Is Sharing Caring?
How CRISPR-Cas systems affect rates of Horizontal Gene Transfer
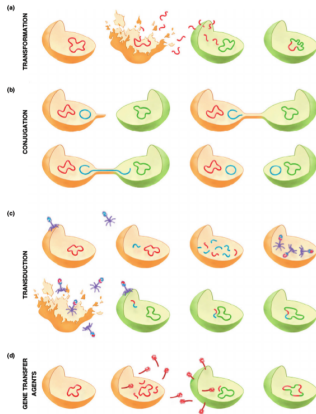
Siddharth Reed

Golding Lab, McMaster University

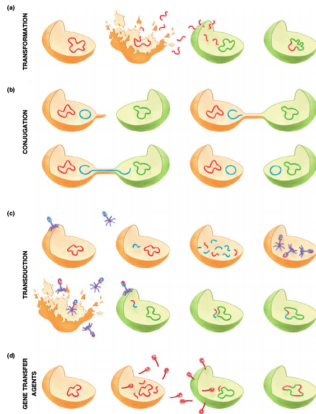February 18, 2022

Background
●●●●●●●

My Project
○○○○○○

Results
○○○○○○○○○○○○

Moral of the Study
○○○○○○

Background

What is Horizontal Gene Transfer?

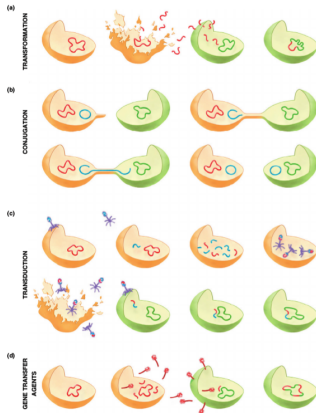

Figure 1: [1]

▶ Transformation

What is Horizontal Gene Transfer?

Figure 1: [1]

- ▶ Transformation
  - ◇ Incorporation of free-floating DNA into the genome

What is Horizontal Gene Transfer?

Figure 1: [1]

- ▶ Transformation
  - ◇ Incorporation of free-floating DNA into the genome
- ▶ Conjugation

What is Horizontal Gene Transfer?

Figure 1: [1]

- Transformation
  - ◇ Incorporation of free-floating DNA into the genome
- Conjugation
  - ◇ Transfer of DNA through cell-cell connections
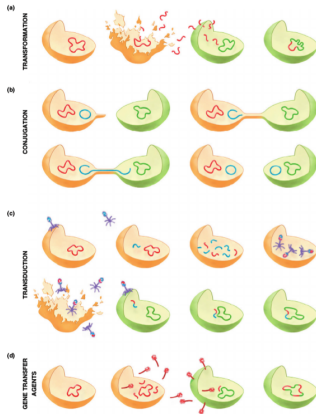
What is Horizontal Gene Transfer?

Figure 1: [1]

- ▶ Transformation
  - ◇ Incorporation of free-floating DNA into the genome
- ▶ Conjugation
  - ◇ Transfer of DNA through cell-cell connections
- ▶ Transduction
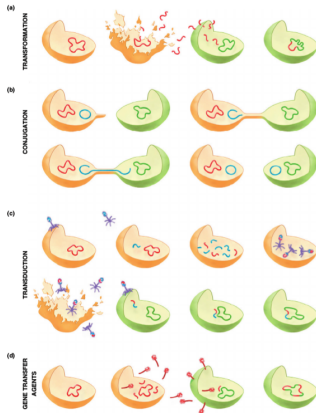
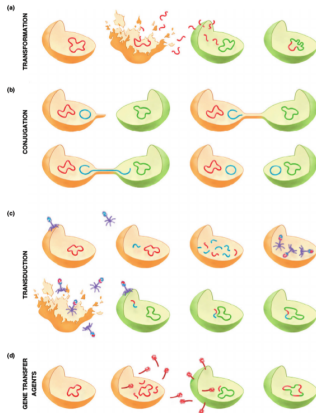# What is Horizontal Gene Transfer?

Figure 1: [1]

- ▶ Transformation
  - ◇ Incorporation of free-floating DNA into the genome
- ▶ Conjugation
  - ◇ Transfer of DNA through cell-cell connections
- ▶ Transduction
  - ◇ Transfer of DNA through phage

What is Horizontal Gene Transfer?

Figure 1: [1]

▶ Transformation
  ◇ Incorporation of free-floating DNA into the genome
▶ Conjugation
  ◇ Transfer of DNA through cell-cell connections
▶ Transduction
  ◇ Transfer of DNA through phage
▶ Rates of HGT can be impacted by

What is Horizontal Gene Transfer?

Figure 1: [1]

- Transformation
  - ◇ Incorporation of free-floating DNA into the genome
- Conjugation
  - ◇ Transfer of DNA through cell-cell connections
- Transduction
  - ◇ Transfer of DNA through phage
- Rates of HGT can be impacted by
  - ◇ Amount of exogenous DNA/cell density/phage density

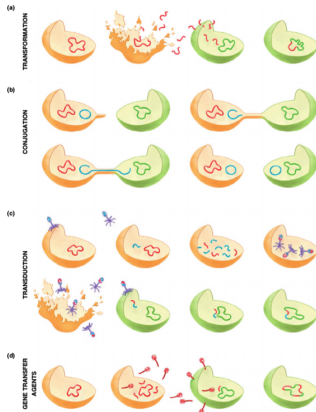# What is Horizontal Gene Transfer?

Figure 1: [1]

- ▶ Transformation
  - ◇ Incorporation of free-floating DNA into the genome
- ▶ Conjugation
  - ◇ Transfer of DNA through cell-cell connections
- ▶ Transduction
  - ◇ Transfer of DNA through phage
- ▶ Rates of HGT can be impacted by
  - ◇ Amount of exogenous DNA/cell density/phage density
  - ◇ Selective pressures

What is Horizontal Gene Transfer?

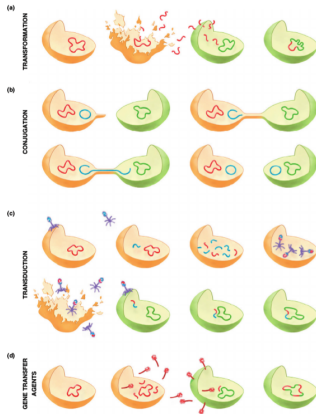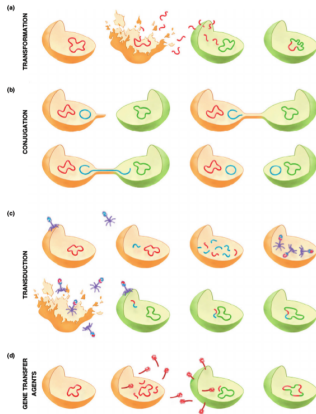Figure 1: [1]

- Transformation
  - ◇ Incorporation of free-floating DNA into the genome
- Conjugation
  - ◇ Transfer of DNA through cell-cell connections
- Transduction
  - ◇ Transfer of DNA through phage
- Rates of HGT can be impacted by
  - ◇ Amount of exogenous DNA/cell density/phage density
  - ◇ Selective pressures
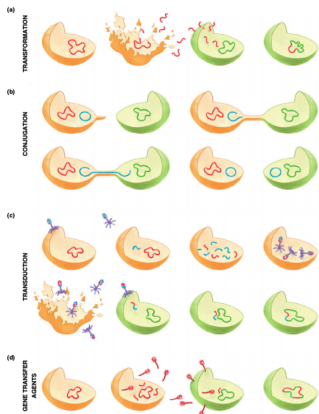 - ◇ Metabolic costs

What is Horizontal Gene Transfer?
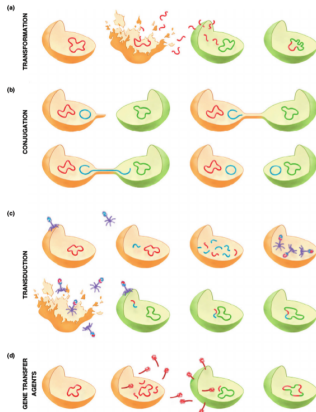
Figure 1: [1]

▶ Transformation
  ◇ Incorporation of free-floating DNA into the genome
▶ Conjugation
  ◇ Transfer of DNA through cell-cell connections
▶ Transduction
  ◇ Transfer of DNA through phage
▶ Rates of HGT can be impacted by
  ◇ Amount of exogenous DNA/cell density/phage density
  ◇ Selective pressures
  ◇ Metabolic costs
  ◇ Sequence compatibility

What is A Network?

Figure 2: [2]

► Useful mathematical abstraction of real world system
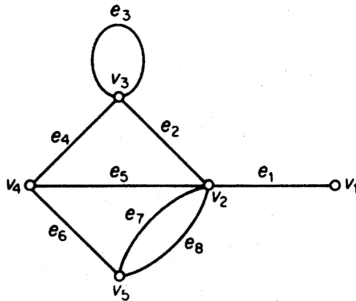
What is A Network?

Figure 2: [2]

- ▶ Useful mathematical abstraction of real world system
- ▶ Nodes represent objects

What is A Network?
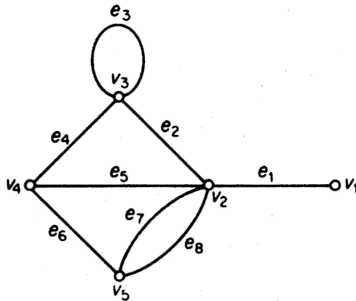
Figure 2: [2]

- ▶ Useful mathematical abstraction of real world system
- ▶ Nodes represent objects
- ▶ Edges represent relationships

What is A Network?
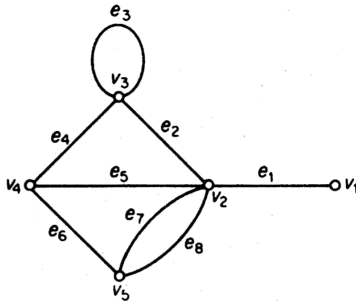
Figure 2: [2]

- ▶ Useful mathematical abstraction of real world system
- ▶ Nodes represent objects
- ▶ Edges represent relationships
- ▶ Nodes and edges can have attributes
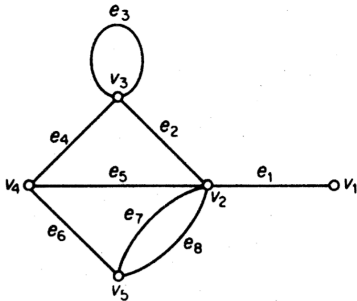
What is A Network?
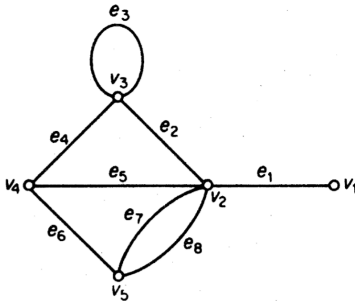
Figure 2: [2]

- ▶ Useful mathematical abstraction of real world system
- ▶ Nodes represent objects
- ▶ Edges represent relationships
- ▶ Nodes and edges can have attributes
- ▶ **Nodes are OTUs, edges are inferred HGT rates**

What are CRISPR-Cas systems?

Figure 3: [3]

▶ Adaptive immune system in bacteria

What are CRISPR-Cas systems?

Figure 3: [3]

- ▶ Adaptive immune system in bacteria
- ▶ Failed "infection" → spacer acquisition → targeted degradation for next "infection"

What are CRISPR-Cas systems?

Figure 3: [3]

- Adaptive immune system in bacteria
- Failed "infection" $\rightarrow$ spacer acquisition $\rightarrow$ targeted degradation for next "infection"
- Protects against foreign DNA absorption/integration

What are CRISPR-Cas systems?

- Adaptive immune system in bacteria
- Failed "infection" → spacer acquisition → targeted degradation for next "infection"
- Protects against foreign DNA absorption/integration
- Requires CRISPR loci + Cas proteins
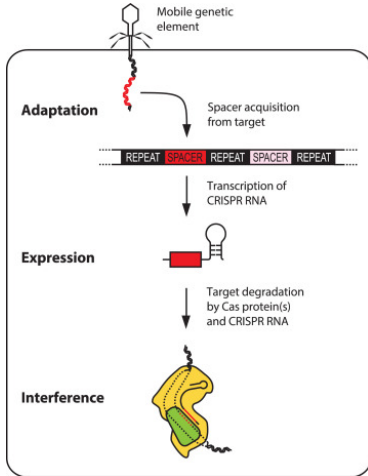
Figure 3: [3]

What are CRISPR-Cas systems?
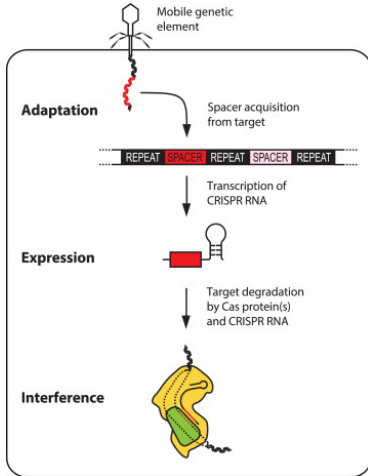
Figure 3: [3]

- ▶ Adaptive immune system in bacteria
- ▶ Failed "infection" $\rightarrow$ spacer acquisition $\rightarrow$ targeted degradation for next "infection"
- ▶ Protects against foreign DNA absorption/integration
- ▶ Requires CRISPR loci + Cas proteins
- ▶ 45% of bacteria have CRISPR loci ($n = 6782$) [4]

Background
○○○○○●○○

My Project
○○○○○○

Results
○○○○○○○○○○○○

Moral of the Study
○○○○○○

Do CRISPR Systems Affect Horizontal Gene Transfer?

Yes

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

▶ 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

- ▶ 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

- ▶ 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 [5]

- ▶ 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 [5]
  - ◇ Can see inhibitory effects of CRISPR on HGT over short evolutionary time scales

- ▶ 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 [5]
  - ◇ 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

CRISPR Cost Complexity and Curbing It

Fitness Cost Factors

- Metabolic maintenance [3]

CRISPR Cost Complexity and Curbing It

Fitness Cost Factors

- ▶ Metabolic maintenance [3]
- ▶ Off-target effects [6]

CRISPR Cost Complexity and Curbing It

Fitness Cost Factors

- Metabolic maintenance [3]
- Off-target effects [6]
- Environmental pressures [7]

CRISPR Cost Complexity and Curbing It

Fitness Cost Factors

- ▶ Metabolic maintenance [3]
- ▶ Off-target effects [6]
- ▶ Environmental pressures [7]
- ▶ Phage density [8]

CRISPR Cost Complexity and Curbing It

Fitness Cost Factors

- Metabolic maintenance [3]
- Off-target effects [6]
- Environmental pressures [7]
- Phage density [8]
- Anti-CRISPR systems [8]

CRISPR Cost Complexity and Curbing It

Fitness Cost Factors

- Metabolic maintenance [3]
- Off-target effects [6]
- Environmental pressures [7]
- Phage density [8]
- Anti-CRISPR systems [8]
- Prophage abundance [9]

Cost Reduction Strategies

CRISPR Cost Complexity and Curbing It

Fitness Cost Factors

- Metabolic maintenance [3]
- Off-target effects [6]
- Environmental pressures [7]
- Phage density [8]
- Anti-CRISPR systems [8]
- Prophage abundance [9]

Cost Reduction Strategies

- Selective CRISPR inactivation [3]

CRISPR Cost Complexity and Curbing It

Fitness Cost Factors

- Metabolic maintenance [3]
- Off-target effects [6]
- Environmental pressures [7]
- Phage density [8]
- Anti-CRISPR systems [8]
- Prophage abundance [9]

Cost Reduction Strategies

- Selective CRISPR inactivation [3]
- CRISPRs get transferred $\implies$ population level immunity [10]

CRISPR Cost Complexity and Curbing It

Fitness Cost Factors

- Metabolic maintenance [3]
- Off-target effects [6]
- Environmental pressures [7]
- Phage density [8]
- Anti-CRISPR systems [8]
- Prophage abundance [9]

Cost Reduction Strategies

- Selective CRISPR inactivation [3]
- CRISPRs get transferred $\implies$ population level immunity [10]
- CRISPR can enhance transduction-mediated HGT [9]

My Project

Background
0000000

My Project
0●0000

Results
00000000000

Moral of the Study
000000

Goals

Within Network Comparisons

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

Background
0000000

My Project
0●0000

Results
00000000000

Moral of the Study
000000

Goals

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

Background
0000000

My Project
000●000

Results
00000000000

Moral of the Study
000000

InDel Rate Estimates
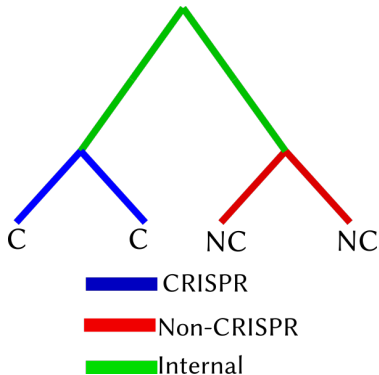
Markophylo [11]

- ▶ Input: a presence/absence matrix of gene families + species tree

InDel Rate Estimates
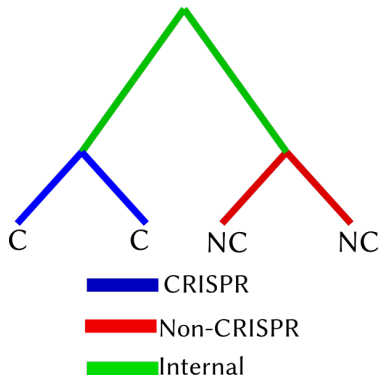
Markophylo [11]

- ▶ Input: a presence/absence matrix of gene families + species tree
- ▶ Branches are partitioned by the OTU having a CRISPR system

InDel Rate Estimates

C   C   NC   NC

■ CRISPR
■ Non-CRISPR
■ Internal

Markophylo [11]

- ▶ Input: a presence/absence matrix of gene families + species tree
- ▶ Branches are partitioned by the OTU having a CRISPR system
- ▶ Internal branches ignored

InDel Rate Estimates
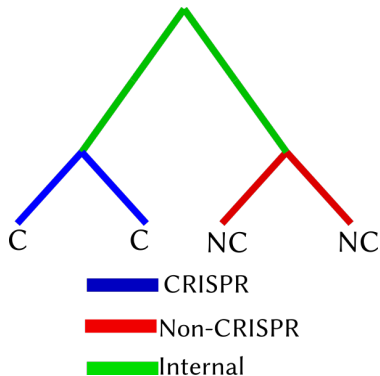
C   C   NC   NC

■ CRISPR

■ Non-CRISPR

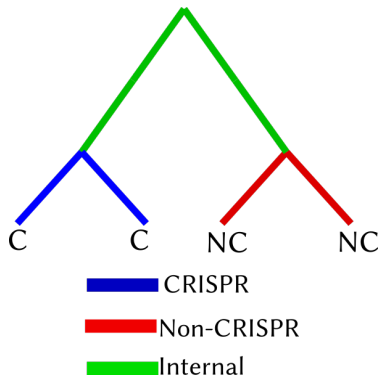■ Internal

Markophylo [11]

- ▶ Input: a presence/absence matrix of gene families + species tree
- ▶ Branches are partitioned by the OTU having a CRISPR system
- ▶ Internal branches ignored
- ▶ Gene birth-death rates are estimated for each branch partition

Background
○○○○○○○

My Project
○○○●○○

Results
○○○○○○○○○○○○

Moral of the Study
○○○○○○

Network Sampling
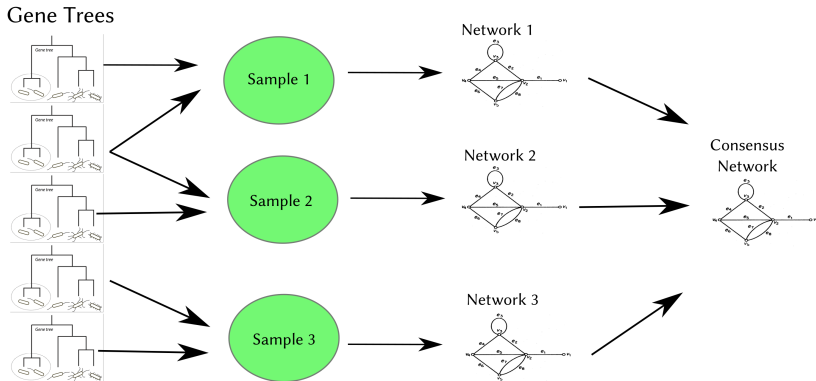
Network Statistics

- **Mean Node Degree:** $\frac{1}{|N_u|} \sum_{uv}^{N_u} w_{uv}$ where N_u is the set of nodes incident to u

Network Statistics

- ▶ **Mean Node Degree:** $\frac{1}{|N_u|} \sum_{uv}^{N_u} w_{uv}$ 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 [12]

Network Statistics

- ▶ **Mean Node Degree:** $\frac{1}{|N_u|} \sum_{uv}^{N_u} w_{uv}$ 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 [12]
- ▶ **Network 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. $A \in [-1, 1]$. [13]

Network Statistics

- ▶ **Mean Node Degree:** $\frac{1}{|N_u|} \sum_{uv}^{N_u} w_{uv}$ 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 [12]
- ▶ **Network 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. $A \in [-1, 1]$. [13]
- ▶ **Network Modularity:** $Q = \frac{1}{2m} \sum_{uv}^{W} [W_{uv} - \frac{k_u k_v}{2m}] \delta(u, v)$ where m is the total edge weight, 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. $Q \in [-1, 1]$ [14]

Background
0000000

My Project
000000●

Results
000000000000

Moral of the Study
000000

Workflow (per genus)

Background
0000000

My Project
000000●

Results
000000000000

Moral of the Study
000000

Workflow (per genus)

Get Fastas

Background
0000000

My Project
000000●

Results
000000000000

Moral of the Study
000000

Workflow (per genus)

Get Fastas

Filter
MGEs

Background
0000000

My Project
000000●

Results
000000000000

Moral of the Study
000000

Workflow (per genus)

Background
0000000

My Project
000000●

Results
000000000000

Moral of the Study
000000

Workflow (per genus)

Background
○○○○○○○

My Project
○○○○○●

Results
○○○○○○○○○○○○

Moral of the Study
○○○○○○

Workflow (per genus)

Workflow (per genus)

Background
ooooooo

My Project
ooooo●

Results
oooooooooooo

Moral of the Study
oooooo

Workflow (per genus)

Background
0000000

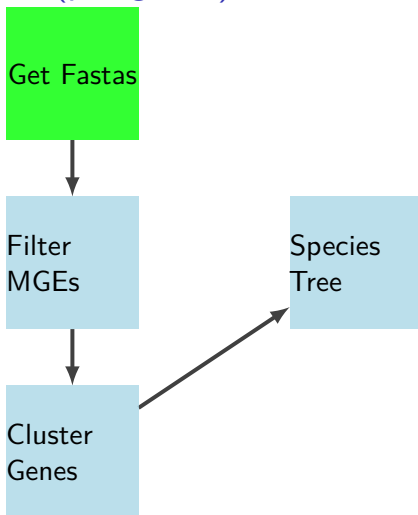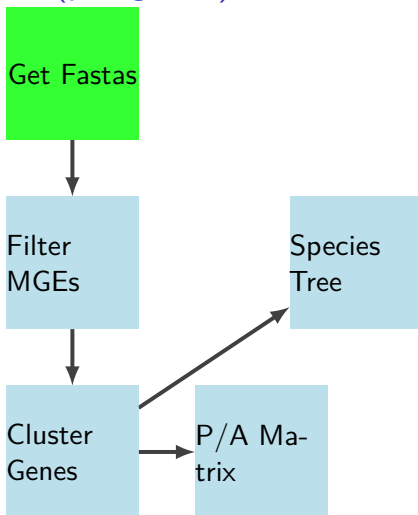My Project
000000●

Results
000000000000

Moral of the Study
000000

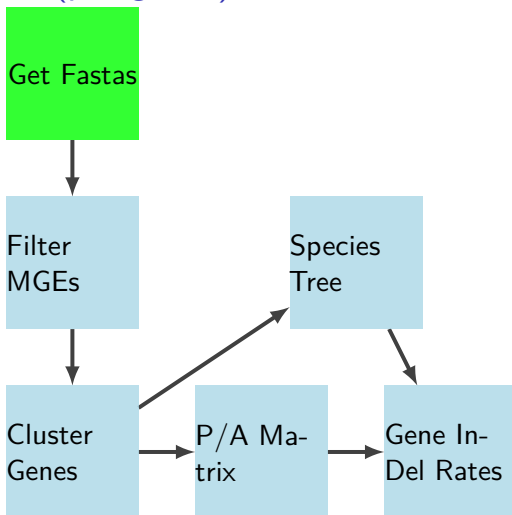# Workflow (per genus)

Workflow (per genus)

Workflow (per genus)
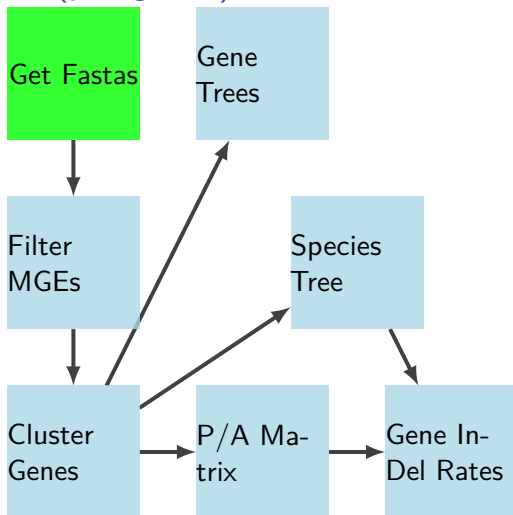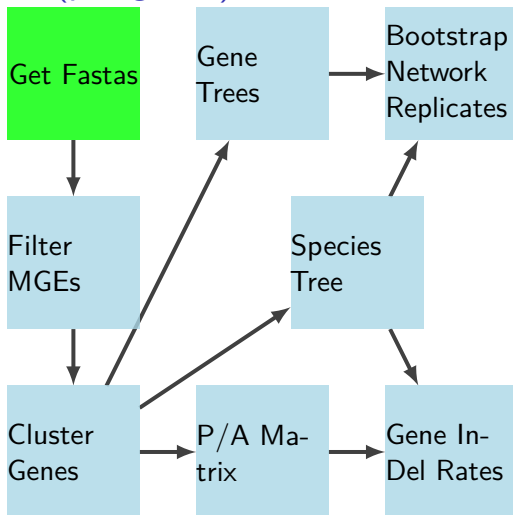
Background
○○○○○○○

My Project
○○○○○●

Results
○○○○○○○○○○○○○

Moral of the Study
○○○○○○

Workflow (per genus)

Background
0000000

My Project
000000

Results
●00000000000

Moral of the Study
000000

Results

Example "Consensus" Network

Background
0000000

My Project
000000

Results
000●000000000

Moral of the Study
000000
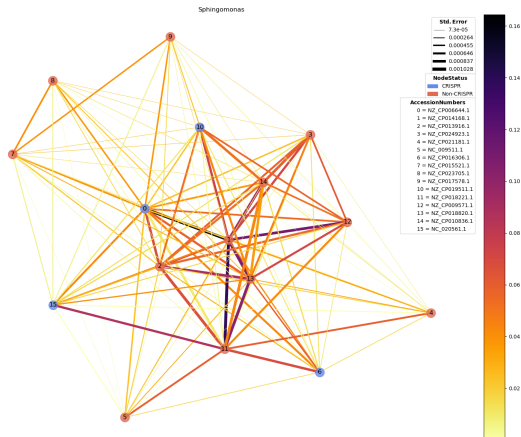
Genus Size Distribution
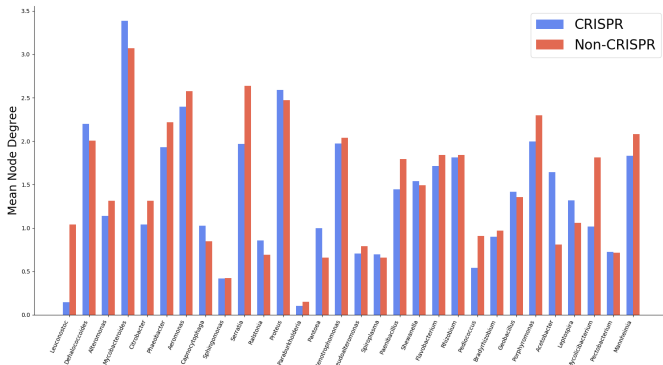
Total = 115 Genera

Genus Size

Mean Node Degree

Gene Indel Rates

Background
○○○○○○○

My Project
○○○○○○

Results
○○○○○●○○○○○○

Moral of the Study
○○○○○○

Gene Indel Rate Vs. Fraction of CRISPR OTUs

Background
0000000

My Project
000000

Results
0000000●0000
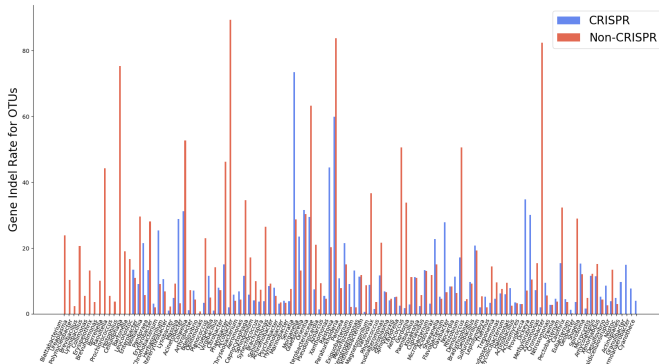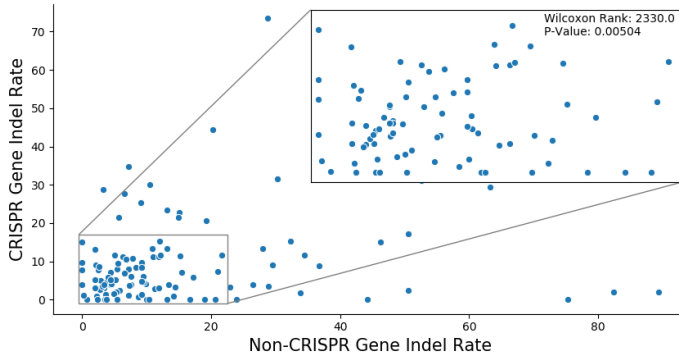
Moral of the Study
000000
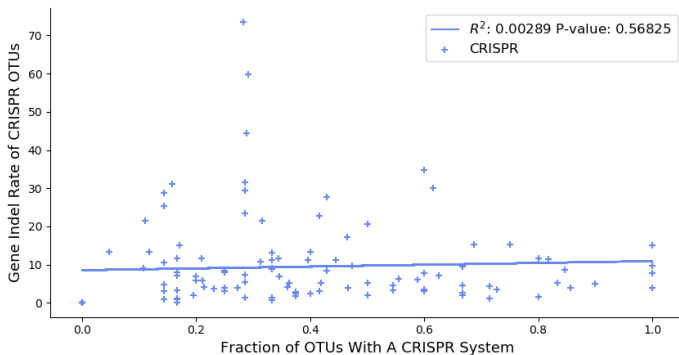
Gene Indel Rate Vs. Fraction of CRISPR OTUs
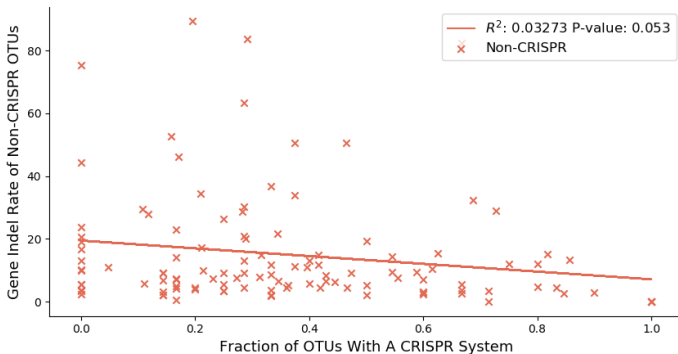
Mean Node Weighted Clustering Coefficient
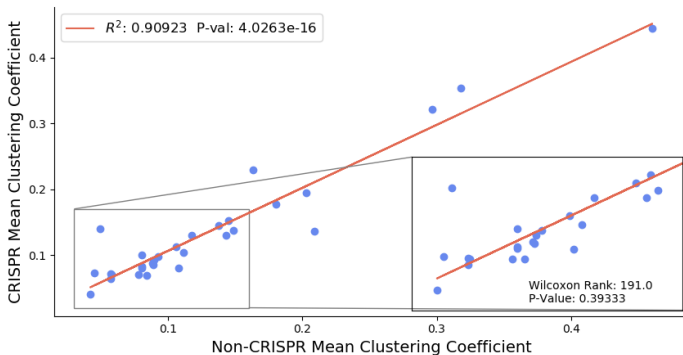
Assortativity Distributions
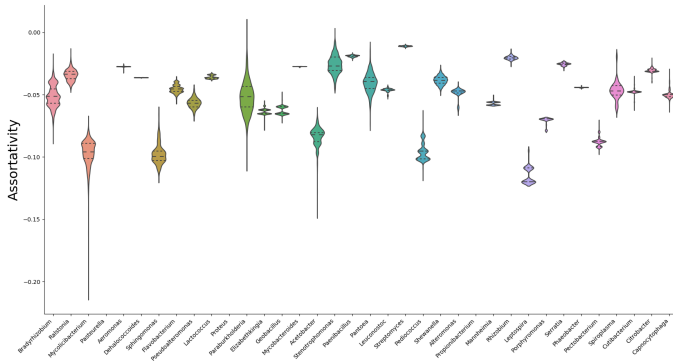
Modularity Distributions

Indel Rate Pair Plot

Background
0000000

My Project
000000

Results
00000000000
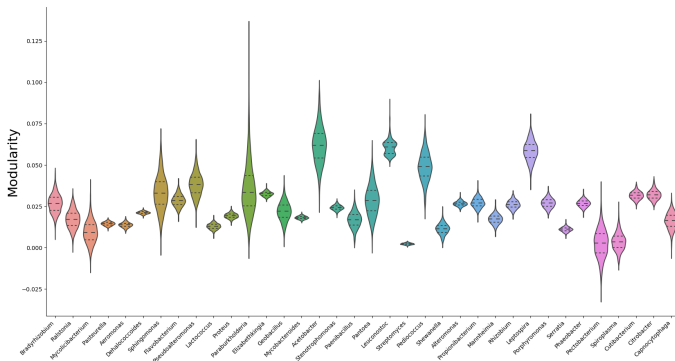
Moral of the Study
●00000
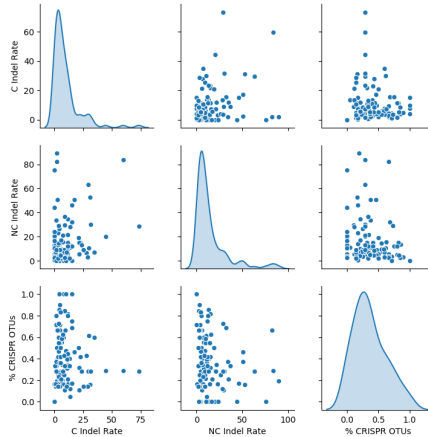
Moral of the Study

Findings

- ▶ Large variation in HGT rate between genera.

Findings

- Large variation in HGT rate between genera.
- CRISPR systems $\sim$ associated with lower HGT rates

Findings

- ▶ Large variation in HGT rate between genera.
- ▶ CRISPR systems $\sim$ associated with lower HGT rates
  - ◇ Prominent exceptions exist

Findings

- ► Large variation in HGT rate between genera.
- ► CRISPR systems $\sim$ associated with lower HGT rates
  - ◇ Prominent exceptions exist
  - ◇ High mixing between CRISPR and non-CRISPR OTUs

Findings

- Large variation in HGT rate between genera.
- CRISPR systems $\sim$ associated with lower HGT rates
  - ◊ Prominent exceptions exist
  - ◊ High mixing between CRISPR and non-CRISPR OTUs
- Population level effects of CRISPR systems may decrease HGT rates

Findings

- Large variation in HGT rate between genera.
- CRISPR systems $\sim$ associated with lower HGT rates
  - ◊ Prominent exceptions exist
  - ◊ High mixing between CRISPR and non-CRISPR OTUs
- Population level effects of CRISPR systems may decrease HGT rates
- Interplay of CRISPR systems and HGT is complex and warrants further study

Background
0000000

My Project
000000

Results
000000000000

Moral of the Study
000●000

Possible Future Directions

- **Intergenic comparisons:** What if we analyze networks with multiple genera e.g. ones that share a microbiome?

Background
oooooooo

My Project
oooooo

Results
ooooooooooooo

Moral of the Study
oooeooo

Possible Future Directions

- **Intergenic comparisons:** What if we analyze networks with multiple genera e.g. ones that share a microbiome?
- **Inferring direction:** Inferring direction of transfer $\rightarrow$ more analytic tools available

Possible Future Directions

- **Intergenic comparisons:** What if we analyze networks with multiple genera e.g. ones that share a microbiome?
- **Inferring direction:** Inferring direction of transfer $\rightarrow$ more analytic tools available
- **CRISPR Label:** binary (presence) $\rightarrow$ continuous (activity) e.g. array length, transciptomic data, etc.

Background
0000000

My Project
000000

Results
00000000000

Moral of the Study
000●000

Possible Future Directions

- **Intergenic comparisons:** What if we analyze networks with multiple genera e.g. ones that share a microbiome?
- **Inferring direction:** Inferring direction of transfer $\rightarrow$ more analytic tools available
- **CRISPR Label:** binary (presence) $\rightarrow$ continuous (activity) e.g. array length, transciptomic data, etc.
- **Gene function analysis:** How are HGT dynamics different for different functional groups?

Possible Future Directions

- ▶ **Intergenic comparisons:** What if we analyze networks with multiple genera e.g. ones that share a microbiome?
- ▶ **Inferring direction:** Inferring direction of transfer $\rightarrow$ more analytic tools available
- ▶ **CRISPR Label:** binary (presence) $\rightarrow$ continuous (activity) e.g. array length, transciptomic data, etc.
- ▶ **Gene function analysis:** How are HGT dynamics different for different functional groups?
- ▶ **Transfer of CRISPR systems:** How do CRISPR systems get transfered around?

Background
ooooooo

My Project
oooooo

Results
ooooooooooooo

Moral of the Study
oooo●oo

Conclusion

Is Sharing Caring?

Background
○○○○○○○

My Project
○○○○○○

Results
○○○○○○○○○○○○

Moral of the Study
○○○●○○

Conclusion

Is Sharing Caring?

▶ Yes, for researchers

Background
0000000

My Project
000000

Results
000000000000

Moral of the Study
000●00

Conclusion

Is Sharing Caring?

- ▶ Yes, for researchers
- ▶ Jury's still out for bacteria

Background
○○○○○○○

My Project
○○○○○○

Results
○○○○○○○○○○○○○

Moral of the Study
○○○○●○

Thanks

Thank you to

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

Code Availability

All code written written for this project is available at https://github.com/DJSid dharthVader/Undergrad_Thesis

Bibliography I

1. Popa O, Dagan T. Trends and barriers to lateral gene transfer in prokaryotes. *Current Opinion in Microbiology* 2011; **14**:615–623.

2. Bondy JA, Murty USR. *Graph theory with applications*. Wiley; 2002.

3. Rath D, Amlinger L, Rath A, Lundgren M. The CRISPR-cas immune system: Biology, mechanisms and applications. *Biochimie* 2015; **117**:119–128.

4. Grissa, I. and Drevet, C. and Couvin, D. CRISPRdb. 2017.

5. Zambelis A, Dang UJ, Golding GB. Effects of CRISPR-cas system presence on lateral gene transfer rates in bacteria. 2015.

Bibliography II

6. Stern A, Keren L, Wurtzel O, Amitai G, Sorek R. Self-targeting by CRISPR: Gene regulation or autoimmunity? *Trends in Genetics* 2010; **26**:335–340.

7. Dzidic S, Bedeković V. Horizontal gene transfer-emerging multidrug resistance in hospital bacteria. *Acta pharmacologica Sinica* 2003; **24**:519—526.

8. Bondy-Denomy J, Davidson AR. To acquire or resist:the complex biological effects of CRISPR-cas systems. *Trends Microbio* 2014; **22**:218–25.

9. Watson BNJ, Staals RHJ, Fineran PC. CRISPR-cas-mediated phage resistance enhances horizontal gene transfer by transduction. Bondy-Denomy J, Gilmore MS, eds. *mBio* 2018; **9**.

Bibliography III

10. Godde JS, Bickerton A. The repetitive DNA elements called CRISPRs and their associated genes: Evidence of horizontal transfer among prokaryotes. *Journal of Molecular Evolution* 2006; **62**:718–729.

11. Dang UJ, Golding GB. Markophylo: Markov chain analysis on phylogenetic trees. *Bioinformatics* 2016; **32**:130–132.

12. Onnela JP, Saramaki J, Kertesz J, Kaski K. Intensity and coherence of motifs in weighted complex networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 2005; **71**:065103.

13. Newman ME. Assortative mixing in networks. *Phys Rev Lett* 2002; **89**:208701.

Bibliography IV

14. Newman ME. Analysis of weighted networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 2004; **70**:056131.