БИОМИКА/BIOMICS

http://biomics.ru

METAGENOMIC ANALYSIS OF RNA VIRUS SEQUENCES - PANACEA FOR RAPID SCREENING OF EFFECTIVE PRIMERS FOR VIRAL DIAGNOSIS

A.E. Kalashnikov[1,2], E.A. Gladyr[1]

[1]*Ernst Institute of Animal Husbandry, Federal Educational and Research Agency, Dubrovitsy, aekalashnikov@yandex.ru,*
[2]*All Russian Research Institute of Animal Breeding, Lesnye Polyany, Ministry of Agriculture of Russian Federation*

Resume

Here we describe a method for genomics that includes data filtering and final preparation of the viral consensus sequences. The outputs could be therefore used for the calculation of primers for viral detection. This method allows a simultaneous analysis of multiple sequences to obtain the variability at each site at whole-genome level or at the most complete viral consensus. We show two examples for diagnosing RNA viruses infecting bees and cattle, which we have successfully tested on the biological samples collected in Russia.

Key words: retroviruses, metagenomics, virus variability, *Apis mellifera*, *Bos taurus*, leukemia virus, mutation kinetics

Introduction

Currently, actively developing information technologies (IT) are used to solve some important problems in the field of genetics of viruses and their molecular diagnostics. Among the tasks, there are sequence analysis of nucleotides or proteins, detection of conserved and variable regions, and the thermodynamic and other factor analysis. These tasks could be components of IT pipelines allowing the selection of specific oligonucleotides and fluorescent probes to specific regions of the viral genomes [Barnes, 2007].

We selected two viral systems: *i)* Bovine leukemia virus (BLV), which causes a highly contagious infection that occurs in a latent form, that is striking in a number of key species of farm animals and represents a high risk [Rola-Łuszczak *et al*., 2013] and *ii)* genetic variants of honeybee viruses, which include the Kashmir bee virus (KBV), Acute bee paralysis virus (ABPV), Chronic bee paralysis virus (CBPV), Deformed bee virus (DWV), Sacbrood bee virus (SBV), Israel acute paralysis virus (IAPV), Black queen cell virus (BQCV), Slow paralysis virus (SPV). The infection with these viruses is a highly contagious for honeybees. The infection occurs in a latent form in the entire bee colony [Berenyi et al., 2006; DeMiranda, Genersch, 2010]. For effective control of infection, it is necessary to develop new modern diagnostic techniques. Diagnosis with classical methods such as ELISA and RID cannot provide sufficient analytical sensitivity and reliability [Rola-Łuszczak *et al*., 2013].

Materials and Methods

Objects

For the BLV viral system, 65 blood samples from the infected animals of black-pied cattle breed were collected from farms of the Moscow region of Russia. For genetic variants of honeybee viruses, a total of 270 adult bees from 90 bee colonies were collected from tick-borne defeat apiaries in the Ryazan region of Russia.

DNA and RNA isolation

For the nucleic acids isolation we used “Amplisens DNA-sorb-B” DNA isolation Kit (Central Research Institute of Epidemiology, Russia), “Amplisens RIBO-zol-B” RNA isolation Kit (Central Research Institute of Epidemiology, Russia), 0.5M Na_2EDTA, pH 8.0 (Amresco, USA), buffer 1 (*see Note 1*), buffer 2 (*see Note 2*).

Reverse transcription and real-time RT-PCR

For the reaction of reverse transcription we used “MMLV RT kit” (Evrogen), “qPCRmix-HS SYBR kit”, “qPCRmix-HS kit” (Evrogen, Russia), agarose electrophoresis grade (Lonza, Switzerland-USA), 10 mg/ml ethidium bromide Biotechnology Grade (Amresco, USA), 100+ bp ladder (Evrogen, Russia).

Primers and probes for detection of virus-infected *Apis mellifera* and BLV detections (all in 5'→3' direction)

Primers and fluorescent probes were synthesized by the Eurogen (Russia). Primer sequences for bee virus detection and reference genes:

ABPV (amplicon size 157 bp):
Forward TCAACCAGGCTATAACCAAC
Reverse TCTTGAAAATGCTTCACCAA

BQCV (amplicon size 130 bp):
Forward CCGTTAGTCCTCAACAGACT
Reverse AAACATGGAGCATAGTACGG,

CBPV (amplicon size 177 bp):
Forward ATTGTGAAGCCCAAAACCTG
Reverse GAACCGCTTCGGTGGTAATA

KBV (amplicon size 153 bp):
Forward GCAAATTGATGCTCCTAATG
Reverse TCATAGTCTAATGGGGCAAG

DWV (amplicon size 204 bp):
Forward ATATGCTTTTCCAGGTCCAT
Reverse CTCGCTTCTTCTCTTCACTC

SBV (amplicon size 186 bp):
Forward GGTGACCTTCATCCAGTATC
Reverse GGGTTTCATTTTCGATAACC

IAPV (amplicon size 105 bp):
Forward AGATGTAAGCATGCAGATCC
Reverse CTGAGATCTTCAGGACCAAA

SPV (amplicon size 201 bp):
Forward GTGGGCTGATAGGATGTTAC
Reverse ATCAATGTTATGCGTGGAAG

ACTB [Chen *et al,* 2005] (amplicon size 182 bp):
Forward AGGAATGGAAGCTTGCGGTA
Reverse AATTTTCATGGTGGATGGTGC

Primer sequences for detection of (BLV) and reference genes:

gene *ACTB* (amplicon size 193 bp):
Forward TGTTACAGGAAGTCCTTTGC
Reverse TGGCTGTCCATTCAAAATAA
Probe: FAM-AGGTGATCGCTTTTGTGTAA-BHQ1

gene *GAPDH* (amplicon size 212 bp):
Forward GACCACTTTGTCAAGCTCAT
Reverse GATGGAAACATGTGGAAGTC
Probe: FAM-AGAGGAAGAGTTCCTCAGCT-BHQ1

gene *GAG* (amplicon size 199 bp):
Forward TCTTAGAAGTCAATATCAAA
Reverse ATAACTAAGGGAGTCAATAA
Probe: R6G-CTCCCTACTCGTCCTTCAGT-BHQ1

gene *ENV* (amplicon size 137 bp):
Forward CCCTTTCAGAATTCGAGCTG
Reverse CTGTAAGGCCTGAGGACTGG
Probe: Cy5-CGCGATCGACCTATTCCTAA-BHQ2

Software for primer selection

For primer calculation and selection we used the following software:
Ugene (*ugene.net*);
ClustalX (ClustalO) (*http://www.ebi.ac.uk/Tools/msa/clustalo/*);
Muscle (*http://www.ebi.ac.uk/Tools/msa/muscle/*);
Primer3 (*http://primer3.sourceforge.net/releases.php*);
Primer3Plus (*http://primer3plus.com/*);
UnaFold (*http://unafold.rna.albany.edu/*);
Primer-BLAST (*https://www.ncbi.nlm.nih.gov/tools/primer-blast/*).

Equipment

In the experiments for nucleic acids isolation, amplification and reverse transcription has been used Vortex V-3 (Elmi, Latvia), Centrifuge 5424R (Eppendorf, Germany) with 45-position rotor, Vacuum pump (Russia), Laminar sterile box BMB-II-S (Laminar Systems, Russia), Spectrofotometer Novaspec Plus Visible (Amersham Biosciences, Sweden), DNA thermocycler CFX Touch with CFX Manager version 3.1 (Bio-Rad Laboratories, USA), Gel scanning system BDA BRD5 (Biometra, Germany).

Methods for assembling PCR reactions, as well as amplification parameters, will be described below, together with a discussion of choosing experimental conditions in the chapter of the Results and Discussion.

Results and Discussion

Algorithm for analysis of bees *Apis mellifera* infected with RNA-containing viruses

The test assays appeared more effective than previously developed for the detection of bee viruses COLOSS (*www.coloss.org*), detected 20-85% of positive specimens. The average frequency of the proviral DNA and RNA of DWV were not less than 75% and for CBPV more than 80%. The identified range of the relative concentration of RNA ranged from $1.14\pm0.689*10^{-2}$ to $5.41\pm2.08*10^{-1}$. The identified range of the relative concentration of proviral DNA of bees ranged from $6.41\pm2.23*10^{-6}$ to $3.95\pm1.50*10^{-2}$ at a frequency of 0.2 to 0.9. The active form of the virus was detected in the prevailing number of bee colonies and with greater concentrations.

The metagenomic search and analysis of nucleotide sequences of viruses (N>650 of nucleotide sequences which was included in GenBank database) were performed in the program Ugene 1.11.3 (*see Note 3*), by using algorithms ClustalX (*see Note 4*) and Muscle (*see Notes 5, 6*) [1, 2]. The differences for the better between the new and previous methods are methods used in this work. This methods were that the nucleotide sequences have been filtered according to the maximum differences by geographic origins of isolating, as well as the maximum genetic subdivision in population.

For each virus there were identified several areas with low frequency of mutations. This areas were used for the design of primers and probes using Primer3 or Primer3Plus includes WEB-online or offline version of programs (parameters: "qPCR" (for bee viruses) or "probe" for BLV) (*see Note 7*). The specificity of all variants of the oligonucleotides for BLV and bee viruses detection in the system of scoring matches has been verified in the BLAST-score system and on clinical sample test. Minimum 15-35 variants of primer pairs for one target was assembled from BLAST for the synthesis of BLV and bee viruses (data not shown), for all targets, for all situations) was calculated (*see Note 8*). The number of matches was calculated in PrimerBlast without replacements in PrimerBlast, divided into the total number of matches (BSc) (*Table 1*). The final variant of primers and (or) probes is indicated in the table. For each of these options, there were at least 3-7 variants of the tested primer pairs (the results of testing them are not shown). For all pairs of primers, PCR was carried out under various conditions of optimization, according to the PCR guide Bio-Rad and Applay-Biosystems, melting the amplification products in the presence of the intercalating dye.(Real-Time PCR Applications Guide – Gene-Quantification, Bio-Rad, Real-time PCR handbook – Gene-Quantification, Applied Biosystems).

Table 1.

The primer sequences and PCR conditions for bee virus detections

Nos	Virus name	Ta, ^{0}C	Tm, ^{0}C	BSc
1	ABPV	54.7	77.5	1.10
2	BQCV	54.0	84.5	3.03
3	CBPV	56.8	72-85	1.50
4	KBV	56.8	82.0	1.11
5	DWV	55.0	80.4	1.18
6	SBV	54.7	76.0	2.39
7	IAPV	56.4	75.0	2.58
8	SPV	56.4	75.0	4.00
9	ActB [Chen *et al*., 2005]	55-65, opt 59	80.5	0.507

Proviral DNA isolation from adult bees was carried out individually. DNA and RNA concentration was measured using a spectrophotometer. The reverse transcription was performed using "MMLV RT kit" with the amount the enzyme equal to 50 U/reaction. The concentration of primers for the amplification of the reference and target genes was 100 nM, and 200 pmol respectively. The temperature of a specific DNA product melting was determined by SYBR fluorescence detection (see Table 1).

The amplification of bee viruses targets was performed for proviral DNA and RNA of *Apis mellifera* and BLV viruses in active form (after the reverse transcription step). Detection of fluorescence for SYBR or TaqMan probes was performed starting from the formation temperature of the product in the calculated melting temperature. Amplification was performed by using the algorithm with the time of primer annealing and product elongation of 30 and 45 seconds respectively (40 cycles) using the "SYBR qPCRmix-HS kit".

The average value of the relative amount of viral RNA and integrated proviral DNA was calculated using the Bio-Rad user manual [Mackay, 2002]. Appropriate tests have been set up in the quantitative detection by SYBR Green I or TaqMan system relative to the reference genes *ACTB* and/or *GAPDH*. The analytical sensitivity of the tests was amounted to 1024 times dilution (final concentration was 0.07 mg/L). The measured concentration of proviral DNA of BLV in samples after the separation was represented 196-1544 and 38-1344 of RNA ng/μL respectively [Livak *et al*., 2001].

Later the samples were normalized to a concentration of 50 μg/mL. The specificity of the primers was verified by melting a curve desired product and by results the electrophoresis in 2 % agarose gel stained with ethidium bromide (data not shown).

Algorithm of diagnostics of bovine leukemia virus (BLV)

The efficiency of the designed primers and fluorescent probes was tested on samples of whole blood of cows of black-pied cattle breed from the farms of the Moscow Region and on RID positive for isolated DNA and RNA. For all samples, the concentration of DNA and RNA was determined and samples below 15 μg/mL were excluded from further analysis. Titration was performed to determine the optimum concentration of primers and specific probes, the range of variation was 0.1-0.8 μM (100-800 nM). It was found that the optimal target primer concentrations are 400 nM, the reference genome is 200 nM and the fluorescent probes are 300 nM.

Correlation analysis is based on the values of the threshold cycle (*Ct*) is the concentration of the DNA was carried out. The analysis showed a 100 % reliable result (no negative for the reference sample gene), which is observed at concentrations above 100 μg/mL. There was a decrease in the value of Ct with an increase in the concentration of DNA in the isolated sample. For 39 samples, comparative DNA

and cDNA testing was performed. It was shown that cDNA provides 100 % confidence in the result (no false negative tests), while DNA is only 84 %. Also, when testing cDNA, lower Ct values were observed for both the reference and target genes.

The efficacy of detecting proviral BLV DNA when compared with "nested" PCR was 26 %, while for RT-PCR, 88 %. In general, proviral RNA was detected in a larger number of cases (more than 99 %). Thus, the effectiveness of the test was 62 % more effective than compared with the classical PCR methods. Determination of the analytical sensitivity of the test on native blood samples with a minimum concentration of target RNA and DNA was performed. When the initial sample was diluted to a given concentration, the PCR was effective up to $8.2*10^{-3}$ (analytical sensitivity $8.8*10^{-3}$ μg/mL) from the initial concentration.

Alignment of nucleotide sequences was carried out in the program Ugene1.11.3 using algorithms in ClustalX and Muscle [1-2]. Design of primers and fluorescent probes were created by the Primer3 algorithm for "SYBRGreen" "probe" (s*ee Note 7*). Further, when primers with the help of factor sorting (Excel) exclude the options that were in the area of increased kinetics of mutations. Including were removed options that did not gain sufficient points in BLAST-score Ufold and the system of internal check at Primer3.

The formation of dimers and secondary structures was tested by UnaFold. The specificity of all oligonucleotides tested the scoring system of the matches in the BLAST system. For the synthesis and PCR selected the best 15-35 variants. To isolate the proviral DNA of BLV, the DNA extraction kit and DNA/RNA isolation was used.

Before isolating, blood leukocyte cells were preincubated with 1:10 0.5 M Na_2EDTA in a volume ratio, mixing 1:6 with buffer 1 and incubated for 1 hour at 0°C. The resulting hemolyze mix was centrifuged at 1500 g for 15 minutes. The pellet was resuspended in buffer 2 added in a ratio of 0.8/1 of the original volume of whole blood. The "MMLV RT Kit" was used for the reverse transcription reaction. To carry out the real-time PCR (RT-PCR), "qPCRmix-HS" was used. The calculated lengths of amplicons *ACTB*, *gapdh, gag, env* were 193, 212, 199, 137 bp respectively.

PCR was carried out on the DNA thermocycler and the fluorescence data were analyzed using the software.

The results of amplification were also visualized using electrophoresis in 3 % agarose gel followed by staining DNA with ethidium bromide of 0.02 %, fixing the image using system image rendering.

When assembling the multiplex reaction, the necessary protocol correction and repeated optimization of the primer and probe concentrations was carried out. Ultimately, the effective primer concentration of the reference gene turned out to be 100 nM, for the target genes was 300 nM, for the fluorescent probes was 300 nM.

The concentration DNA and RNA of BLV was determined on spectrophotometer using a quartz cuvette of 150 μL volume. After isolation the DNA concentration was normalized to the level of 50 μg/ml, and RNA in the reverse transcription reaction was added to 500 ng. The PCR of BLV was run in real time using an optimized algorithm (95^0C 5 min, {95^0C 15 sec, 53.6^{0}C 1 min, 72^0C 45 sec} x 5 steps, {95^0C 15 sec, 53.6^{0}C 1 min, 72^0C 45 sec} x40 steps with fluorescence detection) in the reaction volume of 30 μL.

It was confirmed that the primers to BLV allow to obtain the calculated length of amplicons (electrophoresis data are not shown).

Notes

1. Resolving 109.54 g sucrose (Pancreac, USA), 5 mL 1 M Tris-HCl (pH 7.6) (Amresco), 5 mL 1 M $MgCl_2$ (Amresco), *milli*Q water (Millipore, USA) to 1000 mL.
2. Resolving 50 mL 0.5 M Na_2EDTA (pH 8.0) (Amresco), 15 mL 5 M NaCl (Helicon, Russia), *milli*Q water to 1000 mL.
3. The script for RNA blast search is included.
4. Scoring matrix blosum, open/end gap penalty = 10, extending/separation gap penalty = 0.05.
5. The script is integrated in Ugene.
6. The data not shown, and available at [1-2].
1. This is for TaqMan preinstall configuration, we have differed in divalent cations imposed in concentration of 3.0.

$$BSc = \frac{freq_{TargetPrimer1} + freq_{TargetPrimer2}}{N_{taxviuses}} + freq_{PRIMERBLAST} + Ufold + Tm$$

2. where freq(N) is the number of matches without replacements, tax - the total number of sequences in the taxon, Ufold - a coincidence Ufold (0/1 points).

Conclusion

There are presented the two methods of detecting disseminated RNA-containing viruses of cattle and bees, which are of economic importance, are presented. The methods are based on modern PCR in real time. With the help of the described computer programs it is possible to quickly and efficiently select the primer variants, evaluate their selectivity and selectivity, and test the conditions for their application in molecular diagnostics of DNA and RNA viruses. The analytical sensitivity of the tests was from 0.07 mg/L.

Acknowledgement

The work was supported from RSF project "Study, conservation and management of biodiversity as a basis for obtaining animals for healthy, safe and high quality food" № 14-36-00039.

References

1. Alignment database (2016) bee viruses. *https://goo.gl/4RXgwf*
2. Alignment ftp database (2016) BLV. *https://goo.gl/4RXgwf*
3. Barnes M.R. Bioinformatics for Geneticists: A Bioinformatics Primer for the Analysis of Genetic Data. 2007. P.1-576.
4. Berenyi O., Bakonyi N., Derakhshifar I., Koglberger H., Nowotny N. Occurrence of six honeybee viruses in diseased Austrian apiaries // *Appl. Envir. Microbiology.* // 2006. V.72. P.2414-2420. doi: 10.1128/AEM.72.4.2414-2420.2006
5. Chen Y.P., Higgins J.A., Feldlaufer M.F. Quantitative real-time reverse transcription-PCR analysis of deformed wing virus infection in the honeybee (*Apis mellifera* L.) // *Appl. Environ. Microbiology*. 2005. V.71. P.436-441. doi: 10.1128/AEM.71.1.436-441.2005
6. DeMiranda J.R., Genersch E. Deformed wing virus // *J. Intertebr. Pathology*. 2010. V.103. P.48-61. doi: 10.1016/j.jip.2009.06.012
7. Livak K.J., Schmittgen T.D. Analysis of relative gene expression data using real-time quantitative PCR data // *Methods*. 2001. V.25. P.402-408. doi: 10.1006/meth.2001.1262
8. Mackay I.M. Real-time PCR in the microbiology laboratory // *Clin. Microbiol. Infections*. 2002 . V.10. P.190-212. doi: 10.1111/j.1198-743X.2004.00722.x
9. Rola-Łuszczak M, Finnegan C, Olech M, Choudhury B, Kuźmak J. Development of an improved real time PCR for the detection of bovine leukaemia provirus nucleic acid and its use in the clarification of inconclusive serological test results // *J. Virol. Methods*. 2013. V.189. P.258-64. doi: 10.1016/j.jviromet.2013.02.014

МЕТАГЕНОМЫЙ АНАЛИЗ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ВИРУСОВ РНК — ЯВЛЯЕТСЯ ЛИ ЭТО ПАНАЦЕЕЙ ДЛЯ БЫСТРОГО ПОИСКА ЭФФЕКТИВНЫХ ПРАЙМЕРОВ ДЛЯ ВИРУСНОЙ ДИАГНОСТИКИ

Калашников А.Е.[1,2], Гладырь Е.А.[1]

[1]Всерооссийский институт животноводства им. Л.К. Эрнста, Федеральное агентство по вопросам образования и науки, Дубровицы, aekalashnikov@yandex.ru,
[2]Всероссийский научно-исследовательский институт животноводства, Лесные Поляны, Министерство сельского хозяйства Российской Федерации

Резюме

В настоящей работе мы описали новый метод геномики, который включает фильтрацию данных и окончательную подготовку вирусных консенсусных последовательностей для анализа. Поэтому результаты могут использоваться и для расчета праймеров с целью обнаружения вирусов. Этот метод позволяет проводить одновременный и одномоментный анализ множества последовательностей для получения информации об изменчивости в каждом сайте на уровне целого генома или при работе с наиболее полном консенсусом - на уровне популяции вирусов. Мы приводим два примера для диагностики РНК-вирусов, заражающих пчел, и вируса крупного рогатого скота, которые мы успешно протестировали на биологических образцах, собранных в России.

Ключевые слова: ретровирусы, метагеномика, изменчивость вирусов, *Apis mellifera*, *Bos taurus*, вирус лейкемии, изменчивость мутаций