UDC 34.15.31

RECOMBINANT YELLOW FEVER VIRUS WHICH EXPRESSES HETEROLOGOUS PROTEIN DURING REPLICATION

Keyer V.V., Shevtsov A.B., Kassymkhanova A.S., Baltabekova A.Zh., Shustov A.V.

National Center for Biotechnology
13/5, Korgalzhyn road, Astana, 010000, Kazakhstan
keerv@mail.ru

ABSTRACT

Yellow fever virus (YFV) is a model representative in the genus Flavivirus, which along with the YFV includes other human pathogens such as the tick-borne encephalitis virus, Zika virus, etc. The flaviviruses are prevalent worldwide and pose a risk to the population of Kazakhstan. Although the YFV itself is not present in Kazakhstan, a vaccine strain of the YFV may serve as a convenient model to study aspects of molecular virology of the flaviviruses. The vaccine strain 17D of the YFV was used to construct cDNA copy of a full-length viral genome. The cDNA was cloned in an *E.coli* plasmid. The 5'-end of the genome is placed under control of SP6 RNA polymerase promoter. This allows producing the virus' genomic RNA using in vitro transcription. The 3'-end of the genome is engineered to be fused with antigenomic ribozyme (RBZ) of the hepatitis D virus. The RBZ cleaves itself off upon transcription to yield the correct 3'-end of the YFV genome. Also, a gene encoding green fluorescent protein (GFP) was inserted into the YFV genome in a position preceding structural proteins. An activity of foot-and-mouth disease autoprotease 2A was utilized to cleave off the GFP from viral polyprotein to ensure correct processing of the YFV proteins. Precautions were undertaken to preserve the 5'-cyclization signal which important for replication.

The obtained molecular clone proved infectious. The virus (1612.YFV/GFP) was rescued from the plasmid by using in vitro transcription and RNA transfection. Thus an experimental system to produce recombinant flaviviruses was generated which may be advanced to produce live vaccines based on the YFV vector.

Key words: Yellow fever virus (YFV); Molecular clone; Flavivirus; Heterologous insert; Green fluorescent protein; HDV antigenomic ribozyme.

INTRODUCTION

Yellow fever in humans is an acute disease presenting with fever, headache, jaundice (hence the name "yellow"), vomiting, etc. The disease is caused by a virus (yellow fever virus, YFV) and transmitted by mosquitoes, among which one species Aedes aegypti plays a predominant role. Depending on the virus strain, a fraction of

patients may develop symptoms of sever visceral or neurological disease and the latter leads to death of ~50% of the patients with the sever disease [1].

The virus is thought to have an evolution origin in Africa where it is endemic to equatorial and tropical regions, although now it is endemic also to Central and South America. Epidemics of yellow fever caused mass deaths and devastation in the past. To date, yellow fever is controlled by vaccination and mosquito-control measures. Yellow fever in humans may be prevented by vaccination with a live attenuated vaccine. There are no virus-specific anti-viral drugs (etiologic treatment) for yellow fever and the treatment in hospitals relies on supportive interventions.

In 2017 WHO launched an ambitious international program called the Eliminate Yellow Fever Epidemics (EYE) Strategy [2]. Although the goal of a complete eradication is not attainable as the virus circulates in wildlife nonhuman primates, being transmitted by a variety of sylvatic mosquitoes. The goal of the program is to stop mass outbreaks by vaccination and halt human-to-human transmission cycles in endemic regions.

Yellow fever belongs to a class of diseases which have imposed so substantial impact on public health and social life that it shaped political borders and set marks in songs and folklore. Modern phylogenetic studies reveal that the YFV originated in Africa, and was introduced to Americas in early 1600-s during slave trade era. In Americas and Caribbean, the disease named “Yellow Jack” or “Black Vomit” with 20 to 50% mortality ravaged the populace. Although, some nations benefited form the disease, namely the Caribbean state Republic of Haiti owes its very existence to yellow fever: in 1801 the 25,000-strong army of Napoleon landed on the island of Haiti (which before happen to be a French colony) to suppress the uprising for independence. However, the main enemy of the French army became Yellow Jack which killed 22000 soldiers and officers; small amount of survivors (12%) had enough strength only to evacuate artillery and horses from the island. The French decided to leave, and Haiti proclaimed itself an independent republic in 1804. During Spanish-American War (yr. 1898) for every soldier who died in battle other 13 (!) died of yellow fever. During this war, scientific investigations of the disease were led by U.S. army physician Walter Reed and proved that limiting a population of mosquitoes is effective to lower incidence of yellow fever, which strategy was then used to protect workers during construction of the Panama Canal.

Yellow fever remains important public health problem in our days: a cessation of the outbreaks of is not in sight. WHO reported 26,356 cases during 1988-2007, and the most recent epidemics include 20,000 cases with 1,000 deaths in Cameroon (in 1990) and ~6000 cases in Angola and Congo in 2016 [3].

Diseases caused by flaviviruses in Kazakhstan

Though the YFV is not present in Kazakhstan there are other flaviviral infections threatening the population of the country. Flaviviruses (representatives of the gen. Flavivirus, family Flaviviridae) are causative agents of dangerous human diseases including yellow fever, tick-borne encephalitis, Zika fever, West Nile fever, dengue, etc. Majority of flaviviruses are arboviruses and can infect vertebrates and arthropods; transmission occurs via bites of bloodsucking mosquitoes or ticks. The most prevalent and the most medically important flaviviral infection in Kazakhstan is tick-born encephalitis (TBE) (far eastern variant, also known as Russian spring-summer encephalitis, RSSE) [4]. TBE is endemic in boreal forests in the north-east and south mountain regions. Karshi virus (which is a different member of the TBE serocomplex) exists in the south of the country [5]. Serological investigations indicate a hidden circulation of West Nile fever among residents of West-Kazakhstan living near a delta

of Volga river. There is a threat of importation of other flavivirus infections, in particular Zika virus which is widely present in countries of Southern and Eastern Asia [6].

Genome organization and proteins of the YFV

Flaviviruses are RNA-containing viruses which replication occurs in cytoplasm without reverse transcription, with a help of the virus-encoded RNA-dependent RNA polymerase (RdRp). The YFV genome is a RNA molecule of + (positive) polarity meaning that it is directly translated on ribosomes (fig. 1A [7], with changes). It is 10842 nt-long. The RNA is 5'-terminally capped with an (m7)GpppA(m) structure but unlike cellular mRNAs flavivirus genomes are not polyadenylated at 3'-ends. Nearly all the genome encodes a polyprotein which open reading frame (10233 nt, 3411 codons) is the only long ORF in the genome. The ORF is flanked with noncoding regions (NCRs).

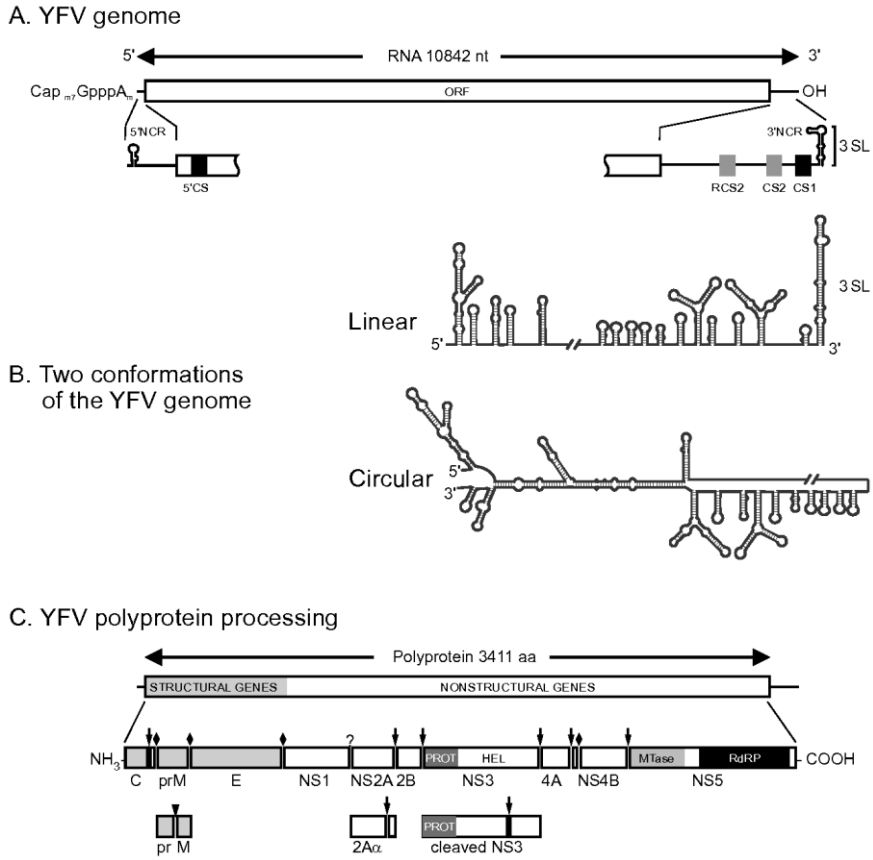


Fig 1. [7], with changes. A. YFV genome organization. NCR, noncoding regions. SL-stem-loop structure. CS-cyclization sequences. B. Two conformations of the YFV genome. In the circular conformation distinct motifs within the 5'NCR and beginning of the capsid gene base-pair with complementary motifs from the 3'NCR. C. Processing of the YFV proteins. Functional domains in selected nonstructural proteins are labeled. NS3 domains: PROT, protease; HEL, helicase. NS5 domains: MTase, methyltransferase; RdRP, RNA-dependent RNA-polymerase. Designations above the cleavage sites indicate involved proteases: downward arrow, viral serine protease NS3; diamond, cellular signal peptidase; triangle indicates cleavage in prM by cellular protease furin; question mark, unknown enzyme cleaves the NS1-2A junction

Purified genomic RNA of the YFV is infectious, i.e. upon transfection into susceptible cells it gives rise to live virus. This feature is common to all RNA-positive viruses and simplifies studies on their molecular virology. Genetic engineering of the

YFV as a model flavivirus provides for dissecting functions of viral components and serves for utilization of the perfectly characterized 17D strain as a vector for live attenuated vaccines.

The 5'NCR (and a sequence of the ORF in proximity to the 5'NCR) and 3'NCR regions contain signals important for replication of genomic RNA. In the sequence within the YFV genomic RNA encoding the first 25 aa of the capsid protein, a motif GGCGTCAATATGGT is present, termed the cyclization sequence (5'CS), which can base-pair with the reverse-complementary motif (named CS1) in the 3'NCR (nt 10749-10762). Also two other motifs with similar capability of base-pairing were noted in the 3'NCR (CS2 and RCS2). In fact, it is long considered that the flaviviral genomes are capable of "cyclization" and formation of a structure which is called panhandle, in which 5'- and 3'-ends of the genomes lay in proximity. Modeling of RNA secondary structures reveals two possible conformations of the YFV genome, linear and circular (Fig. 1B) in which the CS-motifs either participate in formation of stem-loops or form a "handle" of the panhandle structure. This behavior is typical to regulatory elements found in long coding RNAs of prokaryotes and eukaryotes called RNA-switches and such RNA-switch (including the CS-motifs) is believed to regulate the replication.

All viral proteins are encoded in a form of a polyprotein in which individual functional proteins follow in the order: C-prM-E-NS1-NS2A-NS2B-NS3-NS4A-NS4B-NS5 (Fig. 1C). The polyprotein is cotranslationally and posttranslationally cleaved into functional proteins. The cleavage is mediated by cellular and viral proteases.

The YFV virion consists of three structural proteins: the capsid protein (C) forms a complex with the genomic RNA called nucleocapsid which occupies the central part of the virion. Proteins prM and E are surface glycoproteins which are embedded into lipoprotein shells.

The YFV encodes nonstructural proteins NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5 which are not constituents of virions, rather the nonstructural (NS) proteins play a role in various processes in virus life cycle. The NS participate in processing of the viral polyprotein, synthesis of (-) and (+) strands of viral RNA, packaging of the genomic RNA into new virions. All nonstructural proteins are needed for replication, although it is not exactly known how they assemble to form a replicative complex (RC). Study of the RCs of flaviviruses is complicated by a fact that they are associated with membrane structures which are formed in infected cells as a result of remodeling of membranes of endoplasmic reticulum (ER). With the protein NS1, two unresolved puzzles in biology of flaviviruses exist. Unlike other NS, the NS1 is a secretory glycoprotein. The NS1 precursor protein has a signal peptide (SP) for transport into the endoplasmic reticulum (ER). The NS1 is translocated through a translocon on the ER membrane into ER cisterns [8]. On the other hand, presence of the NS1 is necessary for replication (synthesis of (-)strand of the virus RNA does not occur without the NS1). The NS1 is rapidly glycosylated in the ER, and it is surprising that mutations of glycosylation sites in the NS1 inhibit the replication [9, 10]. Ultrastructural studies revealed colocalization of the NS1 with double-stranded RNA (dsRNA) which is a replicative intermediate. The NS1 interacts with the transmembrane proteins NS4A and NS4B [11, 12, 13, 14]. Unresolved problem is how the NS1 interacts with the other proteins in the RC across membrane and what is the NS1 biological role? Flaviviruses with mutations in the NS1 were found which are capable of intracellular replication (in a form of the virus RNAs) although they are defective in packaging (formation of new virions). Interaction of the NS1 with virus glycoproteins residing in lipid shells of virions has been observed [15].

Another enigma of the NS1 is related to its processing. Study of cleavage of the YFV polyprotein allowed identification of proteases responsible for cleavages in all

junctions between functional proteins, except the junction NS1-NS2A. It is known that processing of structural proteins (C, prM, E) is carried out by the cellular signal peptidase; maturation of virions requires activity of the cellular protease furin (this protease resides in the Golgi complex). Junctions between all nonstructural proteins, exception the junction NS1-NS2A, are cleaved by the viral protease NS3. A protease which carries out cleavage between the NS1 and NS2A has not yet been identified. This project aims to identify this protease.

The proteins NS2A and NS2B are transmembrane proteins which interact with each other [16]. They also colocalize with other RCs and are needed for replication [17, 18, 19]. The proteins NS2A and NS2B interact with the NS3 helicase/protease [19]. There is sufficient evidence that the NS2A participates in packaging of the genomic RNA into virions [17, 19].

The protein NS3 is commonly referred to as viral protease. *In vivo* NS3 occurs in a complex with the NS2B which is a co-factor bound to the ER membrane [20]. Also the NS2B-NS3 complex associates with NS4A [21]. Domain of serine protease is located at N-terminus of the NS3 protein. The NS3 protease cleaves viral polyprotein at various junctions between nonstructural proteins (but not the NS1-NS2A junction). This cleavage leads to formation of the functional nonstructural proteins. C-terminal domain of the NS3 has activity of RNA helicase and nucleoside triphosphatase and this domain binds to protein NS5 (viral RNA polymerase) [22].

The proteins NS4A and NS4B are transmembrane proteins (NS4A has 3 TD and NS4B has 3 TD) which interact with each other [19]. Exact functions of the NS4A and NS4B are unknown. There is indication that both of these proteins play a role in remodeling of the ER membranes and in formation of virus-induced membrane structures called vesicular packets in which RCs are localized [23, 24, 25]. The NS4A and NS4B bind to NS1 which interaction presumably involves hydrophobic domain of the NS1 [12].

The protein NS5 is the largest (~105 kDa) and the most evolutionarily conserved protein of flaviviruses [26]. The NS5 has methyltransferase activity which is needed to form a cap structure at the 5'-end of the viral (+)strand RNA [26] and the activity of RNA-dependent RNA polymerase (RdRp). Function of the NS5 as the catalytically active component of replicase is confirmed by presence of amino acid triplet GDD which is characteristic to RdRp [27, 28]. C-terminal domain of NS3 interacts with N-terminal domain of the NS5 [29] and a number of cellular proteins interact with the NS5 as well [30].

Use of yellow fever virus as a vector to express heterologous proteins

At present the highest interest in a use of flaviviruses as vectors for expression of heterologous proteins reside in a field of vaccine development. The YFV strain 17D has been in use as a live vaccine against yellow fever since 30-th years of the XX century and has excellent records of efficacy and safety. Importantly, it is now possible to industrially produce 17D in compliance with GMP/GDP standards. The 17D is considered to be an efficient backbone in development of vectored vaccines against various pathogens. Examples of such use include chimeric viruses which are the YFV with prM/E exchanged for glycoproteins from dengue virus (DEN) [31]. A different chimeric virus was developed by the exchange of the prM/E genes for the ones derived from Japanese encephalitis (JE) virus. This technology (prM/E gene replacement) was named ChimeriVax by the company Sanofi Pasteur. Several vaccines based on this principle are under clinical trials. To date, there is a scarcity of examples of utilization of flaviviruses (also including the YFV) as vectors for tasks apart from development of vaccines.

In this paper we describe YFV genome modified to express a marker protein (GFP) during intracellular replication. The described recombinant YFV with the insertion of GFP in the polyprotein in proximity to N-terminal position is genetically stable.

MATERIALS AND METHODS

Bacterial strains and genetic engineering

E.coli strain DH5α was used for cloning and preparative isolation of plasmids. Direct cloning of PCR products using a kit pGEM-T Easy Cloning system (Promega) was used for intermediate cloning. Genetic engineering was done using standard procedures such as restrictase digestion - ligation of DNA fragments, and PCR.

Synthetic DNA fragments

A PCR was used to synthesize gene for YFV capsid (C) protein with a sequence which is codon-optimized using human codon usage table. A set of oligonucleotides (60-mers) was devised which contain the oligonucleotides corresponding to parts of the C-gene nucleotide sequence; the oligonucleotides alternate in order "sense" ->"antisense" ->"sense" -> etc. The neighboring primers have overlapping sequences and are able to hybridize with each other with a calculated annealing temperature 60°C. This set of oligonucleotides is called "internal" primers. One pair of "flanking" primers which target at the ends of the synthetic fragment, were also devised and used for the PCR. Synthesis of the DNA fragment was carried out by the two-rounds PCR. In the first round, PCR amplification was carried out with a mixture of all internal primers. Resulting products of stochastic synthesis were used as a template in the round II PCR which was done with one pair of the flanking primers. The round II PCR mixture was separated by electrophoresis in a 1% gel of low-melting-point agarose. The fragment of an expected length was isolated from the gel and cloned into pGEM-T (Promega). Inserts in the clones were sequenced by both strands to confirm the correct sequence.

In vitro transcription

A plasmid carrying a cloned cDNA copy of the YFV genome in amount of 5 μg was digested with the restriction enzyme XhoI, which linearizes the plasmid. The DNA was extracted with a phenol-chloroform mixture and precipitated with alcohol. A reaction mixture for synthesis of RNA consists of 12.5 μl of water, 10 μl of 5X buffer for SP6 RNA polymerase, 5 μl of DTT solution (10 mM), 2.5 μl of a 10 mM synthetic analogue of cap m7G(5')ppp(5')G, 10 μl of a 10 mM solution of ribonucleoside triphosphate, 10 μl of linearized DNA solution in water (200 ng/μl). To the mixture, 1 μl of RNase inhibitor (RiboLock, ThermoFisher Scientific) and 2.5 μl of SP6 RNA polymerase (ThermoFisher Scientific) were added. The mixture was incubated at 41°C for 1 hour. To confirm effectiveness of the synthesis, an aliquot of the reaction mixture was applied to gel electrophoresis in 1% agarose.

Transfection of RNA and rescue of virus

BHK-21 cells were grown in a complete growth medium (DMEM High Glucose, supplemented with 10% fetal bovine serum (FBS), 1% antibiotics penicillin/streptomycin, 1% vitamin complex). Media on growing cultures were changed every 24 hours.

Synthetic RNA was transfected into the BHK-21 culture, which reached 70-90%

confluency (5x10^5 cells grown in a well of a 6 well plate). Transfection was performed with cationic liposomes (reagent Lipofectamine 2000, LifeTechnologies). An aliquot (10 µl) of Lipofectamine 2000 was dissolved in 150 µl of OpiMEM medium. In a different tube, to 150 µl of OpiMEM medium, 10 µl of synthesized RNA was added (RNA content in the reverse transcription reaction mixture was ~60 ng/µl, the reaction mixture was used without purification). After the solutions were incubated for 15 minutes, the RNA solution was mixed with the solution of lipofectamine, followed by additional incubation for 5 minutes at room temperature. An RNA-lipid complex was added to the monolayer of cells washed with PBS. The transfected culture was incubated for 24 hours at 37°C in an atmosphere of 5% CO_2, after which the medium was changed to the complete growth medium.

Producing infectious foci under agar coating

The culture of BHK-21 cells was plated into wells of a 6 well plate (5x10^5 cells/well) and the plate was incubated for 4 h at 37°C in a CO_2 incubator. Tenfold dilutions of the sample of virus were prepared, starting from 1/10... to 1/10^6. Dilutions were prepared in PBS buffer containing 1% horse serum, the presence of serum is necessary to prevent nonspecific adsorption of the virus. The medium was removed from the wells. 200 µl of each dilution was added to the wells. After 1 hour, the inocula was removed and the wells were filled with a molten agar medium, prepared by mixing of 1 volume of the complete growth medium, 1 volume of 2X MEM medium, 1 volume of molten 1.5% agarose. The medium was allowed to soludify in the wells. The plate was incubated at 37°C in a CO_2 incubator for 120 h, with periodically observations of the monolayers under a fluorescent microscope. After formation of plaques, the monolayers were stained with crystal violet.

Virus titration by Reed-Muench

Titration of the virus using a method of Reed-Muench was done in 96-well plates. BHK-21 cells were seeded into wells (5x10^3 cells per well) and the plate was incubated at 37°C in a CO_2 chamber with periodic monitoring of the monolayers density, until the cells form a monolayer with confluency of ~80%. In eight tubes dilutions of a sample of the virus were prepared, starting from 10^(-1)... to 10^(-8). Dilutions were prepared in PBS + 1% horse serum. The medium was removed from the wells. Dilutions of the virus were added to the wells of the plate, 100 µl of dilution per well. One dilution was used to infect 12 wells in a row. Thus, eight ten-fold dilutions were used to infect rows A-H of the plate. After 1 h, the inocula were removed and the wells were filled with a complete growth medium. The plate was incubated at 37°C in a CO_2 incubator for 48 hours, and at 48 h post-infection the wells were examined for presence of infection. Numbers of infected and uninfected wells in each row (A-H) were counted. Virus titers were counted in ID50 (50% infectious dose) units according to the Reed-Muench formula.

RESULTS AND DISCUSSION

Production of YFV genome

The vaccine strain YFV 17D was used in this work. PCR amplifications using a cDNA as a matrix were done to produce fragments of the YFV genome. The initial cDNA was obtained by reverse transcription of viral RNA using stochastic priming (dN6 primer). Seven pairs of primers were used to amplify the fragments listed in table

1 (pairs 1-7). The PCR products were cloned by T/A-ligation into the pGEM-T vector using a direct cloning kit, which resulted in obtaining of clones listed in table 2.

Table 1. Primers used in this work

Primer ID	Sequence (5'->3')[1]	Place of annealing on the YFV genome[2]	Orientation	Pair of primers
S_NotAcl	ccacgcggccgcGGAAAACGTTAGAGTCGCATA	685-705	S	1
AS_Nsi	GATGATGCATCTCTCTCCAC	1664-1645	AS	1
S_Nsi	AGAGATGCATCATCTTGTCG	1651-1670	S	2
AS_Mlu	GTACACGCGTGTGGTGAACA	2956-2937	AS	2
S_Mlu	CCACACGCGTGTACATGGAC	2943-2962	S	3
AS_Nhe	CTATGCTAGCTGGATCCAAA	5468-5449	AS	3
S_Nhe	TCCAGCTAGCATAGCCGCTA	5455-5474	S	4
AS_Nae	CACAGCCGGCCATTGTGCCC	6710-6691	AS	4
S_Nae	AATGGCCGGCTGTGGATATC	6697-6716	S	5
AS_Aat	GGATGACGTCAGCCTCCAGG	8411-8392	AS	5
S_Aat	GGCTGACGTCATCCTCCCAA	8398-8417	S	6
AS_Cla	TGTCATCGATGGGCCGGACC	9665-9646	AS	6
S_Cla	GCCCATCGATGACAGGTTCG	9652-9671	S	7
AS_SfiXba	ccacggccatctaggccAACCTCTAGACCCCGTCTTTCTA	10717-10695	AS	7
SdelXho[3]	GCCAGATGTTCT(T)GAGAAACTGGAATTG	8200-8227	S	
ASdelXho[3]	CAATTCCAGTTTCTC(A)AGAACATCTGGCATG	8227-8197	AS	

Notes: [1] – Sequences of target sites on the YFV genome shown in capital letters. Linkers containing sites NotI and SfiI shown in lower case letters; NotI and SfiI sites are underlined;
[2] – sequence of the strain 17D virus from vaccine Flavimun (Genbank: JN628280) was used to design primers;
[3] – primers used for site-directed mutagenesis of the YFV genome. In the SdelXho primer, the nucleotide in parentheses is different from the prototype sequence (change C->T). In the ASdelXho primer, the change G->A is shown in parentheses.

Table 2. Clones containing the YFV genome fragments

Clone ID[1]	Length of fragment, bp
pA (NotI,AclI-NsiI)	969
pB (NsiI-MluI)	1288
pC (MluI-NheI)	2512
pD (NheI-NaeI)	1244
pE (NaeI-AatII)	1703

pF (AatII-ClaI)	1251
pH (ClaI-XbaI,SfiI)	1051
Note: [1] – restriction sites at the ends of the fragment indicated in parentheses.	

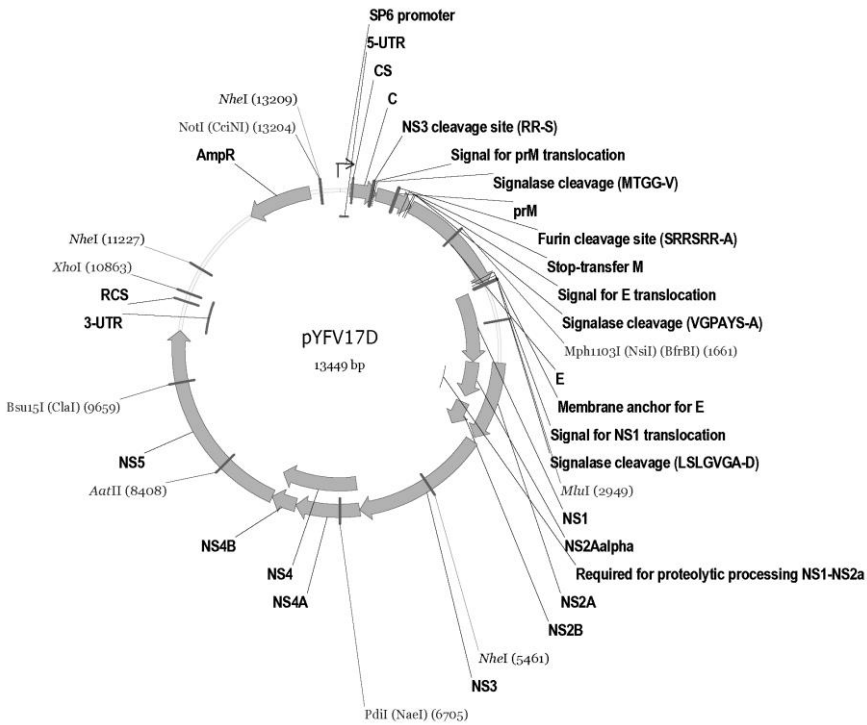
Inserts in the clones pA-pH were sequenced using the M13Forward and M13 Reverse primers, as well as YFV-specific sequencing primers which were synthesized to cover the entire length of the YFV genome. Sequencing of an insert in the clone pE showed that the fragment E contains XhoI restriction site (CTCGAG) which is conserved in various isolates and strains of the YFV. The XhoI site within the YFV sequence was needed to be mutagenized because XhoI restriction enzyme was planned to be used for linearization of the ultimate construct before in vitro transcription. Site-directed mutagenesis was done by a PCR in two rounds. Primers S_NaeI, AS_AatII and additional primers SdelXho and ASdelXho (table 2) were used. Using DNA of the clone pE as a matrix, the first (I) round PCRs were performed: S_NaeI+ASdelXho (1531 bp) and SdelXho+AS_AatII (212 bp). In the second (II) round of PCR, the products 1531 bp and 212 bp were mixed and used to amplify with flanking primers S_NaeI+AS_AatII. The product was cloned into the pGEM-T vector, the resulting clone is designated pEmut. Sequencing the pEmut confirmed that the XhoI site has changed, although the encoded amino acid sequence did not (Leu-Glu).

The inserts in clones pA... pH together cover the central portion of the YFV genome (10018 nt or 92% of a full length of the genome). The inserts were excised from seven plasmids, and for this purpose restriction enzymes NotI+NsiI were used to digest the clone pA; ClaI+SfiI were used to digest the clone pH; other plasmids were digested using restrictases listed in Table 2. A mixture of the excised inserts was ligated in presence of a vector having sticky NotI-SfiI ends. In initial experiments, the vector pGEM-T was utilized for cloning. In these experiments all attempts to obtain a fused fragment A-H failed. The authors suggested that a cause of failure may be instability of the A-H fusion fragment (which resembles a long part of the YFV genome) in the high-copy-number plasmid pGEM. Further attempt was made to assemble the fragment A-H in a different vector, which was chosen from low-copy-number plasmids (pET32a, Novagen). In order to direct this assembly, the original plasmid pET32a was modified. In pET32a, a short linker was engineered between the BlpI and Bsp120I sites, which linker carries only two restriction sites: NotI and SfiI. During this cloning step unwanted sequences of the original vector (such as T7 promoter, a polylinker and T7 terminator) were removed. Insertion of the fusion fragment A-H into the low-copy-number plasmid was successful; it resulted in production of a plasmid designated pAH. Correct assembly of pAH was verified by restriction analysis.

Sequences of the 5'- and 3'-ends of the YFV genome were produced using de novo synthesis of DNA fragments. For this purpose, a DNA fragment which includes a part of the YFV 17D genome from the first nucleotide (A1) to the AclI site (length of 694 nt of the YFV-specific sequence) was synthesized in a constructive PCR. This synthetic fragment contains the 5'-noncoding region of the viral genome (5'-NCR), the gene for capsid protein (C gene) and a part of the prM gene. Also the synthetic fragment includes SP6 RNA promoter (ATTTAGGTGACACTATAG) which is placed immediately upstream of the first nucleotide of the 5'-NCR (A1). Finally, the devised fragment has the 5’-proximal NotI site which was later utilized to insert the synthetic fragment into the plasmid pAH. The synthetic fragment of the 5’-end of the viral genome was inserted using the NotI+AclI sites, and thus the correct 5'-end of a cDNA copy has been built. A resulting plasmid was designated pYFV5AH.

The 3'-end of the YFV genome was also produced as a de novo-synthesized fragment. The latter fragment (length 499 bp) at its 5' end encompasses a part of the 3'-

noncoding region of the YFV genome (3'-NCR), from the XbaI site to the last nucleotide in the viral RNA (T-3'). The fragment was designed to incorporate a sequence of antigenomic ribozyme (RBZ) of the hepatitis D virus (HDV). RBZ is placed immediately downstream of the YFV 3'-terminal nucleotide (T-3'). RBZ sequence in a RNA form has a remarkable enzymatic activity (hence the term ribozyme): this RNA moiety is a highly specific RNase which is capable of cleaving itself from a remainder of RNA molecule. This cleavage affects only one phosphodiether bond, immediately preceding the RBZ sequence (|GGG…, where the bar indicates cleavage). This ribozyme only works *in cis* (i.e. RBZ cuts off itself from the remainder of RNA genome, but the released RBZ does not have enzymatic activities and does not further cleave RNAs). During an in vitro transcription, long RNA molecule is synthesized which contains the full-length YFV genome and is followed with the RBZ element. Self-elimination of the RBZ element leaves the correct 3'-end of the YFV genome. Also, a linker CTCGAGGGCCGCCATGGCC which carries the XhoI and SfiI sites was attached to the 3'-end of the RBZ sequence for a purpose of linearization and cloning. Presence of the XhoI site which is unique in an ultimate construct, allows using XhoI for linearization before the in vitro transcription. The SfiI site was used for cloning: to insert the synthetic 3'-terminal fragment into the plasmid pYFV5AH. Insertion of the latter fragment was done by the XbaI+SfiI sites. A result of the described cloning steps is a plasmid pYFV17D which genetic map is shown in figure 2.



Genes of functional proteins, proteolysis sites, and restriction sites used to assemble the plasmid are depicted.

Fig. 2. Genetic map of the plasmid pYFV17D

The plasmid pYFV17D carries the full-length YFV genome (10862 nt) as the cDNA copy. In the ultimate construct, the cDNA copy is engineered under control of the SP6 RNA-polymerase promoter. This means that the +1 nucleotide from which the RNA-polymerase starts transcription (guanine) immediately precedes the first nucleotide of the YFV genome (adenine). Using the plasmid pYFV17D as a template in

an in vitro transcription reaction, it is possible to synthesize the genomic RNA (10863 nt, including the added G). The RBZ element downstream of the YFV 3'-end cleaves itself off to produce the correct 3'-end.

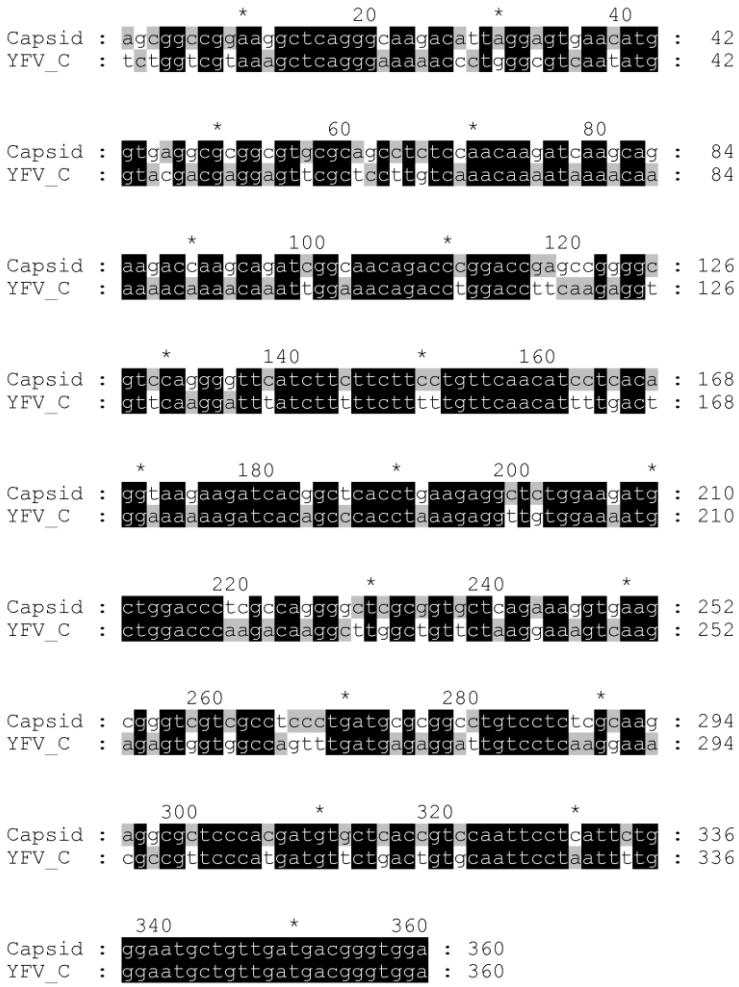
Cloning of a GFP gene into the YFV genome

In order to simplify observation of replication of the virus in cell cultures, a gene encoding green fluorescent protein (GFP) was cloned into the YFV genome (into pYFV17D). The eGFP (enhanced GFP) gene was PCR-amplified from a plasmid in a laboratory collection. With the planned insertion, an important consideration is that positioning of GFP in nonstructural proteins NS1-NS5 which are parts of replicative machinery is very likely to lead to disruption of biological functions of the replicative machine and production of a nonviable virus. Therefore, the insertion of GFP was devised into the genome region encoding structural proteins C-prM-E. The virus needs the structural proteins to produce virions, although the structural proteins are not necessary for replication of the viral RNA. If the insertion of GFP renders a viral protein nonfunctional, the product of insertion into the structural proteins will be a replicon. The term replicon defines a genome which is capable of intracellular replication, but cannot form virions. During proteolytic processing of a YFV polyprotein, junctions between the proteins prM-E and E-NS1 are cleaved by cellular signal peptidase, and this cleavage occurs on the luminal side of membranes of the endoplasmic reticulum. The insertion of GFP between prM and E, or E and NS1, presents a challenge: it is necessary to ensure proper protein processing by signal peptidase. Based upon the described limitations, it was decided to place GFP before of the first (by order in the polyprotein) functional protein which is capsid (C) protein. To ensure synthesis of the functionally active C protein, a sequence of autoprotease 2A of the foot-and-mouth disease virus (FMDV 2A) was positioned between the amino acid sequences of GFP and C. With this design, during translation of RNA-genome of the recombinant virus, the polyprotein GFP-2A-C-prM-E-NS1- ... -NS5 is synthesized. The autoprotease 2A is a peptide (17 aa-long) which interacts with a translating ribosome so that no peptide bond is formed between the C-terminus of the FMDV 2A and the rest of the polyprotein and the GFP-2A fragment is released (GFP-2A| C-…). Termination of transcription does not occur and the ribosome continues to synthesize the rest of the polyprotein (C-NS5).

One important problem which has to be addressed during the insertion is that at the beginning of the C gene highly conserved sequences are present including the so-called cyclization signal (CS motif). In genomes of flaviviruses transmitted by mosquitoes (including the YFV), at the beginning of the reading frame there is a conservative motif (GGCGTCAATATGGT) called the cyclization signal (5'-CS). It was shown that 5'-CS interacts with a reverse complement sequence present in the 3'-NCR at position 10749-10762 nt. As a result of the complementary pairing of the 5'- and 3'-ends of the genomic RNA, this molecule acquires a form of a non-covalently closed ring ("panhandle structure"). Presence of the cyclization signals in their natural positions is an absolute requirement for replication of flaviviruses, including the YFV. To preserve the 5'-CS and neighboring sequences (which may also play a role in the YFV replication), a 75 bp-long sequence beginning the C gene was left in its natural position relative to the 5'-NCR. Accordingly, the GFP-2A aminoacids were inserted after the first 25 aa of the capsid protein.

To avoid possible homologous recombination between the 75 nt fragment of the C gene (before the GFP gene) and the homologous region in the full-length C gene (after the GFP gene), it was decided to change the nucleotide sequence of the full-length C gene by codon optimization. Procedure of codon optimization is replacement of

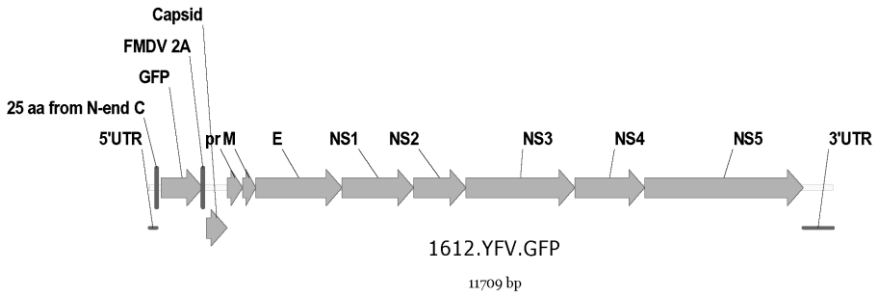
initially present codons by the synonymous codons, providing that an amino acid sequence of an encoded protein is not changed. Here to mention that codon usage in the YFV genes is neither optimal for protein expression in mammalian cells, nor in mosquito cells; this is probably because in a natural infectious cycle the virus must infect both vertebrates and invertebrates. Using the Vector NTI software package, the codon usage in the C gene was optimized to fit to codon frequencies of human genes. The resulting sequence was used to synthesize the artificial C gene. A comparison of the natural and codon-optimized sequences of the C gene is shown in figure 3.



YFV_C, natural gene for capsid(C) protein; Capsid, codon-optimized sequence of the C gene.

Fig. 3. Comparison of natural and codon-optimized sequences of the C gene

The synthetic DNA fragment (designated as the ‘Capsid’ in Figure 3) was used to replace the natural Cgene. An ultimate result of the described genetic engineering is a plasmid p1612.YFV.GFP. A genetic map of the virus 1612.YFV.GFP is shown in figure 4.

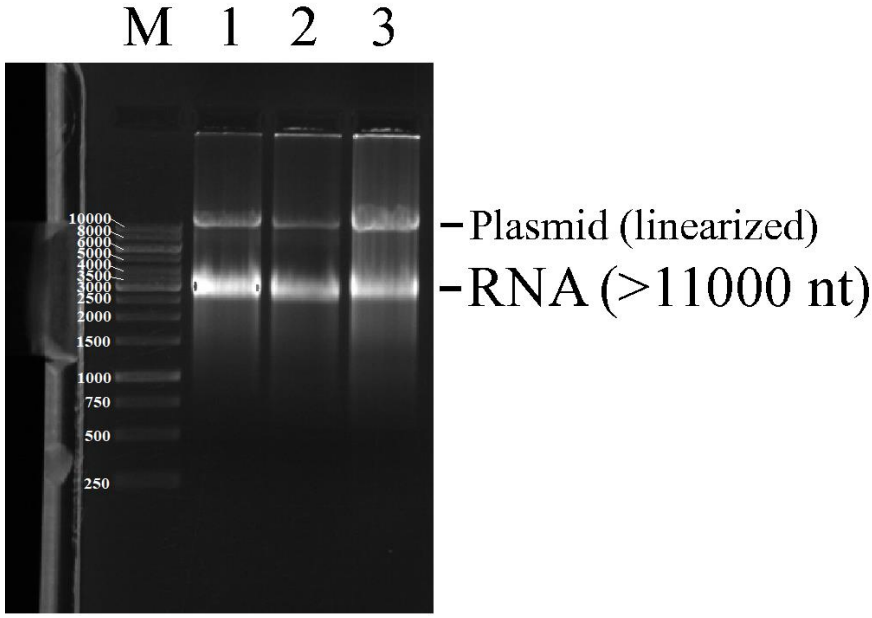


A codon-optimized version of C gene (designated 'Capsid') is used in place of the natural C gene to prevent recombination between sequences present upstream and downstream of the GFP gene.

Fig. 4. Genetic map of the genome 1612.YFV.GFP

Rescue of the 1612.YFV.GFP virus

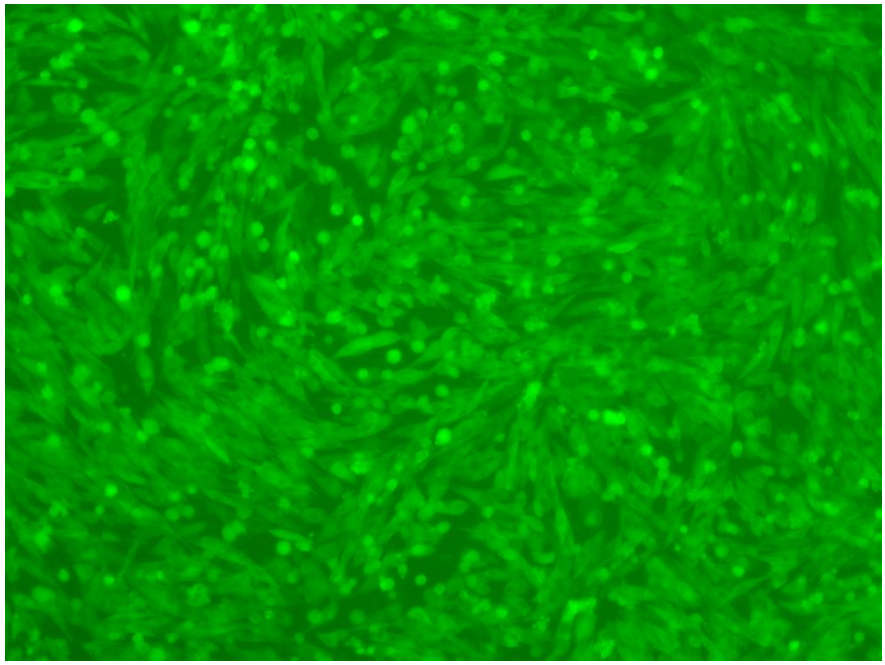
RNA of the virus 1612.YFV.GFP was synthesized. After in vitro transcription, the reaction mixture was separated by electrophoresis in a 1% agarose gel (figure 5). In an agarose gel under non-denaturing conditions, RNA moves like a band with an apparent mobility ~3000 bp (compared to a ladder of DNA fragments); ssRNA has higher mobility compared to dsDNA molecules of similar sizes because of more compact 3D packing of the RNA molecule.



Lanes: M, O'GeneRuler 1 kb (Fermentas); 1, RNA 1612.YFV.GFP; 2-3, two other synthetic RNAs.

Fig. 5. RNA synthesized using in vitro transcription

Synthetic genomic RNA of the virus 1612.YFV.GFP was transfected into a culture of BHK-21 cells using cationic liposomes (Lipofectamine 2000); the transcription mixture was used for transfections without purification. The first visual signs of virus infection (GFP fluorescence) appeared at 24 h post transfection (p.t.). By 36 h p.t. all cells were GFP-positive. A picture of the infected monolayer at 48 h p.t.is shown in figure 6.

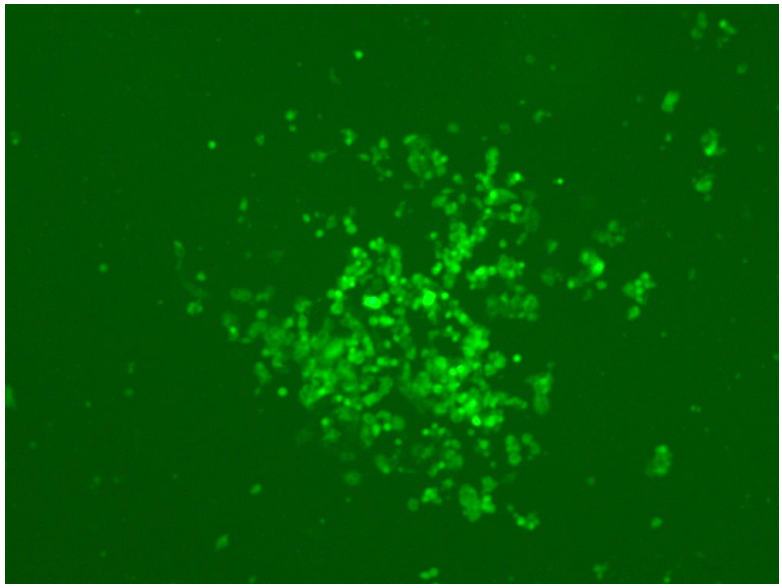


All cells show fluorescence of GFP. The photo was taken 48 hours after transfection.

Fig. 6. Photo of culture BHK-21 transfected the 1612.YFV.GFP virus' RNA

Virus 1612.YFV.GFP is cytopathic

Further observation of the transfected culture showed that the virus has a cytopathic effect (CPE). Large fraction of cells in the culture died by 72 h p.t. Upon transfection, samples of culture medium from the transfected culture were collected every 24 hours, and the media in culture vessels were replaced with a complete medium. The collected samples were used to produce infectious foci under an agar coating. Serial dilutions of each sample (from 1/10 to 1/10^6) were prepared and each dilution was used to infect a monolayer of BHK-21 cells grown in a well of a 6-well plate. The infected monolayers were covered with a molten agar-containing medium. Under the agar coating, the virus can spread only from an infected cell to neighboring cells, and thus groups of infected cells (infectious foci) are formed. A photograph of one infectious focus is shown in figure 7.



The photo was taken 48 hours after infection.

Fig. 7. Infectious focus in a monolayer grown under agar coating

Because of a cytopathic effect (CPE) of the virus, a fraction of cells in foci die and detach from the plate. Thus, in monolayers zones free of cells (plaques) are formed. At 120 h after infection the agar coatings were removed and the plates were stained with crystal violet. A photograph of the plaques is shown in figure 8.

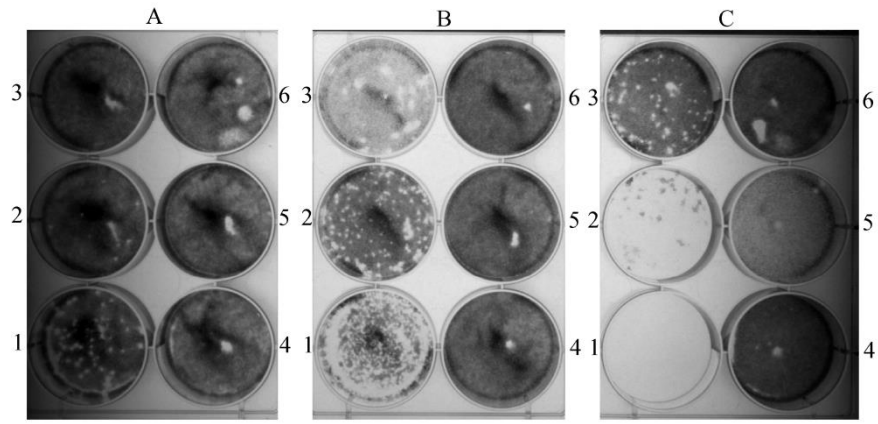


Plate A, wells were infected with the virus 1612.YFV.GFP collected fat 24 h p.t.; plate B, infection with the virus collected at 48 h p.t.; plate C, infection with the virus collected at 72 h p.t. Wells (numbered 1-6) were infected with dilutions from 1/10 (well 1) to 1/10^6 (well 6).

Fig. 8. Plaques formed under agar coating

During incubation of the transfected culture for 3 days the virus propagates to infect all cells in the culture. As more cells start to produce the virus' progeny, the virus titers grows in sequential samples. The increase in the titers is evident in three plates presented at figure 8.

Nevertheless, an investigation of the infected monolayers under a fluorescent microscope showed that counting of plaques does not give correct estimates of the titers. In used cell line BHK-21, the virus 1612.YFV.GFP did not kill 100% of infected

cells. Many GFP+ foci were composed from live cells which were not supposed to detach and make visible plaques. Here be noted that CPE of a given virus is highly dependent on a used cell culture (also cell line and a subclone). It is possible that BHK-21 cells from the authors' collection have some degree of resistance to the CPE of the YFV. To accurately measure amounts of the virus in the collected samples, we decided to use a different method of titering which is very reliable, i.e. method of Reed-Muench.

Determination of virus titers

The Reed-Muench method implies infecting of a large number of wells (in a 96-well plate) with dilutions of the virus, including dilutions at a physical limit (~1 or <1 infectious units in a volume of inoculum). When a dilution contains ~1 of the infectious units (virion) per well, a probability of the well being infected follows Poisson distribution. It is possible to calculate a dilution for which 50% of wells are infected. A number inverse to this dilution gives the virus titer, in units of 50% infectious dose (ID50). For flaviviruses, titers by Reed-Muench (in ID50) are in good agreement with titers determined by other methods, including a physical titer (a number of virions in 1 ml). The following titers were determined for the 1612.YFV.GFP virus using the Reed-Muench method (figure 9).

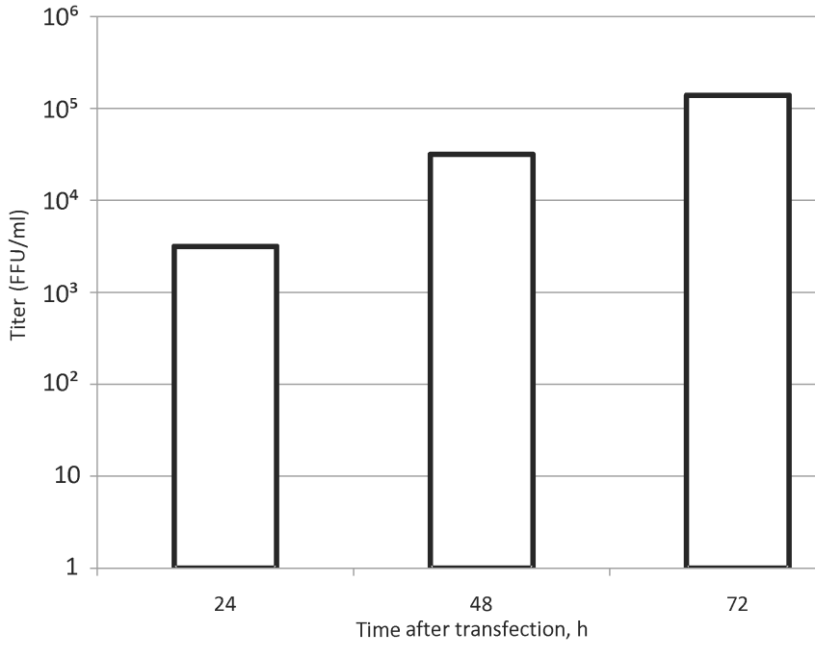


Fig. 9. Accumulation of the virus 1612.YFV.GFPin BHK-21 culture transfected with the viral RNA

CONCLUSION

A cDNA copy of the full-length genome of the yellow fever virus (YFV) was assembled from DNA fragments. Into the YFV genome, in a position preceding genes for structural proteins, a GFP gene was inserted, to provide convenient means to track infection in vivo. The virus expressing GFP during intracellular replication was rescued from the cDNA copy. The YFV is a model representative virus in the genus Flavivirus and the YFV is often used in molecular virology studies. The produced cDNA can be used in molecular virology studies and for genetic engineering, including creation of live vaccines based on the YFV vector.

Acknowledgements

This work was supported by the Ministry of Education and Science of Kazakhstan (grant №AP05130607).

REFERENCES

1. Yellow fever. Available at: http://www.who.int/news-room/fact-sheets/detail/yellow-fever.

2. Eliminate Yellow fever Epidemics (EYE): a global strategy, 2017-2026. Available at: http://apps.who.int/iris/bitstream/handle/10665/272408/9789241513661-eng.pdf.

3. Yellow fever outbreak Angola, Democratic Republic of the Congo and Uganda 2016-2017. Available at: http://www.who.int/emergencies/yellow-fever/en/.

4. Hay J., Yeh K.B., Dasgupta D., Shapieva Z., Omasheva G., Deryabin P., Nurmakhanov T., Ayazbayev T., Andryushchenko A., Zhunushov A., Hewson R., Farris C.M., Richards A.L. Biosurveillance in Central Asia: Successes and Challenges of Tick-Borne Disease Research in Kazakhstan and Kyrgyzstan. *Front. Public Health.*, 2016, no. 4, pp. 35-40.

5. Grard G., Moureau G., Charrel R.N., Lemasson J.J., Gonzalez J.P., Gallian P., Gritsun T.S., Holmes E.C., Gould E.A., de Lamballerie X. Genetic characterization of tick-borne flaviviruses: new insights into evolution, pathogenetic determinants and taxonomy. *Virology*, 2007, no. 25, pp. 80-92.

6. Lim S.K., Lim J.K., Yoon I.K. An Update on Zika Virus in Asia. *Infect. Chemother*, 2017, no. 49, pp. 91-100.

7. Matthias Kalitzky, Peter Borowski. *Molecular Biology of the Flavivirus.* Horizon Scientific Press, 2006, 380 p.

8. Rastogi M., Sharma N., Singh S.K. Flavivirus NS1: a multifaceted enigmatic viral protein. *Virol. J.*, 2016, no. 13, pp. 131. https://doi.org/10.1186/s12985-016-0590-7.

9. Tajima S., Takasaki T., Kurane I. Characterization of Asn130-to-Ala mutant of dengue type 1 virus NS1 protein. *Virus Genes*, 2008, no. 36, pp. 323-329.

10. Fan J., Liu Y., Yuan Z. Critical role of Dengue Virus NS1 protein in viral replication. *Virol. Sin.*, 2014, no. 29, pp. 162-169.

11. Edeling M.A., Diamond M.S., Fremont D.H. Structural basis of Flavivirus NS1 assembly and antibody recognition. *Proc. Natl. Acad. Sci. USA*, 2014, no. 111, pp. 4285-4290.

12. Youn S., Li T., McCune B.T., Edeling M.A., Fremont D.H., Cristea I.M., Diamond M.S. Evidence for a genetic and physical interaction between nonstructural proteins NS1 and NS4B that modulates replication of West Nile virus. *J. Virol.*, 2012, no. 8, pp. 7360-7371.

13. Youn S., Ambrose R.L., Mackenzie J.M., Diamond M.S. Non-structural protein-1 is required for West Nile virus replication complex formation and viral RNA synthesis. *Virology Journal*, no. 10, pp. 339. http://dx.doi.org/10.1186/1743-422x-10-339.

14. Chatel-Chaix L., Fischl W., Scaturro P., Cortese M., Kallis S., Bartenschlager M., Fischer B., Bartenschlager R. A combined genetic-proteomic approach identifies residues within Dengue virus NS4B critical for interaction with NS3 and viral replication. *J. Virol.*, 2015, no. 89, pp. 7170-7186.

15. Scaturro P., Cortese M., Chatel-Chaix L., Fischl W., Bartenschlager R. Dengue Virus Non-structural Protein 1 Modulates Infectious Particle Production via Interaction

with the Structural Proteins. *PLoS Pathog.*, 2015, no. 11. 26562291.

16. Wu R.H., Tsai M.H., Tsai K.N., Tian J.N., Wu J.S., Wu S.Y., Chern J.H., Chen C.H., Yueh A. Mutagenesis of Dengue Virus Protein NS2A Revealed a Novel Domain Responsible for Virus-Induced Cytopathic Effect and Interactions between NS2A and NS2B Transmembrane Segments. *J. Virol*., 2017, no. 91. 28381578.

17. Leung J.Y., Pijlman G.P., Kondratieva N., Hyde J., Mackenzie J.M., Khromykh A.A. Role of nonstructural protein NS2A in flavivirus assembly. J. Virol., 2008, no. 82, pp. 4731-4741.

18. Xie X., Gayen S., Kang C., Yuan Z., Shi P-Y. Membrane Topology and Function of Dengue Virus NS2A Protein. *J Virol.*, 2013, no. 87, pp. 4609-4622.

19. Kümmerer B., Rice C.M. Mutations in the yellow fever virus nonstructural protein NS2A selectively block production of infectious virus. *J. Virol.*, 2002, no. 76, pp. 4773-4784.

20. Li K., Phoo W.W., Luo D. Functional interplay among the flavivirus NS3 protease, helicase, and cofactors. *Virol. Sin.*, 2014, no. 29, pp. 74-85.

21. Luo D., Vasudevan S.G., Lescar J. The flavivirus NS2B-NS3 protease-helicase as a target for antiviral drug development. *Antiviral. Res*., 2015, no. 118, pp. 148-158.

22. Junaid M., Angsuthanasombat C., Wikberg J.E., Ali N., Katzenmeier G. Modulation of enzymatic activity of dengue virus nonstructural protein NS3 nucleoside triphosphatase/helicase by poly(U). *Biochemistry (Mosc)*, 2013, no. 78, pp. 925-932.

23. Zou J., Xie X., Wang Q.Y., Dong H., Lee M.Y., Kang C., Yuan Z., Shi P.Y. Characterization of dengue virus NS4A and NS4B protein interaction. *J. Virol.*, 2015, no. 89, pp. 3455-3470.

24. Apte-Sengupta S., Sirohi D., Kuhn R.J. Coupling of replication and assembly in flaviviruses. *Curr. Opin. Virol.*, 2014, no. 9, pp. 134-142.

25. Zmurko J., Neyts J., Dallmeier K. Flaviviral NS4b, chameleon and jack-in-the-box roles in viral replication and pathogenesis, and a molecular target for antiviral intervention. *Rev. Med. Virol.*, 2015, no. 25, pp. 205-223.

26. Klema V.J., Ye M., Hindupur A., Teramoto T., Gottipati K., Padmanabhan R., Choi K.H. Dengue Virus Nonstructural Protein 5 (NS5) Assembles into a Dimer with a Unique Methyltransferase and Polymerase Interface. *PLoS Pathog*., 2016, no. 12. 26895240.

27. Surana P., Satchidanandam V., Nair DT. RNA-dependent RNA polymerase of Japanese encephalitis virus binds the initiator nucleotide GTP to form a mechanistically important pre-initiation state. *Nucleic Acids Res*., 2014, no. 42, pp. 2758-2773.

28. Zhao Y., Soh T.S., Zheng J., Chan K.W., Phoo W.W., Lee C.C., Tay M.Y., Swaminathan K., Cornvik T.C., Lim S.P., Shi P.Y., Lescar J., Vasudevan S.G., Luo D. A crystal structure of the Dengue virus NS5 protein reveals a novel inter-domain interface essential for protein flexibility and virus replication. *PLoS Pathog*., 2015, no. 11. 25775415.

29. Teramoto T., Balasubramanian A., Choi K.H., Padmanabhan R. Serotype-specific interactions among functional domains of dengue virus 2 nonstructural proteins (NS) 5 and NS3 are crucial for viral RNA replication. *J. Biol. Chem.*, 2017, no. 292, pp. 9465-9479.

30. Le Breton M., Meyniel-Schicklin L., Deloire A., Coutard B., Canard B., de Lamballerie X., Andre P., Rabourdin-Combe C., Lotteau V., Davoust N. Flavivirus NS3 and NS5 proteins interaction network: a high-throughput yeast two-hybrid screen. *BMC Microbiol*, 2011, no. 11, pp. 234. https://doi.org/10.1186/1471-2180-11-234.

31. Azevedo A.S., Gonçalves A.J., Archer M., Freire M.S., Galler R., Alves A.M. The synergistic effect of combined immunization with a DNA vaccine and chimeric yellow fever/dengue virus leads to strong protection against dengue. *PLoS One*., 2013, no. 8. 26678487.