NATIONAL CENTER FOR CASE STUDY TEACHING IN SCIENCE

Phylogenetic Forensics, HIV, and a Dentist



by
Elizabeth R. Everman
School of Biological Sciences
University of Oklahoma, Norman, OK

Introduction

In 1981, cases of a new, deadly disease marked the beginning of a medical mystery that resulted in widespread fear, intense social stigma, and discrimination. This period of time became known as the "AIDs crisis." Despite decades of scientific debunking of misconceptions and development of antiviral therapies that are highly effective at reducing transmission risk, stigma and discrimination continue to persist.

HIV is a sensitive topic and may bring up intense feelings in yourself and your colleagues. During your engagement with this case study, please remember to be respectful. Our goal is to approach this case from an objective, scientific perspective and to consider how genetic evidence can be used to solve disease transmission mysteries.

Begin your study by reading the following article from *The New York Times:*

- Altman, L.K. (1993, June 6). AIDS and a dentist's secrets. <https://www.nytimes.com/1993/06/06/weekinreview/aids-and-a-dentist-s-secrets.html>

Your instructor will guide you through the rest of this case study with a slide presentation.

Case copyright held by the **National Science Teaching Association (NSTA)**. Originally published November 2, 2026. Please see our **usage guidelines**, which outline our policy concerning permissible reproduction of this work. *Credit:* Licensed image © Nikolay Antonov | Dreamstime.com, ID 145761332.

Part I – DNA Replication and Genetic Variation

The molecule responsible for shaping the structure and function of all organisms has four critically important characteristics: it is able to *store information,* accurately *replicate*, be *transmitted*, and *vary* among individuals and species. The processes that influence each of these attributes help us understand things like why calico cats are usually female and why human men are more often colorblind, but they can also be applied in forensic analysis. Consider Table 1 below. Explain the four major characteristics of DNA, and then look up the forensic application.

Question

1. Recall the four major characteristics of DNA. How can these characteristics be used in forensics?

Characteristic	*Describe an Example*	*Uses in Forensics*
Information Storage		Forensic DNA intelligence:
Transmission		Paternity testing:
Replication		
Variation		Phylogenetic forensics:

This case study will focus on phylogenetic forensics, a branch of evolutionary forensics that traces the movement and mutation of DNA sequences from one person (or organism) to the next. One of the most famous applications of phylogenetic forensics is in tracing the transmission of HIV (human immunodeficiency virus), a virus which if left untreated can progress to AIDs (Verma et al., 2020; Abecasis et al., 2018). HIV is a rapidly evolving virus that mutates essentially as soon as it infects a new individual (Abecasis et al., 2018). We will see that these mutations tend to follow predictable frequencies, allowing investigators to establish and study a transmission network that can be used to find "patient zero." In particular, we will consider the case of "the Florida dentist," Dr. David Acer, who infected his patients with HIV in the late 1980s.

Before we can embark on our forensic analysis of this case, we need more information on how variation arises, persists, and spreads. If you've read this far, you probably have an understanding of the basic building blocks of genetic material and the molecular process that allows DNA to be passed from one generation to the next. Although DNA replication is semiconservative and a template is used to ensure that an exact copy is made, mistakes can still occur.

Question

2. What aspects of DNA structure and replication increase the fidelity or correctness of newly replicated DNA sequences?

Despite the ability of DNA polymerase (DNA pol III in prokaryotes; DNA pol ε and δ in eukaryotes) to remove incorrect bases, the wrong base is still occasionally incorporated in the replicating DNA strand. Approximately one out of every 100,000 nucleotides added is incorrect (Pray, 2008). Genomes (the entire collection of nuclear genetic material) are quite large. Human genomes have more than six billion base pairs. Without proofreading, each time a cell divides would result in ~120,000 molecular typos (Pray, 2008)!

Proofreading allows 99% of these incorrect nucleotides to be removed. When an incorrect base is added to the growing chain, the 3′ –OH group is slightly misaligned, which causes DNA pol to stall. While the polymerase is paused, the stability of the new nucleotide is tested by the polymerase. If hydrogen bonds are not formed properly and the base is misaligned, the 3′ to 5′ exonuclease activity of DNA pol III removes the base (Reha-Krantz, 2010). A second process called mismatch repair involves enzymes that can recognize deformities in the chromatin structure of DNA (divots and bulges in the DNA double helix) (Pray, 2008). Places where the double helix deviates from 2 nm diameter or where the 30 nm chromatin fibers are distorted are corrected by these mismatch repair enzymes that excise (cut out) and replace incorrect nucleotides.

Questions

3. What are the consequences of DNA replication mistakes that are not fixed by proofreading or mismatch repair?

4. Some types of mutations occur more frequently than others. In general, transitions are more likely to occur than transversions. Define transversion and transition. Which type is more likely to be missed by proofreading?

   Transition:

   Transversion:

5. The starting sequence below (Figure 1) has been mutated four different ways to illustrate transition and transversion mutations. Identify which mutation has occurred at each substitution.

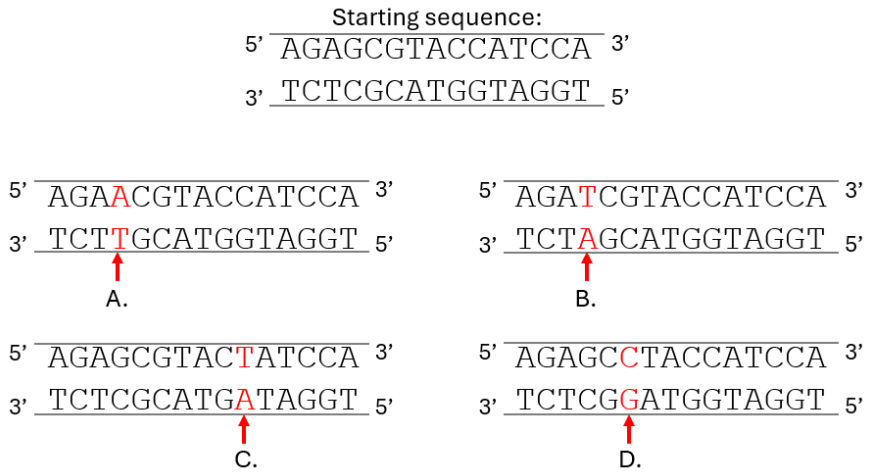


Figure 1. Transition and transversion mutations.

A:

B:

C:

D:

Transitions are about twice as likely to be observed (Azi et al., 2022) because the substitution involves a change between bases that have a similar structure or overall size. They are less likely to cause the 3′-OH to be misaligned, less likely to cause DNA pol to stall, and less likely to distort the double helix. All of these smaller changes contribute to transitions being more likely to be missed by proofreading and mismatch repair mechanisms. As a result, transition mutations occur more frequently than transversion mutations.

Part II – Tracing the Transmission of HIV Using DNA Replication Errors

Although transitions are more likely to escape proofreading in general, the relative proportions of each type of substitution can vary across the genome and between species. In fact, some of these departures from expected frequencies are so unusual, they can be used as diagnostic tools in forensic analysis.

In the case of the virus that causes HIV, we know that A to G transitions are the most common type of mutation, followed by A to C transversions (Figure 2). The basic prediction of transition and transversion rates, as well as the deviations from those predictions result in patterns that can be used to track the transmission of viruses that have high mutation rates, such as HIV.

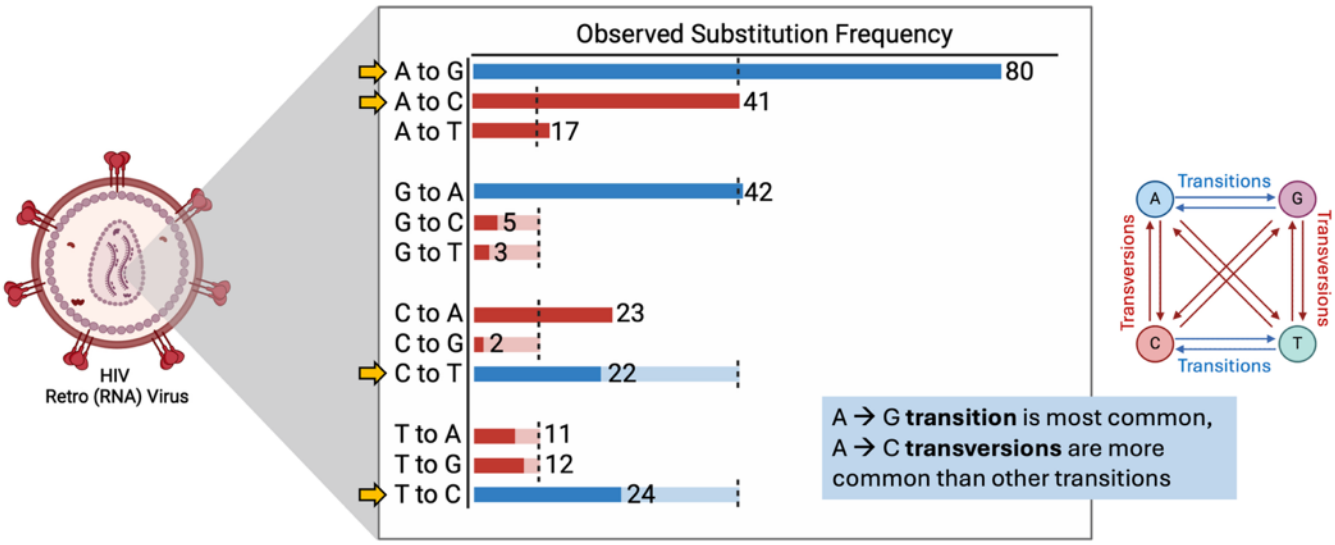


Figure 2. Mutation frequency in HIV has a distinct pattern that deviates from the simple prediction that transitions are always more common than transversions. The base substitution is shown on the left and the expected number of each type of substitution is indicated with the dashed lines (out of 282 substitutions). The red bars indicate transversions and the blue bars indicate transitions. The transition from A to G occurred an average of 80 times in a comparison across many samples, which is about twice as frequently as expected. A to C transversions are nearly three times more common than expected and occur more frequently than C to T transitions (which are about half as common as expected). Adapted from Table 1 in Hillis et al. (1994).

In addition to having a unique pattern of mutations, most organisms and viruses accumulate mutations gradually. This means that with every new infection, one or a few new mutations might occur. This gradual accumulation of mutations that each have a predicted probability of occurring (Figure 2) allows geneticists to track the transmission of a pathogen. Ultimately, this allows the reconstruction of a phylogeny or a genetic tree for the HIV strains that can be used to predict the order of and likelihood of infection.

So why does this departure from predicted mutation patterns matter? To prepare for this case study, you read an article that reported on the dentist Dr. David Acer who infected at least five of his patients with HIV in the late 1980s (Altman, 1993). There were no instances of overt intentional exposure (assault or consensual interactions), so how these patients were infected is still a mystery.

Question

6. In the United States, the transmission of HIV from doctor to patient is extremely rare; in fact, Dr. Acer seems to be the only healthcare professional that has infected his patients without clear intention of doing so (Danila et al., 1991). How can we be certain the HIV strains that infected his patients originated from him? What information do we need?

Part III – Solving the Transmission Mystery

The first step in figuring out whether Dr. Acer infected his patients using sequence similarity and mutations was isolating and amplifying an informative region of the HIV genome. Researchers investigating the case focused on the gene *gp120* (glycoprotein 120), which codes for a protein that makes up the envelope of the virus. This protein plays an important role; without a functional copy of this gene, the virus cannot infect new host cells. As a result, this gene mutates more slowly (about three times lower) than a non-coding region in the virus genome because any mutations that break the *gp120* gene are lost (because the virus fails to infect and replicate) (Geller, 2015). Only mutations that do not break the gene function are retained and can be used to trace the transmission of the virus.

When DNA is extracted from human samples (or any other organism), the sample contains DNA from the subject as well as genetic material from bacteria and viruses carried by the individual. Comparisons of specific sequences from multiple samples requires that the non-target DNA be either diluted or that the target sequence be amplified.

PCR (polymerase chain reaction) allows for amplification of the target sequence by taking advantage of what we have learned about DNA replication in the cell. PCR requires only a subset of the enzymes and molecular components that are required in a cell.

Questions

7. Fill out the table below to summarize what components are needed for DNA replication in a cell versus PCR. Indicate whether the component is required for DNA replication in a cell or PCR (or both).

Component	*Description or Function*	*Required for DNA replication in a cell?*	*Required for PCR?*
Helicase			
Single-strand binding proteins			
Primase			
RNA primers			
DNA primers			
DNA polymerase			
Ligase			
Okazaki Fragments			
dNTPs			
Template DNA			

The key innovation that allows PCR to target specific sequences for amplification is the development of DNA primers that are complementary to the sequences surrounding the target. In a cell, RNA primers provide starting points that allow the entire genome to be copied. In PCR, only the region between the primers is copied. Primer design involves many considerations, such as the number of G and C bases included in the primers, whether the primer has any off-target binding, and the temperature that should be used for annealing during the PCR reaction. These considerations are beyond the scope of this case study, so we will focus on the base pairing rules alone. One tool that is especially useful for primer design is NCBI BLAST Primer Design (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi).

Questions

8. The sequence below encodes a section of the *gp120* gene (sequence accession AF236859 from GenBank). The 5′ to 3′ strand is shown in bold; the 3′ to 5′ strand is shown in regular font. Where will the forward and reverse primers bind? (Hint: look for a match to the forward primer on the 5′-3′ strand and the reverse primer on the 3′-5′ strand).

Forward primer: 5′ – ACCCCAACCCGCAAGAAATA – 3′

Reverse primer: 5′ – TCCTGAGGGTGGTTGAAAGC – 3′

5′ - TGTGTGGAGAGATGCAGATACCACCCTATTTTGTGCATCAGATGCCAAATCACATGTGACAGAAG
3′ – ACACACCTCTCTACGTCTATGGTGGGATAAAACACGTAGTCTACGGTTTAGTGTACACTGTCTTC

CACACAATGTCTGGGCCACACATGCCTGTGTACCCACAGACCCCAACCCGCAAGAAATACACCTGGAAAA
GTGTGTTACAGACCCGGTGTGTACGGACACATGGGTGTCTGGGGTTGGGCGTTCTTTATGTGGACCTTTT

TGTAACAGAAAATTTTAACATGTGGAAAAATAACATGGTAGAGCAGATGCAGGAGGATGTAATCAGTTTA
ACATTGTCTTTTAAAATTGTACACCTTTTTATTGTACCATCTCGTCTACGTCCTCCTACATTAGTCAAAT

TGGGAGCAAAGTCTAAAGCCATGTGTAAAGTTAACTCCTCTCTGCGTTACTTTAAATTGTACCAATGCTA
ACCCTCGTTTCAGATTTCGGTACACATTTCAATTGAGGAGAGACGCAATGAAATTTAACATGGTTACGAT

ATTTGACCAATGCTAATTTGACCAATGCTAATAACATAACCAATGTAGAAAATATAACAGATGAGGTAAG
TAAACTGGTTACGATTAAACTGGTTACGATTATTGTATTGGTTACATCTTTTATATTGTCTACTCCATTC

TAATCAGATCTGAAAATCTCACAAACAATGTCAAAACCATAATAGTGCACCTTAATAAATCTGTAGAAAT
ATTAGTCTAGACTTTTAGAGTGTTTGTTACAGTTTTGGTATTATCACGTGGAATTATTTAGACATCTTTA

CAATTGTACCAGACCCTCCAACAATACAAGAACCAGTATAACTATAGGACCAGGACAAGTATTCTATAGA
GTTAACATGGTCTGGGAGGTTGTTATGTTCTTGGTCATATTGATATCCTGGTCCTGTTCATAAGATATCT

ACAGGAGACATAATAGGAGATATAAGGAAAGTATCTTGTGAGCTTAACGGAACAAAATGGAATGAAGTTT
TGTCCTCTGTATTATCCTCTATATTCCTTTCATAGAACACTCGAATTGCCTTGTTTTACCTTACTTCAAA

TAAAACAGGTAAAGGAAAAATTAAAAGAGCACTTTAATAAGAACATAAGCTTTCAACCACCCTCAGGAGG
ATTTTGTCCATTTCCTTTTTAATTTTCTCGTGAAATTATTCTTGTATTCGAAAGTTGGTGGGAGTCCTCC

AGATCTAGAAATTACAATGCATCATTTTAGTTGTAGAGGGGAATTTTTCTATTGCAATACAACACAACTG
TCTAGATCTTTAATGTTACGTAGTAAAATCAACATCTCCCCTTAAAAAGATAACGTTATGTTGTGTTGAC

GACAATTGGAGAAGTGAATTATATAAATGTAAAGTAGTACAAATTGAACCACTAGGAATAGCACCCACCA
CTGTTAACCTCTTCACTTAATATATTTACATTTCATCATGTTTAACTTGGTGATCCTTATCGTGGGTGGT

GGGCAAAGAGAAGAGTGGTGGAGAGAGAAAAAAAA – 3′
CCCGTTTCTCTTCTCACCACCTCTCTCTTTTTTTT – 5′

PCR takes place in a thermocycler, which cycles the reaction between three key temperatures that lead to DNA denaturation, primer annealing, and primer extension. The temperatures are important; they control the stability of the dsDNA and allow it to be unwound and separated, effectively doing the job of helicase. The primers need to anneal (bind) to the correct places. Primer extension mimics DNA replication most closely. The special heat-resistant type III TAQ polymerase adds nucleotides to the 3′ end of the primers following base paring rules.

Question

9. Fill in the plot below as we go over each step in a PCR cycle.

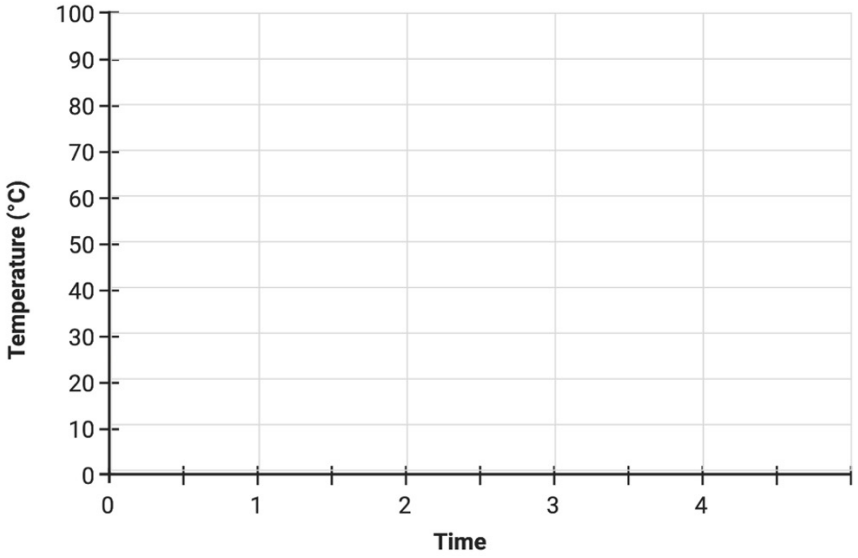


Figure 3. PCR cycle.

Once PCR is complete, each sample has millions of copies of the target sequence. In order to know the order of nucleotides in each sample (and to know what mutations exist among the patients) these PCR samples are sequenced so that the actual DNA sequence of the target region can be compared.

Sequencing is another molecular technique that directly relates to DNA replication. Just like in a cell and in PCR, DNA polymerase replicates the template DNA using dNTPs and primers. In addition to regular dNTPs, the sequencing reaction also includes ddNTPs (dideoxynucleotides). These ddNTPs have been chemically modified in two ways. First, a fluorescent tag is added to distinguish A, T, C, and G bases. Second, the hydroxyl group (OH) that normally appears at the 3′ position is replaced with a single hydrogen. This second modification prevents additional nucleotides from being added to the ddNTP. This sequencing approach, called chain termination sequencing, generates many short DNA fragments that vary in length.

Question

10. Using what you know about the chemical structure of nucleotides, identify the chemical modification that causes chain termination. In Figure 4 below, which is a regular nucleotide, and which is a dideoxynucleotide?

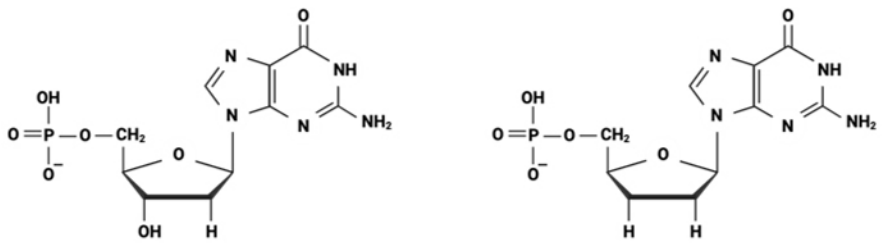


Figure 4. Chain termiation.

The fragments are then sorted by size using gel electrophoresis with the smallest fragment only being one fluorescent nucleotide long and the longest fragment being the length of the template with one fluorescent nucleotide at the end. The size-ordered fragments are scanned with a fluorometer and the color signal is used to determine the identity of the last nucleotide in the sequence.

Sequencing is followed by alignment. In this analysis, the sequenced copies of *gp120* from each of Acer's patients with HIV were lined up and the differences between each person's version of the HIV gene were identified and counted up.

Question

11. In the sequence alignment below, compare each patient to Acer's sequence. Count up the differences as we walk through this example in class.

David Acer	A	G	T	T	G	A	C	T	T	C	T	C	A	G	G	T	A	T	T
Patient A	G	A	T	T	A	G	A	T	T	A	T	C	G	A	G	T	G	T	T
Patient B	G	A	T	T	A	A	C	T	T	C	T	C	C	A	G	T	G	T	T
Patient C	G	G	C	T	G	G	C	T	T	C	T	C	G	G	G	T	A	A	T
Patient D	G	T	T	C	T	A	G	T	C	G	T	G	A	T	T	T	G	T	T
Patient E	G	A	T	T	A	G	A	T	T	C	T	C	G	A	G	T	C	T	T
Patient F	G	T	T	T	T	A	G	T	C	C	T	C	A	T	G	T	G	T	T
Patient G	G	G	C	T	G	G	C	T	T	C	T	C	G	G	G	T	G	T	T

Figure 5. Sequence alignment.

12. Which of the patients were most likely infected by David Acer? Remember to first consider the *number* of changes and then the *type* of changes. A-G transitions and A-C transversions are expected to occur within the infection timeframe while G-T transversions in HIV are rare and less expected within the infection timeframe.

Conclusion

Transmission of infections like HIV from doctor to patient is rare, but it is apparent from DNA analysis that this transmission did occur in the case of David Acer and his patients. While the intent behind the HIV infections remains a mystery (Acer died before criminal charges could be filed), the process explored in this case study (based on the work of evolutionary biologists and geneticists working at the University of Texas at Austin) has become the standard approach for characterizing the transmission of HIV. Abecasis et al. (2018) provides a list of 23 cases that used phylogenetic forensic analysis of HIV. In some cases, the defendant was charged, in other cases the results were inconclusive, and there are also instances of the defendant being exonerated. Now that you know a bit more about how this phylogenetic forensics approach works, consider the following questions.

Questions

13. HIV has such a high mutation rate that a single individual could host thousands of genetic variants. Is analysis of one HIV gene enough to generate a solid conclusion? Consider cases that are trying to establish accidental transmission vs. intentional transmission in a criminal case.

14. Although phylogenetic forensic analysis is the standard tool for investigating HIV transmission patterns (Abecasis et al., 2018), it is not the only application of this method. What are some other uses of phylogenetic forensics in criminal or epidemiological studies?

References

Abecasis, A.B., M. Pingarilho, & A.-M. Vandamme. (2018). Phylogenetic analysis as a forensic tool in HIV transmission investigations. *AIDS* 32(5): 543–54. <https://doi.org/10.1097/QAD.0000000000001728>

Altman, L.K. (1993, June 6). AIDS and a dentist's secrets. *The New York Times*. <https://www.nytimes.com/1993/06/06/weekinreview/aids-and-a-dentist-s-secrets.html>

Aziz, R., P. Sen, P.K. Beura, S. Das, D. Tula, M. Dash, N.D. Namsa, R.C. Deka, E.J. Feil, S.S. Satapathy, et al. (2022). Incorporation of transition to transversion ratio and nonsense mutations, improves the estimation of the number of synonymous and non-synonymous sites in codons. *DNA Research* 29(4), dsac023. <https://doi.org/10.1093/dnares/dsac023>

Danila, R.N., K.L. MacDonald, F.S. Rhame, M.E. Moen, D.O. Reier, J.C. LeTourneau, M.K. Sheehan, J. Armstrong, M.E. Bender, & M.T. Osterholm. (1991). A look-back investigation of patients of an HIV-infected physician: public health implications. *The New England Journal of Medicine* 325(20): 1406–11. <https://doi.org/10.1056/NEJM199111143252003>

Geller, R., P. Domingo-Calap, J.M. Cuevas, P. Rossolillo, M. Negroni, & R. Sanjuán. (2015). The external domains of the HIV-1 envelope are a mutational cold spot. *Nature Communications.* 6, 8571. <https://doi.org/10.1038/ncomms9571>

Hillis, D.M., J.P. Huelsenbeck, & C.W. Cunningham. (1994). Application and accuracy of molecular phylogenies. *Science* 264: 671–7. <https://doi.org/10.1126/science.8171318>

Pray, L. (2008). DNA Replication and causes of mutation. *Nature Education* 1(1), 214. <https://www.nature.com/scitable/topicpage/dna-replication-and-causes-of-mutation-409/>

Reha-Krantz, L.J. (2010). DNA polymerase proofreading: multiple roles maintain genome stability. *Biochimica et Biophysica Acta* 1804(5): 1049–63. <https://doi.org/10.1016/j.bbapap.2009.06.012>

Verma, A.S., V. Kumar, M.K. Saha, S. Dutta, & A. Singh. (2020). HIV: biology to treatment. In: S.K. Saxena & S.M.P. Khurana, eds. *NanoBioMedicine*. Springer Singapore, Singapore. Pp. 167–97. <https://doi.org/10.1007/978-981-32-9898-9_7>

Internet references accessible as of November 2, 2026.