THE UNIVERSITY of WISCONSIN MADISON

High Throughput Computing to Study the Functions of Plant Genes

Edgar Spalding
Department of Botany

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.

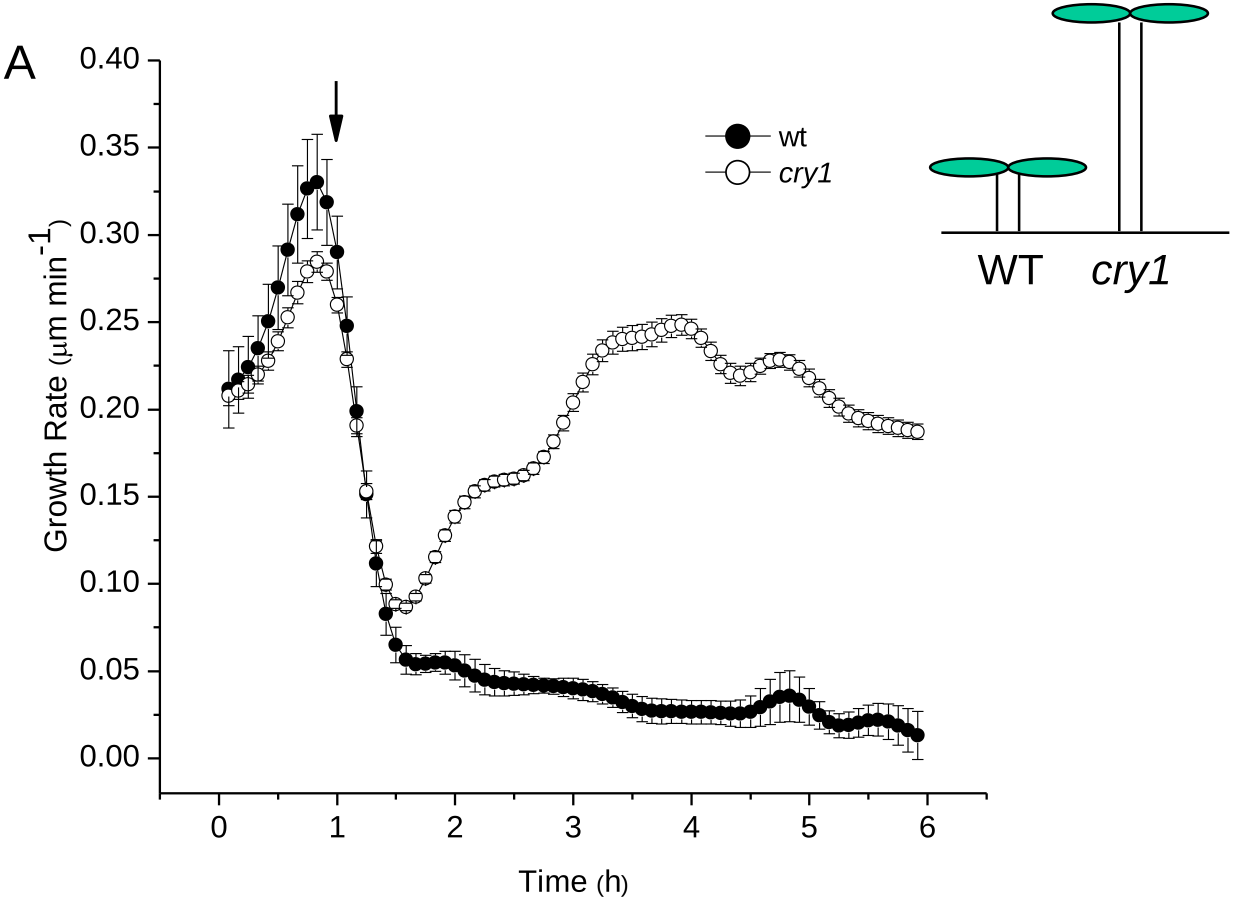
(Video removed)

24,999 genes

25,000 genes

Hypocotyl Growth Inhibition Induced by Blue Light

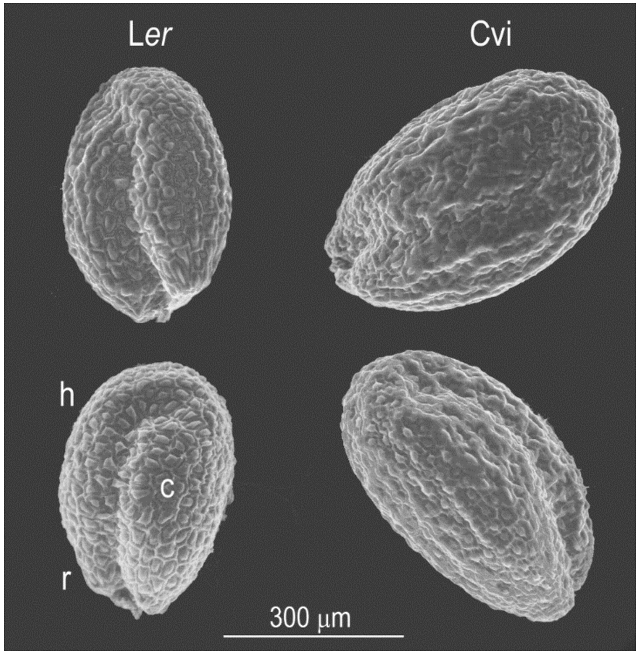
Video removed

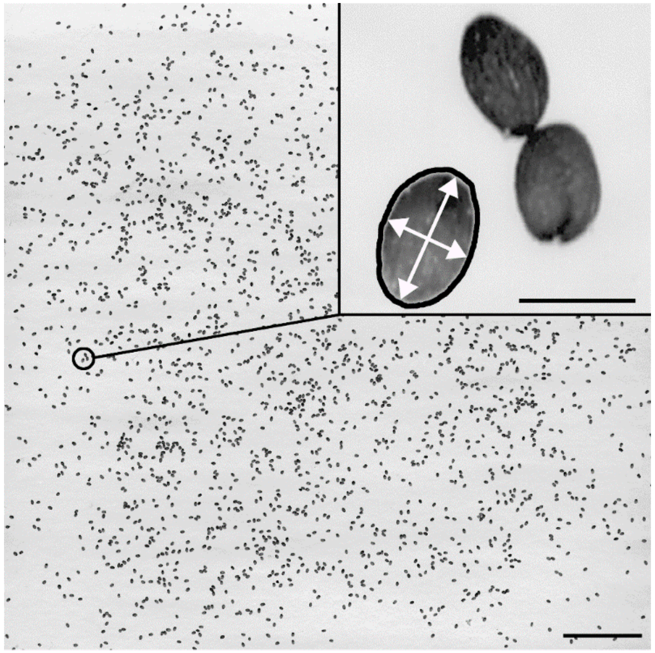


We developed image processing algorithms for measuring seedling growth and development in order to quantify effects of genetic differences (i.e. phenotypes of mutants) in space and time. Our purpose was to answer questions about one or two genes.

Natural Phenotypic Variation

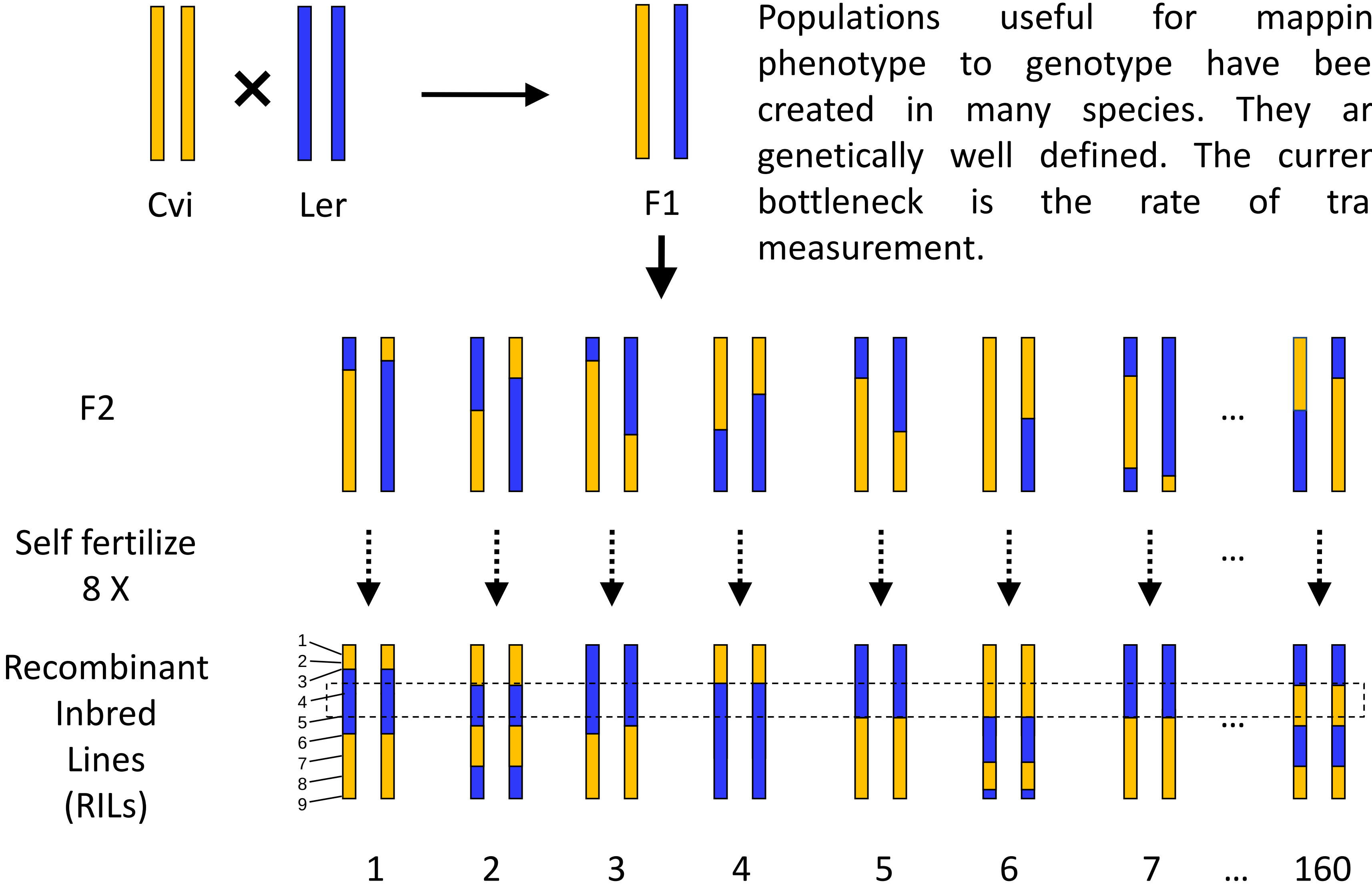
within 160 L*er* X CVI recombinant inbred lines enables QTL mapping





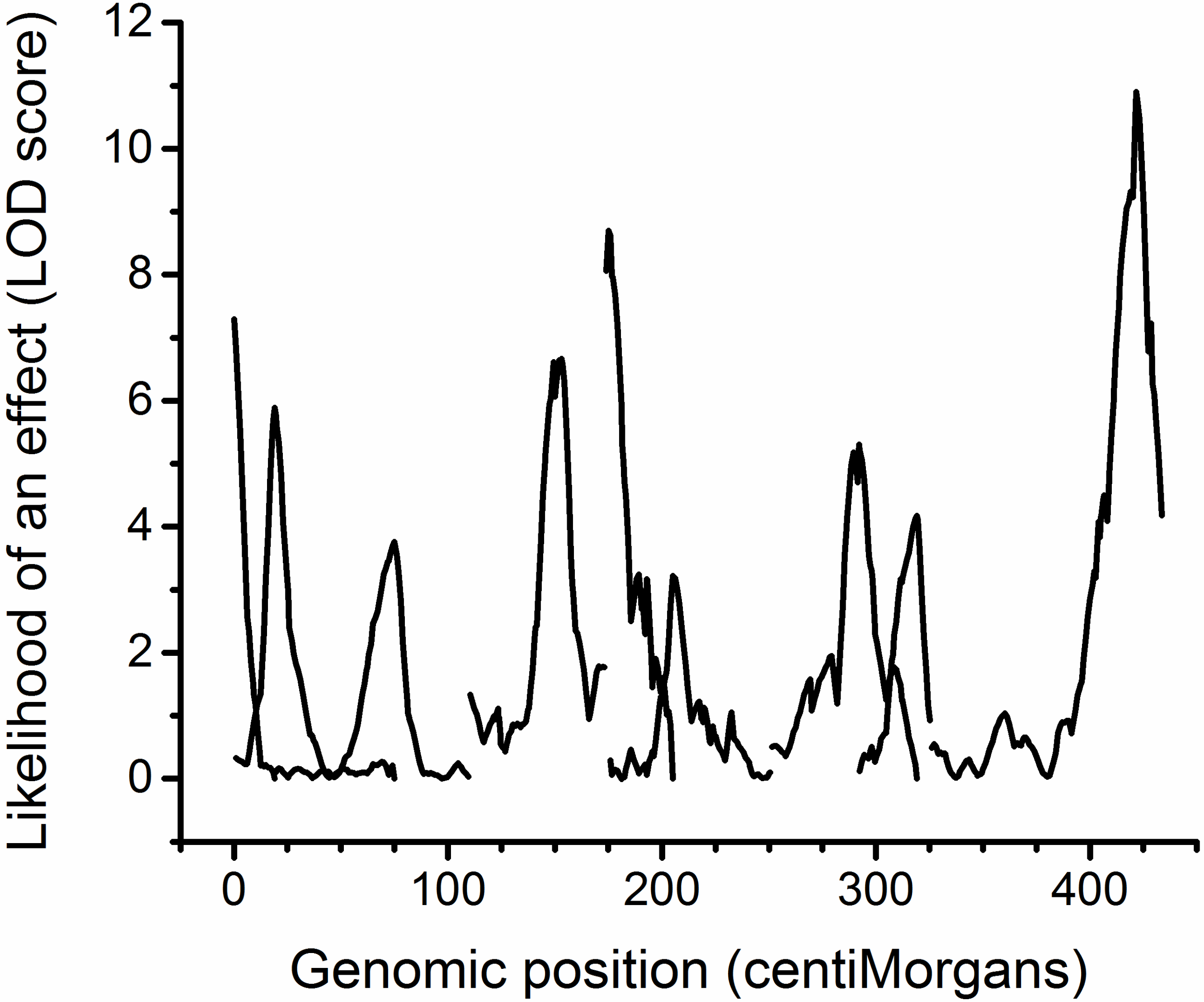
Why not / How to scale up to the whole genome level?

Populations useful for mapping phenotype to genotype have been created in many species. They are genetically well defined. The current bottleneck is the rate of trait measurement.



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

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



Root Gravitropism

(video removed)

Banks of computer-controlled CCD cameras

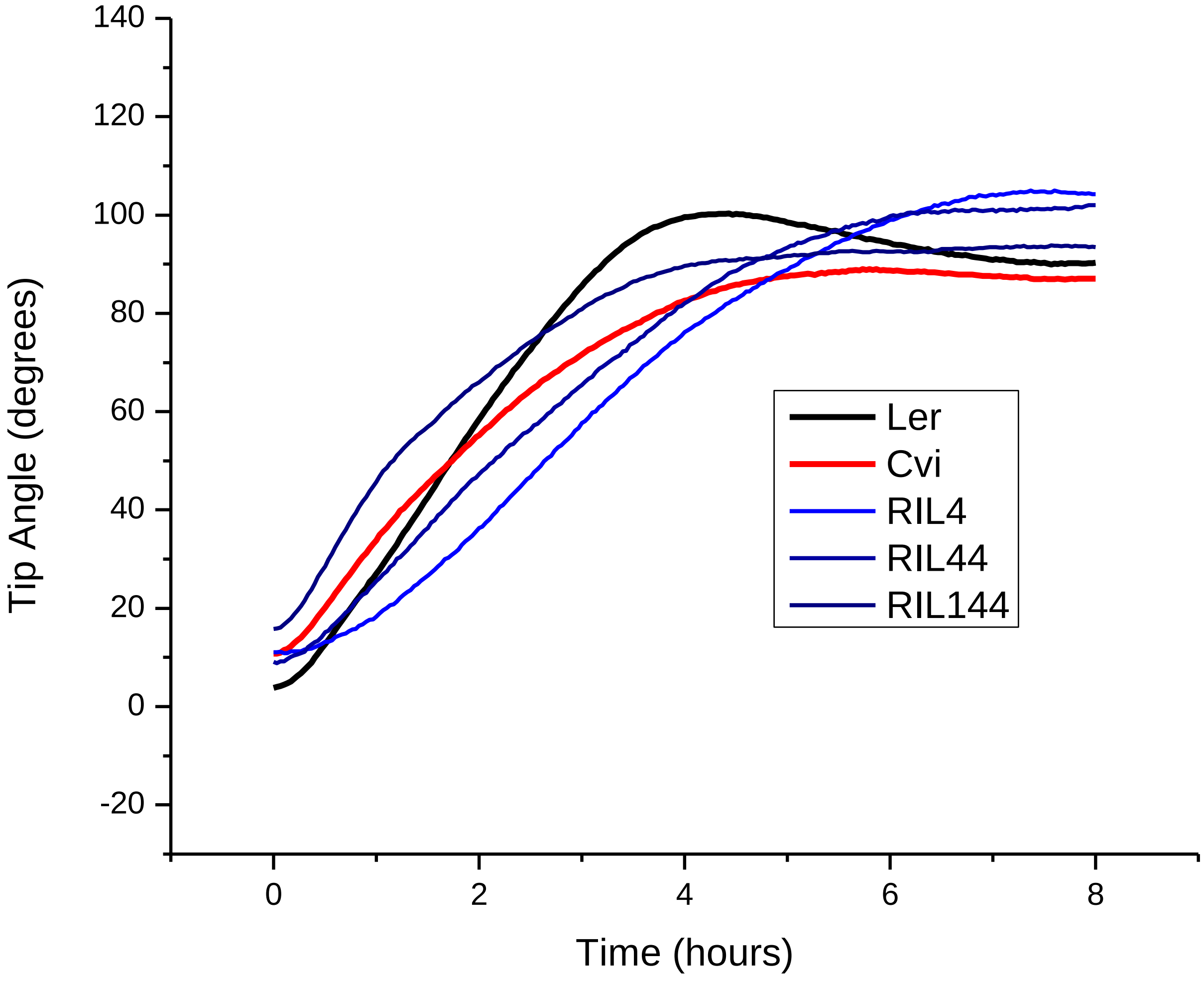
Images are automatically submitted as HTCondor jobs for distributed processing

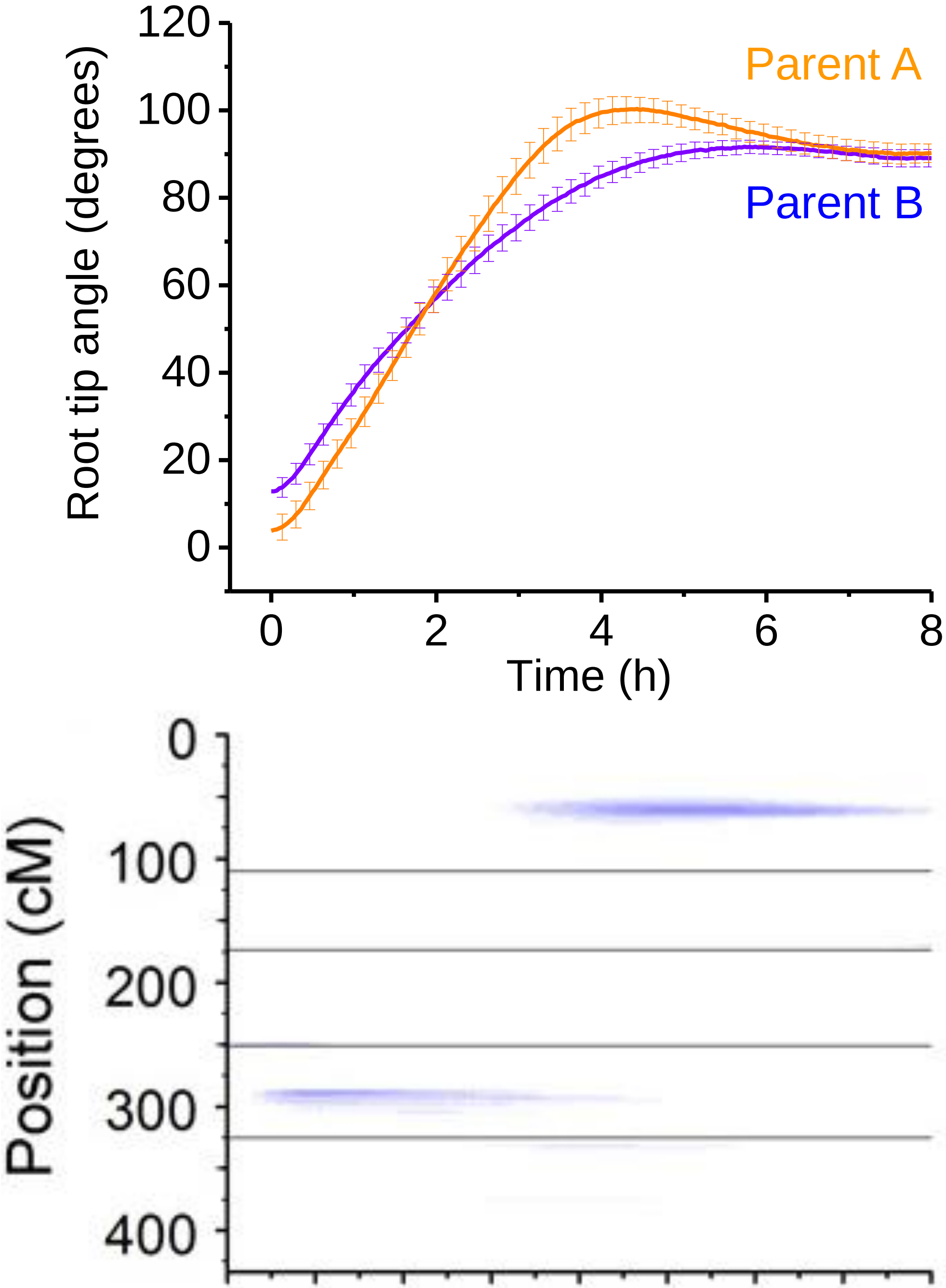
CENTER FOR
HIGH THROUGHPUT
COMPUTING

Miron Livny
a.k.a Captain Condor
Director of CHTC

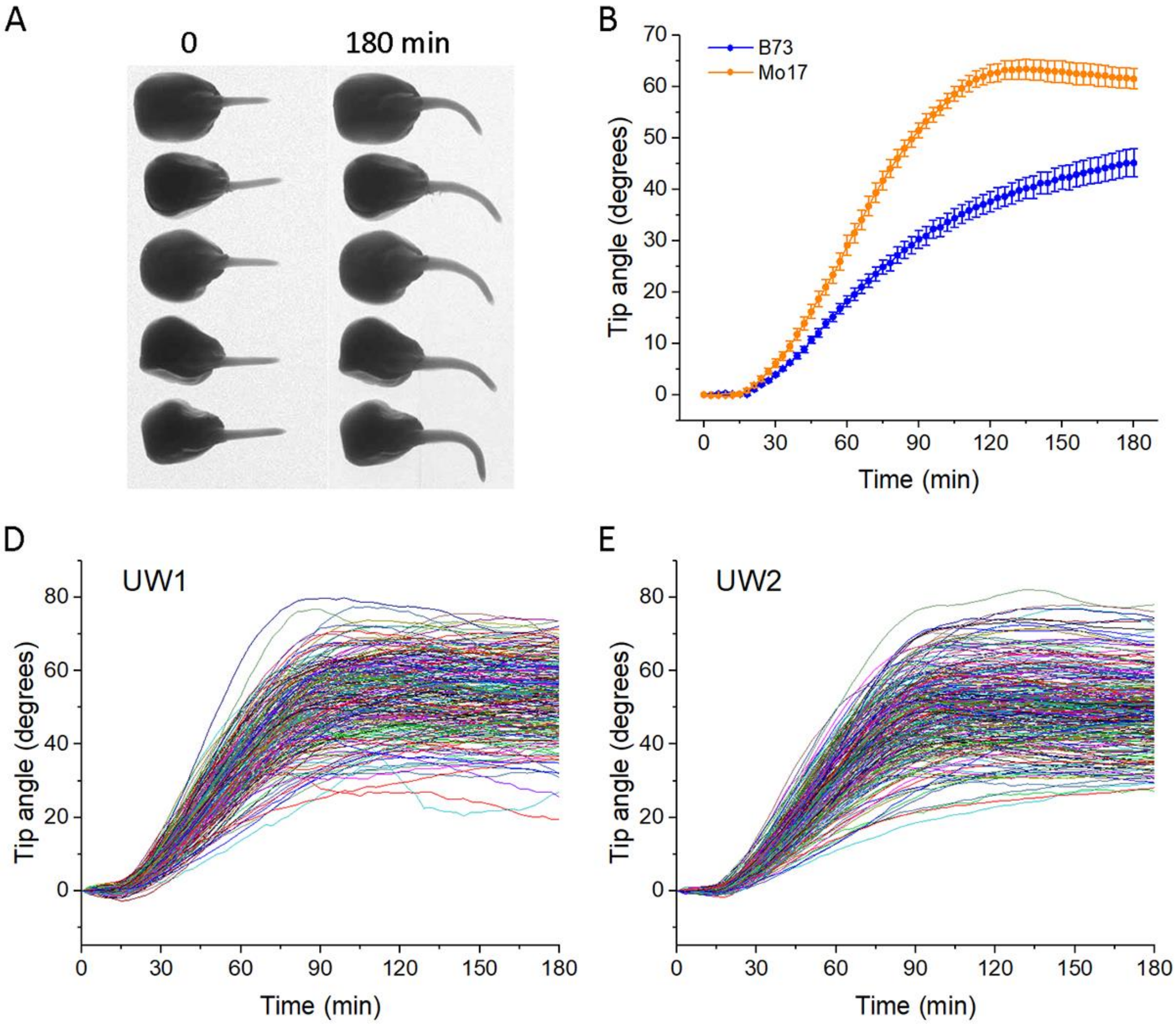
Machine Vision to Study Natural Genetic Variation

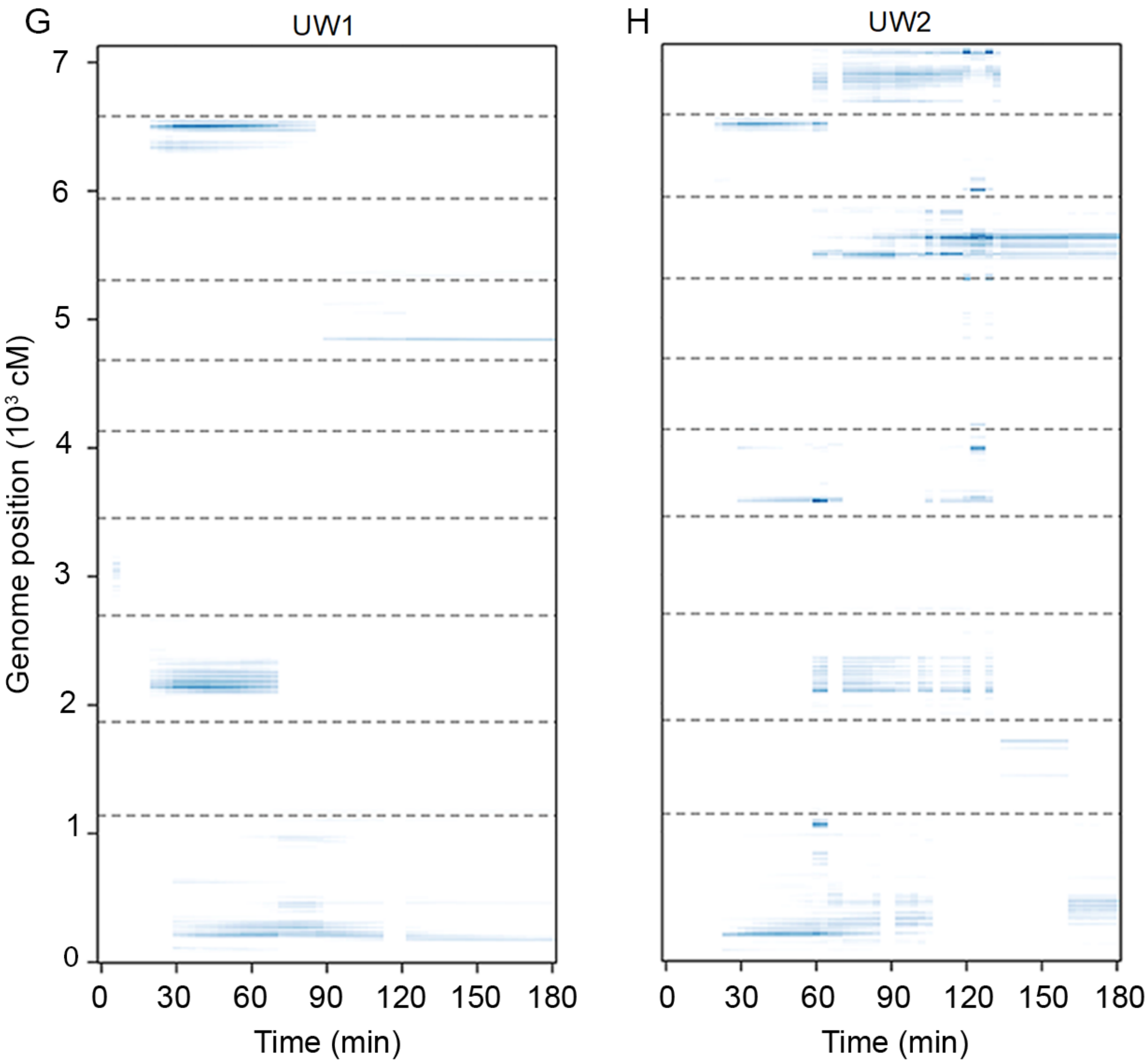
160 L*er* X CVI recombinant inbred lines for QTL mapping




Root tip angle (degrees)
120
100
80
60
40
20
0
Parent A
Parent B
0
2
4
6
8
Time (h)
Position (cM)
0
100
200
300
400

Now let’s do the same thing with corn




G
UW1
H
UW2
Genome position (10³ cM)
0
1
2
3
4
5
6
7
Time (min)
0
30
60
90
120
150
180
Time (min)
0
30
60
90
120
150
180

Now we have identified stretches of DNA that influence the process of gravitropism, in two species.

Those stretches of DNA are too long to be immediately useful because they encompass too many genes to test.

Arabidopsis,
about 500
genes

Maize,
about 5000
genes

7 genes were in both sets!

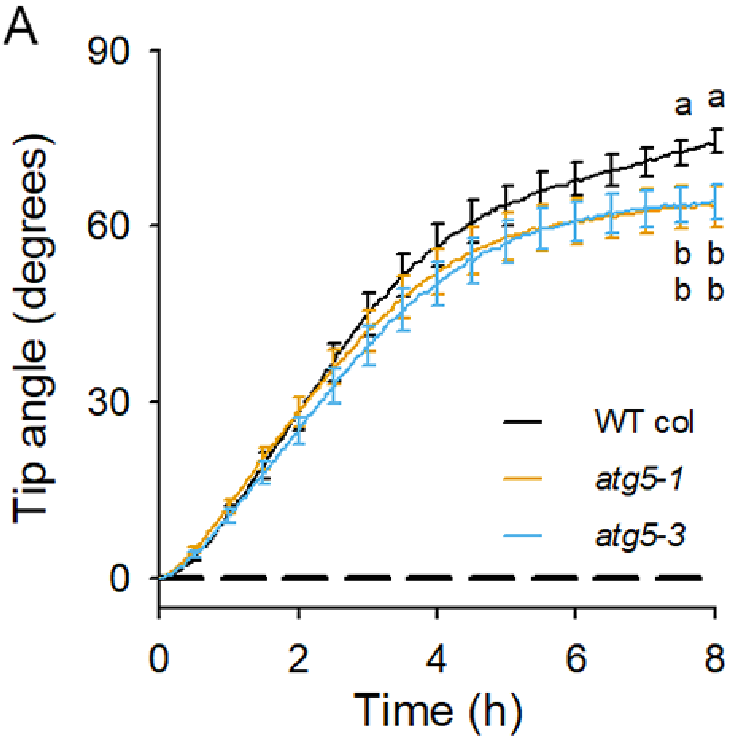
We went from 10^3 to 10^0 candidates

Maize QTL	Maize gene	Arabidopsis QTL	Arabidopsis gene	Identity (%)	Gene name	Gravitropism phenotype
Chr1-2	Zm00001eb025170	4@40.3	At4g15130	61	CCT2	yes
Chr5-1	Zm00001eb216660	1@64	At1g21980	64	PIP5K1	no
Chr5-2	Zm00001eb230080	5@61	At5g17290	61	ATG5	yes
Chr5-2	Zm00001eb230070	5@61	At5g17310	84	UGP2	yes
Chr8-2	Zm00001eb344960	3@17	At3g24140	48	FAMA	yes
Chr10-1	Zm00001eb423930	1@64	At1g22275	41	ZYP1	not tested
Chr10-2	Zm00001eb431340	5@61	At5g16860	55	none	not tested

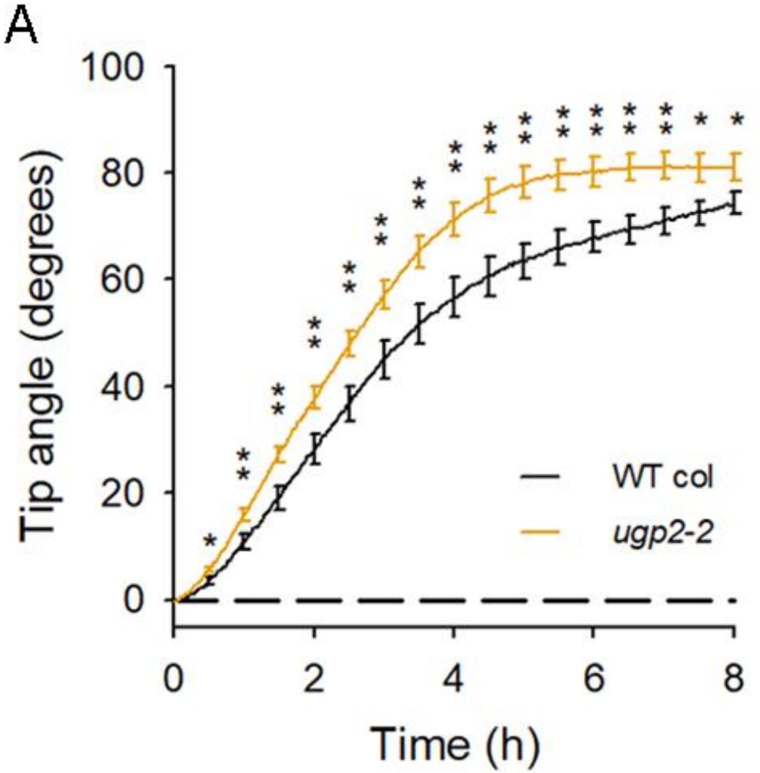
Table 1. Bidirectional best hit genes residing within Arabidopsis and maize gravitropism QTL.

These results indicate that we identified new gravitropism genes

ATG5



UGP2



The value of two large-scale experiments was much greater than the sum of their separate values

Value (A,B) > ValueA + ValueB

Automated computing (and data acquisition) makes these large-scale experiments possible.