

# Is Sharing Caring?

Elucidating the Effects of the  
Presence of CRISPR-Cas Systems  
on Rates of Horizontal Gene  
Transfer Using Network Analysis

Siddharth Reed  
MolBiol 4C12 Thesis



Golding Lab,  
Biology Department,  
McMaster University

April 3, 2019

# Table of Contents

---

1. Background
2. My Project
3. Results
4. Conclusion

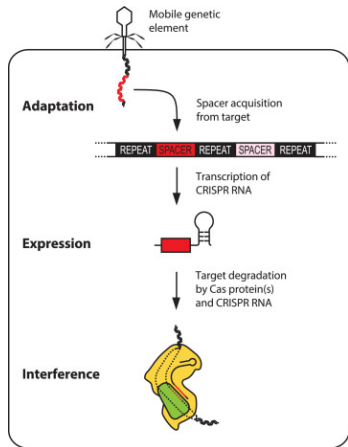
# CRISPR-Cas systems

# What Are They?

---

# What Are They?

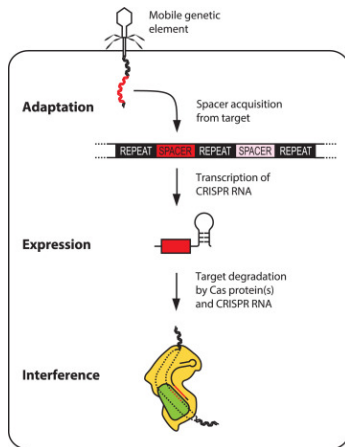
- Adaptive Bacterial Immune System



(Rath et al., 2015)

# What Are They?

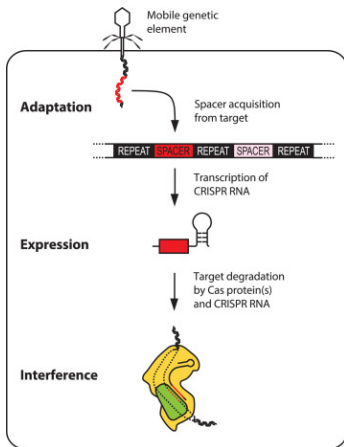
- Adaptive Bacterial Immune System
- Failed “infection” → spacer acquisition → targeted degradation for next “infection”



(Rath et al., 2015)

# What Are They?

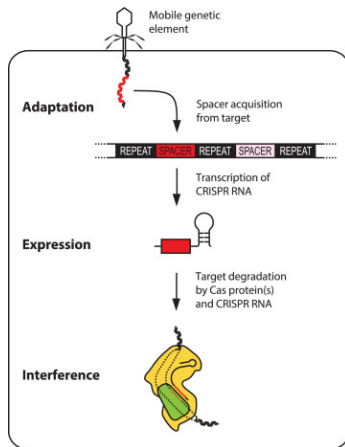
- Adaptive Bacterial Immune System
- Failed “infection” → spacer acquisition → targeted degradation for next “infection”
- Protects against foreign DNA



(Rath et al., 2015)

# What Are They?

- Adaptive Bacterial Immune System
- Failed “infection” → spacer acquisition → targeted degradation for next “infection”
- Protects against foreign DNA
- Requires Cas proteins and CRISPR loci

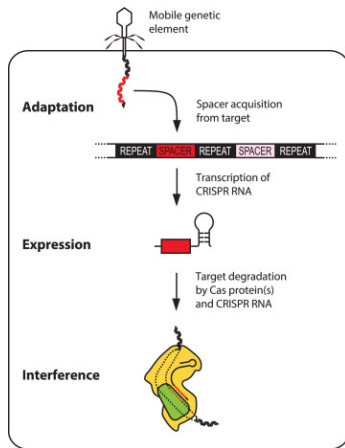


(Rath et al., 2015)



# What Are They?

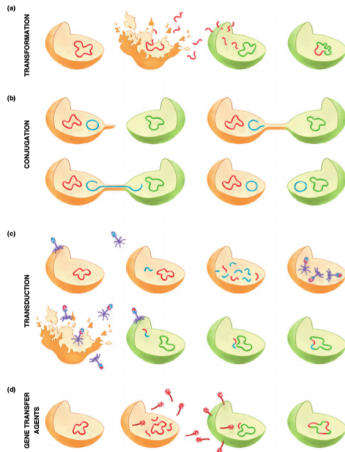
- Adaptive Bacterial Immune System
- Failed “infection” → spacer acquisition → targeted degradation for next “infection”
- Protects against foreign DNA
- Requires Cas proteins and CRISPR loci
- 45% of bacteria have CRISPR loci ( $n = 6782$ ) (Grissa, I. and Drevet, C. and Couvin, D., 2017)



(Rath et al., 2015)

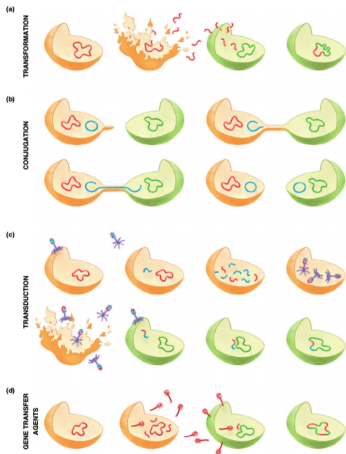
# Horizontal Gene Transfer

# Horizontal Gene Transfer Mechanisms



(Popa and Dagan, 2011)

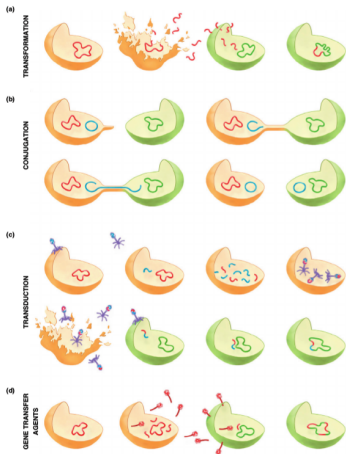
# Horizontal Gene Transfer Mechanisms



- Transformation: Incorporation of free-floating DNA into the genome (Popa and Dagan, 2011)

(Popa and Dagan, 2011)

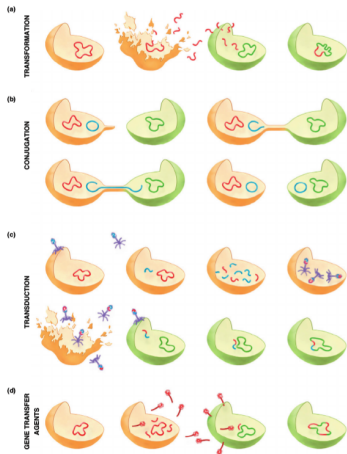
# Horizontal Gene Transfer Mechanisms



- Transformation: Incorporation of free-floating DNA into the genome (Popa and Dagan, 2011)
- Conjugation: Transfer of DNA through cell-cell connections (Popa and Dagan, 2011)

(Popa and Dagan, 2011)

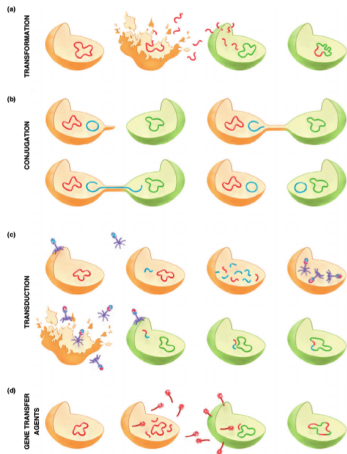
# Horizontal Gene Transfer Mechanisms



(Popa and Dagan, 2011)

- Transformation: Incorporation of free-floating DNA into the genome (Popa and Dagan, 2011)
- Conjugation: Transfer of DNA through cell-cell connections (Popa and Dagan, 2011)
- Transduction: Transfer of DNA through phage (Popa and Dagan, 2011)

# Horizontal Gene Transfer Mechanisms



(Popa and Dagan, 2011)

- Transformation: Incorporation of free-floating DNA into the genome (Popa and Dagan, 2011)
- Conjugation: Transfer of DNA through cell-cell connections (Popa and Dagan, 2011)
- Transduction: Transfer of DNA through phage (Popa and Dagan, 2011)
- **CRISPR-Cas directly affects HGT** (Popa and Dagan, 2011)

# Rate Influencing Factors

---



# Rate Influencing Factors

---

- Amount of exogenous DNA/cell density/phage density

# Rate Influencing Factors

---

- Amount of exogenous DNA/cell density/phage density
- Selective pressures

# Rate Influencing Factors

---

- Amount of exogenous DNA/cell density/phage density
- Selective pressures
- Metabolic costs

## Rate Influencing Factors

---

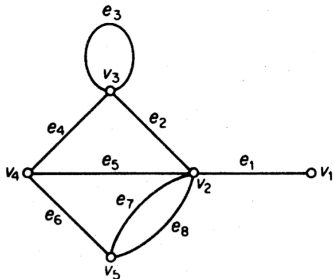
- Amount of exogenous DNA/cell density/phage density
- Selective pressures
- Metabolic costs
- Sequence compatibility

# Phylogenomic Networks

# What is A Network?

---

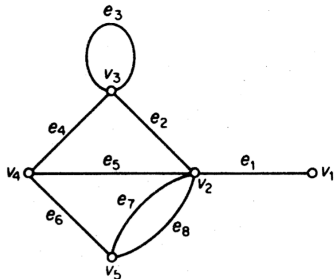
# What is A Network?



- Useful mathematical abstraction of real world system

(Bondy and Murty, 2002)

# What is A Network?

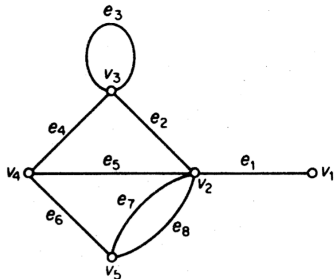


- Useful mathematical abstraction of real world system
- Nodes can have attributes

(Bondy and Murty, 2002)



# What is A Network?



- Useful mathematical abstraction of real world system
- Nodes can have attributes
- Edges can have weights

(Bondy and Murty, 2002)

# Do CRISPR Systems Affect Horizontal Gene Transfer?

Yes

# CRISPR Cost Complexity and Curbing It

# CRISPR Cost Complexity and Curbing It

---

- Cost trade off factors:
  - Cost Reduction Strategies

# CRISPR Cost Complexity and Curbing It

---

- Cost trade off factors:
  - Metabolic maintenance (Rath et al., 2015)
- Cost Reduction Strategies

# CRISPR Cost Complexity and Curbing It

---

- Cost trade off factors:
  - Metabolic maintenance (Rath et al., 2015)
  - Off-target effects (autoimmune) (Stern et al., 2010)
- Cost Reduction Strategies

# CRISPR Cost Complexity and Curbing It

---

- Cost trade off factors:
  - Metabolic maintenance (Rath et al., 2015)
  - Off-target effects (autoimmune) (Stern et al., 2010)
  - Environmental pressures (Dzidic and Bedeković, 2003)
- Cost Reduction Strategies



# CRISPR Cost Complexity and Curbing It

---

- Cost trade off factors:
  - Metabolic maintenance (Rath et al., 2015)
  - Off-target effects (autoimmune) (Stern et al., 2010)
  - Environmental pressures (Dzidic and Bedeković, 2003)
- Cost Reduction Strategies
  - Selective CRISPR inactivation (Rath et al., 2015)

# CRISPR Cost Complexity and Curbing It

---

- Cost trade off factors:
  - Metabolic maintenance (Rath et al., 2015)
  - Off-target effects (autoimmune) (Stern et al., 2010)
  - Environmental pressures (Dzidic and Bedeković, 2003)
  - Phage virulence/density (Bondy-Denomy and Davidson, 2014)
- Cost Reduction Strategies
  - Selective CRISPR inactivation (Rath et al., 2015)

# CRISPR Cost Complexity and Curbing It

---

- Cost trade off factors:
  - Metabolic maintenance (Rath et al., 2015)
  - Off-target effects (autoimmune) (Stern et al., 2010)
  - Environmental pressures (Dzidic and Bedeković, 2003)
  - Phage virulence/density (Bondy-Denomy and Davidson, 2014)
  - Anti-CRISPR systems (Bondy-Denomy and Davidson, 2014)
- Cost Reduction Strategies
  - Selective CRISPR inactivation (Rath et al., 2015)

# CRISPR Cost Complexity and Curbing It

- Cost trade off factors:
  - Metabolic maintenance (Rath et al., 2015)
  - Off-target effects (autoimmune) (Stern et al., 2010)
  - Environmental pressures (Dzidic and Bedeković, 2003)
  - Phage virulence/density (Bondy-Denomy and Davidson, 2014)
  - Anti-CRISPR systems (Bondy-Denomy and Davidson, 2014)
- Cost Reduction Strategies
  - Selective CRISPR inactivation (Rath et al., 2015)
  - CRISPRs themselves can be transferred  $\implies$  population level immunity (Godde and Bickerton, 2006)

# CRISPR Cost Complexity and Curbing It

- Cost trade off factors:
  - Metabolic maintenance (Rath et al., 2015)
  - Off-target effects (autoimmune) (Stern et al., 2010)
  - Environmental pressures (Dzidic and Bedeković, 2003)
  - Phage virulence/density (Bondy-Denomy and Davidson, 2014)
  - Anti-CRISPR systems (Bondy-Denomy and Davidson, 2014)
  - Prophage abundance (Watson, Staals, and Fineran, 2018)
- Cost Reduction Strategies
  - Selective CRISPR inactivation (Rath et al., 2015)
  - CRISPRs themselves can be transferred  $\implies$  population level immunity (Godde and Bickerton, 2006)

# CRISPR Cost Complexity and Curbing It

---

- Cost trade off factors:
  - Metabolic maintenance (Rath et al., 2015)
  - Off-target effects (autoimmune) (Stern et al., 2010)
  - Environmental pressures (Dzidic and Bedeković, 2003)
  - Phage virulence/density (Bondy-Denomy and Davidson, 2014)
  - Anti-CRISPR systems (Bondy-Denomy and Davidson, 2014)
  - Prophage abundance (Watson, Staals, and Fineran, 2018)
- Cost Reduction Strategies
  - Selective CRISPR inactivation (Rath et al., 2015)
  - CRISPRs themselves can be transferred  $\implies$  population level immunity (Godde and Bickerton, 2006)
  - CRISPR can enhance transduction-mediated HGT (Watson, Staals, and Fineran, 2018)

## Previous Findings

---

## Previous Findings

---

- Gophna et al. (2015) found no relation between the presence of CRISPR systems and HGT over short evolutionary timescales



## Previous Findings

---

- Gophna et al. (2015) found no relation between the presence of CRISPR systems and HGT over short evolutionary timescales
  - Assume all singletons arose from HGT

## Previous Findings

---

- Gophna et al. (2015) found no relation between the presence of CRISPR systems and HGT over short evolutionary timescales
  - Assume all singletons arose from HGT
  - Used GC% to identify HGT

## Previous Findings

---

- Gophna et al. (2015) found no relation between the presence of CRISPR systems and HGT over short evolutionary timescales
  - Assume all singletons arose from HGT
  - Used GC% to identify HGT
- Contradicted by a former undergraduate thesis student

## Previous Findings

---

- Gophna et al. (2015) found no relation between the presence of CRISPR systems and HGT over short evolutionary timescales
  - Assume all singletons arose from HGT
  - Used GC% to identify HGT
- Contradicted by a former undergraduate thesis student
  - Can see inhibitory effects of CRISPR on HGT over short evolutionary time scales

## Previous Findings

---

- Gophna et al. (2015) found no relation between the presence of CRISPR systems and HGT over short evolutionary timescales
  - Assume all singletons arose from HGT
  - Used GC% to identify HGT
- Contradicted by a former undergraduate thesis student
  - Can see inhibitory effects of CRISPR on HGT over short evolutionary time scales
  - Higher gene indel rates for CRISPR containing OTUs than non-CRISPR containing outgroups

# My Project

# Objectives

---

# Objectives

---

## Within Network Comparisons

For genera with CRISPR containing OTUs, compare the node statistics of CRISPR containing OTUs to non-CRISPR containing OTUs.



# Objectives

---

## Within Network Comparisons

For genera with CRISPR containing OTUs, compare the node statistics of CRISPR containing OTUs to non-CRISPR containing OTUs.

## Gene Indel Rates vs. Network Statistics

Compare gene Indel rates to node/network statistics for CRISPR containing and non-CRISPR containing OTUs

## Workflow (Per Genus)

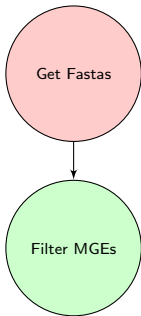
---



Get Fastas

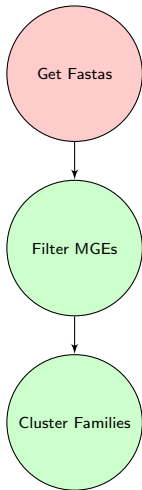
# Workflow (Per Genus)

---



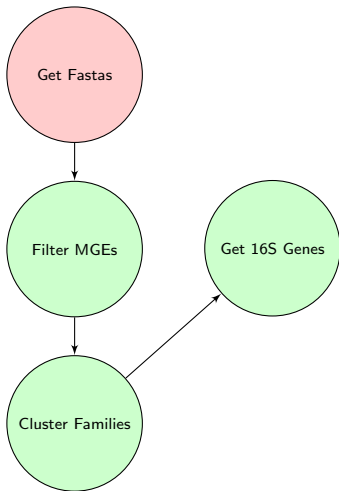
# Workflow (Per Genus)

---

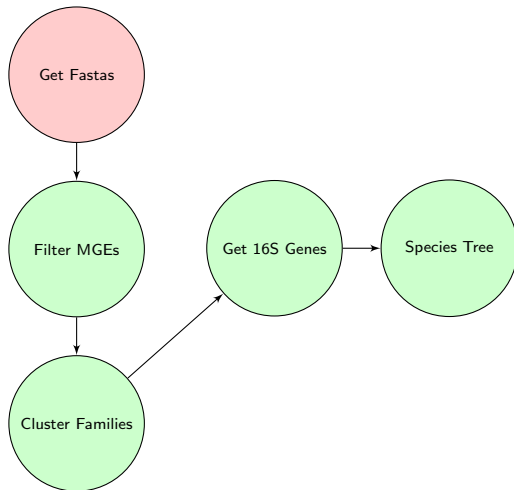


# Workflow (Per Genus)

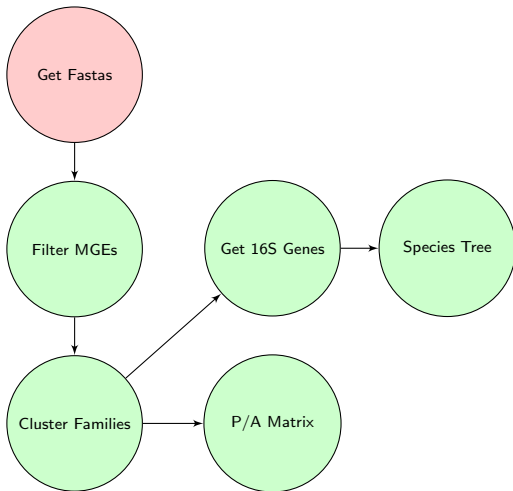
---



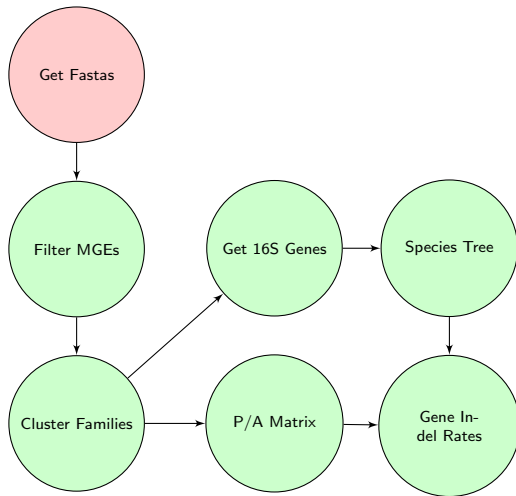
# Workflow (Per Genus)



# Workflow (Per Genus)

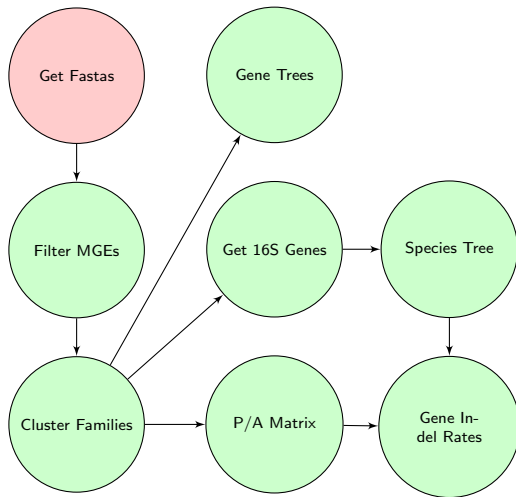


# Workflow (Per Genus)

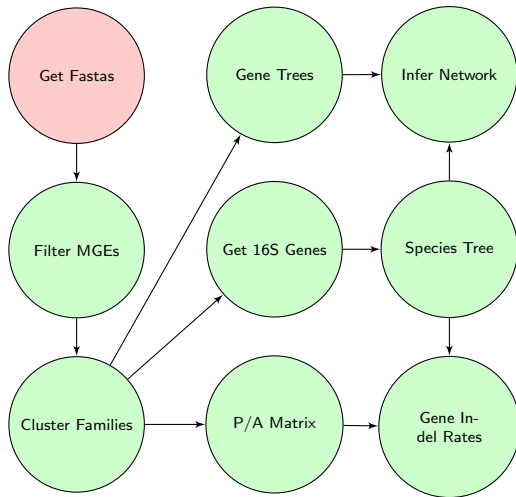




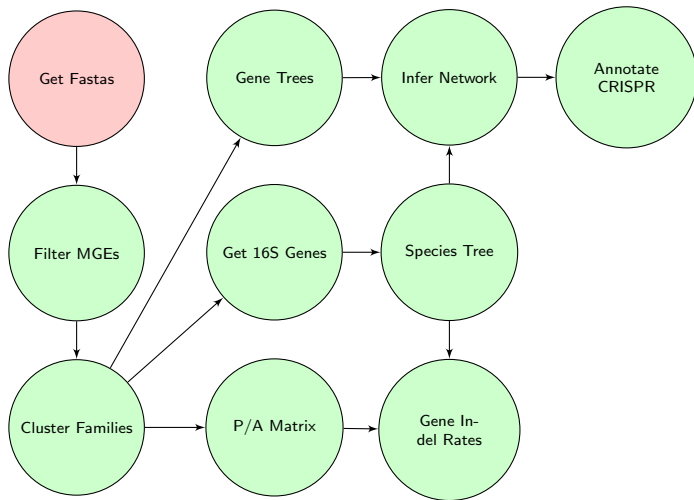
# Workflow (Per Genus)



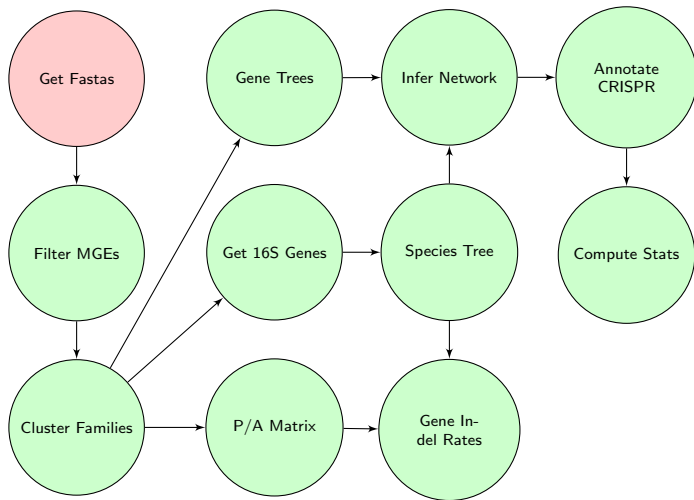
# Workflow (Per Genus)



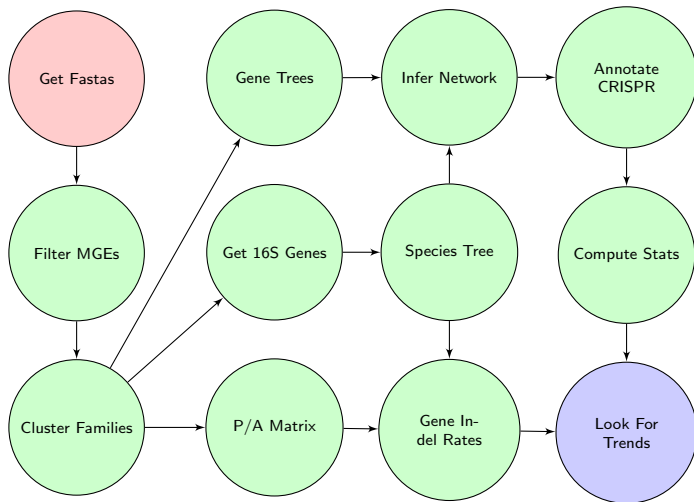
# Workflow (Per Genus)



# Workflow (Per Genus)

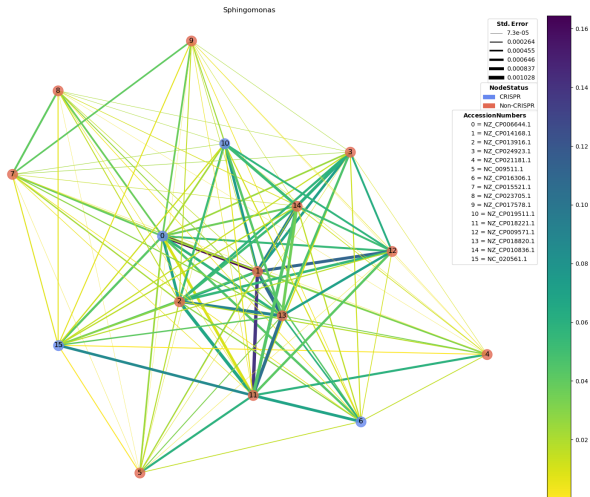


# Workflow (Per Genus)

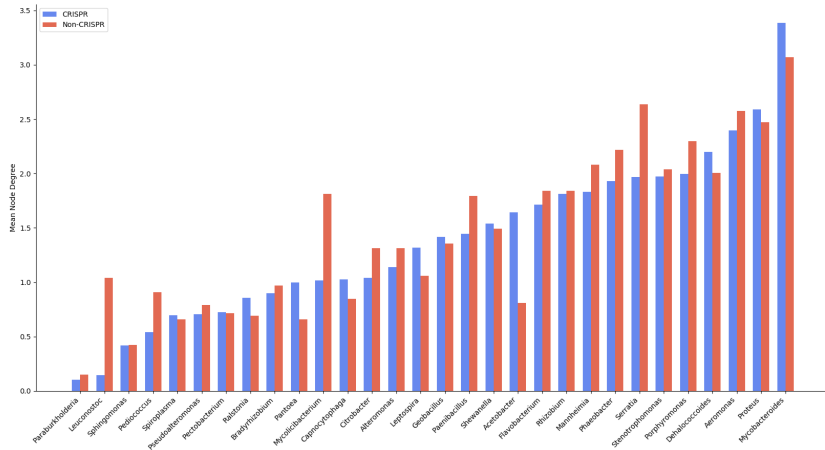


# Results

# Example “Consensus” Network

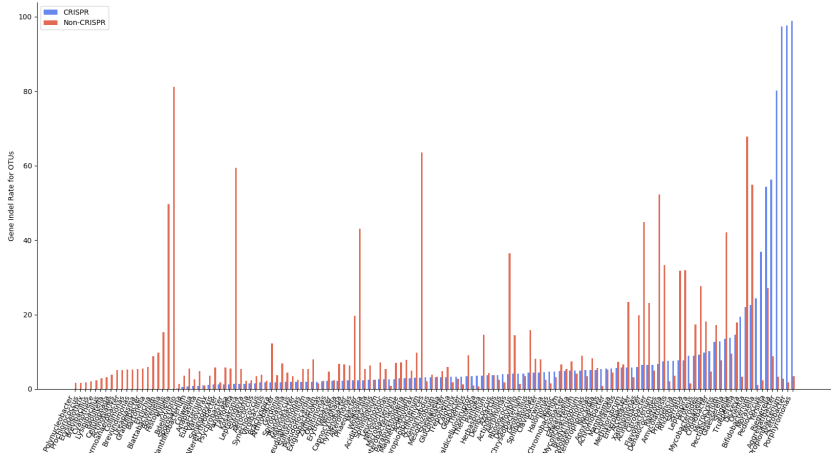


# Mean Node Degree

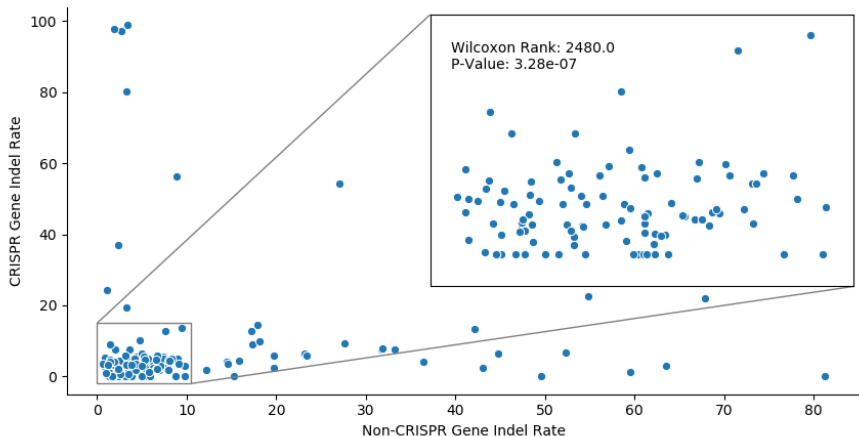




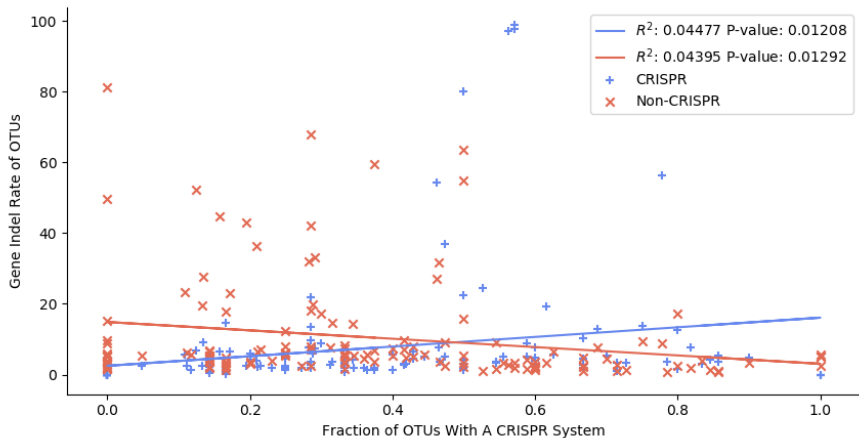
## Gene Indel Rates



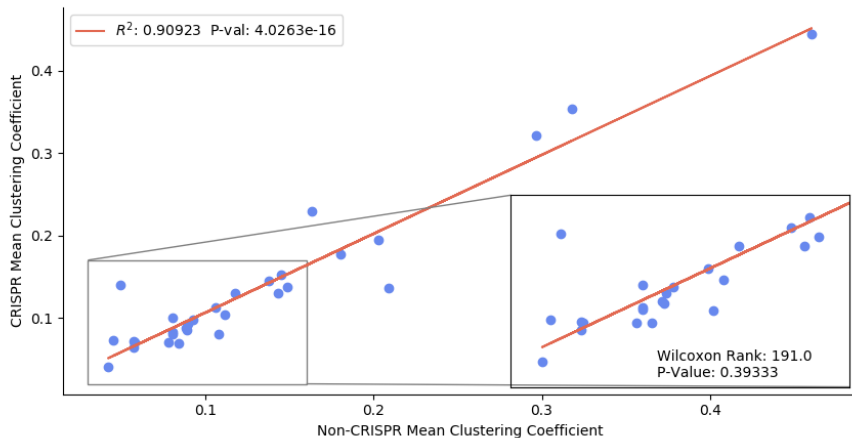
# Gene Indel Rates



# Gene Indel Rate Vs. Fraction of CRISPR OTUs



# Mean Node Weighted Clustering Coefficient



# Conclusion

# Findings

---

# Findings

---

- Large variation in HGT rate between genera.

# Findings

---

- Large variation in HGT rate between genera.
- CRISPR-Cas systems broadly associated with lower HGT rates, with prominent exceptions



## Findings

---

- Large variation in HGT rate between genera.
- CRISPR-Cas systems broadly associated with lower HGT rates, with prominent exceptions
- Population level effects of CRISPR-Cas systems may decrease HGT rates

# Findings

---

- Large variation in HGT rate between genera.
- CRISPR-Cas systems broadly associated with lower HGT rates, with prominent exceptions
- Population level effects of CRISPR-Cas systems may decrease HGT rates
- Interplay of CRISPR-Cas systems and HGT is complex and warrants further study

# Possible Future Directions

---

## Possible Future Directions

---

- **Inferring direction:** Directed networks have a host of available analytic tools Undirected networks do not

## Possible Future Directions

---

- **Inferring direction:** Directed networks have a host of available analytic tools Undirected networks do not
- **Continuous CRISPR activity:** Labeling nodes by estimated CRISPR activity (array length, transcriptomic data, etc.)

## Possible Future Directions

---

- **Inferring direction:** Directed networks have a host of available analytic tools Undirected networks do not
- **Continuous CRISPR activity:** Labeling nodes by estimated CRISPR activity (array length, transcriptomic data, etc.)
- **Gene function analysis:** Considering the transfer dynamics of different functional classes of genes

## Possible Future Directions

---

- **Inferring direction:** Directed networks have a host of available analytic tools Undirected networks do not
- **Continuous CRISPR activity:** Labeling nodes by estimated CRISPR activity (array length, transcriptomic data, etc.)
- **Gene function analysis:** Considering the transfer dynamics of different functional classes of genes
- **Studying movement of CRISPR systems:** Studying how frequently CRISPR systems themselves are transferred from arrays, *Cas* genes

## Possible Future Directions

---

- **Inferring direction:** Directed networks have a host of available analytic tools Undirected networks do not
- **Continuous CRISPR activity:** Labeling nodes by estimated CRISPR activity (array length, transcriptomic data, etc.)
- **Gene function analysis:** Considering the transfer dynamics of different functional classes of genes
- **Studying movement of CRISPR systems:** Studying how frequently CRISPR systems themselves are transferred from arrays, *Cas* genes
- **Intergenic comparisons:** Combine any set of fasta files from OTUs for analyzing transfer dynamics



## Possible Future Directions

---

- **Inferring direction:** Directed networks have a host of available analytic tools Undirected networks do not
- **Continuous CRISPR activity:** Labeling nodes by estimated CRISPR activity (array length, transcriptomic data, etc.)
- **Gene function analysis:** Considering the transfer dynamics of different functional classes of genes
- **Studying movement of CRISPR systems:** Studying how frequently CRISPR systems themselves are transferred from arrays, *Cas* genes
- **Intergenic comparisons:** Combine any set of fasta files from OTUs for analyzing transfer dynamics
- **Considering bacterial ecology and environments:** Consider geographically close OTUs or differences between networks due to environmental factors

# Is Sharing Caring?

# Is Sharing Caring?

Yes, for researchers

# Is Sharing Caring?

Yes, for researchers  
Jury's still out for bacteria

# Thanks

---

Thank you to

- Dr. G. Brian Golding
- Dr. Ben Evans
- The Golding lab
  - Caitlin Simopoulos
  - Daniella Lato
  - Zachery Dickson
  - Sam Long
  - Geoge Long
  - Lucy Zhang
  - Brianne Laverty
  - Nicole Zhang
- Everyone here for listening



All code used for this project is available at [https://github.com/DJSiddharthVader/thesis\\_SidReed](https://github.com/DJSiddharthVader/thesis_SidReed)

## References (1)

---



Grissa, I. and Drevet, C. and Couvin, D. (2017). *CRISPRdb*.  
<http://crispr.i2bc.paris-saclay.fr/>. Online; accessed 22  
October 2018.



Rath, Devashish et al. (2015). “The CRISPR-Cas immune system:  
Biology, mechanisms and applications”. In: *Biochimie* 117. Special  
Issue: Regulatory RNAs, pp. 119–128. ISSN: 0300-9084.



Popa, Ovidiu and Tal Dagan (2011). “Trends and barriers to lateral  
gene transfer in prokaryotes”. In: *Current Opinion in Microbiology*  
14.5. Antimicrobials/Genomics, pp. 615–623. ISSN: 1369-5274.



Bondy, J. A. and U. S. R. Murty (2002). *Graph theory with  
applications*. Wiley.



Stern, Adi et al. (2010). “Self-targeting by CRISPR: gene regulation  
or autoimmunity?” In: *Trends in Genetics* 26.8, pp. 335–340. ISSN:  
0168-9525.

## References (2)

---



Dzidic, Senka and Vladimir Bedeković (2003). “Horizontal gene transfer-emerging multidrug resistance in hospital bacteria”. In: *Acta pharmacologica Sinica* 24.6, pp. 519–526.



Bondy-Denomy, J. and A. R. Davidson (2014). “To Acquire Or Resist: The Complex Biological Effects Of CRISPR-Cas systems”. In: *Trends Microbio.* 22.4, pp. 218–25.



Watson, Bridget N. J., Raymond H. J. Staals, and Peter C. Fineran (2018). “CRISPR-Cas-Mediated Phage Resistance Enhances Horizontal Gene Transfer by Transduction”. In: *mBio* 9.1. Ed. by Joseph Bondy-Denomy and Michael S. Gilmore.



Godde, James S. and Amanda Bickerton (June 2006). “The Repetitive DNA Elements Called CRISPRs and Their Associated Genes: Evidence of Horizontal Transfer Among Prokaryotes”. In: *Journal of Molecular Evolution* 62.6, pp. 718–729. ISSN: 1432-1432.

## References (3)

---



Zhang, Quan and Yuzhen Ye (Feb. 2017). “Not all predicted CRISPR–Cas systems are equal: isolated cas genes and classes of CRISPR like elements”. In: *BMC Bioinformatics* 18.1, p. 92. ISSN: 1471-2105.



Makarova, K. S. et al. (2011). “Evolution and classification of the CRISPR–Cas systems”. In: *Nat. Rev. Microbiol.* 9.6, pp. 467–477.



Guimaraes, L. C. et al. (2015). “Inside the Pan-genome - Methods and Software Overview”. In: *Curr. Genomics* 16.4, pp. 245–252.



Rasko, David A. et al. (2008). “The Pangenome Structure of Escherichia coli: Comparative Genomic Analysis of E. coli Commensal and Pathogenic Isolates”. In: *Journal of Bacteriology* 190.20, pp. 6881–6893. ISSN: 0021-9193.



## References (4)

---



Berglund, Björn (2015). “Environmental dissemination of antibiotic resistance genes and correlation to anthropogenic contamination with antibiotics”. In: *Infection Ecology & Epidemiology* 5.1, p. 28564.



Kunin, V. et al. (2005). “The net of life: reconstructing the microbial phylogenetic network”. In: *Genome Res.* 15.7, pp. 954–959.



Ravenhall, Matt et al. (May 2015). “Inferring Horizontal Gene Transfer”. In: *PLoS Computational Biology* 11.5, pp. 1–16.



Onnela, J. P. et al. (2005). “Intensity and coherence of motifs in weighted complex networks”. In: *Phys Rev E Stat Nonlin Soft Matter Phys* 71.6 Pt 2, p. 065103.



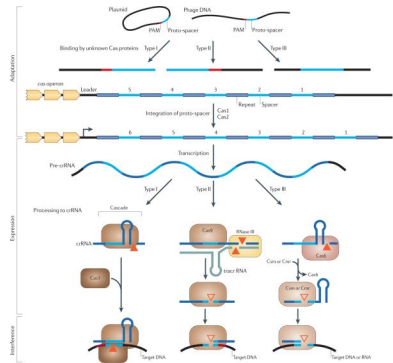
Newman, M. E. (2002). “Assortative mixing in networks”. In: *Phys. Rev. Lett.* 89.20, p. 208701.



– (2004). “Analysis of weighted networks”. In: *Phys Rev E Stat Nonlin Soft Matter Phys* 70.5 Pt 2, p. 056131.

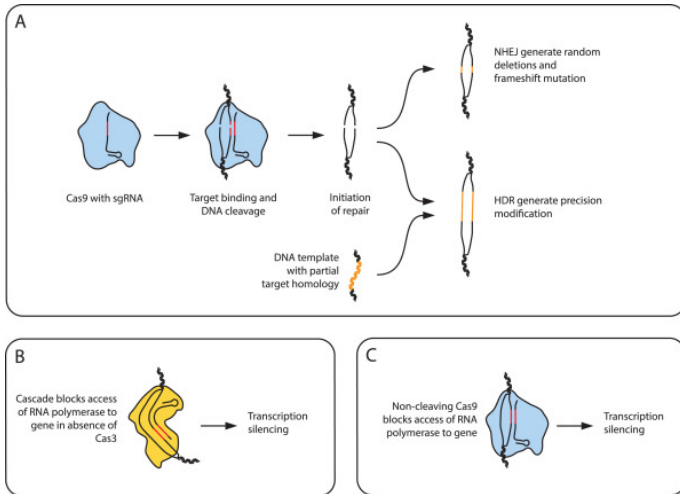
# Diversity & Ubiquity

- 45% of bacteria have CRISPR loci ( $n = 6782$ ) (Grissa, I. and Drevet, C. and Couvin, D., 2017)
- 3 Main Types, multiple sub types (Bondy-Denomy and Davidson, 2014)
- CRISPR arrays represent unique life history of an organism
- 11% – 28% are false or orphaned CRISPR loci (Zhang and Ye, 2017)



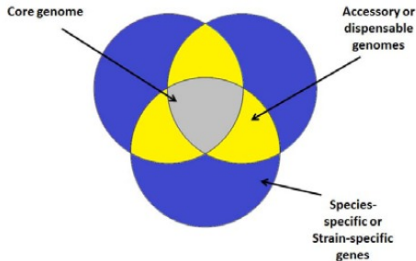
(Makarova et al., 2011)

# CRISPR Biotech Application

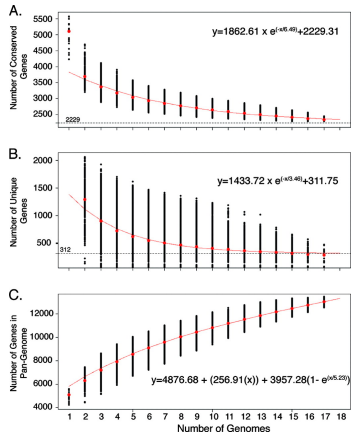


(Rath et al., 2015)

# Pan-Genomes

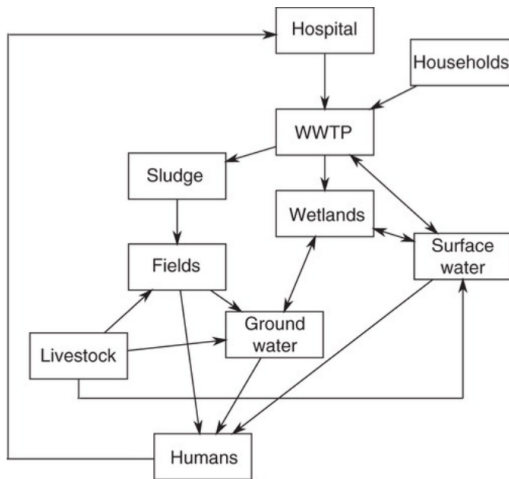


(Guimaraes et al., 2015)



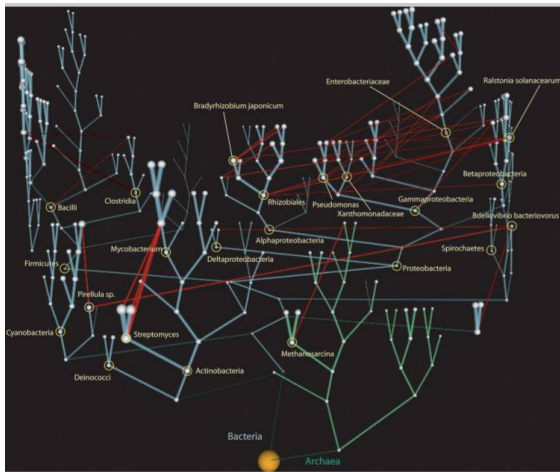
(Rasko et al., 2008)

# HGT Applications



(Berglund, 2015)

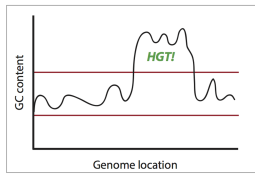
# Prokaryotic “Net of Life”



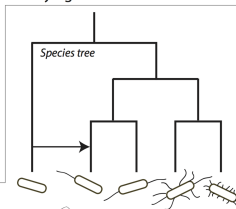
(Kunin et al., 2005)

# Phylogenomic Network Construction

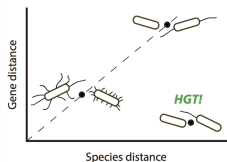
## 1. Parametric methods



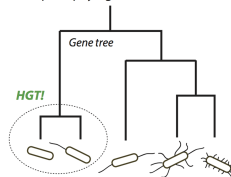
## 2. Phylogenetic methods



## 2a. Implicit phylogenetic methods

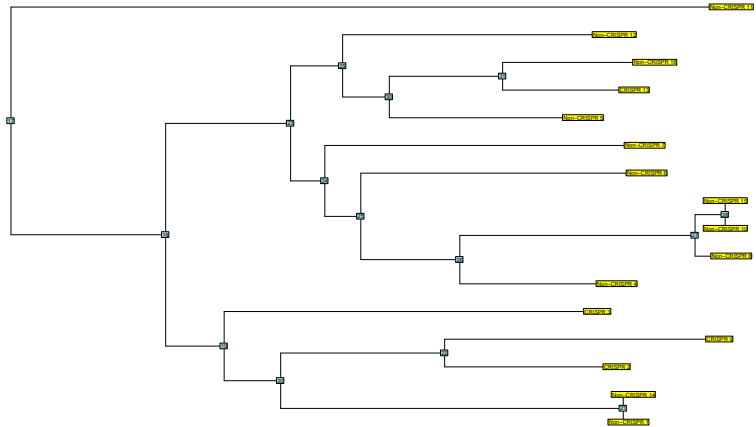


## 2b. Explicit phylogenetic methods



(Ravenhall et al., 2015)

# Sphingomonas Species Tree

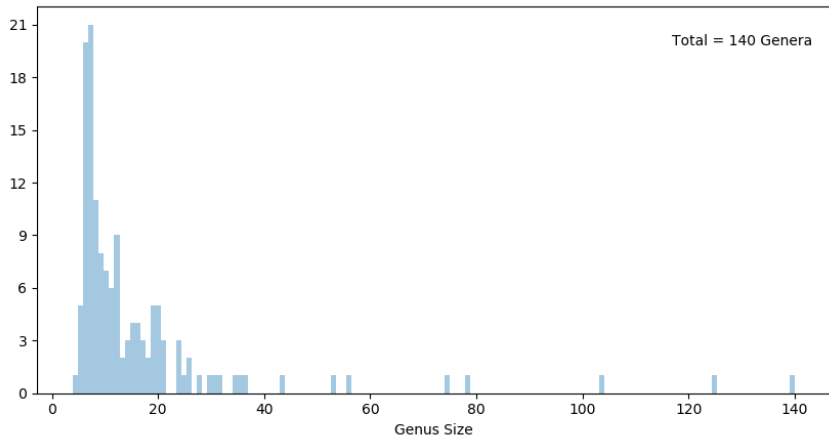




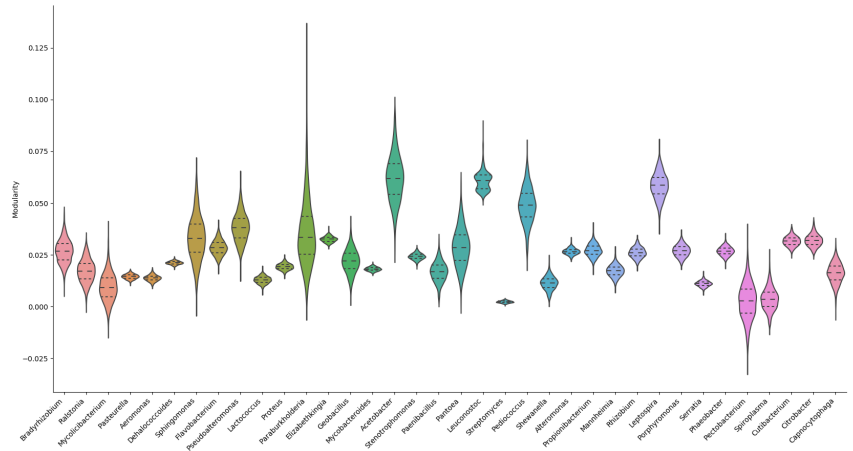
## Network Statistics

- **Average Node Degree:**  $\frac{1}{|N_u|} \sum_{uv} w_{uv}$  where  $N_u$  is the set of nodes incident to  $u$
- **Node Clustering Coefficient:**  $\frac{1}{k_u(k_u-1)} \sum_{vw}^{T(u)} (\hat{w}_{uw} \hat{w}_{vw} \hat{w}_{uv})^{\frac{1}{3}}$  where  $T(u)$  is the set of triangles containing  $u$  (Onnela et al., 2005)
- **Node Assortativity:**  $A = \frac{Tr(M) - ||M^2||}{1 - ||M^2||}$  Where  $M$  is the mixing matrix of a given attribute and  $||M||$  is the sum of all elements of  $M$ .  $A \in [-1, 1]$ . (Newman, 2002)
- **Network Modularity:**  $Q = \frac{1}{2m} \sum_{uv} [W_{uv} - \frac{k_u k_v}{2m}] \delta(u, v)$  where  $m$  is the total weight of all edges,  $k_u$  is the degree of  $u$  and  $\delta(u, v)$  is 1 if  $u$  and  $v$  both have or do not have CRISPR systems and 0 otherwise.  $Q \in [-1, 1]$  (Newman, 2004)

# Genus Size Distribution



# Modularity Distributions



# Assortativity Distributions

