Gene Reports 29 (2022) 101709

Contents lists available at ScienceDirect

Gene Reports

journal homepage: www.elsevier.com/locate/genrep

Co-evolving pairs of complementary nucleotide sequence regions containing compensatory divergences and polymorphisms in rotavirus genomes

Kohei Oshima [a], Reina Setaka [a], Haruka Inui [a], Yuki Kobayashi [b], Yoshiyuki Suzuki [a,*]

[a] Graduate School of Science, Nagoya City University, 1 Yamanohata, Nagoya-shi, Aichi-ken 467-8501, Japan
[b] Nihon University Veterinary Research Center, Fujisawa-shi, Kanagawa-ken 252-0880, Japan

ARTICLE INFO

Edited by Dominic Voon

Keywords:
Bundling signal
Co-evolution
Compensatory divergence
Compensatory polymorphism
Rotavirus
Secondary structure

ABSTRACT

Group A rotavirus (RVA), possessing a genome of 11 segmented double-stranded RNAs, is considered to undergo selective packaging, in which one copy of each genomic segment is incorporated into a virion. The selective packaging is thought to be mediated by formation of supra-molecular complex for 11 positive-sense RNAs (+RNAs) through inter-segmental interaction of complementary nucleotide sequence regions, termed the bundling signal. Here, genomic sequences for mammalian and avian RVA isolates were analyzed to identify co-evolving pairs of complementary nucleotide sequence regions containing compensatory divergences and polymorphisms, allowing for existence of slightly deleterious mutations partially disrupting complementarity. In the predicted secondary structure of +RNAs, intra-segmental co-evolving pairs mostly constituted stems, and were involved in formation of pan-handle structure in the VP1, VP3, and NSP4 segments and stem-loop structure at the 5′ terminal region in the VP7 segment, at the 3′ terminal region in the VP6 and NSP4 segments, and in the middle of the NSP2 segment. Inter-segmental co-evolving pairs between the 5′ terminal region each in the VP3 and NSP3 segments, the middle and 5′ terminal regions in the NSP1 and NSP2 segments, respectively, and the 5′ and 3′ terminal regions in the NSP2 and NSP3 segments, respectively, tended to constitute loops, suggesting that these pairs may interact as bundling signals.

1. Introduction

Group A rotavirus (RVA) is an etiological agent of acute gastroenteritis in young mammals and birds (Bishop et al., 1973). The genome of RVA consists of 11 segmented, double-stranded RNAs (dsRNAs), each of which encodes viral protein 1 (VP1), VP2, VP3, VP4, VP6, VP7, non-structural protein 1 (NSP1), NSP2, NSP3, NSP4, or NSP5/6 (Desselberger, 2014). Due to a low particle-to-plaque forming unit ratio (two-five) and an equimolar synthesis of 11 dsRNAs as well as 11 negative-sense RNAs (-RNAs) in infected cells, RVA is considered to undergo selective packaging; i.e., one copy of each genomic segment is incorporated into a virion (Hundley et al., 1985; Patton, 1990; Joklik and Roner, 1995).

During the replication cycle of RVA, genomic segments are incorporated into a virion in the form of 5′ capped, 3′ non-polyadenylated, positive-sense RNA (+RNA) (McDonald et al., 2016). The +RNAs of genomic segments are known to form stem-loop structure typically within ~200 nucleotides at the 5′ and 3′ terminal regions and pan-handle structure via 5′ and 3′ terminal long-range interaction (Tortorici et al., 2006; Li et al., 2010). Upon incorporation into a virion, 11 +RNAs are thought to be bundled as a supra-molecular complex through inter-segmental interaction of complementary nucleotide sequence regions, termed the bundling signal (Goto et al., 2013; Borodavka et al., 2017; Fajardo et al., 2017). Bundling signals may be a target for antiviral drugs to inhibit the replication cycle of RVA. In addition, compatibility of bundling signals between RVA strains may be an

Abbreviations: -RNA, negative-sense RNA; +RNA, positive-sense RNA; dsRNA, double-stranded RNA; *i*, number of isolates; IAV, influenza A virus; INSD, International Nucleotide Sequence Database; *j*, smallest integer equal to or greater than $i \times q$; N.A., not applicable; NJ, neighbor-joining; NSP, non-structural protein; *p*, proportion of co-varying pairs among all pairs of complementary mono-nucleotide positions; *q*, proportion of isolates allowed to be non-complementary; RVA, group A rotavirus; VP, viral protein; *x*, length of pairs of complementary nucleotide sequence regions; X, unknown features; *y*, number of co-varying pairs.

* Corresponding author at: Graduate School of Science, Nagoya City University, 1 Yamanohata, Nagoya, Aichi 467-8501, Japan.
E-mail address: yossuzuk@nsc.nagoya-cu.ac.jp (Y. Suzuki).

https://doi.org/10.1016/j.genrep.2022.101709
Received 22 July 2022; Received in revised form 21 October 2022; Accepted 3 November 2022
Available online 8 November 2022

indicator for occurrence of reassortment (McDonald et al., 2016). However, the nucleotide sequences of bundling signals in RVA are largely unknown.

Complementary nucleotide sequence regions that are functionally relevant may co-evolve neutrally or nearly neutrally through compensatory mutations under the constraint for maintaining complementarity (Kimura, 1983, 1985). Compensatory mutations in bundling signals may suppress reassortment between wild-type and mutant strains, and may eventually promote speciation (Matthijnssens and Van Ranst, 2012; McDonald et al., 2016). Although reassortment is known to occur within mammalian RVA as well as within avian RVA (Ward et al., 1990; Schumann et al., 2009), mammalian and avian RVAs are considered to have evolved independently along with their hosts without reassortment (Ito et al., 2001; Trojnar et al., 2009). Thus, inter-segmental co-evolving pairs of complementary nucleotide sequence regions containing compensatory divergences between mammalian and avian RVAs were considered as candidates for bundling signals in a previous study (Suzuki, 2014).

It should be noted, however, that compensatory mutations in bundling signals may cause not only compensatory divergences between populations but also compensatory polymorphisms within populations (Kimura, 1983, 1985). In addition, mutations partially disrupting complementarity in bundling signals may exist within populations, because generally the proportion of slightly deleterious mutations may attain ~10 % (Fay et al., 2001). The purpose of the present study was to analyze genomic sequences for mammalian and avian RVA isolates to identify co-evolving pairs of complementary nucleotide sequence regions containing compensatory divergences and polymorphisms, allowing for existence of slightly deleterious mutations partially disrupting complementarity.

2. Materials and methods

2.1. Sequence data

2.1.1. Smaller data set

Nucleotide sequences of 11 genomic segments for 88 mammalian and two avian RVA isolates, which were analyzed in a previous study (Suzuki, 2014) to identify co-evolving pairs of complementary nucleotide sequence regions containing compensatory divergences between mammalian and avian RVAs without allowing for existence of slightly deleterious mutations disrupting complementarity, were re-analyzed in the present study (supplementary Table S1). Multiple sequence alignment has been performed for each genomic segment, and lack of reassortment between mammalian and avian RVAs has been confirmed by phylogenetic analysis; i.e., mammalian and avian RVA isolates always formed distinct clusters in the phylogenetic tree (Suzuki, 2014).

2.1.2. Larger data set

To examine the effect of the number of genomic sequences on efficiency of identifying co-evolving pairs of complementary nucleotide sequence regions, the complete nucleotide sequences of 11 genomic segments for 241 mammalian and 12 avian RVA isolates were retrieved from the International Nucleotide Sequence Database (INSD) on February 22, 2022 (supplementary Table S2). For each genomic segment, multiple sequence alignment was performed for the total of 253 nucleotide sequences using the computer program MAFFT (version 7.475) (Katoh et al., 2002) with default settings. Phylogenetic tree was constructed by the neighbor-joining (NJ) method with the p distance, which was known to produce reliable topologies when closely related sequences were analyzed (Nei and Kumar, 2000; Yoshida and Nei, 2016) using MEGA (version 10.1.8) (Kumar et al., 2018), and was visualized using FigTree (version 1.4.4) (Rambaut, 2018). Reliability of interior branches was assessed by the bootstrap method with 1000 re-samplings. In the phylogenetic trees, mammalian and avian RVAs always formed distinct clusters, except for VP4 (supplementary Figs. S1–S11), in which one reassortment event may be inferred to have occurred shortly after divergence of mammalian and avian RVAs. Nevertheless, lack of reassortment between mammalian and avian RVAs was largely supported (Ito et al., 2001; Trojnar et al., 2009).

2.2. Data analysis

2.2.1. Identifying co-evolving pairs of complementary nucleotide sequence regions containing compensatory divergences between mammalian and avian RVAs

Multiple alignments of nucleotide sequences for 11 genomic segments of mammalian and avian RVA isolates were analyzed to identify co-evolving pairs of complementary nucleotide sequence regions containing compensatory divergences between mammalian and avian RVAs. For this purpose, pairs of complementary mono-nucleotide positions within and between genomic segments were detected by examining complementarity for each of mammalian and avian RVA isolates. Here, only U-A and C-G were regarded as complementary, though U-G may form a wobble base pair, to reduce false-positives in identifying co-evolving pairs of complementary nucleotide sequence regions that were functionally relevant. In addition, proportion q of mammalian and avian RVA isolates was allowed to be non-complementary to take into account existence of slightly deleterious mutations partially disrupting complementarity. Since the proportion of slightly deleterious mutations may attain ~10 % (Fay et al., 2001), q was set to be 0 and 0.1. In practice, among number i of mammalian or avian RVA isolates, j, which was the smallest integer equal to or greater than $i \times q$, were allowed to be non-complementary. Among the pairs of complementary mono-nucleotide positions, co-varying pairs were defined as the pairs in which complementary bases observed in mammalian RVA isolates were different from those observed in avian RVA isolates. The proportion of co-varying pairs among all pairs of complementary mono-nucleotide positions was designated p.

The pairs of complementary mono-nucleotide positions detected above should include not only interacting pairs but also non-interacting pairs. Since compensatory mutations maintaining complementarity may be neutral or nearly neutral for interacting pairs, interacting pairs may tend to become co-varying pairs during evolution. As a result, the proportion of co-varying pairs among interacting pairs may become greater than that among non-interacting pairs, and p may represent an intermediate value. Thus, among the pairs of complementary nucleotide sequence regions, which were composed of two or more contiguous pairs of complementary mono-nucleotide positions, those containing a greater number of co-varying pairs than expected from p may be co-evolving and interacting. For each pair of complementary nucleotide sequence regions with the length of x bases, a statistical test was conducted to examine whether the number of co-varying pairs contained (y) was significantly greater than expected from p based on the binomial probability. Bonferroni correction for multiple testing was performed with the total number of the pairs of complementary nucleotide sequence regions with the length of x bases for ensuring 5 % family-wise false-positive rate (Sokal and Rohlf, 1995).

2.2.2. Identifying co-evolving pairs of complementary nucleotide sequence regions containing compensatory polymorphisms within mammalian and avian RVAs

Multiple alignments of nucleotide sequences for 11 genomic segments of mammalian and avian RVAs were also analyzed to identify co-evolving pairs of complementary nucleotide sequence regions containing compensatory polymorphisms within mammalian RVA. For this purpose, pairs of complementary mono-nucleotide positions within and between genomic segments were detected by examining complementarity for each of mammalian RVA isolates, allowing for non-complementarity in j isolates as described above (Fay et al., 2001). Among the pairs of complementary mono-nucleotide positions, co-varying pairs were defined as the pairs in which each of two or more

different pairs of complementary bases were observed in more than j mammalian RVA isolates. The proportion of co-varying pairs among all pairs of complementary mono-nucleotide positions was designated p. Pairs of complementary nucleotide sequence regions containing a greater number of co-varying pairs than expected from p were identified as described above (Sokal and Rohlf, 1995). Co-evolving pairs of complementary nucleotide sequence regions containing compensatory polymorphisms within avian RVA were identified in a similar manner.

2.2.3. Predicting secondary structures for co-evolving pairs of complementary nucleotide sequence regions

Intra-segmental and inter-segmental co-evolving pairs of complementary nucleotide sequence regions that are functionally relevant may constitute stems and loops, respectively, in the secondary structure of +RNAs for genomic segments. In particular, inter-segmental co-evolving pairs that were considered to be functionally relevant may be candidates for bundling signals. Secondary structure was predicted for +RNA of each genomic segment for human isolate Wa (Wyatt et al., 1980; Wentzel et al., 2013) and pigeon isolate PO-13 (Minamoto et al., 1988; Ito et al., 2001), which were prototypes of mammalian and avian RVAs, respectively, based on the minimum free energy using RNAfold (version 2.4.18) (Gruber et al., 2008) with default settings. In addition, consensus secondary structure was predicted taking into account sequence variation among 241 mammalian and 12 avian RVA isolates based on the minimum free energy using RNAalifold (version 2.4.18) (Gruber et al., 2008) with default settings. Since the length of nucleotide sequences was limited to 3000 bases in RNAalifold, the 5′ and 3′ terminal 1500 bases each were concatenated for the VP1 segment. The predicted consensus secondary structures for mammalian and avian RVAs were similar to the predicted secondary structures for Wa and PO-13, respectively, unless otherwise stated below.

3. Results

3.1. Smaller data set

Genomic sequences for 88 mammalian and two avian RVA isolates were analyzed to identify co-evolving pairs of complementary nucleotide sequence regions containing compensatory divergences between mammalian and avian RVAs. When slightly deleterious mutations disrupting complementarity were not allowed to exist by setting $q = 0$, seven pairs were identified as co-evolving (supplementary Table S3). After eliminating overlaps, three pairs remained; two intra-segmental pairs with six and three bases long within the VP7 segment and one inter-segmental pair with six bases long between the NSP2 and NSP3 segments (supplementary Table S4). Both of the intra-segmental pairs constituted stems in the secondary structure of the VP7 segment in Wa (supplementary Fig. S12) and PO-13 (supplementary Fig. S13), suggesting that the VP7 segment may form functional stem-loop structure at the 5′ terminal region. In the secondary structure of the NSP2 and NSP3 segments, the inter-segmental pair was involved in intra-segmental base-pairings in Wa (supplementary Fig. S12) but loops in PO-13 (supplementary Fig. S13), suggesting that the pair may interact as a bundling signal at least in avian RVA. These results were the same as reported in Suzuki (2014). When slightly deleterious mutations partially disrupting complementarity were allowed to exist by setting $q = 0.1$, one intra-segmental pair with six bases long within the VP3 segment was identified as co-evolving (supplementary Tables S6 and S7), in addition to one inter-segmental pair between the NSP2 and NSP3 segments identified above. The intra-segmental pair within the VP3 segment constituted a stem in PO-13 (Fig. 1). In addition, the pair constituted a stem in the consensus secondary structure for mammalian RVA (supplementary Fig. S14), suggesting that the pair may be involved in formation of pan-handle structure in the VP3 segment.

Co-evolving pairs of complementary nucleotide sequence regions were also identified as the pairs containing compensatory polymorphisms within 88 mammalian RVA isolates. When q was set to be 0, six intra-segmental and three inter-segmental pairs were identified as co-evolving (supplementary Tables S7 and S8). All of six intra-segmental pairs, including one pair with six bases long within the VP1 segment, two pairs with three and two bases long within the VP6 segment, and three pairs with three bases long within the NSP4 segment, constituted stems in Wa (Fig. 1, supplementary Fig. S12), supporting formation of pan-handle structure in the VP1 segment and stem-loop structure at the 3′ terminal region in the VP6 and NSP4 segments. In contrast, all of three inter-segmental pairs, i.e., one pair each between the VP1 and NSP3 segments, VP2 and NSP3 segments, and NSP3 and NSP5/6 segments, were only two bases long (supplementary Tables S7 and S8). As for the pairs between the VP1 and NSP3 segments and VP2 and NSP3 segments, positions 115–116 of the VP1 segment and positions 365–366 of the VP2 segment were both paired with positions 665–666 of the NSP3 segment. In addition, although positions 115–116 of the VP1 segment and positions 365–366 of the VP2 segment were involved in loops, positions 665–666 of the NSP3 segment were involved in intra-segmental base-pairings (supplementary Fig. S12). As for the pair between the NSP3 and NSP5/6 segments, positions 515–516 of the NSP3 segment were involved in a loop, but positions 145–146 of the NSP5/6 segment were involved in intra-segmental base-pairings (supplementary Fig. S12). Thus, these pairs may not interact as bundling signals. When q was set to be 0.1, six intra-segmental pairs were identified as co-evolving; five pairs within the VP6 segment and one pair within the NSP4 segment (supplementary Tables S9 and S10). All of six pairs constituted stems in Wa (Fig. 1), supporting formation of stem-loop structure at the 3′ terminal region in the VP6 segment and pan-handle structure in the NSP4 segment.

Co-evolving pairs of complementary nucleotide sequence regions were further identified as the pairs containing compensatory polymorphisms within two avian RVA isolates. When q was set to be 0, one intra-segmental pair with 11 bases long within the NSP4 segment was identified as co-evolving (supplementary Tables S11 and S12). The pair constituted a stem in PO-13 (Fig. 1), supporting formation of stem-loop structure at the 3′ terminal region in the NSP4 segment. The analysis with $q = 0.1$ was not applicable, because of the small number of genomic sequences analyzed (two).

3.2. Larger data set

In the analysis of genomic sequences for 241 mammalian and 12 avian RVA isolates to identify co-evolving pairs of complementary nucleotide sequence regions containing compensatory divergences between mammalian and avian RVAs with $q = 0$, two pairs were identified as co-evolving (supplementary Tables S13 and S14). Both of these pairs, i.e., one intra-segmental pair within the VP7 segment and one inter-segmental pair between the NSP2 and NSP3 segments, were identified in the analysis of smaller data set. When q was set to be 0.1, one intra-segmental pair within the VP3 segment and one inter-segmental pair between the NSP2 and NSP3 segments were identified as co-evolving (Tables 1 and 2), as observed in the analysis of smaller data set. However, two inter-segmental pairs between the VP3 and NSP3 segments and NSP1 and NSP2 segments were additionally identified as co-evolving (Tables 1 and 2). As for the pair between the VP3 and NSP3 segments, positions 34–37 in both Wa and PO-13 of the VP3 segment and positions 158–161 in Wa and positions 201–204 in PO-13 of the NSP3 segment tended to constitute loops (Fig. 1), suggesting that the pair may interact as a bundling signal. As for the pair between the NSP1 and NSP2 segments, positions 172–176 in Wa and positions 167–171 in PO-13 of the NSP2 segment were involved in loops, but positions 664–668 in Wa and positions 783–787 in PO-13 of the NSP1 segment were involved in intra-segmental base-pairings (supplementary Figs. S12 and S13). However, in the consensus secondary structure for the NSP1 and NSP2 segments in mammalian RVA, the pair constituted loops (supplementary Fig. S15), suggesting that the pair may interact as

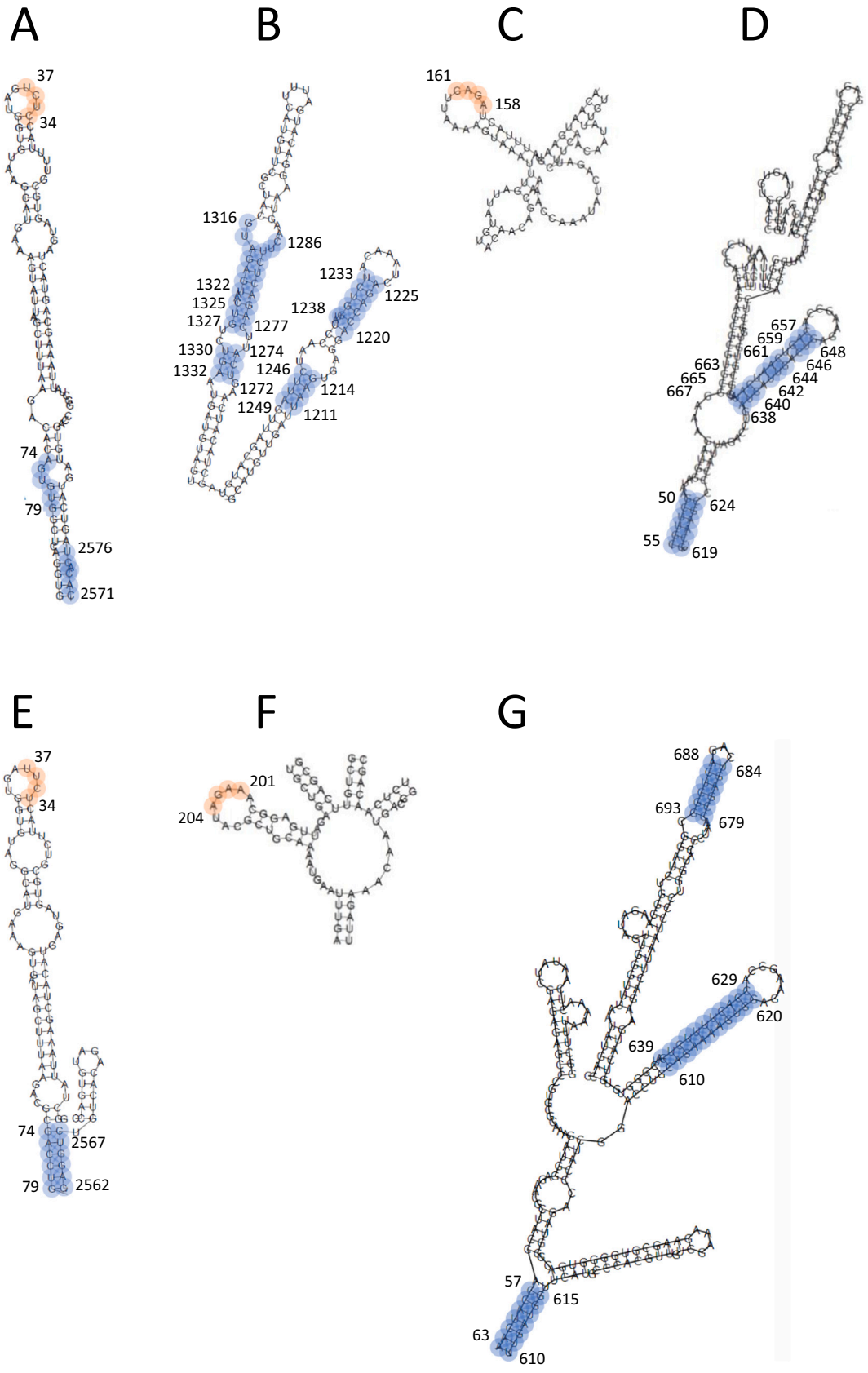

Fig. 1. Local secondary structures around nucleotide positions involved in intra-segmental (blue) and inter-segmental (orange) co-evolving pairs of complementary nucleotide sequence regions in the VP3 (A), VP6 (B), NSP3 (C), and NSP4 (D) segments for Wa and in the VP3 (E), NSP3 (F), and NSP4 (G) segments for PO-13. The ends of the regions are labeled with position numbers. Secondary structure was predicted for the entire nucleotide sequence of +RNA for each genomic segment based on the minimum free energy using RNAfold (version 2.4.18) (Gruber et al., 2008) with default settings. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Numbers of pairs of complementary nucleotide sequence regions in 241 mammalian and 12 avian RVA isolates obtained with $q = 0.1$.

	y[b]									
x[a]	8	7	6	5	4	3	2	1	0	Total
8	**0**[c]	**0**	**0**	**0**	**0**	**0**	**0**	0	1	1
7	N.A.[d]	**0**	**0**	**0**	**0**	**0**	**0**	1	7	8
6	N.A.	N.A.	**0**	**0**	**0**	**2**	0	2	25	29
5	N.A.	N.A.	N.A.	**0**	**0**	**3**	2	10	355	370
4	N.A.	N.A.	N.A.	N.A.	**0**	**5**	3	68	3057	3133
3	N.A.	N.A.	N.A.	N.A.	N.A.	6	36	607	31,104	31,753
2	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	834	9617	726,921	737,372
1	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	155,699	10,534,646	10,690,345

[a] Length of pairs of complementary nucleotide sequence regions.
[b] Number of co-varying pairs of complementary mono-nucleotide positions.
[c] Numbers of pairs of complementary nucleotide sequence regions identified as co-evolving are bold-faced.
[d] Not applicable.

Table 2
Nucleotide positions and sequences involved in co-evolving pairs of complementary nucleotide sequence regions in Wa and PO-13 identified in the analysis of 241 mammalian and 12 avian RVA isolates with $q = 0.1$.

x[a]	y[b]	Genomic segment	Positions in Wa	Sequence in Wa	Positions in PO-13	Sequence in PO-13	Genomic segment	Positions in Wa	Sequence in Wa	Positions in PO-13	Sequence in PO-13
6[c]	**3**	**NSP2**	**43–48**	**5'AGCCAU3'**	**41–46**	**5'CGUUAU3'**	**NSP3**	**1050–1055**	**5'AUGGCU3'**	**1068–1073**	**5'AUAACG3'**
6	**3**	**VP3**	**74–79**	**5'AGUGUG3'**	**74–79**	**5'GACCUG3'**	**VP3**	**2571–2576**	**5'CACACU3'**	**2562–2567**	**5'CAGGUC3'**
5	3	VP3	74–78	5'AGUGU3'	74–78	5'GACCU3'	VP3	2572–2576	5'ACACU3'	2563–2567	5'AGGUC3'
5	**3**	**NSP1**	**664–668**	**5'UAAAU3'**	**783–787**	**5'AUGGU3'**	**NSP2**	**172–176**	**5'AUUUU3'**	**167–171**	**5'ACCAU3'**
5	3	NSP2	43–47	5'AGCCA3'	41–45	5'CGUUA3'	NSP3	1051–1055	5'UGGCU3'	1069–1073	5'UAACG3'
4	3	VP3	74–77	5'AGUG3'	74–77	5'GACC3'	VP3	2573–2576	5'CACU3'	2564–2567	5'GGUC3'
4	**3**	**VP3**	**34–37**	**5'CUCU3'**	**34–37**	**5'UCUU3'**	**NSP3**	**158–161**	**5'AGAG3'**	**201–204**	**5'AAGA3'**
4	3	NSP1	664–667	5'UAAA3'	783–786	5'AUGG3'	NSP2	173–176	5'UUUU3'	168–171	5'CCAU3'
4	3	NSP1	665–668	5'AAAU3'	784–787	5'UGGU3'	NSP2	172–175	5'AUUU3'	167–170	5'ACCA3'
4	3	NSP2	43–46	5'AGCC3'	41–44	5'CGUU3'	NSP3	1052–1055	5'GGCU3'	1070–1073	5'AACG3'

[a] Length of co-evolving pairs of complementary nucleotide sequence regions.
[b] Number of co-varying pairs of complementary mono-nucleotide positions.
[c] Co-evolving pairs of complementary nucleotide sequence regions retained after eliminating overlaps are bold-faced.

a bundling signal at least in mammalian RVA.

Co-evolving pairs of complementary nucleotide sequence regions were also identified as the pairs containing compensatory polymorphisms within 241 mammalian RVA isolates. When q was set to be 0, one intra-segmental pair within the VP6 segment and one inter-segmental pair between the NSP3 and NSP5/6 segments were identified as co-evolving (supplementary Tables S15 and S16), as observed in the analysis of smaller data set. However, two intra-segmental pairs within the VP4 and NSP1 segments were additionally identified as co-evolving. Although positions 217–218 and positions 958–959 of the VP4 segment did not constitute a stem, positions 1529–1530 and positions 1550–1551 of the NSP1 segment constituted a stem in Wa (supplementary Fig. S12), supporting formation of stem-loop structure at the 3′ terminal region in the NSP1 segment. When q was set to be 0.1, three intra-segmental pairs within the VP6 segment were identified as co-evolving (supplementary Tables S17 and S18), as observed in the analysis of smaller data set.

Co-evolving pairs of complementary nucleotide sequence regions were further identified as the pairs containing compensatory polymorphisms within 12 avian RVA isolates. When q was set to be 0, three intra-segmental pairs within the NSP4 segment were identified as co-evolving (supplementary Tables S19 and S20). All of three pairs constituted stems in PO-13 (Fig. 1), supporting formation of pan-handle structure and stem-loop structure at the 3′ terminal region in the NSP4 segment. When q was set to be 0.1, two intra-segmental pairs within the NSP4 segment were identified as co-evolving (supplementary tables S21 and S22), as observed with $q = 0$. Two intra-segmental pairs within the NSP1 and NSP2 segments were additionally identified as co-evolving. Although positions 1427–1429 and positions 1536–1538 of the NSP1 segment did not constitute a stem in PO-13 (supplementary Fig. S13), positions 334–342 and positions 415–423 of the NSP2 segment constituted a stem in the consensus secondary structure for avian RVA (supplementary Fig. S16), supporting formation of stem-loop structure in the middle of the NSP2 segment.

4. Discussion

4.1. Improving efficiency for identifying co-evolving pairs of complementary nucleotide sequence regions

In a previous study, genomic sequences for 88 mammalian and two avian RVA isolates, corresponding to the smaller data set in the present study, were analyzed to identify co-evolving pairs of complementary nucleotide sequence regions containing compensatory divergences between mammalian and avian RVAs, without allowing for existence of slightly deleterious mutations disrupting complementarity (Suzuki, 2014). In the present study, slightly deleterious mutations partially disrupting complementarity were allowed to exist ~10 % among mammalian and avian RVA isolates (Fay et al., 2001). Since the p value obtained with $q = 0.1$ was greater than that obtained with $q = 0$, co-evolving pairs identified with $q = 0.1$ tended to be longer and contain more compensatory divergences than those identified with $q = 0$. Co-evolving pairs of complementary nucleotide sequence regions were also identified using compensatory polymorphisms within mammalian and avian RVAs. Compensatory polymorphisms may be useful for identifying bundling signals that were not only incompatible but also compatible (Heiman et al., 2008; Hutchinson et al., 2010; Boyce et al., 2016; McDonald et al., 2016) or different in genomic locations (Gavazzi et al., 2013; Gerber et al., 2014) between mammalian and avian RVAs. In addition, the data set was updated to include genomic sequences for 241 mammalian and 12 avian RVA isolates, corresponding to the larger data set in the present study. Some pairs of complementary nucleotide

sequence regions that were not identified as co-evolving in the analysis of smaller data set could be identified in the analysis of larger data set. Conversely, however, some of the co-evolving pairs that were identified in the analysis of smaller data set were missed in the analysis of larger data set. Since more sequence variants were included in the larger data set than in the smaller data set, it may be critical to distinguish neutral or nearly neutral polymorphisms from slightly deleterious mutations for efficiently identifying co-evolving pairs in the larger data set.

4.2. Intra-segmental co-evolving pairs of complementary nucleotide sequence regions

In the predicted secondary structure of +RNAs, intra-segmental co-evolving pairs of complementary nucleotide sequence regions mostly constituted stems within the 5′ and 3′ terminal ~200 nucleotides of genomic segments (Li et al., 2010). In mammalian RVA, the pairs appeared to be involved in formation of pan-handle structure in the VP1 and NSP4 segments and stem-loop structure at the 5′ terminal region in the VP7 segment and at the 3′ terminal region in the VP6 and NSP4 segments. In avian RVA, the pairs appeared to be involved in formation of pan-handle structure in the VP3 and NSP4 segments and stem-loop structure at the 5′ terminal region in the VP7 segment, at the 3′ terminal region in the NSP4 segment, and in the middle of the NSP2 segment. Interestingly, in both mammalian and avian RVAs, some of the pairs appeared to support formation of loops by means of constituting stems at their bases, implying functional importance of loops as well as stems (Boyce et al., 2016; Suzuki, 2016).

4.3. Inter-segmental co-evolving pairs of complementary nucleotide sequence regions

In the predicted secondary structure of +RNAs, inter-segmental co-evolving pairs of complementary nucleotide sequence regions between the 5′ terminal region each in the VP3 and NSP3 segments, the middle and 5′ terminal regions in the NSP1 and NSP2 segments, respectively, and the 5′ and 3′ terminal regions in the NSP2 and NSP3 segments, respectively, tended to constitute loops, suggesting that these pairs may interact as bundling signals. These hypotheses may be tested by examining effects of mutations disrupting and restoring complementarity in these pairs on efficiency of genome packaging (Gavazzi et al., 2013; Komoto et al., 2020).

On the other hand, some of inter-segmental co-evolving pairs did not appear to be involved in loops. It should be noted, however, that in viroplasms where bundling of +RNAs is thought to occur, NSP2 binds to +RNAs and unwinds intra-segmental helices to facilitate inter-segmental interaction (Taraporewala and Patton, 2001; Borodavka et al., 2017). Thus, some of inter-segmental co-evolving pairs predicted to be involved in intra-segmental base-pairings based on the minimum free energy (Gruber et al., 2008) may interact in viroplasms. In fact, it was found that the NSP5/6 segment preferentially bound to the VP3, VP6, NSP1, and NSP3 segments in the presence of NSP2 (Borodavka et al., 2017). In particular, positions 495–507 of the NSP3 segment and positions 88–99 of the NSP5/6 segment were found to interact. In the present study, one inter-segmental co-evolving pair identified in mammalian RVA was comprised of positions 515–516 of the NSP3 segment and positions 145–146 of the NSP5/6 segment, which were closely located to the interacting positions described above. However, in the predicted secondary structure based on the minimum free energy (Gruber et al., 2008), positions 145–146 of the NSP5/6 segment were involved in intra-segmental base-pairings, although positions 515–516 of the NSP3 segment were located in a loop. Nevertheless, positions 145–146 of the NSP5/6 segment appeared to be located in a loop in the presence of NSP2 (Borodavka et al., 2017). These observations suggested a possibility that the inter-segmental co-evolving pair between the NSP3 and NSP5/6 segments may also interact as a bundling signal.

4.4. Small number of candidates for bundling signals

Since at least 10 bundling signals are required for bundling 11 genomic segments of RVA, the number of candidates for bundling signals identified in the present study appeared to be relatively small. It should be noted that influenza A virus (IAV), which possesses an eight-segmented -RNA genome, is considered to undergo selective packaging in a similar manner to RVA (Noda et al., 2006; Fournier et al., 2012). For IAV, however, it has been reported that at least one genomic segment was missing in most viral particles (Brooke et al., 2013), suggesting that bundling signals may not be fully equipped to ensure bundling of all genomic segments. Thus, both selective and random packaging (semi-selective packaging) may operate in IAV (Suzuki, 2016). If this was also the case for RVA, the number of bundling signals may be smaller than 10.

Nevertheless, it was still likely that the method adopted in the present study had a relatively low sensitivity for identifying bundling signals, because pan-handle structure and 5′ and 3′ terminal stem-loop structures, which were supposed to exist in 11 genomic segments (Tortorici et al., 2006; Li et al., 2010), were not identified in some genomic segments. In the method, bundling signals were assumed to be inter-segmental pairs of complementary nucleotide sequence regions, which were co-evolving under the constraint for maintaining complementarity. In addition, bundling signals should be sufficiently long without gaps and contain significant amount of compensatory divergences or polymorphisms to be detected by statistical analysis. However, these assumptions may not always be satisfied. It has been proposed for IAV that many regions throughout each genomic segment interact with multiple regions in other segments to construct a large inter-segmental interaction network (Le Sage et al., 2020). The network was considered to be flexibly reorganized in response to mutations without reducing replication efficiency. If this was also the case for RVA, bundling signals may not co-evolve because the constraint for maintaining complementarity in each bundling signal may be relatively weak. Bundling signals having additional functions, such as protein-coding, also may not co-evolve, because compensatory mutations may not be neutral or nearly neutral (Boyce et al., 2016; Suzuki, 2016). In addition, some bundling signals may be short and contain gaps (Li et al., 2010; Le Sage et al., 2020). It may be necessary to further improve the method by relaxing the assumptions for identifying missing bundling signals in RVA.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.genrep.2022.101709.

CRediT authorship contribution statement

Kohei Oshima: Methodology, Investigation, Visualization, Writing – original draft. **Reina Setaka:** Methodology, Investigation, Visualization, Writing – original draft. **Haruka Inui:** Validation. **Yuki Kobayashi:** Conceptualization. **Yoshiyuki Suzuki:** Conceptualization, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare no conflict of interest.

Data availability

Data will be made available on request.

Acknowledgements

The authors thank two anonymous reviewers for valuable comments. This work was supported by JSPS KAKENHI Grant Number JP19K12221 and AMED Grant Number JP21fk0108120h0703 to Y.S.

References

Bishop, R.F., Davidson, G.P., Holmes, I.H., Ruck, B.J., 1973. Virus particles in epithelial cells of duodenal mucosa from children with acute non-bacterial gastroenteritis. Lancet 302, 1281–1283.

Borodavka, A., Dykeman, E.C., Schrimpf, W., Lamb, D.C., 2017. Protein-mediated RNA folding governs sequence-specific interactions between rotavirus genome segments. elife 6, e27453.

Boyce, M., McCrae, M.A., Boyce, P., Kim, J.T., 2016. Inter-segment complementarity in orbiviruses: a driver for co-ordinated genome packaging in the reoviridae? J. Gen. Virol. 97, 1145–1157.

Brooke, C.B., Ince, W.L., Wrammert, J., Ahmed, R., Wilson, P.C., Bennink, J.R., Yewdell, J.W., 2013. Most influenza a virions fail to express at least one essential viral protein. J. Virol. 87, 3155–3162.

Desselberger, U., 2014. Rotaviruses. Virus Res. 190, 75–96.

Fajardo Jr., T., Sung, P.-Y., Celma, C.C., Roy, P., 2017. Rotavirus genomic RNA complex forms via specific RNA-RNA interactions: disruption of RNA complex inhibits virus infectivity. Viruses 9, 167.

Fay, J.C., Wyckoff, G.J., Wu, C.-I., 2001. Positive and negative selection on the human genome. Genetics 158, 1227–1234.

Fournier, E., Moules, V., Essere, B., Paillart, J.-C., Sirbat, J.-D., Isel, C., Cavalier, A., Rolland, J.-P., Thomas, D., Lina, B., Marquet, R., 2012. A supramolecular assembly formed by influenza a virus genomic RNA segments. Nucleic Acids Res. 40, 2197–2209.

Gavazzi, C., Yver, M., Isel, C., Smyth, R.P., Rosa-Calatrava, M., Lina, B., Moules, V., Marquet, R., 2013. A functional sequence-specific interaction between influenza a virus genomic RNA segments. Proc. Natl. Acad. Sci. U. S. A. 110, 16604–16609.

Gerber, M., Isel, C., Moules, V., Marquet, R., 2014. Selective packaging of the influenza a genome and consequences for genetic reassortment. Trends Microbiol. 22, 446–455.

Goto, H., Muramoto, Y., Noda, T., Kawaoka, Y., 2013. The genome-packaging signal of the influenza a virus genome comprises a genome incorporation signal and a genome-bundling signal. J. Virol. 87, 11316–11322.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R., Hofacker, I.L., 2008. The Vienna RNA websuite. Nucleic Acids Res. 36, W70–W74.

Heiman, E.M., McDonald, S.M., Barro, M., Taraporewala, Z.F., Bar-Magen, T., Patton, J. T., 2008. Group a human rotavirus genomics: evidence that gene constellations are influenced by viral protein interactions. J. Virol. 82, 11106–11116.

Hundley, F., Biryahwaho, B., Gow, M., Desselberger, U., 1985. Genome rearrangements of bovine rotavirus after serial passage at high multiplicity of infection. Virology 143, 88–103.

Hutchinson, E.C., von Kirchbach, J.C., Gog, J.R., Digard, P., 2010. Genome packaging in influenza a virus. J. Gen. Virol. 91, 313–328.

Ito, H., Sugiyama, M., Masubuchi, K., Mori, Y., Minamoto, N., 2001. Complete nucleotide sequence of a group a avian rotavirus genome and a comparison with its counterparts of mammalian rotaviruses. Virus Res. 75, 123–138.

Joklik, W.K., Roner, M.R., 1995. What reassorts when reovirus genome segments reassort? J. Biol. Chem. 270, 4181–4184.

Katoh, K., Misawa, K., Kuma, K.-I., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. Nucleic Acids Res. 30, 3059–3066.

Kimura, M., 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge, New York, Melbourne.

Kimura, M., 1985. The role of compensatory neutral mutations in molecular evolution. J. Genet. 64, 7–19.

Komoto, S., Fukuda, S., Murata, T., Taniguchi, K., 2020. Reverse genetics system for human rotaviruses. Microbiol. Immunol. 64, 401–406.

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.

Le Sage, V., Kanarek, J.P., Snyder, D.J., Cooper, V.S., Lakdawala, S.S., Lee, N., 2020. Mapping of influenza virus RNA-RNA interactions reveals a flexible network. Cell Rep. 31, 107823.

Li, W., Manktelow, E., von Kirchbach, J.C., Gog, J.R., Desselberger, U., Lever, A.M., 2010. Genomic analysis of codon, sequence and structural conservation with selective biochemical-structure mapping reveals highly conserved and dynamic structures in rotavirus RNAs with potential cis-acting function. Nucleic Acids Res. 38, 7718–7735.

Matthijnssens, J., Van Ranst, M., 2012. Genotype constellation and evolution of group a rotaviruses infecting humans. Curr. Opin. Virol. 2, 426–433.

McDonald, S.M., Nelson, M.I., Turner, P.E., Patton, J.T., 2016. Reassortment in segmented RNA viruses: mechanisms and outcomes. Nat. Rev. Microbiol. 14, 448–460.

Minamoto, N., Oki, K., Tomita, M., Kinjo, T., Suzuki, Y., 1988. Isolation and characterization of rotavirus from feral pigeon in mammalian cell cultures. Epidemiol. Infect. 100, 481–492.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, Oxford, New York.

Noda, T., Sagara, H., Yen, A., Takada, A., Kida, H., Cheng, R.H., Kawaoka, Y., 2006. Architecture of ribonucleoprotein complexes in influenza a virus particles. Nature 439, 490–492.

Patton, J.T., 1990. Evidence for equimolar synthesis of double-stranded RNA and minus-strand RNA in rotavirus-infected cells. Virus Res. 17, 199–208.

Rambaut, A., 2018. FigTree. http://tree.bio.ed.ac.uk/software/figtree.

Schumann, T., Hotzel, H., Otto, P., Johne, R., 2009. Evidence of interspecies transmission and reassortment among avian group a rotavirus. Virology 386, 334–343.

Sokal, R.R., Rohlf, F.J., 1995. Biometry, 3rd edition. W. H. Freeman and Company, New York.

Suzuki, Y., 2014. A possible packaging signal in the rotavirus genome. Genes Genet. Syst. 89, 81–86.

Suzuki, Y., 2016. Co-evolution in a putative bundling signal of bluetongue and epizootic hemorrhagic disease viruses. Genes Genet. Syst. 91, 283–288.

Taraporewala, Z.F., Patton, J.T., 2001. Identification and characterization of the helix-destabilizing activity of rotavirus nonstructural protein NSP2. J. Virol. 75, 4519–4527.

Tortorici, M.A., Shapiro, B.A., Patton, J.T., 2006. A base-specific recognition signal in the 5' consensus sequence of rotavirus plus-strand RNAs promotes replication of the double-stranded RNA genome segments. RNA 12, 133–146.

Trojnar, E., Otto, P., Johne, R., 2009. The first complete genome sequence of a chicken group a rotavirus indicates independent evolution of mammalian and avian strains. Virology 386, 325–333.

Ward, R.L., Nakagomi, O., Knowlton, D.R., McNeal, M.M., Nakagomi, T., Clemens, J.D., Sack, D.A., Schiff, G.M., 1990. Evidence for natural reassortants of human rotaviruses belonging to different genogroups. J. Virol. 64, 3219–3225.

Wentzel, J.F., Yuan, L., Rao, S., van Dijk, A.A., O'Neill, H.G., 2013. Consensus sequence determination and elucidation of the evolutionary history of a rotavirus Wa variant reveal a close relationship to various Wa variants derived from the original Wa strain. Infect. Genet. Evol. 20, 276–283.

Wyatt, R.G., James, W.D., Bohl, E.H., Theil, K.W., Saif, L.J., Kalica, A.R., Greenberg, H.B., Kapikian, A.Z., Chanock, R.M., 1980. Human rotavirus type 2: cultivation in vitro. Science 207, 189–191.

Yoshida, R., Nei, M., 2016. Efficiencies of the NJp, maximum likelihood, and bayesian methods of phylogenetic construction for compositional and noncompositional genes. Mol. Biol. Evol. 33, 1618–1624.