RECOMMENDED CITATION
Cabral GB, Ferreira JLP, de Souza RP, Cunha M, Figueiredo C, Brígido LFM. Simple protocol for population (Sanger) sequencing for Zika virus genomic regions [Submitted]. Mem Inst Oswaldo Cruz E-pub: 11 July 2017. doi: 10.1590/0074-02760170248.

Running title: Sanger protocol for Zika Virus

Simple Protocol for population (Sanger) sequencing for Zika virus genomic regions

Gabriela Bastos Cabral[1], João Leandro de Paula Ferreira[1], Renato Pereira de Souza[2], Mariana Cunha[2], Cristina Figueiredo[3], Luís Fernando de Macedo Brígido[1*]

1 NDSS, Retrovirus Laboratory, Virology Center, Adolfo Lutz Institute, São Paulo, Brazil.

2. NDTV, Virology Center, Adolfo Lutz Institute, São Paulo, Brazil

3. NDR, Virology Center, Adolfo Lutz Institute, São Paulo, Brazil

Disclaimer: The authors did not identified any potential conflicts of interest

[*] **Contact information for corresponding author**

Ave. Dr. Arnaldo 355, 01246-902

Virology Center, Retrovirus Laboratory, Sao Paulo, Brazil

Fone +55 11 30682983

E-mail address: lubrigido@gmail.com

Word Count: Abstract, 286; Manuscript 2096

Abstract

The Flaviviridae Zika (ZIKV), first identified in Uganda, initially noticed in Brazil in 2015, was probably introduced in the Americas from Oceania. The pathogenic potential associated to the Brazilian outbreak made the recognition and characterization of this infection a priority. Genetic sequences from ZIKV has allowed to study the molecular signatures of the agent and trace its expansion. Next-generation sequencing (NGS) is increasingly used to understand the genetic diversity of different agents, but sanger sequencing may be an alternative for many applications as it is an easy to use, robust, affordable, rapid, and specific tool to obtain sequences. **Objective:** The aim of this study is to develop a simple protocol of populational (sanger) genetic sequencing of the envelope and NS5 regions of the Zika virus **Materials and methods:** Six isolates of Zika virus, obtained from supernatant of infected Vero cells and clinical sample (urine) were evaluated in this study. One of the samples was previously sequenced by NGS. A one-step PCR reaction were performed for the amplification of complete envelope and partial NS5 region of the ZIKV using SuperScript®III One-step RT-PCR system with Taq High Fidelity, followed by a cycle sequence reaction using big dye terminator. **Finding:** The method was effective for amplification and obtaining the genetic sequences of the 6 samples for envelope region and 5 from NS5 region. Although sequences from NGS can provide additional information, nucleotide alignments and phylogenetic trees topology show that Sanger sequencing may be as informative as NGS for some applications. **Conclusion**: The present study provides a simple protocol to amplify and sequence segments of the Zika genome. As currently only a few zika sequences are available, this method can facilitate the production of additional sequences in resource limited settings.

Key words – Zika virus, sanger sequencing, PCR, phylogeny

Nº Finep: 0219/2016

Introduction

Zika virus (ZIKV) is a member of the genus Flavivirus, of the Flaviviridae family. ZIKV was first isolated in Uganda, in 1947, from a sentinel rhesus monkey. Since then, sporadic cases of human infection and isolation from mosquitoes of the genus Aedes has been reported in Africa and Asia. Zika virus has been considered an emergent pathogen since 2007, when an epidemic was reported in Micronesia (Faria et al. 2016).

Zika virus was first identified in the Americas in March 2015, during an outbreak of exanthematic disease in the State of Bahia, Brazil (Campos et al., 2015). In September 2015, investigators in Brazil observed an increase in the number of infants born with microcephaly in areas were Zika virus had been reported and, by mid-February 2016, more than 4300 cases of microcephaly had been notified. Due to technical limitations in serological tests, diagnosis of Zika virus infection in humans is mainly based on RNA detection in serum or plasma and in urine samples, where the viral RNA can be found longer than in plasma. Specific antibody detection is mostly hampered due to serological cross-reactivity with other circulating flaviviruses such as dengue virus or yellow fever virus (Lanciotti RS et al. 2008, Tappe et al. 2014, Tappe et al. 2015).

Studies of molecular epidemiology have shown that the epidemic ZIKV strain in Brazil belongs to the Asian lineage. Most cases the Zika genetic sequences were obtained from next-generation sequencing. Molecular evaluation is important for monitoring viral evolution, vaccine development; improvement of diagnostic

assays, as well as it may help in establishing non vectorial, especially sexual transmission networks. Although NGS may provide details of the quasiespecies diversity of RNA viruses, NGS is also used for the generation of a consensus, major variant sequence that can be used in phylogenetic studies. This major variant can also be obtained by population 9sanger) sequencing. The aim of this study is to develop a simple, robust protocol for genetic sequencing of important genetic regions of the Zika virus, as the envelope and NS5 regions, by Sanger and evaluate if phylogenetic information generated by sequences produced with this protocol, including tree topology, are comparable to that obtained by previous published NGS sequences.

Materials and Methods

Collection of samples

Six isolates of Zika virus, obtained from supernatant of infected Vero cells and one clinical sample (urine) were evaluated in this study. Among those, two of the samples were from a donor and recipient pair of a reported zika virus transmission at a plasma transfusion, the recipient sample previously sequenced by NGS (Barjas-Castro, 2016). Viral isolation was performed from 20 µl of each serum samples, isolated into flasks containing monolayer cultures of C6/36 cells. The cell culture was incubated for nine days at 28 °C. Indirect immunofluorescent antibody (IFA) tests were performed using flavivirus polyclonal antibodies, as described (Glubler 1984).

Nucleic acid extraction

The sample obtained from urine was extracted in duplicate using the same protocol described below, used directly and after concentration of 1mL of urine by centrifugation (21,000 x g) for 1 hour at 4^{o}C.

The Zika viral RNA was extracted from cell culture and urine by QIAGEN (QIAmp® viral RNA mini kit, Qiagen, Hilden, Germany) according manufacture's protocol. The amplification of ZIKA envelope and NS5 partial region was performed by a nested PCR, with specific primers designed for optimizing the amplification of each region, based on reference sequences available in Genbank (KU647676.1, KU509998.1, KU681082.2, KJ776791.1, KR815990.1, KR815989.1, KR816336.1, KU497555.1, KU365778.1, KU365777.1, KU365779.1, KU365780.1, KU232301.1, KU232300.1, KU232298.1).

The protocol allowed the amplification of the entire of Zika virus envelope region (nucleotide position 873-2370 related Genbank accession number NC_012532.1) and, partial amplification of NS5 region, (nucleotide position 8958-9756 related Genbank accession number NC_012532.1). The primers designed for the envelope included: first step: Zika1_out_Forward: AGCAGCAGCTGCCATCGCTTG (777 – 797 pb) and Zika2_out_Reverse: GTACCT GTCCCTCCAGGCTTC (2478 – 2458 pb), yielding a fragment with a 1,701 Kb. For the nested PCR, the primers used were Zika3_Inner_Foward: GATACTGCTGATTGCCCCGGCATA (843 – 866 pb) and Zika4_Inner_Reverse:

TTCTTTGAGAAGTCCACCGAGCAC (2414 – 2391 pb), generating a fragment of 1,571 Kb. The Zika NS5 partial region primers used in the first, one-step PCR, were Zika1_out_foward: (8891–8910pb): TGAGAGGAGAGTGCCAGAGT and Zika2_out_reverse: (9843 – 9864): ATAAAGGAGCTGCCACATTTG with 973 pb. The nested PCR, the primers used were Zika3_inner_foward (8958–8976): TGGAAAGGCCAAGGGCAGC and Zika4_inner_Reverse (9736 – 9756): GTGGCGGCAGGGAACCACAAT, generating a fragment of 0,798 Kb.

The ZIKA envelope and NS5 partial amplification

For the amplification of complete envelope and partial NS5 region of the ZIKV, a RT-PCR reaction, using SuperScript® III One-step RT-PCR system with Platinum Taq High Fidelity (Life Technologies, USA) was performed with 10 µl of viral RNA, 10 µM of each primers (forward and reverse) and water, resulting in a 50 µl final volume of the reaction. For the amplification of the envelope coding region we started with a step of retrotranscription (55°C for 30min and 94°C for 5min), followed by 35 amplification cycles: 18 cycles at 94°C for 30s, 56°C for 30 s, 68°C 2min30s; 17 cycles at 94°C for 30s, 60°C for 30 s, 68°C 2min30s and a final extension at 68°C for 10 min. For the amplification of partial NS5 coding region, we started with a step of retrotranscription (50°C for 30min and 94°C for 5min), followed by 35 amplification cycles at 94°C for 30s, 53°C for 30 s, 68°C 1min30s and a final extension at 68°C for 10 min.

The Nested-PCR for both regions were performed with Go Taq® Green Master Mix

2X (Promega Biosciences, CA), 10 µM of each primers and 2.5 µl volume of the product of the first amplification. The reactions were carried out at 94°C for 3min followed by 35 cycles of at 94°C for 30 s, 55°C for 30 s, 72°C 2min and a final extension at 72°C for 10min.

Cycle sequence of envelope and partial NS5 region Zika Virus

A unique fragment, of approximately 1,571 bp was obtained after amplification of envelope coding region. It was sequenced using eight primers to cover the entire envelope region. Four primers were designed to cover the 800 bp unique fragment obtained from amplification of partial NS5 region. Primers to envelope and NS5 region were detailed in table 1.

Each sequencing reaction was performed using 0,5 µl of BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems) and 1,6 µl for each primer (1µM) in 10 µl final volume of the reaction. To cover the envelope and partial NS5 regions 8 and 4 primers were used, respectively (table1). Ethanol/acetate was used for precipitation. Sequence chromatograms were obtained on automated sequence analyzers ABI 3130XL (Applied Biosystems). All amplicons were sequenced on both directions. Sequences were manually edited using Sequencher 4.7 software (Gene Codes, USA).

Ethical approval

This study was approved by Ethical and Research Committees of Adolfo Lutz Institute, Secretary of Health of São Paulo, Brazil and participating institutions.

Results

The protocol proved effective for Zika virus sequencing from both cell culture supernatants and clinical samples. Seven sequences of complete envelope (MF048802, MF048803, MF048804, MF048805, MF048806, MF048807 and MF048808 GenBank accession number) and six sequences, of partial NS5 region of the Zika virus were obtained in this study (MF077458, MF077459, MF077460, MF077461, MF077462 and MF077463 GenBank accession number). The sequences showed 1515 and 667 base pairs, respectively for envelope and NS5 Zika regions. These sequences were aligned using BioEdit and showed the envelope region was more conserved than the NS5 sequences , with a significant lower percentage of nucleotide substitutions (0.26 x 0.95, respectively p =0.042, Fisher two-tailed). Moreover, two sequences (MF077463 and MF077459) presented ambiguities (R = A or G and Y= C or T) in the NS5 region at positions 495 and 196, respectively. The ambiguity at sample MF077463 is synonymous, with both nucleotides coding for a L(lysine) at position 165 at the NS5 protein whereas the ambiguity R in the sample MF077459 leads to a non-synonymous amino acid substitution at positions 66 (coding for Histidine or Tirosine) (Figure 1a and 1b).

The sequences, alignment in Bioedit, could be grouped by homology in 3 groups congruent for both env and NS5 region: ENV: A) MF048804 - MF048805, B) MF048806 and C) MF048802, MF048803, MF048807 and MF048808 (urine sample). For NS5: A) MF077560 - MF077561) B) MF077462 and C) MF077463, MF077458 and MF077459.

According to the identity score (BLAST/NCBI) were observed the group A showed 100% of identify with other three samples from Brazil, two of which from the same sample, sequenced by NGS and previously published (KR872956.1, KX321639.1 and KX197192.1GenBank accession number); Group B presented 99% of identify with more than 20 sequences, including sequences form French Guiana and Puerto Rico in the envelope and NS5, respectively (KU758871.1 and KU758870.1; KX087101.3, KY075934.1, KX601168.1 GenBank accession number); Group C were observed homology in the envelope region of 100% with 18 sequences: 4 from Nicaragua (KY765327.1, KY765326.1, KY765325.1, KY765324.1), 2 from Honduras (KX262887, KY785414.1), 6 from French Polinesia (KX447520.1, KX447519.1, KX447518.1, KX447513.1, KX447510.1 and KX369547.1), 2 from Florida, USA (KY325479 and KY325465.1) and 3 from Brazil (KY014313.2, KX576684.1, KX280026.1 and KX81122.1). The NS5 region presented homology of 99% with more than 30 sequences from America Central (KY693679.1 – Peru, KY785452.1 – Honduras, KY606272.1 – Mexico) and Brazil (KY811222.1, KY576684.1, KY280026.1 and KY559015.1).

These results corroborate previously published data about the spread of the ZIKV Asian lineage that was first reported in the outbreaks of the Pacific Islands (Musso et al. 2014). Phylogenetic trees (Figure 2a, envelope and 2b, NS5 partial regions) may illustrate these findings.

We further evaluate the informative potential of our population sequences of partial genome using envelope region of complete NGS sequences along our sanger generated sequences, to document if these regions were useful to provide phylogenetic information to use in molecular epidemiological studies. For that end we compared our sequences to those obtained by NGS sequencing. We included a sample previously sequenced by NGS (Genbank acession gbKU321639.1) in a study to document probable transfusion-transmission Zika virus in Brazil (Barjas-Castro, 2016). We evaluated here the recipient-donor pair. The sequences derived by Sanger (receptor / donor), extracted and amplified more than one separate reaction, all showed a phylogenetic similarity that support its relationship described in the work Barjas-Castro (figure 3). The complete envelope (1515 pb) was sufficient to support the relationship between samples. In the original work, NGS was performed only for receptor sample and a homology of 99.8% between the sequence pair was performed using ten partial nucleotide sequences from the donor's ZIKV strain, with lengths ranging from 173 to 427 base pairs for each fragment. In our envelope and NS5 segments, we observed a homology of 100%.

The phylogeny tree below (figure 3) shows the relation between the same sample obtained by NGS and Sanger.

Discussion

The present study provides a simple sanger protocol to amplify segments of the Zika genome and present evidence that suggests that the genetic information from about 20% of its genome (Full envelope and partial NS5) may provide phylogenetic signal that allows molecular epidemiological studies useful to monitor viral evolution. One major limitation of the study was limitations in the availably of RNA from confirmed samples, and it is important to further evaluate the performance directly from additional clinical samples, as blood and urine. The protocol is of a relative low cost with the use of a lower input of most expensive items, as Big dye solution, even using more expensive high fidelity taq and superscript III and considering market prices, the cost of molecular biology reagents were below US$ 80 for envelope and US$ 50 for NS5 regions.

The majority of Zika sequences available is provide by NGS, this technic may provide information on viral diversity and are pivotal in the analysis of quasiespecies extension and structure. However this tool is cost effective in specialized, core laboratories working with high quality samples and adequate bioinformatics support, a situation not commonly available in clinical and public health laboratories, especially in resource constrained settings. Although population sequencing commonly misses minor variants present in the diverse RNA viruses quasiespecies population, sanger sequencing is an alternative, easy to use, robust, affordable, rapid, and specific tool to obtain sequences from the major variant. It may therefore be an important alternative methodology to NGS.

There are currently only a few available zika sequences despite the large number of cases since the beginning of the epidemic and the method here described can facilitate the acquisition of genetic sequences.

Acknowledgments

We thank to the Adolfo Lutz Institute and FINEP for providing funds to do this research.

Sequence data

Sequences are available at GenBank with accession numbers: MF048802 to MF048807 and MF077458 to MF07763.

Author Disclosure Statement

The authors have not identified any potential conflict of interest

Funding

FINEP: 0219/2016

References

Barjas-Castro ML, Angerami RN, Cunha MS, Suzuki A, Nogueira JS, Rocco IM, Maeda AY, Vasami FG, Katz G, Boin IF, Stucchi RS, Resende MR, Esposito DL, de Souza RP, da Fonseca BA, Addas-Carvalho M. Probable transfusion-transmitted Zika virus in Brazil. Transfusion. 2016 Jul;56(7):1684-8.

Campos GS, Bandeira AC, Sardi SI. Zika Virus Outbreak, Bahia, Brazil.Emerg Infect Dis. 2015; 21(10):1885-6.

Faria NR, Azevedo Rdo S, Kraemer MU, Souza R, Cunha MS, Hill SC, et al. Zika virus in the Americas: Early epidemiological and genetic findings. Science. 2016; 352(6283):345-9.

Gubler DJ, Kuno G, Sather GE, Velez M, Oliver A. Mosquito cell cultures and specific monoclonal antibodies in surveillance for dengue viruses. Amer.J. trop. Med. Hyg. 1984; 33:158-165.

Lanciotti RS, Kosoy OL, Laven JJ, Velez JO, Lambert AJ, Johnson AJ, Stanfield SM, Duffy MR. Genetic and serologic properties of Zika Virus associated with and epidemic, Yap State, Micronesia, 2017. Emerg Infect Dis. 2008; 14(8): 1232-9.

Musso D, Rouault E, Teissier A, Lanteri MC, Zisou K, Broult J, Grange E, Nhan TX, Aubry M. Molecular detection of Zika virus in blood and RNA load determination during the French Polynesian outbreak. J Med Virol. 2016.

Tappe D, Nachtigall S, Kapaun A, Schnitzler P, Günther S, Schmidt-Chanasit J. Acute Zika virus infection after travel to Malaysian Borneo, September 2014.
Emerg Infect Dis. 2015; 21(5):911-3.

Tappe D, Rissland J, Gabriel M, Emmerich P, Gunther S, Held G, Smola S, Schmidt-Chanasit J. First case of laboratory-confirmed Zika virus infection imported into Europe. Euro Surveill. 2014; 19(4).

Table **1.** Primers used for cycle sequencing of Envelope and NS5 regions of Zika Virus

Primers	Position	Region
EZ3IFS: gAT ACT gCT gAT TgC CCC ggC ATA	843 – 866	Envelope
EZ5IFS: ATg ACC ggg AAg AgC ATC CAg	1243 - 1263	Envelope
EZ6IFS: Agg CAA ACT gTC gTg gTT CTA	1624 - 1644	Envelope
EZ7IFS: CTT ACA TTg TCA TAg gAg TCg	2024 - 2044	Envelope
EZ4IRS: TTC TTT gAg AAg TCC ACC gAg CAC	2414 – 2391	Envelope
EZ8IRS: g TCC CCA AAT ggT ggA TCA AgT	2001 - 2022	Envelope
EZ9IRS: TgC gTC CTT gAA CTC TAC CAg	1594 - 1614	Envelope
EZ10IRS: CT CCC TTT gCC AAA AAg TCC ACA	1183 - 1205	Envelope
ZNS3IFS: Tgg AAA ggC CAA ggg Cag C	8958 - 8976	NS5
ZN5IFS: CAg TCA gTg gAg ATg ATT gC	9542 - 9562	NS5
ZNS4IRS: gTg gCg gCA ggg AAC CAC AAT	9736 - 9756	NS5
ZN6IRS: TgT CCg CTC CCC CTT Tgg TCT	9354 - 9374	NS5

Figure Legends

Fig1. Chromatogram of the NS5 region of zika virus showing the ambiguities found in the samples (MF077463 and MF077459) of the study. **A)** Chromatogram and alignment (nucleotide and amino acid) of part of the sequence showing the ambiguity (R = A or G) in the sample MF077463 at position 495 and this ambiguity is synonymous, with both nucleotides coding for a L(lysine) at position 165, **B)** Chromatogram and alignment (nucleotide and amino acid) of part of the sequence showing the ambiguity (Y = C or T) in the sample MF077459 at position 196 leads to a non-synonymous amino acid substitution at positions 66 (coding for Histidine or Tirosine).

Fig2. Bayesian phylogenetic tree of Zika Virus Envelope and NS5 genes

sequences. **A)** Bayesian phylogenetic tree of envelope gene sequences of Brazil ZIKV using GTR+G+I model with seven (isolated from cell culture and one from urine) Brazilian ZIKV strains obtained in this study (MF048802, MF048803, MF048804, MF048805, MF048806, MF048807 and MF04880) and 17 reference ZIKV sequences from GenBank (two references from the African lineage [accession numbers:HQ234500.1 djLC002500.1], six references from the Asian lineage [HQ234499.1, EU545988.1, KU681082.1, KU509998, KJ776791 and KU647676] and nine others Brazilian Zika. Bootstrap values are indicated at nodes. **B)** Bayesian phylogenetic tree of NS5 gene sequences of Brazil ZIKV using GTR+G+I model with six (isolated from cell culture) Brazil ZIKV strains (MF077458, MF077459, MF077460, MF077461, MF077462 and MF077463) and 17 reference sequences from GenBank , with two references from the African lineage [accession numbers:HQ234500.1 djLC002500.1], six references from the Asian lineage [HQ234499.1, EU545988.1, KU681082.1, KU509998, KJ776791 and KU647676] and nine others from Brazil. Bootstrap values are indicated at nodes.

Fig3. Bayesian phylogenetic tree of Zika virus. The analysis includes 19 env regions from complete nucleotide sequences available at the National Center for Biotechnology Information (NCBI), eleven from African strain, six from Asian strain and two sequences (*MF048804 – *MF048805) from Brazil produced by our sanger protocol shows similar topology as those made by NGS (gbKU321639.1). Bootstrap values are indicated at nodes.

Fig1.

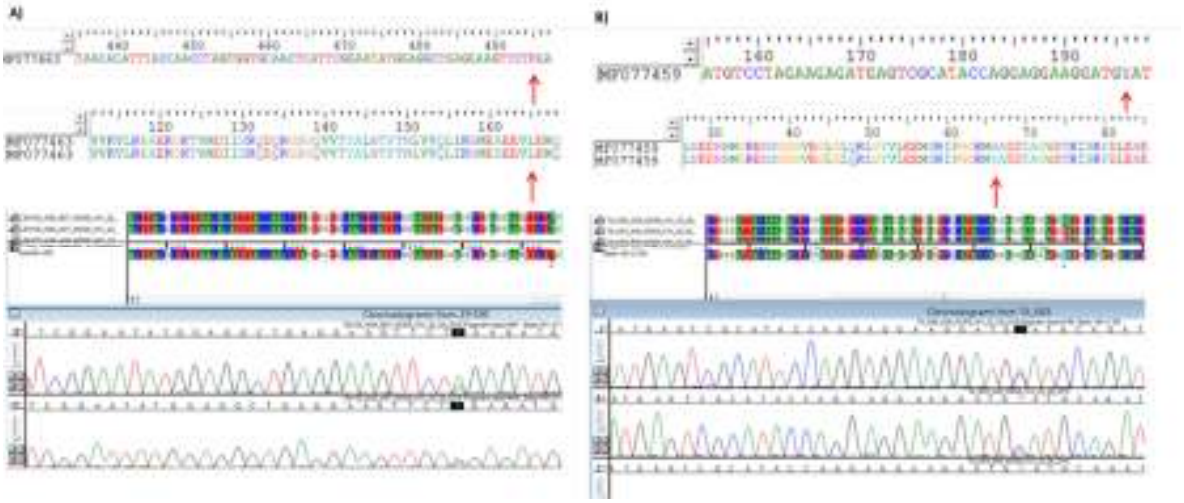


Fig2.

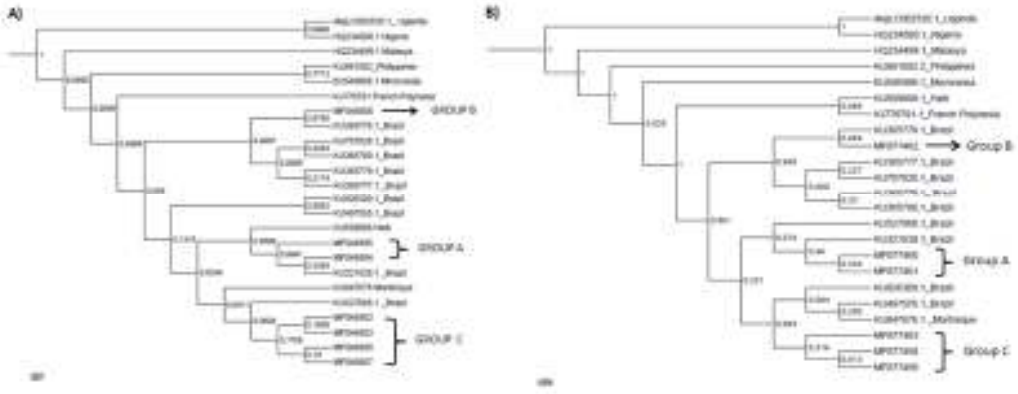
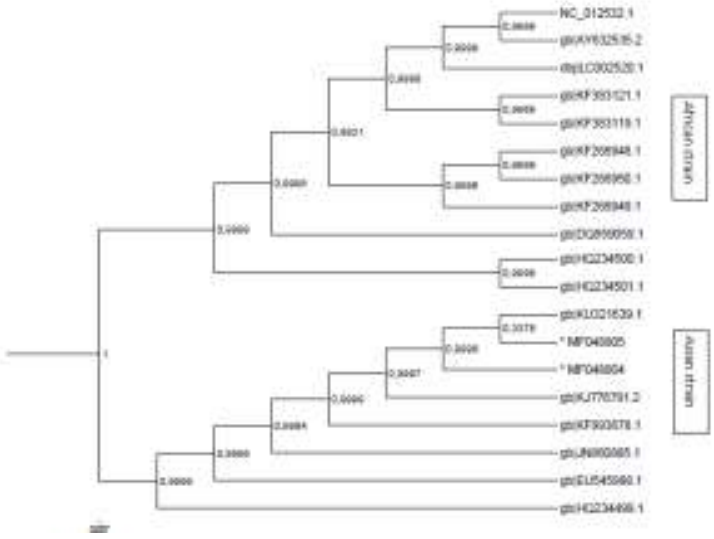


Fig3.


NC_012532.1
gb|AY632535.2
dbj|LC002520.1
gb|KF383121.1
gb|KF383119.1
gb|KF268948.1
gb|KF268950.1
gb|KF268949.1
gb|DQ856058.1
gb|HQ234500.1
gb|HQ234501.1
gb|KU021639.1
* MF046805
* MF046804
gb|KJ770791.2
gb|KF923078.1
gb|JN860085.1
gb|EU545060.1
gb|HQ234499.1
African strain
Asian strain