NSF Advisory Committee 2016

Table of Contents

Click Page to Navigate

Meeting Agenda

Tuesday, December 20th

- 8:30-9:30am "Analysis of genes affecting plant regeneration and transformation in poplar" (Steve Strauss)
- 9:30-10:00am "High throughput culture and imaging plans" (Anna Magnuson)
- 10:00-10:30am "High throughput fluorescence imaging" (Chris Draves of Middleton Spectral Vision)

BREAK

- 11:00-11:45am "Culture medium and transformation plans" (Amy Klocko lead, with Cathleen Ma and Steve Strauss)
- 11:45-12:00pm "Epigenetic regulators of regeneration and transformation (RT) for expanding the reach of GWAS" (Michael Nagle)
- 12:00-1:00pm "Social science and education/outreach plans" (Troy Hall lead, with Jay Well)

Thursday, December 22nd

- 8:30-9:30am "Analysis of genes affecting plant regeneration and transformation in poplar" (Steve Strauss, repeat presentation)
- 9:30-10:30am "Image analysis system plans" (Fuxin Li)

BREAK

- 11:00-12:00pm "Data visualization and statistical analysis" (Yuan Jiang)
- 12:00-1:00pm "GWAS approach, experience, plans" Directed by Wellington Muchero

Home

Steve Strauss presents…

"Analysis of Genes Affecting Plant Regeneration and Transformation in Poplar"

Analysis of genes affecting plant regeneration and transformation in poplar

Steve Strauss, PI

Kickoff Advisory Meeting / December 2016

Agenda

- Check-in, introductions and logistics
- Advisory meeting goals and agenda review
- Science goals
- People and organizational plans
- Work flow and budget overview
- Major near-term activities underway and planned
- Risk analysis

Meeting goals

- To present and discuss our plans, and solicit critiques and suggestions for improvement
- To discuss emerging opportunities, new technologies, and new ideas prior to project initiation
- **Output**: Summary report of criticisms, suggestions, planned actions

Advisory committee meeting agenda

- **Tuesday, December 20th**
- 8:30-9:30am "Grant introduction and overview of plans, organization" (Steve Strauss)
- 9:30-10:30am "High throughput culture and imaging plans" (Anna Magnuson lead, Middleton Spectral)

BREAK

- 11:00-12:00am "Culture medium and transformation plans" (Amy Klocko lead, with Cathleen Ma, Michael Nagle and Steve Strauss)
- 12:00-1:00pm "Social science and education/outreach plans" (Troy Hall lead, with Jay Well)

Advisory committee meeting agenda

- **Thursday, December 22nd**
- 8:30-9:30am "Grant introduction and overview of plans, organization" (Steve Strauss)—for those of you who did not attend the first presentation
- 9:30-10:30am "Image analysis system plans" (Fuxin Li)

BREAK

- 11:00-12:00pm "Data visualization and statistical analysis" (Yuan Jiang)
- 12:00-1:00pm "GWAS approach, experience, plans" (Wellington Muchero)

Agenda

- Check-in, introductions and logistics
- Advisory meeting goals and agenda review
- Science goals
- People and organizational plans
- Work flow and budget overview
- Major near-term activities underway and planned
- Risk analysis

Basic science ideas behind work - 1

- The capacity for regeneration of transgenic plants (aka "transformation" or "RT") remains a major obstacle to broad, low cost use of transgenic methods for research and biotechnology
- Little is know about why species and genotypes vary so widely in their amenability to transformation
- The ability to accurately phenotype plants during RT is a major barrier to understanding and analysis
  - Developments in imaging and image analysis may be game changers

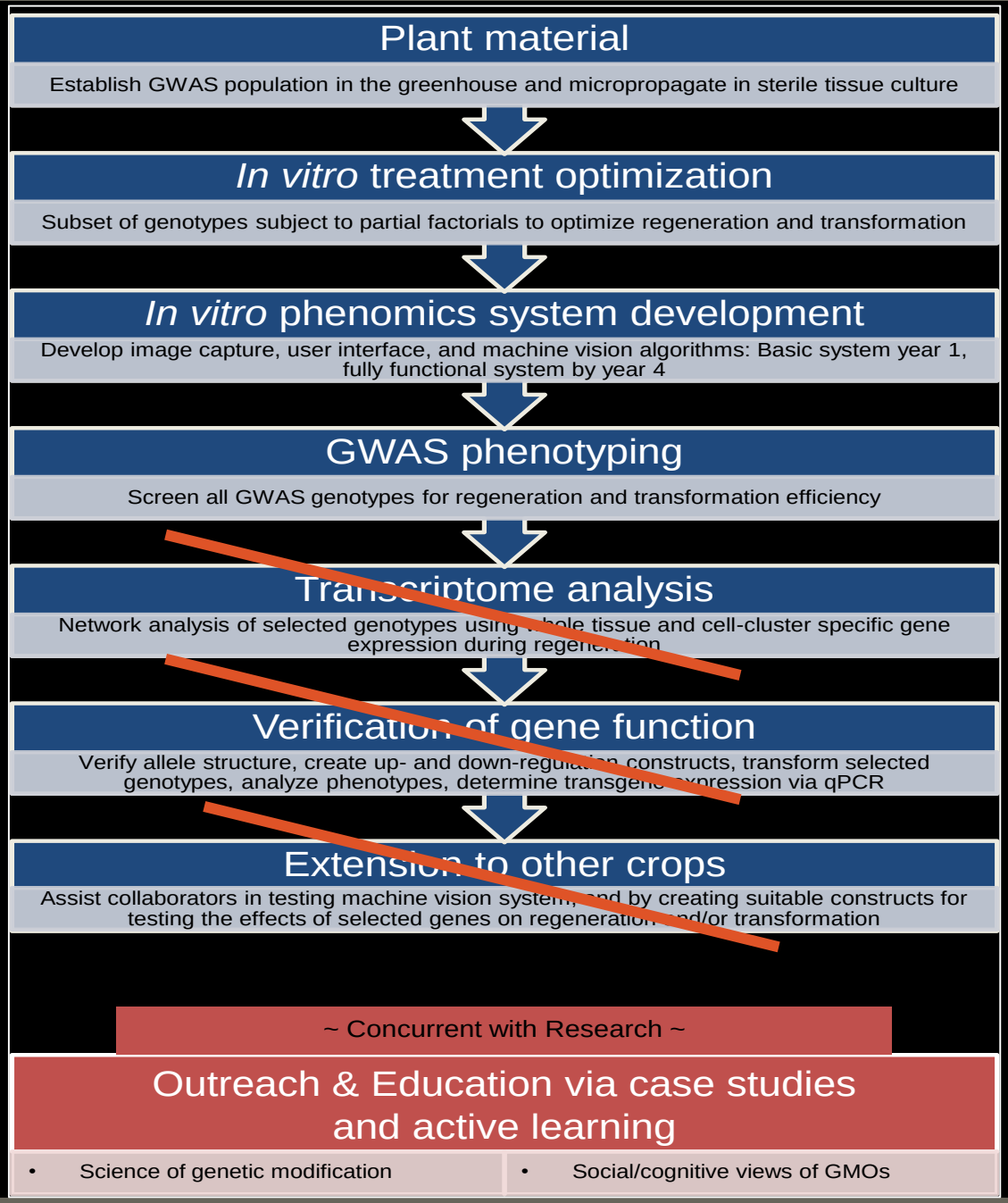
Basic science ideas behind work - 2

- Poplars are good model systems due to their extensive in vitro biology, and genomic resources
  - Reference genome, resequenced association population, low LD among wild trees
- GWAS may enable genes that control various part of the RT process to be identified, and thus the relevant physiological processes inferred, further studied, and the genes possibly employed as reagents to improve RT
- Cognitive approaches to education and outreach may empower teachers and students to better understand—and thus make better decisions as citizens, activists, and professionals—about complex GMO issues

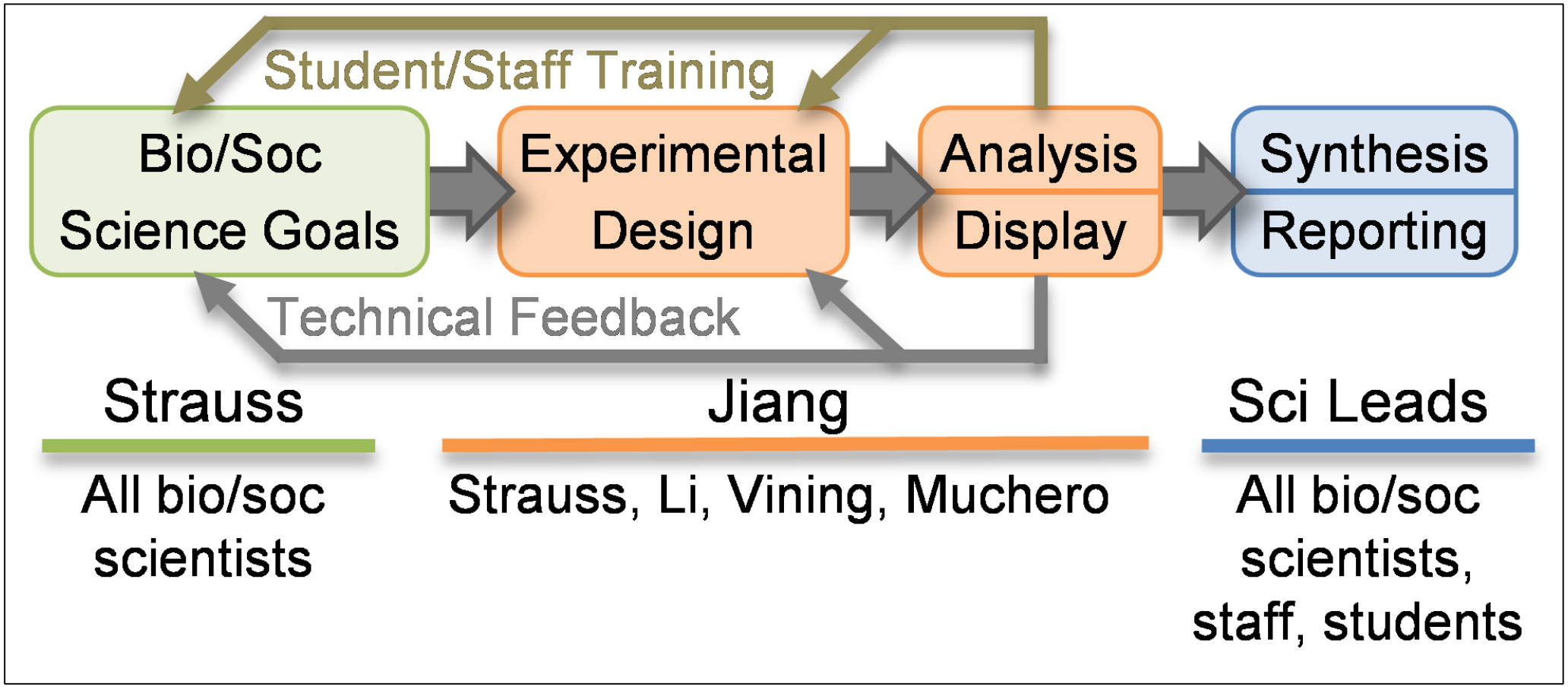
Project objectives in brief

- Explore a variety of RT methods to maximize variation in RT responses
- Develop new phenomic tools, including an image capture and generalizable machine-vision system, to precisely determine in vitro phenotypes
- Using GWAS, map sets of alleles that are associated with variation in RT frequency
- Study cognitive processes with respect to GE crops, develop case studies and new teaching materials, and deliver them to rural and underserved communities in the Pacific Northwest

Sequence of activities



Emphasis on statistically valid design and analysis – and training thereof



Major changes from submitted proposal - 1

- 40% budget reduction (7 M to 4 M)
- Reduced social science / education (5 years to 3 years)
- Reduced machine vision (4 years to 3 years)
- Reduced student help to carry out greenhouse/in vitro work
- Removal of candidate gene characterization/confirmation
- Removal of transcriptome study of regeneration
- Removal of work to extend RT genes and image system to aid transformation of other crop species

Major changes from submitted proposal - 2

Realizations about scale and biological complexity of work

- Availability of additional resequenced wild cottonwood genotypes (from ~1,000 to 1,300)
- Need to work with as uniform starting materials as possible (post dormancy best?)
- Screen for specific regeneration behaviors and do GWAS on subpopulations
- Cannot maintain many hundreds of genotypes in vitro, thus must use sterilized greenhouse (or possibly field) materials
- Need to better and systematically explore in vitro conditions for poplar (expanded in vitro optimization from 1 year to 2 years)
- Manpower is limiting given cuts, new realities

Agenda

- Check-in, introductions and logistics
- Advisory meeting goals and agenda review
- Science goals
- People and organizational plans
- Work flow and budget overview
- Major near-term activities underway and planned
- Risk analysis

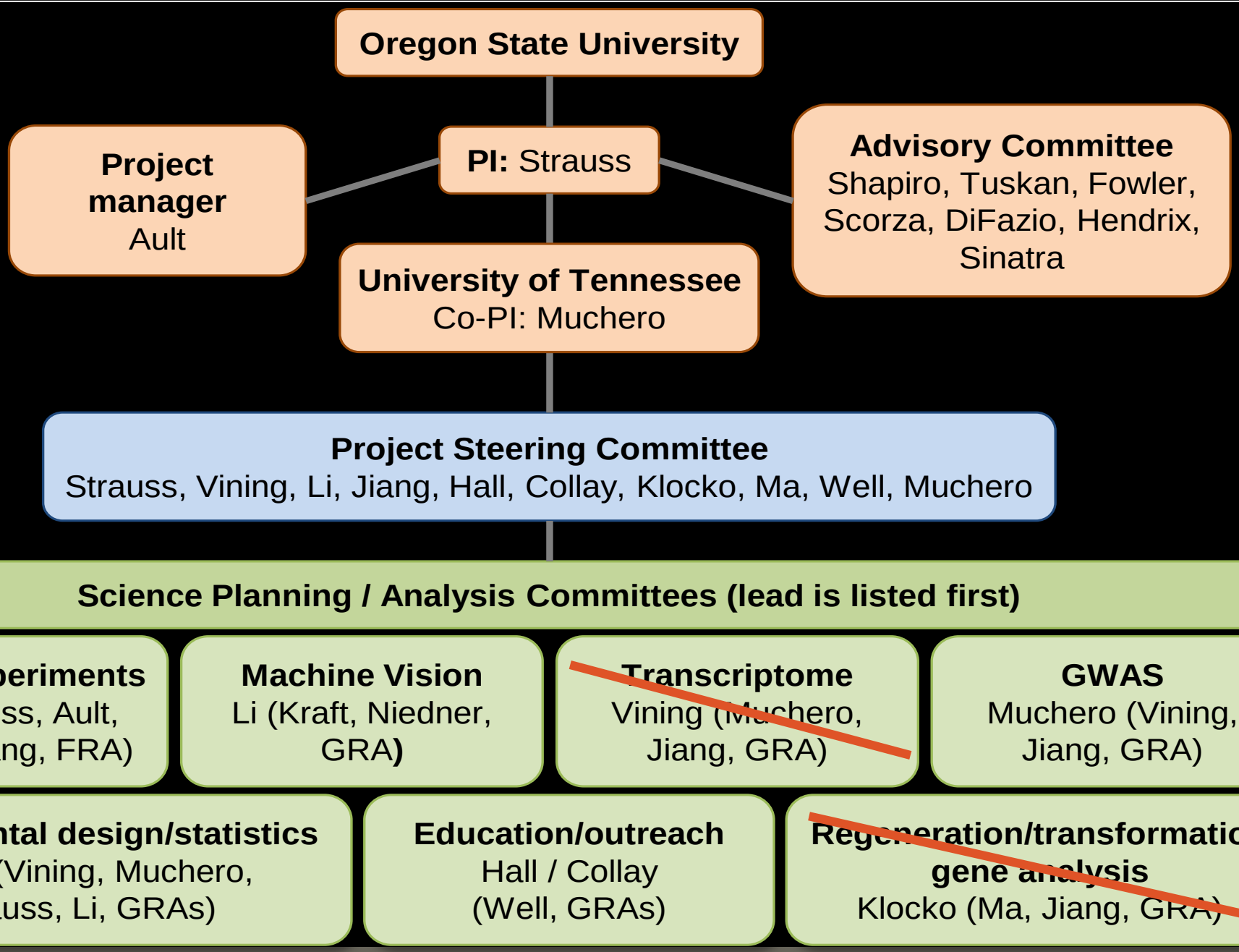
Leadership staff

Last Name	First Name	Role	Institution	Email
Strauss	Steven	PI	OSU	Steve.Strauss@OregonState.Edu
Li	Fuxin	coPI	OSU	Fuxin.Li@OregonState.Edu
Hall	Troy	coPI	OSU	Troy.Hall@OregonState.Edu
Jiang	Yuan	coPI	OSU	Yuan.Jiang@OregonState.Edu
Well	Jay	Key Pers.	OSU	Jay.Well@OregonState.Edu
Muchero	Wellington	Key Pers.	Univ. TN	mucherow@ornl.gov
Klocko	**Amy**	**Key Pers.**	**OSU**	**Amy.Klocko@OregonState.Edu**

Advisory committee

a. Tuskan – GWAS and poplar biology
   http://www.esd.ornl.gov/PGG/tuskan_bio.htm
b. Scorza – crop *in vitro* biology and transformation
   http://www.ars.usda.gov/pandp/people/people.htm?personid=5036
c. Hendrix – transcriptome, network analysis, non-coding RNAs
   http://biochem.science.oregonstate.edu/People/david-hendrix
d. Fowler – plant developmental and cellular biology
   http://bpp.oregonstate.edu/fowler
e. Shapiro – machine vision
   http://homes.cs.washington.edu/~shapiro/
f. Sinatra – science education
   http://rossier.usc.edu/faculty-and-research/directories/a-z/profile/?id=35
g. Gordon-Kamm, Pioneer/DuPont/Dow, In vitro regeneration
   https://www.researchgate.net/profile/William_Gordon-Kamm
h. Wayne Parrott – In vitro regeneration/transformation, biotech outreach
   http://parrottlab.uga.edu/parrottlab/

Org chart
(as submitted)



Outreach collaborations

- Transgenic materials and accompanying technical information for case studies / laboratory exercises

- **Simplot** – Innate potatoes (Doug Cole)
- **Monsanto** – diverse crops they produce (herbicide resistant soy, insect resistant corn likely candidates) (Bill Moar)
- **Pioneer/DuPont ?** – Gene edited *waxy* corn (Bill Gordon-Kamm)

NEWS
Five things to know today, including a landslide that will hamper NW

SPORTS BUSINESS
Custom sneakers inch closer with $333 Adidas 3D Runner

TECHNOLOGY
What Lattice CEO told employees about the pending acquisition

LOG
Nev
Por

MOST **POPULAR**

INDUSTRY NEWS › FOOD & LIFESTYLE

$4M OSU genetic engineering grant includes a Monsanto connection

Nov 22, 2016, 2:20pm PST **Updated** Nov 22, 2016, 3:02pm PST

INDUSTRIES & TAGS Food & Lifestyle

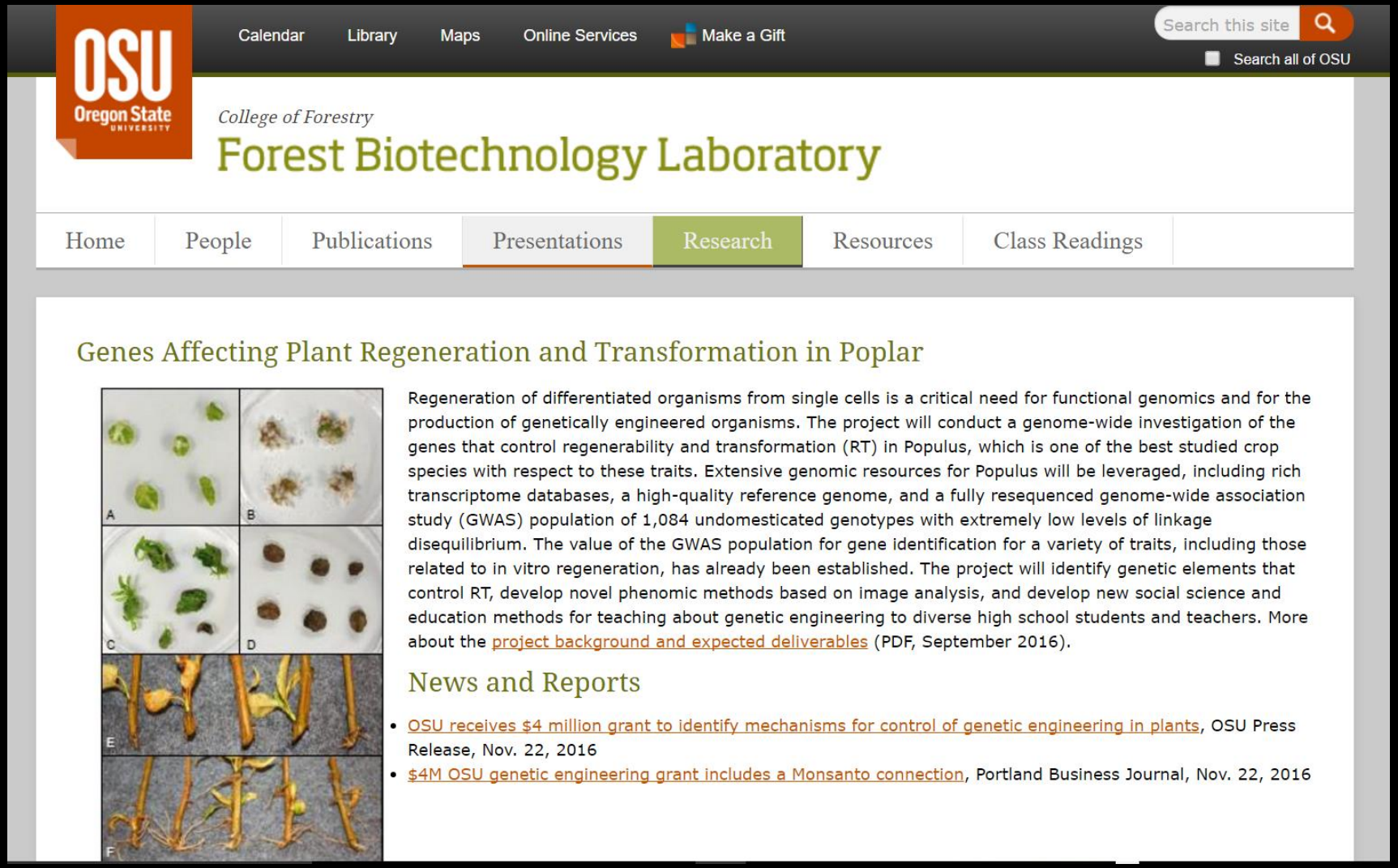
Communications plan - 1

- For general communications, frequent email, text message, and Skype among project participants.
- **Weekly meetings** between the PI and project manager on budgets and workflows, and PI/coPI lab review meetings.
- **Regular monthly meetings** with reports of progress and plans with all members of the Project Steering Committee and associated graduate students (Fig. A-1).
- **Shared cloud server** for exchanging files and analysis results (Box). Development of project database in this or other platforms.

Communications plan - 2

- **All experimental design, statistical analysis, and bioinformatics plans will be presented to and vetted by the quantitative sciences committee led by coI Jiang.**
  - **Jiang and the committee will also take part in analysis, suggest analysis techniques and software, and train students and staff.**
- A **project web site** will be used for informing the public about the project and project goals, methods, personnel, selected data, and making public communications available (conference talks and posters, education materials, links to SMILE materials).
  - **Twitter and Facebook accounts to follow**
- **Full day review and planning meetings at the start and midpoint** with the external advisory committee

Project web site



http://people.forestry.oregonstate.edu/steve-strauss/genes-affecting-plant-regeneration-and-transformation-poplar

Agenda

- Check-in, introductions and logistics
- Advisory meeting goals and agenda review
- Science goals
- People and organizational plans
- Work flow and budget overview
- Major near-term activities underway and planned
- Risk analysis

Work plan - 1

Table 1. Project management and deliverables

PI Strauss and the part-time program manager will take part in most elements thus contributions are not specifically identified in research activities

tasks	year 1	year 2	year 3	year 4	year 5
in vitro and greenhouse activities					
establish population mapping poplulations in greenhouse via rooting of cuttings	X				
screen population for in vivo traits (rooting, shooting, callus)	X	X			
establish maping populations in vitro from greenhouse collections		X	X		
create epigenetic reprogramming constructs for transient expression	X				
in vitro treatment optimization (~6 genotypes) (hormones, Agro, timing, explant types, epigen transgenes, etc)	X	X			
rapid GWAS poplulation screens for optimization responses, identification of subpopulations		X			
screen GWAS for regeneration and transformation phenotypes			X	X	X

Work plan - 2

tasks	year 1	year 2	year 3	year 4	year 5
in vitro phenomics system development					
develop and refine image capture system	X	X	X		
develop and refine user interface	X	X	X		
develop and refine machine vision algorithm	X	X	X		
screen subpopulations (concurrent with system development)	X	X	X		
present imaging system at P4G conference and in publication-s			X		

Work plan - 3

tasks	year 1	year 2	year 3	year 4	year 5
data analysis and publication					
establish experimental design and data statistical analysis pipelines for optimization studies (Huang)	X				
analysis of optimization data	X	X	X		
establish analysis pipelines for GWAS data (Huang/Muchero)		X	X	X	
GWAS analysis of in vivo rooting, callus, adventitious shoots, transformation			X	X	X
bioinformatic analysis of in vivo and in vitro trait associations			X	X	X
overlay candidate SNPs onto poplar gene network			X	X	X
identification of key SNPs in network and functional interpretations			X	X	X
analysis of output from optimization, publication of results	X	X	X		
publication of GWAS results			X	X	X

Work plan - 4

tasks	year 1	year 2	year 3	year 4	year 5	
social science and outreach						
Audience assessment of K12 students and teachers		X				
Background curriculum, case study development		X	X			
Curriculum delivered to SMILE teachers, feedback		X	X	X		
Social Science GRA/coI studies, assesses workshops		X	X	X		
SMILE teachers deliver curriculum to Math/Science clubs		X	X	X		
Social science GRA/coI studies, assesses SMILE clubs		X	X	X		
Culminating GMO activity at high school college connection				X		
Social science GRA/coI studies, assesses culminating activity				X		
Case study GMO curriculum delivered to urban classrooms		X	X			
Social science GRA/coI studies, assesses urban curriculum		X	X			
Pubication of survey and teaching results by social media and conferences			X	X		

Summary of project budget – people 1

PI/coPIs/Key – 0.6 to 3 months per year

- Strauss – 5 years (PI)
- Jiang – 5 years (statistics)
- Li – 3 years (machine vision)
- Hall – 2 years (social science)
- Well – 3 years (education/outreach)
- Muchero – 2 years (GWAS)

- Klocko / new postdoc (gene constructs and bioinformatics) – (1 year for 6 months, 2 years full-time near end)

Summary of project budget – people 2

- Technicians – all years
  - 2-3 months / year - Project manager, Anna Magnuson
  - 4-6 months/year – Cathleen Ma (25 years experience)
  - Full-time – new hire
- Graduate Research Assistant – 3 years each
  - Machine vision (first three years)
  - In vitro/GWAS (last three years)
  - Social science/education (middle three years)
- Student aides (undergraduate)
  - 5-30 months per year

Summary of project budget – major items - 1

- Growth chambers – 164 K
- Custom imaging system – 90 K
- Custom Petri dish transfer trays/system – 5-10 K
- Mediaclave and mediajet (large batch media prep/pour) – 42 K
- Centrifuge and ultra freezer – 18 K

Summary of project budget – major items - 2

- Services/supplies, 5 years – 262 K
- Participant support costs, 3 years – SMILE – 112 K
- Grad student tuition, 5 years – 184 K
- Personnel, 5 years – 2.0 M
- Indirects, 5 years – 1.1 M
- **Total cost, 5 years – 4.0 M**

Agenda

- Check-in, introductions and logistics
- Advisory meeting goals and agenda review
- Science goals
- People and organizational plans
- Work flow and budget overview
- Major near-term activities underway and planned
- Risk analysis

Major upcoming project activities – 1

- Set up web site - DONE
- Set up Twitter and Facebook accounts - Winter 2017
- Set up project databases and flows - ONGOING (Box sites established)
- Hiring of new transformation tech/Fac Res Assistant (UNDERWAY)
- Pre-launch, online advisory committee meetings - ONGOING
- Identify and order growth chambers - UNDERWAY
- Identify vendor, design and order imaging system – ONGOING
- Barcode systems and hardware purchase – ONGOING
- Identify / negotiate large scale media prep system- SPRING-SUMMER 2017

Major upcoming project activities – 1

- Movement of materials into greenhouse, including new genotypes - DONE, to be redone/expanded
- Planning and execution of in vitro optimization studies - ONGOING
- Transgene-perturbation constructs and experiments planned/initiated - UNDERWAY
- Preliminary in vivo rooting GWAS experiment - UNDERWAY
- In vivo shoot/callus response experiment - PLANNING
- Rapid screening for regeneration response, prelim GWAS experiment, using majority of GWAS population (up to 1,300 genotypes) - PLANNING

Agenda

- Check-in, introductions and logistics
- Advisory meeting goals and agenda review
- Science goals
- People and organizational plans
- Work flow and budget overview
- Major near-term activities underway and planned
- Risk analysis

Major project risks

- Reliable function of hyperspectral imaging system
- Efficiency of data/image transfers and organization/ease of use
- Dealing with hyperspectral data amounts, information extraction, interpretation by biologists
- Adequate staff and resources for planned optimization and GWAS work
- Difficulty in scheduling/delays to get collaborative work and training done
- Trimming goals to fit with staff and technical capacity, on the fly…
- How to best leverage education/outreach work for largest impact

Advisory committee meeting agenda

- **Thursday, December 22nd**
- 8:30-9:30am "Grant introduction and overview of plans, organization" (Steve Strauss)—for those of you who did not attend the first presentation
- 9:30-10:30am "Image analysis system plans" (Fuxin Li)

BREAK

- 11:00-12:00pm "Data visualization and statistical analysis" (Yuan Jiang)
- 12:00-1:00pm "GWAS approach, experience, plans" (Wellington Muchero)

Home

Anna Magnuson presents…

"High Throughput Culture and Image Plans"

ANNA MAGNUSON

High Throughput Culture and Image Plans

Work Flow

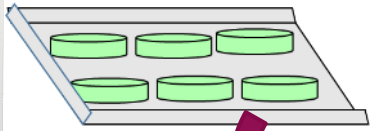
Field

Greenhouse

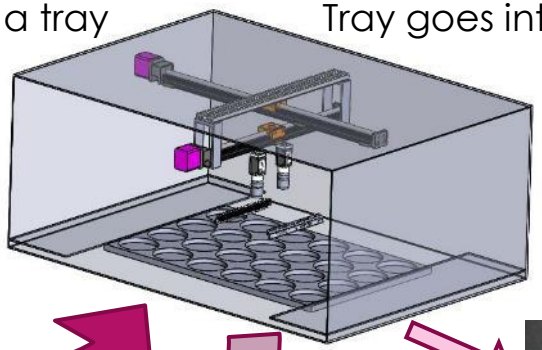
In vitro

Dishes stored in growth chambers

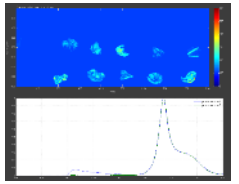
Transferred to a tray



Tray goes into imaging system

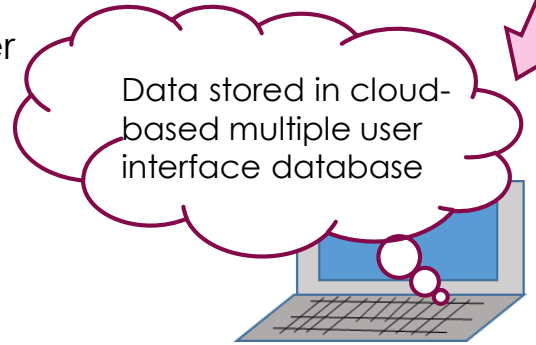


RGB image and spectral data output



Dishes get returned to chamber

Data stored in cloud-based multiple user interface database



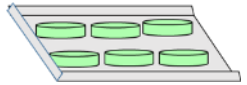
Designing the System

Logistics

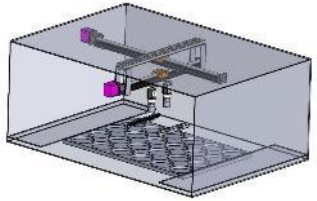
- Organizing Samples

- Large scale culture

- Efficient transfer



- Efficient imaging



- Managing and analyzing data

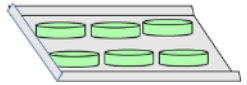
Designing the System

Logistics

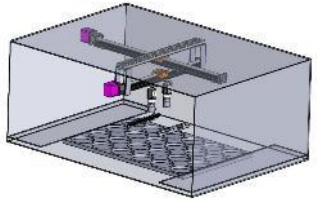
- Organizing Samples

- Large scale culture

- Efficient transfer

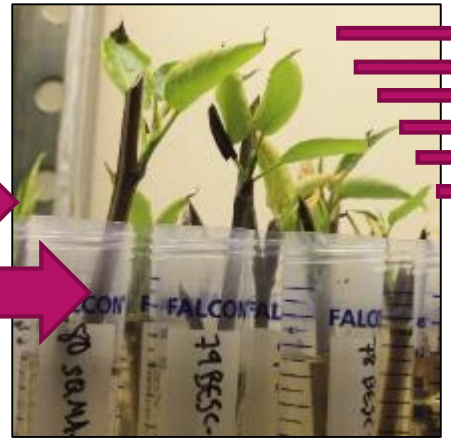


- Efficient imaging



- Managing and analyzing data

Organizing the Samples



- ~2,000 established GWAS trees currently in Benton county
- 4 cuttings/tree for 500 genotypes have already been taken for rooting in the greenhouse
- Of greenhouse cuttings, tissue is taken for culture work
- Each genotype will undergo a multitude of trials under different conditions with different media
- We're striving for as many as 2,000 genotypes, and 20 GWAS studies
- <u>Challenging numbers and goals!</u>

Organizing the Samples

- As sample numbers increase, room for mistakes also increase
- With such large scaling of samples, we need an efficient and organized way to track an experiment from start to finish
- With each petri dish that gets imaged, we need to know (and have easy access to):
  - Genotype
  - Construct info
  - Treatments
  - Conditions
  - Other?

Barcoding for Sample Management

- Currently we are envisioning a barcode system
- Barcodes could be read with a typical, cheap handheld barcode scanner—and labels easily printed from an excel spreadsheet through a barcode printer
- Can track materials from field to greenhouse to specific Petri dishes and experiments
- The barcodes would be affixed to the cover of each petri dish and contain information about the experiment – or a random code
- Barcodes would be visible in the RGB image produced by the imaging system
- The information would then be easily accessible and guaranteed matched to that sample both while the dish is in circulation in the chambers, as well as later one once we only have the images

Barcoding for Sample Management—Outstanding Questions

- Some outstanding questions:
  - Will we be able to use barcodes in the field and on greenhouse plants, as well as on petri dishes?
  - What information will each barcode contain?
  - Can we fit all the information we would like within each barcode? Or would we need to continue to supplement with handwritten notes?
  - Can we fit all the information on a barcode small enough to fit on the petri dish in a visible manner?
  - What problems have others encountered using similar systems?
- Your experience/views?

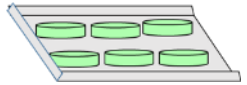
Designing the System

Logistics

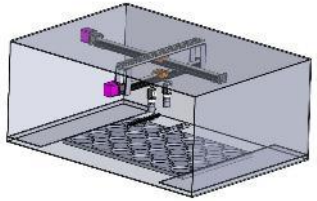
- Organizing Samples

- Large scale culture

- Efficient transfer



- Efficient imaging



- Managing and analyzing data

Large Scale Culture—What to consider?

- Efficiency when handling large numbers
  - Shelves easily accessible (not too tall)
  - Easy to get samples in/out—good shelf design
  - Need to fit in a lab as close to the tissue culture room as possible—same floor a must
  - Enough chambers so that our experiments aren't limited by lack of available shelf space
  - Easy data export
- Cutting risk of anticipated problems
  - Needs to prevent condensation on lids of dishes
  - Need to be compatible with other parts of the system (transferring trays) as to cut down on risk of dropping or mishandling samples
  - Needs to be reliable—has to have power source compatible with backup generator in lab

Large Scale Culture Requires Large Scale Culture Facilities

Minimum Requirements

- Proper airflow specifically designed to cut condensation
- Total growing area across all chambers must accommodate ~1,500 petri dishes—To accommodate 3 GWAS experiments at any given time
- Compatible with existing lab space
    - Run on single phase power
    - Minimal drainage requirements
    - Fit through doors
    - Minimal BTU output
- Programmable light cycles
- Control panel with data logging and export
- Good references

Chambers need airflow specifically designed to remove condensation from dishes

Petri dish clouded by condensation

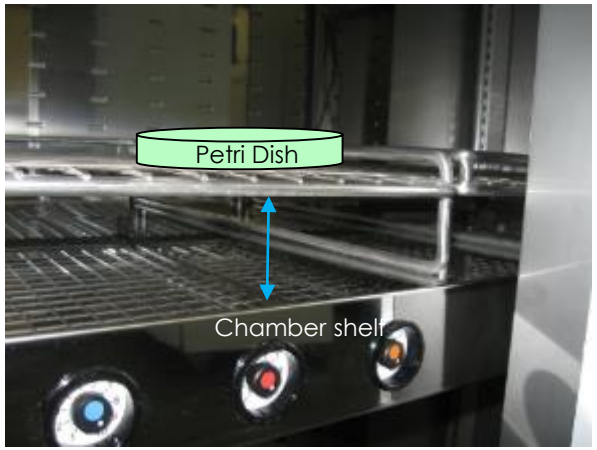
- To image efficiently and to ensure sterile conditions *in vitro*, there can be NO condensation buildup on the lids of petri dishes
- Dishes need to remain sterile (keep lid on) because we'll want to image the same dish multiple times
- Condensation builds up when a temperature differential exists between the top and the bottom of the dish. This happens when there is improper airflow and shelf design inside the chamber

The Problem

Typical growth chamber shelves have lights directly below the shelf, which heats up the underside of the petri dish—seen here

If the dishes sit directly on the shelf (seen here) then air only flows over the lids, causing the temperature differential and thus condensation

This company's solution: For extra charge, include trays in chamber that allow air to pass underneath, cooling bottom of dish



Designed for No Condensation

BioChambers Tissue Culture Chamber
Model TCC-74

Slats in shelf to allow air to flow upward onto the bottom of the dishes

Lights positioned well below bottom of the shelf to keep heat off the dishes and exposed to disperse heat evenly

6.0 AIR DISTRIBUTION

6.1 **Air Flow:** Vertical uniformly upward through an aluminum open channel floor to minimize or eliminate condensation in the tissue culture dishes.

6.2 **Fan Speed:** Adjustable from 60% to 100% (85% or higher recommended, temperature gradients increase at lower fan speeds). Fan speed can be programmed in the VNET controller enabling researchers to vary the airflow through the plants.

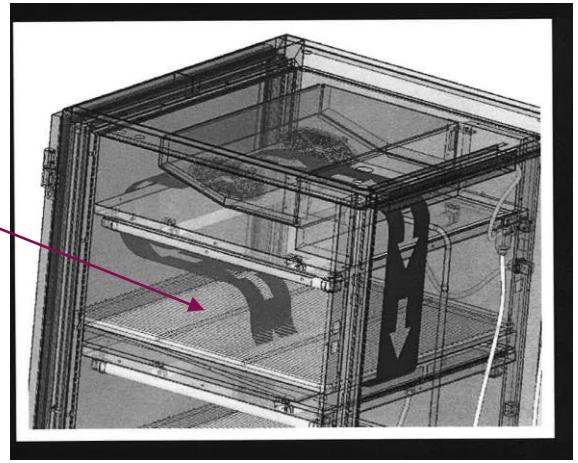
6.3 **Fresh Air:** Filtered fresh air with a manually adjustable vent: 1.7m^3/min (60ft^3/min).

Designed for No Condensation

Percival Scientific Tissue Culture Chamber

Air comes down the back of the chamber and gets dispersed evenly through a perforated diffuser below the shelf, then drawn upward underneath the dish

Again, note that the lights sit far below the shelf



Up close look at Percival shelf— And a positive review

Wire shelf design for maximum airflow— compatible with tray design

Perforated air diffuser, distributes air evenly beneath shelf for temperature control

A Percival customer was willing to show proof that his petri dishes showed no condensation

Comparisons and Other Considerations

	Percival	Weiss	BioChambers
Model	CU-41L5	SCG-120	TCC-74
Exterior Dimensions	41"W x 33.6" D x 77.2" H but needs 6" on all sides for vent.	57"W x 32"D x 78"H but needs 11" on left to vent, 6" on right.	100"W x 35"D x 111"H but can bring height down to 105"
Shelf Dimensions	36.3" x 27" / 922mm x 685mm	1320mm x 680mm	2463mm x 635mm
Shelf Style	Wire facing front to back, shelves can come out	Metal smooth top of light bank with grid wire shelf sitting on top for airflow	Aluminum horizontal slats
Number of chambers in quote	7	6	3
Drainage Requirements	Drains out bottom, reference says almost zero condensation, he has no floor drains and has never needed them	Two drains: condensation drain 5 ft above ground, tube out back of chamber for condensation. Almost no condensation. Bottom drain is in case of any internal water (plant tips over, humidity option running, etc.)	Floor drain needed, or small pump to pump water into sink if drainage becomes a problem.
Electrical Requirements	115 V single phase 60 Hz --comes with two ground cords + one 7 amp cord	230 V +/- 10% - 6%, single phase, 50/60 Hz, 13 amps--Mains protection required	120/208V three phase 60Hz

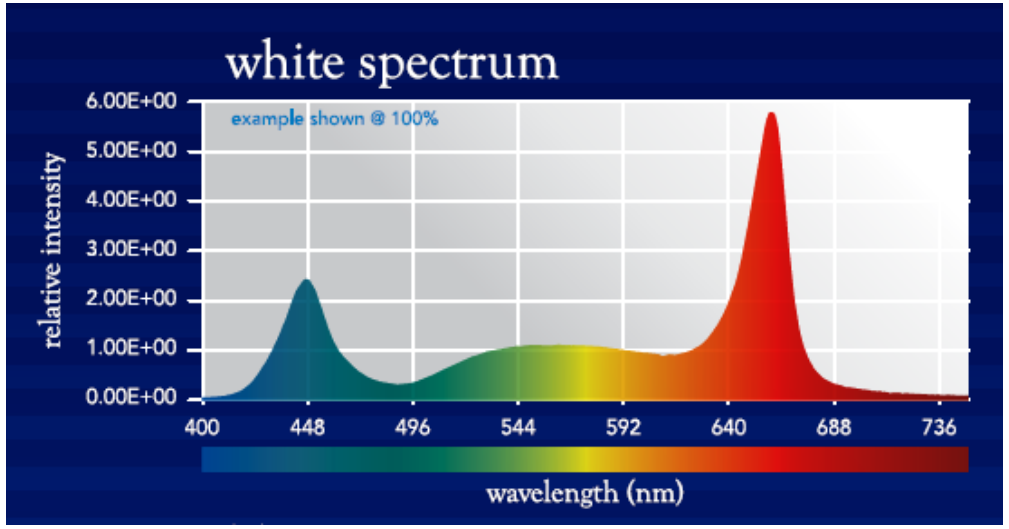
Comparisons and Other Considerations Cont.

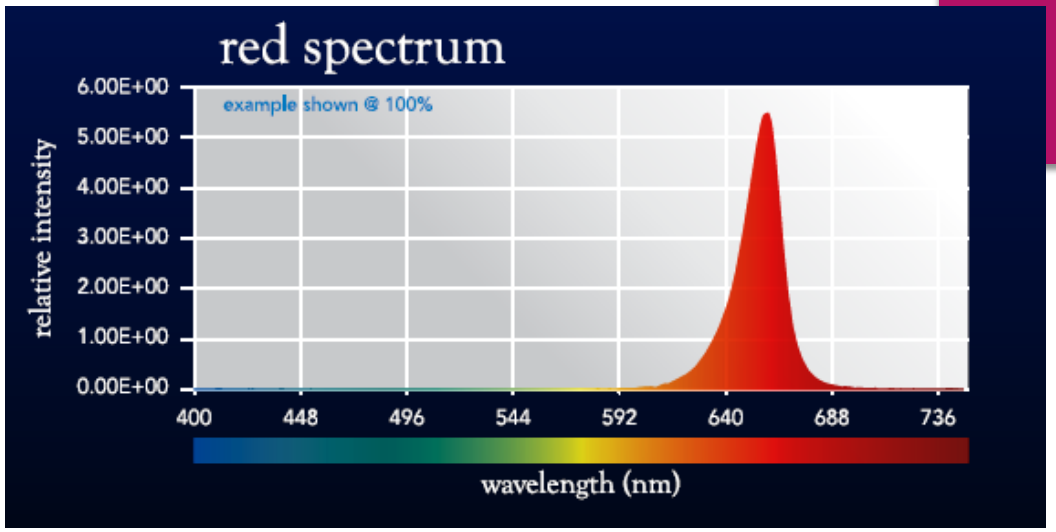
Light Intensity	220 umol/m2/s	100 umol/m2/s two bulb option, 140 three bulb option, much higher on LED option	400 umols/m2/s
Heat projection	3,400 BTU/hr and needs ceiling mounted evap coil	5,000 BTU/hr and CAN NOT operate at temps above 80 deg.--may need more duct work installed	3,500 BTU/hr-- can NOT operate at above 95 deg F.
Build Time	3 months	2 months	4 - 4.5 months
Control Panel	Basic but can be upgraded to log and export data	Keeps very basic temperature log, no export options	Keeps running data log that can be exported to a computer, needs ethernet connection to local area network. Two static IP addresses. Tons of control options, can do multi-day environmental change simulations.
References	Enthusiastic	Luke Warm	Enthusiastic
Warranty	2 years parts, labor, transportation	1 year parts and labor	2 years parts and labor, years 3-5 offer remote diagnostic services, five year warranty on VFD display.
Price Quote, shipping included	**$99,855**	Fluorescent light: $**140,580**, LED light: $**185,220**	**$161,934**

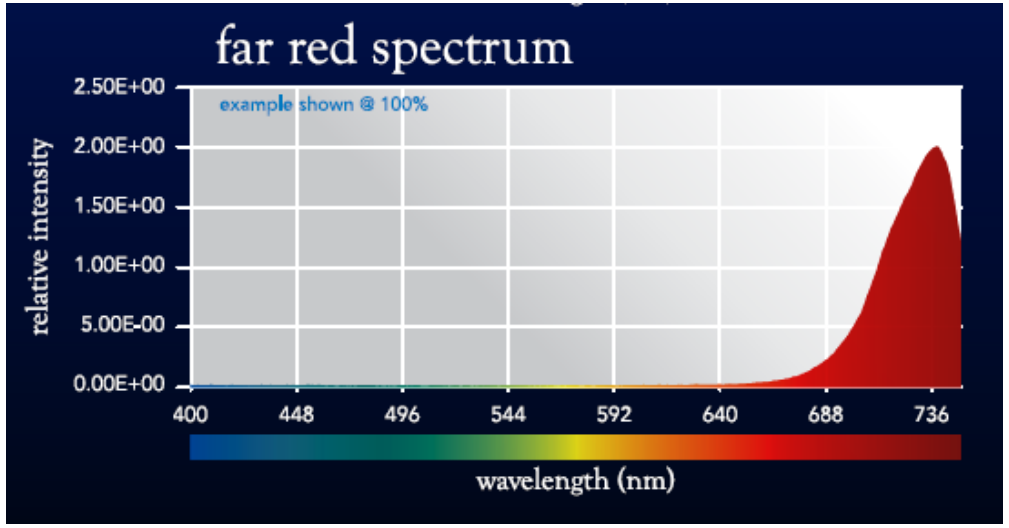
Interesting lighting option for chambers

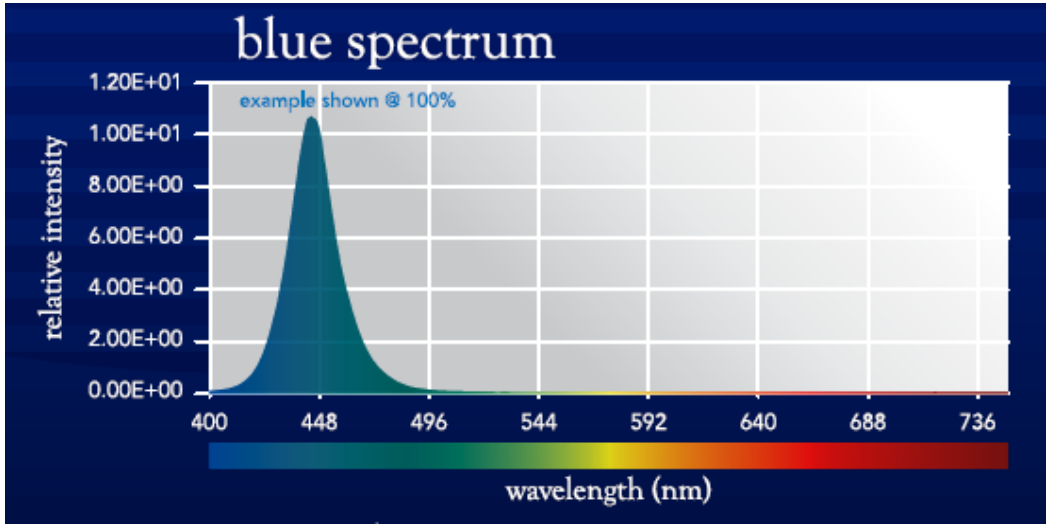
- Percival chambers have extra LED light options in several different spectrum ranges
  - White
  - Red
  - Far-red
  - Blue

- Each chamber has two banks of lights per shelf that can be controlled individually and scaled up/down from 10-100%.









Our current quote is for 3 LED chambers with the white+far red option, and 5 standard fluorescent lighting chambers. We chose white+far red because adjusting the far red light intensity has known effects on shoot elongation. Perhaps with this option we could stimulate shoot growth without transferring explants to special media, but just by adjusting the chamber settings.

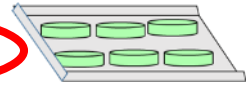
Designing the System

Logistics

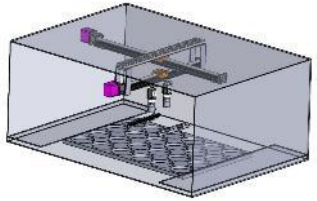
- Organizing Samples

- Large scale culture

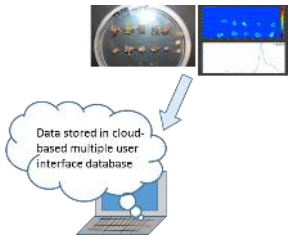
- Efficient transfer



- Efficient imaging



- Managing and analyzing data

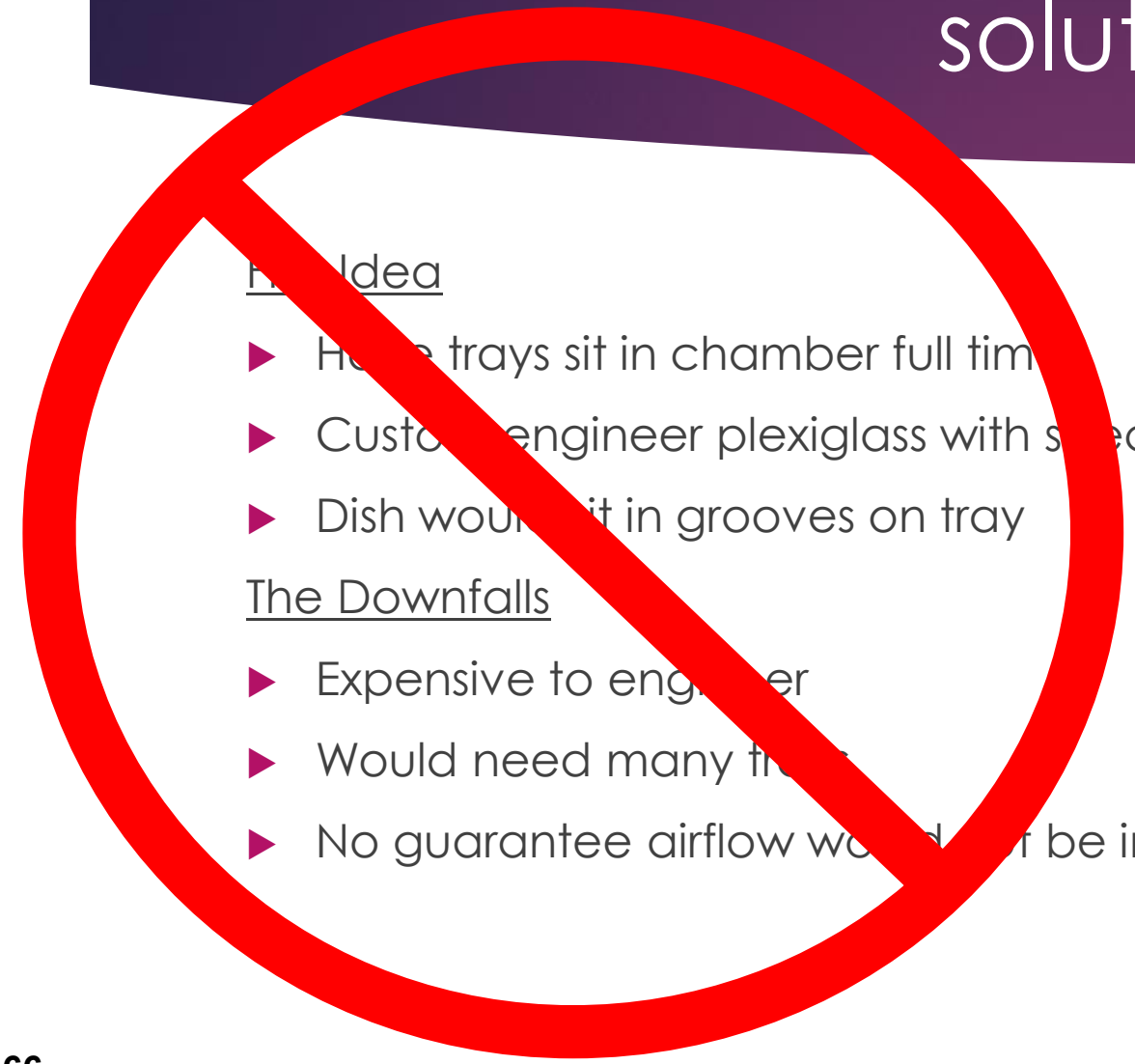


Efficient Transfer—How do we successfully get dishes into imaging system?

Considerations

- Can not inhibit airflow beneath dish in chamber
- Needs to be easily handled to reduce risk of dropping and mishandling samples
- Needs to hold a large enough number of samples to significantly increase efficiency
- Needs to be compatible with imaging system
- Can't inhibit images or image analysis

Efficient Transfer—Working towards a solution



<u>H[illegible] Idea</u>

- H[illegible]e trays sit in chamber full tim[illegible]
- Custo[illegible]engineer plexiglass with s[illegible]ecific size holes for airflow around dish
- Dish woul[illegible]it in grooves on tray

<u>The Downfalls</u>

- Expensive to eng[illegible]er
- Would need many tr[illegible]
- No guarantee airflow wo[illegible]d[illegible]t be inhibited

Efficient Transfer—Working towards a solution

Second Idea

- Trays made with metal mesh bottom and aluminum siding—much like a window screen
- Would sit permanently in chambers

The downfalls

- Tray not as secure
- Mesh background would not be compatible with image analysis

Efficient Transfer—Working towards a solution

Current Idea

- One tray, does not sit in chamber
- Feet hook on to shelf and tray lays flush to shelf
- Device to pull dishes onto tray, tray then locks dishes in

Outstanding considerations

- Material that has no reflection when imaged
- Material that doesn't cool dishes and cause immediate condensation
- Need prototype to test mechanics

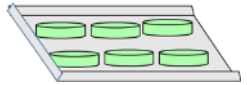
Designing the System

Logistics

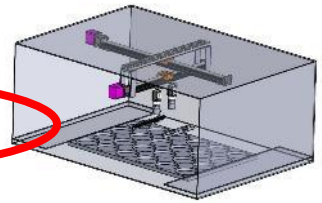
- Organizing Samples

- Large scale culture

- Efficient transfer



- Efficient imaging



- Managing and analyzing data

Efficient Imaging—How do we design the system?

- At first we began with a small company in Corvallis, we thought it would be something we could handle, design, and counsel ourselves
- We soon realized we needed more expertise, a company that had experience in advanced imaging and high throughput systems
- We've since been working with two companies, PhenoKey and Middleton Spectral Vision. You'll be hearing from Middleton shortly.
- Each company has been great to work with and has provided us with much guidance and brainstorming custom solutions to fit our needs

Efficient Imaging—What to consider?

- Speed
- Image precision and consistency
- Resolution
- Good user interface
- RBG and spectral data (including filtered versions of latter?), need to be output and stored separately for ready retrieval and analysis ?
- Compatible with cloud-based export for multiple user access and downstream data analysis

Efficient Imaging—science

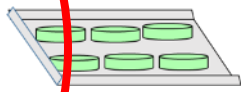
- Clear RGB image output compatible with software image analysis
  - Images must output into a common image format for multiple users
- Must capture/filter specific fluorescence signals for fluorescent protein reporters and stress indicators (GFP and other FPs, chlorophyll)
- High resolution images to be able to look at detailed tissue morphology to aid downstream interpretation, image analysis annotation
- Discovery algorithms as part of imaging software, and compatibility with machine vision and statistical analysis with same goals

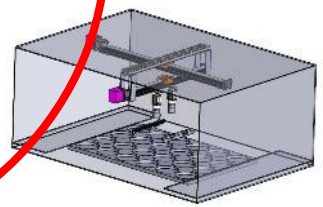
Middleton Spectral Vision will give a detailed walk through of their ideas and design capabilities

Designing the System

Logistics

- Organizing Samples
- Large scale culture
- Efficient transfer
- Efficient imaging
- Managing and analyzing data





Key Mechanics for Compatibility

- In order for everything to make sense, all parts of the system must be compatible
- This has meant considering the whole scope of work flow when making design decisions

- Tray design must be able to latch onto/work well with the chamber shelves, have access to all dishes, get dishes on/off easily
- Imaging system must accommodate tray shape/size
- Tray must be built to hold dishes is a way that is compatible with imaging system requirements (such as camera position)
- Tray can't interfere with chamber or imaging functions

A lot to consider

- How will the dishes sit in the chamber?
- Will the chamber have a door/shelf configuration that makes all the dishes accessible from straight forward?
- Does chamber have shelves that slide out?
- How many dishes will tray hold? Enough to maximize efficiency, but not more than easily handled
- How does the tray physically interact with the chamber?
- How does the tray get handled by the user? Handles? Separate moving parts?
- How does the tray fit in the imaging system?
- Will the tray materials be compatible with images/camera flash?

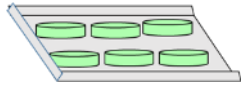
Designing the System

Logistics

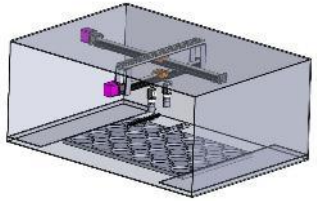
- Organizing Samples

- Large scale culture

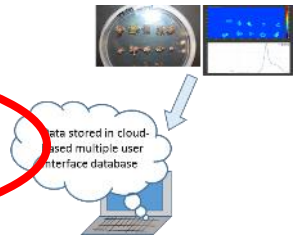
- Efficient transfer



- Efficient imaging



- Managing and analyzing data



Managing data

Many significant challenges

- Huge amounts of sequencing data for GWAS
- Thousands of RGB images
- Thousands of very large spectral data files
- Machine vision outputs
- Various analysis files
  - Statistical analysis and data display images
  - Large GWAS outputs

Managing Data—searching for Goldilocks solution

- Multiple databases
- Will be very large, can unlimited cloud based storage work? Will we need extra server space?
  - Forestry with plenty…..
- How to link databases together?
  - Can all genomic data remain separate or must be imported?
- How to make user interface as efficient as possible?
- How to design for efficiency? Multiple user coordination
- Budgeting for it – database manager not explicit in NSF budget

Major management activities Winter-Spring 2017

	Major management activities Winter-Spring 2017																			
	January				February				March				April				May			
Growth Chamber Purchases	X	X	X	X																
Chamber delivery and testing																	X	X		
Petri dish transfer system design/tested	X	X	X	X	X	X														
Imaging system design and purchase	X	X	X	X	X	X	X	X												
Data management system structure			X	X	X	X	X	X	X	X	X	X	X	X						
Barcode system tests and development	X	X	X	X	X	X	X	X												

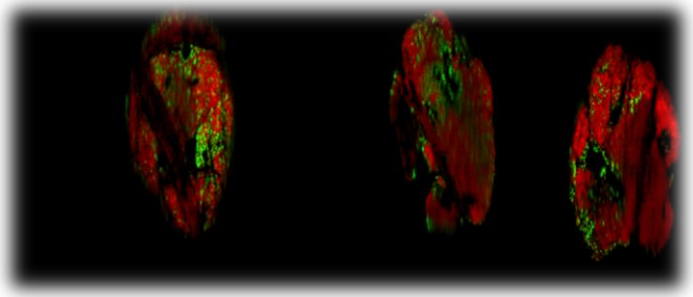
Home

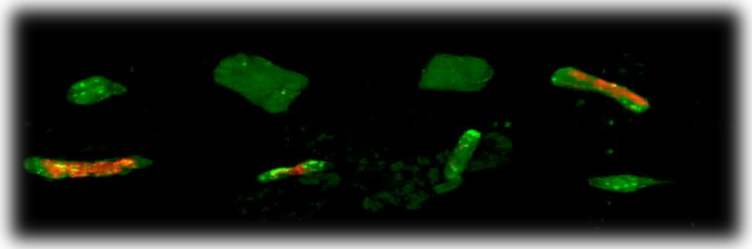
Chris Draves presents…

“High Throughput Fluorescence Imaging”

High Throughput Fluorescence Imaging

Chris Draves



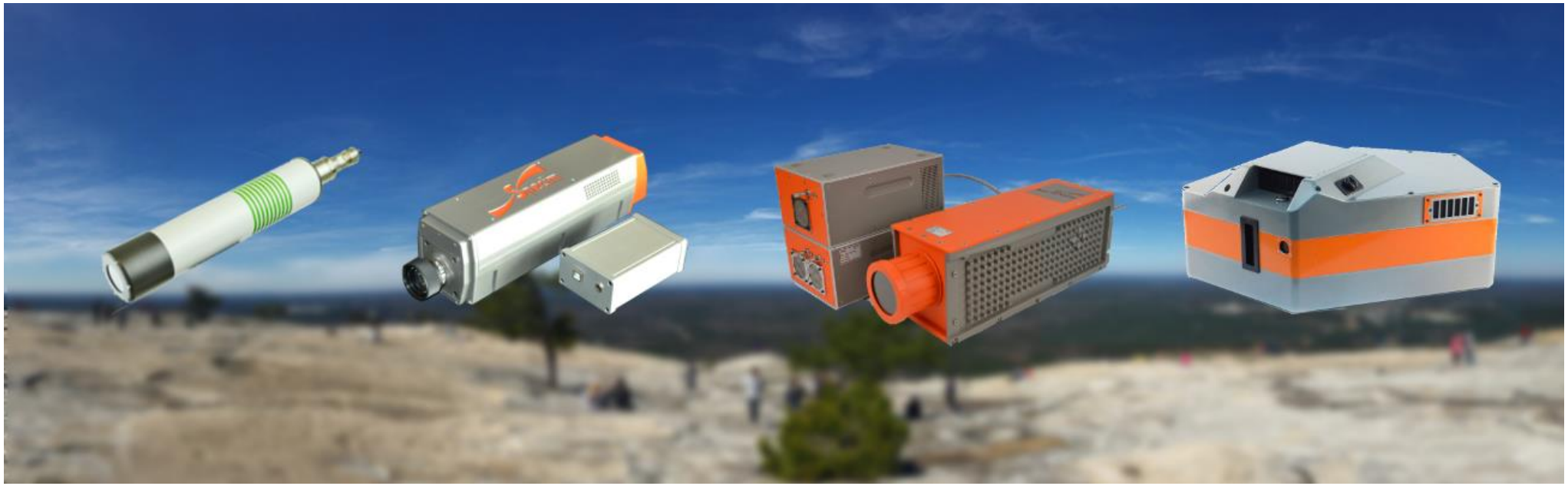


middleton SPECTRAL VISION

About Middleton

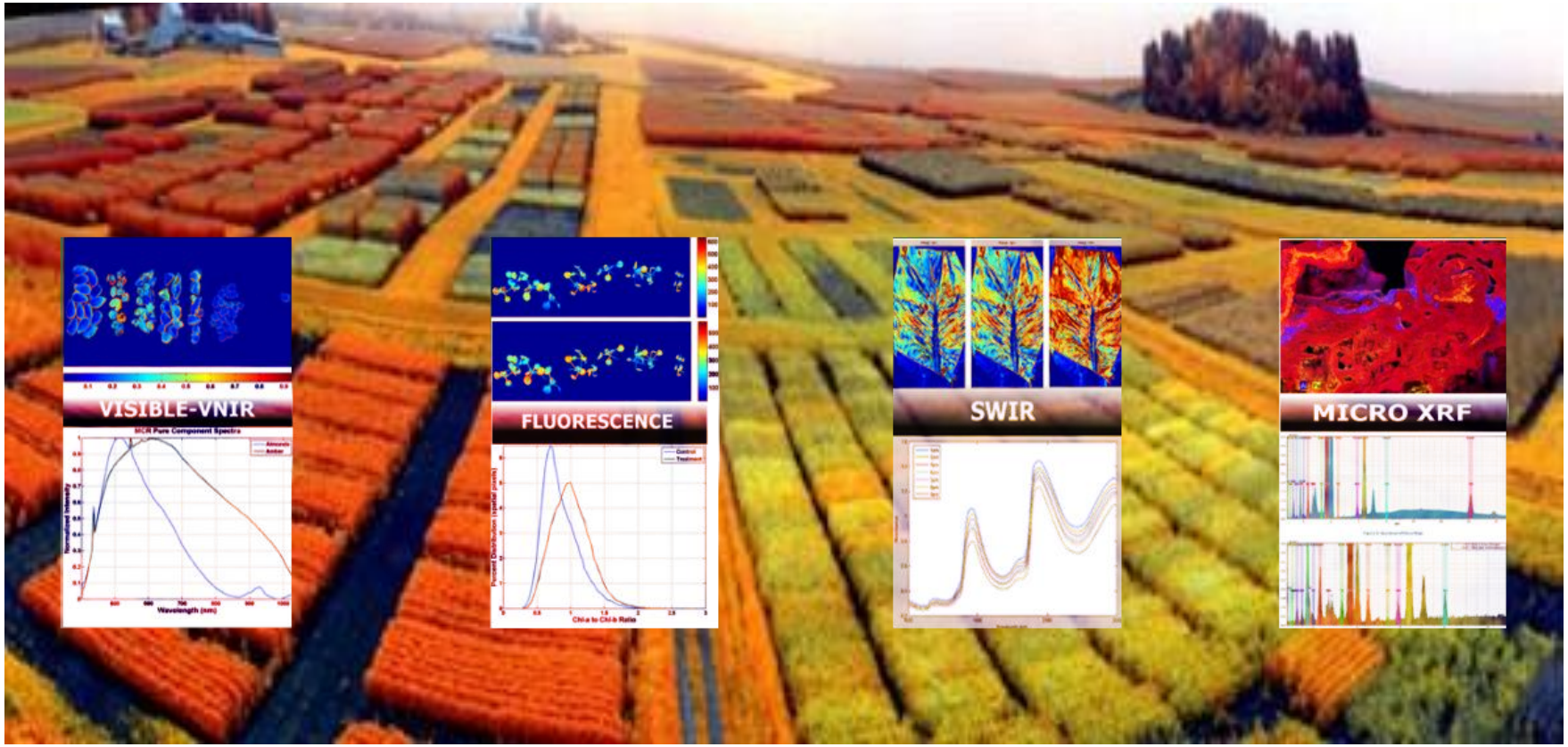
- The company is based in Middleton, WI, a suburb Madison, WI

Complete portfolio of spectral cameras



middleton SPECTRAL VISION

Complete portfolio of Imaging solutions

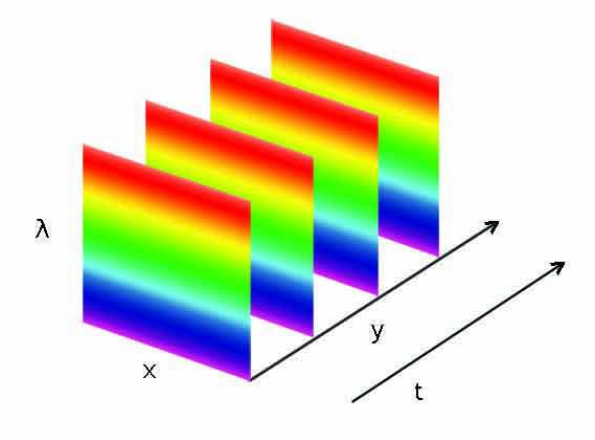


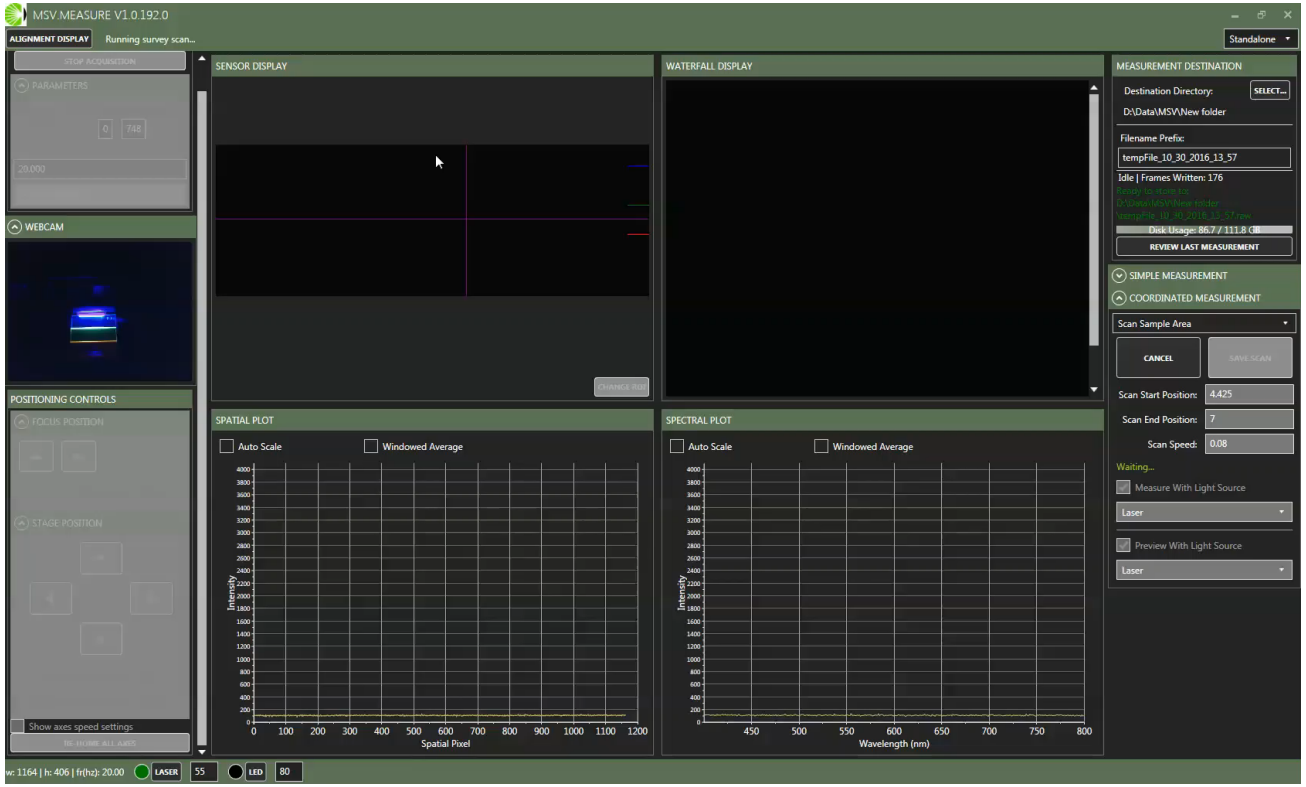
middleton SPECTRAL VISION

The technology – Hyperspectral + Fluorescence Imaging

What is hyperspectral imaging?

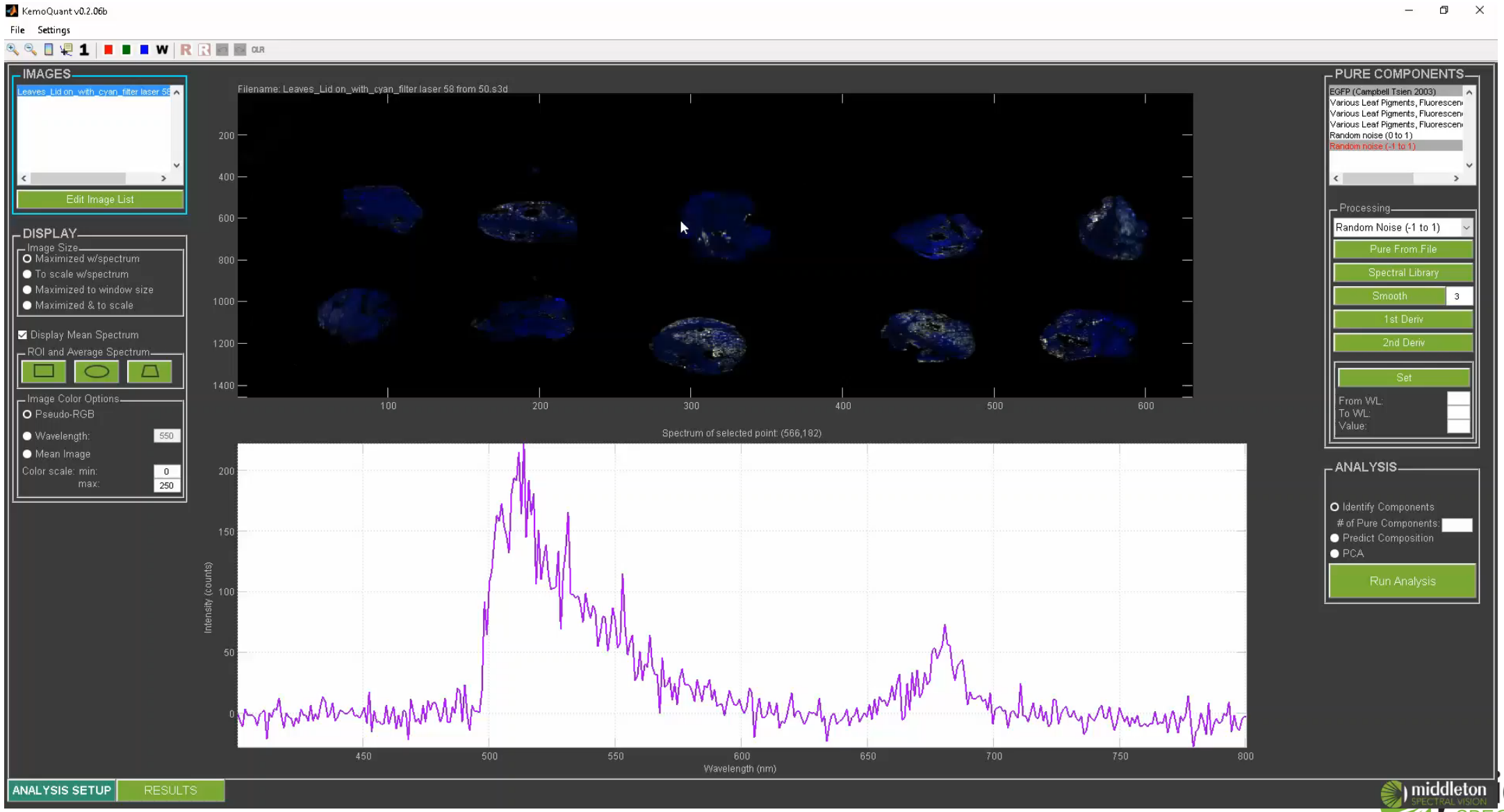
- Full spectral data simultaneously,
- with spatial line scanning over time



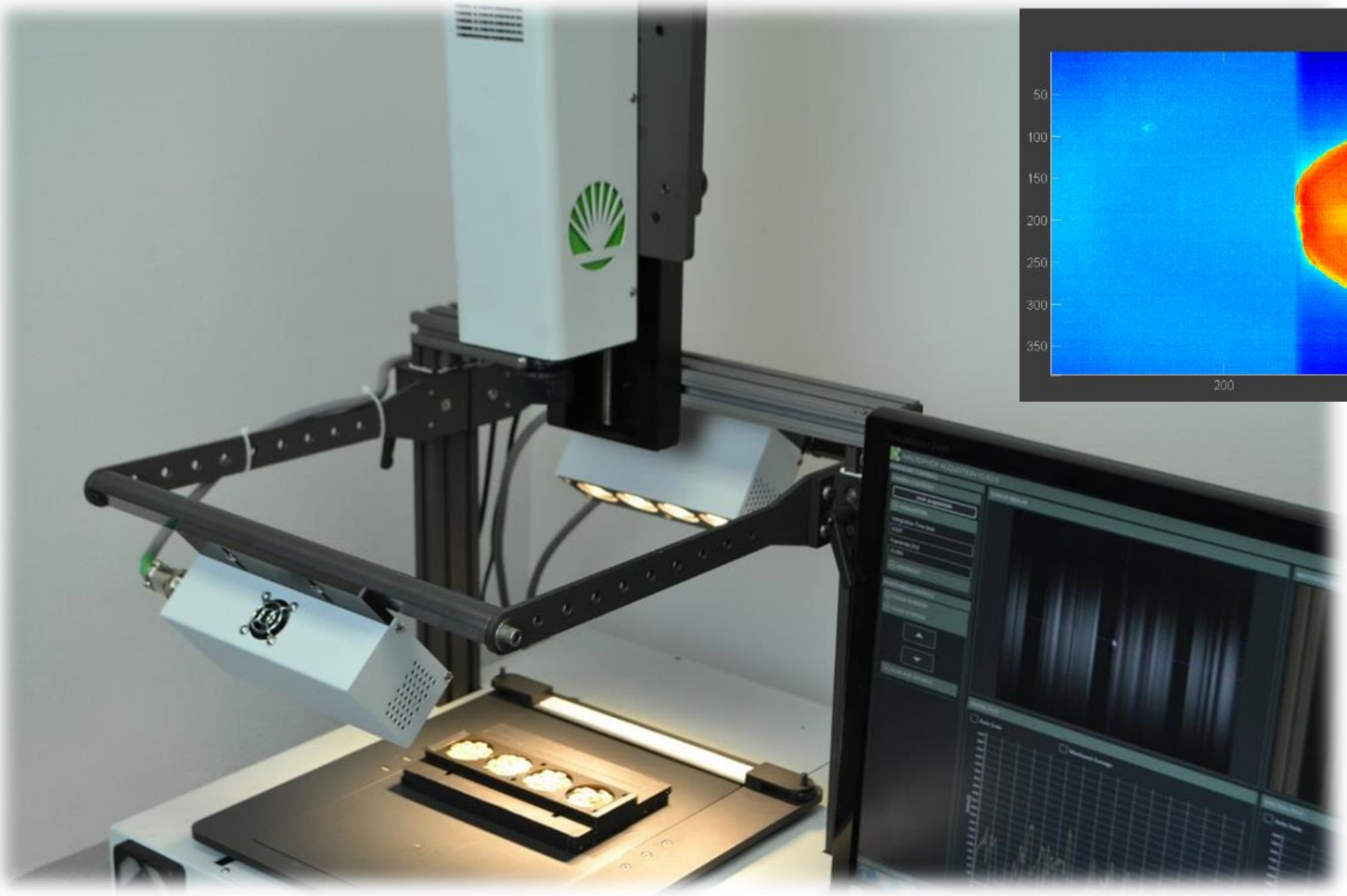


middleton SPECTRAL VISION

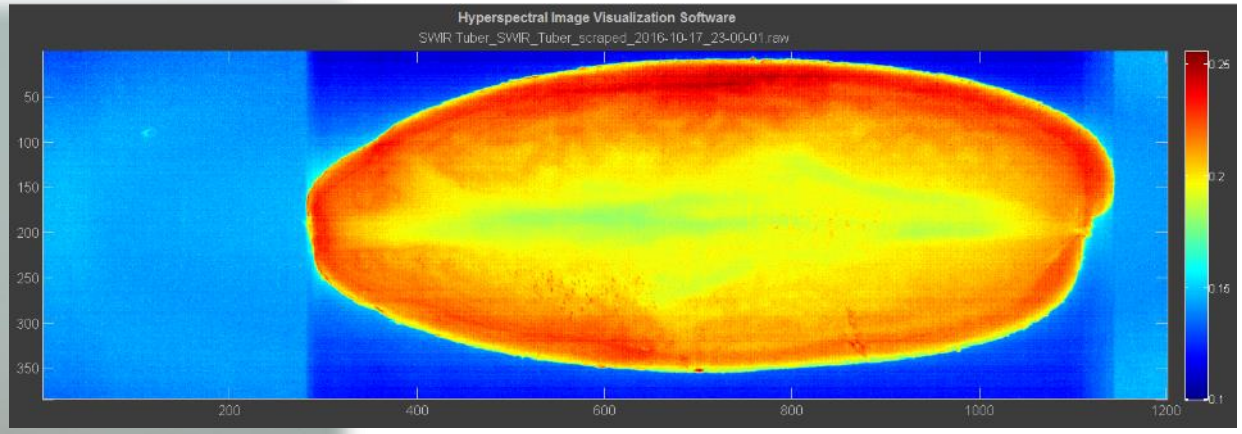
Working with a hyperspectral image



Plant Imaging Systems - Lab



ViaSpec II™ laboratory scanner



SWIR Image of a tuber

middleton SPECTRAL VISION

Multiscale Imaging System

ViaSpec™ Multiscale Plant Imaging System

middleton SPECTRAL VISION

ViaSpec™

middleton SPECTRAL VISION

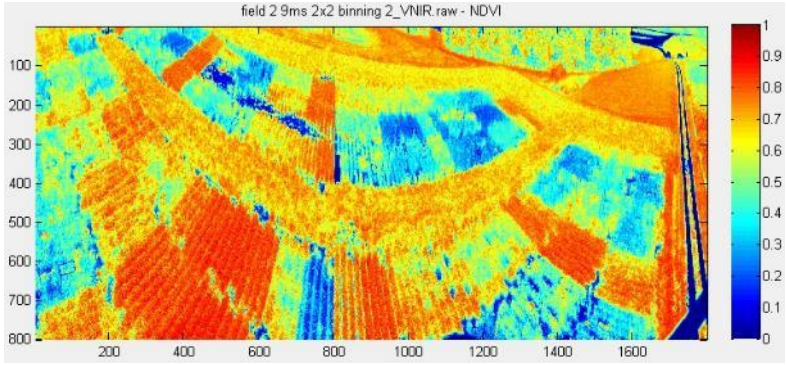
Remote sensing - Ground based hyperspectral solutions

3 camera set up
VNIR, NIR and Thermal

Fenix broad range (400-2500nm)with rotary scanner

middleton SPECTRAL VISION

Field Analysis software

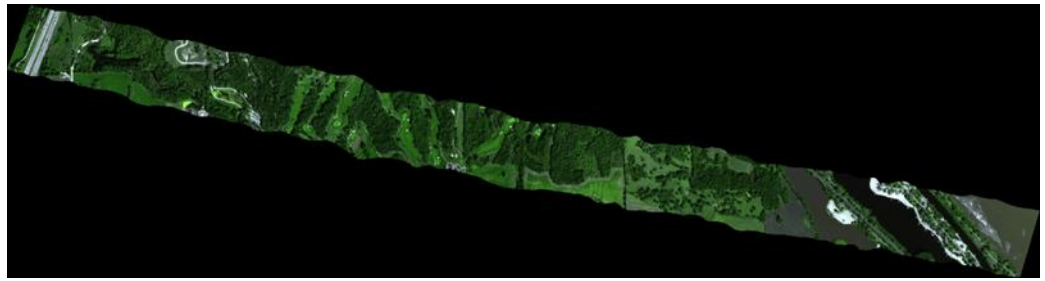


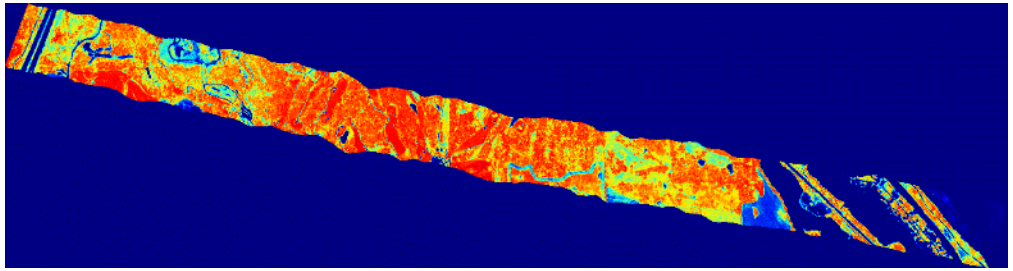
The software has several modes for visualizing the data, to assist in the analysis:

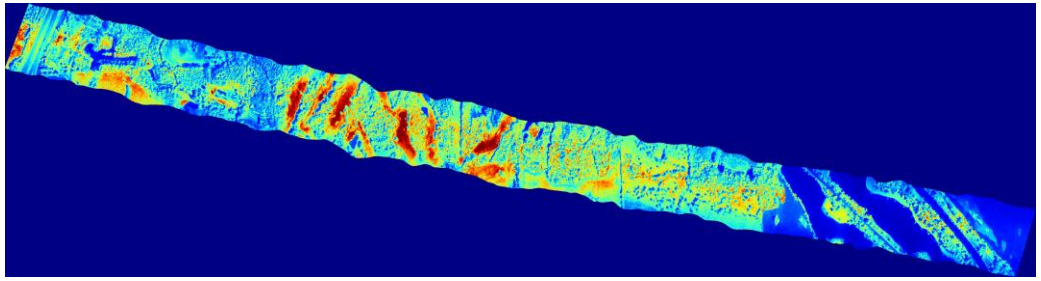
- raw & reflectance image views
- spectral display
- vegetation index image view
- slideshow of vegetation index to see changes over time
- 2-16 adjacent (tiled) vegetation index images for comparison
- plot of average vegetation index for field regions selected over
- time plot of environmental parameters over time

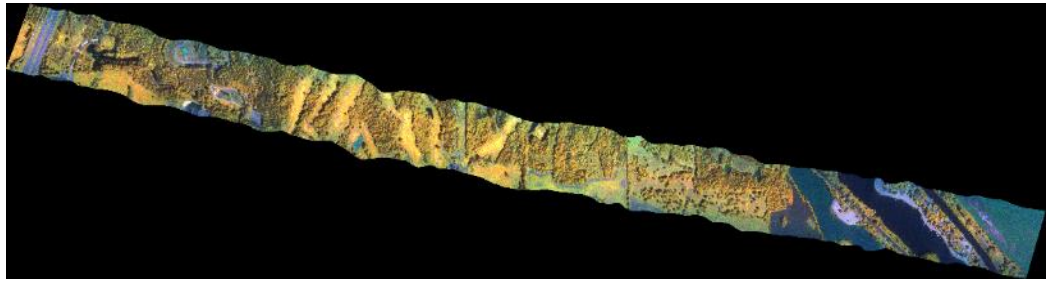
middleton SPECTRAL VISION

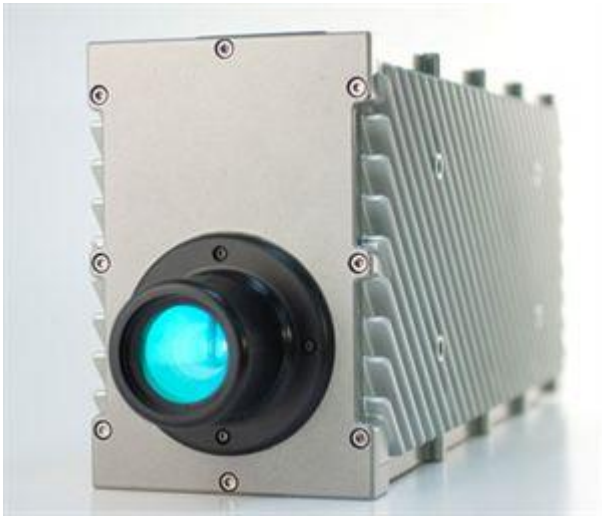
Remote sensing – Airborne











middleton SPECTRAL VISION

High through put phenotyping

High-throughput phenotyping in plant breeding

Annual phenotyping by imaging and other means:

Hundreds of thousands of plants (each big company)

Thousands of genes studied by transgenics

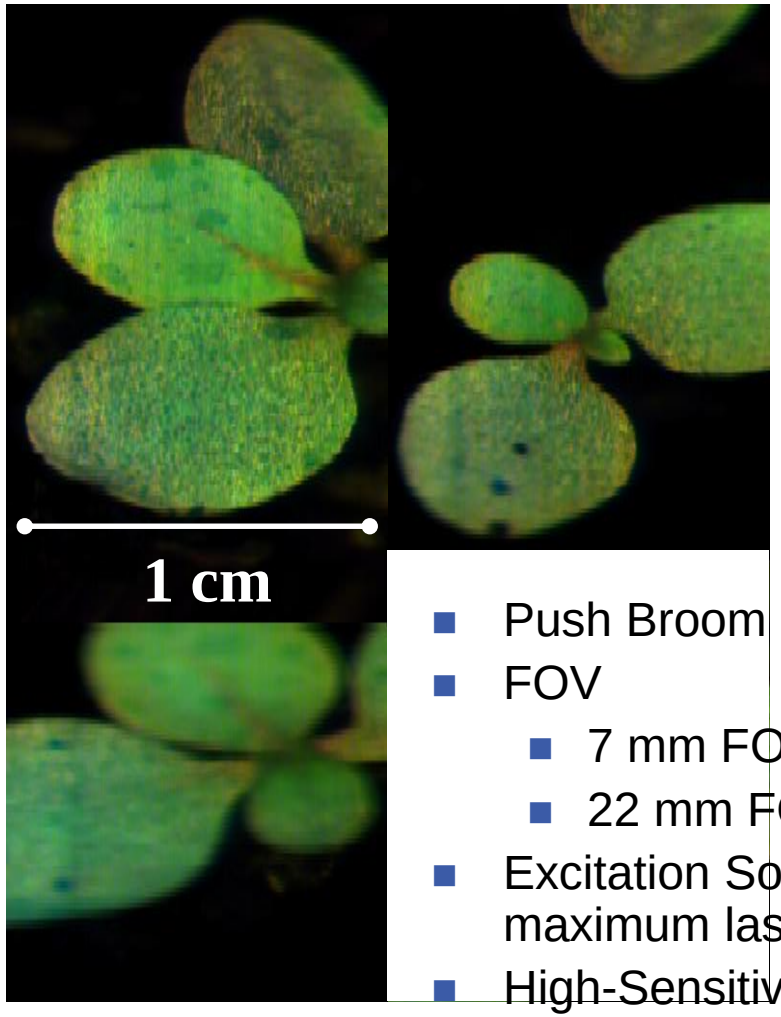
Goal: gene discovery for *economically important traits*

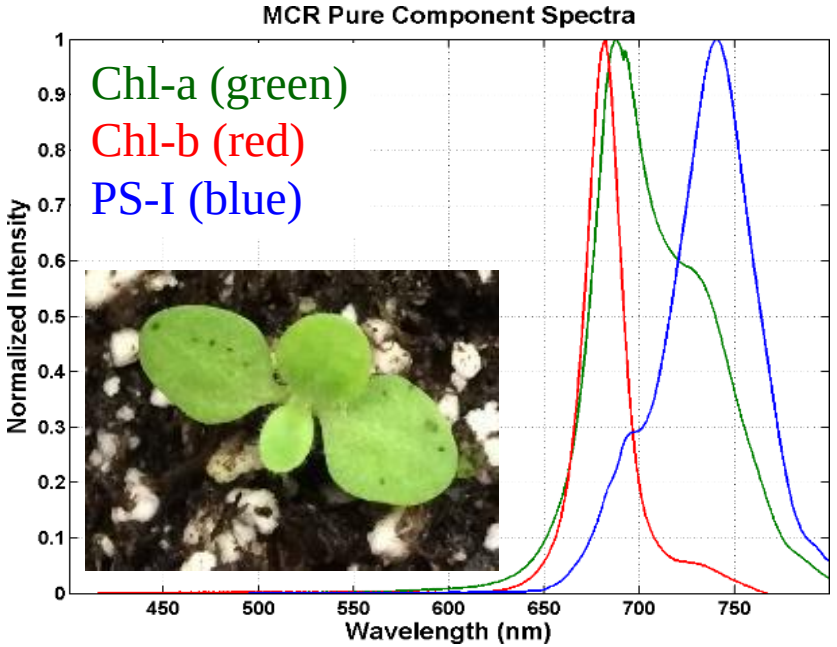
Patenting (€, $)

Markku Keinänen, University of Eastern Finland, SPECIM distributor training, 3 – 5 June, 2015

middleton SPECTRAL VISION

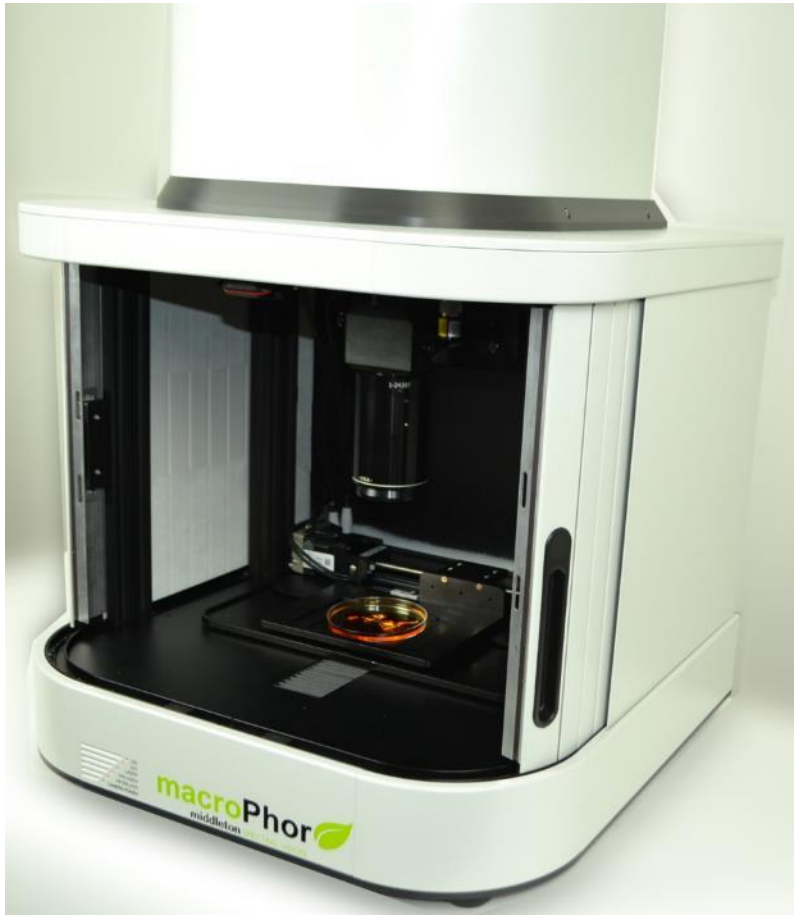
Fluorescence Hyperspectral Imaging of Photosynthetic Pigments





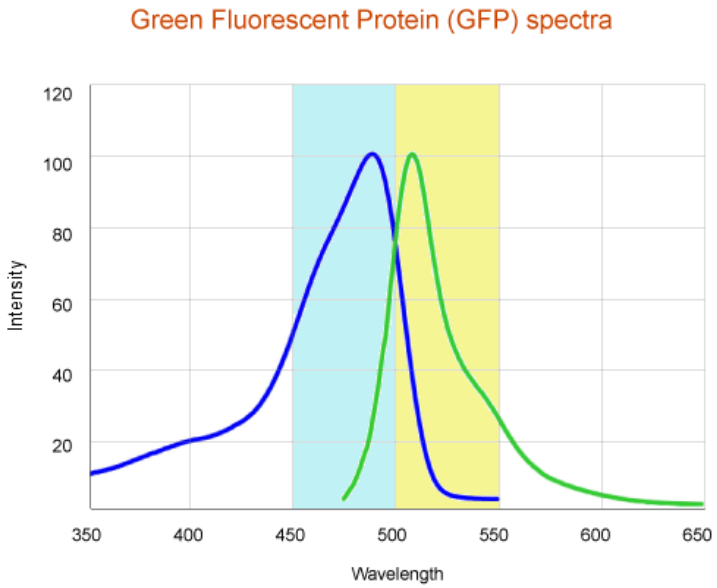
- Push Broom Design
- FOV
  - 7 mm FOV at 34 mm working distance
  - 22 mm FOV at 3 in. working distance
- Excitation Source options 405 nm, 488 nm , 640 nm – maximum laser output 30mW
- High-Sensitivity VNIR sCMOS Hyperspectral Camera
- Optical Resolution: 7.5 nm

MacroPhor ™



middleton SPECTRAL VISION

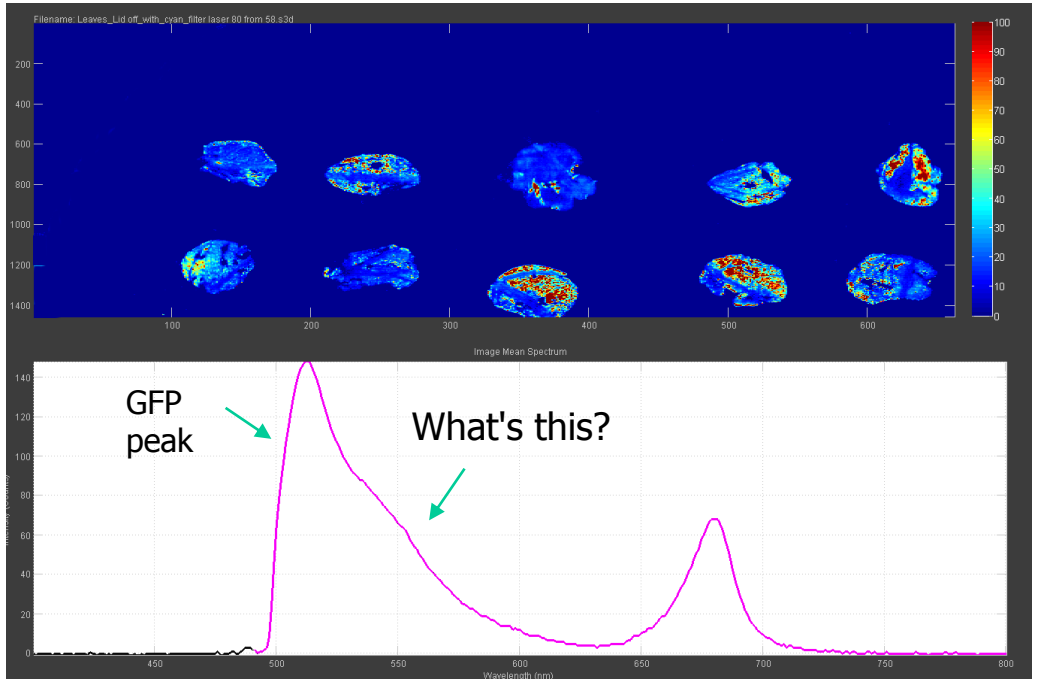
Why hyperspectral fluorescence?



- Current fluorescence imaging techniques are filter based
- An excitation filter is selected to produce an emission spectrum

The problem?

- Spectral crosstalk
- Missing other spectral features that might be useful to the analysis

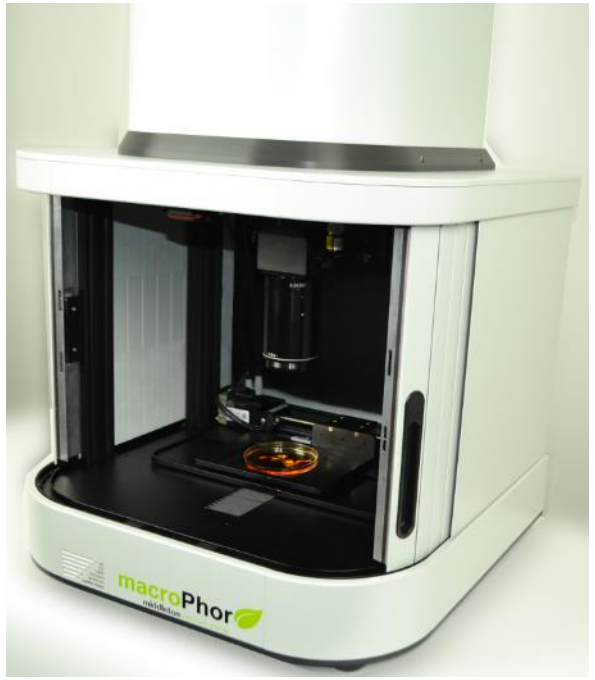


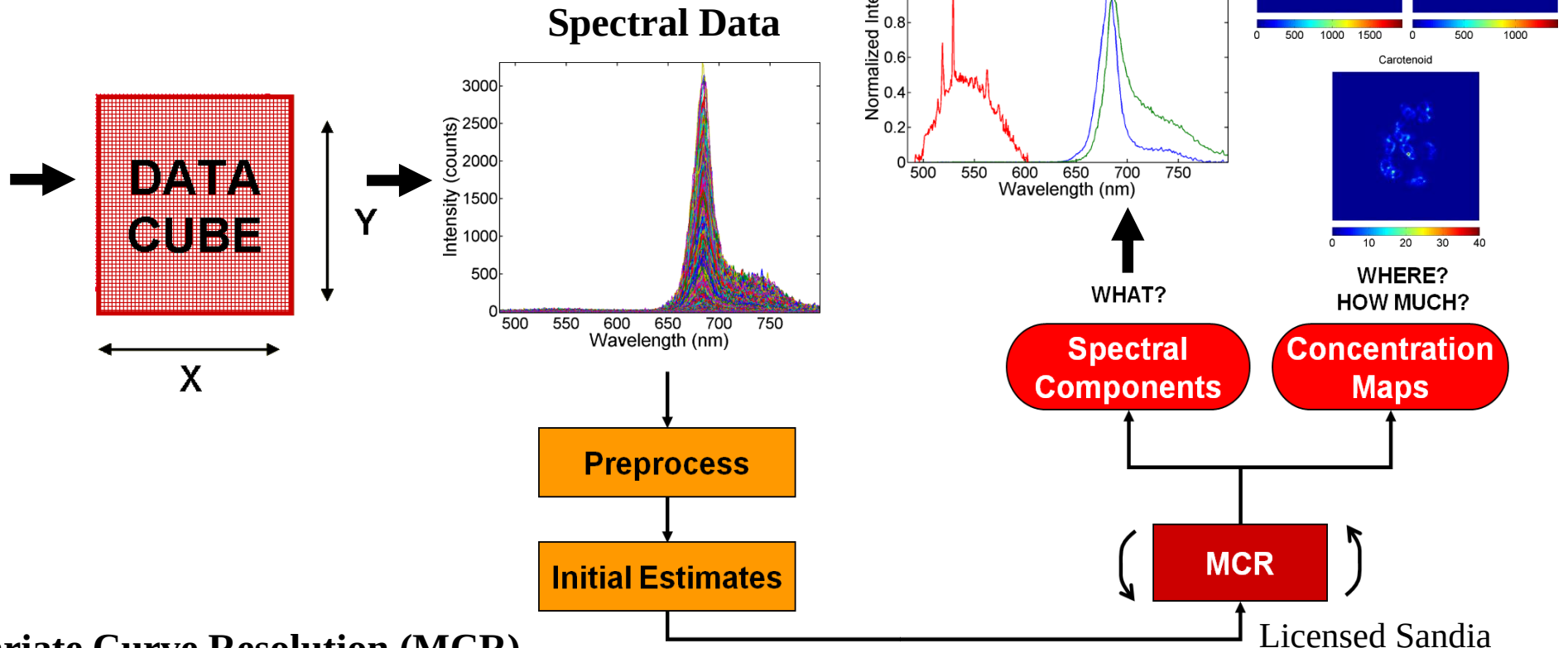
- The hyperspectral image contains all the wavelength information
- Allows you to see features that would not be present with a filter based system
- Using multivariate analysis techniques yields quantitative data

middleton SPECTRAL VISION

Multivariate Curve Resolution (MCR) Analysis Discovery Algorithms

MacroPhor ™





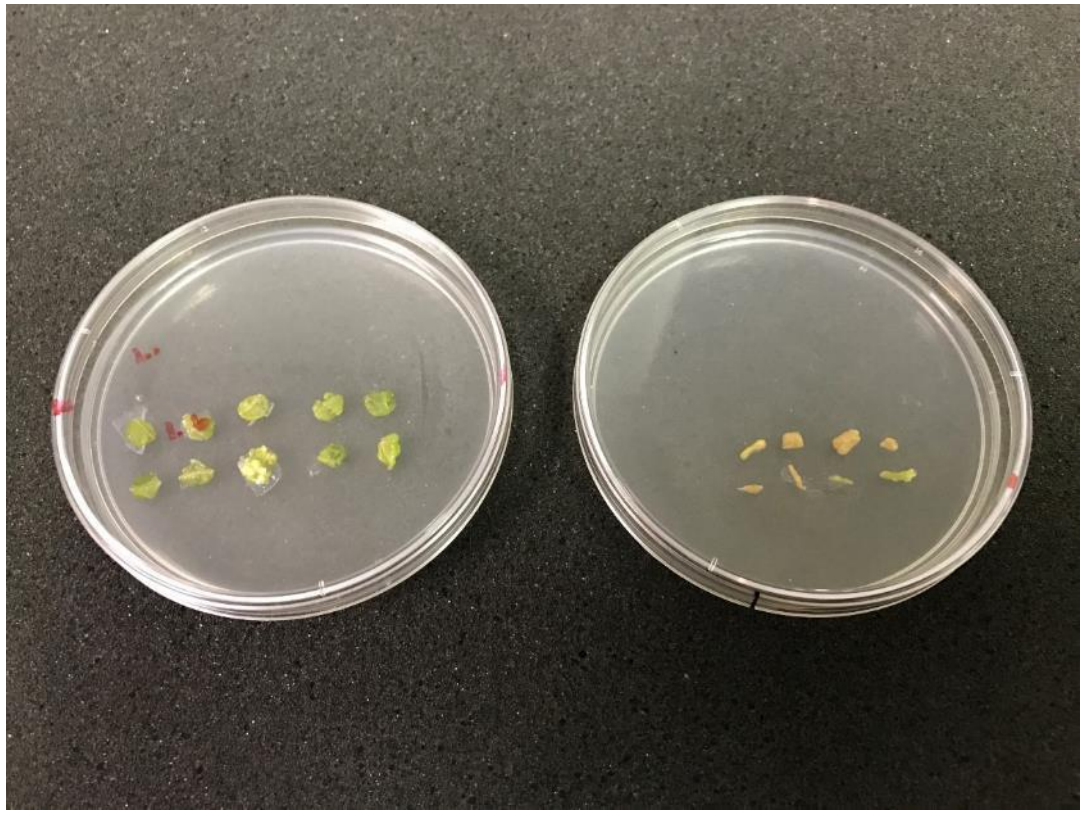
Multivariate Curve Resolution (MCR)

- Linear additive model: $D = CS^T + E$
- Solve $D = CS^T$, constrained alternating (rigorous) least squares methods
- Non-negativity constrained pure components (S) and concentrations (C)
  - Exception: Negativities to model spectral shifts and broadenings

Licensed Sandia National Laboratories MCR Algorithms

middleton SPECTRAL VISION

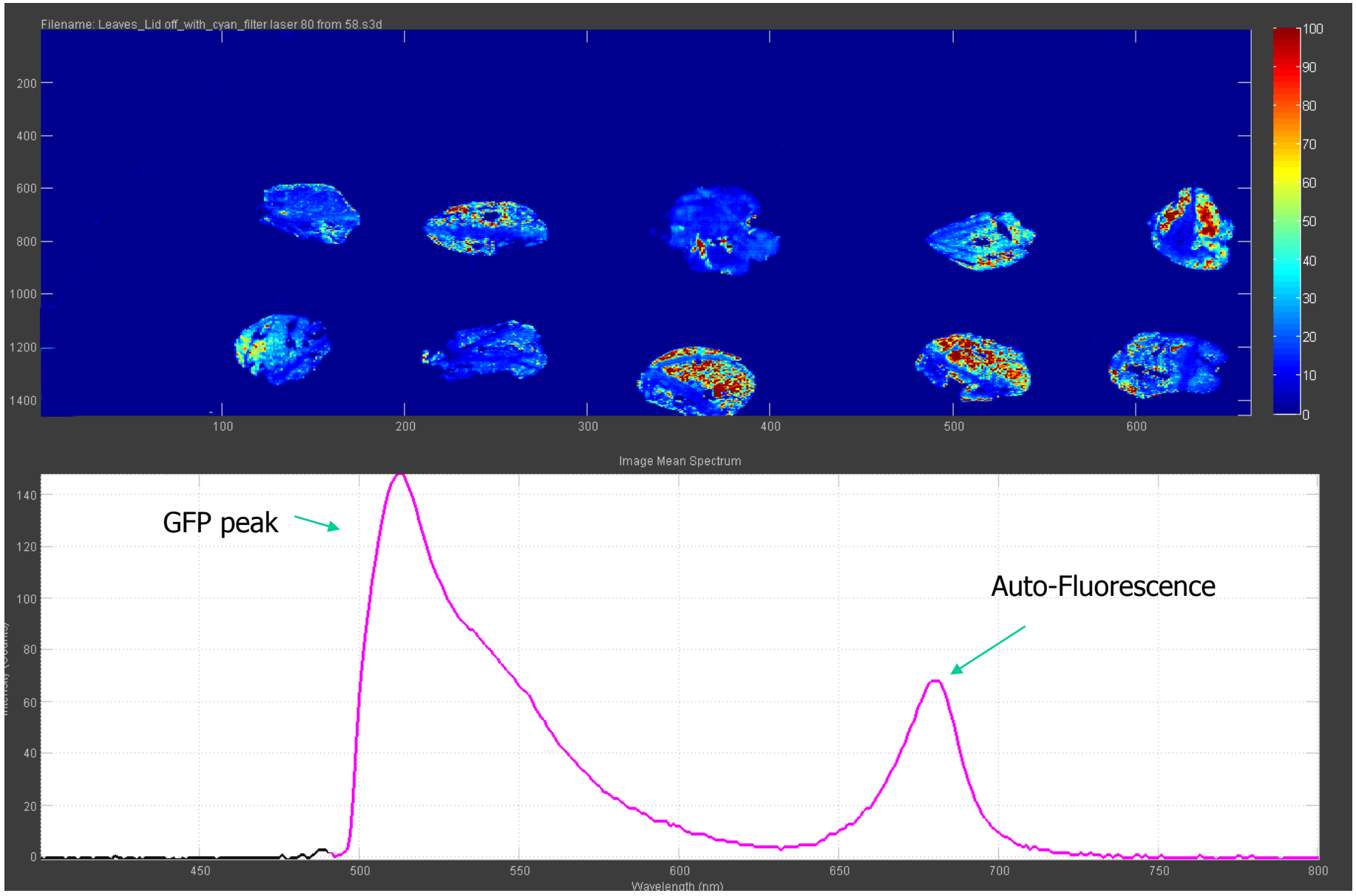
Goals for working together with the Strauss lab



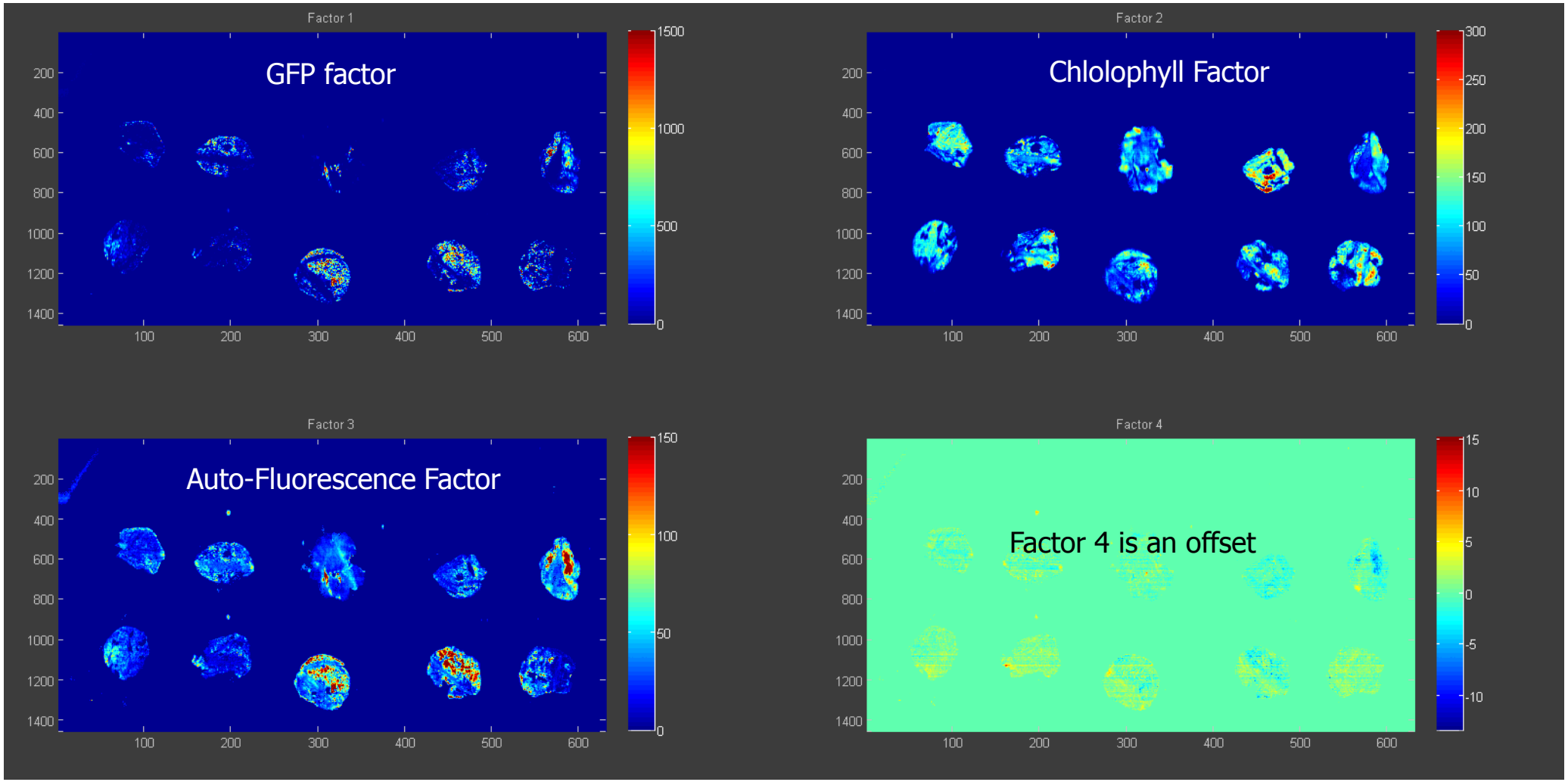
- Provide a versatile system that can image up to 24 sealed petri dishes on a single tray
  - RGB Imaging
  - Fluorescence Imaging
  - VNIR Imaging
- Export Data in an open source file format
- Provide software analysis tools for single plant or petri dish measurements

middleton SPECTRAL VISION

Leaves– mean image



The mean image was separated into three factors using a MCR algorithm



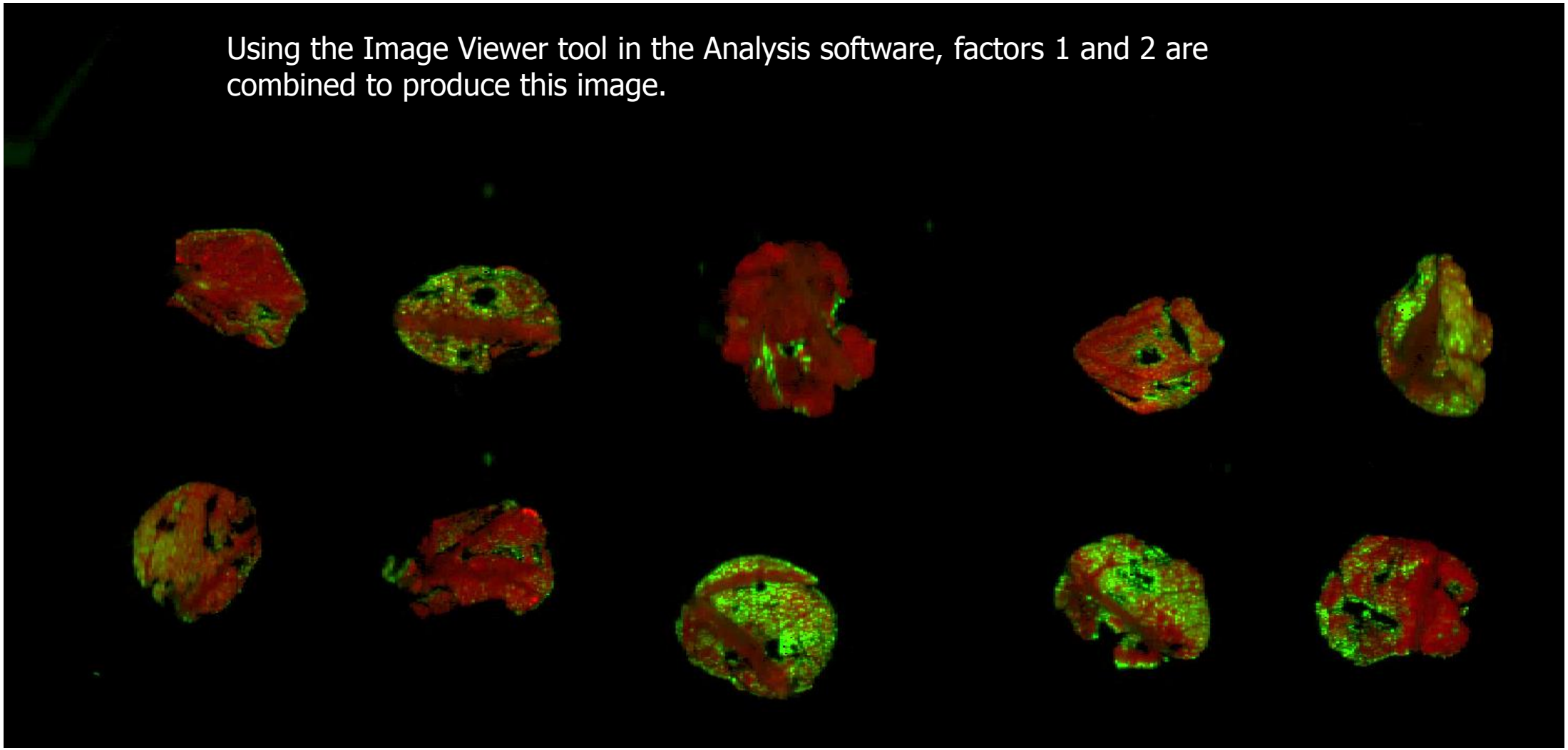
Concentration Maps

middleton SPECTRAL VISION

Spectral Plots

middleton SPECTRAL VISION

Leaves with lid on – GFP and chlorophyll components



middleton SPECTRAL VISION

High throughput imaging system proposal

- Light tight imaging chamber
- High sensitivity fluorescence spectral camera
- 488nm laser line illumination
- X-Y actuator system to position the camera
- Z focus motor
- Sample tray holder for up to 24 petri dishes
- Acquisition and analysis software

middleton SPECTRAL VISION

Contact Information

Info@middletonspectral.com
presenter@middletonspectral.com
www.middletonspectral.com

middleton SPECTRAL VISION

Home

Amy Klocko, Cathleen Ma, & Michael Nagle present…

"Culture Medium and Transformation Plans"

Culture medium and transformation plans

Amy Klocko
Cathleen Ma
Michael Nagle

Presentation overview

- Rationale
- Project goals
- Overviews of regeneration and transformation
- Initial studies of the plant material
  - Rooting
  - Performance in standard culture conditions
- Project plans
- Proposed timeline

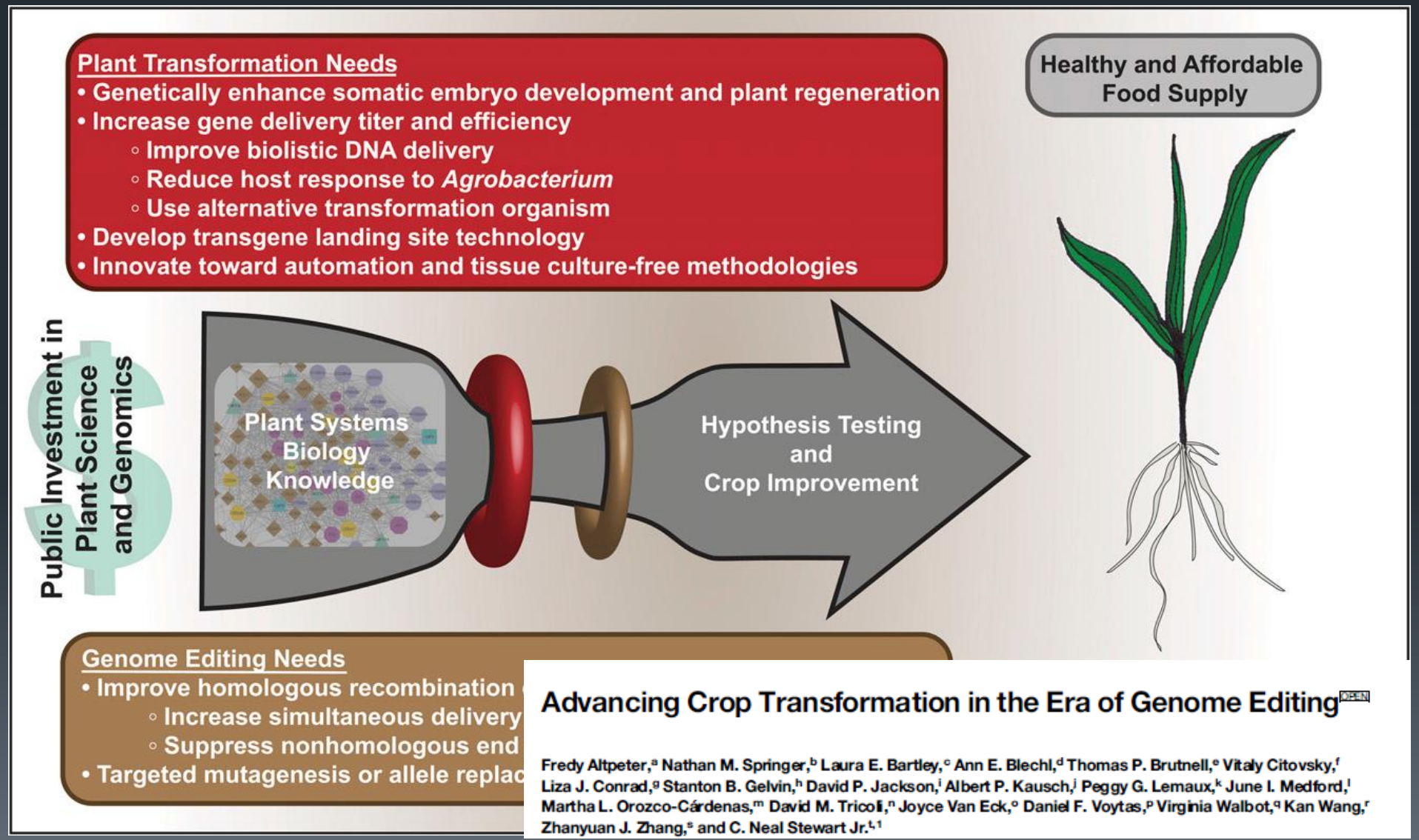
Rationale

- GE technology can be used to produce high value traits in crops
- Commercially important species (rice, wheat) remain extremely challenging to transform as do cultivars of transformable crops (corn)
- We are seeking a biological understanding of the transformation/regeneration (RT) process
- Genes and culture conditions leading to improved RT could be applied to recalcitrant species

Project goals

- Establish uniform greenhouse population of cuttings
- Study *in vivo* rooting methods and *in vivo* callusing
- Optimize *in vitro* growth and culture of poplar
- Perform genome wide association study (GWAS) of RT

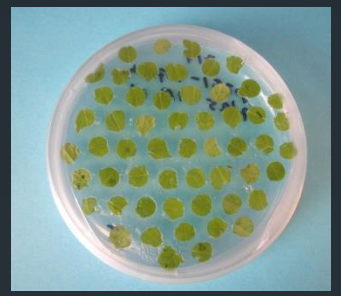
There is great interest in improving plant transformation for research and crop production



Direct versus indirect regeneration *in vitro*

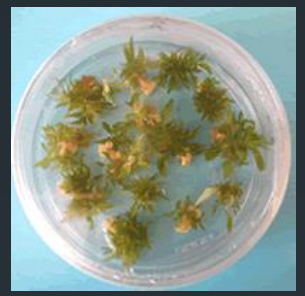
Indirect regeneration is used during plant transformation

Indirect regeneration (has an initial callus stage)



explants

calli



shoots

elongation

rooting

Direct regeneration: immediate growth of shoots (or roots) from explants

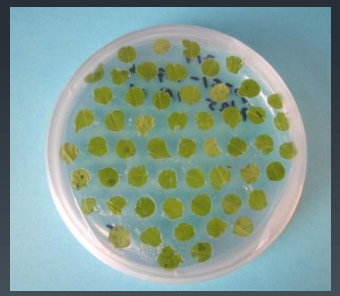
We intend to identify tools for improving the R/T of recalcitrant species/cultivars

Examples of bottlenecks during *in vitro* regeneration

Overview of Agrobacterium-mediated plant transformation

DNA transfer followed by indirect regeneration of shoots

Agrobacterium with plasmid



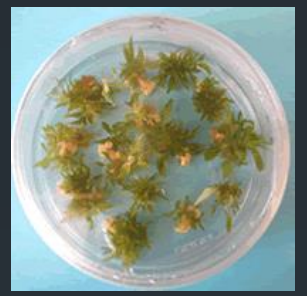
explants

plasmid transfer

stable incorporation of plasmid DNA

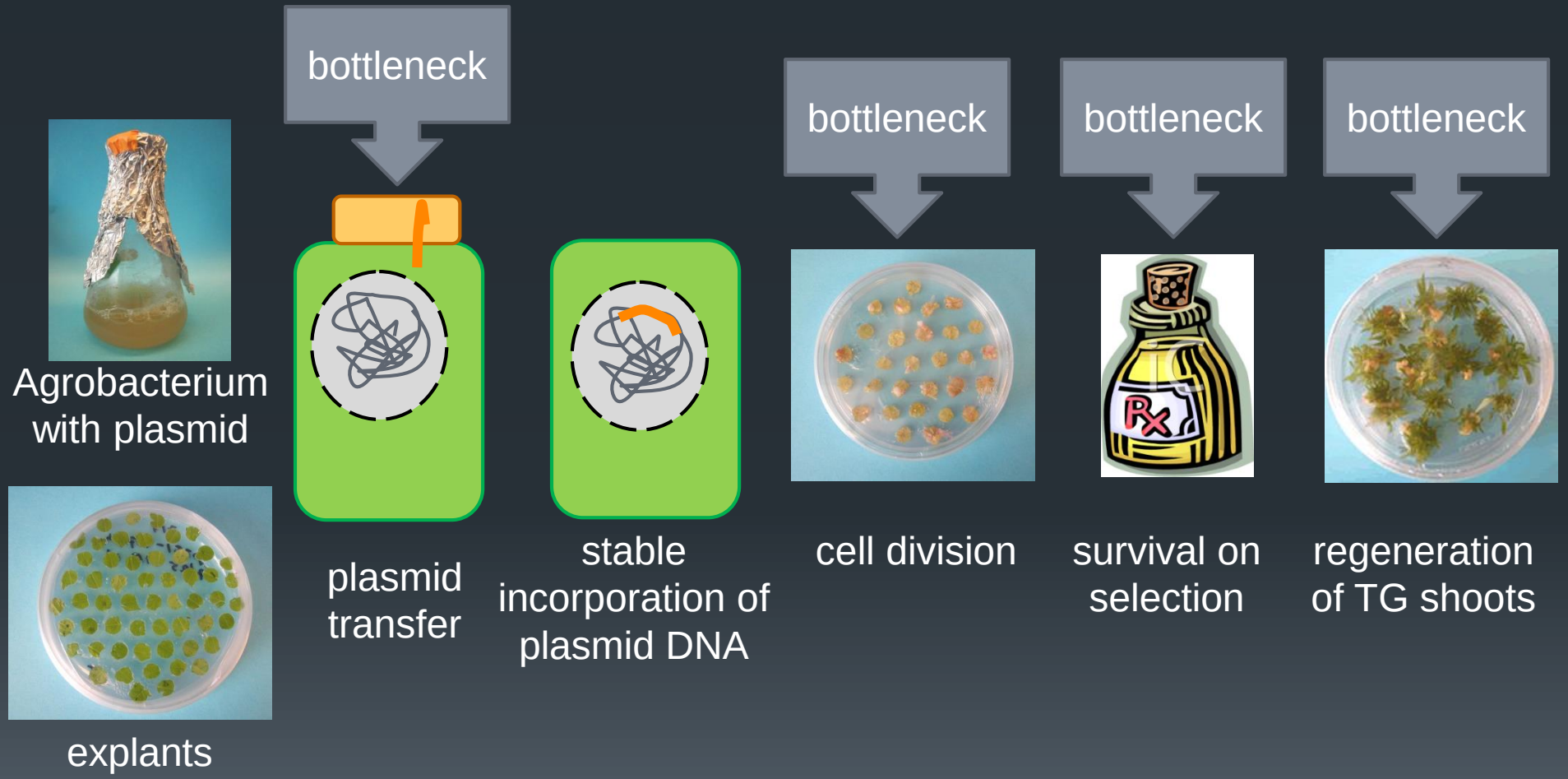
cell division

survival on selection



regeneration of TG shoots

Overview of Agrobacterium-mediated plant transformation

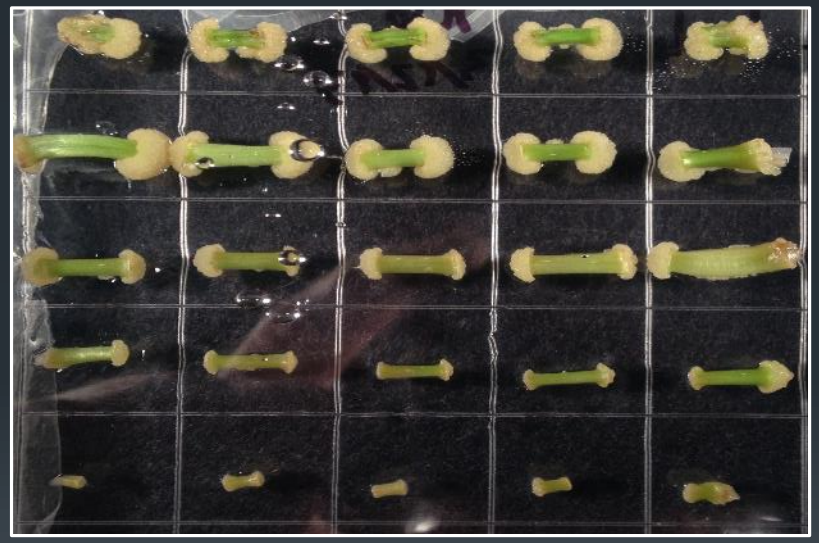


Examples of bottlenecks during *in vitro* regeneration

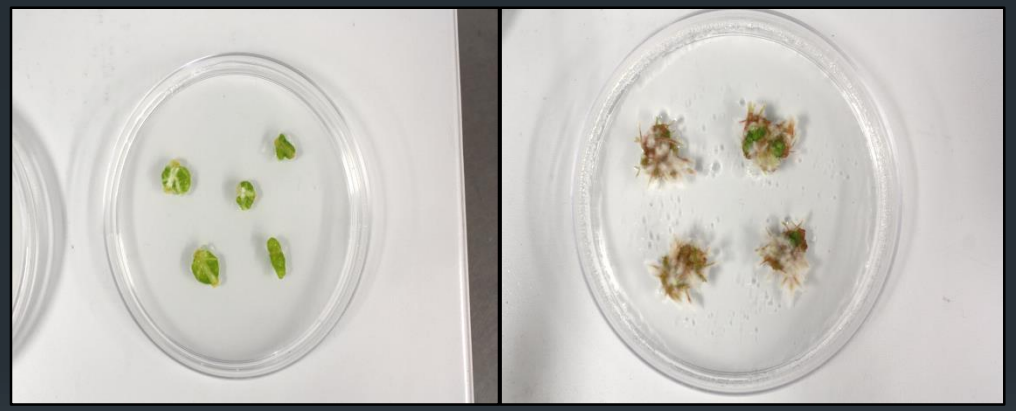
We have over 2,000 varieties of wild *P. trichocarpa* available for source tissue

Study population has a known diverse response to callusing treatment

Stems producing calli



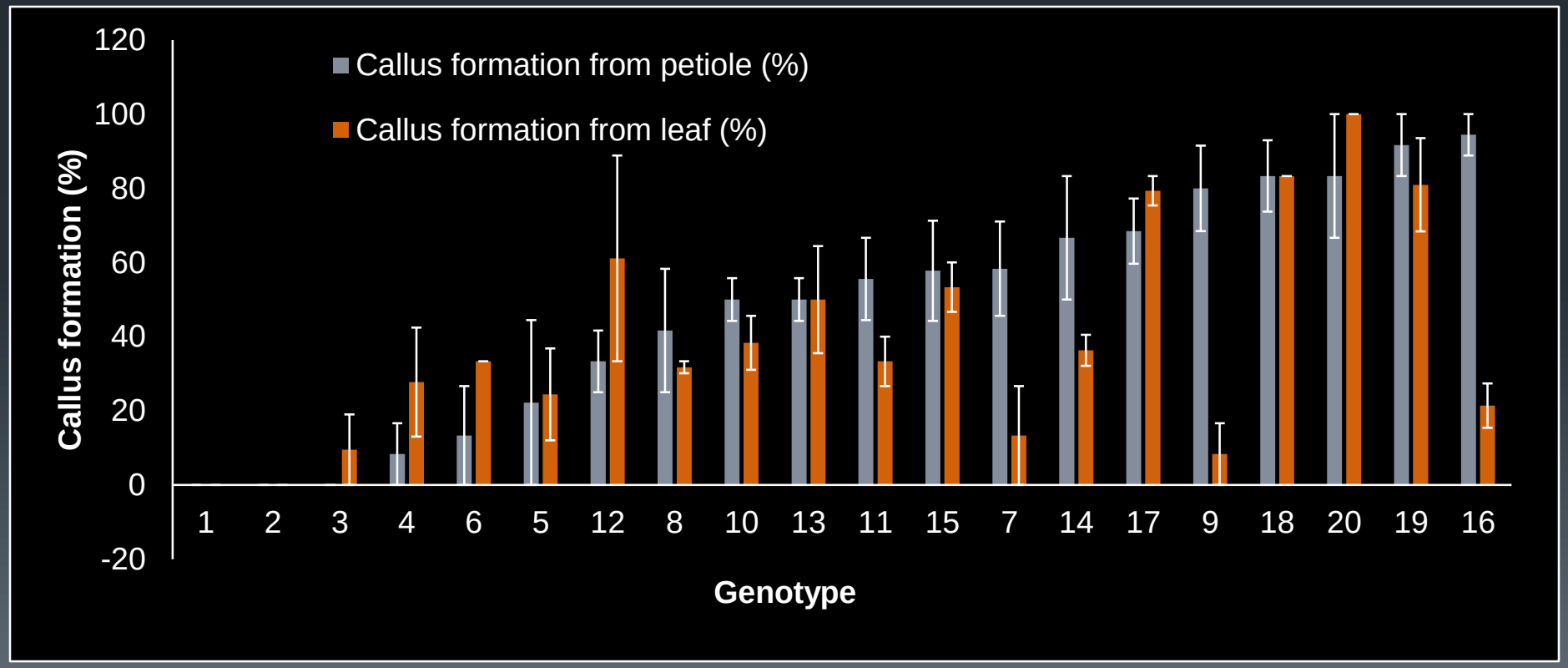
Leaf disks with calli or roots



(left is Muchero and Tuskan, unpubl.)

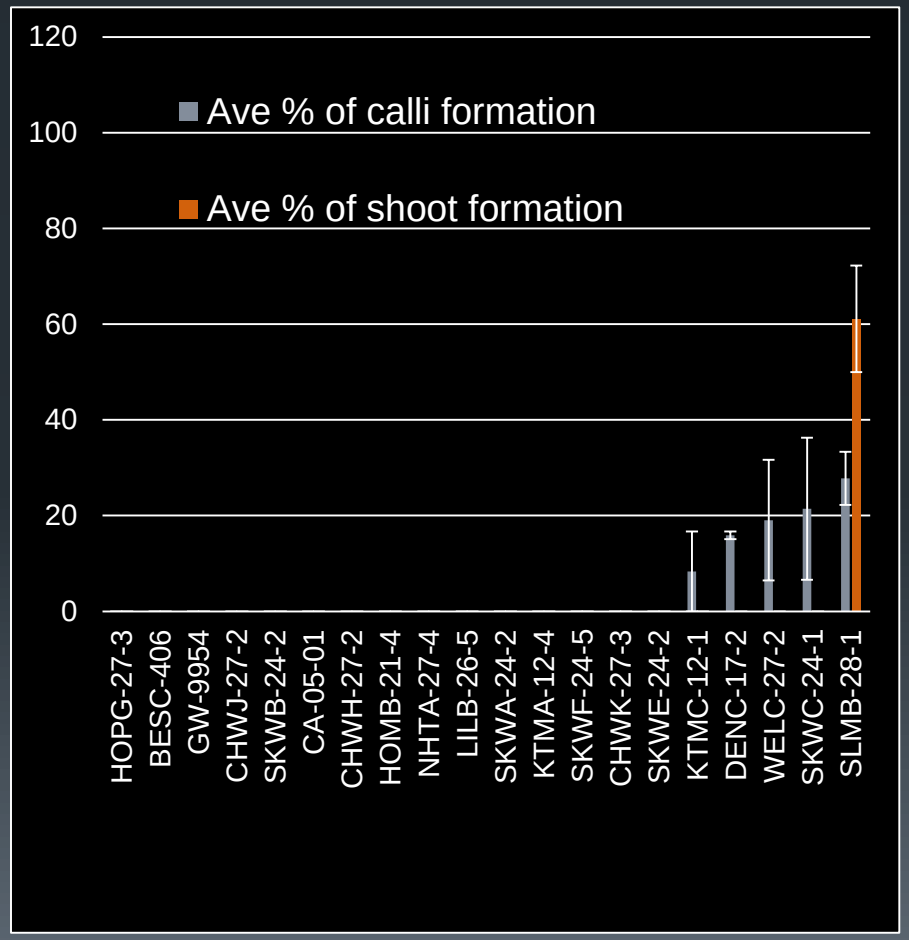
Subset of clones show diverse responses to our standard callus induction treatment

Initial test of 20 clones on standard callus induction medium showed diversity between and within clones

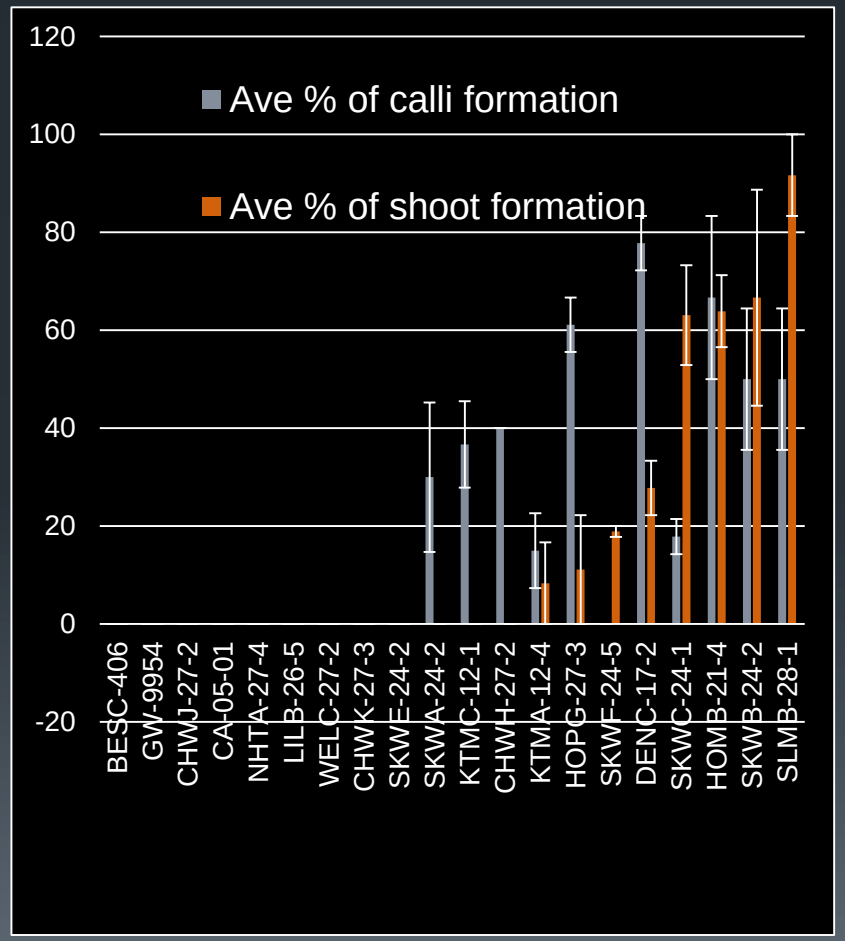


Subset of clones show diverse responses to indirect shoot induction

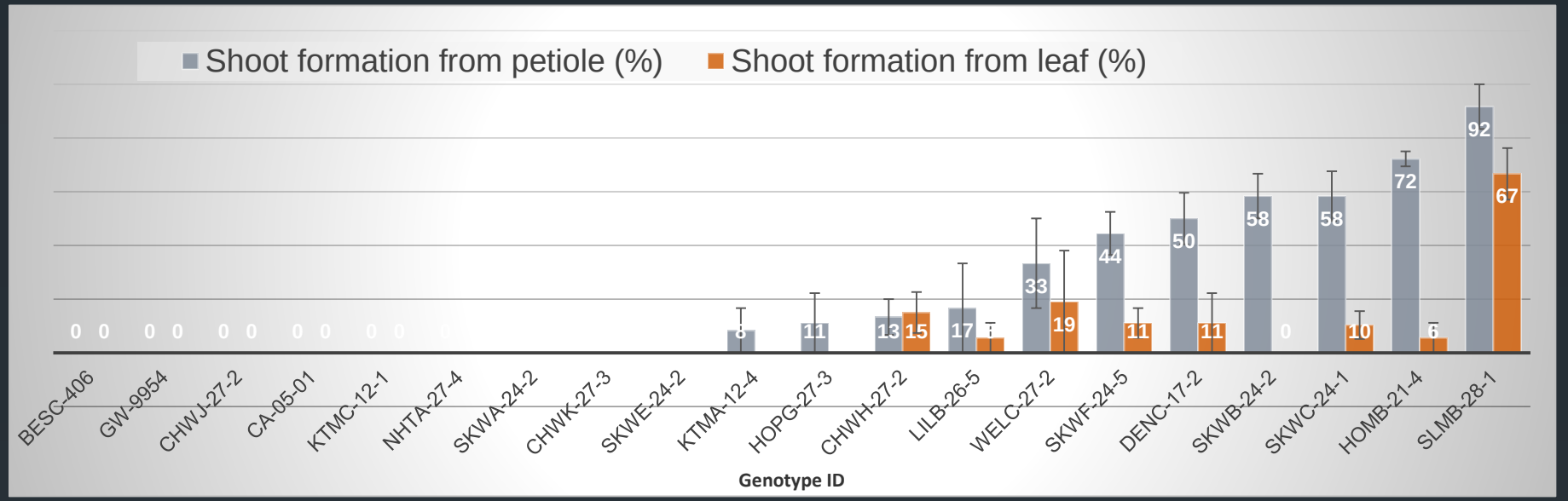
Leaf explants 20 d on SIM1



Petiole explants 20 d on SIM1



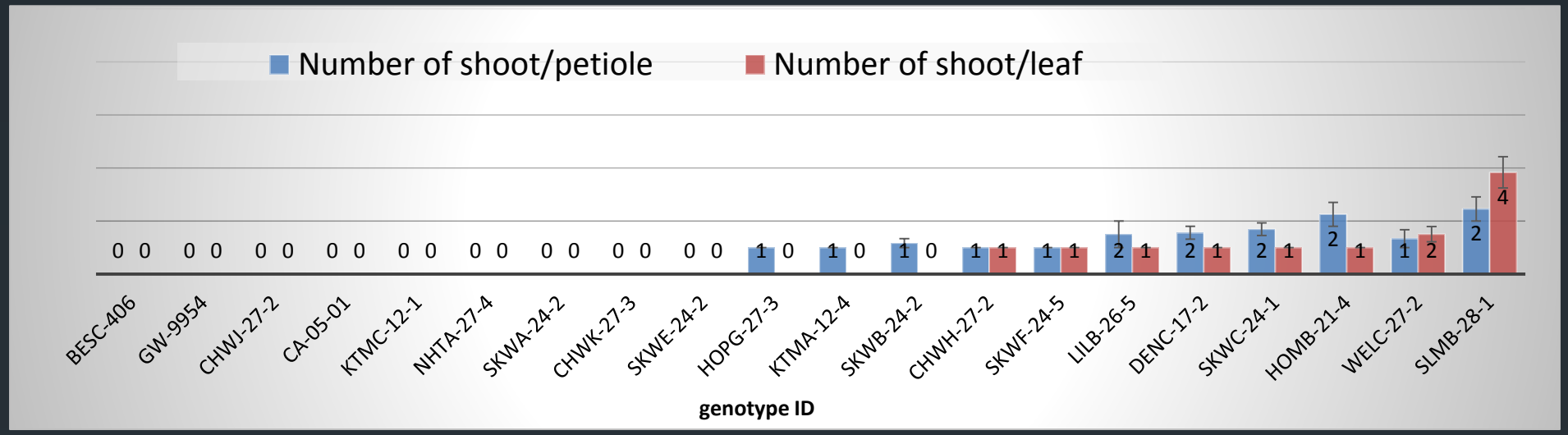
Petiole explants respond better than leaf discs during direct shoot regeneration



SLMB 28-1 petiole explant

SLMB 28-1 leaf explant

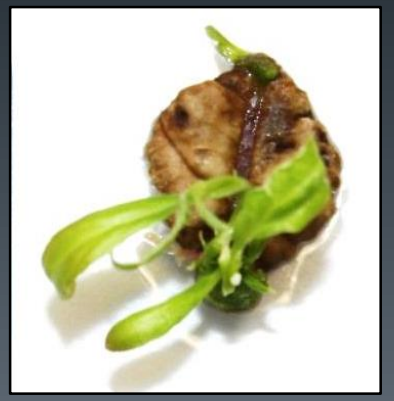
Direct shoot regeneration led to few shoots per explant



HOMB 21-4 petiole explant

SLMB-28-1 leaf explant

SLMB-28-1 leaf explant



Our findings with standard *in vitro* culture highlights the need for improved *in vitro* culture conditions for poplar

In vitro treatments will be tested in two phases

- Phase 1: exploration/investigation
  - Goal: find conditions which maximize response variability
  - Approach: test of 8 main factors on a small number of genotypes
  - Data can also be used for development and testing of image analysis
- Phase 2: GWAS
  - Goal: map the response of genotypes to treatments
  - Approach: test a large number (500+) of genotypes on a small number of conditions for each trait
  - Conditions will be chosen based on heritability, genetic influence, and physiological distinctness
  - Data are to be collected automatically and used for GWAS mapping of traits

Phase 1: exploration of the *in vitro* world with wild poplar

We propose to test eight factors

1. Basal media
2. Hormones
3. Light quality (intensity and ratio)
4. Charcoal
5. Antioxidants
6. Stress/ethylene
7. Agro strains
8. Gene treatment

Each factor will be broken down into individual tests

Proposed treatments for each factor

1. Basal media: 36 treatments
2. Hormones: 115 treatments (BAP, kinetin, NAA, 2-4D, TDZ)
3. Light quality (intensity): 9 treatments
4. Charcoal: 5 treatments
5. Antioxidants (PVP, PVPP, VC, citric acid etc.): 5 chemicals x 5 concentrations
6. Stress/ethylene (low TM, darkness, and AgNO3): 3 ways x 3 levels
7. Agro strains (AGL1, C58, LAB4404, EH101, EH105): 5 treatments
8. Gene treatment (BBM,WUS2, and more): 5 treatments

Initial test of rooting in soil and water as a pilot study for regenerative ability

- Tested 523 genotypes of poplar
- Allowed for synchronization of materials and greenhouse establishment of field cuttings
- Cuttings were collected in spring 2015 and stored at 4 degrees for ~5 months
- 2 blocks in soil, 2 in water
- Data were scored by hand and are being used as a training set for computerized annotation of root tissue

Material had some diversity of morphology prior to intentional rooting

Morphology	# of genotypes
no roots	420
1-2 roots	82
3-5 roots	17
7-11 roots	18
No buds open	522
Some buds open	10
All buds open	5

Plant material was growing well after 3 weeks

Most cuttings rooted after 3 weeks in water

B1

B2

Extensive root growth after 6 weeks in water

B1

B2

Rooting data are being used for training of analysis software

Week 3 images are undergoing analysis

Grey background was not ideal – changed to black

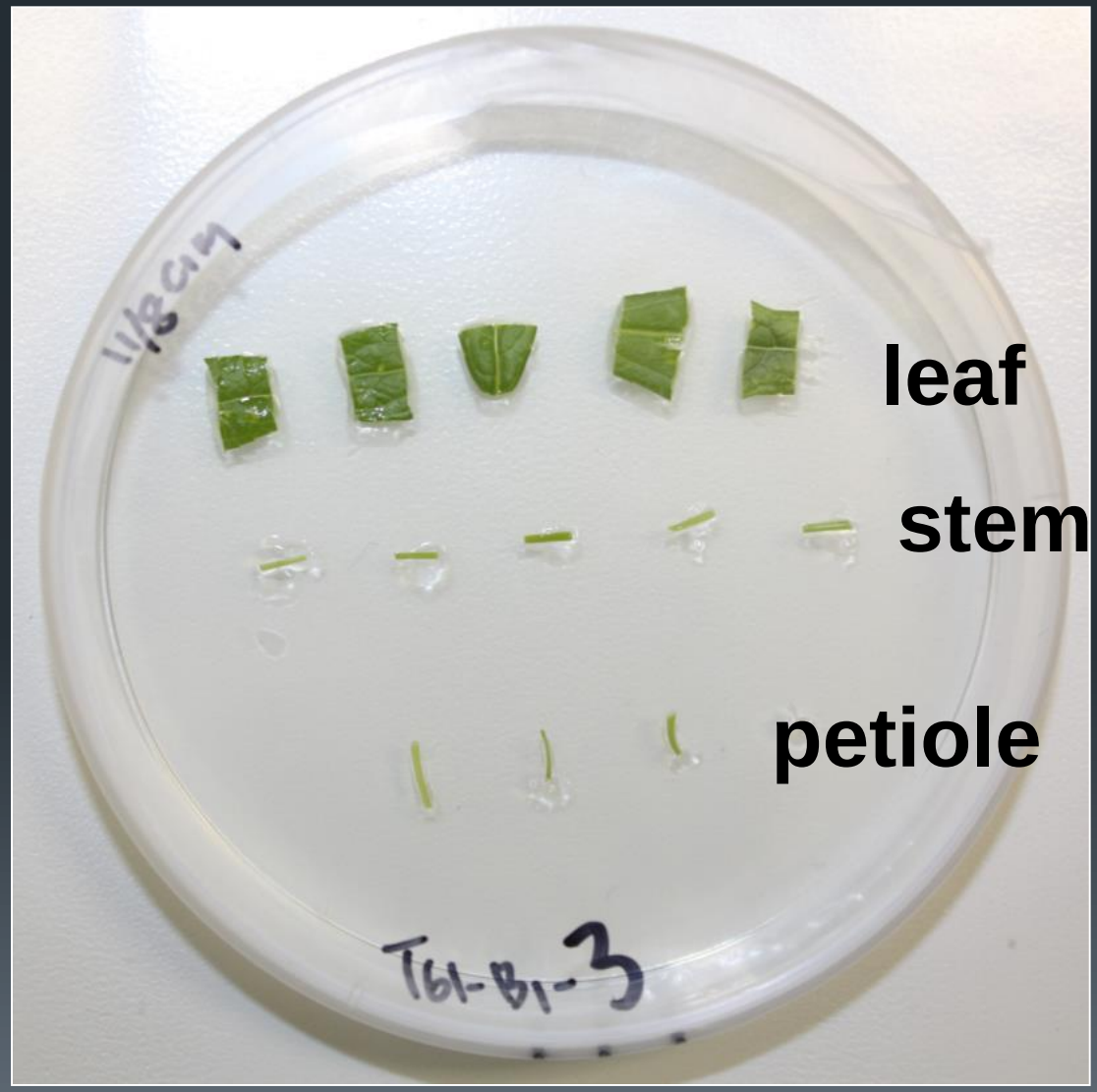
Overlap of samples causes analysis problems – future imaging will space out samples

Automated detection of roots is under development

Basal media testing

- We are testing 36 different basal media on two poplar clones
- Group I: ammonium nitrate, 0.25X, 0.5X, 1X, 1.5X
- Group II: potassium nitrate, 0.5X, 1X, 1.5X
- Group III: mesonutrients, 0.5X, 1X, 1.5X
- Group IV: micronutrients, 1X
- Group V: iron, 1X

We are testing three different explant types: leaves, stems, and petioles



Two clones are being tested

Explants from two clones after 21 days on standard callusing medium

We intend to test additional clones and to publish our findings regarding improved medium for poplar *in vitro* regeneration

Diversity of callus, root and explant morphology observed across media

Explants from one clone after 21 days on CIM, 14 days on SIM

1.5X ammonium nitrate .5X mesonutrients	.5X nitrates 1.5X mesonutrients	.5X ammonium nitrate

These observations illustrate the need for automated and quantified data collection

Gene expression as a reagent to better understand RT

Goal: to better understand regeneration and/or stable transformation by expression of select genes

Literature searches have identified several potential candidates to test

Expression of BBM and WUS improve monocot transformation

Plant Cell Advance Publication. Published on September 6, 2016, doi:10.1105/tpc.16.00124

BREAKTHROUGH REPORT

Morphogenic Regulators *Baby boom* and *Wuschel* Improve Monocot Transformation

Keith Lowe[a], Emily Wu[a], Ning Wang[a], George Hoerster[a], Craig Hastings[a], Myeong-Je Cho[b], Chris Scelonge[a], Brian Lenderts[a], Mark Chamberlin[a], Josh Cushatt[a], Lijuan Wang[a], Larisa Ryan[a], Tanveer Khan[a], Julia Chow-Yiu[a], Wei Hua[a], Maryanne Yu[b], Jenny Banh[b], Zhongmeng Bao[a], Kent Brink[d], Elizabeth Igo[d], Bhojaraja Rudrappa[c], PM Shamseer[c], Wes Bruce[e], Lisa Newman[a], Bo Shen[a], Peizhong Zheng[f], Dennis Bidney[a], Carl Falco[a], Jim Register[a], Zuo-Yu Zhao[a], Deping Xu[a], Todd Jones[a] and William Gordon-Kamm[a,1].

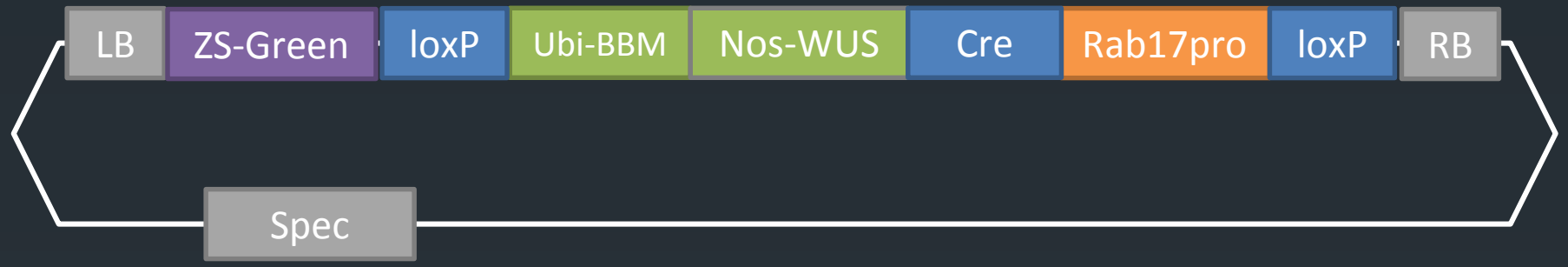
BBM and WUS genes used were from corn

Ubipro:Bbm plus nospro:Wus2 led to high transformation rates

Constructs lead to successful transformation of maize, sorghum, rice, and sugarcane

We received test constructs from Pioneer

Simplified construct diagram for drought inducible BBM and WUS expression



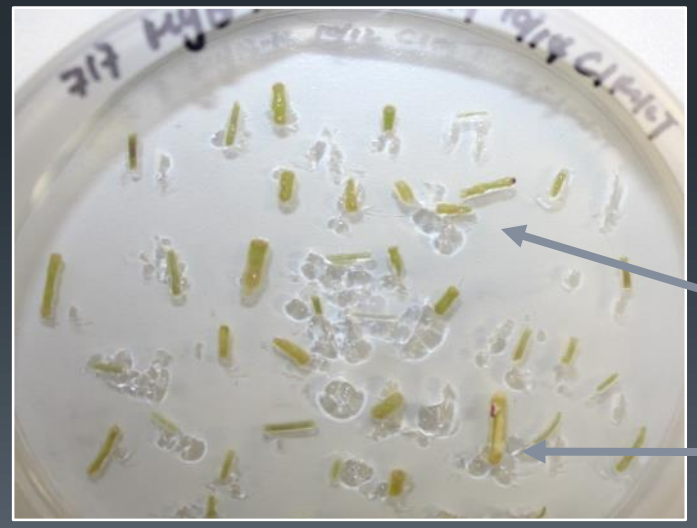
Constructs can be used to test many useful features

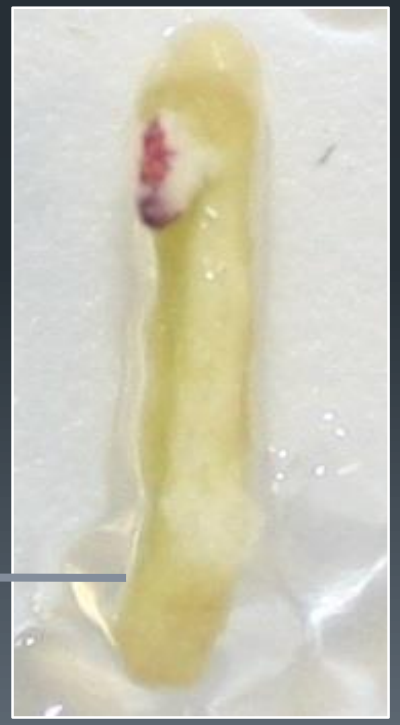
- BBM and WUS for improving RT
- Rab17 promoter, drought inducible, for controlling transgene expression
- Cre/lox excision system for transgene removal
- We are testing these various construct elements

We are also testing transformation markers

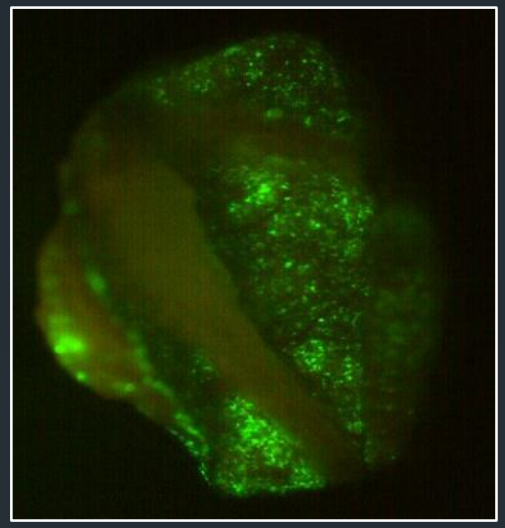
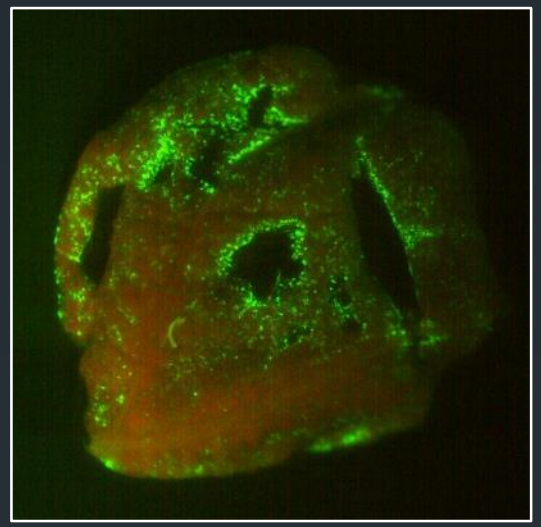
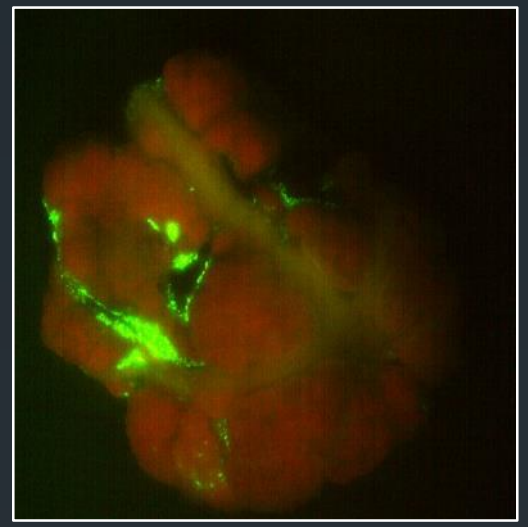
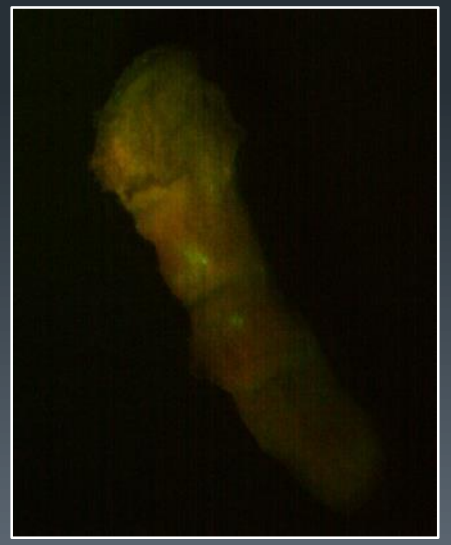
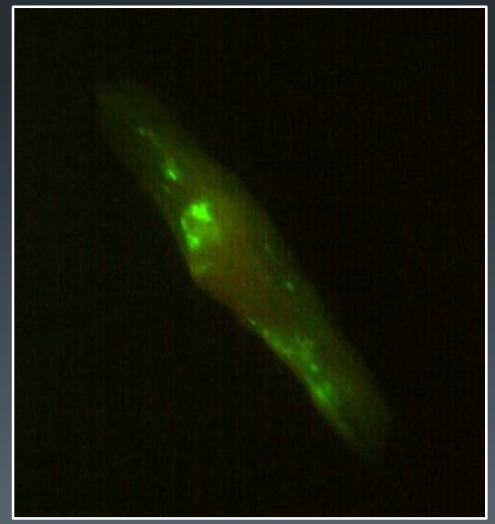
- Constructs from Pioneer and other sources
- Comparison of GFP and anthocyanin for marking stably transformed cells
- Comparing sensitivity, stability, efficiency, and health of transformants
- GFP: visible only by fluorescence imaging methods
- Anthocyanin: visible with white light

Anthocyanin signal visible in explants after 17 days





GFP signal visible in explants after 17 days

Phase 2: GWAS mapping of traits

- Traits to be mapped
  - Indirect regeneration
    - Callus formation
    - Shoot formation after callus
  - Direct regeneration
    - Shoot formation from explants
    - Root formation from explants
  - Stable transformation
    - Observation of GFP and/or anthocyanin producing tissues
  - Shoot regeneration following gene treatments (BMM etc)
  - We also intend to compare callus and root formation *in vitro* and *in vivo*

Phase 2 will involve a large scaling up of our experimental process

- Requires growth chambers (purchasing process ongoing)
- Plates will be imaged by our automated system (under development)
- Data are to be collected by automated image analysis (under development)

- We intend to undertake an early training run to score callus formation *in vitro*
  - Earliest trait that we score
  - Data will be used to help select clones for future trait analysis

Proposed timeline for *in vitro* testing

Activity	Pre-Year 1	Year 1	Year 2	Year 3	Year 4	Year 5
Collection of field materials	X	X				
Test of rooting and callusing in greenhouse (*in vivo*)		XX				
Phase 1: Testing of 8 culture conditions	XX	XXXX	XXXX			
Vector creation and marker testing	XX	XXXX				
Selection of conditions and clones for phase 2		XXXX	XXXX			
Large-scale experiments for GWAS (1 in year 1, then 2 per year)		XX	XXXX	XXXX	XXXX	XXXX

Home

Michael Nagle presents…

"Epigenetic Regulators of Regeneration and Transformation (RT) For Expanding the Reach of GWAS"

Epigenetic regulators of regeneration and transformation (RT) for expanding the reach of GWAS

By Michael Nagle
December 20, 2016

Rational for targeting epigenetic regulators

- Epigenetic variation long thought to affect RT competence
- Does epigenetic perturbation affect rate of RT in poplar?
- Is there sufficient genetic variation to enable mapping of related genes?

Does epigenetic state condition response to meristem induction genes?

- Will study effect of BBM, WUS overexpression in prelim studies and possibly GWAS (Klocko ppt)
- Limited response to *Wus2, Bbm* in some genotypes[1]
  - 17/50 inbred maize Pioneer lines were still completely recalcitrant
  - 16/50 had <1% transformation efficiency
- Epigenetic regulators broadly affect expression of thousands of genes, including *Wus-like* genes
  - Targeting these regulators may help overcome recalcitrance in more genotypes
- Reviewed literature: Data show effects of knockout of various epigenetic regulators in Arabidopsis
  - Loss of transgene silencing ability[2]
  - Increased speed of shoot regeneration and/or total # shoots regenerated[3]

1. Lowe, Plant Cell 2016
2. Chan, Science 2004
3. Shemer, Plant Sci 2015

Review of literature on epigenetic regulators of RT

- Reviewed literature from 1996-2016
  - Google Scholar to find articles by searching for regeneration-related keywords (often inside body text) and find newer articles that cite articles of interest
- 32 articles with information on leaf disc → callus → shoot transitions
  - To understand epigenetic regulators, an understanding of downstream processes is vital
- 23 articles on epigenetic regulators of these transitions and effects of mutations in transgenic Arabidopsis

  Highlights:
  - 1 mutant has ability to maintain regenerative capacity at old age of plant
  - 3 mutants have faster shoot regeneration but no increase in total shoot mass
  - 4 mutants have increased regeneration and shoot differentiation
    - Some mutations increase both callus and shoot formation rates, while others increase one and decrease the other
  - Other epigenetic regulators are known to affect RT-related genes, but effects of mutations on callus/shoot formation have not been reported

Background: Roles of DNA methylation in regeneration, transformation (RT)

- Types of DNA methylation affect RT efficiency differently
  - Promoter and coding region of *WUSCHEL* are methylated at various positions
  - Different methylation patterns: different expression patterns
- De novo methylases are RNA-guided
  - Add methylation where there is none (i.e. on a transgene)
- Maintenance methylases are guided by parent strand methylation
  - Add methylation to daughter strands to maintain patterns over generations

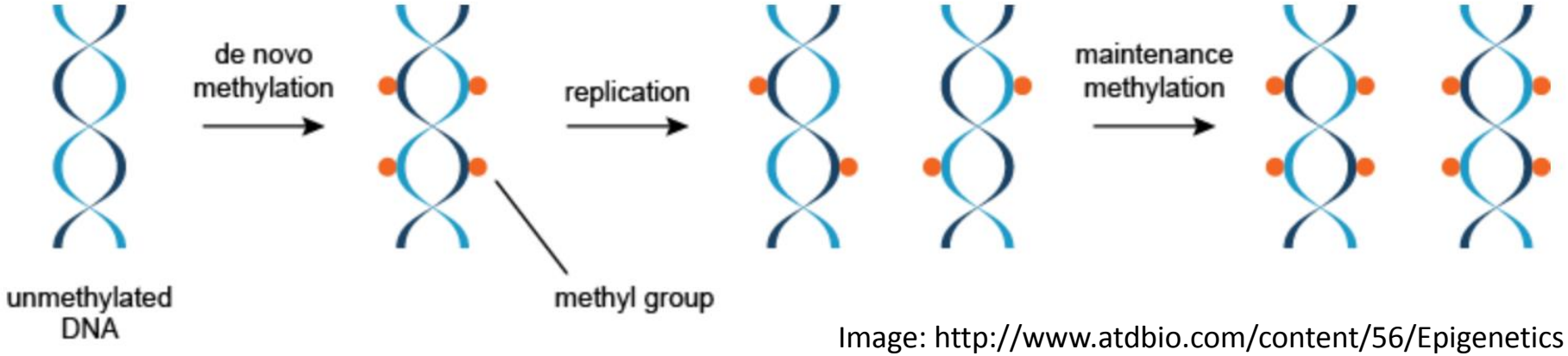
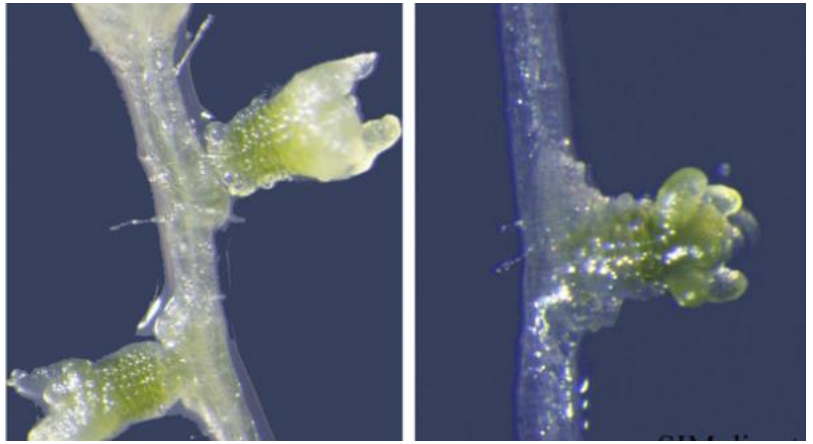


Image: http://www.atdbio.com/content/56/Epigenetics

Candidate Gene Targets: *CHROMOMETHYLASE 3-like*

- Pre-incubation on callus induction media (CIM) induces competency to regenerate shoots
  1. Auxin in CIM promotes rapid cell division
  2. Daughter strands are not methylated quickly enough to keep *WUS* repressed
  3. Loss of methylation on *WUS* promoter → increased *WUS* expression
- Mutant *cmt3* regenerates more shoots with no pre-incubation
  - Methylation is lost more quickly

Arabidopsis loss-of-function mutants



Left: Shoots regenerated after callus induction (CIM→SIM)
Right: Shoots regenerated directly from root (SIM direct)

Competency for shoot regeneration from Arabidopsis root explants is regulated by DNA methylation

Or Shemer[a], Udi Landau[a], Héctor Candela[b], Assaf Zemach[c], Leor Eshed Williams[a,*]

[a] *The Robert H. Smith Institute of Plant Sciences & Genetics in Agriculture, The Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot, Israel*
[b] *Instituto de Bioingeniería, Universidad Miguel Hernández, Campus de Elche, 03202 Elche, Spain*
[c] *Department of Molecular Biology and Ecology of Plants, Tel Aviv University, Tel Aviv 69978, Israel*

Candidate Gene Targets:
DOMAINS REARRANGED METHYLTRANSFERASE-like

- In Arabidopsis, *DRM1-2* repress genes including *WUSCHEL* by methylating regions in promoters, introns, sometimes coding DNA[1]
- Loss-of-function mutants *drm1-2*:
  - higher *WUSCHEL* expression
  - increased rates of shoot regeneration[1]
  - failure to repress transgenes silencing[2]

RNA Silencing Genes Control de Novo DNA Methylation

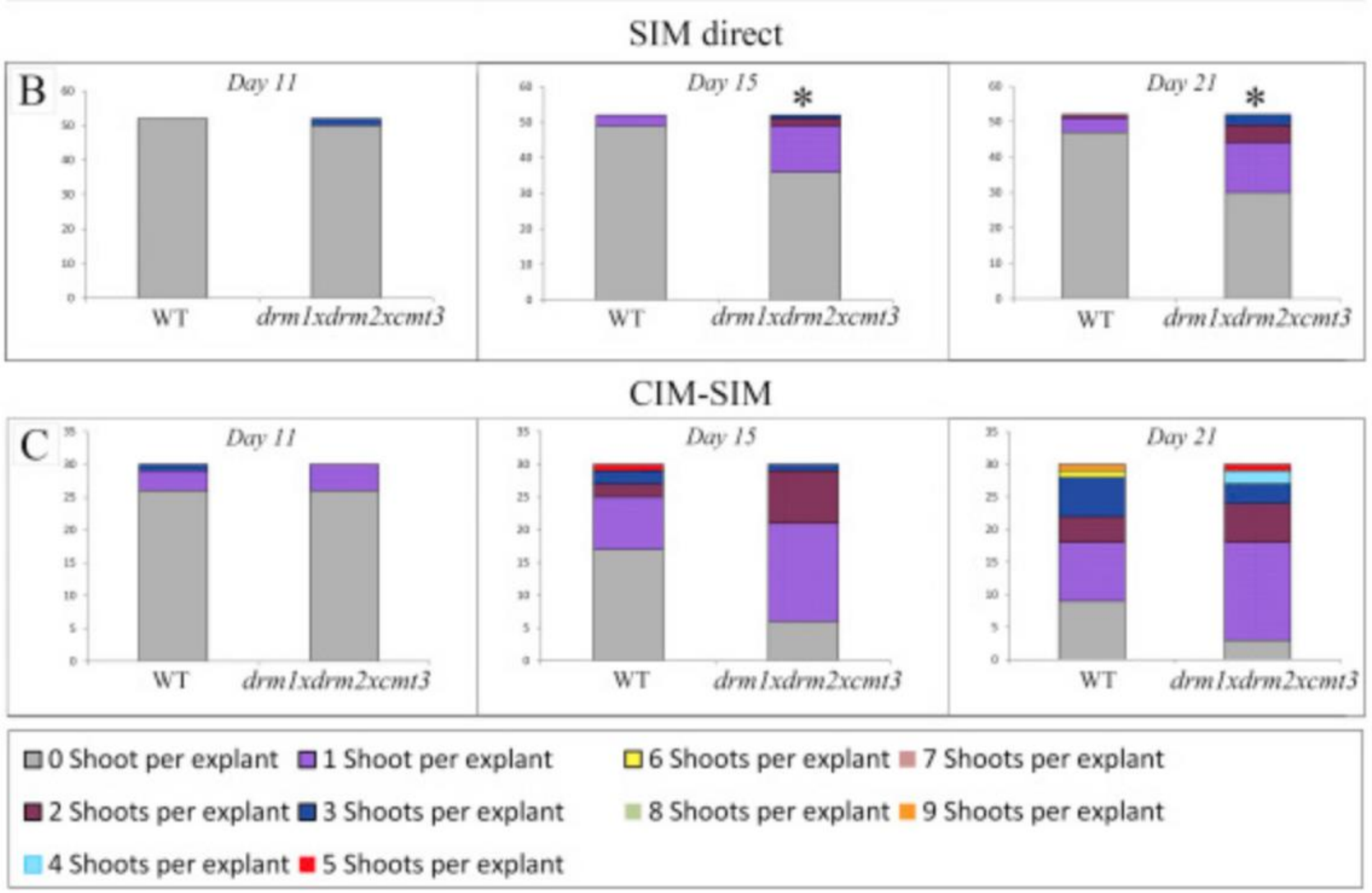
Simon W.-L. Chan,[1] Daniel Zilberman,[1] Zhixin Xie,[2]
Lisa K. Johansen,[2] James C. Carrington,[2] Steven E. Jacobsen[1]*

Arabidopsis overexpressing late flowering gene, with and without DMR1-2

Left: plants with functional DMR1-2 silence the transgene
Right: the late-flowering transgene is not silenced in *dmr1-2* mutants

1. Shemer, Plant Sci 2015
2. Chan, Science 2002

Increased shoot regeneration rates in *cmt3*, *drm1-2* mutants



Shemer, Plant Sci 2015

Proposed experiments

- Bioinformatic studies of poplar homologs underway – homology unclear for some
- Transformation of *P. trichocharpa* to knock down DRM and CMT3 homologs using RNAi
  (or possibly CRISPR/Cas9)
- Preliminary studies of effects in a small number of genotypes prior to possible larger scale use

Next Steps

- Experiment to determine how function of these epigenetic regulators in RT is conserved between Arabidopsis, Populus
  - Design and assembly of RNAi or CRISPR/Cas9 constructs to target possible DRM1-2, CMT3 homologs
  - Agrobacterium-mediated transformation of *P. trichocharpa* to knock down/out targets
  - Regeneration from tissue culture, measuring of shoot regeneration
- If results promising, expand and possibly also look at methylation patterns of *WUSCHEL* in transgenics with altered regeneration (possible thesis study)
  - DNA extraction, bisulfate treatment, amplification of *PtWus* promoter, 5-methyl C sequencing

Home

Troy Hall & Jay Well present…

"PGRP Outreach and Social Science"

PGRP
Outreach and Social Science

Troy Hall

Forest Ecosystems & Society

Jay Well

Science and Math Investigative Learning Experience (SMILE)

Team:

- Troy Hall

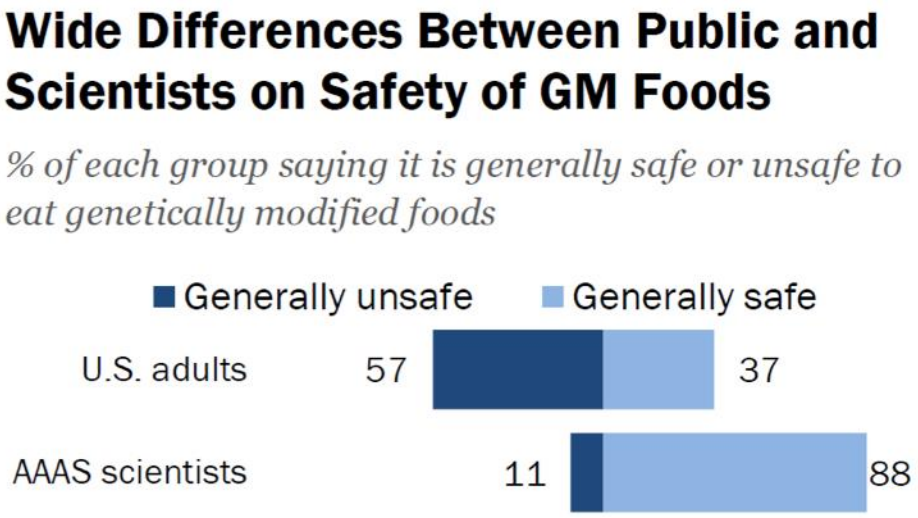
- Jay Well

PGRP Outreach and Social Science

- Brief background
- Our goals and broad approach
- Introduction to SMILE
- Sequencing of activities
- Initial look at the psychological theory
- Questions and discussion

Context – what's the issue?

- Overall lack of STEM engagement in the US
- Decreasing public scientific literacy + increasing science complexity
- Increasing gaps between scientific and public understandings
- Public policy driven by simplistic understandings of issues based on incomplete or one-sided messages



Specific Goals

- Increase high school teachers' content area knowledge, confidence and access to materials for teaching about genetics in society (emphasis on GMO crops)
- Increase learners' abilities to think critically and introspectively about genetic technology
- Increase students' ability to apply scientific knowledge to address complex socio-scientific problems.

Our broader impacts & social science approach

- Combine secondary school science outreach and social science

- Result:
  - Psychologically sound, effective materials & tools
  - Empowered teachers
  - Scientifically literate, engaged students

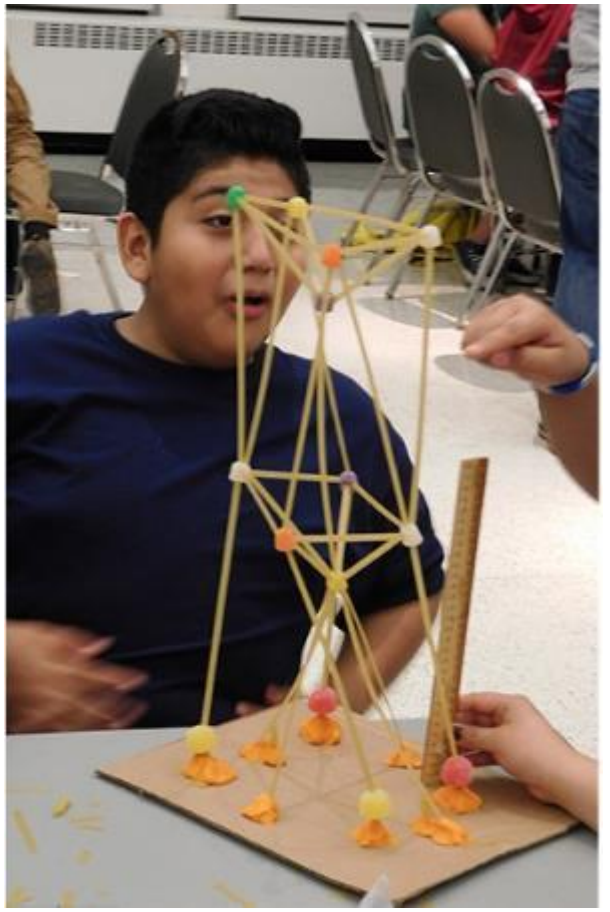
Our broader impacts approach

- Development of two case studies for high school science clubs and classrooms
- Iterative development process
  - Front-end and formative evaluation steps → develop and refine case studies, using SMILE clubs
- Quasi-experimental design
  - Quasi-experimental pre-test, post-test assessment (summative evaluation), using classes

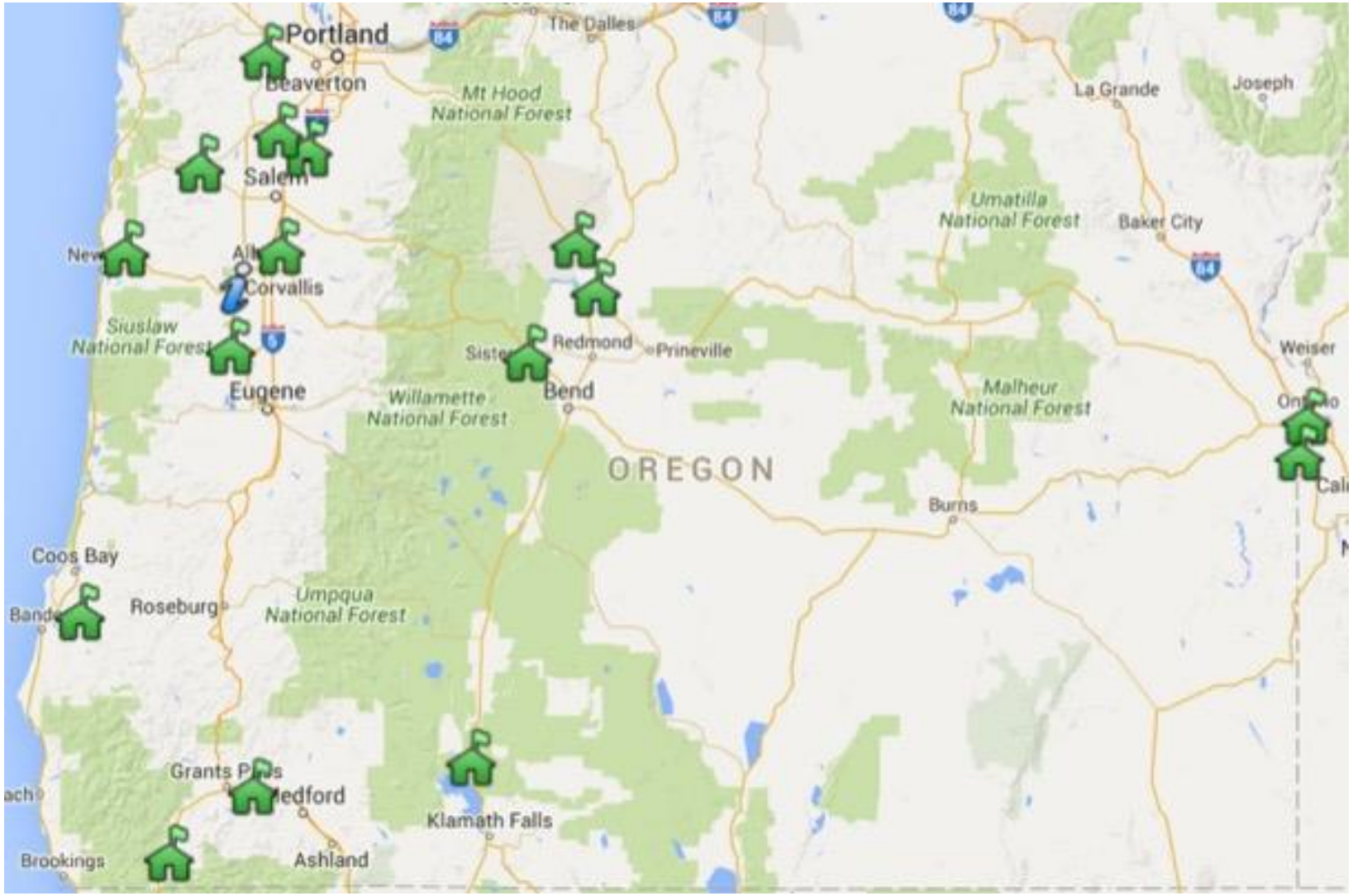
What is the SMILE Program?

Mission: to increase underrepresented students' success in STEM degree programs and careers, and to deliver high-quality teacher professional development.

- An after-school science and math club for students in grades 4-12
- An enrichment program for students from groups historically underrepresented in math, science, and engineering
- A fun and supportive place for students to get encouragement to stay in school and attend college.



SMILE is a statewide program

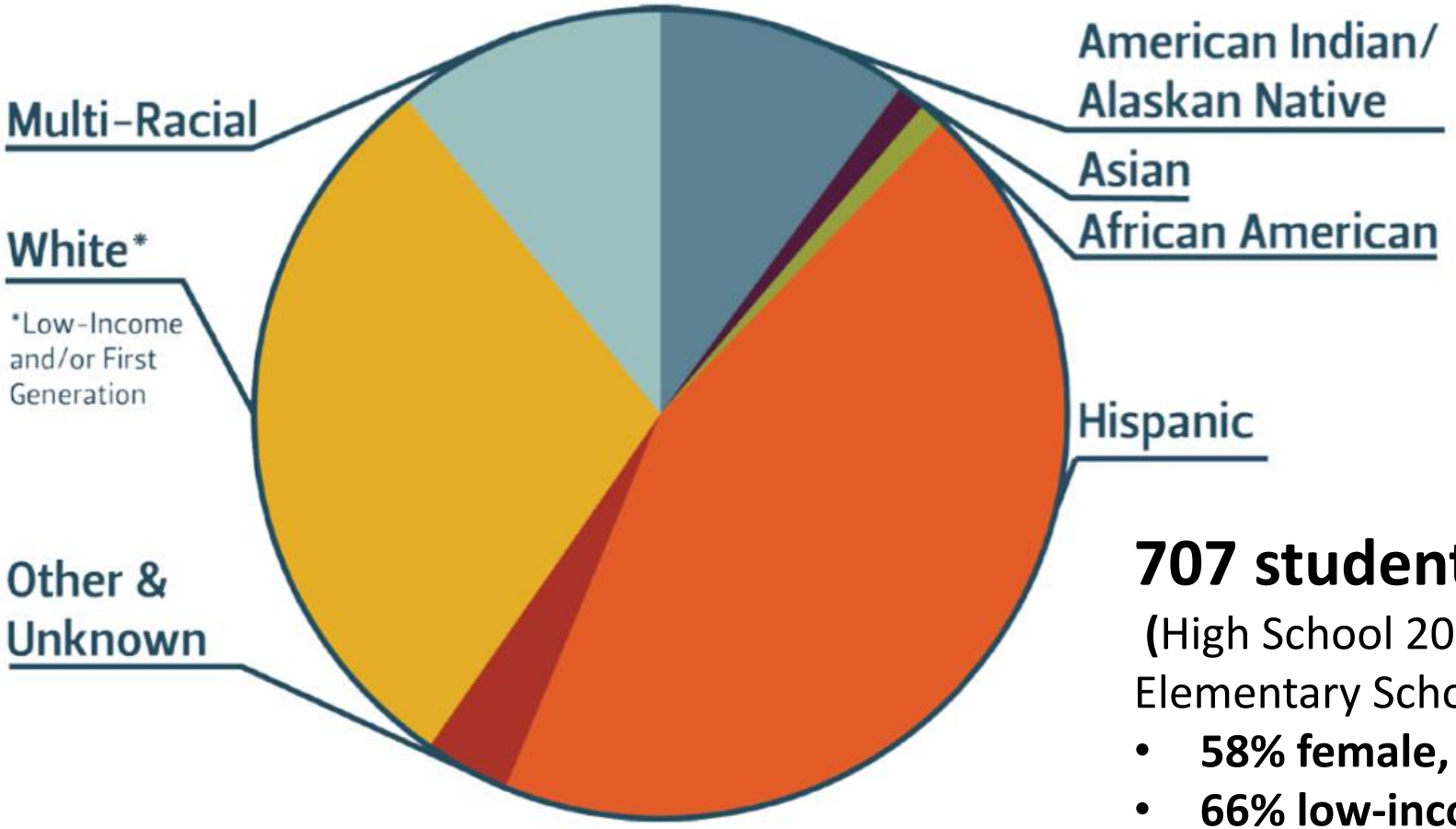


SMILE Facts

- Founded in 1987
- Located in rural underserved communities
- STEM content focus
- Severed over 17,800 Students
- Served over 1,600 Teachers

Currently in 18 communities and 39 clubs

SMILE serves diverse students

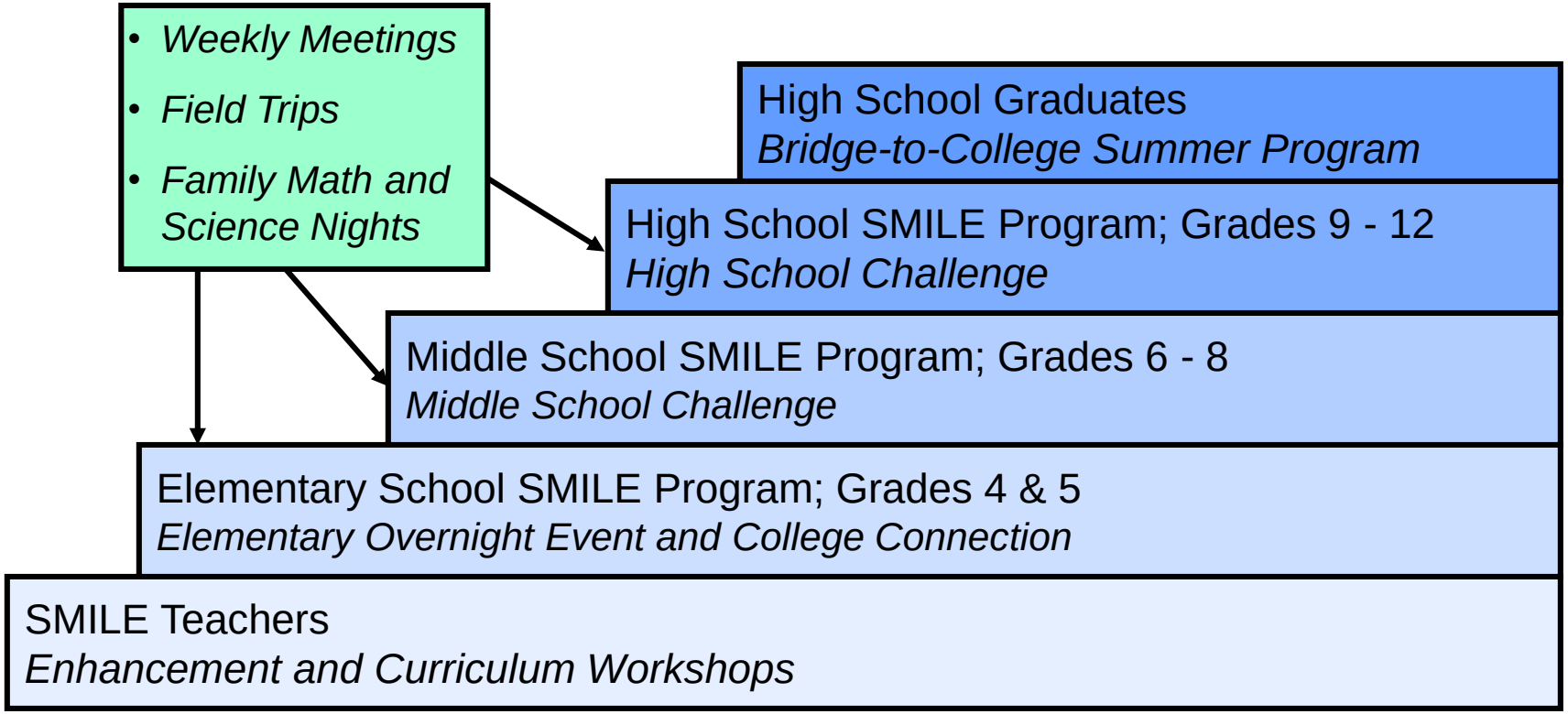


707 students in 2016

(High School 202, Middle School 322, Elementary School 183)

- **58% female,**
- **66% low-income**
- **81% first generation to college**
- **100% of seniors graduated in 2016**

SMILE Structure and Programing

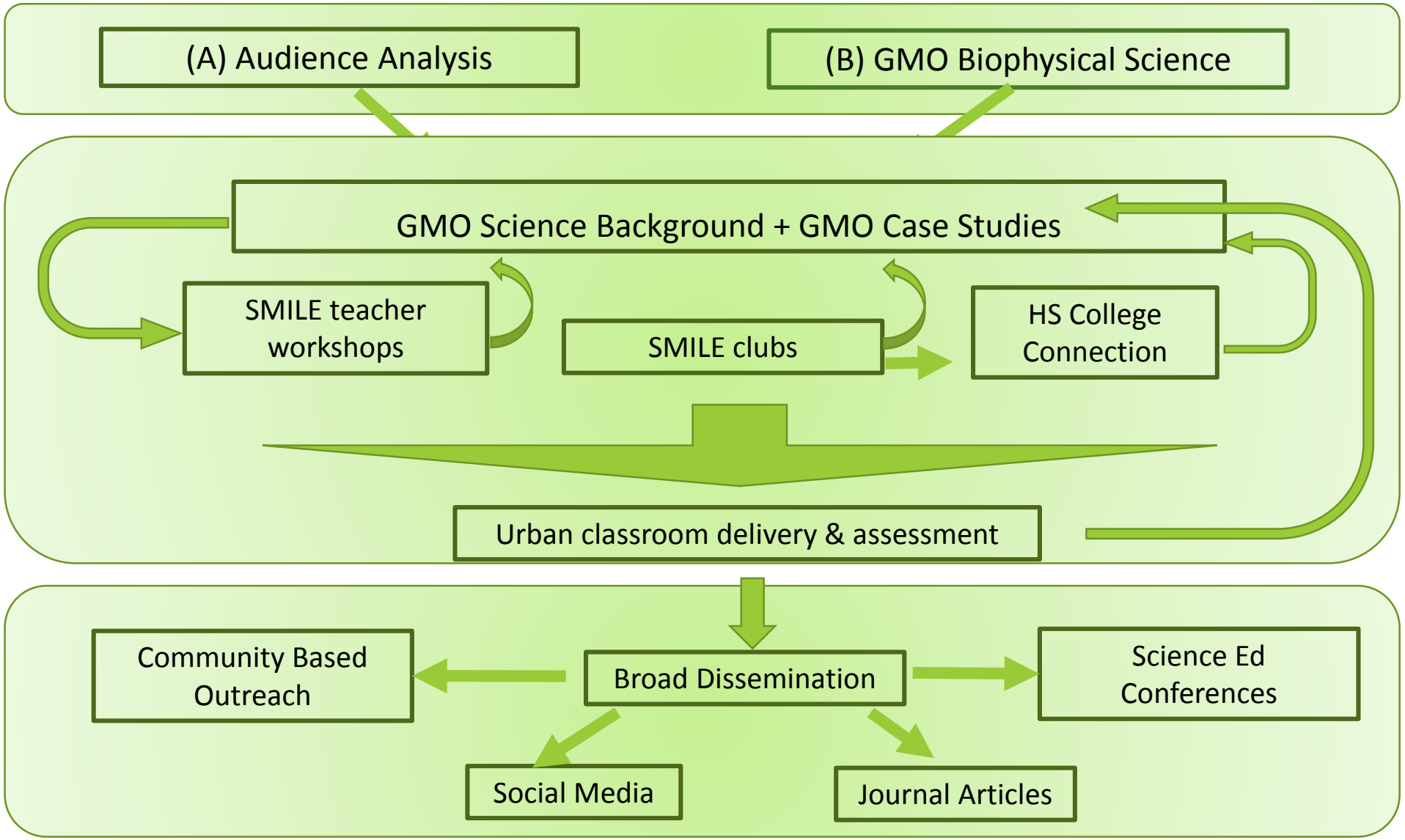


Broader Impacts & Social Science Overview

Phase I: Audience Assessment & Content Development (Years 1-2)

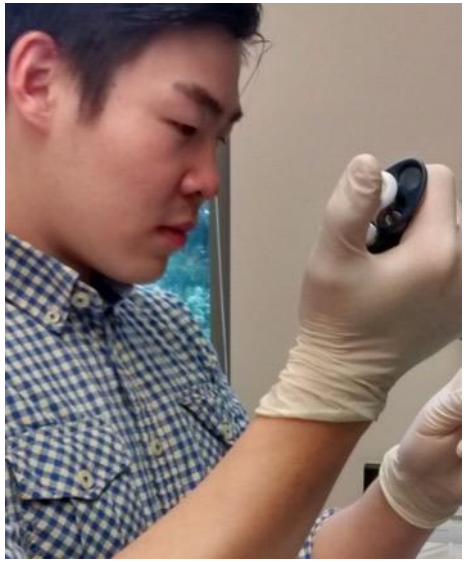
Phase II: Case Study Development: Assess, Improve, Evaluate (Years 2-4)

Phase III: Disseminate (Years 4-5)



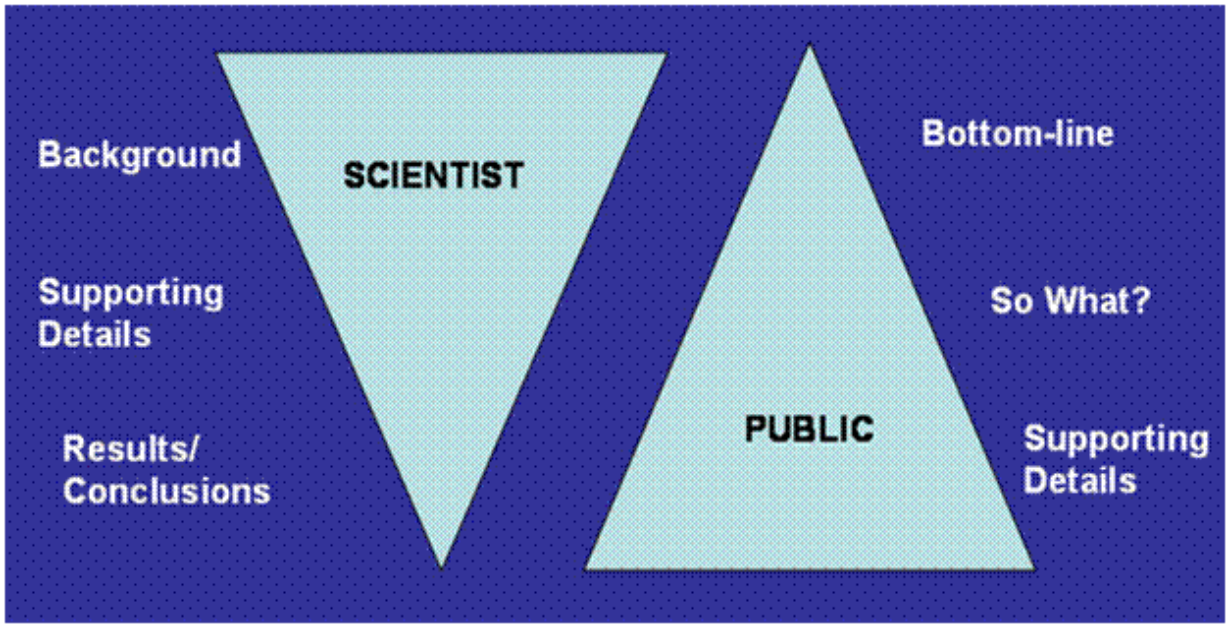
Phase I: Audience Assessment and Content Selection (Years 1-2)

- Audience assessment to inform case studies
  - SMILE workshop with all teachers (August 2017)
  - All SMILE clubs (fall 2017)
  - Explore **teachers'** and **students'** conceptualizations (beliefs, emotions, attitudes) regarding genetic technologies & GMOs



Phase I: Audience Assessment and Content Selection (Years 1-2)

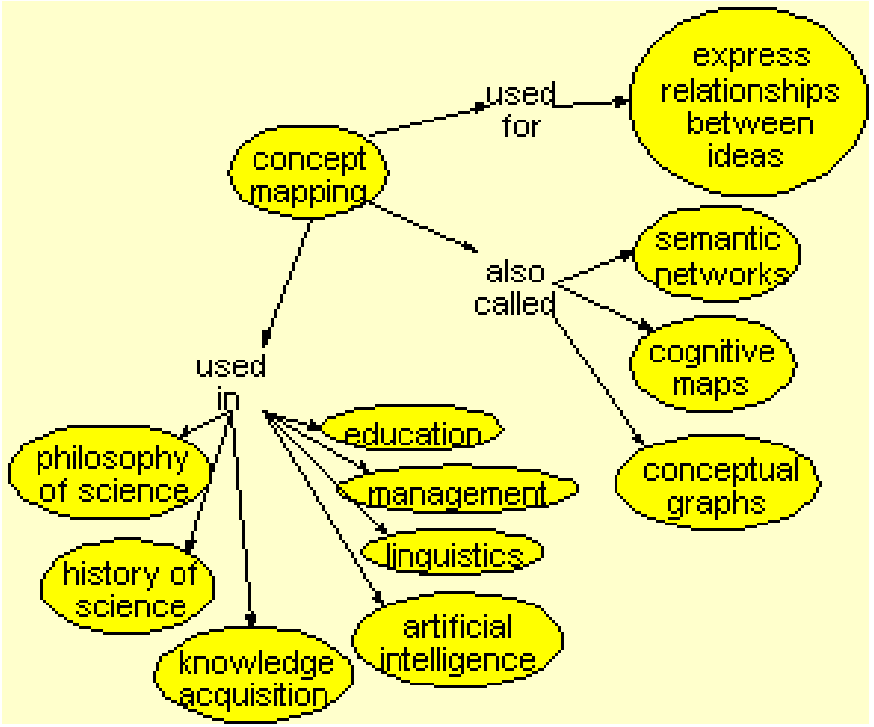
- Review literature on effective techniques to increase student engagement and ability to think critically



AAAS

Phase I: Audience Assessment and Content Selection (Years 1-2)

- Evaluate data collection methods (interviews, think-aloud, cognitive mapping, etc).
- Develop techniques to characterize integrative complexity of mental models
- Interface with GMO biophysical science staff on content and activities for case studies
- Review existing case studies



Rob Kremer, U Calgary

Phase I: Audience Assessment and Content Selection (Years 1-2)

- Outcomes
  - Best practices to guide case content, format, and activities
  - Reliable measurement instruments for later pre-/post-tests
  - Identification of key understandings and misconceptions to inform case study development

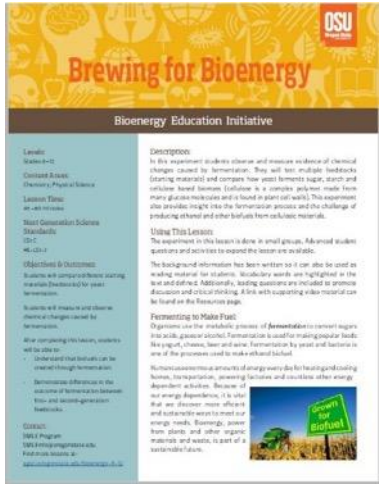
Phase II: Case study development (Years 2-4)

Our curriculum development process is in partnership with SMILE teachers

1. Establish audience assessment
2. Develop case studies
3. Pilot and evaluate case studies in SMILE afterschool clubs
4. Use teacher and student feedback to improve case studies
5. Prepare for publication and other forms of dissemination

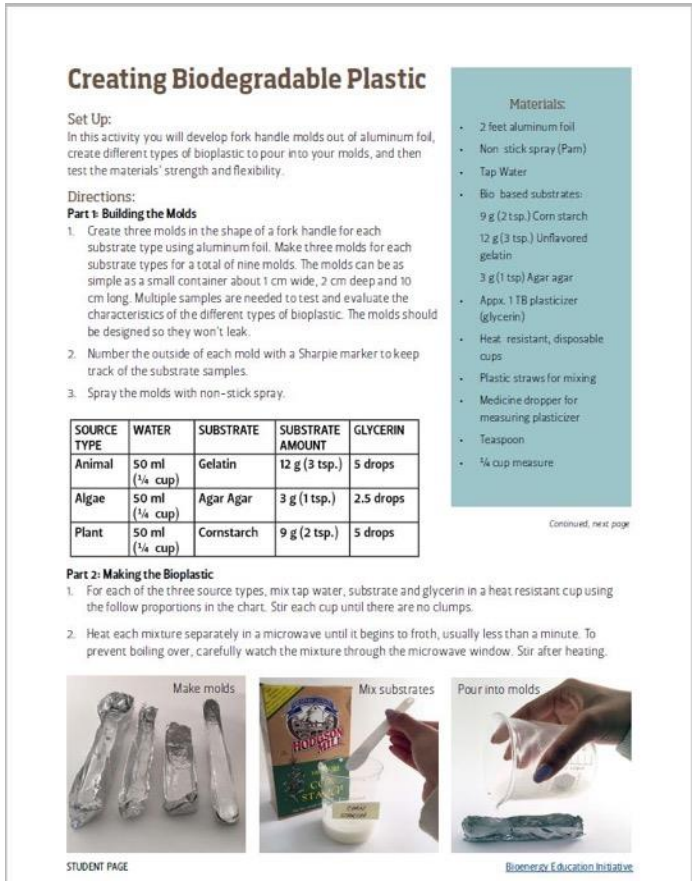
Developing curriculum in partnership with teachers is more likely to be used in the classroom

- Lessons are developed with strong STEM connections
- K-12 educators become more comfortable with content
- Researchers gain a better understanding of the needs of K-12 educators
- Resulting lessons are classroom-ready



Teachers pilot lessons and provide feedback to improve lessons

- Easy to read format
- Teacher and student versions
- Introductory Power Point presentations
- In-depth background reading
- Student worksheets
- Multiple scaffolding options
- Photos and diagrams support lesson directions and concepts



Creating Biodegradable Plastic

Set Up:
In this activity you will develop fork handle molds out of aluminum foil, create different types of bioplastic to pour into your molds, and then test the materials' strength and flexibility.

Directions:

Part 1: Building the Molds

1. Create three molds in the shape of a fork handle for each substrate type using aluminum foil. Make three molds for each substrate types for a total of nine molds. The molds can be as simple as a small container about 1 cm wide, 2 cm deep and 10 cm long. Multiple samples are needed to test and evaluate the characteristics of the different types of bioplastic. The molds should be designed so they won't leak.
2. Number the outside of each mold with a Sharpie marker to keep track of the substrate samples.
3. Spray the molds with non-stick spray.

SOURCE TYPE	WATER	SUBSTRATE	SUBSTRATE AMOUNT	GLYCERIN
Animal	50 ml (¼ cup)	Gelatin	12 g (3 tsp.)	5 drops
Algae	50 ml (¼ cup)	Agar Agar	3 g (1 tsp.)	2.5 drops
Plant	50 ml (¼ cup)	Cornstarch	9 g (2 tsp.)	5 drops

Materials:

- 2 feet aluminum foil
- Non stick spray (Pam)
- Tap Water
- Bio based substrates:
  - 9 g (2 tsp.) Corn starch
  - 12 g (3 tsp.) Unflavored gelatin
  - 3 g (1 tsp) Agar agar
- Appx. 1 TB plasticizer (glycerin)
- Heat resistant, disposable cups
- Plastic straws for mixing
- Medicine dropper for measuring plasticizer
- Teaspoon
- ¼ cup measure

Continued, next page

Part 2: Making the Bioplastic

1. For each of the three source types, mix tap water, substrate and glycerin in a heat resistant cup using the follow proportions in the chart. Stir each cup until there are no clumps.
2. Heat each mixture separately in a microwave until it begins to froth, usually less than a minute. To prevent boiling over, carefully watch the mixture through the microwave window. Stir after heating.

Make molds | Mix substrates | Pour into molds

Case studies will be linked to existing science, math, and literacy standards

Next Generation Science Standards (NGSS)

- Based on the Nation Research Councils Frameworks for K-12 Science Educators (2011)
- Supports learning in three dimensions:
  - Practices
  - Crosscutting Concepts
  - Disciplinary Core Ideas

Common Core

- Research and evidence based
- Aligned with college and career expectations
- Based on application of content through higher order thinking skills
- Internationally benchmarked

NGSS Three dimensions: Practices, Crosscutting Concepts & Core Ideas defined:

1. Scientific & Engineering Practices

1. Asking questions (for science & defining problems (for engineering)
2. Developing and using models
3. Planning & carrying out investigations
4. Analyzing & interpreting data
5. Using mathematics & computational thinking
6. Constructing explanations (for science) & designing solutions (for engineering)
7. Engaging in argument from evidence
8. Obtaining, evaluating, & communicating information

2. Crosscutting Concepts

1. Patterns
2. Cause & effect: Mechanism & explanation
3. Scale, proportion, & quantity
4. Systems & system models
5. Energy & matter: Flows, cycles, & conservation
6. Structure & function
7. Stability and change

3. Disciplinary Core Ideas

Physical Sciences

PS 1: Matter & its interactions

PS 2: Motion & stability: Forces & interactions

PS 3: Energy

PS 4: Waves & their applications in technologies for information transfer

Life Sciences

LS 1: From molecules to organisms: Structures & processes

LS 2: Ecosystems: Interactions, energy, & dynamics

LS 3: Heredity: Inheritance & variation of traits

LS 4: Biological evaluation: Unity & diversity

Earth and Space Sciences

ESS 1: Earth's place in the universe

ESS 2: Earth's systems

ESS 3: Earth and human activity

Engineering, Technology, & the Applications of Science

ETS 1: Engineering design

ETS 2: Links among engineering, technology, science, & society

Example of NGSS Performance Expectation

Students who demonstrate understanding can:

HS-LS3-1. **Ask questions to clarify relationships about the role of DNA and chromosomes in coding the instructions for characteristic traits passed from parents to offspring.** [*Assessment Boundary: Assessment does not include the phases of meiosis or the biochemical mechanism of specific steps in the process.*]

The performance expectation above was developed using the following elements from the NRC document *A Framework for K-12 Science Education*:

Science and Engineering Practices

Asking Questions and Defining Problems
Asking questions and defining problems in 9-12 builds on K-8 experiences and progresses to formulating, refining, and evaluating empirically testable questions and design problems using models and simulations.

- Ask questions that arise from examining models or a theory to clarify relationships.

Disciplinary Core Ideas

LS1.A: Structure and Function

- All cells contain genetic information in the form of DNA molecules. Genes are regions in the DNA that contain the instructions that code for the formation of proteins. *(secondary) (Note: This Disciplinary Core Idea is also addressed by HS-LS1-1.)*

LS3.A: Inheritance of Traits

- Each chromosome consists of a single very long DNA molecule, and each gene on the chromosome is a particular segment of that DNA. The instructions for forming species' characteristics are carried in DNA. All cells in an organism have the same genetic content, but the genes used (expressed) by the cell may be regulated in different ways. Not all DNA codes for a protein; some segments of DNA are involved in regulatory or structural functions, and some have no as-yet known function.

Crosscutting Concepts

Cause and Effect

- Empirical evidence is required to differentiate between cause and correlation and make claims about specific causes and effects.

Connections to other DCIs in this grade-band: N/A

Articulation of DCIs across grade-bands:
MS.LS3.A ; **MS.LS3.B**

Common Core State Standards Connections:
ELA/Literacy -

RST.11-12.1 Cite specific textual evidence to support analysis of science and technical texts, attending to important distinctions the author makes and to any gaps or inconsistencies in the account. *(HS-LS3-1)*

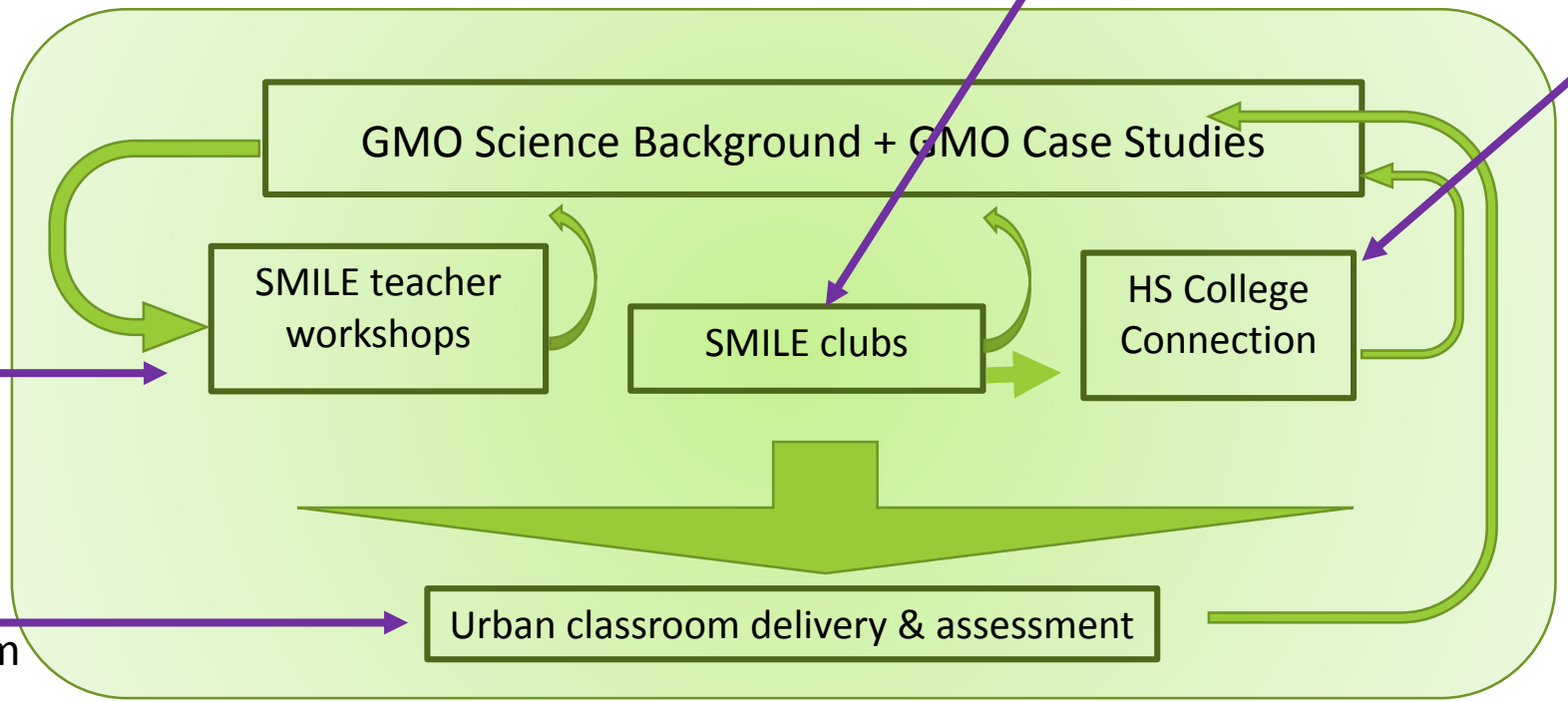
RST.11-12.9 Synthesize information from a range of sources (e.g., texts, experiments, simulations) into a coherent understanding of a process, phenomenon, or concept, resolving conflicting information when possible. *(HS-LS3-1)*

Phase II: Case study development

- Formative assessment
  - Debriefings/teacher workshops after implementation of each case study
  - Refinement of case studies

- Delivered by Teachers
- Audience: Club students

- Delivered by SMILE Faculty & GRA
- Audience: Club students

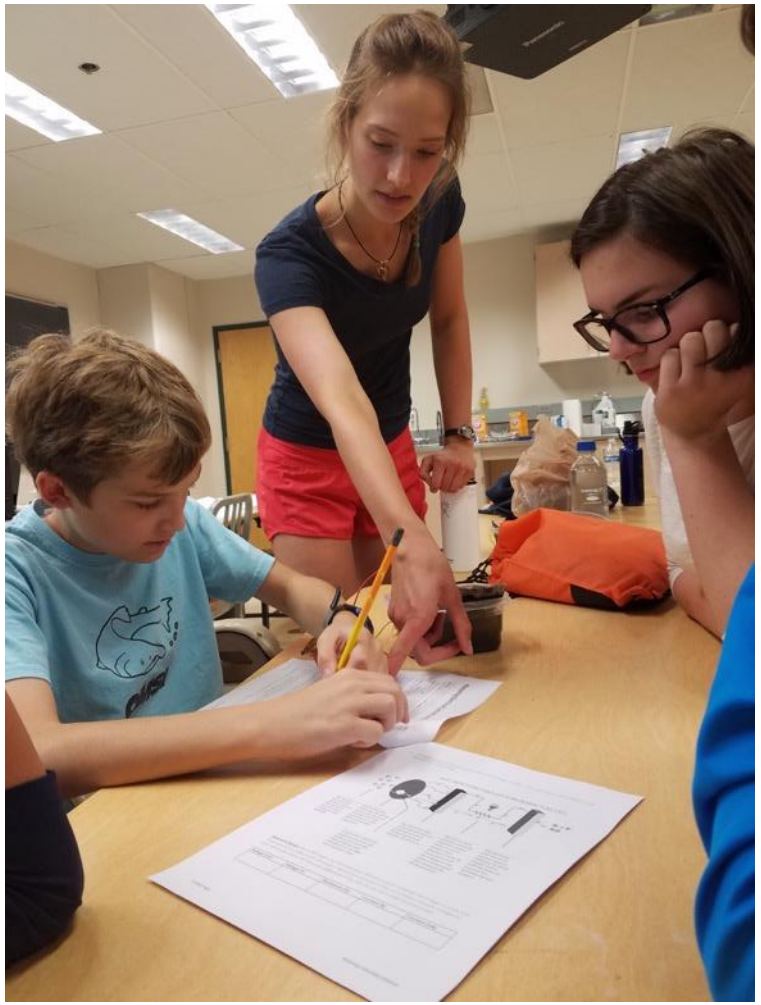


- Delivered by SMILE Faculty & GRA
- Audience: Teachers

- Delivered by Jay
- Audience: Classroom students

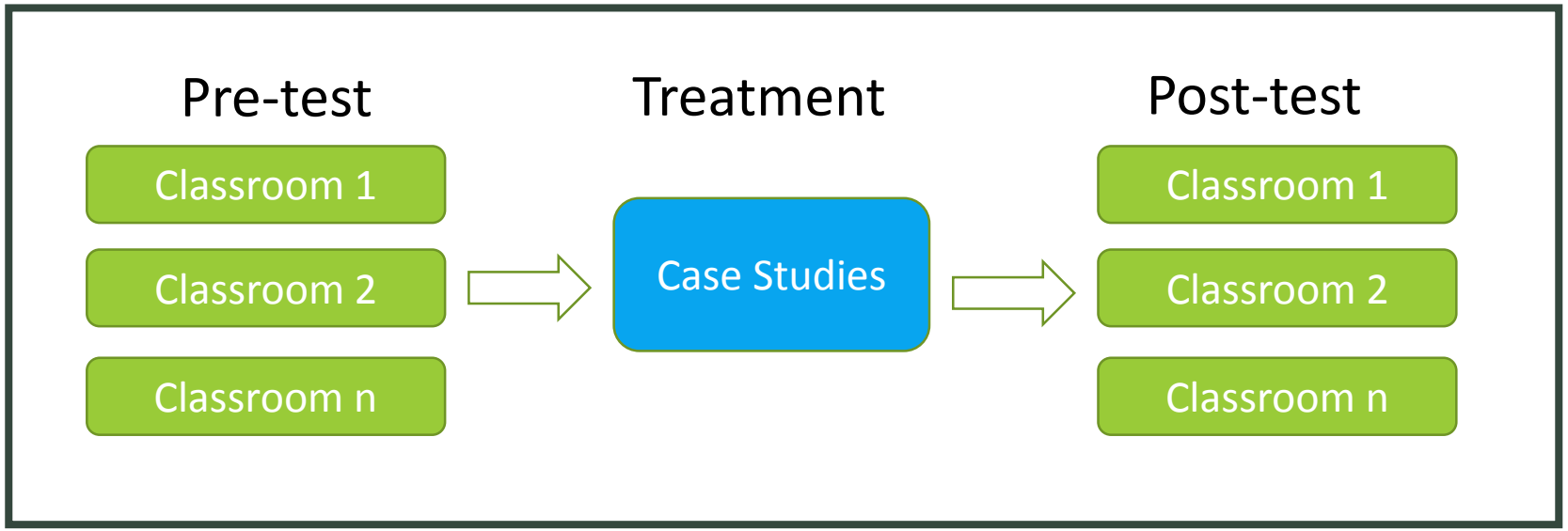
What will a case study consist of

- Focus on a specific GMO
- Sociciscientific based instruction that combine hands-on lab based science with social, political and economic issues
- Consist of 3-5 50 minute lessons
- Focus on student skill development critical thinking, reflection, and discussion
- Utilize existing curriculum when appropriate



Phase II: Summative Evaluation

- Quasi-experimental design
  - Pre-tests/post-test with teachers (before/after workshops)
  - Pre-tests/post-tests with club members (before/after implementation of case studies)



In-depth look at the psychological science

- Hypotheses: Case study curriculum will result in
  - More accurate knowledge about GMOs among students and teachers
  - Increased instructional self-efficacy among teachers
  - Increased complexity of mental models of GMOs (students & teachers)
  - More differentiated attitudes about GMO applications

In-depth look at the psychological science

- The “deficit model” is dead

Dubaiteachmeet.com

In-depth look at the psychological science

Cognitions

("I GMOs are scary"
"GMOs can help alleviate hunger"
"I don't like GMOs")

In-depth look at the psychological science

Naturalnews.com

Cognitions

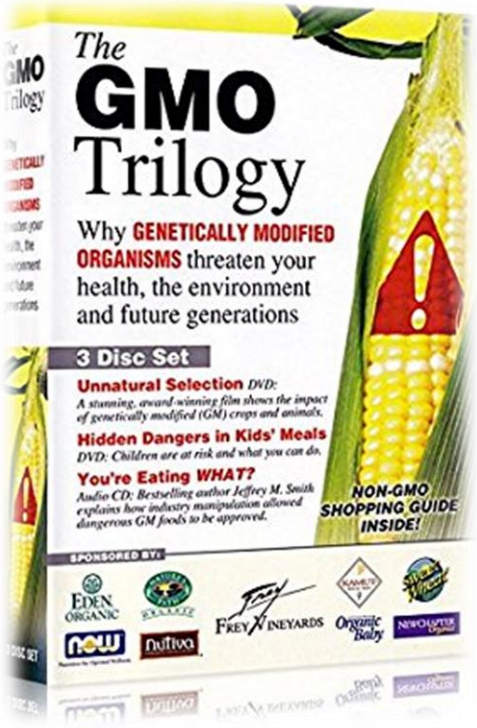
… have different sources

… include emotions

Flaxen saxon

Parentcenterhub.org

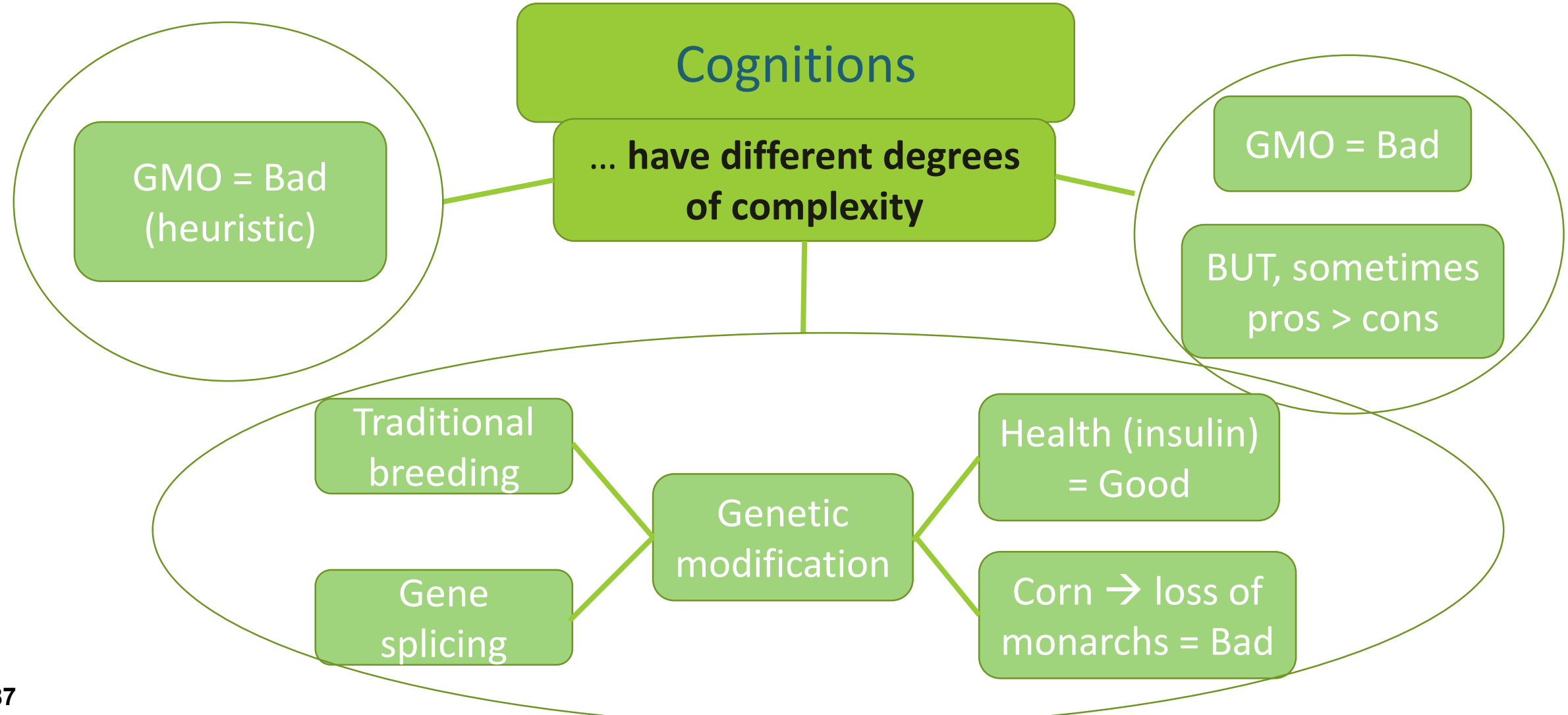
BLS



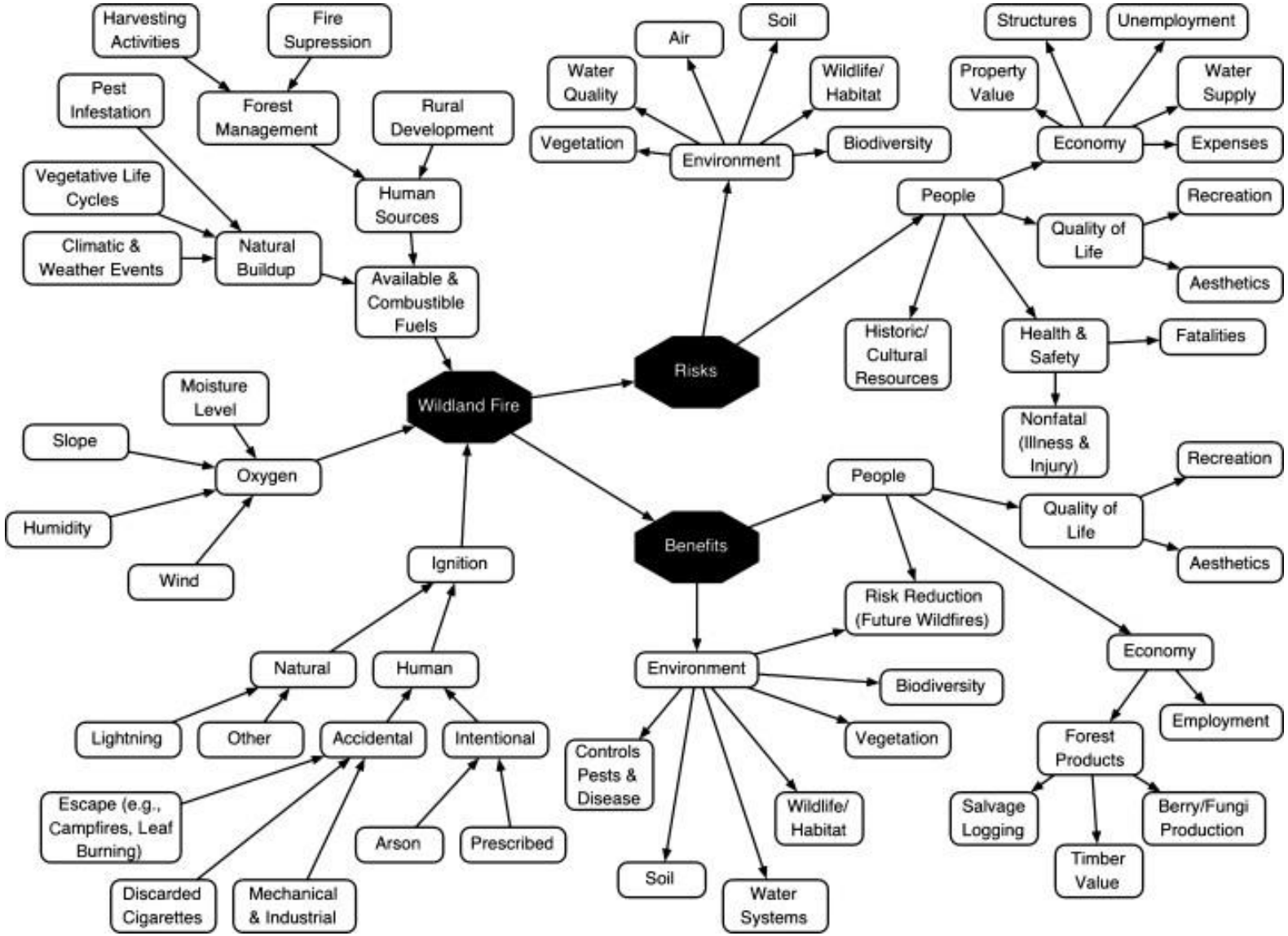
Amazon.com

Naturalnews.com

In-depth look at the psychological science



Example cognitive map



In-depth look at the psychological science

Cognitions

… have different sources

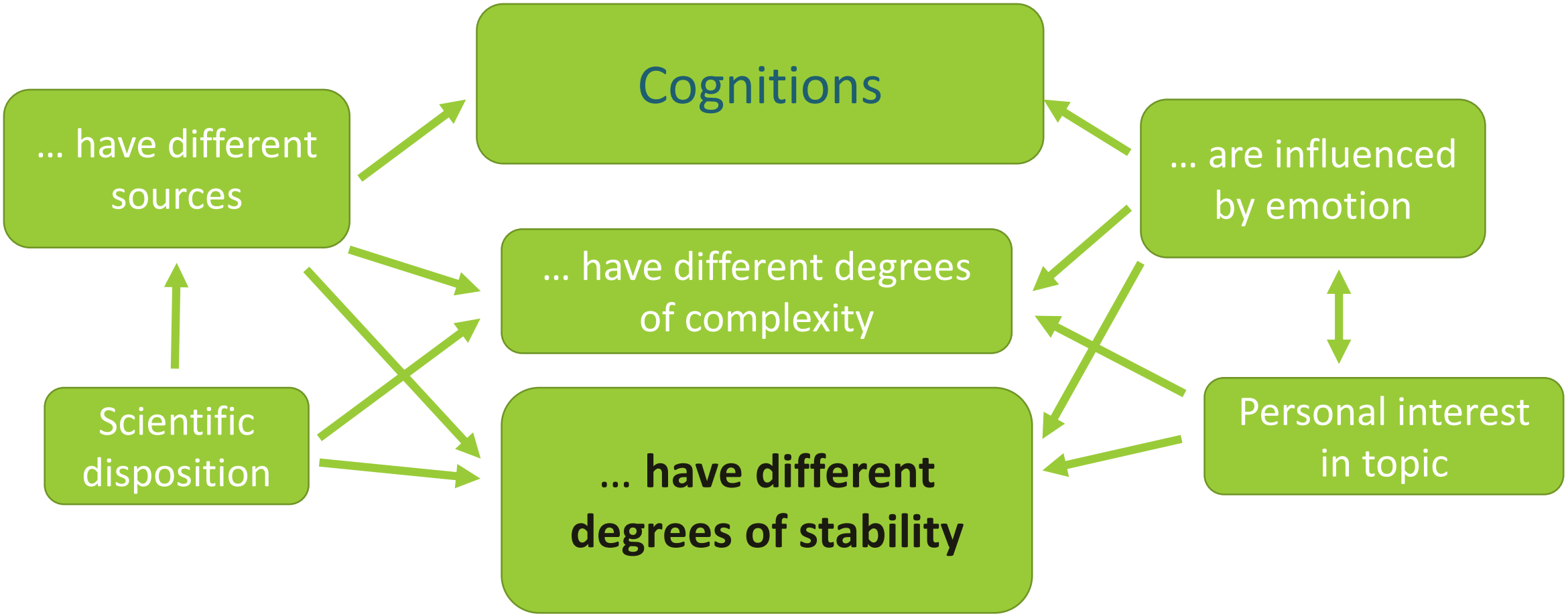
… have different degrees of complexity

… are influenced by emotion

Scientific disposition

Personal interest in topic

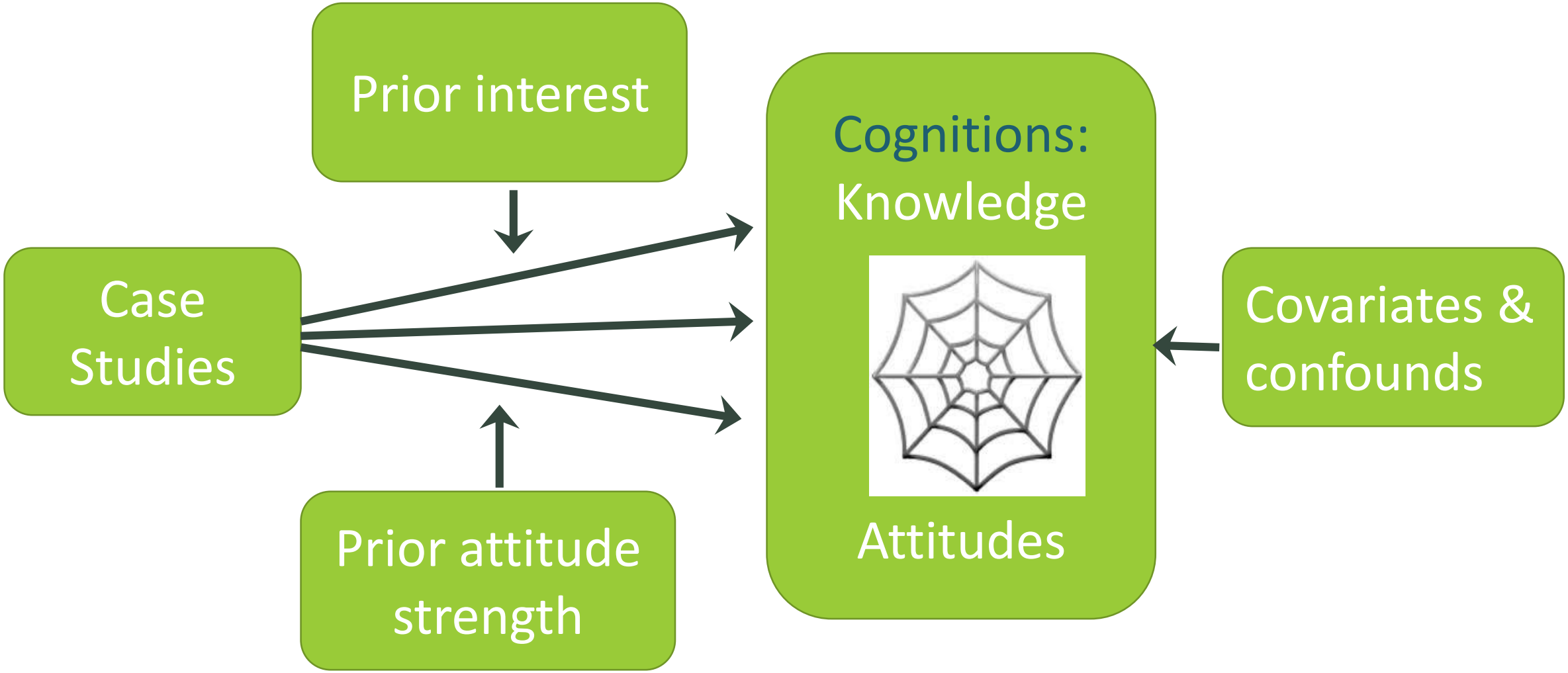
In-depth look at the psychological science



In-depth look at the psychological science

- Case studies will
  - Draw on students' interests
  - Address misconceptions and emotions existing due to prior experiences & information
  - Promote open-minded processing of information

In-depth look at the psychological science



Phase III: dissemination (Years 4-5)

- Write up findings – journal articles
- Finalize all case study materials & publish
- Present at conferences
- Teacher professional development workshops

Outreach and Social Science Project Timeline

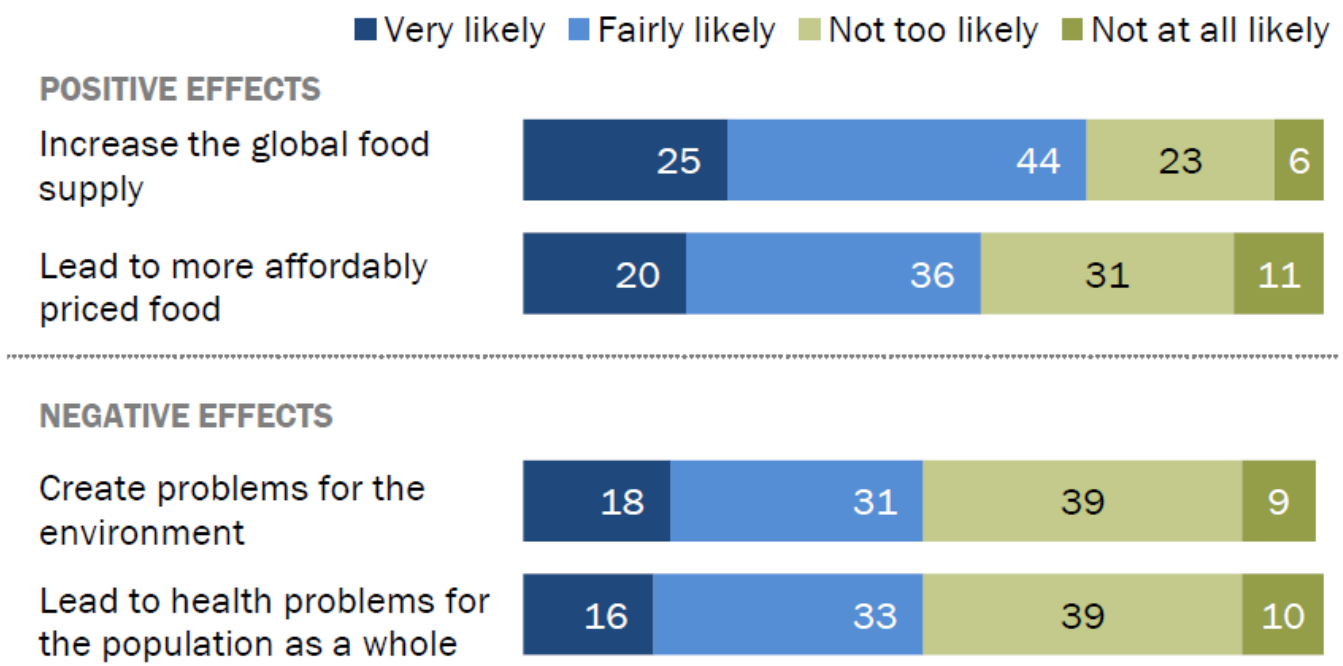
1. Literature Review
   - Winter-Spring 2017
2. Social Science Graduate Student Hired
   - June 2017
3. IRB Approval
   - July 2017
4. Introduction of Project to SMILE Teachers
   - August 2017
5. Audience Assessment with SMILE Clubs
   - Fall 2017
6. Case Study Selection
   - Winter 2018
7. Assessment with Teachers
   - January 2018
8. Case Study Development
   - Winter-Summer 2018
9. Present Case Studies to SMILE Teachers
   - August 2018
10. First Case Study Implementation
    - Fall 2018
11. Second Case Study Implementation
    - Winter 2018
12. Post-survey with SMILE Teachers and Students
    - Spring 2019
13. High School Challenge
    - Spring 2019 or 2020
14. Refine Case Studies
    - Summer 2019
15. Urban School Delivery
    - Fall 2019
16. Publications and Dissemination
    - Spring-Summer 2020

Discussion questions

- We should develop some questions to guide discussion with advisory group:
  - We do not have a control group – is that a fatal flaw?
  - Which control variables are most important to include (given limited bandwidth for pre-test surveys)?
  - How might we go about investigating students' self-awareness about how they process science?

Americans think both positive and negative effects of genetically modified foods are likely

% of U.S. adults who say how likely it is that genetically modified foods will ...

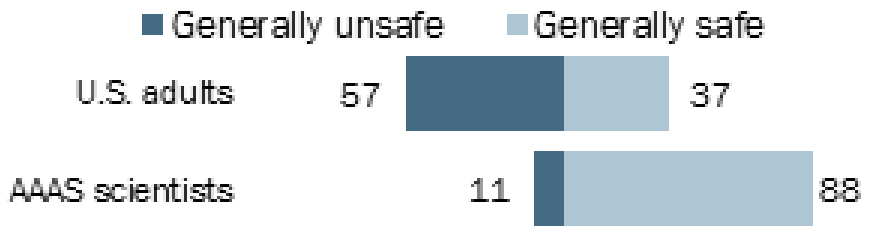


Note: Respondents who did not give an answer are not shown.
Source: Survey conducted May 10-June 6, 2016.
"The New Food Fights: U.S. Public Divides Over Food Science"

PEW RESEARCH CENTER

Eating Genetically Modified Foods

% of each group saying it is generally safe or unsafe to eat genetically modified foods

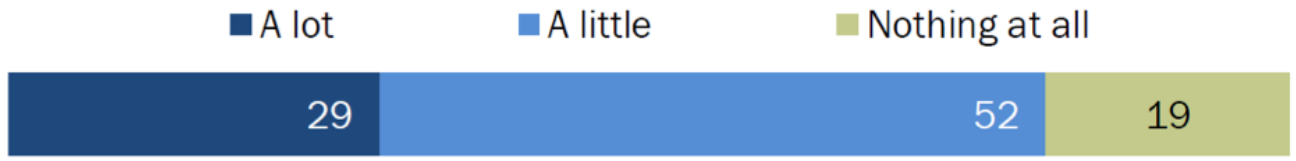


Survey of U.S. adults Aug. 15-25, 2014. Q38. AAAS scientists survey Sept. 11 – Oct. 13, 2014.Q28 Those saying don't know or giving no answer are not shown

PEW RESEARCH CENTER

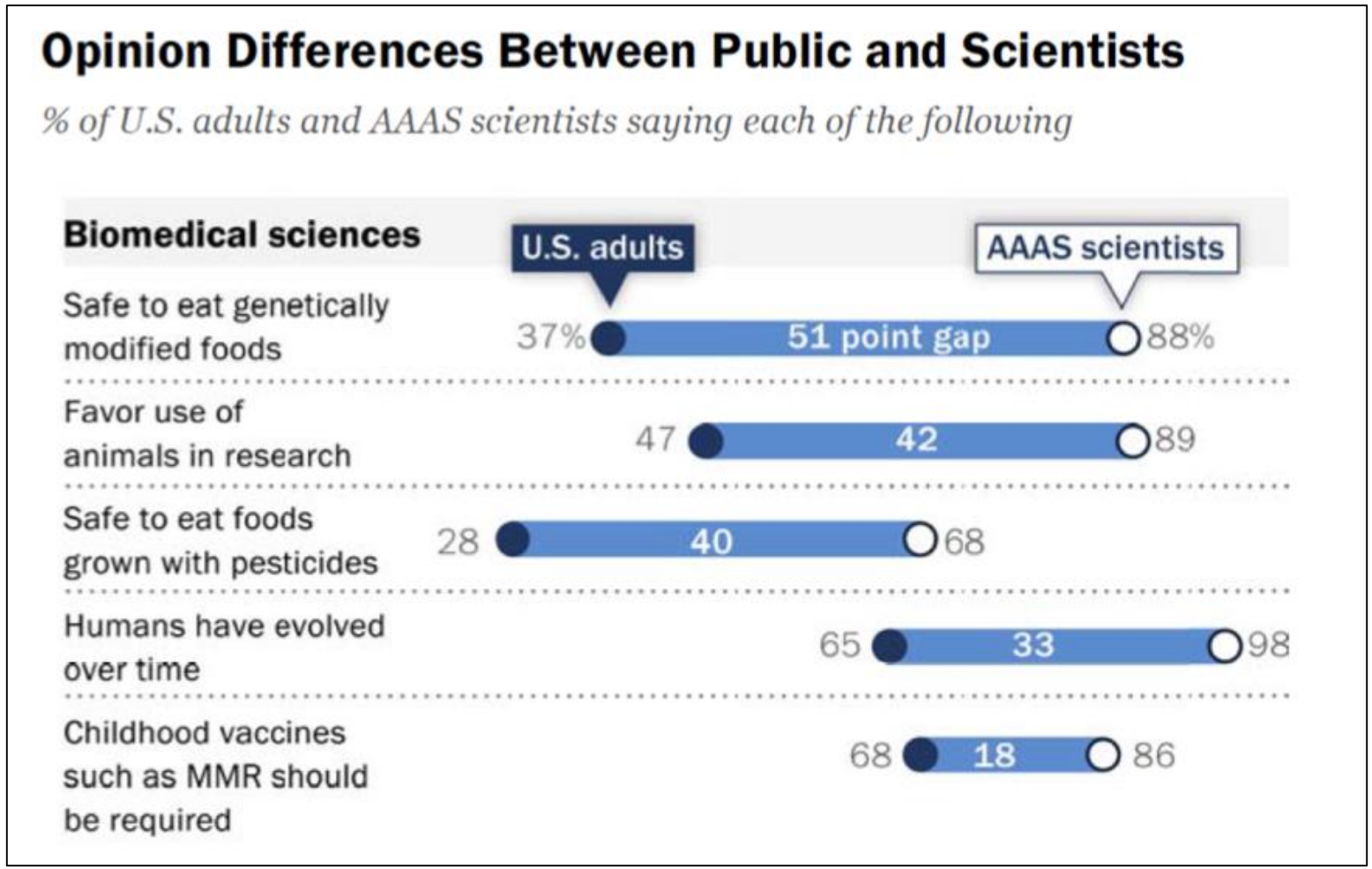
Some 29% of Americans have heard a lot about GM foods; 19% have heard nothing

% of U.S. adults who say they have heard or read ____ about foods with genetically modified ingredients

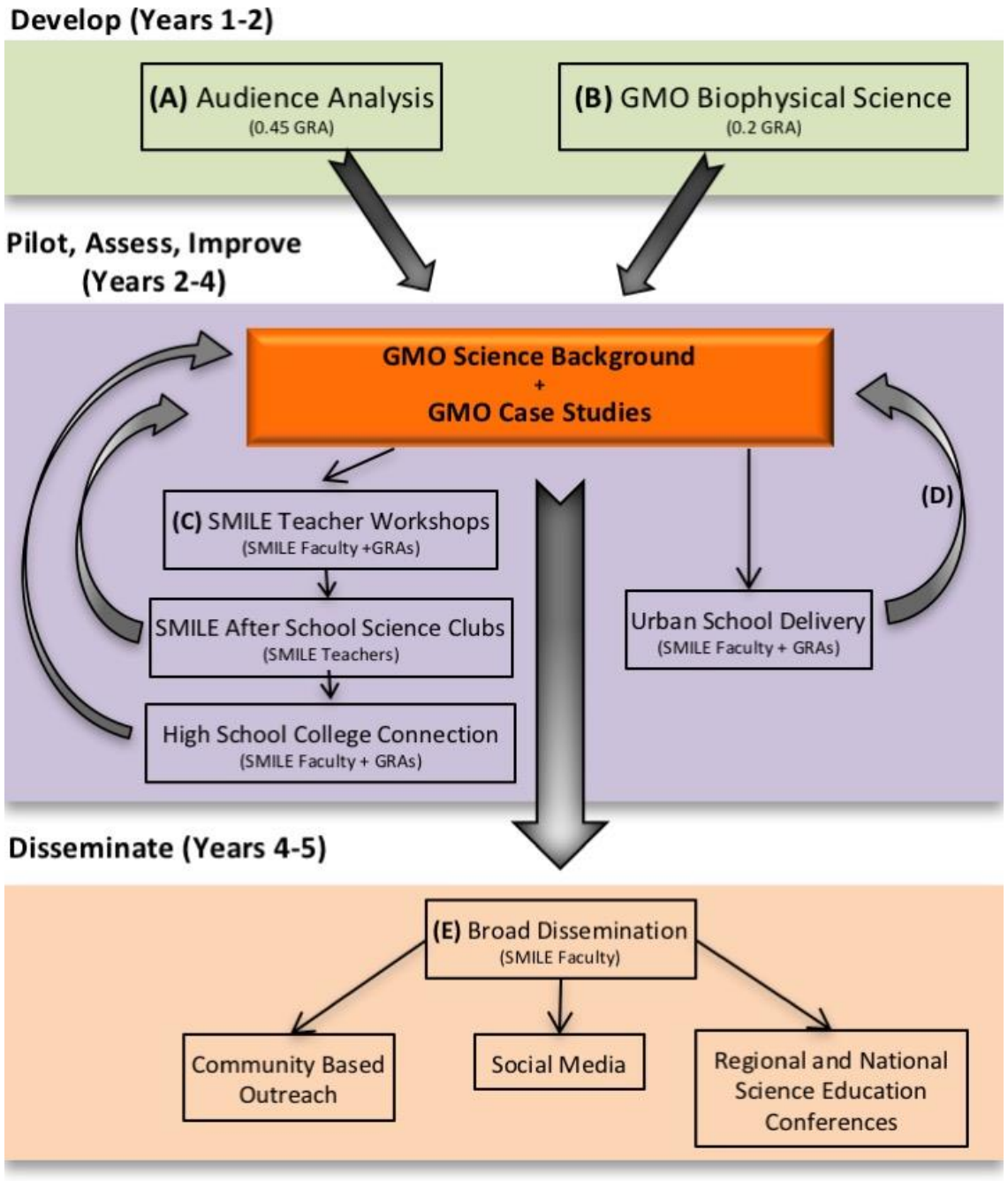


Note: Respondents who did not give an answer are not shown.
Source: Survey conducted May 10-June 6, 2016.
"The New Food Fights: U.S. Public Divides Over Food Science"

PEW RESEARCH CENTER


Opinion Differences Between Public and Scientists
% of U.S. adults and AAAS scientists saying each of the following
Biomedical sciences
U.S. adults
AAAS scientists
Safe to eat genetically modified foods
37%
51 point gap
88%
Favor use of animals in research
47
42
89
Safe to eat foods grown with pesticides
28
40
68
Humans have evolved over time
65
33
98
Childhood vaccines such as MMR should be required
68
18
86

Broader Impacts Overview



Home

Fuxin Li, Zheng Zhou, & Jialin Yuan present…

"Automatic Explants Segmentation From Images"

Automatic explants segmentation from images

Fuxin Li

Zheng Zhou

Jialin Yuan

School of Electrical Engineering and Computer Science

Dec. 22 2016

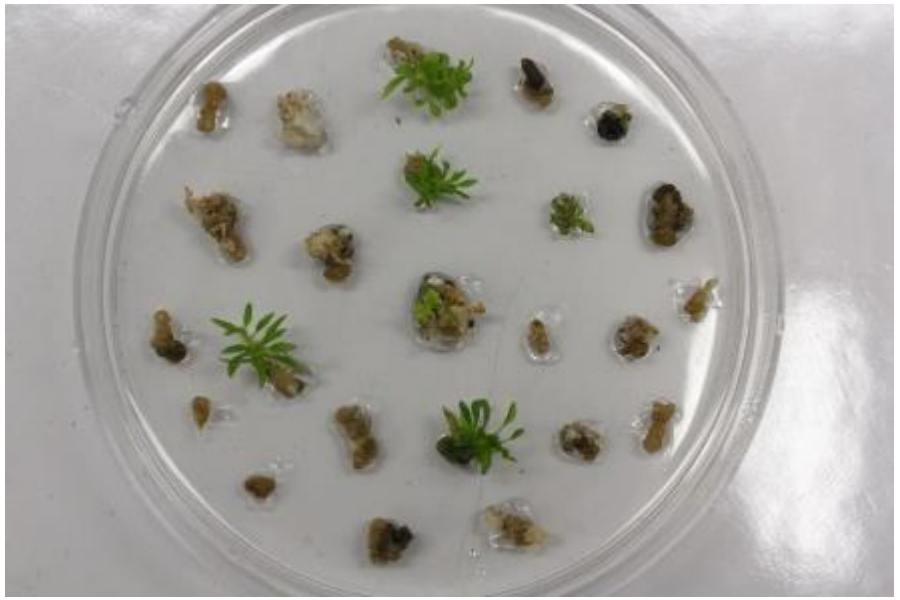
Oregon State UNIVERSITY

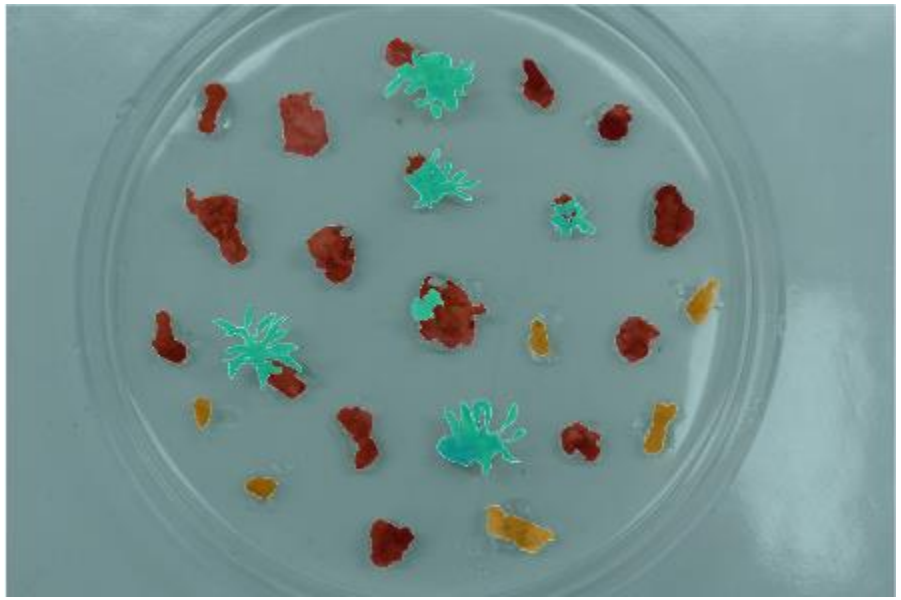
Agenda

- **Project Goals**
- Background info
  - Semantic Segmentation
  - Instance-Level Segmentation
- Project Plans
- Proposed Timeline

Goal of Image Data Analysis

- Recognizing different types of explants (e.g. shoots, roots, etc.), different phenotypes, segment them exactly and count them
- Develop statistics from object recognition and segmentation for GWAS analysis
- Enable users to easily customize “what is an object of interest”

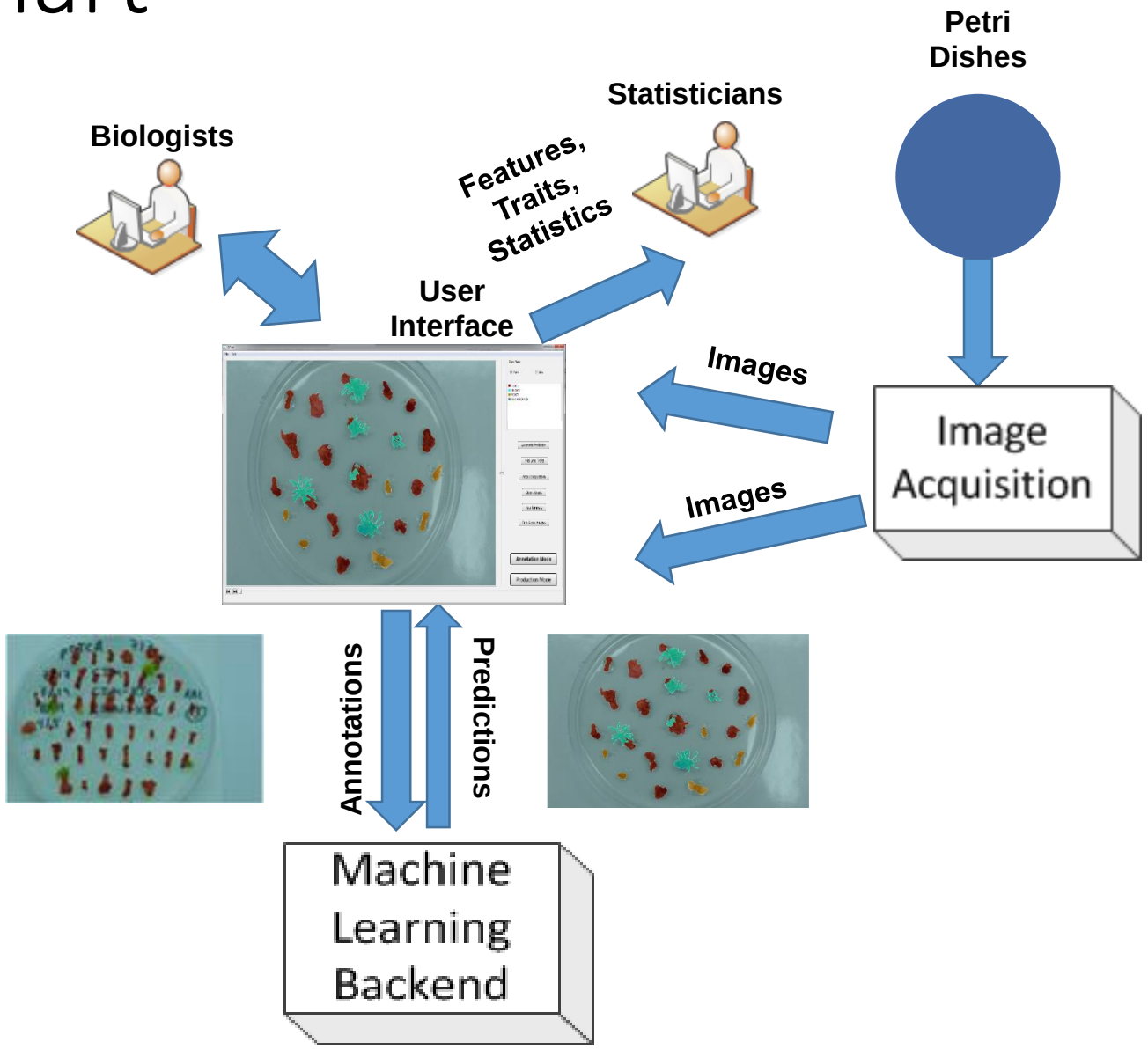




Project Approach

- We propose to automate the image analysis and annotation/segmentation
  - ❑ Experts annotate explant types
  - ❑ Systems learn from those annotations
  - ❑ System makes predictions on new Petri dishes
  - ❑ Expert correct those predictions – creating new annotations
  - ❑ All under a single user interface
- The system is hoped to be general
  - Covers both conventional imaging and hyperspectral images
  - Applicable to new plants
  - Or general instance-level semantic segmentation

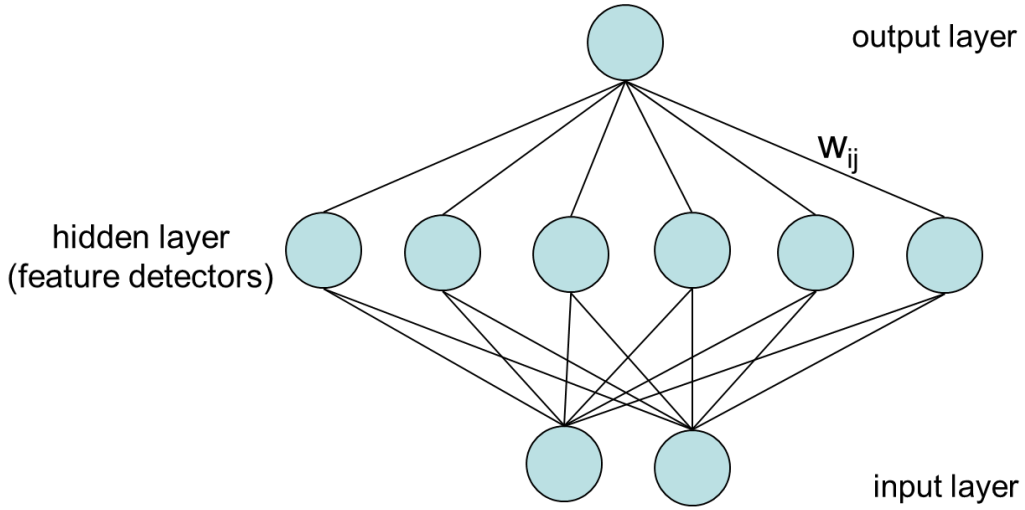
Flow chart


Petri Dishes
Statisticians
Biologists
Features, Traits, Statistics
User Interface
Images
Image Acquisition
Images
Annotations
Predictions
Machine Learning Backend

Agenda

- Project Goals
- **Background info**
  - Deep Learning
  - Semantic Segmentation
  - Instance-Level Segmentation
- Project Plans
- Proposed Timeline

Cutting Edge of Machine Learning: Deep Learning in Neural Networks



Engineering applications:

- Computer vision
- Speech recognition
- Natural Language Understanding
- Robotics

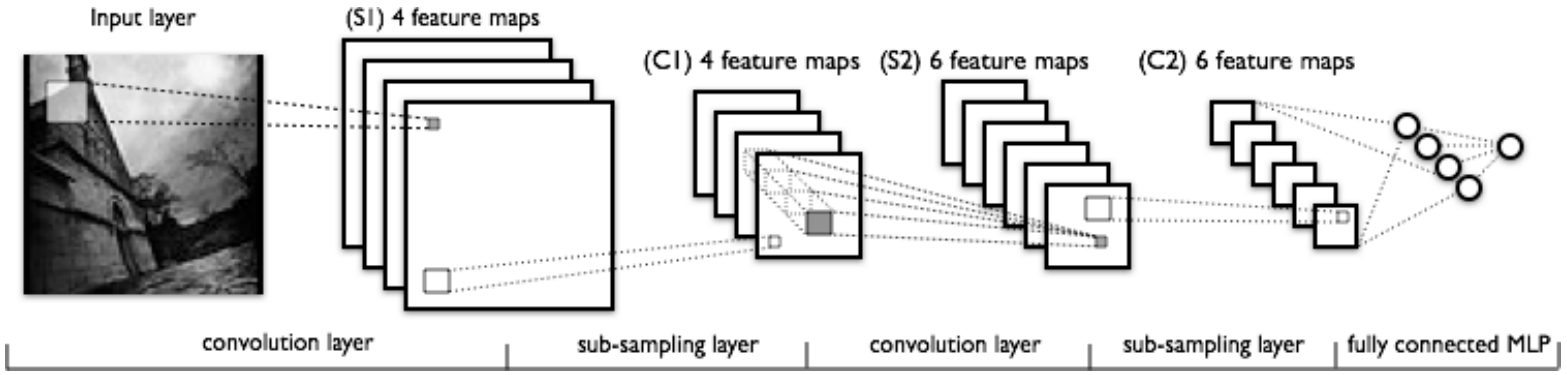


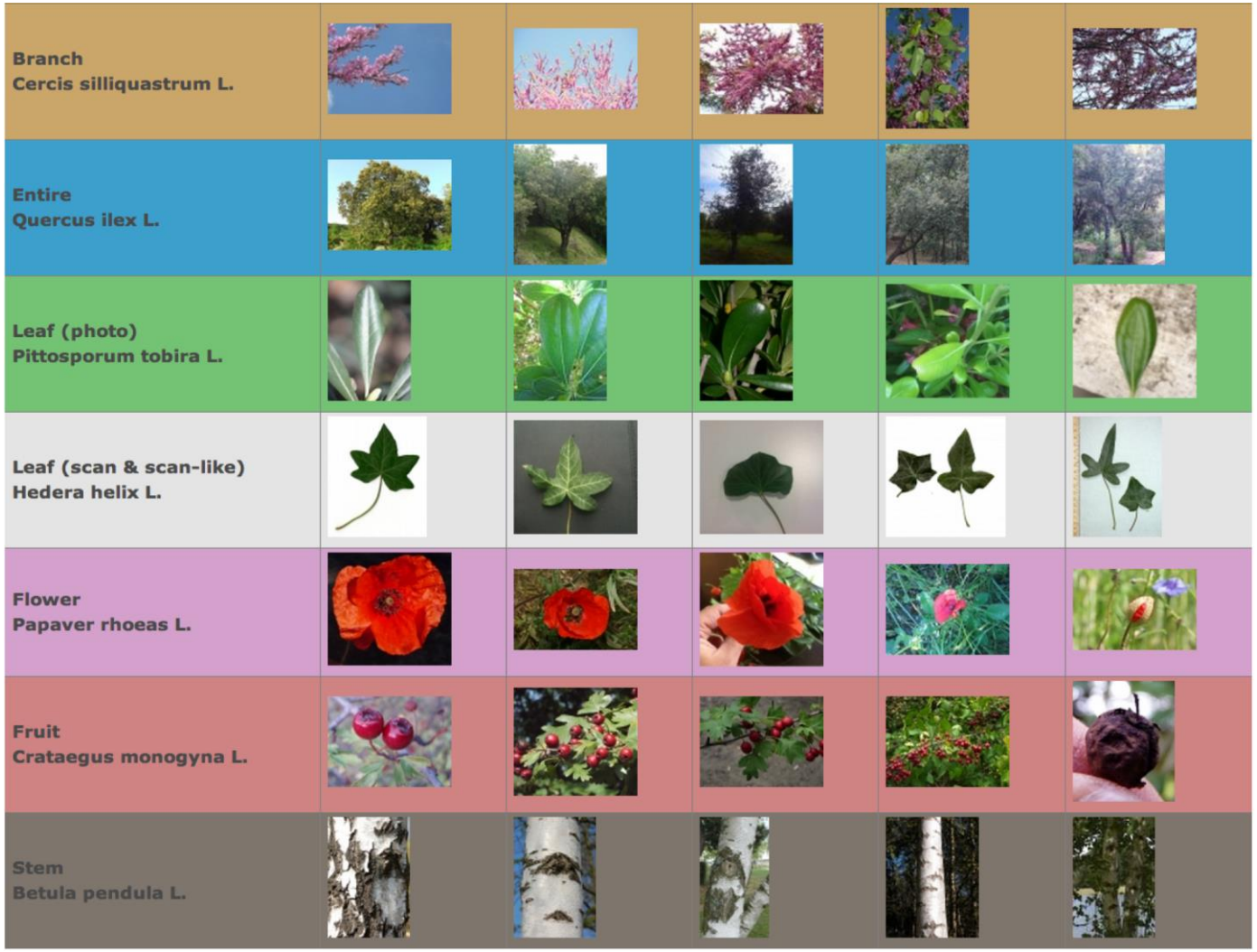
Image Recognition

- ImageNet challenge:
  - 1M natural images, 1000 categories (e.g. cars, SUVs, monitor, airplane, etc.)
  - **Deep learning** algorithms (3.6% error) surpassing human accuracy (5% error)

Relevance to Phenotyping

PlantCLEF 2015

- 113,205 pictures from 1,000 species of trees, herbs and ferns
- Branch, entire, leaf, flower, fruit, stem all available
- Goal is only categorization



PlantCLEF 2015 results

	Run name	Key-words	Score
Deep Convolutional Networks	SNUMED INFO run4	5-fold GoogLeNet Borda+	0,667
	SNUMED INFO run3	5-fold GoogLeNet Borda	0,663
	QUT RV run2	GoogLeNet SoftMax	0,633
	QUT RV run3	GoogLeNet Norm & SoftMax	0,624
	SNUMED INFO run2	GoogLeNet Borda+	0,611
	INRIA ZENITH run1	GoogLeNet Max Pool.	0,609
	SNUMED INFO run1	GoogLeNet Borda	0,604
	INRIA ZENITH run3	Fusion GoogLeNet & Fisher Vectors	0,592
	QUT RV run1	GoogLeNet Sum Pool.	0,563
	ECOUAN run1	CNN Sum Pool.	0,487
Non-Deep Approaches	INRIA ZENITH run2	hand-crafted features + Fisher Vectors	0,300
	MICA run2	Hand-crafted feat. + SVM	0,209
	MICA run1	Hand-crafted feat. + SVM	0,203
	MICA run3	Hand-crafted feat. + SVM	0,203
	SABANCI run2	PCANet (not pretrained)	0,162
	SABANCI run1	PCANet (not pretrained)	0,160
	SABANCI run3	PCANet (not pretrained)	0,158
	UAIC run1	CBIR (LIRE)	0,013

Semantic Segmentation

- Given an image, identify the category and spatial extent of all relevant objects

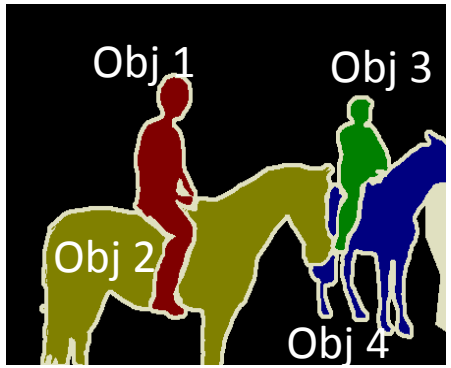
Given:

Image

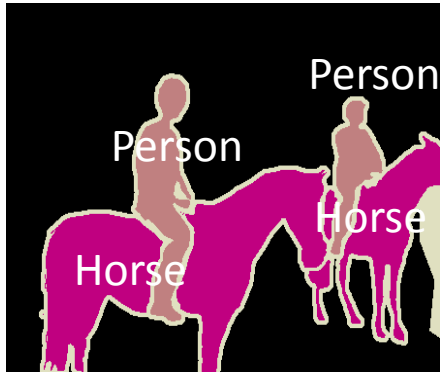
Proposals

Goal:

Instance Label

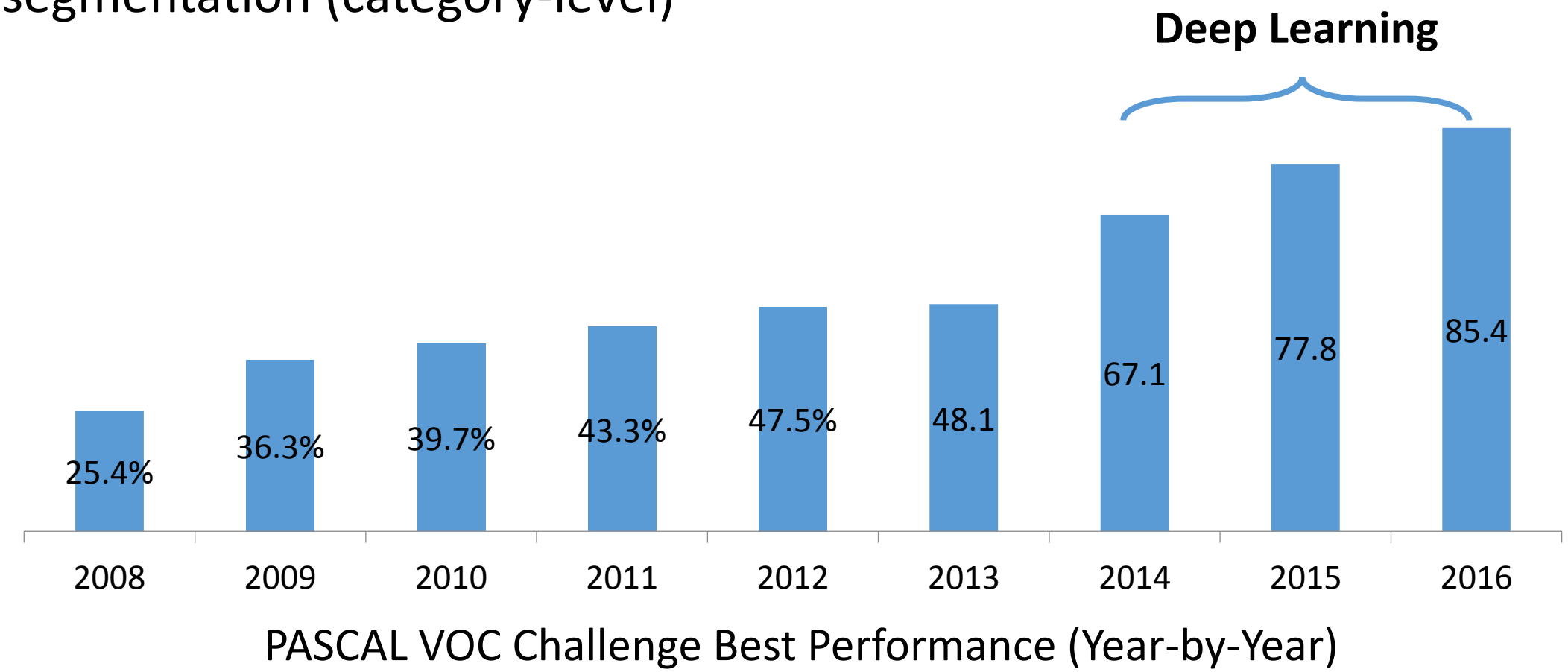


Category Label

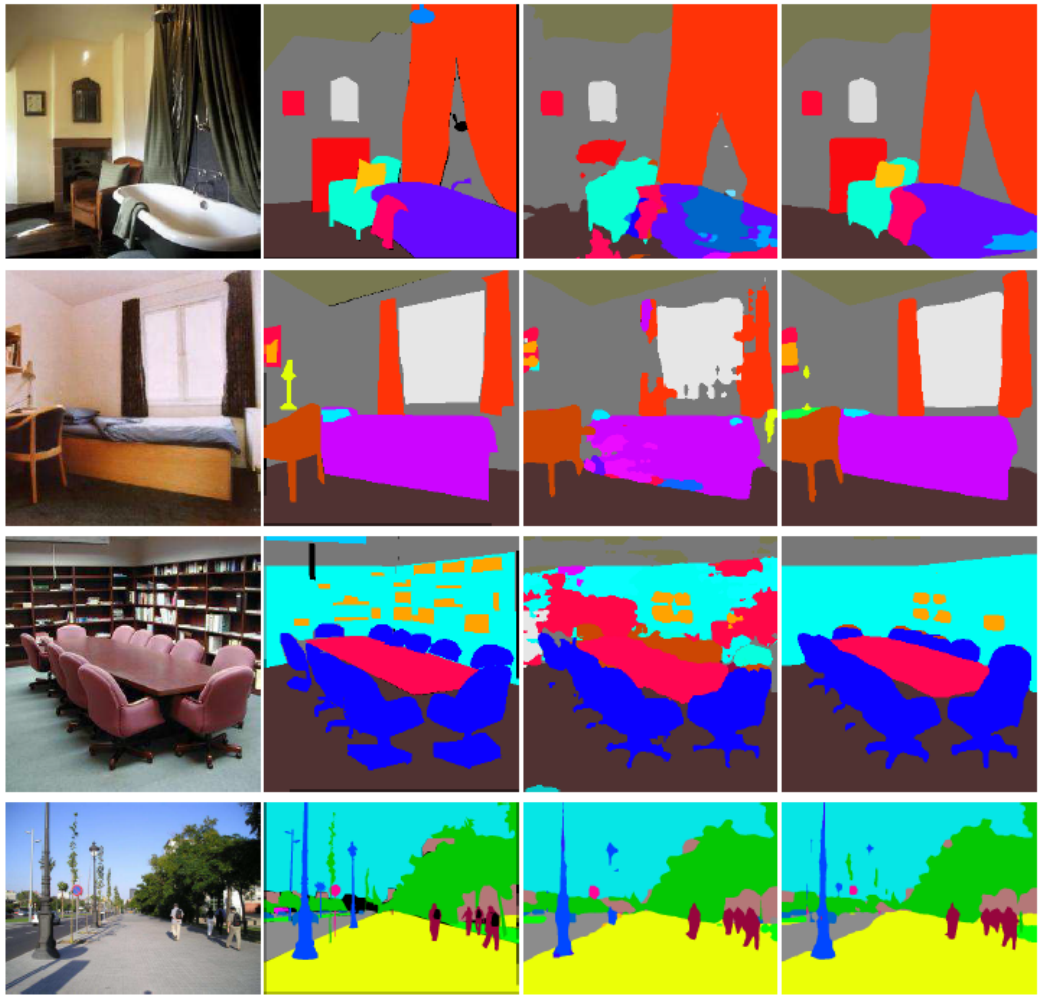


Current Status

- Deep Learning has greatly improved capabilities of semantic segmentation (category-level)



State-of-the-art in Semantic Segmentation

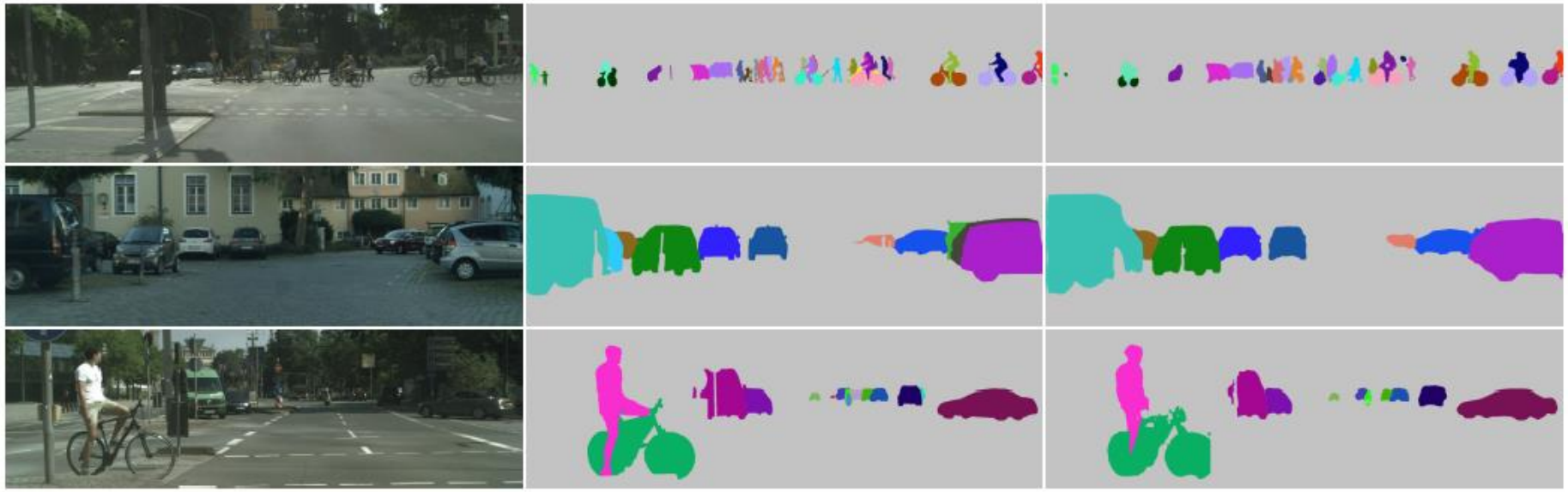


(a) Image (b) Ground Truth (c) Baseline (d) PSPNet

(a) Image (b) Ground Truth (c) Baseline (d) PSPNet

Instance-level Segmentation

- More interesting to us is instance-level segmentation
- Here state-of-the-art is not bad, but not so great yet



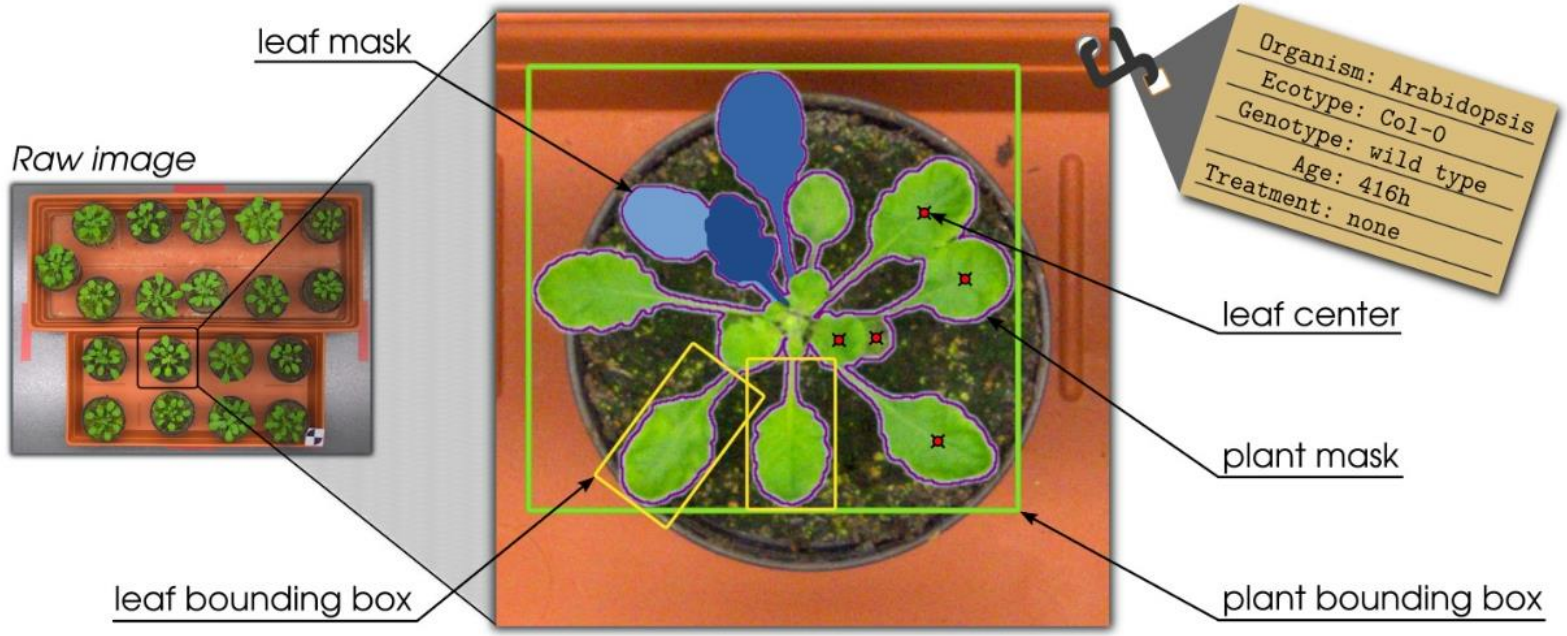
Instance-level Segmentation

- Average pixel-level precision still less than 20% in a very difficult dataset

name	fine	coarse	16-bit	depth	video	sub	AP	AP 50%	AP 100m	AP 50m
Shape-Aware Instance Segmentation	yes	no	no	no	no	2	17.4	36.7	29.3	34.0
DWT	yes	no	no	no	no	2	15.6	30.0	26.2	31.8
InstanceCut	yes	yes	no	no	no	no	13.0	27.9	22.1	26.1
RecAttend	yes	no	no	no	no	4	9.5	18.9	16.8	20.9
Pixel-level Encoding for Instance Segmentation	yes	no	no	yes	no	no	8.9	21.1	15.3	16.7
R-CNN + MCG convex hull	yes	no	no	no	no	2	4.6	12.9	7.7	10.3
Instance-level Segmentation of Vehicles by Deep Contours	yes	no	no	no	no	2	2.3	3.7	3.9	4.9

Plant Phenotyping Datasets

- CVPPP datasets:
  - Arabidopsis
  - Tobacco plants
- Instance-level annotation
  - Every leave is annotated separately
  - Only leaves



Biggest Difficulty: Annotation

- Instance-level annotation is difficult!
  - Also diverse for different plants
- Anecdotal note from Cityscapes insider:
  - Each road scene takes between 1 – 1.5 hours to completely annotate instance-wise
  - Plants are not much easier (individual leaves/roots all similarly colored)
- A sample result taken from the CVPPP dataset paper:

Dataset by start date	Subjects	Period	Total images	Resolution per plant	Annotated plants	Annotated images
23.01.2012	20	18 days	34560	5 megapixel	4	26
16.02.2012	20	20 days	38400	5 megapixel	2	13
15.05.2012	20	18 days	34560	5 megapixel	2	8
10.08.2012	20	30 days	57600	5 megapixel	5	36

Agenda

- Project Goals
- Background info
  - Deep Learning
  - Semantic Segmentation
  - Instance-Level Segmentation
- **Project Plans**
- Proposed Timeline

Specific Goals to be Achieved in the Project

- Enable fast, efficient annotation
- Enable collection of large, annotated datasets
  - With big data, deep learning is already very good in recognition problems
- Integrated interface easy to use by biologists
- Plant biology is an ideal platform
  - Incentives to annotate well (pixel-level)
  - Different-yet-related needs for different researchers
  - Relatively long response period (takes days to obtain a new batch of images)

Project Approach

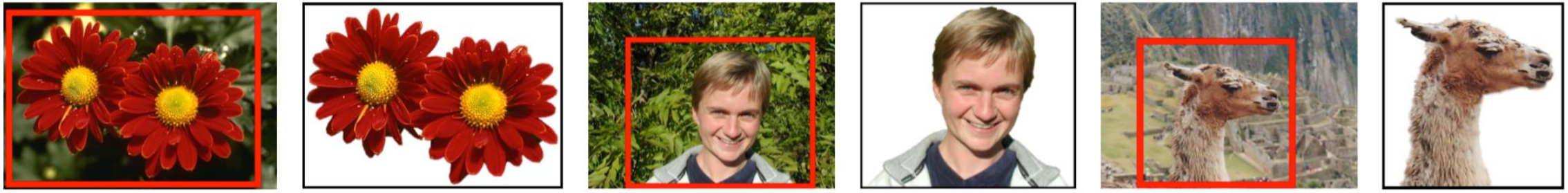
- Deep learning based semantic segmentation interface
- Interactively takes user annotation and trains underlying deep learning system
- On the annotation side, utilize a more traditional segmentation-based annotation scheme for interactive speed
  - Deep learning predictions (unary potentials) + graph-cut
- The annotations are then collected in batch and used to fine-tune the underlying deep learning system
  - Training takes a long time, but likely does not matter in this application

Modules

- Interactive Segmentation Module (Zheng, Jialin)
- User Interface (Zheng)
  - User interactions with the interactive segmentation module
  - Also generates biological traits and features (connect to statisticians)
- Semantic Segmentation Module (Jialin)
  - Receives input from interactive segmentation to fine-tune model

Interactive Image Segmentation

- GrabCut (2004) – Foreground Segmentation



- Drawback: Only uses color information
  - Difficult to segment if there is no color information
- Plan: Utilize CNN features as a basis for graph-cut based interactive segmentation algorithm
  - Quickly incorporate user feedback (no CNN learning)
  - Not limited to color – shape, texture can be captured by CNN features

GrabCut Illustration

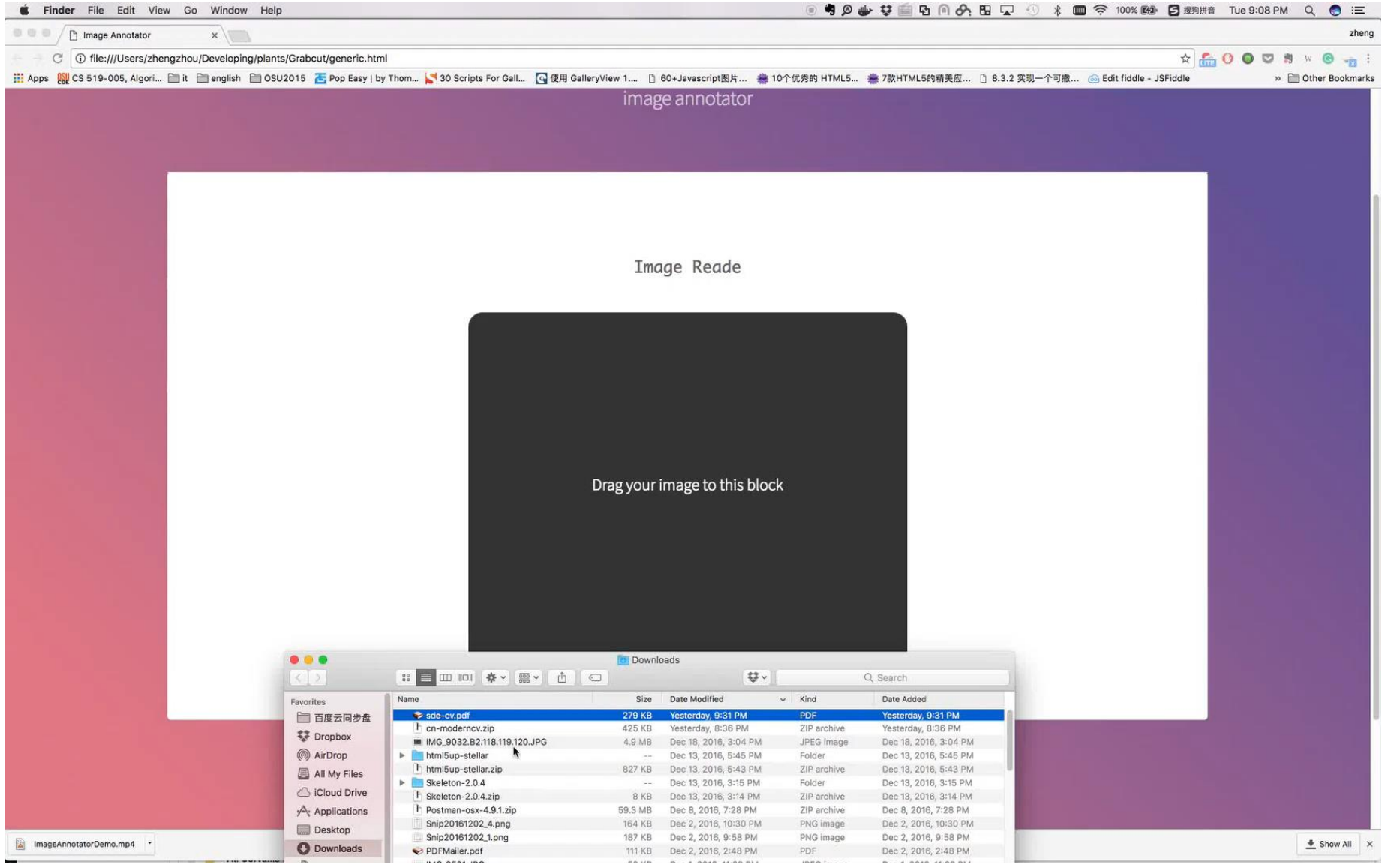
CNN-assisted interactive segmentation

- Two forms of energy functions added to a graphical model inference problem
- Unary potentials (pixel predictions) – FCN will make such kind of predictions
- Shape priors – higher-order potentials of what the shape of the object should look like
  - CNN could inform which shape prior(s) to use
  - Utilize the ideas in the DeepMask paper (Pinheiro et al. 2015)

User Interface

- We propose a web-service based UI
- The classification and segmentation tasks can be done on server side (more computation power)
- User interaction will be browser-based
  - ❑ Users will not need to install several toolboxes and libraries by themselves (difficult for non-CS people!)

User Interface Mock-up



Statistician Connection

- Metadata collection (barcodes, timestamps, etc.)
- UI Support for exporting statistics
- Traits information sent to statisticians
  - Categorization
  - Area of each part
    - e.g. Root, Stem, Leave
  - Count information
    - How many roots
    - How many leaves
  - Shape information
- Potential approach: CNN features directly to statisticians to compare with GWAS results

User Interface Deployment

- The project will probably be deployed on OSU servers
- Users can easily access services from server through web-based user interface
- Potential extension to multiple crop species

Semantic Segmentation

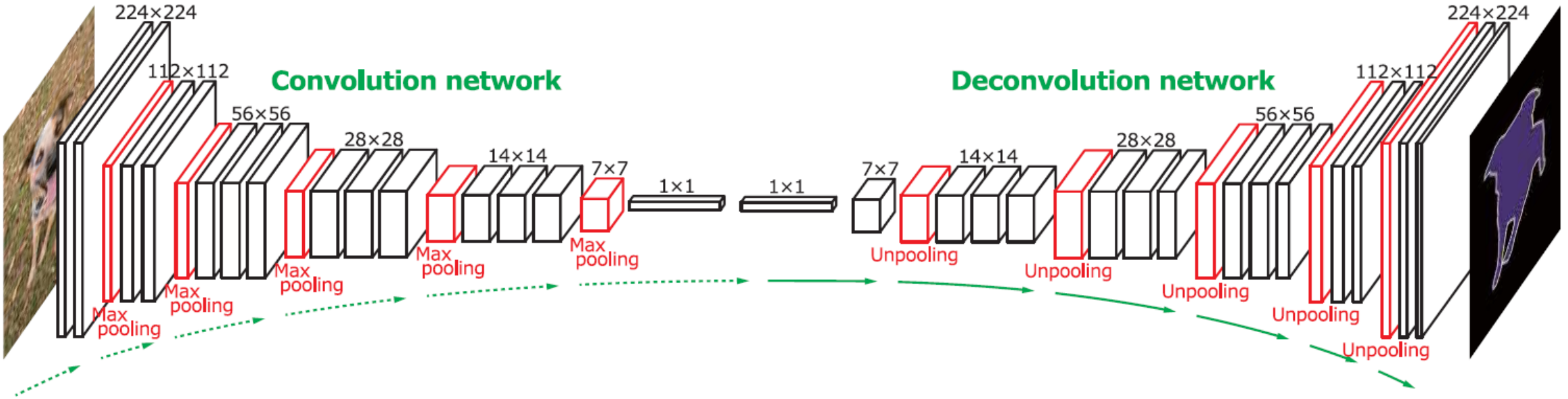
- Initial k-means bootstrap using color segmentation
- Good for basic separation of stems, leaves and roots
- Not good for:
  - Root counting
  - Leave counting
  - Accurate segmentation of roots
  - More subtle situations

Semantic Segmentation

- A Fully-Convolutional Neural Network (FCN) will be the starting deep learning model in this project
- FCN will use the image annotations received from biologists as training examples to tune its parameters.
- On new images, the model sends the testing result to the biologists so that users are able to correctly modify them.
- In the second year, we plan to develop our own instance-level semantic segmentation algorithm and switch to it

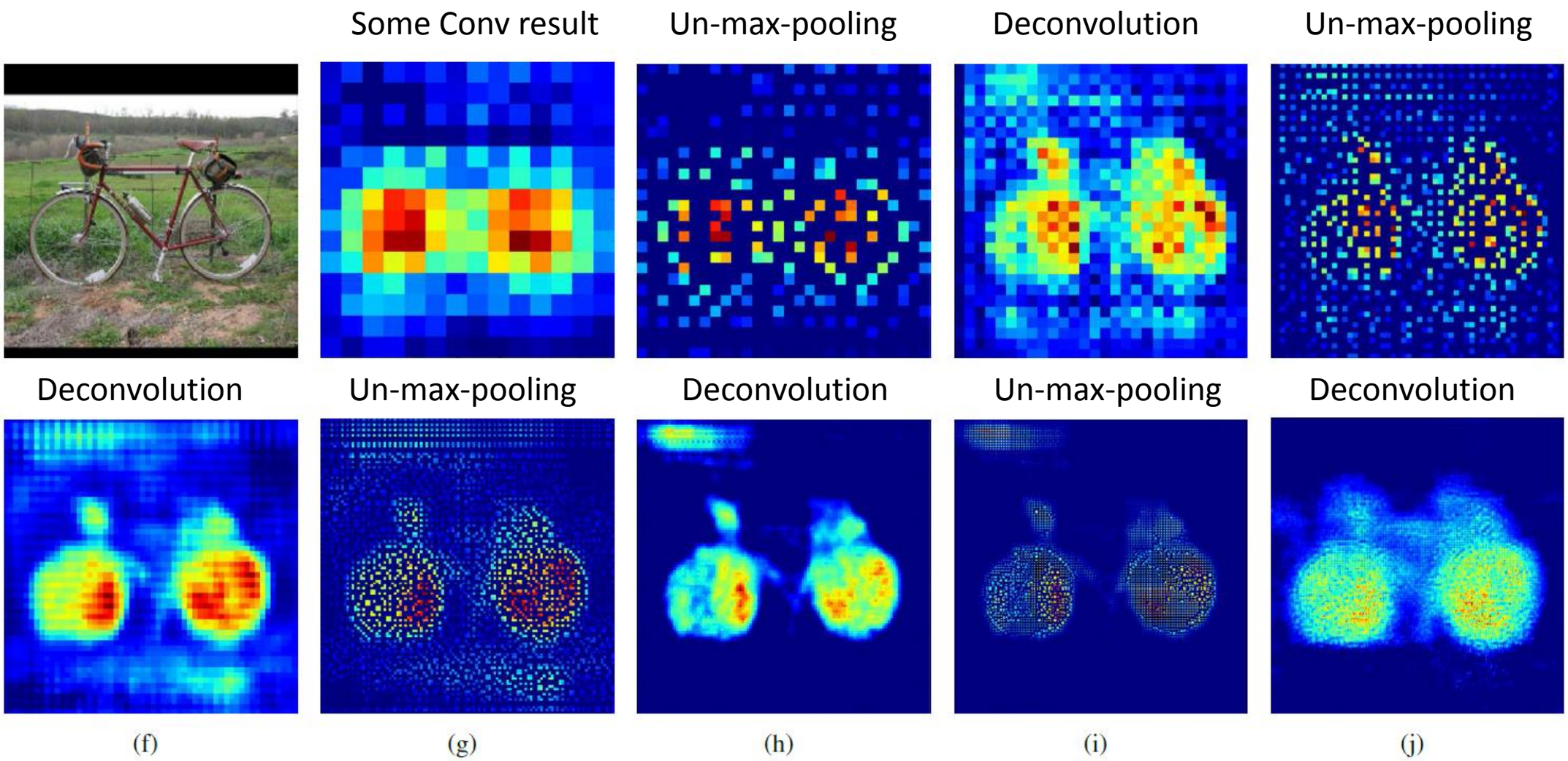
Fully Convolutional Network

- Convolution followed by Deconvolution (reverse operation of convolution)



J. Long, E. Shelhamer, T. Darrell. Fully Convolutional Networks for Semantic Segmentation. CVPR 2015
H. Noh, S. Hong, B. Han. Learning Deconvolution Network for Semantic Segmentation. ICCV 2015

Deconvolution – Making Precise Predictions



Hyperspectral Images

- There has not been a large-scale hyperspectral image dataset that we know of
- However, recent work has shown that CNN also has significant advantages over traditional methods (Yu et al. 2017)
- Hyperspectral images (VNIR, NIR, SWIR, etc.) be thought of as additional channels in the image
- No significant issues to train/test on hyperspectral images with enough training data

Project Plans Conclusion

- Goal: Interactive segmentation tool combined with deep learning
  - Allows fast, accurate annotation capabilities across species and visible traits
  - Allows fast, exact learning and prediction of traits and phenotypes
  - Compare results with GWAS analysis and support other statistical analysis
- Innovative components:
  - Interactive Segmentation
    - Novel approach with deep learning combined with interactive segmentation
  - User Interface
    - Web-based UI allowing biologists to easily access, annotate and organize data
  - Semantic Segmentation
    - Novel approach in instance-based segmentation

Proposed Timeline

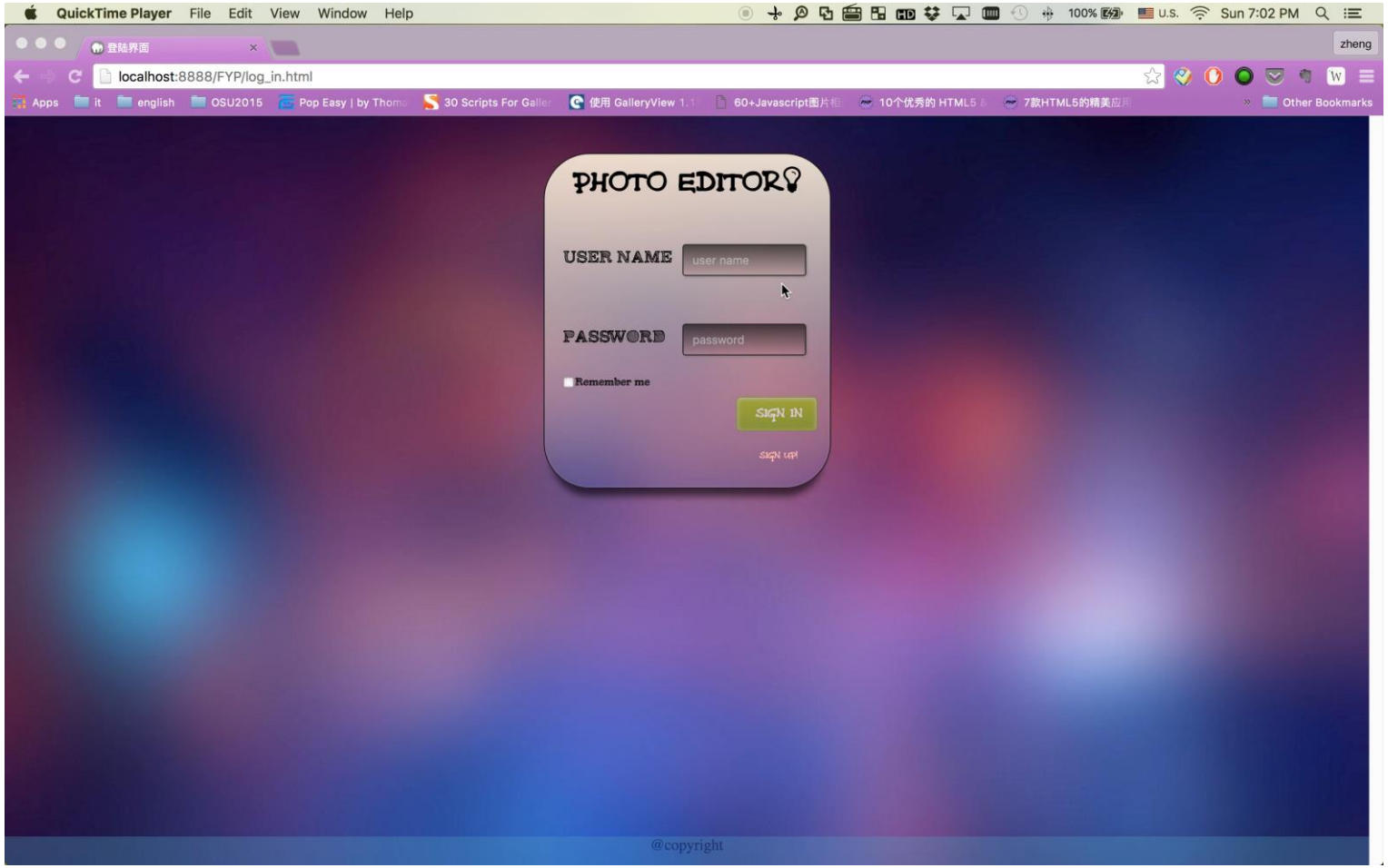
	Year 1	Year 2	Year 3
Create and User-Test Image Acquisition System	XX		
Create basic analysis system, user interface and backend	XXXX		
Refine the accuracy of the system and fuse all sensors		XXXX	X
Develop mid-level image signature methodology, full user interface, system documentation and web deployment		XX	XXXX
Outreach Effort, Publication, Demo of System		X	XXXX

Thank you!

Comments and discussions welcome

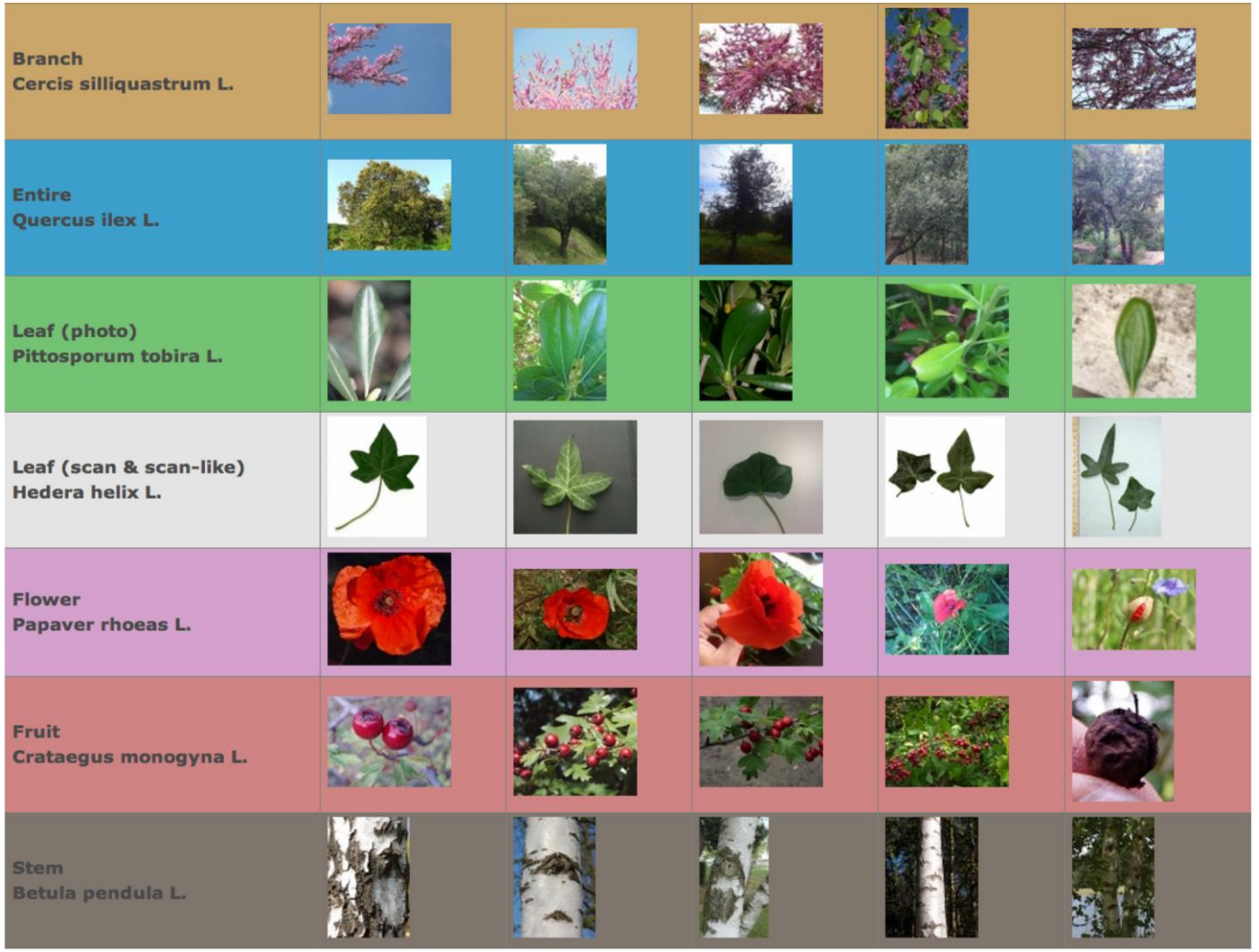
OUR Existing UI design

HTML5 CANVAS DEMO



PlantCLEF 2015

- 113,205 pictures from 1,000 species of trees, herbs and ferns
- Branch, entire, leaf, flower, fruit, stem all available



PlantCLEF 2015 results

	Run name	Key-words	Score
Deep Convolutional Networks	SNUMED INFO run4	5-fold GoogLeNet Borda+	0,667
	SNUMED INFO run3	5-fold GoogLeNet Borda	0,663
	QUT RV run2	GoogLeNet SoftMax	0,633
	QUT RV run3	GoogLeNet Norm & SoftMax	0,624
	SNUMED INFO run2	GoogLeNet Borda+	0,611
	INRIA ZENITH run1	GoogLeNet Max Pool.	0,609
	SNUMED INFO run1	GoogLeNet Borda	0,604
	INRIA ZENITH run3	Fusion GoogLeNet & Fisher Vectors	0,592
	QUT RV run1	GoogLeNet Sum Pool.	0,563
	ECOUAN run1	CNN Sum Pool.	0,487
Non-Deep Approaches	INRIA ZENITH run2	hand-crafted features + Fisher Vectors	0,300
	MICA run2	Hand-crafted feat. + SVM	0,209
	MICA run1	Hand-crafted feat. + SVM	0,203
	MICA run3	Hand-crafted feat. + SVM	0,203
	SABANCI run2	PCANet (not pretrained)	0,162
	SABANCI run1	PCANet (not pretrained)	0,160
	SABANCI run3	PCANet (not pretrained)	0,158
	UAIC run1	CBIR (LIRE)	0,013

Machine Learning Backend: Convolution

Sobel filter

-1	0	+1
-2	0	+2
-1	0	+1

Gx

+1	+2	+1
0	0	0
-1	-2	-1

Gy

Convolution

H11	H12	H13
H21	H22	H23
H31	H32	H33

*

F11	F12	F13	F14	F15	F16
F21	F22	F23	F24	F25	F26
F31	F32	F33	F34	F35	F36
F41	F42	F43	F44	F45	F46
F51	F52	F53	F54	F55	F56
F61	F62	F63	F64	F65	F66

=

G11	G12	G13	G14	G15	G16
G21	G22	G23	G24	G25	G26
G31	G32	G33	G34	G35	G36
G41	G42	G43	G44	G45	G46
G51	G52	G53	G54	G55	G56
G61	G62	G63	G64	G65	G66

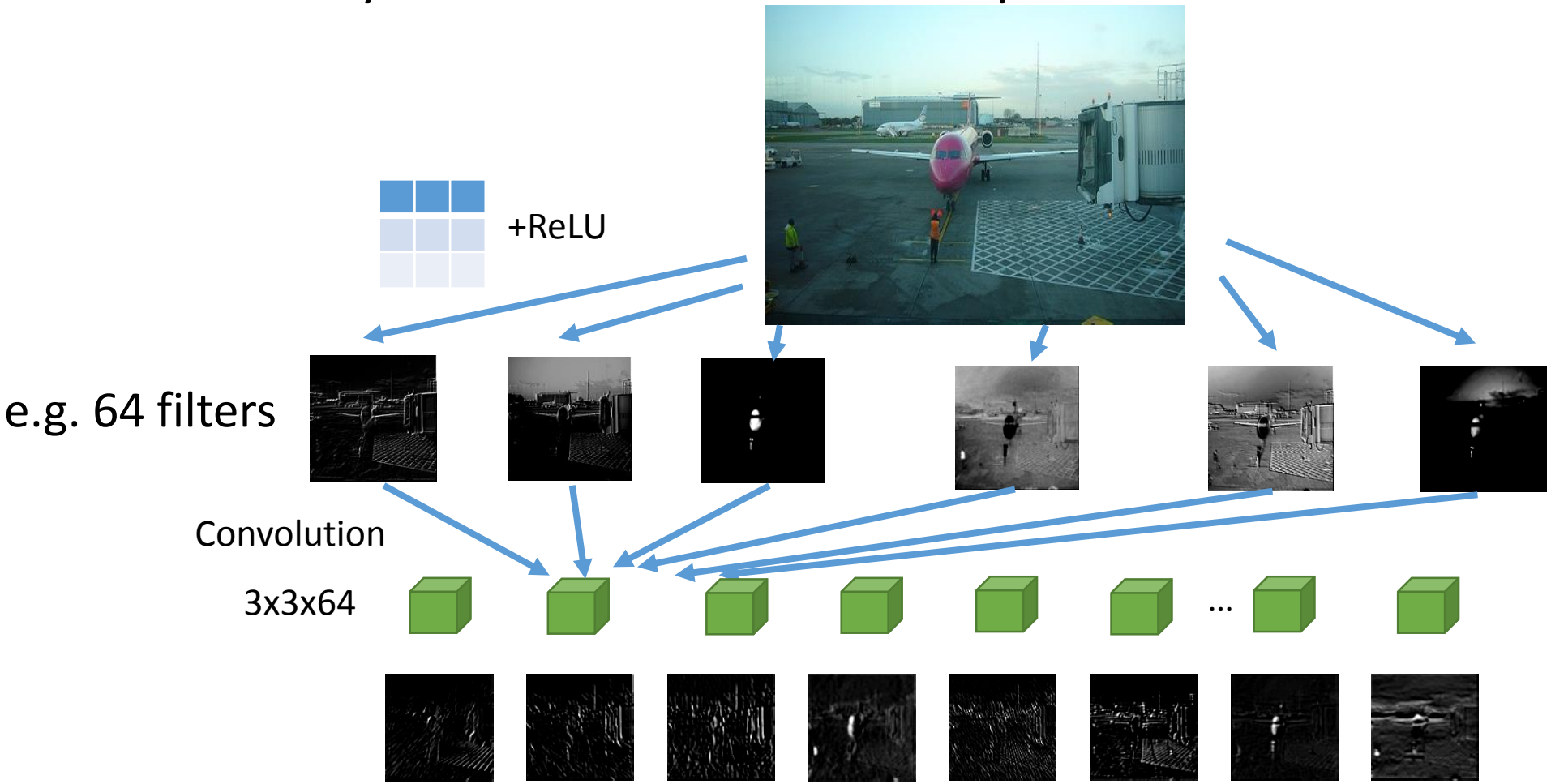
I

Convolution

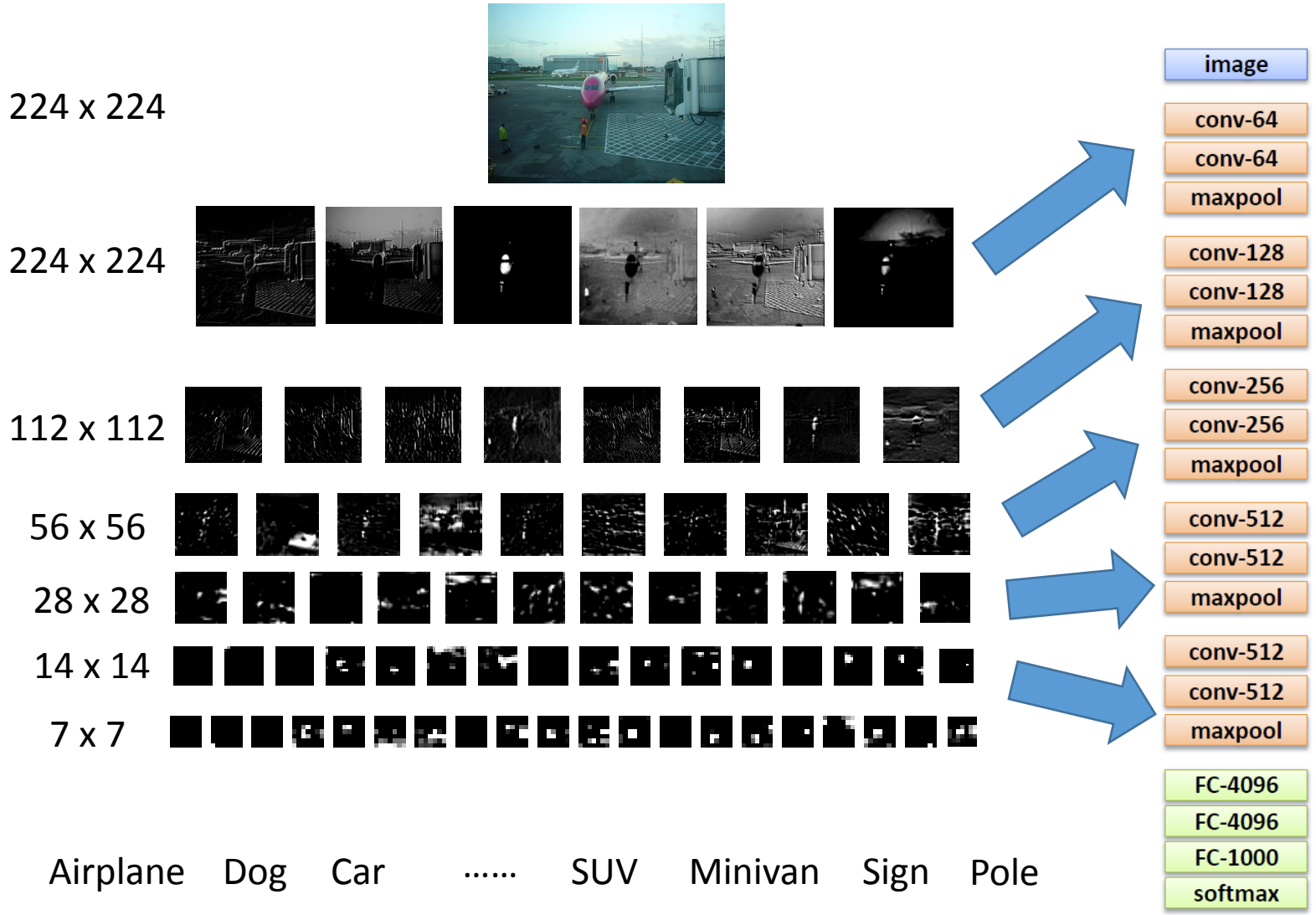
$I * Gx$

Convolutional Networks

- Learn many convolutional filters to operate on the data



Deep Convolutional Networks



(Simonyan and Zisserman 2014)

Dissecting Segments

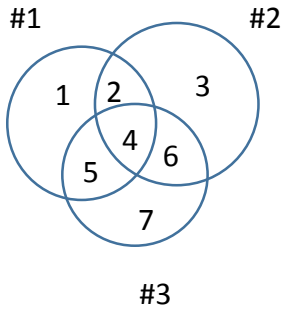
Image

Seg #1:
Chair 0.53
Person 0.29

Seg #2:
Chair 0.23
Person 0.36

Atomic regions:

Venn diagram



Seg #3:
Chair 0.34
Person 0.54

Seg #4:
Chair 0.19
Person 0.43

Generating the Overlap Statistic

- Parametrize on atomic regions:

S_1 

θ_{11}: % of chair
θ_{12}: % of person

S_2 

θ_{21}: % of chair
θ_{22}: % of person

S_3 

θ_{31}: % of chair
θ_{32}: % of person

S_4 

θ_{41}: % of chair
θ_{42}: % of person

$\theta_{ij}|S_i|$ = num. of category c_j ground truth pixels in S_i

A_1: 

$$\mathrm{V}_{chair}(A_1;\theta) = \frac{|A_1 \cap Chair|}{|A_1 \cup Chair|} = \frac{All\ Chair\ pixels\ in\ A_1}{|A_1| + All\ Chair\ outside\ A_1} = \frac{\sum_{S_i \in A_1} \theta_{ij}|S_i|}{\sum_{S_i \in A_1} |S_i| + \sum_{S_i \notin A_1} \theta_{ij}|S_i|}$$

Find Most Consistent Configuration

Seg #1: Chair 0.53
Person 0.29

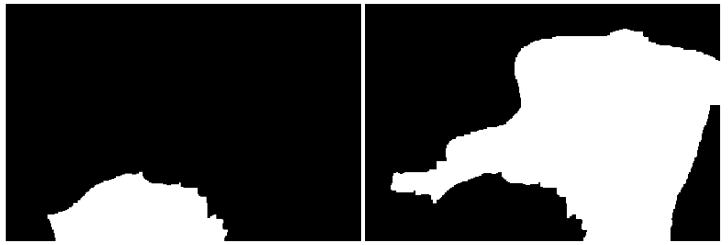
Seg #2: Chair 0.23
Person 0.36

Seg #3: Chair 0.34
Person 0.54

Seg #4: Chair 0.19
Person 0.43

Configuration #1

Person Chair



Configuration #2

Person Chair

Composite Statistical Inference

- Match computed overlap to predicted overlap

Jointly over all categories + all segments!

$$\min_{\theta} \sum_{j=1}^{c} \sum_{i=1}^{m} \beta_{ij} \left(V_j(A_i;\theta) - \widehat{V}_j(A_i) \right)^2$$

"Generated" statistic

Predicted from the regressor

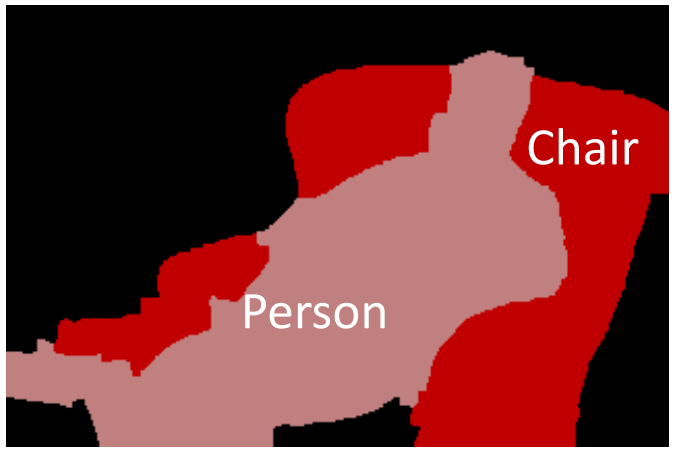
- A_i: segments
- $\widehat{V}_j$: Predicted overlap with category

Li, Carreira, Lebanon and Sminchisescu. CVPR 13

Joint optimization

- θ map after joint optimization on all objects:

Person



Chair

Multiple Objects from Same Category

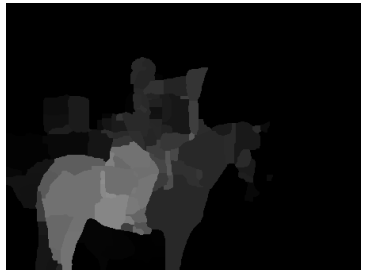
- Assume geometric prior on number of objects
- Enumerate 1 object, 2 objects, etc.
  - Compare posterior value

$n_{horse} = 1$

Posterior: -40.133

$n_{horse} = 2$ Posterior: -35.889

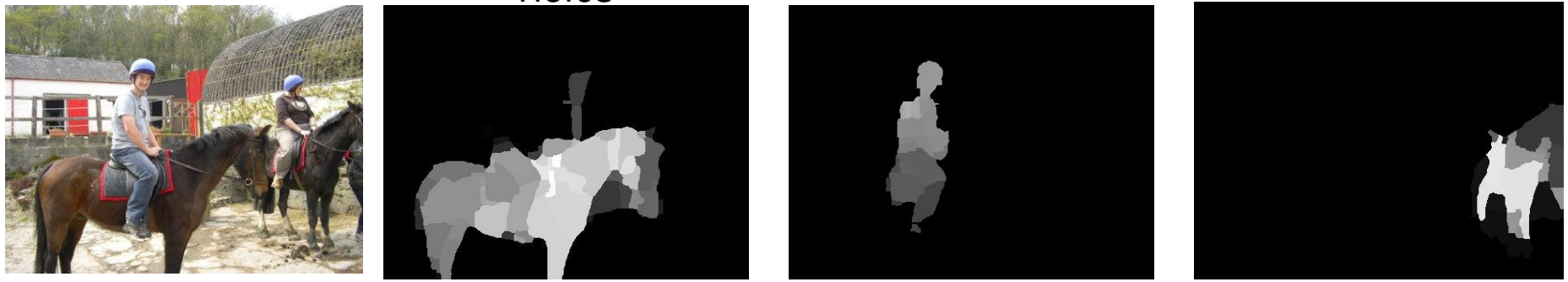
$n_{horse} = 3$ Posterior: -47.600



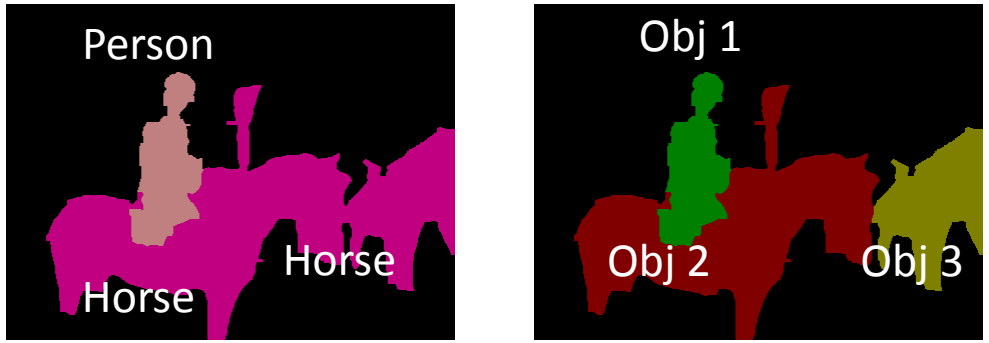
Remember:

Joint Optimization

- θ map after joint optimization:



- Final segmentation:



Overlap+CSI is almost enough

- **150 Noisy** segments, **no noise** in overlap
  - Report PASCAL VOC score (pixel-wise AP)

Best Proposal	CSI
84.2%	**90.7%**

Results: with predicted overlap

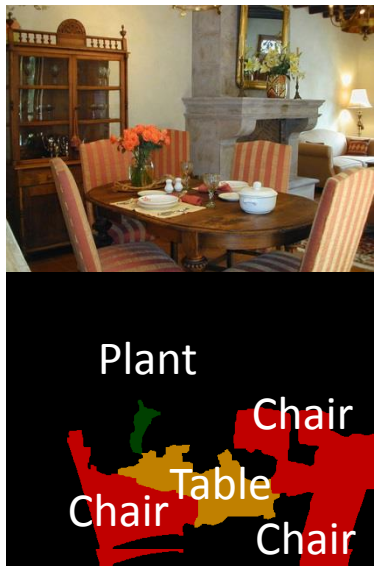
Segmentation Results: VOC2012 BETA

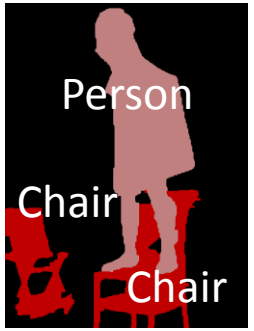
Competition "comp5" (train on VOC2012 data)

Average Precision (AP %)

	mean	aero plane	bicycle	bird	boat	bottle	bus	car	cat	chair	cow	dining table	dog	horse	motor bike	person	potted plant	sheep	sofa	train	tv/ monitor
O2P_SVRSEGM_CPMC_CSI [?]	**47.5**	64.0	**32.2**	45.9	34.7	46.3	59.5	**61.7**	49.4	14.8	47.9	31.2	42.5	51.3	**58.8**	**54.6**	34.9	54.6	34.7	50.6	42.2
NUS_DET_SPR_GC_SP [?]	47.3	52.9	31.0	39.8	**44.5**	**58.9**	**60.8**	52.5	49.0	**22.6**	38.1	27.5	**47.4**	52.4	46.8	51.9	**35.7**	**55.3**	**40.8**	**54.2**	**47.8**
BONN_O2PCPMC_FGT_SEGM [?]	47.0	**65.4**	29.3	**51.3**	33.4	44.2	59.8	60.3	52.5	13.6	**53.6**	**32.6**	40.3	**57.6**	57.3	49.0	33.5	53.5	29.2	47.6	37.6
BONNGC_O2P_CPMC_CSI [?]	45.4	59.3	27.9	43.9	39.8	41.4	52.2	61.5	**56.4**	13.6	44.5	26.1	42.8	51.7	57.9	51.3	29.8	45.7	28.8	49.9	43.3
BONN_CMBR_O2P_CPMC_LIN [?]	44.8	60.0	27.3	46.4	40.0	41.7	57.6	59.0	50.4	10.0	41.6	22.3	43.0	51.7	56.8	50.1	33.7	43.7	29.5	47.5	44.7
OptNBNN-CRF [?]	11.3	10.5	2.3	3.0	3.0	1.0	30.2	14.9	15.0	0.2	6.1	2.3	5.1	12.1	15.3	23.4	0.5	8.9	3.5	10.7	5.3

- CSI does well on high-interaction objects





Home

Yuan Jiang presents…

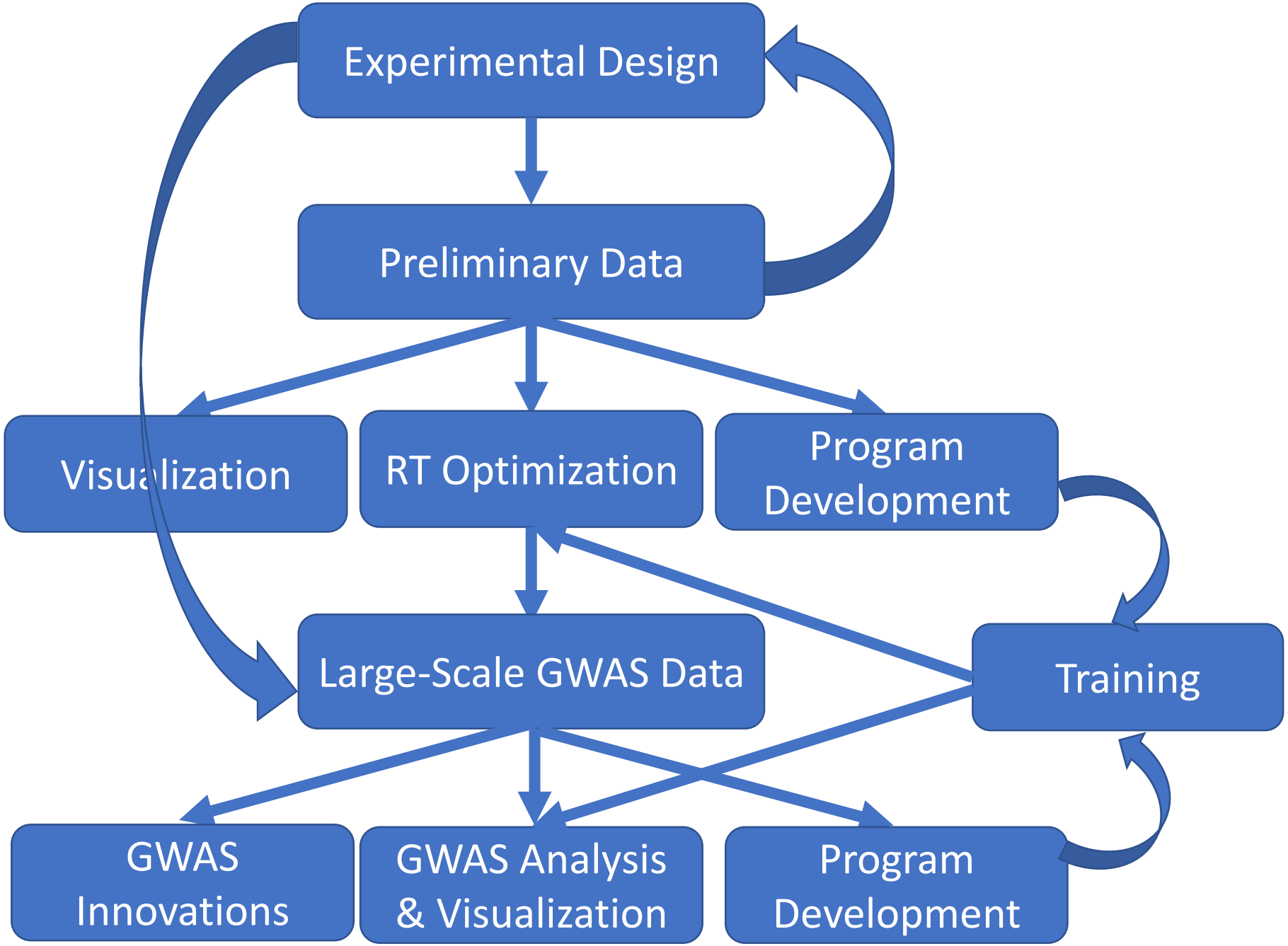
"Data Visualization and Statistical Analysis"

Data Visualization and Statistical Analysis

Yuan Jiang
Department of Statistics
Oregon State University, Corvallis, OR

Overview

- My Role
- Data Visualization
  - Design
  - Traits
  - Models
  - GWAS
- Statistical Analysis
  - Design
  - Standard GWAS Analysis
  - Innovations for GWAS Analysis


Experimental Design
Preliminary Data
Visualization
RT Optimization
Program Development
Large-Scale GWAS Data
Training
GWAS Innovations
GWAS Analysis & Visualization
Program Development

Preliminary Data

- Genotype: 537 undomesticated genotypes with low levels of linkage disequilibrium
- Treatment: soil or water
- Replicate: 2 blocks for each treatment
- Traits:
  - Height: tallest shoot of new growth in cm
  - Diameter: 2 inches above soil level in soil/2 inches from bottom of stick in water of the woody part
  - Root scoring: no root, 1-2 roots, 3-5 roots, >5 roots

Data Visualization: Summary

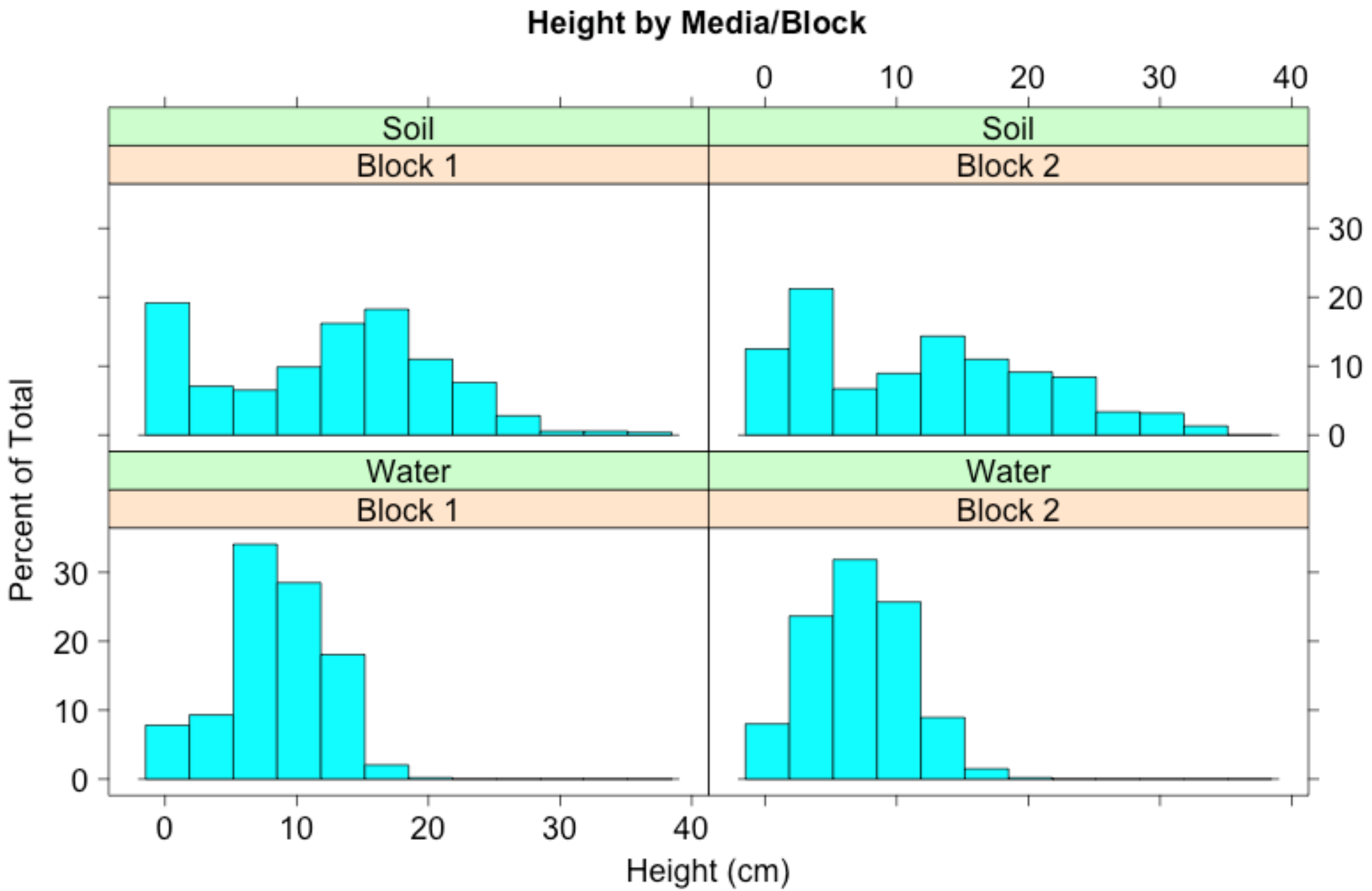
- Traits
  - Distributions
  - Associations
- Models
  - Explorative plots
  - Diagnostic plots
- GWAS
  - Q-Q plot for p-values
  - Manhattan plot

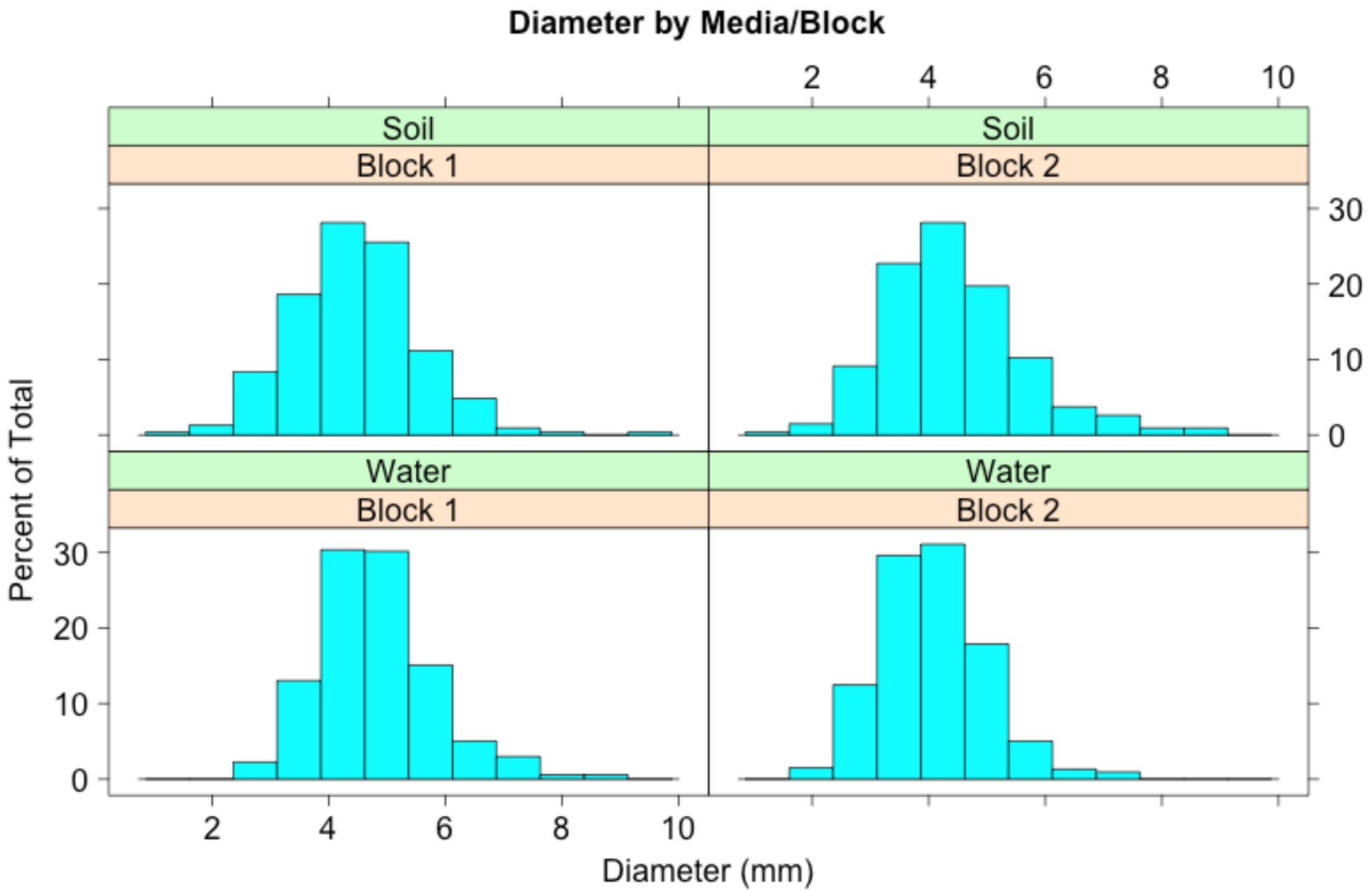
Data Visualization: Traits

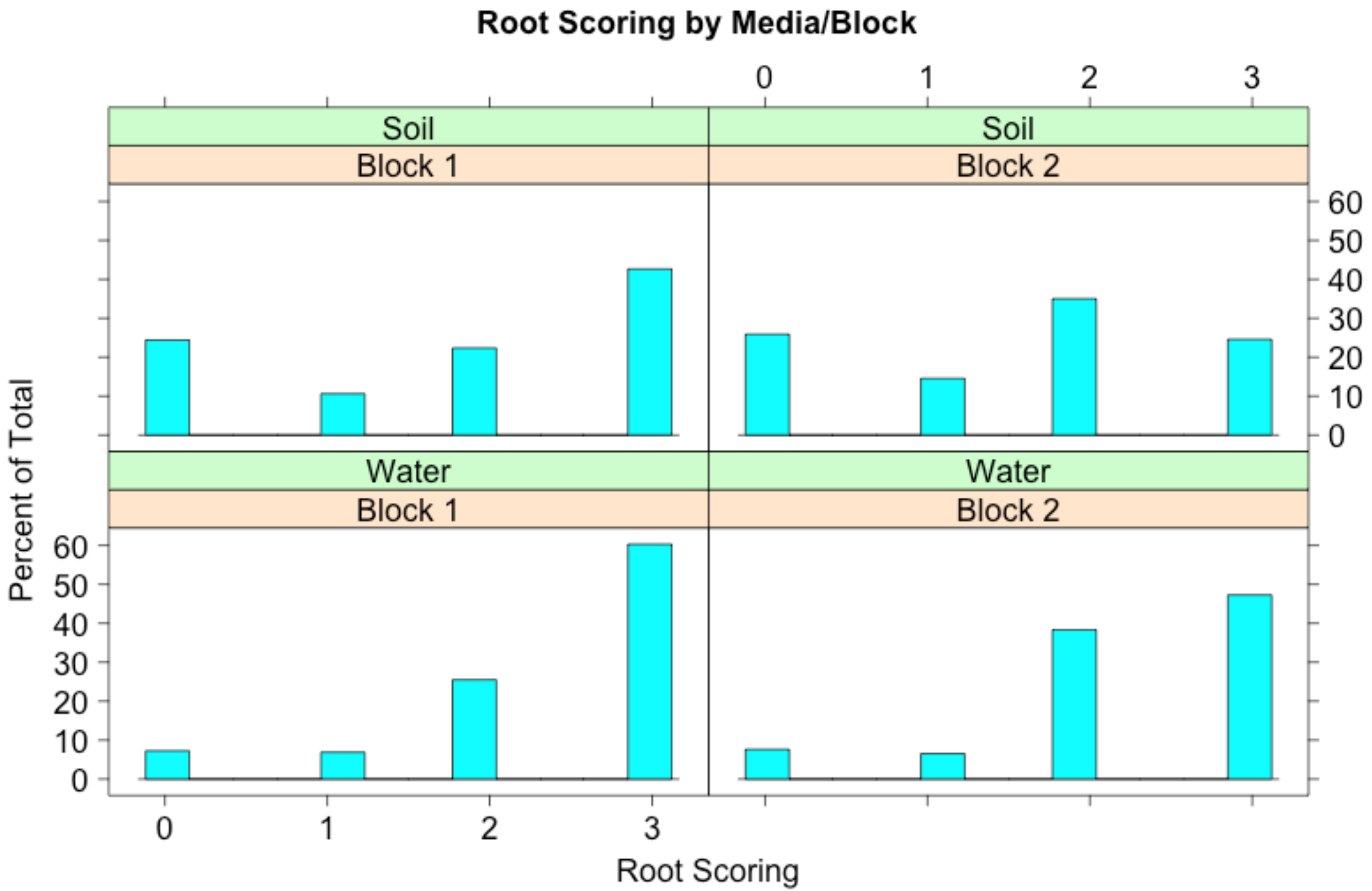
- Distributions
  - Normality
  - Skewness
  - Outlier
  - Transformation
- Associations
  - Scatter plot
  - Scatter plot matrices
  - Linear or nonlinear
- Other Patterns
  - PCA
  - Cluster analysis

Data Visualization: Traits

- Distributions
  - Normality
  - Skewness
  - Outlier
  - Transformation
- Associations
  - Scatter plot
  - Scatter plot matrices
  - Linear or nonlinear
- Other Patterns
  - PCA
  - Cluster analysis

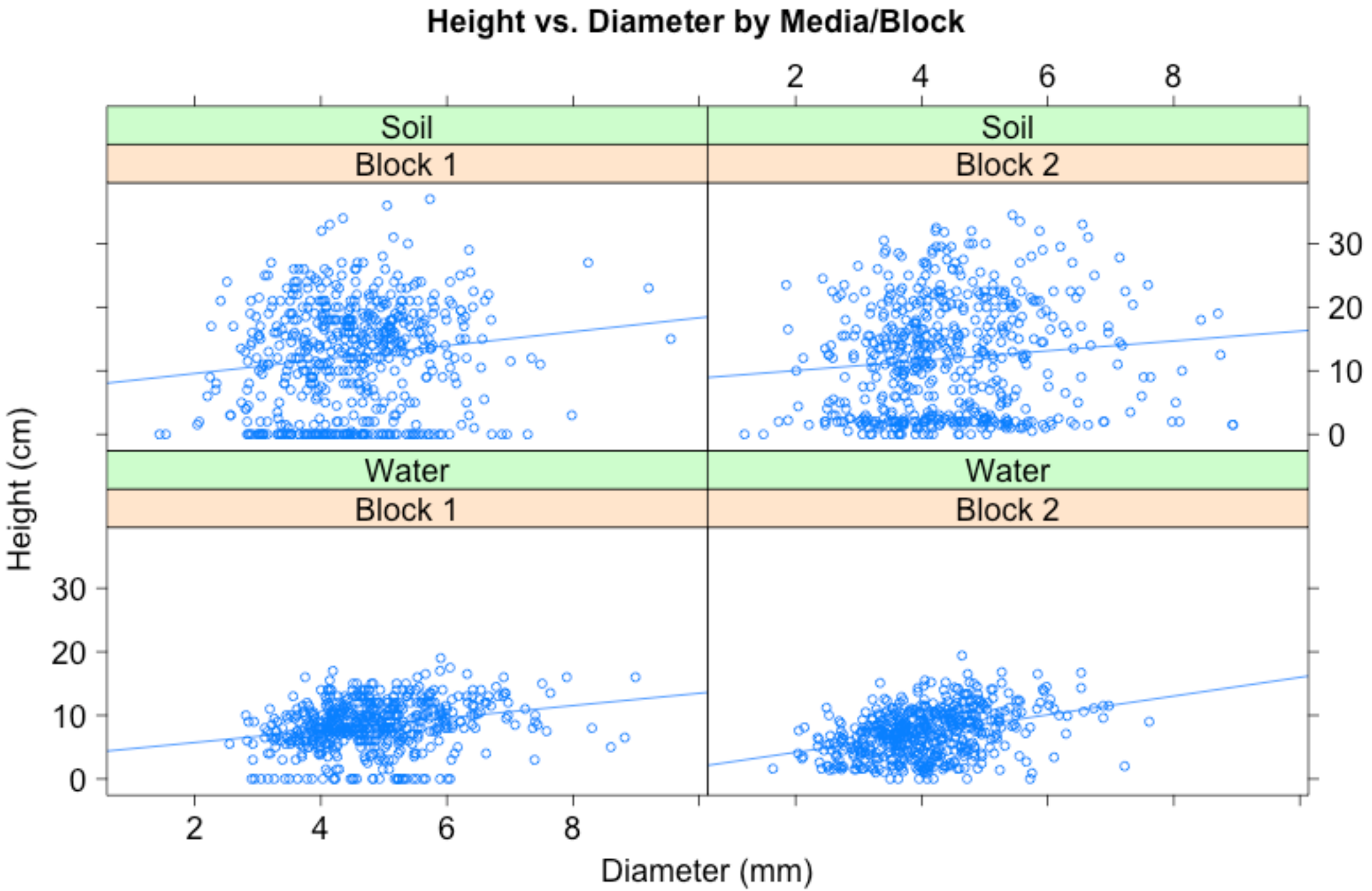

Height by Media/Block
Soil
Block 1
Soil
Block 2
Water
Block 1
Water
Block 2
Percent of Total
Height (cm)
0
10
20
30
40

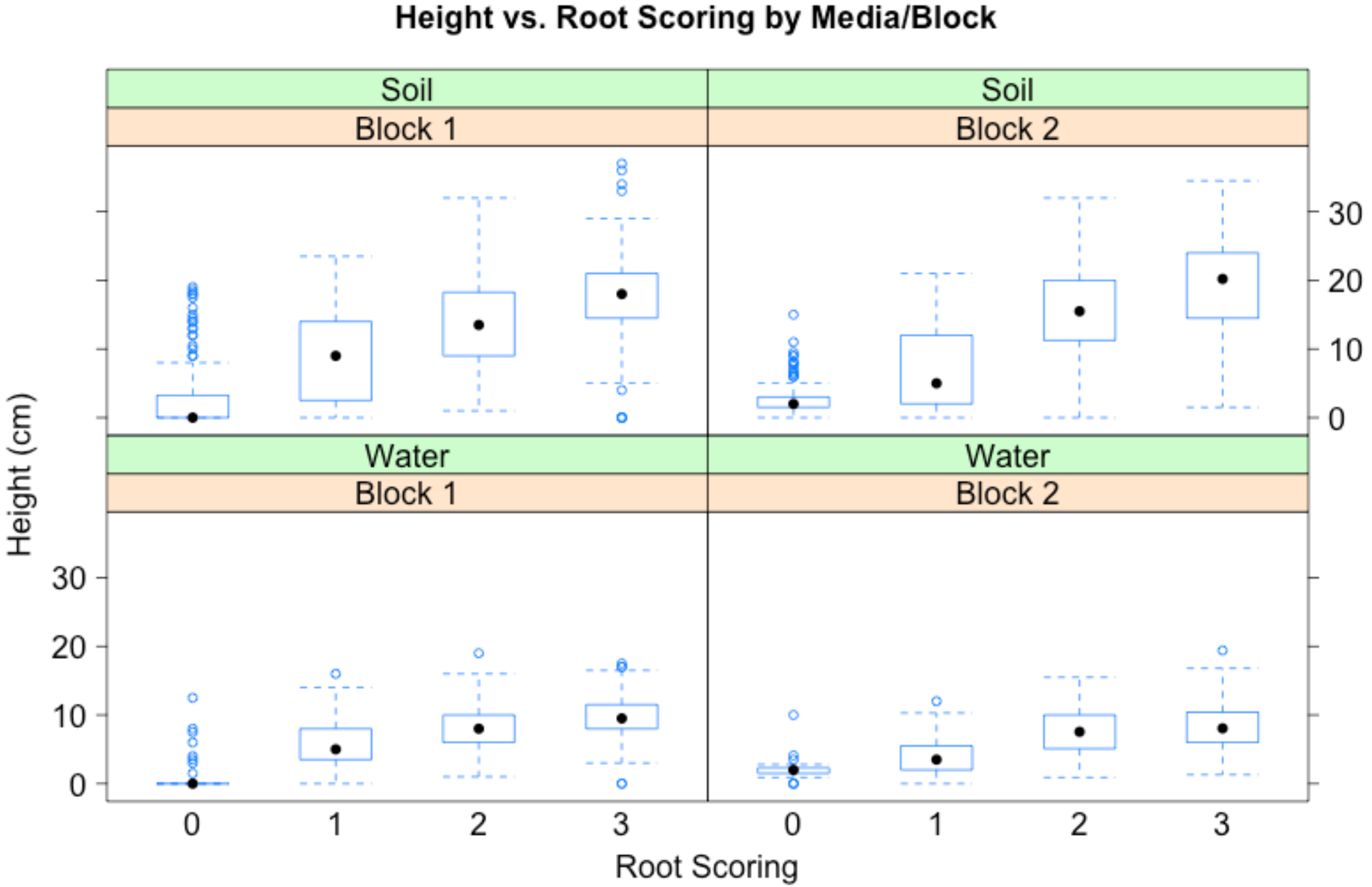

Diameter by Media/Block
Soil
Block 1
Soil
Block 2
Water
Block 1
Water
Block 2
Percent of Total
Diameter (mm)
0
10
20
30
2
4
6
8
10

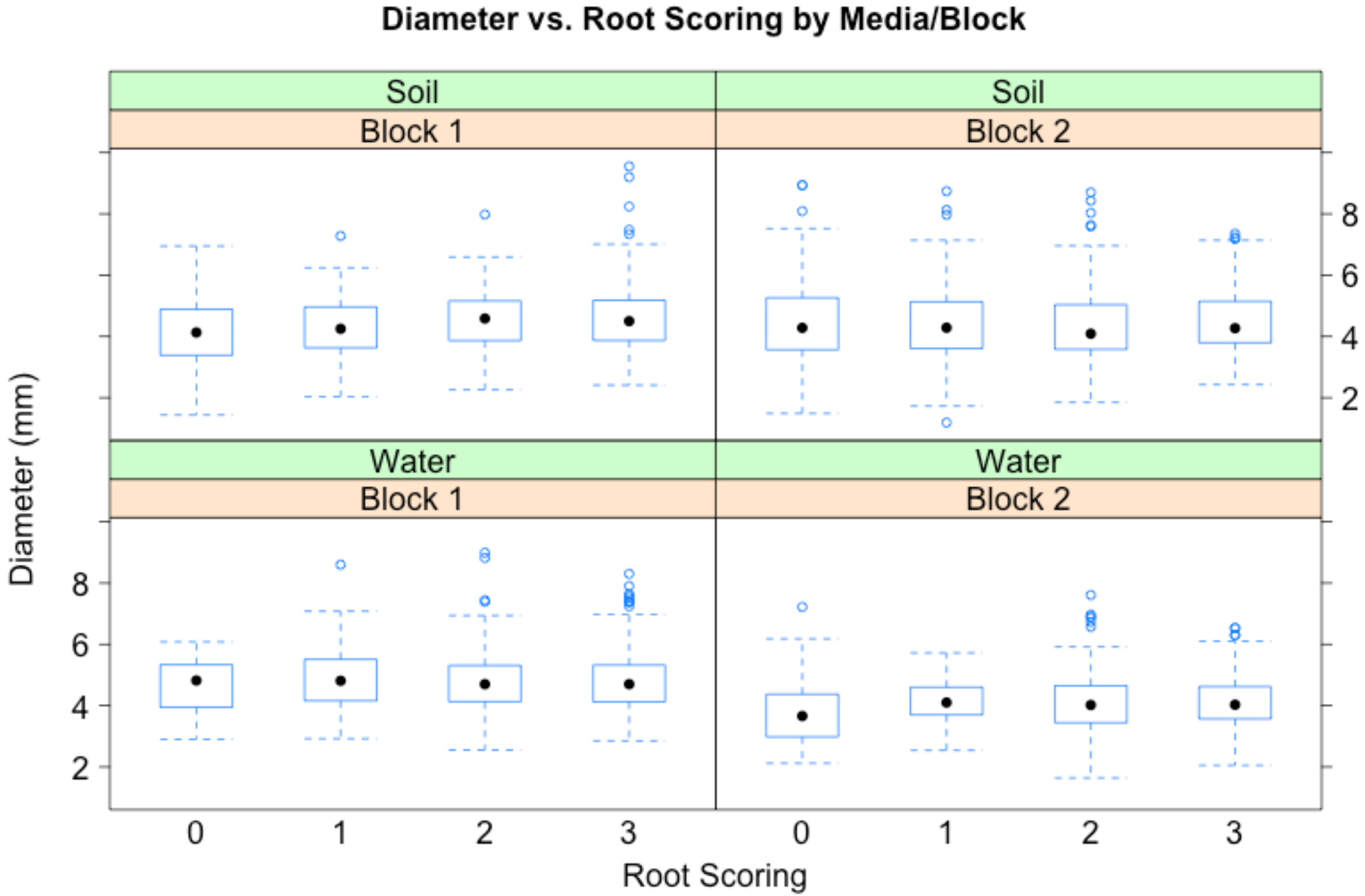

Root Scoring by Media/Block
Soil
Block 1
Soil
Block 2
Water
Block 1
Water
Block 2
Percent of Total
0
10
20
30
40
50
60
0
1
2
3
Root Scoring

Data Visualization: Traits

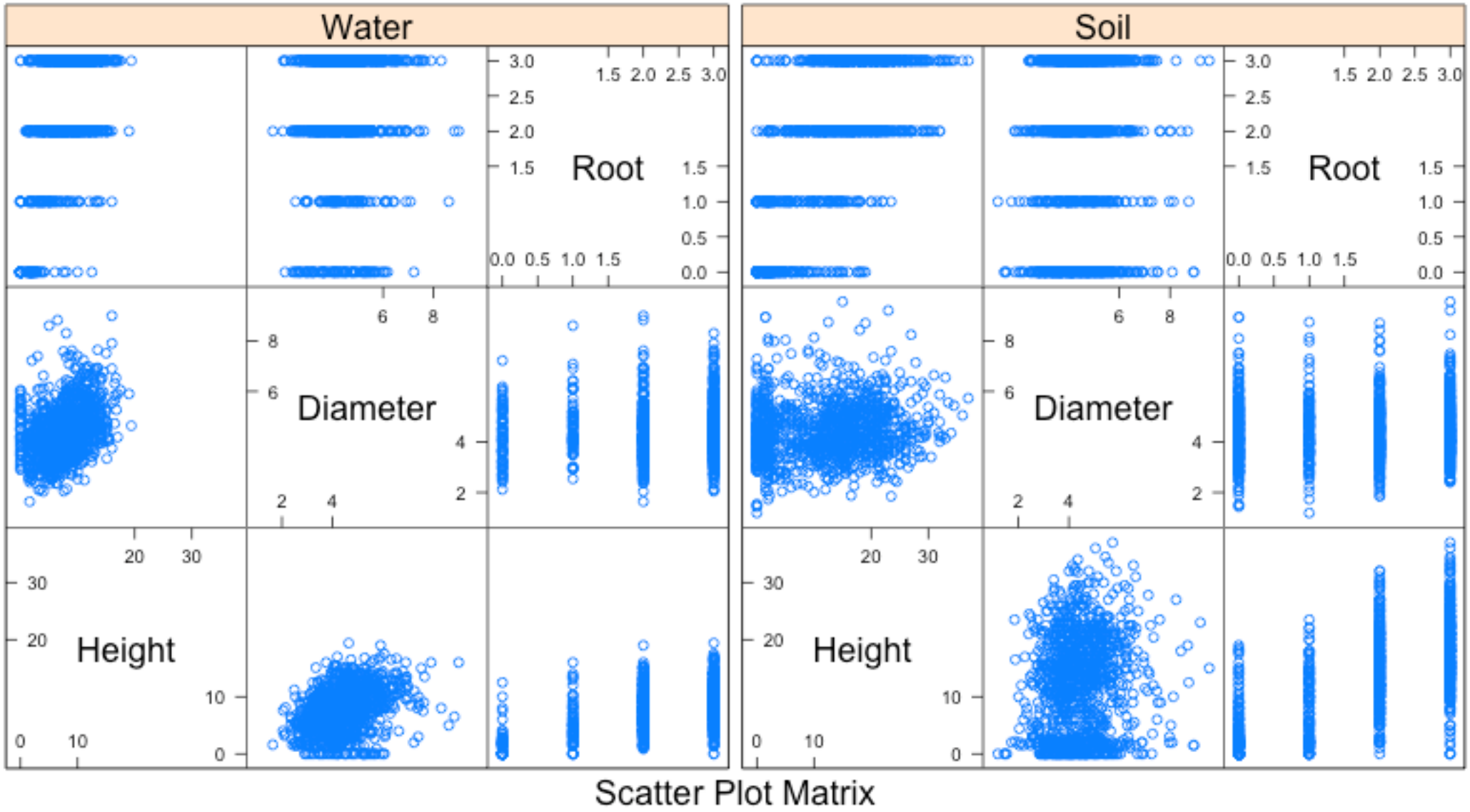
- Distributions
  - Normality
  - Skewness
  - Outlier
  - Transformation
- Associations
  - Scatter plot
  - Scatter plot matrices
  - Linear or nonlinear
- Other Patterns
  - PCA
  - Cluster analysis


Height vs. Diameter by Media/Block
Soil
Block 1
Soil
Block 2
Water
Block 1
Water
Block 2
Height (cm)
Diameter (mm)
0
10
20
30
2
4
6
8


Height vs. Root Scoring by Media/Block
Soil
Block 1
Soil
Block 2
Water
Block 1
Water
Block 2
Height (cm)
0
10
20
30
0
1
2
3
Root Scoring


Diameter vs. Root Scoring by Media/Block
Soil
Block 1
Soil
Block 2
Water
Block 1
Water
Block 2
Diameter (mm)
2
4
6
8
0
1
2
3
Root Scoring

Scatter Plot Matrices by Media



Scatter Plot Matrix

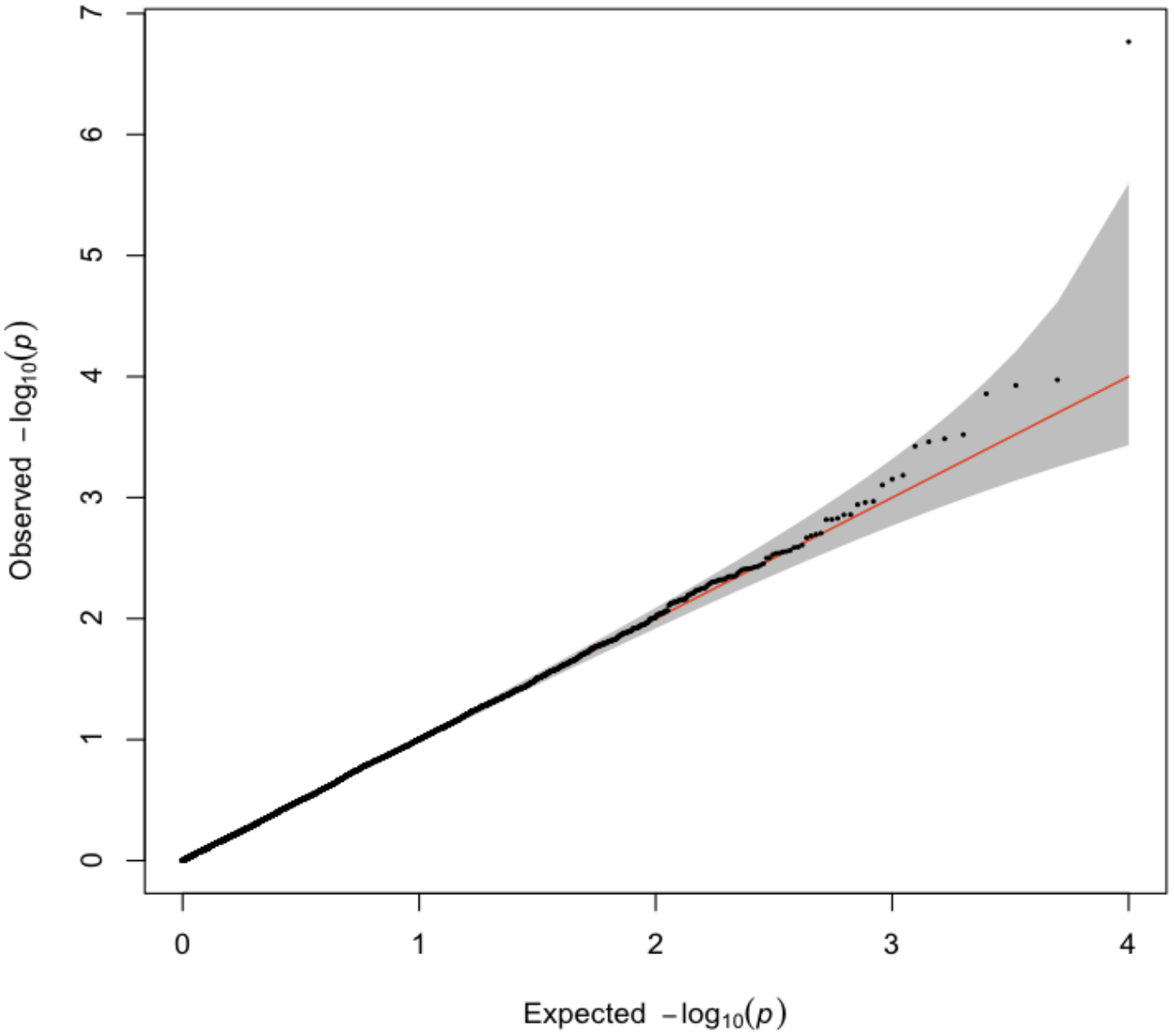
Data Visualization: Models

- Pre-modeling plots: Exploration (examples shown earlier)
  - Distribution plots for transformations
  - Scatter plots for linear or nonlinear relationship
  - PCA
  - Cluster analysis
- Post-modeling plots: Diagnostics (examples shown later)
  - Residual plots
  - Q-Q plots for residuals
  - Q-Q plots for random effects in mixed effects models
  - More…

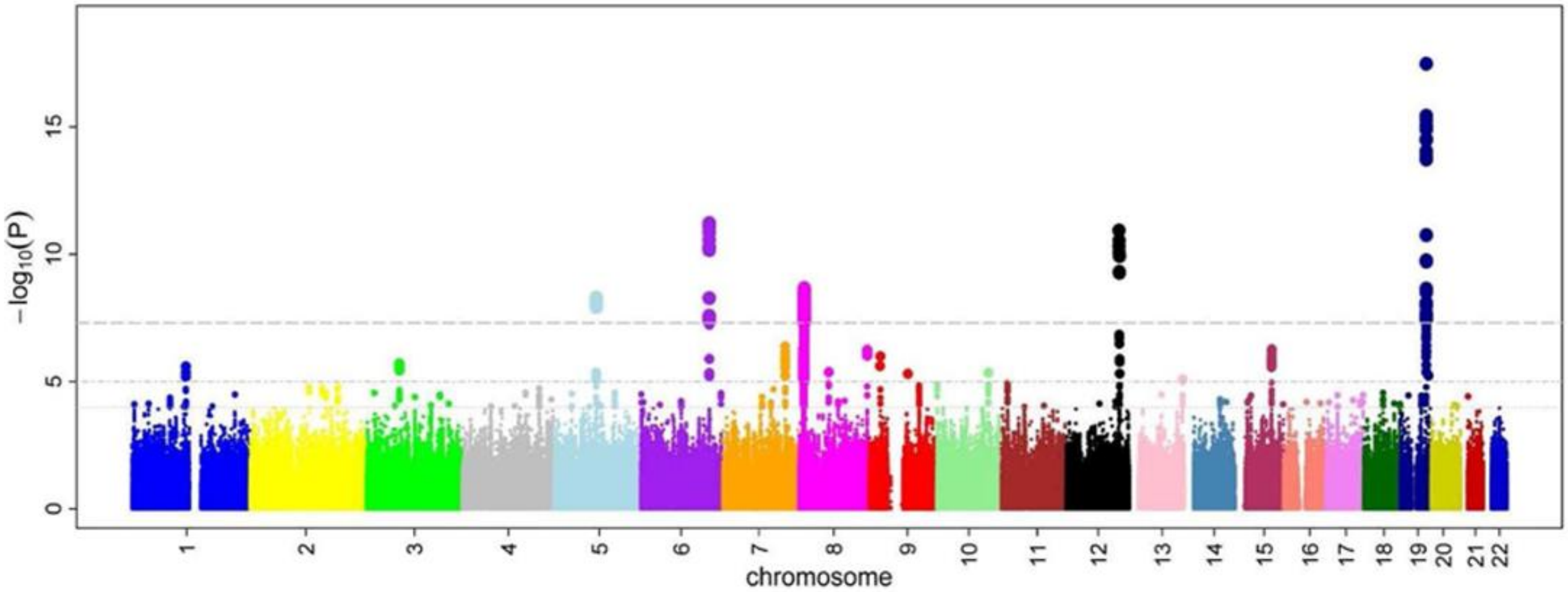
Data Visualization: GWAS

- Q-Q plot for p-values
- Manhattan plots
- See examples from other studies in the following slides

Q-Q Plot for GWAS p-values



Manhattan Plot



Statistical Analysis

- Design of Experiments
  - RT Optimization Studies
- Standard GWAS Analysis
  - Also refer to Wellington's presentation
- Innovations in GWAS Analysis
  - SNP-set Association Analysis
  - Synthetic traits
  - Multiple traits
  - Nonparametric or semiparametric methods

RT Optimization

- Objective
  - To identify a series of treatments that maximize trait heritability
- Genotype population
  - 1,000+ undomesticated genotypes
  - low levels of linkage disequilibrium
- Environmental factors (treatments)
  - Growing media
  - Hormones
  - Light quality
  - Antioxidants
  - Stress
- Phenotypic variance = genotypic variance + environmental variance
- Heritability = genotypic variance/phenotypic variance

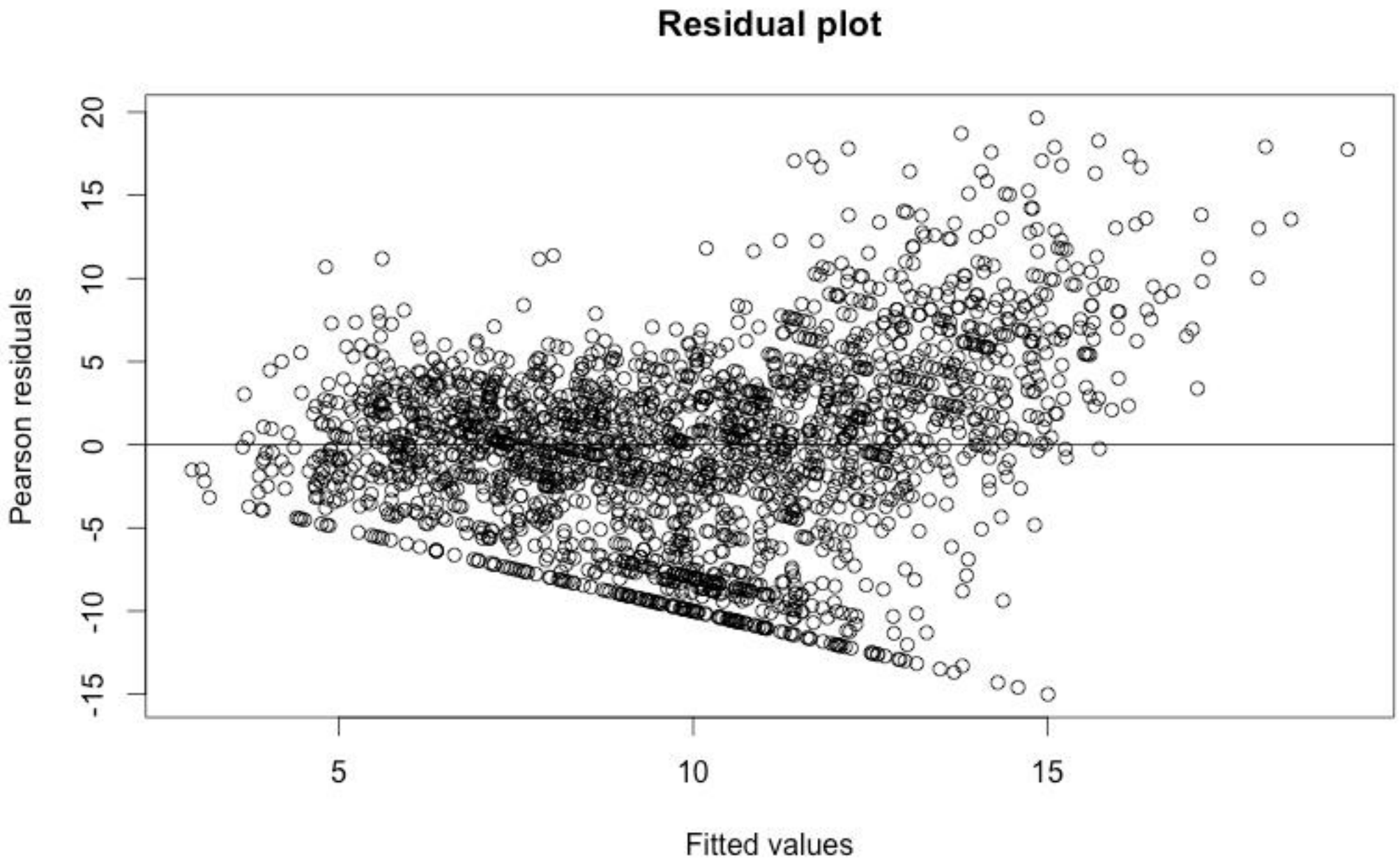
RT Optimization

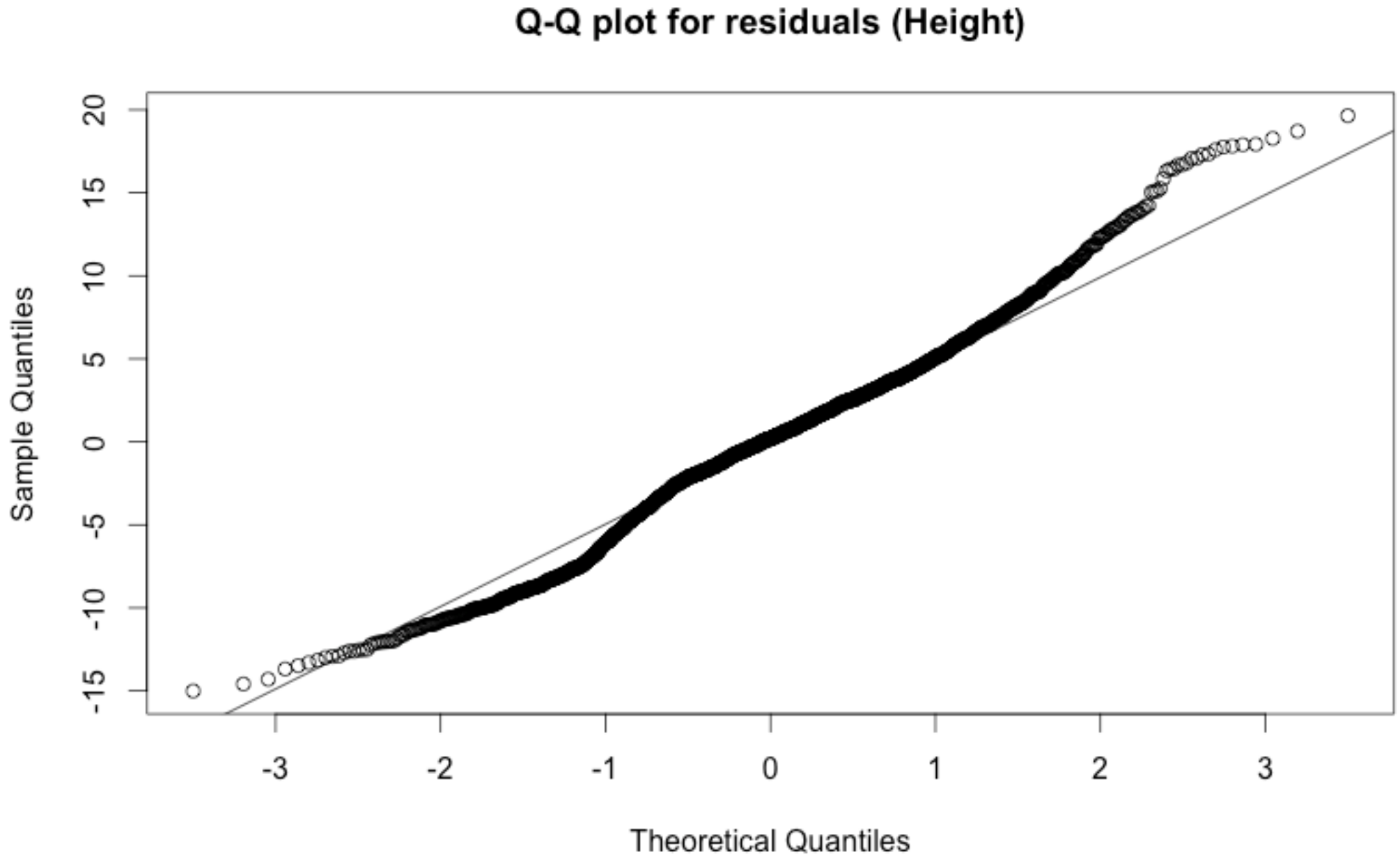
- Linear mixed effects models (LMM) for quantitative traits
  - Height (cm), diameter (mm)
  - Normality, transformation
  - Also known as variance components models
- Generalized linear mixed effects models (GLMM) for qualitative traits
  - Binary (rooting vs. no roots)
  - Ordinal (root scoring: no root, 1-2 roots, 3-5 roots, >5 root)
  - Count (shoot numbers: 0, 1, 2, 3, …)
- Other covariates
  - Plant size not well controlled (larger plants may root faster)
  - Other experimental conditions that may have been measured

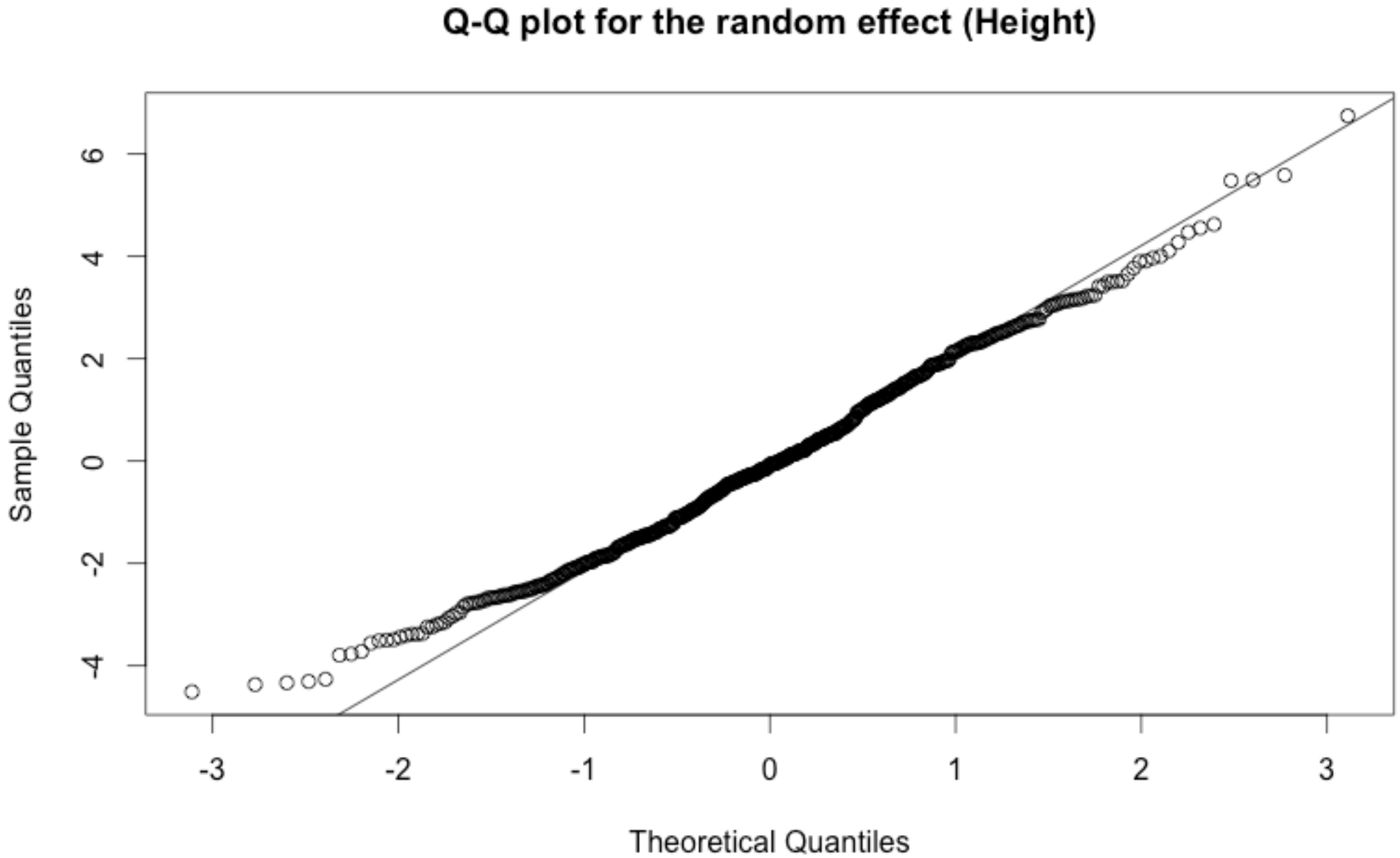
Heritability (Height): 18.1%

LMM: Height ~ Media + Block + (1 | Genotype)

Fixed Effects	Estimate	Std. Error	P-value
(Intercept)	8.2338	0.2555	0.000
Media (Soil)	4.2599	0.2593	0.000
Block (2)	-0.8015	0.2593	0.002
Random Effects	**Name**	**Variance**	**Std. Dev.**
Genotype	(intercept)	7.971	2.823
Residual		36.096	6.008


Residual plot
Pearson residuals
20
15
10
5
0
-5
-10
-15
5
10
15
Fitted values

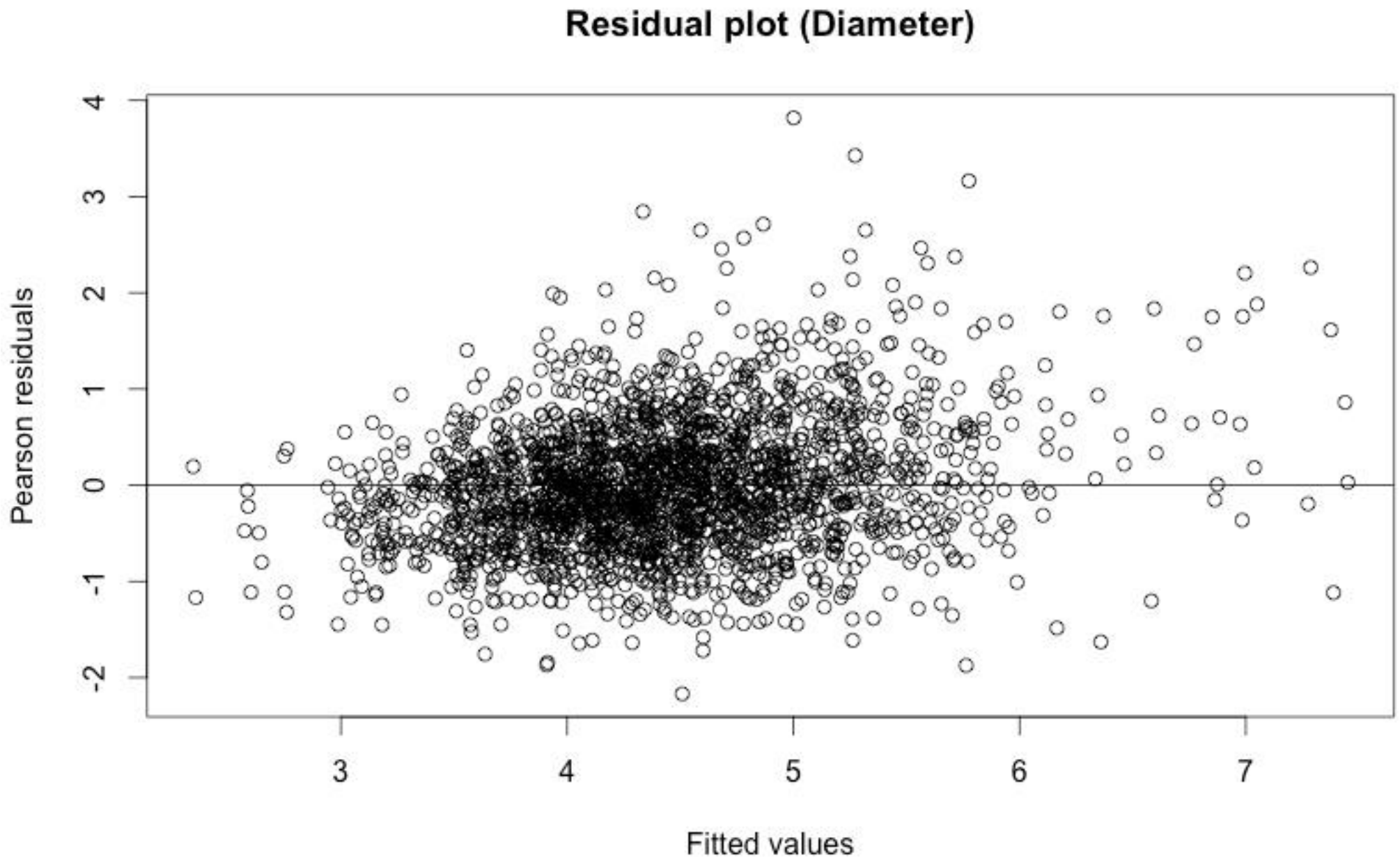

Q-Q plot for residuals (Height)
Sample Quantiles
Theoretical Quantiles
20
15
10
5
0
-5
-10
-15
-3
-2
-1
0
1
2
3

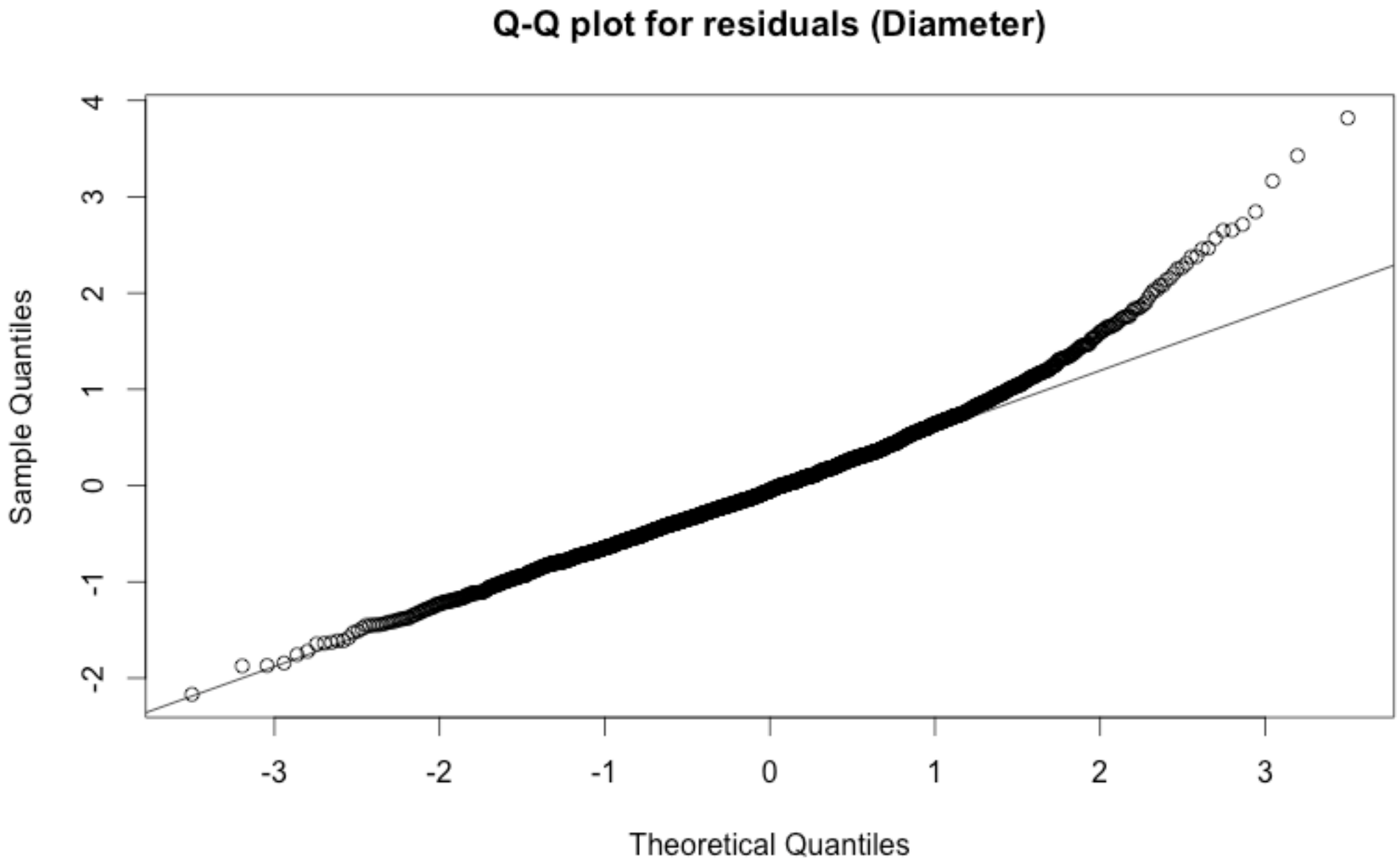

Q-Q plot for the random effect (Height)
Sample Quantiles
Theoretical Quantiles
-4
-2
0
2
4
6
-3
-2
-1
0
1
2
3

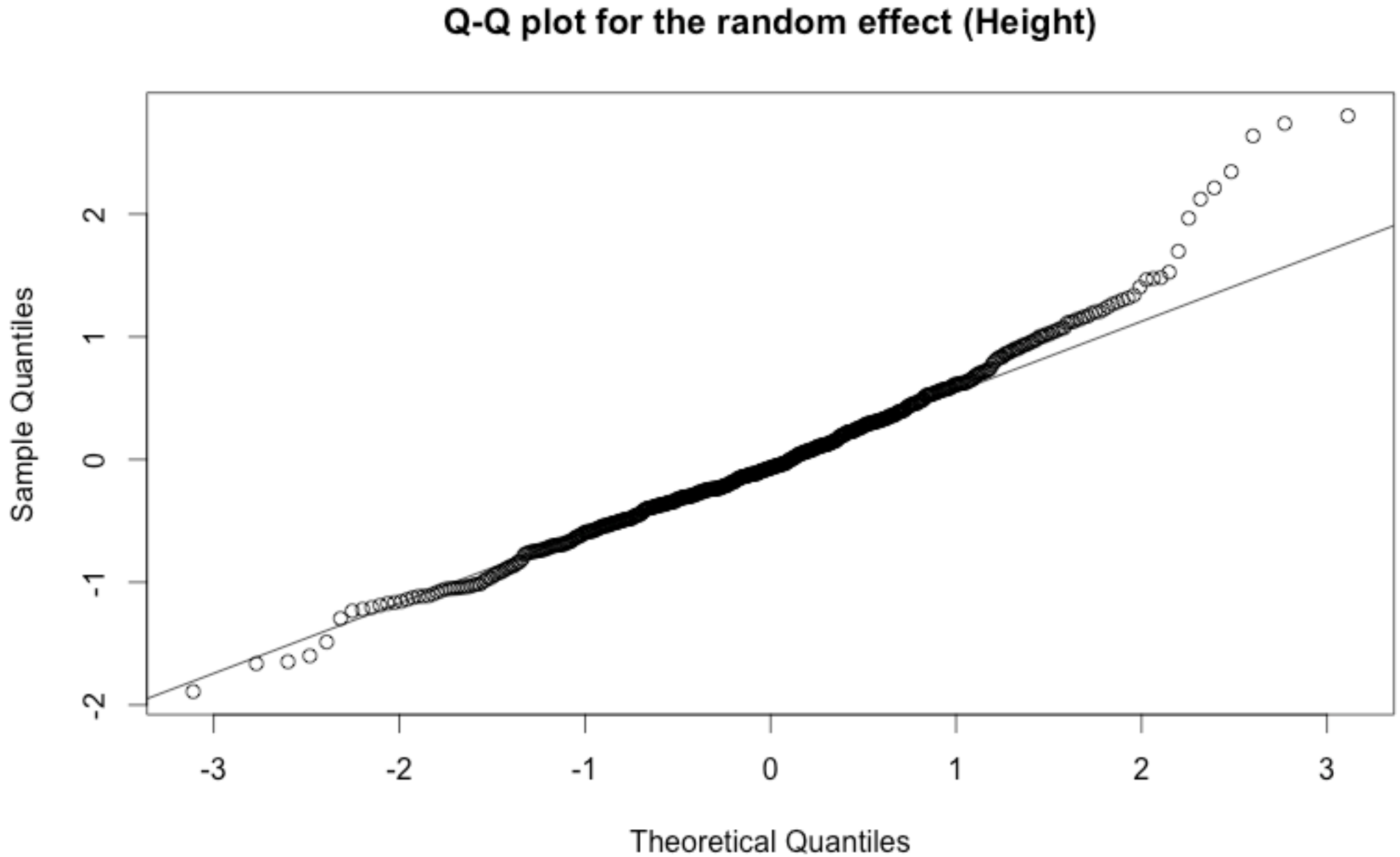
Heritability (Diameter): 48.7%

LMM: Diameter ~ Media + Block + (1 | Genotype)

Fixed Effects	Estimate	Std. Error	P-value
(Intercept)	4.63838	0.04309	0.0000
Media (Soil)	0.01194	0.03307	0.7181
Block (2)	-0.40118	0.03307	0.0000
Random Effects	**Name**	**Variance**	**Std. Dev.**
Genotype	(intercept)	0.5569	0.7462
Residual		0.5871	0.7662


Residual plot (Diameter)
Pearson residuals
4
3
2
1
0
-1
-2
3
4
5
6
7
Fitted values


Q-Q plot for residuals (Diameter)
Sample Quantiles
Theoretical Quantiles
-2
-1
0
1
2
3
4
-3
-2
-1
0
1
2
3


Q-Q plot for the random effect (Height)
Sample Quantiles
Theoretical Quantiles
-2
-1
0
1
2
-3
-2
-1
0
1
2
3

Heritability (Root-binary): 15.7%

GLMM: Root (=0/>0) ~ Media + Block + (1 | Genotype), Link: Probit Function

Fixed Effects	Estimate	Std. Error	P-value
(Intercept)	1.62129	0.08592	<2e-16
Media (Soil)	-0.85491	0.07845	<2e-16
Block (2)	-0.04349	0.07172	0.541
Random Effects	**Name**	**Variance**	**Std. Dev.**
Genotype	(intercept)	0.1868	0.4322
Environment		1	

Heritability (Root-ordinal): 12.9%

GLMM: Root (0/1/2/3) ~ Media + Block + (1 | Genotype), Link: Probit Function

Fixed Effects	Estimate	Std. Error	P-value
0/1	-1.58784	0.05871	< 2e-16
1/2	-1.20285	0.05435	< 2e-16
2/3	-0.29020	0.04833	1.91e-09
Media (Soil)	-0.67217	0.05053	< 2e-16
Block (2)	-0.27234	0.04945	3.64e-08
Random Effects	**Name**	**Variance**	**Std. Dev.**
Genotype	(intercept)	0.1483	0.3851
Environment		1	

Heritability: Combined vs. Within Media

Combined	Height	Diameter	Root (binary)	Root (ordinal)
Var(G)	7.971	0.5569	0.1868	0.1483
Var(E)	36.096	0.5871	1	1
Heritability	18.1%	48.7%	15.7%	12.9%
Media: Water				
Var(G)	2.712	0.5537	14.86	0.2203
Var(E)	11.348	0.3557	1	1
Heritability	19.3%	60.9%	93.7%	18.1%
Media: Soil				
Var(G)	20.20	0.8909	0.09824	0.1698
Var(E)	53.78	0.4269	1	1
Heritability	27.3%	67.6%	8.9%	14.5%

Heritability: Root (adjusted by H+D)

Heritability = 12.0%

Root (binary)	Name	Variance	Std. Dev.
Genotype	(intercept)	0.1358	0.3685
Environment		1	

Heritability = 9.8%

Root (ordinal)	Name	Variance	Std. Dev.
Genotype	(intercept)	0.1088	0.3299
Environment		1	

Standard GWAS Analysis

- Standard GWAS Methods
  - Quantitative trait
  - Binary trait
  - Ordinal trait
  - Count trait
- P-value Visualization
  - Q-Q plot
  - Manhattan plot
- Statistical Significance
  - Bonferroni
  - FDR

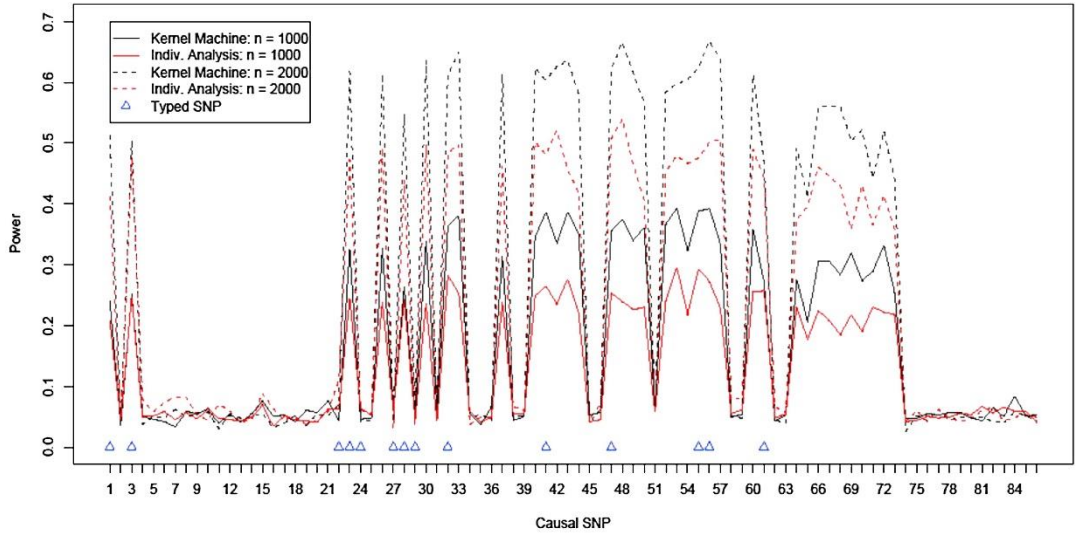
Innovations for GWAS Analysis

- SNP-set Association Analysis
- Synthetic Traits
- Non-/Semi-parametric Methods
- Multiple Traits

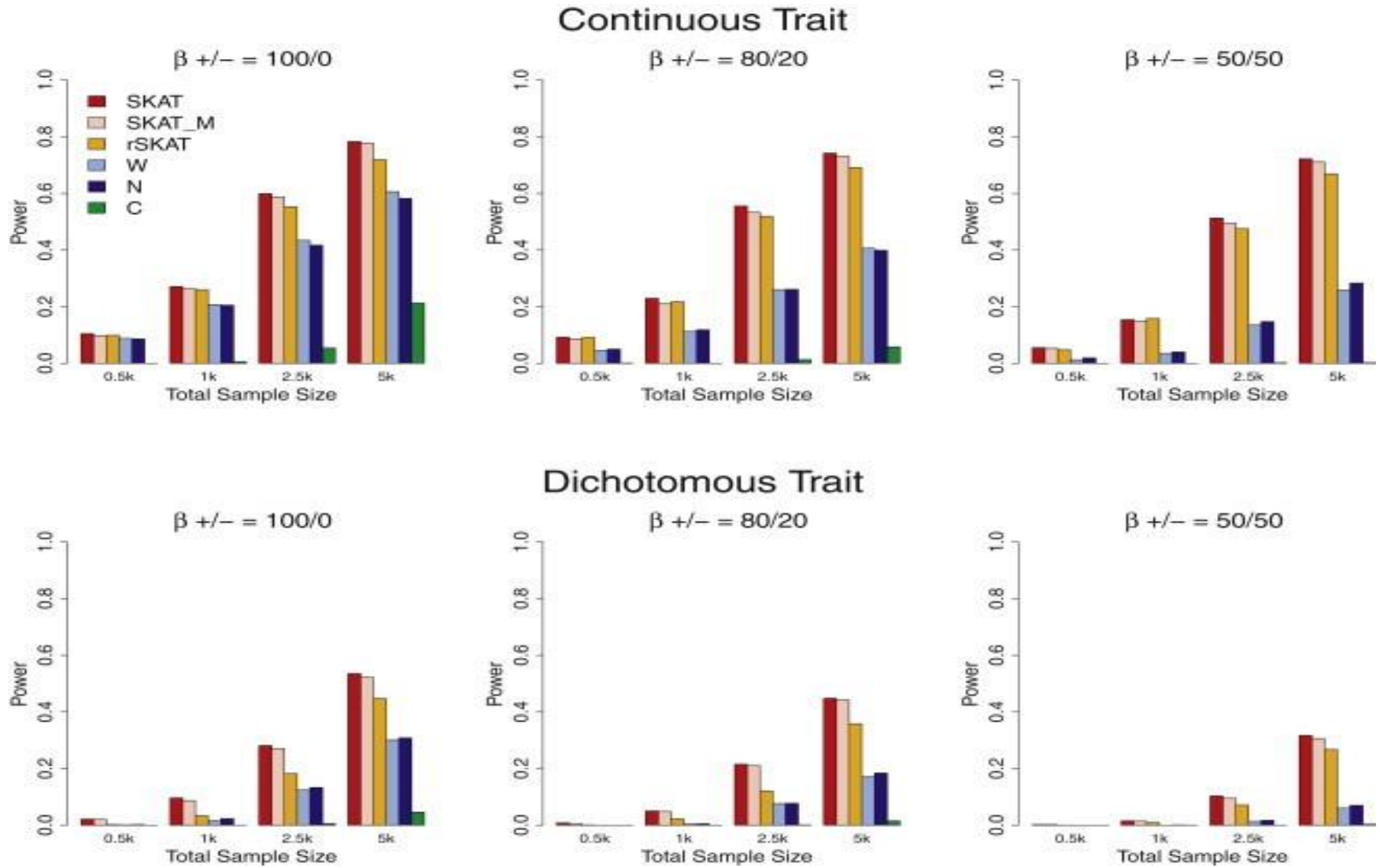
SNP-set Association Analysis

- SNPs grouped into sets of interest
  - Genes, functional pathways, or other genetic units
- Fewer hypothesis tests to be performed
- Improved statistical power after Bonferroni correction
- Examples
  - Combination of p-values: Fisher's method, min-P method
  - Principal component methods: PCA, weighted PCA, sparse PCA
  - Kernal machine methods: SKAT (SNP-set Kernel Association Tests)
- Success in human GWAS
  - New discovery of risk genes for breast cancer (FGFR2) (Wu et al. 2010)

Power Comparison (SKAT vs. Bonferroni)



Power Comparison (SKAT vs. Others)



Synthetic Traits

- Large library of traits measured from the machine-vision algorithms
- Synthetic traits by combing multiple traits using statistical tools
  - PCA: High heritability (variance)
  - Sparse PCA: A portion of key traits instead of all traits (Jolliffe et al. 2003, Zou 2006, Shen and Huang 2008)
  - Other multivariate methods
- Relationship with known traits
- Simple biological interpretation
- GWAS conducted for synthetic traits

Non-/Semi-parametric Tests

- Non-/semiparametric association tests (Zhu et al. 2012 and Jiang et al. 2014)
  - No distributional assumptions
  - Higher power
  - Robustness
- Success in human GWAS
  - New discovery of risk genes for bipolar disorder (RPGRIP1L)
  - New discovery of risk genes for addiction to opiates (PCDH9 for white men and EML2 for white women)
  - Results collected by GWAS Catalog, a database maintained by NHGRI and EBI

Application to Addiction Studies

Table 7. Significant SNPs in the genome-wide association study of a single substance dependence from association tests. op: opiates; oth: other drugs

Chr	SNP	MAF	Gene	*p*-values			
				T	$T_{W,1}$	$T_{W,2}$	$\hat{T}_{IPW}$
			op				
Black men							
2	rs2377339	0.019	NCK2	**1.1e-8**	**1.1e-9**	**1.4e-9**	**8.2e-9**
16	rs2042360	0.066	–	9.2e-7	**6.5e-8**	4.3e-7	9.6e-7
17	rs17544779	0.017	–	**5.6e-8**	6.3e-6	1.8e-6	**4.6e-8**
White men							
13	rs9529180	0.111	PCDH9	1.5e-7	4.6e-7	**4.9e-8**	1.1e-7
13	rs9540995	0.112	PCDH9	2.2e-7	7.0e-7	**5.9e-8**	1.5e-7
13	rs9529185	0.111	PCDH9	1.6e-7	4.7e-7	**5.2e-8**	1.1e-7
Black women							
5	rs2441010	0.012	–	1.0e-7	1.1e-4	8.2e-5	**7.6e-8**
7	rs2528381	0.084	UBE2D4	1.9e-5	**5.1e-8**	2.9e-5	1.6e-5
7	rs1182398	0.014	UBE3C	1.9e-7	**5.6e-8**	1.2e-6	1.1e-7
10	rs7911634	0.011	PCDH15	7.2e-5	**2.7e-9**	3.1e-6	6.6e-5
14	rs17197261	0.020	OR10G3	1.3e-5	**4.5e-8**	1.4e-3	1.0e-5
White women							
19	rs3745816	0.016	EML2	2.2e-5	**4.4e-11**	2.0e-5	1.3e-5
19	rs4445998	0.015	EML2	1.2e-5	**1.2e-11**	2.4e-5	6.7e-6
19	rs1545040	0.020	EML2	1.5e-3	**5.7e-8**	2.5e-3	1.1e-3
			oth				
Black women							
11	rs11603357	0.041	–	2.5e-7	**2.6e-8**	**1.1e-8**	1.5e-7
White women							
17	rs3098945	0.187	ANKRD13B	4.5e-6	**1.8e-8**	6.0e-7	1.1e-6

Multiple Traits

- Correlated or similar traits, e.g., *in vivo* vs. *in vitro* rooting, and regeneration vs. transformation rate, may have a similar genetic basis
  - Subcategories within a human-annotated category of traits
  - Longitudinal structure of observations (e.g., shoot length over time)
- Methods
  - Generalized Kendall's Tau (Jiang and Zhang 2011, Zhu et al. 2012, Jiang et al. 2014)
  - Scaled Multiple-phenotype Association Test (Roy et al. 2003 and Schifano et al. 2013)
- Merits
  - Flexibility: Traits can be mixture of quantitative and qualitative types
  - Higher power compared to single-trait association tests
- Success in human GWAS
  - New risk genes for smoking and drinking
  - New risk genes for dependencies to cocaine, marijuana, and opiates (ADCY4 for black men, CTNNB1 & MMP16 for white men, RASAL2 & PCDH15 for black women, MPV17 for white women)
  - Results collected by GWAS Catalog, a database maintained by NHGRI and EBI

Application to Substance Dependencies

Table 8. Significant SNPs in the genome-wide association study of multiple substance dependencies

Chr	SNP	MAF	Gene	*p*-values			
				T	$T_{W,1}$	$T_{W,2}$	$\hat{T}_{IPW}$
Black Men							
2	rs2377339*	0.019	NCK2	1.1e-06	**6.2e-08**	**1.4e-07**	9.0e-07
5	rs251133	0.406	STARD4-AS1	5.3e-07	5.2e-06	2.8e-05	**4.2e-07**
5	rs10483285	0.037	ADCY4	2.4e-03	**1.3e-07**	5.0e-05	2.0e-03
White Men							
3	rs4016435	0.042	CTNNB1	7.3e-07	6.2e-07	**1.5e-07**	**2.6e-07**
8	rs1477908	0.177	MMP16	1.1e-05	2.3e-05	**2.3e-07**	4.1e-06
Black Women							
1	rs2175254	0.035	RASAL2	2.6e-05	**4.1e-07**	1.0e-05	1.7e-05
8	rs10504824	0.014	WWP1	1.1e-06	**9.1e-09**	**2.7e-07**	5.9e-07
8	rs17609515	0.014	CPNE3	1.1e-06	**9.1e-09**	**2.7e-07**	5.9e-07
10	rs7911634*	0.011	PCDH15	1.7e-04	**1.1e-08**	1.3e-05	1.6e-04
White Women							
2	rs16866493	0.011	–	6.1e-04	**1.9e-07**	5.2e-04	3.3e-04
2	rs878167	0.010	–	1.3e-04	**4.8e-08**	1.0e-04	6.4e-05
2	rs6731600	0.039	–	2.1e-05	9.7e-06	**7.1e-08**	5.2e-06
2	rs6721762	0.039	MPV17	3.2e-05	1.1e-05	**2.3e-07**	8.7e-06
11	rs955396	0.068	TOLLIP/MUC5B	4.4e-05	1.5e-06	**9.3e-08**	4.4e-05
19	rs3745816*	0.016	EML2	5.2e-05	**8.8e-10**	1.7e-04	4.6e-05
19	rs4445998*	0.015	EML2	5.4e-05	**3.8e-10**	3.1e-04	4.6e-05
19	rs1545040*	0.020	EML2	6.7e-04	**1.6e-07**	2.4e-03	6.8e-04

NOTE: The symbol * indicates that the same SNP is also found by single-trait analysis in Table 7.

Home

Wellington Muchero presents…

"Genome-Wide Association Mapping Studies (GWAS) for Identifying Mutations Underlying Tissue Regeneration"

Genome-Wide Association Mapping Studies (GWAS) for identifying mutations underlying tissue regeneration

Wellington Muchero

Project goals

- Establish uniform greenhouse population of cuttings
- Study *in vivo* rooting methods and *in vivo* callusing
- Optimize *in vitro* growth and culture of poplar
- Perform genome wide association study (GWAS) of RT

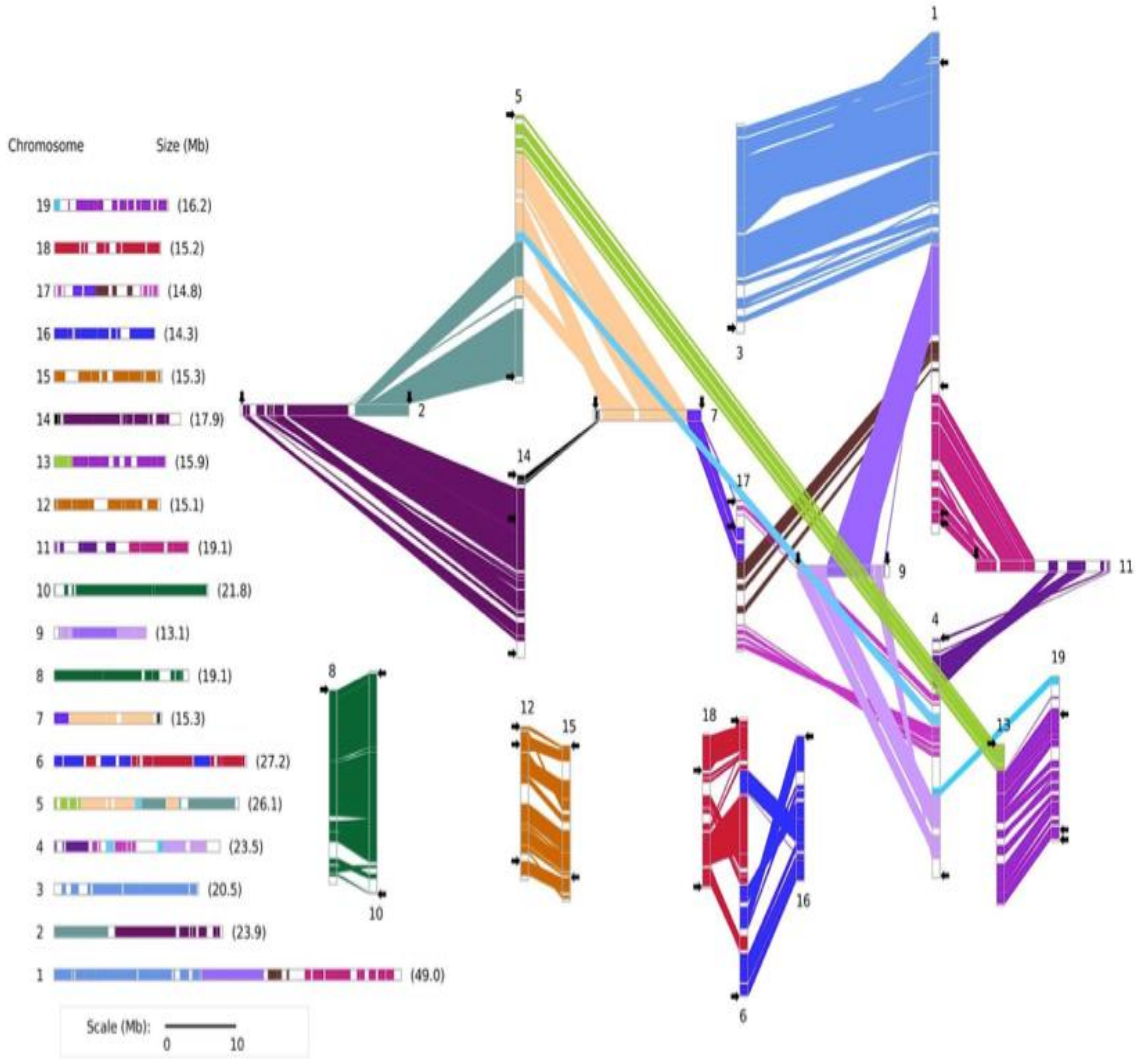
Outline

1. **Phenotype-to-Genotype correlations in *Populus*,**
2. **The *Populus* whole-genome resequencing project,**
3. **GWAS Pipeline and Future Innovations,**
4. **The power of GWAS: Examples from on-going projects,**
5. **Preliminary results,**
6. **Computational Resources and training opportunities.**

Why use *Populus* to link genes to phenotypes?

- ***Undomesticated*** Long-lived ***perennial*** harbors wide phenotypic variation in almost any trait measured,
- Relatively small genome (~480 Mb) amenable for high coverage sequencing for high-confidence SNP calling,
- Rapid Linkage Disequilibrium decay allows for single-gene mapping resolution ***(Slavov et al. 2012)***,
- Ability to ID genes with ***no a priori candidate selection*** ***(Evans et al. 2014 Nature Genetics (Bud phenology) ; Muchero et al. 2015 BMC Genomics (cell wall chemistry)***

The *Populus* reference genome



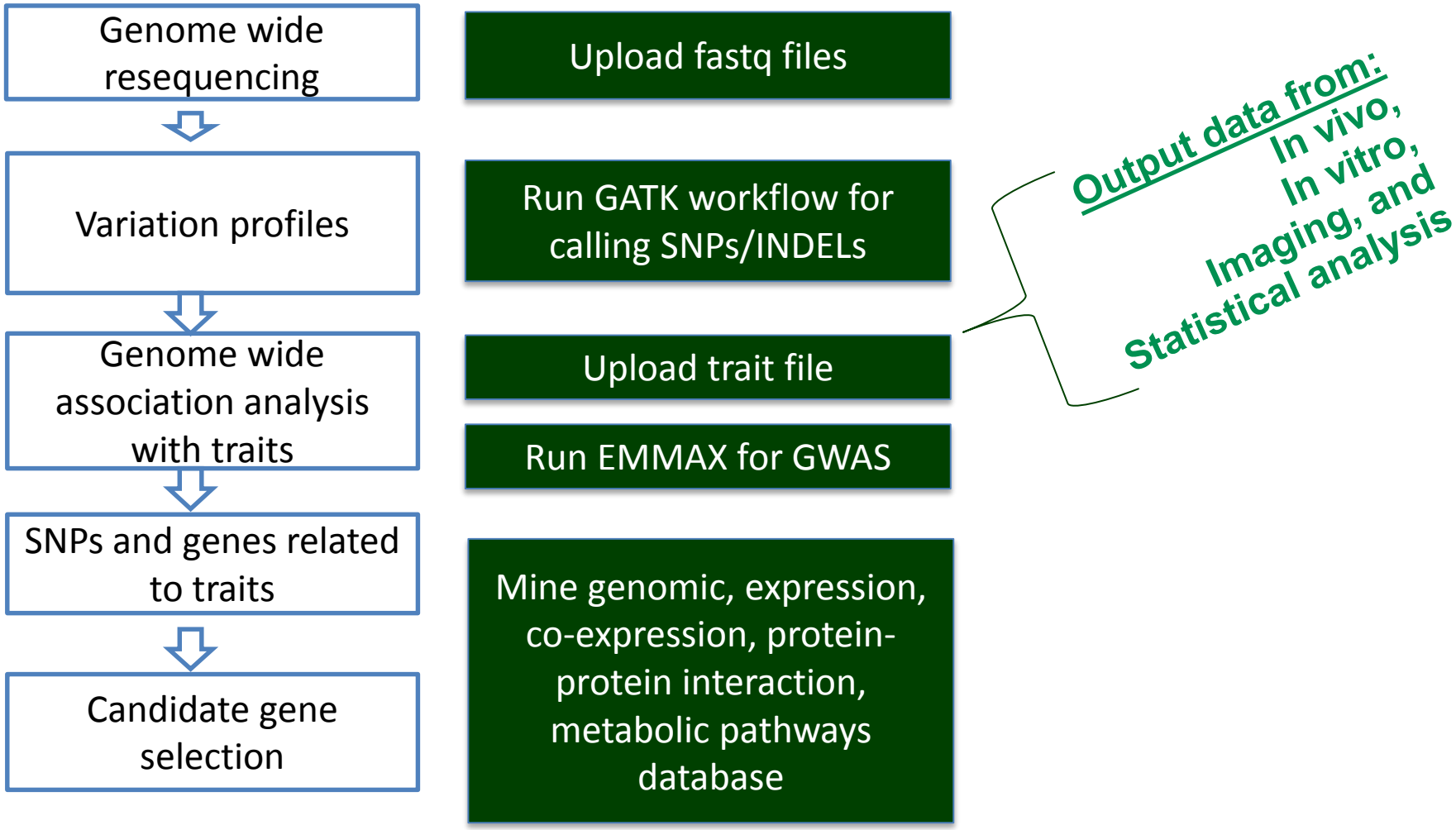
Tuskan et al. 2006

Populus GWAS re-sequencing project

- ~1,400 unrelated *P. trichocarpa* genotypes collected in Western USA/Canada,
- After excluding closely related individuals, 1,084 genotypes were re-sequenced to minimum 15X depth,
- > 48 Million single nucleotide and indel polymorphisms identified by alligning to reference genome,
- That's about 1 mutation every 10 bases,
- 8.2 Million-SNP GWAS panel was selected after eliminating extremely rare variants,
- > 26K of these are located in unmapped scaffolds of the reference genome,
- Only 104 gaps >10 Kilo-bases in main assembly given resulting in almost complete coverage of ~45,000 genes.

Slavov et al. 2012 New Phytologist

Current GWAS Pipeline



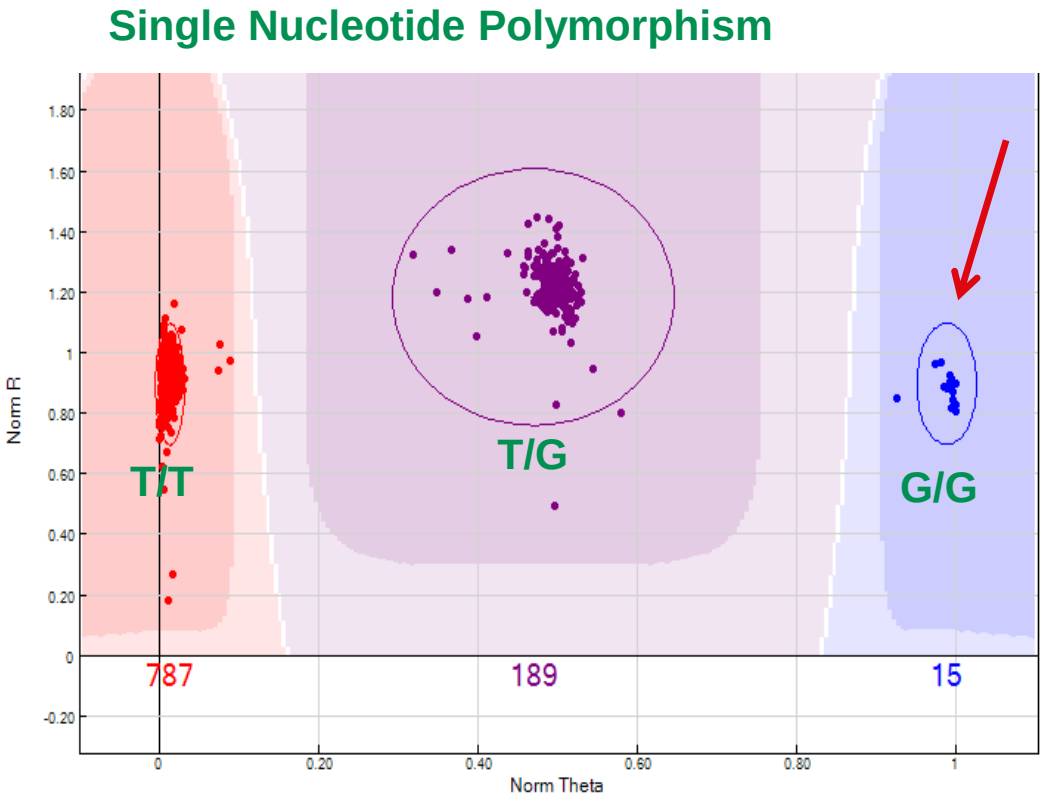
Innovations for GWAS Analysis

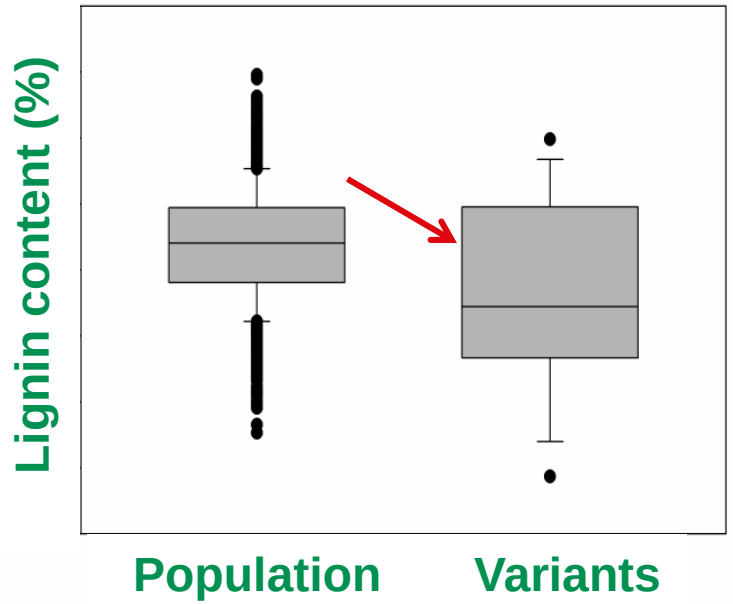
- Current SNP panel is based on Nisqually reference,
- Leverage on-going Pan and Core genome assemblies for non-reference-based variant calling,
- Preliminary results suggest up to 70, 000 genes in PAN genome compared to 45,000 in reference.

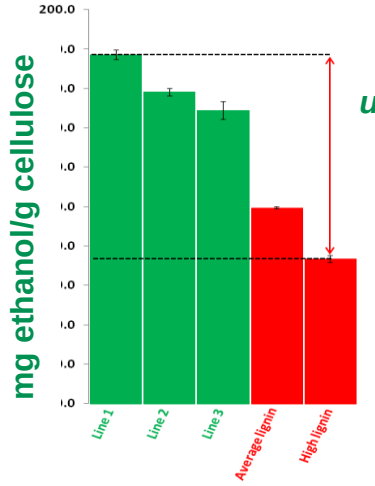
Innovations for GWAS Analysis

- SNP-set Association Analysis (Jiang)
- Synthetic Traits (Jiang)
- Multiple Traits (Jiang)
- Nonparametric/Semiparametric Methods (Jiang)

The Power of GWAS in *Populus*: Finding the unexpected



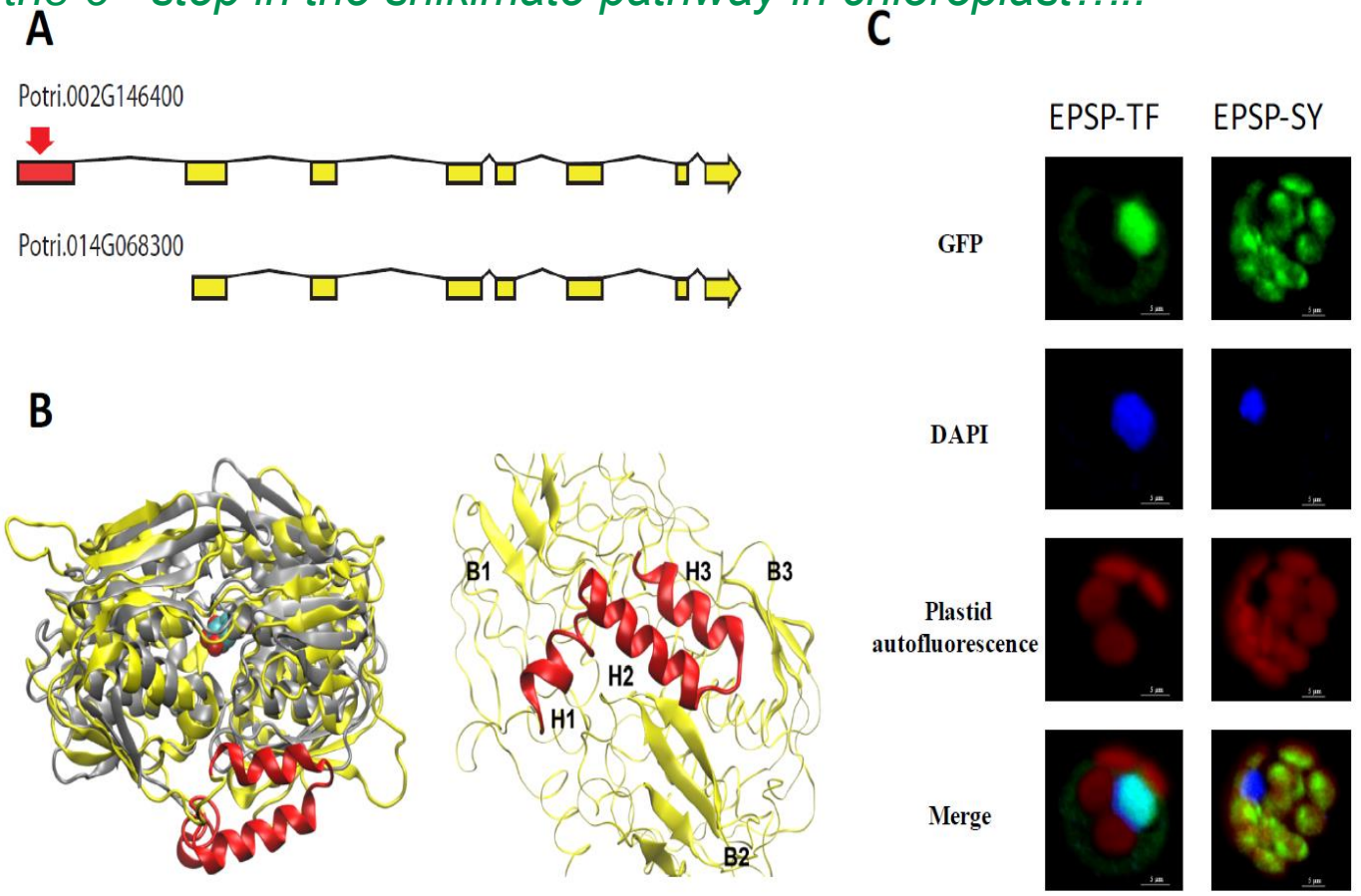




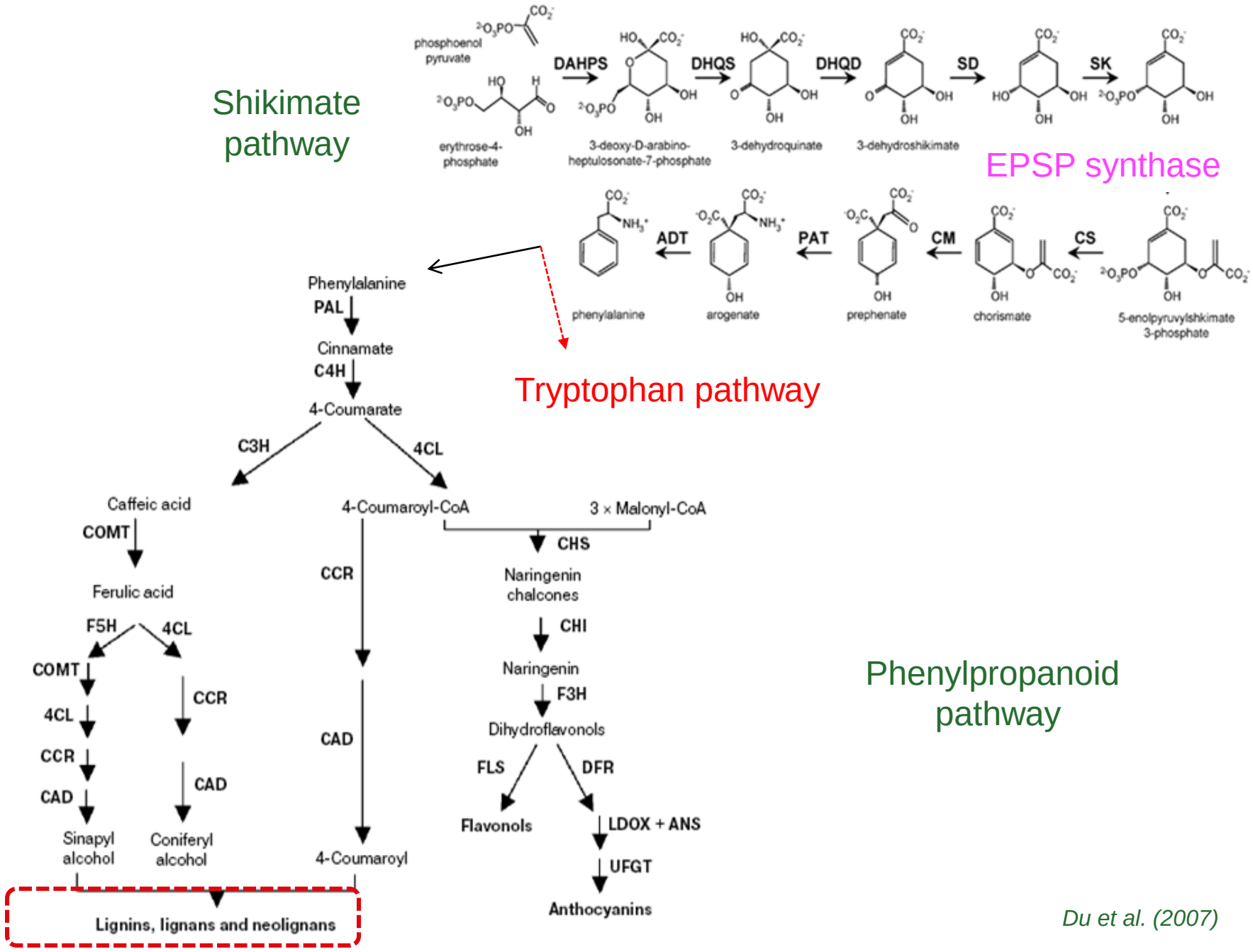
up to 250% increase

Novel regulatory function for an EPSP-like gene

5-enolpyruvylshikimate-3-phosphate (EPSP) synthase has only been known to catalyze the 6th step in the shikimate pathway in chloroplast…..

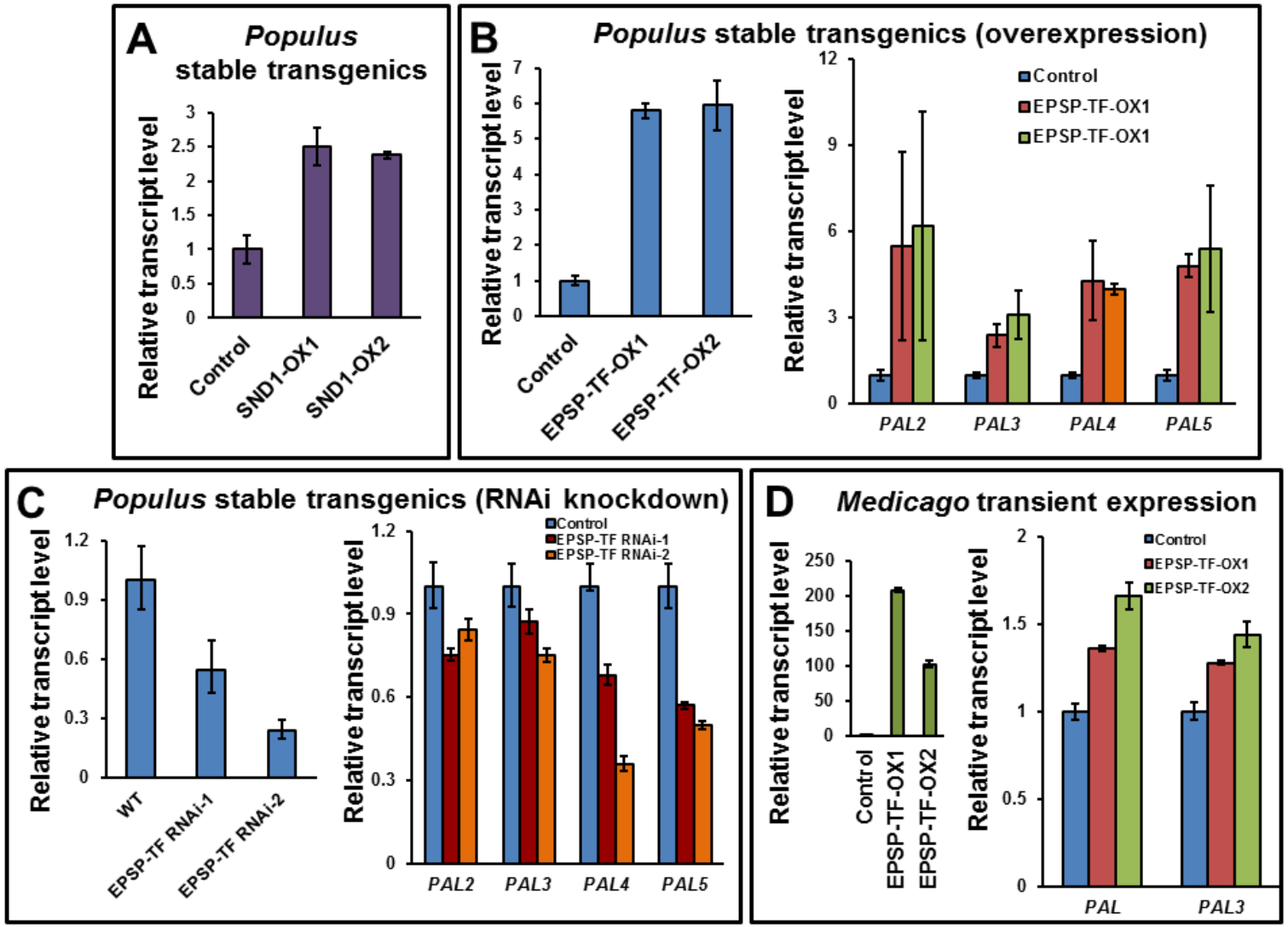


….but Populus isoform (EPSP-TF) exhibits nuclear localization


Shikimate pathway
phosphoenol pyruvate
erythrose-4-phosphate
DAHPS
3-deoxy-D-arabino-heptulosonate-7-phosphate
DHQS
3-dehydroquinate
DHQD
3-dehydroshikimate
SD
SK
EPSP synthase
5-enolpyruvylshikimate 3-phosphate
CS
chorismate
CM
prephenate
PAT
arogenate
ADT
phenylalanine
Tryptophan pathway
Phenylalanine
PAL
Cinnamate
C4H
4-Coumarate
C3H
4CL
Caffeic acid
COMT
Ferulic acid
F5H
4CL
COMT
4CL
CCR
CAD
Sinapyl alcohol
CCR
CAD
Coniferyl alcohol
4-Coumaroyl-CoA
3 × Malonyl-CoA
CCR
CAD
4-Coumaroyl
CHS
Naringenin chalcones
CHI
Naringenin
F3H
Dihydroflavonols
FLS
DFR
Flavonols
LDOX + ANS
UFGT
Anthocyanins
Lignins, lignans and neolignans
Phenylpropanoid pathway

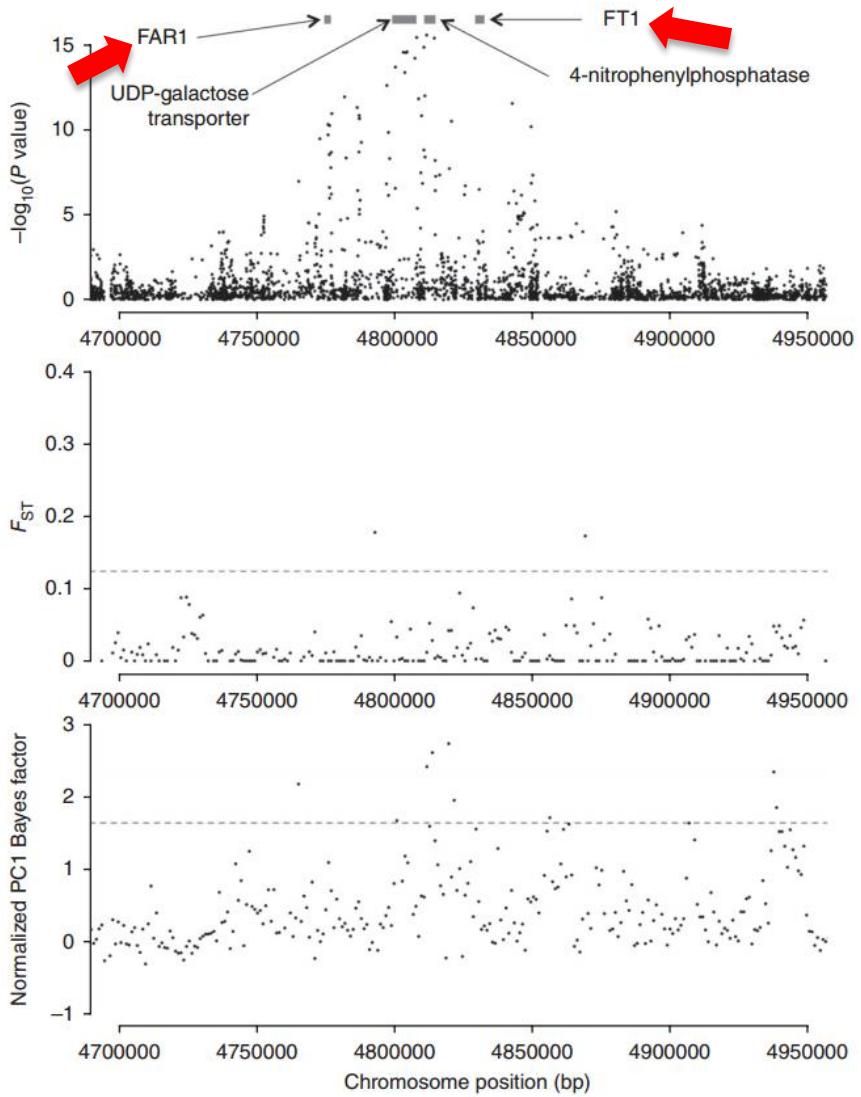
Du et al. (2007)

Transgenic validation of GWAS EPSP-TF



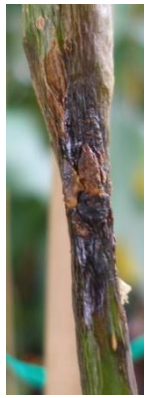
Muchero, et al. U.S. Patent No. 20,150,353,948

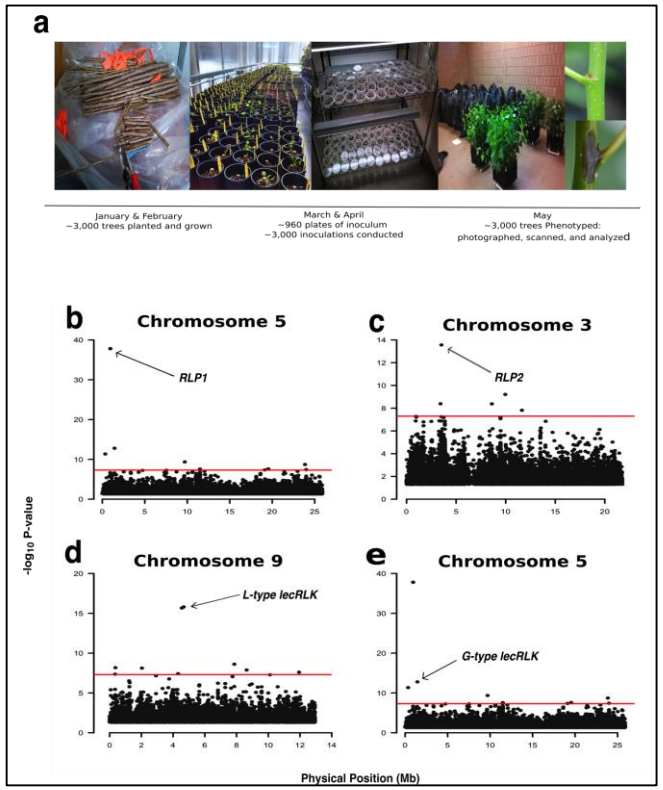
Mapping determinants of phenological traits

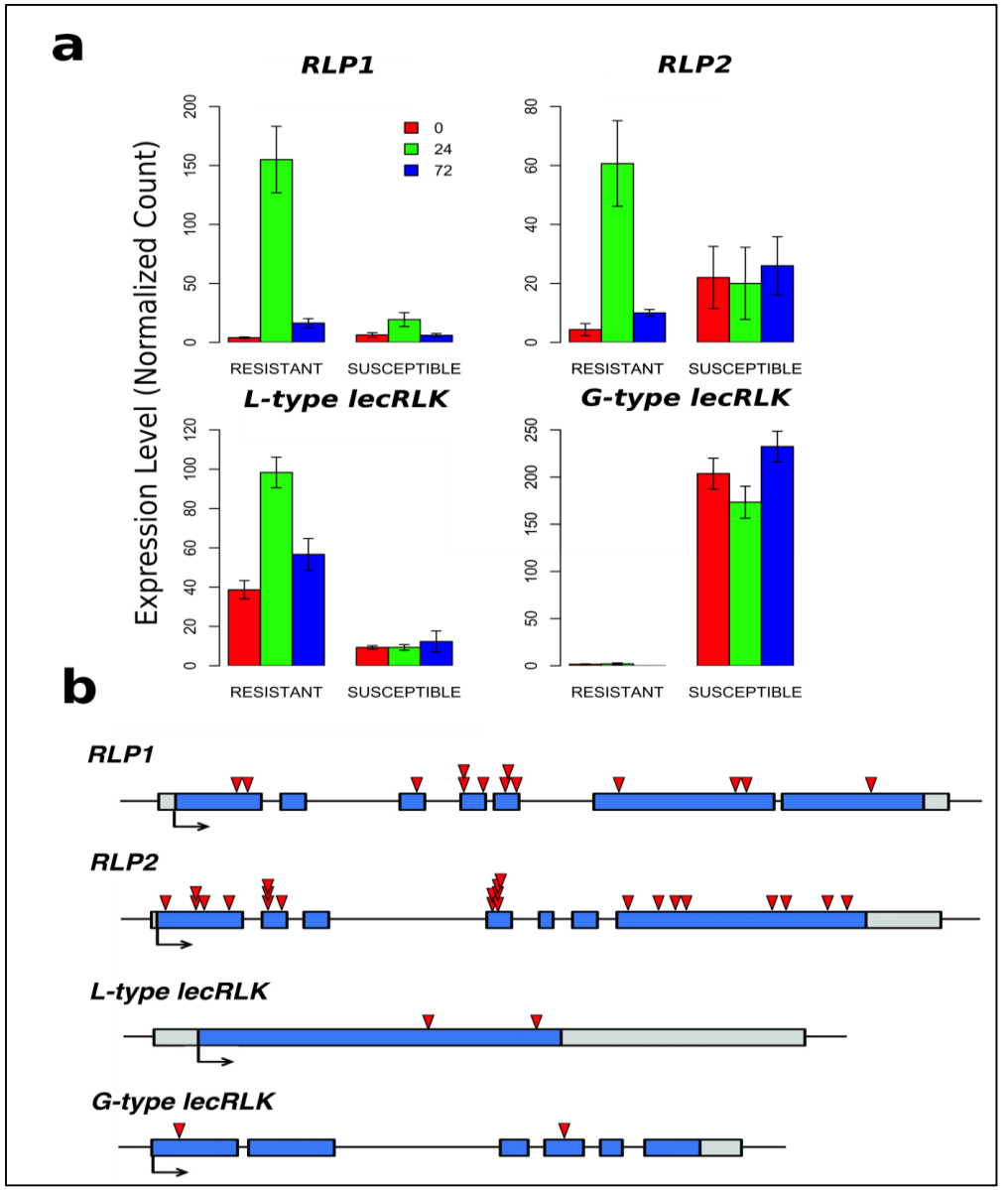


Population genomics of Populus trichocarpa identifies signatures of selection and adaptive trait associations: Evans et al. (Nature Genetics (2014)

Mapping disease resistance alleles





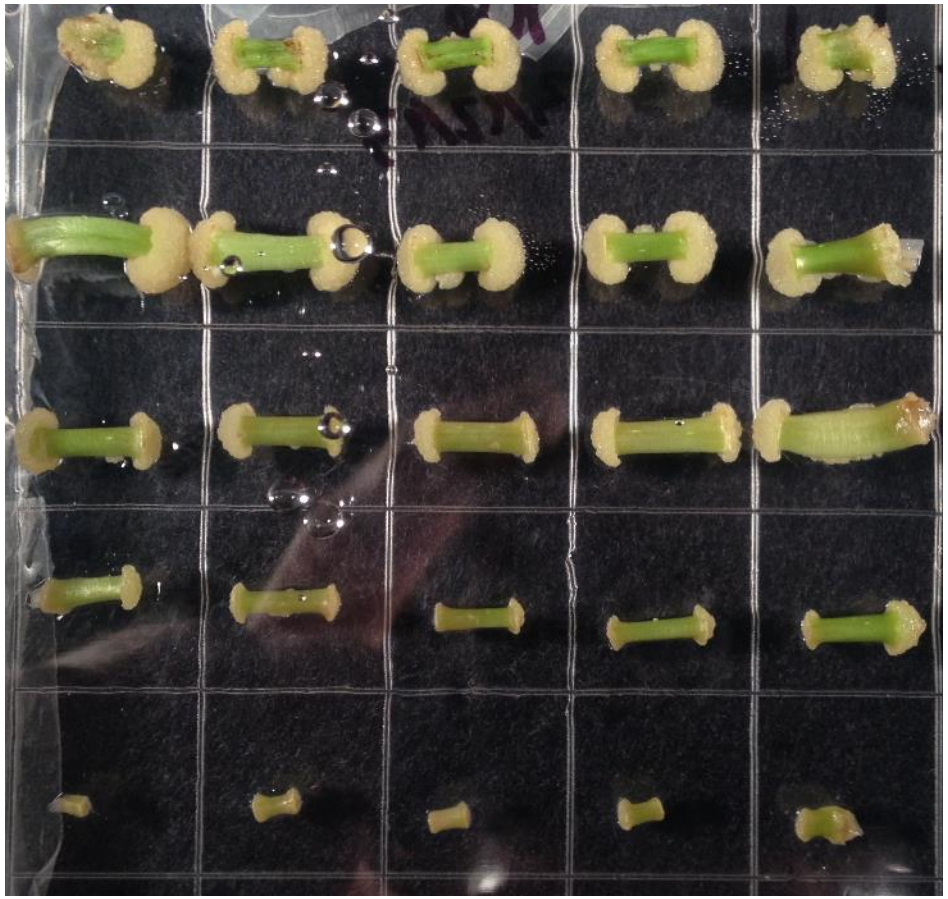


Confronting emerging pathogens: Genomics-empowered approaches to protecting ecosystem health: Muchero et al. Nature (revisions requested)

Phase 2: GWAS mapping of traits

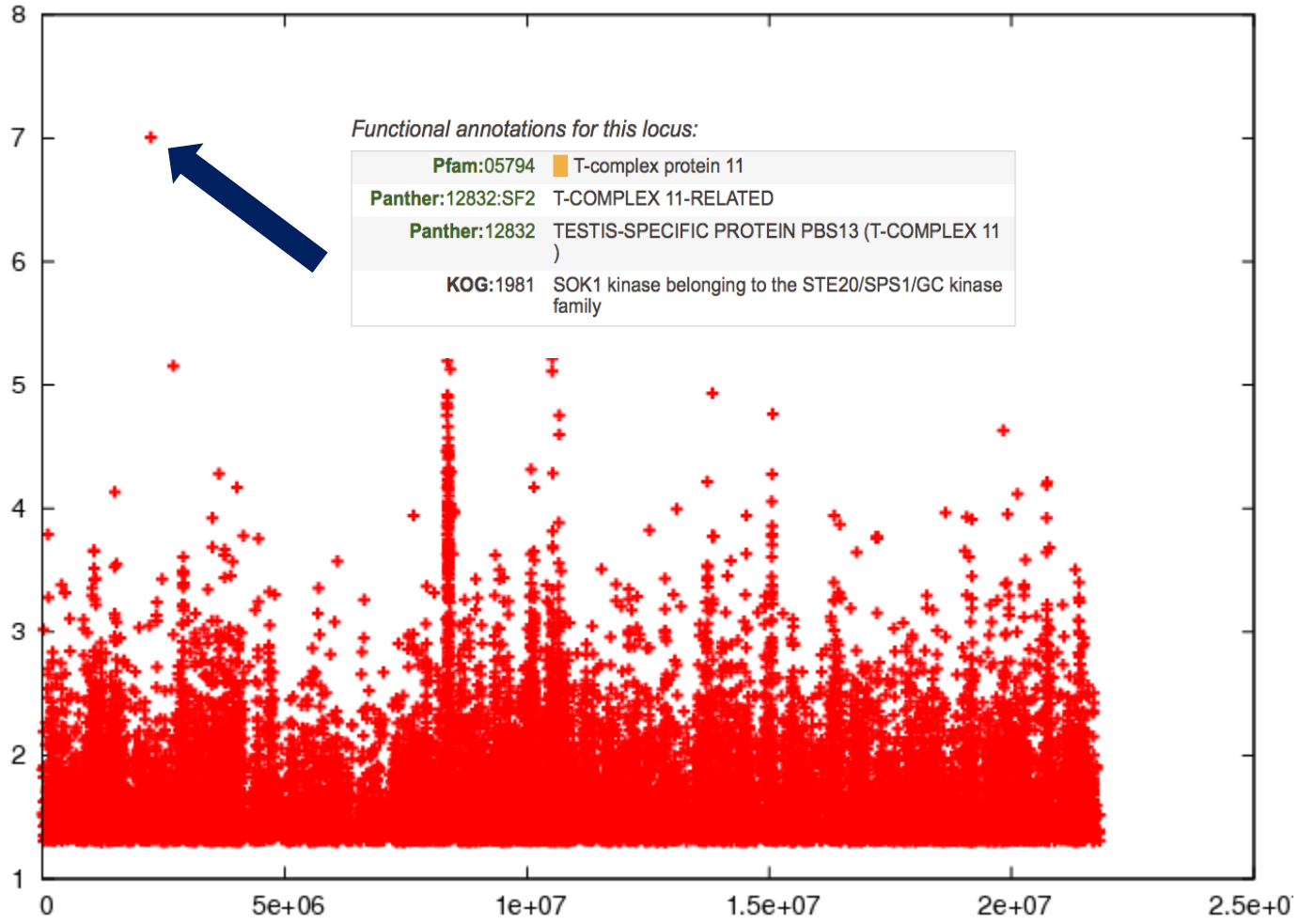
- Traits to be mapped
  - Indirect regeneration
    - Callus formation
    - Shoot formation after callus
  - Direct regeneration
    - Shoot formation from explants
    - Root formation from explants
  - Stable transformation
    - Observation of GFP and/or anthocyanin producing tissues
  - Shoot regeneration following gene treatments (BMM etc)
  - We also intend to compare callus and root formation *in vitro* and *in vivo*

Callus Formation: Preliminary GWAS Results



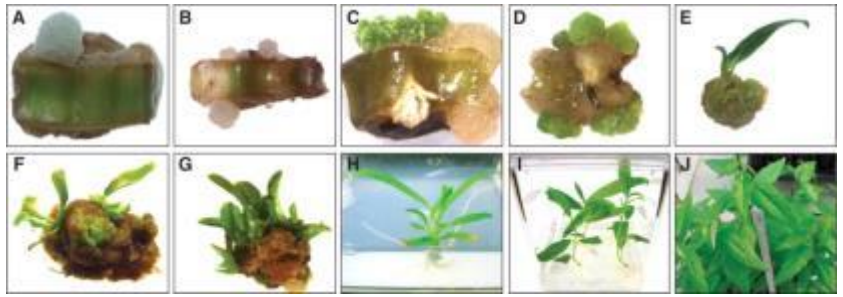
- 350 genotypes were assessed for petiole callus formation,
- 5 petioles were plated in 4 replicate plates,
- Callus formation was visually scored,
- GWAS was perfomed using EMMAX.

Callus Formation: Preliminary GWAS Results



Mammalian Sterile 20–Like Kinases in Tumor Suppression: An Emerging Pathway

Callus Formation: Preliminary GWAS Results

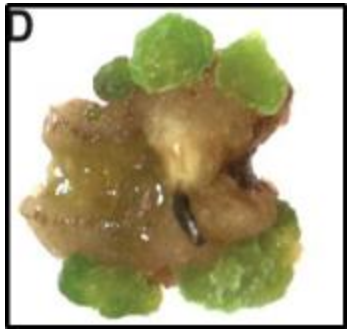


Candidate Genes

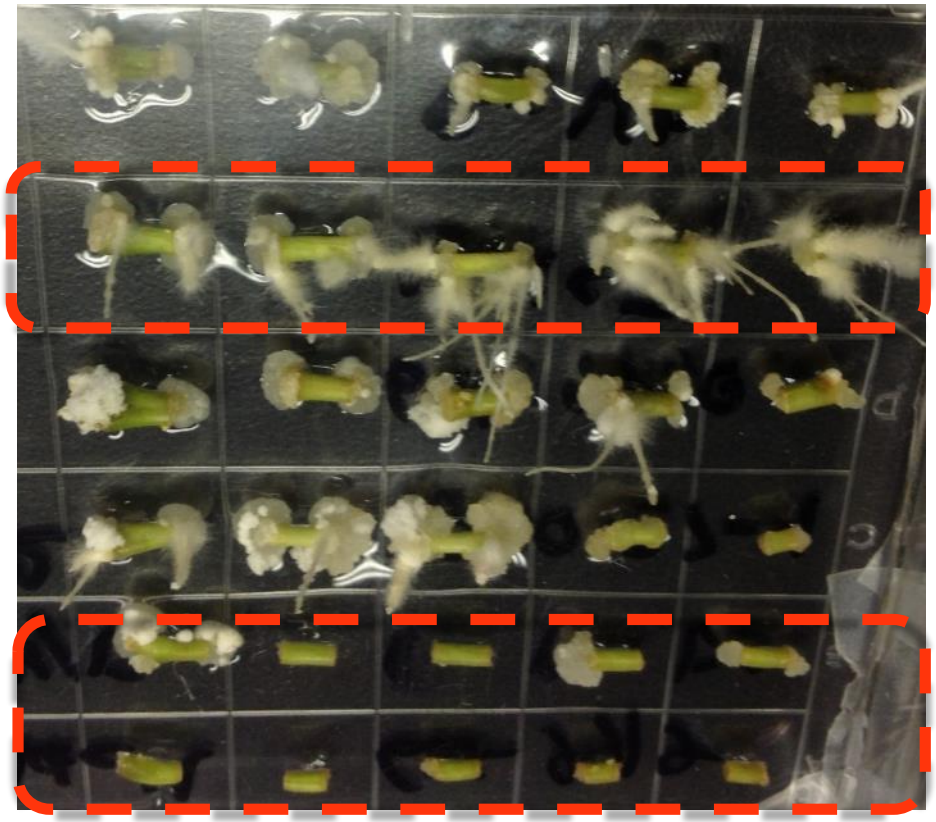
Potri.009G066100	Mitogen-activated protein kinase 3
Potri.003G018500	SOK1 kinase belonging to the STE20/SPS1/GC kinase family
Potri.006G222700	*no functional annotations for this locus*
Potri.018G014800	CHLOROPLAST NUCLEIOD DNA-BINDING-RELATED
Potri.004G118700	TARGETING PROTEIN FOR XKLP2
Potri.012G083800	HISTONE DEACETYLASE RPD3
Potri.015G023600	Ubiquitin-associated/translation elongation factor
Potri.008G208200	Rapid ALkalinization Factor (RALF)

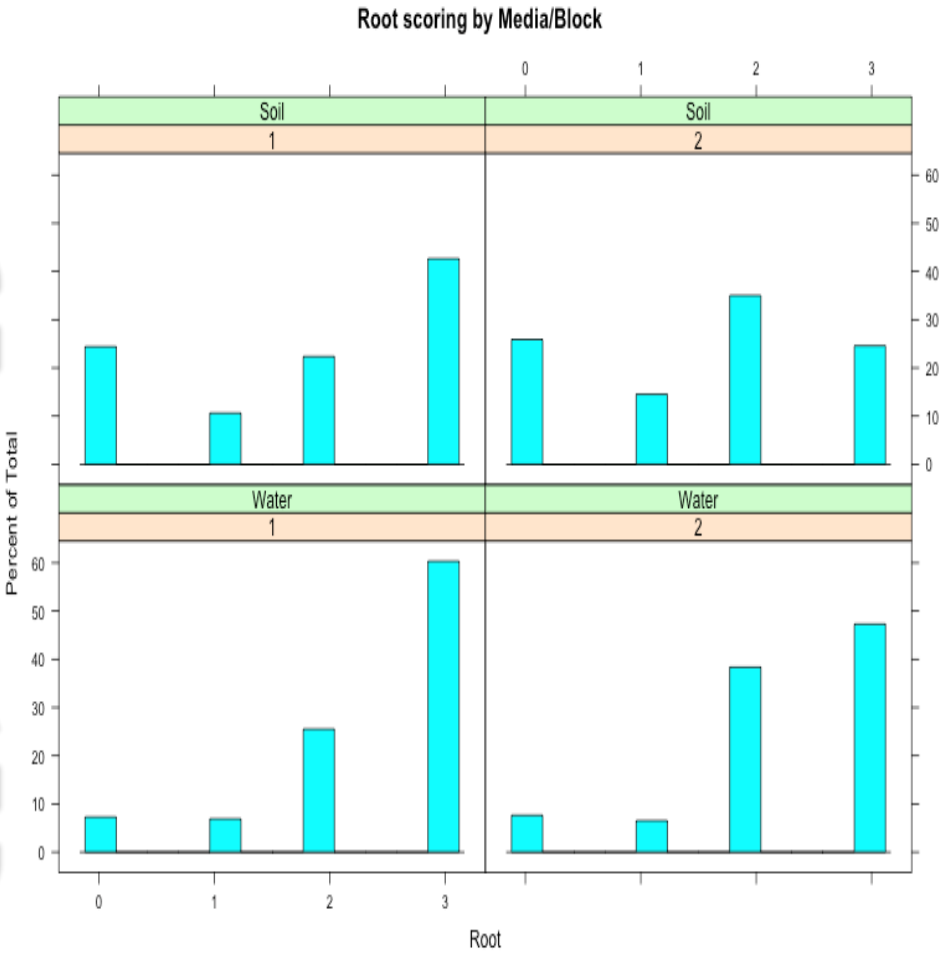
Putative Function

- Mitogen-activated protein kinase signaling pathway in human tumors.
- Mammalian Sterile 20–Like Kinases in Tumor Suppression gene
- Progressive squamous cell lung cancer
- MS-27-275, with marked in vivo antitumor activity against human tumors
- SacRALF1, a peptide signal from the grass sugarcane, is potentially involved in the regulation of tissue expansion



Natural variation in rooting rate





Muchero, Tuskan, Gunter (ORNL)

Jiang and Ma (OSU)

Natural variation in rooting rate

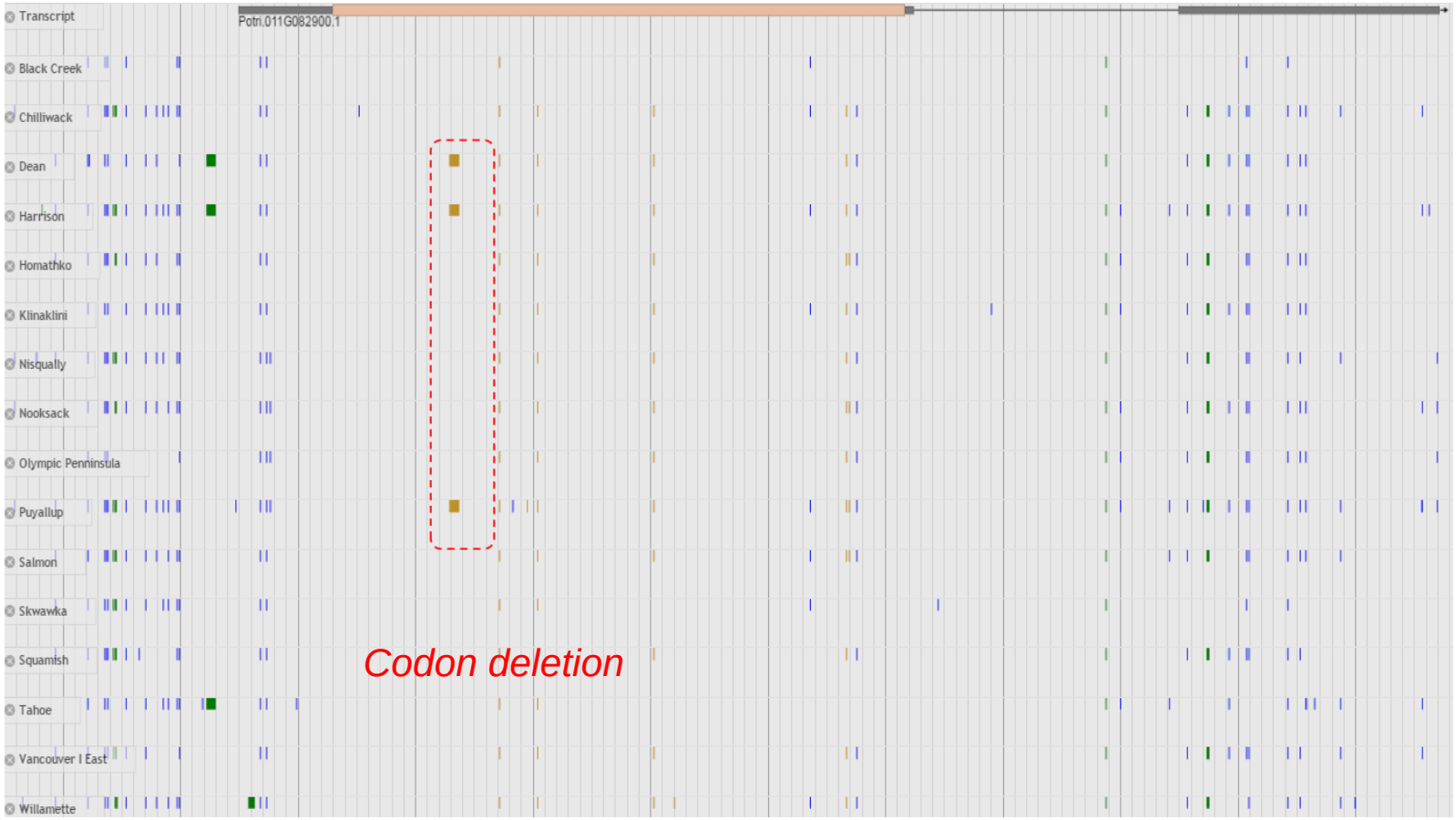
- Averaged data from 4 soil and water OSU experiments,
- Performed basic allele test against 8.2 million SNPs,
- Correction for multiple testing using Bonferroni method.

Caiping Ma (OSU)

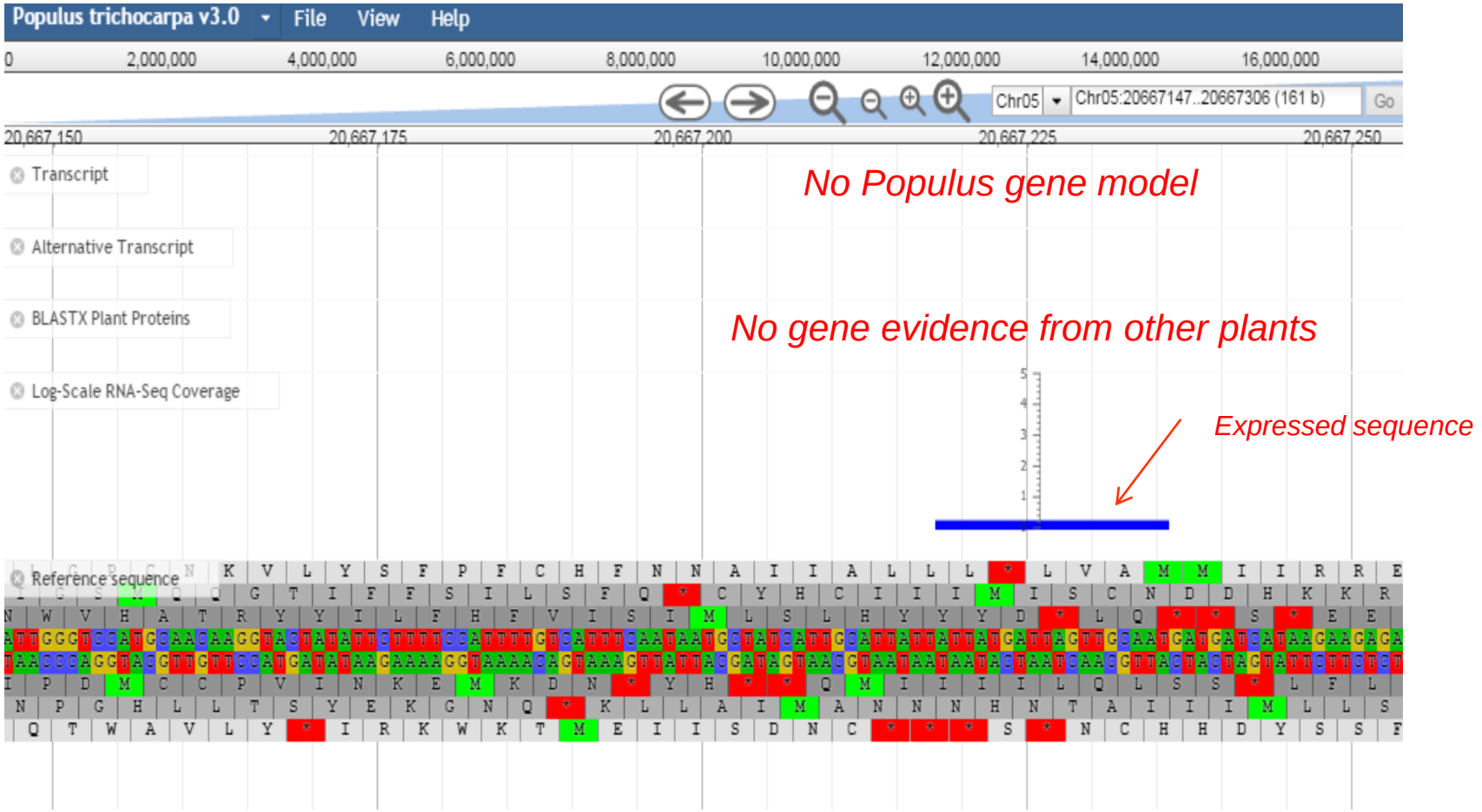
Preliminary GWAS results: rooting rate

Skp2-like F box domain Protein is a target of miR394 in stem cell maintenance

"A Protodermal miR394 Signal Defines a Region of Stem Cell Competence in the Arabidopsis Shoot Meristem (Knauer et al. Dev. Cell 24: 125-132")



Preliminary GWAS results: rooting rate



Mapping resolution allows identification of regulatory elements in intergenic regions

Data Management

- **NSF Funded UT-ORNL Joint Institute for Computational Science (offers student and visitor training), (Kraken system, > 98,000 cores),**
  - **GWAS analysis,**
  - **visualization,**
  - **output storage**
- **On-going collaboration with Joint Genome Institute for genomic and transcriptome sequencing and bioinformatics,**
  - **Sequence data analyses**
  - **Variant calling**
  - **Sequence data and variant storage**
- **ORNL is a leading institution in proposed DOE exascale computing project (ECP) and houses TITAN (3rd fastest computer in the world).**
  - **Computationally intensive simulations**

Thank you to all of our collaborators and advisory committee members, and thank you to the National Science Foundation– Plant Genome Research Program: Award Number 1546900 (CFDA 47.074)