WIREs RNA WILEY

Compositional biases in RNA viruses: Causes, consequences and applications

Eleanor R. Gaunt | **Paul Digard** [ID]

Department of Infection and Immunity, The Roslin Institute, The University of Edinburgh, Edinburgh, UK

Correspondence
Eleanor R. Gaunt, Department of Infection and Immunity, The Roslin Institute, The University of Edinburgh, Easter Bush Campus, Midlothian EH25 9RG, UK.
Email: elly.gaunt@ed.ac.uk

Edited by Jeff Wilusz, Editor-in-Chief

Abstract

If each of the four nucleotides were represented equally in the genomes of viruses and the hosts they infect, each base would occur at a frequency of 25%. However, this is not observed in nature. Similarly, the order of nucleotides is not random (e.g., in the human genome, guanine follows cytosine at a frequency of ~0.0125, or a quarter the number of times predicted by random representation). Codon usage and codon order are also nonrandom. Furthermore, nucleotide and codon biases vary between species. Such biases have various drivers, including cellular proteins that recognize specific patterns in nucleic acids, that once triggered, induce mutations or invoke intrinsic or innate immune responses. In this review we examine the types of compositional biases identified in viral genomes and current understanding of the evolutionary mechanisms underpinning these trends. Finally, we consider the potential for large scale synonymous recoding strategies to engineer RNA virus vaccines, including those with pandemic potential, such as influenza A virus and Severe Acute Respiratory Syndrome Coronavirus Virus 2.

This article is categorized under:
  RNA in Disease and Development > RNA in Disease
  RNA Evolution and Genomics > Computational Analyses of RNA
  RNA Interactions with Proteins and Other Molecules > Protein-RNA Recognition

KEYWORDS
dinucleotides, mutation bias, selection bias, viral genome composition

1 | INTRODUCTION

Synonymous or "silent" nucleotide substitutions in a genome are nucleotide changes that do not result in an amino acid change. The impact of synonymous changes are nevertheless far from phenotypically silent, and have been shown to affect the encoded transcripts in several ways, including mRNA secondary structure (Kudla et al., 2009; Shabalina et al., 2006), mRNA splicing (Warnecke et al., 2009), mRNA stability (Presnyak et al., 2015), microRNA targeting (Birnbaum et al., 2012; Brest et al., 2011), co-translational protein folding (Pechmann & Frydman, 2013), and, in the

context of RNA virus transcripts and genomes, cellular sensing of pathogens (Takata et al., 2017). Synonymous changes are possible due to the phenomenon of codon degeneracy; the potential for one amino acid to be encoded by multiple nucleotide triplets. Codon degeneracy allows viruses a degree of genome plasticity that, by facilitating the evolution of overlapping open reading frames (ORFs) permits the generation of sometimes astonishing genome compression. Nevertheless, while roughly one third of bases in a coding sequence could undergo substitution synonymously, this does not happen, and bases in positions that would theoretically support silent substitution are not randomly represented. In some cases, this is because of superimposed functional elements, such as overlapping ORFs or *cis*-acting RNA signals. As examples, internal ribosome entry sites (IRESs), first discovered in the *Picornaviridae* (Jang & Wimmer, 1990; Trono et al., 1988), are RNA structures formed over hundreds of nucleotides that enable cap-independent translation initiation, and silent mutations can impair IRES function. Likewise, orthomyxoviruses package the correct selection of segments through RNA–RNA interactions partly mediated by the coding regions of the genes and again, synonymous mutations can be functionally deleterious (Li et al., 2021). However, synonymous recoding of virus genomes while avoiding known functional elements can nevertheless significantly attenuate virus replication, indicating further constraints on genome sequence acting at a more global level. At this genome level, preferential selection of particular nucleotides at synonymous sites has been previously identified to result in various types of compositional bias, including nucleotide bias (Auewarakul, 2005; Balzarini et al., 2001; Berkhout et al., 2002; Grantham et al., 1980; Jenkins et al., 2001; Kapoor et al., 2010; Lobo et al., 2009; Müller & Bonhoeffer, 2005; Rothberg & Wimmer, 1981; Shackelton et al., 2006; van der Kuyl & Berkhout, 2012; van Hemert et al., 2007; van Hemert & Berkhout, 2016), codon usage bias (Adams & Antoniw, 2004; Aragonès et al., 2010; Bahir et al., 2009; Belalov & Lukashev, 2013; Berkhout et al., 2002; Bouquet et al., 2012; Butt et al., 2014; Cai et al., 2009; Chen, 2013; D'Andrea et al., 2011; Fu, 2010; Grantham et al., 1980; Haas et al., 1996; He et al., 2017; Jenkins et al., 2001; Jenkins & Holmes, 2003; Kumar et al., 2016; Li et al., 2012; Liu et al., 2010; Nougairede et al., 2013; Plotkin & Dushoff, 2003; Rothberg & Wimmer, 1981; Tao et al., 2009; van Hemert et al., 2007; Wong et al., 2010; Zhao et al., 2003; Zhong et al., 2007), dinucleotide bias (Antzin-Anduetza et al., 2017; Atkinson et al., 2014; Coffin et al., 1995; Di Giallonardo et al., 2017; Gaunt et al., 2016; Karlin et al., 1994; Kunec & Osterrieder, 2016; Rima & McFerran, 1997; Rothberg & Wimmer, 1981; Shackelton et al., 2006; Simmonds et al., 2015; Tao et al., 2009; Tulloch et al., 2014; Upadhyay et al., 2014; Washenberger et al., 2007; Witteveldt et al., 2016) and codon pair bias (Coleman et al., 2008; Gao et al., 2015; Le Nouën et al., 2014; Leifer et al., 2011; Li et al., 2018; Martrus et al., 2013; Mueller et al., 2010; Ni et al., 2014; Wang et al., 2015; Yang et al., 2013). Each of these are elaborated below, along with our current understanding of the underlying mechanisms by which these biases are generated.

While the focus of this review is on genome compositional biases of RNA viruses, often the leading research in a specific area has been undertaking using a DNA virus as a model system, and so where appropriate this research is also described. It is important to note that the concepts discussed have been evaluated using diverse virus systems, often with fundamentally different replication strategies. Exposure to cellular factors is expected to vary depending on where in the cell a virus replicates, the extent of protection of viral genomes from the cellular environment by nucleoproteins, the kinetics of virus replication, as well as the host species and the cell type infected. Nevertheless, all viruses produce mRNAs that are translated in the cytoplasm, so some generalities are likely to exist, as well as differences.

2 | TYPES OF GENOME COMPOSITIONAL BIAS

2.1 | Nucleotide bias

If all bases were represented equally in a genome, each would be recorded at a frequency of 25%. However, biases in individual base frequencies are seen across all genomes, including viral. This is often facilitated by codon degeneracy. Of 20 amino acids, 2 are encoded by a unique codon (Met, Trp); nine by two codons (Phe, Tyr, His, Gln, Asn, Lys, Asp, Glu, Cys); Ile is encoded by three codons; five amino acids are encoded by four codons (Val, Pro, Thr, Ala, Gly) and three are encoded by six codons (Leu, Arg, Ser). Representation of each of the degenerate codons can be highly skewed. For example, across the HIV-1 genome, ~37% of bases are adenine, and adenines are heavily selected for at degenerate positions (Kypr & Mrázek, 1987). This bias is at least partly induced by the cellular factor APOBEC3G (Sheehy et al., 2002), which deaminates cytidine to uridine in the negative sense ssDNA produced during virus replication as an intrinsic antiretroviral defense. Uridine mimics thymine, and so when positive sense DNA is synthesized during genome replication, this reverse complement strand incorporates adenine in place of guanine. This is, in other words, ***mutationally driven***. Conversely, enrichment for adenine at specific sites is thought to reduce the impact of ribosomal frame-shift events due to introduction

of out-of-frame stop codons, as modeled using bacterial genomes (Abrahams & Hurst, 2017) (i.e., driven by **selection**). Other types of nucleotide biases are also described, such as the 70% GC content of the rubella virus genome (Zhou et al., 2012), largely attributed to the use of C bases at degenerate positions(Zhou et al., 2012). Contrarily, extensive C to U mutations (in comparison to other base changes) are seen in the genome of SARS-CoV-2 (Rice et al., 2020; Simmonds, 2020). The mechanisms driving these latter two biases are, at present, poorly understood.

2.2 | Codon usage biases

Usage of degenerate codons is nonrandom, with some codons used frequently and others rarely. Codon preferences vary by host and by viral species, and even by gene. In humans, codon usage biases are stronger in genes that are more highly expressed. The greater exposure of the transcripts from these genes to the drivers of selection may generate stronger biases (Urrutia & Hurst, 2003). Commonly expressed genes use codons which are decoded by abundant tRNAs, whereas during stress the tRNA pool changes to increase abundance of rare tRNAs, as stress response genes are more likely to use rare codons (Torrent et al., 2018). Within-gene biases are also evident; for example, evolutionarily constrained exonic splice enhancer sites demonstrate different codon usage patterns to other coding regions (Savisaar & Hurst, 2018).

In virology, how well a virus reflects the codon usage of its host can be calculated using the Codon Adaptation Index (CAI) metric. Key to genome composition variation is how long a virus has been adapting to its host; for example, a virus that has recently switched host may change its genome composition profile as it adapts to a new host (Babayan et al., 2018; Greenbaum et al., 2008). In CAI scoring, the most frequently used codons score highly and rare codons score below 1. The scores can then be averaged across an ORF or a proteome. CAI scores vary between −1 and +1, with higher scores representing more frequently used codons with respect to the host (Sharp & Li, 1987). Viral genomes display codon usage biases, but these do not necessarily mimic their host. This may arise as a consequence of nucleotide biases; for example, HIV-1 and rubella virus display very different codon usage profiles to each other and the human genome as a result of the nucleotide biases they exhibit (van der Kuyl & Berkhout, 2012; Zhou et al., 2012). Genome architecture and virus ecology may also be important for driving codon usage preferences, as codon usage biases may be more evident in segmented and aerosol-borne viruses compared with vector-borne viruses (Jenkins & Holmes, 2003), as vector-borne viruses must also be able to replicate in their invertebrate hosts (Fros et al., 2021). Within viral genomes, codon usage preferences may also vary. For some large DNA viruses, distinct temporal phases of infection occur; usage of rare codons in late genes of large DNA viruses has been proposed as a mechanism of gene expression regulation (Shin et al., 2015; Zhou et al., 1999). In the SARS-CoV-2 genome, E ORF and ORF10 encode a high proportion of disfavoured codons, whereas in other genes, codon usage is more reflective of the human host (Digard et al., 2020; Rice et al., 2020).

While the reason(s) underlying codon preferences are somewhat speculative, successive codons encoding the same amino acid are more likely to use the same degenerate base and so the same tRNA, possibly allowing for faster recycling of tRNAs, if tRNA diffusion away from the ribosome happens slower than the rate of translation (Cannarozzi et al., 2010). A nonexclusive alternative is that use of rare codons slows translational rate, which in turn can affect how a protein folds (Kimchi-Sarfaty et al., 2007).

2.3 | Dinucleotide biases

In 1981 it was first proposed that nucleotide and codon preferences might be explained by dinucleotide biases (Nussinov, 1981). A dinucleotide is defined as two adjacent nucleotide bases joined by a phosphate bridge, on the same strand of nucleic acid (i.e., in *cis*). Given the four bases of RNA—adenine (A), cytosine (C), guanine (G) and uracil (U)—all possible combinations give rise to 16 possible dinucleotides (Figure 1a). The conventional notation for dinucleotides of, for example "CpG," refers to a cytosine 5′ to a guanine base and joined by a phosphate ("p") bridge (Figure 1b).

In a given sequence, how often a given dinucleotide would occur if nucleotide sequence was random can be calculated by simply multiplying observed base frequencies together. By then counting how many times the chosen dinucleotide occurs in a given sequence, over- or under-representation of any dinucleotide can be calculated. This is referred to as the observed: expected (O:E) ratio, represented by the formula:
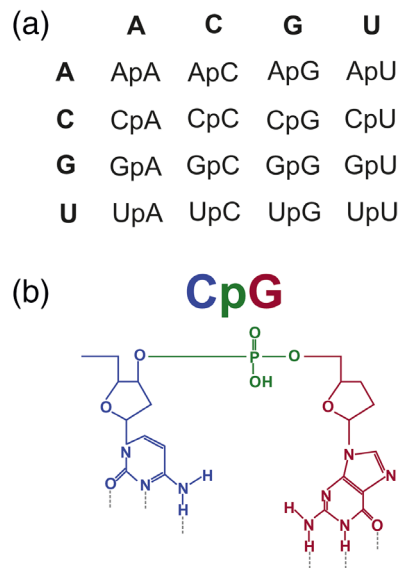
$$O:E = f(XpY)/f(X).f(Y)$$

FIGURE 1 (a) There are 16 possible dinucleotide compositions in RNA. (b) Schematic of CpG motif, with "p" referring to the phosphate bridge (green) joining the cytosine (C) (blue) and guanine (G) (red) bases

where X and Y represent the two nucleotides of choice. A ratio of above 1 indicates that the observed frequency is higher than expected, and so the dinucleotide is over-represented, whereas anything below 1 indicates an under-represented dinucleotide. As an example, consider the CpG O:E ratio for the A/Puerto Rico/8/1934 (PR8) strain of H1N1 influenza A virus (IAV). In PR8, there are 3298 cytosines and 2595 guanines out of a total genome size of 13,588 nucleotides (summed across 8 segments). Thus, the frequencies of C and G are 0.243 and 0.191 respectively. There are 285 CpG motifs out of 13,581 dinucleotides, or an observed CpG ratio of 0.021.

$$\text{CpG O}:\text{E ratio for the PR8 IAV strain} = 0.021/(0.243 \times 0.191) = 0.453.$$

This simplest method of calculating dinucleotide representation does not take into consideration potential sources of exogenous bias such as amino acid composition and codon bias, although software accounting for such factors has been released (Simmonds, 2012); in our experience of analyzing viral genomes, the results delivered by different models are very similar.

2.4 | CpG dinucleotides

Vertebrate genomic dinucleotide composition has been studied since the 1960s, when the striking observation was made that CpG motifs are under-represented in vertebrate genomes (Swartz et al., 1962; Josse et al., 1961). The human genome has a CpG O:E ratio of around 0.25 (Bird, 1980), similar to other mammalian species (Jabbari et al., 1997) (i.e., CpGs occur at a quarter of the frequency one would expect, given individual cytosine and guanine frequencies in the human genome). Little if any CpG suppression is seen in the genome of invertebrates (Josse et al., 1961; Simmonds et al., 2013) (Figure 2), although CpG suppression is seen in plant genomes (Bougraa & Perrin, 1987; Ibrahim et al., 2019).

In vertebrates, genomic CpG suppression is thought to have arisen due to the epigenetic regulation of transcription occurring in part through the methylation of cytosines in the CpG conformation. Methylated cytosines are prone to undergo spontaneous deamination and so conversion to thymine (i.e., TpG), which is proposed to have resulted in a loss of CpG motifs from vertebrate genomes over evolutionary time (Cooper & Krawczak, 1989). Methylation of cytosines in invertebrate genomes is restricted or entirely absent (Bird & Tweedie, 1995), providing an explanation for the contrasting lack of CpG suppression in these organisms. The reasons for CpG suppression in plant genomes are unclear, as they do not support methylation (Bougraa & Perrin, 1987).
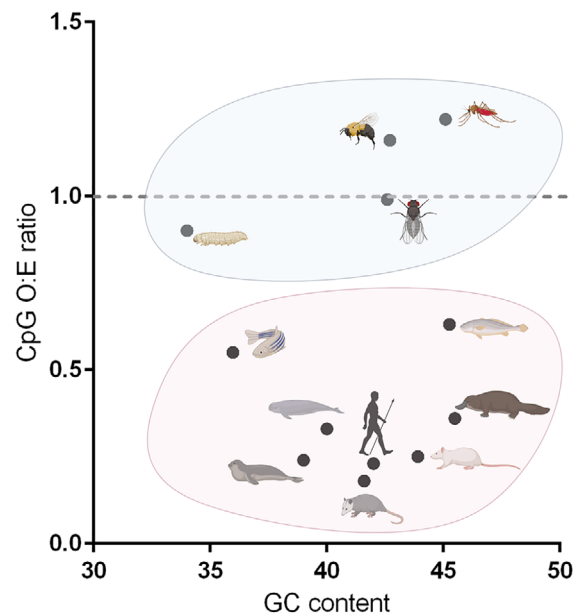
FIGURE 2 GC content vs CpG ratio for various invertebrate (blue circle) and vertebrate (pink circle) species. In blue from left to right: *Spodoptera exempta* (African armyworm), *Drosophila melanogaster* (fruit fly), *Bombus bombus* (bumble bee), *Anopheles gambiae* (mosquito). In pink from left to right: *Danio rerio* (zebrafish), Halichoerus spp (seals), Phocoena spp (porpoise), *Didelphis virginiana* (opossum), *Homo sapiens* (human), *Rattus norvegicus* (brown rat), *Takifugu rubripes* (pufferfish), *Ornithorhynchus anatinus* (platypus)

In the 1990s it was reported that the genomes of small, but not large, viruses infecting eukaryotes also under-represent CpG (Karlin et al., 1994). A more detailed analysis (Simmonds et al., 2013) showed that generally in viruses of mammals, single stranded RNA (ssRNA) viruses under-represent CpG, whereas dsRNA and large DNA viruses do not (Simmonds et al., 2013). The under-representation of CpG in the IAV PR8 genome described above is therefore characteristic of its class of RNA viruses. By comparison, CpG suppression is less apparent or entirely absent in invertebrate viruses (Simmonds et al., 2013). Viral CpG content can therefore be approximated using the genome type-based Baltimore classification of viruses (Baltimore, 1971) except in the case of dsDNA viruses, where size matters (Simmonds et al., 2013). Viruses under-representing CpG in their genomes include the groups of +ssRNA, −ssRNA, small dsDNA, ssDNA (which generally have small genome sizes), positive sense ssRNA reverse transcriptase viruses, and dsDNA reverse transcriptase viruses, while those that do not are dsRNA and large dsDNA viruses (Figure 3a). Overall, for RNA viruses, the extent of CpG bias is considered to be reflective of host (Simmonds et al., 2013). The mechanistic underpinnings giving way to varied rates of CpG suppression are likely to vary between, and even within, different Baltimore group virus classifications due to the differing cellular environments each type of viral genome is exposed to, as well as the different ways in which viruses regulate the cellular environment.

2.5 | UpA dinucleotides

Dramatic dinucleotide suppression in the genome of vertebrates is unique to CpG. However, the TpA dinucleotide is modestly under-represented in the genomes of vertebrates, invertebrates (Simmonds et al., 2013) and plants (Bougraa & Perrin, 1987). Both RNA and DNA viruses mimic their host by displaying moderate suppression of the UpA dinucleotide (Di Giallonardo et al., 2017), but to varying extents (Figure 3b).

2.6 | Codon pair bias

During translation, a ribosome decodes two codons simultaneously, and so as well as codon usage, codon order is also important. Some codon pairs are used more frequently than others, and this is considered as a separate phenomenon of
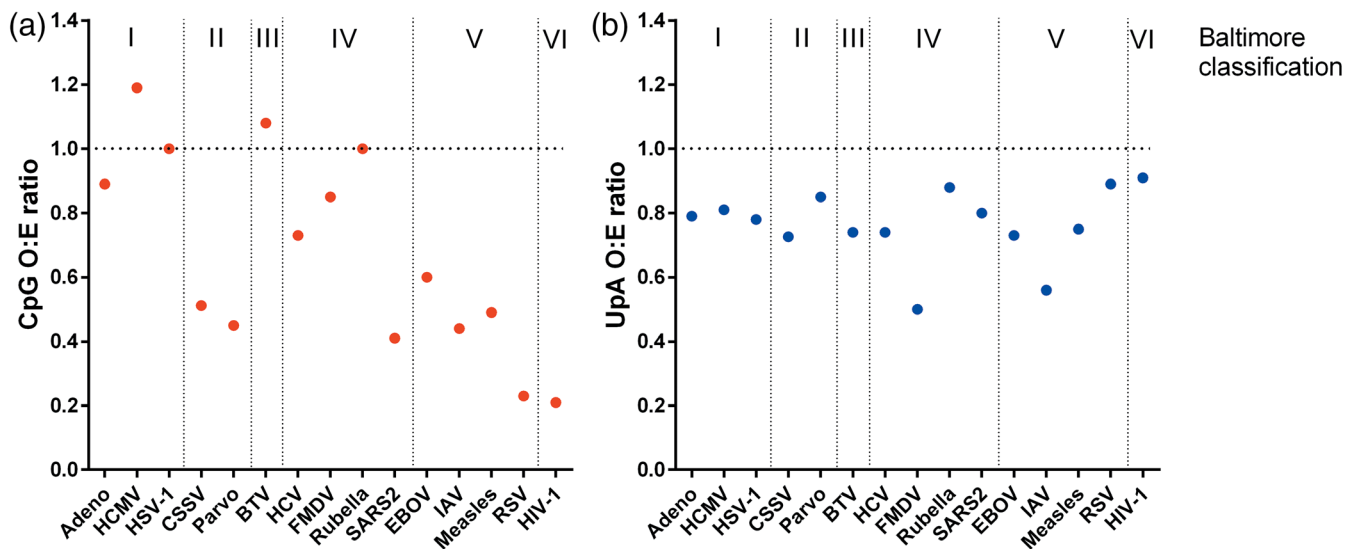
FIGURE 3 Under-representation of CpG dinucleotides (a) and UpA dinucleotides (b) in the genomes of representative viruses. Abbreviations are Adeno, human adenovirus 2; HCMV, human cytomegalovirus; HSV-1, herpes simplex virus 1; parvo, parvovirus; BTV, bluetongue virus; HCV, hepatitis C virus; FMDV, foot and mouth disease virus; SARS2, severe acute respiratory syndrome coronavirus 2; EBOV, ebola virus; IAV, influenza A virus; RSV, respiratory syncytial virus; HIV-1, human immunodeficiency virus 1. The Baltimore classifications are **I** dsDNA; **II** ssDNA; **III** dsRNA; **IV** +ssRNA; **V** −ssRNA; **VI** rtRNA

TABLE 1 Most strongly avoided codon pairs across bacteria, archaea and eukaryotes

Codon pair	% of organisms which avoid it	O:E ratio
UUC GCA	86	0.570
GGG GGU	83	0.460
UUC GAA	82	0.590
CUU AUG	79	0.529
GCU AUG	76	0.590
ACU AUG	73	0.611
GUU AGC	73	0.529
CUU AGU	73	0.521
UUC GCG	72	0.559
GUU AUG	72	0.611

Source: Adapted from Tats et al. (2008).

"codon pair bias." Codon pairs may occur at different frequencies to those expected given the individual codon frequencies within a proteome (Buchan & Stansfield, 2005; Irwin et al., 1995) and in many organisms, some codon pairs are heavily underused, or "disfavoured." The phenomenon was first described in 1985 in *Escherichia coli* (Yarus & Folley, 1985) and has since been summarized for three domains of life (bacteria, archaea, and eukaryotes) (Tats et al., 2008) (Table 1).

Codon pair biases impact translation elongation rate (Gamble et al., 2016). In bacteria, over-represented codon pairs are translated more slowly than under-represented codon pairs (Irwin, Heck, and Hatfield 1995). Conversely, in eukaryotic cells, 17 specific codon pairs impede translation (Table 2), and reversing their order abrogates the effect (Gamble et al., 2016). These 17 codon pairs were all associated with wobble decoding interactions—that is, a non-Watson–Crick interactions between the third base of the codon and the first base of the tRNA anticodon. None of these codon pairs are common to those listed in Table 1.

Codon pair biases have also been linked with determining efficiency of protein folding and the co-ordinated expression of functionally grouped proteins (reviewed in Novoa & Ribas de Pouplana, 2012).

TABLE 2 Codon pairs which are inefficiently translated and associated with wobble decoding

Codon pair	First codon wobble	Second codon wobble
AGG CGA	—	I·A
AGG CGG	—	—
AUA CGA	—	I·A
AUA CGG	—	—
CGA AUA	I·A	—
CGA CCG	I·A	U·G
CGA CGA	I·A	I·A
CGA CGG	I·A	—
CGA CUG	I·A	U·G
CGA GCG	I·A	U·G
CUC CCG	—	U·G
CUG AUA	U·G	
CUG CCG	U·G	U·G
CUG CGA	U·G	I·A
GUA CCG	—	U·G
GUA CGA	—	I·A
GUG CGA	—	I·A

Note: I·A, inosine base pairing with adenine; U·G, uracil base pairing with guanine.
Source: Adapted from Gamble et al. (2016).

The first study of codon pair bias deoptimization of a virus genome determined that in poliovirus, artificially introduced rare codon pairs (relative to host) were translated more slowly (Coleman et al., 2008); this finding has been recapitulated in other virus systems including Marek's disease herpesvirus (Eschke et al., 2018) and IAV (Groenke et al., 2020).

We have described four different types of bias observed in genomes of organisms and the viruses that infect them—nucleotide bias, codon bias, dinucleotide bias and codon pair bias. Let us reconsider the HIV-1 genome—the A base is highly over-represented, occurring with a frequency of ∼37% (Kypr & Mrázek, 1987). If we did not know the underlying mechanism causing this bias, we may have difficulty determining which type of bias we were looking at, because all four may look similar (Figure 4). In order to deconvolute these types of bias, we need to understand the underlying mechanisms underlying their presence in more detail.

3 | DRIVERS OF VIRAL GENOME COMPOSITIONAL BIAS

As described above, genomic composition biases may arise through a variety of evolutionary selection pressures, both positive and negative. These potential drivers of bias are summarized below and in Figure 5.

3.1 | Biases driven by factors influencing translational rate

The efficiency with which different codons and codon pairs are translated in resting cells compared with stressed cells (e.g., during virus infection) varies depending on the tRNA pool available (Buchan et al., 2006). In a study that examined translational efficiency of a library of 217 synonymously recoded GFP sequences, codon usage and GC content of genes were both found to influence translational efficiency, mRNA splicing efficiency and mRNA subcellular localization (Mordstein et al., 2020). In addition, in resting cells, high GC content of a gene increases its transcriptional rate (Kudla et al., 2006). Whether these features influence the translational efficiency of viral genes, and whether viral genes have evolved specific composition traits to regulate transcription and translation, is unknown, but the hypotheses are reasonable. Use of codons or codon pairs which require wobble decoding is known to increase the likelihood of mistranslation events (Patil et al., 2012),
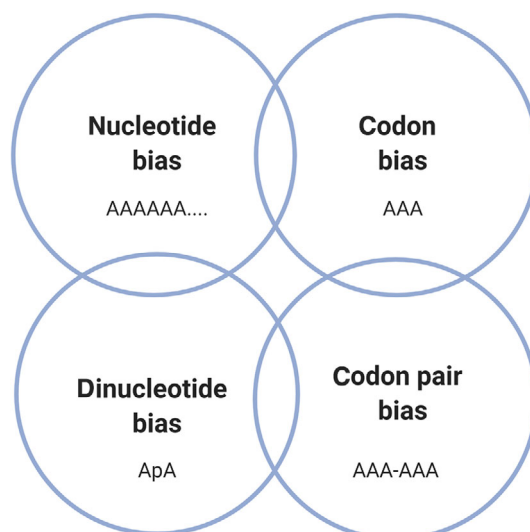
FIGURE 4 Four types of bias are described in the genomes of organisms and the viruses they are infected with
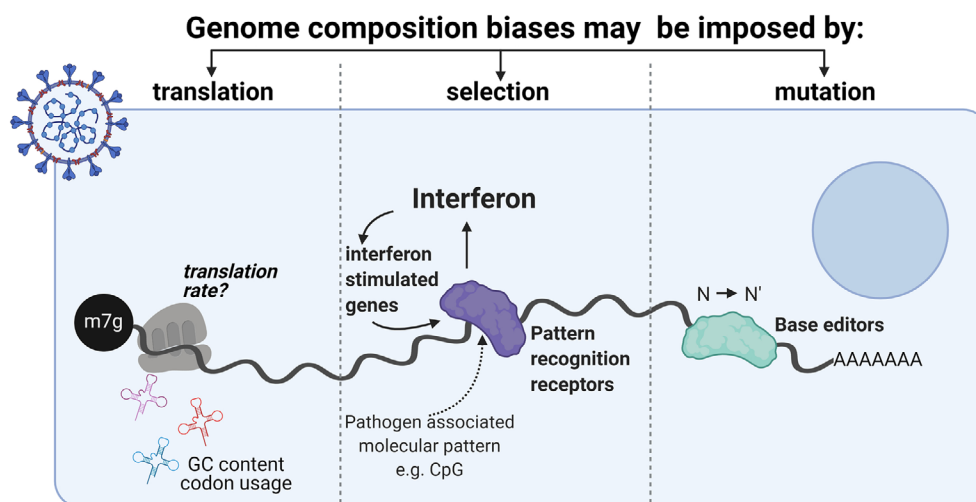

FIGURE 5 Compositional biases in viral genomes may be driven by three types of evolutionary pressure—Translational, selection and mutational. Translationally derived biases arise due to the different translational efficiencies of transcripts with varying composition in different cell conditions (e.g., resting vs. stress). Biases driven by selection arise through viral genomes avoiding encoding specific motifs that may be recognized by components of the innate immune response. Biases driven by mutation arise through editing of viral genomes or transcripts by host cell proteins

and mistranslation events are more frequent during cellular stress (Mohler & Ibba, 2017). Wobble decoding contributes to increased access to alternative reading frames (Drummond & Wilke, 2009; Ou et al., 2019), and so may be relevant for viruses which encode overlapping reading frames, but whether these events are physiologically important for viral replication is also unknown. Translational fidelity can nevertheless shape virus evolution (Ou et al., 2019); for example, some mitochondrially replicating mitoviruses avoid use of tryptophan codons, which mirrors avoidance of their use by the host fungi organelle mitochondrial genome (Nibert, 2017). RNA modifications (e.g., m^6A methylation) may also regulate translation (reviewed elsewhere; Roundtree et al., 2017) and the frequency of such modifications is related to biases in individual base frequencies.

3.2 | Biases driven by factors influencing mutation

Mutations arise in viral genomes either through the actions of host cell editors (i.e., direct *mutation)*, or by copying errors that then become fixed in the viral genome (*selection*). We have already considered the A-rich genome of HIV-1,

and understand that this has arisen due to the *mutational* activities of the cellular protein APOBEC3G. Similarly, the cellular proteins of the adenosine deaminase acting on RNA (ADAR) family convert adenosine to inosine; evidence for ADAR acting on virally derived nucleic acids was first reported in the genome of vesicular stomatitis virus (O'Hara et al., 1984) but has since been identified in the genomes of a range of other viruses (Samuel, 2012). There are numerous other APOBEC and ADAR family members with potential to act on viral genomes (Christofi & Zaravinos, 2019). The observation that the SARS-CoV-2 genome is extremely uracil-rich (Rice et al., 2020; Simmonds, 2020) has been speculatively attributed to the editing efforts of cellular mutators such as APOBEC (originally reported to edit DNA, but also reported to act on RNA; Sharma et al., 2016) and ADAR (Simmonds, 2020; Di Giorgio et al., 2020), but could also be attributable to an as-yet unidentified cellular protein.

3.3 | Biases driven by factors influencing selection

Selection pressure might arise also due to the activities of a cellular protein that, for example, recognizes a specific viral motif or pathogen-associated molecular pattern (PAMP). In general, recognition of a viral PAMP by a host cell protein (or a "pattern recognition receptor"; PRR) triggers type I interferon signaling; these PRRs may themselves be upregulated by interferon, and in this case are known as interferon stimulated genes (ISGs) (reviewed in [Kumar et al., 2011]). The concept of PAMPs being recognized by PRRs during the innate immune response was first hypothesized by Charles Janeway in 1989 (Janeway, 1989). As he predicted, the first PRR identified was Xa21, a gene that protects rice from bacterial infection (described in 1995) (Song et al., 1995). Of the many current examples of PRRs, some recognize specific viral nucleic acid signatures and thus may contribute to driving genome compositional biases. The 10 Toll-like receptors (TLRs) identified in humans are heavily evolutionarily conserved across vertebrates (Oshiumi et al., 2008) and some can recognize pathogen nucleic acids. The clearest example of this relevant to compositional biases is that TLR9 recognizes unmethylated CpG motifs in DNA (Bauer et al., 2001; Krug et al., 2004; Tabeta et al., 2004), and genomic suppression of CpG in murine herpesvirus 68 to evade detection by TLR9 has been reported (Pezda et al., 2011). Examples for RNA viruses are less clear-cut, but TLR7 recognizes purine-rich viral ssRNA (Gantier et al., 2008; Zhang et al., 2016). Thus, deselection of these PAMPs over evolutionary time may be due to the selection pressures applied by these PRRs, as well as as-yet-unidentified cellular factors.

3.4 | Mechanistic understanding of how viral CpGs are selected against

The suppression of CpG dinucleotides in the genomes of viruses and their hosts illustrates a fascinating contrast between mutational versus selection pressure. As described above, over evolutionary time the deamination of methylated CpG motifs in vertebrate genomes has resulted in their removal by mutation (biases driven by mutation). Viral mimicry of genomic CpG suppression was hypothesized to be due to aberrant CpG frequency sensing by an as yet unidentified PRR (Atkinson et al., 2014), and thus CpG motifs had been deselected in viral genomes (biases driven by selection). This hypothesis was strengthened in 2017, when a breakthrough paper reported that the product of the cellular ISG, zinc-finger antiviral protein (ZAP) senses CpG motifs in viral RNA (Takata et al., 2017). ZAP has long been identified as a suppressor of some but not all viruses by inducing degradation of specific viral mRNAs through an unknown targeting mechanism (Gao et al., 2002; Guo et al., 2007; Bick et al., 2003; Zhu et al., 2011). This more recent study used the HIV-1 genome as a model system in which to synonymously enrich CpG frequencies, and while the mutant virus was replication defective in normal cells, that defect was fully abrogated in a ZAP knockout system (Takata et al., 2017). Similarly, enrichment of CpGs in the echovirus 7 genome also caused a replication defect, that could be restored by ZAP knockout (Odon et al., 2019). Similarly, an inhibitory role for ZAP against human cytomegalovirus has been shown, which correlated with CpG-content dependent inhibition of viral Immediate Early 1 protein expression (Lin et al., 2020), further strengthening evidence that ZAP acts as an antiviral PRR though sensing high CpG frequencies in viral mRNAs.

ZAP is encoded on the ZC3HAV1 gene, which generates multiple isoforms via alternative splicing. Two isoforms are expressed to levels readily detectable by western blotting: the long (ZAPL) and short (ZAPS) forms (Li et al., 2019). From the N terminus, both major isoforms incorporate four zinc fingers implicated in RNA binding (Guo et al., 2004), a TiPARP Homology (TPH) domain, also containing a zinc finger (Kerns et al., 2008), and a WWE domain predicted to mediate interactions with proteins that facilitate post-translational conjugations (Aravind, 2001). In comparison with

ZAPL, ZAPS lacks the catalytically inactive poly(ADP-ribose) polymerase (PARP)—like domain, which enhances antiviral activity against an alphavirus and a retrovirus (Kerns et al., 2008). ZAPL is considered to be the constitutively expressed isoform, whereas ZAPS is an ISG which itself triggers IFN (Hayakawa et al., 2011; Ryman et al., 2005; Marcello et al., 2006) and is implicated in CpG recognition (Takata et al., 2017). Accordingly, here we only consider ZAPS (and refer to it simply as "ZAP"). The original paper reporting ZAP as a CpG sensor demonstrated the specific binding of ZAP at CpG sites using cross-linking followed by immunoprecipitation (CLIP) and sequencing (Takata et al., 2017). Crystallographic resolution of the structure of the N-terminus of ZAP bound to CpG motif-containing RNA revealed that the four zinc fingers of ZAP fold in a specific architecture to enable extensive RNA interactions which were diminished by mutation either of RNA CpG sites, or of ZAP at the zinc finger motifs (Luo et al., 2020; Meagher et al., 2019).

Following ZAP recognition of CpG-containing RNA, antiviral activity arises by inhibition of virus gene expression, either by mRNA degradation and/or inhibition of translation (Guo et al., 2007; Zhu et al., 2011). ZAP may inhibit translation by disrupting interactions between the translation initiation factors eIF4A and eIF4G (Zhu et al., 2012). ZAP also recruits transcripts to stress granules (Law et al., 2019). Degradation of viral mRNA is thought to occur through multiple routes, including via recruitment of the RNA exosome complex and/or the major cytoplasmic exoribonuclease, Xrn1 (Guo et al., 2007; Goodier et al., 2015; Todorova et al., 2014; Zhu et al., 2011). ZAP directly interacts with several exosome components, and their depletion by siRNA knockdown resulted in diminished antiviral activity by ZAP (Guo et al., 2007), confirming an essential role for the exosome in ZAP-mediated RNA degradation. During exosome-mediated RNA degradation, mRNAs must be deadenylated and then decapped to yield a monophosphorylated RNA, which can then also be digested by Xrn1 (Chang et al., 2019). Interactions between ZAP and poly-A specific ribonuclease (PARN) may direct deadenylation of the mRNA, while interactions between Xrn1 and the decapping enzymes necessary for $5' \rightarrow 3'$ RNA degradation are indirect, via the RNA helicase DDX17(Zhu et al., 2011). Xrn1 also digests endonucleolytically cleaved RNAs (Gatfield & Izaurralde, 2004), but it is not definitively known whether ZAP binding leads to internal mRNA cleavage events. In support of this possibility, ZAP binds to and its inhibitory activity against CpG-enriched transcripts is dependent on the cellular protein KHNYN, which unlike ZAP, does possess endonuclease activity (Ficarelli et al., 2020; Ficarelli et al., 2019). This is summarized (Figure 6).

How ZAP feeds back into the interferon pathway is uncertain. ZAP has been shown to interact with the cytoplasmic PRR RIG-I and to augment innate immune signaling in response to a variety of artificial RNA stimuli (Hayakawa et al., 2011). This study was performed prior to ZAP's identification as a CpG sensor however, and focussed on recognition of 3′-triphosphate RNA moieties; it remains to be determined if CpG-rich RNA signals through the same mechanisms.

Alternatively, ZAP-mediated innate immune responses may themselves be mediated through interactions with ZAP's cofactor TRIM25. ZAP is directly bound by TRIM25, itself an RNA binding protein and also an E3 ubiquitin ligase (Zou & Zhang, 2006), and this interaction is required for ZAP's antiviral activity (Li et al., 2017; Zheng et al., 2017). TRIM25 binds ZAP through TRIM25's SPRY domain (a protein interaction module characterized by a
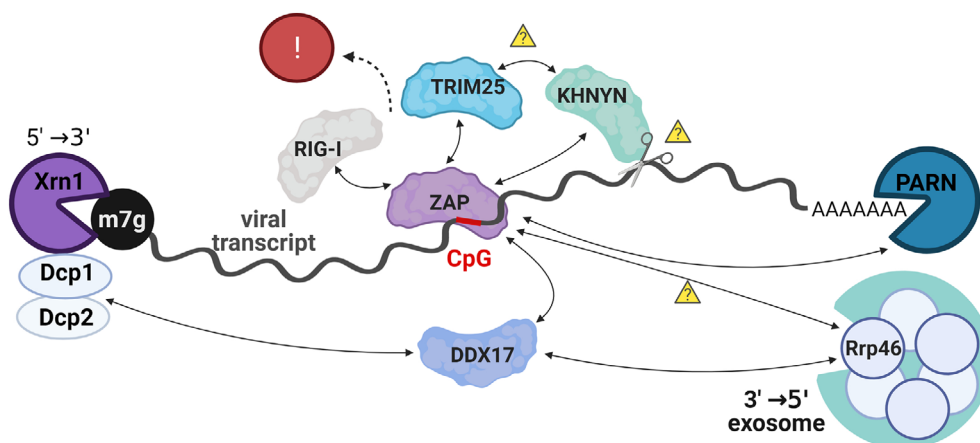

FIGURE 6 Possible mechanisms by which ZAP activity leads to viral transcript degradation. CpG motifs in viral RNA (red) are bound by the cytoplasmic PRR ZAP, which can lead to recruitment of 5′ decapping enzymes (Dcp1/2 complex), the 3′ deadenylation enzyme PARN and potentially the KHNYN RNA endonuclease, followed by 5′–3′ degradation mediated by Xrn1 and/or 3′–5′ degradation mediated by the RNA exosome. Interactions between ZAP and RIG-I and/or TRIM25 may also lead to innate immune signaling

sequence repeat; D'Cruz et al., 2013) and ubiquitinates ZAP, although ubiquitination is not required for ZAP antiviral activity (Choudhury et al., 2017). TRIM25 was originally understood to be essential for activation of the RIG-I-dependent pathway for interferon activation (Gack et al., 2007), but recently it was shown that RIPLET and not TRIM25 ubiquitinates RIG-I, and that RIPLET is sufficient for the ubiquitination and activation of RIG-I (Cadena et al., 2019). It is therefore unclear how important TRIM25 (or by extension, the interaction between ZAP and TRIM25) is during virally induced activation of the interferon response.

CpG suppression may be more nuanced than the blanket genome-wide suppression described above, which has consequent implications for the mechanisms of, and viral counteractivity to, ZAP. In the genomes of Betaherpesviruses, immediate early genes suppress CpG, whereas this is not seen in the rest of the genome (Lin et al., 2020). The authors hypothesized that immediate early gene product(s) are able to abrogate ZAP activity, thus removing any selection against high CpG frequencies in viral genes that are activated at later timepoints during infection. Conversely, in the SARS-CoV-2 genome, CpG is over-represented in E (envelope) ORF and in ORF10, whereas other genes—as expected—suppress CpG (Digard et al., 2020; Rice et al., 2020). Why these ORFs are able to buck the trend seemingly imposed on the rest of the genome is unknown; possibly, high CpG frequencies invite turnover by ZAP, thereby regulating protein production. Alternatively, these ORFs may have been acquired through recombination events and had an ancestral origin not previously subject to the same translational, mutational or selection pressures.

3.5 | CpG context may be an important driver of biases imparted by selection

For ZAP to function as an innate immune sensor and/or effector for foreign RNAs containing high CpG content, there must be a mechanism to limit activation of the system by cellular RNAs that also contain CpG dinucleotides (as all do). Since ZAP recognizes CpG motifs in ssRNA, it is possible that secondary structure of RNA—i.e., CpG context, is an important factor in determining whether CpG motifs can be recognized by ZAP, and there is some evidence indicating this. First, in the crystallography paper characterizing ZAP-RNA binding, the optimal binding motif for ZAP on RNA was found to be $C(n_7)G(n)CG$ (Luo et al., 2020). ZAP was found to bind to multiple sites on an RNA, and in considering the stoichiometry of RNA degradation complex recruitment, the authors concluded that owing to the relatively small size of ZAP relative to RNA degradation complexes, several bound ZAP molecules must be required for this. Therefore the number and spacing of CpG dinucleotides is likely to be important.

Context effects for CpG deselection have also been identified in an evolutionary context. Greenbaum et al., found that since the emergence of the 1918 H1N1 pandemic strain of IAV in humans, CpG motifs have gradually been lost from the viral genome as it became endemic in humans. They asked whether specific nucleotides were more likely to flank the CpGs that were deselected, by measuring the relative frequencies of (C/G)CG(C/G), (A/U)CG(A/U), (A/U)CG (C/G) and (C/G)CG(A/U) in H1N1 genomes over time (Greenbaum et al., 2009). No reduction in (C/G)CG(C/G) motifs was seen, whereas all three of the other motifs declined in frequency, with the strongest reduction seen in the (A/U)CG (A/U) motif. The authors speculated that the severe disease attributed to infection with the 1918 virus was caused by the aberrantly high CpG frequency present in the viral genome provoking a cytokine storm.

A similar observation has been recapitulated in vitro. Using echovirus 7 as a model system, a replicon was recoded to maintain CpG frequency ($n = 51$) but add AACGAA or UUCGUU motifs (Fros et al., 2017). The UUCGUU mutant was fivefold more impaired than a CpG enriched transcript (Fros et al., 2017). Thus, there is a growing body of evidence that CpG context is important for innate sensing.

3.6 | UpA dinucleotide sensing as a driver of bias

Two possible explanations have been put forward to date to explain genomic UpA suppression. First, it was originally reported in 1981 (and subsequently verified) that UpA dinucleotides are cleaved by the cellular ISG RNaseL (Wreschner et al., 1981; Karasik et al., 2021), which could explain their deselection over evolutionary time. However, the authors further reported that RNaseL also cleaves RNA at UpU dinucleotides, and TpT/ UpU are generally not under-represented in animal genomes or in the viruses that infect them, so the specificity and impact of RNaseL on genomic TpA/UpA content is questionable. So far, one study using echovirus 7 as a model found that the reduced replication of an artificially UpA-enriched virus could be rescued by RNaseL removal (Odon et al., 2019), but it appears that the pathway is not specific to RNaseL, as ZAP depletion also complemented the defect in virus replication. While both
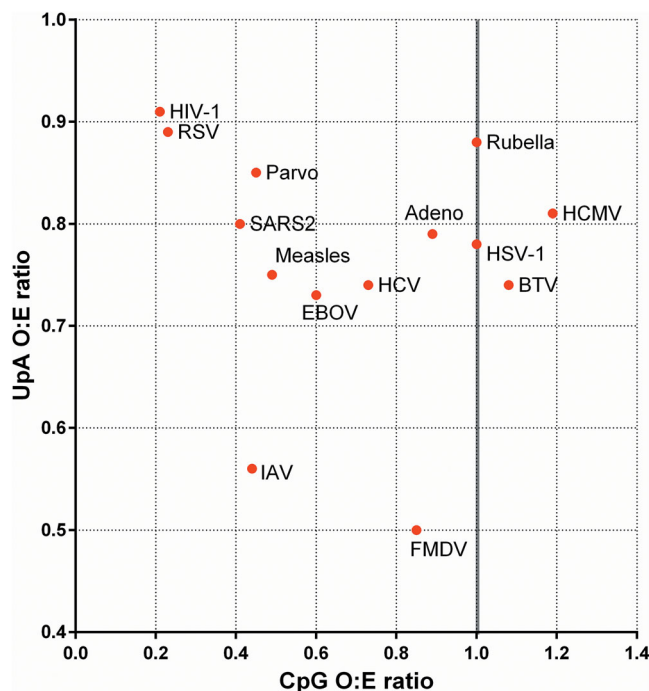
FIGURE 7   Comparison of CpG and UpA suppression in the genomes of various viruses. RNA viruses: BTV, bluetongue virus; EBOV, ebola virus; FMDV, foot and mouth disease virus; HCV, hepatitis C virus; RSV, respiratory syncytial virus; SARS2, severe acute respiratory syndrome coronavirus 2. DNA viruses: adeno, adenovirus; HCMV, human cytomegalovirus; HSV-1, herpes simplex virus 1; Parvo, canine parvovirus 2

CpG and UpA dinucleotide suppression may be driven by co-regulated factors of the interferon response, the extent of CpG and UpA suppression within a virus genome do not necessarily correlate (Figure 7).

The second nonexclusive idea to explain TpA/ UpA suppression is the propensity for this dinucleotide to introduce a stop codon. Stop codons are encoded by UAG, UAA and UGA nucleotide triplets, and so deselection of UpA motifs in the first and second codon positions reduces the risk of aberrant stop codon introduction. However, 6 of 10 disfavoured codon pairs encode a UpA motif cross the codon boundary (Table 1), and so deselection of UpA motifs in this context is evident and may therefore be important for translation regulation. Therefore, multiple constraints may be acting which, together, reduce UpA representation in the genomes of organisms and their infecting viruses.

4 | CONSEQUENCES OF ALTERING VIRAL GENOME COMPOSITIONAL BIASES

To study the biological relevance of under-represented nucleotides, dinucleotides, codons and codon pairs, synonymous recoding has been undertaken for a wide range of viruses. These are summarized (Table 3). In these studies, deoptimization to alter sequence composition in a direction away from that of the host, or optimization to recode viral sequence to look more like host genome has been undertaken. Generally, deoptimization of any of these parameters results in virus attenuation, whereas optimization usually does not improve replication.

4.1 | Codon pair bias recoding

The first study (published in 2008) to draw significant attention to the subject of large scale genome recoding examined the effects of modifying codon pair bias in the poliovirus genome (Coleman et al., 2008), where deoptimization of codon pairs resulted in virus attenuation, and the extent of recoding correlated with the extent of attenuation. The authors found that introduction of disfavoured codon pairs decreased protein translation rates (assayed using a luciferase reporter construct) and yielded viruses that were attenuated in mice, but still offered protection from homologous virus

TABLE 3 Synonymous recoding strategies which have been applied to RNA viruses are summarized

Virus	Recoding strategy	Region recoded	Findings	References
Adeno-associated virus	Codon pair bias deoptimization	Rep	The negative regulatory signal imparted on adenovirus by AAV was diminished, and so adenovirus replication was enhanced	Sitaraman et al. (2011)
Dengue virus	Codon pair bias deoptimization to match insect bias	E/NS3/NS5	Mutants grow well in insect cells but not well (if at all) in mammalian cells. LD50 was $10^{2-3.5}$ fold up in mice	Shen et al. (2015)
	Bioinformatic analyses showed that the above recoding strategy also increased CpG frequency		This re-analysis suggested that attenuation of viral replication in mammalian cells might result from increased CpG content rather than increased codon-pair bias	Simmonds et al. (2015)
Echovirus 7	CpG or UpA dinucleotide bias optimization and deoptimization	VP3/1 and/or 3B/C/D	CpG enrichment in two regions caused a 7000-fold reduction in replication; UpA enrichment caused a 30-fold reduction in cells. Removal of CpGs and UpAs increased replication, with removal of both increasing virus titres 10-fold in cells	Atkinson et al. (2014)
Foot and mouth disease virus	Codon pair bias deoptimization	P1 capsid	10^3-fold increase in the vaccine safety margin compared with WT virus	Diaz-San Segundo et al. (2016)
Human cytomegalovirus	CpG dinucleotide deoptimization	IE1	Reporter constructs with elevated CpG content triggered ZAP induction	Lin et al. (2020)
Human immunodeficiency virus 1	Codon pair optimization and deoptimization	Gag and pol	No observed effects of optimization; deoptimization reduced replication titre in cells. Deoptimized but not optimized virus reverted following passage	Martrus et al. (2013)
	Increased CpG frequency	Gag	Up to 10^2-fold defect in replication in cells	Antzin-Anduetza et al. (2017)
Influenza A virus	Codon pair bias deoptimization	PB1, NP and HA	10^1-fold reduction in titre in cells	Mueller et al. (2010)
	Codon pair bias deoptimization	HA and NA	10^5-fold attenuation in mice and clinical attenuation in ferrets	Yang et al. (2013) and Broadbent et al. (2016)
	CpG and UpA dinucleotide deoptimization	NP	10^{1-2}-fold reduction in titre in cell culture and disease attenuation in mice	Gaunt et al. (2016)
	CpG and codon pair bias deoptimization	NA	Codon pair bias dramatically decreased replication whereas increased CpG dinucleotides did not	Groenke et al. (2020)

(Continues)

TABLE 3   (Continued)

Virus	Recoding strategy	Region recoded	Findings	References
Poliovirus	Codon usage bias deoptimization	Capsid	65 fold reduction in virus titre in cells	Burns et al. (2006)
	CpG and UpA dinucleotide deoptimization	Capsid	Up to a 10^3-fold reduction in virus titre in cells	Burns et al. (2009)
	Codon usage optimization and deoptimization	Capsid	Little effect with codon optimization; deoptimization reduced virus titre in cells and mice	Lauring et al. (2012)
	Codon pair bias deoptimization	Capsid	Replication defect correlated with extent of mutagenesis in cells	Coleman et al. (2008)
Porcine reproductive and respiratory syndrome virus	Codon pair bias deoptimization	GP5	A 10-fold replication defect in cells, 10^3-fold decrease in virus titre in pigs	Ni et al. (2014)
	Codon pair bias deoptimization	NSP9	10^4-fold replication defect in cells, no evidence of infection in pigs	Gao et al. (2015)
Potato virus Y	CpG and UpA dinucleotide deoptimization	Nonstructural genes	Up to 10^3-fold defect (CpG) or 10^6-fold defect (UpA) in systemic spread	Ibrahim et al. (2019)
Respiratory syncytial virus	Codon pair deoptimization	Various combinations, with the most extensive recoding extending to all ORFs except M1 and M2	Multiple $\log_{10}$-fold reduction in titre of various mutants in cells, mice and African Green Monkeys	Le Nouën et al. (2014)
	Codon deoptimization by altering codon usage to be consistent with human	NS1 and NS2	Modest replication attenuation in cells and mice	Meng et al. (2014)
Simian immunodeficiency virus	Nucleotide optimization towards nucleotide frequencies in macaque	Gag and pol	10^2-fold decrease in replication in cells; recoding in polymerase only had no effect	Vabret et al. (2014)
Vesicular stomatitis virus	Codon pair bias optimization and deoptimization	Polymerase	Optimization resulted in a modest replication defect in cells and 10^{2-3}-fold deficit in mice. Deoptimized virus could not be recovered	Wang et al. (2015)

challenge. Following on from this work, several papers (by the same research group and others) expanded this concept by applying the same recoding strategy to other viruses (Table 3).

The potential for codon pair bias recoding as a vaccine development strategy has also been demonstrated in work using IAV as a model system. The PR8 strain of H1N1 IAV (the backbone of which is used to make live attenuated IAV vaccines) was recoded to increase disfavoured codon pair usage. Three viral genome segments (2 [PB1], 4 [HA], and 5 [NP]) were modified and tested separately or in combination for their effects on viral growth characteristics in cells and vaccine potential in mice. Single or combinatorial segment modifications all displayed around 10-fold defect in multicycle replication assays in vitro. However, in BALB/c mice, the triple reassortant had a 3000-fold reduction in virus titre at 24 h post-infection. The triple reassortant virus was further tested for its 50% protective dose (PD$_{50}$; i.e., the inoculum dose required to protect from infection upon challenge), displaying a 50% lethal dose (LD$_{50}$)/PD$_{50}$ ratio 1000-fold higher than that of wildtype PR8 virus. This result, and others from the same lab (Yang et al., 2013;

Broadbent et al., 2016) emphasized the potential of large scale genome recoding as an approach to live-attenuated vaccine development.

A question often posited when large scale recoding of viruses to either mimic or deviate from the patterns seen in host genomes is considered, is what happens in the case of vector-borne viruses that replicate in both invertebrate and vertebrate hosts. This was investigated for dengue virus, which replicates effectively in both the main insect vector *Aedes aegypti*, and in humans (Olson et al., 1996). Recoding of the dengue virus genome to align its codon pair usage in favor with insect genome preferences resulted in a virus that replicated as well as wildtype in insect cells, but experienced a 1–2 $\log_{10}$ decrease in replication in some mammalian cells. In mice, this recoding resulted in a 2–3 $\log_{10}$ increase in LD_{50}. Curiously, the recoded virus replicated normally in BHK-21 cells.

4.2 | Dinucleotide recoding

Large scale recoding of virus genomes using dinucleotide deoptimization was first reported in 2006, using poliovirus as a model system (Burns et al., 2006). In this paper, the authors set out to recode poliovirus by deoptimizing codon usage, but observed that in the process, they introduced 207 CpG dinucleotides across the capsid region. In doing this, virus titres were reduced 65-fold.

The same group went on to specifically study the impact of CpG and UpA enrichment on poliovirus replication. Addition of CpG or UpA were both found to diminish replication, and when both dinucleotide frequencies were simultaneously increased, the effects were found to be synergistic (Burns et al., 2009).

These first papers investigating dinucleotide deoptimization were confounded by lack of corrections for nucleotide and codon usage biases, which as noted above (Figure 4), are inter-related. More recent works have enriched CpG and UpA dinucleotides without altering nucleotide or codon frequencies. The introduction of CpGs or UpAs into the echovirus 7 genome in a way that controlled for these other variables (Atkinson et al., 2014) found that CpG introduction more strongly reduced virus fitness than UpA introduction. Conversely, using IAV as a model system, UpA introduction was more detrimental to virus replication than CpG addition. This IAV work also demonstrated that a sub-clinical dose of CpG-enriched virus protected from challenge with a potentially lethal dose of the wildtype PR8 strain in mice (Gaunt et al., 2016), directly demonstrating the potential of dinucleotide deoptimization as a vaccine development strategy.

CpG and UpA dinucleotide optimization and deoptimization have been characterized in the genomes of various other viruses (Table 3), but the bulk of these studies were undertaken prior to the discovery of ZAP as a CpG sensor. The defect imparted by CpG enrichment has been abrogated by ZAP knockout in echovirus 7 (Odon et al., 2019) and HIV-1 systems (Ficarelli et al., 2020; Takata et al., 2017), although for echovirus 7 the defect was also relieved by RNaseL knockout, and CpG enrichment in HIV-1 impacted splicing events. The role of ZAP in CpG sensing of viral RNAs requires further clarification.

4.3 | Is codon pair bias an artifact of dinucleotide bias?

Of the top 10 most avoided codon pairs across bacteria, archaea and eukaryotes, three contain a CpG motif at the codon boundary and six contain a UpA motif (Table 1). Thus, the two phenomena of codon pair bias and dinucleotide bias are interlinked and at one extreme could simply be two ways of measuring the same effect. This has proven to be a contentious issue (Simmonds et al., 2015; Futcher et al., 2015). Deconvoluting the two has been achieved using the echovirus 7 system to make a panel of mutants which were either codon pair bias deoptimized or dinucleotide bias deoptimized, without altering the other parameter. Using this system, codon pair bias did not impact virus replication kinetics, whereas dinucleotide composition did (Tulloch et al., 2014). This finding was supported by a bioinformatics study from an independent laboratory which reached the same conclusion (Kunec & Osterrieder, 2016). However, when the same authors from this latter study used IAV as a model system to experimentally test their predictions, they found that codon pair bias was far more important than dinucleotide bias, using IAV as a model system (Groenke et al., 2020). In this latter report, the authors found that codon pair deoptimization resulted in diminished mRNA stability (Groenke et al., 2020); however, the codon pair deoptimization resulted in increased UpA dinucleotide frequencies, and as UpA is reported to be cleaved by RNaseL (Wreschner et al., 1981; Karasik et al., 2021), this could also explain the outcome. A bioinformatics study that used nucleotide patterns of viruses to predict host species found the two features to be

discrete, but that dinucleotide bias was far more accurate than codon pair bias in identifying viral host species (Babayan et al., 2018).

The field therefore remains divided about whether codon pair and dinucleotide bias are synonyms or are discrete phenomena. The confusion between codon pair bias and dinucleotide bias is compounded by a proportion of the described studies not including control constructs—e.g., re-ordering codons without altering dinucleotide frequencies (Atkinson et al., 2014; Fros et al., 2017; Gaunt et al., 2016; Ibrahim et al., 2019; Odon et al., 2019). Such controls are imperfect—perhaps, for example, a deoptimized virus has inadvertently introduced a mutation that alters an uncharacterized RNA functional element, whereas the recoding in the control virus did not. Such controls are nevertheless still helpful for strengthening mechanistic conclusions, even if redundant for purely empirical attempts to attenuate a virus. Furthermore, not all studies fully investigate the mechanism of attenuation—for example, few have assessed translation rates (impacted by codon pair biases) or RNA turnover (impacted by dinucleotide biases). Ultimately, better control strategies are needed to deconvolute these two phenomena properly. Distinction—or not—between dinucleotide bias and codon pair bias can be made if we fully understand the mechanism(s) by which these biases attenuate virus propagation.

The discovery of ZAP as a CpG sensor provides the opportunity for researchers to validate CpG enrichment studies. If CpG enrichment results in a defect which can be abrogated with ZAP knockout, the impairment phenotype can sensibly be concluded to be a result of ZAP activity and therefore the introduction of CpGs (rather than an unintended side effect such as introduction of disfavoured codon pairs). Publications on this so far report mixed results (Ficarelli et al., 2020; Odon et al., 2019; Ficarelli et al., 2019) wherein ZAP is not the only sensor whose depletion results in fitness reconstitution, or off-target effects of CpG enrichment are seen. Nevertheless, ZAP knockout looks like a promising test of whether CpG enrichment is the key to why a CpG-enriched virus has a replication defect. However, no such "rescue system" exists for codon pair bias studies. Perhaps, if the limitation is in translational efficiency, tRNA supplementation would rescue the system, but we are not aware of any study attempting this.

5 | DISCUSSION: POTENTIAL FOR DINUCLEOTIDE MODIFICATION AS A VACCINE DEVELOPMENT STRATEGY

The detailed observations on the large-scale recoding of RNA virus genomes has enthused researchers to repeatedly suggest that these methods may offer a potential live attenuated vaccine development strategy, as described above for both codon pair bias and dinucleotide bias deoptimization.

A critical consideration for live attenuated vaccine development, regardless of the virus system being explored, is virus yield. For a successful vaccine candidate, it must be possible to produce that vaccine virus in high amount. However, the described large-scale recoding strategies, while attenuative and in some cases protective from heterologous virus challenge, result in marked defects in virus production levels. The discovery of ZAP as a CpG sensor could provide a potential route to circumvent this issue. Vaccine candidate viruses can simply be grown in ZAP knockout systems, thus recapitulating wildtype virus titres—assuming other unintended effects of mutagenesis (e.g., on genome replication and/or packaging as discussed above) are avoided.

Let us consider the example of IAV as a candidate virus for which a large scale recoded virus could be developed for vaccination. IAV live attenuated vaccines are most commonly produced in embryonated hen's eggs, although there is a motion to switch production to cell culture-based systems (Perdue et al., 2011). A CpG enriched virus could be—and has been (Gaunt et al., 2016)—produced. One can therefore envisage synthesis of a CpG enriched IAV that replicates to wildtype levels in a CpG-sensor knockout system (whether this virus is manufactured in a cell culture system or in embryonated hen's eggs; there is emerging technology for creating gene edited chickens; Long et al., 2019; Idoko-Akoh et al., 2018). This CpG-enriched virus may offer enhanced immunogenicity (Gaunt et al., 2016) and so the amount of vaccine virus required per dose might also be reduced.

IAV offers a very attractive vaccine target for which synonymous recoding must be a serious consideration. For IAV live attenuated vaccines, PR8 strain—which is nonpathogenic in humans—is used as the backbone, and it is straightforward to switch in synonymously recoded segments into this backbone using well established reverse genetics systems (Fodor et al., 1999; Neumann et al., 1999). By contrast, if we were to recode viable segments of a recombinant virus such as SARS-CoV-2 and use that as a vaccine (Digard et al., 2020), there is the risk that this virus can recombine and revert to virulence even with large scale recoding. For poliovirus, the same caveats apply, as well as concerns that this is a neurotropic virus and a CpG-enriched virus may not be subject to the same replicative losses in the immunoprivileged replication sites (Gao et al., 2015).

It is critical that we fully understand the mechanism of attenuation imparted by synonymous recoding before we apply this technology to vaccine development. For example, we can ask whether the addition of CpGs really allowed greater visibility of the virus to the innate immune system because of those additional CpG motifs, or has some unpredicted defect in replication been introduced—such as disruption or creation of an alternative open reading frame, packaging signals, splice junction, etc.? A candidate vaccine virus may be a lot closer to reversion than it appears in phenotyping as all these examples could potentially be overcome by a single reversion mutation. If this happened when a dinucleotide modified strain were used as a live attenuated vaccine this could create a virus adept at replication in humans, but which is highly immunogenic and therefore pathogenic.

6 | CONCLUSION

Three drivers shape the genome composition of viruses—translation, mutation and selection. These result in four types of bias—nucleotide, codon, dinucleotide and codon pair. Systematic recoding of viral genomes to disrupt these frequencies almost universally leads to virus attenuation. Synonymous recoding offers a highly attractive vaccine development strategy with the potential to overcome the yield issues currently thwarting current live attenuated vaccine production efforts. CpG dinucleotide deoptimization alone has available a rescue system in which a vaccine virus could be amplified to wildtype virus titres. No such system exists (yet) for any other deoptimization strategy. However, further work must be undertaken to fully understand the mechanisms impacted by this recoding before we can consider using this approach commercially.

AUTHOR CONTRIBUTIONS
Eleanor Gaunt: Conceptualization; data curation; formal analysis; writing-original draft; writing-review & editing.
Paul Digard: Conceptualization; writing-original draft; writing-review & editing.

DATA AVAILABILITY STATEMENT
Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID
Paul Digard https://orcid.org/0000-0002-0872-9440

RELATED WIREs ARTICLE
lncRNAs regulate the innate immune response to viral infection

REFERENCES
Abrahams, L., & Hurst, L. D. (2017). Adenine enrichment at the fourth CDS residue in bacterial genes is consistent with error proofing for +1 frameshifts. *Molecular Biology and Evolution*, *34*, 3064–3080.
Adams, M., & Antoniw, J. (2004). Codon usage bias amongst plant viruses. *Archives of Virology*, *149*, 113–135.
Antzin-Anduetza, I., Mahiet, C., Granger, L. A., Odendall, C., & Swanson, C. M. (2017). Increasing the CpG dinucleotide abundance in the HIV-1 genomic RNA inhibits viral replication. *Retrovirology*, *14*, 49.
Aragonès, L., Guix, S., Ribes, E., Bosch, A., & Pintó, R. M. (2010). Fine-tuning translation kinetics selection as the driving force of codon usage bias in the hepatitis A virus capsid. *PLoS Pathogens*, *6*, e1000797–e1000797.

Aravind, L. (2001). The WWE domain: A common interaction module in protein ubiquitination and ADP ribosylation. *Trends in Biochemical Sciences*, *26*, 273–275.

Atkinson, N. J., Witteveldt, J., Evans, D. J., & Simmonds, P. (2014). The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. *Nucleic Acids Research*, *42*, 4527–4545.

Auewarakul, P. (2005). Composition bias and genome polarity of RNA viruses. *Virus Research*, *109*, 33–37.

Babayan, S. A., Orton, R. J., & Streicker, D. G. (2018). Predicting reservoir hosts and arthropod vectors from evolutionary signatures in RNA virus genomes. *Science*, *362*, 577–580.

Bahir, I., Fromer, M., Prat, Y., & Linial, M. (2009). Viral adaptation to host: A proteome-based analysis of codon usage and amino acid preferences. *Molecular Systems Biology*, *5*, 311.

Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriological Reviews*, *35*, 235–241.

Balzarini, J., Camarasa, M. J., Pérez-Pérez, M. J., San-Félix, A., Velázquez, S., Perno, C. F., De Clercq, E., Anderson, J. N., & Karlsson, A. (2001). Exploitation of the low fidelity of human immunodeficiency virus type 1 (HIV-1) reverse transcriptase and the nucleotide composition bias in the HIV-1 genome to alter the drug resistance development of HIV. *Journal of Virology*, *75*, 5772–5777.

Bauer, S., Kirschning, C. J., Häcker, H., Redecke, V., Hausmann, S., Akira, S., Wagner, H., & Lipford, G. B. (2001). Human TLR9 confers responsiveness to bacterial DNA via species-specific CpG motif recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 9237–9242.

Belalov, I. S., & Lukashev, A. N. (2013). Causes and implications of codon usage bias in RNA viruses. *PLoS One*, *8*, e56642.

Berkhout, B., Grigoriev, A., Bakker, M., & Lukashov, V. V. (2002). Codon and amino acid usage in retroviral genomes is consistent with virus-specific nucleotide pressure. *AIDS Research and Human Retroviruses*, *18*, 133–141.

Bick, M. J., Carroll, J. W., Gao, G., Goff, S. P., Rice, C. M., & MacDonald, M. R. (2003). Expression of the zinc-finger antiviral protein inhibits alphavirus replication. *Journal of Virology*, *77*, 11555–11562.

Bird, A. P. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Research*, *8*, 1499–1504.

Bird, A., & Tweedie, S. (1995). Transcriptional noise and the evolution of gene number. *Philosophical Transactions: Biological Sciences*, *349*, 249–253.

Birnbaum, R. Y., Josephine Clowney, E., Agamy, O., Kim, M. J., Zhao, J., Yamanaka, T., Pappalardo, Z., Clarke, S. L., Wenger, A. M., Nguyen, L., Gurrieri, F., Everman, D. B., Schwartz, C. E., Birk, O. S., Bejerano, G., Lomvardas, S., & Ahituv, N. (2012). Coding exons function as tissue-specific enhancers of nearby genes. *Genome Research*, *22*, 1059–1068.

Bougraa, M., & Perrin, P. (1987). CpG and TpA frequencies in the plant system. *Nucleic Acids Research*, *15*, 5729–5737.

Bouquet, J., Cherel, P., & Pavio, N. (2012). Genetic characterization and codon usage bias of full-length hepatitis E virus sequences shed new lights on genotypic distribution, host restriction and genome evolution. *Infection, Genetics and Evolution*, *12*, 1842–1853.

Brest, P., Lapaquette, P., Souidi, M., Lebrigand, K., Cesaro, A., Vouret-Craviari, V., Mari, B., Barbry, P., Mosnier, J. F., Hébuterne, X., Harel-Bellan, A., Mograbi, B., Darfeuille-Michaud, A., & Hofman, P. (2011). A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nature Genetics*, *43*, 242–245.

Broadbent, A. J., Santos, C. P., Anafu, A., Wimmer, E., Mueller, S., & Subbarao, K. (2016). Evaluation of the attenuation, immunogenicity, and efficacy of a live virus vaccine generated by codon-pair bias de-optimization of the 2009 pandemic H1N1 influenza virus, in ferrets. *Vaccine*, *34*, 563–570.

Buchan, J. R., Aucott, L. S., & Stansfield, I. (2006). tRNA properties help shape codon pair preferences in open reading frames. *Nucleic Acids Research*, *34*, 1015–1027.

Buchan, R., & Stansfield, I. (2005). Codon pair bias in prokaryotic and eukaryotic genomes. *BMC Bioinformatics*, *6*, P4.

Burns, C. C., Campagnoli, R., Shaw, J., Vincent, A., Jorba, J., & Kew, O. (2009). Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons. *Journal of Virology*, *83*, 9957–9969.

Burns, C. C., Shaw, J., Campagnoli, R., Jorba, J., Vincent, A., Quay, J., & Kew, O. (2006). Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. *Journal of Virology*, *80*, 3259–3272.

Butt, A. M., Nasrullah, I., & Tong, Y. (2014). Genome-wide analysis of codon usage and influencing factors in chikungunya viruses. *PLoS One*, *9*, e90905.

Cadena, C., Ahmad, S., Xavier, A., Willemsen, J., Park, S., Park, J. W., Oh, S.-W., Fujita, T., Hou, F., Binder, M., & Hur, S. (2019). Ubiquitin-dependent and -independent roles of E3 ligase RIPLET in innate immunity. *Cell*, *177*, 1187–200.e16.

Cai, M. S., Cheng, A. C., Wang, M. S., Zhao, L. C., Zhu, D. K., Luo, Q. H., Liu, F., & Chen, X. Y. (2009). Characterization of synonymous codon usage bias in the duck plague virus UL35 gene. *Intervirology*, *52*, 266–278.

Cannarozzi, G., Schraudolph, N. N., Faty, M., von Rohr, P., Friberg, M. T., Roth, A. C., Gonnet, P., Gonnet, G., & Barral, Y. (2010). A role for codon order in translation dynamics. *Cell*, *141*, 355–367.

Chang, C.-T., Muthukumar, S., Weber, R., Levdansky, Y., Chen, Y., Bhandari, D., Igreja, C., Wohlbold, L., Valkov, E., & Izaurralde, E. (2019). A low-complexity region in human XRN1 directly recruits deadenylation and decapping factors in 5′–3′ messenger RNA decay. *Nucleic Acids Research*, *47*, 9282–9295.

Chen, Y. (2013). A comparison of synonymous codon usage bias patterns in DNA and RNA virus genomes: Quantifying the relative importance of mutational pressure and natural selection. *BioMed Research International*, *2013*, 406342.

Choudhury, N. R., Heikel, G., Trubitsyna, M., Kubik, P., Nowak, J. S., Webb, S., Granneman, S., Spanos, C., Rappsilber, J., Castello, A., & Michlewski, G. (2017). RNA-binding activity of TRIM25 is mediated by its PRY/SPRY domain and is required for ubiquitination. *BMC Biology*, *15*, 105.

Christofi, T., & Zaravinos, A. (2019). RNA editing in the forefront of epitranscriptomics and human health. *Journal of Translational Medicine*, *17*, 319.

Coffin, R. S., Howard, M. K., & Latchman, D. S. (1995). Altered dinucleotide content within the latently transcribed regions of the DNA of alpha herpes viruses—Implications for latent RNA expression and DNA structure. *Virology*, *209*, 358–365.

Coleman, J. R., Papamichail, D., Skiena, S., Futcher, B., Wimmer, E., & Mueller, S. (2008). Virus attenuation by genome-scale changes in codon pair bias. *Science*, *320*, 1784–1787.

Cooper, D. N., & Krawczak, M. (1989). Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Human Genetics*, *83*, 181–188.

D'Andrea, L., Pintó, R. M., Bosch, A., Musto, H., & Cristina, J. (2011). A detailed comparative analysis on the overall codon usage patterns in hepatitis A virus. *Virus Research*, *157*, 19–24.

D'Cruz, A. A., Babon, J. J., Norton, R. S., Nicola, N. A., & Nicholson, S. E. (2013). Structure and function of the SPRY/B30.2 domain proteins involved in innate immunity. *Protein Science: A Publication of the Protein Society*, *22*, 1–10.

Di Giallonardo, F., Schlub, T. E., Shi, M., & Holmes, E. C. (2017). Dinucleotide composition in animal RNA viruses is shaped more by virus family than by host species. *Journal of Virology*, *91*, e02381–e02416.

Diaz-San Segundo, F., Medina, G. N., Ramirez-Medina, E., Velazquez-Salinas, L., Koster, M., Grubman, M. J., & de los Santos, T. (2016). Synonymous deoptimization of foot-and-mouth disease virus causes attenuation *in vivo* while inducing a strong neutralizing antibody response. *Journal of Virology*, *90*, 1298–1310.

Digard, P., Lee, H. M., Sharp, C., Grey, F., & Gaunt, E. (2020). Intra-genome variability in the dinucleotide composition of SARS-CoV-2. *Virus Evolution*, *6*, veaa057.

Drummond, D. A., & Wilke, C. O. (2009). The evolutionary consequences of erroneous protein synthesis. *Nature Reviews. Genetics*, *10*, 715–724.

Eschke, K., Trimpert, J., Osterrieder, N., & Kunec, D. (2018). Attenuation of a very virulent Marek's disease herpesvirus (MDV) by codon pair bias deoptimization. *PLoS Pathogens*, *14*, e1006857.

Ficarelli, M., Antzin-Anduetza, I., Hugh-White, R., Firth, A. E., Sertkaya, H., Wilson, H., Neil, S. J. D., Schulz, R., & Swanson, C. M. (2020). CpG dinucleotides inhibit HIV-1 replication through zinc finger antiviral protein (ZAP)-dependent and -independent mechanisms. *Journal of Virology*, *94*, e01337–e01319.

Ficarelli, M., Wilson, H., Galão, R. P., Mazzon, M., Antzin-Anduetza, I., Marsh, M., Neil, S. J. D., & Swanson, C. M. (2019). KHNYN is essential for the zinc finger antiviral protein (ZAP) to restrict HIV-1 containing clustered CpG dinucleotides. *eLife*, *8*, e46767.

Fodor, E., Devenish, L., Engelhardt, O. G., Palese, P., Brownlee, G. G., & García-Sastre, A. (1999). Rescue of influenza A virus from recombinant DNA. *Journal of Virology*, *73*, 9679–9682.

Fros, J. J., Dietrich, I., Alshaikhahmed, K., Passchier, T. C., Evans, D. J., & Simmonds, P. (2017). CpG and UpA dinucleotides in both coding and non-coding regions of echovirus 7 inhibit replication initiation post-entry. *eLife*, *6*, e29112.

Fros, J. J., Visser, I., Tang, B., Yan, K., Nakayama, E., Visser, T. M., Koenraadt, C. J. M., van Oers, M. M., Pijlman, G. P., Suhrbier, A., & Simmonds, P. (2021). The dinucleotide composition of the Zika virus genome is shaped by conflicting evolutionary pressures in mammalian hosts and mosquito vectors. *PLoS Biology*, *19*, e3001201.

Fu, M. (2010). Codon usage bias in herpesvirus. *Archives of Virology*, *155*, 391–396.

Futcher, B., Gorbatsevych, O., Shen, S. H., Stauft, C. B., Song, Y., Wang, B., Leatherwood, J., Gardin, J., Yurovsky, A., Mueller, S., & Wimmer, E. (2015). Reply to Simmonds et al.: Codon pair and dinucleotide bias have not been functionally distinguished. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, E3635–E3636.

Gack, M. U., Shin, Y. C., Joo, C.-H., Urano, T., Liang, C., Sun, L., Takeuchi, O., Akira, S., Chen, Z., Inoue, S., & Jung, J. U. (2007). TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature*, *446*, 916–920.

Gamble, C. E., Brule, C. E., Dean, K. M., Fields, S., & Grayhack, E. J. (2016). Adjacent codons act in concert to modulate translation efficiency in yeast. *Cell*, *166*, 679–690.

Gantier, M. P., Tong, S., Behlke, M. A., Xu, D., Phipps, S., Foster, P. S., & Williams, B. R. G. (2008). TLR7 is involved in sequence-specific sensing of single-stranded RNAs in human macrophages. *The Journal of Immunology*, *180*, 2117–2124.

Gao, G., Guo, X., & Goff, S. P. (2002). Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein, *297*, 1703–1706.

Gao, L., Wang, L., Huang, C., Yang, L., Guo, X.-K., Yu, Z., Liu, Y., Yang, P., & Feng, W.-h. (2015). HP-PRRSV is attenuated by de-optimization of codon pair bias in its RNA-dependent RNA polymerase nsp9 gene. *Virology*, *485*, 135–144.

Gatfield, D., & Izaurralde, E. (2004). Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in Drosophila. *Nature*, *429*, 575–578.

Gaunt, E., Wise, H. M., Zhang, H., Lee, L. N., Atkinson, N. J., Nicol, M. Q., Highton, A. J., Klenerman, P., Beard, P. M., Dutia, B. M., Digard, P., & Simmonds, P. (2016). Elevation of CpG frequencies in influenza A genome attenuates pathogenicity but enhances host response to infection. *eLife*, *5*, e12735–e12735.

Giorgio, D., Salvatore, F. M., Torcia, M. G., Mattiuz, G., & Conticello, S. G. (2020). Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Science Advances*, *6*, eabb5813.

Goodier, J. L., Pereira, G. C., Cheung, L. E., Rose, R. J., & Kazazian, H. H., Jr. (2015). The broad-spectrum antiviral protein ZAP restricts human retrotransposition. *PLoS Genetics*, *11*, e1005252.

Grantham, R., Gautier, C., Gouy, M., Mercier, R., & Pavé, A. (1980). Codon catalog usage and the genome hypothesis. *Nucleic Acids Research*, *8*, r49–r62.

Greenbaum, B. D., Levine, A. J., Bhanot, G., & Rabadan, R. (2008). Patterns of evolution and host gene mimicry in influenza and other RNA viruses. *PLoS Pathogens*, *4*, e1000079–e1000079.

Greenbaum, B. D., Rabadan, R., & Levine, A. J. (2009). Patterns of oligonucleotide sequences in viral and host cell RNA identify mediators of the host innate immune system. *PLoS One*, *4*, e5969.

Groenke, N., Trimpert, J., Merz, S., Conradie, A. M., Wyler, E., Zhang, H., Hazapis, O.-G., Rausch, S., Landthaler, M., Osterrieder, N., & Kunec, D. (2020). Mechanism of virus attenuation by codon pair deoptimization. *Cell Reports*, *31*, 107586.

Guo, X., Carroll, J.-W. N., MacDonald, M. R., Goff, S. P., & Gao, G. (2004). The zinc finger antiviral protein directly binds to specific viral mRNAs through the CCCH zinc finger motifs. *Journal of Virology*, *78*, 12781–12787.

Guo, X., Ma, J., Sun, J., & Gao, G. (2007). The zinc-finger antiviral protein recruits the RNA processing exosome to degrade the target mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 151–156.

Haas, J., Park, E.-C., & Seed, B. (1996). Codon usage limitation in the expression of HIV-1 envelope glycoprotein. *Current Biology*, *6*, 315–324.

Hayakawa, S., Shiratori, S., Yamato, H., Kameyama, T., Kitatsuji, C., Kashigi, F., Goto, S., Kameoka, S., Fujikura, D., Yamada, T., Mizutani, T., Kazumata, M., Sato, M., Tanaka, J., Asaka, M., Ohba, Y., Miyazaki, T., Imamura, M., & Takaoka, A. (2011). ZAPS is a potent stimulator of signaling mediated by the RNA helicase RIG-I during antiviral responses. *Nature Immunology*, *12*, 37–44.

He, W., Zhang, H., Zhang, Y., Wang, R., Lu, S., Ji, Y., Liu, C., Yuan, P., & Su, S. (2017). Codon usage bias in the N gene of rabies virus. *Infection, Genetics and Evolution*, *54*, 458–465.

Ibrahim, A., Fros, J., Bertran, A., Sechan, F., Odon, V., Torrance, L., Kormelink, R., & Simmonds, P. (2019). A functional investigation of the suppression of CpG and UpA dinucleotide frequencies in plant RNA virus genomes. *Scientific Reports*, *9*, 18359.

Idoko-Akoh, A., Taylor, L., Sang, H. M., & McGrew, M. J. (2018). High fidelity CRISPR/Cas9 increases precise monoallelic and biallelic editing events in primordial germ cells. *Scientific Reports*, *8*, 15126.

Irwin, B., Heck, J. D., & Hatfield, G. W. (1995). Codon pair utilization biases influence translational elongation step times. *Journal of Biological Chemistry*, *270*, 22801–22806.

Jabbari, K., Cacciò, S., Païs de Barros, J. P., Desgrès, J., & Bernardi, G. (1997). Evolutionary changes in CpG and methylation levels in the genome of vertebrates. *Gene*, *205*, 109–118.

Janeway, C. A. (1989). Approaching the asymptote? *Evolution and Revolution in Immunology*, *54*, 1–13.

Jang, S. K., & Wimmer, E. (1990). Cap-independent translation of encephalomyocarditis virus RNA: Structural elements of the internal ribosomal entry site and involvement of a cellular 57-kD RNA-binding protein. *Genes & Development*, *4*, 1560–1572.

Jenkins, G. M., & Holmes, E. C. (2003). The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Research*, *92*, 1–7.

Jenkins, G. M., Pagel, M., Gould, E. A., Zanotto, P. M. d. A., & Holmes, E. C. (2001). Evolution of base composition and codon usage bias in the genus Flavivirus. *Journal of Molecular Evolution*, *52*, 383–390.

Josse, J., Kaiser, A. D., & Kornberg, A. (1961). Enzymatic synthesis of deoxyribonucleic acid. VIII. Frequencies of nearest neighbor base sequences in deoxyribonucleic acid. *Journal of Biological Chemistry*, *236*, 864–875.

Kapoor, A., Simmonds, P., Lipkin, W. I., Zaidi, S., & Delwart, E. (2010). Use of nucleotide composition analysis to infer hosts for three novel Picorna-like viruses. *Journal of Virology*, *84*, 10322–10328.

Karasik, A., Jones, G. D., DePass, A. V., & Guydosh, N. R. (2021). Activation of the antiviral factor RNase L triggers translation of non-coding mRNA sequences. *Nucleic Acids Research*.

Karlin, S., Doerfler, W., & Cardon, L. R. (1994). Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses? *Journal of Virology*, *68*, 2889–2897.

Kerns, J. A., Emerman, M., & Malik, H. S. (2008). Positive selection and increased antiviral activity associated with the PARP-containing isoform of human zinc-finger antiviral protein. *PLoS Genetics*, *4*, e21.

Kimchi-Sarfaty, C., Jung Mi, O., Kim, I.-W., Sauna, Z. E., Calcagno, A. M., Ambudkar, S. V., & Gottesman, M. M. (2007). A 'silent' polymorphism in MDR1 gene changes substrate specificity. *Science*, *315*, 525–528.

Krug, A., French, A. R., Barchet, W., Fischer, J. A. A., Dzionek, A., Pingel, J. T., Orihuela, M. M., Akira, S., Yokoyama, W. M., & Colonna, M. (2004). TLR9-dependent recognition of MCMV by IPC and DC generates coordinated cytokine responses that activate antiviral NK cell function. *Immunity*, *21*, 107–119.

Kudla, G., Lipinski, L., Caffin, F., Helwak, A., & Zylicz, M. (2006). High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biology*, *4*, e180.

Kudla, G., Murray, A. W., Tollervey, D., & Plotkin, J. B. (2009). Coding-sequence determinants of gene expression in Escherichia coli. *Science*, *324*, 255–258.

Kumar, H., Kawai, T., & Akira, S. (2011). Pathogen recognition by the innate immune system. *International Reviews of Immunology*, *30*, 16–34.

Kumar, N., Bera, B. C., Greenbaum, B. D., Bhatia, S., Sood, R., Selvaraj, P., Anand, T., Tripathi, B. N., & Virmani, N. (2016). Revelation of influencing factors in overall codon usage bias of equine influenza viruses. *PLoS One*, *11*, e0154376.

Kunec, D., & Osterrieder, N. (2016). Codon pair bias is a direct consequence of dinucleotide bias. *Cell Reports*, *14*, 55–67.

Kypr, J., & Mrázek, J. A. N. (1987). Unusual codon usage of HIV. *Nature*, *327*, 20–20.

Lauring, A. S., Acevedo, A., Cooper, S. B., & Andino, R. (2012). Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. *Cell Host & Microbe*, *12*, 623–632.

Law, L. M. J., Razooky, B. S., Li, M. M. H., You, S., Jurado, A., Rice, C. M., & MacDonald, M. R. (2019). ZAP's stress granule localization is correlated with its antiviral activity and induced by virus replication. *PLoS Pathogens*, *15*, e1007798.

Le Nouën, C., Brock, L. G., Luongo, C., McCarty, T., Yang, L., Mehedi, M., Wimmer, E., Mueller, S., Collins, P. L., Buchholz, U. J., & DiNapoli, J. M. (2014). Attenuation of human respiratory syncytial virus by genome-scale codon-pair deoptimization. *Proceedings of the National Academy of Sciences of the United States of America*, *111*, 13169–13174.

Leifer, I., Hoeper, D., Blome, S., Beer, M., & Ruggli, N. (2011). Clustering of classical swine fever virus isolates by codon pair bias. *BMC Research Notes*, *4*, 521.

Li, M., Kao, E., Gao, X., Sandig, H., Limmer, K., Pavon-Eternod, M., Jones, T. E., Landry, S., Pan, T., Weitzman, M. D., & David, M. (2012). Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. *Nature*, *491*, 125–128.

Li, M. M. H., Aguilar, E. G., Michailidis, E., Pabon, J., Park, P., Wu, X., de Jong, Y. P., Schneider, W. M., Molina, H., Rice, C. M., & MacDonald, M. R. (2019). Characterization of novel splice variants of zinc finger antiviral protein (ZAP). *Journal of Virology*, *93*, e00715–e00719.

Li, M. M. H., Lau, Z., Cheung, P., Aguilar, E. G., Schneider, W. M., Bozzacco, L., Molina, H., Buehler, E., Takaoka, A., Rice, C. M., Felsenfeld, D. P., & MacDonald, M. R. (2017). TRIM25 enhances the antiviral action of zinc-finger antiviral protein (ZAP). *PLoS Pathogens*, *13*, e1006145.

Li, P., Ke, X., Wang, T., Tan, Z., Luo, D., Miao, Y., Sun, J., Zhang, Y., Liu, Y., Hu, Q., Xu, F., Wang, H., & Zheng, Z. (2018). Zika virus attenuation by codon pair deoptimization induces sterilizing immunity in mouse models. *Journal of Virology*, *92*, e00701–e00718.

Li, X., Min, G., Zheng, Q., Gao, R., & Liu, X. (2021). Packaging signal of influenza A virus. *Virology Journal*, *18*, 36.

Lin, Y.-T., Chiweshe, S., McCormick, D., Raper, A., Wickenhagen, A., DeFillipis, V., Gaunt, E., Simmonds, P., Wilson, S. J., & Grey, F. (2020). Human cytomegalovirus evades ZAP detection by suppressing CpG dinucleotides in the major immediate early genes. *PLoS Pathogens*, *2020*(01), 07.897132.

Liu, X., Wu, C., & Chen, A. Y. (2010). Codon usage bias and recombination events for neuraminidase and hemagglutinin genes in Chinese isolates of influenza A virus subtype H9N2. *Archives of Virology*, *155*, 685–693.

Lobo, F. P., Mota, B. E. F., Pena, S. D. J., Azevedo, V., Macedo, A. M., Tauch, A., Machado, C. R., & Franco, G. R. (2009). Virus-host coevolution: Common patterns of nucleotide motif usage in Flaviviridae and their hosts. *PLoS One*, *4*, e6282.

Long, J. S., Idoko-Akoh, A., Mistry, B., Goldhill, D., Staller, E., Schreyer, J., Ross, C., Goodbourn, S., Shelton, H., Skinner, M. A., Sang, H., McGrew, M. J., & Barclay, W. (2019). Species specific differences in use of ANP32 proteins by influenza A virus. *eLife*, *8*, e45066.

Luo, X., Wang, X., Gao, Y., Zhu, J., Liu, S., Gao, G., & Gao, P. (2020). Molecular mechanism of RNA recognition by zinc-finger antiviral protein. *Cell Reports*, *30*, 46–52.e4.

Marcello, T., Grakoui, A., Barba-Spaeth, G., Machlin, E. S., Kotenko, S. V., Macdonald, M. R., & Rice, C. M. (2006). Interferons α and λ inhibit hepatitis C virus replication with distinct signal transduction and gene regulation kinetics. *Gastroenterology*, *131*, 1887–1898.

Martrus, G., Nevot, M., Andres, C., Clotet, B., & Martinez, M. A. (2013). Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture. *Retrovirology*, *10*, 78.

Meagher, J. L., Takata, M., Gonçalves-Carneiro, D., Keane, S. C., Rebendenne, A., Ong, H., Orr, V. K., MacDonald, M. R., Stuckey, J. A., Bieniasz, P. D., & Smith, J. L. (2019). Structure of the zinc-finger antiviral protein in complex with RNA reveals a mechanism for selective targeting of CG-rich viral sequences. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 24303–24309.

Meng, J., Lee, S., Hotard, A. L., & Moore, M. L. (2014). Refining the balance of attenuation and immunogenicity of respiratory syncytial virus by targeted codon deoptimization of virulence genes. *mBio*, *5*, e01704–e01714.

Mohler, K., & Ibba, M. (2017). Translational fidelity and mistranslation in the cellular response to stress. *Nature Microbiology*, *2*, 17117–17117.

Mordstein, C., Savisaar, R., Young, R. S., Bazile, J., Talmane, L., Luft, J., Liss, M., Taylor, M. S., Hurst, L. D., & Kudla, G. (2020). Codon usage and splicing jointly influence mRNA localization. *Cell Systems*, *10*, 351–62.e8.

Mueller, S., Coleman, J. R., Papamichail, D., Ward, C. B., Nimnual, A., Futcher, B., Skiena, S., & Wimmer, E. (2010). Live attenuated influenza virus vaccines by computer-aided rational design. *Nature Biotechnology*, *28*, 723–726.

Müller, V., & Bonhoeffer, S. (2005). Guanine-adenine bias: A general property of retroid viruses that is unrelated to host-induced hypermutation. *Trends in Genetics*, *21*, 264–268.

Neumann, G., Watanabe, T., Ito, H., Watanabe, S., Goto, H., Gao, P., Hughes, M., Perez, D. R., Donis, R., Hoffmann, E., Hobom, G., & Kawaoka, Y. (1999). Generation of influenza A viruses entirely from cloned cDNAs. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 9345–9350.

Ni, Y. Y., Zhao, Z., Opriessnig, T., Subramaniam, S., Zhou, L., Cao, D., Cao, Q., Yang, H., & Meng, X. J. (2014). Computer-aided codon-pairs deoptimization of the major envelope GP5 gene attenuates porcine reproductive and respiratory syndrome virus. *Virology*, *450–451*, 132–139.

Nibert, M. L. (2017). Mitovirus UGA(Trp) codon usage parallels that of host mitochondria. *Virology*, *507*, 96–100.

Nougairede, A., De Fabritus, L., Aubry, F., Gould, E. A., Holmes, E. C., & de Lamballerie, X. (2013). Random codon re-encoding induces stable reduction of replicative fitness of Chikungunya virus in primate and mosquito cells. *PLoS Pathogens*, *9*, e1003172.

Novoa, E. M., & Ribas de Pouplana, L. (2012). Speeding with control: Codon usage, tRNAs, and ribosomes. *Trends in Genetics*, *28*, 574–581.

Nussinov, R. (1981). Eukaryotic dinucleotide preference rules and their implications for degenerate codon usage. *Journal of Molecular Biology*, *149*, 125–131.

Odon, V., Fros, J. J., Goonawardane, N., Dietrich, I., Ibrahim, A., Alshaikhahmed, K., Nguyen, D., & Simmonds, P. (2019). 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 Research*, 47, 8061–8083.

O'Hara, P. J., Nichol, S. T., Horodyski, F. M., & Holland, J. J. (1984). Vesicular stomatitis virus defective interfering particles can contain extensive genomic sequence rearrangements and base substitutions. *Cell*, 36, 915–924.

Olson, K. E., Higgs, S., Gaines, P. J., Powers, A. M., Davis, B. S., Kamrud, K. I., Carlson, J. O., Blair, C. D., & Beaty, B. J. (1996). Genetically engineered resistance to dengue-2 virus transmission in mosquitoes. *Science*, 272, 884–886.

Oshiumi, H., Matsuo, A., Matsumoto, M., & Seya, T. (2008). Pan-vertebrate toll-like receptors during evolution. *Current Genomics*, 9, 488–493.

Ou, X., Cao, J., Cheng, A., Peppelenbosch, M. P., & Pan, Q. (2019). Errors in translational decoding: tRNA wobbling or misincorporation? *PLoS Genetics*, 15, e1008017.

Patil, A., Chan, C. T. Y., Dyavaiah, M., Rooney, J. P., Dedon, P. C., & Begley, T. J. (2012). Translational infidelity-induced protein stress results from a deficiency in Trm9-catalyzed tRNA modifications. *RNA Biology*, 9, 990–1001.

Pechmann, S., & Frydman, J. (2013). Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nature Structural & Molecular Biology*, 20, 237–243.

Perdue, M. L., Arnold, F., Li, S., Donabedian, A., Cioce, V., Warf, T., & Huebner, R. (2011). The future of cell culture-based influenza vaccine production. *Expert Review of Vaccines*, 10, 1183–1194.

Pezda, A. C., Penn, A., Barton, G. M., & Coscoy, L. (2011). Suppression of TLR9 immunostimulatory motifs in the genome of a Gammaherpesvirus. *The Journal of Immunology*, 187, 887–896.

Plotkin, J. B., & Dushoff, J. (2003). Codon bias and frequency-dependent selection on the hemagglutinin epitopes of influenza A virus. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 7152–7157.

Presnyak, V., Alhusaini, N., Chen, Y. H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K. E., Graveley, B. R., & Coller, J. (2015). Codon optimality is a major determinant of mRNA stability. *Cell*, 160, 1111–1124.

Rice, A. M., Morales, A. C., Ho, A. T., Mordstein, C., Mühlhausen, S., Watson, S., Cano, L., Young, B., Kudla, G., & Hurst, L. D. (2020). Evidence for strong mutation bias towards, and selection against, U content in SARS-CoV-2: Implications for vaccine design. *Molecular Biology and Evolution*, 38, 67–83.

Rima, B. K., & McFerran, N. V. (1997). Dinucleotide and stop codon frequencies in single-stranded RNA viruses. *Journal of General Virology*, 78, 2859–2870.

Rothberg, P. G., & Wimmer, E. (1981). Mononucleotide and dinucleotide frequencies, and codon usage in poliovirion RNA. *Nucleic Acids Research*, 9, 6221–6229.

Roundtree, I. A., Evans, M. E., Pan, T., & He, C. (2017). Dynamic RNA modifications in gene expression regulation. *Cell*, 169, 1187–1200.

Ryman, K. D., Meier, K. C., Nangle, E. M., Ragsdale, S. L., Korneeva, N. L., Rhoads, R. E., MacDonald, M. R., & Klimstra, W. B. (2005). Sindbis virus translation is inhibited by a PKR/RNase L-independent effector induced by alpha/beta interferon priming of dendritic cells. *Journal of Virology*, 79, 1487–1499.

Samuel, C. E. (2012). ADARs: Viruses and innate immunity. *Current Topics in Microbiology and Immunology*, 353, 163–195.

Savisaar, R., & Hurst, L. D. (2018). Exonic splice regulation imposes strong selection at synonymous sites. *Genome Research*, 28, 1442–1454.

Shabalina, S. A., Ogurtsov, A. Y., & Spiridonov, N. A. (2006). A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Research*, 34, 2428–2437.

Shackelton, L. A., Parrish, C. R., & Holmes, E. C. (2006). Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *Journal of Molecular Evolution*, 62, 551–563.

Sharma, S., Patnaik, S. K., Taggart, R. T., & Baysal, B. E. (2016). The double-domain cytidine deaminase APOBEC3G is a cellular site-specific RNA editing enzyme. *Scientific Reports*, 6, 39100.

Sharp, P. M., & Li, W.-H. (1987). The Codon Adaptation Index—A measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*, 15, 1281–1295.

Sheehy, A. M., Gaddis, N. C., Choi, J. D., & Malim, M. H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature*, 418, 646–650.

Shen, S. H., Stauft, C. B., Gorbatsevych, O., Song, Y., Ward, C. B., Yurovsky, A., Mueller, S., Futcher, B., & Wimmer, E. (2015). Large-scale recoding of an arbovirus genome to rebalance its insect versus mammalian preference. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 4749–4754.

Shin, Y. C., Bischof, G. F., Lauer, W. A., & Desrosiers, R. C. (2015). Importance of codon usage for the temporal regulation of viral gene expression. *Proceedings of the National Academy of Sciencesof the United States of America*, 112, 14030–14035.

Simmonds, P. (2020). 'Rampant C→U hypermutation in the Ggenomes of SARS-CoV-2 and other coronaviruses: Causes and consequences for their short- and long-term evolutionary trajectories. *mSphere*, 5, e00408–e00420.

Simmonds, P. (2012). SSE: A nucleotide and amino acid sequence analysis platform. *BMC Research Notes*, 5, 50.

Simmonds, P., Tulloch, F., Evans, D. J., & Ryan, M. D. (2015). Attenuation of dengue (and other RNA viruses) with codon pair recoding can be explained by increased CpG/UpA dinucleotide frequencies. *Proceedings of the National Academy of Sciences of the United States of America*, 112, E3633–E3634.

Simmonds, P., Wenjun, X., Kenneth Baillie, J., & McKinnon, K. (2013). Modelling mutational and selection pressures on dinucleotides in eukaryotic phyla –selection against CpG and UpA in cytoplasmically expressed RNA and in RNA viruses. *BMC Genomics*, 14, 610.

Sitaraman, V., Hearing, P., Ward, C. B., Gnatenko, D. V., Wimmer, E., Mueller, S., Skiena, S., & Bahou, W. F. (2011). Computationally designed adeno-associated virus (AAV) Rep 78 is efficiently maintained within an adenovirus vector. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 14294–14299.

Song, W.-Y., Wang, G.-L., Chen, L.-L., Kim, H.-S., Pi, L.-Y., Tom, H., Gardner, J., Wang, B., Zhai, W.-X., Zhu, L.-H., Fauquet, C., & Ronald, P. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science*, *270*, 1804–1806.

Swartz, M. N., Trautner, T. A., & Kornberg, A. (1962). Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids. *Journal of Biological Chemistry*, *237*, 1961–1967.

Tabeta, K., Georgel, P., Janssen, E., Xin, D., Hoebe, K., Crozat, K., Mudd, S., Shamel, L., Sovath, S., Goode, J., Alexopoulou, L., Flavell, R. A., & Beutler, B. (2004). Toll-like receptors 9 and 3 as essential components of innate immune defense against mouse cytomegalovirus infection, *101*, 3516–3521.

Takata, M. A., Gonçalves-Carneiro, D., Zang, T. M., Soll, S. J., York, A., Blanco-Melo, D., & Bieniasz, P. D. (2017). CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature*, *550*, 124–127.

Tao, P., Dai, L., Luo, M., Tang, F., Tien, P., & Pan, Z. (2009). Analysis of synonymous codon usage in classical swine fever virus. *Virus Genes*, *38*, 104–112.

Tats, A., Tenson, T., & Remm, M. (2008). Preferred and avoided codon pairs in three domains of life. *BMC Genomics*, *9*, 463.

Todorova, T., Bock, F. J., & Chang, P. (2014). PARP13 regulates cellular mRNA post-transcriptionally and functions as a pro-apoptotic factor by destabilizing TRAILR4 transcript. *Nature Communications*, *5*, 5362.

Torrent, M., Chalancon, G., de Groot, N. S., Wuster, A., & Babu, M. M. (2018). Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions. *Science Signaling*, *11*, eaat6409.

Trono, D., Pelletier, J., Sonenberg, N., & Baltimore, D. (1988). Translation in mammalian cells of a gene linked to the poliovirus 5′ noncoding region. *Science*, *241*, 445–448.

Tulloch, F., Atkinson, N. J., Evans, D. J., Ryan, M. D., & Simmonds, P. (2014). RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. *eLife*, *3*, e04531.

Upadhyay, M., Sharma, N., & Vivekanandan, P. (2014). Systematic CpT (ApG) depletion and CpG excess are unique genomic signatures of large DNA viruses infecting invertebrates. *PLoS One*, *9*, e111793.

Urrutia, A. O., & Hurst, L. D. (2003). The signature of selection mediated by expression on human genes. *Genome Research*, *13*, 2260–2264.

Vabret, N., Bailly-Bechet, M., Lepelley, A., Najburg, V., Schwartz, O., Verrier, B., & Tangy, F. (2014). Large-scale nucleotide optimization of simian immunodeficiency virus reduces its capacity to stimulate type I interferon in vitro. *Journal of Virology*, *88*, 4161–4172.

van der Kuyl, A. C., & Berkhout, B. (2012). The biased nucleotide composition of the HIV genome: A constant factor in a highly variable virus. *Retrovirology*, *9*, 92.

van Hemert, F., & Berkhout, B. (2016). Nucleotide composition of the Zika virus RNA genome and its codon usage. *Virology*, *13*, 95–95.

van Hemert, F. J., Berkhout, B., & Lukashov, V. V. (2007). Host-related nucleotide composition and codon usage as driving forces in the recent evolution of the Astroviridae. *Virology*, *361*, 447–454.

Wang, B., Yang, C., Tekes, G., Mueller, S., Paul, A., Whelan, S. P. J., & Wimmer, E. (2015). Recoding of the vesicular stomatitis virus L gene by computer-aided design provides a live, attenuated vaccine candidate. *mBio*, *6*, e00237–e00215.

Warnecke, T., Weber, C. C., & Hurst, L. D. (2009). Why there is more to protein evolution than protein function: Splicing, nucleosomes and dual-coding sequence. *Biochemical Society Transactions*, *37*, 756–761.

Washenberger, C. L., Han, J.-Q., Kechris, K. J., Jha, B. K., Silverman, R. H., & Barton, D. J. (2007). Hepatitis C virus RNA: Dinucleotide frequencies and cleavage by RNase L. *Virus Research*, *130*, 85–95.

Witteveldt, J., Martin-Gans, M., & Simmonds, P. (2016). Enhancement of the replication of hepatitis C virus replicons of genotypes 1 to 4 by manipulation of CpG and UpA dinucleotide frequencies and use of cell lines expressing SECL14L2 for antiviral resistance testing. *Antimicrobial Agents and Chemotherapy*, *60*, 2981–2992.

Wong, E. H. M., Smith, D. K., Rabadan, R., Peiris, M., & Poon, L. L. M. (2010). Codon usage bias and the evolution of influenza A viruses. Codon usage biases of influenza virus. *BMC Evolutionary Biology*, *10*, 253.

Wreschner, D. H., McCauley, J. W., Skehel, J. J., & Kerr, I. M. (1981). Interferon action—Sequence specificity of the ppp(A2′p)nA-dependent ribonuclease. *Nature*, *289*, 414–417.

Yang, C., Skiena, S., Futcher, B., Mueller, S., & Wimmer, E. (2013). Deliberate reduction of hemagglutinin and neuraminidase expression of influenza virus leads to an ultraprotective live vaccine in mice. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 9481–9486.

Yarus, M., & Folley, L. S. (1985). Sense codons are found in specific contexts. *Journal of Molecular Biology*, *182*, 529–540.

Zhang, Z., Ohto, U., Shibata, T., Krayukhina, E., Taoka, M., Yamauchi, Y., Tanji, H., Isobe, T., Uchiyama, S., Miyake, K., & Shimizu, T. (2016). Structural analysis reveals that toll-like receptor 7 is a dual receptor for Guanosine and single-stranded RNA. *Immunity*, *45*, 737–748.

Zhao, K.-N., Liu, W. J., & Frazer, I. H. (2003). Codon usage bias and A+T content variation in human papillomavirus genomes. *Virus Research*, *98*, 95–104.

Zheng, X., Wang, X., Fan, T., Wang, Q., Fan, Z., & Gao, G. (2017). TRIM25 is required for the antiviral activity of zinc finger antiviral protein. *Journal of Virology*, *91*, e00088–e00017.

Zhong, J., Li, Y., Zhao, S., Liu, S., & Zhang, Z. (2007). Mutation pressure shapes codon usage in the GC-rich genome of foot-and-mouth disease virus. *Virus Genes*, *35*, 767–776.

Zhou, J., Liu, W. J., Peng, S. W., Sun, X. Y., & Frazer, I. (1999). Papillomavirus capsid protein expression level depends on the match between codon usage and tRNA availability. *Journal of Virology*, *73*, 4972–4982.

Zhou, Y., Chen, X., Ushijima, H., & Frey, T. K. (2012). Analysis of base and codon usage by rubella virus. *Archives of Virology*, *157*, 889–899.

Zhu, Y., Chen, G., Lv, F., Wang, X., Ji, X., Xu, Y., Sun, J., Wu, L., Zheng, Y. T., & Gao, G. (2011). Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 15834–15839.

Zhu, Y., Wang, X., Goff, S. P., & Gao, G. (2012). Translational repression precedes and is required for ZAP-mediated mRNA decay. *The EMBO Journal*, *31*, 4236–4246.

Zou, W., & Zhang, D.-E. (2006). The interferon-inducible ubiquitin-protein isopeptide ligase (E3) EFP also functions as an ISG15 E3 ligase. *Journal of Biological Chemistry*, *281*, 3989–3994.