Open Access Library Journal

Primers for Dengue Virus Strains Based on Their Sequence Variability

Raghavendra Krishnappa, Rajendran Paramasivam

Department of Microbiology, Karpagam University, Coimbatore, India
Email: raghavendrakbt@gmail.com

Received 26 February 2015; accepted 13 March 2015; published 18 March 2015

Abstract

Since DENV-1, 2, 3, and 4 determine the strains for dengue virus, their gene sequence can be used as marker for diagnosis, amplifying and genotyping subtypes in molecular screening reaction which includes RT-PCR, real-time RT-PCR, nucleic acid sequence-based amplification, microsphere-based duplexed immunoassay, and DNA microarrays. There are many gene based PCR diagnostic kits available for screening and quantifying dengue virus, which one to choose? Decisions on choosing the diagnostic kit are debatable, mainly because of sequence variation of endemic dengue virus, which emphasizes us to use region specific primer diagnostic kits for isolating dengue of the prevailing country. But if diagnoistic industry focuses on the homologous regions obtained after aligning sequences, each representing the country of origin, we can design primers which can be used to detect dengue strains from any country of origin. Gene based diagnostics kits should have primers that should be covered for all entries present in the NCBI database with 100% total coverage similarity and specifically only to which they were designed for.

Keywords

NCBI, EBI, PCR, Primers, Dengue, Diagnostics, DENV-1, DENV-2, DENV-3, DENV-4, MSA-ClustalO

Subject Areas: **Genetics, Microbiology**

1. Introduction

Dengue is caused by dengue virus (DENV), a mosquito-borne flavivirus. DENV is a single stranded RNA positive-strand virus of the family Flaviviridae, genus Flavivirus. This genus includes also the West Nile virus, Tick-borne Encephalitis Virus, Yellow Fever Virus, and several other viruses which may cause encephalitis. DENV causes a wide range of diseases in humans, from a self limited Dengue Fever (DF) to a life-threatening syndrome called Dengue Hemorrhagic Fever (DHF) or Dengue Shock Syndrome (DSS). There are four antigenically different serotypes of the virus: DENV-1, DENV-2, DENV-3 and DENV-4. Here, a serotype is a group of

How to cite this paper: Krishnappa, R. and Paramasivam, R. (2015) Primers for Dengue Virus Strains Based on Their Sequence Variability. *Open Access Library Journal*, **2**: e1104. http://dx.doi.org/10.4236/oalib.1101104

viruses classified together based on their antigens on the surface of the virus. These four subtypes are different strains of dengue virus that have 60% - 80% homology between each other, therein hinting us for the potential marker regions. The major difference for humans lies in subtle differences in the surface proteins of the different dengue subtypes. Infection induces long-life protection against the infecting serotype, but it gives only a short time cross protective immunity against the other types. The first infection causes mostly minor disease, but secondary infection has been reported to cause severe diseases (DHF or DSS) in both children and adults which is called Antibody-Dependent Enhancement.

2. Dengue Virus RNA Genome

DENV is a 50 nm virus enveloped with a lipid membrane. There are 180 identical copies of the envelope (E) protein attached to the surface of the viral membrane by a short transmembrane segment. The virus has a genome of about 11000 bases that encodes a single large polyprotein that is subsequently cleaved into several structural and non-structural mature peptides. The polyprotein is divided into three structural proteins, C, prM, E; seven nonstructural proteins, NS1, NS2a, NS2b, NS3, NS4a, NS4b, NS5; and short non-coding regions on both the 5′ and 3′ end. The structural proteins are the capsid (C) protein, the envelope (E) glycoprotein and the membrane (M) protein, itself derived by furine-mediated cleavage from a prM precursor. The E glycoprotein is responsible for virion attachment to receptor and fusion of the virus envelope with the target cell membrane and bears the virus neutralization epitopes. In addition to the E glycoprotein, only one other viral protein, NS1, has been associated with a role in protective immunity. NS3 is a protease and a helicase, whereas NS5 is the RNA polymerase in charge of viral RNA replication. Data provided in **Table 1** maps DENV genes on its genomic and protein sequence which is first basic step researches have to perform for primer design.

Table 1. Structural and non-structural region of the dengue virus mapped on NCBI nucleotide and protein sequence, DQ181806 is of DENV-2 strains.

NCBI ID nucleotide—DQ181806	NCBI ID protein—ABA61185	Region
109..438	5..114	Flavivirus capsid protein C (C)
451..708	119..204	Flavivirus polyprotein propeptide
712..936	206..280	Flavivirus envelope glycoprotein M (prM)
937..1824	281..576	Flavivirus glycoprotein, central and dimerization domains
1849..2118	585..674	Immunoglobulin-like domain III (C-terminal domain) of Flavivirus envelope glycoprotein E
2131..2421	679..775	Flavivirus envelope glycoprotein E, stem/anchor domain (E)
2425..3489	777..1131	Flavivirus non-structural Protein NS1 (NS1)
3508..4131	1138..1345	Flavivirus non-structural protein NS2A (NS2a)
4141..4521	1349..1475	Flavivirus non-structural protein NS2B (NS2b)
4573..5025	1493..1643	Peptidase S7, Flavivirus NS3 serine protease (NS3)
5065..5496	1657..1800	DEAD-like helicases superfamily
5086..5475	1664..1793	DEAD-like helicases superfamily. A diverse family of proteins involved in ATP-dependent RNA or DNA unwinding. This domain contains the ATP-binding region
5584..5913	1830..1939	Helicase superfamily c-terminal domain; associated with DEXDc-, DEAD-, and DEAH-box proteins, yeast initiation factor 4A, Ski2p, and Hepatitis C virus NS3 helicases; this domain is found in a wide variety of helicases and helicase related proteins
6385..6816	2097..2240	Flavivirus non-structural protein NS4A (NS4a)
6826..7566	2244..2490	Flavivirus non-structural protein NS4B (NS4b)
7732..8235	2546..2713	FtsJ-like methyltransferase
8320..10257	2742..3387	Flavivirus RNA-directed RNA polymerase (NS5)

3. Background

Dengue is endemic in more than 110 countries [1]. It infects 50 to 100 million people worldwide a year, leading to half a million hospitalizations [2] and approximately 12,500 - 25,000 deaths [3] [4]. There are many diagnostic kits available for screening and quantifying dengue virus, which one to choose? Decisions on choosing the diagnostic kit are debatable one, mainly because of sequence variation of endemic dengue virus.

4. Methods

Total of 257 countries listed from online source (http://www.listofcountriesoftheworld.com/) were searched in NCBI Nucleotide database for dengue virus complete genome using the search criteria "dengue virus 'strain type' 'region' complete genome" in Nucleotide Database (http://www.ncbi.nlm.nih.gov/nuccore/ 29th Oct 2014). Out of 257 countries only 36, 44, 38 and 14 countries had submitted the complete genome sequence of dengue 1, 2, 3 and 4 virus respectively in to NCBI database.

Total of 547 complete genomic sequences were obtained from 36 countries for DENV-1. Sequence alignment with in the country using online EBI MSA-ClustalO (http://www.ebi.ac.uk/Tools/msa/clustalo/) tool provided 98.58 similarities (shown in **Table 2**). When some of the sequences (AY277666.2, JX669475.1, HM631853.1, GQ868570.1, DQ672560.1, JN903581.1, AB204803.1, KJ189369.1, FJ898437.1, KJ189367.1, EU081281.1, HQ891316.1, JN638336.1, GU131842.1) from 14 countries were picked randomly and MSA performed we found 93.71 similarities.

For DENV-2 total of 721 sequences were obtained from 44 countries. Sequence alignment with in the country using EBI MSA-ClustalO tool provided 98.78 similarities (shown in **Table 2**). When some of the sequences (HM582106.1, JX669488.1, FJ639718.1, AF204178.1, GQ868552.1, AY702038.1, EU920847.1, GQ398268.1, AF169678.1, EU920843.1, GQ868516.1, FJ882594.1, KF041237.1, KC294223.1, GQ868600.1, HM582107.1, EU081180.1, FJ906958.1, HM582117.1, FJ898466.1, FM210246.2) from 21 countries were picked randomly and MSA performed we found 93.48 similarities.

For DENV-3 total of 693 sequences were obtained from 38 countries. Sequence alignment with in the country using online public EBI MSA-ClustalO tool provided 98.34 similarities (shown in **Table 2**). When sequence (AY496873.2, GU131878.1, FJ639731.1, KF954948.1, GU131953.1, AY744679.1, FJ644564.1, AB189128.1, FJ898442.1, JF937652.1, KF041259.1, JF808129.1, KJ189301.1, KF955468.1, EU081225.1, FJ882574.1, DQ675533.1, FJ744740.1, AB214882.1, FJ898474.1, JQ045695.1) from 21 countries were picked randomly and MSA performed we found 96.15 similarities.

And finally for DENV-4 total of 95 sequences were obtained from 14 countries. Sequence alignment with in the country using EBI MSA-ClustalO tool provided 98.38 similarities (shown in **Table 2**). When some of the sequences (JQ513345.1, JN638572.1, GQ868585.1, JF262780.1, AY618993.1, FJ882592.1) from 6 countries were picked randomly and MSA performed we found 93.05 similarities.

5. Conclusions

Currently it is very tough to dig out the dengue virus from online NCBI database based on region of isolation and year of occurrence, which implies not just to dengue virus but to all organisms. But user can try out "Nucleotide Advanced Search Builder" using "dengue virus 1 (organism name)" "complete genome" and "brazil (country name)" to get the list of dengue virus 1, complete genome reported from Brazil.

From MSA, we get a score of 6.29%, 6.52%, 3.85% and 6.95% of genomic sequence variability existing within DENV-1, 2, 3 and 4 strains reported from different countries. And MSA scores for DENV-1, 2, 3 and 4 within the country are 98.58, 98.78, 98.34, and 98.38, respectively, which are towards the greater similarity side. While designing strain specific primers irrespective of country of origin, researches have to focus on 93.71, 93.48, 96.15 and 93.01 of similarity region for DENV-1, 2, 3 and 4 respectively. If we take DENV-1 into consideration, and if we need to have primer pair to detect DENV-1 reported from any country of origin (which falls in known 36 countries), primers have to have 100% similarity hits to all 547 complete genomic sequences from online public NCBI BLAST (http://blast.st-va.ncbi.nlm.nih.gov/Blast.cgi) hit and please avoid cross homology with other organism. So, the designed primer set works to detect dengue 1 strain from any country of origin. And again if researches are interested in detecting strains DENV-1, 2, 3 and 4 based on region specific criteria, MSA dissimilarities of 6.29, 6.52, 3.85 and 6.95 are the region to look for respectively for marker selection.

Table 2. MSA score for DENV strains with in the region. Hits from the other countries which are less than 4 counts were not used for multiple sequence alignment (MSA).

Countries for dengue 1	Hits	MSA-ClustalO
Mexico	115	99.38
Cambodia	89	98.76
Vietnam	67	99.24
Singapore	62	99.91
Nicaragua	45	99.49
Thailand	33	97.54
Brazil	31	97.65
Puerto Rico	19	99.41
Colombia	17	99.33
China	14	95.72
India	7	98.12
French Polynesia	6	99.94
Sri Lanka	6	99.77
Japan	5	94.48
Argentina	4	98.85

Countries for dengue 2	Hits	MSA-ClustalO
Nicaragua	191	99.19
Vietnam	63	99.2
Thailand	62	97.79
Puerto Rico	55	98.64
Brazil	47	98.04
Cambodia	47	97.63
Venezuela	40	98.76
Peru	25	98.16
Mexico	23	99.45
Japan	19	99.24
Colombia	18	98.58
French Guiana	14	98.6
Indonesia	14	97.1
China	10	95.4
Pakistan	9	99.14
Singapore	8	98.95
Tonga	8	99.84
Cuba	6	99.92
Martinique	6	99.69
India	5	97.84
American Samoa	4	99.78
Dominican Republic	4	99.98
Samoa	4	99.71

Countries for dengue 3	Hits	MSA-ClustalO
Timor-Leste	4	99.77
French Polynesia	18	99.74
Pakistan	6	99.59
Puerto Rico	4	99.58
Peru	47	99.57
Venezuela	113	99.56
Nicaragua	128	99.54
Mexico	3	99.39
Paraguay	5	99.33
Brazil	89	99.3
Vietnam	9	99.29
Cambodia	66	99.26
Bangladesh	9	99.17
Singapore	47	99.17
Taiwan	16	99.06
Sri Lanka	13	98.18
India	4	97.93
China	14	95.84
Colombia	18	95.84
Indonesia	29	87.75

Countries for dengue 4	Hits	MSA-ClustalO
Venezuela	41	99.46
Brazil	19	98.86
Colombia	8	99.79
Thailand	6	94.63
Cambodia	4	99.17

Since DENV-1, 2, 3, and 4 determine the strains for dengue virus, their gene sequence can be used as marker for diagnosis, amplifying and genotyping subtypes in molecular screening reaction which includes RT-PCR, real-time RT-PCR, nucleic acid sequence-based amplification, microsphere-based duplexed immunoassay, and DNA microarrays. Conserved region obtained after aligning 4 sequences of different strains occurring in that region help in identifying strains, if the sample is of many strains, than in that case five primers in one PCR well, with proper optimization will lead to identification of all dengue 4 strains in one well. One of the primers (which can be either forward or reverse) can be common for all strains and other four primers obtained from their conserved region, which are specific for their strains. By this, the experimental design will be able to genotype/ amplify all the strains in single PCR well. Use of SyBr Green in Real Time PCR will be very economical for the above mentioned experimental design. For more specificity, user can switch over to Taqman probes; probes can be deigned on conserved region and a primer pair designed on aligned common region, but it will increase the number of primers as well as expenditure. In case of diagnosis, it is always better to select the strains pertaining to the region of occurrence for accurate region specific markers (primers).

References made below will tell us why primer NCBI BLAST result of 100% total coverage similarity hit to the sequence for which they were designed for are very cricial in selecting primers for detection, which is again critical for diagnostic market. In many applications of PCR the template DNA is a mixture of homologous genes, it is important to find out how a primer-template mismatch can affect the accurate interpretation of the results and the effect of a single internal primer-template mismatch is variable but can greatly decrease PCR efficiency, depending on its position and on the primer used [5]. Such an effect of primer-template mismatches can distort the data used in drawing conclusions on the structure of a microbial community in a PCR-based approach, due to bias in the estimation of the relative proportions between groups possessing and not possessing mismatches with the primers. The presence of mismatches could also bias real-time PCR assays by leading to an underestimation of the actual gene copy number if measured against a perfectly matched standard. The effects of mismatches depend on numerous factors, such as oligonucleotide length and the nature and position of the mismatches. Several studies have investigated the effects of primer-template mismatches at the 3′ end of the primer sequence, and it has been demonstrated that PCR was prevented by a single mismatched base at the 3′ end [6]-[9]. Hence, mismatches located closer to the 3′ end of the primers were more critical for PCR efficiency. For the forward primers, a single mismatch at position -5, -6, or -8 of the 3′ end was sufficient to lead to an underestimation of 1 log of the gene copy number [5].

In contrast to single mismatches, the effect of several internal mismatches on PCR has been more widely described in the literature [10] as reported that a few mismatches reduced the sensitivity of detection of *Giardia*, as seen by an increase in the cycle threshold values. Recently [11] using an equal mixture of two strains as template, showed a preferential amplification of the strain that perfectly matched the primer sequence compared to the amplification of the strain exhibiting three mismatches close to the 5′ end of the primer sequence. A more-pronounced reduction in PCR efficiency with an increasing number of mismatches is shown in the literature [12].

References

[1] Ranjit, S. and Kissoon, N. (2010) Dengue Hemorrhagic Fever and Shock Syndromes. *Pediatric Critical Care Medicine*, **12**, 90-100. http://dx.doi.org/10.1097/PCC.0b013e3181e911a7

[2] Whitehorn, J. and Farrar, J. (2010) Dengue. *British Medical Bulletin*, **95**, 161-173. http://dx.doi.org/10.1093/bmb/ldq019

[3] Varatharaj, A. (2010) Encephalitis in the Clinical Spectrum of Dengue Infection. *Neurology India*, **58**, 585-591. http://dx.doi.org/10.4103/0028-3886.68655

[4] Knoop, K.J., Stack, L.B., Storrow, A. and Thurman, R.J. (2010) Tropical Medicine. Atlas of Emergency Medicine. 3rd Edition, McGraw-Hill Professional, New York, 658-659.

[5] Bru, D., Martin-Laurent, F. and Philippot, L. (2008) Quantification of the Detrimental Effect of a Single Primer-Template Mismatch by Real-Time PCR Using the 16S rRNA Gene as an Example. *Applied and Environmental Microbiology*, **74**, 1660-1663. http://dx.doi.org/10.1128/AEM.02403-07

[6] Ayyadevara, S., Thaden, J.J. and Shmookler Reis, R.J. (2000) Discrimination of Primer 3'-Nucleotide Mismatch by Taq DNA Polymerase during Polymerase Chain Reaction. *Analytical Biochemistry*, **284**, 11-18. http://dx.doi.org/10.1006/abio.2000.4635

[7] Baker, G.C., Smith, J.J. and Cowan, D.A. (2003) Review and Re-Analysis of Domain-Specific 16S Primers. *Journal of*

Microbiological Methods, **55**, 541-555. http://dx.doi.org/10.1016/j.mimet.2003.08.009

[8] Huang, M.M., Arnheim, N. and Goodman, M.F. (1992) Extension of Base Mispairs by Taq DNA Polymerase: Implications for Single Nucleotide Discrimination in PCR. *Nucleic Acids Research*, **20**, 4567-4573. http://dx.doi.org/10.1093/nar/20.17.4567

[9] Kwok, S., Kellogg, D.E., McKinney, N., Spasic, D., Goda, L., Levenson, C. and Sninsky, J.J. (1990) Effects of Primer-Template Mismatches on the Polymerase Chain Reaction: Human Immunodeficiency Virus Type 1 Model Studies. *Nucleic Acids Research*, **18**, 999-1005. http://dx.doi.org/10.1093/nar/18.4.999

[10] Guy, R.A., Xiao, C. and Horgen, P.A. (2004) Real-Time PCR Assay for Detection and Genotype Differentiation of Giardia Lamblia in Stool Specimens. *Journal of Clinical Microbiology*, **42**, 3317-3320. http://dx.doi.org/10.1128/JCM.42.7.3317-3320.2004

[11] Sipos, R., Székely, A.J., Palatinszky, M., Révész, S., Márialigeti, K. and Nikolausz, M. (2007) Effect of Primer Mismatch, Annealing Temperature and PCR Cycle Number on 16S rRNA Gene-Targetting Bacterial Community Analysis. *FEMS Microbiology Ecology*, **60**, 341-350. http://dx.doi.org/10.1111/j.1574-6941.2007.00283.x

[12] Smith, S., Vigilant, L. and Morin, P.A. (2002) The Effects of Sequence Length and Oligonucleotide Mismatches on 5′Exonuclease Assay Efficiency. *Nucleic Acids Research*, **30**, 111-121. http://dx.doi.org/10.1093/nar/gnf110