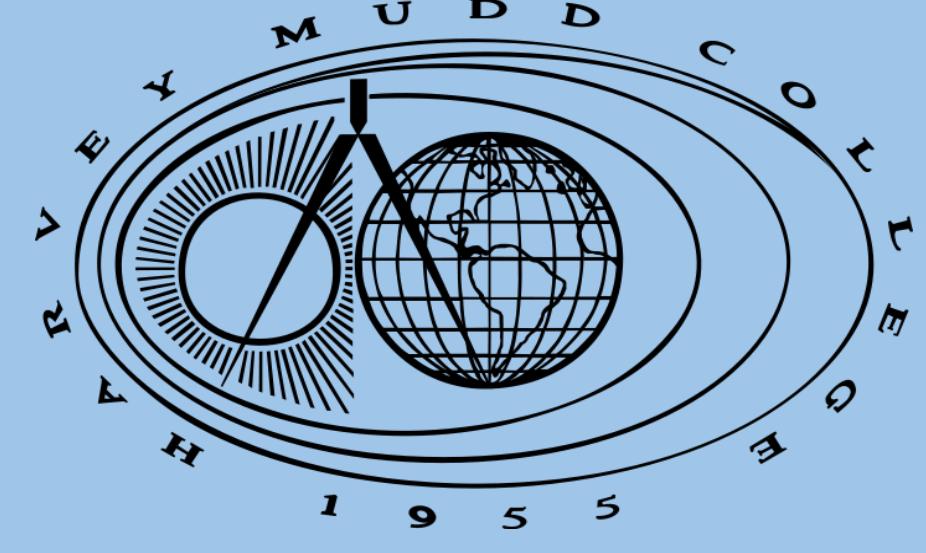# Identifying Horizontal Gene Transfer Events

Kevin Heath, Zunyan Wang, Eliot Bush

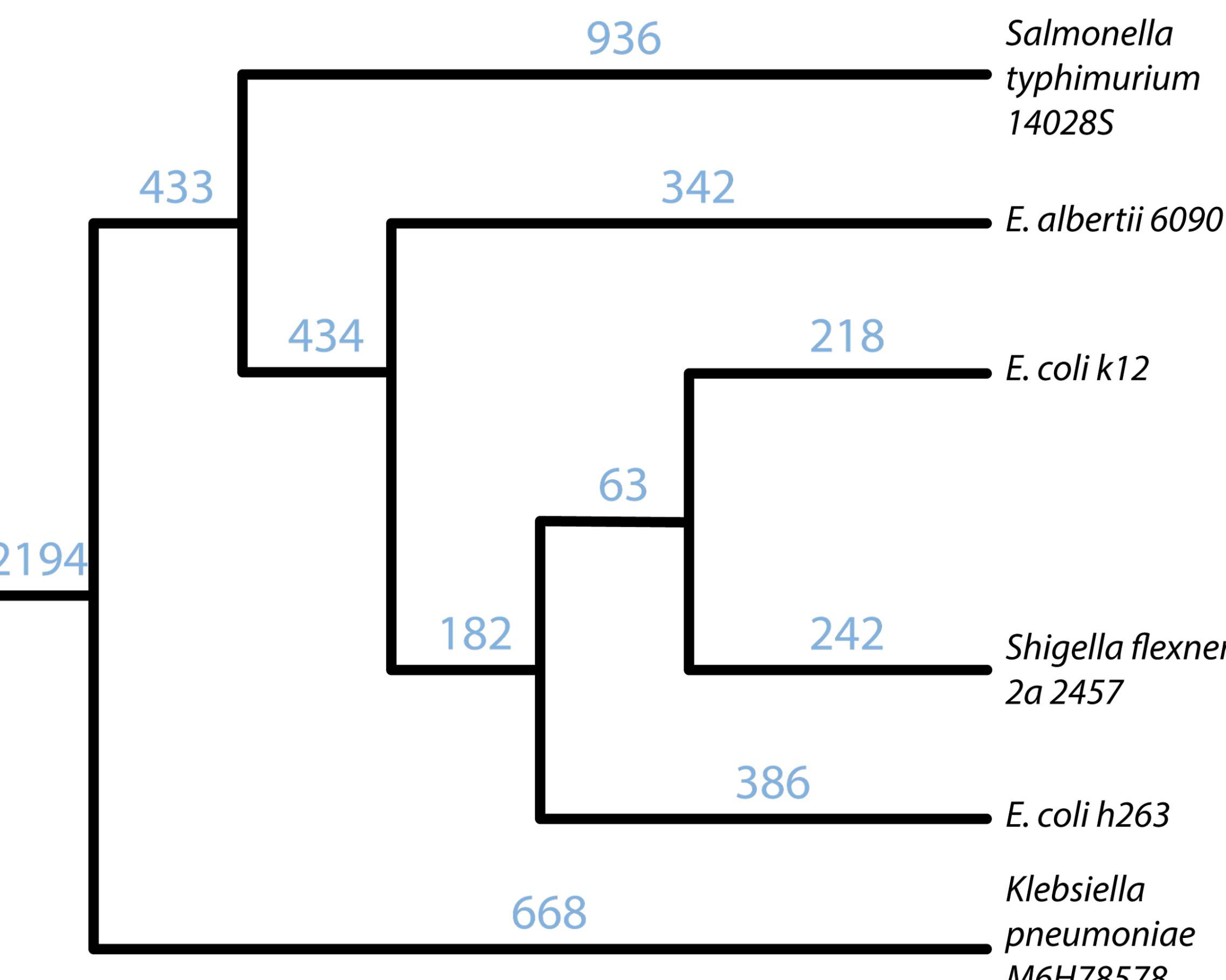Harvey Mudd College, Department of Biology

hhmi

Introduction

Horizontal gene transfer (HGT) is the transfer of genes from one organism to another unrelated organism. HGT events play an important role in the rapid evolution of bacteria and contribute greatly to the diversity of bacterial strains.

Many studies have investigated the horizontal transfer of important functional genes, such as those conferring antibiotic resistance or virulence. However, the problem of identifying all the HGT events during the evolutionary history of a clade of bacteria has been tackled less frequently. We are developing an algorithm to find what HGT events have occurred, how many genes were transferred in a single event, and when the event took place.

General Algorithm

Preprocessing Bacterial Genomes

SiLix output   Gene <-> Species Map

Gene order   Species   Gene Identifiers

↓

Pairwise Adjacency Information

↓

Calculate Pairwise Distance Matrix

↓

Greedily merge gene families with zero cost

Identifying HGT Events

In order to organize HGT events, we first need to identify genes in different strains that are identical or highly similar. If we suspect two or more of these sets of genes – commonly called gene families – entered an ancestral bacterial genome at similar locations and are commonly found adjacent to each other in most living descendants, it is likely that these families were part of a single larger HGT event rather than independent events. We group these gene families together and use them to identify other gene families that belong to the same event.

We developed an algorithm to calculate the cost of putting two gene families together in a group. Rearrangement of a genome incurs high evolutionary costs, so lower costs correspond to a higher likelihood that the two families belong to the same HGT event.

Start
- Two gene families, A & B
- Related groups
- MRCA

Recurse on left & right branches

Leaf? → Yes

No

Groups still exist?

Find all adjacent gene family pairs at each node

Yes

No

Find the next two adjacent gene families (C and D) and their groups

Output Cost

Output 0

Is D in this group?

Does the group with C have the same MRCA as A?

No

Yes

Skip this group and try the next group.

Keep this group
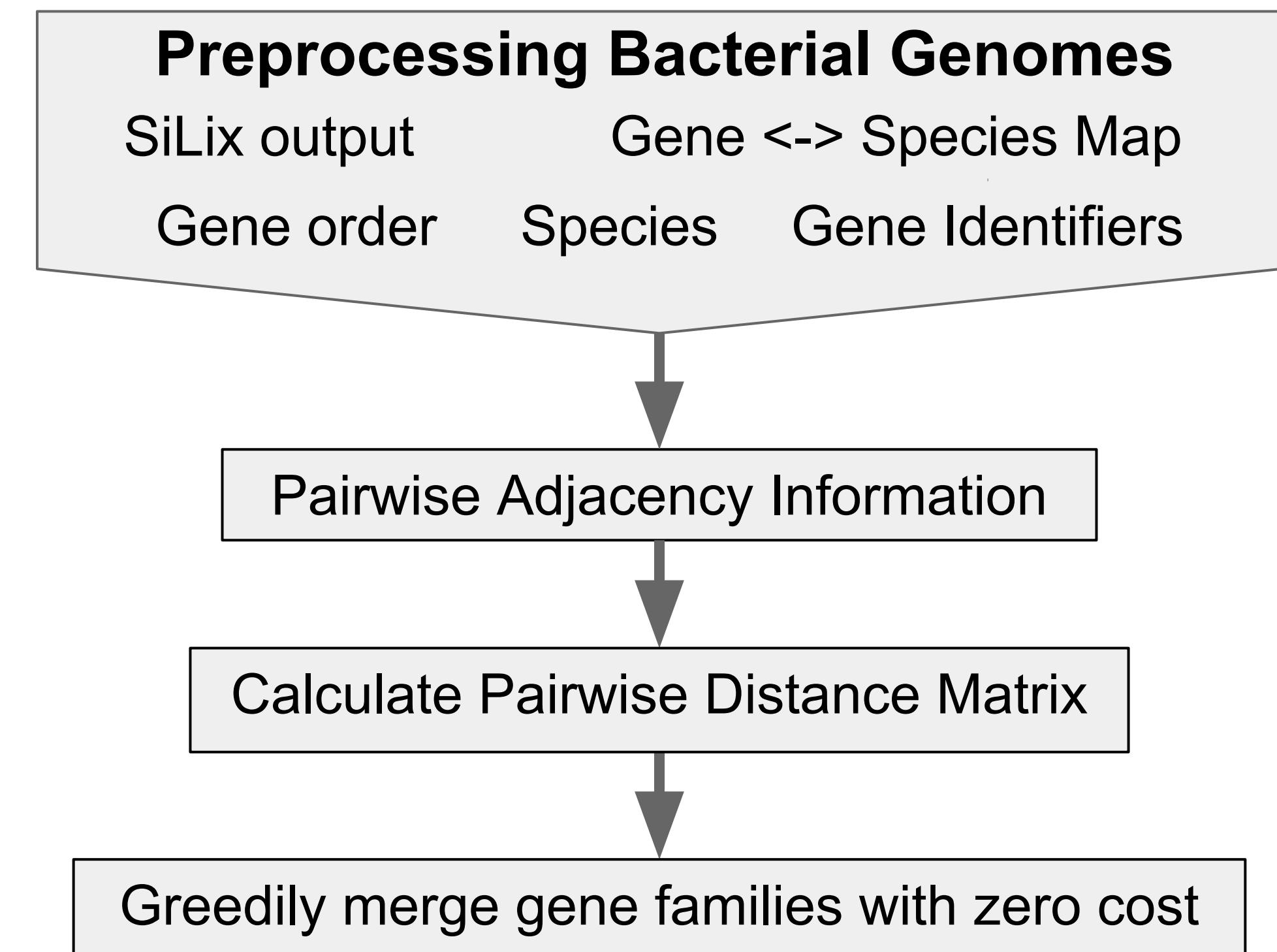
No   Yes

Preliminary Results

We tested our algorithm on a subset of six species. Using only stringent merging (merging only if the pairwise distance was zero), we calculated how many horizontal gene transfer events (HGT) occurred at each node in the phylogenetic tree.

Group Size	# of Families	Group Size	# of Families
1	4512	11	14
2	665	12	16
3	280	13	10
4	184	14	11
5	109	15	10
6	81	16	8
7	47	17	6
8	38	18	4
9	36	19	3
10	18	20+	52

936 — *Salmonella typhimurium 14028S*

433

342 — *E. albertii 6090*

434

218 — *E. coli k12*

63

182

242 — *Shigella flexneri 2a 2457*

2194

386 — *E. coli h263*

668 — *Klebsiella pneumoniae M6H78578*

Our stringent approach to identifying HGT events shows promising results. Out of 12,469 original gene families, 7,957 families were stringently merged into groups with two or more families.

The GAD acid fitness island (AFI) is a well-studied set of 12 genes that help E. coli survive in extremely acidic environments. Three of our six test species have the AFI, and our algorithm has a quarter of the AFI genes as a single HGT event so far.

Future Work

The primary focus of future research will be developing an algorithm to identify horizontal gene transfer events and merging gene families into groups in the non-stringent case. We hope to identify all genes associated with gene islands such as the AFI and group them into one HGT event. The overall goal of this project is to identify HGT events across 98 bacterial strains for further analysis.
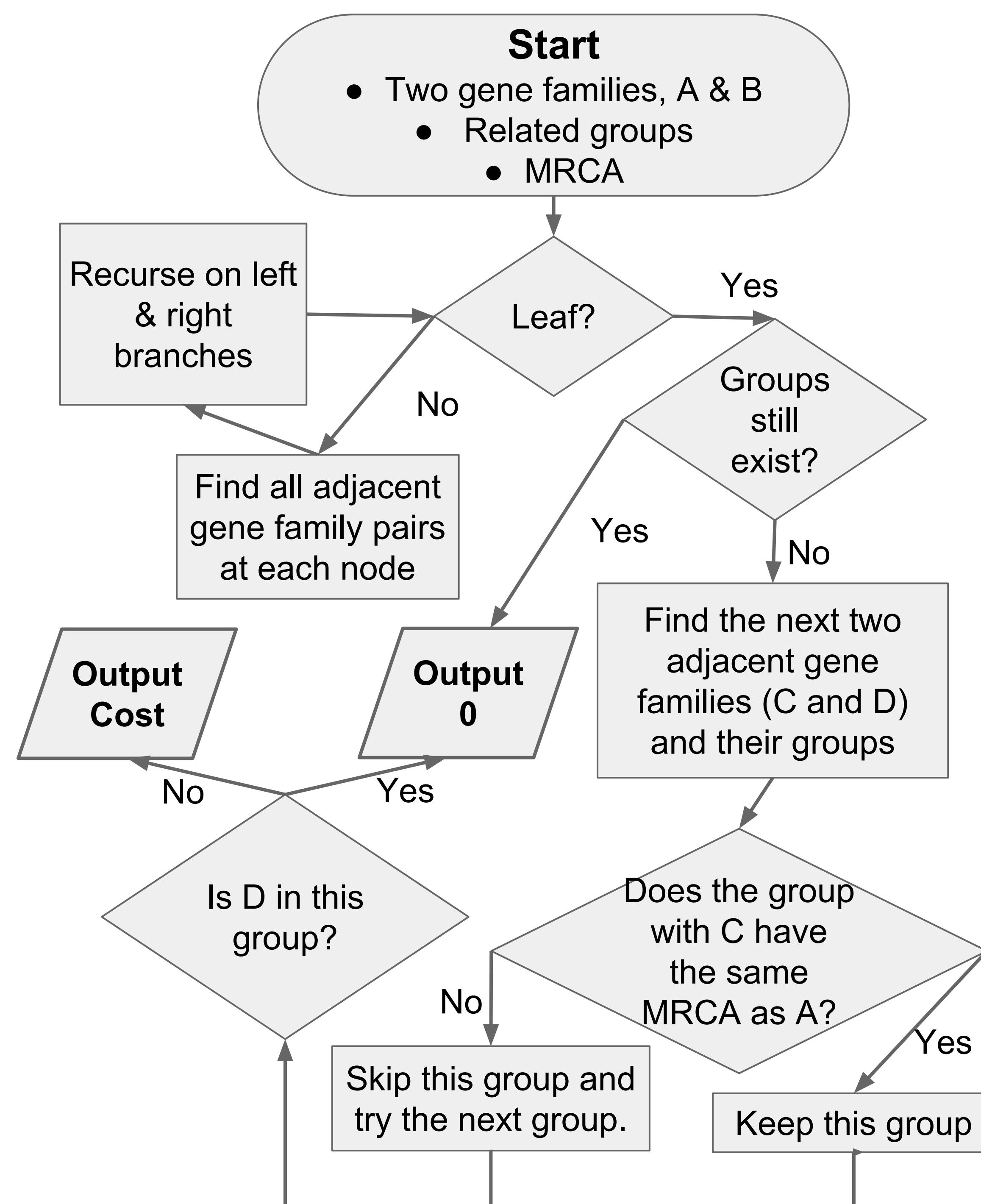
Works Cited

Bergholz, Teresa M., Cheryl L. Tarr, Lisa M. Christensen, David J. Betting, and Thomas S. Whittam. 2007. "Recent Gene Conversions between Duplicated Glutamate Decarboxylase Genes (gadA and gadB) in Pathogenic Escherichia Coli." *Molecular Biology and Evolution* 24 (10): 2323–33. doi:10.1093/molbev/msm163.

Becq, Jennifer, Cécile Churlaud, and Patrick Deschavanne. 2010. "A Benchmark of Parametric Methods for Horizontal Transfers Detection." *PLoS ONE* 5 (4): e9989. doi: 10.1371/journal.pone.0009989.

Acknowledgements