Nucleic Acids Research, 2005, Vol. 33, No. 6 ***e65***
doi:10.1093/nar/gni064

Species-independent detection of RNA virus by representational difference analysis using non-ribosomal hexanucleotides for reverse transcription

Daiji Endoh*, Tetsuya Mizutani[1], Rikio Kirisawa[2], Yoshiyuki Maki[3], Hidetoshi Saito[3], Yasuhiro Kon[4], Shigeru Morikawa[1] and Masanobu Hayashi

Laboratory of Veterinary Radiology, School of Veterinary Medicine, Rakuno Gakuen University, Ebetsu 069-8501, Japan, [1]Special Pathogens Laboratory, Department of Virology 1, National Institute of Infectious Diseases, Musashimurayama 208-0011, Japan, [2]Laboratory of Veterinary Microbiology, School of Veterinary Medicine, Rakuno Gakuen University, Ebetsu 069-8501, Japan, [3]Genosys Division, Sigma-Aldrich Japan, Ishikari 061-3241, Japan and [4]Laboratory of Anatomy, Graduate School of Veterinary Medicine, Hokkaido University, Sapporo 060-0818, Japan

Received December 10, 2004; Revised February 26, 2005; Accepted March 18, 2005

ABSTRACT

A method for the isolation of genomic fragments of RNA virus based on cDNA representational difference analysis (cDNA RDA) was developed. cDNA RDA has been applied for the subtraction of poly(A)$^+$ RNAs but not for poly(A)$^-$ RNAs, such as RNA virus genomes, owing to the vast quantity of ribosomal RNAs. We constructed primers for inefficient reverse transcription of ribosomal sequences based on the distribution analysis of hexanucleotide patterns in ribosomal RNA. The analysis revealed that distributions of hexanucleotide patterns in ribosomal RNA and virus genome were different. We constructed 96 hexanucleotides (non-ribosomal hexanucleotides) and used them as mixed primers for reverse transcription of cDNA RDA. A synchronous analysis of hexanucleotide patterns in known viral sequences showed that all the known genomic-size viral sequences include non-ribosomal hexanucleotides. In a model experiment, when non-ribosomal hexanucleotides were used as primers, *in vitro* transcribed plasmid RNA was efficiently reverse transcribed when compared with ribosomal RNA of rat cells. Using non-ribosomal primers, the cDNA fragments of severe acute respiratory syndrome coronavirus and bovine parainfluenza virus 3 were efficiently amplified by subtracting the cDNA amplicons derived from uninfected cells from those that were derived from virus-infected cells. The results suggest that cDNA RDA with non-ribosomal primers can be used for species-independent detection of viruses, including new viruses.

INTRODUCTION

Identifying the causative agent of an infectious disease is the cornerstone for its eventual control. For example, the outbreak of severe acute respiratory syndrome (SARS) was controlled after the identification of the causative agent coronavirus (SARS-CoV) (1). Developments in molecular biological approaches in recent years have led to the identification of many unknown pathogens. Once a fragment from the agent's genome has been isolated and sequenced, standard genomic walking techniques are used to extend the known sequence, and computer homology searches can then be used to identify the likely phylogenetic relationship of the agent with other known organisms (2). Additionally, sequences of some viruses, such as SARS-CoV, have altered during transmission, and this may prevent the detection of the virus by a PCR method (3,4). Thus, a detection method that is not based on the known sequence is essentially required as an alternative method to the normal PCR method.

Representational difference analysis (RDA) is one of the most reliable methods for identifying new agents since it does not require prior knowledge of the agent's class (5). The technique is based on PCR enrichment of DNA fragments that are present in agent-infected cells but absent in normal cells. Using RDA, Chang *et al*. (6) isolated two DNA fragments

*To whom correspondence should be addressed. Tel: +81 11 388 4847; Fax: +81 11 387 5890; Email: dendoh@rakuno.ac.jp

from a Kaposi's sarcoma (KS) lesion in an AIDS patient. The determination of sequences of these fragments resulted in the discovery of the KS-associated herpes virus (7). Despite the fact that RDA has been developed for detecting agents with a DNA-based genome, it can be used to detect the presence or absence of RNA in a sample by generating a cDNA intermediate (cDNA RDA) to amplify the RNA (8). Since a large quantity of ribosomal RNAs interfere with cDNA RDA, the cDNA intermediate should be synthesized from poly(A)$^+$ RNA. Therefore, it is difficult to detect RNA viruses from virus-infected cells by cDNA RDA because many viruses have no poly(A) at the end of the genome. If cDNA RDA can be applied to total RNA without interference with ribosomal RNAs, the virus genome can be amplified from total RNA of virus-infected cells by cDNA RDA.

The selection of poly(A)$^+$ RNA by using an oligo(dT) column followed by oligo(dT) priming can eliminate the influence of ribosomal RNAs on cDNA synthesis. Primers that are specific to a viral genome also efficiently eliminate the influence of ribosomal RNAs. However, prior knowledge of the virus genome is required for the construction of specific primers. In this study, based on the fact that cDNA can be primed with a mixture of oligomers, we constructed a set of oligomers that was inefficient for priming ribosomal RNAs but that normally primed most of the genome of an RNA virus (9,10). Based on the frequency distribution of hexanucleotides in ribosomal RNAs and viral sequences in current public databases, we determined a mixture of 96 hexanucleotides that rarely prime ribosomal RNAs but can prime all the known mammalian viruses listed in public databases. The results of this study show that species-independent detection of viral RNA from infected cells is possible.

MATERIALS AND METHODS

Design and synthesis of a primer mixture

A rat primary transcript, including 18S, 5.8S and 28S ribosomal RNA (V01270.1), was selected for hexanucleotide frequency analysis of ribosomal RNAs. Genomic sequences of SARS-CoV (AY291315) and bovine parainfluenza virus 3 (BPI3, NC_002161) were also selected as representatives of RNA virus (11). We created three programs, GREG, GAS and OSC (produced by C's Labs, Sapporo, Japan and licensed by Sigma–Aldrich Co. Ltd, St Louis, MO), connected to a MySQL database server (version 4.0.20). The program GREG transformed FASTA-formatted sequence to a text-formatted sequence and inserted it into a table in the MySQL database (GREG table). Using these programs, we designed a mixture of hexanucleotide primers. First, sequences were divided into hexamers, and each hexamer was classified into a pattern of hexanucleotide sequence in which $4^6 = 4096$ patterns were included. The program GAS generated frequency distributions of hexanucleotides in the sequences in the GREG table, extracted the progeny with their probabilities of hexanucleotide patterns and inserted them into a table (GAS table) of the database. We made four sets of GREG and GAS tables for rat ribosomal RNAs, a satellite repeat, BPI3 and SARS-CoV (Table 1). The program OSC listed the differences between the two GAS tables, i.e. the program selected oligomer patterns that exist in BPI3 but not in ribosomal RNAs.

Table 1. Probabilities of hexanucleotide patterns in ribosomal RNAs (V01270), a satellite-repeat (V00125), SARS coronavirus (SARS-CoV, AY291315) and bovine parainfluenza virus 3 (BPI3, NC_002161)

Pattern	Probability ($\times 10^{-3}$)			
	Ribosomal RNA	Satellite repeat	BPI3	SARS-CoV
TCTCTC	13.54	1.28	0.25	0.11
AGAGAG	7.53	0.64	0.63	0.16
GAGAGA	7.11	1.92	0.25	0.11
CTCTCT	6.85	0.64	0.33	0.11
TCTGTC	5.67	0.00	0.28	0.25
CTTTCT	4.68	0.32	0.43	0.67
TCTTTC	4.55	0.00	0.46	0.52
TCTCTG	3.55	0.00	0.48	0.21
GTCTCT	3.43	0.64	0.28	0.21
TGTTAA	0.01	0.00	0.43	0.96
GGTCTA	0.01	0.32	0.15	0.16
ATATAT	0.00	0.00	0.96	0.13
GTGCAC	0.00	0.00	0.00	0.27
TAGTAT	0.00	0.00	0.38	0.16
GATATC	0.00	0.00	0.25	0.13
ATACTA	0.00	0.00	0.30	0.28
TATAGT	0.00	0.00	0.35	0.17
TATATA	0.00	0.00	0.81	0.05
ACTATA	0.00	0.00	0.61	0.31

The hexanucleotide patterns are aligned according to the probabilities in ribosomal RNAs. The 10 highest and lowest frequent patterns are listed in the table. A full version of the table, including all patterns, is supplemented online.

The frequency of hexamer patterns in each RNA sequence was transferred into a table of Microsoft Access. A total of 96 hexamer patterns, including those having very low frequencies or those that did not appear in ribosomal RNAs, were synthesized and mixed for use as an RT primer (Table 2, non-ribosomal hexanucleotides).

Database for mammalian viral genomic data

To estimate the frequencies of priming sites with the selected hexanucleotides, we prepared a MySQL table that included sequence data of reported viral genomes as follows. First, we downloaded all viral sequences from the FTP site of EMBL database (release date 30 June 2004) and entered them into tables (EMBL data table) of the MySQL database. The table included EMBL ID, title, annotation and sequence as fields. The annotation field included taxonomic classification of the origin of the data. We separated the words included in the annotation field into taxonomic words, such as family names, and inserted them into a new table (taxonomic table), in which EMBL IDs and taxonomic words were included. We then selected EMBL IDs of mammalian viruses with any of the viral family names listed in Table 5 from the taxonomic table. Next, sequences of mammalian viruses were divided into groups according to their species presented in the taxonomic table. To determine the targets for hexanucleotide analysis, EMBL IDs having the longest sequence in the species were selected as the estimated genomic sequence of the virus, and then, a new table named 'Sequences of viral species' was prepared. Although the longest sequence from each species was selected, some of these sequences were very short. Therefore, we eliminated sequences that were shorter than half of the common genomic size of each viral family. The resultant 1791 viral sequences were inserted into a table titled 'Genomic sequence of viral species'. The frequency of the non-ribosomal hexanucleotides was determined in sense and complementary

Table 2. Hexanucleotide patterns of non-ribosomal hexanucleotides

Motif	Motif	Motif	Motif	Motif	Motif	Motif	Motif
GATATC	GATACT	CGATAT	ACTACT	ATAGTC	CTTAGT	ACTAAG	AACTTA
TAGTAT	CGTATA	GTATAC	TAACGA	CTAGTA	CTTACA	GCATAC	ATAACG
TATAGT	GTATAG	AATCCA	CGACTA	GTACTA	TTATGC	CAATAT	ATGTTA
TATATA	CGGTTA	TAGCAC	TACTAG	TAAGTT	ATACGC	ACCGTA	TGGTAT
ATACTA	AATAGT	ATATCG	AGTAGT	ATATCC	CGCTTA	GTGCTA	TGCGTA
ATATAT	CGCATA	AATATT	GTTAAC	TCGATA	TAACGC	ACGCTA	GGATAT
GTGCAC	ATTACG	TATAGC	GTCTAC	GTACCA	GGTCAT	ATGTCG	CATAGC
ACTATA	TTAACA	CTTGTA	TACAAG	GTATCA	CTCATA	AGCTTA	CATACT
CGTAAT	AGTATC	TAGTCG	TACCAG	ATACTC	AATTTG	CGACAT	CGGATA
CTATAC	TGTTAA	GTAGAC	TGGATT	ACATTA	CTGGTA	GCTATA	TTACTA
TATACG	ACTATT	CTATAG	TCGTTA	ATATTG	TTCATG	GCTATG	ACTCGT
TATGCG	TAACCG	TAGCTA	ATAGTA	CGTCTA	GCGATA	TGTAAG	TAAGGT

strands of the sequences included in the table titled 'Genomic sequence of viral species'.

Culture of virus-infected cells and RNA extraction

The strains SARS-CoV and BPI3 were Frankfurt1 and BN-1, respectively (12–14). SARS-CoV and BPI3 were propagated by serial infection of Vero E6 and MDBK cells, respectively (15).

Vero E6 cells were routinely subcultured in 75 cm^2 flasks in DMEM (Sigma–Aldrich) supplemented with 0.2 mM/ml L-glutamine, 100 U/ml penicillin, 10 μg/ml streptomycin and 5% (v/v) fetal bovine serum (FBS) and maintained at 37°C in an atmosphere of 5% CO_2. For experimental use, the cells were split once in 25 cm^2 flasks and cultured until they reached 100% confluence. Prior to the virus infection, the culture medium was replaced with 2% FBS containing DMEM. SARS-CoV, which was isolated as Frankfurt1 and kindly provided by Dr J. Ziebuhr (16), was used in the present study. The viral infection was established in the cells with a multiplicity of infection (m.o.i) of 10. The infection of cells with SARS-CoV virus and the subsequent treatment of SARS-CoV RNA were restricted in the P4 area in the National Institute of Infectious Disease, and the work with SARS-CoV was performed in accordance with the rules for infectious pathogens that have been notified by the National Institute of Infectious Disease.

MDBK cells were maintained in Eagle's minimum essential medium (Sigma–Aldrich) supplemented with 5% FBS in a humidified atmosphere of 5% CO_2 at 37°C. The cells were infected with BPI3 at an m.o.i of 0.1. The work on infection of BPI3 was performed in accordance with the rules for pathogens that have been notified by Rakuno Gakuen University.

Extraction and *in vitro* synthesis of RNA

Total RNA was isolated using Trizol (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. Prior to cDNA synthesis, contaminated genomic DNA in the extracted RNA was digested with RNase-free DNase I (Promega, Madison, WI) at 37°C for 1 h. RNA was extracted serially with phenol and chloroform, precipitated with ethanol according to the standard protocol and subsequently used as a control RNA.

For the synthesis of a model RNA, the entire molecule of pCIneo plasmid was transcribed *in vitro* from a T7 promoter. The synthesized 5.4 kb RNA was treated with RNase-free DNase I (Promega), extracted with phenol/chloroform, precipitated with ethanol and subsequently used as a test RNA. After quantitation, the test and control RNAs were mixed to estimate the sensitivity of cDNA RDA in various conditions.

cDNA RDA

First-strand cDNA was synthesized from the mixed RNA with non-ribosomal hexanucleotides by using a double-stranded cDNA synthesis kit (Invitrogen) according to the manufacturer's protocol, i.e. the total RNA was diluted to 1 μg per μl and mixed with dNTPs, the non-ribosomal hexanucleotides, 5× reaction buffer, 0.1 M DTT and an RNase inhibitor. Reverse transcriptase (Superscript II, Invitrogen) was added, and the mixture was incubated at 50°C for 60 min. Second-strand cDNA was synthesized with *Escherichia coli* DNA polymerase (Invitrogen), *E.coli* DNA ligase (Invitrogen) and RNaseH (Invitrogen) at 16°C for 2 h. Double-stranded cDNA was digested with Dpn II, and the resultant fragments were extracted from the digest by using a silicon-membrane-based purification kit (Gene Elute Purification Kit; Sigma–Aldrich).

Linker-derived amplification of DNA fragments and selective amplification steps of cDNA RDA were performed according to the method described by Hubank and Schatz (17). Briefly, 0.1 μg of Dpn II-digested double-stranded cDNA was ligated with RBam24 and RBam12 linkers (5). An aliquot of 1 μl of the ligation solution was diluted with *Taq* mixture (10 μl of 10× *Taq* buffer, including 15 mM $MgCl_2$, and 0.2 mM each of dNTPs). The mixture was preheated to 72°C, and then, *Taq* polymerase (Promega) was added and the mixture was incubated at 72°C for 5 min to synthesize a complementary strand against the overhanging region of RBam24. This was immediately followed by a denaturation step (94°C for 2 min) and 20 cycles of PCR (94°C for 1 min and 72°C for 8 min) to non-specifically amplify 200–800 bp Dpn II-digested cDNA fragments with linkers (amplicons). After amplification, amplicons were redigested with Dpn II and purified with a silicon-membrane-based purification kit to eliminate the spliced linkers. Some amplicons, including the test RNA, were religated with JBam24 and JBam12 linkers. Amplicons with the second linkers were mixed with a large quantity of amplicons without the test RNA sequences.

The mixture was precipitated with ethanol and 3 M sodium acetate, dissolved in 4 µl of 3× EE buffer {30 mM EPPS [*N*-(2-hydroxyethyl)piperazine-*N*′-3-propanesulfonic acid], 3 mM EDTA, pH 8.0} and covered with mineral oil. After the mixture was heated to 99°C for 4 min, 1 µl of 5 M NaCl was added, and the solution was incubated at 67°C for 21 h. During this incubation period, amplicons from normal cellular RNA with a linker included in the test amplicons were hybridized with those from uninfected cells without a linker. After hybridization, the reaction mixture was diluted to 100 µl with reaction buffer of *Taq* DNA polymerase (Promega) as described above. Amplicons having the linker sequence at both ends were amplified by a denaturation step (94°C for 2 min) and 20 cycles of PCR (94°C for 1 min and 72°C for 3 min).

RNAs extracted from virus-infected cells were subjected to cDNA RDA as described above, by using amplicons synthesized from RNA obtained from uninfected cells.

Southern blot hybridization

cDNA RDA-derived fragments in quantities ranging from 100 ng to 2 µg were separated on agarose gels, blotted onto a Biodyne nylon membrane (Pall Co. Ltd, Port Washington, NY) by capillary transfer in 20× SSC (3 M sodium chloride and 0.3 M sodium citrate) for 16 h and fixed to the membrane by baking in an oven at 80°C for 30 min. pCIneo was used as a probe for the *in vitro* synthesized RNA. A random primer labelling kit (BcaBest Labeling kit, Takara Shuzo Co. Ltd, Kyoto, Japan) was used for labelling with ^{32}P and hybridized in SuperHybPlus hybridization solution (Sigma–Aldrich) according to the manufacturers' protocol.

Although amplicons are cloned into a vector to identify its sequence, according to 'Cartagena Protocol on Biosafety Text of the Protocol', amplicons from SARS-CoV-infected cells should not be cloned into any plasmid. Therefore, by performing hybridization, we identified that the sequences of amplicons derived from SARS-CoV-infected cells were identical to SARS-CoV. To determine probes for SARS-CoV, the PCR products predicted from the genomic sequence and sizes of cDNA RDA products were amplified from the SARS-CoV genome. These PCR products were purified, labelled with DIG and independently hybridized using a slit of the nylon membrane blotted with amplicons derived from SARS-CoV-infected cells. DIG labelling and hybridization were carried out according to the manufacturer's instructions provided with the DIG-hybridization kit (F. Hoffmann-La Roche Ltd, Diagnostics Division, Basel, Switzerland). Hybridization was carried out with DIG-labelled DNA probes at 65°C for 16 h in DIG-hybridization solution.

Cloning and sequence analysis

The cDNA RDA-derived fragments were respliced with Dpn II, separated on agarose gels, extracted with a silicon-membrane-based purification kit and cloned into pSPORT1 (Invitrogen). Plasmids that included cDNA RDA-derived fragments were selected by colony PCR and purified with a silicon-membrane-based purification kit. Three clones of the plasmid were sequenced along with the fragment to detect PCR errors. Sequences of the cDNA RDA-derived fragments were determined using a DYEnamic ET-terminator kit (Amersham Biosciences Corp., Piscataway, NJ) and an ABI Prism 310 sequencer (PerkinElmer Life and Analytical Sciences, Inc., Boston, MA) with M13 forward and reverse primers (Takara Shuzo). We used the BLAST program on NCBI to determine sequences that were homologous to the isolated fragments.

RESULTS

Frequency distribution of hexanucleotides within the ribosomal sequences

To predict the major RNA molecule in cellular RNA, we synthesized cDNA from RNA extracted from normal bovine cells, synthesized amplicons by using random primers, spliced them with Dpn II and finally subcloned them into pSPORT1. Among the sequences of 30 selected clones, the sequences of 25 clones and 5 clones were highly homologous to those of ribosomal RNA and a 1399 bp satellite repeat (GenBank accession no. V00125), respectively. Based on this result, ribosomal and satellite sequences were determined for analysing hexanucleotide frequency. Since the reported ribosomal sequences of mammals are highly homologous to each other, we selected rat premature 18S, 5.8S and 28S ribosomal sequences as ribosomal sequences for hexanucleotide frequency analysis. Genome sequences of SARS-CoV and BPI3 were selected as representatives of RNA viral sequence. Table 1 shows the probabilities [number of the patterns/(length of sequence × 2)] of hexanucleotides in the human ribosomal RNA, V00125 satellite repeat, BPI3 and SARS-CoV. If a random sequence is assumed, frequencies would be distributed according to the Poisson's distribution, and the average of frequency would be the same as the variance of frequency. Although the average/variance ratios of all the four sequences were <1, the ratio of ribosomal RNA was smallest in these sequences (Table 3). This value suggests that the probabilities of hexamer patterns of ribosomal RNA were strongly biased from random sequence. Based on histograms of probabilities, the distribution of the probabilities of hexamer patterns in ribosomal RNA differed greatly from those of V00125 satellite repeat or BPI3 (Figure 1). It should be noted that 8 hexanucleotides did not exist and over 90 hexanucleotides were rare in the ribosomal sequence. To determine primer sequences that do not prime ribosomal RNA but prime viral RNA, the probabilities of hexamer patterns in ribosomal RNA and satellite repeat were calculated. The hexamer patterns were then realigned in an ascending order according to the sum of probabilities, and the 1st to 96th patterns were selected as non-ribosomal hexanucleotides, i.e. we selected 96 rarest hexanucleotide patterns (Table 2, non-ribosomal hexanucleotides) in major transcripts in normal mammalian cells on the assumption that ribosomal RNA and transcripts from satellites are the most frequent transcripts in normal cells.

Table 3. Average and variances of frequencies of hexanucleotide patterns in ribosomal RNAs, a satellite repeat (V00125), bovine parainfluenza virus 3 (BPI3) and SARS coronavirus (SARS-CoV)

	Ribosomal RNAs	Satellite repeat	BPI3	SARS-CoV
Average	26.8	1.9	10.6	18.6
Variance	2156.5	2.5	96.0	188.1
Average/variance	0.012	0.76	0.11	0.099

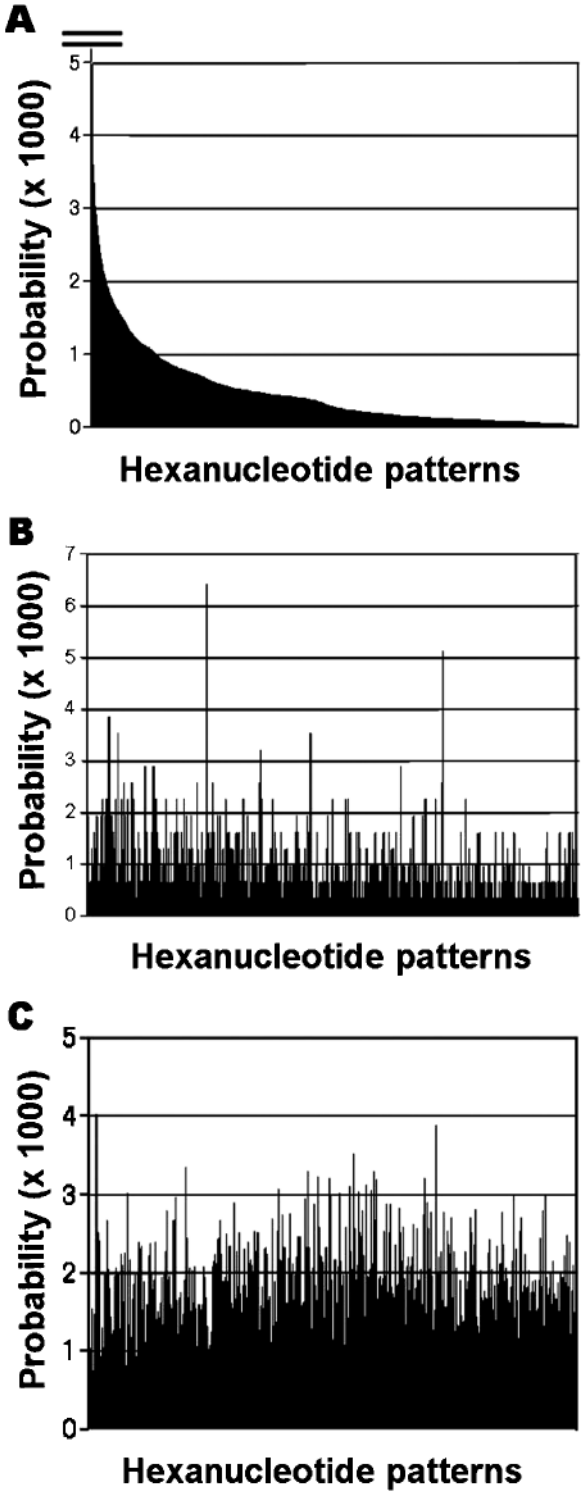

Figure 1. Histogram of probabilities of hexanucleotide patterns in ribosomal RNAs (**A**), a satellite repeat (**B**), BPI3 and SARS-CoV (**C**). From each sequence, the frequency was determined for each hexanucleotide pattern by using programs described in Materials and Methods. Probability was calculated as (frequency of a motif/total length of sequence). For simple comparison, the probabilities of ribosome RNAs and satellite repeat or SARS-CoV and BPI3 were added for each hexanucleotide pattern. Hexanucleotide patterns were aligned according to their probability in the ribosomal RNAs. When the probabilities of hexanucleotide patterns were larger than 0.005, the data were omitted.

The rarity of non-ribosomal hexanucleotides in comparison with non-V00125-satellite and random hexamers was confirmed in reported sequences of mammalian ribosomal RNAs (Table 4). The probabilities were relatively low in all the mammalian ribosomal sequences reported in GenBank. It is considered that non-ribosomal hexanucleotides inefficiently primed ribosomal RNAs in reverse transcription. On the other hand, the probabilities of non-ribosomal, non-V00125-satellite or random hexamers were different among the satellite sequences.

Frequencies of non-ribosomal hexanucleotides in viral genomic sequences

In addition to the inefficiency to prime ribosomal RNAs, efficient priming of V00125-satellite-repeat and BPI3 with non-ribosomal hexanucleotides is shown in Figure 1. To predict priming efficiency in many viruses, the probability of non-ribosomal hexanucleotides in known viral genomes was estimated. In 1791 viral sequences in 'Genomic sequences of viral species', the median probabilities of non-ribosomal hexanucleotides of all known viral sequences is $13.2–37.6 \times 10^{-3}$ (Table 5). When the average probabilities were calculated in viral families, the minimum probability was 3.7×10^{-3} in Herpesviridae and the maximum probability was 44.8×10^{-3} in Poxviridae (Table 5). These median values of probabilities in viral genomes were greater than those of ribosomal RNAs (Table 4) and comparable with those of non-V00125 and random hexamers (Table 5). These data suggest that non-ribosomal hexanucleotides prime cDNA synthesis in most viruses.

Model experiment for differentiated amplification of viral RNA by cDNA RDA

The database analysis suggests that viral RNAs were efficiently primed by non-ribosomal hexanucleotides in comparison with ribosomal RNAs. We investigated the effect of non-ribosomal hexanucleotides on cDNA synthesis by using *in vitro* synthesized plasmid RNA (artificial RNA) and total cellular RNA, including ribosomal RNAs. An autoradiogram of ^{32}P-labelled double-stranded cDNAs that were synthesized using non-ribosomal hexanucleotides or a random primer and subsequently separated by agarose gel electrophoresis is shown in Figure 2. When test and total cellular RNAs were reverse transcribed individually, the efficiencies of cDNA synthesis from artificial RNA using random and non-ribosomal hexanucleotides were almost similar; however, the efficiency of cDNA synthesis of cellular RNA was markedly lower in a non-ribosomal hexanucleotide-primed cDNA sample than in a random primer-primed cDNA sample (Figure 2A). In the mixed samples of test and total cellular RNAs, ribosomal RNAs were inefficiently reverse transcribed with non-ribosomal hexanucleotides. The total incorporated counts in non-ribosomal hexanucleotides-primed cDNAs decreased with a decrease in the proportion of the test RNA in the mixed RNA samples (Figure 2A). It is considered that the efficiency of reverse transcription depends on the proportions of the test RNA in the mixed RNAs. When approximately the same counts of cDNAs were loaded on the gel, synthesis of ribosomal RNA-derived cDNA (ribosomal cDNA) was obvious in random primer-primed cDNAs (Figure 2B).

Table 4. Total frequencies and probabilities of sense and complementary sequences of non-ribosomal, non-V00125 and random hexanucleotides in reported sequences of mammalian ribosomal RNAs and satellite repeat

Accession	Molecule	Species	Sequence length	Non-ribosomal Frequencies	Probabilities ($\times 10^{-3}$)	Non-V00125 Frequencies	Probabilities ($\times 10^{-3}$)	Random Frequencies	Probabilities ($\times 10^{-3}$)
X00686	18S rRNA	*Mus musculus*	1869	27	7.22	101	27.02	85.4	22.85
X82564	45S pre rRNA gene	*M.musculus*	22 118	281	6.35	897	20.28	974.8	22.04
M11188	18S rRNA gene. Complete	*Rattus norvegicus*	1920	27	7.03	107	27.86	87	22.66
M10098	18S rRNA gene. Complete	*Homo sapiens*	1969	27	6.86	108	27.43	89.6	22.75
U13369	Complete repeating unit	*H.sapiens*	42 999	203	2.36	1569	18.24	1743.4	20.27
M27830	28S rRNA gene. Complete	*H.sapiens*	1955	9	2.30	53	13.55	82.4	21.07
X14345	5′ external transcribed spacer of pre-ribosomal RNA	*H.sapiens*	3627	2	0.28	108	14.89	176.6	24.35
AY265350	18S rRNA gene. Complete	*S.scrofa*	2302	25	5.43	114	24.76	105.4	22.89
V01270	18S 5.8S and 28S rRNA	*R.norvegicus*	8647	71	4.11	351	20.30	405	23.42
X01117	18S rRNA sequence	*R.norvegicus*	1874	27	7.20	100	26.68	84.8	22.63
AJ311674	Partial 18S rRNA gene	*Dasypus novemcinctus*	1824	27	7.40	99	27.14	83.2	22.81
AJ311675	Partial 18S rRNA gene	*Erinaceus europaeus*	1825	27	7.40	101	27.67	82.4	22.58
AJ311673	Partial 18S rRNA gene	*Equus caballus*	1824	25	6.85	100	27.41	82.8	22.70
X06778	18S rRNA	*Oryctolagus cuniculus*	1863	30	8.05	103	27.64	84.6	22.71
AJ311678	Partial 18S rRNA gene	*Vombatus ursinus*	1845	28	7.59	101	27.37	81.8	22.17
AJ311676	Partial 18S rRNA gene	*Monodelphis domestica*	1847	28	7.58	100	27.07	81.6	22.09
AJ311677	Partial 18S rRNA gene	*Didelphis virginiana*	1847	28	7.58	100	27.07	81.6	22.09
AJ311679	Partial 18S rRNA gene	*Ornithorhynchus anatinus*	1850	28	7.57	102	27.57	82.8	22.38
V00125	Repeated unit of bovine 1.715 satellite DNA	*Bos taurus*	1399	0	0.00	0	0.00	66.6	23.80
J00036	Thymus satellite I	*B.taurus*	1402	3	1.07	5	1.78	64.2	22.90
U59381	Tetranucleotide microsatellite repeat	*H.sapiens*	1032	2	0.97	29	14.05	35	16.96
AY153482	Microsatellite PDE6B sequence	*S.scrofa*	1895	20	5.28	64	16.89	82.6	21.79
AF298194	Clone pW-1 microsatellite III	*H.sapiens*	1524	17	5.58	31	10.17	71	23.29
AY339973	Clone 1 satellite sequence	*Canis familiaris*	1200	17	7.08	41	17.08	52.6	21.92
U53349	Chromosome 10 microsatellite	*R.norvegicus*	1070	21	9.81	32	14.95	36.2	16.92
AJ295050	Alpha satellite 5CEN-P7	*H.sapiens*	1382	31	11.22	61	22.07	58.8	21.27
AB023433	Satellite sequence	*R.norvegicus*	2845	86	15.11	136	23.90	125	21.97
AY145450	Microsatellites D6Wum35 and D6Wum34	*M.musculus*	1704	71	20.83	103	30.22	73.4	21.54
AF181667	Polymorphic microsatellite sequence	*H.sapiens*	3650	201	27.53	213	29.18	164.4	22.52
AF259760	Microsatellite MNS-87 and MNS-88	*Ovis aries*	1074	86	40.04	38	17.69	48.2	22.44
U10629	Chromosome 4 satellite	*H.sapiens*	1060	116	54.72	48	22.64	55.4	26.13

Sequences of the non-V00125 and random hexamers are supplemented online. Probabilities of the hexanucleotide-sets in 2464 reported satellite repeat are also supplemented online.

Table 5. Number of species and maximum, minimum and median probabilities [median (minimum − maximum)] ($\times 10^{-3}$) of non-ribosomal, non-V00125-satellite-repeat and random hexamers hexanucleotides in known viral species

Family	Number of species	Non-ribosomal	Non-V00125	Random hexamers
Adenoviridae	25	15.3 (6.5–28.3)	25.0 (19.4–47.1)	23.8 (22.8–47.6)
Arenaviridae	8	17.0 (13.1–19.2)	26.0 (23.3–27.6)	24.5 (23.8–25.1)
Astroviridae	7	17.4 (14.5–19.2)	23.9 (22.4–25.6)	24.3 (23.5–25.2)
Bornaviridae	1	20.0 (20.0–20.0)	21.8 (21.8–21.8)	23.7 (23.7–23.7)
Bunyaviridae	26	22.8 (11.8–30.7)	26.2 (21.8–29.7)	23.7 (23.1–25.3)
Caliciviridae	23	13.2 (7.1–19.3)	23.5 (19.1–26.5)	24.1 (23.4–25.4)
Coronaviridae	51	24.2 (23.9–31.4)	25.3 (25.2–29.9)	23.6 (23.5–24.3)
Filoviridae	3	21.9 (20.8–22.6)	24.9 (24.4–26.2)	25.0 (24.5–25.1)
Flaviviridae	53	14.1 (8.6–23.4)	22.4 (19.0–25.9)	24.0 (21.6–25.2)
Hepadnaviridae	14	16.6 (9.7–27.2)	21.6 (18.0–25.3)	24.0 (21.3–24.5)
Herpesviridae	31	15.0 (3.7–29.2)	22.6 (16.4–27.2)	23.5 (22.8–24.5)
Iridoviridae	3	17.7 (14.1–25.2)	23.3 (20.8–29.3)	23.5 (21.9–23.6)
Orthomyxoviridae	1029	17.6 (6.3–29.7)	24.3 (16.2–32.9)	24.3 (20.0–27.9)
Papillomaviridae	123	24.7 (7.7–34.3)	24.9 (18.7–28.9)	23.2 (21.1–25.0)
Parvoviridae	59	20.5 (6.3–35.9)	25.0 (19.5–31.2)	23.8 (21.6–25.1)
Picornaviridae	90	19.3 (7.1–30.6)	24.3 (19.0–29.0)	24.0 (22.6–25.1)
Polyomaviridae	16	18.4 (14.0–28.5)	22.0 (20.6–26.4)	23.6 (22.3–24.7)
Poxviridae	21	37.6 (8.2–44.8)	26.9 (22.7–32.0)	22.9 (21.3–24.0)
Reoviridae	89	28.4 (12.9–38.4)	26.2 (19.2–31.5)	23.8 (20.7–25.8)
Retroviridae	86	17.0 (9.9–27.6)	22.1 (18.2–26.8)	23.3 (21.6–24.6)
Rhabdoviridae	14	16.8 (9.2–25.3)	22.7 (19.0–26.3)	23.6 (23.1–24.2)
Togaviridae	19	20.0 (7.6–24.3)	23.3 (20.1–25.1)	23.7 (22.7–24.5)

Sequences for species were selected according to their length as described in Materials and Methods. The total probability in each species was calculated as the summation of probabilities of all non-ribosomal hexanucleotides in sense and complementary sequences of a viral genomic sequence that was selected as described in Materials and Methods. The probabilities of non-ribosomal hexanucleotides in 1791 virus genomic sequences are supplemented online.

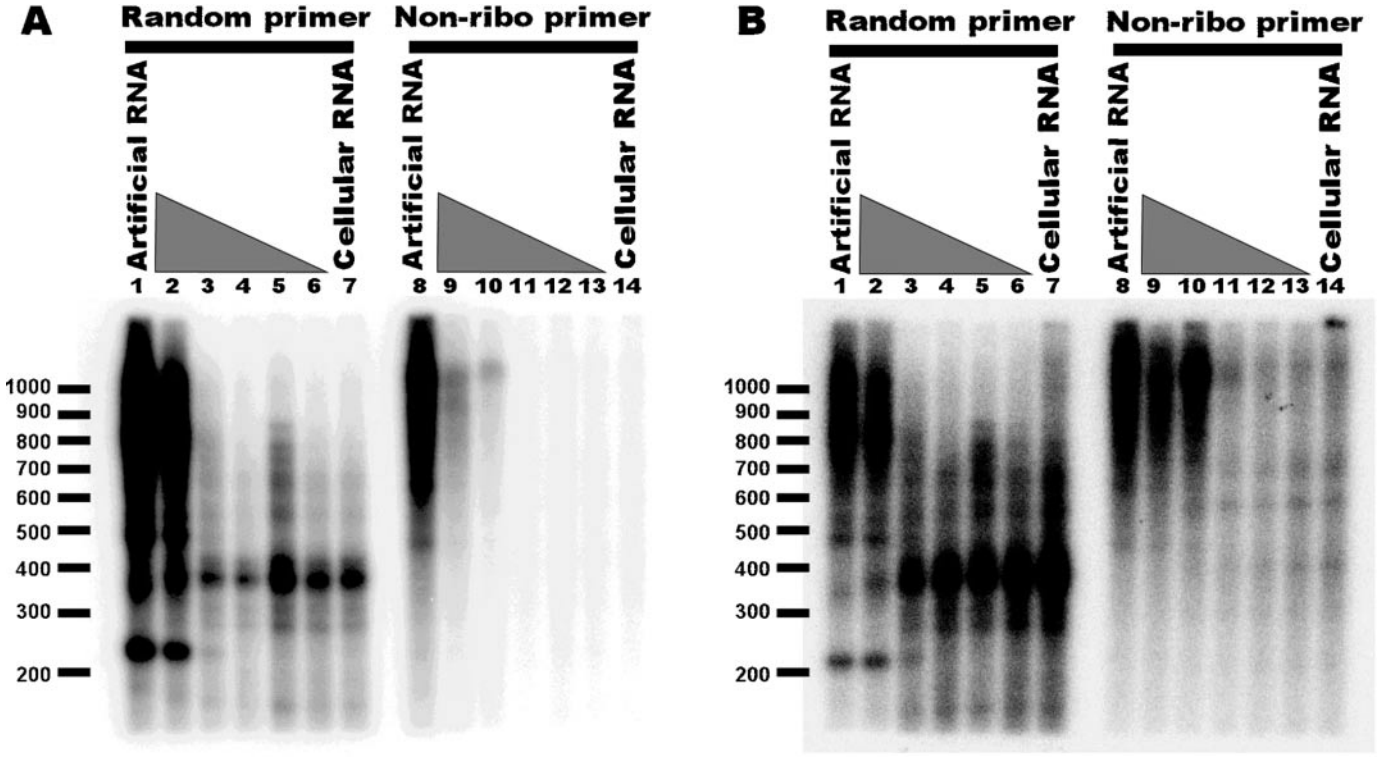

Figure 2. Autoradiogram of ^{32}P-labelled double-stranded cDNA synthesized from mixtures consisting of artificial RNA and total cellular RNA. *In vitro* transcribed RNA was synthesized from pCIneo plasmid and mixed with total cellular RNA extracted from rat2 cells in weight proportions 1:0 (lanes 1 and 8), 1:1 (lanes 2 and 9), 1:10 (lanes 3 and 10), 1:100 (lanes 4 and 11), 1:300 (lanes 5 and 12), 1:1000 (lanes 6 and 13) and 0:1 (lanes 7 and 14). One microgram of mixed RNA was reverse transcribed using random (lanes 1–7) or non-ribosomal (lanes 8–14) hexanucleotides and a second-strand cDNA was then synthesized with RNaseH, DNA polymerase and DNA ligase according to the method described in Materials and Methods. One-tenth of the volume of synthesized cDNAs was loaded on agarose gel (**A**). Loaded volumes were corrected to include the same amounts of ^{32}P in each sample (**B**). Positions and sizes (bp) of markers are present on the left.

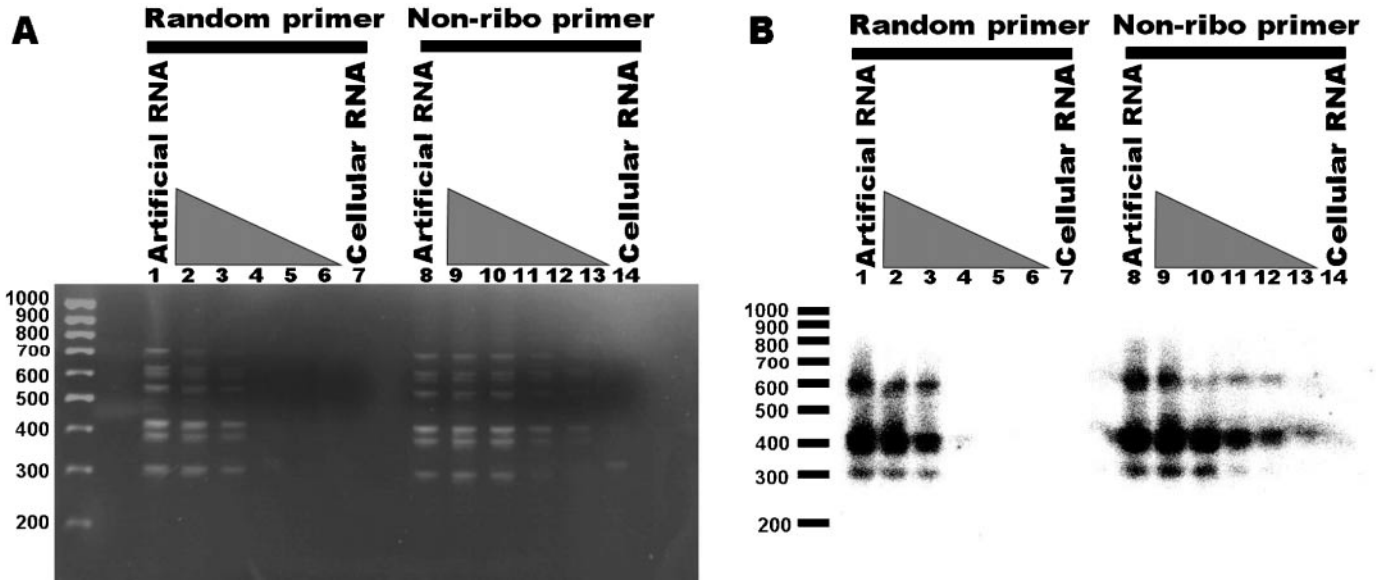

Figure 3. Agarose gel electrophoresis of RDA products (**A**) and its hybridized autoradiogram (**B**). *In vitro* transcribed RNA and total cellular RNA were mixed as described in the legend to Figure 2. Double-stranded cDNAs were synthesized and subjected to RDA as described in Materials and Methods. One-twentieth of the volume of the amplified products was separated on 3% agarose gel and stained with ethidium bromide (**A**), blotted onto a nylon membrane and hybridized with ^{32}P-labelled pCIneo (**B**). Positions and sizes (bp) of markers are present on the left.

On the other hand, the synthesis of ribosomal cDNA was not obvious in the cDNA primed with non-ribosomal hexanucleotides. A relatively large cDNA derived from artificial RNA could be observed even when a smaller proportion of test RNA was mixed with cellular RNA and non-ribosomal hexanucleotide-primed samples. These data suggest that the relative amount of test RNA-derived cDNA is greater in non-ribosomal hexanucleotides-primed samples than in random primer-primed samples.

After the first round of cDNA RDA, amplified fragments were observed on agarose by staining with ethidium bromide. The bands that corresponded to artificial RNA could be observed in 1:0, 1:1 and 1:10 (test:total cellular RNAs) mixtures when cDNAs were primed with a random primer. On the other hand, these bands could be observed in 1:0 to 1:300 RNA mixtures when cDNA was primed with non-ribosomal hexanucleotides (Figure 3A). This amplification of test RNA was confirmed by hybridization with pCIneo, which was a template for *in vitro* RNA synthesis (Figure 3B). The hybridized bands were observed even in a lane corresponding to 1:1000 RNA mixture. When cDNA was primed with a random primer, the hybridized bands could not be observed in 1:100, 1:300 and 1:1000 RNA mixtures. These data suggest that the lower limit of the test RNA amplification decreased at least 30 times when non-ribosomal hexanucleotides were used for reverse transcription when compared with the data obtained by using a random primer.

Detection of BPI3 and SARS-CoV sequences from infected cell RNA

To amplify virus sequence from infected cells, we subtracted amplicons derived from uninfected cells from those derived from virus-infected cells. Amplicons with linkers derived from the infected cells were mixed with amplicons without linkers

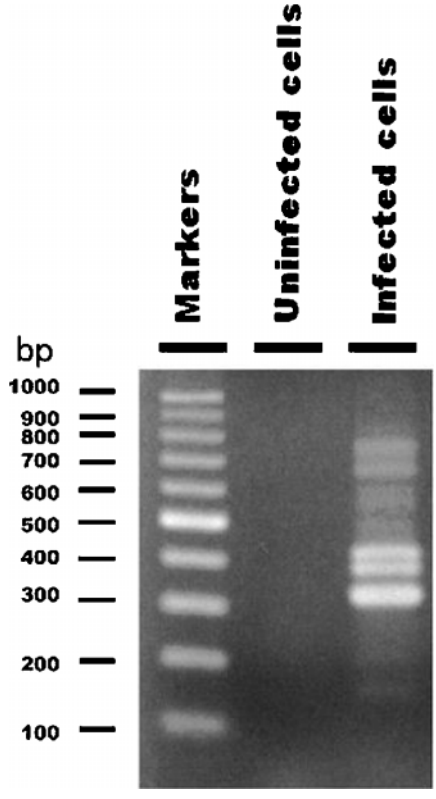

Figure 4. Agarose gel electrophoresis of RDA products from RNA extracted from bovine parainfluenza virus 3-infected cells. Double-stranded cDNA was synthesized from RNA of bovine parainfluenza virus 3-infected MDBK cells and subjected to RDA. Mock-infected cells were used for the synthesis of driver amplicons for RDA. One-twentieth of the volume of the amplified products was separated on 3% agarose gel and stained with ethidium bromide. RDA product from the uninfected control cells was used as a negative control.

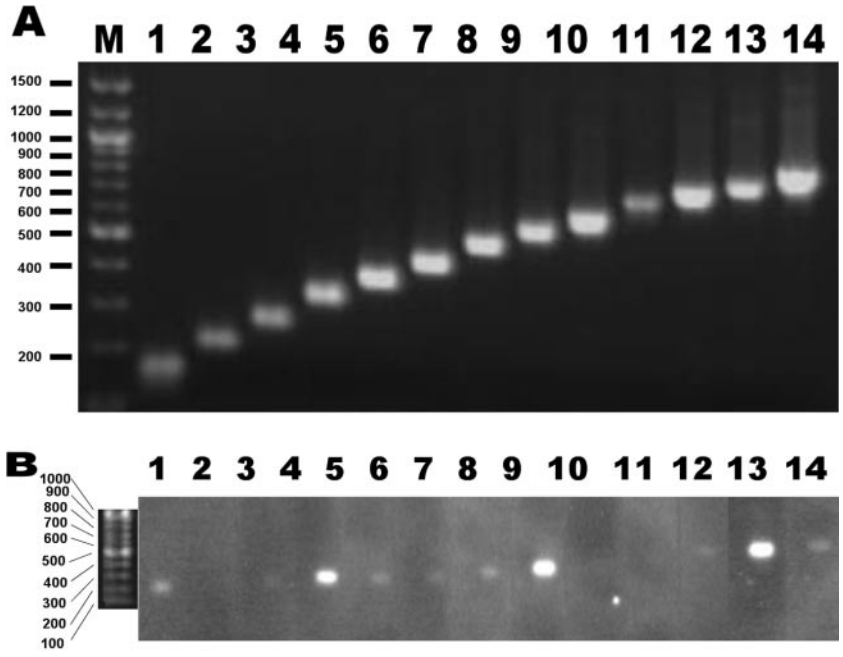

Figure 5. Agarose gel electrophoresis of RDA products with PCR products used for probes for hybridization (**A**) and a hybridized fluorogram (**B**). RNA was extracted from SARS-CoV-infected cells and subjected to RDA according to the method described in Materials and Methods. Mock-infected cells were used for the synthesis of driver amplicons for RDA. One-twentieth of the volume of the amplified products was separated on 3% agarose gels and blotted on a Nylon membrane. The membrane was then cut into slits that contained the lane showing the presence of DNA. On the other hand, the PCR fragments predicted to be amplified in the RDA reaction were amplified and subsequently ascertained by agarose gel electrophoresis (A). The amplified genomic fragments of SARS-CoV were Dig-labelled and used as probes for hybridization to each slit of the Nylon membrane containing the RDA product. Hybridization was performed in separate hybridization bags. After washing with 1× SSC and 0.1% SDS solution, the hybridized probes were detected on a fluorogram (B). Positions and sizes (bp) of markers are present on the left.

derived from uninfected cells at the ratio of 1:100 before a hybridization step of cDNA RDA. At the end of the first round of cDNA RDA, the ladders of amplified fragments were generated from BPI3-infected cells when non-ribosomal hexanucleotides were used (Figure 4). No cDNA RDA-derived bands were obvious when a random primer was used for reverse transcription (data not shown). Amplified fragments from cDNA RDA of BPI3-infected cells were cloned into pSPORT1 plasmid, and the sequences were determined. The sequences of all the cloned fragments from cDNA RDA were identical to the sequence of the BPI3 genome. These data suggest that cDNA RDA with non-ribosomal hexanucleotides enables the identification of the BPI3 genome sequence from infected cells.

Similar to BPI3, cDNA fragments derived from SARS-CoV were also amplified from SARS-CoV-infected cells (Figure 5). Viral origin of the amplified fragments was confirmed by hybridization (Figure 5B) and PCR amplification by SARS-CoV-specific primers (Figure 5A). These results indicate that genomic fragments of SARS-CoV can also be isolated by this method.

DISCUSSION

Reduction of influence of ribosomal RNA on cDNA synthesis by non-ribosomal hexanucleotides

It is well known that 3–5% and 30% of cellular RNA are estimated to be messenger and ribosomal RNA, respectively. The frequency of ribosomal RNA has been estimated by competitive PCR and real-time PCR to be 10 000 copies per cell (18). The amount of ribosomal RNA has been reported to be 1000-fold greater than that of frequently transcribed mRNA, such as beta actin and G6PDH. Thus, the repertoire of cDNA would be strongly affected by ribosomal RNA. This influence of ribosomal RNA has been avoided by oligo(dT) selection. Alternative strategy for the elimination of ribosomal RNA from total RNA, however, has not been developed. Therefore, there have been only a few applications of cDNA RDA for non-poly(A) RNA, such as those in viral genomes. In this study, we developed a new strategy for the elimination of ribosomal RNA in cDNA RDA through the construction of a hexanucleotide mixture and demonstrated its efficiency in cDNA RDA.

The main purpose of PCR is specific amplification of a gene. However, methods using multiple primers for simultaneous gene amplification have been recently employed for DNA chip methods. In order to simultaneously detect multiple genes by using a DNA chip, mixed oligonucleotides have been used as a primer for reverse transcription. This usage of mixed primers was based on the assumption that the specificity of primers was equal to the summation of the specificity of each primer. Thus, we searched for primers that do not prime ribosomal RNAs by frequency analysis of hexanucleotides. The frequency and distribution of oligonucleotides in mammalian and viral genomes have been studied to search for common motifs that might be used for controlling cellular functions (19–23). Volinia *et al*. (22) found sets of common decamers that can be used for the control of transcription control signals or for the common amplification of viruses. Programs for frequency

analysis have been useful for searching common oligonucleotides in many subsets of sequences. On the other hand, database programming is useful not only for searching common oligonucleotides but also for searching oligonucleotides that do not exist in a subset of sequences. In this study, we found that there were hexanucleotide patterns that were rare or that did not exist in ribosomal RNA sequences. We also showed that sequences of 96 selected non-ribosomal hexanucleotides are normally present in known viral sequences (Table 5).

Improvement of detection efficiency of extracellular RNA on cDNA RDA

In the experiment for the determination of the effect of non-ribosomal hexanucleotides, we used a mixture of artificially synthesized and total cellular RNAs. The probabilities of hexanucleotide patterns in the sequence of pCIneo was different from ribosomal RNA (18.4×10^{-3}, 22.4×10^{-3} and 23.0×10^{-3} for non-ribosomal, non-V00125 and random hexamers, respectively). The artificially synthesized RNA included 30 priming sites for non-ribosomal hexanucleotides and was efficiently reverse transcribed (Figure 2). In the model experiments, cDNA RDA with non-ribosomal hexanucleotides efficiently reverse transcribed and specifically amplified the extracellular test RNA in the mixed RNA (Figure 3). cDNA RDA-derived detection of the artificial RNA was 30-fold more sensitive when non-ribosomal hexanucleotides were used than when random hexamers were used. These results suggest that non-ribosomal hexanucleotides dramatically improve the detection efficiency of cDNA RDA.

Application of non-ribosomal hexanucleotides for viral detection

The common existence of non-ribosomal hexanucleotides in known viral genomes (Table 5) and the improved sensitivity for the amplification of a non-ribosomal sequence in the mixed RNAs (Figure 3) suggest that non-ribosomal hexanucleotides could be used for non-specific detection of a viral sequence in infected cells. However, the sensitivity for sequence detection may be low when compared with that of common PCR using specific primers. Thus, the usage of this method might be restricted to infected cells that contain many copies of a virus.

The copy number of RNA viruses is dependent on the virus species, host cells and replicative and productive state of viruses. In our experiment, we detected 3 ng of contaminated RNAs in 1 μg of total RNA. Although this sensitivity is considerably lower than that of normal PCR, our experiments on SARS-CoV and BPI3 suggest that the sensitivity of cDNA RDA with non-ribosomal hexanucleotides is sufficient to detect these viruses in productively infected cells. Additionally, this method can be applied to most productive viruses, since sense and complementary sequences of non-ribosomal hexanucleotides were found to exist in most of the known virus sequences in GenBank (Table 5). In conclusion, this method could be applied as an alternative method for the detection of any emerging viruses.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

This work was funded by a grant from the Ministry of Economy, Trade and Industry of Japan (to D.E.) and by a grant from the Ministry of Environment of Japan. Funding to pay the Open Access publication charges for this article was provided by New Industry Creative Type Technology R&D Promotion Program in the Hokkaido Bureau of Economy, Trade and Industry.

Conflict of interest statement. None declared.

REFERENCES

1. Rota,P.A., Oberste,M.S., Monroe,S.S., Nix,W.A., Campagnoli,R., Icenogle,J.P., Penaranda,S., Bankamp,B., Maher,K., Chen,M.H. *et al.* (2003) Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*, **300**, 1394–1399.
2. Gao,S.J. and Moore,P.S. (1996) Molecular approaches to the identification of unculturable infectious agents. *Emerg. Infect. Dis.*, **2**, 159–167.
3. Holmes,E.C. and Rambaut,A. (2004) Viral evolution and the emergence of SARS coronavirus. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, **359**, 1059–1065.
4. Pavlovic-Lazetic,G.M., Mitic,N.S. and Beljanski,M.V. (2004) Bioinformatics analysis of SARS coronavirus genome polymorphism. *BMC Bioinformatics*, **5**, 65.
5. Lisitsyn,N., Lisitsyn,N. and Wigler,M. (1993) Cloning the differences between two complex genomes. *Science*, **259**, 946–951.
6. Chang,Y., Cesarman,E., Pessin,M.S., Lee,F., Culpepper,J., Knowles,D.M. and Moore,P.S. (1994) Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science*, **266**, 1865–1869.
7. Staskus,K.A., Zhong,W., Gebhard,K., Herndier,B., Wang,H., Renne,R., Beneke,J., Pudney,J., Anderson,D.J., Ganem,D. *et al.* (1997) Kaposi's sarcoma-associated herpesvirus gene expression in endothelial (spindle) tumor cells. *J. Virol.*, **71**, 715–719.
8. Hubank,M. and Schatz,D.G. (1994) Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acids Res.*, **22**, 5640–5648.
9. Tian,H., Cao,L., Tan,Y., Williams,S., Chen,L., Matray,T., Chenna,A., Moore,S., Hernandez,V., Xiao,V. *et al.* (2004) Multiplex mRNA assay using electrophoretic tags for high-throughput gene expression analysis. *Nucleic Acids Res.*, **32**, e126.
10. Eldering,E., Spek,C.A., Aberson,H.L., Grummels,A., Derks,I.A., de Vos,A.F., McElgunn,C.J. and Schouten,J.P. (2003) Expression profiling via novel multiplex assay allows rapid assessment of gene regulation in defined signalling pathways. *Nucleic Acids Res.*, **31**, e153.
11. Bailly,J.E., McAuliffe,J.M., Skiadopoulos,M.H., Collins,P.L. and Murphy,B.R. (2000) Sequence determination and molecular analysis of two strains of bovine parainfluenza virus type 3 that are attenuated for primates. *Virus Genes*, **20**, 173–182.
12. Iwai,H., Morioka,A., Shoya,Y., Obata,Y., Goto,M., Kirisawa,R., Okada,H. and Yoshino,T. (1998) Protective effect of passive immunization against TNF-alpha in mice infected with Sendai virus. *Exp. Anim.*, **47**, 49–54.
13. Mizutani,T., Fukushi,S., Saijo,M., Kurane,I. and Morikawa,S. (2004) Importance of Akt signaling pathway for apoptosis in SARS-CoV-infected Vero E6 cells. *Virology*, **327**, 169–174.
14. Mizutani,T., Fukushi,S., Saijo,M., Kurane,I. and Morikawa,S. (2004) Phosphorylation of p38 MAPK and its downstream targets in SARS coronavirus-infected cells. *Biochem. Biophys. Res. Commun.*, **319**, 1228–1234.
15. Mizutani,T., Fukushi,S., Murakami,M., Hirano,T., Saijo,M., Kurane,I. and Morikawa,S. (2004) Tyrosine dephosphorylation of STAT3 in SARS coronavirus-infected Vero E6 cells. *FEBS Lett.*, **577**, 187–192.
16. Thiel,V., Ivanov,K.A., Putics,A., Hertzig,T., Schelle,B., Bayer,S., Weissbrich,B., Snijder,E.J., Rabenau,H., Doerr,H.W. *et al.* (2003) Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.*, **84**, 2305–2315.
17. Hubank,M. and Schatz,D.G. (1994) Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acids Res.*, **22**, 5640–5648.

18. Monk,R.J., Meyuhas,O. and Perry,R.P. (1981) Mammals have multiple genes for individual ribosomal proteins. *Cell*, **24**, 301–306.
19. Gambari,R., Volinia,S., Nesti,C., Scapoli,C. and Barrai,I. (1994) A set of Alu-free frequent decamers from mammalian genomes enriched in transcription factor signals. *Comput. Appl. Biosci*., **10**, 501–508.
20. Scapoli,C., Rodriguez-Larralde,A., Volinia,S. and Barrai,I. (1993) Enrichment of oligonucleotide sets with transcription control signals. III: DNA from non-mammalian vertebrates. *Comput. Appl. Biosci*., **9**, 647–651.
21. Scapoli,C., Rodriguez-Larralde,A., Volinia,S., Beretta,M. and Barrai,I. (1994) Identification of a set of frequent decanucleotides in plants and in animals. *Comput. Appl. Biosci*., **10**, 465–470.
22. Volinia,S., Scapoli,C., Gambari,R., Barale,R. and Barrai,I. (1991) A set of viral DNA decamers enriched in transcription control signals. *Nucleic Acids Res*., **19**, 3733–3740.
23. Volinia,S., Scapoli,C., Gambari,R., Barale,R. and Barrai,I. (1992) Enrichment of oligonucleotide sets with transcription control signals. II: Mammalian DNA. *Nucleic Acids Res*., **20**, 551–556.