

# **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 Sourced and Curated Data
  - Worldwide scientists are encouraged to submit their own data
  - Application process for submissions

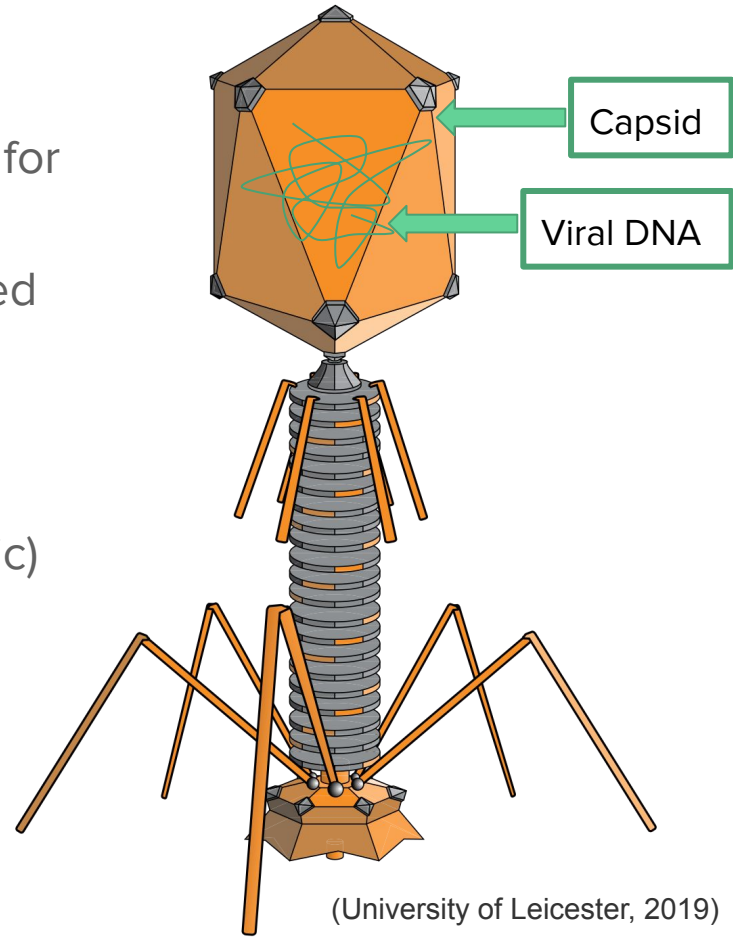
The screenshot displays the IMG/VR website interface. At the top, there is a navigation bar with the logo and search options. Below the navigation bar, there are several sections: 'Viral Outlets', 'Viral Operations/ Taxonomic Units (VOTUs)', 'Ecosystems', and a world map showing viral hotspots. A table at the bottom provides a detailed breakdown of viral hotspots by habitat type.

Engineered	Environmental	Host-associated	Habitat Type
88527	275338	122614	2382
Bioreactor	2884	Air	2429
Bioremediation	3202	Aquatic	2227
Bioremediation	3333	Terrestrial	2186
Built environment	3880	Unclassified	4
Food production	92	Arthropods	8764
Lab enrichment	1817	Erp	100
Lab synthesis	92	Chordata	100
Mineral	82	Echinodermata	11
Solid waste	8780	Fish	260
Unclassified	1	Fungi	1820
Wastewater	20265	Human	11212
		Insecta	281
		Invertebrates	261
		Mammals	10811
		Mollusc	1023
		Mollusca	899
		Plants	1442
		Porifera	81
		Protists	4
		Protozoa	13
		Reptile	4
		Tunicates	89
		Unclassified	200

(IMG/VR, 2019)

# 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

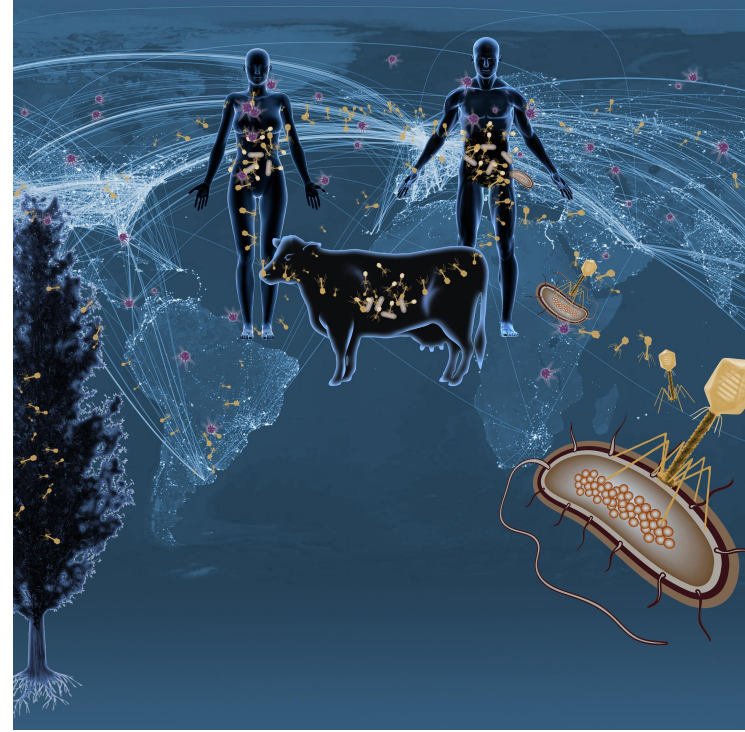
- IMG/VR: A public database for viral genomics
- **Successful in its Content Domain and Biologically Relevant**
- Offers Current Information and Links to Other Databases
- Browsing the Database
- A Professional Database with Limited Navigation
- Summary

# Successfully Proves its Content Domain

- **Mission: “annotation, analysis, and distribution” of genome and microbiome datasets**
  - Species content includes viruses
    - 8,389 cultivated viruses
  - **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

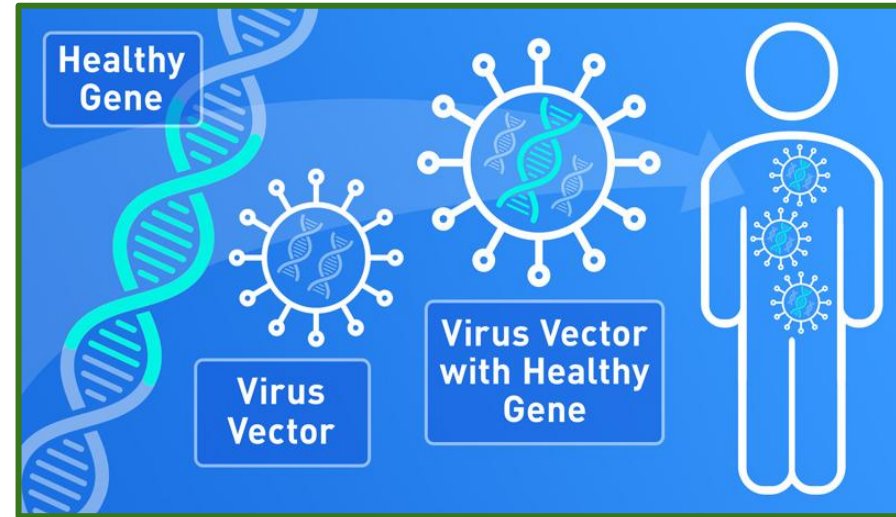


(IMG/VR, 2019)



# 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



(FDA, 2019)

# 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

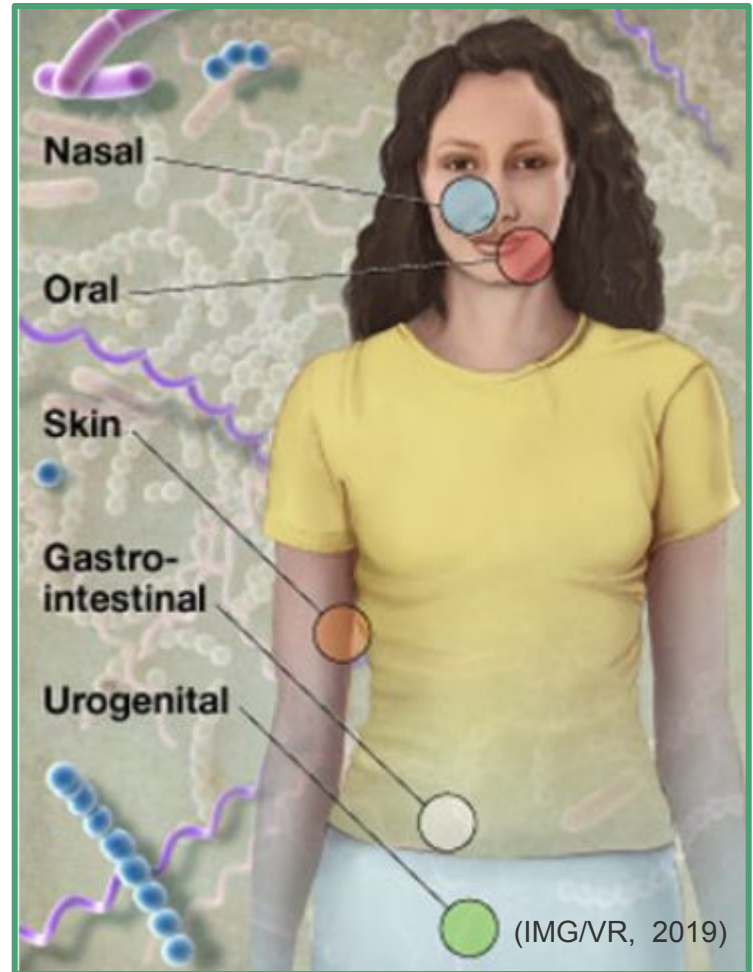
- Within IMG/VR are links to...
  - a. **NCBI BLAST** → comparing genomic information across species and viral strains
  - b. **IMG/M** → analyzing genomic information of bacteria, archaea, eukarya, and plasmids as well
  - c. **IMG/M ER** → analyzing genomic information of microbes
  - d. **IMG/ABC**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

# Outline

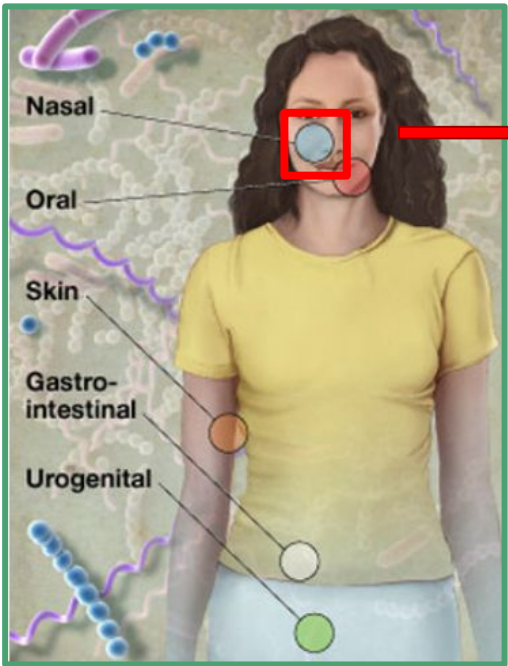
- IMG/VR: A public database for viral genomics
- Successful in its Content Domain and Biologically Relevant
- Offers Current Information and Links to Other Databases
- **Browsing the Database**
- A Professional Database with Limited Navigation
- Summary

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

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

**EXPORT**

# Environmental and Host Navigation Tools

<b>Engineered</b> 40537	<b>Environmental</b> 575538	<b>Host-associated</b> 123684	<b>Habitat Type</b> 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**

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

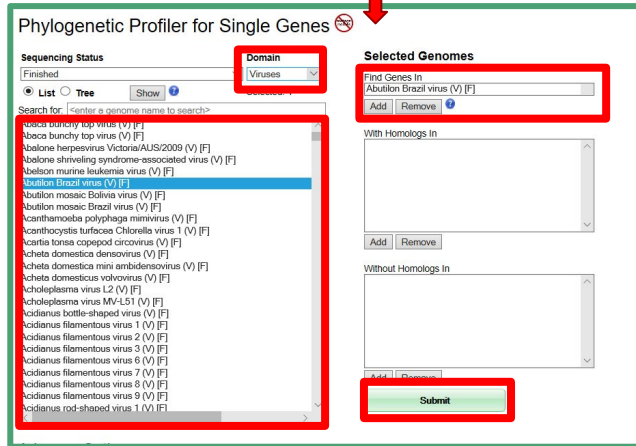
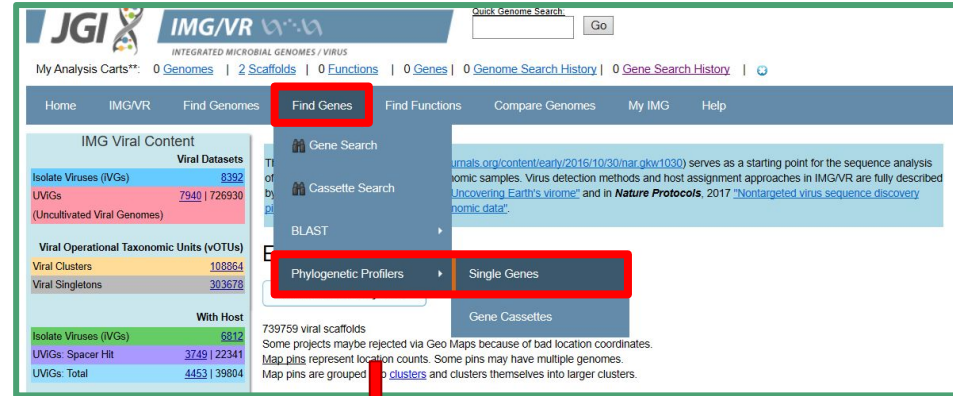
1. Select the data
2. Add the data to your cart
3. Click on the export tab

The screenshot displays the IMG/VR interface for a scaffold in a cart. At the top, it says "1 scaffold(s) in cart". Below this are several tabs: "Scaffolds in Cart", "Upload & Export & Save", "Function Profile", "Histogram", "Kmer Analysis", and "Phylogenetic Distribution". The "Scaffolds in Cart" tab is active. Below the tabs, there is a filter section with "Filter column:" set to "Gene Count" and "Filter" set to "text". An "Export" button is highlighted with a red box. Below the filter section, there are navigation controls: "Page 1 of 1", "<< first < prev 1 next > last >>", and "All" dropdown. Below these are "Column Selector", "Select Page", and "Deselect Page" buttons. A table with one row is shown, with the "Select" checkbox highlighted by a red box. The table columns are: Select, Scaffold ID, Scaffold Name, Genome ID, Genome, Gene Count, Sequence Length (bp), GC Content, Perc VPFs, and Viral Cluster. The data row contains: , C1008267, (empty), 700000619, Human buccal mucosa microbial communities from NIH\_USA - visit 2\_subject 763759525 (\*) (MER-FS) (assembled), 11, 6042, 0.40, .45, and vc\_88807. Below the table, there is another "Export" button and navigation controls. At the bottom, there are buttons for "Add Genomes of Selected Scaffolds to Cart", "Add Genes of Selected Scaffolds To Cart", "Toggle Selected", "Select All", "Clear All", and "Remove Selected".

Select	Scaffold ID	Scaffold Name	Genome ID	Genome	Gene Count	Sequence Length (bp)	GC Content	Perc VPFs	Viral Cluster
<input checked="" type="checkbox"/>	C1008267		700000619	Human buccal mucosa microbial communities from NIH_USA - visit 2_subject 763759525 (*) (MER-FS) (assembled)	11	6042	0.40	.45	vc_88807

# 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”



# 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 organized
- Database allows for analysis of viral genomics that 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>