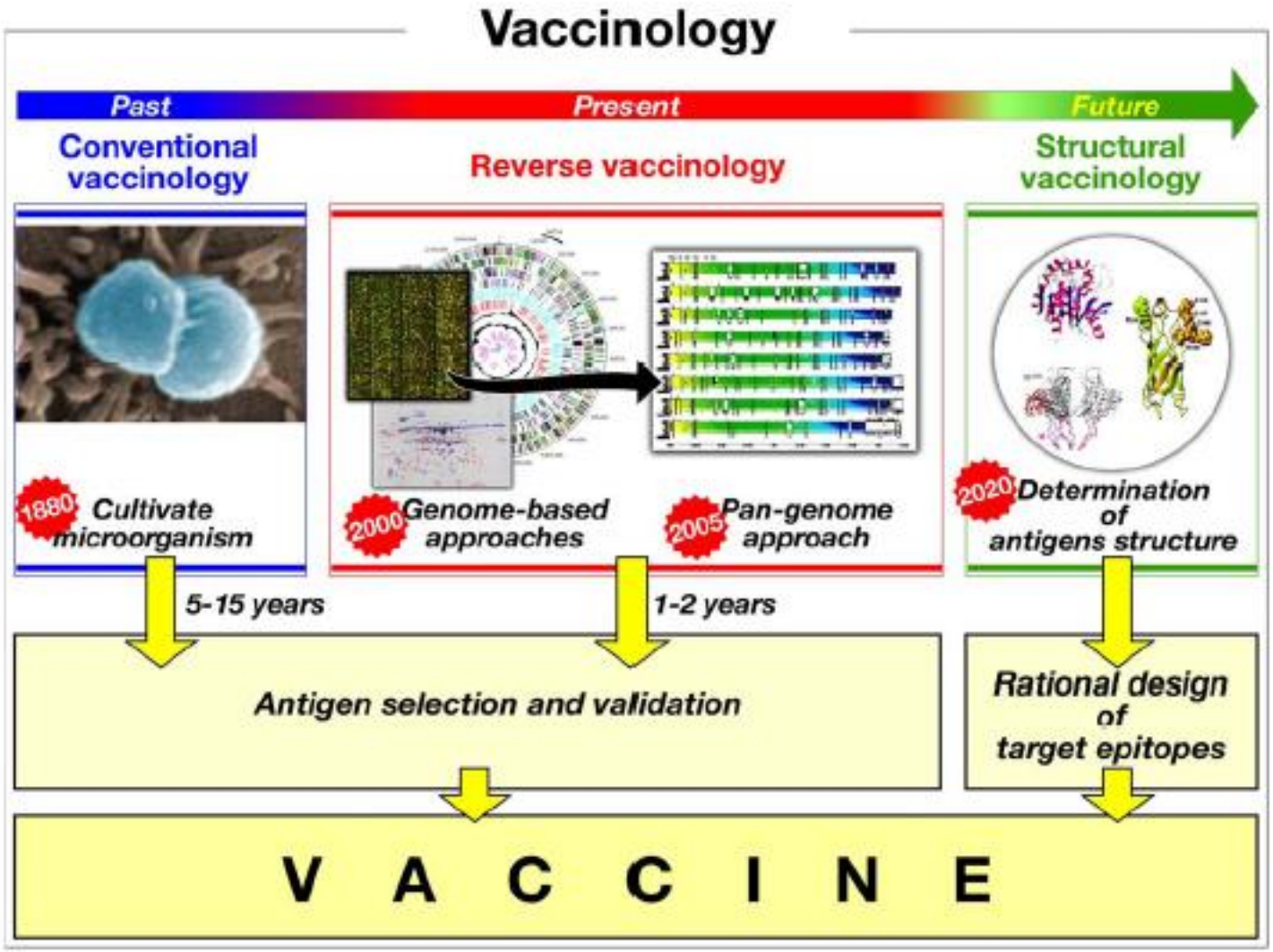
Mapping population & genotype/serotype diversity: game changers in designing viral vaccines

Dr. Urmila Kulkarni-Kale, FMASc
Bioinformatics Centre
Savitribai Phule Pune University
Pune 411007. India
urmila@bioinfo.net.in
Urmila.kulkarni.kale@gmail.com

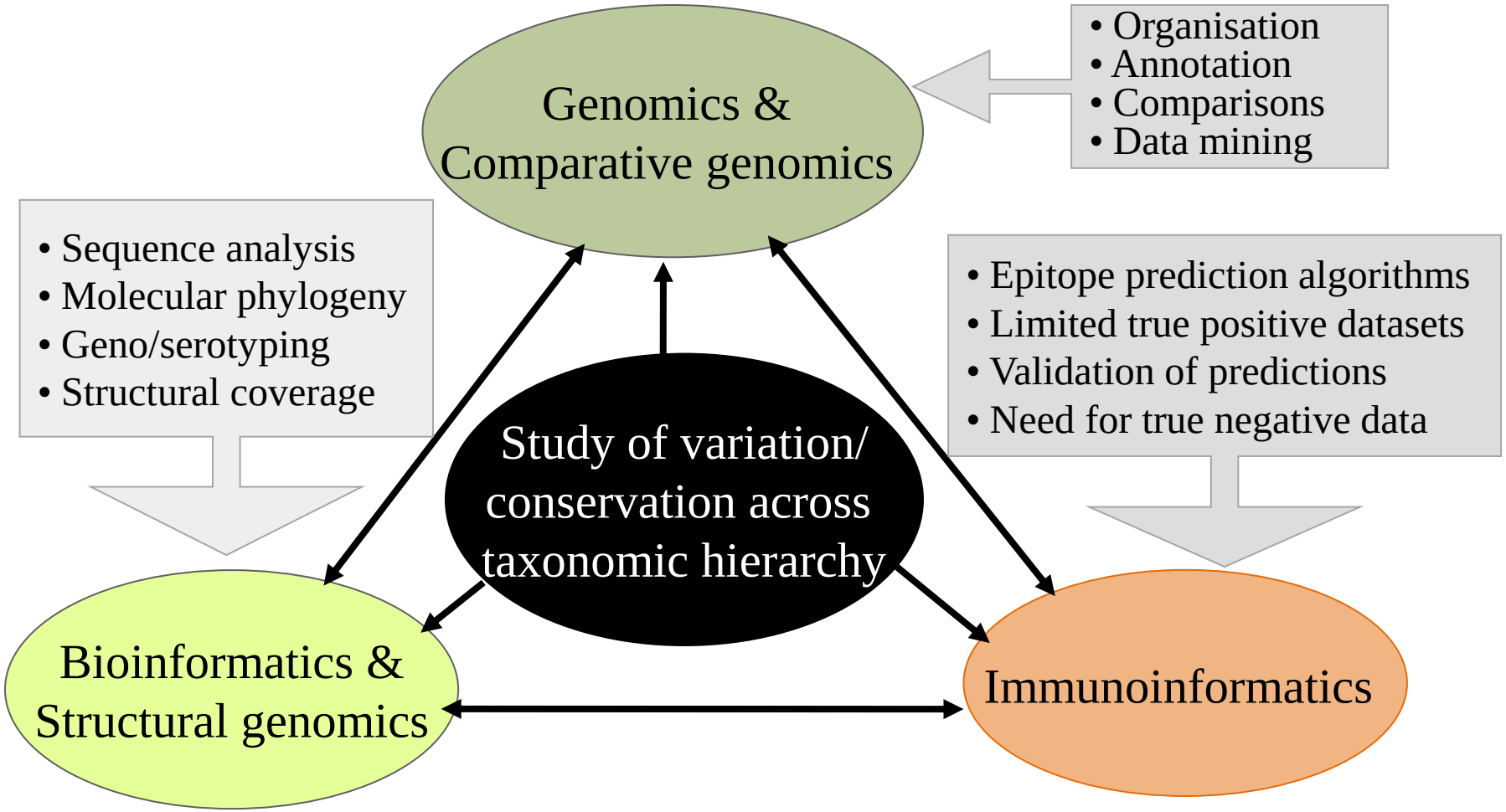
Reverse Vaccinology Approach



Serruto & Rappuoli, FE BS Letters, 580 (2006) 2985–2992

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Genome 2 Vaccinome: Opportunities & Challenges

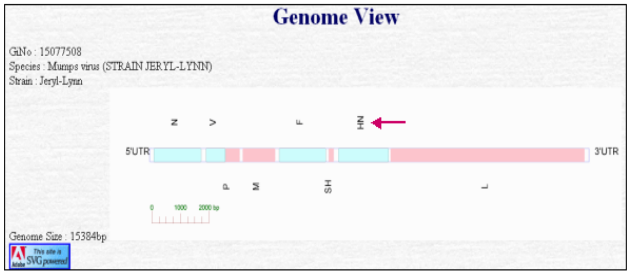


Kulkarni-Kale et al., CBIO, 2012. Volume 7 (4), 454-466.

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Mumps Virus: antigenic diversity & strain specificity

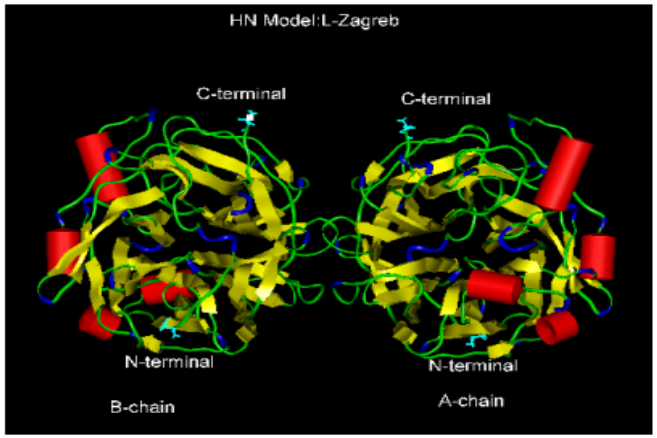
Mumps Virus:at a glance
Source: VirGen database



Order:	Mononegavirales
Family:	*Paramyxoviridae*
Subfamily:	*Paramyxovirinae*
Genus:	Rubulavirus
Species:	Mumps virus

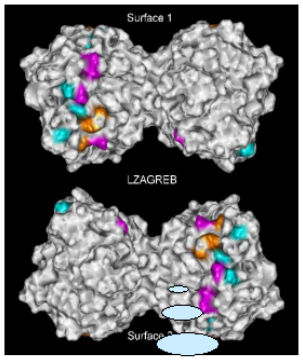
Genome: -ve sense ssRNA
Genotypes: 10: A→J (SH gene)
Known antigenic proteins:
F & HN

L-Zagreb Model: Dimer



L-Zagreb HN:
Kabsch & Sander secondary rendering

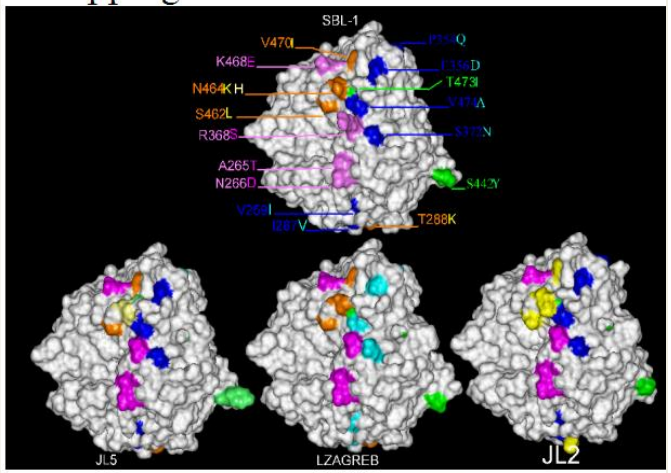
LZAGREB: SAS



- Total mutations: 24
  - Side1: 16
  - Side2: 8
  - On surface: 23
  - Buried: 1
- Dimer interface free of mutations
- Sequentially apart but spatially localised
- Region encompasses predicted conformational epitopes

potential site for antibody recognition as well as neutralisation

Mapping mutations on 3D structure



Colour: according to majority

Funded by: Serum Institute of India, Pune
Kulkarni-Kale et al., 2007, Virology, 359(2):436-46.

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Study of variations at different levels of Biocomplexity

- Strains/isolates of a virus
- Serotypes/genotypes of a virus
- Viruses that belong to same genus
- Viruses that belong to same family

How similar is similar? How different is different?

Correlate: genotype with phenotype

Implications in:
Host-Virus interactions
Rational design of vaccines & drugs
Development of diagnostics

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, University of Pune

Molecular Phylogeny Analysis (MPA):

permits study of similarities within the group and differences between the groups

- Integral part of sequence analysis in bioinformatics
- Applications:
  - Evolution of gene(s) in a group of species
  - Evolution of species
  - Assignment of genotype/serotype, strains
  - Map emergence of drug resistance
  - Prioritization of vaccine candidates

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, University of Pune

MPA: steps

- Types of models
  - Distance based (UPGMA, NJ)
  - Character based (Maximum parsimony)
  - Probabilistic (Likelihood)

- Define a question
- A set of sequences
- Multiple sequence alignments
- Selection of a model
- Use of clustering method(s)
- Generate consensus tree
- Statistical models to assess tree topology(ies)
- Analysis of inferred tree(s) → Assign geno-/serotypes

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Limitations of MPA methods

- Positions of IN-DELs in MSA impact model of evolution
  - Errors in alignment increases as sequence similarity decreases
- Assumption of character-based MPA methods
  - Sites evolve independently
- Different methods result into different trees
  - Becomes a matter of interpretation
- Need to repeat analysis with every New sequence
  - Time consuming and tedious

- Size of data in post-genomic era
- Computational complexity and memory requirements
- Time requirements (as length & number of sequences increase)

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Alternate Alignment-free Methods for MPA

- Composition vector based CVTree Method
  (Qi et al., 2004)
- Feature Frequency Profile (FFP)
  (Sims et al., 2008)
- Advantages
  - Simple, faster
  - Applications demonstrated for clustering & phylogeny
- Disadvantages
  - Takes only frequency in account (not the context)
  - Misclassification and alternate tree topology

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Proposed RTD-based approach

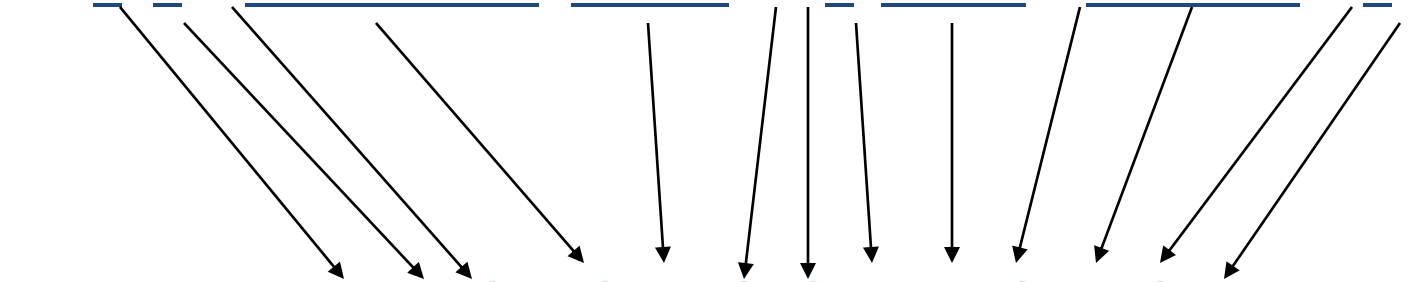
- Based on the concept of Return Time Distribution in stochastic modeling
- Return Time (RT): Time required for the reappearance of particular state without its appearance in between
- Alignment free

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Computing RTD for 'A'

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

CTACACAACTTTGCGGGTAGCCGGAAACATTGTGAATGCGGTGAACA



1-1-0-10-5-0-0-1-5-0-7-0-1

Return times for 'A'

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'

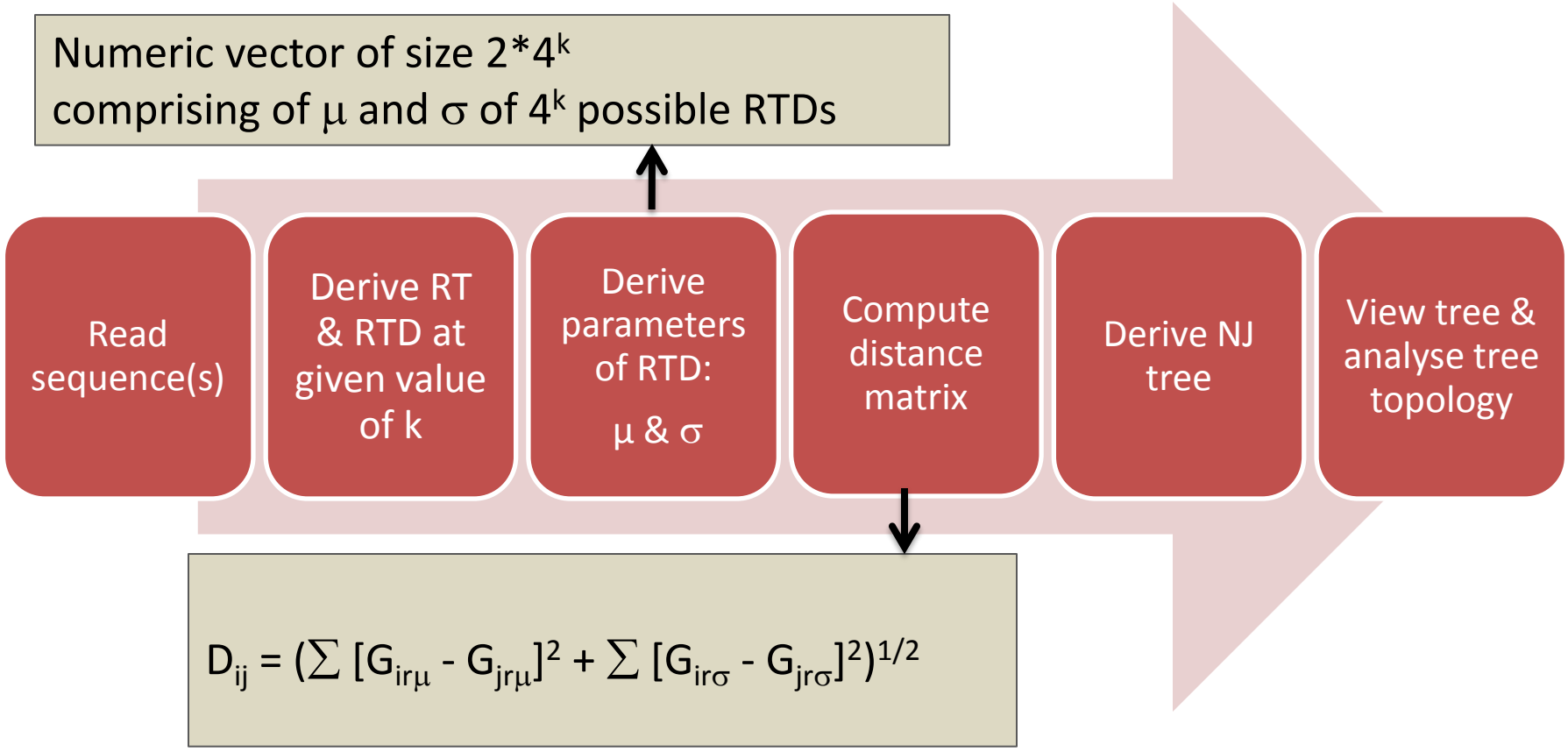
$\mu(A) = 2.38$ and $\sigma(A) = 3.27$

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

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

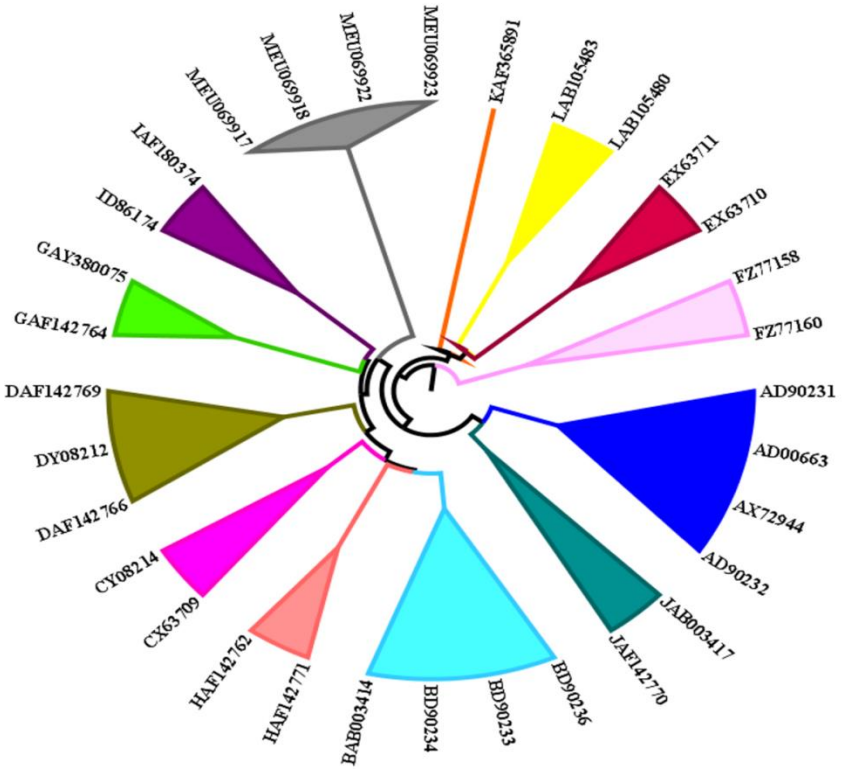
RT→RTD →Distance matrix

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



© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

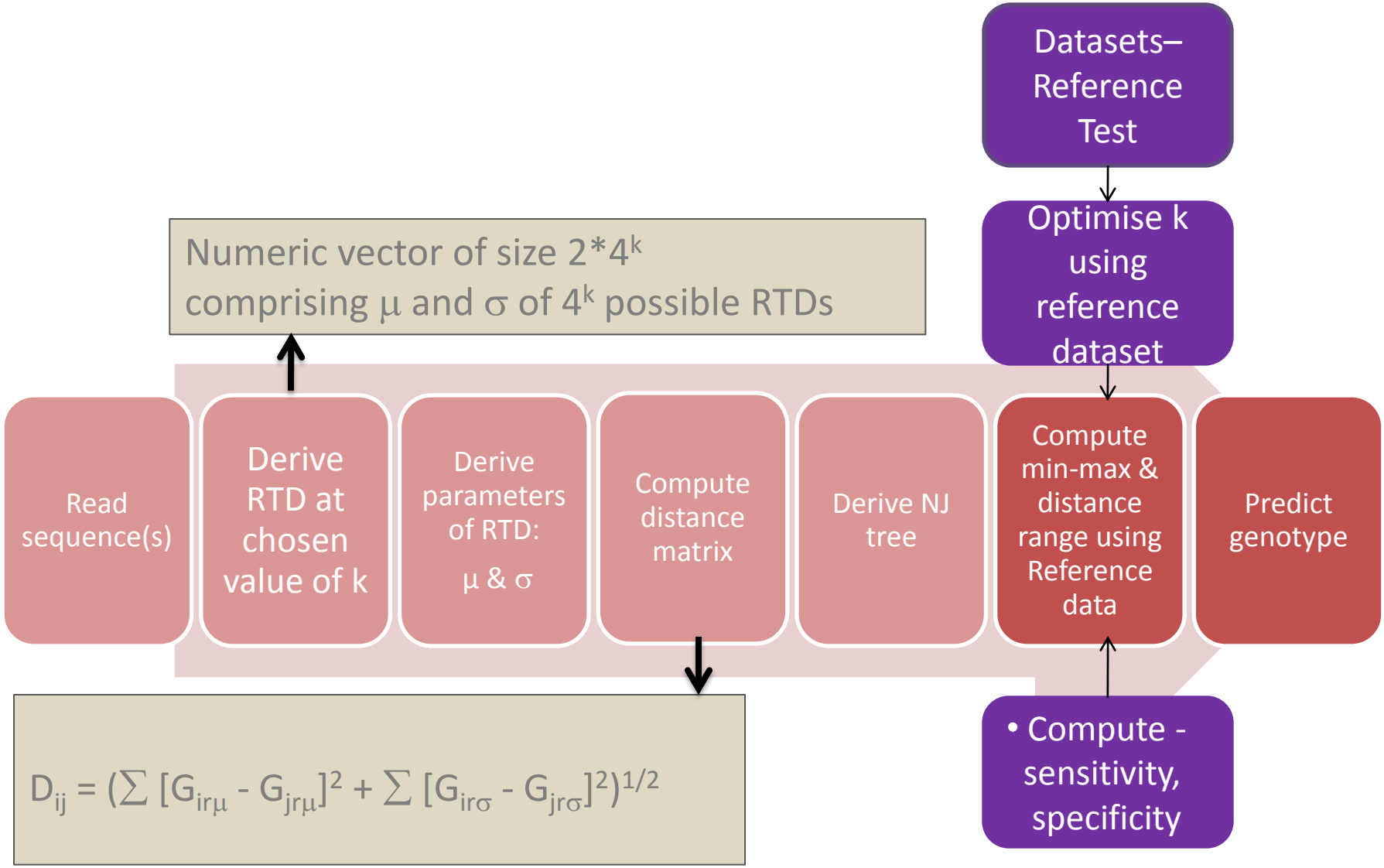
Clustering of Mumps Viruses using sequences of SH & RTD at K=4



Reference data: Mumps Virus: Known genotypes (A-L)

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

RTD → MPA → Genotyping



© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Mumps: Datasets used

- Data source: GenBank

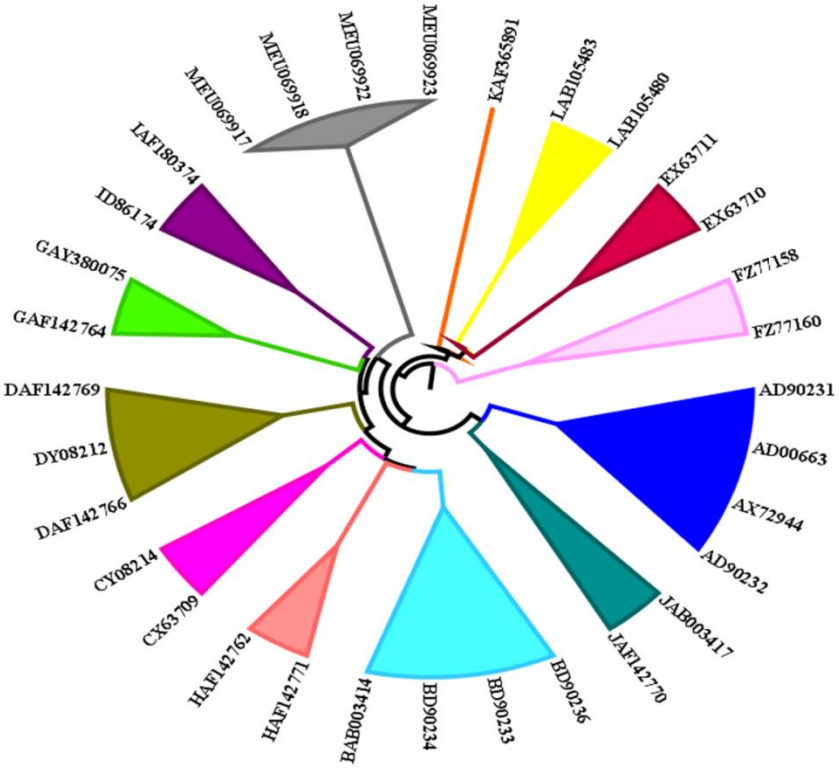
- Reference dataset:
  28 sequences of known genotypes
- Test dataset 1:
  96 entries with known genotypes
- Test dataset 2:
  380 entries
- True negative dataset:
  Non-SH Mumps sequences
  Non-Mumps SH sequences
  Non-Mumps, Non-SH sequences

Genotype	Reference dataset	Test dataset 1	Test dataset 2
A	4	-	22
B	4	-	63
C	2	3	9
D	3	-	32
E	2	-	1
F	2	49	8
G	2	20	158
H	2	11	26
I	2	-	15
J	2	13	44
K	1	-	-
L	2	-	2
Total	**28**	**96**	**380**

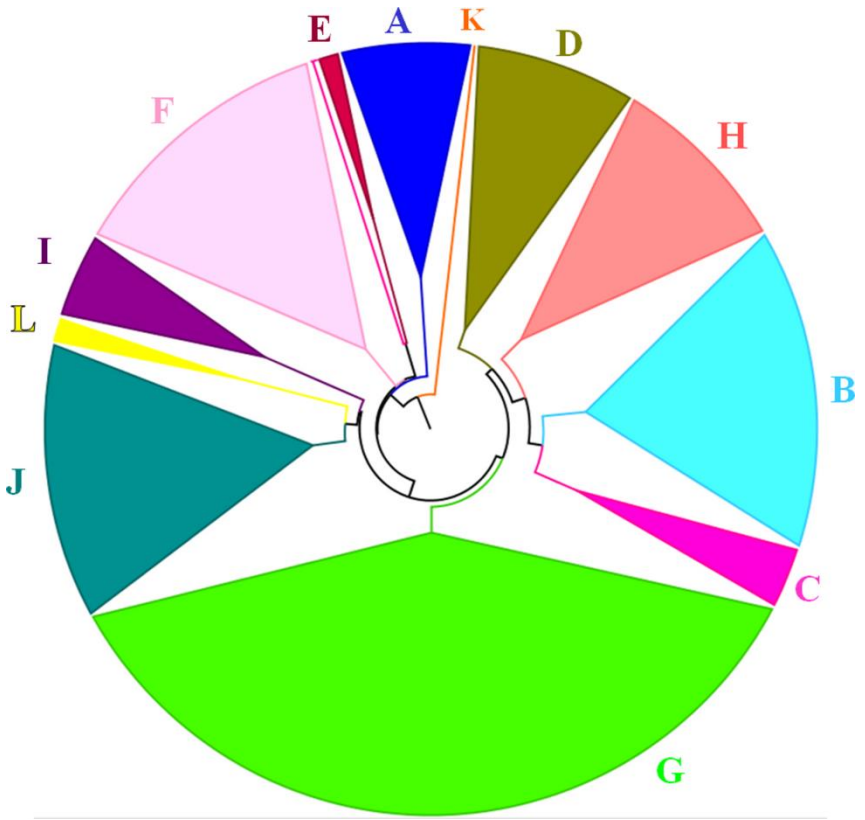
© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Clustering of SH using RTD at K=4

Reference data

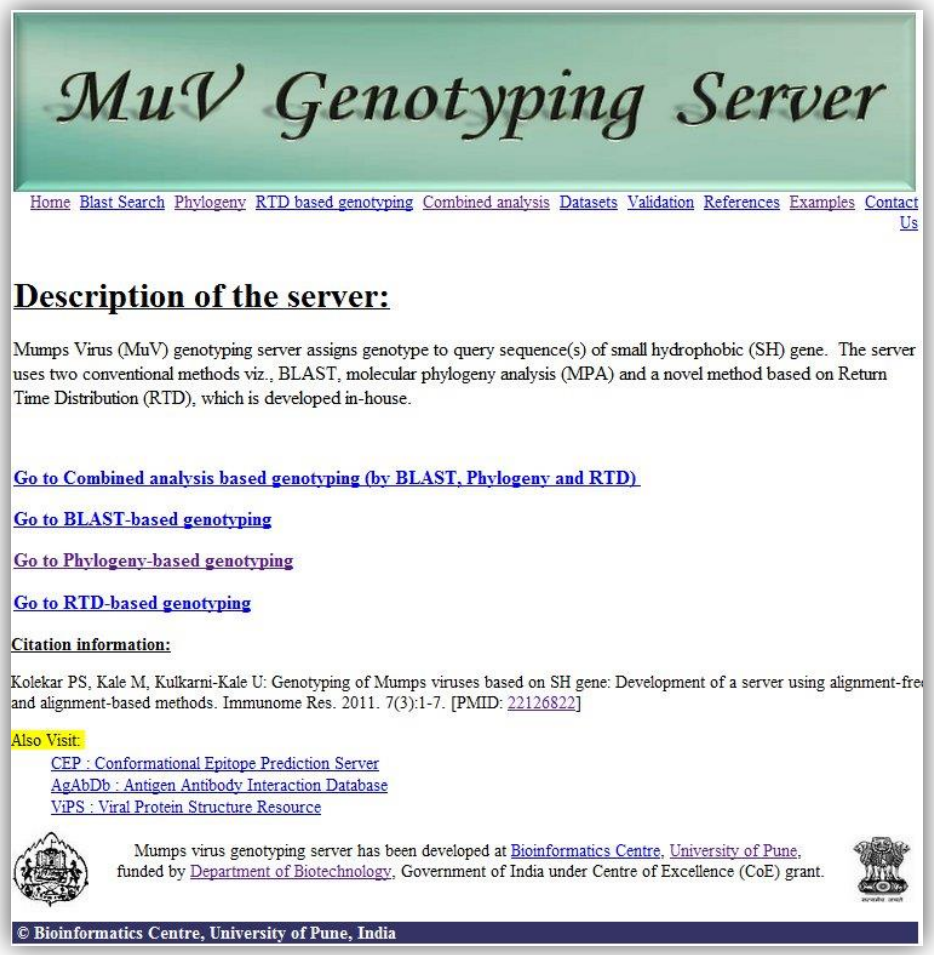


Test dataset 2



© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Genotyping of *Mumps viruses*



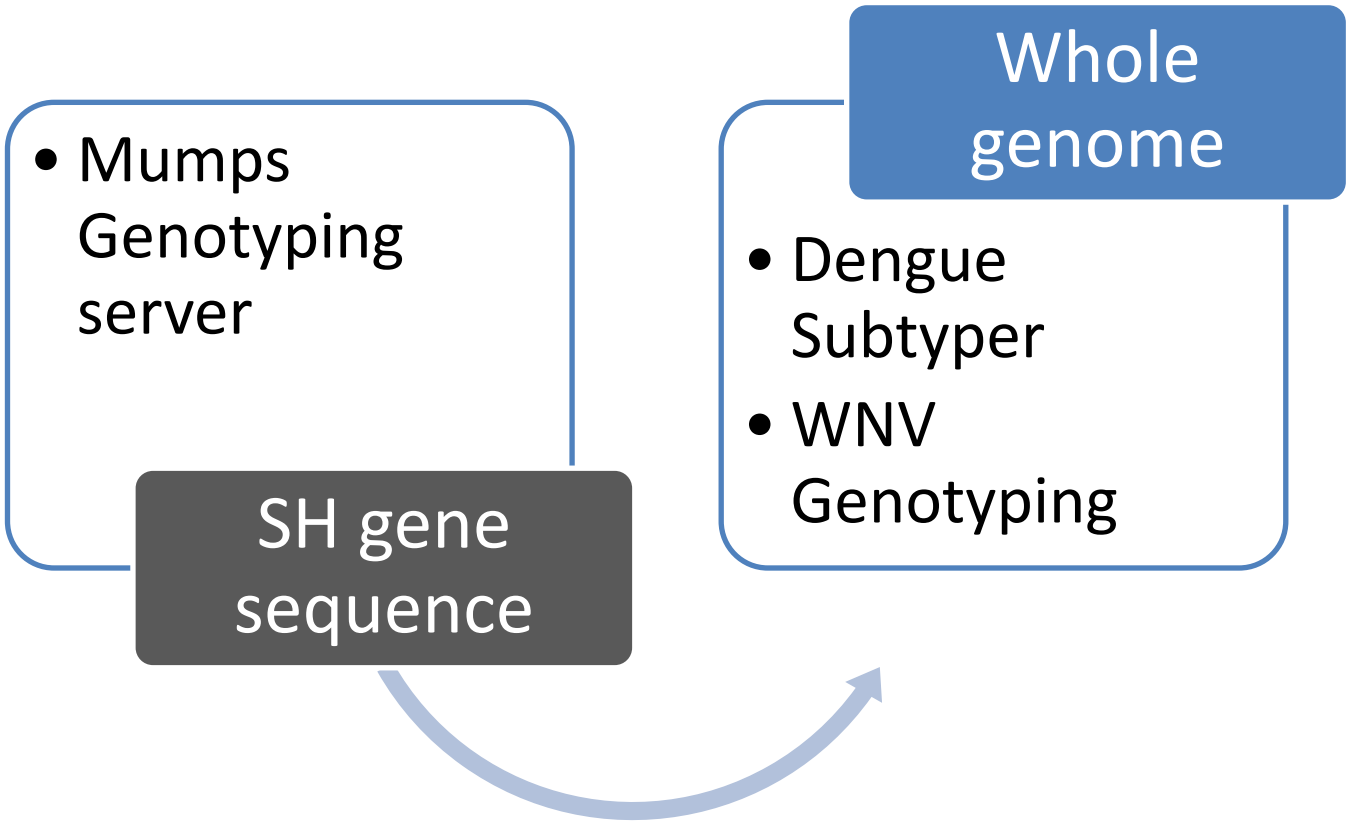
- Known genotypes: 15
- Input : SH gene
- Optimum k=4
- Sensitivity : 98.95%
- Specificity : 100%

- Kolekar et al (2011) Immunome Res, 7(3):1-7

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

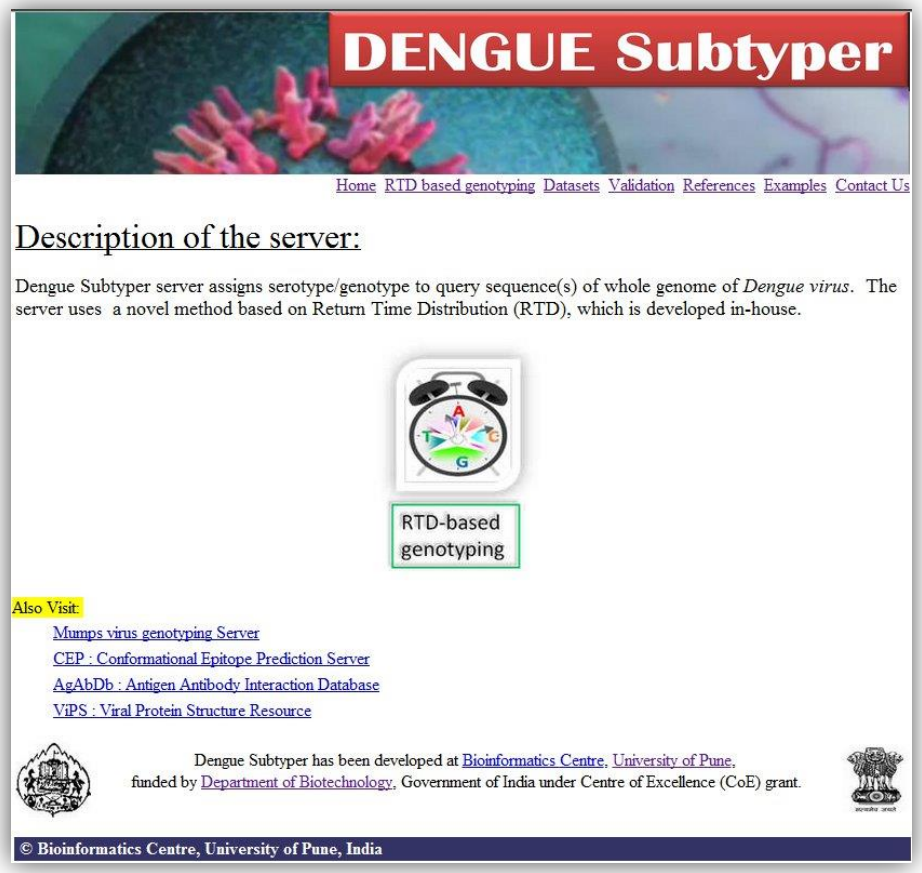
© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

RTD for MPA, Serotyping, Genotyping, & Clustering



© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Subtyping of *Dengue viruses*



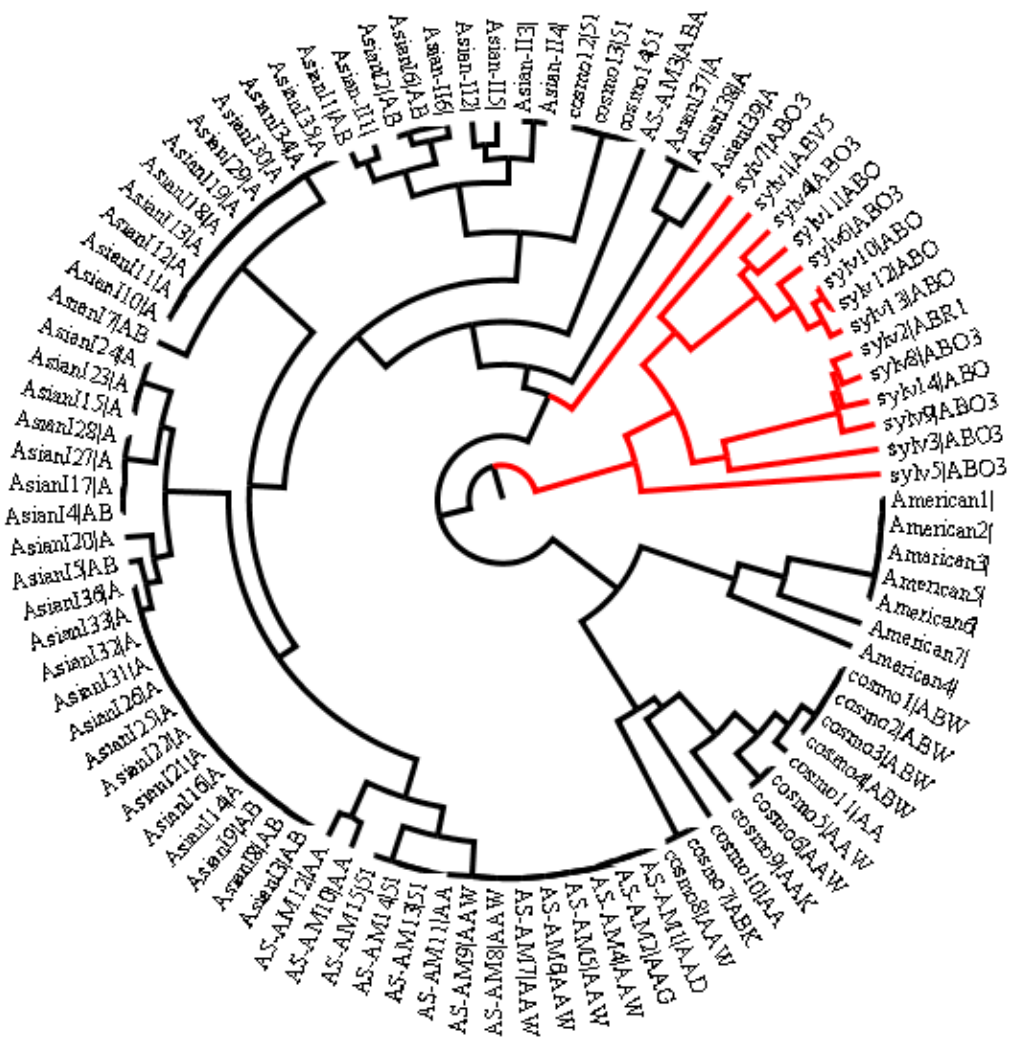
- Input : Whole genome
- Optimum k=5
- Sensitivity : 100%
- Specificity : 100%

Kolekar et al (2012) *Mol Phyl Evol* . Molecular Phylogenetice & Evolution. 2012 Nov;65(2):510–22.

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

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

RTD-based clustering of urban and sylvatic DENV-2 using sequences of Envelope glycoprotein (egp)



Dengue-2 virus serotype is divided into 6 genotypes viz. American, American-Asian, Asian-I, Asian-II, Cosmopolitan and sylvatic.

These genotypes are categorized into urban (endemic/epidemic) and sylvatic types based on their host transmission.

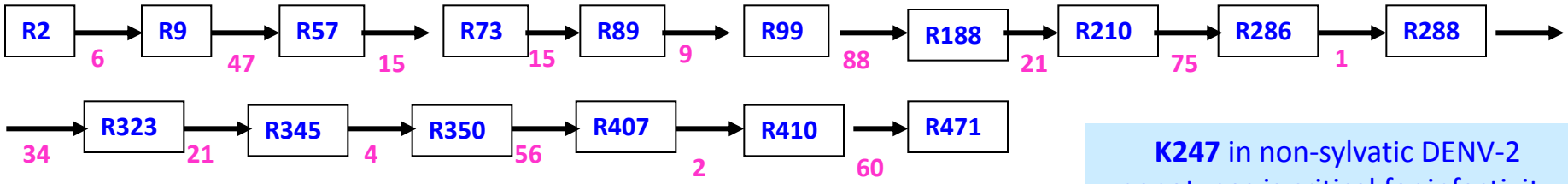
Urban viruses infects humans while sylvatic viruses infects non-human hosts.

RTD-based informative residues viz. N, I and R obtained by WEKA helps in clustering of Dengue-2 wrt host specificity at K=1.

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

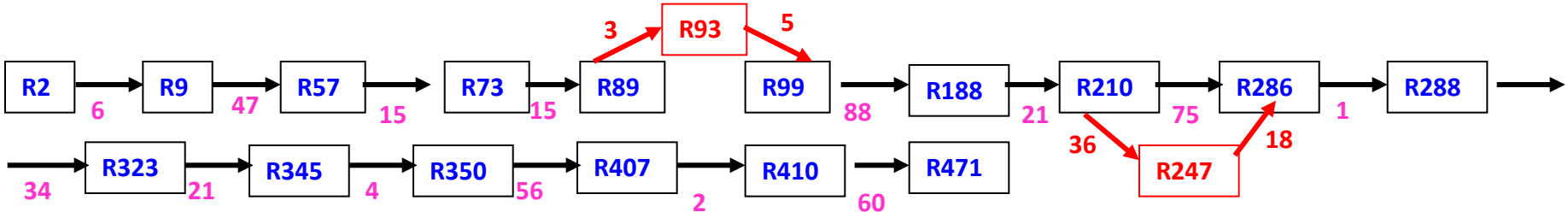
Application of RTD to predict host-specificity
RT of R mapped on 3D structure of E protein

RTD of R residue in non-sylvatic genotypes



K247 in non-sylvatic DENV-2 genotypes is critical for infectivity in humans; while Sylvatic strains shows **K247R** mutation

RTD of R residue in Sylvatic genotype



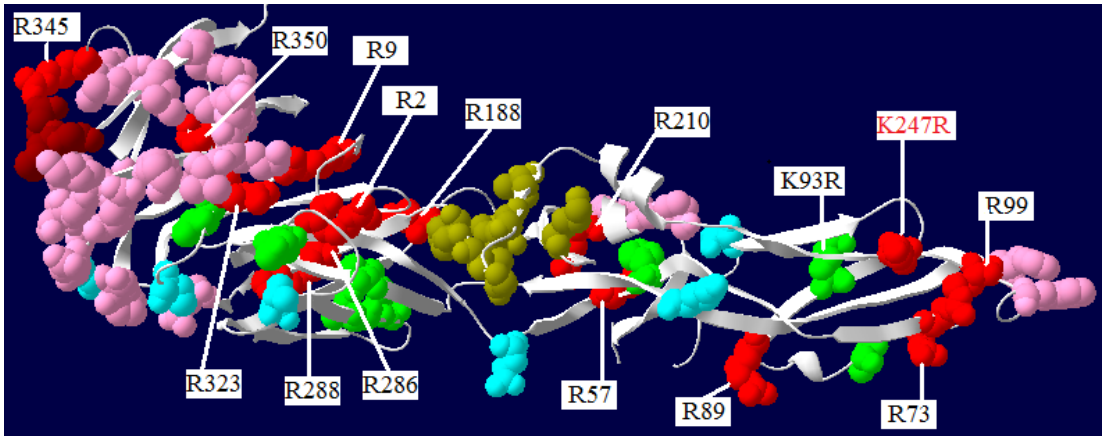
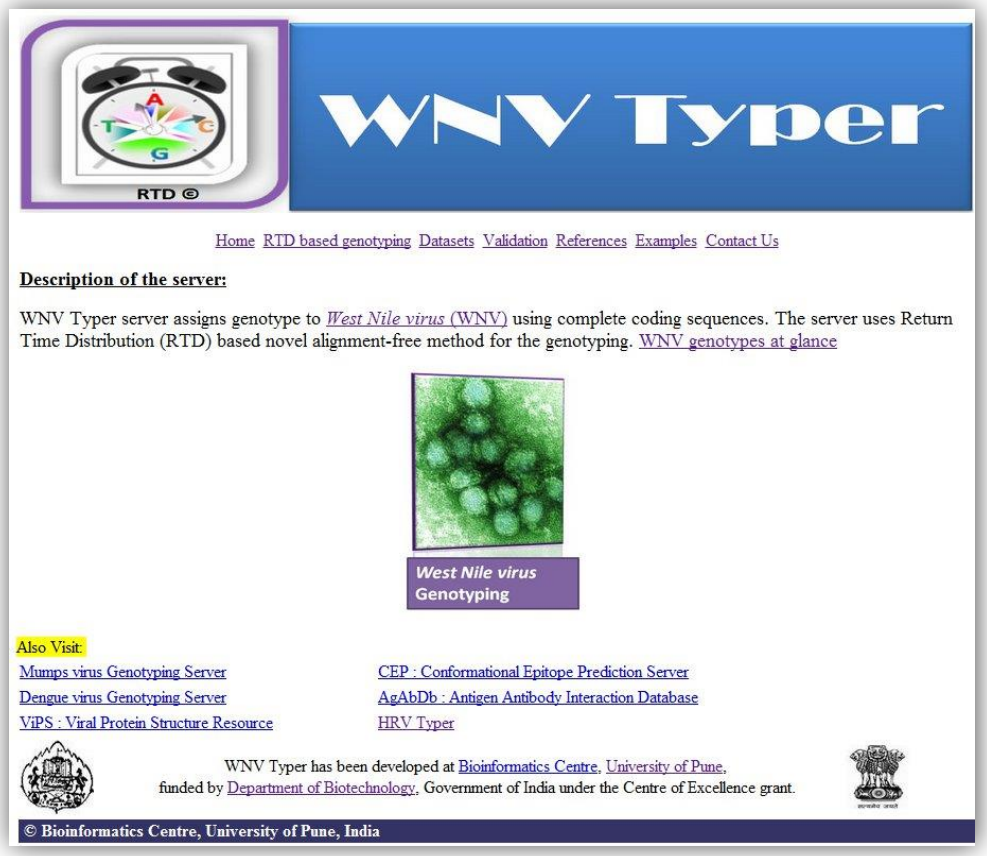


Figure shows mapping of R residues on the E protein structure [1TG8] of American strain (Tonga/EKB194/1974)]

Known epitopes, Ligand-binding residues & receptor-binding residues, Evolutionary trace residues reported to be binding site and novel are shown.

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Genotyping of *West nile viruses*

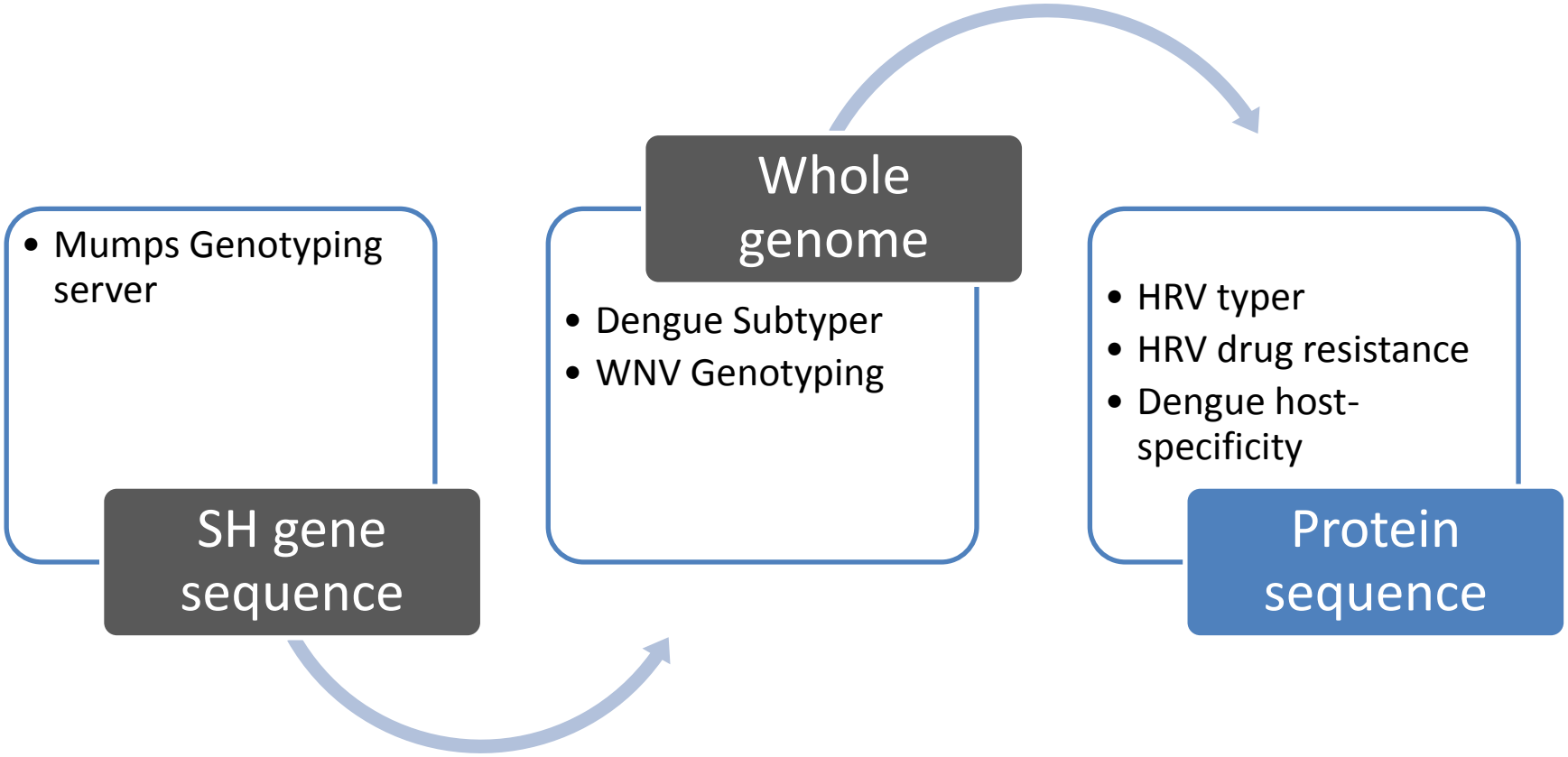


- Input : Whole genome
- Optimum k=7
- Sensitivity : 100%
- Specificity : 100%

Kolekar et. Al., **Journal of Virological Methods** 2014. 198:41-55.

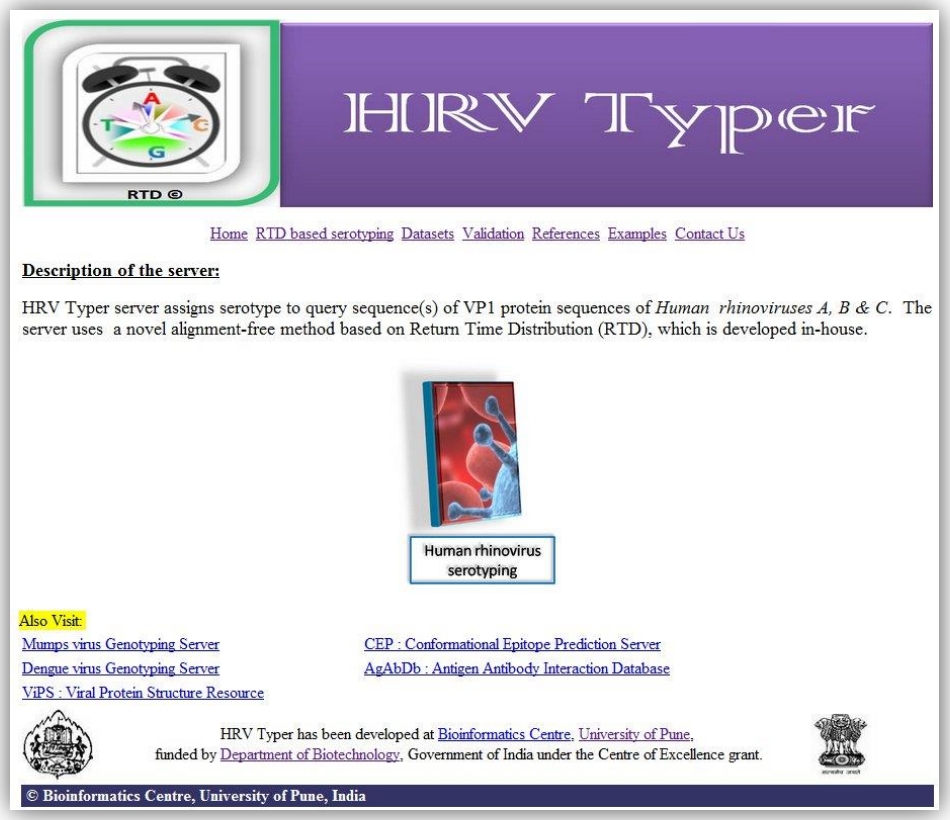
Available at: http://bioinfo.net.in/wnv/homepage.html

RTD for MPA, Genotyping, Serotyping, & **Clustering**



© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Genotyping of *Human rhinoviruses*



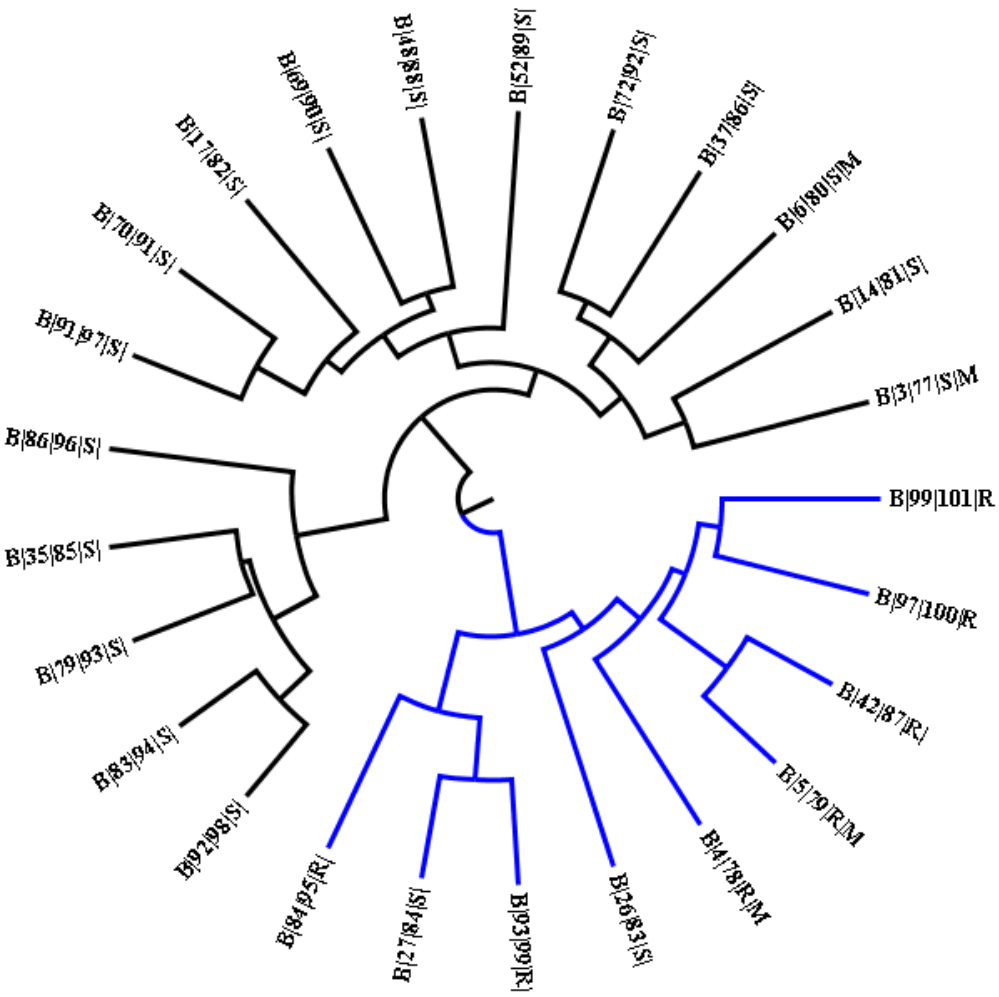
- Input : VP1 protein
- Optimum k=1
- Sensitivity : 100%
- Specificity : 100%

Manuscript under revision

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

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

RTD-based clustering of HRV-B using VP1 (k=1)
Clustering of drug-resistant & sensitive serotypes



HRV-B serotypes are subdivided into Pleconaril–sensitive and resistant serotypes (B-4, -5, -42, -84,-93, -97 and –84) serotypes.

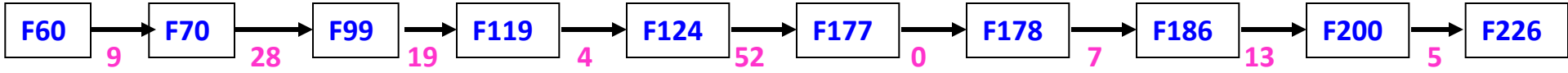
RTD-based informative residues viz. F,P,R,E,S,L,I obtained by WEKA-improves discrimination of pleconaril-sensitive & resistant serotypes.

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

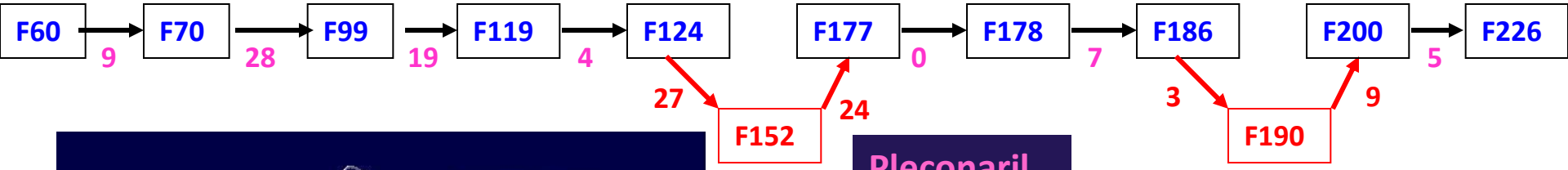
Computation of RTD for F residue and mapping on 3D structure of VP1

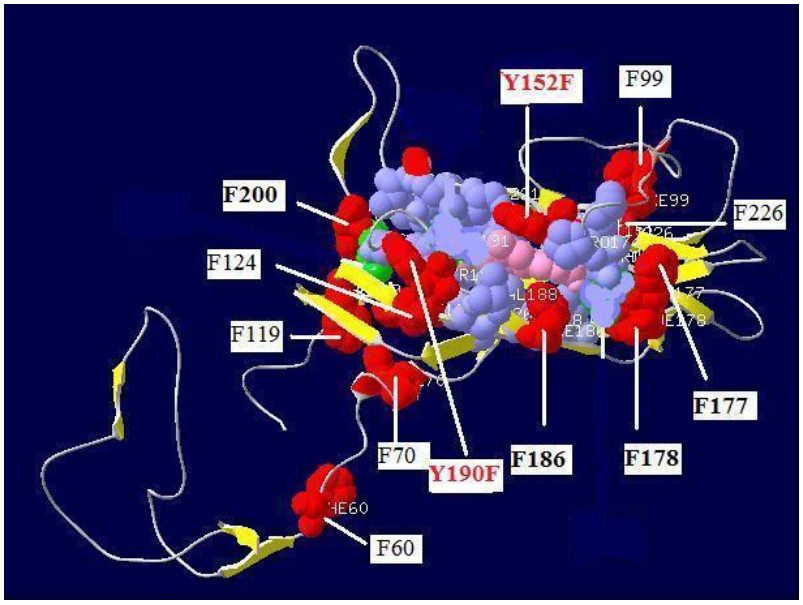
```
>sp|P03303|568-856
GLGDELEEVIVEKTKQTVASISSGPKHTQKVPILTANETGATMPVLPSDSIETRTTYMHFNGSETDVECFLGRAACVHVT
EIQNKDATGIDNHREAKLFNDWKINLSSLVQLRKKLELFTYVRFDSEYTILATASQPDSANYSSNLVVQAMYVPPGAPNP
KEWDDYTWQSASNPSVFFKVGDTSRFSVPYVGLASAYNCFYDGYSHDDAETQYGITVLNHMGSMAFRIVNEHDEHKTLVK
IRVYHRAKHVEAWIPRAPRALPYTSIGRTNYPKNTEPVIKKRKGDIKSY
```

RTD of F residue in pleconaril-sensitive serotype



RTD of F residue in pleconaril-resistant serotype





Pleconaril

Drug-binding site

Figure shows mapping of F residues on the VP1 structure of HRVB-14 serotype [1NCQA].
Phe (F) residues are localized at and near drug (pleconaril)- binding site are shown.

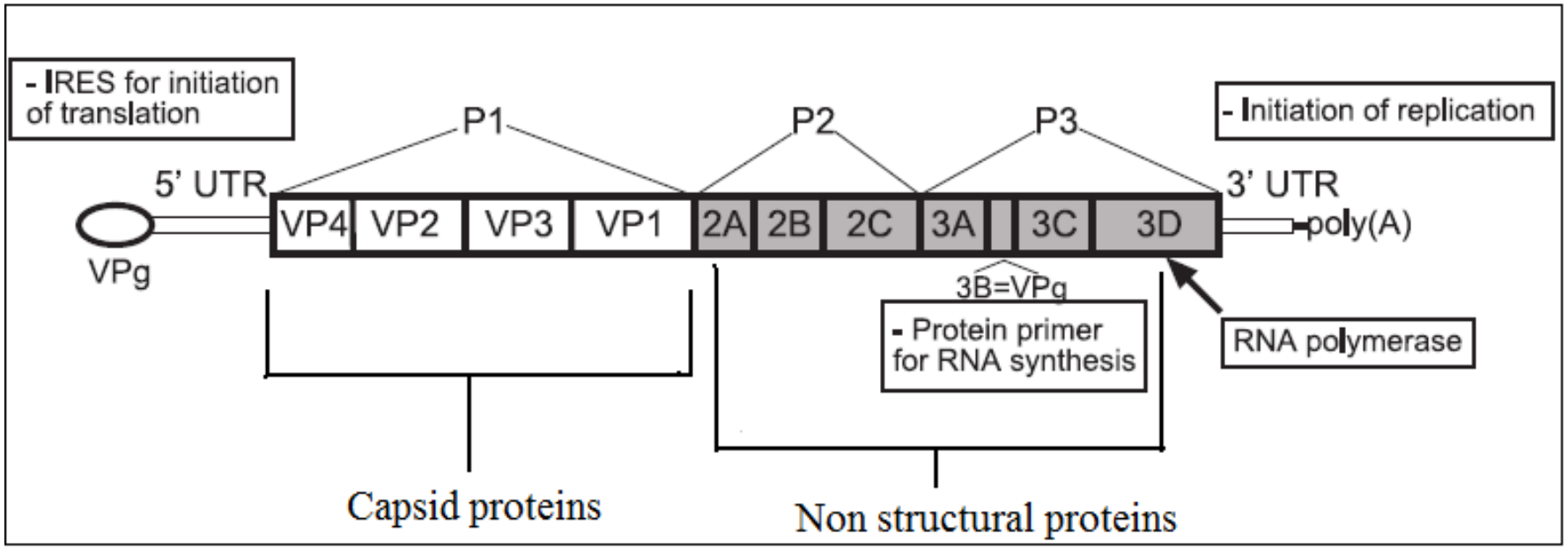
© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Population genomics: Rhinoviruses

Genome organization

Genome Characteristics:

- The genome contains a 5'-UTR, an open reading frame and a 3'-UTR.
- Genome encodes 4 structural and 7 non-structural proteins.



Structural proteins:**VP1-VP4**

Non-structural proteins: **2A**(proteinase: cleaves P1/P2 junction, shutoff of cap-dependent translation), **2B, 2C & 3A** (vesicle formation & negative strand synthesis),**3B** VPg (primer for 3D polymerase), **3C** (proteinase), **3D** (RNA-dependent RNA polymerase)

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Materials: Software(s)/program(s)/server(s) used

Multiple sequence alignment	• **MUSCLE program** in **MEGA 5.05** • **GUIDANCE server:** confidence scores for alignments (Penn et al., 2010).
Inference of genetically distinct clusters	• **STRUCTURE 2.3.3** • **LIAN 3.5**
Recombination & selection pressure analysis	• **Recombination: RDP4** • **Selection pressure: Site methods: SLAC,FEL, IFEL; Branch-site methods: MEME, BSR**

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Results: HRV clusters obtained at k=7

7 distinct lineages includes:

- HRV-B(Magenta)
- 4 sublevel subpopulations within HRV-A viz. pure A (blue), A1(yellow),A2 (red) ,A3 (green):
- A3 represents newly proposed HRV-D (subpopulation A3).
- 2 sublevel subpopulations within HRV-C viz. HRV-C1 (Orange) & HRV-C2 (Cyan).

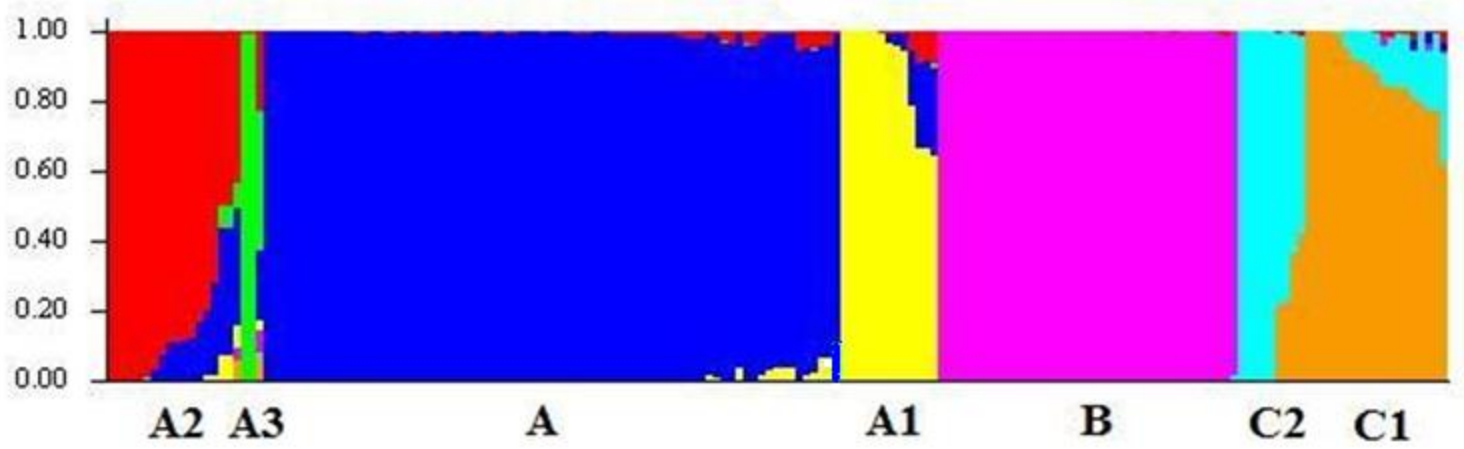


Figure 2 - Seven clusters of *Rhinoviruses* obtained by Bayesian-based approach using admixture model at K=7. The A1, A2, A3, C1 and C2 show the admixed individuals. They are color coded based on the proportion of membership scores with respective sub-populations. Waman et al., 2014

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Results: Key highlights

Genetic diversity using STRUCTURE program	• ***Rhinovirus*** **species is an ensemble of seven genetically distinct lineages** • **HRV-A : four lineages, HRV-C: two lineages, HRV-B is homogeneous**
Evidence of recombination	• **Intra-species recombination is prominent in HRV-A and –C and lead to diversification.** • **Inter-species recombination is limited to HRV-C members**
Evidence of episodic positive selection	• **Episodic positive selection was detected and corroborates with the antigenicity.** • **It was found responsible for emergence of new lineages in HRV-A**

Waman et al., 2014

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Post-genomic Rational Vaccine Design

- Perform genome-based comparisons
- Genotype and/or Serotype populations
- Study viral population for emergence of new subtypes
- Map epitopes & mutations on 3D structures
- Prioritize candidate vaccine

© Dr. Urmila Kulkarni-Kale, Vaccines 2015
Bioinformatics Centre, S.P. Pune University

Publications

- Waman VP, Kolekar PS, Kale MM, Kulkarni-Kale U (2014) Population Structure and Evolution of Rhinoviruses. PLoS ONE 9(2): e88981. doi:10.1371/journal.pone.0088981
- Kolekar P, Hake N, Kale M, Kulkarni-Kale U, WNV Typer: A server for genotyping of West Nile viruses using an alignment-free method based on a return time distribution. Journal of Virological Methods 2014. 198:41-55.
- Kolekar P, Kale M, Kulkarni–Kale U. Alignment–free distance measure based on return time distribution for sequence analysis: applications to clustering, molecular phylogeny and subtyping. Molecular Phylogenetice & Evolution. 2012 Nov;65(2):510–22.
- Kulkarni-Kale, U, Waman, V., Raskar, S, Mehta, S, & Saxena, S (2012) Genome to vaccinome: role of bioinformatics, immunoinformatics & comparative genomics. Current Bioinformatics, 7(4), 454-466.
- Kolekar PS, Kale M, Kulkarni–Kale U. Genotyping of Mumps viruses based on SH gene: Development of a server using alignment–free and alignment–based methods. Immunome Research. 2011 Nov 30;7(3):1–7.
- Kolekar, P. S., Kale M. M. and Kulkarni-Kale, U., "'Inter-Arrival Time' Inspired Algorithm and its Application in Clustering and Molecular Phylogeny", AIP Conference Proceedings (2010). 1298(1):307-312. ISBN 978-0-7354-0854-8. [Conference proceedings]
- Kolekar, P. S., Kale M. M. and Kulkarni-Kale, U., (2011). Molecular Evolution & Phylogeny: What, When, Why & How?, Computational Biology and Applied Bioinformatics, Heitor Silverio Lopes and Leonardo Magalhães Cruz (Ed.), ISBN: 978-953-307-629-4, InTech Publishers. [Book Chapter]

Acknowledgements

PhD Students:

- Mr. Pandurang Kolekar
- Ms. Vaishali Waman
- Ms. Sunitha Manjari (CDAC)

Collaborators:

- Dr. Mohan Kale, Statistics Dept., SPPU
- Dr. Elin Kure, Radium Hospital, Oslo, Norway
- Dr. Sangeeta Sawant, Bioinformatics Centre, SPPU

Funding:

- CoE: Dept. of Biotechnology (DBT), Govt. of India (GoI)
- CoE: Dept. of Electronics & Information Technology (DeitY), MCIT, GoI
- INCP: Indo Norwegian Collaboration Program
- UGC UPE Phase II
- DST PURSE program
- DBT-BINC & DBT-BET fellowship programs