University.

Most participants were infected with subtype B HIV-1 (96.2%). Samples from 6 participants (1.9%) infected with subtype C virus were not linked to an MRN and no further information was available. Of the 4 participants (1.3%) infected with the circulating recombinant form CRF01_AE, two self-identified as Asian and the remaining two had no further information available. Two participants (0.6%) infected with subtype A virus originated from Kenya (**Figure 1A**).

In univariate analysis, participants whose mode of HIV transmission was heterosexual or other ($p = 0.04$) or who were recently infected ($p = 0.02$) were more likely to be in a cluster (**Table 3**). Participants in clusters also tended to have a younger age (mean 37.4 years) compared to those who were not (mean 42.2 years, $p = 0.01$). MSM were significantly less likely to be in a cluster

($p = 0.02$). A borderline significant association was seen between having any nucleoside reverse transcriptase inhibitor (NRTI) resistance mutation and being in a cluster ($p = 0.05$). A multivariate logistic regression model did not converge as there were too few participants within each cluster.

HIV Drug Resistance

Of 305 participants for whom drug resistance data were available, 65% had at least one drug resistance mutation (**Table 4**). The most frequent NRTI resistance mutations were M184V (49.5%), T215Y (26.9%), and M41L (20.3%). The most frequent non-nucleoside reverse transcriptase inhibitor (NNRTI) mutations were K103N (15.4%), G190E (3%), and K101E (2.6%). The most frequent protease inhibitor (PI) mutations were at positions 46 (8.9%), 54 (7.2%) and 82 (26%).

Table 5 summarizes results of univariate and multivariate analyses for drug resistance mutations for 97 participants who had complete information on risk factors. Comparison of risk factors between those included and excluded from analysis revealed no differences between groups except a higher proportion of White Hispanics included in the model ($p = 0.03$). Risk factors associated with drug resistance included being male, having chronic HIV infection, being diagnosed with HIV at an earlier time period, having genotyping at an earlier time period, having lower viral load, and having exposure to either NRTIs or NNRTIs. After multivariate adjustment, factors independently associated with drug resistance of any class included being male, genotyping at an earlier time period, lower viral load, and exposure to NNRTIs.

DISCUSSION

Phylogenetic analyses of a convenience sample of HIV-1 *pol* sequences obtained as part of routine clinical care of outpatients in the San Mateo County AIDS Program demonstrated the existence of a small number of small transmission clusters, most often paired transmissions. Unadjusted risk factors associated with being in a cluster included younger age, heterosexual or other mode of transmission, and recent infection; MSM was associated with a lower likelihood of being in a cluster.

The high prevalence of drug resistance (65.3%) in this population most likely reflects the high burden of long-standing HIV disease (89.3% of participants) and ART exposure (89.2% exposed to any therapy), and highlights the utility of genotypic testing in determining safe initial regimens for newly-diagnosed cases. Those who were treatment-naïve had nearly four times the prevalence of mutations compared to similar studies [37-43], which may be attributable to the high proportion of recently-infected participants (33.3%) receiving therapy. Adjusted risk factors independently associated with drug resistance were being male, earlier time period of genotyping and earlier time period of diagnosis; the latter factors are likely related to long-standing infection as well as more limited and less potent ART available during earlier periods of the US HIV epidemic. Interestingly, HIV viral load was also lower among patients with drug resistance; this may reflect the lower overall viral fitness of drug resistant strains harboring deleterious mutations.

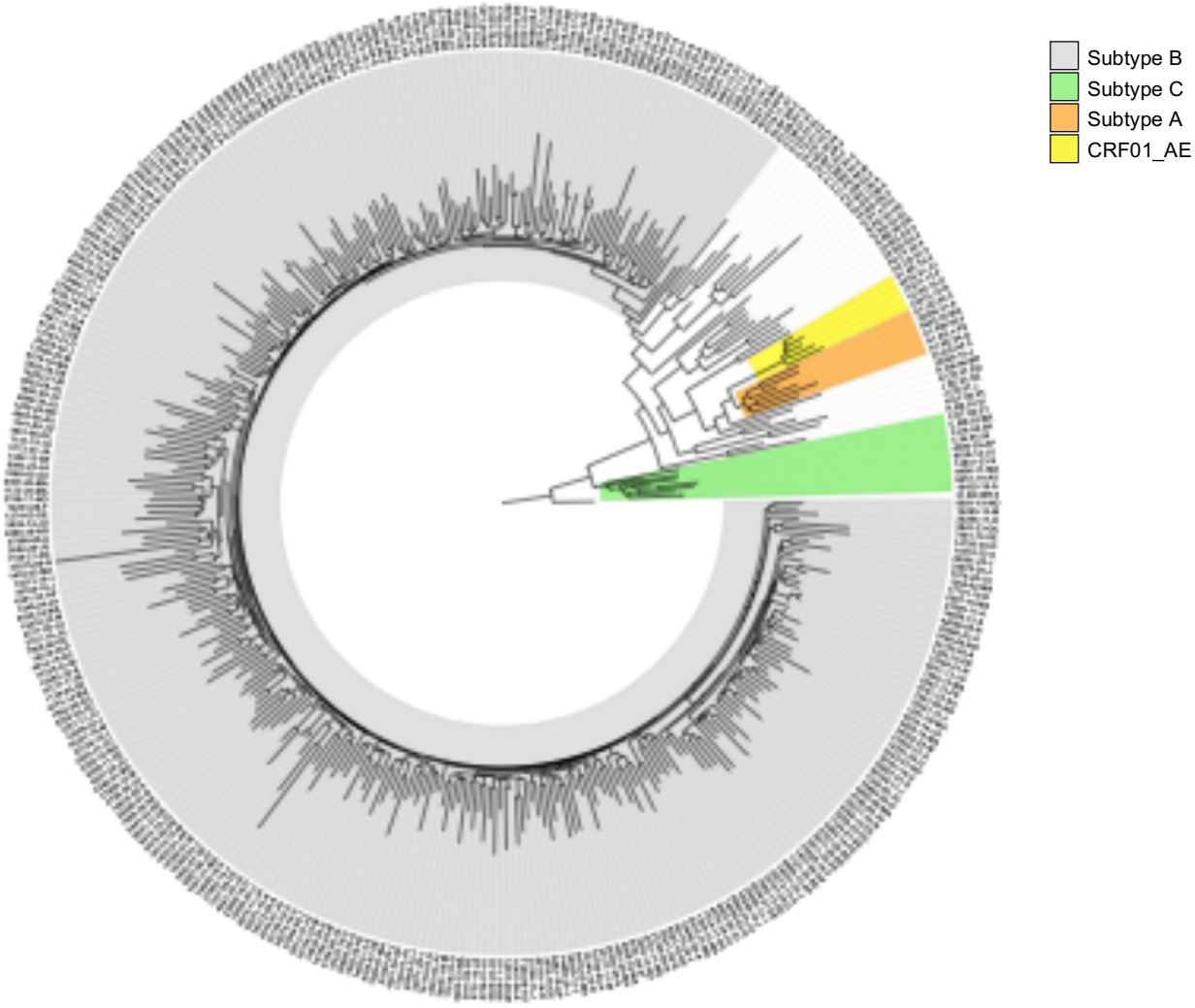
This study has a number of limitations and potential biases. Patients in the San Mateo County AIDS Program have a different racial and ethnic composition than the overall population in San Mateo County; notably, Black or African Americans accounted for 28.0% of the study

population while comprising only 3.1% of the County population, and Asians accounted for only 4.3% of the study population while they comprise 24% of the County population [44]. Patients also have lower income than the county median [45], highlighting the inequalities inherent in the US epidemic. There was also incomplete availability of epidemiological and clinical data in part due to destruction or remote storage of archived medical records, limiting the statistical power of clustering and drug resistance analyses.

Finally, the sampling frame is restricted. There is some evidence that participants in the study were infected or tested positive to HIV in other US states or other countries, and also that individuals from neighboring counties, such as San Francisco County, moved to San Mateo to receive HIV treatment. In studies comprised of predominantly MSM, large transmission clusters have been identified. The lack of large clustering in the present study despite nearly 40% of participants being MSM may suggest the existence of discreet transmission clusters in other geographic locations. The lack of clustering among monolingual White Hispanics, especially those who are not US-born, may reflect their transitory migration and infection with HIV along the Mexico-California border, or, alternatively, a selection bias related to immigration status as this has been independently associated with delayed presentation in HIV [46, 47].

As more HIV genotypic data become routinely available for molecular epidemiologic analyses, initial treatment options and public health approaches to ART implementation can be optimized to avoid early virologic failure. Our finding of over 4% of patient sequences in the San Mateo AIDS Program of a non-B subtype, including subtypes A, C, and recombinant CRF01_AE, provides evidence for migration of African and Asian HIV strains into the community. Despite aggressive epidemiologic surveillance, HIV awareness campaigns, and behavior change programs, the rate of new HIV infections has increased or remained constant in many US communities. High-resolution methods of genotypic surveillance for subtypes and recombinant viruses will identify potential new sources of drug resistance, viral diversity, and drivers of the US epidemic.

FIGURE 1A



Maximum likelihood phylogenetic trees of 306 HIV-1 *pol* sequences sampled from patients in San Mateo County. The trees were constructed under the GTR + I + G model of evolution in PhyML using sequences isolated in San Mateo County between 1996-2006. **A)** Sequence data presented in circular tree format. HIV-1 M-group reference strains (subtypes A-D, F-H, J, K) and common circulating recombinant forms (CRFs) were obtained from the Los Alamos HIV Sequence Database. Highlighting indicates clustering of San Mateo sequences with subtype B (gray), C (green), A (orange), and CRF01_AE (yellow) reference sequences, reflecting genetic similarity and a potential common origin with these viruses. **B)** Sequence data presented in ladder tree format. Bootstrap values indicate a statistically supported (highly probable) or statistically possible transmission cluster.

FIGURE 1B

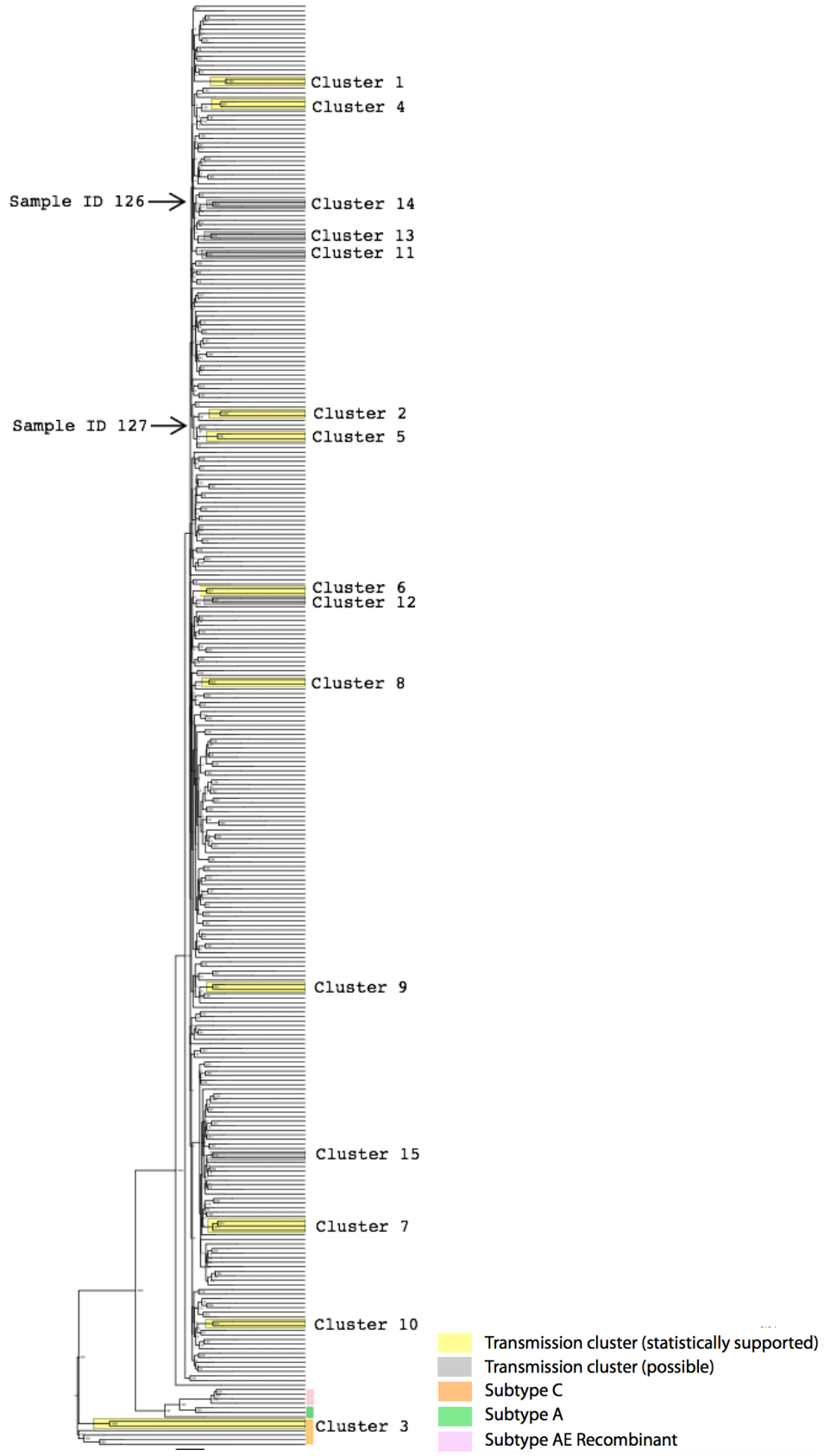

Cluster 1
Cluster 4
Sample ID 126
Cluster 14
Cluster 13
Cluster 11
Sample ID 127
Cluster 2
Cluster 5
Cluster 6
Cluster 12
Cluster 8
Cluster 9
Cluster 15
Cluster 7
Cluster 10
Cluster 3
Transmission cluster (statistically supported)
Transmission cluster (possible)
Subtype C
Subtype A
Subtype AE Recombinant

TABLE 1. Baseline characteristics of 241 participants in the San Mateo County AIDS Program with available demographic information (1996-2010)

	Number (%)	Missing
Gender		0
Male	192 (79.7)	
Female	49 (20.3)	
Age		0
Age at time of sample, mean (SD)	41.7 (8.6)	
Race		9
White non Hispanic	100 (43.1)	
Black or African American	65 (28.0)	
White Hispanic	46 (19.8)	
Asian	10 (4.3)	
American Indian or Native Alaskan	6 (2.6)	
Pacific Islander	2 (0.9)	
More than one	3 (1.3)	
Mode of transmission		64
MSM	71 (40.1)	
Heterosexual and other	49 (27.7)	
IDU	30 (17.0)	
MSM and IDU	27 (15.3)	

TABLE 2. Characteristics of participants in transmission clusters identified by phylogenetic analysis

Cluster number	Bootstrap value	Gender	Age at time of sample	Year of diagnosis	Race	Primary Language	Mode of transmission
Statistically supported							
1	1000	Male	30	2004	White Hispanic	Spanish	Heterosexual
		Female	22	2004	White Hispanic	Spanish	Heterosexual
2	1000	Male	27	2007	White Hispanic	Spanish	Heterosexual
		Unknown	-	-	-	-	-
3	1000	Unknown	-	-	-	-	-
		Unknown	-	-	-	-	-
4	997	Male	34	2001	White non Hispanic	English	IDU
		Female	32	1992	White non Hispanic	English	IDU
5	994	Female	40	-	Black or African American	English	IDU
		Female	46	1988	Black or African American	English	Heterosexual
6	991	Male	37	1995	White non Hispanic	English	MSM and IDU
		Male	50	1995	White Hispanic	Bilingual	MSM
7	983	Male	59	-	Black or African American	English	Heterosexual
		Unknown	-	-	-	-	-
		Female	47	-	Black or African American	English	-
8	982	Unknown	-	-	-	-	-
		Male	39	-	White non Hispanic	-	MSM
9	982	Unknown	-	-	-	-	-
		Male	31	1997	White Hispanic	Bilingual	MSM
10	976	Unknown	-	-	-	-	-
		Male	37	1994	White non Hispanic	English	MSM AND IDU
Statistically possible							
11	947	Unknown	-	-	-	-	-
		Unknown	-	-	-	-	-
12	927	Male	45	-	White non Hispanic	English	Heterosexual
		Female	37	1995	White non Hispanic	English	Heterosexual
13	877	Male	28	-	White Hispanic	-	MSM
		Male	40	-	Black or African American	English	Heterosexual
14	833	Female	26	2001	White Hispanic	-	Heterosexual
		Male	47	1994	White Not Hispanic	English	Heterosexual
15	765	Male	38	1993	Black or African American	English	IDU
		Female	30	1992	Black or African American	English	Heterosexual

TABLE 3. Comparison of factors associated with transmission clustering

	N	In cluster N (%)	Not in cluster N (%)	Univariate analysis OR (95% CI)	*P*-value
Male	241	14 (63.6)	178 (81.3)	0.4 (0.2–1.0)	0.9
Age at sample (year)	241			-	0.01
Race	241				
White not Hispanic		8 (36.4)	92 (42.0)	0.8 (0.3–2.0)	0.6
Black or AA		7 (31.8)	58 (26.5)	1.3 (0.5–3.3)	0.6
White Hispanic		6 (27.3)	40 (18.3)	1.7 (0.6–4.6)	0.3
Other		0	20 (9.1)	-	0.2
Mode of infection	177				
MSM		2 (12.5)	69 (42.9)	0.2 (0.04–0.9)	0.02
Heterosexual and other		8 (50.0)	41 (25.5)	2.9 (1.0–8.3)	0.04
IDU		4 (25.0)	26 (16.2)	1.7 (0.5–5.8)	0.5
MSM and IDU		2 (12.5)	25 (15.5)	0.8 (0.2–3.6)	1.0
Recent infection[1]	160	5 (33.3)	12 (8.3)	5.5 (1.6–18.9)	0.01
Year of HIV diagnosis	163			-	0.5
Year of sample	315			-	0.9
CD4 cells	96				
Absolute (cells/mm^3) mean (SD)		286 (263.9)	308 (190.9)	-	0.7
Percent mean (SD)		13.7 (8.5)	16.8 (9.5)	-	0.3
HIV viral load (log_{10} copies/mL) mean (SD)	158	4.4 (0.7)	4.4 (4.0)	-	1.0
Treatment					
Ever NRTI	166	11 (84.6)	136 (88.9)	0.7 (0.1–3.4)	0.6
Ever NNRTI	162	1 (7.7)	43 (28.9)	0.2 (0.03–1.6)	0.2
Ever PI	161	9 (69.2)	98 (66.2)	1.1 (0.3–3.9)	1.0
Drug resistant mutations	305				
Any drug class		18 (58.1)	181 (66.1)	0.7 (0.3–1.5)	0.4
NRTI		13 (41.9)	165 (60.2)	0.5 (0.2–1.0)	0.05
NNRTI		6 (19.4)	64 (23.4)	1.0 (0.8–1.1)	0.6
PI		7 (22.6)	74 (27.0)	0.8 (0.3–1.9)	0.7

[1] GART sample taken less than 1 year from HIV diagnosis

Table 4. Summary of demographic, clinical and drug resistance mutation data for 305 participants in the San Mateo County AIDS Program (1996-2010)

	Number (%)	Missing (%)
Age at sample (year)		64 (21.0)
19–29	16 (6.6)	
30–39	81 (33.6)	
40–49	101 (41.9)	
50–67	43 (17.8)	
Recent infection (sample < 1 year from diagnosis)		146 (47.9)
No	142 (89.3)	
Yes	17 (10.7)	
Year of HIV diagnosis		142 (46.6)
1981–1989	26 (16.0)	
1990–1999	100 (61.4)	
2000–2008	37 (22.7)	
Year of sample		0
1997–1999	82 (26.9)	
2000–2002	107 (35.1)	
2003–2005	72 (23.6)	
2006–2008	44 (14.4)	
CD4 cells		
Absolute count (cells/mm^3), mean (SD)	305.4 (199.0)	209 (68.5)
Percent (%), mean (SD)	16.5 (9.4)	207 (67.9)
HIV viral load ($\log_{10}$ copies/mL) mean (SD)	4.1 (0.9)	147 (48.2)
Treatment		
Naïve	18 (10.8)	139 (45.6)
Ever NRTI	147 (88.6)	139 (45.6)
Ever AZT	110 (67.9)	143 (46.9)
Ever NNRTI	44 (27.2)	143 (46.9)
Ever PI	107 (66.5)	144 (47.2)
Drug resistance mutation		
All participants (N = 305)		0
Any drug resistant mutation	199 (65.3)	
Any NRTI resistant mutation	178 (58.4)	
Any NNRTI resistant mutation	70 (23.0)	
Any PI resistant mutation	81 (26.6)	
Resistance mutation in 2 drug classes	100 (32.8)	
Resistance mutation in 3 drug classes	15 (4.9)	
Treatment naïve (N = 18)		139 (45.6)
Any drug resistant mutation	4 (22.2)	
Any NRTI resistant mutation	4 (22.2)	
Any NNRTI resistant mutation	2 (11.1)	
Any PI resistant mutation	2 (11.1)	
Resistance mutation in 2 drug classes	0	
Resistance mutation in 3 drug classes	2 (11.1)	
Known prior treatment (N = 148)		139 (45.6)
Any drug resistant mutation	107 (72.3)	
Any NRTI resistant mutation	98 (66.2)	
Any NNRTI resistant mutation	36 (24.3)	
Any PI resistant mutation	46 (31.1)	
Resistance mutation in 2 drug classes	55 (37.2)	
Resistance mutation in 3 drug classes	9 (6.1)	

Table 5. Risk factors associated with drug resistance mutations

	Any drug mutation				Any NRTI mutation			
	Univariate		Multivariate		Univariate		Multivariate	
	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
Male	3.2 (1.1–9.3)	0.03	6.0 (1.4–26.1)	0.02	2.4 1.0–5.8	0.04	3.6 (0.9–14.0)	0.07
Age at sample (year)	-	0.06	0.98 (0.9–1.1)	0.7	-	0.02	1.0 (0.9–1.0)	0.4
Race								
White not Hispanic	0.6 (0.3–1.5)	0.3	N/A		0.7 (0.4–1.5)	0.4	N/A	
Black or AA[1]	0.7 (0.2–1.8)	0.4	N/A		0.8 (0.4–1.7)	0.5	N/A	
White Hispanic	2.1 (0.7–6.4)	0.2	N/A		2.9 (0.96–8.5)	0.05	N/A	
Other	2.3 (0.3–20.9)	0.7	N/A		0.6 (0.2–2.4)	0.2	N/A	
Mode of infection								
MSM	1.0 (0.4–2.4)	1.0	N/A		1.2 (0.6–2.4)	0.6	N/A	
Heterosexual and other	1.6 (0.6–4.5)	0.4	N/A		1.2 (0.5–2.6)	0.7	N/A	
IDU	0.4 (0.1–1.4)	0.2	N/A		1.0 (0.4–2.4)	0.9	N/A	
MSM and IDU	1.3 (0.4–4.4)	0.8	N/A		0.6 (0.3–1.5)	0.5	N/A	
Recent infection[2]	0.1 (0.04–0.5)	0.003	0.5 (0.02–12.1)	0.7	0.2 (0.06–0.7)	0.01	1.2 (0.06–25.4)	0.9
Year of HIV diagnosis	-	0.04			-	0.04		
1981–1989			0.9 (0.2–5.8)	0.8			0.5 (0.03–7.5)	0.4
1990–1990			0.6 (0.2–2.5)	0.4			1.5 (0.2–9.5)	0.3
2000–2008			1.0	-			1.0	-
Year of sample	-	0.0006			-	0.0001		
1997–1999			0.1 (0.02–0.8)	0.2			0.1 (0.02–0.8)	0.2
2000–2002			0.06 (0.008–0.4)	0.01			0.07 (0.01–0.4)	0.007
2003–2005			0.6 (0.1–3.3)	0.09			0.6 (0.1–3.0)	0.09
2006–2008			1.0	-			1.0	-
CD4 cell absolute (cells/mm^3)	-	0.7	N/A		-	0.5	N/A	
CD4 cell percent (%)	-	0.7	N/A		-	0.5	N/A	
Viral load ($\log_{10}$ copies/mL)	-	0.006	2.6 (1.1–6.3)	0.03	-	0.04	1.7 (0.8–3.5)	0.2
Ever NRTI	33.0 (4.0–273.7)	<0.0001	2.4 (0.1–57.3)	0.6	24.4 (3.0–200.8)	0.0001	4.1 (0.2–105.1)	0.4
Ever NNRTI	3.6 (1.1–11.6)	0.02	4.5 (1.0–19.5)	0.045	2.7 (1.0–7.4)	0.06	2.8 (0.8–10.0)	0.1
Ever PI	1.7 (0.7–4.0)	0.2	N/A		1.4 (0.6–3.2)	0.5	N/A	

Table 5 continued.

Variable	Any NNRTI mutation				Any PI mutation			
	Univariate		Multivariate		Univariate		Multivariate	
	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
Male	2.2 (0.5–10.4)	0.5	N/A		3.4 (0.7–16.1)	0.1	N/A	
Age at sample (year)	-	0.2	N/A		-	0.3	N/A	
Race								
White not Hispanic	1.6 (0.6–4.4)	0.3	N/A		1.0 (0.4–2.3)	0.9	N/A	
Black or AA	0.6 (0.2–2.2)	0.5	N/A		0.4 (0.1–1.3)	0.1	N/A	
White Hispanic	0.4 (0.1–1.7)	0.2	N/A		2.2 (0.8–5.7)	0.1	N/A	
Other	4.4 (0.8–23.5)	0.1	N/A		1.3 (0.2–7.7)	0.7	N/A	
Mode of infection								
MSM	0.8 (0.3–2.3)	0.7	N/A		2.0 (0.8–4.9)	0.1	N/A	
Heterosexual and other	1.3 (0.4–3.9)	0.6	N/A		0.6 (0.2–1.7)	0.3		
IDU	0.5 (0.1–2.7)	0.7	N/A		0.6 (0.2–2.3)	0.5		
MSM and IDU	1.5 (0.4–5.3)	0.5	N/A		0.9 (0.3–3.2)	1.0		
Recent infection	0.7 0.1–3.3)	0.6	N/A		-	0.02	N/A	
Year of HIV diagnosis	-	0.2	N/A		-	0.003	N/A	
1981–1989								
1990–1990								
2000–2008								
Year of sample	-	0.01	N/A		-	0.0009	N/A	
1997–1999								
2000–2002								
2003–2005								
2006–2008								
CD4 cell absolute (cells/mm^3)	-	0.4	N/A		-	0.2	N/A	
CD4 cell percent (%)	-	0.8	N/A			0.5	N/A	
Viral load ($\log_{10}$ copies/mL)	-	0.04	N/A		-	0.6	N/A	
Ever NRTI	-	0.1	N/A		-	0.03	N/A	
Ever NNRTI	21.7 (6.2–76.1)	<0.0001	N/A		1.7 (0.7–4.4)	0.3	N/A	
Ever PI	0.5 (0.2–1.5)	0.2	N/A		5.1 (1.6–16.4)	0.003	N/A	

[1] AA: African American

[2] GART sample taken less than 1 year from HIV diagnosis

REFERENCES

1. Stanford HIV Drug Resistance Database. http://hivdb.stanford.edu.
2. Hirsch MS, Gunthard HF, Schapiro JM, et al. Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society-USA panel. Clin Infect Dis. Jul 15 2008;47(2):266-285.
3. Gray RR, Tatem AJ, Lamers S, et al. Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. AIDS. Sep 10 2009;23(14):F9-F17.
4. Guo H, Guo D, Wei JF, et al. First detection of a novel HIV Type 1 CRF01_AE/07_BC recombinant among an epidemiologically linked cohort of IDUs in Jiangsu, China. AIDS Res Hum Retroviruses. Apr 2009;25(4):463-467.
5. Watanaveeradej V, DeSouza MS, Benenson MW, et al. Subtype C/CRF01_AE recombinant HIV-1 found in Thailand. AIDS. Sep 26 2003;17(14):2138-2140.
6. Brenner B, Roger M, Routy J, et al. High rates of forward transmission events after acute/early HIV-1 infection. J Infect Dis. Apr 1 2007;195(7):951-959.
7. Smith DM, May SJ, Tweeten S, et al. A public health model for the molecular surveillance of HIV transmission in San Diego, California. AIDS. Jan 14 2009;23(2):225-232.
8. Hall HI, Song R, Rhodes P, et al. Estimation of HIV incidence in the United States. JAMA. Aug 6 2008;300(5):520-529.
9. CDC. HIV/AIDS surveillance report, 2005. 2007;17:1-46.
10. Grant RM, Hecht FM, Warmerdam M, et al. Time trends in primary HIV-1 drug resistance among recently infected persons. JAMA. Jul 10 2002;288(2):181-188.
11. Magis-Rodriguez C, Lemp G, Hernandez MT, Sanchez MA, Estrada F, Bravo-Garcia E. Going North: Mexican migrants and their vulnerability to HIV. J Acquir Immune Defic Syndr. May 1 2009;51 Suppl 1:S21-25.
12. Levy V, Prentiss D, Balmas G, et al. Factors in the delayed HIV presentation of immigrants in Northern California: implications for voluntary counseling and testing programs. J Immigr Minor Health. Jan 2007;9(1):49-54.
13. Levy V, Page-Shafer K, Evans J, et al. HIV-related risk behavior among Hispanic immigrant men in a population-based household survey in low-income neighborhoods of northern California. Sex Transm Dis. Aug 2005;32(8):487-490.
14. Bonhoeffer S, Chappey C, Parkin NT, Whitcomb JM, Petropoulos CJ. Evidence for positive epistasis in HIV-1. Science. Nov 26 2004;306(5701):1547-1550.
15. Chen R, Quinones-Mateu ME, Mansky LM. HIV-1 mutagenesis during antiretroviral therapy: implications for successful drug treatment. Front Biosci. Jan 1 2005;10:743-750.
16. Liu Y, Nickle DC, Shriner D, et al. Molecular clock-like evolution of human immunodeficiency virus type 1. Virology. Nov 10 2004;329(1):101-108.
17. Robertson DL, Sharp PM, McCutchan FE, Hahn BH. Recombination in HIV-1. Nature. Mar 9 1995;374(6518):124-126.
18. Philpott S, Burger H, Tsoukas C, et al. Human immunodeficiency virus type 1 genomic RNA sequences in the female genital tract and blood: compartmentalization and intrapatient recombination. J Virol. Jan 2005;79(1):353-363.
19. Piontkivska H, Hughes AL. Between-host evolution of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1: an approach based on phylogenetically independent comparisons. J Virol. Nov 2004;78(21):11758-11765.

20. Tovanabutra S, Watanaveeradej V, Viputtikul K, et al. A new circulating recombinant form, CRF15_01B, reinforces the linkage between IDU and heterosexual epidemics in Thailand. AIDS Res Hum Retroviruses. Jul 2003;19(7):561-567.
21. Kuritzkes DR, Lalama CM, Ribaudo HJ, et al. Preexisting resistance to nonnucleoside reverse-transcriptase inhibitors predicts virologic failure of an efavirenz-based regimen in treatment-naive HIV-1-infected subjects. J Infect Dis. Mar 15 2008;197(6):867-870.
22. Wensing AM, van de Vijver DA, Angarano G, et al. Prevalence of drug-resistant HIV-1 variants in untreated individuals in Europe: implications for clinical management. J Infect Dis. Sep 15 2005;192(6):958-966.
23. Truong HM, Grant RM, McFarland W, et al. Routine surveillance for the detection of acute and recent HIV infections and transmission of antiretroviral resistance. AIDS. Nov 14 2006;20(17):2193-2197.
24. Truong HM, Kellogg T, Schwarcz S, et al. Frequent international travel by men who have sex with men recently diagnosed with HIV-1: potential for transmission of primary HIV-1 drug resistance. J Travel Med. Nov-Dec 2008;15(6):454-456.
25. Hayman A, Moss T, Simmons G, et al. Phylogenetic analysis of multiple heterosexual transmission events involving subtype b of HIV type 1. AIDS Res Hum Retroviruses. May 20 2001;17(8):689-695.
26. Ou CY, Ciesielski CA, Myers G, et al. Molecular epidemiology of HIV transmission in a dental practice. Science. May 22 1992;256(5060):1165-1171.
27. Ciccozzi M, Montieri S, Salemi M, et al. An outbreak of HIV-1 subtype G among Italian injecting drug users. AIDS. May 31 2007;21(9):1213-1215.
28. Salemi M, De Oliveira T, Ciccozzi M, Rezza G, Goodenow M. High-Resolution Molecular Epidemiology and Evolutionary History of HIV-1 Subtypes in Albania. PLoS ONE. Jan 1 2008;3(1):e1390.
29. Slatkin M, Maddison WP. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics. Nov 1 1989;123(3):603-613.
30. Hue S, Clewley JP, Cane PA, Pillay D. HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. AIDS. Mar 26 2004;18(5):719-728.
31. Hué S, Pillay D, Clewley JP, Pybus OG. Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. Proc Natl Acad Sci USA. Mar 22 2005;102(12):4425-4429.
32. Thompson, J.D., D.G. Higgins, and T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res, 1994. 22(22): p. 4673-80.
33. Hall, T., BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser., 1999. 41: p. 95-98.
34. Posada, D. and K.A. Crandall, MODELTEST: testing the model of DNA substitution. Bioinformatics, 1998. 14(9): p. 817-8.
35. PAUP: Phylogenetic Analysis Using Parsimony (*and Other Methods) [computer program]. Version v4.0d64: Sinauer Associates, Inc. Publishers, Sunderland, MA; 1998.
36. Guindon S, Lethiec F, Duroux P, Gascuel O. PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. Jul 1 2005;33(Web Server issue):W557-559.

37. Yerly, S., et al., The impact of transmission clusters on primary drug resistance in newly diagnosed HIV-1 infection. AIDS, 2009. 23(11): p. 1415-23.
38. Alteri, C., et al., Characterization of the patterns of drug-resistance mutations in newly diagnosed HIV-1 infected patients naive to the antiretroviral drugs. BMC Infect Dis, 2009. 9: p. 111.
39. Novak, R.M., et al., Prevalence of antiretroviral drug resistance mutations in chronically HIV-infected, treatment-naive patients: implications for routine resistance screening before initiation of antiretroviral therapy. Clin Infect Dis, 2005. 40(3): p. 468-74.
40. Payne, B.A., et al., Low prevalence of transmitted antiretroviral drug resistance in a large UK HIV-1 cohort. J Antimicrob Chemother, 2008. 62(3): p. 464-8.
41. Wensing, A.M., et al., Prevalence of drug-resistant HIV-1 variants in untreated individuals in Europe: implications for clinical management. J Infect Dis, 2005. 192(6): p. 958-66.
42. Yerly, S., et al., Drug resistance mutations in HIV-1-infected subjects during protease inhibitor-containing highly active antiretroviral therapy with nelfinavir or indinavir. Antivir Ther, 2001. 6(3): p. 185-9.
43. Yerly, S., et al., Transmission of HIV-1 drug resistance in Switzerland: a 10-year molecular epidemiology survey. AIDS, 2007. 21(16): p. 2223-9.
44. U.S. Census Bureau. 2008 American Community Survey 1-year estimates. 2008
45. County of San Mateo, Fact Sheet AIDS, Healthy San Mateo 2010: Health Status Indicators for San Mateo County, California, County of San Mateo.
46. Levy, V., et al., Factors in the delayed HIV presentation of immigrants in Northern California: implications for voluntary counseling and testing programs. J Immigr Minor Health, 2007. 9(1): p. 49-54.
47. Levy, V., et al., HIV-related risk behavior among Hispanic immigrant men in a population-based household survey in low-income neighborhoods of northern California. Sex Transm Dis, 2005. 32(8): p. 487-90.

Chapter 3: Implications of HIV-1 Viral Diversity in Mother-to-Child Transmission of HIV: Molecular Epidemiology and a Review of the Literature

INTRODUCTION

In most wealthy or industrialized countries, the frequency of mother-to-child transmission (MTCT) of HIV-1 has been reduced to less than 2% as a result of prevention strategies including access to highly active antiretroviral therapy (HAART), replacement of breast-feeding, elective caesarean section, or a combination thereof [1]. However, in many resource-limited countries with high levels of endemic infection, transmission of HIV-1 from mother to child is recognized as a leading cause of infant and child mortality, where in the absence of antiretroviral treatment (ART) and prophylaxis, more than 25% of infants born to HIV-positive women become infected with HIV [2, 3]. MTCT may be influenced by a number of maternal characteristics, including high plasma viral load, low CD4+ T-lymphocyte count, other cervico-vaginal infections at delivery, and concentration of virus in breast-milk [4-6]. In settings where prolonged breast-feeding is practiced, infection rates may rise to nearly 40 percent. Societal risk factors for mother-to-child transmission include lack of access to prevention-of-mother-to-child-transmission (PMTCT) services and high-risk sexual behavior or injection drug use during the second and third trimesters of pregnancy [7].

Consensus conventions adopted to classify the presumed timing of HIV transmission distinguish between in utero, intrapartum/early postpartum, and late postpartum infant infection. Infants are classified as having acquired in utero infection on the basis of PCR detection of HIV-1 RNA or DNA from infant blood samples obtained within 72 hours of birth. Infants negative by a nucleic acid amplification assay at birth but in whom infection is detectable by 2-6 weeks are classified as intrapartum or early postpartum transmission. In these infants, transmission of virus is thought to have occurred during very late gestation, labor, or delivery, potentially though oral and mucosal exposure to maternal virus in blood, cervical secretions, or colostrum via early breast feeding. Infants with undetectable infection by screening assays within 6-8 weeks of birth who later acquire infection are deemed to have late postpartum transmission presumed to be related to breast-feeding [8. 9].

It is hypothesized that infant infection acquired at each of these times may result from distinct modes of viral transmission associated with different virus sub-populations, or quasi-species, derived from the diverse assortment of viruses present during chronic infection in the mother. Whether cell-associated provirus or viral RNA from blood, cervical secretions, or breast-milk is the primary source of infection in each of these settings is not well understood. The early observation that in HIV-positive mothers giving birth to twins, the first-born twin is much more likely to acquire infection than the subsequently born twin, supported the idea that initial delivery-related trauma and contact with maternal blood and secretions is important in facilitating MTCT [10], although more recent data are conflicting [11]. When maternal ART is initiated in the second trimester at 24-28 weeks, less than 2% of infants acquire HIV infection, suggesting that early infection in utero occurs much less frequently than infection in later stages [12]. The most striking reductions in MTCT are seen when both maternal HIV RNA is suppressed by ART and post-partum ART prophylaxis is given to the newborn. Without ART there is more than a two-fold increase in transmission for every log increase in maternal viral load [13]. Transmission from mothers with less than 1,000 copies/ml of plasma HIV RNA is unusual [14]. Similarly, there is an independent association between lower CD4 cell count and reduced transmission, suggesting that maternal immune response plays a role in limiting MTCT.

The extent to which viral genetic and evolutionary characteristics influence the risk of MTCT of HIV remains unclear. Investigation of the genetic composition and evolution of HIV in the context of MTCT can elucidate viral risk factors for transmission, infection, and disease progression. As HIV sequence data and sophisticated molecular biology tools have become more widely available, analyses of these data have provided a more complete understanding of the factors influencing MTCT and enabled development of additional prevention strategies.

SELECTION AND BOTTLENECKS IN MTCT OF HIV

Cases of MTCT of HIV represent opportunities to identify the virologic characteristics associated with transmission. Some studies have investigated specific characteristics of viral genes while others have analyzed the diversity of HIV-1 quasi-species, comparing the range of viral diversity in maternal samples to those obtained from the infant. These studies suggest that infection in the infant is usually initiated by a single or limited number of maternal viruses, indicating the presence of a bottleneck in vertical HIV transmission [15]. Such a bottleneck may provide an opportunity for intervention strategies in MTCT and transmission in general.

The extensive diversity of viruses present in prolonged HIV-1 infection within individuals, including pregnant women, is the result of continuing viral expansion, selection, and evolution in the face of immune and/or drug pressures. In chronic infection up to 10^{12} new virions are produced daily within infected individuals [16], with a high rate of evolutionary change driven primarily by the poor fidelity of both reverse transcriptase and RNA polymerase. The rate of erroneous nucleotide substitution during reverse transcription is approximately one nucleotide per 10^5 base pairs [17]. This is accompanied by the absence of missense correction mechanisms in the reverse transcription and synthesis of the viral genome.

The viral diversity generated by mutation, nucleotide substitution and point mutation, insertion/deletions, and duplication, is acted upon by selective pressure from the host humoral and cellular immune systems and, when present, antiretroviral drug therapy. In addition, strand switching by the viral reverse transcriptase enables frequent recombination events. A robust host immune response exerts a strong selective force on the viral population, and patients with longer asymptomatic phases typically show evidence of greater positive selection [18, 19]. Positive selection is identified when the per-site rate of non-synonymous mutation (d_N) exceeds that of synonymous mutation (d_S) [18, 20, 21]. Vertical HIV transmission events are shaped by extensive diversity in maternal virus, genetic bottlenecks in MTCT, and evidence of founder effects in the infected infant resulting in relative viral homogeneity of early infant infection, although the extent of diversity may vary by mode of transmission [15, 22]. MTCT may result in transmission of a minor subset of maternal viruses with limited viral diversity in infant viral sequences compared to the maternal sequences [23-25]. In some settings there is evidence that infection in utero may involve transmission of a greater diversity of maternal viral variants compared to transmission at delivery [23, 26].

The association between the strength of positive selection, progression of maternal disease, and HIV transmission are not well-defined. Numerous small studies suggest that viral variants transmitted from mother-to-child are derived from a minority virus population which has

effectively escaped the maternal immune response [23-31]. In addition, transmitted viruses may derive from the major maternal variant in compartments including blood, placenta, or cervical secretions in the birth canal. Differences between maternal and infant immunogenetics, particularly HLA specificity in the recognition of CTL epitopes by maternal and infant immune responses, likely play a role in selection. Additionally, some transmitted viruses may have a replication advantage in the infant and may therefore be naturally selected [32]. Alternatively, success of a particular HIV variant may result from stochastic effects unrelated to selection [33]. In a phylogenetic analysis of maternal and infant C2V3 envelope sequences, Ceballos et al. found that during early pregnancy a single minor viral variant was transmitted to the infant followed by subsequent evolution, suggesting either a selection process or a stochastic event, whereas in later maternal infection (during the final trimester or via breast feeding), maternal and infant sequences were intermingled, suggesting repeated transmission of multiple viral variants [33]. This analysis did not detect positive selection ($d_N/d_S > 1$) in the maternal and infant sequences, suggesting that stochastic effects accounted for the viral quasispecies that were successfully transmitted.

ROLE OF VIRAL RECOMBINATION AND DIVERSITY IN MTCT OF HIV

HIV has an estimated recombination rate of three events per genome per replication, one of the highest rates of all organisms [34]. Recombination occurs as a result of host cell superinfection by distinct viruses [35-37]. Inter- and intra-subtype recombination are the major driving forces of HIV genetic diversity [38, 39] and have resulted in a wide distribution of circulating recombinant forms which contribute significantly to the global pandemic (accounting for >25% of infections in some regions) [38, 40-42]. Recombination events allow viruses to escape immune pressure, avoid accumulation of deleterious mutations, or jump between adaptive peaks [43-46]. Recombinant viruses may also have fitness and/or transmission advantages, potentially enhancing observed rates of MTCT [47]. Quan et al. reported that defective HIV provirus with a lethal mutation in env can be rescued by superinfection with wild-type HIV or a second replication-defective virus with a lethally-mutated capsid protein [48]. These findings suggest that defective noninfectious HIV-1 variants, which constitute a large proportion of *in vivo* HIV populations [49], may constitute a major part of the HIV-1 reservoir.

Within geographic settings where multiple subtypes circulate and recombine, differences have been observed in the timing and rate of transmission based on the subtype of specific genes. It has been suggested that geographic differences in rates of MTCT may be related to the genetic diversity of HIV in different settings [50-51]. HIV-1 subtypes and recombinant forms may have functional and phenotypic differences, including chemokine coreceptor usage, replication efficiency and hence viral load that may lead to differences in rates of vertical transmission [52-61]. Studies examining the rate of MTCT by maternal subtype or recombinant infection have yielded discrepant findings. Pádua et al. were unable to detect specific genetic forms in env or nef sequences associated with MTCT or a significant difference in RNA viral load by viral subtype, but found a greater diversity of genetic forms among non-transmitting mothers [51]. The latter suggests that increased maternal immune pressure may successfully limit transmission. Renjifo et al. found that MTCT was more common among mothers infected with viruses that included a subtype C envelope, compared to viruses with subtype A or D (or A/D recombinant) envelope in Tanzania [61]. In a separate study based on the C2-C5 envelope and 5‘LTR regions,

Renjifo et al. found that some intersubtype recombinant viruses are preferentially transmitted during breast feeding [60]. In Kenya, Yang et al. found MTCT to be more common among mothers infected with subtype D or A/D recombinant viruses compared to subtype A [62].

Whether and how recombinant viruses have a transmission advantage in MTCT warrants further investigation, particularly with the increasing prevalence of circulating recombinant forms of HIV-1. Studies interpreting the effect of viral genotype on the risk on MTCT may be enhanced by increased sample sizes, adjustment of confounding variables, and improved detection of recombinants on a population level [47, 53, 59-62, 64]. However, the success of ART and prevention limits each of these as infant infection becomes less frequent in study populations.

PHYLOGENETICS AND COMPARTMENTALIZATION IN MTCT OF HIV

Phylogenetic analyses of HIV genetic sequences allow computation of the genetic relatedness of multiple virus variants within or between individuals and across populations. These analyses can uncover trends within an epidemic and also provide insight into the origins, timing, and demographic history of transmitted viruses. Phylogenetic comparisons of infant and maternal sequences from various compartments are crucial to understanding the viral evolution in these tissues and their respective contributions to the risk of MTCT.

Compartmentalization of HIV infection involves the formation of distinct genetic populations in specific organs, cells or tissues. Compartments of special importance to MTCT are the maternal genital tract, placenta, and breast milk. Multiple studies have shown that HIV variants from the genital tract appear distinct from the blood [65-77], although HIV-1 viral load is typically lower in these compartments compared to plasma [78-80]. Localized inflammation, co-infections, physical and cellular barriers, incomplete penetration of antiretroviral drugs into the genital tract or local immune responses may account for independent and divergent evolution within the maternal genital tract compared to plasma [66-68, 81, 82].

Bull et al. analyzed HIV-1 RNA and cell-associated HIV-1 DNA (env) from the blood and genital tract of women with chronic HIV-infection and reported low diversity of genital-tract specific phylogenetic clades, particularly from the cervix, possibly indicating bursts of viral replication or the proliferation of infected cells in compartments [79]. However, the absence of tissue-specific genetic features and the phylogenetic overlap of genital tract HIV clades with those from the blood suggest that HIV-1 flow is not restricted between the genital tract and blood, and that viral evolution may not occur independently within the two compartments [79].

HIV compartmentalization between blood and breast milk is also not well-understood. Breast milk transmission is a major source of pediatric HIV infection [83, 84], yet the few studies of viral compartmentalization in breast milk provide contradictory results [85-88]. Limited exchange of virus and viral characteristics is thought to occur between the plasma and breast milk, where even in the absence of antiretroviral therapy HIV-1 viral load is 10-100 fold lower than in plasma [78]. In addition, immunologic elements, such as HIV-specific T-cells, antibodies, cytokines, and chemokines appear to be highly compartmentalized in breast milk [89-91]. For example, Becquart et al. detected compartmentalization of the humoral IgG response to HIV in the mammary gland [92]. Heath et al. recently examined the compartmentalization of

HIV-1 between breast milk and blood in envelope sequences from 13 breast-feeding women, uncovering substantial viral exchange despite significantly reduced HIV-1 load in milk [93]. Specifically, genetic compartmentalization in breast milk was only detected in one of six subjects with contemporaneously-collected samples available. The authors suggest that virologic selection in breast milk does not account for the genetic bottleneck associated with mother-to-child transmission [93]. Further studies will clarify the extent to which potential tissue-specific virologic divergence may influence the risk of transmission or the characteristics of the transmitted variant.

IMPLICATIONS FOR DRUG RESISTANCE AND VACCINE DESIGN IN MTCT OF HIV

The ability of HIV to accumulate and exchange drug-resistance via single nucleotide mutations and recombination presents a clinical dilemma in the use of ART, particularly single dose nevirapine to prevent MTCT. Furthermore, once drug resistance has been selected it may persist in circulating viral RNA and within latent reservoirs of proviral DNA [46]. The persistence of maternal drug resistance mutations following drug exposure during pregnancy presents significant challenges to the design of optimal antiretroviral regimens that prevent MTCT without compromising the efficacy of HAART for mothers [94]. Short-course peripartum regimens in resource-poor settings, maternal single-dose nevirapine, and short-course zidovudine significantly reduce MTCT (37-77% compared to no intervention) [95-101]. However, up to two-thirds of mothers treated with these regimens develop viral resistance to nonnucleoside reverse-transcriptase inhibitor (NNRTI) drugs [102]. Universal HAART, already a mainstay of maternal HIV treatment in developed or resource-rich countries, could mitigate some of the burden of drug resistance associated with peripartum regimens and has been recommended for pregnant and breast-feeding mothers in resource-poor settings [94].

An understanding of HIV molecular evolution and diversity is also important in the design of HIV vaccine strategies [21]. There is a compelling rationale to develop a preventive HIV vaccine for use in infants to prevent MTCT via breast milk and provide a foundation for life-long immunity [103]. Several advances in HIV vaccine research and development, including the partial protection imparted by the ALVAC-AIDSVAX vaccine in Thailand [104], identification of a novel HIV-1 vaccine target expressed on envelope protein [105], and vaccine-induced control of simian immunodeficiency virus in rhesus monkeys [106], have given momentum to renewed efforts to develop an HIV vaccine. The genetic bottlenecks observed in MTCT provide a model for selective transmission, which may inform the design of vaccines, the identification of target antigens, and potentially, the inclusion of infants and pregnant women in preventive vaccine trials. An understanding of the bottleneck(s) in transmission and specific characteristics of transmitted viruses coupled with innovative molecular technologies and data-driven analytic methods, advances in HIV vaccinology, and recruitment of scientific talent are important steps towards achieving a successful HIV vaccine [107].

CONCLUSION AND FUTURE DIRECTIONS

Substantial progress has been made in understanding the evolution of HIV and the factors influencing the risk of MTCT. Translation of these scientific advances, primarily in the use of antiretroviral drugs to prevent MTCT of HIV, has led to successful interventions and significant

reductions in infant infection where preventive strategies have been made available. However, critical questions regarding the impact of molecular evolution and extensive genetic diversity of HIV on MTCT remain unanswered. Molecular epidemiologic studies of viral characteristics that contribute to the risk of vertical transmission may inform drug and vaccine prevention efforts. Further research to identify the selective factors governing which variants are transmitted, how the compartmentalization of HIV variants in different cells and tissues contributes to transmission, and the influence of viral diversity and recombination on the risk of MTCT, may provide insight into new treatment and prevention strategies and the development of an effective HIV vaccine.

REFERENCES

1. Newell ML, Coovadia H, Cortina-Borja M et al. Mortality of infected and uninfected infants born to HIV-infected mothers in Africa: a pooled analysis. Lancet 364, 1236–1243.
2. Mofenson LM. Prevention in neglected subpopulations: prevention of mother-to-child transmission of HIV infection. Clin Infect Dis 2010; 50:130-48.
3. Bulterys M, Wilfert CM. HAART during pregnancy and during breast feeding among HIV-infected women in the developing world: has the time come? AIDS 2009; 23:2473-2477.
4. Garcia PM, et al. Maternal levels of plasma human immunodeficiency virus type 1 RNA and the risk of perinatal transmission. Women and Infants Transmission Study Group. N Engl J Med 1999; 341:394-402.
5. Mofenson LM, et al. Risk factors for perinatal transmission of human immunodeficiency virus type 1 in women treated with zidovudine. Pediatric AIDS Clinical Trials Group Study 185 Team. N Engl J Med 1999; 341:385-93.
6. Koulinska IN, et al. Risk of HIV-1 transmission by breastfeeding among mothers infected with recombinant and non-recombinant HIV-1 genotypes. Virus Res 2006; 120:191-198.
7. Bulterys M, Landesman S, Burns D, Rubinstein A, Goedert J. Sexual behavior and injection drug use during pregnancy and vertical transmission of HIV-1. JAIDS 1997; 15:76-82.
8. Kourtis AP, Bulterys M, Nesheim S, Lee F. Understanding the timing of HIV transmission from mother to infant. JAMA 2001; 285:709-712.
9. Kourtis AP, Lee FK, Jamieson DJ, Bulterys M. Mother-to-child transmission of HIV-1: timing and implications for prevention. Lancet Infect Dis 2006; 6:726-732.
10. Goedert JJ, et al. High risk of HIV-1 infection for first-born twins. The International Registry of HIV-Exposed Twins. Lancet 1991; 338:1471-1475.
11. Biggar RJ, et al. The Risk of Human Immunodeficiency Virus–1 Infection in Twin Pairs Born to Infected Mothers in Africa. J Infect Dis 2003; 188:850–855.
12. Connor EM, et al. Reduction of maternal-infant transmission of human immunodeficiency virus type-1 with zidovudine treatment. N Eng J Med 1994; 331:1173-1180.
13. Guevara H, Casseb J, Zijenah LS, Mbizvo M, Oceguera LF, Hanson C, Katzenstein DA, Hendry MR. Maternal HIV-1 antibody and vertical transmission in subtype C virus infection. JAIDS 2002; 29:435-440.

Ioannidis J, et al. Perinatal transmission of human immunodeficiency virus type 1 by pregnant women with RNA virus loads <1000 copies/mL. J Infect Dis 2001; 183:539-545.

15. Delwart, E. et al. Homogeneous quasispecies in 16 out of 17 individuals during very early HIV-1 primary infection. AIDS 2002; 16:189–195.
16. Perelson AS, Neumann A, Markowitz M et al. HIV-1 dynamics in vivo: virion clearance rate, infected cell life span, and viral generation time. Science 271: 1582-1586.
17. Temin HM. Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. PNAS 1993; 90:6900-6903.

18. Ross HA & Rodrigo AG. Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. J Virol 2002; 76:11715-11720.
19. Williamson, S. Adaptation in the env gene of HIV-1 and evolutionary theories of disease progression. Mol Biol Evol 2003; 20:1318–1325.
20. Bonhoeffer S, Holmes EC, Nowak MA. Causes of HIV diversity. Nature 1995; 376:125.
21. Rambaut A, Posada D, Crandall KA, Holmes EC. The causes and consequences of HIV evolution. Nature Reviews Genetics 2004; 5:52-61.
22. Zhu, T. F. et al. Genotypic and phenotypic characterization of HIV-1 in patients with primary infection. Science 1993; 261: 1179–1181.
23. Dickover RE, Garratty EM, Plaeger S and Bryson, YJ. Perinatal transmission of major, minor, and multiple maternal human immunodeficiency virus type 1 variants in utero and intrapartum. Journal of Virology 2001; 75:2194–2203.
24. Wolinsky SM, Wike CM, Korber BT, Hutto C, Parks WP, Rosenblum LL, Kunstman KJ, Furtado MR, Muñoz JL. Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. Science 1992; 255:1134-1137.
25. Ahmad N, Baroudy BM, Baker RC, Chappey C. Genetic analysis of human immunodeficiency virus type 1 envelope V3 region isolates from mothers and infants after perinatal transmission. J Virol 1995; 69:1001-12.
26. Renjifo B, et al. In-utero transmission of quasispecies among human immunodeficiency virus type 1 genotypes. Virology 2003; 307:278-282.
27. Blish CA, Blay WM, Haigwood NL, Overbaugh J. Transmission of HIV-1 in the face of neutralizing antibodies. Curr HIV Res 2007; 5:578–587.
28. Kampinga GA, et al. Primary infections with HIV-1 of women and their offspring in Rwanda: findings of heterogeneity at seroconversion, coinfection, and recombinants of HIV-1 subtypes A and C. Virology 1997; 227:63–76.
29. Kliks S, et al. Genetic analysis of viral variants selected in transmission of human immunodeficiency viruses to newborns. AIDS Res Hum Retroviruses 2000; 16:1223–1233.
30. Verhofstede C, et al. Diversity of the human immunodeficiency virus type 1 (HIV-1) env sequence after vertical transmission in mother-child pairs infected with HIV-1 subtype A. J Virol 2003; 77:3050–3057.
31. Wu X Parast AB, et al. Neutralization escape variants of human immunodeficiency virus type 1 are transmitted from mother to infant. J Virol 2006; 80:835-844.
32. Kuhn L, et al. Maternal versus paternal inheritance of HLA class I alleles among HIV-infected children: consequences for clinical disease progression. AIDS 2004; 18:1281-1289.
33. Ceballos A, et al. Lack of viral selection in human immunodeficiency virus type 1 mother-to-child transmission with primary infection during late pregnancy and/or breastfeeding. J Gen Virol 2008; 89:2773-82.
34. Zhuang J, et al. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. J Virol 2002; 76:11273–11282.
35. Jung A, et al. Multiply infected spleen cells in HIV patients. Nature 2002; 418:144.
36. Jost S, et al. A patient with HIV-1 superinfection. N Engl J Med 2002; 347:731–736.
37. Koelsch, K. K. et al. Clade B HIV-1 superinfection with wild-type virus after primary infection with drug-resistant clade B virus. AIDS 2003; 17:11–6.

38. Crandall KA, Templeton AR. The Evolution of HIV. Baltimore: The Johns Hopkins University Press; 1999. p. 153–176.
39. McVean G, Awadalla P, Fearnhead P. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics 2002; 160:1231–1241.
40. Robertson DL, Sharp PM, McCutchan FE, Hahn BH. Recombination in HIV-1. Nature 1995; 374:124–126.
41. Pandrea I, et al. Analysis of partial pol and env sequences indicates a high prevalence of HIV type 1 recombinant strains circulating in Gabon. AIDS Res Hum Retroviruses 2002; 18:1103-1116.
42. Essex M and M'Boup S. Regional Variation in the African Epidemics. AIDS in Africa (second ed.). Kluwer Academic/Plenum Publishers. 2002.
43. Liu SL, et al. Selection for human immunodeficiency virus type 1 recombinants in a patient with rapid progression to AIDS. J Virol 2002; 76:10674-10684.
44. Nájera R, Delgado E, Pérez-Alvarez L, Thomson MM. Genetic recombination and its role in the development of the HIV-1 pandemic. AIDS 2002; 16:3-16.
45. Kellam P, Larder BA. Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance. J Virol 1995; 69:669-674.
46. Morris A, et al. Mosaic structure of the human immunodeficiency virus type 1 genome infecting lymphoid cells and the brain: evidence for frequent in vivo recombination events in the evolution of regional populations. J Virol 1999; 73:8720–8731.
47. Koulinska IN, Villamor E, Chaplin B, Msamanga G. Transmission of Cell-Free and Cell-Associated HIV-1 Through Breast-Feeding. JAIDS 2006 41(1)1: 93-99.
48. Quan Y, Liang C, Brenner B, Wainberg M. Multidrug−Resistant Variants of HIV Type 1 (HIV−1) Can Exist in Cells as Defective Quasispecies and Be Rescued by Superinfection with Other Defective HIV−1 Variants. J Infect Dis 2009; 9:1479-1483.
49. Li Y, Kappes JC, Conway JA, et al. Molecular characterization of human immunodeficiency virus type 1 cloned directly from uncultured human brain tissue: identification of replication−competent and−defective viral genomes. J Virol 1991; 65:3973–85.
50. Odaibo GN, Olaleye DO, Heyndrickx L, et al. Mother-to-child transmission of different HIV-1 subtypes among ARV naïve infected pregnant women in Nigeria. Rev Inst Med Trop S Paulo 2006; 48:77–80.
51. Pádua E, et al. Potential impact of viral load and genetic makeup of HIV type 1 on mother-to-child transmission: characterization of env-C2V3C3 and nef sequences. AIDS Res Hum Retroviruses 2009; 25:1171-7.
52. Bjorndal A, et al. Phenotypic characteristics of human immunodeficiency virus type 1 subtype C isolates of Ethiopian AIDS patients. AIDS 1999; 15:647–653.
53. Hu DJ, et al. Viral load differences in early infection with two HIV-1 subtypes. AIDS 2001; 15:683–691.
54. Jeeninga RE, et al. Functional differences between the long terminal repeat transcriptional promoters of human immunodeficiency virus type 1 subtypes A through G. J Virol 2000; 74:3740-3751.
55. Kaleebu P, et al. Effect of human immunodeficiency virus (HIV) type 1 envelope subtypes A and D on disease progression in a large cohort of HIV-1-positive persons in Uganda. J Infect Dis 2002; 185:1244–1250.

56. Kanki PJ, et al. Human immunodeficiency virus type 1 subtypes differ in disease progression. J Infect Dis 1999; 179:68–73.
57. Neilson JR, et al. Subtypes of human immunodeficiency virus type 1 and disease stage among women in Nairobi, Kenya. J Virol 1999; 73:4393–4403.
58. Tscherning C, et al. Differences in chemokine coreceptor usage between genetic subtypes of HIV-1. Virology 1998; 241:181–188.
59. Blackard JT, Renjifo B, Chaplin B, Msamanga G, Fawzi W, Essex M. Diversity of the HIV-1 long terminal repeat following mother-to-child transmission. Virology 2000; 274:402-411.
60. Renjifo B, et al. Differences in perinatal transmission among human immunodeficiency virus type 1 genotypes. J Hum Virol 2001; 4:16–25.
61. Renjifo B, et al, Tanzanian Vitamin, HIV Study Group. Preferential in utero transmission of HIV-1 subtype C compared to subtype A or D. AIDS 2004; 18:1629-1636.
62. Yang C, et al. Genetic diversity of HIV-1 in western Kenya: subtype-specific differences in mother-to-child transmission. AIDS 2003; 11:1667–1674.
63. Murray MC, et al. Effect of human immunodeficiency virus (HIV) type 1 viral genotype on mother-to-child transmission of HIV-1. J Infect Dis 2000; 181:746–749.
64. Tapia N, et al. Influence of human immunodeficiency virus type 1 subtype on mother-to-child transmission. J Gen Virol 2003; 84:607–613.
65. Poss M, Martin HL, Kreiss JK, Granville L, Chohan B, et al. Diversity in virus populations from genital secretions and peripheral blood from women recently infected with human immunodeficiency virus type 1. J Virol 1995; 69:8118–8122.
66. Poss M, Rodrigo AG, Gosink JJ, Learn GH, de Vange Panteleeff D, et al. Evolution of envelope sequences from the genital tract and peripheral blood of women infected with clade A human immunodeficiency virus type 1. J Virol 1998; 72:8240–8251.
67. Kovacs A, Wasserman SS, Burns D, Wright DJ, Cohn J, et al. Determinants of HIV-1 shedding in the genital tract of women. Lancet 2001; 358:1593–1601.
68. Wright TC, Jr., Subbarao S, Ellerbrock TV, Lennox JL, Evans-Strickfaden T, et al. Human immunodeficiency virus 1 expression in the female genital tract in association with cervical inflammation and ulceration. Am J Obstet Gynecol 2001; 184:279–285.
69. Ellerbrock TV, Lennox JL, Clancy KA, Schinazi RF, Wright TC, et al. Cellular replication of human immunodeficiency virus type 1 occurs in vaginal secretions. J Infect Dis 2001; 184:28-36.
70. Kemal KS, Foley B, Burger H, Anastos K, Minkoff H, et al. HIV-1 in genital tract and plasma of women: compartmentalization of viral sequences, coreceptor usage, and glycosylation. Proc Natl Acad Sci USA 2003; 100:12972–12977.
71. De Pasquale MP, Leigh Brown AJ, Uvin SC, Allega-Ingersoll J, Caliendo AM, et al. Differences in HIV-1 pol sequences from female genital tract and blood during antiretroviral therapy. J Acquir Immune Defic Syndr 2003; 34:37–44.
72. Adal M, Ayele W, Wolday D, Dagne K, Messele T, et al. Evidence of genetic variability of human immunodeficiency virus type 1 in plasma and cervicovaginal lavage in ethiopian women seeking care for sexually transmitted infections. AIDS Res Hum Retroviruses 2005; 21:649–653.
73. Tirado G, Jove G, Reyes E, Sepulveda G, Yamamura Y, et al. Differential evolution of cell-associated virus in blood and genital tract of HIV-infected females undergoing HAART. Virology 2005; 334:299–305.

74. Philpott S, Burger H, Tsoukas C, Foley B, Anastos K, et al. Human immunodeficiency virus type 1 genomic RNA sequences in the female genital tract and blood: compartmentalization and intrapatient recombination. J Virol 2005; 79:353–363.
75. Sullivan ST, Mandava U, Evans-Strickfaden T, Lennox JL, Ellerbrock TV, et al. Diversity, divergence, and evolution of cell-free human immunodeficiency virus type 1 in vaginal secretions and blood of chronically infected women: associations with immune status. J Virol 2005; 79:9799–9809.
76. Andreoletti L, Skrabal K, Perrin V, Chomont N, Saragosti S, et al. Genetic and phenotypic features of blood and genital viral populations of clinically asymptomatic and antiretroviral-treatment-naive clade a human immunodeficiency virus type 1-infected women. J Clin Microbiol 2007; 45:1838–1842.
77. Kemal KS, Burger H, Mayers D, Anastos K, Foley B, et al. HIV-1 drug resistance in variants from the female genital tract and plasma. J Infect Dis 2007; 195:535–545.
78. Semrau K, Ghosh M, Kankasa C, Sinkala M, Kasonde P, et al. Temporal and lateral dynamics of HIV shedding and elevated sodium in breast milk among HIV-positive mothers during the first 4 months of breast-feeding. J Acquir Immune Defic Syndr 2008; 47:320–328.
79. Bull ME, Learn GH, McElhone S, Hitti J, Lockhart D, et al. Monotypic human immunodeficiency virus type 1 genotypes across the uterine cervix and in blood suggest proliferation of cells with provirus. J Virol 2009; 83:6020-6028.
80. Dyer JR, Gilliam BL, Eron JJ, Jr, Grosso L, Cohen MS, et al. Quantitation of human immunodeficiency virus type 1 RNA in cell free seminal plasma: comparison of NASBA with Amplicor reverse transcription-PCR amplification and correlation with quantitative culture. J Virol Methods 1996; 60:161-170.
81. Si-Mohamed A, Kazatchkine MD, Heard I, Goujon C, Prazuck T, et al. Selection of drug-resistant variants in the female genital tract of human immunodeficiency virus type 1-infected women receiving antiretroviral therapy. J Infect Dis 2000; 182:112-122.
82. Min SS, Corbett AH, Rezk N, Cu-Uvin S, Fiscus SA, et al. Protease Inhibitor and Nonnucleoside Reverse Transcriptase Inhibitor Concentrations in the Genital Tract of HIV-1-Infected Women. J Acquir Immune Defic Syndr. 2004; 37:1577-1580.
83. Fowler MG, Lampe MA, Jamieson DJ, Kourtis AP, Rogers MF. Reducing the risk of mother-to-child human immunodeficiency virus transmission: past successes, current progress and challenges, and future directions. Am J Obstet Gynecol. 2007; 197:3-9.
84. Kuhn L, Aldrovandi GM, Sinkala M, Kankasa C, Semrau K, et al. Effects of Early, Abrupt Weaning for HIV-free Survival of Children in Zambia. N Engl J Med 2008; 359:1859.
85. Becquart P, Chomont N, Roques P, Ayouba A, Kazatchkine MD, et al. Compartmentalization of HIV-1 between breast milk and blood of HIV-infected mothers. Virology 2002; 300:109-117.
86. Becquart P, Courgnaud V, Willumsen J, Van de Perre P. Diversity of HIV-1 RNA and DNA in breast milk from HIV-1-infected mothers. Virology 2007; 363:256–260.
87. Henderson GJ, Hoffman NG, Ping LH, Fiscus SA, Hoffman IF, et al. HIV-1 populations in blood and breast milk are similar. Virology 2004; 330:295–303.
88. Andreotti M, Galluzzo CM, Guidotti G, Germano P, Altan AD, et al. Comparison of HIV type 1sequences from plasma, cell-free breast milk, and cell-associated breast milk viral

populations in treated and untreated women in Mozambique. AIDS Res Hum Retroviruses 2009; 25:707–711.
89. Sabbaj S, Edwards BH, Ghosh MK, Semrau K, Cheelo S, et al. Human immunodeficiency virus-specific CD8+ T cells in human breast milk. J Virol 2002; 76:7365–7373.
90. Sabbaj S, Ghosh MK, Edwards BH, Leeth R, Decker WD, et al. Breast milk-derived antigen-specific CD8+ T cells: an extralymphoid effector memory cell population in humans. J Immunol 2005; 174:2951–2956.
91. Nickle DC, Jensen MA, Shriner D, Brodie SJ, Frenkel LM, et al. Evolutionary indicators of human immunodeficiency virus type 1 reservoirs and compartments. J Virol 2003; 77:5540-5546.
92. Becquart P, Hakim H, Benoit G, Sépou A, Michel D, Bélec L. Compartmentalization of the IgG immune response to HIV-1 in breast milk. AIDS 1999; 13:1323-1331.
93. Heath L, et al. Restriction of HIV-1 Genotypes in Breast Milk Does Not Account for the Population Transmission Genetic Bottleneck That Occurs following Transmission. PLoS One 2010; 5:e10213.
94. Becquet R, et al. Universal antiretroviral therapy for pregnant and breast−feeding HIV−1−infected women: towards the elimination of mother−to-child transmission of HIV−1 in resource−limited settings. Clin Infect Dis 2009; 49:1936-1945.
95. Dabis F, Msellati P, Meda N, et al. 6−month efficacy, tolerance, and acceptability of a short regimen of oral zidovudine to reduce vertical transmission of HIV in breastfed children in Cote d'Ivoire and Burkina Faso: a double−blind placebo−controlled multicentre trial. Lancet 1999; 353:786–92.
96. Dabis F, Bequet L, Ekouevi DK, et al. Field efficacy of zidovudine, lamivudine and single-dose nevirapine to prevent peripartum transmission of HIV. The ANRS 1201/1202 Ditrame Plus study, Abidjan, Cote d'Ivoire. AIDS 2005; 19:309-18.
97. Guay LA, Musoke P, Fleming T, et al. Intrapartum and neonatal single−dose nevirapine compared with zidovudine for prevention of mother−to−child transmission of HIV−1 in Kampala, Uganda: HIVNET 012 randomised trial. Lancet 1999; 354:795–802.
98. Lallemant M, Jourdain G, Le Coeur S, et al. Single−dose perinatal nevirapine plus standard zidovudine to prevent mother−to−child transmission of HIV−1 in Thailand. N Engl J Med 2004; 351:217–28.
99. Petra study team. Efficacy of three short−course regimens of zidovudine and lamivudine in preventing early and late transmission of HIV−1 from mother to child in Tanzania, South Africa, and Uganda (Petra study): a randomised, double−blind, placebo−controlled trial. Lancet 2002; 359:1178–86.
100. Shaffer N, Chuachoowong R, Mock PA, et al. Short−course zidovudine for perinatal HIV−1 transmission in Bangkok, Thailand: a randomised controlled trial. Lancet 1999;353:773-80.
101. Wiktor SZ, Ekpini E, Karon JM, et al. Short−course oral zidovudine for prevention of mother−to−child transmission of HIV−1 in Abidjan, Cote d'Ivoire: a randomised trial. Lancet 1999;353:781-5.
102. Arrive E, Newell ML, Ekouevi DK, et al. Prevalence of resistance to nevirapine in mothers and children after single−dose exposure to prevent vertical transmission of HIV-1: a meta-analysis. Int J Epidemiol 2007; 36:1009–21.

103. Luzuriaga K, Dabis F, Excler JL, Sullivan J. Vaccines to prevent transmission of HIV-1 via breastmilk: scientific and logistical priorities. Lancet 2006; 368: 511-521.
104. Rerks-Ngarm S, et al. Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand. N Eng J Med 2009; 361:2209-2220.
105. Walker L, et al. Broad and Potent Neutralizing Antibodies from an African Donor Reveal a New HIV-1 Vaccine Target. Science 2009; 326:285-289.
106. Hansen S, et al. Effector memory T cell responses are associated with protection of rhesus monkeys from mucosal simian immunodeficiency virus challenge. Nature Med 2009; 15:293-299.
107. Koff W. Accelerating HIV vaccine development. Nature 2010; 464:161-162.

CONCLUSION

Molecular epidemiology, an intersection of biostatistics, computer science, evolutionary and molecular biology, traditional epidemiology, and technology innovation, is an exciting and powerful approach to characterization, surveillance, and clinical care for human diseases. The investigations described here provide examples of the sophisticated methods and powerful insights that molecular epidemiologic principles can bring to the field of HIV infection and public health. In particular, these approaches and insights are relevant for developing and developed, resource-limited and resource-abundant environments, and the power of these approaches, while universal, is dependent on multi-institutional and multi-disciplinary collaboration. The unique properties of HIV as an extraordinarily variable pathogen present critical challenges to antiretroviral treatment, HIV vaccine development and patient care. However, the variability of HIV also offers a substantial genetic footprint enabling investigation of transmission between regions and populations, between individuals, and vertically from mother to child.

As generation of large volumes of data continues to accompany biomedical progress, increasingly interdisciplinary approaches will be necessary to structure and manage these data in order to glean and bring to bear robust and reproducible insights on the challenges of human disease. The studies and literature review described here are intended not only to add to existing knowledge regarding HIV infection, but also to encourage this continued and necessary innovation and cross-disciplinary collaboration.