Title
Multi-Strain Virus-Host Dynamics from HIV to Phage

Author
Weinberger, Ariel

Publication Date
2011

Peer reviewed|Thesis/dissertation

Multi-Strain Virus-Host Dynamics from HIV to Phage

By

Ariel Dov Weinberger

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Biophysics

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor Harold Lecar (Chair)
Professor Arthur Reingold
Professor Wayne Getz
Professor Jillian Banfield

Spring 2011

Multi-Strain Virus-Host Dynamics from HIV to Phage

by

Ariel Dov Weinberger

Abstract

Multi-Strain Virus-Host Dynamics from HIV to Phage

by

Ariel Dov Weinberger

Doctor of Philosophy in Biophysics

University of California, Berkeley

Professor Harold Lecar, Chair

This dissertation uses mathematical modeling to probe the causes and consequences of multi-strain virus-host coexistence from the applied public health realm in which a second strain of HIV accelerates human mortality to the basic science realm in which persisting, previously dominant viruses drive the evolution of immunological memory in single-celled Bacteria and Archaea. In both applications, population-scale models are built from the ground up, utilizing experimentally measured parameters of virus and host molecules interacting within a cell to predict how virus and host populations coevolve across time. Model predictions are shown to match time-series data of virus-host dynamics in human hosts in the case of HIV and in prokaryotic hosts in the case of phage. Further, both sets of models generate experimentally testable hypotheses, suggesting therapeutic interventions against two major drivers of human mortality: HIV and pathogenic, antibiotic-resistant bacteria.

The first area of application, and the focus of Chapters 2 and 3, is HIV. No one understands why, in about 50% of HIV infections in the West, a more deadly HIV strain emerges late in infection. The new strain, known as X4, differs from its predecessor, known as R5, because X4 only infects $CD4^+$ T cells displaying the receptor CXCR4, while R5 only infects $CD4^+$ T cells displaying the receptor CCR5. Due to the apparent health and anti-HIV immunity of the approximately 10% of Europeans lacking a functional CCR5 receptor, some researchers have touted anti-R5 therapy as an alternative to current anti-HIV drug cocktails. Chapter 2 uses simulations of a novel mathematical model to show how anti-R5 treatment alone may accelerate X4 emergence and resultant immunodeficiency. As an alternative, I show that CCR5 blockers may be more successful in combination with effective HAART therapy or, should they become available, CXCR4 blockers. Chapter 3 probes why X4 only emerges during late-stage HIV infection, showing how X4 persists for many years at low levels, avoiding competitive exclusion to the initially fitter R5 Virus, through X4's unique, low-level infection of naïve CD4+ T cells. In this chapter, I derive a minimal target-cell based model for dual R5, X4 HIV infection in which late-stage switches (bifurcations) to X4 autonomously occur. In this simplified model, an analytic switch condition is probed, allowing us to theoretically predict how different interventions modulate the time to X4 emergence, and providing a compelling explanation for why 50% of Western HIV patients never actually switch to X4 Virus.

In Chapter 4, the focus turns to understanding the evolution of adaptive antiviral immunity—such as the T cell immunity targeted by HIV—in its most elemental setting: single-celled prokaryotes possessing the CRISPR immune system. A novel mathematical model of virus-microbe coevolution is derived to understand why prokaryotes with CRISPR-encoded specific immunity conserve old immune sequences for thousands of microbial generations despite compact prokaryotic genomes with high DNA deletion rates and rapid viral mutation, which makes old CRISPR sequences far less likely to provide immunity against current viruses. Matching metagenomic reconstructions of CRISPR sequences across time in both bacterial and archaeal populations, the model shows how CRISPRs' immunological memory protects against measured blooms of persisting, low-abundance viral sequences. Thus, CRISPR may be the first immune system tuned viruses persisting through lysogeny or remigration.

Chapter 1

Introduction: Viral Diversity from Archaea to Humans

Ariel D. Weinberger

Whether infecting the smallest bacteria or the largest mammals, viruses face a daunting survival challenge. To endure, viruses must find and colonize external host organisms to reproduce before natural forces take their toll. Yet, lacking an internal reproductive capability has not prevented viruses from flourishing: they represent the most abundant and genetically diverse entities of the biosphere (Edwards and Rohwer, 2005). Viral promiscuity and reproductive efficiency come at a cost. With insufficient hosts for each extant virion, viruses inevitably compete with one another in the hunt for productive hosts. How do so many genetically diverse viruses survive given the limited resources which nature offers?

One would expect the relative death of hosts to impart a genetic bottleneck on viral populations, driving viral lines less suited to a given environment extinct at the hands of more locally efficient competition. Extinctions of less fit exploiters were first experimentally noted over eighty years ago, when the ecologist Georgii F. Gause found that a particular set of experimental conditions could only support one of two lines of competing paramecia (Gause, 1932). In the race for limited resources, one line continually exploited a disproportionate share of environmental nutrients, eventually driving the less-fit paramecia to extinction in a process that came to be known as 'competitive exclusion.' Importantly, by changing experimental conditions Gause was able to reverse the winning and losing strains. Gause's experimental demonstrations matched the long-run theoretical predictions of the well-known Lotka-Volterra equations mathematically idealizing predator-prey dynamics (Gause, 1934). The confluence of theory and laboratory experiment supporting the unique success of a single exploiter in a given environment—whether that exploiter is a virus predating organismal prey or an organism exploiting resources—led to the growing acceptance of competitive exclusion as a general principle in ecology (Hardin, 1960).

Yet, even early on, it was noted that nature did not always bottleneck exploiter diversity. Among the competing phytoplankton of the sea, many species were shown to coexist despite each of these organisms exploiting the same limited abundance solar and chemical nutrients in relatively well-mixed (i.e., spatially homogeneous) open waters. G. Evleyn Hutchinson termed the disparity between measured planktonic aquatic diversity and the predictions of the competitive exclusion principle as 'the paradox of the plankton' (Hutchinson, 1961). A number of mechanisms have since been proposed for the lack of competitive exclusion in natural environments (Wilson, 1990). These explanations range from resource fluctuations, which delay convergence to single exploiter steady-states, to spatial heterogeneity, which allows for ecological sub-niches in which distinct species persist.

This dissertation offers a more fundamental explanation for coexistence in the viral world, showing how inherent genotypic diversity among host cells preserves viral diversity in two distinct model systems. Thus, even in spatially homogeneous, idealized systems converging to steady states, distinct viral genotypes may optimally reproduce in distinct host lines. Such virus-host specializations are either virus or host driven: either *viral lines* only efficiently colonize a fraction of the hosts (viral tropism differences) or *host lines* only efficiently resist a fraction of the viruses (host immunity differences). Respectively, Chapters 2 and 3 show how a more deadly and treatment-resistant phenotype of HIV emerges because some CD4+ T cells lack a key receptor required by the less deadly wild-type HIV strain of HIV, while Chapter 4 analyzes

how differential antiviral immunity among otherwise clonal prokaryotes controls phage diversity in natural and laboratory populations.

Viral Multi-Tropism in HIV

The Human Immunodeficiency Virus type-1 (HIV) has been implicated in over 30 million deaths. The virus is so fatal, because it targets and depletes $CD4^+$ T cells, 'helper' immune cells critical for orchestrating and stimulating the overall immune response (Haase, 1999). No one understands why in about 50% of late-stage Western HIV infections (the rate is curiously much lower in sub-saharan Africa), a new, more deadly strain known as X4 emerges (Regoes and Bonhoeffer, 2005). X4, differs from the initial R5 Virus, because X4 only infects $CD4^+$ T cells displaying the CXCR4 receptor, while R5 infects $CD4^+$ T cells displaying CCR5. Unlike CCR5, the CXCR4 receptor is also found on naïve $CD4^+$ T cells, allowing X4 to deplete a second set of critical immune cells, accelerating immunodeficiency and death (Moore et al., 2004).

A small percentage of humans lack a functional CCR5 receptor due to a 32 base-pair deletion in the gene coding for CCR5 (Sabeti et al., 2005). In addition to remaining generally side-effect free, CCR5Δ32/Δ32 individuals are surprisingly immune to both strains of HIV, only being infected by X4 in rare cases (Moore et al., 2004; Sheppard et al., 2002). Their apparent heath and immunity has led the FDA to begin approving small inhibitor molecules to selectively block the CCR5 receptor, with some researchers touting anti-R5 therapy alone as a potentially safer alternative to current anti-HIV drug cocktails (Fatkenheuer et al., 2005; Gulick et al., 2008; Veazey et al., 2003). In fact, a well-publicized bone-marrow transplant (Hutter et al., 2009) from a CCR5-negative individual transformed a previously HIV-positive transplantee to being HIV free for the past three years (Allers et al.). The success and promise of anti-CCR5 treatments aside, an open question is whether these therapies might push HIV toward the competing, more deadly X4 variant in dually-infected individuals (Westby and van der Ryst, 2005).

To probe whether anti-CCR5 therapies promote X4 emergence and attendant early immunodeficiency in patients harboring even small traces of X4 Virus, the second chapter of this thesis presents a novel mathematical model of dual-strain HIV infection. The model uniquely determines R5 and X4 tropism parameters through *ex vivo* flow cytometry measurements of the per-cell CXCR4 and CCR5 concentrations on both naïve and memory CD4+ T cells. Simulations show how anti-R5 treatments alone may in fact accelerate deadly X4 emergence and attendant immunodeficiency in individuals harboring even trace, undetectable levels of X4 Virus. The early switch occurs through X4 infecting a large fraction of CD4+ T cells previously preferentially infected by R5 Virus. Fortunately, the use of CXCR4 blockers or HAART in conjunction with CCR5 blockers appears to eliminate this risk of accelerated immunodeficiency. With X4 on the rise throughout sub-Saharan Africa, this chapter's results caution against the increased use of anti-CCR5 monotherapy.

While Chapter 2 shows how the late-stage emergence of X4 may complicate new treatment paradigms, a more basic question is why X4 merely emerges during late-stage HIV infection? Can X4 really persist at low levels for years all the while avoiding extinction to an initially fitter R5, or does X4 accumulate *de novo* via mutation in each individual? Chapter 3

uses a combination of mathematical theory and simulation to show how X4 can coexist at low levels for long periods of time alongside the wild-type R5 Virus by capitalizing on low-level virus production from the niche of generally quiescent but CCR5- naïve CD4+ T cells. The conditions under which switches to X4 occur are mathematically derived, yielding a minimal target-cell based model for dual R5, X4 HIV infection in which late-stage switches to X4 Virus autonomously occur (i.e., bifurcations). In this simplified model, an analytic switch condition is found and used to show how different treatment schemes modulate the time to X4 emergence. The theoretical results derived in this chapter explain the predictions of Chapter 2.

Specific Immunity in Bacteria and Archaea

While the coexistence of both X4 and R5 in 50% of Western HIV patients arises due to intrinsic coreceptor differences among HIV's CD4 target cells, one can find high levels of viral diversity in phages infecting phenotypically identical prokaryotes only differing in a non-protein coding region of DNA. This prokaryotic DNA region contains the newly discovered Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), a microbial adaptive immune system found in approximately 40% of sequenced Bacteria and approximately 90% of sequenced Archaea. CRISPR loci provide prokaryotes with specific immunity by serially acquiring new short DNA sequences (there can be hundreds per locus) from invading viral and plasmid genomes, using these foreign-derived sequences to cleave and target matching sequences in subsequent genomic invasions. In Chapter 4, I describe a computational model that tracks genetic coevolution within CRISPR loci, explaining how viral diversity emerges from host immune diversity. With dominant viral lines more likely to interact with host cells and thus insert their sequences into the CRISPR pangenome, hosts more rapidly gain immunity to frequent viruses. Model results show how such increased immunity imparts negative frequency-dependent selection on the viral population, preventing dominant viral lineages from commonly excluding their competitors.

Strikingly, viral diversity is not matched among the old immune elements of host CRISPR loci. In fact, the oldest immune elements are generally identical across a population, implying that—contrary to the predictions of the prevailing 'Kill the Winner' paradigm in microbial ecology (Rodriguez-Valera et al., 2009; Thingstad and Lignell, 1997)—current hosts are the descendants of a single ancestral lineage. And more striking than the lack of diversity among old CRISPR sequences, is that old immune elements are preserved in the first place. Ancestral spacers appear to be maintained by hosts for thousands of microbial generations—immunological memory—despite compact prokaryotic genomes with high DNA deletion rates and rapid viral mutation, which makes old CRISPR sequences far less likely to be effective against current viruses. The model I built of virus-host coevolution explains why old spacers emerge clonal and conserved. Matching new metagenomic time-series reconstructions of CRISPR loci across bacterial and archaeal populations, model simulations capture CRISPR's unique immune memory protecting against measured blooms of persisting, low-abundance viral sequences. By maintaining genomically costly old immunities, CRISPR may the first immune system tuned against viruses persisting through lysogeny or remigration.

Implications

This dissertation applies data-driven mathematical models to understand the drivers of viral diversity in both HIV and phage. With two diverse immunological applications and a focus on the predictions of theoretical models, I feel that the work adds important contributions to the well-established field of Evolutionary Dynamics. But the eventual goal is for these and future works to have import in the burgeoning field of Evolutionary Medicine. That is, beyond the theoretical question of how pathogenic diversity is maintained is the medical question of whether such diversity may render the previous century's paradigms of infectious disease control—antibiotics and vaccination—unfit for the majority of pathogens (Lipsitch and O'Hagan, 2007). Most pathogens, as the failure to vaccinate against HIV implies, may simply mutate too facilely to be controlled by even vaccination schemes targeting a number of viral antigens (Telford, 2008). An alternative therapeutic approach may be to utilize pathogenic diversity against itself, by challenging pathogenic variants with more efficient, but ultimately less pathogenic, strains. CRISPR offers a prime candidate. Insofar as CRISPR loci prevent insertions by plasmids as well as viruses, in principle they could block the plasmid-borne transmission of pathogenicity islands and antibiotic resistance. In fact, a recent paper has shown a strong inverse correlation between antibiotic resistance and CRISPR presence across almost 50 bacterial lines (Palmer and Gilmore). Thus, the basic science question of what selects for CRISPR-driven immunity over other competing immune systems may have tremendous practical applications. One envisions replacing antibiotic resistant, CRISPR-negative bacteria with safe, treatable strains containing CRISPR loci. The model presented in Chapter 4 offers one potential way of doing so, arguing that CRISPR is disproportionately found in Archaea, because CRISPR's unique immunological memory is critical against the long-lived, non-lytic phages in archaeal environments. A potential way to impose similar selection for immunological memory in Bacteria would be via phage therapy with persistent, temperate phages. In that way, temperate phage therapy may offer a way to select for CRISPR-laden Bacteria impervious to the deadly spread of antibiotic resistance and pathogenicity. Thus, better understanding the co-dynamics of host and virus and more critically, perhaps, the internal dynamics among competing viruses may offer a relatively unpaved path to potential new treatment paradigms.

References

Allers, K., Hutter, G., Hofmann, J., Loddenkemper, C., Rieger, K., Thiel, E., and Schneider, T., Evidence for the cure of HIV infection by CCR5{Delta}32/{Delta}32 stem cell transplantation. Blood.

Edwards, R.A., and Rohwer, F., 2005. Viral metagenomics. Nat Rev Microbiol 3, 504-10.

Fatkenheuer, G., Pozniak, A.L., Johnson, M.A., Plettenberg, A., Staszewski, S., Hoepelman, A.I., Saag, M.S., Goebel, F.D., Rockstroh, J.K., Dezube, B.J., Jenkins, T.M., Medhurst, C., Sullivan, J.F., Ridgway, C., Abel, S., James, I.T., Youle, M., and van der Ryst, E., 2005. Efficacy of short-term monotherapy with maraviroc, a new CCR5 antagonist, in patients infected with HIV-1. Nat Med 11, 1170-2.

Gause, G.F., 1932. Experimental studies on the struggle for existence. Journal of Experimental Biology 9, 389-402.

Gause, G.F., 1934. Experimental Analysis of Vito Volterra's Mathematical Theory of the Struggle for Existence. Science 79, 16-7.

Gulick, R.M., Lalezari, J., Goodrich, J., Clumeck, N., DeJesus, E., Horban, A., Nadler, J., Clotet, B., Karlsson, A., Wohlfeiler, M., Montana, J.B., McHale, M., Sullivan, J., Ridgway, C., Felstead, S., Dunne, M.W., van der Ryst, E., and Mayer, H., 2008. Maraviroc for previously treated patients with R5 HIV-1 infection. N Engl J Med 359, 1429-41.

Haase, A.T., 1999. Population biology of HIV-1 infection: viral and CD4+ T cell demographics and dynamics in lymphatic tissues. Annu Rev Immunol 17, 625-56.

Hardin, G., 1960. The competitive exclusion principle. Science 131, 1292-7.

Hutchinson, G.E., 1961. The paradox of the plankton. American Naturalist 95, 137-145.

Hutter, G., Nowak, D., Mossner, M., Ganepola, S., Mussig, A., Allers, K., Schneider, T., Hofmann, J., Kucherer, C., Blau, O., Blau, I.W., Hofmann, W.K., and Thiel, E., 2009. Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. N Engl J Med 360, 692-8.

Lipsitch, M., and O'Hagan, J.J., 2007. Patterns of antigenic diversity and the mechanisms that maintain them. J R Soc Interface 4, 787-802.

Moore, J.P., Kitchen, S.G., Pugach, P., and Zack, J.A., 2004. The CCR5 and CXCR4 coreceptors--central to understanding the transmission and pathogenesis of human immunodeficiency virus type 1 infection. AIDS Res Hum Retroviruses 20, 111-26.

Palmer, K.L., and Gilmore, M.S., Multidrug-Resistant Enterococci Lack CRISPR-cas. MBio 1.

Regoes, R.R., and Bonhoeffer, S., 2005. The HIV coreceptor switch: a population dynamical perspective. Trends Microbiol 13, 269-77.

Rodriguez-Valera, F., Martin-Cuadrado, A.B., Rodriguez-Brito, B., Pasic, L., Thingstad, T.F., Rohwer, F., and Mira, A., 2009. Explaining microbial population genomics through phage predation. Nat Rev Microbiol 7, 828-36.

Sabeti, P.C., Walsh, E., Schaffner, S.F., Varilly, P., Fry, B., Hutcheson, H.B., Cullen, M., Mikkelsen, T.S., Roy, J., Patterson, N., Cooper, R., Reich, D., Altshuler, D., O'Brien, S., and Lander, E.S., 2005. The case for selection at CCR5-Delta32. PLoS Biol 3, e378.

Sheppard, H.W., Celum, C., Michael, N.L., O'Brien, S., Dean, M., Carrington, M., Dondero, D., and Buchbinder, S.P., 2002. HIV-1 infection in individuals with the CCR5-Delta32/Delta32 genotype: acquisition of syncytium-inducing virus at seroconversion. J Acquir Immune Defic Syndr 29, 307-13.

Telford, J.L., 2008. Bacterial genome variability and its impact on vaccine design. Cell Host Microbe 3, 408-16.

Thingstad, T.F., and Lignell, R., 1997. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. Aquat. Microb. Ecol. 13, 19-27.

Veazey, R.S., Klasse, P.J., Ketas, T.J., Reeves, J.D., Piatak, M., Jr., Kunstman, K., Kuhmann, S.E., Marx, P.A., Lifson, J.D., Dufour, J., Mefford, M., Pandrea, I., Wolinsky, S.M., Doms, R.W., DeMartino, J.A., Siciliano, S.J., Lyons, K., Springer, M.S., and Moore, J.P., 2003. Use of a small molecule CCR5 inhibitor in macaques to treat simian immunodeficiency virus infection or prevent simian-human immunodeficiency virus infection. J Exp Med 198, 1551-62.

Westby, M., and van der Ryst, E., 2005. CCR5 antagonists: host-targeted antivirals for the treatment of HIV infection. Antivir Chem Chemother 16, 339-54.

Wilson, J.B., 1990. Mechanisms Of Species Coexistence: Twelve Explanations For Hutchinson's 'Paradox Of The Plankton': Evidence From New Zealand Plant Communities. New Zealand Journal of Ecology 13, 17-42.

Chapter Two

Accelerated Immunodeficiency By Anti-CCR5 Treatment in HIV Infection

Ariel D. Weinberger, Alan S. Perelson, Ruy M. Ribeiro & Leor S. Weinberger

INTRODUCTION

Left untreated, the human immunodeficiency virus type-1 (HIV) generally targets and severely depletes a patient's $CD4^+$ T cells over a period of 10-12 years leading to AIDS onset and death (Douek et al., 2003; Ho et al., 1995; Morgan et al., 2002; Wei et al., 1995). HIV's infection of a $CD4^+$ T cell begins when the virus' outer envelope protein gp120 binds to a cell's CD4 receptor and subsequently to one of two cellular coreceptors, CCR5 or CXCR4 (Berger et al., 1999; Moore et al., 2004). Viral-coreceptor binding enables fusion of the viral and target-cell membranes, allowing HIV to inject its retroviral material into the cell. HIV strains that use CCR5 as a coreceptor are termed R5 viruses, while those that bind CXCR4 are called X4 viruses.

R5 virus is predominant during early infection where X4 virus has rarely been observed, independent of the route of viral transmission (Casper et al., 2002; Cornelissen et al., 1995; Harouse et al., 2003; Moore et al., 2004). Importantly, X4 *alone* is generally unable to infect humans: individuals homozygous for a 32 base-pair deletion in CCR5, CCR5Δ32, are almost entirely immune to HIV (Moore et al., 2004). However, in approximately 50% of progressing HIV patients, X4 virus emerges late in infection, overtaking R5 virus as the dominant viral strain. The R5-to-X4 switch is strongly associated with a poor clinical prognosis for the patient: it occurs with a steep loss in $CD4^+$ T cell counts and accelerated AIDS onset. In fact, X4's emergence explains earlier work noting a "phenotypic switch" to a more virulent viral phenotype in many late-stage HIV patients (Blaak et al., 2000; Penn et al., 1999; Richman and Bozzette, 1994; Schuitemaker et al., 1992; Spijkerman et al., 1995).

The mechanisms causing R5's early dominance and the subsequent R5-to-X4 switch are poorly understood, however multiple lines of evidence suggest that CCR5's higher cell-surface density on activated and recently activated memory $CD4^+$ T cells enable R5 to infect more of this crucial cellular population than X4. CCR5's cell-surface density has been shown to determine the efficiency of R5 infection (Lin et al., 2002), possibly because multiple CCR5 receptors act in a cooperative, concentration-dependent manner to facilitate infection (Kuhmann et al., 2000). R5 virus' level of infection is thus highest among $CD62L^-$ effector memory $CD4^+$ T cells (Gondois-Rey et al., 2006), where CCR5's cell surface density is highest. CXCR4's cell-surface density is similarly positively correlated with X4's emergence (Lin et al., 2005), but CXCR4's per-cell density on memory $CD4^+$ T cells is lower than that of CCR5 (Lee et al., 1999), giving R5 an advantage on these cells. On dually-positive $CCR5^+$, $CXCR4^+$ $CD4^+$ T cells, the coreceptors compete for association with CD4 (Lee et al., 2000), which should lend R5 an advantage given CCR5's higher per-cell surface density on dually-positive cells (Lee et al., 1999).

Thus, R5 virus' early advantage may stem from CCR5's greater *per-cell* surface density on activated and recently activated 'effector' memory $CD4^+$ T cells (Lee et al., 1999; Lee et al., 2000). These 'effector' memory $CD4^+$ T cells are the crucial virion-producing populations as evidenced by snapshots taken during SIV infection, which show approximately five times as many virions surrounding infected, activated effector memory $CD4^+$ T cells as around infected, quiescent $CD4^+$ T cells (Zhang et al., 2004). Moreover, Li et al. show that $CD4^+$ T cells positive for Ki67 (a marker that is displayed after late G1 cell-cycle progression and indicates T cell 'activation') produce over 90% of the virions during the chronic phase of SIV infection (Li et al., 2005). This may also explain why X4 has trouble initiating infection when R5 virus is absent:

CXCR4's per-cell density on the most crucial memory $CD4^+$ T cell population is simply too low (Lee et al., 1999).

The perplexing question underlying the R5-to-X4 phenotypic switch is therefore: how does a switch to X4 occur if R5 virus is simply better at infecting memory $CD4^+$ T cells? Since the R5-to-X4 switch only occurs during late infection, it is reasonable that there exists an early selection pressure in favor of R5 virus, which is mitigated over the course of infection. In support of this hypothesis, Ribeiro and colleagues recently proposed the idea that increasing target-cell activation over the course of dual infection causes X4 to eventually outcompete R5 (Ribeiro et al., 2006).

A critical prediction of the Ribeiro framework is that CCR5 blockers (small-molecule pharmaceuticals that bind CCR5 and thereby obstruct R5 virus' ability to infect a $CD4^+$ T cell) successfully reduce overall viral loads, decrease cellular activation levels, and inhibit X4 emergence. This prediction is critical since a central question is whether CCR5 blockers lend X4 virus an advantage and promote clinically deleterious switches to X4 during dual R5 and X4 infection. However, *in vivo* trials of the CCR5 inhibitors *CMPD 167* and *maraviroc* showed CCR5 blockers actually increasing X4 viral loads and decreasing R5 viral loads (approximately reciprocally) in dually-infected patients (Westby et al., 2006; Wolinsky et al., 2004).

Given the recent CCR5 clinical trial data, we analytically probed how changing target cell activation levels could produce a switch and whether such models could account for documented increases in X4 viral load after anti-CCR5 treatment. Our model builds upon (Ribeiro et al., 2006), but in our generalized setup the R5-to-X4 switch can occur even if the fraction of activated naïve $CD4^+$ T cells increases at a slower rate than the fraction of activated memory $CD4^+$ T cells. In this more general setting, we rigorously show how the R5 to X4 switch occurs and find that CCR5 blockers often do accelerate X4's emergence and attendant immunodeficiency. Fortunately, the results also show that when CXCR4 inhibitors or HAART are given along with CCR5 inhibitors, X4 emergence is unlikely to be accelerated and is instead often delayed.

MODELS

In the following three models, all variables are capitalized and represent concentrations per microliter (1/μl). Specifically, in Model 1, T represents the concentration of uninfected $CD4^+$ T cells, and (without loss of generality) is given an initial value of 1000 $CD4^+$ T cells/μl. In Models 2 and 3, T is split into uninfected naïve (N) and memory (M) subpopulations, each with an initial value of 500 $CD4^+$ T cells/μl. I_4 and I_5 reflect the concentrations of *abortively, latently, and productively* infected $CD4^+$ T cells by X4 and R5, respectively, and in Model 3, we analogously define N_4, M_4, M_5 (see below). V_4 and V_5, each given initial values of 1000 virions/ml, represent X4 and R5 viral load concentrations.

Defining the parameters, λ is the rate of thymus production of $CD4^+$ T cells and has units cells/(μl•day), k_4 and k_5 are the respective infection rate coefficients for X4 and R5 infection of $CD4^+$ T cells and have the units μl/(virions•day). All remaining parameters have units 1/day.

These include: d_T, the death rate of uninfected CD4$^+$ T cells in Model 1, set to λ/T_0 to allow for steady-state pre-infection, and d_n and d_m the analogous death rates of uninfected CD4$^+$ T cells in Models 2 and 3, also set so that equilibrium exists pre-infection. Additionally, δ is the death rate of infected CD4$^+$ T cells, p is the rate of viral production by *activated* infected cells, and c is the viral clearance rate. a_n and a_m are required to satisfy Equation 2 (below) and represent the fractions of *activated* naïve and memory CD4$^+$ T cells as a function of CD4, the total number of uninfected and infected CD4$^+$ T cells per microliter. Thus, in Models 1 and 2, CD4 = $T + I_4 + I_5$, and, analogously, in Model 3 CD4 = $N + M + N_4 + M_4 + M_5$. Since the total concentration of CD4$^+$ T cells changes over time, a_n and a_m vary over the course of infection.

Because over 99% of infected cells are defectively infected (Haase, 1999) and because such non-productively infected cells are indistinguishable from uninfected cells, we make the simplifying assumption that a_n and a_m also approximate the fractions of *infected* naïve and memory CD4$^+$ T cells that are activated. Thus, in Models 1 and 2, a_nI_4 and a_mI_5 represent the concentrations of *activated* X4 and R5 infected cells, respectively. Analogously, in Model 3, a_nN_4, a_mM_4, and $a_m\mathrm{M}_5$ represent the concentrations of *activated* X4-infected naïve, X4-infected memory, and R5-infected memory CD4$^+$ T cells, respectively. In our models, it is only these activated subpopulations of infected cells that produce virions. We thus multiply the concentration of activated infected cells (e.g., $a_m\mathrm{M}_5$) by p, the rate of viral production (per-day) from a *productively*-infected (i.e., activated and infected) cell, yielding the respective total concentration of virions produced each day by a given infected cell type.

Model 1: Single Target Cell Compartment

We first extended the basic model of viral dynamics (Ho et al., 1995; Wei et al., 1995) to two viral strains, to test whether this simplified, one-compartment model can generate a representative R5-to-X4 switch.

$$\begin{aligned}
\dot{T} &= \lambda - (k_4V_4 + k_5V_5)T - d_TT \\
\dot{I}_4 &= k_4V_4T - \delta I_4 \\
\dot{I}_5 &= k_5V_5T - \delta I_5 \\
\dot{V}_4 &= pa_nI_4 - cV_4 \\
\dot{V}_5 &= pa_mI_5 - cV_5
\end{aligned} \qquad \text{(Model 1)}$$

Here we make X4's viral production dependent on the fraction of activated naïve CD4$^+$ T cells a_n, but not on a_m. One reason for this simplification is that R5 out-competes X4 for dually-positive memory CD4$^+$ T cells (Roy et al., 2005). Furthermore, the vast majority of CXCR4-positive T cells are in the naïve subset, where CXCR4's cell surface density is also highest (Lee et al., 1999). Since Model 1 lumps all CD4$^+$ T cells into a single target-cell compartment, and because across all lymphocytes CXCR4's median per-cell surface density is almost three times as high as that of CCR5 (Lee et al., 1999), we also assume $k_4 > k_5$. As above, this does not imply that X4 *productively* infects more target cells than R5 at the beginning of infection, since very few naïve cells are activated early in infection (Hazenberg et al., 2000). Importantly, given the simplifications employed, the purpose of Model 1 is not to represent the actual dynamics of

coreceptor tropism in HIV infection, but to rigorously explore an activation-based R5 to X4 switch in the simplest setting.

Model 2: Two Target Cell Compartments

To account for the fact that in reality naïve and memory $CD4^+$ T cells are disjoint target cell compartments, we subsequently build upon Model I and divide T into N and M.

$$\begin{aligned}
\dot{N} &= \lambda + (1-2f)a_n N - k_4 V_4 N - d_N N \\
\dot{M} &= 2fa_n N + a_m M - k_5 V_5 M - d_M M \\
\dot{I}_4 &= k_4 V_4 N - \delta I_4 \\
\dot{I}_5 &= k_5 V_5 M - \delta I_5 \\
\dot{V}_4 &= p a_n I_4 - c V_4 \\
\dot{V}_5 &= p a_m I_5 - c V_5
\end{aligned} \qquad \text{(Model 2)}$$

The equations in this system are analogous to those in Model 1 but the uninfected $CD4^+$ T cell population is now split into uninfected naïve (N) and memory (M) subpopulations. Additionally, f is defined to be the fraction of naïve cells activated via the conventional Ag-TCR interaction, which divide and differentiate into $CD45RO^+$ memory cells. The rest of the activated cells are assumed to have been upregulated via cytokines or other Ag-TCR independent processes and thus remain phenotypically naïve ($CD45RA^+$) (Suarez et al., 2002; Unutmaz et al., 1994; Unutmaz et al., 1995). We note that non-Ag mediated activation of naïve $CD4^+$ T cells is not absolutely necessary for our models' primary conclusions of strain coexistence and phenotypic switching at clinically-representative time-points (i.e., 3-6 years post-infection); we include this activation term for the added realism it brings to the model.

In Model 2, X4 is only able to infect naïve $CD4^+$ T cells, a simplification we employ because of the data in (Lee et al., 1999) showing that the per-cell density of CCR5 is significantly higher than that of CXCR4 on memory $CD4^+$ T cells. Moreover, a recent paper finds that on dually-positive $CXCR4^+$, $CCR5^+$ $CD4^+$ T cells, R5 generally outcompetes X4 (Roy et al., 2005), arguably because of CCR5's higher surface density (Lee et al., 2000). Finally, naïve $CD4^+$ T cells have been found to be preferentially depleted during X4 infection (Nishimura et al., 2005).

Model 3: Two Target Cell Compartments with Viral Competition

Because in practice X4 actually infects both naïve and memory $CD4^+$ T cells, in our final model, Model 3, we extend the two-compartment setup of Model 2 to allow X4's infection of memory $CD4^+$ T cells:

$$
\begin{aligned}
\dot{N} &= \lambda + (1-2f)a_n N - k_{N4}V_4 N - d_N N \\
\dot{M} &= 2fa_n N + a_m M - k_{M4}V_4 M - k_{M5}V_5 M - d_M M \\
\dot{N}_4 &= k_{N4}V_4 N - \delta N_4 \\
\dot{M}_4 &= k_{M4}V_4 M - \delta M_4 \\
\dot{M}_5 &= k_{M5}V_5 M - \delta M_5 \\
\dot{V}_4 &= p(a_n N_4 + a_m M_4) - cV_4 \\
\dot{V}_5 &= pa_m M_5 - cV_5
\end{aligned}
\qquad \text{(Model 3)}
$$

In this model, k_{N4} and k_{M4} are the infection rate coefficients of X4 on naïve (N) and memory (M) $CD4^+$ T cells, respectively, and k_{M5} is the infection rate coefficient of R5 on memory $CD4^+$ T cells. k_{N4}, k_{M4}, and k_{M5} all have units µl/(virions•day). N_4 and M_4 are the concentrations of *abortively, latently, and productively* infected naïve and memory $CD4^+$ T cells, respectively, by X4 virus, and M_5 is the concentration of *abortively, latently, and productively* infected memory $CD4^+$ T cells by R5 virus. All other parameters, variables, and initial conditions have been defined above. Because CCR5 is far more strongly expressed on memory $CD4^+$ T cells than is CXCR4 (Lee et al., 1999), we set $k_{M5} >> k_{M4}$. Conversely, CXCR4 is more highly expressed on naïve $CD4^+$ T cells than it is on memory $CD4^+$ T cells (Lee et al., 1999), making $k_{N4} >> k_{M4}$.

Generalized Conditions for a_n and a_m

HIV is associated with increasing levels of $CD4^+$ T cell activation (Hazenberg et al., 2000; Hazenberg et al., 2003; Mohri et al., 2001; Sachsenberg et al., 1998). Curve fitting *in vivo* data from (Hazenberg et al., 2000), Ribeiro et al. (Ribeiro et al., 2006) found that the fractions of phenotypically-activated ($Ki67^+$) naïve (a_n) and memory $CD4^+$ T cells (a_m) have the following inverse relationships to the total $CD4^+$ T cell count per microliter (denoted CD4 in the equations below):

$$
\begin{aligned}
a_n(CD4) &= \frac{10}{CD4} - .0095 \\
a_m(CD4) &= \frac{10}{CD4} + .05
\end{aligned}
\qquad \text{Eq. (1)}
$$

Rather than restrict ourselves to an analysis based on Eq. (1), we only assume that a_n and a_m obey three general conditions for all $CD4^+$ T cell counts:

$$
\begin{aligned}
&i)\ a_n(CD4) < a_m(CD4) \\
&ii)\ a_n'(CD4) < 0,\ \ a_m'(CD4) < 0 \\
&iii)\ \frac{d}{dCD4}\left(\frac{a_n(CD4)}{a_m(CD4)}\right) < 0
\end{aligned}
\qquad \text{Eq. (2)}
$$

In other words, i) the fraction of activated cells is assumed to always be higher among memory $CD4^+$ T cells than among naïve $CD4^+$ T cells, ii) both fractions are assumed to be increasing as

$CD4^+$ T cell counts decline, and iii) as $CD4^+$ T cells are depleted, the fraction of activated naïve $CD4^+$ T cells increases *relative* to the fraction of activated memory $CD4^+$ T cells. Importantly, the relative fraction a_n/a_m can increase even when a_n increases at a slower rate than a_m in response to $CD4^+$ T cell decline.

RESULTS

A Single Compartment Model Generates an R5 to X4 Switch Without Coexistence

In single target-cell compartment susceptible-infectious (SI) models such as Model 1, the ecological principle of *competitive exclusion* generally applies (Ball et al., 2007). Thus, while Model 1 can produce an R5 to X4 switch in a clinically representative timeframe, it necessarily manifests competitive exclusion (Fig 1a). The lack of steady-state coexistence in Model 1 is significantly different from data, which show long-term coexistence of R5 and X4 variants in post-switch individuals (Philpott et al., 2001). Moreover, X4's emergence late in infection—well into quasi-steady state—is very difficult to achieve in this single compartment framework, because X4 could have been rendered extinct via *competitive exclusion* prior to the late-stage switch (Weinberger and Perelson, *manuscript in preparation*).

Two Target Cell Populations Can Produce R5 and X4 Coexistence

In order to prevent the species with the higher effective reproductive ratio from dominating *exclusively*, which contradicts observed results (Philpott et al., 2001), Model 2 splits the target cell population into naïve and memory $CD4^+$ T cells, and, for simplicity, assumes that X4 solely infects naive cells and that R5 only infects memory cells. The dual-target cell compartment nature of Model 2 makes coexistence possible (Weinberger and Perelson, 2009, *manuscript in preparation*). Thus, while Model 2 can also produce an R5-to-X4 switch at a clinically representative time, it is able to maintain R5 and X4 coexistence post-switch (Figure 1B).

However, CCR5 inhibition cannot produce a transient increase in X4 (Figure 1C). This is because in models that restrict X4 and R5 to infecting distinct target cell populations (e.g., Model 2), X4 does not infect any of the (memory) $CCR5^+$ T cells that are made refractory to R5 infection by CCR5 inhibition. Quantitatively, Eq. (2) stipulates $a_n'(CD4) < 0$, so when CCR5 is inhibited and memory $CD4^+$ T cell counts rise, a_n decreases and the rate of viral production from an X4-infected cell ($p*a_n$) is lowered. Due to the lack of competition, the number of X4-infected cells does not increase to compensate for the decreased per-cell virion production rate, so X4 viral loads decrease (Figure S1). This result is in contrast to recent studies on dually-infected *rhesus macaques* and humans, which demonstrate clear increases in X4 virus after R5 virus is selectively suppressed through the use of a CCR5 inhibitor (Westby et al., 2006; Wolinsky et al., 2004). To produce a temporal X4 increase upon R5 inhibition and to maintain coexistence in contradistinction to Model 1, we need a multi-compartment model where X4 infects both naïve and memory CD4+ T cells.

Two Target Cell Compartments With Viral Competition Allow Coexistence and Match Existing Data

In Model 3, our final and most biologically detailed model, we include naïve and memory $CD4^+$ T cell compartments. Since CXCR4 is found on a large number of memory $CD4^+$ T cells, we allow for X4's infection of memory as well as naïve $CD4^+$ T cells (Figure 2A). Thus, Model 3 serves as a union of the two previous models: it includes the X4 and R5 strain competition found in Model 1 and it also includes the separate target cell compartments of Model 2, which allowed for X4's persistence prior to a switch and the coexistence of strains afterward. Model 3 produces X4-to-R5 switches at clinically representative times of 1000-2000 days and also maintains coexistence post-switch in two types of parameter regimes, the "non-competitive" and "competitive" regimes, whose distinctions are elaborated upon below (Figure 2B).

Given that X4 and R5 viruses can coexist in the disjoint two-compartment model, it is reasonable to conjecture that coexistence is a feature of this extended model as well. To show this formally, we define R_{eff4} and R_{eff5} to be the *effective* reproductive ratios of X4 and R5 virus, respectively, which are given by:

$$R_{eff4}(t) = p*(a_n(CD4(t)*k_{N4}*N(t) + a_m(CD4(t))*k_{M4}*M(t))/(c*\delta_i)$$

$$R_{eff5}(t) = p*(a_m(CD4(t))*k_{M5}*M(t))/(c*\delta_i) \qquad \text{Eq. (3)}$$

$$R_{eff4}/R_{eff5}(t) = (k_{N4}/k_{M5})*(a_n(CD4(t))/a_m(CD4(t)))*(N(t)/M(t)) + k_{M4}/k_{M5}$$

The *effective* reproductive ratio, R_{eff}, is thus a time-dependent function for the average number of infected cells produced by an average infected cell at a given point, t, in time. R_{eff} generalizes R_0, the *basic* reproductive ratio, which evaluates the average infectivity only at the initial time point.

Solving the necessary and sufficient conditions for an R5-to-X4 switch, $d/dt(V_4(t^*)) > d/dt(V_5(t^*))$ and $V_4(t^*) = V_5(t^*)$, we see that a switch occurs in Model 3 if and only if:

$$(a_n(CD4(t^*))/a_m(CD4(t^*))) > (M_5(t^*)-M_4(t^*))/N_4(t^*) \qquad \text{Eq. (4)}$$

But $a_m > a_n$ for all time, so, in particular, at the switch time t* we have $a_n(CD4(t^*))/a_m(CD4(t^*)) < 1$. The right-hand side of Equation (4) must therefore be less than 1, meaning that at the switch point $N_4(t^*) + M_4(t^*) > M_5(t^*)$. Thus, at the switch point t* there are more X4-infected $CD4^+$ T cells than R5-infected $CD4^+$ T cells. This implies that X4 had a higher effective reproductive ratio at some earlier point, t**. However, $R_{eff4}(t^{**}) > R_{eff5}(t^{**})$ does not imply that $R_{eff4}(t) > R_{eff5}(t)$ for all t > t**: Equation (3) implies that when *N* decreases faster than *M and* when the resulting decrease to *N/M* is less than the increase to a_n/a_m, R_{eff5} increases relative to R_{eff4} (i.e., R_{eff4}/R_{eff5} decreases). But the condition for steady-state coexistence of X4 and R5 is $R_{eff4} = R_{eff5}$, so by enabling R_{eff5} to rebound relative to R_{eff4} post-switch, the dual-compartment nature of Model 3 makes coexistence possible.

We can grasp the switch threshold in (4) more easily by substituting in the particular equations of (1), yielding the following switch condition (where the right-hand side is positive):

$$CD4(t^*) < 200\frac{N_4(t^*) + M_4(t^*) - M_5(t^*)}{M_5(t^*) - M_4(t^*) + .19N_4(t^*)} \qquad \text{(Eq. 5)}$$

Importantly, Equations (3) and (5) imply that, with the exception of changes to k_{M5}, modulating parameters to accelerate $CD4^+$ T cell decline hastens an R5 to X4 switch while changing parameters to mitigate $CD4^+$ T cell decline hinders a phenotypic switch. Thus, successful antiretroviral therapy will generally inhibit X4's emergence. However, because R5 and X4 are now in competition, CCR5 inhibitors generate more complicated kinetics.

CCR5 Inhibitors Can Accelerate X4 Emergence: the Need for CXCR4 Inhibitors or HAART

CCR5 inhibitors decrease k_{M5}, causing R5's viral load to decline, and, as a result, memory $CD4^+$ T cell counts to increase. The question we sought to answer is whether X4 infects sufficiently many of these R5-immune memory $CD4^+$ T cells to counteract the increase in $CD4^+$ T cells from CCR5 inhibition. We hypothesized that X4's ability to infect memory CD4 T cells would depend on k_{M4}, and with a sufficiently large k_{M4} (the "competitive regime"), X4 would infect a non-negligible fraction of newly R5-immune cells which and an increase in X4 would ensue. The temporal increase in X4 viral loads would thus cause greatly increased X4 infection of naïve $CD4^+$ T cells, which yields accelerated naïve $CD4^+$ T cell depletion. Indeed, numerical simulations (Text S1 contains more information on how these simulations were done) show that successful CCR5 blockage results in *accelerated* AIDS onset across much of parameter space (Figure 2C). This result does not change when Model 3 is extended to include the loss of virions due to the infection of new target cells (Figure S2). Importantly, the early immunodeficiency after effective CCR5 blockage is due to accelerating X4 emergence and increasing X4 viral loads as the efficacy of CCR5 inhibition increases in the "competitive regime" (Figure 2D).

Conversely, if k_{M4} is sufficiently small (the "non-competitive regime"), X4 does not infect a sufficient number of dually-positive memory $CD4^+$ T cells upon CCR5 blockage. This causes the uninfected memory $CD4^+$ T cell population to increase during anti-CCR5 therapy, yielding a drop in a_n/a_m and hindering a potential switch to X4 as well as immunodeficiency (Figure 2C, small k_{M4} regime). The latter result is to be expected from Model 2, because a weak k_{M4} can be approximated by a complete lack of competition. Thus, a single parameter, k_{M4}, controls the efficacy of anti-CCR5 therapy in dually infected HIV patients, highlighting the need for circumspection in prescribing these treatments.

Given that CCR5 inhibitors accelerate R5-to-X4 switching and immunodeficiency across the wide swath of parameter space in which k_{M4} is relatively large, the question arises as to whether CCR5 inhibitors are similarly deleterious when used in conjunction with CXCR4 inhibitors, which reduce k_{M4}. Simulations show that adding a CXCR4 inhibitor with an efficacy of at least 5% is sufficient to prevent accelerated AIDS onset in the "competitive regime" (Figure 3A). Because X4 emergence is due to an increase in the relative fraction a_n/a_m of activated naïve

to memory $CD4^+$ T cells, we also simulated whether a generic antiretroviral therapy such as HAART, which increases $CD4^+$ T cell counts and reduces a_n/a_m, also prevents the accelerated X4 emergence that CCR5 inhibitors can engender. The results of dual-treatment with HAART and CCR5 inhibitors are analogous to those shown for dual-treatment with CXCR4 and CCR5 inhibitors, proving that a relatively modest additional HAART therapy (with an efficacy above 7%) obviates the risk of CCR5 inhibitors accelerating immunodeficiency in the "competitive regime" (Figure 3B). Finally, generalizing across k_{M4} and k_{M5} parameter space shows that when treatment efficacies are sufficiently strong (e.g. 80% efficacies) dual treatment with CXCR4 inhibitors does not accelerate immunodeficiency relative to untreated individuals (Figure 3C). Similarly, dual-treatment with CCR5 inhibitors and HAART does not accelerate immunodeficiency relative to untreated individuals (Figure 3D).

DISCUSSION

Here we present a mathematical model of dual-strain R5 and X4 HIV *in vivo* dynamics and show that CCR5 inhibitors can accelerate the emergence of X4 virus and immunodeficiency. Two equivalent R5-to-X4 switch conditions were found: either the ratio of the relative fractions of activated naïve and memory $CD4^+$ T cells (a_n/a_m) must surpass a threshold (Eq. 4) or, equivalently, $CD4^+$ T cell counts must drop below a critical value (Eq. 5). The resultant "phenotypic" switch yields a drastic loss in $CD4^+$ T cell counts, due to X4's depletion of R5-immune naïve $CD4^+$ T cells. Of significant clinical importance, our results show that, across much of parameter space, CCR5 inhibitors may force an *early* switch to X4 virus, greatly accelerating $CD4^+$ T cell depletion and AIDS onset. However, CCR5 inhibitors do not appear to have the deleterious effect of accelerating X4 emergence and immunodeficiency when they are used in conjunction with CXCR4 inhibitors or HAART.

The result that CCR5 blockers alone may promote X4 emergence is supported by data from a study on dually-infected *rhesus macaques* injected intravenously with the CCR5 inhibitor CMPD 167 (Wolinsky et al., 2004). After beginning treatment, two out of three primates manifested a transient increase of several logs in X4 viral load, essentially canceling the decrease in R5 viral load. Moreover, a clerical error in a recent study on the effect of the CCR5 inhibitor *maraviroc* on R5-only patients resulted in a dually-infected patient mistakenly being included in the trial (Fatkenheuer et al., 2005). That patient saw no change in total viral load as the X4 viral load increased upon CCR5 inhibition (Westby et al., 2006).

While CCR5 inhibitors alone may accelerate X4 emergence and AIDS onset, there is still good reason to consider their utility as part of a multi-therapy cocktail. Recent clinical data from the MOTIVATE 1&2 trials show that CCR5 inhibitors together with optimized background therapy yield larger increases in CD4 counts and larger reductions in viral loads when compared with optimized background therapy alone (Gulick et al., 2008). Our model simulations strongly support this result, showing that across much of parameter space, employing CCR5 inhibitors together with HAART lengthens the time to AIDS when compared with the time to AIDS under HAART alone (Figure S3).

But even if CCR5 inhibitors are a helpful component in a diversified anti-HIV therapy, one has to wonder about the greater immunological cost associated with blocking this chemokine receptor. A recent meta-population analysis of West Nile Virus (WNV) prevalence in four US states found that CCR5Δ32 homozygotes are approximately four times more likely to develop symptomatic WNV as are those with the wild-type receptor (Lim et al., 2008). Previous murine models have suggested a mechanism by which CCR5 confers protective advantage against symptomatic WNV: CCR5 may promote the transfer of leukocytes to a WNV-infected individual's brain, aiding in immune control of encephalitis (Glass et al., 2005). CCR5's potential protective advantage against symptomatic WNV may also help explain why CCR5Δ32/Δ32 is relatively common (5-14%) among European Caucasian cohorts, but near absent in African populations (Sabeti et al., 2005), the latter being at a far greater risk of contracting WNV.

Additionally, it is important to consider the prospect that CCR5 inhibition may lead to HIV evolving to bind to an entirely new coreceptor during early infection. A precedent for the evolution of new lentiviral coreceptor tropisms exists: the SIV endemic to *red-capped mangabeys* (RCMs) can utilize CCR2b rather than CCR5 (Chen et al., 1998). This is likely because a large percentage (estimated at over 80%) of RCMs are homozygous for a 24 base-pair deletion in the gene for CCR5, and CCR5Δ24/Δ24 cells cannot be transfected with R5 virus (Chen et al., 1998). The ability of SIVrcm to use CCR2b occurs despite almost all other known SIVs utilizing CCR5 exclusively *in vivo* (Moore et al., 2004). Δ24 appears to be an ancient deletion: it has been found in both *red-capped mangabeys* and *sooty mangabeys*, species which diverged more than 10,000 years ago (Chen et al., 1998). It is therefore possible that in the long-run HIV may evolve entirely new coreceptor usages in response to coreceptor inhibition.

Acknowledgements

We thank H. Dahari, H. Green, P. Hraber, L. Rong, R., A. Singh, and W. Turner for extensive discussions and are grateful to members of the Weinberger (UCSD) and Getz (UC Berkeley) laboratories and the UCSD Center for AIDS Research for helpful comments. ADW especially wishes to thank P. Bickel, W. Getz and H. Lecar for very insightful feedb

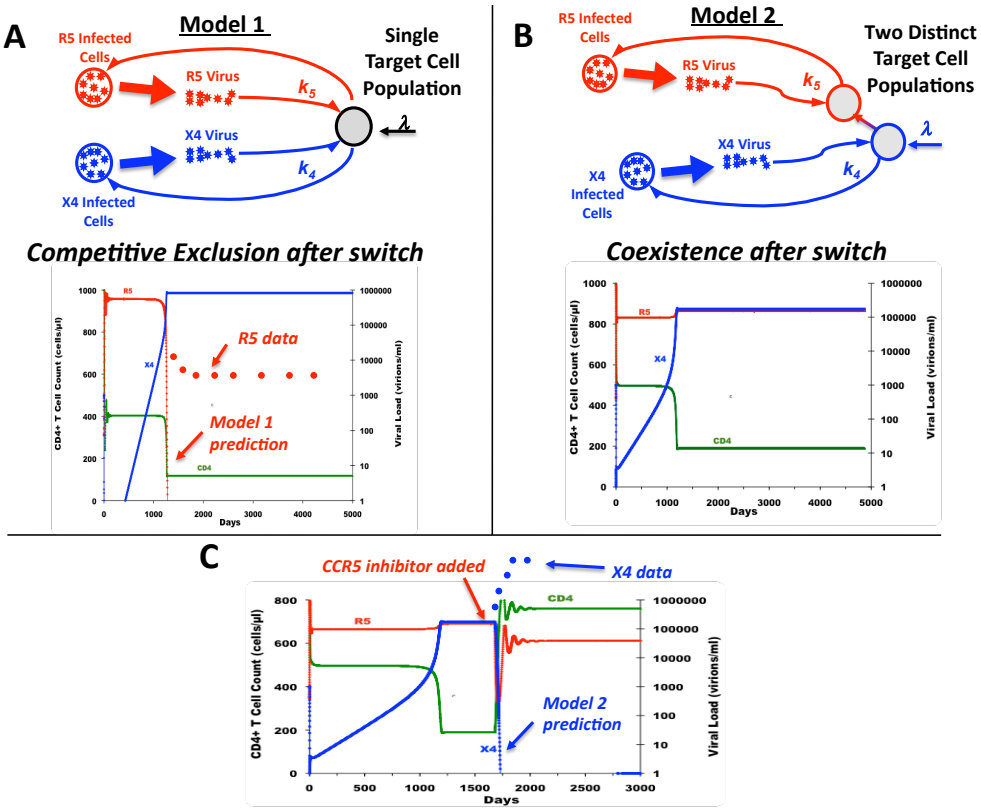
Figure Legends

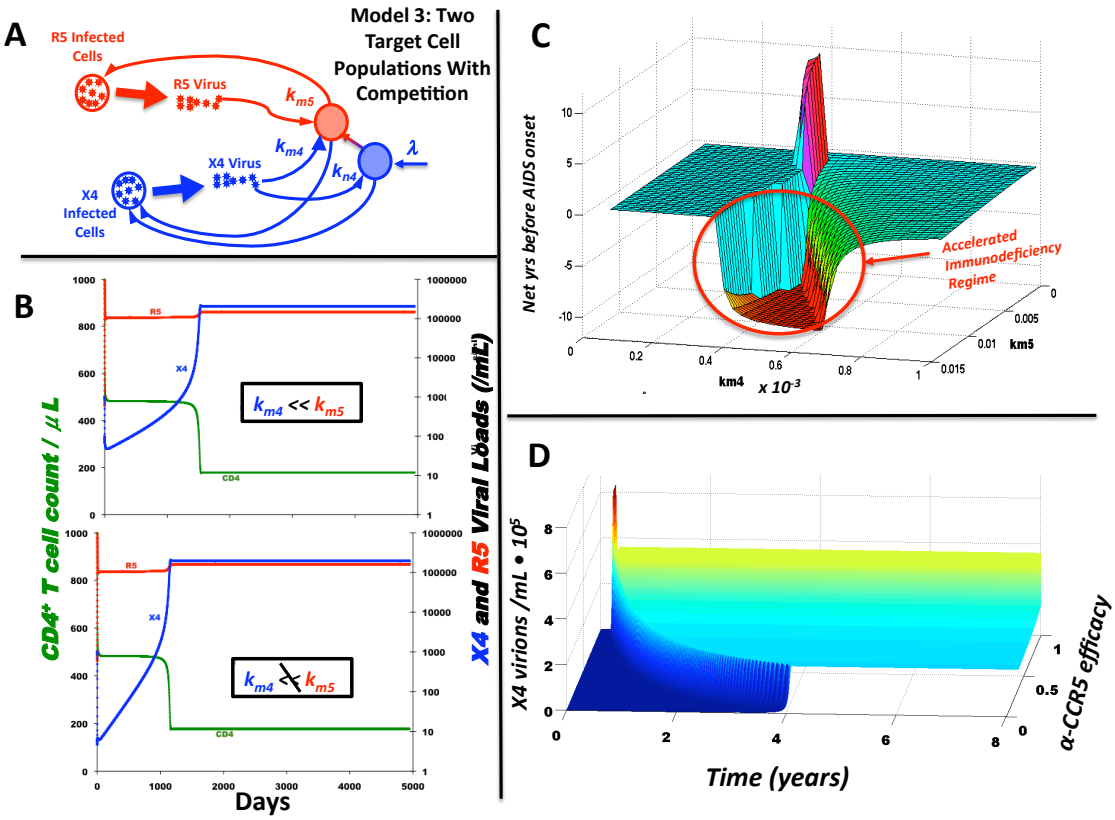
Figure 1: Models assuming a single target-cell population for X4 and R5, and models where X4 and R5 infect distinct populations cannot account for X4 and R5 data. (A) A pedagogical model (schematic and numerical simulations) that is oversimplified to account for only a single target-cell population generates a competitively exclusive R5-to-X4 switch where R5 virus is cleared (i.e. goes extinct) following the switch, contrary to *in vivo* data (Philpott et al., 2001). The model simulates a "phenotypic" switch occurring at a clinically representative time of 3-4 years post HIV-1 infection, and, notably, yields a concomitant decline in $CD4^+$ T cell counts. The parameters used are λ = 33 cells/(μl•day), c = 23/day, p = 5750/day, δ = 0.7/day, k_4 =5•10^{-4} μl/(virions•day), and k_5 = 10^{-4} μl/(virions•day). (B) A model restricting R5 and X4 to disparate target cell compartments can generate a clinically representative R5 to X4 switch over a large parameter regime and also exhibit coexistence of R5 and X4 post-switch. However, as shown in (C), such models cannot account for *in vivo* data showing that R5 inhibitors increase X4 levels (Wolinsky et al., 2004). In (C), we apply a CCR5 blocker with a drug efficacy of 0.9, starting at t=180 days. The model restricts R5 and X4 to independent target cell compartments, and, given the absence of viral competition, always generates strong suppression of X4 in response to CCR5 inhibition. Simulations in (B) and (C) are shown for a representative parameter regime: λ = 33 cells/(μl•day), c = 23/day, p = 2000/day, f = 0.8, δ = 0.5/day, k_4 = 0.0012 μl/(virions•day), and k_5 = 0.0034 μl/(virions•day).

Figure 2: A model with two target-cell populations and viral competition in one of these populations matches *in vivo* data and predicts that R5 inhibitors accelerate AIDS onset. (A) A schematic of Model 3, a competitive model with two target-cell populations. (B) The model exhibits a coreceptor switch at approximately 1000-1400 days post-infection in two types of parameter regimes: the "non-competitive regime" in (B, upper panel) and the "competitive regime" (B, lower panel). For the "non-competitive regime," parameter values are: λ=33 cells/(μl•day), c = 23/day, p = 2100/day, f = 0.8, δ = 0.5/day, k_{N4} = 0.00108 μl/(virions•day), k_{M4} = 4•10^{-5} μl/(virions•day), and k_{M5} = 0.0068 μl/(virions•day), while in the "competitive regime" we change k_{M4} to 5•10^{-4} μl/(virions•day) and k_{N4} to 0.001 μl/(virions•day) (decreased in order to keep X4 in check). (C) The net effect of a CCR5 inhibitor with 80% efficacy on the time to AIDS across different k_{M4} and k_{M5} levels (i.e. different competitive and non-competitive regimes). AIDS onset is defined as the time at which $CD4^+$ T cell counts fall below 200 cells/μl: negative values represent accelerated times to AIDS-onset relative to no treatment. R5 inhibitors clearly accelerate AIDS-onset for a large fraction of parameter space. (D) Time-dependency of X4 emergence as a function of CCR5 inhibitor efficacy (α-CCR5 efficacy) in the "competitive regime." Increased CCR5 inhibitor efficacy accelerates X4 emergence and increases X4's viral set point.

Figure 3: Combination treatment with HAART or CXCR4 inhibitors can prevent CCR5 inhibitors from accelerating AIDS-onset. (A) CXCR4 inhibitors with an efficacy of at least 5% prevent CCR5 inhibitors from accelerating time to AIDS in the "competitive regime." Relative to no treatment, the net time to AIDS onset is positive (i.e. AIDS onset is never

accelerated) for any CCR5 inhibitor accuracy if CXCR4 inhibitors have an accuracy of at least 10%. (B) Similarly, HAART with an efficacy of at least 7% prevents CCR5 inhibitors from accelerating time to AIDS (i.e., having a negative net time) in the competitive regime. (C) Net time to AIDS with dual-treatment using CCR5 and CXCR4 inhibitors, each with 80% efficacy, across different values of k_{M4} and k_{M5} relative to no treatment. The net time to AIDS onset is never negative for dual treatment with CCR5 and CXCR4 inhibitors (i.e. AIDS onset never accelerated) when the inhibitors have strong efficacies. (D) The net effect of dual-treatment with CCR5 inhibitors and HAART, each at 80% efficacy, on the time to AIDS at different values of k_{M4} and k_{M5}. The net time to AIDS onset is never negative for dual treatment with CCR5 inhibitors plus HAART (i.e. AIDS onset never accelerated) when the inhibitors and HAART have strong efficacies.


A
Model 1
R5 Infected Cells
R5 Virus
k_5
Single Target Cell Population
λ
X4 Virus
k_4
X4 Infected Cells
Competitive Exclusion after switch
R5 data
Model 1 prediction
CD4+ T Cell Count (cells/µl)
Viral Load (virions/ml)
Days
B
Model 2
R5 Infected Cells
R5 Virus
k_5
Two Distinct Target Cell Populations
λ
X4 Virus
k_4
X4 Infected Cells
Coexistence after switch
CD4+ T Cell Count (cells/µl)
Viral Load (virions/ml)
Days
C
CCR5 inhibitor added
X4 data
Model 2 prediction
CD4+ T Cell Count (cells/µl)
Viral Load (virions/ml)
Days


A
Model 3: Two Target Cell Populations With Competition
R5 Infected Cells
R5 Virus
X4 Virus
X4 Infected Cells
B
CD4+ T cell count / μL
X4 and R5 Viral Loads (/mL)
Days
C
Net yrs before AIDS onset
Accelerated immunodeficiency Regime
km4
km5
D
X4 virions /mL • 10^5
Time (years)
α-CCR5 efficacy

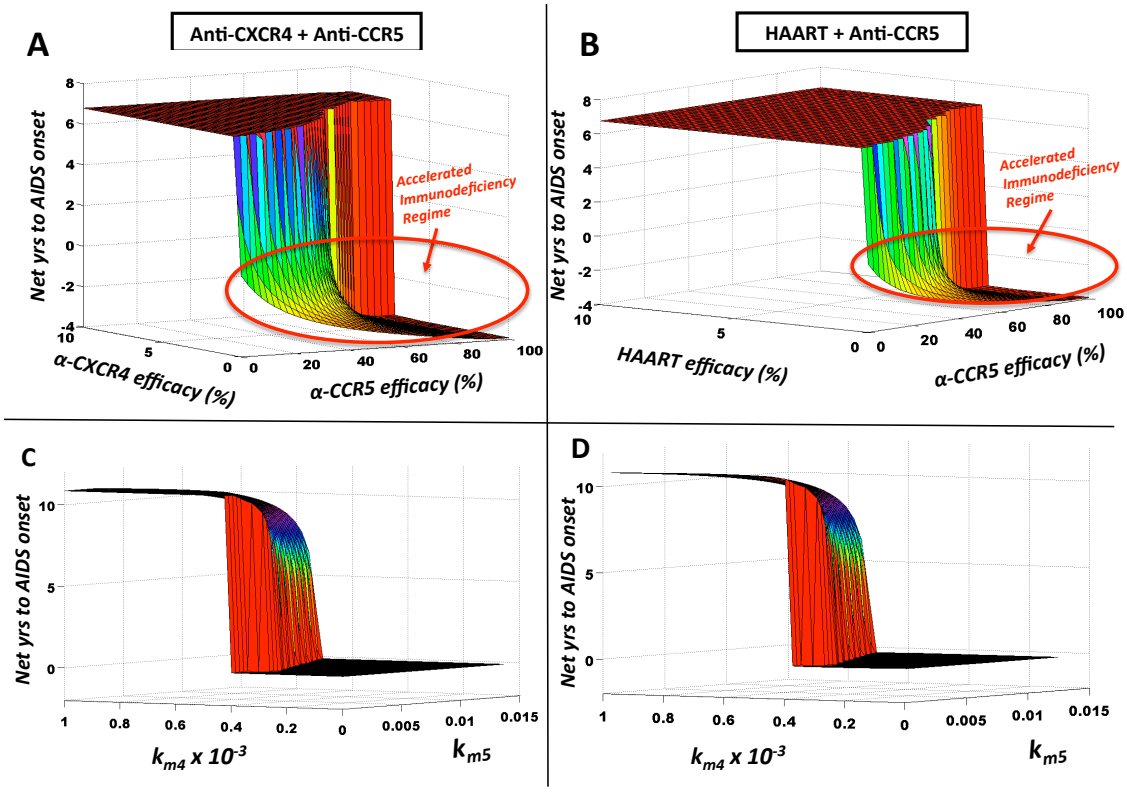
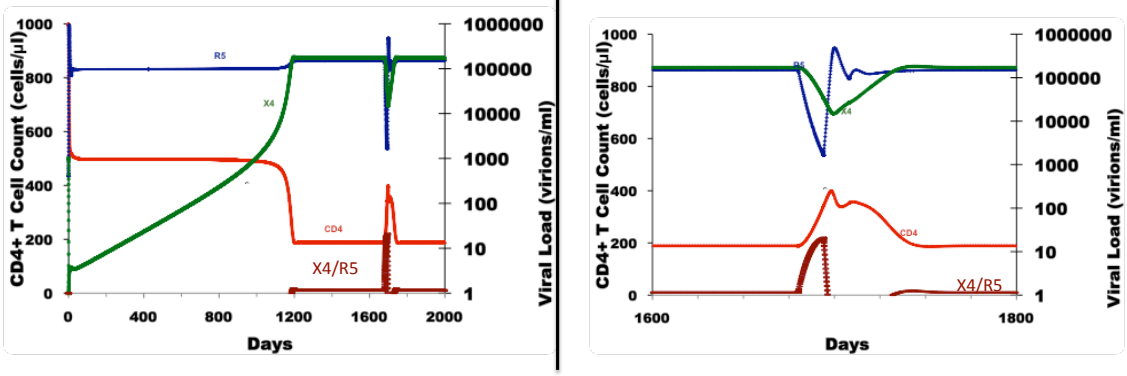

A
Anti-CXCR4 + Anti-CCR5
Net yrs to AIDS onset
Accelerated Immunodeficiency Regime
α-CXCR4 efficacy (%)
α-CCR5 efficacy (%)
B
HAART + Anti-CCR5
Net yrs to AIDS onset
Accelerated Immunodeficiency Regime
HAART efficacy (%)
α-CCR5 efficacy (%)
C
Net yrs to AIDS onset
k_{m4} x 10^{-3}
k_{m5}
D
Net yrs to AIDS onset
k_{m4} x 10^{-3}
k_{m5}

Fig. S1

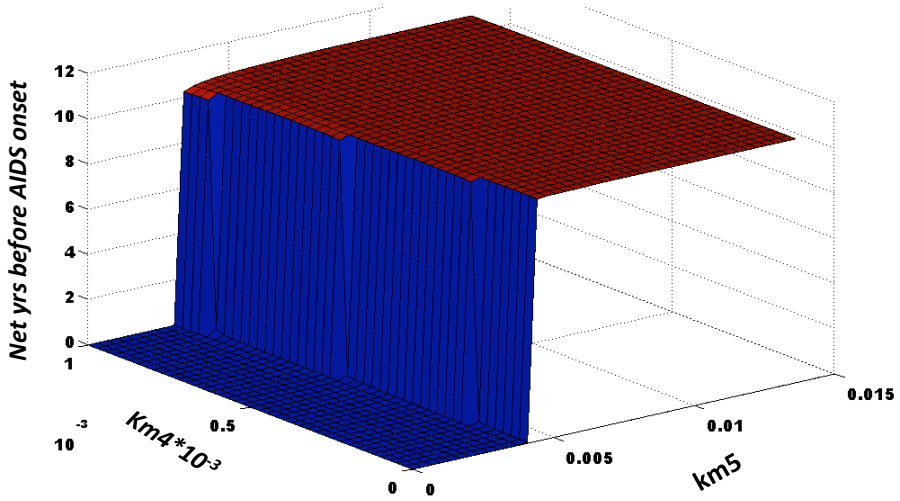
Close-up view treatment period



For **Model 2**, we simulate the anti-CCR5 treatment schedule of Wolinsky et al. (2004, J. Virol.) at steady-state time after an R5 to X4 switch. We fit the drug efficacy in that study to be approximately .9 (*data not shown)*, which is the value we use here (although the conclusions carry over analogously for other efficacies). Contrary to the Wolinsky data, in Model 2, X4 decreases after anti-CCR5 treatment. Importantly, X4/R5 is also considered here. While this function initially increases one-log beyond its steady-state value during anti-CCR5 treatment, it is actually suppressed below steady-state after CD4 counts rise during treatment. Thus, in Model 2, anti-CCR5 treatments eventually decrease the *relative* frequency of X4 in a patient .

1

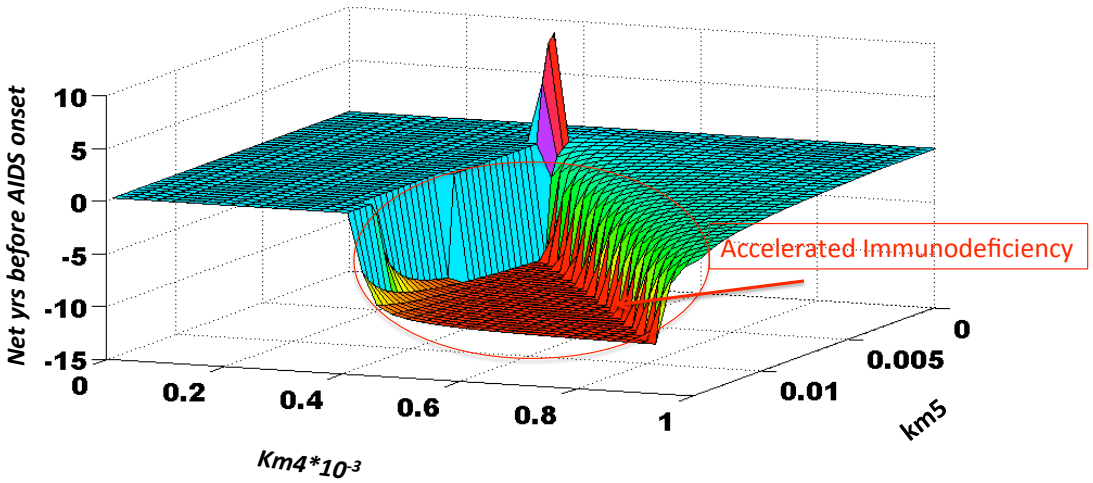
Fig. S2



The effect of employing CCR5 inhibitors in addition to HAART on the net time to AIDS (relative to using HAART alone). Across much of the below parameter space, CCR5 inhibition in addition to HAART ***extends*** time to immunodeficiency, showing the utility of anti-CCR5 treatment as part of a multi-therapy cocktail.

2

Fig. S3



The effect of CCR5 inhibitors alone on the net-time to AIDS onset in an extended version of Model 3, where loss of virions due to infection of target cells is now considered. The resulting figure is analogous to Figure 2c of the main text, and is noteworthy for having a large region in which the net time to AIDS onset is negative (i.e., accelerated AIDS onset due to CCR5 inhibition).

3

References

Ball, C.L., Gilchrist, M.A., and Coombs, D., 2007. Modeling within-host evolution of HIV: mutation, competition and strain replacement. Bull Math Biol 69, 2361-85.

Berger, E.A., Murphy, P.M., and Farber, J.M., 1999. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. Annu Rev Immunol 17, 657-700.

Blaak, H., van't Wout, A.B., Brouwer, M., Hooibrink, B., Hovenkamp, E., and Schuitemaker, H., 2000. In vivo HIV-1 infection of CD45RA(+)CD4(+) T cells is established primarily by syncytium-inducing variants and correlates with the rate of CD4(+) T cell decline. Proc Natl Acad Sci U S A 97, 1269-74.

Casper, C.H., Clevestig, P., Carlenor, E., Leitner, T., Anzen, B., Lidman, K., Belfrage, E., Albert, J., Bohlin, A.B., Naver, L., Lindgren, S., Fenyo, E.M., and Ehrnst, A.C., 2002. Link between the X4 phenotype in human immunodeficiency virus type 1-infected mothers and their children, despite the early presence of R5 in the child. J Infect Dis 186, 914-21.

Chen, Z., Kwon, D., Jin, Z., Monard, S., Telfer, P., Jones, M.S., Lu, C.Y., Aguilar, R.F., Ho, D.D., and Marx, P.A., 1998. Natural infection of a homozygous delta24 CCR5 red-capped mangabey with an R2b-tropic simian immunodeficiency virus. J Exp Med 188, 2057-65.

Cornelissen, M., Mulder-Kampinga, G., Veenstra, J., Zorgdrager, F., Kuiken, C., Hartman, S., Dekker, J., van der Hoek, L., Sol, C., Coutinho, R., and et al., 1995. Syncytium-inducing (SI) phenotype suppression at seroconversion after intramuscular inoculation of a non-syncytium-inducing/SI phenotypically mixed human immunodeficiency virus population. J Virol 69, 1810-8.

Douek, D.C., Picker, L.J., and Koup, R.A., 2003. T cell dynamics in HIV-1 infection. Annu Rev Immunol 21, 265-304.

Fatkenheuer, G., Pozniak, A.L., Johnson, M.A., Plettenberg, A., Staszewski, S., Hoepelman, A.I., Saag, M.S., Goebel, F.D., Rockstroh, J.K., Dezube, B.J., Jenkins, T.M., Medhurst, C., Sullivan, J.F., Ridgway, C., Abel, S., James, I.T., Youle, M., and van der Ryst, E., 2005. Efficacy of short-term monotherapy with maraviroc, a new CCR5 antagonist, in patients infected with HIV-1. Nat Med 11, 1170-2.

Glass, W.G., Lim, J.K., Cholera, R., Pletnev, A.G., Gao, J.L., and Murphy, P.M., 2005. Chemokine receptor CCR5 promotes leukocyte trafficking to the brain and survival in West Nile virus infection. J Exp Med 202, 1087-98.

Gondois-Rey, F., Biancotto, A., Fernandez, M.A., Bettendroffer, L., Blazkova, J., Trejbalova, K., Pion, M., and Hirsch, I., 2006. R5 variants of human immunodeficiency virus type 1 preferentially infect CD62L- CD4+ T cells and are potentially resistant to nucleoside reverse transcriptase inhibitors. J Virol 80, 854-65.

Gulick, R.M., Lalezari, J., Goodrich, J., Clumeck, N., DeJesus, E., Horban, A., Nadler, J., Clotet, B., Karlsson, A., Wohlfeiler, M., Montana, J.B., McHale, M., Sullivan, J., Ridgway, C., Felstead, S., Dunne, M.W., van der Ryst, E., and Mayer, H., 2008. Maraviroc for previously treated patients with R5 HIV-1 infection. N Engl J Med 359, 1429-41.

Haase, A.T., 1999. Population biology of HIV-1 infection: viral and CD4+ T cell demographics and dynamics in lymphatic tissues. Annu Rev Immunol 17, 625-56.

Harouse, J.M., Buckner, C., Gettie, A., Fuller, R., Bohm, R., Blanchard, J., and Cheng-Mayer, C., 2003. CD8+ T cell-mediated CXC chemokine receptor 4-simian/human

immunodeficiency virus suppression in dually infected rhesus macaques. Proc Natl Acad Sci U S A 100, 10977-82.
Hazenberg, M.D., Stuart, J.W., Otto, S.A., Borleffs, J.C., Boucher, C.A., de Boer, R.J., Miedema, F., and Hamann, D., 2000. T-cell division in human immunodeficiency virus (HIV)-1 infection is mainly due to immune activation: a longitudinal analysis in patients before and during highly active antiretroviral therapy (HAART). Blood 95, 249-55.
Hazenberg, M.D., Otto, S.A., van Benthem, B.H., Roos, M.T., Coutinho, R.A., Lange, J.M., Hamann, D., Prins, M., and Miedema, F., 2003. Persistent immune activation in HIV-1 infection is associated with progression to AIDS. Aids 17, 1881-8.
Ho, D.D., Neumann, A.U., Perelson, A.S., Chen, W., Leonard, J.M., and Markowitz, M., 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature 373, 123-6.
Kuhmann, S.E., Platt, E.J., Kozak, S.L., and Kabat, D., 2000. Cooperation of multiple CCR5 coreceptors is required for infections by human immunodeficiency virus type 1. J Virol 74, 7005-15.
Lee, B., Sharron, M., Montaner, L.J., Weissman, D., and Doms, R.W., 1999. Quantification of CD4, CCR5, and CXCR4 levels on lymphocyte subsets, dendritic cells, and differentially conditioned monocyte-derived macrophages. Proc Natl Acad Sci U S A 96, 5215-20.
Lee, S., Lapham, C.K., Chen, H., King, L., Manischewitz, J., Romantseva, T., Mostowski, H., Stantchev, T.S., Broder, C.C., and Golding, H., 2000. Coreceptor competition for association with CD4 may change the susceptibility of human cells to infection with T-tropic and macrophagetropic isolates of human immunodeficiency virus type 1. J Virol 74, 5016-23.
Li, Q., Duan, L., Estes, J.D., Ma, Z.M., Rourke, T., Wang, Y., Reilly, C., Carlis, J., Miller, C.J., and Haase, A.T., 2005. Peak SIV replication in resting memory CD4+ T cells depletes gut lamina propria CD4+ T cells. Nature 434, 1148-52.
Lim, J.K., Louie, C.Y., Glaser, C., Jean, C., Johnson, B., Johnson, H., McDermott, D.H., and Murphy, P.M., 2008. Genetic deficiency of chemokine receptor CCR5 is a strong risk factor for symptomatic West Nile virus infection: a meta-analysis of 4 cohorts in the US epidemic. J Infect Dis 197, 262-5.
Lin, Y.L., Mettling, C., Portales, P., Reynes, J., Clot, J., and Corbeau, P., 2002. Cell surface CCR5 density determines the postentry efficiency of R5 HIV-1 infection. Proc Natl Acad Sci U S A 99, 15590-5.
Lin, Y.L., Portales, P., Segondy, M., Baillat, V., de Boever, C.M., Le Moing, V., Reant, B., Montes, B., Clot, J., Reynes, J., and Corbeau, P., 2005. CXCR4 overexpression during the course of HIV-1 infection correlates with the emergence of X4 strains. J Acquir Immune Defic Syndr 39, 530-6.
Mohri, H., Perelson, A.S., Tung, K., Ribeiro, R.M., Ramratnam, B., Markowitz, M., Kost, R., Hurley, A., Weinberger, L., Cesar, D., Hellerstein, M.K., and Ho, D.D., 2001. Increased turnover of T lymphocytes in HIV-1 infection and its reduction by antiretroviral therapy. J Exp Med 194, 1277-87.
Moore, J.P., Kitchen, S.G., Pugach, P., and Zack, J.A., 2004. The CCR5 and CXCR4 coreceptors--central to understanding the transmission and pathogenesis of human immunodeficiency virus type 1 infection. AIDS Res Hum Retroviruses 20, 111-26.

Morgan, D., Mahe, C., Mayanja, B., Okongo, J.M., Lubega, R., and Whitworth, J.A., 2002. HIV-1 infection in rural Africa: is there a difference in median time to AIDS and survival compared with that in industrialized countries? AIDS 16, 597-603.

Nishimura, Y., Brown, C.R., Mattapallil, J.J., Igarashi, T., Buckler-White, A., Lafont, B.A., Hirsch, V.M., Roederer, M., and Martin, M.A., 2005. Resting naive CD4+ T cells are massively infected and eliminated by X4-tropic simian-human immunodeficiency viruses in macaques. Proc Natl Acad Sci U S A 102, 8000-5.

Penn, M.L., Grivel, J.C., Schramm, B., Goldsmith, M.A., and Margolis, L., 1999. CXCR4 utilization is sufficient to trigger CD4+ T cell depletion in HIV-1-infected human lymphoid tissue. Proc Natl Acad Sci U S A 96, 663-8.

Philpott, S., Weiser, B., Anastos, K., Kitchen, C.M., Robison, E., Meyer, W.A., 3rd, Sacks, H.S., Mathur-Wagh, U., Brunner, C., and Burger, H., 2001. Preferential suppression of CXCR4-specific strains of HIV-1 by antiviral therapy. J Clin Invest 107, 431-8.

Ribeiro, R.M., Hazenberg, M.D., Perelson, A.S., and Davenport, M.P., 2006. Naive and memory cell turnover as drivers of CCR5-to-CXCR4 tropism switch in human immunodeficiency virus type 1: implications for therapy. J Virol 80, 802-9.

Richman, D.D., and Bozzette, S.A., 1994. The impact of the syncytium-inducing phenotype of human immunodeficiency virus on disease progression. J Infect Dis 169, 968-74.

Roy, A.M., Schweighardt, B., Eckstein, L.A., Goldsmith, M.A., and McCune, J.M., 2005. Enhanced replication of R5 HIV-1 over X4 HIV-1 in CD4(+)CCR5(+)CXCR4(+) T cells. J Acquir Immune Defic Syndr 40, 267-75.

Sabeti, P.C., Walsh, E., Schaffner, S.F., Varilly, P., Fry, B., Hutcheson, H.B., Cullen, M., Mikkelsen, T.S., Roy, J., Patterson, N., Cooper, R., Reich, D., Altshuler, D., O'Brien, S., and Lander, E.S., 2005. The case for selection at CCR5-Delta32. PLoS Biol 3, e378.

Sachsenberg, N., Perelson, A.S., Yerly, S., Schockmel, G.A., Leduc, D., Hirschel, B., and Perrin, L., 1998. Turnover of CD4+ and CD8+ T lymphocytes in HIV-1 infection as measured by Ki-67 antigen. J Exp Med 187, 1295-303.

Schuitemaker, H., Koot, M., Kootstra, N.A., Dercksen, M.W., de Goede, R.E., van Steenwijk, R.P., Lange, J.M., Schattenkerk, J.K., Miedema, F., and Tersmette, M., 1992. Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytotropic to T-cell-tropic virus population. J Virol 66, 1354-60.

Spijkerman, I.J., Koot, M., Prins, M., Keet, I.P., van den Hoek, A.J., Miedema, F., and Coutinho, R.A., 1995. Lower prevalence and incidence of HIV-1 syncytium-inducing phenotype among injecting drug users compared with homosexual men. Aids 9, 1085-92.

Suarez, A., Mozo, L., and Gutierrez, C., 2002. Generation of CD4(+)CD45RA(+) effector T cells by stimulation in the presence of cyclic adenosine 5'-monophosphate-elevating agents. J Immunol 169, 1159-67.

Unutmaz, D., Pileri, P., and Abrignani, S., 1994. Antigen-independent activation of naive and memory resting T cells by a cytokine combination. J Exp Med 180, 1159-64.

Unutmaz, D., Baldoni, F., and Abrignani, S., 1995. Human naive T cells activated by cytokines differentiate into a split phenotype with functional features intermediate between naive and memory T cells. Int Immunol 7, 1417-24.

Wei, X., Ghosh, S.K., Taylor, M.E., Johnson, V.A., Emini, E.A., Deutsch, P., Lifson, J.D., Bonhoeffer, S., Nowak, M.A., Hahn, B.H., and et al., 1995. Viral dynamics in human immunodeficiency virus type 1 infection. Nature 373, 117-22.

Westby, M., Lewis, M., Whitcomb, J., Youle, M., Pozniak, A.L., James, I.T., Jenkins, T.M., Perros, M., and van der Ryst, E., 2006. Emergence of CXCR4-using human immunodeficiency virus type 1 (HIV-1) variants in a minority of HIV-1-infected patients following treatment with the CCR5 antagonist maraviroc is from a pretreatment CXCR4-using virus reservoir. J Virol 80, 4909-20.

Wolinsky, S.M., Veazey, R.S., Kunstman, K.J., Klasse, P.J., Dufour, J., Marozsan, A.J., Springer, M.S., and Moore, J.P., 2004. Effect of a CCR5 inhibitor on viral loads in macaques dual-infected with R5 and X4 primate immunodeficiency viruses. Virology 328, 19-29.

Zhang, Z.Q., Wietgrefe, S.W., Li, Q., Shore, M.D., Duan, L., Reilly, C., Lifson, J.D., and Haase, A.T., 2004. Roles of substrate availability and infection of resting and activated CD4+ T cells in transmission and acute simian immunodeficiency virus infection. Proc Natl Acad Sci U S A 101, 5640-5.

Chapter Three

Persistence and Emergence of X4 Virus in HIV Infection

Ariel D. Weinberger and Alan S. Perelson

Introduction

The previous chapter showed how the presence of low-level X4 Virus can lead to accelerated immunodeficiency and patient mortality under anti-CCR5 monotherapy. Yet, how does X4 Virus persist for long periods at low levels in the face of an initially more dominant R5 Virus, thereby endangering potential treatment interventions? The following chapter, much of which was first published in the journal Mathematical Biosciences and Engineering in 2011, probes this question.

X4's persistence is especially surprising given that it is initially less fit than the wild-type R5 Virus. Thus, X4 viruses are rarely seen during early infection, where R5 viruses predominate, whatever the route of infection (Casper et al., 2002; Cornelissen et al., 1995; Harouse et al., 2003; Keele et al., 2008; Moore et al., 2004). Moreover, individuals homozygous for a 32 base pair deletion in the allele for CCR5 (CCR5Δ32/Δ32) are almost entirely immune to HIV infection (Moore et al., 2004).

In vitro competition assays between R5 and X4 virus usually result in X4 dominance (Moore et al., 2004). Since about fivefold more lymphocytes are $CXCR4^+$ rather than $CCR5^+$ (Lee et al., 1999), one wonders why X4 is unable to dominate *in vivo*. A compelling explanation for R5's *in vivo* dominance and the basis for our models is CCR5's disproportionate presence on activated and recently activated memory $CD4^+$ T cells. Memory $CD4^+$ T cells can often be distinguished from their naïve precursor cells, because memory cells display the cell surface receptor CD45R0 (Blaak et al., 2000). Naïve cells generally display the receptor CD45RA, which is modified to its isoform CD45RO after an antigen 'naïve' CD4 T cell encounters its cognate antigen, thereby activating it into an effector memory cell.

Using the distinct cell surface receptors of naive and memory cells as well as antibodies that specifically bind to CCR5 and CXCR4, respectively, Lee et al. estimated the per-cell concentrations of CCR5 and CXCR4 molecules on naïve and memory T cells, respectively (Lee et al., 1999) (Table 1). The authors went further, dividing both naïve and memory cell populations into activated and quiescent subsets, based on whether the cells also expressed the receptor CD62L, which is displayed by naïve and memory cells in quiescent states (Hengel et al., 2003). Using quantitative fluorescence-activated cell sorting (QFACS), they found an average of 4741 R5 antibody-binding sites on $CD62L^+$ $CD45RO^+$ quiescent memory cells with only 1,013 X4 binding sites on this cell population. Among highly activated memory $CD62L^-$ $CD45RO^+$ $CD4^+$ T cells the difference is even more pronounced, with 9,576 R5 binding sites and only 505 X4 binding sites (Table 1). Conversely, the authors measured virtually no R5 antibody binding sites on naïve $CD45RA^+$ $CD4^+$ T cells on which X4 binding sites dominate. In general, as Table 1 shows, CXCR4 is more common on naïve and quiescent cells, while CCR5 dominates in the effector memory population.

As a result of CCR5's higher per-cell density among memory cells, which are more likely to be activated than naive cells (Douek, 2003; Mohri et al., 1998), R5 viruses may have an advantage on the whole over X4 viruses. Comparative snapshots of $CD4^+$ T cells during SIV infection show approximately five times as many virions surround infected, activated $CD4^+$ T cells as surround infected, phenotypically-quiescent $CD4^+$ T cells (Zhang et al., 2004).

Moreover, phenotypically-activated ($Ki67^+$) $CD4^+$ T cells produce over 90% of the virions during the chronic phase of SIV infection (Li et al., 2005).

Previous mathematical models have analyzed several hypotheses to explain how X4 can emerge during late-stage HIV despite R5 having an advantage in infecting the most activated and productive HIV target cells (Callaway et al., 1999; Joly and Pinto, 2005; Regoes and Bonhoeffer, 2002; Regoes and Bonhoeffer, 2005; Ribeiro et al., 2006; Wodarz and Nowak, 1998; Wodarz et al., 1999). An initial model (Regoes and Bonhoeffer, 2002) pursued the idea that antiretroviral treatment disproportionately inhibits R5 virus, precipitating a switch to X4. However, this cannot explain the documented emergence of X4 virus in treatment-naïve individuals (Coakley et al., 2005). Other models (Callaway et al., 1999; Wodarz and Nowak, 1998; Wodarz et al., 1999) analyzed the impact of differential immune responses on phenotypic switching, but these immune-based models utilize specific assumptions that current data argue against. Wodarz et al. (Wodarz et al., 1999) neglects the fact that over 90% of productive R5 infection occurs in $CD4^+$ T cells, not macrophages (Haase, 1999). The model by Wodarz and Nowak (Wodarz and Nowak, 1998) cannot explain the disproportionate increase in X4 viral loads (VLs) after CD8 depletion (Harouse et al., 2003). Finally, the model by Callaway et al. (Callaway et al., 1999) appears inconsistent with the fact that the greatest correlate of disease progression in HIV patients, and the only consistent difference between pathogenic and non-pathogenic lentiviral infections, is increased immune activation, including increased cytotoxic T-lymphocyte (CTL) activation (Giorgi et al., 1999; Silvestri et al., 2007). Since X4 onset is strongly correlated with disease progression, an active cytotoxic immune response is more likely a cause or consequence of X4 emergence than an inhibitor.

Here, the co-occurrence of X4 emergence and immune cell activation is explored in mathematical detail, using current data to derive conditions under which increased target-cell activation over the course of dual R5, X4 HIV infection drives a late-stage switch to X4 virus. As in the above studies, the switch to X4 in our models is the result of progressive HIV infection altering the fitness landscape in favor of X4. Yet, the mechanism altering the fitness landscape is different here, as we use *in vivo* derived data to justify conditions showing changing T cell activation rates directly changing the fitness landscape in favor of X4. Building upon previous studies arguing that target-cell activation drives the switch to X4 (Ribeiro et al., 2006; Weinberger et al., 2009), we derive a minimal target-cell activation-based model for understanding multi-tropism and its attendant immunodeficiency in HIV.

Results

Generalized conditions for a phenotypic switch

Curve fitting a data set measured *in vivo* (Hazenberg et al., 2000a), which determined the fractions of activated T cells using the cell-cycle activation marker (Ki67), Ribeiro et al. [25] found that the fractions of naïve cells that are activated (a_n) and the fractions of memory cells that are activated (a_m) obey the following inverse relationships with respect to the total $CD4^+$ T cell count:

$$a_n(CD4) = \frac{10}{CD4} - .0095$$
$$a_m(CD4) = \frac{10}{CD4} + .05 \qquad \text{(Eq. 1)}$$

Here, CD4 denotes the total number of uninfected and infected $CD4^+$ T cells per microliter of blood.

During HIV infection, $CD4^+$ T counts decline, causing both a_n and a_m to increase. The increase in a_n lets X4 virus benefit from CXCR4's strong presence on activated naïve $CD4^+$ T cells (Lee et al., 1999) (Table 1), allowing for a switch.

The last chapter claimed that a generalization of (1) would allow switches to occur even if the fraction of activated naïve $CD4^+$ T cells increases at a slower rate than the fraction of activated memory $CD4^+$ T cells. Specifically, the previous chapter (Weinberger et al., 2009) examined the possibility of a phenotypic switch when

$$i)\ a_n(CD4) < a_m(CD4)$$

$$ii)\ a_n'(CD4) < 0, \quad a_m'(CD4) < 0$$

$$iii)\ \frac{d}{d(CD4)}\left(\frac{a_n(CD4)}{a_m(CD4)}\right) < 0 \qquad \text{Eq. (2)}$$

Here we rigorously justify these conditions. To do so, we note that throughout infection a far greater fraction of $CD4^+$ memory cells are activated than naive $CD4^+$ lymphocytes (Douek, 2003; Mohri et al., 1998). Thus, we set $a_n < a_m$. Furthermore, increased immune activation is strongly correlated with $CD4^+$ T cell decline in HIV patients (Hazenberg et al., 2003a; Hazenberg et al., 2003b; Paiardini et al., 2008; Silvestri et al., 2007) and this increased activation is manifested in both naïve and memory $CD4^+$ T cells (Hazenberg et al., 2000a; Mohri et al., 1998), so $a_n'(CD4) < 0$ and $a_m'(CD4) < 0$. To justify the final condition in Eq. (2), note that:

$$\frac{d}{dCD4}\left(\frac{a_n(CD4)}{a_m(CD4)}\right) = \frac{a_n'(CD4)a_m(CD4) - a_m'(CD4)a_n(CD4)}{(a_m(CD4))^2}$$

This derivative is negative if and only if

$$a_n'(CD4)\, a_m(CD4) - a_m'(CD4)\, a_n(CD4) < 0$$

Because $a_m'(CD4) < 0$, this is true if and only if:

$$a_n'(CD4) / a_m'(CD4) > a_n(CD4) > / a_m(CD4) \qquad \text{Eq. (3)}$$

Clearly, Eq. (1) is a particular system satisfying Eq. (2), because in Eq. (1) we have $a_n < a_m$ and $a_n' = a_m' < 0$. So in justifying Eq. (2) we are allowing for a larger class of models.

To justify Eq. (3) and thus the final condition of Eq. (2), we note that if a_n' is a larger fraction of a_m' than a_n is of a_m, Eq. (3) holds. We have already shown $a_n < a_m$., implying that:

$$1 - a_m < 1 - a_n \qquad \text{Eq. (4)}$$

Thus, the fraction of naïve cells that is quiescent is greater than the fraction of memory cells that is quiescent. We let n_t and m_t represent the total numbers (i.e., activated + non-activated) of naïve and memory $CD4^+$ T cells, respectively. Because naïve and memory cell counts are initially similar and because R5 virus disproportionately depletes memory $CD4^+$ T cells (Gondois-Rey et al., 2006; Veazey et al., 2000), we assume that $n_t > m_t$ during R5 infection, implying:

$$m_t (1 - a_m) < n_t (1 - a_n) \qquad \text{Eq. (5)}$$

Furthermore, many of the newly activated memory $CD4^+$ T cells were previously quiescent naïve $CD4^+$ T cells activated by interaction with antigen. These additions to a_m also increase a_n by reducing the number of quiescent naive $CD4^+$ T cells. Thus, given the large measured differences between the fractions of activated naïve and memory $CD4^+$ T cells, we argue that discrepancies between the rates of increase of activated naïve and memory $CD4^+$ T cells will often be relatively small. That is, we claim that in many cases (a_n'(CD4) / a_m'(CD4)) > (a_n(CD4) / a_m(CD4)), which is equivalent to Eq. (3). Data sets such as the one from which Eq. (1) was derived, give us evidence that this is reasonable. We note that Eq. (3) clearly holds when a_n'(CD4) ≤ a_m'(CD4) < 0, that is, when the fraction of activated naïve $CD4^+$ T cells increases at least as quickly as the corresponding fraction of memory cells, as $CD4^+$ T cells decline. Such a scenario is obviously to the increasing benefit of X4 virus in an activation-based model. In fact, this idea was used to explain the switch in Ribeiro et al. (Ribeiro et al., 2006). Yet, because $a_n < a_m$, Eq. (3) is even satisfied in certain cases in which 0 > a_n'(CD4) > a_m'(CD4) (i.e., when a_n increases at a *slower* rate than a_m in response to $CD4^+$ T cell decline). Such a broadened scenario would occur if $a_n << a_m$ and a_n' only slightly less negative than a_m'. Of course, in situations where a_n increases far slower than a_m in response to $CD4^+$ T cell decline, Eq. (3) would likely not hold.

Model 1: One Target Cell Population Yields A Competitively Exclusive Switch

In the preceding chapter (Weinberger et al., 2009), we began by extending the basic model of viral dynamics (Ho et al., 1995; Wei et al., 1995) to the simplest dual-strain framework, denoted Model 1 there and below. Through simulations, we showed that R5-to-X4 switches arise from this model, but claimed that such switches are beset by *competitive exclusion*, given the single-compartment nature of that model. Competitive exclusion is not consistent with *in vivo* data, which show X4 and R5 coexisting post-switch (Wolinsky et al., 2004). Here, we analytically show that *competitive exclusion* is the result of Model 1 and further show that accelerated emergence of X4 virus due to anti-CCR5 treatment is a basic result of strain competition for target-cells and is present in even the simplest of competitive models.

$$\dot{T} = \lambda - (k_4 V_4 + k_5 V_5)T - d_T T$$
$$\dot{I}_4 = k_4 V_4 T - \delta I_4$$
$$\dot{I}_5 = k_5 V_5 T - \delta I_5$$
$$\dot{V}_4 = p a_n I_4 - c V_4$$
$$\dot{V}_5 = p a_m I_5 - c V_5$$

(Model 1)

In this model, all variables (capitalized) are concentrations per microliter (1/μl), λ has the units cells/(μl•day), k_4 and k_5 have the units μl/(virions•day), and the remaining parameters have units 1/day. Specifically, T represents the concentration of uninfected $CD4^+$ T cells, and (without loss of generality) is given an initial value of 1000 $CD4^+$ T cells/μl. I_4 and I_5 reflect the concentrations of $CD4^+$ T cells *abortively, latently, and productively* infected by X4 and R5 viruses, respectively; V_4 and V_5, describe X4 and R5 virus concentrations. λ is the rate of production of $CD4^+$ T cells and k_4 and k_5 are the respective infection rate coefficients for X4 and R5 infection of $CD4^+$ T cells. Also, d_T is the death rate of uninfected $CD4^+$ T cells and is set equal to λ/T_0 to allow for steady-state pre-infection, δ is the death rate of infected $CD4^+$ T cells, p is the rate of viral production by *activated* infected cells, and c is the viral clearance rate. a_n and a_m are required to satisfy Equation (2) and represent the fractions of *activated* naïve and memory $CD4^+$ T cells for a given value of CD4. Since CD4 represents the total number of uninfected and infected $CD4^+$ T cells per microliter, $CD4 = T + I_4 + I_5$.

We assume that when activated cells become infected they produce virus at rate p per cell. In our model, it is only these activated infected cells that produce virus. We thus multiply $a_n I_4$ and $a_m I_5$ by p, to obtain the total concentrations of virions produced each day. In a more complex model, one could allow a small amount of viral production from infected resting cells. Importantly, the products $a_n I_4$ and $a_m I_5$ assume that infected cells are no more likely to be activated than uninfected cells. This is because infection in our model is not necessarily productive, and in general most infections have been measured to be non-productive (Haase, 1999).

Given the per-cell concentrations of CCR5 and CXCR4 recorded shown in Table 1, we assume that X4 virus only productively infects *naïve* $CD4^+$ T cells and thus make X4's viral production dependent on a_n, but not a_m. Conversely, we use the same dataset to justify making R5's production dependent on a_m, but not a_n. Because CXCR4's median cell surface density is almost three times as high as that of CCR5 across all lymphocytes (Lee et al., 1999), we also assume $k_4 > k_5$. As above, this does not imply that X4 *productively* infects more target cells than R5 at the beginning of infection, since very few naïve cells are activated early in infection (Hazenberg et al., 2000a).

Deriving a Switch Threshold for Model 1

In analyzing Model 1, we first determine how many *productively* infected cells each strain has at a given point in time. Let R_{eff4} and R_{eff5} be time-dependent functions for the average number of infected cells that an average X4 and R5 infected cell produces. R_{eff4} and R_{eff5} are thus "*effective reproductive ratios*," in contrast to the "*basic reproductive ratios*," R_{04} and R_{05}, which evaluate R_{eff4} and R_{eff5} at the initial time point. The equations for R_{eff4} and R_{eff5} are

R_{eff4} (t)= $p \cdot a_n$(CD4(t)) $\cdot k_4 \cdot T(t) / (c \cdot \delta)$
R_{eff5} (t)= $p \cdot a_m$(CD4(t)) $\cdot k_5 \cdot T(t) / (c \cdot \delta)$ (Eq. 6)

$$\frac{\mathrm{R_{eff4}}}{\mathrm{R_{eff5}}}(CD4) = \frac{k_4}{k_5} \cdot \frac{a_n(CD4)}{a_m(CD4)}$$

We note that while R_{eff4} and R_{eff5} are functions of t (time), the explicit time-dependencies of R_{eff4} and R_{eff5} cancel in the quotient R_{eff4}/R_{eff5}. This allows us to explore and subsequently differentiate R_{eff4}/R_{eff5} as a function of CD4 alone.

Initially, we assume that $R_{eff4} < R_{eff5}$, because at the large $CD4^+$ T cell counts prevalent during early infection $a_m >> a_n$, implying that $a_m \cdot k_5 > a_n \cdot k_4$ despite the fact that $k_4 > k_5$ (i.e., we assume that *initially* the disparity between a_n and a_m is greater than the disparity between k_4 and k_5). With a higher effective reproductive ratio, R5 virus is more efficient and dominates early, consistent with observation. As infection progresses, Eq. (2) shows that the relative fraction of activated naïve cells increases as $CD4^+$ T cells decrease. This yields

$$\frac{d}{dCD4}\left(\frac{\mathrm{R_{eff4}}}{\mathrm{R_{eff5}}}\right) = \frac{k_4}{k_5} \cdot \frac{d}{d(CD4)}\left(\frac{a_n(CD4)}{a_m(CD4)}\right) < 0 \qquad \text{Eq. (7)}$$

In other words, if Eq. (2) holds, lowering $CD4^+$ T cell counts preferentially benefits X4 by increasing its fitness relative to that of R5 virus. This accounts for the possibility of a switch at low $CD4^+$ T cell counts. Here we show that when CD4 counts decrease enough for R_{eff4} / R_{eff5} to go above 1, a switch to X4 virus occurs at a future time point. Conversely, if R_{eff4} / R_{eff5} never increases beyond 1, a switch to X4 cannot occur, potentially explaining why 50% of patients do not exhibit a switch to X4 Virus during HIV infection.

In order for X4 virus to overtake R5 virus at time t*, the following conditions are necessary and sufficient: $d/dt(V_4(t^*)) > d/dt\ (V_5(t^*))$ and $V_4(t^*) = V_5(t^*)$. Solving these switch equations simultaneously yields a necessary and sufficient switch condition for this model:

$$a_n(CD4(t^*))/a_m(CD4(t^*)) > I_5(t^*)/I_4(t^*) \qquad \text{Eq. (8)}$$

Equation (8) describes the threshold *at which* the switch to X4 occurs. We can find an earlier necessary and sufficient threshold for a_n/a_m above which a *future* switch to X4 is guaranteed to occur. To do so, we note that because $a_n/a_m < 1$ for all time, the right-hand side of Eq. (8) must be less than one. Thus, $I_4(t^*) > I_5(t^*)$ is a necessary condition for a switch. In biological terms, when X4's advantage, manifested in a greater number of infected cells, outweighs R5's advantage, manifested in a higher target-cell activation level (i.e., a higher probability that its infected cells are productively infected), the switch occurs.

For I_4 to overtake I_5 at t*, a necessary condition is that at some earlier time point, t**<t*, the rate of growth of X4-infected cells was higher than that of R5-infected cells. Hence, R_{eff4}(t**) > R_{eff5}(t**) is a necessary condition for the R5 to X4 switch to occur. In fact, R_{eff4}(t**) > R_{eff5}(t**) is also a *sufficient* condition for the R5 to X4 switch. Because Eq. (8) implies that R_{eff4}

is always increasing relative to R_{eff5} as $CD4^+$ T cells decline, $R_{eff4}(t^{**}) > R_{eff5}(t^{**})$ means that $R_{eff4}(t) > R_{eff5}(t)$ for all $t \geq t^{**}$. Thus, if $R_{eff4}(t^{**}) > R_{eff5}(t^{**})$, X4 will eventually overtake R5.

An R5 to X4 switch always results in the eventual extinction of R5 in Model 1. This is because coexistence at steady state means:

$$d/dt(T) = d/dt(I_4) = d/dt(I_5) = d/dt(V_4) = d/dt(V_5) = 0 \text{ and } I_4, I_5, V_4, V_5 \neq 0 . \qquad \text{Eq. (9)}$$

$d/dt(V_4) = d/dt(V_5) = 0$ means

$$p = (c \cdot V_4) / (a_n \cdot I_4) = (c \cdot V_5) / (a_m \cdot I_5) ,$$

so

$$V_4 = (V_5 \cdot a_n \cdot I_4) / (a_m \cdot I_5) \quad . \qquad \text{Eq. (10)}$$

Moreover, since $d/dt(I4) = d/dt(I5) = 0$,

$$\delta = (k_5 \cdot V_5 \cdot T) / I_5 = (k_4 \cdot V_4 \cdot T) / I_4$$

implying that

$$I_4 = (k_4 \cdot V_4 \cdot I_5) / (k_5 \cdot V_5) . \qquad \text{Eq. (11)}$$

Plugging Eq. (10) into Eq. (11) tells us that a necessary and sufficient coexistence condition is $(k_4 / k_5) \cdot (a_n / a_m) = 1$. By Eq. (6), this is equivalent to the coexistence iff $R_{eff4} = R_{eff5}$.

Thus, a necessary and sufficient condition for a switch to occur at some point $t^* > t^{**}$ and for R5 to approach extinction is:

$$R_{eff4}(t^{**}) > R_{eff5}(t^{**}) \qquad \text{Eq. (12)}$$

$R_{eff4}(t^{**}) > R_{eff5}(t^{**})$ means that a_n / a_m increases beyond k_5 / k_4. Because a_n / a_m increases as $CD4^+$ T cells decline, it is the level of $CD4^+$ T cell depletion engendered by HIV that is directly implicated in the model's switch. To quantify this for the measured functions of a_n and a_m given in Eq. (1), we substitute Eq. (1) into Eqs. (6) and (12) to yield the following necessary and sufficient threshold beyond which a switch is guaranteed to eventually occur:

$$CD4(t^{**}) < 200 \frac{k_4 - k_5}{k_5 + .19 k_4} \qquad \text{Eq. (13)}$$

Since $k_4 > k_5$, the quotient on the right hand side is positive. Hence, at a CD4 count below a threshold, the switch condition is satisfied, guaranteeing that X4 will eventually take over. With the exception of changes to k_4 and k_5, it is clear from Eq. (13) that all changes to the model's parameters that accelerate $CD4^+$ T cell depletion accelerate an R5 to X4 switch. Conversely, mitigating the level of infection and consequent $CD4^+$ T cell depletion lengthens the time until the switch occurs. Because the partial derivative of the right side of Eq. (13) is positive

with respect to k_4, increasing k_4 also accelerates the switch by increasing the right hand side while decreasing CD4 counts through heightened X4 infection (Fig. 1). Yet, the partial derivative of Eq. (13) with respect to k_5 is negative, meaning that both right and left sides of the equation decrease in response to higher levels of k_5, making it initially unclear as to whether increasing k_5 promotes a switch to X4 virus.

CCR5 Inhibitors Can Promote Switches to X4 virus in a Single Compartment Model

In general, reducing k_5—as occurs in CCR5 inhibitor treatments—increases k_4/k_5, but it also decreases a_n/a_m by increasing $CD4^+$ T cell counts through decreased R5 infection. By Eq. (6), R_{eff4}/R_{eff5} is the product of k_4/k_5 and a_n/a_m, so the question is whether the increase to k_4/k_5 is greater than the decrease to a_n/a_m. If so, R_{eff4}/R_{eff5} increases in response to lowering k_5, implying that anti-CCR5 treatments can accelerate switches to X4.

We examined how modulating k_5 affects the switch to X4 virus. When a_n and a_m are defined as in Eq. (1), increasing k_5 from $1 \cdot 10^{-4}$ μl/(virions•day) to $1.5 \cdot 10^{-4}$ μl/(virions•day) accelerates the time at which X4 emerges (i.e., it increases R_4/R_5) (Fig. 2a, upper panel). However, increasing k_5 even further to $3 \cdot 10^{-4}$ μl/(virions•day) prevents a switch (Fig. 2a, lower panel). In fact, the model predicts a steady state with high X4 viral loads only at intermediate values of k_5: increasing k_5 beyond a threshold blocks X4 emergence (Fig. 2b). To understand why increasing k_5 beyond a threshold prevents a switch to X4 Virus, we note that large values of k_5 (e.g. $k_5=3 \cdot 10^{-4}$) allow R5 to infect the vast majority of $CD4^+$ T cells, leaving few uninfected R5 target cells. This causes *diminishing returns* in the number of new $CD4^+$T cells that can be infected through further increases to k_5. As a result, when k_5 is initially large and k_5 is further increased, the increase to a_n/a_m from further CD4 T cell declines is unlikely to outweigh the decrease to k_4/k_5, causing a decrease in R_4/R_5 and inhibiting X4 emergence. Thus, if k_5 is initially large and a CCR5 inhibitor only partially decreases k_5—keeping us in the high k_5 *diminishing returns* regime—the increase to k_4/k_5 from decreasing k_5 can outweigh the decrease to a_n/a_m from the small increase in $CD4^+$ T cell counts, increasing R_{eff4}/R_{eff5} and *promoting* a switch to X4 (Fig. 2c). Significantly, these switches to X4 are prevented by combination therapies such as HAART (Fig. 2d, left panel) or combined CCR5, CXCR4 inhibition (Fig. 2d, right panel), which combat both R5 and X4 virus equally.

Model 1 is thus a simplified model in which we can rigorously see that competition for target cells may make anti-CCR5 treatment a risky proposition. And while this simplified model seems to imply that CCR5 inhibitors do not do any damage insofar as the steady-state $CD4^+$ T cell count is not lowered as a result of X4 emergence (Fig. 2c), the reality, as described more faithfully in Model 3 below, is that X4's emergence uniquely depletes the naïve $CD4^+$ T cell population, which serves as the pipeline for new memory $CD4^+$ T cells.

Model 2: Coexistence, but no competition

Having analyzed a simplified one-compartment switch-inducing model in detail, we are left with the problem of competitive exclusion. This all-or-nothing result is inconsistent with data that shows the possibility of coexistence after a phenotypic switch (Philpott et al., 2001). We previously claimed that maintaining distinct target-cell populations for R5 and X4 viruses is sufficient to produce coexistence (Weinberger et al., 2009). Here we rigorously show this.

$$\dot{N} = \lambda + (1-2f)a_n N - k_4 V_4 N - d_N N$$
$$\dot{M} = 2fa_n N + a_m M - k_5 V_5 M - d_M M$$
$$\dot{I}_4 = k_4 V_4 N - \delta I_4$$
$$\dot{I}_5 = k_5 V_5 M - \delta I_5$$
$$\dot{V}_4 = pa_n I_4 - cV_4$$
$$\dot{V}_5 = pa_m I_5 - cV_5$$

(Model 2)

The equations in this system are analogous to those in Model 1 but the uninfected $CD4^+$ T cell population is now split into uninfected naïve (N) and memory (M) subpopulations. The target cell death rates, d_n and d_m , are defined analogously to d_T in Model I, ensuring that both subsets of the uninfected $CD4^+$ T cell population are in equilibrium pre-infection. Additionally, f is defined to be the fraction of naïve cells activated by antigen, which then divide and differentiate into $CD45RO^+$ memory cells. The rest of the activated cells are assumed to have been upregulated via cytokines or other antigen-T cell receptor independent processes and thus remain phenotypically naïve ($CD45RA^+$) (Suarez et al., 2002; Unutmaz et al., 1994; Unutmaz et al., 1995). Again, for simplicity it is assumed that X4 virus solely infects naive cells and that R5 virus only infects memory cells. Since the target cell population is now split, the effective reproductive ratios of R5 and X4 become functions of distinct target cell populations:

$$R_{eff4}\,(t) = p * a_n(CD4) * k_4 * N(t)/(c * \delta_i)$$

$$R_{eff5}\,(t) = p * a_m(CD4) * k_5 * M(t)/(c * \delta_i) \qquad \text{Eq. (14)}$$

$$R_{eff4}/R_{eff5}\,(t) = (k_4/k_5) * (a_n(CD4)/a_m(CD4)) * N(t)/M(t)$$

An immediate result is that $k_4 > k_5$ is no longer required for a switch to occur. In fact, if R5 depletes most of its target cells, X4 virus will have an advantage even when X4 has a lower infection rate coefficient. That is, $a_n * k_4 * N > a_m * k_5 * M$ is possible even when $a_n * k_4 < a_m * k_5$.

Because of the differential target-cell compartments, after a phenotypic switch R_{eff5} can rebound and increase relative to R_{eff4}, a fact that could not occur in the above single-compartment model. This occurs because when X4 viral loads burgeon during a switch, X4 encounters an untapped naïve target cell pool, while most memory target cells have already been depleted by R5 infection. This means that the naïve $CD4^+$ T cell population will decrease more rapidly than the corresponding memory cell population after an R5-to-X4 switch. If the resulting decrease in N/M is greater than the increase in a_n/a_m that results from the lowered $CD4^+$ T cell counts post-switch, then R_{eff4}/R_{eff5} decreases by Eq. 14. Since R_{eff4}/R_{eff5} can decrease after a switch in a two-compartment model, coexistence is now possible (Supporting Figure 1).

The equations for V_4 and V_5 in the two compartment model are identical to those in the single-compartment model, so the same switch condition persists (found by setting $V_4=V_5$ and $d/dt(V_4)>d/dt(V_5)$). Thus, as in Model 1, a switch occurs if and only if a_n/a_m goes above the threshold in (8), or, equivalently, if and only if there is sufficient $CD4^+$ T cell depletion. Thus, modulating parameters to increase $CD4^+$ T cell decline accelerates an R5 to X4 switch, while down-regulating infection, for example via drug intervention, inhibits X4 incidence. This result clearly extends to changes in k_5, as X4 and R5 are independent viruses here so that X4 receives

no advantage from a weakened R5 virus. Moreover, having a $CD4^+$ T cell threshold for an R5 to X4 switch means that despite R5's ability to increase after X4 depletes the naïve CD4 population post-switch, R5 is not likely to overtake X4 post-switch, because doing so requires an increase in $CD4^+$ T cell counts.

While this two-compartment model can produce switching and coexistence, it is oversimplified in assuming that X4 cannot infect any memory $CD4^+$ T cells. In fact, despite being outcompeted by R5 for $CCR5^+$, $CXCR4^+$ $CD4^+$ T cells, X4 productively infects certain memory $CD4^+$ T cells, predominantly those that are resting, $CD62L^+$ (Gondois-Rey et al., 2002; Lee et al., 1999).

Model 3: Two Compartments with Competition

In order to account for the observed competition of X4 and R5 viruses for the infection of memory $CD4^+$ T cells, which allows X4 to increase in response to CCR5 inhibition as seen in *in vivo* experiments (Wolinsky et al., 2004), and in order to prevent *competitive exclusion* of the less fit viral strain, we combine Models 1 and 2, allowing X4's infection of both naïve and memory $CD4^+$ T cells:

$$\begin{aligned}
\dot{N} &= \lambda + (1-2f)a_n N - k_{N4}V_4 N - d_N N \\
\dot{M} &= 2fa_n N + a_m M - k_{M4}V_4 M - k_{M5}V_5 M - d_M M \\
\dot{N}_4 &= k_{N4}V_4 N - \delta N_4 \\
\dot{M}_4 &= k_{M4}V_4 M - \delta M_4 \\
\dot{M}_5 &= k_{M5}V_5 M - \delta M_5 \\
\dot{V}_4 &= p(a_n N_4 + a_m M_4) - cV_4 \\
\dot{V}_5 &= pa_m M_5 - cV_5
\end{aligned} \qquad \text{(Model 3)}$$

In this model, k_{N4}, k_{M4}, and k_{M5}, are the infection rate coefficients of X4 and R5 for naïve (N) and memory (M) $CD4^+$ T cells, while N_4, M_4, and M_5 are the infected cell concentrations corresponding to the originating target cell and infecting viral strain. All other parameters, variables, and initial conditions have been previously defined. Because CCR5 is far more strongly expressed on memory $CD4^+$ T cells than is CXCR4 (Lee et al., 1999) (Table 1), we set $k_{M5} >> k_{M4}$. Conversely, CXCR4 is more highly expressed on naïve $CD4^+$ T cells than on memory $CD4^+$ T cells (Lee et al., 1999), making $k_{N4} >> k_{M4}$.

Given that X4 and R5 were already shown to coexist in Model 2 due to decreases in N/M post-switch, coexistence results in this extended dual compartment model as well (Weinberger et al., 2009). To show this graphically, we note that the relative fitness of X4 to R5 is given by:

$$R_{eff4}/R_{eff5}\,(t) = (k_{N4}/k_{M5})*(a_n(CD4(t))/a_m(CD4(t)))*(N(t)/M(t)) + k_{M4}/k_{M5} \quad \text{Eq. (15a)}$$

Simulations show R_{eff4}/R_{eff5} rising to a local maximum before the R5 to X4 switch only to decrease post-switch due to an X4-driven decrease in N/M. R_{eff4}/R_{eff5} eventually fixates at the value 1, allowing for coexistence of R5 and X4 strains post-switch (Supporting Figure 2).

In the previous chapter (Weinberger et al., 2009), we derived the following switch conditions for Model 3:

$$(a_n(\mathrm{CD4}(t^*))/a_m(\mathrm{CD4}(t^*))) > (M_5(t^*)-M_4(t^*))/N_4(t^*) \qquad \text{Eq. (15b)}$$

$$CD4(t^*) < 200\frac{N_4(t^*)+M_4(t^*)-M_5(t^*)}{M_5(t^*)-M_4(t^*)+.19N_4(t^*)} \qquad \text{Eq. (16)}$$

As in the preceding models, Equation (16) implies that, with the exception of changes to k_{M5}, modulating parameters to accelerate $CD4^+$ T cell decline hastens an R5 to X4 switch while changing parameters to mitigate $CD4^+$ T cell decline hinders a phenotypic switch. Thus, successful antiretroviral therapy will generally inhibit X4's emergence. However, as one might predict from our results in Model 1**,** because R5 and X4 are in competition, CCR5 inhibitors can generate more complicated kinetics. In fact, the utility of CCR5 inhibitors depends on the strength of the competition between X4 and R5 virus, which is modulated by X4's infection rate coefficient for memory $CD4^+$ T cells, k_{M4}. In the "non-competitive" regime in which $k_{M4} << k_{M5}$, X4's viral set-point is a *monotonically decreasing* function of the CCR5 inhibitor's efficacy (Fig. 3a, left panel). This is because, due to the low value of k_{M4}, X4 is unable to infect the majority of target cells blocked from R5 infection, allowing CD4 counts to rise and causing a drop in activation levels. Conversely, in the "competitive" regime where k_{M4} is closer in value to k_{M5} but still less than k_{M5} to remain consistent with FACS data [44], X4's viral set-point is a *monotonically increasing* function of the CCR5 inhibitor's efficacy (Fig. 3a, right panel).

In the "competitive regime" the steady-state $CD4^+$ T cell count is not decreased by CCR5 inhibition (Fig. 3a, right panel), which might lead one to suspect that these treatments are safe in this regime as well. Critically, however, the $CD4^+$ T cell count crashes far sooner in the "competitive regime" when anti-CCR5 treatment is employed (Fig. 3b). Thus, CCR5 inhibitors may accelerate immunodeficiency in patients with a competitive X4 virus.

Why is this the case? CCR5 inhibitors decrease k_{M5}, causing R5's viral load to decline, and memory $CD4^+$ T cell counts to increase. X4 is now able to infect some of these newly generated memory $CD4^+$ T cells, but X4's ability to do so depends on k_{M4}. With k_{M4} sufficiently large (the "competitive regime"), X4 infects a non-negligible fraction of newly generated memory $CD4^+$ cells, increasing X4's viral load. But with a larger viral load, X4 is now better able to infect its main target cell pool: *naïve* $CD4^+$ T cells, the untapped target cell reserve where CXCR4 is highly present. The slight increase in memory $CD4^+$ T cell counts due to CCR5 inhibition thus causes severe and accelerated depletion of the *naïve* $CD4^+$ T cell population in this "competitive regime" (Fig. 3c). Thus, a single parameter, k_{M4}, controls the efficacy of anti-CCR5 therapy in dually infected HIV patients, highlighting the need for circumspection in prescribing these treatments.

DISCUSSION

In this chapter, a set of mathematical models for dual R5, X4 infection in HIV has been rigorously derived from a multi-strain version of the basic model of viral dynamics. The models were analyzed to show how an increase in the *ratio* of the fractions of activated naïve and memory $CD4^+$ T cells (a_n/a_m) can trigger an R5 to X4 "phenotypic" switch in dually infected individuals. Importantly, this allows for phenotypic switching even when the fraction of activated naïve $CD4^+$ T cells increases at a slower rate than the corresponding fraction of memory $CD4^+$ T cells (as long as the *ratio* of the two fractions increases beyond a threshold). Our models also help explain why 50% of patients do not manifest a noticeable switch to X4 virus: their relevant parameter regimes may simply keep a_n/a_m below the threshold. Finally, anti-CCR5 treatment is shown to promote X4 virus even in the simplest competitive framework (Model 1).

While we predict that R5 blockers can promote X4 emergence in dual infection, we find that non-CCR5 specific, antiretroviral therapies such as HAART have the opposite effect. This prediction is in fact supported by a recent clinical trial on 15 women with X4 virus prior to undergoing HAART (Philpott et al., 2001). During HAART, the patients showed marked increases in $CD4^+$ T cell counts as well as a correlated reversion in viral tropism toward CCR5. Other groups have also found that antiretroviral treatment inhibits X4 virus (Equils et al., 2000; Melby et al., 2006; Skrabal et al., 2003). However, one group claims that HAART promotes R5 to X4 switching (Delobel et al., 2005). Delobel and colleagues' conclusion arises from an analysis of the genotypes of peripheral blood mononuclear cells of patients on HAART for more than three years with viral loads below detection. Notwithstanding the overall utility of testing coreceptor usage when viremia levels are below detection, the genotypic algorithms used by Delobel et al. (Delobel et al., 2005) do not always correctly predict the actual coreceptor usage (Low et al., 2007). To avoid such errors, the Philpott study, which found preferential suppression of X4 strains in patients on HAART, used a phenotypic MT-2 cell line characterization and a direct HOS-CD4 coreceptor binding assay in addition to a genotypic V3 analysis (Philpott et al., 2001).

An activation-induced switch and its deleterious clinical effects are also consistent with the proposed new paradigm of lentiviral pathogenicity, which argues that immune overactivation is the distinguishing characteristic of symptomatic lentiviral infection in new hosts as opposed to asymptomatic lentiviral infection in natural hosts (Chakrabarti et al., 2000; Hazenberg et al., 2000a; Hazenberg et al., 2003b; Mohri et al., 2001; Paiardini et al., 2008; Sachsenberg et al., 1998; Silvestri, 2005; Silvestri et al., 2007; Silvestri et al., 2003). Evidence of a correlation between immune activation and disease progression is also supported by the fact that T cell activation levels are lowered almost immediately following successful HAART (Hazenberg et al., 2000a).

One might argue that phenotypic switching in HIV has little to do with target-cell activation levels and is instead the result of cumulative mutations that occur over the course of HIV. One would have two reasons for such an argument. First, given its status as a retrovirus, HIV is extremely prone to mutation (Mansky and Temin, 1995). Second, it takes very few mutations to go from R5 to X4 virus. For example, Ho et al. showed that in *rhesus macaques* ten amino acid changes in the V3 loop of an X4-tropic virus are sufficient to modify viral coreceptor usage to CCR5 (Ho et al., 2005). That said, many V3 mutations yield viruses with lower fitness,

implying that fitness troughs exist between R5 and X4 variants (Pastore et al., 2006). Perhaps as a result, mutation between R5 and X4 strains does not seem to be common *in vivo* (Farber and Berger, 2002; Trkola et al., 2002). When drugs are employed to selectively block CCR5 in cases of R5-only infection, HIV's method of escape is not to evolve tropism for CXCR4 but to find a novel way of binding CCR5 despite the blockage (Tilton et al.; Trkola et al., 2002). Finally, X4 is simply outcompeted by R5 during early dual-infection, arguing in favor of an early exogenous selection pressure toward R5, which is mitigated over the course of infection in switching patients.

Current data are insufficient to test our conclusion that the efficacy of CCR5 blockers in dual infection depends on k_{M4}, because k_{M4} has an unknown value. The importance of testing whether CCR5 inhibitors have only partial regimes of utility stems from the fact that these treatments, in contrast to HAART and traditional antiretroviral drugs, are mainly non-toxic (Tsibris and Kuritzkes, 2007). In fact, CCR5Δ32/Δ32, a natural deletion mutation that prevents CCR5 expression, is found in 1% of humans with no known side-effects. The question associated with CCR5 blockers is whether they promote X4 in R5's stead.

This is because X4 virus quickly depletes R5-immune naïve $CD4^+$ T cells, compounding the earlier immunodeficiency that R5 engendered in the memory compartment. Naïve $CD4^+$ T cells are the source for new memory cells and a prime defense against unseen infections: hence, the victim of a phenotypic switch gets the worst of both worlds—memory $CD4^+$ T cell loss by R5 followed by naïve $CD4^+$ T cell loss by X4—greatly lessening the chance of survival.

Methods

Models 1-3 were first solved numerically using the program Berkeley Madonna. We used the Rosenbruck algorithm for solving stiff ODEs with the parameters given in the Figures. Other than k_{N4}, k_{M4}, k_{M5} and f, these parameters have been estimated from *in vivo* measurements. Further, V_4 and V_5 were each given initial values of 1000 virions/ml, as in (Ribeiro et al., 2006; Weinberger et al., 2009), which reflects experiments in macaques in which high viral doses are given to ensure infection (Wolinsky et al., 2004).

Since the purpose of the first two models is to motivate the added complexity in Model 3 and since our main conclusions are taken from Model 3, we offer a justification of the parameter values for Model 3. In particular, λ, the rate at which naïve $CD4^+$ T cells emigrate from the thymus, has been shown to remain relatively constant during HIV infection (Hazenberg et al., 2000b). Following a recent theoretical analysis (Weinberger et al., 2003), we set λ to the constant value of 33 cells/(μl•day). The viral clearance rate, c, has been directly measured to have an average value of 23/day (Ramratnam et al., 1999). The rate of virion production by productively infected cells, p, was set to 2100/day, which is line with the in vivo measure in (Haase et al., 1996) but smaller than the value reported in (Chen et al., 2007). Finally, we set the infected cell death rate δ to 0.5/day, following the measurements in (Perelson et al., 1996), although values as high as 1.0/day are also feasible (Markowitz et al., 2003). The final four parameters k_{N4}, k_{M4}, k_{M5} and f have unknown values, but can nonetheless be substantially restricted. The fraction of naïve $CD4^+$ T cells that are activated by antigen, f, is between 0 and 1. Furthermore, the FACS data summarized in Table 1 leads us to restrict the infection rate

coefficients as follows: $k_{N4} >> k_{M4}$, $k_{M5} >> k_{M4}$. We chose exact values for these 4 parameters, subject to the above constraints, by repeated simulations of our Models so as to produce the general dynamics of long-term HIV infection, including the common phenotypic switch.

We also note that because simulations require an exact form for a_n and a_m, we used the particular form fit in (1). Of course, the analysis throughout this paper shows that we can apply any equations which satisfy (2), with obvious parameter adjustments.

Subsequent to these simulations, we reproduced our work in MATLAB (with the stiff solver, ode23s) so that we could generate three-dimensional plots and show that the switch is *accelerated* when CCR5 is blocked in "competitive" regimes (i.e., those situations in which k_{M4} is relatively large). All code is available upon request.

Acknowledgements

We thank U. Andrews, H. Dahari, H. Green, P. Hraber, M. Plucinski, L. Rong, and R. Starfield for helpful discussions. We also thank two anonymous reviewers for feedback that greatly improved this manuscript. ADW wishes to acknowledge L. Weinberger, W. Getz, H. Lecar, and R. Ribeiro for extensive and informative suggestions.

Portions of this work were done under the auspices of the U. S. Department of Energy under contract DE-AC52-06NA25396 and supported by NIH grants AI28433, RR06555, and P20-RR18754 (ASP). ADW was also supported by a California HIV/AIDS Research Program Dissertation Award.

Table 1: CCR5 and CXCR4 Expression Patterns on Lymphocytes

Cell Type	Mean % $CXCR4^+$	Mean % $CCR5^+$	Median 12G5 (X4) antibody-binding sites per cell	Median 2D7 (R5) antibody-binding sites per cell
All lymphocytes	69.6	14.1	1,572	593
$CCR5^+$ or $CXCR4^+$ only			3,387	8345
$CD4^+$,$CD45RA^+$,$CD62L^+$ *	91.3	2.1	3386	0
$CD4^+$,$CD45RA^+$,$CD62L^-$ *	40.8	0	902	0
$CD4^+$,$CD45RO^+$,$CD62L^-$ *	36.8	74.9	505	9576
$CD4^+$,$CD45RO^+$,$CD62L^+$ *	63.9	48.5	1013	4741
GALT $CD4^+$ **	45	60	N/A	N/A
Blood/Lymph Node $CD4^+$ **	80	5	N/A	N/A

Note that CD62L, L-selectin, is a marker displayed by quiescent $CD4^+$ T cells, naïve or memory (Hengel et al., 2003).

* Human PBMCs as measured in (Lee et al., 1999).
** From *rhesus macaques* (Veazey et al., 2000)

Figure Legends

Figure 1. Simulations of Model 1 show that except for changes to k_5, increasing virulence accelerates the R5 to X4 switch.

In (a) we simulate a common clinical outcome in which a spontaneous phenotypic switch occurs after 3-4 years, with a concomitant decline in $CD4^+$ T cell counts. The parameters used are λ= 33 cells/(µl•day), c= 23/day, p = 5750/day, δ=0.7/day, k_4=5•10^{-4} µl/(virions•day), and k_5 = 10^{-4} µl/(virions•day). The subsequent subfigures represent small modifications of the parameter regime in (a). Respectively, we modify: X4's infection rate coefficient k_4 in (b), the viral clearance rate c in (c), and the rate of viral production from infected cells p in (d). As *Equation 4* shows, parameter changes—with the exception of k_5—which enhance infection accelerate the switch (bottom panels of (b), (c), (d)), while those that dampen infection hinder switching (top panels of (b), (c), (d)). In each top panel (i.e., decreased infection regime), X4 stays below detection for all 5000 days of the simulation.

Figure 2. Even in the simplest target-cell competition model (Model 1), 'diminishing returns' can cause CCR5 blockers to promote X4 Emergence (unless noted, parameters have values from 1a).

While Fig. 1 implies that increasing the level of infection accelerates R5 to X4 switches, changes to k_5 present more complicated kinetics. In (a) the model is simulated at two increased levels of k_5 relative to (1a). As expected from Fig. 1, an initial increase of the R5 infection rate coefficient to k_5 = 1.5•10^{-4} µl/(virions•day) accelerates X4 emergence (a, top panel). However, increasing k_5 more significantly to 3•10^{-4} µl/(virions•day)—prevents X4's onset (a, bottom panel). In (b), X4 is shown to go extinct as k_5 crosses a threshold. In (c), where k_5=3•10^{-4} µl/(virions•day), anti-CCR5 therapy with efficacy below ~70% actually promotes X4 emergence. This occurs because k_5 is not decreased beyond the diminishing returns regime. In (d) combining reverse transcriptase and protease inhibitors (representing the anti-HIV drug cocktail known as HAART), which target R5 and X4 strains equally, is shown to prevent X4 emergence and increase CD4 counts in Model 1. Similarly, in (e) combining anti-CCR5 and anti-CXCR4 therapy is shown to prevent X4 emergence and increase CD4 counts in Model 1.

Figure 3. In the more biologically representative Model 3, CCR5 inhibitors also fail to suppress X4 emergence and immunodeficiency in "competitive regimes."

Model 3 is simulated in two representative parameter regimes, labeled "non-competitive" and "competitive," respectively. In the non-competitive regime, the parameter values are: λ=33 cells/(µl•day), c=23/day, p=2100/day, f=0.8, δ=0.5/day, k_{N4}=0.00108 µl/(virions•day), k_{M4}= 4•10^{-5} µl/(virions•day), and k_{M5}=0.0068 µl/(virions•day). In the competitive regime, k_{M4} is *increased* to 5•10^{-4} µl/(virions•day) and k_{N4} is correspondingly *decreased* to 0.001 µl/(virions•day) to keep X4 in check, but all other parameters are held constant. The crucial distinction between the two regimes, which display similar dynamics in the absence of anti-CCR5 treatment, is that in the non-competitive regime k_{M4} is relatively small, preventing X4 from infecting many $CCR5^+$, $CXCR4^+$ memory $CD4^+$ T cells after CCR5 inhibition. Thus, in the "non-competitive regime" (a, left panel) as we increase the efficacy of a permanent CCR5 inhibitor first given at t=180 days and maintained for all future time, steady-state $CD4^+$ T cell counts rise and steady-state X4

levels fall. However, in the competitive regime (a, right panel), anti-CCR5 treatment does not depress steady-state X4 levels and it does not increase steady-state $CD4^+$ T cell counts, either. While steady-state X4 levels and $CD4^+$ T cell counts remain unchanged under this treatment schedule, in (b) the *time* at which this steady-state occurs is shown to be earlier. That is, in the competitive regime, CCR5 inhibitors *accelerate* $CD4^+$ T cell depletion. In (c), this accelerated outcome is seen to be the result of a temporal gain in memory $CD4^+$ T cell counts triggered by an initially effective anti-CCR5 treatment. The "competitive" X4 virus then increases by infecting newly generated memory cells. This strengthens X4 infection, after which it can severely deplete naïve $CD4^+$ T cells, and counteract any initial gains in memory $CD4^+$ T cells.

Figure 1a

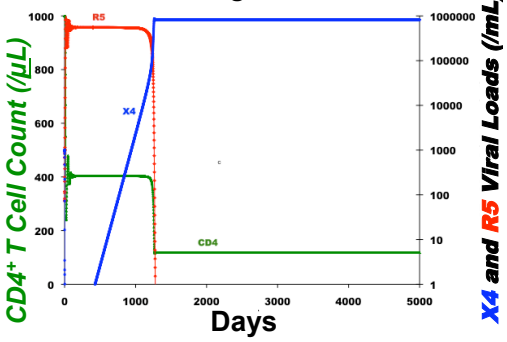


Figure 1b

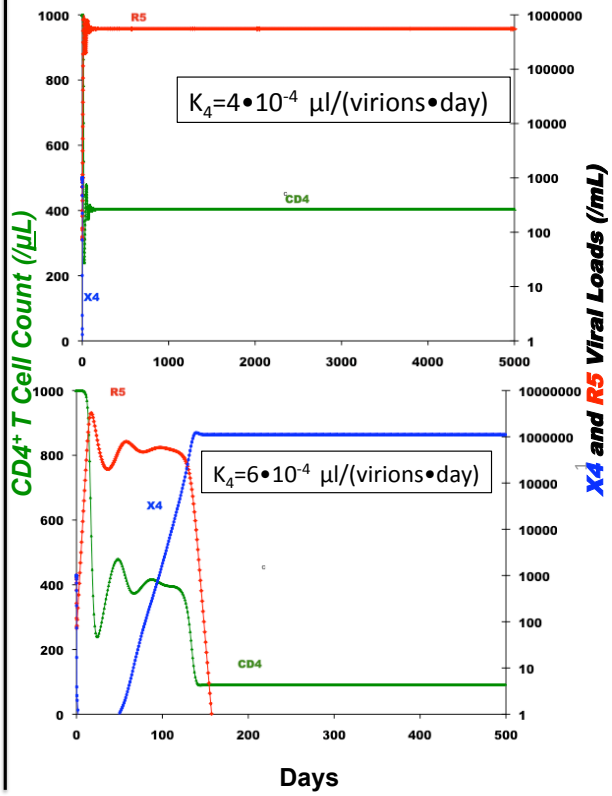


Figure 1c

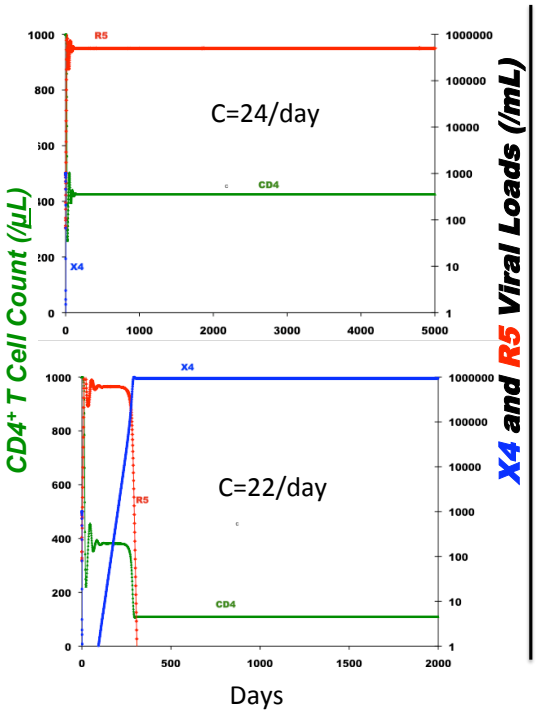


Figure 1d

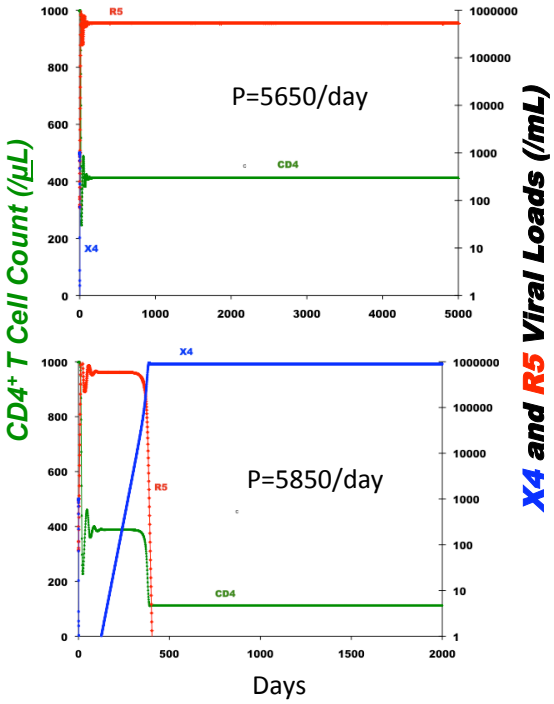


Figure 2a

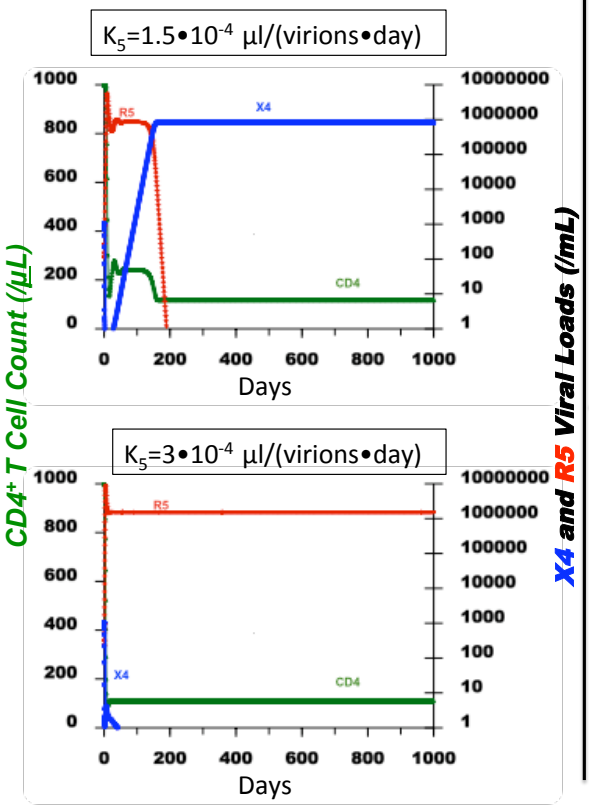


Figure 2b

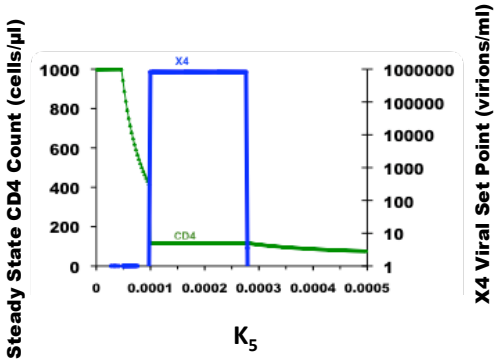


Figure 2c

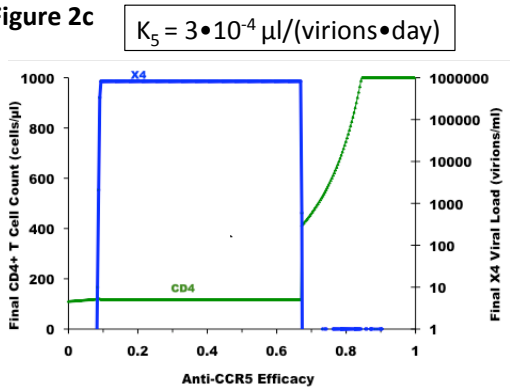


Figure 2d

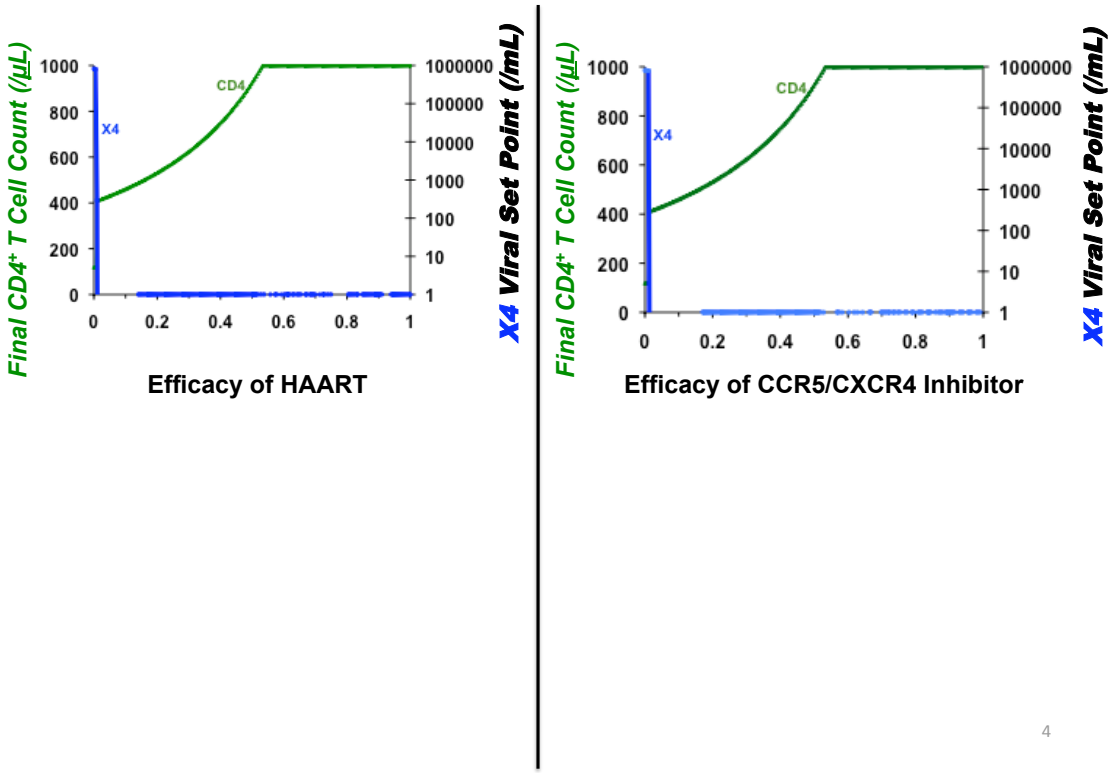

Final CD4+ T Cell Count (/µL)
1000
800
600
400
200
0
X4
CD4
0
0.2
0.4
0.6
0.8
1
1000000
100000
10000
1000
100
10
1
X4 Viral Set Point (/mL)
Efficacy of HAART
Final CD4+ T Cell Count (/µL)
1000
800
600
400
200
0
X4
CD4
0
0.2
0.4
0.6
0.8
1
1000000
100000
10000
1000
100
10
1
X4 Viral Set Point (/mL)
Efficacy of CCR5/CXCR4 Inhibitor
4

Figure 3a

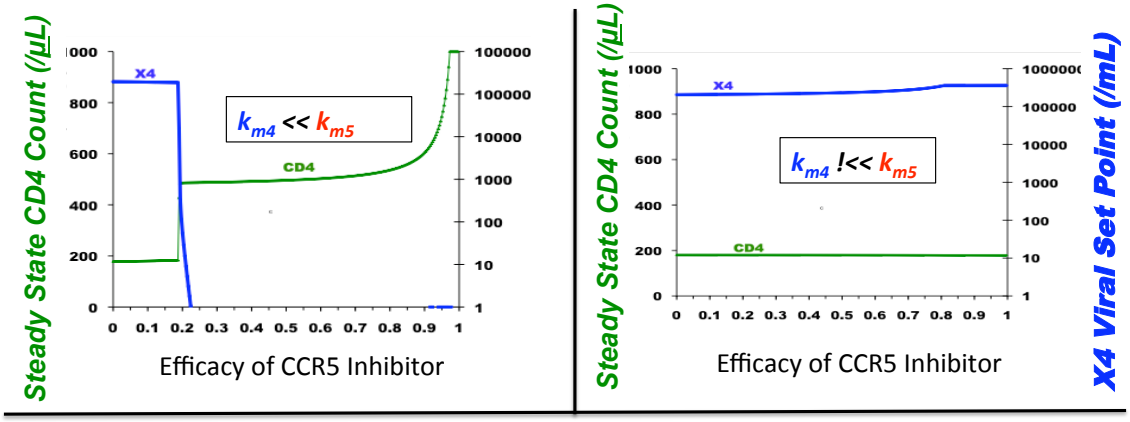


Figure 3b

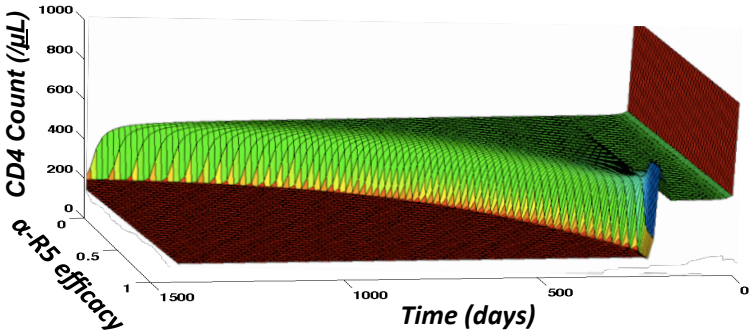
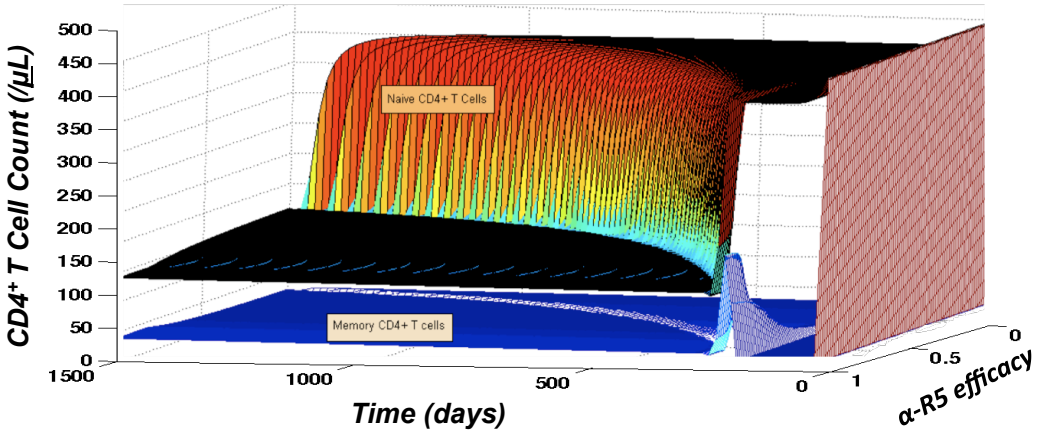
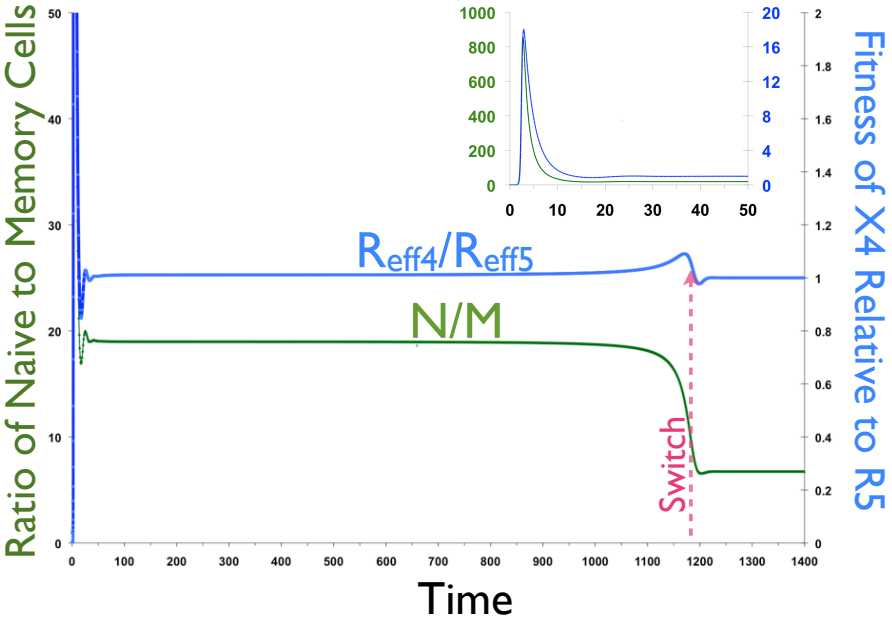


Figure 3c


CD4+ T Cell Count (/µL)
500
450
400
350
300
250
200
150
100
50
0
Naive CD4+ T Cells
Memory CD4+ T cells
1500
1000
500
0
Time (days)
1
0.5
0
α-R5 efficacy

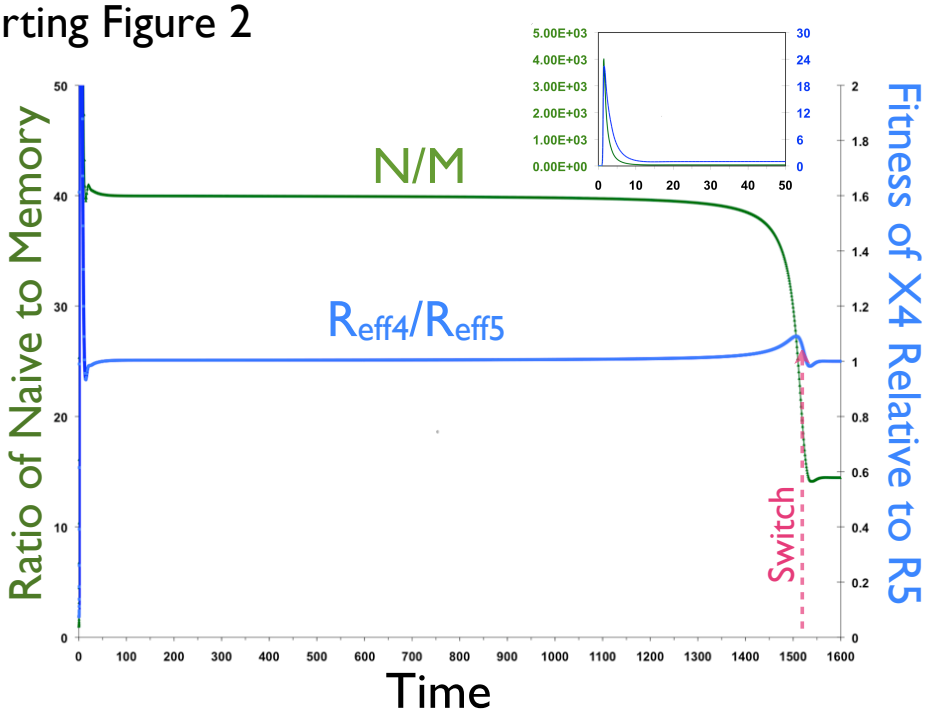
6

Supporting Figure 1



Evolution of N/M and R_{eff4}/R_{eff5} in Model 2. A representative simulation of Model 2 illustrates how the fraction of naïve to memory CD4+ T cells (N/M) increases rapidly during the initial phase of HIV infection (inset shows a close-up of the first 50 days), decreasing to a quasi-set point that slowly drops throughout the pre-switch period. With N/M relatively constant in the period before the switch, R_{eff4}/R_{eff5} slowly increases due to a_n/a_m rising as CD4 counts decrease. N/M decreases post-switch when X4 depletes naïve CD4+ T cells, allowing for coexistence in Model 2 as R_{eff4}/R_{eff5} drops to 1.

Supporting Figure 2



Evolution of N/M and R_{eff4}/R_{eff5} in Model 3. A representative simulation of Model 3 shows the same pattern as observed above in Model 2. The fraction of naïve to memory CD4+ T cells (N/M) increases rapidly during the initial phase of HIV infection (inset shows a close-up of first 50 days), decreasing to a quasi-set point that slowly decreases throughout the pre-switch period. With N/M relatively constant in the period preceding the switch, R_{eff4}/R_{eff5} slowly increases as a_n/a_m rises from HIV-driven reductions in the CD4 count. Again N/M decreases post-switch as X4 depletes naïve CD4+ T cells, allowing for coexistence in Model 3 when $R_{eff4}/R_{eff5} = 1$.

References

Blaak, H., van't Wout, A.B., Brouwer, M., Hooibrink, B., Hovenkamp, E., and Schuitemaker, H., 2000. In vivo HIV-1 infection of CD45RA(+)CD4(+) T cells is established primarily by syncytium-inducing variants and correlates with the rate of CD4(+) T cell decline. Proc Natl Acad Sci U S A 97, 1269-74.

Callaway, D.S., Ribeiro, R.M., and Nowak, M.A., 1999. Virus phenotype switching and disease progression in HIV-1 infection. Proc Roy Sci B 266, 2523-30.

Casper, C.H., Clevestig, P., Carlenor, E., Leitner, T., Anzen, B., Lidman, K., Belfrage, E., Albert, J., Bohlin, A.B., Naver, L., Lindgren, S., Fenyo, E.M., and Ehrnst, A.C., 2002. Link between the X4 phenotype in human immunodeficiency virus type 1-infected mothers and their children, despite the early presence of R5 in the child. J Infect Dis 186, 914-21.

Chakrabarti, L.A., Lewin, S.R., Zhang, L., Gettie, A., Luckay, A., Martin, L.N., Skulsky, E., Ho, D.D., Cheng-Mayer, C., and Marx, P.A., 2000. Normal T-cell turnover in sooty mangabeys harboring active simian immunodeficiency virus infection. J Virol 74, 1209-23.

Chen, H.Y., Di Mascio, M., Perelson, A.S., Ho, D.D., and Zhang, L., 2007. Determination of virus burst size in vivo using a single-cycle SIV in rhesus macaques. Proc Natl Acad Sci U S A 104, 19079-84.

Coakley, E., Petropoulos, C.J., and Whitcomb, J.M., 2005. Assessing chemokine co-receptor usage in HIV. Curr Opin Infect Dis 18, 9-15.

Cornelissen, M., Mulder-Kampinga, G., Veenstra, J., Zorgdrager, F., Kuiken, C., Hartman, S., Dekker, J., van der Hoek, L., Sol, C., Coutinho, R., and et al., 1995. Syncytium-inducing (SI) phenotype suppression at seroconversion after intramuscular inoculation of a non-syncytium-inducing/SI phenotypically mixed human immunodeficiency virus population. J Virol 69, 1810-8.

Delobel, P., Sandres-Saune, K., Cazabat, M., Pasquier, C., Marchou, B., Massip, P., and Izopet, J., 2005. R5 to X4 switch of the predominant HIV-1 population in cellular reservoirs during effective highly active antiretroviral therapy. J Acquir Immune Defic Syndr 38, 382-92.

Douek, D.C., 2003. Disrupting T-cell homeostasis: how HIV-1 infection causes disease. AIDS Rev 5, 172-7.

Equils, O., Garratty, E., Wei, L.S., Plaeger, S., Tapia, M., Deville, J., Krogstad, P., Sim, M.S., Nielsen, K., and Bryson, Y.J., 2000. Recovery of replication-competent virus from CD4 T cell reservoirs and change in coreceptor use in human immunodeficiency virus type 1-infected children responding to highly active antiretroviral therapy. J Infect Dis 182, 751-7.

Farber, J.M., and Berger, E.A., 2002. HIV's response to a CCR5 inhibitor: I'd rather tighten than switch! Proc Natl Acad Sci U S A 99, 1749-51.

Giorgi, J.V., Hultin, L.E., McKeating, J.A., Johnson, T.D., Owens, B., Jacobson, L.P., Shih, R., Lewis, J., Wiley, D.J., Phair, J.P., Wolinsky, S.M., and Detels, R., 1999. Shorter survival in advanced human immunodeficiency virus type 1 infection is more closely associated with T lymphocyte activation than with plasma virus burden or virus chemokine coreceptor usage. J Infect Dis 179, 859-70.

Gondois-Rey, F., Grivel, J.C., Biancotto, A., Pion, M., Vigne, R., Margolis, L.B., and Hirsch, I., 2002. Segregation of R5 and X4 HIV-1 variants to memory T cell subsets differentially expressing CD62L in ex vivo infected human lymphoid tissue. AIDS 16, 1245-9.

Gondois-Rey, F., Biancotto, A., Fernandez, M.A., Bettendroffer, L., Blazkova, J., Trejbalova, K., Pion, M., and Hirsch, I., 2006. R5 variants of human immunodeficiency virus type 1 preferentially infect CD62L- CD4+ T cells and are potentially resistant to nucleoside reverse transcriptase inhibitors. J Virol 80, 854-65.

Haase, A.T., 1999. Population biology of HIV-1 infection: viral and CD4+ T cell demographics and dynamics in lymphatic tissues. Annu Rev Immunol 17, 625-56.

Haase, A.T., Henry, K., Zupancic, M., Sedgewick, G., Faust, R.A., Melroe, H., Cavert, W., Gebhard, K., Staskus, K., Zhang, Z.Q., Dailey, P.J., Balfour, H.H., Jr., Erice, A., and Perelson, A.S., 1996. Quantitative image analysis of HIV-1 infection in lymphoid tissue. Science 274, 985-9.

Harouse, J.M., Buckner, C., Gettie, A., Fuller, R., Bohm, R., Blanchard, J., and Cheng-Mayer, C., 2003. CD8+ T cell-mediated CXC chemokine receptor 4-simian/human immunodeficiency virus suppression in dually infected rhesus macaques. Proc Natl Acad Sci U S A 100, 10977-82.

Hazenberg, M.D., Otto, S.A., Hamann, D., Roos, M.T., Schuitemaker, H., de Boer, R.J., and Miedema, F., 2003a. Depletion of naive CD4 T cells by CXCR4-using HIV-1 variants occurs mainly through increased T-cell death and activation. AIDS 17, 1419-24.

Hazenberg, M.D., Stuart, J.W., Otto, S.A., Borleffs, J.C., Boucher, C.A., de Boer, R.J., Miedema, F., and Hamann, D., 2000a. T-cell division in human immunodeficiency virus (HIV)-1 infection is mainly due to immune activation: a longitudinal analysis in patients before and during highly active antiretroviral therapy (HAART). Blood 95, 249-55.

Hazenberg, M.D., Otto, S.A., van Benthem, B.H., Roos, M.T., Coutinho, R.A., Lange, J.M., Hamann, D., Prins, M., and Miedema, F., 2003b. Persistent immune activation in HIV-1 infection is associated with progression to AIDS. AIDS 17, 1881-8.

Hazenberg, M.D., Otto, S.A., Cohen Stuart, J.W., Verschuren, M.C., Borleffs, J.C., Boucher, C.A., Coutinho, R.A., Lange, J.M., Rinke de Wit, T.F., Tsegaye, A., van Dongen, J.J., Hamann, D., de Boer, R.J., and Miedema, F., 2000b. Increased cell division but not thymic dysfunction rapidly affects the T-cell receptor excision circle content of the naive T cell population in HIV-1 infection. Nat Med 6, 1036-42.

Hengel, R.L., Thaker, V., Pavlick, M.V., Metcalf, J.A., Dennis, G., Jr., Yang, J., Lempicki, R.A., Sereti, I., and Lane, H.C., 2003. Cutting edge: L-selectin (CD62L) expression distinguishes small resting memory CD4+ T cells that preferentially respond to recall antigen. J Immunol 170, 28-32.

Ho, D.D., Neumann, A.U., Perelson, A.S., Chen, W., Leonard, J.M., and Markowitz, M., 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature 373, 123-6.

Ho, S.H., Shek, L., Gettie, A., Blanchard, J., and Cheng-Mayer, C., 2005. V3 loop-determined coreceptor preference dictates the dynamics of CD4+-T-cell loss in simian-human immunodeficiency virus-infected macaques. J Virol 79, 12296-303.

Joly, M., and Pinto, J.M., 2005. CXCR4 and CCR5 regulation and expression patterns on T- and monocyte-macrophage cell lineages: implications for susceptibility to infection by HIV-1. Math Biosci 195, 92-126.

Keele, B.F., Giorgi, E.E., Salazar-Gonzalez, J.F., Decker, J.M., Pham, K.T., Salazar, M.G., Sun, C., Grayson, T., Wang, S., Li, H., Wei, X., Jiang, C., Kirchherr, J.L., Gao, F., Anderson, J.A., Ping, L.H., Swanstrom, R., Tomaras, G.D., Blattner, W.A., Goepfert, P.A., Kilby, J.M., Saag, M.S., Delwart, E.L., Busch, M.P., Cohen, M.S., Montefiori, D.C., Haynes, B.F., Gaschen, B., Athreya, G.S., Lee, H.Y., Wood, N., Seoighe, C., Perelson, A.S., Bhattacharya, T., Korber, B.T., Hahn, B.H., and Shaw, G.M., 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A 105, 7552-7.

Lee, B., Sharron, M., Montaner, L.J., Weissman, D., and Doms, R.W., 1999. Quantification of CD4, CCR5, and CXCR4 levels on lymphocyte subsets, dendritic cells, and differentially conditioned monocyte-derived macrophages. Proc Natl Acad Sci U S A 96, 5215-20.

Li, Q., Duan, L., Estes, J.D., Ma, Z.M., Rourke, T., Wang, Y., Reilly, C., Carlis, J., Miller, C.J., and Haase, A.T., 2005. Peak SIV replication in resting memory CD4+ T cells depletes gut lamina propria CD4+ T cells. Nature 434, 1148-52.

Low, A.J., Dong, W., Chan, D., Sing, T., Swanstrom, R., Jensen, M., Pillai, S., Good, B., and Harrigan, P.R., 2007. Current V3 genotyping algorithms are inadequate for predicting X4 co-receptor usage in clinical isolates. Aids 21, F17-24.

Mansky, L.M., and Temin, H.M., 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J Virol 69, 5087-94.

Markowitz, M., Louie, M., Hurley, A., Sun, E., Di Mascio, M., Perelson, A.S., and Ho, D.D., 2003. A novel antiviral intervention results in more accurate assessment of human immunodeficiency virus type 1 replication dynamics and T-cell decay in vivo. J Virol 77, 5037-8.

Melby, T., Despirito, M., Demasi, R., Heilek-Snyder, G., Greenberg, M.L., and Graham, N., 2006. HIV-1 coreceptor use in triple-class treatment-experienced patients: baseline prevalence, correlates, and relationship to enfuvirtide response. J Infect Dis 194, 238-46.

Mohri, H., Bonhoeffer, S., Monard, S., Perelson, A.S., and Ho, D.D., 1998. Rapid turnover of T lymphocytes in SIV-infected rhesus macaques. Science 279, 1223-7.

Mohri, H., Perelson, A.S., Tung, K., Ribeiro, R.M., Ramratnam, B., Markowitz, M., Kost, R., Hurley, A., Weinberger, L., Cesar, D., Hellerstein, M.K., and Ho, D.D., 2001. Increased turnover of T lymphocytes in HIV-1 infection and its reduction by antiretroviral therapy. J Exp Med 194, 1277-87.

Moore, J.P., Kitchen, S.G., Pugach, P., and Zack, J.A., 2004. The CCR5 and CXCR4 coreceptors--central to understanding the transmission and pathogenesis of human immunodeficiency virus type 1 infection. AIDS Res Hum Retroviruses 20, 111-26.

Paiardini, M., Frank, I., Pandrea, I., Apetrei, C., and Silvestri, G., 2008. Mucosal immune dysfunction in AIDS pathogenesis. AIDS Rev 10, 36-46.

Pastore, C., Nedellec, R., Ramos, A., Pontow, S., Ratner, L., and Mosier, D.E., 2006. Human immunodeficiency virus type 1 coreceptor switching: V1/V2 gain-of-fitness mutations compensate for V3 loss-of-fitness mutations. J Virol 80, 750-8.

Perelson, A.S., Neumann, A.U., Markowitz, M., Leonard, J.M., and Ho, D.D., 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science 271, 1582-6.

Philpott, S., Weiser, B., Anastos, K., Kitchen, C.M., Robison, E., Meyer, W.A., 3rd, Sacks, H.S., Mathur-Wagh, U., Brunner, C., and Burger, H., 2001. Preferential suppression of CXCR4-specific strains of HIV-1 by antiviral therapy. J Clin Invest 107, 431-8.

Ramratnam, B., Bonhoeffer, S., Binley, J., Hurley, A., Zhang, L., Mittler, J.E., Markowitz, M., Moore, J.P., Perelson, A.S., and Ho, D.D., 1999. Rapid production and clearance of HIV-1 and hepatitis C virus assessed by large volume plasma apheresis. Lancet 354, 1782-5.

Regoes, R., and Bonhoeffer, S., 2002. HIV coreceptor usage and drug treatment. J Theor Biol 217, 443-57.

Regoes, R.R., and Bonhoeffer, S., 2005. The HIV coreceptor switch: a population dynamical perspective. Trends Microbiol 13, 269-77.

Ribeiro, R.M., Hazenberg, M.D., Perelson, A.S., and Davenport, M.P., 2006. Naive and memory cell turnover as drivers of CCR5-to-CXCR4 tropism switch in human immunodeficiency virus type 1: implications for therapy. J Virol 80, 802-9.

Sachsenberg, N., Perelson, A.S., Yerly, S., Schockmel, G.A., Leduc, D., Hirschel, B., and Perrin, L., 1998. Turnover of CD4+ and CD8+ T lymphocytes in HIV-1 infection as measured by Ki-67 antigen. J Exp Med 187, 1295-303.

Silvestri, G., 2005. Naturally SIV-infected sooty mangabeys: are we closer to understanding why they do not develop AIDS? J Med Primatol 34, 243-52.

Silvestri, G., Paiardini, M., Pandrea, I., Lederman, M.M., and Sodora, D.L., 2007. Understanding the benign nature of SIV infection in natural hosts. J Clin Invest 117, 3148-54.

Silvestri, G., Sodora, D.L., Koup, R.A., Paiardini, M., O'Neil, S.P., McClure, H.M., Staprans, S.I., and Feinberg, M.B., 2003. Nonpathogenic SIV infection of sooty mangabeys is characterized by limited bystander immunopathology despite chronic high-level viremia. Immunity 18, 441-52.

Skrabal, K., Trouplin, V., Labrosse, B., Obry, V., Damond, F., Hance, A.J., Clavel, F., and Mammano, F., 2003. Impact of antiretroviral treatment on the tropism of HIV-1 plasma virus populations. Aids 17, 809-14.

Suarez, A., Mozo, L., and Gutierrez, C., 2002. Generation of CD4(+)CD45RA(+) effector T cells by stimulation in the presence of cyclic adenosine 5'-monophosphate-elevating agents. J Immunol 169, 1159-67.

Tilton, J.C., Amrine-Madsen, H., Miamidian, J.L., Kitrinos, K.M., Pfaff, J., Demarest, J.F., Ray, N., Jeffrey, J.L., Labranche, C.C., and Doms, R.W., HIV type 1 from a patient with baseline resistance to CCR5 antagonists uses drug-bound receptor for entry. AIDS Res Hum Retroviruses 26, 13-24.

Trkola, A., Kuhmann, S.E., Strizki, J.M., Maxwell, E., Ketas, T., Morgan, T., Pugach, P., Xu, S., Wojcik, L., Tagat, J., Palani, A., Shapiro, S., Clader, J.W., McCombie, S., Reyes, G.R., Baroudy, B.M., and Moore, J.P., 2002. HIV-1 escape from a small molecule, CCR5-specific entry inhibitor does not involve CXCR4 use. Proc Natl Acad Sci U S A 99, 395-400.

Tsibris, A.M., and Kuritzkes, D.R., 2007. Chemokine antagonists as therapeutics: focus on HIV-1. Annu Rev Med 58, 445-59.

Unutmaz, D., Pileri, P., and Abrignani, S., 1994. Antigen-independent activation of naive and memory resting T cells by a cytokine combination. J Exp Med 180, 1159-64.

Unutmaz, D., Baldoni, F., and Abrignani, S., 1995. Human naive T cells activated by cytokines differentiate into a split phenotype with functional features intermediate between naive and memory T cells. Int Immunol 7, 1417-24.

Veazey, R.S., Mansfield, K.G., Tham, I.C., Carville, A.C., Shvetz, D.E., Forand, A.E., and Lackner, A.A., 2000. Dynamics of CCR5 expression by CD4(+) T cells in lymphoid tissues during simian immunodeficiency virus infection. J Virol 74, 11001-7.

Wei, X., Ghosh, S.K., Taylor, M.E., Johnson, V.A., Emini, E.A., Deutsch, P., Lifson, J.D., Bonhoeffer, S., Nowak, M.A., Hahn, B.H., and et al., 1995. Viral dynamics in human immunodeficiency virus type 1 infection. Nature 373, 117-22.

Weinberger, A.D., Perelson, A.S., Ribeiro, R.M., and Weinberger, L.S., 2009. Accelerated immunodeficiency by anti-CCR5 treatment in HIV infection. PLoS Comput Biol 5, e1000467.

Weinberger, L.S., Schaffer, D.V., and Arkin, A.P., 2003. Theoretical design of a gene therapy to prevent AIDS but not human immunodeficiency virus type 1 infection. J Virol 77, 10028-36.

Wodarz, D., and Nowak, M.A., 1998. The effect of different immune responses on the evolution of virulent CXCR4-tropic HIV. Proc Roy Sci B 265, 2149-58.

Wodarz, D., Lloyd, A.L., Jansen, V.A., and Nowak, M.A., 1999. Dynamics of macrophage and T cell infection by HIV. J Theor Biol 196, 101-13.

Wolinsky, S.M., Veazey, R.S., Kunstman, K.J., Klasse, P.J., Dufour, J., Marozsan, A.J., Springer, M.S., and Moore, J.P., 2004. Effect of a CCR5 inhibitor on viral loads in macaques dual-infected with R5 and X4 primate immunodeficiency viruses. Virology 328, 19-29.

Zhang, Z.Q., Wietgrefe, S.W., Li, Q., Shore, M.D., Duan, L., Reilly, C., Lifson, J.D., and Haase, A.T., 2004. Roles of substrate availability and infection of resting and activated CD4+ T cells in transmission and acute simian immunodeficiency virus infection. Proc Natl Acad Sci U S A 101, 5640-5.

Chapter Four

Persisting Low-Abundance Viruses Shape Microbial CRISPR-based Immunity

Ariel D. Weinberger, Christine L. Sun, Mateusz M. Pluciński, Vincent J. Denef, Brian C. Thomas, Philippe Horvath, Rodolphe Barrangou, Wayne M. Getz & Jillian F. Banfield

Introduction

The previous chapters explored how, in HIV infection, the R5 and X4 strains are able to coexist due to phenotypic differences on the surfaces of CD4+ HIV target cells: i.e., the unique lack of the CCR5 coreceptor on the subset of naïve CD4+ T cells. Yet, even non-coding, purely genotypic diversity among host cells can impart diversifying selection on infecting viral populations. A clear example of this emerges in a hypermutative, non-coding region found in the genomes of ~40% Bacteria and ~90% of Archaea (Horvath and Barrangou; Marraffini and Sontheimer).

Known as Clustered Regularly Interspaced Short Palindromic Repeats **(**CRISPR), CRISPR loci provide prokaryotes with adaptive immunity against viruses and plasmods through continual acquisitions of new short DNA sequences from invasive viral and plasmid genomes. Utilizing adjacent CRISPR-associated (Cas) proteins (Makarova et al., 2006), CRISPRs iteratively incorporate 21-72 base-pair 'spacer' sequences excised from targeted genomic regions known as 'proto-spacers' in invading viruses and plasmids (Barrangou et al., 2007; Horvath and Barrangou; Marraffini and Sontheimer; Mojica et al., 2009; van der Oost et al., 2009). New spacers are added at the 5' ends of unidirectionally expanding CRISPR loci between highly synonymous 23-47 base-pair 'repeat' sequences (Barrangou et al., 2007; Deveau et al., 2008; Horvath et al., 2008). Once transcribed and cleaved into functional spacer-repeat units of CRISPR RNA (crRNA), individual spacers confer immunity when any crRNA binds its complementary region during attempted invasions (Brouns et al., 2008). Viruses escape CRISPR targeting through mutations in their proto-spacers or 'proto-spacer associated motifs' (PAMs) (Barrangou et al., 2007; Deveau et al., 2008). Hosts respond by accumulating new spacers: a coevolutionary 'arms race' ensues (Andersson and Banfield, 2008; Deveau et al., 2008; Makarova et al., 2006). Unlike other adaptive immune systems, CRISPRs incorporate new immunities unidirectionally, leaving unprecedented, but incomplete, serial recordings of virus-host interactions. We combined metagenomic reconstructions of CRISPR loci across time with a newly developed multi-scale model of virus-host dynamics to analyze CRISPR-based coevolution in natural populations. Our results link documented metagenomic patterns of old-end locus clonality to an evolutionary advantage against persistent viruses, explaining why CRISPRs preserve old immune potential despite rapid viral mutation.

Metagenomic reconstructions of CRISPR loci across a natural population of archaeal I-plasma in an acid mine drainage (AMD) biofilm reveal that most spacers are shared among distinct cells and do not match currently sampled viruses. Only polyclonal new-end spacers appear to target reconstructed viruses (Fig. 1A). Increased clonality at CRISPR old-ends has previously been observed in bacterial *Escherichia coli* and archaeal *Sulfolobus islandicus* populations (Diez-Villasenor et al.; Held et al., 2010). We tested whether hosts maintain old-end spacers over multi-year periods—thousands of microbial generations—despite rapid viral coevolution. Identifying a G-plasma population that could be tracked metagenomically in the AMD system for five years, we documented old-end conservation and uniformity in two CRISPR loci (Fig. 1B). Old-end conservation occurs despite deletions of single and multiple spacer-repeat units, events also noted in previous analyses (Deveau et al., 2008; Horvath and Barrangou; Horvath et al., 2008; Tyson and Banfield, 2008). With new spacers more likely to provide immunity against current co-evolving viruses (Andersson and Banfield, 2008), selection

should cause spacer deletions to preferentially accumulate at the old-ends, eventually purging old-ends entirely. That old-ends endure in the face of selection is surprising.

Probing whether old spacers are conserved because they still confer immunity to persisting viruses, we reconstructed viral genomes from microbial community data and mapped G-plasma spacers onto these viral genomes (Appendix). While most matching spacers occur at the new-ends of G-plasma loci, some spacers with perfect identity to reconstructed viral sequences occur in older, more clonal regions at all sampled times (Figs. 1B, 1C).

To understand the coevolutionary dynamics driving old-end clonality and to probe whether rare matches of older spacers to current viruses drive old-end conservation, we developed a population-scale mathematical model of CRISPR-encoded immunity (Fig. S1: please see Appendix for Chapter 4's Supporting Figures and Tables). The model relies on two fundamental features of CRISPR: unidirectional spacer addition in hosts and mutation in viruses. Unique to host-pathogen models, the independent variable is not time but the number of virus-host interactions. Each model iteration is the period during which a parameterized, large number of virus-host interactions occurs (Table S1). Such 'interaction-based' modeling generates robust CRISPR spacer patterns across wide ranges of parameter space, avoiding the common pitfall of model sensitivity to system-specific and often-unknown ecological rate parameters. An alternative way of avoiding unknown parameterizations, used in population genetics models, assumes constant population sizes. In contrast, an interaction-based framework enables tracking relative changes in virus and host abundance levels, linking the population genetics of CRISPR immunity to population dynamics, and enabling inferences of viral blooms and host bottlenecks.

The model divides host and viral populations into 'strains.' Virions whose proto-spacers are all identical are assigned to a single strain, as are hosts with fully identical spacers (Fig. S1). Given the preset large number of interactions during each model iteration, well-mixing is assumed, with virus-host interactions distributed according to the products of host and viral strain frequencies. This selects for viral strains that productively infect dominant host strains. In agreement with laboratory experiments (Deveau et al., 2008), the more spacers host and viral strains share, the more likely CRISPRs are to prevent productive infections (Table S1). Importantly, new host and viral strains are stochastically created throughout a simulation, as hosts unidirectionally add random spacers and viruses mutate random proto-spacers. Similarly, old host and viral strains go extinct. To track the coevolution of host and viral populations across time, the model takes metagenomic snapshots of host and viral strains after each iteration, capturing patterns of CRISPR-driven immunity as they emerge.

Reconstructing model snapshots across thousands of iterations reveals how, in the absence of spacer deletions, CRISPR loci robustly converge to the pattern of old-end clonality and new-end diversity (Fig. 2A Left Panel). Old-end uniformity emerges despite initially high levels of old-end diversity (Fig. 2A Right Panel). As old-ends evolve from diversity to uniformity, an intermediate stage is predicted in which the old-ends are grouped into several sub-populations, each distinguished by its initial spacers (Fig. 2A Middle Panel; also Fig. 1A). Old-end sub-populations have similarly been reported in metagenomic reconstructions from natural environments (Held et al., 2010; Tyson and Banfield, 2008). Importantly, in all snapshots, virus-

host coevolution results in only newly added spacers matching current viruses, as in reconstructed I-plasma loci (Fig. 1A).

To capture the coevolutionary dynamics purging old-end spacer diversity, we tracked the relative frequencies of all host spacers across all iterations, using distinguishing ancestral spacers to monitor host sub-populations. Gradual declines in old-end diversity occur throughout simulations (Fig. S2), consistent with the role genetic drift plays in purging diversity. Yet, in addition to gradual fixations, model results demonstrate rapid 'selective sweeps' of individual host sub-populations. Maximizing the average silhouette width (Rousseeuw, 1987), a machine learning technique, we clustered host strains into an optimal number of sub-populations every 100 iterations (Appendix). Following the evolution of these sub-populations reveals key iterations in which ancestral diversity is abruptly lost (Fig. 2B). Verifying clustering-detected sweeps, spacers in a recently clonalized locus position were tracked through the period during which the predicted sweep occurs. A single spacer, unique to one diversifying host sub-population (Appendix), rapidly sweeps this position (Fig. 2C).

Selective sweeps are surprising given a paradigm in microbial ecology known as 'kill the winner' (Thingstad and Lignell, 1997) in which selection for viruses targeting abundant hosts prevents rapid fixations of individual host lines. Host diversity should thus be maintained (Rodriguez-Valera et al., 2009). While model simulations do show viral targeting of dominant host lines (Fig. 2C), further analyses reveal that the viral proto-spacer mutations which suppress the 'winning' hosts do not always occur before all competing host sub-populations have gone extinct (Fig. S4). Thus, despite 'kill the winner' dynamics, CRISPR-based systems result in aperiodic selective sweeps, depleting host diversity (Fig. 2B).

While unidirectional spacer addition explains the emergence of old-end clonality, random spacer deletions could eliminate old-end uniformity by deleting old-ends entirely. We probed why spacer deletions—which, by selection, should preferentially accumulate at the old-ends—do not purge old-end uniformity. To do so, we extended the model, allowing a preset fraction of spacer additions to occur with the loss of a randomly-sized, contiguous spacer block from a random starting point in the locus. A combined add/loss mechanism is consistent with experimental evidence indicating that spacer deletion occurs via homologous recombination and data showing that losses often occur with simultaneous new-end spacer additions (*9*).

The ratio of spacer addition to loss rates determines the emergent lengths of CRISPR loci upon which old-end uniformity hinges. Clonal old-ends are maintained when only 5% of spacer additions occur with deletions of one or more spacers (Fig. 3A). However, allowing spacer deletion events to occur in 50% of additions breaks the pattern of old-end uniformity by removing old-ends entirely (Fig. 3B). Given experimental data showing that many CRISPR loci maintain clonal old-ends over time (Figs. 1B, C), model results predict that the rate of spacer deletion is maintained below a critical threshold in many natural systems.

To understand why selection may preserve old-end uniformity, we compared the fitness of host strains under both low and high-loss regimes. While a low-loss rate (5%) produces consistently high levels of host immunity and thus fitness (Fig. 3A Lower Panel), dramatic dips in host immunity are observed when the probability of spacer deletion is increased to 50% (Fig.

3B Lower Panel). Troughs in the fraction of interactions in which hosts are immune predict host population bottlenecks due to viral blooms (Fig. S5). Tracking CRISPR loci through the first predicted viral bloom, model results show that immunity is primarily conferred by two contiguous older spacers (Fig. 3B). These older spacers were previously lost from many host strains after most viruses mutated the corresponding two proto-spacers (Fig. S6). Once host lines lacking these spacers became dominant, viruses preserving these proto-spacers proliferate (Fig. S6). Importantly, despite sharing two old proto-spacers, the blooming viruses are diverse new mutants (Fig. S7), as are the surviving hosts (Fig. 3B).

Remarkably, metagenomic time series reconstructions (August, 2006) captured a bloom of AMDV3b Virus (depth of sampling ~800X), coincident with a crash of its G-plasma host (Fig. 3D). AMDV3b blooms despite preexisting old-end spacers in G-plasma matching AMDV3b (Fig. 1C). Further, the bloom is highly polyclonal. Correspondingly, there is no evidence of diminished CRISPR diversity among bloom-surviving G-plasma hosts. In fact, two G-plasma sub-populations, differentiated by distinct old-end spacers, precede and survive the crash. Natural systems data thus contrast with expectations from laboratory challenge experiments, which predict that surviving hosts only differ among new-end spacers (Barrangou et al., 2007; Deveau et al., 2008). The absence of a bloom-driven bottleneck in host diversity is understandable if, as in model predictions, older spacers common to diverse loci, rather than new-end spacers found in few loci, confer immunity. A bloom of diversified viruses sharing several old spacers would also explain the bloom's surprising polyclonality.

By protecting against blooms of ancestral viral sequences, model and metagenomic data suggest that CRISPR's unique immune memory makes it optimal for environments in which viruses persist for long periods or commonly remigrate from adjacent regions. CRISPR-based immunity may thus be more prevalent in biofilms than in dilute ocean environments (Sorokin et al.). Immunity against persistent viruses may also explain CRISPR's presence in 90% of sequenced Archaea, which have disproportionately been sampled from extreme environments where viruses tend not to lyse their hosts (Prangishvili et al., 2006). More generally, proviral latency is a viral persistence strategy and among the greatest barriers to eradicating pathogens, as HIV demonstrates (Finzi et al., 1999). By maintaining old immunities, CRISPR may represent the first known immune system tuned against viruses persisting through lysogeny or remigration.

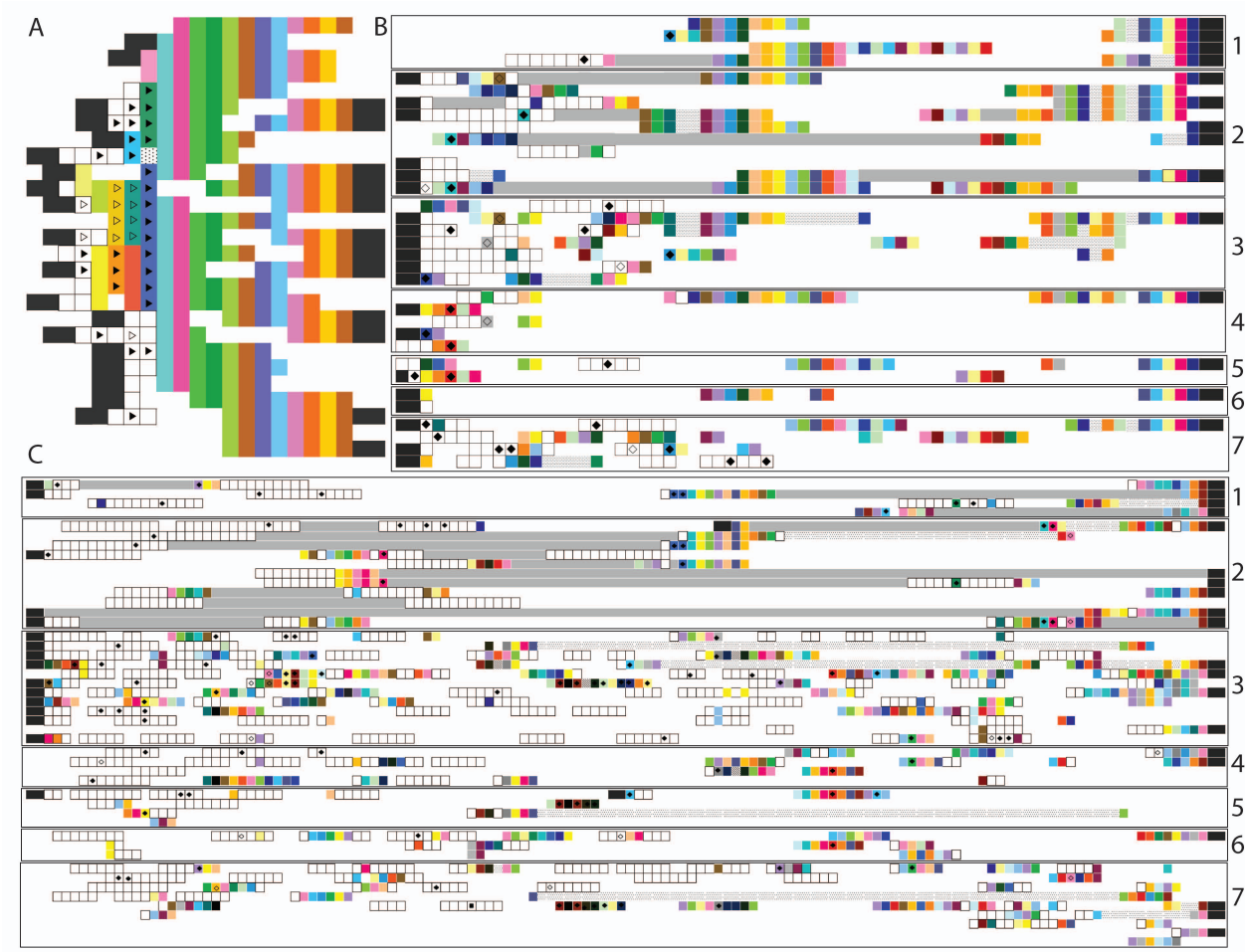
We thank H. Green for programming assistance and T. Cooke, R. Starfield, and L. Weinberger for stimulating discussions. ADW was partially supported by a US Department of Defense NDSEG Fellowship and the Biophysics Graduate Group. Funding was provided by the Department of Energy under contract numbers DE-FG02-07ER6450 and DE-FG02-05ER64134 and the Army Research Office Award # W911NF-10-1-0046.

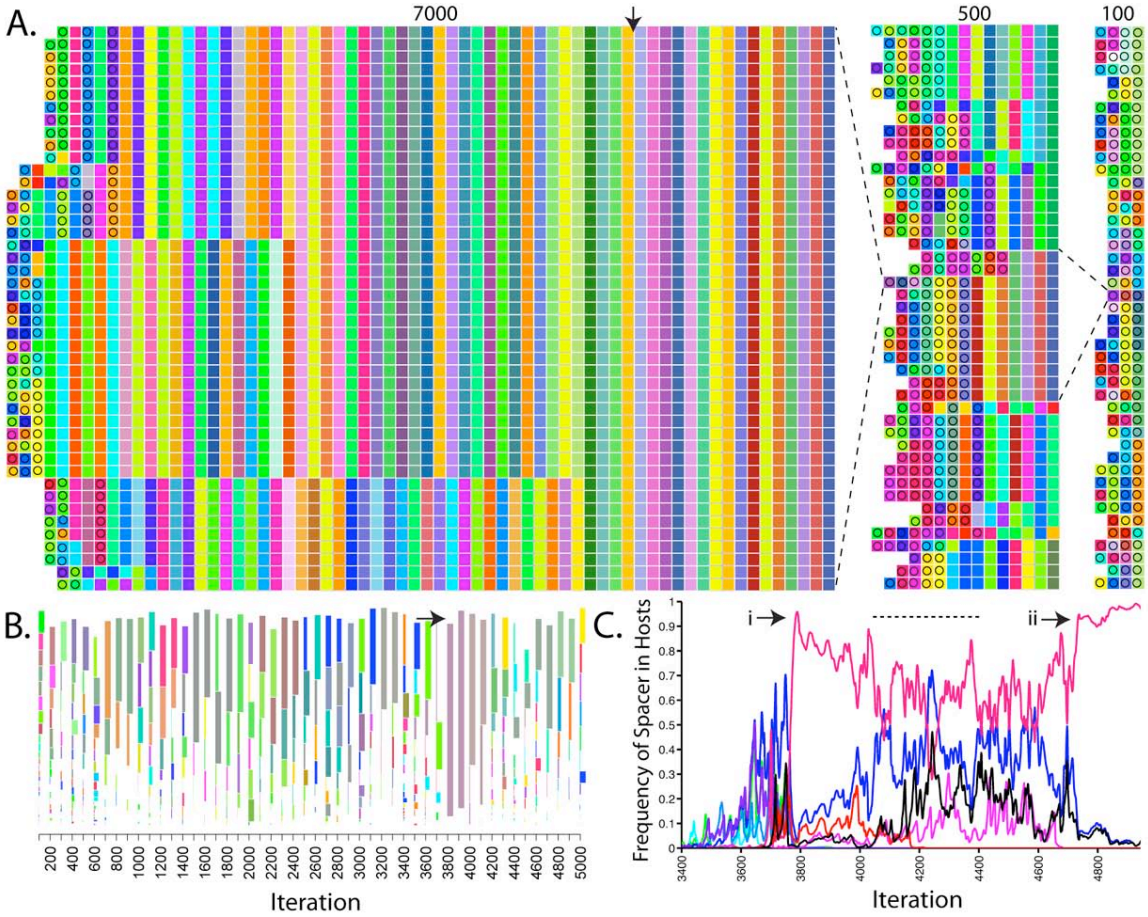
Figure Legends

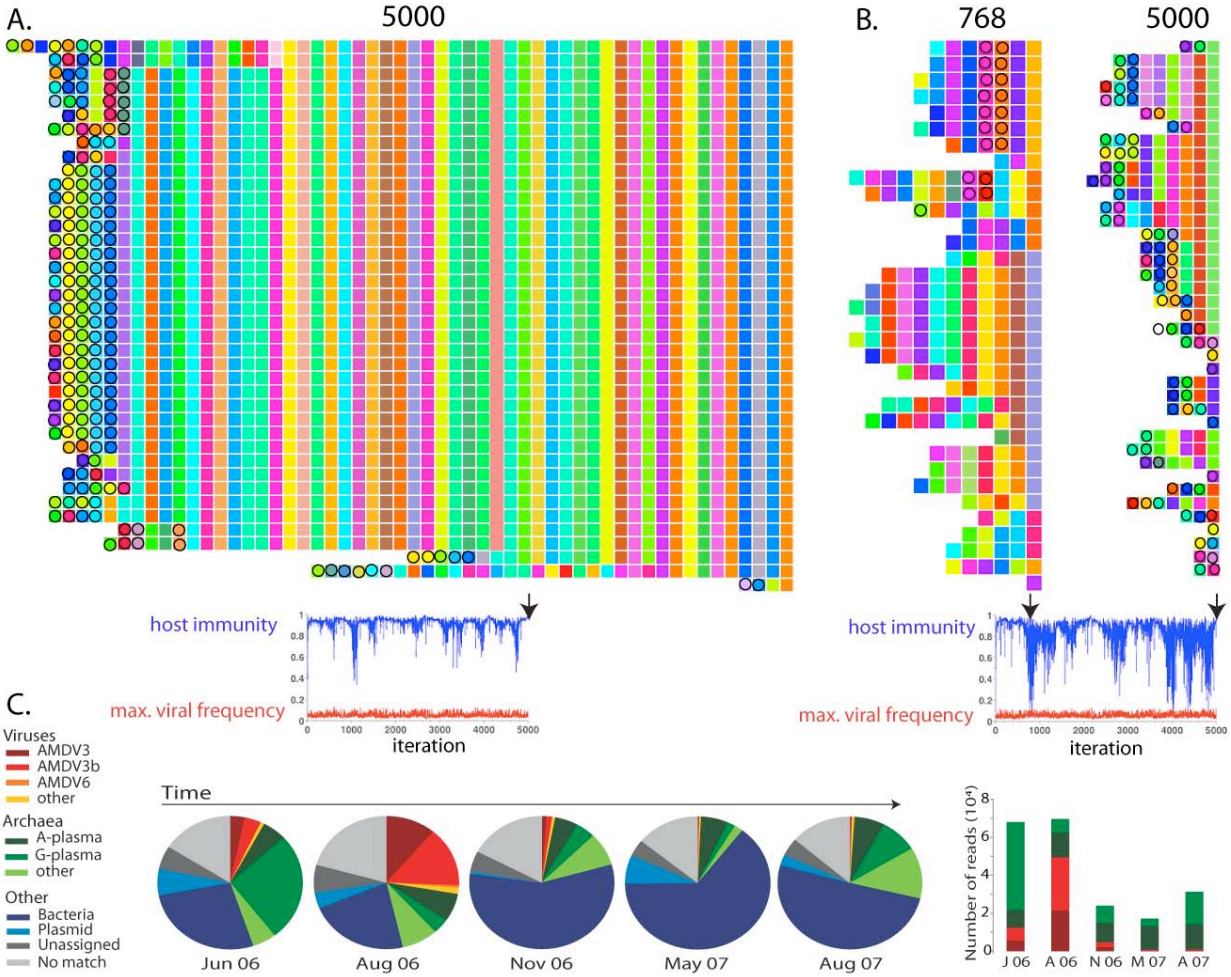
Fig. 1. Old-end clonality and conservation documented in archaeal CRISPR loci. Metagenomic reconstructions of CRISPR loci, with repeats removed, according to spacer (boxes) ordering in reads. Identical spacer sequences share the same color, excepting white boxes denoting unique spacers. Black shows flanking genome; white space is unsampled. **(A)** I-plasma CRISPR loci. Black triangles indicate spacers targeting reconstructed AMDV5 viruses (filled: perfect matches; open: imperfect matches). **(B)** and **(C)** CRISPR loci of a G-plasma population sampled in 2002 (1), 2005 (2), June 2006 (3), August 2006 (4) November 2006 (5), May 2007 (6) and August 2007 (7). Unsequenced DNA between paired reads is shown in grey. Overlapping 454 spacer patterns are condensed (SOM). Dotted boxes indicate probable spacer deletions. Diamonds indicate matches to reconstructed AMDV3b viruses (filled: perfect matches; open: imperfect matches).

Fig. 2. **Evolutionary dynamics driving old-end clonality in CRISPR loci.** (**A**) Computational reconstructions showing the 45 most frequent CRISPR loci at the 100th, 500th and 7000th iterations of a representative simulation (no spacer deletion). Circles indicate spacers perfectly matching any of the 300 most frequent viral strains in that iteration. To preserve space, 128 clonal columns were removed in iteration 7000 prior to the divergence of sub-populations (arrow). (**B**) Hosts are clustered into sub-populations (SOM), distinguished by color, every 100 iterations tracking diversity across time. Cluster heights represent the summed frequencies of strains, widths show the number of distinct strains, and the combined height of clusters in an iteration reflects the fraction of immune virus-host interactions. A marked loss of host diversity occurs prior to iteration 3800 (→), after which the sweeping sub-population diversifies (Fig. S3). (**C**) The frequencies of all spacers at one locus position tracked through the clustering-predicted sweep (i →). Oscillations (dashed line) are caused by viral mutations (Fig. S4) against the sweeping host sub-population before all competing hosts go extinct. A second sweep purges the remaining diversity at this locus position (ii →).

Fig. 3. Old-end conservation protects hosts against resurgent blooms of old viral sequences. The lower panels in (**A**) and (**B**) plot host immunity (blue) against maximum viral strain frequency (red) in each iteration. **(A)** 5% of additions occur with random spacer deletions, preserving old-end conservation and uniformity. Only new-end spacers target current viruses and host immunity is maintained. **(B)** 50% of spacer additions occur with deletions, purging old-end conservation and uniformity. Importantly, depletions in host immunity occur, indicating viral blooms. During the predicted bloom at iteration 768, two older spacers confer immunity against the top 300 viral strains. Host sub-populations lacking these two spacers survive by targeting less frequent viral sub-populations, revealing host and viral 'immunity clouds' (Fig. S8). (**C**) Community composition for the 2006–2007 time series samples via metagenomic sampling. A bloom of AMDV3b (red) occurs in August 2006 coincident with the depletion of its G-plasma host (bright green). Several preexisting old-end spacers target AMDV3b (black diamonds in Fig. 1C).


A
B
1
2
3
4
5
6
7
C
1
2
3
4
5
6
7


A.
7000
500
100
B.
Iteration
C.
i
ii
Frequency of Spacer in Hosts
Iteration


A.
5000
B.
768
5000
host immunity
max. viral frequency
iteration
host immunity
max. viral frequency
iteration
C.
Viruses
AMDV3
AMDV3b
AMDV6
other
Archaea
A-plasma
G-plasma
other
Other
Bacteria
Plasmid
Unassigned
No match
Time
Jun 06
Aug 06
Nov 06
May 07
Aug 07
Number of reads (10⁴)
J 06
A 06
N 06
M 07
A 07

References

Andersson, A.F., and Banfield, J.F., 2008. Virus population dynamics and acquired virus resistance in natural microbial communities. Science 320, 1047-50.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709-12.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J., 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321, 960-4.

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S., 2008. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J Bacteriol 190, 1390-400.

Diez-Villasenor, C., Almendros, C., Garcia-Martinez, J., and Mojica, F.J., 2010. Diversity of CRISPR loci in Escherichia coli. Microbiology 156, 1351-61.

Finzi, D., Blankson, J., Siliciano, J.D., Margolick, J.B., Chadwick, K., Pierson, T., Smith, K., Lisziewicz, J., Lori, F., Flexner, C., Quinn, T.C., Chaisson, R.E., Rosenberg, E., Walker, B., Gange, S., Gallant, J., and Siliciano, R.F., 1999. Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. Nat Med 5, 512-7.

Held, N.L., Herrera, A., Cadillo-Quiroz, H., and Whitaker, R.J., 2010. CRISPR Associated Diversity within a Population of Sulfolobus islandicus. PLoS ONE 5.

Horvath, P., and Barrangou, R., CRISPR/Cas, the immune system of bacteria and archaea. Science 327, 167-70.

Horvath, P., Romero, D.A., Coute-Monvoisin, A.C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barrangou, R., 2008. Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. J Bacteriol 190, 1401-12.

Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V., 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct 1, 7.

Marraffini, L.A., and Sontheimer, E.J., CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet 11, 181-190.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C., 2009. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology 155, 733-40.

Prangishvili, D., Forterre, P., and Garrett, R.A., 2006. Viruses of the Archaea: a unifying view. Nat Rev Microbiol 4, 837-48.

Rodriguez-Valera, F., Martin-Cuadrado, A.B., Rodriguez-Brito, B., Pasic, L., Thingstad, T.F., Rohwer, F., and Mira, A., 2009. Explaining microbial population genomics through phage predation. Nat Rev Microbiol 7, 828-36.

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53-65.

Sorokin, V.A., Gelfand, M.S., and Artamonova, II, Evolutionary dynamics of clustered irregularly interspaced short palindromic repeat systems in the ocean metagenome. Appl Environ Microbiol 76, 2136-44.

Thingstad, T.F., and Lignell, R., 1997. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. Aquat. Microb. Ecol. 13, 19-27.
Tyson, G.W., and Banfield, J.F., 2008. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ Microbiol 10, 200-7.
van der Oost, J., Jore, M.M., Westra, E.R., Lundgren, M., and Brouns, S.J., 2009. CRISPR-based adaptive and heritable immunity in prokaryotes. Trends Biochem Sci 34, 401-7.

Chapter Five

Future Directions: Viral Multi-Tropism as a Deadly Bet-Hedging Adaptation

Ariel D. Weinberger

The body of this dissertation focused on *how* viral diversity emerges in two diverse model systems, HIV and phage. In this short concluding chapter, I describe my current work aimed at understanding *why* viral diversity is so pervasive despite the likelihood of a diversified virus killing more host cells and thus accelerating host mortality. With infected hosts dying more quickly, hosts should have fewer opportunities to transfer the virus to uninfected susceptibles, which should reduce viral prevalence at the population scale. I am interested in finding the advantages that viral multi-tropism likely proffers to counteract its apparent cost in accelerating host mortality.

Previous work has suggested that multi-tropism evolved as a viral persistence mechanism in light of unpredictable target cell stochasticity (Weitz et al., 2005). Support for this idea comes from the prokaryote-phage system in which *Bordetella* bacteria are known to reversibly switch phenotypes to prevent the expression of a receptor bound by the bacteriophage BPP-1 (Doulatov et al., 2004; Liu et al., 2002). In response, the BPP-1 phage has evolved an elegant tropism diversification mechanism (*mtd*), which allows a small fraction of its progeny to infect the mutant bacteria, thereby preventing phage extinction (Medhekar and Miller, 2007).

Such a diversification of phenotypes in response to environmental stochasticity is a well-studied evolutionary strategy known as diversified bet-hedging (Cohen, 1966), but the advantage of a virus attacking new sets of target cells in a multicellular organism is unclear. A successful virus needs to avoid killing its host before sufficient transmission to new hosts has occurred. From Influenza to Picornaviruses to HIV, multi-tropism has been linked to increased virulence and accelerated death of the host organism (Basler and Aguilar, 2008; Nishimura et al., 2005; Penn et al., 1999; Weinberger et al., 2009; Whitton et al., 2005). Thus, the X4/R5 system in HIV provides a model system to analyze the benefits of deadly multi-tropism across a population of multicellular hosts.

To understand the selective benefit that X4 may offer a quasispecies of HIV viruses, it is instructive to compare populations in which X4 Virus is prevalent with other populations in which the strain is relatively rare. Despite emerging in almost 50% of late-stage Western HIV infections, X4 is rarely found outside of the Western subtype-B clade (e.g., subtype-C infections, which account for the majority of the world's HIV infections) (Cecilia et al., 2000; McCormack et al., 2002). Furthermore, X4 is extremely rare among all non-human primates infected with Simian Immunodeficiency Virus (SIV), from which HIV emerged zoonotically (Moore et al., 2004). One wonders why X4 Virus is only prevalent in subtype-B.

A potential differentiating factor between Western and non-Western populations is the disproportionate prevalence of the CCR5 receptor in African populations, potentially offering R5 a unique foothold in Africa. Recent studies demonstrate that $CCR5^+$ T cells are consistently high in HIV-negative residents of Uganda, including those of European (Italian) descent (Clerici et al., 2000). Conversely, CCR5 levels are not upregulated in HIV-negative Italian residents, including those of Ugandan descent. Thus, environment, rather than genetics, appears to play the critical role in determining CCR5 expression levels. Because CCR5 is only found on activated $CD4^+$ T cells (Moore et al., 2004), a fascinating conjecture is that sanitation and public health initiatives reduce CCR5 levels in the West by decreasing the level of environmental pathogens. Based on these studies and the work in previous chapters, my hypothesis is that X4 Virus may

provide HIV with a fitness advantage during early infection in environments where CCR5 levels are depressed. Endogenous stochastic fluctuations in these low CCR5 levels—driven, for example, by variation in β-chemokines that bind to CCR5 receptors—may favor quasispecies that include CXCR4 tropism (X4) as a way to lessen the likelihood of viral clearance in newly infected hosts. Thus, building on the work in Chapters 2 and 3, I am probing the hypothesis that multi-tropism offers viruses a bet-hedging advantage in environments driving severe within-host target cell fluctuations.

To test the hypothesis that multi-tropism improves HIV's invasability in noisy CCR5 environments, I focus on dual-tropic R5/X4 Virus (i.e., strains which can bind to *both* CCR5 and CXCR4) rather than X4 as an alternative to R5 Virus. These dual-tropic R5/X4 strains can infect $CXCR4^+$ target cells during CCR5-poor generations and thereby reduce the probability of viral extinction events in newly infected hosts. My decision to focus on R5/X4 rather than X4 during early infection is supported by the 102 viral genomes from early infection sequenced by the CHAVI consortium. No X4-only variants are found in that cohort but there are a number of R5/X4 viruses (Alan Perelson, email communication of data forthcoming for publication). Moreover, R5/X4 strains have been shown to be evolutionary intermediates between R5 and X4 and R5/X4 strains mutate into X4 throughout infection (Tasca et al., 2008).

Thus, I am applying a joint theoretical and computational approach to test the following two aims:

<u>Aim 1:</u> Construct a two-stage within-host model using measured, varying CCR5 levels to test if multi-tropism increases the median viral lifetime in individuals exposed to HIV.

The models derived in Chapters 2 and 3 assume that HIV has successfully invaded a new host. In order to test the hypothesis that multi-tropism improves fitness by increasing viral invasion, I clearly need to test whether invasion occurs. So, I am adding an individually-based stochastic model, which incorporates measured CCR5 fluctuations to test invasion. This model's output will then be input to the previously constructed deterministic model to look at questions such as how quickly hosts are killed if invasion occurs (more specifically, the time to 200 CD4s/ml—AIDS onset—which is proportional to the time to host death).

A stochastic model is necessary for early infection, because there are few infected cells in a new host, so any early-infection model needs to account for individual variation, or *demographic stochasticity* (Lloyd-Smith et al., 2005). However, when invasion occurs, the virus has access to innumerable target cells (e.g., in the gut and lymph nodes), which enables us to employ a "large-numbers," average-based deterministic model for the second stage of infection.

To model viral entry, I will code the stochastic analogue of my deterministic model using Gillespie's Algorithm (SSA), an efficient individual-based Monte Carlo procedure. I will modify the algorithm to also account for *environmental stochasticity* at the point of viral entry by varying the level of CCR5+ CD4+ HIV target cells. The viral load resulting from the early infection model—which is zero in the event of viral extinction—will serve as the initial viral load in the previously constructed deterministic framework. In running this model, I will also

vary the fractions of virions that are initially R5/X4 (as opposed to R5) from 0 to 1, and run at least 10,000 simulations for each R5/X4 fraction. In that way, I can quantify the median lifetime of an infection in a new host for different initial tropism fractions. My hypothesis is that some intermediate, nonzero fraction of multi-tropism (i.e., the "sweet-spot") increases the median infection lifetime relative to R5 alone.

<u>Aim 2:</u> Computationally test if minimizing fluctuations in CCR5 or increasing proviral latency in R5 Virus lead to R5/X4 clearance.

Using the Model developed in Aim 1, I intend to computationally test the hypothesis that the optimal fraction of R5/X4 variants in a population of viruses scales with the variance in CCR5 levels. I hope to get theoretical bounds on the minimum variance in CCR5 levels needed to support bet-hedging adaptations of R5/X4 Virus and to clear X4 Virus theoretically by going below this minimum. Moreover, if R5/X4 variants evolved to protect against rare stochastic depletions in depressed CCR5 levels, proviral latency would be an alternative viral persistence strategy that could maintain an R5-only virus (Razooky *et al., manuscript in preparation*). I will therefore simulate the model for different levels of proviral latency to see if increasing the level of latency clears R5/X4 and ultimately X4 Virus. The methods used will be similar to those described in Aim 1.

Implications

Approximately 1% of Europeans are homozygous for a 32 base-pair deletion in the gene for CCR5; these individuals have no obvious health defects and are almost entirely immune to HIV (Moore et al., 2004). This natural paradigm has led to the development of a number of promising anti-R5 therapies from bone marrow transplants (Hutter et al., 2009) to anti-CCR5 gene-therapies (Feng et al., 2000; Perez et al., 2008) and CCR5-blocker monotherapies (Moore et al., 2004). However, Chapter 2 showed that anti-R5 interventions may accelerate X4 emergence and immunodeficiency in patients who also harbor X4 Virus. Thus, it is critical to develop strategies to eliminate X4 from the subtype-B population, and this should be possible given that X4 appears unfit to survive in other viral clades.

At the conclusion of the introduction, I mentioned a phage therapy intervention aimed at selecting against bacterial strains able to acquire both pathogenicity and antibiotic resistance. I see that as a potential application of my CRISPR work. Similarly, one can think of applying an improved understanding of how X4 invades Western populations. Thus, a long-term, Darwinian Medicine goal for probing the mechanisms through which X4 has become prevalent in subtype-B is to predict interventions that might curb X4's fitness advantage and thereby drive it out of existence even in the West. This would open the door to a new class of potentially non-toxic treatments based on the paradigm of blocking CCR5. More generally, given the inefficacy of vaccines and drug interventions in the face of highly diverse pathogenic populations (Lipsitch and O'Hagan, 2007; Telford, 2008), elucidating mechanisms to reduce viral diversity may enable the application of well-studied techniques toward halting infectious disease spread.

References

Basler, C.F., and Aguilar, P.V., 2008. Progress in identifying virulence determinants of the 1918 H1N1 and the Southeast Asian H5N1 influenza A viruses. Antiviral Res 79, 166-78.

Cecilia, D., Kulkarni, S.S., Tripathy, S.P., Gangakhedkar, R.R., Paranjape, R.S., and Gadkari, D.A., 2000. Absence of coreceptor switch with disease progression in human immunodeficiency virus infections in India. Virology 271, 253-8.

Clerici, M., Butto, S., Lukwiya, M., Saresella, M., Declich, S., Trabattoni, D., Pastori, C., Piconi, S., Fracasso, C., Fabiani, M., Ferrante, P., Rizzardini, G., and Lopalco, L., 2000. Immune activation in africa is environmentally-driven and is associated with upregulation of CCR5. Italian-Ugandan AIDS Project. AIDS 14, 2083-92.

Cohen, D., 1966. Optimizing reproduction in a randomly varying environment. J Theor Biol 12, 119-29.

Doulatov, S., Hodes, A., Dai, L., Mandhana, N., Liu, M., Deora, R., Simons, R.W., Zimmerly, S., and Miller, J.F., 2004. Tropism switching in Bordetella bacteriophage defines a family of diversity-generating retroelements. Nature 431, 476-81.

Feng, Y., Leavitt, M., Tritz, R., Duarte, E., Kang, D., Mamounas, M., Gilles, P., Wong-Staal, F., Kennedy, S., Merson, J., Yu, M., and Barber, J.R., 2000. Inhibition of CCR5-dependent HIV-1 infection by hairpin ribozyme gene therapy against CC-chemokine receptor 5. Virology 276, 271-8.

Hutter, G., Nowak, D., Mossner, M., Ganepola, S., Mussig, A., Allers, K., Schneider, T., Hofmann, J., Kucherer, C., Blau, O., Blau, I.W., Hofmann, W.K., and Thiel, E., 2009. Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. N Engl J Med 360, 692-8.

Lipsitch, M., and O'Hagan, J.J., 2007. Patterns of antigenic diversity and the mechanisms that maintain them. J R Soc Interface 4, 787-802.

Liu, M., Deora, R., Doulatov, S.R., Gingery, M., Eiserling, F.A., Preston, A., Maskell, D.J., Simons, R.W., Cotter, P.A., Parkhill, J., and Miller, J.F., 2002. Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. Science 295, 2091-4.

Lloyd-Smith, J.O., Schreiber, S.J., Kopp, P.E., and Getz, W.M., 2005. Superspreading and the effect of individual variation on disease emergence. Nature 438, 355-9.

McCormack, G.P., Glynn, J.R., Crampin, A.C., Sibande, F., Mulawa, D., Bliss, L., Broadbent, P., Abarca, K., Ponnighaus, J.M., Fine, P.E., and Clewley, J.P., 2002. Early evolution of the human immunodeficiency virus type 1 subtype C epidemic in rural Malawi. J Virol 76, 12890-9.

Medhekar, B., and Miller, J.F., 2007. Diversity-generating retroelements. Curr Opin Microbiol 10, 388-95.

Moore, J.P., Kitchen, S.G., Pugach, P., and Zack, J.A., 2004. The CCR5 and CXCR4 coreceptors--central to understanding the transmission and pathogenesis of human immunodeficiency virus type 1 infection. AIDS Res Hum Retroviruses 20, 111-26.

Nishimura, Y., Brown, C.R., Mattapallil, J.J., Igarashi, T., Buckler-White, A., Lafont, B.A., Hirsch, V.M., Roederer, M., and Martin, M.A., 2005. Resting naive CD4+ T cells are massively infected and eliminated by X4-tropic simian-human immunodeficiency viruses in macaques. Proc Natl Acad Sci U S A 102, 8000-5.

Penn, M.L., Grivel, J.C., Schramm, B., Goldsmith, M.A., and Margolis, L., 1999. CXCR4 utilization is sufficient to trigger CD4+ T cell depletion in HIV-1-infected human lymphoid tissue. Proc Natl Acad Sci U S A 96, 663-8.

Perez, E.E., Wang, J., Miller, J.C., Jouvenot, Y., Kim, K.A., Liu, O., Wang, N., Lee, G., Bartsevich, V.V., Lee, Y.L., Guschin, D.Y., Rupniewski, I., Waite, A.J., Carpenito, C., Carroll, R.G., Orange, J.S., Urnov, F.D., Rebar, E.J., Ando, D., Gregory, P.D., Riley, J.L., Holmes, M.C., and June, C.H., 2008. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. Nat Biotechnol 26, 808-16.

Tasca, S., Ho, S.H., and Cheng-Mayer, C., 2008. R5X4 viruses are evolutionary, functional, and antigenic intermediates in the pathway of a simian-human immunodeficiency virus coreceptor switch. J Virol 82, 7089-99.

Telford, J.L., 2008. Bacterial genome variability and its impact on vaccine design. Cell Host Microbe 3, 408-16.

Weinberger, A.D., Perelson, A.S., Ribeiro, R.M., and Weinberger, L.S., 2009. Accelerated immunodeficiency by anti-CCR5 treatment in HIV infection. PLoS Comput Biol 5, e1000467.

Weitz, J.S., Hartman, H., and Levin, S.A., 2005. Coevolutionary arms races between bacteria and bacteriophage. Proc Natl Acad Sci U S A 102, 9535-40.

Whitton, J.L., Cornell, C.T., and Feuer, R., 2005. Host and virus determinants of picornavirus pathogenesis and tropism. Nat Rev Microbiol 3, 765-76.

Appendix

Supporting Material for Chapter 4: Persisting Low-Abundance Viruses Drive Microbial CRISPR-based Immunity

Ariel D. Weinberger, Christine L. Sun, Mateusz M. Pluciński, Vincent J. Denef, Brian C. Thomas, Philippe Horvath, Rodolphe Barrangou, Wayne M. Getz & Jillian F. Banfield

Mathematical Model of CRISPR Locus Evolution

Model Aims and Assumptions. Like any model of a complex natural system, the mathematical model of CRISPR-virus coevolution outlined in the main text necessitated simplifying assumptions. Here we describe these assumptions in detail, offering biological and mathematical support for each. We further amend the model to demonstrate that our conclusions are robust to the several simplifications applied.

In building the model, we aimed to computationally reconstruct the long-run coevolution of CRISPR loci and CRISPR-targeted viruses in natural, unculturable systems, filling in the gaps in our 'six-year' metagenomic fossil record. Our specific aim was to explain why CRISPR loci maintain specific immunities for thousands of microbial generations (immunological memory). Because CRISPR loci unidirectionally incorporate new viral spacers at only one end, a natural timeline of 'new' and 'old' ends exists within a CRISPR locus. Thus, we also sought to capture the dynamics by which new-end hyper-diversity is purged from CRISPR loci over time. We did so with a simple, frequency-based model in which hosts and viruses are divided into strains based on their ordered sets of spacers and proto-spacers, respectively. Hosts with fully identical spacers are grouped into one host strain and viruses with identical proto-spacers are grouped into one viral strain. Our selection of frequency based models follows the tradition of population genetics models in which populations are assumed to be very large and ecological fluctuations in population size are assumed to have no impact on processes influencing changes in frequencies of genotypes from one generation to the next. Thus the model we construct cannot address issues relating to stochastic walks of strains to extinction, but to include ecological (i.e. demographic considerations) requires information that is not explicitly available from metagenomic data. Of course, we can still infer that if a strain's frequency declines dramatically over several generations, then its numbers to must have declined too unless the whole population itself underwent explosive growth in the same time period.

As outlined in the main text and detailed more thoroughly in Section 3, the model is then divided into discrete, non-overlapping iterations. During each iteration, large numbers of host and viral strains interact, leading to stochastic mutations in host and viral strains and strong frequency-dependent selection for productively infecting viruses and immune hosts. At the conclusion of each iteration, the model takes a "snapshot" of the surviving host and viral genotypes. The evolution of these metagenomic snapshots was analyzed across thousands of iterations. We now justify the main model assumptions.

i) Virus and Host Populations Coexist at Large Abundances

Empirical support for the long-run coexistence of host and virus in natural microbial systems comes from two metagenomic studies in distinct environments. In the first study, Rodriguez-Brito et al., tracked consistently high loads of virus and host in four aquatic regions across a year-long period (Rodriguez-Brito et al.). Similarly, in our study, we reconstructed the relative abundances of archaeal CRISPR loci and viruses in an acid mine drainage system across the last two years of our six-year metagenomic time series experiment (Fig. 3c). In each

sampling, we recovered both host and viral genomes, implying that even if local extinctions occur, the populations are re-seeded from elsewhere in the system prior to subsequent sampling.

With host and virus continually extant in this system, one can wait long enough such that any preset number of interactions occurs in an iteration. We thus define a model iteration as the (variable-duration) period in which a parameterised number of interactions occurs. Generally 10^6 was utilised as the number of interactions per iteration, but we modified this number to probe model robustness. In truth, the number of interactions per iteration only affects the number of mutants that arise per iteration, it does not directly affect a strain's *relative* fitness and thus its frequency in the model. Our goal in prescribing 10^6 interactions per iteration is simply to assure a large number of viral mutants in each iteration. More precisely, Drake et al., note that the per-genome mutation rate of DNA viruses is approximately $3 \cdot 10^{-3}$ mutations per replication (Drake et al., 1998). That is, independent of genome length, 10^6 interactions allows viruses approximately $3 \cdot 10^3$ mutants per generation. Given that viral proto-spacers are defined as regions immediately preceding a specific 2-3 base-pair nucleotide motifs (PAMs), a back of the envelope estimate argues that $1/16^{th}$ to $1/64^{th}$ of viral genomes is proto-spacer material. So 10^6 interactions yields on the order of 100 proto-spacer mutations per iteration, as parameterised in the model. We used 10^6 interactions and thus produced approximately 100 viral proto-spacer mutations to assure that the majority of the 50 viral proto-spacer loci in our simulations had a chance to evolve. Again, stipulating 10^6 interactions is essentially a time-normalized way of representing the assumption of large virus and host abundances, meaning that in any short period of time there are manifold virus-host interactions and mutations. This approach is also robust from the point of view of population size. Large populations will have more interactions in a given time interval, but since we are not monitoring time directly, only numbers of interactions, our calculation of the number of mutations in each iteration of the modeling has been 'normalized' for population size.

By nondimensionalising the model in terms of time, such 'interaction-based' iterations no longer require parameterisations of unknown, system-dependent rates. For example, the average rate at which viral particles contact and infect host cells is poorly estimated even in the well-studied case of HIV in a human host; such a task is clearly extremely daunting in generally uncultivable microbes in diverse, open natural habitats. An interaction-number-based iteration also appropriately defines host and virus mutation 'rates' as the respective *probabilities* in which each virus-host interaction results in a host spacer addition or a viral proto-spacer mutation. Importantly, these probabilities are driven by the basic genetic architectures of host and phage and should thus be greatly system-independent, as is indicated by the relatively constant per-genome mutation rates of DNA microbes (Drake et al., 1998).

ii) Virus and Host Strains Interact via Exact Frequencies

We excluded the effects of random sampling in the interaction stage of the initial model for two reasons: a) strong selection has been experimentally demonstrated in CRISPR-virus interactions and b) empirical microbial systems consist of large numbers of host and virus. Importantly, while ignoring sampling noise in choosing the numbers of each host and viral strain that interact, we did consider a related 'demographic stochasticity' in the mutation stage of the

model, choosing the number of mutants of each host and viral parent strain from a Poisson Distribution.

The *Law of Large Numbers* guarantees that as a population grows large, the relative frequency at which a particular event is sampled approaches the true probability of the event in question, minimizing the impact of genetic drift. Given the consistently high number of viruses and hosts repeatedly sampled in the Rodriguez-Brito study (Rodriguez-Brito et al.), the total population sizes of viruses and hosts are kept large in our model. Thus, any host or viral strain with a non-negligible frequency has an abundance far from zero and is not susceptible to drift-driven extinctions. Hosts with low frequencies (less than 10^{-6}) are cleared from the model, with the exception of new mutants, which we now consider.

Among new mutants, one would ordinarily need to include drift because mutants should arise at frequency 1/N (where N is the respective host or viral population size at that time). With initially low frequencies prior to 'invading', new mutants are susceptible to drift-driven extinctions. In our model, however, new mutants are initialized at moderate frequencies based on their parent fitnesses, allowing us to ignore sampling noise. In essence, by the time we consider a new mutant to exist, we have taken into account the fact that a much larger proportion of mutants arose but never made it through one infection cycle. Further, as we explain below in the section on mutant frequency initialization, we do this because of the extremely strong selection measured in CRISPR-virus interactions (Barrangou et al., 2007; Deveau et al., 2008). That is, population genetics shows that the probability of a new mutant avoiding extinction due to drift (i.e., 'establishing') in a large population, is solely a function of its selection advantage (Barrett et al., 2006).

iii) Virus and Host Strains Are Well-Mixed (mass-action)

The model assumes that host or viral strains are distributed uniformly in space, even though, in reality, strain specific spatially-distinct reservoirs are likely to exist.

Our model ignores spatial heterogeneity for two reasons. First, we built a first-pass general model for natural systems whose ecologies are unknown and whose topographies can vary. Thus, we have no *a priori* knowledge of the system-specific spatial heterogeneities to incoporate into the model. Further, a previous step allowed us to wait for an extremely large number of interactions to occur, reducing the effects of spatial noise in sampling viruses and hosts. Second, spatial heterogeneity in fact strengthens the major result of this paper—i.e., that old-end spacers are conserved against rare resurgences of low-abundance viruses—because spatial heterogeneity (reservoirs) is a primary mechanism for viral persistence (Rainey and Travisano, 1998; Wilson, 1990).

iv) No Back-Mutation When Proto-spacers Mutate (Infinite Alleles Model)

With more than 30 base-pairs per proto-spacer, and only a single base-pair mutation sufficient in many cases to allow for viral evasion of CRISPR, individual proto-spacers posses innumerable degrees of freedom through which to mutate around CRISPR-based immunity. This is especially so given that single base-pair mutations in either the proto-spacer or flanking proto-spacer adjacent motif (PAM) region are in some cases sufficient to prevent CRISPR-mediated

immunity (Barrangou et al., 2007; Deveau et al., 2008; Marraffini and Sontheimer). With so many mutational adaptations available, including many synonymous mutations, the probability of a strain mutating back to an antecedent becomes negligible. Thus, as in Kimura's well-known Infinite Alleles Model (Kimura and Crow, 1964), we assume that each viral mutant arises from a bin of limitless allele types, choosing a unique, unseen proto-spacer during each viral mutation.

v) Hosts Add and Delete Spacers, but Do Not Mutate Spacers

While host CRISPR loci add and delete spacers in the model, for two reasons we do not consider the mutation of base-pairs within a host spacer. First, selection acts against mutations of host spacers because they weaken spacer/proto-spacer binding and thus antiviral immunity (and by the above infinite alleles argument, hosts are very unlikely to randomly make the same mutations as the viral strains). Even if host spacer mutations were selectively neutral in not reducing CRISPR-spacer binding (e.g., because the Gibbs Free Energy increase is neglible from single base-pair mutations), on average it would take these neutral mutations N generations to drift to fixation, where N is the population size (Barrett et al., 2006). In microbial systems, N can be on the order of 10^8.

The second and more significant reason for ignoring host spacer mutation is that host genomes (10^6-10^7 base-pairs) are orders of magnitude larger than viral genomes (10^3-10^5 base-pairs). Yet, larger genomes do not accumulate mutations any quicker than shorter genomes, as Drake et al., has noted (Drake et al., 1998). The *per-genome* mutation rate of a DNA-based microbe—virus or host—is effectively constant empirically at approximately .003 mutations per replication (Drake et al., 1998). Thus, the probability of mutating any spacer within a CRISPR locus is two to three orders of magnitude lower than the probability of mutating any proto-spacer within a virus.

vi) Host Spacer Deletions Only Occur During Additions

As noted in the main text, the model assumption that spacer deletion only occurs when a strain adds spacers is consistent with experimental data. Indeed, the current hypothesized mechanism by which CRISPR loci delete spacers is homologous recombination (Horvath and Barrangou; Marraffini and Sontheimer), which should result in spacers being added and deleted from a strain concomitantly. To that end, spacer deletions are often observed on strands exhibiting recent spacer additions (Deveau et al., 2008). Further linking the genomic insertion and deletion processes of microbes, Palmer and Gilmore captured an example in which the CRISPR locus of an Enterococcus line is deleted during the incorporation of an adjacent genomic island (Palmer and Gilmore). Coupled insertions and deletions may be a key mechanism by which prokaryotes keep their genome lengths below an empirical threshold of approximately 13 MB, despite continually updating their genomes with new horizontally transferred regions (Kuo and Ochman, 2009). Finally, a coupled addition and deletion process may also explain why deleted spacers are frequently empirically observed despite small deletions not providing hosts much in the way of selection benefits. Population genetic theory predicts that such deleterious mutations should be lost in large microbial populations.

vii) Mutants Initialized at Fraction of Parent Strain Frequencies

From a modeling perspective, the advantage of initializing new host and viral mutant at a moderate fraction (parameterized at 10%) of their parent strains' frequencies is to prevent the fittest variants from immediately being lost due to model clearance of low-frequency strains. We are thus only looking at the subset of beneficial mutants that initially avoid genetic drift and converge to moderate, parent-based frequencies in their first generation. As we noted in ignoring sampling noise in the interaction step of the model, the justification for doing so is threefold: a) the experimentally demonstrated strength of selection for immune hosts and infective viruses in CRISPR-virus interactions and b) the extremely large microbial and phage popaulation sizes and c) the parameterized large number of interactions per model iteration.

Population genetics predicts that the probability of invasion of a new mutant in large populations is monotonically increasing as a function of that mutant's selection coefficient (i.e., its fitness advantage relative to the other strains of that iteration). Specifically, working from the diffusion equation, Kimura approximated the probability, P, of a new mutant avoiding extinction as $P = (1-e^{-2s})/(1-e^{-2sN})$, where N is the population size and s the selection coefficient of the mutant(Kimura, 1964). Hence, with N large as in our microbe-phage system and $s>>1/N$, the probability of a new mutant 'establishing' (i.e., avoiding extinction) is $P = 1-e^{-2s}$. A mutant's establishment probability is thus an increasing function of its fitness advantage in our system (note that this is sometimes estimated in the literature as $P\approx 2s$, an assumption that holds under weak to moderate selection (Barrett et al., 2006)). Because CRISPR-virus systems have been measured to be under extremely strong selection (Barrangou et al., 2007; Deveau et al., 2008), the above equation predicts that beneficial host and viral mutants are likely to be established due to selection dwarfing drift. For simplicity we allow all mutants to establish, because mutant establishment only allows the new mutants to compete in the clonal interference step of the model. If these mutants do not offer increased fitness (for example, viruses which mutate proto-spacers not incorporated by any hosts have no selection advantage), they will be quickly be outcompeted by fitter clones and cleared.

After establishment, the relevant question is what the establishment probability should be. That is, each iteration contains many interactions through which mutants will grow to new frequencies based on their varying fitnesses. To be clear, a host's fitness is the fraction of viruses that the host strain is immune to, while a virus' fitness is the fraction of hosts that the virus can productively infect. To simplify model calculations, we rely on the fact that new mutants inherit most of their parental strains' fitness, having only changed a single module of the fitness-determining CRISPR or proto-spacer regions. So if the parent strain has a high relative fitness, its mutants should as well (in fact, excepting deletions, by construction mutants are at least as fit as their parents). We thus assumed that the mutants initially inherited their fitness from their parents. Given that mutants arise at $1/N$ and that they do so at an unknown point in the midst of long iterations, we initialized mutants at only a fraction of their parents' frequency (generally 10%, but varied to ensure model robustness).

viii) Frequency Thresholds Below Which Strains Are Cleared, Excepting New Mutants in 'Emergence-Periods'.

These coupled assumptions of a minimum strain frequency and a mutant emergence period more rapidly reproduce long-run patterns by accelerating the mutational turnover of host

and viral strains. Specifically, the frequency threshold speeds the clearance of 'older' strains, which are not being selected for and are thus depressed in frequency. To prevent fit new mutants from premature clearance before selection has the chance to increase their frequencies above the threshold, we used an 'emergence period' during which new mutants are not subject to model clearance even when low in frequency. We define the emergence period to be the number of iterations it takes a new mutant that survives its first round of replication to produce a 'burst size level' of progeny. Fit mutants with burst size progeny are then strongly selected for to grow to a represented frequency within the model. The emergence period thus allows new mutants a chance to increase in frequency due to selection.

To account for the demographic stochasticity that low frequency mutants face, we employed a randomly distributed (Poisson) emergence period, choosing individual values of clearance-free iterations for each mutant. To assure that the emergence period alone did not bias our long-run results by excessively down-regulating new mutant clearance due to drift, we also ran the model without a emergence period. Removing the emergence period—without changing any other model parameter—does not change long-run model dynamics (Supplementary Fig. 9).

Thus, an implicit model assumption is that while virus and host population sizes remain large throughout a simulation, the particular virus and host strains making up these large populations turnover rapidly. This assumption is directly supported by the conclusions of the above-mentioned metagenomic tracking experiment in four aquatic environments. That study found high-level mutational turnover at the 'strain' (genotypic) level of virus and host coupled with stability at the 'species' (i.e., total abundance) levels (Rodriguez-Brito et al.).

Importantly, this is not the reason that sweeps are observed in model simulations. The particular simulation shown in Figure 2c has an average of eight host mutants added per generation, each of which (generally) incorporates a distinct spacer. The existence of distinct, beneficial mutant lines results in the classical clonal interference ('kill the winner') patterns observed in Figure 2c and predicted in the literature (Gerrish and Lenski, 1998). Sweeps occur from what is known as the multiple mutation effect (Desai and Fisher, 2007; Desai et al., 2007), events in which a single host line makes a succession of beneficial mutations and gains immunity to the vast majority of viral strains at that time point (Supplementary Figure 4).

Alternative Model Formulations

Neutral, Large Population Model Does Not Reproduce Empirical Patterns

Our population-genetic model uses strong selection(Barrett et al., 2006) to reproduce the empirical spacer patterns found in the long-run metagenomic data. The question is whether these empirical patterns are also consistent with a null (neutral) model in which selection plays not role in CRISPR locus evolution. Drift alone—if given long enough—will indeed drive a single allele to fixation in a previously polymorphic locus. Is that why old-end clonality is observed in empirical CRISPR loci, which have presumably evolved for millions of generations?

Laboratory and natural systems data showing an extremely high strength of selection in CRISPR loci indicate otherwise. Tyson and Banfield found that hosts with genotypically identical non-CRISPR regions incorporated distinct new-end spacers, showing that the CRISPR

locus is evolving at a far higher rate than other genomic regions, a fact unlikely to occur under neutral evolution (Tyson and Banfield, 2008). Further, as noted above, strong selection for immune hosts and infective viruses was also measured in laboratory CRISPR asays (Barrangou et al., 2007; Deveau et al., 2008).

For completeness, we sought to test how CRISPR loci would evolve without selection. That is, we amended the model by stipulating that spacers provide no immunity, but still maintained host spacer addition and viral proto-spacer mutation at the same frequencies as before. The wild-type CRISPR locus did not show any added spacers after thousands of iterations, consistent with the fact that a new mutant requires N generations to fix by drift alone. Given that microbial communities have large values of N, drift is unsurprisingly mitigated in its effects.

Other Model Formulations Less Applicable to Metagenomic Data

While we employed a frequency-based population-genetic model with discrete, non-overlapping generations, one envisions alternative modeling paradigms that combine ecologically dependent virus-host population dynamics with random host and viral mutation. Such 'eco-evolutionary' models have been pursued in well-studied systems such as HIV-T cell dynamics in human hosts (Althaus and De Boer, 2008); here we describe why metagenomic, natural systems data are not yet ripe for such fine-scale modeling.

In order to completely include population dynamics in a coevolutionary model, one needs to understand both the ecological dynamics of the environment being simulated (e.g., how frequently nutrients are being pumped into and out of a system) and the kinetics through which virus and host populations increase, decrease and interact (e.g., how well such nutrients are utilized by the microbes). And even if one had measurements of the extrinsic, environmental variables, his results would almost certainly be difficult to generalize to new, unstudied environments. The bigger problem, however, is that one presently cannot measure the relevant rates of virus-host growth and interaction, because naturally occurring microbes are generally unculturable in the laboratory. In fact, a major utility of metagenomics is that it allows us to get snapshots of otherwise unobservable microbes (Schleper, 2008). But these snapshots only offer relative, frequency-based information about virus-host dynamics, given that metagenomics is a sampling-based technique (i.e., a genomic scoop is taken from an environment). As in our model, one can metagenomically infer a viral bloom but never metagenomically quantify how many viruses actually bloomed. That is why we employed a frequency-based framework.

That said, one could simplify the ecological dynamics immensely and assume host and virus abundances change according to simple functions (e.g., logistically to keep populations bounded) and interact in traditional 'Lotka-Volterra' fashion albeit with mutation. While one would still not know the rate at which viruses interact with hosts, one could non-dimensionalise to remove unknown time dependencies. This is essentially what we did in our system, although we openly discretised the system as any numerical solver would do anyway. Our approach also combines many mutations into a single iteration, essentially a non-dimensionalised variation of the Tau-leaping technique used to accelerate the Doob-Gillespie algorithm (Gillespie, 2007). We could not have employed a non-dimensionalised variant of Gillespie's individual-interaction

algorithm. The reason is the sheer size of the microbial system and the number of interactions one needs to simulate (one by one) to generate long-run patterns. Gillespie's algorithm has a complexity that scales exponentially with the number of particles in the system, making it virtually unusable for co-evolutionary simulations with millions if not billions of microbes.

Model Algorithm

Unique Spacers Represented as Unique Positive Integers. The model represents unique spacers with unique positive integers starting from 1. Whenever a virus mutates a proto-spacer, the resulting proto-spacer mutation is given the value of the *first* positive integer that has yet to be used in the model run.

Host 'Strains' Distinguished by Their Ordered Spacer Sets. Host genotypes are divided into 'strains' based on the set of ordered spacers in their CRISPR loci: all hosts with fully identical spacers are placed in the same strain. Ordering is incorporated into the model, because spacer addition is unidirectional. Incidentally, one may be able to reliably differentiate among host strains by only considering the strains' spacers, because the CRISPR locus has been shown to be among the fastest mutating genomic regions in Bacteria and Archaea (Tyson and Banfield, 2008). Otherwise clonal populations have distinct spacers.

Viral 'Strains' Distinguished by Their Ordered Proto-spacer Sets. Each viral strain is analogously distinguished by the ordered list of proto-spacers in its genome. This is again a reasonable way to distinguish viral genotypes, given the increased mutational pressure on proto-spacer elements, which determine viral fitness against a CRISPR-laden host. For simplicity, we assume that all viral strains contain a fixed, tunable number (S) of proto-spacers. While the order of proto-spacers in a particular virion is not thought to play a significant role in infectivity, we maintain the initial ordering of proto-spacers for consistency of representation of strain types. Thus, proto-spacer mutations are placed in the slots that their parent proto-spacers occupied. For example, if the first mutant in the simulation mutates proto-spacer i ($1 \leq i \leq S$), the viral mutant is stored as the strain $(1,\ldots,i\text{-}1, S\text{+}1, i\text{+}1,\ldots.S)$, which like all viral genotypes in the simulation has length S (conversely, host genotyopes have a varying length depending on how many spacers have been added).

Immune vs. Productive Interactions. The model defines an immune virus-host interaction as one in which the host survives and the virion is cleared, and a productive interaction as one in which the host dies with a large, parameterized number of virions 'bursting' from the cell. In that sense, the interactions in an iteration can be disjointly partitioned into immune and productive subsets. An interaction is either immune or productive, but never both. We tracked the fraction of immune interactions across iterations in Figure 3, giving us a metric of the average fitness of CRISPR-mediated immunity in distinct parameter (spacer deletion) regimes.

Parameters: Table S1 Gives a Summary of Commonly Used Values

k = Number of interactions per iteration.
S = Number of proto-spacers in viral strains.
P_{V_mut} = Expected fraction of virus-host interactions in which viruses mutate a *random* proto-spacer.
P_{B_add} = Expected fraction of virus-host interactions in which hosts add a spacer (the particular spacer added is chosen according to the frequency of proto-spacers in the viral population).
P_{B_lose} = Expected fraction of host spacer additions in which the host deletes a randomly-sized contiguous block of spacers from a random starting position in its CRISPR locus.

$f(n)$ = Fraction of virus-host interactions which is productive, when virus and host share $n \geq 0$ spacers.

The function f is monotonically decreasing with n: the more spacers shared, the more likely the host is to be immune. This is consistent with both laboratory data and theory. Experiments in the *Streptococcus Thermophilus* CRISPR model system have shown that when host and viral strain share two spacers rather just one, host immunity is increased by two orders of magnitude (Deveau et al., 2008). We incorporated Deveau et al.'s data in parameterising f. Importantly, one can also predict the additional immunogenicity that extra matching spacers provide from the theory of stochastic processes. To provide immunity, a spacer must be transcribed and cleaved into CRISPR RNA after which it must contact and bind its corresponding viral proto-spacer *before* the virus integrates its DNA into the host. More shared spacers means more opportunities to find corresponding viral proto-spacers, reducing the average time until the *first* such binding occurs (i.e., the mean waiting time until the first event of this Poisson Process is inversely proportional to the number of shared spacers).

B_{prod} = Number of virions 'bursting' from a productively-infected host cell.
i_v = Fraction of the parent strain's frequency that each *viral* mutant strain is initialized with. This is done because mutants are at least as fit as their parent strains by construction in the model.
i_B = Fraction of the parent strain's frequency that each *host* mutant strain is initialized with.
g = Average (of Poisson distributed) emergence period given to each new host and viral mutant strain. During their emergence periods, mutant strains are not subject to clearance.
V_{min_freq} = Minimum frequency threshold below which any viral strain no longer in its emergence iterations is cleared.
B_{min_freq} = Minimum frequency threshold below which any host strain no longer in its emergence iterations is cleared.
V_{list_max} = Maximum number of viral strains beyond their initial new mutant 'emergence-periods.' The lowest frequency, non-emergence period strains beyond this threshold are cleared.
B_{list_max} = Maximum number of host strains beyond their initial new mutant 'emergence-periods.'

Initialization

The system is generally initialized with one non-immune host strain having no spacers and one viral strain possessing spacers 1 through S. This initial condition was also relaxed, with no substantive effect on the long-run dynamics. For example, we started the simulation shown in Fig. 2 of the main text with two immune host lines: one host strain started with the spacer 1 and the other possessed the spacer 2.

Model Iteration: Virus-Host Interactions Leading to Mutation-Selection

Step 1: Host and Viral Strains Interact According to Well-Mixing. In each iteration, the model assumes that host and viral strains are well-mixed, distributing the k interactions among host-virus 'strain-pairs' according to the products of host and viral strain frequencies. Thus, viruses preferentially interact with more frequent host lines and, similarly, host strains disproportionately face frequent viral strains. The number of interactions among different strain

pairs is stored in an **Interactions** matrix, whose $(i,j)^{th}$ entry is the number of interactions between the ith viral strain and the jth host strain in the current iteration.

Mathematically, for each viral strain ($\mathbf{V}_i$) and each host strain ($\mathbf{B}_j$), the number of interactions between $\mathbf{V}_i$ and $\mathbf{B}_j$ in the tth iteration is:

$$[\mathbf{Interactions}(t)]_{ij} = k \bullet (\mathrm{Freq}(\mathbf{V}_i(t\text{-}1))) \bullet (\mathrm{Freq}(\mathbf{B}_j(\mathrm{t}\text{-}1)))$$

Note that $\mathrm{Freq}(\mathbf{V}_i(t\text{-}1))$ and $\mathrm{Freq}(\mathbf{B}_j(t\text{-}1))$ are the frequencies of $\mathbf{V}_i$ and $\mathbf{B}_j$ at the end of the previous iteration. When t=1, the initializing frequencies are used.

Step 2: New Strain Frequencies Determined by Fitness in Last Iteration. For each $(\mathbf{V}_i, \mathbf{B}_j)$ in a given iteration, the number of spacers shared by the virus-host strain pair is counted. That number of shared spacers is placed in the $(i,j)^{th}$ entry of the **Immunity** matrix at iteration **t**.

f is the parameterized function deciding what fraction of strain-pair interactions is productive given the number of shared spacers. The more spacers a virus-host strain pair share, the more likely the host is to be immune. Using the number of interactions assigned to a strain-pair and the number of shared spacers between the two strains, we define the number of productive interactions for each $(\mathbf{V}_i, \mathbf{B}_j)$ strain-pair in the tth iteration to be:

$$[\mathbf{Productive\ Interactions}(t)]_{ij} = [f([\mathbf{Immunity}(t)]_{ij})] \bullet [\mathbf{Interactions}(t)]_{ij}$$

Because immune interactions are defined to be those interactions that are not productive, the corresponding matrix of **Immune Interactions** during iteration t is:

$$\mathbf{Immune\ Interactions}(t) = \mathbf{Interactions}(t) - \mathbf{Productive\ Interactions}(t)$$

A viral strain's frequency in the *current* iteration depends on its ability to have productively infected hosts during the *previous* generation. Conversely, a host strain's frequency in the *current* generation depends on its ability to have survived viral infection during the *previous* generation. Given the definition of the **Interactions** matrix based on viral and host frequencies in the previous iteration, the model determines initial viral and host strain frequencies for the current iteration (the prime symbol ' denotes the fact that these frequencies will be renormalized as mutants arise in subsequent stages of the iteration and old strains go extinct):

$$Freq(V_i(t')) = \frac{\sum_j [\mathbf{Productive\ Interactions(t)}]_{i,j}}{\sum_{i,j} [\mathbf{Productive\ Interactions(t)}]_{i,j}}$$

$$Freq(B_j(t')) = \frac{\sum_i [\mathbf{Immune\ Interactions(t)}]_{i,j}}{\sum_{i,j} [\mathbf{Immune\ Interactions(t)}]_{i,j}}$$

Step 3: Unidirectional Host Spacer Addition. Each host strain $\mathbf{B_j}$ gives rise to $\mathbf{N_{Bj}(t)}$ mutant strains, which each unidirectionally incorporate random viral proto-spacers during the interactions. We calculate the number of mutant progeny lines for each $\mathbf{B_j}$ according to the formula:

$$N_{Bj}(t) = Poisson\left(P_{B\_add} \bullet k \bullet Freq(B_j(t))\right)$$

This equation makes host spacer addition more likely to occur in more frequent host strains. Because host strain frequencies are determined by host immune profiles, the model assumes that immune strains are more likely to add spacers than non-immune strains. This is consistent with experimental data showing that non-immune hosts have little opportunity to add spacers before productive viral infection overwhelms the cell (Barrangou et al., 2007). Immune strains, however, should be able to incorporate new viral proto-spacers after CRISPR and its associated protein machinery (Cas proteins) successfully prevent infection.

For strains in which $\mathbf{N_{Bj}(t)}$>=1, the model creates all $\mathbf{N_{Bj}(t)}$ mutant strains independently. For each of these mutants, the particular spacer added is chosen randomly according to the distribution of viral proto-spacers in that iteration. The model chooses directly from proto-spacer distributions, because the well-mixed interaction hypothesis ensures that the relative frequencies of proto-spacers in the population match the relative frequencies of the proto-spacers in the viral strains that each host strain interacts with. Finally, the chosen spacers are unidirectionally inserted into the leftmost position of each mutant's locus, so each mutant of $\mathbf{B_j}$ has the ordered form (new spacer, $\mathbf{B_j}$).

The new mutant strain originating from $\mathbf{B_j}$ is given the initial frequency i_B • Freq($\mathbf{B_j(t)}$). The model makes a mutant strain's initial frequency a fraction of its parent's value, because mutant and parent are almost identical in spacer sets and thus in immune profiles, fitness. The large number of interactions per iteration allows the mutant sufficient interactions to approach a sizable fraction of its parent's frequency. Finally, each new mutant strain is given an individualized clearance-free random emergence period of Poisson(g) iterations.

Step 3b: Random Spacer-Adding Mutants Lose Blocks of Spacers. The old-ends of CRISPR loci become clonal over time in the absence of spacer deletions, yet we know that spacer deletions must occur to prevent unchecked locus growth. Moreover, most empirically observed loci have regions where deletion events are evidenced by missing spacers and missing blocks of spacers relative to the spacer patterns in other cells. We sought to understand whether old-end spacers are retained to prevent compromises of host immunity. To do so, in some model runs (e.g., those shown in Fig. 3 of the main text), we allowed an expected fraction of spacer adding host mutants to delete spacers during the addition process. In model runs that include deletions, a random number is chosen between 0 and 1 for each mutant strain that adds a new spacer in the previous step. If that number is below P_{B_lose}, the strain loses a randomly-sized contiguous block of spacers from a random starting point in its locus. The new strain is then initialized in the same way all mutant strains were in the previous step.

Step 3c: Renormalization of Strain Frequencies. Whether or not deletions occur, the frequencies of all host strains are renormalized to sum to 1. Consistent with our above well-mixing

assumption, a population-wide renormalization assumes that all resident strains suffer equally from the emergence of new mutants.

Step 4: Mutation of Viral Proto-spacers. For each viral strain $\mathbf{V_i}$, the resulting number $\mathbf{N_{Vi}(t)}$ of mutant strains that each mutate a proto-spacer during the interactions of the tth iteration is:

$$N_{Vi}(t) = Poisson\left(P_{V\_mut} \bullet k \bullet Freq(V_i(t))\right)$$

As with the hosts, this formula makes viral proto-spacer mutations most likely to occur in the viral strains of highest frequency. Because the frequency of a viral strain is a function of the number of productive infections it causes, this formula effectively makes viral proto-spacer mutation most likely to occur in productively infecting virions. The apparent quickness of action of CRISPR immunity makes such a hypothesis plausible. Moreover, when a virus productively infects a host there are far more opportunities for copying errors, mutations.

If $\mathbf{N_{Vi}(t)}$>=1, we create the $\mathbf{N_{Vi}(t)}$ mutant lines of $\mathbf{V_i}$ independently. For each of these strains, the particular proto-spacer mutated is chosen randomly from the set of proto-spacers possessed by $\mathbf{V}_i$. The new proto-spacer is represented with the first positive integer yet to be used in the model run and placed in the slot occupied by its parent proto-spacer. Thus, the first viral mutant strain created has the form (1,…,i-1, S+1, i+1,…,S), where i is the particular proto-spacer mutated.

Analogous to mutant initialization in the hosts, each viral mutant strain originating from $\mathbf{V_i}$ is given the initial frequency i_V • Freq($\mathbf{V_i(t)}$) and an individualized clearance-free random emergence period of Poisson(g) iterations.

Finally, the frequencies of all viral strains are renormalized to sum to 1, consistent with the well-mixing assumption.

Step 5: Clear Lowest Frequency Strains and Take Metagenomic Snapshots. In natural systems, unbounded numbers of distinct host and viral strains cannot exist. There are 'fit' genotypes and those likely to suffer stochastic extinction when their frequencies become sufficiently low. Given that unfit strains cannot persist indefinitely and the computing necessity of keeping the lists of active viral and host strains manageable, the model clears the least frequent viral or host strains, excepting new mutants in their emergence iterations as justified above.

To clear non-emergence period, low frequency strains, the model extracts all host and viral strains that exceed their initial clearance-free emergence iterations. This is done separately in both viral and host populations. The model then lists all non-emergence period strains in descending frequency order in both virus and host populations. Any host strain with a frequency below B_{min_freq} and any viral strain with a frequency below V_{min_freq} is cleared from its respective population. To keep the number of strains in the simulation bounded and to account for limited resources in nature, the model subsequently clears the lowest frequency non-emergence period host and viral strains beyond the B_{list_max} and V_{list_max} thresholds, respectively.

For both host and viral populations, the model collects the surviving strains, including "emergence-period" strains, and renormalizes frequencies so that the strain frequencies sum to 1

in host and virus populations, respectively. For each surviving $\mathbf{B_j}$, the frequency of $\mathbf{B_j}$ at the end of the *t-th* iteration is defined to be Freq($\mathbf{B_j(t)}$). Similarly, for each surviving $\mathbf{V_i}$, the frequency of $\mathbf{V_i}$ at the end of the *t-th* iteration is set to Freq($\mathbf{V_i(t)}$).

Finally the model takes a metagenomic snapshot of all surviving host and viral strains, capturing both spacer content and relative strain frequencies. In all but the final model iteration, the simulation returns to step 1.

Clustering Analysis

To quantify how host diversity changes across time, computationally generated loci were clustered into an optimal number of sub-populations at frequent iterations. The partitioned sub-populations and their progeny were then tracked across time. Results show how diversity is abruptly lost during selective sweeps (see the 3800th iteration in Fig. 2B), after which diversity is regained coincident with new-end diversifications within the sweeping sub-population (see the final thousand iterations in Fig. 2B and fig. S3). The regained host diversity is captured because the clustering algorithm compares host strains at all locus positions in determining an optimal number of sub-populations for an iteration. This is in contrast to the spacer dynamics plots (e.g., Fig. 2C and fig. S2), which show diversity at a single locus position across time. That locus position will eventually become clonal in the absence of deletions.

Determining an Optimal Number of Clusters. The clustering algorithm begins by determining the number of clusters in the first snapshot (T=100). To do so, the model counts the number of distinct spacers at each locus position (column) of the aligned host strains in the snapshot. Because the number of viral proto-spacers is large—there are approximately 900 proto-spacers at steady state when the viral genomes are set to have 50 proto-spacers—and hosts sample relatively infrequently (1-10 times per iteration) from this list, the algorithm assumes that two different host strains will not incorporate the same spacer. Thus, when two hosts share the same spacer at a given locus position, the algorithm assigns them to the same lineage and assumes all previous locus positions have the same spacers as well.

For each locus column in the snapshot, an independent clustering diagram is generated by simply dividing the hosts into sub-populations based on their spacer in that column. In other words, if column 2 only contains the spacers 3 and 4, all hosts with the spacer 3 in column 2 are placed into one cluster and all hosts with the spacer 4 in column 2 are placed into the other cluster. Using the Hamming distance, which calculates the proportion of shared elements (spacers) in two aligned strains, we were able to determine an average silhouette width(Rousseeuw, 1987) for each clustering diagram (i.e., for each column). The silhouette is a cluster validation technique, which gives a value of how well-clustered a population is. By maximizing this metric, the column that optimally partitions the first snapshot's host strains by lineage is chosen. Subsequent snapshots are clustered in the same way, with the exception that only newer columns (i.e., to the left of previous clustering columns) are compared for a maximal silhouette width. With the column upon which clustering is calculated always to the left of clustering columns from previous iterations, we are able to determine parent clusters for each cluster by checking their spacer in the previous iteration's clustering column.

Representing Clusters Across Time. After making all clusters at each snapshot, the algorithm colors from the final snapshot backwards. All distinct clusters with no progeny are assigned distinct random colors (to maximize the contrast between distinct clusters). Parent clusters are then given the average color of their progeny clusters. In each iteration, the clusters are then displayed. Each cluster's height reflects the summed frequency of all strains within it and the clutser's width reflects the total number of strains in that cluster. The combined height of all clusters in an iteration represents the fraction of virus-host interactions that is immune.

Benefits of Population Genetic Model

Recently, He and Deem(He and Deem) utilized a well-known HIV differential equation model(Nowak and May) to model spacer diversity within CRISPR loci of pre-stipulated lengths. Their eukaryotic construct relies on the assumption that CRISPR-immunized Bacteria and Archaea hunt and kill viruses just as cytotoxic $CD8^+$ T cells target HIV virions. In the absence of prokaryotic killer cells, viruses are assumed to grow exponentially (i.e., as would dividing particles). However, this would lead to the troubling prediction that virus population levels decrease when host population levels uniformly increase (i.e., when all host strains increase by the same factor, which may roughly occur after an influx of resources). Our metagenomic tracking the viral bloom from an initially large host population does not support this prediction (Fig. 3C). Further, while He and Deem's model generates the reduced old-end diversity well-documented in natural and laboratory CRISPR loci, reduced old-end diversity does not reflect a unique prediction of their model. It emerges across all stochastic models in which ancestral variants cannot mutate. By noise alone, old lines stochastically go extinct as time progresses, and with no mechanism to re-insert diverse spacers at old positions, old-end diversity decreases. The relevant question is thus not whether experimentally documented old-end clonality occurs, but i) how it occurs (e.g., selection vs. drift) and ii) why old-end spacers are at all preserved given a known proclivity of prokaryotes to disproportionately delete, over insert, small genetic elements(Kuo and Ochman, 2009). Capturing the coevolutionary dynamics driving old-end clonality, our results track the dynamics through which hosts lose diversity, capturing both gradual declines in diversity (fig. S2) and rapid selective sweeps (Fig. 2C). By allowing an equilibrium locus length to emerge from the ratio of addition to loss parameters (i.e., avoiding pre-stipulated lengths), our model explains observed metagenomic patterns of old-end clonality, showing how CRISPR old-ends provide immunity against blooms of ancestral viral sequences. An emergent CRISPR locus length enables follow-up studies linking the optimal length of a CRISPR locus to the level of viral persistence within a microbial community.

Evidence in Favor of CRISPR-based Immunity in G-plasma

The measured AMDV3b viral bloom (Fig. 3C) occurs despite the presence of preexisting spacers in host loci targeting AMDV3b long before the bloom (see black diamonds in Fig. 1C). Given the failure of these spacers to prevent the viral bloom, one might conjecture that the reconstructed CRISPR loci of G-Plasma are non-functional. We simulated the evolution of CRISPR loci under the assumption that spacers provide no antiviral immunity. Results show that when CRISPR loci evolve by predominantly neutral, drift-driven processes (i.e., spacers confer no immune, selective advantage), emergent loci contain a minimal number of spacers

after thousands of iterations, with no old-end uniformity. In fact, the dominant host strain to emerge is often spacer-less. In contrast, sampled G-plasma loci contain tens of spacers, exhibit the dichotomous patterns of old-end uniformity and new-end diversity, and show sequential additions of new-end spacers (Figs. 1B and 1C), allowing us to infer that G-Plasma CRISPR loci are active and under strong selection, providing antiviral immunity.

Repeat Sequences for Metagenomic Data

I-plasma locus in Fig. 1A repeat sequence: GTATCAATTCCTTATAGGGACGATTTATAG

G-plasma locus in Fig. 1B repeat sequence: ATTTCAGAAAAACTAGTTAGTATGGAAG

G-plasma locus in Fig. 1C repeat sequence: GTTAGAATCTTATTTAGAAAGTTTCAAAG

Metagenomic Methods

Sample collection. For the 2006 – 2007 time series study, biofilms were sampled from the acid mine drainage solution – air interface at the C +75 m location in the Richmond Mine (Iron Mountain, CA - 40° 40′ 38.42″ N and 122″ 31′ 19.90″ W (Elevation ~ 3,100′)) in June, August, and November 2006, as well as May and August 2007. Environmental parameters of this site at the times of sampling have been reported previously (Denef et al., 2010). Samples were transferred to dry ice on site and stored at -80 °C.

DNA extraction, preparation and sequencing of metagenomic libraries. For each biofilm from the C +75 m location, high molecular weight DNA was extracted from a 1 g subsample using procedures described previously (Lo et al., 2007). Preparation of shotgun metagenomic libraries and pyrosequencing using the 454 Genome Sequencer FLX-Titanium system were performed at the W. M. Keck Center for Comparative and Functional Genomics (University of Illinois, Urbana-Champaign, IL) according to manufacturer's instructions (454 Life Sciences, Branford, CT) (Margulies et al., 2005). Signal processing and base calling were performed using the bundled 454 Data Analysis Software version 2.0.00.

Metagenomic data analyses. Sequencing reads from the five libraries were co-assembled using Newbler (GSassembler v. 2.0.01, Roche) using default parameters except for a 95% nucleotide identity and 40 nt minimum overlap requirement. Replicated reads were identified using a previously described protocol based on CD-HIT clustering (Gomez-Alvarez et al., 2009) (> 95 % identity, > five identical bases at the start of the read, no equal length requirement). Within each CD-HIT cluster, reads that shared the same start position on the assembled contigs were identified and removed except for the longest read. Additional filtering of reads containing ambiguous bases, resulted in a total of 990,386 reads (~350 Mbp). A second assembly, using identical parameters, was performed using this filtered reads dataset.

For community profiling, read assignment to previously identified genomic sequence bins was performed by blastn analysis (e-value cutoff of e^{-20}) using a database of contigs previously assembled and binned from four other Richmond Mine biofilm samples: 5-way, collected in March 2002 (Simmons et al., 2008; Tyson et al., 2004), UBA and UBA filtrate collected in June 2005 (Baker et al., 2010; Lo et al., 2007), and UBA-BS collected in November 2005 (Dick et al., 2009).

Contigs representing virus genome fragments were identified based on (a) similarity to previously identified virus contigs recovered from the same system, (b) extreme high depth of sequence coverage (in the case of AMDV3b), (c) assembly curation into genome fragments with detectable sequence similarity to the known viruses, and (in all cases) (d) targeting of the genome sequence fragments by CRISPR spacers. Viruses were determined to replicate in specific hosts based on extensive targeting of their genomes by spacers from host-specific CRISPR loci. Curation of contigs containing reads identified as viral was carried out using Consed (Gordon et al., 1998). Contigs were then imported into GSMapper and extended manually and joined, where appropriate, so that regions fragmented by elevated sequence divergence could be condensed. Cases of extreme divergence were treated as separate contigs. Locations where genomic datasets were fragmented by gene content differences were noted, and

the information used as part of the binning procedure. Viral genomes related to the previously studied AMD viruses but that assembled separately were distinguished. For example, the deeply sampled AMDV3b genome is related to the previously reported AMDV3 population and also to a shallowly sampled AMDV3c population (results not shown) that is also present in the C +75m dataset.

Strainer (Eppley et al., 2007) was used to visualize single nucleotide polymorphism patterns and other forms of variation. This made use of the ".ace" file generated by GSMapper and read re-mapping step that corrects for homopolymer errors during import into Strainer.

Processing of sequencing reads for CRISPR analysis. CRISPR spacer analysis was performed on individual sequencing reads rather than contigs generated from an automated assembly. Sanger reads (mate-paired ~ 800 bp sequences from each end of an ~ 3 kb clone) from the 5-way, UBA, UBA-BS, and UBA filtrate datasets and 454 reads from the C +75 m series were used in the reconstruction of both G-plasma CRISPR loci (data are separated by time points in Fig. 1). Any 454 reads containing at least one ambiguous base ("N") were removed. Using a custom ruby script, the ends of each 454 read were trimmed until a base passed 15/20 NQS (neighborhood quality standard) (Altshuler et al., 2000), with a variation described in (Brockman et al., 2008). Cross_match (developed by P. Green, University of Washington) was used to remove any remaining B adaptor sequences (from library construction). Phred (Ewing and Green, 1998; Ewing et al., 1998) was used to trim the Sanger sequencing reads and Cross_match was used to filter vector sequence.

CRISPR data analysis. Sequencing reads that sampled the CRISPR loci were identified based on the presence of specific repeat sequences (see below). Custom ruby scripts were used to extract CRISPR spacer sequences from 454 and Sanger sequencing reads. We allowed for variation in the repeat sequences to avoid omitting spacer sequences due to errors in sequencing (e.g., homopolymer runs). Spacers were grouped using blastclust (using parameters of 85% identity and 85% length overlap) to remove duplication of groups due to sequencing error. Custom ruby scripts were used to array CRISPR spacers back onto sequencing reads. Assembly of each locus was manually performed in Microsoft Excel based on overlapping spacer patterns and sampling of the flanking genome on part of the read or its mate pair (in the case of Sanger reads). Where possible, 454 reads were arrayed so that patterns of sequential spacers matched locus regions defined based on Sanger reads. For data presentation in Figs. 1B and 1C, unique patterns defined by multiple overlapping 454 reads were condensed to report the longest possible sequence of spacers.

Detection of spacer matches. Spacer matches were detected using blastn, with parameters for short sequences (G = 2, E =1, F = F). Perfect matches signify exact matches (100% identity across entire length of spacer) while imperfect matches require at least 85% identity across at least 85% of the spacer. The databases used in the blast searches were composed of AMDV3b sequences recovered in this study. While the database used to detect imperfect matches only contained contig sequences, the database used to detect perfect matches also included the individual sequencing reads that comprised each of the contigs.

Analysis of community composition in C +75 m time series data. For each individual sample, each read was assigned to a sequence bin (organism or virus type) based on blastn analysis (cutoff < e^{-20}). The unassigned category indicates similarity to contigs in the AMD sequence database with unknown affiliation. Note that, as described previously (Denef et al., 2010), changes in solution pH occurred at the sampling site over the time period studied. This altered the overall community composition, particularly the relative abundances of Bacteria and Archaea.

Modeling

The mathematical model of CRISPR-driven virus-host coevolution was programmed and simulated in the MATLAB programming environment (version 7.7). Images of CRISPR loci (i.e., spacer patterns) were then produced in R (version 2.11), coloring loci generated during MATLAB simulations.

MATLAB simulations were programmed to record all spacers in all CRISPR loci at all model iterations. These population-wide metagenomic recordings are stored as numerical matrices, in which each row contains the spacers of a distinct host strain. To distinguish among spacers, distinct spacers are stored as distinct numbers. Importing these matrices into R, we mapped each unique spacer (i.e., number) to a unique color producing colored spacer patterns. The 'cluster' package in R was used to track the evolution and diversity of host subpopulations across time.

MATLAB Code Used To Generate Simulations

```
%% Run initial parameters
clear all
run parameters

%% Program


for t=1:T

   disp(t)


   %Step 1: Interactions Occur

   if t==1
   [vspacer1,bspacer1]=frequencycnt(Vmati,Bmati,vps,bmax);
   end

   % Immuneweight is a matrix whose entries are
   % the numbers of shared host and viral spacers
   [immuneweight]=isimmune(Bmati, Vmati,vps, bmax);

   % Prodinteracts is matrix that represents the number of productive
   % interactions (based on immunity data).
   [interacts,
prodinteracts]=prodinteractsfun(num_interacts,Bmati(:,2),Vmati(:,2),immuneweight);

   immuneinteracts = interacts-prodinteracts;

   % The number of interactions and productive interactions are saved for
   % future analysis.
   num_prod_interacts(t)=sum(sum(prodinteracts));
   num_interacts1=sum(sum(interacts));

   %Calculate the fraction of immune interactions
   %in this generation
   immunepercent(t)=(num_interacts1-num_prod_interacts(t))/num_interacts1;

   %Adjust Numbers of Bacteria Based on
   %Interactions
   Btemps1= sum(immuneinteracts,1)';
   B_noninteracts=max(Bmati(:,2).*B - sum(interacts,1)', 0);
   Btemps=Btemps1;
   B=sum(Btemps);
```

```
Bmati(:,2)=Btemps/B;
%New Bacterial Population Size
Btotal(t) = B;


if Btotal(t)<=0
   disp('All Bacteria are Dead')
   disp(['The Remaining Number of Viruses is: '...
      num2str(V)])
   break
end

%Adjust Numbers of Viruses Based on
%Interactions
Vtemps=sum(prodinteracts,2);
Vmati(:, 2) = Vtemps/sum(Vtemps);
V=burst*sum(Vtemps);
%New Viral Population Size
Vtotal(t) = V;

if V<=0
   disp('All Virions are Dead')
   disp(['The Remaining Number of Bacteria is: '...
      num2str(B)])
   break
end

%End of Step 2

% Step 3: Determine is each bacteria strain gains, loses, or remains
% unchanged.  If the strain mutates, then pick
% a spacer to add.


if finddaughters=='Y'
   mut_factor=1;
   adds=poissrnd(Padd*num_interacts*(mut_factor.*Bmati(:,2))');
else
   adds=poissrnd(Padd*num_interacts*Bmati(:,2)');

end

addbtotal(t)=sum(adds);

% If there are bacterial mutations, the code proceeds.
if sum(adds)~=0
```

```
%Preallocate extra rows for all bacterial
%adds
next_row=size(Bmati,1);
Bmati = [Bmati' zeros(sum(adds), size(Bmati,2))']';
match=0;

for i=find(adds)

   bactold=Bmati(i,:);

   %Set Frequency of mutants, change
   %parents' frequency accordingly
   mutant_freq=bactold(2)*init_freqb;
   %           Bmati(i,2)=max(Bmati(i,2)-(2e-4)*adds(i),0);
   %mutant_freq=bactold(2)*(init_freqb/adds(i));
   %Bmati(i,2)=bactold(2)*(1-init_freqb);

   for j = 1:adds(i)
      % bact is the information for the bacterial strain that will be
      % adding a spacer.

      % The bacterial strain that was chosen to mutate, adds a
      % spacer.

      if t==1
         bact=bactmutate(bactold,vspacer1,bmax);
      else
         bact=bactmutate(bactold,vspacermatrix(:,t-1),bmax);
      end

      if match==0
         next_row=next_row+1;
      end


      % If the finddaughters protocol is activated, the new bacterium
      % will be assigned an identity to
      % match its parent strain.
      if finddaughters=='Y'
         if isodd(bact(1))
            bcnt=bcnt1;
         else
            bcnt=bcnt2;
         end
```

```
            [Bmati,bcnt,whofromtemp, match]=addbacttypedaughters...
               (Bmati, bact,bmax, bcnt,mutant_freq, next_row);

            if isodd(bact(1))
               bcnt1=bcnt;
            else
               bcnt2=bcnt;
            end
            % A family tree that explains where the new mutant came
            % from is recorded.
            whofromb=[whofromb' [whofromtemp t]']';
         else
            [Bmati,bcnt,whofromtemp, match]=addbacttype...
               (Bmati, bact,bmax, bcnt,mutant_freq, next_row);
            whofromb=[whofromb' [whofromtemp t]']';
         end
      end
   end
end

Bmati(:,2)=Bmati(:,2)/sum(Bmati(:,2));

%  Step 4: Mutate Viruses

mutates=poissrnd(Pmut*num_interacts*Vmati(:,2)');
addvtotal(t)=sum(mutates);

% If there are any viral mutations, the code proceeds.
if sum(mutates)~=0

   mutant_block = zeros(sum(mutates), size(Vmati,2));

   %Set Columns 1,3,4 of mutant block
   mutant_block(:,1)=vcnt+1:vcnt+sum(mutates);
   vcnt = vcnt+sum(mutates);
   %mutant_block(:,2)=init_freqv;%burst/V;
   mutant_block(:,3)=zeros(sum(mutates),1);
   mutant_block(:,end)=poissrnd(2,sum(mutates),1);
   %mutant_block(:,4)=poissrnd(3,sum(mutates),1);
   %mutant_block(:,4)=nbinrnd(1,1/3,sum(mutates),1);

   %We randomly selects mutated spacer index for each
   %mutant
   spacerindices = (0:sum(mutates)-1)'*size(Vmati,2)+4+ceil(S_0*rand(sum(mutates),1));
```

```
    %Set remaining columns of mutant_block
    %before mutations
    index=1;
    j=1;
    for i=find(mutates)

        mutant_block(index:index+mutates(i)-1,2)= Vmati(i,2)*init_freqv;
        mutant_block(index:index+mutates(i)-1,4)= Vmati(i,1);
        mutant_block(index:index+mutates(i)-1,5:end-2)=ones(mutates(i),1)*Vmati(i,5:end-2);
        mutant_block(index:index+mutates(i)-1,end-1)=ones(mutates(i),1)*Vmati(i,end-1);
%1.02*rand(mutates(i),1)
        whofromv=[whofromv' [mutant_block(index:index+mutates(i)-1,1)
ones(mutates(i),1)*[Vmati(i,1) t]]']';
        index = index+mutates(i);
        j=j+1;
    end

    mut_counter=mut_counter+1;
    mutant_block=mutant_block';
    spacerslost=mutant_block(spacerindices);
    mutant_block(spacerindices)=vps(2)+mut_counter:vps(2)+mut_counter+sum(mutates)-1;
    mut_counter=mut_counter+sum(mutates)-1;
    Vmati = [Vmati' mutant_block]';

    spacertree=[spacertree' [spacerslost mutant_block(spacerindices)
t*ones(numel(spacerslost),1)]']';

  end

  Vmati(:,2)=Vmati(:,2)/sum(Vmati(:,2));

  % End of Step 4

  %Step 5: Clearance

  %Clear Old Bacteria Less Than 1e-6
  Bmati=Bmati(((Bmati(:,3) > Bmati(:,end) & Bmati(:,2)>=1e-6) ...
    | (Bmati(:,3) <= Bmati(:,end) & Bmati(:,2) >=0)), :);


  if size(Bmati,1) > 500
    Bmatold = Bmati(Bmati(:,3) > Bmati(:,end), :);
    [bfreqs, bfreqsindex]=sort(Bmatold(:,2),'descend');
```

```
    top300b = bfreqsindex(1:min(size(Bmatold,1),300));
    Bmati=[Bmatold(top300b,:); Bmati(Bmati(:,3) <= Bmati(:,end), :)];
end

%Renormalize
Bmati(:, 2) = Bmati(:,2)/sum(Bmati(:,2));



%Clear Old Viruses Less Than 1e-6
Vmati=Vmati(((Vmati(:,3) > Vmati(:,end) & Vmati(:,2)>=1e-6) ...
    | (Vmati(:,3) <= Vmati(:,end) & Vmati(:,2) >= 0)), :);


%Clear Old, Unfit Viruses
if size (Vmati,1) > 500
    Vmatold = Vmati(Vmati(:,3) > Vmati(:,end), :);
    [vfreqs, vfreqsindex]=sort(Vmatold(:,2),'descend');
    top300v = vfreqsindex(1:min(size(Vmatold,1),300));
    Vmati=[Vmatold(top300v,:); Vmati(Vmati(:,3) <= Vmati(:,end), :)];
end

%Renormalize
Vmati(:, 2) = Vmati(:,2)/sum(Vmati(:,2));



% Step 5: Diversity Indices of Iteration


%Find all spacer frequencies in the bacterial and viral
% populations

% Frequencycnt gets spacer frequencies.

[vspacermatrix(:,t), bspacermatrix(:,t)]=...
    frequencycnt(Vmati,Bmati,vps,bmax);

% Time is incremented in Bmat and Vmat
Bmati=Bmatobject(Bmati,bps,'timestep');
Vmati=Vmatobject(Vmati,vps,'timestep');
% End of Step 5

%Check Bact, Virus Diversities

%Sort Bmati, Vmati in descending order
```

```
    [bfreqs, bfreqsindex]=sort(Bmati(:,2),'descend');
    Bmati=Bmati(bfreqsindex,:);
    [vfreqs, vfreqsindex]=sort(Vmati(:,2),'descend');
    Vmati=Vmati(vfreqsindex,:);

    if immunepercent(t) < 0.8
        bottle_ind=bottle_ind+1;
        bottle_count(bottle_ind,:)=[bottle_ind t];
        Bmat_bottle{bottle_ind} = Bmati;
        Vmat_bottle{bottle_ind} = Vmati;
    end
    %

    if mod(t, 50)==0
        Bmatsave{t/50} = Bmati;
        Vmatsave{t/50} = Vmati;
    end

end



% Initial Values
B=1E8;
V=1E9;
num_interacts = B;

% Number of Timesteps
T=5000;


Padd=5*B^-1;
S_0=50;
Pmut=150*B^-1;

%% Internal Parameters

% Maximum number of Mutants over Simulation
S=round(Pmut*B*T)+1000;
% Maximum Number of Bacteria Spacers
bmax=700;
% Maximum Number of Gained/Lost Virus Spacers
vmax=S_0;
```

```
%Counters for Bacterial, Viral Population over
%time
Btotal = zeros(1,T+1);
Vtotal = zeros(1,T+1);
B0= B;
V0= V;
%tuner=100;
%counter = zeros(1,T+1);
%counter(1) = .999;

% Total number of mutations in bacteria
addbtotal=zeros(1,T);
lossbtotal=zeros(1,T);
addvtotal=zeros(1,T);
mut_counter=0;
burst=200;
num_protos = zeros(1,T);

% Matrices of parameters
vps=[vmax; S_0; S];
bps=bmax;
%% Initial Bacteria Structure

% Bacteria Matrix
bcnt=1;
Bmati=[bcnt 1 0 0 zeros(1, bmax) poissrnd(0)];
Bmatsave = cell(T/50,1);
Bmat_bottle = cell(T,1);
bottle_ind=0;
bottle_count=zeros(T,2);

%% Initial Virus Structure

% Virus Matrix
vcnt=1;
Vmati=[vcnt 1 0 0 1:S_0 1 poissrnd(0)];
Vmatsave = cell(T/50,1);
Vmat_bottle = cell(T,1);

%% Creating Save Matrices
num_prod_interacts=zeros(1,T);
virulence=zeros(1,T);
deadB=zeros(1,T);
deadV=zeros(1,T);
life_max_V=zeros(1,T);
life_max_B=zeros(1,T);
```

```
max_freqb=zeros(1,T);
max_freqv=zeros(1,T);
DivB_shannon=zeros(1,T);
DivV_shannon=zeros(1,T);
S_B=zeros(1,T);
S_V=zeros(1,T);

whofromb=[];
whofromv=[];
spacertree=[];
finddaughters='N';

vspacermatrix=sparse(S_0+S,T);
bspacermatrix=sparse(S_0+S,T);

topvstrains=zeros(3,T);
topbstrains=zeros(3,T);

immunepercent=zeros(1,T);

init_freqv=1e-1;
init_freqb=1e-1;
 function [vspacermatrix_temp, bspacermatrix_temp]=frequencycnt(Vmat,Bmat,vps,bmax)
% This function calculates several types of spacer frequencies for later
% use.


% Vhas is a row vector whose entires are all of the spacers that have
% been added by any virus.
%Vmat=Vmat(Vmat(:,2)>1e-3,:);
vhas=nonzeros(unique(Vmat(:,5:end-2)))';
nhas=length(vhas);
vspacermatrix_temp=sparse(vhas,ones(nhas,1),ones(nhas,1),vps(3)+vps(2),1, nhas);

% The frequency of all viruses that have gained
% each spacer i is summed.
for i=vhas
   hascount=sum(Vmat(:,5:end-2)==i,2);
   vspacermatrix_temp(i)=sum(Vmat(:,2).*hascount);
end

% Bhas is a list of all of the spacers that have been gained by any
% memeber of the bacterial population.
%Bmat=Bmat(Bmat(:,2)>1e-2,:);
bhas=nonzeros(unique(Bmatobject(Bmat,bmax,'bhas')))';
bhas=bhas(bhas>0);
```

```
nhas=length(bhas);
bspacermatrix_temp=sparse(bhas,ones(nhas,1),ones(nhas,1),vps(3)+vps(2),1,
nhas);x=rand(5000);

for i=bhas
   hascount=sum(Bmatobject(Bmat,bmax,'bhas')==i,2);
   bspacermatrix_temp(i)=sum(Bmat(:,2).*hascount);
end

end



function [interacts, prodinteracts]=prodinteractsfun(num_interacts, Bprobs, Vprobs,
immuneweight)

   interacts=num_interacts*(Vprobs*(Bprobs'));

   % number of successful infections in a unimmune host
   Plive = 1e-9;
   prodinteracts=(1-Plive)*interacts.*(immuneweight==0);

   % The number of successful interactions that occur in bacteria with
   %  immunity is calculated below.

  %One new spacer
  index=round(immuneweight)==1;
  prodinteracts(index)=interacts(index).*(1e-5) ;

   %Two New Spacers
   index=round(immuneweight)==2;
   prodinteracts(index)=interacts(index).*(1e-7);

   %Three new spacers
   index=round(immuneweight)==3;
   prodinteracts(index)=interacts(index).*(1e-8);

  %Four new spacers
   index=round(immuneweight)==4;
   prodinteracts(index)=interacts(index).*(1e-9);

   %Five New Spacers
   index=round(immuneweight)==5;
   prodinteracts(index)=interacts(index).*(1e-10);

   %Six New Spacers
```

```
    index=round(immuneweight)==6;
    prodinteracts(index)=interacts(index).*(1e-11);

    %Seven or More New Spacers
    index=round(immuneweight)>=7;
    prodinteracts(index)=interacts(index).*(1e-12);

end




function [immunecnt,immunemat]=isimmune(Bmat, Vmat, vps, bmax)

    % Isimmune creates a matrix immunecnt with the number of
    % number of spacers shared by the  virus and bacteria.
.

    % v and b are the numbers of virus types and bacteria types, respectively.
    v=size(Vmat,1);
    b=size(Bmat, 1);

    immunemat=zeros(v,b);
    immunecnt=zeros(v,b);

    % bindex contains the indices of bacteria who have added a spacer.

    Bhas=Bmatobject(Bmat,bmax,'bhas');
    bindex=find(Bhas(:,1))';

    for j=bindex
        bact=Bmat(j,:);
        for i=1:v
            vact=Vmat(i,:);

            vlacks=nonzeros(Vmatobject(vact,vps,'vlacks'))';
            vhas=nonzeros(Vmatobject(vact,vps,'vhas'))';

            bhas=nonzeros(Bmatobject(bact,bmax,'bhas'))';
            bhasoriginal=bhas(bhas<=vps(2));
            bhasnew=bhas(bhas>vps(2));

            immunecnt(i,j)=length(bhasoriginal)-...
                length(intersect(bhasoriginal,vlacks))+...
```

```
            length(intersect(bhasnew,vhas));
        end
    end

    immunemat=logical(immunecnt);

end





function [output]=Bmatobject(Bmat,bps,request,input1)
    % This function is a program that takes a set
    % of parameters (bps is only bmax here) and a request for
    % data and creates an output.  This allows
    % any function to read any part of Bmat.

    switch request
        case 'deathlist'
            % List of viruses which have made it to
            % their "lifetime minimum" and are too
            % infrequent to continue growing.
            output=logical(Bmat(:,2)) & (Bmat(:,3)==Bmat(:,4)) & (Bmat(:,2)<=1E-3);
        case 'maturedeath'
            % List of viruses which are above their
            % lifetime minimum and have become too
            % infrequent to survive.
            output=logical(Bmat(:,2)) & (Bmat(:,3)>Bmat(:,4)) & (Bmat(:,2)<=1E-5);
        case 'blost'
            output=Bmat(:,bps+5:end-1);
        case 'bhas'
            output=Bmat(:,5:bps+4);
        case 'changelost'
            if size(input1,2)~=bps(1)+1
                error('Error.  Incorrect number of spacers for blost.')
            elseif size(input1,1)~=1
                error('Error.  Too many bacteria.')
            else
                output=[Bmat(1:4) Bmat(5:bps+4) input1];
            end
        case 'changehas'
            if size(input1,2)~=bps(1)
                error('Error.  Incorrect number of spacers for bhas.')
            elseif size(input1,1)~=1
```

```
            error('Error.  Too many bacteria.')
        else
            output=[Bmat(1:4) input1 Bmat(bps+5:end)];
        end
    case 'changefreqs'
        Bmat(:,2)=input1;
        output=Bmat;
    case 'timestep'
        lives=logical(Bmat(:,2));
        Bmat(lives,3)=Bmat(lives,3)+1;
        output=Bmat;
    case 'createlife'
        Bmat(:,3)=0;

        %Bmat(:,4)=nbinrnd(1,.4);
        %Bmat(:,4)=poissrnd(3);

        output=Bmat;
  end
end

function [Bmatnew, bcnt,whofromtemp, match]=addbacttype(Bmat,bact,bps,bcnt,numnew,
next_row)
  % The new bacteria is given an initial lifetime.
  % bact=Bmatobject(bact,bps,'createlife');
  % The list of spacers from bact are removed for comparison to the
  % spacers in Bmat.
  bact(3)=0;
  match=0;

% If the new mutant is unique, then it is added into Bmat.
  if match==0
    % The new mutant is given its own strain number.
    bcnt=bcnt+1;
     % The new mutant's origin is given (who it mutated from).
    whofromtemp=[bcnt bact(1)];
    bact(4)=bact(1);
    bact(1)=bcnt;
    % The number (not frequency) of new mutants is determined by the
    % size of the mother strain.
    bact(2)=numnew;
    bact(end)=poissrnd(2);
    Bmat(next_row,:)=bact;
  end

  Bmatnew=Bmat;
```

```
end
function [output]=Vmatobject(Vmat,vps,request,input1)
   % This function is a program that takes a set
   % of parameters (called vps) and a request for
   % data and creates and output.  This allows
   % any function to read anytime of Vmat.

   switch request
      case 'deathlist'
         output=logical(Vmat(:,2)) & (Vmat(:,3)==Vmat(:,4)) & (Vmat(:,2)<=1E-3);
      case 'maturedeath'
         output=logical(Vmat(:,2)) & (Vmat(:,3)>Vmat(:,4)) & (Vmat(:,2)<=1E-6);
      case 'vlacks'
         output=Vmat(:,5:vps(1)+4);
      case 'vhas'
         output=Vmat(:,5:end);
      case 'changelacks'
         if size(input1,2)~=vps(1)
            error('Error.  Incorrect number of spacers for vlacks.')
            output='error';

         elseif size(input1,1)~=1
            error('Error.  Too many viruses.')
            output='error';
         elseif sum(input1>vps(2))
            error('Error.  Adding a gained spacer to the lacks list.')
            output='error';
         else
            output=[Vmat(1:4) input1 Vmat(vps(1)+5:end)];
         end
      case 'changehas'
         if size(input1,2)~=vps(1)
            error('Error.  Incorrect number of spacers for vhas.')
            output='error';

         elseif size(input1,1)~=1
            error('Error.  Too many viruses.')
            output='error';
         elseif sum(nonzeros(input1)<=vps(2))
            error('Error.  Adding a lost spacer to the gained list.')
            output='error';
         else
            output=[Vmat(1:4) Vmat(5:vps(1)+4) input1];
         end
      case 'changefreqs'
         Vmat(:,2)=input1;
```

```
        output=Vmat;
    case 'timestep'
        lives=logical(Vmat(:,2));
        Vmat(lives,3)=Vmat(lives,3)+1;
        output=Vmat;
    case 'createlife'

        Vmat(:,4)=poissrnd(3, size(Vmat,1),1);

        Vmat(:,3)=0;
        output=Vmat;
  end
end

function bactnew=bactmutate(bact, spacerprobs, bmax)

  % The spacer to be added is chosen based on the bacterial strains
  % interactions with viruses.

  %Identify the spacers of nonzero frequencies
  a=find(spacerprobs);

  %Get their relative frequencies
  b=spacerprobs(logical(spacerprobs));
  spacerprobs=cumsum(b/sum(b));

  % The spacers that the bacterium already posesses are identified for
  % later use.
  bhas=nonzeros(Bmatobject(bact,bmax,'bhas'))';

  % Numhas in the number of spacers that the bacteria already has.
  numhas=length(bhas);

  % choosespacer is a random number used to choose the new spacer.
  choosespacer=rand(1);

  % The chosen spacer that is added is the spacer with the greatest
  % cumsum value that is less than choosespacer.
  chosen=find(spacerprobs<=choosespacer);

  % If the value of choosespacer lies within the first cumsum "bin," then
  % the chosen value becomes 0, so that the chosen spacer will be the
  % first one on the list.
  if isempty(chosen)
    chosen=0;
  end
```

```
    % The spacer is identified.
    newspacer1=chosen(end)+1;

    %We go back to a to get the actual spacer
    newspacer=a(newspacer1);

    % The new spacer is added to bhas.
    bhas=[bhas newspacer zeros(1,bmax-numhas-1)];
    % The program Bmatobject updates bact to contain the new spacer.
    bactnew=Bmatobject(bact,bmax,'changehas',bhas);
end
```

Table S1 – Table of parameters used in model

Symbol	Value (Range Probed)	Description
K	10^6 (10^5-10^8)	Interactions per iteration.
S	50 (1-300)	Proto-spacers per viral genome.
P_{v_mut}	.003 (10^{-4} - $3 \cdot 10^{-3}$)	Expected frequency of interactions in which viruses mutate a random proto-spacer. For DNA microbes this has been measured at ~.003 mutations per genome per replication(Drake et al., 1998).
P_{b_add}	8 (1- 10)	Expected frequency of interactions in which hosts unidirectionally add a random spacer. On average, 8 strains add a spacer, driving clonal interference ('kill the winner') and multiple-mutation driven sweeps.
P_{b_lose}	0 (0-1)	Expected frequency of spacer additions in which hosts delete a random spacer block.
$f(n)$	$10^{(-4+n)}$ n>0 $1\text{-}10^{-9}$ n=0	Given n shared spacers, probability virus-host interaction is productive (i.e., viruses burst and host dies).
b_{prod}	200	Viral 'burst size' of productively infected host cells. Used to demonstrate inferred changes in relative abundances.
i_V	0.1 (.01-0.5)	Fraction of parent strain's frequency that each viral mutant is initialized with. Because CRISPR immunity is genetic, fitness is inherited from parent strains.
i_B	0.1 (.01-0.5)	Fraction of parent strain's frequency that each host mutant is initialized with.
G	3 (0-3)	Average of Poisson-distributed clearance-free iterations given to each new host and viral mutant strain to accelerate mutational turnover.
V_{min_freq}	10^{-6} (10^{-8}-10^{-3})	Frequency threshold below which viral strains beyond their emergence iterations are cleared.
B_{min_freq}	10^{-6} (10^{-8}-10^{-3})	Frequency threshold below which host strains beyond their emergence iterations are cleared.
V_{list_max}	300 (100-1000)	Maximum number of surviving viral strains beyond their emergence iterations.
B_{list_max}	300 (100-1000)	Maximum number of surviving host strains beyond their emergence iterations.

Table S2. Spacer sequences from I-plasma CRISPR locus (Fig. 1A). Total of 68 unique spacer sequences were detected. Perfect and imperfect matches to AMDV5 are noted.

Spacer ID	Spacer Sequence	Match to AMDV5
1	AATTTGACATAGGTAATGCTCAAGCTTACTGTGAATT	
2	CTATCAATGTCTAACTCTTTATCATTTTTACTATCT	
3	GAATGATATTTACCGTACAATGAAACAAAATAATCT	
4	GCTTACTAAAAGGTTATTCCATAGAACAATCAATGCT	
5	TCTAACTAAATTCAGTAATTTCATATATGCCTCATATAT	
6	ATACATCCTATACCTATATGGCGACCAGTAAGCAAT	Perfect
7	AATAATAAACGCCTGTTACTGCTCCTAGCAATCCAAT	Perfect
8	CTGAGGGAATAGTCATTAATACAGGCTCAAATTCTCA	Perfect
9	CTGATATGAACTTAGTTTTGCCACTTCCTGACCTTCCAAT	Perfect
10	CTGTGAATATACAATCAATTGCTCAAACTTTAACAAC	
11	CTTTAGTAGGTGGAAGCGTAACATCTGGTACAGCAAG	Imperfect
12	TTTGACAATACGCAAAATGGGTAAAATGAAAAAAG	Perfect
13	TATTGAATAATCCCAACATTGTGATTGAATGTAAGAC	
14	ACATCTGGTACAGCAAGTACTGGACTTGCATTTTTA	Imperfect
15	TAAGCCGATTTACGAACTTTATACTGATGTTTCAGCAGA	Perfect
16	TACAGTATTCAGTATCACGGAATCGCTGTAGGTTTCT	
17	TAGAACTATCTTGAATGGAAGAGTAAACTCTAAACT	Perfect
19	TTGACAAAAGTGTGATTATCTACCTGACAATAATA	Perfect
20	TCAGAATTGCTCAAAGACGTACCTAATTATCTTGTATT	Perfect
21	TCAGGACAGACTGTAACTGAGCCTATAACTTTCACATC	
22	TCTATTCTCAAGGATTACATAGTTCAAGAAGGAGAGA	
23	TCTCATCAAAAGTATTAGGAAAATAATAATAATTAC	Perfect
24	TCTCTTTCTGGTAATCAGATATCTTGCCTGAATTATA	
25	TCTTGTACCAGATTCTACCAGATTATAACTCGGTTGG	
26	ATTGTTCAAATAGAGAATTTTATAGTCTCCTATTTCA	
27	CTTTAACAACTGCAAATACTCCTCAGGCTTTACCCTCA	
28	TGATTCTTTCAATATCCTTTTAAGATTACTCTCAAA	
29	TGGAATAACCTTTTAGTAAGCACAGAATCTTTAATGCAA	
30	TGTCCGTATCCCTCTGGACATTGTGCAGTATTTAA	
31	TTATCTGCTCTGATGTGCTATGATTCACAATTGCTTT	Imperfect
32	TTCTACTCAATCATCAATCATATTATATGTAGTAAT	
33	TTTAATTTGCTTTGCTGGATAGCCAGTTATATTCAA	
34	TTTCTATAAATCAGGTTACGGTCATTTATCGTTGAATCA	
35	TATACAATCAATTGCTCCAACTTTAACAACTGCAAA	
36	GGGGGGAAATTCTGAATCTTGGGATGTTGAACAAA	Imperfect
37	TTTACACTTATCCACGTGTTAAAAGGACTGAATATCTT	
38	TTATCTGCTCTGATGTGCTAAGATTCACAATTGCTTT	Imperfect
39	GCGGCAAAATGTCATCTTTTAAGACTTTAGGTAATTT	Perfect
40	ACGTCTATGCCCGATAATCCCCCAACAATATCCATA	Perfect
41	GATTTACGGTTGATTACAAATTTAATGACAATGTA	
42	TACCTATAGGAGATGTTGTAGTTGTCAGTCCTATTGA	
43	AATTTACCTATTAAGCCTGAGACGATTAAAGTTCTA	Imperfect

44	ACCTATAGGAGATGTTGTAGTTGTCAGTCCTATTGA	
45	TTATAATCCTCATTAAAGTAATAATGTCTCATTAAT	Perfect
46	CTCCGTCACTTCGTGAGCGATCCCGTATCCGGCTATGT	
47	CCTAACTGGGTATGACATACAGAACCATGCTTTTGC	
48	TAGAATCCATCAGACAGAATCCAGAAGGTCTAGCTA	
49	TATAATGAAATTGCGAGGTAAAAAAGCATGAACGAA	
50	TACAGTATTCAGTATCACGGAATCGCTGTTGGTTTCT	
51	TACAGTATTCAGTATCACGGAATCGCTGTTGGTTTCTG	
52	CAGCATTCCCATAGACACATCTAGGAACGCAATATTTCC	
54	ATGATGCAAGAGCTGCTTATAGCTTGTGGATAGACGT	
57	TTTGCTATTGCGGTTATTTGGCAATAAGCTCAATCCT	
58	CCGTAGCAACCGTTCAACCGAATGAGCCCGTGATAAG	
60	CGAATCGCTAGAATTATCTGCATAGACTATAGTGGA	
62	CTCCCCATACGAGATGTTCCTGGAAATGGAATTCA	
63	ACATTCCCATACTTTTGTATAATTACTGTTAATATC	
64	TAAGTGATTTTCACGATATACAGAAAAATGTAGATG	
65	TCTGCAATCCTTACATTCCCATACTTTTGTATAAT	
66	CTCCCCTTACGAGATGTTCCGGAAAATGGAATTCA	
67	CCGTAGCAACCGGTCAACCGAATGAGCCCGTGATAAG	
68	CGAATCGCTAGAATGATCTGCATAGACTATAGTGGA	
69	GAATTTAGAGAAATTAATTGAGGCAACTGATATATT	
70	TTTGCTATTGCGGTTATTTGGCCATAAGCTCAATCCT	
71	CCTGAAACTGATTCTACCACAAAACTTCCATCATATCG	
72	CCTTGAAGCATATGAGGTAGAACAACCCCCAGTATTCTCA	
73	TATAATGAAATTGCGAGGTAAAAAAAGCATGAACGAA	
74	TACAGTATTCAGTATCACGGAATCGCTGTTGGTTTCCT	

Table S3. Spacer sequences from G-plasma CRISPR locus (Fig. 1B). Total of 399 unique spacer sequences were detected. Perfect and imperfect matches to AMDV3b are noted.

Spacer ID	Spacer Sequence	Match to AMDV3b
1	TATAAAGTATCCGTTATACTTTAAACTCTTATATCCTTTAC	
2	TATGATAATTATCATGATAACTATTAACACTATTACAAG	
3	CGGCGGAAAACCTTCGAATAAAGCAATCTTTTCTCTATCAATG	
4	AATTCTTAAATTTCATTATTATCATACCTCTTTTTCTTC	
5	TTGCTGTAGGGCGCGTTTAATAGGTCTGTGTAGTATC	
6	TATAAAGTATCCGTTATACTTTAAACTCTTATATCCTTTAC	
7	CATGGATATACCCCATACTCTTTCCATAGAACCTCTGCAC	
8	CATTTCCTTAATCTCATTTTCTAATCCATTGATCTCATCCAC	
9	AGTACCGAGCGGGAGATTTATGTTTACTCAATGGGTATC	
10	TATCTCTCTAAAGTCAGCCTTATTTAAAAAAAACACGACTTGAG	
11	TATATACATTAATAATACACGTGAGACTATGTGAGGATTT	
12	TCCAGAAATATCCGTCAGATGCCTATGTTTTCCAAGCCTATGAT	
13	AGATCTTTTACTCTTTCCAATGTAGCTTCCACTACATCAG	
14	TTTAGATCGTTCAGTCTCCTGTATTCTTCATCGTTTAGTGC	
15	CCTGAAATGCCTCTTTTTTACTTACGATCTCCCCATACGATC	
16	TCCTTTATTAGTAGTTAAAATATATGATTCTACTTCCGCAT	
17	AGTACCGAGCGGGAGTATTTATGTTTACTCAATGGGTATC	
18	TATCTCTCTAAAGTCAGCCTTTAATTTAAAAAACACGACTTGAG	
19	TCTGACAGTTCTCGACCAAAAGTAATTACATTATATTTCT	
20	ATGCTTACTATATGTTGCATCTATAATATCCTTATATATTCC	
21	TCTGACCCGTCAGAAAATGTATCAATTTCCGCTTTTAAATC	
22	TTCGTTTTAAAATGCGGTTTTAGTATATTATAAAGCATATT	
23	CATGGATATACCCCATACTCTTTCCATAGAACCTCTGCAC	
24	CATTTCCTTAATCTCATTTCTAATCCATTGATCTCATCCAC	
25	TTCTGCAATGTGTCTGTACATAGATACCACCTTTCTTTGTGTGT	
26	TCGGATTCCTTCATCCCTTTGATCCATGTCTGAAGTGGATC	
27	TTCGTTTTCCATTCTGTTCACCTAGTTCCGGTCTCTGG	
28	TAGACATTGGTAAGGGTTCACTCTGATTATTACACCAGT	
29	TATGATAATTATCATGATAACTATTAACACTTATTACAAG	
30	TAGACATTGGTAGGGTTCACTCTGATTATTACACCAGT	
31	TAAGGTAAAATCCTTCCGGATGATCAACGGCGAACTTTTCT	
32	TGTTAAGTAAATACGTTACCTTCAAATAAAATGAAATCTTTG	
33	TAAATCATACGGAGACCAGAAATACGACCTGTCGTATTTCTT	perfect
34	TTCTTCAGTCATAATTATGTTATCAGTAACGTAAGTG	
35	CATACTTTCTCTCCTCTCCAATCTCAAAATCCAGGGCTTT	
36	CAGTGCACTCAATGTCGTGTTGGTCGTGCTTGGAACGTCGCT	
37	TGTAAGTATGGGAGTGTTCAGTACGAATATTTTGTTATTCAT	
38	AATATCACCACCATTTTACAAAAAGCATTCCCTTATCAGT	
39	ACATTAGCTTTTCTAATCAAGTTATCCTCACTCAATATTATT	
40	ATAATTCTATTGTAAAGCATATACACCAATGAGTATAT	
41	TCCAGAAAATATCCGTCAGATGCCTATGTTTTCCAAGCCTATGAT	
42	AGATCTTTTACTCTTTCCAATGTAGCTTCCACTACATCAG	

43	ATCATACGGTGTCCAGAAATATGTAATGGTCATATTTTT	
44	TGTCACTTTTTGCCGTCTTGCGTCAATGTCCTTGACACAT	
45	TTTCTGATTCTGTTTGTCTATACCGAACAGCTTTTTCAATCTGATC	
46	CATCAGGTTTTGCTTTTCGTATCCCATCAAAACACCTCCATG	
47	TCCAGAAATATGAATGGTCATATTTTTCCATAGGCTTTAC	
48	TCCATAACGTATTCGTGTCCATCGAATGAAAATTCGATATT	imperfect
49	TTTCCAACTCGTCATAAACCTTCTTCGACAATATTCCT	
50	TCAAAATAATGATTTTTCCTCTAATCCTCTTCCTCTTGCGG	
51	TCATTATTTTATGATCTTCTTTCTTGTCCTCTTTTTGTTCT	
52	TAGACATTGGTACGGGTTCACCTCTGTAATACTTACGACCAGT	
53	AGTAGTGGTAGGGGATTACCTTCTAATGTCCGGTG	
54	ACACATACTTTGTCAATGCTCTTATGGTAACTTTT	
55	AATATCACCACCATTTACAAAAGCATTCCCTTATCAGT	
56	AGATCTTTTACTCTTTTCCAATGTAGCTTCCACTACATCAG	
57	CCTGAAATGCCTCTTTTTACTTACGATCTCCCCATACGATC	
58	ATTTCCTCATCAGTTCCTTTTACTTCAGTTGTTTTATTT	
59	CTGAAATGATTTCACCGTAGTCTCTGCTTCCTTCCGTTTCT	
60	TTCTATTTCAACCTCTTCACTCATTTTAACCTTTGTTATTGCC	
61	TTCTCTTATGGTAACCTTTACACCATCTTTTCTCAATTCC	
62	TATATACATTAATAATACACGTGAGACTATGTGAGGATTT	
63	TAAGGTAAAATCCTTCCGGATGATCAACGGCGAACTTTTTCT	
64	TGTAAGTAAATACGTTACCTTCAAATAAAATGAAATCTTTG	
65	ATCTGAATGTTGTGTAACAAAAGAGAAAAATCCATAGTCAT	
66	TTTGTCAATGCTCTTATCGTCACTTTTTTGCCGTCTTGC	
67	ACCTCCTCTCTTTTTGCTTCCCTTCTGGCATTCCAAAAAT	imperfect
68	TTGCTGTAGGGCGCGTTTAATAGGTCTGTGTAGTATC	
69	ATTCTAGTATTCACTATCTGATATTCCAAATCGTTGTCTTC	
70	AGGTCAAAAATGTGAGAATGCAGTAAAGAAAGCTAACAGATTG	
71	AAGCCCAATCCGGGATAATATAGTATTACATCCTCCAAT	
72	CATACCGCCCATCCCTCAGCTTTGTGAGACCTGAGTGCGT	
73	ATCTGAATGTTGTGTAACAAAAGAGAAAAAATCCATAGTCAT	
74	ATTTCCTCATCAGTTCCTTTACTTCAGTTGTTTTATTT	
75	AATATCACCACCATTTTACAAAAGACATTCCCTTATCAGT	
76	TTCGTTTTAAAATGCGGTTTTAGTATATTATAAAGACATATT	
77	TCGGATTCCTTCATCCCTTTGATCCATGTCTGAAGTGGATC	
78	TAGACATTGGTAGGGTTCACTCTGATTATTACACCAGT	
79	TAAGGTAAAATCCTTCCGGATGATCAACGGCGAATTTTTCT	
80	TGTAAGTAAACTACGTTACCTTCAAATAAAATGAAATCTTTG	
81	ATGCTTACTATATGTTGCATCTATAATATCCTTATATATTCC	
82	CTCACATATTCCACTAATTTTCCCTTTGTAATTTCCTTCATTTCT	
83	TAAATCGTCTTTTGCTTACTACTACTACTTTTTTACCGAT	
84	CAGGTGTGAGTACTGTAACGATGCACGCTGCTGAAAAT	perfect
85	TGTAAGTAAATACGTTACCTTCAAATAAAATGAAATCTTTG	
86	AGTCGTCCCAGTCGTATGGGTTCTCAAAGTCCACCCGTGCG	perfect
87	TGTAAGTATGGGAGTGTTCAGTACGAATATTTTGTTATTCAT	
88	CCCCATAATATTCCTTTTCCCGTCATTGTCATTTTTCACC	
89	TCCTTTTCTGCTTTCTTCGTCCTGCCGCTCCTATATTCCGGTG	

90	TTGCAATTTTGTCTCTTTTTCATTTACATTTCTGTTTC	
91	CATACGAAATCAATTCTTCAATTCTCTATTTTTTGTCCCT	
92	CGGGATTTTTCAGGTTATTTTTGGATTTCTGAGGAAAAT	
93	TCTGACCCGTCAGAAAATGTATCAATTTCCGCTTTTAAATC	
94	TTCGTTTTAAAATGCGGTTTTAGTATATTATAAAGCATATT	
95	TCGGATTCCTTCATCCCTTTTGATCCATGTCTGAAGTGGATC	
96	TAAGGTTAAAATCCTTCCGGATGATCAACGGTCGAATTTTCT	
97	CACCAATCATCCATGCAAGATCAGAAGGGATTTCAGATTCCTTC	perfect
98	TATCAATTTCTGTGCAGTAGATGTATTCATTTTTTGCTGTTTC	
99	CTGAAATGATTTCACCGTAGTCTCTGCTTCCTTCCGTTCT	
100	TCTATTTCAACCTCTTCACTCATTTTAACCTTTGTTATTGCC	
101	TATATGTCATGAATTGTCATCACATTAGCTTGTTTTATGTAGG	
102	TCCAGAAAATATCCGTCAGATGCCTATGTTTTCCAAGCCTATGAT	
103	CTCATTAAATCCGCTTTCATATTATCAGGATTATATACC	
104	ATTTCCTCATCAGTTCCTTTTACTTCAGTTGTTTTATTT	
105	TCCAGAAAATATCCCGTCAGATGCCTATGTTTTCCAAGCCTATGAT	
106	AATATCACCACCATTTACAAAAAGCATTCCCTTATCAGT	
107	ATTTCCTCATCAGTTCCTTTTACTTCAGTTGTTTATT	
108	TGAAATTAAAGAATTTGATGAAAATAAAGATTTTACTTGAT	
109	ACAAAGAAATAAGGTATAAAAATCTAATAACCGTCACGGATTTTT	
110	TTTCCATCGGAAAGCATATCAATTTCCGCTTTCAAAATCGAT	
111	TTCTCTTATGGTAACCTTTACACCATCTTTTCTCAATTCC	
112	CTCATTAAATCCGCTTTCATATTATCAGGATTATATACC	
113	CCTGAAATGCCTCTTTTTTACTTACGATCTCCCCATACG	
114	ATTTTCCATTTTTCAATTCACTTCCATCCCTCATTTCAGGG	imperfect
115	TATATATTCAAAATGACATGACGTTCCGCAATTTTTACATT	
116	ACTTTATCAGCAATGCACTTGCATCTGAACAGGCATTG	
117	TGTAGTTTAAAAGACTACAATTTAAAATTTACAGGATCAC	
118	TGTGTTAAACACGGCTCTGACCACTTGATTTATCTGACT	
119	TTCCATGCCTCAAATTTCTATTTTTTCCATCTCATACC	
120	CTCATTAAATCCGCTTTTCATATTATCAGGATTTATATTACC	
121	AATATCACCACCATTTTACAAAAAAGCATTCCCTTATCAGT	
122	AGTACCGAGCGGGAGATTTATGTTTACTCAATGGGTATC	
123	TCGGATTCCTTCATCCCTTTTGATCCATGTCTGAAGTGG	
124	ACATTAGCTTTTCTAATCAAGTTATCCTCACTCAATATTATT	
125	CGGCGGAAAACCTTCGAATAAAGCAATCTTTTCTCTATCAATG	
126	AATTCTTTAAATTTCATTATTATCATACCTCTTTTTCTTC	
127	TAAGGTAAATCCTTCCGGATGATCAACGGCGAATTTTTCT	
128	ATAATTCTATTGTAAAAGCATATACACCAATGAGTATAT	
129	ATGTAAAAGAATATAAAGTTATTCATAAGTTAAATTCGTTT	
130	TATAAGAGAGATTGTTACATTGTTTGTTGTTGAAGTATC	
131	TGCAAATTCCCTTATGAACTCTTCCGAGCTCTTTTTACT	
132	TTCTTTCTGCCTTCCTTCGTTCAAATGCTAAATCATACGGTG	
133	CATACGAAATCAATTCTTCAATTCTCATTTTTTTGTCCCT	
134	CGGGGATTTTTCAGGTTATTTTTGGATTTTCTGAGGAAAAAT	
135	TCCTTTTCTGCTTTTCTTCGTCCTGCCGCTCCTATATTCGGTG	
136	TTGCAATTTTGTCTCTTTTTTCATTTCATTTTCTGTTTTC	

137	CTGCAGAAACCTCTTACGTAGGTGCCTTTTCTGCTGTGTGCAGG	
138	TAAGGTAAAATCCTTCCGGATGATCAACGGCGAATTTTTCT	
139	CTCATTAATCCGCTTTCATATTATCAGGATTATATACC	
140	AGATCTTTTACTCTTCCAATGTTAGCTTCCACTACATCAG	
141	TTTAGATCGTTCAGTCTTCCTGTATTCTTCATCGTTTAGTGC	
142	AGTCGTCCCAGTCGTATGGGTTCTCAAAGTCCACCCGTGCG	perfect
143	TGTAAGTATGGGAGTGTTCAGTACGAATATTTGTTATTCAT	
144	CACCCTTTGAGATCCTCTTTACTTACAACCACTACCTTT	
145	TTCCGTCAGATAGCTTATCAATTTCCGCTTTCAAATCGAT	
146	AATATCACCACCATTTACAAAAGACATTCCCTTATCAGT	
147	CAGCATTTCAGGCTTGTAGGGATGGTTGTAGCAAACGT	
148	TCGGATTCCTTCATCCCTTTGATCCATGTCCTGAAGTGGATC	
149	TCTGACAGTTCTCGACCAAAGTAATTACATTATATTTCT	
150	TATAAAGAATTATTATAATAATATATATAACCTGAACTTACG	
151	TAATGCCCAGTTTCTCTCTCTGTTTGACTAAGTTCTTTAGCTT	perfect
152	CACTTTTACTTTTTTAAAATATCACGGCAGATACCGTGTCAGT	
153	TAGTATTCACTATCTGATATTCCAAATCGTTGTCTTCGCAAT	
154	TTAAATAAGATAAATTAATAAAATTGAAATTTGAACGATT	
155	CCTATATACAATGTATCTAATTTTTACAATTACGGCATCT	
156	TCCAATGTTTTTCTTTCCCGTATTTTTCCCTTCAAAATAC	
157	TACGCCGCCATCGGTCCAGTTATGGCCATCCCAGTAAT	
158	TATTTCTTATTGTTTTAACGCAGTTTGACGTCTAATTT	
159	CATCTAATCTTATATGTCGTCCAATTTTTGCCGCTG	
160	ATACGGTGACCAGAAATATGAGCGGTCATATTTTTTCCAC	
161	TTATATTTCTTATGGGGTAACAAGTTTATACTTGTTGCCATTT	
162	TTCGTTTTAAAATGCGGTTTTAGTATATTATAAAGCATATT	
163	CATGGATAGTACCCCATACTCTTTCCATAGAACCTCTGCAC	
164	TATCTCTCTAAAAGTCAGCCTTATTTAAAAAAACACGACTTGAGAAT	
165	CGGGGATTTTTCAGGTTATTTTTGGATTTCTGAGGAAAAT	
166	TCTGACCCGTCAGAAAATGTATCAATTTCCGCTTTTAAAATC	
167	TACAATTTTCCCTTGATTGTCCCTCATTTCTACTCTC	
168	CCTTTTTCCGAACTTGCCTTCGAAAAGCCTTTCCTCTTTTCT	
169	TTCGATACTTTTCACCGTCCTCTGTTTCTGATTCCAGT	
170	TGCAAAATTCCCTTATGAACTCTTCCGAGCTCTTTTTACT	
171	TTCTTTTCTGCCTTCCTTCGTTCAAATGCTAAATCATACGGTG	
172	CATACGAAATCAATTCTTCAATTCTCATTTTTGTCCCT	
173	TTTAAAAGAAATATTAAAAGACGTTATTAAATAAATATAATAAAAT	
174	AGGTCAAAATCTATTTTTTATTCCGAAGGTATATATATGCT	
175	TTCAATAATATTTCTTATTCTTCTTTTCAATCAATTCTTCAT	
176	AGAAATTTTACCATCTGTCTTCTCTGGCTTTGTCTCTTCTGG	
177	ATGAACTCTTCCGAGCTCTTTTTACTCCAAATGCAACCCC	
178	CATATTCCTGTTGTTGTGTCAGGGGCAGCAGACTGCAATATTCT	
179	AATTCTTTAAAATTTCATTATTATCATACCTCTTTTTCTTC	
180	ATCCTCGAAGAGTTCAGAAATGACGTCATTTTGCTCCTATCTAT	
181	TTGGTGTTATACTTGACCCTTGCAATCACGTTGCTCTGCC	
182	CTTAGTTTCTTCTGCGTGAAGCTCAATTGTATTGCCTGATC	
183	TCGTGAAGATCGCCGTTCAGAATGCCGATGTCTCTGTTGTT	

184	TTGCTTTGTTTCATCAACCCAGATGGTCATTTCGGCTATTG	
185	CAGCATTTCAGGCTTGTAGGGATGGTTGTAGCAAACGT	
186	ATTTCTTTCCAACTTCAACAGTATTTCTTTATTCTTCTTTT	
187	TTGCGGAAGTTCCTCTGCGAGTTTTCTCAACTCGGCGAATTT	
188	TCGTAGCGGCATTTATCTATTTTTCCTTTATCTTTTTATT	
189	CGGCGGAAAACCTTCGAATAAAAGCAATCTTTTCTCTATCAATG	
190	AATTCTTTTAAAATTTTCATTATTATCATACCTCTTTTTTCTTC	
191	TTCAGATTTTCCTGAATCTTTTCGTCCTTGCGGACAAGAGC	
192	TCTTTCAAATTCACTTCCATCCCAATCCGGGATAATATAGTATT	imperfect
193	CATACGAAATCAATTCTTCAAATTCTCATTTTTTTGTCCCT	
194	CCTTTTTTGATAGGTCGGTTTTTTGTTTAAGATTTTTCTTATCGAT	
195	TTACAACCGTTCAGGTCCTCTTTAGTCACTACCACTACCTTTT	
196	ATATTTTTTCAACATCGCAAATGTTAAGTCAGGTAGAT	
197	CATAATTTCTTTTAAATATTTATTCCATTCGCTGGGTGTAT	
198	TGTCCATAACATAGTCATGTCCATCGAATGAAAATTCGATGT	imperfect
199	TAAGGTAAAATCCTTCCGGATGATCAACGGCGAATTTTCT	
200	AATCTTCTCCATTTAGATCGTCGAAATACTTTTCAGCATC	
201	TTCTCCATCCTTCCTCTCCTGGTCAAAGGACTTCATGTATG	perfect
202	TGCCCCACACAATTTTTCCTTCAAGATCCTCCCCACTTTCT	
203	ACTTTTTCTTTTGTTATTTCCTTTTCCATTTTCTCACTT	
204	TTCTCCATCCTTCCTCTCCTGGTCAAAAGGACTTCATGTATG	perfect
205	TGCCCCACACAATTTTTCCTTCAAGATCCTCCCCCACTTTCT	
206	TTTTTATGCGTGTTATTTTTTTCTTATTTATTTTTACATT	
207	TGTACTGGAAGTCTAATCCGTCTTCCTGTGGGAGTTCCTCTGCG	
208	TATCTCTCTAAAGTCAGCCTTATTTAAAAAACACGACTTGAG	
209	ACATTAGCTTTCTAATCAAGTTATCCTCACTCAATATTATT	
210	CGGCGGAAACCTTCGAATAAAGCAATCTTTTCTCTATCAATG	
211	AATTCTTTAAATTTCATTATTATCATACCTCTTTTCTTC	
212	TTGCTGTAGGGCGCGTTTAATAGGTCTGTGTAGTATCAT	
213	AATATCACCACCATTTACAAAAAGCATTCCCTTATCAGT	
214	TTGCTGTAGGGCGCGTTTTAATAGGTCTGTGTAGTATCAT	
215	TTTTAAATCGTCTTTACATACTACTACAACCTTTTT	
216	ACATTTTAATCCCTCCTCCTAATTAAGCCTTCATTTACTTG	
217	TTCATACGAAGTCTGAACCCTTGGTTCTCTCCTTTTCCTTT	
218	CATGGATATACCCCATACTCTTTCCATAGAACCTCTGCACAT	
219	AGTAGTGGTAGGGGATTACCTTCTAATGTCCGGTG	
220	CTTATTTTAACCTTTGTTATGGCTAATTCAGTATCGCTT	
221	CACTCTATTAAACTTATCAATTTAGCCCGTGTTATTTTCTTTC	
222	ATATTAAAAGGTGTTATCCCTTTCTGCTTTTTCTTCGT	
223	CACATATCCCTTCAGCTTCTCCATCGCTCTTGTCACTGCAT	
224	TCGGAACAGTGGCGTTCCGGTGCGGGTTGTAAATATATATCT	
225	TCATTATTTTTATGATCTTCTTTTCTTGTTCCTCTTTTTGTTCT	
226	ACACATACTTTGTCAATGCTCTTATGGTAACTTTTT	
227	TATGATAATTATCATGATAACTATTAACACTATTACAAG	
228	TAAGGTAAAATCCTTCCGGATGATCAACGGCGAATTTTCT	
229	AATCATCAGTGCCTTTTATATACTTTGTCAGCACCCTT	
230	CATACGAAATCAATTCTTCAATTCTCATTTTTGTCCCT	

231	TGCCCGCTTTGAGATTTCCATCGAACACTTTTTTTGCATTATG	perfect
232	AATATTTTCATAAGTTGCATTAATTTCAACTTGTAAGGTTAAATT	
233	AATCATCAGTGCCTTTTACATACTTTGTCAGTGCCC	
234	TCCCATGAATGCCTAAACTCCTTATATTACATCCGTT	
235	AATTTTCTGTGTCCAGAAGGTCATCGTTCCCACTCGCCAG	
236	TGCATACTTCGGCAAACACTTTATCAGCAATGCACTTGCATCTG	
237	CCTTTCATATCCCAGAATACTCTTTTTTCCATAGTTGTATAT	perfect
238	TTGTTATTTCATTCATTTCTATTTCCTTCATCATTTTTAC	
239	TTATTTAAAATAGATGAAACTTCTTCACCTATTGATAATCTT	
240	TTATCTTCTTCAGCCTTTCTCCTTTGAAATGACAGATCAT	perfect
241	TAACCATCGCCTGTACCGGTTTCAACATACCCAATATAGC	
242	TGTAGTTTAAAAAGCTACAATTTAAAAATTTCAGGATCAC	
243	TGTGTTAAACACGGCTCTGACCACTTGATTTATCTGACT	
244	TTCCCATGCCTCAATTTCTATTTTTTCCATCTCATACC	
245	TCCTTCTTCACTGTCAGGGTCTTACCAAACAGCTCCTC	perfect
246	ACCGCATCTAATCTTATATGTCGACCAATTTTTGCCACTG	
247	ACATTTTTGATTATGAACTTTATTTGTATTGAAAATAT	
248	TGCGATAAAGATCGCAAAAAAATCCCGCTATTGCCAGTCT	
249	ATAATTCTATTGTAAAAGCATATACACCAATGAGTATAT	
250	TCAAAACTTCTTTCCCCAATAATTGCCTAATGAAATTGT	
251	TAACCATCGCCTGTACCGGTTTCAACATACCCAATATAGC	
252	CACTTCCATCCCTCATTCAGGGATAATATAGTATTACAT	perfect
253	TATTTTTTCCATAGACTTACATATTCCTTTATTTCATTC	
254	AATATTTTCTTCAGTCATAATTATGTTATCAGTAACGTAAGTG	
255	TGACCTTTTACCTCTTCTAACGTGTGAGGTCTGAACTTTCC	
256	TGCATCTGAACAGGCAATCACAGGACAGACTAAAGCCTGAT	
257	TGTTCTCATTGGGAGTTTCCTTCCCGATCTCTTTCTCTTCT	
258	TTTTAGATCGTTCAGTCTCCTGTATTCTTCATCGTTTAGTGC	
259	TCACTTTTTCCTTTGTTATTTCCTTCATTTCCTTTTCCATTTC	
260	TATACTTTCTCTCCTCTCCAATCTCAAAATCCAGGGCTTTC	perfect
261	TGCGTTCCTATCCCGCCATCGGTCCAGTTATGGCCATCCCAGT	perfect
262	CGTCCAATTCAACCTCTATACGGTCAAAAAAGTCTTTAAT	
263	TTTCCTCCTCTCTTACGATATTCAATAATAATTTTATTAG	
264	CATCTACTTCCATTCATGTCCACATCTTGGACATTTTACAAT	
265	TCAGATTATTTTCCTGAATCTTTTCCTTGCGGACAAGAGC	
266	TTTTCTGCATGAGAAGAAATAGTATTCCAGTGTTTGAT	
267	ATTAATTGTAAGTATACAGGTTAATTTATTTAAATCAT	
268	TTTAAAGAATATTAAAGACTTATTAAATAATATAAT	
269	AGGTCAAATCTATTTTTATTCCGAAGGTATATATATGCT	
270	TTCAATAATATTTCTTTATTCTTCTTTTCAATCAATTCTTC	
271	AGAAATTTACCATCTGTCTTCTCTGGCTTTGTCTCTTCTGG	
272	CGTCCCAATTCAACCTCTATACGGTTACAAAAAGTCTTTTAAT	
273	CGGGGATTTTTTCAGGTTATTTTTGGATTTCTGAGGAAAAT	
274	TACAATTTTCCCTTGATTGTCCCTCATTTCTACTCTC	
275	TCTAATAACCGTCACGGATTTTTTCACATTAAAAATAAC	perfect
276	CTCACCTATTGATAAATCTTTGTCAGTTCCCTTATTGTAACT	
277	TTCGATACTTTTCACCGTCCTCTGTTTCTGATTCCAGT	

278	AATTCTTTAAATTTCATTATTATCATACCTCTTTTTCTTC	
279	ATTTCTTTCCAACTTCAACAGTTATTTCTTTATTCTTCTTT	
280	TCATTATTTTTATGATCTTCTTTCTTGTCCTCTTTTTTGTTCT	
281	CACCTAAAACGCAGAGATTTTACAGAATATTCTTTTTTGCTCG	
282	TCATTTTTTTAGACCCTCCTTCGGCAAAATATCGTGTCAATGCTC	
283	TTGCTTGTTCATCAACCCAGATGGTCATTTCGGCTATTG	
284	ATTTCTTTCCAACTTCAACAGTATTTCTTTATTCTTCTTT	
285	CGAATGAAAATTCCATGTGATATGTGTTATCCCTTTCCGCTT	imperfect
286	TTTTCCATAGACTTACATATTCCTTTATTTCATTCATTTTC	
287	TTTAACAAAGAAATAAGGTATAAAATCTAATAACCGTC	
288	AATTCTTTAAATTTCATTATTATCATACCTCTTTTTTCTTC	
289	TGAATCTTAATTCATGTTTTTTTATTTTGAAATACGATTGCTTTT	
290	TGTAAGTAAATACGTTACCTTCAAATAAAATGAAATCTTTTG	
291	TACAATTTTCCCTTGATTGTCCCTCATTTCTACTCTCAAT	
292	TACTCTTGCAATCACGTTGCTCTAGCCAGTCCACGTCTTC	
293	TTTGCATTATGTTTATCACGGTAGCTGTCAGTCTCCATGAAC	perfect
294	TTTTCCATGTTCTCAATCACTTCCATCCCAATCCGGGAT	
295	TATATAATGTTACATTGTAAATAAAATAAAGTAAAAATTAAT	
296	TCCAGAAAATATCCGTCAGATGCCTATGTTTCCAAGCCTATGAT	
297	TCAAACTTCTTTCCCCAATAATTGCCTAATGAAATTGT	
298	TGTAGTTTAAAAGCTACAATTTAAAATTTCAGGATCAC	
299	TATATTCCTTTATCTCATTCATTCCAACTTCCTTCATTTCCTTTT	perfect
300	TGCCCCACACAATTTTTCCTTCAAGATCCTCCCCCACTTTCT	
301	CCCAAGACACTGCAAGTGCCTTAAAGATACTCTGTAACC	
302	TATATACATTAATAATACACGTGAGTACTATGTGAGGATT	
303	TCGGATTCCTTCATCCCCTTTTGATCCATGTCTGAAGTGGATC	
304	TTCGTTTTCCATTCTGTTCACCTAGTTCCGGTCTCTGG	
305	TGGGGCTTTTACTCCATCCTCAAAAATCCCCTGTCCTCTTCATC	
306	TTTACGAAGGAGAGCGGAAGAAAATAATACATATCACATCG	
307	ATTCTAGTATTCACTATCTGATATTCCAAATCGTTGTCTTC	
308	AAGCCCAATCCGGGATAATATAGTATTACATCCTCCAAT	
309	CTTTCTTCTGTATTTTTTGGGCATGGTACTCATGAACAC	
310	CATGATATACCCCATACTCTTTCCATAGAACCTCTGCAC	
311	CATTCCTTAATCTCATTTTCTAATCCATTGATCTCATCCAC	
312	TATCTCTCTAAAGTCAGCCTTATTTAAAAAAACACGACTTGAG	
313	CCAAAAATATGAATGGTCATATTTTTTCCATAGACTTACATATTC	
314	CTTACATATTCTGCTAATTTTCCCTTCGTAATTTCCTTCAT	
315	TCATATTTTTTCCACAAATTTACATATTCCTTTATTTC	
316	AGAAAATGATTTGCCTGACCCTAAATTACCATAAAAATGTAT	
317	TTCCTTAACGCTTTCTGCAACAACAACATCGGGAAGCTCT	
318	CTGTTCAATACTTCGAAATTTCTTTCCAACTTCAACAATATT	
319	CCTCTTTGTCAATCACTTTATACTTTCTCTCCTCTCCAATC	perfect
320	TGATCTACTGCTTCTGGTCCAAGTTTTATCAAGGCGTAGCC	
321	ATGAACTCTTCCGAGCTCTTTTTACTCCAATGCAACCCC	
322	TCAAACTTCTTTCCCCAATAATTGCCTAATGAAATTGT	
323	TGTAGTTTAAAAAGCTACAATTTAAAAATTTCAGGATCAC	
324	TATCTCTCTAAAGTCAGCCTCATTTAAAAAAACACGACTTGAG	

325	CCCCATAATATTTCCTTTTCCCGTCATTGTCATTTTTTCACC	
326	TTCATACGAAGTCTGAACCCTTGGTTCTCTCCTTTTCCTTT	
327	ACATTTTTAATCCCTCCTCCTAATTAAGCCTTCATTTACTTG	
328	TTTTTCCATAGGCTTACATATTCCTTTATTTCATTCATTTC	
329	ACTTTGAAGGTCTGAGCAGTATTAGCTTTTGTTACGTC	
330	TTTCCATTTTTCCGTTTTCCATCTTTCAATTCACTTCCATCC	
331	TGTACACGTATTTTTCGATTCTTGTCTTGTCTGACATCTTAAT	
332	CTGAATTGAGATATTTCCCAATTCCATTTTTCTTTTTTTCATTT	
333	CATACGAAATCAATTCTTCAATTCTCATTTTTTGTCCCT	
334	CTTTATTTCATTCATTTCAATTTCCTTCATTTCCTTTTCC	
335	TAATTTTTTACAATTACCGCATCTAATCTTATATGTCGTCC	
336	CATCTGTCCAGTTATGGCCATCCCAGTGGTCCAGGTTGGTGTT	
337	TGCCACGAATGTTTCGTAGTTGGTGATCGTTACGTCGTGAAG	
338	TCGGCAGATATTTTGTCAATGCTCTTATGGTAACTTTTTTGCC	
339	TGTTCATAACATAGTCATGTCCATCAAAAGAAAATTC	
340	TTCGATATTAAAGGTGTTATCCTTTTCTGCTTTTCTTCGTAAAG	
341	TCATGATACCATTCATCACTTTTGTATTCAAACGGTTGT	
342	TCCTCCGTAAAGCTGACAAATCGTACGGAGACCAGAAATATG	
343	TTTCAGGTCCTCTTTACTTACTACCACTACCTTTTTACCG	
344	TGGGGATTCCACCTAAAACGCGATTTCAGAATATTCTTTTTT	
345	TACAGCGATGCACGTAGGTCAATGTCAAATCAAAATTTGC	
346	TTTTGAAAAAAAATATCTTAGCGTTCAGGACATTTTTGACG	
347	TTTACATATTCCTTTATTTCATTCATTTCAATTTCCTTCATTTC	
348	TCATCAATTTTAACTTCCTCACTCATTTTTACCTTTAGTATGG	
349	TGTACAATACCACTTATTGGTTTAGCTGGAACTTTGTGATGCT	
350	CCCTCATTCAGGGATAATATAGTATTACATCCTTGTATAT	perfect
351	CTTGCAGGATTAGGTGCATACTTCGGCAAACACTTTATCAGC	
352	TTGCATCATAGTACGTTATATTTCCATTCGATTCTATAAT	
353	TTAAATACAATTTTATAATATCTAAATGCGGGACATGAGATT	
354	ACTATGTAATGAACATAGTCATGGAACAAAGCACAAG	
355	TTTTGTTATTTCAGTCAGTTCATGTTCTGTTTTCATTTTTTAG	
356	AACTTCAACAATATTTCGTTATTCTTCTTTTCAATCAATTCT	
357	TCATTTTCTCATTTTTCTTTTTTTCACTTTCCATTTTCTG	
358	TATGTAATGGGATTCATCCAGAACCAGAACATCCCACT	
359	CTCCAATGCAACCCCATAATATTTCCTTTTCCCGTCATTGT	
360	TCCTCCTCTCTTACGATATTCAATAATAATTTTATTAGCGG	
361	TCCCTCAATTTTGTGAGACCTGAGTGCGTTCCTACACCTC	
362	CATATTCCTGTTGTTGTGTCAGGGGCAGCAGACTGCAATATTCT	
363	TCTGTCTCAAGTAGTTCTTCATTCCCGCTGGCCAGTATTTTCG	
364	TCCATGCCTCAGTTGCTATTTTTTCCATCTCATACC	
365	TTCATACGAAGTCTGAACCATTGGTTCTCTCATTTTCCTTT	
366	CATTTCACTTAATCTCATTTTCTAATCCATTGATCTCATCCAC	
367	ATTATTTCACCTTTTTAATTTCCATATTGGGCTCATCG	
368	ATTAAAATTAACGGTATTTCCATAAGAAAAATTAAATT	
369	TTCATATTGAGTTAAATATTTTATATTTATAAATTTATAT	
370	ATCTGAATGTTGTGTAACAAAAGAGAAAAAATCCATAGTCAT	
371	ACTCATTTTAACCTTTGTTATGGCTAATTCAGTATCGCTTAG	

372	CAGGATAATATAGTATTACAAGGTAGAATATAAATGTTTTG	imperfect
373	TTCCTTTTCCATTTTCTTTTCACCTATTTTCCGGTCTC	
374	TCCCTCATTCAGGGATAATATAGTATTACATCCTTATATAT	
375	TTCTTGCTTGCCACGAATGTTTCGTAGTTGGTGATC	
376	ACTTTGAAGGTCTTGGCAGTATTAGCTTTTGTTACGTC	
377	CATGGATATACCCCATACTCTTTCCCATAGAACCTCTGCAAC	
378	TATCTCTCTAAAGTCAGCCTTATTTAAAAATACACGACTTGAG	
379	TAAGGTAAATCCTTCCGGATGATCAACGGCGAATTTTTCT	
380	TGTAAGTAAATACGTTACCTTCAAATAAAATGAAATCTTTGT	
381	ATGCTTACTATATGCTTGCATCTATAATATCCTTATAT	
382	TCCTTATCCATTTTCTTTTCACCTATTTTCCGGTCTC	
383	TCCCTCATTCTGTGGATGATATAGTATTACATCCTTATATAT	
384	TTTCCATTTTACCGTTTACCATCTTTCAATTCACTTCCATCC	
385	TGTACACGTATATTACGATTCTTGTCTTGTCTGACATCTTAAT	
386	TTCCATGCCTCAATTTCTATTTTTTCCATCTCATACC	
387	TTCGTTTTCCATTCTGTTCACGTAGTTCCGGTCTCTGGATG	
388	TAATCTATGTAGGCTTTATAACCTTCTGCGTGGTTTTCCAACTT	
389	CGTCAATTGTCTTGACACATAGACGTGCTTTGTGCTTTGT	
390	CCCTTTTAAGGTCAAATCTATTTTTTATCCCGAAGGT	
391	ATATTCTTTTACTTTTTCCTTTGTTATTTCCTTCATTTCCTTTT	
392	CTGCTTCCTTCAGTTTCTGATCCGTGTATAAGAACATATCG	
393	TCAAACGGCTTTCCCCAATAATTGCCTAATGAAATTGTAT	
394	TAGATATATTTTTTTCATATCGTAAATTTTCCGTATC	
395	TTCAGAGTGAGTATCATTTTCTCACCCTTACTGTACTCTTAC	
396	TTCATTTCATTTTCTGTTTTCATTTTTTAGACCCTCCTTCGGC	
397	TCCATCCCAATCCGGGATAATATAGTATTACATCCTTATAT	
398	TTTCCATTTTTCCGTTTTCCATTTTCTCAATTCACTTCCATCC	
399	TGTGTTAAACACGGCTCTGACCATTGATTTATCTGACT	

Table S4. Spacer sequences from G-plasma CRISPR locus (Fig. 1C). Total of 1093 unique spacer sequences were detected. Perfect and imperfect matches to AMDV3b are noted.

Spacer ID	Spacer Sequence	Match to AMDV3b
1	TAATGTAGCAATGACCGCCGAATGAAAATTCCATGTG	
2	GCCATGAATCGAGGAAATTTGTTCTGTCCTCTACGAA	
3	AGAAAGGCTACGCAGGAAAGAGATCACATTTTCTCA	
4	GGTTACTTGGAACTGGAGACGTTCCGGATAGATCCT	
5	GTCTATTCCGTCCACTAGCGCACCTGCGATGATAAAA	
6	AAAAATACCCTTATCGGGGCATCATAGCAAATACATA	
7	TCGTAAGTGCCCTCCTGCGTAAATCGTCCGCCAGGT	
8	TATCTGTGACAGTCACGGATTCTCTCGGGGCATCCTCA	
9	CTTGTCTGGGGCATCAATGTAATTCCATTTGGGCACC	
10	AGTTAAATATGCTTGTAGAGTTGCAATATCGATATC	
11	TTTTATAAATAAAACCGGCTCCGTAAACGATGCTCG	
12	TCGAGGAACCTGCAAATAAGAGGATAAGCGACCTGA	
13	GAGAGATTGAACTGCGACACCGATGCATAGAATGTG	perfect
14	TTGCCACAGATATTACACAAATAGAGCCTGTGCTGG	
15	CTTGCGAATCCAGGGCATAAGCCAGGTGTTGTATCTCC	
16	CCCTAGCTGCCATGATCATGTCCTTCCTGATCGAAA	
17	AGTTGTGAGGCTTTGCTTTCCTTGCAGCAACTCTCT	
18	TCCATATTGGCTCTTATTCTTGCTCATTCTGTAGCAA	
19	CTAATACCTTTGTAAGTCCCAGATGCGAATATATTAA	
20	AAAATATGAACGATCATAATATTTCCAGAGACTTAC	
21	CTTCAGGCCGTACTTAGGATCAGTTCCTGTAATTTT	imperfect
22	GATTTCTTTTTTTAATCCGTAATACCTTATTACG	
23	CATTCCTAAAAAACATATTTTCAGATCTGATTATAT	
24	GTAGGCCTATCCGAACTGTCTAGGGGATTCATTTTTTT	
25	AAAATAATAATAGAGACACATAGCAAACAATTGCAT	
26	TATGTTCTATAGTATCTTTTACGTAAAGCTGTAGATC	
27	AAAGGCTAGGAATCCGACAGCCCCCACCAATATAAT	
28	CCTTTGAAAAAATTCTATTGAAGATTGAGAAAAATT	
29	TTACCTTTCTTATTAAATTCCATATATTCAACATCT	
30	CCATTCGGGATGTACATAGAAGAATACAGCATGCCC	
31	GGTATCCATTCAAAACGCCATTCCACCAACTATAAGG	
32	TTATGCCGGTGGTGGCGTACGTATGCGTGGATGTG	perfect
33	TTCGATTCGGTCACAACCTGCGAGGGAGCCACTCCT	
34	TGATGGTATGTTTCCTTGTTTGCCTCAGGTTCTCAT	
35	ATTCCAACTCTATATTTGATTGTTGTGGCAGGAGTC	
36	CCAGACCTATGTTAATATCACATCTTTCCAGTTTAA	
37	GTATTGAAGTGCTGGAATCTGAAATAGAGGATTTAA	
38	AAAGAGATTGCTAGGCATACACACACACATAATTTTTC	
39	GTGAATAAGGACATACCGTCCATCTCTCAAATTTTTGTGA	
40	GTGGTATGTTACCTGTTCGATGTAGTTTCCTGTGTTT	
41	CTGATCCAGAAACAGAAGCTTTCCTTTCTGACAATCC	
42	AAGAGAGCACAGGGAAAATAGATGGTTCGATGACTG	

43	CTGCTTTTAATATAGTAGTAATTGTGTCATTATATA	
44	TTCTATACTCCAAGGTTTCACAACCATACTGCCAGAC	
45	ATTATTTTTATAACAAATACCGTAATCCCAATCAT	
46	TTGTGATCATTACCATGAATCATTGGAACTATTTGAA	
47	TATGGACATGTCTCAGCCATATAAAAATGTTTGGA	
48	GTGATTTGCTTTCCGGAGTGGTGTTCCCAATGATT	
49	AAGTAACATATGTCCCGATTATAATTAAAACATAA	
50	TTTTTTCACCTCTTATTAAGCCTGGCCCAAATATGA	
51	ATTGATAATCTTCGTTAATTCTCTTATCGTCATTTGT	imperfect
52	GGTAAAGGAATAGTGGATCTTGTTGCAAGGTTCATC	
53	CATAAGTCGCTTGTCATCGTGGACGGAAAAGGAAAT	
54	GATGGAAGAGCACTAAATTGTATCTGCTGTCCAATAT	
55	GGTGCCTTCCGTTCCCTGTCTTCTCAGGATAGACTAGG	
56	TTAGTTCTCCGGGAAAATACAACGTAACCCTGCAGA	
57	AACTTTGCTGTTACTTGCAGCCATTCCGAATATAATC	
58	TCATTAGGGATGTTGGAACTGTCTAACGCTTCACTGC	perfect
59	ACTATAGCAATCCCTACAGACGCGGAGAGCGGAACATAC	
60	GTATAGTCAAGTGGCCATGGGAAAGGTATCCGACACT	
61	TATTTTCTTTGACCGTCGATAAAATAGGGTTTTTAAAA	
62	CTAACAACTAAAGTAACATATGTCCCGATTATAATT	
63	GAAATATGAACGATCATAATATTTCCACAGTTTTACA	
64	GGAAGAATTTAAAGACAATGAACCTATCAAAATTGAAA	
65	GTTTTTCCGTATGGACTATTTAACATCATTTTATAA	
66	GAGCACAGGTCCAATCTGGAACGCAAATGCCTGGGCA	
67	GACTGAAAGTGAGGATAGGAAGCCACTCCCCGAAGA	
68	CCTATTCCGGAATCCATCATCCACGGCTTGTCTATC	
69	AGTAGCCATATTATGATGGAGTAGCCCCAGTACGGT	
70	TAAGTTGTCAGGAAGAGGAGGTGGTATATTCTGATA	
71	TCCTGACACTCATTAACAACAACATTGTAACCAGCATC	
72	GCAGAACAATTCAAACTTCGTCACCATTGAGGATGTCA	perfect
73	TGGATTGATCAAATGCTCCTGCGGTGAAGAAATCAGA	
74	AAAAGCACTTTCCAGTGTGTAAGCACCATTTCCACCC	
75	TATGCCCTCTCTATGTAGTCCATATCCATAGTAGTT	
76	GACAATCCACTGGTATATCTACATCATCAGTTAACA	
77	GATGACTGAGGAGATGCTGAGTATGATGCCCGAGGAAGA	
78	GCCAATGAACCATCTCACGAACCGGTACCAGAGAT	
79	GAATGCCGATTGTCTTGTTCTCCCTGTAAGTAATCG	perfect
80	GCAAATAGACAGTATACTTGAATTTATAGGAGATCTC	
81	GCTGATAGCAGCTGCAAACTTCTTCACAAGCATAATC	
82	GGTCTTCATTCTCTCTATTGACGTTTTCCCTGCTTATA	imperfect
83	GTAATACATGCATATATATAAACCTATCTGTCCTTTA	
84	TATAGTCAAGTGGCCATGGGAAAGGTATCCGACACT	
85	CAATTCCTGATACAAATTGAGATCTCAGCTCTCCTG	perfect
86	GATTGCTTTGGTACGGAGGTGGTCTCTCAGTCTGGTT	
87	GGTCCCCCTGTGGTACTCGAACTGCTCCTTGAATACCAG	
88	TATGGTGTATGTAGATCCCTGATAGCCAGTGAACGT	
89	GAATTCGTAATGGAGTGGTGACATAACCAGTTCCAAA	

90	GCGATGGATCGACTCAGATCGGAGTATCAGGACAAA	
91	GAATTGGCCAGATATAATTATCTAAGGTGAGTTTATA	perfect
92	GGAGTTGAGCAAGACGGAACTTTTCCGGCTTGCAGTG	
93	TGTATTTGAGCTTTGGCCTTCCGAGAATGTCCCTCT	perfect
94	TAATCCCCTTTCTGATTTCTTCACCGCAGGAGCATT	
95	CCACATCTACAAGCTCGTATGCTTCGCTCCACAGCCT	
96	CTACCACACTGACGGCAGATGCAACGGCATCA	perfect
97	AGCGATTCAATATATTCCTTCAAAATTGGATAATCA	
98	GTTGTAAGACCAACATCTTTGGCCTTTGATAGTATGT	
99	TCGTCTCCTTTTTGTTTGTTTGATTTCTCTTTTGTTT	
100	TCCTTCCTCTCCTGGTCAAAGGACTTCATGTATGCC	perfect
101	AATTATTGAAAGTCGTTCCAATGAACCCGGCCACATT	
102	GGATACTGGATATTCTATGCACCCTTACCATCCTGG	
103	GATTCATTTCCATCATACAGCAAGCTCTGCATATCTG	
104	TCATTTCCTTTTCCATTTCTCAATTCACTTCCATCC	
105	AGTGTCCAGGATCTTGATGGAGTATCTCCAGAACCAA	
106	GTAAAGAGAGTTGCTGCAAGGAAAGCAAAGCCTCA	
107	AGCTATATCGATACTTCAACCTACAAAAACCCGATC	
108	ATTATCTCATACTCGACATAAAGAAGCACCCCTGACATAA	
109	TCGTGCCTTCCTCTCCTTTATGGGAGCTGTCCTTCT	
110	ATCTTATAGTATGGCTGATAGCCACATGATAAGATT	
111	TTATAGTAAGAGTGAAATTACTATGTATTGATTT	
112	GGGCCATGAATCGAGGAAATTTGTTCTGTCCTCTACGAA	
113	AGAAAGGCTACGCAGGAAAGAGATCACTATTTTCTCA	
114	TTTTATAAATAAAACCGGCTCCGTAAAACGATGCTCG	
115	GGTCCAGGAAGCGGTTCGAGTGGGCAACCATGATCTT	
116	TAGCCCTGATTGCTTTGGTACGGAGGTGGTCTCTCAG	
117	AATATATAGAGGAATACCAGAATATTCCGTTATATT	
118	CTGATTGTTAAAAAACAAAAATAACAATTCCAATCATA	
119	TCCCTATCCATCCCTCAAAGCTGTACTTCAATTTTC	perfect
120	AAGACTGTTTTATGATGTCTTGCGATCTATAGATCGTA	
121	ACTGTGCACATAATGGGAATGGAATCCAGAAGGTCT	
122	GGACAGAAATCTATGCGTCCAGGTAATGGCTAGAATG	
123	CAAAGAAGATTATCCATTGTCGTTGAGGGCAATAGA	
124	GCCAGCGCTGAAAGTGAAGCCGGCTGGACATTCTCTG	
125	GTGTCGAAGAGTTCAGCTCTGAGCCGGGATAAGGAT	perfect
126	ATATTATCATCCAATTCAACTTCTATAGTGTTGAAAA	
127	CGGCATACGTTGAAGGAATATTTTGTTTGTCAGAATCCT	
128	TTCTAATCATTACCTGGACGCATAGATTTCTATCCA	
129	TATAGCTTAAACGTTAATGTCGGCTTTTTAAATTAGC	
130	TACTTCTACTAATGACACTACGGGTAAAAATAATA	
131	TTCTCACAGATGGGGCAGATCCTTCGAAGTAACTCTC	
132	TAATCCCAATCAAATGATATAAAAGAATCAATAACT	
133	TCTTGCGTCAATTGCCTTGACACATAAACATGCCTTG	
134	TGTATTTTCTCTTGCGTGTTCTTGGAGACTTGCCGT	
135	GTTTCTCTTCCTGTAAGATTCCTCTGCCCCTGGTAGA	
136	TTGTTTAAAAGAACTGAATAATTTTAATATTATTG	

137	GAAGAAAACAAAGCAGAAGATGGAAGTCAACGACCC	
138	GTAGATAAATCAATAACTTTTATTAGGATTATTATAGG	
139	TCTTTTTTGAGATTAGGTTAAGACCAGGGCATATGACA	
140	GCCCTACAGCCGACTTGAATAGCCTACTGGATATTTATTG	
141	TAAAAGAGAGTTGCTGCAAGGAAAGCAAAGCCTCA	
142	GATCCTTTCCTGCAGCATCCTGTATTCCTGGAGGTA	
143	GACGATTACAAGCTTTATAAAGGACTTCCTTAATGAG	
144	GAATATCACAAAGAATGGGATGGACATGATGTTGAGA	
145	AAATTAAGCAGATATCCTAATATGCCGCCGATGAACG	
146	TGTTGCCCCGATGAACCCTATTACGCTGCTTGTGCT	
147	GGTTAATTCGATCTCCTTCACGGAGAGCTGGCCTTTA	
148	TATTGTTATTTGTTTAGTCATCATACACATAATAAA	
149	CTGCGCTGCCCATGAAGAACAGAGGAAGGGCAAAAG	
150	GCATGGATAGGATATGGAATACAGGTCAAAATAATGATT	
151	ATTCTAATCCTCCTTCTATCAACACTTACGCTGTTT	
152	AGTTACTGTTACCCCCGATACACCGAATAAATCAGA	
153	GAAGAACATGATGATCCAATCAAGTGGAAGATCTTTGT	
154	GTTGTTTGTGTCGTATGTTTCAAGGCCAGATATCGCA	
155	AACAGCCCCAGGGGGCAGAAGCAGGAGATGCCAAAA	
156	GTCATCAGGGAAGAGATATAAGCACAGTTTCAGTT	
157	CTATGGCTCCAACGAACACTATGATGCCCAGGATAACC	perfect
158	CATTTTTCTCATCATTGTTCTTTTCCGATCGACGAT	
159	AGAGAGAGCCTCCAGTATGATACTGTCAATTTCATCA	
160	CAGGAAAGGAAAGAGACGCAAAACTGCAAGAGTGAAGG	
161	TCATCTTCGGTCAAAAAAATCAATACATAGTAAATT	
162	TCTATAATACACATTTTAAATTCGTCAGTTCAACCG	
163	TTCGGTATTTTGCGAATAAATCTTCAACTTTTCCATT	
164	GTGATGGCAATCCTGAAGGAAGACCTGGAGCAAATC	
165	ATTATTACTTTTCCCTGACTATAGGTCAGAGATGACAA	
166	TAACTAAAAGGGATAGAAAGGTAACAAAAGAATATAA	
167	CAGCTCAACGAATTGAGATATCTGGTAAGAAATATT	
168	CAAGAAATCCAACACACACACACATGAAACAAAAAA	
169	ATGATTTTAAATAAACTTAACCTGTATACTTTAGGAG	
170	TGAATAACTATCGGATAGTCCACTAAAATAATAACA	
171	TATCAAAGGTTATCACTTTATTAAAATGAAATAAAT	
172	TCGTCCAGCTGCTCCTATATTCGGTGTCCAAAAAATA	
173	CCATACATTTTCTGTATTTCAATCAATTTTTTATTT	
174	TTCCATCAATAGGGAGGGGTATACATCTTCCCCCGTA	
175	GATGATTTTAAATAAACTTAACCTGTATACTTTAGGAG	
176	AGTTGAATTACAGGTATTACCGAGCTTTGAGAATGT	imperfect
177	GTGAACTACTCCGCCGGAACAATTGATTTCAATATT	perfect
178	GAAACTCCTCAATCGCTTTTGAATGATGTCCTGATA	
179	ATCATCAGTGCCTTCCAATTCATCCGCCATCTTTTGC	
180	TACGGCACAAGCCAAGTTGAATTACAGGTATTGCCG	perfect
181	AAAGAGCATTTCTGTGCAGCAAAATATCGATTATATC	
182	CCTATAAGGTCCACATAGACCTTATGCTTAACATTC	
183	GAGAGTCTGCTTCTTCTTGCCTACATCATAAAATATG	perfect

184	TTATGAAGCTGATAAGCCAGTCATTAAATCTGCCTTAA	perfect
185	CCAAAAAACCAGAAAACAGAGCAAAACTCTTACAA	
186	TCGTACTTACTCTGGCATACAGCCCCACTGTAATTCT	perfect
187	GATACGGTAAACAATTTTATAAAATGATGTTAAATAG	
188	TAGACATTAAATGACCCATTATTATAATATGTGTTAT	
189	ACGCTGAACACCAATGTAACATGGTCCTATTTCAACA	
190	TGATCTATCTGTTGGATTAACTGCCCCTGTTCGGAAG	
191	TTTGTTGTATAATATAATATTAATGAATTATTCCA	
192	GCGATTTGCGGTCAAATAAGATTCTAATTGATATTTT	
193	ATAAAAGTCAGTATCCTTCAACACTGGCTCTACTGA	
194	TGAAGACTCCTGTTCCCACTGTGCCTGAGGATCCTGG	
195	GAACTTAGGATTTTGATTGATGTAATAATCGTAACG	
196	ACAAACTTAGTAGTGGTTGTTGTAGATGGTAATGCA	
197	AGATGTGGACACGAAATGAAAGTGATGAGGATCATTT	
198	ATGGTCTGAAATGATGAGGCTTCCTGAACCTGATCG	
199	TAAACATAGGAACAGGATCAGGAGCAGTAAGCGTTTT	
200	TCTATATCCCTGATGAGAAAGAAATTGGGAAGTTTA	
201	ATTCACTTCCTGGTTGAACCTGGAATTCGGATTTTT	
202	CTTGTTTTTGTTCATCGCCATCTGGGATCTCATTTT	
203	GCATAACGTTTCTTCGCTCTTTCAGGGATAGCCCTT	
204	AACGTATAAATATGCTGTTGTCTTTTCAGATAATCTT	
205	GGAAAAATATCCGACACGCGAGTGCCAACGGAAAAGGATC	
206	GGTAACTGACCCAGTCGATATTTTGCGAATAAATCTT	
207	TGCGATTACAAGGGGCTTAACCCCAAGATACGACAGA	
208	TGGAGATAGTTATCCAGAACACAGGCGACACCGCAT	
209	TGTCTGCTTTGAAGAGTATTTTTGTTGTGGGATTGA	
210	TCTCAATCTCAATCAAGTAATTTTCTCCGTATTAT	
211	CTTGTCTGGGCATCAATGTAATTCCATTTGGGCACC	
212	GAGTTAAATATGCTTGTAGAGTTGCAATATCGATATC	
213	TAGCAAGGAAACTGTGTTATTGGCAAGACGTGCATA	
214	ATTAAATACAGTAGCATATGAGAACATATCCACGTT	
215	TACTATATACTATATTATAAGGTAGTATATAAACTC	
216	TCTATCTCTCGTATTCTATCACTATGTATATCCGCA	
217	ATGGTCTGAAAATGATGAGGCTTCCTGAACCTGATCG	
218	ACAGAGATTATCAGCAGGAAGCAGACGCATTGAATAA	
219	CTATATACTCCCATTCCTTTATTATGTATTTTCTCTG	
220	AGTGGTATTGCATTAACAAATAGATCAAACATCCACA	
221	CTAATTCTGTATATCATTTTCCACCTCCTTTTCTCTT	
222	CTTTCAGCGGGCCCGTTACTGGTGTTTCGTTGCTCTT	
223	GTTATTGGATTGAGTCCGGGTGATGCAGGGGTTTTCATAT	
224	GGATGAAATACCAGATGCACCTCTCCGCCAGGTCCAT	imperfect
225	TAGATAACCTAAAAATAGTTTATTCCTTGGCACTGA	perfect
226	CCAGATCAAGAATGCGGAGCAGGTCTACAATACAACA	perfect
227	GCGTGCAATGGACGAGATTGAGGATTGGGATATCTTC	perfect
228	TGCTTGGCCCTTCTCGGTGTTGTTGATGGTCTCCTGC	perfect
229	GATTCATTATATGAAGTAACGGTTATTCCATTATAGG	perfect
230	AGCATGCTGGACCGACGGAGTGAAGGTCTTCATTGGG	

231	GAATGATTGTTCTGTCTGGATATCTCCTGATTCGTG	
232	AAAATAATAAATAGAGACACATAGCAAACAATTGCAT	
233	AAAGGCTAGGAATCCGACAGCCCCCCACCAATATAAT	
234	GGGCCCTCACGCGCGAAAATCTCCTCAATAGAAAG	
235	TATCTTCTCTAAAGAAGAATCTAAAGAGCAATCGAAA	
236	TATGCCGTGTCACCACTCGTTACACTATATATATTCCA	perfect
237	GGATCAGGCCTATGCCCATTGTCCGGGTGTACTTCTG	perfect
238	TATATCTGATAAGAAAAGAAGATCCAGAGGATGCT	
239	GTGCATAGGTATGAATGACCGTGTTTCCCTGCAATA	perfect
240	GATTCAGGCGCTTGTTCATTGTCTGGTCATGCTTTA	perfect
241	TGTATTAGCATAAAAAGTAAAATAAAGCCTTTAAACCC	
242	AAATATCCTGCGTTTTCCCTCTATACCTTCTACATCTTT	
243	CTTTATGATACTTCCCACAAGATCCATTCTTGTACT	
244	GTCTAACGGTATGATCGGAAGTGTGCAGCTGCTGGT	
245	GATGTCTTGGACGCTTTGCATAGAGTCAAGATTCAG	
246	CATTTCCATATCTTAGGAGATGGCTGGCTATCTCCCA	
247	TGCCACCCAGAGGCTGAACTATTCGATATATGTAAAA	
248	GAAAGATGAGATATACCATACAGCACGCAAAAAAAGA	
249	TCTATTTCCAGATCAGGAAGCCTTACTATTCTTTTA	
250	TTTTACTGGATTTAAGCCTGTCATAGACATACTGGAT	
251	GTAACTACGGAAGAGGGCTAACTTCGCCCTGAACATCG	
252	TGATTCTTTCAGTGCTTCAGAGAATTTGGAGTTAAC	
253	TACACCTTATCTCATACTTCGGAAAAGCTTCCATCA	
254	ATCTATTTACAATTGCAATTGATCCAACGATCCCATC	
255	GTCTTAAACGGTTCACATTACATCGATGTTCAGATA	
256	ACGATTCGGCTGAGGAGATTATACCTCCCAGAATGGA	
257	CATAAGTCGCTTGTCATCGTGGACGGAAAAGGAAT	
258	GGATTCTTCTCTGAACAAGGATTCGAATACTACTTCG	
259	CGTATTCCTGCATCCTTCGTATGTTTTCCCGACTGA	
260	AAAGAACCGATTATCACATTTGGAATACCCAATATC	
261	GAAGTCTTTCATGGAAGTCCATTTTTCCGGGAGAAC	
262	TTTTTAATACGGATATTTTTACTCCAAAACCATGCA	
263	CCAGCTCAATTTCCGATGTCAGGATAGGCATAAAATTGC	
264	TTTTTAATACGGATATTTTTACTCCAAAAACCATGCA	
265	TACACACTATTTAACATCATTTTGTAGAATTGTCTA	
266	GGACAAAACAATAACGAGAGCAACGATGATCGTGAGCC	perfect
267	CTGGTAAGAATAACCTACAAGGTTGAAATCATGCAC	
268	GATGAAATACCAGATGCACCTCTCCGCCAGGTCCAT	perfect
269	GTCATCAGGGAAGAGATATAAGCACAGTTTCAAGTT	
270	AAAATAATAATAGAGACACATAGCAAAACAATTGCAT	
271	TATGTTCTATAGTAATCTTTTACGTAAAGCTGTAGATC	
272	GAAAGGCTAGGAATCCGACAGCCCCCACCAATATAAT	
273	GTTGTATGAGTATGCGGATCCGCCCAGTACTGAGGA	
274	GTTGAGCAGAATCCACAATACGTATCCCAGTGGGCA	
275	CGACTTTCACAGAACCACATCAGTCCTTCCTCGGTCA	
276	GCTGCGAAAGAGGTTGCATCTAAAACAGGAAGGAGCCT	
277	TAGGGATATGCCAGGATACCTGTCCTGTAGGGAGA	

278	CATTTAAATCTAGAGCATCCGTCATTGCAGCTGCAA	
279	ATGGGCTTCAGGAAGCGATAACCATCCCAGGAACAA	perfect
280	TATTCTTTATTATACGATATGTTGGTCCAATCTATT	
281	CATACTGTGAATCCGGATAGGTTGAGAAAATAACTAC	
282	GATTAACTCTCCAGTTCGATCCGGTTGAGAAAAATCT	
283	GACTCACATTGAGGAGAAGCACCCGGAGAATGTAGTA	
284	GATAATCTAGGGGGCGTGAACAAAGGGGATACAGTA	
285	TCTAAATATATTTAATTTTGATTGCTCTTTTATCTTCC	
286	GAAGGGGATTTTTAAATATATTAATAATTGTTTTATTA	
287	GTAAATTACCCATACAGCCACCCTCCAGTCATATCA	
288	ACTAGGAAAACCGGAAAATTCCGGAACTCAAATTTTAC	
289	GATTAATTCTTCCGCAGTAGATGTTACTGTCAATCCT	
290	TCTACATCTTTTCTTGTGGTGCATGGGATGGGCAGGG	
291	ACAAACAAAATATTTGATAATTTGTGCAGGATCTTGA	
292	GATATCCCTTATCAGAAGGAACAACAACCAGATATT	
293	GATAACCCAATTGCCTTTAGCACAACAACATTTAACCTA	
294	ACCAATGACGATTTCAACCACAATGTCCTCATCGTAG	
295	TGTATGCAGAGCCGCCGATTGTCCATGAATAAATAA	
296	GAAAATATCCGACACGCGAGTGCCAACGGAAAAGGATC	
297	GTAACTGACCCAGTCGATATTTTGCGAATAAATCTT	
298	GATCAATATCTACATAAAGTTCAATACTCATTATTAT	
299	TCATATGGTTCTTTGAGAGTTACAATCACACTTCAAAA	imperfect
300	GTATTGGATTGAGTCCGGGTGATGCAGGGTTTTCATAT	
301	GAGAAAGGCTACGCAGGAAAGAGATCACATTTTCTCA	
302	TAATTGCGTTTTGAAACACATAAAGGGATTTAGTAGT	
303	TATAGAGTATGAACTGTATAAATATATGCTTAGCCA	
304	GTCCCCGATAATATGGAAATGCGGCCCATTGCTGCAGT	
305	GTAATTTTAGTTATTACATTATTGTTTAAATTCATTT	
306	TTATGTGTTGTGATTATATCATCAATTTTATGTTTA	
307	TCTATCAGGACGAACTCCCCCATCACTCCCCATTGATT	
308	TATTTTCTTTTGACCGTCAATAAAATAGGTTTTAAAAC	
309	AATCTCAAAAAAGATGGCAATAAAGGAATCTTAGAA	
310	CTGGAGGAACTAGCTCTTTTTGTAGTTGCCATTAGT	
311	GCCAACCGAGTAGATCAGCTTTATGACAAGGTATACT	
312	CTTTCAACAAGCTTTCCTTTTACATATATCGAATA	
313	GGAATGTGTATCCTCCGGATTGCGTTACTTTCACA	
314	GTGAAACTACCTTGATGTGATCTGTCCCGCAGAGACCG	
315	GAAATCCAGTGTACTTTCCAAATAGGCAATATCGGA	
316	TTTCTTATTCCTCCAAATTTATTGTTTTTTGCT	
317	GTAAGTTTTTAGAGTCCTTAATTATTCAATATATTCAACA	
318	GTACAAATATATATATTACACCGCTAATCGTTAAAAA	
319	GTAGATAAATCAATAACTTTATTAGGGATTTATTTATAGG	
320	TTCCTGGTTCATCAACCCACGTAATGTCAGCATTAT	
321	GAACCAATCTTTTGTCTTTTTATATTGCATTAACACGT	
322	ATAAGACAAACTCTCCCATCACACCCATTATTGAGA	
323	ATCATTTCATTTTGCAGGTATGTGCCATGAGACCTGA	
324	GAATTTATCCCGATGGAATTTTGCAAAGAAATATAATAAA	

325	GGGTGTTATGATAGCCATAAACGGAAGACCTAAA	
326	GTTTTCTGTTTTACGAATCTCTTTAGAATTGTCATTCT	
327	TAGGTATAAGCCCAAAGATAAGCGAATCTGACCTTT	
328	CCATTGACTGAACTTGAATATCGTGGTGAGCTGATGA	
329	GGTTGTATGAGTATGCGGATCGCCAGTACTGAGGA	
330	TTGAGCAGAATCCACAATACGTATTCCCAGTGGGCA	
331	GGTGTTATGATAGCCATAAACGGTAAGACCTAAA	
332	GTTTTTCTGTTTTACGAATCTCTTTAGAATTGTCATTCT	
333	GGCCAGTCAGTCTGAAAATGCTGTCCTTGCTGAATCA	
334	GTATAAATCGCTGAGCTTTTGAAGAACTGACTTTCT	
335	CTCCTAAATTTCTTGAGAATGTGAACCACGACCTGA	
336	TATAAAATGAATTATAGAGAAAAATCGTATATGGCAA	
337	TCTAAATATATTTAATTTTGATTGCTCTTTATCTTCC	
338	CATTGATCAACATAGATATGGTTCCCTGGAACAGAT	
339	TGCATTTTCCAACCAATCTAAGCCCAATAATGAAAT	
340	TGTTATGTTAGTATTCATGACTGGGTTCGGCTCGATT	
341	AAGACTGTTTATGATGTCTTGCGATCTATAGATCGTA	
342	CTACATAGTCGAATACAGCTATACTTGCAACCAGGA	
343	GACAAACGTCCAGAGAGGACTTCTGCTTCGATTCATGG	
344	TCGATCAGATGAATGAGAACCTGAAGCAAACAAGGA	
345	GGTTACTAGAATCCGTCTTCGTCAGTTCCACTCTTAA	
346	AAGATAAATAATTTTTAACTGAAAAATTCATTCAAA	
347	CCTCAAAACTGTACCCAAGTCCGGCTTGGTCGTTTGG	
348	TTAACCTGGATAATCTGGTGTCATCTGCGTCAAGAG	
349	GACAGGTGCATTTATTCTGGTTTGCAACCAAACTGACA	
350	GGAGATAACCAAAATCATAAGGTCCGACGATCAAATCCG	
351	ATTCTTTCCTAACGTCCAATTTTCAAGAGAACTGATA	
352	GCCCATGTCATCTTTGAGGTTGTGTTCATGTTATCCT	
353	TATGTACCCATTGCCTGGACTGCCTCAGATTTATTT	perfect
354	GGTTGAAATCGTCATTCGTTGGAACTGATCTCAATTCT	
355	TAGTGGAAAATGCACCTCTTACATTTACTATACTTC	
356	TCATCTATTCTTACCAAGATTAAAGATTATATAAAT	
357	CCAATGAACCATCTCACGAACCGGTACCAGAGATG	
358	TAGTGGAAAATGCACCTCTTACACTTTTACTATACTTCC	
359	GTCATCTATTCTTACCAAGATTAAAGATTATATAAAT	
360	GCCAGTCAGTCTGAAAAATGCTGTCCTTGCTGAATCA	
361	GGTATAAATCGCTGAGCTTTTGAAGAACTGACTTTCT	
362	CTCCTAATTTCTTGAGAATGTGAACCACGACCTGA	
363	TGTGTCGAGAGAATGCTTCCAGGGCTGCAAGCTTGA	
364	GACAAGACGGATATCAGAACTGGTACCTATTTTTACCG	
365	CATCTATTGCTTTATATATTTCAAAGGACAGTTCAT	
366	CAAAACATATTCTTGCAAAATCCCTCTTTTCAACACC	
367	GTGCCTTCCGTTCCCTGTCTTCTCAGGATAGACTAGG	
368	ATTGATAAGTGCTTACAGTATATTGCGACCAAGATTT	
369	CGGCTACACTCTCATGGGAGTTGTCGTTTTTCTTCCT	
370	ACAGGACAGGAAGTGGTTTTCCACGCCTTCAACAACAT	
371	CAGTGAATGTTGCCGTTGGGGGCTTCGGAGAAAGAGA	

372	GATGTGAATGCAGGATTGCCAGTTCGTGAAACCTCA	perfect
373	TTCGGTATTTTGCGAATAAAATCTTCAACTTTTCCCATT	
374	TTCTGCCGTTACATTCGGATCCGGGAGAGGCATTTAG	perfect
375	GAGATTTCTGATTGCTGTGTTGTAAGGTTCTGGAGGC	
376	GCTGGTATATCTACATCATCAGTTAACATAATATGATA	
377	TATCTTTTTTCACCTCTTATGCCTGGCCCAAATATGA	
378	CTTGAAGTGTGCAGCTTCCATGTCCTTTGCTTTTAG	
379	TCGTTTTTGTAGTCTATGGCCTTTGCGAATTCACTC	
380	AGTATTTCCCACCACTACCAGCCTCCAAAATGTCAT	
381	AATACTTTAGATGACAAAATCGTAAGAACAAACTTT	
382	GTAGTGGGAACGGTCGAAAATGCCCGCCCTTTTTTTT	
383	TTTCTGAGATTCCTTCTCTTTTTGTCTCTTTATTCATT	
384	CAAAACATATTCTTGCAAATCCCTCTTTTCAACACC	
385	TCGTCTCCTTTTGTTTGTTGATTTCTCTTTTGTTTG	
386	TCCTTCTCTCCTGGTCAAAGGACTTCATGGTATGCC	imperfect
387	ATGATCGATATTGACACAAAAAGCATGAAGGTAATG	
388	CAGCATCTGGCGGAACAGGTTCATTTGCGTATGCA	
389	CAGCTGAAGCATCGGAGGTTCGTAAAAGGTTCATGACA	perfect
390	GCCTAATAAAACTGTTGCTATACATTACATCTACATTA	
391	GACATCTGTAGTAGTCTTGTTTGCAGATATGGCAAGA	
392	GGATTGTTTTTCGATTGTTCTTAACGAATCCACCA	
393	CCTTACTGGCCAGAACCTTAGGGACTTCGAGTTATTT	
394	TAAGAGCATTGACAAAGTATGGGTCGAGTGAACATGA	
395	TGTCCAAGCAAGGAAGGTCAGTGGAACCACATCTTGG	
396	GCATCCCTATCAGCAGAGAGTTAAATCCTGAATAGT	
397	GCGGGACAGGGGCGAATGCTCGAAGATCTGAGCAATC	
398	TCCGTATGCAACGCCATTATGGAATTATATAATAATA	
399	TGGATCATATGAGATCGTGTCGATGCACCCCAATAT	
400	CTTACAACCAATGTTATATAAGAACCAATTCCAATC	
401	TGCCACCCAGAGGCTGAACTATTCGATATATGTAAA	
402	TCGTCACTTGACTGTATGGACTCACCGTGAGAACTTG	
403	TTTCTGCTGAAGTGACCGGAGGCACATATCCGTTCA	
404	TACGATCTGACAGTTGTATTTCTCGATCTTATATGGA	perfect
405	GCAGTACAATGCATCTGAACTGTACAGGACATTTATG	
406	ATCAATAACAGGAAATATAATACGGTTCAAAACTTAG	
407	ATTCGTCATAAAAAGCATCAGCCCGTCTAAGATCAAA	
408	GGAATCTTGGTTCTGGAGATACTCCATCAAGATCCTG	
409	GTGAATTGGCTGTCCTGAATTGGAATCAAGAAATGTC	
410	GAACAGATCCTGGAATCTTGGTTCAAGTGATCTACCTA	
411	GTCCACTAACACTAACACACCTGCGATGATAAAAATGA	
412	ATAGGGAAAAGTGTGGAATGTGGAGGGAAAAATGA	
413	GCTGTCCTGTTCTCTTTTGCTCGATCTCAGTCCGCCTC	
414	TTAATATCGAATTTTCATTCGATAGGCATCACTATGT	
415	GACAGTAATTGGGCAAAACCGAACGGCAGAATTGAAA	
416	GTAATTTTAGTTATTACATTATTGTTTTAAAATTCATTTT	
417	ATTATATCGGTTACATATTTCCAGTTTTCCTTAATATA	
418	CATGTACCGGCGTTCCTGAGAGATTGAACTGCGACA	perfect

419	TTCATCGATAGGGGAGCTCCTTATCAAAAATATTCCTC	
420	ATAGTAAATGATGCAGGGATTTTAACTTCTCTCCCC	
421	CAGACTATTGATTACATACTCTTTCCTGAAATATTC	
422	TAGGCCTATCCGAACTGTCTAGGGGATTCATTTTTT	
423	TTTAATTGCGATAACACAACAAAAAGGTGTACCGTT	
424	TAGAACTATCTGTCCGTCTACTCCACTCATTTGAGG	
425	TCCAGATCTGAATCTTCTGATATGGAAAATAATAA	
426	CGGGGTAGGTGATCGATCGTTCTTTGGGTCGGTCTGGGC	
427	GGCAAGGAAACCTACCAAAAGAACAGGGTTTCCAAGGA	
428	GTAAAGGAAACCGAGAGGGCAGGATAGATAGTGAAGT	
429	TCAGCTCGAAATATTTAGGTTCTATTTCTGACAGCTC	
430	TCCAATAAGACAAACTCCCCCATCACTCCCATGATTG	
431	GTCATAGACATACTGGATATTCTCAGGATGTGCCTCC	
432	GTTCAGGAAGATACTGTATCTTTTGCCCTTTTGGAG	
433	ATGAGCTCTTTGTATTAGGATTGAAAACAAGGGATGG	
434	TTTCCTGGTTTATCAACCCACGTAATGTCAGCATTA	
435	ATGGACATCTCTGCGATTGTCAGGGCATGAATGTCA	perfect
436	ATAAGTTTTTAGAGTTCTTAATTATTCAATATATTCAA	
437	GTCTGCTTGGAGGTACACTGGGGCTACGCATACTCTGCCA	imperfect
438	GTTTAAAGTGTTCATTGTAATTAAGAGCGTGATATTT	perfect
439	TATGGTTCCGCAGGATGGAGCAATGGACTGGGATCG	perfect
440	CTTTTGTTCTTCAACAGGTGCAGGTGATCCCTTGAA	
441	ATGCTTCCACAGGCATCAGTAAATGATGCGGGAAATT	
442	GACAGTAATTGGGCAAAAACCGAACGGCAGAATTGAAA	
443	GATCAGTTCTAACTGCGTTATTCTCGATCTATGTTGA	
444	GGACAGAACAATCATTCTGGTGTGCAATCAGACCGAC	
445	GATAATAATTGACCGAACCCGAACGGTAATATTGAGA	
446	GAAGCATGGTACCTGAATAAATCCGTCTATGTGATA	
447	GATACAGCAATTCCTTTTAGTTATTCGTTCATATTTTTT	
448	GGAGATATTGCCACGTCCTCTGGTCCATTCTCTCCTG	
449	GGGAGCGCAAATATGAGTCCTATTATTGTGGCTACCC	
450	GTGGGAGCAGAGGACAATACACGCGTGTTTTGGAAA	
451	TGTTATTTTCTTCCCTTTGCCTTTTTTGTCAAATTCC	
452	CAGGAAATACTGATCTCTGGAGTCGTCCTACTTTTAC	
453	TCTATTCCGTCCACTAGCGCACCTGCGATGGATAAAA	
454	GTCGTAAGTGCCCTCCTGCGTAAAATCGTCCGCCAGGT	
455	TATCTGTGACAGTCACGGATTCTCTCGGGGGCATCCTCA	
456	CCAGGGTTTTCTGATTTTCAGGCCTATGAAAATCTCT	
457	GACAATCGGATCGTCGCATCAACCCACCATAATCCC	
458	CCGTACTCGGCTCATATTCCGTGATGCAGTTTACGT	
459	AGTACCTCATTCTCGGGAAGCAATATGCCCTCATGAC	
460	ATGATGCCCTGGACTTCCTGGCACTCGTCGGAGTTGT	
461	AATATCAGGGATACGATGGCGACTGCAACATATCC	
462	GATCTCGAACACTTTTATATCTCCCTATCCCGATATA	
463	TTTGGGCCGGGGTTTAAACTTGAAGGGATTTTTAAATAT	
464	CTGGGACTGGATTGGCAAATGGCAGGGGATTATATA	
465	TCAATATCTAATTTGTCAATTCTTTTGTAATTATT	

466	CATAACGTTTTCTTCGCTCTTTCAGGGATAGCCCTT	
467	GACCGATTCCAACTCTATACTTGATTGTCATGGCAGG	
468	TTCACCTCCGATTCTTTCAGACCTGCCAAAAATTCTTC	
469	GACAATACCACTGAGTTGTCGTGAGATGATACGATT	
470	GGACAGATGGGACTTTGAAATTGAGGAAGAATAACTA	
471	ATCATAAGATTTTTTAAGGTCGTAGCGGCATTTATC	
472	ATCAGTTCTAACTGCGTTATTCTCGATCTATGTTGA	
473	GCAGAACAATCATTCTGGTGTGCAATCAGACCGAC	
474	GATGCAATCCCCCAAAGAAACCAGAAAACAGAGCAAA	
475	AACCATGTCACGCTTGACATCGAGAGATTGGTACTT	
476	TGTCTACGTACAATCGATCTCTGCCTTCCTCAAAT	perfect
477	TCAATATCTAATTTGTCAATTTCTTTTGTTAATTATT	
478	CTCTATGTAGTCCATATCCATAGTAGTTTTTCCTGA	
479	ATGATCGATATTGACACAAAAAGCATGAAGGTAATGG	
480	GCTTTCTTAGGGGTGCAAATTTTTCTGTGTTCAAAAG	
481	CTGATCCCCTATTTACAACTATATTCTCGGTATCATT	
482	GATAATGTAGGGGGTGTGAACAGAGGGGATACTGTA	
483	TCTTTCTGATTCGATTGCTCAGATCTTCGAGCATTCG	
484	TCCTCTTGCTCCAGCACTATGCCTTCGTGGTGCTCC	
485	ATTCCGTCCGAATCAATTCACCTGATCCAATTCAGA	
486	CCTAATAAACTGTTGCTATACATTACATCTACATTA	
487	ACATCTGTAGTAGTCTTGTTTGCAGATATGGCAAGA	
488	GATTGTTTTTCGATTGTTCTTAACGAATCCACCAGG	
489	ATTTTTATCGCAATGGAGATCGATAAGGCAATATAATT	
490	AGCAGTGAGGGGGAGCCGTCACGTGACGGAAACTCCC	
491	TGAGGGCAACTGGATCAGGCTGGAACGAACCATCCTGA	
492	GACAATCGGATCGTCGCATCAACCCACCATAATCCA	
493	GAAATCCATCGGGGAATTCCAAACCCTTCGAAATGAGA	
494	ATAAACGCTAAATCCTGCAATTGGACAGTATTGTGT	imperfect
495	AATTACATATCGGTAGATCCAGCGGACATCATTTCT	
496	TCATATGGTTCTTTGAGAGTTACAATCACACTTCAAAG	imperfect
497	TATTGGATTGAGTCCGGGTGATGCAGGGTTTTCATAT	
498	TAGATAACCTAAAAATAGTTTTATTCCTTGGCACTGA	
499	TGACCTTGCCTGGGCCGTCTTTCTGAGGTATTCCCA	
500	GATGATTCCAATGAGAACATTTGCGATCATGACCATTG	
501	GTCTTTCAATTTATCATATTCAGATTTTAGGAATTTCC	
502	GATTATCGATTTTCCGAATTAAACGTTAAAAATTTT	
503	TGAAACATACGACACAAAACAACAGGATGCTCAAAAAA	
504	TTTCGGACTCTTCTGCCCCGGATGGCTGAGCTTCCATGC	
505	ACTATCTCATCCGCTCGGTATTCCCAATGCCTCAGA	
506	AACGTATAAATATGCTGTTGTCTTTTTCAGATAATCTT	
507	ACCGATTCCAACTCTATACTTCGATTGTCATGGCAGG	
508	TTCACCTCCGATTCTTTCAGACCTGCCAAAATTCTTC	
509	TCTCTATTTGCTTCCAGCACACCTATCCTGCAAAA	
510	GTAGAACTTTAGATCCGTCCTGATGGCCAATATAGAT	
511	GGAATATTTGTATAACAAATAGTCCGATGTTCGGCTGA	
512	GGCCAAAGGATGGAGGGTTAATCAGATCCTTCCAATT	

513	GAGTCCTTTCTGGGTCAGACAGGAGCAGCTGTTATTCT	
514	GTGGGGCAACAGATTGGTTTTCGTAAGGAAAATATTA	
515	GTTCCCACTATGACGCACCTCTTTATTTTTAAGAA	
516	GGTTAACTCCAAATTCTCTGAAGCACTGAAAGAATCA	
517	CATTCCTGAAAAACATATTTTCAGTTTTGATTATAT	
518	TTCTATGAACAAGGAGAAATCGCATACCTCCTCTTCA	
519	GTATCTTCTCTAAAGAAGAATCTAAAGAGCAATCGAAA	
520	TTCTTCCTTGTGCTCATATTATCACTAGTTTAGCTTG	
521	GGACAGAAATCTATGCGTCCAGGTAATGGCTAGAAT	
522	GCAAAGAAGATTATCCATTGTCGTTGAGGGCAATAGA	
523	GCCTTTTTTTTTGCAATTTTGAAAGTCTTCAAAAATGC	
524	GGTGGCTTGGTAGCTCACTGCAGGGCTTGCTGTTTCG	
525	TTTTTTGTTTCATGTGTGTGTGTTGGATTTCTTGG	
526	TTTACGTCGCATATTATACTTTGATTAAATGGTAAAA	
527	TGATACACTACACGTGCCTTGCACTCTGCTGCATGCG	
528	ATTATTAGCTTTCCCTGGCTATAGGTCAGAGATGAC	
529	TGTATGGCCTTCCATCCTCCATATACACCTATAGTACT	
530	TTAATTTTTTTATGATTTTTCATTGTCTTTTTTAAAATT	
531	CTCTTCCTGAATGCCTCCCTAACGACTCCAGAAAAC	
532	GTATAACATGCGTGCACGGGAAGCCGAACAATACGATC	perfect
533	CGTTTTGAGAAATGGCAAACGAAGCAGATATTTCAG	
534	GTAATATTGATAAGGGGGAGTGCCGTAAGAACCATA	
535	CATGTACGGCGTTCCTGAGAGATTGAACTGCGACA	
536	ATCATTTCATTTGCAGGTATGTGCCATGAGACCTGA	
537	GAATTTATCCCGATGGAATTTGCAAGAAATATATAAA	
538	GGCAAATTACTGCTTCTGCAAGATGTAATTATTTTTGA	
539	GACTCTTTCAGGCCGCCTTGCACCTTGAAGAAAGAGA	perfect
540	TATTATTCGATTTATTTCTGTGGTTTTCCCTCTTCT	
541	TAGATAACCTAAAAAATAGTTTATTCCTTGGCACTGA	
542	GAACAGCCCCAGGGGGCAGAAGCAGGAGATGCCAAAA	
543	GGGTCTGTGGAAACAGATTTTGGAGGCTGGAAAACA	
544	GGTAATTGGACTGGGTATCCTCGTCCTATTCTGAGG	
545	TATCTTCTCTAAAAGAAGAATCTAAAGAGCAATCGAAA	
546	GTCTATCTCTCGTATTCTATCACTATGTATATCCGCA	
547	GATGGTCTGAAATGATGAGGCTTCCTGAACCTGATCG	
548	GTCTTTCAGACATGTGGGGATTCATTGGCTCACTACCT	
549	GATATTTGATTCATGATTTTGATAAAATGAAAACTATAG	
550	GATATTCCTTGAATTCCACTGGCAAGGAGAGAAAATGAC	
551	CCAATACAGGGACTGGATGTTGTAATTTCAGATACA	
552	GTCATATGGTTCTTTGAGAGTTTACAATCACACTTCAAAG	imperfect
553	TATTGGATTGAGTCCGGGTGATGCAGGGTTTTTCATAT	
554	TAGATAACCTAAAATAGTTTATTCCTTGGCACTGA	
555	AACAGCCCCAGGGGGCAGAAGCAGGAGTATGCCAAAA	
556	GTAATTCACTGCTTCTGGCAATGCTTTGCTACACTG	
557	CTAATTGATATTGTTCATTCGGTAGATAATGTTTTA	
558	TGGAGGAACTCAAGAAGAGTCCTTCCCCTAAAACAAG	
559	TGCTTATATCTTCAGTCATATTCCCTCTTCACCTCCT	perfect

560	GAAGATGAGATATACCATACAGCACGCAAAAAGA	
561	TAAACCCGACAGACATAAAAAGCAAGGATTTGCAGA	
562	AATACATTGGCATTTGAACGAAGGAAAGCAGAAAGAG	
563	TAATCCCATCGTCTATATGTATGACTGGAGGCAATCC	
564	GAGAATGTGAATGACGTAATCGGCAAAGAAATAAGG	
565	CTTAGAAAAAATCTCCATTGTGCTGCAGGAATTCCC	
566	GTCTCAATCTCAATCAAGTAATTTTTCTCCGTATTAT	
567	GGATGCAATCCCCCAAAGAAACCAGAAAACAGAGCAAA	
568	TGTCTACGTACAATCGATCTCTGCCTTCCTCAAATG	perfect
569	AACATAAGCGACCGGAATATAGACCGGTTGAAAAGG	
570	GCTGGAAGCTGATTTCGTTTTTGTAAACTCACCATAACA	
571	AAGAGATTCGAGAGAGAGAAGTACGTCGAGAACAAACG	
572	TTTTAATTCGACAATCAAACATACGGTTGACCCTGT	
573	GATATATTTTTCTGCAGTATTCGTAAGATATAATCA	
574	CAATGGGCATGATGTATGATGTAAGCATTGTCTGGTCT	
575	TAGATATATTTGTCCGCTTACTCCACTCATTTGAGG	
576	TATACGATTTTCTCTATAATTCATTTTATAAGGTAA	
577	TTTGATTTTAACACCAAAAAATGCCCATATCAAAAC	
578	GTCTAATGACCTTGCCTGATAGGTCGAATAGCTGCTG	
579	GTAACTGACCTAAGCGATATTTTTGAAATAAATCTT	
580	CTTTCTCCCAGGTGATCTCATGGTTCTCCTTTGCGTG	imperfect
581	TGGGCCAACAAACTCGGATTGGAGGGTACAAATCTT	
582	AACGTATAAATATGCTGTTGTCTTTTATATACTCTT	
583	GATCCATAGAGGTATATCTCCCAGCAGGAGAAAACT	
584	CTAGCACTTTGGCATAAAGATTATAAGTGCTTCCATTC	
585	GTGTCATAAAGCTGATCTACTCGGTTGGCAGGAAGTT	
586	TGTATACAGGGTCATATGCTGAAGTAGAATCATATA	
587	TTCTAAAATCCCTTTATTGCTATTGCTATCTTTTTTGA	
588	TTTTTTATTCGATTGAAAAAACTATACATGGCAATCAC	
589	GATGTTCATCTTTATCAAATTTCAATTAGGAAAAAAC	
590	GCAAAGGTGAACACAAGATAAGTGCCACCGATGGATA	
591	GATAATATGGAAATGCGGCCCATTGCTGCAGTGTATA	
592	CTGAAGTGTGAGCTTGCTGCTGCTCCGAGTCCGATTA	
593	ACGCAGTCGGTGGGGGATCTGTCGAATCGGTTTCCAC	
594	GATTATCGATTTTCCGAATTAACGTTAAAAATTTT	
595	TGAAACATACGACACAAACAACAGGATGCTCAAAAA	
596	GTCTGAATTCCCCATTCGAAAAGGTTTATGAGTTGA	
597	GTTTATTGACATGCTGCACCTTCAGGTATTTTTCCA	
598	GTTCAGCCACTGCTGAAAGTATCGTCAGCAAAACTGT	perfect
599	GTTGAAATCGTCATTCGTTGGAACTGATCTCAATTCT	
600	GTATAATGCTGTGATATGAACCTTGCAGTTCATACT	
601	TAACACGGAAGAGGGCTAACTTCGCCCTGAACATC	
602	GGAACTCGTGCTGGTTGCAGGCTTGACCTGCTGGGATGA	
603	TTATGAATACCAATATAACATGGTCTAACTTTTACAAT	
604	CTTTACCAAGCAAACTAATTAGACAAAATTCCTTTC	
605	GTAATACACACATATATATAAACCTATCTGTCCTTT	
606	GTTACTGCTGTGCTGAAATCCACGGTTTGCCCTACA	

607	TGTCATCTTTGCTGTTGTGTTCATGTTATCCTCTCTA	
608	ATGCTACTTCCACAGTTCCACTCACACCTATAGTGTCAT	
609	CAGTACTTACATCGATGATTTTTCCAAATGTCCCTA	perfect
610	ATGCAAAACACAACAACATCCCAAAGCCCAATCACC	
611	GTTCTATATAGTTTCCTGTGTTTACTGGGCTTTGGGA	
612	TTGACTTTCTTGTGCACACACACACATGAAAAAATCA	
613	GTTCCACCAACTATAAGGATGATGAATTTCCTGGTTA	
614	GTCTTTTGTTTGTTTTAATCAGTCTGGCCTTCTTTC	
615	GAAGTCTTTCGTGAAAGTCCATTTTCCGGGTGAAGA	
616	TCCACGATTGGCAACGGAGAATACAGTGTGACATACG	imperfect
617	ATGATTCCAATGAGAACATTTGCGATCATGACCATTG	
618	GTCTTTCAATTTATCATATTCAGATTTAGGAATTTCC	
619	GGATTATCGATTTTCCGAATTAAACGTTAAAAATTTT	
620	TGAAACATACGACACAAACAACAGGATGCTCAAAAAA	
621	GTCTGAATTCCCCCATTCGAAAAAGGTTTATGAGTTGA	
622	GCATGATAAGATTTCATAACAAGCATAATGCTGACAT	
623	TCGATCTCCCTTATGTTGTCGTAGTGGATATCCGTCA	
624	GCGGCAACCACACAAAATATGAACTCAAGAAACCACA	
625	GATGTAAGCTCAGAATCGGAAATTGAAAATAATCAGT	
626	AAATTGAAATAAACCAGATTTCCTTAAATATTTCTCC	
627	TTCTACATCTTTGCAACACTGGCAATGTTAGAACTC	
628	TTATGCTTCGTTTCCCACTCCTGCCACCCATCCATTC	
629	AATAAAAGACCTTCATGGACACGAACTAGAAACAAT	
630	CCTATAAGGTCCACATAGACCTTATGCTTTAACATTC	
631	CTGATTTTTATTTTCATTTAACATAATAAGATAATT	
632	GTTGAAATTTTGGGTGAAAGTTTGCCAGGTGTATTGAT	
633	GAATGATATACTCGGCTGCGTGGATACCATATCATA	
634	ATTCTGTATATTGATCTCCCTACAGGACAGGTATCCTG	
635	GGTTACTTGGAACTGGAGACGTTCCGGATAGATCCTG	
636	GTGAAACATACGACACAAACAACAGGATGCTCAAAAA	
637	GGTTTATTGACATGCTGCACCTTCAGGTATTTTTCCA	
638	ACAGAGATTATCAGCAGGAAGCTAGACGCATTGAATAA	
639	CATTCTTGATTTGCCTAAATCCCTTAATCTGGATTTC	
640	GTCAAATCGCCAACCTCTTTCAGAATTTCTTGACTATT	
641	TCTTTATGTTTGTCCTACCCCAGAATTCATCCAGAA	
642	GTGGAAGCATTGGCATGATTACCACGCTCCTTCATT	
643	GATCAGGCCTATGCCCATTGTCCGGGTGTACTTCT	perfect
644	GTATATCTGATAAGAAAAAGAAGATCCAGGAGGATGCT	
645	CGAATAACGTTATCAGGTTTAACTTCCGCTCTCATG	
646	TTCAATAGACATCTCCTCCCGTAATACACGCATATAT	
647	TTCATATTTCTGGGCACCGTATGATTTGTCAGCTTTA	
648	TCGTCCAGCTGCTCCTATATTCGGTGTCCAAAAATA	
649	CCATACATTTTCTGTATTTCAATCAATTTTATTT	
650	TTCCATCAATAGGGAGGGGTATACATCTTCCCCGTA	
651	ACTACAATCGGTCGAAATATGCAATGTTCGAATCCGT	
652	TCATGAATGATGACATAAAACGAGTATGTCCCCGATG	
653	GGGGAGGGGTCAGGGCCGGTGGAAATCCTGGATTGT	perfect

654	CTCAAACATCCTTACATGTTCAGATTTTCTCGGTATG	
655	CCAGGGTTTTCTGATTTTCAGGCCTATGAAATCTCT	
656	GGACAAGACGGATATCAGAACTGGTACCATTTTTACC	
657	GCTTGTCTGGGCATCAATGTAATTCCATTTGGGCACC	
658	TCTCAAGACATGTACGCTCCAGACGGATACAAAGCA	
659	GGATTCGATGAGCACATTTGCGCTCCTCGATCTTGG	
660	GCAAATATGAAGCTCTGCTGTATATCTTGACTTTAA	
661	ATCATCGCCTCCCTCATATGGGGACTGTCAGTATTGA	
662	TAATTCGCAAATCAATAAATGCATATTTTATTTT	
663	GTCTTTGGGTCTGGTGAAAGCGAGAATGAAAGGATA	
664	GATCCATTGCAGTCTAGACACTCATGAGAAAGCCTAG	
665	TCCCTCATTCAGGGATAATATAGTATTACATCCTTATA	
666	AACCGTGCCAGTGAGCGTCGTGAACGGCGGATATCA	
667	TTGCTGATGAGGGAGCATCTGTTCCTTACGATATTGG	perfect
668	TGCTCCTGCCCCGAGGTTATAGAATAATATGAAGTGA	perfect
669	CATACGTTTATTGTTTTTAATTACAAATATTGTTAT	
670	ACGAATAACGATTTCAACCGGAACGTGCTTATTGTA	
671	CGTTTTCCATTTCCGATCCCCCGATCTGCAGAGCCT	
672	ACAAAAGAAGATATCTGTCGTCTGACCTTCGATTCC	
673	GTGCTATCGAATGAATAGGTAAAGCTAGAGGAAGTG	
674	GACAATCAGGGTGATGCTGTCTTGTCTAACGGTATGA	
675	GAAAGAAGAAGCTTTCCAGGATGGACTGGAAAACAGGTT	
676	GAAGAACACACTAAATTATTGTATTATGGCGACAATC	
677	ATAAGGTTAGATGTGCTAATTGTAAAAAATTTGATAC	
678	TAATACCTTATTACGTTATTTGATATAAACATTGTTA	
679	TGAGGCCCTCACGCGCGAAAATCTCCTCAATAGAAAG	
680	TTTTAAATAACCTCCATACCAATAATACCGATAACAA	
681	GCTTACTCCAACTCATTTGAGGTCCAGAGAGTCTAATTA	
682	TATCAACTGCTTTATTCCATTTTTCATGTTTTTTA	
683	TGAATAAGGACATACCGTCCATCCCTCAATTTTGTG	
684	TTCGATATTTTTGCGAATAAATTCTTCAACTTTTCCATT	
685	TTTCACATTCAATGTCAAAGAAGGGAGTGTAGATATC	
686	TATAATCCTGTATCAGTTTTTTGATCTTCTTCTTAGT	
687	GCCCCGTGAGAAAAAGGTTACCACCCAGAGGTTACCA	
688	ACTTCAGAAAAAGACCAAATATCACGTGACATTCAACA	
689	GAATTTTTCTAAATCAATCTCGGTCATTTCTTTATT	
690	TCGAGTGTTAGGTTCATTGTTCCTTTTTTTCTGTC	
691	GTGGAATTCTCGTGGTTGGGAGCTCCGGGGTATTTTTT	
692	GACGACAATCTGCCGCAGACACTGGCTGTTTTCACATC	
693	AAAAACAGAGCGGGCATTTTCGACCGTTCCCACTAT	
694	AATGTGTTCAATCCCAATTACATCTCCCTAAATCCAG	
695	AACCAATCTTTGTCTTTTTATATTGCATTAACACGT	
696	GGATTCCGCCACCCCTGTATTCCCTGCCGTTGCCACA	
697	GCTCCTTACTGCGATGTAGAGGATAAAATAGGAATGTT	
698	ATGGTCTGAAATGATGAGGCTTACCTGAACCTGATCG	
699	ACGAATAACGATTTTCAACCCGGAACGTGCTTATTGTA	
700	GTCAGCCCCGGACCTCACATACTCAAACACAAACATA	

701	GGTGATCACATCGCCATAATGGGAAAGGTTAAAGGA	
702	GAAAAATGATCAGTTGAAGTTTGGAAGGGCAGTTTG	
703	GTGAAACTACCTTGATGTGATCTGTCCGCAGAGACC	
704	GGAAATCCAGTGTACTTTCCAAATAGGCAATATCGGA	
705	GTTTCTTATTCCTCCAAATTTATTGTTTTTTGCTGGG	
706	AGCTATATCGATACTTCAACCTACAAAACCCGATC	
707	GGCCCATGTCATCTTTGAGGTTGTGTTCATGTTATCCT	
708	GTATGTACCCATTGCCTGGACTGCCTCAGATTTATTT	imperfect
709	TACTATTATACACACACATTTTTTATTTTTTATTATT	
710	TTTCTCAATTCACTTCCATCCCTCAACCAGGGATAA	imperfect
711	AGAATCTTTCCGGCCATATTTGCAATTAATGATTTAT	
712	TTTTTTGCACAATATCCCCTCACAACTTGAGATTTA	
713	GGGTGATCACATCGCCATAATGGGAAAGGTTTAAAGGA	
714	GGAAAAATGATCAGTTGAAAGGTTTGGAAGGGCAAGTTT	
715	CCAATTAGTCTTGATGACTCTTATGGAAAAGTCATA	
716	GAAGAAGTACAGGACAATCAGATACGATATAAGATT	
717	TGATATGTGACCTGCTCAATGTAGTTTCCTGTGTTTA	
718	TACATTCCATTGATACTGTTTGCAATAGGATACAGT	
719	GGAAAAAAACAAGACCCAATCTTGAGGAACATCGACA	
720	GTTGTCAGCTCATGCCCAGAAGGATGTGGGAGAATTCG	
721	GACAGAACAATCATTCTGGTGTGCAATCAGACCGAC	
722	TCTTTTGCGATTCCTTCTCTTTTGTCTCTTTATTCA	
723	AATTTAATCAAAACGTCATCAACGGATTTATTCCATT	
724	GCATTTCCAAGACTAATTCATCGTAAAAAGCATCAGC	
725	GTTGTATAGATGTCTCGCAATCTCCTCATCGATCACGG	
726	GTCATCACCACGGATGGGGATGGAATGGATGGAATGGT	
727	GGAAAAAAACAAAGACCCAATCTTGAGGAACATCGACA	
728	TTCCATTTTGATACTTAAAATCTAAATTTGAATATA	
729	CTTCCTCTTGCGGAAGTTCCTCTGCCAGCTTTTCTTA	
730	TAATCCAACGTTCCTTCCCATTGAGATCCGTGAATA	
731	AACCAATCTTTTGTCTTTTTATATTGCATTAACACGT	
732	TTATCAAGACTATTCTATCGCTTGATCGTTTTGAGT	
733	GTTCGTGGGCAAAGGTGCTACCCAACGGAGATGTGGGG	
734	GTACTTCCAGAATATGCCTGAATACCGATACCATGT	
735	GAGAATGGATTCGAAGGTGTTGATTAAAATGGAGAACA	
736	GGAGAACCACGTTGAAGAAATCCATGCCTCAGATCCT	
737	GGTACGGATGATTGGTTATGTTCTCCCAATCGTTATAT	
738	GGTATCCCCATATCGTATATTTTGTCTATTTTTCTAAA	
739	TTCACCTCCGATTCTTTCAGACCTGCCAAAAAATTCTTC	
740	TCTCTATTTGCTTCCAGCACACCTATCCTGACAAAAA	
741	TAAAATAAATAGTTACCGAATAAATTTGATTTTTGTTA	
742	TAATTTCTCTTATTTTTGATAGCCTGATAAACTGTT	
743	GGCATCACTGTAAAACTGTATTTCGTCTAAAAGCAA	
744	CTTTTTTTTCTTCAACAGGTGCAGGTGATCCCTTGA	
745	ATTTTATCGACGGTCAAAGAAAATACGGTTTTGAAAT	
746	GAATTTATTCCTATGGAATTTGCCAGAAATATATAGA	
747	TGACTATGTCATGAACATATGCATGCGACAAAAGTGTG	

748	TATGATAAGATTTCATAACAAGCATAATGCTGACAT	
749	GTAGAACTATCTGTCCGTCTACTCCACTCATTTGA	
750	GTCCAGATCTGAATCTTCTGATATGGAAAATAATAAGG	
751	GGGTAGGTGATCGATCGTTCTTTGGGTCGGTCTGGGC	
752	GCAAGGAAACCTACCAAAAGAACAGGGTTTCCAAGGA	
753	GGAGACTGACAGCTACCATGATAAACATAATGCAAAAA	perfect
754	CAATGCAGTAGGAGCAGGGAATTATATCTCAATCAAC	
755	GTATATCTGATAAGAAAAAGAAGATCCAGAGGATGCT	
756	GGTCTGAATATTTTGTAAAAACTCCTGAATAAACTATT	imperfect
757	TGATCTCATCACACGAGCTTACAGTGCAATGGAGGAA	
758	CAAAGAAGCCGGTAAACGGGAGGAAACCGAGGAAGTGA	
759	GGAAATAGTTTCTTTAGAGCCTGTAAAAATTATTTCTT	
760	GTCTCTTAGCTTTGTTAGTCCTGAGTGCGTTCCTATCC	
761	GTAATTTCTCTTATTTTTGATAGCCTGATAAACTGTT	
762	GAAGACGTGAATGACGTAATCGGCAAAGAAATAAGGT	
763	CAATTTTCTACATGTTTGATTATGAACTTTATTTGTA	
764	ATTAAATCTGCTTGGTTATATGGAAAATTGTATCGC	
765	GGGTGACAGAAAAAAACATCCTGTGGAACGTATATATC	
766	ACATGGAAGGTGCAGCTGCAGAGCACCCAGGTTCTGA	
767	TAAAAATGAATATACATTATTACCTCCTAAATATAT	
768	TACGATCTGACAGTTTGTATTTCTCGATCTTATATGGA	imperfect
769	CAGTACAATGCATCTGAACTGTACAGGACATTTATG	
770	ATTCGTCATAAAAAGCATCAGCCCGTCTAAGATCAAAA	
771	GTTAAATCTTACTGTAATTTAAATCAGTAAAAGTT	
772	TTCTTGACATTCATTAACAACAACATTAGTAACCAGC	
773	GAGTAAGATCCCGACACACGTCCAGAGAGGTCTTCT	
774	AGGTTTCCTCGGACCCAGCCCTAACTATCTCATCAA	
775	TCCATTAACACTCTGAATTGTTGGATCCTTCATTTT	
776	CTACTAATTAAAAATAAAAATAACATTCCAATCATA	
777	TCGATACGGTCTACGTATTTTGTAAGAGTGGGGACT	
778	TGCTGAAGCGTCCCATACTGGTGCTGTGCTTAGGTTG	
779	CTATTGGTTTGTTGTCCTAATTTTAGTTGCTATTTTCT	
780	GGTTTAGACCGAGTACCTTTTCTCCCCTGCATTGTGCC	
781	ACTAAAAAATGAAGGCTTAATTAGGAGGAGGGATTAA	
782	GGATCTGTTAGGTAGATCACTTGAACCAAGATTCCA	
783	CTGTTTCAGGTGAATGATCGTTCTCGGGAGCTCCGG	
784	GAAATCATCGGGGAATTCCAAACCCTTCGAAATGAGA	
785	TCTCTATTTGCTTCCAGCACACCTATCCTGACAAAA	
786	TAAAATAAATAGTTACCGAATAAATTTGATTTTGTTA	
787	GCCTTACTGGCCAGAACCTTAGGGACTTCGAGTTATTT	
788	GGAGAGATTGAACTGCGACACCGATGCATAGAATGT	
789	GATATTATAACCAAGTTCCTGTGCCATTCTAATCATT	
790	GCACTTCAATACGGGAAACCCAGATCACGCACAACAA	
791	ACTAAATCCACTGGTATATCTACATCATCAGTTAACA	
792	ATGACTGAGGAGATGCTGAGTATGATGCCGAAGGAAGA	
793	TGAATAATAAAATATATATTGATTAGACATTGCTAAAA	
794	CTAACAACTAAAGTAACATATGTCCCGATTAAAATTA	

795	AATGGTAAAATTGAGAACGGATAATTAATTACCGCC	
796	TTTTATTCATTTTAACATCACTTTTACAATTATTCCT	
797	TCTCATTTTTACATCACATTCCCGGATTCTTTTTT	
798	AAAGAAGAAGCTTTCCAGGATGGACTGGAAACAGGTT	
799	GTGGCTTGGTAGCTCACTGCAGGGCTTGCTGTTTCG	
800	TTTTTTGTTTCATGTGTGTGTGTTTGGATTTTCTTGG	
801	TGTTACATTGGTGTTCAGCGTGGGATTCGGTTCAATA	
802	GGATCCATGGCAGAGAGCACGTACTTCAACTGATTTA	
803	GTCCACGTCTTCGTAGACTGGAACTCTGTATATTTT	perfect
804	CAGTACAATGCATCTGAACTCTATAGGACATTCATAC	
805	TTAGACAAGACAGCATCACCCTGACCGTCCGGCTTGA	
806	GGAATGGAGAAAACAGAAAAAGGAAAAAGAGGAACA	
807	GCCAGCGCTGAAAGTGAAGCCGGCTGGACATTCTCT	
808	GGTGTCGAAGAGTTCAGCTCTGAGCCGGGATAAGGAT	imperfect
809	ATTTCTTCGGCGGTTTTTACAACTTTTTGTATTTTCTA	
810	GAATGCCTCAGTCCTGACTGTTCCAACAGCTCAATT	
811	GTTGTAATTGTCCAGCCATCAGATCACTGTCATGACCA	
812	TCACTCTTTCAGATATGAGTTCTCGCTCAAATTGAGC	
813	TGACCGCAGAACCCCCAAATTCAAGACAGATATTCG	
814	TTTCCCCATGTCCATATCAATCATAGTTAATCCAAAT	
815	TGAGTACAAATATTCCCCTATACCTGAAATGGAAAAC	
816	AGCCTGTCATCACCACGGATGGGATGGAATGGATGGAA	
817	AATATATCAAATGTCCAACCGTTTAAGCCGCCGAAAA	
818	TTCCCACGATCTGCCATGAATCGAAGCAGAAGACCT	
819	TTTAAAAAATAATGTTGAAAGCCATAAACATTATATG	
820	GTCTTTTTGCCAACCCATTTTGCCGCCACAGCCGCAA	
821	ATATACAAAGTATCTAATTTTTTACAATTAGCACATC	
822	CAAGAAATAGGGAACAGTATGTGCACGAATACAGAA	
823	GGAAAGAGCGAAGAAAACGTTATGTGGCTCATCTGAAAA	
824	GGTGAACTACTCCGCCGGAACAATTGATTTCAATATT	perfect
825	GTATGTCTATGACAGGCTTAAATCCAGTAAAAGGGAG	
826	ACCGATTCCAACTCTATACTTGATTGTCATGGCAGG	
827	GATGGCAAAAATGCCTGCTTTTCCAACAATCAACTATA	
828	GAGCCCGATTGCCATCATGAGGCCAAGGAACAGATTTG	
829	GCCAAAAGGATGGAGGGTTAATCAGATCCTTCCAATTC	
830	ATTATTAGCTTTCCCTGACTATAGCTCACAGATGAC	
831	AGTAACCTTCCCATTCTTGTTCAGAGGATTTGCATAA	
832	AATACATGATGTTTGCAAGTGTACTGGATGATCTGAC	perfect
833	CATCAAAATCCCTGTCCTCTTTGTCAATCACTTTATAC	perfect
834	CAGAATTGTACCTTCATATCTTGTTTCAAAGACTCC	
835	GGATATCGCAGTAGGTAAAGGGGGGTGCCTGTAATCACGC	
836	TATCCAGTAGACTCACAGCCTGCATTACCCTGTCATC	
837	GTTGATCCGTCAGGAATAGAGGCAGGGATACTTGCCA	
838	GGTTGGGTTATAAATGTGGTGTAAAGGAAGACAAGGAA	
839	AACCAATCTTTTGTCTTTTTATATTGCATTTAACACGT	
840	GTGTTATGTTAGTATTCATGACTGGGTTCGGCTCGATT	
841	GAAATGCTATCATTAAAACCAGTAGCGTATCGCATAGA	

842	GAGATTCTTTACTATCCATCCGTCTTCCAGGAAAAA	
843	CAGCACTTGCATAAAGATTATAAGTGCTTCCATTCG	
844	TGTCATAAAGCTGATCTACTCGGTTGGCAGGAAGTT	
845	TTTTTATTCGATTGAAAAAACTATACATGGCAATCAC	
846	TATTTTCGAGATGAAAGTCACCGATGGAAAGGTTTC	
847	GTCATCCAGGTAGGGGACATATCAGGAATCATAT	
848	GTCTCACATTCACNATCAAAGAAAGGAGTATAGACA	
849	AGTTTTAGATCATTCTTTATTTGACTTAACAACAAAA	
850	GATTGGTATATCGGTGATTGCAACTATTGTAATCATA	
851	GAGGTATGCAGTTCGATTACCGTCATCACAGAAAGT	
852	TTTCTCAATTCACTTCCATCCCAATTCGGGATAATA	perfect
853	GCAGACATCCCTCAAGGATGTTTCGACCCCAGCCAT	
854	TCGTCCTCTTCCTGGAAATTTCTGTATTGGGCCCTT	perfect
855	GGATAGTAACCAGTCCAGCAGCCCATTACGAATACT	
856	ACCCCAAGAGAGAGGAGTATTCCCCCTCCAAACATCA	
857	TTACGGATGTCTGCCGGCCAGAGACCGGAAAATAGG	perfect
858	TGAGTATGCGCACCCAATATAAAGCCGTAATAATTA	
859	TTACCATCGCAATGAATATCAGTGCTGCAAGATCGA	
860	CTTGAAGTATGGGGTATCCGGTCTCGGGTATGAACT	
861	TAATTATTCAATATATTCAACAAATAACACTAATATA	
862	TAGAAATAATACGCAATTTCCGTTTTTTAAGACCAGA	
863	TTAAGAATAGGTTAGAGTTCTTAGGGTGTGGCTGGTA	
864	TAAATTCTAACGGGTCATCAACAATAGGAATATATATA	perfect
865	GAAATATTCCGGATATGTGCTGTATGATTTGAGGTG	
866	ACCAATCGCAACCATTTCAAAACCAAGTTTCTTTTGC	
867	GAGATCGAATCAATTCATCTGGATCTCTATATTTCT	
868	ATTTATTTGGGAGAGAACGCAACGGGAAGGAAAATA	
869	GGCTGCCTGTCCCCTGTGGTACTCGAACTGCTCCTT	
870	GTATCATTTCGTCCTCGTCTCTTCTAAATAATAGAT	
871	ATTCTCCAGGGATCATCATGCAGCAATGCTCTGATCTGA	
872	AAACTCAAAGAGACCGTCCTAGAACTGGAAACACATG	
873	CAGTGTACGAGGGGCATTCATGTGGGGAGAAAAATT	
874	GTTCCAGGGCGGAGTAGGATTTTCCACTTCCCCTTGG	
875	ATATAAGCTCTCTGAGTATTACAACTCAGGTTTTGG	
876	CGTAATTAAGCCTTCATTTAAGTATTCCTTAAAACA	
877	TCTAAATATATTTAATTTTGATTGCTCTTTATCTTC	
878	TTATAATCATGGGATTTGTCCTCATATTCATTAGTA	perfect
879	TTTGTGCTGGTTGCTCCTCCCGCAAAGCTTCTTTGT	
880	TAAAGCAGCTGAAGGAGTCGAAACACAGGAAGCAGA	
881	GAGGGGGACGACGAAAGGCCAATAGGATTTCCAACA	
882	TTTGAATTTCTCCCAGGGATGAATATCCTGGAAGGT	
883	GTGAAGAAATCTCTCACGGAGGGGTCAATTAATGCA	
884	TACGACAGATCTGGGGCTGTTGGTAATGAGTAAGTTC	
885	TAGGACAGATCTGGGGCTGTTGGAAGCTTGAAGGAG	
886	GTTTTCAGTGTCGAGTCAGTACTCATCTTCAAGTTTT	
887	TAATTAAAATTTTTTTAATAGGTGATTAAAAATGCA	
888	AGAATCTTTTTCAGCAAATTCACGCGGTTGGAGCGC	

889	AATATAGGAGCGGCAGGACGAAGGAAGGCAGAAAAG	
890	ATTTGCAAAGCTGAGATCTCCAGGGACGAATATTAT	
891	TGGATGATGGATTCCGGAATAGGAGGAATGTTCAAA	
892	GGCGAAGTTATATCAGTTCTAAAGAGGCACATAAAT	
893	TGGATATTGTGGTTGATTCAATGAATGCATCAATTGA	
894	GGTGATCCTTGGCACAAAAACCTATTCCTATCCCAA	perfect
895	CATCATCTTCCCAATTTCATTTATAACTTCCATCCTA	
896	CCCTTTCCCCGACAAATGATGGCTCTAGCAAGTTTGC	
897	TGTATAAGAACATATCGTCCATCCCTCAATTTTGTG	
898	TCCGTCAGGCCTGATCGTGAAGGTGGCATATCCATCC	perfect
899	CTGTTCCGATCACCATTGAAGAATACTCAATCGGAA	
900	TCCGGTATTCCAGGAGGCCACAGGGCAGTTTTCTCT	
901	TGATGTTCGAGAATGCATATCCCTGATATTTGTATGA	
902	ATGTTCCAAAGCGTAATACTCGAACCTGTCAGATCTT	
903	TTTGACTGTCTGTGTTAACGCTCCTTTTTTCATTTCT	
904	TTACATTAAAATTATCAGTAAATAATGCTACTTATA	
905	ACAACAGCAGGATATTTCAGTGTGGATGCCGATCTA	
906	ATACTCGAATATCGATACAAACATGACCGATTCATT	
907	TATCAATCGTAGCCCAGTCGTTTAGATATAGAAATC	
908	AATTTCCCTAATGATATGTAATTATTTGATTCGTATG	
909	GTTGAGAGTCATCTGGATATGTTTTGCCATGGTGCT	
910	GTGAATCGAGGAAATTTGTTCTGTCCTCTACGAAAAA	
911	TGATAGGGATAATGTCGTTAATCGGGCTAGGAGCAGC	
912	CTAATACTCGTCACGACACTAATATTCATGTCATACA	
913	AACATCTTGCACCTCCATGCGGCAGATCTGTACTGA	
914	GTTGGATACCGTGCCTTAATTGCACCTAGGATGGAG	
915	TGATTTCTGCCTTCATTTGGGCATTTTCCAGTGAAAG	
916	ACACGTATGTATATGGCGAAGTACCGCCTGATATAA	
917	TGAAAAACTTTGTAGAAACTGCATGTATCAATCTCCA	
918	TTCCTGAAATGTAGAATATTTTGGTGAGAATTTAGT	
919	GGGGCTGTTGGGTTGCAGTATCTCCATGAGTGTTAC	
920	GACAGATGGGACATTGAAATTGAGGAAGAATAACTATA	
921	TGACCCTTGGGGAATCGTACCCCTGTCACTCCCGAA	
922	GAACGTGACTGAATTGTCACCCCAATTTCAGGTGAA	
923	GATGAACTGCGGGCTTGGCTTGCTATAATACTTTAGG	
924	GACTTCGCCATTCCCGCGATCTGCCATGAATCGAAG	
925	TTGTTGTGGTTGTTCCTAATGAGCGTGTCAAGGTTC	
926	CATTTTTTTTGTCCTGTTTTTCCATATGGTGAATTTA	
927	AATACATTTCAGTTTAATAACTTCACAACATTCAAA	
928	CTTTCTCGCCTCGTAAGCCTCGTAAAATGCAATTAAA	
929	TTGCGTTCTTTCCCAGATTGACTGGATAGCTTGTGAA	
930	ATGCATTGATCTTACCTAATCTCAAAAAAGATGGCAA	
931	GAATTTATTCCTATGGATTTGCCAGAAATATATAGA	
932	GGAAAGCACACTGGATTTCAGGATAAATGAAATAAA	
933	TTAACCTGTATAGTCGGATATCTGAACCCAAGCCTGT	
934	GCAAATTCCTCTTTTCGACACCCCGAAAAAGAGAAGC	
935	TGAATTGCGATATTTCCGAATTCCGTTTTTCTTTTT	

936	GCCGAATGCGTCCGTGACTGTTAGCTGCACTGTGTA	
937	TTGAACAGGATATCAAGGCCTCGCAACTTTATAGCA	
938	GAATGCGGCCGATCAGTGGGTCTTCCAGATCAACATA	
939	GATCTCCTTTCTGAGCCTTCTCCATTTTTGCTTTGT	
940	GTATGAAGGAGGCGATGATTGGGGCGAAGACTCCATAT	
941	GAACCCAATCACCGTCAACCCGACGGACATAATTGCC	
942	TGTTTTCTCGCCCTTTCCAAGGATGACTCTTACATA	
943	TGAATAAGGACATATCTCCCGTCGTTCAGCTTTGTG	
944	CTTTATAACTCCATTCTGATCCTCGACAAGCTGGGT	
945	TTTACTCTGCCTCCGTGAAAAACTTTGTAGAAACTG	
946	GAACGATAAAACGGTAAATTGTTGATTGGTATATTAC	
947	CTTTTGCACGGCTGACAAATTATGATTGGCAATTAAA	
948	AATAGTCTCTTTAGGAATTGCAGGTAGCTTTTCCTG	
949	GACAGATGGGACATTGAAATTGAGGAAGAATAACTA	perfect
950	TATTAGGCGGATCTCCTGCGTTCACATATTCATGGA	
951	GATACTAATTGAGCGAAGCCGAACGGTAATATTGAG	
952	GAGAATCCTACAACACAGCATATCCGGAGATGGATT	
953	ATTTAACTCACTCCACCCGTGGTGCAACAAGACCTGT	
954	ATAGTATTCAATCGCTCCTATTGGATCTACCCAGAA	
955	TGCACATGCTTTGTCCTGATATGCTGGTCAGAGTCGA	
956	GACAGTGTATATGCTATTGATGAAAAACAAAAGTAGA	
957	AATTATCGATTTTCTGAATTAAACGTTAAAAATTTT	
958	GAGCAAGATCCCGACACACGTCCAGAGAGGTCTTCT	
959	GTCTCTCTGTGCAAGAATTCCCATATGGTGAGTATG	
960	ATAATAATTTTCACCGTCATATAGACAACCGCCATACA	
961	GGAGAGGAAGGATGGAGAACCCGTCCTTGGGAAGGA	perfect
962	TGAAAAATATCTGAATATGTATTGTTTAGATAAAAATT	
963	CGAAGAAAAAGCACCACTGTTGATCAACACATGGTT	
964	CAATATTCTTTTTATAAAATACCATTTGCAGGTTCA	perfect
965	AAGAGGGAAAGATGCTGAAATACCTTCTGAGAAGA	
966	TTTAAACATTTCTCAATTCACTTCCATCCCAATTCGG	
967	CAAAATTATTTCTGTTCCCAACGAAGCAGCTTTGCTG	
968	GACTATAAACTAAGGACGATAACAATAAATAAATTA	
969	AATAAATTGAGTGGGAATGCCGCATTTCTAAAAAAC	
970	TTTTATGGTTCTTTTTTGCCTCAGTCAGGATATCAG	
971	GAAATATGAATGGTCATATTTTTTCCACAGATTTAC	
972	ATAAAATTATTCCCTGAATAACTATCACTTAGTCCA	
973	TCTCCTATCCAGGACAGAATATAATATTCGTCCCTG	
974	AATGCATACATAAAACTTCCGATACCAACTAATCTT	
975	GTGATAACATTTCCGATGATGTCGGCAGTGTATACGATG	
976	TTTTTATTCGATTGAAAAAAACTATACATGGCAATCAC	
977	ATGTCATCTTTATCAAATTCAATTAGGGAAAAAAAC	
978	TGCACTCTGCTGCATGCATATGTCCATAACATAGTC	
979	TTCTATGCCAATGGTGAGATCGCATATTTGCTCTTCA	
980	TGTATGCATGATTGAACAACCAAGTTACCGATATACA	
981	TGTCTTCTCAGGATAGACTAGGTCCATCCATTCAACA	
982	TGGATATCAAGAACACAAGCCTATCATTCACCCTCAA	

983	CGGATCCTCCCGTAAAATGGACAGTAAGACACACCG	
984	CGCTCTCTGGAAGTTCATTGCTGGCAGAAAACATCG	
985	ATGTATATCCAGTCATTCGTTGTCAATCCTAAGAACG	
986	GAATATGTTACAAGACATGAAGGGGCTTCATTCCAA	
987	TTGTCAGCTCATGCCCAGAAGGATGTGGGAGAATTC	
988	TGCAACAGGGCTTATTTAATGAGTAGTTCTTGATTG	
989	GTAATTCAGGTTAAGGAATCCGATGGAAAGACAGCG	
990	TTTTCCTCTTCTCTGCCAGTGCCCTCAGGAATGTCG	perfect
991	TGATAACATTTCCGATGATGTCGGCAGTGTATACGATG	
992	GTTCGTATGTGATCTTATTCATGTGTTTCTTCCTCCT	
993	GAAGAAATGAGCCTAATTTTGCAAGGATACGAAGAA	
994	AATACTATTATCCAGTTTGCAGCGAGCCAGTTTAAG	
995	TTTAGGCTTAGGTTCTGACTGCTCTACGTTTTCACT	
996	ACGGATATTCGTGTTGTCTCAGGTGCAACATCTGTA	
997	TGATCCGGAGCTCCCGAGAACGATCATTCACTTGAA	
998	TAGCAATTTCATCAGAATAGTATTTAATCAAATCAGA	
999	TCTTCAACACGGATATTTTAACGCCAAAACCATGCAAT	
1000	CTATTTGTTGGGTCACTGAATGAGATGATCGCGGTG	
1001	TTCCCATCTCTCCCATATACACAAGGGCATAGATCGA	perfect
1002	ATGCTTGTTGAACTCGTTTCGTTTCCCAGTGCTAGA	
1003	TAAAGGCACGTTCTCTGTGTATGCTGTTAGTGATTTA	
1004	GAACTGATAAAAAGAAATATTACAAATCCTATCAATA	
1005	CTTATCCTTGAGCTTCATTACGTTTGCCGCAAGCTGC	perfect
1006	ATCCTGACCAGCGAAATAGAAATGCTGGAACAGTCAGGA	
1007	TATATTTCTTATTTGATTCAAAGCAGATTGACGTCTA	
1008	TTTGCTTCCATTTGGTCCCTTGATTCTATGGGATCCA	perfect
1009	TTAGTATGGCTAATTCCGTATCGCTTAGACGCTTCA	
1010	TTCAATGAATGATTCATCGGGCACTGTAATGAATGTA	perfect
1011	TTTTATCTAACCCATACAGATATATTTTCATTTTTT	
1012	GATTAGGTCAGCTCTCATGTCGTCAATTCCTAATCTA	
1013	AATACTTTAGATGACAAAATCGTAAGAACAAACTTTC	
1014	AGGGTAAATATCGTGCCTATGAGCCTGTATTCCACG	
1015	GGAAAAACCCTTTCCCAAACAACAGCCAATTTATGC	
1016	TCCATTTTCTATTGGCATGATAGTGCCGTTTGTTTCA	
1017	ATTTGCTGCGCTGTCTGATGTATTTGCCGTCTGGTA	
1018	TGATATACCCCTCCCCTCCCTTAATTAATTCAAGGCA	
1019	AAGGGCAGCAGGATCGAATTCTCCCCGGATCTTGAT	
1020	CGGTATGTTCGAACCCTCCAGGTATTCCACTGCTGT	
1021	TTGAATTTGCTTGGTTATATGGAAAATTGTATCGCA	
1022	TACGACTGGGACGACTGGGGAGACTACTGCAAGTCCA	perfect
1023	AACAATTCTGGCGTCTGCATGCTCTCCAATTCAAGA	
1024	GGCATTGTTAATGTTATTTATATAGTCATTGTGTTA	
1025	AAAGTTTTTCTCAAAATAATCAATATAAAAGTTATTA	
1026	CAACCATGAAGTGAGAAACTTTCTTGAAGAATTAAG	
1027	TCAATCACAATCTCATTCTTATCATTAGGTATTTTTT	
1028	ATGAAGACACTTTAAACAATAAATTTCCCATTGAATTG	
1029	CTGCAGCCGAGGCAACGACCCTTGTATTGAGCTGGTC	

1030	TGGATAGCCGTTTAGGCCAGCAACCTTTCGAACAAC	
1031	GGTCGAATAGTCTACATTTTGAGGAGGAGAAACGACA	
1032	TCTCAATCTCAATCAAGTAATTTTTCTCCGTATTAT	
1033	TCATTCNAATAGTCATTATTATTTCACCTTTTTAAT	
1034	ATTTCAACAGCTCCATGGCCCCAAACGAGTATCTGA	
1035	TTAAATTAAATGAGGCAATAAGCCTGCTATGTAACT	
1036	AATTTATTCAAAACCTCATCAACCGCTTTATTCCATT	
1037	CAGGCAGATTGAGTTGGAAATTTGGGGTATTGAAGCATG	
1038	TTGGGCTTCTGACGATCATCAGCATGCTTCGAACGGA	
1039	AAAGATTATAATCAAATTACCGATCACGATTTCCAATA	
1040	GAATTTTGAGAGCACACCATCTCACCCCTGTTGTTG	
1041	TTATAGTAAGAGTGAAAATTTACTATGTATTGATTTTT	
1042	ATAACCCAATTGCCTTTAGCACAACAACATTTAACCTA	
1043	TTGTTTACAGTGCAAAATTATGTGAATACTCATCATTA	
1044	TCCCATTCTATGTTATACATTATATCACTCTCTTTTCTA	
1045	CAAAACTTGAGCCATCATACCTCTTCAAATAAGGAA	
1046	ATGGAAAAATCAGCGAACCCTTCAGGACTAAAGGGA	
1047	ACATCTCCATTAGGCAGGACCTTCCCGACAAACTGGA	
1048	AAAAGTGTGGAATACCTTTATGAAAAAGACGGAGA	
1049	TATGAAGGAGGCGATGATTGGGGCGAAGACTCCATAT	
1050	TTTGCTCTGCCTCCGTGAAAAACTTTGTAGAAACTG	
1051	GGGTGACTATGGAGATCAATTGTTAGAATACCAGGG	
1052	TTTCTTTTTGTGCAGCCTCGAACCCTTCCCTAAGTT	
1053	TGATCTATCTGTTGGATTAACTGCCCCTGTTGGAAG	
1054	TTTGTTGATAATATAATATTTAATGAATTATTCCCA	
1055	CAAAAACCATCAAAGCGATCTTTTTTTCCGCCCCATC	
1056	TCCTTCCTCTCCTGGTCAAAGGACTTCATGTATGCCC	perfect
1057	GAGTAATCGTGGGGATGTCCTTGACAACTGCTATATG	
1058	GAGTGTCCAGGATCTTGATGGAGTATCTCCAGAACCAA	
1059	TCATTTTGTGACAGATATCACATTTCTCTGGAATGA	
1060	GCATCCCTATCAGCAGAGAGTTAATCCTGAATAGT	
1061	TTCTGCCTTCCTTCGTAAAGCTGACAAATCATACGGT	
1062	ACATGACATCATGATACAAATGTCATATCCGATATA	
1063	ACAGGACAGGTATCATGGCACGTTCCTAAGATAGATA	
1064	GGAAGTATGAGGTAACAGGGAAAGAGGACAGGGATTTT	
1065	TATTCACGTTCCTGTTCTGGCTCAGCCTTCTTTCAAT	
1066	ACCAGTGGTTCACGAATGGAACTGCGATCGGTGGCC	
1067	TGTACAATGAAGTCCAGTACATCAATTCATCTGTTC	
1068	CTCTTCGTGCTGTCCACGCTGATATTCCTGAGCAAGA	
1069	ATGTCTTAAGAGTTTATAAAGTTTGTCGGATTTGAA	
1070	TTTTGGAAATGGGGGGTATCCTACCATTGGCACACCCC	
1071	TTTATGGTTCTTTTTTGCCTCAGTAAGGATATCAGTC	
1072	TTGATCTCCCATGTATAATTGTACGGGGAAGTACCGC	
1073	ATATACCTTGGTCAGAACGCATCAGGAAACCTAATA	
1074	ATTTGCAAAGCTGAGGTCTCCAGGGACGAATATTAT	
1075	TTCATAGAGATATATTTCCTCAGTAGAGCTGTCATTC	
1076	TATCAAAAATATTCCTCTCATTCCTTGCCTCGATTG	

1077	GTCATATTCTCAAGCCTGGGCAATCTGGGAATTGCC	
1078	GAGAAAAAGTGTCCTGATGAACACAAATCTTTTTTA	
1079	TGCCTTAATTGCACCTAGGATGGAGTGATCTGAATGG	
1080	TTTTCAGTGGATCACCGGACATTAGAAGGTAATCTCC	
1081	ATTATTGAAATTTTGTCAGATTGACGAAATGTTACG	
1082	AACAAAAATACAATGTGAAGAAGCAGGCGGAGTATG	
1083	TAGAGTTAAGCAAAGACGGAAACCGTAATGGACAAAC	
1084	GTTCATAATCTGTGAACAGGATATCGTTCCAGGACAAA	
1085	ATTTTTGCCGCTGACCTTAGCATTTAAACATTTCTTT	
1086	GAGGACGTCGATGACGTAATCGGCAAAGAAATAAGA	
1087	CCTCCCTTTCGCCCAGGATATAGAAGAAAATAACTTC	
1088	GCACCAATCATAGCAAGCGTGATCGCAGGAATCACA	
1089	CCAAAATGTCATTATTAATATTCTCAATCACAATAATA	
1090	CATACTCGCACACAGTGTATCTGGAAACAAATCTCA	
1091	TGATATACCCCTCCCCTCCCTTAATTAAATTCAAGGCA	
1092	GATACAAGAGTGCATTTGATGCATATAGCTCCTGCAG	
1093	GTACGACAGATCTGGGGCTGTTGGTGGTGAGTAAGTTC	

Table S5. AMDV3b contig sequences. Contigs were assembled with 454 sequencing reads from the C +75m time series dataset.

```
>AMDV3b_Contig17850
GGTCATAAGATGATAACCGATAGTAGTATAAAAACATTACTTCAATGATTACTTTGGTGTCATAGTTAAAAATGATT
ATATACTTCTTAAGGTTATATCATAATATGTATGAAGACGGAAAAAGAAAACTGTAATCCGTAGCTTCAGGCTGGAT
GAAGACTTGAGCAATTGGATGGAAGAAAAGCATATATCCATAGGCAAGTTGATCAACAGCCTGGTTCGGGAATACAT
GGAAAAAGAACTGGGAAAAGAAGACCCTCACGCAAGCAAAAAAGACCTCATGACATAGATGACATAAGTAAGTATAA
AACTTTGTTTTTTGAAGAATTAATTTTTGGTTATTTAATAAATATGGATAGATAGATGTAAGGACACATGGCATTAC
ATGTTATAGAGAACTTGTATTTTCCTTATTTTTTAGAATGCACGTGTCATGAAAAAAAGCATTCTCTATAAGAGATA
AAATAATAAAAAATAACCAAATAACCAAATTACCGCACGTTATAATATATTTGGTTATTCTTATTTTCCCTTTATAA
TACTGCTTCGTTTGAAGATCAATAACCAAATAACCAAATAACCATAGTTTTCACTATACTTGTTGTTCTCGTAAAAT
TGGTTATTTGGTTATTCGTCTCTTAGCCAATCAGAAAAACACGGCTTCCGCATACATAACCAAATTTCTGAAAACTC
GGTTTTTGACTATATTTGTGGTTATTCACTTTCCAGAAAACATCTAAAATACGGGTATTCTGGAGATAACCATTTTT
CGCAGAGCCTTTGTTTTCCGACATATTAGAGCGATTAACTCCAAATAGGATATAAGGTTTTTGCTACCCCAGGAATA
CGGTATGATCCCTTGTAGAATCCCGTATCAAGAAACTGTATGCTCTTTCTTGCTCTTGTCATCCCTACATACAGAAT
GCGGTATTCCTCAGGTGTCATGGCATCTGCAATCATCTGTGGAAGATCAAGGGTGAGCATCACATGGTCAAACTCAA
GTCCCTTTGCTGCATGGATTGTATCACAGTAAACAACAGGAACGTTCTCCATCAGTAGCCTTTTTACATTGATCTTA
ACTCTGTCAGGAATATCCATCCTGTTGATTATCTCAACCGGGCCCATCTTCTCTCTGAGCATATCGTATGAAAAAAT
CGTTGGCTCCTTTTTCACTTTTGCCTTTGTCCCCCTTACCCATAGCGAAGCAGGCGAATGCTCCACTATATAGGCAA
ACTCACTGGCATCTATGCTGGGCCATTTCCTCATTATCTCCGCAATCCTGAAAGTCTGCTGTGTCCATCCGTTTCCG
TACCCGTGTTCATGGTTTATCGGTATCACCGCATAGGGAAGCTGATTGGCTATATTCTTTGCCAGATTTGTTGTCCT
GCACAGTATTGCCTTTGAACCATCCAGCTTTACCATATCCCTGAGGGAATACATGTCAATCTTTGTTTCTCCAAGCC
CTATGTATGGCATTCGCATCACCGTGCCCTCTGACAATGCCATTTCTGCCAGTTTCCTCAGATTATCCCCATACCTG
TGTGATTCCCCAAGTATTTCATACCTGTCTGCCCGCTTTACAAATTCCTTAACATCTATCCCCCTGAATCCATATAT
GCTCTGATGAGGATCCCCAACAATGTATCGTTTTTCACATTTGATATTGTCCACTATCTTCCACATCAATGGGGATA
AGTCCTGGGCTTCATCTATGTAGAGGTATCTCTTGTAATAGGTACGCTGCGATCCTATCTCCAGCACATCAGTATAA
TCAAATTTTTTAATCTCATCCTTGTATTTTTTCCATGCATCGAACAGATACCGCATGGACATCCTTTCATTCATCGG
CTGATACGGCTTTTTCATGGTGTTTATTACTGTGTCATAGTATCTGGAGAACCTTTCCCAGTCAGTTTCACCATCCT
CAAGGATTCTCATCCCTTTGGGATATGTCAGTCCGTATTTGTCGCAAAATGCCTTTATGTCTTTACTGTCCACGAAT
GGTCCGATACCGTTCAAAGAGGACATGGCACTATGGATAGTGGCAACATGGTGTTTATCTGCCTCTATCCTCTCCCT
TGCACCCTTAGCCATGCTTCTGTTGTACGTCAGGTACAGAAAATCATCGGTTCCATTCTGCTTTATCAGTTGTTCAA
TCCTGTGTGTTTTTCCTGTGCCTGGAGGCCCTATGATTGTATAAGATTCCATCTTTATGTCATGTACTGTGTTTATA
AAAAACTTCTTAGAGAAATTTTATATGCTTTTAAATTATGCATGGTTATGCAGTTGAAATTGAAAGTGAAAAAATAC
GGTAAATCATACGTTCTGGTGATACCGAAAGGCATCATTGATTCCACTGATGTCAGACCTGGCACGATACTTACCGT
TGAAATTGAGTGATCCAATGCTGCATCGTGAAACGATCGGTGATCATGTTGTTGTTATTCATGCAGATCCGCATAAG
GATAAACTCAGGATTGAGATTATAATAGACGGAAAAAGTGAGGTATGGGAGAAGATAGACAGAGAGGATCTGGATGA
TAACCCTGCTCTATCTAAGCTGATAAATGCTGGCTTCTATGAGGTATGGGAGCATTTTCTTCAGAGCAAAGCAGATT
CCCCTGCTTCAATTCTTGCCCATGTGTCTAAGGTGATCTCGTACGATCACCAGTTCTCCGCGGACATATACTTTGGA
GATAATGTTTACAGGCTTGACAAGGATGGGTTCATATCATCAAAATCATTTGTGATATGGTATCTCGGCACATTTCA
CCGCATTCCATTGATAAAGGAGGAGGAATGGAAGGAATTTGTTGCTGAAATTCTTGCAATGGCCGAAGTATCATTCT
ATGATCCCCTTGGCGTTGATATGATGGGAATGCTTCTTGACATGATGAGGCAATCCGAGATACACAATGATTTCTGC
GATAGGGTAGCAGGAATGTACAGTCCTGGGGGAGCGAAGTCATGGTTTGTGTATATCAAGGAGAACAATTTTTCCCT
CTATGTCCCATCTGGAGTGCTTCAGTCCATACGGCACAGGTCAGATATGAAAACAAAGGCTTTCAAGAATTACTGGG
CTCCGTTTCTGATAGACAACAGGTATGTTCATGCTCTCGGCACGGCAGGAACCTATGAGAAAAGACCATCCAGCAAG
TTTTGGGTTCTGTCGGTTGAGAAGATAGCTGAGGTTGATGCATCCATAGCTGAGATACTCGCAAATACCATAGACTG
CCAGAAAGAATGTGCATTGGAGGTAATACAATGATAGAATTAGATGATAATGAAGTAAATAAGGAACTGTTAGACAT
GCTGTCTGCCATAGAGAGAAAAAAGAAATGGCTTGTGTGCAATATTGATGGCAAAGATGCCCTGATTCCTGACACAA
GGGATGCCCTTGGGCATGATCTGTTTGGAGTTGAGAAATTCCAGTGCCATTCCCCAAGCCCTGATCTGATTGTTTTA
GCATGTCCCCTGAAATTCATCGCAAACAGATATGCCTGGGGCATGTTCACAAAGAGCATGAAATATTCGGATGTGGA
CTTCAATACAATGTCCCTGAGACCTGGGATAATAGAGGGCAATCAGGTCCATATCAGCAGGAAAGAGGAGGTTCACC
CATGATCCCCACAGTAAAACTCCTTGAGCCCCGCATATTCATTGTCTGTGAAAGGTGTCAAATAATTGGAATGTTCA
GGACAGAGCAGATAGATGAAGCCGCCGATTGCCTTGAGAACCACATGTTCCAGCTTCCCACTCATGCAGTGGAATCG
GCAGTCATGGAGGTGGAAGTGTGATCCCCGCAACCCTCAATCCCCGCCTCCTTGCCCTCTACGGATGCGAGATGCCC
AAGCATCATGTGAGGGAGTGGATGCATTCGTATCATGTGAGGAGAGAGTTAAGGAAAGCAATGGAAGTGATGGACAC
```

CATCGAATTGGAAAAGAGTCTGGAGGCCATGAAATGACTTGTGAATGCGGGCATATAAAGGCTCAGCATCAGTTCAA
TGGGAAATTCTGTTTTCTATGCAGCTGCACAAAATACAGAAAGCAAACTAAAAATGAATCCACTCCGGATGAAGCCT
ACCTTGACCGCAATCTTGCAGTCCAGGCATTTGGGAAATTAGCAATGCAAAACGGATTCAATGTCGGAATAAAACCT
GACGGAGAGTATTCGATCCTATACATTGATCTCCCAACCGGTCAAGTATCCTGGCATATTCCCAAGGATGAGCTGAT
AGATGGATTTCCCGAGTATAAAGGCAAGTGGGATGGTCATGATCTTGAGACTAAGAGGAATCGACTGAAAGCGTATA
TGGAGGCCCTGAAATGAACAGAAAAATTCTCACCAATTGGCTTCGTGCTGGCTTAGCATTGCAGCTGGTATGCGAGT
TTGACGGGTGCACAAGGAAAGCGGAATACAGCATCACGAGGTGGCATGGCCTCCATATAGTGGTATGCAGGGAGCAT
TTCCCTCGGATCGGCGAACTGGATGCAATGGATGAGGAGGCCCTGAAATGACATCTAAGGGAACCGCAGCTGAAGGC
CTTTTATTAAAAATGCTCCTGGAATATGGGATTGAGTGTGAGCACGGGTCTCCATCAAGACCAGGGGACATCATAGT
TCCACCAAACGTCATAATCGAGGTCAAGGATCCCAAGGAGAAGTCGTTTTCGTTTAGATCTCATTCAGGAAAGGGGG
AGCAGCAGTGGAAAGCGCTTAAAGAAAAACATGAAAAATTCCCATGGCTGGAAATTTATTACGTCATAAGATTCGCA
AGGAAAGAATGGCGCTGTTTCGACTTCCCACTGGAACCAAAACCATTGAAATTGAAAGATGGATTCGAATTGAAATT
GTTGTTTAACCTCCTACTTAGCCGGCAGGGAGCATTAATGGCAGCAGTGGGTGGTGATGCAGTATAGAGATCGAAAA
CATAGCCCAGCAGGACATCCAGGCAAATGATCCAACGGACGTGATGATCCTGCATATCCTCACGAAATACCAGATGC
ACCTCTCCGCCAGGTCCATAAGCCGCACCATAGAGAGAATCTACGGGGAGGAAATCTCACCATACACGATTGCCATC
CATCTCCGCCATGTGCCCTGTGAATCCACTGAAACCAATATCCGAAAGTATTCTCTCAAGACGTGACATCCTCCCAC
CCCTGAAGAGGTGGGTTTTCACGCTCGATGTGATAAAAATAGTAAACCTATTAAACTCTGAATTATAACTATAGATA
GTCTTGGGAGCATGTAAAAGTATGGCCTTGGGAGACTCTATAGCTTTCGCATGCTCCCAGGAACAACCAAAAGGAGG
AAGGAATTTCCGGGATCAGCTCCTGATGCTTTACCAGGCAAGGTGGATGCTTTTCCCGCCTGTTTCCGGAAACAACA
TGAAAAAATTGAGAGGAATATACGGAATAGCAATAGCCCTTCTCGTTATCGTCAGCGTCGCTCTTGGAGCGTACATA
TTTAGTGCTGATGCGAGTACAGGGCCATCCAGCACGGGCCTTAACACCCTTGATGGATATGGATTAGCCAATGAGGC
GAGCGCAGTCTGGGCACTGAACGATTCAGGACATCAGGCACACTGGTATTCTCTATCTTTCACCGTGGATAAAACAA
CCAATAGCATAACATTCGCGACGCCGCTGAATGAGAAGATCTCACTCATAATTGTGCAGACAAAGAACGTGAGCCAG
GGAGTGTATCACTTGTTACAGAACGGCCTTCTGTTCACATACACAACATTCAGTACCACAGACCTCAGCGGAAGCCT
GTCAGACGCATACATGTACTTCGGTGAACCGGTCAATGCCACGAGCAGCAATGCAATAGCAGATAAGGGAATAACCG
ATTACGCACTGAACCAGAGCCTTTTCGACGAACAGGCATCCTACTTCGGGAACTCAGTGCAGCTGCAGATCCTGCCA
TACCTGGCATCCAACTTTGCAAGCACGTCCCAGTATGCTTTCTACTTGAATGAATCAAAGAATGCCACCGGCGTATA
CAACAGTGCATCAATAACATTTTCACAGTATTTCGAGTACACATCTGGACAGCAATATCCACTTGTGACATACATAT
CAGCGTTCGCACTGGTCATACTGATCATAGCAATATGGGTGACATACTATGCATCCCCTGAATACGATGGCGGAGAA
GAATCCAGAGCGGCAGTCTTCCAGACCCGCAAGGAAATGGGAGTCACAATCGCAGGTCTTGCAGGTGTTGTCGGCAT
ACTATTAATCGAAGGATTCCTGGGCAACCTGACGCCACTGGGAGGCTGGGGAGCAGCAATAGGATCGCTCTTCTTCT
TCGGCCTGTTTACATATGCCTTTACCGAGGTGCCACAGAGACAGAAGTACACCCGGACAATGGGCATAGGCCTGATC
GGGATGGTTGTCGGCCTCATAGTCAACATGTTCTGGCCATTCGCCAGCATCGTGTATAACTTCATGACCTCCGGCAA
CGTCATAGCCATGATCTACGGATGGCTTGCAGTGCTGTTCTTCATAGCAATGGCATACTACGGTCTCATTGACACGA
AGAGGTACAAGCTGCACCCAAGGCACGTGGCGATGAAAGCCAGGAAACTGAAGGCGTAACCCATGACAGAAGACATG
CAGACGGACGTCACCAATGCTGAGGGATACGCAGGCTGGGATGCAGATGACATTCTGCTCCTGGTAGGCGTATCCAT
CATTGCCGTTGTCGGTCTGGCATTCTGGCTCTACACAAGCCAGGTGGCACAGATAGGTGTGGGGCTCGTAAGCTCCT
TCACCTCTCCGGTCTCGGCAGCATTCAACGGCCTGGCGTCGCTGATTACGAGCTTCTTCAAATGGGCAGGGCATGCG
GTCAGTTCGATGATCATCATGACCCCTTCGAGGATCCACATGAGCTTTGCATTCCATCAACAGGAGAGGATCCTCGC
AATAACCATCTCGGGACTGTGATAATCACGTCCCTCGCCAAAAATATACCAAAAGCCGCCTCCCTGACAGATGTCGA
GGATGGCGGCTACATTTTTCTATTCATTATTGGGGCAGTTATCATCCTGGGGATCATTTTCTATCCGGCACAGATAG
GGAGGATATTCTCGGGACTGTTCTCATCATTTCTTGCCCCTGCATTGGCATTCATCGGAGGAATCGCCAATTTTGTG
TGGTCCCTTGTCCAGGGCGTTCTTGGACTCGTGGGAAAGGCAGCCACCGGAGCAGGCACATTCGTGGGAAAGGCAGC
CACCGGAGCAGGCACCTTCGTGGGAAAGGCAGCCACCGGAGCAGGCACATTCGTCTGGAACCACACCATCGGGGCTC
TCTTCACCATAGGAGGCATATCATGACAGAAGAGAGGAGCGGCCTCGGCCTCATTGTCATTGGCATCATGATGCTCC
TCTACGTCATGGGAGCACTGGTGACCGGATCACTGGGCCTTGCAGTATTCCTGGGATCCATAACAGCCCATTTCACC
CCCCTGACCCCATATGCTGCCATAGCAATGTGGACAGGCATCTATATCCTGGTGATGATAACAGTGCTTGGCCTCGG
CATATACATGGAGGTATCAAATTGAATGCAAGGGAAGAACAGAGGAAGGTAGAAGAGGACAGGAGAGACAGGTATCT
CTCGGCATTAGAAGGATTCGCACGCAGCATCACCAGCAAGCACATAATAAGAAAAGAAGGAAACAAAATTGAAAAGT
GAAAGCATGGAAAACATGGAAGTGATTGTAGAATGAAAGATTTTGAATACGGAATGAAAGAACTTAAAGTAGCATCG
CAATTAAATGGAAGAGATATGAAAAAATATTATTCCGAAAATGGAGTGAGATTGCTTGATGAATAAGAAAACAATAG
CCTTAATTTTTTCACTCATTATGATTGGTGTTCCGATAGTTACTATGTTTGCGGACATGCCGAGTTTTAGTAACGGA
ACAAGTACTCCTATTTCTACACCAATCACGTCATCCGCAATAGCAAGTTCATATTCGAATACGGCTTACCTGAATAC
AACAGTTTCGTATTATCACAATGTTCTACAGAGTAATTCAGGTGTGAGTCTTGTAACGATGCCCGCTGCTGAAAATA
GTGGATACTCCACAACTTCATATAGTGATCCTTCTCAGGTATTTAGCATATCAAGTAACGGAGCAGAGACAACGACA
TCATTTACTCTTCCTTCTTCACAATTCTGGGTTGCTGCGAACACAAATACTGTCCAATTGAAGGATTTAGCGTTTAT

GGGGCAATATTCTGAACAAACTAAACACATTGACAGGTGCAACTTATTATGTGTACACACCAGGAATTATATACGTA
AATCTTACTGTTCCAGCTACAAACGGTTTTACAGGAGGAACAGTTTCGTGGTCAACAAGTGCGTTTAGTCAAGATTA
TAGTACCAGTCCTCACGTGATGTATGTTGATCCAAAATGGAATATACATGGTGTTACAACTGGTATAACCGCATATG
GTTCTTTAACAGTTACAATCACACTCCAAAGTACAGCAATTACAGCACTTACAGGCCCTTTGAATCCTTTATATAGC
TGGGTATCAGGTGGAAGTTCATCTGAAAACGGTGTCGCCACCTCTACAGATCAAAGTGGGAACACATATACTACTCA
GACGTCCATTGCTGATGAGGGAGCATATGTTCCTTACGAAATTGGTTATCATTTTAATTCTGCGTCTGCATCTTGGT
CATATTCAGACGTAACTTCATTTAACGTTGTAACTGAGAATGAGGAATACGCTACGAATGCGGTTTACAACGGGGTA
ACGGAAGCTGTGCCATATACATTTACAGGTGCGGTCTCTGTTGAATCTATCACATTTATTCCGGAAGCAACGCCAAA
TTTAGCATTAACCAGTGGTGGTGGATGCGCAGTTGAATATTATATTTCCAATATTGAACAGGCACAATCAACTTCGA
CGACTCAAACATTCACTGGAGCCCAGAATCCAACACAGAACCAGAACTGGTGGAATTCGACTTTAAATTTTACCCTA
ACCGATCCCAGCGGAGCTTTAAACAATCCATCCTTAGCCACAACCTCTGGCACAATTAACGCTGAAACAACGATGAC
ATGGTCAATAAACCACATTTTAAGCGCTGGGTACGGGCAGGGCGCTATGTTCGATACCGTATATTATCATGGCACAG
ACATCACGTCGGGTTCGTATAATCCTTCAAGCCAGTATAATTCATTATCGCAATCCACGTCCGAAGTAATCGACAGC
AGCACATCAGTATCGTACACTTTCGACGAGGCCATAAACTTCGCCCCGACCCTCTTATCCCACAACGTGGTCTGGAC
CGGCACATCAGATACCGCAAAAGTCTACGTCAACGTGTCTCAGGCAGGATGGACAACCGAACCGGAGCAACTCGTGA
TCCACTGGGGCAATTCCGCTTCAACCACACAGACCCTCACTGCGTCAGCACCCGCTCAGTTTGTTGCCACGCACACT
TATTCCACCACTGGTACAAAAGATATCTCGTTCCAATTGACAAATATACCCGATTCCTCATCCATAACAGCTGGAAG
CACATCCGTTCCCGCATTGTCATATACCATATCTATCAATCCCTCGCCATCACCAGGGCAATACGCTCAATTGAACA
GCGCAAATGATATCTACCTCAATTTCTCAACGGTGCATGACATCGTATCCCGCGTAACCCTCAGTGTATCCGGATCC
ACTGTAGCGACATATTCACCGGATTCACAGACAGGAACACTAATAGCTACTCCTTCGCAGGTTCCGGAAATCGGTTC
GACTGCCTTTACGGCAACATGGATATTCGATGGTGCGAATTTCTCGTATTCTGTTACCTACGGAGCTCCGGAATACC
CATTATTGAACAGTCCATACATAACCACGATGTTCGGCACGAACGCCACCCATTCCTACCCGATCACCATCTCAAGT
CCTCCAAGCGGAACAGGAGATTATCAGCAACTCATAACCTTAAGTCCATCCACCACAGGTCACGTTCCAAGCTATTA
CGGCATCAACTCGCAGGCAAGCAATTTCCAGATTGCCCTTCCAAACGGCACCCTGTTGTATACCTGGATTCAAAATT
TCAACAGCACTTCTCTGACCATGTGGGTAAAAATGCCATACGGCACAAGCCAAGTTGAATTACAGGTATTGCCGAGC
TTTGAGAATTTACTATCTGCAACTGGATATATCGGTTTAGCGGATTCACCAACTGATAACGGAGCATATGTCTTTCC
TGCATATTGGTCTGGATATACCCCTATATCTGATTTCTATAACGAGTCTGGAATAAAGCTAACAGAATACAACTACA
ATGGGATAAGGGTAGTTGGTTTGTTTGGATATAATGACATCAATGCTCCAGTGCTGGAGTTTAATAAAGATATCCCA
GCGGGAAGTTATATTGTTTATTCCAAAGCTCTGCATCTGGCGAATCTTACAGGTGGATTAGGTGCATCTAATGGGTT
TGTGGGTTTGTATGGCACAGATGGGACAGGTGTAGATGCGGAAATGGGTTTTGGGTCCTCATTTATCTCTTTAGCTT
ATTTGAATTCAACTGCTAATAAGGTGGACATAAACCAGCAGGGAAAAGCAGTAAGCGAATGGGTTAATGCAACTTTA
ACATATAATGCTACCCAATATTTTTCATCTACTTTTCAAGCTGGTTCAACTTCCTACACAGGCAGTGCGGATGCTAA
CACTATACCTATAGATACATCTAACATGCAAATAGGTGTTGGTCTTGTCACGACGACTGGGATGATGAATTCCAATC
AATTTACCAGTATGAGTTTGCCATTTCCTATGTAGCTTCAATGCCCACCTACACCATCGGTTCAGTAACCATCGGTT
CAGGTTCAGTATTTCAGGCGAATGCGACCACATTCACAAAAGCCCCATATTCCACGGTCGCCAATTATCAACCGAAC
CCGCTTGAGCAGATTTACACATATAATATATACGATTCGTTCAGCGACAATTACATAACGCTCTTGAGTAACTCCTC
ATGGACATTCCTGCAGGATTCGGGATATGCCAGTTATGATCCAACGACGTACGCCCTTACCTTCAGTGGAGTATCCG
GCATAGGCGAAGTATCGGCGTCATATTTGGAGCCGTCACAACAGCTTGGGCAACCGTCATTTGTCACCCTGTCCCTG
TCCGCATATGGATCATCACTGTTCAGTGGATTCAGCTATTCCGTTGCCTACGTTCCGTATTCGTCAAATACAACCTA
TTACGATAATGGGACCTCGTCGGATGTCCAGATTCCGTTTGGTTCGACCATGAACATCACGGTATTTGATGCGTGGA
ATCAACCCGTATATTCCAAGACAGGCATACTTGTTGACCAGTCGACGATGATAGTCCCGATATCCCTGACCAATCTC
GTTACGTCGGTATCCTTCCAGTTCATCAATACCACGGCGGCACAGGTGGGCATCACGTCCAATGGAATCACCCAAGT
GGAAAACGGTTTTACCTCGTTTTATGCGGCAAATGATACCTCTTATACGTTTTCAGCATCCATTTTTGACGCATCCT
TAGGAAAGAACGTCAATTACACCGGTTCGTTCACTCCTAAGAAGCCGTCCTATACGATCTACATCAATGCAACAGCA
CCACTCGCTGAAATAATCCTGAATGCCAACGCCTATCCTGGGTCCCAGCTGGGAGAGCTCTCTTCATCCGGATCGGA
TCGGGTATTGATGACAGTGGATGGTATTCCAACGGATTTGGGATCGTCCTACGTGGGCTTCATTGGACAGATAATCA
GTGTGCGTATTTTCACGGTGCTGAACCAGACCCTCGCAACCGCTTCAATTCCCCTTTCCCTGCCATTCCAGACACAG
ACGATCAATATCCAGCGACCAAGCTGGAATTTCCAGATCAAGAATGCGGAGCAGGTCTACAATACAACATCGCCGTT
GGCAACTGAAATTATCAACCTCACGGACACATCCACCAATACCAGCTACACGACAAGCGACATGGTCGGGAACGTTC
TATCAATGTATCTGGCATCCGGATCGTATCACATCTATTTGCATGATAATGCAACATTCAAACAGAATTTCACACTA
TCCAATAATACTTATTATATCGTATTCGGTCAGCAGCTGCTTACTGTGACGCAATACAATCAGATCATCAGCAAGAT
ATACGGAAATACGGCGGGATTACAGATCTTACCCATTAATGCACCGTCACAGATCGTCCAGGGCCAGAAAACGCAGC
TGCAATTCCAGGTCGATTACGCAAATGGAACACAGATATCTGGTGCTGCACTGTCATCATTCTTGGCGAATTCCACG
ATCACCCTGAACAATGGATCTGCATCGCTTCTTCTCACGCCGTATGTATTGAATAACATTCTTTACGCCAACCTTAC
CGGTCCATATCCTGGTTCTTACACCGTGAAGGTTGTGGGAGGCTTCAATGATGCCGGAACACCAGTGGGAGGGCATT
ACTCATATCCGCTGACATCCATAACAGTTGCACAGGATTCAGGATTACATCTCTATATCTCTGGGCCCACGACGGTT

CAGGCAAACACGACAAATACATACTACCTGCAATTTCAGTATGGTAACGGTTCATTGCTCAGTTCTACTGCATCATC
CAGTCTCCTTGCCAATCTCACAGCCTCGGTTGGTTCAACACATCTAACCCCGTCCACGATCGGCAACGGTGAATACA
GTGTCACATATTCTCCAACGACACCCGGGGAGTTCACGATCCTCGTATCTGGTTATTTCTTACACAGCGGTGTGAAT
CTCACGACATCCTCGTTCTATCCTGTTTCAACCATATCACAAACCTCAATATTGGTTGTCTTGCCAGTCGATACGCC
TACGACTGTCCTTGAAAACACAACAACATCATATGTCTTCAATGTGGAATTTCCCAACGGCACACTCCTGAACGATT
CGCAGCTGTCGACGATTGTTTCACATTCCACCCTCACACTGGAAAACTCCGCAGTCATAACGCCAACGATACAGGTA
TATGATGCGAAGATATACATCAATTTCACGCTCCCATCGACTGGGTATTACACGCTTACATGGCAGTCATCGATCAC
CCACGGACAGAAATACACGCTGATATACTCCCAGATGGTACACGCTTTCGCTGTGATACAATCATCCTATGGGCTTC
AGGAAGCGATAACCATCCCAGGAACAATCCAGGTATCCCATGCCACCACCGGAACCATCATTTTCACCCTGCTGAAT
TCATCCACGCCTAACACCATGACTCCCAATGCAGCACAGACGGGATACTTGATCGCAAACACAACCTTGGAGCTCCT
ATATCATGGGAAGTACGTGGCAATCATACTCCCATACTACGAGCAACCAGGGGAGGCATCATTTACCGTCAATGAAT
CAGTCACCGGGACCGGTTATTCAATTCTCTCGATCACGCATCCCACTGATATCTCCGGAAAGAATGTGTCATTCACA
GGCGAATCCAGGTCTTTTATCGTATCCACAGTGAACCCGGCCAACCCCCCAACACCAATAACAAGCCAGAGCATAAC
GGCATTCCTCACAAGCCCAGCAGCCCTGGCCATAGAGTTCATCATTACCCTGCTGGGTTTCCTCTCATGGTTGAGAC
CAAAGATACGGAAGAAGCACCAGGCAGAAGAATCAGATCTCAATCAGGCCACATTGGTGGTTGAGGGGAACATCGCA
ACCAAGATACTGGCTCGCCAGCCCTTAACGGAGATGGAACAGTTTTGGTGGGACGCAATTCCCAAGGATGTGAGGAA
GAAACTCATGAGGGAGGGGACAGCAGGTCAGAGATCCCTGTTCCCACACGTCAGGTCCCGATCACAATCAGGAGGAA
ATAAATGAAAATTAGACAAAAATTAAGAGGAAAATACATGAAAAGAAGAAATAAAGGAATTCTCATAATTGGCATAA
TGATCCTTGCCCTGTCAATTTTATCGGCTGCAGGATCACCTCATATGCAGATACCCACACATAGGGCAACGCTGGCT
TCTTATTCGGTACCAGACAGCACATCAAATAATACATCCAATGTCACACACAACGCCAGCAATCCCCTCTACACCGA
GATCAACATGAACCTATCGGCTGATTATTCATCCCTCCAGCTGTATGTAACGGCACAGCCAGGTATCGTATCACCGA
TGCTCACCCATTTCTCTGTCACTCTTCGTGCCCCGGGAAACTCATCATATGAGATAACCGATTCAGCCCTTAACGGA
GGATCCAGCATACTTAAGCAGGGCACATTCTCTTACGTGACAACATTCAGTCTCATCGGTAATAATTCAGGAGTGGC
AAATTTCACATTCTATATAACATCCGGGACGCTGCACCACACTTCCAGATTTTCGTATCTGCTCATATTCATGACCC
CGGTCAATTATATCAATTACGAGAAGGCAAAACTTGCCCCACTGGGCCAGCTCACCTACACCCAGGGAGCAGAGATA
GGAGCAGGAGGAATCGTCCTGTCGCTGGGACTGTACAGGTTATTCCTTCCAGTGATGAAGAAAAAGATCCAGAGGGA
TAACAAGAAAGGAGGTTTGATATTACTTGTCAAAAACTGATGAAACTGATGAACAAAACGAACAGATGAACCCAAAC
GAGGGAGGGCCATTCTCACCAAGAAGATTGATCTCAAGAAACAAGAGGGAAGTGGTCCCGGATGCAAGCAAATTGAT
CATCACCAATATCCAGGACAAGGAAGGATGGGGACCTGCCTTCAGGAGATGGTTTTTACGGAATAATGGTCCGGTGG
TCTTGCTCGTCATGTTCATCCTTGCCATGAGCTATTATGTCCCCTCAAAATTCTACCCTCAGGCAACGGAACTTGTA
TTCGGTATGGTATTCTTCGGACTCATCGGATCCATGACAGCTGCATACGAATACGAGAAGAGGCAGGACCAGTCACA
CAATACTTTTGTGGATCTCATCCAGATTAAGCTGCATGAGACTGTTGTGGGCAGGGAAAAGGGAAAGGACGGGAACG
TCGAGACAAAGGTAATAGTTCCTAAAATAACAGAGAGGGCTCTCTATCTGGCAAAGGACTATATGTTCGAAGGAAAG
GCAGGCGATCCATATAAGATCGAGAAATACAACTGCCAGATCGTAGGGACATTTGGAAAAATCATCGATGTAAGTAC
TGTGGATGAAACTGGAAGAATCATCATAGGTGAAGGGGATGGCGACCTTCCGTCGGGCCTTATCGTGAAGGTGGCAT
ATCCATCCAGATCCGATCTCTCGGATGAGGTGAAAAGGGCCGACAAGCTTCGGAAGGAAGGGGCCATAAGGGAGGAT
CAGTTCCTGCAGTTCAAGGCCATGGTGGATCAATACAACCTCTGGAGGGATAAGATCATCGAATACGCTCAAGTAGA
AGGGATAAGCATGATCGACATAGACAAGATGAACAAGAAGCAGAGATCCTTCTTTGTCGCACTGGCCAAGGACAGCA
TGAGATACTGGAACGGAACTGACCAGGATCCAATACCTCTCAAGGACTGGAAACAGATGGATCCCTCGGTCGCACTG
CCTAAACTCCTGCAGATCATGAGGGATTACAACGACATAAGAGCAGAGAAGCTCGATCTCCTGATAAACAGGGAATC
AGATAAGATGGAGGCAGAACTGGGGGCAACAATAAATCTCATGAGGGCAACCGGACAGTCCGAAGAAGCAATAACCC
AGTTCCTTGTGGAGCAGAGGAAGGGCCTTGTATCCGTATCCAGCAGAACAGAAGAGGAAATGGAGGGAGACATCAAT
GCCGGATCAGATGAATAAGCACCCACCAGCAAGCATGGCCCCTGAGGTCAAGAGGGAACTCATATCCCTTCTGGTGA
GAACGGAATGGATCCGGATCTCGTTGATATGCTCGGTCACATGCAGCCCTACAAAATCCAACTGTTTGACCGAATAT
ACCAGGGATATGGACACAGACCAGACTATATAATCGAAAGGGCAAGGGAGCTCCTGCATGTCGATATGAACAAGCCA
CACATCAAGAGGCTCCTGGAGAAGGATGAGAATAAGGGCGTGTGGGATGAGGAGAAATTCCTGGAATTCGCATTCAA
ATATTACAGGAAAGCTCTGGCACTGGAATTCATCAAGACAGAAATGCATGAATACATGAACAGCCTTGCATCATCAT
CGGATAGGTCCAGCTTCATTGTAGCATCACAACCGGGGAAGGTGAGGATAGTTTGATTCCGGGAGATCTGACATTTG
ATCCAAATCAGCAGGAGGCCATACACAGACGATCTGCATGCCTTGATTTCATGATAGAGGTTGTGAGAAGGGGAAAC
GGCATTCATACCAACGATGCACAGATACTCACCCAGCTTGTCAGTAATCTATACCGTATCCTTCCGGTGCACAACAC
AAATCCCCAGGAGATAGGAGAAAGGTTCGAGGAAGATCGGTTCAAATACGTGAAGAAGATTGAGAATCTTGCACAGG
ACCCAACACTCGAATCAGACGAATACGCATTCAAGGTATGGGCAGAAACGGAAAAATTCATAGAACGCTGCACAACA
TTCATGAATGATACTCTATTCGTAAAGGTGCAGAAGAAGAAGAAACCAATAACGGAGGGAGCAGGATATGACCAGAG
ATCGTGATTTTGTACTGCCAGAATATGTCACACTGGATAGGGACGGCGAGAACGGGCATAATGTCATCTTATTGGGT
GCTCCCAGGCAGGGCAAGACATACCTTTTGAAATATCTGACATACAGGGCATGTATCGGAAGGGATCTTTACGATGA
CGATAAGAGCAACATCGAGCCTCAGGCATGCATCTGGAGGGCAAGGACCCATGATAGGTATCTTGAATTCTTCAAGC

TCGGCATCGGTATTCTGGCGATCCCATCAGGAGGAGATTTCACACTTATCCGGATCAATAATGATGATTCCAAGGTG
GAGATCACAGTTGATGACCTCGAAGAACAGGGAATGGAATACTTTGTCTATGATTCCCCCCAGGATGTCGTGGATCA
TCTCGTTCCTGGCAAGGTAATATGTATCCTCTTCCCTGGAAATCTCATGGAGGAAGCCGAATTCTACGCTGATCTTG
GGGAGAAAATAAATCTGAGGCAGTCCAGGCAATGGGTACATATGGTGATCGATGAGGCCGGTGATCTCTTCCCTCCA
TTCAACCAGCACACAATGAAGACCCAGGAGAAATTTATAAATGCAGCCTCAGACTTTTCCAAGAATCTCGTCAACTG
CATCATAGCCTCTCACGAGATGTCCGGACTGGACTGGAGACTTGGCCAGAAATTCCCCTGGACCATCTACAAGAAGG
GAGCAACGAGAAGAAAAAGAACAGGTGCAGCACCTCAGATACTCCGTCAGGATTATATCAATAGGCTTGAGAAAAAG
GAGGCCATCATAACCTTTGGAGCATTCTACGATAAATTTACTTTTCCCATGATACCCGAGGATAAGAGACTGGATTA
CAAACTGACAACCATAAAGAGAACGTTCGTAATAGAGAGTGAGGGAAACGTAACTTTCAGCGAGAAGGTGGGGAGAT
AATCACCAATCGCCGAACTTTTTCTCTTCAGGAGAGCTGAGATCTCAATTTGTATCAGGAATTGTGGCCAGAATTTG
TGGACCACAAATTATGATGAAACTAACATGAAAACGGCAGGTCATATAATACATTGAAGCCCTGAAACAAGGGCTTA
ATTTGATTATAAACTCACCTTAGATAATTATATCTGGCCAATTCAGGCCTAAATGGAAAAAATATTAATTGAACATC
ACTTAAGTTACCTTATGGATATTGGTTACATCCGGATCTCAAAGGGATCTGAAAATAAGGAAAACCAGGAAGAGGCC
ATACTGAAATATTTTCATATCACAAGGAAGGAGGTCAAGTTCTTCTCGGATGTTATCACCGGAGAATCCCCAGCAGA
TGAGAGACCAGGTTTCATTGAAATGAAATCATACATGGATGAGGATCCCGATGGGAAACTCTATGTGTATGAGATAT
CGAGGCTCGGGAGGGACCACGCCGAGACGGTATGGTTGGTATCTCAGCTAAATAAGAATTACCCAGGGAGGAAGATC
TTCTCTGCATCACCAATGGAGTCATGGATGAATATAGAAA

>AMDV3b_Contig17817
AAATACGACTTCACGAGATGCCAGAGTAAATTGGAAGATATAGAAGTGATGAAATGAGTCTCGAAAAATTATCAAAA
CATGAGGAAAGAATAGGGGACAGGTTCTACTGCCAGCGCTGTTTTCATGACAGTGAACTGCATGATTTCAAAAGGAA
ATGGAAAAATAAAAGCCGCTGCCACGGATGGAAGCCCGGAAAGACCTTCGAGGAACGTTGTGAATGTCAGGAATTCG
TATCAGTTAAGGAGGCCCTGAAATGATAATAACAAAGAAACAGGCACAGGCGTGGAATATATTTCGGAGTCAGCTCT
GCGATTTTCATGAGCGATTCCCTGATCAGACGGCATCCATGTACTCATGTAGGGACTGCATGATAATCCAGGACAAA
TGCGGTTGCCGTGAAGATGAAGGATGCGAATATTGTGAGCCTGATATGGCAAGATTGGCAAGGTCAATCATAGGGAG
GAGATGAATTGTGAGAGGGAAAAATATTATTGAGCAATAAGAAATAATATAAGCATGAAATGATTACGTATTTGACA
ATAATGAAATCAGTATTCATGTACTATGGAGGGAAATTCCATCAATTGAAAGACATAATTTCGATATTACAGGATCA
CATGAATTCATTTGATGCAGTTGTCGACGTGTTTGGTGGTTCTGGTAAGGTTCTCCTCAATTTACCTGATGAATGGA
AGAAGATGAAGATCTATAACGATATTGATTCTGATCTGTATGTAACGTTCAAGGTTCTTCAGGATCCTTTGAAGAGA
AACATGCTTATCAAGAAGCTCAGGGTTGCATTTGCCCATGAGAAGATCTTCATGGAGATGAGGGATAAAAGATATCC
TGGAGATGTGGAAACGGCCTTCAGGCTGATATACTTGCAGACGTATTCATTCATGGGTGATGGCAGATCATTTGGAA
GGAGGTTCAAGGGCAACAAGCAGTCAAGGTTCACCATAGAGAACTTTGTGTATGTGAGGGACTGGACCATAGAGCAT
ATGGATTTCCGGGATCTGATGAAGAGGTACTCGAAGCCCAGGGTGTTGTTTTACCTGGACCCCCCATACCTGTCGTC
AGGAAAGAAGTACAGGCACAGTTTCAGCATTGATGACCTCAGGGATTTGAAGCAGAAAATGGATGGACATCCAGGTT
CATATCTCCTGAATCTGTCATCATTCGACGAAGGTATGGAGGAGATATTTGGAAAGCCACAGAAAGTGATCGACTAT
GCCAACCAACTGGACCACTACGGAAGGAATAGATGGGGATGTCGGGTATTGGTGGAGGTTTATTTGAATTAAGAATT
ATAGCGACCTTAGAACCTCCGATGCTTCAGCTGAGGAGAGTATGTCAGGAGAAGATGGAGGTCATTGAGGAAATGAA
AAGTGAAGAAAAAATAGTTAGATGTAAGAACTGCGGCATGGAAATCGAAAAATGCGATTGTGGGTATGTAGGCCCTC
ATTGGGTACATCGGACCGAAGGCAATAGGGAGTTATTCTGTTCAGGATCTCACACTGTGGCAGAACCAGAGGAGGCC
CTGAAATGAGCCTCCATAATGGAACTGCCGCAGAAACGGAACGAAAGAAGATGCTCGAAGCCGAAGGGCACGTAATC
CTGAATCTTGGACAGAATCAGTACGGAGACCTGGAGGATGTCTGCCATAATGTCATCATCGAGGTAAAGACATCAAA
GCATTCAAGGCGTGATCTCTCAAGGAAGGAAAAGCTCCAGATGGAAAACCTCCTGAAGCTGCTTCCGTACAGGGAGA
TCAGATATGACATCCGCTTCCATGGAAACAATCATCATTCTGCCATCTGGCAGGTTGAATTCCCCGATAGGATTGTG
AGTTCATTTAAGATTAGGGCAACTATGAACCCCCCTGAAAATAAATCCCCTGAAATCAATTCTAAATCAATGACAAA
TAAAAATAAGAAGGTGAAACCATGAAAGAACCAGTCTCTGTCGCAATGAATAAGAAGGATAGAGATACAGTAAGTCG
AGAAAGAAAAGAAAGAAGGGAAGGTGCAGAAATGACAAAACTTAATATGCACACAGAGAGTAGACGGGGCATAATCG
CTCCGCCAGGAGAGGAGGTCAACTACCAACTGCTAAAGCATGTTGGCTTGTAACTCTTATCAGGCTACAATAGGCTG
GCTGACAGCAGCCCTGTTCCTG

>AMDV3b_Contig17486
CATTAGAAGGATTCGCACGCAGCATCACCAGCAAGCACATAATAAGAAAAGAAGGAAACAAAATTGAAAAGTGAAAA
CATAGAAAACATGGCAAAGATTAATAAGTTAGATACAAGTATCAGGAGCATGGAGGATGAGACCAAGATAAGGGTCA
GTTCGATCACCAGGAAAAGGCTCTCCAAGGCTAAGGCTCTGGGAGAGTTCAATTCATACGATGATCTGATAAACAAC
CTGCTTGATCAGGTTCCGATAACTGCTCCGCCAGGAGAGGAGGTGAAGAAATGAATACGTTTAGTGATGAAGAAATA
AAGGGACGCAGGAGAGGATTGTTCAGAAAATCGAAACTTATATCTAAGGAAAAAAAGACAATTATACCGGAGCAACC
AAAAGGTCTGAATCCGGATAGATATTATACTCAGTATACTTTGACATACTTAGAAAATAATTTCAAACGGCTTGGGT

ATGATTATAAAGGAAGAGAAAAACAATTTATGTTCGAAATGATAAATAAAACAGGACTACCCATAAATTTACAGTTT
CTGAATGACAATGGTGGTATTACAGGTGAAACTACAATCCATGCAAAATCAGGGAATGCAACAATACAATTAAGGAA
TTGGTACCCGTCTGTATCGGGGATAGAAATGTTTTTGCACGAGCTTGCACATGCTATAGATGGAATTTTATACTGGA
GAGAGAACTCTGGAGATCTGTCCCGAATGAAAGAACATGGACAAAGTTGGAAAATGAGAGCCGTAGAACTTGGTGCA
GTGCCAGAGAGATCACTTATGATATGGGATCGAAACAATGAAGGTTATTTATAATGAATAAATCAAAAGCCATTTCG
ATTTTTTTTGCTTTCCTGATGCTTTCCGGCATCGTTGTCATGGTCAATTACATTCCTACTGATTACATCCACACTGA
TTACAACCATACTGTTTCCACAGCGACTGCCTCCCCAACGAGTTCCACGAACACGGTATACCTGAACTCGACGGTTG
CCTACTGGCAGAATGTGGGGATAGCAAACTCTGGCACAACTACTACAACGGTACCATATGCTGAAAATAGTTCAGAG
AACACTGCTAATTTTGGATTTGCAGTAGCCATGAGTTCAAACAATTTGCAGGCATCTGCATCGGGGACAATCGGTTC
ATATTATTATTACCAGGATTCAGCCATTACTCTTTATTCTCTTGCCATTCCCACCCTGTCTTACGCTATCACGGATA
GTTCTACCCACGTCACAACATATTACGATAGTGTCGGCACGGTGTATGCCAACATATCTGTAAACGGGAATACTCAA
TCGTGGAGCCATACTTTCAATATAAATAATGATGTGGGATCAAGTGGACCAAGTGGATCCACTGAACCAGCTTACAT
GTATACGGAAATATCTCCCCATTGGTCGATGGGTTCGGCAAACCCACTTACTTTTTCTATTTCTATCACCGAAACAA
ATGATGGAGGCAATTCTGCTTATGGTTTATATGAGACAGGCAATGCATATACATTTGCTGGTAATGCAACCACAAAT
GCAGTGGAGGGGCTTTCGGTGGCACCAACTCAGACATCGGTTCCCTTTGTTTACGGCTGGCACTTCAATTCAGCTTC
CTCAGACTATTCCCCAGCAGGTGATCTGACCTCATATAATGTAGTTTGGTCGTCAAACTATGCCGGTGAATTCACCA
CCAACGGACAGACATCGGCACAGGAAACTTCAGGTACAATAGGCGGTTCTCTGTCATCTTTTTCAACGTCCTTCAAC
CCCAATGGTAATGATCCGCCAGCAGTCAGCAATTCAGTATCCAGCTGGACATACTCTTATACCCTTGAATCCCAAGT
CCAAGCCCAGCAAGTCACAGGCTACGCACAATCATACTCAGGTGCTCAAAACCTATCACAGAACCAGAACTGGTGGA
ATTCGACTTTAAATTTTACCCTAACCGATCCCAGCGGAGCTTTAAACAATCCATCCTTAGCCACAACCTCTGGCACA
ATTAACGCTGAAACAACGATGACATGGTCAATAAACCACATTTTAAGCGCTGGGTACGGGCAGGGCGCTATGTTCGA
TACCGTATATTATCATGGCACAGACATCACGTCGGGTTCGTATAATCCTTCAAGCCAGTATTATTCATTATCGCAAT
CCACGTCCGAAGTAATCGACAGCAGCACAT

>AMDV3b_Contig17792
TACGTCATGGGAGCACTGGTGACCGGATCACTGGGCCTTGCAGTATTCCTGGGATCCATAACAGCCCATTTCACCCC
CCTGACCCCATATGCTGCCATAGCAATGTGGACAGGCATCTATATCCTGGTGATGATAACAGTGCTTGGCCTCGGCA
TATACATGGAGGTATCAAATTGAATGCAAGGGAAGAACAGAGGAAGGTAGAAGAGGACAGGAGAGACAGGTATCTCT
CGGCATTAGAAGGATTCGCACGCAGCATCACCAGCAAGCACATAATAAGAAAAGAAGGAAACAAAATTGAAAAATGA
AAGCATGGAAAACATGGAAGGCATGGCAAAAGCTAACAAGTTAGATACAAGTATCAGGAGCATGGAGGATGAGACCA
AGATAAGGGTCAGTTCGATCACCAGGAAAAGGCTCTCCAAGGCTAAGGCCATGGGAGAGTTCAATTCATACGATGAT
CTGATAAACAACCTGCTTGATCAGGTTCCGATAACTGCTCCGCCAGGAGAGGAGGTGAAGAAATGAATAAAAAATAA
TAGCACTCTTAGGAGTTGCTTTGATGCTTGGAACTGCTTTTGGAACTATGATATTGTATATGCCGAGTTTCTCAGAT
GGGAGTAATAATCCTATATCTTCACCTGTGACTGCATCTACCATATCAAGCTCGTATTCCAACACAGCTTACCTGAA
TTCAACAGCGTCATATTGGCATAGTGTTAATCTCTCTAATACCGGTTCTTCGAGTTTTGTCATTGGAATAAGCCAAA
ATAGTTCTTATTCTGCATATCGCTTTGCATCATCTAATTCGTGGGCTTTAGCTTCCAATGACCAGAGTGCATCTTGG
ACACAAACGTTTTCAGGTGCTCCTAATTATGTTTATCTTACCAATGCGAACTCTATCAGTATTCCAGACATATTTTT
TGCACTTGGTTCGTATGTCATTGCTGGGTATGGTGTTTCTTATGTATATCAGGAGTTAGGCACAGTTTACGCCAATC
TTACGGTTGATGGATATACTGAAACATGGTCACATACTTTTAATTTTAACGAACCGGCAGATTCTTATTACGCATTT
ATTGACCCACAATGGAATTTTTATATTAATACATCACTTTTTAGAATCTCTAATTTTGCTTCAGCCTCACTCTCTGT
TGTCGCAACAGCTGATGGAACTGCTTATCTTGATGCACCACATGATTTTGCCACAACAACGGCTAATGCTGGGGAAA
ACATAATCCTCACAGAAACAACAACCCCATCTGTTGATAGTGTCAACATGTTTCAGGGATATCATTGGAATCAAGCT
TCCGTAACATATTCTCCGGCTTCAGACTTAACTTCATATTCCATACATTTAAGTTTACCTTATTCCACGACTCAAGT
AACATATAATGGTAATACATATACAGGTACAACAGATGTTATAACAGGATCACTTAGCAGTTTTACATTTGATGTTA
ATATTCCAATCAGCCAACCACTTTCAGATATAGCAGCAGCGAGCTTTGTTACTGCTTCGTATTATCTTGAGGATCAG
GTTCAGGCCCAGCAAGTCACAGGCTACGCACAATCATACTCAGGTGCTCAAAACCCATCACAGAACCAGAACTGGTG
GGATTCGACTTTAAGTTTTACCCTAACCGATCCCAGCGGAGCTTTAAACAATCCATCCTTAGCCACAACCTCTGGCA
CAATTAACGCTGAAACAACGATGACATGGTCAATAAACCACATTTTAAGCGCTGGGTACGGGCAGGGCGCTATGTTC
GATACCGTATATTATCATGGCACAGACATCACGTCGGGTTCGTATAATCCTTCAAGCCAGTATAATTCATTATCGCA
ATCCACGTCCGAAGTAATCGACAGCAGCACATCAGTATCGTACACATTCGACGAGGCCATAAACTTCGCCCCGACCC
TCTTATCCCACAACGTGGTCTGGACCGGCACATCAGATACCGCAAAAGTCTACGTCAACGTGTCTCAGGCAGGATGG
ACAACCGAACCGGAGCAACTCGTGATCCACTGGGGCAATTCCGCTTCAACCACACAGACCCTCACTGCGTCAGCACC
CGCTCAGTTTGTTGCCACGCACACTTATTCCACCACTGGTACAAAAGATATCTCGTTCCAGCTGGTCAATGTTCCCG
ATTCTTCCCTGATCACGGCGGGTAGTGCGTCCGTGTCGACATTACCCTATCAGATCAGCATCACACCAACACCGACA
CCCGCTCCCTACACCCAGCTGAAAACAGGCGAACCCATATCACTCACGTTCTCCACGGTACATGACATTGTTTCAGG
TGTTACCCTCTCCATTTCCGGTTCTGTTGTGAAGGCATTCTCACCGGATTCGCAGACCGGTTCACTGACCGCCAACG

TGTCTCAGGCTGGGCAGACTGCCTTCACTGCCGTGTGGTCATTCGACGGTGCATCGTTCTCTTATTCGGTCCCGTAT
GGTGCGCCGGAGTACCCGACATTGAACTCGCCATACATAACAACGATGTTCGGCACGAACGCCACCCATTCCTACCC
GATCACGATCTCAAGTCCTCCAAGCGGAACAGGAGATTATCAGCAACTCATAAACATCCCCAATCCATCGAGTTACG
GCATAAATTCGGCAGTGAGTAATTTCGAGATCGCATACACCAACGGAACACCGGTATATTCTTGGATTCAATCCTAC
AATTCCACTTCGATTATGGTATGGAGCAAGATCGCCTACGGAACGTCTTCGCTTAGTTTGGAGGTGTTTCCGAGCTT
TGAGAATTTACTGAGTGCAACAAGATATATTGGAGAAGCACCACAATTAAGCTCGACATATGCAGAGTATAATAATT
TTAATTACACATTTACAAACAATACCACAAATGCCACATATTCATATCCCTTGACATTATCGGGCGTTCCATCTGGA
AATGGCACATATCAGCAGTTCATAAATCTTCCAGACCCTGTGAATTACGGCATAAACTCAAAAGGTTCAAATATACA
GTTTACCGCATCAAATGGAACAATCCTTTACGCATGGATTCAAAATATAAATAGCACATCAATGCAGGTATGGATTA
AGAATTATAATGGGAGTTCAACTATAATAATGCAGGTATTCCCGAGTTCTGAGAATGTTCTGAATGCAACTGGGCAT
GTGGGAGCTTATCCTACAATAACTTCTGATTATGGAGAATACTTTAATGCACCATTAGTATTTGGTAATTCAACGTC
ACCTAATGCGTGGGATTTTGCAGGAACAAGTTTGCCATCTGGACTTGTTTCCTCTGAATCTCAATCAGGAATTGTTT
CATTTAATAATAGTCTCAGCATCGTTATGAATCCAGATTTACCAACAGGAAGTTATGGATATGTAGATACATTGAAT
ACGTATGGGCAGAATACAACTTTTATCACTGATGCCAATACAAATTCTTTTGATACTACTAATCTCAGAGCTGTGAC
GCTTGGTCTGGTACAAAATATAAGTTACTGGTCAAATTACCCTGAAAACGAGACTGGATTTGGTTGGATGGGAGATT
CGTATAAGCAGAATACAATGAGTGTAGTACAGACAAGTGCAAATACAACATTATATAGTGTTGTAAGTAATAGTAAT
TATAATGTATGGGGAATTAATTACAGTGATTCCAAGATGTCGGCAACTTATAATGGAATAACCGTTGCTTCATATAA
TGAATCTGGAAATTATCACGCCGTTTTAAGCGATCAACTCAACCCATCCATAAGTGGGTATTCTTCAATAGATTTAG
TATTCACTTATTTAATAGCTATTCACAGACCATCTTCGATGCCCACCTACACCATCGGTTCAGTAACCATCGGTTCA
GGTTCAGTATTTCAGGCGAACGCAACGACGTTCACGAAAGCGCCCTATTCAACGGTCGCCAATTATCAACCAAACCC
GCTTGAGCAGATTTACACATATAATATATACGATTCGTTCAGCGACAATTACATAACGCTCTTGAGTAACTCCTCAT
GGACATTCCTGCAGGATTCGGGATATGCCAGTTATGATCCAACAATGCACGCCCTTACCTTCAGTGGAGTATCCGGC
ATAGGCGAAGTATCGGCGTCATATTTGGAGCCGTCACAACAGCTTGGGCAACCGTCATTTGTCACCCTGTCCCTCTC
CGCATACGGATCATCACTGTTCAGTGGATTCAGCTATTCCGTTGCCTACGTTCCGTATTCGTCAAATACAACCTATT
ACGATAATGGGACCTCGTCGGATGTCC

>AMDV3b_Contig17737
CAAGAAGTCAGGCATGGGCACAAAGGCTTTCCGAAACTTCTGGGCACCATACCTGAAGATAAGGGAAAGGGATGTAG
CCCACAGGATGGGCACATCTATCCCATTTGACCGCAGGATAATGAAGAAATTCTGGGTGCTTGACATGGAGAAGCTG
ACCGAACAGGAGGTTTCACTTGCAGGCATCCTTGAGAACGTCATAAACTGCGAGGAGGAGAAAAGAGGAGGAGATAA
GCTATGCCGAAAAATGAGGAAAAAATGGAACTGAAAACGGATGCGGAAGTTGATCAGGAAGTGTCAGACATGCTGTC
TGCCATAGAGAGAAAAAAGAAATGGCTTGTGTGCAATATTGATGGCAAAGATGCCCTGATTCCTGACACAAGGGATG
CCCTTGGGCATGATCTGTTTGGAGTTGAGAAATTCCAGTGCCATTCCCCAAGCCCTGATCTGATTGTTTTAGCATGT
CCCCTGAAATTCATCGCAAACATATATGCCTGGGGCATGTTCACAAGGAGCATGAAATATTCGGATGTGGACTTCAA
TACAATGTCCCTGAGACCTGGGATAATAGAGGGCAATCAGGTCCATATCAGCAGGAAAGAGGAGGTTCACCCATGAT
CCCCACAGTAAAACTCCTTGAGCCCCGCAGTGGAAGTGTGATCCCCGCAACCCTCAATCCCCGCCTCCTTGCCCTCT
ACGGATGCGAGATGCCCAAGCATCATGTGAGGGAGTGGATGCATTCGTATCATGTGAGGAGAGAGTTAAGGAAAGTA
ATGGAAGTGATGGACACCATCGAATTGGAAAAGAGTCTGGAGGTCCTGAAATGAACAGAAAAATTCTCAATTGGCTT
CGTGCTGGCTTAGCATTGCAGCTGATATGCGAGTTTGACGGATGCACAAGGAAAGCGGAATACAGCATCACGAGGTG
GCATGGCCTCCATATAGTGGTATGCAGGGAGCATTTCCCTCGGATCGGCGAACTGGATGCAATGGATGAGGAGGAAG
TGAAATGATCCACATAGGTGGCTACTGGGGGAGGAAATACGACTTCACGAGATGCCAGAGTAAATTGGAAGATATAG
AAGTGATGAAATGAGTCTCGAAAAATTATCAAAACATGAGGAAAGAATAGGGGACAGGTTCTACTGCCAGCGCTGTT
TTCATGACAGTGAACTGCATGATTTCAAAAGGAAATGGAAAAATAAAAGCCGCTGCCACGGATGGAAGCCCGGAAAG
ACCTTCGAGGAACGTTGTGAATGTCAGGAATTCGTATCAGTTAAGGAGGCCCTGAAATGATAATAACAAAGAAACAG
GCACAGGCGTGGAATATATTTCGGAGTCAGCTCTGCGATTTTCATGAGCGATTCCCTGATCAGACGGCATCCATGTA
CTCATGTAGGGACTGCATGATAATCCAGGACAAATGCGGTTGCCGTGAAGATGAAGGATGCGAATATTGTGAGCCTG
ATATGGCAAGATTGGCAAGGTCAATCATAGGGAGGAAATATTCGGAAAACCACAGAAAGTGATCGACTATGCCAATC
GACTGGACCACAATGAAAGGAATAGATGGGGATGCGGGTATTGGTGGAGGTTTATTTGAATTTTCATCTTCGCTATT
ATAGGATATGCGAATACTTATATACTTCTTCGCTGTTCCTATTTTGATGAATATGTCTGAAGTTTCCAGGGAGGCTC
ACCGCTTCTGGTCTGCTAATGGGAACGTCAAACGAGTGTCGGAAATATTCGGTTACGTCCGACGCTATCTCCAACGG
CTGCTGGCATGATACAGTTACAGGAAGCGATTTTGATCTCATTGCACACGGACAGGTTCGTATCATGCCACAGGGCT
GGGAGAGATTTTTGAGTTTTACTAAGGAGGTTCTTTGAAATGGAACATAAAAAAATAAGTGAGTTGTCAGGACTCAC
AAGGATATTTCCCAGAAGGGCATTAAATTTTTGGAATGCCAGAAGGGAAGCAAAAAGGCTGGAATACCATCCAAGAG
ATAATGGGAGCGCTGGCCAAATCGGGCCATTTCCTGAGGGAAAAAAAAGAACGAACATACAGGCACAAACTATCGGA
GGGATGGAAAATTGATGGATAAGATTAAAAAGGTAAGGGAACTCCAAGAAAGGCTACACAGGAAAGAGATCACGCTT
TCGCAGTTCGAGGAAGCCCTCGACGATGTCCTTGAATGCGATCCGGCACGCCCAGATTCCCGCAGGAGGGGGCAAGA

GTCCAGGTAAAATGTCAACACTGGAGGAATTTTGATGACTTGTAAAAAGTGTAACGGCAGCGGAGGCATAGGATACT
ATGAGAACCAATCCCCATTGGGAAGTGGACTGAACTGGCCCATGTATATGGAGGATGTATGTCCTGAGTGTGTTGAG
AAGGGAATCTGCCCTAAGTGCGACAAGCCAATGAAATGGAACGAAGAACACGATGCATATTTCTGTCCTGTTCACGG
ATGGGATGGTTGTGGGAAAATGGAGGCTGAGTTCTGATGCCAGATAATAAGAATTTTGAACCGATGGAAACCATGGT
TCTCCAGATTGGTACATTCTTCATGGTCATCCCAATTTCTTCATTCAGGAGGGAAGAGAAATTATGACTCACAGAAT
ACACAACGGAAACAATATAGAAGTCAAGGCGTACGACCTGTTTCCAGATCCTGAACAGACCAATCAGAAAATTAAAT
GGAAGAAGAACAGGAGATTCAAGGATGGCAAGTTTCCTCAAGGAAGGCCATATAGGAGGGGAAAGACATGGAAGAGA
TAAGTGGAAAAGAAAAGAAAGAGGAGGCCCTGAAATGACATCTAAGGGAACCGCAGCTGAAGGCCTTTTATTAAAAA
TGCTCCTGGAATATGGGATTGAGTGTGAGCACGGGTCTCCATCAAGGCCAGGGGACATCATAGTTCCACCAAACGTC
ATAATCGAGGTCAAGGATCCCAAGGAGAAGTCATTTAGCTTTAAATCACACTCCGCCAAAGGGGAAGCTCAGTGGAA
GGCTCTCAGGGAGAAGCAGGAAAAGTTCCCCTGGCTTGAGGTCTATTATATTGTGAGATTCGGACGGAAGGAATGGC
GCTGTTTCGACTTCCCACTGGAACCAAAACCATTGAAATTGAAAGATGGATTCGAATTGAAATTGTTGTTTAACCTC
CTACTTAGCCGGCAGGGAGCATTAATGGCAGCAGTGGGTGGTGATGCAGTATAGAGATCGAAAACATAGCCCAGCAG
GACATCCAGGCAAATGATCCAACGGACGTGATGATCCTGCACGTCCTGATGAAATATAAGATGCACCTCTCCGCCAG
GTCCATAAGCCGCACCATAGAGAGAATCTACGGGGAGGAAATCTCACCATACACGATCGCCATCCACCTCCGCCATG
TGCCCTGTGAATCCACTGAAACCAATATCCGAAAGTATTCTCTCAAGACATCAGATCATCAAATAAAAATAGTAAAC
CTATTAAACTCTGAATTATAACTATAGATAGTGGGGAGCATGCGAAAGCTATAGAGTCTCCCAAGACCATACTTTTA
CATGCTCCCAGGAACAACCAAAAGGAGGAAGGAATTTCCGGGATCAGCTCCTGATGCTTTACCAGGCAAGGTGGATG
CTTTTCCCGCCTGTTTCCGGAAACAACATGAAAAAATTGAGAGGAATATACGGAATAGCAATAGCCCTTCTCGTTAT
CGTCAGCGTCGCTCTTGGAGCGTACATATTTAGTGCTGATGCGAGTACAGGGCCATCCAGCACGGGCCTTAACACCC
TTGATGGATATGGATTAGCCAATGAGGCGAGCGCAGTCTGGGCACTGAACGATTCGGGCCACCACGCACACTGGTAT
TCTCTATCGTTTACCGTGGATAAAACAAAGAATAGTATAACGTTCTCCACGCCGCTGCACGAGAAAATATCCCTTAT
TATAGTGCAGACAAAGAACGAGAGCCAGGGCGTATACAACTTACTGCAGAACGGCCTGCTGTTTACATATACGACAT
TCAGCAGCACAGGCCTCAGCGGAAGCCTGTCAGACGCATACATGTACTTCGGTGAACCGGTCAATGCCACGAGCAGC
AATGCAATAGCAGATAAGGGAATAACCGATTACGCACTGAACCAGAGCCTTTTCGACGAACAGGCATCCTACTTCGG
GACCTCA

>AMDV3b_Contig17522
ATAGTTCAACCATACTCTCCATACCTCGTACAGGTCCTCAGGAAAGCCATAGAGAGAGGAGGTTGAGGGAAAGATGA
ACAAGACGACATTTCGGGATTTTGGTCGTTACTATTCTGCTGCTTGCCACAATGATCCTGTACCTTCTCATTCTGGA
GGCCCTGAAATGATAATAACAAAGAAACAGGCACAGGCGTGGAATATATTTCGGAGTCAGCTCTGCGATTTTCATGA
GCGATTCCCTGATCAGACGGCATCCATGTACTCATGTAGGGACTGCATGATAATCCAGGACAAATGCGGTTGCCGTG
AAGATGAAGGATGCGAATATTGTGAGCCTGATATGGCAAGATTGGCAAGGTCAATCATAGGGAGGATGTCACGAGAA
GATGGAGGTCATTGAGGAAATGAAAAGTGAAGAAAAAATAGTTAGATGTAAGAACTGCGGCATGGAAATCGAAAAAT
GCGATTGTGGGTATGTAGGCCCTCATTGGGTACATCGGACCGAAGGCAATAGGGAGTTATTCTGTTCAGGATCTCAC
ACTGTGGCAGAACCAGAGGAGGCCATGAAATGAGCCTCCATAATGGAACTGCCGCAGAAACGGAACGAAAGAAGATG
CTCGAAGCCGAAGGGCACGTAATCCTGAATCTCGGCCAGAATCAGTACGGAGACCTGGAGGATGTCTGCCATAATGT
CATCATCGAGGTAAAGACATCGAAGAACGGGAGGCGTGATCTCTCAAGGAAGGAAAAGCTCCAGATGGAAAACCTCC
TGAAGCTGCTTCCGTACAGGGAGATCAGATATGACATCCGCTTCCATGGAAACAATCATCATTCTGCCATCTGGCAG
GTTGAATTCCCCGATAGGATTGTGAGTTCATTTAAGATTAGGGCAACTATGAACCCCCCTGAAAATAAATCCCCTGA
AATCAATTCTAAATCAATGACAAATAAAAATAAGAAGGTGAAACCATGAAAGAACCAGTCTCTGTCGCAATGAATAA
GAAGGATAGAGATACAGTAAGCGTCTCGGTGATTCTAATAATGCTCGCCGCTGGCATCTATCTCGCCCTGAGATTCG
GCATCCTTAACGGATTTTTCAACATCATACAGCACATCATCAGATTGCTGTCAATGGGAGGGGCATGAAACCATGAA
GTTAGACTATCAGATGGACGGTGAAGTCATCAGATCAATCAAGAGGTCCAGGACAAGGACCCTAATCCTTAAGTATC
TTTGGGACAAATACCCTGAAAAAGCATATCTCTCTGAAATTACAAGATTCATCCATTCTGACGTTTCCAATATCCAT
GGAGCACTGAACGGAATGCTCCATGGAAGATGGGCTCTTGAATTATCATTGGTCTCACTGGGACTGGTTGATAAAAT
CCACCTGAAAAACCATGTTTACTACAGGTTGAACAATACAAGAATACAGGACATCCAGGATATTCTGAATCTGATAA
ACGCCAAATCGGTCATGGAGGTGTTAGAATGACCCAAATGGAGATAATTGAACCCATCCAGGCAAATGATCCAACGG
ACGTGATGATCCTGCATATCCTCACGAAATACCAGATGCACCTCTCCGCTAGGTCCATAAGCCGCACCATAGAGAGA
ATCTACGGGGAGGAAATCTCACCATACACGATCGCCATCCATCTCCGCCATGTGCCCTGTGAATCCATTGAAACCAA
TATCCGAAAGTATTCTCTCAAGACGTGACATCCTCCCACGCCTAAAGAGCGTGGGCTTCCTGCTTCAACCACTTCAC
TTGCTTTATCCCCTCCTTCCGAAGAAGGGATAGAGTAGAGGTGATTGTGTCCACAGGCGTAAATTCCCTACTGTCCG
TAGGTACTCCTGTTTTTATTAAACCGAATCTCTTTATGTTTATTGCTGCGTTCCATTCCCTGTCATGTCTGGTATGG
CATTCCGGACATGTCCATGTTCTCTGGTTGAGTTTCAGGTTGTAGACATATCCGCATTTTGAACACATCTTTGAGGA
TGGATCAAACCTTCCTATCTCTATTATATTTTTCCCTCTCACAATTGCCTTGGTTTTCAGCATTGTGAAAAAGGTAG
ACCATCCTGCATCTGCTATGCTTCCTGCAAGATGGTGATTCTTTACCATACCTGATACGTTCAGGTCTTCTGTTATT

ATTGTGTCGTATTGTCGTGTAAGGGCATCTGATACCTTGTTATGGAAATCCATTCTCTGATTGGCTACGTGTTCCTG
CAACCTTGCTACCTTTATCCTTGCCTTGTTCCTGTTTTTTGATCCCTTCCTCTTTCTGGATAGCTGTCTCTGAGCTT
TTTTGAGCTTCTTTGTAGATTTCTTCAGGTTCTTTGGATTCCCTATCTGTATTCCATCGGATGTTGTAAGGAAAGCC
GTTACTCCCATGTCTATGCCGGTTGAGGTATCAGCCTTTATTTCCATGTGTTCAGGTGCTTCCGTTGTTGTTTCAGC
CAGAACTGAAACATAGTATTTTCCCGATGGAGTCTTCGAAATGGTAAGTGAACGCATTTCACCATGGATCTCCCTGT
GTTTGAAAATGCGTATTGGTACCTCGAATTTTGGGATTATCAGACGGTTGCCCTCAATCTTGAAATGCTGGGGTATT
GAAAATGACCCGCCGTTTGCCTTCTTCTTGAATACCGGATATTTGCCCGTCTTCTTGAAAAAACTGTGAAACGCACT
GTCCAGATGCATCAGAACCTGCTGAAGGCTCTGGGAATTGATCTCATTGAGCCATTCCCTTTCCTTCTTCAATGGAG
GCAAAAGCAATGCCATCTCCTTGTACGACATACCTTTTCCCGTTTCAGCATATCTCTTATTTCGCAGATCAAGGAAA
TAGTTCCACACAAACCTGCAAGATCCGATGTGCTTCTCAATGAGTGTTTCTTGATCCTCATCGGGATACGTCCGGAA
TTTGTAGGCTCTAAACATGGTAGTTATGTAAACACGTTAGTGTTATTAAGATTTCTGTTCGCTCCCATCCCACCCCT
GAAGAGGTGGGTTTTCACGCTCGATGTGATAAAAATAGTAAATCTATTAAACTCTGAATTATAACTATAGATAGTCT
TGGGAGCATGTAAAAGTATGGCCTTGGGAGACTCTATAGCTTTCGCATGCTCCCAGGAACAACCAAAAGGAGGAAGG
AATTTCCGGGATCAGCTCCTGATGCTTTACCAGGCAAGGTGGATGCTTTTCCCGCCTGTTTCCGGAAACAACATG

>AMDV3b_Contig17821
ATTGAACAGTCCATACATAACCACGATGTTCGGCACGAACGCCACCCATTCCTACCCGATCACCATCAACAGCCCTC
CAAGTGGTTCAGGATACTACCAGCAACTCATAACCTTAAGTCCATCCACCACAGGTCACGTTCCAAGCTATTACGGC
ATCAACTCCCAGTCAAGCAATTTCCAGATTGCCCTTCCAAACGGCACCCTGTTGTATACTTGGATTCAATCTTTCAA
TAGCACTTCTCTGTCTATGTGGGTAAAAATGCCATATGGCACAGGTCAGGTTGAATTGCAGGTATTCCCGAGTTTCG
AGAATTTATTATCTGCAACTGGATATATTGGAGAAGCACCACAGTTATCAAGCACTTATGGAGAATATAATAATTTA
ATGGATGTTTTTAATAAATCACTTATATCTCAAAATTTAAGTTATACAAGGCAAGGCATGTATGGAAGTTCTAATTT
AATAGGATATATAGCAATTATATCGTCAATAAATGAATATGGATATATGCAATTACAAAATAGTTCAAATAATTATA
ATCCACAAATACATGATTCTTCAAATTATACGTATTTTGGAATATGGGACGGTGGTGGATTTGCAGGTTCTACAAAT
TTATCTATTTTTAGTTTCTACAATTAAAC

>AMDV3b_Contig17633
TTCTCGGTTGTCTGCGATGACGATGCGAAGGGCCTTATGGTTATCCTATTGGTGAAATAATTCACCGGTGGCGTTAT
GCCTGATGGCATCTGGTAATCGATGCTCTGATACGTAGTGGAACCGATCTGCGATGATGGTTCGATCAAAGTGACCT
GAACGGACGTGAATCCTGTGACATCCTCAATGGTGACGAAGTTTGAATTGTTCTGCAGGGGCAGATACGTTGACGGA
TAGATGTTCGACAGCCTCCATGTTGAATTGAATATGACCGTAACATAGTTCGTTGAGGGCGCTATGGGGATATTGTA
GGTGTCATAGACATAGGTGGCATTGTAGTTTCCCGTAGGCGCACCCGAAAATGTCGATGTCGTCGTGGCATTCGCCT
GAAATACTGAACCTGAACCGATGGTTACTGAACCTGAACCGATGGTGTAGGTGGGCATTGATGCAATATAACTGCTT
GTAAAGAGAGCCCTGACAAAGATGCCATAAGGAGGAGAGTTATCTGTAGCAATGGAAAAATAATTAAATTCTTCAGG
TGTCATTCCAGAAGCTGTGTAATAACTCGTATAATTAAATCCAAAAGTTGATTGAGTTGAGTTATATATATTCAAGG
ATAGAGTTGCATTATAATTTTTACCGCTAATGCCTGTATTTGAATAAAAAGAATTATTGTCTCCTGCAATATTGTAA
TTAACGTGATAGTTTCCCTGAACAATAGAGTAACTGTTCGATTGATATAAGGAACTTGTTGGAAAAAGACTTGTCAA
ACCAAATTGATCATAAGGGTCGGATGATGCAGATGATGAAAACATATTTCCATATAAATTATACGCTATTGGAAAAT
TTATAAGGGATATGTTCGATGCGGAAGTGTATATTCCACCATATACGGCGGTACTACCTGTTGGTGCTGTCACATAA
ATTCCATTGGATAGATATGTGATAGTAGCATTCGTTCCGAAGGTGTGAAACTCTGACGGTAATGAACTCAAATCTGC
AAAACTCTGATAGAATGGAAACACTTTTTTACCATTATCTTCGGGACTGTTCCATTTGCCCACATACCCAGTTGCAC
TCAAAACATTCTCAAAGCTCGGCAATACCTGTAATTCAACTTGGCTTGTGCCGTATGGCATTTTTACCCACATTGTC
AGGGAAGTGGAATTGTAAGATTGTATCCAAGTGTATAAGAGGCTTCCATTGGAATCGGCTATGTAAAAGTTAGATGA
TTTGCTATTAATCCCATACTGCGAAGGATTTGATAGCGTTATGAGTTGCTGGTAATCCCCTGTGCCTGACGGTGGCG
AATTAATTGTAATTGGATATTGGTGAGTCAAATTACTTCCTATAGACGCATCTGGATTTTCCTGTAGATACCCCGTT
GCACTCAACAGATTCTCGAAACTCGGGAATACCTGCATTATTATAGTTGAACTCCCATTATAATTCTTAATCCATAC
CTGCATTGATGTTGAATTTATACTTTGAATCCATGCATATAATAGAGTTCCGTTTGAGCTTGTAAACTGGATATTTG
AACCTTTTGAGTTTATGCCGTAATTCACAGGGTCTGGAAGATTTATGAACTGCTGATATGTGCCATTTCCAGATGGG
ACACCCGATAATGTCAGGGGATATGAATATGTGGTATTTGTGGTATTGTTTGTAAATGTGTAATTAAAATTATTATA
CTCTGCATATGTCGAGCTTAATTGTGGTGCTTCTCCAATATATCCTGTTGCAC

>AMDV3b_Contig17787
TGGTCGAACCAAACGGAATCTGGACATCCGACGAGGTCCCATTATCGTAATAGGTTGTATTTGACGAATACGGAACG
TAGGCAACGGAATAGCTGAATCCACTGAACAGTGATGATCCATATGCGGAGAGGGACAGGGTGACAAATGACGGTTG
CCCAAGCTGTTGTGACGGCTCCAAATATGACGCCGATACTTCGCCTATGCCGGATACTCCACTGAAGGTAAGGGCGT
GCGTCGTTGGATCATAACTGGCATATCCCGAATCCTGCAGGAATGTCCATGAGGAGTTACTCAAGAGCGTTATGTAA

TTGTCGTCGAACGAATCGTATATATTATATGTGTAGATCTGCTCAAGCGGGTTCGGTTGATAATTGGCGACCGTTGA
ATAGGGCGCTTTCGTGAACGTCGTGGCATTCGCCTGAAATACTGAACCTGAACCGATGGTTACTGAACCTGAACCGA
TGGTGTAGGTGGGCATGGCAATGTTAGGTATGCTAAAATAATATGTAGCATTCCAGTTATCCCTTATGATAACTGGA
TAAATAAGTGAACCTGTTAATGCTTCCGTATATGTCAATACCGTGCTGTTAATTGATACATATTGTGTAATATTTCC
ATTCGAAACATTTTCCCATATTGAATATAGCCTGTATGCACCATTGTAATTGCCAGTTGAAGGCGAATTATTAGGAA
ATATCTGACTTGTAAACGATGATTCACCATCTGATGTTACTCCACCTGATAAGCCACCAAAATTTTGAAACACCAGG
GGCTGACCAGTTGCATTGGCTTGAAAATGAAAGGATGCATATAGATATGAACGATATAATGTTCCATTTGCAGAAAA
ACTTGTAGAAACTCGTGAATCTTCCCTATTTCCAGAGGCATTATATCCATATTCGTATAATCCGTTTTTTAACGTAA
CGTTATGTATGTTAGTCAATCCTTTAGATGTTGTAAAATTAGTGTAAAAATTCATTGGGTTTCCAGAAAATACACGA
TATCCGTTGTCATACTCTGCATATGTCGAGCTCAACTCTGGTGCTTCTCCAATATACCCCGTTGCATTCAAGACATT
TTCAAAGCTCGGTAATACCTGTAATTCCACTTGGCTTGTATTGTAAGGCATTTTTACCCACATGGTCAGAGAAGTGC
TATTGAAAGATTGAATCCAAGTATATAGCAATGTGCCGTTTGGAAGGGCAATCTGGAAATTGCTTGCCTGCGAGTTG
ATGCCGTAATAGCTTGGAACGTGACCTGTGGTGGATGGACTTAAGGTTATGAGTTGCTGGTAGTATCCTGAACCACT
TGGAGGGCTGTTGATGGTGATCGGGTAGGAATAGGTGGCGTTCGTAGGTGCTTCTCCCACATACCTCGTTGCACTAA
ACAAATTCTCAAAACTCGGGAATACCTGCATTATTATAGTTGAACTTGAATTATAATTTTTGACCCATACCTGCATT
GATGTGCTATTTATATTTTGAATCCATGCGTAAAGGATTGTTCCATTTGATGCGGTAAACTGTATATTTGAACCTTT
TGAGTTTATGCCGTAATTCACAGGGTCTGGAAGATTTATGAACTGCCAATATGTGCCATTTCCAGATGGAACGCCCG
ATAATGTCAAGGGATATGAATATGTGGCGTTCGCAGACTGCAATACTGTAACCTTATTGCTGTATTCAGCAGGTGTT
AGATTCGAACCATACTGGAAAGTGATCGACTGGGAATAGCCATACTGATCGACTGCCTGCCATGTAACGAGAAGAGG
AGCGGATACGGGCTGGGTGAAATCGTATTTGACTGTTCCTGATGTCAGATCGGGCTGGAACCTGTCCTCAAGGACGC
CGTTGACTGTAAGGTTTATTATTGTCATCTTTGCCGATGTCTGAGTGCTGTATGCGAGAGACATCAAAACGGAGGTG
CCAACGGACGAATAATCAACTGGGCTTGGCGTGAGGGAGAAGGGGACAACATTCACCGTTGCTGCCTGTGTCGAAGA
GTTCAGCTCTGAGCCGGGATAAGGATTTCCGTTGGGATCATTGTGAACATACCAGTCGATAATTTTCTGCCCCGTTG
AATGGAAATAGAAGGAGGGCGACGCAAGGGATTCAGACGAAGTGGGCGATGAATGGATGTAAGGCTTGCCATTGGCG
CTTACGGCAAGGTTCTGTTGCTCCCCTAATATGACCTCTGTGGAATTCGCATACAGGTCAAGAGGCGTGAGGGGGTT
TACAGGCGTTGCTGCGCTGCTGTTCACAATGGAAGGATGGTAGTTGATGAACTCATCGAAATTGTATGAAACTGCGG
ATGAGCTGTCCACAACTACGGCTGTTCCCTGGCTAACCGAATTGTACTCAGCCGAAGGCGCAAAAGAGGATGATGAA
ATATCCGAACCAGAATAGAAAATCCTGTCAAAGGCCGGAGACTGTCCGTAGCCTACAGAAAGGATATGGTTTATGCT
CCATGTCATTGTTGTTTCCGCATCTATTGTTCCGGATGTCGTTGACAGCGACGGATTGTTCAATGCACCTGAAGGAG
GCGACAACGTAAAGGAGGAAGATGAGTTCCACCAATTCTGAAGCTCACTTGGGTCCTGAGACGAACTATAAGACTGT
GTTGTAGATGTTTTCTGAGCCTGAACTTGGGAGGAAAGATAGTAGTACACGTTCCACGTGGGACCGTTACCGTTGCT
ATCTGTTGTACCTGCCGCACCAAGAGGTGGATCAGCATTCGGAGAAAATGTAATCGAGAAGTGTGAAAGGTTACCTG
TAAATGTATATGGCACTGACCTGGAAAGCCCATTGTAGCTGGCTGTTGTGGATGCTGATGGATTCTCCGTCACAATA
TTGTAGGAAATCAGATCGGATGGAGGGGAGATAGTAATTGTAGTAGAGCCGTAAGTCCATCCGTAAGCCAACGGAGT
TGAACCACTCGAGGGTTCGGAAGAAACAATAGAATTGGAAGCCGATGAAAGAGTGTGGGAGGAACCCAGATAATTTC
CGTTGATTGTGTATCCACTATCCGGTGTTGCCTCCAGCTGGATAATGGCAGAACCGGCGTACCAGCGGTTGATCGCA
GGGATAGTGAGACTGATCTCAACATAGAAAGGGGCTAATGATTCGTAATTAAAAGTATATGAGGGGACTGAAGGCGA
AAGGGAGCTGCCGCCAGAAGTAATTACCTTCATCGAAACGGTACCAATATTATAATCGGAAGAAAGAGCGGACACAA
CATCATCCGGAGACGGAAACACGAAGTAAAAATGCATCAAAGAACCTCCGGCAAAATAGAAAGATGGCGTAGAAATG
TTAAGATAATCATTTCCACCAGCGGCTGTCCATGCTGTGCTGGTGGAGCTTCCAGCCGAATAGGAAAAAGGAGTAGC
CGAACTATTGGCGTTGATTGGGAGGGTAACAGTCGTTATGCCAGAGTTTGCCAGATCAACGGCATGCCAATATGACG
CTGTTGAATTCAGGTAAGCTGTGTTGGAATACGAGCTTGAAATGGTAGATGCAGTCACAGCTGAAGATATAGGATTA
TTACTCCCATCTGAGAAACTCGGCATATACAATATCATAGTTCCAAAAGCAGTTCCAAGCATCAAAGCAACTCCTAA
GAGTGCTATTATTTTTTTATTCATTCACTTCACGATCTCCAGTCGCAAGGAAGCCCACCTCTTCAGATGTGGGAGGA
ATTGCGACACTCTTATATATTGAGTGTATATATCAATATTGTGCTCAAGACCCTTCAGATAAAACTGCTTCCCCATG
ATGAACAGCAGGGAGCATTGGTTGATACCTTCATCAAATTCAACAGTGCCTGCAACTTTGTTTCCAGGATAGCGTTT
GAAGGGAAGCTGTACAACAAGGTTATTCTCCAGAAAATCACGTACAGGGACATCAGGGAGAGGTTCGGTCTTGCTGC
ACAGCTTGCCATCAGGGTCATAGCAAAGGTTGTTGAAACCTATAAGGCTGATAAGACGTCTTTCCATGAGTTCAGGG
ATTTTGGTTCCATTGTATATGATCAGAGGATACTGAGCTTCAAGGGTGTTGATGAGGTGAGCATCAGTACCACCAGT
GGAAGGGTAAGAGTACCCATAACCATTGGAAAATATGGTGATATACCTTTTGAAAGGATCAGGGGTCAGTGTGATCT
CATCAGGAAAAACAAAATGTTCTACCTCATGGTTGCCGTTGAAGTTCCTGAAGAACCCATCATCAAACCGAAGGACA
TCATGGGAGTTGACATGGGCATCGTCAATATTGCGGTGGATTCTACTGGGAAGTACTATTCCGGTGATGAGATCAAT
GAGGTCAGGGAACACAATGCTGATCTCCGTTCCCGTCTCCAGTCTGTAGGGACCAAATCCGCAAAACGCCATCTGAA
GAAACTCTCCGGAAAGGAGGGCAGGTTCGCCAGAACCACCAACCACATAATATCTAAAGAGATCGTTCAGAAAGCCA
AAGGCACCTCTTCTGCGATTGCCATTGAAAACCTCAGTGGCATAAGAATGAGGACAACCGTTAGGAAAGGCAACAGG
TACATCCATAATTCGTGGGCTTTCAACCAGCTCCGGTCGTTCATTGAATACAAGGCAAGGGAAGCAGGAATCCCTGT

CATTGTTGTGGACCCACACAACACCAGCAGGGAATGTCCCCATTGCCATGCCATAAGCAGGAAGAACAGGCTGGAAC
AATCCAGGTTCAAATGTATTTCCTGTGGGTTTGAGGGAGAAGCGGATTTCATCGCCTCCCTCAACATCAGGAACAGG
GCTGCTGTCAGCCAGCCTATTGTAGCCTGATAAGAGTTACAAGCCAACATGCTTTAGCAGTTGGTAGTTGACACCAC
CACTGCTCTGGACTTCTATCTGGAGGGGCTTCTTCGGCAGAATAACCTGATTGCCCTCTATTATCCCAGGTCTCAGG
GACATTGTATTGAAGTCCACATCCGAATATTTCATGCTCCTTGTGAACATGCCCCAGGTATATCCGTTTTGCGATGA
ATTTCAGGGGACATGCTAAAACAATCAGATCAAGGCTTGGGGAATGGCACTGGAATTTCTCAACTCCAAACAGATCA
TGCCCAAGGGCATCCCTTGTGTCAGGAATCAGGGCATCTTTGCCATCAATATTGCACACAAGCCATTTCTTTTTCTC
TCTATGGCA

>AMDV3b_Contig17307
TTCAGGCGTCACCCTCTCCATTTCCGGTTCTGTTGTGAAGGCATTCTCACCGGATTCGCAGACCGGTTCACTGACCG
CCAACGTGTCTCAGGCTGGGCAGACTGCCTTCACTGCCGTGTGGTCATTCGACGGTGCATCGTTCTCTTATTCGGTC
CCGTATGGTGCGCCGGAGTACCCGACATTGAACTCGCCATACATAACAACGATGTTCGGCACGAACGCCACCCATTC
CTACCCGATCACGATCTCAAGCCCTGCAAGTGGTTCAGGAAATTACCAGCAACTCATAAACATCCCCAATCCATCGA
GTTACGGCATAAATTCGGCAGTCAGCAATTTCGAGATCGCATACACCAACGGAACACCGGTATATTCTTGGATTCAA
TCCTACAATTCCACTTCGATTATGGTATGGAGCAAGATCGCCTACGGAACGTCTTCGCTTAGTTTGGAGGTGTTTCC
GAGCTTTGAGAATTTACTGAGTGCAACAGGATATATTGGAGAAGCACCACAATTAAGCTCGACATATGCAGAGTATA
ATAATTTTAATTACACATTTACAAACAATACCACAAATACCACATATTCATATCCCCTGACATTATCGGGTGTCCCA
TCAGGCACAGGGTTTTACCAGCAATTAATAACAATCAACAATCCATCTGATTATAATATAAACTCAAATGGTTCAAA
TATCCAGTTTACAAGCTCAAACGGAACTCTATTATATGCATGGATTCAAAGTATAAATTCAACATCAATGCAGGTAT
GGGTCAAAAATTATAATTCAAGTTCAACTATAATAATGCAGGTATTCCCGAGTTCTGAGAATTTACTATCTGCAACA
GGATATGTTGGTGAGCAGGGTACAGTAAACAATAATGGCATCCTTGTATTCCCATATTTTGCATATTTTTACGGCAA
TACAATTCACCAATCATCTATAGGCGGATTGAGTTATACAGGAACACTAACGAATAACGTATTACAGATAAATGACA
ACGGTTCAGATGCTATGTCCGTAAGCAACAATCTCTTAGCAAATGCTAATGATACATTTATATTTGGAGCAAATTAT
GATACAGGGACTTCCAGTCATTCTATATCACAGGGATCAACAGGGGGAACGTATTACTTAGGAACAGAGTTTTTTGG
AAATTCTGAATATAATACAGCTTGGTTTAATGGCACTCTTAATCCGGAATTAGTTTATTATGCTGCCAATGGAGGCT
CTGGTTCTTCTGGTGCTTCAACGAATACTTTTGAGAATTATATGCTTAATAATCATTTTACTTCAAGTGGTTATGAT
TATGGCTATTTCTTTGTAAACAACACATTGCAAGAAACACTATGGTATACAACAGGTCATGTTCCCGGCTTTGGGCT
TGGTAATCCAATTGGATTTTACACACAACAAAAGTCTGTTGCATATTTACAGTATCAGTATATTGCAATCCTTACAA
ATACAACAAACAATTCAATGCCCACCTACACCATCGGTTCAGGTTCAGTAACCATCGGTTCAGGTTCAGTATTTCAG
GCGAATGCCACGACGACATCGACATTTTCGGGTGCGCCTACGGAAAACTACAATGCCACCTATGTCTATGACACCTA
CAATATCCCCATAGCGTCCTCAACGAACTATGTTACGGTCATATTCAATTCAACATGGAGGCTGTCGAACATCTATC
CGTCAACGTATCTGCCCCTGCGGAACAATTCAAACTTCGTCACCATTGAGGATGTCACAGGATTCACGTCCGTTCAG
GTAACT

>AMDV3b_Contig17646
AGGTGTTTCCGAGCTTTGAGAATTTACTGAGTGCAACAAGATATATTGGAGAAGCACCACAATTAAGCTCGACATAT
GCAGAGTATAATAATTTTAATTACACATTTACAAACAATACCACAAATACCACATATTCATATCCCCTGACATTATC
GGGTGTCCCATCTGGAAATGGCACATATCAGCAGTTCATAAATCTTCCAGACCCTGTGAATTACGGCATAAACTCAA
ATGGTTCAAATATCCAGTTTACAAGCTCAAACGGAACTCTATTATATGCATGGATTCAAAGTATAAATTCAACATCA
ATGCAGGTATGGGTCAAAAATTATAATTCAAGTTCAACTATAATAATGCAGGTATTCCCGAGCTTTGAGAATTTATT
TTCATCGACAGGATATTTGGGCGAGGCACCACAGTTATCAAGCACTTATGGAGAATATAATAATTTAATGGATGTTT
TTAATAAATCACTTATATCTCAAAATTTAAGTTATACAGGGCAACGCATGTATGGAAGTTCTAATTTAATAGGATAT
ATAGCAATTATATCGTCAATAAATGAATATGGATATATGCAATTACAAAATAGTTCAAATAATTATAATCCACAAAT
ATATGATTCTTCAAATTATACGTATTTTGGAATATGGGACGGTGGTGGATTTGCAGGTTCTACAAATTTATCTATTT
TAGTTTCACAATTAAACTTTGTTTCAATTGTTGGACTGAAAAATTCAAATGGATACAATTGGTATTATCAGGCTAAT
GATATAAGTGGAAATTATCAATATAATGTATTCAACAATGGAAATCATTGGGTCATAGATAGTGGAAGTATAAGTGT
CACTTTAAGATATTTAATAGGCATAGATGCACCAAATGGGGCAATGCCCACCTACACCATCGGTTCAGGTTCAGTAT
TTCAGGCGAATGCCACGACGTTCACGAAAGCGCCCTATTCAACGGTCGCCAATTATCAACCGAACCCGCTTGAGCAG
ATTTACACATATAATATATACGATTCGTTCAGCGACAATTACATAACGCTCTTGAGTAACTCCTCATGGACATTCCT
GCAGGATTCGGGATATGCCAGTTATGATCCAACGACGTACGCCCTTACCTTCAGTGGAGTATCCGGCATAGGCGAAG
TATCGGCGTCATATTTGGAGCCGTCACAACAGCTTGGGCAACCGTCATTTGTCACCCTGTCCCTGTCCGCATATGGA
TCATCACTGTTCAGTGGATTCAGCTATTCCGTTGCCTACGTTCCGTATTCGTCAAATACAACCTATTACG

>AMDV3b_Contig17647

GTACCCGACATTGAACTCGCCATACATAACAACGATGTTCGGCACGAACGCCACCCATTCCTACCCGATCACGATCT
CAAGCCCTCCAAGCGGAACAGGAGATTATCAGCAACTCATAAACATCCCCAATCCATCGAGTTACGGCATAAATTCG
GCAGTCAGCAATTTCGAGATCGCATACACCAACGGAACACCGGTATATTCTTGGATTCAATCCTACAATTCCACTTC
GATTATGGTATGGAGCAAGATCGCCTACGGAACGTCTTCGCTTAGTTTGGAGGTGTTTCCGAGCTTTGAGAATTTAC
TATCTGCAACTGGATATATTGGAGAAGCACCACAGTTATCAAGCACTTATGGAGAATATAATAATTTAATGGATGTT
TTTAATAAATCACTTATATCTCAAAATTTAAGTTATACAGGGCAAGGCATGTATGGAAGTTCTAATTTAATAGGATA
TATAGCAATTATATCGTCAATAAATGAATATGGATATCTACAATTACAAAATAGTTCAAATAATTATAATCCACAAA
TATATGATTCTTCAAATTATACGTATTTTGGAATATGGGACGGTGGTGGATTTGCAGGTTCTACAAATTTATCTATT
TTAGTTTCACAATTAAACTTTGTTTCAATTGTTGGACTGAAAAATTCAAATGGATACAATTGGTATTATCAGGCTAA
TGATATAAGTGGAAATTATCAATATAATGTATTCAACAATGGAAATCATTGGGTCATAGATAGTGGAAATATAAATG
TCACTTTAAGATATTTAATAGGCATAGATGCACCAAATGGGGCAATGCCCACCTACACCATCGGTTCAGTAACCATC
GGTTCAGGTTCAGTATTTCAGGCGAATGCCACGACGACATCGACATTTTCGGGTGCGCCTACGGGAAACTACAATGC
CACCTATGTCTATGACACCTACAATATCCCCATAGCGTCCTCAACGAACTATGTTACGGTCATATTCAATTCAACAT
GGAGGCTGTCGAACATCTATCCGTCAACGTATCTGCCCCTGCGGAACAATTCAAACTTCGTCACCATTGAGGATGTC
ACAGGATTCACGTCCGTTCAGGCCACTTTGATCGAACCATCATCG

>AMDV3b_Contig17826
CCCCATATGCTGCCATAGCAATGTGGACAGGCATCTATATCCTGGTGATGATAACAGTGCTTGGCCTCGGCATATAC
ATGGAGGTATCAAATTGAATGCAAGGGAAGAACAGAGGAAGGTAGAAGAGGACAGGAGAGACAGGTATCTCTCGGCA
TTAGAAGGATTCGCACGTAGCATCACCAGCAAGCACATAATAAGAAAAGAAGGAAACAAAATTGAAAAGTGAAAACA
TAGAAAACATGGCAAAGATTAATAAGTTAGATACAAGTATCAGGAGCATGGAGGATGAGACCAAGATAAGGGTCAGT
TCGATCACCAGGAAAAGGCTCTCCAAGGCTAAGGCTCTGGGAGAGTTCAATTCATACGATGATCTGATAAACAACCT
GCTTGATCAGGTTCCGATAACTGCTCCGCCAGGAGAGGAGGTGAAGAAATGAATACGTTTAGTGATGAAGAAATGAA
CATGTTACTAAAAAATCCATTCGTTAGAACTGTAAACATTGATGAAAGAGTTGATAATTTCTGGGAAGTCAATAGAA
TGTGGGTTCCTTCTGGATATCAGAAAAAAGGACAGAATGAATATTGCCCAATAGAAATGACTACTGATAAGATTTTT
AAAAGATATACGTTGGGAGTAGTAGGTCAAGGGCTTGAAACAGAAGATTACTCAGTTATGGATACCGAGAATCCTTA
CTTTGATGTTCCTCTACTTCAATTTCTTGATATGTATGAAAACAGAATGATACCAATAGAAATGACAAAGAGACAAA
GAATATTGACAAGAATCAAAGACATAGGACTTTGCAGTTTTGCTTTTATTGGATTAAGTATCTTGTTCCTTATGATT
TTCTATTTAGTATTCCACCCTGTGAGTATGCCACATAGTTTACCTTATTTCTATGGAAAGTGATTATATGTTAAACT
ATTTCTTACCTTATGGGATTTACGAAAACTTGCCTTTTCATTATATTTTCTTTACACAACACAAAGGAGTATGGTTT
TAAATGAATAAAAAAATAATAGCACTCTTAGGAGTTGCTTTGATGCTTGGAACTGCTTTTGGAACTATGATATTGTA
TATGCCGAGTTTCTCAGATGGGAGTAATAATCCTATATCTTCACCTGTGACTGCATCTACCATTTCAAGCTCGTATT
CCAACACAGCTTACCTGAATACAACAGTGTCATATTGGCATGGTGTAGATCTCTCTAATACCGGTTCTTCGAGTTTT
AATATAGGAATAAGCCAAAATAGTTCTTATTCTGCATATAGCTTTGGATCATCTGGTTCGTGGGGTTTAGCTTCCAA
TGACCAGAGTGCATCTTGGACACAAACGTTTTCAGGTGCTCCTAATTATGTTTATCTTACCAATGCGAACTCTATCA
GTATTCCAGACATATTTTTTGCACTTGGTTCGTATGACAGTGCTGGGCATGTTTTTCAGGAGTTAGGCACAGTTTAC
GCCAATCTTACGGTTGATGGATATACTGAAACATGGTCACATACTTTTAATTTTAACGAACAGGGAGATTCTTATTA
CGCATTTATTGACCCACAATGGAATTTTTTTAGTGGTACACCACTTTTTACAATCTCTAATTTTGCTTCAGCCTCAC
TCTCTGTTGTCGCAACAGCTGATGGAACTGCTTATGTTGATGCACCATATGATTTTGCCACAACAACGGCTAATGCT
GGGGCAAACACAATCCTCACAGAAACAACAACCCCATCTTTTGCTGGTGTCAACATGTTTCAGGGATATCATTGGAA
TCAAGCTTCCATAATATATTCTCCGGCTTCAGACTTAACTTCATATTACATACATTTAAGTTTACCTTATTCCACGA
CACAAGTAACATATAATGGTAATACATATACAGGTACAACAGATGTTATACCAGGATCACTTAGCAGTTTTACATTT
GATGTTAATAATCCAATCAGCCAACCAATAGGAGACACAGCAGGATATAGTGTCTCAGCTTCATATTATCTTGAGGA
TCAGGTTCAAGCCCAGCAGGTCACAGGCTACGCACAATCATACTCAGGTTCCCAGAATCCAACACAGAACCAGAACT
GGTGGAATTCGACTTTAAGTTTTACCCTAACCGATCCCAGCGGAGCTTTAAACAATCCATCCTTAGCCACAACCTCT
GGCACAATTAACGCTGAAACAACAATGACATGGAGCATAAACCATATCCTTTCTGTAGGCTACGGACAGTCTCCGGC
CTTTGACAGGATTTTCTATTCTGGTTCGGATATTTCATCATCCTCTTTTGCGCCTTCGGCTGAGTACAATTCGGTTA
GCCAGGGAACAGCCGTAGTTGTGGACAGCTCAACCGCAGTTTCATACAATTTCGATGAGTTCATCAACTACCATCCT

>AMDV3b_Contig17669
TCATTTAACGTTGTAACTGAGAATGAGCAATACGCTACGAATGCGGTTTACAACGGGGTAACGGAAGCTGTGCCATA
TACATTTACAGGTGCGGTCTCTGTTGAATCTATCACATTTATTCCGGAAGCAACGCCAAATTTAGCATTAAGTTATG
GTGGTGGATACACAGTTGAATATTATATTTCCAATATTGAACAGGCACAATCAACTTCGACGACTCAAACATTCACT
GGAGCCCAGAACCCGACACAACAACCAGAACTGGTGGAATTCGACTTTAAGTTTTACCCTAACCGATCCCAGCGGAG
CTTTAAACAATCCATCCTTAGCCACAACCTCTGGCACAATTAACGCTGAAACAACGATGACATGGAGCATAAACCAT
ATCCTTTCTGTAGGCTACGGACAGTCTCCGGCCTTTGACAGGATTTTCTATTCTGGTTCGGATATTTCATCATCCTC

TTTTGCGCCTTCGGCTAAGTACAATTCGGTTAGCCAGGGAACAGCCGTAGTTGTGGACAGCTCAACCGCAGTTTCAT
ACAATTTCGATGAGTTCATCAACTACCATCCTTCCATTGTGAACAGCAACGTAGCAACGCCTGTAAACCCCCTCACG
CCTCTTGACCTGTATGCGAATTCCACAGAGGTCATATTAGGGGAGCAACAGAACCTTGCCGTAAGCGCCAATGGCAA
GCCTTACATCCATTCATCCGCCACTTCGTCTGAAT

>AMDV3b_Contig17833
CATTCGTATCATGTGAGGAGAGAGTTAAGGAAAGCAATGGAAGTGATGGACACCATCGAATTGGAAAAGAGTCTGGA
GGCCCTGAAATGAACAGAAAAATTCTCACCAATTGGCTTCGTGCTGGCTTAGCATTGCAGCTGATATGCGAGTTTGA
CGGGTGCACAAGGAAAGCGGAATACAGCATCACGAGGTGGCATGGCCTCCATATAGTGGTATGCAGGGAGCATTTCC
CTCGGATCGGCGAACTGGATGCAATGGATGAGGAGGCCCTGAAATGATAATAACAAAGAAACAGGCACAGGCGTGGA
ATATATTTCGGAGTCAGCTCTGCGATTTTCATGTGCGATTCCCTGATCAGACGGCATCCATGTACTCATGTAGGGAC
TGCATGATAATCCAGGACAAATGCGGTTGCCGTGAAGATGAAGGATGCGAATATTGTGAGCCTGATATGGCAAGATT
GGCAAGGTCAATCATAGGGAGGAGATGAATTGTGAGAGGGAAAAATATTATTGAGCAATAAGAAATAATATAAGCAT
GAAATGATTGCGTATTCGACAATAATGAAATCAGTATTCAGGTACTATGGAGGGAAATTCCATCAATTGAAAGACAT
AATTTCGATATTACAGGATCACATGAATTCATTTGATGCAGTTGTCGACGTGTTTGGTGGTTCTGGTAAGGTTCTCC
TCAATTTACCTGATGAATGGAAGAAGATGAAGATCTATAACGATATTGATTCTGATCTGTATGTAACGTTCAAGGTT
CTTCAGGATCCTTTGAAGAGAAACATGCTTATCAAGAAGCTCAGGGTTGCATTTGCCCATGAGAAGATCTTCATGGA
GATGAGGGATAAAAGATATCCTGGAGATGTGGAAACGGCCTTCAGGCTGATATACTTGCAGACGTATTCATTCATAG
GTGATGGCAGATCATTTGGAAGGAGGTTCAAGGGCAACAAGCAGTCAAGGTTCACCATAGAGAACTTTGTGTATGTG
AGGGACTGGACCATAGAGCATATGGATTTCCGGGATCTGATGAAGAGGTACTCGAAGCCCAGGGTGTTGTTTTACCT
GGACCCCCCATACCTGTCGTCAGGAAAGAGGTACAGGCGCAGTTTCAGCATTGATGACCTCAGGGATTTGAAGCAGA
AAATGGATGGACATCCAGGTTCATATCTCCTGAATCTGTCATCATTCGACGAAGGTATGGAGGAAATATTCGGAAAA
CCACAGAAAGTGATCGACTATGCCAATCCAACGGACCACAACGGAAGGAATAGATGGGGATGCGGGTATTGGTGGAG
GTTTATTTGAATTTTCATCTTCGCTATTATAGGATATGCGAATACTTATATACTTCTTCGCTGTTCCTATTTTGATG
AATATGTCTGAAGTTTCCAGGGAGGCTCACCGCTTCTGGTCTGCTAATGGGAACGTCAAACGAGTGTCGGAAATATT
CGGTTACGTCCGACGCTATCTCCAACGGCTGCTGGCATGATACAGTTACAGGAAGCGATTTTGATCTCATTGCACAC
GGACAGGTTCGTATCATGCCACAGGGCTGGGAGAGATTTTTGAGTTTTACTAAGGAGGTTCTTTGAAATGGAACATA
AAAAAATAAGTGAGTTGTCAGGACTCACAAGGATATTTCCCAGAAGGGCATTAAATTTTTGGAATGCCAGAAGGGAA
AAGAAGAGAATGGAATTGTACAATAGACTCGTAGATGCAGATCTTTCCCTGAAGGGTCTACTGCAGAGGGCAGCTTA
TGAAGGCCTCATTGAATCAAGAACAGTCTCGATACACGAACCAAGGCATGTTCCTAATTCCCTGGCATCGATACAAA
AGCACAAATCCGATATTTTACGAACCTCCGAGCAAGAAAGCTCCGACCTTCAGGGAGGAGATAGAATAATGACAACA
TATGAAACATGGGAAAAAACCATGATGGAAAGAATAGCCAGCATCAACAATGCCGTTAAAGGCAGGCGAAGCCAATG
AGTGAACTGAGGATAAGGGAGTTTCCAGTAGAGGTTTTGGAATGATCCACGAATTTTACTACGGAACTTCGTGGCAC
TGGCCTCTGCCCTATCTTCCACGAGTGATGCTTTCATACAACATGCTGAGGAAGATAAAGACTCCATGGAAGGTCGA
CATTCCATTCATGATGGATTCCGGTGCATTTGCCGTAATTCAGAAGTATGGAAAATATCCTTTCACTCCAGAGGAAT
ACGCATCTGGAATACAGAAATGGCATCCAGACATAGCATGGACAATGGACTATCCCTGTGAGCCTTTCGTGCAAAAG
AGGGGAGACTACAATCCAAGGAAATCTCAGGAAATGACAATAGATAATCAGATCAGACTACTGGACTTGAATACCAA
TACTCAGATGGTCGTTCAGGGATGGACAGTTCAGGATTATCTTGAAAATCTTGATAGAATCAAGGAGCAAGGGCTGC
TCACTGAGCATCTTGGCATCGGTTCTGTATGCAGGAGAGGGCAGAACCGGGAAATAGCCCGCATAGTAAGGGCAATC
CATAACAATGTGCCATCATGGGTTAAGCTTCATGGTTTCGGCATAAAGGTTTCTGTCCTGAAAGATACGGATGCCAG
ATTCTATCTACACAGCGTGGACAGCCAGTCATGGGGCTATGAAATGAGATATGGGGATTGGCTGAGAGGTCAGTACA
ACGGCAAGACATGGAAAGACAAAGTCCCCAATCTAAAATCCTACGTGTCAAGGATAGAATCGATGCTGAATCCTCTT
GAACCATTGTATGCGGAGGCCCTGAAATGAGCCTCCATAATGGAACTGCCGCAGAAACGGAACGAAAGAAGATGCTC
GAAGCCGAAGGGCACGTAATCCTGAATCTTGGACAGAATCAGTACGGAGACCTGGAGGATGTCTGCCATAATGTCAT
CATCGAGGTAAAGACATCAAAGCATTCAAGGCGTGATCTTTCCAGAAAGGAGAAGGAACAGATGACAAACCTCCTCA
TGCTGAGACCCTACCGAACGATCAGGTATGACATCCGGTTTTCAGGCAATGGCCATCACGGCCCAGTATGGCAGTGT
GAACATCCGGACAGGGTGTTAAGTTCATTCAAGCTTAACACAAATATCAACACCCATCCGTTTAATGCTCCAGCGAT
TTCTAAATCGGTCTTGGATAAAAATAGGAAGGCGAAAAATGGAATACATACTGAAAAAAAGAAGAATCCGGAAAAAA
ACAAAGCCATGGAGGCCCAAAAGTGGTAGTATATAAATTATGGTGGAACTTTCTTCCACATGTATGGAGGAAATGTC
TATTAAAGAAGGTGAGGCATTGATCTGGATATGGATTAAATGCAAATGTGGACACAGGGCGTGGAGTCATCCTGTCC
GCACCAAGTTTAACGAGAAGCCATGCCTCCACCACACAAAGAAGCAAGACGTTGCTGGTTACATATATCACAGCCGT
CCATGGCCTGATATAAGAGGATATTGCCCCTGCGATGATTTTGAGCCTGCCGGTAGATACCAGAGGGGAATATAGAC
CGGGAATGATGCCCGACAACATAGAGTTCAGGAACTGGACGGATGAGTGAAAAAATGACGGAAGAGTTTAAAACATT
TGACTGGGGAAGTAAAATGATAATGGATGAACCCAAAATACTGGTATGGAGATGCCTTAACTGTAATTATGTCGGTT
CTTCAGAAGATGGCACAAGAAAACCAACAAAAAAGAGTTACGCAATGATTGGGCACTTCGAGAATTTGGAGGACTAT
TATGGAAAGGCATGGATCACTGGTACGATTATCCTACTCATTCTTTTATGGGTGAATATGTTAAAAAATTTCAATGG

ATAAACGATATGAAACAACACACAAAAGAATAACAAAGGTGAAGAAATGAATACGTTTAGTGATGAAGTGTTTAGAA
TCACGCAACAAAATGCGCCACAGTATATTAGATATCTGTATAATCTTCAATGGAAAGAAAAATCTTTTAAACCAAAT
TTCAAAAAACACACACGGCAAAGAGCAAAGATAATAGGTAATACTATTTATTTAGATTCTAATTTTGAAACCATCGA
AAAGCGCAACGAAGTAGCTCTTCACGAACTTGCACATTATATATCCAATGATAAATATGGATTAAGAAACAATCCTC
ACGATGATTATTTTTTTCTGGTACTTAATGAGTTGAATGGAAAAGATATAAGGTATAAGTATGATAAAATGGGGGTT
ATGAAACTTGAAAAATAAAAGAGTGTTAAGATCGCTCCGAAATAAAAAATTGATTGCTATTATTGGTTGTTTGTTAA
TGTTTTCTGGTGTTTTTGGATTGATCCAATACATCCCTGTTTATTCAGGGAACCCTTTGCAAACACAATCCACCAAT
CCCGGACTTTCCATCTCTAATAATACGGCCTATGTGAATACAACAGTGTCTTATTATCATAATGTTTTGCAGAGCAA
TTCCGCTGTGAGTACGGTTAATATGCCAGTCGCTGAAGATAGTAGTTATTCTACATTCGGTGTCGATAACCCAACTT
ATGTTTTTACCACTTCGAGTAATGGTTTATGGAAAACTGCAAGTTTCACAACGGGGATCTCTCAATTTTGGGGAGGC
TCAAACTCCAACACGGTTCAACTTAAGGATCTGGCATTTATGGGGCAATATTCTGACCAAATAGAGACATCGACAGG
TACAAATTATTATGTATATACTCCAGGGAACATATATGTAAATCTCACAGTTCCTACAACAGGGGGATTTACAGGCG
GGACATTGTCATGGTCAACTACTTCGTTTAGTCAAGACTACAGTTCTTCTCCACAGGTGATGTATGTTAATCCAGTA
TGGAATATATATAGTGTAACGAGTGGTGACACGGCATACGGAAGCCTCACGGTTACCATTGCACTTCAAAGTGTTGA
CATCTCAGGATTAACAGGTGCTGCCAATTCGTATTCTTGGGCTACGGGAGGAAGTTCATCACAAAATGGAATAGTTA
CATCCACTGACCAAACAGGAAATGCCTATACATCTCAAACAGCATTGCCATCCGAGGATAACGTGCCTTATGAAATT
GGGTATCATTTCAACTCTGCCTCAGCTTCTTGGTCCTATTCAGATGTAACTTCATTTAACGTTGTAACTGAGAATGA
GCAATATGCAACAAACGCTGTTTACAATGGAGTTACAGAATCCGCACCATACACATTTACAGGTGTAGTTTCTGTTG
AATCAATTACATTTGATCCAGAAGCCACACCGAACGTGGCGGACGTTGCCGGTTATACAGTTCAATACTACATATCA
AATATAGAGCAGGAACAATTCACTACAACAACACAATCTTACTCAACATCACAGACCCCAACGCAGAATCAAAATTG
GTTTAATTCATCAATGTCTCTATCTTTAAGTCCACCACCAGGTGCGATAAATAATCCATCTTTAGCAACCACATCGG
GAACAGTTAATGCCGAAACAACTATGACATGGTCAATAAACCACATTTTAAGCGCTGGGTACGGGCAGGGCGCTATG
TTCGATACCGTATATTATCATGGCACAGACATAACATCTGGAAATTATGATCCGTCAAGCCAGTATAATTCATTATC
GCAATCCACGCCCGAAGTAATCGACAGCAACACATCAGTATCGTACACATTCGACGAGGCCATAAACTTCGCCCCGA
CCCTCTTATCCCACAACGTGGTCTGGACCGGCACATCAGATACCGCA

>AMDV3b_Contig17844
TGTTTCAATTGTTGGACTGAAAATTCAAATGGATACAATTGGTATTATCAGGCTAATGATATAAGTGGAAATTATCA
ATATAATGTATTCAACAATGGAAATCGTTGGGTCATAGATAGTGGAAGTATAAGTGTCACTTTAAGATATTTAATAG
GCATAGATGCACCAAATGGGACAATGCCCACCTACACCATCGGCACGGGCAGTGTGTTTCAGGCGAATGCCACGACG
TTCACGAAAGCGCCCTATTCAACGGTCGCCAACTACCAACCGAACCCGCTTGAGCAGATTTACACATATAATATATA
CGATTCGTTCAGCGACAATTACATAACGCTCTTGAGTAACTCCTCATGGACATTCCTGCAGGATTCGGGATATGCCA
GTTATGATCCAACGACGCACGCCCTTACCTTCAGTGGAGTATCCGGCATAGGCGAGGTTTCAGCGTCCTATCTGGAG
CCGTCACAACAGCTTGGGCAACCGTCATTTGTCACCCTGTCCCTGTCCGCATACGGATCATCACTGTTCAGTGGATT
CAGCTATTCCGTTGCCTACGTTCCGTATTCGTCAAATACAACCTATTACGATAATGGGAC

>AMDV3b_Contig17193
TGAATTCAACTGCTAATAAGGTGGACATAAACCAGCAGGGAAAAGCAGTAAGCGAATGGGTTAATGCAACTTTAACA
TATAATGCTACCAAATATTTTTCATCTACTTTTCAAGCTGGTTCAACTTCCTACACAGGCAGTGCGGATGCTAACAC
TATACCTATAGATACATCTAACATGCAAATAGGTGTTGGTCTTGACGGGACGAGTGGGGATTATGAATTCCAATCAA
TTTACCAGTATGAGTTTGCCATTTCCTATGTAGCTTCAATGCCCACCATCTCAATAGGTTCTGCAAGTACATCTGGC
TGGTCGGTGAATGGTATATCTGGAATTACCAGTTTTTCAGCTCCTAATTTATCTTCAAACAAATTGCCATTTTTTTA
TGCTTATTCATCAAATGGGGATTATATAGATAGAAATTTAAGTGTATATAACATTAAATCTATTACATTTTTTGCCT
ATACGACCGGTCTTGGTGATGTATTTTTGTTCGCAAATTCAACAGGGGCAGGTCAAATGATGAGATTAGATGGAAGG
GCGGGATATCCGGATACAGCAATAGCCTCAACTACTAATTGGACAACATGGGAATCCCCTCCAAACGATTATCAGGC
AAGTACGCATAATTGGTTTGAAATACATTTACAGAATAACTCTAATATTTGGTCAGTATATGTAAATAATCTTTCAA
ATCCATCTGGTACATATGGCAAATTTGTGGAAAATTACACAATGAGTTATAATGGTACATACTTCGGATATGGTGGT
GATGGTGATGGTTCATCACATATTACCTATTTCTCTCCCTTAATGTTTAATGCTCCTGAATATGTACCTGTTGCGAT
GCCCACCTACACCATCGGTTCAGTAACCATCGGTTCAGGTTCAGTATTTCAGGCGAATGCCACGACGTTCACGAAAG
CGCCCTATTCAACGGTCGCCAATTATCAACCGAACCCGCTTGAGCAGATTTACACATATAATATATACGATTCGTTC
AGCGACAATTACATAACGCTCTTGAGTAACTCCTCATGGACATTCCTGCAGGATTCGGGATATGCCAGTTATGATCC
AACGATGCACGCCCTTACCTTCAGTGGAGTATCCGGCATAGGCGAAGTATCGGCGTCATATTTGGAGCCGTCACAAC
AGCTTGGGCAACCGTCATTTGTCACCCTGTCCCTCTCCGCATATGGATC

>AMDV3b_Contig17714

CATTCGTATCATGTGAGGAGAGAGTTAAGGAAAGCAATGGAAGTGATGGACACCATCGAATTGGAAAAGAGTCTGGA
GGCCCTGAAATGAACAGAAAAATTCTCACCAATTGGCTTCGTGCTGGCTTAGCATTGCAGCTGATATGCGAGTTTGA
CGGGTGCACAAGGAAAGCGGAATACAGCATCACGAGGTGGCATGGCCTCCATATAGTGGTATGCAGGGAGCATTTCC
CTCGGATCGGCGAACTGGATGCAATGGATGAGGAGGCCCTGAAATGATAATAACAAAGAAACAGGCACAGGCGTGGA
ATATATTTCGGAGTCAGCTCTGCGATTTTCATGTGCGATTCCCTGATCAGACGGCATCCATGTACTCATGTAGGGAC
TGCATGATAATCCAGGACAAATGCGGTTGCCGTGAAGATGAAGGATGCGAATATTGTGAGCCTGATATGGCAAGATT
GGCAAGGTCAATCATAGGGAGGAGATGAATTGTGAGAGGGAAAAATATTATTGAGCAATAAGAAATAATATAAGCAT
GAAATGATTGCGTATTCGACAATAATGAAATCAGTATTCAGGTACTATGGAGGGAAATTCCATCAATTGAAAGACAT
AATTTCGATATTACAGGATCACATGAATTCATTTGATGCAGTTGTCGACGTGTTTGGTGGTTCTGGTAAGGTTCTCC
TCAATTTACCTGATGAATGGAAGAAGATGAAGATCTATAACGATATTGATTCTGATCTGTATGTAACGTTCAAGGTT
CTTCAGGATCCTTTGAAGAGAAACATGCTTATCAAGAAGCTCAGGGTTGCATTTGCCCATGAGAAGATCTTCATGGA
GATGAGGGATAAAAGATATCCTGGAGATGTGGAAACGGCCTTCAGGCTGATATACTTGCAGACGTATTCATTCATAG
GTGATGGCAGATCATTTGGAAGGAGGTTCAAGGGCAACAAGCAGTCAAGGTTCACCATAGAGAACTTTGTGTATGTG
AGGGACTGGACCATAGAGCATATGGATTTCCGGGATCTGATGAAGAGGTACTCGAAGCCCAGGGTGTTGTTTTACCT
GGACCCCCCATACCTGTCGTCAGGAAAGAGGTACAGGCGCAGTTTCAGCATTGATGACCTCAGGGATTTGAAGCAGA
AAATGGATGGACATCCAGGTTCATATCTCCTGAATCTGTCATCATTCGACGAAGGTATGGAGGAAATATTCGGAAAA
CCACAGAAAGTGATCGACTATGCCAATCCAACGGACCACAACGGAAGGAATAGATGGGGATGCGGGTATTGGTGGAG
GTTTATTTGAATTTTCATCTTCGCTATTATAGGATATGCGAATACTTATATACTTCTTCGCTGTTCCTATTTTGATG
AATATGTCTGAAGTTTCCAGGGAGGCTCACCGCTTCTGGTCTGCTAATGGGAACGTCAAACGAGTGTCGGAAATATT
CGGTTACGTCCGACGCTATCTCCAACGGCTGCTGGCATGATACAGTTACAGGAAGCGATTTTGATCTCATTGCACAC
GGACAGGTTCGTATCATGCCACAGGGCTGGGAGAGATTTTTGAGTTTTACTAAGGAGGTTCTTTGAAATGGAACATA
AAAAAATAAGTGAGTTGTCAGGACTCACAAGGATATTTCCCAGAAGGGCATTAAATTTTTGGAATGCCAGAAGGGAA
AAGAAGAGAATGGAATTGTACAATAGACTCGTAGATGCAGATCTTTCCCTGAAGGGTCTACTGCAGAGGGCAGCTTA
TGAAGGCCTCATTGAATCAAGAACAGTCTCGATACACGAACCAAGGCATGTTCCTAATTCCCTGGCATCGATACAAA
AGCACAAATCCGATATTTTACGAACCTCCGAGCAAGAAAGCTCCGACCTTCAGGGAGGAGATAGAATAATGACAACA
TATGAAACATGGGAAAAAACCATGATGGAAAGAATAGCCAGCATCAACAATGCCGTTAAAGGCAGGCGAAGCCAATG
AGTGAACTGAGGATAAGGGAGTTTCCAGTAGAGGTTTTGGAATGATCCACGAATTTTACTACGGAACTTCGTGGCAC
TGGCCTCTGCCCTATCTTCCACGAGTGATGCTTTCATACAACATGCTGAGGAAGATAAAGACTCCATGGAAGGTCGA
CATTCCATTCATGATGGATTCCGGTGCATTTGCCGTAATTCAGAAGTATGGAAAATATCCTTTCACTCCAGAGGAAT
ACGCATCTGGAATACAGAAATGGCATCCAGACATAGCATGGACAATGGACTATCCCTGTGAGCCTTTCGTGCAAAAG
AGGGGAGACTACAATCCAAGGAAATCTCAGGAAATGACAATAGATAATCAGATCAGACTACTGGACTTGAATACCAA
TACTCAGATGGTCGTTCAGGGATGGACAGTTCAGGATTATCTTGAAAATCTTGATAGAATCAAGGAGCAAGGGCTGC
TCACTGAGCATCTTGGCATCGGTTCTGTATGCAGGAGAGGGCAGAACCGGGAAATAGCCCGCATAGTAAGGGCAATC
CATAACAATGTGCCATCATGGGTTAAGCTTCATGGTTTCGGCATAAAGGTTTCTGTCCTGAAAGATACGGATGCCAG
ATTCTATCTACACAGCGTGGACAGCCAGTCATGGGGCTATGAAATGAGATATGGGGATTGGCTGAGAGGTCAGTACA
ACGGCAAGACATGGAAAGACAAAGTCCCCAATCTAAAATCCTACGTGTCAAGGATAGAATCGATGCTGAATCCTCTT
GAACCATTGTATGCGGAGGCCCTGAAATGAGCCTCCATAATGGAACTGCCGCAGAAACGGAACGAAAGAAGATGCTC
GAAGCCGAAGGGCACGTAATCCTGAATCTTGGACAGAATCAGTACGGAGACCTGGAGGATGTCTGCCATAATGTCAT
CATCGAGGTAAAGACATCAAAGCATTCAAGGCGTGATCTTTCCAGAAAGGAGAAGGAACAGATGACAAACCTCCTCA
TGCTGAGACCCTACCGAACGATCAGGTATGACATCCGGTTTTCAGGCAATGGCCATCACGGCCCAGTATGGCAGTGT
GAACATCCGGACAGGGTGTTAAGTTCATTCAAGCTTAACACAAATATCAACACCCATCCGTTTAAATGCTCCAGCGA
TTTCTAAATCGGTCTTGGATAAAAATAGGAAGGCGAAAAATGGAATACATACTGAAAAAAAGAAGAATCCGGAAAAA
AACAAAGCCATGGAGGCCCAAAAGTGGTAGTATATAAATTATGGTGGAACTTTCTTCCACATGGTATGGAGGAAATG
TCTATTAAAGAAGGTGAGGCATTGATCTGGATATGGATTAAATGCAAATGTGGACACAGGGCGTGGAGTCATCCTGT
TCCGCACCAAGTTTAACGAGAAGCCATGCCTCCACCACACAAAAGAAGCAAGACGTTGCTGGTTACATATATCACAG
CCGTCCATGGCCTGATATAAGAGGATATTGCCCCTGCGATGATTTTGAGCCTGCCGGTAGATACCAGAGGGGAATAT
AGACCGGGAATGATGCCCGACAACATAGAGTTCAGGAACTGGACGGATGAGTGAAAAAATGACGGAAGAGTTTAAAA
CATTTGACTGGGGAAGTAAAATGATAATGGATGAACCCAAAATACTGGTATGGAGATGCCTTAACTGTAATTATGTC
GGTTCTTCAGAAGATGGCACAAGAAAACCAACAAAAAAGAGTTACGCAATGATTGGGCACTTCGAGAATTTGGAGGA
CTATTATGGAAAGGCATGGATCACTGGTACGATTATCCTACTCATTCTTTTATGGGTGAATATGTTAAAAAATTTCA
ATGGATAAACGATATGAAACAACACACAAAAGAATAACAAAGGTGAAGAAATGAATACGTTTAGTGATGAAGTGTTT
AGAATCACGCAACAAAATGCGCCACAGTATATTAGATATCTGTATAATCTTCAATGGAAAGAAAAATCTTTTAAACC
AAATTTCAAAAAACACACACGGCAAAGAGCAAAGATAATAGGTAATACTATTTATTTAGATTCTAATTTTGAAACCA
TCGAAAAGCGCAACGAAGTAGCTCTTCACGAACTTGCACATTATATATCCAATGATAAATATGGATTAAGAAACAAT
CCTCACGATGATTTATTTTTTTCTGGTACTTAATGAGTTGAATGGAAAAGATATAAGGTATAAGTATGATAAAATGG

GGGTTATGAAACTTGAAAAATAAAAGAGTGTTAAGATCGCTCCGAAATAAAAAATTGATTGCTATTATTGGTTGTTT
GTTAATGTTTTCTGGTGTTTTTGGATTGATCCAATACATCCCTGTTTATT

>AMDV3b_Contig17705
TACTTAATGGAAGGCAATTTCATCGACCCTCATGTAGGTGCGGCAATCGTTGCATCTCCGCCTTCCTCCTGCTGTCG
TCTTTCTCTCCCACGTCCTTCCGCACTTTGGGCATGTGACATATACACCCGTATGATAGCGGGCATATTTCTCGTTC
CTTACGATACGGAACTCTTCCAGAGCTCCGGGCGGAAGAACCAGATTTCCTGACTGATCGAACATCACTGGATCATC
TCCTTGTACTTCTTCCATCCTATTTGCACAAGAAGGAGGAAATTGTCCCATAGATCGGAAATGAGGGGGATCACGGC
ACGTCACCCGATATTTCCGCATCAGGGAGCTCTTCATACGAAGTACCGCCGTAACCTGTGTAGCTGCGGAGTGCCCT
TCCTAATGCCCTTGTCTGTGCCATCTCTCTCAAAAATTTATGCATGTTCCGGTTCTGGACGTTTTCACTGTTTGCTG
TGCCCACTCCATTGGTCGGTTTTGACATATCTTCATAGGCCATCTTCTGGACTTCAGGAGACAACTTGGATATGGTT
GCGAGCTTGTCAGCTGATATCTTCGGGAATACCTTCGCCTCCGCAACCCATGTATTTGTGCTCTCATCATAGCGATC
CGTGATCTCTATGCGCTGATACCCGATCTTCCCTGCCATGAACTGGAGCCCGGCAACCTTGATGTACGCATTGTCAT
TGAGCTTCATGAAGAACAGGTTTGCCATCTCCGGAGGGAGCCCTGCAATCTTTGCGATGGCAAGTTCGCTTCCCGGT
TCCACTGTGAATTTTGATTTCTCGCTGTTCTGCACTGCTCCATGTGGTTCCTGTTTAGCGGGTATCCCCTGCCTATC
CTGAACTATCCTTCCAACCACGAAGTCAAGATCGCCATCATGGACAAGGCAGTCAATAAGCGATGAGGCTTCCGATT
TTGTGAGCTCACCGATGGGAAAGCCAAGGACTTTCTCGACGCCTGCGATGTCCTCATTTGCCTCAGCCATCCTGTTA
AGAATCAGGCTTTTCTGTGCATTCGTCATTGCTGTGGAGGGAATCATTTCACATCCCTTTCCGATTTGAATTTCTCA
GTCAGGGTTTCATCAATCTGCAGAACCTGCTTCAGAAGATCAGCATCCATTGACTCTCCGTATAATTTGATAGTCCA
CTGGTGCTGGCCTTTGGCATTCGTCCCAAGTTCAATTGACGCTGGCTGGGAAACAACATTAATTTCCCTTACTTCTG
TGGCTGGCGATTCTACGGTGCTATTCACGATAGCACCTCCTTATAGACTTGTGAGGTCATAAGATGATAACCGATAG
TAGTATAAAAACATTACTTCAATGATTACTTTGGTGTCATAGTTAAAAATGATTATATACTTCTTAAGGTTATATCA
CAATATGTCTGAAGACGGAAAAAAGAAAACTGTAATCCGTAGCTTCAGGCTGGATGAAGACTTGAGCAATTGGATGG
AAGAAAAGCATATATCCATAGGCAAGTTGATCAACAGCCTGGTTCGGGAATACATGGAAAAAGAACTGGGAAAAGAA
GACCCTCACGCAAGCAAAAAAGGCGACAAAAAAGAACGATTTTAATTCTCGGGCTGTGAACATGGATCATGAAAACC
TGCATGCATTTTCATGTGTGTGTGCGTCTTTTTTTGTGTACCCCCCTTGGAAATGCCAAAAAAATGAAAAATGGGAA
AACACGGAAAAATTAATTTTTGGTTATTTAATAAATATAGATAGATAGATGTAAGGACACATGGCATTACATGTTAT
AGAGAAGTTGTATTTTCCTTATTTTTTAGAATGCACGTGTCATGAAAAAAAGCATTCTCTATAAGGGGTAAAATAAT
AAAAAATAACCAAATAACCAAATTACCGCGCGTTACAATATATTTGGTTATTCTTATTTTCCCTTTATAATACTGCT
TCGTTTGAAGATCAATAACCAAATAACCAAATAACCATAGTTTTCACTATACTTGTTGTTCTCGTAAAATTGGTTAT
TTGGTTATTCGTCTCTTAACCACTCAGAAAAACACGGCTTCCGCATACATAACCAAATTTCTGAAAACTCGGTTTTT
GACTATATTTGTGGTTATTCACTTTCCAGAAAACATCTAAAATACGGGTATTCTGGAGAT

>AMDV3b_Contig17679
TGGAATGGAGGAAATAAGAGAATACATAAATTTATGGAAAAGGTACGATAGTGCATACTTTTGGACGCCGTACGATT
TGGCGTTTCAAAGGAGAAAAGCGGAAGAGGAGAACTCATATCACATCGAATTTTCATTCAATGGTCATAACTACGTT
ATGACCATAGTCATGAAACAATCTACAAGGAATGTTTACGTGACAAGGACATTGATGAAAGACAGCAGGCGCGTGAC
GATAAGAGAATTGACAAAGATTATCAATAGGTGAGAAGAAAATGGAAAAGAAAATATACAGAGTTCCAGTCTACCTT
CCATACTAACTAGTAATTTTGAAACAGGGCATACATGAAGTCCTTTGACCAGGAGAGGAAGGATGGAGAACCCGTCC
TTGGGAAGGAAAACCTTGACCGTGTTATGAACAATGCCAGATGGCTGATCAACCAGATAACTAATAAGATTGGAGAG
GCGCAATAAAGGAGGTAATATTTCAGAAGAGAAGAGGGAACCTGTCTGAATATTTTGTAAAAACTCCTGAATAAACT
ATTTTTAGGTTATCTATGGTTAAACTAATTCTAGATAACTTTAAATAGTTGTATAGAATATACAACTATGGAAAAAA
GAGTATTCTGGGACATGAAAGGTCCTGAAGGGTTTTTTTATAAGGGCTCAAAGTACTAATGAATATGAGGACAAATC
CCATGATTATAATGGAGATAAAATAGATGGAGCGGTTTGTGCAATTGAAATAAATAATCCAGATATCACACTGGAAG
AGTGCAAGCAGGCCTTTAAGGAAGATGGCTTGGAATGTTATATCGAGGAGAATGAGTTCATGGTCTTATTTGAAGGA
TACAAATATGAAGAAATAGAGGACGGTGTCATAGCAAAAATGGATGCTATAAGGAAGATCTGGAATTTGAAAACTGG
GAATGTTATATTTGATATGGAAATTGAAGGGTGTTATTATTGAAATATTTGGGAAAGGATGTAGTAGGTAGAAGGAA
GGATAGATCATTTAGGAGAAAT

>AMDV3b_Contig17162
TACGAACCGTATGTACTCTGAGAGGGTGAGGCCTCTTGCTTCTGCCCTCTCCCTGAGCTTCTCAATGTCTGCATCCG
TTAACCGGGTGCACACTCTTGTTTTCTGCATGTTCTCTCACTGGTCAACGGGATCAGGGAGATCAAAAAATCCCTGA
ACCTCTTTTTCGGTTAAAAGTCCTCCAGTAGGACAAATATATCATTCGTCTCTTTTCCACAATAAATGCACAGTTCC
GGTGCATCTGAAAGAGGACTTCTTTCCCAAGCTTCAACTACTTCCACTACTCTCTCGTTTTTCAGGTTCTTTTCGTG
TTCTTTGCACACATATCCTTCAAACATTTTATCACCTTAATGTAGTATATCATAGTTGTATATAAACTTGTCGTCTA
ATAAAGCAACAATATAACCTCATATAAATAAAAATTATTCCTCTTTCTATCTTTAAATATGTATTGTACAACCTATC

CATGAGATCTCCAATCGATTCGTTATCAGGATATATAAATATCCAGTTCACGACATGACAAAATGTGTATTTTGTCA
TGCAATTGATTTACAGATTTGAGAGAACCTGGTGACATCCTCCCACCCCTGAAGAGGTGGGTTTTCACGCTCGATGT
GATAAAATCAGTAATCCATTATTATCTGGCGTGGCTCCGGTTCGTTACATCCACAGAGCATGCATTTCCTGTTTGAT
GGCGAACCACCAATCCAGCCTTGAGAATCCAGGTTAAAGTGATACGTTTGCCCATCCATGAATCGCAGGAGATGCCC
GCAGTTCGAACAGAATGCGCCTCTCATCTTCGGTCAGACCTCCATTCCAAAACTATCATTCAACCACTCCTTGTTAA
AAAGAGAAGAAGCTTTACATTCTCTGACTTGTTTATTATAATCGTTTAACTTTTTAATACAAAACTTTTCACTACGG
CCTTTCTTAAGCGCATCTCCTATATATTCTTCAATAAAAATCATCATTCTAAAATCGTTCATTTCACTCAATCCTCC
TCATAGTTACTTGGGGCATTGGATAAGTCTTCTTTTTTCTTTAGCAATTTACTTTTTATCAACGCTATAATTGCCTC
AATAGCTTCGTTATCATCTGCTCCTACCTTTATGTCAATAGTTTCCCATTTTAGTAAAAATGATTTTGATGGTTCTA
ATTTCTCTATTTTCGAGGTATTTTCTACATATTGAGCGGCTTCATTCGGATTCATTTCACTCAATCCTCCTCATAGT
CATTCAAATCGATGAATCTCCTGCCGTCTATCTCTGTGTATTGGGGAAAGATACGAATCTCTCCTGTGCAGCTCTGC
ATCTTTCGTAGGAGAGCTTTCGTAAATTTCTCTTCATCTCTCCATGTGATTTTGGATGGCATCAGAAAACCTCCACA
AATTCTATGTCCATTATATCAAAGTCATCAGGGAACTTTTGAGCCTTGACATACCTCTCGGCATTATCCTTGCTTGC
AAATGCCTTCACCACTTTCCAGCAAGCATTGCAATTCTTTCCGTTAAATTTATCCAATTCTATTTTAGACACTACAT
AGATTATCATCAGAACCCCTCCAAGAACGCCTTTACCTTCTTGTAATCGATGAGAAGG

>AMDV3b_Contig17175
ACAACTTCACTCATGAAAGAAAGTATGATCCAGATGCTGATGGTGGTGATTTTCTTTCCATCGTAACGAGGGCACAC
TCCTTCGCTCACTTTGCTGTATATCCAGAAACATTGATAGAACCACTGATTAAGGCAGGATGTCCTAAAGAAGTATG
TTCAAAGTGTGGAAAACCAAAAATGCCTTTCAGGGAGAAGACGGGAAGGACTTATGATTCAAGGAATGGTGAATATA
CGGATGAAGGAAAACATATAAGCAGAAAAATAGTAGAGACTAACCATGTATCTGAATCTTCAGTCTTTAGAACGGGC
TTAATACAGGAAACAAAAACAGTTATGAAACCATCATGTTCCTGTAATGCGAATTTTATTTCAGGGACAGTCCTTGA
CCCCTTCAATGGCTCAGGAACTACCATGCTCGTTGCCCGTAATCTTGGGAGATCAGCCATAGGTATTGATGTAAATC
CTATCTACAGTGAAATCGTCAAAAAGAGGTTGCAGTGGGGTATGGGGCTGGACATTGAGCATGAGATAATTGAGAAG
GTGCTGTCATGAACCTTTACAAATGTCTGTAGTAGGAGCAGATGATGAGGAATAATTTTTATTTATATGGGGTTATA
TTGTTGCCTTATTAGACGACAAGTTTATATACAACTATGATATACACATTAAGAGGTGAAATAATGTTTGAAGGATA
TGTGTGCAAAGAACACGAAGAAAACCTGAAAAACGAGAGAGTAGTGGAAGTAGTTGAAGCTGGGGAAAAAAGTCCTC
TTTCAGATGCACCGTTGCAGTGCATCTATTGCGGAAAGGAGACGAATGATATATTTGTCCTACTGGAGGACTTTTAA
CCGAAAAAGAGGTAAAAAAAATGCAAAAAGAATATGACGATCAGGGATTTTTTGATCTCCCTGATCCCGTTGACCAG
TGAGAGAACATGCAGAAAACAAGAGTGTGCACCCGGTTAACGGATGCAGACATTGAGAAGCTCAGGGAGAGGGCAGA
AGCAAGAGGCCTCACCCTCTCAGAGTACATACGGTTCGTACTCGAGAGCGGGCTGTAATGATCGAGTTGAATAAATT
AAAAATATTTAAAAGTTTTAATTTTCCTTCCAACACCGGAAAATAAATTTAATCTTTTTTCTCTTATAGGTTCTTCG
TTTGGTGTGTGGGGATCTCCCTTCGAACCCCCAGCCATGCCTTGATTAGAATTTCCACAAATTCCTGCAGTCCCACC
GTTTCCAGTCATTGTGCTTAACTGTCCACAAAGAACTCCTTTCGGGACACTCAAAAGAGGGAGGGGTGATATATAAA
TAAAGACGTCAATTTTCAAGTATAGGTTTTTGGGTACAGTAACCTCTGGACAGTTGACCTTTTGATCTTTAAA

>AMDV3b_Contig17858
TGGCCAGGAGCAGGCCAGATGATAAGTGGATCCAGGAATCCAGCAATCCAGGGCACAAGGGATATACAAGGGCCTAC
ATCTTCAAGCTGTACGGAGAAAAAGCATTCAATGATGATGCAAACCACACAATCAAGCCGGAATATCTGGAGAAGGC
CATCGAACACGCAAAGGAGAACCATGAGATAACGTGGGAGAAGAGATTGGTCAGGGCAAGGACGCTCAAGGATCTGA
GCAGGAAGAAAGTCGCATACGGTGAGCCGAAGGACGAAGCCGTAAAGCAGGCAAAGGAGCTGAGAGATAAGGGAGAA
ACATTCTAAAATCGTTCATTTCACTCAATCCTCCTCATAGTCATTCAAATCGATGAATCAAGAGACTGTTAAACGTA
GAGGCGTCAAGCCAATTATAGACTCACTCCTTTAGGTAATAAAAGTAGCTATCTATCTGCAGCATTGAATATTTATT
ACTTCTCGGATTTTTTACAGGACCACCCGTTATTTTACGAATGAAATACCCAGTACTTTCGTCTTCAAGTTCCATCG
AGTAAACCCTATACCCAAATAGCATTAGGAAGGGATTGAACAGCATTACATTCCAACTGAACGCATATAAGCCAAAT
ATTGTTACGAGAACCAGAAGCAACAACATACCCTTTAGGCCTCCCACAACTGCTACTGAGATAACTGACACAGCATA
GGCCAAAATGTAACTGGAAATAACACTTAATATCTCTCTCGTTTCTTTAACACTTTTTTCATCGGTTGTCCTATCGG
TATCAAGATGCTTAAGGTAAGTTAACCATAGGATAAACCCACAAACAGATATCGCTGTCATTGCGATTAGAAATGAT
GAAAATGCTTTATTCAAAAGCGTATTTGATCCATATGGAATACCGTAGATTATAAGTGATGCTATGAATGTAGGAAG
CATATTAGACAGGTAAAAGAATATCTTCCGACTTAGCCCAATATTAGTCATACTATTGACTGACGTTGTATTCTAAA
TAAATATTTGTAGTTATGTGGATTGATAACATAAAAATAACAAACTTAAATAATTTAGGCAATAATTTAAAGGTCGC
CCATGGTCTCTTCCAATTTTCTTAAAGCTCTGTCTTCCATTTCCTGCGCAATCCTCTTGGCTTCCTCGAGAGATTTT
GCATAACTGACATACTCTCCTCAGCTGAAGCATCGGAGGTTCGTAAAGGTTCATGACAGCACCTTCTCAATTATCTC
ATGCTCAATGTCCAGCCCCATACCCCACTGCAACCTCTTTTTGACGATTTCACTGTAGATAGGATTTACATCAATAC
CTATGGCTGATCTCCCAAGATTACGGGCAACGAGCATGGTTGTGCCAGAACCGTTGAAAGGGTCGAGGACAATACCA
TCTTTGGGACATCCTGCCTTAATCAGTGGTTCTATCAATGTTTCTGGATATACAGCAAAGTGAGCGAAGGGGTGAGG

CATTGTTGGGATTGAGAAGAAATCCCCTCCGTCAGCATCTGGATCATACTTTCTTTTATGAGTGAAGTTGTGTCCTG
CGTTGGAAACCCATGACTTATGCCTTACAGCAGGTTCTTTTAACACGTCTCCAGGATTAGCACCGTTCTCGTTGTAT
GGCTTTGTGTGGAGGGGATCATCATACCTACCATTTCTGTTCACTGCCATATCATGCTTTGTTTTCATGTTATTCAG
AATCTTGGGGATATTCCCTGTGTTCGTTGTAAGTGCGAAGTCAGCGGTTTTTCCAGATTGCAATTGATTAGGTATGT
TCTTCTGAGATATTCGCTTTATACTTCCTTCTGAAAGCGGTTTTCTGACTGAATCGAGATCAAAATAATACTTTCCA
CTCTTAGAGAAGAAGAAAATGTATTCCCATTTGTTGCTGAACCTGTCTCTGACGCTTGAAGGCATCCCGTTGGGCTT

>AMDV3b_Contig17532
CTACATATTGAGCGGCTTCATTCGGATTCATTTCAATCATTCCTCCTCATAGTCATTCAAATCGATGAATCTCCTGC
CGTCTATCTCTGTGTATTGGGGAAAGATACGAATCTCTCCTGTGCAGCTCTGCATCTTTCCTGACATACTCTCCTCA
GCTGAAGCATCGGAGGTTCTAAGGTCGCTATCGTTCATAACGTTGCCTCTTCAGGGATATCAGTGTTCCCTTTGTAC
GTATCCCTGTCTACGTACAATCGATCTCTGCCTTCCTCAAATGCTGGACGCAATGCTATGTTGAACGAAGCATTGGC
ATCGGCATGATCAACGTGCCCACACTGGCACTTAAAACTCTTTCCTGAACGGTTGCCTATCTGTCCACACCTTGAAC
ACTCCTTTGAAGTGTGGTATGGATCGACGTAGACAACAGGTATTCCAAGTAGTCTGGCTTTGTATTCTATCATCTGT
TCTAATTGATAGAAAGACCAGCTATTCAGGGAATACCTGAAATTCCTGCCTGACTTTGCATTTCTTATTCCCTGCAA
ATACTCAAGCTTTATTCCCATCCCATCCTGTTTCGCTTCCTGCACTATCTTTTTCGATACCTTGTGATTTATGTCTT
TCACTATCCTGTTTTCCCTGTTTCTTATCTCTTTTACCTTCCTGTATTTCCCGTTCTTCTGGAGATTCTTTCTCATG
TTCCTGTATTTCTTATGGACGTGATCTGCCTTCTTCCCAAGTTTTGTGACCTTTCCTGTCAATGGATTTGCTAAAAC
TGCTATGTGCCCTGTAGTGTTGCGGTCTATGCCAAGATATCCCTTCGGATTAATCATCGCATCCTCAGATACAGATA
CAGAAACGTATGCAAATTCCTTATCAATCTCGATCTGGTTTATCTTCAGAAAATCGTTTCTGAAGTGGTAGCCCAGT
TTAAGTTTAAGCGATGGGATTGTTATTGCCCGTTCTGCATGACTGAACTTGATTCCCTGTGCGGGTATGGTGAGCTT
GACACTTGTGATAACTTTTGCTTCTCTGTTTCTGCTATATTTCCTGAGAATCTGGTTTGATATGACCGCCTTGAGAC
CGAATTGCTTGACGTCTTTTGAAGAGAGAGTTCTGTGTTCAATGGCAAACTCTGCGACTTCTCTTGCTTTTTTTAGT
TCTGTGCTGAAATCCATTCCATGTTTTATTTTGAATGTCAATATCATACTTTCTTTTGCTCCTCTATGTATTTTTTG
ACTACTTCAAGCTCGAGTTGAATAAATTAAAAATATTTAAAAGTTTTAATTTTCCTTCCAACACCGGAAAATAAATT
TAATCTTTTTTTCTCTTATAGGTTCTTCGTTTGGTGTGTGGGGATCTCCCTTCGAACCCCCAGCCATGCCTTGACCG
GACTCTTTCAGCCCGCCTTGCACCTGTTGGGGTTCTCTTCCATTTTGCGACGCCTTTCATCGCATTCTCATGGTAGA
TCAGGCGATCACGTTTGGTTATCCTACGTTTGGTTATCCTACGTTTGGGTATCCTTCTCGTTTTTGCGACGCCTTTC
ATCGCATTCTCATGGTAGATCAGGCGATCACGTTTGGGTATCCTTCTCGTTTTTCTTGTTGCCATTATTTTCCTCCT
CTTTCTTTTGTGCTCTCTCTAATATTTGATTGAATTAATCTTCTGATACGTTTTCGCTCCCTGGGGCGATGTTTGTT
GAACGCCTCAAGTGCGGCCAGTTTCCTATCAACCTCTCCCTTTCCGTAATGCTTATCTGCCTCGTGTATTGCCTTCA
TCTGCTCCTTTTCAGGCGAATGTAAATTGAAGCCATACTTGGAGAGGCTCCCCTTATGCCTTATTCTTATCTTTGCC
AATCAATCACCTTTCCTGCTCGACTTCGCTCCAATATAGATCATGCCCAGCGACAGTATGAGTGTGAGGGCATATAC
CAGATCCATGAAGAGCGTTATGGACGGCATTGCTGTGAACCTTCCTGCCACATCGTATATTCCAGTCCCTATCCATC
CCTCAAAGCTGTACTTCAATTTTCCCTCGATAGCCGCCACTATCACATAGAAAAAAAGAAGCACACCTATGAGGAAC
AAGAGTGGCCTCATATCCTCACAGGTGGCTGAAGAAGTTCGTTATGGCGTGTGCAATGCCGTTGAACACTGCTGTCA
CGGGTGCGAAGAAGCTCGACACCATGCTCAGGACTGAGTTCCCAAGCTGCTGGAACATCCCCGCAAGCCCCTGCTGC
CCTGCCGGCGTTGTGATGGCGAAATACAGGGACAGACCGACTATCAGTATCACAAAAAATACGATACCAAAGGCGAT
CAGCTTCACTGCCCCCATATTCTCACCTCAGGTCCGTCCTCTTGGCGAACACGAATGATGCCGCAAGCAGCATGAAT
ATTATCAGCCCGTAGAACAGATACAGTATCCATCCCAGTGATCCTGTGGTCGCAAGGTTCGCAAGGAACTGCCCATC
CGGTATGAACACCTCAACAACGTATGCGCCCAGCAGATAGCCAAGTATGACAATGCCGAGGGTGTAGTCATACCTGC
GCCTTCTGTTTATGGCAGAGTATGCGTAGCCCAGTGTACCTCCAAGCAGAAGGCCCATTATCACACCGCCGGAACCA
AGAGCAAACATGCCCGTGCTTCCAAGGAAGTAGGATATGGACACAACTATTGCCATCACTATTATGCCGATGACTAT
GCCTGCAATTTCCCTGCTCTCAACCATCATGCCCTTTTCATGTTTCCTTGACATTATCTCTCCCTCCTTCTTGGCTT
CTTGGCATTCCCGCTCTGGCTCCAACGACACTATCTCCCTCCTTCTTGGCATTCCCGCTATGGCTCCAACGAACACT
ATGATGCCCAGGATAACCATGGCGATCTGCGATACTGTTATGACGTTTATGTCAAATGGCGCTCCGTAGGACTGTGT
TATCTCGACTGACAGATGGTTTGTCGATGCCAGCGCTGCTGAGTTGTTCAGTGTTATCCATATTGTGCCGTAGTATC
CGAAGCTGCCGGTAAGCATTCCAAGCGCATTCATTTCAACTGTCTTGTTCAGGTTGTTAATGTTGGCTGCCCTGCTG
TATATCACAGTGGAAGACCAGATGTTAGCCTCTATGACTCCCTTGTCTGTCAGTGCATGTGTGGATGTGTCGTTGAC
CTGCTTGCCGAAGACCACTCCCGCTGTCTGAAGCGCAAAGTTGTTTGTTTGACTGGCGCCGGTTGCATTCAATACCG
ATGTCTGCACGTTCAGCGTCGTGAACAGGTAGGATCGGTTCAGAAGCCCACCAACAGTGTAGTCAGGATTCTTCTGG
AATGCCACAATGTTCGTTGCGTTGAACCCATCAGGGAAACTGGCGGTCACAGTGTCTCCAGAGAGGCTCAATTTCGC
AGCCTCCCATGTGGGCGCTTTCCCGGTGCCGTTATAGCCGTACACCACTCCGATGTCAGACCCAAGACCGACGGTGC
TCGTCAGATTGCTTTTCCCGGTTGCCGCCAGTCCCTGAAATGACAGGCCGATGACAGCACCGGACAAAACAATAACG
AGAGCAACGATGATCGTGAGCCATCGCTTTTCTGTTATATTCTTTCCGTACATTCGTTTCTCACTTTTCCAAGCATC
TGATATGATGCTACAAAAAGACTATCTATACGCATAAGAAGTATATTAATTCTTCGGTTTCTGCTGAGACCGTATTT

TTCGGCTTCCCGTGCACGCATGTTATACAGCATTCCTTTAAGGAAAAAATCGTTCGACCTTTCATAAACAAATGTCG
CTACGACATTTTAATCATAAATGTACGTAAAAACTAAAAACAACCTCACCATACCATATTGATGACAACTGCAATAT
GGAAAGCAGCCAATCGGCAGGCCACAAGGATCCTCAGATATGGTGAGGTTCGGGAGATACTCGCTAACCTCGAGCTG
GAGCCTACCGGGGCATGGGCCAGAAAATTCTCTACCAGCAAGGGCACCGTCCAGATATGGCTCCAGGCACTGCTCTT
CACTGGAATGAGGTTTCACGAACTGGCAATCCTGCATTCACATCCGGAGCTGCTCCAGGATGACGGGACGATCATGC
TGGCCAAGAACATATTTTATGATGTGGGTAAGAAGAAGCAGACTCTTAGGGAAAGAGTCGTCTATCTTTCCGATATG
GGACTGGATATAGTGAATAGGTTCTTTGACGTTCTGCCGTTACATTCGGATCCGGGAGAGGCATTTAGATACCTCGA
CGCTGTCCTGACTGCCACAGGAAAGAAGCTGGGATACCCTGAGAGGACGATTACTTACAGGGAGAACAAGACAATCG
GCATTCAGGAAATCGAGAAAAACGGCAAAAGGATGGCCGTGAGGCAGACTACCCCGGTAGAAAGGACACTCGTCACC
AATGGCCTGTCCGTCCGTGTCTTCAGAAAAACATGGAACTCGTGGCTCGTGAACTACTTCAAGAATGACATTCATGC
CCTGACAATCGCAGAGATGTCCATGGGGCATACCGATCAGACAGCGTTGAAGCATTATCTGACGCTGCAGTTTGATG
ATGATGACATGAAGGATATCGAGAAGGCTGTTAAAGGGTTCGGGCTGCACCCGGGCTGGAGAGAAAAAAAGAAAAAT
CAACAGCAGAACCCGCAAAATAGGTTATTATAAATGAATAATTTAGAAAGCCAATAACATAGGTTAATGTCTACATG
AATATTTGATCAGGTTCATCATCGTGGTTCGGCTTCTGCACGACCTTGACAGTCTGCTTAACCCTCTCGCCATCGGG
ATATTCATTCCGTGAGTTGATATCATACATATCCCTTATGATCTGGTCCACTTCCCGAATCACGCTCTCAGGATCCT
TGTCATACACAATGGGATTGGTGACCTGGGACTGGTCGAATATCTCCTTCAGTGGAAGCTCTCCGAACTCTGCGCTC
ATATCCGAAGATATGTTCTCGTTTAGCAGTTCCGTCAGGGGATGCAGCGGCACCATGCCAAGGTCACGCCTCACTTC
ACCCTTTCCGTTGGGTCGAGATAACGTGCAACAGCCGTGTATTTGAGCTTTGGCCTTCCGAGAATGTCCCTCTTTCC
GG

>AMDV3b_Contig17701
CTTCTCCTCACTCTTTGGAGATTCCACGCCTTCACTCTTTGCGTAGTAGAAAGGCTTGGCCATTAGTTGAAGAGCAA
CAGGCCTGTATCTCTCTGCGATGCCTGAAGCATCAATCTTTGCGGCTATTCCCTTGAGATAAGTAAGATAATCATCC
TGAGAGATCATTTCACATCCCTTTCCGATTTGAATTTCTCACTCAGGGTTTCGTCAATCTGCAGAACCTGCTTCAGA
AGATCCTCATCCATTGACTCTCCGTATAATTTGATAGTCCACTGGTGCTGGCCTTTGGCATTCGTCCCAAGTTCAAT
TGACGCTGGCTGGGAAACAACATTAATTTCCCTTACTTCTGTGGCTGGCGATTCTACGGTGCTATTCACGATAGCAC
CTCCATTTCTTTGGTTTGCGAGACCATAATGATAGATAACACAAGTAGCTTATAAATGTTATGATTGACGCATAACT
TTCGATAATGAGCATAAAGTTTATTAACTTATGACGGTTAAGTTTTACAGATAAGATGACAGAGGATGAAAAGCTTG
AACGAATAGGTCTCAACCTGAGTAAAAATCTTAATGACTTTCTCAGGAATCATGGGATAGGCCTTTCTGAATTTTAC
AGGAAGGCGGGAGAAGATTTTTTACCTCGCTACCTTGAGAAAATTCAGGAAGAATTAAAAAAGCTATGATGGTTTGT
GACGAATGCGGGAAAACGTACCCGCTACCGCAAGATAAGGAGGAGTCCTGATGTGTGATATCTGTGATACCGACGGA
GAATACTATACATGTAACTGTCCTGATTGTGAGAGATGCGGGGAAAGGGTATGCCCAACCTGCCATCACATTTCAAT
GATGTATGAAATATGGCCTGATGACGATGAGGATGAGGAGGACCTGGCACCACATCCAAACCCTGCTAAAGATGGAG
GTAAGCAGAAATGAATAGAGAAGAGAAACATATTCCGGACAGGAATAATTCTCAAAAGGATAATTCGAATCGAGAAT
TTAACCTTACCGCAATCCGTACAGAAGCCGATTTTTTCGAATTTCAGGATAAGTGTAATGTAGAAGGGATCTTCGGC
ATTAAAAAACGGCTTGAAGTGCTCTCCAAAGTGGATATTTCTCCGGAAATGAGAAAATATATTATAGACTTTTACAG
GGATGACGCAATAAATAATCAATATGATCAGGAAGCCCTGGACGCATTCAGGAAGACATTCAAGAGCTTTTCCAATC
TCTATTTCCCGGGTCCCATTTATATTGTTGAAGCGCTGAACAAGAGATTTCATTTTCGTACGGTCCCCGAGATAGGA
ACGGACAAAGAGCCCGTCTATTACTACAACGGCCAGATATACGAGCGGGCAGAGGAGACCATAAAGGCGGAAGCACA
TAGGGAATACATACGGCAATGGAAGGAGATGCTTGCCATTTCTGATGATAAGGAGTTAAATTCACGGCTTAAGAATG
CCCTTGACAGGGGGCCATCCGCCAATGACATAAATGAAGTGATTTCCATGATCCGCCGGACCACATTCACCCATGAC
GAGATGAACCCTGGAAGCCATATTCCGTTCCTTAATGGATTATTGAACCTTCAGACAGGGAAGCTTGAACCATTCAG
GGCCGATCTGTTCTACACTTATCAGATCAATGCCAGCCTTCTTGATCGCATTGTAACACTGAAAGACGTTCCCATGT
TCTCAGGGCTTCTGCACACGGCGTTTTATCGGCGGGACATTCCCACGGTTCTCTCATACTTTGCATATTCCTTTTAT
CCTGACCTCCCGGTCCACAAAACGCTATTCATTCTCGGCAGGGAGAGGATAGGGAAGGGAACCCTCGTGAGAATCAT
TCAAGGACTCATGCCGAAGGGGTCAGGTTCCATATCACTGGCCCGGTTGCTGACCTCAGATAGATTTCAGTTCACCG
GAATTGAGGGCAAGAACCTTCTGATCGACAGCGAGACCAAGAGGAAATTCAAGAGAGGAACCATCATGGAGTGGTCC
GCATTCTGTAATCTTTTCGGAAAGGATGTTCTGAGCGTCGAGCCTAAGGGCCATGAAGCCCATGACTATGTTAGTAA
AGCAAAAGGGATCTTCCTGGGAAACCTTCCATTCATTCCCATCGACAGTCCGCCGGCTATATCCCGAATGCTTGTTG
TTGTTACGCGGGATGAAAGGCCAAAGAAGGTCATTCCAGATCTCGACAAGAAGATACTTGACAGCGAGAGGGACCAG
ATAGCAACACTGTTGATGCAGATACTTTTCAAGCTCATGGACCGTGATTTTGTTTTTCCCGGCCAACTGTCCGATGA
TGAAACCGCAGAGCTGATCGAAAAACTGGCGGATCCTGTTGAAAATTTCATTGAGGAGCAGACAGAATACGACAAGG
AAAGCACCGTTCCTGTAGACGAAGCACATGAAAGGTTTGAAGAGTGGTGCAAGAGCAAAGGCATTCCGACCATATCA
AGACAGACATTCGTTAAGAAATTTGGCAGGACATACCAGAAAAAGAGGATTGGAACGCATGGAAAGGGCGGATACGT
CTTCTTGAATTGCATACTCTATGAAGAGGATCTGGAACTCAAAGCCAAGACCACAGATCAAGTTGGAACCCAGTCGA
GTTTTTCAGAAACCCTGAAAATAAGGGATTCTGACAACAGCAAGGATGGGTTCCAACATGCGTCCCATGACCCTTCG

CGTGTACGGGAGGAAAATGCCGATCATGATCATGATATAAAGGTTAAGGGACACAAGTTGGAACCTGATGAAGAAAC
TTCCGGAAGCCCTGAAAATAAGGCTCCCGTAGAAACTAAAAAAGGTTCCAACTTGTACCAGAGCCATGAGGATGTCC
GTCCTGGGTATTCCAATAATTCATCGTCTGGCCCTGTAAATCCCGATGTCCAGAATGAAAACTCCGGATTTTACAGG
TACCGTGTCCTGAAAGATTTCTCCTTGTATGGACGGACATATCATCCCGGAATTGAGTTCGTGTTCCTCAAAAAACT
CGATGGTGAGGTCAGTAAGGGTAACCTTCAACTTGTGGAAGAAGCAAGTACCGTGAAAAATTCCACTAAAGATGACA
AAGACCACAAAGACATGGAAAATGAAAAGGATAAGGAAGTCAGGGCATACATCGACAATCTGAAAAGGCAATTATCG
GAGAACCCAAAGAAGGCCTTATTCGATTTTGTTGAGAACGAAGCGCCAAAGACAAAGTATCGTTCCATGACGGCAAA
AGCAATCCATGACATGATCCCTGACCCGGAAATGACAGTAAAGCGTGTTTATGATCTGTGTGAGGAACTCACCAGGG
AAGGGGCATTCATAAAGAACAAAGCCGGTGCATATTCCGTAAATGCAGAATTTTTGAACGGTGGTGGCTACCAGTGA
GATTCATTCAGGCCACGCCCAGGCAGATCCAGGCAATAAAGAACCTGTGCTATAACCGATCCAATATCTATTTTGTT
TATGAATCACTGGAAAGGCTTGGGAAGGATTCGCTATTCCAGCTGAGCATTGGAGACGCTAAGGAGATCATATCTGC
ACTTGTGGGTGATAAACCATGAACAGATCCGGGCTTCTCACTGTTACCCTGGTGCTGGCATCCTGGGCCATATTCCT
AATTCTGATCCTTGCCATGTTCTACCCTCAGGAACTTGCTGCATTCTGGAAATGGCTAATGGCGCTGGGAGGTGAGT
ATCCATGATCTCCCGGTCCAGTGCACTGAGGTTGAGGGGCTTCGATGTCAGCGTAACCTACAGGCGCTCCGATGGTG
CGGTCCTAACAAGGACCGGCACACTAAAGGGCGTGTACAAGAACCTGCTCATTGAAGCCCCGGTCTCAGGCATGAAC
TACCGCATACCACTCATACAGGTCATGGCCGTACGGCCACTGGACCCGATCACAGTCCACAATTTTTTACAGGACGG
CATGGAAGCTGCCTTAAGCTCCGGGAAGGAATGAGAATGAGCATGAGAGAGGAAGCACTCGGGAAGTTTGTTAGCCT
TATTCCCGAAGGCATAAAGGGAAGGCTTAGGGCCATTGAGAAAAGGATAACGGAGGTACTGAAGGAATGAAAACAAG
TGAAAGAAAAGAGGTTAAAAAGGAAATGGAGGGCCCCAGGAACTATGGCAACCCTGAACCCCTCAATACTTCACATA
TTGGGAGATCGGTAACAATAACCTTAGTGAATGGAAGGATTGAATCGGGCAAACTGAAGGCGCTGGGAGCATACATG
TTGAGCATCGAGGGGCCTAACGGAAGGGAACTGATCATAAACAAGGGTCAGATCGTAACGGTGAGTGTGCTATGAAA
TCCCCTGCAATGAAAGGCATACAATCCTCACTTTCCCTATTTTCCAGTCTAAATTCAGGTCATAATGTAAAACAACC
CGTTGACACCCCTTCGAATATGAGAGATACTACAAACAGAACAAAATCACATTCACAGGGTATCGCCAGGGCAGGGG
ATGAGATGAACTGGATTGCATACTGCTCCAGATCCAGTTACGATTATATGAGCAGAAAGCCCACGACTTAAGTCGTG
GGATGAATGCGAAGAGTTTATATAATGTCAACTATTTCTATTAATATGATTAAAGCCCTTAAGGTGCGTTTGTATCC
GACAGAAGAACAGATGGTTCTTCTTGAGAAGCACTTCGGGGCAAACAGGTTTATCTGGAATCATTTCCTTGAAGCAA
GGACGAAATATTATGCAGAAAGCAAGTCAAGGGGAAAGGCACAGGGACTGAACTATTTCCAGACCTGCAAAATGCTA
ACTGAAATGAAGAAGGAGAAGGAATGGCTGTATGAAATCTCCAGCTCTACCCTGCAACAGACCTTGAGGAAGCTGGA
CAATGCATTCACATCATTCTTCAGGAAGAATACTGATTATCCGAAATTCAAGTCGAAGAAGGACAACCAGTATTTTG
TTGTCCCTCAAAATGGGAAGATCGTGGGTAATAGGATAGTGATCCCCAAGTTTCTTGAAGGCATAAAATTCAGGGAC
AAGTTTGACTTCAGTAAAATCAAGGAGATTGACCAGATAGTAGTCACCAAGGATGTGGAACGCTATTACGCATCCAT
TTTCTATGAGTCCGATGAAATCCCTGAAAAAGCTGAGGCAGTTGTAGGGATAGATGTCGGGGTGAGGCACTTCATCA
CAATGTCCGATGGTATTCAGGTAGAACCACTGAACGCATACAGGAAGATGGAGAAAAAACTGAAAAGGGAACAGAGG
AGGCTATCAAGGAAAAAGAAGGGATCGAAGAACAGGAGAAAGCAGATAGTGAAATTGGAGAGATTGCATCAGAAGAT
CAGGGATAGCAGGACAGATTTCAACCACAAGGTGTCATCTGCGATAGCCAAGCACTATGGCATAGTTGTGGTCGAAG
ACCTTAACATTCAGGGCATGGAACGGAACAGGCATCTCTCCGGGAGCATAGTAGATCAGGGATGGCATCAGTTCCTC
TCCATGCTGAAATACAAGATGGAGTGGAGAGAAGCAGAACTGATTAAGATCGGAAGGTTTGAACCGTCATCAAAATT
ATGCTCCGGATGTGGAAACATCAAGCACGAACTGAAGCTGTCAGATCGCACATACTGTTGCGATGTGTGTGGTCTTG
AGATTGATCGTGACCTGAATGCGGCAATAAATATCTGCAACATTGGTCTAATTAAGGTAGGGAAGGACATTCCCGAA
TTCACGCCTGTGGAAAGTGCTACGGCGGCGGAACTCTCAAAAGGAGGTCTACGAGTTGCCACTCTATGAATCAGGAA
CCCCACGTTCTTTAGACGTGGGAGAATGTCAGGGTGAGCATCCTATGACATCTATCAGAAAAATGGAAGTGATGGAC
ACCATCGAATTGGAAAAGAGTCTGGAGGCCCTGAAATGAACAGAAAAATTCTCACCAATTGGCTTCGTGCTGGCTTA
GCATTGCAGCTGATATGCGAGTTTGACGGATGCACAAGGAAAGCGGAATACAGCATCACGAGGTGGCATGGCCTCCA
TATAGTGGTATGCAGGGAGCATTTCCCTCGGATCGGCGAACTGGATGCAATGGATGAGGAGGAAGTGAAATGATCCA
CATAGGTGGCTACTGGGGGAGGAAATACGACTTCACGAGATGCCAGAGTAAATT

>AMDV3b_Contig17883
ATACTTGTTACCGTTTATAATTACTATATGACCCTGATCCAAATGCTTTCTGATATACTGACCTACTGTTTTATAAT
CAGTTAGATATATTTCACGGTTATCTTGCTTATTCGCGAACAACCCTCTCATAAAATAAACCAGGTCATGTCCCTTT
ACTTTGCATTCCGCACAATGATATTCTACACGTTTAAACATTTCTCAATTCACTCCATCCCTCATTCAGGGATAATA
TAGTATTACAAACTAGGATATAAATGTTTTGTTCTGGTCATTTGTTTTCATTGAGGACGATAAAAATAAGATGTTGG
ATAAATCAGTCTGTAATGCCCAGTTTCTCTCTCTGTTTGACTATGTTCTTTATCTTCTCCTTGTTAGGAAGGCTGGT
GGCATCATAGTACGTTATGTTTCCATCCGATTTCACGATGCCGTAATTTACATATCTGTTTTTCCAGGACCCTCCGT
CACCGACTTCTACGATTGTGCCCGCTTTGAGATTTTCCATCGAACACTTTTTTTGCATTATGTTTATCATGGTAGCT
GTCAGTCTCCAGGAACACCCTCTTCAGGCCGTACTTTGGATCAGTTCCTGTAATTTTGGCAATATATACTCCTCCCT
TCCTGGGCTCCCCTTCAGCTGTGATACGCACATGTACTTCCTTTCCCAAGCTTGTTGCTCTTGGTTTCGGCTGTGCT

GCTACCCTGGCTGATGGATTTGGTTTCACTGTGACCTCCCGAATGTTTCCTTCCTTGTCCATCTTGAGGTTTCCAAC
GAATGGGGCATAGAGGTCCTTTGATCTGCTGGTCTCAATGACCCTGACTTTTTCTCCCTTATCTCTCAGCTCCTTTG
CCTCCTTTACGGCTTCGCCCTTCGGCTCACCGTATGGGTCTTTCTTCCTGCTCAGATGTTGCACGTCTTCAAGTGCC
AACTTCCGAGGTATCCCGAATGGCTGTCTTTTCTTCAGTGGATATACTTTCCTGTTCTTTCCTTTTCCTTTGCTTCT
AAATTCTGTCATTTTTCATTCCTCCTCGAATTTTCCTCTGGCCACATAGGCTATATCGCTGTCGATCACACGCTCCA
TCCTCTCATTTCCATTGTACTTGGTTCTTAGGAGGGATAGTTCTGCCATTGTATCCCCTTTCCCATATTTGCTAACG
GCACGTGTGAGTGCCTAACGGATTTAAGTCCAAACTGCTTGACATTCTTTGAGGACCCTTTCTGCTCCTCCTCTATG
TATTACTATTCTTCTTAAAATACTTATCGCAATTCATCTCCTCTCTAAAGATCGGAGCTTTCTTGCTCATAAATTTG
TAACCACGCGTATAAACGTTATGAACGATAGCGACCTTAGAACCTCCGATGCTTTAGCTGAGGAGAGTATGTCAGTG
TCCAACAAAGATTTCCTTTTTGGAATATCAAAATGCAGGGAGGGGGTTTATAAATAAGGAGTGGTATTTTTGACCCT
ACCTTTTTTGGACGGAATTGATATAATTGTAAACAGAAGATTTCTTTATACCCAGATATTGAGAAATCTCCCGTATT
GTTTTTCCTGCGTTCCTCAGTTTTAAGATCTCGTCAACTGACACGTTCCGTCTTTCCAACACAGTGGGCCGTCCTAT
TTTCTTTCCCTCTTTCCTGGCCCGGGCCATGCCATCCAGAACCCTATCCGAAATCATGTCTCTTTCCAGTTCAGCCA
CTGCTGAAAGTATCGTCAGCAAAAACTGTGTCTGCGGATCCTTCTCACCAGGATTCACCCTTATACCTTGGTCAATG
GCTTCAAAAAATTTCCCCTTATCCTTGAGCTTCATTACGTTTGCCGCAAGCTGCATCGCAGATCTTCCGAACCTATC
AATCCTTAGCGTCAGGACTCCATCAAGATTATCAATTTCATCAAATAGTTGCTTTAATGACGGTCTATCTGCTTTAG
AACCGGACGCATAATCTATATATTCCCTGTATTCCCATCCCCGTGCCTTGCAATAATCCCTTAGCCTTATGAGCTGG
TTCTCAGGATTCTGCTCTTTATCCTTCGTACTTACTCTGGCATACAGCCCCACTGTAATTCTTTTCTCTTCCGTCAT
TTTTACCTCTTTTATTTTTATTCACCGCTCATTTAAAAAGGTTTAAGCCCTGTCCTTTGTATCTGGCATATTATTTC
ACCTTGAGTTTTGTGGTATCCGTGCCGATCTTTCTCTCCATTTGCTTCCATTTGGTCCCTTGATTCTATGGGATCCA
CTATGACTACCTTCCCATTCACTAGCCAGACTTGCCTGGCATACTTCTTTCTCGGTTTAATCCTGTATTTTTCCTCT
TCATTTTTCATATTTTTCAATCCGGTATTTAAAGCGGGCAAATTTTCGGCCCAACGTCCCCTCTGGAATCATGTATA
CATGATAATAGGAAGGCAGCCGCCGCAACATGACATTTTCCTCTACAGATGCAGTCGGTGCCGGTTTCTTTGCTTGT
GTCATGTCGATGTTATATGACACTTGGCATGACATCATTTTTCCTCCTCATGATCGGTCTTTCCATCAGGGTCGTTC
TTTCCCTTCTTTTCCAGTTTTTTTGCCGCCTTCTCTGCAAGCATCTCCATCAGATGTTTCATATCGGGAGTGTTCAT
GGGATCGAAAGGGTCCACCTTCTTAATCTCTGCAATGTCCTTTCCGGTGTTGACATAACCCTCACAGTTCGGGCATA
CAGTCCTTTTGGCACCTCCGGTATAGGTCCACGTATAGCTGCAGTGCCTGCAAGTCACCCTGATGCCGGTATCCTCT
TCATCCATCTCCGCTACATAGACTCTGTGATCGCAGTTAGGACACTTTGCCATTCTTCTTGACCCCGTGTAAGCCCA
TGCATAGCCGCACTCGTCACACCTGATTTCTATTGCAGGCTTACCGTTCTTCAAAGAGATCTGGCCACCGTCAAGGA
GATTCAATTCCTCCTGGAACTCCTTGAACCTCTTTTCCATGTATTCCTTTTCTTTGCCTCGTATTTCTCCCTGAGTT
CATCGGATGGCAGCTTGAATTTCACAAGTTTGACTGTTATCTCTGACATGCCTCTGTGTATTCTCGGGTACTTAGCA
AGCGGTCTATCAGGATCAAATGGATTCCTGGAGAGCAATTTCGGCTTCATCAAGCCCTGGACATCGGTTGCCTCGAA
TCGGATATGTACTAGTTTTCTCGACTGCCTTTCAATGAATGATTCATCGGGCACTGTAATGAATGTAATAATCTGCT
TGTACCTGAATCCCTGTGTAAGCATACCGAATATCCTTGCACTTACCTCCTGCCAGAGTCTTGCATTGATTCCAAGC
CCTGCGTCATCAAAAATTATAACAGATCCCGGAGGAAGTCCCAAGTTGACAAGGGCAATGAAATCCTTCACTGTAAA
AACAATCCTGTCCACATTGAAATTCGGATCAACAGCTTCCGCCAGCCTAATGGATGACAGTGATTTGCCCGAGCCTG
TGTCCCCAATAAACAGGGCAATTAGATTCCTGTTTTTCTTTATCCTGTTTCTAATCCACCTTAAAAGCCATATGCTG
TTGGCTTTTTCATCATCTTTTTTCCCGTCTTCATTCTCTCTATTGACGTTTTCCCTGCTTATATCTTCAGTCATATT
CCCTCTTCACCTCCTTCTGTATATGTGTAATGCATTGGAATCAGCAGACCAAGCCTTCCATAGAGCCTTGAAAGAGA
ACCCATCCACGAGAAAAGTATATCATAGCTCTCAGAGACGTATCTTTCCGAAGGATACATTCTTGACGCCTCCATCA
ATTGATTCCTGGCGATTTTGTATTCCCTGGTATACTCCTCATCAAAATAGGACGACATCAGCTCATGAAGAGCCTTG
ATCTTCTGTCCCCACAAAAATACATTTCCAAGATTGTGGGGAGATTCTCCGGGGAATTCAGATACTTTGACAAGCAG
CTCATAAATAACCCTCATGCCGTTTGGTGATTCATTTTCAGTCATTCTGCCGCCTCCTCTTTCTTGTGATAACCATC
ACTGGACAAACAGGCCCGTCTATCGAATACTCATCCTCCTTCTTGGGTTCCTCTTCAAAAAGAGATGAGAGAGGGGG
CGGAAGTATGCCTTCGTCCTCTTCCTGGAAATTTCTGTATTGGGCCCTTAGATAAACGCTATCCTCGATAGATATGC
CCATCAGCTTGAGCCTCCTGAATGCATTCTCCCTCTTCCAGTCCCTTACAGTATTCCTCATTTTTCTTCCACCTCTG
TCTCCTTTTTCCTCTTCTCTGCCAGTGCCCTCAGGAATGTCGGGTTCTGCAGGAGCTCCTCTACCTGAGCCTGATAT
ACCTGGGCCTGATCCTCATCCAGATCCTGTTTTTTGATAATGTCAAAAGCACTGTCTCTGAATGCACTTTCTATGTG
GTTTATCATTTCCTTCGTCTGCCTTGTACCCTCGACTATGGCCTCCACCTTGTTTGTCTGCCTCAGTTTTGCAACCT
CTAAAAGTGCATTCTGCAGCATCTCTGTCAGATCATCCAGTACACTTGCAAACATCATGTATTTGGCACGGTTGTAA
TGAGCCCACGCAATTTCAACAAAATACGGAATTGATGTTCCTGGAAGGTATTTGACTTCAATGCACCAGTAAGTGAC
CCTTCCGTTAATTTTCCCCTGCTCTAAGGTGAGTCCTGTCTTCTTCTCGATTAAAGCCCATATCTCTTCAGAAACAA
ATATACGATCCCAGTCCATTGCTCCATCCTGCGGAACCATAAGGAATGTGCCCGACCTCGTCCTAAGGCTTCTTGCG
TAGGCTGACAATGCAGGTACAAAAATTATAATTGATATTATAGTGTAGACTAGGCCTATAATCCTATATTCAATGTA
ACTAAGGAATATGCCTAAAGTCAATTGGGGGATTAAGGCTTTTTATCAGGCTCCTTTTTGATGTGAATTTCCTTAGG
GAATCCACATTGCCGTGTTCACTGATTACCTTTCTTAATTCATCCTGTATCTCCTCATGAAAGCCAGATTCAAAGGC

CTTCTCCCCTGTGCCTCTCGAGCCATCCCCTGCCCTGCCTGTCCTTCATATAAAGAACGCCTCCCTTTTCTCTTTTT
CGAGGATATATTTCCTTCCGGAAAACATTGCAATTGCAACACCAAATATGGCGGCAACGAAACCCTTTGCAATTCCT
ACTTCATACTGGGAATATGTATAGAGAGCTGCAGGAGGCTT

Fig. S1

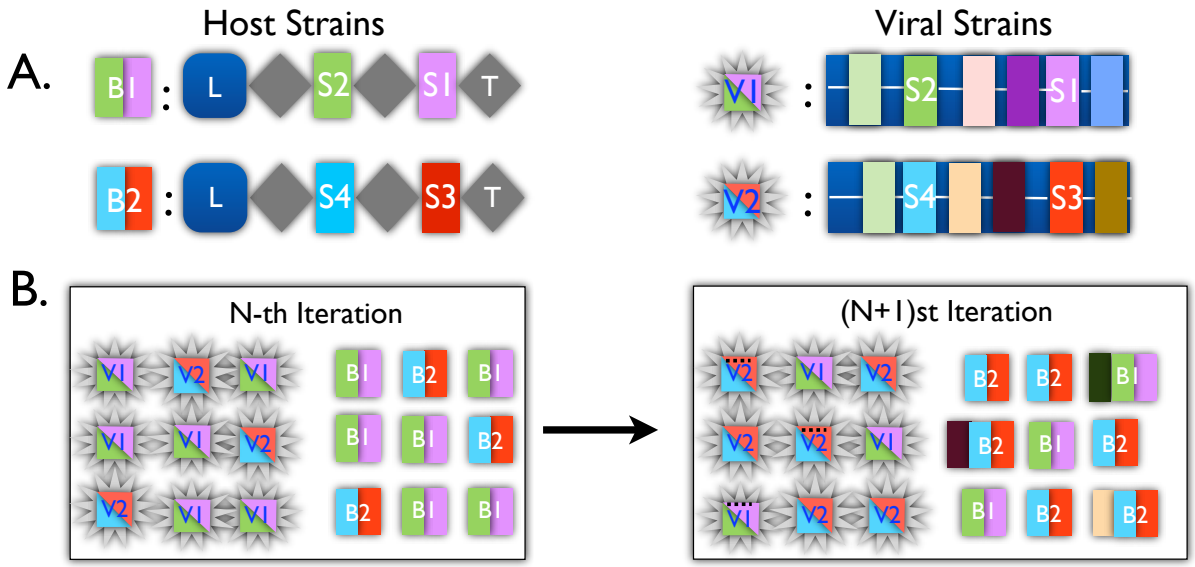


Figure S1: Schematic overview of the interaction-based mathematical model. **(A)** Host strains (rectangles) are defined by spacer content, with virus strains (stars) defined by corresponding proto-spacer sequences. The full mathematical model considers all proto-spacers in defining viral strains, but for ease of display this cartoon only tracks the fitness-impacting viral proto-spacers matching current host spacers. **(B)** Diagram of a representative iteration. Model-stipulated 'well mixing' results in dominant host strains being virally challenged more frequently causing negative frequency dependent selection. Thus, the initially frequent host strain (B1) is depleted by the newly dominant viral strain able to productively infect it (V2). Clouds of host and viral strains emerge as viral strains mutate (dotted black lines) and hosts incorporate random new spacers unidirectionally (new colored bars at left ends of hosts). The model is built to predict the patterns of virus-host coevolution that emerge after thousands of iterations.

Fig. S2

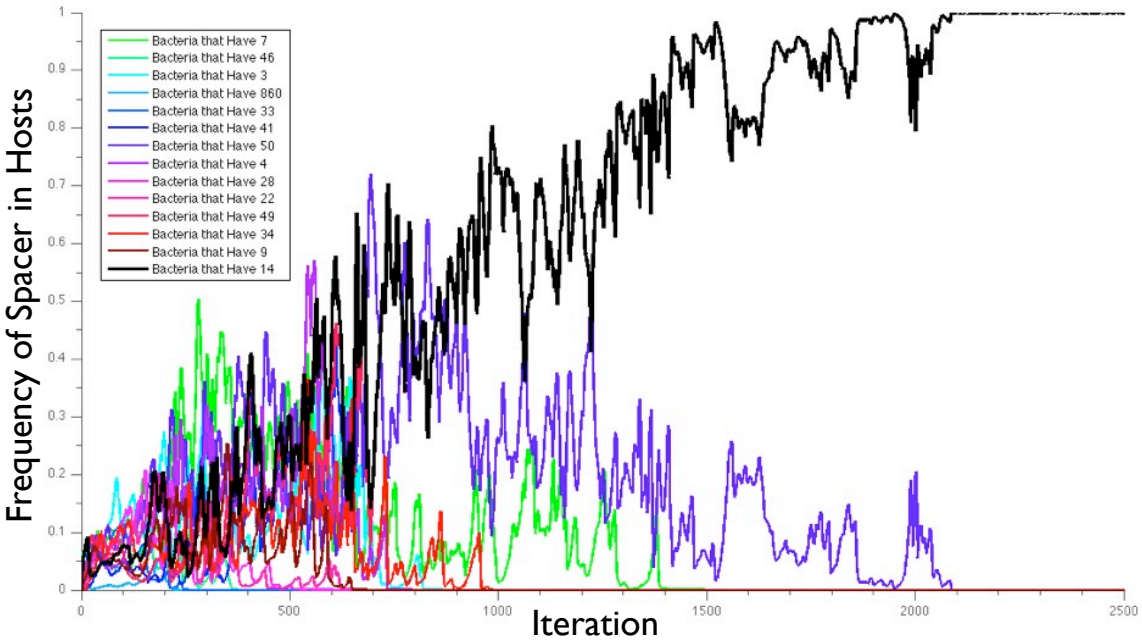


Figure S2. Gradual loss in spacer diversity at one locus position. Tracking 14 spacers incorporated at the second locus position by distinct host lines for the simulation shown in Fig. 2. In contrast to the rapid selective sweep observed at the 104th locus position shown in Fig. 2C, the 2nd locus position is characterized by a gradual fixation of one spacer (lineage). In contrast to the negative frequency dependent selection ('kill the winner') which occurs in a single iteration (fig. S1), across thousands of iterations, positive frequency dependent selection is clearly evident. This occurs as host lines of low frequency go extinct throughout the simulation.

Fig. S3

Figure S3. New-end locus diversifications post-sweep. In Fig. 2B, optimal clustering analysis predicted a selective sweep prior to the 3800th iteration. Yet, by the 4300th iteration, a number of distinct sub-populations were identified by the silhouette-based clustering algorithm. Reconstructions of host loci at 3 representative time points--before the sweep, immediately after the sweep, and 500 iterations after the sweep--shows that the new clusters correctly predict new-end diversifications of the sweeping sub-population. A second selective sweep (T=4800: ii--> in Fig. 2C) selects for the lineage of one of these sub-populations.

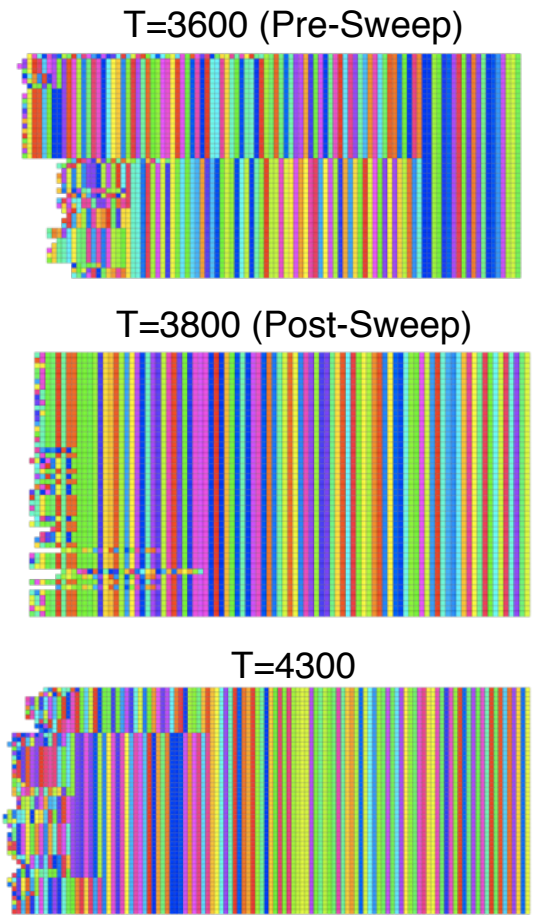


Fig. S4

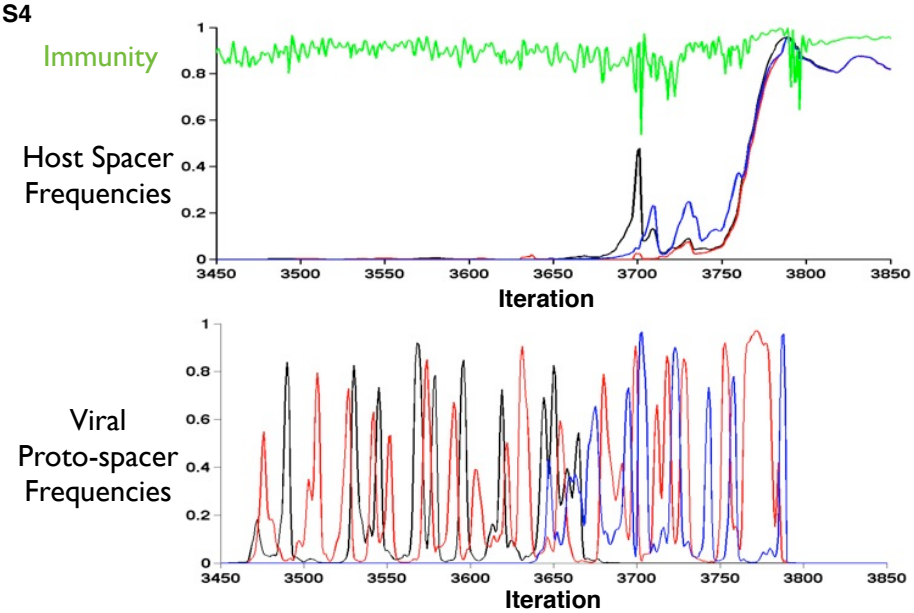


Figure S4. Selective sweep (Fig. 2C) results from viral delay in mutating multiple proto-spacers on one line and is not bottleneck driven. In the main text (Figs. 2B,C), a selective sweep of a host line with a single spacer type (spacer 517434: shown here in black) was shown before the 3800th iteration. In both host (top panel) and viral (bottom panel) populations, we tracked the frequency of this black spacer along with the next two spacers added by its sweeping host line. The frequency of spacer 518987 is shown in red and spacer 544610's frequency is shown in blue. Frequencies of the two corresponding proto-spacers inversely oscillate in the viral population (bottom panel), implying that viruses fail to lose both spacers on a single line until just prior to 3800, after the sweep. The host line thus sweeps due to immunity to both circulating viral clouds. In fact, the sweep occurs during a period of high host immunity (green curve in top panel) and is not induced by a host bottleneck. Immunity declines after viruses mutate all 3 proto-spacers post-sweep.

Fig. S5

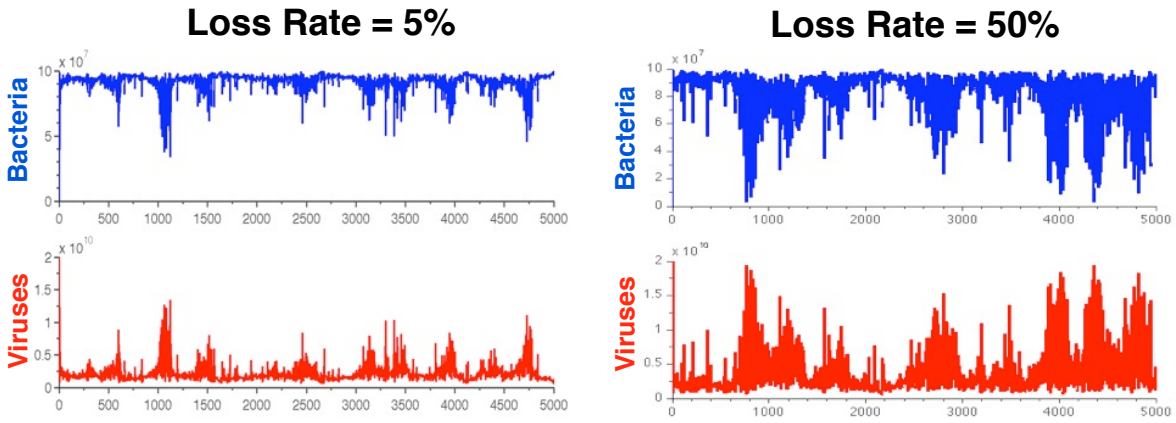


Figure S5. Higher loss rate increases likelihood of inferred viral blooms. Predicted relative abundances for host (blue) and viral (red) populations tracked across iterations. The left panel (low-loss rate regime) shows predicted relative abundances for the simulation in Fig. 3A of the main text, while the right panel (high-loss rate regime) shows predicted relative abundances for the simulation in Fig. 3B. Host levels represent the number of immune host interactions and viral levels the number of productive interactions multiplied by a laboratory-measured viral burst size of 200 virions per interaction.

Fig. S6

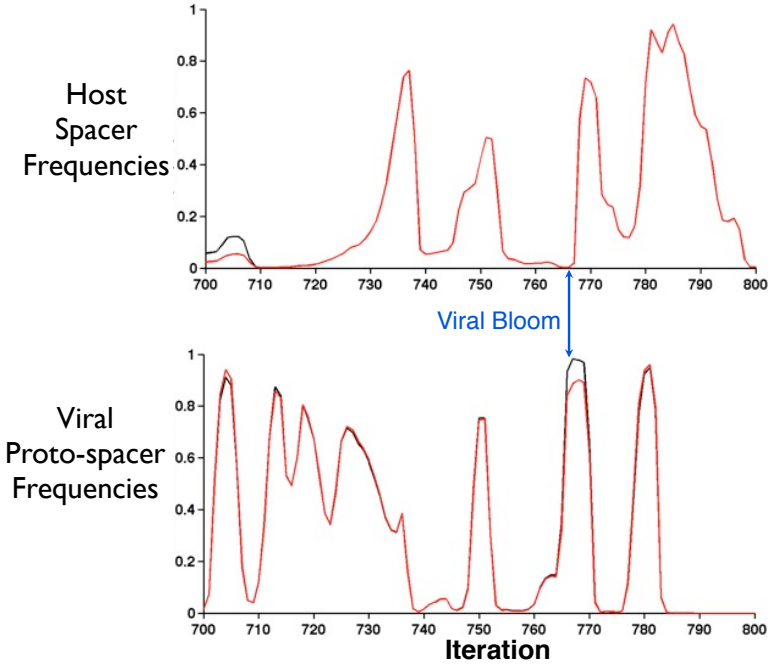


Figure S6. Predicted viral bloom (Fig. 3B, lower panel) occurs due to host spacer deletions. In the main text, a nadir in host immunity was shown at the 768th iteration in Fig. 3B. Hosts with two key older spacers survived this predicted viral bloom. Here we tracked the frequency of these two spacers through the bloom in both host (top panel) and viral (bottom panel) populations. Spacer 39184 is shown in black and spacer 49611 in red. Note that most hosts lose these two contiguous spacers (Fig. 3B) prior to the 740th iteration, when almost all viruses have mutated the corresponding two proto-spacers. Yet, a small remnant viral population maintains these two proto-spacers, proliferating and diversifying against newly un-immune hosts.

Fig. S7

Figure S7. Viral proto-spacer diversity during the predicted bloom (Fig. 3B). Each of the 651 viruses at the predicted bloom (Fig. 3B) is shown along the rows. Horizontal positions along a row represent the virus' 50 aligned protospacers, with distinct protospacers colored distinctly. The strains are then clustered based on their possession of the two older spacers that conferred immunity at the bloom (Figs. 2B and S6). Columns containing these proto-spacers, originally positions 21 and 33, are moved to become the two rightmost columns and clustering is done from the right (this transposition does not affect the results and is done for clarity). Several clear families of viruses are observed in this mosaic. The main family is characterized by closely related mutants, almost all of which share the two rightmost proto-spacers.

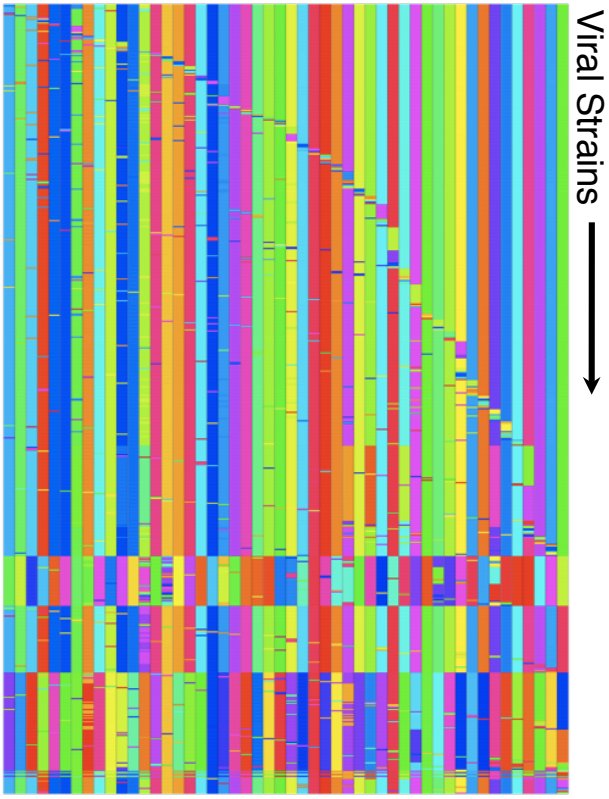


Fig. S8

Figure S8. 'Cloud on cloud' immunity during the predicted viral bloom. The image on the right reflects a clustered heatmap of the virus-host immunity matrix during the predicted viral bloom (768th iteration in Fig. 3B). The rows contain viral strains and the columns show host strains. Each entry of the immunity matrix reflects the number of shared spacers between the respective host and viral strain. Pale yellow color represents no shared spacers (susceptibility), yellow shows one shared spacer, orange two shared spacers, and red three shared spacers. We maximized the silhouette width (SOM text) to cluster both hosts (columns) and viruses (rows) into an optimal number of sub-populations. In that way, virus and host populations are grouped by immunity. Results show that distinct subsets of the host population are immune to distinct viral sub-populations. Importantly, a related group of seven host strains (top center) shares two spacers (orange) with most viruses as predicted in Fig. 3B, which shows matches to the top 300 viral strains. Other host strains survive the bloom due to immunity to less frequent viral strains.

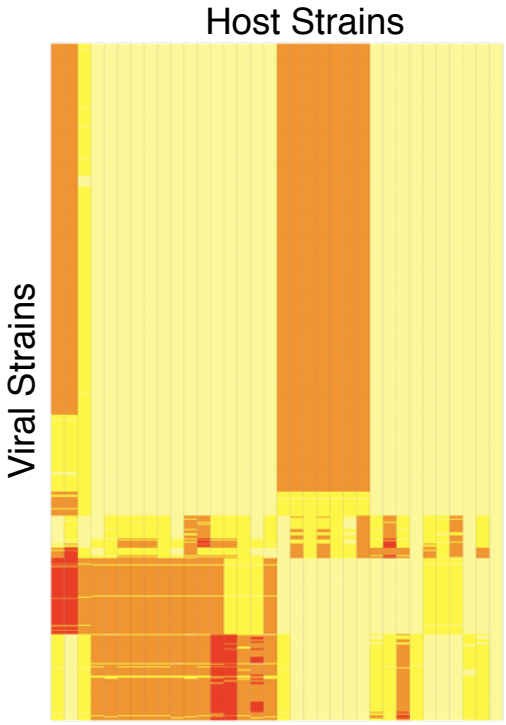
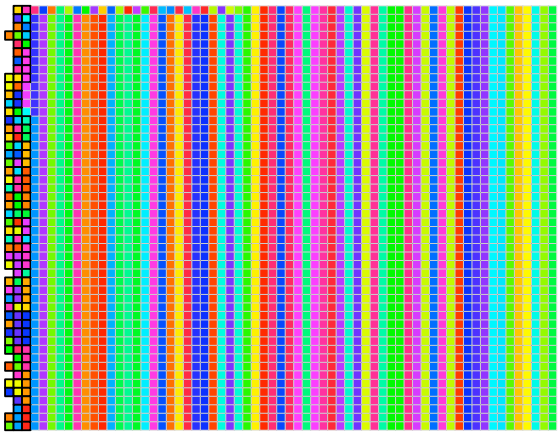
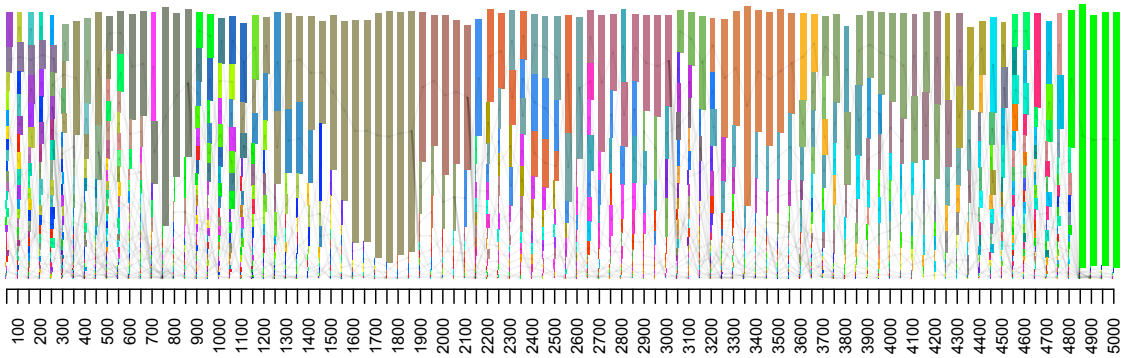


Fig. S9

Fig. S9. Model Simulations With Grace Period Also Show Old-End Clonality and Mutant Sweeps. In the top panel, we plot the top 50 host strains, by frequency, after 5000 model iterations. The remaining parameters are as in Figure 2a, with sufficient host addition to allow for 'kill the winner' dynamics. In the bottom panel, host diversity is tracked every 50 iterations across a simulation as in Figure 2b, using the 'silhouette' technique to choose an optimal number of clusters per iteration. Note that prior to Iteration 4850, a diversity sweep occurs, implying that sweeps are not artifacts due to the grace period preserving new mutants.





References

Althaus, C.L., and De Boer, R.J., 2008. Dynamics of immune escape during HIV/SIV infection. PLoS Comput Biol 4, e1000103.

Altshuler, D., Pollara, V.J., Cowles, C.R., Van Etten, W.J., Baldwin, J., Linton, L., and Lander, E.S., 2000. An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature 407, 513-516.

Baker, B.J., Comolli, L.R., Dick, G.J., Hauser, L.J., Hyatt, D., Dill, B.D., Land, M.L., Verberkmoes, N.C., Hettich, R.L., and Banfield, J.F., 2010. Enigmatic, ultrasmall, uncultivated Archaea. Proc Natl Acad Sci U S A 107, 8806-11.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315, 1709-12.

Barrett, R.D., M'Gonigle, L.K., and Otto, S.P., 2006. The distribution of beneficial mutant effects under strong selection. Genetics 174, 2071-9.

Brockman, W., Alvarez, P., Young, S., Garber, M., Giannoukos, G., Lee, W.L., Russ, C., Lander, E.S., Nusbaum, C., and Jaffe, D.B., 2008. Quality scores and SNP detection in sequencing-by-synthesis systems. Genome Research 18, 763-770.

Denef, V.J., Kalnejais, L.H., Mueller, R.S., Wilmes, P., Baker, B.J., Thomas, B.C., VerBerkmoes, N.C., Hettich, R.L., and Banfield, J.F., 2010. Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. Proc Natl Acad Sci U S A 107, 2383-90.

Desai, M.M., and Fisher, D.S., 2007. Beneficial mutation selection balance and the effect of linkage on positive selection. Genetics 176, 1759-98.

Desai, M.M., Fisher, D.S., and Murray, A.W., 2007. The speed of evolution and maintenance of variation in asexual populations. Curr Biol 17, 385-94.

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S., 2008. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J Bacteriol 190, 1390-400.

Dick, G., Andersson, A., Baker, B., Simmons, S., Thomas, B., Yelton, P., and Banfield, J., 2009. Community-wide analysis of microbial genome sequence signatures. Genome Biology 10, R85.

Drake, J.W., Charlesworth, B., Charlesworth, D., and Crow, J.F., 1998. Rates of spontaneous mutation. Genetics 148, 1667-86.

Eppley, J., Tyson, G., Getz, W., and Banfield, J., 2007. Strainer: software for analysis of population variation in community genomic datasets. BMC Bioinformatics 8, 398.

Ewing, B., and Green, P., 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 8, 186-94.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P., 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8, 175-85.

Gerrish, P.J., and Lenski, R.E., 1998. The fate of competing beneficial mutations in an asexual population. Genetica 102-103, 127-44.

Gillespie, D.T., 2007. Stochastic simulation of chemical kinetics. Annu Rev Phys Chem 58, 35-55.

Gomez-Alvarez, V., Teal, T.K., and Schmidt, T.M., 2009. Systematic artifacts in metagenomes from complex microbial communities. ISME J 3, 1314-1317.

Gordon, D., Abajian, C., and Green, P., 1998. Consed: A Graphical Tool for Sequence,ÄâFinishing. Genome Research 8, 195-202.

He, J., and Deem, M.W., Heterogeneous diversity of spacers within CRISPR (clustered regularly interspaced short palindromic repeats). Phys Rev Lett 105, 128102.

Horvath, P., and Barrangou, R., CRISPR/Cas, the immune system of bacteria and archaea. Science 327, 167-70.

Kimura, M., 1964. Diffusion models in population genetics J. Appl. Probability 1, 177–232

Kimura, M., and Crow, J.F., 1964. The Number of Alleles That Can Be Maintained in a Finite Population. Genetics 49, 725-38.

Kuo, C.H., and Ochman, H., 2009. Deletional bias across the three domains of life. Genome Biol Evol 1, 145-52.

Lo, I., Denef, V.J., VerBerkmoes, N.C., Shah, M.B., Goltsman, D., DiBartolo, G., Tyson, G.W., Allen, E.E., Ram, R.J., Detter, J.C., Richardson, P., Thelen, M.P., Hettich, R.L., and Banfield, J.F., 2007. Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. Nature 446, 537-541.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J.-B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., and Rothberg, J.M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376-380.

Marraffini, L.A., and Sontheimer, E.J., Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature 463, 568-71.

Marraffini, L.A., and Sontheimer, E.J., CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet 11, 181-190.

Nowak, M.A., and May, R.M. Virus Dynamics: Mathematical Principles of Immunology and Virology (Oxford Univ. Press, Oxford, 2000).

Palmer, K.L., and Gilmore, M.S., Multidrug-Resistant Enterococci Lack CRISPR-cas. MBio 1.

Rainey, P.B., and Travisano, M., 1998. Adaptive radiation in a heterogeneous environment. Nature 394, 69-72.

Rodriguez-Brito, B., Li, L., Wegley, L., Furlan, M., Angly, F., Breitbart, M., Buchanan, J., Desnues, C., Dinsdale, E., Edwards, R., Felts, B., Haynes, M., Liu, H., Lipson, D., Mahaffy, J., Martin-Cuadrado, A.B., Mira, A., Nulton, J., Pasic, L., Rayhawk, S., Rodriguez-Mueller, J., Rodriguez-Valera, F., Salamon, P., Srinagesh, S., Thingstad, T.F., Tran, T., Thurber, R.V., Willner, D., Youle, M., and Rohwer, F., Viral and microbial community dynamics in four aquatic environments. ISME J 4, 739-51.

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53-65.

Schleper, C., 2008. Microbial ecology: Metabolism of the deep. Nature 456, 712-4.

Simmons, S.L., Dibartolo, G., Denef, V.J., Goltsman, D.S., Thelen, M.P., and Banfield, J.F., 2008. Population genomic analysis of strain variation in Leptospirillum group II bacteria involved in acid mine drainage formation. PLoS Biol 6, e177.

Tyson, G.W., and Banfield, J.F., 2008. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ Microbiol 10, 200-7.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., and Banfield, J.F., 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428, 37-43.

Wilson, J.B., 1990. Mechanisms Of Species Coexistence: Twelve Explanations For Hutchinson's 'Paradox Of The Plankton': Evidence From New Zealand Plant Communities. New Zealand Journal of Ecology 13, 17-42.