The Effects Of CRISPR-Cas Systems On The Rate Of Horizontal Gene Transfer

A Network Theoretic Approach

Siddharth Reed
Undergraduate Thesis

Golding Lab,
Biology Department,
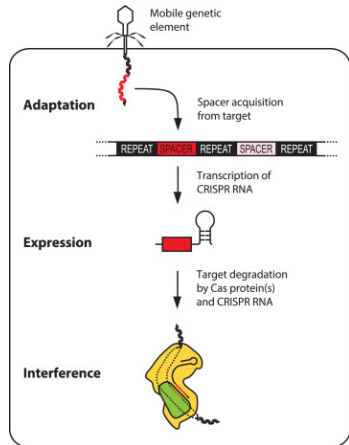McMaster University

January 22, 2019

Table of Contents

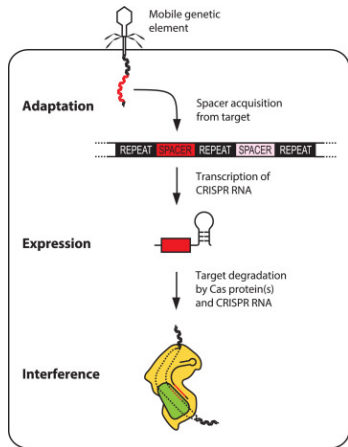# CRISPR-Cas systems

What Are They?

What Are They?
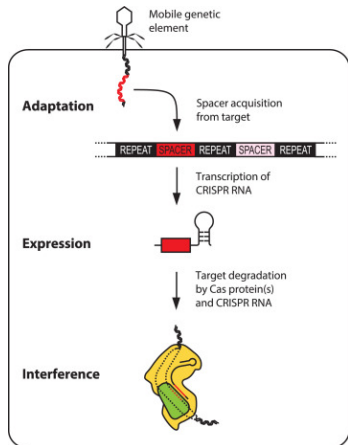
- Adaptive Bacterial Immune System

What Are They?

- Adaptive Bacterial Immune System
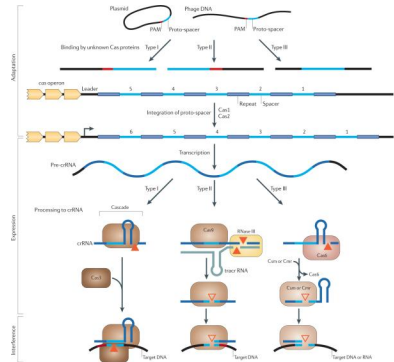- Protects against foreign DNA

1

What Are They?

- Adaptive Bacterial Immune System
- Protects against foreign DNA
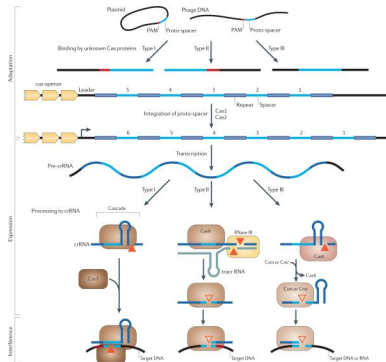- Requires Cas proteins and CRISPR loci

Diversity & Ubiquity

Diversity & Ubiquity

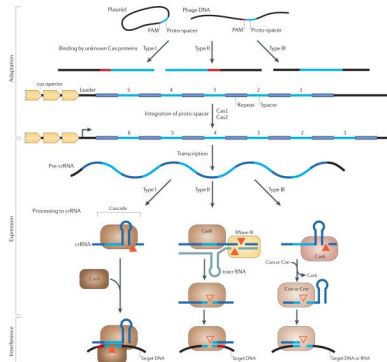- 45% of bacteria have CRISPR loci ($n = 6782$)[2]

Diversity & Ubiquity

- 45% of bacteria have CRISPR loci ($n = 6782$)[2]
- 3 Main Types, multiple subtypes[3]

Diversity & Ubiquity

- 45% of bacteria have CRISPR loci ($n = 6782$)[2]
- 3 Main Types, multiple subtypes[3]
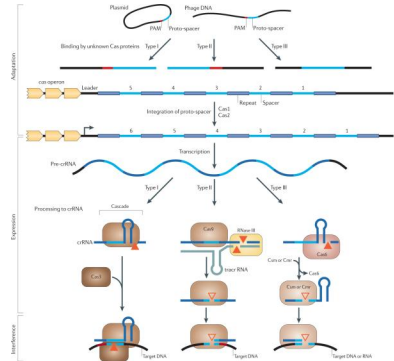- CRISPR arrays represent unique life history of an organism

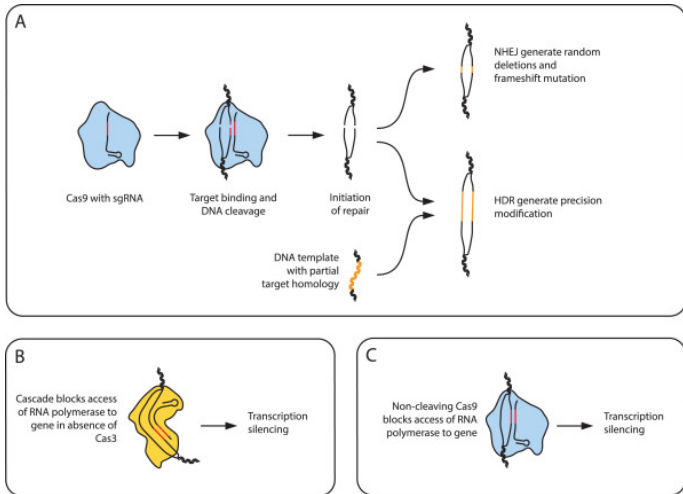# Diversity & Ubiquity

- 45% of bacteria have CRISPR loci ($n = 6782$)[2]
- 3 Main Types, multiple subtypes[3]
- CRISPR arrays represent unique life history of an organism
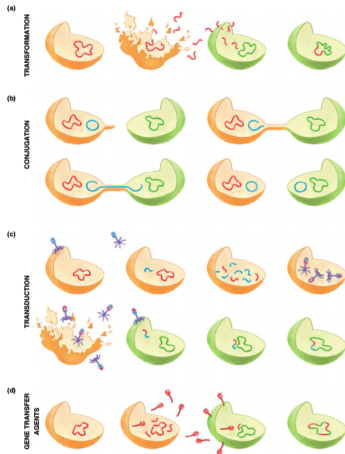- $11\% - 28\%$ are false or orphaned CRISPR loci[4]

Biotech Application

Biotech Application

1

Horizontal Gene Transfer

Mechanisms

6

Mechanisms

- Conjugation: Transfer of DNA through cell-cell connections[6]

6

Mechanisms

- Conjugation: Transfer of DNA through cell-cell connections[6]
- Transformation: Incorportaion of free-floating DNA into the genome[6]

6
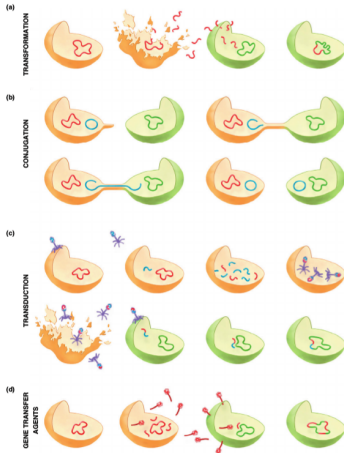
Mechanisms

- Conjugation: Transfer of DNA through cell-cell connections[6]
- Transformation: Incorporaion of free-floating DNA into the genome[6]
- Transduction: Transfer of DNA through phage[6]
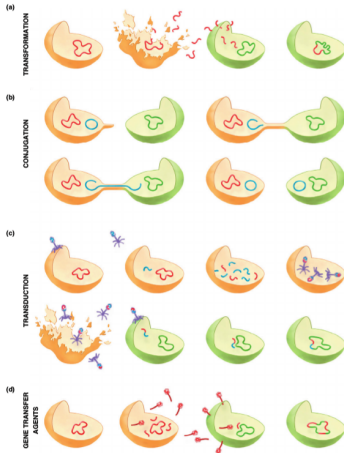
6

Mechanisms

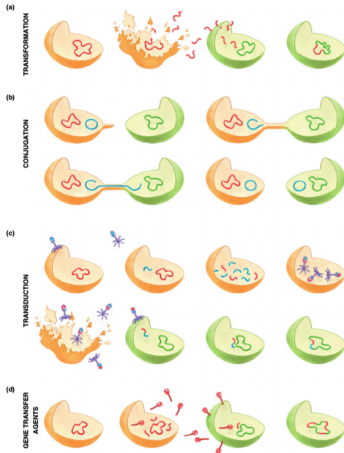- Conjugation: Transfer of DNA through cell-cell connections[6]
- Transformation: Incorportaion of free-floating DNA into the genome[6]
- Transduction: Transfer of DNA through phage[6]
- **CRISPR-Cas directly affects Transduction and Transformation**[6]

6

Pan-Genomes

Pan-Genomes
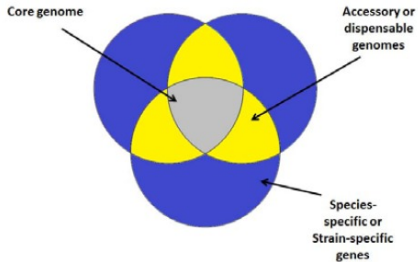
Pan-Genomes
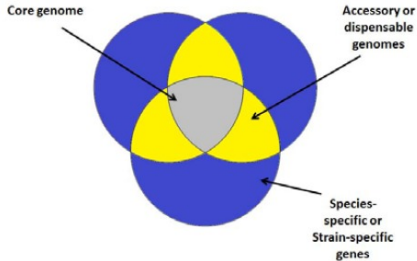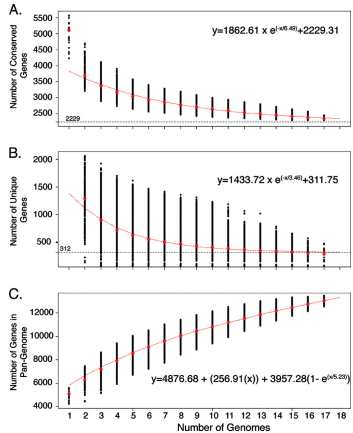
Rate Influencing Factors

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

Applications

Applications

9
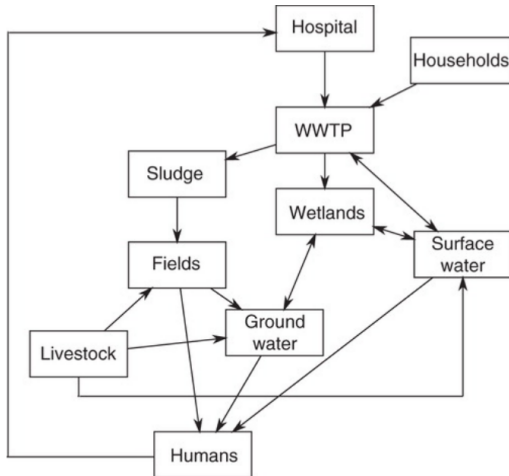
Phylogenomic Networks

What is A Network?

What is A Network?
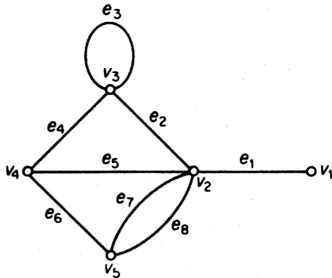
- Useful mathematical abstraction of real world system
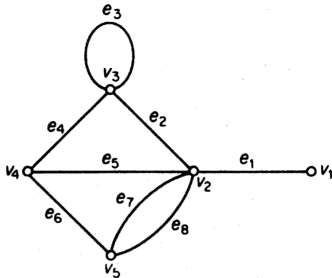
10

What is A Network?

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

10

What is A Network?

- Useful mathematical abstraction of real world system
- Nodes can have attributes
- Directed or Undirected Edges

10
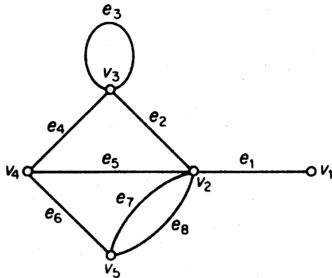
What is A Network?
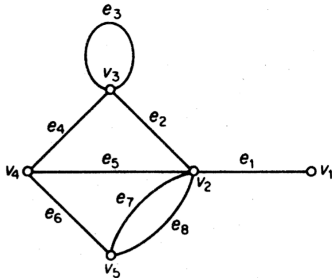
- Useful mathematical abstraction of real world system
- Nodes can have attributes
- Directed or Undirected Edges
- Weighted or Unweighted Edges

10

Prokaryotic "Net of Life"

Prokaryotic "Net of Life"
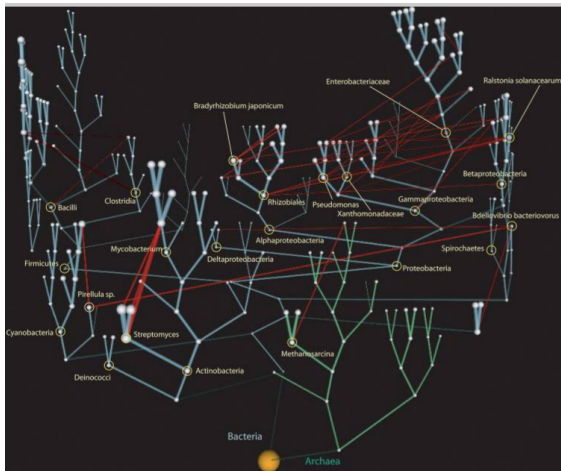
Construction
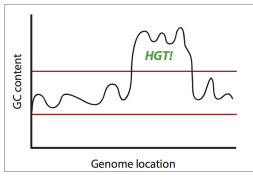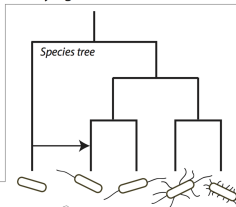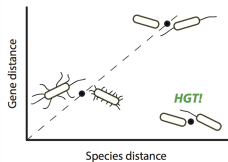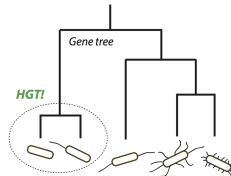
Do CRRISPR Systems Affect Horizontal Gene Transfer?

Yes

CRISPR Cost Complexity

CRISPR Cost Complexity

- Cost tradeoff factors:

CRISPR Cost Complexity

- Cost tradeoff factors:
  - Metabolic maintenace[1]

CRISPR Cost Complexity

- Cost tradeoff factors:
  - Metabolic maintenace[1]
  - Environmental pressures[13]

CRISPR Cost Complexity

- Cost tradeoff factors:
  - Metabolic maintenace[1]
  - Environmental pressures[13]
  - Off-target effects (autoimmune)[14]

CRISPR Cost Complexity

- Cost tradeoff factors:
  - Metabolic maintenace[1]
  - Environmental pressures[13]
  - Off-target effects (autoimmune)[14]
  - Anti-CRISPR systems[3]

CRISPR Cost Complexity

- Cost tradeoff factors:
  - Metabolic maintenace[1]
  - Environmental pressures[13]
  - Off-target effects (autoimmune)[14]
  - Anti-CRISPR systems[3]
  - Phage virulence/density[3]

CRISPR Cost Complexity

- Cost tradeoff factors:
  - Metabolic maintenace[1]
  - Environmental pressures[13]
  - Off-target effects (autoimmune)[14]
  - Anti-CRISPR systems[3]
  - Phage virulence/density[3]
  - Prophage abundance[15]

Curbing CRISPR Cost

Curbing CRISPR Cost

- CRISPRs themselves can be transfered $\implies$ population level immunity[16]

Curbing CRISPR Cost

- CRISPRs themselves can be transfered $\implies$ population level immunity[16]
- Selective CRISPR inactivation[1]

Curbing CRISPR Cost

- CRISPRs themselves can be transfered $\implies$ population level immunity[16]
- Selective CRISPR inactivation[1]
- CRISPR can enhance transduction-mediated HGT[15]

Previous Findings

Previous Findings

- Gophna et al. (2015) found that CRISPR has no effect on HGT over short evolutionary timescales[17]

Previous Findings

- Gophna et al. (2015) found that CRISPR has no effect on HGT over short evolutionary timescales[17]
  - Assume all singletons arrose from HGT

Previous Findings

- Gophna et al. (2015) found that CRISPR has no effect on HGT over short evolutionary timescales[17]
  - Assume all singletons arrose from HGT
  - Used GC% to identify HGT

Previous Findings

- Gophna et al. (2015) found that CRISPR has no effect on HGT over short evolutionary timescales[17]
  - Assume all singletons arrose from HGT
  - Used GC% to identify HGT
- Contradicted by a former thesis student

Previous Findings

- Gophna et al. (2015) found that CRISPR has no effect on HGT over short evolutionary timescales[17]
  - Assume all singletons arrose from HGT
  - Used GC% to identify HGT
- Contradicted by a former thesis student
  - Can see inhibitory effects of CRIPSR on HGT over short evolutionary time scales

Previous Findings

- Gophna et al. (2015) found that CRISPR has no effect on HGT over short evolutionary timescales[17]
  - Assume all singletons arrose from HGT
  - Used GC% to identify HGT
- Contradicted by a former thesis student
  - Can see inhibitory effects of CRIPSR on HGT over short evolutionary time scales
  - Higher gene indel rates for CRISPR containg genera than non-CRISPR containing outgroups

My Project

Hypothesis

Null Hypothesis

Bacterial strains or genera with known CRISPR systems will show no significant differences in network statistics compared to those strains or genera without known CRISPR systems.

Hypothesis

Null Hypothesis

Bacterial strains or genera with known CRISPR systems will show no significant differences in network statistics compared to those strains or genera without known CRISPR systems

Alternative Hypothesis

Bacterial strains or genera with known CRISPR systems will show a significant difference in at least 1 network statistic compared to those strains or genera without known CRISPR systems.

Objectives

Objectives

Within Network Comparisons

For genera with CRISPR conatining strains, compare the node statistics of CRIPSR-containing strain to non-CRISPR-containing strains.

Objectives

Within Network Comparisons

For genera with CRISPR conatining strains, compare the node statistics of CRIPSR-containing strain to non-CRISPR-containing strains.

Between Network Comparisons

For genera with no CRISPR conatining strains, compare the network statistics of mixed to non-CRISPR-containing networks.

Objectives

Within Network Comparisons

For genera with CRISPR conating strains, compare the node statistics of CRIPSR-containing strain to non-CRISPR-containing strains.

Between Network Comparisons

For genera with no CRISPR conating strains, compare the network statistics of mixed to non-CRISPR-containing networks.
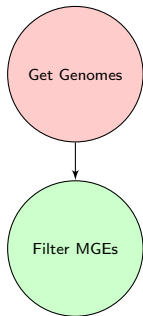
Gene Indel Rates vs. Network Statistics

Compare gene InDel rates to node/network statistics for CRISPR-containing and non-CRISPR-containing strains/genera.
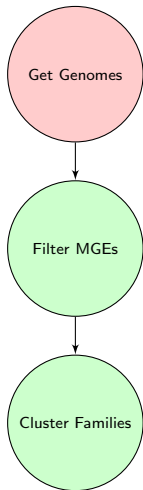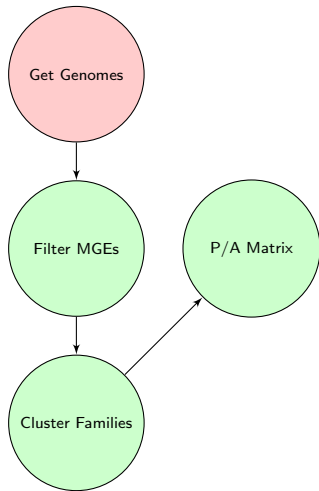
Workflow

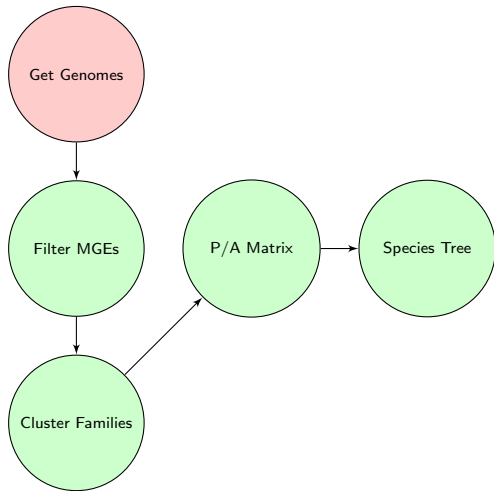Get Genomes

Workflow

Workflow

Workflow
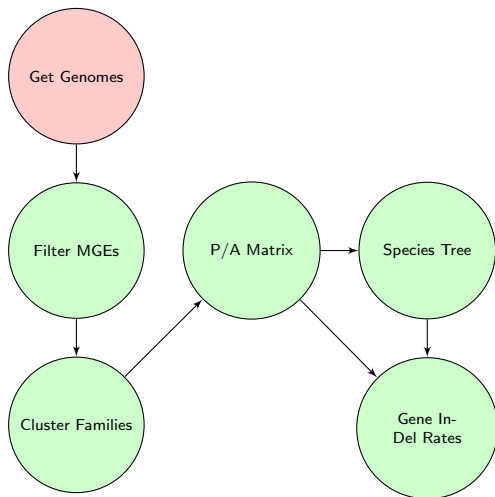
Workflow
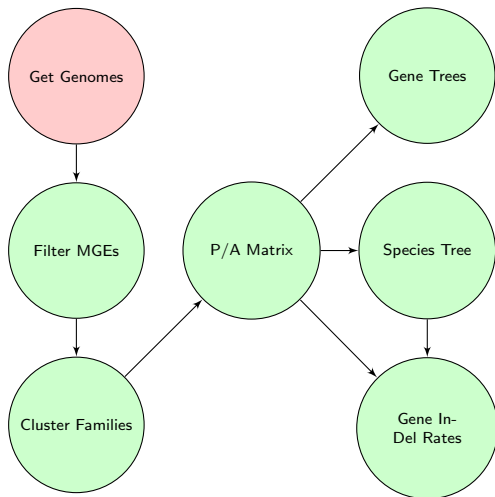
Workflow
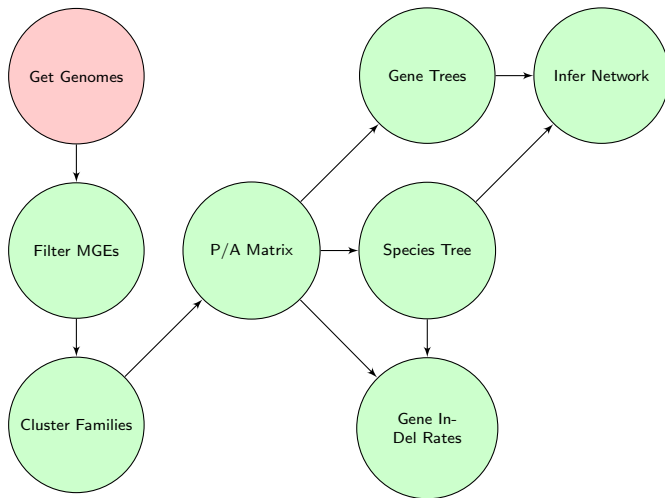
Workflow
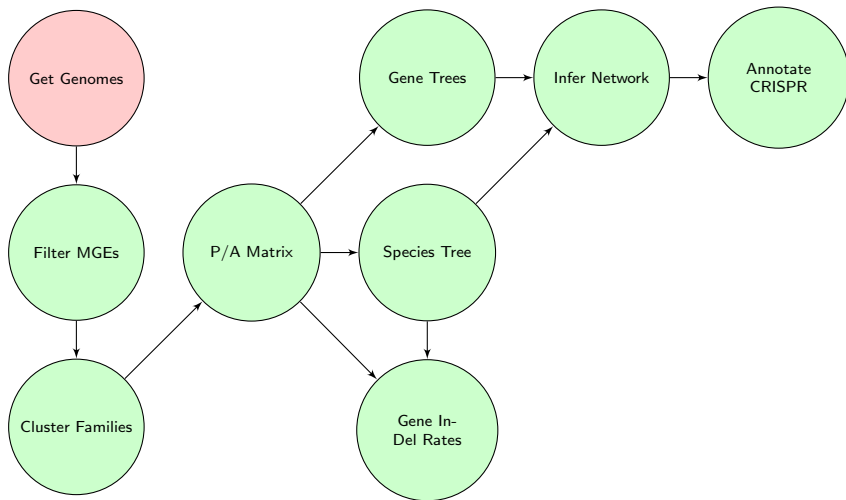
Workflow
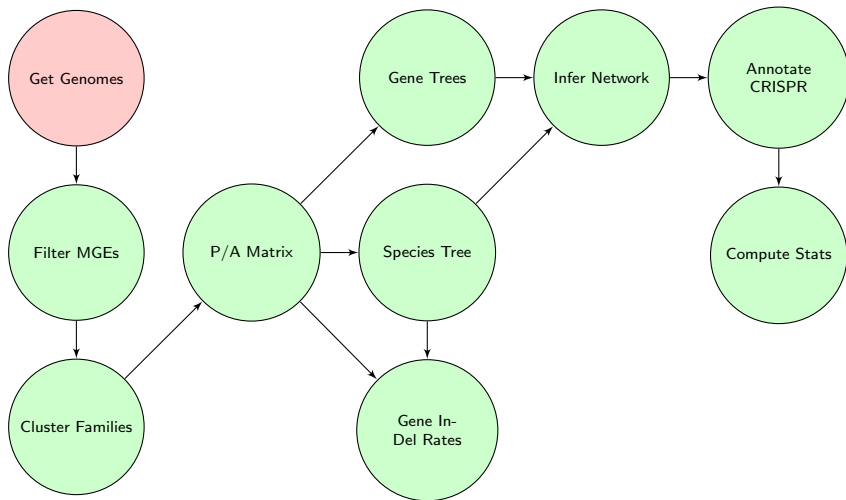
Workflow

Workflow

Workflow
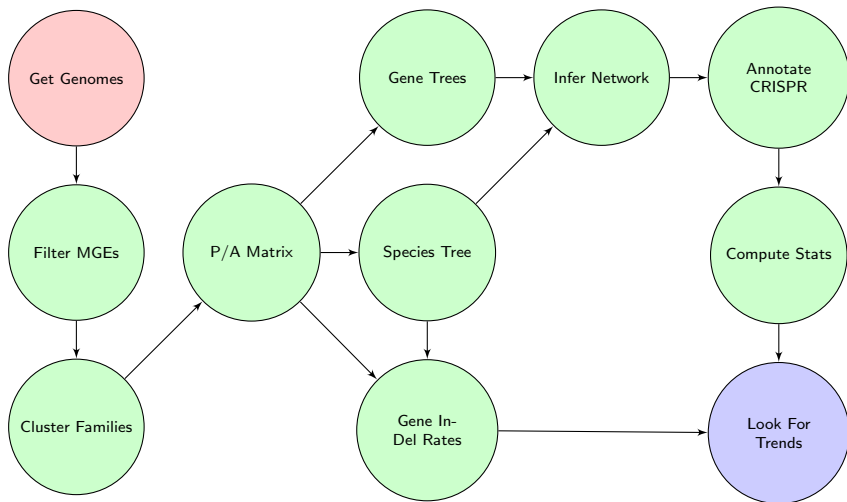
Network Statistics

Network Statistics

- **Total/Average edge weight**: $\sum_i w_i, \frac{\sum_i w_i}{N}$

Network Statistics

- **Total/Average edge weight**: $\sum_i w_i, \frac{\sum_i w_i}{N}$
- **Network Density**: $\frac{2E}{N(N-1)}$

Network Statistics

- **Total/Average edge weight**: $\sum_i w_i, \frac{\sum_i w_i}{N}$
- **Network Density**: $\frac{2E}{N(N-1)}$
- **Network Diameter**: Shortest path between the 2 furthest nodes.

Network Statistics

- **Total/Average edge weight**: $\sum_i w_i, \frac{\sum_i w_i}{N}$
- **Network Density**: $\frac{2E}{N(N-1)}$
- **Network Diameter**: Shortest path between the 2 furthest nodes.
- **Node Closeness Centrality**: $\frac{N-1}{\sum_v d(u,v)}$ where $d(x,y)$ is the length of the shortest path $v \rightarrow u$.

Network Statistics

- **Total/Average edge weight**: $\sum_i w_i, \frac{\sum_i w_i}{N}$
- **Network Density**: $\frac{2E}{N(N-1)}$
- **Network Diameter**: Shortest path between the 2 furthest nodes.
- **Node Closeness Centrality**: $\frac{N-1}{\sum_v d(u,v)}$ where $d(x,y)$ is the length of the shortest path $v \rightarrow u$.
- **Node Clustering Coefficient**: $\frac{2e}{k(k-1)}$ where k is the neighbors and e is the edges between all neighbors.

Network Statistics

- **Total/Average edge weight**: $\sum_i w_i, \frac{\sum_i w_i}{N}$
- **Network Density**: $\frac{2E}{N(N-1)}$
- **Network Diameter**: Shortest path between the 2 furthest nodes.
- **Node Closeness Centrality**: $\frac{N-1}{\sum_v d(u,v)}$ where $d(x,y)$ is the length of the shortest path $v \rightarrow u$.
- **Node Clustering Coefficient**: $\frac{2e}{k(k-1)}$ where k is the neighbors and e is the edges between all neighbors.
- **Edge Weight KL Divergence**:
  $D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x)log(\frac{p(x)}{q(x)})dx$

Network Statistics

- **Total/Average edge weight**: $\sum_i w_i, \frac{\sum_i w_i}{N}$
- **Network Density**: $\frac{2E}{N(N-1)}$
- **Network Diameter**: Shortest path between the 2 furthest nodes.
- **Node Closeness Centrality**: $\frac{N-1}{\sum_v d(u,v)}$ where $d(x,y)$ is the length of the shortest path $v \to u$.
- **Node Clustering Coefficient**: $\frac{2e}{k(k-1)}$ where k is the neighbors and e is the edges between all neighbors.
- **Edge Weight KL Divergence**:
  $D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) log(\frac{p(x)}{q(x)}) dx$
- **Node Associativity**: $\frac{j(j+1)(\overline{k}-\mu_q)}{2E\sigma_q^2}$ where j is the excess degree of the node and $\overline{k}$ is the average excess degree of the node's neighbors and μ_q and σ_q are the mean and standard variation of the excess degree distribution.

Thanks

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

References (1)

Devashish Rath et al. "The CRISPR-Cas immune system: Biology, mechanisms and applications". In: *Biochimie* 117 (2015). Special Issue: Regulatory RNAs, pp. 119–128. ISSN: 0300-9084. DOI: https://doi.org/10.1016/j.biochi.2015.03.025. URL: http://www.sciencedirect.com/science/article/pii/S0300908415001042.

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

References (2)

📄 J. Bondy-Denomy and A. R. Davidson. "To Acquire Or Resist:The Complex Biological Effects Of CRISPR-Cas systems". In: *Trends Microbio.* 22.4 (Apr. 2014), pp. 218–25. DOI: 10.1016/j.tim.2014.01.007.

📄 Quan Zhang and Yuzhen Ye. "Not all predicted CRISPR–Cas systems are equal: isolated cas genes and classes of CRISPR like elements". In: *BMC Bioinformatics* 18.1 (Feb. 2017), p. 92. ISSN: 1471-2105. DOI: 10.1186/s12859-017-1512-4. URL: https://doi.org/10.1186/s12859-017-1512-4.

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

📄 Ovidiu Popa and Tal Dagan. "Trends and barriers to lateral gene transfer in prokaryotes". In: *Current Opinion in Microbiology* 14.5 (2011). Antimicrobials/Genomics, pp. 615–623. ISSN: 1369-5274. DOI: https://doi.org/10.1016/j.mib.2011.07.027. URL: http://www.sciencedirect.com/science/article/pii/S1369527411001111.

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

📄 David A. Rasko et al. "The Pangenome Structure of Escherichia coli: Comparative Genomic Analysis of E. coli Commensal and Pathogenic Isolates". In: *Journal of Bacteriology* 190.20 (2008), pp. 6881–6893. ISSN: 0021-9193. DOI: 10.1128/JB.00619-08. eprint: https://jb.asm.org/content/190/20/6881.full.pdf. URL: https://jb.asm.org/content/190/20/6881.

📄 Björn Berglund. "Environmental dissemination of antibiotic resistance genes and correlation to anthropogenic contamination with antibiotics". In: *Infection Ecology & Epidemiology* 5.1 (2015), p. 28564. DOI: 10.3402/iee.v5.28564. eprint: https://doi.org/10.3402/iee.v5.28564. URL: https://doi.org/10.3402/iee.v5.28564.

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

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

📄 Matt Ravenhall et al. "Inferring Horizontal Gene Transfer". In: *PLOS Computational Biology* 11.5 (May 2015), pp. 1–16. DOI: 10.1371/journal.pcbi.1004095. URL: https://doi.org/10.1371/journal.pcbi.1004095.

📄 Senka Dzidic and Vladimir Bedeković. "Horizontal gene transfer-emerging multidrug resistance in hospital bacteria". In: *Acta pharmacologica Sinica* 24.6 (June 2003), pp. 519–526. ISSN: 1671-4083. URL: http://www.chinaphar.com/1671-4083/24/519.htm.

📄 Adi Stern et al. "Self-targeting by CRISPR: gene regulation or autoimmunity?" In: *Trends in Genetics* 26.8 (2010), pp. 335–340. ISSN: 0168-9525. DOI: https://doi.org/10.1016/j.tig.2010.05.008. URL: http://www.sciencedirect.com/science/article/pii/S0168952510001083.

📄 Bridget N. J. Watson, Raymond H. J. Staals, and Peter C. Fineran. "CRISPR-Cas-Mediated Phage Resistance Enhances Horizontal Gene Transfer by Transduction". In: *mBio* 9.1 (2018). Ed. by Joseph Bondy-Denomy and Michael S. Gilmore. DOI: 10.1128/mBio.02406-17. eprint: https://mbio.asm.org/content/9/1/e02406-17.full.pdf. URL: https://mbio.asm.org/content/9/1/e02406-17.

📄 James S. Godde and Amanda Bickerton. "The Repetitive DNA Elements Called CRISPRs and Their Associated Genes: Evidence of Horizontal Transfer Among Prokaryotes". In: *Journal of Molecular Evolution* 62.6 (June 2006), pp. 718–729. ISSN: 1432-1432. DOI: 10.1007/s00239-005-0223-z. URL: https://doi.org/10.1007/s00239-005-0223-z.

📄 U. Gophna et al. "No evidence of inhibition of horizontal gene transfer by CRISPR-Cas on evolutionary timescales". In: *ISME J* 9.9 (Sept. 2015), pp. 2021–2027.