Evolutionary History of Cotranscriptional Editing in the Paramyxoviral Phosphoprotein Gene

Jordan Douglas\textsuperscript{1,2,\ast}, Alexei J. Drummond\textsuperscript{1,2,3}, and Richard L. Kingston\textsuperscript{3}

\textsuperscript{1} Centre for Computational Evolution, University of Auckland, Auckland 1010, New Zealand \textsuperscript{2} School of Computer Science, University of Auckland, Auckland 1010, New Zealand \textsuperscript{3} School of Biological Sciences, University of Auckland, Auckland 1010, New Zealand

\ast Correspondence: jordan.douglas@auckland.ac.nz

Abstract

The phosphoprotein gene of the paramyxoviruses encodes multiple protein products. The P, V, and W proteins are generated by transcriptional slippage. This process results in the insertion of non-templated guanosine nucleosides into the mRNA at a conserved edit site. The P protein is an essential component of the viral RNA polymerase, and is encoded by a faithful copy of the gene in the majority of paramyxoviruses. However, in some cases the non-essential V protein is encoded by default and guanosines must be inserted into the mRNA in order to encode P. The number of guanosines inserted into the P gene can be described by a probability distribution which varies between viruses. In this article we review the nature of these distributions, which can be inferred from mRNA sequencing data, and reconstruct the evolutionary history of cotranscriptional editing in the paramyxovirus family. Our model suggests that, throughout known history of the family, the system has switched from a P default to a V default mode four times; complete loss of the editing system has occurred twice, the canonical zinc finger domain of the V protein has been deleted or heavily mutated a further two times, and the W protein has independently evolved a novel function three times. Finally, we review the physical mechanisms of cotranscriptional editing via slippage of the viral RNA polymerase.

© The Author(s) 2021. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
1 Introduction

Most viruses possess genes that encode for more than one protein. When these proteins arise from translation of a common nucleotide sequence in differing reading frames, the phenomenon has been termed gene overlap (Barrell et al., 1976) or overprinting (Keese and Gibbs, 1992). In viruses, overprinting has frequently been linked to the strong size constraints that exist on viral genomes (Belshaw et al., 2007), however it has also been considered to confer certain evolutionary advantages (Brandes and Linial, 2016; Sabath et al., 2012). Overprinting by viruses is ubiquitous (Chirico et al., 2010), and it can arise from events occurring during both gene transcription (Brennicke et al., 1999) and the translation of messenger RNA (mRNA; Kozak (2002)).

At the transcriptional level, viruses may employ cotranscriptional RNA editing (Catta neo, 1991), in which nucleotides that are not directly specified by the template are inserted into the viral mRNA during transcription (i.e. the mRNA is no longer a faithful copy of the gene). Viral families that perform this kind of RNA editing include the Paramyxoviridae (Vidal et al., 1990b; Hausmann et al., 1999a), the Filoviridae (Sanchez et al., 1996; Shabman et al., 2014), and the Poxyviridae (Olsper et al., 2015; Rodamians et al., 2015). Cotranscriptional RNA editing also occurs in a variety of prokaryotes (Larsen et al., 2000; Penno et al., 2015). The primary mechanism underpinning cotranscriptional RNA editing is thought to be transcriptional slippage, which allows a nucleic acid polymerase to reiteratively copy a single base (Streisinger et al., 1966; Garcia-Diaz and Kunkel, 2006).

At the translational level, non-canonical initiation, elongation, and termination events are also used as overprinting mechanisms by numerous viral families (Firth and Brierley, 2012; Maia et al., 1996), including the Paramyxoviridae (Giorgi et al., 1983; Curran and Kolakofsky, 1988; Latorre et al., 1998), the Coronaviridae, and the Retroviridae (Brierley and Dos Ramos, 2006). These events include leaky scanning, non-AUG initiation, ribosomal shunting, and ribosomal frameshifting.

In this article we review cotranscriptional RNA editing in the Paramyxoviridae; a family of nonsegmented, negative-sense, single-stranded RNA viruses, within the order Mononegavirales (Amarasinghe et al., 2019; Pringle, 1991; Rima et al., 2018). Cotranscriptional editing of the paramyxoviral phosphoprotein gene (P gene) governs production of up to three proteins: P, V, and W. The editing process involves insertion of one or more non-templated guanosine nucleosides into the mRNA at a conserved edit site (Vidal et al., 1990b; Hausmann et al., 1999a), which stochastically shifts the reading frame. As a result, the P, V, and W proteins share a common N-terminal region (encoded by the gene sequence upstream of the edit site), but possess distinct C-terminal regions (encoded by the gene sequence downstream of the edit site), which allows for differing function.

The P protein (phosphoprotein) is an essential subunit of the viral RNA-dependent RNA-polymerase (RdRp). In contrast, the V and W proteins are non-essential, but may serve as virulence factors. This is quite typical for viral proteins that have arisen by gene overprinting (Rancurel et al., 2009). While most paramyxoviral genomes directly encode the P protein, a minority directly encode the V protein, with the virus consequently becoming completely dependent on P gene editing for viability.

Our review begins with a discussion of virally-directed RNA synthesis in the paramyxoviruses, the overprinting of the P gene, and the organization and function of the P, V, and W proteins. We collate experimental information on the nature of the genome (which of P or V is directly encoded?) as well as the distribution describing the number of guanosine nucleotides inserted into the P gene, and hence the relative abundance of mRNA encoding P, V, and W. To explain this data, we propose a maximum parsimony model for the evolution of the editing system. While the P protein is always produced, due to its highly conserved and critical role in viral replication, V and W are “luxury” proteins whose functional status varies between paramyxoviruses, and which are occasionally lost altogether through retirement of the editing system. Novel functionality is materialising relatively rapidly in this region of the
genome, emphasising the ongoing nature of the evolutionary process. We conclude by reviewing what is known about transcriptional slippage, which provides the mechanism for P gene editing, and its connection with the genomic sequence at the edit site.

2 Paramyxoviral RNA synthesis and the rule of six

The *Paramyxoviridae* appear to infect most vertebrate species (Table 1) and are responsible for a number of serious diseases in both animals and humans. Type species include measles virus (MeV; genus: *Morbillivirus*), mumps virus (MuV; genus: *Orthorubulavirus*), Sendai virus (SeV; genus: *Respirovirus*), and Hendra virus (HeV; genus: *Henipavirus*).

In paramyxoviruses, as for the entire order *Mononegavirales*, gene transcription and genome replication are distinct processes, and both are carried out by the viral RdRP. The catalytic subunit of the RdRP – the viral Large protein (L protein) – performs the basic operation of RNA synthesis and is also responsible for mRNA capping and polyadenylation (Fears and Plemper, 2017). Although the viral and host mRNA are indistinguishable, the strategies used by virus and host to cap and polyadenylate mRNA are quite divergent. Polyadenylation by the paramyxoviral RdRP results from a transcriptional slippage mechanism, resembling that used for P gene editing – the focus of this review. Therefore it has been hypothesised that these two non-templated nucleotide insertion systems share common ancestry, with development of a slippage prone polymerase subsequently enabling overprinting of the P gene (Hausmann et al., 1999a).

The viral single-stranded RNA genome is bound to the nucleocapsid protein, forming a helical protein-nucleic acid complex which encapsulates and protects the genome (Whelan et al., 2004; Fears and Plemper, 2017; Guseva et al., 2019). The nucleocapsid acts as a template for all virally-directed RNA-synthesis. Transcription precedes genome replication, with switching between the two processes believed to be driven by the accumulation of the nucleocapsid protein (Curran and Kolakofsky, 2008; Plumet et al., 2005). When operating as a transcriptase, the RdRP sequentially transcribes the viral genes, releasing capped and polyadenylated mono-cistronic mRNA. When operating as a replicase, the conserved regulatory sequences between genes are ignored, and the RdRP produces a full length copy of the viral genome or antigenome, simultaneously encapsidating it with the nucleocapsid protein (Noton and Fears, 2015).

Each nucleocapsid protein binds six nucleotides of RNA (Gutsche et al., 2015; Alayyoubi et al., 2015; Jamin and Yabukarski, 2017; Webby et al., 2019), and paramyxoviral genomes always conform to the "rule of six" whereby genome length is some multiple of six (Kolakofsky et al., 1998; Calain and Roux, 1993; Kolakofsky et al., 2005). This is hypothesised to result from the requirement to position the promoter sequences required for initiation of RNA synthesis in the correct register, or phase, with respect to the nucleocapsid protein (Le Mercier and Kolakofsky, 2019).

3 Overprinting of the P gene

3.1 Cotranscriptional editing of the P gene

Cotranscriptional editing of the P gene occurs through the insertion (or in certain mutants the deletion; Jacques et al. (1994)) of $m$ guanosines, $G_m$, into the mRNA at a conserved edit site. A $G_{3k+1}$ insertion ($m = 1, 4, 7, \ldots$) shifts the reading frame downstream of the edit site by -1 (or alternatively +2). A $G_{3k+2}$ nucleotide insertion ($m = 2, 5, 8, \ldots$) shifts the reading frame by
-2 (or alternatively +1). A $G_{3k}$ insertion ($m = 0, 3, 6, \ldots$) leaves the reading frame unaltered.

This editing system operates in two different modes (Figure 1). In the P-mode, P is encoded by the unedited gene. V can be derived from a single guanosine insertion $G_1$ and W can be derived from a double insertion $G_2$. This is the situation in MeV (Cattaneo et al., 1989) and SeV (Vidal et al., 1990a). Whereas in the V-mode, V is encoded by the unedited gene, while W can be derived from a single guanosine insertion $G_1$ and P from a double insertion $G_2$. This is the situation in MuV (Paterson and Lamb, 1990). A third edit mode (the W-mode) is conceptually possible, but so far has not been observed.

It is generally assumed that the properties of P/V/W are defined by the reading frame downstream of the edit site, and the actual number of guanosines inserted is immaterial to function (i.e. there is no effective difference between a V protein resulting from a $G_1$ insertion and a V protein resulting from a $G_4$ insertion). This is because the mRNA flanking the edit site encodes an intrinsically disordered region of P/V/W (Habchi and Longhi, 2012; Longhi et al., 2017; Guseva et al., 2019). Any extended sequence of G nucleotides is translated into polyglycine, and while the conformational preferences of polyglycine are still not entirely established (Tran et al., 2008; Ohnishi et al., 2006), the homo-polymeric sequence will be disordered. Therefore small variations in the length of this sequence are likely to be functionally neutral in this context.

3.2 Genome replication and the switching of P gene edit modes

Any switch between edit modes requires a frameshift mutation in the genome, i.e. during genome replication. This mutation must occur at a position upstream of the edit site, but not so far upstream that it disrupts some other function of the encoded P protein. Due to the rule of six, any insertion or deletion (indel) must be rapidly compensated such that the genome length remains divisible by six. Otherwise the replication efficiency of the virus would be severely impacted (Calain and Roux, 1993; Kolakofsky et al., 2005; Skiadopoulos et al., 2003; Sauder et al., 2016). For example, a single nucleotide insertion upstream and proximal to the edit site, accompanied by a single nucleotide deletion elsewhere in the genome, would be sufficient to transit the system from the P-mode to the V-mode. It has recently been noted that using P gene editing as a taxonomic criterion leads to inconsistencies in virus classification (Rima et al., 2018). The necessarily abrupt switching between edit modes suggests one of the reasons why – there are viruses with very closely related genome sequences that have adopted different edit modes (Section 5).

A question that naturally follows is how RNA editing within the P gene is effectively suppressed during genome replication. Based on nucleotide sequencing, many early studies showed that paramyxoviral genomes were homogenous in the region surrounding the P gene edit site (Horikami and Moyer, 1991; Thomas et al., 1988; Cattaneo et al., 1989; Vidal et al., 1990a; Ohgimoto et al., 1990; Southern et al., 1990; Takeuchi et al., 1990; Paterson and Lamb, 1990). This homogeneity could result from the near complete suppression of editing during viral genome replication. Alternatively, it could also arise from extremely inefficient copying of edited anti-genomes of non-hexamer length (Hausmann et al., 1996). In the Ebolaviruses (family: Filoviridae) the viral glycoprotein (GP) gene is edited in a fashion analogous to the paramyxoviral P gene. However, in this case there are no strict constraints on genome length (Weik et al., 2005), and RNA editing at the Ebolavirus GP editing site is observed to occur at appreciable frequency during both transcription and genome replication (Mehedi et al., 2011; Shabman et al., 2014; Volchkova et al., 2011).

Overall the frequency with which paramyxoviral P gene editing occurs during genome replication remains unclear. If its occurrence is non-trivial, then this could be plausibly linked to the transition between edit modes which has occurred multiple times in the evolutionary history of the family (Section 5.4).

3.3 Translational overprinting of the P gene
Remarkably, the P gene can be the locus for further overprinting events. Operations at the translational level, including leaky scanning (Giorgi et al., 1983; Shaffer et al., 2003), non AUG initiation (Boeck et al., 1992; Curran and Kolakofsky, 1988), and ribosomal shunting (Latorre et al., 1998), facilitate production of yet more proteins from the P gene in some paramyxoviruses. While it is not known why the P gene has become the sole locus for both transcriptional and translational overprinting events in the paramyxoviruses, this probably reflects the presence of long intrinsically disordered tracts in the P/V/W proteins (Longhi et al., 2017; Guseva et al., 2019), placing relatively weak constraints on nucleotide sequence evolution in this part of the genome (Kovacs et al., 2010; Rancurel et al., 2009; Jordan et al., 2000).

4 Organisation and function of the proteins resulting from gene editing

4.1 P protein

The phosphoprotein is the largest of the three proteins resulting from P gene editing, and has a range of functions. In complex with the viral L protein, it forms an integral part of RdRP and enables both translocation of the RdRP along its template, (Kingston et al., 2004; Bruhn et al., 2019; Milles et al., 2018; Sourimant et al., 2020) as well as packaging of the nascent RNA genome by the nucleocapsid protein during replication. The phosphoprotein is therefore essential (Curran et al., 1991) and is encoded by all paramyxoviruses.

The N-terminal region (NT) of P is shared with V and W. It is intrinsically disordered but can undergo coupled binding and folding to enable function. One such event involves the highly conserved soyz1 and soyz2 motifs (Karlín and Belshaw, 2012). These two modules, together with internally located sequences, are involved in chaperoning viral nucleocapsid protein monomers during replication by binding to the nucleocapsid protein and blocking the non-specific packaging of cellular RNA (Yabukarski et al., 2014; Milles et al., 2018; Alayyoubi et al., 2015; Guryanov et al., 2016). The NT is also a locus for the recruitment of several host proteins, most prominently STAT1 (signal transducer and activator of transcription 1) in the morbilliviruses and henipaviruses (Ramachandran and Horvath, 2009; Harrison and Moseley, 2020), through which P/V/W can act to inhibit STAT signalling. The functions of the N-terminal region are likely regulated by phosphorylation (Saikia et al., 2008; Sun et al., 2009; Sugai et al., 2012; Young et al., 2019; Pickar et al., 2014; Qiu et al., 2016b). The N-terminal region ranges in size from 109 aa (in APMV-3) to 570 aa (in GH-M74a) in length.

The unique C-terminal region of the phosphoprotein (PCT) is encoded by the sequence following the edit site. It contains an oligomerisation domain (a coiled coil; Burmeister et al. (2000); Tarbouriech et al. (2000); Communie et al. (2013a); Cox et al. (2013); Bruhn et al. (2014); Jensen et al. (2020)) and a nucleocapsid / L protein binding domain (the foot domain, or X domain; Johansson et al. (2003); Kingston et al. (2008); Yegambaran et al. (2013); Blanchard et al. (2004)) which are connected by a highly flexible linker (Longhi et al., 2017; Herr et al., 2019). The C-terminal region of the phosphoprotein binds to both the large protein (Bruhn et al., 2019; Abdella et al., 2020) and the nucleocapsid (Kingston et al., 2004; Habchi et al., 2011; Communie et al., 2013b; Bloyet et al., 2016; Du Pont et al., 2019), and mediates their engagement. The C-terminal regions range in size from 229 aa (in PIV-5) to 386 aa (in CPIV-3).

4.2 V protein

The paramyxoviral V protein is involved in evasion of the innate immune response, and is a major determinant of viral pathogenicity (Patterson et al., 2000; Devaux et al., 2008; Alamares et al., 2010; Schaap-Nutt et al., 2010; Satterfield et al., 2015). V proteins may inhibit both induction of the cellular interferon (IFN) response and IFN signalling through direct interactions with a multitude of host proteins. These functions have been comprehensively reviewed
elsewhere (Audsley and Moseley, 2013; Parks and Alexander-Miller, 2013; Ramachandran and Horvath, 2009). V also regulates viral RNA synthesis (Horikami et al., 1996; Witko et al., 2006; Parks et al., 2006; Nishio et al., 2002, 2005; Pisanelli et al., 2016), although the mechanism underpinning this remains unclear. Although V aids viral replication, it is non-essential (Curran et al., 1991) and is encoded by most but not all paramyxoviruses (Section 5.3). V is therefore considered a “luxury” protein.

The unique C-terminal region of V (VCT) contains a highly conserved cysteine-rich zinc finger domain, which binds two zinc ions (Liston and Briedis, 1994; Li et al., 2006a; Motz et al., 2013). A β-hairpin, anchored at its start and end by zinc-coordinating residues, is the only regular secondary structure within this domain. In some paramyxoviral V proteins, the conserved zinc finger domain immediately follows the edit site sequence. However in others, a linker of widely varying length and composition is observed (maximal length 136 aa, in CPIV-3). Overall, V is the second largest of the P gene proteins: with VCT ranging from 50 aa (in NiV) to 188 aa (in CPIV-3) in length.

The structural basis for V protein function has been investigated in several cases, and there are crystal structures of the full length parainfluenza virus 5 (PIV-5) V protein in complex with host protein DDB1 (DNA damage-binding protein 1; Li et al. (2006a)), and of the PIV-5 VCT in complex with host protein MDA5 (melanoma differentiation-associated protein 5; Motz et al. (2013)). One general conclusion from these studies is that the conformation of the zinc finger domain is overall malleable, and likely partially templated by the binding partner. Additionally, in the complex with DDB1, sequences from both N-terminal and C-terminal regions of the V protein are involved in binding, explaining how V protein activity sometimes arises from the coordinated action of both regions.

It appears that the functional roles of the V protein are evolving quite rapidly. Several observations support this.

First, some highly conserved biological functions of the V protein differ significantly in the way they are implemented. For example, while the vast majority of paramyxoviral V proteins bind STAT family members in order to suppress IFN signalling, the suppression is achieved in extremely diverse fashion. *Morbillivirus* V proteins bind STAT1 via their N terminal region, and STAT2 via their C-terminal region (Devaux et al., 2010; Röthlisberger et al., 2010; Chinnakannan et al., 2014). These binding events inhibit phosphorylation and nuclear translocation of the STATs. In contrast, *Rubulavirusae* V proteins generally bind STAT1 or STAT2 via the C-terminal region alone (Nishio et al., 2002, 2005; Pisanelli et al., 2016), and this leads to the targeted degradation of STATs via the proteosomal pathway. This requires the recruitment of additional host proteins, such as DDB1 (Lin et al., 1998; Andrejeva et al., 2002), that enable the polyubiquitination of STATs.

Second, there are clear examples of species-specific adaptations in V function which must have occurred relatively recently in evolutionary history. Considering STAT signal suppression by the rubulaviruses in more detail, species-specific adaptations of V protein function include (1) a gain in ability to bind and degrade STAT3 by MuV (Puri et al., 2009); (2) a loss of ability to degrade STATs by Human parainfluenza virus 4 (HPIV-4), despite the retention of STAT1/STAT2 binding activity (Nishio et al., 2005); (3) a complete loss of STAT binding activity by Tioman virus (TioPV; Caignard et al. (2013)); and (4) a switch to a mechanism involving mislocalisation rather than degradation of STAT proteins by Mapuera virus (MapV; Hagmaier et al. (2007)).

Overall we emphasise that the V protein is multifunctional and its exact function varies across genera, and among species. These functional adaptations likely reflect the unique selective pressures faced by each virus, associated with its tropism. The rapid molecular evolution of V appears to be linked to its role in mediating binding events and is likely enabled by its high levels of intrinsic disorder.

4.3 W protein

A third protein may also be generated by contranscriptional editing. Unlike P and V, its unique
C-terminal sequence is not conserved across paramyxoviral genera and consequently this protein has been assigned many names (Fontana et al., 2008) including W (Vidal et al., 1990a), D (Pelet et al., 1991; Galinski et al., 1992), PD (Wells and Malur, 2008), and I (Paterson and Lamb, 1990). For the purposes of this review, we use W to denote the protein encoded by the reading frame that encodes neither P nor V, and WCT to denote its unique C-terminal sequence.

There is evidence that W has evolved a function within some paramyxoviral genera. In all cases, W accumulates in the nucleus (Karsunke et al., 2019; Yang et al., 2019; Shaw et al., 2005; Lo et al., 2009; Wells and Malur, 2008). This is the situation for Newcastle disease virus (NDV; genus: *Orthoavulavirus*), Hendra and Nipah virus (HeV and NiV; genus: *Henipavirus*), and Human parainfluenza virus 3 (HPIV-3; genus: *Respirovirus*). Nuclear localisation signals can be identified in the unique region of the W protein (WCT) in each case (Shaw et al., 2005; Karsunke et al., 2019; Audsley et al., 2016a; Wells and Malur, 2008; Smith et al., 2018).

NDV sits alone, and we could not detect a homologous WCT in any other *Orthoavulavirus*. A recent study showed that deleting WCT impaired NDV replication in cultured cells, and this effect was revealed when the full-length W protein was supplied in trans (Yang et al., 2019). However, no detailed function has been assigned to this protein.

The *Henipavirus* W protein has the clearest functional linkages. The W protein influences the course of disease in animal models (Satterfield et al., 2015, 2016), and may play a direct role in subversion of the IFN response (Ciancanelli et al., 2009; Shaw et al., 2005; Keiffer et al., 2020). For example, NiV W can sequester unphosphorylated STAT proteins in the nucleus, via its N-terminal STAT1 binding site and C-terminal NLS, potentially inhibiting IFN signalling. NiV and HeV W were also recently discovered to modulate host gene expression by interacting with the 14-3-3 family of regulatory proteins, an interaction that depends upon phosphorylation of the penultimate serine residue in WCT (Edwards et al., 2020; Fig. 2).

For HPIV-3, in an early study, joint interruption of the V and W open reading frames attenuated viral replication (although individual interruptions had no effect; Durbin et al. (1999)). In interpreting this result, it should be noted that the V protein of HPIV-3 is abnormal, and likely to be expressed in truncated form (Section 5.3). A more recent study also suggests that WCT promotes viral genome transcription and replication, and is potentially also involved in the downregulation of β interferon expression (Roth et al., 2013). The C-terminal regions of HPIV-3, bovine parainfluenza virus 3 (BPIV-3; genus: *Respirovirus*), and caprine parainfluenza virus 3 (CPIV-3; genus: *Respirovirus*) W proteins have strong sequence similarity which is itself suggestive of shared function (Figure 2).

For remaining paramyxoviruses, WCT may not necessarily confer any biological function at all, and the region is often very short (2 aa in SeV, 6 aa in MeV, 11 aa in MuV; Chinnakannan et al. (2014); Horikami et al. (1996); Curran et al. (1991); Paterson and Lamb (1990)). However the W protein could still potentially exert biological effects through its shared N-terminal region, with synthesis of W potentially being more rapid than the synthesis of either P or V.

### 5 Evolution of the cotranscriptional gene editing system

#### 5.1 A maximum parsimony model for the evolution of P gene editing

Across the *Paramyxoviridae* there are differences in edit mode, with a faithful copy of the P gene encoding the P protein in some viruses, and the V protein in others. There are also differences in edit pattern, with the relative abundances of the transcripts encoding P, V, and W varying widely. Relative transcript abundance is defined by the probability distribution $p(G_m)$, where $m$ is the number of guanosines inserted. The most direct source of information about this distribution comes from sequencing the mRNA produced in virally infected cells. However, as Wignall-Fleming et al. (2019) have highlighted, if mRNA preparations are contaminated with anti-genomic RNA, the results may not faithfully reflect the actual abundance of mRNA. Furthermore, several studies have noted that transcript abundance...
varies with time post-infection (Kulkarni et al., 2009; Qiu et al., 2016a). In both cases the proportion of V and W transcripts increased as the infection progressed, though neither the mechanism nor functional implications are understood. Finally, while mRNA abundances are generally assumed to be related to encoded protein abundances, this may not always hold in practice (Liu et al., 2016).

With these caveats noted, the experimentally derived probability distributions (edit patterns) for 26 paramyxoviruses are displayed in Figure 3. The maximum observed insert size is G14 in NiV (Lo et al., 2009). Additional data on mRNA abundance, not displayed in the figure, can be found in the following publications – SeV: Pelet et al. (1991); Kato et al. (1997); NiV: Kulkarni et al. (2009); MeV: Liston and Briedis (1994); Millar et al. (2016); Donohue et al. (2019); NDV: Mebatsion et al. (2001); Yang et al. (2019); BeIv: Audsley et al. (2016b); TevPV: Johnson et al. (2019); Burroughs et al. (2015); HPIV-2: Ohgimoto et al. (1990); MuV: Takeuchi et al. (1990); CeMV: Bolt et al. (1995); PPRV: Mahapatra et al. (2003); PDV: Blixenkrone-Møller et al. (1992); PorPV: Berg et al. (1992).

The fundamental differences between viruses, apparent in Figure 3, reflect evolutionary events which have occurred throughout the history of the family. The following events are minimally required to explain the functional and evolutionary data: i) gain of the editing system, ii) loss of the editing system, iii) evolution of the V protein zinc finger motif and gain of biological function, iv) loss of the V protein zinc finger motif and associated function, v) switching of the edit mode and adaptation of the edit pattern, and vi) acquisition of unique function by the W protein. We estimated the evolutionary history of the Paramyxoviridae and inferred the ancestral lineages where these events occurred as follows: for each event we imputed the occurrence of the event onto branches such that the number of events required to explain the states observed at the leaves in the tree is minimised (Figure 4). This is the maximum parsimony model. An explicit limitation of this model is that it does not account for the full functional diversity of the V protein, which has multiple biological activities (Section 4.2). A maximum parsimony model for the evolution of P gene coding capacity has previously been developed (Jordan et al., 2000), but based on a much sparser data set.

5.2 Acquisition of the editing system and evolution of the V protein

The phosphoprotein gene editing system has not been detected beyond the Paramyxoviridae (Hyndman et al., 2012; Jordan et al., 2000). Therefore, the P gene editing system likely came into existence only once – in the lineage which led to the Paramyxoviridae. This event was coupled with the origin of the V protein; the evolution of its unique zinc binding motif; and the gain of many of its conserved functions (Figure 4). However the timing of these events cannot be resolved.

Cotranscriptional editing also occurs in the closely related Filoviridae family, although in a different gene. This independent adaptation of cotranscriptional editing as an overprinting mechanism may be a consequence of having a slippage prone polymerase, as all members of the order Mononegavirales exploit slippage to polyadenylate their mRNA (Conzelmann, 1998).

5.3 Partial or complete loss of the V protein

Under a maximum parsimony model, the V protein has been lost entirely on two independent occasions, both associated with the loss of the editing system (Figure 4). The C-terminal zinc binding domain has also been deleted, or significantly mutated, on two further occasions.

Loss of the V protein is associated with retirement of the cotranscriptional editing system – in lineage which lead to Human parainfluenza virus 1 (HPIV-1; genus: Respirovirus) and in the lineage which lead to Cedar virus (CedV; genus: Henipavirus). As these viruses once employed the P-mode, loss of the editing system was axiomatically coupled with loss of both V and W protein expression. It is possible that loss of V protein activity preceded loss of the edit system, but this is indeterminate. Retirement of the editing system appears impossible for
viruses employing the V-mode because the P protein is essential for polymerase function.

For both HPIV-1 and CedV, the edit site is not identifiable in the genome and edited mRNA could not be detected experimentally (Matsuoka et al., 1991; Marsh et al., 2012). In HPIV-1, the conserved V protein coding sequence is apparent in the genome however there is no clear mechanism for protein production due to the presence of multiple stop codons in the relevant reading frame (Matsuoka et al. (1991); Figure 5). This suggests that loss of V occurred quite recently in evolutionary history and there has been insufficient time for the sequences to diverge, creating a pseudogene. For CedV, only residual traces of the V protein coding sequence remain (Marsh et al., 2012).

In the case of HPIV-3, the edit site is operational (Galinski et al., 1992) and the zinc finger motif is detectable in the genome by sequence analysis (Figure 5). However, several stop codons between the edit site and the zinc finger prohibit production of the full-length V protein, unless further non-canonical transcriptional or translational mechanisms are invoked (Galinski et al., 1992). There are also two mutations in positions which are directly involved in zinc coordination (Figure 5). This suggests the VCT coding sequence is a pseudogene, similar to the situation in HPIV-1. In protein-based analysis of infected cells, the full V protein was not detected but a truncated variant which lacks the conserved C-terminal region was (Roth et al., 2013). Overall, current evidence suggests that the V protein of HPIV-3 is expressed in a truncated form lacking the canonical zinc binding motif. Its functional status is unclear.

Finally, in the case of the Jeilongviruses, the V protein C-terminal domain has been retained, but with mutation of several critical residues involved in zinc coordination (Figure 5). The C-terminal region does not interact with STAT1 or STAT2 (Audsley et al., 2016b), which is a conserved function of many other paramyxoviral V proteins (Section 4.2). Nonetheless, the Jeilongviral V protein has retained other functions, such as the ability to bind and inactivate the cytoplasmic RNA sensor MDA5 (Audsley et al., 2016b). This finding in particular highlights the multi-functional nature of the V protein, and the limitations of a nomenclature in which its multiple functionalities are not fully explicated.

The loss of the edit system or loss of the full length V protein may have implications for viral pathogenicity, although the interactions between virus and host are extremely complex. CedV (Marsh et al., 2012) causes no known disease, yet is very closely related to HeV and NiV which cause severe and frequently fatal disease in humans (Marsh and Wang, 2012). These viruses target the same family of cellular receptors (Laing et al., 2019) and the loss of V and W has been suggested as a contributor to attenuated virulence of CedV. Contrastingly, HPIV-1 and HPIV-3 are a leading cause of respiratory disease in humans, despite the absence or truncation of the V protein (Schomacker et al., 2012). Of possible significance is that CedV, HPIV-1, and HPIV-3 all produce “C proteins” from the P gene using translational overprinting mechanisms, and these C proteins have established roles as IFN antagonists (Mathieu et al., 2012; Schomacker et al., 2012). Hence there could once have been partial functional redundancy existing between V and C, which allowed for the loss of the V protein while maintaining some ability to evade the interferon system.

5.4 Switching of edit modes and adaption of edit patterns

The P-mode was likely the edit mode of the last common ancestor of the Paramyxoviridae. Under a maximum parsimony model, the editing system has switched to the V-mode four times during evolutionary history (Figure 4). These events occurred in the lineages that lead to: 1) Avian paramyxovirus 11 (APMV-11; genus: Metaavulavirus), 2) the Rubulavirinae subfamily, 3) the Ferlaviruses, and 4) Salem virus (SaIPV; genus: Salemvirus). Edit patterns have been experimentally investigated for three of these four clades: 10 rubulaviruses (Lau et al., 2010; Bowden et al., 2001; Chua et al., 2001; Southern et al., 1990; Ohgimoto et al., 1990; Kawano et al., 1993; Thomas et al., 1988; Paterson and Lamb, 1990; Takeuchi et al., 1990; Kondo et al., 1990), 2 Ferlaviruses (Woo et al., 2014; Kurath et al., 2004), and SaIPV (Renshaw et al., 2000).

In general the edit patterns of viruses that retain the ancestral P-mode (Figure 3, top panel)
are quite different to those of viruses that have subsequently adopted the V-mode (Figure 3, bottom panel). In the former, \( G_0 \) and \( G_1 \) insertions are most frequently observed, while in the latter, \( G_0 \) and \( G_2 \) insertions predominate. It seems clear that edit patterns have co-evolved with edit modes to maintain adequate production of P and V transcripts. In two clades (within the Respirovirus and Henipavirus genera), the edit patterns are long-tailed, and a significant fraction of the transcripts have more than 2 guanosine nucleotides inserted.

The edit pattern of SalPV (Figure 3, bottom panel) appears to be an outlier (Renshaw et al., 2000). The \( G_0 \)-centric distribution resembles those of viruses using the P-mode, and the relative abundance of P transcripts is very low. Given the taxonomic position of SalPV, as the most immediate outgroup of the Morbilliviruses (Figure 4), it could be that this is a virus that has switched edit mode but not yet adaptively evolved the edit pattern.

5.5 Acquisition of unique function by the W protein

Under our model, the W protein has evolved a novel function associated with its unique C-terminal region on 3 independent occasions (Figures 2 and 4): once for NDV (Yang et al., 2019; Karsunke et al., 2019), once for the henipaviral clade comprised of HeV and NiV (Shaw et al., 2005; Lo et al., 2009; Edwards et al., 2020), and once for the respiroviral clade composed of BPIV-3, HPIV-3, and CPIV-3 (Durbin et al., 1999; Pelet et al., 1991). There are varying levels of experimental evidence supporting the existence of a W protein function in these three clades (see Section 4.3). For the remaining paramyxoviruses, W has no known function. Rather, it is more likely that the expression of W is an inevitable byproduct of the editing system; an evolutionary spandrel (Gould and Lewontin, 1979).

For the most part, W transcripts are produced quite rarely (Figure 3). However, this does not appear to be the case for two clades where W has acquired function. Instead, the edit pattern is long-tailed, and the transcript probability \( p(G_{3k+2}) \) of producing a W ranges from 21 to 24% in HeV, NiV, BPIV-3, and HPIV-3 (Lo et al., 2009; Pelet et al., 1991; Galinski et al., 1992), and sometimes even higher in temporal analyses (Kulkarni et al., 2009).

In contrast, production of W is not significantly elevated for NDV (Steward et al., 1993; Mebatsion et al., 2001). The overall proportion of W transcript in NDV is estimated at around 8-9% (Steward et al., 1993; Qiu et al., 2016a; Yang et al., 2019) or as low as 2.4% (Mebatsion et al., 2001). However, experiments studying the effects of W protein knockout on viral replication (Yang et al., 2019), suggest that these low transcript abundances are optimal for fulfilling the unknown biological function of the NDV W protein (Section 4.3).

6 Molecular mechanism of cotranscriptional gene editing

In the Paramyxoviridae, cotranscriptional gene editing results from transcriptional slippage. This same process facilitates overprinting in other viruses (Sanchez et al., 1996; Shabman et al., 2014; Olspert et al., 2015; Rodamilans et al., 2015; Mehedi et al., 2011) and prokaryotes (Larsen et al., 2000; Penno et al., 2015; Mehedi et al., 2011). Slippage sites can also rescue an organism from deleterious frameshift mutations (Tamas et al., 2008).

Transcription has been extensively studied, most recently at the single-molecule level for the RdRP of bacteriophage \( \varphi 6 \) (Dulin et al., 2015a,b) and DNA-dependent RNA polymerases of prokaryotes, eukaryotes, and DNA viruses (Abbondanzieri et al., 2005; Shaevitz et al., 2003; Dangkulwanich et al., 2013; Larson et al., 2012; Skinner et al., 2004; Douglas et al., 2020, 2019). These studies have provided significant insights into the mechanisms underlying transcription elongation.

In this final section, we discuss cotranscriptional editing in the Paramyxoviridae under the framework presented in the single-molecule literature, noting some additional complexities.
which arise from the viral genome being packaged within a nucleocapsid.

### 6.1 Transcription elongation and slippage

Under a simple Brownian ratchet model, transcription elongation can be modelled as a cycle involving three canonical steps (Bar-Nahum et al. (2005); Abbondanzieri et al. (2005); Figure 6, large arrows). First, RNA polymerase steps forward along the template from the pretranslocated to the posttranslocated state, which frees the enzyme’s active site. Second, a complementary nucleoside triphosphate (NTP) binds to the active site. Third, the bound NTP is incorporated onto the 3′ end of the mRNA and pyrophosphate is released, thus restoring the system to the pretranslocated state.

Through backtracking, where the polymerase translocates upstream along the template (Komissarova and Kashlev, 1997; Abbondanzieri et al., 2005), and hypertranslocation, where it translocates downstream (Yarnell and Roberts, 1999), the polymerase can arrive at a catalytically inactive state (Figure 6). These processes can lead to transcriptional pausing (Saba et al., 2018; Artsimovitch and Landick, 2000). In the case of paramyxoviruses, extensive backtracking and hypertranslocation may be inhibited by the presence of nucleoproteins acting as “roadblocks”, analogous to the role played by nucleosomes in eukaryotic DNA transcription (Nuñez, 2012).

Slippage involves the movement of one sequence in the product/template hybrid relative to the other, which can lead to imperfect basepairing. Slippage was hypothesised by Streisinger et al. (1966) as one of the primary mechanisms of indel events. The mechanism is thought to involve formation of a nucleotide bulge near the 3′ end of the mRNA (Garcia Diaz and Kunkel, 2006). If the bulge forms in the nascent strand, an insertion can result, whereas a bulge in the template strand can lead to a deletion.

Based on studies of the behaviour of dsDNA molecules under applied force, Kühner et al. (2007) and Neher and Gerland (2004) hypothesise that slippage occurs in three steps (Figure 6). First, a bulge forms on one side of the hybrid. This initial reaction must overcome a large Gibbs energy barrier. Second, the bulge diffuses along the hybrid. Diffusion is likely to be quite rapid (Woodson and Crothers, 1987), and favoured if Watson-Crick basepairing is maintained in the bulged hybrid. Third, the bulge is absorbed at the other end of the hybrid. While these experiments were performed using DNA/DNA hybrids, the general model is likely to apply to all double helical nucleic acids. However due to the differing structural and dynamic properties of DNA/DNA, DNA/RNA and RNA/RNA hybrids (Bloomfield and Crothers, 2000), the propensity of a given nucleic acid sequence to slip may be very different in each setting.

### 6.2 Stuttering by the paramyxoviral polymerase

Through transcriptional slippage, a single templated nucleotide can be copied multiple times (stuttering). Stuttering is the proposed mechanism of cotranscriptional editing in paramyxoviruses. If correct, this model must explain many of the edit patterns presented in Figure 3. Some of these edit patterns are long-tailed, with the virus producing significant numbers of transcripts with more than 7 guanosine nucleotides inserted. Given the structural and energetic impediments to forming large bulge loops in duplexed nucleic acids (Longfellow et al., 1990; Turner and Mathews, 2010), a model in which these species result from the iterative formation of small bulges appears more realistic than a model invoking the direct formation of bulges of arbitrarily large size. However, this remains an assumption, as bulge formation at the P gene edit site has not yet been structurally and biophysically characterised.

The two distinct modes of editing (i.e. the P-mode and the V-mode) are encoded by quite different sequences (Figure 7).

The edit sites among viruses employing the P-mode are conserved. Using the PROSITE notation (Sigrist et al., 2002), the (genomic-sense) edit motif can be described by U(3,6)–C(2,6). In SeV, for example, the edit site sequence is UUUUUUCcC, where the lower case c is the stutter site i.e. the site reiteratively transcribed from the template resulting in a guanosine

https://mc.manuscriptcentral.com/vevolu
insertion into the mRNA (Hausmann et al., 1999b,a; Vidal et al., 1990b). Under the stuttering model, nucleotides are inserted as follows (Figure 6, left hand side): 1) a 1 nt bulge forms in the 3′ mRNA of the RNA/mRNA hybrid. 2) The bulge is free to diffuse along the hybrid. Although the bulge is thermodynamically disfavoured, it can occur because of U/A and non-canonical U/G basepairing which are maintained throughout diffusion. 3) In no particular order, the bulge is absorbed at the 5′ end and the lower-case c can be transcribed again. Each iteration of these three steps is associated with a $G_1$ insertion.

In contrast, the edit sites across the four clades of the V-mode group are quite distinct from one another. SaIPV is anomalous, and its edit site sequence resembles the P-mode group (Renshaw et al., 2000). This could explain the relatively low amounts of P transcript produced (Figure 3). The Ferlavirus edit site is distinct from all other known edit sites (Woo et al., 2014; Kurath et al., 2004) and the mechanism of guanosine insertion is not clear. Through convergent evolution, APMV-11 and the Rubulavirinae subfamily have similar edit sites (PROSITE: A(3,4)–U(2)–C–U(1,2)–C(4,7); genomic-sense). In the case of MuV, the edit site AAAUUCUCCC has been well characterised (Paterson and Lamb, 1990). Stuttering is proposed to occur in a fashion similar to SeV, however the edit site sequence allows $G_2$ inserts (encoding the P protein) to occur with greater frequency than $G_1$ inserts (encoding the W protein) due to the preferential formation of a 2 nucleotide bulge (Figure 6, right hand side). The iterative formation, diffusion, and absorption of 1 or 2 nucleotide bulges could account for the presence of larger insertions, which occur at quite low frequency (Figure 3).

In principle, transcriptional slippage could be initialised from any one of the states available to the polymerase (backtracked, pretranslocated, posttranslocated, or hypertranslocated; Figure 6). Because the editing process takes a finite time to occur, editing and pausing of the polymerase must be coupled to some extent (Vidal et al., 1990b; Hausmann et al., 1999a; Pelet et al., 1991). However it is not known if editing is associated with prolonged pausing, and the transition of the RdRP to a catalytically-inactive state. There is currently limited experimental data addressing this point. Partial substitution of guanosine triphosphate (GTP) with inosine triphosphate (ITP), in in vitro assays of SeV transcription, significantly enhanced P gene mRNA editing (Vidal et al., 1990b; Curran et al., 1993). As inosine incorporation promotes backtracking and/or pausing in other cellular and viral RNA polymerases (Shaevitz et al., 2003; Larson et al., 2012; Schweihr et al., 2014; Dulin et al., 2015a, 2017), the enhancement of P gene editing could reflect an increased time for editing to occur. However it might also reflect the perturbation of bulge formation and diffusion at the edit site, through the substitution of G:C with I:C pairings. Further experimental investigation of the linkage between editing and pausing is clearly needed.

Slight variation in the edit site sequence perturbs stuttering of the viral RdRP. For instance, when the length of the poly(A) sequence at the SeV edit site was increased, from A(3)–G(6) to A(8)–G(1), the average number of inserts increased dramatically (Hausmann et al., 1999a). Similarly, when the SeV edit site sequence was mutated to resemble that of BPIV-3, its edit pattern changed correspondingly (Hausmann et al., 1999b). These results speak to the primary importance of the genome sequence in governing polymerase stuttering. This is supported by studies on the polyviral RNA editing site, which can be transferred to the genome of an entirely different family of single-stranded RNA viruses, without complete loss of function (Stewart et al., 2019).

The roles that nucleoprotein displacement and the rule of six play during cotranscriptional editing have been investigated (Iseni et al., 2002; Hausmann et al., 1996; Kolakofsky, 2016). Changing the nucleoprotein phase around the edit site sequence (of SeV) resulted in an apparent change in edit pattern (Iseni et al., 2002). We computed the expected nucleoprotein phase at the edit site of each virus under the rule of six model. Although nucleoprotein displacement may play a role in editing, the nucleoprotein phase at the edit site does not appear to be well conserved (Figure 7).
7 Conclusion

The paramyxoviral P gene is subject to overprinting at both the transcriptional and translational levels. Here we have reviewed cotranscriptional editing of the P gene, which results in production of an essential protein (P), that is absolutely required for viral replication, as well as “luxury” proteins (V and W), that can aid viral replication by interfering with host defences (Figure 1). Consistent with their role, the V and W proteins are undergoing relatively rapid functional diversification. We have compiled the genomic sequences at the P gene edit site (Figure 7) as well as all existing quantitative data on the gene editing that occurs during transcription (Figure 3).

Based on the latter data, we have constructed an evolutionary model which incorporates some basic notions of protein function, and describes the minimal set of events required to account for the observed variations in the editing process (Figure 4). As structural and functional data on the P, V, and W proteins continues to accumulate, it should be possible to elaborate this model to incorporate the specific functional roles of P, V, and W.

Although transcriptional slippage provides the accepted physical mechanism for insertion of non-templated bases into the P gene, many aspects of this process remain ill-defined. Slippage at the edit site depends on bulge loop formation in the duplex RNA, however the structural and energetic behaviour underlying this process remains uncertain. It is also unclear how slippage is coordinated with either canonical or non-canonical steps of the transcription elongation pathway (Figure 6). Better models of the slippage process would help define some of the physical constraints that exist on the evolution of the remarkable gene overprinting system of the paramyxoviruses.

8 Virus abbreviations

AchPV 1-2 Achimota viruses 1-2 AnaPV Anaconda paramyxovirus APMV 2-13 Avian paramyxoviruses 2-13 APV A-C Antarctic penguin viruses A-C AsaPV Atlantic salmon paramyxovirus GH-M74a Ghanaian bat henipavirus BeiV Beilong virus BPIV-3 Bovine parainfluenza virus 3 CDV Canine distemper virus CedV Cedar virus CeMV Cetacean morbillivirus CPIV-3 Caprine parainfluenza virus 3 FdlV Fer de Lance virus FeMV Feline morbillivirus HeV Hendra virus HPIV 1-4 Human parainfluenza viruses 1-4 JPV J-virus MenPV Menangle virus MeV Measles virus MojV Mojiang virus MosPV Mossman virus MapV Mapuera virus MuV Mumps virus NarPV Nariva virus NDV Newcastle disease virus NiV Nipah virus PDV Phocine distemper virus PIV-5 Parainfluenza virus 5 PorPV Porcine rubulavirus PPIV-1 Porcine parainfluenza virus 1 PPRV Peste-des-petits-ruminants virus RPV Rinderpest virus SalPV Salem virus SeV Sendai virus SosPV Sosuga virus SunCV Sunshine coast virus SV-41 Simian virus 41 TePV Tenvirus ThkPV 1-3 Tuhoko viruses 1-3 TioPV Tioman virus TlmPV Tailam virus TupPV Tupaia virus

9 Algorithms and data availability

Sequences were aligned by M-Coffee (Wallace, 2006) and treated with subsequent manual adjustment using AliView (Larsson, 2014). Phylogenetic tree built with BEAST 2 (Bouck et al., 2019) from an alignment of the L protein, and a relaxed clock model Drummond et al. (2006). Sequence database accession numbers, P/V/W sequences, L alignment, and BEAST 2 input/output files are available at https://github.com/jordandouglas/ParamyxovirusSlippageEvolution.

10 Funding

https://mc.manuscriptcentral.com/vevolu
This work was supported by the University of Auckland Doctoral Scholarship.

References

Elio A Abbondanzieri, William J Greenleaf, Joshua W Shaevitz, Robert Landick, and Steven M Block. Direct observation of base-pair stepping by rna polymerase. *Nature*, 438(7067):460–465, 2005.

Ryan Abdella, Megha Aggarwal, Takashi Okura, Robert A Lamb, and Yuan He. Structure of a paramyxovirus polymerase complex reveals a unique methyltransferase-ctd conformation. *Proceedings of the National Academy of Sciences*, 117(9):4931–4941, 2020.

Judith G Alamares, Subbiah Elankumaran, Siba K Samal, and Ronald M Iorio. The interferon antagonistic activities of the v proteins from two strains of newcastle disease virus correlate with their known virulence properties. *Virus research*, 147(1):153–157, 2010.

Maher Alayyoubi, George P Leser, Christopher A Kors, and Robert A Lamb. Structure of the paramyxovirus parainfluenza virus 5 nucleoprotein–rna complex. *Proceedings of the National Academy of Sciences*, 112(14):E1792–E1799, 2015.

Gaya K Amarasinghe, María A Ayllón, Yíming Bào, Christopher F Basler, Sina Bavari, Kim R Blasdell, Thomas Briese, Paul A Brown, Alexander Bukreyev, Anne Balkema Buschmann, et al. Taxonomy of the order mononegavirales: update 2019. *Archives of virology*, 164(7):1967–1980, 2019.

J Andrejeva, E Poole, DF Young, S Goodbourn, and RE Randall. The p127 subunit (ddb1) of the uv-dna damage repair binding protein is essential for the targeted degradation of stat1 by the v protein of the paramyxovirus simian virus 5. *Journal of virology*, 76(22): 11379–11386, 2002.

Irina Artsimovitch and Robert Landick. Pausing by bacterial rna polymerase is mediated by mechanistically distinct classes of signals. *Proceedings of the National Academy of Sciences*, 97(13):7090–7095, 2000.

Michelle D Audsley and Gregory W Moseley. Paramyxovirus evasion of innate immunity: Diverse strategies for common targets. *World journal of virology*, 2(2):57, 2013.

Michelle D Audsley, David A Jans, and Gregory W Moseley. Nucleocyttoplasmic trafficking of nipah virus w protein involves multiple discrete interactions with the nuclear import and export machinery. *Biochemical and biophysical research communications*, 479(3):429–433, 2016a.

Michelle D Audsley, Glenn A Marsh, Kim G Lieu, Mary Tachedjian, D Albert Joubert, Lin-Fa Wang, David A Jans, and Gregory W Moseley. The immune evasion function of j and beilong virus v proteins is distinct from that of other paramyxoviruses, consistent with their inclusion in the proposed genus jeilongvirus. *Journal of General Virology*, 97 (3):581–592, 2016b.

B Bankamp, EN Lopareva, JR Kremer, Y Tian, MS Clemens, R Patel, AL Fowlkes, JR Kessler, CP Muller, WJ Bellini, et al. Genetic variability and mrna editing frequencies of the phosphoprotein genes of wild-type measles viruses. *Virus research*, 135(2):298–306, 2008.

Gil Bar-Nahum, Vitaly Epshtein, Andrei E Ruckenstein, Ruslan Rafikov, Arkady Mustaev, and Evgeny Nudler. A ratchet mechanism of transcription elongation and its control. *Cell*, 120(2):183–193, 2005.

MD Baron, MS Shaila, and T Barrett. Cloning and sequence analysis of the phosphoprotein...
gene of rinderpest virus. *Journal of general virology*, 74(2):299–304, 1993.

Bart G Barrell, GM Air, and CA Hutchison. Overlapping genes in bacteriophage φx174. *Nature*, 264(5581):34–41, 1976.

Robert Belshaw, Oliver G Pybus, and Andrew Rambaut. The evolution of genome compression and genomic novelty in ma viruses. *Genome research*, 17(10):1496–1504, 2007.

Mikael Berg, Bernt Hjertner, Jorge Moreno-López, and Tommy Linné. The p gene of the procine paramyxovirus lpmv encodes three possible polypeptides p, v and c: the p protein mrna is edited. *Journal of general virology*, 73(5):1195–1200, 1992.

Laurence Blanchard, Nicolas Tarbouriech, Martin Blackledge, Peter Timmins, Wilhelm P Burmeister, Rob WH Ruigrok, and Dominique Marion. Structure and dynamics of the nucleocapsid-binding domain of the sendai virus phosphoprotein in solution. *Virology*, 319(2):201–211, 2004.

Merete Blixenkrone-Möller, Bhaskar Sharma, Tamas M Varsanyi, Aizhung Hu, Erling Nor rby, and Jan Kövamees. Sequence analysis of the genes encoding the nucleocapsid protein and phosphoprotein (p) of phocid distemper virus, and editing of the p gene transcript. *Journal of general virology*, 73(4):885–893, 1992.

Victor Bloomfield and Donald M Crothers. *Nucleic acids: structures, properties and functions*. Number 574.192 B52. 2000.

Louis-Marie Bloyet, Joanna Brunel, Marion Dosnon, Véronique Hamon, Jenny Erales, Antoine Gruet, Carine Lazert, Christophe Bignon, Philippe Roche, Sonia Longhi, et al. Modulation of re-initiation of measles virus transcription at intergenic regions by pxd to ntail binding strength. *PLoS pathogens*, 12(12):e1006058, 2016.

Ronald Boeck, J Curran, Y Matsuoka, R Compsans, and D Kolakofsky. The parainfluenza virus type 1 p/c gene uses a very efficient gug codon to start its c’protein. *Journal of virology*, 66(3):1765–1768, 1992.

Gert Bolt, Søren Alexandersen, and Merete Blixenkrone-Møller. The phosphoprotein gene of a dolphin morbillivirus isolate exhibits genomic variation at the editing site. *Journal of general virology*, 76(12):3051–3058, 1995.

Remco Bouckaert, Timothy G Vaughan, Joëlle Barido-Sottani, Sebastián Duchène, Mathieu Fourment, Alexandra Gavryushkina, Joseph Heled, Graham Jones, Denise Kühnert, Nicola De Maio, et al. Beast 2.5: An advanced software platform for bayesian evolutionary analysis. *PLoS computational biology*, 15(4):e1006650, 2019.

Timothy R Bowden, Marcel Westenberg, Lin-Fa Wang, Bryan T Eaton, and David B Boyle. Molecular characterization of menangle virus, a novel paramyxovirus which infects pigs, fruit bats, and humans. *Virology*, 283(2):358–373, 2001.

Nadav Brandes and Michal Linial. Gene overlapping and size constraints in the viral world. *Biology Direct*, 11(1), May 2016. doi: 10.1186/s13062-016-0128-3. URL https://doi.org/10.1186/s13062-016-0128-3.

Axel Brennicke, Anita Marchfelder, and Stefan Binder. RNA editing. *FEMS microbiology reviews*, 23(3):297–316, 1999.

Ian Brierley and Francisco J Dos Ramos. Programmed ribosomal frameshifting in hiv-1 and the sars–cov. *Virus research*, 119(1):29–42, 2006.

Jessica F Bruhn, Katherine C Barnett, Jaclyn Bibby, Jens MH Thomas, Ronan M Keegan,
Daniel J Rigden, Zachary A Bornholdt, and Erica Ollmann Saphire. Crystal structure of the nipah virus phosphoprotein tetramerization domain. *Journal of virology*, 88(1): 758–762, 2014.

Jessica F Bruhn, Anne L Hotard, Christina F Spiropoulou, Michael K Lo, and Erica Ollmann Saphire. A conserved basic patch and central kink in the nipah virus phosphoprotein multimerization domain are essential for polymerase function. *Structure*, 27(4):660–668, 2019.

Wilhelm P. Burmeister, Nicolas Tarbouriech, Joseph Curran, and Rob W.H. Ruigrok. *Nature Structural Biology*, 7(9):777–781, September 2000. doi: 10.1038/79013. URL https://doi.org/10.1038/79013.

Amy L Burroughs, Mary Tachedjian, Gary Crameri, Peter A Durr, Glenn A Marsh, and Lin-Fa Wang. Complete genome sequence of teviot paramyxovirus, a novel rubulavirus isolated from fruit bats in australia. *Genome Announcements*, 3(2), 2015.

Grégory Caignard, Marianne Lucas-Hourani, Kevin P Dhondt, Jean-Louis Labernardière, Thierry Petit, Yves Jacob, Branka Horvat, Frédéric Tangy, and Pierre-Olivier Vidalain. The v protein of tioman virus is incapable of blocking type i interferon signaling in human cells. *PLoS One*, 8(1):e53881, 2013.

Philippe Calain and Laurent Roux. The rule of six, a basic feature for efficient replication of sendai virus defective interfering rna. *Journal of virology*, 67(8):4822–4830, 1993.

Roberto Cattaneo. Different types of messenger rna editing. *Annual review of genetics*, 25(1):71–88, 1991.

Roberto Cattaneo, Karin Kaelin, Knut Baczko, and Martin A Billeter. Measles virus editing provides an additional cysteine-rich protein. *Cell*, 56(5):759–764, 1989.

Senthil K Chinnakannan, Barbara Holzer, Beatriz Sanz Bernardo, Sambit K Nanda, and Michael D Baron. Different functions of the common p/v/w and v-specific domains of rinderpest virus v protein in blocking ifn signalling. *Journal of General Virology*, 95(1): 44–51, 2014.

Nicola Chirico, Alberto Vianelli, and Robert Belshaw. Why genes overlap in viruses. *Proceedings of the Royal Society B: Biological Sciences*, 277(1701):3809–3817, 2010.

Kaw Bing Chua, Lin-Fa Wang, Sai Kit Lam, Gary Crameri, Meng Yu, Terry Wise, David Boyle, Alex D Hyatt, and Bryan T Eaton. Tioman virus, a novel paramyxovirus isolated from fruit bats in malaysia. *Virology*, 283(2):215–229, 2001.

Michael J Ciancanelli, Valentina A Volchkova, Megan L Shaw, Viktor E Volchkov, and Christopher F Basler. Nipah virus sequesters inactive stat1 in the nucleus via a p gene encoded mechanism. *Journal of virology*, 83(16):7828–7841, 2009.

G. Communie, T. Crepin, D. Maurin, M. R. Jensen, M. Blackledge, and R. W. H. Ruigrok. Structure of the tetramerization domain of measles virus phosphoprotein. *Journal of Virology*, 87(12):7166–7169, April 2013a. doi: 10.1128/jvi.00487-13. URL https://doi.org/10.1128/jvi.00487-13.

Guillaume Communie, Johnny Habchi, Filip Yakubarski, David Blocquel, Robert Schneider, Nicolas Tarbouriech, Nicolas Papageorgiou, RW Ruigrok, Marc Jamin, Malene Ringkjøbing Jensen, et al. Atomic resolution description of the interaction between the nucleoprotein and phosphoprotein of hendra virus. *PLoS Pathog*, 9(9):e1003631, 2013b.

Karl-Klaus Conzelmann. Nonsegmented negative-strand rna viruses: genetics and
manipulation of viral genomes. *Annual review of genetics*, 32(1):123–162, 1998.

Robert Cox, Todd J Green, Sangeetha Purushotham, Champion Deivanayagam, Gregory J Bedwell, Peter E Prevelige, and Ming Luo. Structural and functional characterization of the mumps virus phosphoprotein. *Journal of virology*, 87(13):7558–7568, 2013.

J. Curran, R. Boeck, and D. Kolakofsky. The sendai virus p gene expresses both an essential protein and an inhibitor of RNA synthesis by shuffling modules via mRNA editing. *The EMBO Journal*, 10(10):3079–3085, October 1991. doi: 10.1002/j.1460-2075.1991.tb07860. x. URL https://doi.org/10.1002/j.1460-2075.1991.tb07860.x.

Joseph Curran and Daniel Kolakofsky. Ribosomal initiation from an acg codon in the sendai virus p/c mrna. *The EMBO journal*, 7(1):245–251, 1988.

Joseph Curran and Daniel Kolakofsky. Nonsegmented negative-strand rna virus rna synthesis in vivo. *Virology*, 371(2):227–30, 2008.

Joseph Curran, Thierry Pelet, Jean-Philippe Jacques, and Daniel Kolakofsky. Rna synthesis and mrna editing in paramyxovirus infections. In *Regulation of Gene Expression in Animal Viruses*, pages 125–139. Springer, 1993.

Manchuta Dangkulwanich, Toyotaka Ishibashi, Shixin Liu, Maria L Kireeva, Lucyna Lubkowska, Mikhail Kashlev, and Carlos J Bustamante. Complete dissection of transcriptional elongation reveals slow translocation of rna polymerase ii in a linear ratchet mechanism. *Elife*, 2:e00971, 2013.

Patricia Devaux, Gregory Hodge, Michael B McCchesney, and Roberto Cattaneo. Attenuation of v-or c-defective measles viruses: infection control by the inflammatory and interferon responses of rhesus monkeys. *Journal of virology*, 82(11):5359–5367, 2008.

Patricia Devaux, Andrew W Hudacek, Gregory Hodge, Jorge Reyes-del Valle, Michael B McCchesney, and Roberto Cattaneo. Stat1-blind measles virus cannot control inflammation and is attenuated in rhesus monkeys. *Journal of Virology*, 2010.

Ryan C Donohue, Christian K Pfaller, and Roberto Cattaneo. Cyclical adaptation of measles virus quasispecies to epithelial and lymphocytic cells: To v, or not to v. *PLoS pathogens*, 15(2):e1007605, 2019.

Jordan Douglas. UglyTrees: a browser-based multispecies coalescent tree visualiser. *Bioinformatics*, 2020.

Jordan Douglas, Richard Kingston, and Alexei J Drummond. Approximate bayesian computation of transcriptional pausing mechanisms. *BioRxiv*, page 748210, 2019.

Jordan Douglas, Richard Kingston, and Alexei J Drummond. Bayesian inference and comparison of stochastic transcription elongation models. *PLoS computational biology*, 16(2): e1006717, 2020.

Alexei J Drummond, Simon YW Ho, Matthew J Phillips, and Andrew Rambaut. Relaxed phylogenetics and dating with confidence. *PLoS biology*, 4(5): e88, 2006.

Venice Du Pont, Yi Jiang, and Richard K Plemper. Bipartite interface of the measles virus phosphoprotein x domain with the large polymerase protein regulates viral polymerase dynamics. *PLoS pathogens*, 15(8):e1007995, 2019.

David Dulin, Igor D Vilfan, Bojk A Berghuis, Susanne Hage, Dennis H Bamford, Minna M Poranen, Martin Depken, and Nynke H Dekker. Elongation-competent pauses govern the
fidelity of a viral rna-dependent rna polymerase. *Cell reports*, 10(6):983–992, 2015a.

David Dulin, Igor D Vilfan, Bojk A Berghuis, Minna M Poranen, Martin Depken, and Nynke H Dekker. Backtracking behavior in viral rna-dependent rna polymerase provides the basis for a second initiation site. *Nucleic acids research*, 43(21):10421–10429, 2015b.

David Dulin, Jamie J Arnold, Theo van Laar, Hyung-Suk Oh, Cheri Lee, Angela L Perkins, Daniel A Harki, Martin Depken, Craig E Cameron, and Nynke H Dekker. Signatures of nucleotide analog incorporation by an rna-dependent rna polymerase revealed using high-throughput magnetic tweezers. *Cell reports*, 21(4):1063–1076, 2017.

Anna P Durbin, Josephine M McAuliffe, Peter L Collins, and Brian R Murphy. Mutations in the c, d, and v open reading frames of human parainfluenza virus type 3 attenuate replication in rodents and primates. *Virology*, 261(2):319–330, 1999.

Megan R Edwards, Mikayla Hoad, Sofiya Tsimbalyuk, Andrea R Menicucci, Ilhem Mes saoudi, Jade K Forwood, and Christopher F Basler. Henipavirus w proteins interact with 14-3-3 to modulate host gene expression. *Journal of Virology*, 2020.

K Falk, WN Batts, A Kvellestad, G Kurath, J Wiik-Nielsen, and JR Winton. Molecular characterisation of atlantic salmon paramyxovirus (aspv): a novel paramyxovirus associated with proliferative gill inflammation. *Virus research*, 133(2):218–227, 2008.

Rachel Fearns and Richard K Plemper. Polymerases of paramyxoviruses and pneumoviruses. *Virus research*, 234:87–102, 2017.

Andrew E Firth and Ian Brierley. Non-canonical translation in rna viruses. *The Journal of general virology*, 93(Pt 7):1385, 2012.

Judith M. Fontana, Bettina Bankamp, and Paul A. Rota. Inhibition of interferon induction and signaling by paramyxoviruses. *Immunological Reviews*, 225(1):46–67, October 2008. doi: 10.1111/j.1600-065x.2008.00669.x. URL https://doi.org/10.1111/j.1600-065x.2008.00669.x.

Mark S Galinski, Roberta M Troy, and Amiya K Banerjee. Rna editing in the phosphoprotein gene of the human parainfluenza virus type 3. *Virology*, 186(2):543–550, 1992.

Miguel Garcia-Diaz and Thomas A Kunkel. Mechanism of a genetic glissando*: structural biology of indel mutations. *Trends in biochemical sciences*, 31(4):206–214, 2006.

Colomba Giorgi, Benjamin M Blumberg, and Daniel Kolakofsky. Sendai virus contains overlapping genes expressed from a single mrna. *Cell*, 35(3):829–836, 1983.

Stephen Jay Gould and Richard C Lewontin. The spandrels of san marco and the panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 205(1161):581–598, 1979.

Sergey G Guryanov, Lassi Liljeroos, Prasad Kasaragod, Tommi Kajander, and Sarah J Butcher. Crystal structure of the measles virus nucleoprotein core in complex with an n-terminal region of phosphoprotein. *Journal of virology*, 90(6):2849–2857, 2016.

Serafima Guseva, Sigrid Milles, Martin Blackledge, and Rob WH Ruigrok. The nucleoprotein and phosphoprotein of measles virus. *Frontiers in microbiology*, 10:1832, 2019.

Irina Gutsche, Ambroise Desfosses, Grégory Effantin, Wai Li Ling, Melina Haupt, Rob WH Ruigrok, Carsten Sachse, and Guy Schoehn. Near-atomic cryo-em structure of the helical measles virus nucleocapsid. *Science*, 348(6235):704–707, 2015.
Johnny Habchi and Sonia Longhi. Structural disorder within paramyxovirus nucleoproteins and phosphoproteins. *Mol. BioSyst.*, 8(1):69–81, 2012. doi: 10.1039/c1mb05204g. URL https://doi.org/10.1039/c1mb05204g.

Johnny Habchi, Stéphanie Blangy, Laurent Mamelli, Malene Ringkjøbing Jensen, Martin Blackledge, Hervé Darbon, Michael Oglesbee, Yaoling Shu, and Sonia Longhi. Characterization of the interactions between the nucleoprotein and the phosphoprotein of henipavirus. *Journal of Biological Chemistry*, 286(15):13583–13602, 2011.

K Hagmaier, N Stock, B Precious, K Childs, L-F Wang, S Goodbourn, and RE Randall. Mapuera virus, a rubulavirus that inhibits interferon signalling in a wide variety of mammalian cells without degrading stats. *The Journal of general virology*, 88(Pt 3):956, 2007.

Angela R Harrison and Gregory W Moseley. The dynamic interface of viruses with stats. *Journal of Virology*, 94(22), 2020.

Stéphane Hausmann, Jean-Philippe Jacques, and Daniel Kolakofsky. Paramyxovirus RNA editing and the requirement for hexamer genome length. *Rna*, 2(10):1033–1045, 1996.

Stephane Hausmann, Dominique Garcin, Christophe Delenda, and Daniel Kolakofsky. The versatility of paramyxovirus mrna polymerase stuttering. *Journal of virology*, 73(7):5568–5576, 1999a.

Stéphane Hausmann, Dominique Garcin, Anne-Sophie Morel, and Daniel Kolakofsky. Two nucleotides immediately upstream of the essential a6g3 slippery sequence modulate the pattern of g insertions during sendai virus mrna editing. *Journal of virology*, 73(1):343–351, 1999b.

N Herr, MN Webby, EMM Bulloch, M Schmitz, and RL Kingston. Nmr chemical shift assignment of the c-terminal region of the menangle virus phosphoprotein. *Biomolecular NMR assignments*, 13(1):195–199, 2019.

SANDRA M Horikami and SA Moyer. Synthesis of leader mrna and editing of the p mrna during transcription by purified measles virus. *Journal of virology*, 65(10):5342–5347, 1991.

Sandra M. Horikami, Sherin Smallwood, and Sue A. Moyer. The sendai virus v protein interacts with the NP protein to regulate viral genome RNA replication. *Virology*, 222(2):383–390, August 1996. doi: 10.1006/viro.1996.0435. URL https://doi.org/10.1006/viro.1996.0435.

Timothy H Hyndman, Rachel E Marschang, James FX Wellehan Jr, and Philip K Nicholls. Isolation and molecular identification of sunshine virus, a novel paramyxovirus found in Australian snakes. *Infection, Genetics and Evolution*, 12(7):1436–1446, 2012.

Frédéric Iseni, Florence Baudin, Dominique Garcin, Jean-Baptiste Marq, Rob WH Ruigrok, and Daniel Kolakofsky. Chemical modification of nucleotide bases and mrna editing depend on hexamer or nucleoprotein phase in sendai virus nucleocapsids. *Rna*, 8(8):1056–1067, 2002.

J.P. Jacques, S. Hausmann, and D. Kolakofsky. Paramyxovirus mrna editing leads to g deletions as well as insertions. *The EMBO Journal*, 13(22):5496–5503, November 1994. doi: 10.1002/j.1460-2075.1994.tb06884.x. URL https://doi.org/10.1002/j.1460-2075.1994.tb06884.x.

Marc Jamin and Filip Yabukarski. Nonsegmented negative-sense mrna viruses—structural data bring new insights into nucleocapsid assembly. In *Advances in virus research*, volume 97, 2021.
pages 143–185. Elsevier, 2017.

Malene Ringkjøbing Jensen, Filip Yabukarski, Guillaume Communie, Eric Condamine, Caroline Mas, Valentina Volchkova, Nicolas Tarbouriech, Jean-Marie Bourhis, Viktor Volchkov, Martin Blackledge, et al. Structural description of the nipah virus phosphoprotein and its interaction with stat1. *Biophysical Journal*, 2020.

Kenth Johansson, Jean-Marie Bourhis, Valerie Campanacci, Christian Cambillau, Bruno Canard, and Sonia Longhi. Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the c-terminal domain of the nucleoprotein. *Journal of Biological Chemistry*, 278(45):44567–44573, August 2003. doi: 10.1074/jbc.m308745200. URL https://doi.org/10.1074/jbc.m308745200.

Rebecca I Johnson, Mary Tachedjian, Bronwyn A Clayton, Rachel Layton, Jemma Bergfeld, Lin-Fa Wang, and Glenn A Marsh. Characterization of teviot virus, an australian bat borne paramyxovirus. *Journal of General Virology*, 100(3):403–413, 2019.

I King Jordan, Ben A Sutter IV, and Marcella A McClure. Molecular evolution of the paramyxoviridae and rhabdoviridae multiple-protein-encoding p gene. *Molecular biology and evolution*, 17(1):75–086, 2000.

David Karlin and Robert Belshaw. Detecting remote sequence homology in disordered proteins: Discovery of conserved motifs in the n-termini of mononegavirales phosphoproteins. *PLoS ONE*, 7(3):e31719, March 2012. doi: 10.1371/journal.pone.0031719. URL https://doi.org/10.1371/journal.pone.0031719.

Julia Karsunke, Sandra Heiden, Magdalena Murr, Axel Karger, Kati Franzke, Thomas C. Mettenleiter, and Angela Römer-Oberdörfer. W protein expression by newcastle disease virus. *Virus Research*, 263:207–216, April 2019. doi: 10.1016/j.virusres.2019.02.003. URL https://doi.org/10.1016/j.virusres.2019.02.003.

Atsushi Kato, Katsuhiko Kiyotani, Yuko Sakai, Tetsuya Yoshida, Tatsuo Shioda, and Yoshiyuki Nagai. Importance of the cysteine-rich carboxyl-terminal half of v protein for sendai virus pathogenesis. *Journal of Virology*, 71(10):7266–7272, 1997.

Mitsuo Kawano, Masato Tsurudome, Naohiro Oki, Machiko Nishio, Hiroshi Komada, Haruo Matsumura, Shigeru Kusagawa, Hisataka Ohta, and Yasuhiko Ito. Sequence determination of the p gene of simian virus 41: presence of irregular deletions near the rna-editing sites of paramyxoviruses. *Journal of general virology*, 74(5):911–916, 1993.

Paul K Keese and Adrian Gibbs. Origins of genes:” big bang” or continuous creation? *Proceedings of the National Academy of Sciences*, 89(20):9489–9493, 1992.

Timothy R Keiffer, Michael J Ciancaneli, Megan R Edwards, and Christopher F Basler. Interactions of the nipah virus p, v, and w proteins across the stat family of transcription factors. *Msphere*, 5(6), 2020.

Richard L Kingston, Damon J Hamel, Leslie S Gay, Frederick W Dahlquist, and Brian W Matthews. Structural basis for the attachment of a paramyxoviral polymerase to its template. *Proceedings of the National Academy of Sciences of the United States of America*, 101(22):8301–8306, 2004.

Richard L Kingston, Leslie S Gay, Walter S Baase, and Brian W Matthews. Structure of the nucleocapsid-binding domain from the mumps virus polymerase; an example of protein folding induced by crystallization. *Journal of molecular biology*, 379(4):719–731, 2008.

Daniel Kolakofsky. Paramyxovirus rna synthesis, mrna editing, and genome hexamer phase: https://mc.manuscriptcentral.com/vevolu
Daniel Kolakofsky, Thierry Pelet, Dominique Garcin, Stéphane Hausmann, Joseph Curran, and Laurent Roux. Paramyxovirus RNA synthesis and the requirement for hexamer genome length: the rule of six revisited. Journal of virology, 72(2):891–899, 1998.

Natalia Komissarova and Mikhail Kashlev. Transcriptional arrest: Escherichia coli RNA polymerase translocates backward, leaving the 3’ end of the RNA intact and extruded. Proceedings of the National Academy of Sciences, 94(5):1755–1760, 1997.

Sachin Kulkarni, Valentina Volchkova, Christopher F Basler, Peter Palese, Viktor E Volchkov, and Megan L Shaw. Nipah virus edits its p gene at high frequency to express the v and w proteins. Journal of virology, 83(8):3982–3987, 2009.

Matthew H Larson, Jing Zhou, Craig D Kaplan, Murali Palangat, Roger D Kornberg, Robert Landick, and Steven M Block. Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase ii. Proceedings of the National Academy of Sciences, 109(17):6555–6560, 2012.
Anders Larsson. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30(22):3276–3278, August 2014. doi: 10.1093/bioinformatics/ btu531. URL https://doi.org/10.1093/bioinformatics/btu531.

Patrizia Latorre, Daniel Kolakofsky, and Joseph Curran. Sendai virus y proteins are initiated by a ribosomal shunt. *Molecular and Cellular Biology*, 18(9):5021–5031, 1998.

Susanna KP Lau, Patrick CY Woo, Beatrice HL Wong, Annette WP Wong, Hoi-Wah Tsoi, Ming Wang, Paul Lee, Huifang Xu, Rosana WS Poon, Rongtong Guo, et al. Identification and complete genome analysis of three novel paramyxoviruses, tuhoko virus 1, 2 and 3, in fruit bats from china. *Virology*, 404(1):106–116, 2010.

Susanna KP Lau, Patrick CY Woo, Ying Wu, Annette WP Wong, Beatrice HL Wong, Candy CY Lau, Rachel YY Fan, Jian-Piao Cai, Hoi-Wah Tsoi, Kwok-Hung Chan, et al. Identification and characterization of a novel paramyxovirus, porcine parainfluenza virus 1, from deceased pigs. *Journal of General Virology*, 94(10):2184–2190, 2013.

Philippe Le Mercier and Daniel Kolakofsky. Bipartite promoters and rna editing of paramyxoviruses and filoviruses. *RNA*, 25(3):279–285, 2019.

Ti Li, Xiujuan Chen, Kenneth C. Garbutt, Pengbo Zhou, and Ning Zheng. Structure of DDB1 in complex with a paramyxovirus v protein: Viral hijack of a propeller cluster in ubiquitin ligase. *Cell*, 124(1):105–117, January 2006a. doi: 10.1016/j.cell.2005.10.033. URL https://doi.org/10.1016/j.cell.2005.10.033.

Zhuo Li, Meng Yu, Hong Zhang, Danielle E Magoffin, Philippa JM Jack, Alex Hyatt, Hai Yan Wang, and Lin-Fa Wang. Beilong virus, a novel paramyxovirus with the largest genome of non-segmented negative-stranded rna viruses. *Virology*, 346(1):219–228, 2006b.

Grace Y Lin, Reay G Paterson, Christopher D Richardson, and Robert A Lamb. The v protein of the paramyxovirus sv5 interacts with damage-specific dna binding protein. *Virology*, 249(1):189–200, 1998.

Peter Liston and Dalius J Briedis. Measles virus v protein binds zinc. *Virology*, 198(1): 399–404, 1994.

Yansheng Liu, Andreas Beyer, and Ruedi Aebersold. On the dependency of cellular protein levels on mrna abundance. *Cell*, 165(3):535–550, 2016.

Michael K Lo, Brian H Harcourt, Bruce A Mungall, Azaibi Tamin, Mark E Peeples, William J Bellini, and Paul A Rota. Determination of the henipavirus phosphoprotein gene mrna editing frequencies and detection of the c, v and w proteins of nipah virus in virus-infected cells. *Journal of General Virology*, 90(2):398–404, 2009.

Carl E Longfellow, Ryszard Kierzek, and Douglas H Turner. Thermodynamic and spectroscopic study of bulge loops in oligoribonucleotides. *Biochemistry*, 29(1):278–285, 1990.

Sonia Longhi, Louis-Marie Bloyet, Stefano Gianni, and Denis Gerlier. How order and disorder within paramyxoviral nucleoproteins and phosphoproteins orchestrate the molecular interplay of transcription and replication. *Cellular and Molecular Life Sciences*, 74(17): 3091–3118, 2017.

Madhuchhanda Mahapatra, Satya Parida, Berhe G Egziabher, Adama Diallo, and Tom Barrett. Sequence analysis of the phosphoprotein gene of peste des petits ruminants (ppr) virus: editing of the gene transcript. *Virus Research*, 96(1-2):85–98, 2003.

Ivan G Maia, Karin Séron, Anne-Lise Haenni, and Françoise Bernardi. Gene expression from viral rna genomes. *Plant molecular biology*, 32(1-2):367–391, 1996.
Glenn A Marsh and Lin-Fa Wang. Hendra and nipah viruses: why are they so deadly? *Current opinion in virology*, 2(3):242–247, 2012.

Glenn A Marsh, Carol De Jong, Jennifer A Barr, Mary Tachedjian, Craig Smith, Deborah Middleton, Meng Yu, Shawn Todd, Adam J Foord, Volker Haring, et al. Cedar virus: a novel henipavirus isolated from australian bats. *PLoS pathogens*, 8(8):e1002836, 2012.

Cyrille Mathieu, Vanessa Guillaume, Valentina A Volchkova, Christine Pohl, Frederique Jacquot, Ren Yih Looi, Kum Thong Wong, Catherine Legras-Lachuer, Viktor E Volchkov, Joel Lachuer, et al. Nonstructural nipah virus c protein regulates both the early host proinflammatory response and viral virulence. *Journal of virology*, 86(19):10766–10775, 2012.

Y Matsuoka, J Curran, T Pelet, D Kolakofsky, R Ray, and RW Companys. The p gene of human parainfluenza virus type 1 encodes p and c proteins but not a cysteine-rich v protein. *Journal of virology*, 65(6):3406–3410, 1991.

Teshome Mebatsion, Stefan Verstegen, Leonarda TC De Vaan, Angela Römer-Oberdörfer, and Carla C Schriër. A recombinant Newcastle disease virus with low-level v protein expression is immunogenic and lacks pathogenicity for chicken embryos. *Journal of virology*, 75(1):420–428, 2001.

Masfique Mehedi, Darryl Falzarano, Jochen Seebach, Xiaojie Hu, Michael S Carpenter, Hans-Joachim Schnittler, and Heinz Feldmann. A new ebola virus nonstructural glycoprotein expressed through ma editing. *Journal of virology*, 85(11):5406–5414, 2011.

EL Millar, LJ Rennick, B Weissbrich, J Schneider-Schaulies, WP Duprex, and BK Rima. The phosphoprotein genes of measles viruses from subacute sclerosing panencephalitis cases encode functional as well as non-functional proteins and display reduced editing. *Virus research*, 211:29–37, 2016.

Sigrid Milles, Malene Ringkjøbing Jensen, Carine Lazert, Serafima Guseva, Stefiannia Ivashchenko, Guillaume Communie, Damien Maurin, Denis Gerlier, Rob WH Ruigrok, and Martin Blackledge. An ultraweak interaction in the intrinsically disordered replication machinery is essential for measles virus function. *Science advances*, 4(8):eaat7778, 2018.

Carina Motz, Kerstin Monika Schuhmann, Axel Kirchhofer, Manuela Moldt, Gregor Witte, Karl-Klaus Konzelm, and Karl-Peter Hopfner. Paramyxovirus v proteins disrupt the fold of the rna sensor mda5 to inhibit antiviral signaling. *Science*, 339(6120):690–693, 2013.

Baibaswata Nayak, Sachin Kumar, Peter L Collins, and Siba K Samal. Molecular characterization and complete genome sequence of avian paramyxovirus type 4 prototype strain duck/hong kong/d3/75. *Virology journal*, 5(1):124, 2008.

Richard A Neher and Ulrich Gerland. Dynamics of force-induced dna slippage. *Physical review letters*, 93(19):198102, 2004.

Machiko Nishio, Dominique Garcin, Viviane Simonet, and Daniel Kolakofsky. The carboxyl segment of the mumps virus v protein associates with stat proteins in vitro via a tryptophan-rich motif. *Virology*, 300(1):92–99, 2002.

Machiko Nishio, Masato Tsurudome, Morihiro Ito, and Yasuhiro Ito. Human parainfluenza virus type 4 is incapable of evading the interferon-induced antiviral effect. *Journal of virology*, 79(23):14756–14768, 2005.

Machiko Nishio, Junpei Ohtsuka, Masato Tsurudome, Tetsuya Nosaka, and Daniel Kolakofsky. Human parainfluenza virus type 2 v protein inhibits genome replication by binding to the l protein: possible role in promoting viral fitness. *Journal of virology*, 82(13):6130–6138, 2008.
Sarah L. Noton and Rachel Fears. Initiation and regulation of paramyxovirus transcription and replication. *Virology*, 479:545–554, 2015.

Evgeny Nudler. RNA polymerase backtracking in gene regulation and genome instability. *Cell*, 149(7):1438–1445, 2012.

Shinji Ohgimoto, Hisanori Bando, Mitsuo Kawano, Kousuke Okamoto, Kunio Kondo, Masato Tsurudome, Machiko Nishio, and Yasuhiko Ito. Sequence analysis of p gene of human parainfluenza type 2 virus: P and cysteine-rich proteins are translated by two mRNAs that differ by two nontemplated G residues. *Virology*, 177(1):116–123, 1990.

Satoshi Ohnishi, Hironori Kamikubo, Masayoshi Onitsuka, Mikio Kataoka, and David Shortle. Conformational preference of polyglycine in solution to elongated structure. *Journal of the American Chemical Society*, 128(50):16338–16344, 2006.

Allan Olspert, Betty Y-W Chung, John F Atkins, John P Carr, and Andrew E Firth. Transcriptional slippage in the positive-sense RNA virus family potyviridae. *EMBO reports*, 16(8):995–1004, 2015.

Christopher L Parks, Susan E Witko, Cheryl Kotash, Shuo L Lin, Mohinder S Sidhu, and Stephen A. Udem. Role of V protein RNA binding in inhibition of measles virus minigenome replication. *Virology*, 348(1):96–106, 2006.

Griffith D Parks and Martha A Alexander-Miller. Paramyxovirus activation and inhibition of innate immune responses. *Journal of molecular biology*, 425(24):4872–4892, 2013.

Reay G. Paterson and Robert A. Lamb. RNA editing by G-nucleotide insertion in mumps virus p-gene mRNA transcripts. *Journal of virology*, 64(9):4137–4145, 1990.

John B. Patterson, Diane Thomas, Hanna Lewicki, Martin A. Billette, and Michael BA Oldstone. V and C proteins of measles virus function as virulence factors in vivo. *Virology*, 267(1):80–89, 2000.

Thierry Pelet, Joseph Curran, and Daniel Kolakofsky. The P gene of bovine parainfluenza virus 3 expresses all three reading frames from a single mRNA editing site. *The EMBO journal*, 10(2):443–448, 1991.

Christophe Penno, Virag Sharma, Arthur Coakley, Mary O’Connell Motherway, Douwe van Sinderen, Lucyna Lubkowska, Maria L. Kireeva, Mikhail Kashlev, Pavel V. Baranov, and John F. Atkins. Productive mRNA stem loop-mediated transcriptional slippage: crucial features in common with intrinsic terminators. *Proceedings of the National Academy of Sciences*, 112(16):E1984–E1993, 2015.

Adrian Pickar, Pei Xu, Andrew Elson, Zhuo Li, James Zengel, and Biao He. Roles of serine and threonine residues of mumps virus P protein in viral transcription and replication. *Journal of virology*, 88(8):4414–4422, 2014.

Giuseppe Pisanelli, Maudry Laurent-Rolle, Balaji Manicassamy, Alan Belicha-Villanueva, Juliet Morrison, Bernardo Lozano-Dubernard, Felipa Castro-Peralta, Giuseppe Iovane, and Adolfo Garcia-Sastre. La piedad michoacan mexico virus P protein antagonizes type I interferon response by binding STAT2 protein and preventing STATS nuclear translocation. *Virus research*, 213:11–22, 2016.

Sébastien Plumet, W Paul Duprex, and Denis Gerlier. Dynamics of viral mRNA synthesis during measles virus infection. *Journal of virology*, 79(11):6900–6908, 2005.

C. R. Pringle. The order Mononegavirales. *Archives of Virology*, 117(1-2):137–140, March.
Mamta Puri, Ken Lemon, W Paul Duprex, Bertus K Rima, and Curt M Horvath. A point mutation, e95d, in the mumps virus v protein disengages stat3 targeting from stat1 targeting. Journal of virology, 83(13):6347–6356, 2009.

X Qiu, Q Fu, C Meng, S Yu, Y Zhan, L Dong, T Ren, Y Sun, L Tan, C Song, et al. Kinetic analysis of ma editing of newcastle disease virus p gene in the early period of infection. Acta virologica, 60(1):71–77, 2016a.

Xusheng Qiu, Yuan Zhan, Chunxun Meng, Junqing Wang, LuNa Dong, Yingjie Sun, Lei Tan, Cuiping Song, Shengqing Yu, and Chan Ding. Identification and functional analysis of phosphorylation in newcastle disease virus phosphoprotein. Archives of virology, 161(8):2103–2116, 2016b.

Aparna Ramachandran and Curt M Horvath. Paramyxovirus disruption of interferon signal transduction: Status report. Journal of Interferon & Cytokine Research, 29(9):531–537, 2009.

Corinne Rancurel, Mahvash Khooravi, A Keith Dunker, Pedro R Romero, and David Karlin. Overlapping genes produce proteins with unusual sequence properties and offer insight into de novo protein creation. Journal of virology, 83(20):10719–10736, 2009.

Randall W Renshaw, Amy L Glaser, Hana Van Campen, Frank Weiland, and Edward J Dubovi. Identification and phylogenetic comparison of salem virus, a novel paramyxovirus of horses. Virology, 270(2):417–429, 2000.

Bert Rima, Peter Collins, Andrew Easton, Ron Fouchier, Gael Kurath, Robert A Lamb, Benhur Lee, Andrea Maisner, Paul Rota, and Lin-Fa Wang. Problems of classification in the family paramyxoviridae. Archives of virology, 163(5):1395–1404, 2018.

Bernardo Rodamilans, Adrian Valli, Ares Mingot, David San León, David Baulcombe, Juan J López-Moya, and Juan A García. Rna polymerase slippage as a mechanism for the production of frameshift gene products in plant viruses of the potyviridae family. Journal of virology, 89(13):6965–6967, 2015.

Jason P Roth, Joseph K-K Li, John D Morrey, Dale L Barnard, and Almut H Vollmer. Deletion of the d domain of the human parainfluenza virus type 3 (hpiv3) pd protein results in decreased viral rna synthesis and beta interferon (ifn-β) expression. Virus genes, 47(1):10–19, 2013.

Anne Röthlisberger, Dominique Wiener, Matthias Schweizer, Ernst Peterhans, Andreas Zurbriggen, and Philippe Plattet. Two domains of the v protein of virulent canine distemper virus selectively inhibit stat1 and stat2 nuclear import. Journal of virology, 84(13):6328–6343, 2010.

Jason Saba, Xien Chua, Tatiana V Mishanina, Dhananjaya Nayak, Tricia A Windgassen, Rachel Anne Mooney, and Robert Landick. The elemental mechanism of transcriptional pausing. bioRxiv, page 422220, 2018.

Niv Sabath, Andreas Wagner, and David Karlin. Evolution of viral proteins originated de novo by overprinting. Molecular Biology and Evolution, 29(12):3767–3780, July 2012. doi: 10.1093/molbev/mss179. URL https://doi.org/10.1093/molbev/mss179.

Paramananda Saikia, M Gopinath, and MS Shaila. Phosphorylation status of the phosphoprotein p of rinderpest virus modulates transcription and replication of the genome.
Archives of virology, 153(4):615–626, 2008.

Anthony Sanchez, Sam G Trappier, BW Mahy, Clarence J Peters, and Stuart T Nichol. The virion glycoproteins of ebola viruses are encoded in two reading frames and are expressed through transcriptional editing. Proceedings of the National Academy of Sciences, 93(8): 3602–3607, 1996.

Benjamin A Satterfield, Robert W Cross, Karla A Fenton, Krystle N Agans, Christopher F Basler, Thomas W Geisbert, and Chad E Mire. The immunomodulating v and w proteins of nipah virus determine disease course. Nature communications, 6(1):1–15, 2015.

Benjamin A Satterfield, Robert W Cross, Karla A Fenton, Viktoriya Borisevich, Krystle N Agans, Daniel J Deer, Jessica Graber, Christopher F Basler, Thomas W Geisbert, and Chad E Mire. Nipah virus c and w proteins contribute to respiratory disease in ferrets. Journal of virology, 90(14):6326–6343, 2016.

Christian J Sauder, Vahan Simonyan, Laurie Ngo, Konstantinos Karagiannis, Yu Cong, Cheryl Zhang, Rong Wang, Wells W Wu, Tahir Malik, and Steven A Rubin. Evidence that a polyhexameric genome length is preferred, but not strictly required, for efficient mumps virus replication. Virology, 493:173–188, 2016.

Anne Schaap-Nutt, Christopher D’Angelo, Margaret A Scull, Emerito Amaro-Carambot, Machiko Nishio, Raymond J Pickles, Peter L Collins, Brian R Murphy, and Alexander C Schmidt. Human parainfluenza virus type 2 v protein inhibits interferon production and signaling and is required for replication in non-human primates. Virology, 397(2):285–298, 2010.

Henrick Schomacker, Anne Schaap-Nutt, Peter L Collins, and Alexander C Schmidt. Pathogenesis of acute respiratory illness caused by human parainfluenza viruses. Current opinion in virology, 2(3):294–299, 2012.

Volker Schweikhard, Cong Meng, Kenji Murakami, Craig D Kaplan, Roger D Kornberg, and Steven M Block. Transcription factors tfiif and tfiis promote transcript elongation by rna polymerase ii by synergistic and independent mechanisms. Proceedings of the National Academy of Sciences, 111(18):6642–6647, 2014.

Reed S Shabman, Omar J Jabado, Chad E Mire, Timothy B Stockwell, Megan Edwards, Milind Mahajan, Thomas W Geisbert, and Christopher F Basler. Deep sequencing identifies noncanonical editing of ebola and marburg virus mas in infected cells. MBio, 5(6): e02011–14, 2014.

Joshua W Shaevitz, Elio A Abbondanzieri, Robert Landick, and Steven M Block. Backtracking by single rna polymerase molecules observed at near-base-pair resolution. Nature, 426(6967):684–687, 2003.

Jessica A Shaffer, William J Bellini, and Paul A Rota. The c protein of measles virus inhibits the type i interferon response. Virology, 315(2):389–397, 2003.

Megan L Shaw, Washington B Cardenas, Dmitriy Zamarin, Peter Palese, and Christopher F Basler. Nuclear localization of the nipah virus w protein allows for inhibition of both virus- and toll-like receptor 3-triggered signaling pathways. Journal of virology, 79(10): 6078–6088, 2005.

Christian JA Sigrist, Lorenzo Cerutti, Nicolas Hulo, Alexandre Gattiker, Laurent Falquet, Marco Pagni, Amos Bairoch, and Philipp Bucher. Prosite: a documented database using patterns and profiles as motif descriptors. Briefings in bioinformatics, 3(3):265–274, 2002.
Mario H Skiadopoulos, Leatrice Vogel, Jeffrey M Riggs, Sonja R Surman, Peter L Collins, and Brian R Murphy. The genome length of human parainfluenza virus type 2 follows the rule of six, and recombinant viruses recovered from non-polyhexameric-length antigenic cdnas contain a biased distribution of correcting mutations. *Journal of virology*, 77(1): 270–279, 2003.

Gary M Skinner, Christoph G Baumann, Diana M Quinn, Justin E Molloy, and James G Hoggett. Promoter binding, initiation, and elongation by bacteriophage T7 RNA polymerase a single-molecule view of the transcription cycle. *Journal of Biological Chemistry*, 279(5): 3239–3244, 2004.

Katrina Sleeman, Bettina Bankamp, Kimberly B Hummel, Michael K Lo, William J Bellini, and Paul A Rota. The c, v and w proteins of nipah virus inhibit minigenome replication. *Journal of General Virology*, 89(5):1300–1308, 2008.

Kate M Smith, Sofiya Tsimbalyuk, Megan R Edwards, Emily M Cross, Jyoti Batra, Tatiana P Soares da Costa, David Aragão, Christopher F Basler, and Jade K Forwood. Structural basis for importin alpha 3 specificity of w proteins in hendra and nipah viruses. *Nature communications*, 9(1):1–13, 2018.

Julien Sourimant, Vidhi D Thakkar, Robert M Cox, and Richard K Plemper. Viral evolution identifies a regulatory interface between paramyxovirus polymerase complex and nucleocapsid that controls replication dynamics. *Science Advances*, 6(10):eaaz1590, 2020.

JA Southern, B Precious, and RE Randall. Two nontemplated nucleotide additions are required to generate the p mRNA of parainfluenza virus type 2 since the m RNA genome encodes protein v. *Virology*, 177(1):388–390, 1990.

Michael Steward, I Barry Vipond, Neil S Millar, and Peter T Emmerson. RNA editing in Newcastle disease virus. *Journal of General Virology*, 74(12):2539–2547, 1993.

Hazel Stewart, Allan Olspert, Benjamin G Butt, and Andrew E Firth. Propensity of a picornavirus polymerase to slip on potyvirus-derived transcriptional slippage sites. *The Journal of general virology*, 100(2):199, 2019.

George Streisinger, Yoshimi Okada, Joyce Emrich, Judith Newton, Akira Tsugita, Eric Terzaghi, and M Inouye. Frameshift mutations and the genetic code. In *Cold Spring Harbor Symposium on Quantitative Biology*, volume 31, pages 77–84. Cold Spring Harbor Laboratory Press, 1966.

Akihiro Sugai, Hiroki Sato, Misako Yoneda, and Chieko Kai. Phosphorylation of measles virus phosphoprotein at s86 and/or s151 downregulates viral transcriptional activity. *FEBS letters*, 586(21):3900–3907, 2012.

Dengyun Sun, Priya Luthra, Zhuo Li, and Biao He. Plk1 down-regulates parainfluenza virus 5 gene expression. *PLoS pathogens*, 5(7):e1000525, 2009.

Kaoru Takeuchi, Kiyoshi Tanabayashi, Michiko Hishiyama, Yasuko K Yamada, Akio Yamada, and Akira Sugiura. Detection and characterization of mumps virus v protein. *Virology*, 178(1):247–253, 1990.

Ivica Tamas, Jennifer J Wernegreen, Björn Nystedt, Seth N Kauppinen, Alistair C Darby, Laura Gomez-Valero, Daniel Lundin, Anthony M Poole, and Siv GE Andersson. Endosymbiont gene functions impaired and rescued by polymerase infidelity at poly (a) tracts. *Proceedings of the National Academy of Sciences*, 105(39):14934–14939, 2008.

Nicolas Tarbouriech, Joseph Curran, Rob WH Ruigrok, and Wilhelm P Burmeister. Tetrameric
coiled coil domain of sendai virus phosphoprotein. *Nature Structural & Molecular Biology*, 7(9):777, 2000.

Patricia A. Thibault, Ruth E. Watkinson, Andres Moreira-Soto, Jan F. Drexler, and Benhur Lee. Zoonotic potential of emerging paramyxoviruses. In *Advances in Virus Research*, pages 1–55. Elsevier, 2017. doi: 10.1016/bs.aivir.2016.12.001. URL https://doi.org/10.1016/bs.aivir.2016.12.001.

Sheila M. Thomas, Robert A. Lamb, and Reay G. Paterson. Two mRNAs that differ by two non-templated nucleotides encode the amino coterminally proteins p and v of the paramyxovirus SV5. *Cell*, 54(6):891–902, 1988.

Hoang T. Tran, Albert Mao, and Rohit V. Pappu. Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *Journal of the American Chemical Society*, 130(23):7380–7392, 2008.

Douglas H. Turner and David H. Mathews. Nnndb: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Research*, 38(suppl_1):D280–D282, 2010.

Silvia Vidal, Joseph Curran, and Daniel Kolakofsky. Editing of the sendai virus p/c mRNA by g insertion occurs during mRNA synthesis via a virus-encoded activity. *Journal of Virology*, 64(1):239–246, 1990a.

Silvia Vidal, Joseph Curran, and Daniel Kolakofsky. A stuttering model for paramyxovirus p mRNA editing. *The EMBO Journal*, 9(6):2017, 1990b.

Valentina A. Volchkova, Olga Dolnik, Miguel J. Martinez, Olivier Reynard, and Viktor E. Volchkov. Genomic mRNA editing and its impact on Ebola virus adaptation during serial passages in cell culture and infection of guinea pigs. *The Journal of Infectious Diseases*, 204(suppl_3):S941–S946, 2011.

I. M. Wallace. M-coffee: combining multiple sequence alignment methods with t-coffee. *Nucleic Acids Research*, 34(6):1692–1699, March 2006. doi: 10.1093/nar/gkl091. URL https://doi.org/10.1093/nar/gkl091.

Melissa N. Webby, Matthew P. Sullivan, Kavestri M. Yegambaram, Mazdak Radjainia, Jeremy R. Keown, and Richard L. Kingston. A method for analyzing the composition of viral nucleoprotein complexes, produced by heterologous expression in bacteria. *Virolgy*, 527:159–168, 2019.

Michael Weik, Sven Enterlein, Kathrin Schlenz, and Elke Mühlberger. The Ebola virus genomic replication promoter is bipartite and follows the rule of six. *Journal of Virology*, 79(16):10660–10671, 2005.

Greg Wells and Achut Malur. Expression of human parainfluenza virus type 3 pd protein and intracellular localization in virus infected cells. *Virus Genes*, 37(3):358–367, 2008.

S. P. J. Whelan, J. N. Barr, and G. W. Wertz. Transcription and replication of nonsegmented negative-strand RNA viruses. In *Current Topics in Microbiology and Immunology*, pages 61–119. Springer Berlin Heidelberg, 2004. doi: 10.1007/978-3-662-06099-5_3. URL https://doi.org/10.1007/978-3-662-06099-5_3.

Elizabeth B. Wignall-Fleming, David J. Hughes, Sreenu Vattipally, Sejal Modha, Steve Goodbourn, Andrew J. Davison, and Richard E. Randall. Analysis of paramyxovirus transcription and replication by high-throughput sequencing. *Journal of Virology*, pages JVI–
Table 1: Summary of paramyxovirus taxonomy (Amarasinghe et al., 2019), including notable host species (Thibault et al., 2017).

| Subfamily          | Genus          | Host species        | Type species                                      |
|--------------------|----------------|---------------------|--------------------------------------------------|
| Avulavirinae       | Metaavulavirus | Bird                | Avian parainfluenza virus 2 (APMV-2)               |
|                    | Orthoavulavirus| Bird                | Newcastle disease virus (NDV)                     |
|                    | Paraavulavirus | Bird                | Avian parainfluenza virus 3 (APMV-3)              |
| Rubulavirinae      | Orthorubulavirus| Bat, human, pig    | Mumps virus (MuV)                                |
|                    | Paranrubulavirus| Bat, human, pig    | Menangle virus (MenPV)                            |
| Orthoparamyxovirinae| Aquaparamyxovirus | Fish              | Atlantic salmon paramyxovirus (AsaPV)            |
|                    | Ferlavirus     | Reptile             | Fer de Lance virus (FdLV)                         |
|                    | Henipavirus    | Bat                 | Hendra virus (HeV)                               |
|                    | Jeilongvirus   | Rodent              | Beilong virus (BeiV)                             |
|                    | Morbillivirus  | Cat, dolphin, human| Measles virus (MeV)                              |
|                    | Narmovirus     | Rodent              | Nariva virus (NarV)                              |
|                    | Respirovirus   | Cow, human, rodent  | Sendai virus (SeV)                               |
|                    | Salemvirus     | Horse               | Salem virus (SaV)                                |

Figure 1: Cotranscriptional editing of the P gene. The two observed modes of editing are depicted: these are the P- and V-modes. A single transcript can encode one of P, V, or W depending on the number of guanosines stochastically inserted at the edit site during transcription. While the P, V, and W proteins all share a common N-terminal region (NT), their C-terminal regions (PCT, VCT, and WCT) are distinct.
Figure 2: W protein C-terminal regions (WCT). For the displayed sequences there is experimental data regarding the cellular localisation or function of the W protein, or a W protein homolog in another virus. All numbering is relative to the start of WCT. Sites are coloured by amino acid characteristic if the characteristic is 100% conserved at the alignment position. Under the ClustalX colouring scheme hydrophobic residues are blue, positively charged residues – red, negatively charged residues – magenta, polar residues – green, cysteine – pink, glycine – orange, proline – yellow, and aromatic residues – cyan (Larkin et al., 2007).

Figure 3: Experimentally derived frequency distributions (edit patterns) describing guanosine nucleotide insertion at the P gene edit site. To facilitate comparison, the viruses are grouped by edit mode (P-mode or V-mode). Not included in the figure are several P-mode paramyxoviruses (CedV and HPIV-1) in which P gene editing does not occur, and for which P protein mRNA is the sole species produced. The total proportion of transcripts encoding the three functionally distinct mRNA species are indicated for each experiment. The bulk of the experimental data was obtained by cDNA sequencing, for which the number of sequenced transcripts $n$ is specified. Experimental data for BPIV-3 was obtained by a primer extension method acting directly on the mRNA population, and hence $n$ is not specified. Viral genera indicated in bottom right, see Section 8 for virus names.

Figure 4: Phylogeny of the Paramyxoviridae. Tree created from an alignment of the viral L protein, with Sunshine coast virus (SunCV; Hyndman et al. (2012)) as an outgroup. Coloured rectangles on branches indicate a hypothesised evolutionary event occurring some time in that lineage. Clade posterior supports are shown on the internal nodes. Branches lengths are proportional to time such that there is an average of 1 amino acid substitution per unit of time. See Section 8 for virus names. Tree visualised using UglyTrees (Douglas, 2020).

Figure 5: Cysteine-rich C-terminal regions of the V protein. The first amino acid in each aligned sequence is numbered relative to the start of the V protein. The size of the linker which connects the shared N-terminal region of V to the first aligned position is indicated. The arrows at the top of the alignment indicate residues whose side chains directly coordinate bound zinc ions, based on structural analysis of the PIV-5 V protein (Li et al., 2006a). Asterisks denote stop codons. Sites are coloured by amino acid group if a group is at least 70% conserved at the alignment position (colour scheme indicated in Figure 2). Among paramyxoviruses which have retained the ancestral V protein, the displayed region is invariant at 13 out of 59 positions across the entire group. The tree is the same as that in Figure 4.

Figure 6: State diagrams of Brownian ratchet and slippage models. Plausible stuttering pathways for SeV (accession: AB039658; genomic position: 2783) and MuV (accession: EU884413; genomic position: 2432) are shown, with a RNA/mRNA hybrid of 7 bp in length. The figure depicts single nucleotide insertion (for SeV) or double nucleotide insertion (for MuV). Insertions of other sizes may be possible and a single nucleotide insertion must certainly occur in MuV at low frequency. A nucleoprotein protomer bound to the viral genome (top strand) is depicted by the coloured octagon. Large arrows indicate the canonical transcription elongation pathway, double-ended triangular arrows denote equivalency between two connecting states, and unlabelled arrows describe translocation reactions. While slippage initialises in the pretranslocated state in this diagram, the actual state where this process initialises is unknown.
Figure 7: Edit site sequences in the paramyxoviruses. The sequences of the negative sense (genomic) RNA are displayed. The numbers indicate the genomic position of the first displayed nucleotide. P- and V-modes are denoted by P and V respectively. Nucleoprotein phases are displayed; the first nucleotide within each nucleoprotein protomer is highlighted in black. This tree is the same as that in Figure 4.
Genome encoded

Produced by cotranscriptional editing

mRNA (+)

P-mode

V-mode

Translation

Transcription

Edited site

P protein: soyuz modules

V protein: zinc finger

W protein: C

Protein

Binding domain

Coiled coil

Sizes not to scale
## W protein

**Unique C-terminal region**

### Orthoavulavirus

|   | NDV | NDV |
|---|-----|-----|
| 1 | G A H G R A P K R G T T N V R L N S R E V N P A A E T A R K D R R T K S R P P L E T R A Q T R T Q H I M D N G R S H N Y | Q L V Q P L M L S D Q G R A K T I P L Y L R I M S S H L |

### Henipavirus

|   | HeV | NiV |
|---|-----|-----|
| 1 | G A Q T R S L N M L G R K T C L G R R V V Q P G M F A D Y P P T K K A R V L L R R M S N | G A Q T R N I H L L G R K T C L G R R V V Q P G M F E D H P P T K K A R V S M R R M S N |

### Respirovirus

|   | CIPV-3 | BPIV-3 | HPIV-3 |
|---|--------|--------|--------|
| 1 | G R E K R K N G A K E R E E R D N R E E R K G R R E E R K R R S R L G E E S I Q C P | S Q S P R Q R E R I Q P K Q Q A G T K E K D Q H P W T H A P R D N T L K P T D S A M E R R D P Q E T R A G H S D Q R Y | Q P S T P T Q R G K Q K Y R Q N H Q K H N L H H G I S S T T T P T E T N Q A Q L L Q Q L Q D Q P I Q K N R S E |

|   | CIPV-3 | BPIV-3 | HPIV-3 |
|---|--------|--------|--------|
| 61 | S K S A R P - Q G N Q Q D H R V D P K G R E S H P G T S S T A R A A V E Q N Q A K T L R Y Q H L G R A I Q W D Q A E | Q P S T P T Q R G K Q K Y R Q N H Q K H N L H H G I S S T T T P T E T N Q A Q L L Q Q L Q D Q P I Q K N R S E | |
