MRPrimerV: a database of PCR primers for RNA virus detection

Hyerin Kim[1,†], NaNa Kang[2,†], KyuHyeon An[1], Doyun Kim[2], JaeHyung Koo[2,*] and Min-Soo Kim[1,*]

[1]Department of Information and Communication Engineering, DGIST, Daegu, 42988, South Korea and [2]Department of Brain and Cognitive Sciences, DGIST, Daegu, 42988, South Korea

ABSTRACT

Many infectious diseases are caused by viral infections, and in particular by RNA viruses such as MERS, Ebola and Zika. To understand viral disease, detection and identification of these viruses are essential. Although PCR is widely used for rapid virus identification due to its low cost and high sensitivity and specificity, very few online database resources have compiled PCR primers for RNA viruses. To effectively detect viruses, the MRPrimerV database (http://MRPrimerV.com) contains 152 380 247 PCR primer pairs for detection of 1818 viruses, covering 7144 coding sequences (CDSs), representing 100% of the RNA viruses in the most up-to-date NCBI RefSeq database. Due to rigorous similarity testing against all human and viral sequences, every primer in MRPrimerV is highly target-specific. Because MRPrimerV ranks CDSs by the penalty scores of their best primer, users need only use the first primer pair for a single-phase PCR or the first two primer pairs for two-phase PCR. Moreover, MRPrimerV provides the list of genome neighbors that can be detected using each primer pair, covering 22 192 variants of 532 RefSeq RNA viruses. We believe that the public availability of MRPrimerV will facilitate viral metagenomics studies aimed at evaluating the variability of viruses, as well as other scientific tasks.

INTRODUCTION

Several fatal infectious diseases, such as Middle East respiratory syndrome coronavirus (MERS), Ebola and Zika virus, have recently emerged around the world, and mortality rates are very high in patients who contract these illnesses (1). Many infectious diseases are caused by viral infections, and in particular by RNA viruses. Detection and identification of these viruses are essential for understanding viral disease, and the accuracy and availability of tools designed for this purpose are crucial for effective and efficient virological studies.

The polymerase chain reaction (PCR) is widely used for rapid virus identification due to its low cost and high sensitivity and specificity, as well as the ubiquitous availability of the necessary reagents and equipment. Design of high-quality primers is essential for reliable PCR-based virus detection. During the design process, it is necessary to simultaneously check multiple filtering constraints on primers and perform similarity testing to verify that the designed primers will amplify only target virus rather than off-target sequences. Similarity testing for virus detection is a nontrivial task because, to achieve reliable detection, it is necessary to consider not only the entire genome of the host, but also the genomes of all other viruses, as off-target sequences.

Very few online database resources have compiled high-quality PCR primers for RNA viruses. Most primer databases, including PrimerBank (2,3), RTPrimerDB (4–6) and qPrimerDepot (7), contain primers for general use in real-time PCR and qPCR in specific organisms such as human, mouse, rat, fruit fly and zebrafish but do not contain primers for virus detection. The NCBI Probe Database (http://www.ncbi.nlm.nih.gov/probe/) provides nucleic acid reagents for use in a wide variety of biomedical research applications such as RNAi, PCR and microarray, as well as primers and probes for the detection of some viruses. However, its primers and probes are designed under various different filtering constraints and not validated against other viruses. Thus, they might not be appropriate for use in qPCR experiments requiring a full set of primer pairs that satisfy the same constraints, or in experiments requiring no cross-reactivity with different viruses. VirOligo (8), a database for virus-specific oligonucleotides, initially compiled from the published literature

more than 1637 oligonucleotides for detection of bovine respiratory disease-associated viruses. The last updated version of VirOligo (http://viroligo.okstate.edu/index.html) supports 109 RNA/DNA viruses. However, the database has not been updated with new entries for a number of years (the last update was in 2003). Primer-BLAST (9), one of the most widely used web-based tools for primer design, performs target-specific similarity testing. However, it is not suitable for designing PCR primers for RNA viruses, primarily for two reasons. First, it can perform similarity testing only within the same species. That is, it cannot perform similarity testing of a candidate primer designed to detect a specific virus in human that is related to two different species. For the same reason, it cannot design a primer to detect specific virus(es) in a host infected by multiple viruses. Second, it does not support batch design of primers for multiple segments of a virus, which is essential to ensure accurate detection in multi-phase PCR experiments. Accordingly, this lack of public resources prevents efficient detection of viruses in hosts, representing an obstacle to a more comprehensive understanding of viral disease.

To effectively detect and identify RNA viruses, we present a new database, MRPrimerV, which contains a collection of 152 380 247 high-quality PCR primer pairs for detection of 1818 RNA viruses. These *in silico* primers can detect 100% of RNA viruses in the most up-to-date version of the NCBI RefSeq database (Release 76, http://www.ncbi.nlm.nih.gov/refseq/). The database contains at least one valid primer pair for each of the 7144 coding sequences (CDSs) of these viruses. All primer pairs satisfy the same stringent filtering constraints and have passed rigorous similarity testing against all 101 684 human gene sequences and all RNA virus sequences in the RefSeq database. As a result, every primer pair in MRPrimerV is highly specific for RNA viruses. If a target virus has multiple CDSs, then MRPrimerV ranks the CDSs by the penalty scores of their best primer pairs: the lower the score, the higher the quality of the primer pair. We also consider a primer pair that can detect more number of genome neighbors as a higher quality one. To detect a target RNA virus, users need only pick and use the first primer pair for a single-phase PCR experiment, or the first two primer pairs for two-phase PCR experiments. In addition, we extracted 44 653 genome neighbors from the NCBI Viral Genome Resource, analyzed them along with RefSeq sequences, and compiled the result into the MRPrimerV database. Consequently, MRPrimerV provides the list of genome neighbors that can be detected using each primer pair, covering a total of 22 192 variants of 532 RefSeq RNA viruses.

MRPrimerV DATABASE

A schematic overview of MRPrimerV is provided in Figure 1. In the interest of thoroughness, we used the entire set of RNA virus sequences in the most up-to-date RefSeq database that have at least one CDS. The total number of such viruses is 1818; the total number of non-segmented genomes is 1400; and the total number of segmented genomes is 418. The 418 segmented genomes have many segments, and the total number of segments is 1572. Some viral genomes or segments have many CDSs, and

the total number of CDSs in 1818 viruses is 7144. We used all 7144 CDSs as one of the source databases for MRPrimerV. In addition to viral sequences, primer design for virus detection requires a set of human sequences for similarity testing. To compile high-quality PCR primers, we used all 101 684 human gene sequences in the RefSeq database as another source database for MRPrimerV. We also used the gene sequences of other animal species to detect some viruses. Concerning the MERS-CoV virus involved in recent outbreaks of respiratory illness, we used 26 720 camel (*Camelus dromedarius*) sequences in the RefSeq database as another source of sequences for MRPrimerV. The RefSeq database provides a comprehensive, integrated, non-redundant, well-annotated set of sequences and reference standards for multiple purposes, including genome annotation, gene identification and comparative analyses (10). However, the RefSeq database generally contains one genome per viral species; therefore, other sequence variants and closely related groups of viruses may not be detected using primers designed based on a single reference genome (11). For example, it may be difficult to use primers designed based on the reference HIV-1 sequence to detect the many subtypes of HIV-1. To overcome the limitations of the RefSeq database, we extracted 44 653 genome sequences from the NCBI Viral Genome Resource (https://www.ncbi.nlm.nih.gov/genome/viruses/). These sequences were collected by the NCBI Viral Genomes Project as validated genomes for viral species and indexed as 'neighbors' to reflect well-defined genotypes (12). We performed sequence alignments between the primers designed based on the RefSeq sequences and the genome neighbor sequences, and then inserted information regarding which neighbors can be detected using each primer pair into the MRPrimerV database.

To obtain all feasible and valid primers for RNA virus detection, we applied multiple filtering constraints and performed large-scale rigorous similarity testing against all human gene sequences using the MRPrimer technology (13), which returns all feasible and valid primer pairs existing in RNA virus sequences. MRPrimer performs fairly complex, large-scale processing to simultaneously check filtering constraints and perform similarity testing of all possible sub-sequences in a given database, based on a distributed MapReduce framework, resulting in design of very high-quality primers. qPCR analysis using many primer pairs, along with the corresponding sequencing and comparative analyses, revealed that primer pairs designed by MRPrimer are stable and effective in qPCR experiments (13). For qPCR experiments as well as for single or simultaneous multiple specific RNA virus detection, we applied the same filtering constraints to all primer pairs for all RNA viruses (Table 1). MRPrimerV includes not only SYBR Green primers, but also TaqMan probes to facilitate more reliable detection of viruses. The filtering constraints for TaqMan probe design are summarized in Supplementary Table S1.

We converted the results of MRPrimer processing, along with an annotation database downloaded from GenBank (ftp://ftp.ncbi.nlm.nih.gov/genomes/), to the MRPrimerV database in key–value format. The resultant database contains 152 380 247 *in silico* primer pairs for detection of
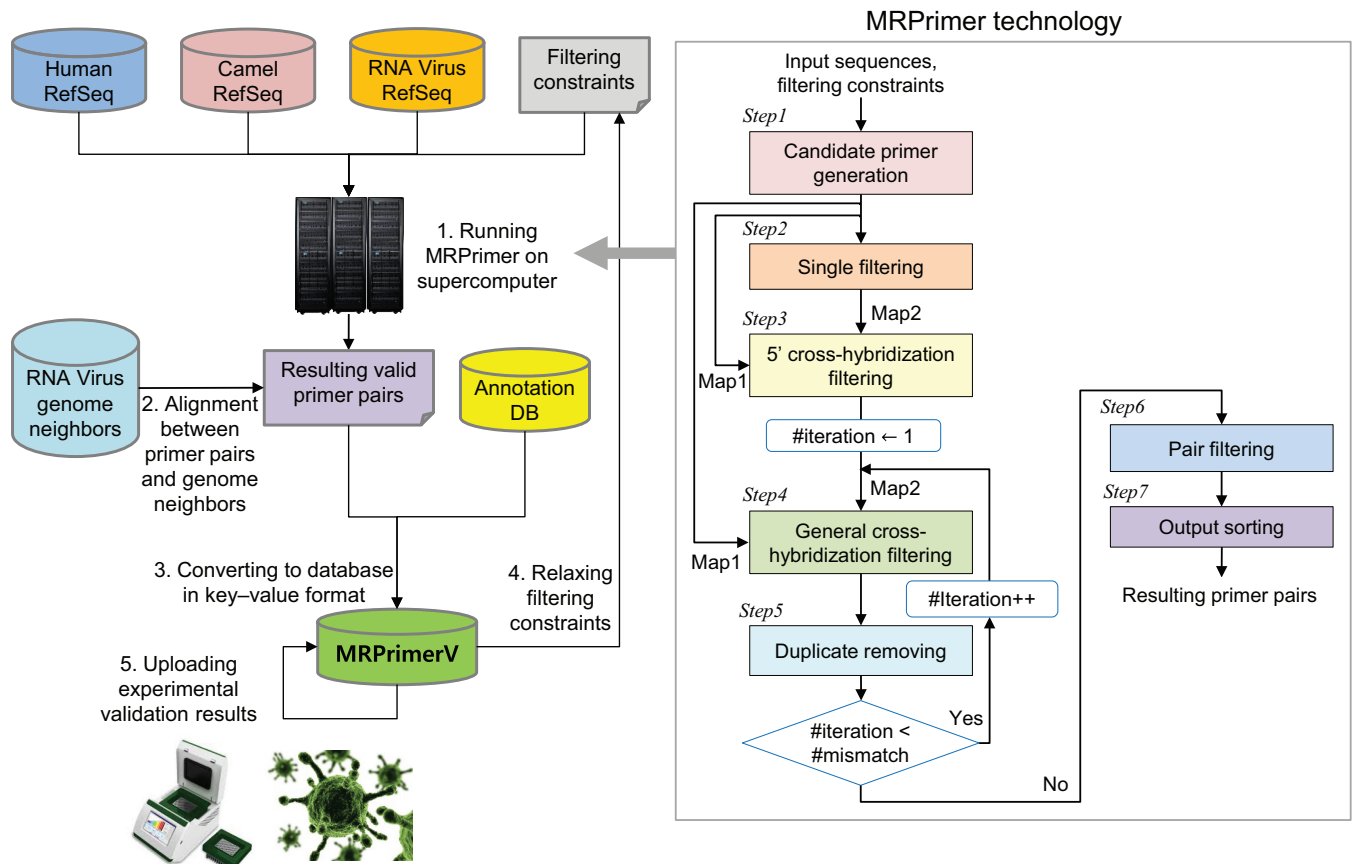
Figure 1. Schematic overview of MRPrimerV. (1) Checking the filtering constraints and performing large-scale rigorous similarity testing based on the MRPrimer technology on a supercomputer for the entire human sequence and RNA virus sequences. (2) Aligning the primer pairs with the genome neighbors to check whether each primer pair can amplify each genome neighbor.(3) Converting the results of processing along with information from the GenBank annotation database into the MRPrimerV database of 152,380,247 high-quality primer pairs. (4) Relaxing filtering constraints to cover 100% of 1,818 RNA viruses. (5) Updating the database with experimental validation results for some viruses.

Table 1. List of filtering constraints used in MRPrimerV

		Value range	
	Parameter	Default	Relaxed
Single primer	primer length	19–23 bp	19–23 bp
	melting temperature (TM)*	58–62°C	57–62°C
	GC content	40–60%	35–65%
	self-complementarity	<5-mer	<9-mer
	3′ self-complementarity	<4-mer	<4-mer
	contiguous residue	<6-mer	<6-mer
	3′ end stability (ΔG)	≥-9 *kcal/mol*	≥-9 *kcal/mol*
	hairpin	<4-mer	<9-mer
Primer pair	length difference	≤3-mer	≤5-mer
	TM difference	≤5°C	≤5°C
	product size	100–500 bp	70–500 bp
	pair-complementarity	<5-mer	<9-mer
	3′ pair-complementarity	<4-mer	<4-mer

*The nearest-neighbor thermodynamic model (14) was used to calculate the melting temperature.
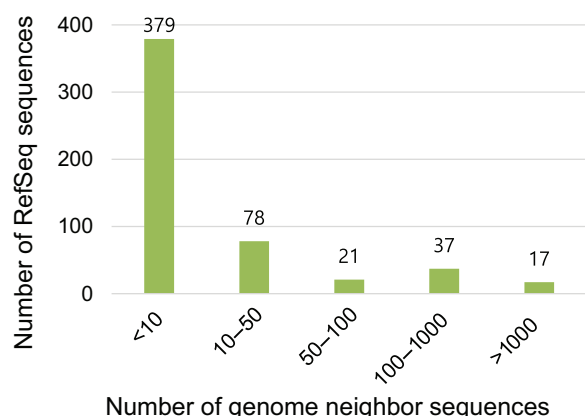
Figure 2. Distribution of the number of RefSeq sequences for each range of numbers of genome neighbors. The majority of RefSeq sequences (379/532, 71.24%) have fewer than ten genome neighbors, and only 0.03% of RefSeq have more than 1000 genome neighbors.

2963 non-segmented genomes or segments. Table 2 shows the statistics of RNA viruses covered by the primer pairs in MRPrimerV. In Table 2, 'CDS-specific primers' indicates primers that can amplify only a specific CDS of a target virus, i.e. they are target-specific in terms of CDS. By contrast, 'virus-specific primers' are those that amplify multiple CDSs of a target virus, but do not amplify CDSs of other viruses, i.e. they are target-specific in terms of virus. Amplicon size using a CDS-specific primer is unique, whereas that using a virus-specific primer might not be unique. MRPrimerV contains not only CDS-specific primers, but also virus-specific primers, because the latter are still target-specific for detection of a specific virus. Moreover, we note that, if valid primers exist for some segments of a virus with a segmented genome, we can detect that virus using those primers even when there are no valid primers for the remaining segments. Under the default constraints in Table 1, use of both CDS- and virus-specific primers could improve the RNA virus coverage ratio from 99.1% to 99.4% (Table 2). To further improve the coverage ratio, we relaxed the filtering constraints (Relaxed in Table 1) and performed MRPrimer processing. As a result, in terms of viral non-segmented genomes or segments, the coverage ratio was improved up to 99.7% (Table 2). All 1818 RNA viruses were completely covered, i.e. they could be detected using MRPrimerV compiled under relaxed constraints.

We validated the primers in MRPrimerV with viral genome neighbor sequences from the NCBI Viral Genome Resource. The 44 653 genome neighbors we used cover a total of 532 RefSeq RNA sequences for viruses infecting human hosts. Figure 2 shows the distribution of the number of RefSeq sequences for each range of numbers of genome neighbors. The majority of RefSeq sequences (379/532, 71.24%) have fewer than 10 genome neighbors, and only 0.03% of RefSeq sequences have more than 1000 genome neighbors. For instance, Human rotavirus B strain Bang373 (NC_021541) has four neighbors, whereas Rotavirus A segment 8 (NC_011502) has 2318 neighbors. We aligned the top-50 primers designed under the default constraints (Table 1) with the genome neighbor sequences and

checked whether each primer pair can amplify each genome neighbor. Table 3 shows the statistics of genome neighbors covered by the top-50 primers in MRPrimerV. The top-50 primers cover 22 192 neighbors out of 44 653 neighbors (49.69%). In the case of HIV-1, the top-50 primers cover 99.83% of neighbors and the top-50 primers for Rotavirus C segment 5 (NC_007570) cover 72.47% of neighbors.

The MRPrimerV database consists of nine key–value tables: one table for PCR primers, one table for TaqMan probes, five partial annotation tables for five query types, one full annotation table for viral genomes, and one full annotation table for viral CDSs. The partial annotation tables are used to identify user input for each query type. The full annotation tables are used to generate the output page. The database is physically stored using Redis (http://redis.io/), an in-memory key–value store that supports various kinds of data structures for various types of values, including string, hash, list, set and sorted set.

Due to a very large number of off-target sequences, including the 101 684 human gene sequences used to perform similarity testing, we had to run MRPrimer on a DGIST supercomputer (Rank #454 in TOP500 Supercomputer, June 2016) for more than 2 weeks to both check multiple filtering constraints and perform large-scale rigorous similarity testing for billions of candidate primers and probes. At present, it is important to perform large-scale processing to obtain high-quality products. For instance, DeepBind (15) performs large-scale deep learning to generate a database of predictive models of the sequence specificities of DNA- and RNA-binding proteins.

DATABASE INTERFACES

MRPrimerV provides two kinds of interfaces, simple search and glossary, which users can employ to search for primer pairs for a target RNA virus. In the simple search interface, users input a target RNA virus (as organism, keywords, GenBank accession, NCBI gene symbol or NCBI Gene ID) and click the search button. MRPrimerV then immediately outputs the best primer pairs for each CDS of the target RNA virus. Especially, in the simple search interface for 'organism,' MRPrimerV supports the query autocomplete feature, so that users can conveniently type the name of their organism of interest in the search box. On the glossary page, lists of RNA viruses are sorted alphabetically, so users can easily browse RNA viruses and obtain the primer pairs for a specific virus by clicking its name. The output page of both interfaces provides a brief description for the virus, a GenBank accession number with a link to the corresponding GenBank web page, brief information regarding the CDS (e.g. gene symbol and gene ID), and detailed information about the top primer pair including the penalty score, forward and backward primer sequences, TaqMan probe, melting temperatures, amplicon size, primer positions and validation results. In addition, MRPrimerV provides brief information and the list of genome neighbors that can be detected using the corresponding primer pair if information if available. Figure 3 shows primer pairs for respiratory syncytial virus (RSV) and mumps virus with the list of 78 genome neighbors for gene symbol L. Although the MRPrimerV database contains 152 380 247 primer pairs, the output web

Table 2. Statistics of RNA viruses covered by primer pairs in MRPrimerV

Filtering constraints	Default		Relaxed	
	Number	Ratio	Number	Ratio
Total number of non-segmented genomes or segments	2972	100%	2972	100%
Number of non-segmented genomes or segments covered by CDS-specific primers	2944	99.1%	2960	99.6%
Number of non-segmented genomes or segments covered by both CDS- and virus-specific primers	2955	99.4%	2963	99.7%
Total number of RNA viruses	1818	100%	1818	100%
Number of viruses covered by MRPrimerV	1817	99.9%	1818	100%

Table 3. Statistics of genome neighbors covered by the top-50 primers in MRPrimerV

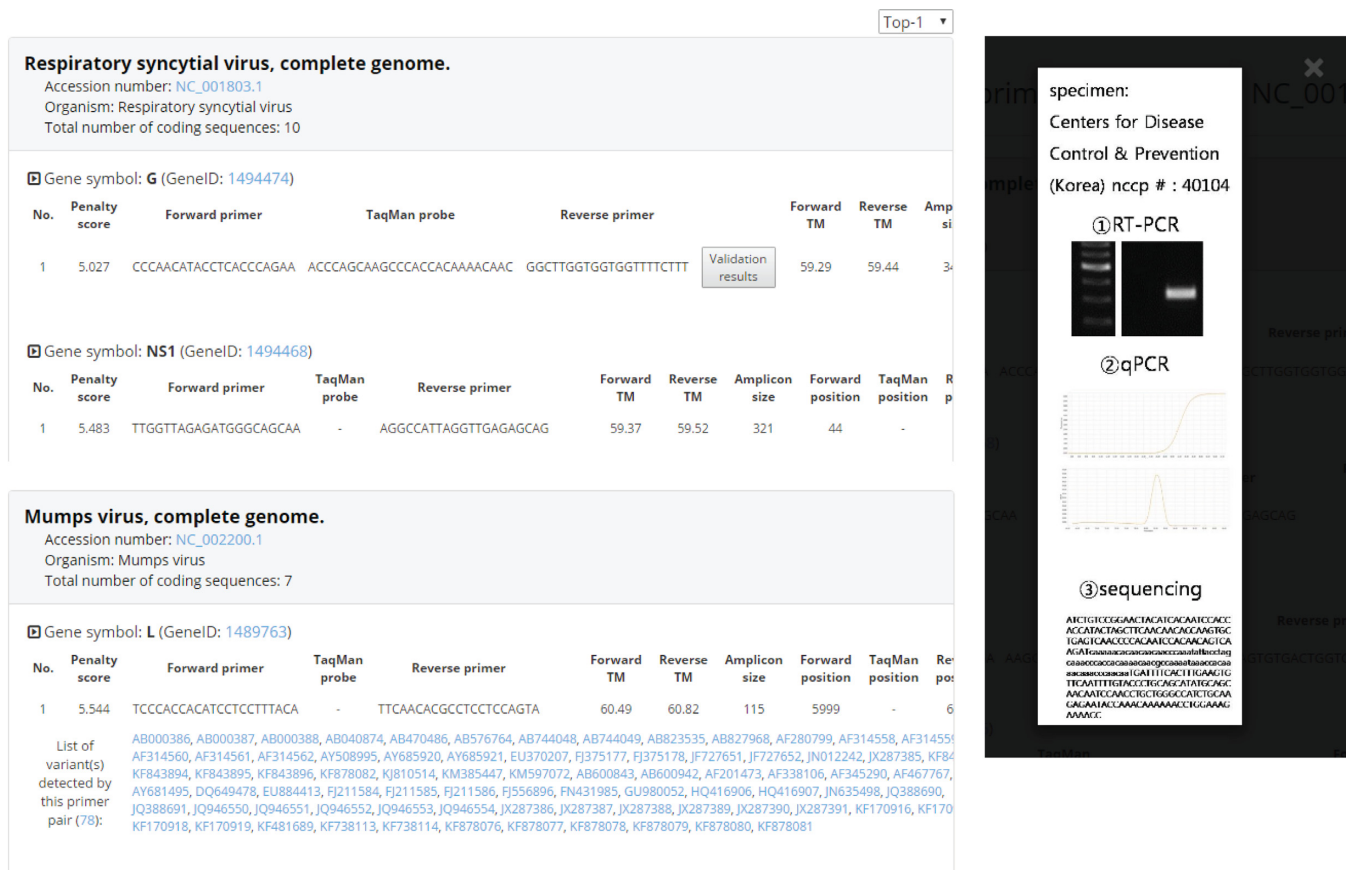|	All 532 RefSeq sequences	HIV-1	Rotavirus C segment 5
Total number of genome neighbors	44 653	762	109
Number of genome neighbors covered by top-50 primers of MRPrimerV	22 192	760	79
Ratio (%)	49.69	99.73	72.47

Figure 3. Output interface of MRPrimerV. Respiratory syncytial virus (RSV) has 10 coding sequences (CDSs); the output page first shows the top primer pair (penalty score: 5.027) for gene symbol G, and, next, the top primer pair (penalty score: 5.483) for gene symbol NS1. The bottom of the figure shows mumps virus with the list of 78 genome neighbors for gene symbol L.The head of the table provides information including a description of the virus, GenBank accession number (with a link to detailed gene information from GenBank), and organism information. The body of the table provides information about coding genes (gene symbol and gene ID) and primer information. If users click the 'Validation results' button, experimental validation data are shown, including specimen information, agarose gel data, qPCR amplification and melting curves, and sequencing data of the qPCR amplicon.

page interface does not show all primer pairs in the results, since it might be inconvenient for users. Users can adjust the number of resultant primer pairs for each CDS among Top-1, Top-5, Top-10 and Top-50 by clicking the dropdown menu in the output web page.

If the target virus has multiple CDSs, then MRPrimerV ranks the CDSs by the penalty scores of their best primer pairs: the lower the score, the higher the quality of the primer pair. We also consider a primer pair that can detect more number of genome neighbors as a higher quality one. The penalty score is calculated according to the method used in Primer3Plus (16). Thus, PCR using the best primer pair of the best (i.e. lowest-scoring) CDS is potentially more sensitive than PCR using the best primer pair from the next-best (i.e. second-lowest-scoring) CDS. In Figure 3, RSV has 10 CDSs; the output page first shows the top primer pair (penalty score: 5.027) for gene symbol G, and, next, the top primer pair (penalty score: 5.483) for gene symbol NS1. In most cases, to detect a target RNA virus, users need only pick and use the first primer pair for a single-phase PCR experiment.

In some cases, users might want to perform multi-phase PCR experiments for more accurate detection of viruses. For example, the World Health Organization (WHO) recommended that three rRT-PCR assays be conducted for routine detection of MERS (17). Because MRPrimerV ranks the CDSs and their top primer pairs by their penalty scores, users can easily select primers for multi-phase PCR experiments. For instance, for two-phase PCR experiments, users need only pick and use the first two primer pairs.

MRPrimerV also supports detection of specific virus(es) for a host infected by multiple viruses because all primer pairs in MRPrimerV were rigorously similarity-tested against not only the human sequence database, but also the entire RNA virus sequence database; at the same time, they all satisfy the same stringent and uniform filtering constraints. For instance, when we performed validation using Influenza A H1N1 and H3N2 viruses, each of which has seven segments and seven corresponding top primer pairs, we observed no cross-reactivity. The primers for H1N1 yielded a single band only for H1N1, but no band for H3N2, and vice versa. We also performed validation using Japanese encephalitis virus (JEV) and Dengue virus (Flavivirus). We observed no cross-reactivity; the JEV primers yielded only a single band for JEV and no band for Dengue virus. Sequencing analysis also confirmed the absence of cross-reactivity.

MRPrimerV contains not only primers and probes, but also validation results for some viruses, in particular 12 RNA viruses from the Centers for Disease Control and Prevention (CDC) of Korea (http://www.cdc.go.kr/CDC/eng/main.jsp). These viruses represent the entire set of viruses maintained by the CDC of Korea (Supplementary Table S2). By clicking the 'validation results' button, users can view validation data including specimen information, agarose gel data, qPCR amplification and melting curves, and sequencing data of the qPCR amplicon obtained using the selected primer pair. The right popup window in Figure 3 shows validation results for RSV.

CONCLUSION

The recent MERS outbreak in Korea, which spread rapidly due to slow and unreliable diagnosis, resulted in a 41% reduction in foreign tourism and decreased the gross domestic product growth rate in 2015 by 0.1% (18). The MRPrimerV database contains 152 380 247 high-quality PCR primer pairs for detection of 1818 viruses, covering 100% of the RNA viruses in the most up-to-date NCBI RefSeq database (Release 76). Because all primers in MRPrimerV were subjected to the same stringent filtering constraints and rigorous similarity testing against all 101 684 human gene sequences, 26 720 camel sequences and all RNA virus sequences, they are all highly target-specific for RNA viruses. The current MRPrimerV database mainly provides primers for detection of RNA viruses that infect human hosts. However, since many human pathogens, such as Zika, JEV, WNV, DENV, MERS-CoV and influenza virus, can also infect other animal species, future work will be directed toward updating the database to include primers against other animal species. MRPrimerV is freely accessible and provides a user-friendly interface. Because the database ranks the CDSs by the penalty scores of their best primer pairs, users can easily select primers for multi-phase PCR experiments to achieve more accurate detection of viruses. MRPrimerV also supports multi-virus detection and TaqMan probes. MRPrimerV contains not only primers, but also validation results for some viruses, and users are invited to send their own experimental validation data for MRPrimerV primers, so that they can be shared by researchers and international health communities. In addition, MRPrimerV provides the list of genome neighbors that can be detected using each primer pair, covering a total of 22 192 variants of 532 RefSeq RNA viruses. Therefore, this database could be used in viral metagenomics studies aimed at evaluating viral variability, as well as other scientific tasks. In the future, we will add a feature that allows users to design primers using virus sequences they provide. We believe that MRPrimerV, as a public database of high-quality primers for RNA virus detection, will aid future efforts to design primers for reliable diagnoses, facilitating effective responses to potential epidemics.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Sands,P., Mundaca-Shah,C. and Dzau,V.J. (2016) The neglected dimension of global security–a framework for countering infectious-disease crises. *N. Engl. J. Med.*, **374**, 1281–1287.
2. Wang,X., Spandidos,A., Wang,H. and Seed,B. (2012) PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res.*, **40**, D1144–D1149.
3. Spandidos,A., Wang,X., Wang,H. and Seed,B. (2010) PrimerBank: a resource of human and mouse PCR primer pairs for gene expression detection and quantification. *Nucleic Acids Res.*, **38**, D792–D799.
4. Lefever,S., Vandesompele,J., Speleman,F. and Pattyn,F. (2009) RTPrimerDB: the portal for real-time PCR primers and probes. *Nucleic Acids Res.*, **37**, D942–D945.
5. Pattyn,F., Speleman,F., De Paepe,A. and Vandesompele,J. (2003) RTPrimerDB: the real-time PCR primer and probe database. *Nucleic Acids Res.*, **31**, 122–123.
6. Pattyn,F., Robbrecht,P., De Paepe,A., Speleman,F. and Vandesompele,J. (2006) RTPrimerDB: the real-time PCR primer and probe database, major update 2006. *Nucleic Acids Res.*, **34**, D684–D688.
7. Cui,W., Taub,D.D. and Gardner,K. (2007) qPrimerDepot: a primer database for quantitative real time PCR. *Nucleic Acids Res.*, **35**, D805–D809.
8. Onodera,K. and Melcher,U. (2002) VirOligo: a database of virus-specific oligonucleotides. *Nucleic Acids Res.*, **30**, 203–204.
9. Ye,J., Coulouris,G., Zaretskaya,I., Cutcutache,I., Rozen,S. and Madden,T.L. (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, **13**, 134–144.
10. Pruitt,K.D., Brown,G.R., Hiatt,S.M., Thibaud-Nissen,F., Astashyn,A., Ermolaeva,O., Farrell,C.M., Hart,J., Landrum,M.J., McGarvey,K.M. *et al.* (2014) RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.*, **42**, D756–D763.
11. Brister,J.R., Ako-Adjei,D., Bao,Y. and Blinkova,O. (2015) NCBI viral genomes resource. *Nucleic Acids Res.*, **43**, D571–D577.
12. Bao,Y., Federhen,S., Leipe,D., Pham,V., Resenchuk,S., Rozanov,M., Tatusov,R. and Tatusova,T. (2004) National center for biotechnology information viral genomes project. *J. Virol.*, **78**, 7291–7298.
13. Kim,H., Kang,N., Chon,K.W., Kim,S., Lee,N., Koo,J. and Kim,M.S. (2015) MRPrimer: a MapReduce-based method for the thorough design of valid and ranked primers for PCR. *Nucleic Acids Res.*, **43**, e130.
14. SantaLucia Jr,J. and Hicks,D. (2004) The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 415–440.
15. Alipanahi,B., Delong,A., Weirauch,M.T. and Frey,B.J. (2015) Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.*, **33**, 831–838.
16. Untergasser,A., Cutcutache,I., Koressaar,T., Ye,J., Faircloth,B.C., Remm,M. and Rozen,S.G. (2012) Primer3–new capabilities and interfaces. *Nucleic Acids Res.*, **40**, e115.
17. Corman,V.M., Muller,M.A., Costabel,U., Timm,J., Binger,T., Meyer,B., Kreher,P., Lattwein,E., Eschbach-Bludau,M., Nitsche,A. *et al.* (2012) Assays for laboratory confirmation of novel human coronavirus (hCoV-EMC) infections. *Euro Surveill.*, **17**, 20334.
18. Lee,S.I. (2015) Costly lessons from the 2015 Middle East respiratory syndrome coronavirus outbreak in Korea. *J. Prev. Med. Pub. Health*, **48**, 274–276.