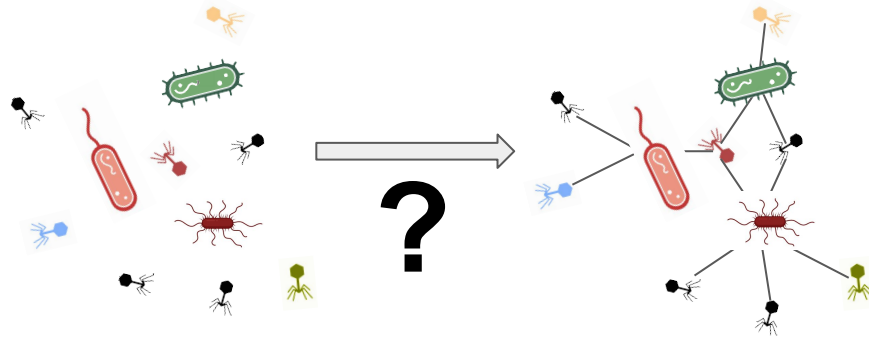


Recovering the missing links: the modeling of virus-host infection networks in silico

Gaylord (Eric) Bastien, PhD Candidate

Duhaime Lab, Ecology and Evolutionary Biology, University of Michigan
CS026 | Microbial ecology and physiology





Quick introduction to viral ecology and knowledge gap

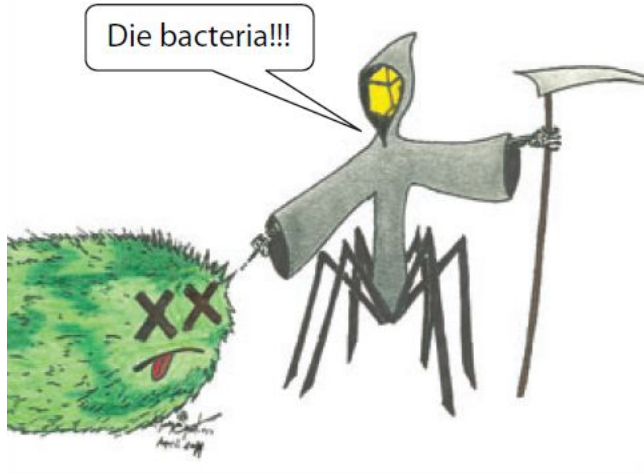


Predicting virus-host ecological interactions *in silico*



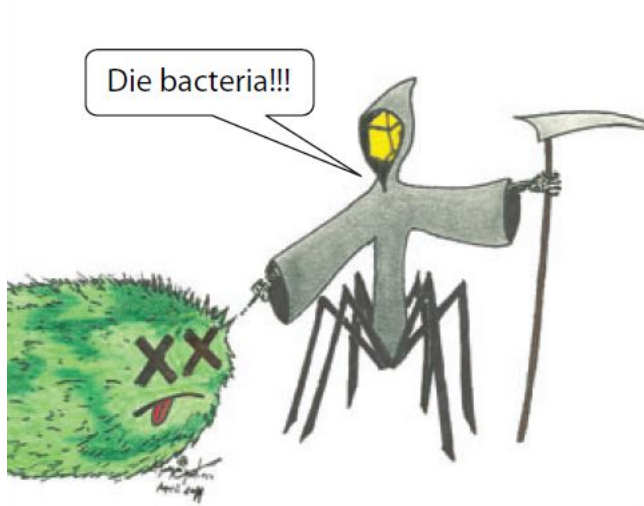
Applying network thinking to study virus-host infection networks

Viruses are central to the ecology and evolution of their hosts



Impact element cycles

Viruses are central to the ecology and evolution of their hosts

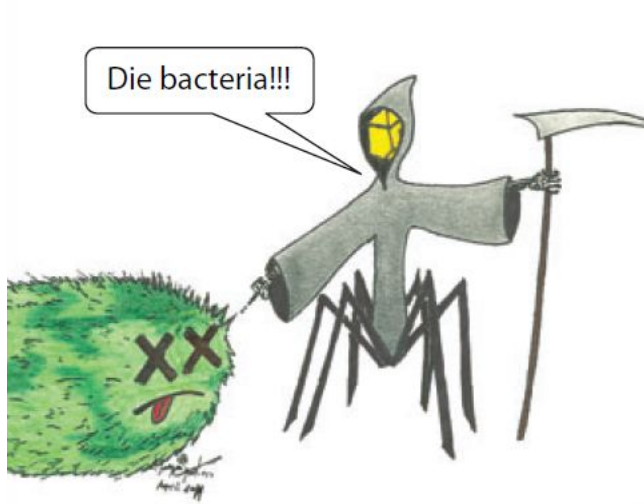


**Impact element
cycles**



**Impact microbial
metabolism**

Viruses are central to the ecology and evolution of their hosts



**Impact element
cycles**

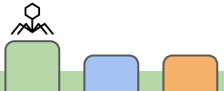
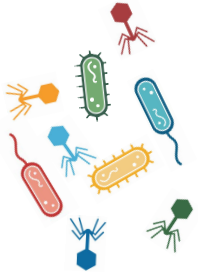


**Impact microbial
metabolism**

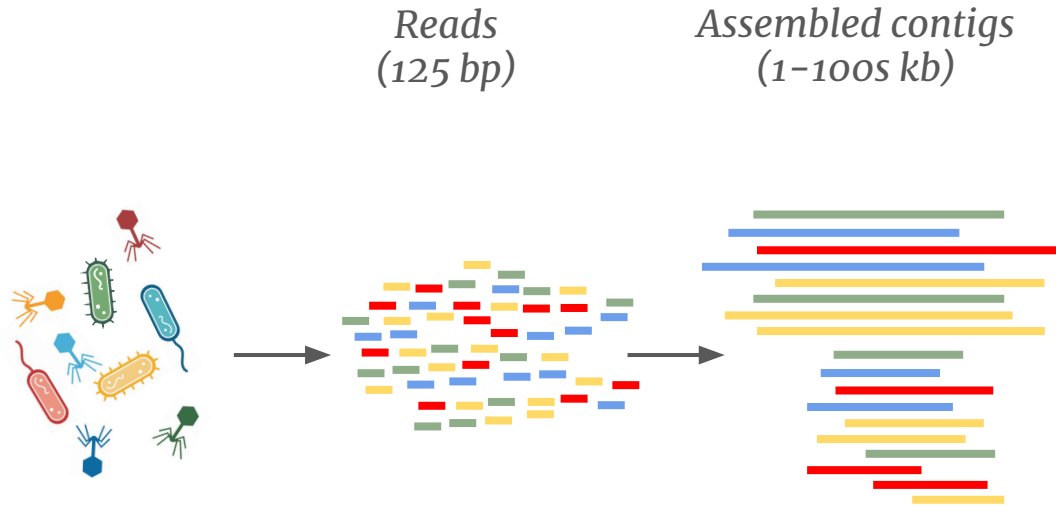


**Impact microbial
evolution**

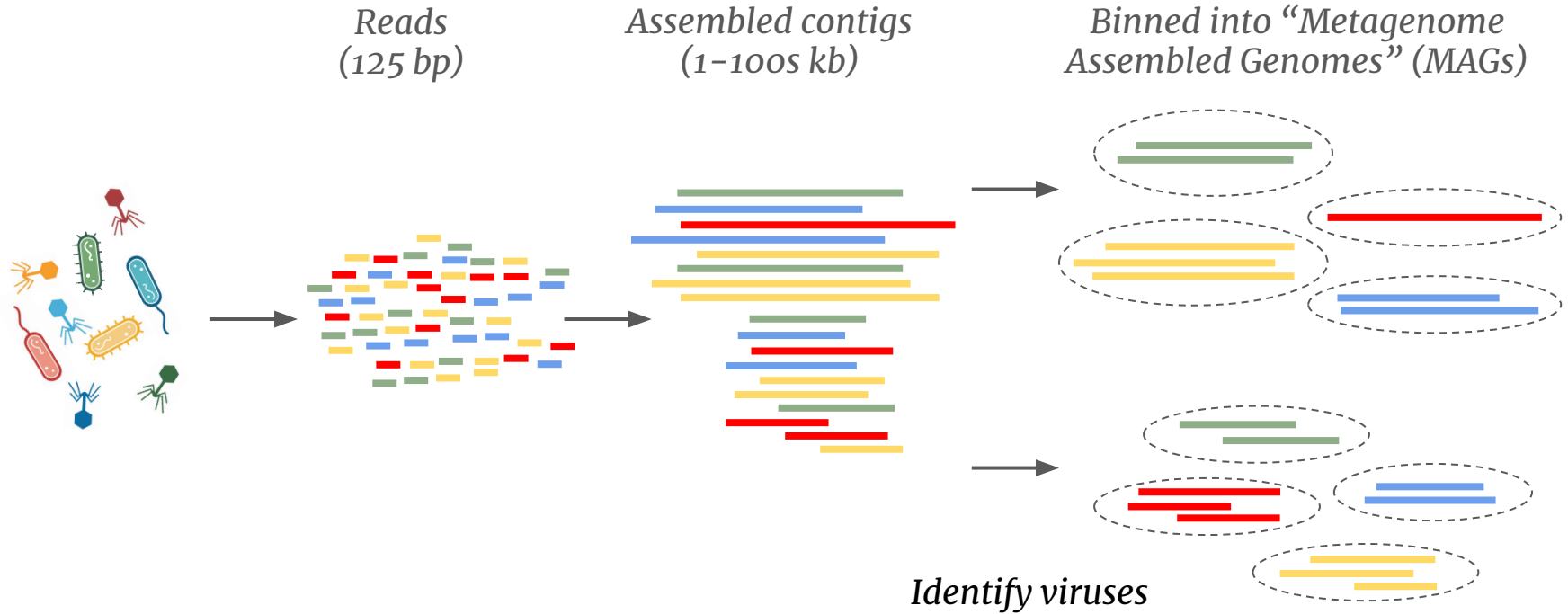
Metagenomics opened an avenue to study microbial diversity



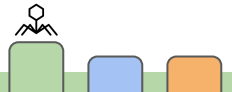
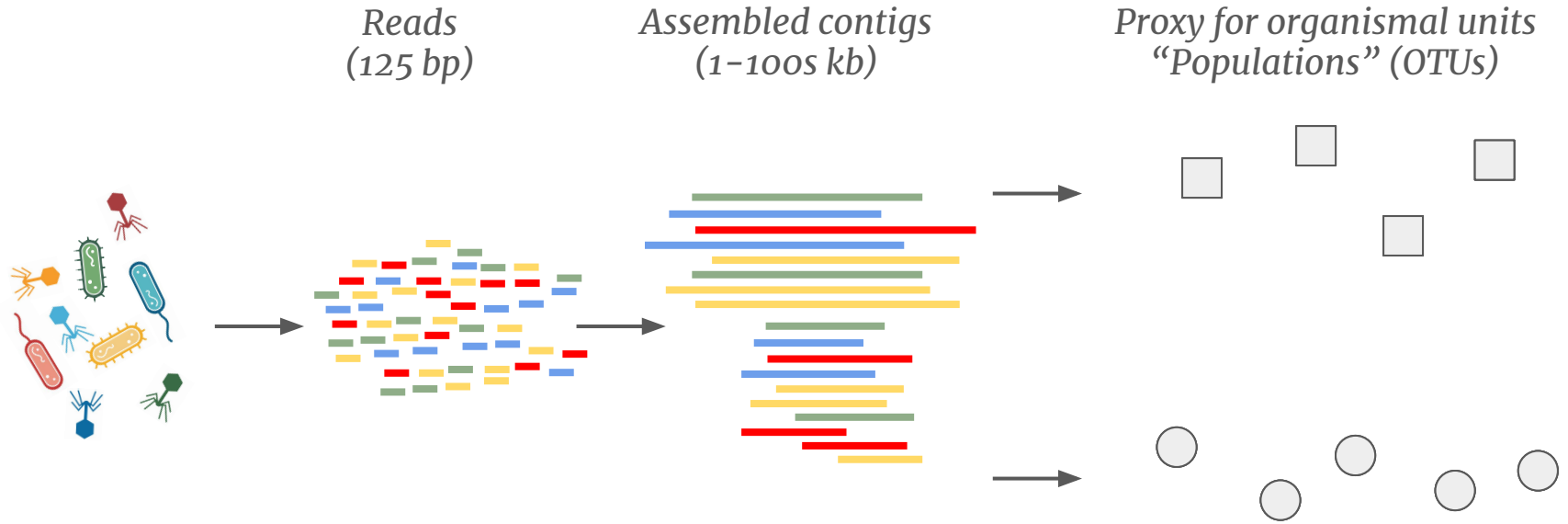
Metagenomics opened an avenue to study microbial diversity



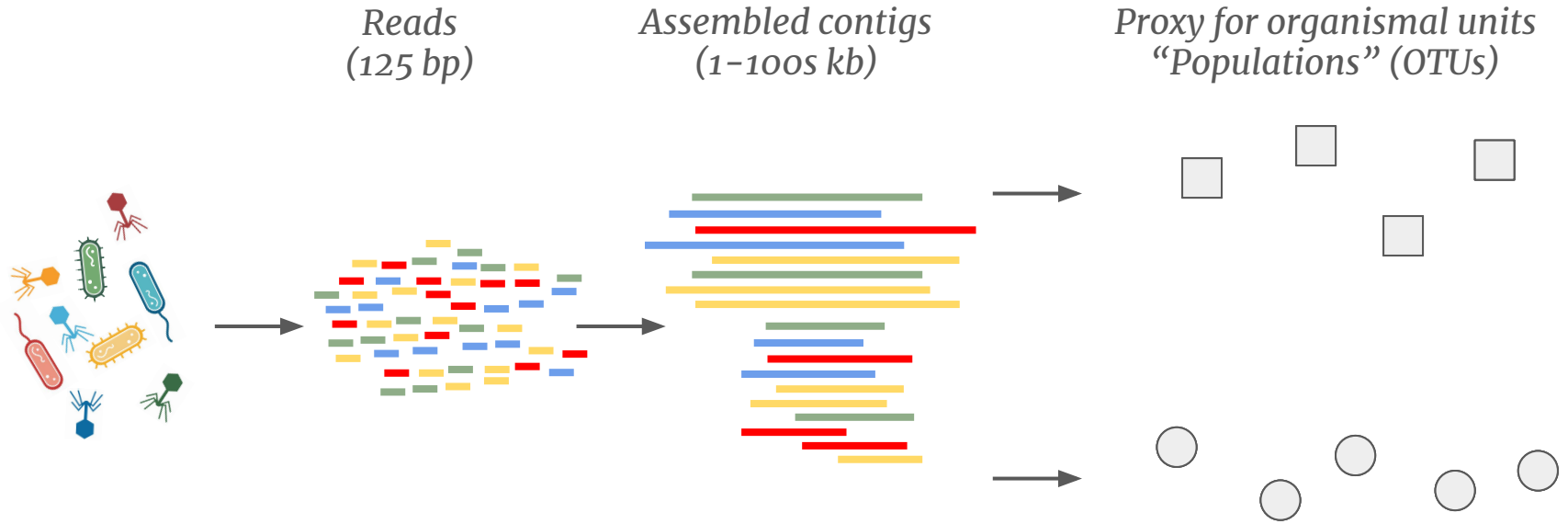
Metagenomics opened an avenue to study microbial diversity



Metagenomics opened an avenue to study microbial diversity



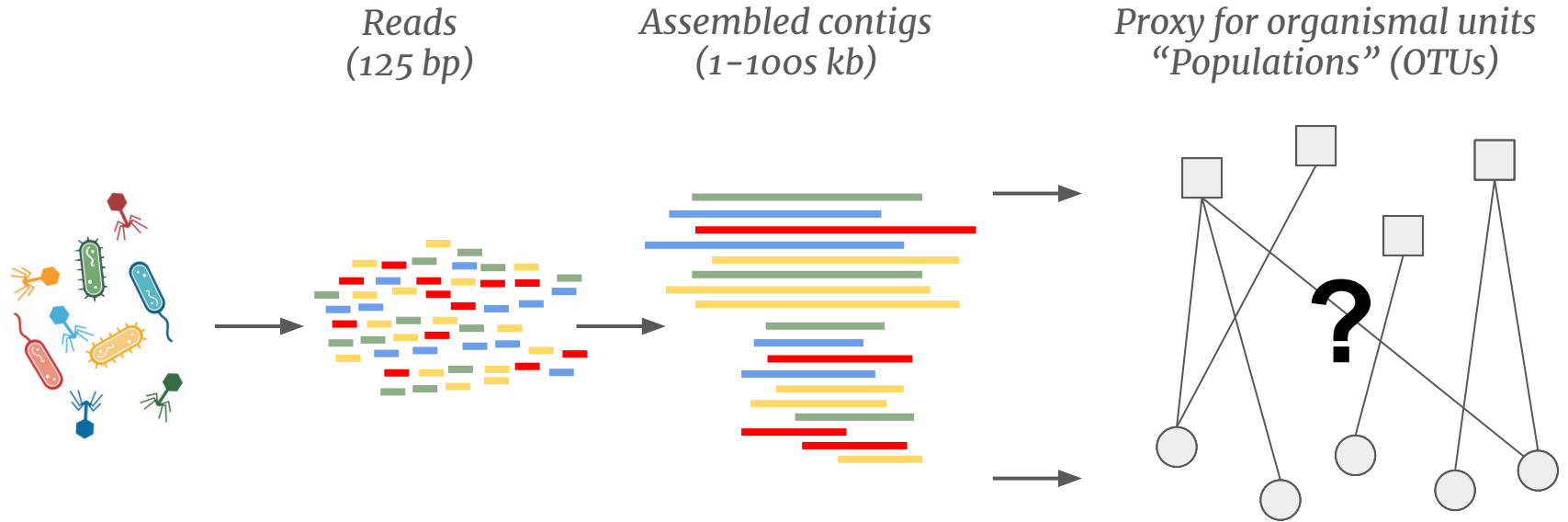
Metagenomics opened an avenue to study microbial diversity



Conservatively, 5 millions uncultivated viruses have been discovered



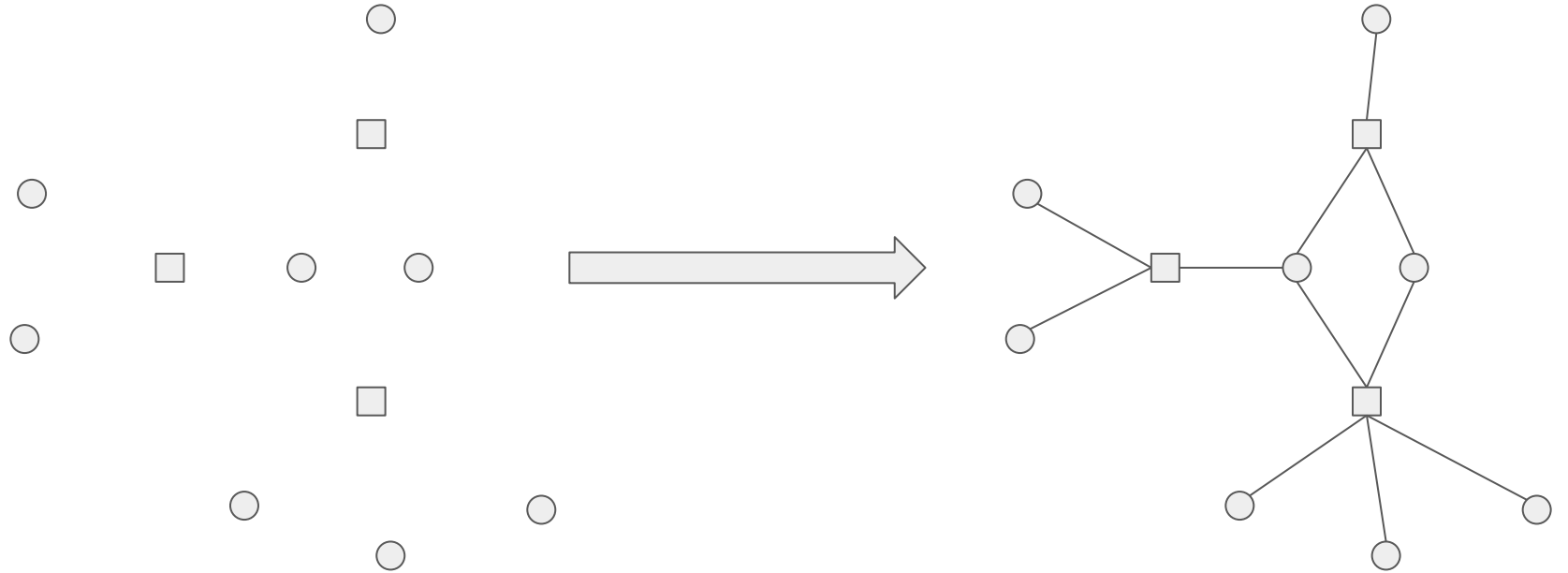
Metagenomics opened an avenue to study microbial diversity



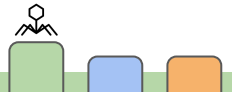
**Conservatively, 5 uncultivated millions
viruses have been discovered**



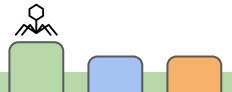
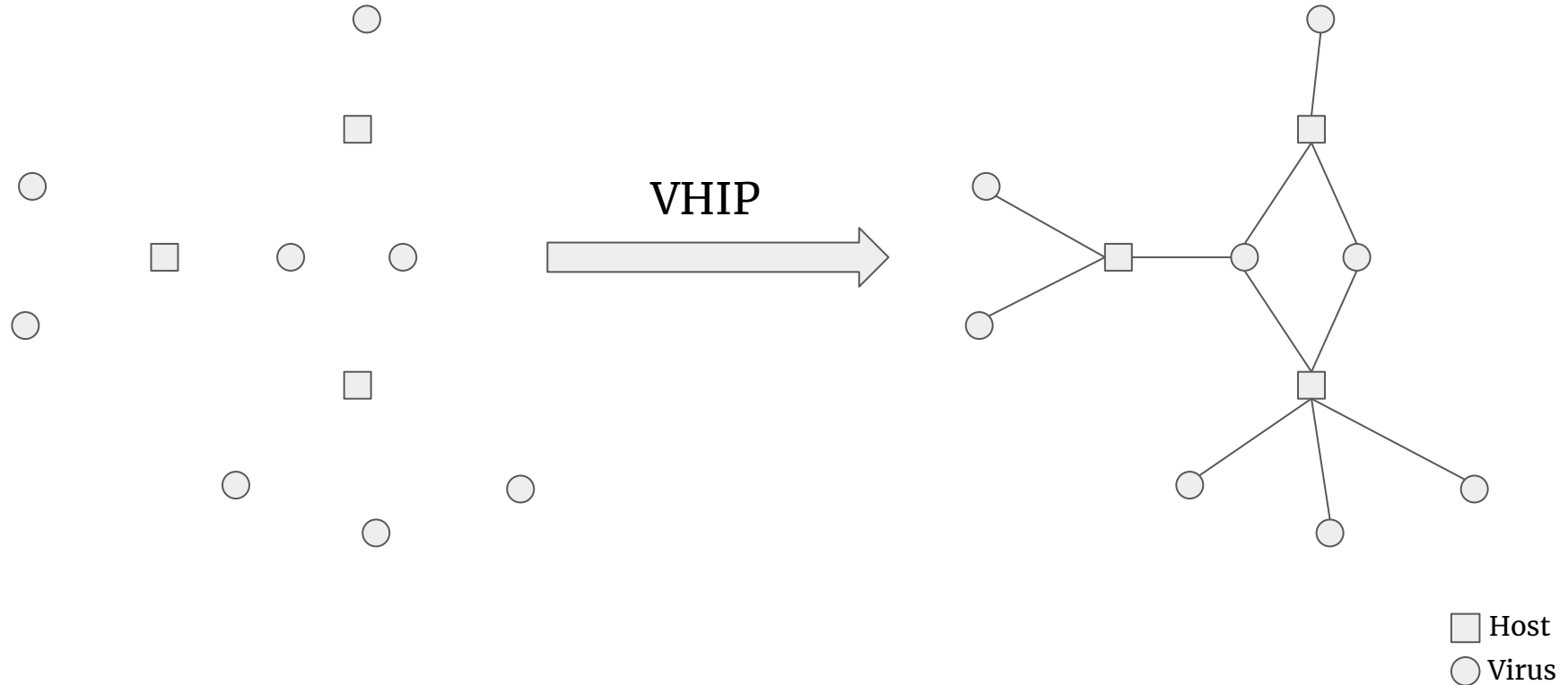
How to recover the missing links?



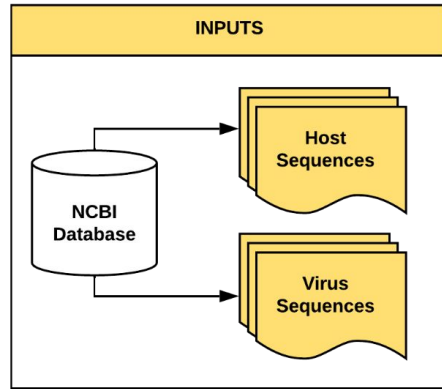
■ Host
○ Virus



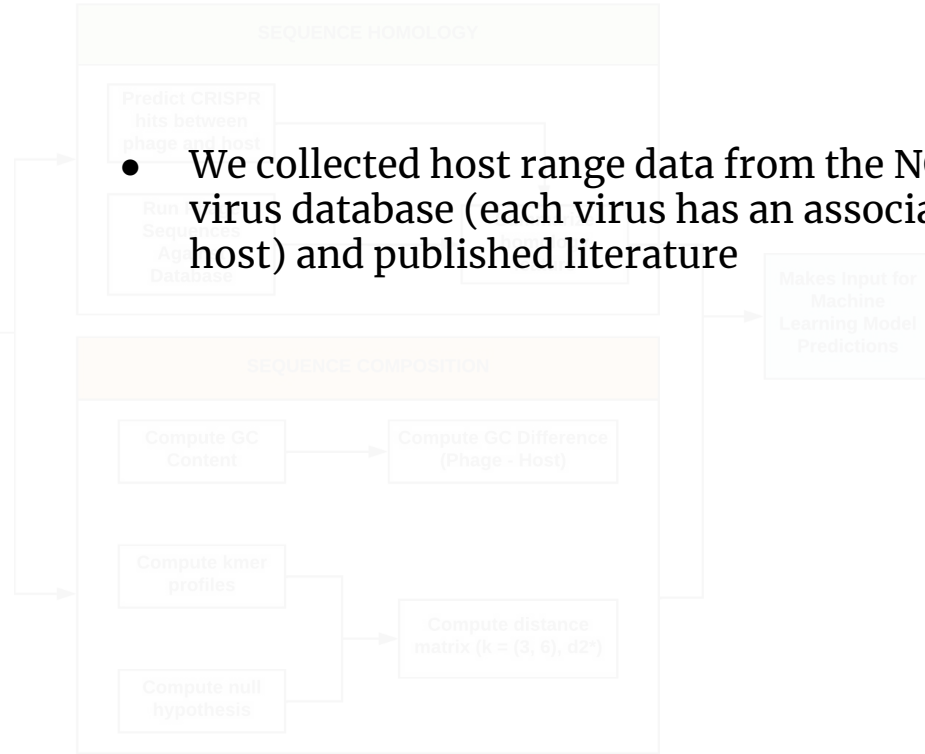
How to recover the missing links?



Design of Virus-Host Interaction Predictor (VHIP)



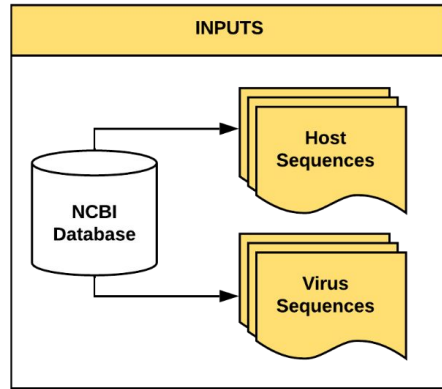
- We collected host range data from the NCBI virus database (each virus has an associated host) and published literature



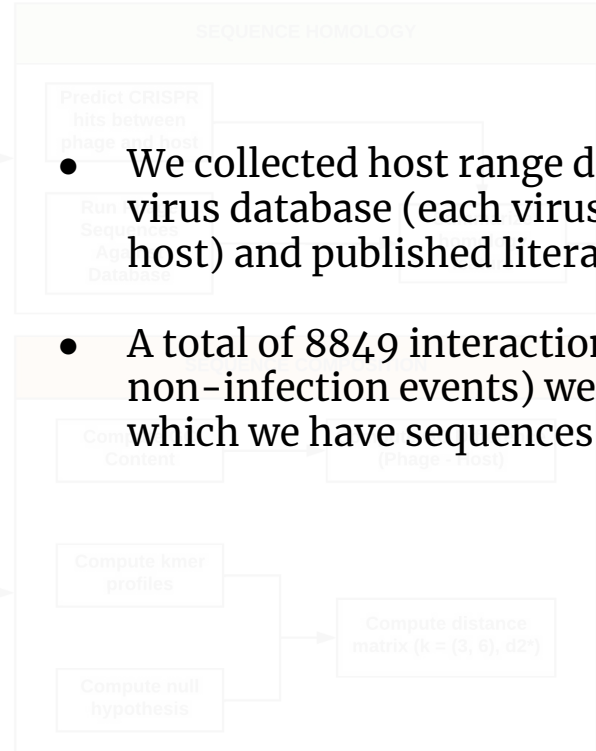
Cecelia
Batterbee



Design of Virus-Host Interaction Predictor (VHIP)



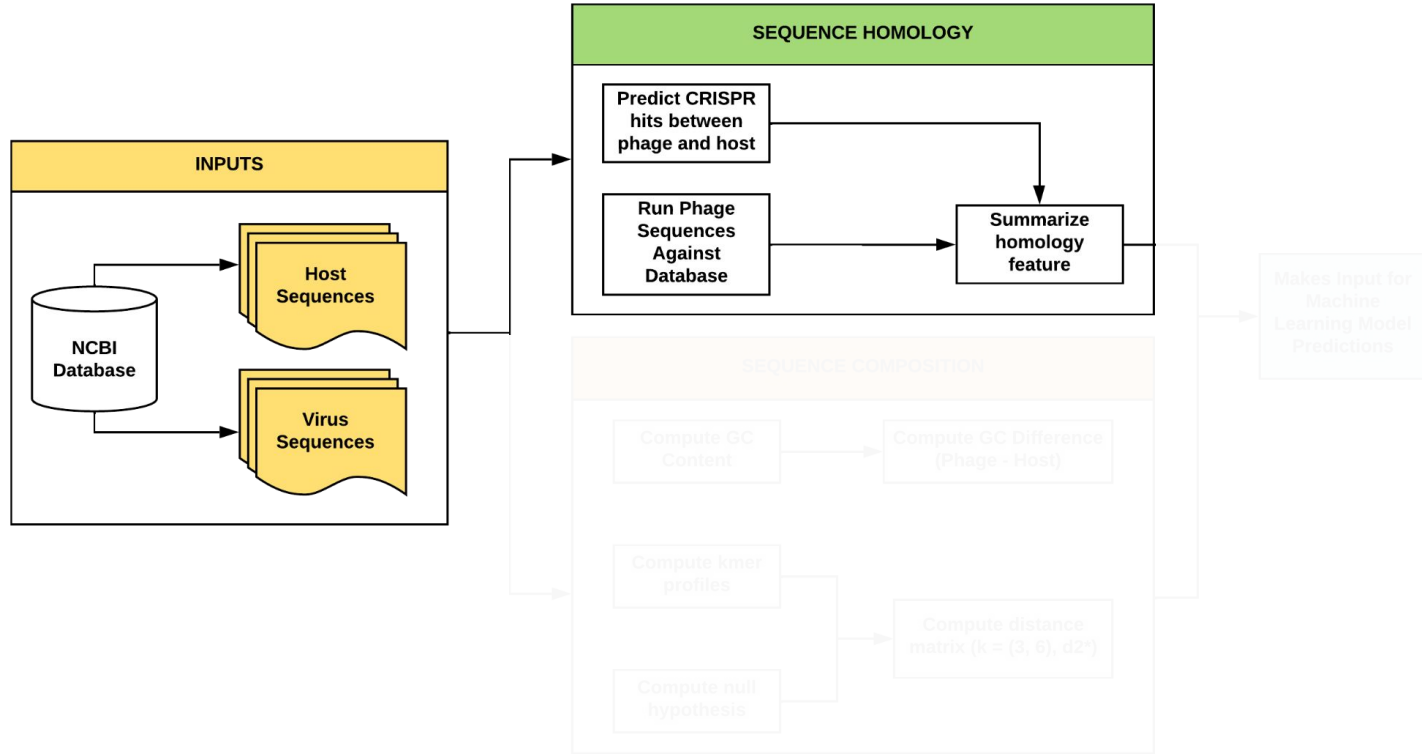
Cecelia
Batterbee



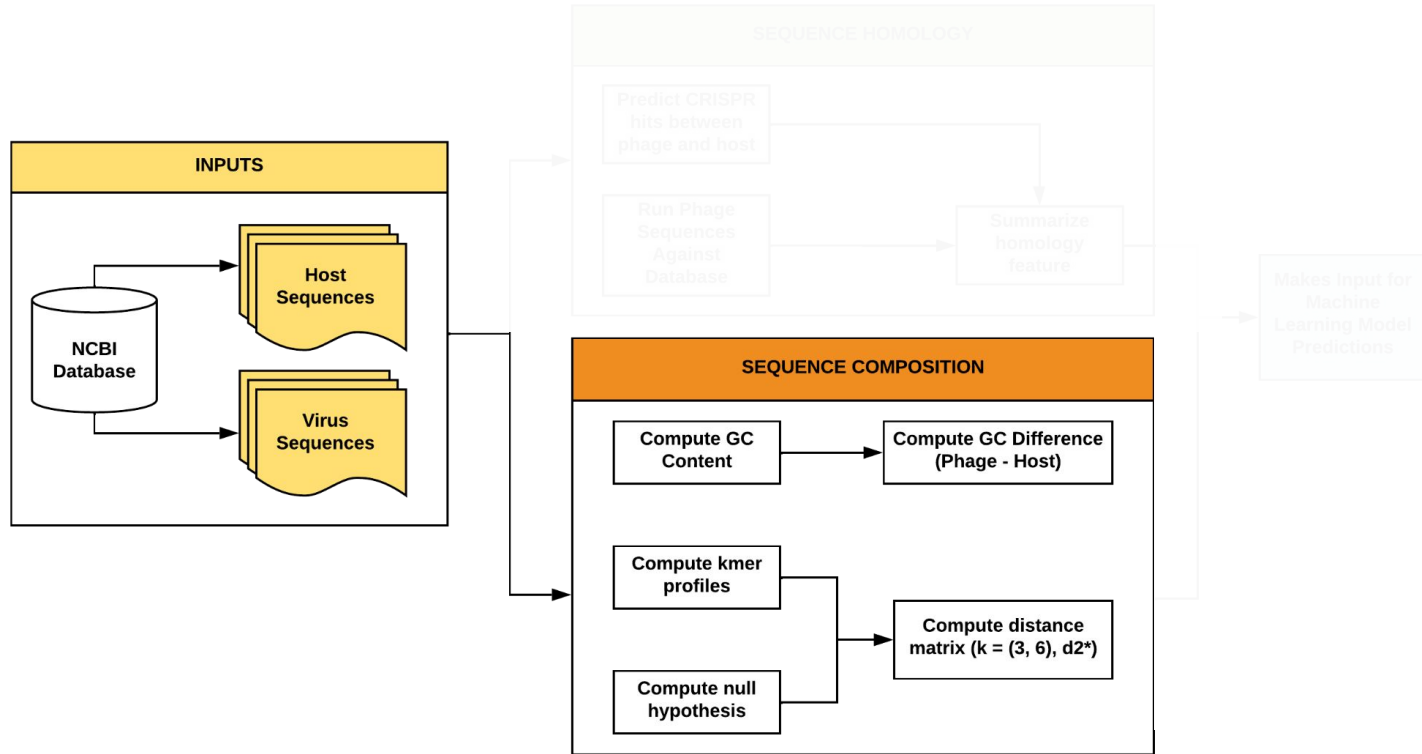
- We collected host range data from the NCBI virus database (each virus has an associated host) and published literature
- A total of 8849 interactions (both infection and non-infection events) were collected from which we have sequences for hosts and viruses



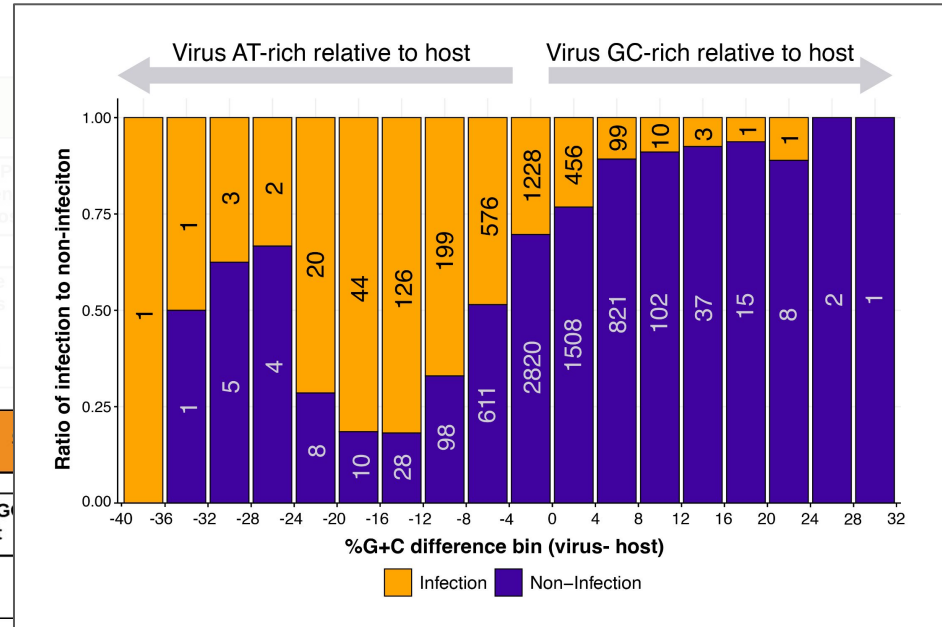
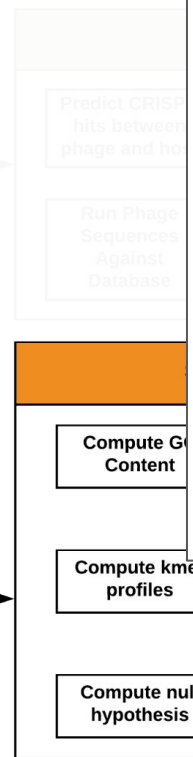
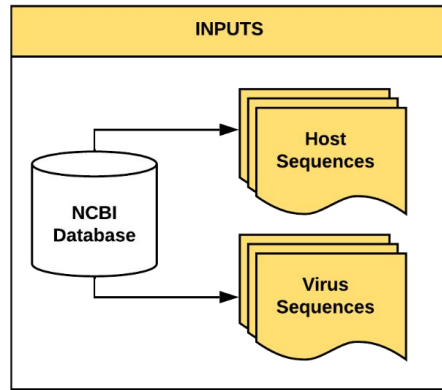
Design of Virus-Host Interaction Predictor (VHIP)



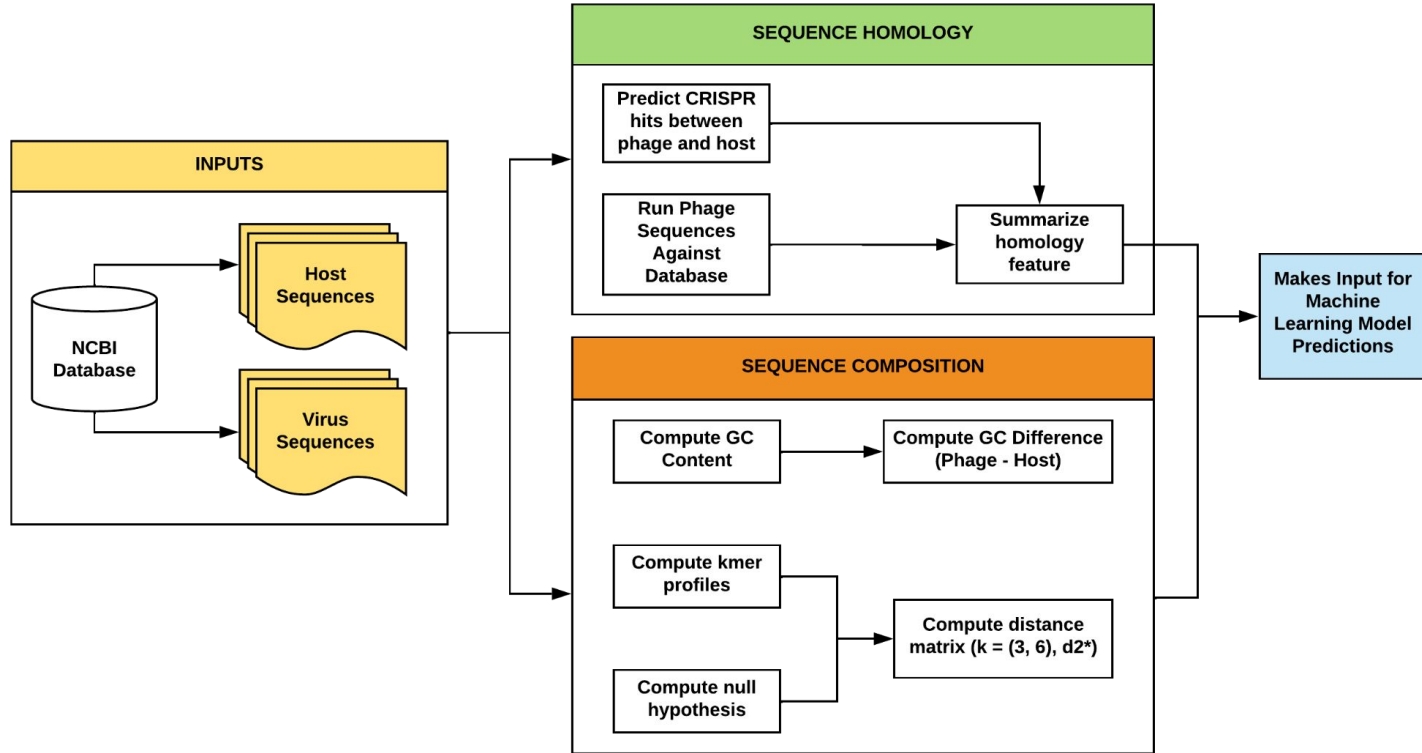
Design of Virus-Host Interaction Predictor (VHIP)



Design of Virus-Host Interaction Predictor (VHIP)



Design of Virus-Host Interaction Predictor (VHIP)



Evaluating the predictions made by VHIP

87% accuracy rate



Evaluating the predictions made by VHIP

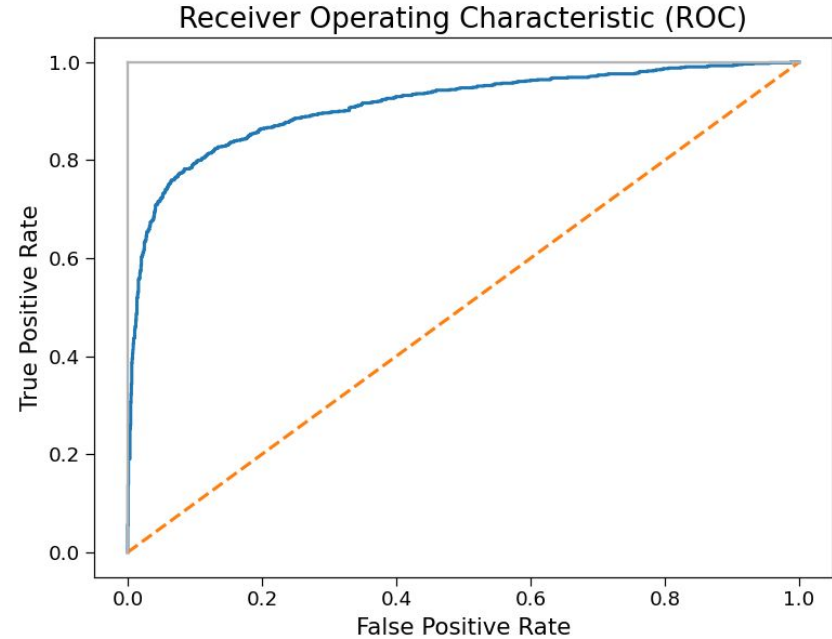
87% accuracy rate

Truth	Non-Infection	True Negative 2884 65.18%	False Positive 430 9.72%
	Infection	False Negative 140 3.16%	True Positive 971 21.94%
		Non-Infection	Infection
		Predicted	

Evaluating the predictions made by VHIP

87% accuracy rate

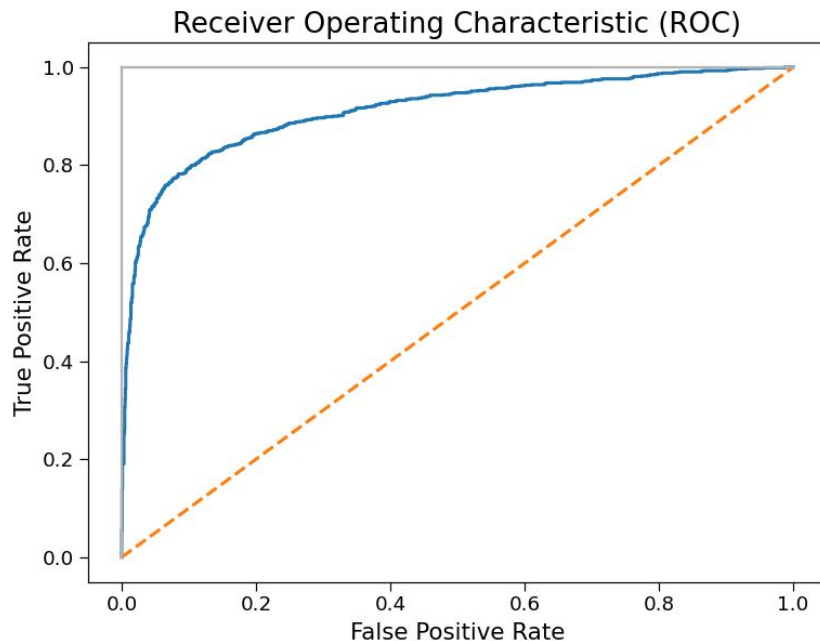
Truth	Non-Infection	True Negative 2884 65.18%	False Positive 430 9.72%
	Infection	False Negative 140 3.16%	True Positive 971 21.94%
		Non-Infection	Infection
		Predicted	



Evaluating the predictions made by VHIP

87% accuracy rate

Truth	Predicted	
	Non-Infection	Infection
Non-Infection	True Negative 2884 65.18%	False Positive 430 9.72%
Infection	False Negative 140 3.16%	True Positive 971 21.94%



*Model will be made available as a Python package on Github
in coming weeks*

Practical application of VHIP on two different microbial systems

Lake Erie



AJ Wing

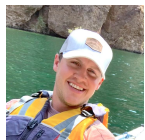


Satellite view of a harmful algal bloom in Lake Erie



Practical application of VHIP on two different microbial systems

Lake Erie



AJ Wing



Satellite view of a harmful algal bloom in Lake Erie

Prairie Potholes



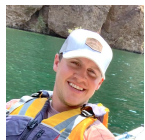
Aerial picture of the prairie Potholes

Dalcin Martins, P. et al. Viral and metabolic controls on high rates of microbial sulfur and carbon cycling in wetland ecosystems. *Microbiome* 6, 138 (2018).



Practical application of VHIP on two different microbial systems

Lake Erie



AJ Wing



Satellite view of a harmful algal bloom in Lake Erie

6408 viruses x 17 hosts = 108,936 possible interactions

Prairie Potholes



Aerial picture of the prairie Potholes

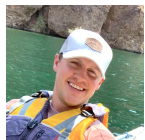
Dalcin Martins, P. et al. Viral and metabolic controls on high rates of microbial sulfur and carbon cycling in wetland ecosystems. *Microbiome* 6, 138 (2018).

2039 viruses x 26 hosts = 52,999 possible interactions



Practical application of VHIP on two different microbial systems

Lake Erie

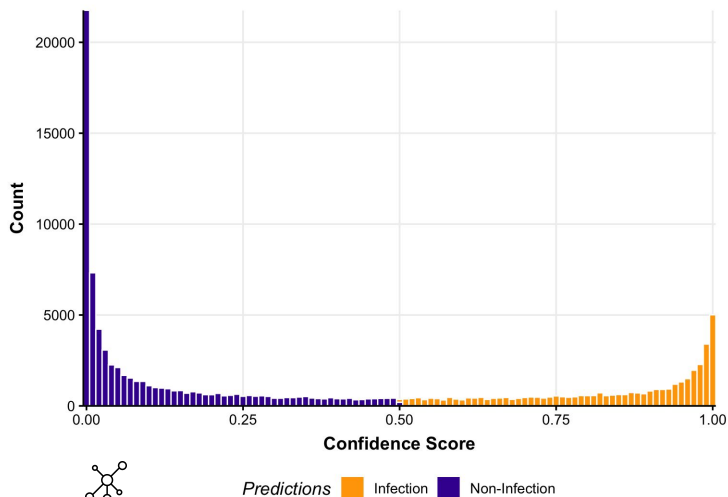


AJ Wing



Satellite view of a harmful algal bloom in Lake Erie

6408 viruses x 17 hosts = 108,936 possible interactions



Prairie Potholes



Aerial picture of the prairie Potholes

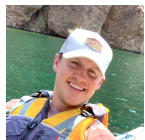
Dalcin Martins, P. et al. Viral and metabolic controls on high rates of microbial sulfur and carbon cycling in wetland ecosystems. *Microbiome* 6, 138 (2018).

2039 viruses x 26 hosts = 52,999 possible interactions



Practical application of VHIP on two different microbial systems

Lake Erie

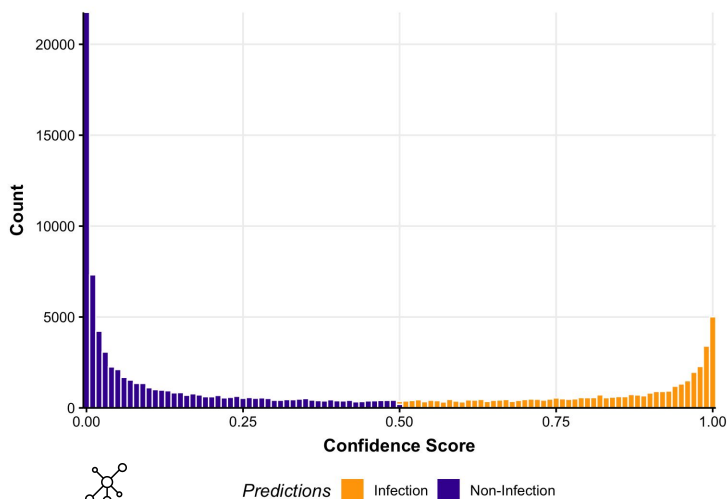


AJ Wing



Satellite view of a harmful algal bloom in Lake Erie

6408 viruses x 17 hosts = 108,936 possible interactions



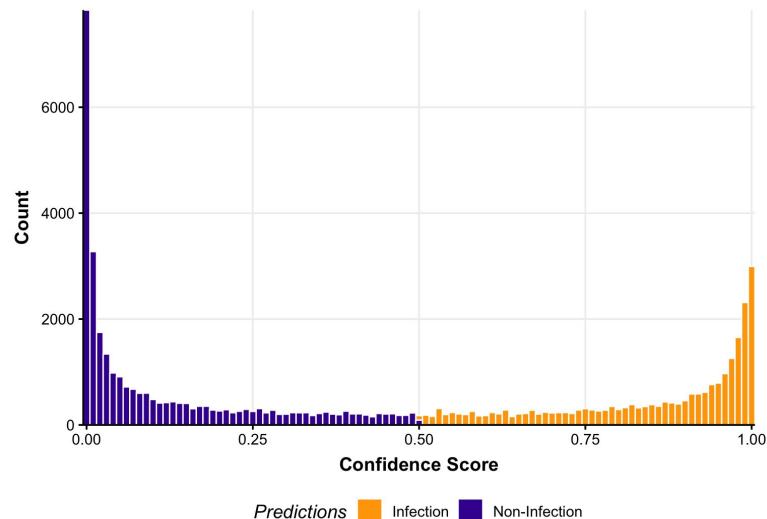
Prairie Potholes



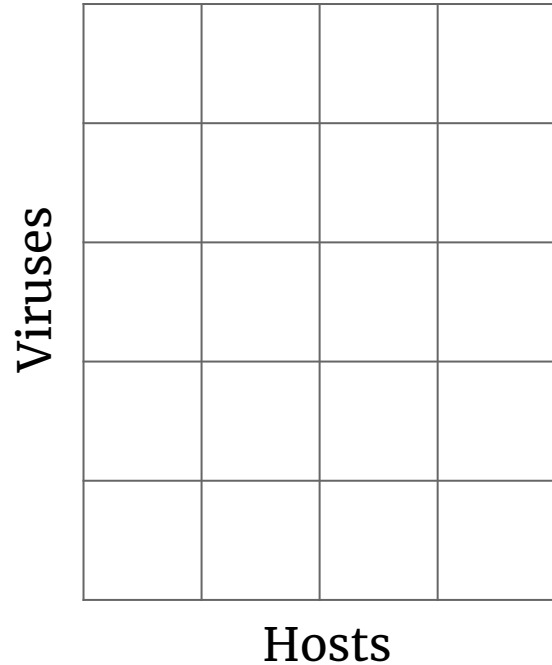
Aerial picture of the prairie Potholes

Dalcin Martins, P. et al. Viral and metabolic controls on high rates of microbial sulfur and carbon cycling in wetland ecosystems. *Microbiome* 6, 138 (2018).

2039 viruses x 26 hosts = 52,999 possible interactions



Practical application of model on two different study systems



- We can visualize the output of VHIP as a matrix, where each row represents a different virus and each column represents a different host



Practical application of model on two different study systems

Viruses	0.7	0.6	0.3	0.9
	0.9	0.2	0.8	0.9
	0.8	0.9	0.1	0.1
	0.9	0.8	0.2	0.1
	0.8	0.2	0.1	0.1
	Hosts			

- We can visualize the output of VHIP as a matrix, where each row represents a different virus and each column represents a different host
- Each cell then contains the confidence score value for that virus-host pair



Practical application of model on two different study systems

Viruses	0.7	0.6	0.3	0.9
	0.9	0.2	0.8	0.9
	0.8	0.9	0.1	0.1
	0.9	0.8	0.2	0.1
	0.8	0.2	0.1	0.1
	Hosts			

- We can visualize the output of VHIP as a matrix, where each row represents a different virus and each column represents a different host
- Each cell then contains the confidence score value for that virus-host pair
 - We can color the cell based on the score value.
 - *Above 0.5 is infection*
 - *Below 0.5 is non-infection*



Practical application of model on two different study systems

Viruses	0.7	0.6	0.3	0.9
	0.9	0.2	0.8	0.9
	0.8	0.9	0.1	0.1
	0.9	0.8	0.2	0.1
	0.8	0.2	0.1	0.1
	Hosts			

- We can visualize the output of VHIP as a matrix, where each row represents a different virus and each column represents a different host
- Each cell then contains the confidence score value for that virus-host pair
 - We can color the cell based on the score value.
 - *Above 0.5 is infection*
 - *Below 0.5 is non-infection*



Practical application of model on two different study systems

Viruses	0.7	0.6	0.3	0.9
	0.9	0.2	0.8	0.9
	0.8	0.9	0.1	0.1
	0.9	0.8	0.2	0.1
	0.8	0.2	0.1	0.1
	Hosts			

- We can visualize the output of VHIP as a matrix, where each row represents a different virus and each column represents a different host
- Each cell then contains the confidence score value for that virus-host pair
 - We can color the cell based on the score value.
 - *Above 0.5 is infection*
 - *Below 0.5 is non-infection*



Practical application of model on two different study systems

Viruses	0.7	0.6	0.3	0.9
	0.9	0.2	0.8	0.9
	0.8	0.9	0.1	0.1
	0.9	0.8	0.2	0.1
	0.8	0.2	0.1	0.1
	Hosts			

- We can visualize the output of VHIP as a matrix, where each row represents a different virus and each column represents a different host
- Each cell then contains the confidence score value for that virus-host pair
 - We can color the cell based on the score value.
 - *Above 0.5 is infection*
 - *Below 0.5 is non-infection*



Practical application of model on two different study systems

Viruses

0.7	0.6	0.3	0.9
0.9	0.2	0.8	0.9
0.8	0.9	0.1	0.1
0.9	0.8	0.2	0.1
0.8	0.2	0.1	0.1

Hosts

- We can visualize the output of VHIP as a matrix, where each row represents a different virus and each column represents a different host
- Each cell then contains the confidence score value for that virus-host pair
 - We can color the cell based on the score value.
 - *Above 0.5 is infection*
 - *Below 0.5 is non-infection*



Practical application of model on two different study systems

Viruses

0.7	0.6	0.3	0.9
0.9	0.2	0.8	0.9
0.8	0.9	0.1	0.1
0.9	0.8	0.2	0.1
0.8	0.2	0.1	0.1

Hosts

- We can visualize the output of VHIP as a matrix, where each row represents a different virus and each column represents a different host
- Each cell then contains the confidence score value for that virus-host pair
 - We can color the cell based on the score value.
 - *Above 0.5 is infection*
 - *Below 0.5 is non-infection*



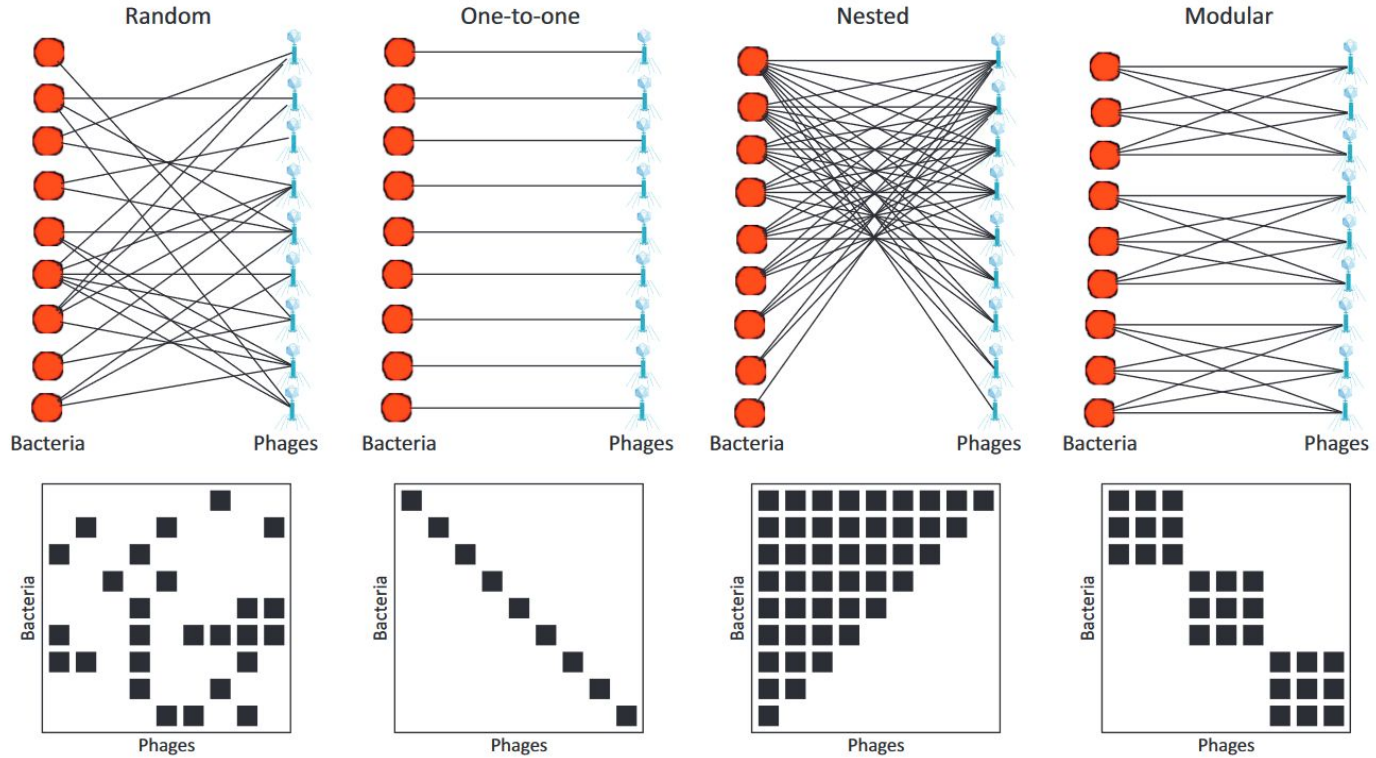
Practical application of model on two different study systems

Viruses	0.7	0.6	0.3	0.9
	0.9	0.2	0.8	0.9
	0.8	0.9	0.1	0.1
	0.9	0.8	0.2	0.1
	0.8	0.2	0.1	0.1
	Hosts			

- We can visualize the output of VHIP as a matrix, where each row represents a different virus and each column represents a different host
- Each cell then contains the confidence score value for that virus-host pair
 - We can color the cell based on the score value.
 - *Above 0.5 is infection*
 - *Below 0.5 is non-infection*
- Rows and columns can be reorganized
 - Rows are organized by descending number of host a virus is predicted to infect
 - Columns are organized by descending susceptibility



Practical application of model on two different study systems



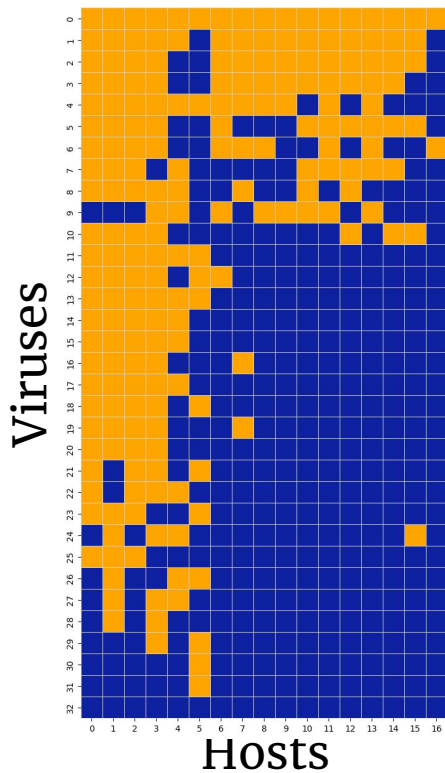
Weitz, J. S. *et al.* Phage–bacteria infection networks. *Trends in Microbiology* **21**, 82–91 (2013).

Hosts



Practical application of model on two different study systems

Lake Erie

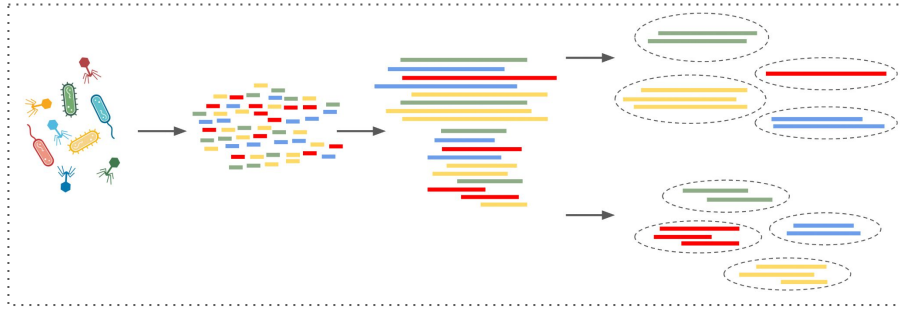


Prairie Potholes



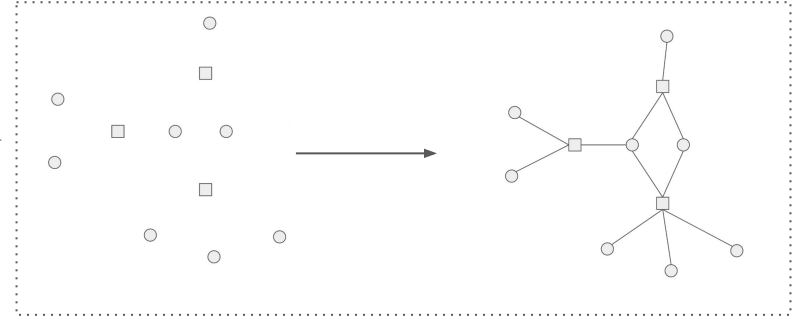
A framework to study virus-host infection networks

Metagenomics pipeline



Recover viral and microbial populations

VHIP output



Predict ecological interactions between viral and microbial populations



Thank you!

Advisors

Dr. Melissa Duhaime Dr. Luis Zaman



Labmates



AJ Wing



Jessica
Choi



Rachel
Cable



Lizy
Michaelson



Morgan
Lindback



Cecelia
Batterbee



gbastien@umich.edu

 [@GEricBastien](https://twitter.com/GEricBastien)

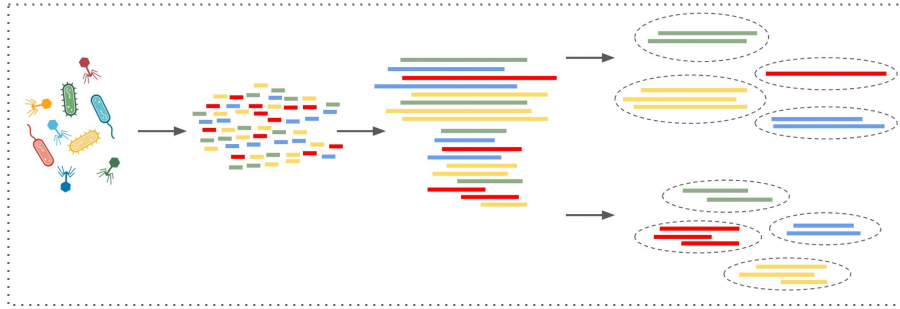
Funding



Polar Programs
#2055455

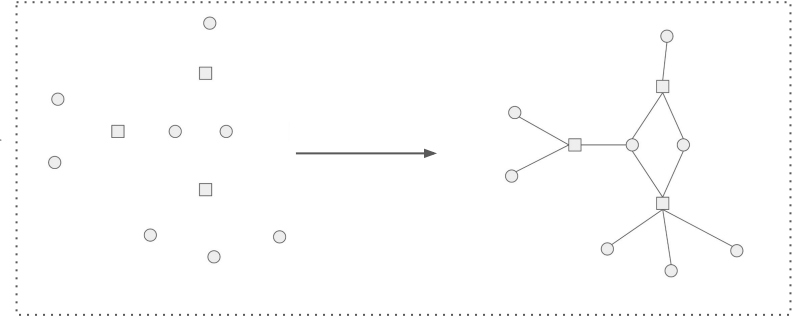
A framework to study virus-host infection networks

Metagenomics pipeline



Recover viral and microbial populations

VHIP output



Predict ecological interactions between viral and microbial populations

