

Integrated Microbial Genomes & Microbiomes/Viruses(IMG/VR): A Database for Professional Analysis of Viral DNA

Christina Dominguez and Aby Mesfin

October 1, 2019

Loyola Marymount University

Outline

- IMG/VR: A public database for viral genomics
- Successful in its Content Domain and Biologically Relevant
- Offers Current Information and Links to Other Databases
- Browsing the Database
- A Professional Database with Limited Navigation
- Summary

Outline

- **IMG/VR: A public database for viral genomics**
- Successful in its Content Domain and Biologically Relevant
- Offers Current Information and Links to Other Databases
- Browsing the Database
- A Professional Database with Limited Navigation
- Summary

IMG/VR is a Public Database for Viral Genomics

- Owned by The Regents of University of Southern California
 - Publically funded
- Secondary Curated Source Data
 - Worldwide scientists are encouraged to submit their own data
 - Application process for submissions

The screenshot displays the IMG/VR (Integrated Viral Genomes) website. At the top, there is a navigation bar with the JGI logo and the text 'IMG/VR INTEGRATED VIRAL GENOMES / VIRUS'. Below this, there are links for 'My Analysis Carts', 'Genomes', 'Scaffolds', 'Functions', 'Genes', 'Genome Search History', and 'Gene Search History'. A search bar is located on the right side of the top bar.

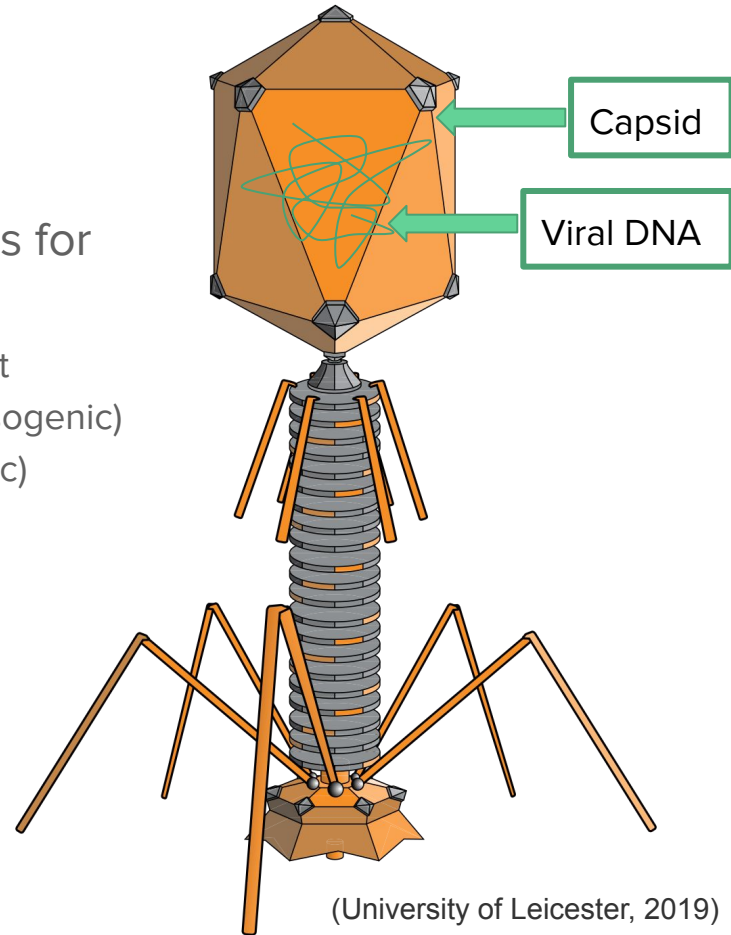
The main content area is divided into several sections:

- IMG Viral Content:** A list of viral datasets including 'Viral Datasets', 'Viral Operational Taxonomic Units (VOTUs)', 'With Host', and 'Quality'. Each category has a corresponding number of entries.
- Ecosystems:** A section with a 'Show Human Body Sites' button and a 'Hint' box explaining how to use the map.
- Map:** A world map showing the distribution of viral genomes, with colored circles indicating the location and count of genomes. A legend below the map shows the 'Ecosystem' for each location.
- Legend:** A table showing the distribution of viral genomes across various ecosystems and habitat types.

Ecosystem	Count	Habitat Type	Count
Engineered	8922	Algae	2559
Bioreactor	2884	Animal	2722
Bioremediation	2067	Aquatic	247821
Bioproduction	2331	Terrestrial	23260
Build environment	2050	Unclassified	4
Food production	29	Birds	593
Lab enrichment	1317	Crustans	126
Lab synthesis	52	Echinodermata	11
Modeled	41	Fish	200
Solid waste	230	Fung	3500
Unclassified	1	Human	8329
Wastewater	2058	Insecta	348
		Invertebrates	287
		Mammals	13261
		Mollusca	2233
		Mollusca	209
		Plants	1323
		Porifera	81
		Protista	8
		Protozoa	10
		Reptilia	4
		Tardigrada	52
		Unclassified	209

Viruses

- Biological agents that insert DNA into host cells for replication
 - Viral capsids contain viral DNA to be injected into host
 - Inserted DNA can be integrated into host genome (lysogenic) or quickly transcribed/translated by host enzymes (lytic)
- Not considered “alive”
 - Missing some of the characteristics of living things
 - Are not made of cells
 - Require a host for reproduction
 - Do not require energy for survival



Outline

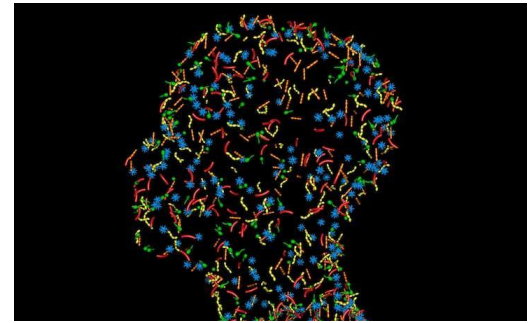
- IMG/VR: A public database for virgial genomics
- **Successful in its Content Domain and Biologically Relevant**
- Offers Current Information and Links to Other Databases
- Browsing the Database
- A Professional Database with Limited Navigation
- Summary

Successfully Proves its Content Domain

- **Mission: “annotation, analysis, and distribution” of genome and microbiome datasets**
 - Species content includes viruses
 - 8,389 cultivated viruses
 - 735,112 viral genomic fragments
 - **ANALYSIS and ANNOTATION**
 - Offers many analytical tools for dataset evaluation ex. BLAST
 - GenBank processed through IMG submission system and annotation pipeline
 - **DISTRIBUTION**
 - Public forum for access to information

Biologically Relevant Database

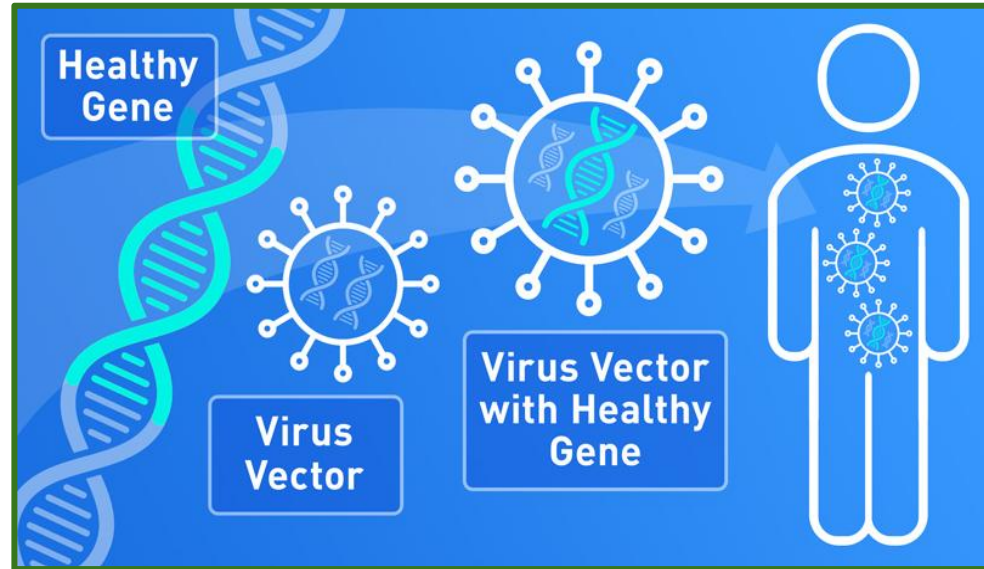
- Comparative analysis between genome datasets
 - Visualize similarities and differences
- Answers biological questions related to viruses in:
 - Human health
 - Animal health
 - Environmental and Crop health
 - Model organisms in research



(Nautilus, 2019)

Applications of Studying Viral Genomics

- Immunology
 - Understanding pathogenic viruses
 - Phage therapy to target pathogenic bacteria in patients
 - Gene therapy
- Biological Research and Genetic Engineering
 - Genetically modifying cells with viruses
 - Viruses can serve as DNA vectors



Outline

- IMG/VR: A public database for viral genomics
- Successful in its Content Domain and Biologically Relevant
- **Offers Current Information and Links to Other Databases**
- Browsing the Database
- A Professional Database with Limited Navigation
- Summary

IMG/VR Offers Current Information

- Up to Date
 - Started in 2016
 - Updated on a quarterly basis
 - Last update: September 2019
- Content Timely
 - Virus research needed as it impacts human health
 - Extends to the scientific community and beyond

IMG/VR links to other databases as well

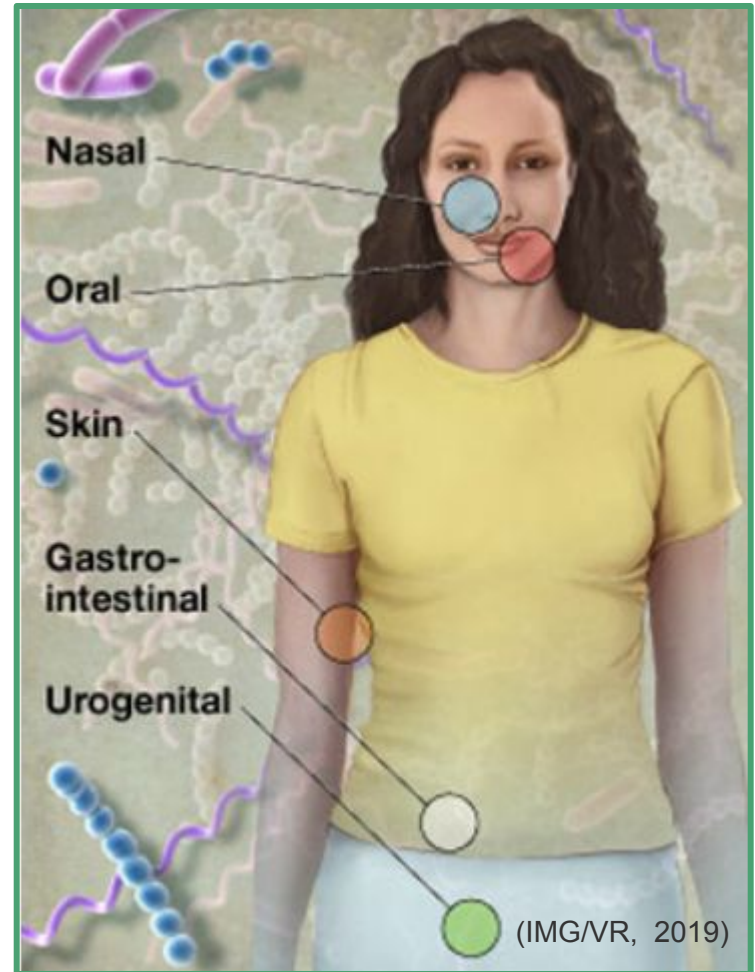
- Within IMG/VR are links to...
 - a. NCBI BLAST → comparing genomic information across species and viral strains
 - b. IMG/M → analyzing genomic information of bacteria, archaea, eukarya, and plasmids as well
 - c. IMG/M ER → analyzing genomic information of microbes
 - d. IMG/ABC → a database of biosynthetic clusters

Outline

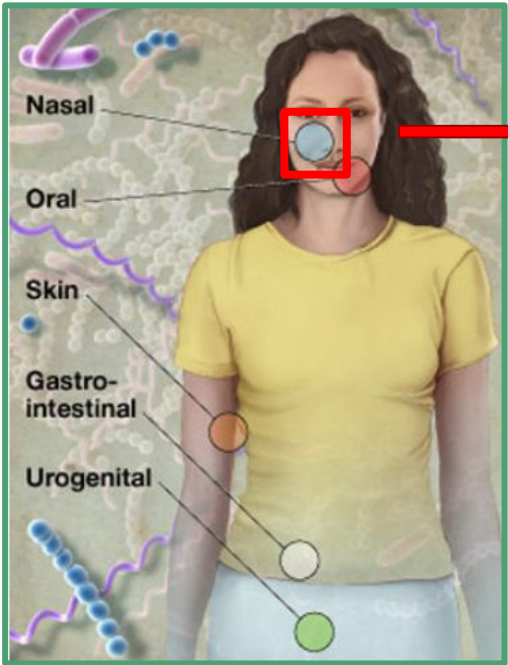
- IMG/VR: A public database for viral genomics
- Successful in its Content Domain and Biologically Relevant
- Offers Current Information and Links to Other Databases
- **Browsing the Database**
- A Professional Database with Limited Navigation
- Summary

Browsing Genomes on IMG/VR

- Viral genomic data is organized
 - Human body navigation tool
 - Nasal, Oral, Gastrointestinal, Skin, and Urogenital
 - Environmental
 - Air, aquatic, terrestrial
 - Host Association
 - Birds, insects, fungi, etc.



Human Body Navigation Tool



Add Selected to Scaffold Cart Toggle Selected Select All Clear All Download Selected in Excel

Filter column: Scaffold ID Filter text Apply

Export Page 1 of 25 << first < prev 1 2 3 4 5 6 7 8 9 10 next > last >> 100

Column Selector Select Page Deselect Page

Select	Scaffold ID	Percent VPFS	Viral Clusters	Host
<input type="checkbox"/>	C1369256	.33	vc_83570	
<input type="checkbox"/>	C1369276	.88	vc_622	Staphylococcus_epidermidis ; Staphylococcus_schleiferi
<input type="checkbox"/>	C1604150	.7	vc_3670	Cutibacterium_granulosum
<input type="checkbox"/>	C1604256	.31	vc_3670	Cutibacterium_granulosum
<input type="checkbox"/>	C1757481	.64	vc_736	Streptococcus_sp_C300
<input type="checkbox"/>	C1757495	.89	vc_736	Streptococcus_anginosus ; Streptococcus_intermedius ; Streptococcus_mitis ; Streptococcus_oralis ; Streptococcus_parasanguinis ; Streptococcus_sp_C300 ; Streptococcus_sp_FDAARGOS_256

Scaffold ID	C1369256
Genome	Human_retroauricular_crease_microbial_communities_from_NIH_USA_visit_2_subject_159591683_replicate_2_*(MFR-FS)_assembled
Topology	linear
Sequence Length (bp)	10387
GC Content	0.34
Read Depth	1
Lineage	Viruses; unclassified; unclassified; unclassified; unclassified; unclassified; vc_83570
Lineage Percentage	0.56
Gene Count	18
Ecosystem	Host-associated
Ecosystem Category	Human
Ecosystem Type	Skin
Ecosystem Subtype	retroauricular_crease
Specific Ecosystem	Unclassified
Habitat	Human
Habitat Type	Host-associated(human)
Perc VPFS	.33
Viral Cluster	vc_83570

EXPORT

(IMG/VR, 2019)

Environmental and Host Navigation Tools

Engineered 40537	Environmental 575538	Host-associated 123684	Habitat Type 7982
Bioreactor 2694	Air 128	Algae 2459	Air 20
Bioremediation 2207	Aquatic 547821	Animal 2777	Aquatic(sediment) 180
Biotransformation 1331	Terrestrial 27585	Annelida 284	Engineered 1351
Built environment 5890	Unclassified 4	Arthropoda 4782	Freshwater 955
Food production 20		Birds 553	Host-associated(human) 2093
Lab enrichment 1317		Cnidaria 120	Host-associated(other) 462
Lab synthesis 92		Echinodermata 11	Host-associated(plants) 362
Modeled 67		Fish 200	Marine 1130
Solid waste 5760		Fungi 1680	Non-marine_saline_and_Alkaline 204
Unclassified 1		Human 81279	Terrestrial(other) 100
Wastewater 20958		Insecta 338	Terrestrial(soil) 990
		Invertebrates 367	Thermal_springs 2
		Mammals 13591	Thermal_springs 129
		Microbial 1070	mixed_environments 4
		Mollusca 406	
		Plants 13423	

Add Selected to Genome Cart Select All Clear All

Filter column: [Domain] Filter [text] Apply

Export Page 1 of 28 << first < prev 1 2 3 4 5 6 7 8 9 10 next > last >> 100

Column Selector Select Page Deselect Page

Select	Domain	Status	Study Name	Taxon OID	Genome Name	Ecosystem	Ecosystem Category	Ecosystem Type	Ecosystem Subtype	Specific Ecosystem	Habitat Type	Habitat (from GOLD)	Depth (in meters)	Viral Count
<input type="checkbox"/>	*	P	Microbial and viral regulation of community carbon cycling across diverse low-oxygen zones	330000908	Marine microbial and viral communities from oxygen minimum zone, Eastern Pacific Ocean - ETNP2014F12-01SV259	Environmental	Aquatic	Marine	Oceanic	Unclassified	Marine	Marine		88
<input type="checkbox"/>	*	P	microbial communities from Monterey Bay, California, United States	330002423	Seawater microbial communities from Monterey Bay, California, United States - 1D	Environmental	Aquatic	Unclassified	Coastal		Marine	seawater	5	18
<input type="checkbox"/>	*	P	Hot spring microbial communities from Yellowstone National Park	330000196	Hot spring thermophilic microbial communities from Obsidian Pool, Yellowstone National Park, USA - site 9 B9	Environmental	Aquatic	Thermal springs	Hot (42-90C)	Unclassified	Thermal_springs	Hot spring		4

Scaffold Detail

Scaffold ID	Ga0066868_10000026
Genome	Marine microbial and viral communities from oxygen minimum zone, Eastern Pacific Ocean - ETNP2014F12-01SV259 (*) (MER-FS) (assembled)
Topology	linear
Sequence Length (bp)	45696
GC Content	0.36
Read Depth	12
Lineage	Viruses, unclassified, unclassified, unclassified, unclassified, unclassified, vc_1110
Gene Count	50
Ecosystem	Environmental
Ecosystem Category	Aquatic
Ecosystem Type	Marine
Ecosystem Subtype	Oceanic
Specific Ecosystem	Unclassified
Habitat	Marine
Habitat Type	Marine
Perc VPFs	.36
Viral Cluster	vc_1110

Add to Scaffold Cart

Add Selected to Scaffold Cart Toggle Selected Select All Clear All

Filter column: [Scaffold ID] Filter [text] Apply

Export Page 1 of 1 << first < prev 1 next > last >> All

Column Selector Select Page Deselect Page

Select	Scaffold ID	Gene Count	Sequence Length (bp)	GC Content	Perc VPFs	Viral Cluster	Host Detection
<input type="checkbox"/>	Ga0066868_10000026	50	45696	0.36	.36	vc_1110	0
<input type="checkbox"/>	Ga0066868_10000055	54	35527	0.43	.28	vc_13822	0
<input type="checkbox"/>	Ga0066868_10000401	20	16027	0.38	.45	vc_4685	0
<input type="checkbox"/>	Ga0066868_10000499	23	14650	0.31	.57	vc_609	0
<input type="checkbox"/>	Ga0066868_10000567	16	13833	0.29	.38	vc_13440	0
<input type="checkbox"/>	Ga0066868_10000580	20	13722	0.36	.45	vc_25593	0
<input type="checkbox"/>	Ga0066868_10000617	18	13414	0.3		vc_38772	0
<input type="checkbox"/>	Ga0066868_10000642	38	13039	0.37	.18	vc_17722	0
<input type="checkbox"/>	Ga0066868_10000646	17	13020	0.43		vc_1822	0
<input type="checkbox"/>	Ga0066868_10000771	19	11836	0.32	.32	vc_13496	0

EXPORT

Genome Information Provided by IMG/VR

- The genomic data offered entails:
 - a. Scaffold ID Number
 - b. Name of the Organism
 - c. Length of the Sequence
 - d. Gene Count
 - e. Guanine and Cytosine Content
 - f. Predicted Host
 - g. Habitat

Exporting Data from IMG/VR as an Tab Delimited Excel Sheet

1. Select the data
2. Add the data to your cart
3. Click on the export tab

1 scaffold(s) in cart

Scaffolds in Cart Upload & Export & Save Function Profile Histogram Kmer Analysis Phylogenetic Distribution

Filter column: [Gene Count] Filter [text]: [] Apply ?

Export Page 1 of 1 << first < prev 1 next > last >> [All] v

Column Selector Select Page Deselect Page

Select	Scaffold ID	Scaffold Name	Genome ID	Genome	Gene Count	Sequence Length (bp)	GC Content	Perc VPFS	Viral Cluster
<input checked="" type="checkbox"/>	C1008267		7000000619	Human buccal mucosa microbial communities from NIH_USA - visit 2 - subject 763759525 (*) (MER-FS) (assembled)	11	6042	0.40	.45	vc_88807

Export Page 1 of 1 << first < prev 1 next > last >> [All] v

Add Genomes of Selected Scaffolds to Cart Add Genes of Selected Scaffolds To Cart

Toggle Selected Select All Clear All Remove Selected

Browsing Single Genes on IMG/VR

1. Select “Find Gene” Tab
2. Select “Phylogenetic Profiler”
3. Select “Single Gene”
4. Choose the Organism
 - a. Choose “Viruses” in the Domain tab
 - b. Click “Show”
 - c. Select the Organism and add to the “Selected Genomes” Tab
 - d. Press “Submit”

The screenshot shows the IMG/VR website interface. The 'Find Genes' tab is highlighted in red. A dropdown menu is open, showing 'Phylogenetic Profilers' and 'Single Genes', both highlighted in red. The 'Single Genes' option is selected. The page also displays 'Viral Datasets' and 'Viral Operational Taxonomic Units (vOTUs)' with counts for various categories.

The screenshot shows the 'Phylogenetic Profiler for Single Genes' interface. The 'Domain' dropdown is set to 'Viruses'. A search list is displayed with 'Abutilon Brazil virus (V) [F]' selected. The 'Selected Genomes' section shows 'Abutilon Brazil virus (V) [F]' added. The 'Submit' button is highlighted in red.

Outline

- IMG/VR: A public database for viral genomics
- Successful in its Content Domain and Biologically Relevant
- Offers Current Information and Links to Other Databases
- Browsing the Database
- **A Professional Database with Limited Navigation**
- Summary

A Professional Database with Limited Navigation

- Requires experience in bioinformatics and general biology
 - Infers professional use
- Difficulty in navigating sources of information
- The single gene data was inaccessible due to a recurring error page or “forbidden” page
- Frequently referred to “Help” page and “FAQ” page



Summary

- Public database for viral genomics
- Successful in its Content Domain and Biologically Relevant
 - Fulfills its mission of “annotation, analysis, and distribution” of genome datasets
 - Potential to answer a wide range of questions in biology
- Easy Navigation between other databases
- Browsing the Database
 - Navigation within database is direct
- Database allows for analysis of viral genomics at a more professional level
 - Requires experience

Acknowledgments

Thank You
Dr. Dahlquist
LMU Biology Department

References

“Bacteriophage.” University of Leicester (2019). Retrieved October 1, 2019 from

<https://www2.le.ac.uk/projects/vgec/highereducation/topics/microbial-genetics-1/bacteriophage>

Integrated Microbial Genomes & Microbiomes/VR. (2019). Retrieved October 1, 2019, from

<https://img.jgi.doe.gov/cgi-bin/vr/main.cgi%7CIMG/VR>

“What Is Gene Therapy? How Does It Work?” FDA (2019). Retrieved October 1, 2019 from

<https://www.fda.gov/consumers/consumer-updates/what-gene-therapy-how-does-it-work>

“What is the Human Microbiome, Exactly?” Nautilus (2019). Retrieved October 1, 2019 from

<http://nautil.us/blog/what-is-the-human-microbiome-exactly>