Structural Variation of Diploid & Polyploid Plant Genomes

JOSHUA A. UDALL[1], THIRUVARANGAN RAMARAJ[2], LOUISA STARK[3]

1 BRIGHAM YOUNG UNIVERSITY, 2 NATIONAL CENTER OF GENOMIC RESOURCES, 3 UNIVERSITY OF UTAH

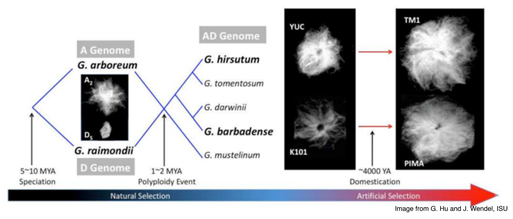
POSTER BY: L. CHANEY, A. SHARP, & D. YUAN

NSF Plant Genome Research Program Award #1339412

Project Information

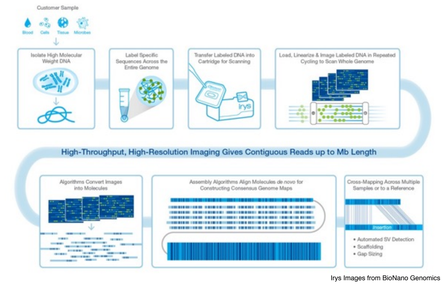
Cotton: a model for understanding genomve evolution in polyploids

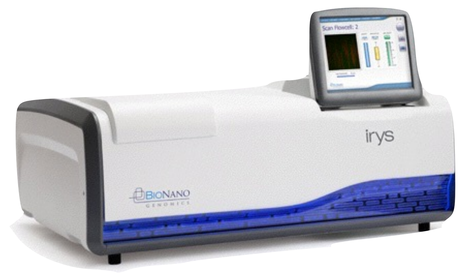
- The Gossypium genus includes >50 species, including cultivated cotton
- A single polyploidization event occurred 1-2mya with the A- and D- diploid genomes, after 5–10 million years of isolated diploid evolution
- There are currently 7 extant allopolyploid species — 2 of which are cultivated and provide >95% of the worlds cotton production
- The evolutionary history and phylogenetic framework provides a unique, monophyletic platform to examine genomic structural variations
- Structural rearrangements can contribute to speciation and phenotypic variability and are traditionally identified through time extensive means



Irys technology (BioNano) will be used to create optical maps

- Extremely long (high-molecular weight) DNA is labeled by incorporation of fluorophore-labeled nucleotides at a specific 7-bp sequence motif throughout the genome.
- Labeled genomic DNA is then linearized in nanochannels and single molecules are imaged
- Labels give a pattern of distances that is specific to their genomic loci. These patterns are used to determine molecule overlap.
- Molecules are assembled into genome maps
- Strategies to compare genome optical maps between multiple species within a phylogenetic framework will be developed





The objective of this project is to investigate structural variations in diploid and tetraploid cotton species by creating genome optical maps.

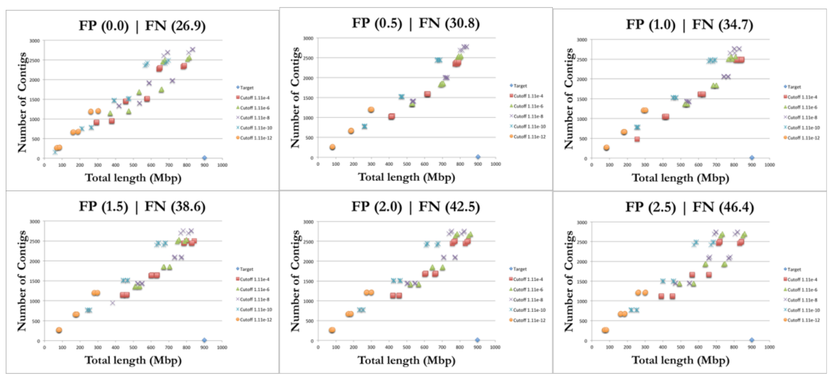
Assembling Genome Optical Maps

OMWare - Software for Optical Maps

- A python module to help scientists use optical map data
- Optical map assemblies are highly sensitive to input parameters
- Runs multiple assemblies with permutations of input parameters
- Optimized to re-use compatible results as much as possible, in order to minimize resource usage
- Acces software via Git Hub: `http://github.com/sharpa/OMWare`

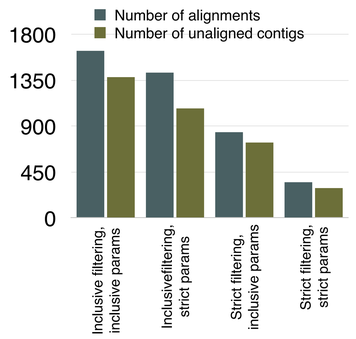
Pairwise parameter search

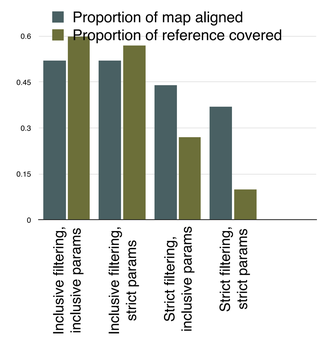
- We generated 315 assemblies with reference genomes available and 395 assemblies without a reference genome



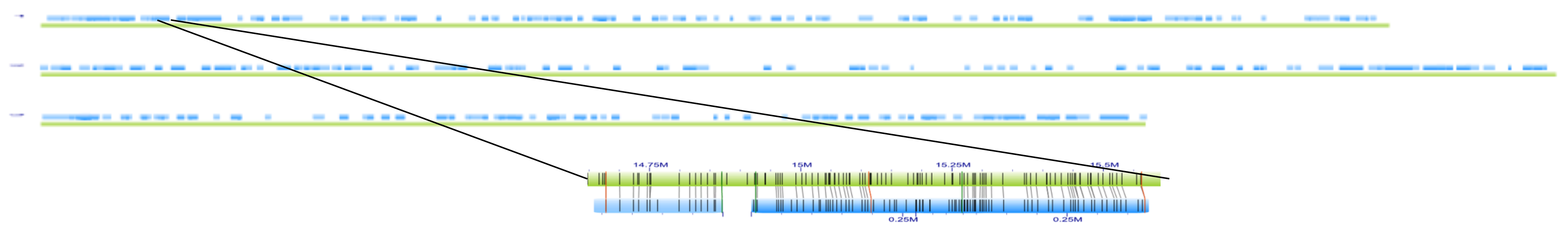
Criteria for best parameters

- We compared Gossypium raimondii map assemblies to a sequence reference genome assembly
- We selected four candidate assemblies, with strict and inclusive data filtering, and inclusive parameters (generated the greatest total length) and strict parameters (maximized length with fewer additional contigs)

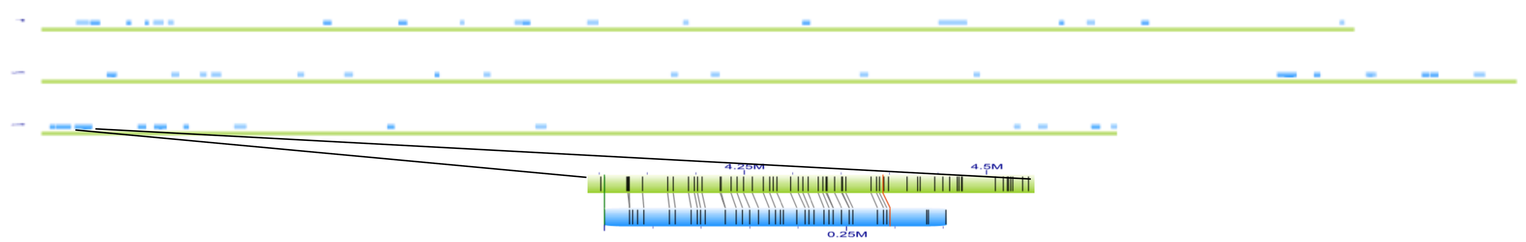




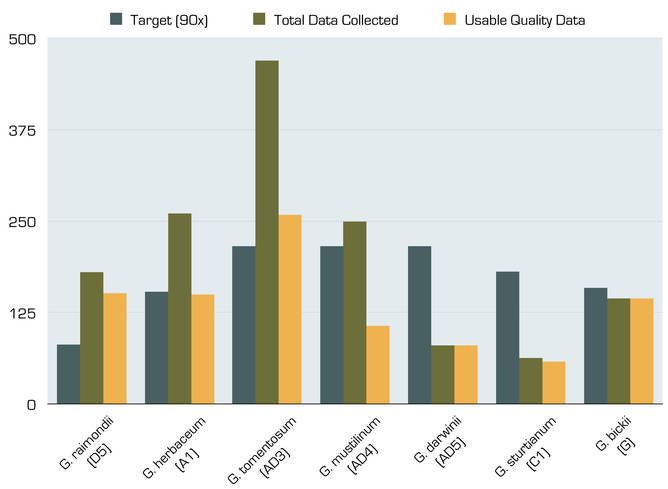
Inclusive data filtering, inclusive parameters



Strict data filtering, strict parameters



Progress of Data Collection



Recent activities

- Over 18 cotton species growing in the green house
- Performed 139 extractions on over 40 taxa
- Run the Irys macine over 50 times
- Collected over 258 Gb of usable data
- >1,000 assemblies run
- 6 parameter searches
- 9+ undergraduate students, 4 graduate students, 2 postdocs have worked on the project
- Over 10 collaborators from different institutions have been involved

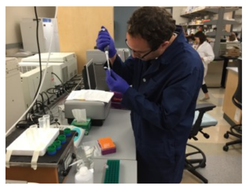
Broader Impacts

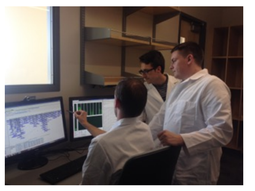
Curriculum development

- Curriculum will be developed with the Genetics Science Learning Center for the Learn.Genetics website
- Initial meetings have taken place to determine content

Research Experience for Teachers (RET)

- Two middle school biology teachers spent 5 weeks with the Udall lab in June — both had an extremely positive experience





"I have an idea on how to teach genetics and evolution better. I believe I also have more interest/passion for the topic, which encourages students" -J Stowers

"This experience expanded my understanding of some of the current real-world research that is taking place in the field of genetics... that I can take back to my students." -M Woolley