

The coreceptor mutation CCR5D32 influences the dynamics of HIV epidemics and is selected for by HIV

Amy D. Sullivan*, Janis Wigginton, and Denise Kirschner†

Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109-0620

Communicated by Avner Friedman, University of Minnesota, Minneapolis, MN, June 27, 2001 (received for review May 30, 2000)

We explore the impact of a host genetic factor on heterosexual HIV epidemics by using a deterministic mathematical model. A protective allele unequally distributed across populations is exemplified in our models by the 32-bp deletion in the host-cell chemokine receptor CCR5, CCR5D32. Individuals homozygous for CCR5D32 are protected against HIV infection whereas those heterozygous for CCR5D32 have lower pre-AIDS viral loads and delayed progression to AIDS. CCR5D32 may limit HIV spread by decreasing the probability of both risk of infection and infectiousness. In this work, we characterize epidemic HIV within three dynamic subpopulations: CCR5yCCR5 (homozygous, wild type), CCR5yCCR5D32 (heterozygous), and CCR5D32yCCR5D32 (homozygous, mutant). Our results indicate that prevalence of HIV/AIDS is greater in populations lacking the CCR5D32 alleles (homozygous wild types only) as compared with populations that include people heterozygous or homozygous for CCR5D32. Also, we show that HIV can provide selective pressure for CCR5D32, increasing the frequency of this allele.

Nineteen million people have died of AIDS since the discovery of HIV in the 1980s. In 1999 alone, 5.4 million people were newly infected with HIV (ref. 1 and <http://www.unaids.org/epidemicupdate/epireport/epireport.html>). (For brevity, HIV-1 is referred to as HIV in this paper.) Sub-Saharan Africa has been hardest hit, with more than 20% of the general population HIV-positive in some countries (2, 3). In comparison, heterosexual epidemics in developed, market-economy countries have not reached such severe levels. Factors contributing to the severity of the epidemic in economically developing countries abound, including economic, health, and social differences such as high levels of sexually transmitted diseases and a lack of prevention programs. However, the staggering rate at which the epidemic has spread in sub-Saharan Africa has not been adequately explained. The rate and severity of this epidemic also could indicate a greater underlying susceptibility to HIV attributable not only to sexually transmitted disease, economics, etc., but also to other more ubiquitous factors such as host genetics (4, 5).

To exemplify the contribution of such a host genetic factor to HIV prevalence trends, we consider a well-characterized 32-bp deletion in the host-cell chemokine receptor CCR5, CCR5D32. When HIV binds to host cells, it uses the CD4 receptor on the surface of host immune cells together with a coreceptor, mainly the CCR5 and CXCR4 chemokine receptors (6). Homozygous mutations for this 32-bp deletion offer almost complete protection from HIV infection, and heterozygous mutations are associated with lower pre-AIDS viral loads and delayed progression to AIDS (7–14). CCR5D32 generally is found in populations of European descent, with allelic frequencies ranging from 0 to 0.29 (13). African and Asian populations studied outside the United States or Europe appear to lack the CCR5D32 allele, with an allelic frequency of almost zero (5, 13). Thus, to understand the effects of a protective allele, we use a mathematical model to track prevalence of HIV in populations with or without CCR5D32 heterozygous and homozygous people and also to follow the CCR5D32 allelic frequency.

We hypothesize that CCR5D32 limits epidemic HIV by decreasing infection rates, and we evaluate the relative contributions to this by the probability of infection and duration of infectivity. To capture HIV infection as a chronic infectious disease together with

vertical transmission occurring in untreated mothers, we model a dynamic population (i.e., populations that vary in growth rates because of fluctuations in birth or death rates) based on realistic demographic characteristics (18). This scenario also allows tracking of the allelic frequencies over time. This work considers how a specific host genetic factor affecting HIV infectivity and viremia at the individual level might influence the epidemic in a dynamic population and how HIV exerts selective pressure, altering the frequency of this mutant allele.

CCR5 is a host-cell chemokine receptor, which is also used as a coreceptor by R5 strains of HIV that are generally acquired during sexual transmission (6, 19–25). As infection progresses to AIDS the virus expands its repertoire of potential coreceptors to include other CC-family and CXC-family receptors in roughly 50% of patients (19, 26, 27). CCR5D32 was identified in HIV-resistant people (28). Benefits to individuals from the mutation in this allele are as follows. Persons homozygous for the CCR5D32 mutation are almost nonexistent in HIV-infected populations (11, 12) (see ref. 13 for review). Persons heterozygous for the mutant allele (CCR5 WyD32) tend to have lower pre-AIDS viral loads. Aside from the beneficial effects that lower viral loads may have for individuals, there is also an altruistic effect, as transmission rates are reduced for individuals with low viral loads (as compared with, for example, AZT and other studies; ref. 29). Finally, individuals heterozygous for the mutant allele (CCR5 WyD32) also have a slower progression to AIDS than those homozygous for the wild-type allele (CCR5 WyW) (7–10), remaining in the population 2 years longer, on average. Interestingly, the dearth of information on HIV disease progression in people homozygous for the CCR5D32 allele (CCR5 D32yD32) stems from the rarity of HIV infection in this group (4, 12, 28). However, in case reports of HIV-infected CCR5 D32yD32 homozygotes, a rapid decline in CD4⁺ T cells and a high viremia are observed, likely because of initial infection with a more aggressive viral strain (such as X4 or R5X4) (30).

The Model

Because we are most concerned with understanding the severity of the epidemic in developing countries where the majority of infection is heterosexual, we consider a purely heterosexual model. To model the effects of the allele in the population, we examine the rate of HIV spread by using an enhanced susceptible-infected-AIDS model of epidemic HIV (for review see ref. 31). Our model compares two population scenarios: a CCR5 wild-type population and one with CCR5D32 heterozygotes and homozygotes in addition to the wild type. To model the scenario where there are only wild-type individuals present in the population (i.e., CCR5 WyW), we track the sexually active susceptibles at time t [$S_{ij}(t)$], where i 5 1 refers to genotype (CCR5 WyW only in this case) and j is either the male or female subpopulation. We also track those who are

*Present address: Centers for Disease Control and Prevention Epidemiology Program Office, State Branch Oregon Health Division, 800 NE Oregon Street, Suite 772, Portland, OR 97232.

†To whom reprint requests should be addressed. E-mail: kirschne@umich.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

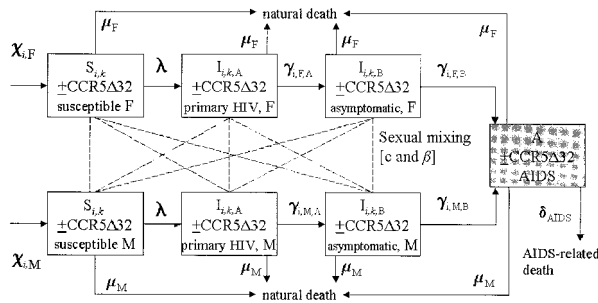


Fig. 1. A schematic representation of the basic compartmental HIV epidemic model. The criss-cross lines indicate the sexual mixing between different compartments. Each of these interactions has a positive probability of taking place; they also incorporate individual rates of transmission indicated as I , but in full notation is $I_{i,j,k \rightarrow i',j',k'}$, where i,j,k is the phenotype of the infected partner and i',j',k' is the phenotype of the susceptible partner. Also shown are the different rates of disease progression, $g_{i,j,k}$, that vary according to genotype, gender, and stage. Thus, the interactions between different genotypes, genders, and stages are associated with a unique probability of HIV infection. M, male; F, female.

HIV-positive at time t not yet having AIDS in $I_{i,j,k}(t)$ where k refers to stage of HIV infection [primary (A) or asymptomatic (B)]. The total number of individuals with AIDS at time t are tracked in $A(t)$. The source population are children, $x_{i,j}(t)$, who mature into the sexually active population at time t (Fig. 1, Table 1). We compare the model of a population lacking the CCR5D32 allele to a demographically similar population with a high frequency of the allele. When genetic heterogeneity is included, male and female subpopulations are each further divided into three distinct genotypic groups, yielding six susceptible subpopulations, $[S_{i,j}(t)]$, where i ranges from 1 to 3, where 1 \leq CCR5W; 2 \leq CCR5 WyD32; 3 \leq CCR5 D32yD32. The infected classes, $I_{i,j,k}(t)$, also increase in number to account for these new genotype compartments. In both settings we assume there is no treatment available and no knowledge of HIV status by people in the early acute and middle asymptomatic stages (both conditions exist in much of sub-Saharan Africa). In addition, we assume that sexual mixing in the population occurs randomly with respect to genotype and HIV disease status, all HIV-infected people eventually progress to AIDS, and no barrier contraceptives are used. These last assumptions reflect both economic and social conditions.

Parameter Estimates for the Model. Estimates for rates that govern the interactions depicted in Fig. 1 were derived from the extensive literature on HIV. Our parameters and their estimates are summarized in Tables 2–4. The general form of the equations describing the rates of transition between population classes as depicted in Fig. 1 are summarized as follows:

$$\frac{dS_{i,j}^{-t}}{dt} \leq x_{i,j}^{-t} \leq m_j S_{i,j}^{-t} \leq I_{i,j,k \rightarrow i,j} S_{i,j}^{-t}, \quad [1]$$

Table 1. Children's genotype

Parents	Mother		
Father	WyW	WyD32	D32yD32
WyW	$x_{1,j}$	$x_{1,j}, x_{2,j}$	$x_{2,j}$
WyD32	$x_{1,j}, x_{2,j}$	$x_{1,j}, x_{2,j}, x_{3,j}$	$x_{2,j}, x_{3,j}$
D32yD32	$x_{2,j}$	$x_{2,j}, x_{3,j}$	$x_{3,j}$

$x_{1,j}$ \leq wild-type children (WyW); $x_{2,j}$ \leq heterozygous children (WyD32); $x_{3,j}$ \leq homozygous children (D32yD32) of gender j . Children's genotypes are determined by using Mendelian inheritance patterns.

$$\frac{dI_{i,j,A}^{-t}}{dt} \leq I_{i,j,k \rightarrow i,j} S_{i,j}^{-t} \leq m_j I_{i,j,A}^{-t} \leq g_{i,j,A} I_{i,j,A}^{-t}, \quad [2]$$

$$\frac{dI_{i,j,B}^{-t}}{dt} \leq g_{i,j,A} I_{i,j,A}^{-t} \leq m_j I_{i,j,B}^{-t} \leq g_{i,j,B} I_{i,j,B}^{-t}, \quad [3]$$

$$\frac{dA^{-t}}{dt} \leq g_{i,j,B} I_{i,j,B}^{-t} \leq I_{i,F,B}^{-t} \leq I_{i,M,B}^{-t} \leq m_A A^{-t} \leq d_A A^{-t}, \quad [4]$$

where, in addition to previously defined populations and rates (with i equals genotype, j equals gender, and k equals stage of infection, either A or B), m_j represents the non-AIDS (natural) death rate for males and females respectively, and m_A is estimated by the average (m_F \leq m_M \leq m_A). This approximation allows us to simplify the model (only one AIDS compartment) without compromising the results, as most people with AIDS die of AIDS (d_{AIDS}) and very few of other causes (m_A). These estimates include values that affect infectivity ($I_{i,j,k \rightarrow i,j}$), transmission ($b_{i,j,k \rightarrow i,j}$), and disease progression ($g_{i,j,k}$) where the i,j,k notation represents the genotype, gender, and stage of infection of the infected partner, and $j \leq j$.

The effects of the CCR5 WyD32 and CCR5 D32yD32 genotypes are included in our model through both the per-capita probabilities of infection, $I_{i,j,k \rightarrow i,j}$, and the progression rates, $g_{i,j,k}$. The infectivity coefficients, $I_{i,j,k \rightarrow i,j}$, are calculated for each population subgroup based on the following: likelihood of HIV transmission in a sexual encounter between a susceptible and an infected ($b_{i,j,k \rightarrow i,j}$) person; formation of new partnerships (c_j); number of contacts in a given partnership (f_j); and probability of encountering an infected individual ($I_{i,j,k} N_j$). The formula representing this probability of infection is

$$I_{i,j,k \rightarrow i,j} \leq \frac{c_j \cdot f_j}{N_j} \left\{ \sum_{i,j,k} b_{i,j,k \rightarrow i,j} \cdot I_{i,j,k} \right\},$$

where $j \leq j$ is either male or female. N_j represents the total population of gender j (this does not include those with AIDS in the simulations).

The average rate of partner acquisition, c_j , includes the mean plus the variance to mean ratio of the relevant distribution of partnership-change rates to capture the small number of high-risk people: $c_j \leq m_j$ \leq ($s_j^2 y_m$) where the mean (m_j) and variance (s_j^2) are annual figures for new partnerships only (32). These means are estimated from Ugandan data for the number of heterosexual partners in the past year (33) and the number of nonregular heterosexual partners (i.e., spouses or long-term partners) in the past year (34). In these sexual activity surveys, men invariably have more new partnerships; thus, we assumed that they would have fewer average contacts per partnership than women (a higher rate of new partner acquisition means fewer sexual contacts with a given partner; ref. 35). To incorporate this assumption in our model, the male contacts/partnership, f_M , was reduced by 20%. In a given population, the numbers of heterosexual interactions must equate between males and females. The balancing equation applied here is $SA_F \cdot m_F \cdot N_F \leq SA_M \cdot m_M \cdot N_M$, where SA_j are the percent sexually active and N_j are the total in the populations for gender j . To specify changes in partner acquisition, we apply a male flexibility mechanism, holding the female rate of acquisition constant and allowing the male rates to vary (36, 37).

Transmission probabilities. The effect of a genetic factor in a model of HIV transmission can be included by reducing the transmission coefficient. The probabilities of transmission per contact with an infected partner, $b_{i,j,k \rightarrow i,j}$, have been estimated in the literature (see ref. 38 for estimates in minimally treated groups). We want to capture a decreased risk in transmission based on genotype (ref. 39, Table 2). No studies have directly evaluated differences in infectivity between HIV-infected CCR5 WyD32

Table 2. Transmission probabilities

HIV-infected partner (i, j, k)		Susceptible partner (i, j)			
		(j to j)	WyW	WyD32	D32yD32
Acute/primary					
WyW or D32yD32	M to F	0.040	0.040	0.00040	
	F to M	0.020	0.020	0.00020	
WyD32	M to F	0.030	0.030	0.00030	
	F to M	0.015	0.015	0.00015	
Asymptomatic					
WyW or D32yD32	M to F	0.0010	0.0010	10 $\times 10^{-26}$	
	F to M	0.0005	0.0005	5 $\times 10^{-26}$	
WyD32	M to F	0.0005	0.0005	5 $\times 10^{-26}$	
	F to M	0.00025	0.00025	2.5 $\times 10^{-26}$	

Listed are the different transmission probabilities ($b_{i,j,k \rightarrow i,j}$) for random sexual mixing between persons where i, j, k is the phenotype of the infected partner and i, j is the phenotype of the susceptible partner. M, male; F, female.

heterozygotes and HIV-infected CCR5 wild types. Thus, we base estimates for reduced transmission on studies of groups with various HIV serum viral loads (40), HTLV-IyII viral loads (41), and a study of the effect of AZT treatment on transmission (29). We decrease transmission probabilities for infecting CCR5D32yD32 persons by 100-fold to reflect the rarity of infections in these persons. However, we assume that infected CCR5D32yD32 homozygotes can infect susceptibles at a rate similar to CCR5WyW homozygotes, as the former generally have high viremias (ref. 30, Table 2). We also assume that male-to-female transmission is twice as efficient as female-to-male transmission (up to a 9-fold difference has been reported; ref. 42) (ref. 43, Table 2).

Given the assumption of no treatment, the high burden of disease in people with AIDS is assumed to greatly limit their sexual activity. Our initial model excludes people with AIDS from the sexually active groups. Subsequently, we allow persons with AIDS to be sexually active, fixing their transmission rates (b_{AIDS}) to be the same across all CCR5 genotypes, and lower than transmission rates for primary-stage infection (as the viral burden on average is not as high as during the acute phase), and larger than transmission rates for asymptomatic-stage infection (as the viral burden characteristically increases during the end stage of disease).

Disease progression. We assume three stages of HIV infection: primary (acute, stage A), asymptomatic HIV (stage B), and AIDS. The rates of transition through the first two stages are denoted by $g_{i,j,k}$, where i represents genotype, j is male/female, and k represents either stage A or stage B. Transition rates through each of these stages are assumed to be inversely proportional to the duration of that stage; however, other distributions are possible (31, 44, 45). Although viral loads generally peak in the first 2 months of infection, steady-state viral loads are established several months beyond this (46). For group A, the primary HIV-infecteds, duration is assumed to be 3.5 months. Based on results from European cohort studies (7–10), the beneficial effects of the CCR5 WyD32 genotype are observed mainly in the asymptomatic years of HIV infection; 7 years after seroconversion survival rates appear to be quite similar between heterozygous and homozygous individuals. We also assume that CCR5D32yD32-infected individuals and wild-type individuals progress similarly, and that men and women progress through each disease stage at the same rate. Given these observations, and that survival after infection may be shorter in untreated populations, we choose the duration time in stage B to be 6 years for wild-type individuals and 8 years for heterozygous individuals. Transition through AIDS, d_{AIDS} , is inversely proportional to the duration of AIDS. We estimate this value to be 1 year for the time from onset of AIDS to death. The progression rates are summarized in Table 3.

Table 3. Progression rates

Genotype	Disease stage	Males/females
WyW	A	3.5
	B	0.16667
WyD32	A	3.5
	B	0.125
D32yD32	A	3.5
	B	0.16667

Shown are the rates of progression, $g_{i,j,k}$ reflecting the different rates at which persons progress through different stages of disease by genotype, gender, and disease stage.

Demographic Setting. Demographic parameters are based on data from Malawi, Zimbabwe, and Botswana (3, 47). Estimated birth and child mortality rates are used to calculate the annual numbers of children ($x_{i,j}$) maturing into the potentially sexually active, susceptible group at the age of 15 years (3). For example, in the case where the mother is CCR5 wild type and the father is CCR5 wild type or heterozygous, the number of CCR5 WyW children is calculated as follows [suppressing (t) notation]: $x_{1,j} =$

$$B_r \frac{O}{k} \frac{F_{S_{1,F}}^{-S_{1,M}} \frac{1}{N_M} \frac{I_{1,M,k}!}{1} F_{-0.5!}^{-S_{1,F}} \frac{1}{N_M} \frac{I_{2,M,k}!}{1} G_1 \\ p_v \frac{S_{-1,F,k}^{-S_{1,M}} \frac{1}{N_M} \frac{I_{1,M,k}!}{1} F_{-0.5!}^{-S_{1,F}} \frac{1}{N_M} \frac{I_{2,M,k}!}{1} G_1$$

where the probability of HIV vertical transmission, $1/2 p_v$, and the birthrate, B_r , are both included in the equations together with the Mendelian inheritance values as presented in Table 1. The generalized version of this equation (i.e., $x_{i,j}$) can account for six categories of children (including gender and genotype). We assume that all children of all genotypes are at risk, although we can relax this condition if data become available to support vertical protection (e.g., ref. 48). All infected children are assumed to die before age 15. Before entering the susceptible group at age 15, there is additional loss because of mortality from all non-AIDS causes occurring less than 15 years of age at a rate of $m_x \times x_{i,j}$ (where m_x is the mortality under 15 years of age). Children then enter the population as susceptibles at an annual rate, $s_j \times x_{i,j}/15$, where s_j distributes the children 51% females and 49% males. All parameters and their values are summarized in Table 4.

Prevalence of HIV

Demographics and Model Validation. The model was validated by using parameters estimated from available demographic data. Simulations were run in the absence of HIV infection to compare the model with known population growth rates. Infection was subsequently introduced with an initial low HIV prevalence of 0.5% to capture early epidemic behavior.

In deciding on our initial values for parameters during infection, we use Joint United Nations Programme on HIV/AIDS national prevalence data for Malawi, Zimbabwe, and Botswana. Nationwide seroprevalence of HIV in these countries varies from 11% to over 20% (3), although there may be considerable variation within given subpopulations (2, 49).

In the absence of HIV infection, the annual percent population growth rate in the model is 2.5%, predicting the present-day values for an average of sub-Saharan African cities (data not shown). To validate the model with HIV infection, we compare our simulation of the HIV epidemic to existing prevalence data for Kenya and Mozambique (<http://www.who.int/yemc-hiv/yfact-sheets/ypdfsykenya.pdf> and ref. 51). Prevalence data collected from

these countries follow similar trajectories to those predicted by our model (Fig. 2).

Effects of the Allele on Prevalence. After validating the model in the wild type-only population, both CCR5D32 heterozygous and homozygous people are included. Parameter values for HIV transmission, duration of illness, and numbers of contacts per partner are assumed to be the same within both settings. We then calculate HIV yAIDS prevalence among adults for total HIVyAIDS cases.

Although CCR5D32yD32 homozygosity is rarely seen in HIV-positive populations (prevalence ranges between 0 and 0.004%), 1±20% of people in HIV-negative populations of European descent are homozygous. Thus, to evaluate the potential impact of CCR5D32, we estimate there are 19% CCR5 yD32 heterozygous and 1% CCR5 D32yD32 homozygous people in our population. These values are in Hardy-Weinberg equilibrium with an allelic frequency of the mutation as 0.105573.

Fig. 3 shows the prevalence of HIV in two populations: one lacking the mutant CCR5 allele and another carrying that allele. In the population lacking the protective mutation, prevalence increases logarithmically for the first 35 years of the epidemic, reaching 18% before leveling off.

In contrast, when a proportion of the population carries the CCR5D32 allele, the epidemic increases more slowly, but still logarithmically, for the first 50 years, and HIVyAIDS prevalence reaches 12% (Fig. 3). Prevalence begins to decline slowly after 70 years.

In the above simulations we assume that people with AIDS are not sexually active. However, when these individuals are included in the sexually active population the severity of the epidemic increases considerably (data not shown). Consistent with our initial simulations, prevalences are still relatively lower in the presence of the CCR5 mutation.

Because some parameters (e.g., rate constants) are difficult to estimate based on available data, we implement an uncertainty analysis to assess the variability in the model outcomes caused by any inaccuracies in estimates of the parameter values with regard to the effect of the allelic mutation. For these analyses we use Latin hypercube sampling, as described in refs. 52±56. Our uncertainty and sensitivity analyses focus on infectivity vs. duration of infectiousness. To this end, we assess the effects on the dynamics of the epidemic for a range of values of the parameters governing trans-

mission and progression rates $b_{i,j,k}$ and $g_{i,j,k}$. All other parameters are held constant. These results are presented as an interval band about the average simulation for the population carrying the CCR5D32 allele (Fig. 3). Although there is variability in the model outcomes, the analysis indicates that the overall model predictions are consistent for a wide range of transmission and progression rates. Further, most of the variation observed in the outcome is because of the transmission rates for both heterosexual males and females in the primary stage of infection ($b_{2,M,A3 i,F}$, $b_{2,F,A3 i,M}$). As mentioned above, we assume lower viral loads correlate with reduced infectivity; thus, the reduction in viral load in heterozygotes has a major influence on disease spread.

HIV Induces Selective Pressure on Genotype Frequency

To observe changes in the frequency of the CCR5D32 allele in a setting with HIV infection as compared with the Hardy-Weinberg

Fig. 2. Model simulation of HIV infection in a population lacking the protective CCR5D

