Human immunodeficiency virus — one of nature's greatest evolutionary machines

Darren Martin[*†] **and Carolyn Williamson**[*]

Transmission of an HIV-like virus from chimpanzees to humans approximately 80 years ago triggered the worldwide AIDS pandemic. Possessing very high mutation and recombination rates, the descendants of this ancestral virus have evolved greatly. Most of this evolution has been in response to selective pressures imposed by human immune responses and has not provided HIV with any significant new biological characteristics. The continuing diversification of HIV variants is a principal obstruction to controlling the virus with drugs and vaccines.

From a human perspective, the human immunodeficiency virus (HIV) story begins around 1930, somewhere in equatorial West Africa, when it is thought that an obscure chimpanzee virus infected a person.[1–3] Unnoticed over the next fifty or so years, this virus's descendants spread among people throughout the world. Even after HIV was discovered in the 1980s, its spread was relentless, despite intense efforts to stop it. Understanding this story, and learning from it, is one of the most difficult and important endeavours that humankind has yet undertaken.

HIV's ancestors

Long before the first HIV infection of humans, biological evolution had already produced in HIV's ancestors the mixture of features that has made the global AIDS pandemic possible. These viruses had the potential to evade our defences, integrate into our genomes, reproduce in our cells and not kill us before exploiting our sexual habits to move between us.

HIV-like viruses have been discovered in more than twenty African primate species belonging to two families (apes and monkeys). Transmission of HIV-like viruses between different primate species has most likely occurred for many centuries.[4] With few exceptions, the HIV-like infections found in their natural primate hosts are not obviously pathogenic. A degree of tolerance (that is, susceptibility without death) may have developed in, for example, chimpanzees if the virus was introduced from a different ape species, spread unabated through the chimpanzee population and killed all susceptible individuals. The small fraction of chimpanzees that survived infection without developing an AIDS-like disease would produce offspring that were themselves tolerant and, in only a few hundred years, infections would no longer be terminal.

If frequent transmission of HIV-like viruses has occurred between the different primate species, a reasonable question to ask is: Why have HIV-like viruses not been transmitted into humans before? It has, in fact, been determined that besides the chimpanzee-to-human transmission incident supposedly around 1930, there have been multiple transmissions of other HIV-like viruses from primates to humans, resulting in two HIV types (HIV-1 and -2) and three HIV-1 groups (HIV-1 group M, group N, and group O).[1,3,5] The HIV-1 group M viruses are responsible for the worldwide AIDS pandemic. Although the group N and O viruses have much in common with the group M viruses, they are either not as deadly or have, for some reason, not spread as successfully.

The virus and its life cycle

Virus particles, otherwise known as virions, are so small they can be visualized only using electron microscopes. Virions are

*Institute for Infectious Diseases and Molecular Medicine, Division of Medical Virology, Faculty of Health Sciences, University of Cape Town, Observatory 7925, South Africa.
†Author for correspondence. E-mail: darren@science.uct.ac.za

merely the suspended-animation phase of virus life cycles — analogous to the seed phase of a flowering plant's life cycle. At this stage of its existence, HIV spends its time being passively transported either between cells within an infected person or from cells in an infected person to cells in an uninfected person. HIV really starts to live only once it attaches to and enters a cell. HIV, as with all viruses, depends on its host's cellular machinery for survival. The interactions of viruses with their hosts is especially intimate: entry of viral particles into host cells involves 'uncoating' — the process by which the protective outer layer(s) of viral particles are stripped away to expose their genomes directly to the intracellular environments of host cells. Viruses in essence become one with host cells, in many cases (including HIV) even merging their genomes with those of their hosts.

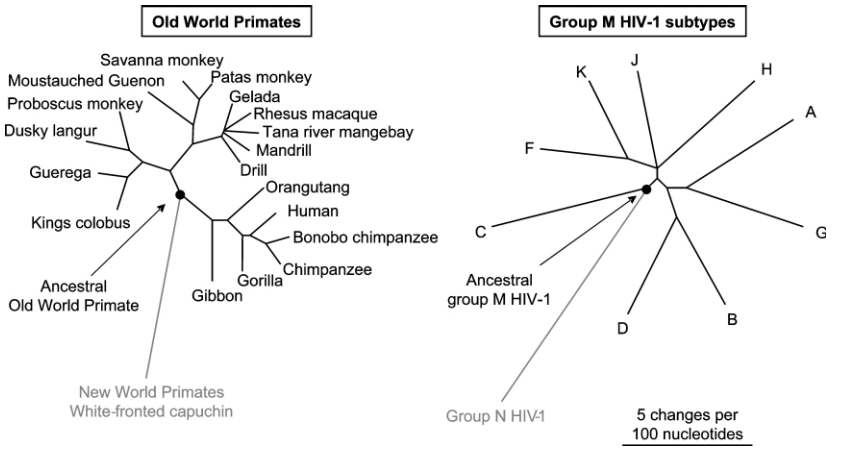

Fig. 1. Comparison of recent primate and group M HIV-1 evolution rates. Evolutionary relationships between primate gammaglobin genes and full-length HIV genomes are represented here in tree form. The tips of the trees' branches represent the various nucleotide sequences used to construct the trees. Names and branches in grey represent New World Primate and HIV-1 group N sequences included for comparison purposes. Dots represent possible tree positions of most recent common ancestors of the Old World primates and group M HIV-1 subtypes. The combined lengths of branches from the point representing the common ancestor to the tips of any one branch are proportional to the number of nucleotide changes that the ancestral sequence underwent during the evolution of the sequence represented by that branch tip. Whereas the primate tree represents 130 million years of evolution, the HIV-1 tree represents 70 years.

Upon infection of a new cell, the HIV's RNA genome must be converted to DNA, the form that integrates into our genome. This process is the reverse of the normal flow of genetic information and is called reverse transcription (hence the name retrovirus, the family of viruses to which HIV belongs). To complete one cycle of reproduction, the DNA must be converted back to RNA, which is then packaged into the viral particle. All organisms on Earth are capable of producing RNA copies of a gene (DNA) using a process called transcription. HIV simply uses the cell's transcription machinery to perform this step of its life cycle. Reverse transcription is, however, not an activity that cells normally perform. HIV must therefore encode its own reverse transcription machinery.

Genetic economy: compact and complex genomes

There is great selection pressure for viruses to be as small as possible. Smaller viruses reproduce faster than larger ones and more of them can fit within an infected cell. The size of any viral genome is close to the absolute minimum number of nucleotides required to encode all the things that viruses need to survive. In stark contrast to our genome, where the overwhelming majority of nucleotides are apparently superfluous, nearly every nucleotide of HIV's genome has a purpose.

Virus genomes generally contain only those genes required for viral reproduction and movement (both from cell to cell within a host, and between hosts). However, viruses do not encode all the proteins they need to complete their life cycles — where possible, they use proteins encoded by their hosts. Virus genomes are thus a greatly compressed set of instructions required to coerce their hosts into performing tasks that are necessary for the virus's survival.

The complexity of these instructions and the amount of encoding they require reflect both the complexity of a virus species' life cycle and the harshness of the niche it has evolved to fill. Whereas HIV has an approximate genome size of ~10 000 nucleotides, the simplest mammalian viruses have genomes of only ~2000 nucleotides. HIV is five times larger than the smallest mammalian viruses because (1) its mode of reproduction requires that it encode some enzymes and proteins that have no suitable human substitutes, and (2) it needs to survive indefinitely our potent immune systems to ensure its eventual transmission to new hosts. As a consequence of immune pressure, the virus has devoted a large proportion of its genome to producing proteins that enable the virus to both evade and suppress host immune systems.

HIV evolution

HIV has a very high mutation rate, which means that it has the potential to evolve quickly. Just how fast HIV evolves can be appreciated by comparing its rate of evolution to that of humans. However, before making this comparison, it is necessary first to consider how to measure evolution rates.

Determination of evolution rates is complicated as they cannot, in most cases, be directly measured and are influenced by a multitude of different factors, mutation rate being only one of them. In addition, the relative importance of these different factors may change drastically during the long evolutionary history of an organism. However, the accurate determination of human and HIV evolution rates are unnecessary for us to get a reasonably good impression of just how different the rates of human and HIV evolution are.

It is possible to use the sequence of nucleotides in the genes, or even the entire genomes of a group of related organisms, to produce an evolutionary tree that describes how the organisms are related to one another. In such a tree, it is important to note, first, that there is a point along one of the branches that represents the common ancestor of all the sequences represented by the tree, and, second, that the tips of branches represent the sequences used to construct the tree. Given this, the combined lengths of the tree's branches from the point representing the common ancestor to the tips of any particular branch are proportional to the number of nucleotide changes that the ancestral sequence underwent during the evolution of the sequence represented by that branch's tip. The evolutionary trees for Old World primates (including humans) and HIV are presented in Fig. 1. You will notice that the human tree is slightly smaller than the HIV tree. What is so startling about these trees is that, whereas the human tree represents 130 million years of evolution, the HIV tree represents only 70 years.

There are a number of reasons why HIV's rate of evolution is so much higher than that of humans. Because mutations usually occur during reproduction, the more frequently an organism reproduces, the higher its mutation rate will be. Whereas the generation times of the apes represented in Fig. 1 are at least a

few years, with any individual having approximately two offspring, the generation time of HIV is approximately 60 hours, with each individual genome producing thousands of offspring.[6]

Far more significant than the difference in HIV and human generation times, however, is the accuracy with which their respective genome-copying machineries operate. Whereas the human copying machinery makes only one error per 10 billion nucleotides copied, the copying machinery of HIV can make up to one error in every 2000 nucleotides copied.[7–9] Unlike the DNA-to-DNA copying enzymes (called DNA polymerases) that are used during the reproduction of most organisms, the RNA-to-DNA copying enzymes (reverse transcriptases) that are found in retroviruses such as HIV can neither 'proofread' their copying efforts nor correct any of the mistakes that they make. The HIV reverse transcriptase is, however, also particularly error-prone, so that even when compared with other retroviruses, HIV has an astonishingly high mutation rate.

HIV-1: the super-mutator

In certain instances, HIV's mutation rate can be so high that nearly every new genome that is produced will contain one or more copying errors. As a consequence, within an infected individual it would be very difficult to find any two viruses that were identical. This assemblage of unique individual mutants is sometimes called a quasi-species. The diversity of different HIV genomes found within a single AIDS patient can far exceed the diversity in entire species of higher organisms (including humans). By infecting individuals with a diverse swarm of mutant viruses, HIV has unique biological and evolutionary advantages over disease-causing organisms such as bacteria and many other viruses that produce infections with genetically uniform populations. Among these advantages is the ability to adapt rapidly to hostile host immune systems and antiretroviral drugs.

The extent of HIV diversity challenges our concept of what constitutes a species. While the diverse population of viruses found within an infected person might be similar in genetic depth to, for example, mammalian species such as humans or lions, the substantially more diverse range of HIVs found worldwide is similar in genetic depth to entire groups of species such as, for example, the primates or cats (Fig. 1). In effect, HIV is not a single species and, as a result, HIVs found worldwide have been split into nine groupings called subtypes (named A through K, excluding I and E). HIV is evolving so fast that within a few decades it will be necessary to split and re-split the existing subtype groups just to keep track of the new viruses that will evolve.

HIV's extremely high mutation rate does, however, present it with some serious problems: (1) a large proportion of mutations will be harmful to the virus, (2) potentially useful mutations may have no immediate value and might be beneficial only in certain circumstances, and (3) useful mutations will most likely occur in genomes that contain other mutations that are harmful. The discovery that HIV had a mutation rate that could result in copying errors in nearly every new genome produced was puzzling. It has long been known that there is a maximum mutation rate beyond which no species can survive. If the mutation rate of a species were to exceed this maximum (approximately one error per reproductive cycle), the accumulation of harmful mutations with every reproduction cycle would result in a progressive and irreversible loss of genetic vitality that would drive the species to extinction. The mutation rate of HIV was believed to be over the theoretical maximum and it was surprising, therefore, that its genome was not experiencing a so-called mutational meltdown.

It has, however, been discovered that certain features of HIV's biology allow it to push the permissible copying error threshold. The mutation rate of HIV is not solely dependent on how often its genome copying machinery makes mistakes. During an infection, HIV produces proteins that influence its mutation rate through interaction with host enzymes that both increase[10] and decrease[11] the prevalence of genome copying errors. Estimates of actual mutation rates (as opposed to genome copying errors) have been as low as one mutation per 30 000 nucleotides copied — that is, one mutation in every three replication cycles.[7] The involvement of multiple host and virus factors means that the actual HIV mutation rate most likely differs from one infection to another.

HIV is able to escape the progressive accumulation of whatever harmful mutations do occur through a process called recombination. Recombination refers to the exchange of genetic material between two or more genomes within the same cell. It provides the virus with a very powerful mechanism by which beneficial mutations can be uncoupled from harmful mutations. Recombination occurs during the RNA-to-DNA copying step of reproduction, when the reverse transcriptase enzyme will use two different HIV RNA genomes to produce a single hybrid DNA genome copy. The recombination rate in HIV is extraordinarily high and exceeds the mutation rate.[12] The two features of HIV's biology that encourage recombinational rearrangement of mutations are, first, the packaging of two separate genomes into every virus particle and, second, the frequent infection of individual cells with more than one virus particle.

HIV-1: the super-evolver?

Given that HIV seems to be one of nature's greatest mutators, it is perhaps interesting that the virus has not evolved any really significant new characteristics since it first began infecting apes and humans. With the exception of one or two genes, the genome structure and overall biology of HIV is still remarkably similar to HIV-like viruses found in apes and monkeys. For example, none of the HIV-like viruses has yet evolved an aerosol or insect transmission strategy. The reliance of HIV and HIV-like viruses on transmission by direct blood-to-blood contact is notably inefficient and presents the virus with the difficult and dangerous need to persist within its hosts for long periods. The phrase 'soap opera evolution' has been used to describe HIV's apparent lack of evolutionary progression — much happens but it never seems to go anywhere.

Evolution is, however, all about strategy. HIV's strategy is not to be a super-evolver but rather a super-mutator, able to survive for long periods within humans. Frequent mutation ensures the virus remains constantly one step ahead of our immune systems. To combat HIV, our immune systems are absolutely reliant on the identification of HIV and HIV-infected cells through the detection of physical features found only in HIV proteins. Changing these physical features whenever they are detected is key to the virus's success. There is an immediate evolutionary advantage for any virus variant that avoids an immune response because it does not possess a targeted physical feature. Such variants will thrive, while others possessing the feature are destroyed.

HIV has evolved a complex external structure of inaccessible pits, hidden grooves and elaborate masking structures that act as camouflage, effectively enabling the virus to hide from our antibody-based immune defences.[13] A high mutation rate allows the virus constantly to change this complex external shield, thus

effectively enabling continued evasion. Large portions of all HIV proteins, including many vital working parts, are vulnerable to detection by our cellular-based immune defences and there is also much pressure for these parts to change. HIV is able continually to escape cellular immune defences mediated by, for example, cytotoxic T lymphocytes, through mutations that alter parts of the virus recognized by these immune cells. Even though many of these immune escape mutations slightly compromise the functionality of the virus's working parts the need for immune escape is absolute and damaged escapees easily out-compete undamaged viruses that are unable to escape. None (or at best very few) of the immune evasion mutations that accumulate during the course of a single infection will have any use when a virus is transmitted to a new host. Therefore, whereas HIV's mutation strategy ensures its survival in the short term, in the long term it creates what is in effect useless evolutionary 'noise'.

The consequences of HIV diversity

Ironically, it is precisely this evolutionary noise that presents the greatest obstacle to both HIV vaccine development and drug treatment. Vaccines designed against one subtype may not protect against another subtype. In fact, primate studies have shown that some experimental vaccines will work only if the viruses used to construct the vaccines are identical to those with which vaccinated animals are subsequently infected.[14] Clearly, it is not possible to make vaccines matched to all existing HIV subtypes and sub-subtypes, let alone vaccines perfectly matched to every HIV variant that exists — the number of variants is countless and is relentlessly growing with every passing hour.

Current vaccine strategies have tried to overcome this problem by: (1) focusing on regions of the genome that are conserved between subtypes; (2) making use of pieces of all the least variable parts of the virus that are detectable by our immune systems; (3) combining the variable parts of different subtypes; (4) using theoretical ancestral and consensus proteins, which are more closely related to the proteins of circulating viruses than the circulating viral proteins are to one another.[15,16] These vaccines are based on the premise that, despite HIV's extreme diversity, there are immune-detectable regions of its genome that cannot significantly change without compromising their functionality. However, the amount of viral diversity that could be controlled by an effective vaccine remains unknown.

Also of concern is that a first-generation vaccine may not completely block infection, and 'breakthrough infection' of vaccinated individuals may result in the eventual emergence and spread of vaccine escape mutants. Such a scenario is evident in regions of the world where drug therapy is common, where as many as 10% of transmitted viruses are resistant to at least one antiretroviral drug.

Although the human genome is the focus of more intense study, the HIV genome is the best-understood of any on Earth. With an approximate size of 10 000 nucleotides, HIV's genome is nearly 300 000 times smaller than the human genome. It is small enough that over 400 full-length genomes have been sequenced, the function of each of its genes is known, and the three-dimensional structures of most of the proteins encoded by these genes have been solved. Despite having a better idea about HIV's inner workings than we do for any other organism, we still have no really fundamental understanding of how the virus works. For example, we are still unable to use our accumulated knowledge to construct a convincing computational model of what happens within an HIV-infected human cell, much less accurately predict the course that an infection will take in a human body. We fortunately do not need to have such an intimate understanding of how HIV works to beat it – we simply need to know enough about it to produce drugs that interfere with its working parts and vaccines that improve our natural defences against it.

1. Gao F., Bailes E., Robertson D.L., Chen Y., Rodenburg C.M., Michael S.F., Cummins L.B., Arthur L.O., Peeters M., Shaw P.M. and Hahn B.H. (1999). Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* **397**, 436–441.
2. Korber B., Muldoon M., Theiler J., Gao F., Gupta R., Lapedes A., Hahn B.H., Wolinsky S. and Bhattacharya T. (2000). Timing the ancestor of the HIV-1 pandemic strains. *Science* **288**, 1789–1796.
3. Sharp P.M., Bailes E., Chaudhuri R.R., Rodenburg C.M., Santiago M.O. and Hahn B.H. (2001). The origins of acquired immune deficiency syndrome viruses: where and when. *Phil. Trans. R. Soc.* **356**, 867–876.
4. Sharp P.M., Gao F., Beer B.E., Hirsch V.M. and Hahn B.H. (2000). Origins and evolution of AIDS viruses: estimating the timescale. *Biochem. Soc. Trans.* **28**, 275–282.
5. Hahn B.H., Shaw G.M., de Cock K.M. and Sharp P.M. (2000). AIDS as a zoonosis: scientific and public health implications. *Science* **287**, 607–614.
6. 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**, 1582–1586.
7. 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**, 5087–5094.
8. Preston B.D., Poiesz B.J. and Loeb L.A. (1988). Fidelity of HIV-1 reverse transcriptase. *Science* **242**, 1169–1171.
9. Roberts J.D., Bebenek K. and Kunkel T.A. (1988). The accuracy of reverse transcriptase from HIV-1. *Science* **242**, 1171-1173.
10. Mangeat B., Turelli P., Caron G., Friedli M., Perrin L. and Trono D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* **424**, 99-103.
11. Mansky L.M., Le Rouzic E., Benichou S. and Gajary L.C. (2003). Influence of reverse transcriptase variants, drugs, and Vpr on human immunodeficiency virus type 1 mutant frequencies. *J Virol.* **77**, 2071-2080.
12. Zhuang J., Jetzt A.E., Sun G., Yu H., Klarmann G., Ron Y., Preston B.D. and Dougherty J.P. (2002). Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J. Virol.* **76**, 11273–11282.
13. Piguet V. and Trono D. (2001). Living in oblivion: HIV immune evasion. *Seminars in Immunology* **13**, 51–57.
14. Barouch D.H., Kuntsman J., Kurada M.J., Schmitz J.E., Santra S., Peyerl F.W., Krivulka G.R., Lifton M.A., Gorgon D.A., Montefiori D.C., Lewis M.G., Wolinsky S.M. and Letvin N.L. (2002). Eventual AIDS vaccine failure in a rhesus monkey by viral escape from cytotoxic T lymphocytes. *Nature* **415**, 272–273.
15. Klein M. (2003). Prospects and challenges for prophylactic and therapeutic HIV vaccines. *Vaccine* **21**, 616–619.
16. Amara R.R. and Robinson H.L. (2002). A new generation of HIV vaccines. *Trends Mol. Med.* **8**, 489–95.