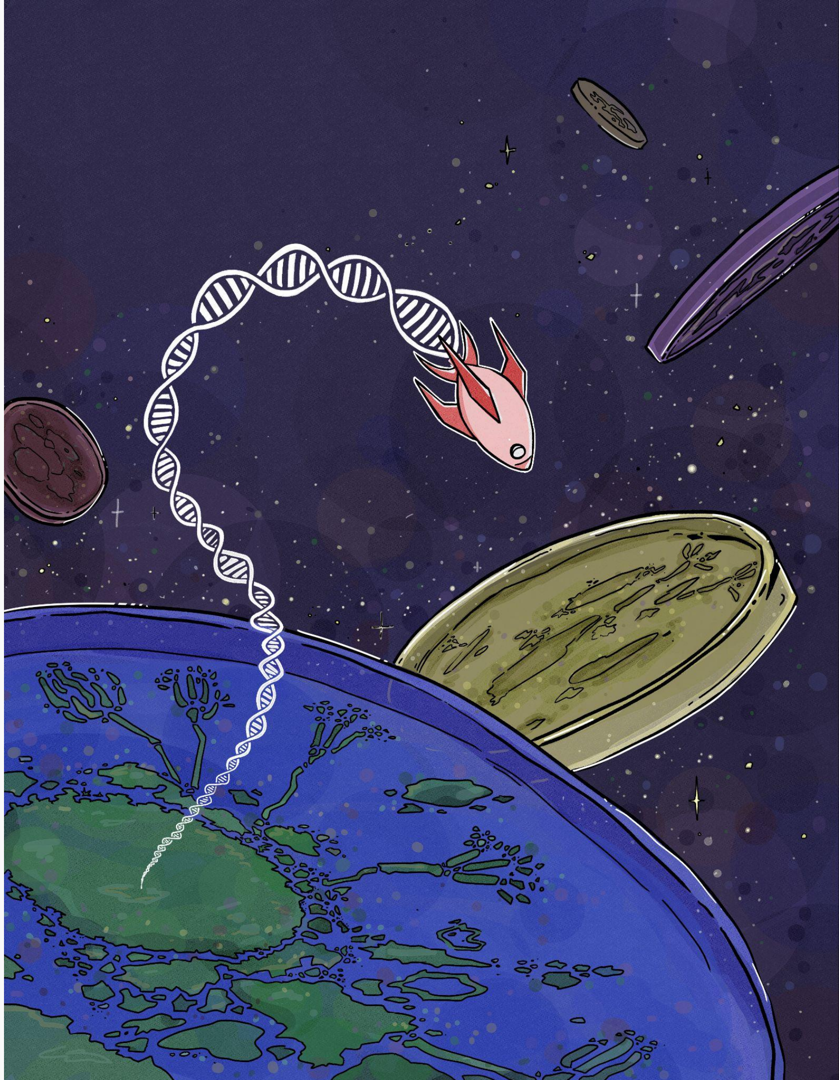
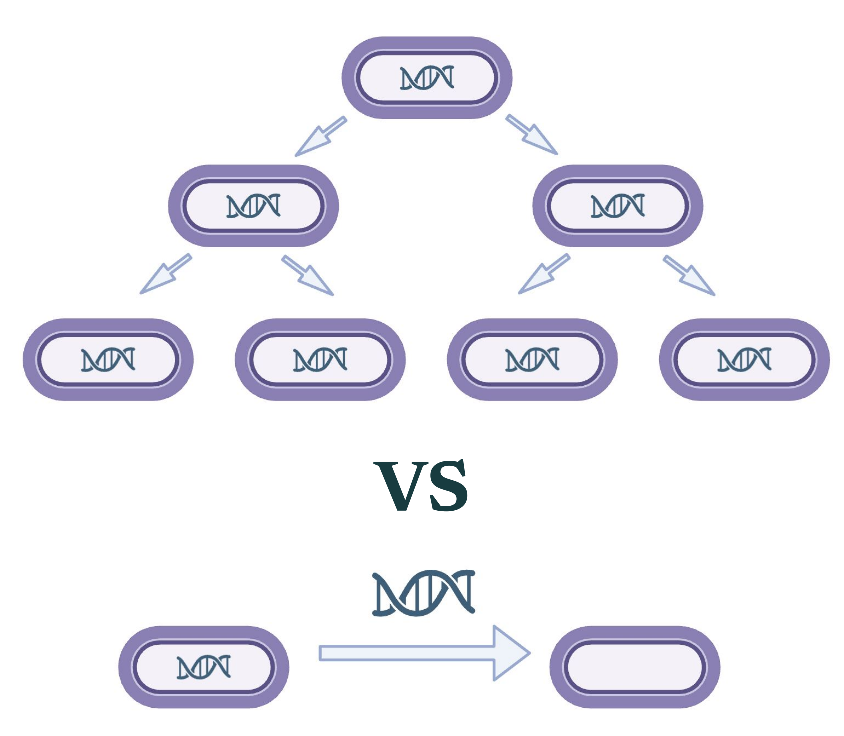
High Throughput Computing for Comparative Genomics on Large Public Datasets

Conor Bendett
Department of Plant Pathology
HTC25



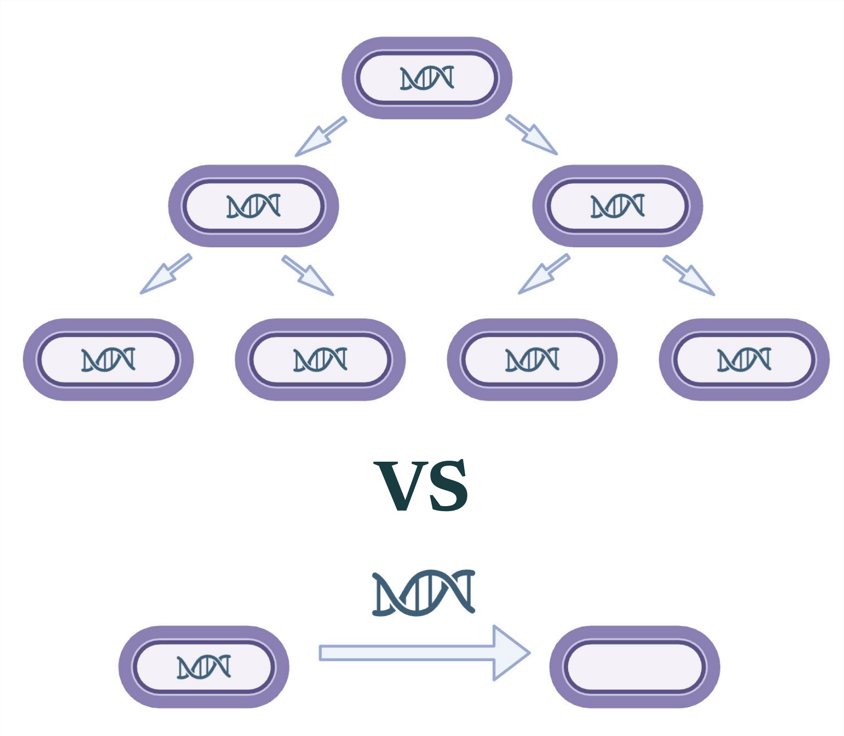
Vertical Gene Transfer



Horizontal Gene Transfer

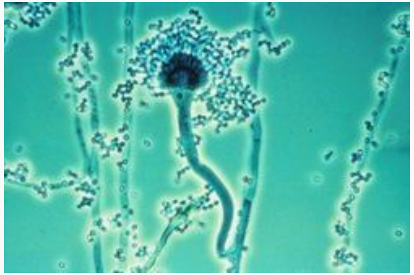
Horizontal gene transfer is a well-studied driver of evolution in bacteria, but we know very little about it in organisms like plants, animals, and fungi.

Vertical Gene Transfer



Horizontal Gene Transfer

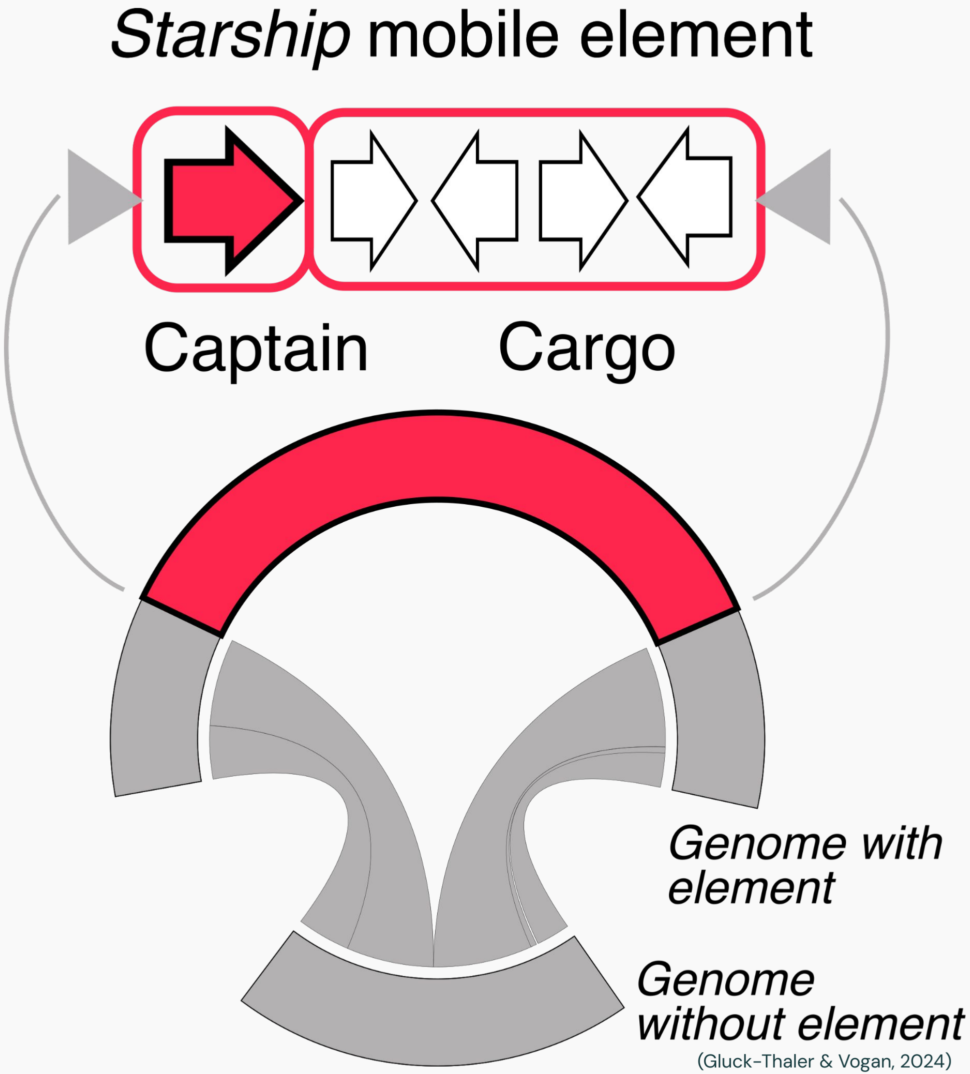
Horizontal gene transfer is a well-studied driver of evolution in bacteria, but we know very little about it in organisms like plants, animals, and **fungi.**



https://en.wikipedia.org/wiki/Amanita_muscaria
https://en.wikipedia.org/wiki/Aspergillus_fumigatus

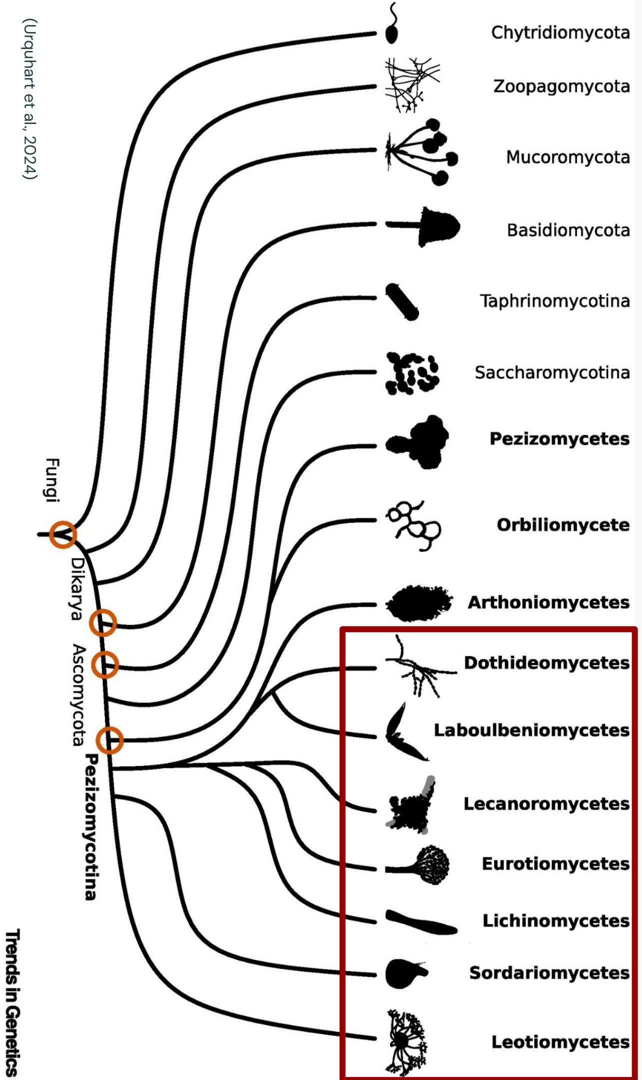
Starships are mobile genetic elements that can horizontally transfer between species and carry adaptive genes.

Detecting *Starships* requires high quality genomic data from different individuals of the same species.



(Gluck-Thaler & Vogan, 2024)

Starships have been found in the group Pezizomycotina, which includes most molds.

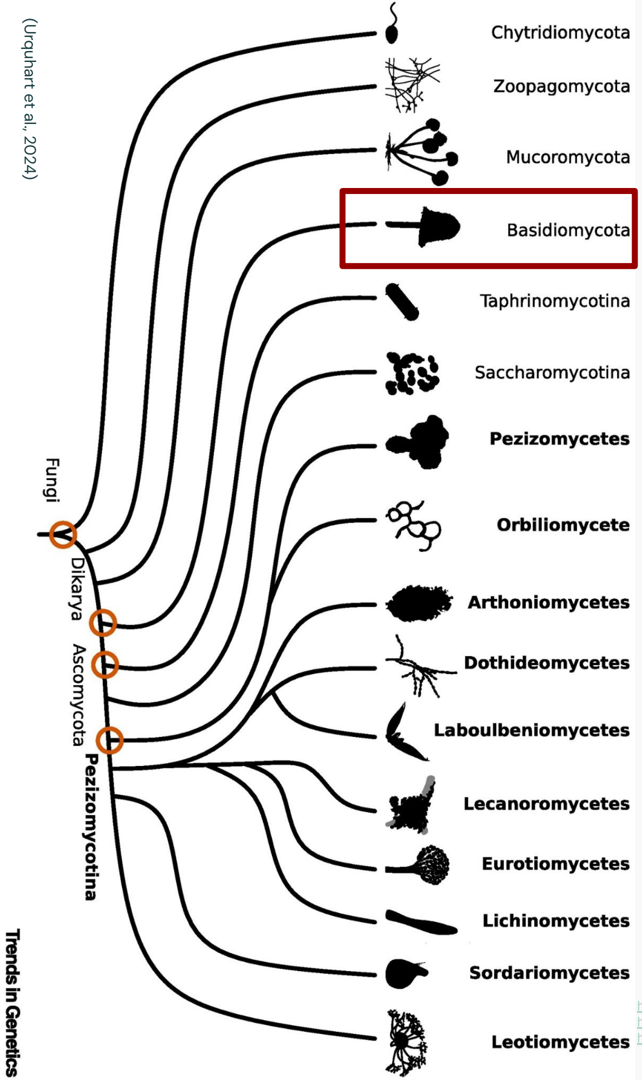


(Urquhart et al., 2024)

Trends in Genetics

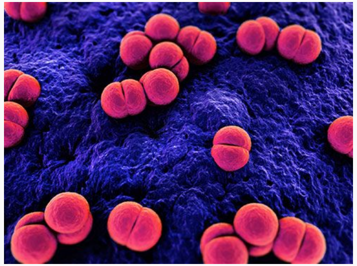
https://apps.lucidcentral.org/pppw_v11/text/web_full/entities/citrus_storage_moulds_197.htm
https://en.wikipedia.org/wiki/Physcia
https://www.amazon.com/Penicillium-Roqueforti-1D-Production-Cheeses/dp/B078YZPWF8

My work has been looking for *Starships* in the Basidiomycota, which includes most 'mushrooms.'



(Urquhart et al., 2024)

Trends in Genetics



https://www.uab.edu/news/research-innovation/fungal-infection-that-causes-deadly-disease-being-studied-at-uab
https://www.first-nature.com/fungi/boletus-edulis.php
https://www.mykoweb.com/CAF/species/Pleurotus_ostreatus.html

Searched 1,047 public genomes for captain genes

STARFISH

CHTC

Found 1,201 putative captains in 431 genomes from 249 species

https://www.inaturalist.org/taxa/47378-Trametes
https://www.mykoweb.com/CAF/species/Sphaerobolus_stellatus.html

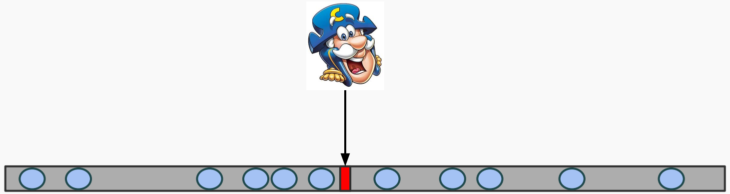
Found 1,201 putative captains in 431 genomes from 249 species

https://www.inaturalist.org/taxa/47378-Trametes
https://www.mykoweb.com/CAF/species/Sphaerobolus_stellatus.html

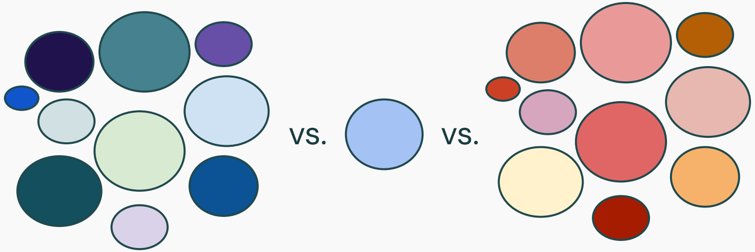
But, existing *Starship* detection tools require more genomic data than is available for these mushrooms.

So, we had to develop a new tool called deltabit.

deltabit



Take all genes in genomic region of interest

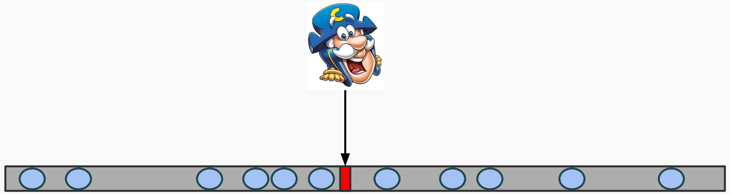


Compare each gene to a set of ingroup and outgroup genes

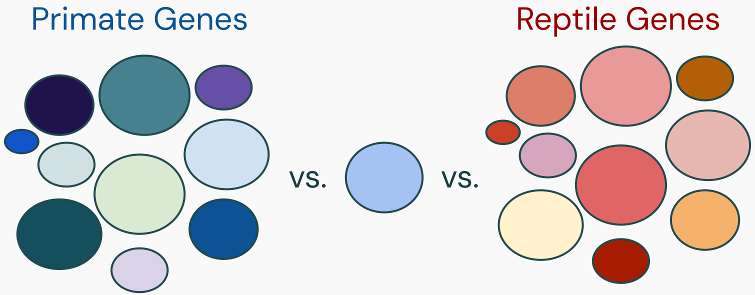
best ingroup score – best outgroup score = deltabit score

deltabit

Example:
Are genes horizontally transferring between humans and lizards?



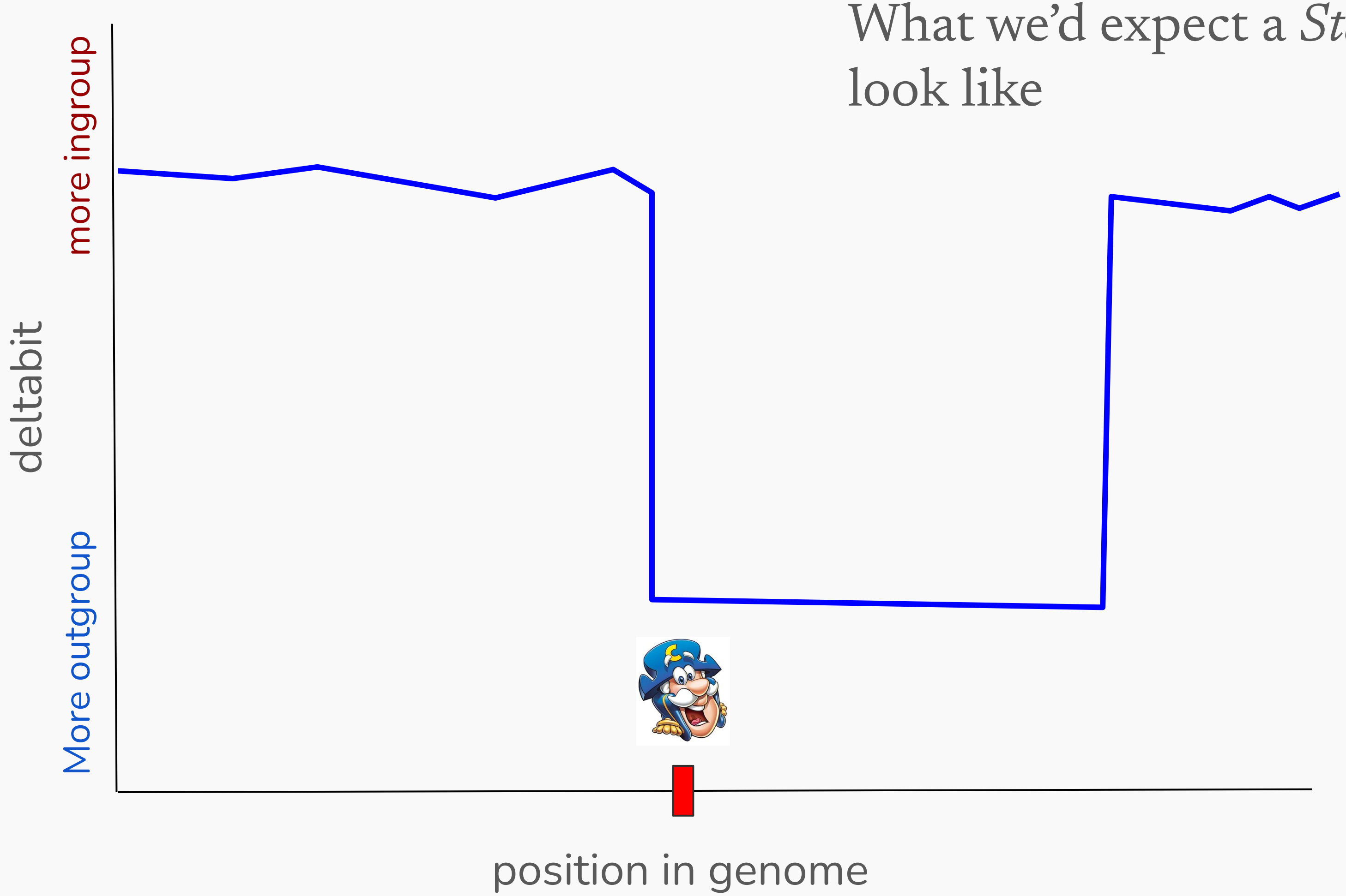
Take all genes in genomic region of interest



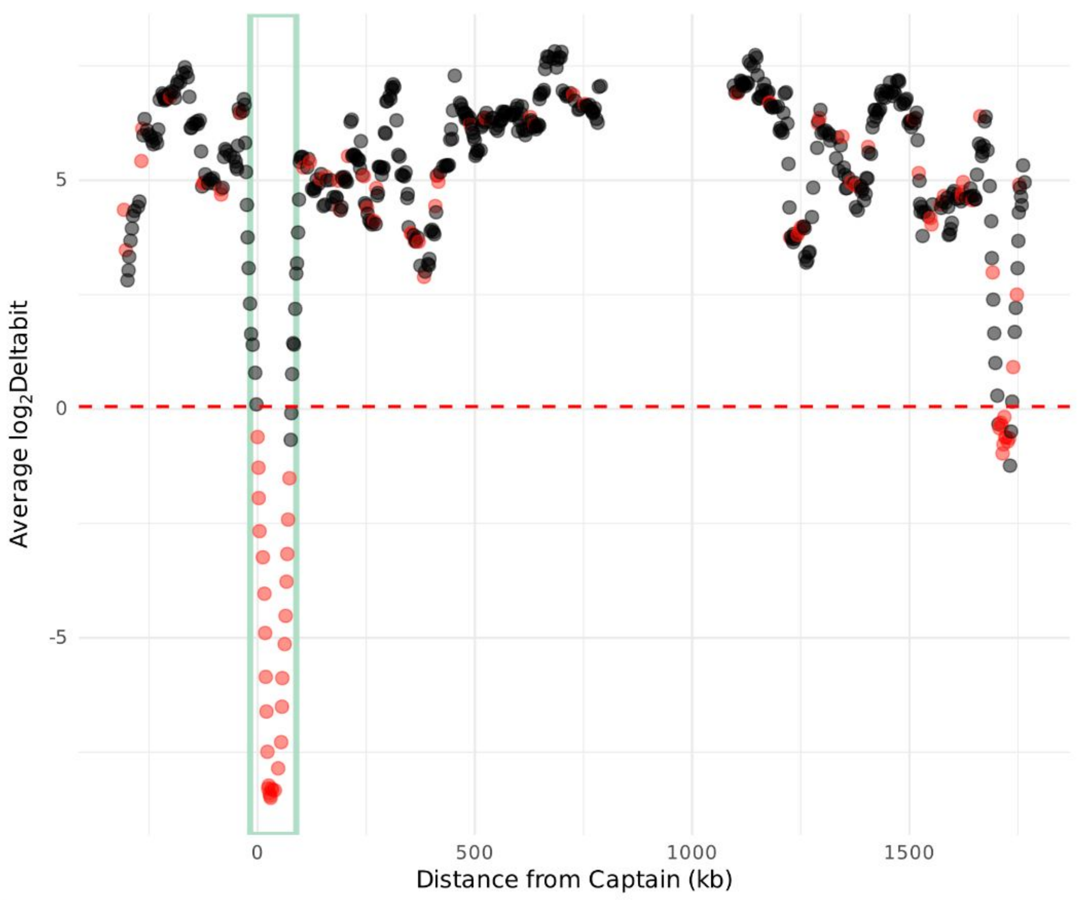
Compare each gene to a set of ingroup and outgroup genes

best ingroup score – best outgroup score = deltabit score

What we'd expect a *Starship* to look like



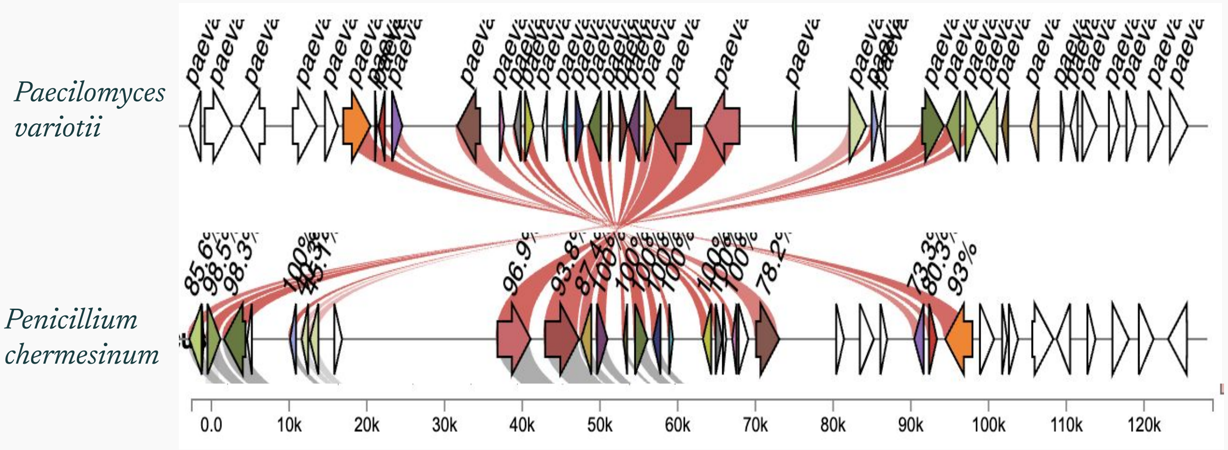
deltabit

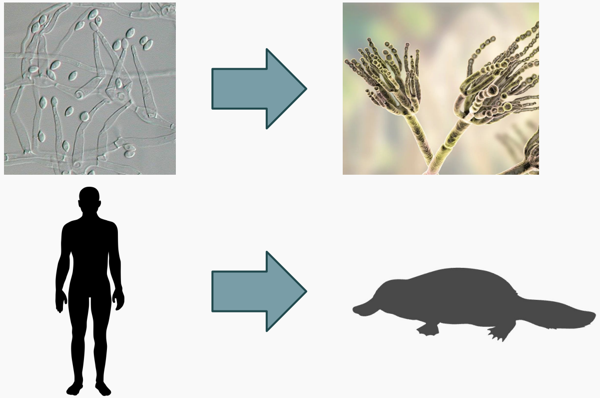


We can then plot all the genes in the region and their deltabit scores. *Starship*-like features will show up as clustered peaks.

This shows *Hephaestus*, a well-studied *Starship* in the fungus *Paecilomyces variotii*, which has been horizontally transferred to/from *Penicillium*.

We can then look for clusters of similar genes in other genomes and find other copies of the *Starship* that may have been horizontally transferred.





>180 million years of evolution between these species.
Similar distance between platypus and human!

Strengths of deltabit compared to other *Starship* detection methods

https://github.com/cbendett/deltabit

- Works for species with limited high-quality genomes
- Doesn’t rely on comparisons to previously identified *Starships*
- Leverages the availability of of public genomes of varying quality
- Directly finds evidence of horizontal transfer (and can identify potential recipients/donors)
- Can detect other *Starship*-like features that horizontally transfer

Strengths of deltabit compared to other *Starship* detection methods

https://github.com/cbendett/deltabit

- Works for species with limited high-quality genomes
- Doesn’t rely on comparisons to previously identified *Starships*
- Leverages the availability of of public genomes of varying quality
- Directly finds evidence of horizontal transfer (and can identify potential recipients/donors)
- Can detect other *Starship*-like features that horizontally transfer

~130 billion pairwise comparisons per captain, ~150 trillion in total

Bioinformatics on the CHTC

- Lots of pipelining small tools; not always the easiest to manage
- Troubleshooting memory/disk space can be time consuming (unpredictable outputs)
- Great support
- Data sharing: Fungal Supergroup Database (FSDB)

Acknowledgements

Advisor

Dr. Emile Gluck-Thaler

Lab Members

Dr. Samuel O'Donnell
Annika Pratt
Noah Brown
Kritika Chopra
Dr. Justin Eagan

CHTC

HTC25

DEPARTMENT OF PLANT PATHOLOGY
COLLEGE OF AGRICULTURAL & LIFE SCIENCES