 doi:10.20944/preprints202104.0484.v1

Review

From RNA world to SARS-CoV-2: the edited story of RNA viral evolution

Zachary W. Kockler [1] and Dmitry A. Gordenin [2]*

[1] Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences, US National Institutes of Health, Research Triangle Park, North Carolina, United States of America; zachary.kockler@NIH.gov
[2] Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences, US National Institutes of Health, Research Triangle Park, North Carolina, United States of America; gordenin@niehs.nih.gov
* Correspondence: gordenin@niehs.nih.gov

Abstract: The current SARS- CoV-2 pandemic underscores the importance of understanding the evolution of RNA genomes. While RNA is subject to the formation of similar lesions as DNA, the evolutionary and physiological impacts RNA lesions have on viral genomes are yet to be characterized. Lesions that may drive the evolution of RNA genomes can induce breaks that are repaired by recombination or can cause base substitution mutagenesis, also known as base editing. Over the past decade or so, base editing mutagenesis of DNA genomes has been subject to many studies, revealing that exposure of ssDNA is subject to hypermutation that is involved in the etiology of cancer. However, base editing of RNA genomes has not been studied to the same extent. Recently hypermutation of single-stranded RNA viral genomes have also been documented though its role in evolution and population dynamics. Here, we will summarize the current knowledge of key mechanisms and causes of RNA genome instability covering areas from the RNA world theory to the SARS- CoV-2 pandemic of today. We will also highlight the key questions that remain as it pertains to RNA genome instability, mutations accumulation, and experimental strategies for addressing these questions.

Keywords: RNA world theory; messenger RNA; Viral RNA; Genome stability; Viral evolution; Hypermutation, APOBEC, ADAR, RNA editing.

1. Introduction

One of the favorite quotes that Miro Radman uses in his presentations is the title of Theodosius Dobzhansky's essay "Nothing in Biology Makes Sense Except in the Light of Evolution" [1]. In his Synthetic Theory of Evolution, Dobzhansky defined two major factors – genetic variation (i.e., mutation and other types of genome instability) and natural selection [2] – that interplay in generating new species. Remarkably high instability levels of RNA genomes accelerate speciation to the levels that often allow documenting evolution in real time. Besides the unique opportunity for researchers, this, at times, represents a considerable threat to the species hosting RNA viral genomes, including ongoing pandemic of SARS-CoV-2. The latter resulted in recent boom of attention to mechanisms of genome instability in RNA viruses.

In today's biological systems, the genetic material making up the genome is primarily DNA. In contrast, a plethora of viruses that infect cellular hosts throughout all kingdoms rely upon RNA as their primary genetic material. Viral RNA, as well as DNA, genomes use the host organism/biological system to template and synthesize proteins to perform all functions necessary for creating new virus particles and for transmitting their genetic information to progeny. Whichever the genetic material of the virus genome, there is the requirement that the genome remains stable to allow for the transmission of viable

 doi:10.20944/preprints202104.0484.v1

genetic material to progeny, and to prevent the extinction of species [3-6]. However, non-catastrophic levels of genome instabilities are instrumental for accumulating beneficial variants to prepare a species to meet the challenges of ever-changing environments and allow for downstream evolution [7-10]. Therefore, a balance between a stable genome and instances of genome instability must be met.

To date, there has been numerous studies into the stability/instabilities of DNA genomes, but the same level of research has not been performed for RNA. This disparity is important to note because RNA genomes are predicted to be a vital key to biological evolution, as prior to the last universal common ancestor (LUCA) the RNA world theory predicts the existence of protocells that used RNA as a genetic material. Further, RNA was also proposed to be used as an enzyme to mediate all metabolic functions (as proteins had not yet evolved) [11-18]. With the reliance upon RNA for the genetic material, as well as for cellular function, there must have been efficient replication of RNA within the cell, but the modes of replication of these protocells are not known. There have been a few proposed mechanisms for self-driven RNA propagation [11-13,19-22], but one such proposed mechanism gained support from a recent communication [22] of the in vitro evolution of a holopolymerase ribozyme that can search and identify a promotor, and perform processive synthesis. This suggests that these protocells may have evolved a ribozyme for efficient RNA synthesis. However, for such a ribozyme to evolve there is the prerequisite for an RNA synthesis mechanism that would not have utilized a ribozyme. To get a better picture for how these protocells replicated, along with their likely sources of genome instability, remnants of the mechanisms of protocell replication may still remain within the genomes of modern RNA viruses.

The modes of replication of modern viruses are known (discussed below) and have been found to have the highest mutation rates per nucleotide among all biological species [23]. Viral RNA genomes are not as stable as DNA genomes, and this could be due to multiple factors including; special features of RNA genomes, RNA virus replication machinery, high selection pressure, and the susceptibility of viral RNA to environmental and/or endogenous lesions [24,25]. Thus, these instabilities of RNA virus genomes in turn should speed up their evolution. These evolutionary insights are especially important in the light of the current (at time of publishing) COVID-19 pandemic caused by the RNA Coronavirus SARS- CoV-2. Just how SARS-CoV-2 evolved to become transmissible between humans is not yet known, but likely a coronavirus infecting an animal host jumped to infect humans [26,27]. Such a jump would require the introduction of variants through non-catastrophic events of genome instability to gain an evolutionary advantage. How these key variants were introduced remains unclear, but in this review, we discuss possible and likely sources of genome instability that introduced these variants as well as highlight key remaining questions from the RNA world to SARS-CoV-2 pandemic.

2. Replication-transcription cycles in RNA viruses

2.1. RNA can be a carrier of genetic information through generations

The first protocells of the RNA world theory had to replicate their genome to pass the genetic material on to progeny, but the mechanism of how they replicated remains unclear. A recent communication from the Unrau Lab [22] describes in vitro evolution experiments that results in a holopolymerase ribozyme that can search and identify a promotor followed by processive elongation synthesis. This would suggest that replication of protocell's genome could proceed via a ribozyme, but this would require the presence of two RNA molecules per protocell. This happening in the very first protocells is unlikely, as a chance of having two copies of RNA -one for the genome and one for the actual replicase ribozyme- arising independently in the same early protocell is low [13,19]. Thus, prior to ribozyme evolution, there must have been a mechanism for replicating RNA independent of ribozymes. However, the development of a model for non-enzymatic replication of RNA genomes is at best in its infancy. Specifically, the problem is that, to date, a long-tract non-enzymatic RNA replication mechanism in nature has yet to be found (reviewed in [11]), which is further compounded by the difficulty to separate long stretches

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 April 2021 doi:10.20944/preprints202104.0484.v1

of replicated RNA strands, should long tract RNA be synthesized. This inability lends to the likelihood of a hypothesis that, instead of a long tract synthesis, a shorter type of synthesis is utilized [13,21]. This idea has its own problems based on the ends of the replicated genome. A non-enzymatic replication would be required to begin at one end and continue through the other end, which would require an improbable standard oligo for all replication events to act as a primer at the very beginning of the genome. While the other problem arises at the terminal end where the last base is added at a low efficiency in what is called the "last base addition problem" [28] due to the imidazolium-bridged dinucleotide intermediates typically requiring two bases to extend, of which is not available at the very end of the template. These two problems can be resolved if the RNA genome template was instead a circle, as there would be no beginning and have no end, so the synthesis could be primed anywhere on the circle and there would be no "last base addition problem". Nevertheless, replication of a ssRNA circle to form a dsRNA circle would require efficient long tract RNA synthesis that, should it be successful, would cause the circle to become highly strained due to the high bending energy of dsRNA. Therefore, the early protocells likely would not have a circular RNA genome.

Even with the stated problems above, Szostak and colleagues proposed a new model for replicating "primordial RNA genomes" through what they call a virtual circular genome [11]. This virtual circular genome contains multiple oligos that cover the entire circular RNA genome. Replication of the virtual circular genome will come after the annealing of the oligos and allow for templated addition of activated monomers, dimers, or trimers to allow for extension of the oligo. These oligos can then switch templates to allow for a continued elongation of the oligos to slowly replicate the entirety of the genome, allowing newly synthesized genetic material to pass on to progeny. Further, this mode of replication would also offer the ability for more than one copy of the RNA genome to be present in the same protocell, opening the door to the evolution of ribozymes, and consistent with the results observed in [22].

The replicating protocells could have set the track for evolution that ultimately (after many evolutionary steps) arrived at where we are today. Though, we do not know the intermediate steps between the protocells to today, it is still possible that remnants of these protocells remain as pathogens. Specifically, RNA viruses and viroids are two types of pathogens that use RNA for their genomes. Viroids are the smallest infectious pathogens known and contain non-protein coding RNAs sized just 200-400 nucleotides [29-31]. However, most of RNA viruses do encode for some of their own proteins, but otherwise rely on cellular transcription and translation systems for the necessary proteins. The proteins that the virus does encode are specific for each virus type including capsid proteins, coat proteins, and RNA replication machinery resulting in different viral structures and modes of viral replication.

2.2. Single-strand and double-strand RNA virus genomes.

Generally, there are three classes of RNA viruses, and the key feature for separating the classes is based on the state in which the viral RNA genome (i.e., RNA packaged into the virion) is present [32,33]. The first class of RNA viruses maintains their genomes as double stranded (ds) RNA while the other two classes are single-stranded (ss). The single-stranded RNA viruses are further separated by the polarity of the genomic strand. The ssRNA genome can be formed by the positive or (+) strand, which also functions as a mRNA for viral proteins, or by a negative or (-) strand [32,33]. With these viruses being maintained differently requires different modes of replication (reviewed in [34]). Specifically, in positive-strand ssRNA viruses, RNA dependent RNA polymerase (RdRp) uses the positive genomic strand as a template to create a new negative strand copy (the anti-genome) that is subsequently used as a template to create large numbers of positive-strand viral RNA genomes. Alternatively, in negative-strand RNA viruses RdRp uses the genomic negative-strand as a template to create positive-strand antigenome, that also serve as mRNAs. RdRp subsequently uses the positive-strand anti-genome as a template to create large numbers of negative-strand viral RNA genomes. dsRNA viruses replicate their

 doi:10.20944/preprints202104.0484.v1

genome differently by generating positive-strand mRNAs (templated by the dsRNA genome) that are also used by RdRp as a template to create dsRNA genomes packaged into new virus particles. It should be noted that each virus type (even dsRNA viruses) relies upon multiple copies of ssRNA as intermediates of replication. This is important because the bases in ssRNA viruses are not protected by hydrogen bonds as they are in dsRNA viruses, therefore these forms may be more prone to lesions. The polarity of the lesions in the resulting genomes depends on the virus class (Figure 1). In positive-strand ssRNA and in dsRNA viruses the predominant ssRNA species is the positive-strand, which is also an mRNA translated into viral proteins (Figure 1 A and C). In negative-strand ssRNA viruses (Figure 1B), the negative RNA strand is more abundant. This bias in strand abundance can affect mutation accumulation bias, which may then be detected as mutation spectra strand bias (see below). Further, these lesions can become breaks in the RNA genomes where breakage of dsRNA genome will still maintain an unbroken strand to hold the genome together, but ssRNA can result in irreversible separation of genome sections. Additionally, should these lesions be encountered by RdRp the replication will stall resulting in incomplete copy that can be subsequently utilized by RNA recombination that often results in genome rearrangements [35].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 April 2021 doi:10.20944/preprints202104.0484.v1

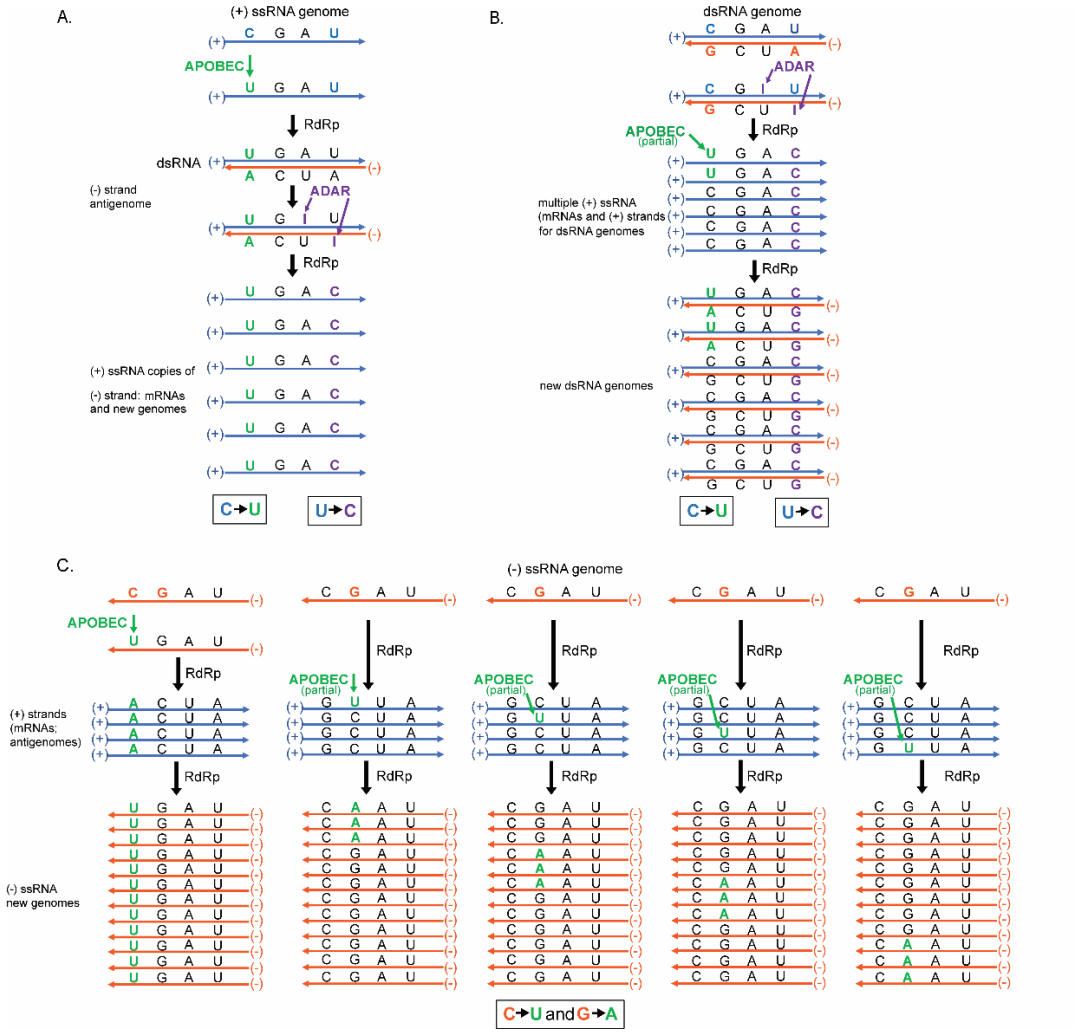

Figure 1. RNA virus genome type and replication mode define mutation strand bias in the progeny.

Presented are examples of mutagenesis with the ssRNA-specific cytidine deaminase APOBEC and the dsRNA-specific adenosine deaminase ADAR starting from a single viral genome infecting a cell. Positive (+) strands are shown in blue. Negative (-) strands are shown in orange. Color codes, same as a strand color, are assigned to original non-mutated nucleotides that will be altered (mutated) in the next steps. APOBEC mutagenesis and resulting mutant nucleotides are shown in green. ADAR mutagenesis in dsRNA and resulting mutant nucleotides are shown in purple. ADAR mutagenesis that could occur in ds parts of folded ssRNA molecules is presumed to be less frequent than ADAR deamination in fully dsRNA and thus not illustrated. Nucleotides that stayed not mutated in the progeny are shown in black through all steps. Predominant classes of mutations in progeny presented as changes in ssRNA genomes (panels A and C) or in coding (+) strands RNA of dsRNA genomes (panel B) are shown in boxes. (**A**) Viruses with positive (+) ssRNA genome. The infecting (+) strand RNA molecule is used as a template by RNA dependent RNA polymerase (RdRp) to synthesize a dsRNA with both (+) and (-) strands. A single dsRNA molecule is subsequently used to generate multiple copies of (+) strand RNA transcripts and/or genomes. A single APOBEC-induced C to U change in the infecting genomic (+) strand ssRNA would amplify in all viral progeny (C to U mutations). An ADAR-induced A to I (inosine) change in the (+) strand dsRNA would not reproduce in genomes of viral progeny. In contrast, an ADAR-induced A to I change in the (-) strand dsRNA would be copied into multiple (+) strand RNA transcripts and thus be amplified in the viral progeny as U to C mutations in genomic (+) strand ssRNA. (**B**) Viruses with double-stranded (ds) RNA genomes. Multiple (+) ssRNA transcripts and/or genome precursors are generated by RdRp. Each (+) ssRNAs precursor is then used to generate a dsRNA genome. Only ADAR-induced A to I mutations in (-) strand are amplified into multiple dsRNA genomes via copies of (+) strands. Since there are multiple (+) strand intermediates, there is a chance of detectable level of C to U APOBEC-induced deamination in a fraction of (+) strands. (**C**) Viruses with negative (-) ssRNA genomes. Several (+) ssRNA transcripts and/or precursors of (-) ssRNA genomes are generated by RdRp that are then used to generate multiple (-) ssRNA genomes. (-) ssRNA genomes of infecting particles as well (+) ssRNA precursors can serve as a substrate for APOBEC mutagenesis. The change (C to U or G to A) recovered by sequencing progeny genomes would be defined by the strand which is deaminated by APOBEC. Multiple C to U mutant molecules will arise from a single deamination in the infecting (-) ssRNA genome. Smaller number of G to A changes would result from each deamination event in a (+) strand precursor, but since there may be multiple precursor copies (shown in the multiple columns), a number of these changes may be comparable with C to U changes.

 doi:10.20944/preprints202104.0484.v1

3. Viral RNA genome rearrangements

Replication of RNA genomes in viruses is rarely perfect, often there is the introduction of errors [36] (discussed in section 4) as well as the incomplete replication of the genome. When a genome is incompletely replicated it is either left unrepaired and degraded, or it is repaired by recombination with another RNA molecule, (reviewed in [35]). Viral recombination occurs at high rates of 2-20% recombination events per 100 nucleotides [37-42], and the rate of recombination is dependent on the fidelity of the RNA dependent RNA polymerase [43-46]. Specifically, RNA dependent RNA polymerases with high fidelity are associated with a low recombination rate, while RNA dependent RNA polymerases with low fidelity are associated with high levels of recombination [43-46]. This is due to the major RNA viral recombination mechanism being initiated by faulty, or incomplete, viral replication.

Replicative RNA recombination begins when the viral RdRp stalls during replication of the viral genome followed by the dissociation of the newly synthesized RNA molecule from the template that subsequently binds to another template where it is used as a primer to begin synthesis (Figure 2A, reviewed in [35]). The synthesis continues through the end of the template to complete the RNA molecule [39,40,47-50]. After RdRp stalling and dissociation, should the RNA find its homologous sequence in another identical RNA genome (Figure 2Ai), the resulting genome will then be identical to the previous template [39,40,47-51]. However, if the molecule that is utilized as the template is a completely different RNA molecule, which is possible due to the high number of mRNAs in the cell along with the possibility of infection by other RNA viruses, it will result in a chimeric RNA molecule containing parts of both RNA sequences (Figure 2Aii). This creates a novel viral genome. If the recombination event includes new beneficial genes, there may be the increased fitness of the virus. Such an increase in fitness can be a major driver of evolution as well as may create new viral disease.

Non-replicative RNA recombination is a much rarer form of RNA recombination, occurring independent of RdRp, where two molecules are joined at their ends to create a chimeric molecule (Figure 2B) [52-62]. These events have been documented by modern sequencing approaches in viruses incapable of replication, but the mechanism for their formation is not yet known. Even so, a possible avenue for research may come from the many cell types that are able to recombine mRNA molecules through RNA splicing or RNA self-splicing to excise introns using small nuclear ribonucleoproteins (snRNPs) or through ribozymes [63-65]. Potentially, non-replicative recombination could use these mechanisms to join two different RNA molecules, instead of its more traditional function of excising introns. Regardless of the mechanism of recombination, what results is the formation of chimeric molecules that can become a novel viral RNA genome that may combine beneficial traits that helps with the virus's overall fitness.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 April 2021 doi:10.20944/preprints202104.0484.v1

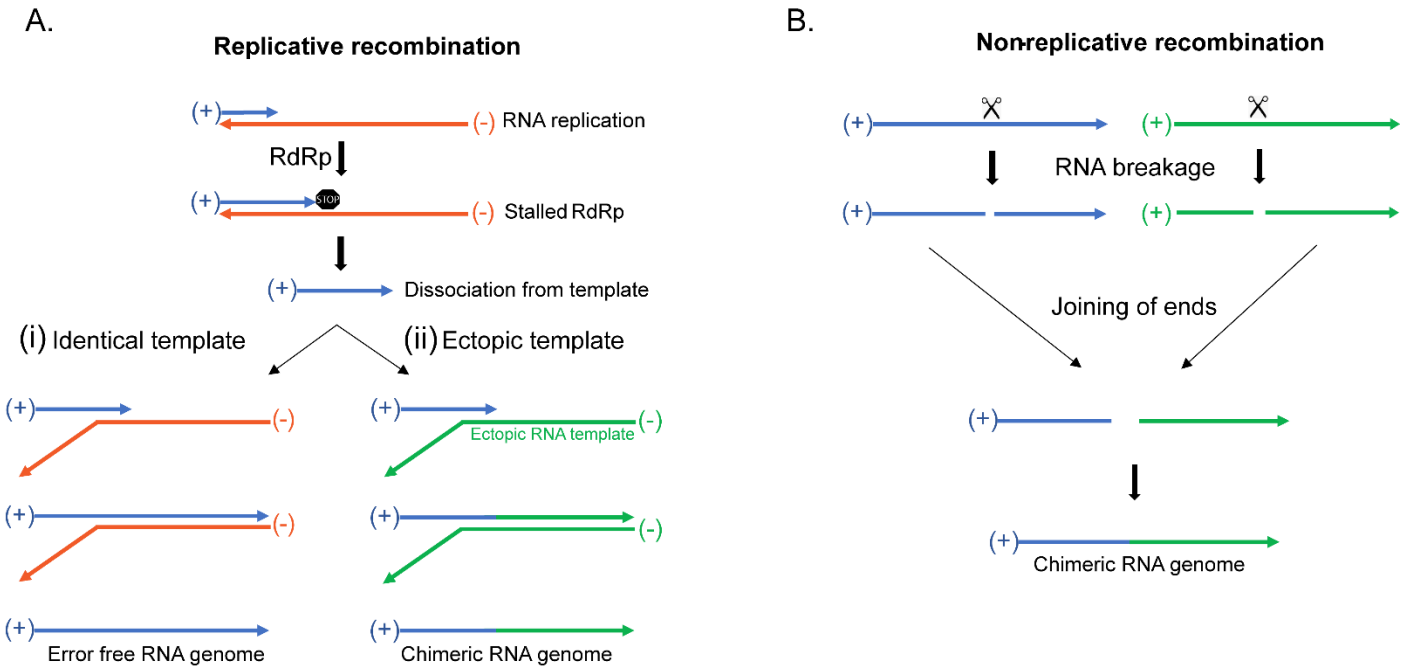

Figure 2. Viral RNA recombination.

(**A**) Replicative recombination begins after incomplete RNA replication resulting in the dissociation from the template and rebinding with another RNA molecule to complete replication. (i) RNA template rebinding at a homologous location in an identical RNA template results in error-free recombination. (ii) RNA template rebinding in an ectopic RNA molecule creates a chimeric molecule. (**B**) Non-Replicative RNA recombination occurs through a yet unknown mechanism, which can involve breakage and joining of two different RNA molecules to create a chimeric RNA molecule.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 April 2021 doi:10.20944/preprints202104.0484.v1

4. RNA replication errors

4.1. The RdRp's sequence variation effect on replication fidelity

RNA viral evolution has resulted in a diverse population of virus types that inevitably contain different combinations of genes within each virus. But there is a crucial factor for RNA virus replication present in all classes of viruses called RNA-dependent RNA polymerase (RdRp) [66,67]. RdRp is most closely related to eukaryotic reverse transcriptases [66], and functions in replicating viral genomes using another viral RNA genome as a template. Though RdRp is conserved throughout RNA viruses, the overall RdRp sequence is highly variable, with some sequence conservation as low as 10% [68,69]. However, within this variable sequence of the RdRp, there are seven domains that contain key conserved residues. These domains are oriented in the order from amino to carboxy terminus G, F, A, B, C, D, and E [68,69] (with some rare exceptions to this order [69]). Together these domains enable the synthesis of new RNA molecules by binding RNA, selecting and stabilizing ribonucleotides, and catalyzing the addition of the ribonucleotides, reviewed in [67]. Within the RdRp, the combination of conserved regions as well as variable regions has been exploited in metagenomics approaches to provide a clue to their evolutionary relationship [66,70]. Specifically, the finding that evolutionarily distant host organisms are infected with related RNA viruses indicate that these viruses did not evolve linearly, but instead are a result of horizontal transfer of sequences [66,70]. Further, these same approaches have informed multiple models of viral evolution from the RNA world to modern day viruses [70].

The variation of the RdRp domains not only allows for the use in evolutionary studies. Variation of the RdRp domains leads to different levels of RNA replication fidelity, which is already orders of magnitude less accurate as compared to DNA replication [71-73]. Within the RdRp domains that are directly involved in ribonucleotide selection or catalysis there are key conserved aspartate and lysine residues in the center of the domain that when disrupted greatly alters the RdRp activity [68,69]. However, the remaining sequence within the domains are not conserved to the same extent, but each specific sequence variation can modify RdRp function by any of the following RdRp functions, including RNA binding activity, selection and stabilization of ribonucleotides, and catalyzing the addition of the ribonucleotides. Together, these sequence variations observed between viruses results in the variable RdRp replication fidelity, reviewed in [67]. Beyond sequence variability, the fidelity of RdRp synthesis is also affected by environmental factors such as pH where a change in the pH from pH 6.5 to pH 8.0 can decrease the fidelity as much as nine times [74]. Further, the presence of nucleoside analogs [75] as well as the presence of divalent metals [47,51] can decrease the fidelity of the RdRp, which have been utilized as antiviral treatments [76-79].

4.2. RdRp's replication fidelity impacts viral evolution

It was argued that the low fidelity of the RdRp drives the evolution of RNA viruses [80,81]. This idea was supported in multiples studies where viruses (IAV, PV, FMDV, Chikungunya virus (CHIKV), and Human enterovirus 71 (EV71)) were exposed to nucleotide analogs (this increases the mutation rate often used as an antiviral strategy [76-79]) and after a few passages, a subpopulation emerged that became resistant through the acquisition of a mutant RdRp that has a higher replication fidelity [82-86]. Together, this suggests that a high mutation rate can mediate the formation of advantageous mutations that can drive evolution, but it also suggests that a higher replication fidelity can result in a more stable virus propagation. The latter notion is supported by viral strains that contained high fidelity RdRp variants continued stable propagation, however they ultimately became attenuated [87,88].

With such a high mutation rate observed in viruses, there is a selection for smaller genomes because a larger genome would have more opportunities for the acquisition of a deleterious mutation resulting in "error catastrophe" [89-91]. Consequently, there is a balance of mutagenesis to be high enough to allow for adaptation, but low enough to be able

 doi:10.20944/preprints202104.0484.v1

to maintain a complex genome and prevent error catastrophe. This is believed to be a selection factor causing a tendency to limit the size of the genome -- for most RNA viruses to be around 15kb in length [90,91]. Nevertheless, the Nidovirales family of viruses have RNA genomes upwards of 30kb (maximum of 41kb) [92,93] which is twice as large as a majority of viruses. A reason for the large viral genome size in Nidovirales remained unclear until the Gorbelyna group identified a sequence encoding a 3′ to 5′ exoribonuclease inside the SARS-CoV nsp14 subunit (also called ExoN). Further, the authors proposed that ExoN allowed for the increase in genome size by proofreading RdRp errors and thereby reducing a chance of error catastrophe [94,95]. Subsequently, the 3′ to 5′ exoribonuclease function of ExoN was found in vitro, and ExoN was also found to be essential for the viability of the alphacoronavirus HCoV-229E [96]. Similar experiments were conducted in ExoN-knockout mutants of two betacoronaviruses, MHV and SARS-CoV viruses [97,98], and revealed that the viruses were still viable, but to a much lower extent as compared to wild type. Also, the ExoN-knockout mutants were deemed to have a "mutator phenotype" as they had a 15- to 21-fold increase, respectively, in mutations, as compared to wild type ExoN strains, approximately reaching the mutation frequency of other "non-nidovirales" viruses [97,98]. Together, this indicated that ExoN may act by proofreading RdRp errors, which was later supported by the findings of that ExoN can excise mismatched nucleotides from a double-stranded RNA substrate [99,100]. In the same work, it was found that ExoN proofreading activity is enhanced 35-fold by the inclusion of nsp10, which suggested the formation of a heterodimer that forms to proofread mismatches [100]. Together this leads to a repair of mismatches incorporated by RdRp resulting in a higher fidelity of RNA synthesis.

Since then, more insights into ExoN have come from structural studies within SARS-CoV to reveal that ExoN (nsp14) physically interacts with more than just nsp10 to form a multimeric enzyme complex involved in replication of RNA. ExoN was described to have an in vitro association with the nsp7/nsp8/nsp12 tripartite complex [101]. nsp8 was proposed to act as a primase as it was shown to be able to synthesize 6-nucleotide long products in vitro [102] to prime synthesis by nsp12, the RdRp of SARS-CoV that lacks a synthesis priming loop, [69,103-106]. However, when nsp8 was studied in conjunction with the tripartate complex, the primase activity was not identified [101,107], but, instead, found a 3′-terminal adenylyl transferase activity that may add a 3′-poly(A) tail to transcripts. Therefore, nsp8 is important for RNA synthesis, but more work is required to understand its role in the context of the tripartite complex. The last subunit of the tripartite complex, nsp7, works in complex with nsp8 and has been proposed to function as a processivity factor [101,108] as well as a primase [102,109]. The exact function of Nsp7 needs to be investigated further. Altogether, though the exact function of the tripartite complex remains unclear, it is known that it aids ExoN in its function in the removal of mismatches and results in fewer mistakes in the newly synthesized viral RNA genome. However, if selection for a higher evolution rate occurs this would open the door to other modes of mutations in ExoN containing viruses. Possibly, these mutations could be introduced through the error-prone bypass of lesions and/or RNA editing.

5. **Lesion-induced mutagenesis in viral RNA genomes**

5.1. Environmental and endogenous RNA lesions and modifications

Viral genomic RNAs as well as cellular RNAs are the subject to environmental and endogenous lesions. These lesions can result in RNA breakage or can block RNA replication leading to repair through recombination, like one-ended breaks in DNA. Broken RNA genomes would be either lost or participate in recombination like events which can in turn produce rearranged genomes (see section 3 and Figure 2). Similar to DNA, RNA base lesions and modifications can be caused by a variety of endogenous and exogenous agents (Table 1 and references in footnotes) however, unlike for DNA, most base lesions in RNA cannot be repaired. The only known exception is for some alkylation products of cytosine

 doi:10.20944/preprints202104.0484.v1

and adenine bases which can be reversed to normal bases by a special class of oxidative demethylases – AlkB in bacteria or ALKBH family in humans [110-112].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 April 2021 doi:10.20944/preprints202104.0484.v1

Table 1. RNA base modifications

Canonical base	Modified base	Possibility of non-enzymatic generation	Enzymatic generation	Enzymatic reversal
Adenine	N1-methyladenosine (m1A)	SN2-alkylation	Not known	AlkB/ALKBH
Adenine	N3-methyladenosine (m3A)	SN2-alkylation	Not known	AlkB/ALKBH
Adenine	N6-methyladenosine (m6A)	Isomerization of m1A	Methyltransferases	AlkB/ALKBH
Adenine	Hypoxanthine base (Inosine ribonucleotide, I)	Low rate spontaneous deamination	Adenine Deaminases Acting on RNA (ADAR)	None
Uracil	Pseudouridine (Ψ)	Not known	Pseudouridine synthase	None
Guanine	N1-methylguanosine (m1G)	SN2-alkylation	Methyltransferases	AlkB/ALKBH
Guanine	N3-methylguanosine (m3G)	SN2-alkylation	Not known	Not known
Guanine	N7-methylguanosine (m7G)	SN2-alkylation	Methyltransferases	Not known
Guanine	O6-methylguanosine (O6mG)	SN1-alkylation	Not known	Not known
Cytosine	N1-methylcytosine (m1C)	SN2-alkylation	Not known	Not known
Cytosine	N3-methylcytosine (m3C)	SN2-alkylation	Not known	AlkB/ALKBH
Cytosine	N5-methylcytosine (m5C)	Not known	Methyltransferases	Not known
Cytosine	Uracil	Spontaneous deamination	APOBEC/AID	None

This Table contains a summary of information compiled from [110,111,113-118]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 April 2021 doi:10.20944/preprints202104.0484.v1

Several enzymatic modifications of RNA bases – pseudouridine (Ψ), N6-methyladenine (m6A), N5-methylcytosine (m5C) and Hypoxanthine (Inosine ribonucleotide (I)) were reported to have physiological functions in RNA viruses ([114] and references therein). These as well as several other base modifications are also present in cellular mRNAs and have also multiple functional consequences [117,119,120] and altogether are referred as the RNA-editome, or as epitranscriptome. For most of RNA base modifications, their full mutagenic potential is yet to be determined. It is even not clear, if they are present in the full-size replicating viral genome or only in non-replicating viral mRNAs. Only two kinds of enzymatic RNA edits – cytidine to uridine (C to U) by APOBEC cytidine deaminases and adenosine to inosine (A to I) by ADAR adenosine deaminases are known to be carried into copies of viral genomes resulting in C to U and A to G mutations, respectively. As discussed in the next section, these two groups of deaminases may have the greatest impact on mutation accumulation in several human RNA viruses.

5.2. Base substitution mutagenesis in RNA viruses

Base substitutions are an important source for viral evolution and population dynamics. They are also a common avenue for viruses to escape the host's adaptive immune system. Thus, it is important to identify mechanisms underlying the generation of base substitutions in viral populations. Usual approaches to understanding the mutagenic mechanisms underlying genome mutations come from a combination of knowledge accumulated in model studies as well as from agnostic documenting features of mutational spectrum that deviate from the spectrum expected, if mutagenesis would be completely random. Such "non-random" features of mutational spectra are also called mutational motifs (by analogy with musical motifs, which combine notes according to the rules of harmony) or mutational signatures (multiple set of features defining uniqueness of an object). This synthetic approach turned to be productive in revealing the mutagenic mechanisms in human cancers [121-123].

Besides significant mechanistic knowledge, the progress in deciphering cancer mutagenic mechanisms was made possible by accumulating many mutation catalogs – complete lists of de novo mutations in genomes of individual human tumors. Since tumor tissue is a mixture of a small number of clones, it is possible to identify mutations that had occurred after a tumor clone, or a small set of clones, have separated from the normal tissue. First, DNA sequence reads from tumor and from normal tissue are mapped against a reference genome. A vast majority of the differences with a reference would be common for normal tissue and for tumor DNA because they come from the common germline of the same individual. The differences that are present only in the tumor but are not detected in the DNA from normal tissue, are compiled to create the tumor's mutation catalog.

A similar strategy that was utilized with human cancers can now be applied to several RNA viruses, especially because of the accumulation of extensive sequencing data, most notably for SARS-CoV-2, which was a subject of a gigantic genome resequencing effort across the world [124] (see also URL https://www.gisaid.org/). However, since viral genomes are small, each genome would contain a small number, or even no, mutations so a mutation catalogue representative of a population (or of a species) could only be built from sequences of multiple genomes. Building such a catalog must be done under the assumption that each genome has been acted upon by similar mutational mechanisms, which may not be the case for all genomes. Beyond this assumption, other very important factors must be considered to ensure the proper development of a virus population mutational catalog. Firstly, an important point is that the mutations in the catalog must represent independent changes rather than a small number of events that are amplified through the development of a population, or through species evolution. A simplified example is shown in (Figure 3) where a viral population starting from a single genome accumulates mutations over nine rounds of copying. Many mutations in these genomes are identical, not because they occurred multiple times, but instead because they stem from a single genome that was then propagated. Therefore, if each of these mutations were treated as independent events, the mutations that occurred in an early generation would

 doi:10.20944/preprints202104.0484.v1

be counted more times than the actual number of independent mutation events that generated these mutations. This problem would not occur in a set of non-identical (non-duplicated) mutations, as each observed change/mutation would come from an independent mutation event, and would represent the spectrum of mutational processes (e.g. set of non-duplicated mutations in a genome at generation 9 on Figure 3). Consequently, to ensure an accurate mutational catalog, each mutation in a catalog must represent an independent change.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 April 2021 doi:10.20944/preprints202104.0484.v1

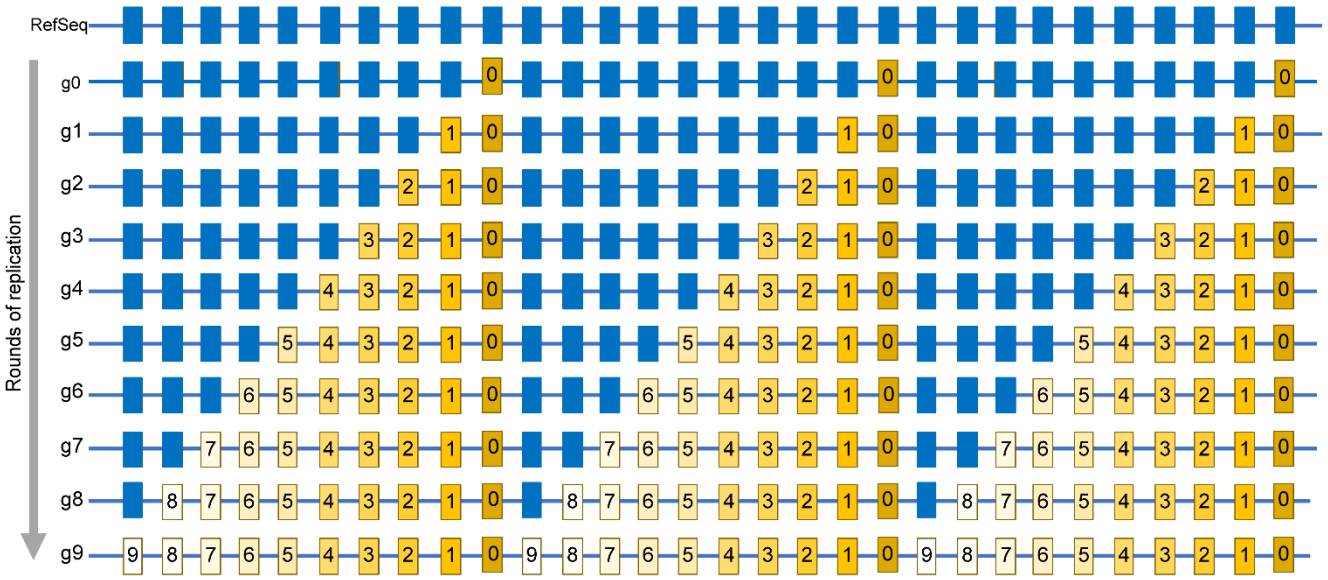

Figure 3. Simplified schematic of mutations accumulation in virus population.

Mutations are identified by comparing a sequence of a viral isolate with a reference sequence (RefSeq). Positions in which mutations are found in at least one isolate are shown by rectangles. Blue rectangles are positions same as in RefSeq. g0 – A genome of virus quasispecies starting a population that may already have some differences from RefSeq. g1 – g9 rounds of replication generating additional mutations, which are numbered same as the generation in which a mutation event had occurred. Mutations occurring in later generations would be present in smaller fractions (reflected by the decreasing yellow color density) within the population. The entire set of independent mutation events would be described by the list in which every mutation is represented only once, regardless of the number of genomes where it is found. In this population, such a list is represented by the g9 genome.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 April 2021 doi:10.20944/preprints202104.0484.v1

Another important factor for the development of a mutational catalog is the knowledge of the original genome sequence from which all other genomes stem. This is required to identify the mutant alleles in every sequenced viral genome as well as to identify the directionality of the mutation events. The latter is especially important for defining mutational events across long term population dynamics or during species evolution. In the cases of identifying mutational signatures in mammalian RNA viruses, a single reference sequence for the entire dataset is not available, so the reference roles are assigned to the sequences in the nodes of phylogenetic trees built for genomes of isolates from a population [125], for distant isolates of a single virus species [126-128] or for several related quasispecies [129]. Each node is taken as a surrogate of the reference sequence for the genomes in the same clade of a phylogenetic tree.

With the surrogate reference sequence established, it can be utilized to develop a mutational catalog. This approach allowed for the detection of C to U changes as a prominent or even the major component of mutagenesis in a wide range of mammalian RNA viruses [130]. Typically, the presence of base substitution type(s) that exceed expectation(s) for random mutagenesis can be usually explained by several reasons such as base-specific RdRp errors, mutagenic lesions, enzymatic lesion direct reversal as well as RNA editing. However, as of now, a single feature of base preference cannot be directly applied to mutation spectra interpretation as the base preferences for RdRp lesions are yet to be determined. While the base specificities for several RNA damaging agents, damage reversal enzymes and RNA editing enzymes, are well established they cannot serve as a single diagnostic indication of a mutagenesis source, because these specificities are often similar between agents (Table 1). Another factor to account for in the analysis of mutagenesis results is strand bias of a particular change. In viruses this bias will depend on a preference of a base modifying factor to single-stranded (ss) or to double-stranded (ds) RNA. It will be also affected by the kind of RNA forming genome of a virus: positive-strand ssRNA, negative-stranded ssRNA, or dsRNA (Figure 1A-C). For example, in positive-strand ssRNA viruses, changes stemming from ssRNA-specific agent modifications of the positive-strand will come into genomes of viral progeny. As for changes in viral progeny caused by a dsRNA specific agent, they will be mostly coming from modifications of the negative-strand (Figure 1A). The base changes shown on all panels of Figure 1 are the same as C to U changes expected from ssRNA-specific cytidine deaminases APOBEC and from guanine (G) to inosine (I) dsRNA-specific adenine deaminases ADAR (Figure 4). Interestingly, the spectra and strand preference of the two most prevailing kinds of changes in hypermutated isolates of human vaccine-derived rubella virus corresponded to the prevailing C to U, and U to C, changes in genomic strand of this positive-strand ssRNA virus shown on Figure 1A [128]. These hypermutated viruses (up to 300 base substitutions in a 9 kb genome) were extracted from granulomas of different children with primary immunodeficiency. Each independent virus isolate stemmed from the attenuated rubella vaccine virus; whose known original sequence was used as a reference to build a mutation catalog from six isolates that contained 993 mutations. C to U, or A to G, changes were the major mutations in the catalog. Such a pattern of mutations in the rubella vaccine virus mutation catalog matches the signature of APOBEC cytidine deaminases and supports the idea that APOBEC enzymes are the major mutator. Recently, It has been established that APOBEC3A (A3A) enzyme has a preference to the unpaired parts (loops) of folded RNA structures in mRNAs of human tumors [131]. Importantly, C to U changes in the positive-strand in hypermutated rubella genomes were the only type of base substitution that showed a statistically significant density increase in predicted RNA-loops over stems - sequences with a potential for self-pairing [132]. This lends support for APOBEC mutation being the source for C to U changes, however both strand bias and unpaired loop preference could be the feature of any agent causing chemical deamination of cytidines in RNA.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 April 2021 doi:10.20944/preprints202104.0484.v1

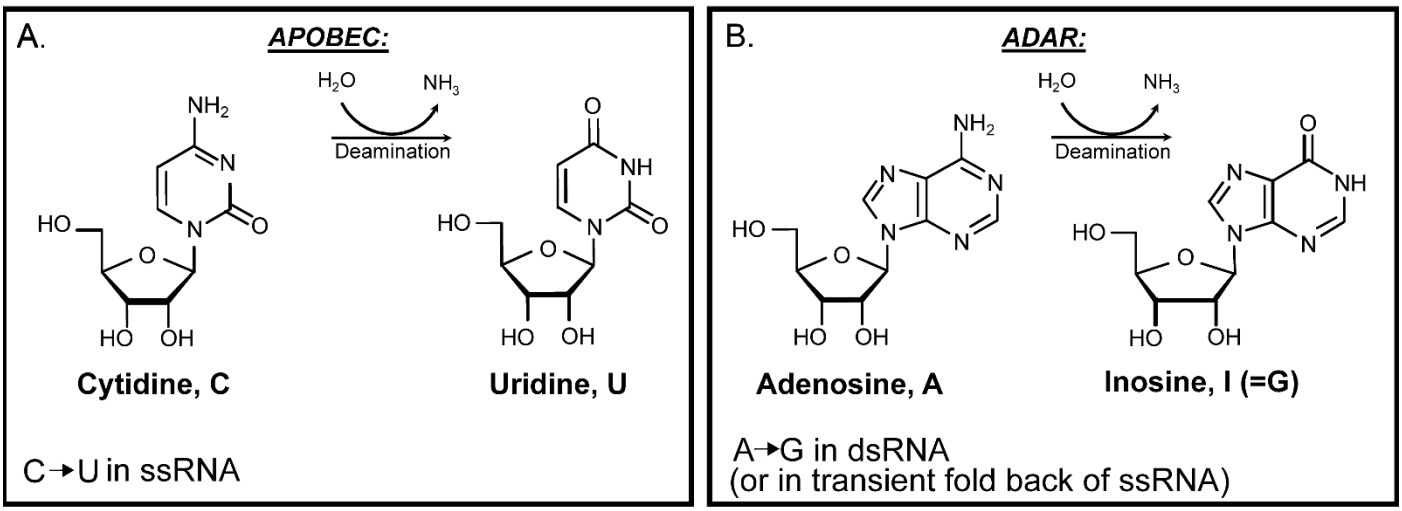

Figure 4. Enzymatic deamination of RNA nucleosides.

(**A**) APOBEC cytidine deaminase. Deamination of cytidine in ssRNA generates uridine resulting in C→U mutation in the RNA virus genome. (**B**) ADAR adenosine deaminase. Deamination of adenosine in dsRNA or in folded and paired ssRNA (forming dsRNA) generates inosine, which after rounds of copying with RdRp is fixed as A→G mutation.

 doi:10.20944/preprints202104.0484.v1

Deamination of cytosines not only occurs in RNA, cytidine deamination in DNA is one of the most frequent spontaneous changes and has a preference to ssDNA [115]. Chemicals, such as nitrites, can actively induce these cytosine deamination and mutation events in ssDNA [133], but APOBEC cytidine deaminases have also been a source for cytidine deamination. It is known that human APOBEC cytidine deaminases have complete preference to ssDNA and ssRNA over ds polynucleotides. All APOBECs show clear preference to immediate nucleotide context surrounding deaminated cytidines in DNA. APOBEC3G has a preference to cCn context (mutated nucleotide capitalized; n – any nucleotide), while APOBEC1 and all other members of APOBEC3 gene cluster prefer tCn deamination motif [134-137]. The preferred DNA deamination motif for APOBEC3A and APOBEC3B was even narrowed to tCa [138]. APOBEC1 was initially discovered as RNA editor and significant evidence had accumulated by now that members of APOBEC3 gene cluster can also edit mRNA [139], however unlike for DNA, detailed editing signatures in RNA are yet to be established. We therefore used APOBEC signature motifs established for DNA to evaluate the mutation spectra in a catalog compiled of hypermutated rubella genomes (Figure 5 shows example for uCa to uUa motif). This method was initially developed for evaluating APOBEC mutagenesis in human cancers [140], however it allows statistical estimate of over-representation with any oligonucleotide motif in mutation datasets [138,141,142]. A fraction of mutations in an oligonucleotide motif among mutations of a given nucleotide is compared with the presence of the same oligonucleotide in the genomic context surrounding mutated bases (see also Figure 5 and legend). We found a high level of enrichment with APOBEC motif uCn and even greater enrichment with more narrow uCa motif which is also the most preferred DNA editing motif for APOBEC3A and APOBEC3B [128]. Unlike for APOBEC editing, there is only a multi-motif ADAR editing web-based Inosine-Predict score tool, which takes into account immediate nucleotide context for every guanosine position as well as the potential to form a secondary structure. This tool was developed for ADAR editing in mRNAs [143]. There was slight, albeit statistically significant increase in Inosine-Predict score for adenine positions involved in A to G mutations (U to C in complementary strand) mutations as compared to non-mutated positions of As (or Us) [128].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 April 2021 doi:10.20944/preprints202104.0484.v1

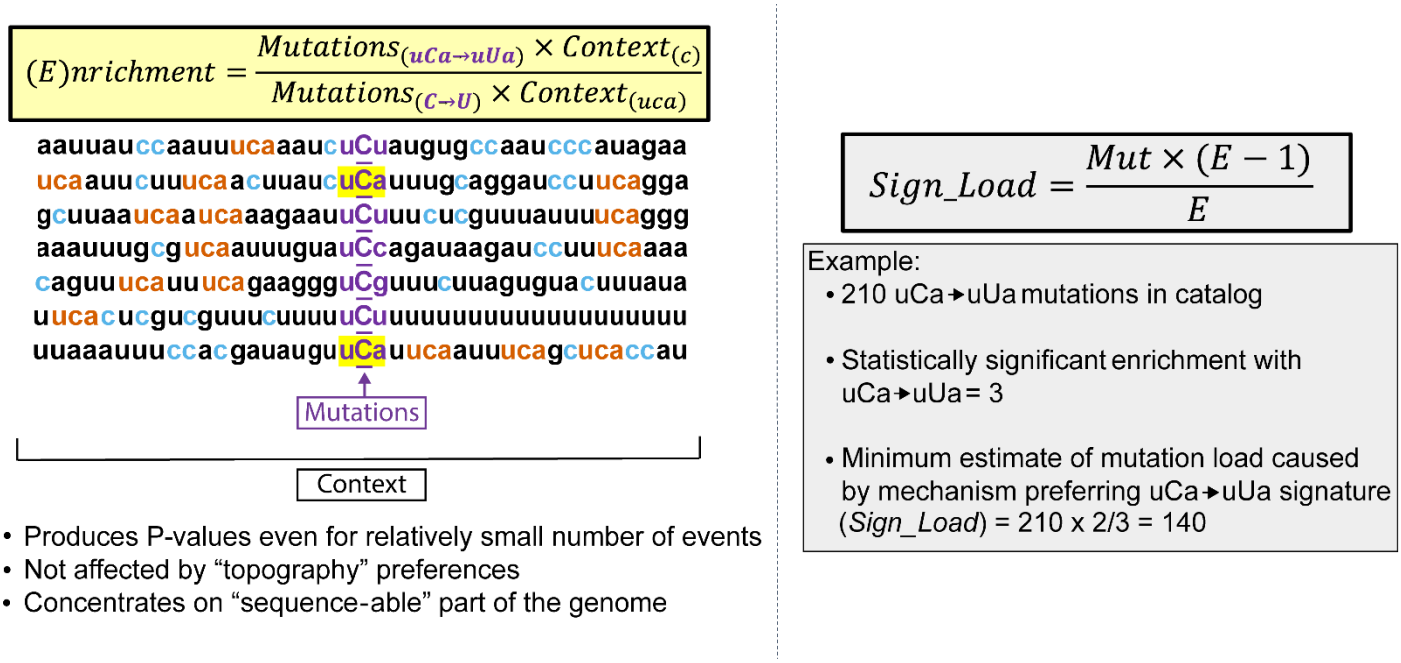

Figure 5. Trinucleotide motif-centered RNA mutational signature analysis.

Shown is the example of analysis for uCa→uUa signature motifs in ssRNA, therefore reverse complements are not included. Counted are all C→U mutations as well as all trinucleotide motif uCa→uUa mutations (5′ and 3′ flanking nucleotides shown in small letters; mutated C shown in capital letters). Also counted are all cytosines (c), represented in blue, and all motif-conforming trinucleotides (uca), represented in orange, in 41 nucleotide contexts centered around mutated cytosines. Enrichment (E) values show the fold-difference between actual fraction of uCa→uUa mutations among C→U mutations in all trinucleotide motifs and the fraction of motif conforming trinucleotides (uca) among all cytosines (c) in the immediate vicinity of mutated cytosines. Counts used for enrichment calculation can be also used for calculating p-values in order to identify trinucleotide mutational motifs with statistically significant enrichment. Statistically significant enrichment values can be used for minimum estimates of a signature-associated mutation load (*Sign Load*).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 April 2021 doi:10.20944/preprints202104.0484.v1

Altogether, APOBEC-like and ADAR-like changes represented 86% of 993 mutations in the catalog from six hypermutated genomes of vaccine derived rubella virus. We then applied similar, but yet extended analytical and statistical evaluation approaches to evaluate mutation load and spectrum accumulated from over 30,000 SARS-CoV-2 genome sequences accumulated during first several months of pandemic [132]. In this analysis, we compared spectrum and signatures with hypermutated isolates of rubella virus. The unique feature of this dataset is that the starting reference sequence is well defined [144], so each difference from the reference is a direct trace of mutation event. However, as in the example on Figure 3, differences from the reference sequence that are identical between sequenced isolates can represent a record of a single mutation event, which was then amplified in the population. In fact, some mutations were found in several thousands of isolates. We therefore used a set of non-duplicated mutations to represent the summary of mutational processes operating in pandemic SARS-CoV-2 population. This set would be a minimum estimate of the independent mutation events list, because repeated occurrence of the same mutation cannot be excluded, especially for mutations increasing speed of spreading within a host. To reduce the effects of functional selection we also analyzed subsets of non-functional (synonymous or non-coding) and functional (non-synonymous). These mutation sets were analyzed in parallel with the catalog of mutations from hypermutated rubella isolates. Analytical tools initially developed for a single APOBEC motif derived from prior experimental studies were extended to statistical evaluation of all 192 possible tri-nucleotide base substitution motifs.

We calculated one-sided Fisher's exact test P-value for each motif and then corrected P-values for multiple hypotheses testing and applied FDR<0.05 threshold. We found that mutational processes with the same signatures that were revealed in hypermutated rubella isolates also may operate in the SARS-CoV-2 pandemic population. The main similarities between SARS-CoV-2 and rubella were: (i) the presence of APOBEC-like signature uCn to uUn in positive strand; (ii) frequent presence of ADAR like A to G and U to C (shown as in positive strand, will correspond to A to G in the negative-strand); (iii) preference of loops vs stems for C to U mutations. Also specific to SARS-CoV-2 were the statistically significant enrichment with the two additional trinucleotide-centered signatures. Firstly, there was enrichment with mutations in cGn to cAn (reverse complement for nCg to nUg in the negative-strand) which could reflect increased frequency of cytosine deamination in CpG motifs and C to U changes in the RNA negative-strand of dsRNA intermediate producing multiple copies of the positive-strand with the complementary G to A change (see example of negative strand mutagenesis in Figure 1A). Since enrichment calculations accounted for the presence of a motif in the genomic background, the depletion of CpG motifs in the viral genome [125,145-148] cannot be due to a single explanation of such an enrichment. Increased frequency of C to U changes in DNA had been connected with frequent cytidine methylation in CpG [149] but can also occur in the absence of methylation in DNA [150,151]. We proposed that increased frequency of cytosine deamination in RNA CpGs could contribute into increased enrichment with this mutational motif [132]. Secondly, there was increased presence of G to U changes in the positive-strand, which could reflect C to A changes in negative-strand. These changes could be caused by increased formation of ROS-induced 8-oxoG within cells or during library preparation [152,153].

Recently Adebali and colleagues performed analysis of large number of SARS-CoV-2 sequenced genomes organized in a phylogenetic tree [125]. Unlike in [132], this study used the nodes of the tree as a reference sequence for each sequence within each node. Despite the two studies having used different approaches for creating representative datasets of independent mutation events in a large collection of SARS-CoV-2 genomes, the main categories of mutation preferences APOBEC-like, ADAR-like, CpG-like and apparent ROS induced mutations were similar. Overall similar conclusions about prevailing mutagenic sources and signatures were in works addressing intra-host variations of SARS-CoV-2 [154,155]. In summary, resequencing of RNA virus genomes suggested major mutational processes generating diversity that can lead to development of new virus

 doi:10.20944/preprints202104.0484.v1

forms. These studies also defined several questions and technical developments that should be addressed in near future.

6. Concluding remarks and future questions.

Emergence of new RNA viral quasispecies pathogenic for humans, especially the SARS-CoV-2 pandemic, triggered massive research efforts to all aspects of RNA virus mechanistic studies. Mechanisms underlying instability of RNA virus genomes are important for better prediction of their evolution, new pathogen emergence, and the development of antiviral drugs. Besides that, understanding biological and molecular mechanics that allows this group to flourish rather than be washed away with catastrophic error rates represents a fundamental question related to general mechanisms of evolution. Below are questions and technological applications that we anticipate being addressed soon.

6.1. Single-molecule sequencing applied to RNA virome in the environment and in single organisms

The sequence of a natural individual viral isolate is usually generated from a reference-based or de novo alignment of multiple small Illumina reads, thus it does not reflect the variations of individual viral RNAs but instead is an average of the total population [156]. However, the recent combination of deep Illumina sequencing, and advanced bioinformatics, allows intra-host variations to be addressed in genomes of a single viral species during an acute infection period [154,155,157,158], so some interhost variation can be revealed. Further, a combination of bioinformatics with metagenomics focusing on a small conserved element of RNA viruses, such as RdRp, made it possible to identify and interrogate the content of multispecies virome [66,159]. Even so, the combination of short reads with metagenomics does have its limitations as it relies on the building arbitrary contigs from short reads, so the entire genome cannot be assessed [160]. Nevertheless, deep sequencing with short reads allowed the identification of multiple new viral species [161], was successfully used to characterize the viromes in different organisms [162,163]; and characterize the viromes in different environmental samples [164]. To overcome these short-read sequencing issues, there are two technologies carrying promise for studies of viral genome instability by generating long reads from individual polynucleotides, Oxford Nanopore Technology and Pacific Biosciences [165]. Each of the two platforms is plagued by a rather high sequencing error rate, but even with the current level of accuracy Oxford Nanopore was used for characterizing viromes [166-168]. Any further increase in sequencing accuracy may cause a revolution with viral genome instability research. While the field awaits this increase of accuracy of the core technologies, there is a way to reduce false positive mutation calls by adding unique molecule identifier (UMI) barcodes added by either limited number of PCR cycles or by ligation to increase the accuracy in both platforms [169].

6.2. Impact of RdRp misincorporation and proofreading onto viral mutation rates

Low accuracy of RdRp, as compared with replicative DNA polymerases, led to the proposal that the major source of viral genome mutations is connected with replication errors [3,170]. Since many viruses have an exonuclease (ExoN or its homologs; see special section above) appearing to proofread RdRp misincorporation, it is important to collect more information about the impact of RdRp proofreading into prevention of hypermutation in RNA viruses. This would be approached by the modification of either, RdRp accuracy or ExoN capability by mutations and/or by endogenous or environmental factors. It is quite possible that the combination of such functional defects can lead to ultra-mutation phenotypes that would function similar to the synergistic hypermutation observed in cellular organisms when mismatch repair and DNA polymerase proofreading defects are combined [171-173]. Further, many antiviral drugs are chain-terminating NTP analogs designed to preferentially affect chain extension by RdRp [76]. However, this chain termination can be counteracted by ExoN [95,174], so search for inhibitors of this enzyme is important for practical applications

6.3. Are there RNA repair mechanisms besides AlkB direct reversal?

It is long known that RNA is more vulnerable for breakage as compared to DNA [175] and is at least just as susceptible as DNA to base lesions (Table 1 and references therein). However, unlike DNA, there is only one well established mechanism to repair RNA base lesions – direct reversal of alkylation. Currently there are no direct indications for the existence of other RNA repair mechanisms. Speculations can still be made based on structural similarity of RNA and DNA resulting in RNA being a substrate or a ligand for common DNA repair enzyme, e.g., RPA [176], but more research will be needed to reveal RNA repair mechanisms of they do indeed exist.

6.4. Impact of environmental RNA lesions onto viral genome instability

DNA base lesions are often an impediment to replicative DNA polymerases and require specialized trans-lesion synthesis (TLS) DNA polymerases to successfully accomplish genome duplication. TLS polymerases are often error-prone and results in mutagenesis, while the lack of TLS can lead to genome rearrangements or to replication failure [177,178]. The same cannot be said about RNA genomes as, per current knowledge, there have been no model studies addressing the impact of environmental RNA damage on structural or sequence integrity of RNA genomes. Further mechanistic studies to fill this gap of knowledge are important for understanding the dynamic world of RNA viruses.

6.5. Impact of endogenous RNA lesions onto viral genome instability

Recent studies summarized in section 5 indicated that adenosine deaminases ADAR and cytidine deaminases APOBEC are the prevailing sources of bases substitutions in several human RNA viruses including SARS-CoV-2. Both types of enzymes are the part of innate immunity, which raises a question about individual levels of RNA virus hypermutation. Interestingly, hypermutation of the vaccine-derived rubella virus was reported in subjects with primary immunodeficiency in adaptive immune system, which could be the reason of excessive activation of innate immunity and, consequently, excessive activation of APOBECs [128]. Another important question is about the level of endogenous hypermutation of RNA viruses in species that may serve reservoirs for the occurrence of new quasispecies. It is interesting that bats, which are a known coronavirus reservoir have multiple (4-7) APOBEC3 homologs while most of other mammalian orders have only one or two versions of APOBEC3 [179,180]. Therefore, studies into these organisms may reveal insights into the formation of novel viruses.

6.6. Experimental models to define signatures of environmental and endogenous mutagenesis in RNA

Defining diagnostic mutational signatures can develop into a multiprong scalable approach to understanding sources and mechanism of mutagenesis in RNA viruses. Mutational signatures turned to be a productive approach for another set of unstable genomes – human cancer. This could be a pure agnostic analysis of large datasets of genome instability catalogs [121,123,181], which can be also combined with prior mechanistic knowledge about different types of mutagenesis [122]. Another approach is to collect knowledge about mutational signatures in defined systems - mammalian [182,183] or microbial [138,141], and then utilize the information to build a specific statistical hypothesis for interrogating datasets of natural variants. High rates of mutation and relative ease of RNA virus genome sequencing can certainly make these approaches productive and scalable.

Author Contributions: Both listed authors (Z.W.K and D.A.G.), contributed to the visualization, writing, and editing of the manuscript.

Funding: This research was supported by the US National Institute of Health Intramural Research Program Project Z1AES103266 to D.A.G.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: We thank Drs. Hamed Bostan and Marcos Morgan for critical reading of the manuscript

Conflicts of Interests: Both listed Authors have declared no competing interests.

References

1 Dobzhansky, T. Nothing in Biology Makes Sense except in the Light of Evolution. The American Biology Teacher, (**1973**), 35, 125-12910.2307/4444260.

2 Dobzhansky, T. Genetics and the origin of species / Theodosius Dobzhansky. Third edition, revised. edn, (**1951**), (Columbia University Press).

3 Domingo, E. Molecular Basis of Genetic Variation of Viruses,Virus as Populations,Esteban Domingo,Academic PressBoston, (**2016**), 35-7110.1016/b978-0-12-800837-9.00002-2.

4 Domingo, E. & Perales, C. Viral quasispecies. PLoS Genet, (**2019**), 15, e100827110.1371/journal.pgen.1008271.

5 Domingo, E., Sabo, D., Taniguchi, T. & Weissmann, C. Nucleotide sequence heterogeneity of an RNA phage population. Cell, (**1978**), 13, 735-74410.1016/0092-8674(78)90223-4.

6 Domingo, E., Sheldon, J. & Perales, C. Viral quasispecies evolution. Microbiol Mol Biol Rev, (**2012**), 76, 159-21610.1128/mmbr.05023-11.

7 Eigen, M. & Schuster, P. The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. Naturwissenschaften, (**1977**), 64, 541-56510.1007/BF00450633.

8 Fornes, J., Tomas Lazaro, J., Alarcon, T., Elena, S. F. & Sardanyes, J. Viral replication modes in single-peak fitness landscapes: A dynamical systems analysis. J Theor Biol, (**2019**), 460, 170-18310.1016/j.jtbi.2018.10.007.

9 Schuster, P. Quasispecies on Fitness Landscapes. Current topics in microbiology and immunology, (**2016**), 392, 61-12010.1007/82_2015_469.

10 Swetina, J. & Schuster, P. Self-replication with errors. A model for polynucleotide replication. Biophys Chem, (**1982**), 16, 329-34510.1016/0301-4622(82)87037-3.

11 Zhou, L., Ding, D. & Szostak, J. W. The virtual circular genome model for primordial RNA replication. RNA, (**2021**), 27, 1-1110.1261/rna.077693.120.

12 Joyce, G. F. & Szostak, J. W. Protocells and RNA Self-Replication. Cold Spring Harb Perspect Biol, (**2018**), 1010.1101/cshperspect.a034801.

13 Szostak, J. W., Bartel, D. P. & Luisi, P. L. Synthesizing life. Nature, (**2001**), 409, 387-390Doi 10.1038/35053176.

14 Jheeta, S. The Routes of Emergence of Life from LUCA during the RNA and Viral World: A Conspectus. Life (Basel), (**2015**), 5, 1445-145310.3390/life5021445.

15 Robertson, M. P. & Joyce, G. F. The origins of the RNA world. Cold Spring Harb Perspect Biol, (**2012**), 410.1101/cshperspect.a003608.

16 Dworkin, J. P., Lazcano, A. & Miller, S. L. The roads to and from the RNA world. J Theor Biol, (**2003**), 222, 127-13410.1016/s0022-5193(03)00020-1.

17 Gilbert, W. Origin of Life - the Rna World. Nature, (**1986**), 319, 618-618DOI 10.1038/319618a0.

18 Sankaran, N. The RNA World at Thirty: A Look Back with its Author. Journal of molecular evolution, (**2016**), 83, 169-17510.1007/s00239-016-9767-3.

19 Orgel, L. E. The origin of life--a review of facts and speculations. Trends Biochem Sci, (**1998**), 23, 491-49510.1016/s0968-0004(98)01300-0.

20 Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. Critical reviews in biochemistry and molecular biology, (**2004**), 39, 99-12310.1080/10409230490460765.

21 Szostak, J. W. An optimal degree of physical and chemical heterogeneity for the origin of life? Philos Trans R Soc Lond B Biol Sci, (**2011**), 366, 2894-290110.1098/rstb.2011.0140.

22 Cojocaru, R. & Unrau, P. J. Processive RNA polymerization and promoter recognition in an RNA World. Science, (**2021**), 371, 1225-123210.1126/science.abd9191.

23 Moya, A., Holmes, E. C. & Gonzalez-Candelas, F. The population genetics and evolutionary epidemiology of RNA viruses. Nat Rev Microbiol, (**2004**), 2, 279-28810.1038/nrmicro863.

24 Gago, S., Elena, S. F., Flores, R. & Sanjuan, R. Extremely high mutation rate of a hammerhead viroid. Science, (**2009**), 323, 130810.1126/science.1169202.

25 Duffy, S., Shackelton, L. A. & Holmes, E. C. Rates of evolutionary change in viruses: patterns and determinants. Nature reviews. Genetics, (**2008**), 9, 267-27610.1038/nrg2323.

26 Abdel-Moneim, A. S. & Abdelwhab, E. M. Evidence for SARS-CoV-2 Infection of Animal Hosts. Pathogens, (**2020**), 910.3390/pathogens9070529.

27 Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. Nat Med, (**2020**), 26, 450-45210.1038/s41591-020-0820-9.

28 Wu, T. & Orgel, L. E. Nonenzymatic template-directed synthesis on oligodeoxycytidylate sequences in hairpin oligonucleotides. J Am Chem Soc, (**1992**), 114, 317-32210.1021/ja00027a040.

29 Flores, R., Gago-Zachert, S., Serra, P., Sanjuan, R. & Elena, S. F. Viroids: survivors from the RNA world? Annu Rev Microbiol, (**2014**), 68, 395-41410.1146/annurev-micro-091313-103416.

30 Flores, R. & Owens, R. A. Viroids (Pospiviroidae and Avsunviroidae),Reference Module in Life Sciences,Elsevier (**2020**) 10.1016/b978-0-12-809633-8.21257-0.

31 Flores, R., Serra, P., Minoia, S., Di Serio, F. & Navarro, B. Viroids: from genotype to phenotype just relying on RNA sequence and structural motifs. Front Microbiol, (**2012**), 3, 21710.3389/fmicb.2012.00217.

32 Koonin, E. V., Senkevich, T. G. & Dolja, V. V. The ancient Virus World and evolution of cells. Biology direct, (**2006**), 1, 2910.1186/1745-6150-1-29.

33 Baltimore, D. Expression of animal virus genomes. Bacteriol Rev, (**1971**), 35, 235-241.

34 Ortin, J. & Parra, F. Structure and function of RNA replication. Annual Review of Microbiology, (**2006**), 60, 305-32610.1146/annurev.micro.60.080805.142248.

35 Agol, V. I. & Gmyl, A. P. Emergency Services of Viral RNAs: Repair and Remodeling. Microbiol Mol Biol Rev, (**2018**), 8210.1128/MMBR.00067-17.

36 Domingo, E. Molecular basis of genetic variation of viruses: error-prone replication. Virus as Populations: Composition, Complexity, Quasispecies, Dynamics, and Biological Implications, 2nd Edition, (**2020**), 35-7110.1016/B978-0-12-816331-3.00002-7.

37 Lai, M. M. C. Genetic-Recombination in Rna Viruses. Current topics in microbiology and immunology, (**1992**), 176, 21-32.

38 Levy, D. N., Aldrovandi, G. M., Kutsch, O. & Shaw, G. M. Dynamics of HIV-1 recombination in its natural target cells. Proceedings of the National Academy of Sciences of the United States of America, (**2004**), 101, 4204-420910.1073/pnas.0306764101.

39 Nagy, P. D. & Simon, A. E. New insights into the mechanisms of RNA recombination. Virology, (**1997**), 235, 1-9DOI 10.1006/viro.1997.8681.

40 Sztuba-Solinska, J., Urbanowicz, A., Figlerowicz, M. & Bujarski, J. J. RNA-RNA Recombination in Plant Virus Replication and Evolution. Annu Rev Phytopathol, (**2011**), 49, 415-44310.1146/annurev-phyto-072910-095351.

41 Urbanowicz, A. et al. Homologous crossovers among molecules of brome mosaic bromovirus RNA1 or RNA2 segments in vivo. J Virol, (**2005**), 79, 5732-574210.1128/Jvi.79.9.5732-5742.2005.

42 Liao, C. L. & Lai, M. M. C. Rna Recombination in a Coronavirus - Recombination between Viral Genomic Rna and Transfected Rna Fragments. J Virol, (**1992**), 66, 6117-6124Doi 10.1128/Jvi.66.10.6117-6124.1992.

43 Kempf, B. J., Peersen, O. B. & Barton, D. J. Poliovirus Polymerase Leu420 Facilitates RNA Recombination and Ribavirin Resistance. J Virol, (**2016**), 90, 8410-842110.1128/Jvi.00078-16.

44 Peersen, O. B. Picornaviral polymerase structure, function, and fidelity modulation. Virus Research, (**2017**), 234, 4-2010.1016/j.virusres.2017.01.026.

45 Woodman, A., Arnold, J. J., Cameron, C. E. & Evans, D. J. Biochemical and genetic analysis of the role of the viral polymerase in enterovirus recombination. Nucleic acids research, (**2016**), 44, 6883-689510.1093/nar/gkw567.

46 Xiao, Y. H. et al. RNA Recombination Enhances Adaptability and Is Required for Virus Spread and Virulence. Cell Host Microbe, (2016), 19, 493-50310.1016/j.chom.2016.03.009.

47 Arnold, J. J., Ghosh, S. K. & Cameron, C. E. Poliovirus RNA-dependent RNA polymerase (3D(pol)). Divalent cation modulation of primer, template, and nucleotide selection. J Biol Chem, (**1999**), 274, 37060-3706910.1074/jbc.274.52.37060.

48 Kirkegaard, K. & Baltimore, D. The mechanism of RNA recombination in poliovirus. Cell, (**1986**), 47, 433-44310.1016/0092-8674(86)90600-8.

49 Romanova, L. I. et al. The primary structure of crossover regions of intertypic poliovirus recombinants: a model of recombination between RNA genomes. Virology, (**1986**), 155, 202-21310.1016/0042-6822(86)90180-7.

50 Simon-Loriere, E. & Holmes, E. C. Why do RNA viruses recombine? Nat Rev Microbiol, (**2011**), 9, 617-62610.1038/nrmicro2614.

51 Arnold, J. J., Gohara, D. W. & Cameron, C. E. Poliovirus RNA-dependent RNA polymerase (3Dpol): pre-steady-state kinetic analysis of ribonucleotide incorporation in the presence of Mn2+. Biochemistry, (**2004**), 43, 5138-514810.1021/bi035213q.

52 Adams, S. D., Tzeng, W. P., Chen, M. H. & Frey, T. K. Analysis of intermolecular RNA-RNA recombination by rubella virus. Virology, (**2003**), 309, 258-27110.1016/s0042-6822(03)00064-3.

53 Austermann-Busch, S. & Becher, P. RNA structural elements determine frequency and sites of nonhomologous recombination in an animal plus-strand RNA virus. J Virol, (**2012**), 86, 7393-740210.1128/JVI.00864-12.

54 Gallei, A., Pankraz, A., Thiel, H. J. & Becher, P. RNA recombination in vivo in the absence of viral replication. J Virol, (**2004**), 78, 6271-628110.1128/JVI.78.12.6271-6281.2004.

55 Gmyl, A. P. & Agol, V. I. Diverse Mechanisms of RNA Recombination. Mol Biol, (**2005**), 39, 529-54210.1007/s11008-005-0069-x.

56 Gmyl, A. P. et al. Nonreplicative RNA recombination in poliovirus. J Virol, (**1999**), 73, 8958-896510.1128/JVI.73.11.8958-8965.1999.

57 Gmyl, A. P., Korshenko, S. A., Belousov, E. V., Khitrina, E. V. & Agol, V. I. Nonreplicative homologous RNA recombination: promiscuous joining of RNA pieces? RNA, (**2003**), 9, 1221-123110.1261/rna.5111803.

58 Holmblat, B. et al. Nonhomologous recombination between defective poliovirus and coxsackievirus genomes suggests a new model of genetic plasticity for picornaviruses. mBio, (**2014**), 5, e01119-0111410.1128/mBio.01119-14.

59 Kleine Buning, M. et al. Nonreplicative RNA Recombination of an Animal Plus-Strand RNA Virus in the Absence of Efficient Translation of Viral Proteins. Genome Biol Evol, (**2017**), 9, 817-82910.1093/gbe/evx046.

60 Raju, R., Subramaniam, S. V. & Hajjou, M. Genesis of Sindbis virus by in vivo recombination of nonreplicative RNA precursors. J Virol, (**1995**), 69, 7391-740110.1128/JVI.69.12.7391-7401.1995.

61 Scheel, T. K. et al. Productive homologous and non-homologous recombination of hepatitis C virus in cell culture. PLoS pathogens, (**2013**), 9, e100322810.1371/journal.ppat.1003228.

62 Schibler, M., Piuz, I., Hao, W. & Tapparel, C. Chimeric rhinoviruses obtained via genetic engineering or artificially induced recombination are viable only if the polyprotein coding sequence derives from the same species. J Virol, (**2015**), 89, 4470-448010.1128/JVI.03668-14.

63 Wilkinson, M. E., Charenton, C. & Nagai, K. RNA Splicing by the Spliceosome. Annu Rev Biochem, (**2020**), 89, 359-38810.1146/annurev-biochem-091719-064225.

64 Zhang, L., Vielle, A., Espinosa, S. & Zhao, R. RNAs in the spliceosome: Insight from cryoEM structures. Wiley Interdiscip Rev RNA, (**2019**), 10, e152310.1002/wrna.1523.

65 Cech, T. R. Self-splicing RNA: implications for evolution. Int Rev Cytol, (**1985**), 93, 3-2210.1016/s0074-7696(08)61370-4.

66 Dolja, V. V. & Koonin, E. V. Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. Virus Res, (**2018**), 244, 36-5210.1016/j.virusres.2017.10.020.

67 te Velthuis, A. J. W. Common and unique features of viral RNA-dependent polymerases. Cell Mol Life Sci, (**2014**), 71, 4403-442010.1007/s00018-014-1695-z.

68 Bruenn, J. A. Relationships among the Positive Strand and Double-Strand Rna Viruses as Viewed through Their Rna-Dependent Rna-Polymerases. Nucleic acids research, (**1991**), 19, 217-226DOI 10.1093/nar/19.2.217.

69 Gorbalenya, A. E. et al. The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage. J Mol Biol, (**2002**), 324, 47-6210.1016/S0022-2836(02)01033-1.

70 Krupovic, M., Dolja, V. V. & Koonin, E. V. Origin of viruses: primordial replicators recruiting capsids from hosts. Nat Rev Microbiol, (**2019**), 17, 449-45810.1038/s41579-019-0205-6.

71 Domingo, E. & Holland, J. J. RNA virus mutations and fitness for survival. Annual Review of Microbiology, (**1997**), 51, 151-178DOI 10.1146/annurev.micro.51.1.151.

72 Drake, J. W. Rates of Spontaneous Mutation among Rna Viruses. Proceedings of the National Academy of Sciences of the United States of America, (**1993**), 90, 4171-4175DOI 10.1073/pnas.90.9.4171.

73 Perrino, F. W., Preston, B. D., Sandell, L. L. & Loeb, L. A. Extension of mismatched 3' termini of DNA is a major determinant of the infidelity of human immunodeficiency virus type 1 reverse transcriptase. Proceedings of the National Academy of Sciences of the United States of America, (**1989**), 86, 8343-834710.1073/pnas.86.21.8343.

74 Eckert, K. A. & Kunkel, T. A. Fidelity of DNA synthesis catalyzed by human DNA polymerase alpha and HIV-1 reverse transcriptase: effect of reaction pH. Nucleic Acids Res, (**1993**), 21, 5212-522010.1093/nar/21.22.5212.

75 Crotty, S. et al. The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. Nat Med, (**2000**), 6, 1375-137910.1038/82191.

76 Seley-Radtke, K. L. & Yates, M. K. The evolution of nucleoside analogue antivirals: A review for chemists and non-chemists. Part 1: Early structural modifications to the nucleoside scaffold. Antiviral Res, (**2018**), 154, 66-8610.1016/j.antiviral.2018.04.004.

77 De Clercq, E. Antivirals and antiviral strategies. Nat Rev Microbiol, (**2004**), 2, 704-72010.1038/nrmicro975.

78 De Clercq, E. & Li, G. Approved Antiviral Drugs over the Past 50 Years. Clin Microbiol Rev, (**2016**), 29, 695-74710.1128/CMR.00102-15.

79 Jordheim, L. P., Durantel, D., Zoulim, F. & Dumontet, C. Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases. Nat Rev Drug Discov, (**2013**), 12, 447-46410.1038/nrd4010.

80 Holland, J. et al. Rapid evolution of RNA genomes. Science, (**1982**), 215, 1577-158510.1126/science.7041255.

81 Novella, I. S. et al. Exponential increases of RNA virus fitness during large population transmissions. Proceedings of the National Academy of Sciences of the United States of America, (**1995**), 92, 5841-584410.1073/pnas.92.13.5841.

82 Pfeiffer, J. K. & Kirkegaard, K. A single mutation in poliovirus RNA-dependent RNA polymerase confers resistance to mutagenic nucleotide analogs via increased fidelity. Proceedings of the National Academy of Sciences of the United States of America, (**2003**), 100, 7289-729410.1073/pnas.1232294100.

83 Sadeghipour, S., Bek, E. J. & McMinn, P. C. Ribavirin-resistant mutants of human enterovirus 71 express a high replication fidelity phenotype during growth in cell culture. J Virol, (**2013**), 87, 1759-176910.1128/JVI.02139-12.

84 Coffey, L. L., Beeharry, Y., Borderia, A. V., Blanc, H. & Vignuzzi, M. Arbovirus high fidelity variant loses fitness in mosquitoes and mice. Proceedings of the National Academy of Sciences of the United States of America, (**2011**), 108, 16038-1604310.1073/pnas.1111650108.

85 Sierra, M. et al. Foot-and-mouth disease virus mutant with decreased sensitivity to ribavirin: implications for error catastrophe. J Virol, (**2007**), 81, 2012-202410.1128/JVI.01606-06.

86 Binh, N. T., Wakai, C., Kawaguchi, A. & Nagata, K. Involvement of the N-terminal portion of influenza virus RNA polymerase subunit PB1 in nucleotide recognition. Biochem Biophys Res Commun, (**2014**), 443, 975-97910.1016/j.bbrc.2013.12.071.

87 Liu, X. et al. Vaccine-derived mutation in motif D of poliovirus RNA-dependent RNA polymerase lowers nucleotide incorporation fidelity. J Biol Chem, (**2013**), 288, 32753-3276510.1074/jbc.M113.484428.

88 Vignuzzi, M., Wendt, E. & Andino, R. Engineering attenuated virus vaccines by controlling replication fidelity. Nat Med, (**2008**), 14, 154-16110.1038/nm1726.

89 Steinhauer, D. A., Domingo, E. & Holland, J. J. Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. Gene, (**1992**), 122, 281-28810.1016/0378-1119(92)90216-c.

90 Drake, J. W. & Holland, J. J. Mutation rates among RNA viruses. Proceedings of the National Academy of Sciences of the United States of America, (**1999**), 96, 13910-13913DOI 10.1073/pnas.96.24.13910.

91 Eigen, M. Error catastrophe and antiviral strategy. Proceedings of the National Academy of Sciences of the United States of America, (**2002**), 99, 13374-1337610.1073/pnas.212514799.

92 Bukhari, K. et al. Description and initial characterization of metatranscriptomic nidovirus-like genomes from the proposed new family Abyssoviridae, and from a sister group to the Coronavirinae, the proposed genus Alphaletovirus. Virology, (**2018**), 524, 160-17110.1016/j.virol.2018.08.010.

93 Saberi, A., Gulyaeva, A. A., Brubacher, J. L., Newmark, P. A. & Gorbalenya, A. E. A planarian nidovirus expands the limits of RNA genome size. PLoS pathogens, (**2018**), 14, e100731410.1371/journal.ppat.1007314.

94 Snijder, E. J. et al. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J Mol Biol, (**2003**), 331, 991-100410.1016/s0022-2836(03)00865-9.

95 Ogando, N. S. et al. The Curious Case of the Nidovirus Exoribonuclease: Its Role in RNA Synthesis and Replication Fidelity. Front Microbiol, (**2019**), 10, 181310.3389/fmicb.2019.01813.

96 Minskaia, E. et al. Discovery of an RNA virus 3 '-> 5 ' exoribonuclease that is critically involved in coronavirus RNA synthesis. Proceedings of the National Academy of Sciences of the United States of America, (**2006**), 103, 5108-511310.1073/pnas.0508200103.

97 Eckerle, L. D. et al. Infidelity of SARS-CoV Nsp14-Exonuclease Mutant Virus Replication Is Revealed by Complete Genome Sequencing. PLoS pathogens, (**2010**), 6ARTN e1000896 10.1371/journal.ppat.1000896.

98 Eckerle, L. D., Lu, X., Sperry, S. M., Choi, L. & Denison, M. R. High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. J Virol, (**2007**), 81, 12135-1214410.1128/Jvi.01296-07.

99 Bouvet, M. et al. In vitro reconstitution of SARS-coronavirus mRNA cap methylation. PLoS pathogens, (**2010**), 6, e100086310.1371/journal.ppat.1000863.

100 Bouvet, M. et al. RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. Proceedings of the National Academy of Sciences of the United States of America, (**2012**), 109, 9372-937710.1073/pnas.1201130109.

101 Subissi, L. et al. SARS-CoV ORF1b-encoded nonstructural proteins 12-16: replicative enzymes as antiviral targets. Antiviral Res, (**2014**), 101, 122-13010.1016/j.antiviral.2013.11.006.

102 Imbert, I. et al. A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. The EMBO journal, (**2006**), 25, 4933-494210.1038/sj.emboj.7601368.

103 Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. Nucleic Acids Res, (**1989**), 17, 4847-486110.1093/nar/17.12.4847.

104 Sevajol, M., Subissi, L., Decroly, E., Canard, B. & Imbert, I. Insights into RNA synthesis, capping, and proofreading mechanisms of SARS-coronavirus. Virus Research, (**2014**), 194, 90-9910.1016/j.virusres.2014.10.008.

105 Kirchdoerfer, R. N. & Ward, A. B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. Nat Commun, (**2019**), 10, 234210.1038/s41467-019-10280-3.

106 Posthuma, C. C., Te Velthuis, A. J. W. & Snijder, E. J. Nidovirus RNA polymerases: Complex enzymes handling exceptional RNA genomes. Virus Res, (**2017**), 234, 58-7310.1016/j.virusres.2017.01.023.

107 Tvarogova, J. et al. Identification and Characterization of a Human Coronavirus 229E Nonstructural Protein 8-Associated RNA 3 '-Terminal Adenylyltransferase Activity. J Virol, (**2019**), 93ARTN e00291-19
10.1128/JVI.00291-19.

108 Zhai, Y. et al. Insights into SARS-CoV transcription and replication from the structure of the nsp7-nsp8 hexadecamer. Nat Struct Mol Biol, (**2005**), 12, 980-98610.1038/nsmb999.

109 Velthuis, A. J. W. T., van den Worm, S. H. E. & Snijder, E. J. The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. Nucleic acids research, (**2012**), 40, 1737-174710.1093/nar/gkr893.

110 Drablos, F. et al. Alkylation damage in DNA and RNA--repair mechanisms and medical significance. DNA Repair (Amst), (**2004**), 3, 1389-140710.1016/j.dnarep.2004.05.004.

111 Feyzi, E. et al. RNA base damage and repair. Current pharmaceutical biotechnology, (**2007**), 8, 326-33110.2174/138920107783018363.

112 Thapar, R. et al. RNA Modifications: Reversal Mechanisms and Cancer. Biochemistry, (**2019**), 58, 312-32910.1021/acs.biochem.8b00949.

113 Pogolotti, A. L., Jr., Ono, A., Subramaniam, R. & Santi, D. V. On the mechanism of DNA-adenine methylase. J Biol Chem, (**1988**), 263, 7461-7464https://doi.org/10.1016/S0021-9258(18)68520-5.

114 Potuznik, J. F. & Cahova, H. It's the Little Things (in Viral RNA). mBio, (**2020**), 1110.1128/mBio.02131-20.

115 Lindahl, T. Instability and decay of the primary structure of DNA. Nature, (**1993**), 362, 709-71510.1038/362709a0.

116 Ge, J. & Yu, Y. T. RNA pseudouridylation: new insights into an old modification. Trends Biochem Sci, (**2013**), 38, 210-21810.1016/j.tibs.2013.01.002.

117 Schaefer, M. R. The Regulation of RNA Modification Systems: The Next Frontier in Epitranscriptomics?, (**2021**), 12, 345.

118 Beranek, D. T. Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. Mutation research, (**1990**), 231, 11-30.

119 Frye, M., Jaffrey, S. R., Pan, T., Rechavi, G. & Suzuki, T. RNA modifications: what have we learned and where are we headed? Nature reviews. Genetics, (**2016**), 17, 365-37210.1038/nrg.2016.47.

120 Ontiveros, R. J., Stoute, J. & Liu, K. F. The chemical diversity of RNA modifications. The Biochemical journal, (**2019**), 476, 1227-124510.1042/BCJ20180445.

121 Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. Nature, (**2013**), 500, 415-42110.1038/nature12477.

122 Roberts, S. A. & Gordenin, D. A. Hypermutation in human cancer genomes: footprints and mechanisms. Nature reviews. Cancer, (**2014**), 14, 786-80010.1038/nrc3816.

123 Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. Nature, (**2020**), 578, 94-10110.1038/s41586-020-1943-3.

124 Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin, (**2017**), 2210.2807/1560-7917.ES.2017.22.13.30494.

125 Azgari, C., Kilinc, Z., Turhan, B., Circi, D. & Adebali, O. The Mutation Profile of SARS-CoV-2 Is Primarily Shaped by the Host Antiviral Defense. (**2021**), 13, 394.

126 Khrustalev, V. V. & Barkovsky, E. V. Unusual nucleotide content of Rubella virus genome as a consequence of biased RNA-editing: comparison with Alphaviruses. International journal of bioinformatics research and applications, (**2011**), 7, 82-10010.1504/IJBRA.2011.039171.

127 Khrustalev, V. V., Khrustaleva, T. A., Sharma, N. & Giri, R. Mutational Pressure in Zika Virus: Local ADAR-Editing Areas Associated with Pauses in Translation and Replication. Frontiers in cellular and infection microbiology, (**2017**), 7, 4410.3389/fcimb.2017.00044.

128 Perelygina, L. et al. Infectious vaccine-derived rubella viruses emerge, persist, and evolve in cutaneous granulomas of children with primary immunodeficiencies. PLoS pathogens, (**2019**), 15, e100808010.1371/journal.ppat.1008080.

129 Simmonds, P. Rampant C-->U Hypermutation in the Genomes of SARS-CoV-2 and Other Coronaviruses: Causes and Consequences for Their Short- and Long-Term Evolutionary Trajectories. mSphere, (**2020**), 510.1128/mSphere.00408-20.

130 Simmonds, P. & Ansari, M. A. Mutation bias implicates RNA editing in a wide range of mammalian RNA viruses. (**2021**), 2021.2002.2009.43039510.1101/2021.02.09.430395 %J bioRxiv.

131 Jalili, P. et al. Quantification of ongoing APOBEC3A activity in tumor cells by monitoring RNA editing at hotspots. Nat Commun, (**2020**), 11, 297110.1038/s41467-020-16802-8.

132 Klimczak, L. J., Randall, T. A., Saini, N., Li, J. L. & Gordenin, D. A. Similarity between mutation spectra in hypermutated genomes of rubella virus and in SARS-CoV-2 genomes accumulated during the COVID-19 pandemic. PLoS One, (**2020**), 15, e023768910.1371/journal.pone.0237689.

133 Chan, K. et al. Base damage within single-strand DNA underlies in vivo hypermutability induced by a ubiquitous environmental agent. PLoS Genet, (**2012**), 8, e100314910.1371/journal.pgen.1003149.

134 Conticello, S. G. Creative deaminases, self-inflicted damage, and genome evolution. Annals of the New York Academy of Sciences, (**2012**), 1267, 79-8510.1111/j.1749-6632.2012.06614.x.

135 Green, A. M. & Weitzman, M. D. The spectrum of APOBEC3 activity: From anti-viral agents to anti-cancer opportunities. DNA Repair (Amst), (**2019**), 83, 10270010.1016/j.dnarep.2019.102700.

136 Harris, R. S. & Dudley, J. P. APOBECs and virus restriction. Virology, (**2015**), 479-480, 131-14510.1016/j.virol.2015.03.012.

137 Refsland, E. W. & Harris, R. S. The APOBEC3 family of retroelement restriction factors. Current topics in microbiology and immunology, (**2013**), 371, 1-2710.1007/978-3-642-37765-5_1.

138 Chan, K. et al. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. Nat Genet, (**2015**), 47, 1067-107210.1038/ng.3378.

139 Lerner, T., Papavasiliou, F. N. & Pecori, R. RNA Editors, Cofactors, and mRNA Targets: An Overview of the C-to-U RNA Editing Machinery and Its Implication in Human Disease. Genes, (**2018**), 1010.3390/genes10010013.

140 Roberts, S. A. et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet, (**2013**), 45, 970-97610.1038/ng.2702.

141 Saini, N. et al. Mutation signatures specific to DNA alkylating agents in yeast and cancers. Nucleic Acids Res, (**2020**), 48, 3692-370710.1093/nar/gkaa150.

142 Saini, N. et al. UV-exposure, endogenous DNA damage, and DNA replication errors shape the spectra of genome changes in human skin. PLoS Genet, (**2021**), 17, e100930210.1371/journal.pgen.1009302.

143 Eggington, J. M., Greene, T. & Bass, B. L. Predicting sites of ADAR editing in double-stranded RNA. Nat Commun, (**2011**), 2, 31910.1038/ncomms1324.

144 Wu, F. et al. A new coronavirus associated with human respiratory disease in China. Nature, (**2020**), 579, 265-26910.1038/s41586-020-2008-3.

145 Grigoriev, A. Mutational patterns correlate with genome organization in SARS and other coronaviruses. Trends in genetics : TIG, (**2004**), 20, 131-13510.1016/j.tig.2004.01.009.

146 Tulloch, F., Atkinson, N. J., Evans, D. J., Ryan, M. D. & Simmonds, P. RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. eLife, (**2014**), 3, e0453110.7554/eLife.04531.

147 Woo, P. C., Wong, B. H., Huang, Y., Lau, S. K. & Yuen, K. Y. Cytosine deamination and selection of CpG suppressed clones are the two major independent biological forces that shape codon usage bias in coronaviruses. Virology, (**2007**), 369, 431-44210.1016/j.virol.2007.08.010.

148 Xia, X. Extreme genomic CpG deficiency in SARS-CoV-2 and evasion of host antiviral defense. Molecular biology and evolution, (**2020**)10.1093/molbev/msaa094.

149 Pfeifer, G. P. Mutagenesis at methylated CpG sequences. Current topics in microbiology and immunology, (**2006**), 301, 259-281.

150 Alsoe, L. et al. Uracil Accumulation and Mutagenesis Dominated by Cytosine Deamination in CpG Dinucleotides in Mice Lacking UNG and SMUG1. Scientific reports, (**2017**), 7, 719910.1038/s41598-017-07314-5.

151 Behringer, M. G. & Hall, D. W. Genome-Wide Estimates of Mutation Rates and Spectrum in Schizosaccharomyces pombe Indicate CpG Sites are Highly Mutagenic Despite the Absence of DNA Methylation. G3 (Bethesda), (**2015**), 6, 149-16010.1534/g3.115.022129.

152 Boo, S. H. & Kim, Y. K. The emerging role of RNA modifications in the regulation of mRNA stability. Experimental & molecular medicine, (**2020**), 52, 400-40810.1038/s12276-020-0407-z.

153 Costello, M. et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic Acids Res, (**2013**), 41, e6710.1093/nar/gks1443.

154 Di Giorgio, S., Martignano, F., Torcia, M. G., Mattiuz, G. & Conticello, S. G. Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. Sci Adv, (**2020**), 6, eabb581310.1126/sciadv.abb5813.

155 Graudenzi, A., Maspero, D., Angaroni, F., Piazza, R. & Ramazzotti, D. Mutational signatures and heterogeneous host response revealed via large-scale characterization of SARS-CoV-2 genomic diversity. iScience, (**2021**), 24, 10211610.1016/j.isci.2021.102116.

156 Chiara, M. et al. Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. Brief Bioinform, (**2020**) 10.1093/bib/bbaa297.

157 Wang, Y. et al. Intra-host variation and evolutionary dynamics of SARS-CoV-2 populations in COVID-19 patients. Genome Med, (**2021**), 13, 3010.1186/s13073-021-00847-5.

158 Lythgoe, K. A. et al. SARS-CoV-2 within-host diversity and transmission. Science, (**2021**), eabg082110.1126/science.abg0821.

159 Wolf, Y. I. et al. Origins and Evolution of the Global RNA Virome. mBio, (**2018**), 9, e02329-0231810.1128/mBio.02329-18.

160 Simmonds, P. Methods for virus classification and the challenge of incorporating metagenomic sequence data. J Gen Virol, (**2015**), 96, 1193-120610.1099/jgv.0.000016.

161 Wolf, Y. I. et al. Doubling of the known set of RNA viruses by metagenomic analysis of an aquatic virome. Nat Microbiol, (**2020**), 5, 1262-127010.1038/s41564-020-0755-4.

162 Sadeghi, M., Tomaru, Y. & Ahola, T. RNA Viruses in Aquatic Unicellular Eukaryotes. Viruses, (**2021**), 13, 36210.3390/v13030362.

163 Simmons, H. E. et al. Deep sequencing reveals persistence of intra- and inter-host genetic diversity in natural and greenhouse populations of zucchini yellow mosaic virus. J Gen Virol, (**2012**), 93, 1831-184010.1099/vir.0.042622-0.

164 Hjelmso, M. H. et al. Evaluation of Methods for the Concentration and Extraction of Viruses from Sewage in the Context of Metagenomic Sequencing. PLoS One, (**2017**), 12, e017019910.1371/journal.pone.0170199.

165 Boldogkoi, Z., Moldovan, N., Balazs, Z., Snyder, M. & Tombacz, D. Long-Read Sequencing - A Powerful Tool in Viral Transcriptome Research. Trends in microbiology, (**2019**), 27, 578-59210.1016/j.tim.2019.01.010.

166 Lovestad, A. H., Jorgensen, S. B., Handal, N., Ambur, O. H. & Aamot, H. V. Investigation of intra-hospital SARS-CoV-2 transmission using nanopore whole genome sequencing. The Journal of hospital infection, (**2021**) 10.1016/j.jhin.2021.02.022.

167 Charre, C. et al. Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterisation. Virus evolution, (**2020**), 6, veaa07510.1093/ve/veaa075.

168 Li, J. et al. Rapid genomic characterization of SARS-CoV-2 viruses from clinical specimens using nanopore sequencing. Scientific reports, (**2020**), 10, 1749210.1038/s41598-020-74656-y.

169 Karst, S. M. et al. High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. Nat Methods, (**2021**), 18, 165-16910.1038/s41592-020-01041-y.

170 Fitzsimmons, W. J. et al. A speed-fidelity trade-off determines the mutation rate and virulence of an RNA virus. PLoS biology, (**2018**), 16, e200645910.1371/journal.pbio.2006459.

171 Schaaper, R. M. Mechanisms of mutagenesis in the Escherichia coli mutator mutD5: role of DNA mismatch repair. Proceedings of the National Academy of Sciences of the United States of America, (**1988**), 85, 8126-813010.1073/pnas.85.21.8126.

172 Schaaper, R. M. Base selection, proofreading, and mismatch repair during DNA replication in Escherichia coli. J Biol Chem, (**1993**), 268, 23762-23765.

173 Morrison, A., Johnson, A. L., Johnston, L. H. & Sugino, A. Pathway correcting DNA replication errors in Saccharomyces cerevisiae. The EMBO journal, (**1993**), 12, 1467-1473.

174 Robson, F. et al. Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. Mol Cell, (**2020**), 79, 710-72710.1016/j.molcel.2020.07.027.

175 Yan, L. W. L. & Zaher, H. S. How do cells cope with RNA damage and its consequences? Journal of Biological Chemistry, (**2019**), 294, 15158-1517110.1074/jbc.REV119.006513.

176 Mazina, O. M. et al. Replication protein A binds RNA and promotes R-loop formation. J Biol Chem, (**2020**), 295, 14203-1421310.1074/jbc.RA120.013812.

177 Chatterjee, N. & Walker, G. C. Mechanisms of DNA damage, repair, and mutagenesis. Environmental and molecular mutagenesis, (**2017**), 58, 235-26310.1002/em.22087.

178 Vaisman, A. & Woodgate, R. Translesion DNA polymerases in eukaryotes: what makes them tick? Critical reviews in biochemistry and molecular biology, (**2017**), 52, 274-30310.1080/10409238.2017.1291576.

179 Jebb, D. et al. Six reference-quality genomes reveal evolution of bat adaptations. Nature, (**2020**), 583, 578-58410.1038/s41586-020-2486-3.

180 Uriu, K., Kosugi, Y., Ito, J. & Sato, K. The Battle between Retroviruses and APOBEC3 Genes: Its Past and Present. (**2021**), 13, 124.

181 Nik-Zainal, S. et al. The genome as a record of environmental exposure. Mutagenesis, (**2015**), 30, 763-77010.1093/mutage/gev073.

182 Riva, L. et al. The mutational signature profile of known and suspected human carcinogens in mice. Nat Genet, (**2020**), 52, 1189-119710.1038/s41588-020-0692-4.

183 Kucab, J. E. et al. A Compendium of Mutational Signatures of Environmental Agents. Cell, (**2019**), 177, 821-836 e81610.1016/j.cell.2019.03.001.

 doi:10.20944/preprints202104.0484.v1