Evolution and diversity of nucleotide and dinucleotide composition in poxviruses

Cristian Molteni, Diego Forni, Rachele Cagliani, Manuela Sironi, Ignacio G. Bravo

JOURNAL OF GENERAL VIROLOGY

RESEARCH ARTICLE

Molteni *et al.*, *Journal of General Virology* 2023;104:001897
DOI 10.1099/jgv.0.001897

MICROBIOLOGY SOCIETY

OPEN ACCESS

Evolution and diversity of nucleotide and dinucleotide composition in poxviruses

Cristian Molteni[1,*], Diego Forni[1], Rachele Cagliani[1], Ignacio G. Bravo[2] and Manuela Sironi[1]

Abstract

Poxviruses (family *Poxviridae*) have long dsDNA genomes and infect a wide range of hosts, including insects, birds, reptiles and mammals. These viruses have substantial incidence, prevalence and disease burden in humans and in other animals. Nucleotide and dinucleotide composition, mostly CpG and TpA, have been largely studied in viral genomes because of their evolutionary and functional implications. We analysed here the nucleotide and dinucleotide composition, as well as codon usage bias, of a set of representative poxvirus genomes, with a very diverse host spectrum. After correcting for overall nucleotide composition, entomopoxviruses displayed low overall GC content, no enrichment in TpA and large variation in CpG enrichment, while chordopoxviruses showed large variation in nucleotide composition, no obvious depletion in CpG and a weak trend for TpA depletion in GC-rich genomes. Overall, intergenome variation in dinucleotide composition in poxviruses is largely accounted for by variation in overall genomic GC levels. Nonetheless, using vaccinia virus as a model, we found that genes expressed at the earliest times in infection are more CpG-depleted than genes expressed at later stages. This observation has parallels in betahepesviruses (also large dsDNA viruses) and suggests an antiviral role for the innate immune system (e.g. via the zinc-finger antiviral protein ZAP) in the early phases of poxvirus infection. We also analysed codon usage bias in poxviruses and we observed that it is mostly determined by genomic GC content, and that stratification after host taxonomy does not contribute to explaining codon usage bias diversity. By analysis of within-species diversity, we show that genomic GC content is the result of mutational biases. Poxvirus genomes that encode a DNA ligase are significantly AT-richer than those that do not, suggesting that DNA repair systems shape mutation biases. Our data shed light on the evolution of poxviruses and inform strategies for their genetic manipulation for therapeutic purposes.

INTRODUCTION

The family *Poxviridae* (order *Chitovirales,* phylum *Nucleocytoviricota*) includes a number of genetically diverse dsDNA viruses that replicate in the cytoplasm. Members of the family infect a wide spectrum of hosts, including insects, birds, reptiles and mammals [1–5]. Based on phylogenetic relationships and host(s) associations, the *Poxviridae* are organized into two subfamilies, *Chordopoxvirinae* and *Entomopoxvirinae*, for vertebrate- and invertebrate- infecting viruses, respectively. Chordopoxviruses are further classified into several genera [6]. Among these, the genus *Orthopoxvirus* includes viruses of great medical relevance, such as variola virus (VARV, the causative agent of smallpox), vaccinia virus (VACV, which was used in the smallpox eradication campaign) and monkeypox virus (MPV), responsible for a recent global outbreak of mpox [7, 8]. Other orthopoxviruses can also infect humans and are usually zoonotically transmitted [9]. Several chordopoxviruses infect wild and domestic animals causing substantial pathology and economic loss (e.g. fowlpox virus, lumpy skin disease virus or orf virus) [10]. Members of other genera within *Chordopoxvirinae* can also infect humans, the most cogent example being molluscum contagiosum virus (MCV), an endemic human-specific pathogen and the only representative of the genus *Molluscipoxvirus* [11]. The subfamily *Entomopoxvirinae* is divided into four genera and includes viruses that establish parasitic or symbiotic relationships with their insect hosts [12–17].

Received 01 August 2023; Accepted 20 September 2023; Published 04 October 2023
Author affiliations: [1]Scientific Institute IRCCS E. MEDEA, Bioinformatics, Bosisio Parini, Italy; [2]Laboratoire MIVEGEC (Univ Montpellier CNRS, IRD), Centre National de la Recherche Scientifique, Montpellier, France.
***Correspondence:** Cristian Molteni, cristian.molteni@lanostrafamiglia.it
Keywords: codon usage bias; CpG; GC content; nucleotide and dinucleotide composition; poxviruses.
Abbreviations: CUB, codon usage bias; G+Ceq, G+C at equilibrium; IE, immediate early; MCV, molluscum contagiosum virus; MPV, monkeypox virus; O/E, observed/expected; OG, orthogroup; pBIC, phylogenetic Bayesian Information Criterion; PCA, principal components analysis; PR, post replicative; RCSU, relative synonymous codon usage; TLR, toll-like receptor; VACV, vaccinia virus; VARV, variola virus; ZAP, zinc-finger antiviral protein.
GenBank accession numbers of all sequences analzyed are reported in Supplementary Materials.
Three supplementary tables and three supplementary figures are available with the online version of this article.

Poxviruses are characterized by large virions and long genomes, ranging in size roughly from 135 to 350 kb [18]. The long-term evolution of chordopoxviruses was characterized by major waves of gene acquisition from host genomes, followed by extensive gene duplications, but also by gene losses [19]. These host–gene capture events have contributed to shape the gene content of extant poxvirus species. Indeed, comparative genomic analyses indicate that extant poxvirus genomes typically harbour a set of core, essential genes plus a variable number of non-core genes, which mainly encode proteins involved in host–virus interactions [18–21]. This arrangement in core and dispensable genes allows poxviruses to show remarkable genome plasticity.

While viral gene content evolves through discontinuous, macromolecular events such as recombination, gene duplication or gene gain/loss, other genomic features such as nucleotide composition and codon usage bias (CUB) are the result of micromolecular events, such as mutation or repair biases. Global and local nucleotide composition can be subject to mutation–selection–drift evolutionary dynamics, if they lead to differential viral fitness [22–24]. This seems to actually be the case, as viral genome compositional features result in differential gene expression, and are likely to contribute to successful infection and evasion of host immune responses [25]. This line of evidence is often described as the 'translational selection hypothesis', proposing that differences in CUB result in functional differences at the gene expression level, ultimately affecting organismal fitness, and being thus subject to natural selection [26]. Indeed, a number of studies have shown that the CUB of viral genes tends to match that of the host [27–34]. Also, changes in viral CUB were reported to modulate the clinical presentation of infection, as in the case of papillomaviruses [35] and of respiratory viruses [36]. Indeed, the link between viral CUB and gene expression is well established, to the extent that synonymous recoding of viral genes to 'deoptimize' CUB has been regarded as a promising strategy for the development of live attenuated vaccines [37–41].

In parallel, a substantial body of evidence reports that CUB in viruses reflects mostly nucleotide and/or dinucleotide composition, patterned by directional mutation events and ultimately subject to drift and/or to historical contingency, rather than arising from translational selection [25, 42–44]. This neutralist view of CUB explains differences in viral genome composition to have evolved to mimic the genomic composition of their hosts' genomes, as animal genomes vary extensively in overall G+C content, but also in terms of dinucleotide representation. In viruses infecting vertebrates, the CpG dinucleotide shows the strongest bias, being highly under-represented [45–48]. This trend towards CpG under-representation is not observed in the majority of viruses infecting invertebrates [47, 49–53]. Conversely, the TpA dinucleotide (UpA in RNA) is under-represented in the genomes of vertebrates and in invertebrates alike [47, 48, 51, 54]. Different hypotheses have been proposed to explain CpG and TpA avoidance in viruses infecting vertebrates [47, 49, 54, 55]. The fitness advantage of dinucleotide biases may arise from a balance between two contrasting trends: (i) an increase in gene expression efficiency, as genes enriched in CpG are more efficiently transcribed and translated [56, 57]; and (ii) an increased exposure to the immune system, as genes enriched in CpG or TpA deviate from the host's compositional standards and could serve as targets for immune mechanisms triggered by the recognition of non-self nucleic acids. Indeed, elements in the vertebrate non-adaptive immune system have evolved to target nucleic acids with compositional features under-represented in their own genome, such as CpG- or TpA-enriched DNA or RNA. This is the case of toll-like receptor 9 (TLR9), which can recognize the presence in the cytoplasm of dsDNA with non-methylated CpG dinucleotides, of RNAse L, which degrades cytoplasmic RNA with a preference for a cleavage after TpA or TpT sites, or of the zinc-finger antiviral protein (ZAP), which hinders the infectious cycle of RNA viruses with elevated frequencies of CpG and/or UpA dinucleotides [58–68]. Although the antiviral activity of RNAse L or ZAP have been mainly studied in the context of RNA virus infections, they may mediate an antiviral action against DNA virus infections by a direct action on the viral mRNAs. Indeed, recent evidence has indicated that the herpesvirus human cytomegalovirus (HCMV, a large dsDNA virus) can be restricted by ZAP through the targeting of viral transcripts [69, 70]. Also, ZAP counteracts the replication of *modified vaccinia virus Ankara* (a highly passaged VACV strain) in human cells [71]. Overall, a sound body of experimental evidence shows a decreased infectivity of DNA and of RNA viruses after synonymous genomic recoding to increase TpA and/or CpG [51, 72–76], and experimental development confirms a decreased fitness of viral genomes with increased TpA and/or CpG content [77].

The compositional features of viral genomes have been studied in great detail for RNA viruses, whereas much less is known about the origin, diversity and evolution of the genomic nucleotide composition of large DNA viruses in general, and of poxviruses in particular. The forces that have shaped the CUB and nucleotide composition of these viruses have remained poorly characterized, and no comprehensive analysis of CpG and TpA composition has been reported for poxviruses. We aimed here to fill these knowledge gaps, by using sequence information of representative poxviruses that belong to distinct viral genera and infect highly diverse hosts. Our results are relevant not only to further our understanding of the major forces driving genome evolution in these viruses, but also to provide clues on possible strategies for the engineering of attenuated strains for vaccination purposes.

METHODS

Viral genome sequences and alignments

A list of 53 complete poxvirus genomes was derived from the ICTV metadata resource (https://ictv.global/vmr). Specifically, we selected all exemplar strains with complete genome sequence information (n=52). In addition, we included a recently described orthopoxvirus, Alaskapox virus (NCBI accession number: MN240300 [78]). The final dataset contained a total of 53 viruses,

Table 1. List of poxvirus names and abbreviations

Subfamily	Genus	Species
Entomopoxvirinae	Alphaentomopoxvirus	*Anomala cuprea entomopoxvirus* (ACEV)
	Betaentomopoxvirus	*Amsacta moorei entomopoxvirus* (AMEV)
	Betaentomopoxvirus	*Adoxophyes honmai entomopoxvirus* (AHEV)
	Betaentomopoxvirus	*Choristoneura biennis entomopoxvirus* (CBEV)
	Betaentomopoxvirus	*Choristoneura rosaceana entomopoxvirus* (CREV)
	Betaentomopoxvirus	*Mythimna separata entomopoxvirus* (LMEV)
	Unclassified	*Diachasmimorpha entomopoxvirus* (DLEV)
	Deltaentomopoxvirus	*Melanoplus sanguinipes entomopoxvirus* (MSEV)
Chordopoxvirinae	Avipoxvirus	*Fowlpox virus* (FPV); *Canarypox virus* (CPV); *Flamingopox virus* (FLMPV); *Penguinpox virus* (PNGPV); *Pigeonpox virus* (PPV); *Turkeypox virus* (TKPV)
	Capripoxvirus	*Sheeppox virus* (SPV); *Goatpox virus* (GPV); *Lumpy skin disease virus* (LSDV)
	Centapoxvirus	*Murmansk microtuspox virus* (MMPV); *Yokapox virus* (YKV)
	Cervidpoxvirus	*Mule deerpox virus* (DPV)
	Crocodylidpoxvirus	*Nile crocodilepox virus* (NCPV)
	Leporipoxvirus	*Myxoma virus* (MYXV); *Rabbit fibroma virus* (RFV)
	Macropopoxvirus	*Western kangaroopox virus* (WKPV); *Eastern kangaroopox virus* (EKPV)
	Molluscipoxvirus	*Molluscum contagiosum virus* (MCV)
	Mustelpoxvirus	*Sea otterpox virus* (SOPV)
	Oryzopoxvirus	*Cotia virus* (COTV)
	Parapoxvirus	*Orf virus* (ORFV); *Bovine papular stomatitis virus* (BPSV); *Grey sealpox virus* (GSEPV); *Pseudocowpox virus* (PCPV); *Red deerpox virus* (RDPV)
	Pteropopoxvirus	*Pteropox virus* (PTPV)
	Salmonpoxvirus	*Salmon gillpox virus* (SMGPV)
	Sciuripoxvirus	*Squirrelpox virus* (SQPV)
	Suipoxvirus	*Swinepox virus* (SWPV)
	Vespertilionpoxvirus	*Eptesipox virus* (EPTPV)
	Yatapoxvirus	*Yaba monkey tumor virus* (YMTV); *Tanapox* (YLDV)
	Orthopoxvirus	*Vaccinia virus* (VACV); *Abatino macacapox virus* (ABMPV); *Akhmeta virus* (AKHV); *Camelpox virus* (CMPV); *Cowpox virus* (CPV); *Ectromelia virus* (ECTV); *Monkeypox virus* (MPV); *Raccoonpox virus* (RAPV); *Skunkpox virus* (SKPV); *Taterapox virus* (TATPV); *Variola virus* (VARV); *Volepox virus* (VPXV); *Alaskapox virus* (AKPV)

representative of two subfamilies and 21 genera (see Tables 1 and S1, available in the online version of this article). Genome sequences were retrieved from the NCBI database.

Nucleotide composition and principal component analysis

In order to determine the dinucleotide composition of poxvirus genomes, we applied the *compseq* tool (https://www.bioinformatics.nl/cgi-bin/emboss/). The observed-over-expected dinucleotide composition (O/E ratio) was defined as the quotient of the observed frequency of a specific dinucleotide divided by the product of the frequencies of the contributing nucleotides. The O/E CpG and TpA ratios were obtained by setting the size of word to 2, the '*calcfreq*' parameter, and by counting words also in the reverse complement of the sequence. For each viral species, CpG and TpA ratios were calculated for the whole genome, by concatenating all coding sequences, as well as for single genes longer than 100 nt. For concatenated sequences, overlapping ORF regions were masked.

The COUSIN (COdon Usage Similarity Index) tool [79] was applied to calculate the G+C content in first, second and third codon position. To estimate the relative synonymous codon usage (RCSU) we used the *seqinr* R package [80]. RSCU is a measure of non-uniform usage of synonymous codons and represents the number of times a particular codon is observed compared to uniform synonymous codon frequency.

Principal component analysis (PCA) was performed using a vector of all RSCU values calculated for concatenated coding sequences using the mixOmics R package [81].

Orthology inference

To investigate CpG distribution in the orthopox genus, we performed orthology inference using the OrthoFinder bioinformatic tool. OrthoFinder is able to recognize hierarchical OrthoGroups (OGs) as groups of genes that have descended from a common gene in the last common ancestor [82]. We only considered OGs present in all orthopoxviruses. OrthoFinder was run using MAFFT as an aligner and FastTree for tree inferences. We obtained a list of 118 orthologous genes, which were used for subsequent analyses (Table S2). The CpG O/E ratios of these genes were used to make a clustered heatmap using the *pheatmap* R package, applying the 'complete linkage' method as clustering algorithm, both for genes and for species.

Gene expression timing and levels

Gene expression levels of VACV were retrieved from a previous study [83]. We based our analyses on the authors' classification and we assigned each VACV (Western Reserve, WR) gene to one of three categories corresponding to their expression during the infection: two early clusters (E1.1 and E1.2) and a post-replicative cluster (PR). Transcript read counts were retrieved from the same study and normalized as suggested by the authors [83]. We performed a Nemenyi non-parametric all-pairs comparison test to verify the equality between the medians of each temporal expression category in relation to CpG gene values. Analyses were carried out in the R environment using the *PMCMRplus* package [84].

Phylogenetic trees, phylogenetic regression analysis and phylogenetic ANOVA

The poxvirus phylogenetic tree was generated using the concatenated alignment of the amino acid sequences of the 25 proteins that are conserved in all poxviruses (https://ictv.global/report_9th/dsDNA/poxviridae). The alignment was generated with MAFFT (v7.427) [85] with default parameters. The tree was generated with IQTREE (v1.6), with a partition per protein, and allowing each partition to have a distinct evolutionary model and a specific evolution rate (*-spp* parameter) [86, 87].

We used Pagel's λ using the BayesTrait software to evaluate the extent to which phylogenetic relatedness explains the variation in the distribution of the continuous trait G+C content values we calculated [88]. This tool allows us to evaluate whether a specific trait varies following the shared ancestry specified by a phylogenetic tree. In particular, we performed two Random Walk (Model A) Markov chain Monte Carlo (MCMC) analyses of the G+C content (log-transformed), by estimating λ or fixing it to 1 [88]. MCMC chains were run for 100 million iterations and the two models were compared by log Bayes Factor from a stepping-stone sampling.

A phylogenetic ANOVA was performed using the phylANOVA function in the *phytools* R package [89–91]. The *P*-value was calculated with 1000 simulations.

Identification of G+C content shifts

To estimate the occurrence of shifts in C+G content during poxvirus evolution, we ran the *l*1ou R package [92]. This tool estimates the number and location of shifts of a continuous trait along a phylogenetic tree, without any a priori assumptions, by using a phylogenetic LASSO (Least Absolute Shrinkage and Selection Operator) method under an Ornstein–Uhlenbeck process [92]. We ran *l*1ou by using an ultrametric version of the tree calculated for the 25 poxvirus conserved proteins, generated with the ape package [92] and the whole genome C+G content of the 53 viral genomes. The best shift configuration was defined by the phylogenetic Bayesian Information Criterion (pBIC), which is a conservative model when searching for adaptive shifts [92].

Substitution rates and recombination analysis

A set of representative C+G- and A+T-rich poxvirus species were selected. We analysed species for which a sufficient number of sequenced genomes were available in the NCBI virus database (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/): clade I cowpox virus (n=32), fowlpox virus (n=21), type I molluscum contagiosum virus (n=16), orf virus (n=23), crocodylepox virus (n=17), sheeppox virus (SPV, n=18) and goatpox virus (GPV, n=12). A list of all NCBI accession identifiers is reported in Table S3. Viral genomes within each viral species were aligned using MAFFT and for each species we selected as a reference strain the one with the oldest known sampling date. Then, for each genomic position, we counted the number of changes that were present in at least two sequences compared to the reference strain. Finally, we selected the total number of G/C to A/T changes and the number of A/T to G/C changes and, because the probability that a specific nucleotide mutates depends on its frequency in the genome, these counts were normalized by the number of the respective nucleotides in the reference genome.

The expected G+C content at equilibrium (G+Ceq) was estimated by dividing the G/C to A/T normalized rates by the sum of G/C to A/T and A/T to G/C rates.

Evidence of recombination events was analysed using ClonalFrameML v.1.12 [93]. ClonalFrameML requires a sequence alignment and a phylogenetic tree as input files. We used the whole genome alignments of the individual viral species mentioned above (intraspecies analysis), the trees generated by IQTREE and the tree mean branch lengths as priors for the M parameter. R/theta (relative rate of recombination to mutation), 1/delta (inverse mean DNA import length) and nu (mean divergence of imported DNA) were estimated.

RESULTS

CpG depletion is limited and varies across and within viral genomes

We have assembled a dataset of 53 complete poxvirus genomes (Table 1). These correspond to all exemplar strains from the ICTV metadata resource (https://ictv.global/vmr), plus the recently described Alaskapox virus. The viruses in the final dataset belong to two subfamilies, *Entomopoxvirinae* (insect-infecting viruses, *n*=8) and *Chordopoxvirinae* (vertebrate-infecting viruses, *n*=45) (Table S1).

To investigate CpG and TpA biases, we calculated the O/E ratio for each dinucleotide, where the expected frequency in a genome is the product of the frequencies of the corresponding nucleotides. Ratios above or below 1 indicate that the dinucleotide is over- or under-represented, respectively, and ratios below 0.78 and above 1.25 are generally considered to define significant depletion and enrichment [55, 94]. The distribution of O/E CpG ratios was very wide both in entomopoxviruses and in chordopoxviruses, while the range for distribution of O/E ratios for TpA was narrower (Fig. 1a). Overall, O/E ratios for CpG and TpA dinucleotides did not deviate from the null expectation, with the exception of a number of chordopoxviruses that displayed significantly lower TpA content than expected, and of a number of entomopoxviruses displaying significantly higher CpG content than expected (Fig. 1a). Previous analyses of vertebrate genomes as well for vertebrate-infecting viruses identified a covariation between genomic G+C content and specific dinucleotide frequencies. Specifically, G+C-rich genomes tended to display a high O/E ratio for CpG dinucleotides, while A+T-rich genomes tended to display a high O/E ratio for TpA [48, 51, 95, 96]. We thus tested whether variation in G+C content correlated with variation of O/E ratios for CpG and for TpA dinucleotides in our poxvirus dataset. Entomopoxviruses presented G+C-poor genomes with low dispersion for overall G+C content (median: 0.197, range: 0.177–0.300), a number of them being enriched in CpG nucleotides above the null expectation, but also with two being depleted (Fig. 1b). Conversely, chordopoxviruses covered a wider range and displayed a bimodal distribution of G+C genomic values (median: 0.328, range: 0.235–0.666). Significant correlations between genomic G+C content and O/E ratios for CpG and TpA were detected for both dinucleotides. Such correlations were primarily driven by G+C-richer chordopoxviruses, which tended to be enriched in CpG dinucleotides, and were significantly depleted in TpA dinucleotides (Fig. 1b).

To gain further insight into poxvirus dinucleotide composition, we calculated O/E ratios for CpG and TpA dinucleotides for all viral ORFs longer than 100 nt. These fine-grain ratio distributions were overall consistent with the results obtained for the corresponding whole genomes. Specifically, the O/E ratios for CpG displayed again a much wider variation than the O/E ratios for TpA, both across and within viral genomes (Fig. 2).

Dinucleotide biases partially depend on gene expression timing

Orthopoxviruses have historically been the subject of detailed investigation compared to poxviruses in other genera because they include several viruses highly pathogenic for humans. We thus exploited the available knowledge to investigate the functional determinants of within-genome variation in CpG content.

Orthopoxviruses are divided into two sister clades: the Eurasian/African orthopoxviruses (Old World orthopoxviruses) and the North American orthopoxviruses (New World orthopoxviruses), with the still unclassified Alaska virus forming a distinct branch located between the these two major clades [78, 97]. Genome composition in orthopoxviruses has evolved through lineage-specific episodes of gene duplication, capture and loss, with extant genomes displaying gene content values between 164 (VARV) and 208 (CPV) [19]. To perform meaningful comparisons, we identified the core gene set of orthopoxviruses, i.e. the largest set of orthologous genes that are present in all extant orthopoxviruses genomes. We identified a core set of 118 orthologous genes shared by all orthopoxviruses (detailed in Table S2). Paired comparison detected no significant differences in O/E ratios for CpG between core and non-core genes within a given genome (Fig. S1). Using the core genes set we next compared O/E CpG among orthopoxviruses by generating a heatmap with dendrograms using all orthologues (Fig. 3). The O/E ratio for CpG was much more similar across genes (columns) than across genomes (rows). The dendrogram for the viral clustering nicely recapitulated the phylogenetic relationships among orthopoxviruses, with New World and Old World viruses being clearly separated. In fact, New World orthopoxviruses tended to have lower O/E CpG than Old World viruses, and this was particularly evident for two clusters of genes that showed values >1 in several Old World viruses and <1 in most New World viruses (Fig. 3). The dendrogram for the gene clustering did not show any clear association between clusters of core genes and the functional annotation (Fig. 3).

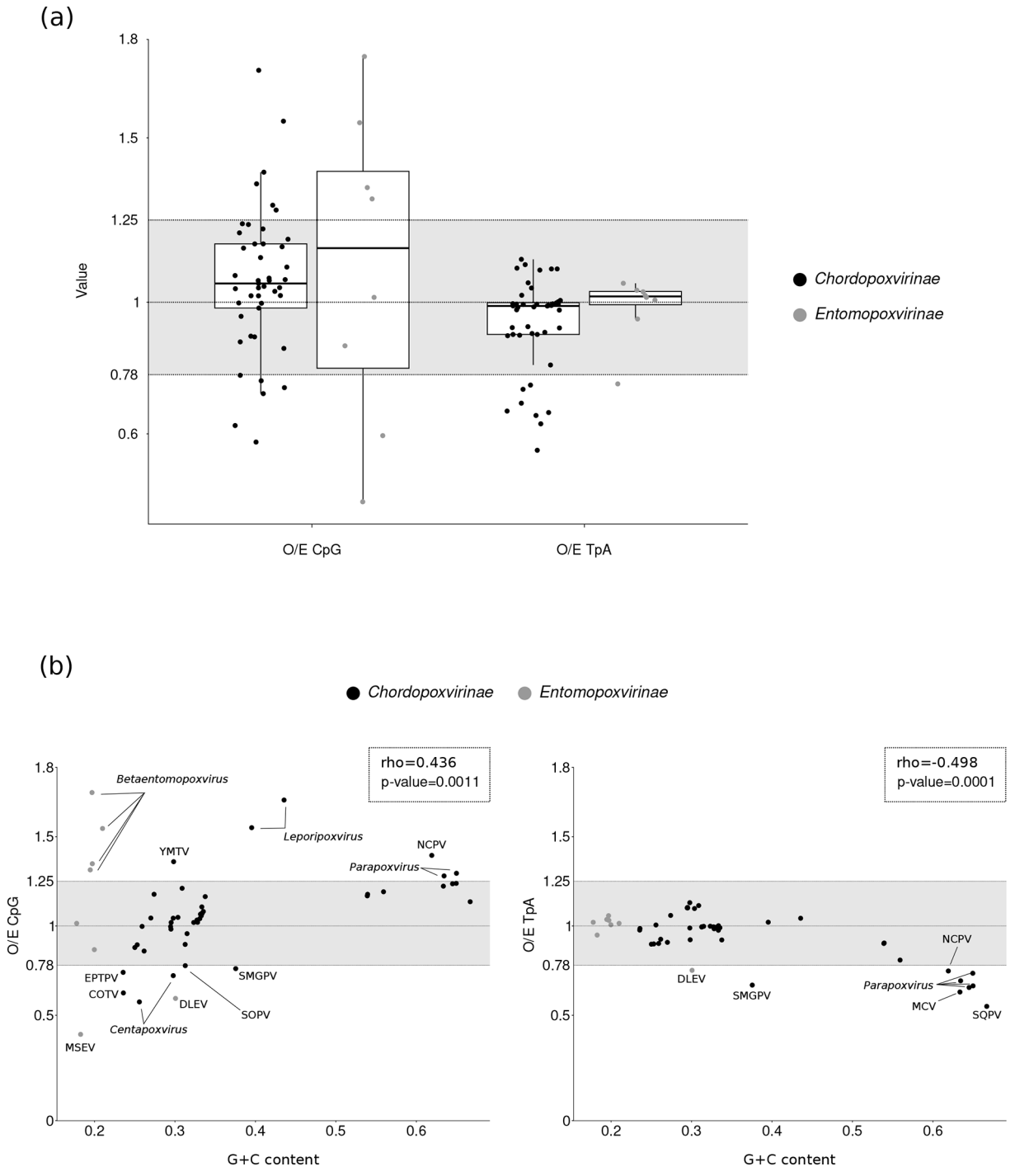

Fig. 1. CpG and TpA dinucleotide content in poxviruses. (a) Boxplots (median, first and third quartiles) representing values of CpG and TpA O/E ratios (observed frequency/expected frequency) of a set of chordopoxvirus (n=45) and entomopoxvirus (n=8) reference genomes. The area in light grey corresponds to the expected range of O/E ratios. (b) Correlation between G+C content and O/E CpG or TpA values in poxvirus genomes. Viruses with over- or under-represented dinucleotides (CpG and TpA) are labelled. Spearman's correlation coefficient (rho) and *P*-value are also shown.

To provide a finer evaluation of the functional implications of O/E CpG and TpA ratios at the gene level, we resorted to focusing the study on VACV. VACV is the type species in the genus *Orthopoxvirus* (https://ictv.global/report_9th/dsDNA/poxviridae). This virus has been intensely investigated and is regarded as a model to study poxvirus biology. Because the functional role of several VACV proteins has been elucidated, we tested whether, more generally, CpG representation was influenced by gene function. In line with the results obtained in the comparison among orthpoxviruses, no significant difference was observed among functional

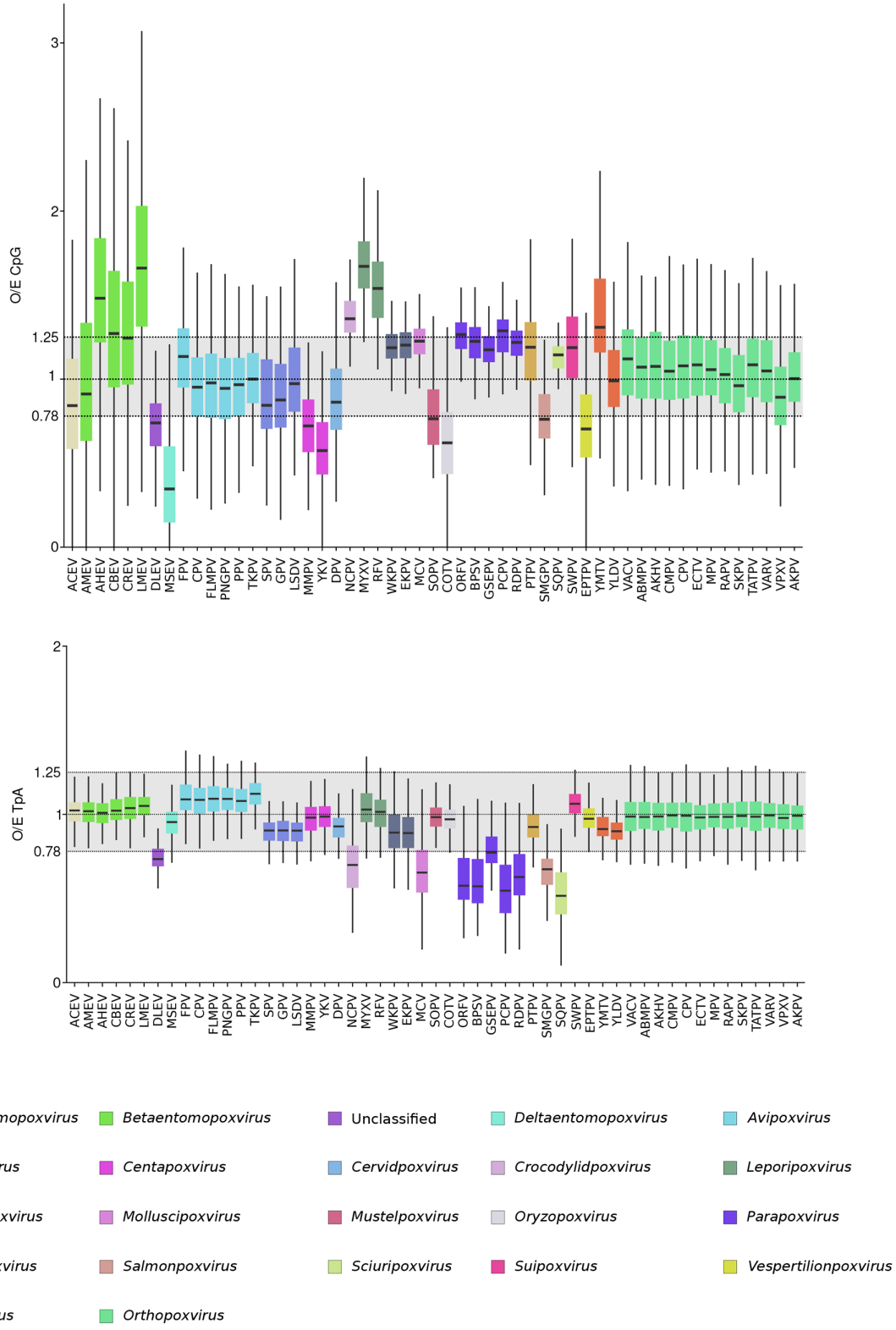

Fig. 2. CpG and TpA dinucleotide content in poxvirus genes. Boxplots (median, first and third quartiles) representing values of O/E CpG and TpA levels calculated for each viral gene. Boxplots are coloured according to each genus in the family *Poxviridae*. The area in light grey corresponds to the expected range of O/E ratios. See Table 1 for a list of virus abbreviations.

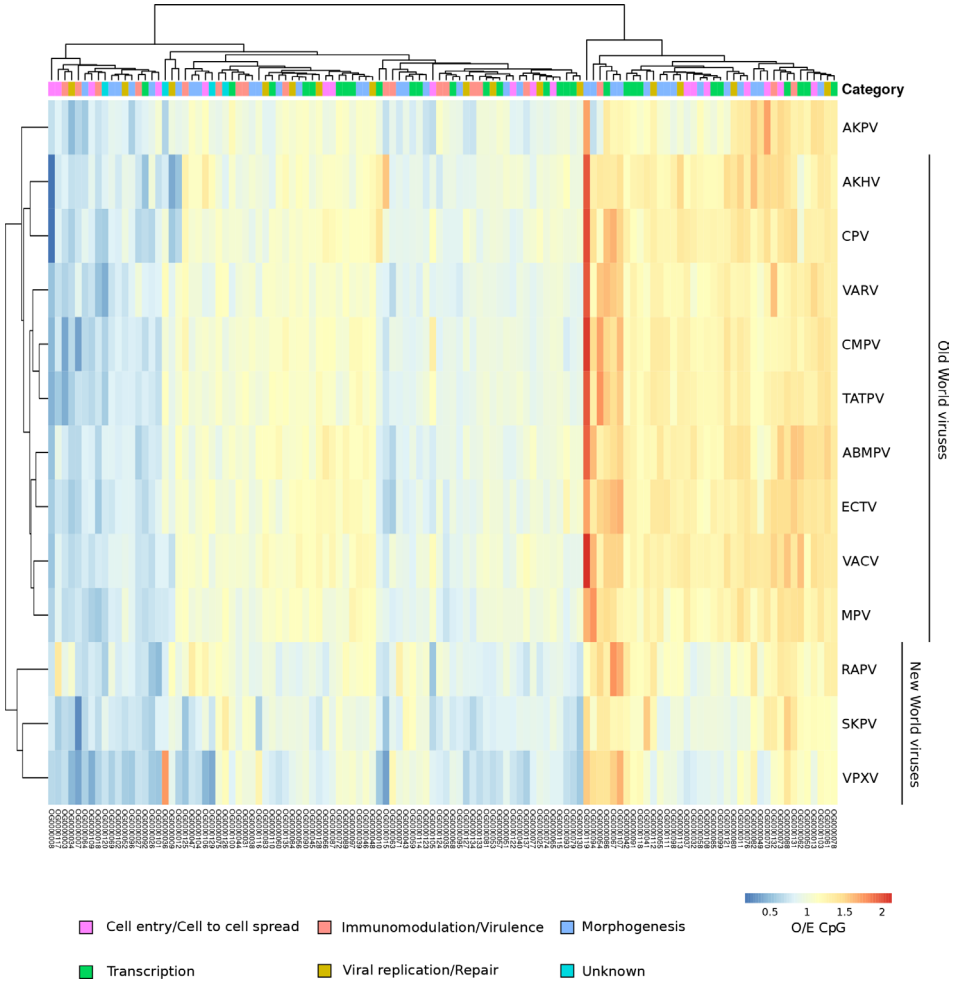

Fig. 3. CpG distribution among orthopoxvirus core genes. Heatmap plot showing the distribution O/E CpG values calculated for 118 orthopoxvirus core genes (see also Table S2). Columns represent orthogroups and rows indicate viral species. Both column and row clustering were applied and coloured bars on the top represent the functional category of each orthogroup. See Table 1 for a list of virus abbreviations.

categories (Fig.S2). As this gene function level analysis did not provide any insight into dinucleotide variation between genes, we resorted to analysing gene expression timing data for VACV genes. Genes in the VACV genome can be classified depending on the variation in their expression levels across the first stages of viral infection, with two clusters of early genes (E1.1 and E1.2) and a cluster of post-replicative genes (PR) [83]. We used these data to infer whether O/E CpG ratios differ between genes as a function of their transcription timing. Our results indicated that O/E CpG ratios in E1.1 genes were significantly lower than for the other two categories of genes expressed at later time points (Fig. 4).

Codon usage bias is strongly influenced by genomic G+C content

Variations in dinucleotide composition and G+C content are tightly related to CUB [25, 42–44]. As a proxy for CUB, we calculated the RSCU for each viral genome, so that each of the 53 poxviruses was represented by a vector of 59 positions, corresponding to the codons encoding the 18 amino acids with synonymous coding. We performed then a PCA to reduce dimensionality on the 59-dimension RSCU vector. Variation along the first component captured a very large portion of the total RSCU variance (87%), with the second component only accounting for 5%. The 53 viruses were distributed along the two components with most entomopoxviruses, excluding DLEV (*Diachasmimorpha entomopoxvirus*), tending to group together (Fig. 5a). Among chordopoxviruses, a relatively good stratification as a function of viral taxonomy was observed. Conversely, vertebrate-infecting poxviruses showed very limited clustering by host species (Fig. 5b). Overall, these results suggest that codon usage in poxviruses is at least partially driven by phylogenetic relationships among viruses, while host taxonomy did not provide any explanatory power for variation in viral CUB.

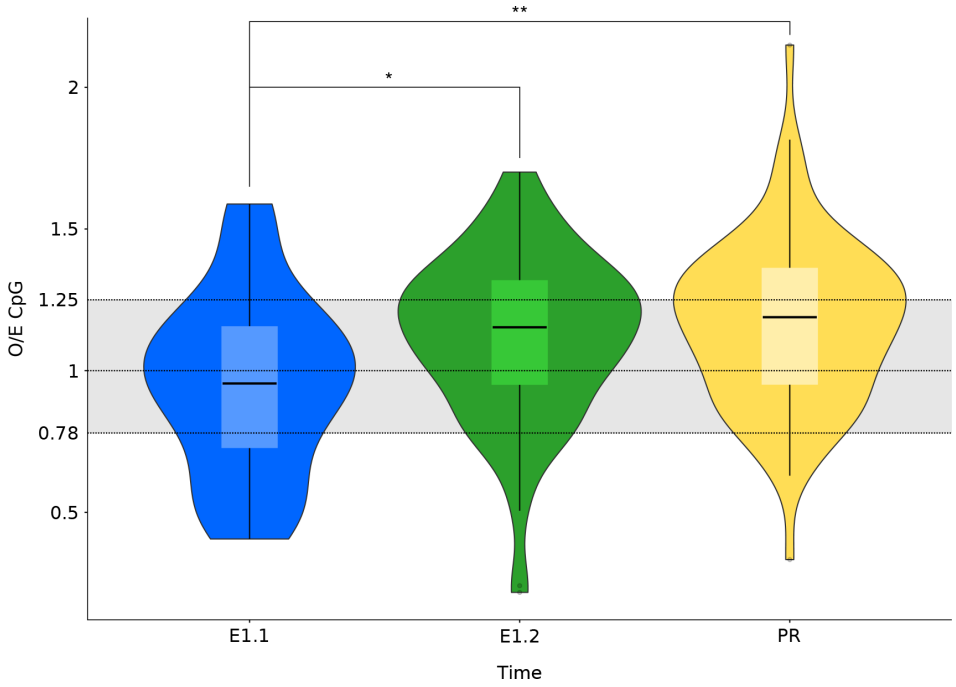

Fig. 4. O/E CpG values of vaccinia virus genes. Violin plot of O/E CpG values for VACV genes grouped by their expression time. E1.1 indicates the earliest expression time, E1.2 a slightly later one and PR to a post-replicative cluster of genes. The area in light grey corresponds to the expected range of O/E ratios.

In viruses, CUB can be strongly influenced by overall G+C content [25, 42, 44, 98]. This is indeed the case in poxviruses, as we found an extremely strong correlation between variation along the first component of the PCA (PC1) and the variation in G+C content in the third codon position (GC3) (Fig. 5c and S3). Likewise, GC12, the average value of G+C content in the first and second position of the codons, was strongly correlated to PC1 (Fig. 5c). Overall, these results suggest that genomic G+C content strongly shapes CUB.

To determine whether CpG dinucleotides influence codon usage in poxviruses, beyond the contribution of overall G+C content, we correlated variation along the second principal component with variation in O/E ratios for CpG. No significant correlation was detected (Spearman's rank correlation, rho=0.07, P=0.59). We also compared the codon frequencies among pairs of synonymous codons that contain or lack a CpG (i.e. we calculated the XCpA/XCpG ratio for Ser, Pro, Thr and Ala codons). Because this ratio is strongly influenced by G+C content, this value was compared with the corresponding XApA/XApG ratio (calculated for Glu, Lys and Gln). A tendency to avoid CpG-ending codons was only observed for viruses showing an overall CpG depletion (melanoplus sanguinipes entomopoxvirus, DLEV, centapoxviruses, salmon gillpox virus, cotia virus and sea otterpox virus) (Fig. 6). Overall, these results confirm that CpG avoidance is not a general feature of poxviruses.

Genomic base composition is mainly driven by mutation biases

The analyses above showed that G+C content is a major determinant of CUB. We thus examined whether a phylogenetic signal was detectable in the distribution of G+C content across the poxvirus phylogeny. For this we generated a phylogenetic tree using the sequences of 25 proteins that are conserved in all poxviruses (https://ictv.global/report_9th/dsDNA/poxviridae) and we calculated Pagel's λ as a proxy for estimating the extent for which variation in a continuous trait (here G+C content) is accounted for by the phylogenetic relatedness between the terminal taxa in the tree [99]. Our estimate of λ was not different from 1 (λ=0.999, Log Bayes Factor=0.22), indicating a very strong phylogenetic signal.

We next assessed whether shifts in genomic base composition occurred during poxvirus evolution. For this, we fitted variation in G+C content as a continuous trait using a phylogenetic lasso method (ℓ1ou) [92]. This method tests for shifts of quantitative traits across a phylogeny and estimates whether jumps or discontinuous events have occurred and in which ancestral nodes of the tree such events are more likely to have taken place. We selected the best-fit shift configuration using pBIC. The ℓ1ou algorithm estimated that five shifts, all of them with strong bootstrap support, occurred during poxvirus evolutionary history (Fig. 7a). One basal shift (labelled 1^+ in Fig. 7a) towards increasing G+C content would have occurred after the split of tetrapod-infecting poxviruses from salmon gill poxvirus. Two more recent shifts towards decreased G+C levels were detected, one (labelled 2^-) that

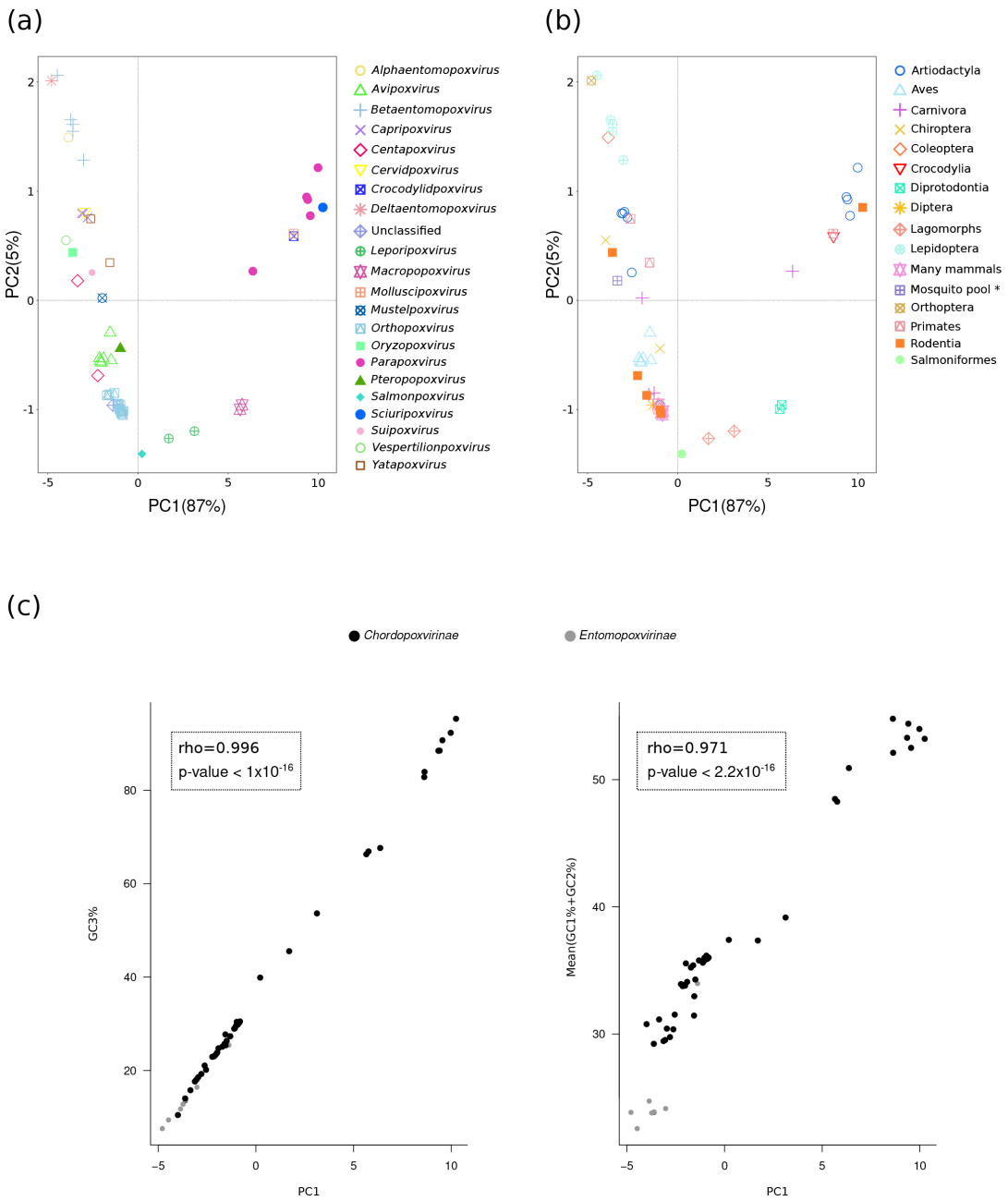

Fig. 5. PCA of poxvirus coding sequences. (a) PCA based on RSCU of the concatenated coding sequence of each poxvirus reference genome. Each virus is labelled by its genus, as shown in the key (DLEV is labelled as unclassified). (b) PCA based on RSCU of the concatenated coding sequence of each poxvirus reference genome. Each virus is labelled by host species, as shown in the key. The asterisk indicates that the natural host is unclear. (c) Correlations between G+C content at third codon positions (GC3) and the first principal component (PC1) (left panel); correlation between the mean value of G+C content at first and second codon positions (GC12) and PC1 (right panel). Viruses are divided by subfamilies, as shown in the key. Spearman's correlation coefficient (rho) and its relative *P*-value are also shown.

would constitute a synapomorphy of the avipox branch and another one (3^-) on a branch leading to several poxviruses, to which MCV and squirrelpox virus splits are basal. This last shift would have been followed by two independent events towards increase G+C content in parapoxviruses (4^+) and in leporipoxviruses (5^+) (Fig. 7a).

The wide distribution in G+C content in poxvirus genomes, together with the frequent switches proposed above, open the question of the underlying molecular mechanisms driving nucleotide composition in these viruses. One possibility is that the main force driving G+C content variation in poxviruses is mutation bias (neutralist hypothesis). This would be in line with the

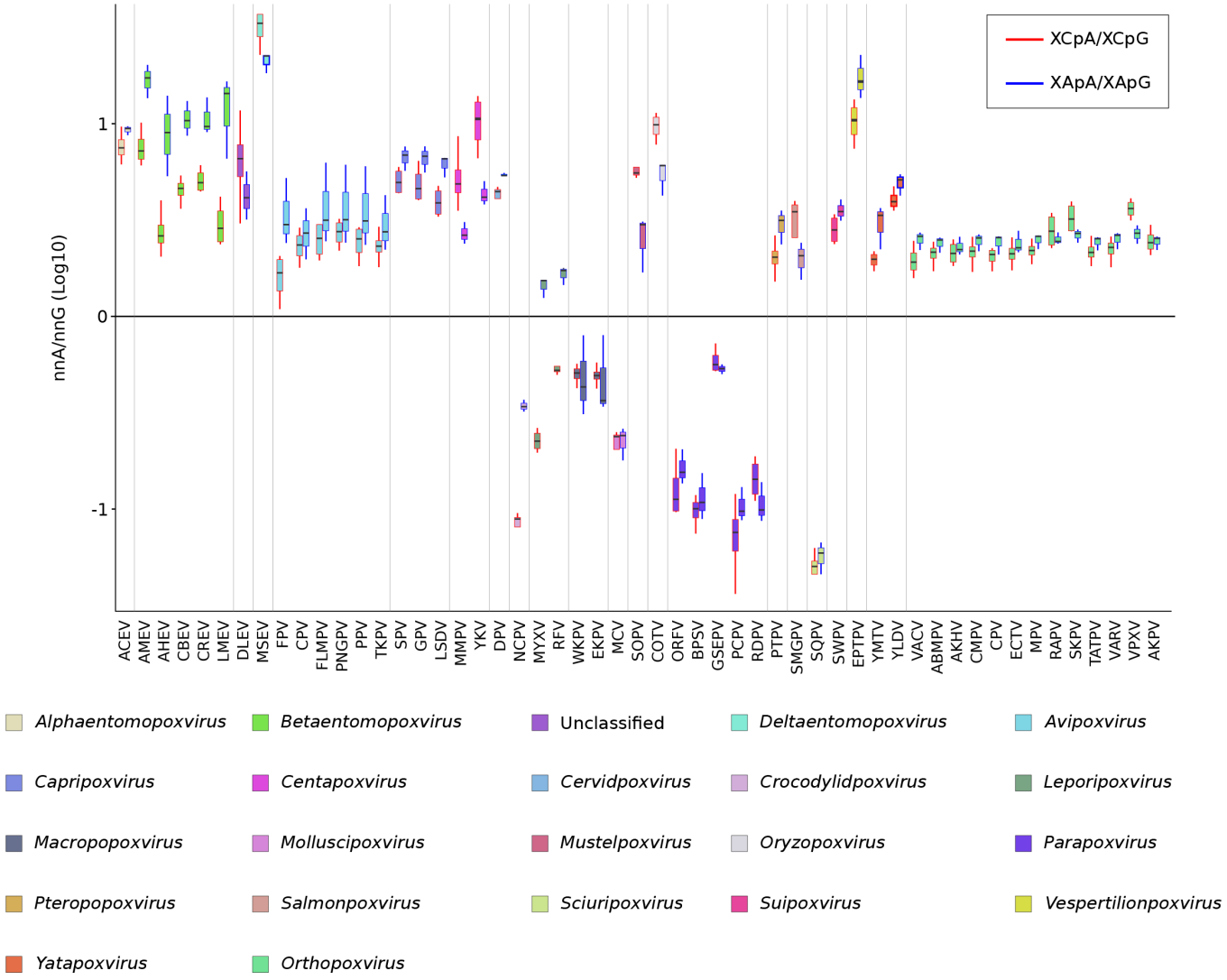

Fig. 6. Codon usage in poxviruses. Boxplot (median, first and third quartiles) representation of the ratio of synonymous codon frequency carrying or not carrying a CpG dinucleotide. The red profile indicates the $\log_{10}$ ratio between XCpA and XCpG codons (calculated for Ser, Pro, Thr and Ala); the blue profile indicates the ratio between XApA and XApG codons (calculated for Glu, Lys and Gln). Boxes are coloured by genus, as shown in the key. Vertical lines separate viral genera. See Table 1 for a list of virus abbreviations.

observation that the G+C content at sites that are likely to be evolving close to neutrality (i.e. third codon positions) shows even more extreme variation among genomes than the average gene G+C content (Fig. 7b). Nonetheless, variations in GC3 might also result from selection for G+C composition (selectionist hypothesis). Under the neutralist hypothesis, in genera or species with G+C-rich genomes, mutation would have been consistently biased towards G/C, while the opposite has occurred in A+T-rich genera or species. Conversely, if selection had played a role in driving G+C content, the probabilities of fixation of A/T→G/C and G/C→T/A mutations would differ, because they confer a differential fitness advantage/disadvantage.

One possibility to choose between these competing explanatory hypotheses is to analyse changes that have been accumulating within species rather than between species. The reason is that, the smaller the fitness effect of a mutation, the weaker the power of natural selection to modulate its frequency in a population over short evolutionary timeframes, so that substitutions at shallow evolutionary levels are more likely to reflect the mutation patterns rather than to result from natural selection [100, 101]. For this, we selected viruses at the extremes of the G+C content range for which a sufficient number of extant sequenced genomes were available. We thus analysed clade I cowpox virus (CPV, *n*=32), fowlpox virus (FPV, *n*=21), type I molluscum contagiosum virus (MCV, *n*=16), orf virus (ORFV, *n*=23), Nile crocodilepox virus (NCPV, *n*=17), sheeppox virus (SPV, *n*=18) and goatpox virus (GPV, *n*=12). MPV and VARV were not included because their mutation spectrum has shown to be largely determined by the mutator activity of human APOBEC3 enzymes [102–105]. VACV and myxoma virus were also excluded because of their peculiar

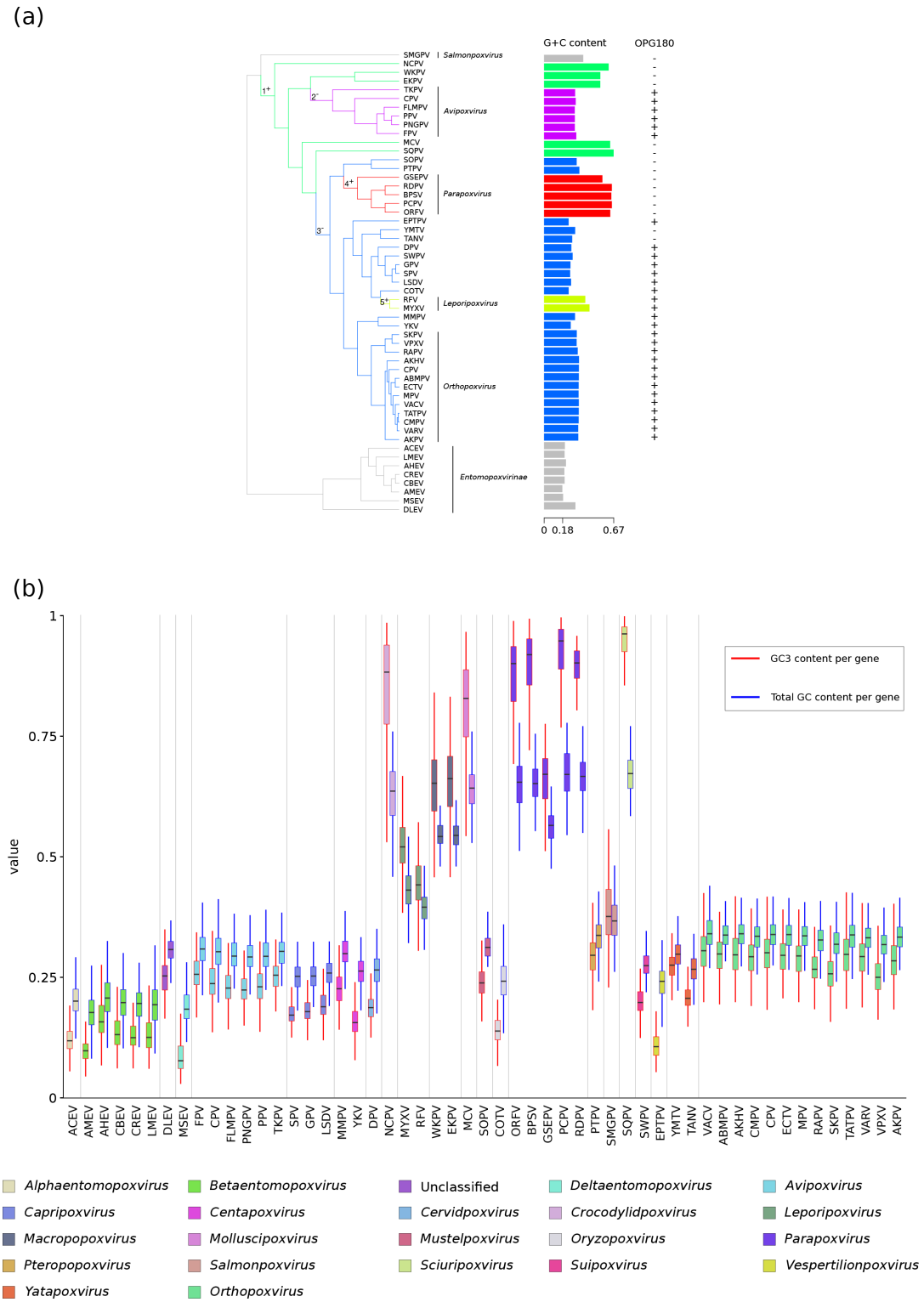

Fig. 7. C+G content changes during *Poxviridae* evolution. (a) G+C content shifts along the *Poxviridae* phylogenetic tree. Each shift is labelled with a number and the signs + (indicating an increase in G+C content) or – (indicating a decrease in G+C content). The phylogenetic tree was built using 25 proteins that are conserved in all poxviruses. Levels of genome-wide G+C content are reported as bars. The presence/absence of OPG180 in chordopoxviruses is also reported. (b) Boxplot (median, first and third quartiles) representation of the C+G content both genome-wide (blue profile) and in third codon positions only (red profile) for all poxvirus genes. Boxes are coloured by genus, as shown in the key. Vertical lines separate viral genera. See Table 1 for a list of virus abbreviations.

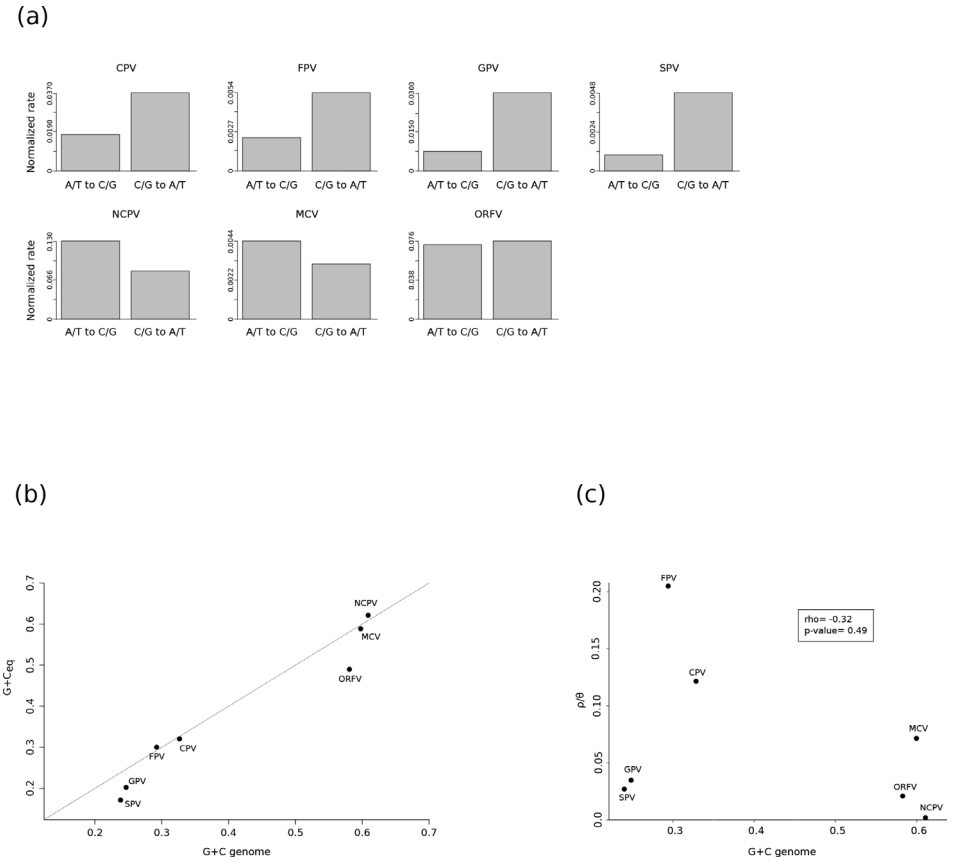

Fig. 8. Mutation frequencies in poxviruses. (a) Normalized counts of G/C to A/T mutations and of A/T to G/C mutations for a set of poxviruses with G+C-rich or G+C-poor genomes. Substitutions were counted if present in at least two viral strains and normalized by the count of the respective nucleotides (see Methods for more details). (b) Correlation plot between the G+C content calculated genome-wide and the expected equilibrium G+C content (G+Ceq), calculated based on the substitution spectrum. (c) Correlation plot between the genome-wide G+C content and the relative rate of recombination to mutation calculated by ClonalFrameML. Spearman's correlation coefficient (rho) and *P*-value are also shown.

evolutionary histories, including the iatrogenic vaccine pressure for VACV and the strong human-driven geographical bottleneck events for myxoma virus [106, 107]. Viral genomes were aligned and we counted the number of G/C→A/T mutations and the number of A/T→G/C mutations. Because the probability that a specific nucleotide mutates also depends on its frequency in the genome, counts were normalized by the number of G/C or A/T nucleotides. In G+C-poor viruses (CPV, FPV, GPV and SPV) the rate of G/C→A/T mutations (rate G/C→A/T) was higher than the rate of A/T→C/G mutations (rate A/T→G/C). The opposite was true for G+C-rich viruses (NCPV and MCV), while ORFV, also G+C-rich, displayed comparable rates (Fig. 8a). We next calculated the expected equilibrium G+C content (G+Ceq) based on the inferred mutational rates. The results indicated that all viruses, with the exclusion of ORFV, have overall G+C content values very close to their anticipated G+Ceq (Fig. 8b). Overall, these results suggest that nucleotide content in poxviruses is primarily determined by mutational biases.

Mutation biases can arise through different mechanisms, including polymerase fidelity, proofreading activity or post-replicative repair [108]. The poxvirus replicative holoenzyme is highly conserved and possesses both 5′-to-3′ DNA polymerase activity and 3′-to-5′ proofreading exonuclease activity [109]. We thus reasoned that the observed mutation biases are unlikely to arise from lineage-specific DNA polymerase bias, and that they might instead derive, at least partially, from repair mechanisms. We thus focused on chordopoxviruses, whose biology has been studied in greater detail compared to entomopoxviruses, and searched the literature for genes involved in genome repair mechanisms. Five such genes were identified, which encode the viral primase (OPG117), the UDP glycosylase (OPG116), the resolvase (OPG149), the nuclease (OPG089) and the ATP-dependent DNA ligase (OPG180) [5, 110–112]. Whereas the former four are present in the genomes of all chordopoxviruses, this is not the case for the DNA ligase, which is a non-essential gene [113–115]. Analysis of DNA ligase presence/absence on the chordopoxvirus phylogenetic tree indicated that most genomes lacking the gene are G+C-rich (Fig. 7a). To formally test for an association between DNA ligase presence/absence and G+C content, we applied a phylogenetic ANOVA [89]. The results indicated that, after accounting

for phylogenetic relationships, genomes encoding the DNA ligase have significantly lower G+C content than genomes with no OPG180 gene (pairwise *t*-values=−7.06, *F*=49.86, *P*=0.007).

Finally, in cellular organisms genomic G+C content has been related to recombination rates [96, 116–119]. We thus addressed whether this is the case in poxviruses. For this we used ClonalFrameML to calculate recombination/mutation rates for the seven viruses analysed above. No evident relationship between recombination rates and G+C content was detected (Fig. 8c).

DISCUSSION

Animal genomes differ widely in G+C content, dinucleotide composition and CUB, both when comparing genomes from different species and when comparing genes within the same genome [102, 120]. Examples at the genomic level are the enrichment in G/C--ending codons in mammal genomes, the global trend in vertebrate genomes to display a lower than expected CpG dinucleotide frequency or the under-representation of the TpA dinucleotide in all animal genomes [120, 121]. The global framework for understanding compositional biases is the mutation–selection–gene conversion–drift hypothesis, albeit it is not always evident to pinpoint the actual differences between neutralist and selectionist approaches, beyond over-simplifications [122]. For our focus here on dinucleotide frequencies, a mutational explanation for the CpG under-representation posits that cytosine methylation, common in vertebrate genomes, increases C→U→T mutation rate, although this process does not explain the CpG depletion observed also in vertebrate mitochondrial genomes [47, 49, 55, 120]. On the other hand, the selectionist explanation for the CpG under-representation posits that it may be the result of the impact of this dinucleotide on transcription and translation [56, 57, 123]. In the case of TpA dinucleotides (UpA in RNA), the depletion observed across the entire domain of animal genomes may result from the preferential cleavage of UpA dinucleotides by cytosolic RNAses [54].

Just like their cellular hosts, viral genomes also display nucleotide, dinucleotide and CUB biases, which do not necessarily match those of the corresponding host(s) genomes, and are the result of the same mutation–selection–drift processes [26]. Irrespective of their ultimate origin, differences in compositional biases between viruses and their hosts can be exploited by cellular innate immune sensing systems evolved to detect non-self nucleic acids. Antiviral responses can be mediated by overall differences between virus and host genomes, as is the case of the role of SCHLAFEN11 in preventing translation of viral mRNAs, often A+T-richer than mammalian host transcripts [124, 125]. The same signature of differential composition allows RNAse L to preferentially degrade A+T-rich viral mRNAs [or viral genomes, in the case of RNA(+) viruses] [62, 68]. The CpG and TpA dinucleotides constitute a particular target of the viral–host immune interaction interface. In vertebrates, the protein ZAP mediates the recognition of CpG- and UpA-rich RNAs, whereas the intracellular receptor TLR9 can selectively trigger an innate immune response in the presence of cytoplasmic DNA containing unmethylated CpGs, often associated with the presence of DNA viruses replicating in the cytoplasm, such as poxviruses [58, 67]. We have thus tried to understand differences in viral dinucleotide biases in general and in poxviruses in particular, as a possible result of these viral-restricting mechanisms. It is important to note first that neither ZAP, nor TLR9 nor any other system specifically recognizing CpG-rich nucleic acids are known in invertebrates, and indeed, invertebrate genomes do not display signatures of CpG depletion [120]. Dinucleotide composition has been intensely investigated for vertebrate RNA viruses, many of which display CpG-depleted genomes [51, 72, 77, 126–128]. In the genomes of small animal dsDNA viruses, such as polyomaviruses and papillomaviruses, CpG dinucleotides are under-represented [129]. However, these viruses are not known to be restricted by ZAP, and their dinucleotide composition can instead be partly accounted for by methylation avoidance and/or by avoidance of TLR9 recognition [63, 130, 131]. In herpesviruses, large dsDNA viruses with genomes comparable in size to those of poxviruses, genetic diversity is very large and so is the diversity of genomic composition biases: alphaherpesviruses do not show CpG depletion, whereas many gammaherpesviruses do, and in betaherpesviruses, specific depletion of CpG is observed in immediate early (IE) genes only [69, 132].

Poxviruses, with their host spectrum ranging from insects to mammals, offer an interesting model to study dinucleotide biases and genome composition in general. Our analysis of dinucleotide composition showed no global tendency towards CpG or TpA depletion. Interestingly, after correcting for G+C content, we found that two of the seven poxvirus genomes showing the lowest CpG content belonged to entomopoxviruses. Moreover, a third CpG-depleted genome was that of salmon gill poxvirus. Because ZAP evolved in tetrapods, it cannot be responsible for shaping the genome composition of entomopoxviruses nor of this virus infecting fishes. All this evidence clearly indicates that selection for avoiding ZAP-mediated restriction is not the main determinant of CpG content in poxviruses.

In our analyses, most chordopoxviruses that infect tetrapods showed very little signal of CpG depletion, if any. It has been proposed that the protein product of the OPG005 viral gene, the C16 protein, can antagonize ZAP by sequestering it to cytoplasmic punctate structures [71]. We thus checked whether loss of the OPG005 gene might explain the low CpG content observed in a few chordopoxvirus genomes (eptesipox virus, cotia virus, sea otterpox virus, murmansk microtuspox virus and yokapox virus). Functional orthologues of C16 were detected in all these poxviruses except sea otterpox virus, suggesting that loss of the ZAP antagonist does not explain CpG depletion in most of these genomes. Further along this line, the modified vaccinia virus Ankara contains a disrupted OPG005 gene, and this virus has been shown to be restricted by ZAP [71]. However, in this derived VACV strain, the mechanism underlying ZAP-mediated restriction is unlikely to be mediated by RNA binding, as ZAP has no effect

on viral DNA, mRNA or protein abundance. Instead, in this viral system, ZAP has been reported to interfere with the viral life cycle at the late steps of virion maturation [71].

Despite the limited evidence of overall CpG depletion in poxvirus genomes, individual poxvirus genes display a wide variability of CpG content. As a comparison, the variation of TpA is much narrower. This is reminiscent of observations in betaherpesviruses, whereby only IE genes are CpG-depleted [69, 132]. Taking advantage of the large body of functional knowledge available for VACV as a model, we show that genes that are expressed at the earliest times in infection present lower O/E CpG ratios than genes expressed at later infection stages. This observation is in line with data reporting that ZAP specifically destabilizes the transcripts of genes expressed at the early time points of HCMV (a betaherpesvirus) infection [70]. A possible explanation for these observations is that, as is the case for HCMV, the ZAP-mediated innate immune response against poxviruses is strongest in the early phases of infection [70], when the viral antagonist (the C16 protein) may be unable to fully block its action. Regardless, it is also worth noting that only a fraction of E1.1 genes in VACV are actually more CpG-depleted than expected, given their composition. As in the case of HCMV, this might be due to the specific targeting by ZAP of a minority of viral transcripts [70]. An alternative, non-mutually exclusive possibility is that CpG dinucleotides are not the only targets of ZAP recognition. Indeed, during HCMV infection, ZAP binds both viral and cellular RNAs in regions showing an elevated frequency of cytosines [70].

In many RNA viruses, CpG depletion also has an effect on CUB. Given the limited tendency to deplete CpG dinucleotides, this is not the case with poxviruses. Also, in these viruses, translational selection (the non-random usage of synonymous codons to optimize translation efficiency) seems to be weak, as no clustering between virus CUB and host taxonomy was observed in the PCA. Indeed, our results are consistent with poxvirus CUB being mainly determined by genomic G+C content.

The range for genomic G+C content is very large in poxviruses. Whereas entomopoxviruses tend to be G+C-poor, chordopoxviruses are highly variable. For instance, two human-specific viruses, VARV and MCV, display large variation in G+C content values, 0.327 and 0.633, respectively. Moreover, we found that G+C content at synonymous and at non-synonymous sites shows highly correlated patterns of variation. Overall, these lines of evidence indicate that CUB in poxviruses is largely determined by overall G+C content, and that the forces shaping G+C content operate at the genome-wide level and over relatively long time periods.

When introducing a phylogenetic perspective into the distribution of G+C content in poxviruses, we found a very strong phylogenetic inertia, and we identified at least five shifts in G+C content that occurred during the evolutionary history of poxviruses. In principle, such changes in G+C content can be due to different forces, including mutation biases and natural selection. To disentangle the effect of these two forces, we analysed within-species diversity so as to capture changes that have occurred during short-term evolutionary periods and are thus less impacted by natural selection. The results clearly indicate that variation in G+C content was largely the result of mutation biases and that most poxviruses are close to the expected G+C content at equilibrium. The only exception to this trend was ORFV (genus *Parapoxvirus*), which infects different *Bovidae* species and can be zoonotically transmitted to humans. The reason(s) why ORFV showed similar A/T→G/C and G/C→A/T mutation rates remain to be clarified and, unfortunately, other viruses in the same genus were insufficiently sampled to be studied and to provide a comparative, paired analysis.

Because poxviruses replicate in the cytoplasm of the infected cells, they encode all or most of the enzymes necessary to duplicate their large genomes [5]. Repair of DNA damage is also largely dependent on viral proteins, most of which are highly conserved at least in chordopoxviruses. Conversely, the ATP-dependent DNA ligase (OPG180) is encoded by a non-essential gene [5]. Despite its dispensability, chordopoxvirus mutants lacking the DNA ligase display enhanced sensitivity to DNA damage caused by UV irradiation or chemical inducers of double strand breaks [110, 115, 133]. It is thus conceivable that loss of the viral ligase impairs the repair of specific lesions, eventually leading to biased mutation patterns, as our data suggest. On the one hand, this hypothesis is further supported by data from bacteria, which also exhibit wide variations in genomic G+C content that are largely dependent on the presence of specific DNA repair systems [134–136]. On the other hand, despite the fact that poxvirus replication occurs in the cytoplasm, cellular DNA ligases have been shown to be recruited to viral replication sites, where they may compensate or replace the activity of the viral enzymes [137, 138]. It has nevertheless been noted that quiescent cells express low ligase levels, possibly suggesting that viral tropism interacts with the presence/absence of a viral ligase gene to modulate mutation biases. Also, viral tissue tropism, transmission mode and host preferences are likely to determine exposure to different environmental mutagens. Thus, further insight into the role of DNA damage and repair in mutation biases will require experimental analysis on poxviruses lacking or expressing a viral DNA ligase and grown in different cell types and/or under different conditions.

In summary, our results indicate that in poxviruses global genomic nucleotide composition is the largest determinant of dinucleotide content (specifically for CpG and TpA dinucleotides) as well as of CUB, and that mutational bias is the leading force shaping genome nucleotide content. This conclusion does not exclude that host immune responses may contribute to the shaping of genome composition in poxviruses. For instance, recent evidence indicates that the evolutionary history of MPV during the global mpox outbreak, as well as the evolution of other human-infecting poxviruses, has been accompanied by APOBEC3-mediated editing (which introduces TC→TT changes and contributes thus directionally to the viral mutational spectrum) [102, 105, 139]. Also, even if interaction with the host immune system has not patterned the genomic composition in poxviruses, it has probably

exerted a selective pressure driving positive selection at individual sites [140, 141]. In this case, however, the signatures of selection are localized in specific positions and will not affect the overall nucleotide or dinucleotide genome composition.

Our study has a number of limitations. First, we restricted analysis to representative poxviruses with complete genome sequences. However, the genetic diversity of these viruses is likely to be wider, as exemplified by several of them having only partially sequenced genomes. Second, host associations and host ranges are unknown or unclear for many poxviruses. For instance, the natural host(s) of VACV, yokapox virus or orthopoxvirus Abatino are unknown [142–144]. Even in the case of MPV, its host range in natural settings remains poorly characterized [145]. Thus, the host associations we used for the PCA should be regarded with caution. Third, the function and expression timing of poxvirus genes has been studied in some detail only for VACV, the prototype virus in the genus *Orthopoxvirus*. As a consequence, we based our analysis of expression timing and dinucleotide composition on data obtained for this virus, which may, or may not, be representative of all poxviruses. Our conclusion will thus require validation using other host systems and possibly poxviruses in different genera. Finally, intraspecific genetic diversity data are relatively scant. As a consequence, mutation biases, G+Ceq and recombination rates could only be estimated for a minority of poxviruses, weakening the generalization of our conclusions. Despite these limitations, our results and interpretation widen our understanding of the natural history of poxviruses and may inform strategies of genetic manipulation and synthetic biology to produce recombinant vaccines and oncolytic viruses, as well as biotechnological tools.

Funding information
This work was supported by the Italian Ministry of Health ('Ricerca Corrente 2023' to M.S.). I.G.B. was funded by the European Union's Horizon 2020 research and innovation programme under the grant agreement CODOVIREVOL (ERC-2014-CoG-647916).

Conflicts of interest
The authors declare no competing interests.

References

1. Gyuranecz M, Foster JT, Dán Á, Ip HS, Egstad KF, *et al*. Worldwide phylogenetic relationship of avian poxviruses. *J Virol* 2013;87:4938–4951.
2. Sarker S, Isberg SR, Moran JL, Araujo RD, Elliott N, *et al*. Crocodilepox virus evolutionary genomics supports observed poxvirus infection dynamics on saltwater crocodile (*Crocodylus porosus*). *Viruses* 2019;11:1116.
3. Alonso RC, Moura PP, Caldeira DF, Mendes MHAF, Pinto MHB, *et al*. Poxviruses diagnosed in cattle from Distrito Federal, Brazil (2015-2018). *Transbound Emerg Dis* 2020;67:1563–1573.
4. Lefkowitz EJ, Wang C, Upton C. Poxviruses: past, present and future. *Virus Res* 2006;117:105–118.
5. Moss B. Poxvirus DNA replication. *Cold Spring Harb Perspect Biol* 2013;5:a010199.
6. McInnes CJ, Damon IK, Smith GL, McFadden G, Isaacs SN, *et al*. (2023) ICTV Virus Taxonomy Profile: *Poxviridae* 2023. *J Gen Virol*;104. Epub ahead of print 31 May 2023.
7. Fenner F, Henderson DA, Arita I, Jezek Z, Ladnyi ID, *et al*. Smallpox and its eradication / F. Fenner... [*et al.*]; 1988. https://apps.who.int/iris/handle/10665/39485
8. Kraemer MUG, Tegally H, Pigott DM, Dasgupta A, Sheldon J, *et al*. Tracking the 2022 monkeypox outbreak with epidemiological data in real-time. *Lancet Infect Dis* 2022;22:941–942.
9. Silva NIO, de Oliveira JS, Kroon EG, Trindade G de S, Drumond BP. Here, there, and everywhere: the wide host range and geographic distribution of zoonotic orthopoxviruses. *Viruses* 2020;13:43.
10. McVey DS. Poxviridae. In: McVey DS, Kennedy M, Chengappa MM and Wilkes R (eds). *Veterinary Microbiology*. Wiley; 2022. pp. 522–532.
11. Chen X, Anstey AV, Bugert JJ. Molluscum contagiosum virus infection. *Lancet Infect Dis* 2013;13:877–888.
12. Takatsuka J, Nakai M, Shinoda T. A virus carries a gene encoding juvenile hormone acid methyltransferase, a key regulatory enzyme in insect metamorphosis. *Sci Rep* 2017;7:13522.
13. Nakai M, Kinjo H, Takatsuka J, Shiotsuki T, Kamita SG, *et al*. Entomopoxvirus infection induces changes in both juvenile hormone and ecdysteroid levels in larval *Mythimna separata*. *J Gen Virol* 2016;97:225–232.
14. Palli SR, Ladd TR, Tomkins WL, Shu S, Ramaswamy SB, *et al*. *Choristoneura fumiferana* entomopoxvirus prevents metamorphosis and modulates juvenile hormone and ecdysteroid titers. *Insect Biochem Mol Biol* 2000;30:869–876.
15. Coffman KA, Hankinson QM, Burke GR. A viral mutualist employs posthatch transmission for vertical and horizontal spread among parasitoid wasps. *Proc Natl Acad Sci USA* 2022;119:e2120048119.
16. Chiu E, Hijnen M, Bunker RD, Boudes M, Rajendran C, *et al*. Structural basis for the enhancement of virulence by viral spindles and their in vivo crystallization. *Proc Natl Acad Sci USA* 2015;112:3973–3978.
17. Woods SA, Streett DA, Henry JE. Temporal patterns of mortality from an entomopoxvirus and strategies for control of the migratory grasshopper (*Melanoplus sanguinipes* F.). *J Invertebr Pathol* 1992;60:33–39.
18. Hendrickson RC, Wang C, Hatcher EL, Lefkowitz EJ. Orthopoxvirus genome evolution: the role of gene loss. *Viruses* 2010;2:1933–1967.
19. Senkevich TG, Yutin N, Wolf YI, Koonin EV, Moss B. Ancient gene capture and recent gene loss shape the evolution of orthopoxvirus-host interaction genes. *mBio* 2021;12:e0149521.
20. Upton C, Slack S, Hunter AL, Ehlers A, Roper RL. Poxvirus orthologous clusters: toward defining the minimum essential poxvirus genome. *J Virol* 2003;77:7590–7600.
21. Hatcher EL, Hendrickson RC, Lefkowitz EJ. Identification of nucleotide-level changes impacting gene content and genome evolution in orthopoxviruses. *J Virol* 2014;88:13651–13668.
22. Akashi H. Codon bias evolution in *Drosophila*. Population genetics of mutation-selection drift. *Gene* 1997;205:269–278.
23. Duret L. Evolution of synonymous codon usage in metazoans. *Curr Opin Genet Dev* 2002;12:640–649.
24. Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* 2011;12:32–42.
25. Mordstein C, Cano L, Morales AC, Young B, Ho AT, *et al*. Transcription, mRNA export, and immune evasion shape the codon usage of viruses. *Genome Biol Evol* 2021;13:evab106.
26. Bulmer M. The selection-mutation-drift theory of synonymous codon usage. *Genetics* 1991;129:897–907.

27. Bahir I, Fromer M, Prat Y, Linial M. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol Syst Biol* 2009;5:311.
28. Lucks JB, Nelson DR, Kudla GR, Plotkin JB. Genome landscapes and bacteriophage codon usage. *PLoS Comput Biol* 2008;4:e1000001.
29. Wong EHM, Smith DK, Rabadan R, Peiris M, Poon LLM. Codon usage bias and the evolution of influenza A viruses. Codon Usage Biases of Influenza Virus. *BMC Evol Biol* 2010;10:253.
30. Zhou J, Xing Y, Zhou Z, Wang S. A comprehensive analysis of Usutu virus (USUV) genomes revealed lineage-specific codon usage patterns and host adaptations. *Front Microbiol* 2023;13:967999.
31. Yu C, Li J, Li Q, Chang S, Cao Y, *et al*. Hepatitis B virus (HBV) codon adapts well to the gene expression profile of liver cancer: an evolutionary explanation for HBV's oncogenic role. *J Microbiol* 2022;60:1106–1112.
32. Qin L, Ding S, Wang Z, Jiang R, He Z. Host plants shape the Codon Usage Pattern of Turnip Mosaic Virus. *Viruses* 2022;14:2267.
33. Kumar N, Kulkarni DD, Lee B, Kaushik R, Bhatia S, *et al*. Evolution of Codon Usage Bias in Henipaviruses is governed by natural selection and is host-specific. *Viruses* 2018;10:604.
34. Wang Q, Lyu X, Cheng J, Fu Y, Lin Y, *et al*. Codon Usage provides insights into the adaptive evolution of Mycoviruses in their associated fungi host. *Int J Mol Sci* 2022;23:7441.
35. Félez-Sánchez M, Trösemeier J-H, Bedhomme S, González-Bravo MI, Kamp C, *et al*. Cancer, Warts, or Asymptomatic infections: clinical presentation matches Codon Usage preferences in human papillomaviruses. *Genome Biol Evol* 2015;7:2117–2135.
36. Chen F, Yang J-R. Distinct codon usage bias evolutionary patterns between weakly and strongly virulent respiratory viruses. *iScience* 2022;25:103682.
37. Lauring AS, Jones JO, Andino R. Rationalizing the development of live attenuated virus vaccines. *Nat Biotechnol* 2010;28:573–579.
38. Le Nouën C, Brock LG, Luongo C, McCarty T, Yang L, *et al*. Attenuation of human respiratory syncytial virus by genome-scale codon-pair deoptimization. *Proc Natl Acad Sci USA* 2014;111:13169–13174.
39. Eschke K, Trimpert J, Osterrieder N, Kunec D, Mocarski E. Attenuation of a very virulent Marek's disease herpesvirus (MDV) by codon pair bias deoptimization. *PLoS Pathog* 2018;14:e1006857.
40. Cheng BYH, Nogales A, de la Torre JC, Martínez-Sobrido L. Development of live-attenuated arenavirus vaccines based on codon deoptimization of the viral glycoprotein. *Virology* 2017;501:35–46.
41. MartinezMA, Jordan-Paiz A, Franco S, Nevot M. Synonymous virus genome recoding as a tool to impact viral fitness. *Trends Microbiol* 2016;24:134–147.
42. Simón D, Cristina J, Musto H. An overview of dinucleotide and codon usage in all viruses. *Arch Virol* 2022;167:1443–1448.
43. Kunec D, Osterrieder N. Codon pair bias is a direct consequence of dinucleotide bias. *Cell Rep* 2016;14:55–67.
44. Daron J, Bravo IG. Variability in codon usage in coronaviruses is mainly driven by mutational bias and selective constraints on CpG dinucleotide. *Viruses* 2021;13:1800.
45. Cooper DN, Gerber-Huber S. DNA methylation and CpG suppression. *Cell Differ* 1985;17:199–205.
46. Karlin S, Ladunga I, Blaisdell BE. Heterogeneity of genomes: measures and values. *Proc Natl Acad Sci USA* 1994;91:12837–12841.
47. Burge C, Campbell AM, Karlin S. Over- and under-representation of short oligonucleotides in DNA sequences. *Proc Natl Acad Sci USA* 1992;89:1358–1362.
48. Simmen MW. Genome-scale relationships between cytosine methylation and dinucleotide abundances in animals. *Genomics* 2008;92:33–40.
49. Bird AP, Taggart MH. Variable patterns of total DNA and rDNA methylation in animals. *Nucl Acids Res* 1980;8:1485–1497.
50. Gentles AJ, Karlin S. Genome-scale compositional comparisons in eukaryotes. *Genome Res* 2001;11:540–546.
51. Simmonds P, Xia W, Baillie JK, McKinnon K. Modelling mutational and selection pressures on dinucleotides in eukaryotic phyla--selection against CpG and UpA in cytoplasmically expressed RNA and in RNA viruses. *BMC Genome* 2013;14:610.
52. Provataris P, Meusemann K, Niehuis O, Grath S, Misof B. Signatures of DNA methylation across insects suggest reduced DNA methylation levels in holometabola. *Genome Biol Evol* 2018;10:1185–1197.
53. Gonçalves-Carneiro D, Takata MA, Ong H, Shilton A, Bieniasz PD. Origin and evolution of the zinc finger antiviral protein. *PLoS Pathog* 2021;17:e1009545.
54. Beutler E, Gelbart T, Han JH, Koziol JA, Beutler B. Evolution of the genome and the genetic code: selection at the dinucleotide level by methylation and polyribonucleotide cleavage. *Proc Natl Acad Sci USA* 1989;86:192–196.
55. Karlin S, Mrázek J. Compositional differences within and between eukaryotic genomes. *Proc Natl Acad Sci USA* 1997;94:10227–10232.
56. Bauer AP, Leikam D, Krinner S, Notka F, Ludwig C, *et al*. The impact of intragenic CpG content on gene expression. *Nucleic Acids Res* 2010;38:3891–3908.
57. Krinner S, Heitzer AP, Diermeier SD, Obermeier I, Längst G, *et al*. CpG domains downstream of TSSs promote high levels of gene expression. *Nucleic Acids Res* 2014;42:3551–3564.
58. Takata MA, Gonçalves-Carneiro D, Zang TM, Soll SJ, York A, *et al*. CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature* 2017;550:124–127.
59. Luo X, Wang X, Gao Y, Zhu J, Liu S, *et al*. Molecular mechanism of RNA recognition by zinc-finger antiviral protein. *Cell Rep* 2020;30:46–52.
60. Bowie AG, Unterholzner L. Viral evasion and subversion of pattern-recognition receptor signalling. *Nat Rev Immunol* 2008;8:911–922.
61. Goonawardane N, Nguyen D, Simmonds P, Schwemmle M. Association of zinc finger antiviral protein binding to viral genomic RNA with attenuation of replication of echovirus 7. *mSphere* 2021;6:e01138-20.
62. Odon V, Fros JJ, Goonawardane N, Dietrich I, Ibrahim A, *et al*. The role of ZAP and OAS3/RNAseL pathways in the attenuation of an RNA virus with elevated frequencies of CpG and UpA dinucleotides. *Nucleic Acids Res* 2019;47:8061–8083.
63. Ficarelli M, Neil SJD, Swanson CM. Targeted restriction of viral gene expression and replication by the ZAP antiviral system. *Annu Rev Virol* 2021;8:265–283.
64. Ficarelli M, Wilson H, Pedro Galão R, Mazzon M, Antzin-Anduetza I, *et al*. KHNYN is essential for the zinc finger antiviral protein (ZAP) to restrict HIV-1 containing clustered CpG dinucleotides. *Elife* 2019;8:e46767.
65. Miyazato P, Matsuo M, Tan BJY, Tokunaga M, Katsuya H, *et al*. HTLV-1 contains a high CG dinucleotide content and is susceptible to the host antiviral protein ZAP. *Retrovirology* 2019;16:38.
66. Kmiec D, Nchioua R, Sherrill-Mix S, Stürzel CM, Heusinger E, *et al*. CpG Frequency in the 5' Third of the *env* Gene Determines Sensitivity of Primary HIV-1 Strains to the Zinc-Finger Antiviral Protein. *mBio* 2020;11:e02903-19.
67. Gonçalves-Carneiro D, Mastrocola E, Lei X, DaSilva J, Chan YF, *et al*. Rational attenuation of RNA viruses with zinc finger antiviral protein. *Nat Microbiol* 2022;7:1558–1567.
68. Cooper DA, Banerjee S, Chakrabarti A, García-Sastre A, Hesselberth JR, *et al*. RNase L targets distinct sites in influenza A virus RNAs. *J Virol* 2015;89:2764–2776.
69. Lin Y-T, Chiweshe S, McCormick D, Raper A, Wickenhagen A, *et al*. Human cytomegalovirus evades ZAP detection by suppressing CpG dinucleotides in the major immediate early 1 gene. *PLoS pathogens* 2020;16:e1008844.

70. Gonzalez-Perez AC, Stempel M, Wyler E, Urban C, Piras A, *et al*. The zinc finger antiviral protein ZAP restricts human cytomegalovirus and selectively binds and Destabilizes viral Ul4/Ul5 transcripts. *mBio* 2021;12:e02683-20.
71. Peng C, Wyatt LS, Glushakow-Smith SG, Lal-Nag M, Weisberg AS, *et al*. Zinc-finger antiviral protein (ZAP) is a restriction factor for replication of modified vaccinia virus Ankara (MVA) in human cells. *PLoS Pathog* 2020;16:e1008845.
72. Burns CC, Campagnoli R, Shaw J, Vincent A, Jorba J, *et al*. Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons. *J Virol* 2009;83:9957–9969.
73. Gaunt E, Wise HM, Zhang H, Lee LN, Atkinson NJ, *et al*. Elevation of CpG frequencies in influenza A genome attenuates pathogenicity but enhances host response to infection. *eLife* 2016;5.
74. Antzin-Anduetza I, Mahiet C, Granger LA, Odendall C, Swanson CM. Increasing the CpG dinucleotide abundance in the HIV-1 genomic RNA inhibits viral replication. *Retrovirology* 2017;14:49.
75. Fros JJ, Dietrich I, Alshaikhahmed K, Passchier TC, Evans DJ, *et al*. CpG and UpA dinucleotides in both coding and non-coding regions of echovirus 7 inhibit replication initiation post-entry. *eLife* 2017;6.
76. Pereira-Gómez M, Carrau L, Fajardo Á, Moreno P, Moratorio G. Altering compositional properties of viral genomes to design live-attenuated vaccines. *Front Microbiol* 2021;12:676582.
77. Lauring AS, Acevedo A, Cooper SB, Andino R. Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. *Cell Host Microbe* 2012;12:623–632.
78. Gigante CM, Gao J, Tang S, McCollum AM, Wilkins K, *et al*. Genome of Alaskapox virus, a novel orthopoxvirus isolated from Alaska. *Viruses* 2019;11:708.
79. Bourret J, Alizon S, Bravo IG. COUSIN (COdon Usage Similarity INdex): a normalized measure of Codon Usage preferences. *Genome Biol Evol* 2019;11:3523–3528.
80. Charif D, Lobry JR. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: Bastolla U, Porto M, Roman HE and Vendruscolo M (eds). *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations*. Berlin, Heidelberg: Springer; pp. 207–232.
81. Rohart F, Gautier B, Singh A, Lê Cao K-A. mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol* 2017;13:e1005752.
82. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 2019;20.
83. Yang Z, Bruno DP, Martens CA, Porcella SF, Moss B. Simultaneous high-resolution analysis of vaccinia virus and host cell transcriptomes by deep RNA sequencing. *Proc Natl Acad Sci USA* 2010;107:11513–11518.
84. Pohlert T. The Pairwise multiple comparison of mean ranks package (PMCMR). *R package* 2014:2004–2006.
85. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.
86. Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res* 2016;44:W232–5.
87. Chernomor O, von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol* 2016;65:997–1008.
88. Pagel M, Meade A, Barker D. Bayesian estimation of ancestral character states on phylogenies. *Syst Biol* 2004;53:673–684.
89. Garland T, Dickerman AW, Janis CM, Jones JA. Phylogenetic analysis of covariance by computer simulation. *Syst Biol* 1993;42:265–292.
90. Harmon LJ, Weir JT, Brock CD, Glor RE, Challenger W. GEIGER: investigating evolutionary radiations. *Bioinformatics* 2008;24:129–131.
91. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol* 2012;3:217–223.
92. Khabbazian M, Kriebel R, Rohe K, Ané C, Hansen T. Fast and accurate detection of evolutionary shifts in Ornstein–Uhlenbeck models. *Methods Ecol Evol* 2016;7:811–824.
93. Didelot X, Wilson DJ. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput Biol* 2015;11:e1004041.
94. Karlin S, Cardon LR. Computational DNA sequence analysis. *Annu Rev Microbiol* 1994;48:619–654.
95. Odon V, Fiddaman SR, Smith AL, Simmonds P. Comparison of CpG- and UpA-mediated restriction of RNA virus replication in mammalian and avian cells and investigation of potential ZAP-mediated shaping of host transcriptome compositions. *RNA* 2022;28:1089–1109.
96. Duret L, Galtier N. The covariation between TpA deficiency, CpG deficiency, and G+C content of human isochores is due to a mathematical artifact. *Mol Biol Evol* 2000;17:1620–1625.
97. Emerson GL, Li Y, Frace MA, Olsen-Rasmussen MA, Khristova ML, *et al*. The phylogenetics and ecology of the orthopoxviruses endemic to North America. *PLoS One* 2009;4:e7666.
98. Simón D, Cristina J, Musto H. Nucleotide composition and Codon usage across viruses and their respective hosts. *Front Microbiol* 2021;12:646300.
99. Pagel M. Inferring the historical patterns of biological evolution. *Nature* 1999;401:877–884.
100. Messer PW. Measuring the rates of spontaneous mutation from deep and large-scale polymorphism data. *Genetics* 2009;182:1219–1232.
101. Hershberg R, Petrov DA. Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet* 2010;6:e1001115.
102. Gigante CM, Korber B, Seabolt MH, Wilkins K, Davidson W, *et al*. Multiple lineages of monkeypox virus detected in the United States, 2021-2022. *Science* 2022;378:560–565.
103. Isidro J, Borges V, Pinto M, Sobral D, Santos JD, *et al*. Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus. *Nat Med* 2022;28:1569–1572.
104. O'Toole Á, Neher RA, Ndodo N, Borges V, Gannon B, *et al*. Putative APOBEC3 deaminase editing in MPXV as evidence for sustained human transmission since at least 2016. *Evol Biol* 2023.
105. Forni D, Cagliani R, Pozzoli U, Sironi M. An APOBEC3 mutational signature in the genomes of human-infecting orthopoxviruses. *mSphere* 2023;8:e0006223.
106. Damaso CR. Revisiting Jenner's mysteries, the role of the Beaugency lymph in the evolutionary path of ancient smallpox vaccines. *Lancet Infect Dis* 2018;18:e55–e63.
107. Kerr PJ, Liu J, Cattadori I, Ghedin E, Read AF, *et al*. Myxoma virus and the Leporipoxviruses: an evolutionary paradigm. *Viruses* 2015;7:1020–1061.
108. Sanjuán R, Domingo-Calap P. Mechanisms of viral mutation. *Cell Mol Life Sci* 2016;73:4433–4448.
109. Greseth MD, Traktman P. The life cycle of the vaccinia virus genome. *Annu Rev Virol* 2022;9:239–259.
110. Templeton CW, Traktman P. UV irradiation of vaccinia virus-infected cells impairs cellular functions, introduces lesions into the viral genome, and uncovers repair capabilities for the viral replication machinery. *J Virol* 2022;96:e0213721.
111. Culyba MJ, Minkah N, Hwang Y, Benhamou O-MJ, Bushman FD. DNA branch nuclease activity of vaccinia A22 resolvase. *J Biol Chem* 2007;282:34644–34652.
112. Senkevich TG, Koonin EV, Moss B. Predicted poxvirus FEN1-like nuclease required for homologous recombination, double-strand break repair and full-size genome formation. *Proc Natl Acad Sci USA* 2009;106:17921–17926.

113. Colinas RJ, Goebel SJ, Davis SW, Johnson GP, Norton EK, *et al*. A DNA ligase gene in the copenhagen strain of vaccinia virus is nonessential for viral replication and recombination. *Virology* 1990;179:267–275.
114. Kerr SM, Smith GL. Vaccinia virus DNA ligase is nonessential for virus replication: recovery of plasmids from virus-infected cells. *Virology* 1991;180:625–632.
115. Parks RJ, Winchcombe-Forhan C, DeLange AM, Xing X, Evans DH. DNA ligase gene disruptions can depress viral growth and replication in poxvirus-infected cells. *Virus Res* 1998;56:135–147.
116. Pouyet F, Mouchiroud D, Duret L, Sémon M. Recombination, meiotic expression and human codon usage. *Elife* 2017;6.
117. Melamed-Bessudo C, Shilo S, Levy AA. Meiotic recombination and genome evolution in plants. *Curr Opin Plant Biol* 2016;30:82–87.
118. Lassalle F, Périan S, Bataillon T, Nesme X, Duret L, *et al*. GC-content evolution in bacterial genomes: the biased gene conversion hypothesis expands. *PLoS Genet* 2015;11:e1004941.
119. Figuet E, Ballenghien M, Romiguier J, Galtier N. Biased gene conversion and GC-content evolution in the coding sequences of reptiles and vertebrates. *Genome Biol Evol* 2014;7:240–250.
120. Karlin S, Burge C. Dinucleotide relative abundance extremes: a genomic signature. *Trends Genet* 1995;11:283–290.
121. Hershberg R, Petrov DA, Nachman MW. General rules for optimal codon choice. *PLoS Genet* 2009;5:e1000556.
122. de Jong MJ, van Oosterhout C, Hoelzel AR, Janke A. Moderating the neutralist-selectionist debate: exactly which propositions are we debating, and which arguments are valid? *Biol Rev Camb Philos Soc* 2023.
123. Picard MAL, Leblay F, Cassan C, Willemsen A, Daron J, *et al*. Transcriptomic, proteomic, and functional consequences of codon usage bias in human cells during heterologous gene expression. *Protein Sci* 2023;32:e4576.
124. Li M, Kao E, Gao X, Sandig H, Limmer K, *et al*. Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. *Nature* 2012;491:125–128.
125. Stabell AC, Hawkins J, Li M, Gao X, David M, *et al*. Non-human primate schlafen11 inhibits production of both host and viral proteins. *PLoS Pathog* 2016;12:e1006066.
126. Greenbaum BD, Levine AJ, Bhanot G, Rabadan R. Patterns of evolution and host gene mimicry in influenza and other RNA viruses. *PLoS Pathog* 2008;4:e1000079.
127. Tulloch F, Atkinson NJ, Evans DJ, Ryan MD, Simmonds P. RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. *Elife* 2014;3:e04531.
128. Di Giallonardo F, Schlub TE, Shi M, Holmes EC. Dinucleotide composition in animal RNA viruses is shaped more by virus family than by host species. *J Virol* 2017;91:e02381-16.
129. Upadhyay M, Vivekanandan P, Burk RD. Depletion of CpG dinucleotides in papillomaviruses and polyomaviruses: a role for divergent evolutionary pressures. *PLoS One* 2015;10:e0142368.
130. King K, Larsen BB, Gryseels S, Richet C, Kraberger S, *et al*. Coevolutionary analysis implicates toll-like receptor 9 in papillomavirus restriction. *mBio* 2022;13:e0005422.
131. White MK, Safak M, Khalili K. Regulation of gene expression in primate polyomaviruses. *J Virol* 2009;83:10846–10856.
132. Lin Y-T, Chau L-F, Coutts H, Mahmoudi M, Drampa V, *et al*. Does the zinc finger antiviral protein (ZAP) shape the evolution of herpesvirus genomes? *Viruses* 2021;13:1857.
133. Kerr SM, Johnston LH, Odell M, Duncan SA, Law KM, *et al*. Vaccinia DNA ligase complements Saccharomyces cerevisiae cdc9, localizes in cytoplasmic factories and affects virulence and virus sensitivity to DNA damaging agents. *EMBO J* 1991;10:4343–4350.
134. Teng W, Liao B, Chen M, Shu W, Faucher SP. Genomic legacies of ancient adaptation illuminate GC-content evolution in bacteria. *Microbiol Spectr* 2023;11:e0214522.
135. Lind PA, Andersson DI. Whole-genome mutational biases in bacteria. *Proc Natl Acad Sci USA* 2008;105:17878–17883.
136. Weissman JL, Fagan WF, Johnson PLF. Linking high GC content to the repair of double strand breaks in prokaryotic genomes. *PLoS Genet* 2019;15:e1008493.
137. Luteijn RD, Drexler I, Smith GL, Lebbink RJ, Wiertz EJHJ. Mutagenic repair of double-stranded DNA breaks in vaccinia virus genomes requires cellular DNA ligase IV activity in the cytosol. *J Gen Virol* 2018;99:790–804.
138. Paran N, De Silva FS, Senkevich TG, Moss B. Cellular DNA ligase I is recruited to cytoplasmic vaccinia virus factories and masks the role of the vaccinia ligase in viral DNA replication. *Cell Host Microbe* 2009;6:563–569.
139. Delamonica B, Davalos L, Larijani M, Anthony SJ, Liu J, *et al*. Evolutionary potential of the monkeypox genome arising from interactions with human APOBEC3 enzymes. *Virus Evol* 2023;9:vead047.
140. Molteni C, Forni D, Cagliani R, Arrigoni F, Pozzoli U, *et al*. Selective events at individual sites underlie the evolution of monkeypox virus clades. *Virus Evol* 2023;9.
141. Molteni C, Forni D, Cagliani R, Mozzi A, Clerici M, *et al*. Evolution of the orthopoxvirus core genome. *Virus Res* 2022;323:198975.
142. Zhao G, Droit L, Tesh RB, Popov VL, Little NS, *et al*. The genome of Yoka poxvirus. *J Virol* 2011;85:10230–10238.
143. Gruber CEM, Giombini E, Selleri M, Tausch SH, Andrusch A, *et al*. Whole genome characterization of *Orthopoxvirus* (OPV) abatino, a zoonotic virus representing a putative novel clade of old world Orthopoxviruses. *Viruses* 2018;10:546.
144. Esparza J, Schrick L, Damaso CR, Nitsche A. Equination (inoculation of horsepox): an early alternative to vaccination (inoculation of cowpox) and the potential role of horsepox virus in the origin of the smallpox vaccine. *Vaccine* 2017;35:7222–7230.
145. Forni D, Cagliani R, Molteni C, Clerici M, Sironi M. Monkeypox virus: the changing facets of a zoonotic pathogen. *Infect Genet Evol* 2022;105:105372.