

Trends Microbiol. Author manuscript; available in PMC 2013 February 08.

Published in final edited form as:

Trends Microbiol. 2011 June; 19(6): 286–294. doi:10.1016/j.tim.2011.03.006.

Two-stepping through time: mammals and viruses

Nicholas R. Meyerson and Sara L. Sawyer

Section of Molecular Genetics and Microbiology, Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX, USA, 78712

Abstract

Recent studies have identified ancient virus genomes preserved as fossils within diverse animal genomes. These fossils have led to the revelation that a broad range of mammalian virus families are older and more ubiquitous than previously appreciated. Long-term interactions between viruses and their hosts often develop into genetic arms races, where both parties continually jockey for evolutionary dominance. It is difficult to imagine how mammalian hosts have kept pace in the evolutionary race against rapidly evolving viruses over large expanses of time, given their much slower evolutionary rates. However, recent data has begun to reveal the evolutionary strategy of slowly-evolving hosts. We review these data, and suggest a modified arms race model where the evolutionary possibilities of viruses are relatively constrained. Such a model could allow more accurate forecasting of virus evolution.

"The microscopic and sub-microscopic parasites can evolve so much more rapidly than their hosts that the latter have little chance of evolving complete immunity to them... The most that the average species can achieve is to dodge its minute enemies by constantly producing new genotypes."

-- J.B.S. Haldane, 1949 [1]

Ancient relationships between mammals and their viruses

Recent studies have unearthed a treasure trove of prehistoric virus 'fossils,' viral genomes or genome segments frozen millions of years ago as integrated copies in the genomes of diverse animal hosts (see [2] and references therein). The fact that these integrated viral fossils can be easily recognized as belonging to modern virus families is stunning, given the fact that modern exogenous viruses have replicated and evolved for many millions of years since these viral fossils were captured [2]. Despite high rates of mutation, the evolution of virus sequence is clearly constrained. This constraint comes partially from intrinsic selective forces that limit virus evolution, such as selection for modulation of pathogenicity to the host, and the structural constraints of the virus itself. Other major constraints on virus evolution come from the diverse immune strategies imposed by hosts. Cumulatively, these constraints act together to limit all aspects of virus evolution, from the swarm of variants produced in a single host to the evolution of expanded host range. These newly identified fossils indicate that constraint on virus evolution may be far greater than has previously been appreciated.

Corresponding author: Sawyer, S.L. (saras@mail.utexas.edu).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

^{© 2011} Elsevier Ltd. All rights reserved.

The discovery of these fossils also confirms that mammals have evolved to their current form within a landscape full of diverse viral threats, a situation that has dramatically influenced their own evolution. Although constrained in many ways, viruses still evolve much more quickly than the hosts that they infect. The response time for evolutionary adaptation by viruses can be nearly instantaneous, while mammals reproduce on the scale of years or decades. The dominant force enabling mammals to counteract the extreme genetic diversity of the viruses that they face is generally thought to be the adaptive immune system. Adaptive immunity employs gene rearrangements performed during the lifetime of an individual host, creating a nearly infinite spectrum of receptors and antibodies that evens the playing field between the host and rapidly evolving viral pathogens. But the first line of defense in fighting infection, and one that is thought to be successful against the vast majority of pathogens encountered, is the hard-wired innate immune system. Innate immunity is executed by genes that must function strictly in the form in which they were inherited. The human genome has approximately 1,000 genes dedicated to defense or immunity, most of which cannot diversify in the course of a single host lifetime [3]. Somehow all of these genes must remain honed against their viral targets even though they are trapped in a slowly evolving mammalian genome [4]. In recent years, molecular evidence has emerged, largely from the HIV field, that describes how hosts respond to infection over evolutionary time. While the strategy of viruses is to rapidly adapt to new challenges, the strategy of hosts is, by necessity, altogether different.

Host immune complexity dramatically limits the escape options available to viruses

Genes involved in immunity and defense become effective against their viral targets by natural selection over many generations, with individuals encoding less effective alleles dying from infection with bias [5]. Once effective immunity alleles become common, viruses are expected to counter-evolve, thereby placing selective pressure back on the host species. These dynamics create an ever-escalating genetic 'arms race' between host and virus that results in the rapid evolution of both [6]. Genetic arms races have been shown to play out predominantly through host and virus proteins that directly interact. For instance, host major histocompatibility complex (MHC) genes encode receptors that present peptides at the cell surface for recognition by T cells. MHC class I genes have been acutely selected for alleles that effectively present viral peptides and, as a result, MHC genes are highly divergent between and within species [4]. Viruses, in turn, evolve under selection for mutations that prohibit presentation of viral peptides by these receptors [7]. In the past ten years, an entire landscape of membrane-bound or cytosolic viral sensors have been discovered to act in innate immunity [8–11]. These include a large number of constitutively expressed proteins called 'restriction factors' that recognize viruses and inhibit their replication directly [12]. In primates, these proteins include tetherin/BST-2 and members of the APOBEC3 and tri-partite motif (TRIM) families [13–15]. For example, the TRIM5a protein interacts with the capsid core of retroviruses as they enter the cytoplasm of an infected cell (Figure 1a). Different primate orthologs of TRIM5a have recognition specificity for different retroviral capsids, and infection is only blocked when interaction occurs [16]. Physical interactions such as the one between TRIM5α and capsid are fertile ground for arms race dynamics [17, 18]. In many cases, viruses also encode proteins that antagonize host immunity pathways [8, 14], and these physical interactions can also be subject to arms races [19-21].

Genetic arms races are fundamentally dialogues of call and response between host and viral genomes. However, since arms races unfold over evolutionary time scales, how do hosts actually compete, given the dramatically different evolutionary rates that usually divide mammalian hosts from their viral pathogens? The answer to this conundrum lies largely in

the complex and multi-faceted nature of the immune system. Novel mutations in viral genomes that allow escape from one arm of the immune system will only be viable if they do not make the virus susceptible to other immune strategies. Viral evolution is further limited by interactions with proviral host proteins. As obligate parasites, viruses survive by hijacking host proteins (called cofactors) for processes such as cellular entry and nuclear trafficking (Figure 1a). For instance, cellular entry of HIV requires the human cell surface receptor CD4. A co-receptor is also required for entry, which for most HIV strains is the chemokine receptor CCR5 (Figure 1a)[22]. In large-scale screens, HIV and influenza have each been shown to require several hundred cofactors for replication in humans cells (see [23, 24] and references therein). It should be noted that allelic versions of cofactors that lack compatibility with viruses are just as effective at blocking infection as potent immunity alleles, and possibly more so. For example, some humans encode a variant allele of CCR5, CCR5\(Delta 32\), where a 32 base pair deletion gives rise to a defective receptor that is not expressed on the cell surface [22]. Those individuals homozygous for this allele are almost completely resistant to HIV infection, and even heterozygous genotypes afford some protection due to reduced expression levels of CCR5. Why this allele exists is unknown, but it is highly relevant in a population that is now infected with HIV. In this respect, cofactors can also be engaged in arms race dynamics with viruses [25, 26]. In such cases, the host genome experiences selection to encode cofactors that are non-optimal for viruses, and viruses are selected to efficiently utilize available host cofactors (Figure 1b). New viral variants will only be viable if they retain all necessary cofactor interactions, and simultaneously avoid fatal interactions with immune system proteins.

Selective pressures exerted by the many host immunity proteins and cofactors in combination funnel viruses into a very small mutational space for escape and adaptation. This is supported by many examples where viruses repeatedly escape through the exact same amino acid change, not just through mutation from one amino acid to any other. For example, in experimental cross-species infections of simian immunodeficiency virus (SIV) from sooty mangabeys (SIVsm) into rhesus macaques, viral replication was initially weak due to restrictive TRIM5 alleles [27]. However, viral escape occurred in four different macaques and in all cases involved the same single amino acid change (R97S) in the capsid protein. In another example, there have been multiple independent cross-species transmissions of SIV from chimpanzees (SIVcpz) and possibly gorillas (SIVgor) to humans, giving rise to HIV-1 groups M, N and O. A single amino acid position in the gag-encoded matrix protein underwent the same mutational substitution (M30R) in all of these crossspecies transmissions [28]. When HIV-1 was passaged back through chimpanzees, this mutation reverted. Similar to these two examples from the HIV field, there are many amino acid positions in the influenza genome that have characteristic mutations depending on the source species, suggesting that the web of selective constraints in each species repeatedly results in viruses acquiring signature adaptive mutations [29]. The focused routes of adaptation seen in these examples are a reflection of the massive and multi-factorial constraints imposed by the host. This serves as an important reminder that viral evolution studies are most meaningful when conducted in animal models, as a greater range of viral mutations can be explored in cell culture assays than will actually be viable in the infected host.

Arms races play out in populations of hosts, which slows the rate of viral adaptation

Each step in an evolutionary arms race begins with selection for an advantageous mutation in a population, either host or viral (Figure 2). Only when adaptive mutations go to fixation in either a host or viral population is the arms race permanently moved onward. This is probably a relatively rare event, and the fixation of a potent mutation in a host or viral gene

could theoretically end an arms race forever, driving the other party to extinction. Instead, the Red Queen hypothesis [30] predicts that selective pressures exerted by hosts and viruses upon one another will often result in oscillating allele frequencies in both populations. This is because selective pressure for viruses to counter-evolve will not be strong until a significant number of potential hosts are of a resistant genotype. This provides an important advantage to the host, because it slows the spread of novel, adaptive viral variants through host populations.

The case of CCR5, the HIV cellular co-receptor mentioned above (Figure 1a), illustrates this concept. Most SIV strains also use CCR5 as a co-receptor, but some sooty mangabeys and red-capped mangabeys encode defective alleles of CCR5 [31, 32]. Despite the fact that these alleles are relatively common, the SIV strains that infect these two species (SIVsmm and SIVrcm) replicate perfectly well, even in individuals that are homozygous for defective CCR5. This is because in both cases these viruses have evolved to use additional coreceptors. This is the type of escape that can be expected in the face of common resistance alleles. In humans, the defective $CCR5\Delta 32$ allele is quite common in some parts of the world, with about 18% of Caucasian individuals being heterozygous and 1% being homozygous [22]. Similar to sooty and red-capped mangabeys, a few CCR5Δ32 homozygous humans have also been reported to be infected with HIV, again through mutations of the HIV surface protein that allow use of an additional co-receptor (CXCR4 in this case) [22]. This viral escape through co-receptor switching also happens in many latestage HIV-1 patients who are wild-type for CCR5, and in patients treated with the CCR5inhibiting drug Maraviroc [33]. In both cases this is presumably because the preferred target cells become more scarce or unavailable. Due to the HIV/AIDS epidemic, there continues to be a rise in the frequency of hypomorphic alleles of CCR5 in Africa [34], suggesting that HIV might continue to evolve the ability to use new co-receptors as highly susceptible hosts diminish.

The scenario just described is but one part of a larger picture. In fact, probably hundreds of host loci additively contribute to viral susceptibility. There are many human genes, acting in diverse cellular pathways, in which genetically-encoded polymorphic variants can influence the outcome of viral exposure or infection [35, 36]. Primates encode entire families of several key proteins involved in virus recognition, such as the APOBEC3, TRIM, MHC, killer cell immunoglobulin-like receptor (KIR), and interferon-inducible transmembrane (IFITM) protein families [9, 13, 37, 38]. Further, each of these loci might have many cocirculating alleles in a population; the human *HLA-B* locus of the MHC has over 800 reported alleles [11]. Diploid hosts can carry two different alleles at each locus, each with distinct viral targets. For dominant immunity factors, this means twice the specificity and therefore increased resistance to infection. Heterozygote advantage has been directly demonstrated at MHC loci in both humans and macaques [39, 40]. Different alleles of TRIM5 found in primate populations have also been shown to have different viral specificities, possibly leading to a scenario where balancing selection operates to maintain multiple alleles in populations [27, 41]. With all of these considerations, it is unlikely that any two individuals in any mammalian population have the exact same genetic immunity profile. A key point is that, when viruses evolve to escape resistant host genotypes in the context of a host population, this will provide them access to only a limited number of new hosts due to the mosaic of host genotypes involved. The broad genetic diversity between immune system components of different individuals should dramatically slow the spread of viral escape variants through host populations, and is yet another weapon in the arsenal of the slowly evolving host.

Single point mutations can lead to major adaptations during evolutionary battles

Because host genes engaged in evolutionary arms races are under strong selective pressure to change and adapt, they often evolve at a faster rate than other genes. In fact, immunity genes are some of the most rapidly evolving mammalian genes, acquiring unexpectedly high numbers of non-synonymous nucleotide substitutions [42–44]. The genes of all three of the retroviral restriction factors diagrammed in Figure 1 have accumulated such mutations more rapidly than the bulk of human genes (Figure 3). Importantly, this pattern indicates that single amino acid substitutions in a protein can be adaptive in the context of an arms race [18, 45, 46]. This is critical to the success of the host because simple point mutations are the most common and abundant form of genetic variation upon which natural selection can act.

The power of point mutations can be well demonstrated with examples from the functional characterization of these same three retroviral restriction factors. Single amino acid changes in TRIM5 α can be highly adaptive, as some allow recognition of different mammalian retroviral capsid types [47–51]. Conversely, single amino acid changes in retroviral capsid proteins can also be adaptive by decreasing susceptibility to TRIM5 α [27, 52–55]. Another restriction factor, tetherin, is a cell surface, membrane-bound protein that prohibits budding of retroviruses as well as filoviruses, herpesviruses, and arenaviruses (Figure 1a) [14]. Viruses encode countermeasures to tetherin, including the SIV antagonist Nef and the HIV-1 antagonist Vpu, and single point mutations in tetherin can modulate sensitivity to these viral antagonists [19–21]. The APOBEC3G protein is a restriction factor with activity against a broad range of viruses, but is neutralized by the HIV/SIV accessory protein Vif (Figure 1a) [56]. A single amino acid change in APOBEC3G can make it insensitive to Vif [57–59], and single amino acid changes in Vif can alter specificity for APOBEC3G [60]. Thus, exquisitely small biochemical changes, at the level of a single amino acid, can enhance or reduce affinity between players in an evolutionary arms race.

In some cases, changes more dramatic than point mutations occur in the context of arms races, and these often end one arms race and start another. For instance, in two different primate lineages *TRIM5* has fused with the gene encoding cyclophilin A to produce a TRIM-Cyp restriction factor with a novel retroviral recognition domain, significantly changing the terms of the arms race between TRIM5α and capsid in those species [17]. In humans, tetherin has acquired a deletion in the binding site for its historical SIV antagonist, Nef, freeing tetherin from that arms race until HIV-1 evolved a novel way to neutralize human tetherin using the viral protein Vpu [20, 21, 61]. Adaptive point mutations in the ligand-binding surface of primate CCR5 that influence HIV and SIV binding are not common. Instead, as previously discussed, *CCR5* seems to fight the arms race through protein downregulation at the cell surface, with frame-shifted null alleles having arisen independently in humans and two additional SIV-infected primate species [31, 32, 62]. While highly effective, major adaptive changes such as these are expected to occur less often than adaptive point mutations, simply because they involve rarer genetic events.

Rock-paper-scissors: are arms races cyclic?

Our studies and those by other groups reveal that host restriction factors could be resampling a small number of biochemical forms repeatedly as arms races play out over evolutionary time. For instance, a single amino acid mutation in human TRIM5 α (R332P) can largely restore to this protein the ability to recognize and restrict HIV, and is therefore an important determinant of retroviral specificity [48, 49]. Interestingly, three amino acids, arginine (R), proline (P), and glutamine (Q), have been repeatedly sampled by natural selection at position 332 over tens of millions of years of primate evolution (Figure 4a) [17]. The R-, P-, and Q-

bearing forms of TRIM5a each target different suites of retroviruses [47–50, 52–55]. Amazingly, R, P, and Q are each encoded by different alleles co-circulating in sooty mangabey populations, while P and Q are each encoded by different alleles found in rhesus macaque populations (Figure 4a) [41, 63]. This pattern might reflect ancient polymorphism that has survived all or some of simian primate speciation, rather than recurrent mutation to the same few amino acids [41]. Long-term maintenance of such polymorphism could occur if there were balancing selection acting on these alleles [41, 63]. However, polymorphism has rarely been shown to survive even a single speciation event. Regardless of how this striking pattern arose, continued selection for three specificities must be invoked to explain it.

It is unknown why TRIM5α has resampled three amino acids (R/P/Q) at position 332 over evolutionary time. Only three amino acids at this position would be necessary if there are just three amino acids possible at the cognate position(s) in capsid. If true, this would represent a highly constrained arms race, one that is perhaps consistent with the many constraints detailed above. The consistent cycling over evolutionary time between three amino acids (R/P/Q) is perhaps indicative of a rock-paper-scissors game where the choice of weapons on both sides is limited, and therefore must be recycled (Figure 4b). The evolutionary histories of other retroviral restriction factors, APOBEC3G and tetherin, also reveal resampling of a small set of amino acids at key positions that critically govern viral specificity (Figure 4a). This is interesting, given the vastly different viral repertoires and modes of action of each of these three restriction factors. Small insertions and deletions in the regions of these critical residues have also been observed [18, 20, 64], but these more dramatic mutations may sometimes change the interaction surface in a way that still fits into the cyclic pattern of recycled biochemical forms, or might introduce a new one.

In reality, physical interactions involve three-dimensional protein surfaces, not single amino acid sites, so changes at single sites must be considered in the context of the rest of the protein. This can be demonstrated with examples from TRIM5a biology, where R332 in the context of the human TRIM5a protein does not restrict HIV, but R332 in the context of sooty mangabey TRIM5a does [41, 49, 65]. It is known that other amino acids in the vicinity of position 332 also contribute to target specificity [47-49, 51, 63]. Several codon positions in TRIM5 show elevated rates of non-synonymous substitution and high levels of polymorphism [17], so arms races with retroviruses probably play out at multiple specificity-determining residues simultaneously. If multiple residues on a binding surface each engage in their own rock-paper-scissors chase, this will create a finite number of unique protein surfaces with unique interaction specificities. This has been demonstrated with primate protein kinase R (PKR), where amino acid combinations at three critical residues determine interaction with the poxvirus antagonist K3L [46]. Importantly, this study and others demonstrate that the rules of engagement between two interacting proteins, even at a complex interaction surface, can be determined. While the rock-paper-scissors model is almost certainly oversimplified, it could at least help begin to define constraints on viral evolution as discussed in more detail below.

Our knowledge of how viruses have responded over time to these cycling evolutionary forms of restriction factors remains weak. Alignments of viral genomes are usually limited to samples that have been collected within the last several decades, a scale that precludes analyzing mutational change through an evolutionary arms race. Importantly, the fact that viral genomes dating back tens of millions of years have now been found to be frozen in their historical form in the genomes of slowly-evolving animals might, for the first time, allow us the opportunity to look at viral adaptation over evolutionary scales. Nonetheless, some relevant data does currently exist. Rapid resampling of amino acids at certain residue positions has been observed in the HIV genome. This has been shown to reflect escape from

MHC presentation in one host, and then due to the fitness cost of this escape mutation, reversion in another host [66]. Importantly, this does support the idea that there are cognate viral forms that correspond to different host immunity alleles. It was recently found that TRIM5 α also restricts herpes simplex virus, but that this effect is highly virus strain-specific [67]. This is again consistent with viral populations having cognate polymorphisms that exhibit escape from certain host alleles. Strain-specific interactions with host cells are commonly observed in the influenza field [68]. Studies that utilize diverse clinical and laboratory isolates of a single virus are powerful and can reveal functional polymorphisms within viral populations, some of which could reflect individual adaptations to different host polymorphisms.

To reiterate, the rock-paper-scissors model describes a scenario where viral adaptation to a particular host genotype is highly constrained due to the delicate interplay of thousands of genetic determinants. Because of the limited number of genetic forms available, we have proposed a modified arms race model where both parties (host and virus) recycle a small number of alleles in a rock-paper-scissors chase. It remains unknown whether such dynamics will describe other systems. Thanks to high-throughput techniques, rapid progress is being made in understanding the host genetics of viral infection. New immunity genes and cofactors are still being discovered on a regular basis. The IFITM proteins, which restrict cellular entry of diverse viruses such as dengue, West Nile, SARS, HIV, and influenza A, were discovered only in 2009 [37, 69, 70]. Signatures of positive selection have been documented in various host cofactor and immunity genes, suggesting that arms race dynamics will describe many other host-virus interactions [25, 46, 71, 72]. It remains to be seen how generally amino acid resampling, potentially reflective of rock-paper-scissors dynamics, occurs in these scenarios.

Predicting viral evolution, will it ever be possible?

It is tempting to speculate that the level of constraint observed may make forecasting of virus evolution possible. Important goals are to predict the evolution of viruses as they spread through populations, the transmission of viruses from one species to another, and the development of drug resistance. Due to the difficulty of these problems, the number of papers where prospective prediction of viral evolution has even been attempted remains small [73, 74]. Many interesting evolutionary scenarios fundamentally require viruses to adapt to new host genotypic landscapes. The entire endeavor would be doomed if viral escape were truly unconstrained with no 'rules' to be found. However, the rock-paper-scissors model says that limited opportunities for escape are possible, at least if we assume that only point mutations will be utilized. Major and more rare genetic events utilized for escape, such as recombination or reassortment between viruses of different species, or the acquisition of novel viral genes, gene domains, or deletions within genes, will likely be impossible to predict. This might preclude prediction of evolution in viruses that experience these phenomena frequently, such as influenza.

Ultimately, forecasting viral evolution will require (i) a model of the selective constraints at play in any particular genotypic environment, and (ii) knowledge of how the virus will respond to each of these constraints. Understanding the selective constraints imposed by different host genotypes is not going to be easy (Box 1). This will require a comprehensive list of immunity and cofactor genes relevant to a particular virus, a catalogue of the major alleles of each, and an understanding of which genes are the most potent barriers to, or facilitators of, infection. Next, different host alleles of each gene will each need to be characterized for their viral specificity. Viral escape from restrictive alleles will also need to be characterized, either empirically or using computational models of the protein-protein interaction interfaces between host and viral proteins. Finally, all of this information has to

be combined. Once such models exist, if a new genotypic environment presents to a virus one major genetic incompatibility in a known cofactor or immunity factor, and viral response to that block has been well characterized, prediction of viral evolution will be possible. Multiple genetic blocks to infection in a particular genetic background will require multiple corresponding adaptations of the virus, making prediction of viral evolution more complex but perhaps not impossible.

Box 1

Outstanding questions

- What is the global repertoire of host antiviral and proviral genes that interact with each virus?
- Which host loci present the most potent genetic barriers to cross-species and individual-to-individual virus transmission?
- How many functionally different alleles circulate at each proviral and antiviral locus in host species, and how are they geographically localized?
- Once restrictive cofactor and immunity alleles are known, how do different viral isolates escape them?
- Once relevant host genes have been described, and viral escape to each defined, can models be derived that predict viral evolution in a new host?

Conclusions

Evolutionary thinking in the HIV field has lead to many significant biological insights. For example, any given retrovirus is expected to be very well adapted to its natural host, and able to evade all aspects of that host's immune system. The opportunities for discovering novel mechanisms of resistance are limited in such systems. However, retrovirologists have clearly demonstrated that cross-species infection assays can lead to a rich description of host and virus genetics. By studying heterologous pairings of viruses and hosts, pairs which have not stayed 'in step' through arms race evolution, large genetic phenotypes of both immunity and virulence have been revealed. Such approaches are, at their core, based on an appreciation for the antiquity of mammalian retroviruses and the fact that mammals have coevolved with similar viruses for tens of millions of years. Now that we realize that many other virus families are ancient and ubiquitous in nature, such approaches should be applied in those fields. Reciprocally, such studies are likely to continue to refine the evolutionary theory of host-virus interactions.

Acknowledgments

We wish to thank Ann Demogines, Lauren Ehrlich, Nels Elde, Andy Ellington, Rob Gifford, Patric Jern, Welkin Johnson, Efrem Lim, Dianne Lou, Paul Rowley, Susan Rozmiarek, Sam Scarpino, and Masa Yamashita for insightful comments and discussions. This work was supported by grants (to SLS) R01-GM-093086 from the National Institutes of Health, 003658-0250-2009 from the Norman Hackerman Advanced Research Program, and 107447-45-RGNT from amfAR, The Foundation for AIDS Research. NRM is supported by a National Science Foundation Graduate Research Fellowship. SLS holds a Career Award in the Biomedical Sciences from the Burroughs Wellcome Fund, and is an Alfred P. Sloan Research Fellow in Computational and Evolutionary Molecular Biology.

References

1. Haldane JBS, Disease and Evolution, La ricercha scientifica, 1949: 19:68–76.

- 2. Johnson WE. Endless forms most viral. PLoS Genet. 2010; 6:e1001210. [PubMed: 21124950]
- 3. Lander ES, et al. Initial sequencing and analysis of the human genome. Nature. 2001; 409:860–921. [PubMed: 11237011]
- 4. Trowsdale J, Parham P. Defense strategies and immunity-related genes. Eur J Immunol. 2004; 34:7–17. [PubMed: 14971025]
- Emerman M, Malik HS. Paleovirology- modern consequences of ancient viruses. PLoS Biol. 2010; 8:e1000301. [PubMed: 20161719]
- Dawkins R, Krebs JR. Arms races between and within species. Proc R Soc Lond B. 1979; 205:489–511. [PubMed: 42057]
- 7. Vogel TU, et al. Major histocompatibility complex class I genes in primates: co-evolution with pathogens. Immunol Rev. 1999; 167:327–337. [PubMed: 10319271]
- 8. Bowie AG, Unterholzner L. Viral evasion and subversion of pattern-recognition receptor signalling. Nat Rev Immunol. 2008; 8:911–922. [PubMed: 18989317]
- 9. Parham P, et al. Primate-specific regulation of natural killer cells. J Med Primatol. 2010; 39:194–212. [PubMed: 20618586]
- 10. Liu S, et al. New developments in the induction and antiviral effectors of type I interferon. Curr Opin Immunol. 2011; 23:57–64. [PubMed: 21123041]
- 11. Worobey M, et al. Point, counterpoint: The evolution of pathogenic viruses and their human hosts. Annu Rev Ecol Evol Syst. 2007; 38:515–540.
- Neil S, Bieniasz P. Human immunodeficiency virus, restriction factors, and interferon. J Interf Cytok Res. 2009; 29:569–580.
- 13. Nisole S, et al. TRIM family proteins: retroviral restriction and antiviral defence. Nat Rev Microbiol. 2005; 3:799–808. [PubMed: 16175175]
- 14. Evans DT, et al. BST-2/tetherin: a new component of the innate immune response to enveloped viruses. Trends Microbiol. 2010; 18:388–396. [PubMed: 20688520]
- 15. Goila-Gaur R, Strebel K. HIV-1 Vif, APOBEC, and intrinsic immunity. Retrovirology. 2008; 5
- 16. Emerman M. How TRIM5alpha defends against retroviral invasions. Proc Natl Acad Sci USA. 2006; 103:5249–5250. [PubMed: 16567650]
- 17. Johnson WE, Sawyer SL. Molecular evolution of the antiretroviral TRIM5 gene. Immunogenetics. 2009; 61:163–176. [PubMed: 19238338]
- Sawyer SL, et al. Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. Proc Natl Acad Sci USA. 2005; 102:2832–2837. [PubMed: 15689398]
- 19. Gupta RK, et al. Mutation of a single residue renders human tetherin resistant to HIV-1 Vpumediated depletion. PLoS Pathog. 2009; 5:e1000443. [PubMed: 19461879]
- 20. Lim ES, et al. Ancient adaptive evolution of tetherin shaped the functions of Vpu and Nef in human immunodeficiency virus and primate lentiviruses. J Virol. 2010; 84:7124–7134. [PubMed: 20444900]
- 21. McNatt MW, et al. Species-specific activity of HIV-1 Vpu and positive selection of tetherin transmembrane domain variants. PLoS Pathog. 2009; 5:e1000300. [PubMed: 19214216]
- 22. O'Brien SJ, Moore JP. The effect of genetic variation in chemokines and their receptors on HIV transmission and progression to AIDS. Immunol Rev. 2000; 177:99–111. [PubMed: 11138790]
- 23. Watanabe T, et al. Cellular networks involved in the influenza virus life cycle. Cell Host Microbe. 2010; 7:427–439. [PubMed: 20542247]
- 24. Bushman FD, et al. Host cell factors in HIV replication: meta-analysis of genome-wide studies. PLoS Pathog. 2009; 5:e1000437. [PubMed: 19478882]
- 25. Zhang ZD, et al. Rapid evolution by positive Darwinian selection in T-Cell antigen CD4 in primates. J Mol Evol. 2008; 66:446–456. [PubMed: 18414925]
- 26. Demogines A, et al. Ancient and recent adaptive evolution of primate non-homologous end joining genes. PLoS Genet. 2010; 6:e1001169. [PubMed: 20975951]
- 27. Kirmaier A, et al. TRIM5 suppresses cross-species transmission of a primate immunodeficiency virus and selects for emergence of resistant variants in the new species. PLoS Biol. 2010; 8:e1000462. [PubMed: 20808775]

28. Wain LV, et al. Adaptation of HIV-1 to its human host. Mol Biol Evol. 2007; 24:1853–1860. [PubMed: 17545188]

- 29. Naffakh N, et al. Host restriction of avian influenza viruses at the level of the ribonucleoproteins. Annu Rev Microbiol. 2008; 62:403–424. [PubMed: 18785841]
- 30. Van Valen L. A new evolutionary law. Evol Theor. 1973; 1:1-30.
- 31. Riddick NE, et al. A novel CCR5 mutation common in sooty mangabeys reveals SIVsmm infection of CCR5-null natural hosts and efficient alternative coreceptor use *in vivo*. PLoS Pathog. 2010; 6:e1001064. [PubMed: 20865163]
- 32. Chen ZW, et al. Natural infection of a homozygous Delta 24 CCR5 red-capped mangabey with an R2b-tropic simian immunodeficiency virus. J Exp Med. 1998; 188:2057–2065. [PubMed: 9841919]
- 33. Mosier DE. How HIV changes its tropism: evolution and adaptation? Curr Opin HIV AIDS. 2009; 4:125–130. [PubMed: 19339951]
- 34. Schliekelman P, et al. Natural selection and resistance to HIV. Nature. 2001; 411:545–546. [PubMed: 11385558]
- 35. O'Brien SJ, Nelson GW. Human genes that limit AIDS. Nature Genet. 2004; 36:565–574. [PubMed: 15167933]
- 36. Woolhouse MEJ, et al. Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nature Genet. 2002; 32:569–577. [PubMed: 12457190]
- 37. Brass AL, et al. The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. Cell. 2009; 139:1243–1254. [PubMed: 20064371]
- 38. Sawyer SL, et al. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. PLoS Biol. 2004; 2:e275. [PubMed: 15269786]
- 39. Carrington M, et al. HLA and HIV-1: Heterozygote advantage and B*35-Cw*04 disadvantage. Science. 1999; 283:1748–1752. [PubMed: 10073943]
- 40. O'Connor SL, et al. MHC heterozygote advantage in simian immunodeficiency virus-infected Mauritian cynomolgus macaques. Science Transl Med. 2010; 2:22ra18.
- 41. Newman RM, et al. Balancing selection and the evolution of functional polymorphism in Old World monkey TRIM5 alpha. Proc Natl Acad Sci USA. 2006; 103:19134–19139. [PubMed: 17142324]
- 42. Kosiol C, et al. Patterns of positive selection in six mammalian genomes. PLoS Genet. 2008; 4:e1000144. [PubMed: 18670650]
- 43. Gibbs RA, et al. Evolutionary and biomedical insights from the rhesus macaque genome. Science. 2007; 316:222–234. [PubMed: 17431167]
- 44. Mikkelsen TS, et al. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature. 2005; 437:69–87. [PubMed: 16136131]
- 45. Hughes AL, Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature. 1988; 335:167–170. [PubMed: 3412472]
- 46. Elde NC, et al. Protein kinase R reveals an evolutionary model for defeating viral mimicry. Nature. 2009; 457:485–489. [PubMed: 19043403]
- 47. Li Y, et al. Removal of arginine 332 allows human TRIM5alpha to bind human immunodeficiency virus capsids and to restrict infection. J Virol. 2006; 80:6738–6744. [PubMed: 16809279]
- Stremlau M, et al. Species-specific variation in the B30.2(SPRY) domain of TRIM5alpha determines the potency of human immunodeficiency virus restriction. J Virol. 2005; 79:3139– 3145. [PubMed: 15709033]
- 49. Yap MW, et al. A single amino acid change in the SPRY domain of human Trim5alpha leads to HIV-1 restriction. Curr Biol. 2005; 15:73–78. [PubMed: 15649369]
- 50. Kaiser SM, et al. Restriction of an extinct retrovirus by the human TRIM5alpha antiviral protein. Science. 2007; 316:1756–1758. [PubMed: 17588933]
- 51. Perron MJ, et al. Two surface-exposed elements of the B30.2/SPRY domain as potency determinants of N-tropic murine leukemia virus restriction by human TRIM5 alpha. J Virol. 2006; 80:5631–5636. [PubMed: 16699044]

52. Hatziioannou T, et al. Retrovirus resistance factors Ref1 and Lv1 are species-specific variants of TRIM5alpha. Proc Natl Acad Sci USA. 2004; 101:10774–10779. [PubMed: 15249685]

- 53. Yap MW, et al. TRIM5alpha protein restricts both HIV-1 and murine leukemia virus. Proc Natl Acad Sci USA. 2004; 101:10786–10791. [PubMed: 15249690]
- 54. Perron MJ, et al. TRIM5alpha mediates the postentry block to N-tropic murine leukemia viruses in human cells. Proc Natl Acad Sci USA. 2004; 101:11827–11832. [PubMed: 15280539]
- Keckesova Z, et al. The human and African green monkey TRIM5alpha genes encode Ref1 and Lv1 retroviral restriction factor activities. Proc Natl Acad Sci USA. 2004; 101:10780–10785.
 [PubMed: 15249687]
- 56. Malim MH. APOBEC proteins and intrinsic resistance to HIV-1 infection. Phil Trans R Soc B. 2009; 364:675–687. [PubMed: 19038776]
- 57. Schrofelbauer B, et al. A single amino acid of APOBEC3G controls its species-specific interaction with virion infectivity factor (Vif). Proc Natl Acad Sci USA. 2004; 101:3927–3932. [PubMed: 14978281]
- 58. Bogerd HP, et al. A single amino acid difference in the host APOBEC3G protein controls the primate species specificity of HIV type 1 virion infectivity factor. Proc Natl Acad Sci USA. 2004; 101:3770–3774. [PubMed: 14999100]
- Xu H, et al. A single amino acid substitution in human APOBEC3G antiretroviral enzyme confers resistance to HIV-1 virion infectivity factor-induced depletion. Proc Natl Acad Sci USA. 2004; 101:5652–5657. [PubMed: 15054139]
- 60. Simon V, et al. Natural variation in Vif: Differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. PLoS Pathog. 2005; 1:20–28.
- 61. Sauter D, et al. Tetherin-driven adaptation of Vpu and Nef function and the evolution of pandemic and nonpandemic HIV-1 strains. Cell Host Microbe. 2009; 6:409–421. [PubMed: 19917496]
- 62. Palacios E, et al. Parallel evolution of CCR5-null phenotypes in humans and in a natural host of simian immunodeficiency virus. Curr Biol. 1998; 8:943–946. [PubMed: 9707408]
- 63. Wilson SJ, et al. Rhesus macaque TRIM5 alleles have divergent antiretroviral specificities. J Virol. 2008; 82:7243–7247. [PubMed: 18480454]
- 64. Song B, et al. The B30.2(SPRY) domain of the retroviral restriction factor TRIM5alpha exhibits lineage-specific length and sequence variation in primates. J Virol. 2005; 79:6111–6121. [PubMed: 15857996]
- 65. Ohkura S, et al. All three variable regions of the TRIM5alpha B30.2 domain can contribute to the specificity of retrovirus restriction. J Virol. 2006; 80:8554–8565. [PubMed: 16912305]
- 66. Delport W, et al. Frequent toggling between alternative amino acids Is driven by selection in HIV-1. PLoS Pathog. 2008; 4:e1000242. [PubMed: 19096508]
- 67. Reszka N, et al. Simian TRIM5α proteins reduce replication of herpes simplex virus. Virology. 2010; 398:243–250. [PubMed: 20060996]
- 68. Kuo RL, et al. Influenza A virus strains that circulate in humans differ in the ability of their NS1 proteins to block the activation of IRF3 and interferon-β transcription. Virology. 2010; 408:146–158. [PubMed: 20934196]
- 69. Huang IC, et al. Distinct patterns of IFITM-mediated restriction of filoviruses, SARS coronavirus, and influenza A virus. PLoS Pathog. 2011; 7:e1001258. [PubMed: 21253575]
- Lu J, et al. The IFITM Proteins Inhibit HIV-1 Infection. J Virol. 2011; 85:2126–2137. [PubMed: 21177806]
- Kerns JA, et al. Positive selection and increased antiviral activity associated with the PARPcontaining isoform of human zinc-finger antiviral protein. PLoS Genet. 2008; 4:e21. [PubMed: 18225958]
- 72. Ortiz M, et al. Evolutionary trajectories of primate genes involved in HIV pathogenesis. Mol Biol Evol. 2009; 26:2865–2875. [PubMed: 19726537]
- 73. Bush RM, et al. Predicting the evolution of human influenza A. Science. 1999; 286:1921–1925. [PubMed: 10583948]
- 74. Bloom JD, et al. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. Science. 2010; 328:1272–1275. [PubMed: 20522774]

75. Cao W, et al. Regulation of TLR7/9 responses in plasmacytoid dendritic cells by BST2 and ILT7 receptor interaction. J Exp Med. 2009; 206:1603–1614. [PubMed: 19564354]

76. Tareen SU, Emerman M. Human Trim5α has additional activities that are uncoupled from retroviral capsid recognition. Virology. 2011; 409:113–120. [PubMed: 21035162]

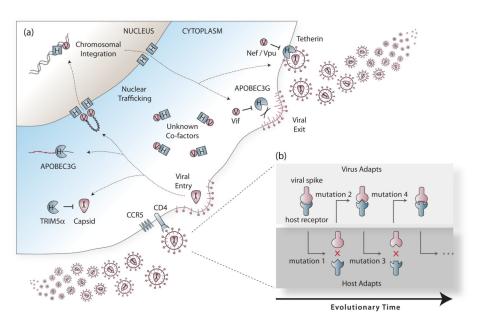


Figure 1. Evolutionary arms races involve interacting host and virus proteins

(a) A schematic of the HIV/SIV lifecycle is shown, illustrating some of the instances where virus proteins (V, red) are known to interact with host proteins (H, blue). In some cases these interactions involve host cofactors (squares and cell surface receptors) that are hijacked by the virus for replication. Hosts also encode antiviral restriction factors (pie shapes) that inhibit viral replication through various mechanisms. For instance, TRIM5a interferes with capsid cores, APOBEC3G hypermutates viral genomes, and tetherin inhibits with virus budding. (b) The cellular receptor, CD4, is used to illustrate a hypothetical arms race scenario. Interaction of the viral spike glycoprotein with CD4 results in virus entry into the cell. Over time, any mutation in CD4 that reduces the strength of this interaction will be preferred by natural selection acting on the host population. Selective pressure will then be placed on the virus population for a mutation in the glycoprotein that re-establishes interaction with CD4. This back-and-forth interplay will result in the rapid fixation of mutations that alter both protein sequences.

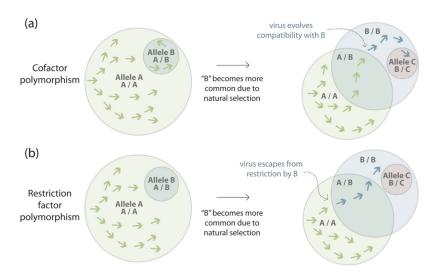


Figure 2. Arms races play out in populations

The circles represent populations of hosts, with the circle sizes representing the relative frequencies of different alleles of a single host gene. The spread of a virus (arrows) through this population of hosts is illustrated. Overlapping regions represent heterozygous hosts and non-overlapping regions represent homozygous hosts. (a) In this schematic, the effect of polymorphism in a cofactor gene is illustrated, where the effect on susceptibility would be predicted to be semi-dominant (similar to the CCR5\Delta 32 mutation discussed in the text). Allele A encodes a cofactor that can be exploited by the virus, while allele B encodes a cofactor variant that is resistant to viral exploitation. The virus will be able to replicate in both A/A and A/B individuals, although presumably less well in A/B heterozygotes. Eventually, allele frequencies in the host population will shift due to the partial protection afforded by the A/B genotype. As allele B becomes more common, there will be selective pressure on the virus to better utilize the B cofactor (blue arrows). A new resistant allele, C, might also emerge or pre-exist in the host population. (b) In this schematic, the effect of polymorphism in a restriction factor gene is illustrated, where allelic versions that restrict a virus are predicted to have a dominant influence on susceptibility to infection. The major allele, A, encodes a restriction factor that is inactive against a circulating virus population. Allele B encodes a restriction factor that is effective against this virus. When allele B is rare, all carriers are assumed to be heterozygous. Because A/B individuals are protected from infection, allele frequencies shift such that allele B is now more common, giving rise to B/B homozygotes. The diminished reservoir of A/A homozygotes places pressure on the virus to 'escape' (blue arrows) restriction by the B restriction factor. A new allele with potency against the virus, C, might also emerge or pre-exist in the host population.

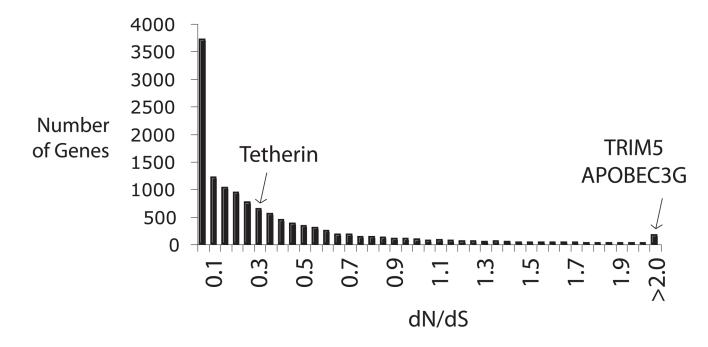


Figure 3. Primate restriction factor genes are rapidly evolving

The distribution shows dN/dS values previously determined for 13,454 human-chimpanzee orthologous gene pairs [44]. Adaptive, gain-of-function mutations commonly arise from point mutations that change an amino acid in the encoded protein. When genes are experiencing sequential rounds of positive selection for new adaptations, as in the arms race scenario, they will retain a higher proportion of non-synonymous mutations (dN) than synonymous mutations (dS) in the domains critical for governing the physical interaction. Domains under positive selection will thus accumulate a characteristic signature of dN/dS > 1. dN/dS values for three primate retroviral restriction factor genes are indicated on this distribution. As expected, restriction factor genes such as TRIM5 and APOBEC3G have some of the highest dN/dS values in the human genome [18, 38]. While it is known that codons within tetherin are evolving under positive selection [19–21], a full length gene analysis reveals a much lower dN/dS value (dN/dS ~ 0.3). Tetherin is either under less intense selective pressure for adaptation than APOBEC3G and TRIM5, or is more constrained by its other cellular roles. Both tetherin and TRIM5a are known to function in host roles other than retroviral restriction [75, 76]. Additional evolutionary constraint comes from the other host proteins with which restriction factors must interact in order to execute restriction [14, 56].

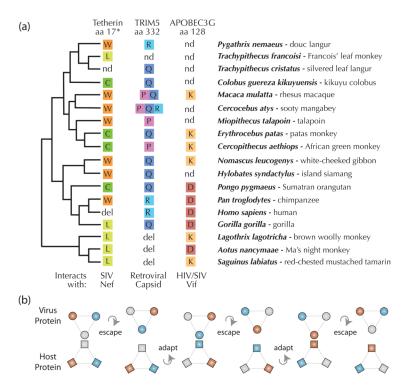


Figure 4. The cyclic arms race: a game of rock-paper-scissors?

(a) Molecular evidence for the rock-paper-scissors model. For each restriction factor, a key residue known to alter specificity for retroviral targets is illustrated, and the amino acid (aa) encoded at that position is reported for a panel of primate sequences available on Genbank. In two cases, SNPs have been noted at these positions, so multiple amino acids are listed. Amino acid coordinates refer to the human protein except in tetherin (*), where amino acid 17 is the chimpanzee coordinate, since amino acids 14–18 are deleted in the human sequence. Abbreviations: nd, not determined; del, deletion; W, tryptophan; L, leucine; C, cysteine; R, arginine; Q, glutamine; P, proline; K, lysine; D, aspartic acid. (b) The illustration shows the physical interaction between a host restriction factor protein (square) and a virus protein (circle). In the rock-paper-scissors model, both the host and the virus have only a small number of biochemical variants with which to compete (3 in this example), because of the many constraints placed on each. These variants are symbolized by different colors. When colors are matched, the host restriction factor successfully inhibits viral infection. However, escape by the virus is also possible. The host population will then respond through selection for an adaptive mutation that re-establishes interaction. The virus might resample previous states (gray circle) that will now be resistant in the context of the new host genotype (orange square).