



Grid Computing to Study the Functions of Plant Genes

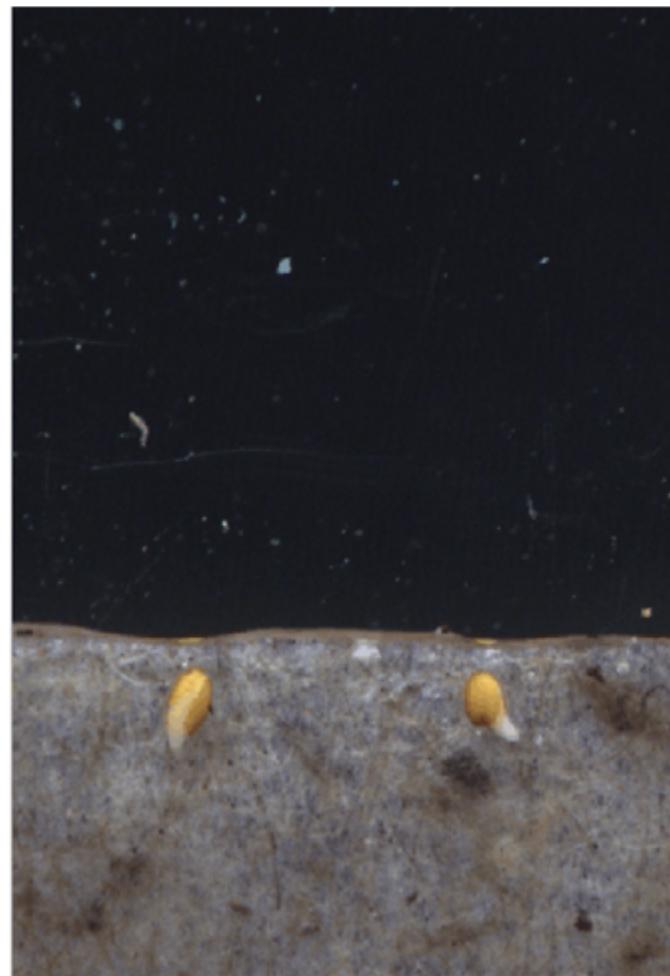
Edgar Spalding
Department of Botany



www.botany.wisc.edu/spalding.htm

A Bit of Background

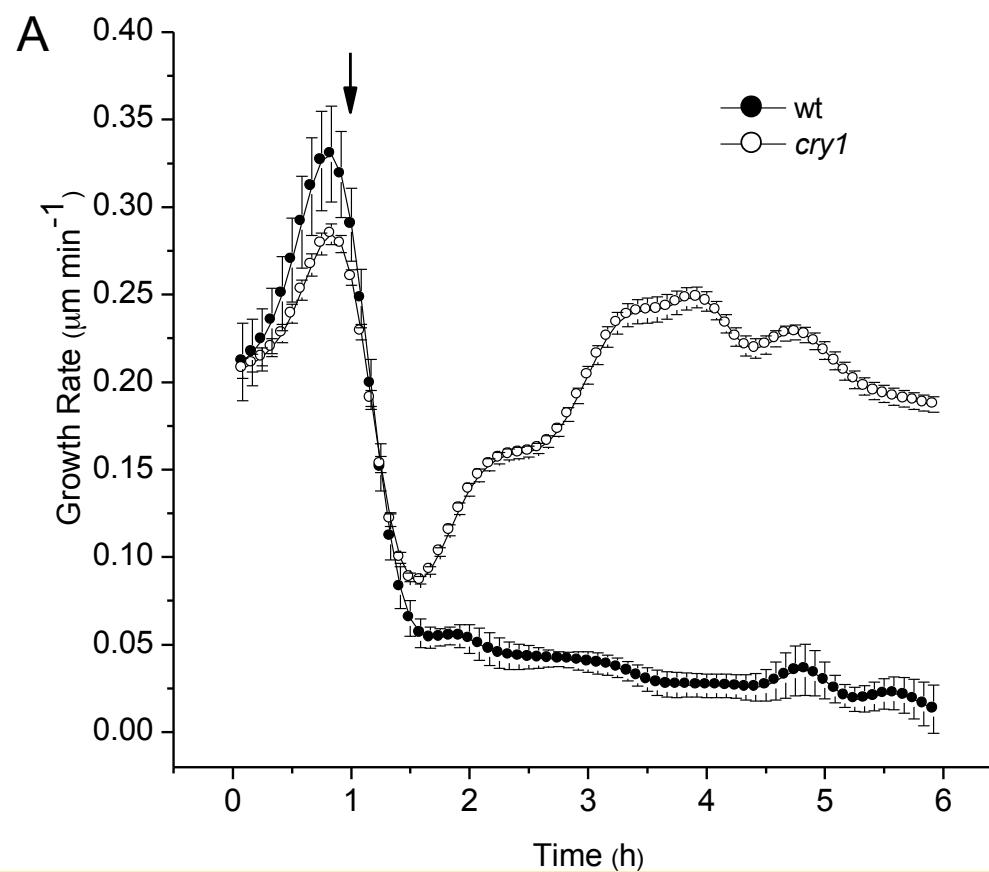
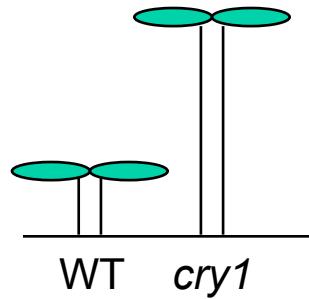
- A major goal in biology is to learn the function of each gene in an organism.
- A proven approach is to compare the behaviors of individuals possessing different versions of that gene.
- Organisms have on the order of 10^4 genes so that makes for a lot of comparisons.



24,999 genes

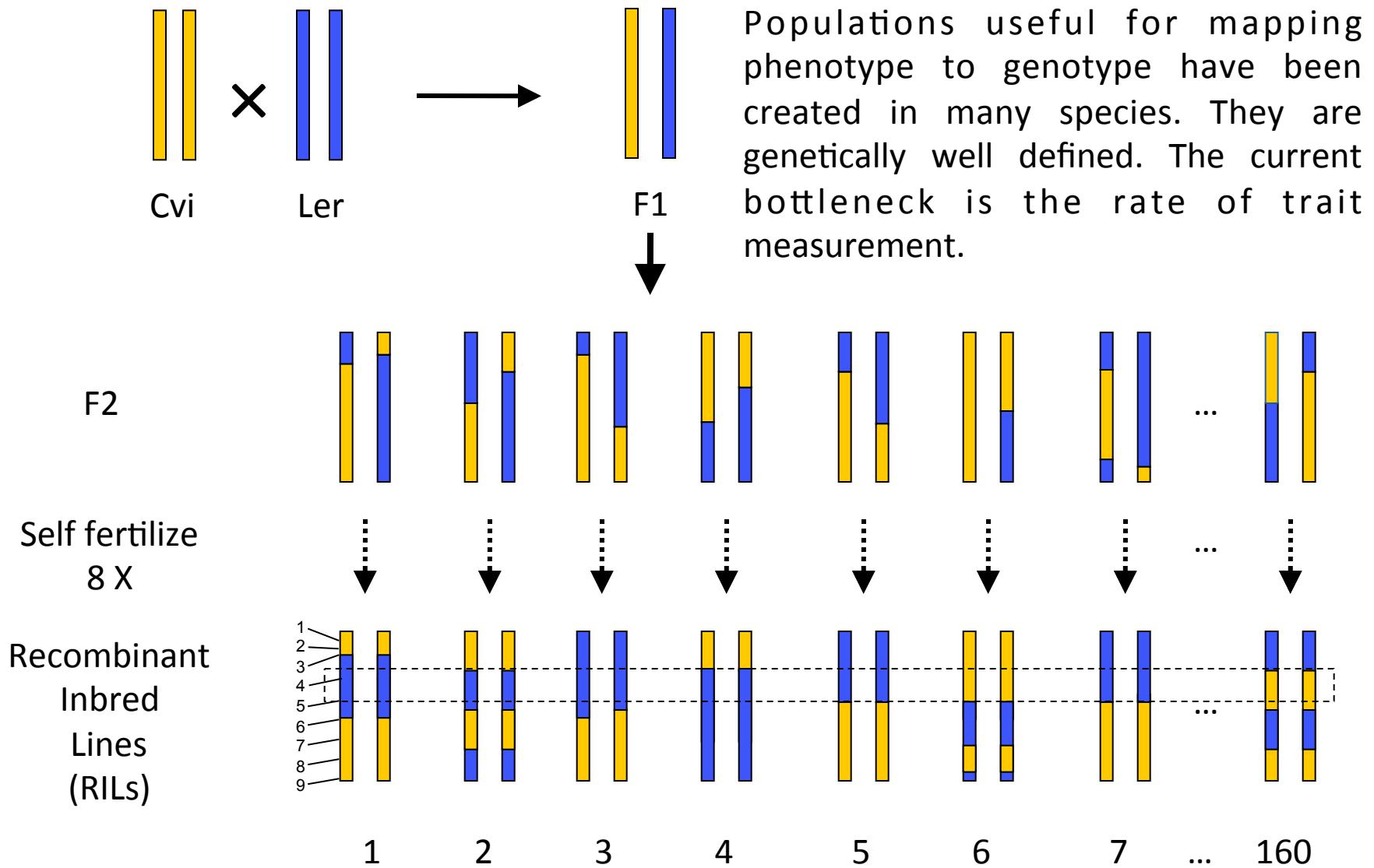
25,000 genes

Hypocotyl Growth Inhibition Induced by Blue Light



We developed image processing algorithms for measuring seedling growth and development in order to quantify effects of genetic differences (i.e. phenotypes of mutants) in space and time. Our purpose was to answer questions about one or two genes.

Why not / How to scale up to the whole genome level?



Alonso-Blanco *et al.* *The Plant Journal* **14**, 259-257 (1998)

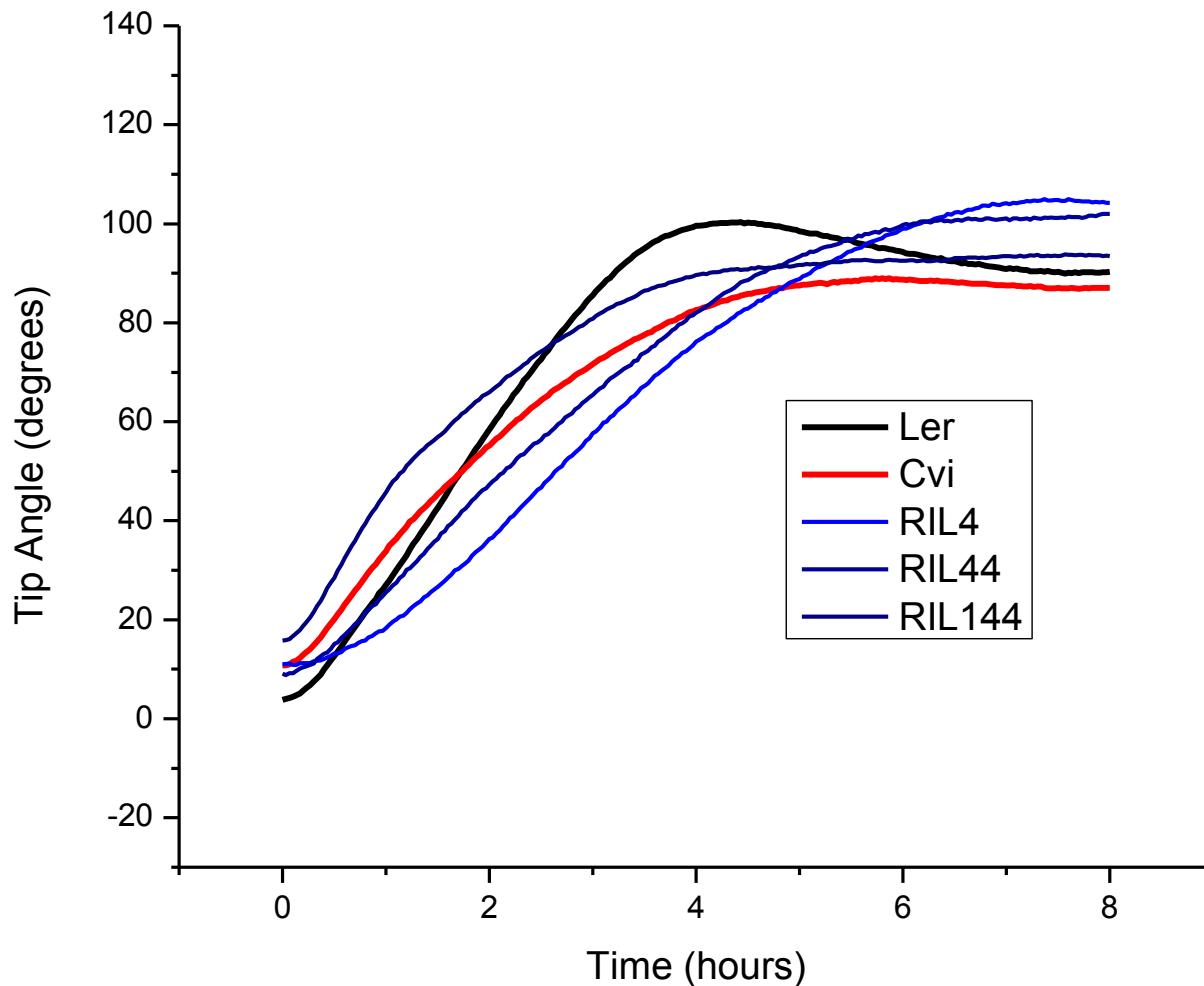
Let's make the process High Throughput
Switch to Root Gravitropism...

because we could study with high resolution, high accuracy, and high throughput

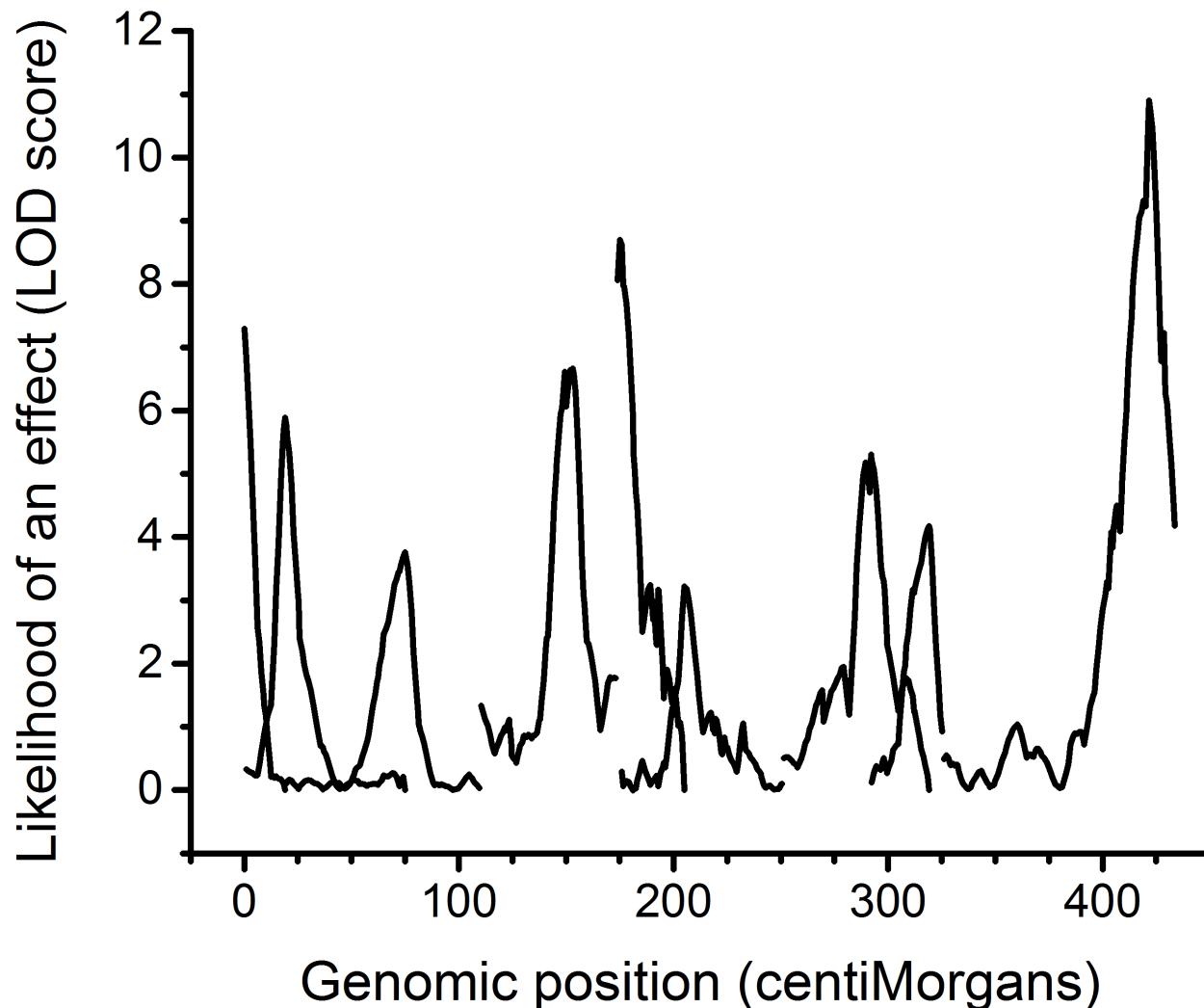


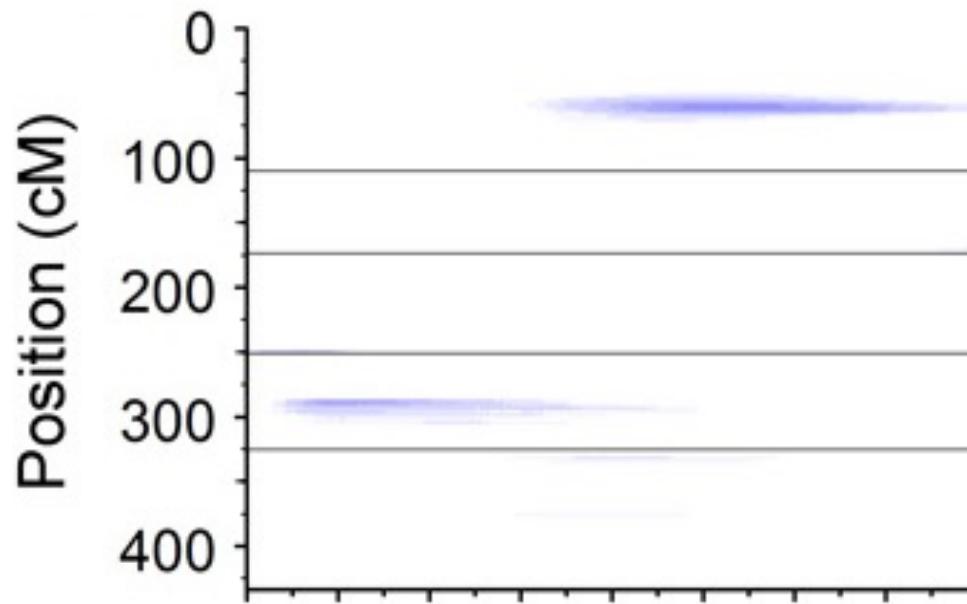
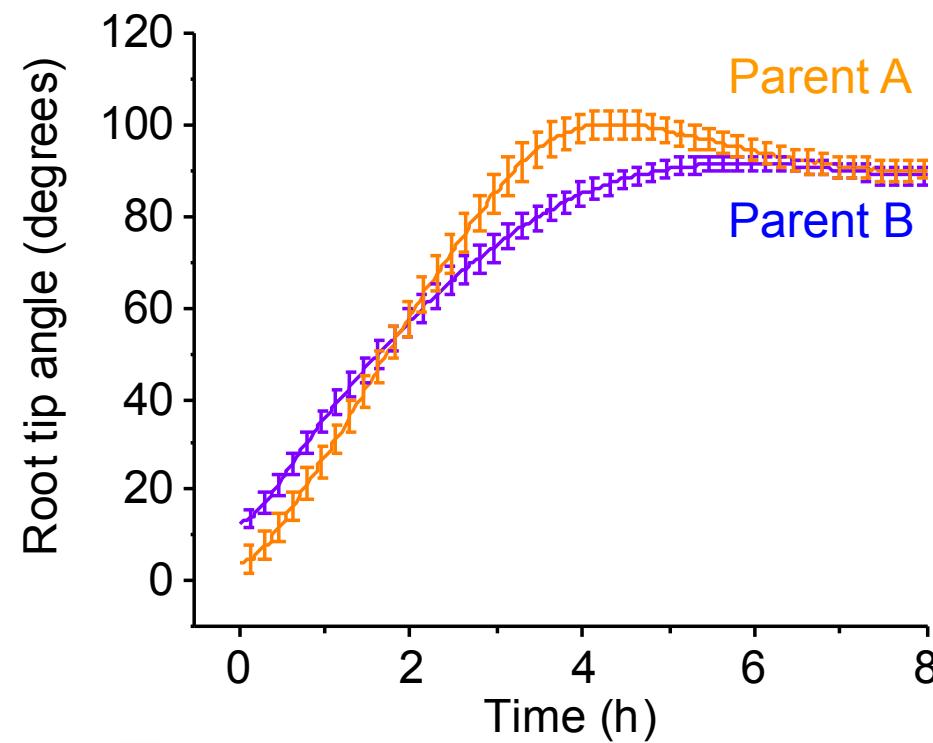
Machine Vision to Study Natural Genetic Variation

160 *Ler* X *CVI* recombinant inbred lines for QTL mapping

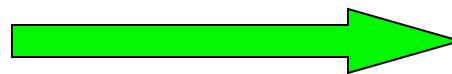


A single QTL plot
(for example - root tip angle at one point in time is the phenotype)





Our Automated Workflow

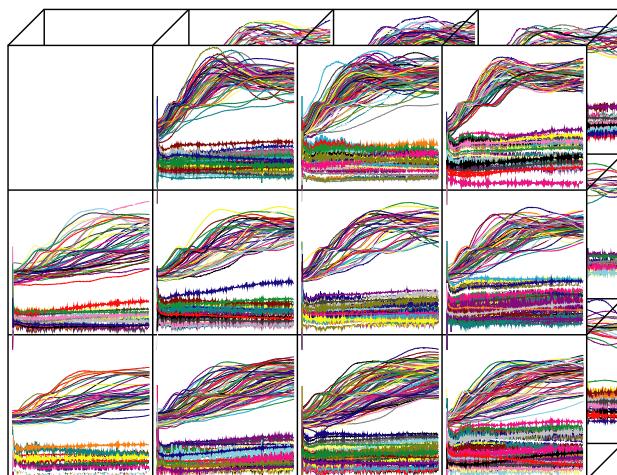


automated image acquisition
generates movies



data mining

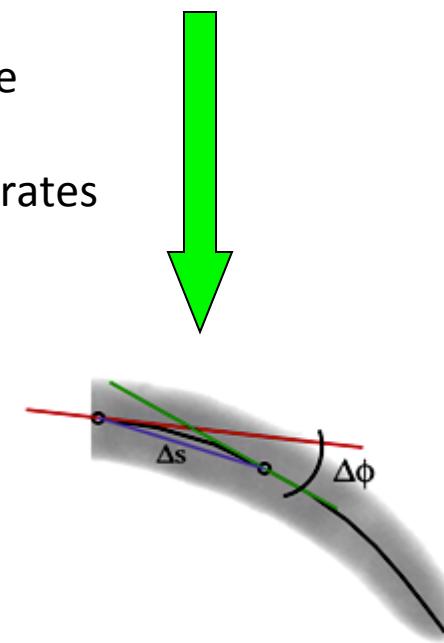
test new
hypotheses



algorithmic feature
extraction and
quantification operates
on each object...



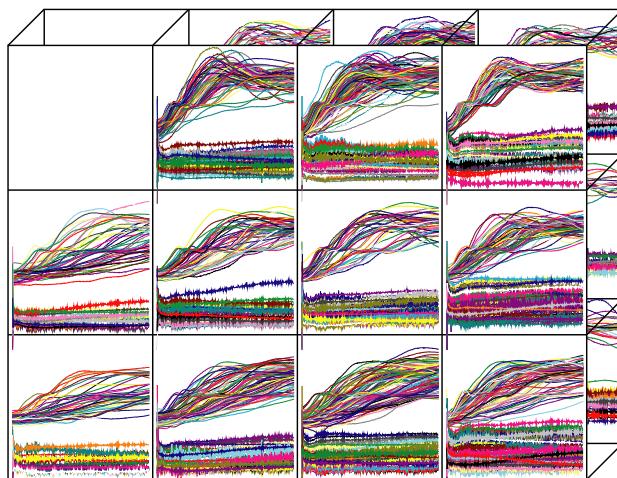
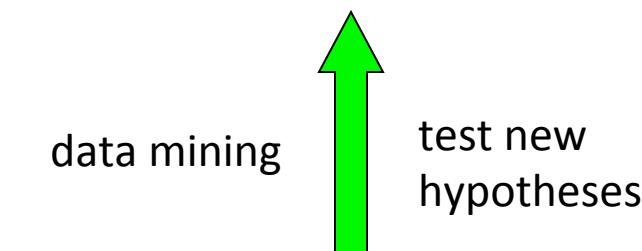
...resulting in large, high
dimension data sets...



Our Automated Workflow



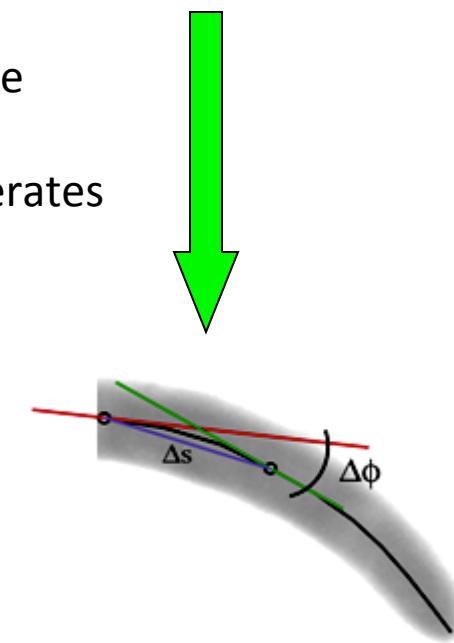
automated image acquisition
generates movies



algorithmic feature
extraction and
quantification operates
on each object...



...resulting in large, high
dimension data sets...



To set the LOD score significance threshold you permute the phenotype against the genotype lots of times, to see what kind of signal ‘chance’ produces.

We chose to perform 25,000 permutations. To distribute this work on a grid, we bundled 5 permutations per job and ran 5000 jobs.

5 permutations
per job

\times 5000 jobs = 25,000 permutations

But we made measurements every 2 min for 8 h, essentially 241 separate expts

5000
permutation
jobs

@ 241 time points

5000
permutation
jobs

5000 jobs @ 241 time points = 1,205,000 jobs

This work was performed by HTCondor *Glide In* from CHTC to OSG

The Key People



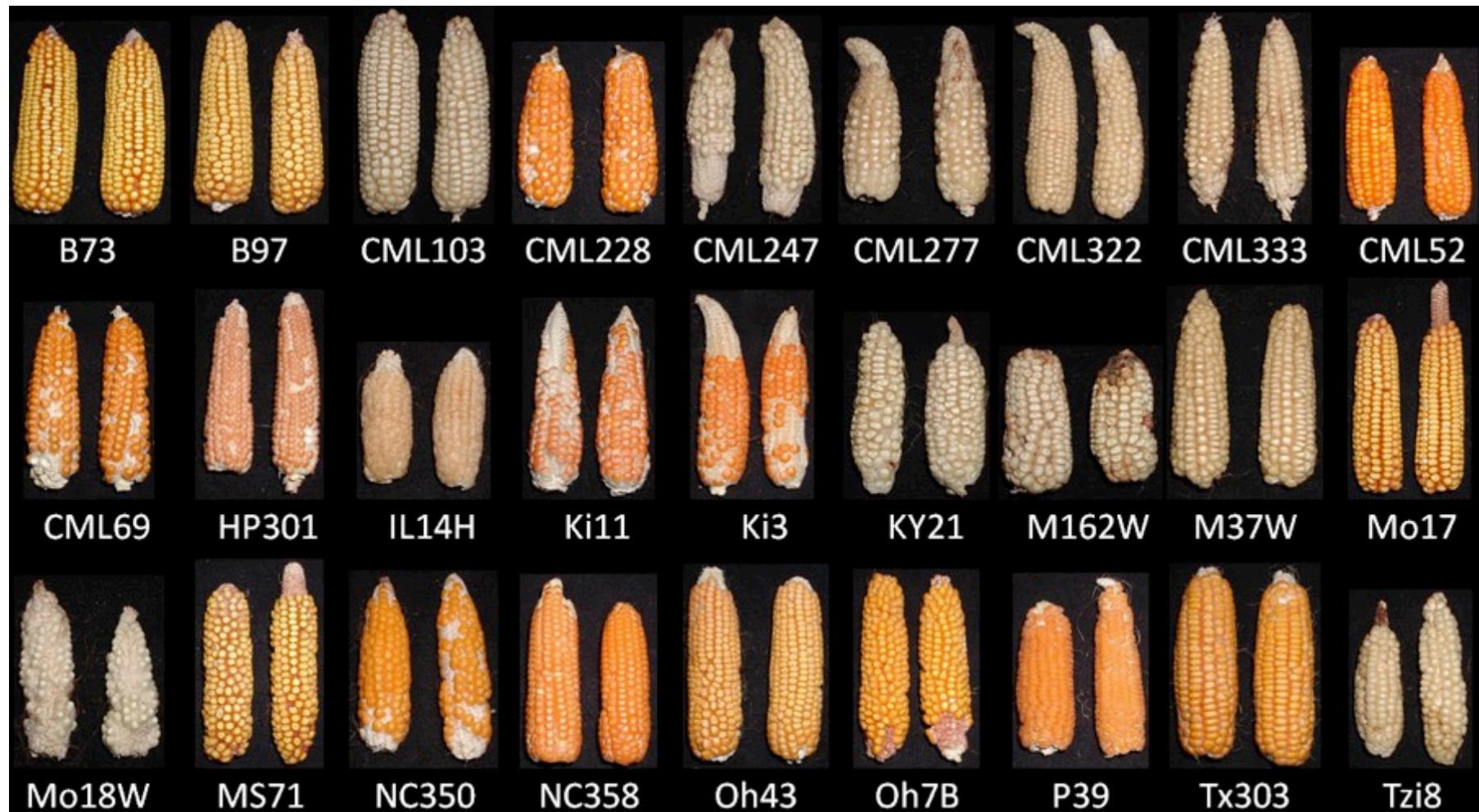
Logan Johnson

Candace Moore

Nobody eats root tips, but we all eat a lot corn



Huge diversity in ear shapes among genetically-distinct maize types



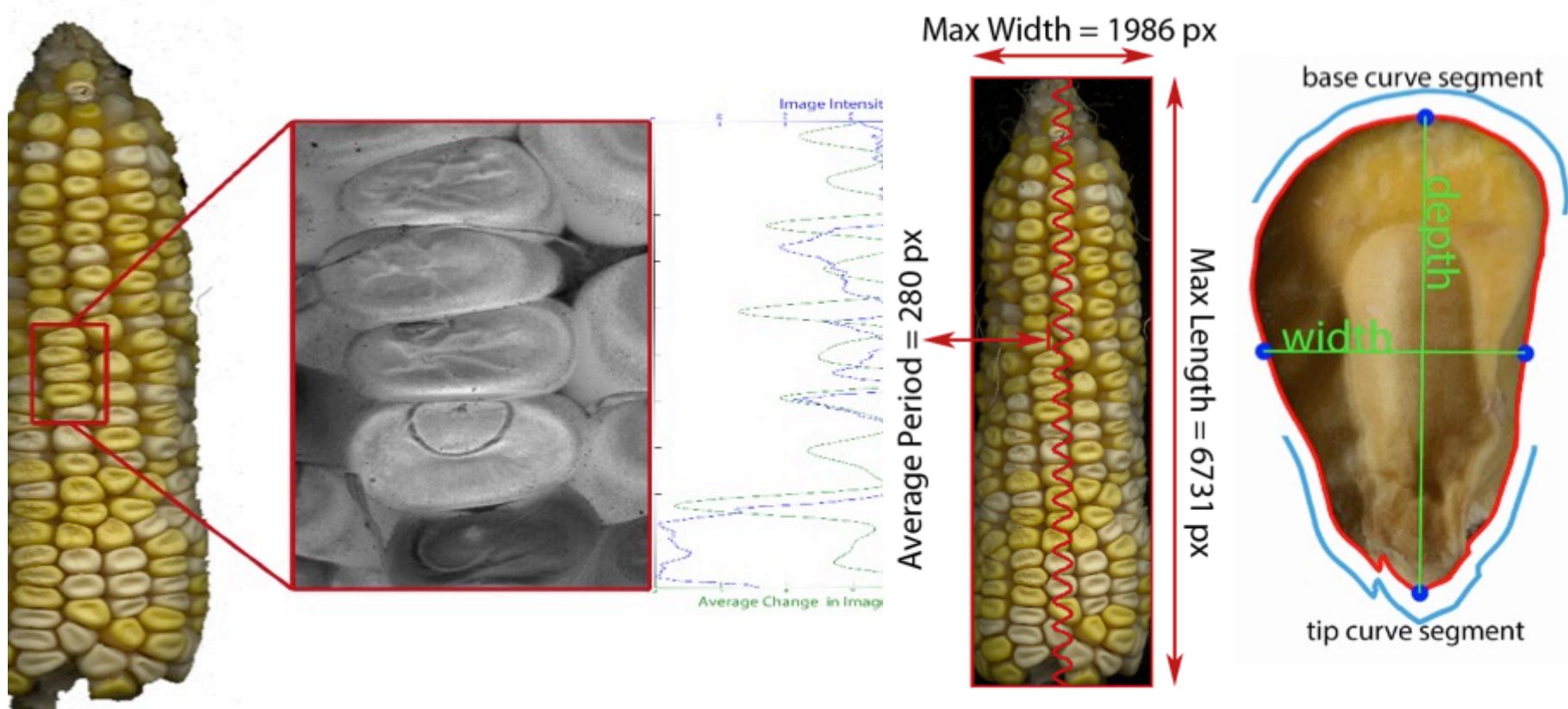
Is this the best way to measure corn ear ‘traits’?



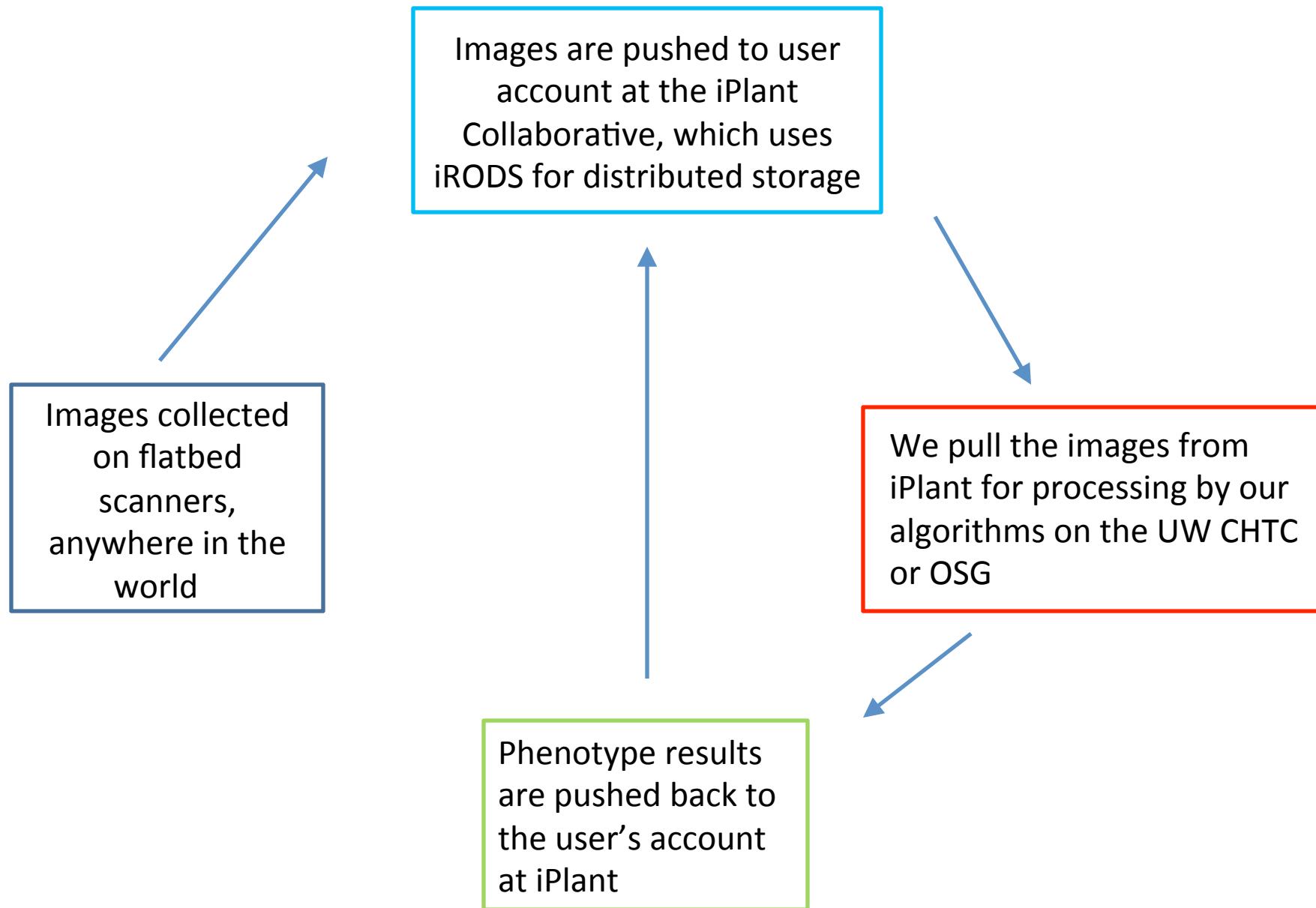
Too funny –

“Geneticist Ed Buckler measures a maize ear for statistical analysis.”

Corn ear and kernel traits measured by image analysis



The current corn ear and kernel workflow



Biology is becoming a high-throughput science

New insights will come from Big Data

People who are comfortable with HTC will win!