

Getting a foothold in PAN-GENOMICS with Triffids

By **Chelsea Matthews**

Supervisors: Prof. Rachel Burton and Dr. Nathan Watson-Haigh

What is a Pan-genome?

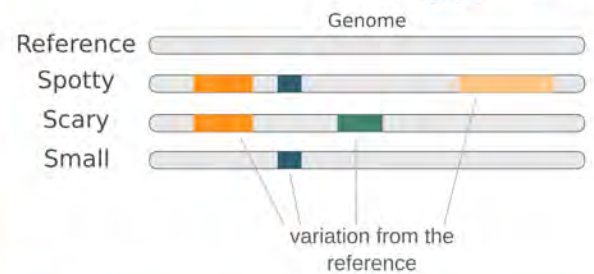
an organised collection of genomic data from multiple members of a single species

What is a Triffid?

a fictional predatory plant capable of locomotion that hunts and kills humans
(see *The Day of the Triffids* by John Wyndham for more information)

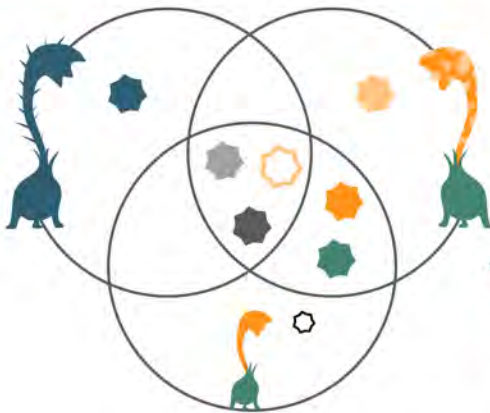
Collection Pan-genome

a collection of genomic sequence from multiple members of the Triffid species



Presence Absence Variation Pan-genome ▲

focus just on gene content
identifies the core genome (genes present in all triffids) and the accessory genome (genes present in only some triffids)



Core Genes



conserved set of genes generally necessary for survival

Accessory Genes



variable set of genes responsible for variations in lifestyle and evolutionary trajectories

Motivation

There are **4 different types of pan-genomic models**. Here we give them names and define them so that when we talk about a particular pan-genome, everyone else knows what we're talking about too.

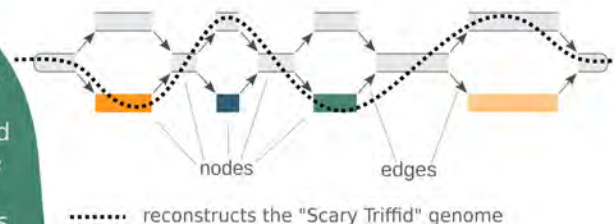
Representative Sequence Pan-genome ◆◆★

a curated collection of genomic sequence with variation represented only once



Graphical Pan-genome ◆◆▲★

breaks genomic sequence into blocks ("nodes") with "edges" indicating how the blocks were connected in the input sequences



Applications

- ◆ reduce reference bias
- improve variant calling accuracy
- ▲ link phenotype with genotype
- ★ alternative to linear reference

Acknowledgements

Thank you to my supervisors for their ongoing support and to the Burton Lab for being so welcoming. Also my partner for trying really hard to be interested.