Metagenomic sequencing for pathogen identification and pathogen discovery

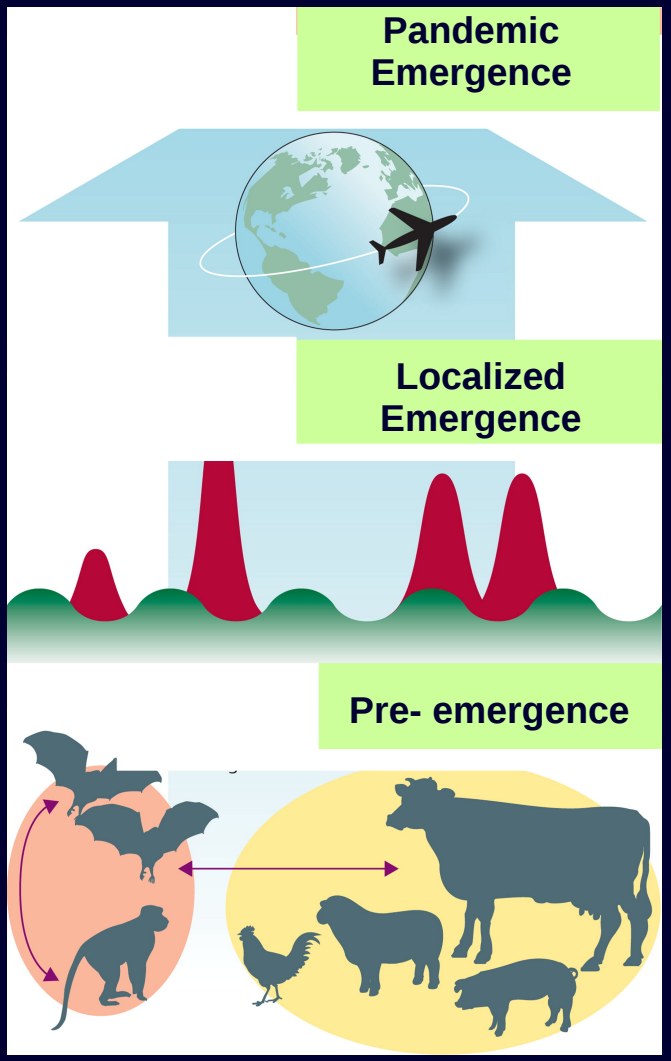
Indian Veterinary 2015

Chitra Pattabiraman

National Centre for Biological Sciences, TIFR, Bangalore

October 27th, 2015

Emergence of novel zoonotic pathogens



SARS, H1N1

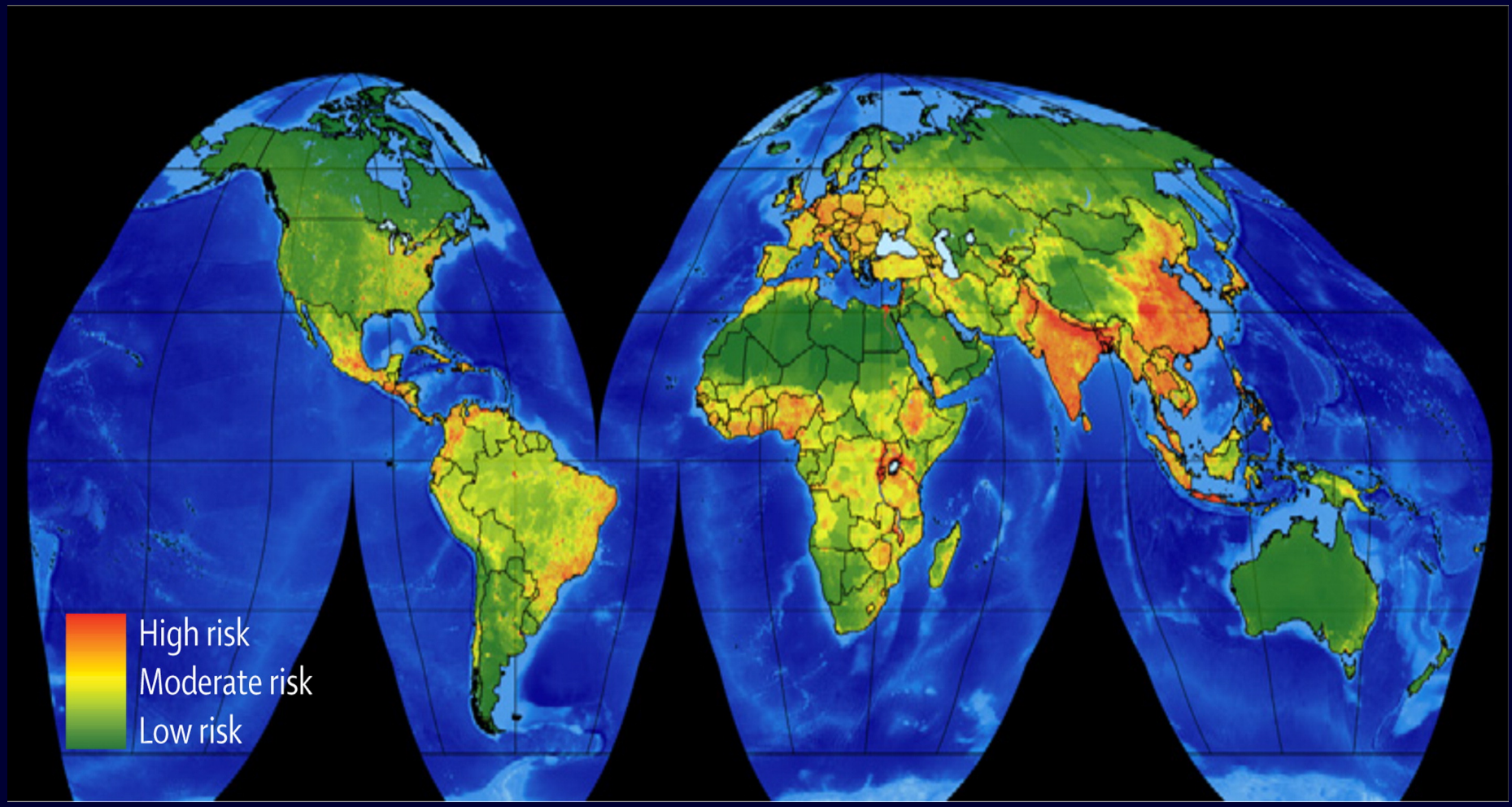
MERS, Ebola (?)

H5N1 (bird flu), Monkeypox

Morse et al., 2012

The Lancet 2012 380, 1956-1965DOI: (10.1016/S0140-6736(12)61684-5)

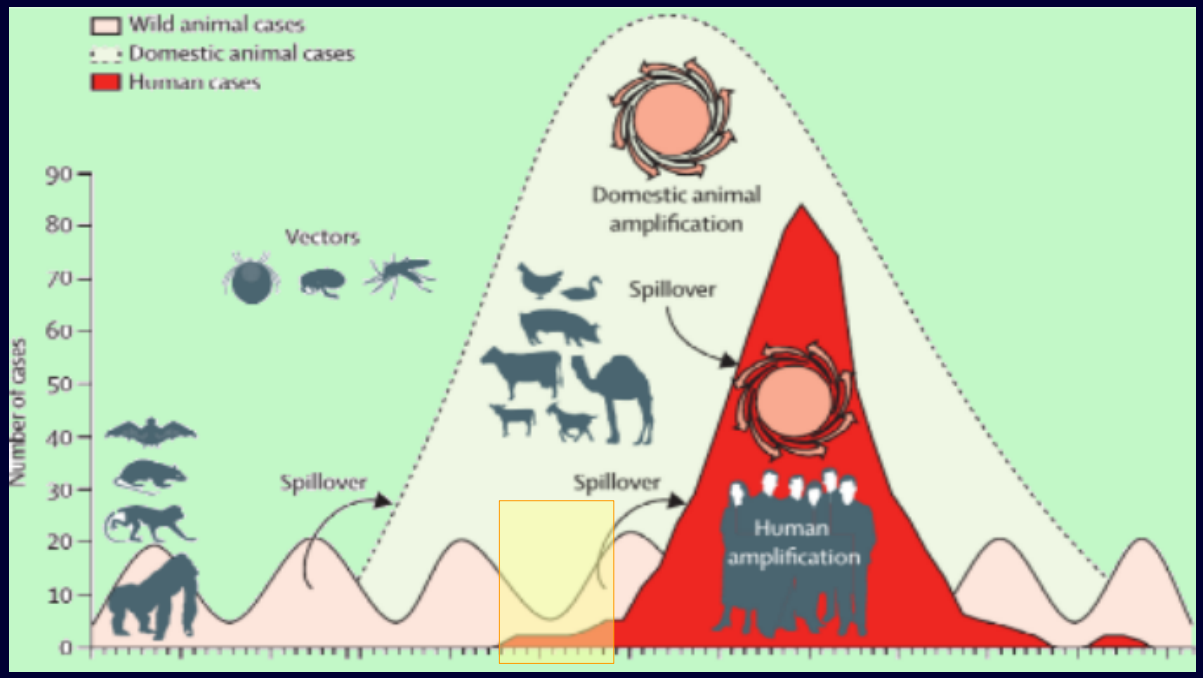
Predicted “hotspots” for the emergence of zoonotic diseases



Morse et al., 2012

The Lancet 2012 380, 1956-1965DOI: (10.1016/S0140-6736(12)61684-5)

A window for surveillance

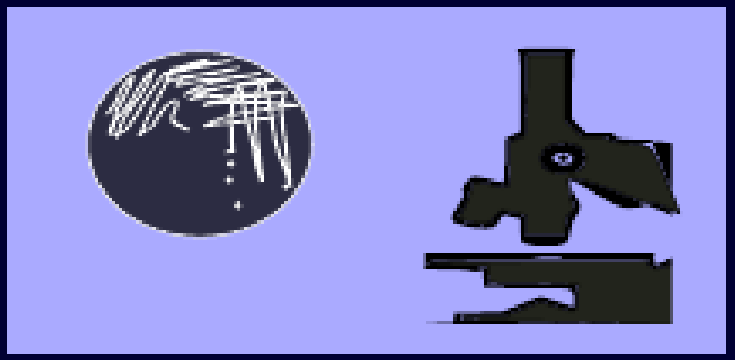


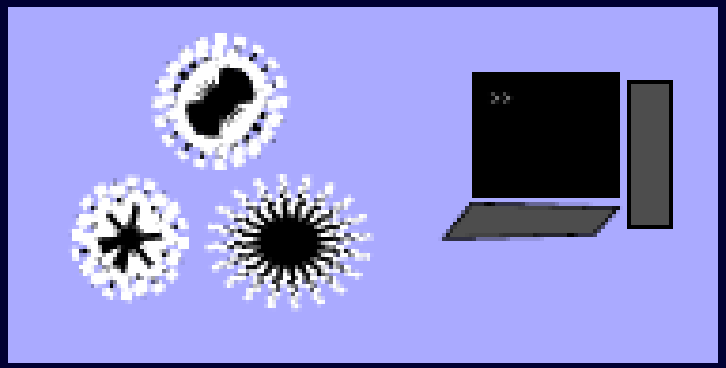
Modified from Karesh WB et al., 2012

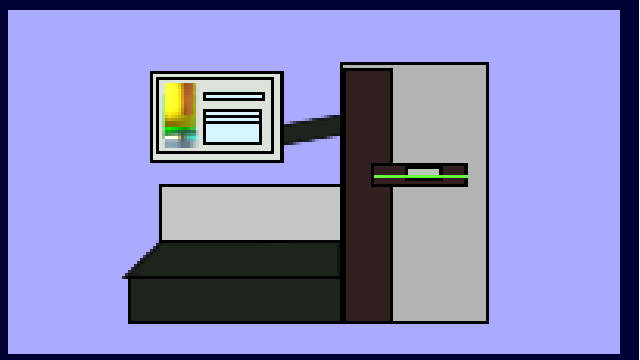
The Lancet 2012 380, 1936-1945DOI: (10.1016/S0140-6736(12)61678-X)

Routes to Pathogen Discovery

Grow it :	**Grow the pathogen from the clinical samples**
Have a look:	**Seeing is believing (Microscopy)**
Read it's genome:	**Use molecular tools from related pathogens**
	Sequence the genome
Read an encyclopedia and find one relevant paragraph:	**Metagenomic Sequencing**



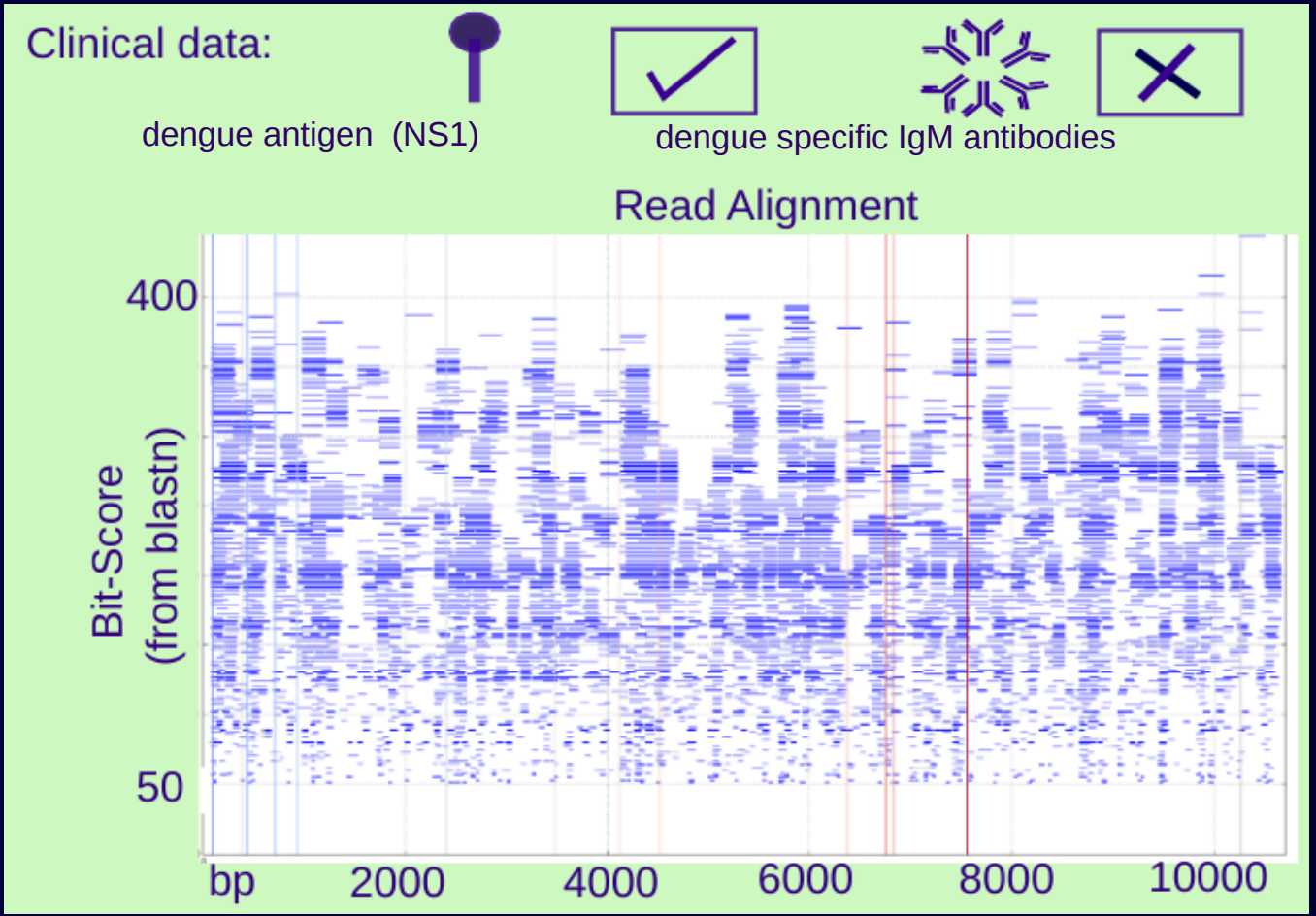




Sequencing from acute fevers

Sample 1

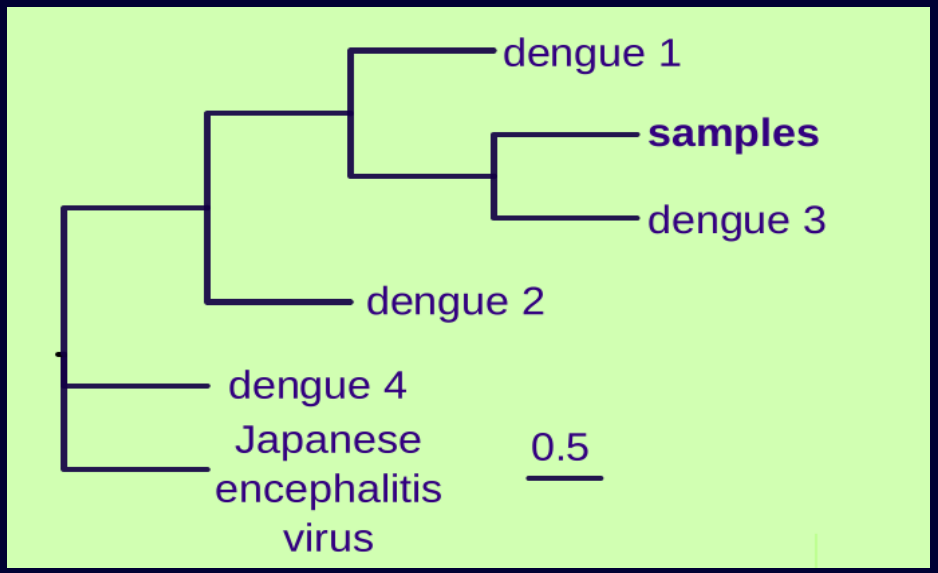
dengue Virus 3 + Japanese encephalitis virus

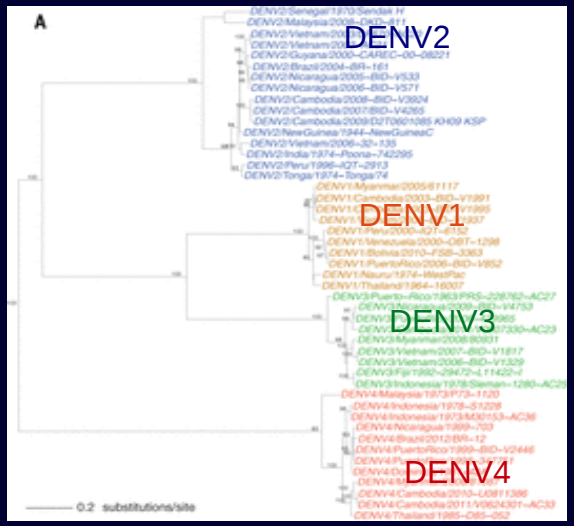


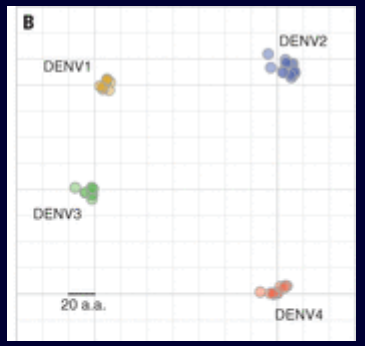
dengue virus 3 genome (10707 bp, 20494 reads)

The complete dengue3 genome

390 Single Nucleotide Polymorphisms, 42 –a silent, 14 – stop codons, 334 – amino acid changes







Katzelnick et al 2015

Sample 2
dengue virus 3

nonstructural protein 3, partial [Dengue virus 3]

Sequence ID: gb|ACJ06059.1| **Length:** 88 **Number of Matches:** 1

▶ See 5 more title(s)

Range 1: 35 to 85 GenPept Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
101 bits(252)	5e-26	Compositional matrix adjust.	48/51(94%)	49/51(96%)	0/51(0%)	+1

```
Query  16   KQNAEPDGPMPELEEEMFKKRNLTIMDLHPGSGKTRKYLPAIVREAIKRRL  168
            + NAEPDGP PELEEEMFKKRNLTIMDLHPGSGKTRKYLPAIVREAIKRRL
Sbjct  35   QTNAEPDGPTPELEEEMFKKRNLTIMDLHPGSGKTRKYLPAIVREAIKRRL  85
```

Sequences from acute and prolonged fevers

Sample 3

Japanese encephalitis virus
Rickettsia sp
Cryptococcus neoformans

Japanese encephalitis virus isolate IND-WB-JE2, complete genome

Sequence ID: **gb|JX072965.1|** Length: 10915 Number of Matches: 1
Range 1: 8599 to 8713

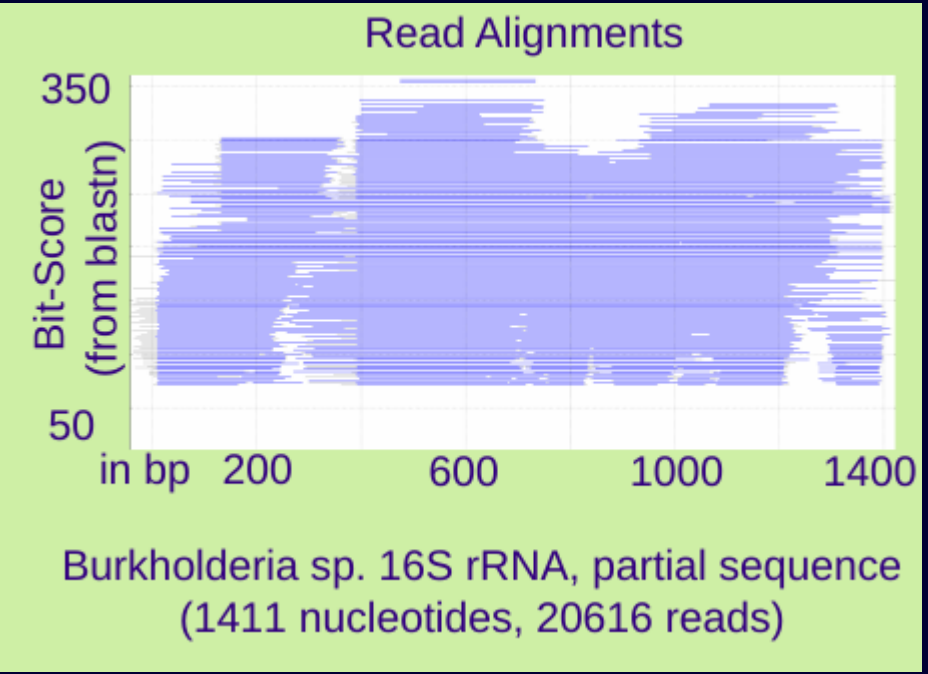
Score	Expect	Identities	Gaps	Strand	Frame
208 bits(230)	1e-50()	115/115(100%)	0/115(0%)	Plus/Plus	

Features:

```
Query  1     ACGGAAGCTATGAAGTGAAGGCTACTGGCTCAGCCAGTTCTCTCGTCAACGGAGTGGTGA  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  8599  ACGGAAGCTATGAAGTGAAGGCTACTGGCTCAGCCAGTTCTCTCGTCAACGGAGTGGTGA  8658
Query  61    AGCTCATGAGCAAACCTTGGGACGCCATTGCCAACGTCACCACCATGGCCATGAC  115
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  8659  AGCTCATGAGCAAACCTTGGGACGCCATTGCCAACGTCACCACCATGGCCATGAC  8713
```

Sample 4

Burkholderia cepacia
Burkholderia multivorans



Acute encephalitis in India

Japanese encephalitis virus
reads=26
Median blast score= 240.26

Cryptococcus neoformans
reads= 38 (large ribosomal subunits)

Cryptococcus neoformans
reads = 984 (large ribosomal subunits)

Alternaria alternata
Many viruses of fungi

Aphid lethal paralysis virus
reads= 78, median blast score=96)
Kadipiro virus
reads=40, blast score= 75.8

Aphid lethal paralysis virus
Reads= 93, blast score = 230

Aphid lethal paralysis virus
reads = 19, blast score = 165.393

Aphid lethal paralysis virus
Reads = 13
median blast score = 159.9

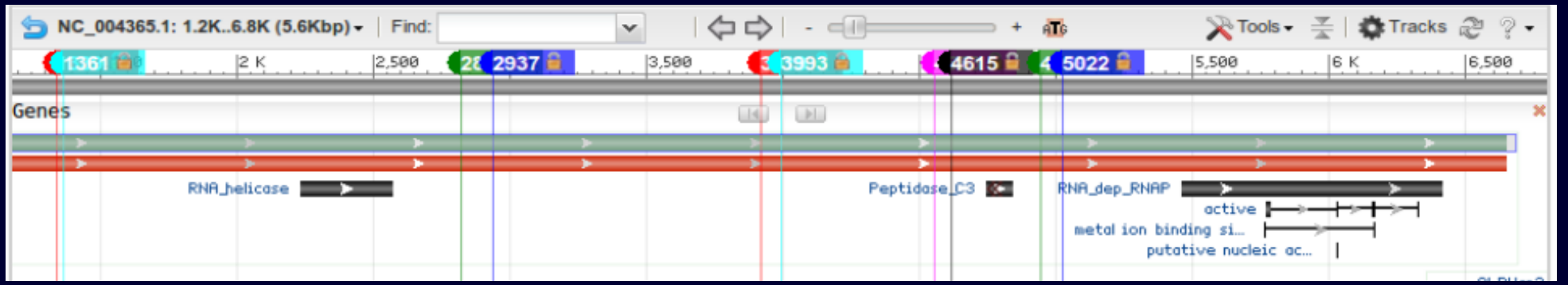
Human Immunodefeciency Virus 1
(reads= 25, median blast score= 100)
Japanese encephalitis virus
(reads = 30, median blast score = 131)
Aphid lethal paralysis virus
(reads = 12, median blast score = 295)
Porcine Reproductive and Respiratory Virus
(reads = 1, 50 nucleotides, blast score = 44.1)

Acute encephalitis in India : sequences from cerebrospinal fluid

Aphid lethal paralysis virus



Viral Zone



ssRNA (+) viruses, (9.8kb)

Dicistroviridae are known to infect invertebrates - bees, aphids etc. and are highly lethal

translated blast results

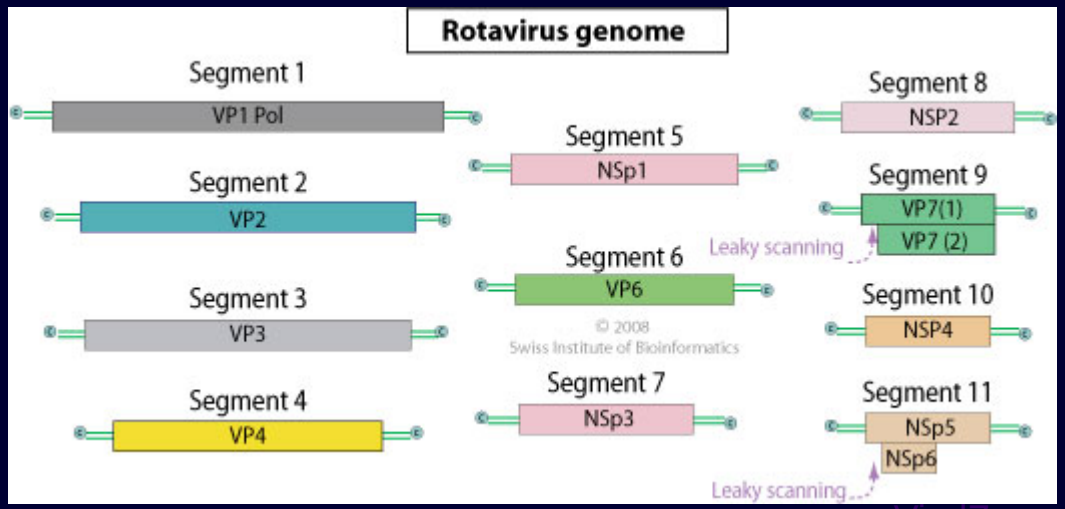
translated
2 RTIETDMTYAQVTEECSMKMASRFNQHMDFTNYLDAYRNP
read
RTIETDMTYAQVTEECSMKMASRFNQHMDFTNYLDAYRNP
ALPV
761 RTIETDMTYAQVTEECSMKMASRFNQHMDFTNYLDAYRNP

Score - 89.0 E-value - 9e-19

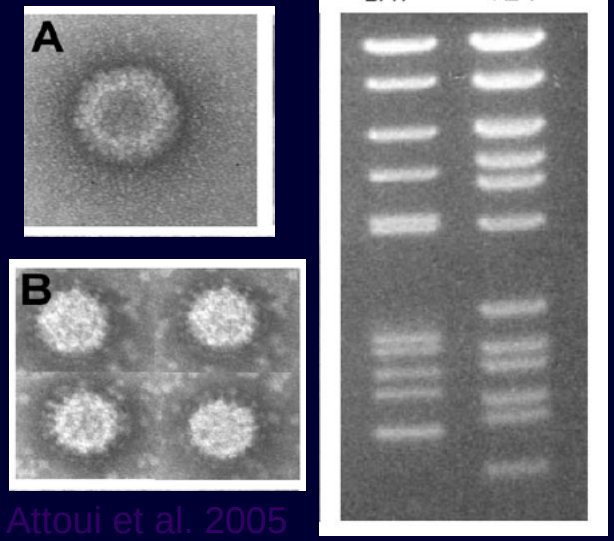
translated
3 EIVPLDVNPDILKENGTFLIYGTEPGRR
read
EIVPLDVNPDILKENGTFLIYGTEPGRR
ALPV
1471 EIVPLDVNPDILKENGTFLIYGTEPGRR

Score - 64.7 E-value - 2e-19

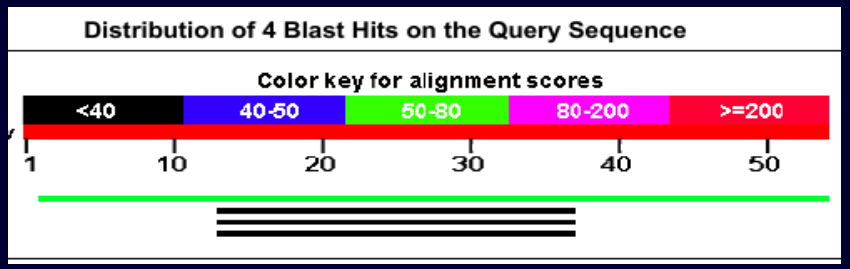
Sequences related to Kadipiro virus (KDV), Reoviridae (dsRNA)



ViralZone



Attoui et al. 2005



Kadipiro virus JKT-7075 segment 10 Vp10 gene, complete cds

Sequence ID: **gb|AF052020.1|** Length: 946 Number of Matches: 1

Range 1: 895 to 946

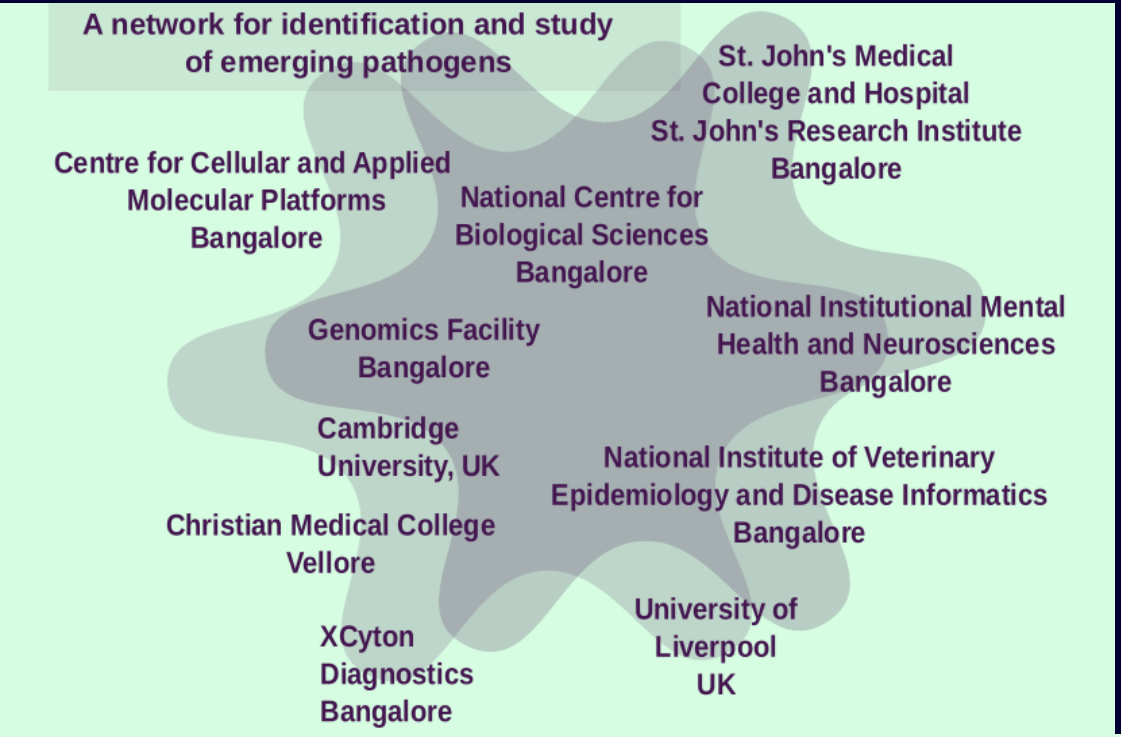
Score	Expect	Identities	Gaps	Strand	Frame
78.8 bits(86)	4e-12()	50/53(94%)	1/53(1%)	Plus/Plus	

Features:

```
Query  2    GGAGCTACCACATTGGCTCGGTATTGGCGCGTTGGCGATTGAATATAACCGAC  54
Sbjct  895  GGATCTACCACATTGGCTCGGTATTGGCGCGTTGGCGACTG-ATATAACCGAC  946
```

Banna Virus: New orbiviruses isolated from patients with unknown fever and encephalitis in Yunnan province. Chin J Virol 6, 27–33. Xu, P., Wang, Y., Zuo, J., Lin, J. & Xu, P. (1990).

Acknowledgements



Dr. Sudhir Krishna

Clinical Collaborators

Dr. Mary Dias, Dr. Anita Shet (St. John's Medical College and Hospital, Bangalore), Dr. V. Ravi (National Institute of Mental Health and Neurosciences, Bangalore), Dr. Asha Abraham, Dr. Shobha Mammen (Christian Medical College, Vellore)

Collaborations:

Dr. Derek Smith, Barbara Muehlemann, Dr. Terry Jones (University of Cambridge), Dr. Sunil Govekar, Dr. Ravikumar BV (XCyton Diagnostics), Dr. Paresh Patel, Dr. Taslimarif Saiyed, Shilpa S, Goutham Atla, Dr. Malali Gowda, (C-CAMP), Dr. Shahid Jameel (India Alliance), Dr. Tom Solomon (University of Liverpool)

Discussions and comments: Dr.Balamurugan, Dr. Rahman (NIVEDI), Dr. Bas Dutilh, Dr. Varadharajan Sundaramurthy, Dr. Aswin Seshasayee,Dr. Sowdhamini, Dr. Ian Lipkin, Dr. Gagandeep Kang, Dr. Ramaswamy, Dr. Ron Fouchier, Dr. Ab Osterhaus, Dr. Greg Winter, Dr. Paul Li, Dr. Xavier de Lamballerie, Dr. Marc Lecuit, Dr. Reety Arora, Shlesha Richhariya, Hardik Gala.

This work was funded by DBT, DBT-RA Postdoctoral Fellowship amd NCBS