

Uncovering the dynamics of genome function at the organismal scale using machine vision and computation



Tessa Durham Brooks, Ph.D.
Doane College
Department of Biology

Anticipation at the dawn of the Genomic Era

“Within the next few years, technologies developed for the Human Genome Project and similar sequencing efforts will revolutionize medicine, agriculture, crimefighting, and other fields.” – Gwynne and Page, Science, 2000



"Today we are learning the language in which God created life."

US President Bill Clinton

Genetronic Replicator – Star Trek: TNG, 1992

The genomic powerhouses

For reference: Humans have about 25,000 genes, 3.2 bil base pairs.



~25,000 genes
100 mil base pairs
Sequenced finished 2000



~20,000 genes
97 mil base pairs
Sequence finished 1998



~22,000 genes
137 mil base pairs
Sequence finished 2000

The problem: at best functional roles have been assigned for 15% of predicted genes of a genome

For reference: Humans have about 25,000 genes, 3.2 bil base pairs.



~25,000 genes
100 mil base pairs
Sequenced finished 2000



~20,000 genes
97 mil base pairs
Sequence finished 1998



~22,000 genes
137 mil base pairs
Sequence finished 2000

Why has determining gene function in multicellular organisms been difficult?



Why has determining gene function in multicellular organisms been difficult?



Our goal: Describe genome function at the organismal scale.

Requirements:

- Observations should be made at sufficiently high spatial and temporal resolution
- Methods should be relatively high-throughput to allow genomic survey
- Observations should be able to be made over time and in many environmental contexts



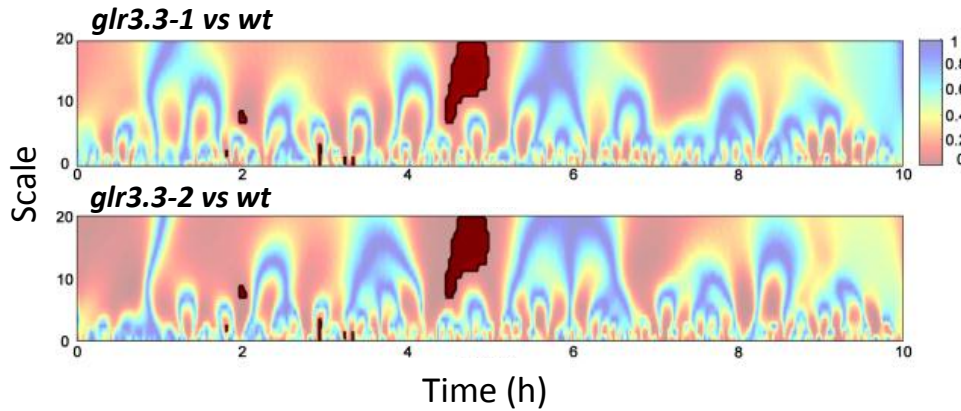
~25,000 genes

Root gravitropism: a model for image analysis approaches in functional genomics

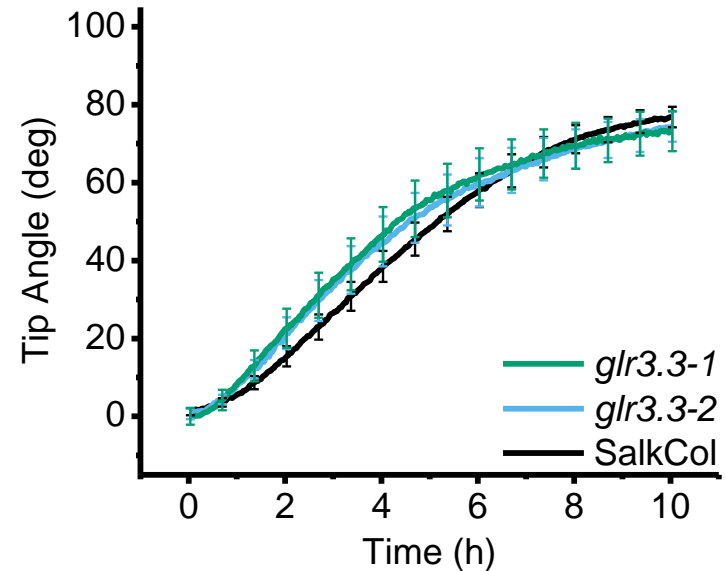
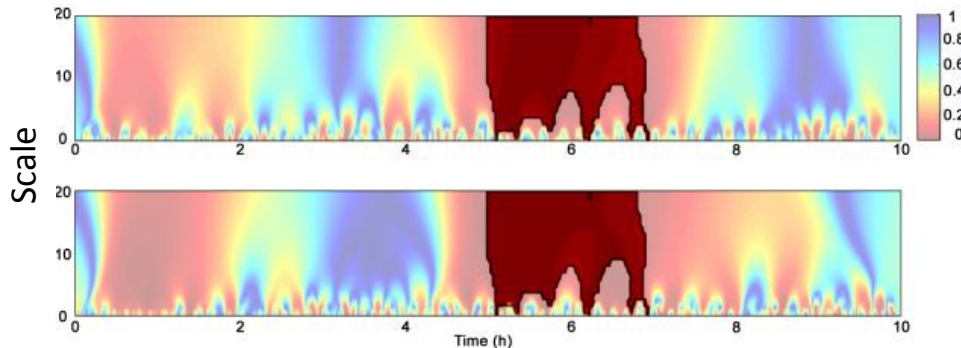


Root gravitropism: a model for image analysis approaches in functional genomics

Second Order (acceleration)



First Order (swing rate)



Developing tools to detect genome function at the organismal scale.

Requirements:

- ✓ Observations should be made at sufficiently high spatial and temporal resolution
- Methods should be relatively high-throughput to allow genomic survey
- Observations should be able to be made over time and in many environmental contexts

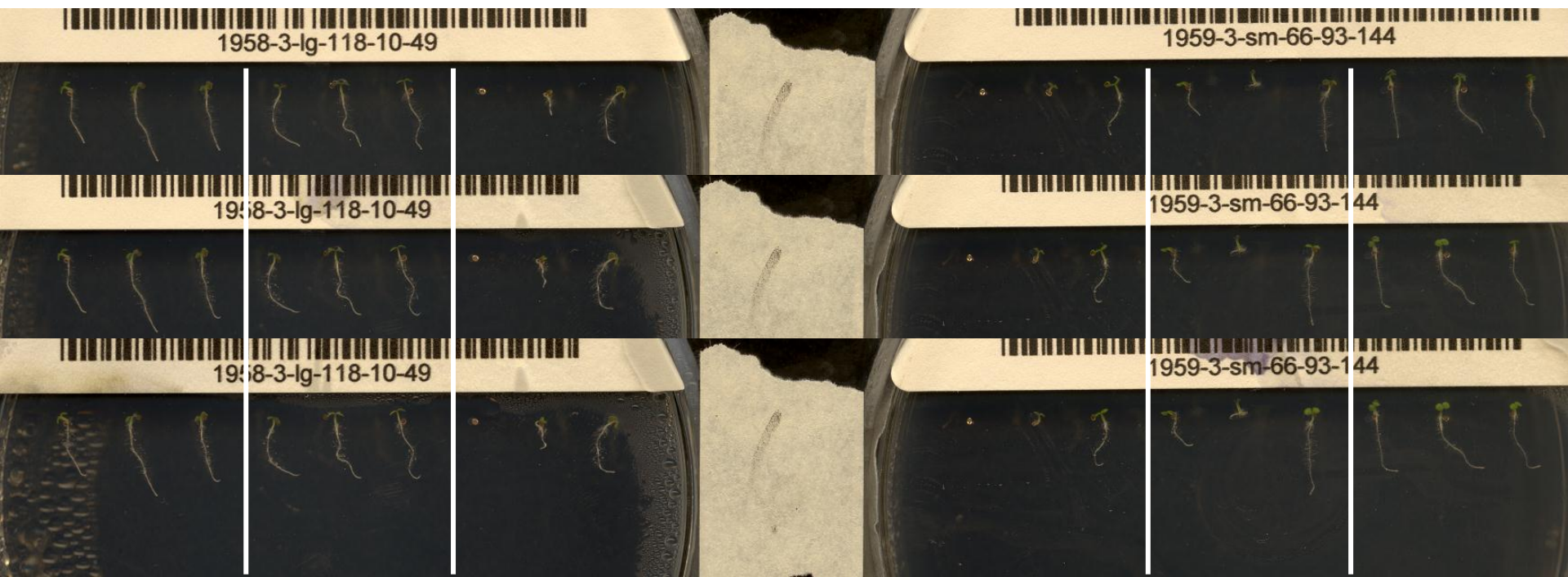


~25,000 genes

Automation and High Throughput ver. 1

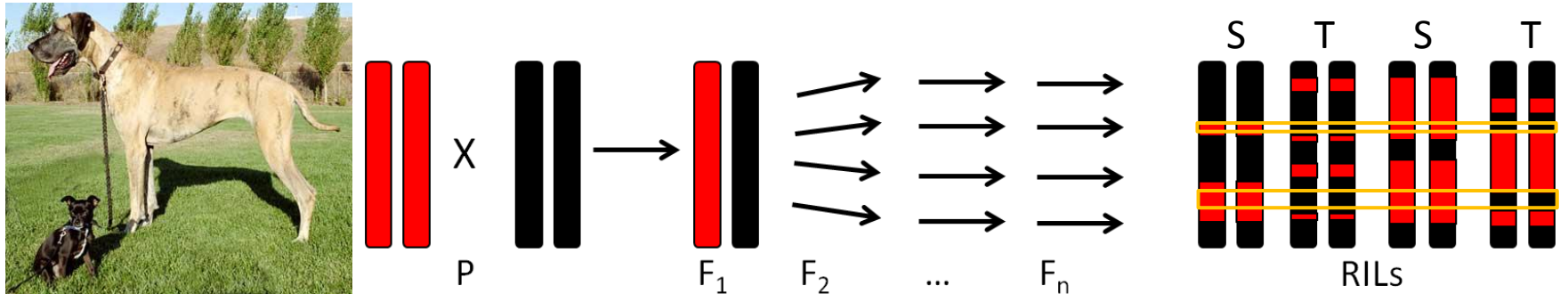


Automation and High Throughput ver. 2



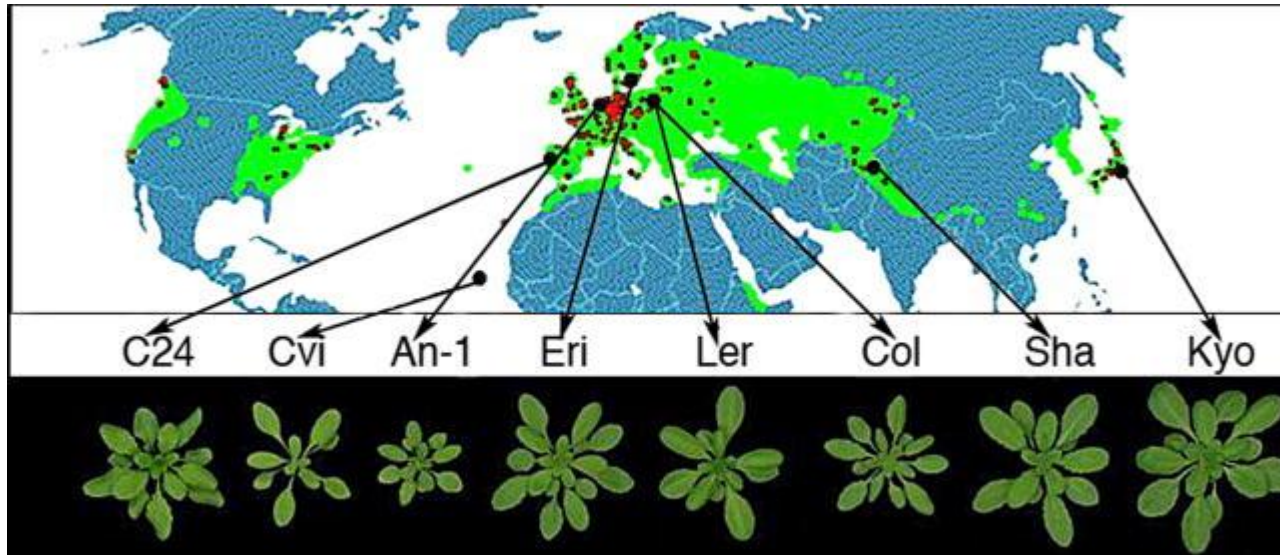
High throughput genetic stocks

- Recombinant inbred lines (RILs)



- Ecotypes (e.g. 1001 genomes project)

QTL Analysis



Matthieu Reymond
Max-Planck Institute

Developing tools to detect genome function at the organismal scale.

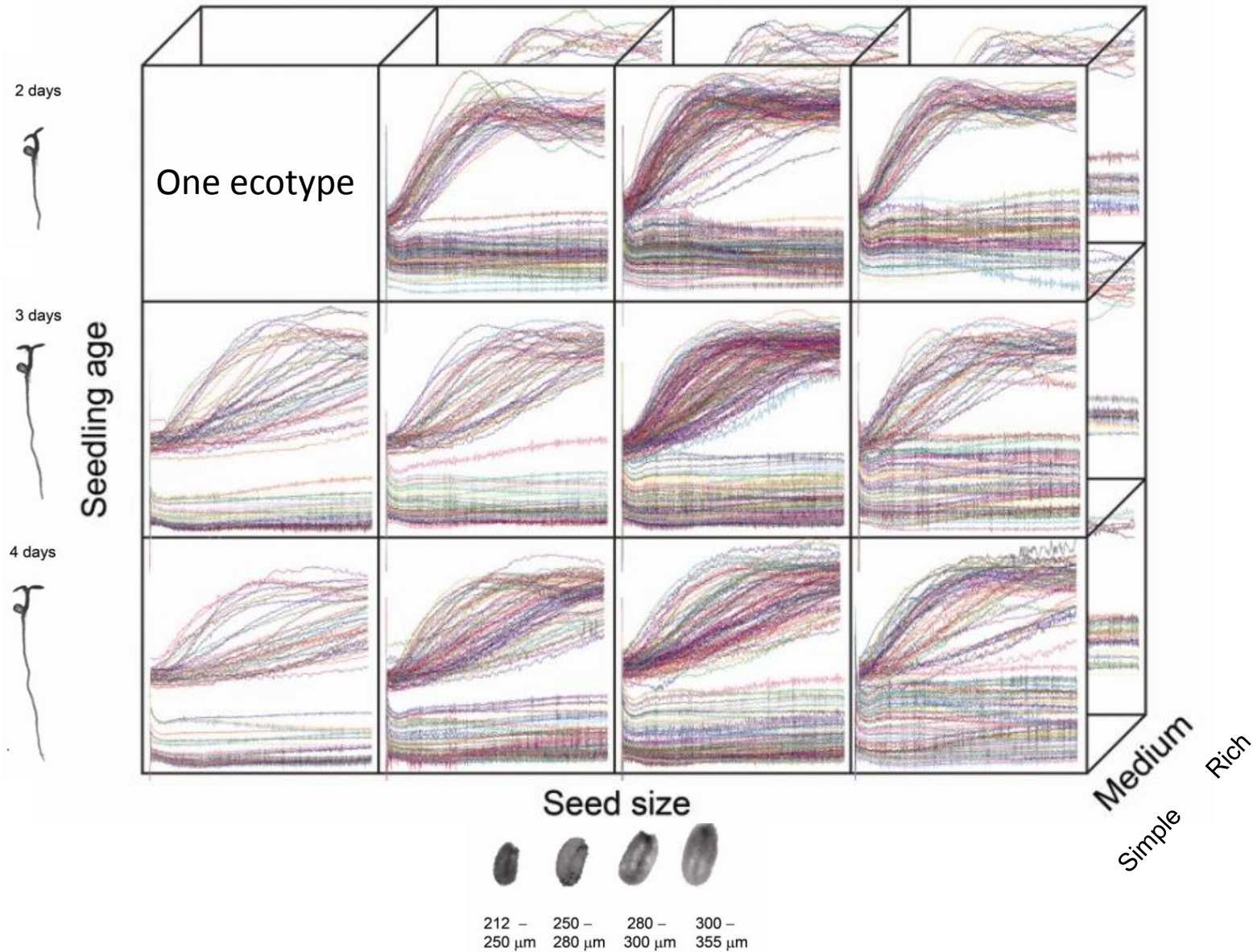
Requirements:

- ✓ Observations should be made at sufficiently high spatial and temporal resolution
- ✓ Method should be relatively high-throughput to allow genomic survey
- Observations should be able to be made over time and in many environmental contexts

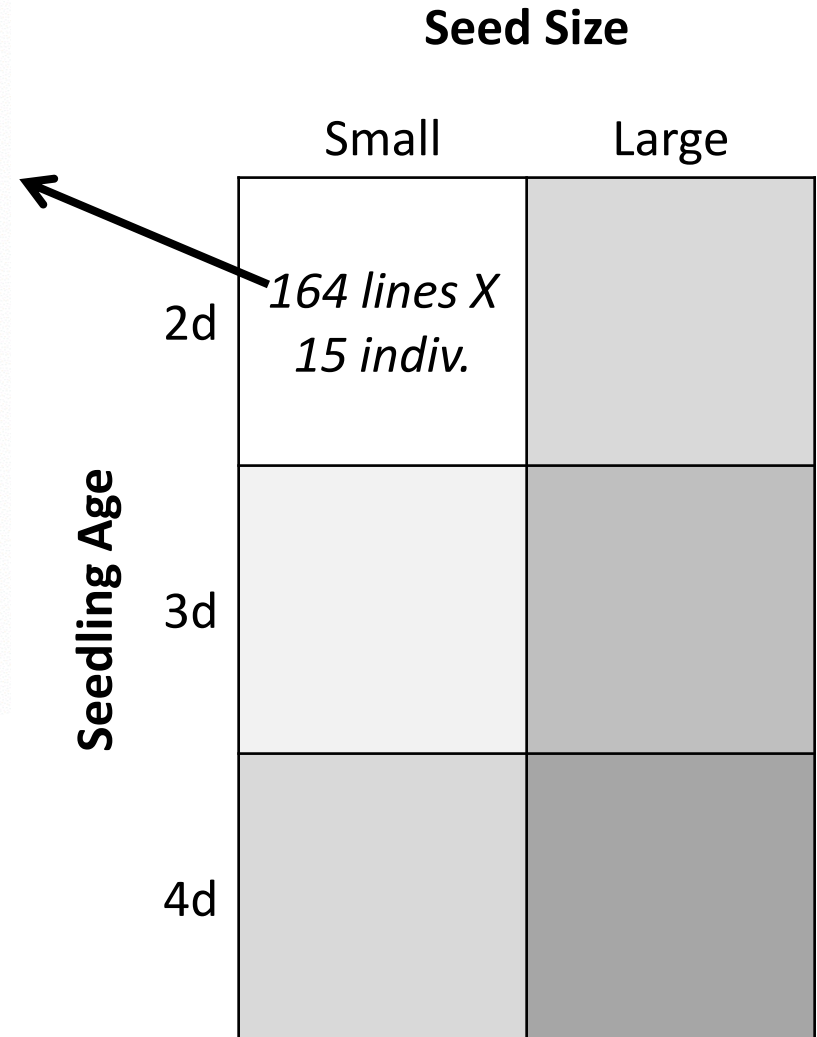
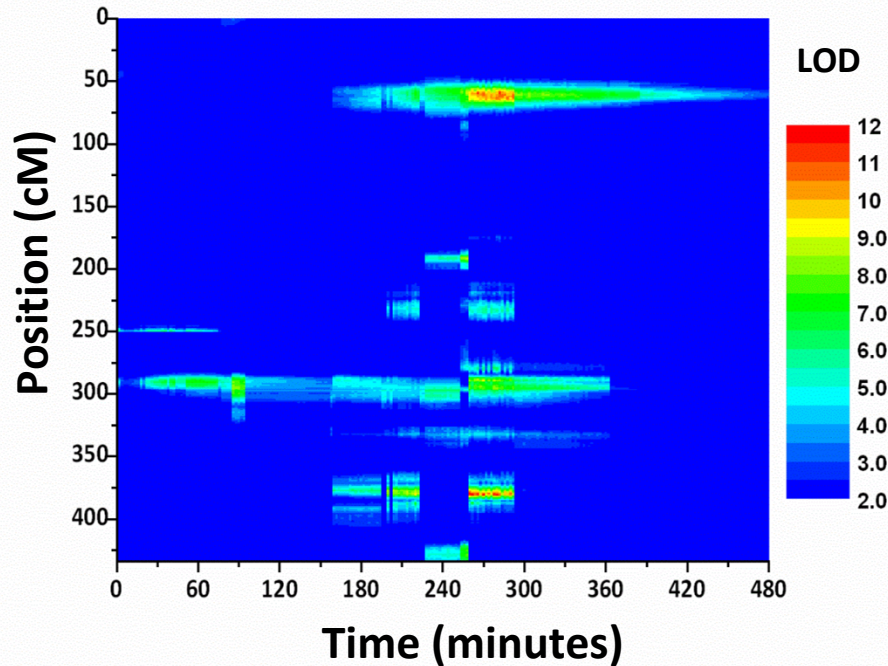


~25,000 genes

Phenotypes are plastic



Genomics analysis in a multi-dimensional condition space



Moore, et al., unpublished result

Developing tools to detect genome function at the organismal scale.

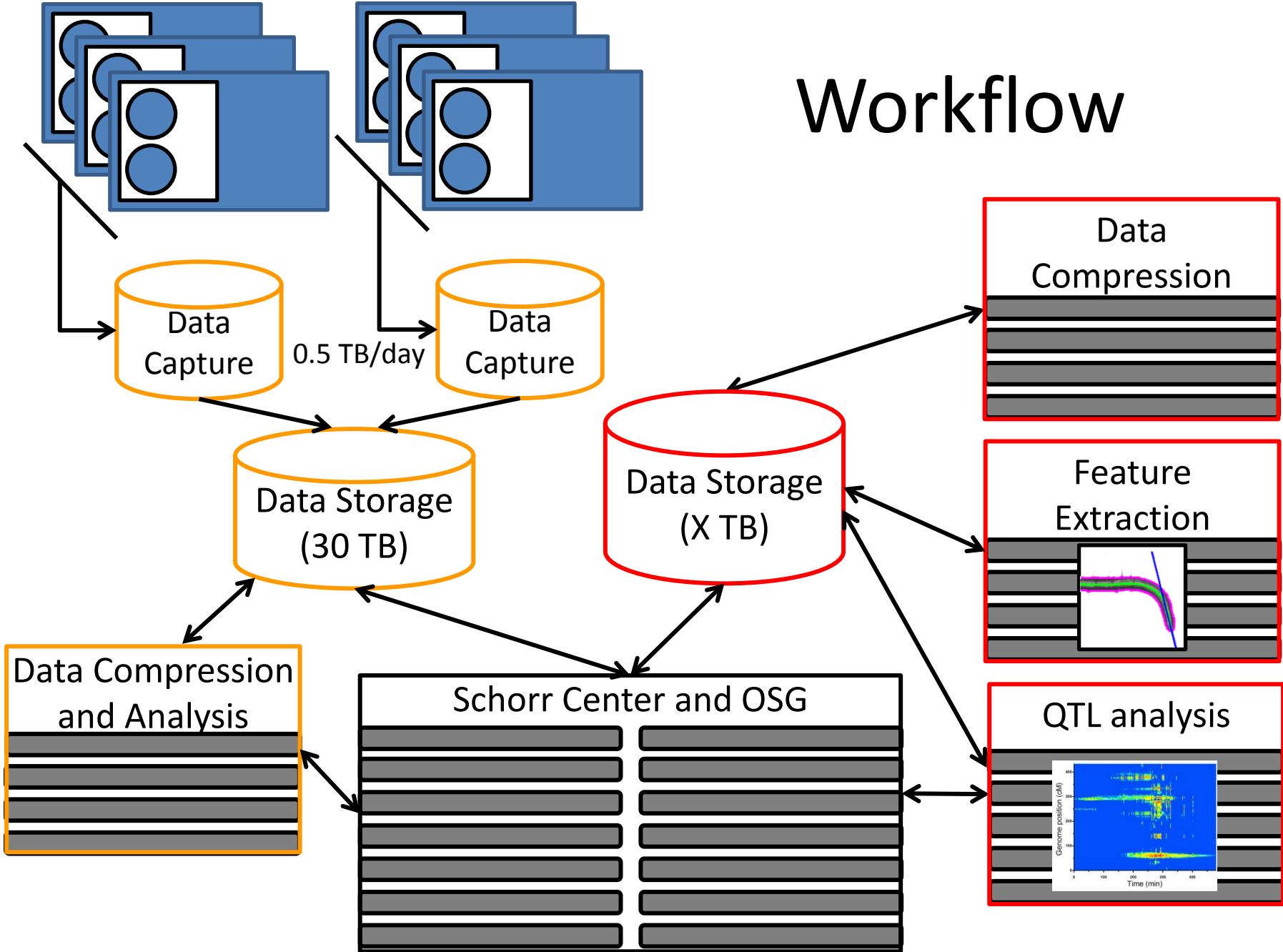
Requirements:



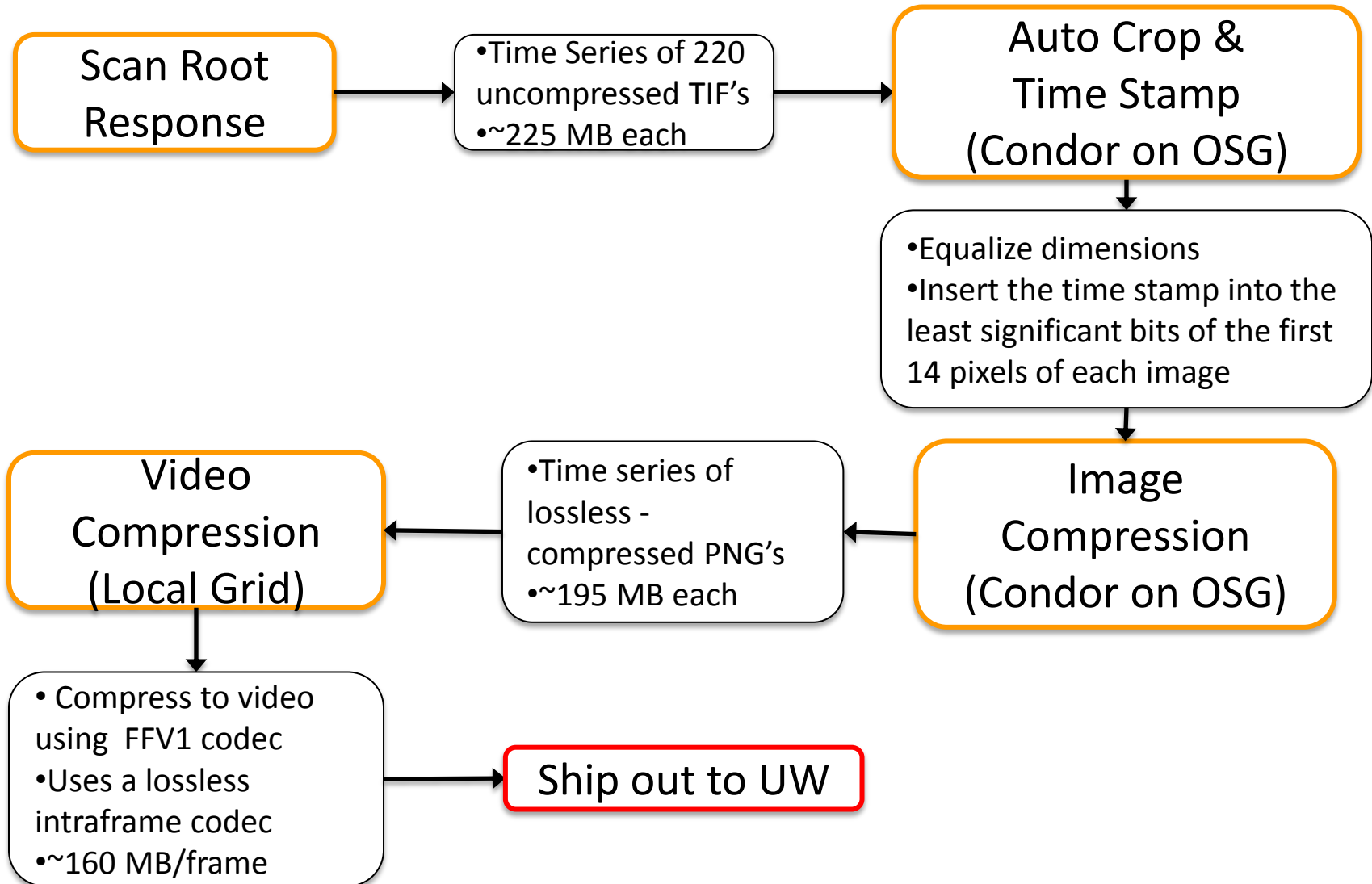
~25,000 genes

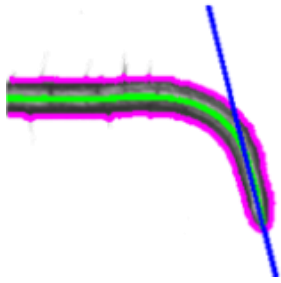
- ✓ Observations should be made at sufficiently high resolution
- ✓ Method should be relatively high-throughput to allow genomic survey
- ✓ Observations should be able to be made over time and in many environmental contexts
- **Cyberinfrastructure must facilitate the above**

Workflow

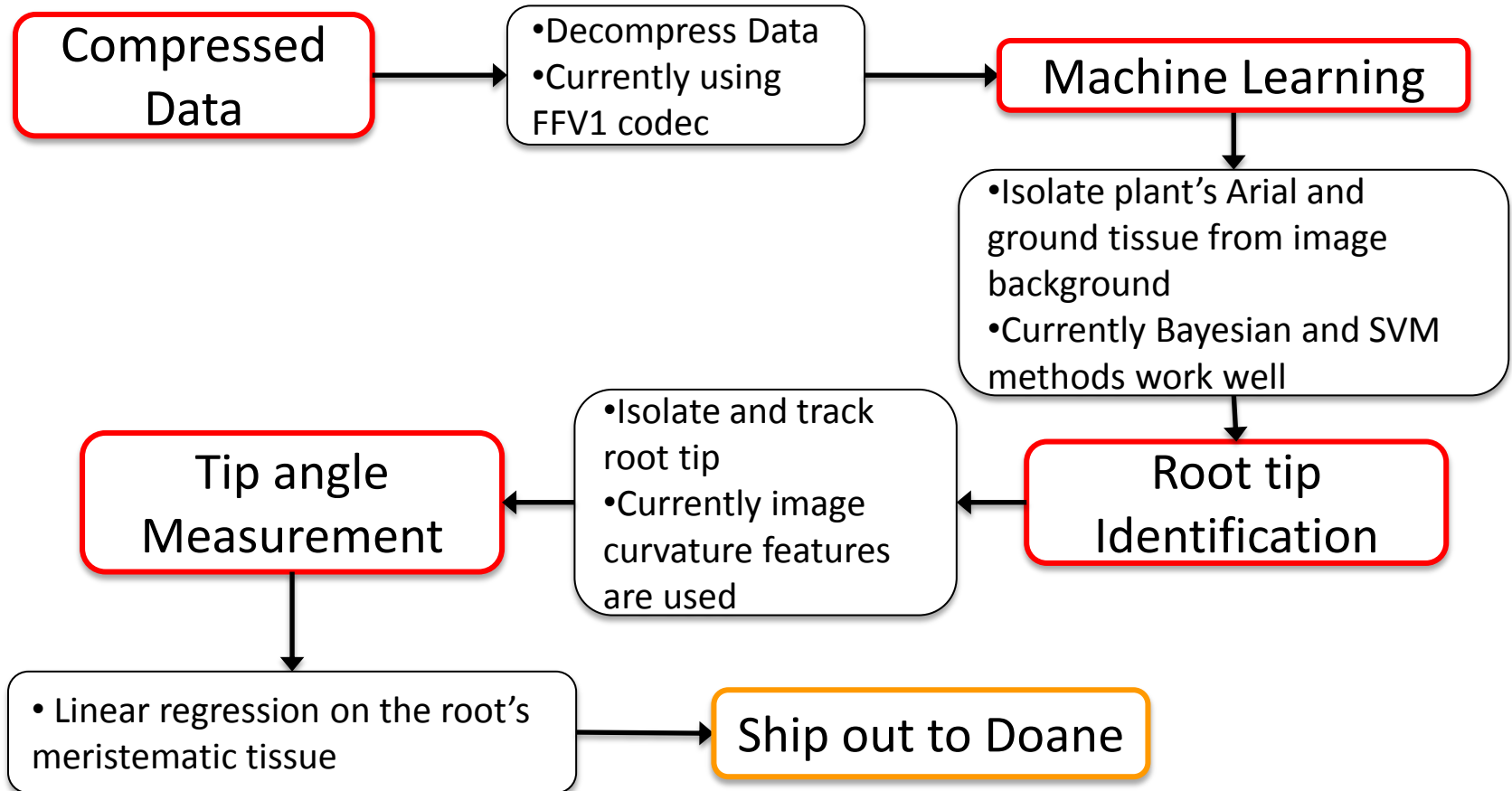


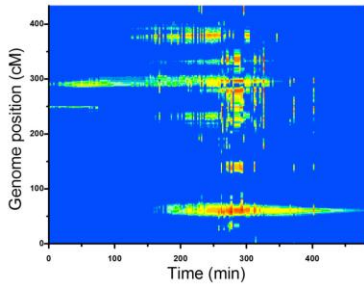
Data Compression



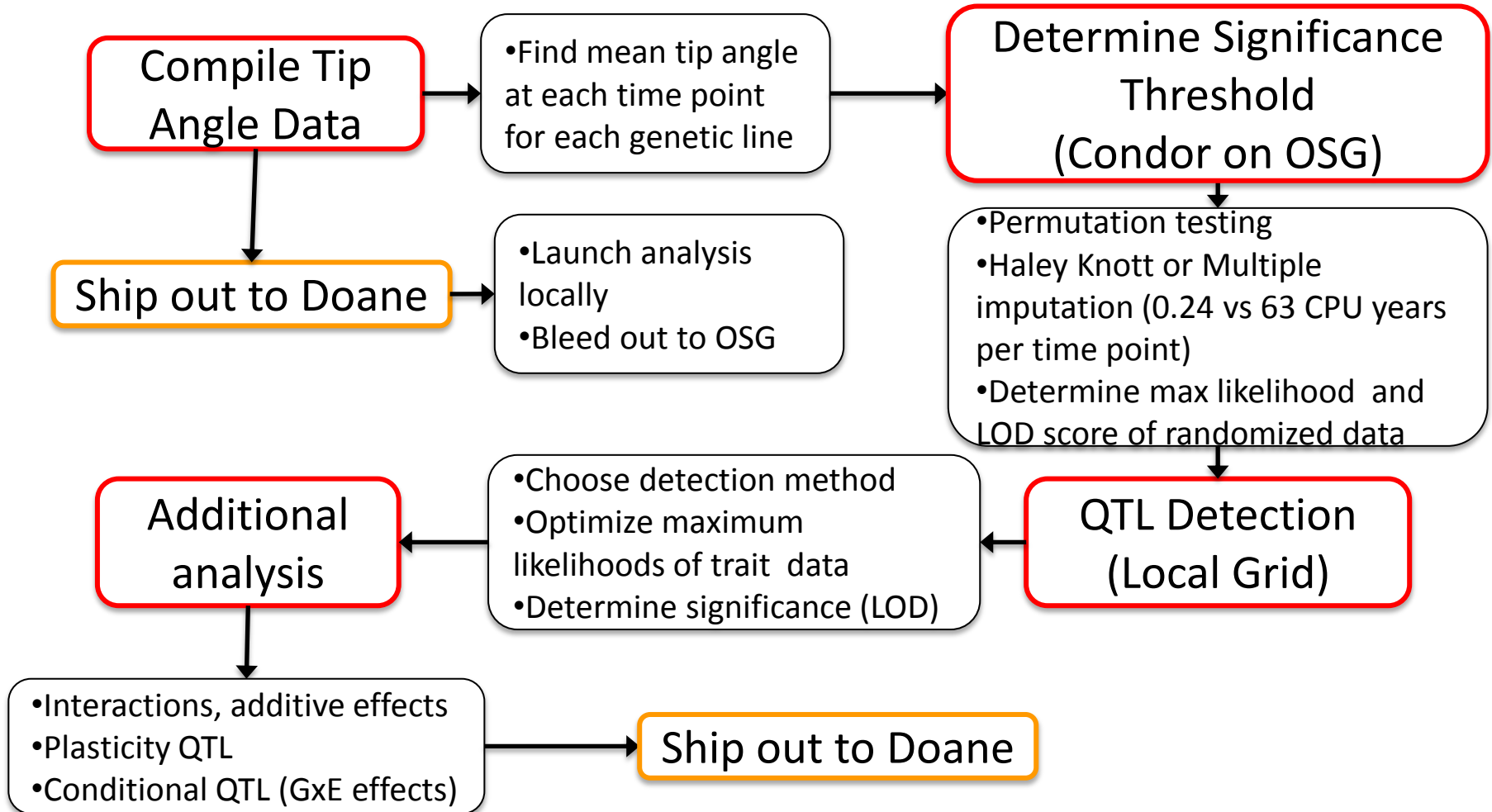


Feature Extraction





QTL Analysis - Association of a phenotypic value (e.g. root tip angle) with a genetic element



Progress and Future Directions

- One small college has collected over 14,500 individual root gravitropic responses in six conditions (32 TB) in RIL population in 6 mo.
- We will finish collection from NILs (near-isogenic lines) - an additional 8,700 individuals, 19 TB in 1-2 mo.
- Begin image analysis and QTL analysis – dataset opens new doors in visualizing genomes
- Interdisciplinary undergraduate training opportunities

Acknowledgments

Doane College

Mike Carpenter (CIO), David Andersen, Dan Schneider

Chris Wentworth (Physics) and Alec Engebretson (IST)

Students: Amy Craig and Brad Higgins (Physics), Autumn Longo and Grant Dewey (Biochemistry), Tracy Guy, Miles Mayer, Halie Smith, Anthony Bieck, Sarah Merithew, Devon Niewohner, Muijj Ghani, Julie Wurdeman (Biology)

University of Wisconsin

Edgar Spalding's Laboratory: Nathan Miller, Candace Moore, Logan Johnson

Miron Livny (CHTC)

UNL – Schorr Center and HCC

David Swanson, Brian Bockelman, Carl Lundstedt

University of Florida

Mark Settles

Computing Resources



Funding



PGRP - 1031416
EPSCoR - URE



NE-INBRE



DOANE
COLLEGE

Questions?

