Full Length Research Paper

Molecular characterization and phylogenetic analysis of Family Adenoviridae from marine environment as inferred from the hexon protein gene

Dhirendra Kumar Singh and Swapna P. Antony*

National Centre for Aquatic Animal Health, Cochin University of Science and Technology, Cochin-682016, Kerala, India.

Accepted 6 April, 2015

The marine environment, covering 70% of the planet's surface, is estimated to contain 10³⁰ viruses. Despite their great importance in the marine environment, virtually very little is known about marine viral biodiversity or the evolutionary relationships of marine and non-marine viruses. The present study is aimed at molecular characterization and phylogenetic analysis of the viral Family Adenoviridae from the marine environment, based on the hexon protein gene. Briefly, 50 L of sea water was purified and concentrated by a combination of tangential flow filtration, flocculation and centrifugation. Metagenomic DNA was extracted from the viral particles and PCR amplification was performed using specific primers designed for hexon protein gene of adenoviruses. Positive amplicons were cloned and sequenced. Sequences were analysed using GeneTool, BioEdit, ExPASy, ClustalW, GeneDoc and Mega 5.0 programmes. Partial hexon protein amplicon of 254 bp fragment encoding hexon protein of 84 amino acids could be recovered from the metagenomc viral DNA (KJ958986). Phylogenetic relationship of the newly identified hexon protein gene was constructed employing neighbour joining method based on already available hexon protein sequences in GenBank. The results of phylogenetic analysis confirmed the basal placement of the newly identified hexon protein gene in the Adenoviridae family. The present study shows that the newly identified adenovirus from marine environment is more closely related to adenoviruses isolated from Ovine and Bovine sources than to other adenoviruses from human sources. The present study provides a clearer insight into adenoviridae present in the marine system.

Key words: Adenovirus, marine viruses, hexon protein, metagenomics, phylogeny.

INTRODUCTION

Viruses are the smallest sub-microscopic forms of life, attacking all organisms on the earth and the most abundant entity in sea water. Currently, best estimation of viral particles indicates concentrations of $\sim 10^6$ viruses ml⁻¹ of surface seawater, $\sim 3 \times 10^6$ viruses ml⁻¹ in the deep sea to $\sim 10^8$ viruses ml⁻¹ in productive coastal waters till now, the total estimation of viruses is about $\sim 10^{30}$ in the oceans (Bergh et al., 1989; Suttle, 2005; Rohwer and Barott, 2013). With higher abundance of viruses it is also seen that their genome sequences are also remarkably diverse. Due to lack of standard culture techniques, our knowledge about viral world is limited (Fuhrman and Campbell, 1998; Paul and Sullivan, 2005). To overcome

these difficulties, new culture-independent approaches came into light which are mainly based on molecular tools and done by amplification of conserved gene present in viruses (Nicholas and Paul, 2011).

Metagenomics is one of the prominent cultureindependent approaches that have revealed an incredible amount of viral genetic diversity, particularly in the marine

^{*}Corresponding author. E-mail: swapnapantony@gmail.com, swapnapantony@cusat.ac.in. Tel: + 914842381120; Fax: + 914842381120.

environment (Breitbart et al., 2002). Metagenomic study shows that approximately 75 % of the sequences in these viral metagenomes did not match any genes in the database (that is, E-value 00.001 to the non-redundant GenBank database) suggesting that most viral diversity remains uncharacterized (Breitbart and Forest, 2005). Studies have shown that some marine viruses show similarity with non-marine viruses (Forest et al., 2000). Analysis of the genome of Roseophage SIO1 (phage that infects the marine heterotrophic bacteria *Roseobacter* SIO67) has shown that marine and non-marine phages are genetically related, but basic life histories may be significantly different.

Family Adenoviridae includes a group of nonenveloped icosahedral viruses consisting of a protein coat or capsid, surrounding a DNA-protein core, ranging from 70 to 100 nm in diameter. The genome consists of double-stranded DNA ranging from 30 to 40 kb depending on the serotype (Bosch et al., 2006; Hundesa et al., 2006; Muscillo et al., 2008; Belsy et al., 2009). The viral proteins are essential for the infection process. Adenovirus with damaged DNA has been successfully shown to infect host cells (Ko et al., 2003; Jothikumar et al., 2005; Eischeid et al., 2009). Family Adenoviridae consists of four genera namely: Mastodenovirus, Aviadenovirus, Atadenovirus and Siadenovirus; however, only genera Mastadenovirus could affect humans (Pond, 2005).

Adenoviruses are able to withstand extreme conditions, including an acidic pH environment of 5-6 and are able to withstand even the harshness of lipid solvents. Hence, they can live outside the host for long periods of time compared to other viruses. Adenovirus has also shown to be resistant to both tertiary treatment and UV radiation (Pond, 2005). Human adenoviruses have been recently listed on the U.S. EPA Candidate Contaminant List (Ko et al., 2003; Griffin et al., 2003; Fon and Lipp, 2005; Xagoraraki et al., 2007) and is being suggested as an index of pollution of human origin in waters, based on its presence in greater numbers and high persistence in surface waters (Linden et al., 2007; Mahony, 2008). Furthermore, human adenoviruses have been shown to be prevalent in marine surface waters and in shellfishes namely: ovsters and clams (Jiang et al., 2001; Pina et al., 1998; Bofill-Mas et al., 2006). However, it has been suggested that all viral genomes detected does not correspond to infectious viral particles and a high proportion of non-infectious viral particles may be expected in the environment (Aragao et al., 2010).

Recent research is focused on using adenoviruses for cancer treatment for effective gene delivery and expression since they demonstrate several advantages over other vectors (Kanera and Hemminki, 2004; Rein et al., 2007, 2011). After subjecting the virus to extensive engineering to remove the genes which control viral replication and thereby also creating room to insert genes of therapeutic interest, adenoviruses become a promising agent for killing tumor cells.

The present study is aimed at molecular characterization and phylogenetic analysis of the viral family *Adenoviridae* from the marine environment, based on the hexon protein gene. The study also analyzes the genetic diversity and phylogenetic relationship of the newly identified adenovirus to the already identified adenoviruses from other sources. This is the first report of metagenomic characterization of adenoviruses from the Cochin Barmouth Region, India.

MATERIALS AND METHODS

Sample collection

About 50 L of surface sea water sample was collected in polyethylene containers from the Cochin Barmouth region, India (at latitude 9° 58' 0" North and longitude 76° 15' 0" East) during September 2013 (Figure 1).

Purification and concentration of samples

The water samples were immediately brought to the lab and filtered through 2 mm mesh and 0.4 µm to remove larger particles. This filtered sample was concentrated using tangential flow filtration system (QuixStandBenchtop System, GE healthcare). At each stage, the fractions were checked for the presence of viral particles using epiflourescence microscopy using SYBR Green stain (Invitrogen).

Flocculation of viral particles

The concentrated samples were further concentrated by flocculation method following previously explained procedure for marine samples by John et al. (2011). Briefly, the viral particles in the retentate were concentrated using 1 ml Fe solution (10 g Fecl₃ I^{-1}). After addition of FeCl₃ solution, shaking was repeated several times and kept for 1 h at room temperature to allow Fe-virus flocculate formation. Filtrate was collected in 10 ml tubes and centrifuged at 14,000 rpm for 15 min at 4°C. The pellets were then dissolved in freshly prepared 0.2 M Ascorbate-0.1M EDTA-Mg buffer (pH 6-7) and stored in the dark at -20°C until use.

Extraction of viral nucleic acids

Viral nucleic acids were extracted using QIAamp MinElute Virus Spin Kit (QIAGEN) following the manufacturer's protocol and was stored at -20°C. Viral nucleic acids were quantified and qualified by spectrophotometry at 260 nm and 280 nm and agarose gel electrophoresis.

PCR amplification

PCR amplification of 1 µl of viral nucleic acid (~200 ng)

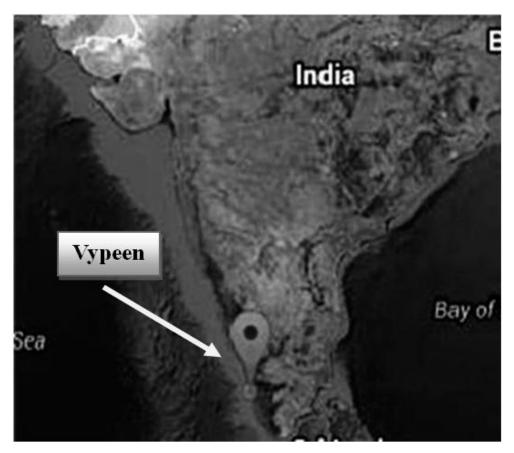


Figure 1. Sampling site - Cochin Barmouth Region, Vypeen, India.

was performed in a 25 μ l reaction volume containing 1x standard Taq buffer (10 mMTris-HCl, 50 mMKCl, pH 8.3), 1.5-3.5 mM MgCl₂, 200 μ M dNTPs, 0.4 μ M of each primer and 1U Taq DNA polymerase. The primer pairs used were Adeno- F (5'-gccgcagtggtcttacatgcacatc-3') and Adeno- R (5'-cagcacgccgcggatgtcaaagt-3'). The thermal profile used for the PCR amplification was 94°C for 2 min followed by 35 cycles of 94°C for 15 s, 60°C for 30 s and 68°C for 30 s and a final extension at 68°C for 10 min. PCR products were analyzed by electrophoresis in 1.5 % agarose gel in TBE buffer, stained with ethidium bromide and visualized under UV light.

Sequencing

Sequencing was performed for both PCR products and cloned plasmids. In the case of PCR products, the samples were subjected to ExoSap treatment for purification and the purified samples were sequenced at Scigenom, India, using both gene specific forward and reverse primers.

pGEMT cloning and plasmid sequencing

The positive PCR amplicons were cloned onto pGEM-T

Easy vector (Promega) following manufacturer's protocol using *E. coli* JM109 strain. Positive clones were screened using vector specific and gene specific primers. Plasmids were extracted using GenElute Plasmid MiniPrep kit (Sigma) from positive clones and were sequenced using vector specific forward and reverse primers at Scigenom, India.

Sequence analysis

The sequences were edited, assembled and analysed using GeneTool and BioEdit software. Gene translation and prediction of deduced proteins were performed with ExPASy (http://www.au.expasy.org/). The nucleotide sequence homology and the translated amino acid sequence comparisons were performed using BLAST algorithm (BLASTn, BLASTp and tBLASTx) of the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/blast). The multiple sequence alignments of nucleotide and amino acid sequence were performed with sequences retrieved from NCBI and multi-aligned using ClustalW and GeneDoc computer programmes. Phylogenetic and molecular evolutionary analyses were conducted by the Neighbor-Joining (NJ) and Maximum-Likelihood (ML) methods

tgccgcagtggtcttacatgcacatcgcggggcggcatacagccgaactggctgtcggtgtg S У М Н IAGR н ТАЕ LAV Ρ 0 W G v ΕI G GG н GQSV IN R RR S v н т D accacgggcctcttgctgcgctgcgaacttgtccagcaactcattcgcacggtcgatcgt \mathbf{LL} Ι R т т т L R C \mathbf{E} L v L D R G Q Q v ctcatgtggacgccgtccccattgcccgcgcatatcgctgtcttcaggaactttgacatc S Ρ LP АН Ι Α v R N т D Ι L M W т Ρ F cacaacatacta RGVL

Figure 2. Nucleotide and amino acid sequences of the partial hexon protein gene of Marine Adenovirus (KJ958986).

using MEGA version 5 (Tamura et al., 2011).

RESULTS AND DISCUSSION

Adenoviruses were first isolated and characterized in 1953 and was named "adenoviruses", after the original tissue (adenoids) in which they were discovered. Till

date, adenoviruses occur worldwide and have been isolated from a large number of different species, majority being from vertebrates including every major class from fishes to mammals. So far, more than 100 members of the adenovirus group have been identified that infect a wide range of vertebrate hosts and count to 100 serotypes, 57 being in humans. The adenoviruses isolated from mammals, especially those from humans, are well characterized. All of these viruses isolated so far contain a linear, double-stranded DNA (dsDNA) genome encapsidated in an icosahedral protein shell with fiber proteins of varying lengths.

Adenoviruses, due to their high prevalence and fairly conserved organization of the central part of their genomes, offer one of the best models for studying the molecular evolution of DNA viruses on a larger time scale. Adenoviruses have been shown to be prevalent in marine surface waters, including marine organisms, which may or may not be infectious (Jiang et al., 2001; Bofill-Mas et al., 2006; Aragao et al., 2010). Although the frequency of reports on marine adenoviruses is increasing worldwide, there are no molecular data concerning the adenoviruses of the Indian Coast. The present study is aimed at molecular characterization and phylogenetic analysis of a marine adenovirus from the Indian Coast, based on the hexon protein gene. This also analyzes the genetic diversity studv and phylogenetic relationship of the newly identified adenovirus to the already identified adenoviruses from other sources. This is the first report of metagenomic characterization of adenoviruses from the Cochin Barmouth Region, India.

In this study, 50 L of initial sample was concentrated down to 12 ml using a combination of tangential flow filtration and flocculation method. Epiflourescence microscopy confirmed the presence and concentration of viral particles by these methods. Number of viral like particles was found to be higher in the final sample after concentration, indicating that the viral purification protocol from seawater was efficient. Also, viral nucleic acids could be successfully isolated from the sample using QIAamp MinElute virus spit kit (Qiagen). The quality of viral nucleic acid obtained (A260/A280) was found to be 1.7 and the concentration obtained was 450 ng/µl, as inferred from spectrophotometry. Diluted viral nucleic acid was subjected to PCR amplification using gene specific primers.

A 254 bp fragment encoding 84 amino acids and possessing homology to the hexon protein gene of family *Adenoviridae* could be obtained from the metagenomic viral DNA (Figure 2). The sequence was submitted in NCBI GenBank under the accession number KJ958986. Translation of the nucleotide sequence was performed using GeneTool and ExPaSy programmes. Results of BLASTn, BLASTp and tBLASTx analysis of the sequences showed similarity to the already submitted hexon protein gene of adenovirus from various sources, thereby confirming the

sequence to belong to Family Adenoviridae.

Analysis of the sequence suggested that the hexon coding sequence was unlike any other known adenovirus hexon sequence so far reported. The sequence showed similarity of 97-100% for query coverage of 12-11% respectively in the case of BLASTn, whereas, BLASTp showed only 67-80% for query coverage of 17% for an Evalue ranging from 0.001-0.007. Sources of adenovirus that showed similarity to the newly identified marine adenovirus includes Simian, Ovine, Bovine, Human, Equine, Canine, Bat, Caprine, Alpaca, Cynomolgus and Sea Lion.

Multiple alignment performed for the nucleotide sequence and amino acid sequence of the hexon protein gene (KJ958986) showed the presence of conserved regions with the sequences (Figure 3a and b). The phylogenetic relationships of adenoviruses from various sources were also established, based on nucleotide sequence and amino acid comparison of the hexon a

	an +	40 +	60	+ 80	+ 100	+ 120	+ 140
Unidentifi :	TECCECNETEETCT	TACATECACATCOCEGECO	echiacadecesa-	CHEGCTGTCGCTGTCG	A		
0111589546 - NTGGCGNCGCCTTCG	итерисскоероностори	тасатосасатососсоро	CONTROCTORODOT	перетского перетре	100200000000000000000000000000000000000	скостастите стисс	TAACANOTTTACAAACCO
gi 1088631 : AIGGCIACICCIICG	ATGATGCCGCAGTGGTCT	TACATGCACATCGCCGG <mark>G</mark> CI	CCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	TENCICACOTGECCIAG	IGCAGTITICCCCAGECRACAG	GROCIACIT <mark>TRAAT</mark> IGG	TAACAAGIIIIAGAAACCO
gi 1088631 : ATGGCGACGCCTTCG	ATGATGCCGCAGTGGTCT	TACATGCACATCGCCGG <mark>G</mark> CI	CCRECTCORES.	TENCICACOTGECCIAG	IGCAGTITICCCCAGECRACAG	GROCIACIT <mark>TRAGT</mark> IGG	TAACAAGIIIIAGAAACCO
gi 1088631 : ATGGCGACCCCGTCG					IGCAGTIT <mark>IGCCO</mark> AGGCAACAG		
gi 1088631 : ATGGCGACCCCTTCG					IGCAGTIT <mark>IGCCO</mark> AGGCAACAG		
gi 3583649 :					IGCAGTICCCCC <mark>CCCCC</mark> ACAG		
gi 3583649 :	TECCECAGIGETCI	TACATECACATEGECGE <mark>A</mark> CI	GENIGOTICEGNET.	CONGRETCORGETOTOR	IGCAGTICGCCC <mark>GC</mark> GC <mark>C</mark> ACAG	OROCIACITORATORGO	NARTANGITINGAARTCC
gi 3583€49 :	TGCCGC2GTGGTCT	TACATGCACATCOCCOGAC	CONTROTICOCNET:	CCUCAGICOCCECTOR	IGCAGTICCCCC <mark>GC</mark> GC <mark>C</mark> ACAG	OROCIACITORATORG	NAR TANGITI MGAARTCC
gi 3583648 :	TGCCGCAGTGGTCT	INCATECACATOGOOGENCI	GENTECTICEGNET.	CONGAGICOCCEPTORS	IGCAGITICCCCC <mark>GCCCC</mark> ACAG	OPOCIACITOPATORG	ANARTANGITINGAARTCC
gi 3583648 :	TGCCGCRGTGTGTCT	INCATECACATEGECGENCI	GGNIGOTICEGNGI.	CONGAGICOGGEICICS	IGCAGTICCCCC <mark>T</mark> CCAACAC	OCCIACIIO/GIAIGG	GAACAACIIIIAGAAACCO
gi 3583648 :	TGCCGCRGTGGTCT	INCATECACATOGOOGENCI	GGRCGCTICEGNGI	RCENCAGICOGGEICIGG	IGCAGTICCCCC <mark>CCCC</mark> ACAG	CCTACTICZCTCIGG	A <mark>GAR</mark> CARCITIRGARACCC
gi 3583648 :	TGCCGCRGIGGICI	INCATECACATE SECOND	GERCECTICEGRET	CONCAGICOGGEICIGG	IGCNETTOSCCO <mark>GO</mark> GO <mark>O</mark> ACHSI	CCCIACITORCICION	H <mark>GAR</mark> CARCITIRGARACCC
					I GCNETT COCCO <mark>GC</mark> OC <mark>O</mark> ACHO		
					IECYELLOSCOCCOCCOCCYCCE		
gi 3583€48 :	IGCCGCRGIGGICI	INCHIGCACATOGOOGENCI	GGRIGOTICEGRII.	CCNGAGICOGGEICTOG	IGCNETTOCCCOGICCNACAG	OCCIACINO CIACINO	A <mark>GAACAAA</mark> TITAGAAACCC
gi 3583€48 :	IGCCGCRGIGGICI	INCATECACATEGECGERC	GGNIGCIICGGNGI	CCNGAGICOGGEICICS	IGCAGTI <mark>C</mark> GCCC <mark>GT</mark> GC <mark>AAC</mark> AG	ONCCINCINCISCON CONCERNING	ACAACAAATTTACAAACCC
gi 3583€48 :	TECCECAETEETCT	INCATECNCATCOCCEGNCI	CENTECTICESSET.	RECEIGAGICOGGEICHOG	IGCNGTICCCCCCTCCNACHC	OCTACINO GIARGO	G <mark>GNACARN</mark> TITHGARGCCC
gi 3583647 :					IGCAGTICGCCC <mark>GI</mark> GCAACAG		
gi 3583647 :					IGCAGTICGCCC <mark>GI</mark> GCAACAG		
gi 3583647 :					IGCAGTICGCCC <mark>GI</mark> GCAACAG		
gi 3583647 :	necceckenergenon	TACATECACATOGOOGEACI	GENTECTICEGNET	CONGACIOCOCETONICE	IGCAGTI <mark>O</mark> GCCC <mark>GT</mark> GC <mark>A</mark> ACAGI	O.CONCINCIATED	GANCHENTTINGENACCO
gi 3583€47 :					IGCAGTI <mark>O</mark> GCCO <mark>G</mark> IGC <mark>AAC</mark> AG		
gi 1088631 : ATGGCGACGCCGTCG					TECNETINGECCO AGEO GAO GE		
	Teccecagreerce:	IACATGCACATCgCcGG C:	legatge te eagt	a CI de GG dige	tgcågIt Gece Ge ACaG:	a A ctactt A 199	g Aàcaa tttaGaaaccc
	200 +	220 *	240	+ 260	+ 280	+ 300	+ 320
Unidentifi : TIG TGG					CAT-ETCO TETOTT AGE		GACATOCOCOCOCOTOCTG-
gi 1589546 : TOCCACACGCICCAG							
gi 1088631 : TOCCARACCICAG							
gi 1088631 : TOCCARAGCTECAG	стессотпленссолены		TRUTTACE-REACC		GENERCENCERCENTTINGENCI	TEREBORES	GACADITOR GERGAGECIGG
gi 1088631 : TOCCRAREGOIGCAG							
gi 1088631 : TOTOLEAEGCTECAE		BACCOCCERE COCCECERE	TROUTAGE -REACO	CONTROCAGONACO GONG	GENERCAR <mark>OCENENTITEE</mark> RCI		

	EGGGEACGCIGG
gi 1088631 : TOCCARACCICCACCTOCCCOTTELCCCACACACCCCCCARACCCCCCACACACCACCCCCCCC	RECECTICE
gi 1088631 : TOCC: AR: @CTCCRECTECCECTTEDCCCREDEARCCCCCCREARACKCCCCREDEARCTACK - #GACCCCCDTECCREDEAGAACAACCECCETTEGEACAACCECCACTARCTTCACAATC	COCOACCOTC
gi 1088631 : TOCORRAGOTECX60T6066TTTECCCCAERAAACCCCCARAACCCCCCAERTACTACA-REACCCCCOTECX6CRA66GengeeXaAcaacCEAETTEGEACATE6GenecCateregenecACTTCACAATT	CEGECACECTCG
gi 3583649 : ACCACCECCTCATETTECCCTTCCTCCCCCCCCCCCCCC	GOGGOGIGCI
gi 3583645 : Accelececcicalenneccecurcerecceerececerecenterecerecerecerecerecerecerecerecerecere	IGCGGCGTGCT
gi 3583649 : AccelececcicAlefrecectroetecececececececececececececececececece	IGCGGCGTGCT
gi 3583648 : AcceleccocicaTeTrecectroeteccoertexCoceckerActAteCtottack-therecertexAcheeccercateAcheeccercateTateccokeckGeTetrecertexTe	GCGGCGTGCT
gi 3583648 : AccesceAcreAnectececrroerecodernesceeesesceAracentececreAcaeConcecerecesceAcaeCeneccreCosceAcaeCeneccesceAraeConcecerecescaeCeneccesceAraeConcecerecescaeCeneccesceAraeConcecerecescaeCeneccerecesceAraeCeneccesceAra	GCGGCGTGCT
gi 3583648 : Actorececterecectroerecodenessoceeskerc: Acatorecectroeerecontorececescatorececescator ActiveConterectroerececescatorececescatorececescatorececescatorececescatorececescatorececescatorececescatorececescatorececescatorececescatorececescatorececescatorececescatorececescatore	GCGGCGTGCT
gi 3583648 : Actorececterecectroeteccoetecectroeteccecker/schologines-therecerecerecerecerectroetecteckerecorecker/o	GCGGCGTGCT
gi 3583648 : Actorececterecectroeteccoetecectoeteccoetecectrotecetecectrotecec	GCGGCGTGCT
gi 3583648 : TCTCARCETTIERCECTECEETTERCECCEERERCECCEERERTECCECEREREECEREREECEEREREECTERCARCECEEREREECTECCARE	ACCCCTCCT
gi 3583648 : ACCACCECCTEATECTECECTTCETECCCETTEACCECERAEACCETTACE-CACTOCECTECECTECECTECECERCECERCECERCECERCE	GCGGCGTGCT
gi 3583648 : AGCONCCOCTORTECTOCCOCTTOTECCOCOTTENCOCCEGENAMECTATACONOTOTIANE-CAGTICACHAGAGECTERAGECONCANDAGECECONCENDETTERACHECONCECONCENDETTERACHECONCECONCENDETTERACHECONCECONCENDETTERACHECONCECONCENDETTERACHECONCECONCENDETTERACHECONCECONCENDETTERACHECONCECONCECONCENDETTERACHECONCECONCENDETTERACHECONCECONCENDETTERACHECONCECONCECONCENDETTERACHECONCECONCENDETTERACHECONCECONCENDETTERACHECONCECONCECONCENDETTERACHECONCECONCECONCECONCECONCECONCECONCECO	GCGGCGTGCT
gi 3583648 : AGCONCCOCTORTECTOCCOCTTOTECCOCOTTONCCOCONACTOTTACC-CARTECOTECA-CARGEOTERAGEOTERAGEOTERAGEOTERAGEOTECACTOCCACARCAGEOTECACARCAGEOTERAGEOTERAGEOTECACARCAGEOTECA	GCGGCGTGCT
gi 3583647 : ACCACCECCTEATECTECECTTOETECCCETTEACCECCACACACACACACACTECTTACA-CACTOCCTACA-CACGETECACCOCACAACACACACACACACACACACACACACACAC	GCGCGGGTGCT
gi 3583647 : AGCONCCOCTORIGCECCTTOETECCOORTENCCOGENAMECENTECCTECENTECECTREE-ENGINE CONSCRETE AGCONECTREECENECECCNECKON TOTTERENT	GCGGCGTGCT
gi 3583647 : AGCONCCOCTORISCICCOCTICOLOCOCOLICACO GENAMACANATACOLOCICITACA - LAGIICOCORA A GONGOCIENACO ACANO XEA ORONII TEACATEC CONCLADA INTERCATO	GCGGGGGTGCT
gi 3583647 : AGCORCECCTERTECTECECTTETECCECTTERCEGERAGECERAGECERTECECTECTTECE-CARTECECTRECAGECERAGECERCA	GCGGCGTGCT
gi 3583647 : AGCORCECCTERTECTECECTTETECCECTTERCEGERAGECERAGECERTECECTECE - AGTICECTREEACGECTERAGECERCAGAECAECAECAECAECAECAECAECAECAECAECAECAECA	GCGGCGTGCT
gi 1088631 : TCGCACCEGCTECZECTECECTTCETECCGENEGACCECCAGEACCCCCAGEACCECETACCA-SEACGECTATCA-SEAGECEACCAACGECAACCAGECAGE	IGC GOC ACCCT GG

cCageg etg GeIGCG tt gt CC Gt gaeCg ga gAca . Ta tCttAca a . Cg I e e eI gC Gt gg gAcaaC G gI et GAcatgg ageae t CIIIGACAI CgeGGe GCI

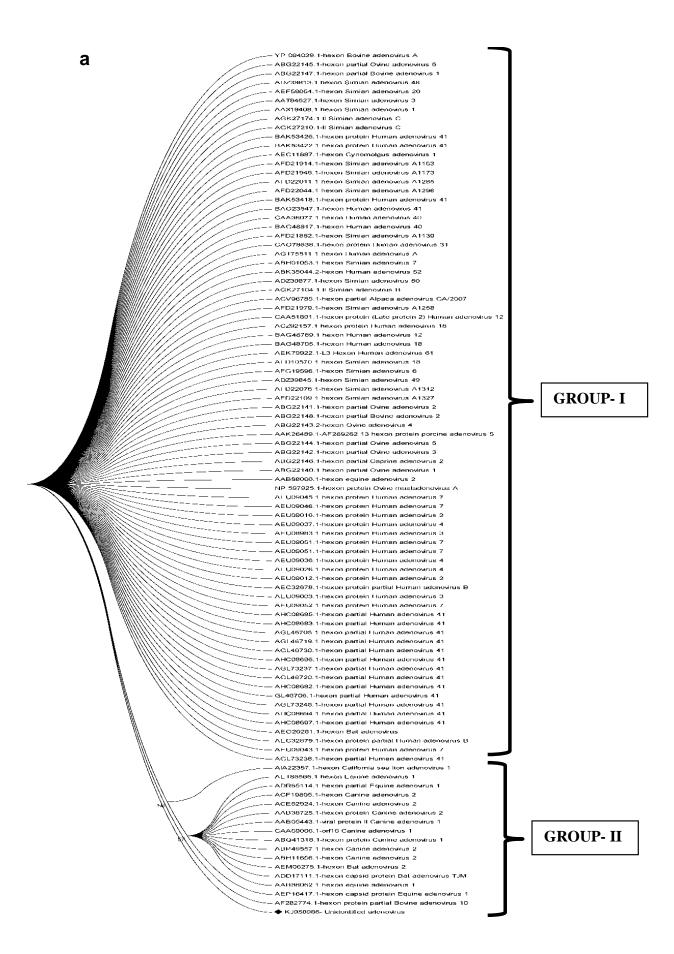
h

b					
Unidentifi	*	20 * VEVE T EGGE	40 * 6 HGQSVENRRRSEHEETTGEL		80 * 100 TPSPLPAH AVERN-PDIRGVI : 84
gi 1935274	: PQWSYMHIAGQDAAEYLS	F <mark>A</mark> LVQFA <mark>Q</mark> AT <mark>S</mark> SYF <mark>KID</mark> NKE	FRNFTVAFTHDVTT <mark>E</mark> R-SQRL <mark>Q</mark>	L-RFVPV <mark>MQ</mark> ED <mark>GQ</mark> YTY <mark>R</mark> TR	FQLAVGDNRVLDMASTYFDIRGTLDRGP : 104
gi 1901444 gi 1732280	: PQWSYMHIAGQDAAEYLS : PQWSYMHIAGQDAAEYLS	ALVQFAQATSSYE <mark>KID</mark> NKE	FRNPTVAPTHDVTTER-SQRLQ	L-REVEV <mark>MQEDGQ</mark> YTY <mark>K</mark> TR L-REVEVMOEDGOYTYKTR	FOLAVGDNRVLDMASTYFDIRGTLDRGP : 104 FOLAVGDNRVLDMASTYFDIRGTLDRGP : 104
gi 1477669	: POWSYMHIAGODAAEYLS	ALVQFACATSSYFKIDNK	FRNPTVAPTHDVTTER-SQRLQ	L-REVENMOEDGCYTYNTR	FQLAVGDNRVLDMASTYFDIRGTLDRGF : 104
gi 1526591	: PQWSYMHIAGQDAABYLS	F <mark>A</mark> LVQFAQATSSYF <mark>KID</mark> NKE	RNPTVAPTHDVTT <mark>P</mark> R-SQRLQ		FQLAVGDNRVLDMASTYFDIRGTLDRGF : 104
gi 1464473 gi 1452805	: POWSYMHIAGODAAEYLS : POWSYMHIAGODAAEYLS	FALVOFACATSSYFRIDNE ALVOFACATSSYFRIDNE	FRNPTVAPTHDVTTER-SQRLQ FRNPTVAPTHDVTTER-SORLO	L-REVEVMQEDGQYTYKTR L-REVEVMOEDGQYTYKTR	FCLAVGDNRVLDMASTYFDIRGTIDRGP : 104
gi 1116103	: POWSYMHIAGODAAEYLS	F <mark>A</mark> lvqfa <mark>q</mark> at <mark>s</mark> syf <mark>kid</mark> nke	FRNPTVAPTHDVTTER-SQRLQ	L-RFVPV <mark>MQ</mark> ED <mark>GQ</mark> YTY <mark>X</mark> TR	F <mark>QLAVGDNRVLDM</mark> ASTYFDIRG <mark>TLDRGF : 104</mark>
gi 3432012 gi 2897190	: PQWSYMHIAGQDAAEYLS: : PQWSYMHIAGQDA <mark>A</mark> EYLS:	ALVOFAQATSSYE <mark>KID</mark> NKE	FRNPTVAPSHDVTTER-SQRL	L-RYVPV <mark>MQEDGO</mark> YTY <mark>K</mark> TR L-RYVPVMCEDGOYTYKTR	FOLAVGDNRVLDMASTYFDIRGTIDRGF : 104 FOLAVGDNRVLDMASTYFDIRGTIDRGF : 104
gi 2239284	: POWSYMHIAGODAABYLS	ALVGFACATNSYFREDNK	FRNPTVAPTHDVTTER-SQRIQ	1-REVENMOEDGOYTENTR	FQLSVGDNRVLDMGSTFFDIRGTLDRGF : 104
gi 3476022	: PQWSYMHIAGQDAAEYLS	ALVQFAQATNSYE <mark>RID</mark> NKE	RNPTVAPTHDVTT <mark>E</mark> R-SQRLQ	L-REVEV <mark>MQ</mark> ED <mark>GQ</mark> YTE <mark>M</mark> TR	FQLSVGDNRVLDMGSTFFDIRGTLDRGP : 104
gi 4990734 gi 3583648	: EQWSYMHIAGQDASEYLS : EQWSYMHIAGQDASEYLS	PGLVQFARAIDIYFSIGNK	FRNEFVAETEDVITDR-SORLI	L-REVEVDREDNISSEVR	FTLAVGDNRVL 87 YTLAVGDNRVLDMASTFFDIRGVL 100
gi 3462149	: FOWSYMHIAGODASEYLS:	pglvçfa <mark>q</mark> at <mark>e</mark> syf <mark>di</mark> gnki	FRNE <mark>MV</mark> APTHDVT <mark>TDR</mark> -SQRL <mark>Q</mark>	l-KIVPVDKED <mark>ST</mark> YFF <mark>K</mark> AR	F <mark>N</mark> LAVGDNRVLDM <mark>A</mark> SCYFDIRG <mark>V : 99</mark>
gi 5654196 gi 5654196	: POWSYMHIAGODASEYLS : POWSYMHIAGODASEYLS	PGLVQFARATDTYF <mark>SI</mark> GNKE	FRNPTVAPTHDVTTDR-SQRLT	I-REVEVDREDTAYSYRVR	FTLAVGDN : 84 FTLAVGD : 83
gi 4990734	PQWSYMHIAGQDASEYLS	PGLVQFARAT <mark>D</mark> TYF <mark>S</mark> IGNK	RNPTVAPTHDVTTDR-SQRLT	L-REVEVDREDTAYSYEVR	FTLAVGDNR 85
gi 4968925 gi 5654195	PQWSYMHIAGQDASEYLS PQWSYMHIAGQDASEYLS	PGLVQFARAT <mark>D</mark> TYF <mark>S</mark> IGNKE	RNPTVAPTHDVTTDR-SQRL <mark>T</mark>	L-REVEVDRED TA YSY <mark>K</mark> VR	FTLAVGDNRVNA 88 FT 78
gi 4968925	: POWSIMHIAGODASEILS : POWSYMHIAGODASEYLS	PGLVQFARAILIIFSIGNKI PGLVQFARATETYFSIGNKI	FRNPIVAPIHDVIIDR-SQRLI FRNPIVAPIHDVIIDR-SQRLI	L-REVEVEREDIAISIEVE	FT: 78 FTL2: 80
gi 4990734	: PQWSYMHIAGQDASEYLS	PGLVQ <mark>L</mark> ARATDTYF <mark>S</mark> LGNKE	RNPTVAPTHDVTTDR-SQRLT	l-rfvpvdred ta ysy <mark>r</mark> vr	FTLAVGDNRL : 86
gi 5654196 gi 4968926	: PQWSYMHIAGQDASEYLS : PQWSYMHIAGQDASEYLS	PGLVQFARATETYFSLGNK PGLVQFARATETYFSLGNK	RNPTVAPTHDVTTDR-SQRIT	L-REVEVDREDTAYSY	FTLAVGDNRV : 86 : 74
gi 4968925	: POWSYMHIAGODASEYLS:	PGLVQFA <mark>R</mark> AT <mark>D</mark> TYF <mark>SI</mark> GNKE	FRNPTVAPTHDVTTDR-SQRLT	L-RFVPVDRED <mark>TA</mark> YSY <mark>S</mark> VR	FRVV : 80
gi 4968925 gi 5654195	: PQWSYMHIAGQDASEYLS : POWSYMHIAGCDASEYLS	PGLVQFARATETYF <mark>SI</mark> GNKF	FRNPTVAPTHDVTTDR-SQRLT	1-REVEVDREDTAYSYEVR	FTLAVGTFVN : 8€ FTLAVGDN8G : 8€
gi 5654195	: PQWSYMHIAGQDASEYLS	PGLVQFARATDTYFSIGNK	FRNPTVAPTHDVTTDR-SQRLT	L-REVEVEREDTAYSYEVR	FTLAVGGN 84
gi 3583648	: PQWSYMHIAGQDASEYLS	PGLVQF <mark>VRATD</mark> TYF <mark>SM</mark> GNKE	RNPTVAPTHDVTT <mark>D</mark> R-SQRL <mark>M</mark>	L-REVEVDRED <mark>NT</mark> YSY <mark>K</mark> VR	Y <mark>T</mark> LAVGDNRVLDM <mark>A</mark> STFFDIRG <mark>V</mark> L : 100
gi 3583647 gi 3583648	 PQWSYMHIAGQDASEYLS PQWSYMHIAGQDASEYLS 	PGLVQFARATETYFSMGNKE	FRATTAPTHDVIIDR-SQRD	L-REVEVEREDNISINVE	YTLAVGDNRVLDMASTFFDIRG <mark>VL : 100</mark> YTLAVGDNRVLDMASTFFDIRG <mark>VL : 100</mark>
gi 3583648	: PQWSYMHIAGQDASEYLS	pglvõfa <mark>rat</mark> dtyf <mark>sl</mark> gnki	FRNPTVAPTHDVT <mark>TDR</mark> -SQRL <mark>T</mark>	l-rfvpvdred <mark>ntc</mark> sy <mark>r</mark> vr	Y <mark>T</mark> LAVGDNRVLDM <mark>A</mark> STYFDIRG <mark>V</mark> L : 100
gi 3583648 gi 3583648	 PQWSYMHIAGQDASEYLS PQWSYMHIAGQDASEYLS 	PGLVQFARATDTYF <mark>SI</mark> GNKE	FRNPTVAPTHDVTTDR-SQRLT	L-REVEVDREDNTYSYRVR	YTLAVGDNRVLDMASTYFDIRG <mark>VL : 100</mark> YTLAVGDNRVLDMASTFFDIRGVL : 100
gi 3583648	POWSYMHIAGODASEYLS	PGLVQFARATETYF <mark>S</mark> IRNK	RNPTVAPTHDVTTDR-SQRLA	L-REVEVDRED <mark>NT</mark> YSY <mark>S</mark> VR	Y <mark>T</mark> LAVGDNRVLDM <mark>A</mark> STYFDIRG <mark>V</mark> L : 100
gi 3583647 gi 3583648	 PQWSYMHIAGQDASEYLS: PQWSYMHIAGQDASEYLS: 	PGLVQFARATDTYF <mark>S</mark> MGNKE	FRNETVAETHDVTTER-SORL	I-REVEADREDNTYSY	YTLAVGDNRVLDMASTFFDIRGVL : 100
gi 3583648	: POWSYMHIAGODASEYLS	PGLVQFARATETTF S IGNKI	FRNPTVAPTHDVTTDR-SQRIM	L-RFVPVDREDNIISIRVR	YTLAVGDNRVLDRASIFFDIRGVL : 100
gi 3583648	: PQWSYMHIAGQDASEYLS	PGLVQFARATDTYF <mark>S</mark> MG <mark>Y</mark> KF	RNPTVAPTHDVTTDR-SQRLM	L-REVEVDRED <mark>NT</mark> YSY <mark>S</mark> VR	YTLAVGDNRVEDMASTFFDIRGVL : 100
gi 3583648 gi 8927594	: POWSYMHIAGODASEYLS : POWSYMHIAGODASEYLS	PGLVQFARATDTYFSMGNKF GLVCFARATDSYFSIGNKF	FRNPIVAPIHDVIIDR-SQRD FRNPIVAPIHDVII <mark>E</mark> R-SQRD	L-REVEVEREDICYTYNTR	YTLAVGDNRVLD <mark>GFSHSLTSARVL</mark> : 100 F <mark>GLT</mark> VGDNRVLDM <mark>G</mark> STYFDIRG <mark>VIDRGF</mark> : 104
gi 1934344	: PQWSYMHIAGQDASEYLS	PGLVQFA <mark>Q</mark> AT <mark>E</mark> SYF <mark>KI</mark> GNKF	FRNPT <mark>V</mark> APTHDVT <mark>TE</mark> R-SQRL <mark>Q</mark>	1-REVEVDRED TÇ YTY <mark>R</mark> TR	F <mark>Q</mark> LAVGDNRVLDM <mark>G</mark> STYFDIRG <mark>T</mark> L DRGF : 104
gi 2529367 gi 1088631	: POWSYMHIAGODASEYLS: : POWSYMHIAGODASEYLS:	PGIVQFA <mark>QA<mark>RKP</mark>YE<mark>HL</mark>GNKE PGLVCFACATETYEKLGNKE</mark>	FRNPTVAPTHERTTERFABACS	C-AFVEVDREDIVYTYNTR I-RFVEVDREDICYSYNTR	F <mark>CLAVGNNRVLDMGRTYFY</mark> IRG <mark>TLDRGF : 105</mark> FCLAVGDNRVLDMASTYFDIRGTLDRGF : 104
gi 1088631	: POWSYMHIAGODASEYLS	PGLVQFA <mark>Q</mark> AT <mark>E</mark> TYF <mark>KI</mark> GNKF	FRNPTVAPTHDVTTER-SQRL	L-RFVPVDRED TQ YSY <mark>K</mark> TR	F <mark>Q</mark> LAVGDNRVLDM <mark>A</mark> STYFDIRG <mark>T</mark> LDRGP : 104
gi 1088631 gi 1088631	: POWSYMHIAGODASEYLS : POWSYMHIAGODASEYLS	PGLVQFA <mark>Q</mark> AT <mark>E</mark> SYF <mark>KI</mark> GNKF	FRNPTVAPTHDVTTER-SQRL	L-REVEVDRED TO YTY <mark>KT</mark> R	FOLAVGDNRVLDMGSTYFDIRGTIDRGF : 104 FOLAVGDNRVLDMGSTYFDIRGTIDRGF : 104
gi 1344672	: FOWSYMHIAGODASEYLS	PGLVQFAQATETYF <mark>KI</mark> GNK	FRNFTVAFTHDVTT <mark>E</mark> R-SQRLQ	L-REVEVDREDTQYTYETR	F <mark>Q</mark> LAVGDNRVLDM <mark>A</mark> STYFDIRG <mark>VIDRGF : 104</mark>
gi 1589546 gi 1088631	: PQWSYMHIAGQDASEYLS : PQWSYMHIAGQDASEYLS	PGLVQFA <mark>Q</mark> AT <mark>E</mark> SYF <mark>KI</mark> GNKE	RNPTVAPTHDVTT <mark>P</mark> R-SQRLQ	L-REVEVDRED TO YTY <mark>K</mark> TR	FQLAVGDNRVLDMGSTYFDIRGTLDRGP : 104 FQLAVGDNRVLDMGSTYFDIRGTLDRGP : 104
gi 1088631	: POWSIMHIAGQLASEILS	PGLVQFAQATESIFKIGNKI PGLVQFAQATESYFKIGNKI	FRNPIVAPIHLVIILR-SQRLQ	L-REVEVEREDICITIER	FOLAVGENRVLENGSTITEERGIEERGP : 104
gi 3799772	: PQWSYMHIAGQDASEYLS	PGLVQFA <mark>R</mark> AT <mark>D</mark> TYF <mark>SH</mark> GNKF	RNPTVAPTHDVTTDR-SQRLT	1-RFVPVDRED TA YSY <mark>B</mark> AR	F <mark>T</mark> LAVGDNRVLDM <mark>A</mark> STYFDIRG <mark>VLDRGF : 104</mark>
gi 3799772 gi 3256590	 PQWSYMHIAGQDASEYLS: PQWSYMHIAGCDASEYLS: 	PGLVQFARAILIIFSIGNRF PGLVQFARATETYFSIGNRF	FRNPIVAPIHDVIIDR-SQRII FRNPIVAPIHDVIIDR-SQRII	1-RFVPVDREDTAYSYRAR	FTLAVGDNRVLDMASTYFDIRG <mark>VLDRGF : 104</mark> FTLAVGDNRVLDMASTYFDIRGVLDRGF : 104
gi 3812837	: PQWSYMHIAGQDASEYLS	PGLVQFA <mark>R</mark> AT <mark>E</mark> TYF <mark>SL</mark> GNKF	FRNPT <mark>V</mark> APTHDVT <mark>TDR-</mark> SQRL <mark>T</mark>	I-RFVPVDKED <mark>TA</mark> YSY <mark>K</mark> TR	F <mark>T</mark> LAVGDNRVLDM <mark>A</mark> STYFDIRG <mark>VIDRGP : 104</mark>
gi 3796928 gi 3415738	 PQWSYMHIAGQDASEYLS PQWSYMHIAGQDASEYLS 	PGLVQFARATETYFSLGNK PGLVQFARATETYFTLGNK	FRNETVAETHDVTTDR-SQRIT	L-REVEVDREDTAYSYNAR	FTLAVGDNRVLDMASTYFDIRGVLDRGP : 104 FTLAVGDNRVLDMASSYFDIRGVLDRGP : 104
gi 1903565	: FQWSYMHIAGQDASEYLS	PGLVQFA RATD TYF <mark>T</mark> GNKF	FRNPTVAPTHDVTTDR-SQRL <mark>T</mark>	L-REVEVDRED TA YSY <mark>K</mark> AR	F <mark>T</mark> LAVGDNRVLDM <mark>A</mark> SSYFDIRG <mark>V</mark> LDRGF : 104
gi 1903565 gi 2705672	 PQWSYMHIAGQDASEYLS PQWSYMHIAGQDASEYLS 	PGLVQFARATDTYFTIGNKE	FRNPTVAPTHDVTTDR-SQRLT	I-REVEVDREDTTYSYRAR	FTLAVGDNRVLDM <mark>A</mark> SSYFDIRG <mark>VLDRGP : 104</mark> FTLAVGDNRVLDM <mark>A</mark> SSYFDIRG <mark>VLDRGP : 104</mark>
gi 313376	: POWSYMHIAGODASEYLS	PGLVQFARATETYFTIGNK	FRNETVAETHDVTTDR-SQRL <mark>T</mark>	L-RFVPVDRED TT YSY <mark>K</mark> AR	FTLAVGDNRVLDMASSYFDIRGVLDRGP : 104
gi 3799771 gi 2591571	: PQWSYMHIAGQDASEYLS: : PQWSYMHIAGQDASEYLS:	PGLVQFARATDTYF <mark>SL</mark> GNKF	FRNPTVAPTHDVTTDR-SQRLT	1-REVEVDREDTAYSYKAR	FTLAVGDNRVLDMASTYFDIRG <mark>VLDRGF : 104</mark> FCLAVGDNRLLDM <mark>G</mark> SSYFDIRG <mark>RLDRGF : 104</mark>
gi 4826657	: POWSYMHIAGODASEYLS	PGLVQFARATETYF <mark>S</mark> IGNK	FRNPTVAPTHDVTTDR-SQRL	L-REVEVDREDTAYSYNAR	FTLAVGDNRVLDMASTYFDIRG <mark>VLDRGF : 104</mark>
gi 3256590 gi 1243756	: POWSYMHIAGODASEYLS : POWSYMHIAGODASEYLS	PGLVQFARAT <mark>DTYFS</mark> IGNK	RNPTVAPTHDVTTDR-SQRLT	L-REVEVDREDTAYSYKAR	FTLAVGDNRVLDMASTYFDIRGVLDRGP : 104
gi 1108259	: POWSYMHIAGODASEYLS	PGLVQFARATDIIF SI GNKI	FRNPTVAPTHDVTTDR-SQRLT	L-REVEVEREDIAISIN	YTLAVGDNRVLDMASTYFDIRGVLDRGP : 104
gi 5321275	POWSYMHIAGODASEYLS	PGLVQFA <mark>R</mark> AT <mark>D</mark> TYF <mark>TL</mark> GNKF	FRNPT <mark>V</mark> APTHDVT <mark>TD</mark> R-SQRL <mark>T</mark>	L-REVEVDRED TA YSY <mark>B</mark> AR	F <mark>T</mark> LAVGDNRVLDM <mark>A</mark> SSYFDIRG <mark>VLDRGP : 104</mark>
gi 2699909 gi 3799770	: PQWSYMHIAGQDASEYLS: : PQWSYMHIAGQDASEYLS:	PGLVQFARATETYFELGNK	FRNPTVAPTHDVILDR-SQRLI FRNPTVAPTHDVILDR-SQRLI	L-REVEVEREDTAYSYNAR L-REVEVEREDTAYSYNAR	FTLAVGDNRVLDMASSYFDIRG <mark>VLDRGF : 104</mark> FTLAVGDNRVLDMASTYFDIRG <mark>VLDRGF : 104</mark>
gi 1903566	: PQWSYMHIAGQDASEYLS	PGLVQFARATDTYF <mark>S</mark> IGNKF	FRNPTVAPTHDVT <mark>TD</mark> R-SQRL <mark>T</mark>	L-REVEVDREETAYSYKVR	FTLAVGDNRVLDM <mark>A</mark> STYFDIRG <mark>VLDRGP : 104</mark>
gi 58561 e gi 5673154	: PQWSYMHIAGQDASEYLS : POWSYMHIAGODASEYLS	PGLVQFARATDTYFSIGNKE PGLVCFARATDTYFSIGNKE	RNETVAETHDVIIDE-SQRLI RNETVAETHDVIIDE-SOBLI	I-REVEVEREDTAYSYKVR	FTLAVGDNRVLDMASTYFDIRG <mark>VLDRGF : 104</mark> FTLAVGDNRVLDMASTYFDIRGVLDRGF : 104
gi 3416039	: POWSYMHIAGODASEYLS	PGLVQFA <mark>R</mark> AT <mark>D</mark> TYF <mark>S</mark> IGNKI	FRNPTVAPTHDVTTDR-SQRL <mark>T</mark>	L-RFVPVDRED TA YSY SV R	F <mark>T</mark> LAVGDNRVLDM <mark>A</mark> STYFDIRG <mark>VLDRGF : 104</mark>
gi 3799772 gi 3799771	: POWSYMHIAGODASEYLS : POWSYMHIAGODASEYLS	PGLVQFARATDTYFSIGNKE	RNPTVAPTHDVTTDR-SQRLT	L-REVEVDREDTAYSYRAR	FTLAVGDNRVLDMASTYFDIRGVLDRGF : 104 FTLAVGDNRVLDMASTYFDIRGVLDRGF : 104
gi 3799771	: POWSYMHIAGODASEYLS	PGLVQFA <mark>R</mark> AT <mark>D</mark> TYF <mark>SL</mark> GNK	RNPTVAPTHDVTTDR-SQRIT	L-REVEVDREDTAYSY AR	FTLAVGDNRVLDMASTYFDIRGVLDRGP : 104
gi 3799770 gi 3303688	: POWSYMHIAGODASEYLS: : POWSYMHIAGODASEYLS:	PGLVQFARATDTYFSIGNKE	FRNPTVAPTHDVTTER-SQRLT	L REVEVOREDTAYSYKAR	FTLAVGDNRVLDMASTYFDIRGVLDRGP : 104 FTLAVGDNRVLDMASTYFDIRGVLDRGP : 104
gi 3416039	: FQWSYMHIAGQDASEYLS	PGLVQFARATETTFS GNR	FRNPTVAPTHDVIIDR-SQRII	L-REVEVDREDTAYSYNY	F <mark>T</mark> LAVGDNRVLDM <mark>A</mark> STYFDIRG <mark>V</mark> LDRGF : 104
gi 3416039	: POWSYMHIAGODASEYLS	PGLVQFARATDTYF <mark>SI</mark> GNKF	FRNPTVAPTHDVT <mark>TD</mark> R-SQRLT	L-REVEVDRED TA YSY <mark>K</mark> VR	FTLAVGDNRVLDMASTYFDIRG <mark>VLDRGP : 104</mark>
gi 4826658 gi 4826657	: POWSYMHIAGODASEYLS : POWSYMHIAGODASEYLS	PGLVQFARATDTYFSUGNKE PGLVOFARATDTYFSUGNKE	RNETVAETHDVTTDR-SQRLT RNETVAETHDVTTBR-SOBLT	I-REVEVORED TA YSYKVR	YTLAVGDNRVLDMASTYFDIRGVLDRGP : 104 FTLAVGDNRVLDMASTYFDIRGMLDRGP : 104
gi 6039268	: POWSYMHIAGODASEYLS	PGLVQFA <mark>R</mark> AT <mark>D</mark> TYF <mark>SI</mark> GNKF	FRNFTVAFTHDVTTDR-SQRLT	L-REVEVDRED TA YSY <mark>K</mark> VR	Y <mark>T</mark> LAVGDNRVLDM <mark>A</mark> STYFDIRG <mark>V</mark> L DRGF : 104
gi 5084524 gi 3336014	 PQWSYMHIAGQDASEYLS: PQWSYMHIAGCDASEYLS: 	PGLVQFARATETYFSIGNK	FRNETVAETHEVTIER-SORLI	I-REVEVOKEDTAYSYKTR	FTLAVGDNRVLDMASTYFDIRGVIDRGP : 104 FTLAVGDNRVLDMASTYFDIRGVIDRGP : 104
gi 3256590	: POWSYMHIAGODASEYLS	PGLVQFARATETYF <mark>ST</mark> GNK	FRNPTVAPTHDVTTDR-SQRLT	I-REVEVDREDTAYSYKTR	FTLAVGDNRVLDMASTYFDIRGVIDRGP : 104
gi 6367741 gi 1088631	: PQWSYMHIAGQDASEYLS : PQWSYMHIAGQDASEYLS	ALVOF <mark>SO</mark> ATETYF <mark>O</mark> IGNKE	FRNPTVAPTHDVT <mark>TE</mark> R-SQRLQ	L-RETEVTQEDTQYAYKVR	FOLTVGDNRVLDMGSTYFDIRG <mark>RLDRGP : 104</mark> FOLAVGDNRVLDMASTYFDIRG <mark>TLDRGP : 104</mark>
gi 5280169	: FOWSYMHIAGODASEYLS:	PGLVQFAQATETYF <mark>KI</mark> GNKI	FRNPTVAPTHDVIIER-SQRIQ	L-REVEVOREDTQYTHETR	F <mark>Q</mark> LAVGDNRVLDM <mark>A</mark> STYFDIRG <mark>T</mark> L DRGF : 104
gi 1088631	: PQWSYMHIAGQDASEYLS: PQWSYMHIAGqdasEyLs;		FRNPT <mark>V</mark> APSHDVT <mark>T</mark> ER-SQRL <mark>Q</mark> frnpt6ap hdvtT r s rl	L-REVEVDREDTQYSY <mark>R</mark> TR vpvd ed v 4 r	FTLAVGDNRILDMASTYFDIRGTLDRGF : 104 avg nr 1dm s fdirg
	· AmorentwedgestArs	ba adra ac Ar eduki	ruptoap nuvti r s ri	vpvded y 4 r	ary hi iom a loting

Figure 3. Multiple alignments of (a) nucleotide sequence, and (b) deduced amino acid sequence of the hexon protein gene of Marine Adenovirus (KJ958986) with other adenoviruses obtained using GeneDoc programme Version 2.7.0. The alignment was performed with ClustalW and edited with GeneDoc software. The three levels of shading indicate different degrees of conservation. Black background and white letters corresponds to 100% conservation, gray background and white letters corresponds to 80% conservation, and gray background and black letters corresponds to 60% conservation.

protein genes. For a robust and reliable phylogenetic analysis, two methods were used, namely: maximum likelihood and neighbor joining methods. Neighbor joining tree was constructed using bootstrap test by 1000 replicates (Figure 4a and b).

Analysis of the results of the ML tree constructed using the nucleotide sequence showed that the tree could be broadly divided into two clusters. Cluster 1 included



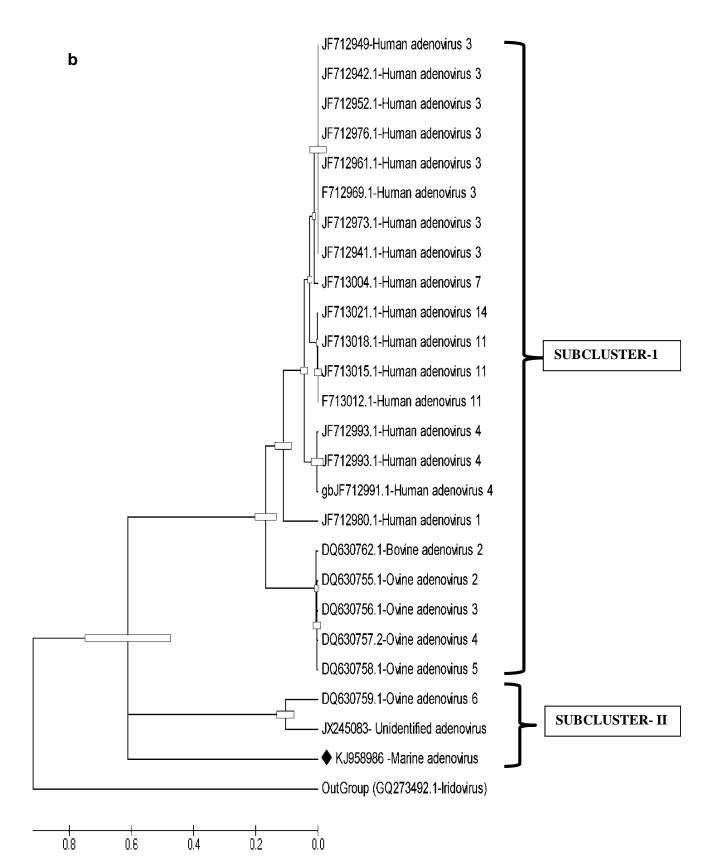


Figure 4. Bootstrapped (a) Neighbor-Joining Tree and (b) Maximum Likelihood Tree obtained using MEGA version 5.0 illustrating relationships between the nucleotide and deduced amino acid sequence of the hexon protein gene of Marine Adenovirus (KJ958986) with other adenoviruses. Values at the node indicate the percentage of times that the particular node occurred in 1000 trees generated by bootstrapping the original nucleotide sequences for the NJ tree.

adenoviruses isolated from various sources and cluster 2 is an outgroup. The present sequence was found to belong to cluster 1, thereby confirming the sequence identity. Cluster 1 could again be divided into 2 subclusters. Subcluster 1 contains human, bovine and canine adenoviruses, and subcluster 2 consists of an ovine adenovirus, an unidentified adenovirus and the adenovirus from the present study.

Analysis of the results of the NJ tree constructed using the amino acid was worth noticing. The tree could be broadly divided into two clusters. Cluster 1 includes adenoviruses isolated from various sources namely: human, bovine, ovine, simian, alpaca and bat; whereas, cluster 2 consists of equine, canine, bat, California sea lion and the newly identified adenovirus. The results of phylogenetic analysis based on the hexon protein sequences showed light onto the basal placement of marine adenoviruses in the family. It is shown that the marine adenoviruses isolated from organisms other than from humans. It was really interesting to see that the adenovirus isolated from sea lion also belonged to cluster 2.

Conclusion

Adenovirus has been considered as a key organism for uncovering new knowledge and providing insight into multiple aspects of animal cell biology and human antiviral defences. It also serves for adenovirus-based gene transduction vectors with engineered receptorbinding domains that promote infection of specific cell types. The viral proteins are potent engineered molecular tools for exploring these cellular processes. In this study, the hexon protein gene of an apparently novel marine adenovirus present in the marine waters off Cochin. India has been reported. The hexon coding sequence, coupled with bioinformatic analysis, demonstrated that the reported marine adenovirus is different from all previously characterized adenoviruses, and might be a novel marine adenovirus. The present study also reports the results of phylogenetic analyses based on the nucleotide and amino acid sequence data of hexon protein of family Adenoviridae. This is the first report of a marine adenovirus strain identified and characterized from the Cochin Barmouth region, India. The identified hexon protein showed less similarity to the already deposited hexon protein sequences of family Adenoviridae in GenBank. Multiple alignments showed the presence of conserved regions within the sequence. The phylogenetic analysis showed that the hexon protein of adenoviruses diverged from a common ancestral sequence into two major groups. The results of phylogenetic analysis based on the hexon protein sequences confirmed the basal placement of marine adenovirus in the family Adenoviridae. Molecular phylogenetic arrangements as inferred from the present study suggest the presence of

single origin for all the viruses belonging to family Adenoviridae. Increasing number of research in the fundamental aspects of adenovirus biology itself indicates the potential scope of the organism and endorses that adenovirus research continues to be a productive and rewarding area of discovery.

ACKNOWLEDGMENTS

The authors are grateful to the Department of Science and Technology (DST) and Ministry of Science and Technology, Government of India for the Research Grant (DST Letter No. DST/INSPIRE Faculty Award/2012). The authors are grateful to Prof. I. S. Bright Singh, Coordinator, NCAAH, CUSAT and Dr. Rosamma Philip, Associate Professor, Department of Marine Biology, CUSAT for all scientific support and for providing the necessary facilities. Thanks are also due to Dr. Valsamma Joseph and Dr. T.P Sajeevan, Assistant Professor, NCAAH, Dr. Jayesh P., Post-Doctoral Fellow, CUSAT and Mr. Afsal V.V., NETFISH-MPEDA, for all the scientific support.

REFERENCES

- Aragao GC, Oliveira DS, Santos MC, Mascarenhas JDP, Oliveira CSD, Linhares ADC, Gabbay YB (2010). Molecular characterization of norovirus, sapovirus and astrovirus in children with acute gastroenteritis from Belem, Para, Brazil. Rev. Pan-AmazSaude 1:149-158.
- Belsy A, Odalys V, Alexander P, Clara S, Angel G, Grehete G, Guelsys G, Luis S, Pedro M, Guadalupe GM, Alina L, Pilar PB, Inmaculada C (2009). Molecular characterization of adenoviral infections in Cuba: report of an unusual association of species D adenoviruses with different clinical syndromes. Arch Virol., 154:619-627.
- Bergh O, Børsheim KY, Bratbak G, Heldal M (1989). High abundance of viruses found in aquatic environments. Nature. 340:467-468.
- Bofill-Mas S, Albinana-Gimenez N, Clemente-Casares P, Hundesa A, Rodriguez-Manzano J, Allard A, Calvo M, Girones R (2006.).Quantification and Stability of Human Adenoviruses and Polyomavirus JCPyV in Wastewater Matrices. Appl and Environ. Microbiol., 72:7894-7896.
- Bosch A, Abad FX, Pinto RM (2006). Survival and transport of enteric viruses in the environment. In: S. M. Goyal (Ed.) "Viruses in Food", Food Microbiology and Food Safety Series. Springer, New York, pp 151-187.
- Breitbart M, Forest R (2005). Here a virus, there a virus, everywhere the same virus? Trends Microbiol., 13: 278–284.
- Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, Azam F, Rohwer F (2002). Genomic analysis of uncultured marine viral communities. Proc. Natl. Acad. Sci. U.S.A., 99: 14250–14255.

- Eischeid AC, Meyer JN, Linden KG (2009). UV Disinfeaction of Adenoviruses: Molecular Indications of DNA Damage Efficiency. Appl. Environ. Microbiol., 75:23-28.
- Fon T, Lipp EK (2005). Enteric Viruses of Humans and Animals in Aquatic Environments: Health Risks, Detection, and Potential Water Quality. Microbiol. Mol. Biol. Rev., 69:357–371.
- Forest R, Segall A, Steward G, Seguritan V, Breitbart M, Wolven F, Azam F (2000). The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with non-marine phages. Limnol. Oceanogr., 45:408–418
- Fuhrman JA, Campbell L (1998). Microbial Microdiversity. Nature. 393:410-411.

Girones R (2006). Identification of Human and Animal Adenoviruses and Polyoma viruses for Determination of Sources of Fecal Contamination in the Environment. Appl. Environ. Microbiol., 72:7886 7893.

Griffin DW, Donaldson KA, Paul JH, Rose JB (2003). Pathogenic Human Viruses in Coastal Waters. Clin. Microbiol. Rev., 16:129–143.

Hundesa A, Maluquer dMC, Bofill MS, Albinana GN,

- Jiang S, Noble R, Chu W (2001). Human Adenoviruses and Coliphages in UrbanRunoff: Impacted Coastal Waters of Southern California. Appl. Environ. Microbiol., 67:179–184.
- John SG, Mendez CB, Deng L, Poulos B, Kauffman AK, Kern S, Brum J, Polz MF, Boyle EA, Sullivan MB (2011). A simple and efficient method for concentration of ocean viruses by chemical flocculation. Environ Microbiol Rep.3:195-202
- Jothikumar N, Cromeans TL, Hill VR, Lu X, Sobsey MD, Erdman DD (2005).Quantitative Real Time PCR Assays for Detection of Human Adenoviruses and Identification of Serotypes 40 and 41. Appl. Environ. Microbiol., 71:3131-3136.
- Kanera A, Hemminki A (2004). Modified adenoviruses for cancer gene therapy. Int. J. Cancer. 110:475-480.
- Ko G, Cromeans TL, Sobsey MD (2003). Detection of Infectious Adenovirus in Cell Culture by mRNA Reverse Transcription-PCR. Appl. Environ. Microbiol., 69:7377-7384.
- Linden KG, Thurston J, Schaefer R, Malley JP Jr (2007). Enhanced UV Inactivation of Adenoviruses under Polychromatic UV Lamps. Appl. Environ. Microbiol., 73:7571-7574.
- Mahony JB (2008). Detection of Respiratory Viruses by Molecular Methods. Clin. Microbiol Rev., 21: 716-747.

- Muscillo M, Pourshaban M, Iaconelli M, Fontana S, Di Grazia A, Manzara S, Fadda G, Santangelo R, La Rosa G (2008). Detection and Quantification of Human Adenoviruses in Surface Waters by Nested PCR, TaqMan Real- Time PCR and Cell Culture Assays. Water Air Soil Pollut 191:83-93.
- Nicholas B, Paul K (2011). Metagenomics and the molecular identification of novel viruses. Vet. J., 190: 191–198.
- Paul JH, Sullivan MB (2005). Marine phage genomics: what have we learned? Curr. Opin. Biotechnol., 16:299–307.
- Pina S, Puig M, Lucena F, Jofre J, Girones R (1998). Viral Pollution in the Environment and in Shellfish:Human Adenovirus Detection by PCR as an Index of Human Viruses. Appl. Environ. Microbiol., 64:3376 – 3382.
- Pond K (2005). Water recreation and disease: plausibility of associated infections: Acute Effects, Sequelae and Mortality. World Health Organization.
- Rein DT, Breidenbach M, David TC (2007). Current developments in adenovirus-based cancer gene therapy. Future Oncol., 2:137–143.
- Rein DT, Volkmer A, Beyer IM, Curiel DT, Janni W, Dragoi A, Hess AP, Maass N, Baldus SE, Bauerschmitz G, Breidenbach M (2011). Treatment of chemotherapy resistant ovarian cancer with a MDR1 targeted oncolytic adenovirus. Gynecol. Oncol., 123:138-146.
- Rohwer F, Barott K (2013).Viral information. Biology & Philosophy. 28:283-297.
- Suttle CA (2005). Viruses in the sea. Nature. 437: 356–361.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods, Mol. Biol. Evol., 28:2731–2739.
- Xagoraraki I, Kuo DH, Wong K, Wong M, Rose JB (2007). Occurrence of Human Adenoviruses at Two Recreational Beaches of Great Lakes. Appl. Environ. Microbiol., 73:7874-7781.