Pervasive Horizontal Transfer of Adeno-Associated Virus Capsid Genes

Robert J. Gifford [1,2]

[1] MRC-University of Glasgow Centre for Virus Research, University of Glasgow, Glasgow, UK, [2] Centre for Epidemic Response and Innovation, Stellenbosch University, Stellenbosch, South Africa

Abstract

Adeno-associated viruses (AAVs) are non-pathogenic DNA viruses with remarkable gene delivery capabilities, making them essential tools in gene therapy and biomedical research. Despite their therapeutic importance, key aspects of AAV natural biology remain unclear, complicating efforts to explain rare AAV-associated diseases and optimize gene therapy vectors. By analyzing sequence data from virus isolates and endogenous viral elements (EVEs), I reveal a striking evolutionary pattern: while AAV sub-lineages, defined by the replication-associated (*rep*) gene, have broadly co-diverged with host groups over millions of years, capsid (*cap*) diversity has been shaped by extensive recombination. One capsid lineage, Mammalian-wide (*M-wide*), has spread horizontally across diverse rep lineages and host taxa through multiple recombination events. Notably, AAVs with *M-wide* capsids - including AAV-4, AAV-12, and bovine AAV (BAAV) - originate from historical adenovirus (Ad) stocks, raising the possibility that laboratory conditions contributed to capsid transfer. Distinguishing natural from laboratory-driven recombination is essential for understanding AAV ecology and its implications for gene therapy. A systematic sequencing effort in human and primate populations is needed to assess the extent of recombinant capsid acquisition, determine the impact of laboratory-driven recombination on circulating AAV diversity, and track ongoing recombination events that could affect vector safety and efficacy.

Main Text

Introduction

Adeno-associated viruses (AAVs) are small, non-enveloped, single-stranded DNA viruses of the *Dependoparvovirus* genus (*Parvoviridae*) [1]. They require co-infection with helper viruses, such as adenoviruses or herpesviruses, for productive replication [2]. The AAV genome encodes two primary genes: *rep* (essential for replication and genome packaging) and *cap* (encoding capsid proteins that determine host range and tissue specificity). Flanked by inverted terminal repeats (ITRs), the genome replicates via a rolling hairpin mechanism mediated by Rep proteins [3].

AAVs are widely used in gene therapy, where viral genes are replaced with therapeutic transgenes while retaining the ITRs and supplying *rep* and *cap* genes in trans [4]. Recombinant AAV (rAAV) vectors are central to treating genetic disorders, with regulatory approval for conditions like hemophilia and inherited retinal diseases, and potential applications for many others [5]. Capsid proteins, which determine host range and tissue specificity, are key targets for modification to improve gene delivery, transduction efficiency, and tropism, as well as to evade neutralizing antibodies from prior AAV exposure [6].

Despite decades of research on rAAVs, AAV natural biology remains poorly understood. Like other parvoviruses, AAVs exhibit high recombination and mutation rates [7], yet their extensive genomic 'fossil record'—comprising EVEs—suggests long-term host-virus associations [8]. Sequence and serological data indicate that diverse AAVs circulate in humans and non-human primates [9], but virus-host interactions remain unclear. While AAVs are generally considered apathogenic [2], or even beneficial [10], reports of AAV-2 linked to unexplained childhood hepatitis [11], along with emerging evidence of potential pathogenicity, underscore the need for deeper insights into AAV ecology and host interactions [12].

A comprehensive understanding of AAV evolutionary dynamics is crucial for both optimizing gene therapies and elucidating AAV-associated disease risks. Here, we show that AAV capsids have undergone extensive horizontal transfer, a process that may influence both viral evolution and gene therapy applications. Our findings also raise critical questions about the role of laboratory-driven recombination in shaping AAV diversity.

Results and Discussion

Phylogenetic analysis of *rep* genes from AAV isolates and EVEs shows that AAVs form a distinct, well-supported subclade within genus *Dependoparvovirus* (**Fig. 1**). Rooted *rep* phylogenies broadly mirror vertebrate evolution at higher taxonomic levels, suggesting long-term host-virus co-divergence. Furthermore, mong placental mammals, AAV clades broadly align with host orders (primates, rodents, bats, artiodactyls), reflecting stable associations. EVE data further support these relationships: the '*dependo.13-cercopithecidae*' locus, an AAV *rep* pseudogene found across multiple cercopithecine species, indicates a primate-AAV association exceeding 20 million years (My), while the '*dependo.46-gliridae*' locus suggests a >30 My association with rodents [8].

Primate AAV clades also show structured host associations. A well-supported subclade primarily comprises AAVs from Asian primates and humans, while taxa outside this group exhibit mixed host associations, including African primates (**Fig. 1, Table 1**). However, these biogeographic patterns are not absolute: some AAVs linked to the Asian clade have been detected in African primates, and AAVs from both clades circulate in humans [9]. The full extent of these interactions remains uncertain due to uneven sampling and unresolved host linkages

Phylogenies based on *cap* genes reveal a strikingly different topology from *rep*, highlighting distinct evolutionary trajectories. While most *cap* lineages align with mammalian host orders, one exception emerges: a single capsid lineage—termed Mammalian-wide-1 (*M-wide*)—is found across multiple *rep* lineages and host taxa, indicating extensive horizontal transfer of AAV capsid genes (**Fig. 1**). Several M-wide-bearing AAVs—including BAAV, AAV-4, AAV-11, AAV-12, and bank vole (*Myodes glareolus*) AAV type 2 (MgAAV-2)—appear in *rep* phylogenies alongside viruses whose *cap* genes align with their respective host groups (artiodactyls, primates, rodents). The most parsimonious explanation is that multiple *rep* lineages independently acquired *M-wide* capsids via recombination.

While some M-wide acquisitions likely occurred deep in evolutionary time, the origins of certain AAV isolates suggest that laboratory-driven recombination may have played a role. AAVs were first identified in the 1960s as contaminants of adenovirus (Ad) stocks. These stocks – which were derived from primate and bovine tissues—had been established a decade earlier, in an era when

few viruses could be propagated outside live hosts. Given the limited virological tools available, AAVs likely remained undetected in early adenovirus cultures for years. These stocks—often derived from mixed biological materials—may have facilitated AAV co-infection and recombination, with helper viruses, high viral loads, and permissive cell types providing ideal conditions for capsid exchange. Notably, several AAVs encoding M-wide capsids—including AAV-4, AAV-12, and BAAV—were originally isolated from these stocks (**Table 1**). Phylogenetic analyses show that their *cap* genes form a distinct, well-supported subclade (**Fig. 1**), suggesting a shared history linked to laboratory culture.

Since the 2000s, AAVs encoding M-wide capsids have been detected in primate centers [13] and dairy cattle [14]. Thus, if they did indeed originate in lab cultures, these viruses seem to have since established themselves in host populations. Notably, one such virus (AAV-11) is >99% identical in the *rep* gene to a virus (AAV-10) that encodes the canonical primate AAV capsid. This suggests that recent recombination events may be actively shaping AAV diversity. If recombination is driving capsid exchange, the functional diversity of M-wide capsids in primate AAVs—spanning epithelial, intestinal, and lymphoid tropisms—suggests this process could have significant biological consequences [5, 6].

While laboratory propagation may have played a role in M-wide's spread, an alternative possibility is that this lineage was already naturally widespread, with its presence in Ad stocks reflecting pre-existing bovine and primate AAV diversity rather than lab-driven recombination. Disentangling these scenarios is key to understanding AAV ecology and evolution [15]. More broadly, the observed host associations of primate AAVs raise questions about whether they reflect long-standing virus-host relationships or if recent ecological shifts and human activity have influenced AAV transmission.

Modern sequencing technologies can facilitate broad-scale characterization and surveillance of AAV diversity. Such efforts would not only clarify the natural ecology and evolution of AAVs but also inform gene therapy applications by identifying circulating diversity and potential immune interactions. Moreover, given that rAAV production can potentially lead to encapsidated contaminant sequences [16], the possibility of therapy-associated recombination as a long-term risk warrants careful evaluation.

Materials and Methods

Adeno-associated virus (AAV) genomic sequences were obtained from public databases and curated in a standardized framework incorporating published genomic data, alignments, and serotype classifications. Endogenous viral element sequences were sourced from a dedicated comparative genomics resource. Phylogenies were inferred using widely used maximum likelihood methods, and recombination signals were assessed through established statistical tests. All sequence data, phylogenies, and analysis pipelines are available in the associated databases.

Acknowledgments

I thank Professor Robert Kotin for his valuable feedback.

Figure 1. Tanglegram showing *rep/cap* phylogenetic discordance and *M-wide* spread. Comparison of *rep* (left) and *cap* (right) phylogenies, highlighting recombination. The *rep* tree is rooted on *Dependoparvovirus* outgroups, while the *cap* tree is midpoint-rooted. Shaded clades represent host taxonomic groups. The *M-wide* capsid lineage (red) spans multiple mammalian orders. Connecting lines illustrate discordant *rep/cap* phylogenies and independent *M-wide* acquisitions. Taxa potentially acquiring *M-wide* are in bold. The darker shaded region highlights a sub-lineage of *M-wide* AAVs from adenovirus stocks. Bootstrap support >70% is marked (*).

Abbreviations: AAV, adeno-associated virus; EPV, endogenous parvoviral element.

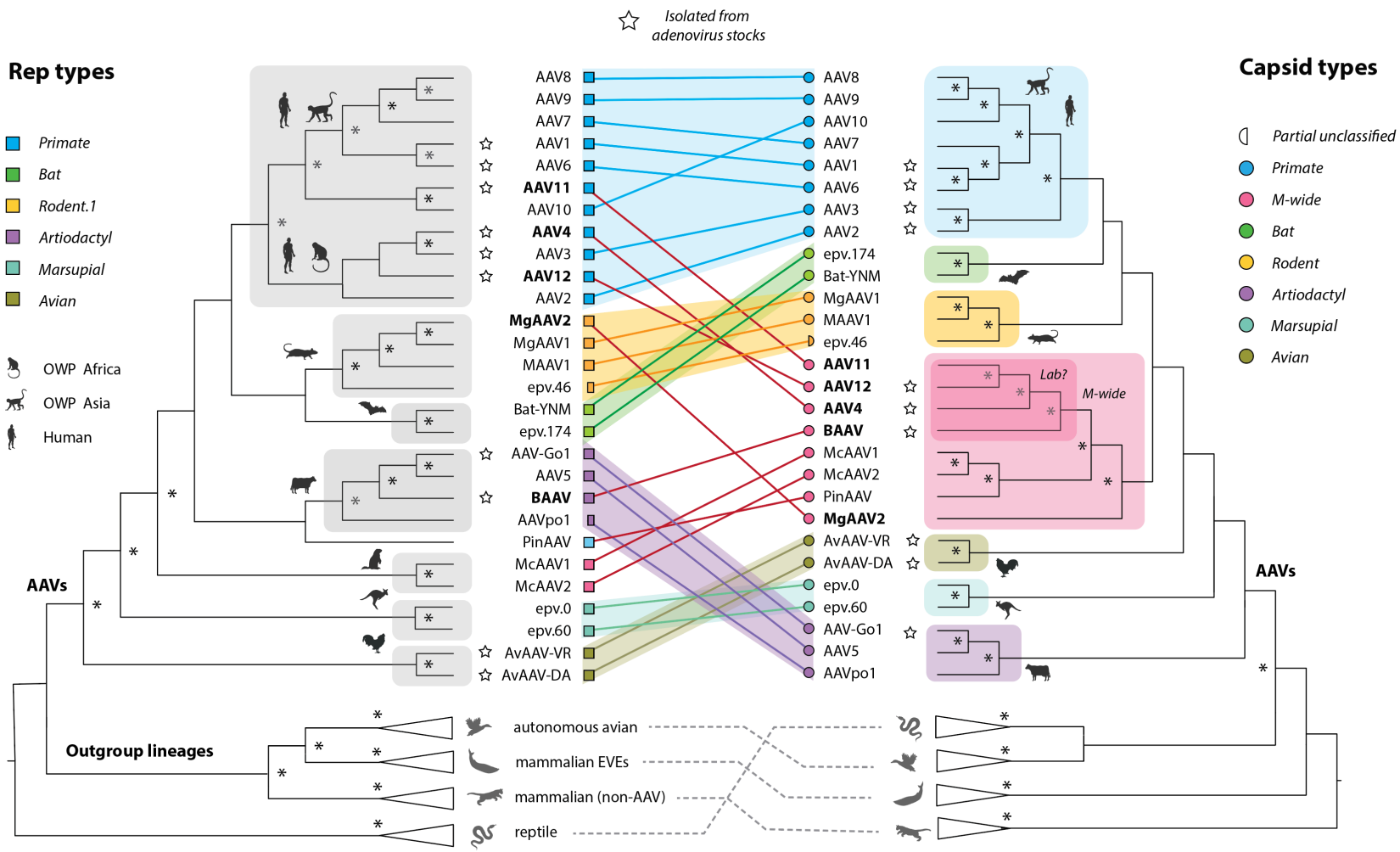
Isolated from adenovirus stocks
Rep types
Primate
Bat
Rodent.1
Artiodactyl
Marsupial
Avian
OWP Africa
OWP Asia
Human
AAVs
Outgroup lineages
AAV8
AAV9
AAV7
AAV1
AAV6
AAV11
AAV10
AAV4
AAV3
AAV12
AAV2
MgAAV2
MgAAV1
MAAV1
epv.46
Bat-YNM
epv.174
AAV-Go1
AAV5
BAAV
AAVpo1
PinAAV
McAAV1
McAAV2
epv.0
epv.60
AvAAV-VR
AvAAV-DA
AAV8
AAV9
AAV10
AAV7
AAV1
AAV6
AAV3
AAV2
epv.174
Bat-YNM
MgAAV1
MAAV1
epv.46
AAV11
AAV12
AAV4
BAAV
McAAV1
McAAV2
PinAAV
MgAAV2
AvAAV-VR
AvAAV-DA
epv.0
epv.60
AAV-Go1
AAV5
AAVpo1
Lab?
M-wide
Capsid types
Partial unclassified
Primate
M-wide
Bat
Rodent
Artiodactyl
Marsupial
Avian
AAVs
autonomous avian
mammalian EVEs
mammalian (non-AAV)
reptile

Table 1. Isolation History & Host Associations of Primate and Artiodactyl Adeno-Associated Viruses

AAV Isolation history			Ad Isolation history					AAV-Host Associations			
Name[1]	**Isolation Source**	**Year**	**Ad Strain(s)**	**Iso. Source**	**Cell culture**	**Year[2]**	**Ad Host**	**Primary Association**	**Seq.[3]**	**Serology[4]**	
										Human	**NHP**
AAV-1	Ad stocks	1965	Ad7	Human	MK, Human	1953	*H. sapiens*	Humans	Yes	++++	+++
AAV-2	Ad stocks	1965	Ad12, Ad7	Human	Human	1953	*H. sapiens*	Humans	Yes	+++++	++
AAV-3	Ad stocks	1965	Ad7	Human	AGMK	1953	*H. sapiens*	Humans?	Yes	+++	++
AAV-4*	Ad stocks	1968	SV15	Macaque	MK, AGMK	1956	*M. mulatta*	OWM Africa	No	+	+++++
AAV-5	Hu. Tissue (Culture)	1982	*n/a*					Humans?	No	+++	++
AAV-6	Ad stocks	1998	Ad5	Human	Human	1953	*H. sapiens*	Humans	No	++++	++
AAV-7	OWM Tissue (PCR)	2004						OWM Asia	Yes	++	+++++
AAV-8	OWM Tissue (PCR)	2004						OWM Asia	Yes	++	+++++
AAV-9	Hu. Tissue (PCR)	2004						OWM Asia	Yes	+++	+++++
AAV-10	OWM Tissue (PCR)	2004						OWM Asia	Yes	+++	+++++
AAV-11*	OWM Tissue (PCR)	2004						OWM Asia	Yes	*n/d*	++
AAV-12*	Ad stocks	2008	SV18	Vervet	n/k	*n/k*	*C. aethiops*	OWM Africa	No	-	*n/d*
BAAV*	Ad stocks	1970	Bovine Ad	Cow	Bovine	1957	*B. taurus*	Cattle	Yes	*n/d*	*n/d*
AAV-Go.1	Ad stocks	2004	Caprine Ad	Goat	Caprine	2004	*C. hircus*	Goat	Yes	*n/d*	*n/d*

Footnote:

1. Asterisks indicate isolates with M-wide capsids.
2. Where multiple adenovirus (Ad) stocks are implicated, the year of the earliest obtained is shown.
3. Independent sequence evidence for presence in the presumed primary host.
4. Serological prevalence in humans and non-human primates is categorized as follows:
    - **+**: Very low
    - **++**: Low
    - **+++**: Moderate
    - **++++**: High
    - **+++++**: Very high
    - **-**: Absent
    - *n/d*: Not determined

References

1. Cotmore, S.F., et al., *ICTV Virus Taxonomy Profile: Parvoviridae.* J Gen Virol, 2019. **100**(3): p. 367-368.
2. Meier, A.F., C. Fraefel, and M. Seyffert, *The Interplay between Adeno-Associated Virus and its Helper Viruses.* Viruses, 2020. **12**(6).
3. Im, D.S. and N. Muzyczka, *The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity.* Cell, 1990. **61**(3): p. 447-57.
4. Hastie, E. and R.J. Samulski, *Adeno-associated virus at 50: a golden anniversary of discovery, research, and gene therapy success--a personal perspective.* Human gene therapy, 2015. **26**(5): p. 257-265.
5. Wang, J.H., et al., *Adeno-associated virus as a delivery vector for gene therapy of human diseases.* Signal Transduct Target Ther, 2024. **9**(1): p. 78.
6. Zolotukhin, S. and L.H. Vandenberghe, *AAV capsid design: A Goldilocks challenge.* Trends Mol Med, 2022. **28**(3): p. 183-193.
7. Duffy, S., L.A. Shackelton, and E.C. Holmes, *Rates of evolutionary change in viruses: patterns and determinants.* Nat Rev Genet, 2008. **9**(4): p. 267-76.
8. Campbell, M.A., et al., *Comparative analysis reveals the long-term coevolutionary history of parvoviruses and vertebrates.* PLoS Biol, 2022. **20**(11): p. e3001867.
9. Gao, G., et al., *Clades of Adeno-associated viruses are widely disseminated in human tissues.* J Virol, 2004. **78**(12): p. 6381-8.
10. Mayor, H.D., et al., *Antibodies to adeno-associated satellite virus and herpes simplex in sera from cancer patients and normal adults.* Am J Obstet Gynecol, 1976. **126**(1): p. 100-4.
11. Tacke, F., *Severe hepatitis outbreak in children linked to AAV2 virus.* Nature, 2023. **617**: p. 471-472.
12. Sant'Anna, T.B. and N.M. Araujo, *Adeno-associated virus infection and its impact in human health: an overview.* Virology Journal, 2022. **19**(1): p. 173.
13. Mori, S., et al., *Two novel adeno-associated viruses from cynomolgus monkey: pseudotyping characterization of capsid protein.* Virology, 2004. **330**(2): p. 375-83.
14. Ng, T.F., et al., *A metagenomics and case-control study to identify viruses associated with bovine respiratory disease.* J Virol, 2015. **89**(10): p. 5340-9.
15. French, R.K. and E.C. Holmes, *An Ecosystems Perspective on Virus Evolution and Emergence.* Trends Microbiol, 2020. **28**(3): p. 165-175.
16. Brimble, M.A., S.M. Winston, and A.M. Davidoff, *Stowaways in the cargo: Contaminating nucleic acids in rAAV preparations for gene therapy.* Mol Ther, 2023. **31**(10): p. 2826-2838.