

VIPERdb v3.0 a platform for viral capsid knowledge discovery and structural predictive properties via machine learning

<http://viperdb.scripps.edu> (V2.0)

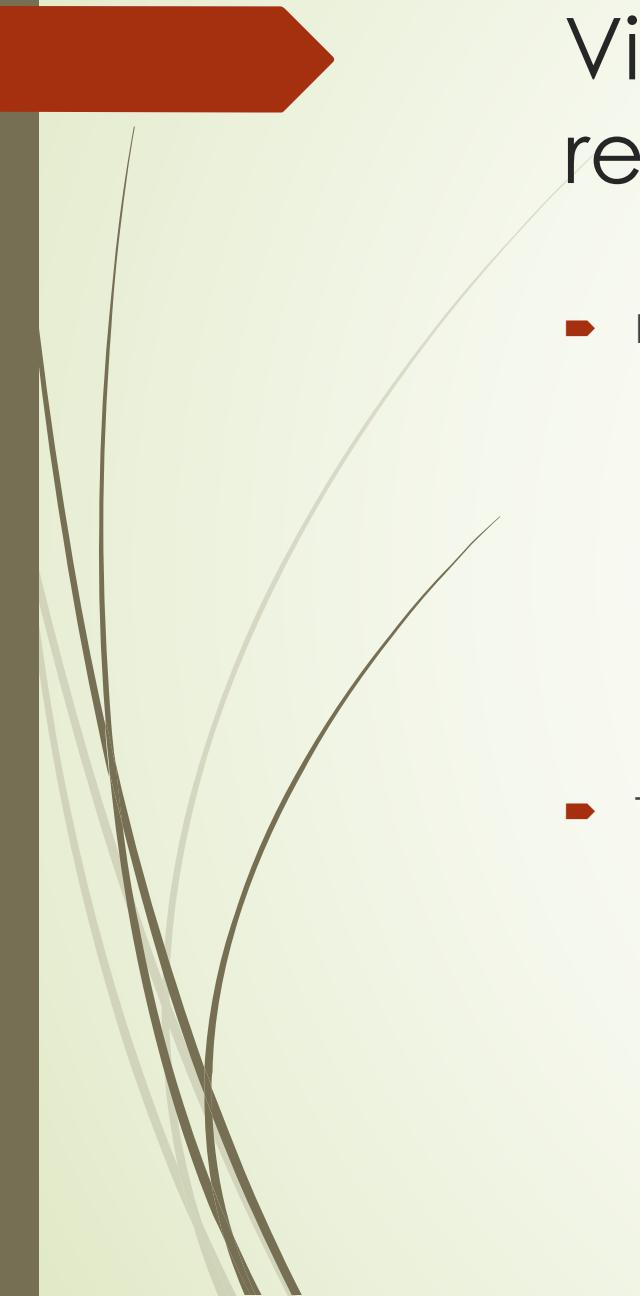
<http://dante.scripps.edu> (V3.0)

Daniel Jorge Montiel Garcia



Outline

- ▶ What is Viperdb
- ▶ Main Viperdb tools.
- ▶ New programming architecture for the version 3.
- ▶ Software specifications of Data engine
- ▶ Live code topic
 - ▶ Generate structural phylogenetic trees using viperdb services and hierarchical classifiers.



Virus Particle Explorer (Viperdb) is curated repository of viral capsid proteins.

- ▶ Features
 - ▶ Complete capsid proteins
 - ▶ Curated repository.
 - ▶ Single orientation for all the capsid proteins (Viperdb orientation).
 - ▶ Structural and physicochemical characteristics for all CPs (Structural descriptors).

▶ **36** analysis tools.

- ▶ Connections with PDB UNIPROT NCBI and ICTV.

- ▶ Technical Characteristics

▶ **250** web services.

▶ **32,558** Lines of source code (PHP, Phyton, javascript and perl) grouped into **275** files.

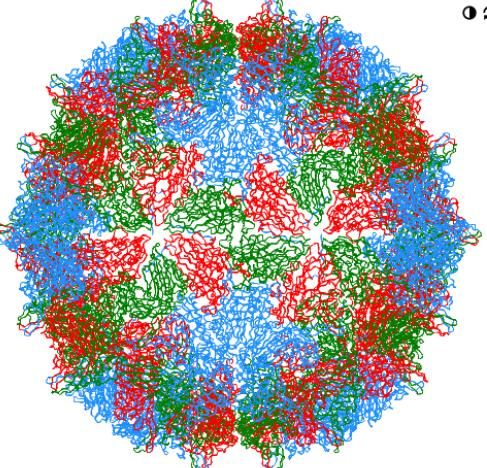
Info_page shows a detailed biological information for each virus on the database

 Virus Particle Explorer v3

About Data Utilities Virus World Viperdb2 Help Contact Us Find a Virus PDB ID Family List Genus List

2BBV Black Beetle Virus (Bbv)

Biodata Illustrations JsMol Capsid Maps MSA Annotations Related Viruses Downloads



Views	19298
Family	Nodaviridae
Genus	Alphanodavirus
Genome	ssRNA positive-strand viruses, no DNA stage
Host	Insect
Method	X-RAY DIFFRACTION
T-Number	3
Resolution	2.80
# of Subunits	180
Net Surface Charge	+420 e ⁻ per virion
Outside SASA	20,210.7 Å ² × 60
Radius	Inner: 102Å Ave: 166Å Outer: 172Å
Deposition Date	1994-06-06
PDB in	Protein Data Bank

Primary Citation

The refined three-dimensional structure of an insect virus at 2.8 Å resolution.

Wery, J.-P., Reddy, V.S., Hosur, M.V., Johnson, J.E.

Custom knowledge tools associated with each virus to analyze the physicochemical properties

2BBV Black Beetle Virus (Bbv)

Biodata Illustrations JsMol Capsid Maps MSA Annotations Related Viruses Downloads ▾

Accessible surface Profiles

Contact Tables

Association Energies

Q-Scores

Construct an Oligomer of this Virus

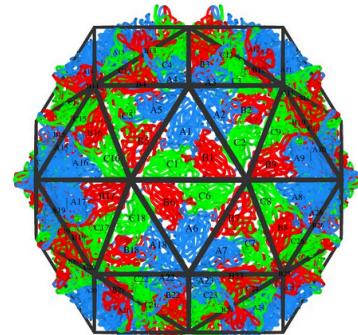
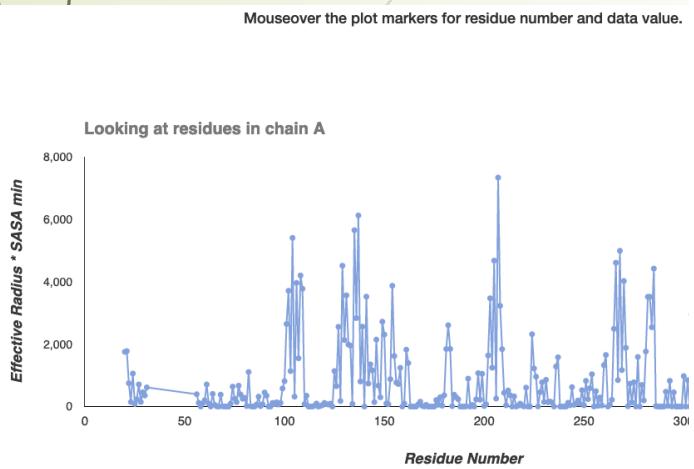
Map a residue on this Virus

Secondary Structure Information

Looking at residues in chain A

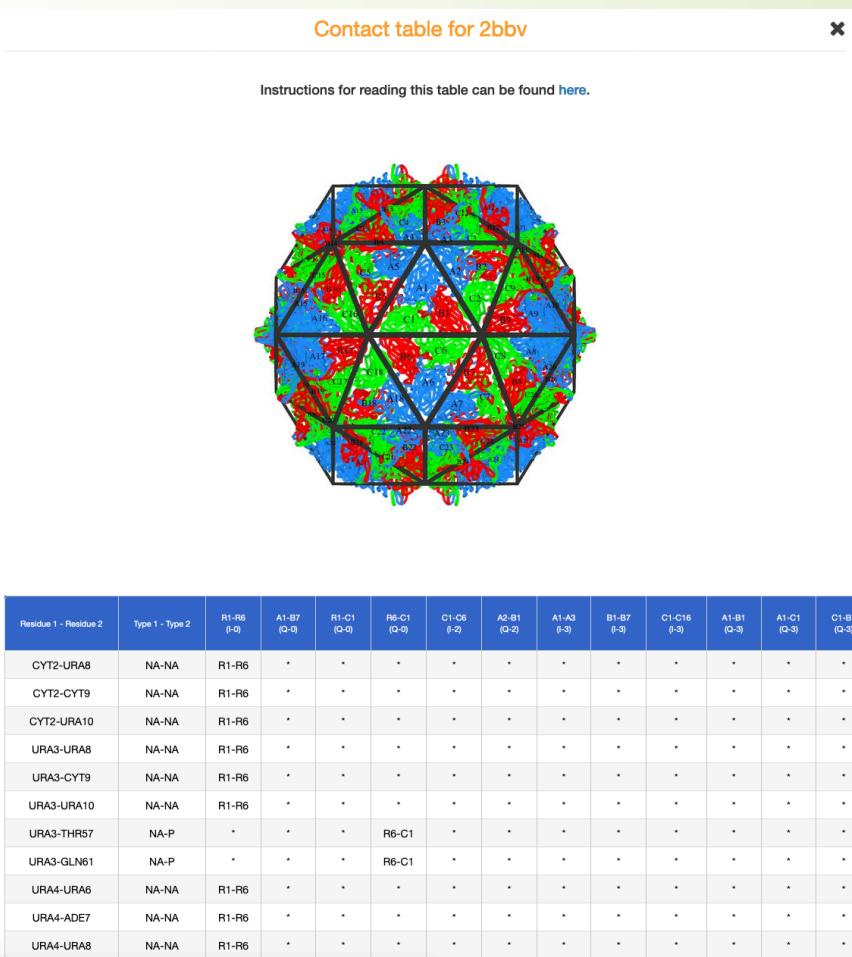
Go to this page

Solvent accessible residues, energy tables and contact maps to analyze the viral capsids



Click on the data column titles for details on the calculated properties (Caution*)

Interface	Symmetry	Association Energy (Kcal/mol)	Buried Surface Area (Å²)	Solvation Energy (Kcal/mol)	Relative strength (Percentage of strongest association energy)
A1-B1	Q-3	-85.7	4233.4	-21.3	100 %
A1-C1	Q-3	-73.0	3633.1	-20.6	85 %
B1-C1	Q-3	-71.5	3594.3	-21.2	83 %
B1-C2	Q-5	-37.4	1965.2	-9.9	44 %
A1-A2	I-5	-36.8	1945.5	-10.6	43 %
A2-B1	Q-2	-36.4	1868.3	-9.6	42 %
B1-C6	Q-6	-27.5	1634.0	-13.1	32 %
C1-C6	I-2	-21.1	1385.8	-11.0	25 %
A1-C6	Q-6	-18.8	1208.6	-9.8	22 %
C1-R1	Q-0	-6.0	546.8	0.2	7 %
R1-R6	I-0	-4.2	784.1	3.4	5 %



More tools for capsid analysis

The screenshot shows the Viperdb2 web application interface. The top navigation bar includes links for About, Data, Utilities, Virus World, Viperdb2, Help, Contact Us, Find a Virus, PDB ID, Family List, and Genus List. On the left, a sidebar lists various tools: Submit to VIPER, Web API's, Family Association Energies, Oligomer Generator, Icosahedral Server, Gallery Maker, Contact Finder, Icosahedral Matrices, Amino Acid Info, Secondary Structure Info, Multiple Structure Sequence Alignment(MSSA), and Anomaly Detection Tool. The main content area displays the "Association Energies by Family" page for the Nodaviridae family. It states that the family has 111 unique interfaces. Below this, a table provides energy values for different interface types. The table has columns for PDB, Gen, 3:, 4:, 5:, 6:, 7:, 8:, and 9:. The data row corresponds to the PDB entry 2q23, which is identified as Alphanodavirus. The values are: 3: A1_A6 (-39.7), 4: A1_A12 (-), 5: A1_A15 (-), 6: A1_B1 (-77.1), 7: A1_B2 (-75.1), 8: A1_B6 (-), and 9: A1_B7 (-).

PDB	Gen	3:	4:	5:	6:	7:	8:	9:
2q23	Alphanodavirus	3 -39.7	-	-	-77.1	-75.1	-	-
		A1_A6	A1_A12	A1_A15	A1_B1	A1_B2	A1_B6	A1_B7

Multiple structure sequence alignment

Sequence Viewer

Legend

Color Coding

Hotspot

Interface

Core

Surface Out

Surface In

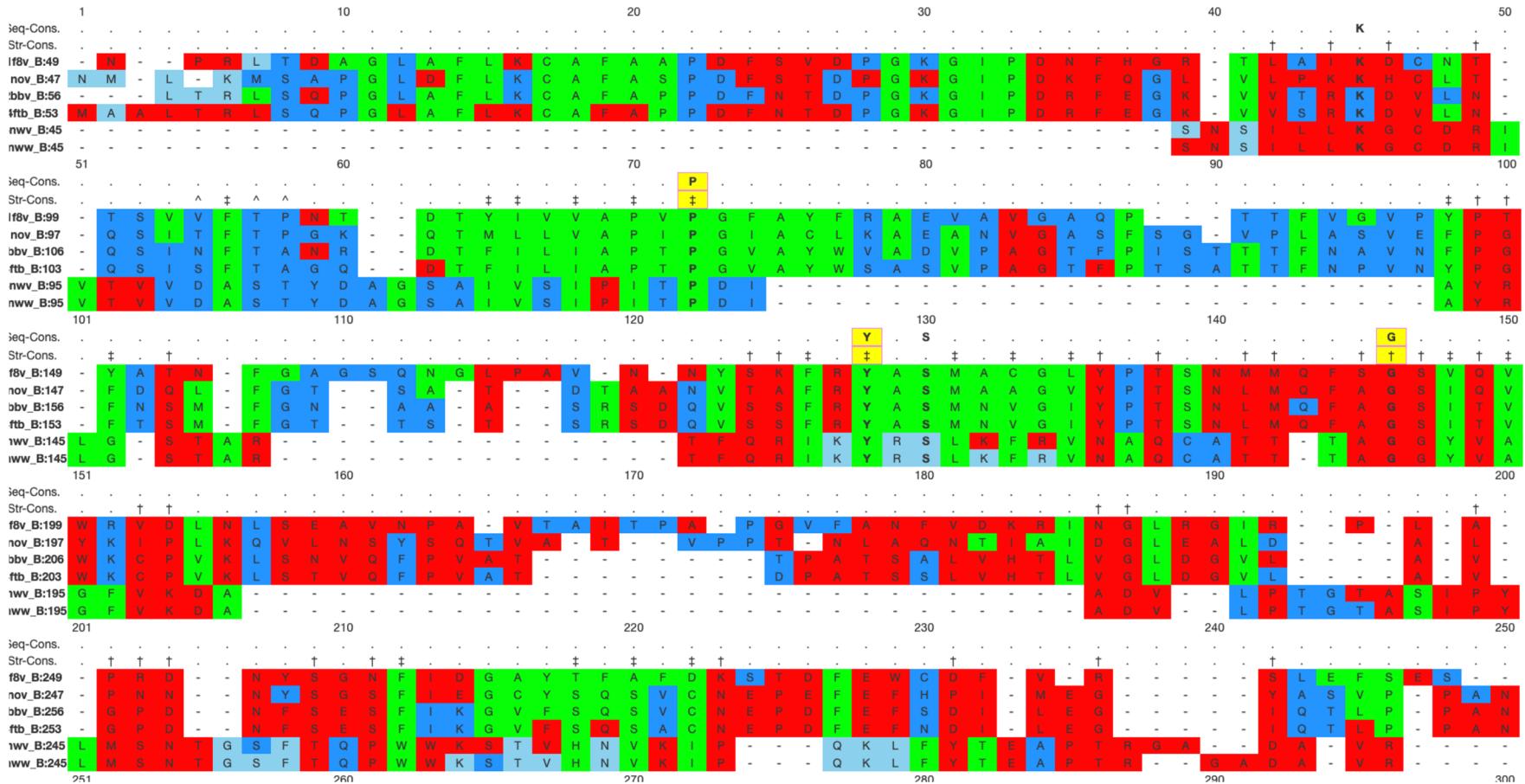
Structural Conservation Symbols

†

‡

^

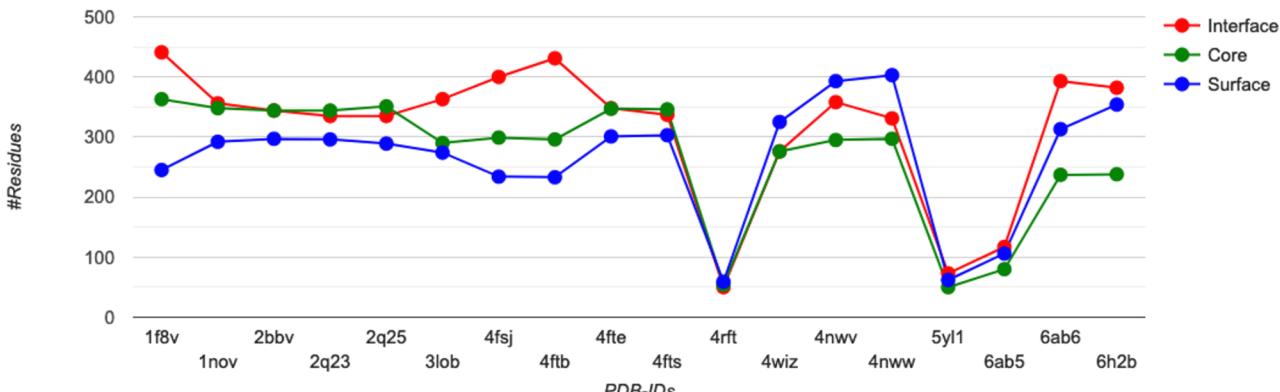
^



Anomaly detection tool

Anomaly detection tool

Family Nodaviridae



PDB	Name	Genus	T Number	Score*	Interface	Core	Surface
1f8v	Pariacoto Virus (PAV)	Alphanodavirus	3	0.0205686670	441	363	245
1nov	NODAMURA VIRUS	Alphanodavirus	3	0.0405384552	356	348	292
2bbv	Black Beetle Virus (BBV)	Alphanodavirus	3	0.0425411868	344	344	297
2q23	Crystal structure of Flock House N363T mutant	Alphanodavirus	3	0.0434402582	335	344	296
2q25	Flock House Virus Coat protein D75N mutant	Alphanodavirus	3	0.0419933893	335	351	289
3lob	Crystal Structure of Flock House Virus calcium mutant	Alphanodavirus	3	0.0544037939	363	290	274
4fsj	CRYSTAL STRUCTURE OF THE VIRUS LIKE PARTICLE OF FLOCK HOUSE	Alphanodavirus	3	0.0416168784	400	299	234
4ftb	CRYSTAL STRUCTURE OF THE AUTHENTIC FLOCK HOUSE VIRUS PARTICL	Alphanodavirus	3	0.0326278004	431	296	233
4fts	Crystal structure of the D75N mutant capsid of Flock House virus	Alphanodavirus	3	0.0405873122	348	347	301
4rft	T=1 SUBVIRAL PARTICLE OF GROPER NERVOUS NECROSIS VIRUS CAPS DELETION MUTANT (DELTA 1-34 and 218-338)	Betanodavirus	1	0.0000366162	50	54	59
4wiz	CRYSTAL STRUCTURE OF GROPER NERVOUS NECROSIS VIRUS-LIKE PARTICLE AT 3.6A	Betanodavirus	3	0.0474126746	276	276	325
4nww	CRYSTAL STRUCTURE OF ORSAY VIRUS-LIKE PARTICLE	Unclassified	3	0.0228576509	358	295	393
4nww	CRYSTAL STRUCTURE OF AN N-TERMINALLY TRUNCATED CAPSID PROTEI OF ORSAY VIRUS	Unclassified	3	0.0209974621	331	297	403
5yl1	T=1 subviral particle of Penaeus vannamei nodavirus capsid protein deletion mutant (delta 1-37 ; 251-368)	Unclassified	1	0.0000570859	73	50	62
6ab5	Cryo-EM structure of T=1 Penaeus vannamei nodavirus	Unclassified	1	0.0005394529	117	80	106
6ab6	Cryo-EM structure of T=3 Penaeus vannamei nodavirus	Unclassified	3	0.0402477195	393	237	313
6h2b	STRUCTURE OF THE MACROBRACHIUM ROSENBERGII NODAVIRUS	Unclassified	3	0.0318393016	382	238	354



Software

NIH Strategic Plan for Data Science

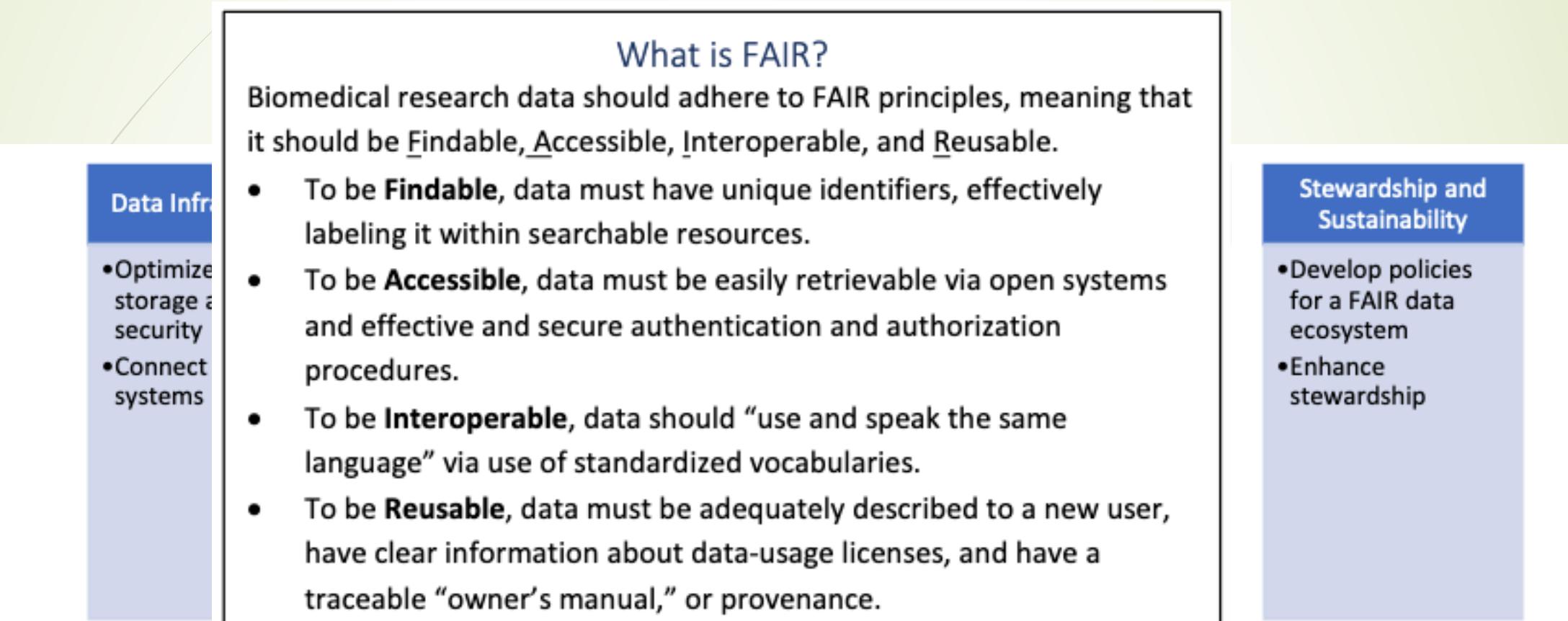
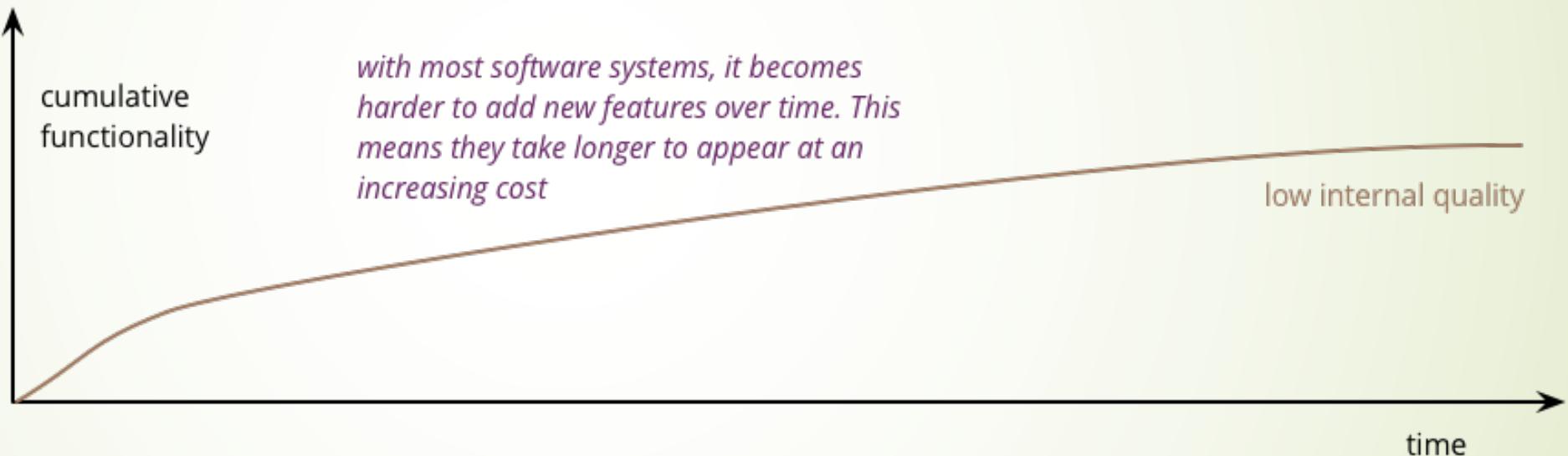
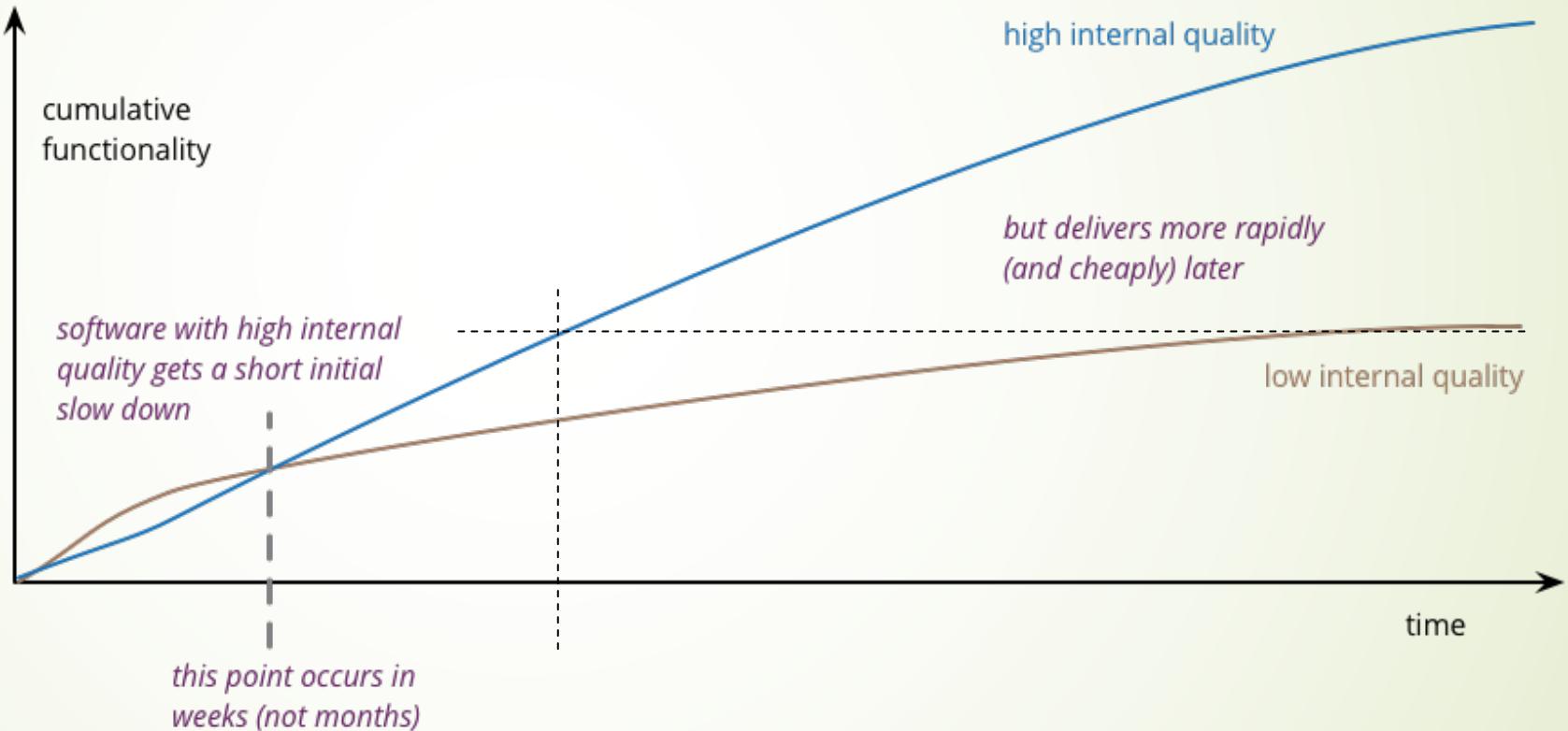


Figure 2. NIH Strategic Plan for Data Science: Overview of Goals and Objectives

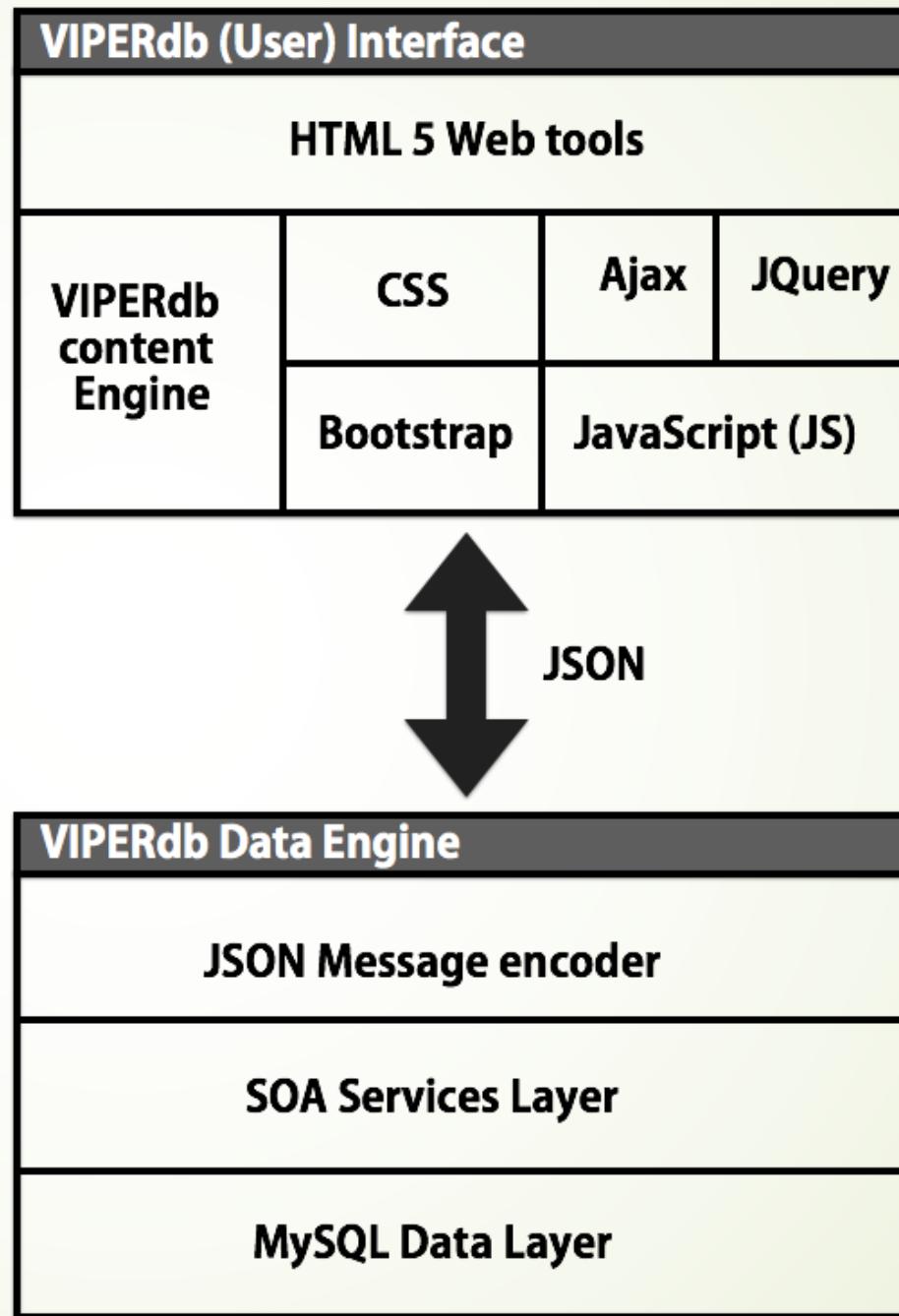
Poor quality software adds additional complexity to the bioinformatics projects



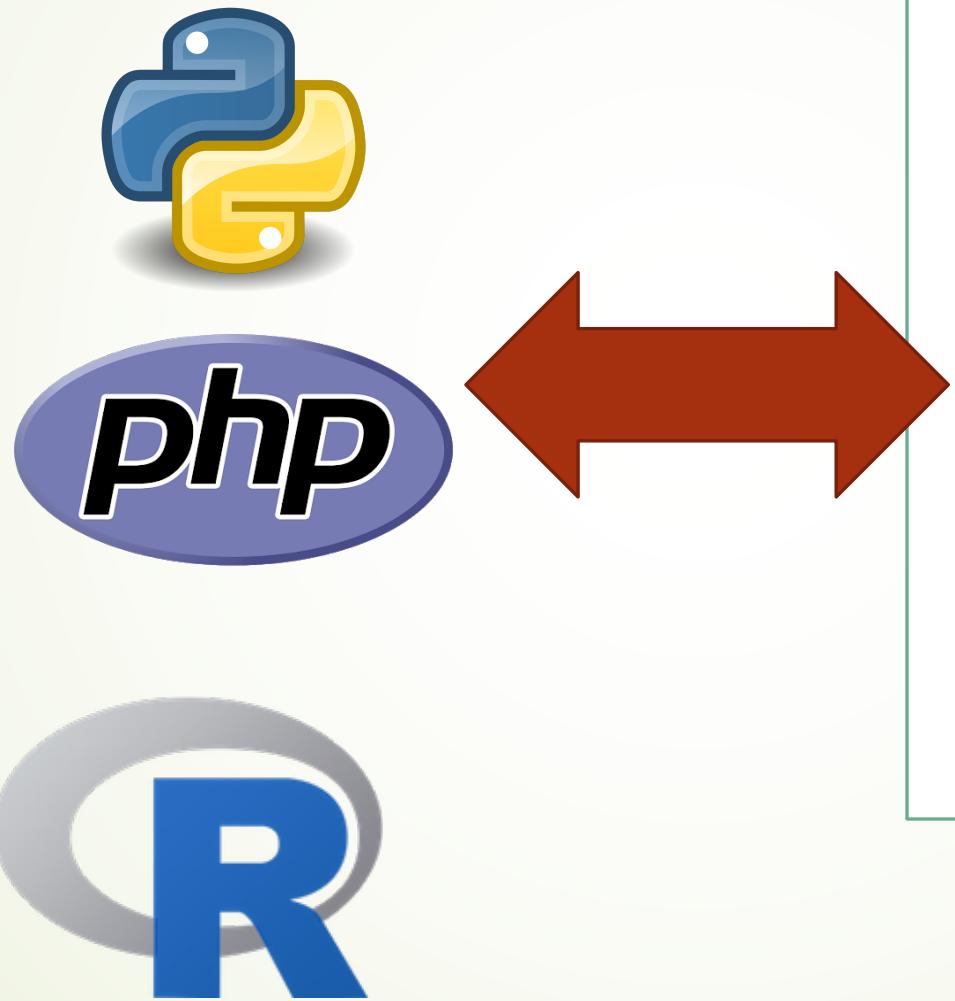
The software quality is seen as a luxury but can make the difference between a successful project and a failure



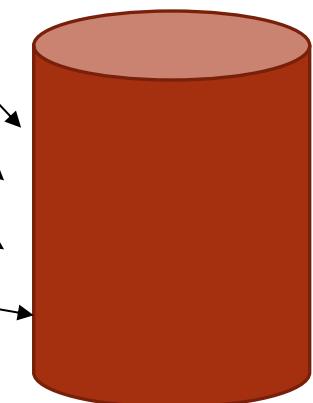
Viperdb v3. architecture



Viperdb Web services



Functionality
Functionality
Functionality
Functionality



The output of a web service is on json format

http://dante.scripps.edu/services/structural_regions.php?serviceName=get_str_family&family=bromoviridae

```
[{"VDB":"amv","name":"Alfalfa Mosaic Virus","genus":"Alfamovirus","tR":"173","tiR":"34","tcR":"27","tsR":"112","tnumber":"1","Pint":0.003904492005044668,"PCor":0.005672904580384517,"PSur":0.017238305804552625,"P":0.00038182520828220883}, {"VDB":"1cwp","name":"Cowpea Chlorotic Mottle Virus (CCMV)","genus":"Bromovirus","tR":"477","tiR":"189","tcR":"98","tsR":"190","tnumber":"3","Pint":0.04893799497323845,"PCor":0.07278466302819085,"PSur":0.05935506022206155,"P":0.21141889453090024}, {"VDB":"1js9","name":"Brome Mosaic Virus","genus":"Bromovirus","tR":"503","tiR":"213","tcR":"84","tsR":"206","tnumber":"3","Pint":0.03942650618503286,"PCor":0.0700653119645588,"PSur":0.0507970528348451,"P":0.14008203426589266}, {"VDB":"1yc6","name":"T=1 Particle of Brome Mosaic Virus","genus":"Bromovirus","tR":"154","tiR":"43","tcR":"38","tsR":"73","tnumber":"1","Pint":0.005445491227565479,"PCor":0.012375471056085703,"PSur":0.02669444119915785,"P":0.00017989522487723932}, {"VDB":"1za7","name":"THE CRYSTAL STRUCTURE OF SALT STABLE COWPEA CHOLOROTIC MOTTE VIRUS AT 2.7 ANGSTROMS RESOLUTION.","genus":"Bromovirus","tR":"481","tiR":"196","tcR":"101","tsR":"184","tnumber":"3","Pint":0.0467283641149306,"PCor":0.07123188773307343,"PSur":0.06073344107782185,"P":0.20215427019146312}, {"VDB":"3j7l","name":"FULL VIRUS MAP OF BROME MOSAIC VIRUS","genus":"Bromovirus","tR":"477","tiR":"193","tcR":"105","tsR":"179","tnumber":"3","Pint":0.04774382641175315,"PCor":0.06809259991524613,"PSur":0.06098292768269951,"P":0.19825557536175606}, {"VDB":"3j7m","name":"VIRUS MODEL OF BROME MOSAIC VIRUS (FIRST HALF DATA SET)","genus":"Bromovirus","tR":"478","tiR":"200","tcR":"103","tsR":"175","tnumber":"3","Pint":0.045228460633073106,"PCor":0.06980710986858087,"PSur":0.06058423930507502,"P":0.19128068736869783}, {"VDB":"3j7n","name":"VIRUS MODEL OF BROME MOSAIC VIRUS (SECOND HALF DATA SET)","genus":"Bromovirus","tR":478,"tiR":197,"tcR":102,"tsR":179,"tnumber":3,"Pint":0.0463684587770926,"PCor":0.0705570109109942,"PSur":0.06098292768269951,"P":0.19951295683134218}, {"VDB":"ccmv_swln_2","name":"CCMV Swollen Form Model 2","genus":"Bromovirus","tR":477,"tiR":166,"tcR":99,"tsR":212,"tnumber":3,"Pint":0.05166452876850997,"PCor":0.07234762930622546,"PSur":0.046109055620436684,"P":0.17234671285030398}, {"VDB":"1f15","name":"Cucumber Mosaic Virus (CMV)","genus":"Cucumovirus","tR":538,"tiR":196,"tcR":125,"tsR":217,"tnumber":3,"Pint":0.0467283641149306,"PCor":0.041089468490015456,"PSur":0.04196018985350474,"P":0.08056539586661685}, {"VDB":"1la7","name":"Tomato Aspermy Virus","genus":"Cucumovirus","tR":540,"tiR":208,"tcR":117,"tsR":215,"tnumber":3,"Pint":0.041800110415045076,"PCor":0.05318358664765122,"PSur":0.04364445858681409,"P":0.09702511401052372}, {"VDB":"5ow6","name":"CryoEM structure of recombinant CMV particles with Tetanus-epitope","genus":"Cucumovirus","tR":534,"tiR":195,"tcR":121,"tsR":218,"tnumber":3,"Pint":0.04707774401805784,"PCor":0.04718500041445691,"PSur":0.0411088696375739,"P":0.0913177372362751}]}
```

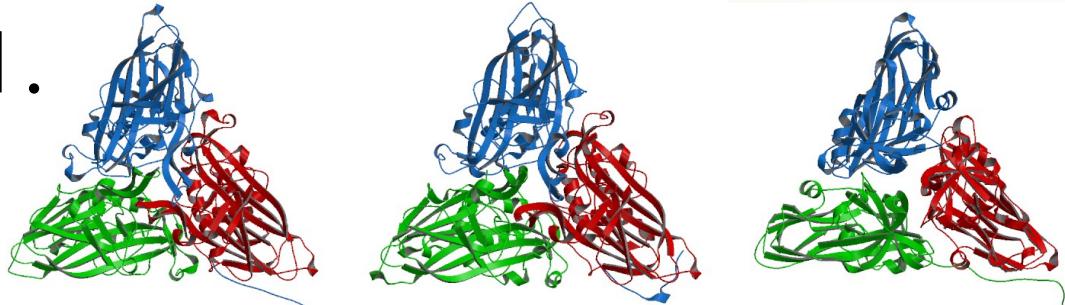


Live code topic

Use hierarchical clustering to build structural phylogenetic trees

Structural Phylogenetic tree

1.



PDB codes for CPs

2.



Virus Particle Explorer v3

Data engine



3. Data Preparation

4. Classification

5. Visualization

Code topic

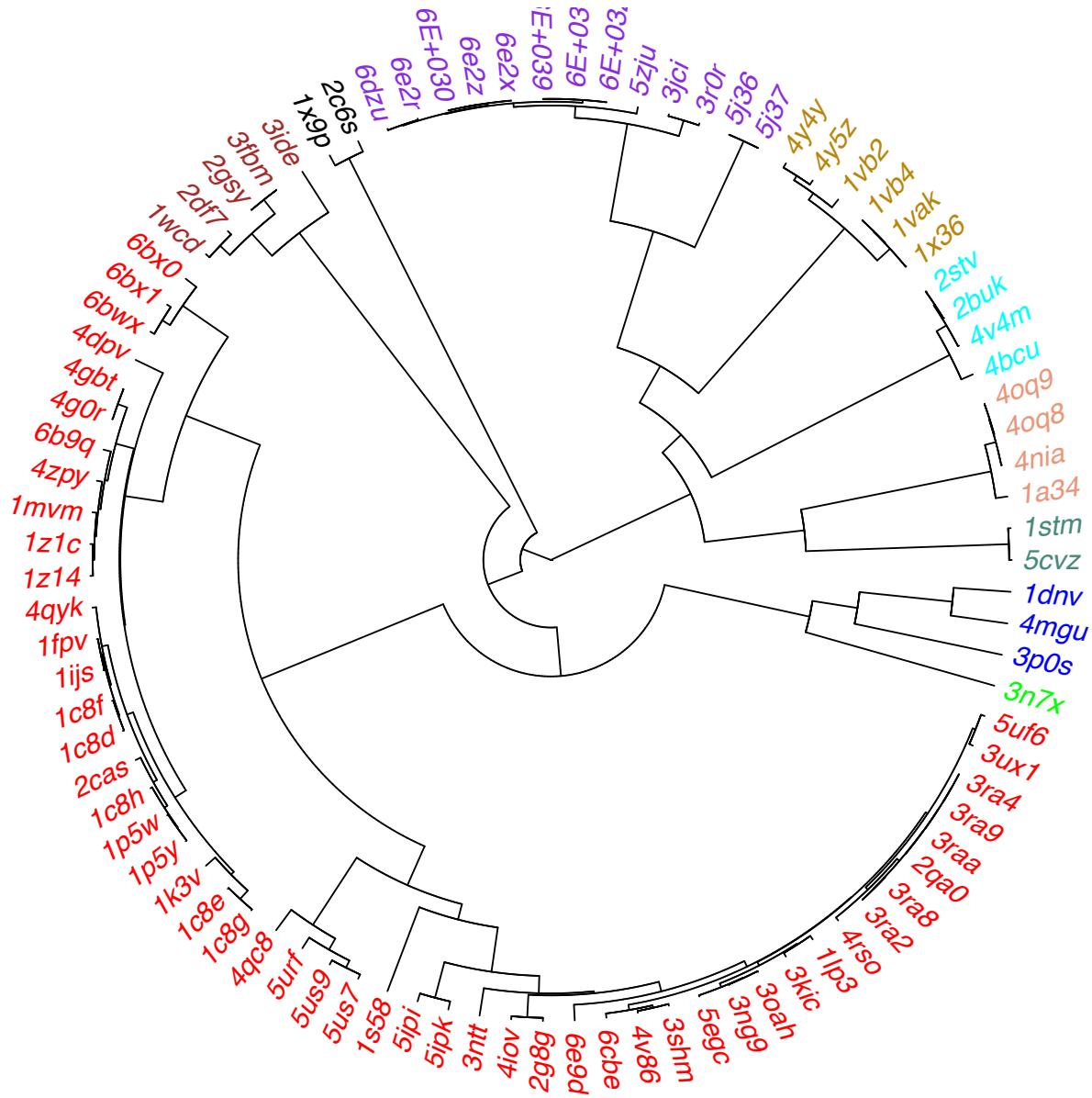
```
library(jsonlite)
library("ape")

jsonFile <-
  as.data.frame(fromJSON("http://dante.scripps.edu/services/structural_alignments.php?lpdb=1f8v,1nov,2bbv,2q23,2q25,4fft
,4wiz,4nww,1cwp,1js9,3j7m"))
row.names(jsonFile) <- jsonFile$V1
colnames(jsonFile) <- append( jsonFile$V1, 'VDB',0)

distm <- as.matrix(jsonFile[2:length(jsonFile)])
distm <- apply(distm,2, as.numeric)
distm <- 100-distm
dd <- as.dist(distm)

hc <- hclust(dd)
hdc <- as.phylo(hc)

colors = c("red", "blue", "green", "blueviolet", "brown", "darkgoldenrod", "cyan", "darksalmon", "aquamarine4", "black",
         "chocolate", "tomato", "snow2", "lightblue1", "yellow", "coral4", "deeppink3", "seagreen4", "cyan3", "orange3")
clus4 = cutree(hc, 3)
plot(hdc, type = "unroot", cex = 1, label.offset = 1, tip.color = colors[clus4])
```





Conclusions

- ▶ Viperdb v3.0 is an extraordinary platform for Machine learning and Data analysis.
- ▶ Web services provide a rich and interactive platform to share functionalities with other bioinformatics developers.
- ▶ Viperdb v3.0 architecture was developed following high quality standards.
- ▶ Viperdb development team continues developing new functionalities and improving the current version.

Thanks



Vijay S. Reddy

Christian Gallardo

Team Members

Current Members



The Scripps Research Institute



Vijay Reddy, Ph.D

Researcher



Phuong Ho

Database Admin



Jack Johnson, Ph.D

Researcher



Irapuato Institute of Technology (ITESI)



Daniel Jorge Montiel García, Ph.D

Researcher



Nelly Beatriz Santoyo Rivera

Human Resources Manager



André Delgado Gómez

Dashboard Tool



Ignacio Cruz Rodríguez

Software Quality Assurance



Brian Alejandro Flores Silva

Natural Language Processing



Alan Misael Mendoza Pérez

Dashboard Tool



Antonio de Jesús Rodríguez García

Natural Language Processing



Past Members



Noemí Villafañá Navarrete

Module Users of Viperdb



Diego Abraham Corona Chavéz

Module Virus World



María Guadalupe Moreno Magdaleno

Module Adenty v1



Francisco Javier Paredes Rocha



Fernando Vargas Bravo



Juan Daniel Zavala Martínez