

Virus Bioinformatics | Let the data speak |

Dr. Urmila Kulkarni-Kale

S. P. Pune University | University of Southeastern Norway | Citadel Precision Medicine

urmila.Kulkarni.kale@gmail.com

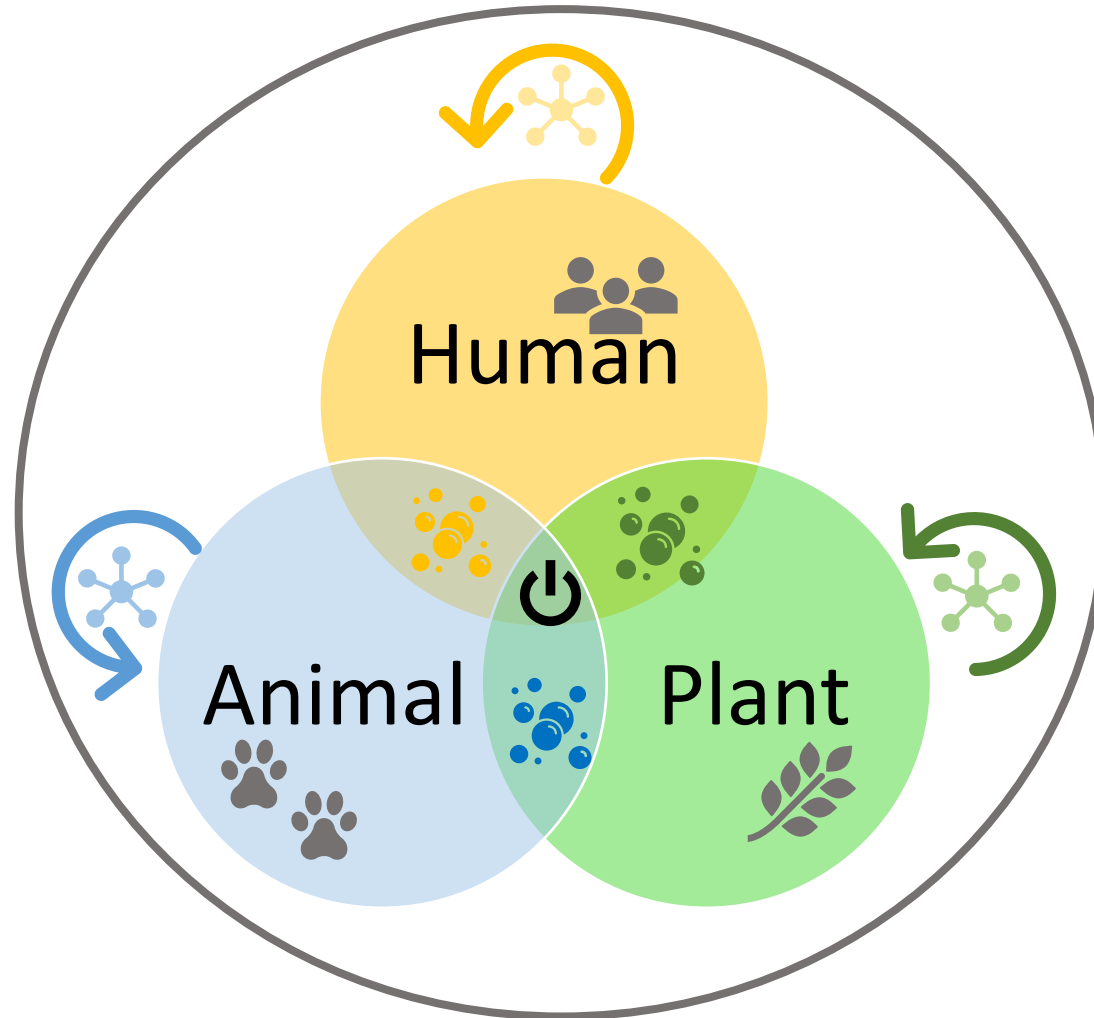
Schedule for the day

 wellcome connecting science		Viral Genomics and Bioinformatics 12 - 17 May 2024, OUCRU, Ho Chi Minh City, Vietnam					
Sunday May 12		Monday May 13	Tuesday May 14	Wednesday May 15	Thursday May 16	Friday May 17	
8:00		Bus to venue	Bus to venue	Bus to venue	Bus to venue	Bus to venue	8:00
8:30		Sequencing demonstration Tan's Team	NGS File Formats & QC Sreenu	Long Read Data Analysis Richard	Metagenomics and Virus Discovery Richard	Build Your Own Pipeline: Reference Mapping	8:30
9:00			Reference Mapping Richard			Build Your Own Pipeline: De novo Assembly	9:00
9:30							9:30
10:00			10:00				
10:30		Break	Break	Break	Break	Break	10:30
11:00		Sequencing demonstration Tan's Team	Reference Mapping cont.	De novo Assembly Sreenu	Introduction to MSA Urmila	Build Your Own Pipeline: Metagenomics	11:00
11:30			Statistics and Visualisation Richard		Tree Building and Visualisation Urmila	Build Your Own Pipeline: Phylogenetics	11:30
12:00	12:00						
12:30						12:30	
13:00	Registration and Lunch	Lunch	Lunch	Course Photograph	Lunch	Lunch	13:00
13:30				Lunch			13:30
14:00	Introductions - WCS, Training Team, Participants	Seminar + short break	Seminar + short break	Seminar + short break	Seminar + short break	Group Presentations	14:00
14:30		Intro to Linux Srikeerthana	Consensus and Variant Calling Richard	De novo Assembly Validation & Statistics Sreenu	Recombination Analysis Urmila		14:30
15:00							15:00
15:30	Break						15:30
16:00	Advanced Learning Skills	Break	Break	Break	Break	Wrap Up	
16:30		BASH Scripting Sreenu	Consensu and Variant Calling cont.	Scaffolding & Gap filling Sreenu	Recombination Analysis cont.	Bus to hotel	16:30
17:00			Group Practical Reference Mapping	Group Practical De novo Assembly	Group Practical Phylogenetics		17:00
17:30							17:30
18:00	Networking and Dinner	Q&A	Q&A	Q&A	Q&A		18:00
18:30		Bus to hotel	Bus to hotel	Bus to hotel	Bus to hotel		18:30
19:00							19:00
		Monday May 13	Tuesday May 14	Wednesday May 15	Thursday May 16	Friday May 17	

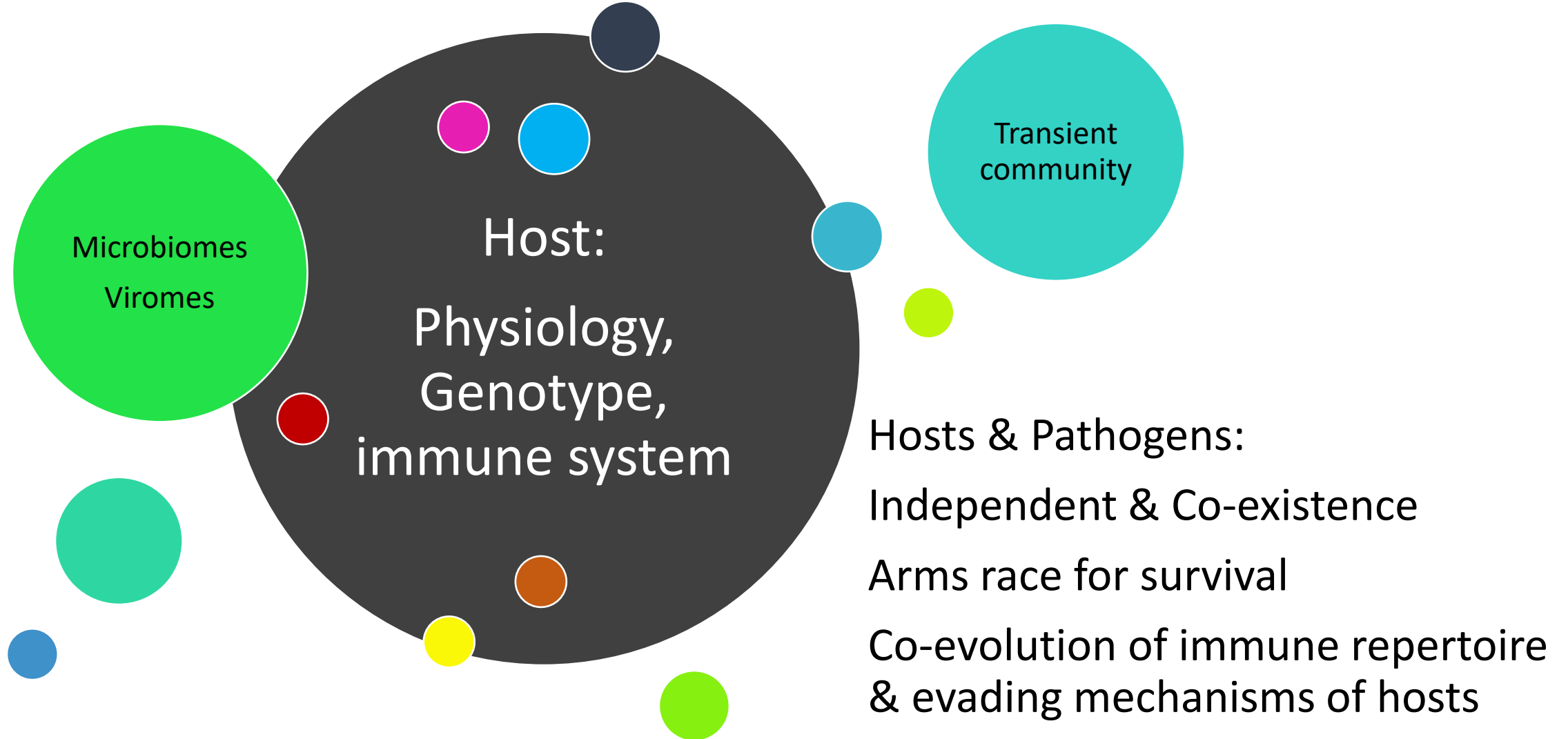
Sequenced a virus/virome?
Assembled a genome/Metagenome?

What next?

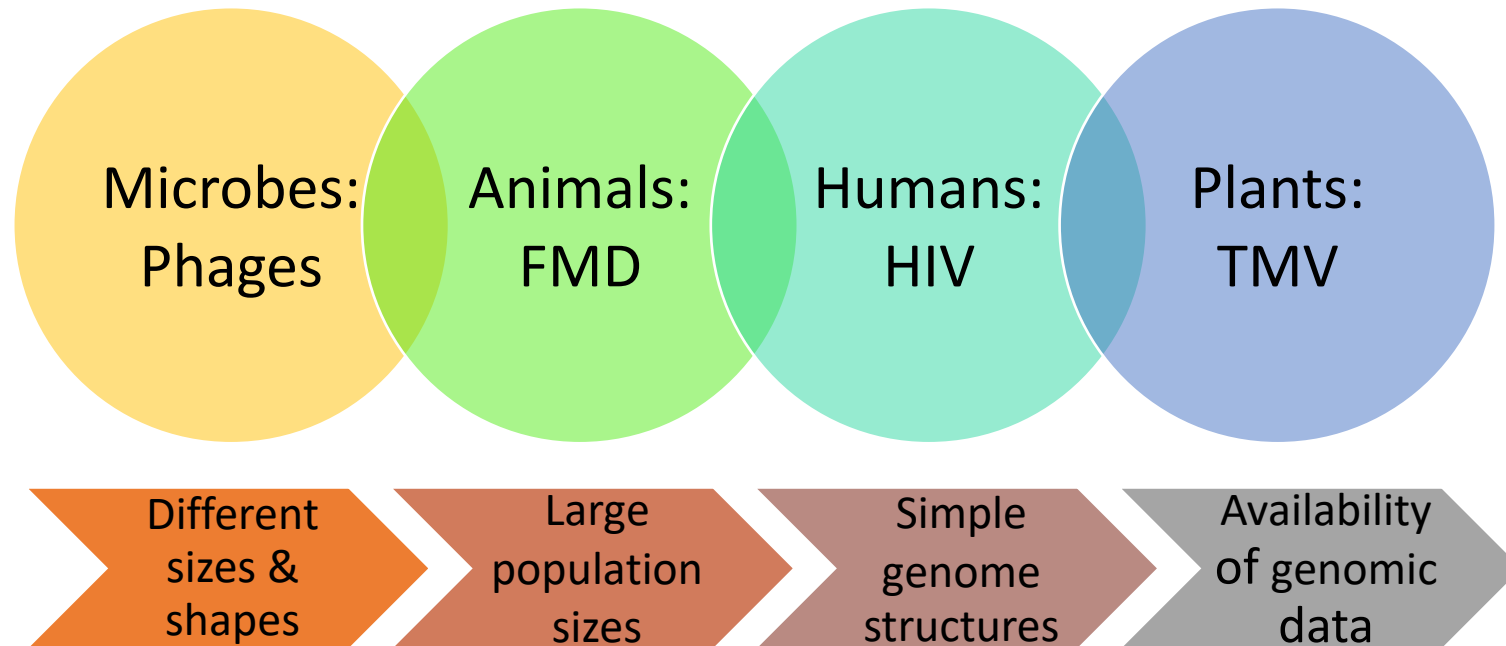
Concept of Health → One Health



The Multiverse / Parallel Universe



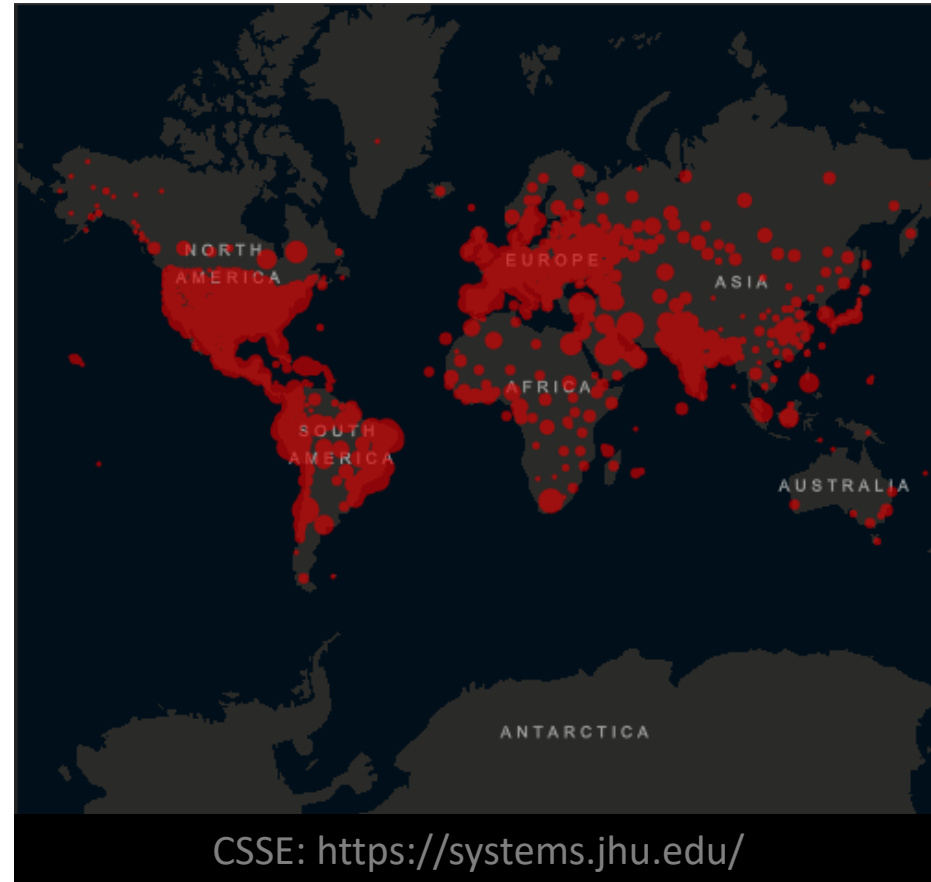
Virus Bioinformatics



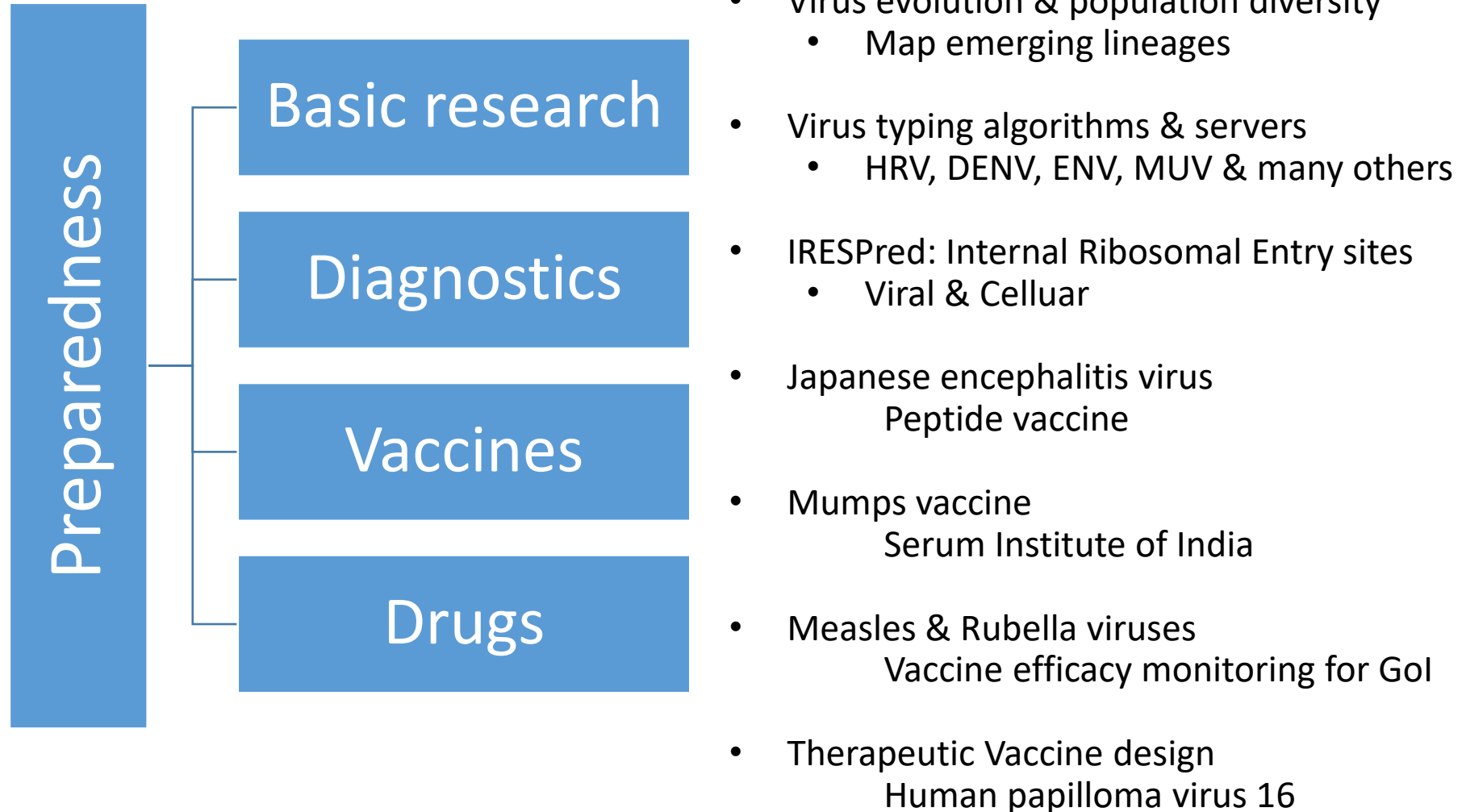
The unknown of known

How do viruses –

- infect hosts?
- cause disease?
- jump species?
- spread & evolve?
- escape the immune system?
- develop drug resistance?

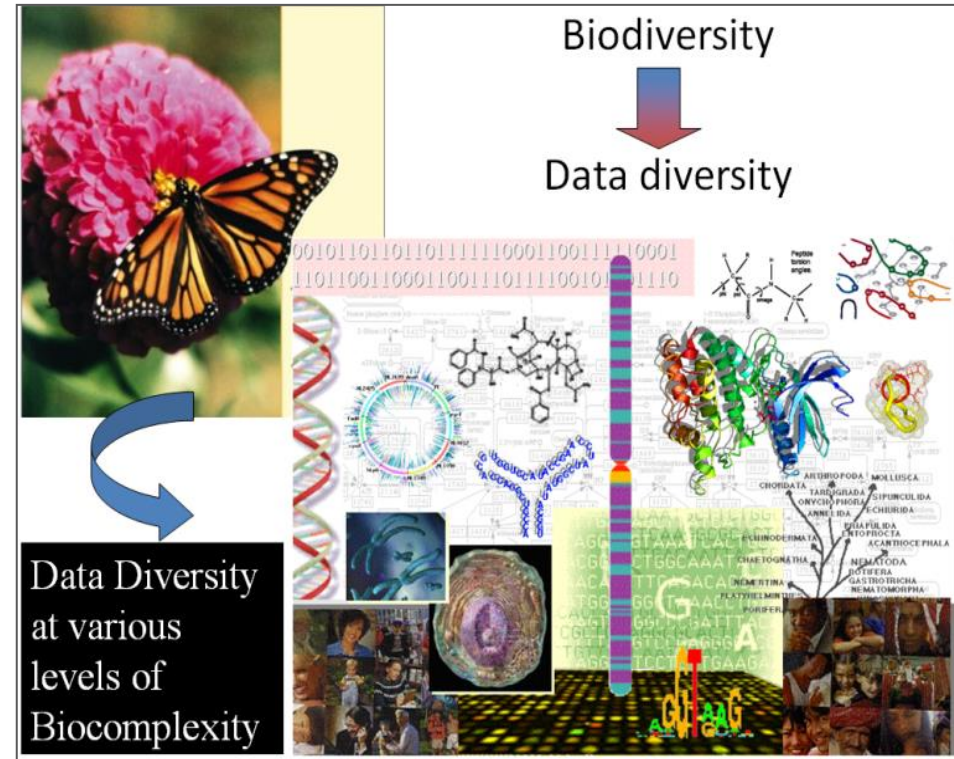
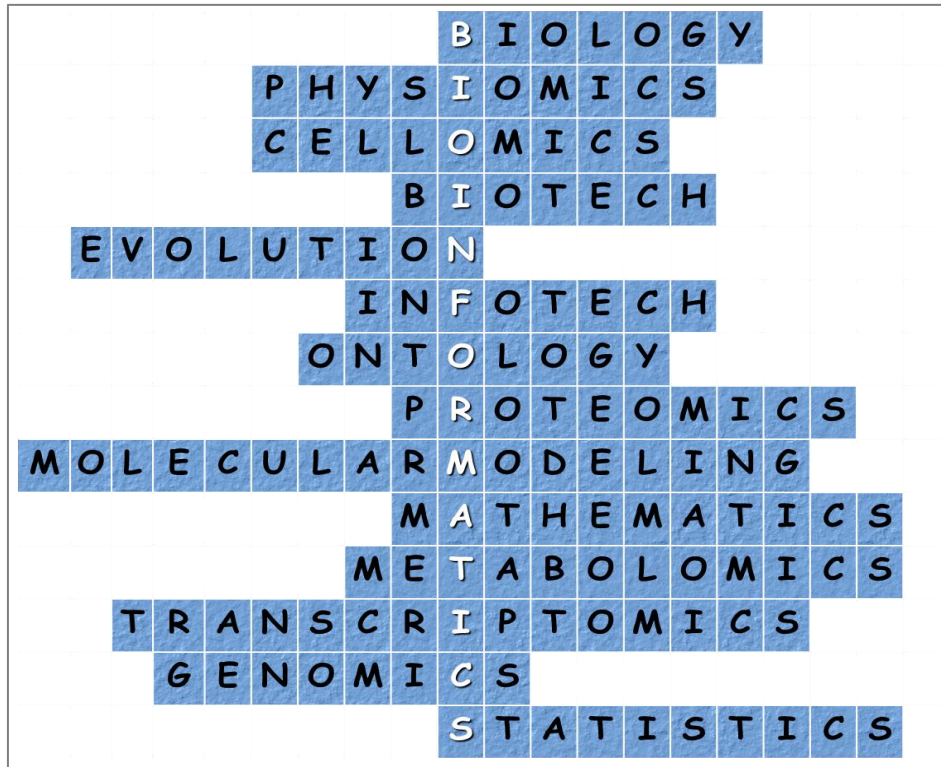


How to combat/control?



Bioinformatics: Bridges many disciplines

Data-driven approach for knowledge discovery in Biology



Using known (data) to explore unknown

- Genome sequences
- Protein sequences
- 3D structures
- Omics data
- Virome data

Big data:

Volume. Velocity. Variety. Veracity. Value.

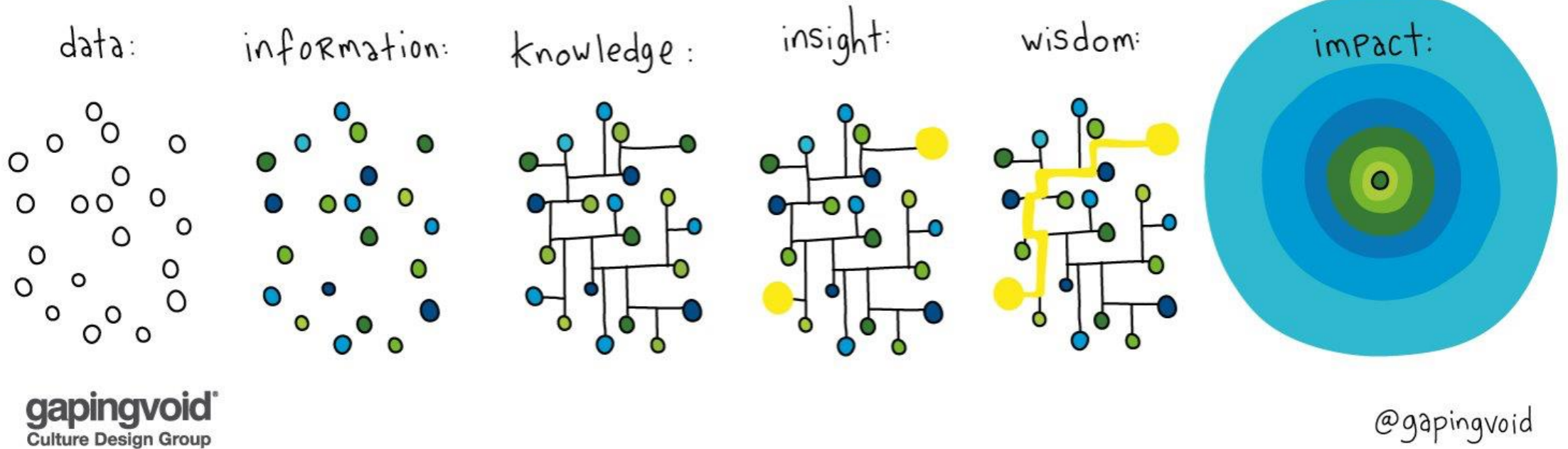
FAIR principles:

Findable. Accessible. Interoperable. Reusable.

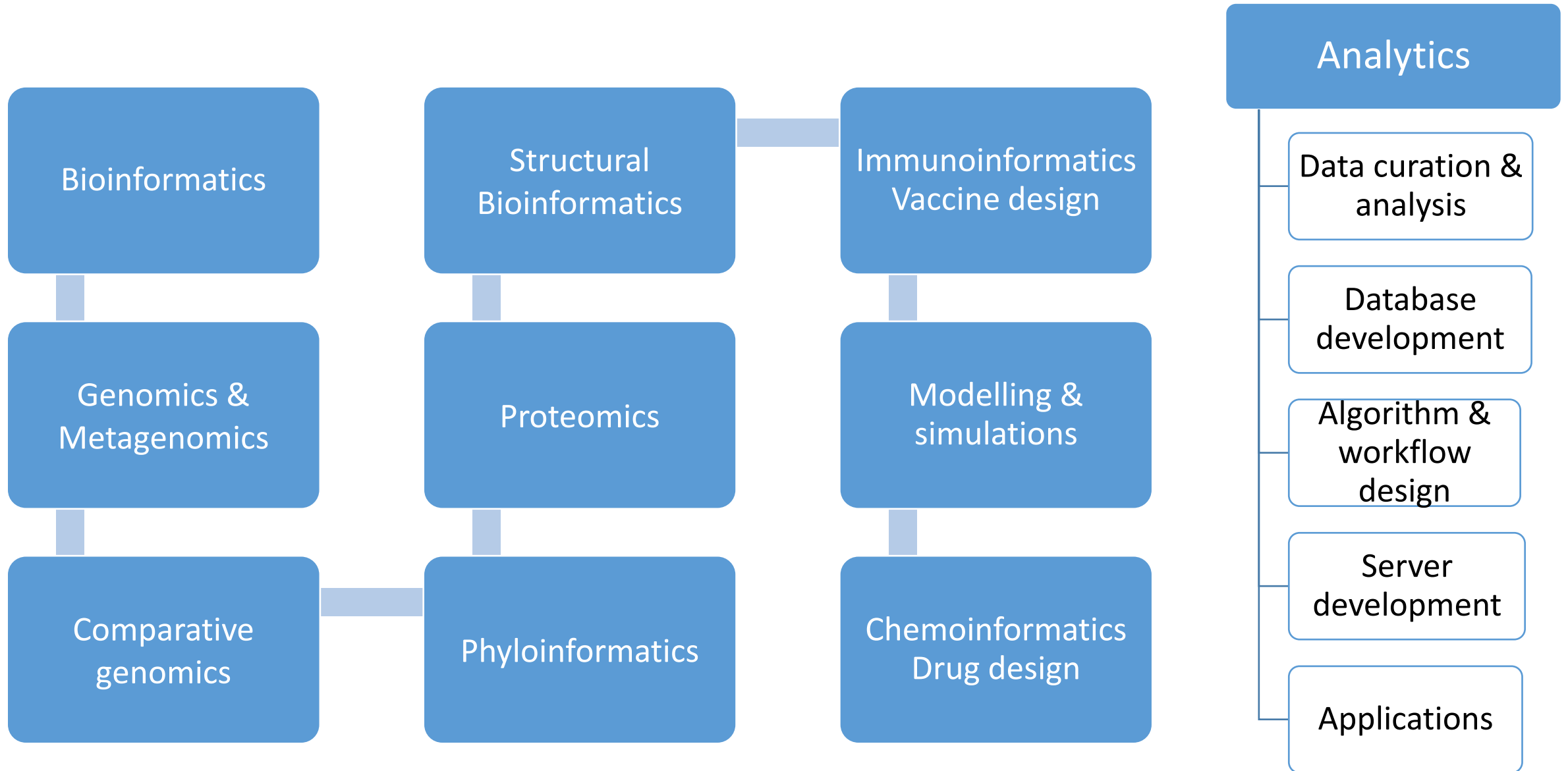


Virus bioinformatics: data led discovery framework

Reimagined using illustrations & philosophy of gapingvoid, a culture design group

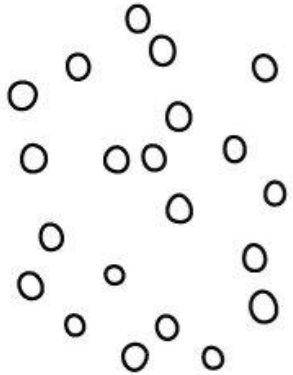


Viral Bioinformatics: Scope & Methods



Data: compilation & curation: HPV16 story

data:



gapingvoid
Culture Design Group

Human papillomavirus type 16 (HPV16), complete genome

GenBank **K02718.1**

[FASTA](#) [Graphics](#)

Go to: ☐

LOCUS PPH16 7904 bp DNA circular VRL 18-MAR-1994
DEFINITION Human papillomavirus type 16 (HPV16), complete genome.
ACCESSION K02718
VERSION K02718.1
KEYWORDS circular; complete genome.
SOURCE Human papillomavirus type 16
ORGANISM [Human papillomavirus type 16](#)

Prototype entry: issues

- Deletion of CAT6903 and Insertion of GAT6952 in L1 gene of prototype.
- First 14 missing nucleotides in E5 gene of prototype led to frame shift during translation

Parton A. Nucleotide sequence of the HPV16 L1 open reading frame. Nucleic Acids Res. 1990 Jun 25;18(12):3631. doi: 10.1093/nar/18.12.3631.

Human papillomavirus type 16, complete genome

NCBI Reference Sequence **NC_001526.4**

[FASTA](#) [Graphics](#)

Go to: ☐

LOCUS NC_001526 7906 bp DNA circular VRL 13-AUG-2018
DEFINITION Human papillomavirus type 16, complete genome.
ACCESSION NC_001526
VERSION NC_001526.4
DBLINK BioProject: [PRJNA485481](#)
KEYWORDS RefSeq.
SOURCE Human papillomavirus type 16
ORGANISM [Human papillomavirus type 16](#)

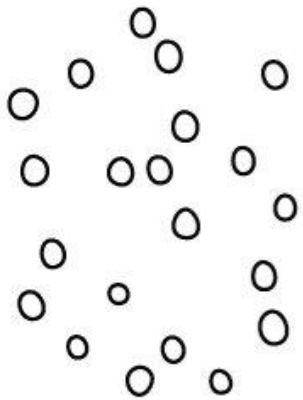
Refseq entry: solution

- Insertion & deletion in prototype led to wrong reporting of mutations
- Proper demarcation of E5 gene boundaries in Refseq led to fixing Frameshift errors (stop codons)

Mane A, Patil L, **Limaye S**, Nirmalkar A, **Kulkarni-Kale U**. 2020. **J Med Virol**. 2020:1-6

Data compilation & curation: BHV1 story

data:



gapingvoid
Culture Design Group

Bovine herpesvirus 1, complete genome

NCBI Reference Sequence: [NC_001847.1](#)

[FASTA](#) [Graphics](#)

Go to: ☐

LOCUS NC_001847 135301 bp DNA linear VRL 13-AUG-2018
DEFINITION Bovine herpesvirus 1, complete genome.
ACCESSION NC_001847
VERSION NC_001847.1
DBLINK BioProject: [PRJNA485481](#)
KEYWORDS RefSeq.
SOURCE Bovine alphaherpesvirus 1 (Infectious bovine rhinotracheitis virus)
ORGANISM [Bovine alphaherpesvirus 1](#)

Refseq entry: issues

- Refseq entry derived from cooper strain (BHV-1.1: a composite of 5 strains - BHV-1.1 and BHV-1.2)
- Spontaneous deletion of US1.67/US2 in genomic region of cooper strain (Campos et, 2016)

Bovine herpesvirus type 1.1 isolate NVSL challenge 97-11, complete

GenBank: [JX898220.1](#)

[FASTA](#) [Graphics](#)

LOCUS JX898220 134896 bp DNA linear VRL 06-MAY-2013
DEFINITION Bovine herpesvirus type 1.1 isolate NVSL challenge 97-11, complete genome.
ACCESSION JX898220
VERSION JX898220.1
KEYWORDS .
SOURCE Bovine herpesvirus type 1.1
ORGANISM [Bovine herpesvirus type 1.1](#)

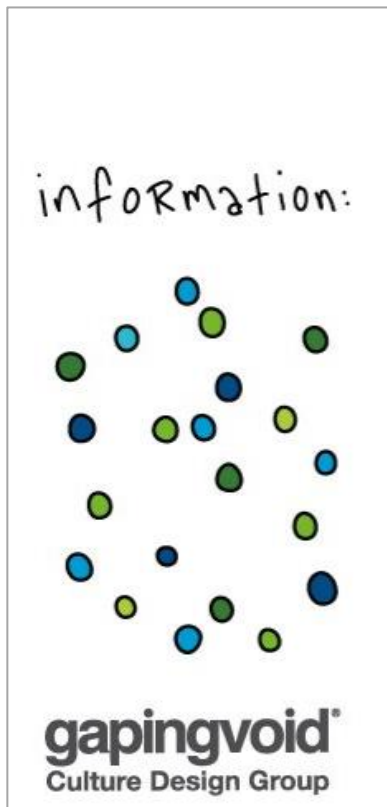
Prototype entry: solution

- Pure BHV-1.1 strain that could be used as reference for proper clustering of sequences into BHV-1.1 and BHV-1.2 sub-types
- No deletion of genomic region

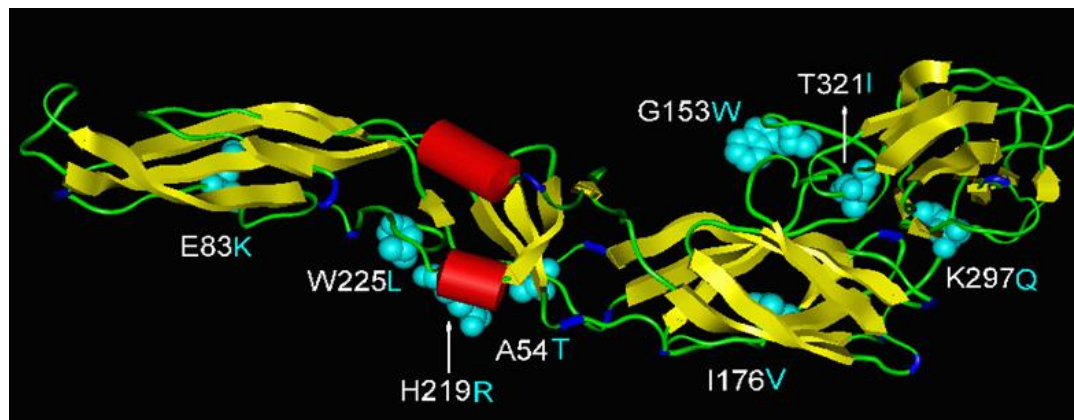
Ongoing project in collaboration with Dr. Suresh Kuchipudi, University of Pittsburgh, USA

Data to Information

Perspectives from similarity (JEV story)



Species & Strain specific variations
Loop1 in TBEV: TLA**EE**HQGG
Loop1 in JEVN: **HNE**KRADSS
Loop1 in JEVS: **HN**KKRADSS



How similar is similar?
How different is different?

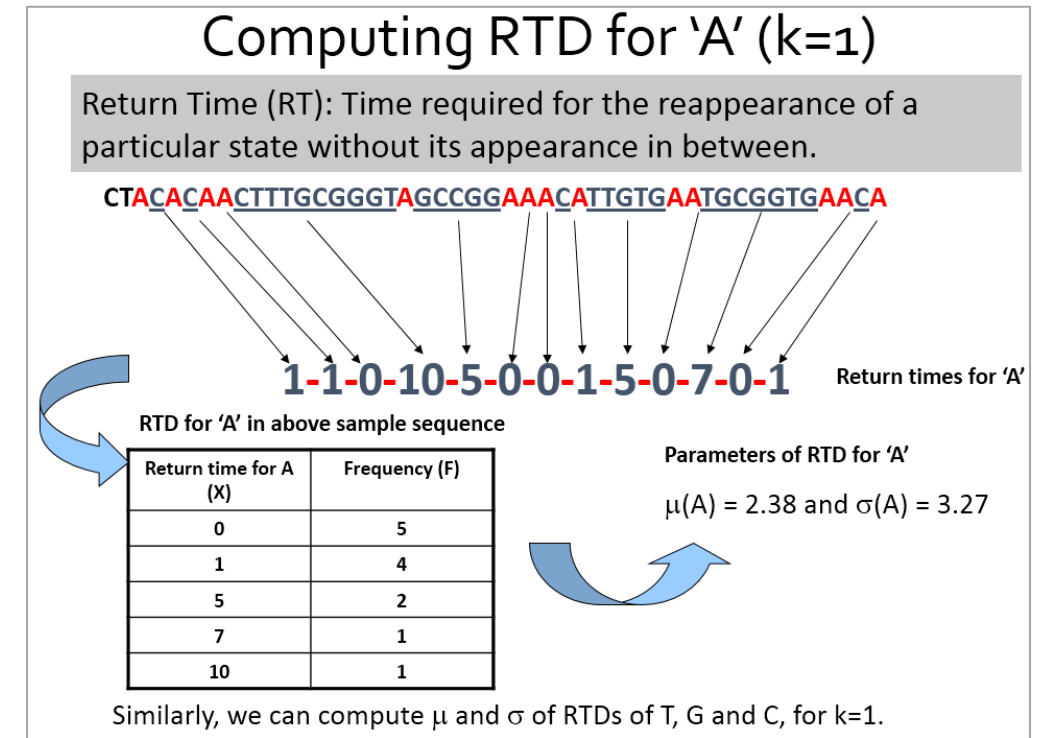
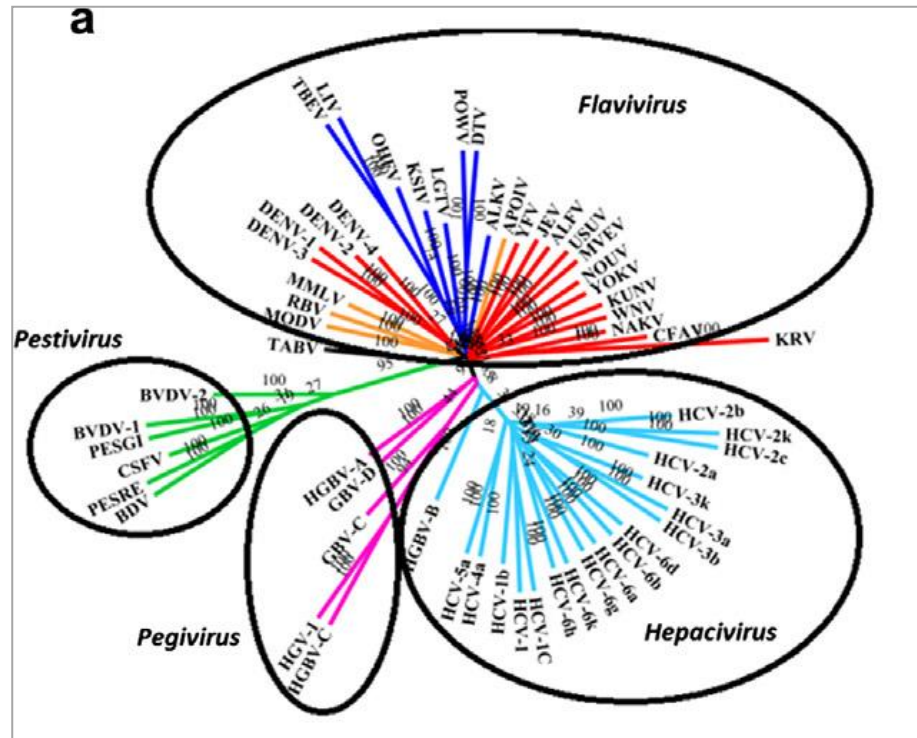
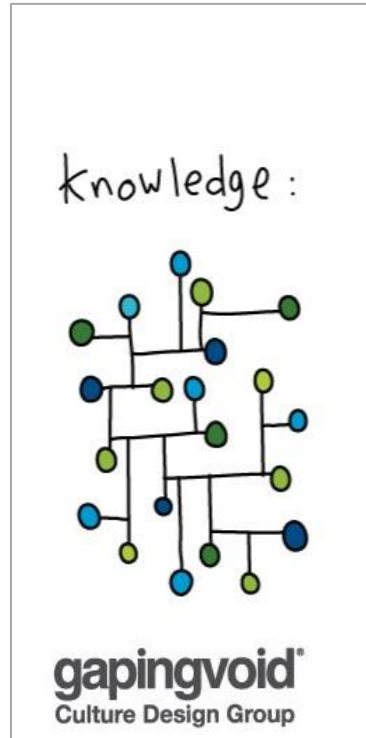
Biodiversity & Biocomplexity:

Isolates, strains, genotypes, serotypes, species & clades

Genus, subfamilies, families

Information to Knowledge

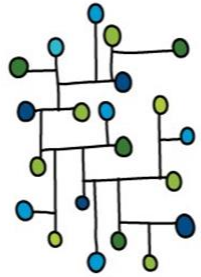
Connecting dots with evolutionary perspective



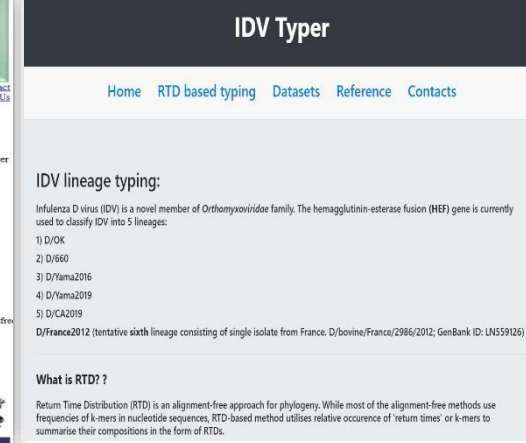
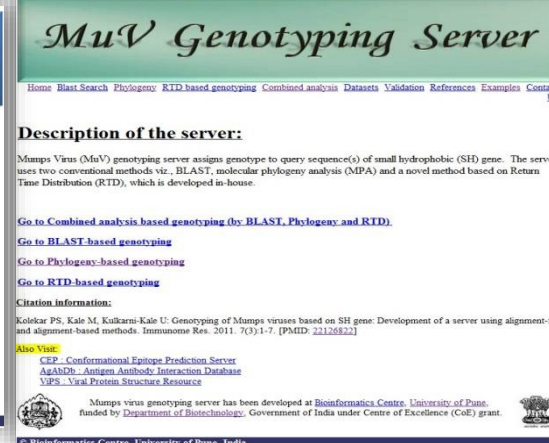
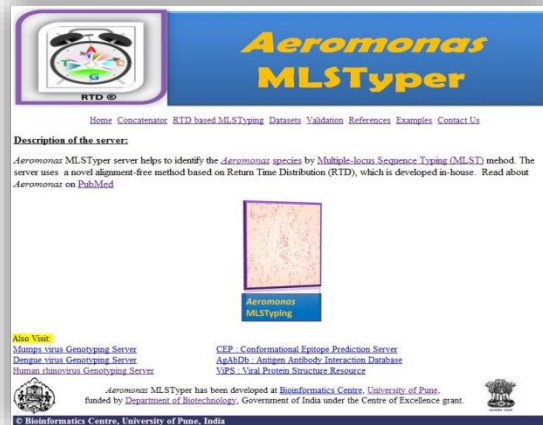
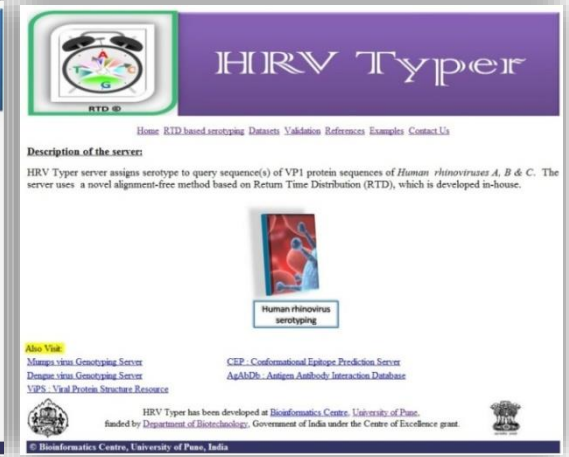
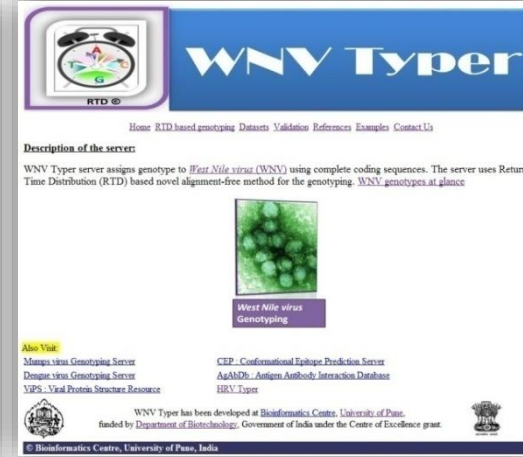
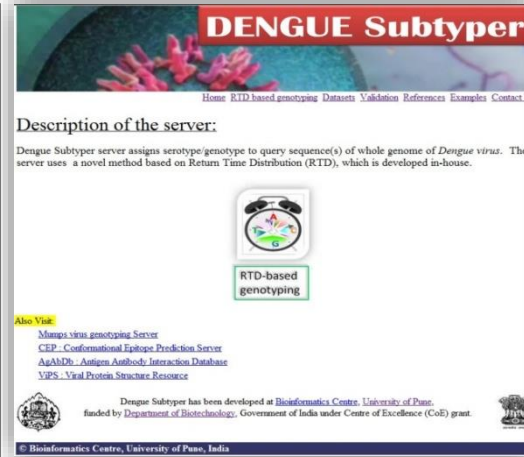
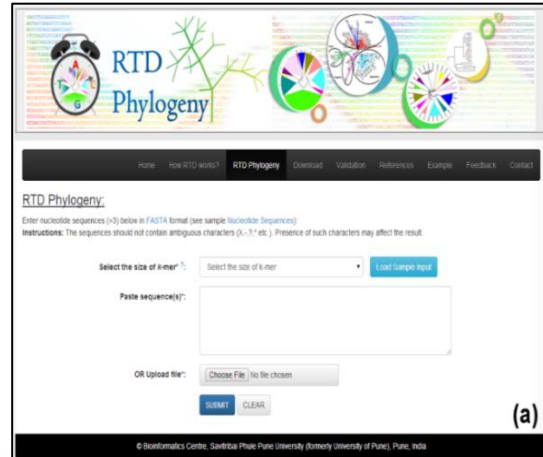
Kolekar P, Kale M, Kulkarni-Kale U. Alignment-free distance measure based on return time distribution for sequence analysis: Applications to clustering, molecular phylogeny and subtyping. *Molecular Phylogenetics Evolution* 2012; 65(2):510-22.

Information Knowledge: Using Data To Discover Algorithms| Servers | Products developed

Knowledge :



gapingvoid
Culture Design Group

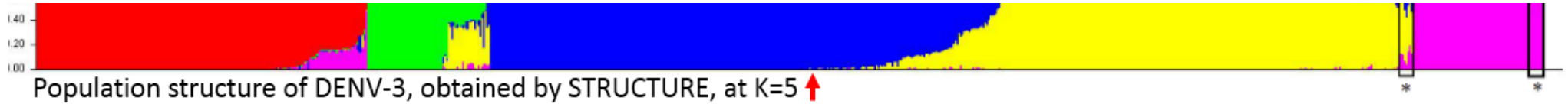
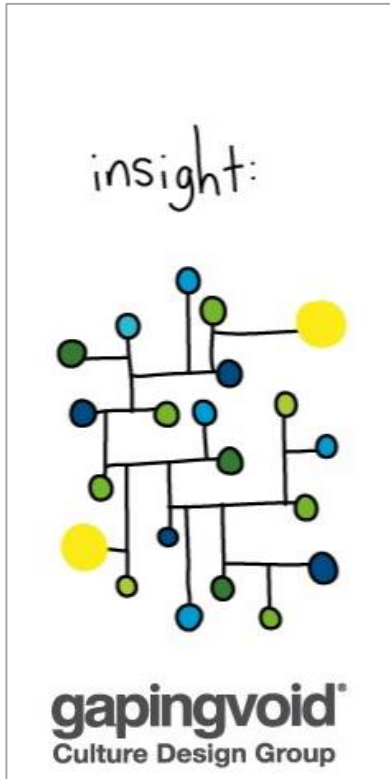


NEXT ONE?

Area: Phyloinformatics, Alignment-free, RTD-based Viral & bacterial typing (Gene/Genome/MLST)
Accuracy, Sensitivity & Specificity: 100% | Collaborator: Dr. Mohan Kale, Dept. of Statistics
Students: Dr. Pandurang Kolekar, Dr. Vaishali Waman, Sanket Limaye, Mr. Nilesh Hake, Mr. Anant Shelke

Knowledge to Insight

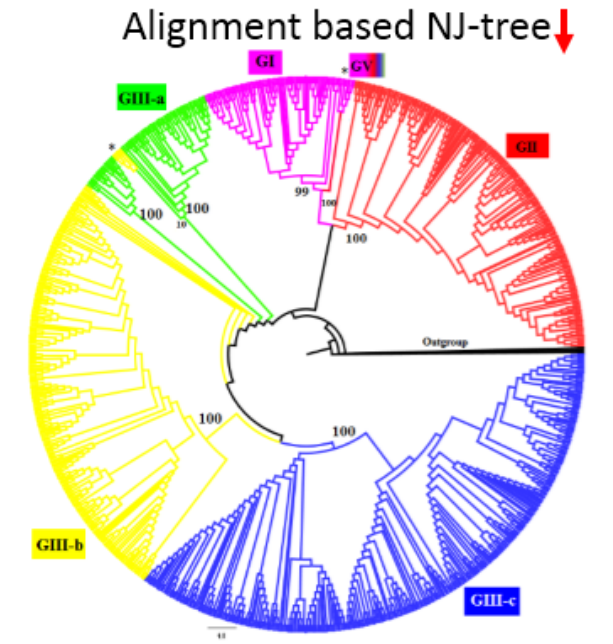
Population diversification perspectives (Denv-3 story)



Genetic structure of DENV-3

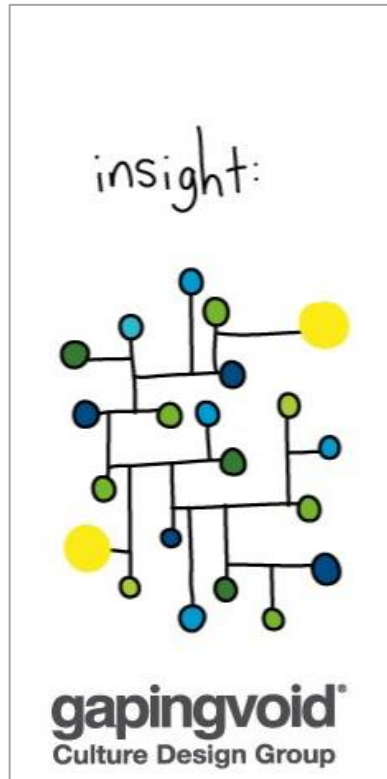
- Known genotypes: 5 (GI, GII, GIII, GIV, GV)
- Input data: 860 complete genomes
- Emerging lineages: GIII subdivided into **3 major subpopulations** (GI, GII, GIII-a, GIII-b & GIII-c).
- Admixed strains in GIII & GV (a reservoir for emerging subpopulations).
- Major factors shaping diversity are Inter-genotype recombination, migration and adaptive evolution

Waman VP, Kale MM, Kulkarni-Kale U. Infection Genetics & Evolution 2017; 49:234-240.

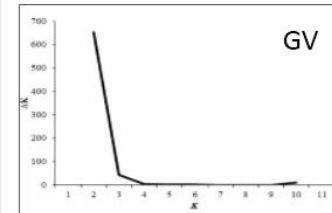
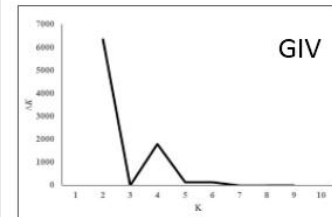
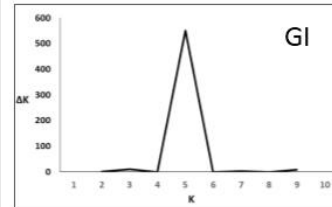


Knowledge to Insight

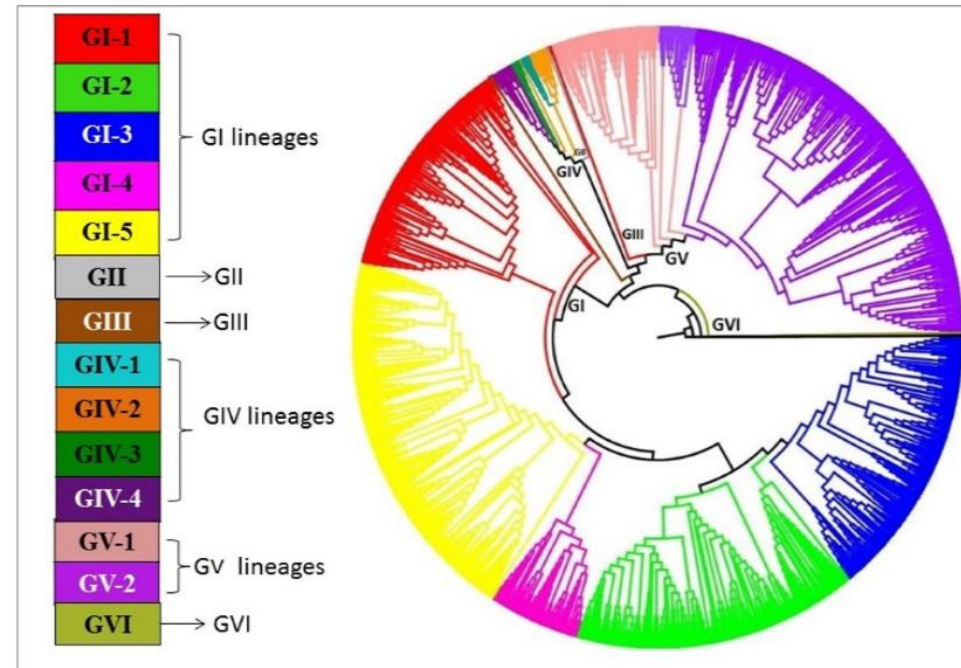
Population diversification perspectives (DENV-1 story)



Plots of K vs ΔK : obtained by STRUCURE program

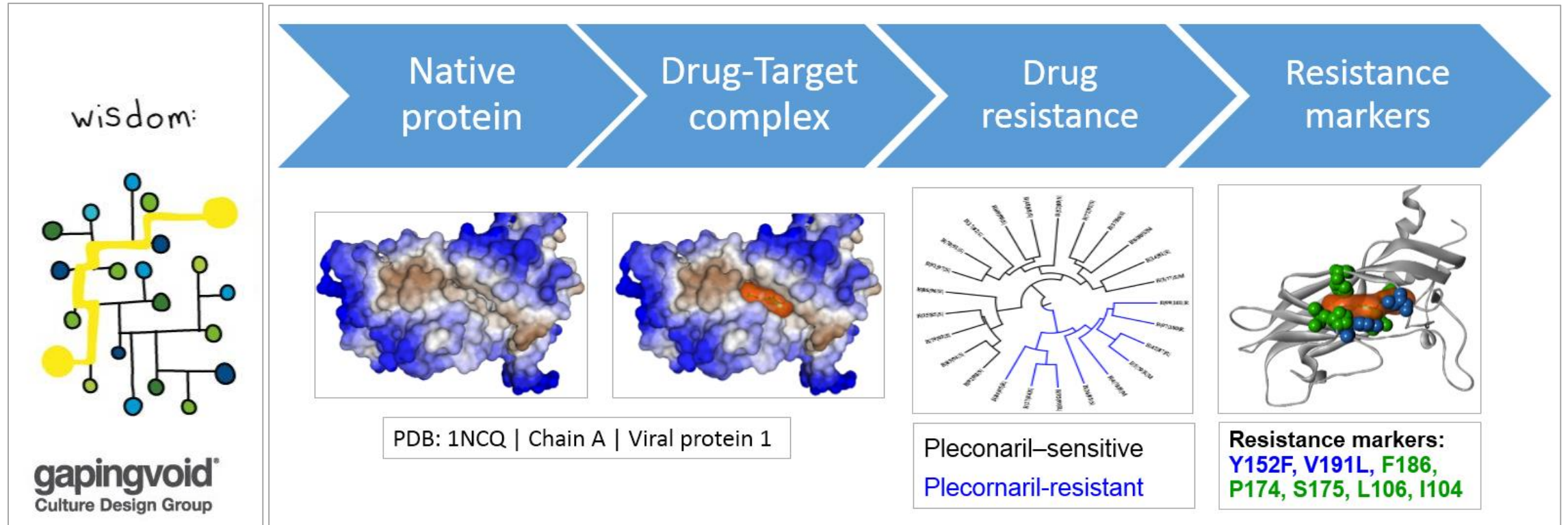


Complete genome-based NJ tree | Data 1670 strains of DENV-1



Insight to Wisdom

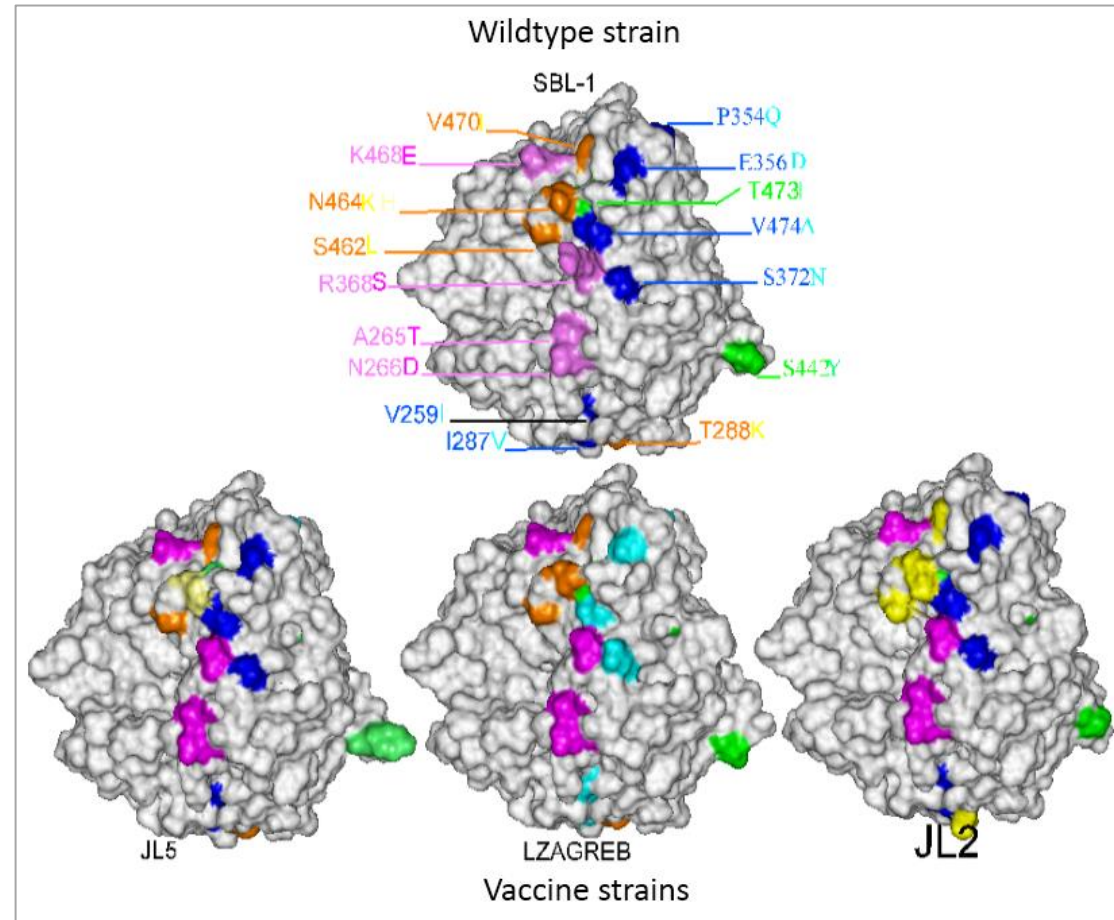
Tracing emergence of drug resistance in Rhinoviruses



Waman et al., Unpublished

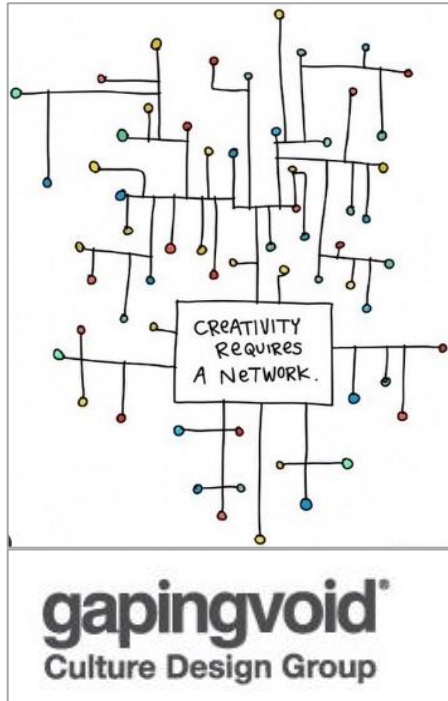
Wisdom to Impact

Societal benefit through translation (Mumps virus story)

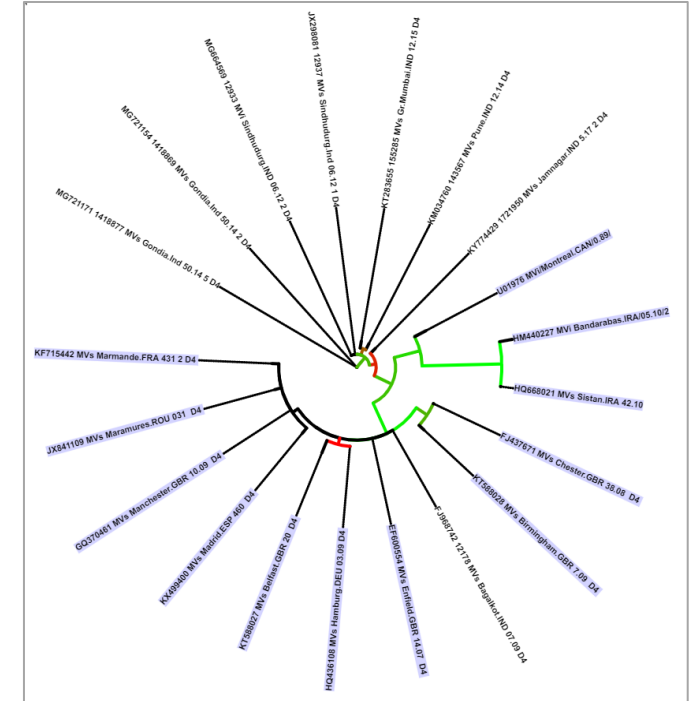


Kulkarni-Kale et al., 2007.
Funded by: Serum Institute of India

Making room for new named strains (Measles story)



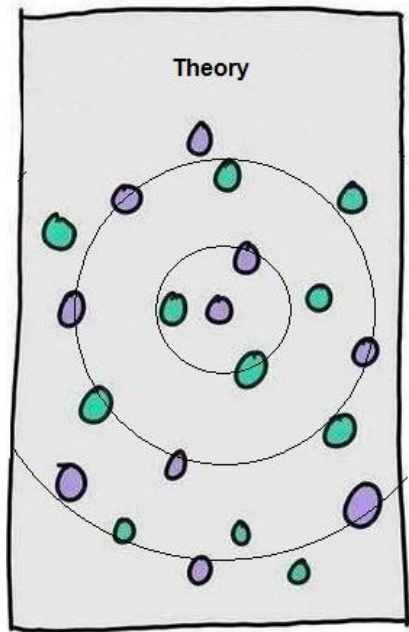
- Indian MeV genomes: 43
- Lineages: D4 & D8 (reported for the first time)
- Only 1 of 8 (Bagalkot.IND/07.09) isolates belongs to WHO named strain of D4 (Enfield lineage in circulation across Europe since 2007)
- Need for more named strain inclusion for understanding lineage diversity across globe.



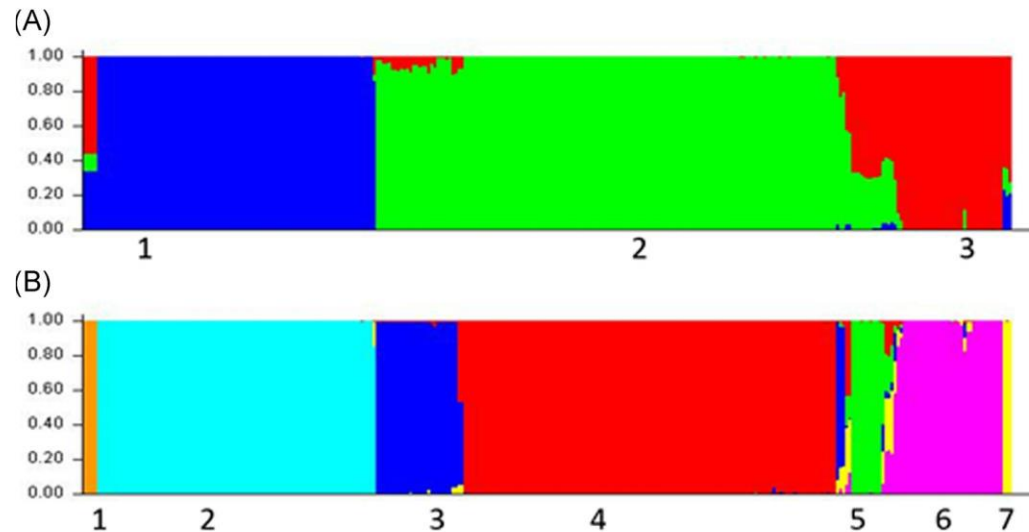
Vaidya SR, Kasibhatla SM, Bhattad DR, et al. of measles viruses in India: Genomic sequencing and comparative genomic Characterization of diversity s studies. *J Infect.* 2020;80(3):301-309. doi:10.1016/j.jinf.2019.11.025

Data to Theory via DIKIW pipeline

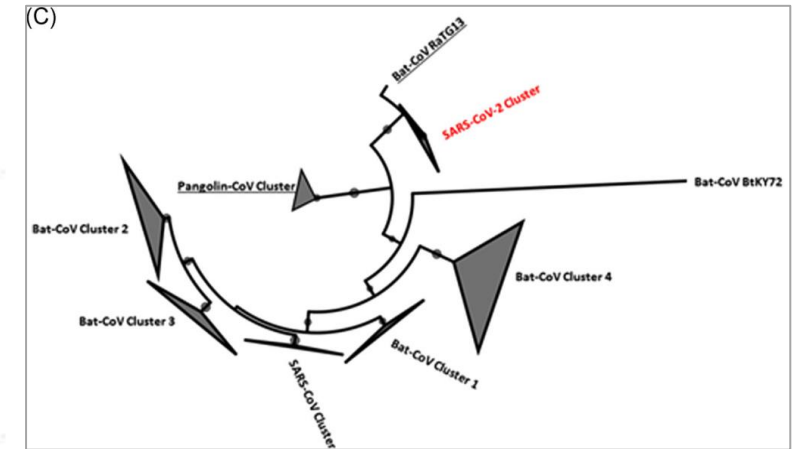
Demystifying emergence of COVID-19 (SARS-CoV-2 story)



Graphic inspired by
gapingvoid
Culture Design Group
Source internet



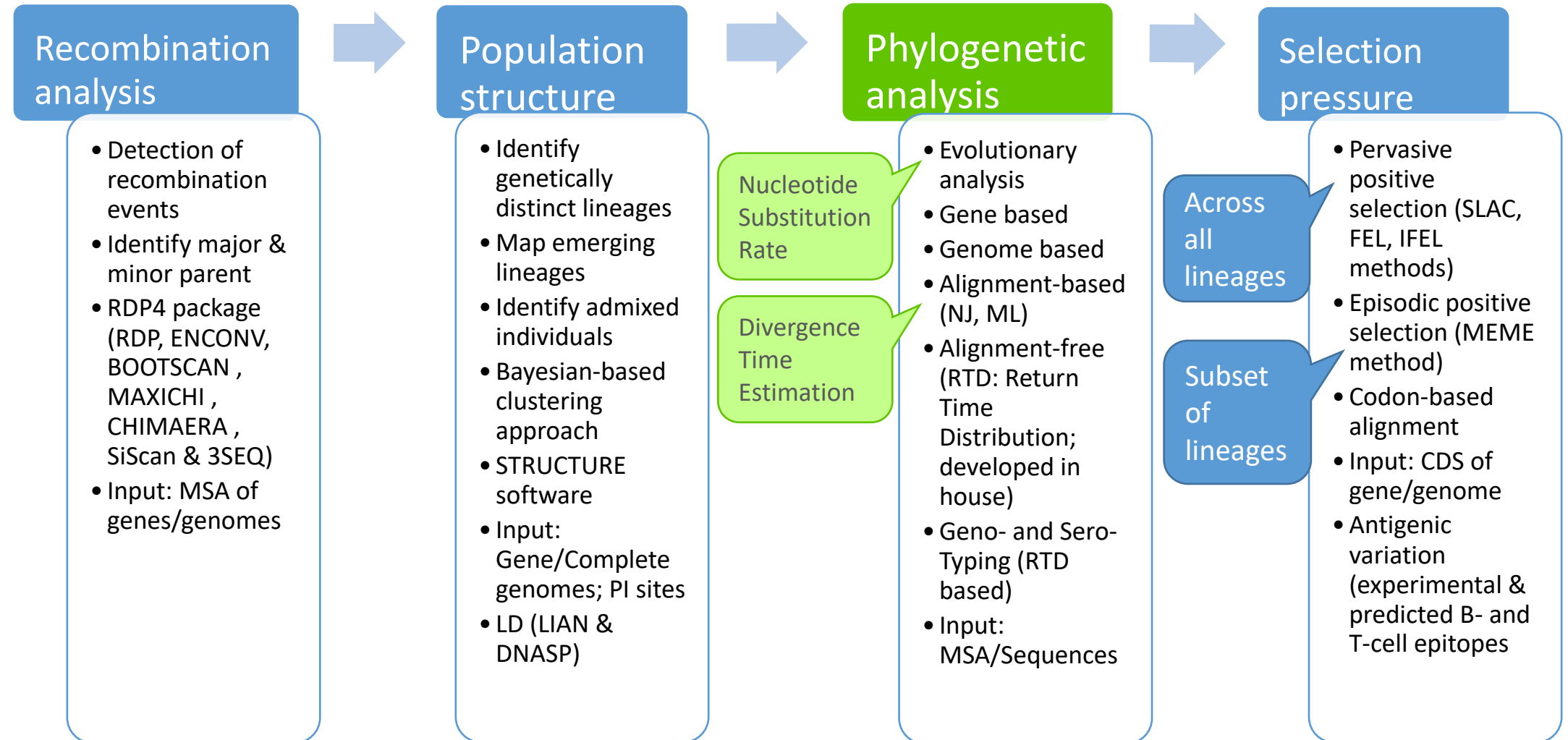
- A: Populations at K=3 wherein the labels 1, 2, and 3 represent SARS-CoV-2, SARS-CoV, and Bat-CoV.
B: Populations at K=7 | labels 1(Pangolin-CoV), 2(SARS-CoV-2), 3(Bat-CoV-Cluster_1), 4(SARS-CoV), 5(Bat-CoV-Cluster_2), 6(Bat-CoV-Cluster_3), 7(Bat-CoV-Cluster_4)



Kasibhatla SM, Kinikar M, Limaye S, Kale MM, **Kulkarni-Kale U**. 2020. Understanding evolution of SARS-CoV-2: A perspective from analysis of genetic diversity of RdRp gene **J Med Virol**. 2020:1-6.

Step-by-step recipe (that works!)

Use curated data & right tool set(s)



Take home lessons

Science is creativity & continuum of ideas!



gapingvoid
Culture Design Group

- Data curation
- Refrain from over-interpretation
 - New lineages?
 - Virulent strain?
- Look for negative data
- Include data from hosts
 - Another dimension
- Collaboration is the key
- Enjoy what you do

Acknowledgements

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
- Serum Institute of India, Pune

International Collaborators

- Dr. Suresh Kuchipudi, Penn State University, USA
- Dr. Andrew Davison, University of Glasgow, UK
- Dr. Venugopal Nair, PirBright Institute, UK
- Dr. Elin Kure, Radium Hospital, Oslo, Norway
- Dr. Mona Sabo, USN, Norway
- Dr. Ashild Andreassen, NIPH, Norway

National Collaborators

- Dr. Sunil Vaidya, NIV
- Dr. Arati Mane, NARI
- Dr. Madhuri Thakar, NARI
- Dr. Mohan Kale, Statistics Dept., SPPU
- Dr. Sangeeta Sawant
- Dr. Abhijeet Kulkarni
- Prof. J. K. Pal
- Ms. Smita Saxena
- Mr. Dattatraya Desai

Ph.D. Students/ Masters' students

- Dr. Pandurang Kolekar
- Dr. Vaishali Waman
- Dr. Sunitha Manjari
- Dr. Kiran Kadam
- Dr. Rajiv Karbhal
- Dr. Vandana Sandhu
- Nilesh Hake
- Dr. Abhijeet Pataskar
- Sanket Limaye
- Anant Shelke
- Meenal Kinikar
- Madhura Tathode
- Hemant Pawar
- 100+ Masters' project students