Control of Human Anelloviruses by Cytosine to Uracil Genome Editing

Anne L. Timmerman,a,b Joanna Kaczorowska,a,b Martin Deijs,a,b Margreet Bakker,a,b Lia van der Hoek a,b

aAmsterdam UMC, University of Amsterdam, Department of Medical Microbiology and Infection Prevention, Laboratory of Experimental Virology, Amsterdam, The Netherlands
bAmsterdam Institute for Infection and Immunity, Amsterdam, The Netherlands

Anne L. Timmerman and Joanna Kaczorowska contributed equally. Author order was determined based on contribution.

ABSTRACT Anelloviruses are the most common viruses infecting humans. Every human carries a nonpathogenic personal anellovirus virome (anellome), yet it is unknown which mechanisms contribute to its stability. Here, we assessed the dynamics and impact of a host antiviral defense mechanism—cytidine deaminase activity leading to C to U editing in anelloviruses—on the stability of the anellome. We investigated anellome sequence data obtained from serum samples collected every 6 months from two healthy subjects followed for more than 30 years. The subjects were infected by a total of 64 anellovirus lineages. Minus-stranded C to U editing was observed in lineages belonging to the Alpha-, Beta-, and Gammatorquevirus genera. The edited genomes were present within virus particles, therefore editing must have occurred at the late stages of the virus life cycle. Editing was favored by 5’-TC contexts in the virus genome, indicating that apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like, catalytic subunit 3 or A3 (APOBEC3) proteins are involved. Within a lineage, mutational dynamics varied over time and few fixations of mutations were detected, indicating that C to U editing is a dead end for a virus genome. We detected an editing coldspot in the GC-rich regions, suggesting that the GC-rich region is crucial for genome packaging, since only packaged virus particles were included in the analysis. Finally, we noticed a lineage-specific reduced concentration after an editing event, yet no clearance. In conclusion, cytidine deaminase activity does not clear anelloviruses, nor does it play a major role in virus evolution, but it does contribute to the stability of the anellome.

IMPORTANCE Despite significant attention on anellovirus research, the interaction between the anellovirus virome and the human host remains unknown. We show the dynamics of APOBEC3-mediated cytidine deaminase activity on anelloviruses during a 30-year period of chronic infection and postulate that this antiviral mechanism controls anelloviruses. These results expand our knowledge of anellovirus-host interactions, which may be important for the design of gene therapies.

KEYWORDS anellovirus, anellome, torque teno virus, torque teno mini virus, torque teno midi virus, APOBEC3, C to U genome editing, virome

Viruses have coexisted with humans for thousands of years, and their symbiosis can range from harmful to beneficial effects on the host fitness (1, 2). Anelloviruses are the most common viruses infecting humans (1, 3). The first anellovirus was discovered in 1997 in the blood of a Japanese patient with hepatitis-like symptoms (4). Many other anellovirus species have subsequently been detected from various biological sources, including blood, nasal secretions, saliva, bone marrow, lymph nodes, pancreatic tissue, feces, semen, and breast milk (5–10). The entire human population is likely infected by anelloviruses, and coinfections with multiple distinct lineages are common (3, 11, 12).
Each individual carries their own personal anellovirus virome (12), referred to as the anellome (3). Despite significant attention, no study has found an association between anellovirus infection and disease (1, 3), and anelloviruses are therefore regarded as commensals.

Anelloviruses have circular, single-stranded negative-sense DNA (ssDNA) genomes which range from 2 to 3.9 kb in length (13–15). Anelloviruses are members of the Anelloviridae family, are extremely diverse, and contain 3 genera infecting humans: Alphatorquevirus (torque teno viruses; TTV), Betatorquevirus (torque teno mini virus [TTMV]) and Gammatorquevirus (torque teno midi virus [TTMDV]). The genomes contain an untranslated guanine-cytosine (GC)-rich region, and 6 or 7 viral proteins are expressed through alternative splicing (11, 16). Although not all the viral protein functions are known, it was revealed that the ORF1 protein, the largest encoded protein, forms the nonenveloped viral capsid (17). Viral protein ORF2 is predicted to play a role in disrupting host immune responses through suppressing NF-κB (18), and the ORF2/2 protein is presumed to be involved in genome replication and expression (19, 20). The TTV-derived apoptosis-inducing protein (TAIP) is believed to be involved in apoptosis induction (21). Other proteins, named ORF1/1, ORF1/2, ORF2/3, and ORF3, are of unknown function (3). Anelloviruses probably replicate via rolling circle replication, and they use hairpins present in the untranslated region in this process (22, 23). The viruses lack a gene encoding a DNA polymerase (24) and are therefore presumably dependent on the host replication machinery in the nucleus for replication.

Because higher anellovirus titers are found in immunocompromised patients (25–28), it is likely that anellovirus loads are partly regulated by the host immune system. Recent findings based on AnelloScan (T7 phage display anellovirus peptide library immunoprecipitation assay) showed that anellovirus proteins elicit varied antibody reactivity (29). Reactivity was mainly observed against the C-terminal region of ORF1. Two studies have shown that TTV DNAs sometimes display evidence of hypermutation, most likely caused by APOBEC3 (apolipoprotein B mRNA-editing enzyme, catalytic subunit 3 or A3) activity, part of the host innate immune mechanism (30, 31).

The APOBEC3 enzymes catalyze the deamination of cytosine (C) into uracil (U) (32), also called C to U editing. The APOBEC3 subfamily consists of seven proteins, A3A, A3B, A3C, A3D, A3F, A3G, and A3H. All these proteins can deaminate viral RNA and/or DNA, and most have higher affinity for single-stranded DNA or RNA (33, 34). A3A, A3C, and A3H are active in the cytoplasm and nucleus, whereas A3B is only active in the nucleus and A3C, A3D, and A3F only in the cytoplasm. APOBEC3 proteins prefer a specific nucleotide context and in general most APOBEC3s deaminate the C in a 5'-TC nucleotide context, whereas A3G favors a 5'-CC nucleotide context for editing (the deaminated C is underlined).

The APOBEC3s play a role in the innate immunity restricting viruses and mobile genetic elements (33, 34). Deamination of Cs in viral genomes leads to decreased viral infectivity because introduced mutations disrupt the function of viral proteins (34). APOBEC3-mediated restriction in viruses was first found in retroviruses, including HIV-1, in which editing occurs during reverse transcription on the negative strand (35, 36). Subsequently, it was found that other RNA and DNA viruses, including coronaviruses, herpesviruses, paroviruses, and human papillomaviruses, can also be targeted by APOBEC3 (37–43). Underrepresentation of APOBEC3 target motifs has been found in the 3 anellovirus genera (37), suggesting a long-term evolutionary conflict between the virus and the host; the dynamics between anelloviruses and host molecular antiviral defense mechanisms is currently unknown.

We hypothesize that APOBEC3-mediated C to U editing of anelloviruses might be a host antiviral mechanism involved in anellome control. Therefore, we searched for edited profiles (>5% of Cs edited to Ts) in deep-sequenced anellovirus lineages which were present in regularly sampled (every 6 months), healthy individuals. We determined which anellovirus lineages were prone to editing (here named “PTE lineages”) and at which time point this editing occurred during the 30 years of follow-up. The
time points with recognizable editing were named editing attack time points (here, “EATs”). By examining the editing in the positive- and negative-stranded anellovirus DNA, we determined at which stage of the viral life cycle (either during rolling circle replication or assembly/release of virions) the genome editing occurred. Moreover, we examined which protein(s) of the APOBEC3 family (A3A-A3H) is/are the most likely candidate(s) to act on anelloviruses and searched for long-term persistence of edited variants to analyze whether these mutations may drive evolution. In addition, we investigated editing coldspots along the genome to unravel whether they represent essential regions for packaging of an anellovirus genome. Finally, we analyzed whether APOBEC3-mediated genome editing can eradicate anellovirus lineages over time.

RESULTS

C to U editing in alpha-, beta-, and gammatorquevirus. Examination of longitudinally collected serum samples provides the opportunity to follow lineage dynamics and the impact of C to U editing. Two study subjects were followed over time, both participating in the Amsterdam Cohort Studies (ACS) on HIV infection and AIDS (44). Subject 1 was a healthy HIV-1-negative male who donated blood every 6 months for over 34 years, providing a total of 55 serum samples from 1985 to 2019 starting at age 41. Subject 2 was also a healthy HIV-1-negative male who also donated blood every 6 months. He began at the age of 35 and was followed from 1987 to 2019 (follow-up of 32 years) for a total of 52 consecutive serum samples. The subjects were unrelated. The anellomes of these two subjects have been described recently (12). A total of 53 anellovirus lineages were found in subject 1 and 11 lineages in subject 2. Lineage calling required an intact ORF1 gene in the viral genome, and these were separated based on 95% identity (12).

To detect APOBEC3-mediated editing, we used a variant caller program, LoFreq (45), that can detect variants in reads aligning to a reference lineage. Subsequently, alternate consensus genomes are constructed using the detected variants. Alternate genomes which displayed ≥5% editing (more than 5% of the total Cs [or Gs] being edited to a T [or A]) were included in further analyses (Table S1). The parent reference lineages of the edited alternate genomes were named “PTE lineages.” In subject 1, 18 out of 53 lineages could be categorized as PTE (33.96%) (Fig. 1A). PTE lineages were found among the TTVs (12 of 31 lineages; 38.71%), TTMVs (5 of 15; 33.3%), and TTMDVs (1 of 7; 14.3%). In subject 2, we detected 2 PTE TTVs out of a total of 7 TTV lineages (25.6%) (Fig. 1B). In total, the C-to-T mutations found in the alternate genomes displayed a mean variant frequency of 0.26 (±0.34) (Table S2). Half of the PTE lineages showed C to T genome editing at more than 1 time point (“EAT”), as shown in Fig. 1C, with TTV-AMS-S1-41 showing the most EATs during the follow-up (Table S1).

To confirm that the observed C to T editing was not a sequencing artifact introduced during either rolling circle amplification and/or Illumina library preparation, we performed 3D-PCR. This is a semi-nested PCR in which the denaturation temperature of the second PCR is variable. At a lower denaturation temperature, templates which are affected by editing and therefore more AT-rich can be amplified, whereas unedited references do not yield PCR products at these temperatures because unedited templates require higher denaturation temperatures (±95°C) for amplification. We designed and performed three 3D-PCR assays for one lineage from subject 1 (TTMV-AMS-S1-44 at sample time point 1 [TP1]) (Fig. 1D). Amplified products from the lowest denaturation temperature, assumed to contain the edited templates, were subsequently cloned and sequenced. Indeed, products obtained in PCR at lower denaturation temperatures contained sequences with C-to-T mutations (Fig. 1E) with the same editing patterns as those obtained via Illumina sequencing (Fig. 1F).

Genome editing at the minus strand of anelloviruses. Editing in anellovirus genomes can theoretically take place on negative-stranded DNA or on both positive- and negative-stranded DNA. To determine the preferred target strand for editing, we compared the frequencies of C-to-T and G-to-A mutations on the minus strand (anellovirus genome). A significantly higher fraction of C-to-T mutations was observed compared to G to A mutations (Fig. 2A) (P < 0.0001, Friedman test followed by Dunn’s multiple-
comparison test). When comparing the C-to-T fraction on the negative/positive DNA strand, per genus, we found significantly higher fractions of C-to-T mutations on the negative strand in all genera (Fig. 2B) (P, 0.0001, Wilcoxon matched-pairs signed-rank test; no statistical test was performed for TTMDV due to low sample number).
Anellovirus replication and genome editing. We hypothesized that if the C-to-T mutations had been equally distributed between both the positive and negative strands, genome editing would likely have been occurring during virus replication in the nucleus, as this is the only cellular location where both negative- and positive-stranded viral DNA are present. The fact that C-to-T mutations were mainly found on the negative strand in anelloviruses leads to the hypothesis that genome editing takes place in the cytoplasm. In theory, this editing could then occur during entry (as described by Stenglein et al., 2010 [46]) or during assembly and release of progeny virus. If C to U editing indeed takes place during virus assembly and/or release, the viral genome within the virus particles would contain Us. Therefore, we analyzed the presence of Us in the packaged virus genomes. The nucleic acids originating from DNase-treated serum from Subject 1 sample TP1 (containing virus TTMV-AMS-S1-44; this sample was also used in the 3D-PCR as shown in Fig. 1) were treated with USER (Uracil-Specific Excision Reagent) enzyme. This reagent consists of Uracil DNA glycosylase and DNA glycosylase-lyase Endonuclease VIII and can catalyze the excision of the uracil base and break the phosphodiester backbone, leading to breakage of the single-stranded anellovirus DNA. If editing only occurs during virus entry, there should be no Us in the packaged virus genomes, and USER enzyme will not affect the 3D-PCR efficiency. We included several controls: synthetic DNA with either no editing (reference), 50% of the Cs changed to Ts (C>T), or 50% of the Cs changed to Us (C>U). As shown in Fig. 3, at a lower denaturing temperature of 73.2°C, PCR products from TP1 without USER treatment contain the edited targets. Importantly, this signal is lost when USER enzyme is used prior to PCR, exactly as it was observed in the C>U synthetic DNA control, proving that the genome inside virus particles contains Us. Notably, at higher denaturation temperatures (75.9°C and 95°C), PCR products from the TP1 sample were visible regardless of USER treatment. This result is expected because this amplification product originates from the unedited reference genome that is also present at TP1, completely matching the result obtained when a 1:1 control mix of synthetic reference DNA (no editing) and synthetic C>U control DNA is used as the input for PCR.

Editing context of APOBEC3-mediated editing. To predict which APOBEC3 proteins are involved, we analyzed the editing context. Since all APOBEC3 proteins favor 5'-CC or 5'-TC contexts for editing, we first calculated editing context dinucleotide motifs. To normalize for genome length and GC%, we calculated the normalized mutation frequency by dividing the number of actual mutations by the maximum possible mutation sites in PTE lineages. Significantly higher frequencies of 5'-TC mutation were found (edited C underlined), followed by 5'-CC dinucleotide context preference (Fig. 4A).
and B) \( P < 0.001 \), Friedman test, followed by Dunn’s multiple-comparison test). We also analyzed the nucleotide downstream of the edited C. Here, a significantly higher frequency of 5'-CT mutations was observed, followed by 5'-CA mutations (\( P < 0.001 \), Friedman test, followed by Dunn’s multiple-comparison test). In addition, editing contexts known to be preferably recognized by individual APOBEC3s were counted (Table S3, adapted from Silvas and Schiffer, 2019 [33]). The highest numbers of editing sites and possible editing sites were found for an A3C/D-preferred context (Fig. 4E). However, when we checked the editing frequencies (the number of edits per potential site as percentages), A3A-preferred contexts showed the highest frequency, followed by A3F-preferred contexts (\( P < 0.05 \), Friedman test followed by Dunn’s multiple-comparison test). The lowest number of preferred contexts frequency was found for A3G.

We also investigated the editing context for each anellovirus genus using nucleotide frequencies of 5 nucleotides (nt) upstream and 5 downstream of the edited C. For all genera, the preferred editing context was defined by two Ts upstream and one T downstream of the edited C (Fig. 5A), yet TTMVs and TTMDVs contained more T-enriched contexts than TTVs. The control context was also determined, which is the nucleotide context surrounding all Cs in the PTE reference lineages. The control context of PTE-TTVs was GC-rich, whereas the control contexts of PTE-TTMVs and PTE-TTMDVs were already T-enriched (Fig. 5B). To examine whether the probability of adjacent T is higher in the mutated context than in the control context, we compared the surrounding T-probability between the control and editing contexts. The probability of surrounding Ts was significantly higher in the mutated context than in the control context for all three genera, thus also for TTMV and TTMDV (\( P < 0.0076 \) for TTV, \( P < 0.0338 \) for TTMV, and \( P < 0.0082 \) for TTMDV, using a paired t test; Fig. 5C), therefore, C to T editing is— for all three genera— favored by an enriched T context.

### FIG 3
Presence of uracil in packaged anellovirus genomes. Ethidium bromide-stained agarose gel with PCR products generated by 3D-PCR targeting TTMV-AMS-S1-44. The primers for product 1 (Fig. 1D to F) were used on nucleic acids from subject 1 collected at start of study (time point 1, TP1). Samples were treated with (+) or without (−) USER enzyme (Uracil-Specific Excision Reagent), which can catalyze the excision of the uracil base and break the phosphodiester backbone of the products. Synthetic oligonucleotides were used as controls containing the DNA sequence of product 1 with 0 mutations (control unedited reference), 6 C-to-T mutations (50% editing; control edited C > T) or 6 C-to-U mutations (50% editing) (control edited C > U). In addition, a 1:1 mixture of products containing the control unedited reference and control edited C-to-U mutations was included (control mix reference + C > U) to mimic the mixture present in the serum sample. Products visible under 100 base pairs (bp) did not have the expected length and were considered primer dimers. Unprocessed agarose gel is shown in Fig. S2. M, 100-bp ladder.

| USER enzyme: | TTMV AMS-S1-44 | Control unedited Reference | Control edited C > T | Control edited C > U | Control mix Reference + C > U | Control NTC | M |
|-------------|----------------|---------------------------|----------------------|----------------------|-------------------------------|-------------|---|
| -           | -              | -                         | -                    | -                    | -                             | -           | - |
| +           | -              | -                         | -                    | -                    | -                             | -           | - |
| -           | +              | -                         | -                    | -                    | -                             | -           | - |
| +           | +              | -                         | -                    | -                    | -                             | -           | - |
| -           | +              | +                         | -                    | -                    | -                             | -           | - |
| +           | -              | -                         | -                    | -                    | -                             | -           | - |

No long-term persistence of edited variants. To examine whether edited genomes are dead ends or variants which drive anellovirus evolution, we analyzed whether there was any long-term persistence of the edited variants during the follow-up. We analyzed TTMV-AMS-S1-41, TTV-AMS-S1-06, TTV-AMS-S1-07, and TTV-AMS-S1-25 lineages with 11, 5, 6, and 6 EATs, respectively. As shown in Fig. 6, the editing patterns were...
different for each time point. Out of the total 608 C-to-U substitutions observed in the four lineages at the first EAT, only 7 were fixed in the lineages: 4 mutations in TTMV-AMS-S1-41 (positions 877, 1,299, 2,196 and 2,252) and 3 in TTV-AMS-S1-25 (positions 984, 1,640, and 1,748). TTV-AMS-S1-06 and TTV-AMS-S1-07 had none of the editing mutations fixed. The low level of fixation supports the hypothesis that C to U editing is a dead end for a viral genome.

Editing coldspots at GC-rich regions. We identified editing cold- and hotspots. At every C nucleotide in the genome, we counted the number of EATs that presented with a T at that position. For TTMV-AMS-S1-41, a hotspot is present at positions 877 and 1,299 (Fig. 7A). A coldspot was found at the GC-rich 2,800 region (described below in more detail), but also around nucleotide position 1,950 (lineage-specific coldspot). A lineage-specific coldspot was also found in TTV-AMS-S1-06 in the 20-to-70 region (Fig. 6B), and TTV-AMS-S1-25 contains a lineage-specific coldspot in the 1,000-to-1,400 region (Fig. 6D). We subsequently analyzed the context of the Cs in these lineage-specific coldspots. Figure 7B shows that the C context in the lineage-specific coldspots displays low upstream and downstream Ts, except for TTMV-AMS-S1-41.

To investigate the editing in the GC-rich untranslated region alongside editing per open reading frame (ORF), we subtracted the GC-rich region and ORFs (ORF1, ORF2, ORFx/2, ORF3, and TAIP; see “Genome annotation” in Materials and Methods) in the genomes of all PTE lineages. The TAIP-encoding region was excluded from further analysis because it was only found in 2 of the 20 PTE lineages. Only the previously

FIG 4 Dimer context analysis of editing in anellovirus genomes. (A to D) Numbers of editing and possible editing sites of dimer contexts and the normalized mutation frequency (total number of edits divided by number of possible edits). The edited Cs are underlined. (E) Numbers of editing and possible editing sites of contexts preferred by each APOBEC3 protein (A3A, A3B, A3C, A3D, A3F, A3G, and A3H). (F) Normalized editing frequency specific to the preferred context of each APOBEC3 protein (see also Table S3). *, P ≤ 0.05; **, P ≤ 0.01; ***, P ≤ 0.001; ****, P ≤ 0.0001 (Friedman test followed by Dunn’s multiple-comparison test). A3, APOBEC3.
mentioned coldspot GC-rich region stood out, with significantly less editing compared to the ORFs (Fig. 7C; \( P < 0.0001 \), using a Kruskal-Wallis test followed by Dunn’s multiple-comparison test). We analyzed whether the GC-rich region, like the lineage-specific coldspots, also harbors an unfavorable low-T context surrounding Cs. Indeed, a high level of Cs and Gs upstream and downstream of a target C is highly unfavorable for APOBEC3 editing (Fig. 7D). We also examined whether C to U editing in each ORF leads to nonsynonymous or synonymous substitutions. More nonsynonymous changes were found, with, on average, 9.2% of the nonsynonymous substitutions introducing stop codons (Fig. 7E and Table S1). This strengthens the hypothesis that C to U genome editing is normally lethal.

**C to U genome editing and anellovirus lineage persistence.** Finally, we analyzed the overall effect of an EAT on a PTE lineage. To quantify genome editing, we counted edited versus unedited aligned reads originating from the same reference PTE lineage. To enable quantification (only for counting purposes), we used de novo assembled scaffolds from SPAdes and filtered out those that could be used as “edited scaffolds” (47). We found 73 scaffolds that contained 5% or more of Cs edited to Ts (mean, 27.9%; minimum, 7%; maximum, 55.2%) (Table S4) corresponding to 17 PTE lineages. Next, we counted the reads aligning to the edited scaffolds, and compared this to the number of reads aligning to the unedited reference of the lineage. Of the 17 PTE lineages, the vast majority (16 lineages) were continuously present with no clearance after EAT (Fig. 8A). Only one lineage, TTMV-AMS-S1-47, may have been cleared after an EAT. This lineage was introduced in the anellome halfway through follow-up in subject 1 (around year 20), subjected to editing in year 28.6 of follow-up and subsequently absent during the following years (>5 years).

Although C to U genome editing seldom led to permanent loss of a PTE lineage, it
is possible that editing controls lineage load. To study this, we determined the normalized read count (reads per million [RPM]) for each PTE lineage before, during, and after an EAT. Interestingly, the highest concentrations were found at an EAT, followed by significantly fewer RPM following an EAT (mean fold change of 0.25, Friedman test followed by Dunn’s multiple-comparison test; \( P = 0.0002 \)) (Fig. 8B), indicating load-controlling activity by C to U editing in a lineage-specific fashion.

**FIG 6** Editing at different editing attack time points (EATs) per lineage. Overview of C to U editing positions displayed on the coding strand of the anellovirus genome per EAT. Y axes indicate years of follow-up for TTMV-AMS-S1-41 (A), TTV-AMS-S1-06 (B), TTV-AMS-S1-07 (C), and TTV-AMS-S1-25 (D).
DISCUSSION

This study analyzed the dynamics and impact of C to U editing on the human anelloome. Since it is not possible to perform functional cell culture analysis, we examined the full spectrum of packaged anellovirus DNA from longitudinally collected serum samples, with 30 years of follow-up. We found that anellovirus genome editing takes place during the later stages of the virus life cycle and that the 5'-TC context is the most common editing context. Within a lineage, limited fixation of mutations was detected, indicating that editing is a dead end for the virus. We found editing coldspots at the intergenic GC-rich regions and noticed lower anellovirus concentrations in a lineage after an editing event.

Evolutionary pressure to escape APOBEC3-mediated C to T editing activity has been...
FIG 8 C to U editing dynamics per PTE lineage during follow-up. (A) Relative proportion of reads aligning to the unedited reference sequence (blue) and the C to U edited scaffolds (orange) of a lineage for each PTE lineage. (B) Read counts (reads per million [RPM]) aligning to the unedited reference before, during, and after an EAT. Means were compared using a Friedman test followed by Dunn’s multiple-comparison test (*, P ≤ 0.05; ***, P ≤ 0.001).
found in DNA and some RNA viruses, including parvoviruses, herpesviruses, human papillomaviruses, coronaviruses, and hepatitis B virus (HBV) (30, 37–41, 43, 48). To date, studies which have reported C to T genome editing in anelloviruses are limited. Ball et al. (31) reported one edited TTV variant out of 93 sequenced clones, using Sanger sequencing, in a hepatitis C virus-infected patient over an interval of 29 months. Tsuge et al. (30) detected C to T genome editing in TTVs from 4 healthy and 4 HBV-infected individuals using 3D-PCR followed by Sanger sequencing, and Poulain et al. (37) described decreased APOBEC3 recognition motifs for some anelloviruses. We provide further evidence that anellovirus C-to-U deamination indeed occurs and can be found in all three anelloviruses genera infecting humans.

It remains unknown which APOBEC3 protein(s) is/are involved in anellovirus genome editing. We detected a preferred editing context of Ts surrounding a mutated C, specifically a 5'-TTCT general context, for all genera. Based on this editing context, A3G is the least likely candidate for genome editing because it favors a 5'-CC context (33). In addition, since we have shown that anellovirus genome editing does not occur on plus-stranded DNA, A3B can also be excluded as candidate because it is only active in the nucleus (34). We saw less editing in GC-rich regions, which have a high potential of forming secondary structures. Earlier research showed that A3A prefers double-stranded DNA (dsDNA), especially hairpin-forming sequences, and even has reduced activity at ssDNA (49–52). This, in combination with the fact that the number of A3A-preferred contexts was lower than that of the other APOBEC3s, allows us to exclude A3A as an editing candidate. The remaining APOBEC3 proteins, A3C, A3D, A3F, and A3H, are localized in the cytosol and prefer a 5'-TC context. Thus, they are most likely responsible for anellovirus genome editing, both individually and combined. These 4 APOBEC3 proteins have also been considered editing candidates in other viruses, including human papillomaviruses (A3C and A3H) (39), HBV (A3C, A3H) (53, 54), and HIV-1 (A3D, A3F, A3H) (55). Candidates A3F, but also A3G, have been previously mentioned as candidates for TTV editing (30).

Editing patterns from PTE lineages were chaotic and fluctuated over time. This may suggest that different APOBEC3 proteins were active or that the editing originated from different tissues where different APOBEC3 were expressed (34). However, the stochastic behavior of cytidine deaminases is more likely to be the cause of the random editing (56, 57). APOBEC3s bind at random positions on the genome, slide, on average, 30 nt several times along the genome, and then detach (random-walk model). During the sliding process, the protein leaves most of the bases untouched, even when a favorable nucleotide context is present. The stochastic behavior of APOBEC3s can lead to single mutations but, due to their repetitive sliding toward the same position, it also causes subsequent mutations (56, 57). This behavior may explain the seemingly random editing pattern.

We detected an editing coldspot in the GC-rich regions. In other viruses, including coronaviruses, herpes simplex viruses, adeno-associated viruses, human endogenous retroviruses, and human papillomaviruses, a reduction in C-to-U mutations has also been found in secondary RNA structures and GC-rich sequences (58, 59). This could be caused by APOBEC3 recognition site depletion. Alternatively, it is possible that the GC-rich region forms a hairpin structure, supporting DNA packaging (60), as seen for parvoviruses (61, 62). Because we analyzed only intact viruses (with packaged genomes), we expect that the signals needed for packaging were intact. Therefore, the identification of the GC-rich region as an editing coldspot is interesting.

We were not able to investigate whether APOBEC3 activity also occurs during virus entry. The lack of a culture system, with the target cells unknown, hampers research on anellovirus-host interaction during virus entry. We also could not identify the discriminating characteristics of PTE lineages versus anellovirus lineages that might escape control by APOBEC3. We identified a total of 20 lineages that were clearly susceptible to editing, yet the other lineages in our two study subjects could not be definitively categorized as “non-PTE” lineages. This is because, with blood sampling every
6 months, EATs may easily have been missed if they occurred during the 6-month interval between samplings.

We hypothesize that C to U genome editing is the force which keeps an anellome under control, with specific lineages culled via EAT when viral loads become targets. A similar effect of APOBEC3 genome editing on viral load has been previously described for HBV, with HBV loads decreasing following C to U genome editing in infected patients (30), and also in APOBEC3G-overexpressed epithelial liver cell lines (38). One possible alternative to specific control of the anellome is that APOBEC3 expression is occasionally upregulated due to stimulation by external factors, such as bacterial or viral infections (reviewed by Covino, Gauzzi, and Fantuzzi [63]). In this case, anellovirus genome editing is a bystander effect. Future studies may focus on the effect of external immune activation and the influence such forces may have on anellovirus diversity, load, and editing.

**MATERIALS AND METHODS**

**Serum samples.** Longitudinally collected serum samples from 2 healthy male donors (subjects 1 and 2) were analyzed. Both subjects are healthy participants in the Amsterdam Cohort Studies (ACS) on HIV infection and AIDS (44) and were chosen based on long-term follow-up (more than 30 years) and the absence of any comorbidities. The ACS was established to examine the prevalence, incidence, and risk factors of HIV-1 infection and included men who have sex with men who live mainly close to or in Amsterdam (44). Besides an HBV infection (subject 1 in 1998), both subjects had no history of blood disease, cancer, autoimmune disease, or neurodegenerative disease. The subjects donated blood roughly every 6 months, with one gap (1996 to 2003). Subject 1 donated at 55 time points, starting at age 41, from 1985 to 2019 (413 months). Subject 2 donated starting at age 35 from 1987 to 2019 (386 months). Serum samples were stored at −80°C. The ACS was approved by the Medical Ethics Committee of the Amsterdam University Medical Center of the University of Amsterdam, the Netherlands (MEC 07/182). All research was performed in accordance with relevant guidelines and regulations. Each donor participated voluntarily, without incentive, and submitted written informed consent upon enrollment.

**Anellovirus analysis and identification of PTE lineages via variant calling.** Isolated nucleic acids of all DNase-treated serum samples were rolling-circle amplified and sequenced using the Illumina MiSeq system (12) ([https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA785545](https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA785545)). Trimmed reads were de novo assembled resulting in the identification of 53 distinct lineages in subject 1 and 11 lineages in subject 2. Lineages were defined using a ≥95% identity threshold, have been published previously (12), and are available in GenBank under accession numbers OL694779 to OL694842. To identify PTE lineages, we searched, per lineage, for variants with editing signatures (≥5% of the Cs or Gs edited). To find these variants, we aligned the trimmed reads to the reference genomes using BWA-MEM (version 0.7.17) and converted the alignments to a BAM file using SAMtools (version 1.6, with -view) (64). GATK (version 4.2.6.0) was subsequently used to assign reads to a single new read group (AddOrReplaceReadGroups with RGID = c, RGLB = lib1, RGPL = ILLUMINA, RPGU = unit1, and RGSM = 20) and to mark duplicates (MarkDuplicatesSpark). Variants were called using LoFreq (version 2.1.5, standard settings) (45), a sensitive variant-caller program based on sequence quality and coverage. Variant tables were created using GATK (VariantsToTable -F CHROM -F VAR -F REF -F ALT -F POS -F AF -F DP -F TYPE -GF AD). An alternate genome from the reference lineage was created based on the variant list using SAMtools (faidx) GATK (CreateSequenceDictionary followed by IndexFeatureFile and FastaAlternateReferenceMaker). Alternate genomes were aligned to the reference genomes using CodonCode Aligner (version 8.0.2) and alternate genomes with ≥5% of Cs replaced by Ts (and/or Gs replaced by As) compared to their corresponding reference genome were defined as “edited” genomes (Table S1). In case the editing criteria were met in the variants, the unedited reference genome was considered a PTE lineage.

**Amplification and sequence analysis of editing using 3D-PCR.** One PTE lineage (TTMV-AMS-S1-44) was selected to confirm editing using 3D-PCR (adapted from Tsuge et al. [30]). A 3D-PCR uses a range of melting temperatures. Because GC-rich products need a higher melting temperature to denature double-stranded DNA compared to AT-rich DNA, AT-rich sequences can selectively amplify at lower denaturation temperatures. Three primer pairs, in a semi-nested setting, were designed in highly edited regions of TTMV-AMS-S1-44. All 3D-PCR primers were located in the ORF1 region (Table S5). The PCR mix contained 2.5 μL isolated nucleic acids, 12.5 μL DreamTaq Green PCR Master Mix (Thermo Fisher Scientific, CATT1081), 0.25 μL forward and 0.25 μL reverse primer (both 20 μM), and 9.5 μL H2O. The initial PCR began at 95°C for 5 min, followed by 35 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 1 min each and a final elongation step at 72°C for 10 min. Semi-nested PCR, containing 2.5 μL of the 1:10 H2O-diluted product of the first PCR, was performed at varying denaturation temperatures: 10°C for 5 min, followed by 35 cycles of 10°C for 30 s, 55°C for 30 s and 72°C for 1 min each and a final elongation step at 72°C for 10 min (see Table S5 for further details on product sizes and denaturing temperatures per each 3D-PCR). Products were separated by electrophoresis on 1.3% agarose gel and cloned into a pCR4-TOPO vector (Invitrogen, CATT45170) according to the manufacturer’s instructions. In short, vector-transformed *Escherichia coli* (Top10 strain) was cultured overnight at 37°C on X-Gal (Thermo Fisher Scientific, CATB1960)-covered plates (LB medium [Becton, Dickinson, CAT241420] 1% carbenicillin [Gibco, CAT10177-012]). Transformed colonies were selected by blue-white screening and the plasmid inserts amplified by
Control of Anelloviruses by C to U Genome Editing

Supplemental Material

Supplemental material is available online only.

FIG S1, TIF file, 0.4 MB.
ACKNOWLEDGMENTS

We thank all participants and coworkers of the Amsterdam Cohort Studies (ACS) on HIV infection and AIDS, a collaboration between the Public Health Service of Amsterdam, the Amsterdam University Medical Center of the University of Amsterdam, the Jan van Goyen Medical Center, and the HIV Focus Center of the DC Clinics. Funding was supported by Netherlands HIV Monitoring Foundation and financially and the Center for Infectious Disease Control of the Netherlands National Institute for Public Health and the Environment.

Data analysis work was performed on a high-performance computing cluster (Lisa Surfsara). We thank SURFsara (www.surfsara.nl) for support in using the Lisa Compute Cluster. We thank Cormac M. Kinsella for his excellent feedback on English writing. We thank Hans van der Spek for interesting discussions and Jeen de Graaf for support in writing the shell scripts.

J.K., A.L.T., M.D., and L.v.d.H., study design. M.B. investigation and resources. A.L.T., J.K., and M.D., C to T genome editing bioinformatics. A.L.T., wet lab experiments and figures. A.L.T. and L.v.d.H., writing of the manuscript with contributions from all authors. All authors approved the submitted version of the manuscript. This research was financially and the Center for Infectious Disease Control of the Netherlands National Institute for Public Health and the Environment.

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Lia van der Hoek (c.m.vanderhoek@amsterdamumc.nl).

We declare no competing interests.

REFERENCES

1. Koonin EV, Dolja VV, Krupovic M. 2021. The healthy human virome: from virus-host symbiosis to disease. Curr Opin Virol 47:86–94. https://doi.org/10.1016/j.coviro.2021.02.002.

2. Parker MT. 2016. An ecological framework of the human virome provides classification of current knowledge and identifies areas of forthcoming discovery. Yale J Biol Med 89:339–351.

3. Kaczorowska J, van der Hoek L. 2020. Human anelloviruses: diverse, omnipresent and commensal members of the virome. FEMS Microbiol Rev 44:305–313. https://doi.org/10.1093/femsre/fuaa007.

4. Nishizawa T, Okamoto H, Konishi K, Yoshizawa H, Miyakawa Y, Mayumi M. 1997. A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. Biochem Biophys Res Commun 241:92–97. https://doi.org/10.1006/bbrc.1997.7765.

5. Maggi F, Pilli M, Fornai C, Andreoli E, Tempestini E, Vatteroni M, Presciutti S, Marchi S, Pietrobelli A, Boner A, Pistello M, Bendinelli M. 2003. TT virus in the nasal secretions of children with acute respiratory diseases: relations to viremia and disease severity. J Virol 77:2418–2425. https://doi.org/10.1128/JVI.77.4.2418-2425.2003.

6. Inami T, Konomi N, Arakawa Y, Abe K. 2000. High prevalence of TT virus DNA in human saliva and semen. J Clin Microbiol 38:2407–2408. https://doi.org/10.1128/JCM.38.6.2407-2408.2000.

7. Itoh Y, Takahashi M, Fukuda M, Shibayama T, Ishikawa T, Tsuda F, Tanaka T, Nishizawa T, Okamoto H. 2000. Visualization of TT virus particles recovered from the sera and feces of infected humans. Biochem Biophys Res Commun 279:718–724. https://doi.org/10.1006/bbrc.2000.4013.

8. Kikuchi K, Miyazawa H, Abe K, Kako M, Katayama K, Fukushima S, Mishiro S. 2000. Indirect evidence of TTV replication in bone marrow cells, but not in hepatocytes, of a subacute hepatitis/aPLASIA nem patient. J Med Virol 61:165–170. https://doi.org/10.1002/(SICI)1096-9071(20005)61:1%3C165::AID-JMV273%3E3.0.CO;2-F.

9. Spandole S, Cimpenineri D, Berca LM, Mihăescu G. 2015. Human anelloviruses: an update of molecular, epidemiological and clinical aspects. Arch Virol 160:893–908. https://doi.org/10.1007/s00705-015-2363-9.

10. Goto K, Sugiyama K, Ando T, Mizutani F, Terabe K, Tanaka K, Nishiyama M, Wada Y. 2000. Detection rates of TT virus DNA in serum of umbilical cord blood, breast milk and saliva. Tohoku J Exp Med 191:203–207. https://doi.org/10.1620/tjem.191.203.

11. Arze CA, Springer S, Dudas G, Patel S, Bhattacharyya A, Swaninathan H, Brugnara C, Delagrave S, Ong T, Kahvejian A, Echelard Y, Weinstein EG, Hajar RJ, Andersen KG, Yozwiak NL. 2021. Global genome analysis reveals a vast and dynamic anellovirus landscape within the human virome. Cell Host Microbe 29:1305–1315.e6. https://doi.org/10.1016/j.chom.2021.07.001.

12. Kaczorowska J, Deijs M, Klein M, Bakker M, Jebbink MF, Sparreboom M, Kinsella CM, Timmerman AL, van der Hoek L. 2022. Diversity and long-term dynamics of human blood anelloviruses. J Virol 96:e0010922. https://doi.org/10.1128/jvi.00109-22.

13. Ninomiya M, Nishizawa T, Takahashi M, Lorenzo FR, Shimogawara T, Okamoto H. 2007. Identification and genomic characterization of a novel human torque teno virus of 3.2 kb. J Gen Virol 88:1939–1944. https://doi.org/10.1099/vir.0.82895-0.

14. Jones MS, Kapoor A, Lukashov VV, Simmonds P, Hecht F, Delwart E. 2005. New DNA viruses identified in patients with acute viral infection syndrome. J Virol 79:8230–8236. https://doi.org/10.1128/JVI.79.13.8230-8236.2005.

15. Miyata H, Tsunoda H, Kazi A, Yamada A, Khan MA, Murakami J, Kamahora T, Shiraki K, Hino S. 1999. Identification of a novel GC-rich 113-nucleotide region to complete the circular, single-stranded DNA genome of TT virus,
54. Baumert TF, Rösler C, Malim MH, Von Weizsäcker F. 2007. Hepatitis B virus DNA is subject to extensive editing by the human deaminase AP0BEC3C. Hepatology 46:682–689. https://doi.org/10.1002/hep.21733.

55. Hultquist JF, Lengyel JA, Refsland EW, LaRue RS, Lackey L, Brown WL, Harris RS. 2011. Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. J Virol 85:11220–11234. https://doi.org/10.1128/JVI.05238-11.

56. Senavirathne G, Jaszczur M, Auerbach PA, Upton TG, Chelico L, Goodman MF, Rueda D. 2013. Single-stranded DNA scanning and deamination by APOBEC3G cytidine deaminase at single-molecule resolution. J Biol Chem 287:15826–15835. https://doi.org/10.1074/jbc.M112.432790.

57. Pham P, Calabrese P, Park SJ, Goodman MF. 2011. Analysis of a single-stranded DNA-scanning process in which activation-induced deoxycytidine deaminase (AID) deaminates C to U haphazardly and inefficiently to ensure mutational diversity. J Biol Chem 286:24931–24942. https://doi.org/10.1074/jbc.M111.342790.

58. Senavirathne G, Jaszczur M, Auerbach PA, Upton TG, Chelico L, Goodman MF, Rueda D. 2013. Single-stranded DNA scanning and deamination by APOBEC3G cytidine deaminase at single-molecule resolution. J Biol Chem 287:15826–15835. https://doi.org/10.1074/jbc.M112.432790.

59. Pham P, Calabrese P, Park SJ, Goodman MF. 2011. Analysis of a single-stranded DNA-scanning process in which activation-induced deoxycytidine deaminase (AID) deaminates C to U haphazardly and inefficiently to ensure mutational diversity. J Biol Chem 286:24931–24942. https://doi.org/10.1074/jbc.M111.342790.

60. Chen J, MacCarthy T. 2017. The preferred nucleotide contexts of the AID/APOBEC cytidine deaminases have differential effects when mutating retroposon and virus sequences compared to host genes. PLoS Comput Biol 13:e1005471. https://doi.org/10.1371/journal.pcbi.1005471.

61. Cotmore SF, Tattersall P. 2013. Parvovirus diversity and DNA damage responses. Cold Spring Harb Perspect Biol 5:a012989. https://doi.org/10.1101/cshperspect.a012989.

62. Cotmore SF, Tattersall P. 2005. Genome packaging sense is controlled by the efficiency of the nick site in the right-end replication origin of paroviruses minute virus of mice and Lulll. J Virol 79:2287–2300. https://doi.org/10.1128/JVI.79.4.2287-2300.2005.

63. Covino DA, Gaucci MC, Fantuzzi L. 2018. Understanding the regulation of APOBEC3 expression: current evidence and much to learn. J Leukoc Biol 103:433–444. https://doi.org/10.1002/JLB.2MR0717-310R.

64. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

65. van der Hoek L, Sol CJA, Snijders F, Bartelsman JFW, Boom R, Goudsmitt J. 2019. Human immunodeficiency virus type 1 RNA populations in faeces with higher homology to intestinal populations than to blood populations. J Gen Virol 77:2415–2425. https://doi.org/10.1093/jmcb/mjab011.

66. Bitinaite J, Rubino M, Varma KH, Schildkraut I, Vaisvila R, Vaishkunaite R. 2007. USER friendly DNA engineering and cloning method by uracil excision. Nucleic Acids Res 35:1992–2002. https://doi.org/10.1093/nar/gkm041.

67. Wagih O. 2017. ggseqlogo: A versatile R package for drawing sequence logos. Bioinformatics 33:3645–3647. https://doi.org/10.1093/bioinformatics/btx469.

68. Woodcroft BJ, Boyd JA, Tyson GW. 2016. OrfM: a fast open reading frame predictor for metagenomic data. Bioinformatics 32:2702–2703. https://doi.org/10.1093/bioinformatics/btw241.

69. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy996.

70. Langmead B. 2010. Aligning short sequencing reads with Bowtie. Curr Protoc Bioinforma Chapter 11:Unit 11.17. https://doi.org/10.1002/0471250953.bi1107s32.

71. Robinson JT, Thovraldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. Nat Biotechnol 29:24–26. https://doi.org/10.1038/nbt.1754.