

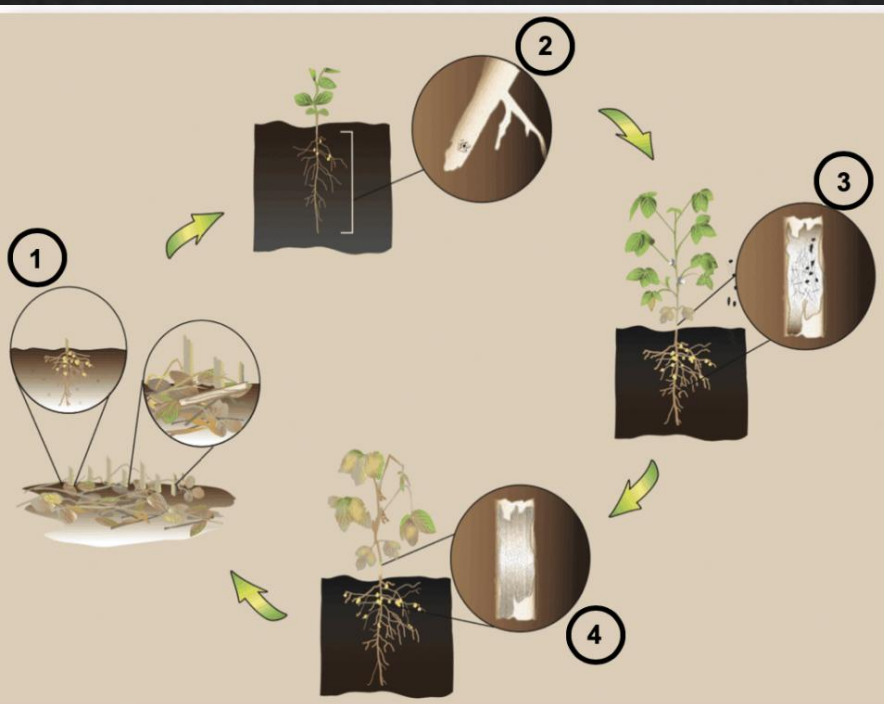
High Throughput Computing for Fungal Genomics

17 July 2026

Annika Pratt

About Me

- ◆ Bachelor of Arts in Biology and Computer Science
Concordia College in Moorhead, MN
- ◆ OSG School in 2024
- ◆ 4th year PhD student in Plant Pathology at UW-Madison



Macrophomina phaseolina

Fungal plant pathogen

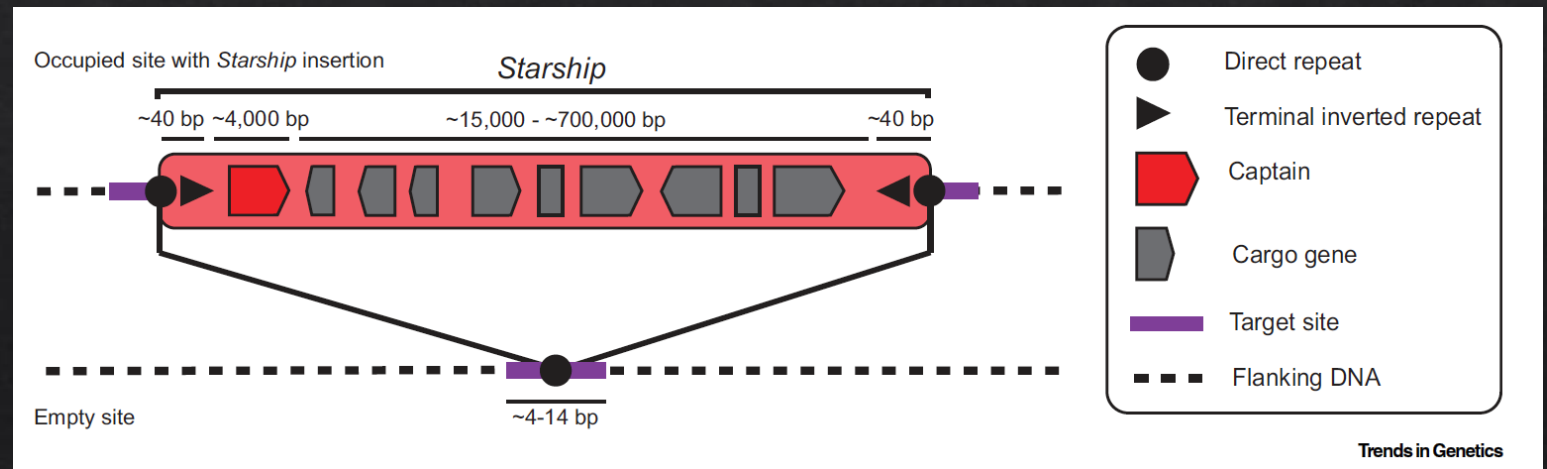
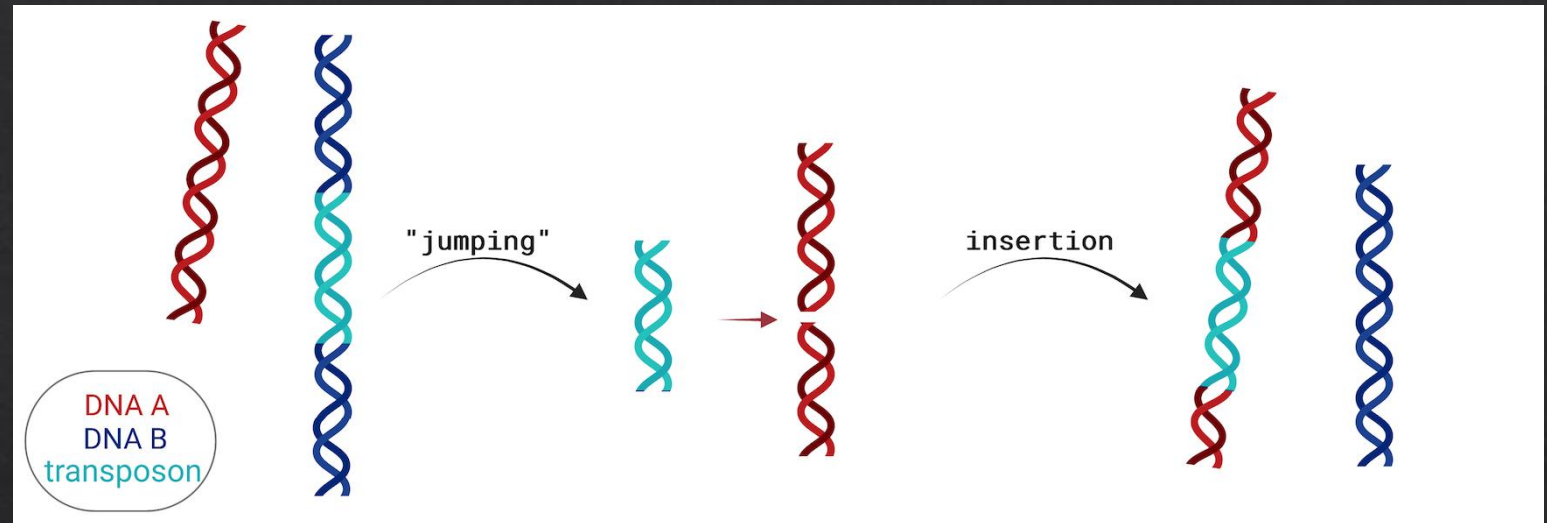
Infects 100 economically important crops

\$660 million in damages in a single season on soybean

We don't know much about *how* it causes disease

Starships

- ❖ Large transposable elements in filamentous fungi
- ❖ Can carry 'cargo' genes between different fungi

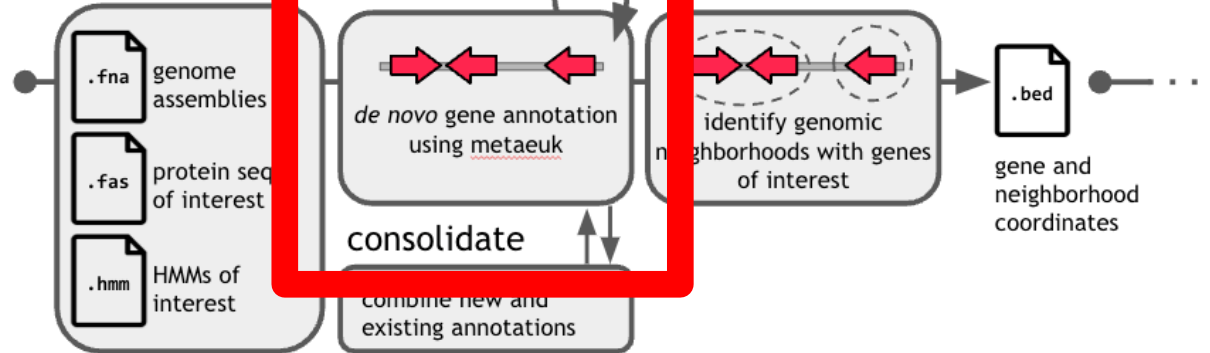


Do *Starships* in *Macrophomina phaseolina* contribute to plant disease?

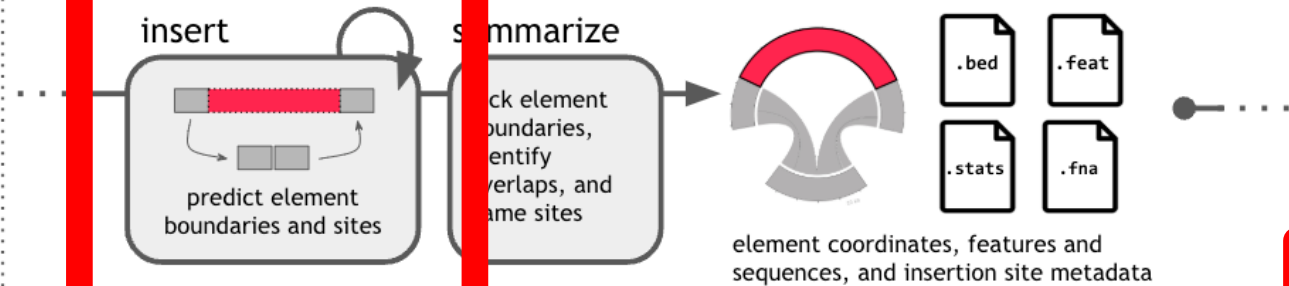
- ◇ 490 genomes
 - ◇ Genome Annotation
 - ◇ EarlGrey ~4 hours per genome
 - ◇ Funannotate ~10-14 hours per genome
 - ◇ InterPro ~8 hours per genome
 - ◇ Plus a few other programs that take several minutes each
 - ◇ *Starship* annotations

Main workflow

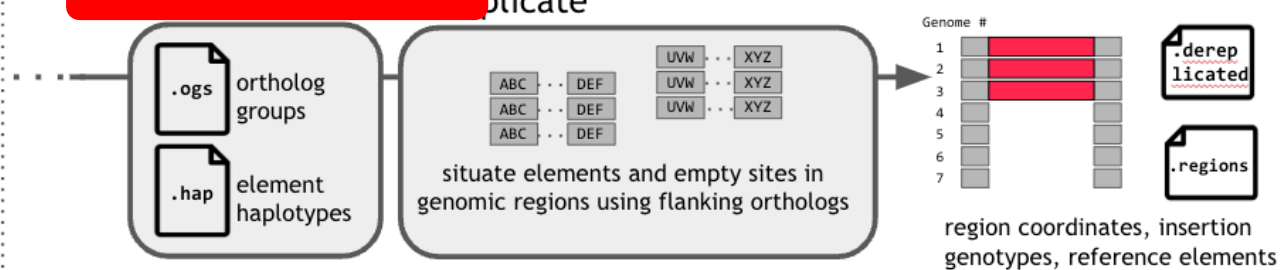
Gene finder:



Element finder:



Region finder:



STARFISH



auxiliary commands

sim	calculate k-mer similarity between captains or elements
group	sort captains or elements into homologous groups using mcl
extend	extend downstream element boundary using BLASTn alignments
augment	<i>de novo</i> gene annotation with metaeuk and a profile database
cargo	define element pangenomes and calculate cargo-based similarities between elements
format	format fasta and gff files with genomIDs
format-ncbi	format ncbi assemblies and assign genomIDs
dereplicate-hood	situate neighborhoods in genomic regions using flanking orthologs

visualization commands

pair-viz	execute element- element/ site/region alignments, with optional circo
locus-viz	execute element/region alignments with nucmer and generate synteny schematics with <u>gggenomes</u>
genome-viz	visualize elements and sites in a genome with circo

Deathstar

- ◇ Background
 - ◇ *Macrophomina phaseolina* can infect a lot of plants, but each individual strain might only infect certain plants
 - ◇ Most strawberry-infecting strains only infect strawberry
- ◇ Collaborators Jade Dilla-Ermita and Peter Henry found
 - ◇ 200 kb region of the *Macrophomina phaseolina* (typical *Starship* size)
 - ◇ 12 genes upregulated during strawberry infection (potentially involved in disease)
 - ◇ Absent in strains that don't infect strawberry (indicative of transposable element)

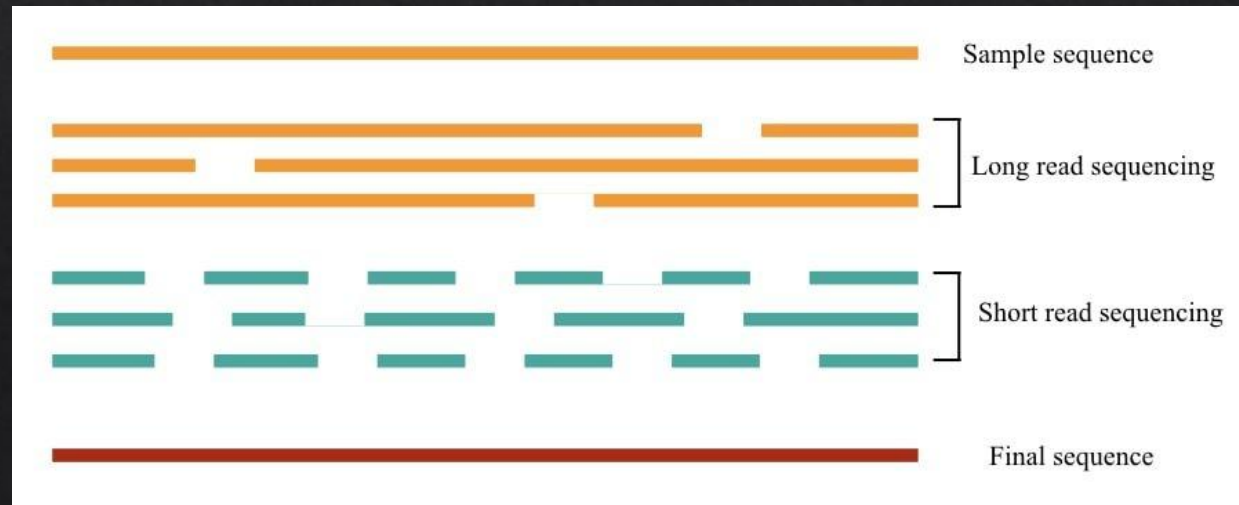
Deathstar

- ◇ Collaborators Jade Dilla-Ermita and Peter Henry found
 - ◇ 200 kb region of the *Macrophomina phaseolina* (typical *Starship* size)
 - ◇ 12 genes upregulated during strawberry infection (potentially involved in disease)
 - ◇ Absent in strains that don't infect strawberry (indicative of transposable element)

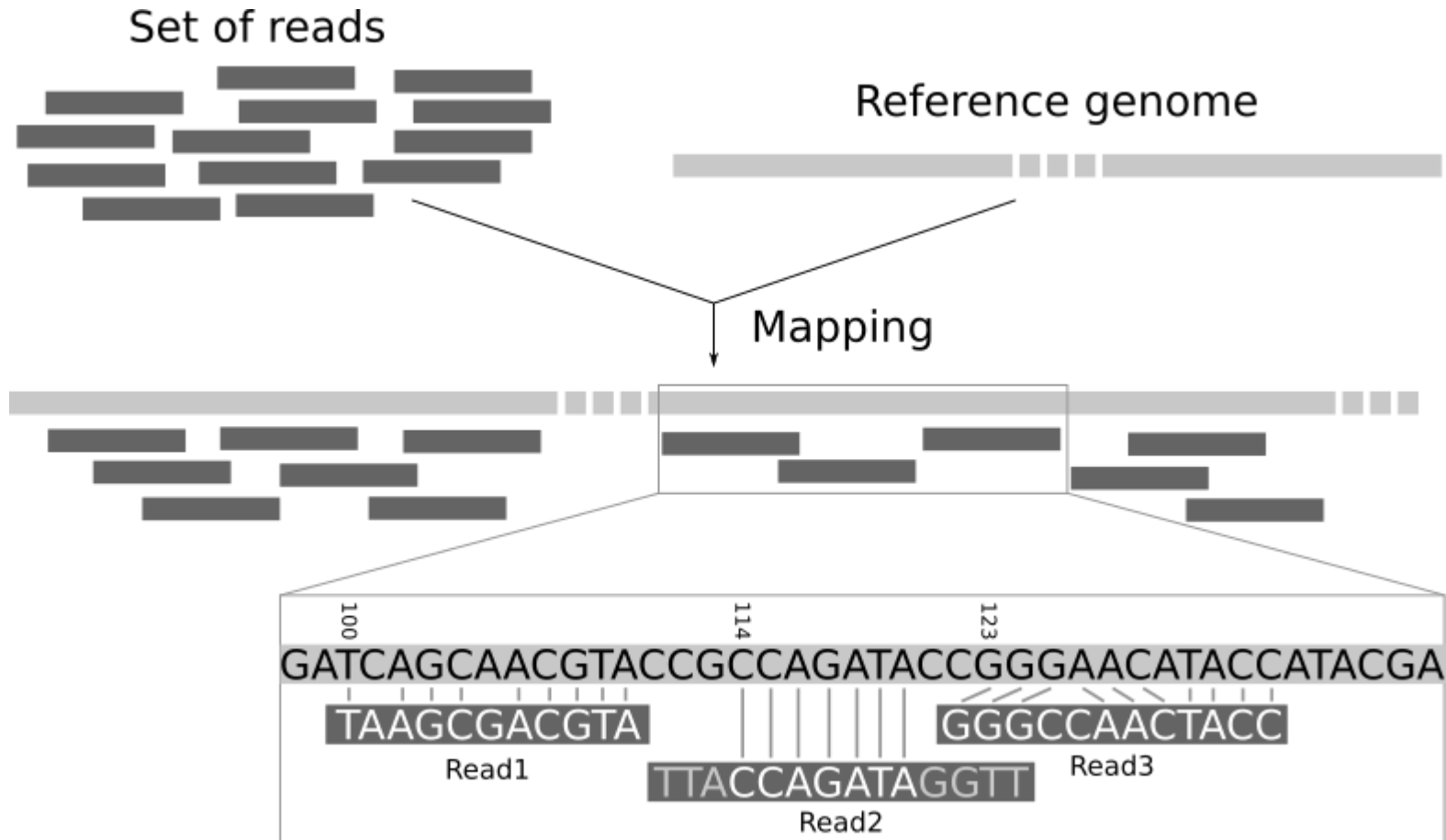
It's a *Starship*!

Is there a strong association between *Deathstar* and strawberry disease?

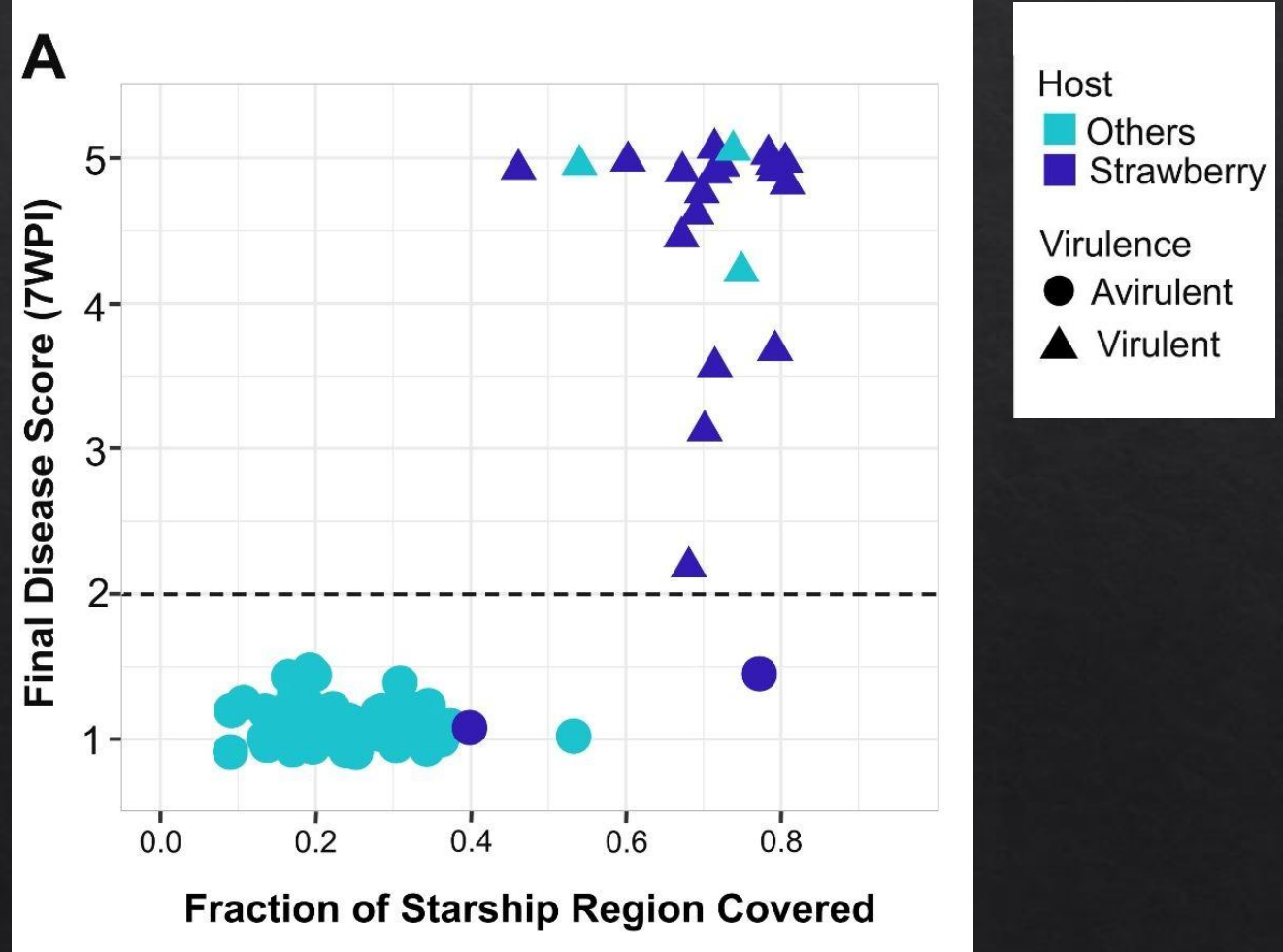
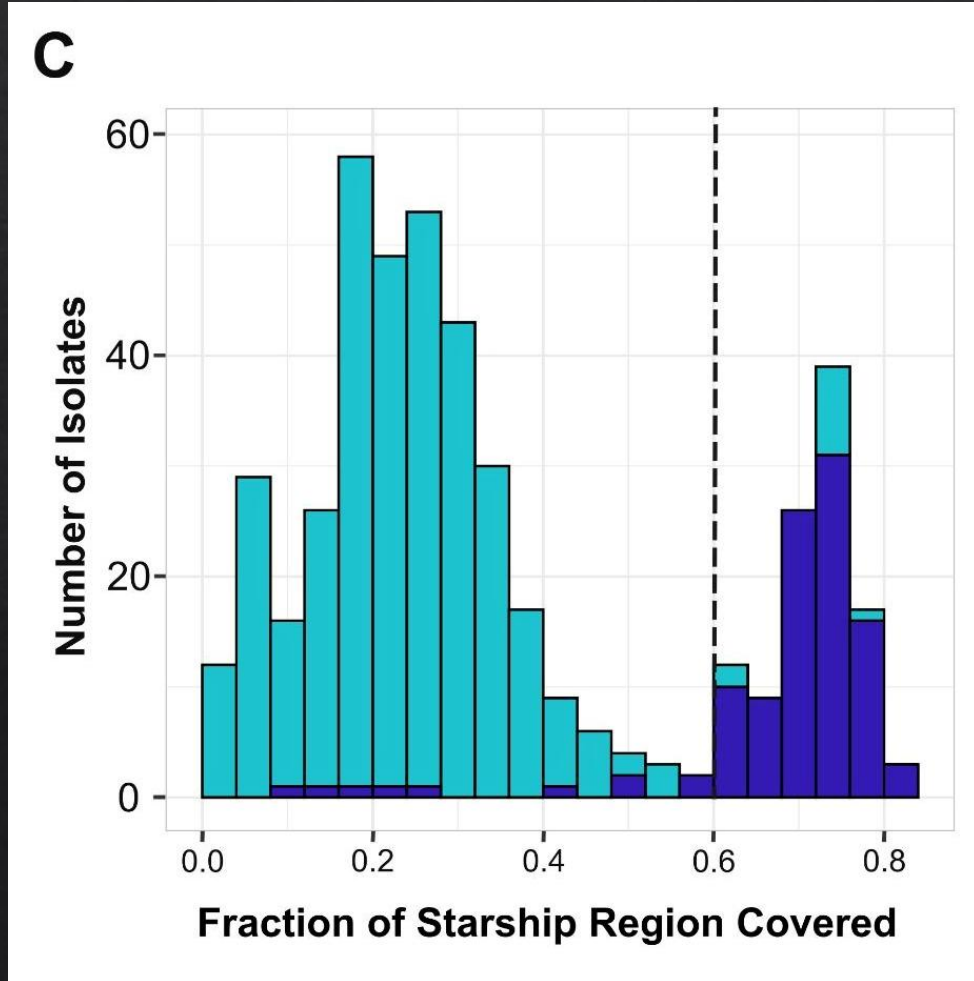
- ◇ Back to the 490 genomes
- ◇ Background
 - ◇ Long read sequencing – relatively expensive, easy to detect *Starships* with starfish
 - ◇ Short read sequencing – relatively inexpensive, difficult to detect *Starships* with starfish
 - ◇ Only 28 of the 490 genomes are long-read, so we can't accurately detect *Deathstar* using starfish



Alternative to starfish: read mapping



Deathstar is associated with strawberry infection



In Conclusion



**My lab uses HTC to run
bioinformatic tools on
many genomes
simultaneously**



**Biggest lesson
learned/piece of
advice:**

If in doubt, ask the facilitators!

Acknowledgements

- ◆ CHTC, CHTC Facilitators, OSG School
- ◆ Funding
 - ◆ Wisconsin Soybean Marketing Board
 - ◆ UW-Madison Department of Plant Pathology
 - ◆ NSF
 - ◆ NIH
- ◆ PIs: Emile Gluck-Thaler and Mehdi Kabbage
- ◆ Lab members (all of them! Too many to list)



OSG

CHTC

Questions?