

Coinfinder: Detecting Significant Associations and Dissociations in Pangenomes

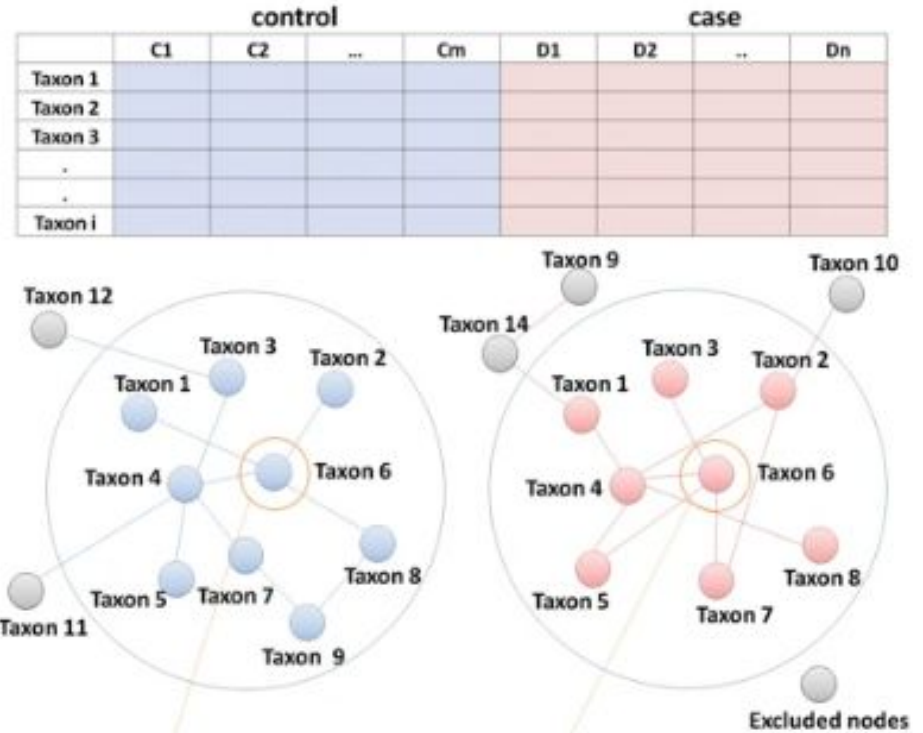
Fiona Jane Whelan, Martin Rusilowicz, James Oscar
McInerney

Gene to Gene Interactions in Pangenome?

- Pangenome = Core + Accessory
- Why even have an Accessory Genome?
 - Antibiotic Resistance
 - Mobile Elements
 - Pathogenicity/ Host adaptation,
- We should expect that genes that are in the accessory genomes should have some relationship with each other

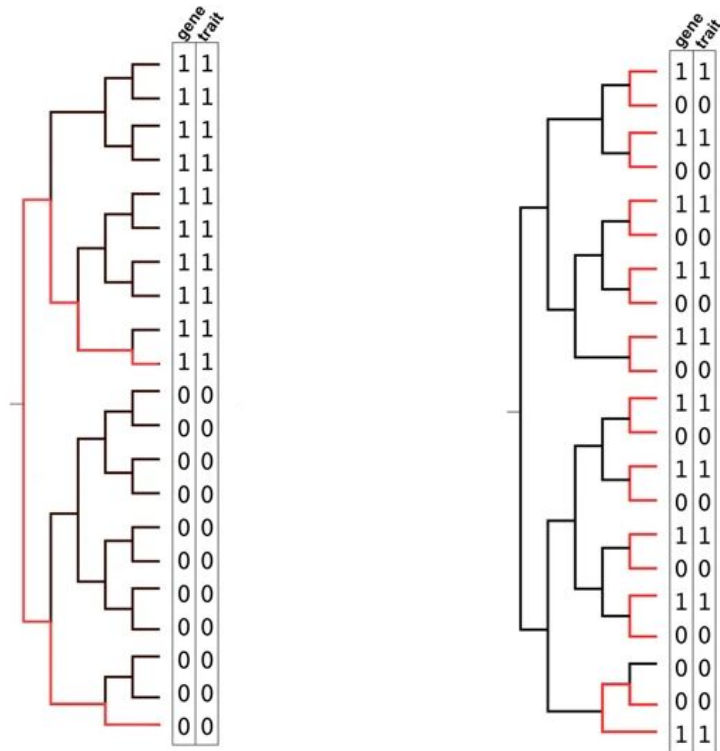
Attempts at Understanding Correlations

- Microbial community level
 - SparCC
 - NetShift



Attempts at Understanding Correlations

- pan-GWAS
 - Scoary



Attempts at Understanding Correlations

- Phyletic Gene-Gene Correlation
 - CoPAP
 - Pantagruel
- Can this be done without having to produce phylogenies? Also, Can we show gene dissociations as well?



Coinfinder Implementation

- Input:
 - Gene_presence_absence.csv
 - Tree inferred from core genome
- Date processing
 - Core Genomes
 - Low abundance gene (user defined cutoff)
- Two modes:
 - Association
 - Dissociation

Co-occurrence

- The probability of seeing gene i in a genome
- Expected number of times seeing genes i and j together in the dataset
- The actual number of times we see genes i and j together in the dataset
- We now have everything to do a significance test

$$P_i = N_i / N$$

$$E_A(ij) = P_i * P_j * N$$

$$O_A(ij) = N_{ij}$$

Bonferroni-corrected Binomial Exact Test

- $P_i = 8/10, P_j = 8/10$
- $E_{ij} = (0.8)(0.8)100 = 64$
- $O_{ij} = 80$
- Bonferroni Alpha = $0.05/100 = 0.0005$
- Compute a probability using binomial distribution = 0.00022366563
 - $0.00022366563 < 0.0005$
 - In this dataset, genes co occur more than expected

[illegible]

Dissociation

- The probability of seeing gene i in a genome
- Expected number of times seeing genes i and j not together in the dataset
- The actual number of times we see genes i and j not together in the dataset
- We now have everything to do a significance test

$$P_i = N_i / N$$

$$E_D(ij) = [P_i(1-P_j) + P_j(1-P_i)] * N$$

$$O_D(ij) = N_i + N_j - 2N_{ij}$$

Bonferroni-corrected Binomial Exact Test

- $P_i = 50/100$, $P_j = 50/100$
- $E_{ij} = (0.5)(0.5)100 = 25$
- $O_{ij} = 100$
- Bonferroni Alpha = $0.05/100 = 0.0005$
- Compute a probability using binomial distribution = 0.000001
 - $0.000001 < 0.0005$
 - In this dataset, genes dissociate more than expected.

[illegible]

Wading Through Results

- Are the changes lineage dependent or lineage independent
 - Why we need an input tree
- Outputs
 - Gene Association/Dissociation Networks
 - Gene Presence Heatmap

Outputs

