

## Article

# Distinguishing Genetic Drift from Selection in Papillomavirus Evolution

Robert D. Burk <sup>1,\*</sup> , Lisa Mirabello <sup>2</sup> and Robert DeSalle <sup>3</sup>

<sup>1</sup> Departments of Pediatrics, Microbiology & Immunology, Epidemiology & Population Health, Obstetrics, Gynecology and Woman's Health, and Albert Einstein Cancer Center, Albert Einstein College of Medicine, Bronx, NY 10461, USA

<sup>2</sup> Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD 20850, USA

<sup>3</sup> Sackler Institute of Comparative Genomics, American Museum of Natural History, New York, NY 10024, USA

\* Correspondence: robert.burk@einsteinmed.edu

**Abstract:** Pervasive purifying selection on non-synonymous substitutions is a hallmark of papillomavirus genome history, but the role of selection on and the drift of non-coding DNA motifs on HPV diversification is poorly understood. In this study, more than a thousand complete genomes representing *Alphapapillomavirus* types, lineages, and SNP variants were examined phylogenetically and interrogated for the number and position of non-coding DNA sequence motifs using Principal Components Analyses, Ancestral State Reconstructions, and Phylogenetic Independent Contrasts. For anciently diverged *Alphapapillomavirus* types, composition of the four nucleotides (A, C, G, T), codon usage, trimer usage, and 13 established non-coding DNA sequence motifs revealed phylogenetic clusters consistent with genetic drift. Ancestral state reconstruction and Phylogenetic Independent Contrasts revealed ancient genome alterations, particularly for the CpG and APOBEC3 motifs. Each evolutionary analytical method we performed supports the unanticipated conclusion that genetic drift and different evolutionary drivers have structured *Alphapapillomavirus* genomes in distinct ways during successive epochs, even extending to differences in more recently formed variant lineages.

**Keywords:** molecular evolution; genetic drift; human papillomavirus (HPV)



**Citation:** Burk, R.D.; Mirabello, L.; DeSalle, R. Distinguishing Genetic Drift from Selection in Papillomavirus Evolution. *Viruses* **2023**, *15*, 1631. <https://doi.org/10.3390/v15081631>

Academic Editor: Joe S. Mymryk

Received: 27 June 2023

Revised: 20 July 2023

Accepted: 21 July 2023

Published: 26 July 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Hundreds of different papillomaviruses have been described [1] encompassing the full range of vertebrate hosts from fish [2] and amphibians [3] to birds [4] and especially mammalian host groups [1]. Several dozen species and more than 200 types have been curated from humans [5–8]. *Alphapapillomavirus* genomes have been the most scrutinized in light of their role in cervical and other cancers [9–11]. The *Alphapapillomavirus* 9 species group of human papillomavirus (HPV) types, and HPV16 in particular, stand out as being most strongly associated with carcinogenesis [8,12–16]. Thus, understanding how their evolution has resulted in these devastating human pathogens [17,18] is important.

Considerable effort has been devoted to assessing whether there are amino acid differences that would readily explain the carcinogenic properties of HPV16, the species *Alphapapillomavirus* 9, and the more encompassing high-risk (HR) clade of *Alphapapillomavirus* [19–22]. Evolutionary selection is typically measured on the basis of rates of nonsynonymous to synonymous substitutions in codons [23–25]. Molecular evolutionary biologists have few other tools to convincingly discover and describe evidence of evolutionary pressures acting at more fundamental genome structural or basic nucleotide composition levels. Thus, current evidence for selection or its alternative, genetic drift, are hampered by few analytical methods. Whole genome analyses of various HPV types do not support the notion that pathogenicity is more strongly connected to nonsynonymous substitutions [26] than, perhaps, to a lack thereof [27–29] or to other kinds of substitution

in papillomavirus genomes [29,30]. Indeed, elevated oncogenic risk from large epidemiological studies is also associated with variations in non-coding regions and with silent substitutions that are non-randomly distributed across the ~8000 bp genome [30–33]. Comparative phylogenetic perspectives on papillomavirus genome evolution may yet reveal how disease phenotypes correlate with higher-level clade membership, with species membership, with viral type, and even with lineage and sublineage distinctions [31,34–37].

Chen et al. [38] generated UPGMA trees based on k-mer spectra encompassing all papillomaviruses that corresponded surprisingly well with classical alignment-based phylogenetic trees. Those results indicated that changes in amino acid sequences are not the only consequence of evolutionary pressure. Beyond codons, the depletion of CpG sites appeared to be phylogenetically structured [38], which is relevant given the association of CpG content with nucleosome formation and for methylation and deamination [31,39–47]. Similarly, host class 3 apolipoprotein B mRNA-editing enzyme (APOBEC3) anti-viral activity should select against TCA and TCT sites [27,28,48–51]. However, the early open reading frames E6 and E7 upregulate APOBEC3 and host methyltransferase activity [49,52–55], but they have strikingly different patterns of mutability. Oncogenic HPV16 non-synonymous single nucleotide polymorphism (SNP) variants are hypovaryable in E7 relative to other open reading frames [28]. In contrast, the E6 locus appears to be able to vary more freely than other early- or late-expressed gene sets [28,56].

In this report, we use phylogenetic approaches to investigate non-coding genomic motifs that have previously been of interest regarding the structure and evolution of papillomavirus genomes. Non-recombining asexual organisms, such as papillomaviruses, under continual mutational burden display some aspects of selective pressure to avoid host primordial defenses, such as cytosine deamination [27,43,57,58]. Here, we analyzed the intrinsic nucleotide composition (i.e., A, C, G and T) already speculated to be of lineage-specific significance [59–62], the number and distribution of CpG sites, sites available to APOBEC3 attack, and strand disparities in APOBEC3 sites. We also studied guanine quadruplexes [63], other guanine-rich motifs (e.g., duplexes) [64], toll-like-receptor 9 (TLR9) stimulatory and suppressing sequence motifs implicated in viral pathogenicity [65–70], high-affinity and non-canonical E2-binding sites [71,72], inverted repeats, perfect palindromes, duplicated regions, and reverse complementary regions [73,74]. The evidence regarding the evolution of DNA motifs in HPV genomes is most consistent with genetic drift.

## 2. Materials and Methods

### 2.1. Virus Genome Data

For higher level analyses at the genus level, all reference genomes for each of the 83 *Alphapapillomavirus* reference types were obtained in Genbank format from the Papillomavirus Episteme database (PAVE, <https://pave.niaid.nih.gov>, accessed on 1 April 2023) [5]. For the purposes of phylogenetic and UPGMA analyses, two datasets were established: one unaligned and without modification and a second aligned dataset in which each genome was informatically processed to remove both non-coding and overlapping coding regions on the basis of annotations embedded in the Genbank format. For phylogenetic and ancestral state reconstructions of *Alphapapillomavirus 9*, reference genomes were obtained from PAVE for each variant of *Alphapapillomavirus 9* types, as well as the variants of HPV18 and HPV45 as outgroup taxa. In order to maximize assessment of variability for Principal Components Analysis within *Alphapapillomavirus 9* variants, whole genomes were obtained from NCBI in which no more than 25 nucleotides were missing or ambiguous in the upstream regulatory region (URR) and annotated for all open reading frames. This included a total of 747 genomes for HPV16, 284 for HPV35, 41 for HPV31, 28 for HPV33, 138 for HPV58, and 35 for HPV67. Many deposited complete genomes for HPV52 were found to be incompletely annotated. Relaxing the annotation requirement permitted expansion of the lineage representation for HPV52 to 191 complete genomes.

## 2.2. Phylogenetic and Cluster Analysis

Phylogenetic trees at the *Alphapapillomavirus* species and type levels were obtained with maximum likelihood (GTR+G) using PAUP 4.0a169 [75] analyzing the concatenated nucleotide sequences from non-overlapping open reading frames for E6, E7, E1, E2, L2 and L1 separately aligned according to the aligned inferred amino acid sequences using the TranslatorX server ([translator.co.uk](https://translator.co.uk), accessed on 1 April 2023) [76]. Because codon models of selection must distinguish between synonymous and nonsynonymous nucleotide substitutions in-frame [77], we excluded all overlapping reading frames from those analyses (for example, but not limited to, where E4 and E2 overlap). In addition to phylogenetic trees, the separate unweighted pair group method with arithmetic mean (UPGMA) trees were constructed based on codon usage, trimer composition, and nucleotide composition. For each of these, Euclidean distances were constructed using reference types from PAVE [5]. Distance matrixes were evaluated with UPGMA in the *fitch* package of Phylip version 3.695 [78]. Trees were visualized with FigTree version 1.4.3 [79].

## 2.3. Enumeration of Sequence Motifs

Code was written in Python 3 leveraging the Biopython module to determine the number and position of nucleotide sequence motifs across each whole and unaligned genome analyzed, both in total and in a sliding window scan of the genomes. The level of resolution (sliding window size of 150 nucleotides) was chosen to be small enough (i.e., half the size) to resolve the two smallest open reading frames (E7 and E4), while also being large enough to capture degrees of freedom on the occurrence of DNA motifs. For example, to the extent that CpG motifs are suppressed in the *Alphapapillomavirus 9* lineages, even if those motifs were randomly distributed, one would expect to find a CpG motif only once every 75 dimers. Each iteration involved a 9-nucleotide widow shift to have at least 30 shifts in E6 and E4. Parameters recovered with regular expression (regex) pattern matching (see Supplementary Document S1) in unaligned genomes as well as base composition on the coding strand (i.e., A, C, T, G), trimer composition on the coding strand, CpG motifs, high affinity E2 binding site motifs on the coding and opposite strand, APOBEC3 binding site motifs on both strands, toll-like receptor 9 (TLR9) stimulatory and suppressing motifs on both strands, guanine quadruplexes on both strands, palindromes, near-palindromic inverted-repeats allowing from 3 to 50 spacing nucleotides, larger and more distant duplicated or duplicated, and, finally, reverse-complemented regions being at least 14 nucleotides long and separated by at least 14 nucleotides. Trimer composition was assessed by way of a normalized Euclidean distance in a sliding window relative to that of the whole genome.

## 2.4. Statistical Analyses

Determinations of confidence limits and statistical significance of local trimer composition, GC content, and the relative number of CpG and APOBEC3 sites in each 150 nt sliding window were assessed by comparison of observed values to those obtained from a same-sized window pseudo-sampled from 1000 randomizations of the entire genome. Assessment of differences in relative proportions of covarying residues and relative proportions of APOBEC sites on opposite strands were accomplished through standard Z score calculations with a Bonferroni correction for multiple comparisons.

Principal Components Analysis (PCA) was accomplished in Python 3 with the SciKit-Learn library, using a standard scaler for preprocessing and Seaborn for visualization.

In order to obtain unbiased branch lengths for each DNA sequence motif and for changes in base composition, ancestral state reconstructions of these continuous variables were fitted to the ML topology using a Brownian motion model in Mesquite version 3.70 [80]. For each motif and branch combination, the number of changes was determined to be significantly different where that exceeded the upper 95% confidence threshold for that motif across all branches. Phylogenetic Independent Contrasts (PIC) of reconstructed ancestral states were calculated in Python 3 with the Denropy library. Pearson product mo-

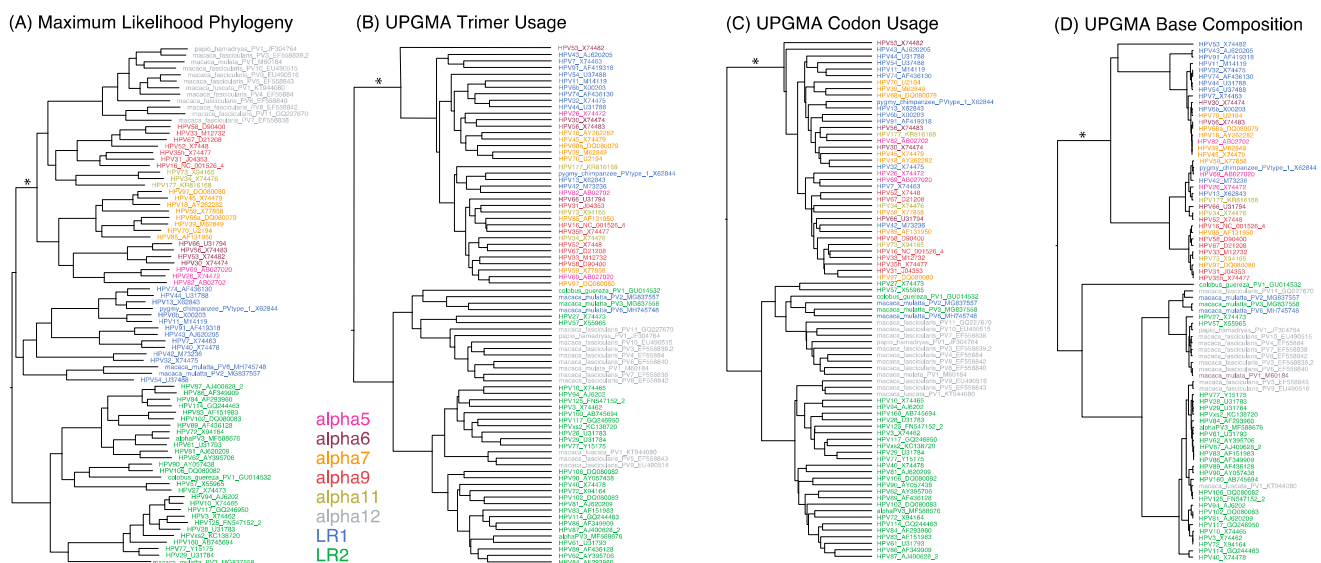
ment correlation  $p$ -values among continuous variable PIC were converted to false discovery (Q) values through the ratio of their relative rank to the total number of comparisons.

Non-overlapping reading frame alignments were examined using codon models with HyPhy [81] for overall rates of nonsynonymous to synonymous substitution (with BUSTED), site-by-site episodic selection (with MEME), and/or changes in the strength of selection (RELAX) using the Datamonkey server [82].

### 3. Results

#### 3.1. Topological Comparisons

Classical molecular evolution methods are based upon the establishment of homology (i.e., alignment) and tree building. Nonetheless, it can be informative to examine how alignment-free analyses compare to an alignment-based phylogeny [38]. In Figure 1, the *Alphapapillomavirus* alignment-based tree (Panel A) is largely recapitulated by the alignment-free trees based on trimer composition (Panel B), codon usage (Panel C), and base compositions (Panel D) where the low-risk 2 (LR2) viruses (green) did not cluster with the low-risk 1 (LR1) (blue) and high-risk (HR) virus types (Figure 1). Measures of topological congruence between trees counting the number of shared and unshared nodes with Matching Pair (MP) distances [83] demonstrated that the trees in Figure 1 are significantly more similar to each other (MP < 390) than to random trees (MP > 549, lower 95% CI = 559,  $p < 0.001$ ). However, all three alignment-free trees placed the non-human primate viruses (*Alphapapillomavirus 12*, grey) within the LR2 clade instead of with the HR group, as found in the alignment-based tree. The discordance between alignment-based and alignment-free trees (Figure 1) could indicate driving features of additional information (e.g., non-coding elements) in the genomes that extend beyond aligned ORFs.



**Figure 1.** Phylogenetic trees of alphapapillomaviruses based on different features. Relatedness of each *Alphapapillomavirus* reference type was inferred from maximum likelihood (ML) phylogenetic analysis of aligned nucleotides in codons that are free to vary, excluding those in overlapping reading frames (A), as well as UPGMA analysis of unaligned genomes for trimer composition (B), codon usage (C), and base composition (D). Branch lengths are proportional to the amount of change. The internal branch uniting the high risk (HR) clade is denoted with an asterisk (\*). HR-HPV types in the Alpha5/6/7/9/11 clade are colored pink, brown, orange, red, and gold, respectively; low-risk (LR) HPV type groups 1 and 2 are shown in blue and green, respectively; and non-human primate HPV types in Alpha12 are shown in grey.

### 3.2. Sliding Window Analyses

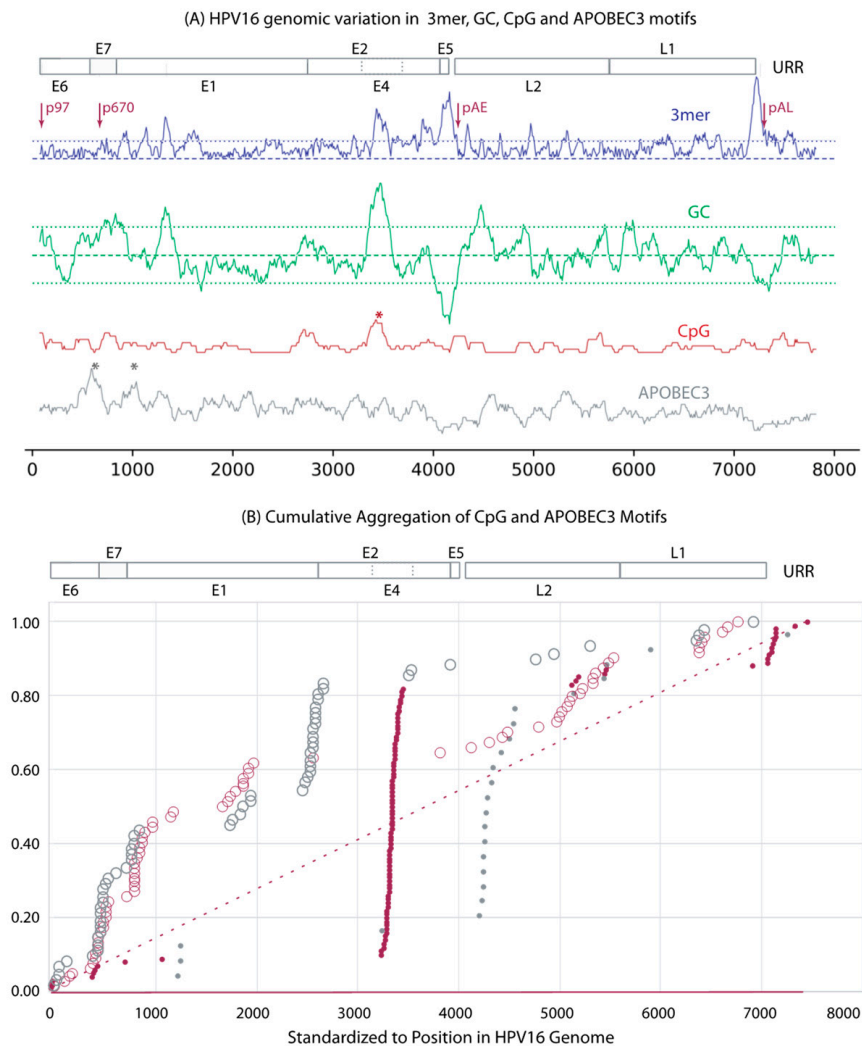
To look for common patterns in unaligned genomes, we evaluated all *Alphapapillomavirus* types ( $n = 83$ ) with sliding window analyses, finding the local number of CpG and APOBEC3 motifs in each 150 bp window while also measuring changes in GC content and local trimer composition; these analyses are mapped to nucleotide positions in the HPV16 reference genome (Figure 2A). For viral types infecting humans (Figure 2B), 70% of the windows with significantly higher concentrations of CpG sites corresponded to E4 (Figure 2B), which also had the highest GC content (green line in Figure 2A) and atypical trimer usage (blue line Figure 2A). For 39 of the *Alphapapillomavirus* types, CpG sites were nonrandomly distributed within E4 itself. A second region of high CpG concentration was found near the late polyadenylation region within the URR (Figure 2B). CpG sites in association with adenosine-rich motifs may be bound by antiviral proteins, such as zinc finger antiviral protein (ZAP) [84]; however, we did not find differences in observed vs. expected numbers of such motifs in genomes of alphapapillomaviruses (data not shown). Concentrations of APOBEC3 recognition sites accumulate linearly above expectations through the early ORFs, with exceptions for regions in E7 and E1 ORFs. The *Alphapapillomavirus* 12 viruses infecting non-human primates exhibit a pattern that is different from those infecting humans (grey plots Figure 2B).

### 3.3. Principal Component Analysis

To further examine the alignment-free clustering noted in Figure 1, we investigated *Alphapapillomavirus* genome content with Principal Components Analysis (PCA). Drift, as predicted by Brownian motion random-walk evolutionary models, results in non-recombining genomes within lineages to diffuse into overlapping regions if given enough time [85–87] (Figure 3 provides a conceptual model in relation to those expectations [85–87]). Thus, under drift, the ability to discriminate lineages diminishes with time (Figure 3), as indicated by recently diverged lineages being well-discriminated (see  $t = 1$  in Figure 3) and more anciently diverged lineages/clades showing broad overlap (see  $t = 3$  in Figure 3). In contrast, a common selection pressure should cause narrow convergence on a distinct optimum in the parameter space [88], and convincing evidence of this would be the convergence of lineages that are not the closest relatives of each other (see light blue and light green lineages at  $t = 3$  in Figure 3B). Using unaligned *Alphapapillomavirus* reference genomes [5], reducing parameter space dimensionality with PCA among the 13 DNA sequence motifs (Supplementary Document S1) and simple base composition (A, C, G, T) showed a number of similarities (Figures 4A and 4B, respectively). PCA incompletely segregated human high-risk HPVs from human LR1 virus types and incompletely segregated LR2 types from NHP-Alpha12 types (NHP), as previously illustrated by alignment-free clustering (Figure 1), and consistent with expectations for anciently diverged clades (see  $t = 3$  in Figure 3A). Indeed, with the exception of *Alphapapillomavirus* 12 and LR2, GC content was significantly different between each clade ( $p < 0.001$  in Mann–Whitney U and Bonferroni post hoc tests).

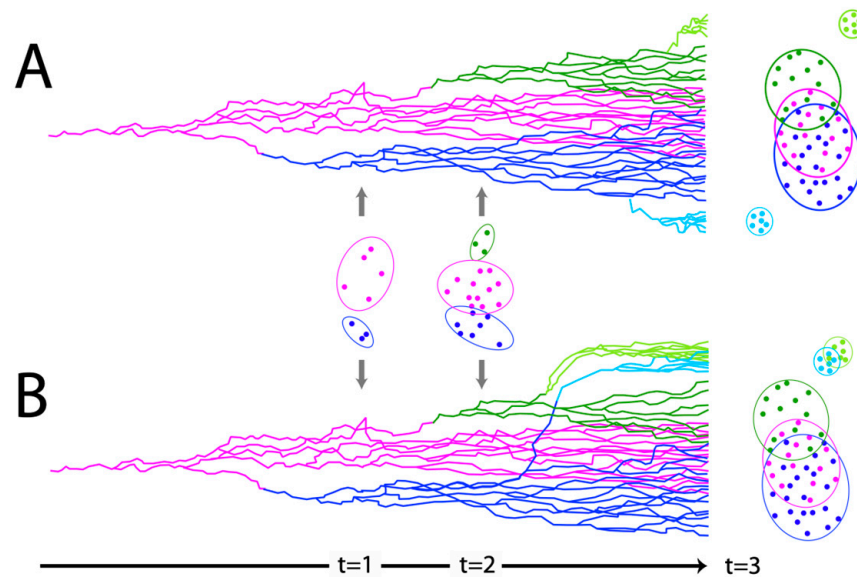
Tens of millions of years since the divergence of the LR1, LR2, and HR clades, clustering still remains that likely reflects their common ancestors. However, diffusion of the phylogenetic pattern is evident (Figure 4). Indeed, this is what is predicted with a Brownian motion drift process in which the displacement from an ancestral condition is the product of the parameter variance and time [89,90] (Figure 3). To evaluate patterns of recent type divergence, we leveraged 1278 complete *Alphapapillomavirus* 9 HPV genomes (from GenBank, August 2022), including types, lineages, sublineages, and SNP variants. PCA of the 13 DNA sequence motifs and of nucleotide base compositions tended to isolate recently diverged types surprisingly well (Figure 5), as also reflected by individual base compositions (Supplementary Document S9). Based on the sequence motifs PCA, the strongest distinctions were for HPV16 (red) and HPV35 (grey), both from each other and from other types, whereas neither HPV52 and HPV67 nor HPV58 and HPV33 were fully discriminated (Figure 5A). Separation of types was more complete based on base composition where

only HPV52 and HPV67 failed to discriminate (Figure 5B), and for which the first principal component largely reflected GC content (Supplementary Document S9). In addition to separating types, PCA of base compositions (Figure 5B) also discriminated among some variant lineages of HPV16, HPV58, HPV31, and HPV67. It should be noted that recently diverged lineages appear to have differentiated in ways that are type-specific (as in Figure 3A), without evidence of common ancestry or common adaptation (as expected in Figure 3B) to a similar ecological niche. In fact, HPV16, the most oncogenic HPV type, best reflects the ancestral nucleotide composition of *Alphapapillomavirus 9* (Supplementary Document S9).



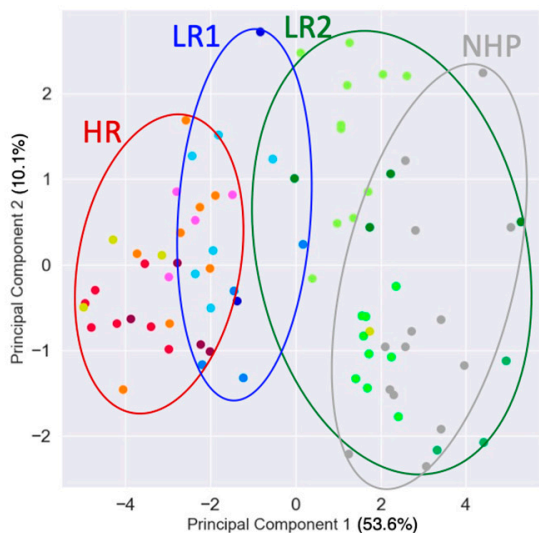
**Figure 2.** Sliding window analyses: (A) HPV16 is shown as a representative genome. The y-axes show the deviation from genome-wide expectations for trimer composition (3mer, purple), GC composition (GC, green), CpG number (CpG, red), and number of APOBEC3 motifs (APOBEC3, grey) across the genome, as determined by sliding window analyses; the average genome-wide expectation for each of these motifs is represented by the dashed line. The 95% confidence limits (dotted lines) are depicted for trimer composition and GC content. Positions with significant clustering of CpG or APOBEC3 sites are marked with asterisks. Promoter and polyadenylation sites are marked with arrows. (B) Cumulative analyses of all 83 *Alphapapillomavirus* reference types showing regions exhibiting significant clustering of CpG (solid circles) and APOBEC3 (open circles) sites. The genome position was standardized against the first codon of E6 in HPV16. Data from HPV *Alphapapillomavirus* types are in red. Data from non-human-primate-specific *Alphapapillomavirus* types are in grey. Red lines indicate expected cumulative distributions should sites be randomly distributed (solid, on the abscissa) and should sites exhibit random clustering (dashed).

### Brownian Motion Model of Neutrally Evolving Non-Recombining Genomes

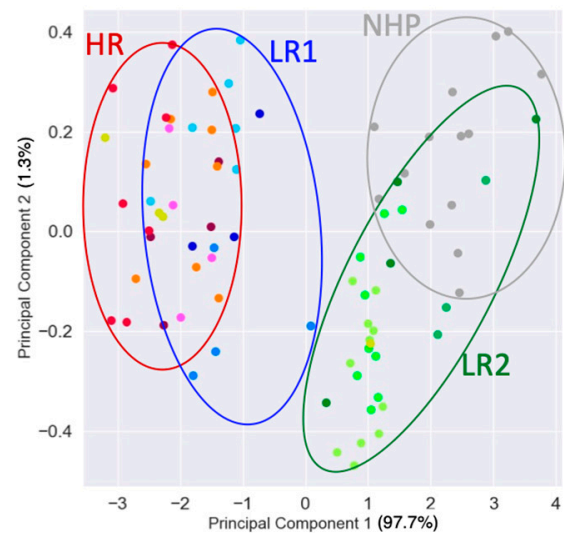


**Figure 3. Model of genetic drift.** Traits evolving under a Brownian motion drift model that readily distinguish recently diverged (asexual, nonrecombining) viral types ( $t = 1$ ) are expected to broaden their distributions through random walk ( $t = 2$ ), resulting in overlapping distributions (A) for more anciently diverged clades ( $t = 3$ ). In contrast, the imposition of a common selective force will draw at least some unrelated clades (B) to a more narrow, distinct, and coincident region of the parameter space (light blue and light green lineages).

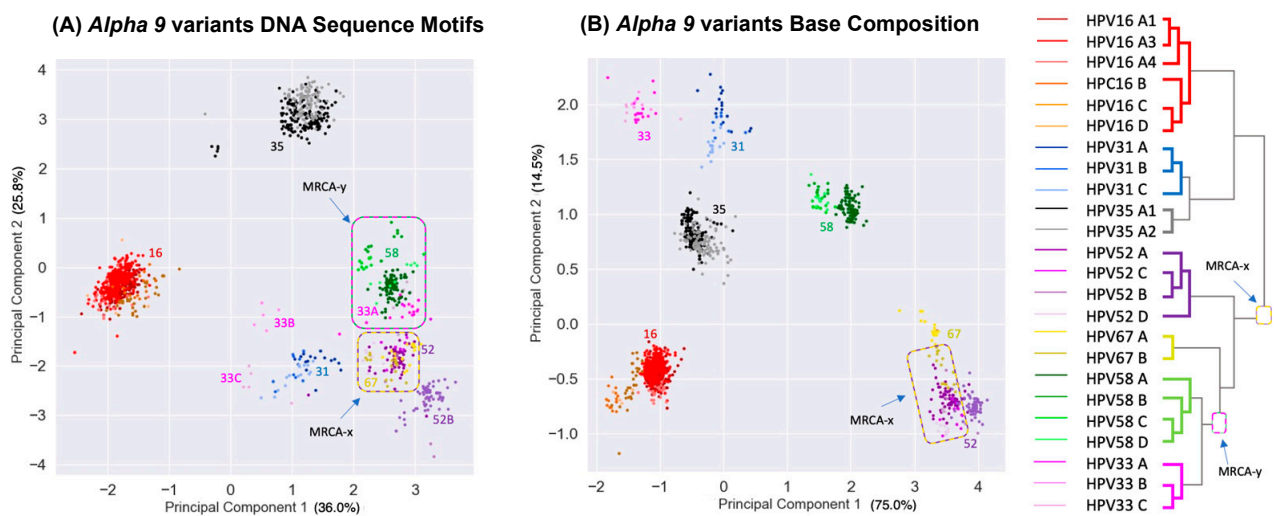
#### (A) *Alphapapillomavirus* DNA Sequence Motifs



#### (B) *Alphapapillomavirus* Base Composition



**Figure 4. Principle component analyses of HPV types.** Plot of second versus first principal components for 83 *Alphapapillomavirus* reference type genomes summarizing variation across 13 DNA sequence motifs (A), and summarizing variation for base compositions alone (B). Each point represents an *Alphapapillomavirus* type genome. Colors correspond to *Alphapapillomavirus* species, as in Figure 1. Ellipses circumscribe each of the two Low Risk (LR1 and LR2), High Risk (HR), and Non-Human-Primate-infecting (NHP) groups. Percentages of variation explained by each principal component are in brackets.



**Figure 5. Principle component analyses of HPV variants.** Plot of second versus first principal components for lineage variants of seven viral types in the *Alphapapillomavirus 9* species summarizing variations across 13 DNA sequence motifs (A), and summarizing variation for base compositions alone (B). Each point represents one of 1278 genomes. Colors correspond to HPV type and variant lineage, as indicated on the phylogenetic tree (at right). Percentages of variation explained by each principal component are in brackets.

The co-location of HPV52 and HPV67 in PCA space (Figure 5A,B) is not explained by recent common ancestry—they do not form a monophyletic group (see tree in Figure 5). An alternative explanation is that both contain features of their most recent common ancestor (denoted MRCA-x on Figure 5), whereas HPV58 and HPV33 have diverged (Figure 5B). A similar case can be made for the co-location of HPV58 and HPV33 genomes—that is, though their base compositions have diverged (Figure 5B), they still overlap in the PCA space based on more complex shared sequence motifs (Figure 5A) consistent with their recent common ancestor (denoted MRCA-y on Figure 5). The PCA results are consistent with Brownian motion [91] acting upon ancestral states of HPV isolates manifesting features of genetic drift [86,90,92].

### 3.4. Ancestral State Reconstruction

To understand how the DNA motifs in HPV genomes have changed over time, we utilized Brownian motion ancestral state reconstruction (see Figure 6, and Supplementary Document S4) [90,93]. Of the 13 DNA sequence motifs, 4 exhibited more ancestral branch change than expected (Figure 6): the number of APOBEC3 recognition sites (11 branches), the number of palindromic regions (8 branches), and the number of CpG sites and TLR9-stimulating motifs on the same branches given that the latter are CpG-rich. HPV16 has the least CpG sites and the most palindromic motifs represented by multiple instances of punctuated change since the *Alphapapillomavirus* MRCA. However, there does not appear to be *Alphapapillomavirus*-wide coordination of significant changes on ancestral lineages as would be expected from a selection pressure that is common to all of the alphapapillomaviruses. In a manner similar to the PCA of recently diverged HPV genomes (Figure 5), ancestral state reconstruction (Figure 6) suggests no obvious common evolutionary trajectory since the *Alphapapillomavirus* MRCA. For example, some lineages show marked decreases in CpG sites whereas others increase. Similarly, both increases and decreases in TLR9 sites are apparent in the ancestral lineages of LR1 and LR2. Thus, each clade seems to have its own specific history with respect to the evolution of these motifs.

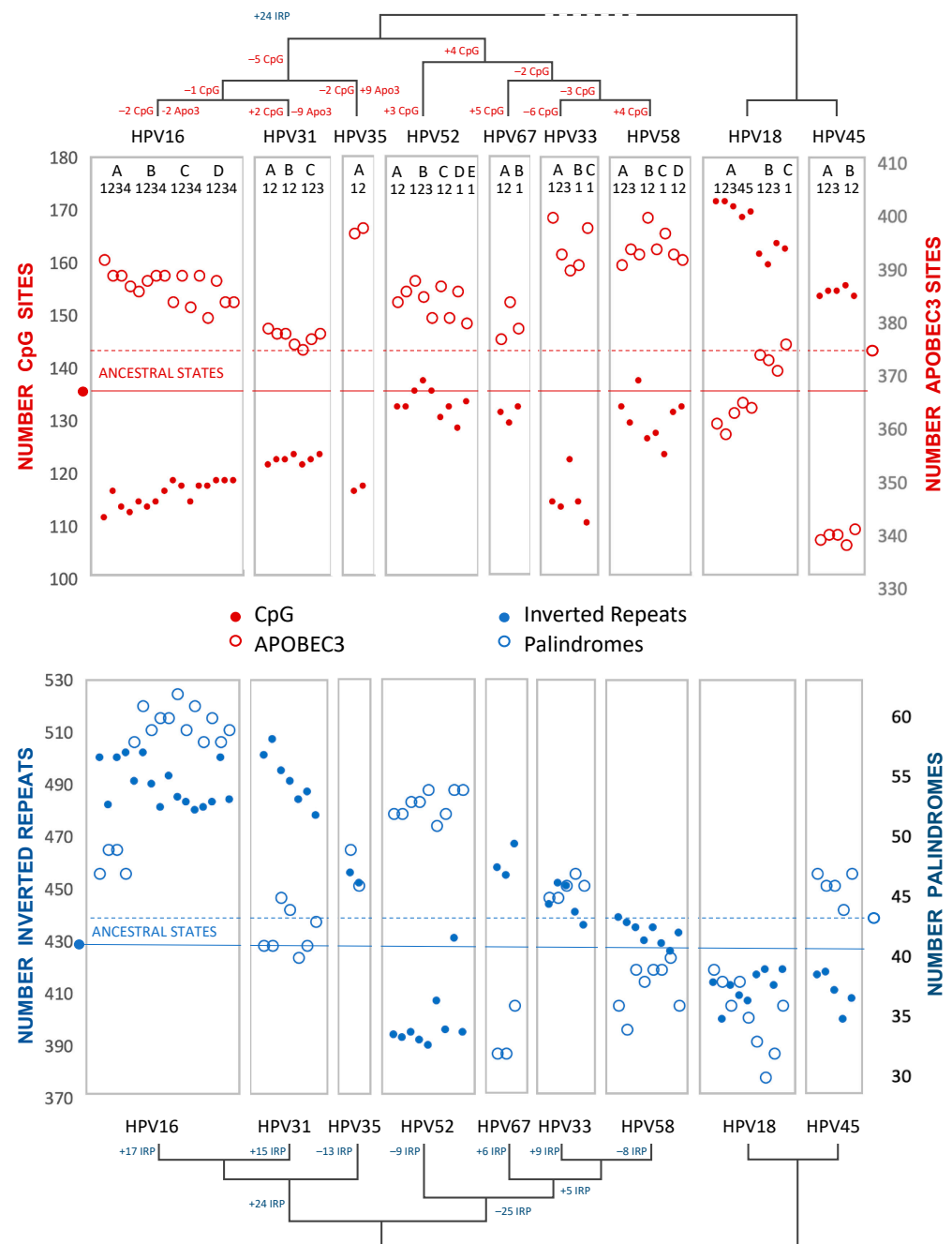




**Figure 6. Ancestral state reconstruction of alphapapillomaviruses.** Significant changes in ancestral states reconstructed on the ML tree for number of CpG sites (CpG), the number of APOBEC3 sites (Apo3), the number of Toll-like receptor 9 stimulating motifs (TLR), and the combined number of inverted repeats and perfect palindromes (IRP) was determined. Colored lines illustrate the ancestral paths from the most recent common ancestor (MRCA) of *Alphapapillomavirus* to the HPV type with the most extreme number of CpG sites (blue), the number of APOBEC3 sites (magenta), the number of Toll-like receptor 9 stimulating motifs (brown), and the combined number of inverted repeats and perfect palindromes (olive). Branch lengths are proportional to total genomic nucleotide change.

As with PCA (above), to explore recent HPV oncogenic type variant divergences, we compared ancestral states for all *Alphapapillomavirus* 9 variant reference genomes from PAVE, including HPV18 and HPV45, as an outgroup to root the *Alphapapillomavirus* 9 tree (Figure 7, and Supplementary Document S5). Brownian motion reconstruction revealed that gradual changes for recently diverged variants have resulted in markedly different trajectories. For example, the number of inverted repeats with potential for secondary structure increases in HPV16/31 but was reduced among HPV52 variants, while the number of perfect palindromes increased in both HPV16 and HPV52 variant lineages (see Supplemental Documents S6–S8). HPV31 and HPV33 have numbers of palindromes that reflect the ancestral condition, implying convergent reduction in HPV67 and HPV58. The number of CpG sites is reduced in variants of HPV16, whereas CpG sites are more numerous in HPV18 and HPV45 than in any *Alphapapillomavirus* 9 type. HPV31 variants, in particular, are closest to the predicted number of ancestral APOBEC3 sites, with increases in these sites for HPV35 and HPV33/58. In the outgroup, HPV45 has markedly fewer APOBEC3 sites in comparison to all other types examined.

These analyses did not suggest a common evolutionary mechanism that would explain how HPV16, HPV18, and HPV58 are among the most prevalent and/or most oncogenic of the HR-HPV viruses. Nevertheless, visual inspection of CpG and APOBEC3 content across *Alphapapillomavirus* 9 variants and the outgroup lineages suggests an inverse relationship (Figure 7). Indeed, with a false-discovery rate of 5%, the extant values for CpG and APOBEC3 motifs on the coding strand of HPV16 and HPV18 variants proved to be inversely correlated: ( $R = -0.655, p = 0.006$ ) ( $R = -0.955, p = 0.00006$ ), respectively. However, this kind of pairwise correlation fails to account for phylogenetic relatedness and the non-independence of closely-related viruses (see also [29]).



**Figure 7.** Ancestral state variation across HPV types, variants, and subvariants in the *Alphapapillomavirus* 9 species. These panels contrast CpG sites with APOBEC3 sites (top), and inverted repeats with palindromes (bottom). The ancestral state inferred for each parameter is represented by a horizontal line. Each boxed group represents an HPV type in phylogenetic order with each point inside a box representing an HPV variant in alphabetical order (i.e., HPV16: A1/A2/A3/A4/B1/B2/B3/B4/C1/C2/C3/C4/D1/D2/D3/D4; HPV31: A1/A2/B1/B2/C1/C2/C3; HPV35: A1/A2; HPV52: A1/A2/B1/B2/B3/C1/C2/D1/E1; HPV67: A1/A2/B1; HPV33: A1/A2/A3/B1/C1; HPV58: A1/A2/A3/B1/B2/C1/D1/D2; HPV18: A1/A2/A3/A4/A5/B1/B2/B3/C; HPV45: A1/A2/A3/B1/B2).

### 3.5. Phylogenetic Independent Contrasts

The inverse pairwise correlation between CpG and APOBEC3 motifs (e.g., HPV18 variants in Figure 7) could result from the close phylogenetic relatedness of variants within lineages compared to the distance between lineages. Phylogenetic Independent Contrasts

(PIC) provides a strategy to tease out these possibilities [86,90]. We looked for phylogenetic correlations among the 13 DNA motifs and base composition (A, C, G, T) in viral genomes at three hierarchical levels: all 83 *Alphapapillomavirus* reference types, 25 types in the human HR clade (*Alpha5/6/7/9/11*), and separately in the variants within *Alphapapillomavirus 7* and *Alphapapillomavirus 9*, as well as in the clade combining *Alphapapillomavirus 5* and *Alphapapillomavirus 6* in order to have similar degrees of freedom for variants.

Using the coding strand, base compositions showed correlations with PIC that are expected from Chargaff's Second Parity Rule [94] (i.e., significant positive correlations for A with T, C with G, and negative for all other combinations), except that in *Alphapapillomavirus 9*, A and T did not exhibit the expected positive correlation (Supplemental Document S10). Additional significant correlations (FDR  $Q \leq 0.01$ ) using PIC on DNA sequence motifs are shown in Supplementary Document S10 at each hierarchical level. Notably, the inverse correlation of CpG and APOBEC3 sites remains significant within the HR and *Alphapapillomavirus 7* clades (Supplementary Document S10). The number of palindromic regions (inverted repeats and perfect palindromes) was inversely related to CpG content across the *Alphapapillomavirus* and HR clades yet was not significant within any HR species subset. The only commonality among recently diverged HR species groups was that TLR9 and APOBEC3 motifs were significantly correlated in the opposite direction in *Alphapapillomavirus 7* and *Alphapapillomavirus 9* (Supplementary Document S10). Taken together, the results of PIC suggest that coordinated changes among the 13 motifs are not uniform through time, as would be expected if alphapapillomaviruses were adapting to the same niche or evolving toward a limited number of selective optimal genome features. This observation is also consistent with drift.

### 3.6. Codon-Based Selection

Lastly, we also evaluated the role of Darwinian (i.e., positive) selection using non-synonymous to synonymous substitution rates (dN/dS). Examining the aligned non-overlapping ORFs resulted in 8676 aligned nucleotides (of which 6165 were variable), representing 2892 aligned amino acid positions (of which 2106 were variable). Using HyPhy [81] with a GTR model, the phylogeny- and genome-wide dN/dS was estimated to be 0.1743, indicating strong purifying selection. Clade-specific differences in the strength of selection were examined with RELAX [95]. The hypotheses with the highest likelihood ratios (LR) included a significant relaxed selection strength (K) in the LR2 clade (K = 0.63; LR = 50.72) and a significant increase in selection in the *Alphapapillomavirus 12* clade relative to all *Alphapapillomavirus* (K = 1.12; LR = 43.26) (Supplemental Document S3). In terms of selection on aligned sites instead of branches, when considering *Alphapapillomavirus* as a group, MEME [25] found purifying selection at 277 amino acid sites ( $p < 0.05$ ) but no sites under diversifying selection after curation with G-blocks [96] (Supplemental Document S2).

## 4. Discussion

In this report, we focus on the role of non-coding DNA motifs and the evolution of alphapapillomaviruses. Rather than recovering patterns of nucleotide evolution that would suggest a common selective framework for alphapapillomaviruses within the cervicovaginal niche, we document non-coding changes occurring in a highly lineage-specific manner. These results are most consistent with directional mutation pressure and neutral evolution (i.e., genetic drift as described by Sueoka [97,98]). Given the rarity of positive Darwinian selection [26] and recombination, new HPV variants emerge containing different combinations of non-coding motifs as we describe from our analysis of 13 higher-order nucleotide motifs. Surprisingly, even nucleotide composition alone discriminated closely related types, e.g., alpha-9 types (compare Figures 3 and 5). The evolution of HPV genomes represents independent yet successful genetic drift away from ancestral niche-adapted genotypes [99] escaping the accumulation of unfavorable mutations and the speed of Muller's Ratchet [100].

In addition to alterations to the number of CpG, APOBEC3, and TLR9 sites, there have been significant changes to the number of palindromic regions, each of which can result from simple changes in overall nucleotide (A, C, G, T) composition [101,102]. A progressive loss of CpG sites was identified within the more prevalent HR-HPV types. This included a 30% reduction from the MRCA of all alphapapillomaviruses to the MRCA of HR-HPV, and, thereafter, an additional 12% decrease from the MRCA of HR-HPV to the MRCA of *Alphapapillomavirus 9*. Globally, HPV16 is the most prevalent type, and it notably possesses less than half the CpG sites inferred to have been present in the MRCA of all *Alphapapillomavirus* types. Undoubtedly the viable lower limit of CpG sites is constrained in part by the overlapping reading frames of E2 and E4, either of which may not be able to vary synonymously without altering the other reading frame non-synonymously (Figure 2). Intriguingly, a second region of increased CpG sites across *Alphapapillomavirus* corresponds to the 5' URR containing a nuclear matrix association region [103–105] where viral genome-wide association studies (VWASs) consistently find SNPs associated with increased carcinogenicity [28,30,32,33].

The diverse family of apolipoprotein B editing and catalytic (APOBEC) enzymes play critical roles in host defense against a wide range of viruses through cytosine deamination of exposed ssDNA. APOBEC3 enzymes are also implicated in precancerous genome instability [51,106–110]. Like methylation, APOBEC3 activity is upregulated in HPV infections [55] and their signature mutation type [111] is associated with viral clearance [27] and the mutational footprint left on neoplastic host cells [48]. Whereas APOBEC3 sites are nonrandomly distributed across many human DNA viruses [27,50], we found more on the transcribed negative-strand than on the coding positive-strand of *Alphapapillomavirus* genomes. This could represent selection to minimize sites on the coding strand, given the tendency of APOBEC3 enzymes to also edit mRNA when they are over-expressed [112], which corresponds well to the inverse correlation of these motifs in Phylogenetic Independent Contrasts (Supplemental Document S10). Similarly, adenosine deaminases acting on RNA (ADAR), host defenses against viral RNA create a regime that would not favor use of adenosine on the coding strand, which could explain a coordinated paucity of antisense APOBEC3 (i.e., AGA/UGA in transcripts) motifs [113].

It was surprising that something as primordial as base composition recapitulated *Alphapapillomavirus* phylogeny and also strongly discriminated *Alphapapillomavirus 9* types by PCA. Clearly, there is more encoded in the genomes of alphapapillomaviruses beyond codon and amino acid usage, alternative splicing of mRNA transcripts [114], and the efficiency afforded by overlapping reading frames [115]. Even simple nucleotide motifs such as tandem repeats of guanine are prone to simultaneous oxidation by glutathione [64], the lack of which has been proposed as a risk factor for the development of cervical cancer [116]. In addition to a tight cytosine balance, nucleotide composition could be under selection on silent substitutions [117]—for example, through effects on the speed, efficiency, and accuracy of transcription and translation [29,118]. Discovering marked differences in the number of palindromic regions (perfect palindromes and inverted repeats) is interesting given that base composition biases lead to higher palindromic content when the bias favors complementary nucleotides (e.g., A and T) [119]. An increase in the number and diversity of locally arranged inverted repeats could be a source of variation in hairpins, cruciforms, and pseudoknots that are open to evolutionary tinkering given their functional significance for transcription, translation, and the influence of miRNA on HPV gene expression [120,121].

The predominance of purifying selection on amino acids is consistent with observations of other dsDNA viruses [122]. We did not find substantial evidence of diversifying selection operating on individual amino acid sites.

Our findings imply something quite different from the notion that *Alphapapillomavirus* diversity resulted from the onset of diversifying selection with ancestral occupation of a new ecological niche [123]. Such a release from stabilizing selection would be marked by evidence of positive selection [124], which is simply not observed. For example, consider the convergent adoption of nonsynonymous mutations (e.g., K417T, E484K, and N501Y) in

the spike protein of multiple lineages of SARS-CoV-2 [125] that are under positive natural selection. The emergence and persistence of new variants based to a large extent on drift could result from genomic “robustness” to the mutation pressures imposed by innate intracellular host restriction factors, especially if accompanied by incomplete purifying selection [126]. Were this the case, one would expect the least mutationally compromised virus to be the most prevalent, and this seems to be the case when comparing HPV16 to other *Alphapapillomavirus* types (Supplementary Data S9). In conclusion, the evolutionary signature recovered in *Alphapapillomavirus* HPV genomes is one of marked purifying selection on amino acid changes (which bodes well for the effectiveness of HPV vaccines), and a lack of convergence towards any single common adaptive peak or strategy. Cytosine deamination patterns and strand-asymmetry are consistent with a directional mutation model of neutral molecular evolution and genetic drift [97,127].

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/v15081631/s1>: S1: Regular-Expression Patterns for DNA Sequence Motifs.doc; S2: MEME sites all Alpha.csv; S3: Selection strength RELAX; S4: HPV Reference Types Ancestral Reconstructions.pdf; S5: HPV *Alphapapillomavirus* 9 Ancestral Reconstructions.pdf; S6: HPV Alpha RefType InvertRepeat.csv; S7: HPV Alpha RefType Duplicated.csv; S8: HPV Alpha RefType Palindromes.csv; S9: HPVAlpha nt changes.pdf; S10: Correlated Phylogenetic Independent Contrasts of DNA Sequence Motifs.

**Author Contributions:** Conceptualization, R.D.B. and R.D.; methodology, R.D. and R.D.B.; software, R.D.; validation, R.D. and R.D.B.; formal analysis, R.D.; investigation, R.D., L.M. and R.D.B.; resources, R.D.B.; data Curation, R.D. and L.M.; writing—original draft preparation, R.D.B.; writing—review and editing, R.D.B., L.M. and R.D.; supervision, R.D.B.; project administration, R.D.B.; funding acquisition, R.D.B. and L.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** The Intramural Research Program of the United States National Institutes of Health and the NCI (grant number: 5U01CA078527; to Burk) supported in part the research in this study, as did the Albert Einstein Cancer Research Center (P30CA013330, PI Ed Chu) and the CFAR at the Albert Einstein College of Medicine (P30AI124414, PI Harris Goldstein).

**Data Availability Statement:** Alignments of nucleotide sequences, amino acid sequences and matrices of continuous variables with corresponding trees can be found at with accession number TB2:S30442at Treebase ([www.trabase.org](http://www.trabase.org)) as accessed on 1 April 2023.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Bernard, H.U.; Burk, R.D.; Chen, Z.; van Doorslaer, K.; zur Hausen, H.; de Villiers, E.M. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology* **2010**, *401*, 70–79. [[CrossRef](#)] [[PubMed](#)]
2. Krabberger, S.; Austin, C.; Farkas, K.; Desvignes, T.; Postlethwait, J.H.; Fontenele, R.S.; Schmidlin, K.; Bradley, R.W.; Warzybok, P.; Van Doorslaer, K.; et al. Discovery of novel fish papillomaviruses: From the Antarctic to the commercial fish market. *Virology* **2022**, *565*, 65–72. [[CrossRef](#)] [[PubMed](#)]
3. Russo, A.G.; Harding, E.F.; Yan, G.J.H.; Selechnik, D.; Ducatez, S.; DeVore, J.L.; Zhou, J.; Sarma, R.R.; Lee, Y.P.; Richardson, M.F.; et al. Discovery of Novel Viruses Associated With the Invasive Cane Toad (*Rhinella marina*) in Its Native and Introduced Ranges. *Front. Microbiol.* **2021**, *12*, 733631. [[CrossRef](#)] [[PubMed](#)]
4. Truchado, D.A.; Williams, R.A.J.; Benitez, L. Natural history of avian papillomaviruses. *Virus Res.* **2018**, *252*, 58–67. [[CrossRef](#)] [[PubMed](#)]
5. Van Doorslaer, K.; Li, Z.; Xirasagar, S.; Maes, P.; Kaminsky, D.; Liou, D.; Sun, Q.; Kaur, R.; Huyen, Y.; McBride, A.A. The Papillomavirus Episteme: A major update to the papillomavirus sequence database. *Nucleic Acids Res.* **2017**, *45*, D499–D506. [[CrossRef](#)]
6. Walker, P.J.; Siddell, S.G.; Lefkowitz, E.J.; Mushegian, A.R.; Adriaenssens, E.M.; Dempsey, D.M.; Dutilh, B.E.; Harrach, B.; Harrison, R.L.; Hendrickson, R.C.; et al. Changes to virus taxonomy and the Statutes ratified by the International Committee on Taxonomy of Viruses (2020). *Arch. Virol.* **2020**, *165*, 2737–2748. [[CrossRef](#)]
7. Muhr, L.S.A.; Eklund, C.; Dillner, J. Towards quality and order in human papillomavirus research. *Virology* **2018**, *519*, 74–76. [[CrossRef](#)]
8. Bruni, L.; Albero, G.; Serrano, B.; Mena, M.; Collado, J.J.; Gómez, D.; Muñoz, J.; Bosch, F.X.; de Sanjosé, S.; ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre). In Human Papillomavirus and Related Diseases in the World. 2021. Available online: <https://hpvcentre.net/statistics/reports/XWX.pdf> (accessed on 26 June 2023).

9. zur Hausen, H. Papillomaviruses and cancer: From basic studies to clinical application. *Nat. Rev. Cancer* **2002**, *2*, 342–350. [[CrossRef](#)]
10. Kidd, L.C.; Chaing, S.; Chipollini, J.; Giuliano, A.R.; Spiess, P.E.; Sharma, P. Relationship between human papillomavirus and penile cancer—implications for prevention and treatment. *Transl. Androl. Urol.* **2017**, *6*, 791–802. [[CrossRef](#)]
11. Sabatini, M.E.; Chiocca, S. Human papillomavirus as a driver of head and neck cancers. *Br. J. Cancer* **2020**, *122*, 306–314. [[CrossRef](#)] [[PubMed](#)]
12. Munger, K.; Baldwin, A.; Edwards, K.M.; Hayakawa, H.; Nguyen, C.L.; Owens, M.; Grace, M.; Huh, K. Mechanisms of human papillomavirus-induced oncogenesis. *J. Virol.* **2004**, *78*, 11451–11460. [[CrossRef](#)] [[PubMed](#)]
13. Schiffman, M.; Rodriguez, A.C.; Chen, Z.; Wacholder, S.; Herrero, R.; Hildesheim, A.; Desalle, R.; Befano, B.; Yu, K.; Safaeian, M.; et al. A population-based prospective study of carcinogenic human papillomavirus variant lineages, viral persistence, and cervical neoplasia. *Cancer Res.* **2010**, *70*, 3159–3169. [[CrossRef](#)] [[PubMed](#)]
14. Byun, J.M.; Jeong, D.H.; Kim, Y.N.; Jung, E.J.; Lee, K.B.; Sung, M.S.; Kim, K.T. Persistent HPV-16 infection leads to recurrence of high-grade cervical intraepithelial neoplasia. *Medicine* **2018**, *97*, e13606. [[CrossRef](#)]
15. Wei, F.; Xia, N.; Ocampo, R.; Goodman, M.T.; Hessol, N.A.; Grinsztejn, B.; Ortiz, A.P.; Zhao, F.; Kojic, E.M.; Kaul, R.; et al. Age-specific prevalence of anal and cervical HPV infection and high-grade lesions in 11 177 women by HIV status: A collaborative pooled analysis of 26 studies. *J. Infect. Dis.* **2022**, *227*, 488–497. [[CrossRef](#)]
16. Arbyn, M.; Weiderpass, E.; Bruni, L.; de Sanjose, S.; Saraiya, M.; Ferlay, J.; Bray, F. Estimates of incidence and mortality of cervical cancer in 2018: A worldwide analysis. *Lancet Glob. Health* **2020**, *8*, e191–e203. [[CrossRef](#)]
17. Schiffman, M.; Herrero, R.; DeSalle, R.; Hildesheim, A.; Wacholder, S.; Cecilia Rodriguez, A.; Bratti, M.C.; Sherman, M.E.; Morales, J.; Guillen, D.; et al. The carcinogenicity of human papillomavirus types reflects viral evolution. *Virology* **2005**, *337*, 76–84. [[CrossRef](#)]
18. Burk, R.D.; Chen, Z.; Van Doorslaer, K. Human Papillomaviruses: Genetic Basis of Carcinogenicity. *Public Health Genom.* **2009**, *12*, 281–290. [[CrossRef](#)]
19. Bouvard, V.; Baan, R.; Straif, K.; Grosse, Y.; Secretan, B.; El Ghissassi, F.; Benbrahim-Tallaa, L.; Guha, N.; Freeman, C.; Galichet, L.; et al. A review of human carcinogens—Part B: Biological agents. *Lancet Oncol.* **2009**, *10*, 321–322. [[CrossRef](#)]
20. Chen, Z.; Schiffman, M.; Herrero, R.; Desalle, R.; Anastos, K.; Segondy, M.; Sahasrabudde, V.V.; Gravitt, P.E.; Hsing, A.W.; Burk, R.D. Evolution and taxonomic classification of human papillomavirus 16 (HPV16)-related variant genomes: HPV31, HPV33, HPV35, HPV52, HPV58 and HPV67. *PLoS ONE* **2011**, *6*, e20183. [[CrossRef](#)]
21. Clifford, G.M.; Howell-Jones, R.; Franceschi, S. Judging the carcinogenicity of human papillomavirus types by single/multiple infection ratio in cervical cancer. *Int. J. Cancer* **2011**, *129*, 1792–1794. [[CrossRef](#)]
22. Chen, Z.; Schiffman, M.; Herrero, R.; DeSalle, R.; Anastos, K.; Segondy, M.; Sahasrabudde, V.V.; Gravitt, P.E.; Hsing, A.W.; Chan, P.K.S.; et al. Classification and evolution of human papillomavirus genome variants: Alpha-5 (HPV26, 51, 69, 82), Alpha-6 (HPV30, 53, 56, 66), Alpha-11 (HPV34, 73), Alpha-13 (HPV54) and Alpha-3 (HPV61). *Virology* **2018**, *516*, 86–101. [[CrossRef](#)]
23. Fitch, W.M. Rate of change of concomitantly variable codons. *J. Mol. Evol.* **1971**, *1*, 84–96. [[CrossRef](#)]
24. Suzuki, Y. New methods for detecting positive selection at single amino acid sites. *J. Mol. Evol.* **2004**, *59*, 11–19. [[CrossRef](#)]
25. Murrell, B.; Wertheim, J.O.; Moola, S.; Weighill, T.; Scheffler, K.; Kosakovsky Pond, S.L. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **2012**, *8*, e1002764. [[CrossRef](#)]
26. King, K.M.; Rajadhyaksha, E.V.; Tobey, I.G.; Van Doorslaer, K. Synonymous nucleotide changes drive papillomavirus evolution. *Tumour Virus Res.* **2022**, *14*, 200248. [[CrossRef](#)]
27. Zhu, B.; Xiao, Y.; Yeager, M.; Clifford, G.; Wentzensen, N.; Cullen, M.; Boland, J.F.; Bass, S.; Steinberg, M.K.; Raine-Bennett, T.; et al. Mutations in the HPV16 genome induced by APOBEC3 are associated with viral clearance. *Nat. Commun.* **2020**, *11*, 886. [[CrossRef](#)]
28. Mirabello, L.; Yeager, M.; Yu, K.; Clifford, G.M.; Xiao, Y.; Zhu, B.; Cullen, M.; Boland, J.F.; Wentzensen, N.; Nelson, C.W.; et al. HPV16 E7 Genetic Conservation Is Critical to Carcinogenesis. *Cell* **2017**, *170*, 1164–1174.e6. [[CrossRef](#)]
29. Shackelton, L.A.; Parrish, C.R.; Holmes, E.C. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *J. Mol. Evol.* **2006**, *62*, 551–563. [[CrossRef](#)] [[PubMed](#)]
30. Pinheiro, M.; Harari, A.; Schiffman, M.; Clifford, G.M.; Chen, Z.; Yeager, M.; Cullen, M.; Boland, J.F.; Raine-Bennett, T.; Steinberg, M.; et al. Phylogenomic Analysis of Human Papillomavirus Type 31 and Cervical Carcinogenesis: A Study of 2093 Viral Genomes. *Viruses* **2021**, *13*, 1948. [[CrossRef](#)] [[PubMed](#)]
31. Mirabello, L.; Yeager, M.; Cullen, M.; Boland, J.F.; Chen, Z.; Wentzensen, N.; Zhang, X.; Yu, K.; Yang, Q.; Mitchell, J.; et al. HPV16 Sublineage Associations With Histology-Specific Cancer Risk Using HPV Whole-Genome Sequences in 3200 Women. *JNCI J. Natl. Cancer Inst.* **2016**, *108*, djw100. [[CrossRef](#)] [[PubMed](#)]
32. Pinheiro, M.; Gage, J.C.; Clifford, G.M.; Demarco, M.; Cheung, L.C.; Chen, Z.; Yeager, M.; Cullen, M.; Boland, J.F.; Chen, X.; et al. Association of HPV35 with cervical carcinogenesis among women of African ancestry: Evidence of viral-host interaction with implications for disease intervention. *Int. J. Cancer* **2020**, *147*, 2677–2686. [[CrossRef](#)]
33. Bee, K.J.; Gradissimo, A.; Chen, Z.; Harari, A.; Schiffman, M.; Raine-Bennett, T.; Castle, P.E.; Clarke, M.; Wentzensen, N.; Burk, R.D. Genetic and Epigenetic Variations of HPV52 in Cervical Precancer. *Int. J. Mol. Sci.* **2021**, *22*, 6463. [[CrossRef](#)] [[PubMed](#)]
34. Kovacic, M.B.; Castle, P.E.; Herrero, R.; Schiffman, M.; Sherman, M.E.; Wacholder, S.; Rodriguez, A.C.; Hutchinson, M.L.; Bratti, M.C.; Hildesheim, A.; et al. Relationships of human papillomavirus type, qualitative viral load, and age with cytologic abnormality. *Cancer Res.* **2006**, *66*, 10112–10119. [[CrossRef](#)] [[PubMed](#)]

35. Berumen, J.; Ordonez, R.M.; Lazcano, E.; Salmeron, J.; Galvan, S.C.; Estrada, R.A.; Yunes, E.; Garcia-Carranca, A.; Gonzalez-Lira, G.; Madrigal-de la Campa, A. Asian-American variants of human papillomavirus 16 and risk for cervical cancer: A case-control study. *J. Natl. Cancer Inst.* **2001**, *93*, 1325–1330. [[CrossRef](#)]
36. Hildesheim, A.; Herrero, R.; Castle, P.E.; Wacholder, S.; Bratti, M.C.; Sherman, M.E.; Lorincz, A.T.; Burk, R.D.; Morales, J.; Rodriguez, A.C.; et al. HPV co-factors related to the development of cervical cancer: Results from a population-based study in Costa Rica. *Br. J. Cancer* **2001**, *84*, 1219–1226. [[CrossRef](#)]
37. Xi, L.F.; Koutsky, L.A.; Hildesheim, A.; Galloway, D.A.; Wheeler, C.M.; Winer, R.L.; Ho, J.; Kiviat, N.B. Risk for high-grade cervical intraepithelial neoplasia associated with variants of human papillomavirus types 16 and 18. *Cancer Epidemiol. Biomark. Prev.* **2007**, *16*, 4–10. [[CrossRef](#)]
38. Chen, Z.; Utro, F.; Platt, D.; DeSalle, R.; Parida, L.; Chan, P.K.S.; Burk, R.D. K-Mer Analyses Reveal Different Evolutionary Histories of Alpha, Beta, and Gamma Papillomaviruses. *Int. J. Mol. Sci.* **2021**, *22*, 9657. [[CrossRef](#)] [[PubMed](#)]
39. Badal, S.; Badal, V.; Calleja-Macias, I.E.; Kalantari, M.; Chuang, L.S.; Li, B.F.; Bernard, H.U. The human papillomavirus-18 genome is efficiently targeted by cellular DNA methylation. *Virology* **2004**, *324*, 483–492. [[CrossRef](#)]
40. Mirabello, L.; Sun, C.; Ghosh, A.; Rodriguez, A.C.; Schiffman, M.; Wentzensen, N.; Hildesheim, A.; Herrero, R.; Wacholder, S.; Lorincz, A.; et al. Methylation of human papillomavirus type 16 genome and risk of cervical precancer in a Costa Rican population. *J. Natl. Cancer Inst.* **2012**, *104*, 556–565. [[CrossRef](#)]
41. Ekanayake Weeramange, C.; Tang, K.D.; Vasani, S.; Langton-Lockton, J.; Kenny, L.; Punyadeera, C. DNA Methylation Changes in Human Papillomavirus-Driven Head and Neck Cancers. *Cells* **2020**, *9*, 1359. [[CrossRef](#)]
42. Wentzensen, N.; Sun, C.; Ghosh, A.; Kinney, W.; Mirabello, L.; Wacholder, S.; Shaber, R.; LaMere, B.; Clarke, M.; Lorincz, A.T.; et al. Methylation of HPV18, HPV31, and HPV45 genomes and cervical intraepithelial neoplasia grade 3. *J. Natl. Cancer Inst.* **2012**, *104*, 1738–1749. [[CrossRef](#)] [[PubMed](#)]
43. Upadhyay, M.; Samal, J.; Kandpal, M.; Vasaikar, S.; Biswas, B.; Gomes, J.; Vivekanandan, P. CpG dinucleotide frequencies reveal the role of host methylation capabilities in parvovirus evolution. *J. Virol.* **2013**, *87*, 13816–13824. [[CrossRef](#)] [[PubMed](#)]
44. Xi, L.F.; Jiang, M.; Shen, Z.; Hulbert, A.; Zhou, X.H.; Lin, Y.Y.; Kiviat, N.B.; Koutsky, L.A. Inverse association between methylation of human papillomavirus type 16 DNA and risk of cervical intraepithelial neoplasia grades 2 or 3. *PLoS ONE* **2011**, *6*, e23897. [[CrossRef](#)]
45. Piyathilake, C.J.; Macaluso, M.; Alvarez, R.D.; Chen, M.; Badiga, S.; Edberg, J.C.; Partridge, E.E.; Johanning, G.L. A higher degree of methylation of the HPV 16 E6 gene is associated with a lower likelihood of being diagnosed with cervical intraepithelial neoplasia. *Cancer* **2011**, *117*, 957–963. [[CrossRef](#)]
46. Clarke, M.A.; Wentzensen, N.; Mirabello, L.; Ghosh, A.; Wacholder, S.; Harari, A.; Lorincz, A.; Schiffman, M.; Burk, R.D. Human papillomavirus DNA methylation as a potential biomarker for cervical cancer. *Cancer Epidemiol. Biomark. Prev.* **2012**, *21*, 2125–2137. [[CrossRef](#)]
47. Feng, C.; Dong, J.; Chang, W.; Cui, M.; Xu, T. The Progress of Methylation Regulation in Gene Expression of Cervical Cancer. *Int. J. Genom.* **2018**, *2018*, 8260652. [[CrossRef](#)]
48. Vartanian, J.P.; Guetard, D.; Henry, M.; Wain-Hobson, S. Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science* **2008**, *320*, 230–233. [[CrossRef](#)]
49. Warren, C.J.; Westrich, J.A.; Doorslaer, K.V.; Pyeon, D. Roles of APOBEC3A and APOBEC3B in Human Papillomavirus Infection and Disease Progression. *Viruses* **2017**, *9*, 233. [[CrossRef](#)]
50. Poulain, F.; Lejeune, N.; Willemart, K.; Gillet, N.A. Footprint of the host restriction factors APOBEC3 on the genome of human viruses. *PLoS Pathog.* **2020**, *16*, e1008718. [[CrossRef](#)]
51. Xu, W.K.; Byun, H.; Dudley, J.P. The Role of APOBECs in Viral Replication. *Microorganisms* **2020**, *8*, 1899. [[CrossRef](#)]
52. Burgers, W.A.; Blanchon, L.; Pradhan, S.; de Launoit, Y.; Kouzarides, T.; Fuks, F. Viral oncoproteins target the DNA methyltransferases. *Oncogene* **2007**, *26*, 1650–1655. [[CrossRef](#)]
53. Au Yeung, C.L.; Tsang, W.P.; Tsang, T.Y.; Co, N.N.; Yau, P.L.; Kwok, T.T. HPV-16 E6 upregulation of DNMT1 through repression of tumor suppressor p53. *Oncol. Rep.* **2010**, *24*, 1599–1604. [[CrossRef](#)]
54. Sen, P.; Ganguly, P.; Ganguly, N. Modulation of DNA methylation by human papillomavirus E6 and E7 oncoproteins in cervical cancer. *Oncol. Lett.* **2018**, *15*, 11–22. [[CrossRef](#)]
55. Wallace, N.A.; Munger, K. The curious case of APOBEC3 activation by cancer-associated human papillomaviruses. *PLoS Pathog.* **2018**, *14*, e1006717. [[CrossRef](#)]
56. Zhe, X.; Xin, H.; Pan, Z.; Jin, F.; Zheng, W.; Li, H.; Li, D.; Cao, D.; Li, Y.; Zhang, C.; et al. Genetic variations in E6, E7 and the long control region of human papillomavirus type 16 among patients with cervical lesions in Xinjiang, China. *Cancer Cell Int.* **2019**, *19*, 65. [[CrossRef](#)] [[PubMed](#)]
57. Woo, P.C.; Wong, B.H.; Huang, Y.; Lau, S.K.; Yuen, K.Y. Cytosine deamination and selection of CpG suppressed clones are the two major independent biological forces that shape codon usage bias in coronaviruses. *Virology* **2007**, *369*, 431–442. [[CrossRef](#)]
58. Van Doorslaer, K. Evolution of the Papillomaviridae. *Virology* **2013**, *445*, 11–20. [[CrossRef](#)] [[PubMed](#)]
59. Liu, P.; Qiu, Y.; Xing, C.; Zhou, J.H.; Yang, W.H.; Wang, Q.; Li, J.Y.; Han, X.; Zhang, Y.Z.; Ge, X.Y. Detection and genome characterization of two novel papillomaviruses and a novel polyomavirus in tree shrew (*Tupaia belangeri chinensis*) in China. *Virol. J.* **2019**, *16*, 35. [[CrossRef](#)] [[PubMed](#)]

60. Latsuzbaia, A.; Wienecke-Baldacchino, A.; Tapp, J.; Arbyn, M.; Karabegovic, I.; Chen, Z.; Fischer, M.; Muhlschlegel, F.; Weyers, S.; Pesch, P.; et al. Characterization and Diversity of 243 Complete Human Papillomavirus Genomes in Cervical Swabs Using Next Generation Sequencing. *Viruses* **2020**, *12*, 1437. [[CrossRef](#)] [[PubMed](#)]
61. Lou, H.; Boland, J.F.; Torres-Gonzalez, E.; Albanez, A.; Zhou, W.; Steinberg, M.K.; Diaw, L.; Mitchell, J.; Roberson, D.; Cullen, M.; et al. The D2 and D3 Sublineages of Human Papilloma Virus 16-Positive Cervical Cancer in Guatemala Differ in Integration Rate and Age of Diagnosis. *Cancer Res.* **2020**, *80*, 3803–3809. [[CrossRef](#)]
62. Liu, W.; Li, J.; Du, H.; Ou, Z. Mutation Profiles, Glycosylation Site Distribution and Codon Usage Bias of Human Papillomavirus Type 16. *Viruses* **2021**, *13*, 1281. [[CrossRef](#)] [[PubMed](#)]
63. Tluczkova, K.; Marusic, M.; Tothova, P.; Bauer, L.; Sket, P.; Plavec, J.; Viglasky, V. Human papillomavirus G-quadruplexes. *Biochemistry* **2013**, *52*, 7207–7216. [[CrossRef](#)]
64. Cadet, J.; Douki, T.; Ravanat, J.L. One-electron oxidation of DNA and inflammation processes. *Nat. Chem. Biol.* **2006**, *2*, 348–349. [[CrossRef](#)] [[PubMed](#)]
65. Krieg, A.M.; Wu, T.; Weeratna, R.; Efler, S.M.; Love-Homan, L.; Yang, L.; Yi, A.K.; Short, D.; Davis, H.L. Sequence motifs in adenoviral DNA block immune activation by stimulatory CpG motifs. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12631–12636. [[CrossRef](#)] [[PubMed](#)]
66. Miller, L.S. Toll-like receptors in skin. *Adv. Dermatol.* **2008**, *24*, 71–87. [[CrossRef](#)]
67. Hasan, U.A.; Zannetti, C.; Parroche, P.; Goutagny, N.; Malfroy, M.; Roblot, G.; Carreira, C.; Hussain, I.; Muller, M.; Taylor-Papadimitriou, J.; et al. The human papillomavirus type 16 E7 oncoprotein induces a transcriptional repressor complex on the Toll-like receptor 9 promoter. *J. Exp. Med.* **2013**, *210*, 1369–1387. [[CrossRef](#)]
68. Pohar, J.; Yamamoto, C.; Fukui, R.; Cajnko, M.M.; Miyake, K.; Jerala, R.; Bencina, M. Selectivity of Human TLR9 for Double CpG Motifs and Implications for the Recognition of Genomic DNA. *J. Immunol.* **2017**, *198*, 2093–2104. [[CrossRef](#)]
69. Cannella, F.; Pierangeli, A.; Scagnolari, C.; Cacciotti, G.; Tranquilli, G.; Stentella, P.; Recine, N.; Antonelli, G. TLR9 is expressed in human papillomavirus-positive cervical cells and is overexpressed in persistent infections. *Immunobiology* **2015**, *220*, 363–368. [[CrossRef](#)]
70. King, K.; Larsen, B.B.; Gryseels, S.; Richet, C.; Kraberger, S.; Jackson, R.; Worobey, M.; Harrison, J.S.; Varsani, A.; Van Doorslaer, K. Coevolutionary Analysis Implicates Toll-Like Receptor 9 in Papillomavirus Restriction. *mBio* **2022**, *13*, e0005422. [[CrossRef](#)] [[PubMed](#)]
71. Rogers, A.; Waltke, M.; Angeletti, P.C. Evolutionary variation of papillomavirus E2 protein and E2 binding sites. *Virology* **2011**, *8*, 379. [[CrossRef](#)]
72. Newhouse, C.D.; Silverstein, S.J. Orientation of a novel DNA binding site affects human papillomavirus-mediated transcription and replication. *J. Virol.* **2001**, *75*, 1722–1735. [[CrossRef](#)] [[PubMed](#)]
73. Gao, R.; Kim, C.; Sei, E.; Foukakis, T.; Crosetto, N.; Chan, L.K.; Srinivasan, M.; Zhang, H.; Meric-Bernstam, F.; Navin, N. Nanogrid single-nucleus RNA sequencing reveals phenotypic diversity in breast cancer. *Nat. Commun.* **2017**, *8*, 228. [[CrossRef](#)] [[PubMed](#)]
74. Ganapathiraju, M.K.; Subramanian, S.; Chaparala, S.; Karunakaran, K.B. A reference catalog of DNA palindromes in the human genome and their variations in 1000 Genomes. *Hum. Genome Var.* **2020**, *7*, 40. [[CrossRef](#)]
75. Swofford, D. PAUP\* (\* Phylogenetic Analysis Using PAUP). 2021. Available online: <https://paup.phylosolutions.com> (accessed on 26 June 2023).
76. Abascal, F.; Zardoya, R.; Telford, M.J. TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **2010**, *38*, W7–W13. [[CrossRef](#)]
77. Monit, C.; Goldstein, R.A.; Towers, G.; Hue, S. Positive Selection Analysis of Overlapping Reading Frames Is Invalid. *AIDS Res. Hum. Retroviruses* **2015**, *31*, 947. [[CrossRef](#)]
78. Felsenstein, J. *PHYMLIP (Phylogeny Inference Package)*; Department of Genome Sciences, University of Washington: Seattle, WA, 2013.
79. Rambaut, A. *FigTree*; Institute of Evolutionary Biology, University of Edinburgh: Edinburgh, UK, 2016.
80. Maddison, W.P.; Maddison, D.R.M. Mesquite: A Modular System for Evolutionary Analysis; 2021. Available online: <https://www.mesquiteproject.org/> (accessed on 26 June 2023).
81. Kosakovsky Pond, S.L.; Poon, A.F.Y.; Velazquez, R.; Weaver, S.; Hepler, N.L.; Murrell, B.; Shank, S.D.; Magalis, B.R.; Bouvier, D.; Nekrutenko, A.; et al. HyPhy 2.5-A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Mol. Biol. Evol.* **2020**, *37*, 295–299. [[CrossRef](#)]
82. Weaver, S.; Shank, S.D.; Spielman, S.J.; Li, M.; Muse, S.V.; Kosakovsky Pond, S.L. Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Mol. Biol. Evol.* **2018**, *35*, 773–777. [[CrossRef](#)] [[PubMed](#)]
83. Bogdanowicz, D.; Giaro, K. Comparing Phylogenetic Trees by Matching Nodes Using the Transfer Distance Between Partitions. *J. Comput. Biol.* **2017**, *24*, 422–435. [[CrossRef](#)] [[PubMed](#)]
84. Luo, X.; Wang, X.; Gao, Y.; Zhu, J.; Liu, S.; Gao, G.; Gao, P. Molecular Mechanism of RNA Recognition by Zinc-Finger Antiviral Protein. *Cell Rep.* **2020**, *30*, 46–52 e44. [[CrossRef](#)]
85. Blanquart, F.; Wymant, C.; Cornelissen, M.; Gall, A.; Bakker, M.; Bezemer, D.; Hall, M.; Hillebregt, M.; Ong, S.H.; Albert, J.; et al. Viral genetic variation accounts for a third of variability in HIV-1 set-point viral load in Europe. *PLoS Biol.* **2017**, *15*, e2001855. [[CrossRef](#)]
86. Harvey, P.H.; Pagel, M.D. *The Comparative Method in Evolutionary Biology*; Oxford Series in Ecology and Evolution; Oxford University Press: Oxford, UK; New York, NY, USA, 1991; 239p.
87. Harmon, L. *Phylogenetic Comparative Methods: Learning from Trees*; 2018. Available online: <https://lukejharmon.github.io/pcm/> (accessed on 26 June 2023).



88. O'Meara, B.C.; Beaulieu, J.M. Modelling Stabilizing Selection: The Attraction of Ornstein–Uhlenbeck Models. In *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 381–393.
89. Edwards, A.W.F.; Cavalli-Sforza, L.L. Reconstruction of evolutionary trees. In *Phenetic and Phylogenetic Classification*; Heywood, W.H., McNeill, J., Eds.; Systematics Association: London, UK, 1964; pp. 67–76.
90. Felsenstein, J. Phylogenies and the Comparative Method. *Am. Nat.* **1985**, *125*, 1–15. [[CrossRef](#)]
91. Einstein, A.; Fürth, R.; Cowper, A.D. *Investigations on the Theory of the Brownian Movement*; Dutton and Company: New York, NY, USA, 1919; 124p.
92. Lynch, M.; Walsh, B. *Genetics and Analysis of Quantitative Traits*; Sinauer: Sunderland, MA, USA, 1998; 980p.
93. Pagel, M.; Cunningham, C. The Maximum Likelihood Approach to Reconstructing Ancestral Character States of Discrete Characters on Phylogenies. *Syst. Biol.* **1999**, *48*, 612–622. [[CrossRef](#)]
94. Rudner, R.; Karkas, J.D.; Chargaff, E. Separation of *B. subtilis* DNA into complementary strands. 3. Direct analysis. *Proc. Natl. Acad. Sci. USA* **1968**, *60*, 921–922. [[CrossRef](#)] [[PubMed](#)]
95. Wertheim, J.O.; Murrell, B.; Smith, M.D.; Kosakovsky Pond, S.L.; Scheffler, K. RELAX: Detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* **2015**, *32*, 820–832. [[CrossRef](#)]
96. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552. [[CrossRef](#)] [[PubMed](#)]
97. Sueoka, N. Directional mutation pressure and neutral molecular evolution. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 2653–2657. [[CrossRef](#)]
98. Jermini, L.S.; Graur, D.; Lowe, R.M.; Crozier, R.H. Analysis of directional mutation pressure and nucleotide content in mitochondrial cytochrome b genes. *J. Mol. Evol.* **1994**, *39*, 160–173. [[CrossRef](#)]
99. Elde, N.C.; Chen, Z.; DeSalle, R.; Schiffman, M.; Herrero, R.; Wood, C.E.; Ruiz, J.C.; Clifford, G.M.; Chan, P.K.S.; Burk, R.D. Niche adaptation and viral transmission of human papillomaviruses from archaic hominins to modern humans. *PLOS Pathog.* **2018**, *14*, e1007352. [[CrossRef](#)]
100. Felsenstein, J. The evolutionary advantage of recombination. *Genetics* **1974**, *78*, 737–756. [[CrossRef](#)]
101. Kypr, J.; Mrázek, J.; Reich, J. Nucleotide composition bias and CpG dinucleotide content in the genomes of HIV and HTLV. *Biochim. Biophys. Acta (BBA) -Gene Struct. Expr.* **1989**, *1009*, 280–282. [[CrossRef](#)]
102. Khrustalev, V.V.; Barkovsky, E.V. Unusual nucleotide content of Rubella virus genome as a consequence of biased RNA-editing: Comparison with Alphaviruses. *Int. J. Bioinform. Res. Appl.* **2011**, *7*, 82–100. [[CrossRef](#)] [[PubMed](#)]
103. Bernard, H.U. Regulation of the transcription of genital human papillomaviruses by CDP, nucleosomes and nuclear-matrix attachment regions. *Papillomavirus Rep.* **2000**, *11*, 73–80.
104. Stunkel, W.; Huang, Z.; Tan, S.H.; O'Connor, M.J.; Bernard, H.U. Nuclear matrix attachment regions of human papillomavirus type 16 repress or activate the E6 promoter, depending on the physical state of the viral DNA. *J. Virol.* **2000**, *74*, 2489–2501. [[CrossRef](#)]
105. Tan, S.H.; Bartsch, D.; Schwarz, E.; Bernard, H.U. Nuclear matrix attachment regions of human papillomavirus type 16 point toward conservation of these genomic elements in all genital papillomaviruses. *J. Virol.* **1998**, *72*, 3610–3622. [[CrossRef](#)]
106. Swanton, C.; McGranahan, N.; Starrett, G.J.; Harris, R.S. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. *Cancer Discov.* **2015**, *5*, 704–712. [[CrossRef](#)] [[PubMed](#)]
107. Salter, J.D.; Bennett, R.P.; Smith, H.C. The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends Biochem. Sci.* **2016**, *41*, 578–594. [[CrossRef](#)] [[PubMed](#)]
108. Hakata, Y.; Miyazawa, M. Deaminase-Independent Mode of Antiretroviral Action in Human and Mouse APOBEC3 Proteins. *Microorganisms* **2020**, *8*, 1976. [[CrossRef](#)]
109. Mourier, T.; Sadykov, M.; Carr, M.J.; Gonzalez, G.; Hall, W.W.; Pain, A. Host-directed editing of the SARS-CoV-2 genome. *Biochem. Biophys. Res. Commun.* **2021**, *538*, 35–39. [[CrossRef](#)]
110. Shapiro, M.; Krug, L.T.; MacCarthy, T. Mutational pressure by host APOBEC3s more strongly affects genes expressed early in the lytic phase of herpes simplex virus-1 (HSV-1) and human polyomavirus (HPyV) infection. *PLoS Pathog.* **2021**, *17*, e1009560. [[CrossRef](#)]
111. Tate, J.G.; Bamford, S.; Jubb, H.C.; Sondka, Z.; Beare, D.M.; Bindal, N.; Boutselakis, H.; Cole, C.G.; Creatore, C.; Dawson, E.; et al. COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **2019**, *47*, D941–D947. [[CrossRef](#)]
112. Sharma, S.; Patnaik, S.K.; Kemer, Z.; Baysal, B.E. Transient overexpression of exogenous APOBEC3A causes C-to-U RNA editing of thousands of genes. *RNA Biol.* **2017**, *14*, 603–610. [[CrossRef](#)] [[PubMed](#)]
113. Chen, J.; Furano, A.V. Breaking bad: The mutagenic effect of DNA repair. *DNA Repair.* **2015**, *32*, 43–51. [[CrossRef](#)] [[PubMed](#)]
114. Berget, S.M.; Moore, C.; Sharp, P.A. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 3171–3175. [[CrossRef](#)] [[PubMed](#)]
115. Sanger, F.; Air, G.M.; Barrell, B.G.; Brown, N.L.; Coulson, A.R.; Fiddes, C.A.; Hutchison, C.A.; Slocombe, P.M.; Smith, M. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* **1977**, *265*, 687–695. [[CrossRef](#)] [[PubMed](#)]
116. Kwasniewska, A.; Tukendorf, A.; Semczuk, M. [Frequency of HPV infection and level of glutathione in serum of women with cervix dysplasia]. *Med. Dosw. Mikrobiol.* **1995**, *47*, 213–218.
117. Kudla, G.; Lipinski, L.; Caffin, F.; Helwak, A.; Zyllicz, M. High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biol.* **2006**, *4*, e180. [[CrossRef](#)]
118. Plotkin, J.B.; Robins, H.; Levine, A.J. Tissue-specific codon usage and the expression of human genes. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 12588–12591. [[CrossRef](#)]

119. Ladoukakis, E.D.; Eyre-Walker, A. The excess of small inverted repeats in prokaryotes. *J. Mol. Evol.* **2008**, *67*, 291–300. [[CrossRef](#)]
120. Tornesello, M.L.; Faraonio, R.; Buonaguro, L.; Annunziata, C.; Starita, N.; Cerasuolo, A.; Pezzuto, F.; Tornesello, A.L.; Buonaguro, F.M. The Role of microRNAs, Long Non-coding RNAs, and Circular RNAs in Cervical Cancer. *Front. Oncol.* **2020**, *10*, 150. [[CrossRef](#)]
121. Bai, L.; Wei, L.; Wang, J.; Li, X.; He, P. Extended effects of human papillomavirus 16 E6-specific short hairpin RNA on cervical carcinoma cells. *Int. J. Gynecol. Cancer* **2006**, *16*, 718–729. [[CrossRef](#)]
122. Wertheim, J.O.; Kosakovsky Pond, S.L. Purifying selection can obscure the ancient age of viral lineages. *Mol. Biol. Evol.* **2011**, *28*, 3355–3365. [[CrossRef](#)] [[PubMed](#)]
123. Bravo, I.G.; Féllez-Sánchez, M. Papillomaviruses. *Evol. Med. Public Health* **2015**, *2015*, 32–51. [[CrossRef](#)]
124. Kapralov, M.V.; Filatov, D.A. Molecular adaptation during adaptive radiation in the Hawaiian endemic genus *Schiedea*. *PLoS ONE* **2006**, *1*, e8. [[CrossRef](#)] [[PubMed](#)]
125. Gobeil, S.M.C.; Janowska, K.; McDowell, S.; Mansouri, K.; Parks, R.; Stalls, V.; Kopp, M.F.; Manne, K.; Li, D.; Wiehe, K.; et al. Effect of natural mutations of SARS-CoV-2 on spike structure, conformation, and antigenicity. *Science* **2021**, *373*, eabi6226. [[CrossRef](#)] [[PubMed](#)]
126. Kustin, T.; Stern, A.; Yeager, M. Biased Mutation and Selection in RNA Viruses. *Mol. Biol. Evol.* **2021**, *38*, 575–588. [[CrossRef](#)] [[PubMed](#)]
127. Mooers, A.O.; Holmes, E.C. The evolution of base composition and phylogenetic inference. *Trends Ecol. Evol.* **2000**, *15*, 365–369. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.