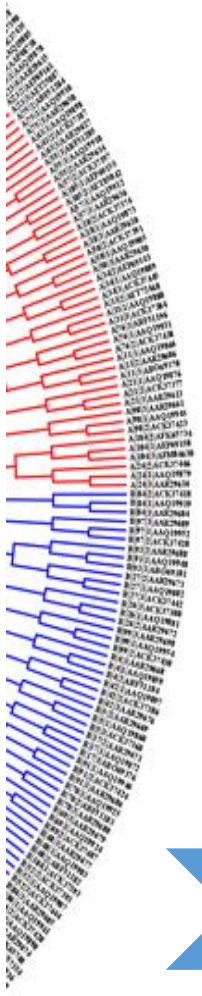
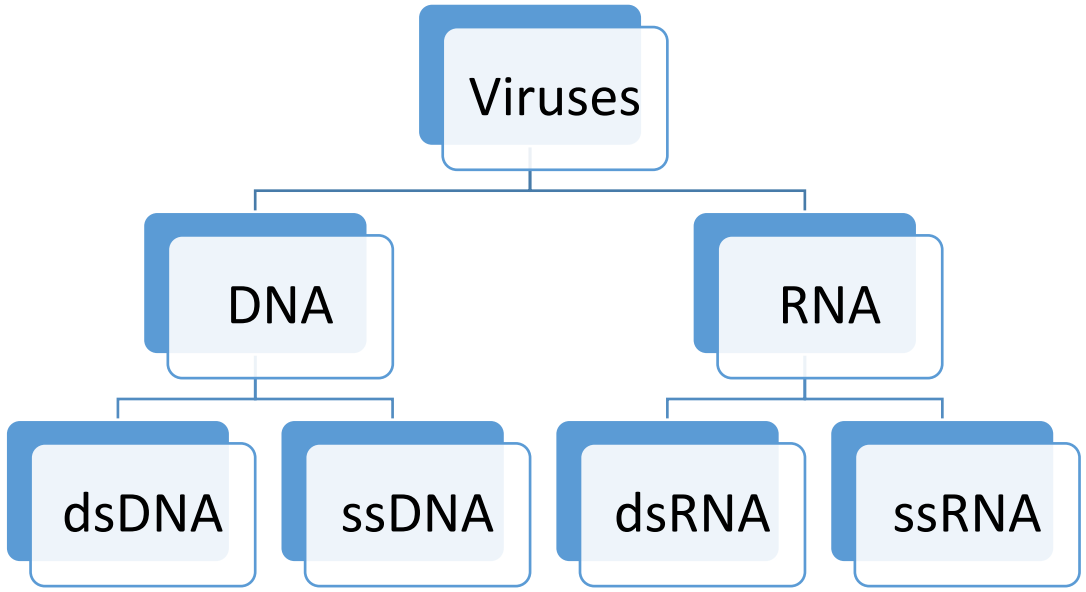
Computational Methods for Typing Viruses

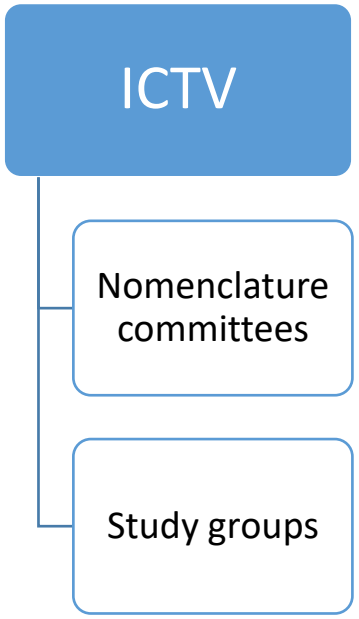
Dr. Urmila Kulkarni-Kale

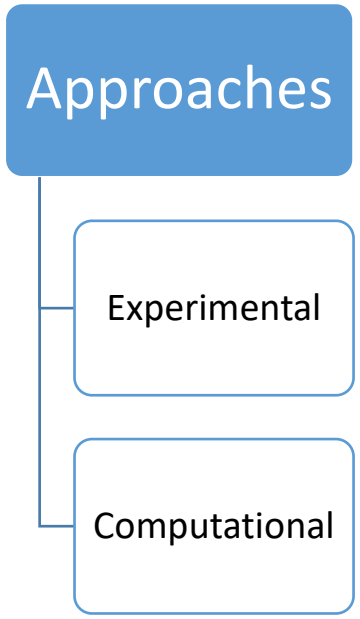
S. P. Pune University| University of Southeastern Norway | Citadel Precision Medicine
urmila.Kulkarni.kale@gmail.com

Virus Taxonomy & Nomenclature



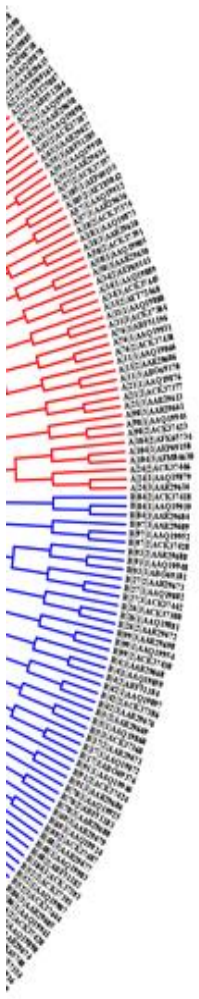






© Dr. Urmila Kulkarni-Kale | SPPU | USN | CPM |

Viral typing: the starting point for characterization



Typing approaches

- Experimental
- Computational

Computational approaches

- Alignment-based
- Alignment-free

Subject to manual interpretation

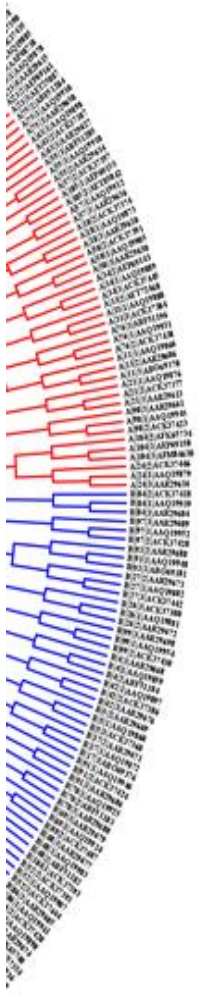
Database searches → MSA → Phylogeny → Assignment of type

Knowledge bases → K-mer statistics → Distance computation → Assignment of type

Assigned based on statistical significance

© Dr. Urmila Kulkarni-Kale | SPPU | USN | CPM |

Virus Typing based on Molecular Phylogeny Analysis : the workflow

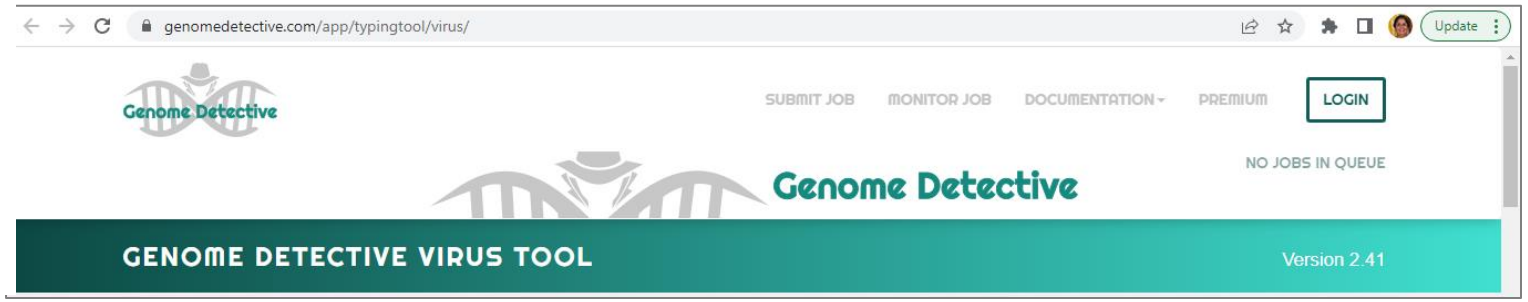


- Define the objective
- Curate a set of reference sequences (known types)
- Carry out Multiple Sequence Alignments
- Undertake phylogenetic analysis:
  - distance-based (NJ: Neighbor-Joining)
  - character-based (MP: Maximum parsimony)
  - Bayesian-based (ML: Maximum likelihood)
- Generate datasets for bootstrapping
- Use clustering method
- Generate consensus tree
- Assess tree topology(ies)
- Analysis of inferred tree(s)
- Assign type to the viral isolate based on proximity to known ones

© Dr. Urmila Kulkarni-Kale | SPPU | USN | CPM |

Genome Detective Virus Tool
http://www.genomedetective.com/app/typingtool/virus/

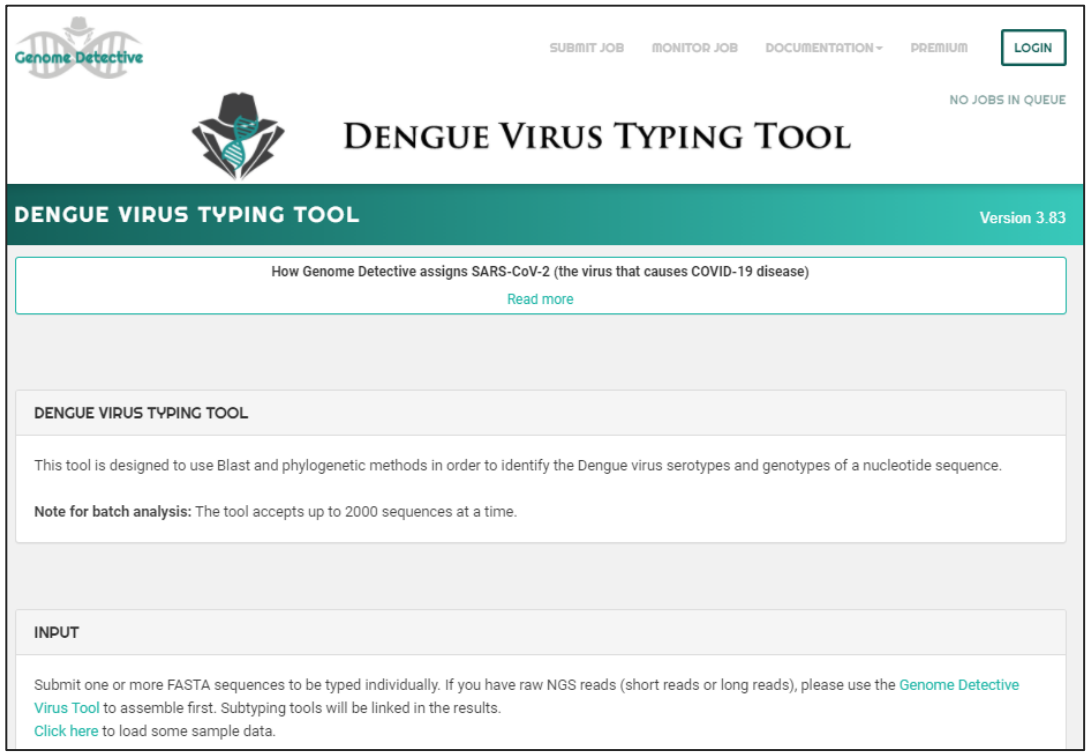
- Assigns taxonomic rank at species level to sequences of eukaryotic viruses & phages
- Ranking is subject to availability of reference genomes in RefSeq
- Currently 11140 distinct taxonomic names are assigned based on 14500 reference sequences
- Subtyping tools, for the identification of subspecies, are available for 19 viruses
- Input: NGS short reads (FASTQ); Contigs; Consensus (FASTA)

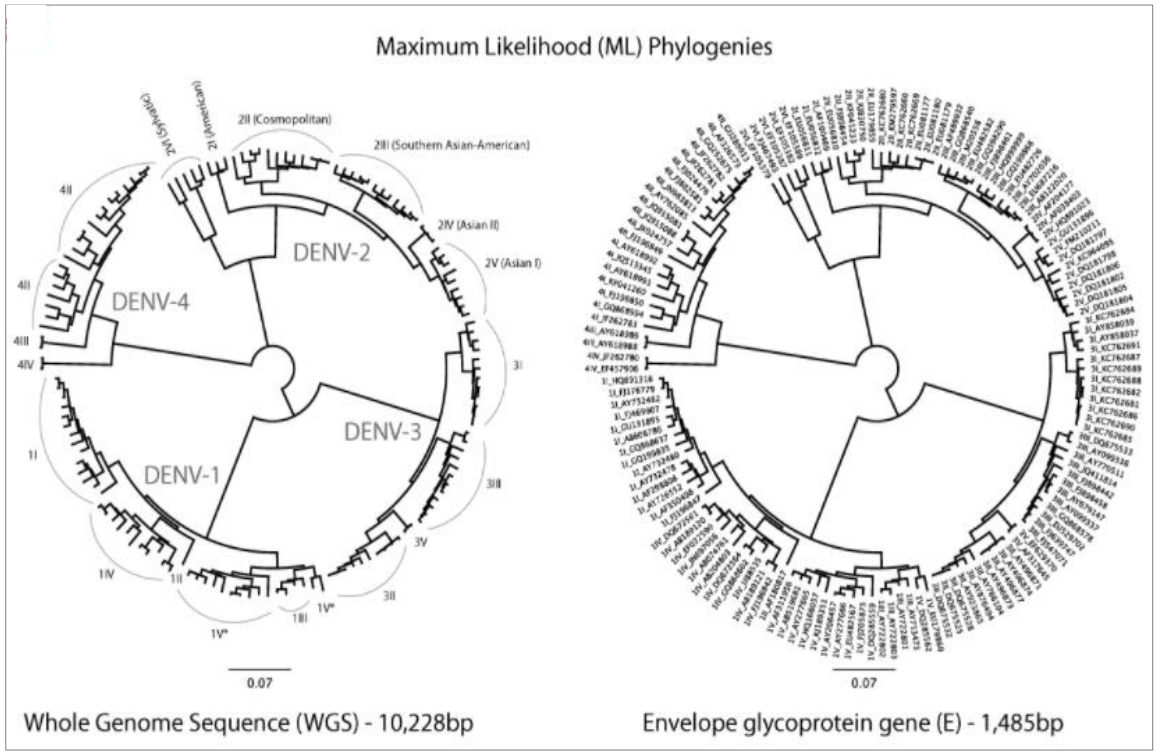


- Chikungunya Typing Tool
- Coronavirus Typing Tool
- Dengue Typing Tool
- Enterovirus typing tool
- HCV Recombinant Typing Tool
- HHV82 Virus Typing Tool
- HTLV-1 and 2 Typing Tool
- Hepatitis A Virus Genotyping Tool
- Hepatitis B Virus
- Hepatitis E Virus Genotyping Tool
- Human Immunodeficiency Virus Typing Tool
- Measles Virus TypingTool
- MonkeyPox Virus
- Norovirus Typing tool
- Panviral Tool
- Rift Valley Fever Virus
- West Nile Virus
- Yellow Fever Virus Typing Tool
- Zika Virus Typing Tool

© Dr. Urmila Kulkarni-Kale | SPPU | USN | CPM |

Dengue Virus Typing Tool@ Genome Detective





https://www.genomedetective.com/app/typingtool/dengue/

Group leader: Dr. Tulio de Oliveira, South Africa

© Dr. Urmila Kulkarni-Kale | SPPU | USN | CPM |

Genome Detective: SARS-CoV-2 Typing Tool

https://www.genomedetective.com/app/typingtool/cov/

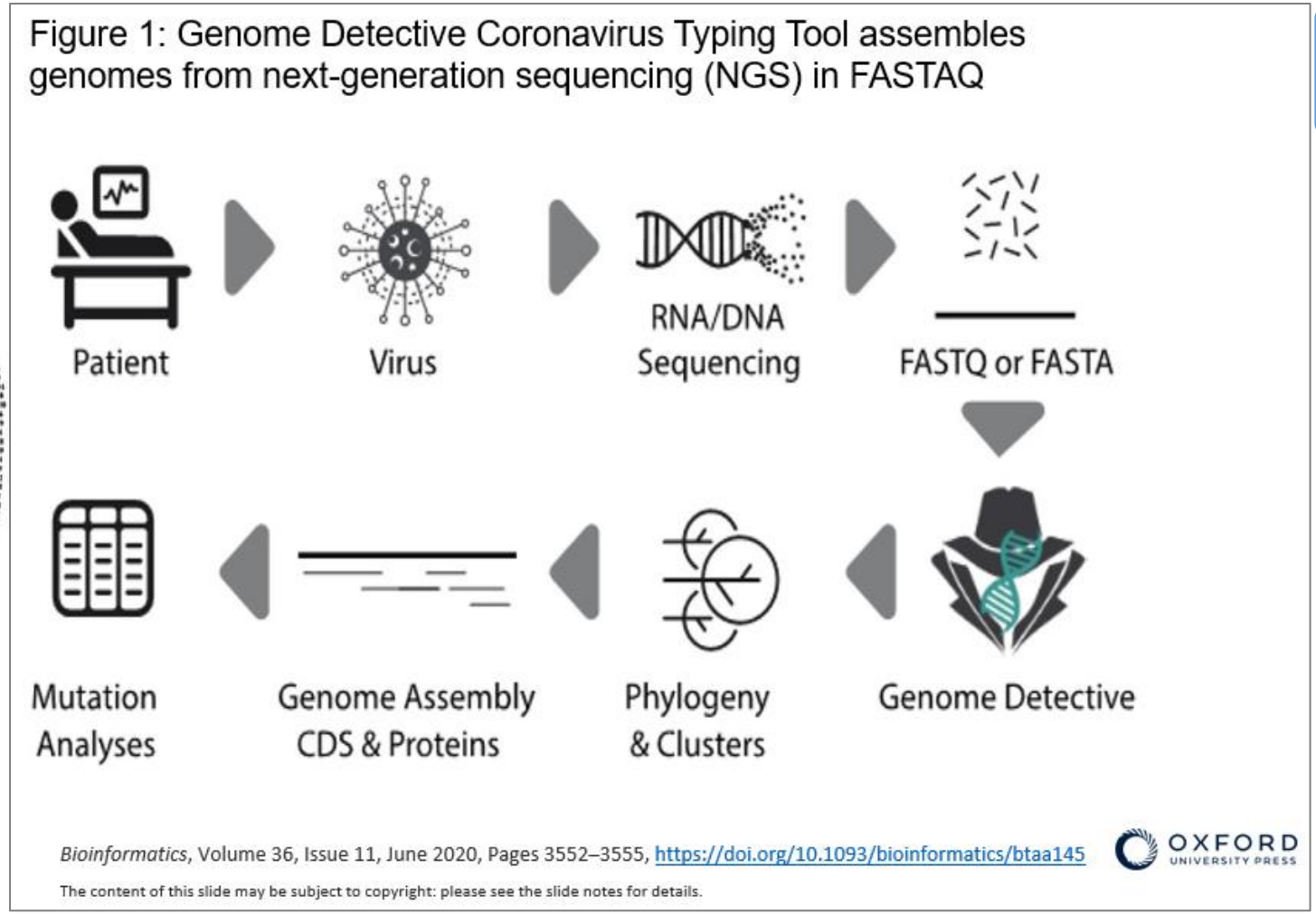


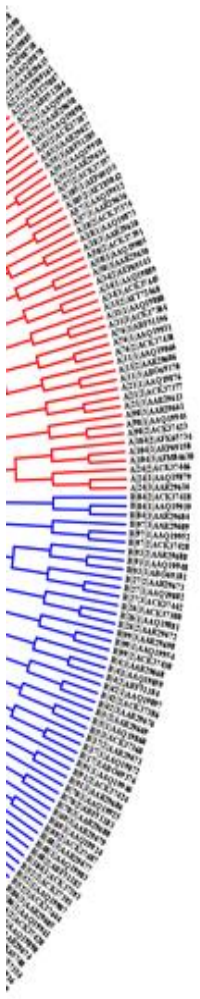
Figure 1: Genome Detective Coronavirus Typing Tool assembles genomes from next-generation sequencing (NGS) in FASTAQ

Bioinformatics, Volume 36, Issue 11, June 2020, Pages 3552–3555, https://doi.org/10.1093/bioinformatics/btaa145

The content of this slide may be subject to copyright: please see the slide notes for details.

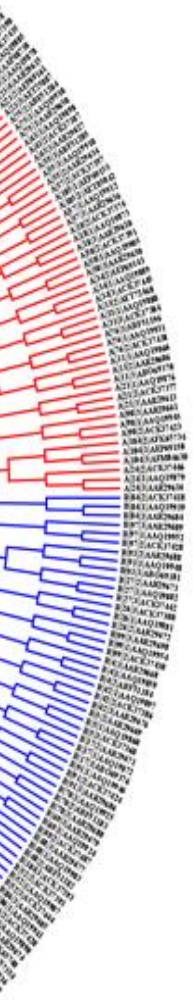
Workflow

- Reference dataset: 431 WGS – aligned using MUSCLE codon alignment, manual edits for CDS boundaries
- Input: NGS data (1Gb) or 2000 assembled genomes;
  Formats: FSATQ/FASTA
- Processes: Species identification, phylogeny construction (ML – PhyML & Bayesian – MrBayes), cluster identification, ADA alignment to identify novel mutations

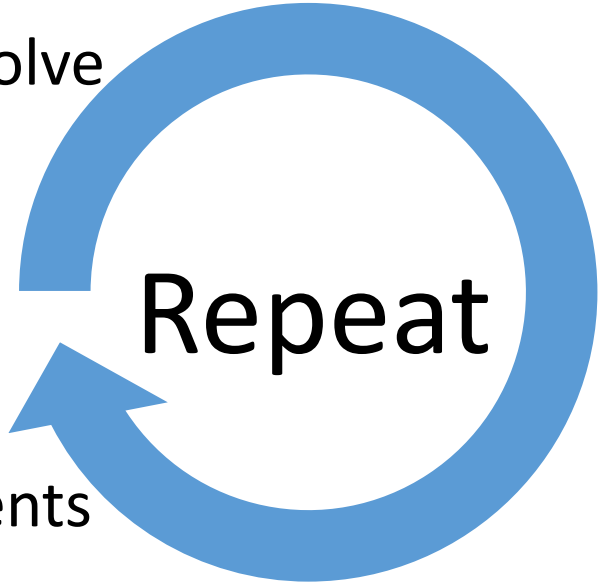


Alignment-free phylogeny and Viral typing: Algorithm & server developments @ SPPU

Molecular Phylogeny Analysis (MPA): Caveats

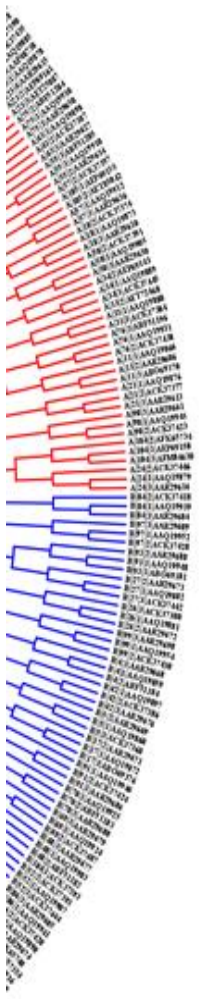


- Heuristic appeal in sequence alignment algorithm
- Positions of IN-DELs in MSA impact model of evolution
  - Errors in alignment increases as sequence similarity decreases
- Assumption of character-based methods: sites evolve independently
- Different methods result into different trees
  - Becomes a matter of interpretation
- Size of data in post-genomic era
- Computational complexity and memory requirements
- Time requirements



© Dr. Urmila Kulkarni-Kale | SPPU | USN | CPM |

Designed Novel Approach: Alignment-free

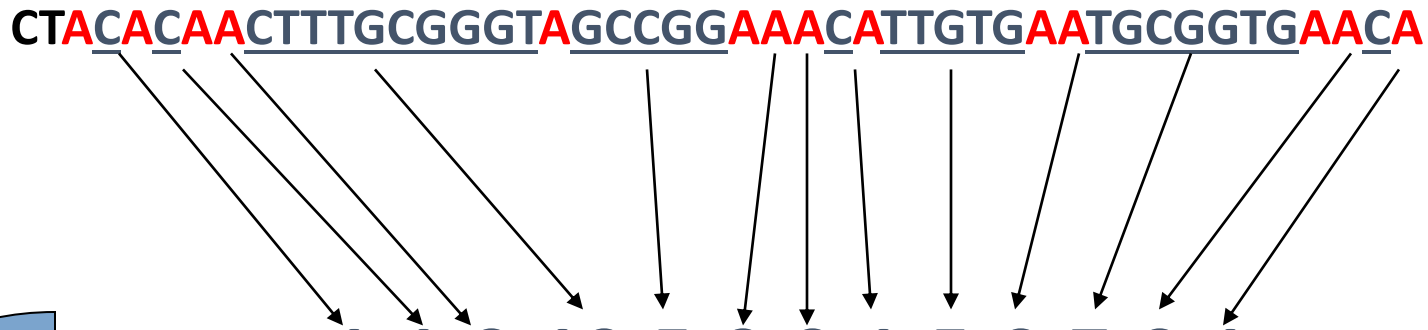


Based on the concept of Return Time Distribution (RTD)

- Accounts for frequency of K-mers and their return times
- Performs equally well in varying levels of sequence similarity
- Fast & accurate with high sensitivity & specificity

Computing RTD for 'A' (k=1)

Return Time (RT): Time required for the reappearance of particular state without its appearance in between.



RTD for 'A' in above sample sequence

Return time for A (X)	Frequency (F)
0	5
1	4
5	2
7	1
10	1

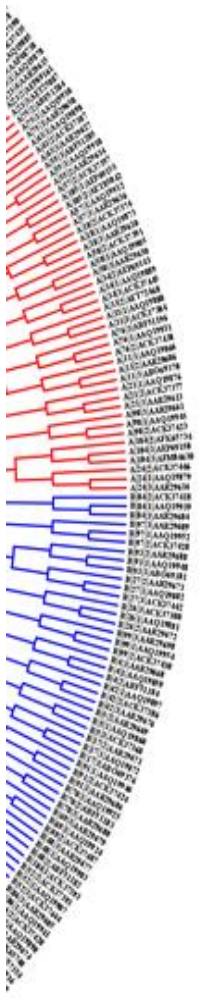
Parameters of RTD for 'A'

$\mu(A) = 2.38$ and $\sigma(A) = 3.27$

Similarly, we can compute μ and σ of RTDs of T, G and C, for k=1.

© Dr. Urmila Kulkarni-Kale | SPPU | USN | CPM |

Alignment-free phylogeny based-on Return Time Distribution



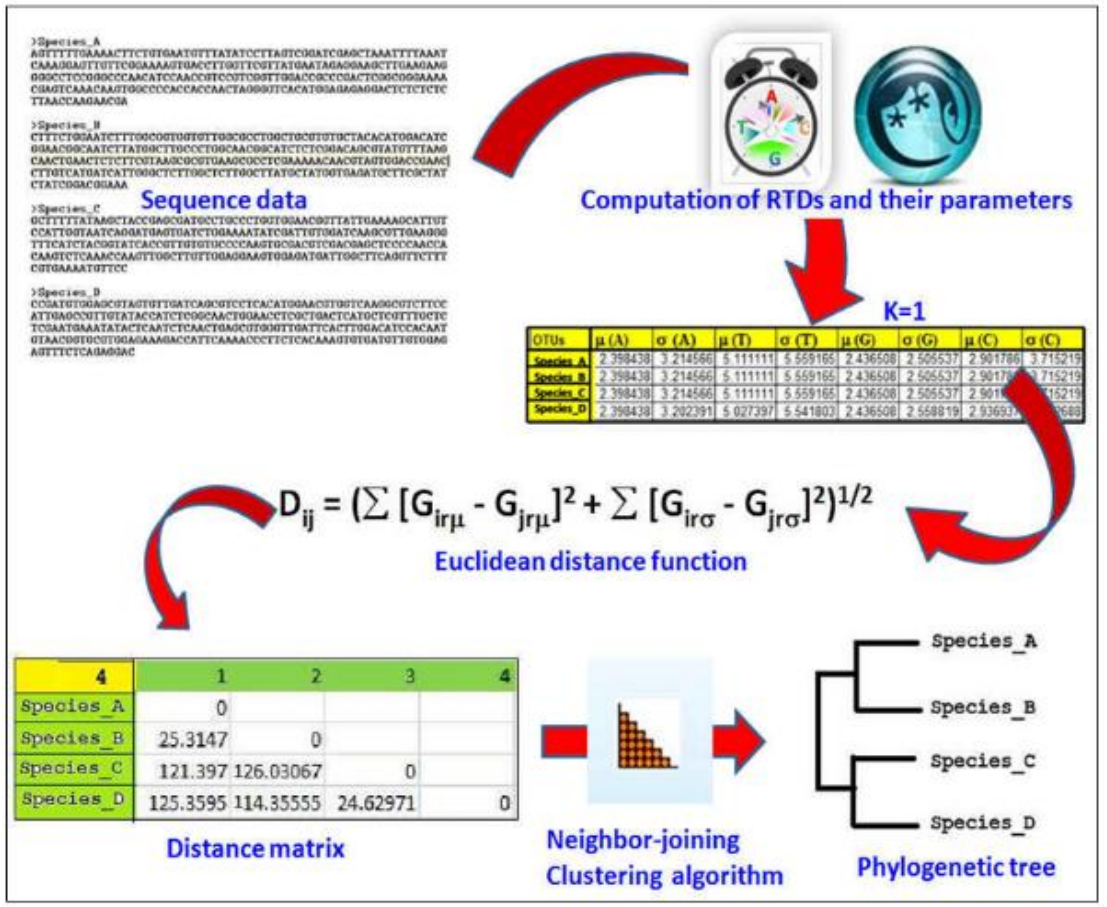
RT→RTD →Distance matrix

The frequency distribution of all such observed RT is termed as RTD of that nucleotide

Numeric vector of size $2*4^k$ comprising of μ and σ of 4^k possible RTDs

Read sequences → Derive RT & RTD at given value of k → Derive parameters of RTD: μ & σ → Compute distance matrix → Derive NJ tree → View tree & analyse tree topology

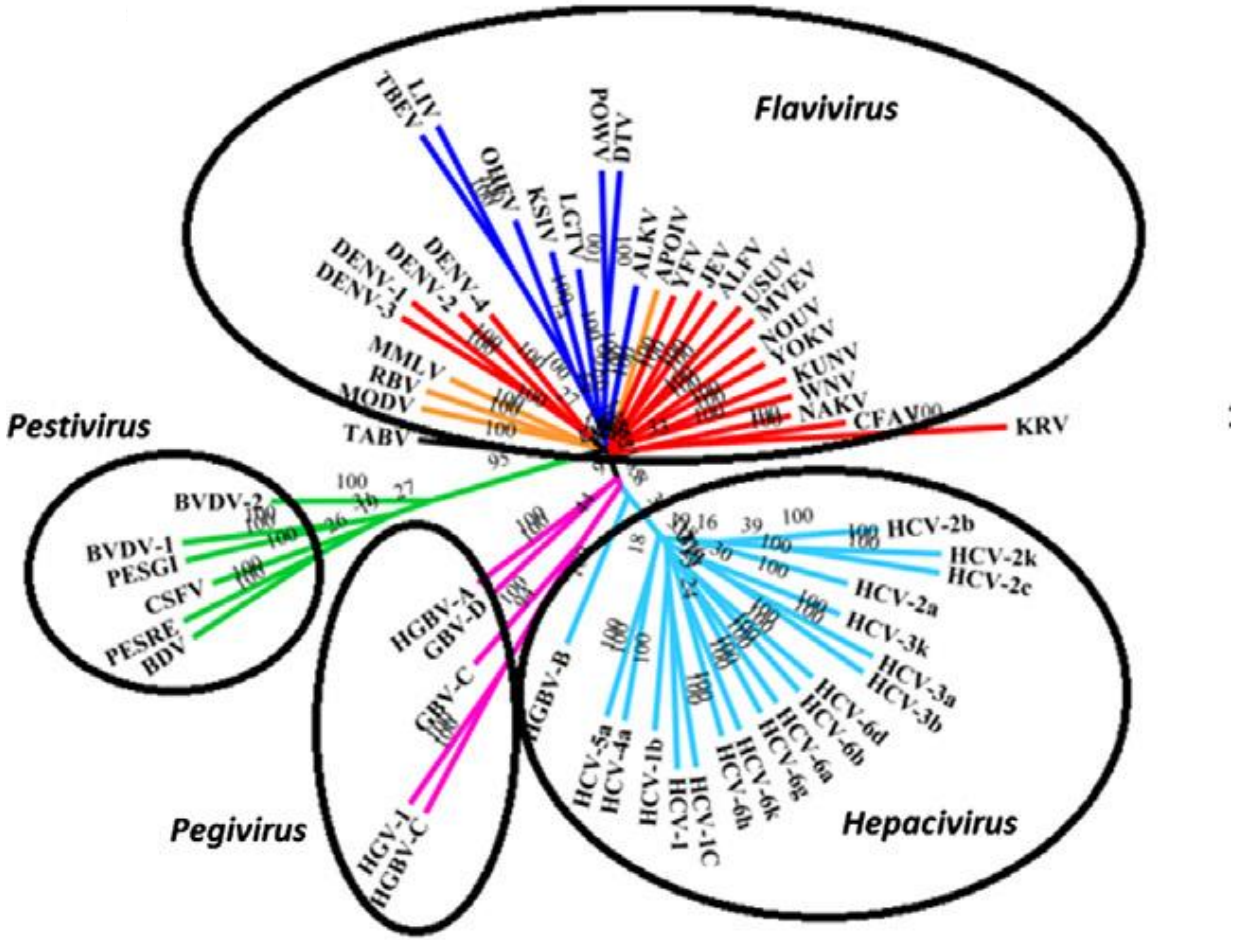
$$D_{ij} = (\sum [G_{ir\mu} - G_{jr\mu}]^2 + \sum [G_{ir\sigma} - G_{jr\sigma}]^2)^{1/2}$$



© Dr. Urmila Kulkarni-Kale | SPPU | USN | CPM |

Clustering & Phylogeny of members of family :*Flaviviridae*

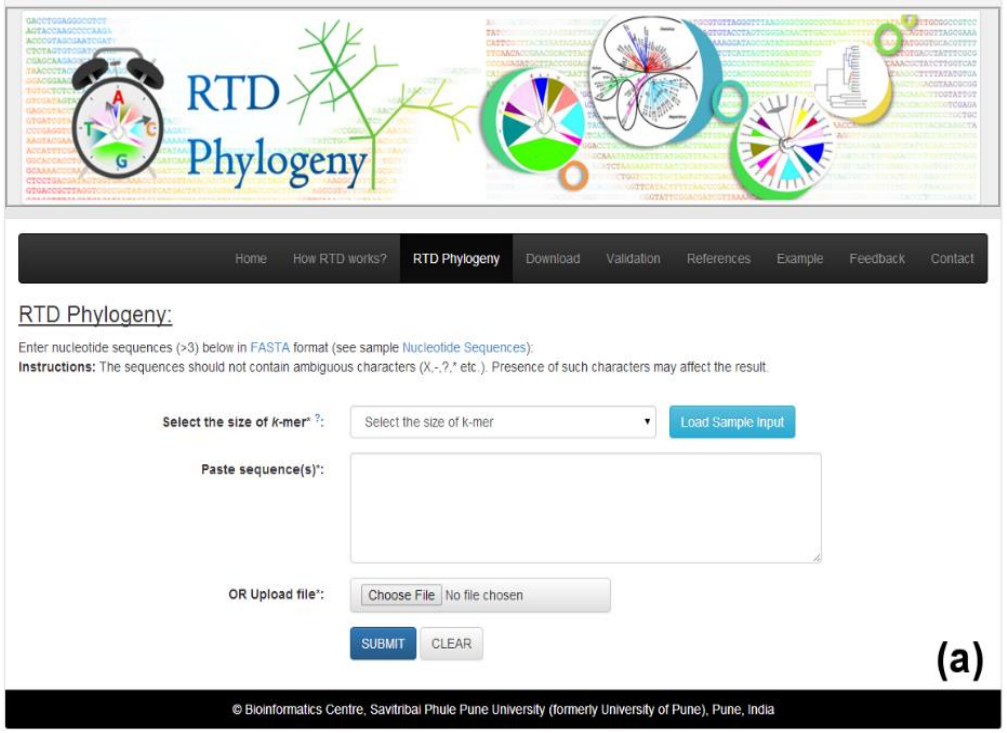
Sequence similarity: 40-86%

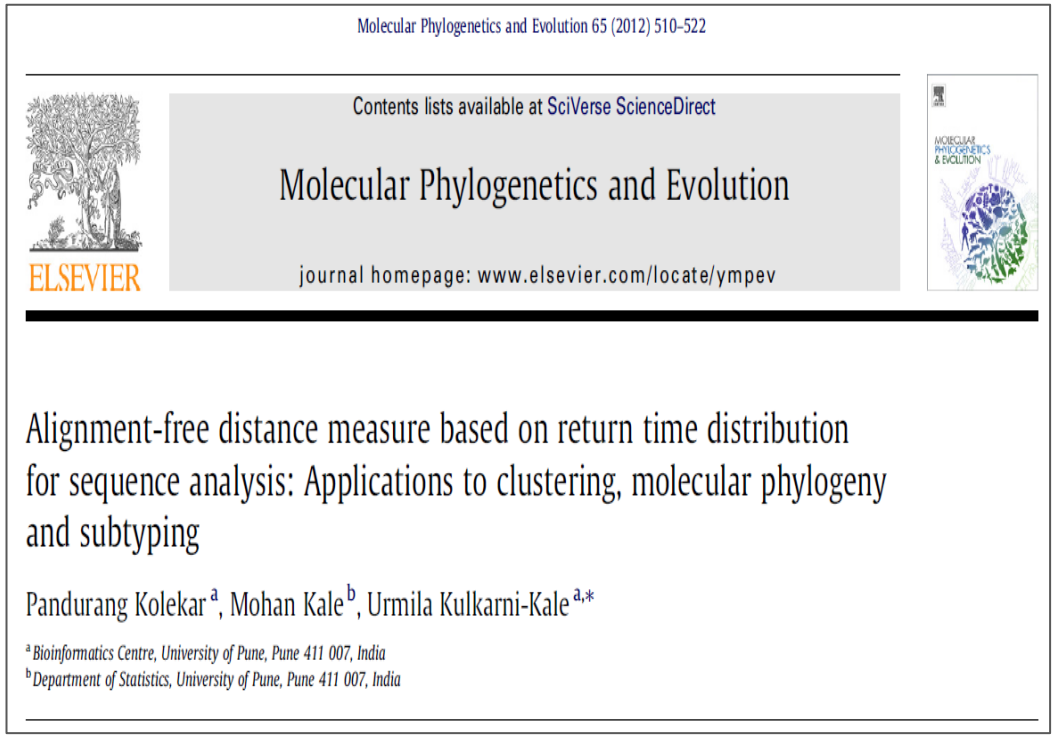


© Dr. Urmila Kulkarni-Kale | SPPU | USN | CPM |

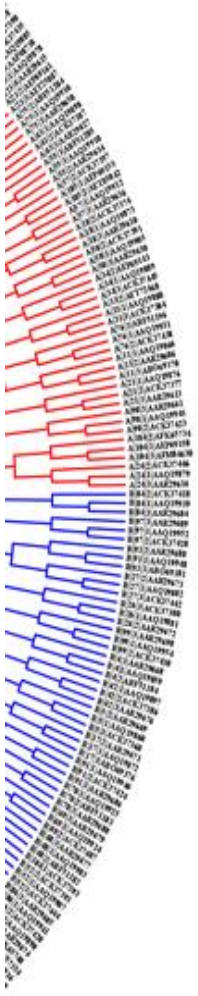
RTD Phylogeny: alignment-free phylogeny server

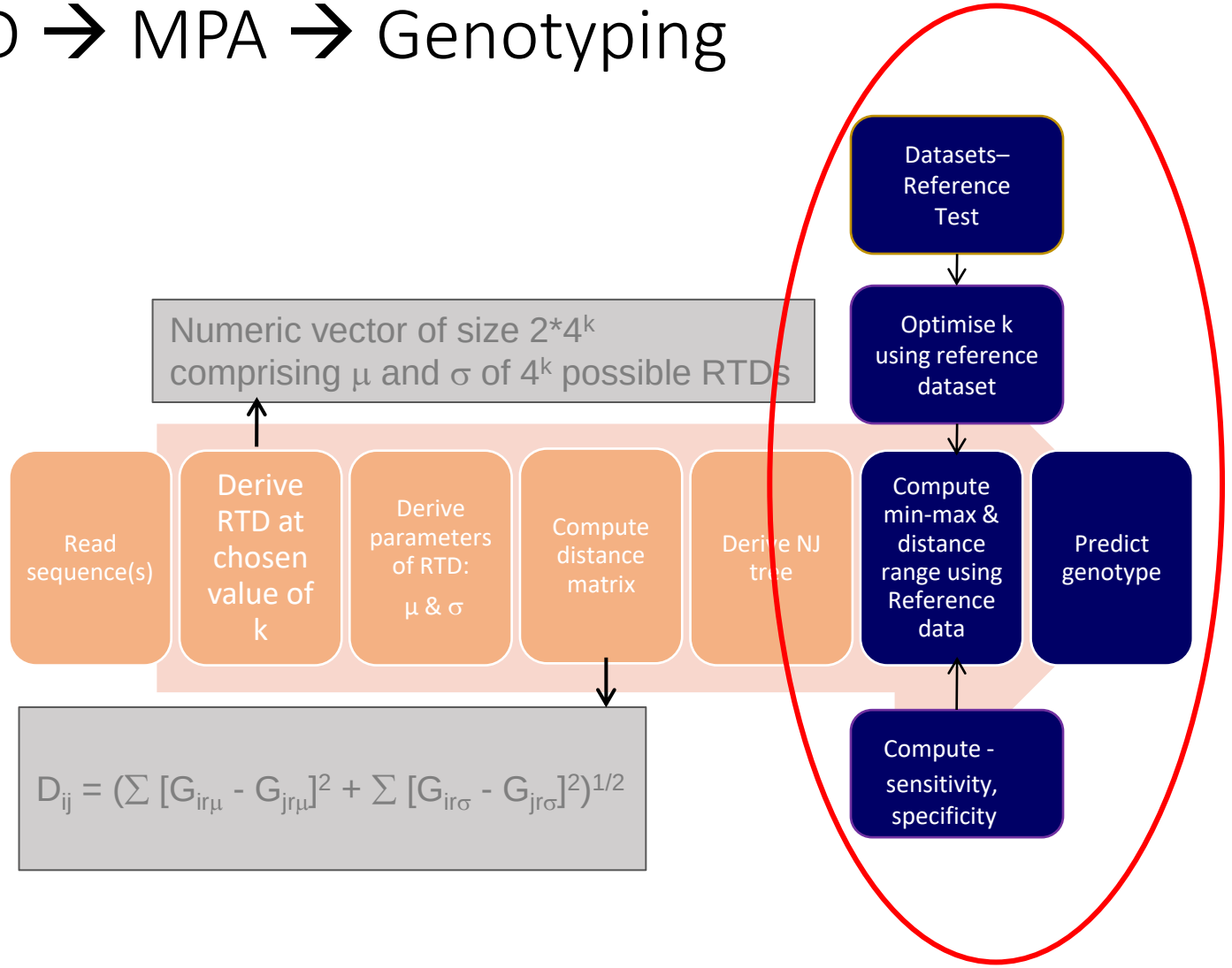
http://bioinfo.unipune.ac.in/RTD/Home.html





RTD → MPA → Genotyping

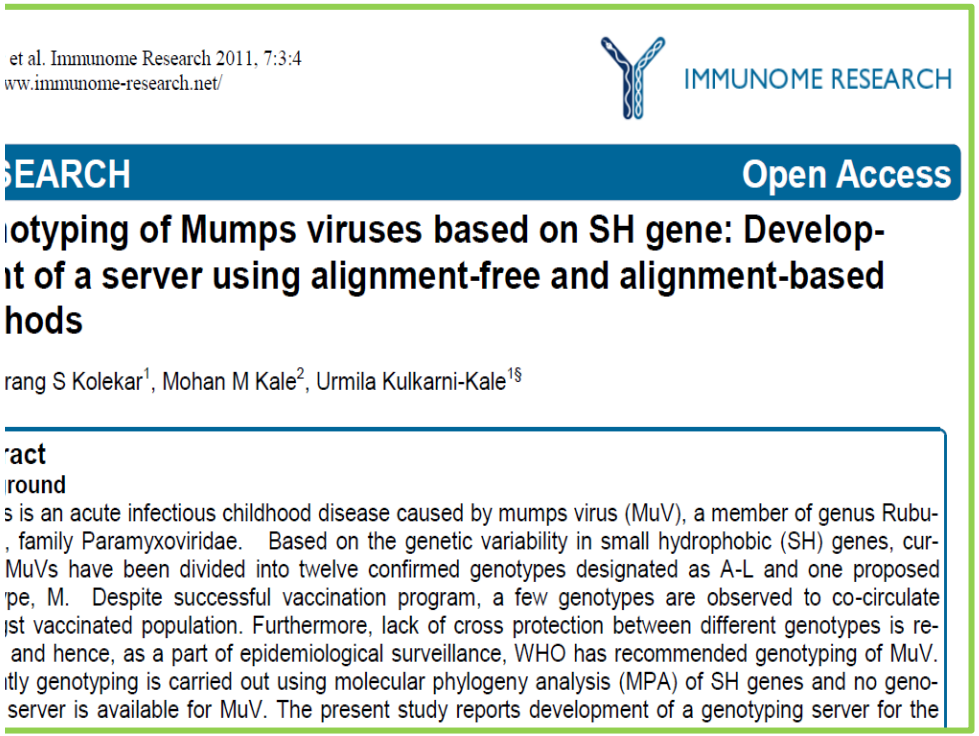




© Dr. Urmila Kulkarni-Kale | SPPU | USN | CPM |

Genotyping of *Mumps viruses*

Genus *Rubulavirus* | family *Paramyxoviridae*



et al. Immunome Research 2011, 7:3:4
ww.immunome-research.net/

IMMUNOME RESEARCH

EARCH Open Access

**otyping of Mumps viruses based on SH gene: Develop-
t of a server using alignment-free and alignment-based
hods**

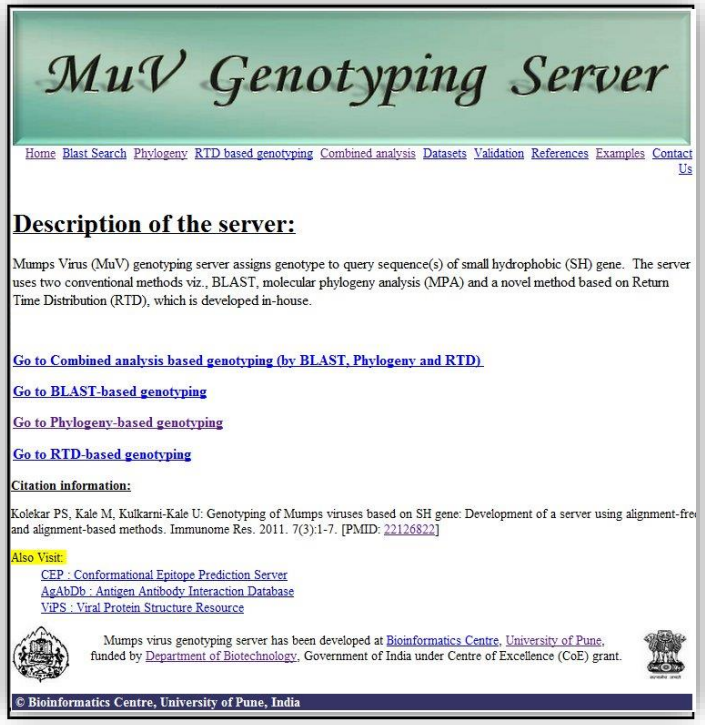
rang S Kolekar[1], Mohan M Kale[2], Urmila Kulkarni-Kale[1§]

ract

round

s is an acute infectious childhood disease caused by mumps virus (MuV), a member of genus Rubu-
, family Paramyxoviridae. Based on the genetic variability in small hydrophobic (SH) genes, cur-
MuVs have been divided into twelve confirmed genotypes designated as A-L and one proposed
pe, M. Despite successful vaccination program, a few genotypes are observed to co-circulate
st vaccinated population. Furthermore, lack of cross protection between different genotypes is re-
and hence, as a part of epidemiological surveillance, WHO has recommended genotyping of MuV.
tly genotyping is carried out using molecular phylogeny analysis (MPA) of SH genes and no geno-
server is available for MuV. The present study reports development of a genotyping server for the

- Known genotypes: 15
- Input : SH gene
- Optimum k=4
- Sensitivity : 98.95%*
- Specificity : 100%
- *Annotation error in GenBank



MuV Genotyping Server

Home Blast Search Phylogeny RTD based genotyping Combined analysis Datasets Validation References Examples Contact Us

Description of the server:

Mumps Virus (MuV) genotyping server assigns genotype to query sequence(s) of small hydrophobic (SH) gene. The server uses two conventional methods viz., BLAST, molecular phylogeny analysis (MPA) and a novel method based on Return Time Distribution (RTD), which is developed in-house.

Go to Combined analysis based genotyping (by BLAST, Phylogeny and RTD)

Go to BLAST-based genotyping

Go to Phylogeny-based genotyping

Go to RTD-based genotyping

Citation information:

Kolekar PS, Kale M, Kulkarni-Kale U: Genotyping of Mumps viruses based on SH gene: Development of a server using alignment-fre and alignment-based methods. Immunome Res. 2011. 7(3):1-7. [PMID: 22126822]

Also Visit:

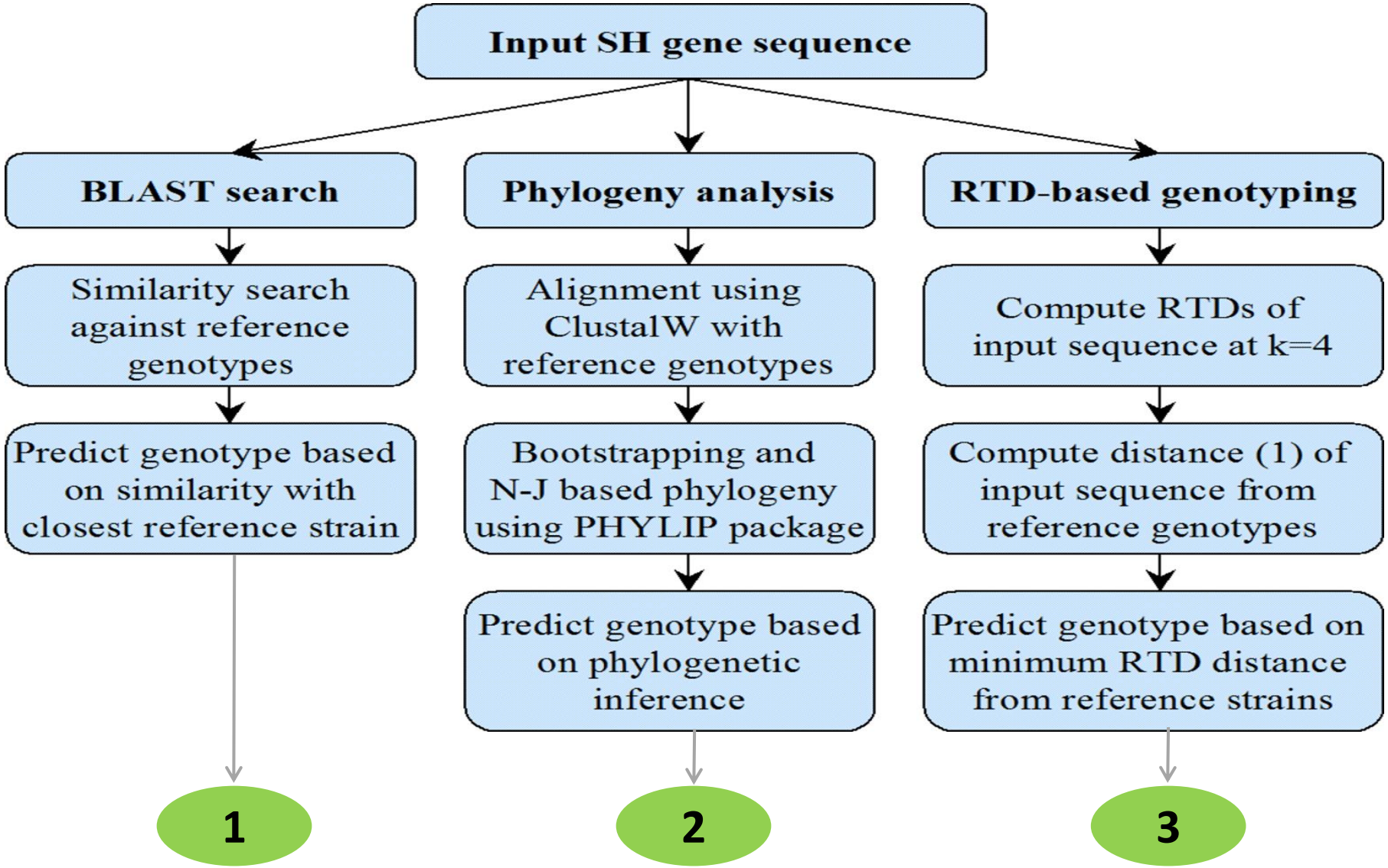
CEP : Conformational Epitope Prediction Server
AgAbDb : Antigen Antibody Interaction Database
ViPS : Viral Protein Structure Resource

Mumps virus genotyping server has been developed at Bioinformatics Centre, University of Pune, funded by Department of Biotechnology, Government of India under Centre of Excellence (CoE) grant.

© Bioinformatics Centre, University of Pune, India

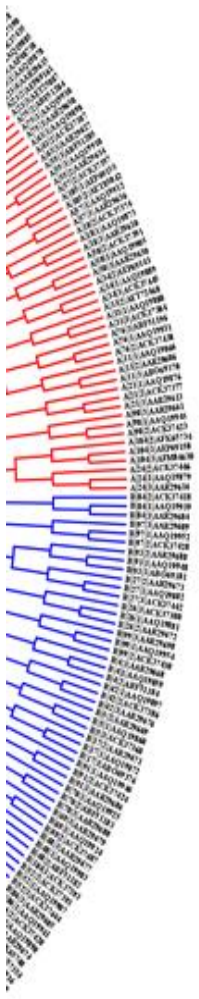
Available at:
http://bioinfo.net.in/muv/homepage.html

How Mumps Genotyping server works



© Dr. Urmila Kulkarni-Kale | SPPU | USN | CPM |

Dengue viruses : Serotypes & Genotypes



Dengue virus (DENV):

Genus: *Flavivirus*; family: *Flaviviridae*

Single-stranded positive sense RNA virus

Simple genome*: ~10.5Kb;* 3 structural and 7 non-structural genes

Four distinct '**serotypes**': DENV-1, DENV-2, DENV-3 and DENV -4

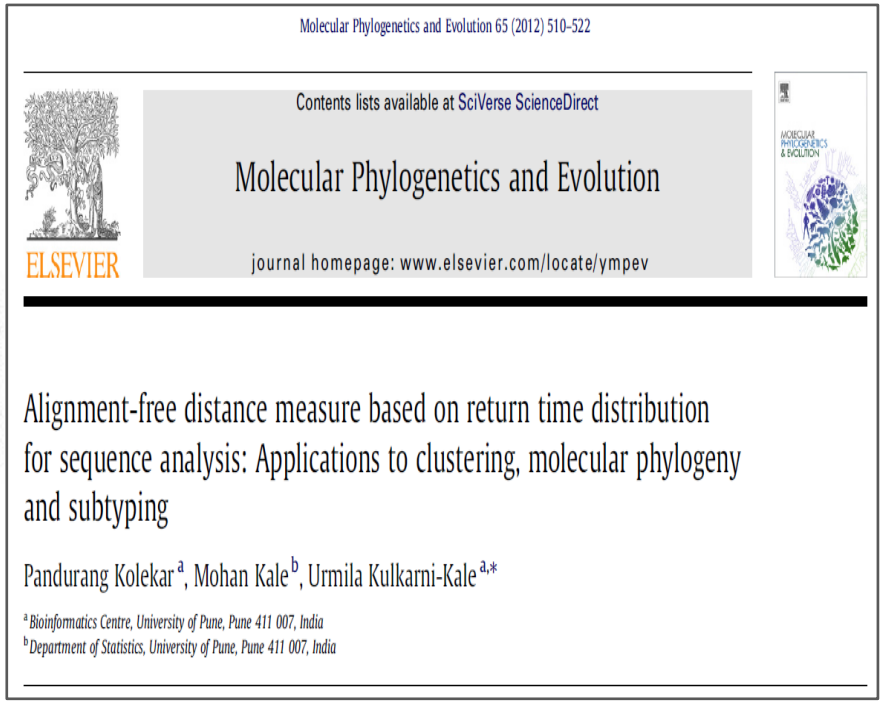
Each serotype is further divided into phylogenetically distinct 'genotypes'

Dengue virus serotype 1	Dengue virus serotype 2	Dengue virus serotype 3	Dengue virus serotype 4
• Genotype I • Genotype II • Genotype III • Sylvatic	• American • Asian-I • Asian-II • Asian/American • Cosmopolitan • Sylvatic	• Genotype I • Genotype II • Genotype III • Genotype IV • Genotype V	• Genotype I • Genotype II • Genotype II • Sylvatic
MS under preparation	Waman et al., 2016	Waman et al., 2017	Waman et al., 2016

© Dr. Urmila Kulkarni-Kale | SPPU | USN | CPM |

Subtyping of *Dengue viruses*

Genus *Flavivirus* | Family *Flaviviridae*



Molecular Phylogenetics and Evolution 65 (2012) 510-522

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

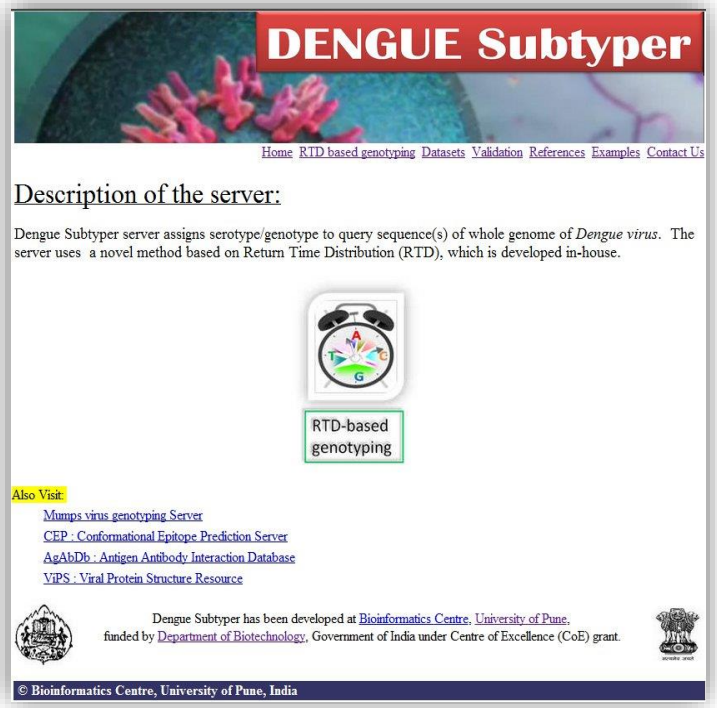
ELSEVIER

Alignment-free distance measure based on return time distribution for sequence analysis: Applications to clustering, molecular phylogeny and subtyping

Pandurang Kolekar [a], Mohan Kale [b], Urmila Kulkarni-Kale [a,*]

[a] Bioinformatics Centre, University of Pune, Pune 411 007, India
[b] Department of Statistics, University of Pune, Pune 411 007, India

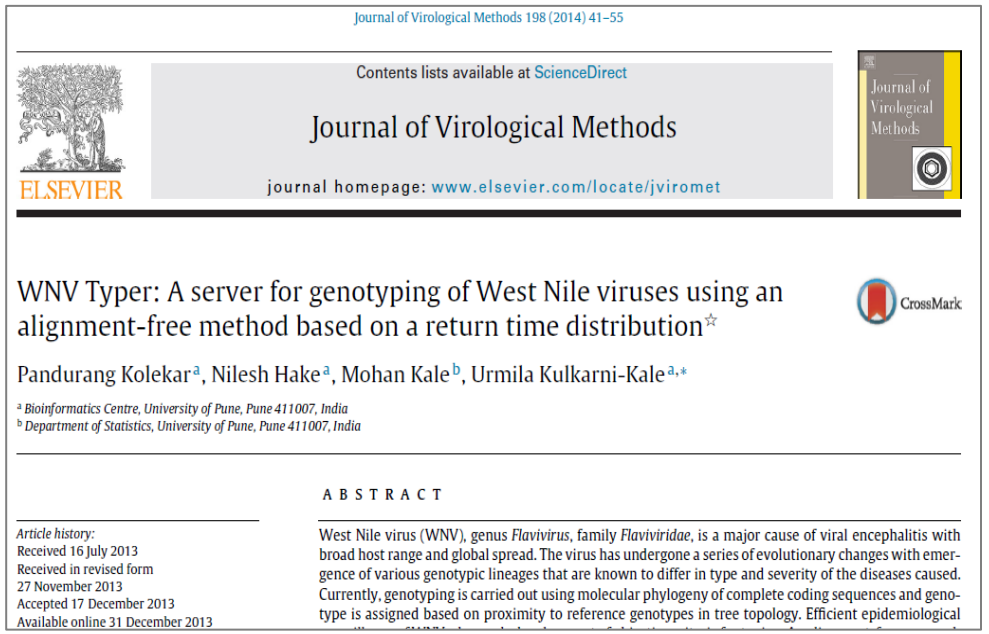
- Input : Whole genome
- Optimum k=5
- Reference dataset: 57
- Test dataset: 371
- Sensitivity : 100%
- Specificity : 100%
- Kolekar et al (2012) Mol Phyl Evol . 65(2):510-22.



Available at:
http://bioinfo.net.in/dengue/homepage.html

Genotyping of *West nile viruses*

Genus *Flavivirus* | Family *Flaviviridae*



Journal of Virological Methods 198 (2014) 41–55

Contents lists available at ScienceDirect

Journal of Virological Methods

journal homepage: www.elsevier.com/locate/jviromet

WNV Typer: A server for genotyping of West Nile viruses using an alignment-free method based on a return time distribution☆

Pandurang Kolekar[a], Nilesh Hake[a], Mohan Kale[b], Urmila Kulkarni-Kale[a,*]

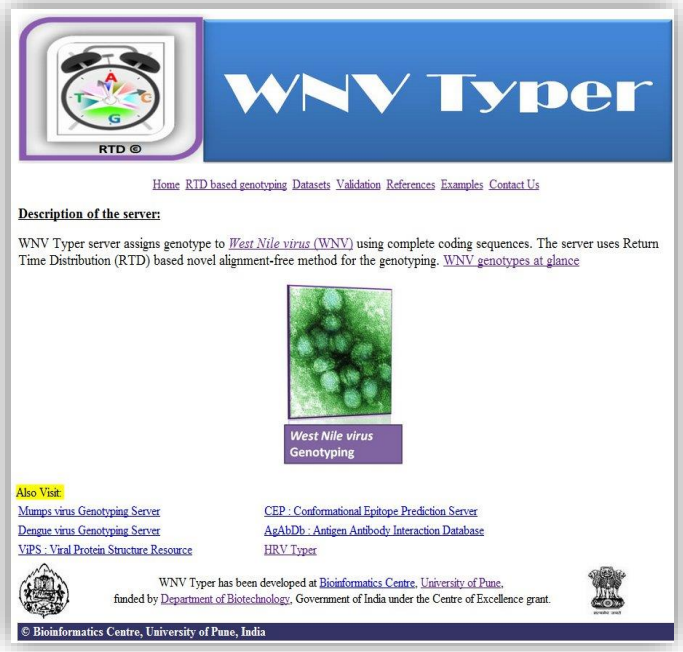
[a] Bioinformatics Centre, University of Pune, Pune 411007, India
[b] Department of Statistics, University of Pune, Pune 411007, India

Article history:
Received 16 July 2013
Received in revised form
27 November 2013
Accepted 17 December 2013
Available online 31 December 2013

ABSTRACT

West Nile virus (WNV), genus *Flavivirus*, family *Flaviviridae*, is a major cause of viral encephalitis with broad host range and global spread. The virus has undergone a series of evolutionary changes with emergence of various genotypic lineages that are known to differ in type and severity of the diseases caused. Currently, genotyping is carried out using molecular phylogeny of complete coding sequences and genotype is assigned based on proximity to reference genotypes in tree topology. Efficient epidemiological

- Input : Whole genome
- Optimum k=7
- Reference dataset: 61
- Test dataset: 566
- Sensitivity : 100%
- Specificity : 100%



WNV Typer

RTD ©

Home RTD based genotyping Datasets Validation References Examples Contact Us

Description of the server:

WNV Typer server assigns genotype to *West Nile virus* (WNV) using complete coding sequences. The server uses Return Time Distribution (RTD) based novel alignment-free method for the genotyping. WNV genotypes at glance

West Nile virus Genotyping

Also Visit:

Mumps virus Genotyping Server
Dengue virus Genotyping Server
ViPS : Viral Protein Structure Resource
CEP : Conformational Epitope Prediction Server
AgAbDb : Antigen Antibody Interaction Database
HRV Typer

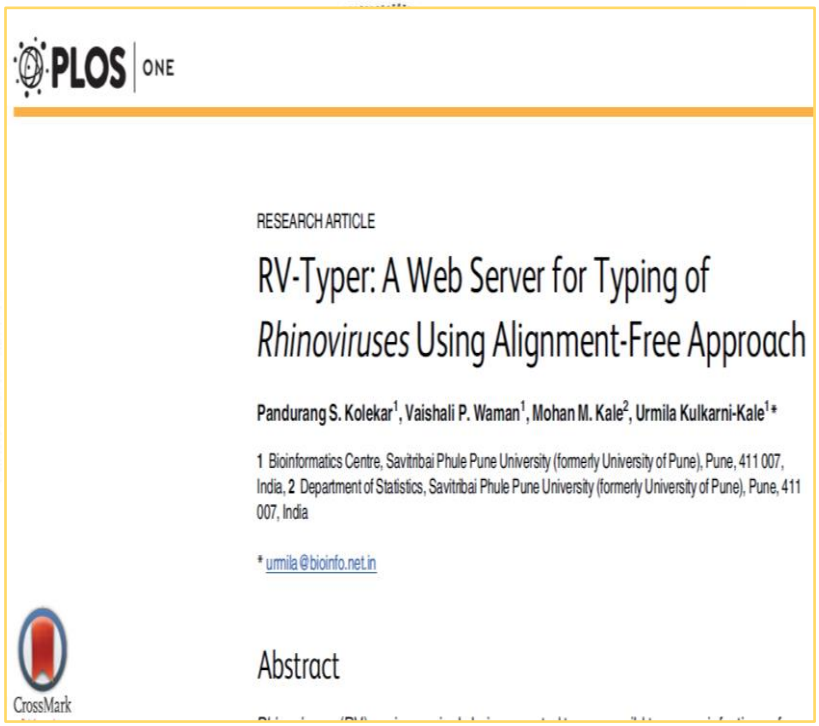
WNV Typer has been developed at Bioinformatics Centre, University of Pune, funded by Department of Biotechnology, Government of India under the Centre of Excellence grant.

© Bioinformatics Centre, University of Pune, India

Available at:
http://bioinfo.net.in/wnv/homepage.html

Genotyping of *Human rhinoviruses*

Genus Enterovirus | Family *Picornaviridae*



PLOS ONE

RESEARCH ARTICLE

RV-Typer: A Web Server for Typing of *Rhinoviruses* Using Alignment-Free Approach

Pandurang S. Kolekar[1], Vaishali P. Waman[1], Mohan M. Kale[2], Urmila Kulkarni-Kale[1]*

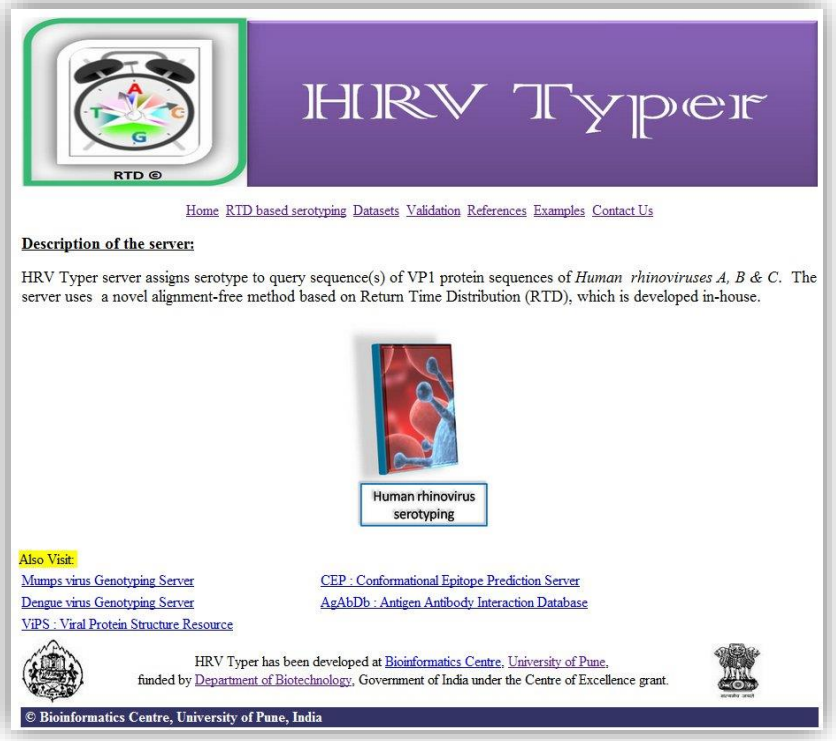
1 Bioinformatics Centre, Savitribai Phule Pune University (formerly University of Pune), Pune, 411 007, India, 2 Department of Statistics, Savitribai Phule Pune University (formerly University of Pune), Pune, 411 007, India

* urmila@bioinfo.net.in

CrossMark

Abstract

- Input : VP1 protein
- Optimum k=1
- Reference dataset: 432
  RV-A (238), -B (83) and -C (111)
- Test dataset: 218
- Sensitivity : 100%
- Specificity : 100%



HRV Typer

RTD ©

Home RTD based serotyping Datasets Validation References Examples Contact Us

Description of the server:

HRV Typer server assigns serotype to query sequence(s) of VP1 protein sequences of *Human rhinoviruses A, B & C*. The server uses a novel alignment-free method based on Return Time Distribution (RTD), which is developed in-house.

Human rhinovirus serotyping

Also Visit:

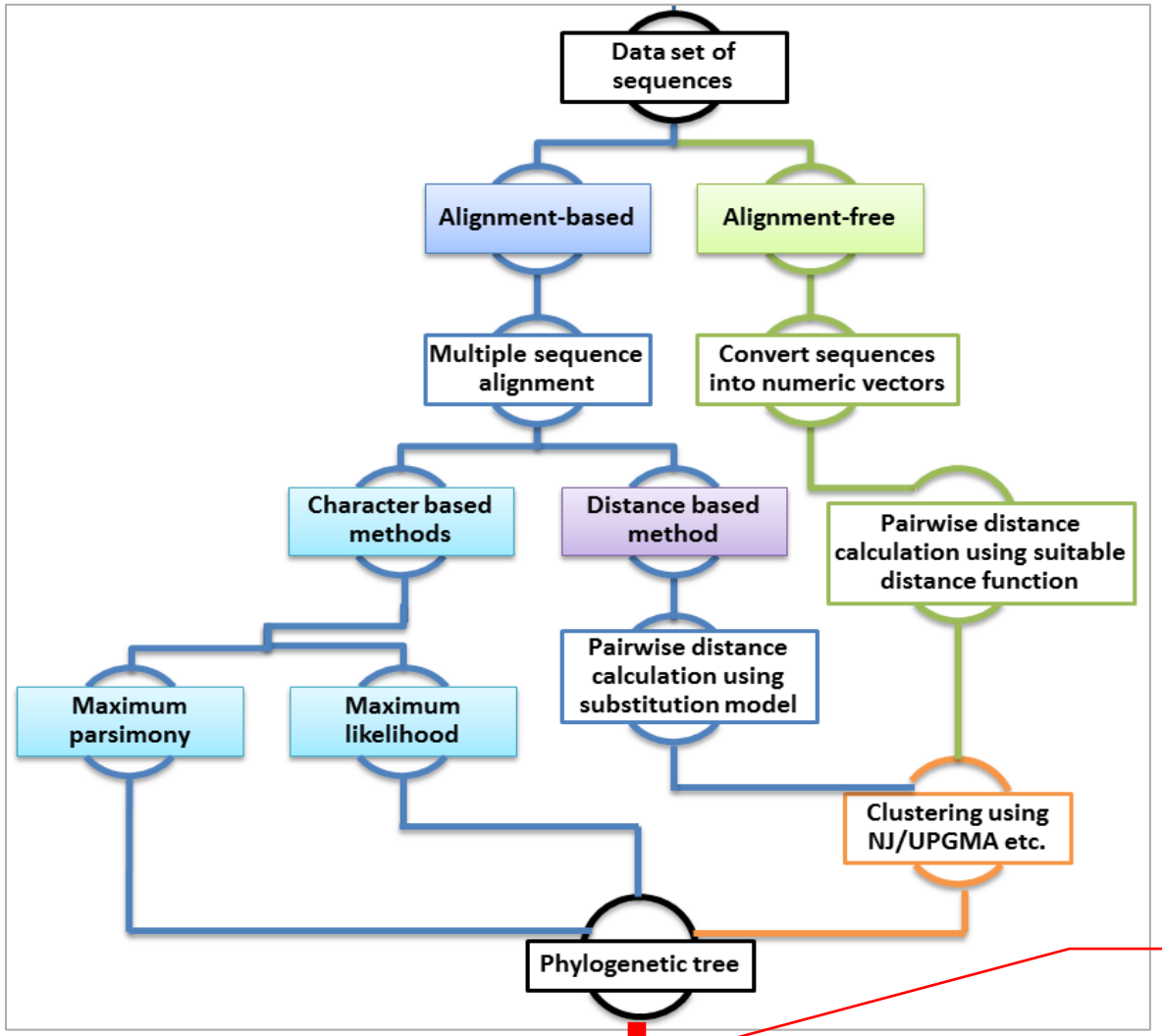
Mumps virus Genotyping Server
Dengue virus Genotyping Server
ViPS : Viral Protein Structure Resource
CEP : Conformational Epitope Prediction Server
AgAbDb : Antigen Antibody Interaction Database

HRV Typer has been developed at Bioinformatics Centre, University of Pune, funded by Department of Biotechnology, Government of India under the Centre of Excellence grant.

© Bioinformatics Centre, University of Pune, India

Available at:
http://bioinfo.net.in/hrv/homepage.html

Molecular Phylogeny Approaches for Viral Typing



Alignment-free Phylogeny:

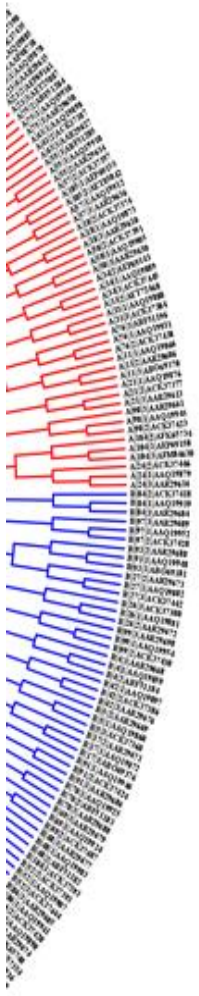
- Quick computations
- High accuracy
- Works equally well with sequences having similarity in the range of 25-99%
- Uses BLOCK-Bootstrap

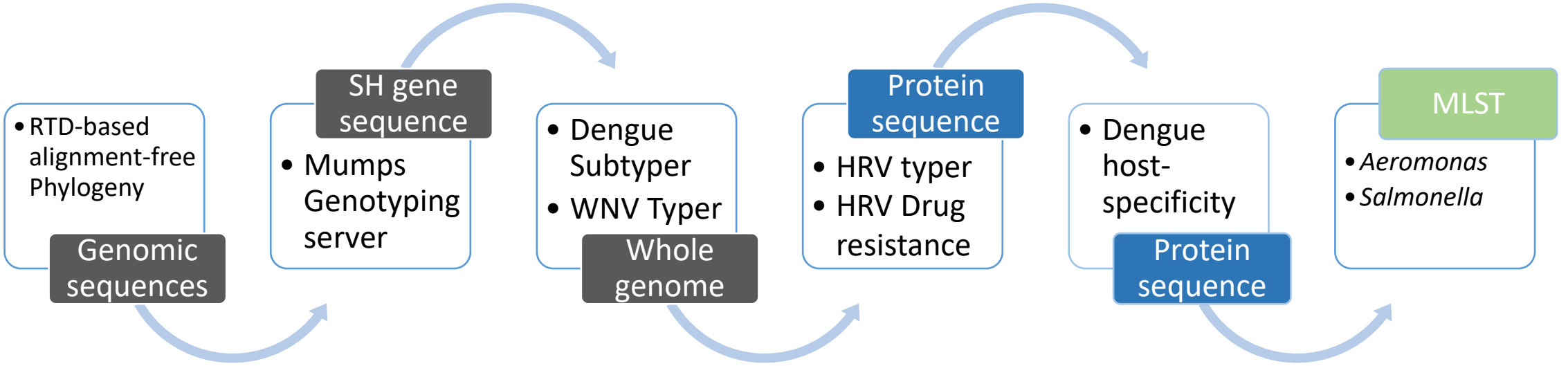
Knowledge-based Severs for Viral Typing

Assign geno- /serotype based on clustering proximity of unknown with known

© Dr. Urmila Kulkarni-Kale | SPPU | USN | CPM |

RTD for Clustering, Molecular Phylogeny, Genotyping, Serotyping & Multi-Locus Sequence Typing





© Dr. Urmila Kulkarni-Kale | SPPU | USN | CPM |