



**Jaypee University of Information Technology  
Solan (H.P.)  
LEARNING RESOURCE CENTER**

Acc. Num **SP04038** Call Num:

**General Guidelines:**

- ◆ Library books should be used with great care.
- ◆ Tearing, folding, cutting of library books or making any marks on them is not permitted and shall lead to disciplinary action.
- ◆ Any defect noticed at the time of borrowing books must be brought to the library staff immediately. Otherwise the borrower may be required to replace the book by a new copy.
- ◆ The loss of LRC book(s) must be immediately brought to the notice of the Librarian in writing.

Learning Resource Centre-JUIT



SP04038



# SEQUENCE SIMILARITIES BETWEEN VIRUSES WITH GENOME AND EST DATABASES

By

**Priyanka Tandon (041539)**



**MAY-2008**



**Submitted in partial fulfillment of the Degree of Bachelor of  
Technology**

**DEPARTMENT OF BIOTECHNOLOGY AND  
BIOINFORMATICS  
JAYPEE UNIVERSITY OF INFORMATION  
TECHNOLOGY-WAKNAGHAT  
SOLAN, H.P.**

## CERTIFICATE

This is to certify that the work entitled, "Sequence similarities between viruses with genome and EST database." submitted by Priyanka Tandon in partial fulfillment for the award of degree of Bachelor of Technology in Bioinformatics of Jaypee University of Information Technology has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

  
Dr. C. D. Poduri

## **ACKNOWLEDGEMENT**

First and foremost we would like to thank our project guide Dr. C.D.Poduri for the invaluable advice and positive encouragement he has provided us throughout the course of this project and helped at all stages during the project and through the preparation of the dissertation. We extend our gratitude towards all our fellow students for making the four years of B.Tech. the most memorable ones. Finally, we appreciate the continuous love and support of our friends and families and always being there for us.



## TABLE OF CONTENTS

	Page no.
Certificate	02
Acknowledgement	03
Table of Content	04 - 05
Index of tables	06
List of abbreviations	06
Abstract	07
Introduction	08-09
1. Japanese encephalitis virus and Chikungunya virus	10-12
Classification	10
Japanese encephalitis virus (JEV)	10
1.3 Life cycle of JEV	11
1.4 Chikungunya virus	11
1.5 Life cycle of CHIK virus	12
2. Influenza A Virus and Bird flu Virus	12-13
2.1 Classification	12
2.2 Influenza A Virus	13
2.3 Bird flu Virus	13
3. Bovine diarrhea virus (BVDV) and Hepatitis C virus (HCV)	14-15
3.1 Classification	14
3.2 Bovine diarrhea virus (BVDV)	14
3.3 Life cycle of BVDV	15
3.4 Hepatitis C virus (HCV)	15

	<b>Page no.</b>
Protocol of the project	17
Conclusion	48
Bibliography	49

## INDEX OF TABLES

	<b>Page no.</b>
<b>Table I</b> Sequence comparison of JEV and CHIK with human and mosquito ESTs	18-19
<b>Table II</b> Sequence comparison of Influenza A virus with human ESTs	119-26
<b>Table III</b> Sequence comparison of Influenza A virus with duck ESTs	27-34
<b>Table IV</b> Sequence comparison of BVDV with bovine ESTs	35-36
<b>Table V</b> Sequence comparison of BVDV with human ESTs	36
<b>Table VI</b> Sequence comparison of BVDV with various databases in NCB	37-44
<b>Table VII</b> Sequence comparison of HCV with human ESTs in TIGR	44
<b>Table VIII</b> Sequence comparison of HCV with various databases in NCBI.	45-47

### LIST OF ABBREVIATIONS:

**JEV:** Japanese encephalitis Virus

**CHIK:** Chickungunya virus

**BVDV:** Bovine diarrhea virus

**HCV:** Hepatitis C virus



## ABSTRACT

Sequence similarities between viruses with genome and EST databases give a better insight into the life cycle and molecular epidemiology of the viruses being studied. Information thus provided aids in the development of better therapeutic agents. The aim of the project is to check for sequence similarities between viruses [JEV and CHIK virus] with genome (*Aedes* sp., *Culex* sp.) and EST (human, porcine, mosquito) database. To check for the sequence similarities between Influenza A virus and bird flu virus with EST (human, mallard) database.

The other viruses which here studied together were bovine diarrhea virus and hepatitis C virus. Online resources were used for carrying out sequence similarity searches. There are contamination warnings in results that were obtained upon sequence comparison of BVDV (accession id NC\_001461) with human ESTs while no contamination was observed in the results obtained upon sequences comparison of HCV (accession id AY051292) with human ESTs. Only two human ESTs showed sequence similarity with HCV.

## INTRODUCTION

A sequence can tell us what the organism can do, about the history of the organism, and gives leverage for understanding how the organism works. With specific sequences we can find out more about structure, function and evolution.

Sequence analysis has become such an important field because the sequence of a biomolecule (DNA or protein) carries a lot of information about the biomolecule's function and history.

1. JEV sequences matched with a number of human ESTs which are probable contamination entries. This in part suggests that the corresponding regions of JEV might also represent vector contamination, necessitating regeneration.
2. Consequently, the CHIK genome/database, InfluenzaA genome/database.

In particular, we usually assume that similar sequences hold a similar function, structure and they have a similar evolution history. Significance of doing sequence similarity is that it helps in doing phylogenetic analysis, annotation of genome and aids in the development of better diagnostics and therapeutics, apart from revealing useful insight into the biology of the organism involved.

Such an exercise makes sense for the simple reason that previous researchers have reported significant of human genome components and human ESTs with a number of viral genome and pathogens.

In the present project our interest was in looking for sequence similarities of viruses and human pathogens (JEV, CHIK, InfluenzaA, Bird flu) initially with human ESTs and subsequently with human genome. Rationale for choosing JEV, CHIK, InfluenzaA and bird flu are the viruses involved in infectious diseases. Very little

information with regard to their pathobiology is available. Consequently any work on these viruses gains prominence. Among these viruses JEV is phylogenitically related to hepatitis C virus. Accordingly the information that is generated using JEV can impart be applied to hepatitis C virus.

Significance of the project as stated earlier lies in the observation that this project leads to identification of molecular pathogenic mechanisms and hence to better diagnostics, therapeutics and vaccines.



## 1. Japanese encephalitis virus and Chikungunya virus:

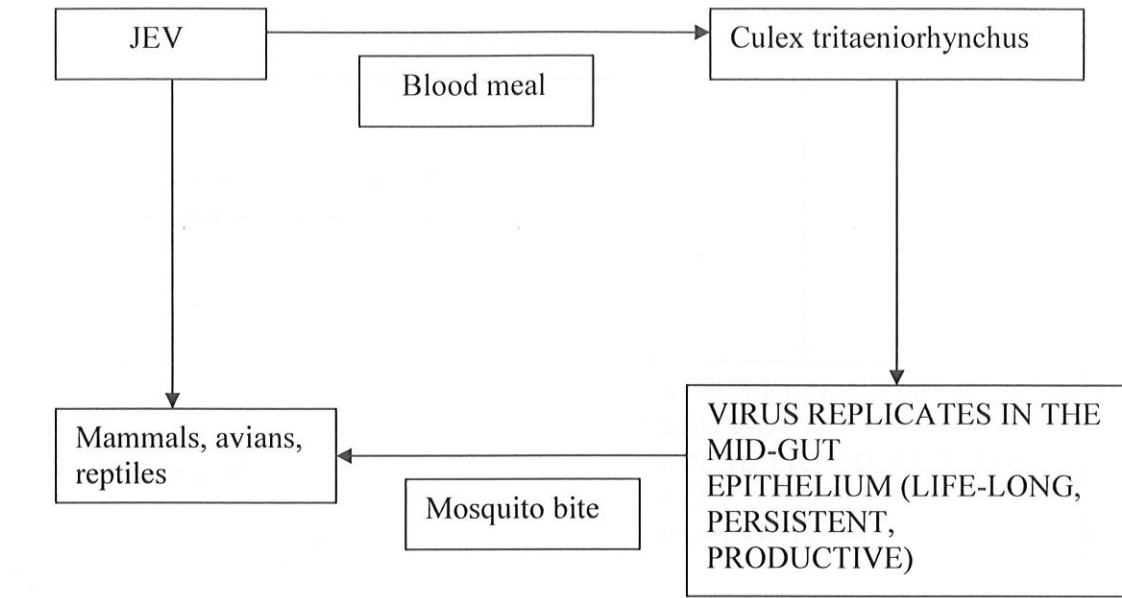
### 1.1 Classification:

	Japanese encephalitis virus	Chikungunya virus
FAMILY	<i>Flaviviridae</i>	<i>Togaviridae</i>
GENUS	Flavivirus	Alphavirus
SPECIES	Japanese encephalitis virus	Chikungunya virus

### 1.2 Japanese encephalitis virus (JEV):

The NCBI accession id of Japanese encephalitis virus is NC 001437. It is classified in the group B arboviruses. Initially it was placed in the family *togaviridae* because of its single stranded positive sense RNA genome but now because of its unique genomic organization, structural and replication strategy, this genus then became the genus flavivirus of the independent family *flaviviridae*. The genome of Japanese encephalitis virus is single stranded, positive sense linear RNA molecule of 45S. The genome consists of about 10976 nucleotides. Its 3' end does not contain a poly (A) tract, but is postulated to form a stem-loop secondary structure; to which some cellular regulatory proteins could interact. Vector of Japanese encephalitis virus is *Culex tritaeniorhynchus*. Acute encephalitis was described in 1981 in Japan.

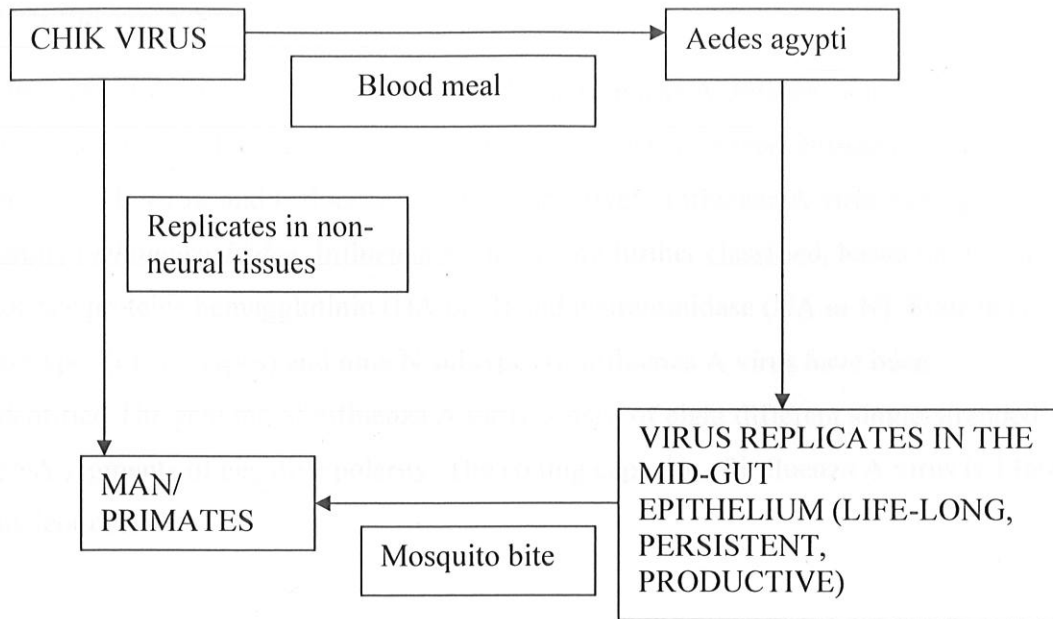
### 1.3 Life cycle of JEV:



### 1.4 Chikungunya virus (CHIK):

The NCBI accession id of chikungunya virus is NC 004162. It is classified in the group A arboviruses. It is placed in the family *Togaviridae* because of its antigenic properties and single stranded positive sense RNA. It is placed in *Alphavirus* genus. Vector of chikungunya virus is *Aedes* species. The genome contains about 11826 nucleotides arranged as single, nonsegmented positive sense strands of RNA. 1st break of this fever-Tanzania 1953 from febrile illness in human and 2nd break of this fever-South Africa 1956. The genomes are polyadenylated at the 3' end and capped with 7-methylguanosine at the 5' end.

**1.5 Life cycle of CHIK virus:**



**2. Influenza A Virus and Bird flu Virus:**

**2.1 Classification:**

	<b>Influenza A Virus</b>	<b>Bird flu</b>
FAMILY	<i>Orthomyxoviridae</i>	<i>Orthomyxoviridae</i>
GENUS	Influenzavirus A	Influenzavirus A
SPECIES	Influenza A virus	Bird flu virus



## 2.2 Influenza A Virus

There are three genera of influenza virus: Influenzavirus A, Influenzavirus B and Influenzavirus C. Each genus includes only one species, or type: Influenza A virus, Influenza B virus, and Influenza C virus, respectively. Influenza A virus belongs to the family *orthomyxoviridae*. Influenza A viruses are further classified, based on the viral surface proteins hemagglutinin (HA or H) and neuraminidase (NA or N). Sixteen H subtypes (or serotypes) and nine N subtypes of influenza A virus have been identified. The genome of influenza A virus consists of eight different single-stranded RNA segments of negative polarity. The coding capacity of influenza A virus is 13600 nucleotides.

## 2.3 Bird Flu Virus

Bird flu is classified as a type A influenza virus and belongs to the family *orthomyxoviridae*. They are single stranded RNA viruses. There are many subtypes of influenza viruses that cause avian flu. The most notable of these avian flu virus subtypes is H5N1. In August 1997, the first cases of human bird flu infection emerged in Hong Kong.

### 3. Bovine diarrhea virus and Hepatitis C virus

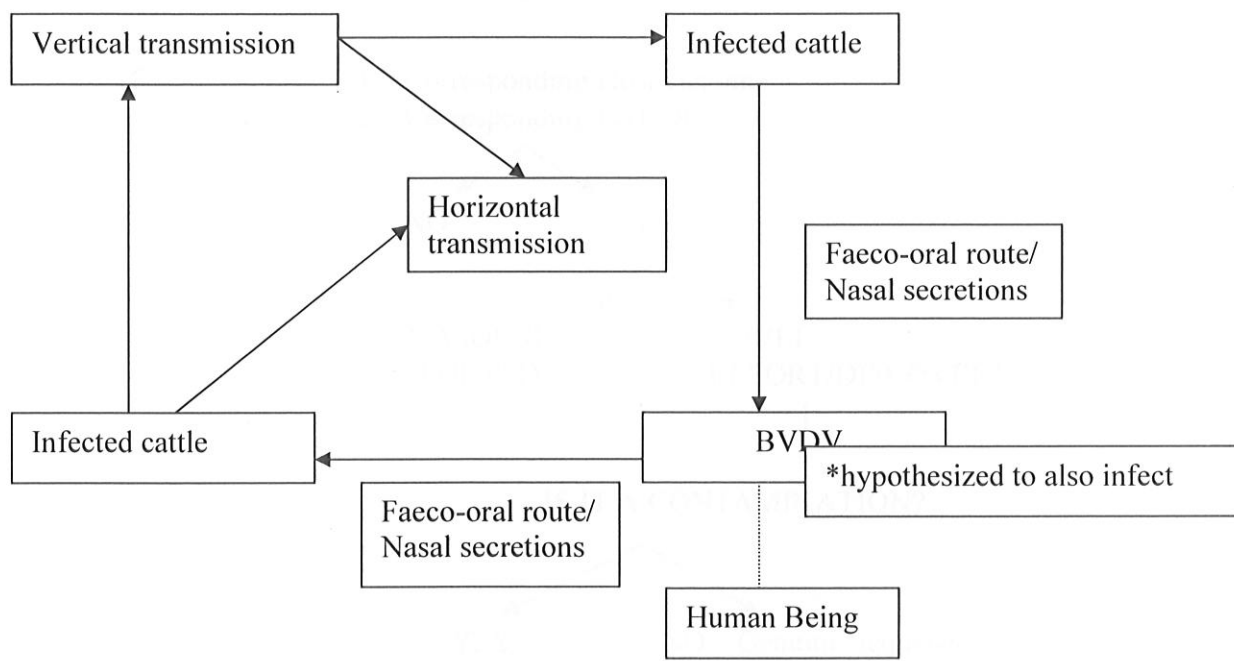
#### 3.1 Classification:

	<b>Bovine diarrhea virus</b>	<b>Hepatitis C virus</b>
FAMILY	<i>Flaviviridae</i>	<i>Flaviviridae</i>
GENUS	Pestivirus	Hepacivirus
SPECIES	Bovine diarrhea virus	Hepatitis C virus

#### 3.2 Bovine diarrhea virus:

**BVDV** is classified into 2 genotypes i.e. 1a, 1b and 2. In the current project the accession id we have taken is NC\_001461. BVDV is classified in the pestivirus genus of the family Flaviviridae, are small enveloped viruses having a single-stranded positive-sense RNA genome. The BVDV genome is approximately 12.3 kb in length with a 5'-nontranslated region (NTR), a single large open reading frame (ORF) and a 3'-NTR lacking a poly(A) tail. The 5'-NTR contains an internal ribosome entry site that initiates translation of BVDV Mrna in a cap-independent manner

### 3.3 Life cycle of BVDV:



\* Current science; Oct 2006; 91#8:996-997

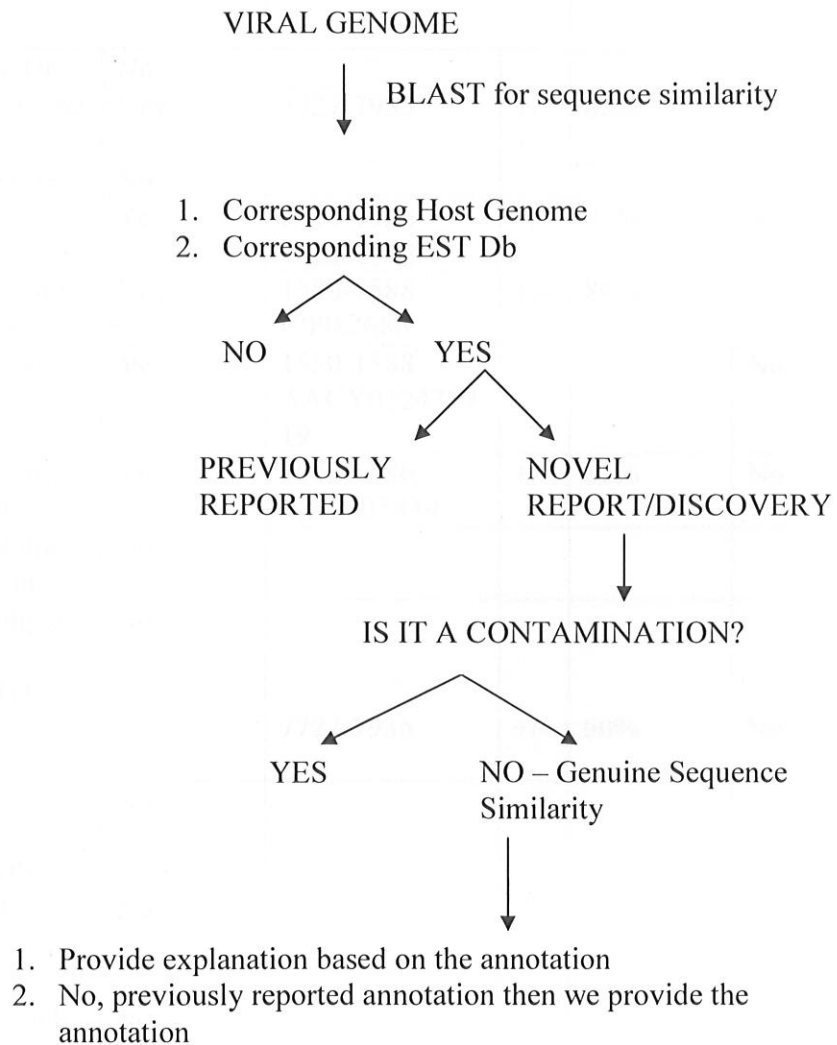
### 3.4 Hepatitis C virus:

HCV is classified into 6 genotypes and 18 subtypes. In the ongoing project we have taken full length genome of the Indian strain of HCV (accession id is AY051292) belonging to genotype 1 and subtype 1c of HCV. Hepatitis C virus (HCV) a positive –strand RNA virus, has been considered to have a poly (U) stretch at the 3' terminus of the genome.



## PROTOCOL OF THE PROJECT:

The broad overview of the project is given in the following flowchart (fig1)



Query Seq.	Blasted with/aligned with	Sequences showing similarity	Position on the viral genome (Accession id if Position on the viral genome (Accession id if similarity found)	+/+ or +/+ or +/-	% similarity	Contamination
<b>CHIK VIRUS NC_004162</b>	Human genome database	No				
	Human EST Db	No				
	Mosquito genome Db	Yes	7727-7935	+/ +	60%	No
	Mouse Genome	No				
	Mouse EST(EST Db)	Yes	6719-7326 CX206572.1	+/-	73%	No
	Non Human Non Mouse ESTDb	Yes	1550-1588 C0982686	+/-	84%	No
	Environmental samplesDb	Yes	1550-1588 AACY0224709 19			No
	Genome survey sequence Db	Yes	1548-1586 CW9203839	+/-	87%	No
	Human Genome plus transcript	No				
	High throughput genome sequence(HTGS)	No				
	Ades Agypti genome Db	Yes	7727-7935	+/-	60%	No
	Sequence Tagged Sites(STS)Db	No				
	Human ALU repeat elements Db	No				
	Protein databank Db	No				
	Reference genomic seq Db	No				
<b>JEV NC_001437</b>	Human genome database	No				
	Human EST Db	No				

	Mosquito genome Db	No				
	Mosquito EST	No				
	Horse genome Db	No				
	Horse EST Db	No				
	Porcine genome Db	No				
	Porcine EST Db	No				
	Non Human Non Mouse ESTDb	No				
	Environmental samplesDb	No				
	Genome survey sequence Db	No				
	Human Genome plus transcript	No				
	High throughput genome sequence(HTGS)	No				
	Sequence Tagged Sites(STS)Db	No				
	Human ALU repeat elements Db	No				
	Protein databank Db	No				

	Reference genomic seq Db	No				
	Nucleotide collection Db	yes	AY453412 (Usutu virus)		78%	
			AY898809 (Alfuy virus)		71%	
			AY490240 ( West nile virus)		69%	
			AY532665 ( West nile virus)		86%	

**TableII**

The following results were obtained upon sequence comparison of InfluenzaA (accession id: AC\_000145) with human ESTs

Query Seq	Blasted with/Aligned with	Sequences showing similarity	Position on the viral genome	+/- strand or -/- strand	% similarity	contamination	Accession id of animal showing similarity
CY011247	Human genomic plus transcript	NO					
	Mouse genomic plus transcript	NO					
	Nucleotide Collection	YES					With influenza A virus
	Reference mRNA seq	NO					
	Reference genomic sequence	NO					

	Expressed sequence tags	NO					
	Genomic Survey Seq	NO					
	High Throughput genomic seq	NO					
	Patent seq	YES		+/-	85%		AAKN0148 88701
	Protein Databank	NO					
	Human ALU repeat elements	NO					
	Sequence Tagged sites	NO					
	Whole-Genome shotgun reads	NO					
	Environmental samples	NO					
<b>CY011246</b>	Human genomic plus transcript	NO					
	Mouse genomic plus transcript	NO					
	Nucleotide Collection	NO					
	Reference mRNAseq	NO					
	Reference genomic sequence	NO					
	Expressed sequence tags	NO					
	Non human Non mouse ESTs	NO					
	Genomic Survey Seq	NO					



	High Throughput genomic seq	YES					AC080074
	Patent seq						
	Protein Databank	NO					
	Human ALU repeat elements	NO					
	Sequence Tagged sites	NO					
	Whole-Genome shotgun reads	YES					AAFR0304 71731  AAGV0108 44711
	Environmental samples	NO					
<b>CY011245</b>	Human genomic plus transcript	NO					
	Mouse genomic plus transcript	NO					
	Nucleotide Collection	YES					
	Reference mRNAseq	NO					
	Reference genomic sequence	NO					
	Expressed sequence tags	NO					
	Non human Non mouse ESTs	NO					
	Genomic Survey Seq	NO					
	High Throughput genomic seq	YES		+/+	85%		AC1644752
	Patent seq	YES					
	Protein	NO					

	Databank						
	Human ALU repeat elements	NO					
	Sequence Tagged sites	NO					
	Whole-Genome shotgun reads	YES		+/+			AAFC03010297
				+/-			AADN02021492
	Environmental samples	NO					
<b>CY011240</b>	Human genomic plus transcript	NO					
	Mouse genomic plus transcript	NO					
	Nucleotide Collection	YES					With influenza A virus
	Reference mRNAseq	NO					
	Reference genomic sequence	NO					
	Expressed sequence tags	YES	1024-1107	+/-	79%		BJ838932
	Non human Non mouse ESTs	YES	1024-1107	+/-	81%		CL840901
	Genomic Survey Seq	YES					
	High Throughput genomic seq	NO					
	Patent seq	YES					
	Protein Databank	NO					
	Human ALU repeat elements	NO					



	Sequence Tagged sites	NO					
	Whole-Genome shotgun reads	NO					
	Environmental samples	YES	1399-1446	+/+	81%		AACY021585946
<b>CY011243</b>	Human genomic plus transcript	NO					
	Mouse genomic plus transcript	NO					
	Nucleotide Collection	YES					
	Reference mRNAseq	NO					
	Reference genomic sequence	NO					
	Expressed sequence tags	NO					
	Non human Non mouse ESTs	NO					
	Genomic Survey Seq	NO					
	High Throughput genomic seq	NO					
	Patent seq	YES					
	Protein Databank	NO					
	Human ALU repeat elements	NO					
	Sequence Tagged sites	NO					
	Whole-Genome shotgun reads	NO					
	Environmental samples	YES					

<b>CY011242</b>	Human genomic plus transcript	NO					
	Mouse genomic plus transcript	NO					
	Nucleotide Collection	YES					With influenza A virus
	Reference mRNAseq	NO					
	Reference genomic sequence	NO					
	Expressed sequence tags	YES					
	Non human Non mouse ESTs	NO					
	Genomic Survey Seq	NO					
	High Throughput genomic seq	NO					
	Patent seq	YES					
	Protein Databank	NO					
	Human ALU repeat elements	NO					
	Sequence Tagged sites	NO					
	Whole-Genome shotgun reads	NO					
	Environmental samples						
<b>CY011241</b>	Human genomic plus transcript	NO					
	Mouse genomic plus transcript	NO					

	Nucleotide Collection	NO					
	Reference mRNAseq	NO					
	Reference genomic sequence	NO					
	Expressed sequence tags	NO					
	Non human Non mouse ESTs	NO					
	Genomic Survey Seq	NO					
	High Throughput genomic seq	NO					
	Patent seq	YES					
	Protein Databank	NO					
	Human ALU repeat elements	NO					
	Sequence Tagged sites	NO					
	Whole-Genome shotgun reads	NO					
	Environmental samples	NO					
<b>CY011244</b>	Human genomic plus transcript	YES					
	Mouse genomic plus transcript	NO					
	Nucleotide Collection	YES					With influenza A virus
	Reference mRNAseq	NO					
	Expressed sequence tags	YES	690-730	+/+	82%		BE066029

			690-730	+/-	82%		BE066024
			690-730	+/+	82%		AA857328
	Non human Non mouse ESTs	NO					
	Genomic Survey Seq	NO					
	High Throughput genomic seq	NO					
	Patent seq						
	Protein Databank	NO					
	Human ALU repeat elements	NO					
	Sequence Tagged sites	NO					
	Whole- Genome shotgun reads	YES					AAFR0302 7744
	Environmenta l samples	NO					

**TableIII:**

The following results were obtained upon sequence comparison of InfluenzaA (accession id: AC\_000145) with duck ESTs

Query Seq	Blasted with/Aligned with	Sequences showing similarity	Position on the viral genome	+/- strand or -/- strand	% similarity	contamination	
CY005231	Human genomic plus transcript	YES	193-246	+/-	77%		NW_923306
	Mouse genomic plus transcript	YES	1158-1200	+/+	81%		NT_039706
	Nucleotide Collection	YES					
	Reference mRNAseq	NO					
	Reference genomic sequence	NO					
	Expressed sequence tags	NO					
	Non human Non mouse ESTs	NO					
	Genomic Survey Seq	NO					
	High Throughput genomic seq	NO					
	Patent seq	NO					
	Protein Databank	NO					
	Human ALU repeat elements	NO					
	Sequence	NO					

	Tagged sites						
	Whole-Genome shotgun reads	NO					
	Environmental samples	NO					
<b>CY005230</b>	Human genomic plus transcript	NO					
	Mouse genomic plus transcript	NO					
	Nucleotide Collection	NO					
	Reference mRNAseq	NO					
	Reference genomic sequence	NO					
	Expressed sequence tags	NO					
	Non human Non mouse ESTs	NO					
	Genomic Survey Seq	NO					
	High Throughput genomic seq	YES	927-971 1158-1200	+/- +/>+	82% 81%		CU469146 BX000532
	Patent seq						
	Protein Databank	NO					
	Human ALU repeat elements	NO					
	Sequence Tagged sites	NO					

	Whole-Genome shotgun reads	YES	184-238	+/-	80%		AAGU01365565
	Environmental samples	NO	1150-1191	+/+	86%		AACY020867595
<b>CY005229</b>	Human genomic plus transcript	YES	1082-1123	+/+	80%		NW_926128
	Mouse genomic plus transcript	NO					
	Nucleotide Collection	NO					
	Reference mRNAseq	NO					
	Reference genomic sequence	NO					
	Expressed sequence tags	NO					
	Non human Non mouse ESTs	NO					
	Genomic Survey Seq	NO					
	High Throughput genomic seq	NO					
	Patent seq	NO					
	Protein Databank	NO					
	Human ALU repeat elements	NO					
	Sequence Tagged sites	NO					
	Whole-Genome shotgun	NO					



	reads						
	Environmental samples	NO					
CY005997	Human genomic plus transcript	NO					
	Mouse genomic plus transcript	NO					
	Nucleotide Collection	NO					
	Reference mRNAseq	NO					
	Reference genomic sequence	NO					
	Expressed sequence tags	NO					
	Non human Non mouse ESTs	NO					
	Genomic Survey Seq	NO					
	High Throughput genomic seq	NO					
	Patent seq	NO					
	Protein Databank	NO					
	Human ALU repeat elements	NO					
	Sequence Tagged sites	NO					
	Whole-Genome shotgun reads	NO					
	Environmental samples	NO					

CY011243	Human genomic plus transcript	NO					
	Mouse genomic plus transcript	NO					
	Nucleotide Collection	NO					
	Reference mRNAseq	NO					
	Reference genomic sequence	NO					
	Expressed sequence tags	NO					
	Non human Non mouse ESTs	NO					
	Genomic Survey Seq	NO					
	High Throughput genomic seq	NO					
	Patent seq	NO					
	Protein Databank	NO					
	Human ALU repeat elements	NO					
	Sequence Tagged sites	NO					
	Whole-Genome shotgun reads	YES	570-606		86%	+/-	AAVX01252123
	Environmental samples	YES					
CY005226	Human genomic plus	NO					

	transcript						
	Mouse genomic plus transcript	NO					
	Nucleotide Collection	NO					
	Reference mRNAseq	NO					
	Reference genomic sequence	NO					
	Expressed sequence tags	YES	2-55		90%	+/-	EG510406
	Non human Non mouse ESTs	NO					
	Genomic Survey Seq	NO					
	High Throughput genomic seq	NO					
	Patent seq	YES					
	Protein Databank	NO					
	Human ALU repeat elements	NO					
	Sequence Tagged sites	NO					
	Whole-Genome shotgun reads	NO					
	Environmental samples						
CY005225	Human genomic plus transcript	NO					
	Mouse genomic	NO					

	plus transcript						
	Nucleotide Collection	YES					
	Reference mRNAseq	NO					
	Reference genomic sequence	NO					
	Expressed sequence tags	NO					
	Non human Non mouse ESTs	NO					
	Genomic Survey Seq	NO					
	High Throughput genomic seq	NO					
	Patent seq	NO					
	Protein Databank	NO					
	Human ALU repeat elements	NO					
	Sequence Tagged sites	NO					
	Whole-Genome shotgun reads	NO					
	Environmental samples	NO					
<b>CY005228</b>	Human genomic plus transcript	NO					
	Mouse genomic plus transcript	NO					
	Nucleotide	NO					

	Collection						
	Reference mRNAseq	NO					
	Reference genomic sequence	NO					
	Expressed sequence tags	NO					
	Non human Non mouse ESTs	NO					
	Genomic Survey Seq	NO					
	High Throughput genomic seq	NO					
	Patent seq						
	Protein Databank	NO					
	Human ALU repeat elements	NO					
	Sequence Tagged sites	NO					
	Whole-Genome shotgun reads	YES					
	Environmental samples	NO					

**Table IV:**

The following results were obtained upon sequence comparison of BVDV (accession id: NC\_001461) with bovine ESTs:

Query Seq	LIST OF ALL BOVINE ESTs MATCHING WITH THE QUERY	REGION OF MATCH ON QUERY SEQ	ANNOTATION OF EST	VEC SCREEN RESULT	REMARK
NC_001461	TC301175	4844-5265	UP Q95J56 BOVIN (Q95J56) J-domain protein	-	No contamination
	NP398115	4844-5265	Bos Taurus DnaJ1 protein mRNA	-	No contamination
	DV786950	1144- 1150	Liver cDNA library Bos Taurus cDNA,mRNA seq		
	DT815591	9966-10264	Bos Taurus Cdna clone, Mrna seq	-	No contamination
	DV793902	892-1111	Bos Taurus Cdna clone, Mrna seq	-	
	DV798853	892-1144	Bos Taurus cdna,Mrna seq	-	
	TC330861	10031-10264	CTCL tumor antigen se57-1 isoform 1 (Bos taurus)	-	No contamination
	DV787667	900-1140	Liver cdna library Bos Taurus cdna,mrna seq		
	DV776761	899-1140	Bos Taurus cdna .mrna		

			seq		
	CF766560	10377-10535	Bos Taurus cdna,mrna seq		
	DV797005	892-1076	Bos Taurus cdna .mrna seq		
	TC373770	3945-4059	Bos Taurus cdna .mrna seq		
	DV798705	892-1076	Bos Taurus cdna .mrna seq		

**Table V:**

The following results were obtained upon sequence comparison of BVDV (accession id: NC\_001461) with human ESTs:

Query Seq	LIST OF ALL BOVINE ESTs MATCHING WITH THE QUERY	REGION OF MATCH ON QUERY SEQ	ANNOTATION OF EST	VEC SCREEN RESULT	REMARK
NC_001461	CV023482	4994--5262	DnaJ(Hsp40) homolog Hs253844	-	No contamination
	THC2461761	4844-5265		-	No contamination
	THC2534509				
	BG546231	9966-10264		-	No contamination
	AV706129	892-1111		-	



**Table VI:**

The following results were obtained upon sequence comparison of BVDV with various databases in NCBI:

Query Seq	Blasted with/Aligned with	Sequences showing similarity	Position on the viral genome	+/- strand or -/- strand	% similarity	description
NC_001461	Human genomic plus transcript	YES	4994-5265	+/+	94%	Homo sapiens DnaJ (Hsp40) homolog, subfamily C, member 14 ( <b>DNAJC14</b> ), mRNA
			5021-5183	+/-	96%	Homo sapiens chromosome 12 genomic contig, alternate assembly (based on Celera assembly) <u>dopamine receptor interacting protein</u>
			12409-12359	+/+	80%	Homo sapiens chromosome 11 genomic contig, reference assembly. <u>leucine zipper protein 2</u>
			12409-12359	+/+	78%	Homo sapiens chromosome 5 genomic contig, alternate assembly (based

	<b>Mouse genomic plus transcript</b>	YES	4994-5260	+/+	94%	Mus musculus DnaJ (Hsp40) homolog, subfamily C, member 14 ( <b>Dnajc14</b> ), Mrna
			5021-5183	+/+	95%	Mus musculus <b>chromosome 10</b> genomic contig, strain <b>C57BL/6J</b>
			5009-5254	+/-	76%	Mus musculus <b>chromosome 9</b> genomic contig, strain <b>C57BL/6J</b>
	<b>Nucleotide Collection</b>	YES				With BVDV
	<b>Refrence mRNAseq</b>	YES	4994-5265	+/+	99%	<b>Bos taurus</b> DnaJ (Hsp40) homolog, subfamily C, member 14 ( <b>DNAJC14</b> ), mRNA
			4994-5265	+/+	95%	<b>Canis familiaris</b> similar to dopamine receptor interacting protein ( <b>LOC474392</b> ), Mrna
			4994-5265	+/+	95%	Homo sapiens DnaJ (Hsp40) homolog, subfamily C, member 14 ( <b>DNAJC14</b> ), mRNA
						<b>Equus caballus</b> similar to J-domain protein Jiv

			4994-5265	+/+	94%	(LOC100051590), Mrna  <b>Pan troglodytes</b> dopamine receptor interacting protein (DNAJC14), mRNA
			4994-5265	+/+	94%	<b>Macaca mulatta</b> similar to dopamine receptor interacting protein, transcript variant 3 (LOC710206), Mrna
			4994-5265	+/+	94%	<b>Mus musculus</b> DnaJ (Hsp40) homolog, subfamily C, member 14 (Dnajc14), mRNA
			4994-5265	+/+	94%	<b>Rattus norvegicus</b> DnaJ (Hsp40) homolog, subfamily C, member 14 (Dnajc14), Mrna
			4994-5265	+/+	92%	<b>Gallus gallus</b> DnaJ (Hsp40) homolog, subfamily C, member 14 (DNAJC14), mRNA  <b>Danio rerio</b> hypothetical LOC555824 (LOC555824), Mrna

			5060-5253	+/+	78%	Drosophila melanogaster CG14650 CG14650-RA, transcript variant A (CG14650), mRNA
			4994-5256	+/+	64%	Drosophila pseudoobscura GA13147-PA (Dpse\GA13147) mRNA, partial cds
			5096-5199	+/+	74%	
			5088-5190	+/+	71%	
	<b>Reference genomic sequence</b>	YES	5021-5183	+/-	96%	Homo sapiens chromosome 12, alternate assembly (based on HuRef), whole genome shotgun sequence
			5099-5199	+/+	74%	Drosophila melanogaster chromosome 3R >gb AE014297.2  ,complete sequence
			5108-5190	+/-	73%	>gb CM000070.2  <b>Drosophila pseudoobscura</b> strain MV2-25 chromosome 2, whole genome

			10072-10110	+/+	84%	shotgun sequence  >gb CP000393.1  <b>Trichodesmium erythraeum</b> IMS101, complete genome
<b>Expressed sequence tags</b>	YES	4994-5265	+/+	99%	<b>Bos taurus</b> cDNA clone B4772 5', mRNA sequence	
		4994-5265	+/+	99%	normal cattle brain <b>Bos taurus</b> cDNA clone RZPDp1056N207 Q 5', mRNA sequence	
		4994-5265	+/-	98%	<b>Bos taurus kidney fetus</b> <b>Bos taurus</b> cDNA clone E1KI012H11 3', mRNA sequence	
		4994-5262	+/+	98%	<b>Ovis aries</b> cDNA, mRNA sequence	
		5000-5265	+/-	98%	<b>Bos taurus</b> cDNA 3', RNA sequence Full Length cDNA from the Mammalian Gene Collection Homo sapiens cDNA 5' similar to	

						BC016941, mRNA sequence
						Bos taurus cDNA clone 9BOV32_D13 5', mRNA sequence
				+/+	98%	Homo sapiens cDNA 5' similar to BC016941, mRNA sequence
				+/+	95%	Homo sapiens cDNA 5', mRNA sequence
				+/+	95%	Homo sapiens cDNA clone IMAGE:6729715 5', mRNA sequence
				+/+	95%	Homo sapiens cDNA clone IMAGE:6568891 5', mRNA sequence
				+/+	95%	Pongo pygmaeus cDNA clone DKFZp469B242 5', mRNA sequence
				+/+	95%	DG11-kidney Canis lupus familiaris cDNA 3', mRNA sequence
				+/+	94%	Sus scrofa cDNA 3', mRNA sequence
						fully-grown mouse oocyte Mus musculus cDNA clone

				+/-	94%	DS172D_G01 5', mRNA sequence
				+/-	94%	Hematopoietic stem cells Mus musculus cDNA, mRNA sequence
				+/+	94%	
				+/+	94%	
	<b>Non human Non mouse ESTs</b>	YES				<p>Bos taurus cDNA clone B4772 5', mRNA sequence</p> <p>normal cattle brain Bos taurus cDNA clone RZPDp1056N207 Q 5', mRNA sequence</p> <p>Bos taurus kidney fetus Bos taurus cDNA clone E1KI012H11 3', mRNA sequence</p> <p>Ovis aries cDNA, mRNA sequence</p>



						Pongo pygmaeus cDNA clone DKFZp469B242 5', mRNA sequence
--	--	--	--	--	--	---

**Table VII**

The following results were obtained upon sequence comparison of HCV (accession id: AY0512992) with human ESTs in TIGR:

Query Seq	LIST OF ALL HUMAN ESTs MATCHING WITH THE QUERY	REGION OF MATCH ON QUERY SEQ	ANNOTATION OF EST	VEC SCREEN RESULT	REMARK
AY051292	AV755731	1542-1601	Homosapiens cDNAclone,mRNAsequence	-	No contamination
	AV758366	1150-1596	Homosapiens cDNAclone BMFAKA 03 ,mRNAsequence	-	No contamination
	R28798	1566-1669	22 week old human fetal liver,mRNAsequence	-	No contamination

**Table VIII**

The following results were obtained upon sequence comparison of HCV (accession id: AY051292) with various databases in NCBI.

Query Seq	Blasted with/Aligned with	Sequences showing similarity	Position on the viral genome	+/- strand or +/- strand	% similarity	description	Accession id of animal showing similarity
AY051292	Human genomic plus transcript	YES	848-884	+/+	86%	Homo sapiens chromosome 4 genomic contig, alternate assembly.	NW 922217.1
			848-884	+/-	86%	Homo sapiens chromosome 4 genomic contig, reference assembly.	NT 022792.17

	<b>Mouse genomic plus transcript</b>	NO				No significant similarity found.	
	<b>Nucleotide Collection</b>	YES				with Hepatitis C virus	
	<b>Refrence mRNAseq</b>	NO				No significant similarity was found.	
	<b>Expressed sequence tags</b>	YES				Homo sapiens CDNA, mRNA sequence	CV350260.1
	<b>Non human Non mouse ESTs</b>	NO				No significant similarity found.	
	<b>Genomic Survey Seq</b>	NO				No significant similarity found.	
	<b>High Throughput genomic seq</b>	YES				Homo sapiens chromosome 4, working draft sequence.	AC114911.1
	<b>Patent seq</b>	YES				With various fragments/sequences of HCV.	
	<b>Protein Databank</b>	NO				No significant similarity found.	
	<b>Human ALU repeat elements</b>	NO				No significant similarity found.	
	<b>Sequence Tagged sites</b>	NO				No significant similarity found.	

	<b>Whole-Genome shotgun reads</b>	NO				No significant similarity found.	
	<b>Environmental samples</b>	NO				No significant similarity found.	

## **CONCLUSION:**

### **1. Results obtained with Chikungunya Virus**

Similarities were found with mouse/mosquito ESTs and genome. As well as similarities with mouse neurosphere normalized CDNA libraries.

### **2. Results obtained with Influenza A virus:**

Similarities with human working draft sequences (chromosome 7)/Environmental samples, human and ESTs.

### **3. Results obtained with Japanese Encephalitis Virus:**

No significant similarity was found.

### **4. Results obtained with Bovine Diarrhea Virus:**

BVDV showed sequence similarity with number of human and animal ESTs; incidentally all these ESTs had the annotation DnaJ (Hsp 40) homolog subfamily C, member 14.

They also showed similarity with ESTs whose annotation include

1. Dopamine receptor interacting protein (chromosome 12)
2. Leucine zipper protein 2(chromosome 11)

### **5. Results obtained with Hepatitis C virus:**

1. It showed sequence similarity with human ESTs over a span of greater than 100 nucleotides with more than one EST.
2. Incidentally annotation of theses ESTs is not yet completed; these ESTs have been listed as singleton.

## BIBLIOGRAPHY

- Altschul,S.F,Gish,W.,Miller,W.,Myers,E. W.&Lipman,D.J(1990)  
“Basic local alignment search tool.” J.Mol.215:403-410

Program used BLAST

- Altschul,Stephen F.,Thomas L.Madden,Alejandro A.Schaffer,Jinghui Zhang, Zheng, Zhang, Webb Miller, and David J. Lipman (1997),”Gapped BLAST and PSI-BLAST: a new generation of Protein database search programs””, Nucleic Acid Res.25:3389-3402

Program used: BLASTN 2.2.14

- Altschul,Stephen F.,Thomas L.Madden,Alejandro A.Schaffer,Jinghui Zhang, Zheng, Zhang, Webb Miller, and David J. Lipman (1997),”Gapped BLAST and PSI-BLAST: a new generation of Protein database search programs””, Nucleic Acid Res.25:3389-3402

- [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)
- [www.tigr.org](http://www.tigr.org)
- [www.who.int](http://www.who.int)
- Encyclopedia of Virology
- Springer Index Of Viruses
- Nature Encyclopedia of human genome(Volume-5)
- Encyclopedia of molecular medical biology