



Characterization of bacterial endosymbionts and systems biology resources for *Diaphorina citri*, insect vector for the citrus greening disease

Stephanie Hoyt^{1,2}, Prashant S Hosmani¹, Mirella Flores-Gonzalez¹, Wayne Hunter³, Susan Brown⁴, Lukas A. Mueller¹ and Surya Saha¹

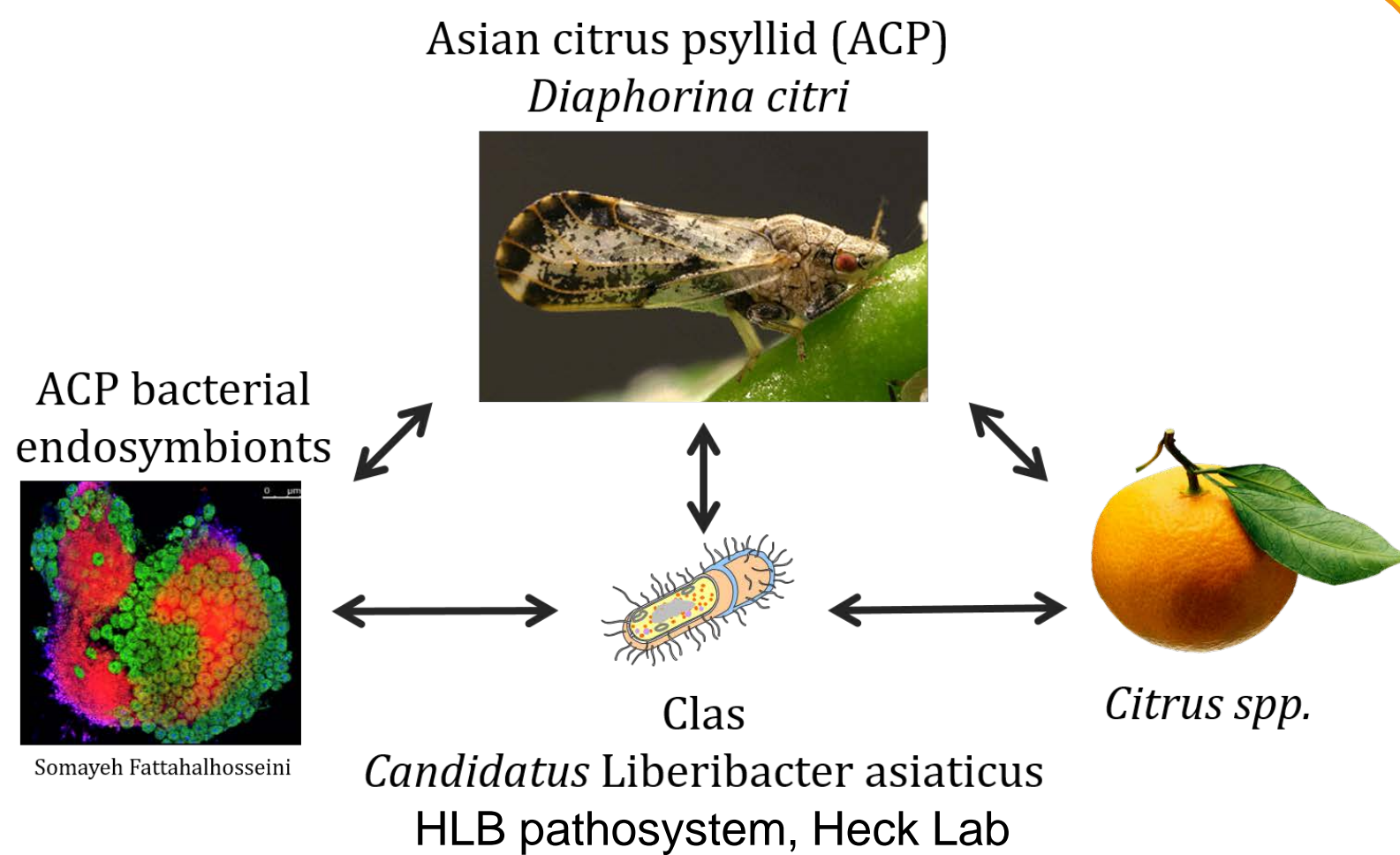
(1) Boyce Thompson Institute, Ithaca, NY, (2) Cornell University, Ithaca, NY, (3) USDA-ARS, U.S. Horticultural Research Laboratory, Fort Pierce, FL, (4) Division of Biology, Kansas State University, Manhattan, KS.

Introduction

The Asian citrus psyllid (*Diaphorina citri* Kuwayama) is the insect vector of the bacterium *Ca. Liberibacter asiaticus* (CLAs), the causal agent for citrus greening or Huanglongbing disease which threatens the citrus industry worldwide. There is evidence that CLAs interacts with the known endosymbionts of *D. citri*: *Wolbachia*, *Ca. Profftella armatura*, and *Ca. Carsonella ruddii*.

Endosymbiont sequences were extracted from the new *D. citri* genome assembly and verified for coverage against Pacbio and Illumina data, as well as against published reference genomes. Scaffolds from SSPACE were annotated with prokka and analyzed using Orthofinder and BRIG for gene copy number and conserved synteny.

We have designed a web portal with genomics, proteomics, transcriptomics and bioinformatics resources for the vector *D. citri* including endosymbionts, the host citrus (*C. clementina* and *C. sinensis*) and multiple pathogens including CLAs for the citrus greening disease.



Workflow

- The previous version of the *D. citri* genome, v1.1, was fragmented.
 - We sequenced the 450Mb psyllid genome to 80X coverage with Pacbio long reads.
 - Canu assembly followed by Dovetail Chicago-based scaffolding was used to create an improved assembly (v2.0) with a contig N50 of 758.7kb and 1906 contigs.
- Long range Hi-C scaffolding was performed by Dovetail to get 13 chromosomes (Diaci v3.0) with 442Mb with N50 of 40,584Kb. Unplaced scaffolds will be assigned to Chr00 pseudomolecule
- Endosymbiont sequences were extracted from the new *D. citri* genome assembly and verified for coverage using the Integrative Genomics Viewer (IGV).
 - All contigs were compared to Pacbio and Illumina data, as well as published reference genomes of the endosymbionts.
 - Any contigs that looked to have been mis-assembled were split into multiple contigs and regions that had no support were removed.
 - The verified contigs were then scaffolded using SSPACE.
- These assemblies were then annotated using prokka, and the annotations were used for orthology analysis with Orthofinder.

Genome	Diaci 1.1	Diaci 2.0	Diaci 3.0
Contigs	161,988	1,906	1,271
Total Length	485 Mb	498 Mb	475 Mb
Longest	1 Mb	4.2 Mb	50 Mb
Shortest	201 bp	6 Kb	1 Kb
Ns	19.3 Mb	4.5 Mb	12 Mb
Contig N50	34,407 bp	758.7 Kb	40 Mb

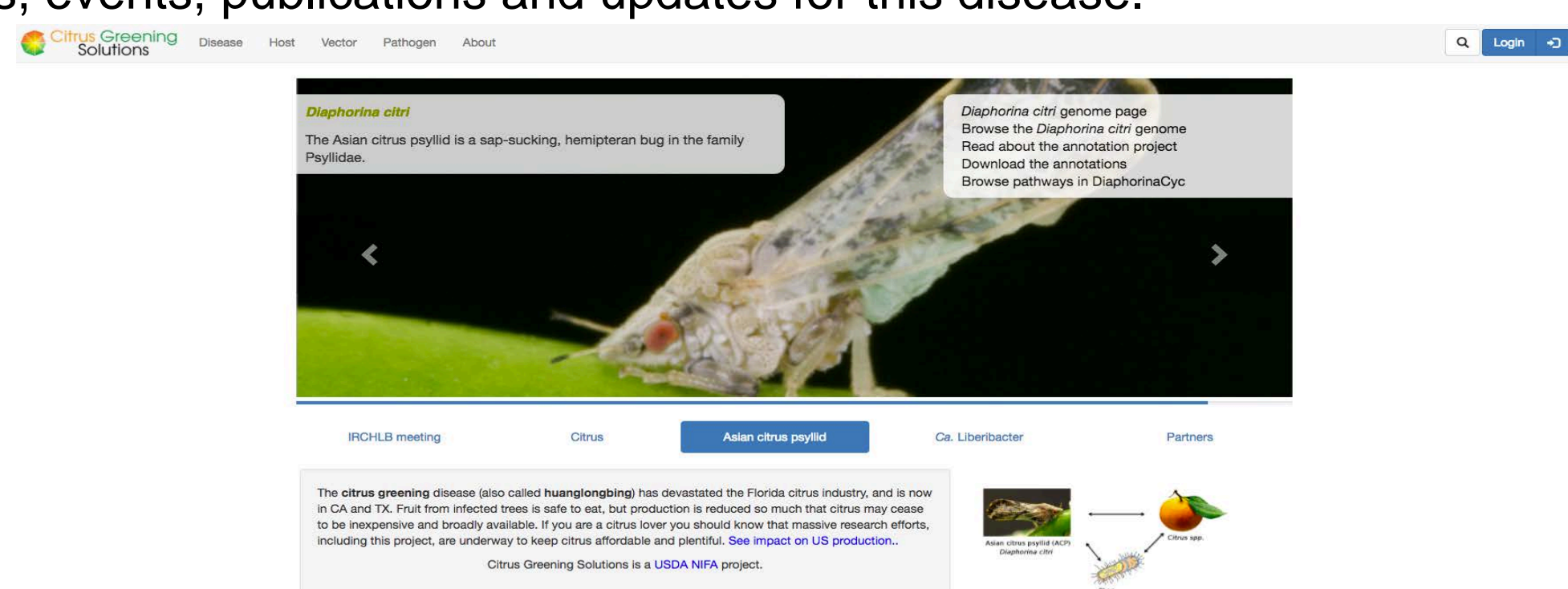
Table 1: Comparison between *D. citri* genome versions

	Wolbachia	Profftella	Carsonella
	10 scaffolds	1 chromosome and 1 plasmid	1 chromosome
Largest	923 Kb	471 Kb	-
Smallest	19 Kb	4.7 Kb	-
Total Size	2 Mb	475.7 Kb	150 Kb

Table 2: Assembly statistics for endosymbiont genomes

CitrusGreening.org

Citrusgreening.org is a web portal with genomics and bioinformatics resources focused on *Diaphorina citri*, *Citrus spp.* and *Candidatus Liberibacter asiaticus*. The portal has news, events, publications and updates for this disease.



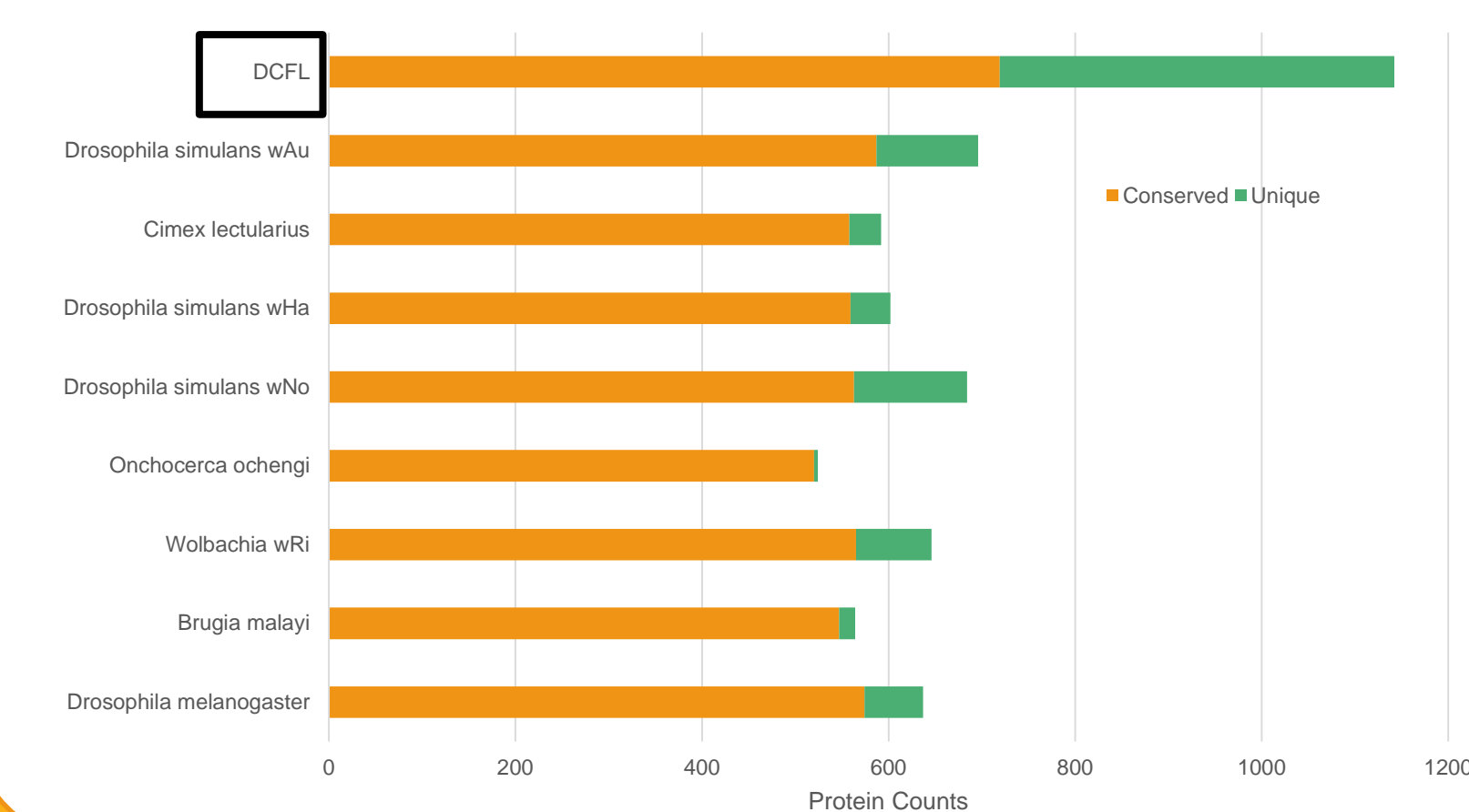
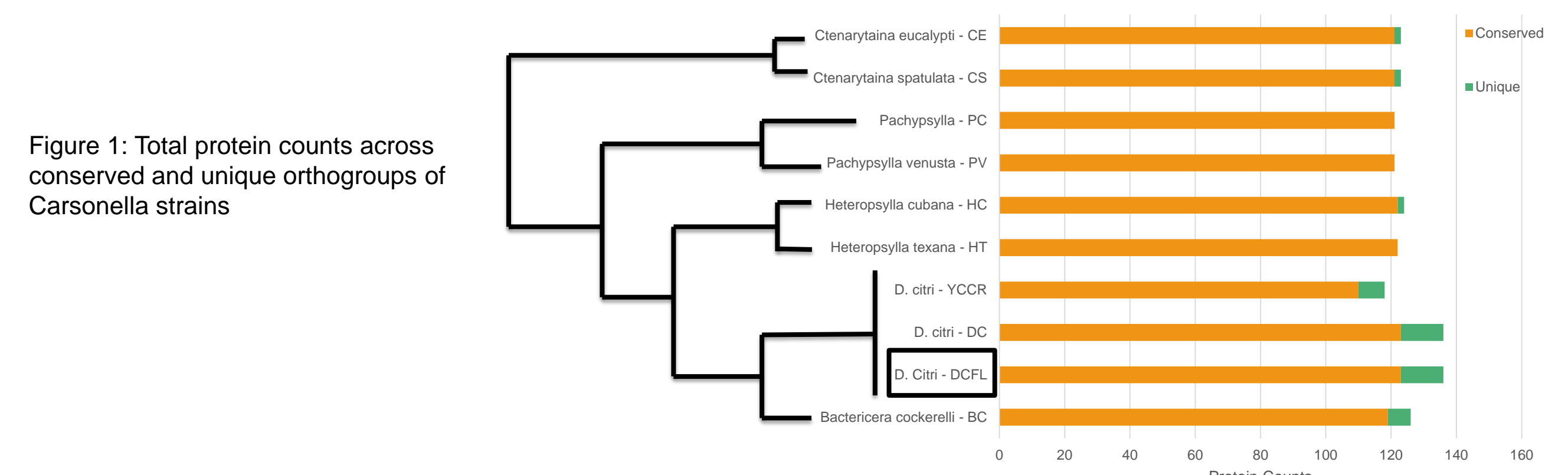
- News
- Events
- Publications
- Annotation updates
- Links
- FTP download

Orthology Analysis

- We used Orthofinder as a way to judge the completeness of our assemblies
 - We compared the orthogroups present in our assemblies to those conserved among reference genomes for each endosymbiont
 - We defined a conserved orthogroup for a given endosymbiont to be one that contains at least one protein from each published genome
 - We defined a unique orthogroup for a given endosymbiont to be one that is present in our assembly, and is not present in at least 50% of the reference genomes
- General Workflow:
 - We used Orthofinder for analysis of references in order to determine median copy number of proteins within each orthogroup
 - We analyzed published references and assemblies together to see which orthogroups were represented by our assemblies

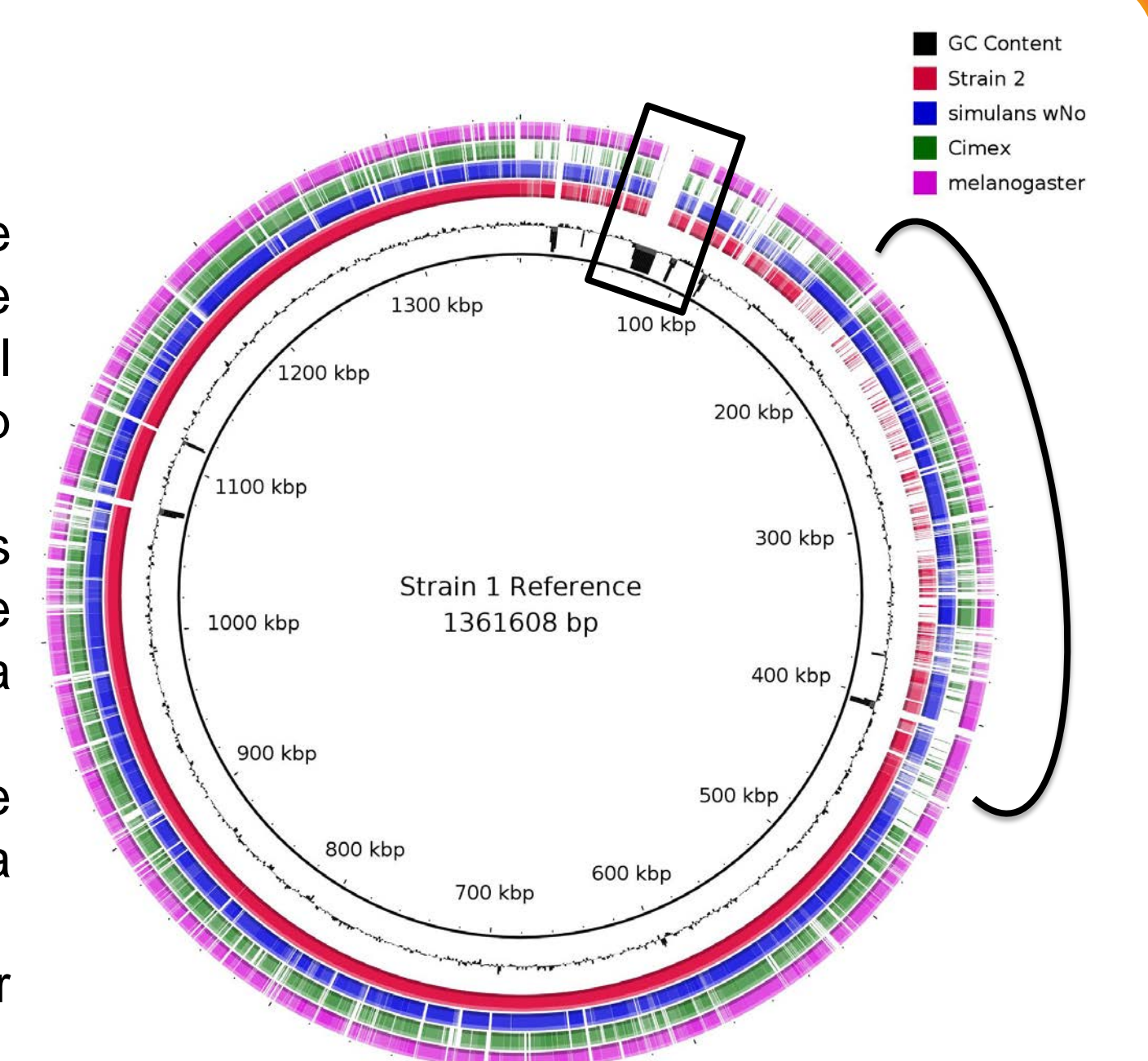
	Wolbachia	Profftella	Carsonella
Number of reference genomes	8	2	9
Total number of conserved orthogroups	559	307	116
Number of conserved orthogroups in our assembly	557	307	106
Number of unique orthogroups	167	-	12

Table 3: Orthology analysis of endosymbiont genomes



Wolbachia Strains

- Scaffolds were removed from the Wolbachia assembly resulting in a large decrease in duplication, but a small decrease in conserved orthogroup coverage.
 - This left the 3 longest scaffolds which together cover the reference proteome defined by Wolbachia genomes
- Based on these results we hypothesize that there are two strains of Wolbachia present in this sample:
 - Strain 1: Scaffolds 1 and 2 cover 534/559 conserved orthogroups
 - Strain 2: Scaffolds 1 and 3 cover 503/559 conserved orthogroups

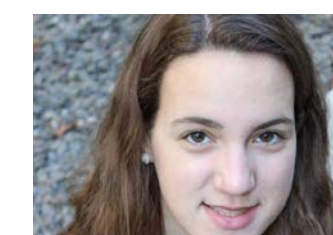


Acknowledgments

Project Partners



Funding provided by



Stephanie Hoyt
shh89@cornell.edu



citrusgreening.org

