

COMMENTARY

Scientific belief in determining the origins of viruses

Robert Friedman*

Department of Biological Sciences, University of South Carolina, Columbia, South Carolina, United States of America

Abstract

Viruses, as a population of subcellular entities of genetic material in a package, are at least an order of magnitude as numerous as that of their cellular hosts, on which they depend. While the particulars of the emergence of the ancestors of viruses are unknown, the evolutionary histories of their genes are possible by the evolutionary methods of gene sequence comparison and phylogenetic reconstruction. These studies depend on nucleotide or protein sequence data and a reference to the underlying process of genetic replication, which is a cause of taxon-specific differences in the rates of genetic mutation and correction of replication errors. Viral types also do not easily fit within the confines of traditional typologies, such as the use of a host preference for a biological marker of their taxonomic assignments. Instead, their population processes are dynamic and vary across taxa, which are causes of genetic change and new viral lineages. Since a naive deterministic model of evolution is not supported in the corpus of modern science, it is instead an ideal to infer knowledge of viral origins by models based on the stochastic processes of molecular evolution. These approaches are the basis for validating and building knowledge on virus evolution that leads to confidence and robustness in scientific findings (a requisite to the formation of a scientific belief).

Keywords: Virus evolution; Phylogenetics; Population genetics; Scientific method***Corresponding author:**Robert Friedman
(bob.friedman.2@gmail.com)**Citation:** Friedman R. Scientific belief in determining the origins of viruses. *Microbes & Immunity*. 2026;3(4):025500130. doi: 10.36922/MI025500130**Received:** December 8, 2025**Revised:** December 25, 2025**Accepted:** December 29, 2025**Published online:** January 13, 2026**Copyright:** © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.**Publisher's Note:** AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

A naive model of virus evolution is expected to base itself on a set of deterministic natural processes, where viral types are classified by mere observation and the underlying population processes as static ecological phenomena. Instead, a modern approach expects that viral populations are investigated by the general models of population dynamics along with the quantitative rigor of stochastic modeling. These dynamics are generated by a variety of ecological and evolutionary factors, including those of mutation, recombination, natural selection, and genetic drift.¹ At the ecological level, these populations show evidence of subdivision, expansion into novel niches, and formation of new variants by the mechanisms of evolution at the genetic level.

Since the rate of genetic replication of viruses is generally higher than that of cellular life, there is an expectation for their higher potential for adaptiveness in nature.² Their cellular hosts, which the viruses depend on for genetic replication, are generally evolving at a lower rate, and therefore, any extant host population is presumed to have settled on

a commensal or pathogenic relationship with them. This commentary is focused on the pathogenic varieties that infect a vertebrate host.

As a result of these challenges, the jawed vertebrate hosts have a recombinational mechanism at the somatic level for the detection of foreign antigens and to counter the molecular diversity as presented by a variety of pathogenic life, including that of viruses. Otherwise, it is expected that the fitness of a host population is reduced and vulnerable to extinction over an evolutionary time scale.

In contrast to this approach, based on biological mechanism, is a non-scientific perspective, as cautioned about in *The Demon-Haunted World*.³ Instead, the scientific method is a system for explanations of natural processes, based on its power to explain natural phenomena and build knowledge from experimental evidence and data samples from material observation (i.e., empirical inquiry). This approach is not innate to the human experience, but instead emerged recently in history as a formalized method for acquiring knowledge of the natural world, often associated with Baconian empiricism. Although the antecedents to modern science are often relegated as belonging to a separate discipline (philosophy), scientific methodology is descended from – and remains grounded in – philosophy and formal logic. A major reason to pursue philosophy is for the construction of an argument for rejection of the sophistries that are continually and rigorously pursued in human society, such as the formation of belief by statements of mere opinion, or the use of naive analogical reasoning without the substance of material evidence.⁴ These flaws are as likely to occur among those trained in the scientific method, whether by theory or as a practitioner, as they are among untrained others.

2. The phylogenetics of viral populations

Phylogenetic analysis is the prevailing method for the study of virus evolution and their populations (*The Tree of Life* is the only illustration in Charles Darwin's *Origin of Species*). This method is inherently probabilistic and ideally based on genetic data. An alternate method is to analyze data of a virus' physical characteristics, such as its morphological traits, but this kind of approach is hampered by relatively fewer data samples and a lesser reliability as a classifier with taxonomic insight, such as caused by the confounding factors of convergence of traits over time among unrelated populations. Therefore, an expectation is that genetic samples provide a more permanent and quantifiable measure of evolutionary relationships.

Consequently, it is known that the deeper and more ancient evolutionary relationships (ancestral lineages)

across a phylogenetic tree are not as tractable for reconstruction as compared to the more recent ones. This is an expectation from the errors inherent in both the methods and sampling of empirical data and the methodology of evolutionary analysis. One method for estimating these kinds of errors is bootstrap analysis, a repeated recomposition of the genetic data, which leads to measurements of confidence for the pattern of branching across a phylogenetic tree. It also reveals the informativeness of the genetic data and its methodological biases, which leads to a robust interpretation about viral populations and their evolutionary history.

A simple probabilistic model of sequence divergence for a genetic region (a linear arrangement of nucleotides) is expected to show that the greater the divergence, the higher the rate of error inherent in the estimation of it. This effect is expected to propagate to the reconstruction of evolutionary history as defined by the underlying pattern of branches in a phylogenetic tree, which leads to the expectation that the ancient lineages are not as reliably constructed. Similarly, it is possible to devise a thought experiment on the evolution of humans as a descendant from within the order of primates. While the morphological analysis of fossilized skeletal and teeth data reveals relationships about the more recent descendants along the human lineage, it is not as informative in revealing their deeper ancestries in the order of primates. One major cause of this problem is a lack of fossilized data (data sparsity), leading to a source of sampling bias. Another source of error is in the convergence of morphologies among unrelated species, such as the use of tooth size as an evolutionary marker, even though unrelated species can occupy the same ecological niche and converge on similar character states over evolutionary time.

Therefore, data sampling for predicting the distant ancestor of humans and chimpanzees, the most closely related extant species to humans, is much less reliable than for the recently extinct populations of modern humans. The compounding effect of these errors over evolutionary time leads to a high rate of error and low reliability for the construction of the ancestral character states. Likewise, the models of genetic data analysis are subject to the same problem of a compounding error in the estimation of past histories in the ancestral lineages.

This problem of historical reconstruction is also observed in the natural languages of human societies, as it is more difficult to reconstruct the divergent syntax and semantics between ancient forms of Latin and the Romance languages of Europe, while it is a simpler investigation to reconstruct the recent history within a single Romance language. For the latter case, large gaps of knowledge

do not exist, and the possibility of false identification of shared traits is less.

3. Empirical analysis of viral genes

In past studies of biological populations, it has been shown that the origin of new lineages is generally not an outcome of a simple and distinct population process, but instead that there are multiple factors that are influential. An example is in the HIV-1 viral type (human), which is classified into subtypes, yet their molecular evolution shows several events where a subtype is more closely related to simian immunodeficiency virus (SIV; non-human primate) than to other HIV-1 subtypes.^{5,6} Therefore, the original classification scheme naively assigned the viral types by an ecological niche (host preference), yet subsequent genetic analyses showed a different result, and the naming scheme by host species is in conflict with their evolutionary relatedness and membership in a specific lineage (topic of systematic taxonomy).^{7,8}

Instead, it is a modern procedure to construct taxonomies based on evolutionary relationships and the corresponding branching pattern in an evolutionary tree, as Charles Darwin envisioned in the single illustration of his book. Specifically, the naming of the virus type as HIV is misleading in the context of genetic inheritance because it would be assumed that it is a single lineage that emerged from a shared ancestor, but instead subtypes of HIV emerged independently from SIV, and therefore in these cases, their genetic characteristics are expected to show greater similarity to an SIV virus than to the other HIV subtypes.

An additional confounding factor is the lack of SIV data samples across the lineage of primates. Since it is not currently feasible to sample all populations of extant primates, a procedure for avoiding sampling bias, any survey of phylogenetic relationships across HIV-1 and SIV has the potential for future revision with the collection of novel data.

Finally, there is the problem of sampling across a genome, as a portion of genetic material may show evidence of a different branching pattern in a phylogenetic tree than another portion, and one cause of this phenomenon is recombination, where the extant viral types of the study have exchanged genetic material among populations. This problem can also occur due to a lack of stationarity in the evolution of genes in a population.

Another case study of viral evolution is in influenza A, with a genome of several genetic segments encapsulated by a protein shell. This type of virus is typically subclassified by two of its segments at the genetic level, represented by

an abbreviation of “H” or “N,” so one possible viral subtype is “H3N2.” These segments may exchange segments by recombination among viral lineages (reassortment), so in these cases, the H and N segments are expected to reflect a mixed ancestry and show nonconcordant phylogenetic trees.^{9,10}

The last example of recombination in viruses is in the spike protein of severe acute respiratory syndrome coronavirus-2.¹¹ The corresponding gene has undergone a high rate of evolution, as expected from its association with the mechanisms of host infection. This process is inferred to have produced the XBB subtype through recombination between two Omicron BA.2 sublineages.¹²

These case studies show the complexity of biological evolution and a dependence on empirical approaches for acquiring knowledge of natural processes. Theory alone can serve as a guide, but not as a validator, which leads to the construction of a scientific belief on an evolutionary finding. A common vice is to favor the formalisms and appearances of a robust scientific report, but this is not sufficient for the formation of knowledge, but instead a vain pursuit. Instead, it is necessary that any communication on the mechanisms of the natural world is grounded in the tenets of empiricism and a reasonable level of skepticism toward novel conclusions. Complex natural phenomena are not typically amenable to our instinct for order and hierarchy, but instead depend on a complexity of expression of ideas, and a patience for experimentation and subsequent interpretation. The latter is a foundation of modern scientific rhetoric for the arrival at a belief about the causes of natural phenomena.

4. Foundations of scientific beliefs

In the context of evolutionary analysis, the above studies show that mutation and recombination are common occurrences of genetic substitution across virus populations. With their high mutability, along with their large population sizes and high rates of replication, it is expected that their evolutionary adaptiveness is at a more rapid pace than that of their vertebrate hosts.² Likewise, the pathogenic viruses show a high resistance to the mechanisms of adaptive immunity in their vertebrate hosts.

This leads to a prediction that the evolutionary reconstruction of the lineages of viruses depends on a much larger sample of data than is typical in the ecologies of cellular life. This problem is confounded further by the force of genetic recombination, both from known events and those not yet observed, leading to sampling biases in the estimation of the parameters for an evolutionary model

and subsequent construction of a phylogenetic tree. Finally, an empirical study depends on ecological knowledge of the host or hosts of preference by the virus. These practices are expected to increase the probability and confidence in the conclusions from any historical reconstruction by an evolutionary method. This perspective is relevant to both policies in public health and the preservation of animal livestock.

Robust empirical analysis of virus populations relies on robust data sampling at the molecular level, stochastic modeling, and reliability in the methods of phylogenetic reconstruction, although science-based belief in the biological sciences is generally less predictive than offered by the precision of models employed in classical physics. Therefore, confidence and theory across the natural sciences are generally dependent on a continual feedback process in the observation of phenomena, a hypothetical statement about the observations and their underlying causes, and experimentation for validation of a hypothesis. This scientific perspective and the limits on acquiring knowledge are the subject of epistemology and for forming beliefs across the Sciences. Otherwise, the endeavor is wholly reduced to a study of technology and subject to the sophistries of humans in constructing a belief based on observation alone.

In society, there is not a widespread education on the stochastic processes of the natural world, nor is there an opportunity for practical experience in an empirical science, so this leads to a common-sense reliance on the speech of experts as measured by their status and class in a society. The problem is this is as likely to form the mere images of reality and to substitute the modern rigors for acquiring knowledge with that of sophistry and mere rhetoric (the basis of mythology). This is inherent across many human pursuits, as cautioned by the pre-Socratic philosopher Protagoras, "Man is the measure of all things."⁴ This idea can be interpreted as a generalized view that we are the sole judges of our perceptions, but nature is nearly unique in that it is not mere reflections and interpretations of our mind (see the myth of Narcissus), but instead that which emerges from the processes of nature. This is an eternal problem of philosophy and a reminder on the rejection of false ideas and ideals,¹³ otherwise expressed in the ancient world as a worship of idols (personifications of abstract thoughts), and reveals a source of a limit in our capacity to acquire knowledge, alongside any suppression of knowledge that is of the same utility as no knowledge at all.

Acknowledgments

None.

Funding

None.

Conflict of interest

The author declares no conflict of interest.

Author contributions

This is a single-authored article.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

1. Fisher R. *The Genetical Theory of Natural Selection*. Oxford, UK: Clarendon Press; 1930.
doi: 10.5962/bhl.title.27468
2. Aiewsakun P, Katzourakis A. Time-dependent rate phenomenon in viruses. *J Virol*. 2016;90:7184-7195.
doi: 10.1128/jvi.00593-16
3. Sagan C. *The Demon-Haunted World: Science as a Candle in the Dark*. New York, NY, USA: Random House; 1996.
4. Vlastos G. Parmenides Theory of Knowledge. In: *Transactions and Proceedings of the American Philological Association*. Baltimore, MD, USA: Johns Hopkins University Press; 1946. p. 66-77.
5. Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med*. 2011;1:a006841.
doi: 10.1101/cshperspect.a006841
6. Carrasco-Hernandez R, Jácome R, López Vidal Y, Ponce De León S. Are RNA viruses candidate agents for the next global pandemic? A review. *ILAR J*. 2017;58(3):343-358.
doi: 10.1093/ilar/ilx026
7. Hennig W. *Grundzüge Einer Theorie Der Phylogenetischen Systematik*. Berlin, Germany: Deutscher Zentralverlag; 1950.
8. Hennig W. In: Davis DD, Zangerl R, editors. *Phylogenetic Systematics*. Urbana, IL, USA: University of Illinois Press; 1999.
9. Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. Evolution and ecology of influenza a viruses. *Microbiol Rev*. 1992;56:152-179.

doi: 10.1128/mr.56.1.152-179.1992

10. Peacock TP, Moncla L, Dudas G, *et al.* The global H5N1 influenza panzootic in mammals. *Nature*. 2025;637: 304-313.

doi: 10.1038/s41586-024-08054-z

11. Li X, Giorgi EE, Marichannegowda MH, *et al.* Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Sci Adv*. 2020;6:eabb9153.

doi: 10.1126/sciadv.abb9153

12. Uriu K, Ito J, Zahradnik J, *et al.* Enhanced transmissibility, infectivity, and immune resistance of the SARS-CoV-2 Omicron XBB.1.5 variant. *Lancet Infect Dis*. 2023;23: 280-281.

doi: 10.1016/S1473-3099(23)00051-8

13. Kant I. In: Paul G, Wood AW, editors. *Critique of Pure Reason*. Cambridge, UK: Cambridge University Press; 1998.