Analysis of Genes Affecting Plant Regeneration and Transformation in Poplar

Steve Strauss, PI

NSF PGRP Advisory Meeting / October 2019

Major staff working on this project

Cathleen Ma
Transformation & Greenhouse Experiments

Kate Peremyslova
GWAS, Transformation Experiments

Julie Kucinski, GWAS, in vitro experiments

Steve Strauss
PI, Professor

Amanda Goddard
Program & Field Manager

Michael Nagle
PhD Candidate, GWAS, Transformation Genes

Jay Well,
SMILE, Education and Outreach

Fuxin Li
Co-PI, Professor, Machine Vision

Jialin Yuan
PhD Student, Machine Vision

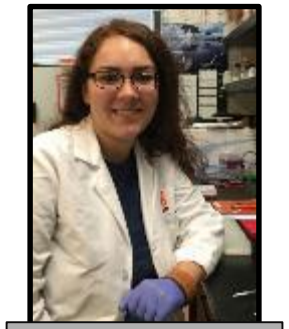
Damanpreet Kaur
PhD Student, Machine Vision

Yuan Jiang
Co-PI, Professor, Statistics

Troy Hall
Co-PI, Professor & Department Head, Education and Outreach

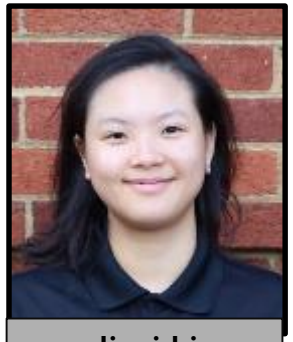
Betsy Emery
PhD Student, Education and Outreach

Other important people working on this project



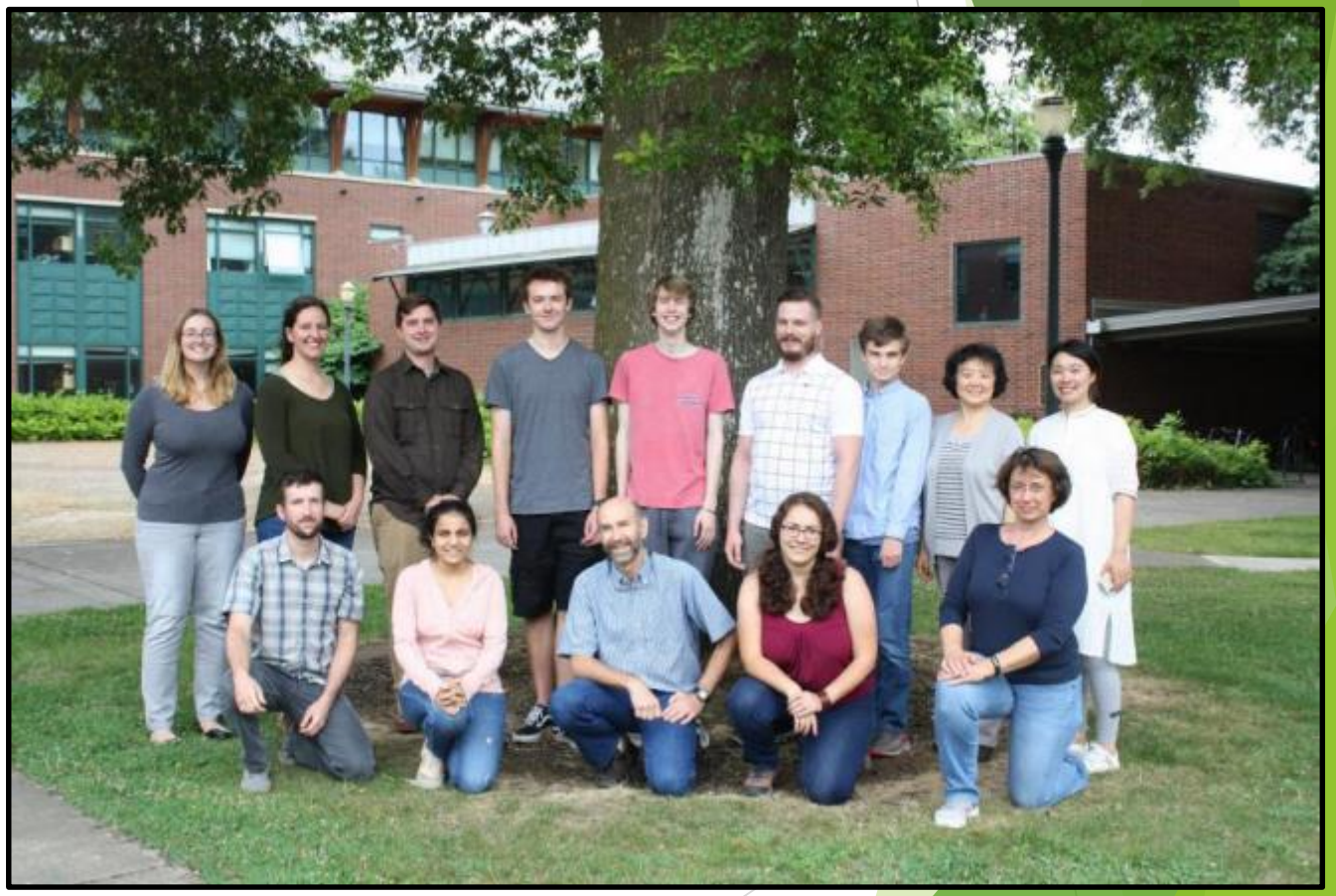
Bahiya Zahl
Undergraduate

Surbhi Nahata
Master's Student



Jiayi Li
Undergraduate

Alyssa Andrews
Undergraduate



Agenda

- Check-in, introductions and logistics
- Advisory meeting goals and agenda review
- Project overview
- Science reports
- Final discussions / responses

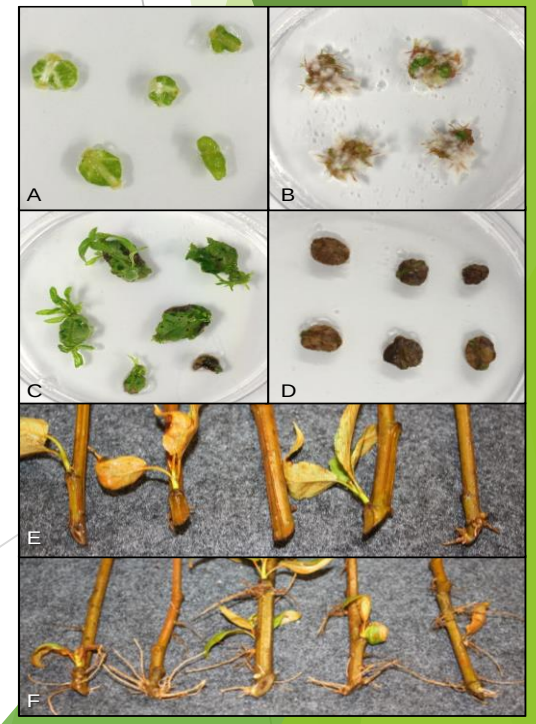
- Appendices at back
  - Organization chart
  - Leadership staff and advisory committee names / emails
  - Products to date (posters and talks)
  - Original project goals and timelines

Advisory committee meeting agenda

Time start	Topic	Speaker/s - Organizer
8:45	Logistics/connections/self-introductions	Goddard (Strauss)
9:00	Overview of project and progress	Strauss (Goddard)
9:30	Phenomics I: Plat materials, *in vitro* and transgenic adaptation studies, GWAS phenotyping methods, greenhouse management	Ma (Peremyslova, Nagle, Jiang, Strauss)
10:15	Phenomics II: Experimental imaging and image analysis pipeline, DEV gene study example/s	Nagle (Ma, Peremyslova, Jiang)
11:00	**BREAK**	
11:15	Phenomics III: Machine vision analysis systems	Li (Yuan, Damanpreet, Nagle)
12:15	**Lunch**	
1:00	GWAS pipeline and results	Nagle (Jiang)
1:45	Advances in integrated analysis of GWAS and eQTN studies in *Populus trichocarpa*	Muchero – ORNL and UT
2:15	**BREAK**	
2:30	Broader impacts: Education and curricula, social science	Hall (Well, Emery)
3:30	Summing up: Review of action items and suggestions	Strauss (Goddard)

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, and a limiting factor for GWAS statistical efficiency
  - Developments in imaging and image analysis may be game changers



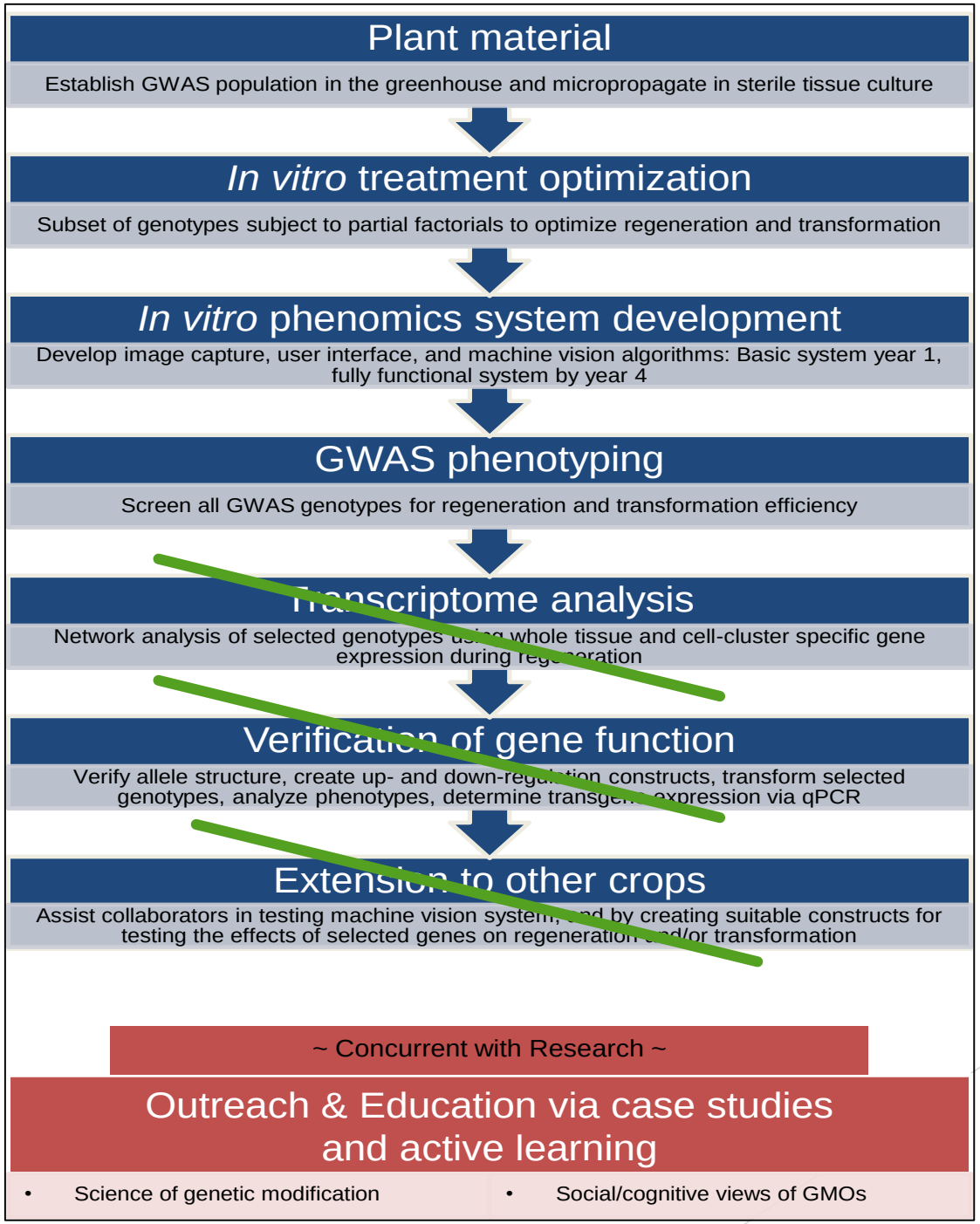
Basic science ideas behind work - 2

- 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 (and thus GWAS “mapability”) 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

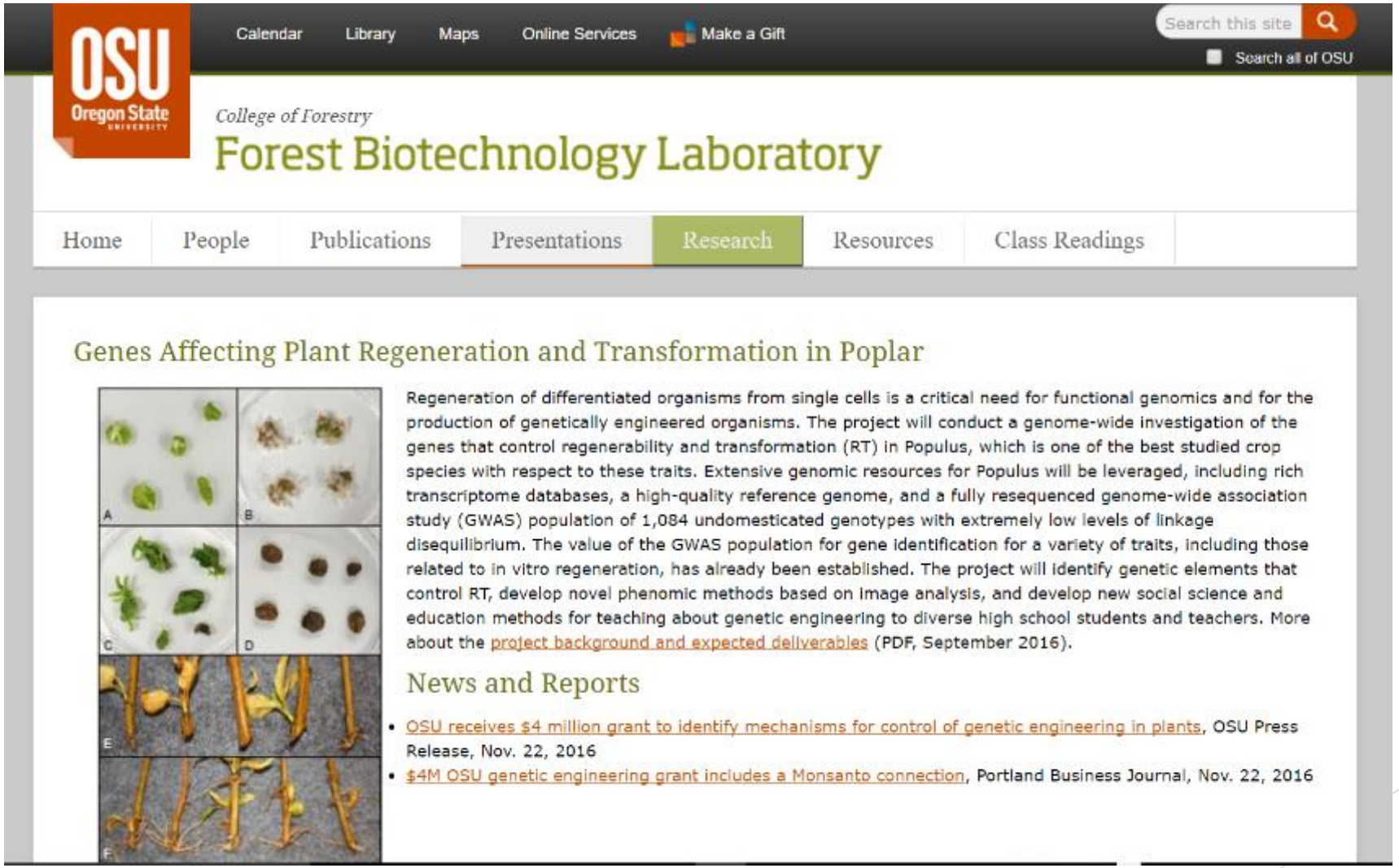
Sequence of activities



Communications and data

- **Weekly meetings** – 1+ hour each (to be extended to two shortly)
- **Monthly meetings** – More general overviews of plans and new initiatives
- **Shared cloud server** for exchanging files and analysis results (Box). Development of project database in this or other platforms.
  - Other sharing and communication platforms considered but not adopted to date
- Simple project web site, plus normal pub/poster web sites
  - Plus twitter announcements
- Core data stored in 2-3 places on cloud and hard drive (Box, Google, external drives)

Project web site



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

Summary of project budget - 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)

- Postdoc (gene constructs and bioinformatics)
  - 1 year full-time, final year

Summary of project budget – People

- Technicians – all years
  - 2-3 months/year - Project manager, Amanda Goddard
  - 4-6 months/year – Cathleen Ma (25 years experience)
  - Full-time – FRA, Kate Peremyslova
  - 6 months/year - Temp, Julie Kucinski
- Graduate Research Assistant – Time remaining
  - Machine vision (16/36 months)
  - In vitro/GWAS (29/36 months)
  - Social science/education (21/36 months)
- Student aides (undergraduate)
  - 3-18 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 ✓
- Laminar flow hood - 5 K ✓
- **~300 K inception to date**

Summary of project budget (Inception to date / original budget shown)

- Services/supplies, 5 years – 111 K / 262 K
- Participant support costs, 3 years – SMILE – 18K / 112 K
- Grad student tuition, 5 years – 72.5 K / 184 K
- Personnel, 5 years – 671 K / 2.0 M
- Indirects, 5 years – 486 K /1.1 M
- **Total cost, 5 years – 1.9 M / 4.0 M**

Major changes from submitted proposal - 1

Realizations about scale and biological complexity of work

- Use of maximal samples of genotypes
  - Statistical information on weakness of GWAS re. potential for false discovery
  - Availability of additional resequenced wild cottonwood genotypes (from ~1,000 to 1,300; ~300 with SNP data yet to be provided)
- *In vivo* source materials for GWAS: Cannot afford to maintain many hundreds of genotypes *in vitro*, thus must use sterilized greenhouse materials and contamination a serious problem
- Logistical issues of sizes of experiments (take many months, management, vigor, uniformity, sterilization of materials)
- Need to better and systematically explore *in vitro* and transformation conditions for efficient GWAS (expanded *in vitro* optimization from 1 year to 2 years)
- In short: Manpower is limiting given cuts, new realities

Major changes from submitted proposal - 2

Realizations about scale and biological complexity of work

- Complexity, effort needed to develop visualization and machine vision tools into routine, portable, efficient, and web-based systems for biologists was underestimated
  - Gap between interests/skills, and manpower, of machine learning staff and biologists
- Computation requirements for machine vision and advanced GWAS, especially with resampling, large SNP datasets, high resolution or hyperspectral images
  - Means for linkages to CyVerse, computing grids, Kbase or others
- Challenges to choosing and interpreting GWAS algorithms from many available, but often with substantial limitations re. sample size, speed, ability to handle complex models and non-normal data, and more
- Challenges to choosing GWAS tools, and interpreting GWAS results re. candidate genes and regulatory motifs, given predominance of non-coding SNPs, locations outside of genes, complexity of real gene regulation
- Long delays in obtaining key reagents such as Agro strains and DEV genes for tests (legal hoops and bottlenecks), and needing to often subclone/modify what we are provided
- In short, manpower and computation limitations serious

Planned publications

- **Regeneration / transformation treatment optimization**
  - Regeneration
  - Genetic variation / heritability
  - Transformation
- **Phenomics**
  - Imaging system pipeline, comparison to human scoring
  - Machine vision annotation system
  - Machine vision prediction algorithms and efficiency
- **GWAS**
  - In vivo
    - Shoot regeneration
    - Root regeneration
  - In vitro
    - Direct regeneration
    - Indirect regeneration
    - Transformation treatments (Agro strains (2), acetysyringone (2), DEV gene (2)

Agenda

- Check-in, introductions and logistics
- Advisory meeting goals and agenda review
- Project overview
- Science reports
- Final discussions / responses

- Appendices at back
  - Organization chart
  - Leadership staff and advisory committee names / emails
  - Products to date (posters and talks)
  - Original project goals and timelines

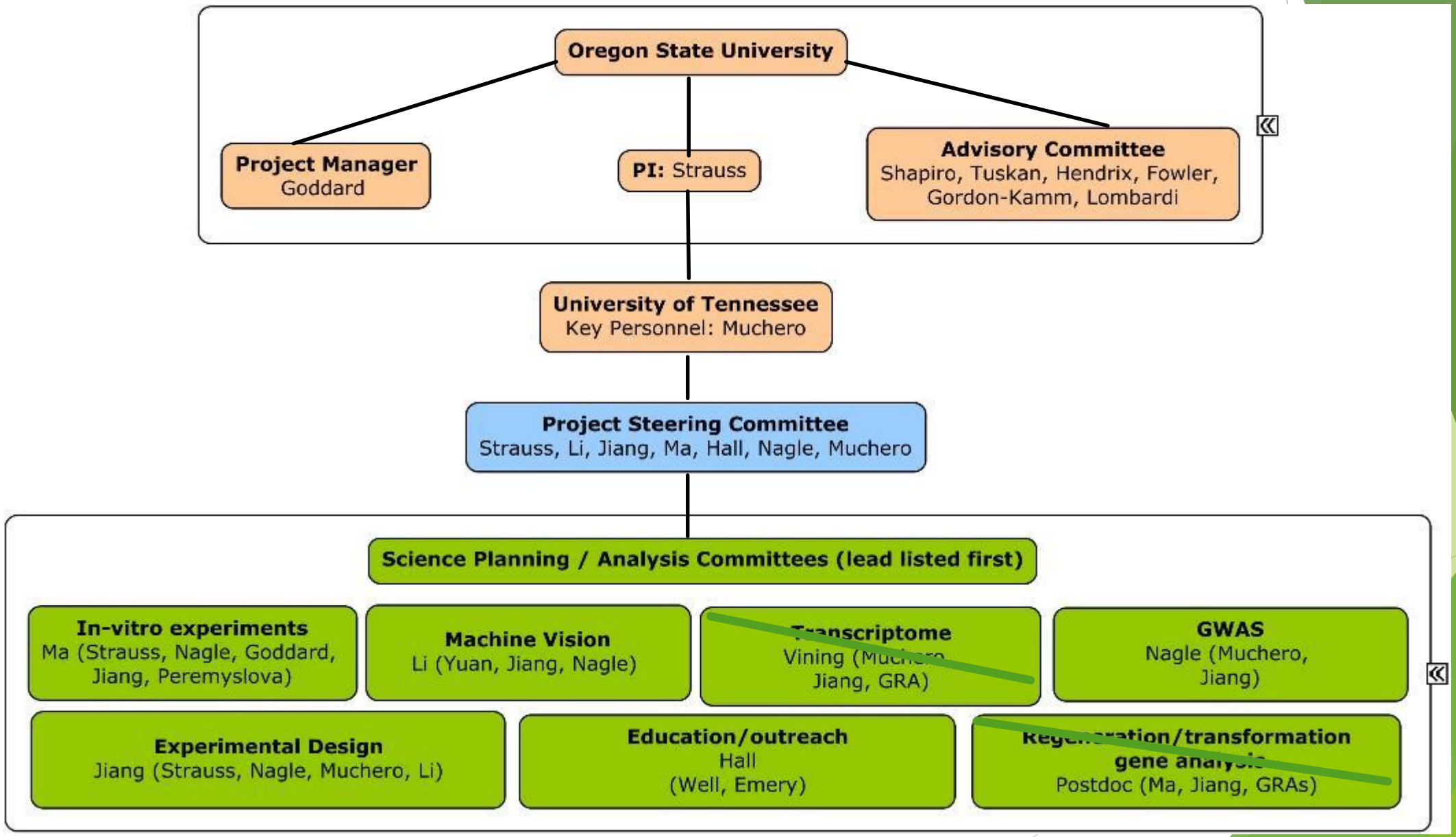
Advisory committee meeting agenda

Time start	Topic	Speaker/s - Organizer
8:45	Logistics/connections/self-introductions	Goddard (Strauss)
9:00	Overview of project and progress	Strauss (Goddard)
9:30	Phenomics I: Plat materials, *in vitro* and transgenic adaptation studies, GWAS phenotyping methods, greenhouse management	Ma (Peremyslova, Nagle, Jiang, Strauss)
10:15	Phenomics II: Experimental imaging and image analysis pipeline, DEV gene study example/s	Nagle (Ma, Peremyslova, Jiang)
11:00	**BREAK**	
11:15	Phenomics III: Machine vision analysis systems	Li (Yuan, Damanpreet, Nagle)
12:15	**Lunch**	
1:00	GWAS pipeline and results	Nagle (Jiang)
1:45	Advances in integrated analysis of GWAS and eQTN studies in *Populus trichocarpa*	Muchero – ORNL and UT
2:15	**BREAK**	
2:30	Broader impacts: Education and curricula, social science	Hall (Well, Emery)
3:30	Summing up: Review of action items and suggestions	Strauss (Goddard)

Appendices

* Organizational chart
* Staff
* Advisory Committee
* Publications / talks to date
* Original plans / schedule

Organizational chart



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.Jian@OregonState.edu
Well	Jay	Key Pers.	OSU	Jay.Well@OregonState.edu
Muchero	Wellington	Key Pers.	Univ. TN	Mucherow@ornl.gov

Advisory committee

a. Tuskan – GWAS and poplar biology
   http://www.esd.ornl.gov/PGG/tuskan_bio.htm
b. Hendrix – Transcriptome, network analysis, non-coding RNAs
   http://biochem.science.oregonstate.edu/People/david-hendrix
c. Fowler – Plant developmental and cellular biology
   http://bpp.oregonstate.edu/fowler
d. Shapiro – Machine vision
   http://homes.cs.washington.edu/~shapiro/
e. Gordon-Kamm, Pioneer/DuPont/Dow, In vitro regeneration
   https://www.researchgate.net/profile/William_Gordon-Kamm
f. Lombardi – Education, broader impacts
   https://sites.temple.edu/slrg/the-team/doug-lombardi/

Products to date -2019

- Regeneration and Transformation in *Populus trichocarpa*
  Invited talk: Forest Tree Workshop, Plant and Animal Genome Meeting, San Diego, CA
  Michael Nagle and others, January 2019
- Analysis of Genes Affecting Plant Regeneration and Transformation
  Poster presented at the NSF PGRP Awardees Meeting, Arlington, VA
  Amanda Goddard, Steve Strauss and others, September 2019
- Advanced phenotypic analysis of in vitro development and transformation for GWAS in Populus: Machine vision analysis of RGB and hyperspectral images
  Poster presented at Society for In Vitro Biology 2019 Meeting, Tampa, FL
  Michael Nagle and others
- Web-based Annotation Tool for Image-based Phenotyping
  Computer Vision Problems in Plant Phenotyping (CVPPP 2019), Long Beach, CA.
  Jialin Yuan, Zheng Zhou, Michael Nagle, Peremyslova Ekaterina, Ali Behnoudfar, Nihar A. Doshi, Ritesh Mewalal, Cathleen Ma, Anna Carlina Magnuson, Yuan Jiang, Steven H. Strauss, and Fuxin Li.

Products to date – 2019 continued

- Web- Based Deep Segmentation Tools for Phenotyping
  Poster presented at: Plant and Animal Genome Meeting, San Diego, CA
  Jialin Yuan and others, January 2019
- Genome-wide association studies of regeneration in Populus with machine vision and hyperspectral phenomics
  Poster presented at: Plant and Animal Genome Meeting, San Diego, CA
  Michael Nagle and others, January 2019

Products to date - 2018

- Next-generation phenomics in support of GWAS to Identify Genes Controlling Development of an imaging-based phenomics system for in vitro GWAS studies of plant regeneration and transformation
  Poster presented at the NSF PGRP Awardees Meeting, Arlington, VA
  Anna Magnuson, Steve Strauss and others, September 2018
- Phenomics pipeline for high-throughput image analysis of *in vitro* plant development
  Poster presented at Annual Society for Plant Biology National Meeting, Montreal
  Anna Magnuson and others, July 2018
- Toward Optimization of in vitro Regeneration and Transformation in Wild Black Cottonwood (*Populus trichocarpa*)
  Poster presented at Society for In Vitro Biology National Meeting, St. Louis, MO.
  Cathleen Ma, Steven H. Strauss and others, June 2018
- Project Overview: Analysis of genes affecting plant regeneration and transformation in poplar
  Invited presentation to SMILE teachers at Teachers Conference, OSU
  Steven Strauss, January 2018
- Identifying the genomic basis of adventitious rooting in Populus
  Genomics of regeneration in plants and animals workshop, Plant and Animal Genome XVI, San Diego, CA
  Steven Strauss, January 2018

Products to date - 2017

- Analysis of Genes Affecting Plant Regeneration and Transformation
  Poster presented at the NSF PGRP Awardees Meeting, Arlington, VA
  Steve Strauss, Brett Pierce, September 2017
- GWAS Identification of Loci Associated with Rooting in Populus
  Poster presented at the IUFRO Tree Biotechnology conference in Conception, Chile, as well as the Society for In Vitro Biology annual meeting in Raleigh, NC
  Steve Strauss, Anna Magnuson / Cathleen Ma, June 2017

Original work plans – regen/transform methods

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

Original work plans – Phenomics and machine vision

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 PAG conference and in publication-s			X			

Original work plans – GWAS & pipelines

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

Original work plans – Broader impacts

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		

THANK YOU FOR LISTENING

Oregon State University

Phenomics I: Plant materials, *in vitro* and transgenic adaptation studies, GWAS phenotyping methods, greenhouse management

October 3, 2019

Cathleen Ma, Kate Peremyslova, and Julie Kucinski

Outline

- Field collection and materials storage
- Plant care and greenhouse management
- *In vivo* stem regeneration study
- *In vivo* rooting study
- *In vitro* regeneration optimization
- *In vitro* transformation optimization
- *In vitro* GWAS regeneration

Field collection and material storage

- Aim is to collect dormant cuttings for GWAS study
- Collection and storing
  - Harvest four 6” cuttings from new growth branches from each genotype
  - Placed cuttings separately in two Ziplock bags/genotype (2 cuttings/bag) with water proof labels
  - Stored in RH and FRL freezers
  - Materials are used for *in vivo* and *in vitro* studies

Date collected	# genotypes collected from lower Marchel	# genotypes collected from clone bank	Total # of genotypes collected from both field sites	Materials storage condition	# genotypes for in vivo stem regeneration and rooting study	# phases
Feb. 2017	833	0	833	4°C	~600	Phase 1-3
Jan. 2018	314	662	976	-10°C	~600	Phase 4-7
Feb. 2019	204	295	499	-10°C	~20	Phase 8

Plants are growing at Lower Marchel and clone bank

Collection at Lower Marchel in 2017

Collection in clone bank in 2018

Collected dormant cuttings

Plant care and greenhouse management

- Aim is to grow healthy and uniform plants for GWAS
- Dormant cuttings were used for *in vivo* stem regeneration and rooting studies, then rooted plants were transplanted in 4x9.5” long tube pots and grew in greenhouse for GWAS study
- ~1,000 genotypes (2 plants/genotype) were grown in two locations

Plant care and greenhouse management – 2 greenhouses

- One set of plants used for the study are randomly growing closed greenhouse supplemented with light 16h at 24°C
  - Plants are used year around by trimming and fertilizing with slow releaser every 3 months
  - Disease and pest monitoring carried out weekly and spray control in timely manner by OSU greenhouse crew
- Another set of plants are randomly growing in open greenhouse with natural light and go dormant in winter
  - The plants also are trimmed and fertilized every 3 months in growing season
  - These plants are used for backup

Plants grown in two greenhouses

Closed greenhouse

Open greenhouse

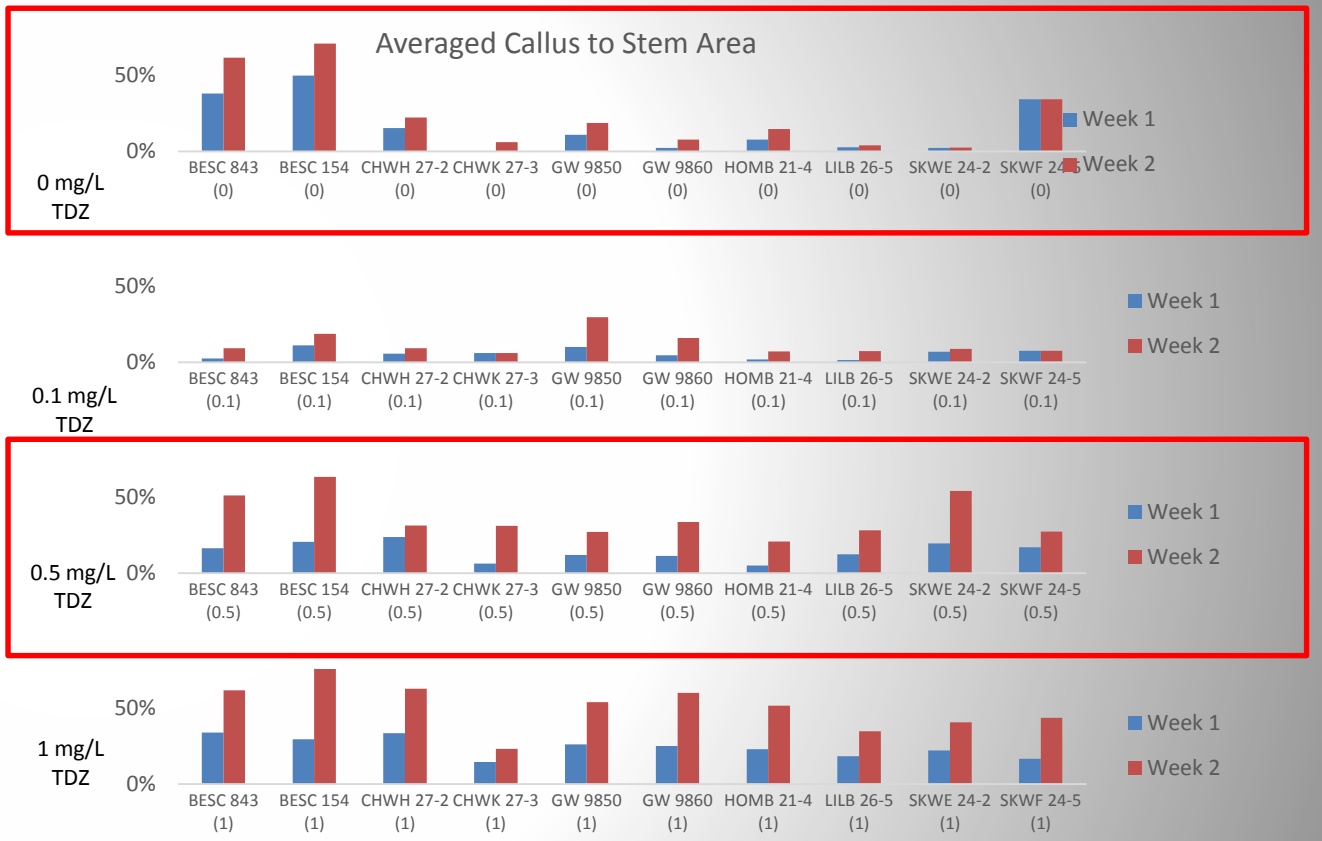
In vivo GWAS study: stem and root regeneration (completed)

- Goal is to discover genes associated with regeneration *in vivo* conditions
- Shoots first covered then roots, though mostly done at same time
- To select optimum hormone TDZ concentration, we first tested 4 levels of TDZ (0, 0.1, 0.5, and 1mg/L) in 10 genotypes for shoot development
  - Dormant cuttings were cut with one bud and planted in 50ml Flacon tube with water in head greenhouse
  - Eppendorf tube with 100 ul different levels of TDZ placed on freshly cut stem tip to hold treatment and maintained moisture for 2 days
  - Repeat application with same amount and levels TDZ for 2 days weekly to promote shoot regeneration
  - Data and imaging collected each week for five weeks
  - Manual score callus and shoot formation at week 3 and 5

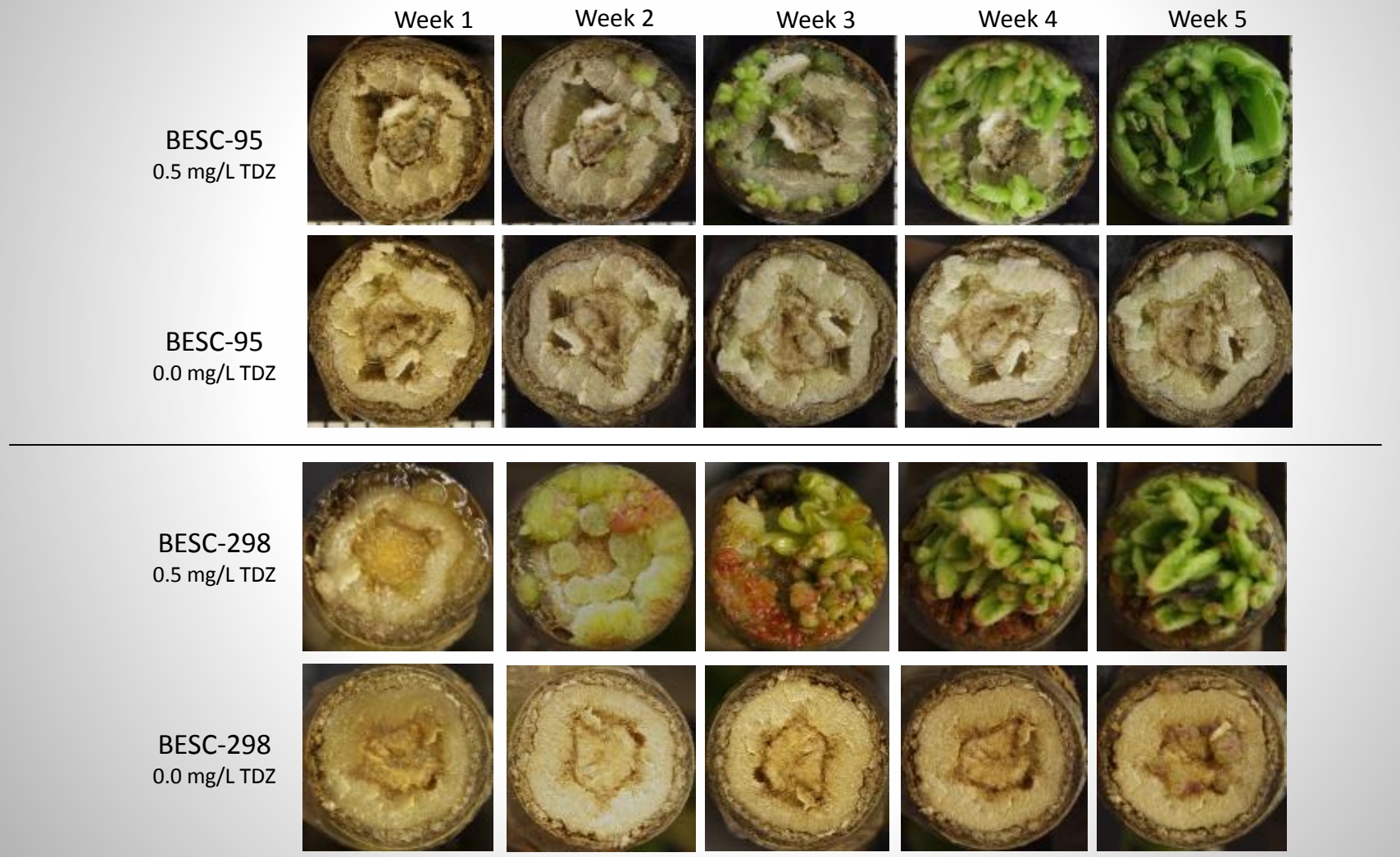
TDZ was tested to aid in callus and shoot formation from stems

0.0 mg/L TDZ

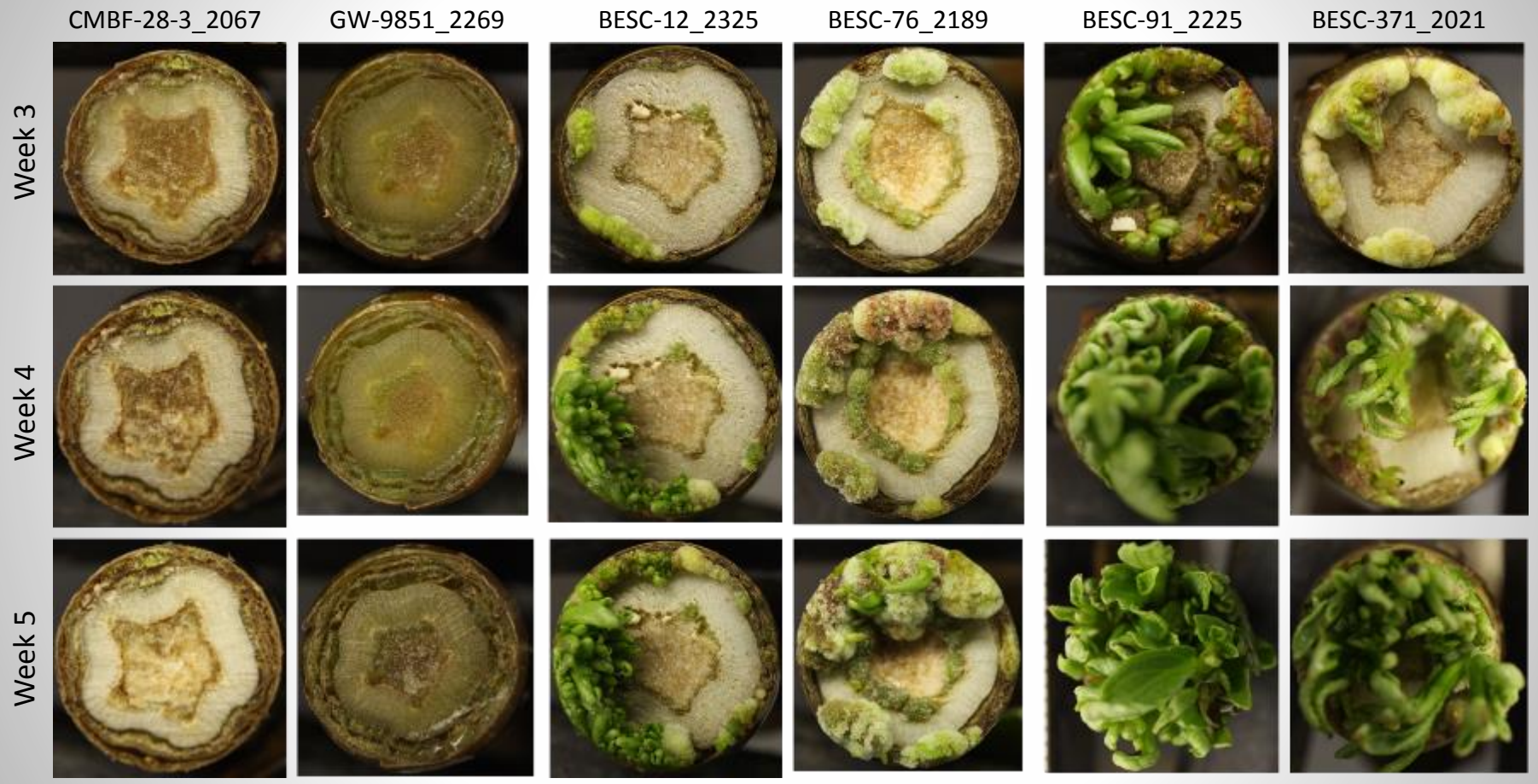
0.5 mg/L TDZ



TDZ (0.5 mg/L) promoted shoots and genetic variation



TDZ (0.5 mg/L) continued: Genetic variation in stem regeneration

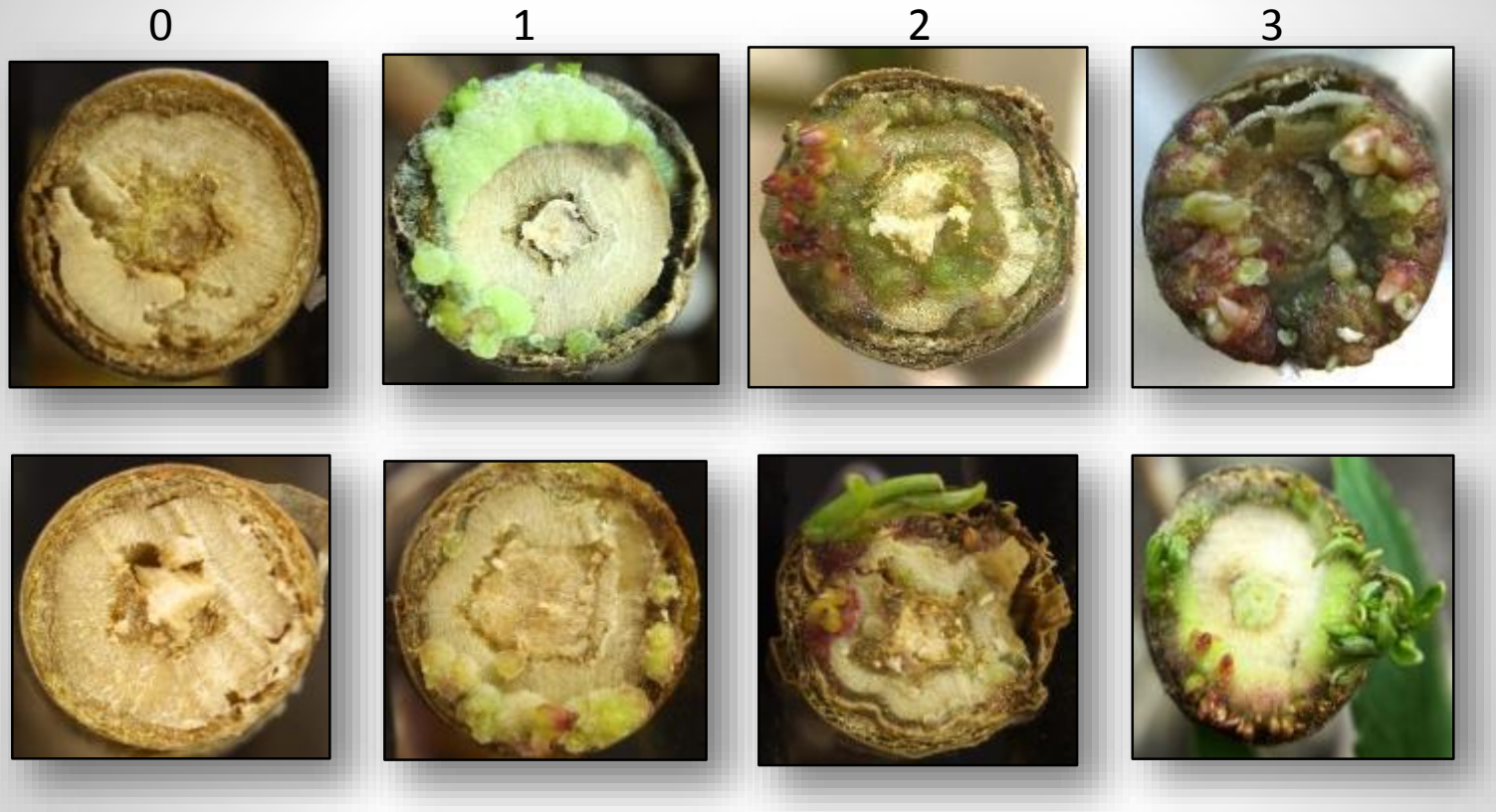


Callus formation scoring system for stem regeneration study

0 1 2 3

0= no callus
1= small callus
2= medium callus
3= large callus

Shoot formation scoring system



0= no shoot
1= 1-2 shoots
2= 3-5 shoots
3= more than 5 shoots

1,278 genotypes studied for *in vivo* stem regeneration over 8 experiments

GWAS Phase	# Genotypes studied	Date of cuttings collected	Starting dates	Dates in soil
1	200	2/8-10/17	8/3/2017	9/6/2017
2	210	2/8-10/17	8/29/2017	10/4/2017
3	200	2/8-10/17	10/5/2017	11/20/2017
4	200	1/16-17/18	3/16/2018	5/4/2018
5	196	1/16-17/18	5/9/2018	6/22/2018
6	198	1/16-17/18	7/25/2018	9/7/2018
7	57 (119 repeat)	1/16-17/18	9/26/2018	11/7/2018
8	17 (183 repeat)	2/6-7/19	3/18/2019	4/22/19
Total	**1,278 (1,580)**			

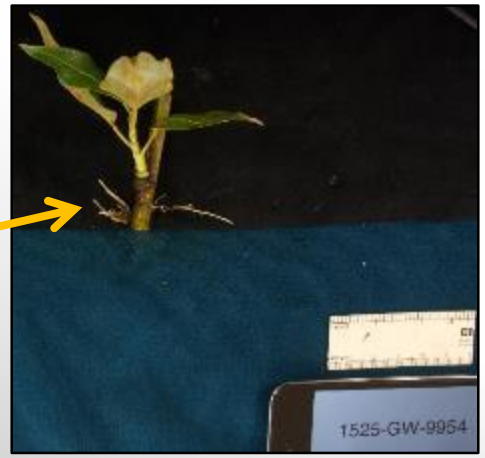
In vivo GWAS study: Rooting (completed)

- Goal is to discover genes associated to rooting *in vivo* condition
- Methods and materials
  - Dormant cutting was cut with one bud and planted in 50ml Flacon tube water without any hormone application in head house
  - 200 genotypes each phase; Total 8 phases including repetition for some genotypes
  - Light is provided by fluorescent tubes with a 16-h photoperiod and temperature is 22-25°C
  - Tubes were filled with fresh tap water every 3-4 days
  - Data and imaging collected each week for five weeks

1,224 genotypes have been used for *in vivo* GWAS rooting

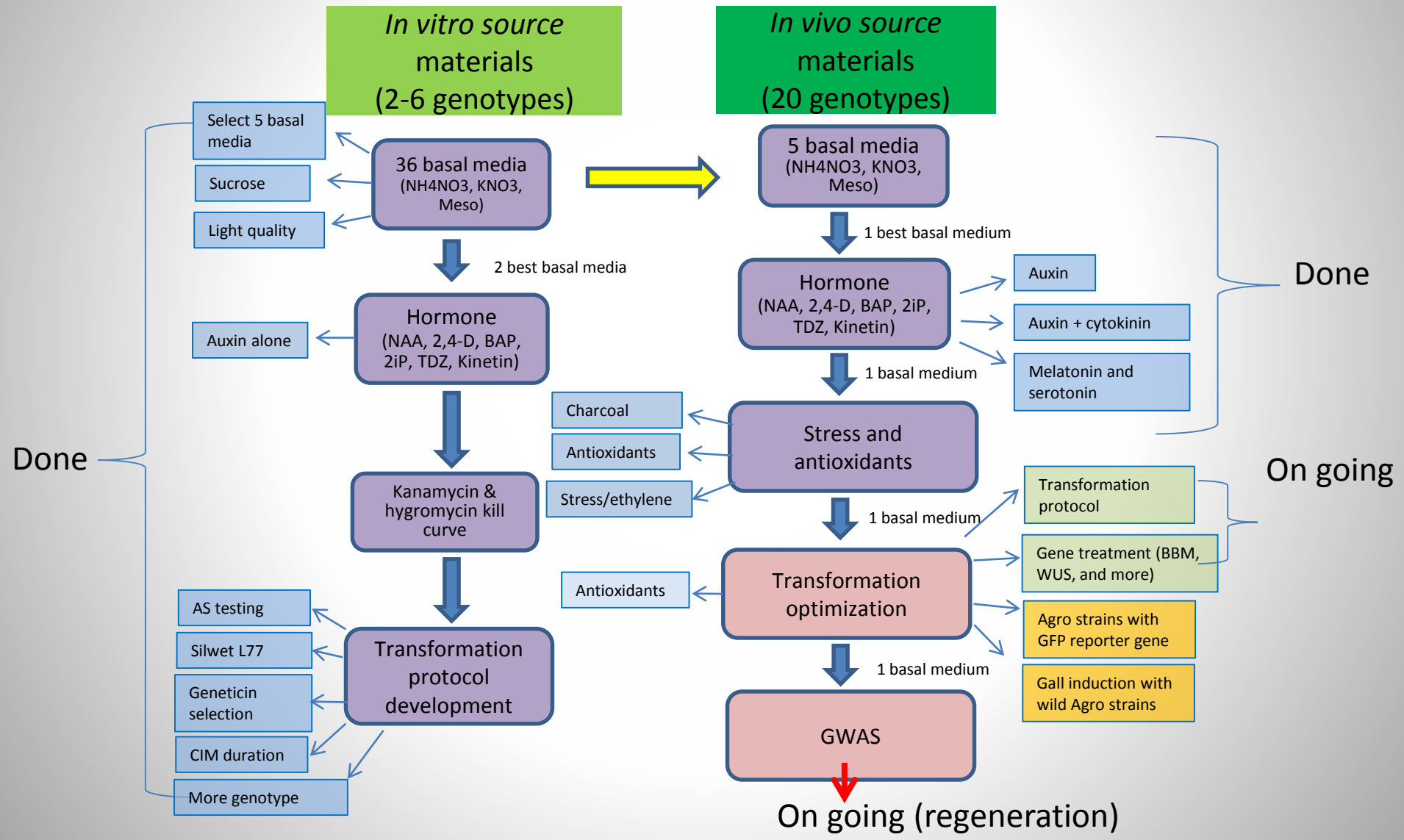
GWAS Phase	# Genotypes studied	Date of cuttings collected	Start dates	Dates in soil
1	195	1/16-17/18	11/13/2018	12/15/2018
2	190	1/16-17/18	12/17/2018	1/28/2019
3	174	1/16-17/18	2/21/2019	3/22/19
4	200	1/16-17/18	3/21/2018	5/4/2018
5	193	1/16-17/18	5/11/2018	6/22/2018
6	198	1/16-17/18	7/27/2018	9/7/2018
7	57 (115)	1/16-17/18	9/28/2018	11/6/2018
8	17 (183)	2/6-7/19	3/19/2019	4/22/19
Total	**1,224 (1,522)**			

Adventitious stem roots



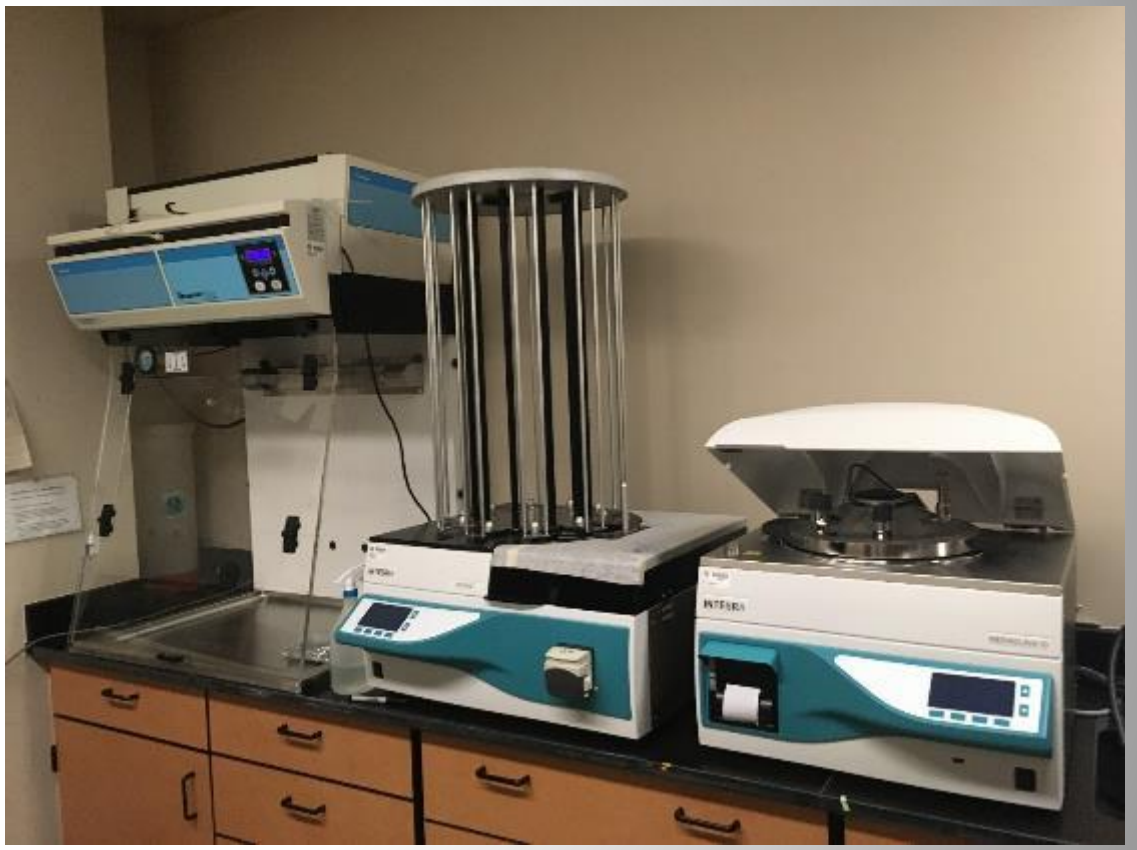
Adventitious basal roots

Goal is to determine optimal treatment and condition to be applied to GWAS to maximize genetic variance and minimize environmental variance



In vitro study equipment purchased

Percival Scientific Growth Chambers:
In order to reduce condensation we used plastic boxes to hold unsealed Petri dishes



Integra mediajet and mediaclave

Twelve optimization experiments including 163 treatments have been completed

Experiment ID	Title of experiment	Goal of experiment	# of treatments	# genotypes tested
E1	Direct vs. indirect regeneration	Study the effect of two regeneration pathways on organogenesis	2	20
E2	36 basal media screen (in vitro)	Screen big range of macro and micro nutrients and study their effect on regeneration	36	2
E3	5 basal media test (in vitro)	Select the best basal medium that gives 50% of genotypes in regeneration	5	5
E3	5 basal media test (in vivo)	Select the best basal medium that gives 50% of genotypes in regeneration	5	20
E4	Varying [sucrose] (in vitro)	Test whether cottonwood grow better on low level of sucrose	8	4
E5	Various auxins (in vitro)	Test what types of auxin at which level is good for regeneration	18	6
E5	Various auxins (in vivo)	Test what types of auxin at which level is good for regeneration	9	20

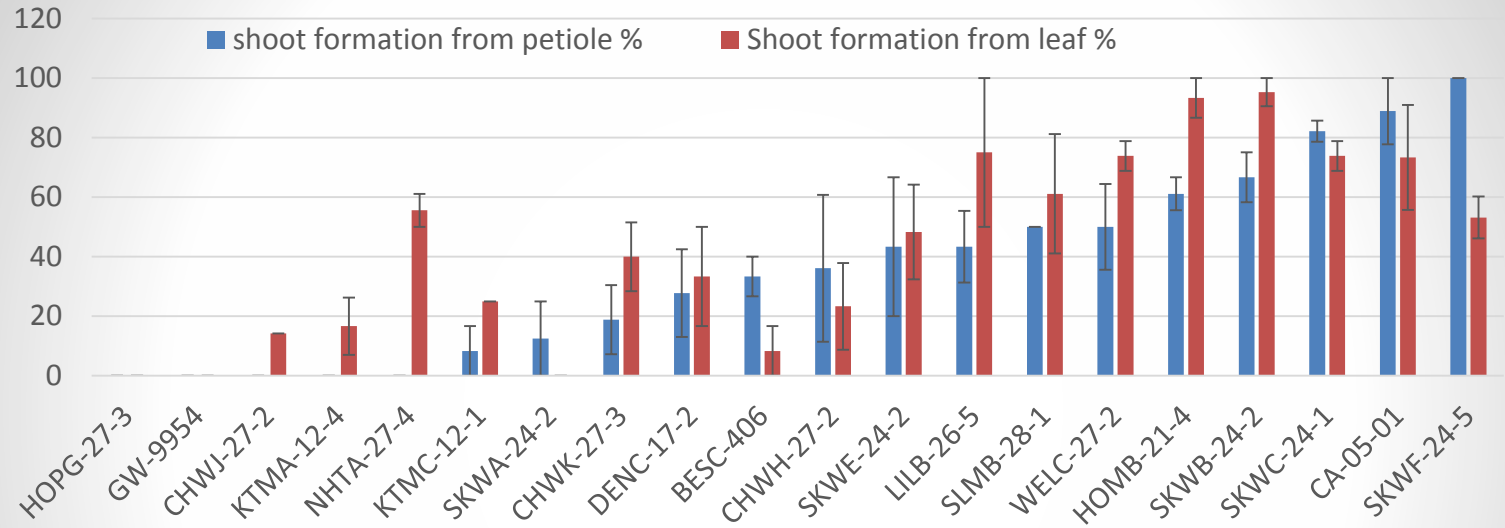
In vitro experiments: Continued

Experiment ID	Title of experiment	Goal of experiment	# of treatments	# genotypes tested
E6	Various auxin, cytokinin combos (in vivo)	Study the effect of various auxin and cytokinin combination on organogenesis	36	18
E7	Melatonin, serotonin effects (in vivo)	Study the effect of newly discovered plant hormone melatonin and serotonin on organogenesis	8	19
E8	PPM (Plant Preservative Mixture) and benomyl effect (in vivo)	Test the efficacy of PPM and benomyl in controlling contamination	7	8
E9	LA (Lipoic Acid) effect (in vivo)	Study the effect of antioxidants LA on organogenesis	7	16
E10	AC (Activated Charcoal and VC (Vitamine C) (Ascorbic Acid) effect (in vivo)	Study the effect of AC and antioxidant VC on organogenesis	7	16
E11	AgNO3 effect (in vivo	Study the effect of ethylene inhibitor AgNO3 on organogenesis	7	16
E12	Light spectrum & intensity effect (in vivo)	Study the effect of different light spectrum and intensity on callus and shoot formation	8	4

Data and image collection and some issue *in vitro* regeneration optimization

- Proportion of callus, shoot and root, average size of callus and mean number of shoots and roots
  - Longest shoots were assessed after 9 weeks (6 weeks on SIM)
- RGB images taken at 0 and 3 weeks on CIM and 2, 4, and 6 weeks on SIM
- We used 3 explant types (leaf, stem, and petiole) for the study
- We found the leaf and petiole often necrotic after sterilization
- Therefore we have only used stem explants for *in vitro* GWAS study

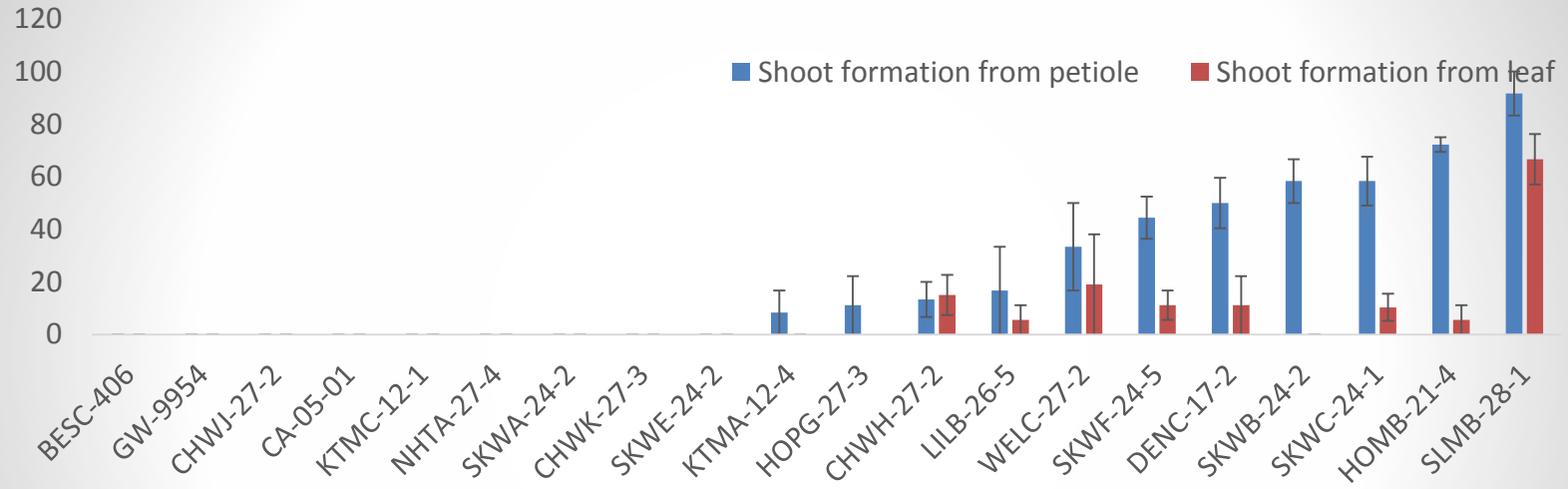
Indirect regeneration system gave 90% response among genotypes tested (CIM then SIM)



CA-05-01 petiole explant

LILB 26-5 petiole explant

55 % genotypes formed shoots through direct regeneration (direct to SIM)



SLMB 28-1 petiole explant

SLMB 28-1 leaf explant

We selected auxin and cytokinin types and concentrations for the testing based on literature in *Populus trichocarpa*

Treatment	NAA (mg/L)	BAP (mg/L)
1	0.5	0.5
2	1	0.5
3	2	0.5
4	0.5	1
5	1	1
6	2	1

Treatment	NAA (mg/L)	2iP (mg/L)
1	0.5	0.5
2	1	0.5
3	2	0.5
4	0.5	1
5	1	1
6	2	1

Treatment	NAA (mg/L)	Kinetin (mg/L)
1	0.5	0.5
2	1	0.5
3	2	0.5
4	0.5	1
5	1	1
6	2	1

Treatment	2,4-D (mg/L)	BAP (mg/L)
1	0.01	0.5
2	0.05	0.5
3	1	0.5
4	0.01	1
5	0.05	1
6	0.1	1

Treatment	2,4-D (mg/L)	2iP (mg/L)
1	0.01	0.5
2	0.05	0.5
3	1	0.5
4	0.01	1
5	0.05	1
6	0.1	1

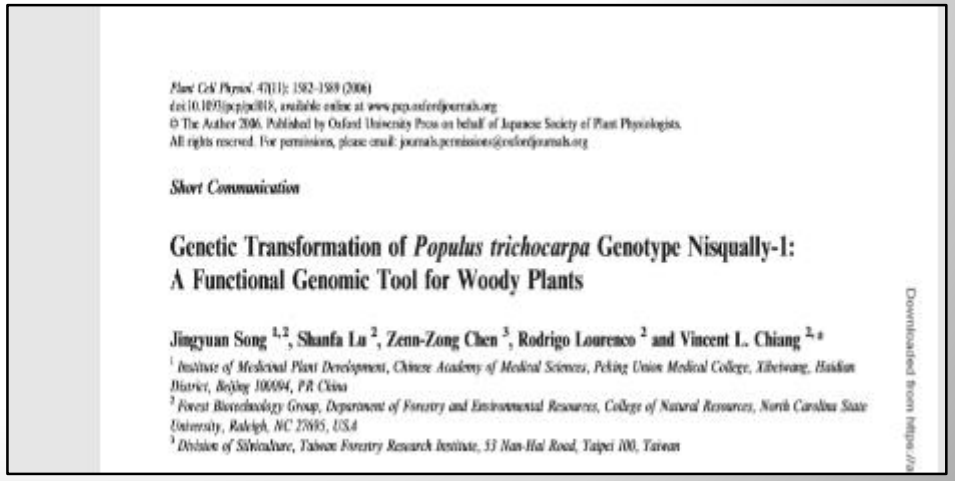
Treatment	2,4-D (mg/L)	Kinetin (mg/L)
1	0.01	0.5
2	0.05	0.5
3	1	0.5
4	0.01	1
5	0.05	1
6	0.1	1

Plant Molecular Biology Reporter **22:** 1–9, September 2004
© 2004 *International Society for Plant Molecular Biology. Printed in Canada.*

Publish by Abstract

***Agrobacterium*-Mediated Transformation of the Genome-Sequenced Poplar Clone, Nisqually-1 (*Populus trichocarpa*)**

CAIPING MA, STEVEN H. STRAUSS, and RICHARD MEILAN*
Department of Forest Science, Oregon State University, Corvallis, OR 97331-5752



Plant Cell Physiol. 47(11): 1582–1589 (2006)
doi:10.1093/pcp/pcl018, available online at www.pcp.oxfordjournals.org
© The Author 2006. Published by Oxford University Press on behalf of Japanese Society of Plant Physiologists.
All rights reserved. For permissions, please email: journals.permissions@oxfordjournals.org

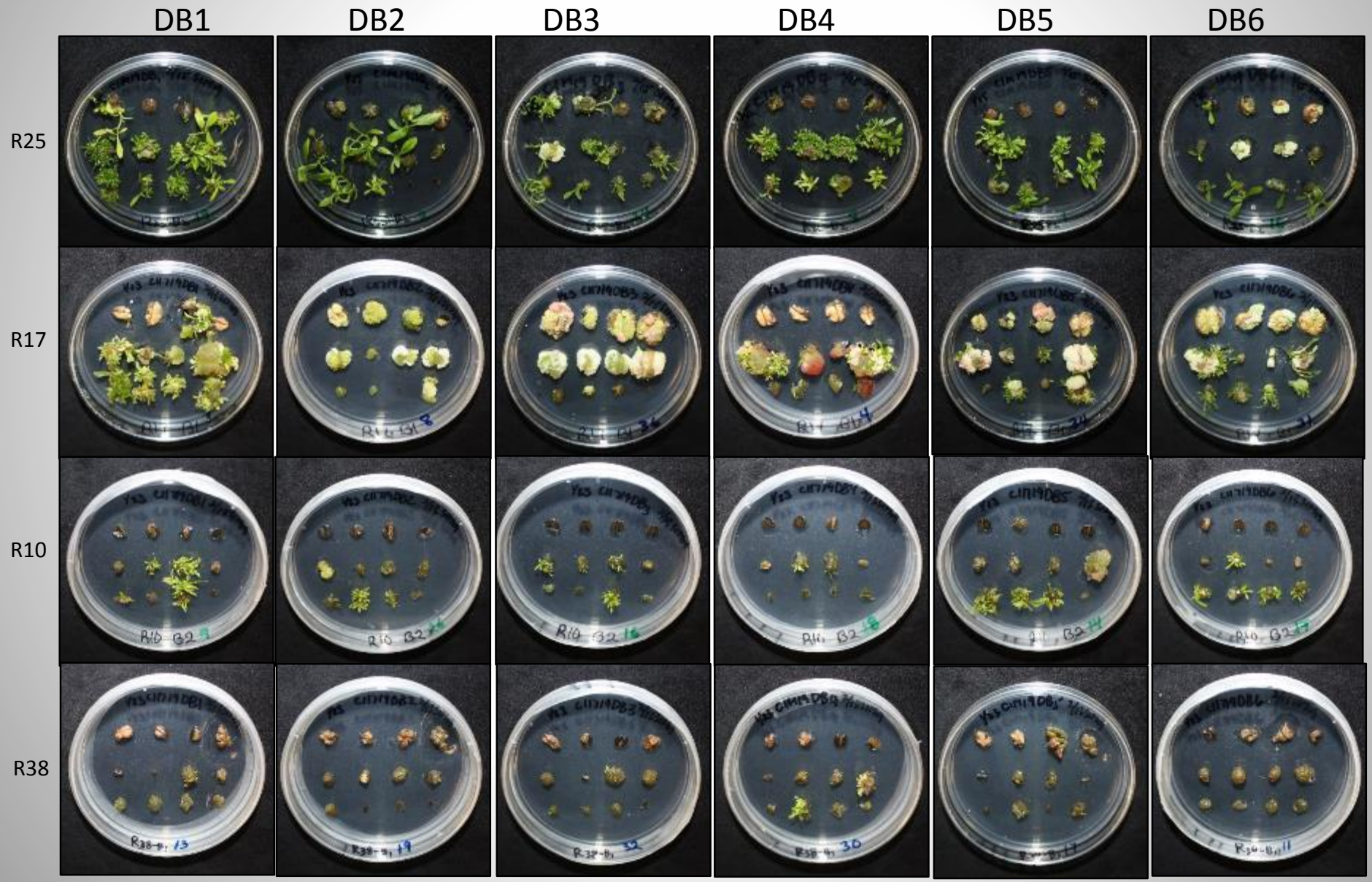
Short Communication

Genetic Transformation of *Populus trichocarpa* Genotype Nisqually-1: A Functional Genomic Tool for Woody Plants

Jingyuan Song [1,2], Shanfa Lu [2], Zeno-Zong Chen [3], Rodrigo Lourenco [2] and Vincent L. Chiang [2,*]
[1] *Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Xiheiwang, Haidian District, Beijing 100094, PR China*
[2] *Forest Biotechnology Group, Department of Forestry and Environmental Resources, College of Natural Resources, North Carolina State University, Raleigh, NC 27695, USA*
[3] *Division of Silviculture, Taiwan Forestry Research Institute, 53 Nan-Hai Road, Taipei 100, Taiwan*

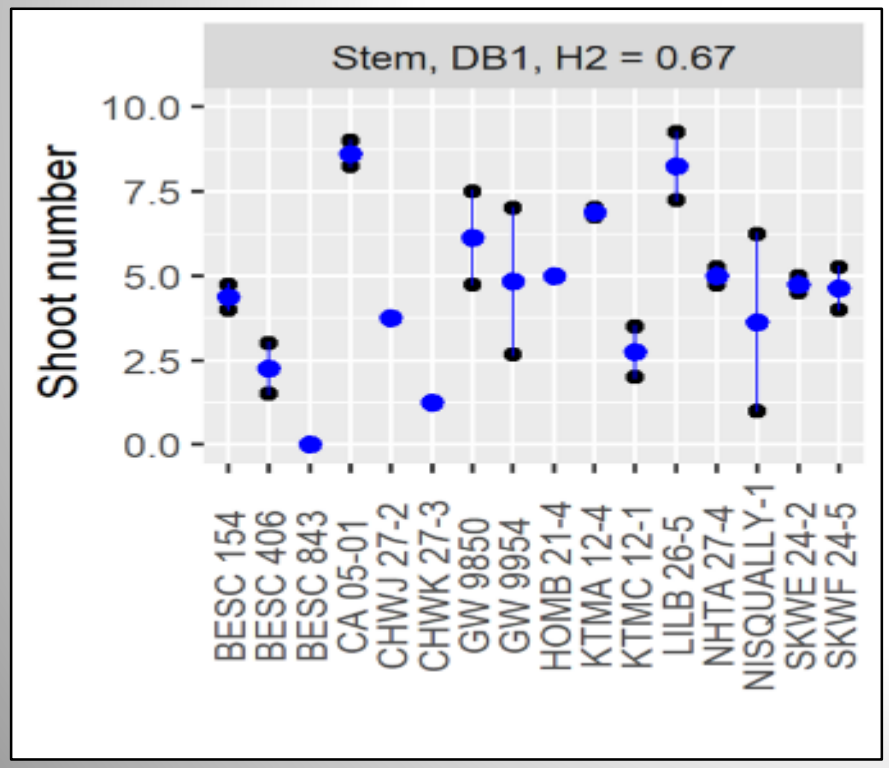
Downloaded from https://a

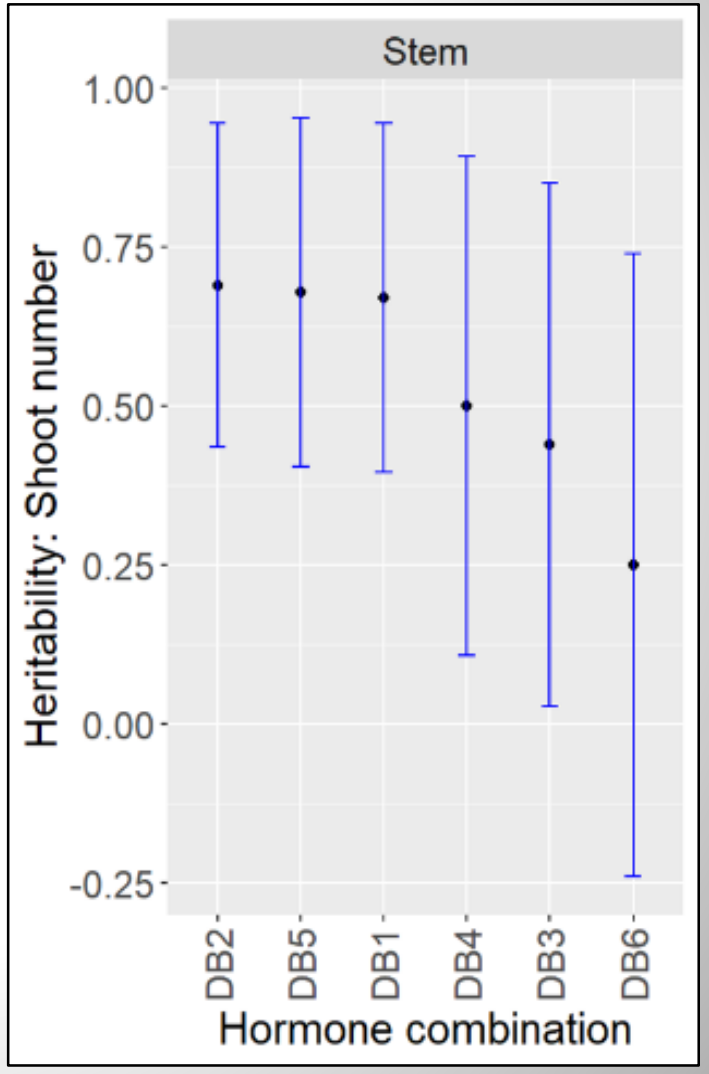
DB hormone combinations showed high genotype and low explant variance



DB1 hormones gave high H^2 and much genotype variance

Treatment	DB1	DB2	DB3	DB4	DB5	DB6
2.4-D (mg/L)	0.01	0.05	0.1	0.01	0.05	0.1
BAP (mg/L)	0.5	0.5	0.5	1	1	1





Transformation optimization

- Goal is to test various factors that influence transformation rate
- Methods and materials
  - *In vivo* stem explants (2-3mm)
  - Agrobacterium strain AGL1 containing 2X35S::eGFP or DS-Red
  - 1% Tween 5 min, 70% ethanol 5 min, 20% bleach 20 min (10 min in vacuum), and 4 washes
  - 2-20 genotypes, 20 explants/plate, 3-4 plates/treatment/genotype
- Data and images taken
  - RGB and hyperspectral imaging after 3 and 7 weeks
  - Manual score callus and shoot formation, GFP callus and shoot production at week 3 and 7

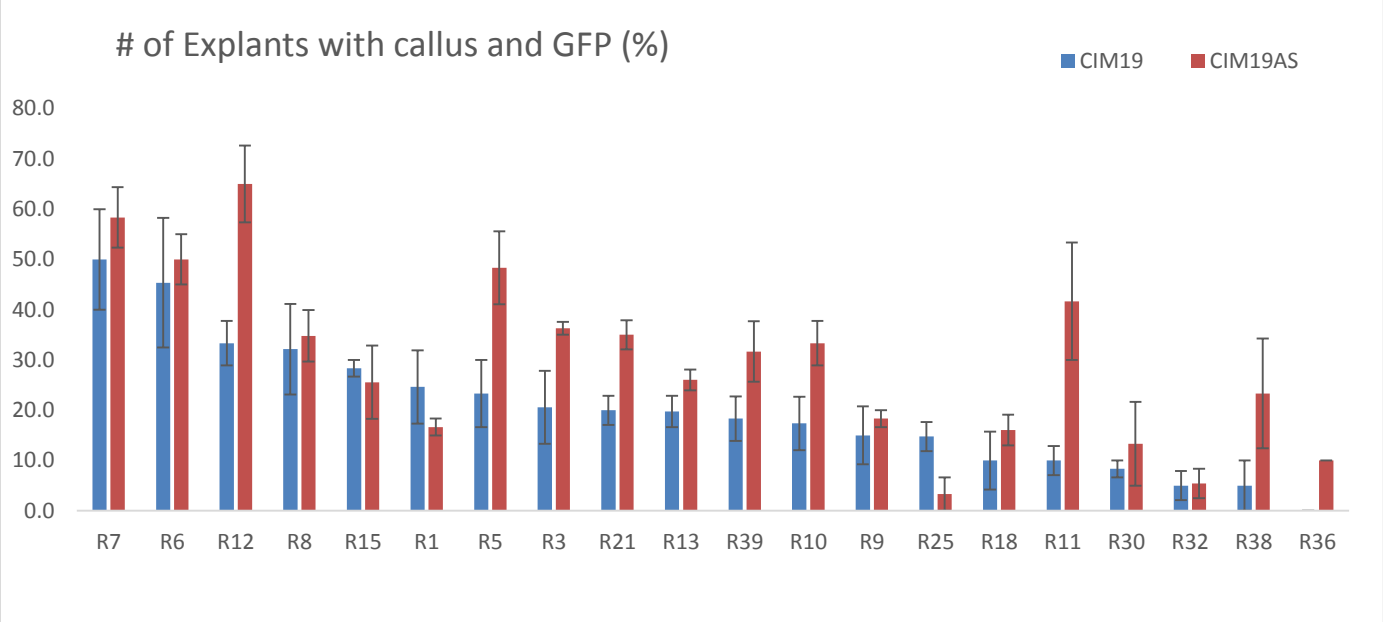
14 transformation optimization experiments including 54 treatments done or nearly so

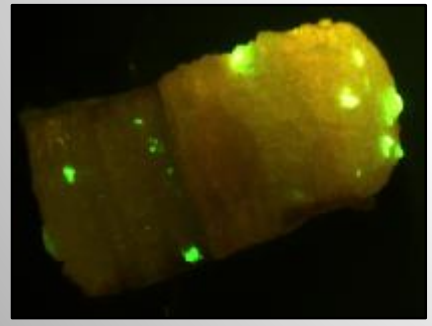
Experiment ID	Title of experiment	Goal of experiment	# of treatments	# genotypes tested
T1	AS vs. no AS	Test whether acetosyringone (AS) enhances Agrobacterium-mediated transformation in cottonwood	2	20
T2	Antioxidants (Lipoic acid)	Test if lipoic acid can reduce explants browning and increase transformation rate	4	20
T3	Pre-culture vs. no pre-culture	Test if preculturing on CIM affects up-taking of transgenes, thus affects transformation rate	2	20
T4	Virulence gene	Investigate whether providing Agrobacterium with a plasmid containing virulence gene augments the efficiency of transfer of the T-DNA (transferred DNA)	3	4
T5	Spectinomycin kill curve with no Agro infection	Select which concentration can be used in transformation	6	2
T6	Spectinomycin kill curve with Agro infection	Select optimal concentration for transgenic callus and shoot selection	6	2
T7	Different concentrations Agro	Determine which concentration is effective for gene delivery, but not damage the explants (browning)	4	4

Transformation experiments continued

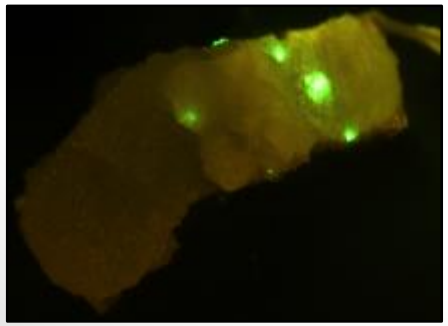
Experiment ID	Title of experiment	Goal of experiment	# of treatments	# genotypes tested
T8	Different levels of Silwet L77	Determine if/which level of this surfactant helps to Improve transformation frequency	4	4
T9	Different duration of vacuum during Agro inoculation	Determine if vacuum during agro infiltration helps to Improve transformation efficiency	4	5
T10	Different levels of Break Thru-233	Determine if/which level of this surfactant helps to Improve transformation rate	5	5
T11	Duration on CIM after washing	Test whether duration on CIM after washing affect shoot regeneration and transformation	3	20
T12	Different CIM s	Test which CIM will affect shoot regeneration and transformation rates and give highest heritability	4	20
T13	Agro strains	Test different Agro strains efficiency in gene transformation	3	20
T14	Wild Agrobacterium gall induction	Investigate what/how cottonwood genotypes are susceptible to wild Agro, thus they are easy for *Agrobacterium*-mediated transformation	4	20

Acetosyringone (AS) enhanced GFP callus formation

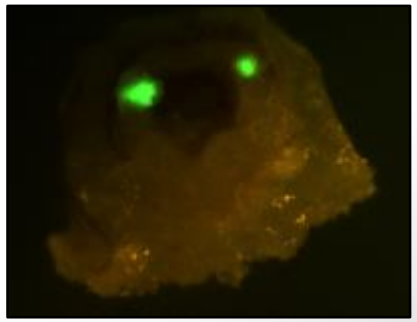




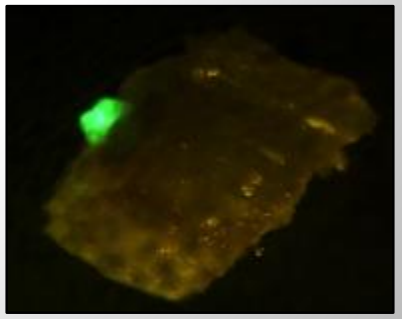
R7 AS



R21 AS

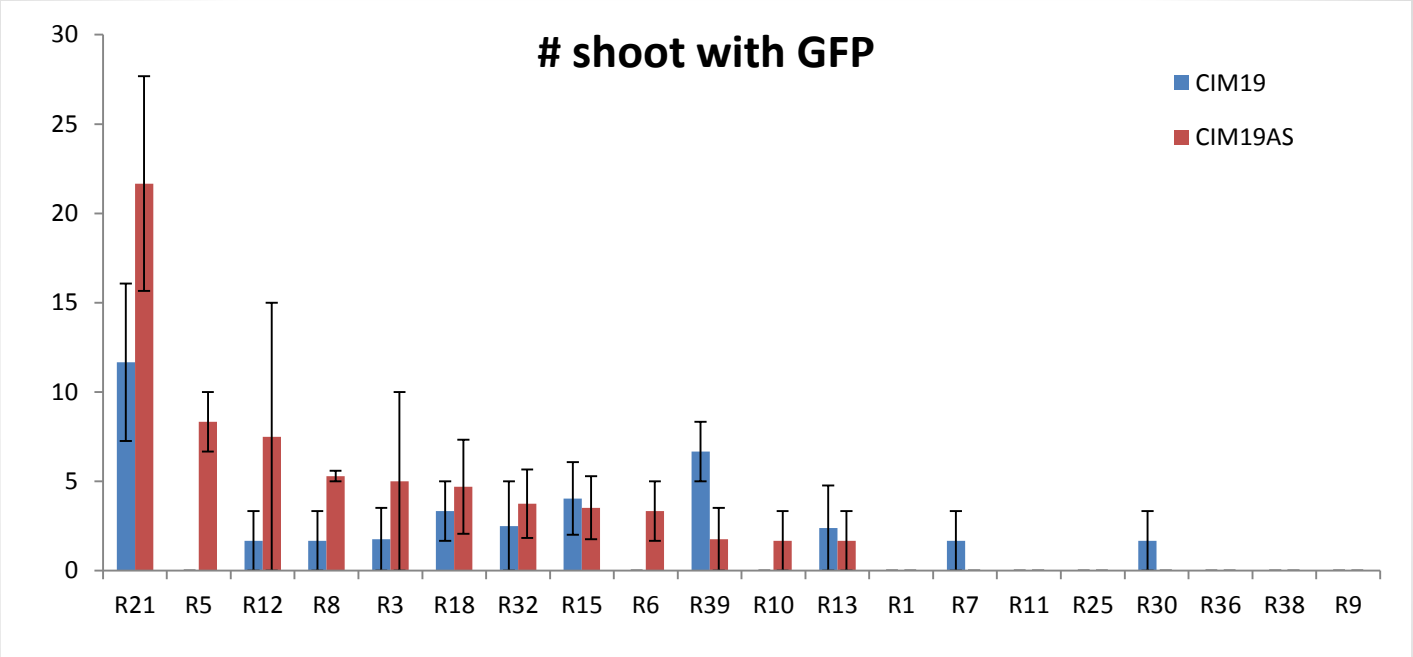


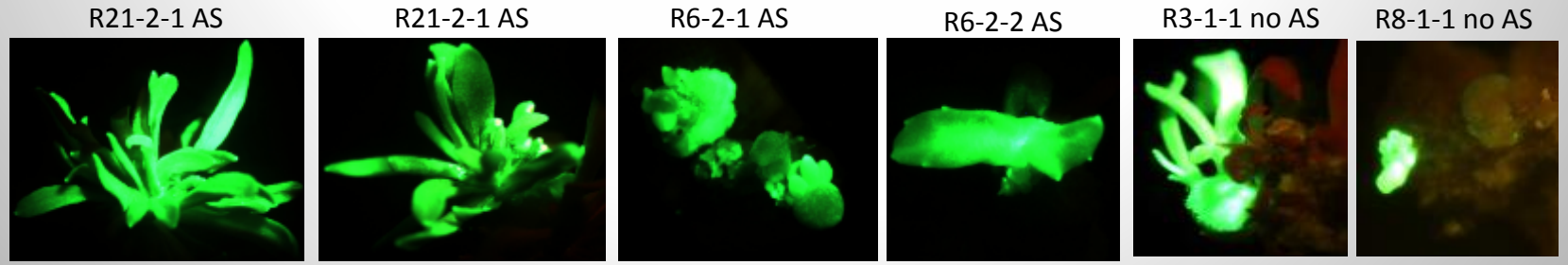
R5 AS



R12 AS

AS also enhanced transgenic shoot formation

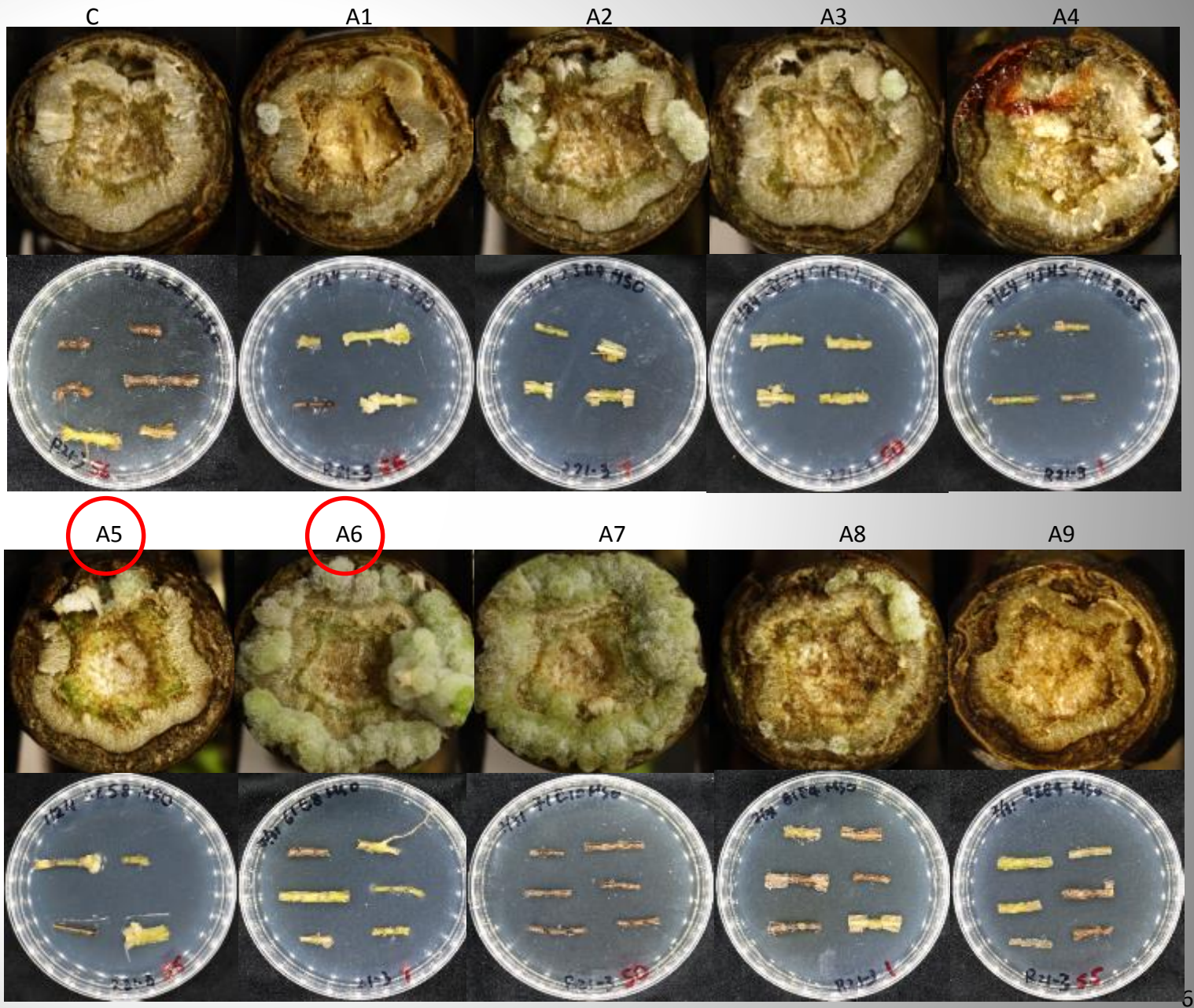




Crown gall induction under *in vivo* and *in vitro* conditions

- 9 wild *Agrobacterium* strains were provided by Dr. Jeff Chang in Department of Botany & Plant Pathology at OSU
- These wild Agro strains could naturally induce crown gall (tumor) on many plant species
- Our goal is to investigate what/how cottonwood genotypes are susceptible to wild Agro, thus learning if they are suitable for *Agrobacterium*-mediated transformation
- If sufficient genetic variance seen, may employ for GWAS
- Studies ongoing to test H^2 and best inoculation methods
  - Two candidate strains preliminarily identified in small scale *in vivo* and *in vitro* studies

Callus formation from genotype SKWB24-2 inoculated with 8 wild agro strains and control (water) under *in vivo* vs *in vitro* conditions



In vitro GWAS for callus and shoot regeneration

- Goal is to use the optimal treatment and condition from optimization experiments for over 1,000 genotypes
- Best conditions:
  - Stem explants (4 mm)
  - Two treatments:
    - Direct : 7 weeks in SIM under light
    - Indirect: 3 weeks CIM in dark and 4weeks SIM in light
  - MS medium containing lower salts and 2.5% sucrose
  - 2,4-D 0.01mg/L and BAP 0.5mg/L in CIM and 0.13mg/L TDZ in SIM
  - 12 explants/plate
  - Two plates/treatment/genotype
- Data and images taken:
  - RGB and hyperspectral imaging after 3 and 7 weeks
  - Manual score contamination at week 3 and 7

Sterilization methods used and tested

	Spray with 70% ethanol	1% Tween	Green cure fungicide	70% ethanol	Bleach With vacuum	Bleach at shaking	SuperShock With vacuum	hydrogen peroxide	Sterile H20
1.	No	5 min	No	5 min	5 min, 15%,	10 min, 15%	No	No	4 times
2.	No	5 min	No	5 min	10 min, 15%	5 min, 15%	No	No	4 times
3.	No	5 min	No	5 min	10 min, 15%	10 min, 15%	No	No	4 times
4.	No	5 min	No	5 min	10 min, 20%	10 min, 20%	No	No	4 times
5.	Yes	5 min	No	5 min	No	No	Yes, 10 min; Then 10 min shaking	No	4 times
6.	Yes	5 min	No	5 min	No	No	Yes, 10 min; Then 10 min shaking	Yes	4 times
7	Yes	No	5 min	5 min	10 min, 20%	10 min, 20%	No	No	4 times
8	Bleach white 1 min	No	No	70% Isopropyl Alcohol	No	No	Yes, 10 min; Then 10 min shaking	Yes	No
9	Bleach white 1 min	No	No	70% Isopropyl Alcohol	No	No	Yes, 10 min; Then 10 min shaking	Yes	Once
10	No	5 min	No	5 min	10 min, 20%	10 min, 20%	No	No	4 times

Methods 1-9: cut stem in sterile water; Method 10: cut stem in no water to reduce bacteria spread

1,169 (including duplicate and repeat) genotypes have been tested for *in vitro* GWAS regeneration

Phase	Date of experiment	# of genotypes	# of plates	Sterilization method
A	1/30 - 2/1/19	56	224	1
B	2/13 - 2/15/19	57	228	2
C	2/27 - 3/1/19	56	224	2
D	3/6 - 3/8/19	56	224	2
E	3/13 - 3/15/19	56	224	2
F	4/10-4/12/19	56	224	3
G	4/17-4/19/19	60	240	4
H	4/24-4/26/19	60	240	4
I	5/1-5/3/19	60	240	4
J	5/8-5/10/19	60	240	4
K	5/15-5/17/19	60	240	4
L	5/22-5/24/19	60	240	4
M	6/5-6/7/19	60	240	4
N	6/19-6/21/19	60	240	4
O	6/26-6/28/19	60	240	4
P	7/2/2019	24	96	4, 5
Q	7/10-7/12/19	60	240	4, 6, 7
R	7/26/2019	24	96	4, 8
S	8/9/2019	24	96	4,9
T	8/30/2019	20	80	4
U	9/4 and 9/6/2019	40	160	4.10
V (repeat)	9/11-9/13/2019	60	240	10
W (repeat)	9/26-9/26/19	40	160	10
		1169	**4676**	

Over 70% genotypes have 1-2 plates with zero or less than half explants (6) contaminated

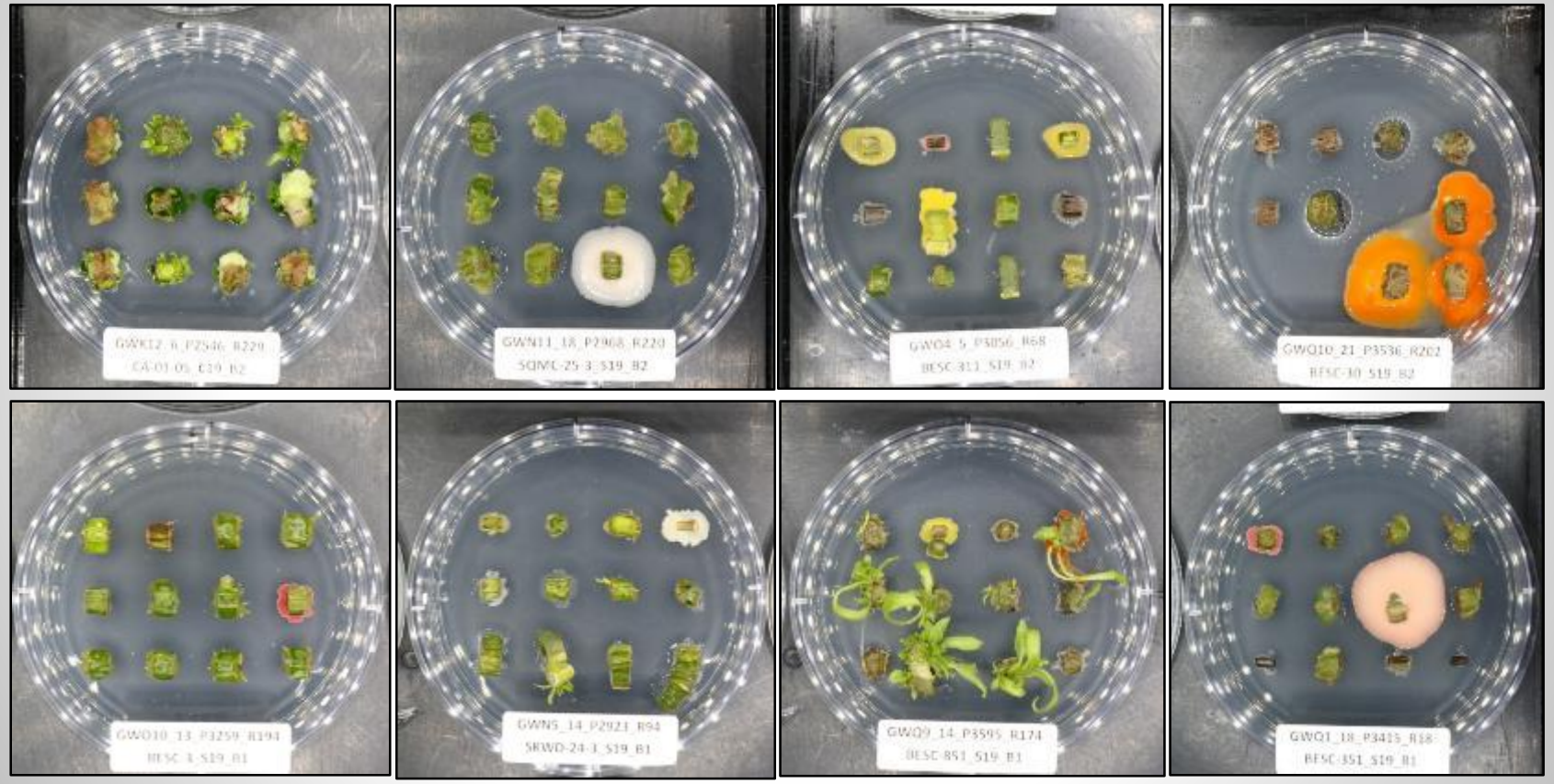
Indirect (CIM-SIM)

Phase	2 plates	1 plate	Total # genotypes	Total # genotypes cultured
A	33	18	51	56
B	47	9	56	57
C	45	9	54	56
D	26	21	47	56
E	31	14	45	56
F	18	18	36	55
G	27	15	42	60
H	24	17	41	60
I	30	19	49	60
J	15	22	37	60
K	29	19	48	60
L	25	27	52	60
M	17	22	39	60
N	12	22	34	60
O	14	28	42	60
P	12	7	19	24
Q	16	20	36	60
Total	**421**	**307**	**728**	**960**
%	**43.9**	**32.0**	**75.8**	

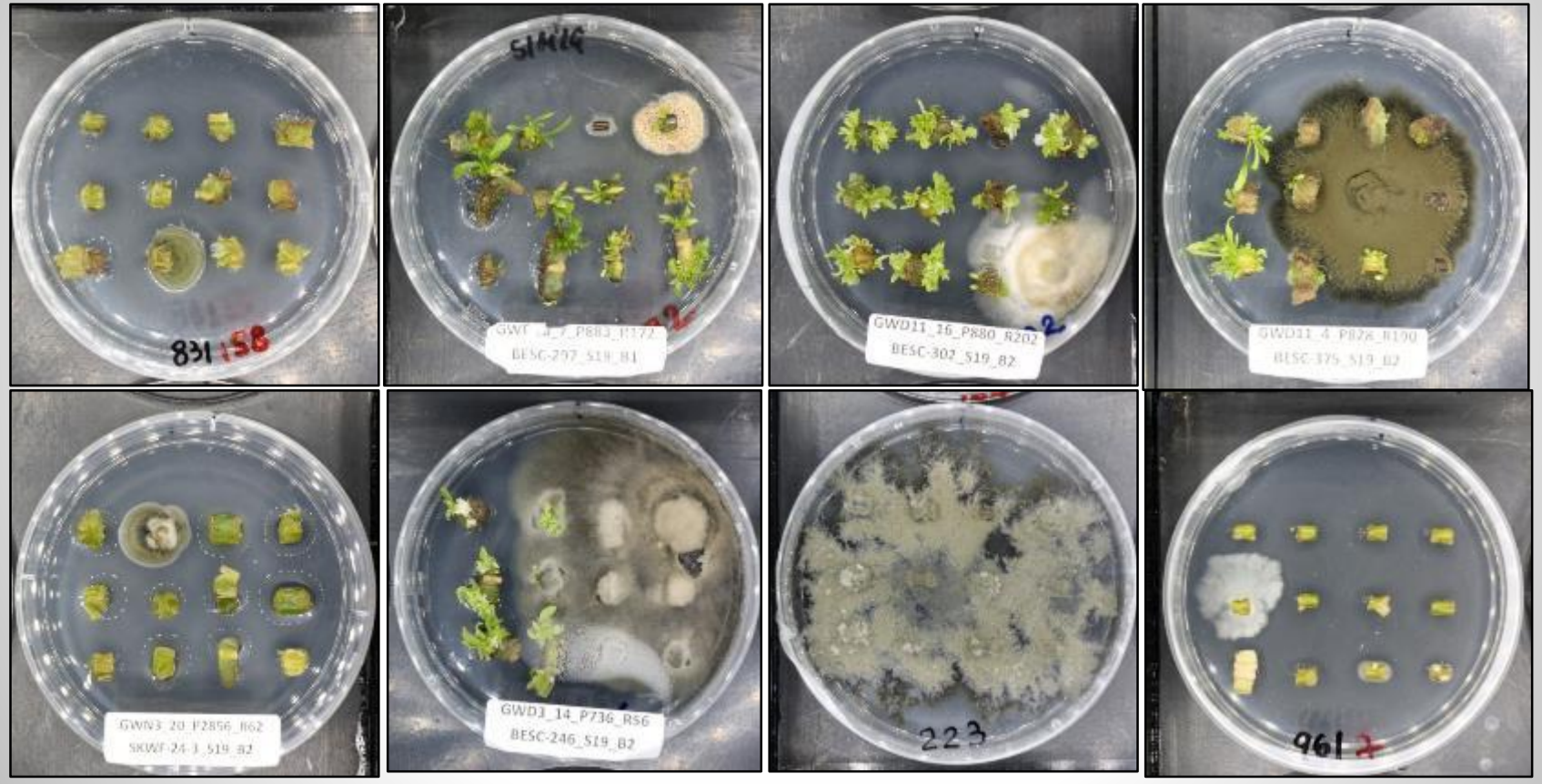
Direct (SIM)

Phase	2 plates	1 plate	Total # genotypes	Total # genotypes cultured
A	20	20	40	56
B	47	10	57	57
C	36	14	50	56
D	28	20	48	56
E	30	17	47	56
F	22	15	37	55
G	30	13	43	60
H	43	8	51	60
I	46	11	57	60
J	17	21	38	60
K	25	21	46	60
L	20	24	44	60
M	17	18	35	60
N	12	14	26	60
O	14	12	26	60
P	9	10	19	24
Q	18	20	38	60
Total	**434**	**268**	**702**	**960**
%	**45.2**	**27.9**	**73.1**	

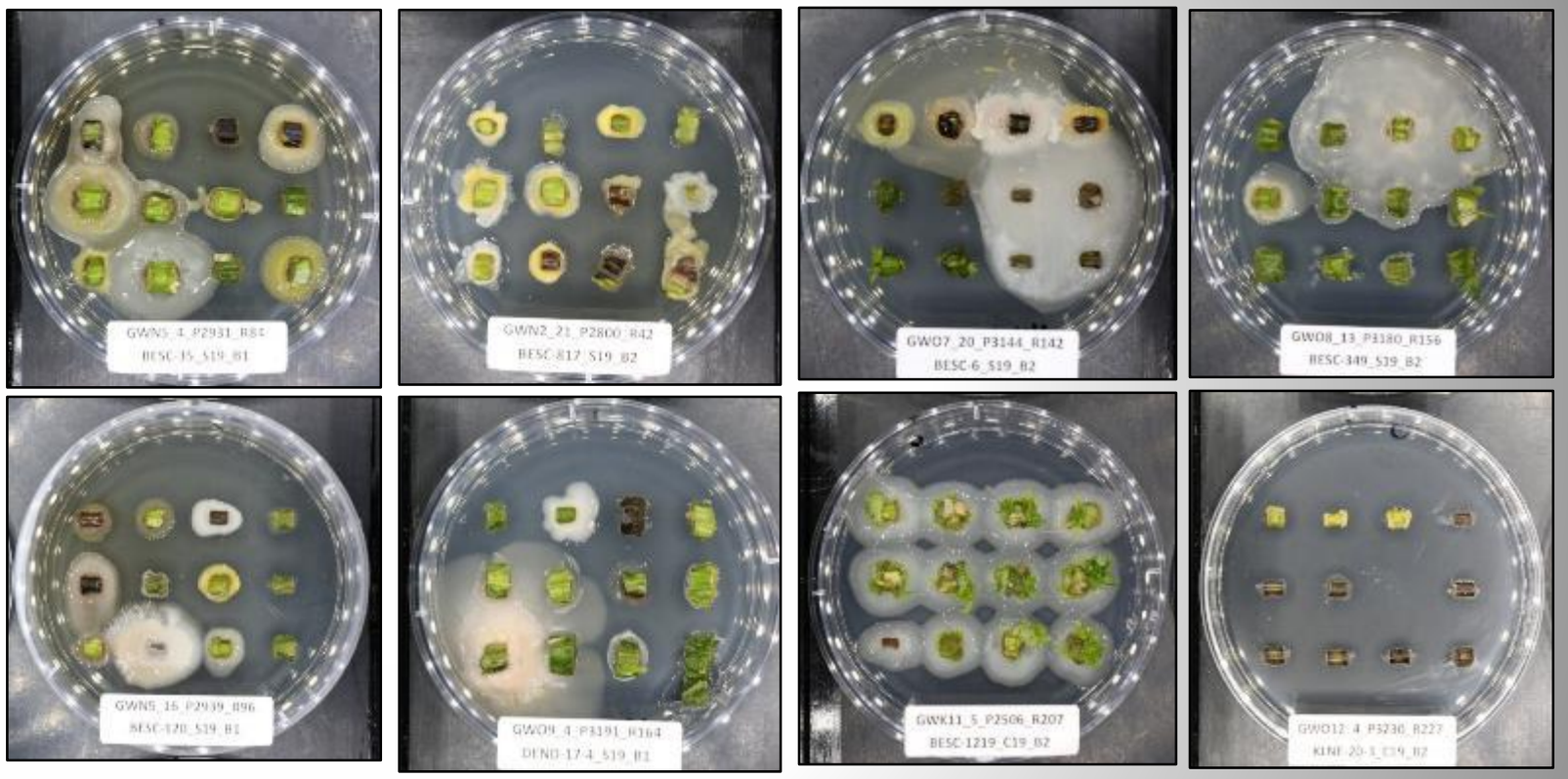
Examples of bacterial contamination: Wide range of types and size



Examples of fungus contamination

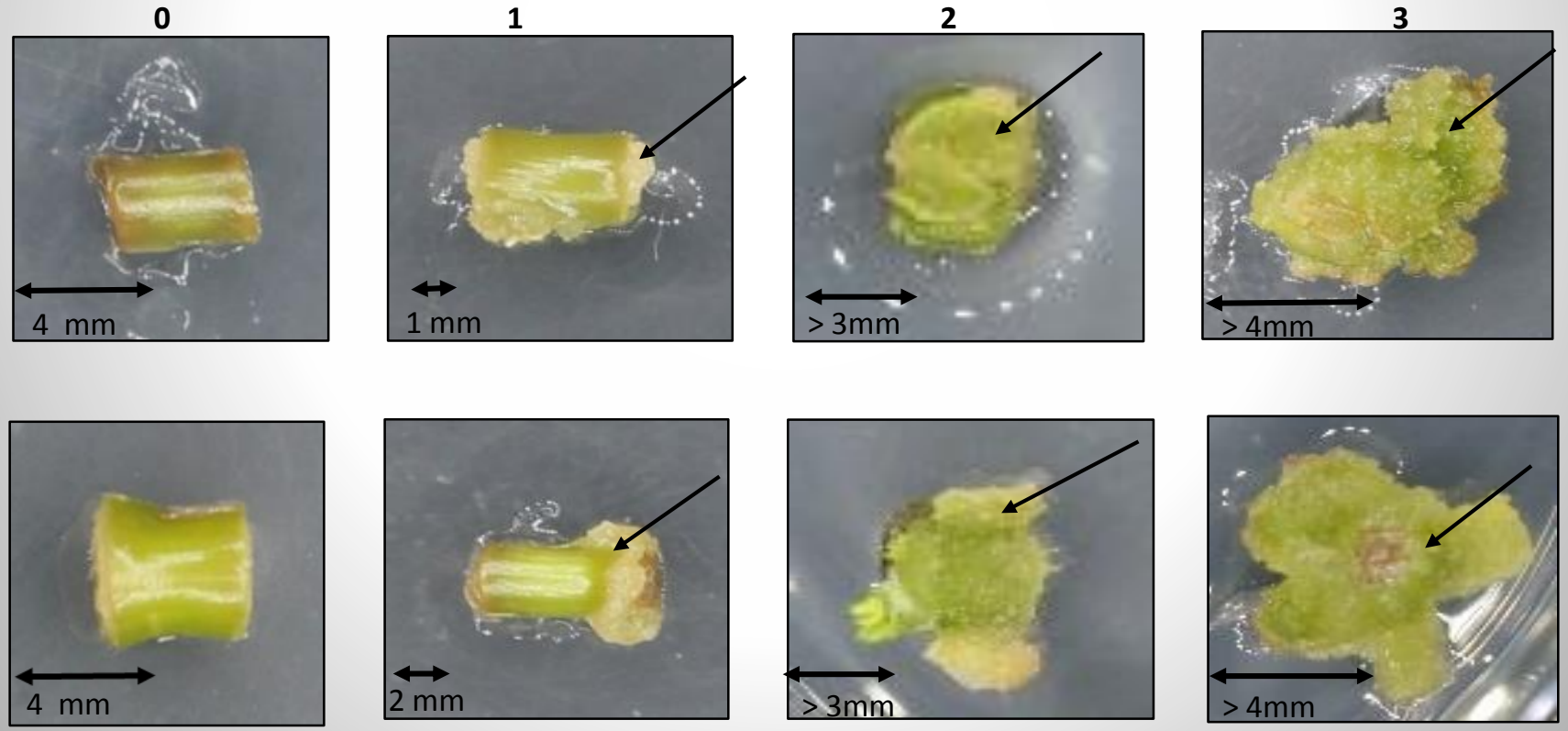


Plates that have more than 6 contaminated explants will be excluded from analysis, and all individual explants scored for presence and extent



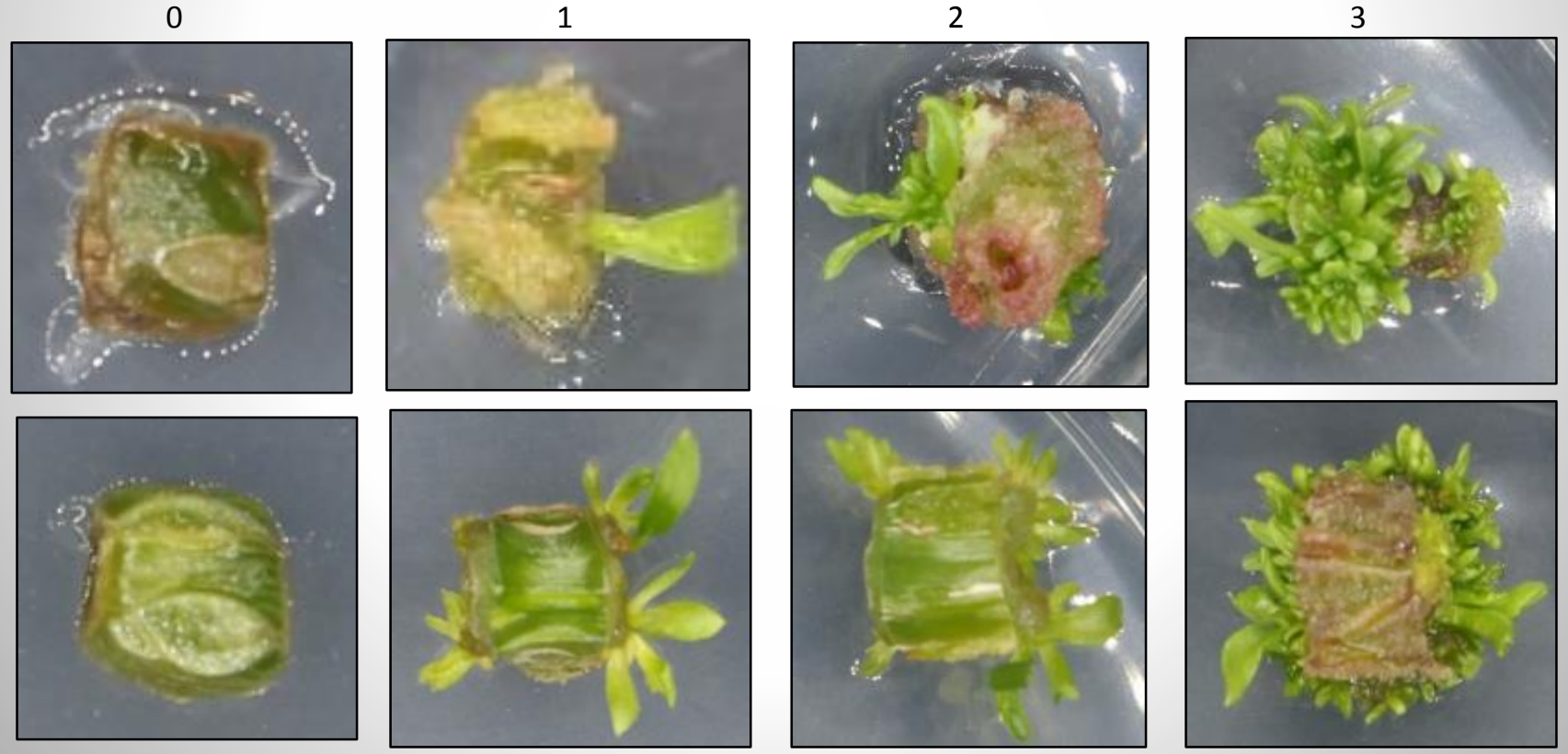
Scoring system for callus formation at week 3

Scale	Definition	% Area
0	No callus	0
1	Less than 50% of the original explant area	<50%
2	Greater than 50% of the original explant area	>50%
3	Greater than or equal to the original explant area	≥100%



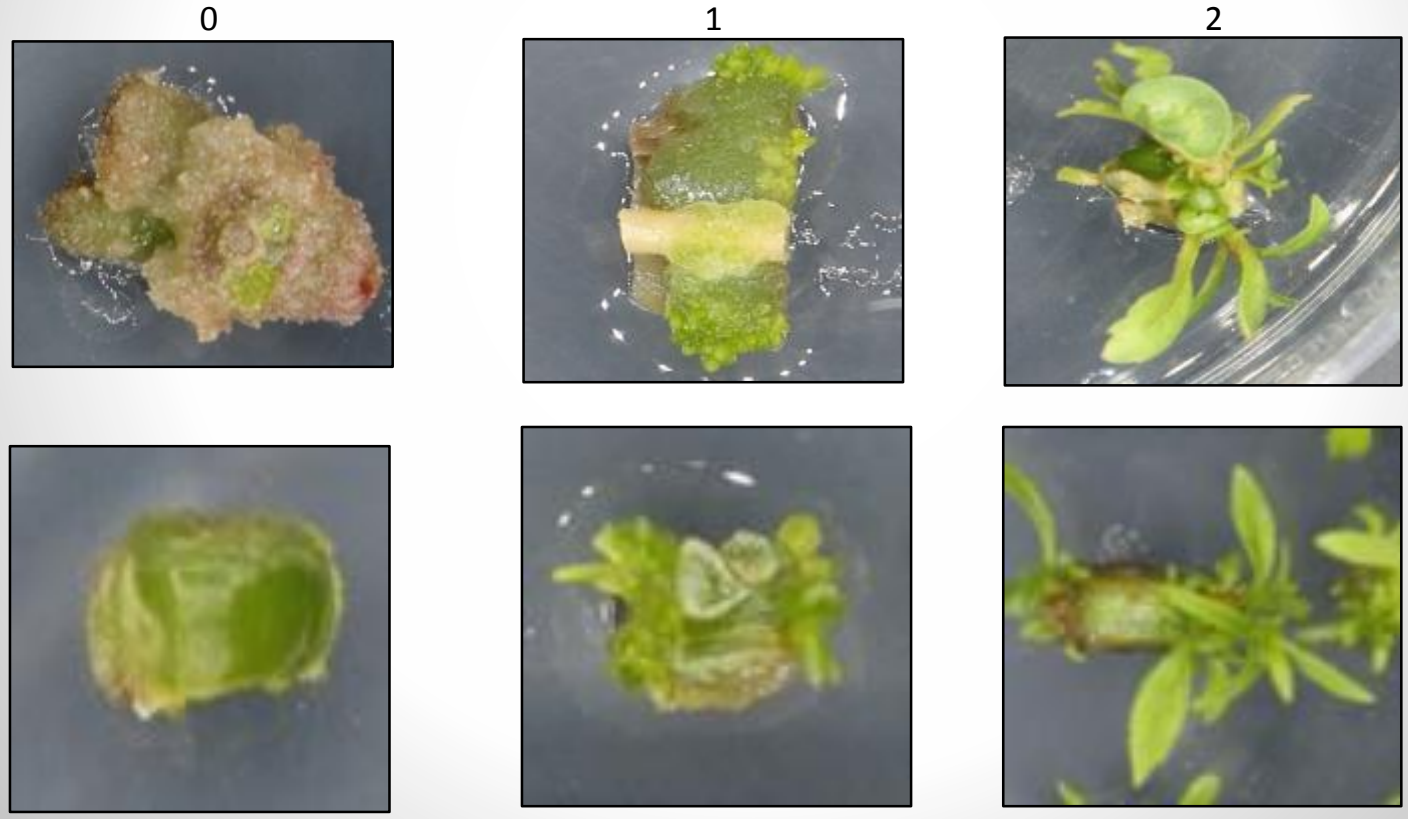
Scoring system for shoot formation at week 7

Scale	Definition
0	No shoots
1	1-3 shoots
2	4-10 shoots
3	>10 shoots



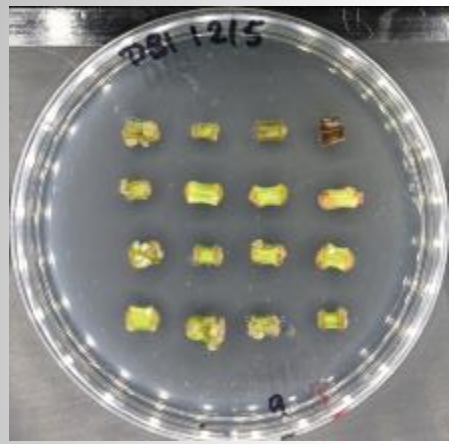
Scoring system for shoot length at week 7

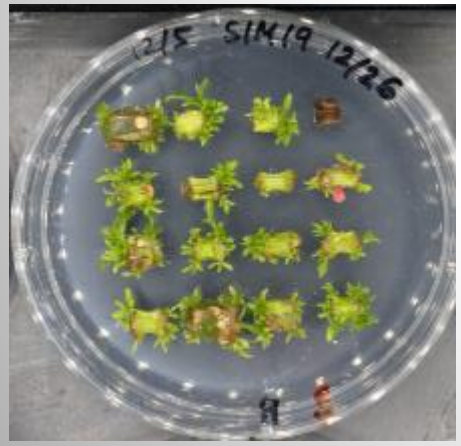
Scale	Definition
0	0 (no shoot)
1	1-5 mm
2	>5 mm



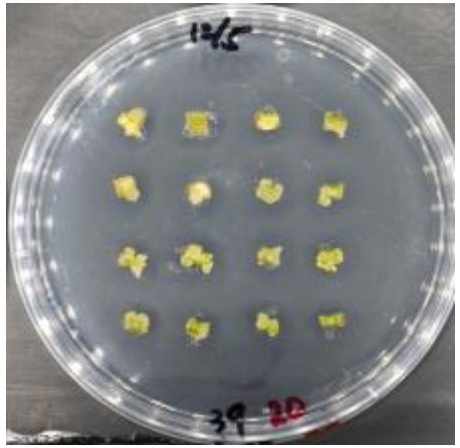
Good response genotypes

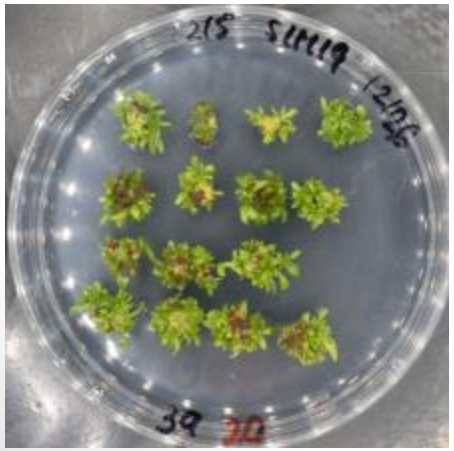
CA-05-01



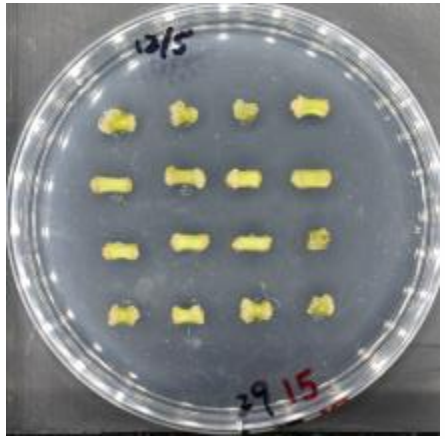


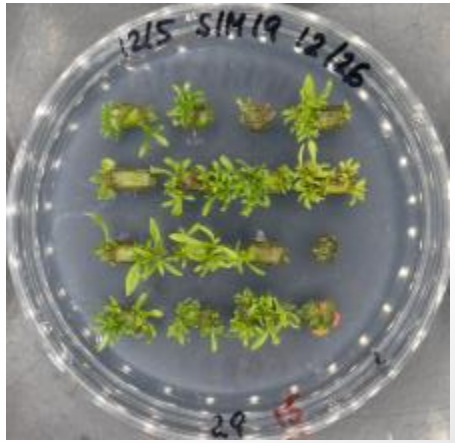
SLMB-28-1



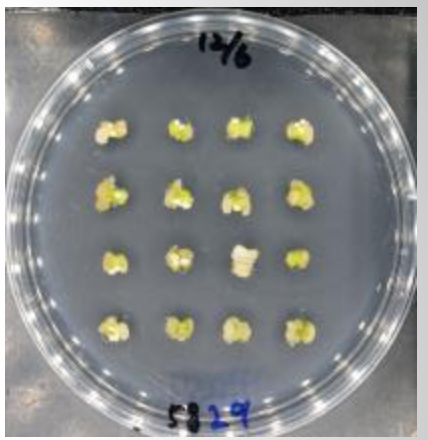


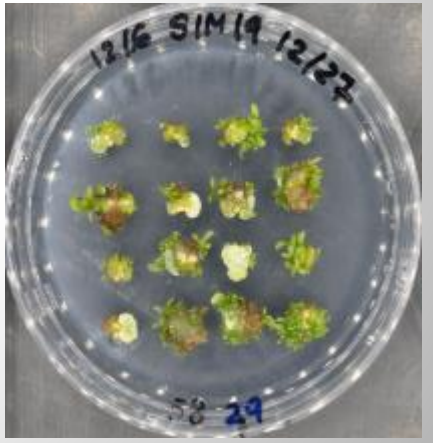
NISQUALLY





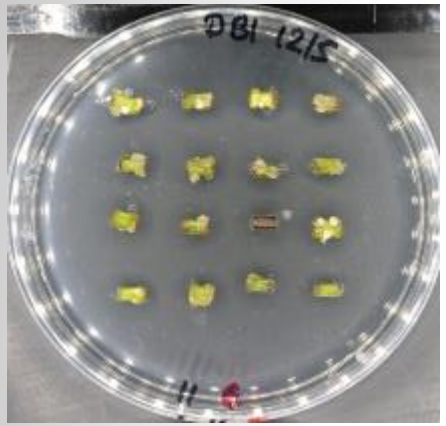
BESC-226



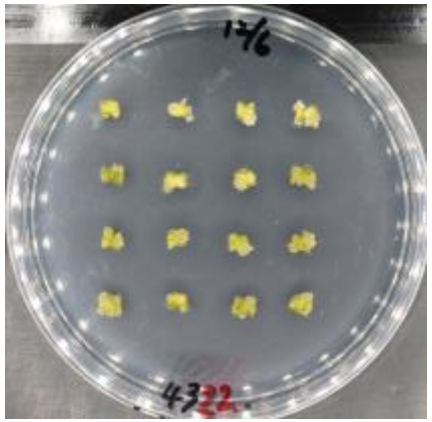


Poor response genotypes

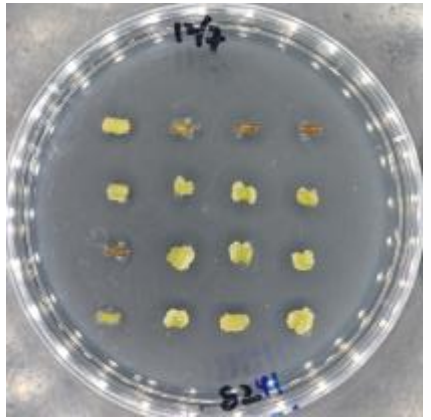
CHWJ-27-2



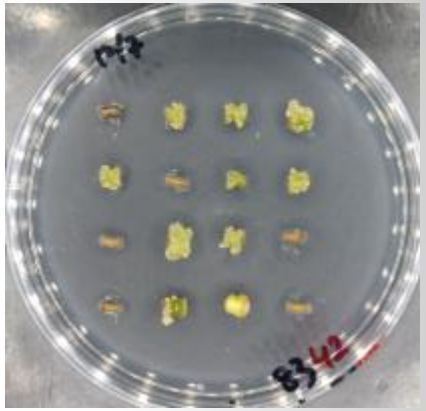
CHWG-27-2

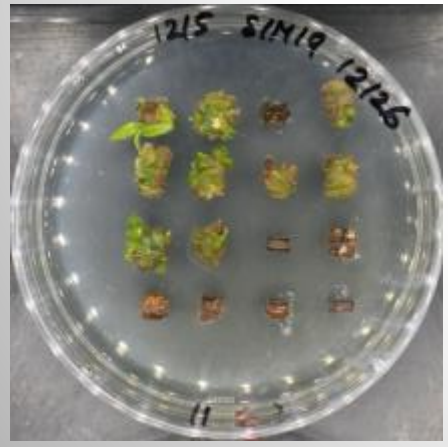


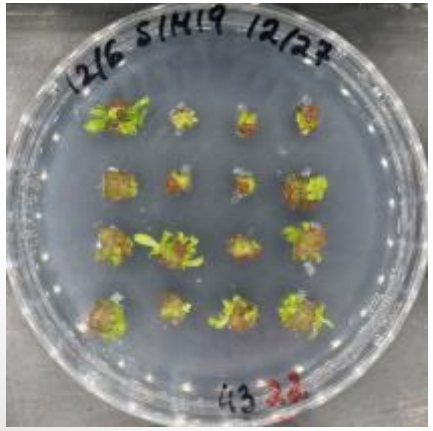
BESC-869

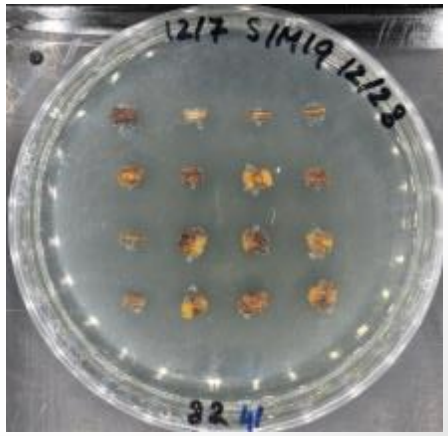


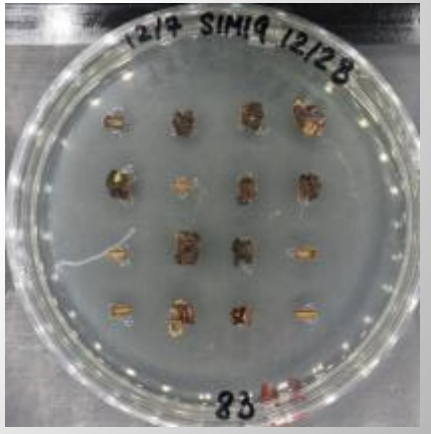
GW-9857











Future plans

- Use of auxotrophic Agro to avoid wash step in GWAS?
  - LBA4404-thy is kindly provided by Corteva for GWAS transformation. This Agro strain has thymidine synthetase KOd so it dies when explants are transferred to media without exogenous thymidine. It will allow us to skip the washing step in the transformation process and reduce labor and time; thus speeding up the transformation procedure
- Low levels of antibiotics in medium in transformation help to control bacterial contamination
  - Need to assess effect on regeneration
- Retest 3-4 effective wild types Agro for gall induction under *in vitro* condition with longer stem explants and more concentrated Agro solution
  - To conduct GWAS for relative susceptibility to wild Agro based on gall growth
- Use of smaller media cells for regeneration and transformation GWAS?
  - We will test 96-well plates to assess if it can save money on media and allow more GWAS conditions to be assessed (based on chlorophyll and fluorescent reporter signals)
- Expect to begin transformation GWAS in late fall to winter

Publication goals / fall-winter manuscripts

- Regeneration and transformation systems
  - Regeneration optimization
  - Genetic variation and heritability
  - Transformation optimization and heritability
  - Imaging system and data analysis pipeline
- GWAS
  - *In vivo* stem regeneration
  - *In vivo* root regeneration
  - *In vitro* callus and shoot regeneration (direct and indirect)

THANK YOU FOR LISTENING

Oregon State University

Phenomics II: Experimental imaging and image analysis pipeline, DEV gene study example/s

Michael Nagle

NSF PGRP advisory meeting

Oct. 3, 2019

Background and overview of phenomics workflow

- *macroPhor Array* used for high-throughput RGB and hyperspectral imaging
- Large volume of data to organize and manage
- Manual scoring of phenotypes and the transition toward automated, high-throughput, objective methods
  (machine vision, hyperspectral, and the intersection of both)
- Transformation optimization experiments:
  - To demonstrate phenomics workflows (to be refined and used in GWAS)
  - To discuss challenges and plans for transformation optimization experiments themselves

Presentation Overview

I. Data acquisition and management

II. Introduction to transformation optimization experiments

III. Analysis of phenotype data, by phenotyping method

   A. Manual scoring

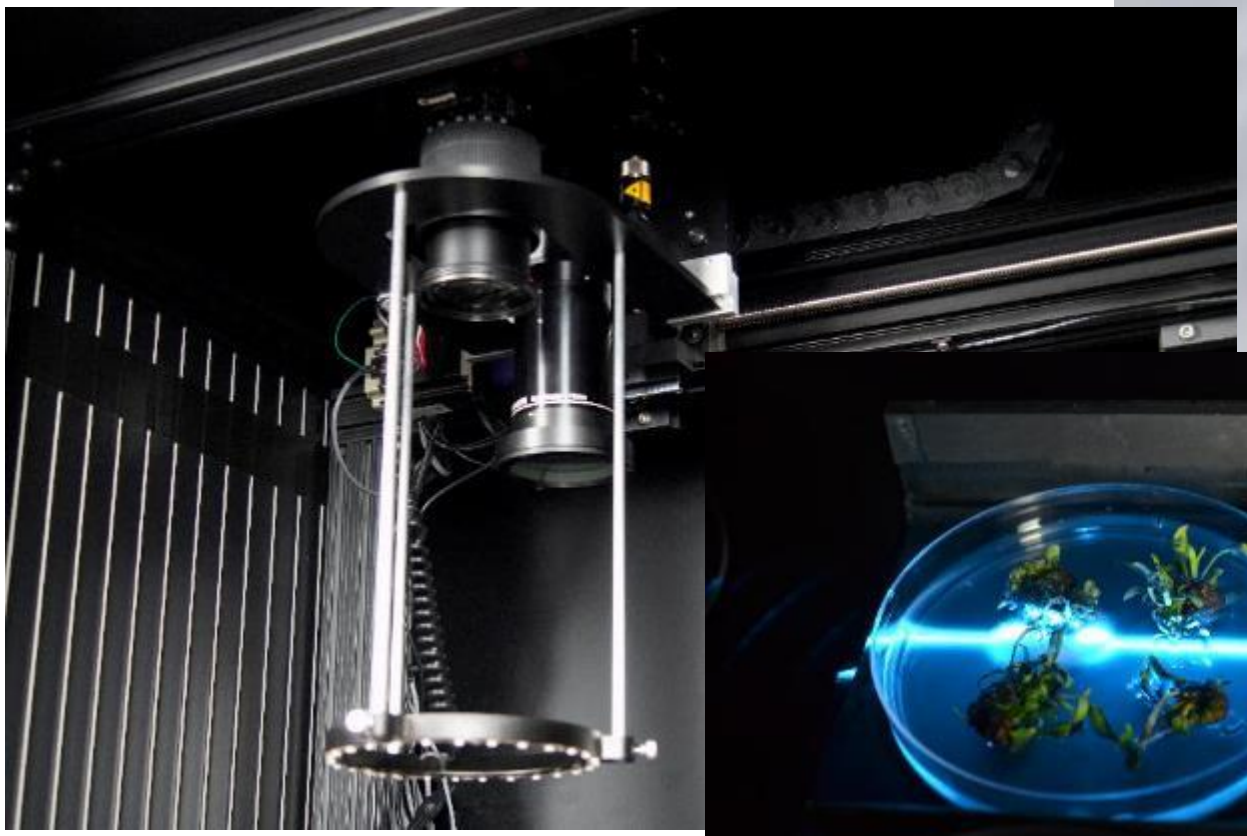
   B. Machine vision data

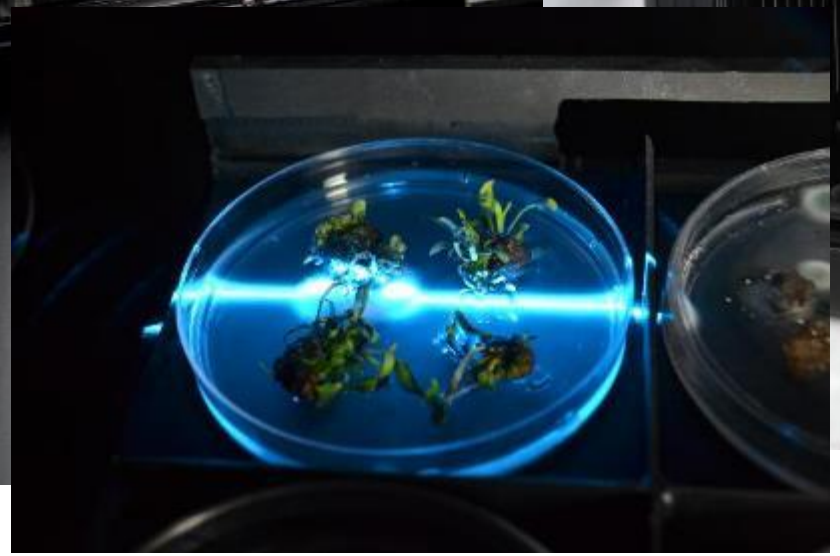
   C. Hyperspectral image data

   D. Integration of machine vision and hyperspectral data

macroPhor Array™

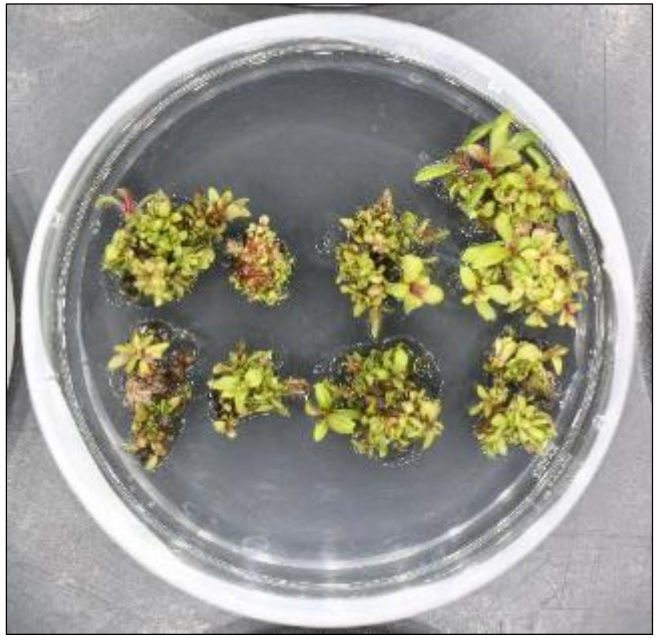
Custom instrument for high-throughput hyperspectral & RGB imaging

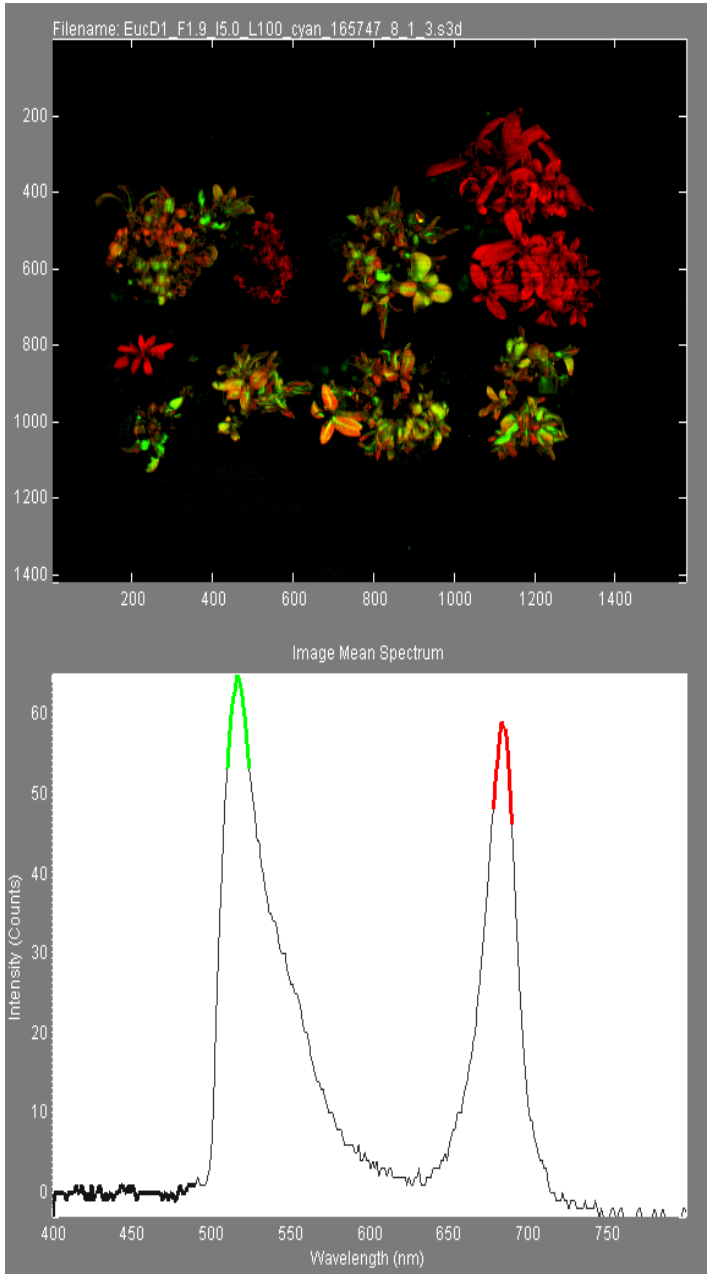




RGB and hyperspectral images are captured

- Hyperspectral images contain a spectrum for each pixel
  - False color applied to certain wavelengths for inspection (right, top)
  - Mean image spectrum shown (right, bottom)
- Standard RGB images
- Ongoing work to align images and integrate hyperspectral analysis with machine vision segmentation of RGB





Unique filenames for each image are connected to biological data through a dictionary

- CSVs contain key linking specific tray and plate ID #s to biological data (e.g. genotype and hormone/gene treatment)
- *macroPhor Array* saves RGB and hyperspectral images for each plate with filenames indicating tray ID and position of plate on tray:

Tray ID and hyperspectral camera settings input by user for each tray

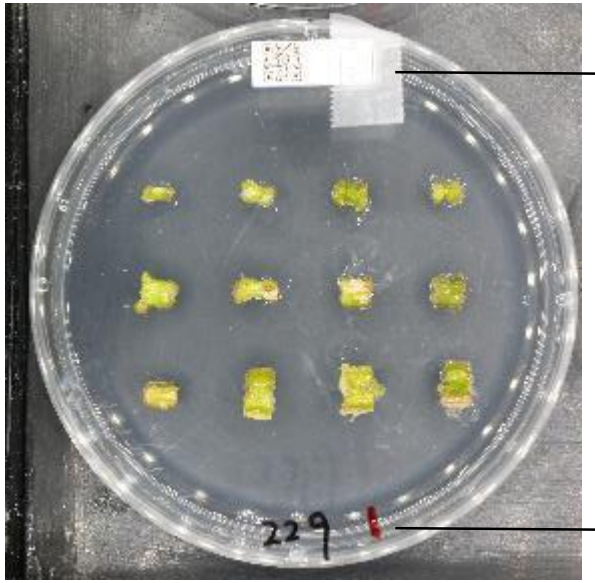
Appended by *macroPhor Array* for each plate

GWA10_I2.0_F1.9_L80_193846_0_0_1

GWAS phase A	Tray #10	2ms integration (exposure)	1.9" focus position	Laser power 80	Time	Row	Column	# plates on tray imaged so far

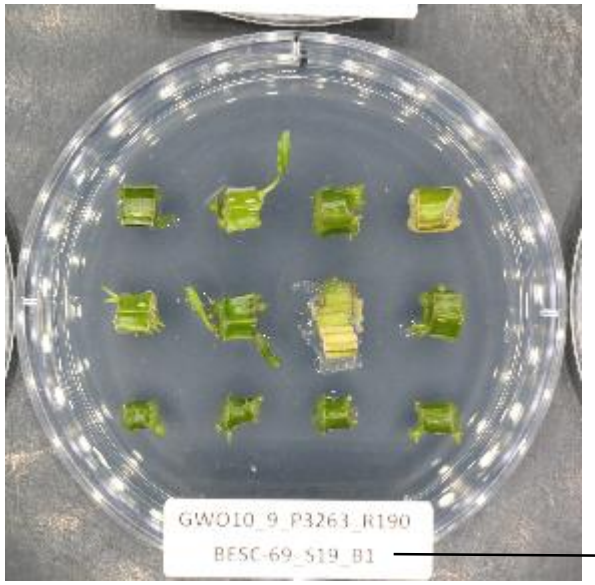
Labeling system to reduce human error risk

- Filename-dictionary scheme requires operator to place plates in specific positions and name trays appropriately
- Potential for labeling system to automate away need for managing filenames and dictionary **or at least provide redundancy**
- Options for labeling systems:
  - 1d barcode (very limited information, e.g. serial number)
  - QR code (attempted)
  - Alphanumeric – currently in use



Size, focus, readability issues lead to errors in reading

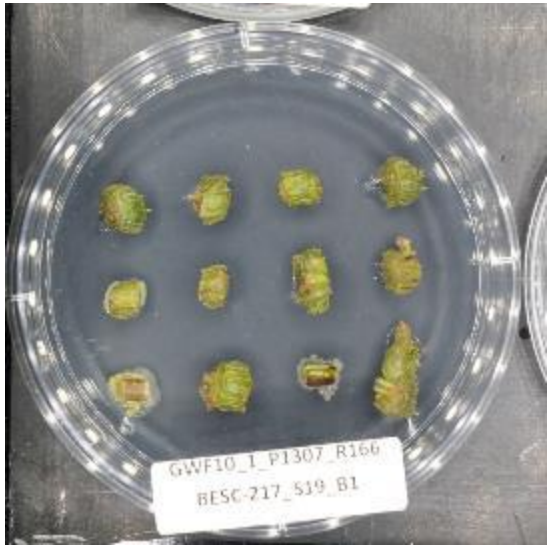
Labeling began in GWAS phase 4. Prior, handwritten numeric IDs



Extra redundancy: Dictionary keys **and** IDs Readable enough with imaging settings for plants

(Automated) high-throughput screening of data for errors

Human error opportunity	Detection method
Crooked plates	Manual sweep probably quickest
Wrong camera settings	• Check integration time and focus in hyperspectral metadata header (.hdr) • Compare chroma standards to measure laser strength?
Plates placed on wrong tray/slot -OR- wrong tray ID in filename	Compare labels and filenames to keys and IDs in dictionary



- Machine vision reading of labels to speed things up?
  - Time to write code and run vs checking manually?
  - High error rate of machine vision a concern
    - Redundancy of information within labels reduce risk exponentially?

Data storage and backup

- Local copies on hard drives:
  - Failure of two 8TB SeaGate drives (no data lost)
  - All images taken since acquiring imager in Apr. 2018 stored locally indefinitely (to continue?)
- Cloud backups:
  - Team members can search, view, download
  - Current backups for all images on Box
  - Starting Google Drive backups for redundancy (with cloud-cloud sync)

Individual hard drive	Status	Capacity	Cost
Seagate 1	Full	8TB	$149
Seagate 2	Full	8TB	$149
Seagate 3	Failed	8TB	$149
Seagate 4	Failed	8TB	$149
Western Digital 1	Full	10TB	$204
Western Digital 2	25% full	10TB	$204
Internal solid state	Usually ~50% full	4TB	$600

Presentation Overview

I. Data acquisition and management

II. Introduction to transformation optimization experiments

III. Analysis of phenotype data, by phenotyping method

A. Manual scoring

B. Machine vision data

C. Hyperspectral image data

D. Integration of machine vision and hyperspectral data

In preparation for GWAS of transformation:

Experiments to identify heritable treatments affecting transformation

- Effect of treatment on trait
  - Another heritable phenotype?
  - Unmask QTLs hidden by upstream recalcitrance to transformation/regeneration
- Enhancers of transformation itself
  - Chemical treatments to enhance transformation itself (e.g. acetosyringone, Sil-wet)
  - Agrobacterium strains and virulence (helper) plasmids
- Enhancers of regeneration
  - Hormone treatments (main experiments, for optimization papers, complete)
  - Developmental (DEV) genes as regulators of regeneration (and embryogenesis)
- Relevant both to GWAS project and GREAT TREES industry/academia co-operative on enhancing transformation/regeneration

Background: DEV genes to enhance regeneration… and GWAS of regeneration

- Overexpression of developmental genes (DEV genes) enhances shoot regeneration in plants including Arabidopsis, monocots, soybean, poplar
- Area of research rapidly progressing, expanding to additional genes, species
- Potential for DEV gene overexpression to unmask variation in GWAS

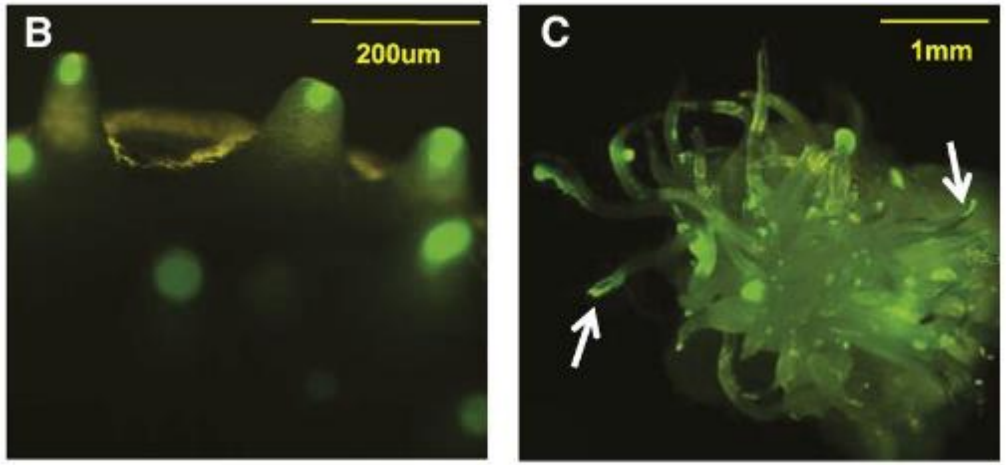
The Plant Cell, Vol. 28: 1998–2015, September 2016, www.plantcell.org © 2016 American Society of Plant Biologists. All rights reserved.

BREAKTHROUGH REPORT

Morphogenic Regulators *Baby boom* and *Wuschel* Improve Monocot Transformation[OPEN]

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,[c] Julia Chow-Yiu,[a] Wei Hua,[a] Maryanne Yu,[b] Jenny Banh,[b] Zhongmeng Bao,[a] Kent Brink,[d] Elizabeth Igo,[d] Bhojaraja Rudrappa,[e] PM Shamseer,[e] Wes Bruce,[f] Lisa Newman,[a] Bo Shen,[a] Peizhong Zheng,[g] 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]

Somatic embryogenesis in recalcitrant maize lines enhanced by overexpression of WUS (co-transformed with GFP)



frontiers in Plant Science

MINI REVIEW
published: 02 October 2018
doi: 10.3389/fpls.2018.01443

Opportunities for Innovation in Genetic Transformation of Forest Trees

Michael Nagle[1†], Annabelle Déjardin[2†], Gilles Pilate[2] and Steven H. Strauss[1*]*

[1] Forest Ecosystems and Society, Molecular and Cellular Biology, Oregon State University, Corvallis, OR, United States, [2] BioForA, INRA, ONF, Orléans, France

Effects of DEV genes on regeneration: An additional GWAS treatment

Species	Gene
Populus trichocarpa	*LEAFY COTYLEDON 1 (LEC1)*
	LEAFY COTYLEDON 2 (LEC2)
	EARLY BUD BREAK 1 (EBB1)
	WUSCHEL 2 (WUS2)
	BABY BOOM (BBM)
Populus tomentosa	*WUSCHEL 1 (WUS1)*
	WUSCHEL-ASSOCIATED HOMEOBOX 5 (WOX5)
	WUSCHEL-ASSOCIATED HOMEOBOX 11 (WOX11)
Populus trichocarpa	*WUSCHEL 1 (WUS1)*
Helianthus annuus (sunflower)	*WUSCHEL (WUS)*
Gnetum gnomon	
Malus domestica (apple)	
Vitus vinifera (grape)	
Populus trichocarpa	*GROWTH REGULATORY FACTOR 5*

Studied in Strauss Lab pilot studies, to be studied further in high-throughput screens

From Beijing National Forest Academy, to be studied first in high-throughput screens

Corteva plasmids

Cloning in progress

Pilot DEV studies revealed variables which reduce shoot, GFP phenotypes and/or add noise to data

Variable	How to deal with
Age of *in vitro* materials, progressive decline in regeneration ability	**Use of only young *in vitro* materials, or *in vivo* materials from greenhouse**
Necrosis of leaf explants (seemingly randomly)	**Use of stem explants only**
Escape from selection	Switch from kanamycin to geneticin or gentamycin
Agro culture health (proportion dead cells)	**Inoculate all cultures via single colony to starter culture to 50mL culture, simultaneously**
Selectable marker expression, can vary if distance between (promiscuous) promoter and marker varies	**Use minimal promoters, consistent spacing in experimental/control plasmids**
Rate of fluorescent reporter expression	Switch from pRolD:GFP to GmUbi:ZsYellow and GmEFA:DsRed2 (Pioneer)
Incomplete transgene integration	Use of spacers next to T-DNA insertion sites
Readthrough transcription/translation of genes outside T-DNA	Use of ALLSTOP elements

Related to transformation methods

Related to plasmid elements

Presentation Overview

I. Data acquisition and management

II. Introduction to transformation optimization experiments

III. Analysis of phenotype data, by phenotyping method

A. Manual scoring

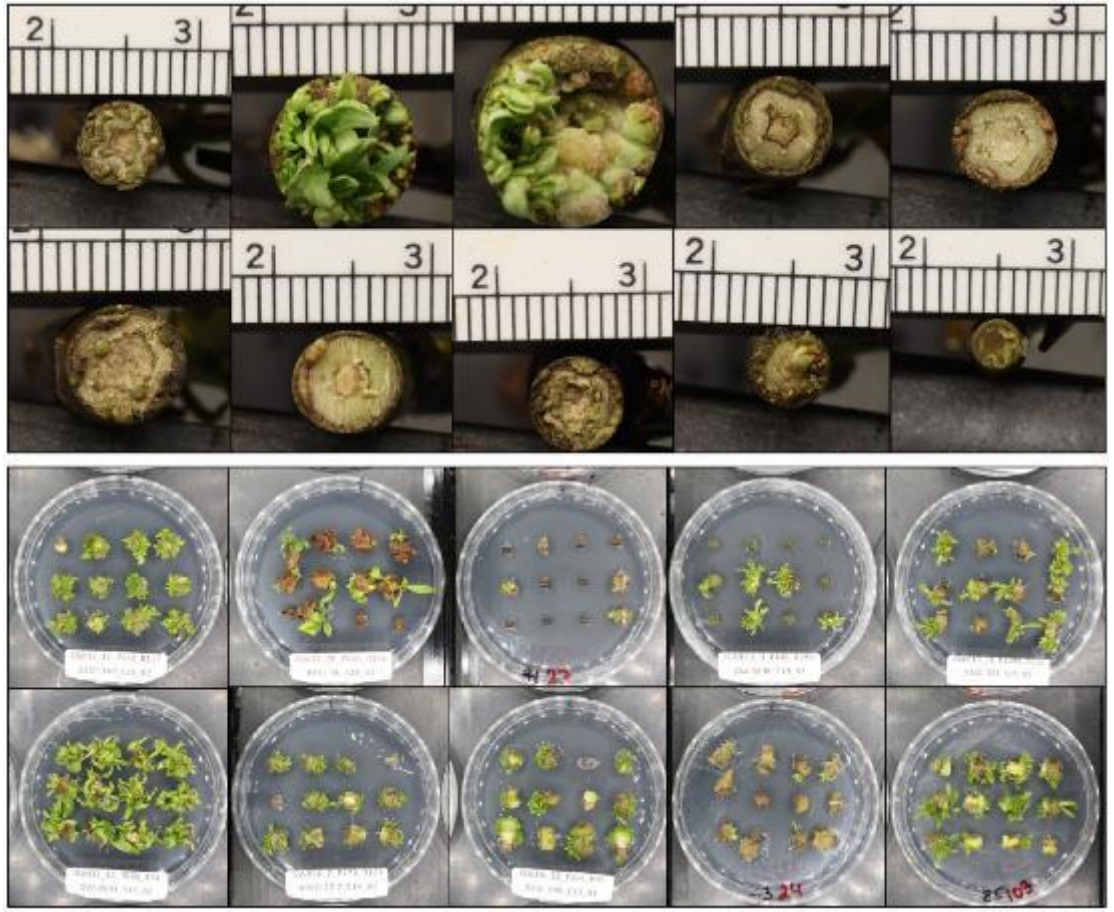
B. Machine vision data

C. Hyperspectral image data

D. Integration of machine vision and hyperspectral data

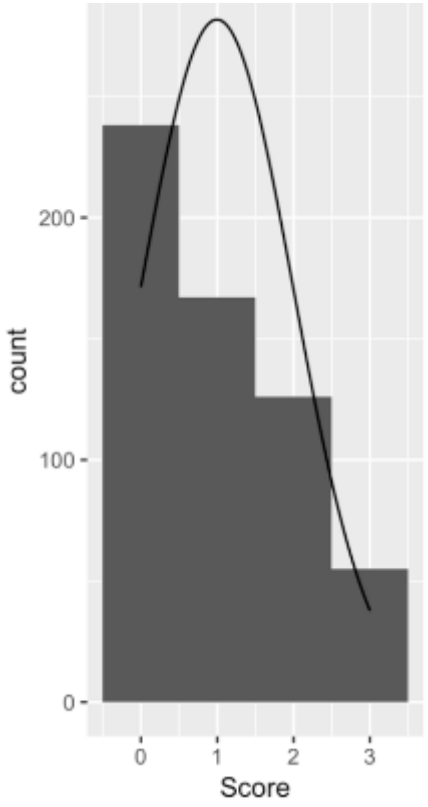
Approaches to manual scoring and consequences for modeling

- Manual scoring complete for:
  - Most stem regeneration GWAS data
  - *In vitro* optimization experiments
  - DEV/Vir gene experiments to date
- Discrete scores by plant/explant:
  - Callus size
  - Callus color
  - Pseudo-count of individual shoots
  - Presence of callus/shoot with fluorescent reporter
- Aggregate statistics over whole plates, derived from discrete scores for each explant
  - e.g. proportion of explants with shoot
  - Smooth over intra-plate variation
  - Coerce data into distribution allowing general models w/o significant (?) normality violation
- Generalized models required if significant normality violation unavoidable

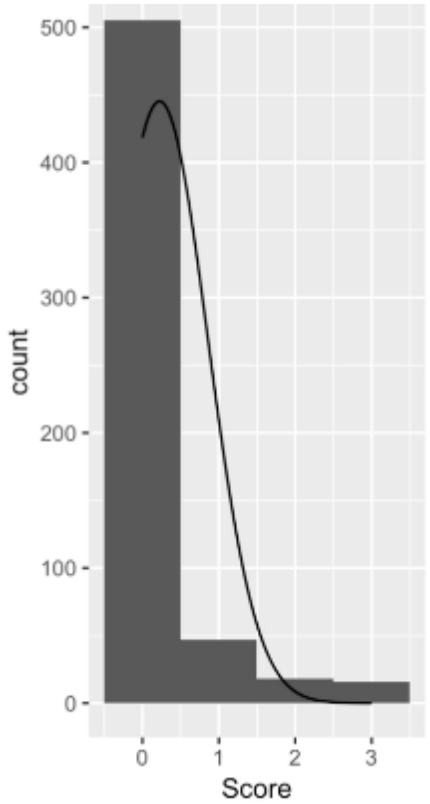


Histograms of stem regeneration data (manual scores)

Callus score at week 3



Shoot score at week 3

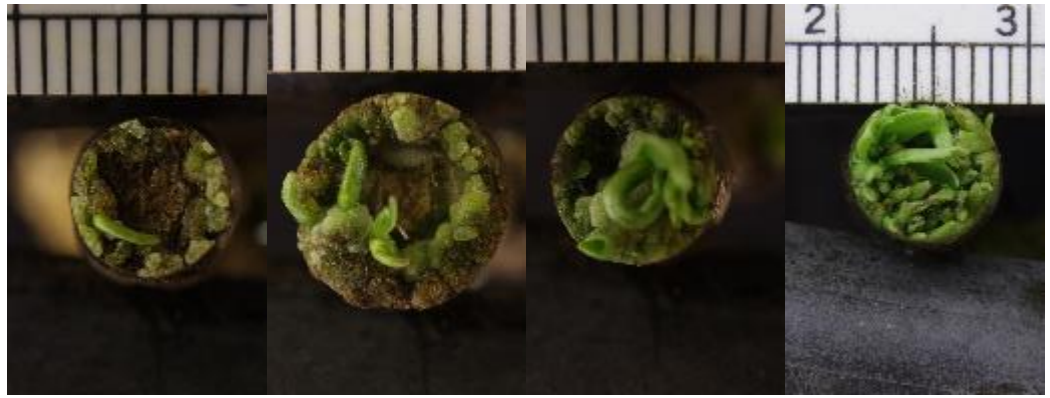


Callus score examples:

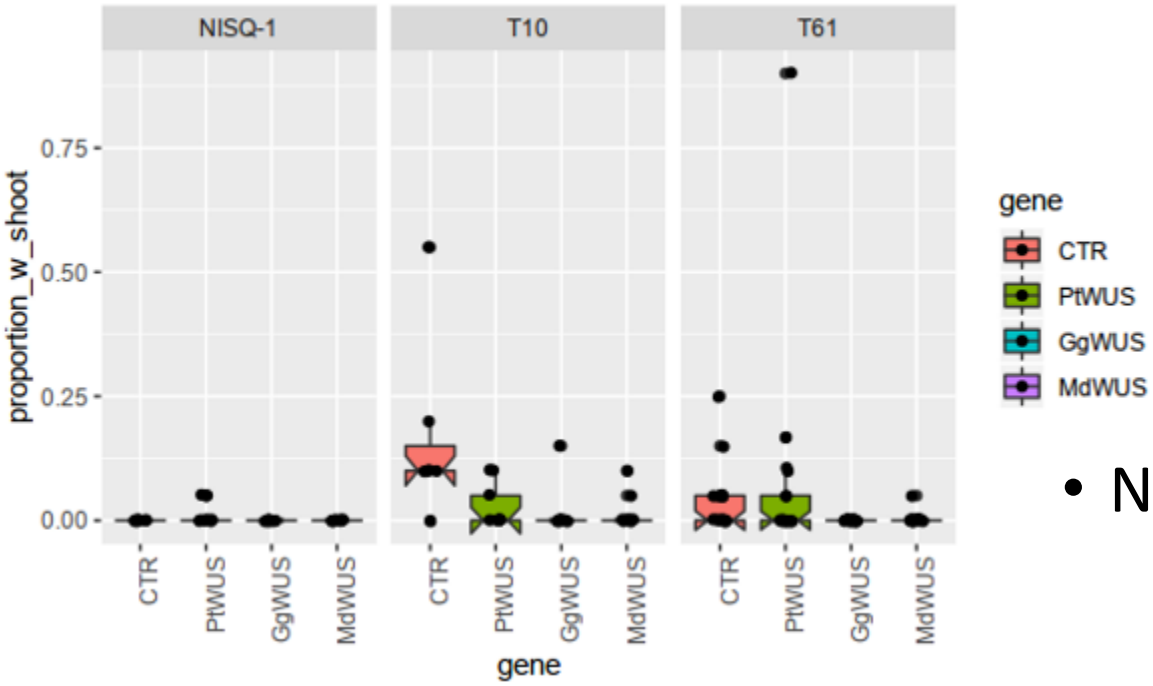
1 2 3 4

Shoot score examples:

1 2 3



Example analysis of manual score data with modeling
(from DEV gene experiment with WUS homologs and superior backbone)



Formula for linear model:

Proportion of explants with callus ~ Gene + Date + Background

Gene treatment	t-value for effect coefficient
P. trichocarpa WUS	0.098
M. domestica (apple) *WUS*	-3.342
G. Gnomen WUS	-1.595

- Negative results not surprising given:
  - Role of WUS in establishing, maintaining shoot *primordia* specifically (Zhang 2017 Plant Cell)
  - Developmental arrest when WUS expressed w/ strong promoter in Arabidopsis (Zuo 2002 Plant J)
- Next: Transient expression?
  WUS coexpressed w/ other genes?

Presentation Overview

I. Data acquisition and management

II. Introduction to transformation optimization experiments

III. Analysis of phenotype data, by phenotyping method

A. Manual scoring

B. Machine vision data

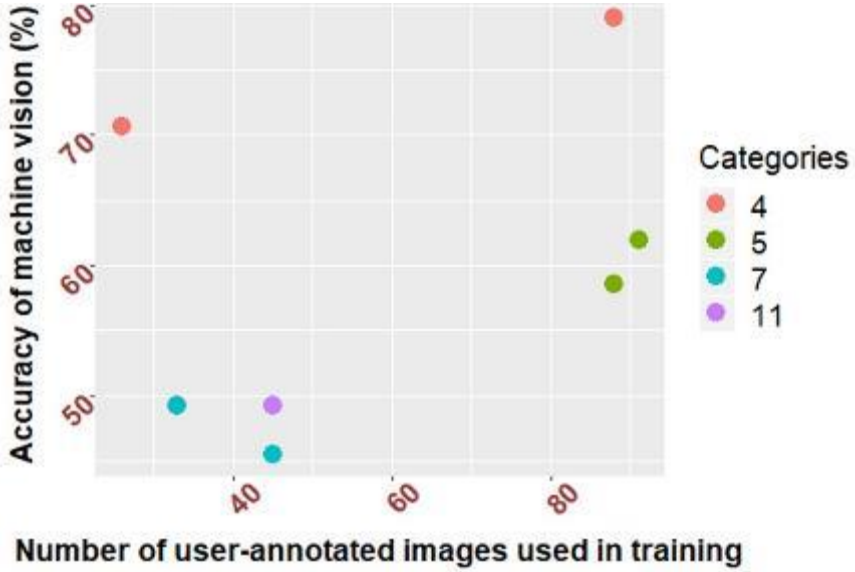
C. Hyperspectral image data

D. Integration of machine vision and hyperspectral data

Machine vision accuracy and precision depends on choice of architecture as well as training and task

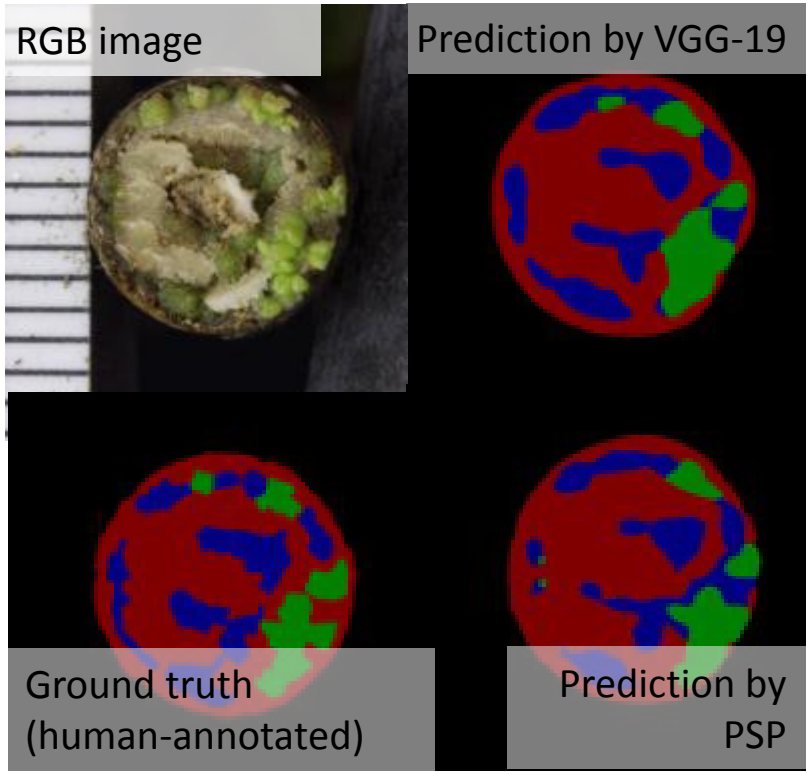
Shoot/callus can be divided into multiple classes by color

Accuracy for VGG-19 models:
Fewer classes, better performance



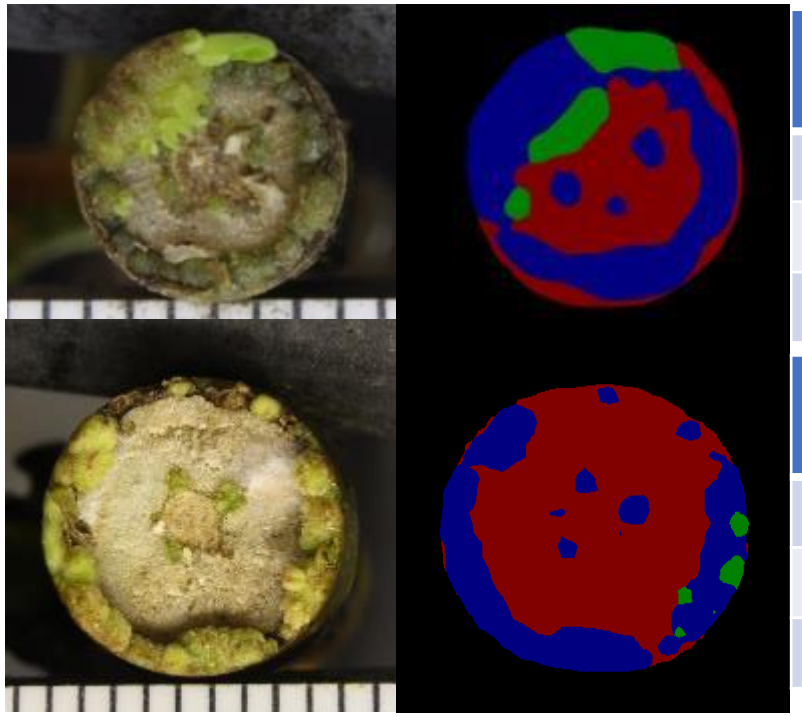
Varying precision/accuracy for different architectures:

- VGG-19 (2014)
- Pyramid scene parsing (2016)
- DeepLab (2018)



In addition to measuring amount of each tissue, can count separate instances

Machine vision task	Biological trait	Statistical distribution
Semantic segmentation	Proportion of total plant area classified as X tissue	Normal or lognormal after dropping zero values
Instance segmentation	Number of unconnected (or individual) shoots	Poisson? (TBD)



Tissue class	Percent of total area
stem	45%
callus	43%
shoot	12%

Tissue class	Connected components
stem	N/A
callus	9
shoot	5

Semantic segmentation phenotype distributions and approaches to modeling detailed in upcoming GWAS presentation

Presentation Overview

I. Data acquisition and management

II. Introduction to transformation optimization experiments

III. Analysis of phenotype data, by phenotyping method

   A. Manual scoring

   B. Machine vision data

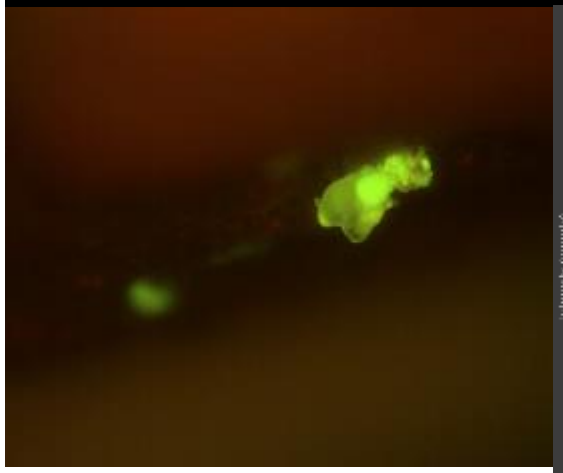
   C. Hyperspectral image data

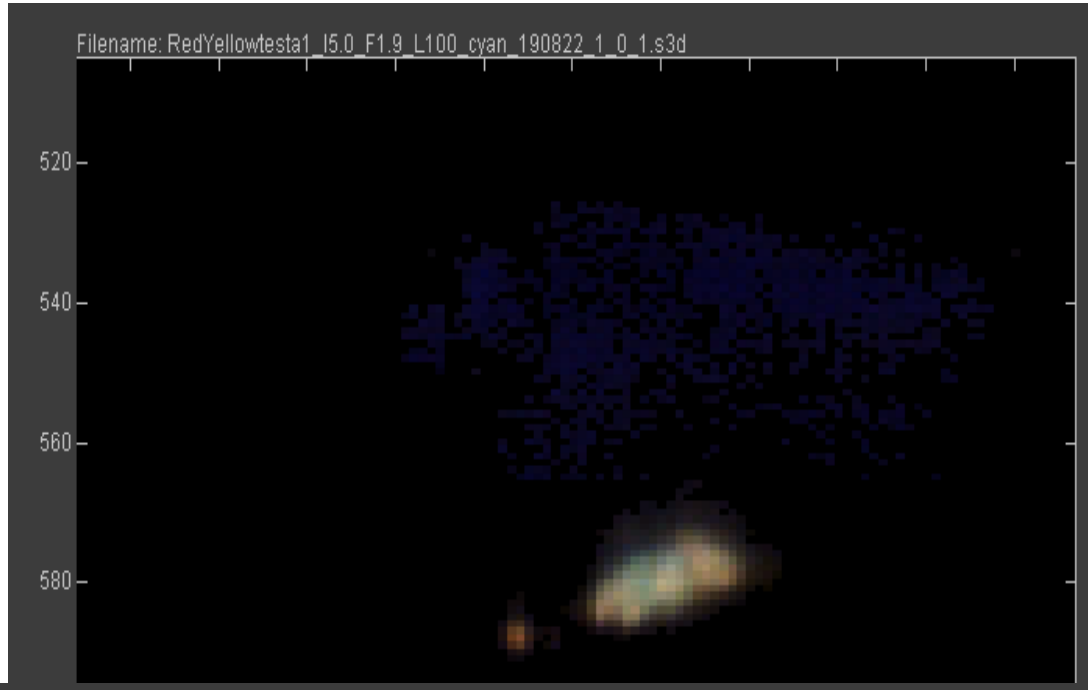
   D. Integration of machine vision and hyperspectral data

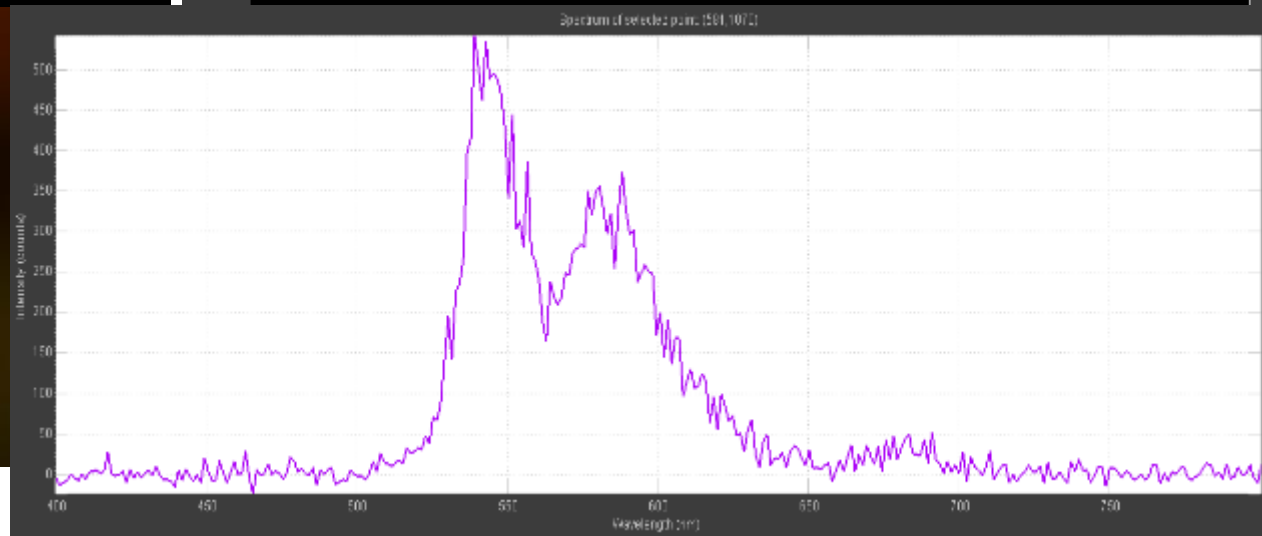
Spectral overlap between fluorescent proteins

Example: DsRed and ZsYellow

T61 harboring *PtWUS1*
Transformed 5/8
RGB image taken 5/29
Fluorescent image taken 6/14
Plate ID: CT2_13; Explant #12



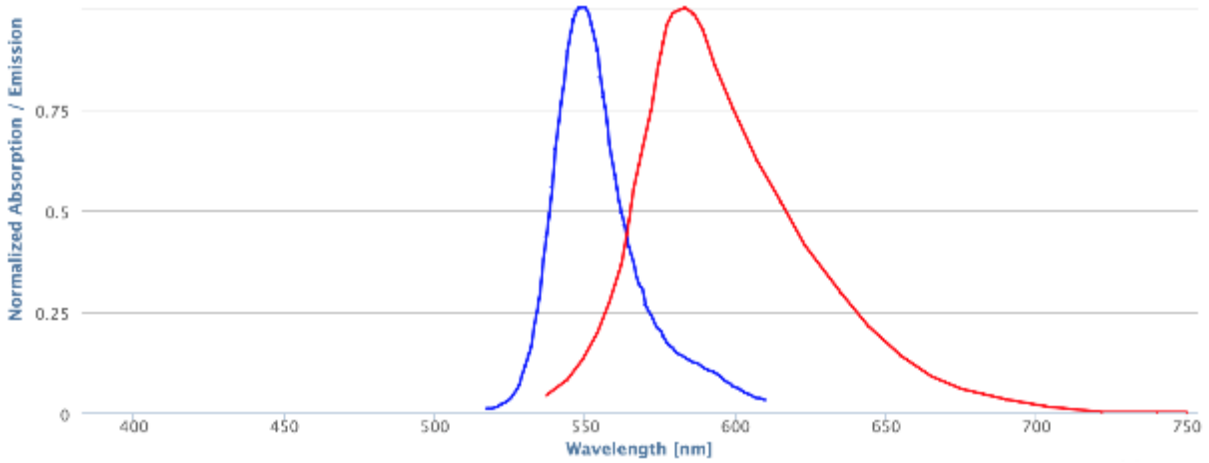


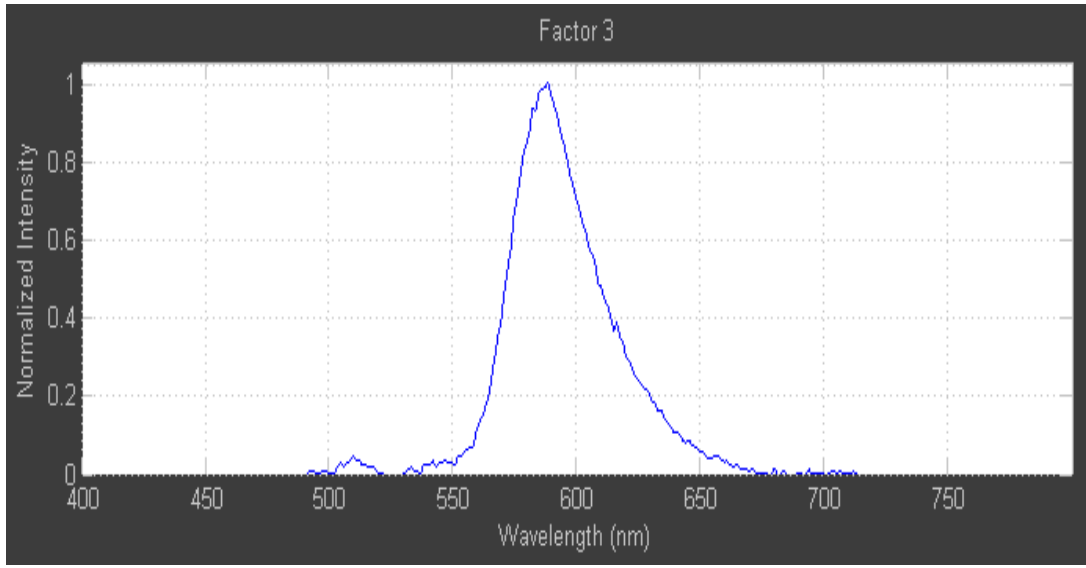


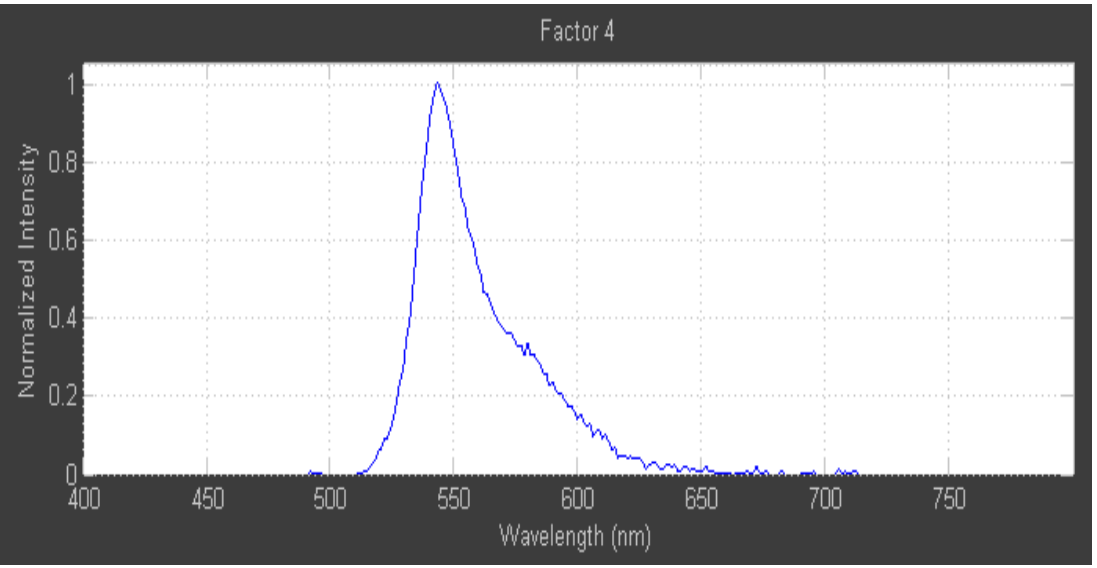
KemoQuant can deconvolute spectra

- Multivariate Curve Resolution (MCR) to deconvolute DsRed and ZsYellow (shown)
- Deconvolution of reporter proteins from chlorophylls as well
- PCA option (KemoQuant and R)

Published spectra for DsRed and ZsYellow

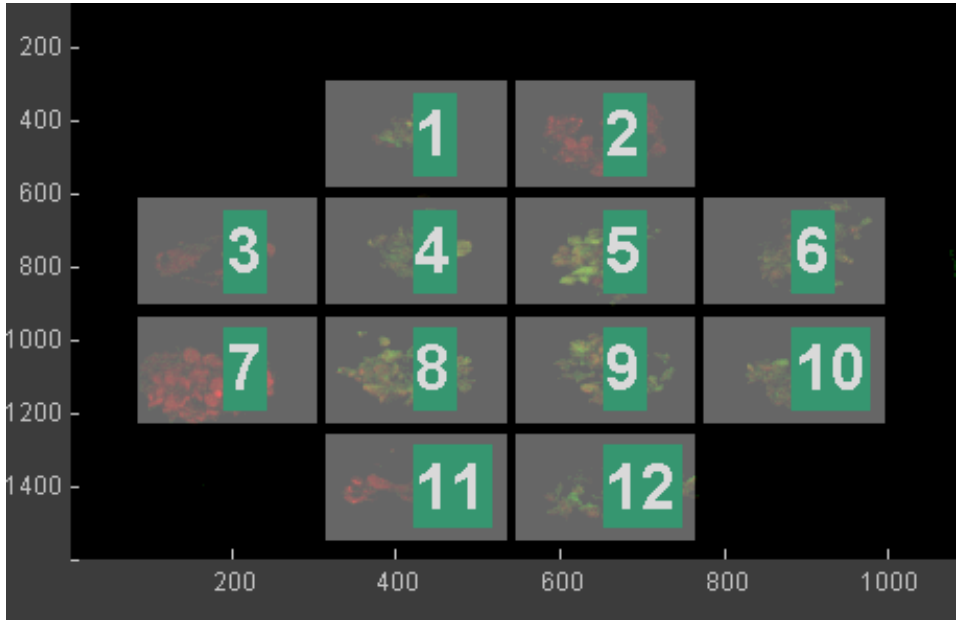


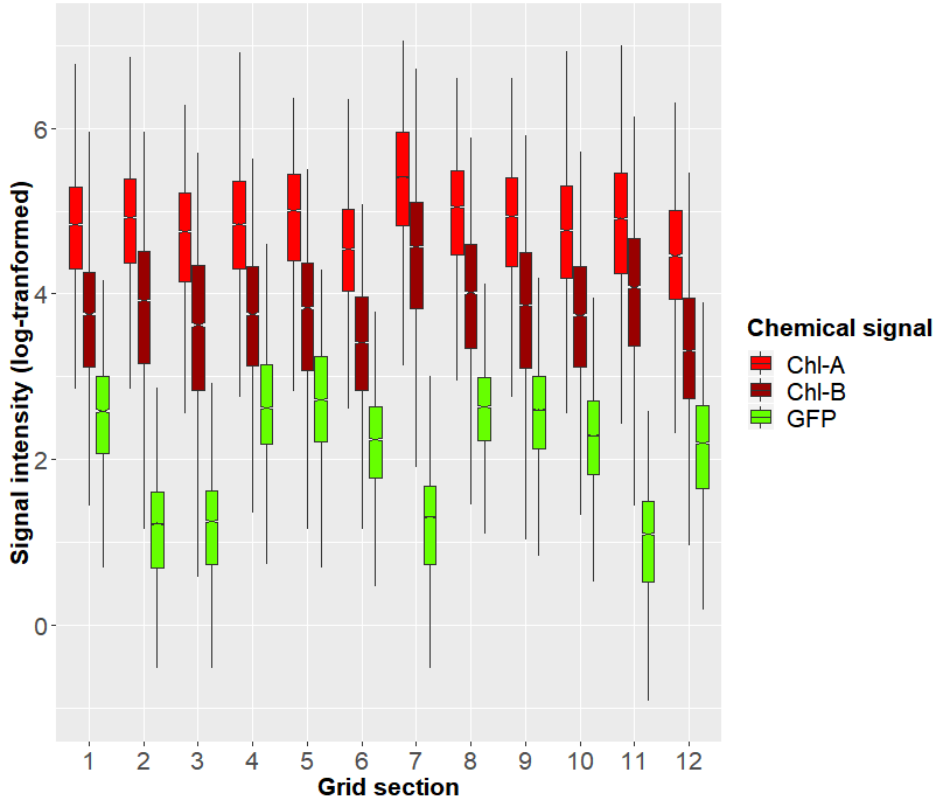




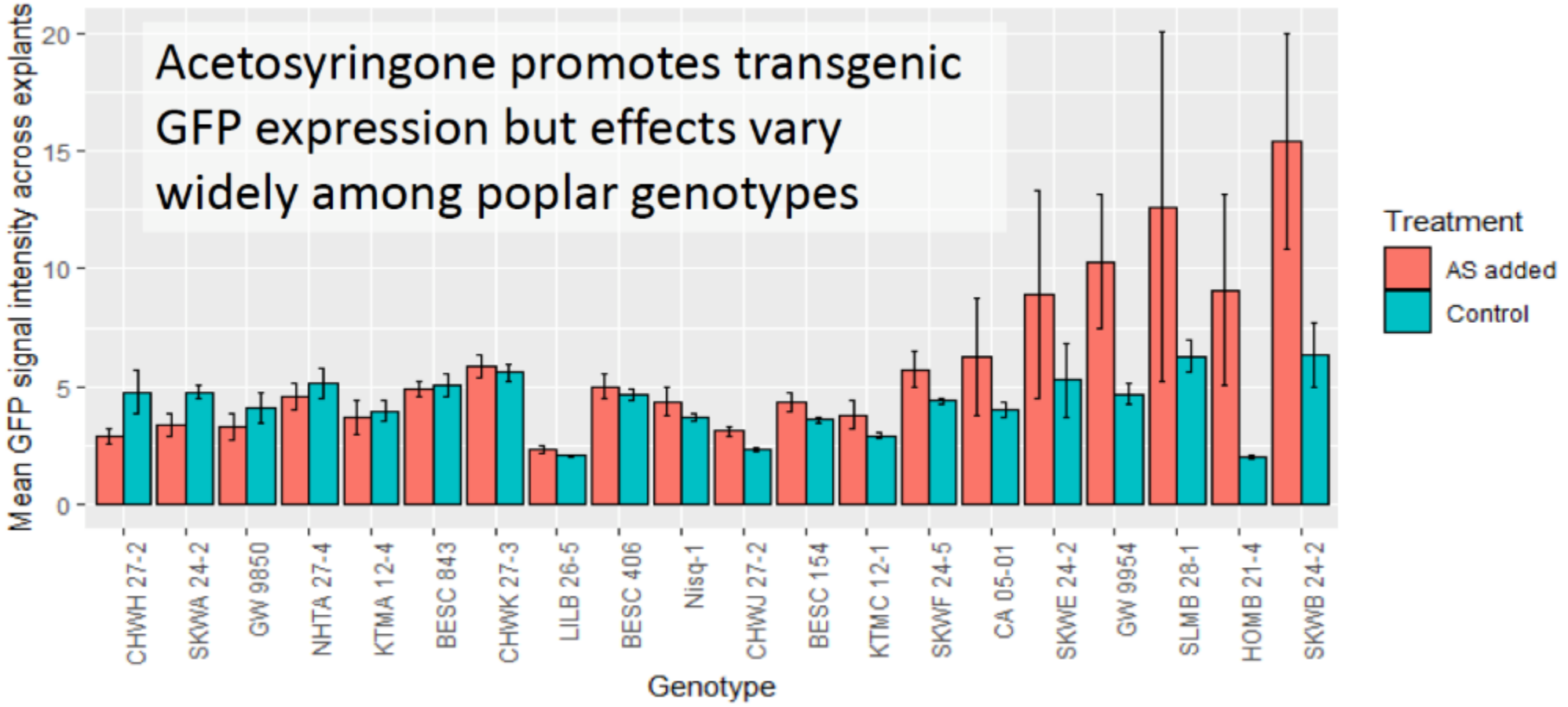
Signals for fluorescent compounds are quantified over each grid item (explant)

- CLS of each pixel's spectrum over each fluorophore's emission spectrum
- KemoQuant or R





Hyperspectral data used for transformation optimization and DEV studies – and next for GWAS



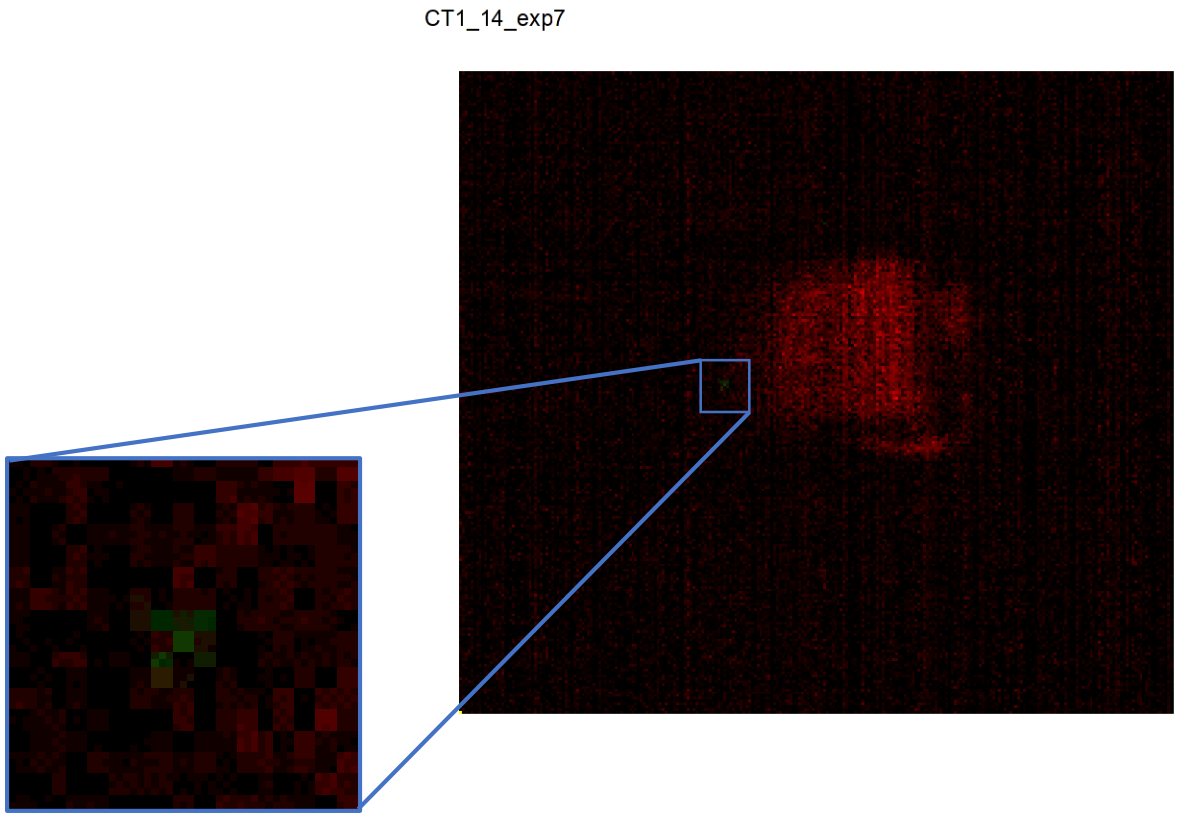
Performance of heuristic suggests hyperspectral analysis recognizes transgenic tissue more reliably than human

- Heuristic – a decision rule that is practical and accurate enough
- Potential for macroPhor Array and R code to perform the task of recognizing transgenic tissue
- Attempted heuristic (tested w/ DEV gene data):
  Transgenic if enough pixels have enough DsRed signal (from CLS)
  - Apparent power of 80% with 5% false positive rate...
  - Assuming human scores are correct
- Inspection of hyperspectral images after classification by heuristic, comparison to manual scores suggests heuristic much more reliable

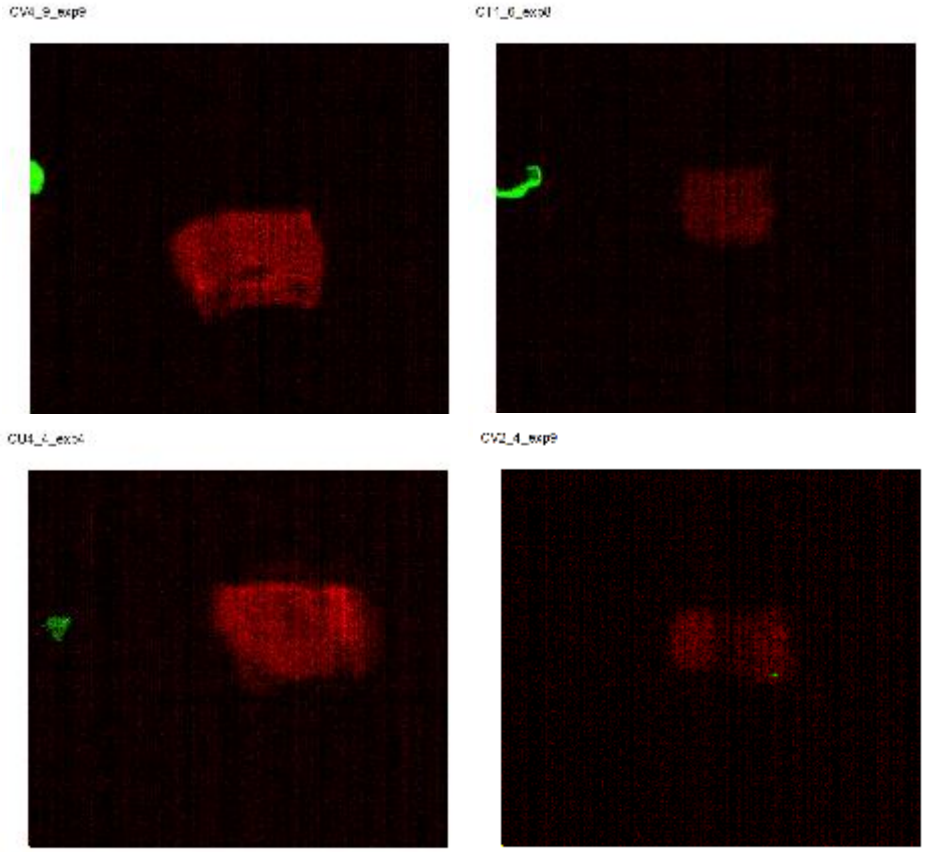
Disagreement between heuristic and manual phenotyping

Hyperspectral analysis may outperform as long as overlap controlled for

"False negatives":
GFP detected by manual phenotyping,
not hyperspectral heuristic

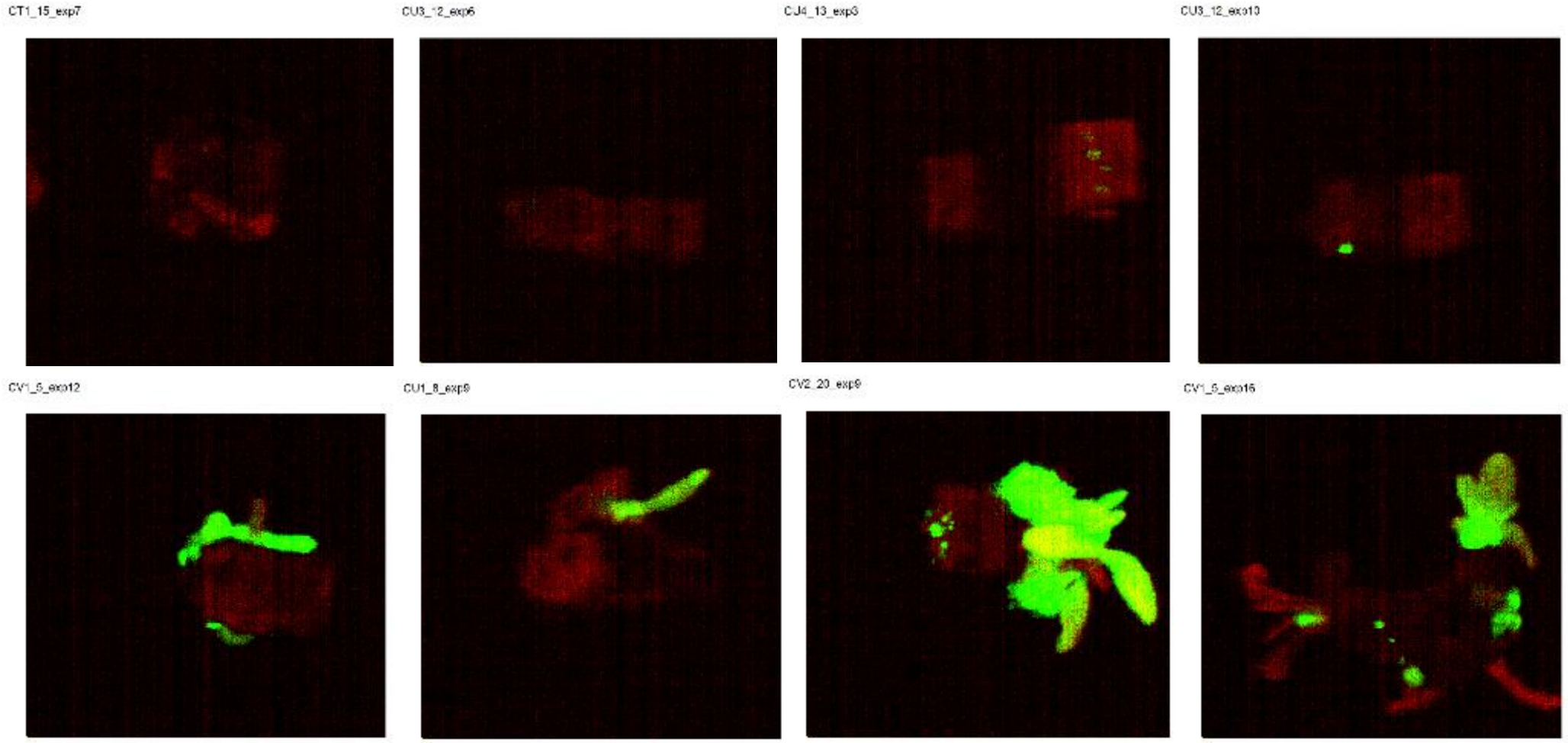


"False positives":
Examples of GFP detected by hyperspectral heuristic,
not manual phenotyping



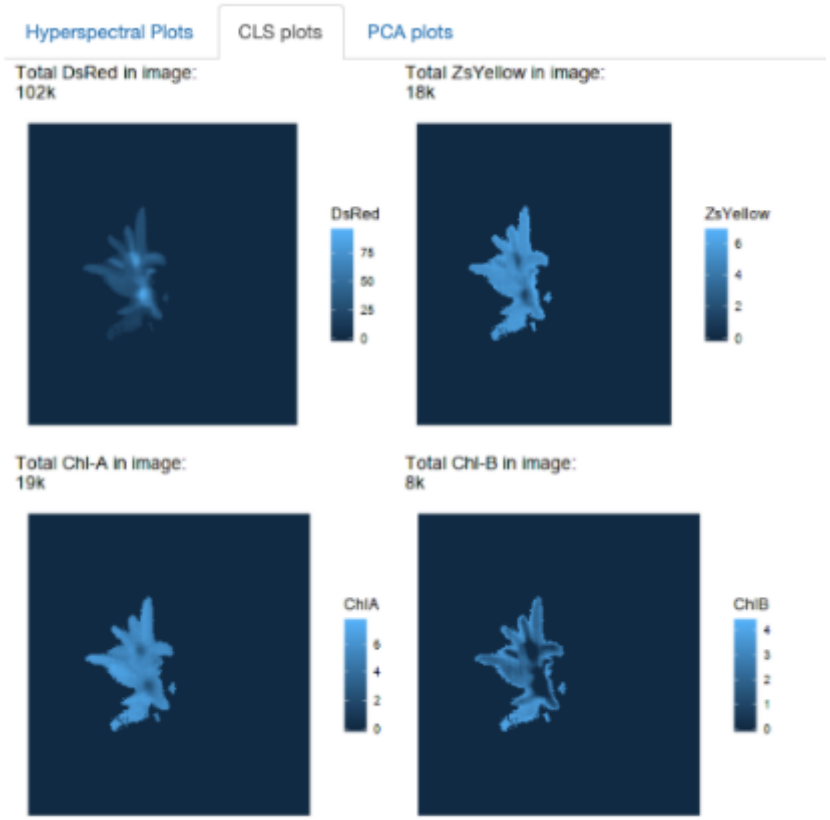
Diversity of phenotypes seen in hyperspectral images

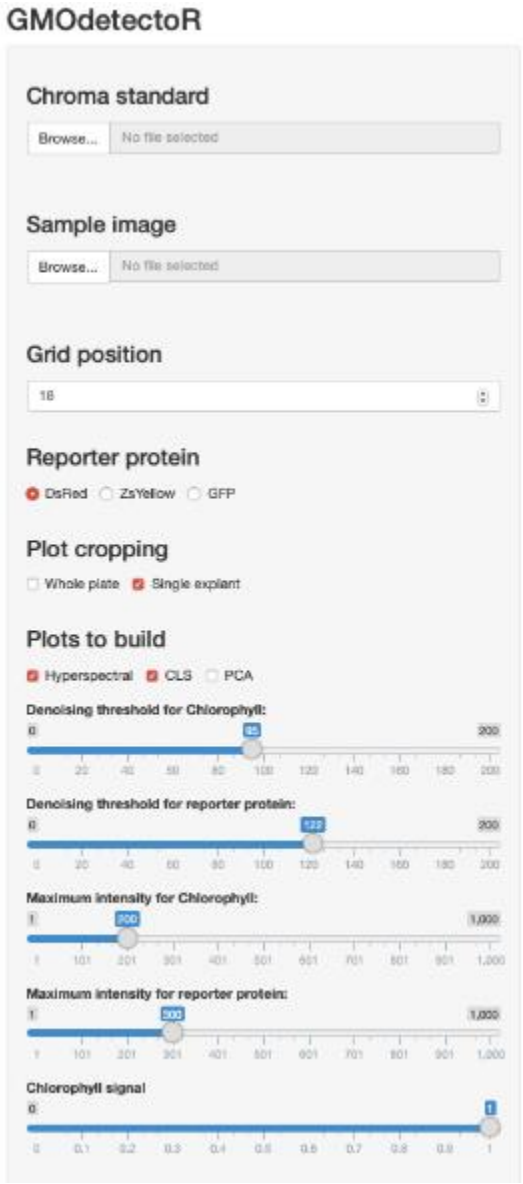
Wide range of fluorescent tissue sizes and types, fluorescence intensity

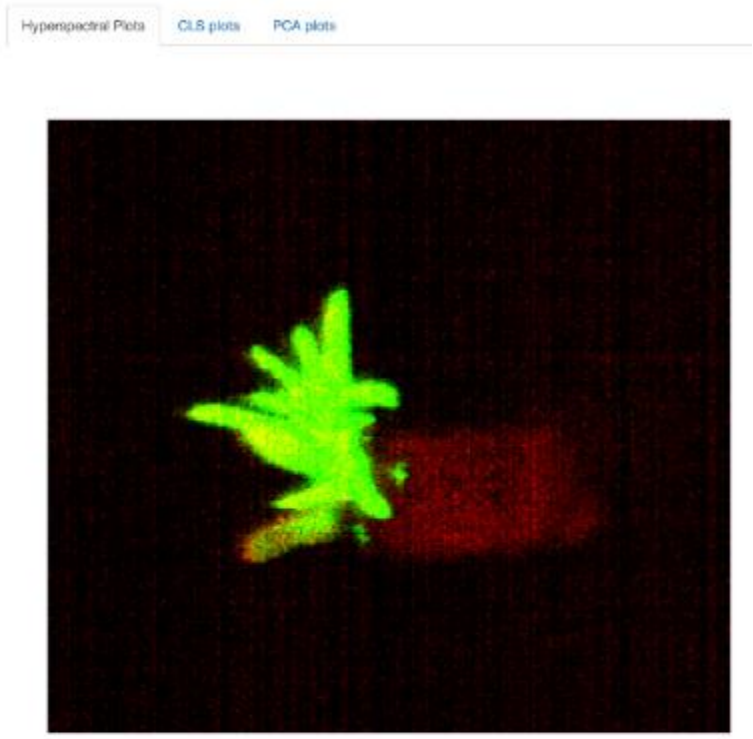


Examples selected to show range of phenotypes

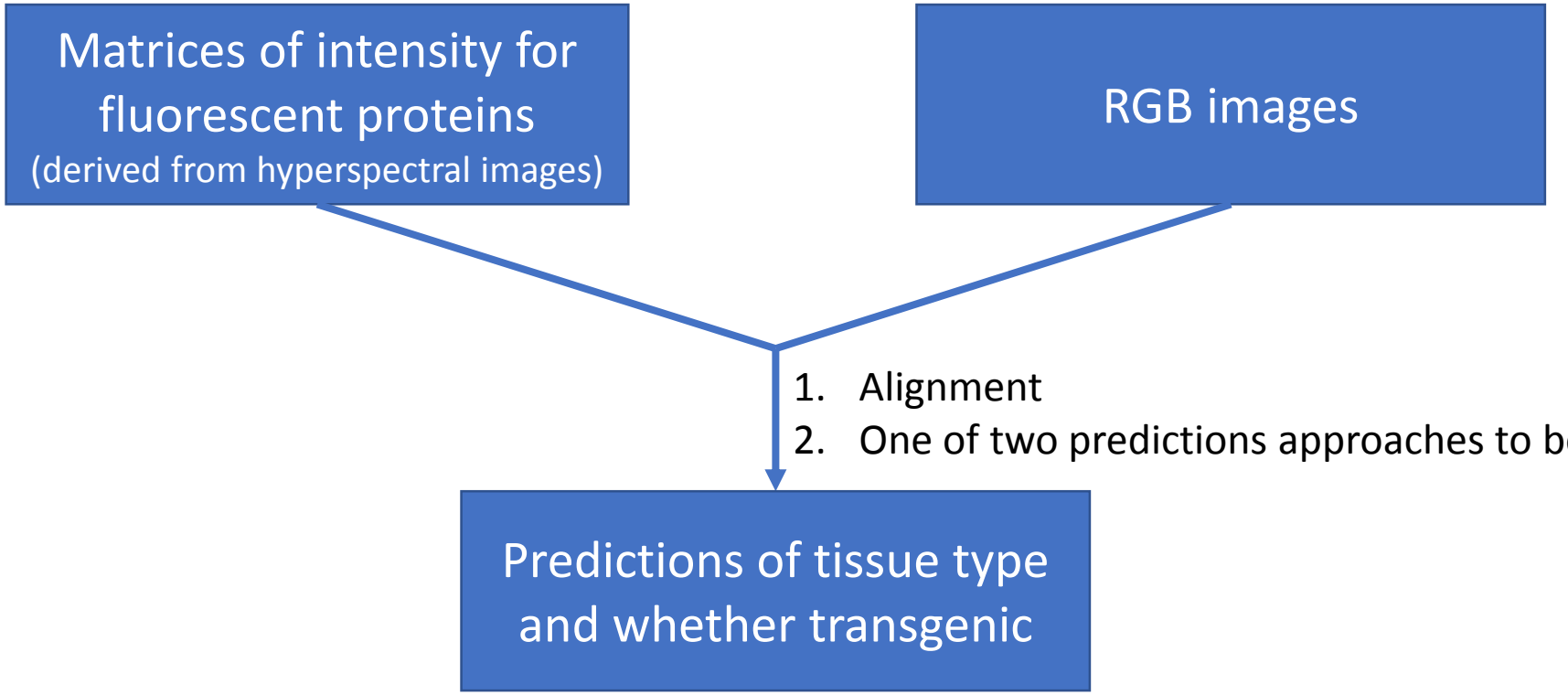
CLS and calculation of cumulative test statistics for fluorophores over select pixels with R







Desire for integration of hyperspectral, machine vision

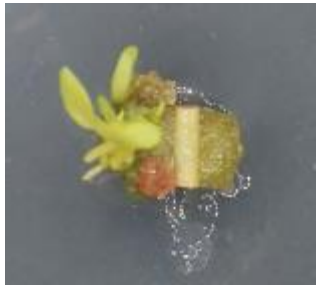


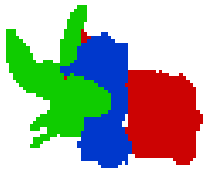
Not illustrated or currently planned:
Integration of RGB, hyperspectral without first reducing hyperspectral data by regression

Dual approaches to integration of hyperspectral, machine vision

Approach 1: Regression of fluorescent proteins after deep learning

- No deep learning applied to hyperspectral data itself
- Regression to measure reporter signal over pixels labeled by deep learning (from RGB images) as X tissue
- Calculation of total reporter in X tissue

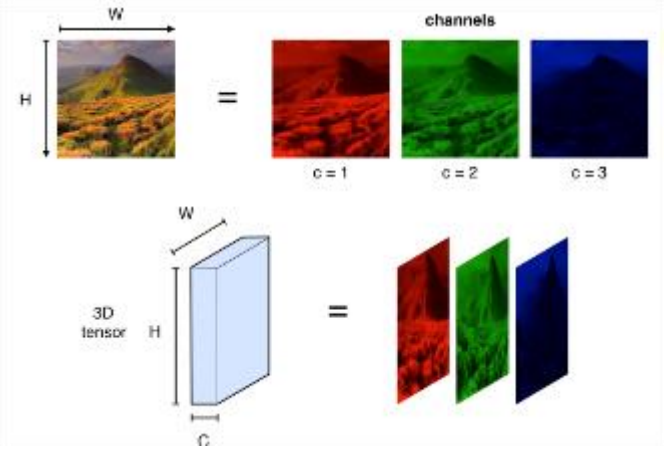




What is total of DsRed test statistics over green area (shoot)?

Approach 2: Deep learning including fluorescent signals from regression

- Stack RGB, fluorophore channels, let neural networks learn from all
- To treat transgenic shoot, nontransgenic shoot as separate classes or within a nested class
- Need for ground truth (annotation)



Beyond R, G, B, extra channels can be added for each fluorescent protein signal (determined by CLS)

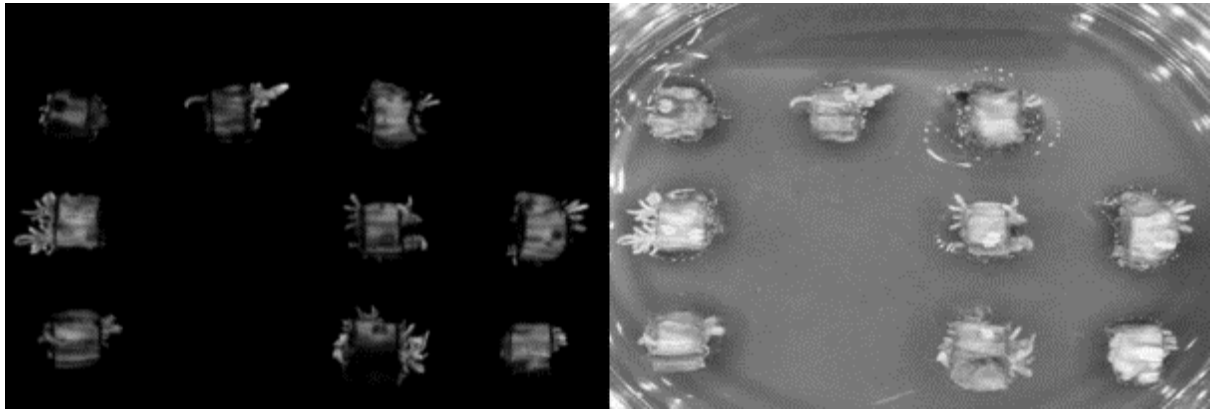
Ongoing work to align RGB and hyperspectral channels

- Differences between hyperspectral, RGB channels:
  - Resolution – necessitates rescaling
  - Camera position – necessitates cropping
- Efforts to apply existing alignment algorithms are underway
- Align “green” channels from both image types:
  - Chlorophyll channel from regression of chlorophyll spectrum hyperspectral images
  - Green channel in RGB images

Example attempt to align images

Hyperspectral

RGB



Summary and next steps

- Machine vision and hyperspectral images can offer greater detail, accuracy, reliability than manual scoring
- Desire to implement fully automated phenotyping based on hyperspectral images for future transformation experiments
- For transformation GWAS (or sooner? Transformation experiments?), integration of hyperspectral and machine vision data to obtain scores of reporter protein signal in specific tissues
- Transformation optimization experiments ongoing:
  - Determining optimum chemical treatments to improve zero-heavy distributions of regeneration phenotypes and aid future DEV experiments
  - Preparing for Agrobacterium strain testing (with and without Vir plasmids)
  - High-throughput screen of DEV plasmids with fully automated phenotyping to begin in October

For material discussed, publications and presentations currently planned

Phenotypes	Outlet	Current status	Next steps	Aim to publish/present
DEV gene paper	Plant Cell, Tissue and Organ Culture? New Phytologist?	Preparing plasmids and plant material for Oct. experiment	High-throughput DEV gene screens, making use of insights from transformation optimization	Spring/Summer 2020? Depends heavily on positive results?
Transformation optimization paper		Experiments underway for Sil-wet, selection, more	Complete phenotyping and statistical analysis	Late 2019 or early 2020
Phenomics paper	Plant Phenomics?	Refining phenomic system, particularly: 1. Integration of machine vision and hyperspectral 2. Deep learning model improvement	Annotation for MV training, then workflow deployment	Late 2019 or early 2020
"Phenomic system for imaging and quantification of in vitro plant regeneration and transformation"	Society for In Vitro Biology 2019		Select treatments and begin GWAS (Winter)	June 2020

Thank you for listening

Scheduled 15 minute break

Oregon State University

Oregon State University

Phenomics III: Machine Vision Analysis Systems

NSF PGRP Advisory Meeting

October 3, 2019

By Jialin Yuan, Damanpreet Kaur, Michael Nagle, Fuxin Li

COLLEGE OF ENGINEERING

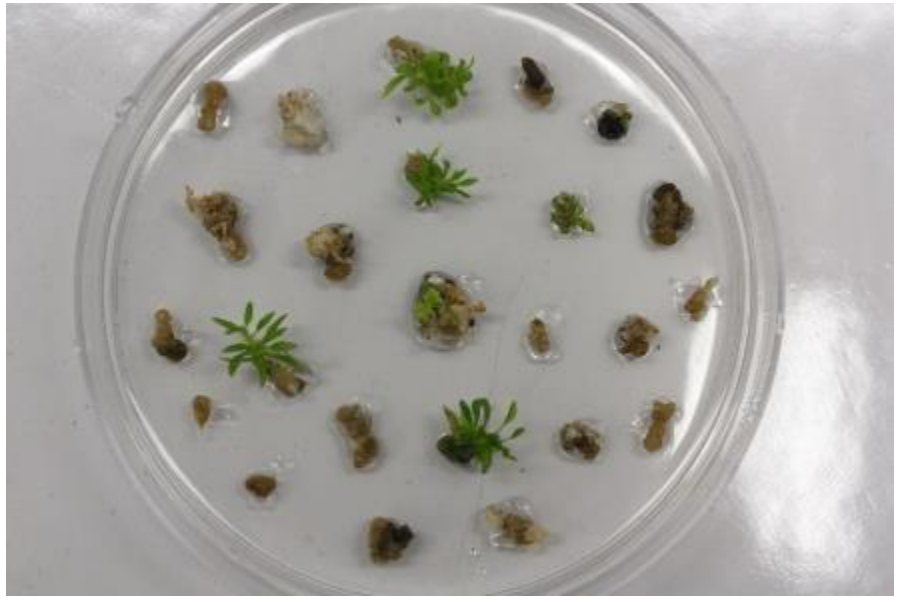
School of Electrical Engineering and Computer Science

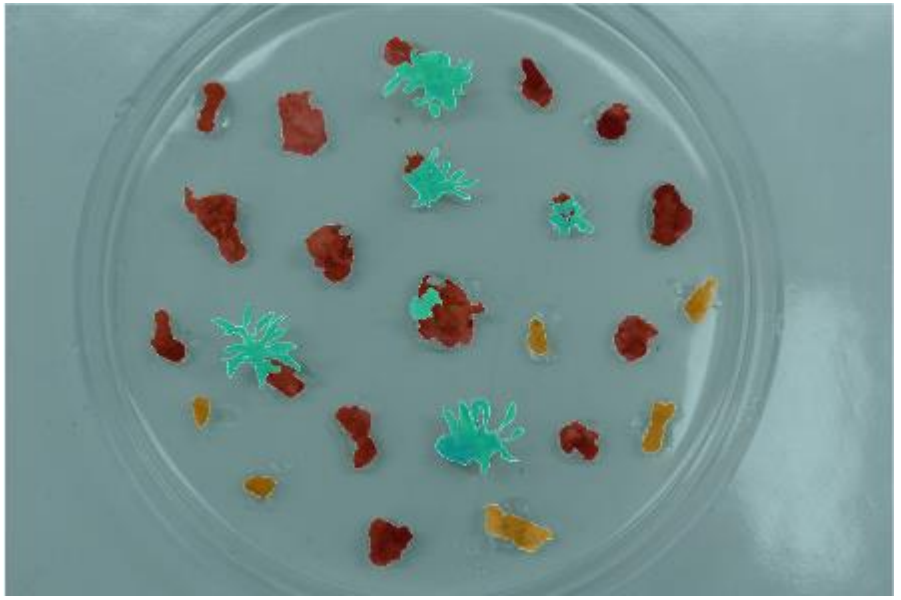
Agenda

- Annotation GUI
- GWAS analysis with Machine Vision (callus / shoot traits)
- Ongoing technical work on segmentation
- Root growing analysis
- Publication plans and future work

Goal of Image Data Analysis (Slide from 2016)

- Recognizing different types of explants (e.g. shoots, roots, etc.), 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”



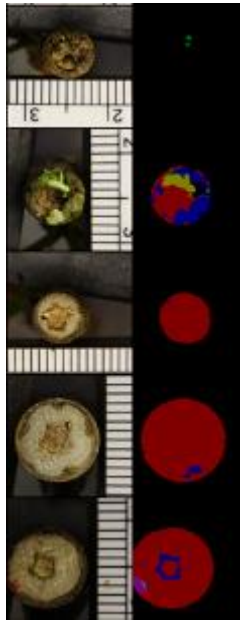


Approach

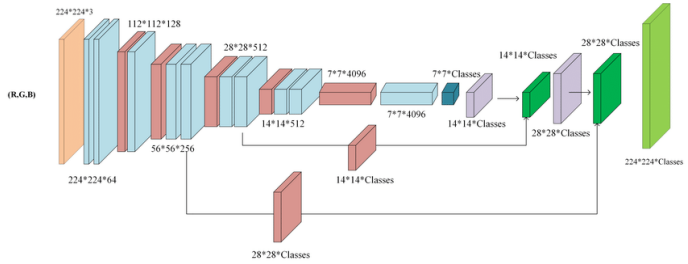
Annotation GUI

Biologists

Dataset of Annotated Images

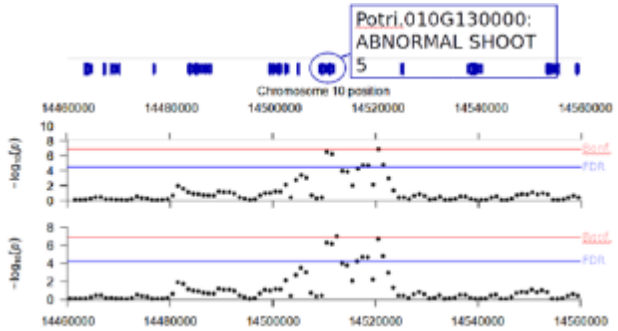


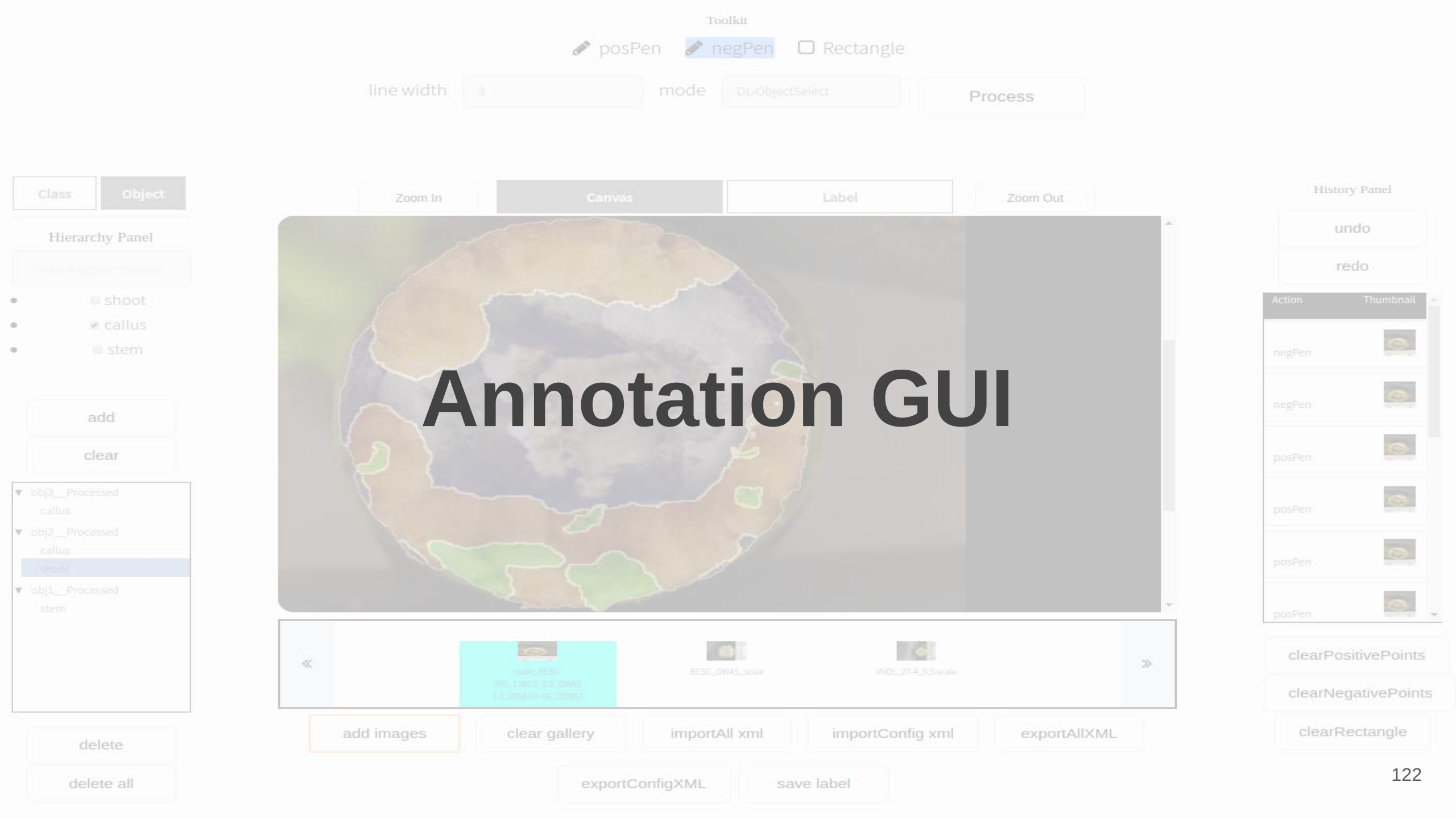
Learn Deep Segmentation Models



Predicted Traits

GWAS Analysis




Annotation GUI

Background: Difficulty in Annotation

- Segmentation-level annotation is difficult
  - Most current approaches use polygons
  - Not easy to draw polygons on plants!

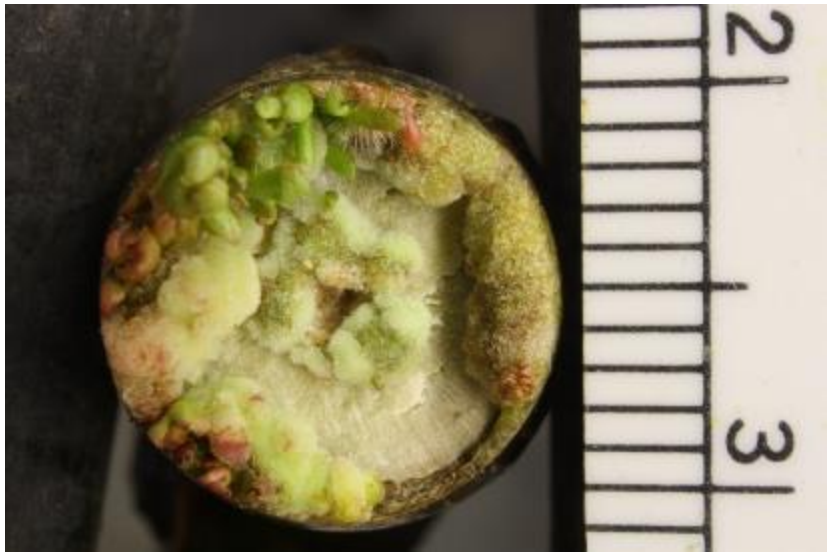
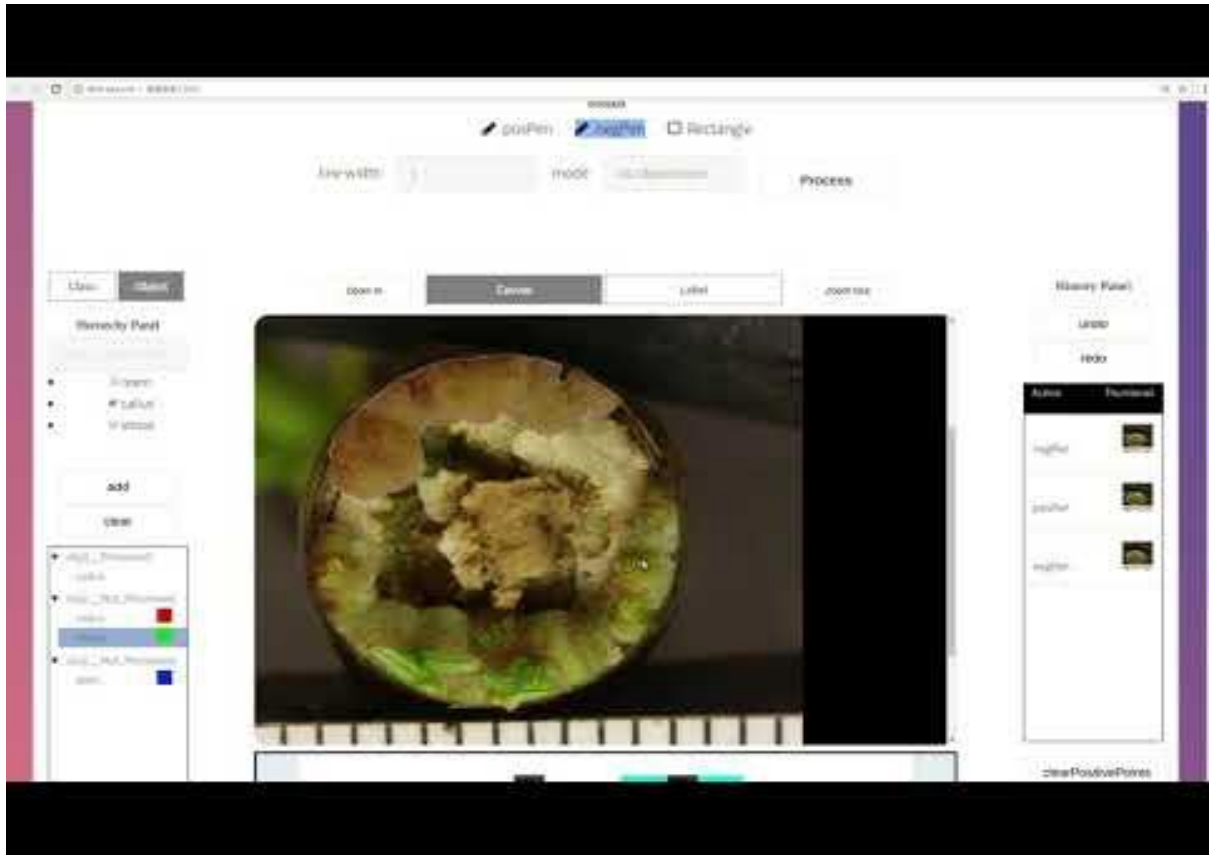


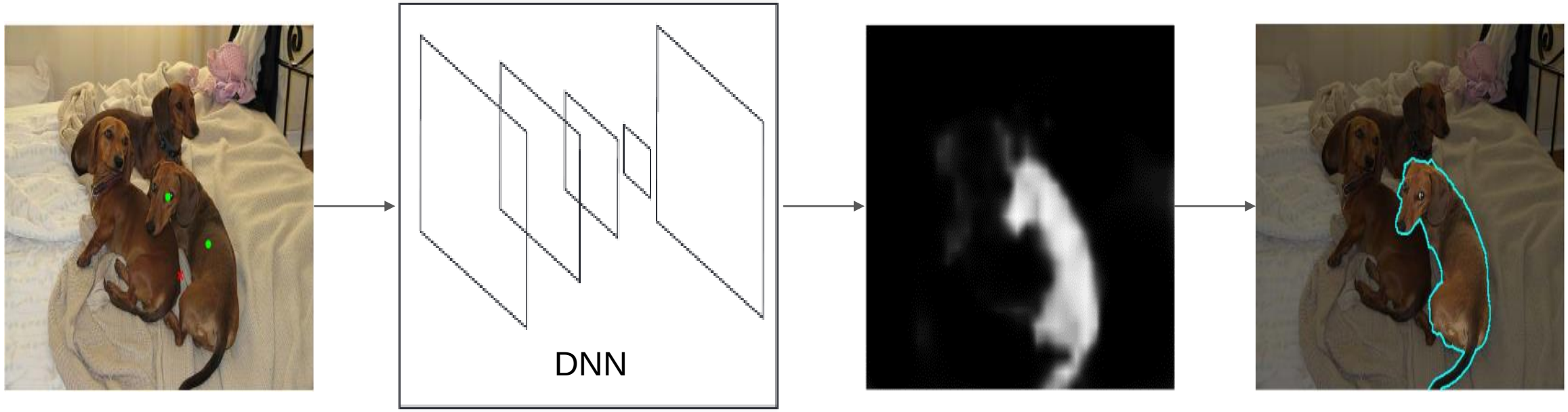
Image Annotation GUI

- Web-based GUI ⇒ No installation, easy to use
- Customizable ⇒ User can specify objects and the properties they have
- Deep interactive object selection[1] ⇒ Good annotation quality and efficient to use



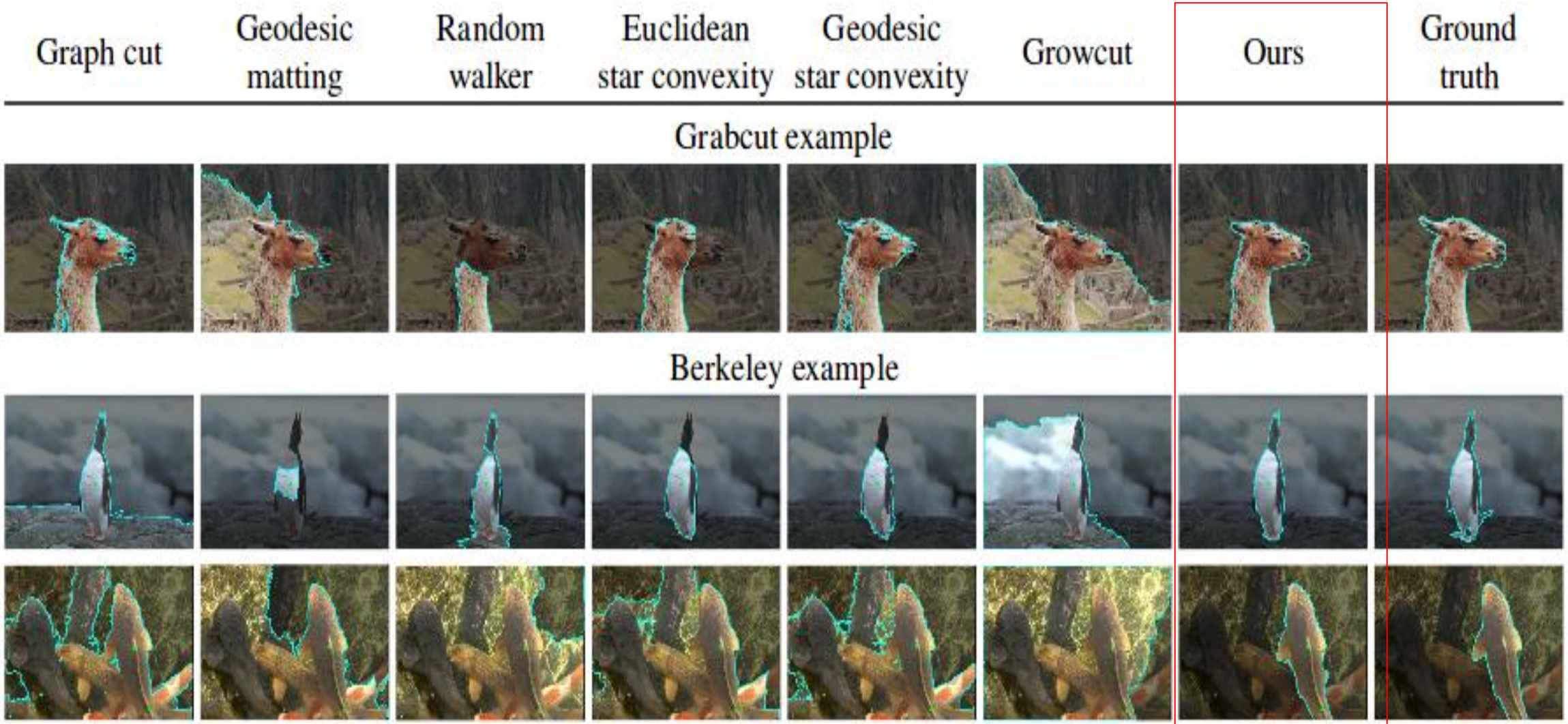
[1] Xu, Ning, et al. "Deep interactive object selection." *CVPR*. 2016.

Deep interactive object selection



Example of selecting an object using the provided user clicks.

Deep interactive object selection



Deep interactive object selection

- The mean **number of clicks** required to achieve a certain accuracy

Segmentation models	Pascal (85% IU)	Grabcut (90% IU)	Berkeley (90% IU)	MS COCO seen categories (85% IU)	MS COCO unseen categories (85% IU)
Graph cut [2]	15.06	11.10	14.33	18.67	17.80
Geodesic matting [1]	14.75	12.44	15.96	17.32	14.86
Random walker [8]	11.37	12.30	14.02	13.91	11.53
Euclidean start convexity [9]	11.79	8.52	12.11	13.90	11.63
Geodesic start convexity [9]	11.73	8.38	12.57	14.37	12.45
Growcut [23]	14.56	16.74	18.25	17.40	17.34
Ours	**6.88**	**6.04**	**8.65**	**8.31**	**7.82**

Annotation Tool Timeline

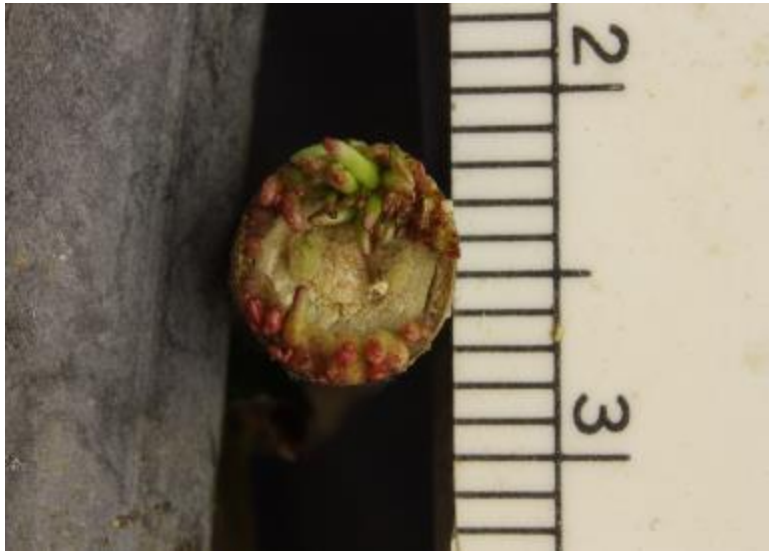
- 2017/1 – 2017/12 Develop the basic annotation tool (V0)
- 2018/1 – 2018/6 Functionality improvements (V1)
  - More than 15 issues fixed, algorithm improvements, V1 is mostly functional
- 2018/6 – 2019/4 Functionality improvements (V2)
  - More than 15 issues fixed, fully functional, used to annotate current dataset
- 2019/5 – now
  - More testing, usability enhancements, preparation for larger-scale deployment

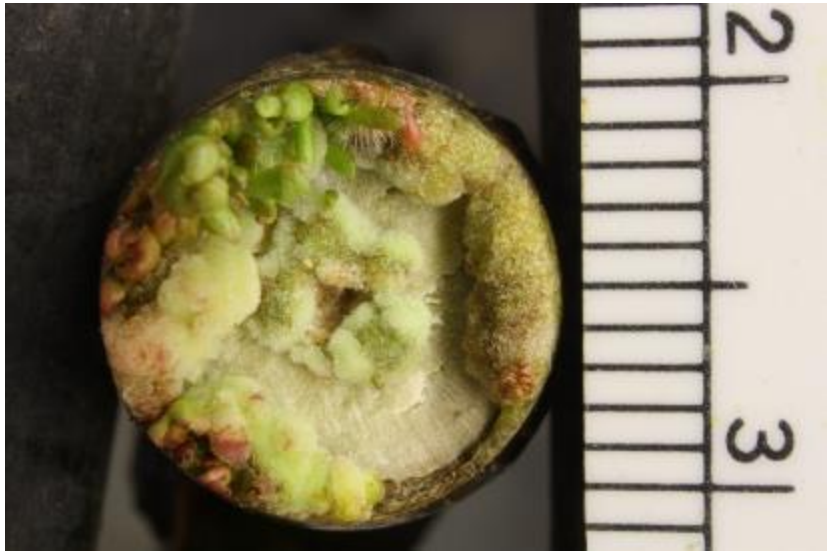
GWAS analysis with Machine Vision (callus / shoot traits)

GWAS Analysis

How well does the plant regenerate?

- Depends on the growth of callus and shoots



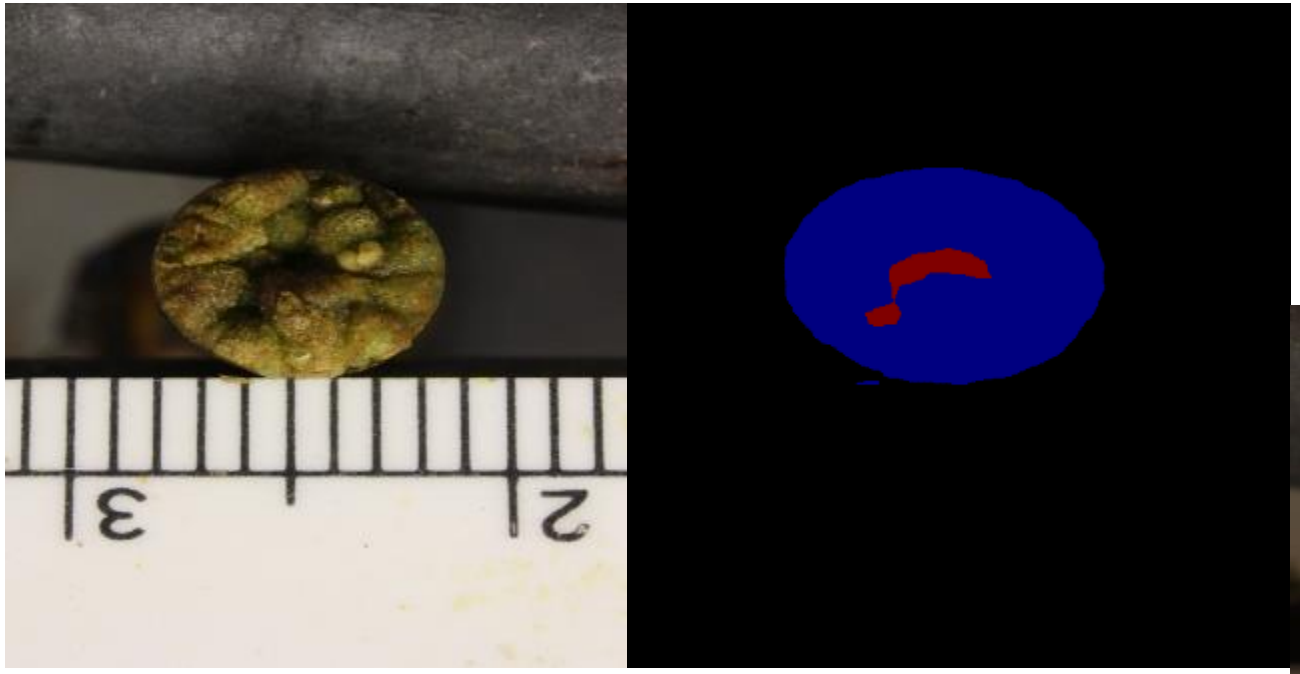


Plant Regeneration Experiment

- Annotated 136 images (120 for training, rest for testing) using the annotation tool
- Learned several deep models to predict callus/stem/shoot areas
    - VGG, PSPNet, DeepLab v3+ tried in the process
    - Settled on DeepLab v3+

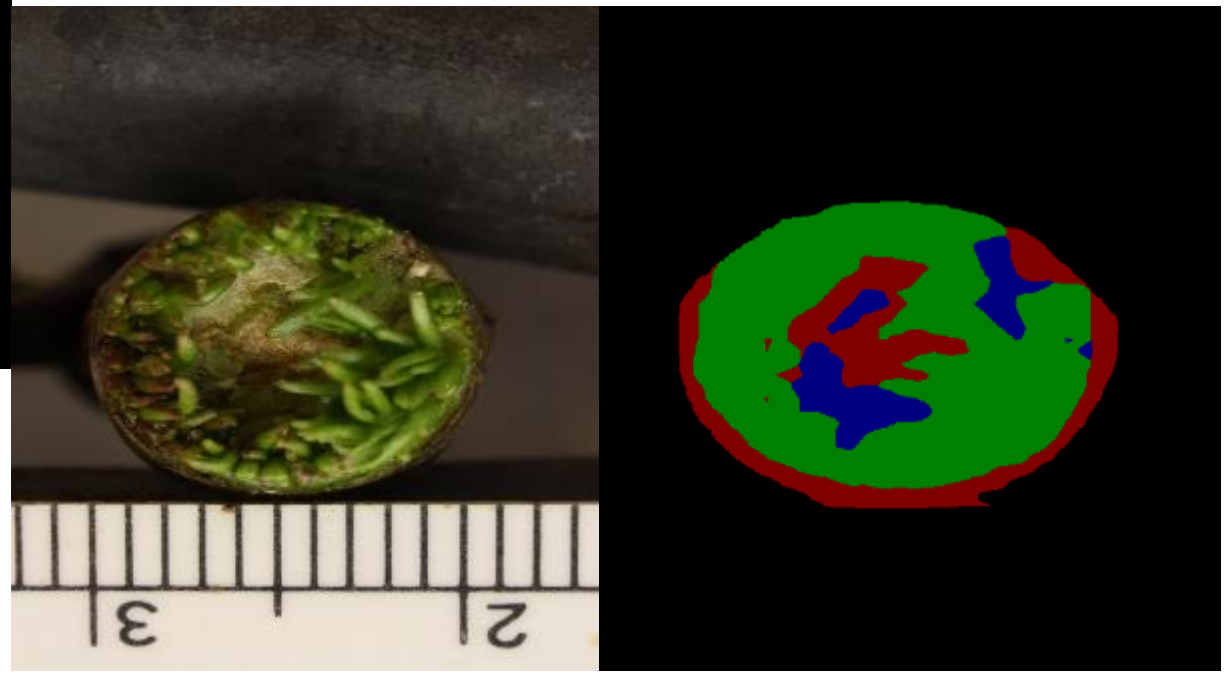
Callus/Stem/Shoot Segmentation

(Deeplab v3+ [2] Model)



Prediction
Category Label

Prediction
Category Label



[2] Chen, Liang-Chieh, et al. "Encoder-decoder with atrous separable convolution for semantic image segmentation." *(ECCV)*. 2018.

Callus/Stem/Shoot Segmentation

(Deeplab v3+ [2] Model)

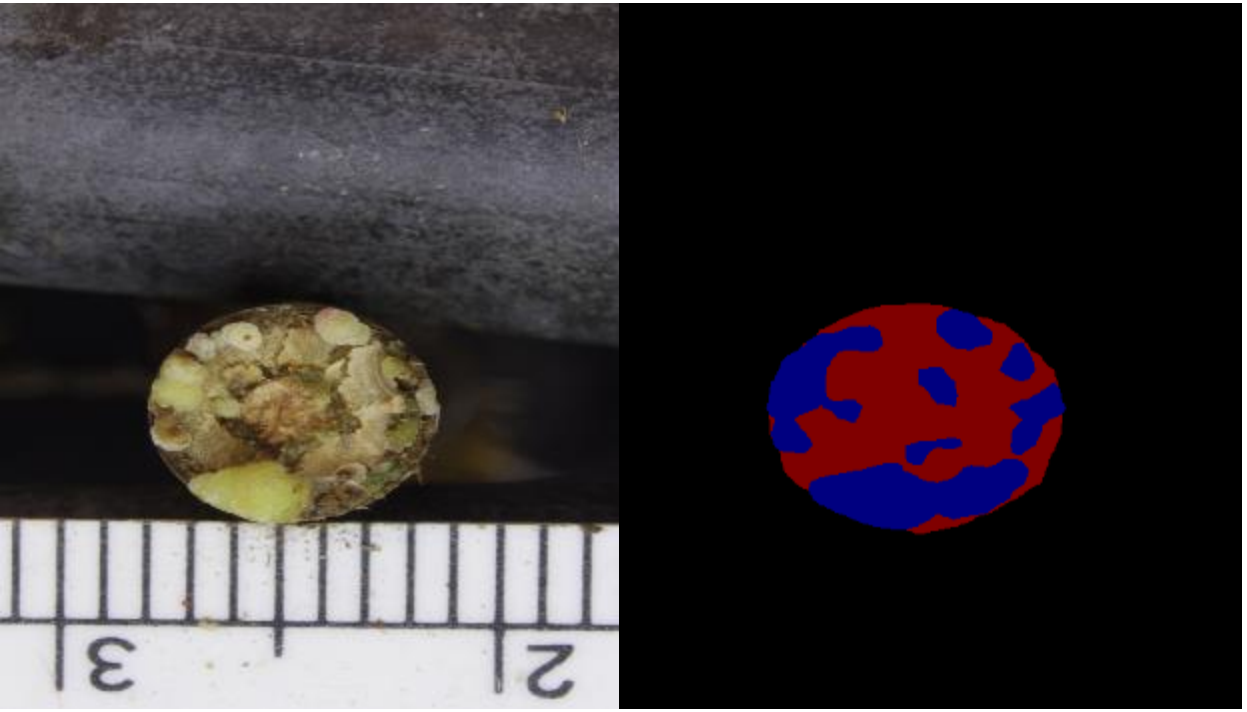
	Background	**Stem**	**Callus**	**Shoot**	**Mean IoU**
Training	99.16%	90.37%	90.12%	87.48%	91.78%
Validation	99.17%	73.87%	77.60%	76.47%	81.78%

Model trained using a 80-20 training-test dataset split.

- Training dataset - 102 images
- Testing dataset - 25 images

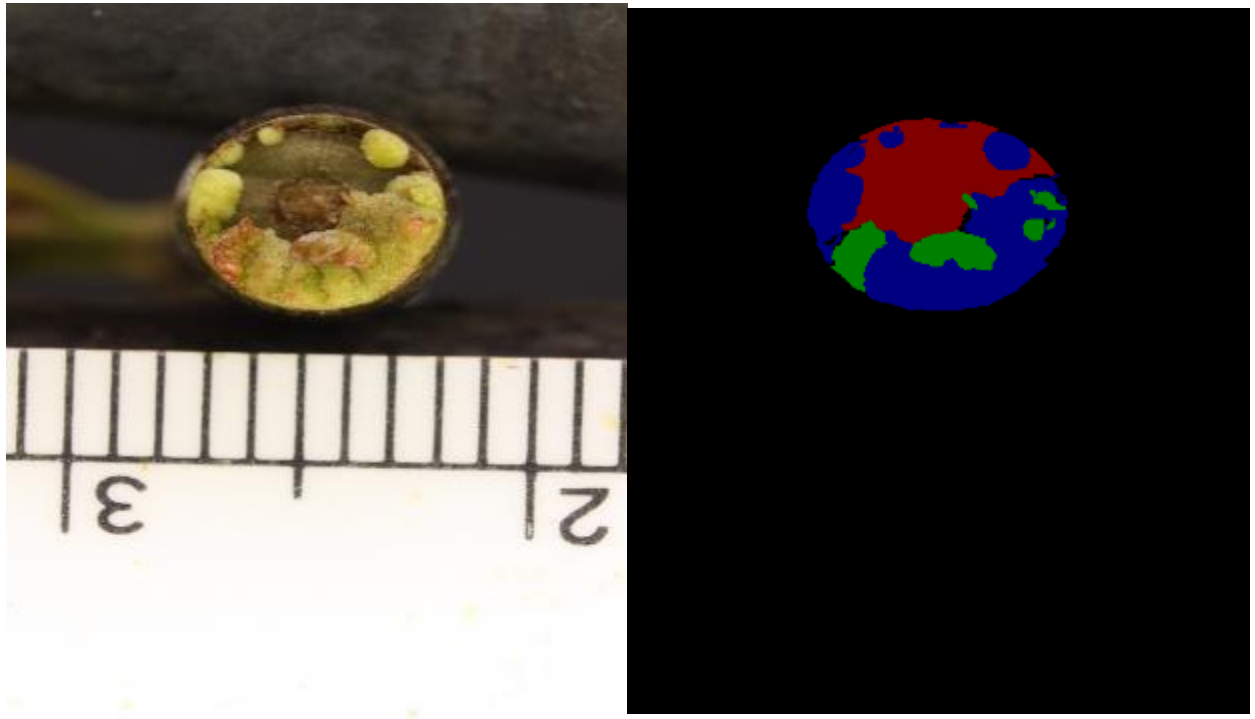
[2] Chen, Liang-Chieh, et al. "Encoder-decoder with atrous separable convolution for semantic image segmentation." *(ECCV)*. 2018.

Deeplab results



Class Name	Area (%)	No. of Connected components
callus	0.43	7
shoot	0	0

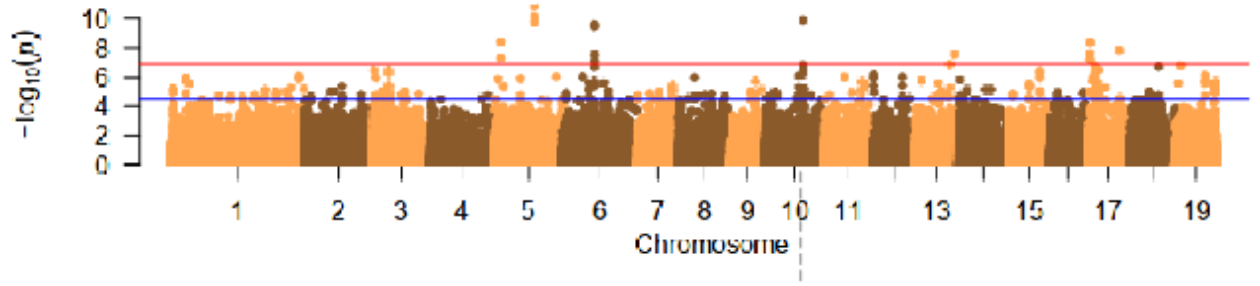
Class Name	Area (%)	No. of Connected components
callus	0.49	5
shoot	0.16	4



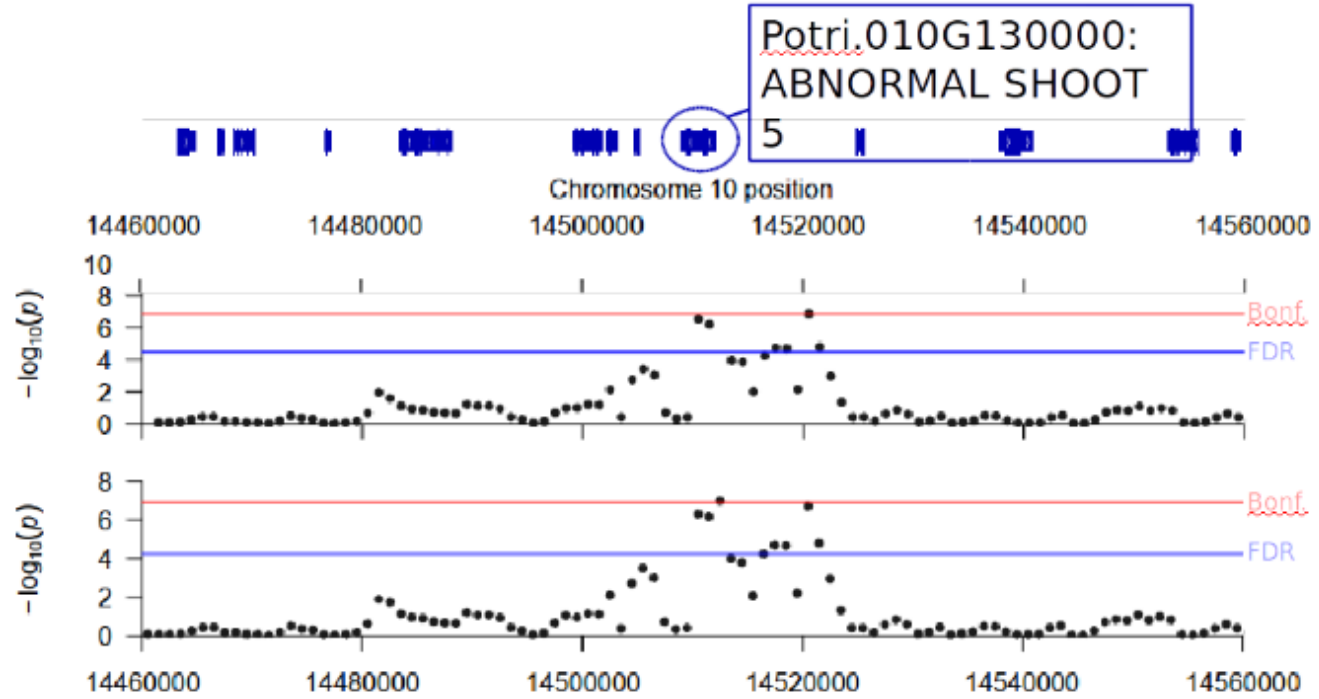
Required Annotations

- Estimate the performance on different training dataset sizes
- Less overfitting by increasing the training dataset size
- When does the performance saturate on test dataset?

GWAS analysis: SKAT test

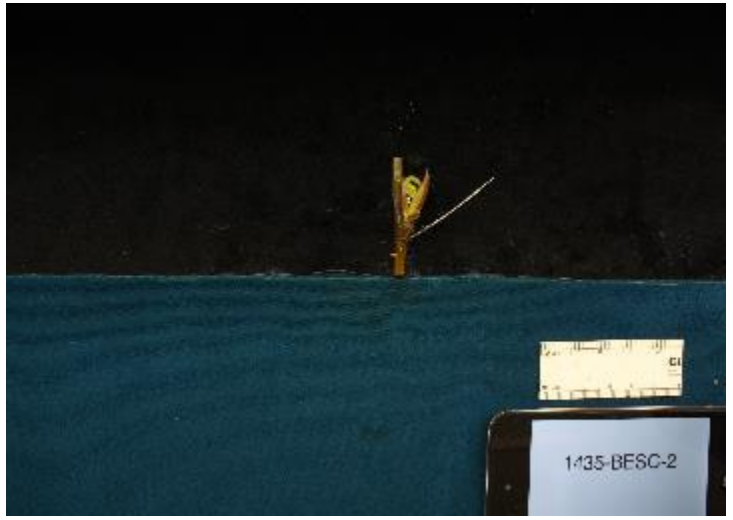


View zoomed to chromosome 10 subsection, aligned to gene track



Association between a known shoot regulator in *Arabidopsis* with computed shoot area

Root Growth Analysis



Week 2

Week 3

Week 4

Interesting traits from machine vision

Leaf size

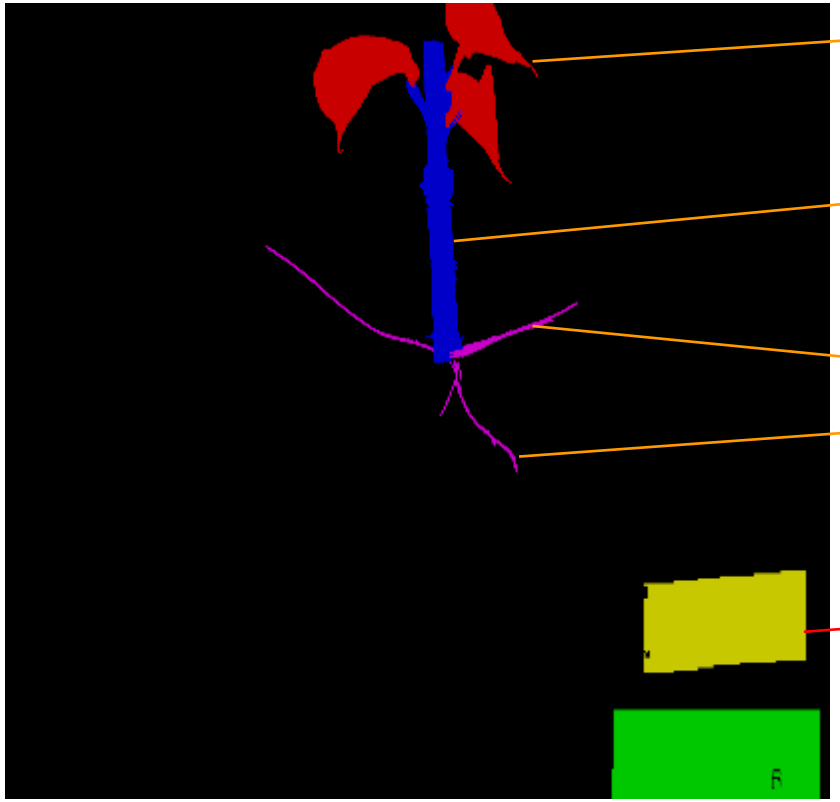
Stem diameter

No. of roots
Roots' length
Roots' type: bazel or lateral

Pixel size in 'cm'

Interesting traits from machine vision

- Machine vision solution: **segmentation**



Leaf size: area in pixel count

Stem diameter: average width

No. of roots: connections of root to stem
Roots' length: count pixel along root
Roots' type: local background color analysis

Pixel size in 'cm': average width of ruler in pixel

Segmentation on the image

- First level segmentation: unsupervised

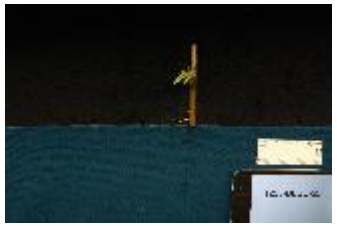
  Background | plant | ruler | label

- Second level segmentation: unsupervised

  Background / leaf / stem / root

Segmentation on root growth images

- No annotation is used!
  - Segment ‘easy examples’ using prior-knowledge
    - Color
    - Location
    - Shape

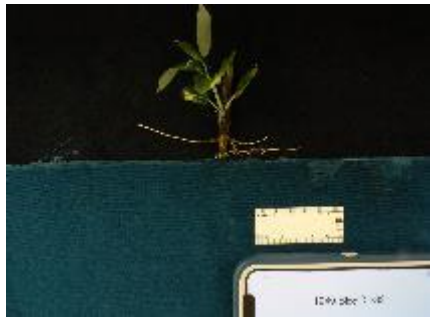


  - Train a segmentation network with collected ‘easy examples’
    - Inconsistent background
    - Roots in different color / shape
    - Leaves in different color
    - Data augmentation from easy examples
      - Rotation
      - Flip
      - Color manipulation

Qualitative results

Leaf size (cm^2)	10.601
Stem diameter (cm)	0.492
# of roots	3
Root length (cm)	[4.264, 3.211, 2.913]
Root type	[Lateral, Lateral, Basal]

Qualitative results



Leaf size (cm^2)	11.377
Stem diameter (cm)	0.553
# of roots	3
Root length (cm)	[5.321, 2.783, 5.175]
Root type	[Lateral, Lateral, Lateral]

Ongoing Technical Work

Interactive Semantic Segmentation

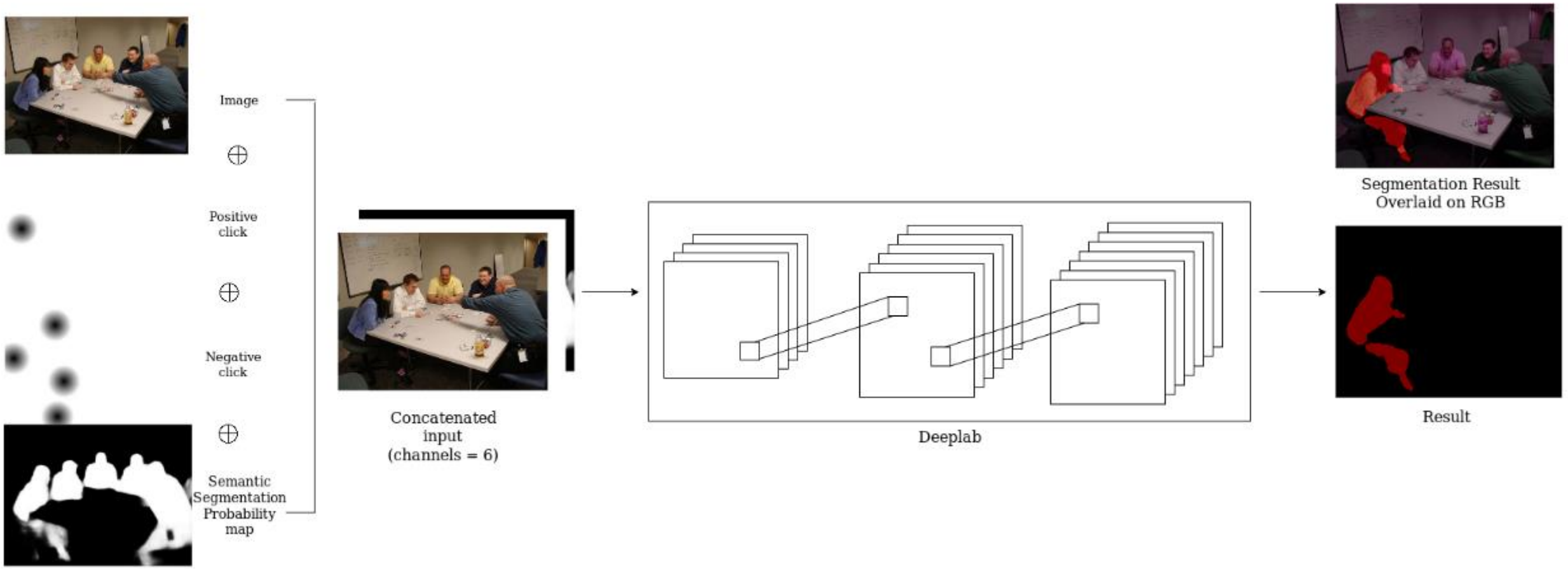
Instance Segmentation

Idea: To understand which object is present in the image at pixel-level.

Semantic-Guided Interactive Segmentation

- Use known semantic segmentation results to guide interactive segmentation
- Semantic segmentation already has good performance, this should make future annotations easier
- Approach: Incorporate semantic prediction results into the deep network for interactive segmentation
- Progress: Good progress on PASCAL VOC dataset, needs integration into the system

Semantic-Guided Interactive Segmentation



Semantic-guided Interactive Segmentation Results

Results:

	Mean IoU (in %)	**Boundary F-measure**
Baseline Interactive Segmentation	73.90	31.20
New Algorithm (using semantic results as a prior)	83.10	74.10

Semantic Segmentation Results

Ground truth image

Semantic result using Deeplab

Probability map

Interactive Segmentation Result

Semantic-guided Interactive Segmentation results

Ground truth

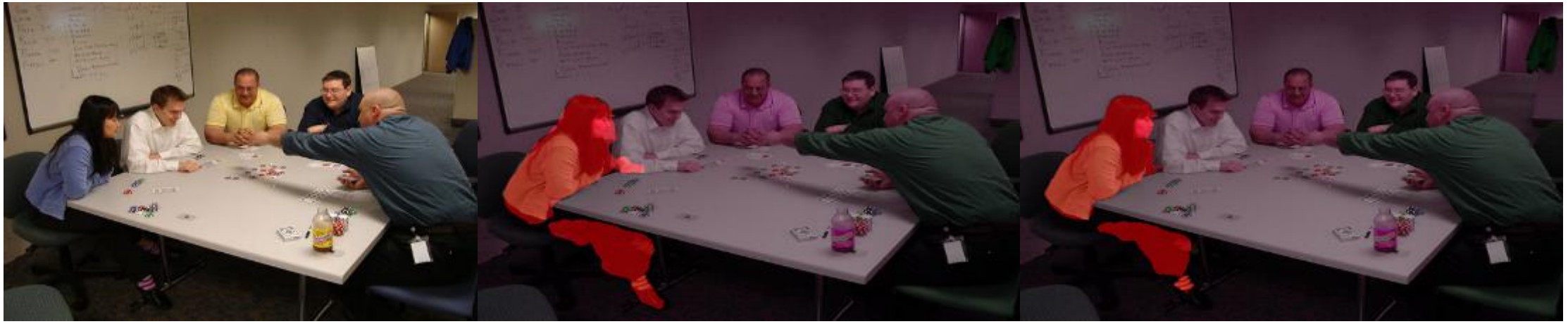
Interactive with Semantic

Mean IoU: 94.6
Boundary F-measure: 77.5

Interactive with Semantic+CRF

Mean IoU: 95
Boundary F-measure: 78.3

Qualitative Results



Ground truth
Semantic+CRF

Interactive with Semantic

Mean IoU: 86.2
Boundary F-measure:65.7

Interactive with

Mean IoU: 86.9
Boundary F-measure:73.4

Qualitative Results

Ground truth

Semantic+CRF

Interactive with Semantic

Mean IoU: 83.4
Boundary F-measure: 66.6

Interactive with

Mean IoU: 82.1
Boundary F-measure: 69.3

Qualitative Results

Ground truth
with Semantic +CRF

Interactive with Semantic

Mean IoU: 91.1
Boundary F-measure: 96

Interactive

Mean IoU: 94.4
Boundary F-measure:97.5

Semantic-guided Interactive Segmentation Results

Incorporating Semantic-guided interactive algorithm into the Annotation system:

- The new algorithm will help in improving the -
  - Efficiency
  - Performance of the annotation system

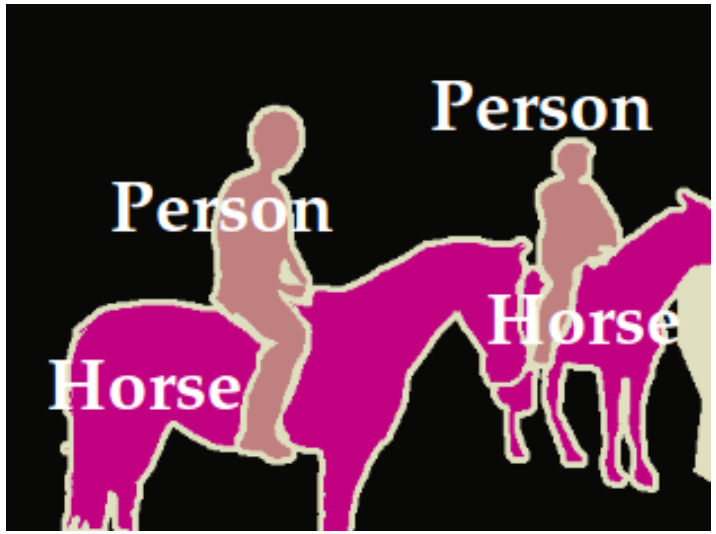
- Reduce the interactive effort on the user part
  - Useful for plant scientists as it requires minimal user input while annotation

Instance Segmentation

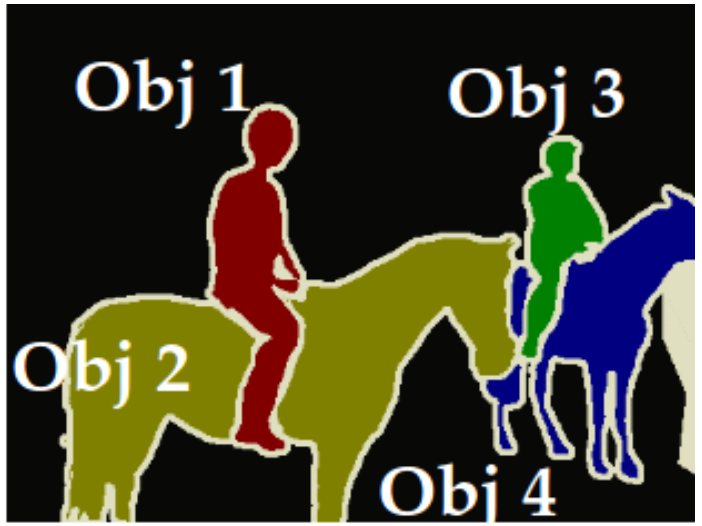
- What is Instance Segmentation?

Image

Category Label



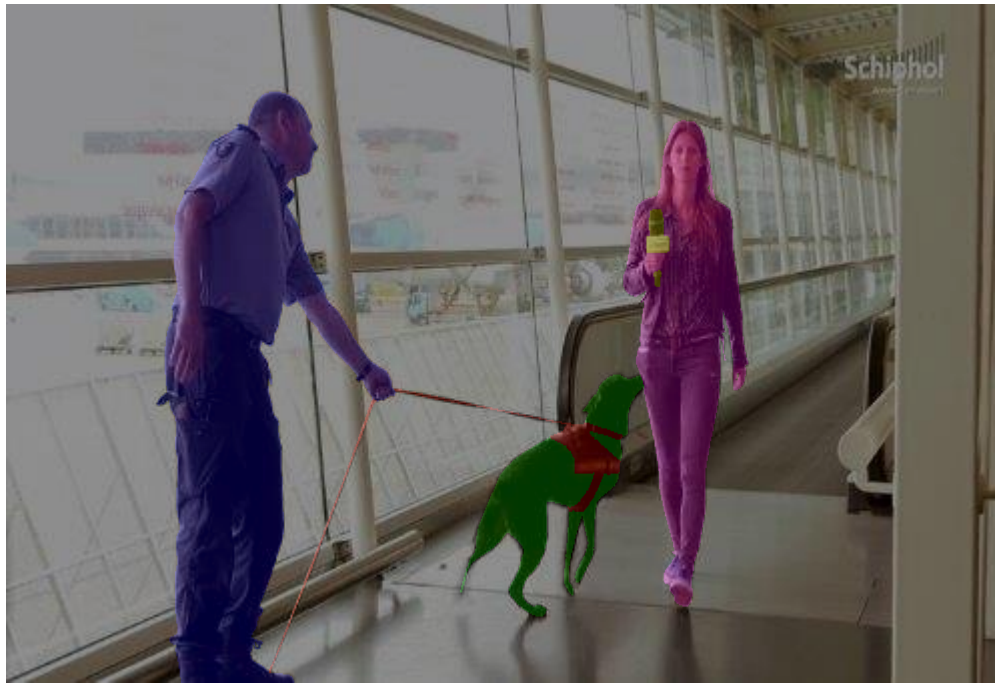
Object Label



From Proposals to No Proposals

- One stage segmentation approaches
  - Significant amount of redundant computation
  - Bottom-up process is difficult to become real-time

Visual of top 50 proposals

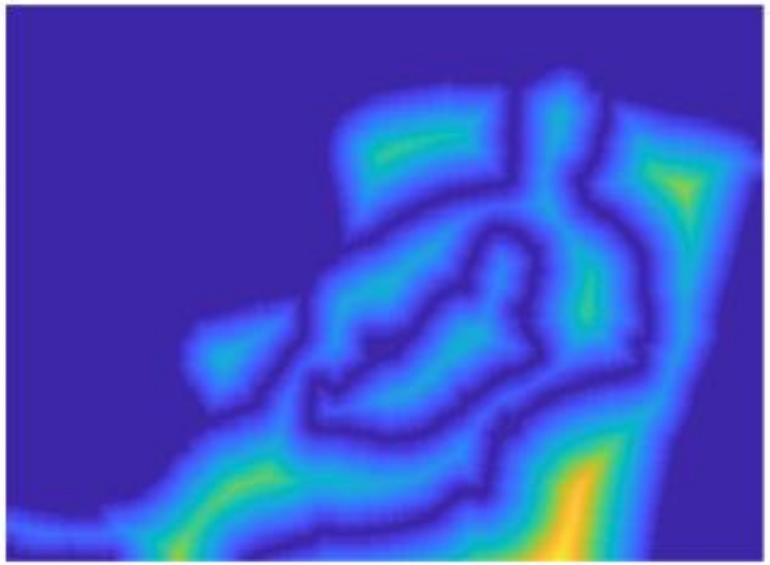


Ground truth annotation

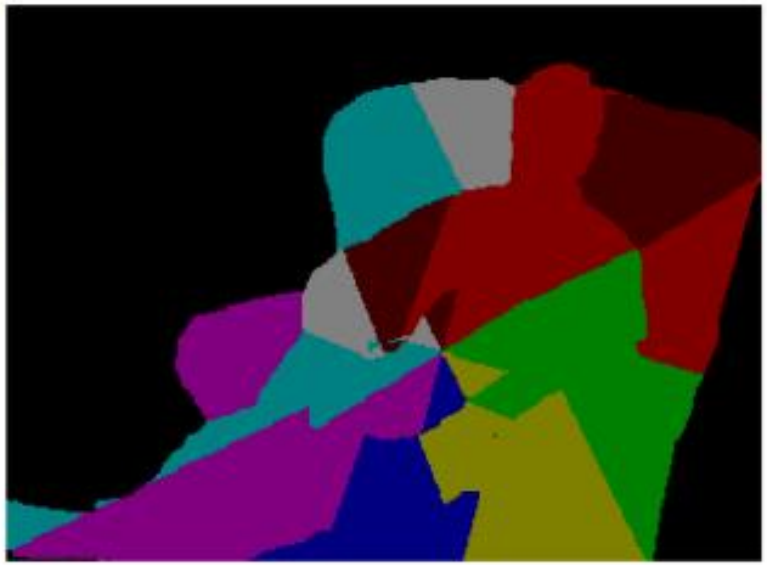
From Proposals to No Proposals

- Predict a surrogate objective
  - Post-processing from the prediction
  - FCN can directly predict surrogate without proposals
  - Those surrogates have issues

Bai & Urtasun 2017



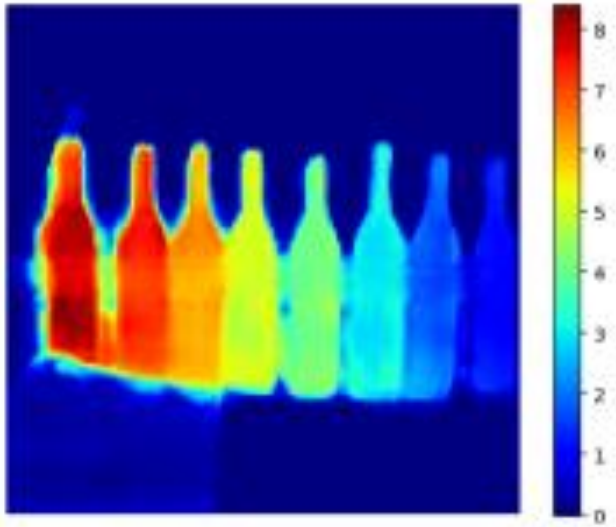
Uhrig et al. 2016



From Proposals to No Proposals

- We proposed to directly predict the instance label
  - Relax the labels to be continuous-valued
  - Directly predict **real-valued** instance labels as a deep network

Input Image



Real Valued Labels

Current Results on PASCAL

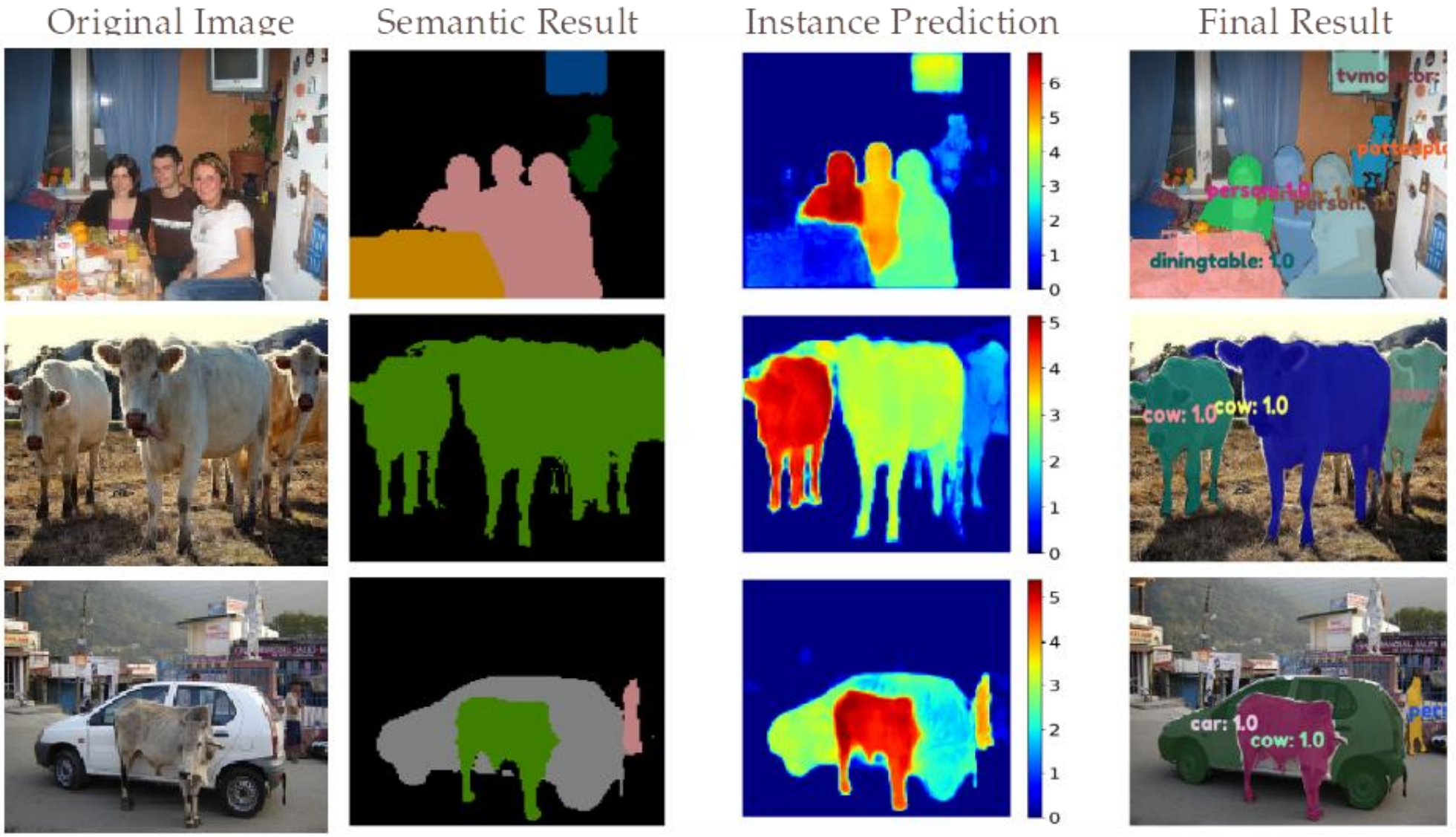
Method	mAP^r					AP^r_{avg}
	0.5	0.6	0.7	0.8	0.9	
SGN[21]	61.4	55.9	49.9	42.1	26.9	47.2
DIN[1]	62.1	53.3	41.5	-	-	-
FCIS[20]	65.7	-	52.1	-	-	-
Embedding[18]*	64.5	-	-	-	-	-
DVIS	63.75	**58.62**	**53.75**	**46.78**	**31.01**	**50.79**

Table 2. AP^r result on the PASCAL VOC 2012 *val.* set. See

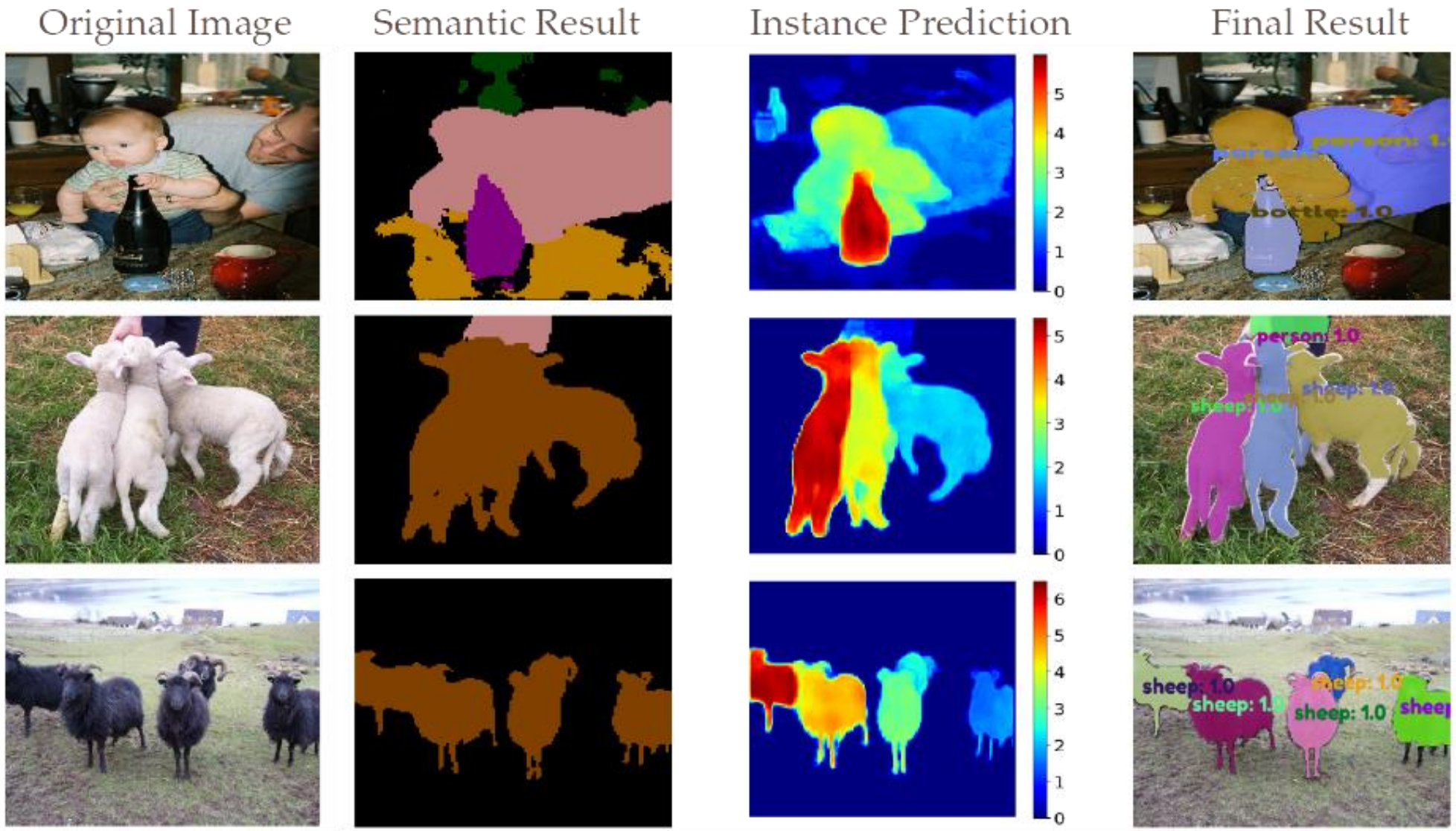
	$IoU_{small\ objects}$
DeepLab-v3+	0.57
Embedding[3]-gt	11.10
DVIS-gt	**33.25**

Table 2. Foreground segmenation on small objects (in size $\leq$ 500 pixels) on PASCAL VOC 2012 *val.* set

Visual Results

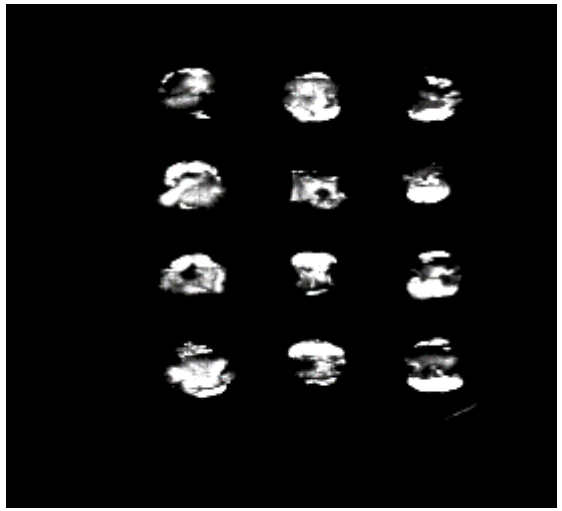


Visual Results

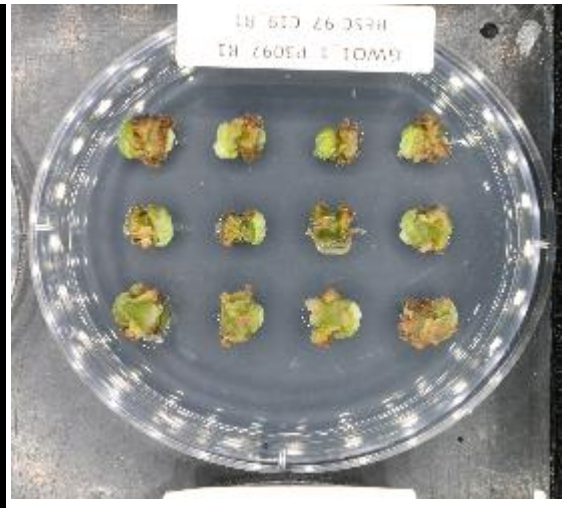


Other work

Aligning hyperspectral image with RGB image

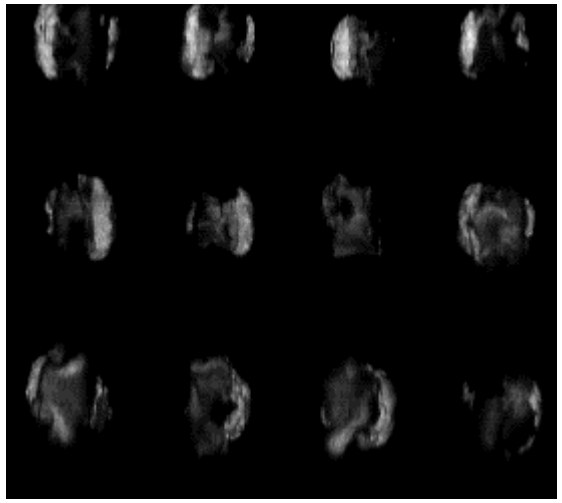


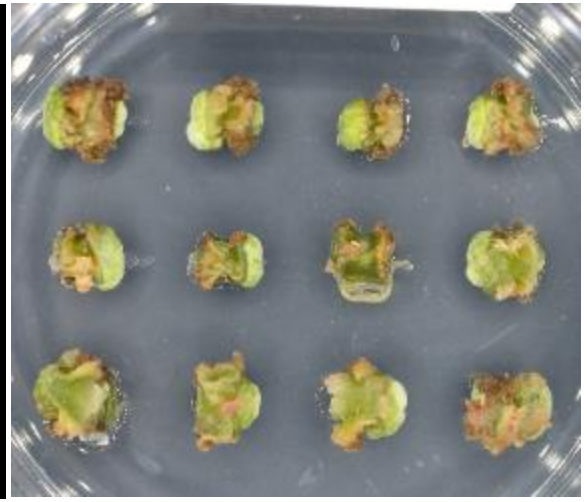
Hyperspectral Image



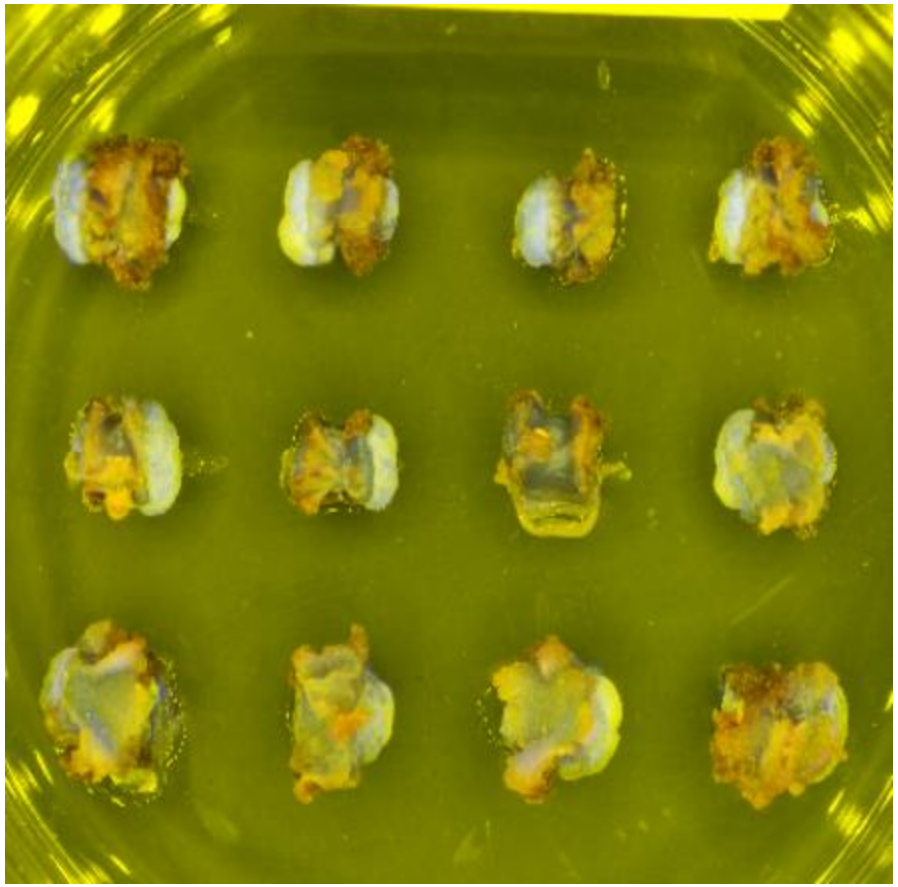
RGB Image

Original





Preprocessed

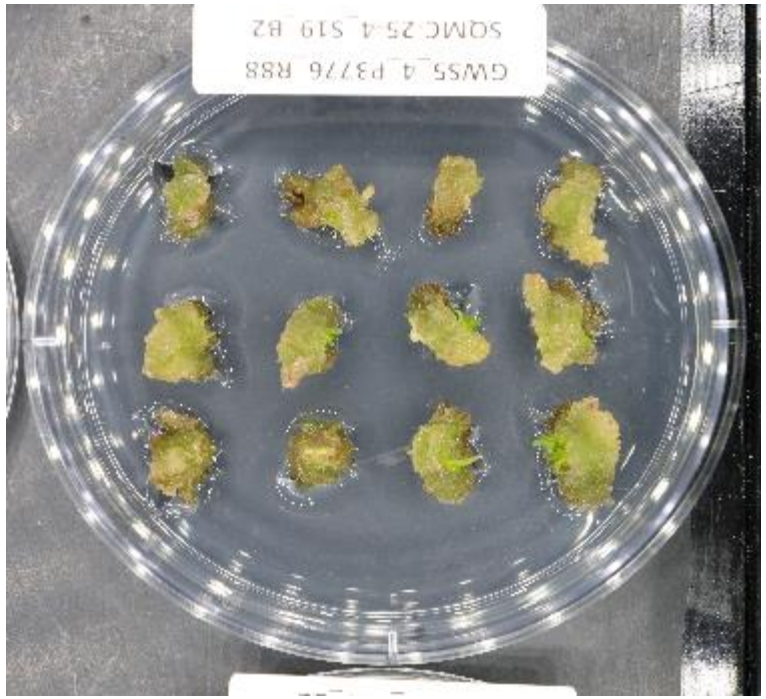


Aligned Image

Aligning hyperspectral image with RGB image

- Will be useful for -
  - Aligning the spectral matrices of explants with the classification images
  - Calculation of green fluorescent protein (GFP) in the tissues
  - Detecting shoot growth based on Chlorophyll in the hyperspectral images

Analyzing explants traits



Grid No	Center	Region area
1	(29.36, 40.68)	0.018
2	(34.87, 62.07)	0.024
3	(22.70, 55.55)	0.013
4	(42.98, 52.95)	0.026
5	(53.88, 34.80)	0.024
6	(51.44, 38.72)	0.022
7	(45.26, 51.28)	0.020
8	(39.31, 40.08)	0.026
9	(52.38, 37.00)	0.024
10	(50.91, 47.38)	0.019
11	(51.89, 47.62)	0.025
12	(50.01, 59.52)	0.036

Publication plan

1. **Instance Segmentation (submitted once to ICCV 19)**

   Conference: CVPR

   Timeline: November '19

2. **Semantic-guided interactive segmentation**

   Conference: CVPR

   Timeline: November '19

3. **Annotation System**

   Journal submission (a plant phenomics journal)

Future Work

1. Annotation System

By 2019

a. Further work on improving the GUI, adding shortcuts and making it more user-friendly

b. Improve documentation and user guide on the annotation GUI, release it to the public

c. To be solved: GPU resources?

Spring 2020

Incorporate the semantic-guided interactive segmentation algorithm for the annotation system

2. Hyperspectral Image System

By 2019

Align FP hyperspectral matrices with RGB image for all grid types

Future Work: Automatic Trait Analysis

- One lesson learned from the entire effort is that training networks is not that straightforward
  - Deep models require significant amount of parameter tuning (a dedicated person tuning for 1-3 weeks, depending on experience)
  - For the goal: fully automate the trait analysis (including model training), several improvements needed
    - Automatic connection to a cloud engine with GPU resources
    - AutoML for tuning the parameters
    - A fee model to accommodate computational costs and software engineering work to setup
  - Additional funding probably needed to achieve that goal
    - Currently starting a Capstone project

Future Work: Hyperspectral Imaging

- We realize that it may not be easy to annotate hyperspectral images via segmentation
  - Some proteins are too simple (e.g. GFP) where partial linear regression or PCA followed by simple thresholding is sufficient
  - Others are too complicated and scattered in high-dimensional hyperspectral data, making it hard to label
  - Maybe necessary to utilize longitudinal analysis to obtain labels (e.g. plant growth after several weeks)
  - How to best integrate hyperspectral data and automatic trait analysis is an unresolved problem

Thank you!

Please enjoy our catered Lunch 12:15-1:00pm

Coded by:

Jialin Yuan

Damanpreet Kaur

Nihar Doshi

Zheng Zhou

Ali Behnoudfar

With input from:

Michael Nagle
Ekaterina Peremyslova
Ritesh Mewalal
Cathleen Ma
Anna Magnuson
Yuan Jiang
Steve Strauss
Fuxin Li

We thank the NSF support from the Plant Genome Research Program on grant IOS-1546900

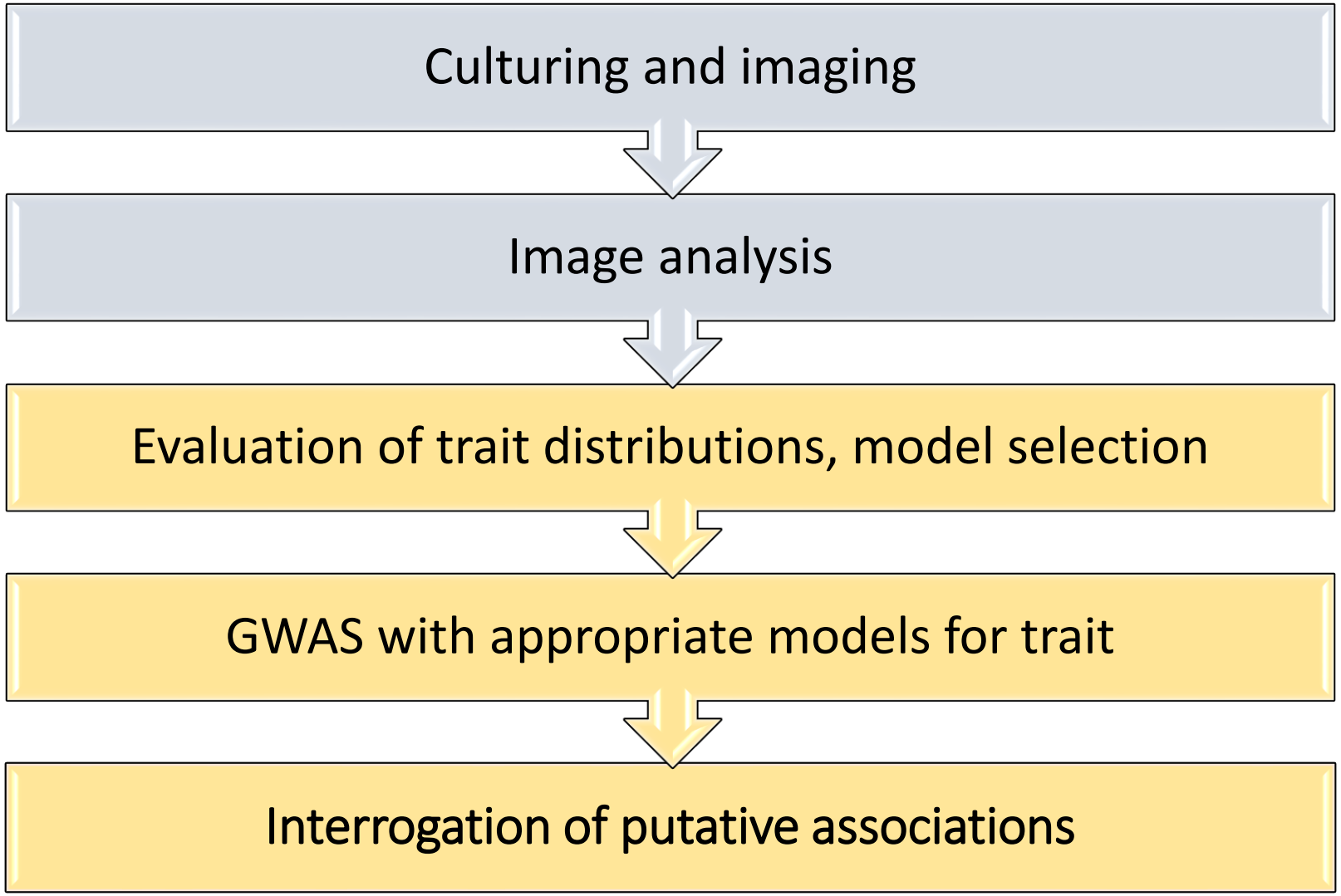
GWAS pipeline and results

Michael Nagle

NSF PGRP advisory meeting

Oct. 3, 2019

From phenotyping to GWAS and beyond

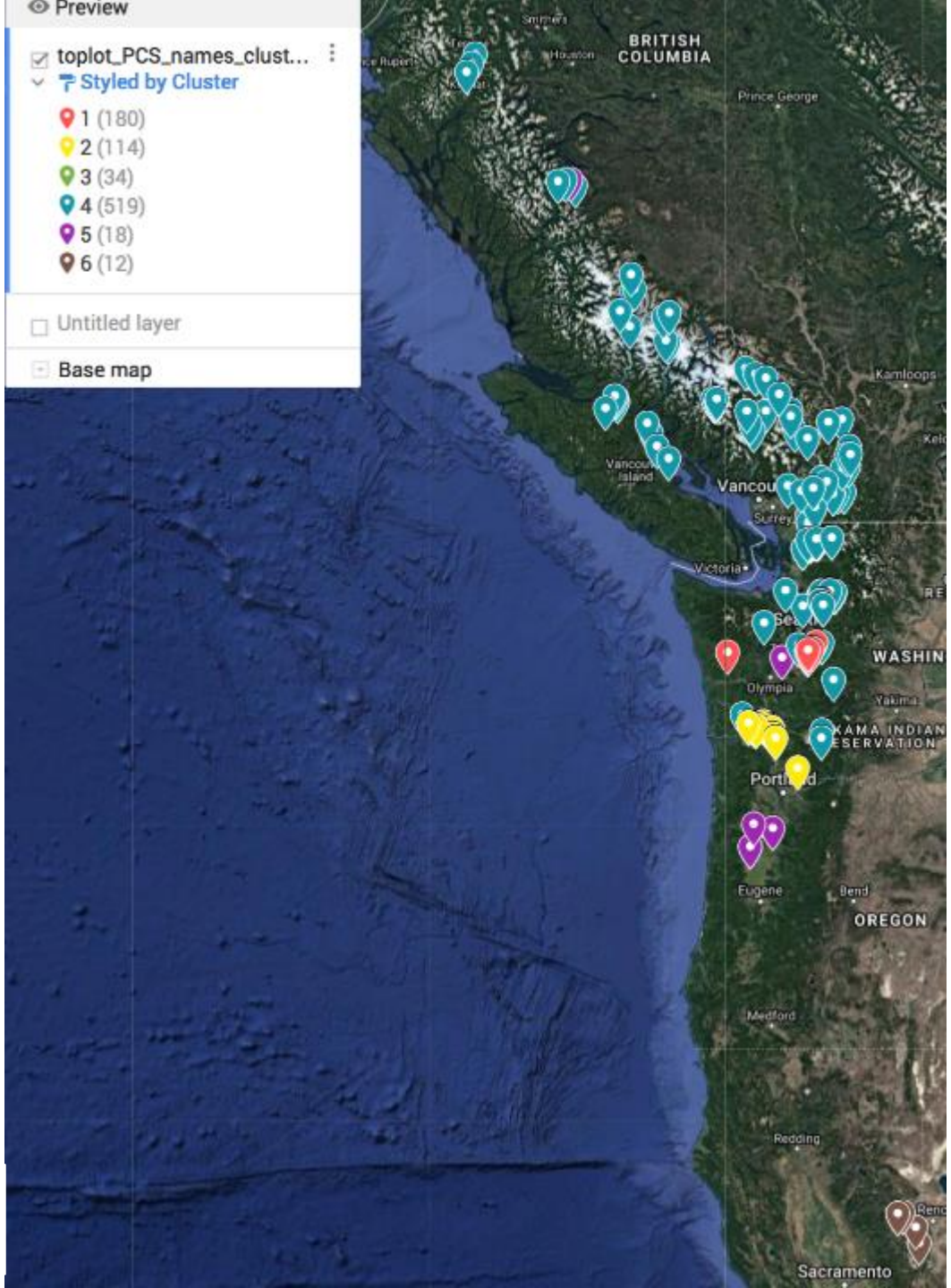


Overview of GWAS and post-GWAS methods

- Appropriate modeling tool depends on distribution of phenotype data (and transformation, if any)
- Resampling: Calculate p-values with true null distribution, rather than depending on approximation to a common distribution
- Desire to resample efficiently over large SNP set motivates use of:
  - GWAS methods combining SNPs
  - High-performance computing
- After GWAS: Is a role for genes implicated by GWAS also supported by evidence from literature, transcriptome, interactome, mutant studies?

Genomic resources for *P. trichocarpa* open doors for genetic discovery

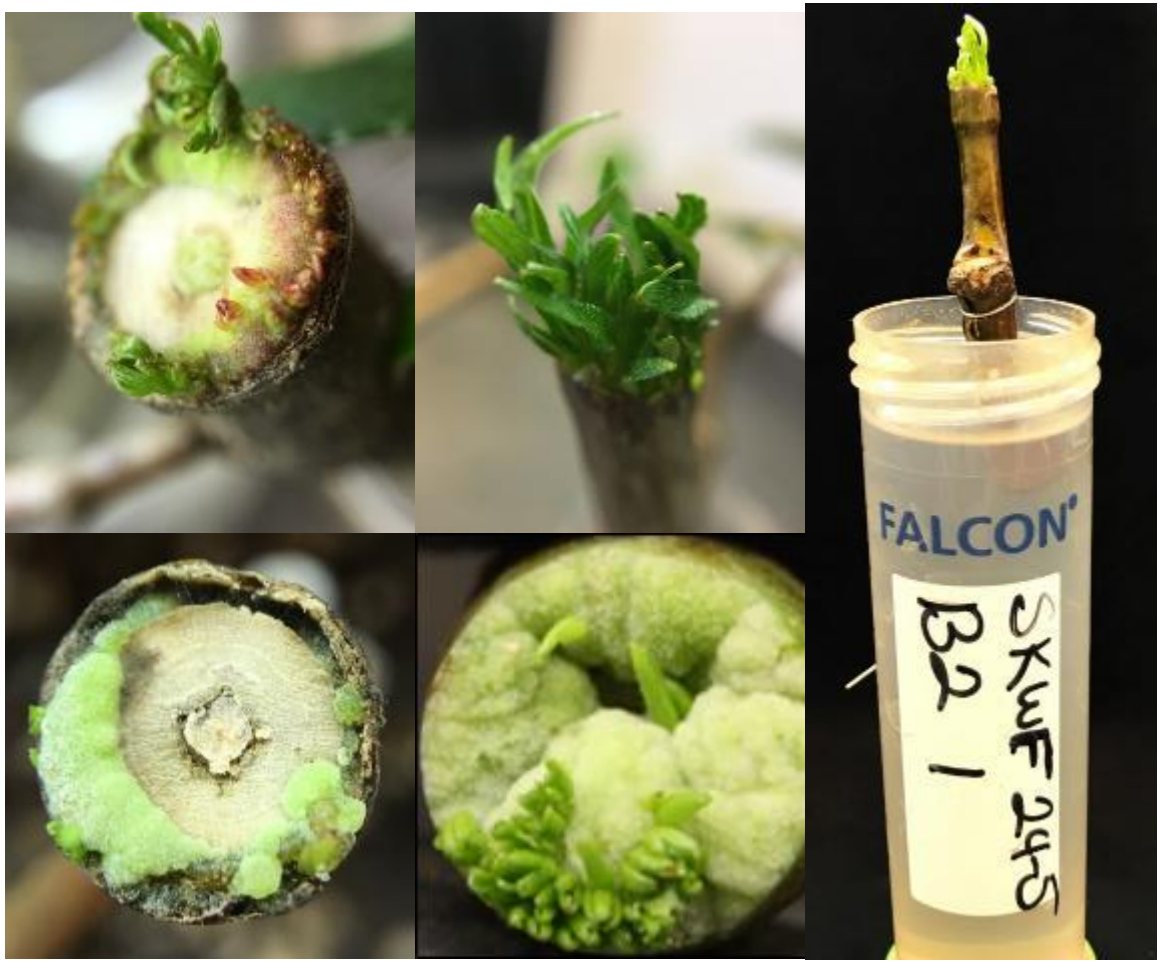
- Current SNP set (released 2016) from Oak Ridge National Laboratory's Bioenergy Research Center
  - 882 genotypes
  - Diversity from California to British Columbia
  - ~28M SNPs
  - ~40x coverage
  - Single reference genotype (Nisqually-1)
- To come:
  additional genotypes and pan-genome



[Interactive Google Earth map of GWAS population, PC clusters]

Ongoing GWAS pipeline testing and refinement using stem regeneration GWAS data

- Project includes GWAS of *in vitro* regeneration, stem regeneration, rooting, transformation and more traits
- Stem regeneration:
  - Wound gives rise to callus, shoot
  - Cytokinin (TDZ) on stem tip encourages regeneration

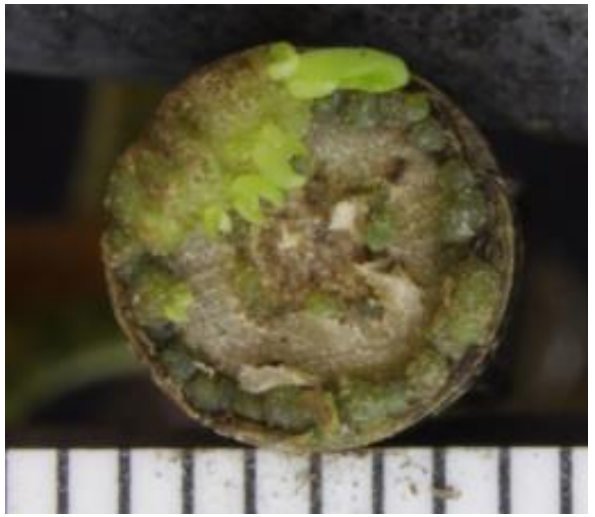


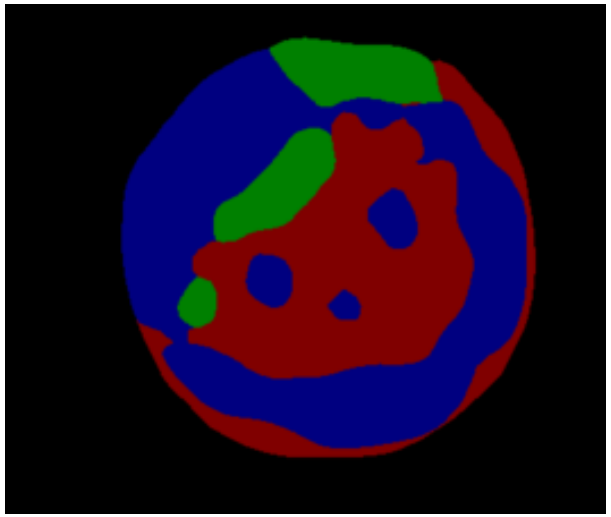
Plant images segmented by machine vision to provide statistics for use in GWAS

Machine vision segmentation of images by tissue class

Calculations of area statistics for each tissue class

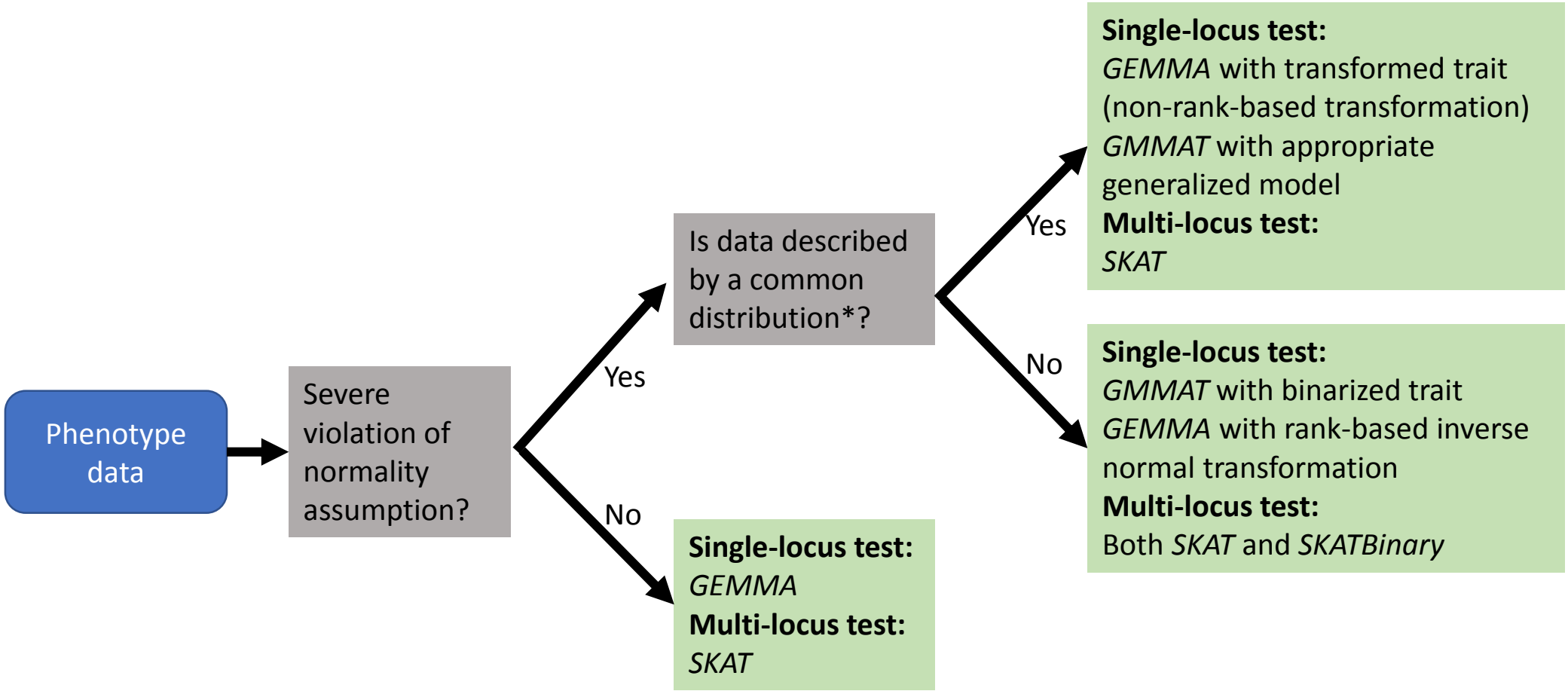
Into GWAS





Tissue class	Percentage of total area
stem	45%
callus	43%
shoot	12%

Current GWAS workflow



*with support in generalized modeling tools in R

We have attempted/performed analyses using these ten GWAS tools

Tool	Type	Notes
PLINK	General linear model	Model phenotype as function of SNPs and PCs
TASSEL5	Mixed linear model	GUI, not written for large SNP sets
GENESIS		Assumes variance of SNP effect coefficient is the same for every SNP
EMMAX		
GEMMA		Precise, efficient, has multivariate option
GMMAT	Generalized mixed linear model	Can build models for certain non-gaussian distributions
SKAT	MLM with kernelized SNP-sets	Tests user-defined SNP groups
Farm-CPU	Alternating mixed/unmixed model	Trouble with large SNP sets
BOLT-LMM	Bayesian mixed linear model	Adds no SNPs to model since heritability calculations require larger, more homogenous population
MLMM		
FaST-LMM	Mixed linear model	Fast GWAS method that is well-established and exact

Key:

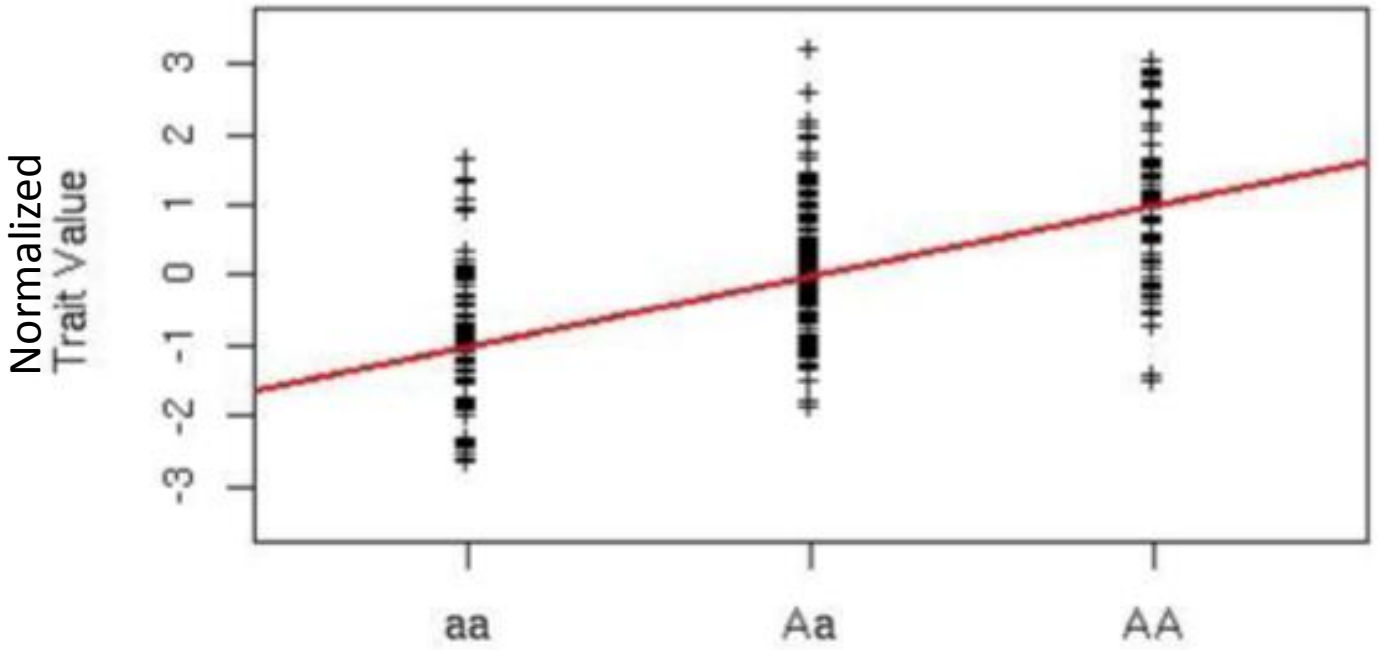
Producing results for full SNP data
Problems due to data input

Linear models in GWAS:
Is there an association between allele and trait?

$$y_i = \mu + G_i\beta + \varepsilon$$

G_i : genotype variant
(0, 1 or 2 copies of alt. allele)

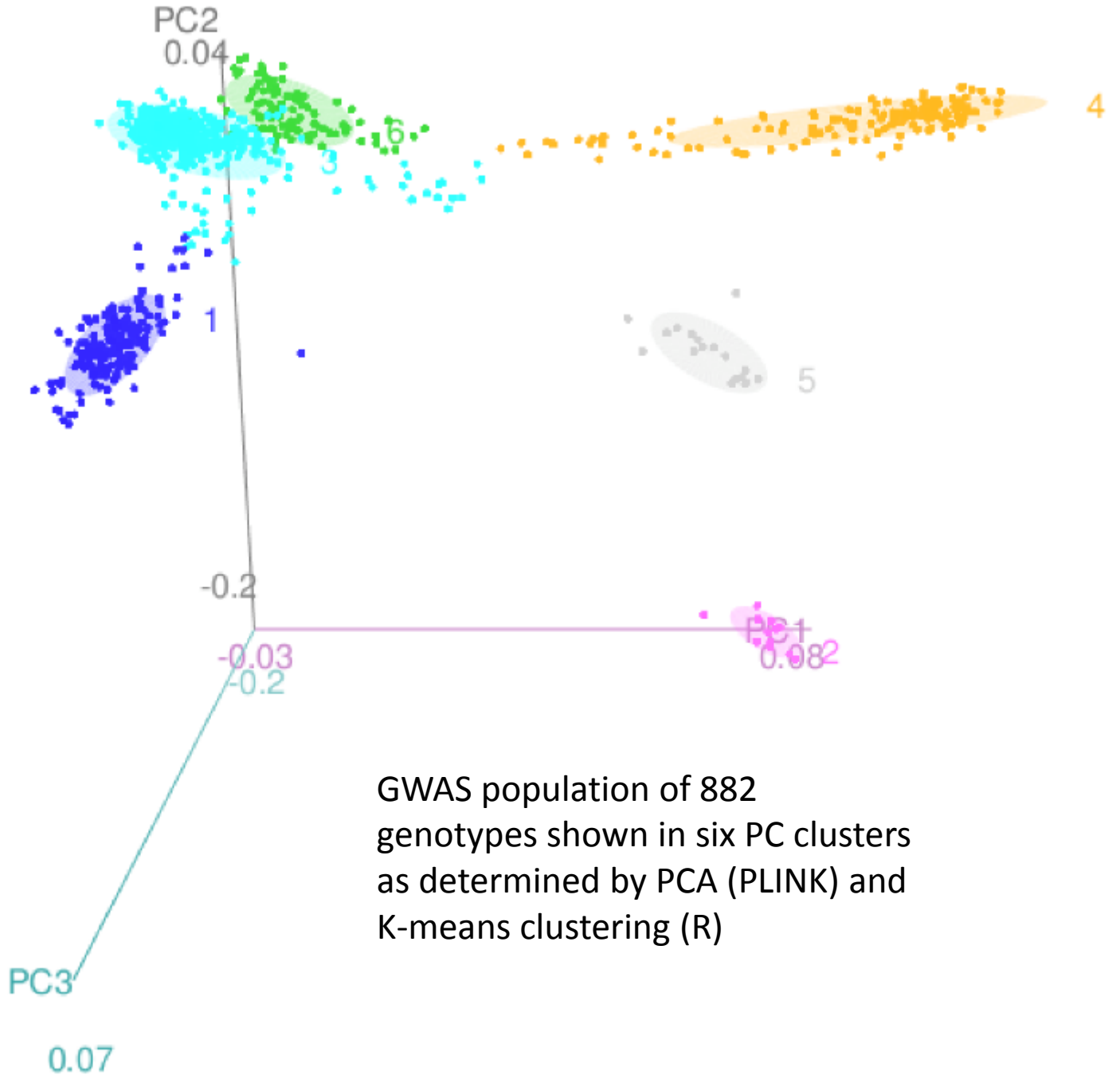
Association between trait and # copies of alternative allele



graphics from UWISG GWAS workshop Aug '18

Note: Dominance models are occasionally used in GWAS

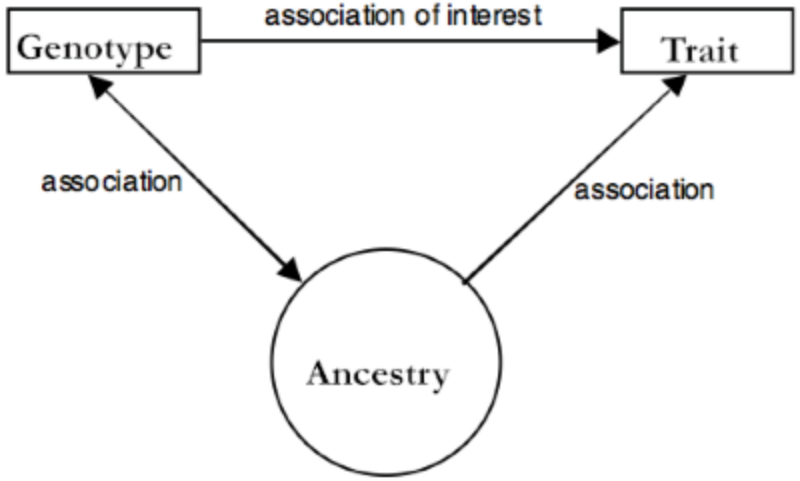
Six clusters of closely related genotypes across three principal components (PCs)



GWAS population of 882 genotypes shown in six PC clusters as determined by PCA (PLINK) and K-means clustering (R)

Controlling for population stratification

- Why an association between trait and SNP? Causal association, or confounding factor of population stratification?
- Approaches to control for stratification:
  A. Multiply p-values by a relatedness coefficient
  B. Split GWAS population into subpopulations
  C. Include population structure in model
     i. Principal components
     ii. Kinship matrix (in mixed model)

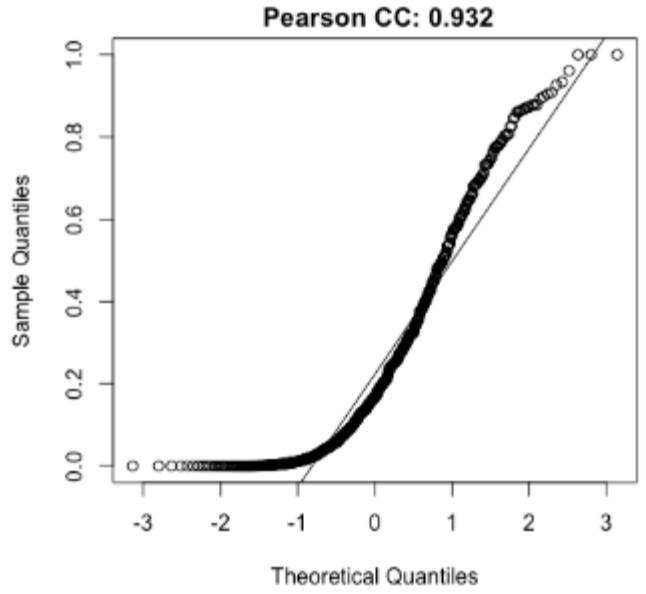


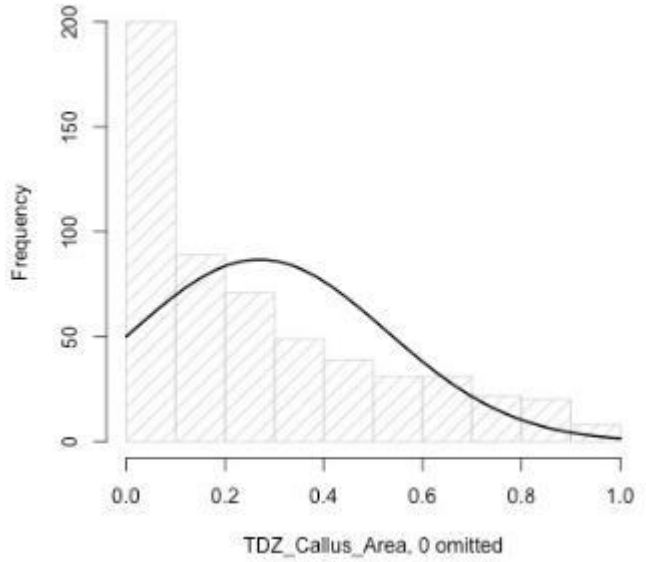
UW Institute for Statistical Genomics

- Overcorrection in highly stratified populations with rapid linkage disequilibrium decay? Alternatives?

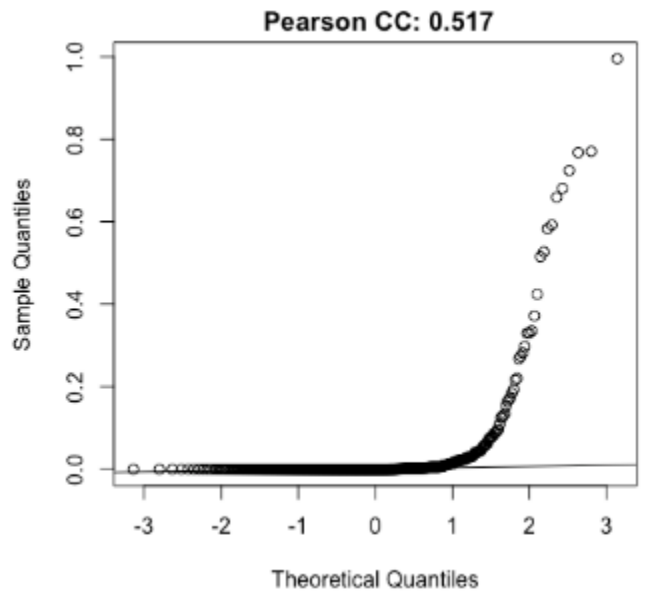
Adherence of traits to normality assumption?

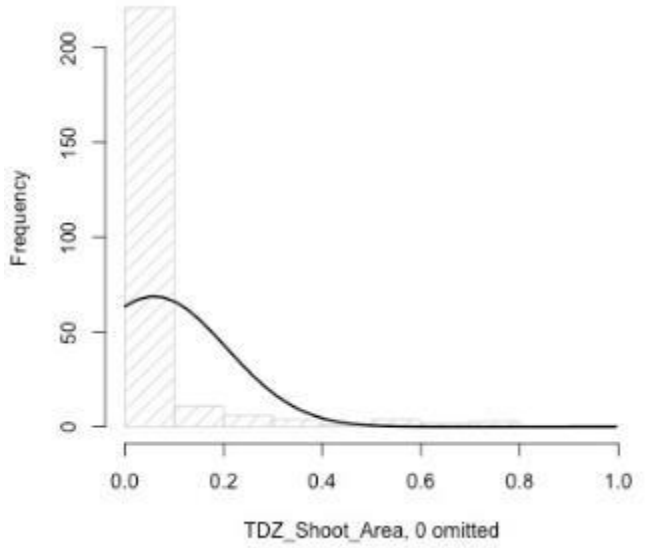
Proportion of stem with callus





Proportion of stem with shoot





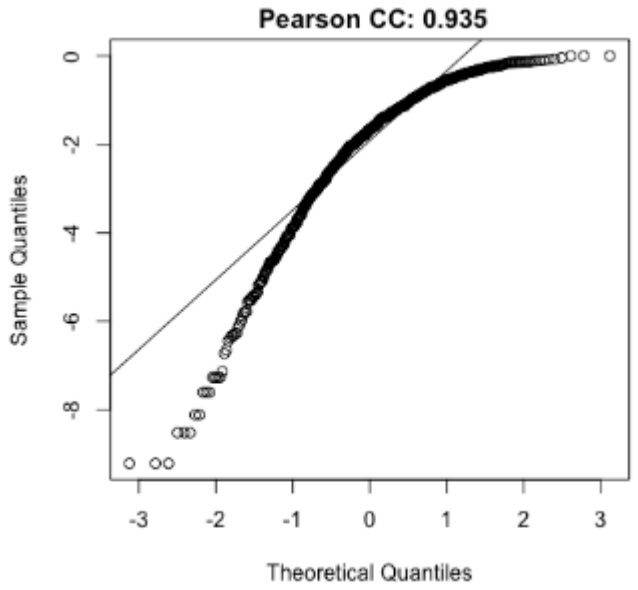
Zero values not shown on histogram:
30 for callus
336 for shoot
(out of 590 total)

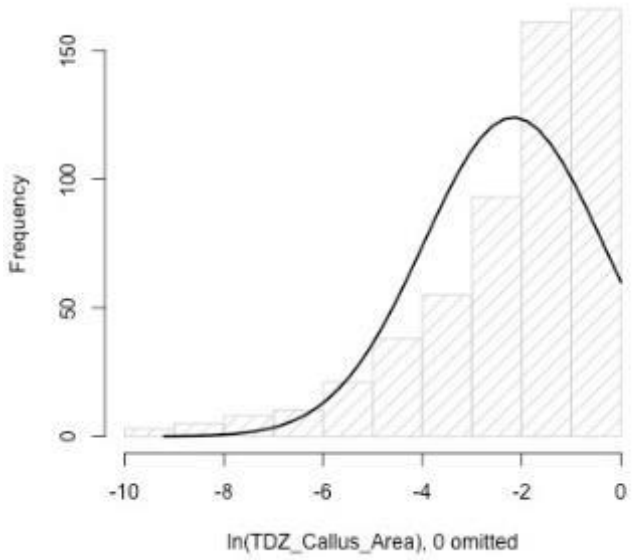
Adherence of traits to normality assumption *after log transformation*?

- Zero values dropped before log-transformation
- With log-transformation, improved adherence to normality assumption for shoot area

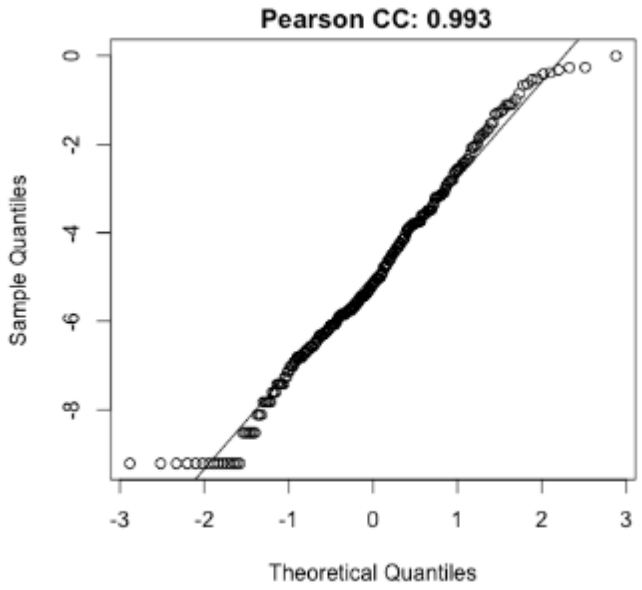
Zero values not shown on histogram:
30 for callus
336 for shoot
(out of 590 total)

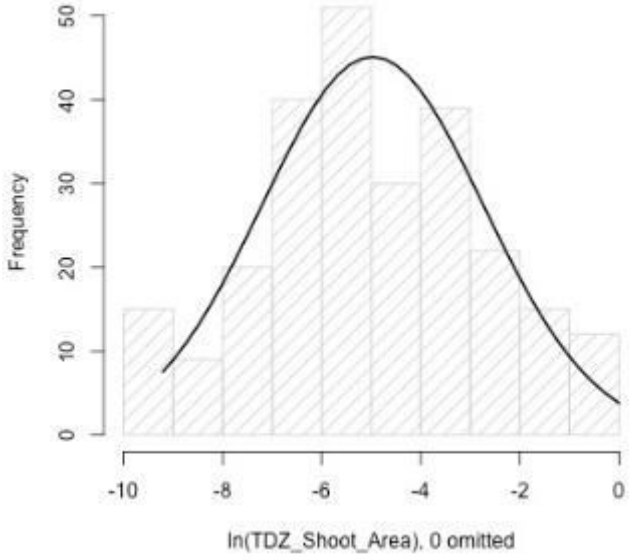
Proportion of stem with callus **(log-transformed)**





Proportion of stem with shoot **(log-transformed)**

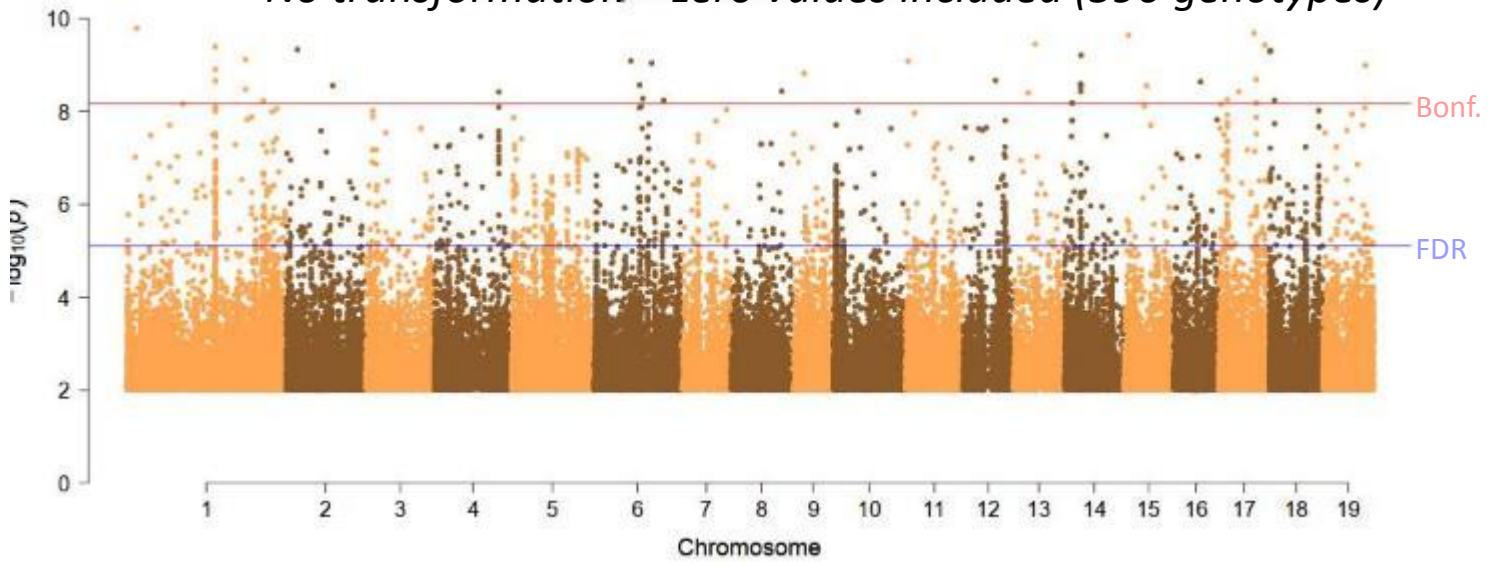




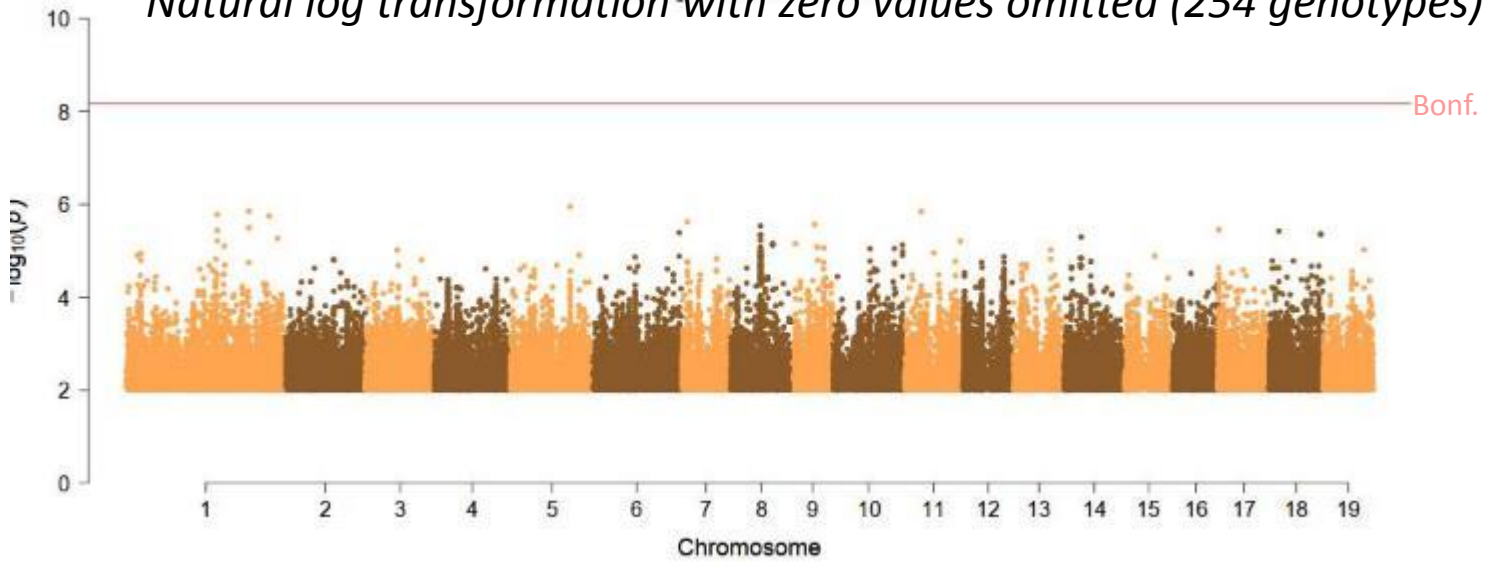
Reducing and coercing data to normality
by dropping zeros and log transformation

- Multiple testing corrections:
  - Bonferroni (red)
  - FDR (blue)
- Alternatives to log transformation…

No transformation – zero values included (590 genotypes)



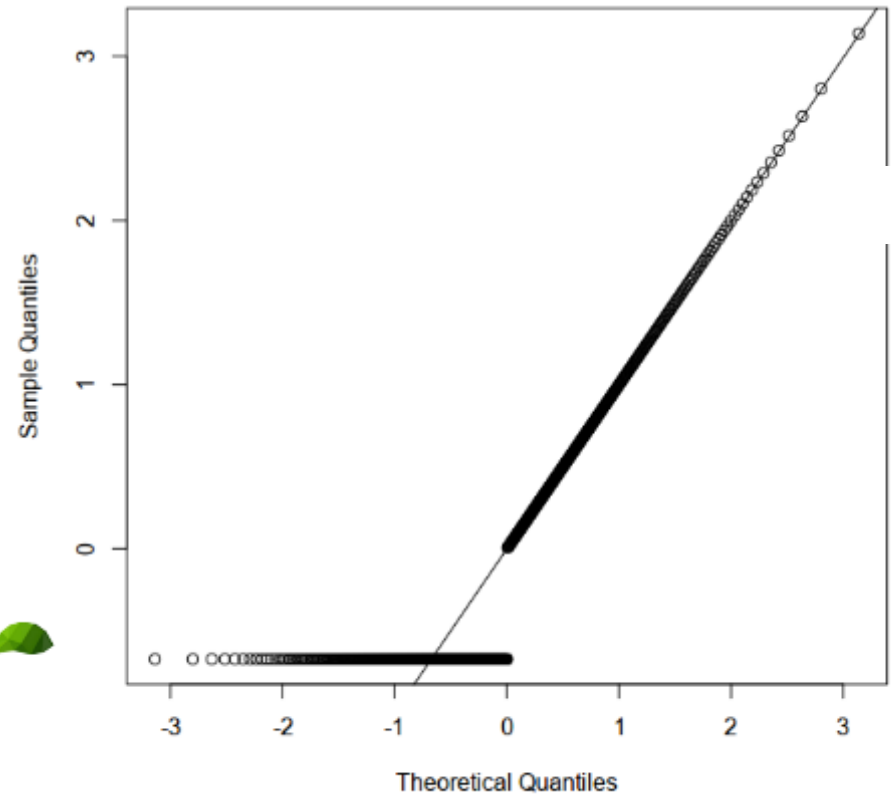
Natural log transformation with zero values omitted (254 genotypes)



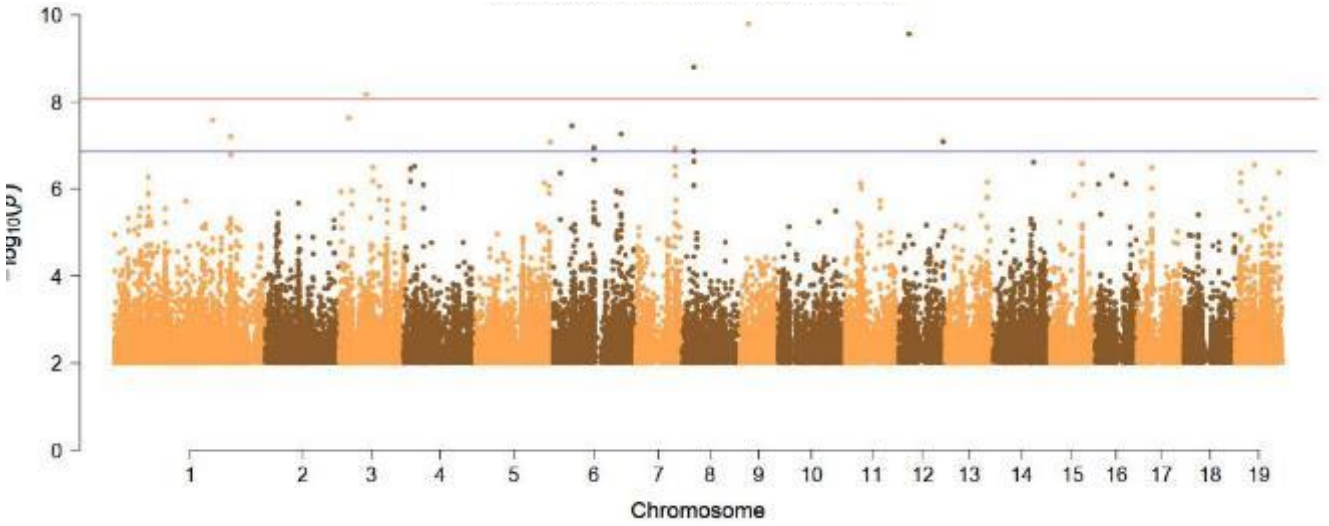
Note: Phenotype data used is from PSPNet with 126 training samples (Jan '19)
– soon to come better models, more training data

For perfect match tc normal distribution, rank-based inverse normal (RB-INV) transformation

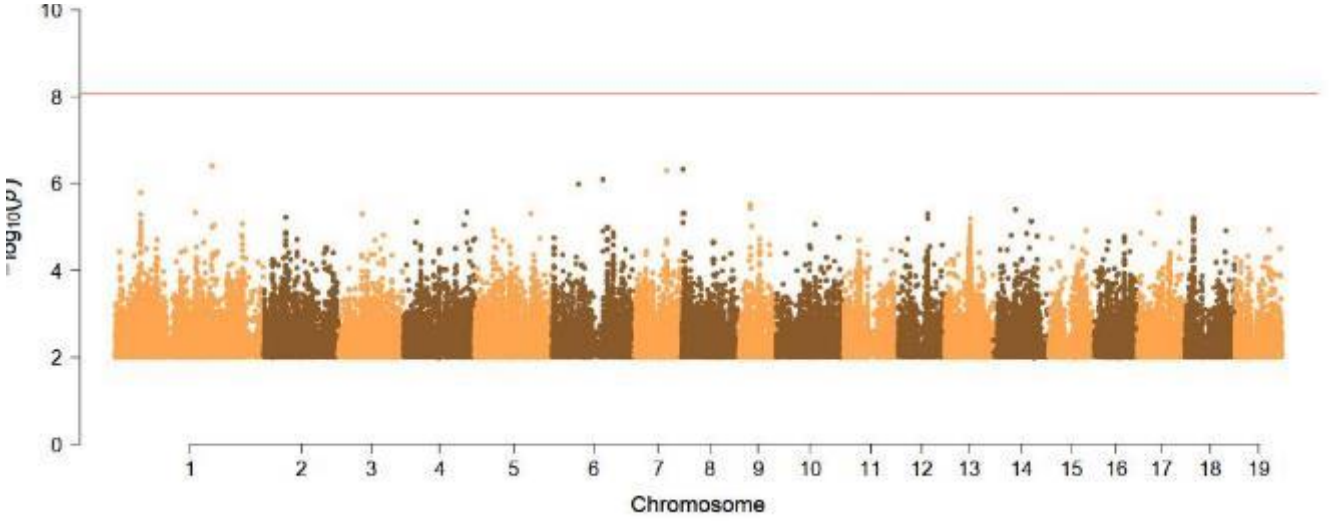
QQ-plot of shoot area phenotype after RB-INV (but before dropping zeros)



RB-INV transf. without zero values omitted (590 genotypes)

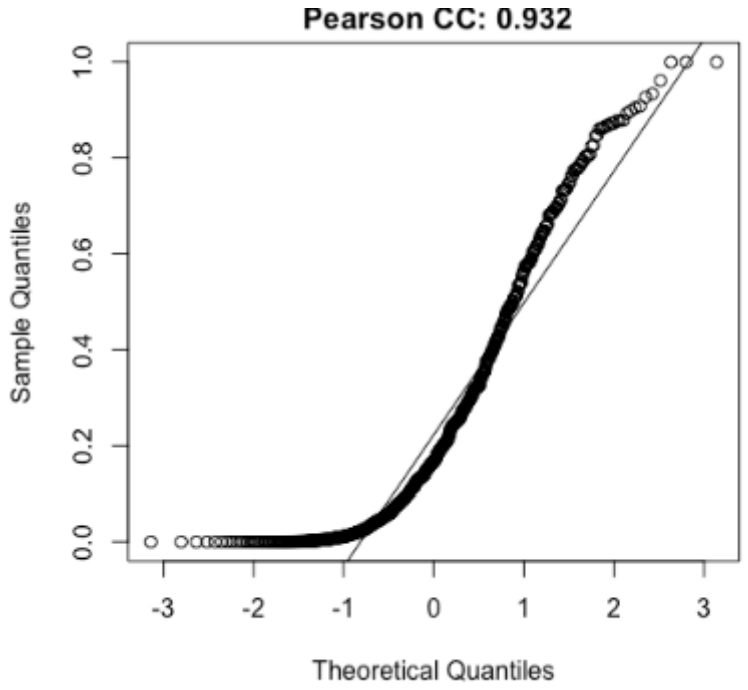


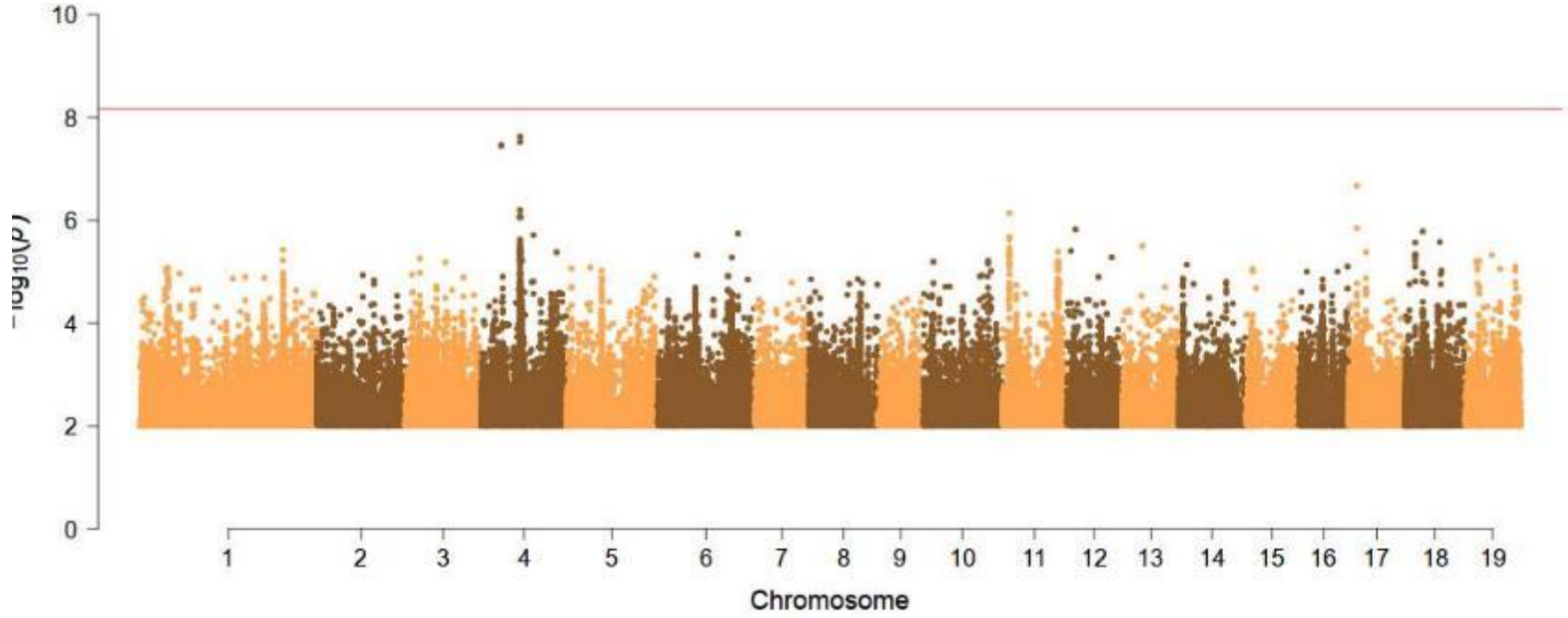
RB-INV transf. with zero values omitted (254 genotypes)



PLINK results for callus area (no transformation)

Proportion of stem with callus



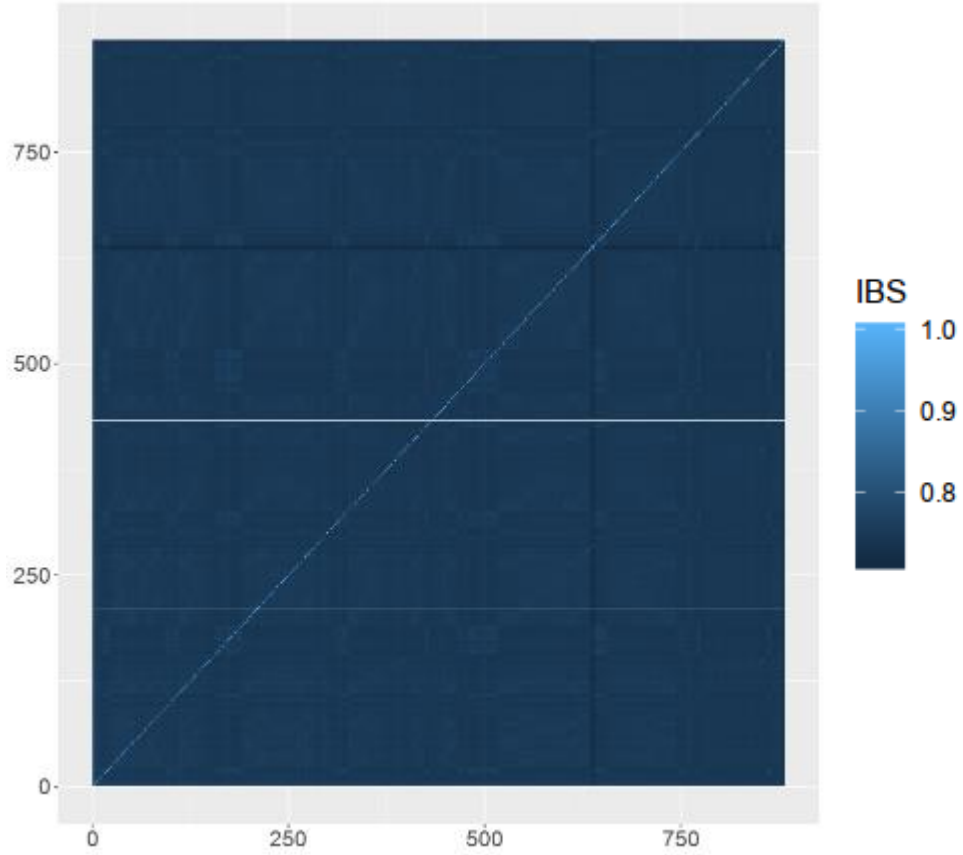


Will more genotypes, increased power, will produce results surviving correction?
Is phenotype normal enough for this model to be valid?

Mixed effect linear models with kinship matrices more conservative than PCs alone

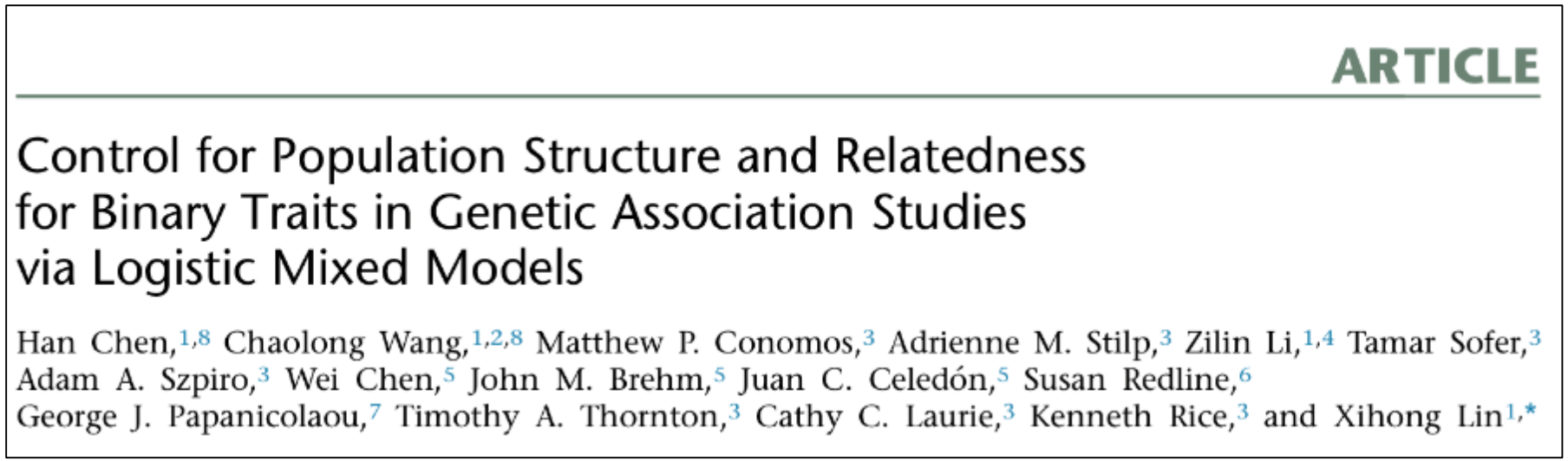
- Kinship matrices capture relationships on finer scale than PCs
- With kinship matrix in model, evidence from closely related individuals is downweighted
- Same kinship matrix can be used for all GWAS methods discussed except
  - PLINK (option not available)
  - SKAT and FaST-LMM (produce their own meeting req.)

Kinship matrix for poplar GWAS population calculated by proportion of SNPs identical-by-sequence (IBS)



Generalized Linear Mixed Model Association Test (GMMAT) offers flexibility

- Generalized linear mixed models (GLMM) for modeling data where errors do not follow normal distribution
- Link function: Model a function of y instead of y
- Flexibility in choosing family of distribution



ARTICLE

Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models

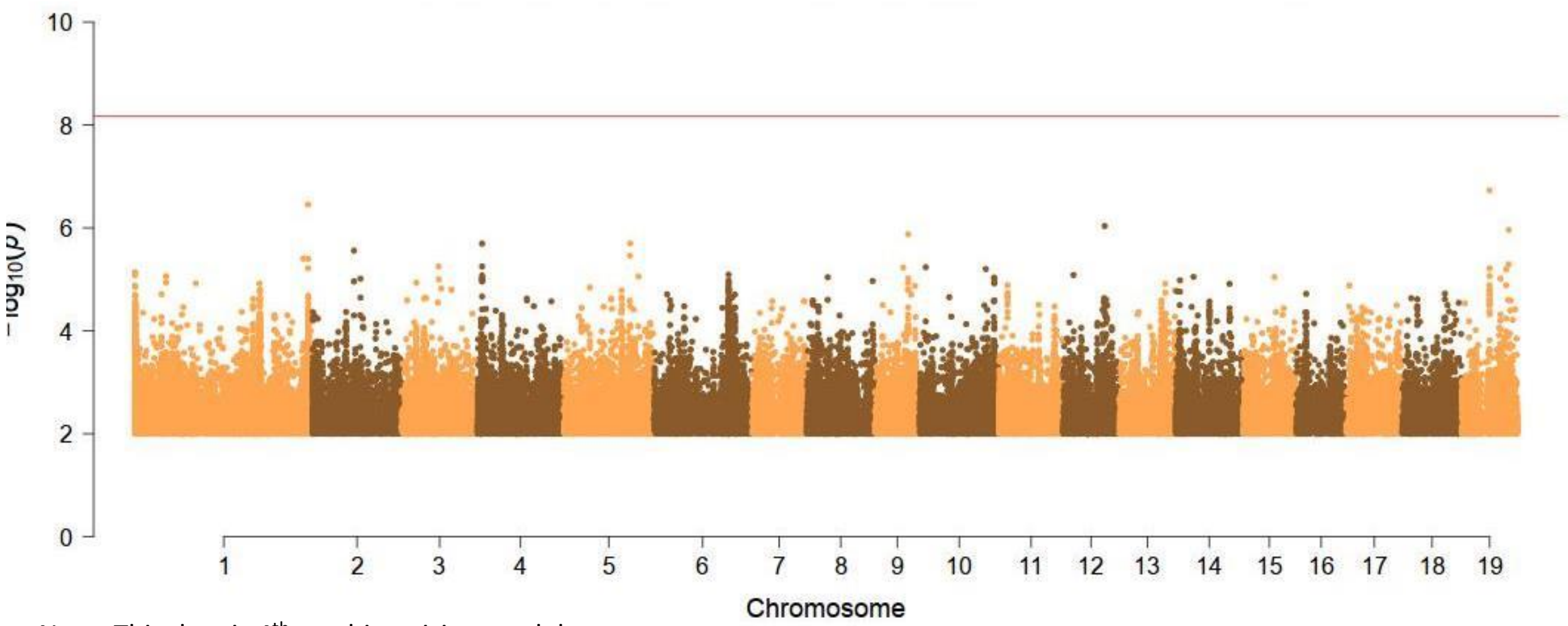
Han Chen,[1,8] Chaolong Wang,[1,2,8] Matthew P. Conomos,[3] Adrienne M. Stilp,[3] Zilin Li,[1,4] Tamar Sofer,[3] Adam A. Szpiro,[3] Wei Chen,[5] John M. Brehm,[5] Juan C. Celedón,[5] Susan Redline,[6] George J. Papanicolaou,[7] Timothy A. Thornton,[3] Cathy C. Laurie,[3] Kenneth Rice,[3] and Xihong Lin[1,*]

In single-locus GWAS, no valid methods are producing results that survive multiple test correction

	Transformation	Link function	Distribution family	Results
GEMMA Linear Mixed Model-	None	Identity	Normal	Statistically invalid, inflated false positives
	Log-transformation*	Identity*	Normal	Inflated false negatives?
GMMAT Generalized Linear Mixed Model	Binarization	Logit	Binomial	Inflated false negatives?
	None*	Log*	Normal	Null model fails to converge
	None*	Identity*	Gamma	Data too right-skewed for proper test

Combinations of link function and distribution family not listed as supported in GMMAT manual

Binarized shoot phenotype tested by GMMAT



Note: This data is 4th machine vision model (126 training samples), binarized shoot

SNP-set sequence kernel association test (SKAT) tests combined effect of SNP groups

- SKAT H_0: No effect of the kernel (**K**) on trait
- **K** calculated by reducing windows of SNPs into statistics representing the frequency of rare alleles and how rare they are

ARTICLE

2011

Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test

Michael C. Wu,[1,5] Seunggeun Lee,[2,5] Tianxi Cai,[2] Yun Li,[1,3] Michael Boehnke,[4] and Xihong Lin[2,*]

ARTICLE

2013

Sequence Kernel Association Tests for the Combined Effect of Rare and Common Variants

Iuliana Ionita-Laza,[1,6,*] Seunggeun Lee,[2,6] Vlad Makarov,[1] Joseph D. Buxbaum,[3,4,5] and Xihong Lin[2,*]

2016

An efficient resampling method for calibrating single and gene-based rare variant association analysis in case–control studies

SEUNGGEUN LEE*, CHRISTIAN FUCHSBERGER

Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA

Multi-SKAT: General framework to test multiple phenotype associations of rare variants

Diptavo Dutta[1,2], Laura Scott[1,2], Michael Boehnke[1,2], and Seunggeun Lee *[1,2]

[1]Department of Biostatistics

[2]Center for Statistical Genetics

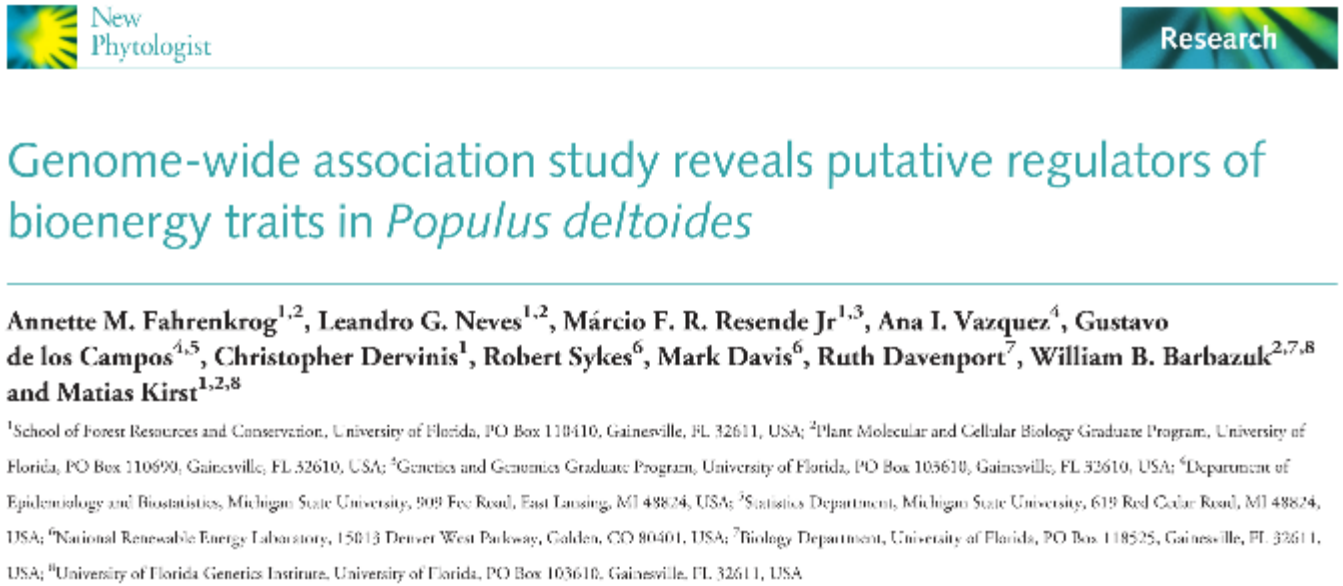
University of Michigan

Ann Arbor, Michigan, USA

2018

Although SKAT most commonly used in human GWAS, has been used in poplar

- Limited statistical power of GEMMA, other single-locus methods
- Rationale for using SKAT: Greater statistical power comes with ability to
  1. detect effects of rare SNPs
  2. test combined SNP effects



New Phytologist

Research

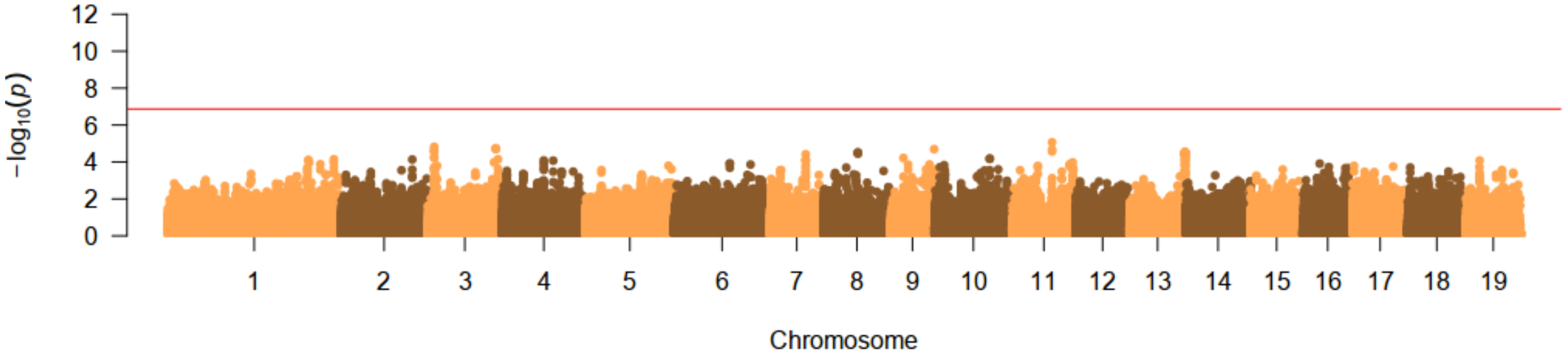
Genome-wide association study reveals putative regulators of bioenergy traits in *Populus deltoides*

Annette M. Fahrenkrog[1,2], Leandro G. Neves[1,2], Márcio F. R. Resende Jr[1,3], Ana I. Vazquez[4], Gustavo de los Campos[4,5], Christopher Dervinis[1], Robert Sykes[6], Mark Davis[6], Ruth Davenport[7], William B. Barbazuk[2,7,8] and Matias Kirst[1,2,8]

[1]School of Forest Resources and Conservation, University of Florida, PO Box 110410, Gainesville, FL 32611, USA; [2]Plant Molecular and Cellular Biology Graduate Program, University of Florida, PO Box 110690, Gainesville, FL 32610, USA; [3]Genetics and Genomics Graduate Program, University of Florida, PO Box 103610, Gainesville, FL 32610, USA; [4]Department of Epidemiology and Biostatistics, Michigan State University, 909 Fee Road, East Lansing, MI 48824, USA; [5]Statistics Department, Michigan State University, 619 Red Cedar Road, MI 48824, USA; [6]National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO 80401, USA; [7]Biology Department, University of Florida, PO Box 118525, Gainesville, FL 32611, USA; [8]University of Florida Genetics Institute, University of Florida, PO Box 103610, Gainesville, FL 32611, USA

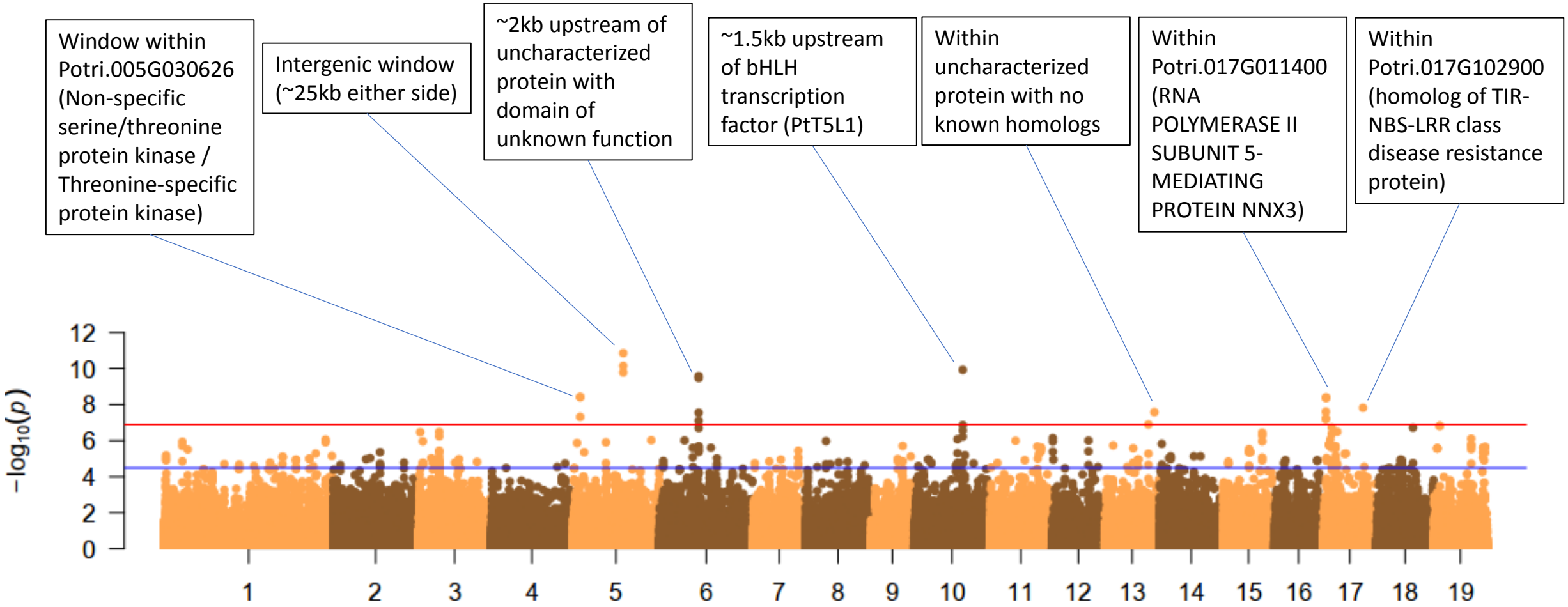
First preliminary results with SKAT shown for callus

(Phenotypes from VGG19 with 590 genotypes, 88 training samples)

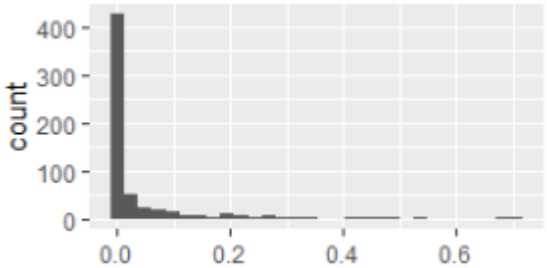
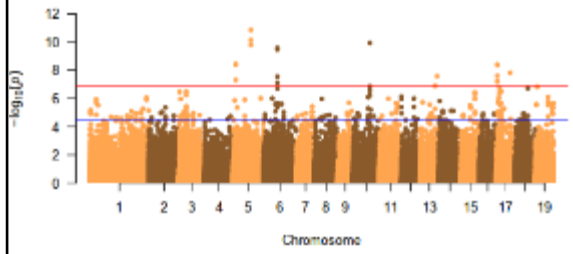
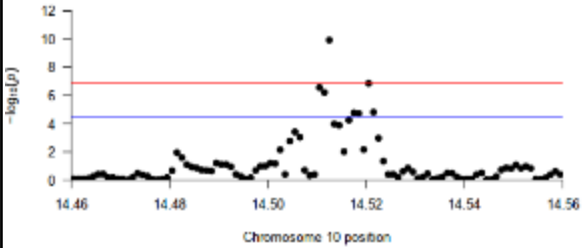
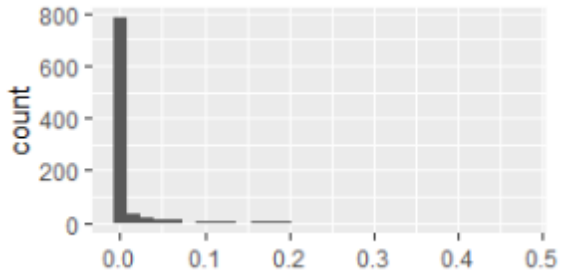
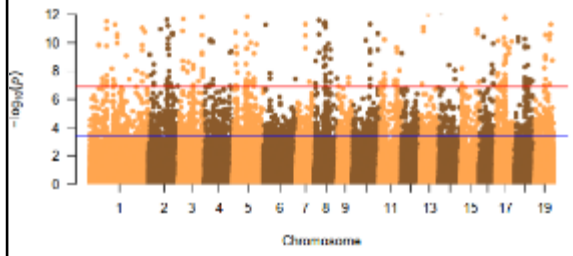
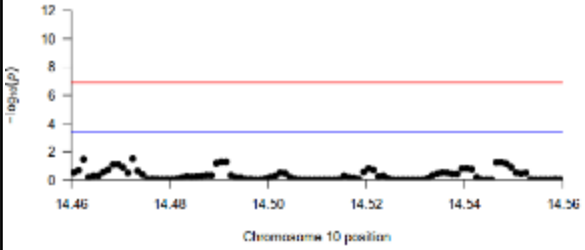


Results for shoot: Many highly significant SNPs are found in intergenic regions or uncharacterized genes

(first SKAT model shown – VGG19 with 590 genotypes, 88 training samples



Wildly different results for SKAT with two phenotype datasets

Model	# Training samples	Mean IoU	# Testing samples	Distribution of phenotype values	Whole genome view	View zoomed to segment of Chr. 10
VGG19	88	79.0	590			
PSPNet	126	80.6	882			

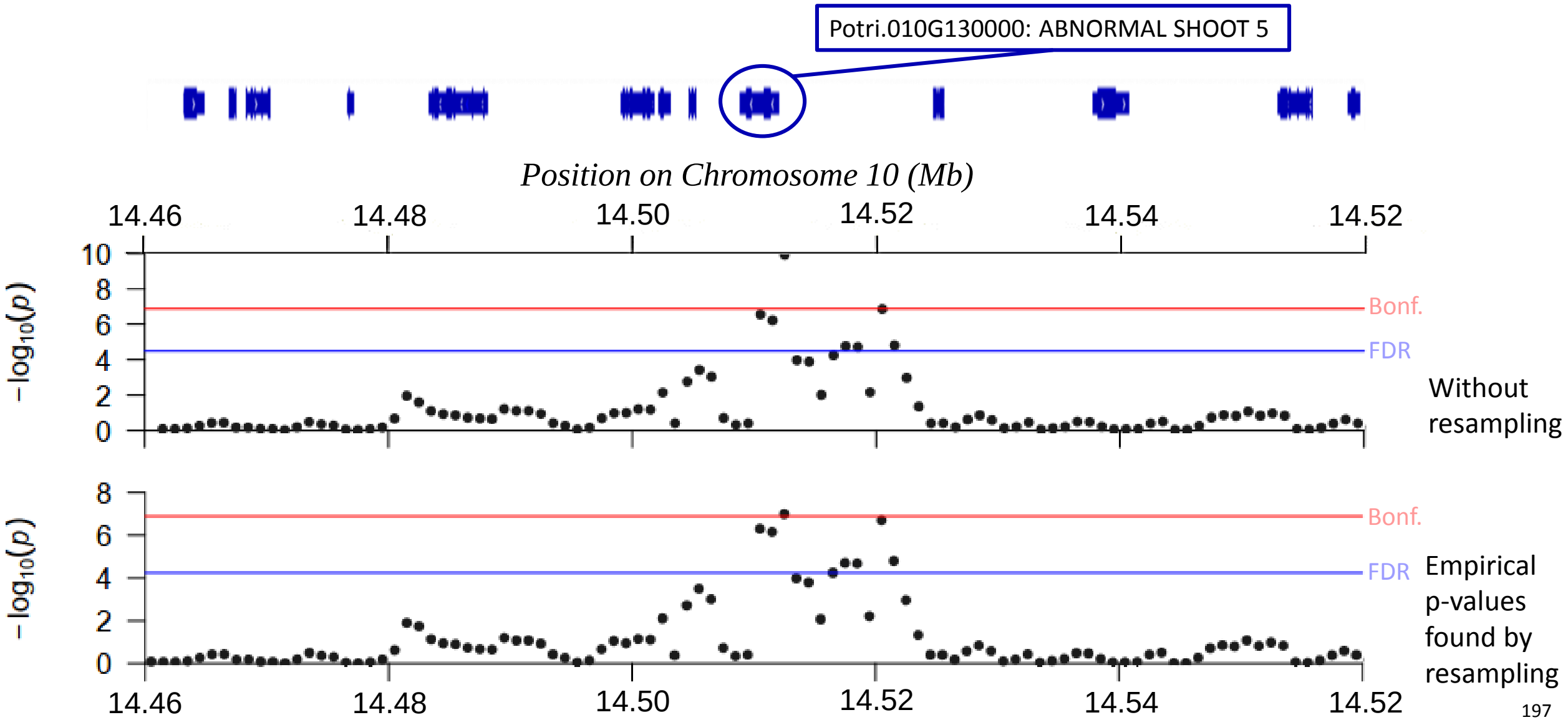
Multiple-testing corrections

- Bonferroni adjustment
  - Depends only on number of tests and confidence level (usually 0.05)
  - Assumes independence of tests
  - Correlations between phenotypes, and between SNPs in linkage disequilibrium
  - Adjusted Bonferroni for N(effective) independent SNPs – less conservative
- False Discovery Rate adjustment (Benjamini-Hochberg)
  - FDR threshold depends on distribution of p-values and varies between traits... many producing no FDR-significant p-values and no ability to calculate a threshold
  - Usually less conservative than Bonferroni, with extent depending on trait

Resampling to allow relaxation of normality assumption

- An alternative to transformation or the use of GLMMs
- Resampling by permutation:
  1. Scramble phenotype data
  2. Repeat test X times, list p-values of effect of SNP on (randomized) phenotype
  3. Where does p-value for true data fall in null distribution?
- Adaptive resampling (AR) to reduce computational burden
  Runtimes for AR with shoot phenotype shown:
  - ~100 CPU hours for SKAT
  - CPU years for single-locus methods

Little change from resampling



According to some sources, resampling may be used as substitute for multiple testing correction

- Correlations between SNPs are preserved during resampling (genotype data not shuffled – only phenotype data)
- Resampling control for familywise error-rate and abolish need for multiple testing correction for SNPs?[1,2]
- Examples from human GWAS[3]:
  - "Empirical p-values<0.017, reflecting Bonferroni correction for 3 independent tests (one per brain region): $\alpha = 0.05/3$, were considered to represent significant association."[3]

1. "Permutation procedures" in PLINK manual, 2017. Broad Institute and collaborators (http://zzz.bwh.harvard.edu/plink/perm.shtml)
2. Gao, X., Becker, L.C., Becker, D.M., Starmer, J.D. and Province, M.A., 2010. Avoiding the high Bonferroni penalty in genome-wide association studies. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, *34*(1), pp.100-105.
3. Mignogna, K.M., Bacanu, S.A., Riley, B.P., Wolen, A.R. and Miles, M.F., 2019. Cross-species alcohol dependence-associated gene networks: Co-analysis of mouse brain gene expression and human genome-wide association data. *PloS one*, *14*(4), p.e0202063.

Next: Speeding up analysis to enable more resampling with single-locus GWAS

Tool	Type	Notes
PLINK	General linear model	Model phenotype as function of SNPs and PCs
TASSEL5	Mixed linear model	GUI, not written for large SNP sets
GENESIS		Assumes σ^2 equal for every SNP
EMMAX		
GEMMA		Precise, efficient, has multivariate option
GMMAT	Generalized mixed linear model	Can build models for certain non-gaussian distributions
SKAT	MLM with kernelized SNP-sets	Tests user-defined SNP groups
Farm-CPU	Alternating mixed/unmixed model	Not written for large SNP sets, need debugging/hack/update
BOLT-LMM	Bayesian mixed linear model	Adds no SNPs to model because estimated H^2_{SNP} has σ^2 too large
MLMM		
FaST-LMM	**Mixed linear model**	**Fast GWAS method that is well-established and exact**

Key:

Producing results for full SNP data
Problems due to data input

Developing capabilities for resampling in single-locus GWAS

- *Requirement 1:*
  Efficient GWAS code –
  FaST-LMM built for high-throughput GWAS
- *Requirement 2:*
  Public high-performance cluster computing resources
- *Requirement 3:*
  Code for parallelization of FaST-LMM with resampling
  (using Apache Spark or similar framework)

FaST-LMM for high speed and accuracy

- Mathematical approach:
  - Algebraic transformation to find uncorrelated SNPs, use these to build kinship matrix
  - Use of a kinship matrix made from *M* SNPs lower than *N* genotypes allows for models to be built more quickly
- Consequences of FaST-LMM innovations:
  - Order of magnitude faster than inexact method EMMAX, nearly 2 orders of magnitude faster than exact method GEMMA
  - Said to produce same results as GEMMA
- Released in 2012, continues to be used in applied GWAS studies, and in methods studies as a standard to compare new methods to

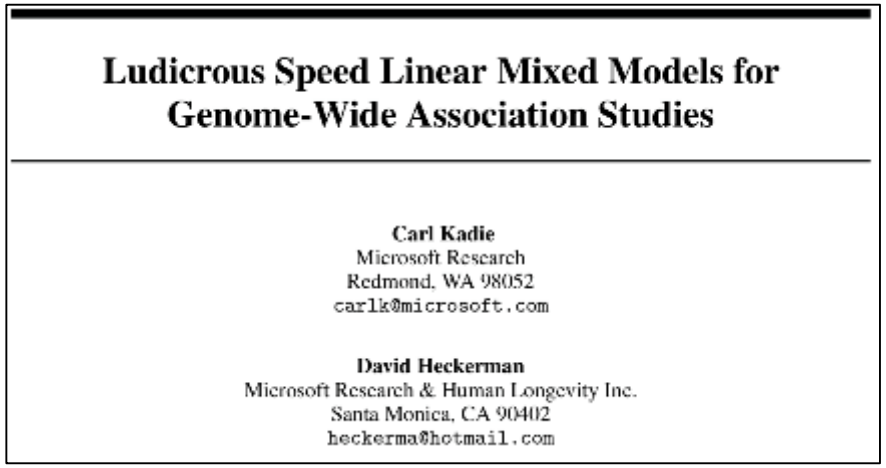
Lippert, C., et al. 2011. FaST linear mixed models for genome-wide association studies. *Nature methods*, *8*(10), p.833.

Public high-performance computing resources

- Open Science Grid (NSF) –
  - Distributed network of clusters at many institutions share jobs
  - Explicitly mentions permutation tests as an example of ideal use (https://support.opensciencegrid.org/support/solutions/articles/5000632058-is-the-open-science-grid-for-you-)
- XSEDE (NSF)
  - Provides access to NSF-owned clusters
  - Competitive proposal approval process with limited resources at first
- SUMMIT (DoE)
  - Most powerful cluster in world with 4,608 nodes featuring high-end CPUs and GPUs
- NSF National Center for Atmospheric Research Computational & Information Systems Lab

Published parallel implementations of FaST-LMM not appropriate

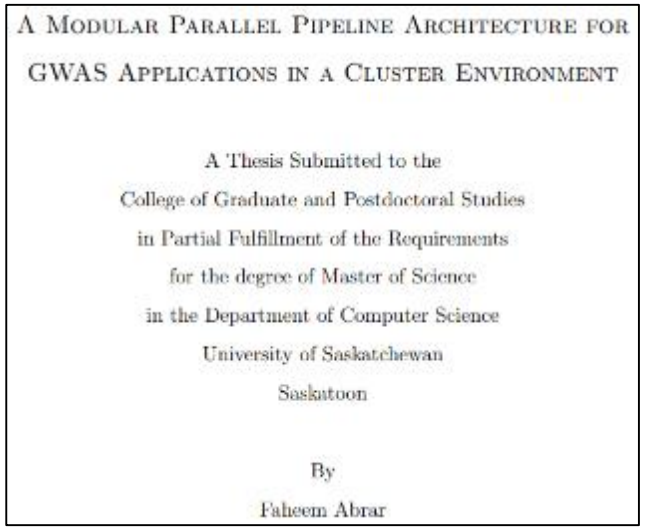
- "Embarrassingly parallel": Can be parallelized simply by dividing data, work among computers
  - The case for GWAS with large numbers of phenotypes
  - Relatively easy with common parallel computing frameworks
- Otherwise, implementations intended for single massive GWAS
  - 2018 implementation by FaST-LMM inventors
  - 2019 Master's thesis



Ludicrous Speed Linear Mixed Models for Genome-Wide Association Studies

Carl Kadie
Microsoft Research
Redmond, WA 98052
carlk@microsoft.com

David Heckerman
Microsoft Research & Human Longevity Inc.
Santa Monica, CA 90402
heckerma@hotmail.com



A Modular Parallel Pipeline Architecture for GWAS Applications in a Cluster Environment

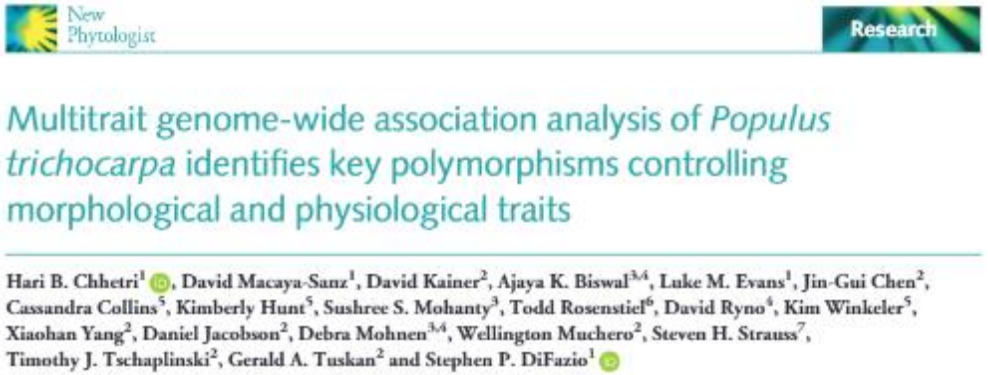
A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in the Department of Computer Science
University of Saskatchewan
Saskatoon

By
Faheem Abrar

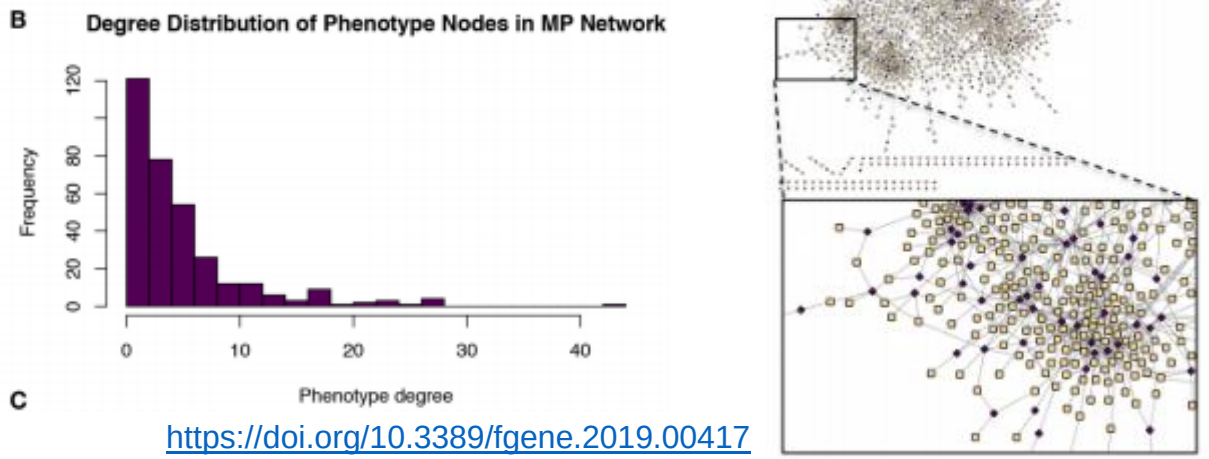
Multi-trait analysis can increase power when correlations exist between traits

– increasingly common in poplar community

- Approach 1: GWAS with multiple response variables (phenotypes)
  - Options built into GEMMA, GMMAT
  - Add-ons for SKAT and FaST-LMM (e.g. Multi-SKAT)
- Approach 2: Network-based methods to analyze relationships between large numbers of phenotypes influenced by many of the same QTLs

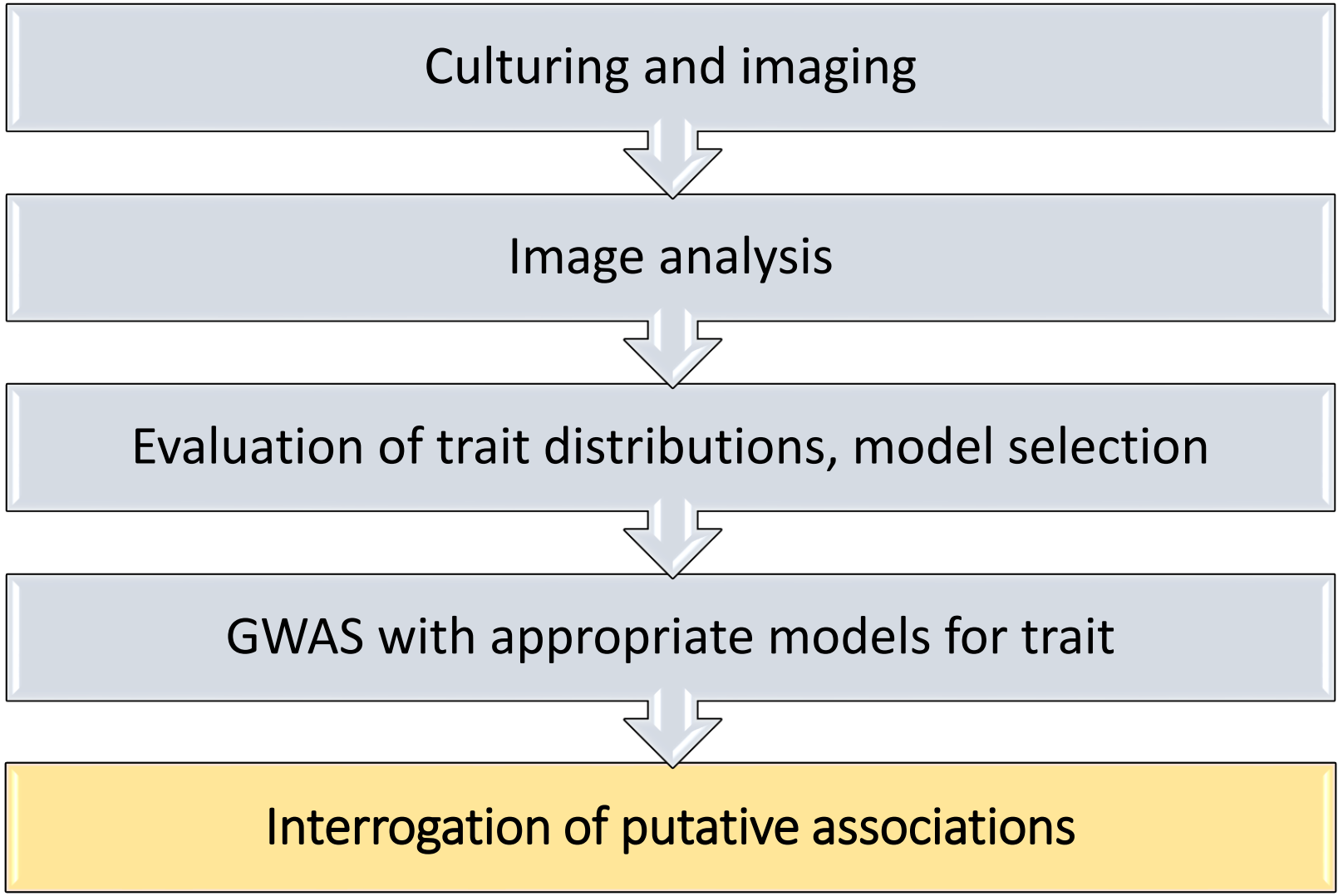


Network between module (cluster of SNPs) and phenotype (MP Network) provides insight into related traits in poplar (Weighill et al. 2019)



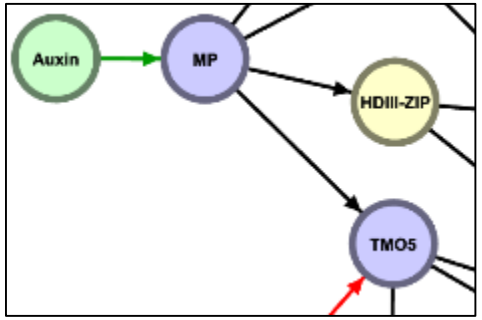
https://doi.org/10.3389/fgene.2019.00417

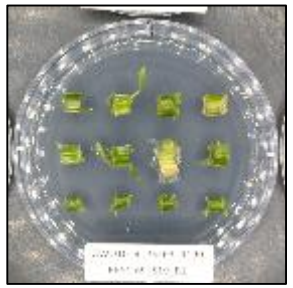
From phenotyping to GWAS and beyond



Approaches to interrogating putative SNPs by making use of available knowledge on homologs

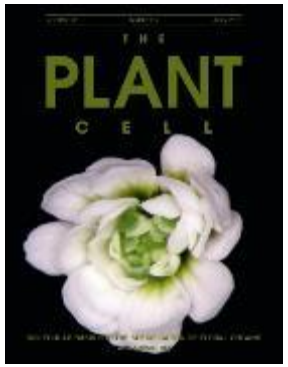
A general approach for SNPs in/near genes uncharacterized in poplar





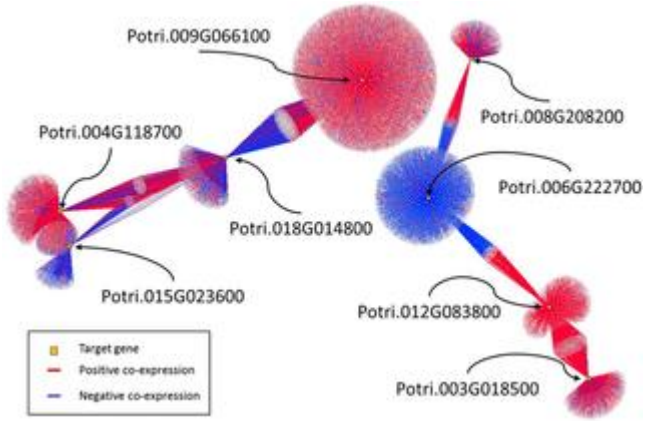
Functional evidence

- **Computational methods:**
  - Epistasis analysis to detect conservation of gene-gene interactions between Arabidopsis and poplar homologs
- **Mutant studies:**
  - If QTL is in/near transcription factor: Transient agroinfiltration with overexpression vectors and qPCR to study downstream genes
  - *In vitro* transformation and regeneration assays with overexpression vectors



Learning about possible homologs

- BLAST or similar alignment tools to find homologs
- Literature review for homologs (basic literature in Arabidopsis, other model/nonmodel plants)
  - Role in cell fate determination?
  - Expression patterns?
  - Role in wider genetic regulatory network?



Transcriptomic evidence

Available poplar transcriptome resources (Phytozome.doe.gov)

- Co-expression data (tissues including leaf, xylem, root from ORNL) Figure above shown from Tuskan et al. 2018
- eQTL analysis
- Promoter analysis to identify conserved regulatory motifs

Example of validation by epistasis:
Effect of PtT5L1 on shoot depends on LHW homolog

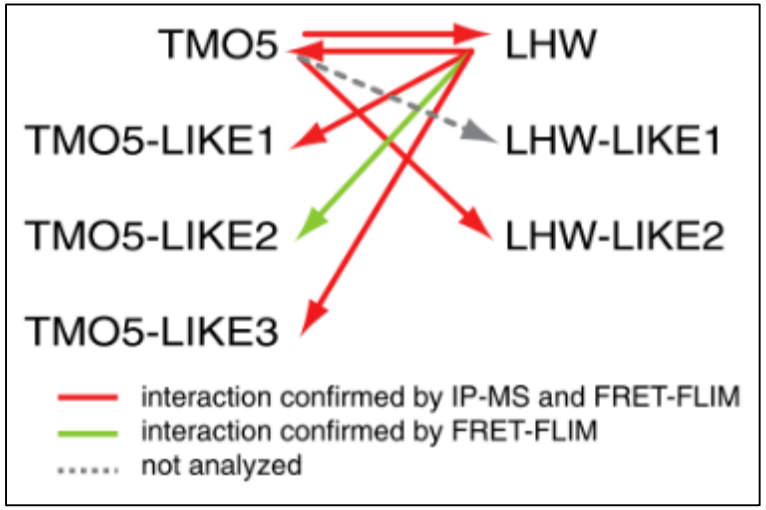
- Limited SNP set including 96 poplar genes (with 1044 SNPs) that are related to Arabidopsis TMO5 or LHW (Smith-Waterman score >200)
- Ran logistic regression on binarized shoot phenotype to avoid violating normality assumption
- Interaction between SNPs in PtT5L1 and PtLHW-LIKE1 (p=3.109e-08)
  - Survives Bonferroni correction (threshold of 5.6e-06)

LR tests for interaction terms between TMO5, LHW subclades

SNP1	SNP2	$\beta_{interaction}$	p_0
Chr01_ 6791671	Chr10_ 14508786	-1.001540	3.109e-08
Chr10_ 14508513	Chr10_ 14508898	-0.934245	1.898e-07
Chr10_ 14508898	Chr10_ 14509560	-0.925768	2.640e-07

Interaction between LHW-LIKE3 homologs and PtT5L1

Epistasis between intragenic SNPs in PtT5L2



1. De Rybel. "A bHLH Complex Controls Embryonic Vascular Tissue Establishment and Indeterminate Growth in Arabidopsis." *Dev. Cell* 2013

Summary of GWAS methodology

- Distribution of phenotype data dictates GWAS tools that can be used for statistically valid tests
- Shoot phenotypes highly non-normal… several options:
  - Transformation
  - Generalized models
  - Resampling
- Validation of associations
  - Insights from literature
  - Interactome (epistasis analysis)
  - Transcriptome (eQTL mapping, co-expression)
  - Potential for mutant studies?

Next steps for GWAS: Analysis and publication

- Implementation of parallel resampling in FaST-LMM
  - Use of high-performance cluster (NSF Open Science Grid, etc)
- Execution of GWAS workflow with final phenotype data for all traits

Phenotypes	Current status	Next steps	Aim to publish
Stem regeneration (callus and shoot)	Completing additional annotations for MV training	Deploy GWAS workflow, interpret results and write paper	Late 2019 / Early 2020
Rooting	Refining MVmodel	Deploy MVmodel and GWAS workflow	Late 2019 / Early 2020
In vitro regeneration (callus and shoot)	Completing phenotyping	Annotation for MVtraining, GWAS workflow deployment	Early 2020
Transformation	Optimizing transformation methods and treatments	Select treatments and begin GWAS (Winter)	Late 2019

Thank you for listening

Oregon State University

Advances in integrated analysis of GWAS and eQTN studies in *Populus trichocarpa*

Jin Zhang
Jay Chen
Jerry Tuskan
Wellington Muchero

THE CENTER FOR BIOENERGY INNOVATION

U.S. DEPARTMENT OF ENERGY | Office of Science

Overview and outline

- Recent improvements to the poplar GWAS panel
- Integrating GWAS and eQTN to reveal transcriptional regulation that control of complex traits in poplar

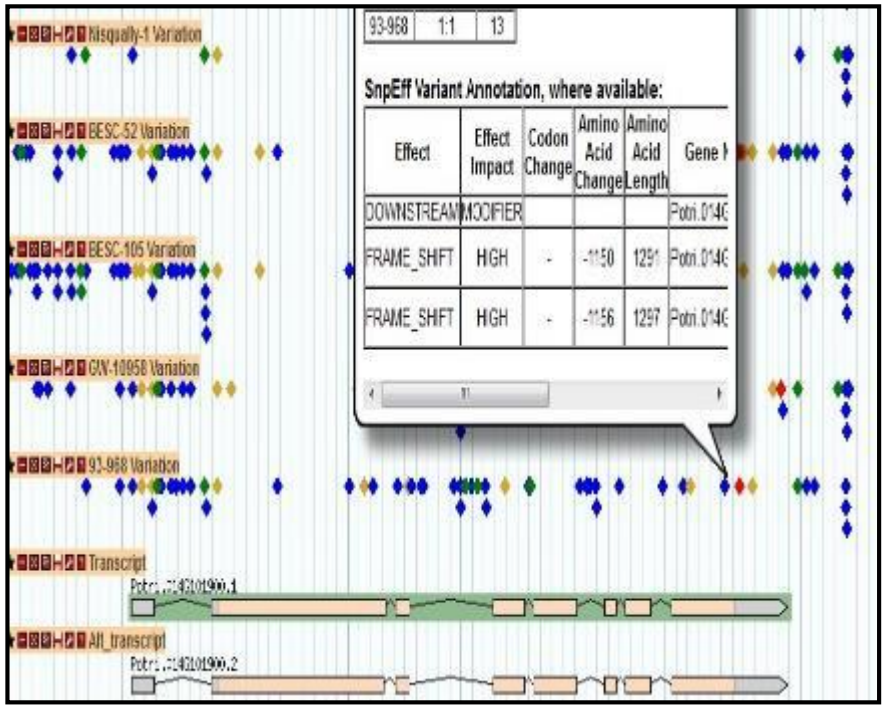
THE CENTER FOR BIOENERGY INNOVATION

Current status of the poplar GWAS panel

Clatskanie, OR (2009)
Coastal Mesic

Corvallis, OR (2009)
Inland Mesic

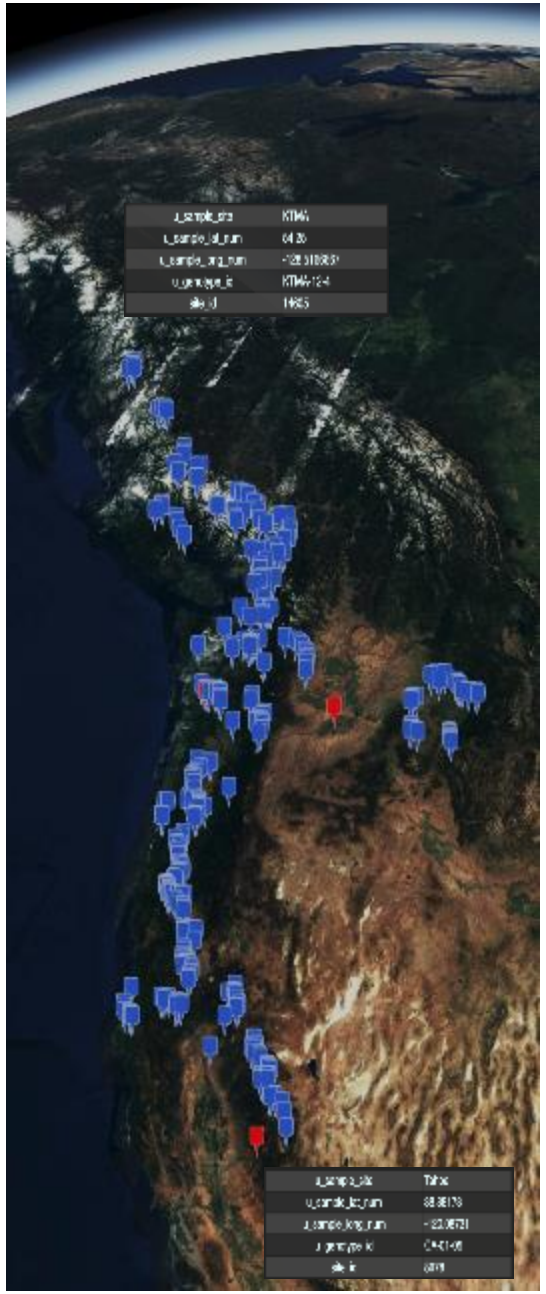
Boardman, OR (2016)
Inland Xeric



Evans et al. Nature Genetics (2014)

n = 917

Current status of the poplar GWAS panel



Clatskanie, OR (2009)
Coastal Mesic

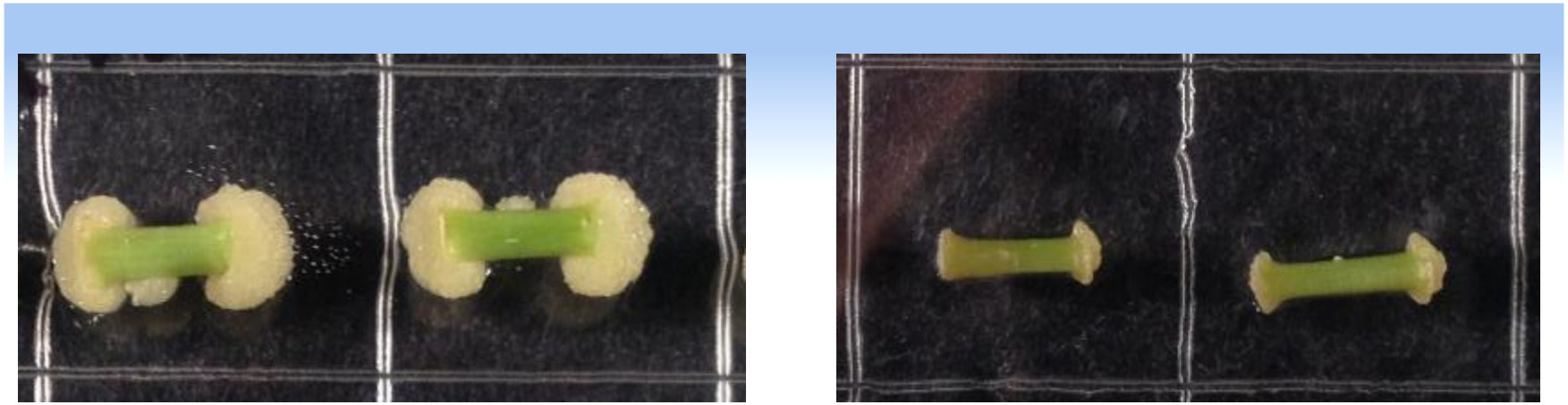
Corvallis, OR (2009)
Inland Mesic

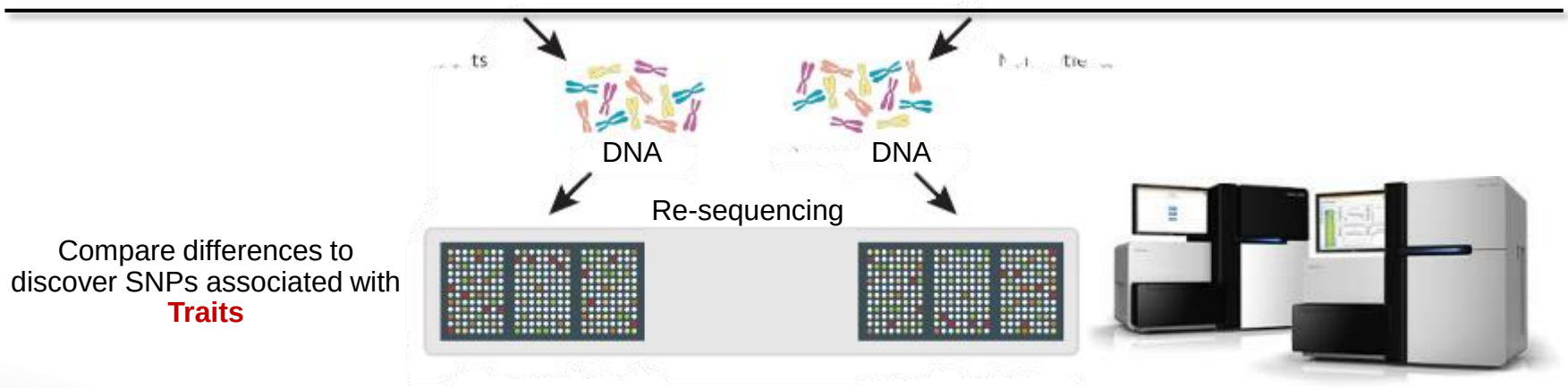
Boardman, OR (2016)
Inland Xeric

- The selected population / panel
  - 1,352 genotypes
  - Southern BC to Northern CA
  - Established in 3 common gardens
- Genotyping
  - Resequenced at a minimum of 18x depth
  - 29 million high-quality SNPs
  - A SNP every 17 bp
  - LD decays on average within 300 bp and in many cases within <20 bp
- Transcriptome
  - RNAseq data was generated for ca. 500 of the 1250 genotypes for leaves, xylem and roots

n = 1,352

Genome-Wide Association Studies (GWAS)

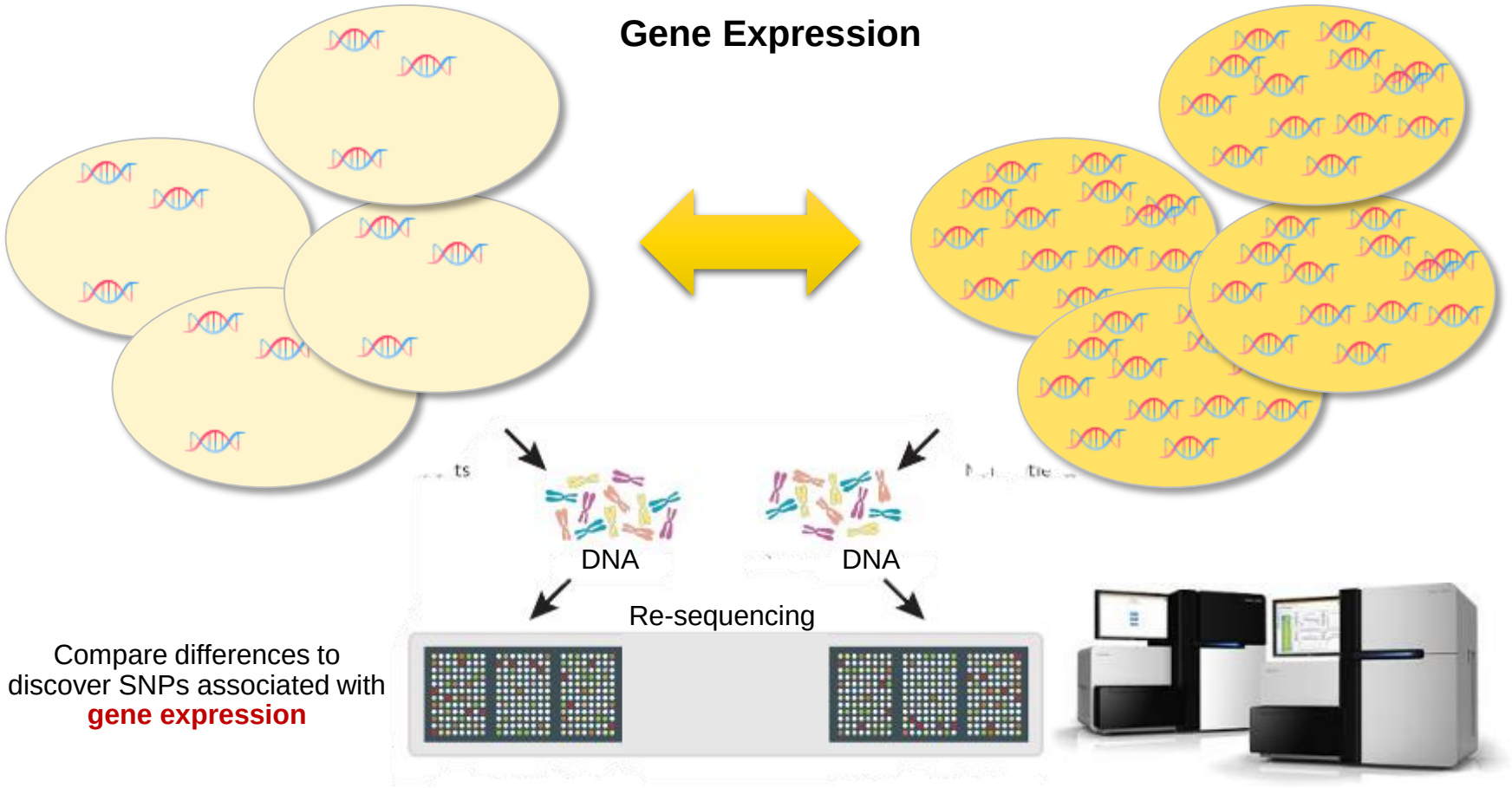




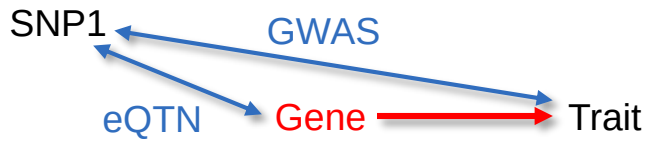
Compare differences to discover SNPs associated with **Traits**

THE CENTER FOR BIOENERGY INNOVATION

expression Quantitative Trait Nucleotide (eQTN)

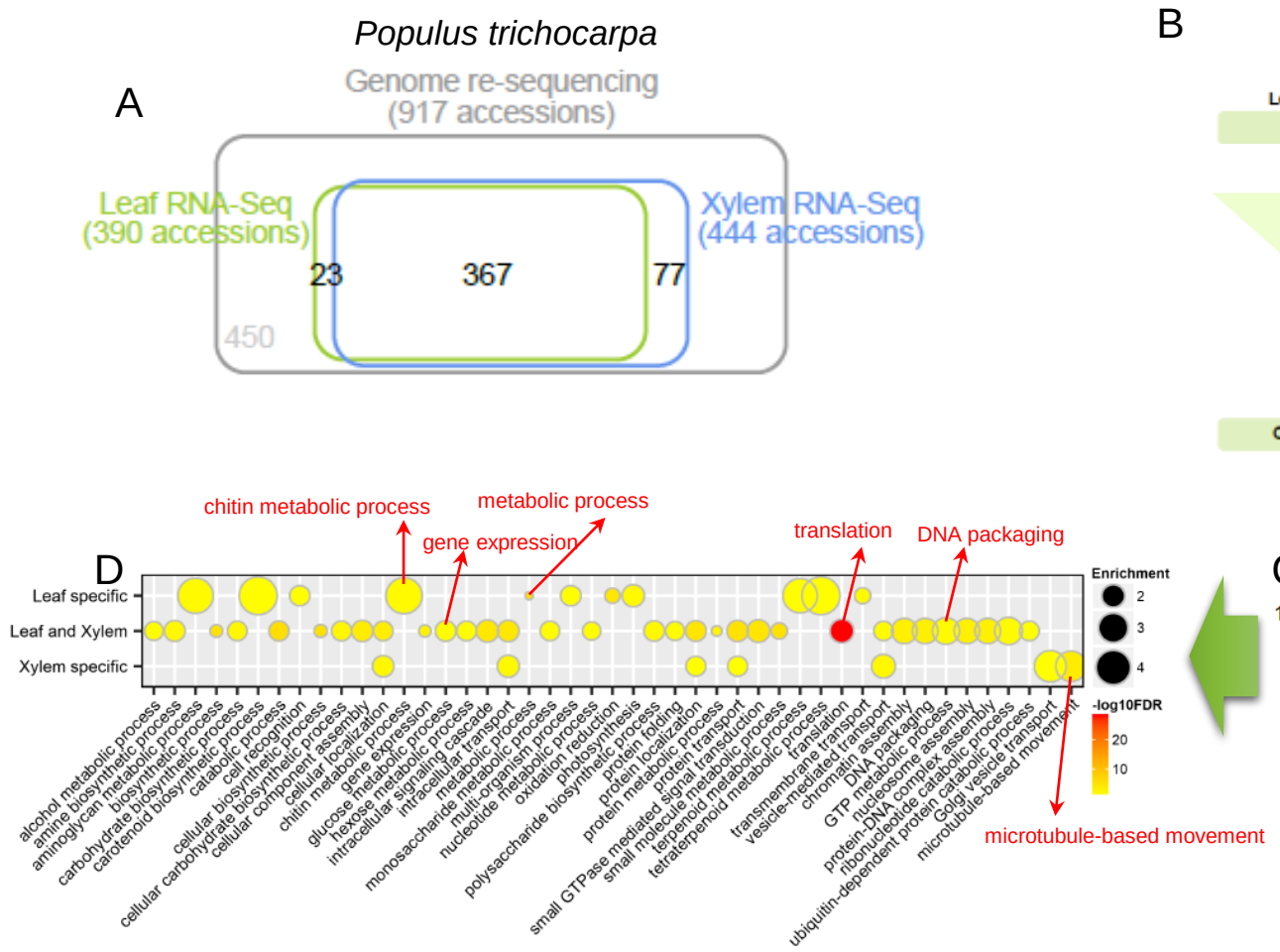


Compare differences to discover SNPs associated with **gene expression**



THE CENTER FOR BIOENERGY INNOVATION

Datasets and data filtering

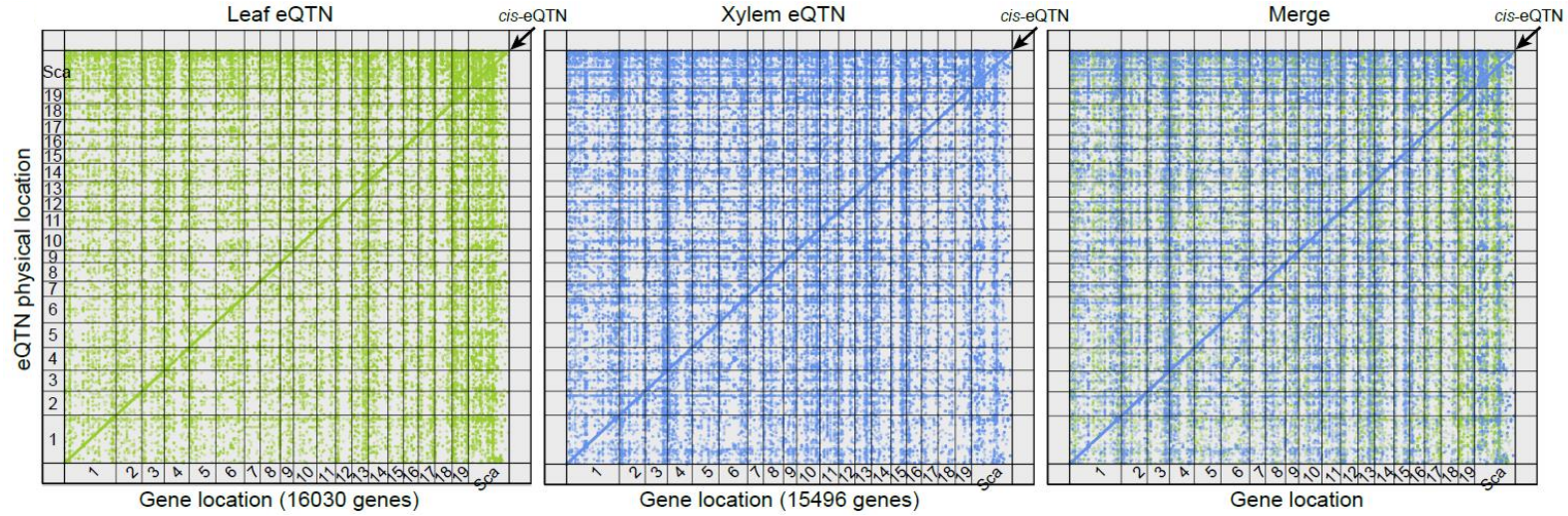


Statistics of eQTN in Leaf and Xylem

A

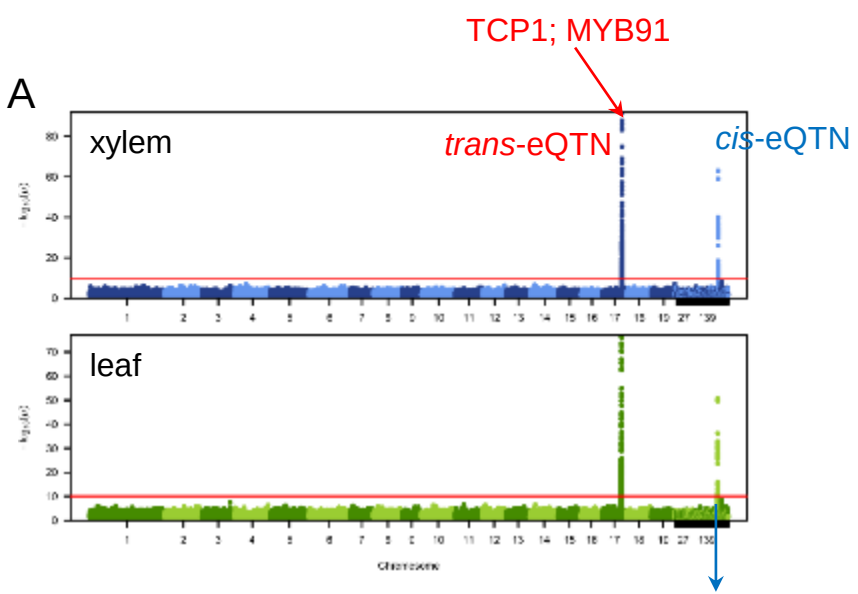
	Leaf		Xylem	
	eQTN	Target Gene	eQTN	Target Gene
***P* < 1E-10**	**790,364**	**6,709**	**1,002,961**	**8,349**
Different Chr	185,851 (23.5%)		215,605 (21.5%)	
Same Chr	604,513 (76.5%)		787,356 (78.5%)	
├ Same Chr (within gene body)	41,051 (6.8%)		63,098 (8.0%)	
├ Same Chr (within 1Mb)	537,326 (88.9%)		682,388 (86.7%)	
└ Same Chr (out of 1Mb)	26,136 (4.3%)		41,870 (5.3%)	

B

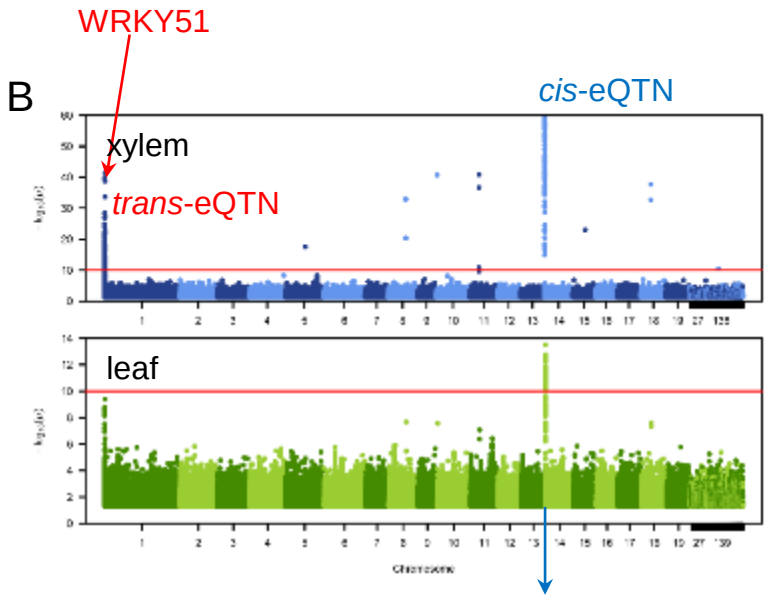


THE CENTER FOR BIOENERGY INNOVATION

Regulation of gene expression by *cis*- and *trans*-eQTN



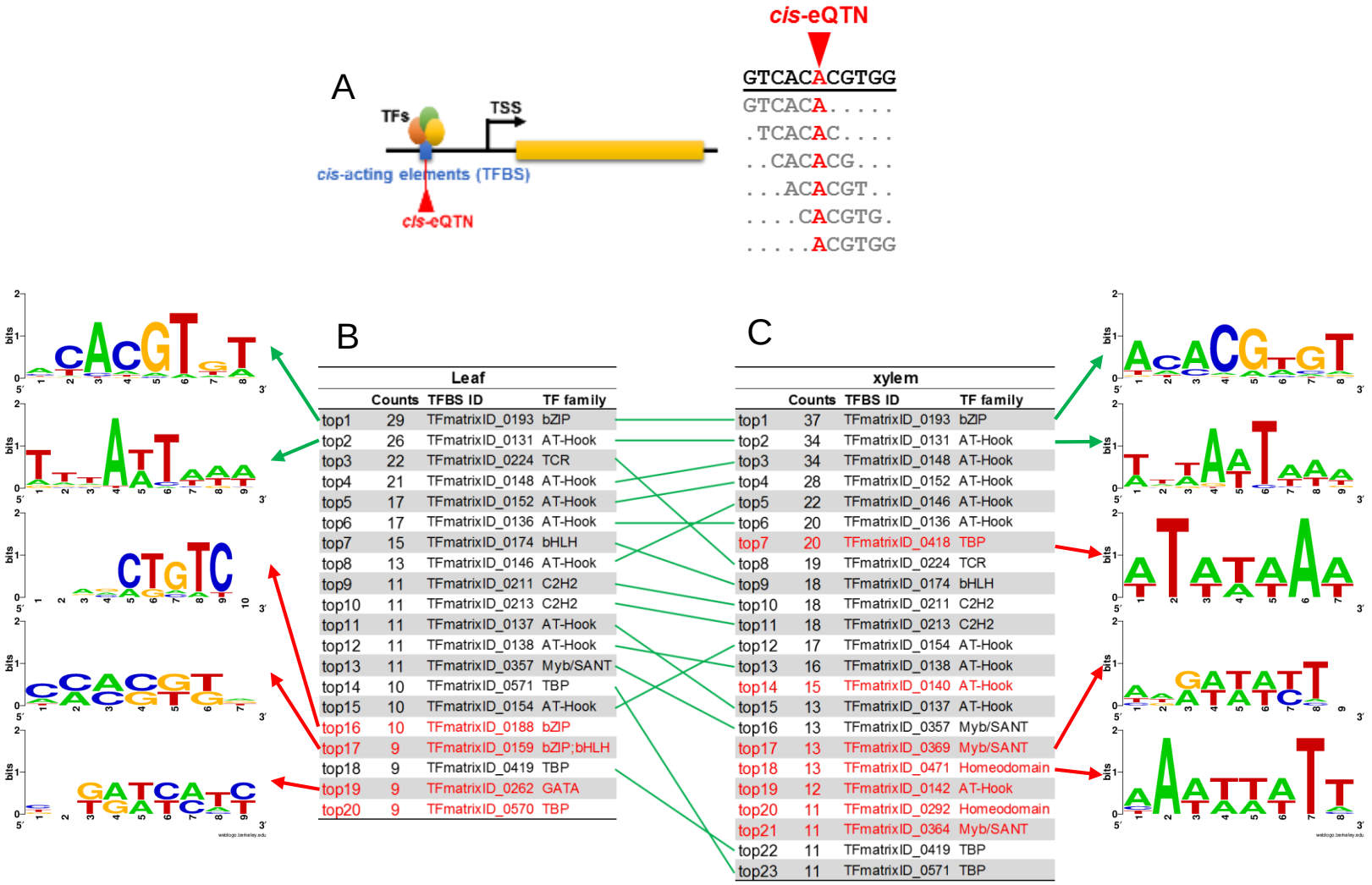
cpn60 chaperonin family protein
T-complex protein 1



HAC12, histone acetyltransferase of the CBP family 12

Encodes an enzyme with histone acetyltransferase activity that can use both H3 and H4 histones as substrates. No single prior lysine acetylation is sufficient to block HAC12 acetylation of the H3 or H4 peptides, suggesting that HAC12 can acetylate any of several lysines present in the peptides.

cis-eQTN



Leaf

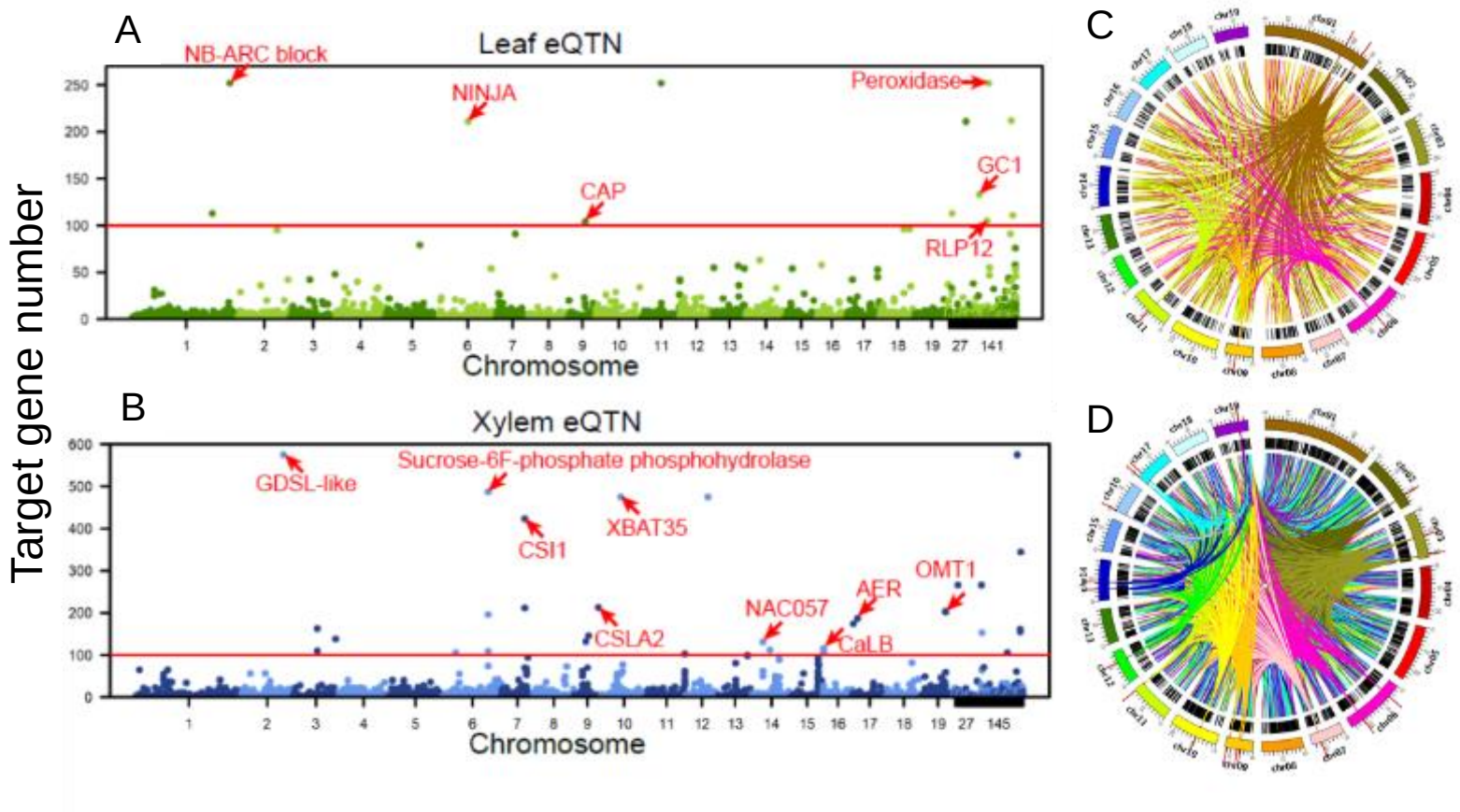
	Counts	TFBS ID	TF family
top1	29	TFmatrixID_0193	bZIP
top2	26	TFmatrixID_0131	AT-Hook
top3	22	TFmatrixID_0224	TCR
top4	21	TFmatrixID_0148	AT-Hook
top5	17	TFmatrixID_0152	AT-Hook
top6	17	TFmatrixID_0136	AT-Hook
top7	15	TFmatrixID_0174	bHLH
top8	13	TFmatrixID_0146	AT-Hook
top9	11	TFmatrixID_0211	C2H2
top10	11	TFmatrixID_0213	C2H2
top11	11	TFmatrixID_0137	AT-Hook
top12	11	TFmatrixID_0138	AT-Hook
top13	11	TFmatrixID_0357	Myb/SANT
top14	10	TFmatrixID_0571	TBP
top15	10	TFmatrixID_0154	AT-Hook
top16	10	TFmatrixID_0188	bZIP
top17	9	TFmatrixID_0159	bZIP;bHLH
top18	9	TFmatrixID_0419	TBP
top19	9	TFmatrixID_0262	GATA
top20	9	TFmatrixID_0570	TBP

xylem

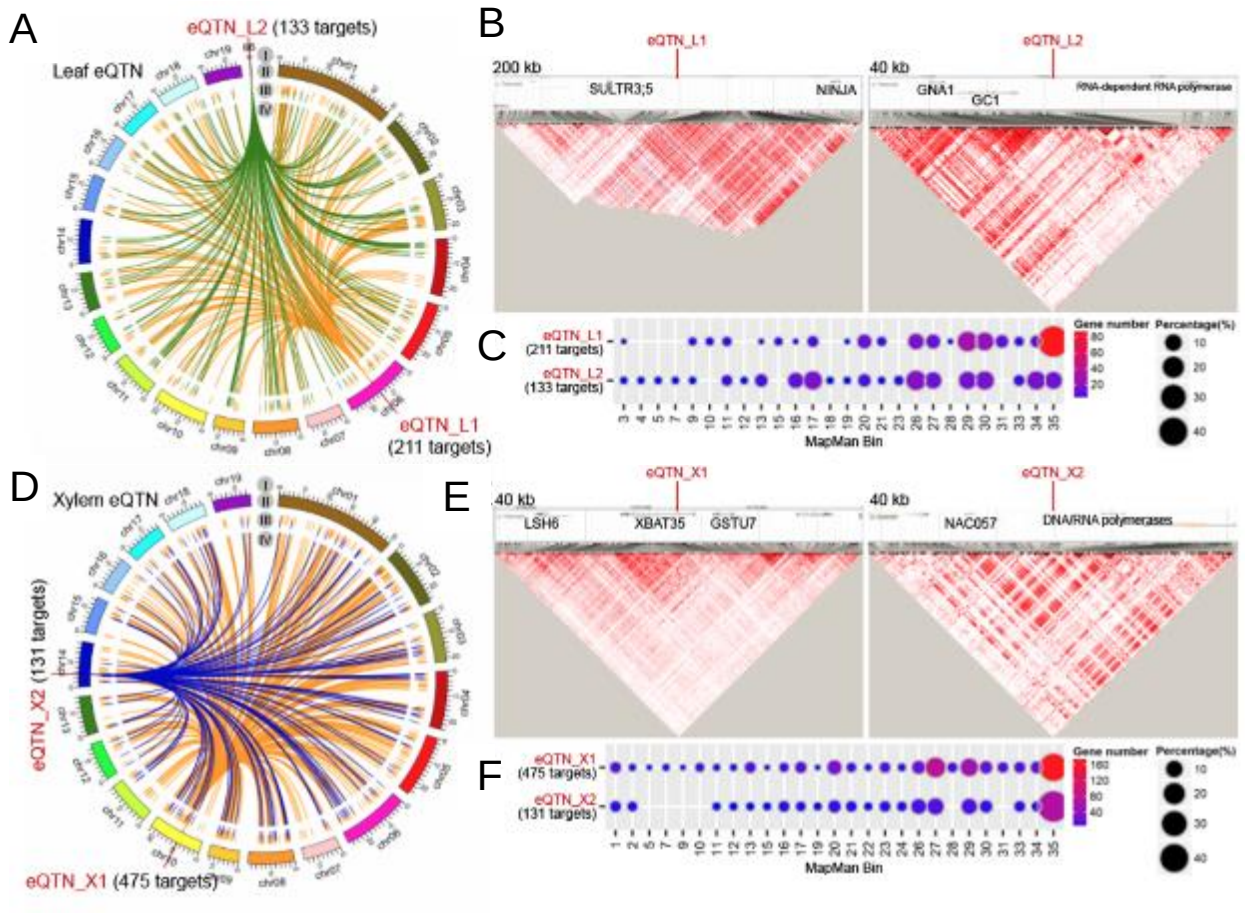
	Counts	TFBS ID	TF family
top1	37	TFmatrixID_0193	bZIP
top2	34	TFmatrixID_0131	AT-Hook
top3	34	TFmatrixID_0148	AT-Hook
top4	28	TFmatrixID_0152	AT-Hook
top5	22	TFmatrixID_0146	AT-Hook
top6	20	TFmatrixID_0136	AT-Hook
top7	20	TFmatrixID_0418	TBP
top8	19	TFmatrixID_0224	TCR
top9	18	TFmatrixID_0174	bHLH
top10	18	TFmatrixID_0211	C2H2
top11	18	TFmatrixID_0213	C2H2
top12	17	TFmatrixID_0154	AT-Hook
top13	16	TFmatrixID_0138	AT-Hook
top14	15	TFmatrixID_0140	AT-Hook
top15	13	TFmatrixID_0137	AT-Hook
top16	13	TFmatrixID_0357	Myb/SANT
top17	13	TFmatrixID_0369	Myb/SANT
top18	13	TFmatrixID_0471	Homeodomain
top19	12	TFmatrixID_0142	AT-Hook
top20	11	TFmatrixID_0292	Homeodomain
top21	11	TFmatrixID_0364	Myb/SANT
top22	11	TFmatrixID_0419	TBP
top23	11	TFmatrixID_0571	TBP

THE CENTER FOR BIOENERGY INNOVATION

trans-eQTN hotspots



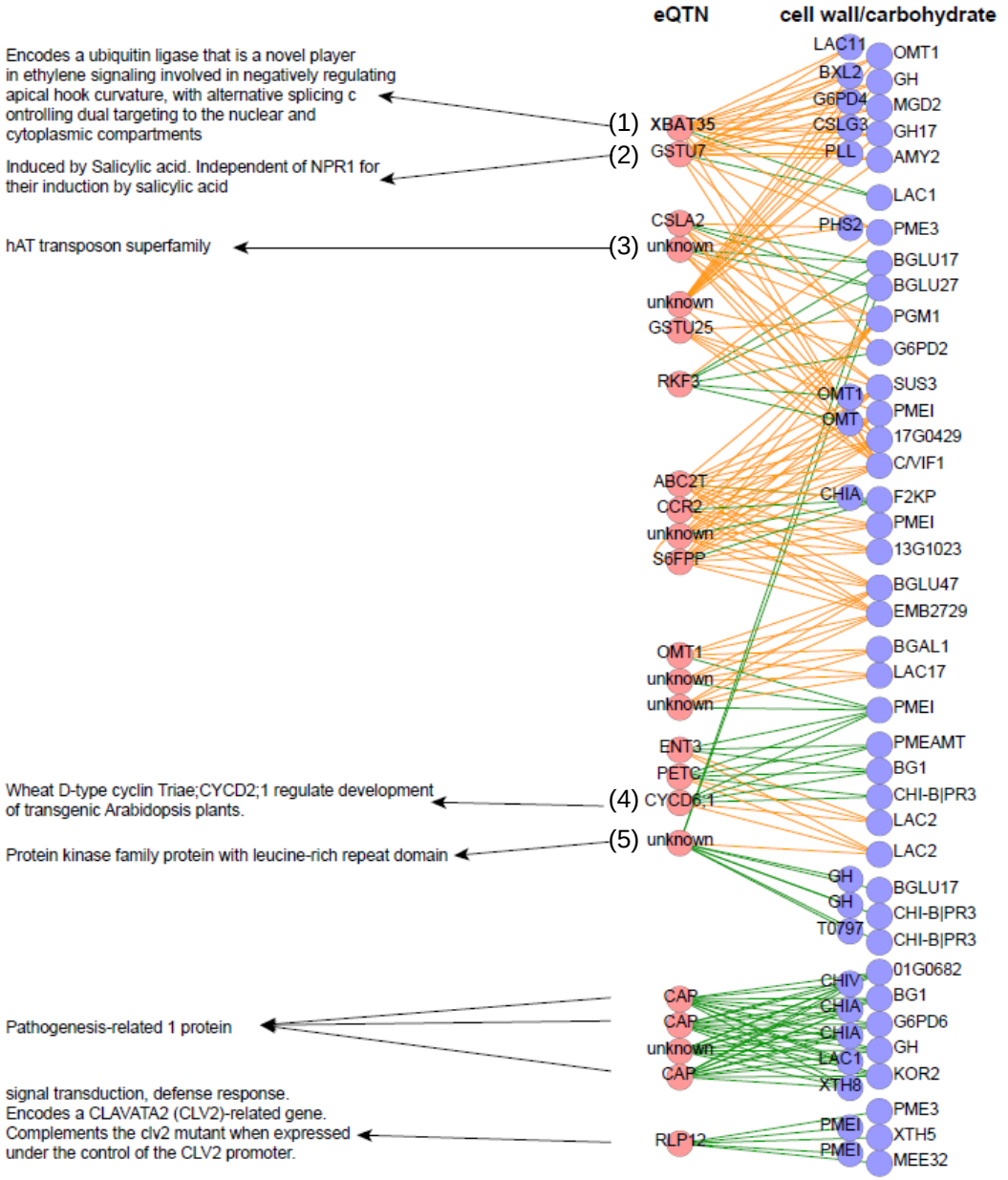
trans-eQTN hotspots



BIN	NAME
1	PS
2	major CHO metabolism
5	fermentation
7	OPP
9	mitochondrial electron transport / ATP synthesis
10	cell wall
16	secondary metabolism
17	hormone metabolism
19	tetrapyrrole synthesis
20	stress
21	redox
22	polyamine metabolism
23	nucleotide metabolism
26	misc
27	RNA
28	DNA
29	protein
30	signalling
31	cell
33	development
34	transport

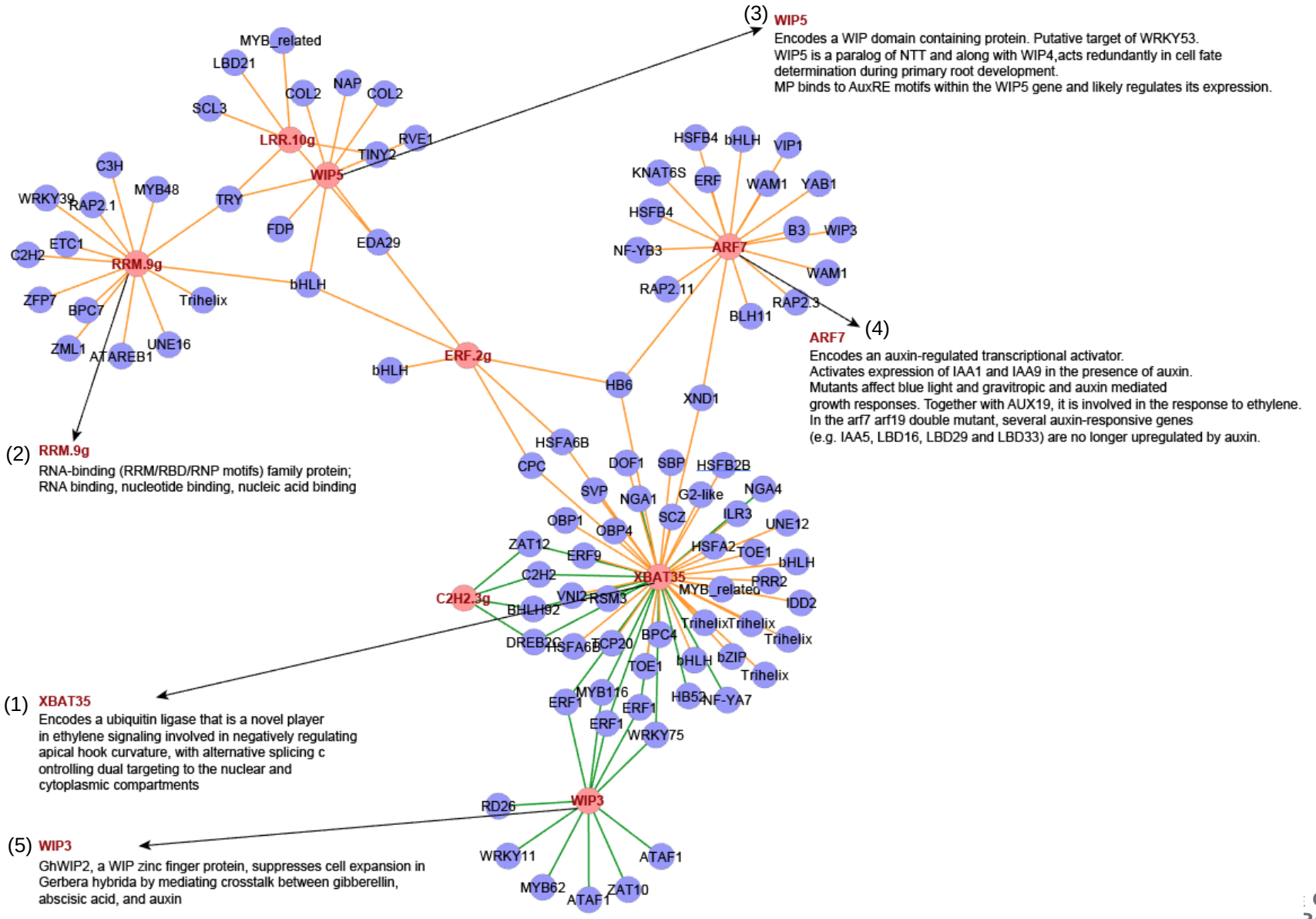
THE CENTER FOR BIOENERGY INNOVATION

cell wall-related eQTN-regulatory network



THE CENTER FOR BIOENERGY INNOVATION

TF-related eQTN-regulatory network



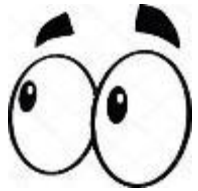
Integration of Phenotypic GWAS and eQTN

Case 1: Drought leaf senescence

A

Score: 0 1 2

GWAS



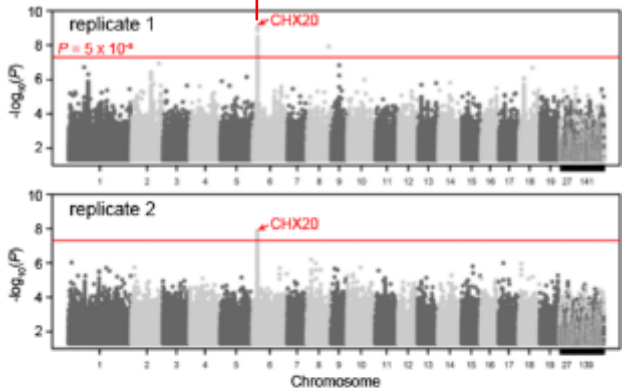
Human eyes
2018

eQTN

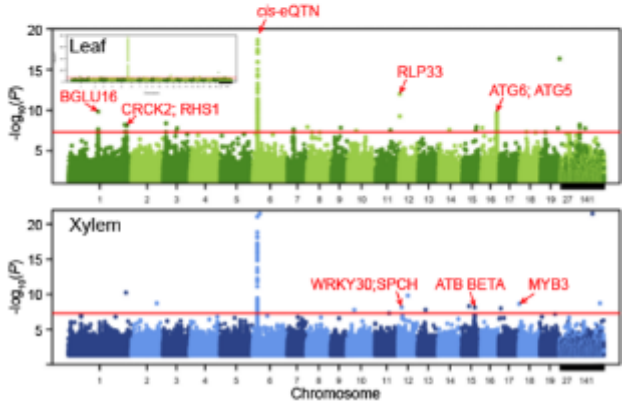
HiSeq
2014

B

cation/H^+ exchanger CHX20: osmoregulation



C



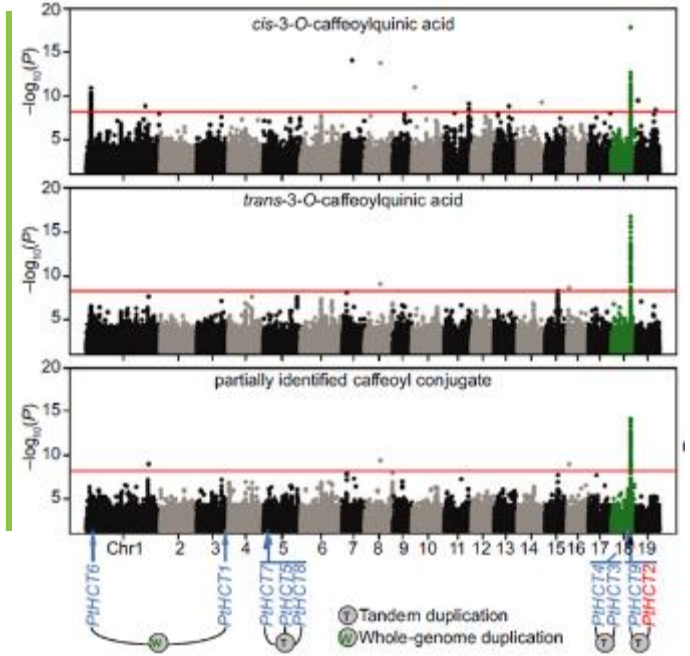
THE CENTER FOR BIOENERGY INNOVATION

Integration of Phenotypic GWAS and eQTN

Case 2: metabolomics

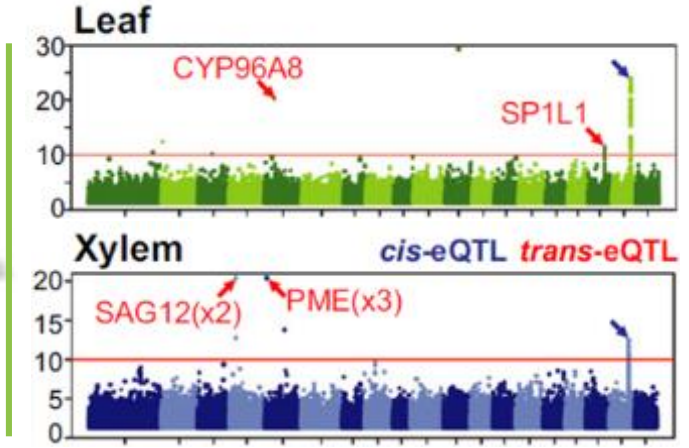
mGWAS A

GC-MS
2012

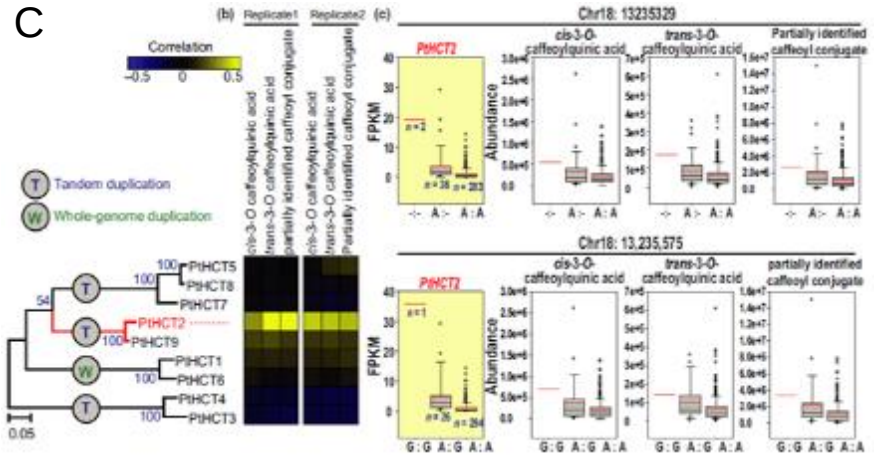


eQTN B

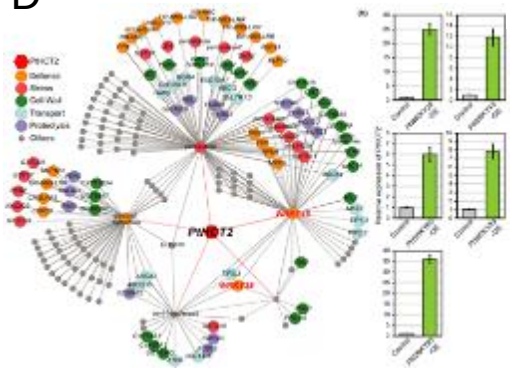
HiSeq
2014



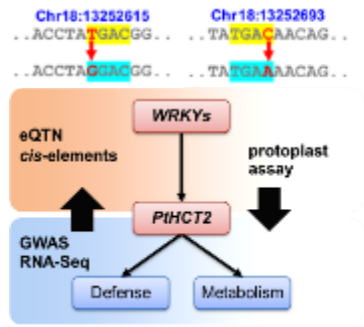
C



D



E



Zhang, J. et al. (2018). GWAS and eQTN analyses reveal roles of *HCT2* in caffeoylquinic acid biosynthesis and its regulation by defense-responsive transcription factors in *Populus*. ***New Phytologist***, 220(2), 502-516.

THE CENTER FOR BIOENERGY INNOVATION

Integration of Phenotypic GWAS and eQTN

Case 3: ionomics

iGWAS

ICP-MS
2012

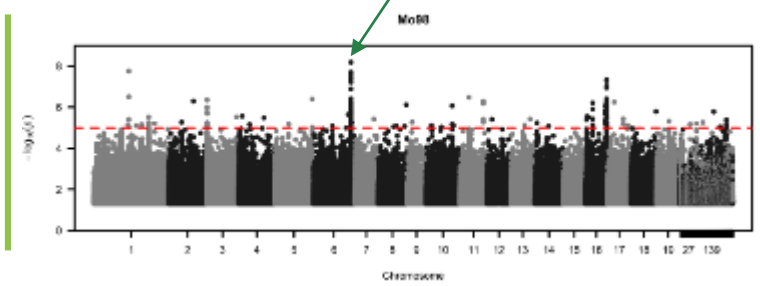
eQTN

HiSeq
2014

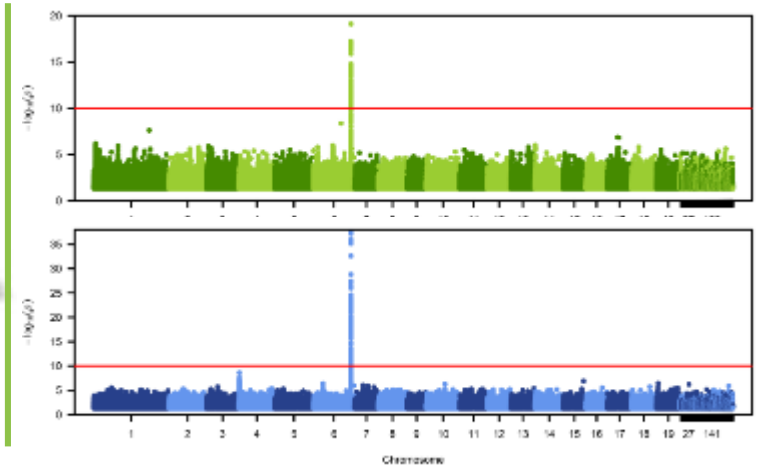
molybdate

A

molybdate transporter 1 (MOT1)

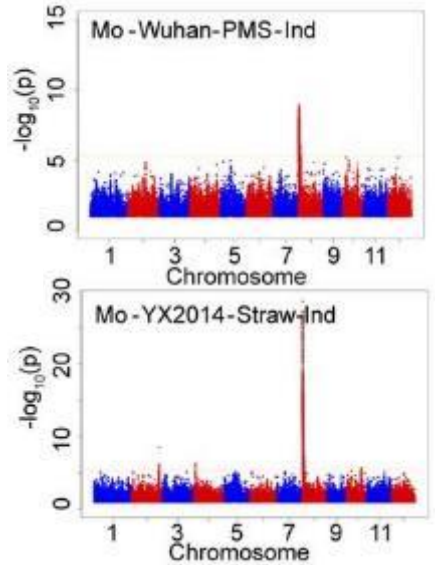


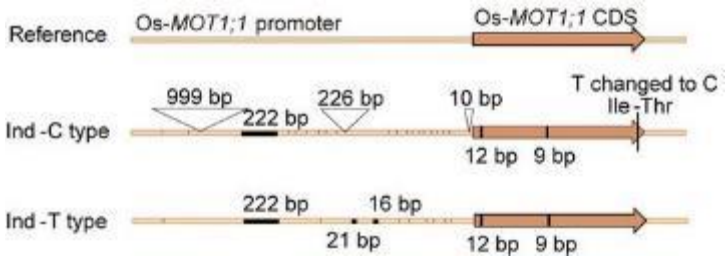
B



C

rice



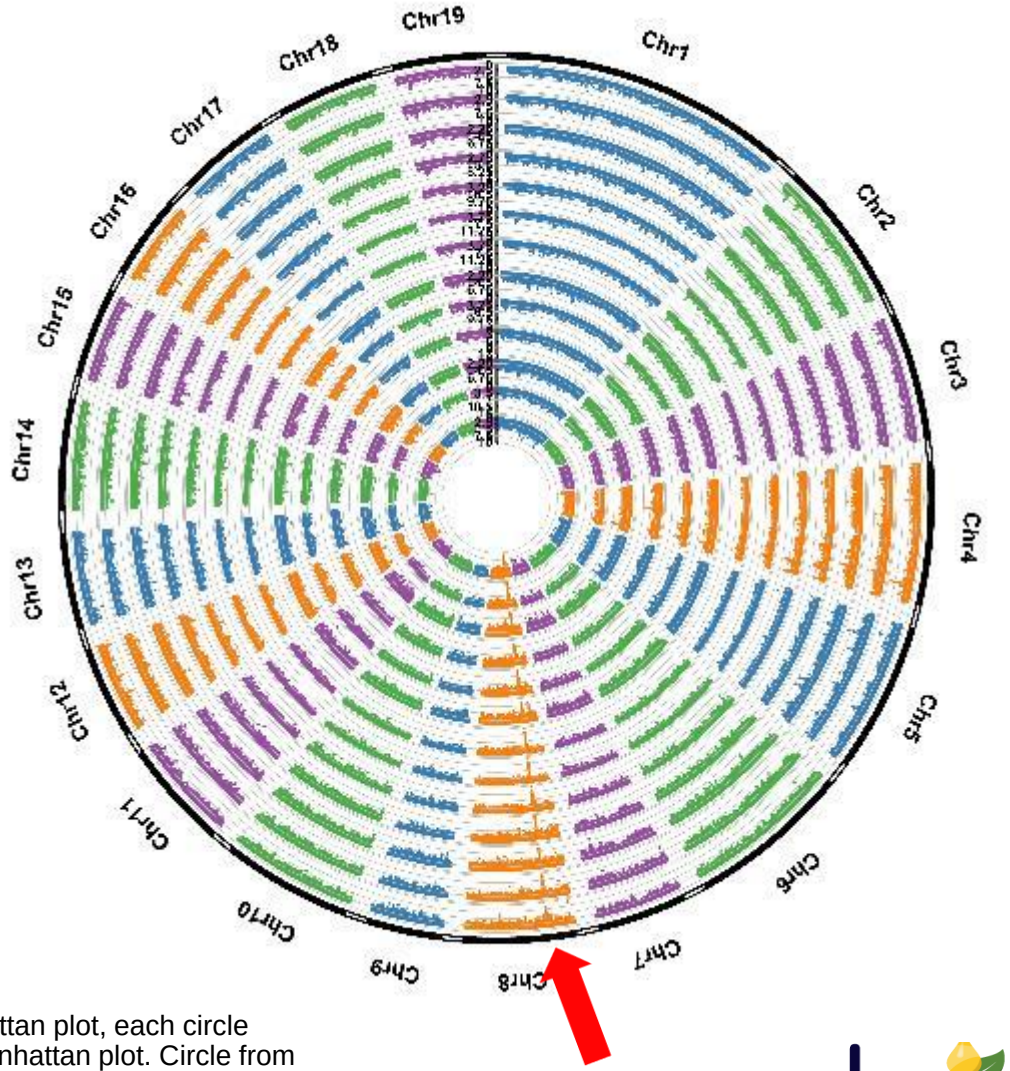


Yang M, et al. (2018) Plant Cell

THE CENTER FOR BIOENERGY INNOVATION

Integration of Phenotypic GWAS and eQTN

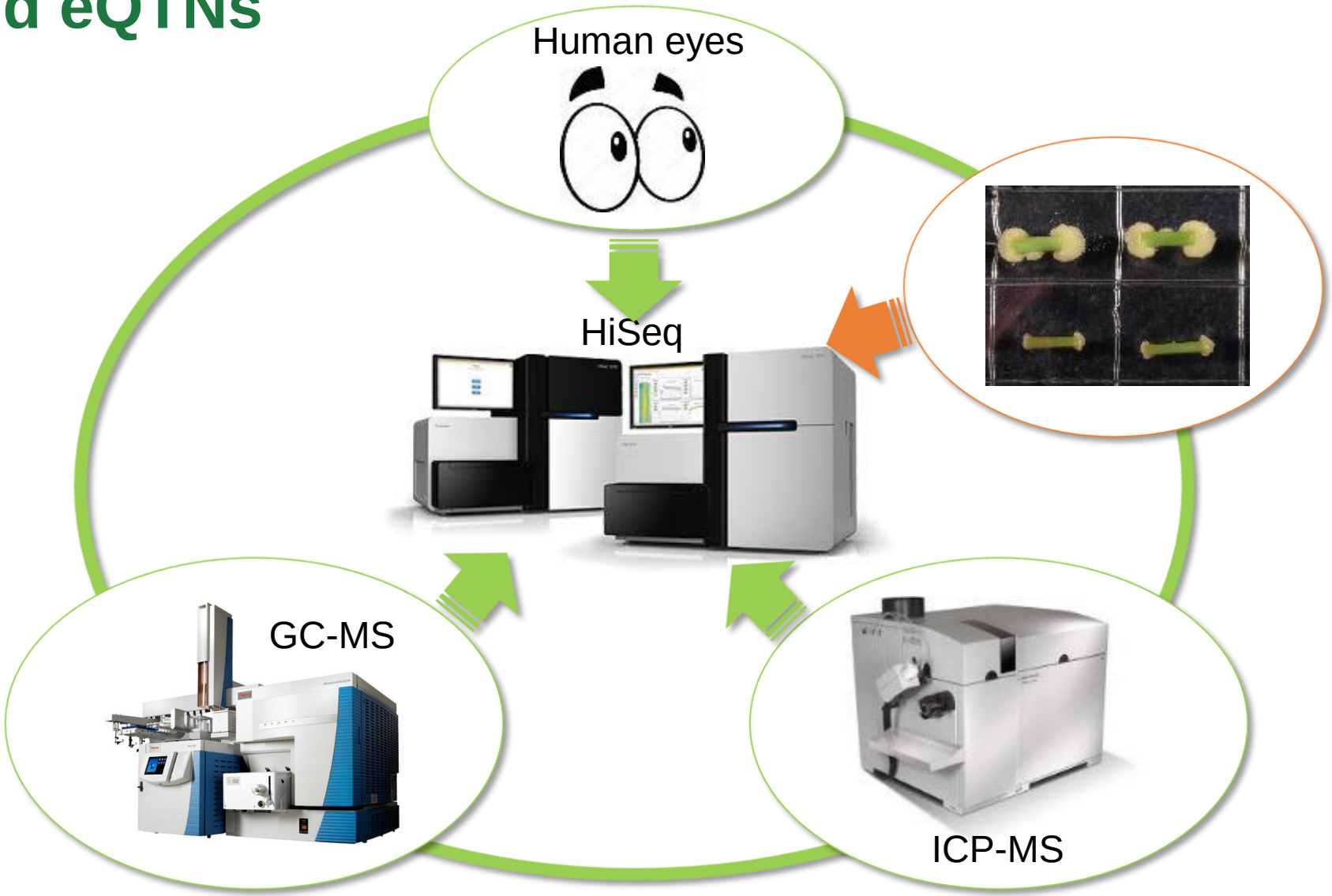
Case 4: Bud break



Circular Manhattan plot, each circle represent a Manhattan plot. Circle from outer layer to inner layer indicates CA2010 (rep1-rep3), CA2013 (rep1-rep3), CL2010(rep1-rep3), CO2010(rep1-rep3) and CO2012(rep1).

THE CENTER FOR BIOENERGY INNOVATION

Future plans: Integrating regeneration GWAS and eQTNs



THE CENTER FOR BIOENERGY INNOVATION

Acknowledgements

ORNL Plant Systems Biology Group: Tony Bryan, Olaf Czarnecki, Lee Gunter, Sara Jawdy, Udaya Kalluri, Jessy Labbe, Raja Payyavula, Priya Ranjan, Jerry Tuskan

ORNL Metabolomics Group: Nancy Engle, Martin Madhavi, Tim Tschaplinski

WVU: Stephen DiFazio, Luke Evans

GreenWood: Brian Stanton, Austin Himes, Carlos Gantz, Kathy Haiby

USDA Forest Service, Institute of Forest Genetics: Valerie Hipkins, Jennifer DeWoody, Tom Blush

JGI: Vasanth Singan, Erika Lindquist, Christa Pennacchio

Thank you for listening

Scheduled 15 minute break

Broader Impacts: Education and Curricula

Jay Well, Assistant Director, SMILE Program
Troy Hall, Professor and Department Head, Forest Ecosystems & Society
Betsy Emery, Graduate Research Assistant, Forest Ecosystems & Society

Today's Presentation:

Update on Curriculum Development – Jay Well

- What we've done
- Changes we made from the original proposal
- Next steps--summative assessment
- Milestones and impact
- Dissemination and publication

Update on Social Science – Troy Hall

- Initial data collection process
- Results from initial data collection
- Next steps – testing final curriculum
- Anticipated publications/presentations

Context of the Study:

Education Side

- Overall lack of STEM engagement among students in the US
- Decreasing public scientific literacy + increasing science complexity
- Increasing gaps between scientific and public understandings of science

Social Science Side

- GM/GE is controversial among adults in the U.S.
- Most studies focus on adult attitudes - what about youth?
- Youth are future decision makers/leaders

Overall Goal: Increase open-minded deliberation about socio-scientific issues around GE Ag

Photo Credit: DesignSpace

Specific Goals Outlined in Proposal:

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

Proposed broader impacts approach…

Partner with SMILE – the Science and Math Investigative Learning Experiences program at Oregon State University

- An after-school science and math club for students in grades 4-12 that focuses on increasing STEM literacy among underrepresented students in rural communities across Oregon

Work with SMILE and the biophysical science team to develop two case studies (2-3, 50-minute lessons each) about genetic modification for high school students

Evaluate and improve curriculum in an iterative process in partnership with SMILE teachers

- Pilot curriculum in SMILE after-school science clubs
- Evaluate curriculum with formative assessment to refine and improve lessons
- Disseminate final lessons broadly through teacher networks

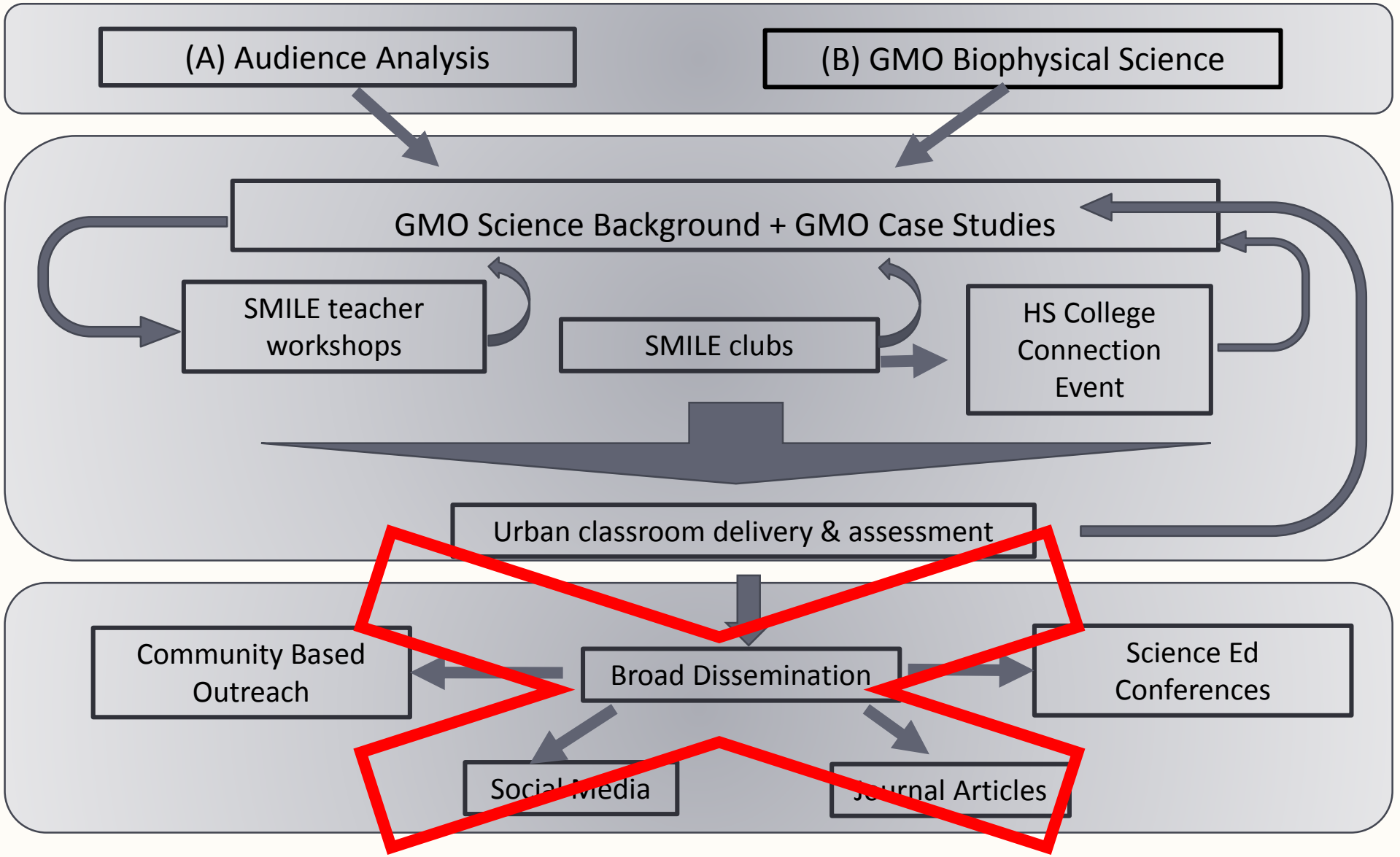
Broader impacts overview

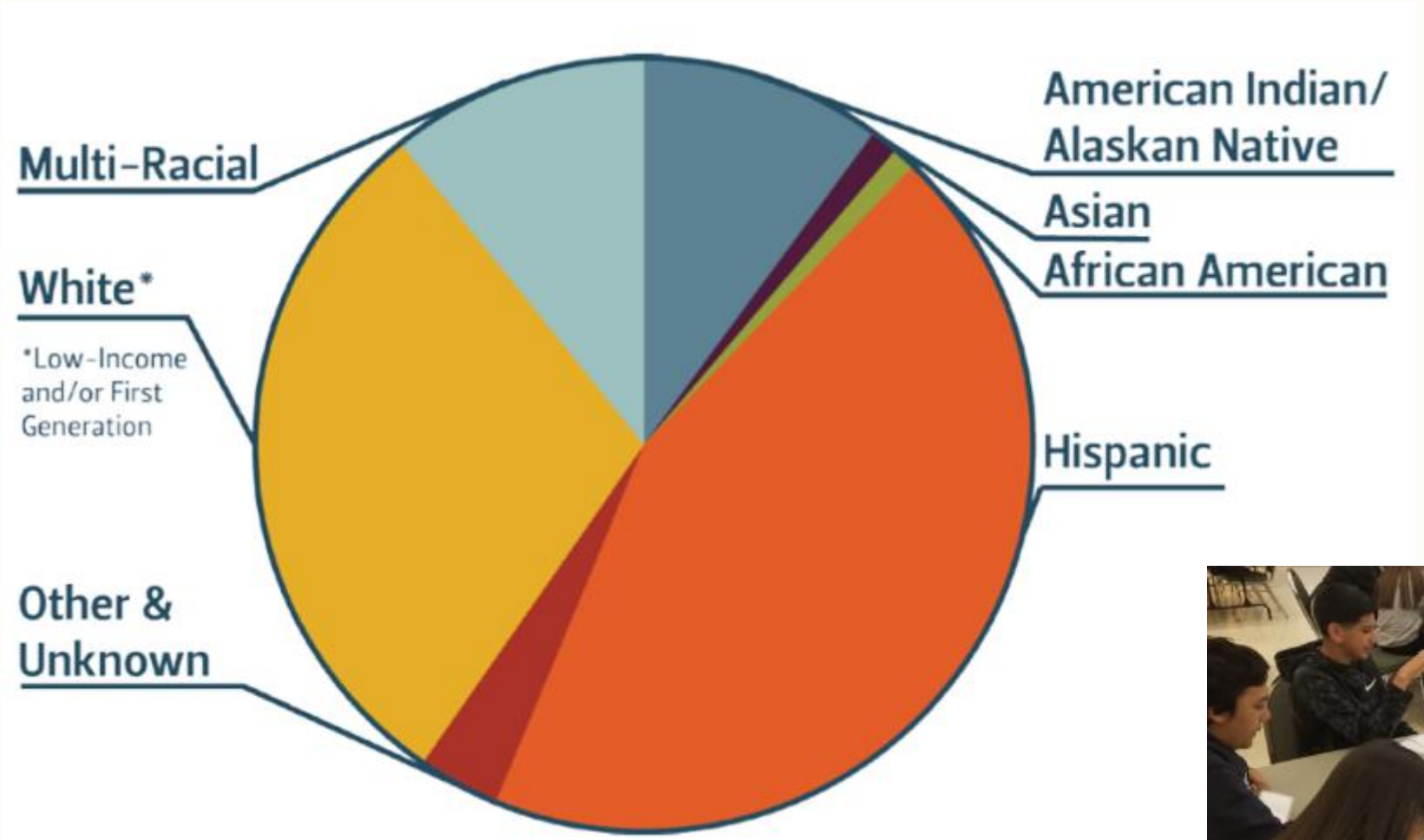
Initial Assessment: Audience Assessment
(COMPLETE)

Formative Assessment: Develop, Assess, Improve, & Evaluate Case Studies
(NEAR COMPLETION)

Summative Assessment:
(IN PROGRESS)

Disseminate
(LIMITED IN SCOPE DUE TO DECREASED NSF FUNDING)





SMILE serves diverse students in rural communities

707 students (HS 202, MS 322, ES 183)

- **58% female,**
- **66% low-income**
- **81% first generation to college**

Initial Assessment: Audience Assessment (Complete)

Surveyed ~40 SMILE teachers to explore existing cognitions (knowledge, beliefs, emotions, attitudes) regarding genetic modification (April 2018)

- Self-assessed knowledge about GE/genetic modification is low
- Many teachers are neutral or don't know about various GE issues
- Teachers with opinions are overall more positive than negative
- Many teachers felt like they weren't knowledgeable enough to teach GE material

Initial Assessment: Audience Assessment (Complete)

Facilitated **4 focus groups** to understand teachers' comfort presenting GE material, strategies they use and challenges they face in teaching controversial material, and their perceptions of student knowledge/attitudes toward GE/GMOs

- Teachers are comfortable using socially controversial material in their classes
- Students have vague understandings of GE/GMOs and do not tend to have strong attitudes about GE/GMOs
- Teachers face a variety of challenges in teaching GE material
  - Lack of detailed knowledge about the topic
  - Low scientific literacy among students requires teachers to provide lots of background to the topic
  - Difficulty finding materials at the appropriate level for their students

Initial Assessment: Content Selection (Complete)

- Reviewed literature about effective techniques to increase student engagement and ability to think critically; incorporated those into case study design
- Reviewed existing case studies and curriculum to understand gaps and how our curriculum can address these gaps
- Coordinated with GMO biophysical science experts on content and activities included in case studies to ensure they are factually correct

Formative Assessment: Content Development (Near completion)

Audience assessment → switch from a case study focus to broader curriculum

Developed 3 introductory lessons (digital and scientific literacy focus) and 5 lessons about GE

Piloted lessons in SMILE afterschool clubs

Used teacher feedback to improve case studies, which will be used in summative assessment

NOTE: The eight one-hour lessons we developed for this grant are available on the "Broader Impacts" page at: http://people.forestry.oregonstate.edu/steve-strauss/genes-affecting-plant-regeneration-and-transformation-poplar

Summative Assessment (In Progress)

Partnering with OSU grad programs: M.S. in Education (Math or Science) and M.S. Agricultural Education

- Curriculum will primarily be implemented by student teachers instead of project staff.
- Project staff will train student teachers on lessons, data collection protocol
- Increases sample size, control over delivery, and efficiency of the project

Will introduce curriculum in high school science classes as part of student teacher placements

- Schools in Portland and other OR communities outside of Corvallis
- Max enrollment: 1,200 high school students (10 student teachers x 4-6 classes x 30 students per class)

Outline of Summative Assessment

Pre-Survey (knowledge and attitudes) → 120 minutes of classroom instruction → Post-Survey (knowledge and attitudes)

Concept map of GE ag

Lesson 1: Eras of Crop Improvement – Overview of plant modification in ag

Lesson 2: Fact Checking in the Digital Age – Overview with specific GE ag examples

Lesson 3: Why Genetically Modify – Three case studies of GE ag

Revisit initial GE ag concept map

Changes in Education Project From Proposal

- Shifted from developing case studies about specific GE products to developing multiple lessons geared towards better understanding the science and perspectives
- Only provided a pre-test during data collection in SMILE clubs due to challenges in club timing and controlling for delivery, attendance
  - Will be using pre/post surveys for data collection during summative assessment
- Working with student teachers to conduct summative assessment in addition to project staff to increase efficacy and efficiency of outcomes
  - Will be training student teachers in advance during Fall/Winter
  - Will allow for increased sample size and greater demographic/geographic footprint of the project

Milestones

Curriculum Development:

1. Fact Checking in the Digital Age
2. Methods of Food Modification
3. GMOs and the Nature of Science
4. Eras of Crop Improvement
5. Investigating the GMO Controversy
6. Why Genetically Modify?
7. A Better Banana
8. GE Labeling and Identification

Teacher Professional Development:

August 2018 January 2019 August 2019 January 2020

High School/Middle School Challenges:

- April 2019
- Modifying associated lessons into a 5-day applied genetics MS/HS curriculum

Summative Assessment:

- School Year 2019/2020: Recruiting schools, teachers, delivery winter/spring

Impact by the numbers (through Aug 2019):

- **Total SMILE Teachers: 125 for 546 contact hours (Workshops and Challenges)**
  - Teacher workshops: 88 SMILE Teachers for 264 contact hours
  - MS teachers at challenges: 22 for 132 contact hours (challenge is 6 hours)
  - HS teachers at challenge: 15 teachers for 150 contact hours (challenge is 10 hours)
- **Total SMILE Students (Challenges): 200 for 1504 contact hours**
  - MS students at challenges: 124 for 744 contact hours (challenge is 6 hours)
  - HS students at challenge: 76 students for 760 contact hours (challenge is 10 hours)
- **Total SMILE Students (Clubs):** 420 for 840 contact hours
  - Assuming each student participated in 2 of 6 lessons

Lessons piloted in multiple SMILE clubs

Lesson	MS Clubs	HS Clubs	Total Clubs	Estimated Students Reached (12 students/club)
Fact Checking in an Era of Fake News	5	7	12	144
Nature of Science	4	3	7	84
Methods of Food Modification	7	5	12	144
Eras of Plant Improvement	3	7	10	120
Why Genetically Modify?	6	6	12	144
Investigating the GMO Controversy	3	6	9	108

Dissemination and publication

- Feedback from multiple sources will be used to refine lessons and put in finalized form
- Look to publish lessons in teacher practitioner journal such as National Science Teacher Associations' *Science Scope* or *Science Teacher*
- Presented the first lesson ("Fact Checking…") at the Oregon Science Teachers Association annual conference in October 2018 (31 K-12 teachers participated)
- Further disseminate lessons through partner teacher networks and conferences

Broader Impacts: Social Science

Troy Hall, Professor and Department Head, Forest Ecosystems & Society
Betsy Emery, Graduate Research Assistant, Forest Ecosystems & Society
Jay Well, Assistant Director, SMILE Program

Theoretical Framework:

Overall goal:

increase students' cognitive complexity

move students from heuristic-based, simple "decisions" about GMOs to more nuanced attitudes based on balanced consideration of multiple potential positive and negative aspects

Research Questions:

***Initial phase*:** What knowledge, beliefs, and attitudes do science club students in rural Oregon have about genetic engineering (GE) and genetically modified foods (GMF)?

***Summative phase*:** How do carefully designed curricular materials affect Oregon high school students' beliefs, attitudes, and cognitive complexity regarding GMF?

Initial Phase: Data Collection in SMILE clubs (Complete)

- Baseline knowledge and attitudes
  - 9 SMILE clubs (5 MS, 4 HS)
  - 125 middle school (MS) and high school (HS) students
  - 73 surveys
  - 102 concept maps
- Developed and refined data collection methods to characterize student cognitions
  - Data collection instruments (e.g., concept mapping protocol, pre-test/post-test survey) will be revised for summative evaluation

Initial Data Collection Methods

Concept Mapping

(Data Analysis Complete)

Focus Question:

What are your thoughts and feelings about genetically modifying the foods humans eat?

Online Survey

(Data Analysis Complete)

Content areas:

Beliefs/knowledge; attitudes

Variables, Measurement, and Analysis

Variable	Measurement	Analysis
Belief Structure	Concept maps	Scoring of map structure & content (topics, elaboration)
Knowledge	6 survey statements: self-assessed knowledge about GE and GMF	Knowledge Index Score
Beliefs	7 survey statements: (dis)agreement with claims about GE and GMF; concept map nodes	Descriptive statistics
Attitudes	3 survey statements: attitudes toward GE applications; concept map valence	Attitude Index (mean level of support); Frequencies (concept maps)

Concept Maps:

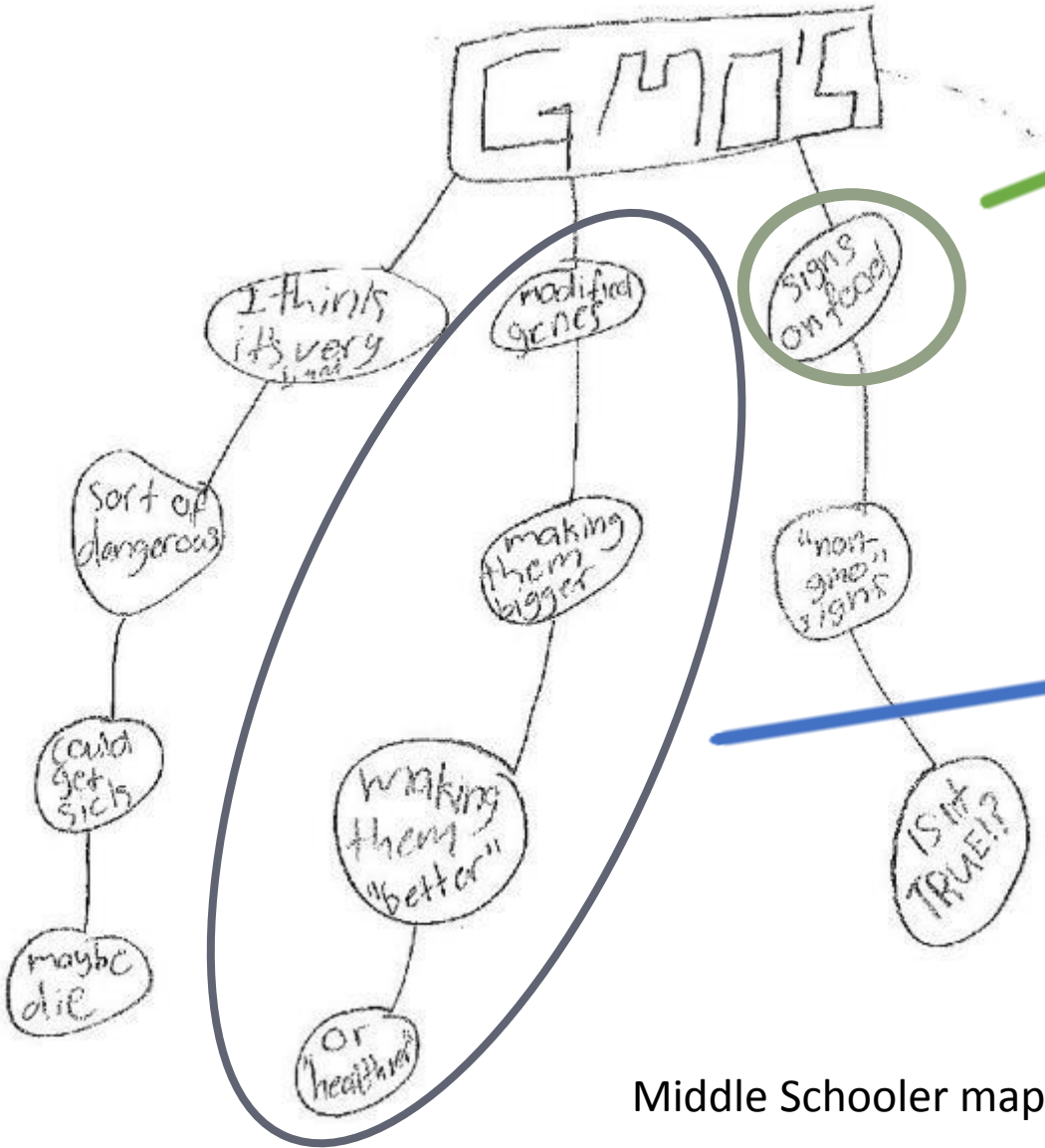
Procedure:

- 15 minutes training on a different topic (social media)
- Instructions seeded 20 concepts in a word bank (e.g., “benefits,” “costs,” “environmental impacts,” etc.)

Analysis to answer 3 questions:

- What are students' unprompted beliefs about GMO foods?
- How are those beliefs structured?
- What are the students’ attitudes toward GMO foods?

Concept Maps – 3 levels of "coding"



Middle Schooler map

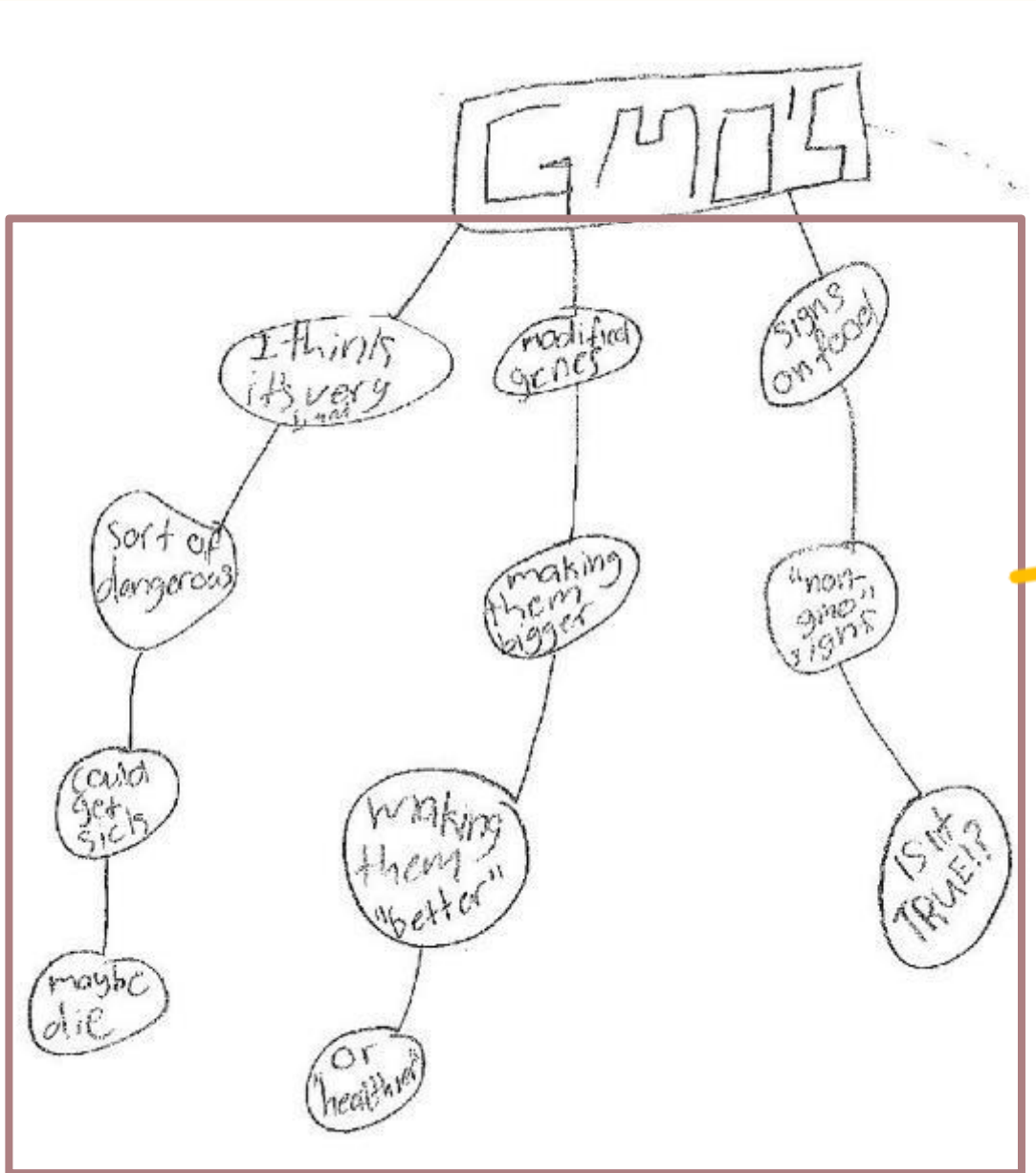
First level: Concept

- **Code**
  - 21 separate topic codes
  - E.g., costs, science, chemicals, agriculture
- **Valence**
  - positive, negative, ambivalent, unknown

Second Level: Cluster

- **Valence**
- **Elaboration** (depth of thinking)

Concept Maps – 3 levels of "coding"



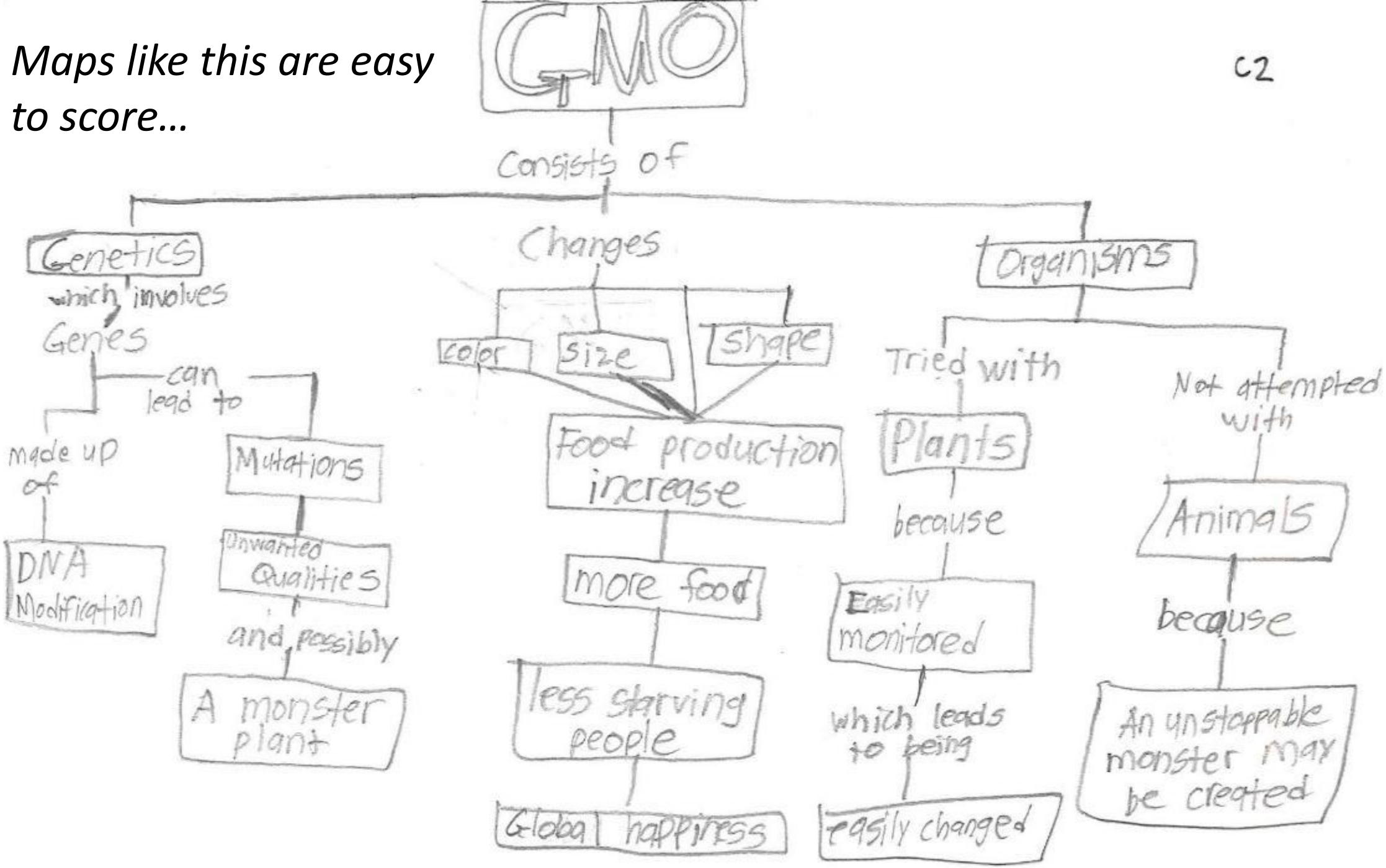
Third Level: Map

- **Map Type (belief structure):**
  - chain, spoke, network
- **Cross-links:**
  - Number of connections between clusters
- **Overall Map Valence:**
  - 0: no apparent valence
  - 1: all positive
  - 2: more positive than negative
  - 3: ambivalent
  - 4: more negative than positive
  - 5: all negative
- **Word Bank Words:**
  - # Words used from the word bank (0-20)

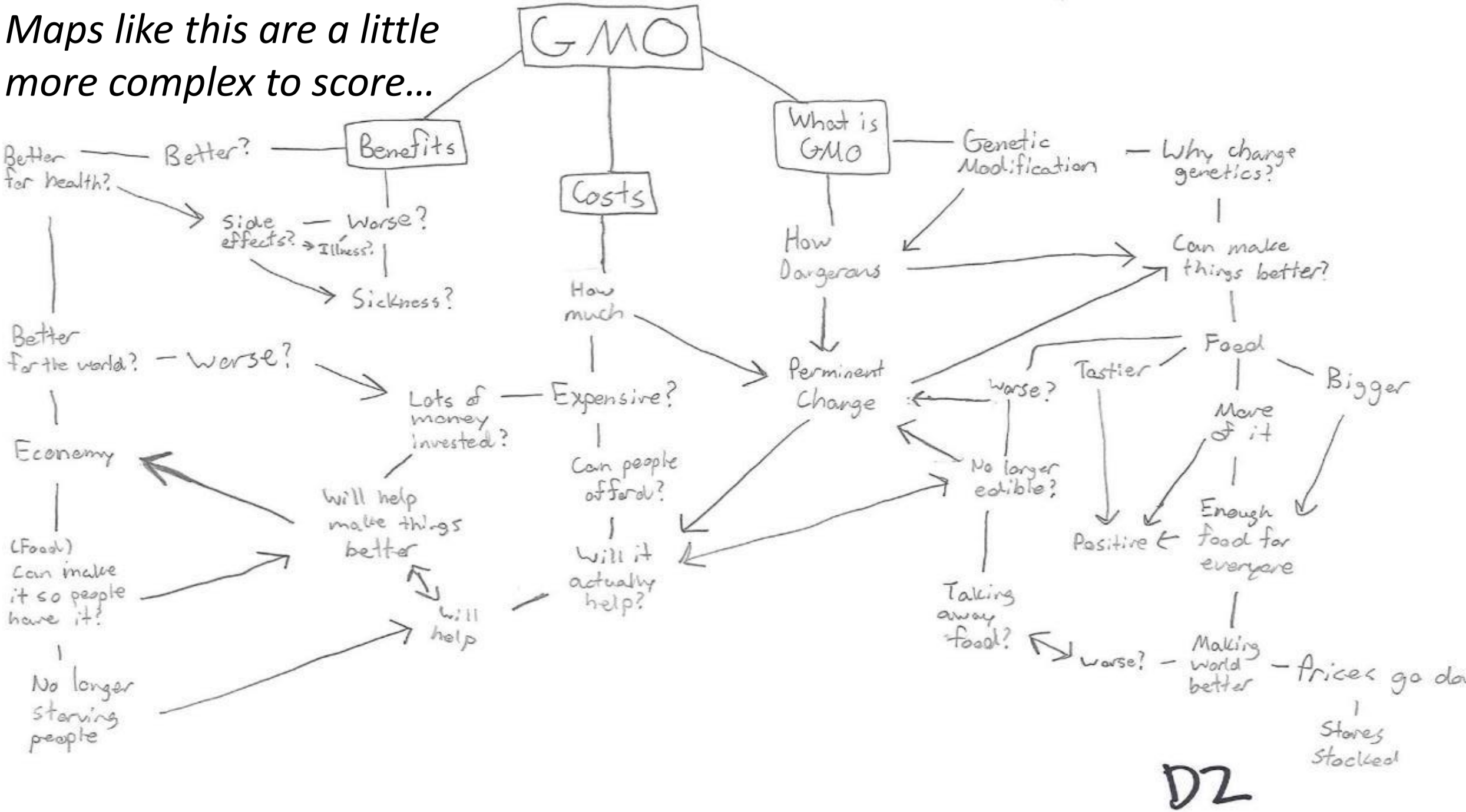
Challenges with Concept Maps

Maps like this are easy to score…

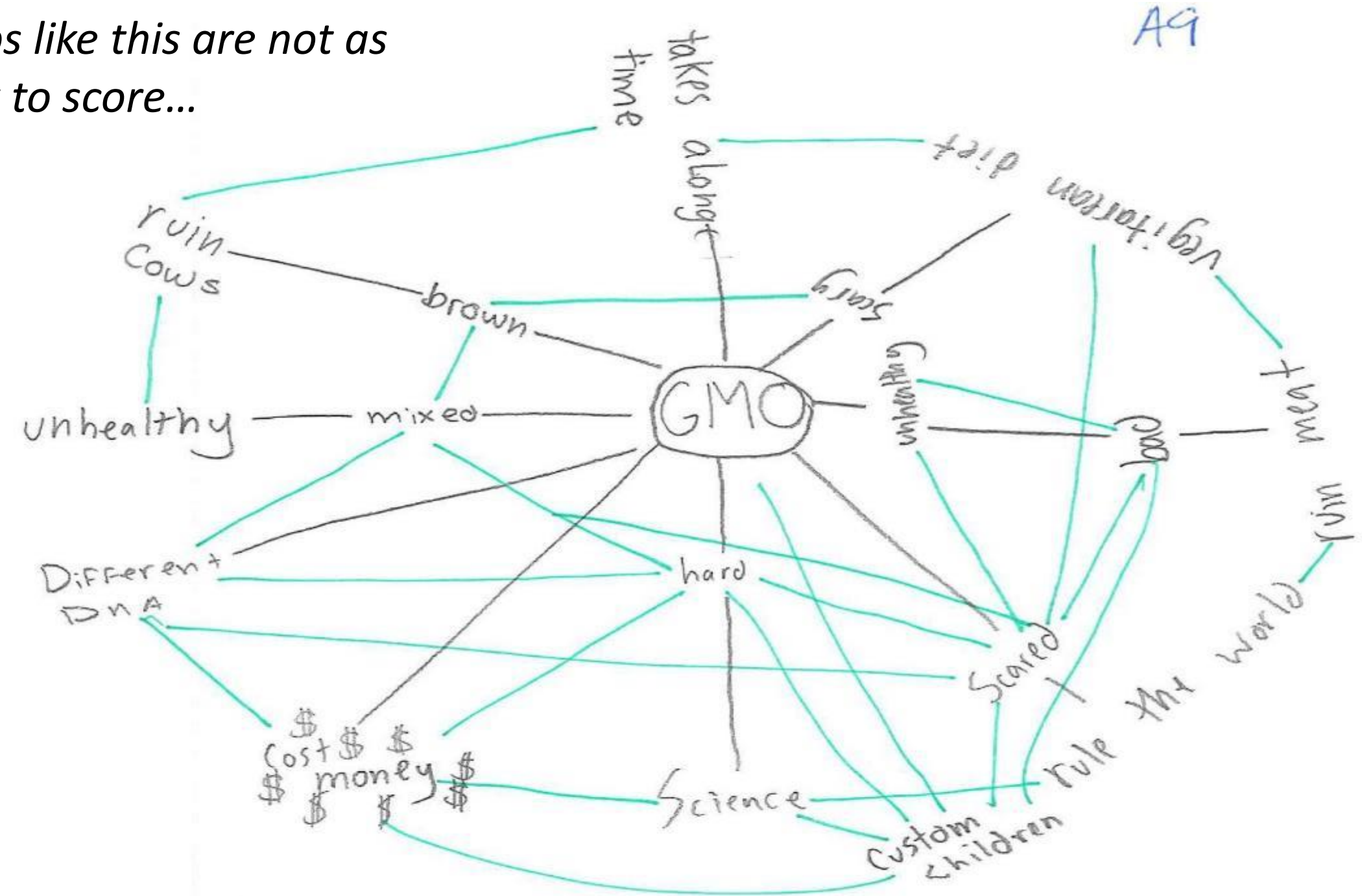
C2



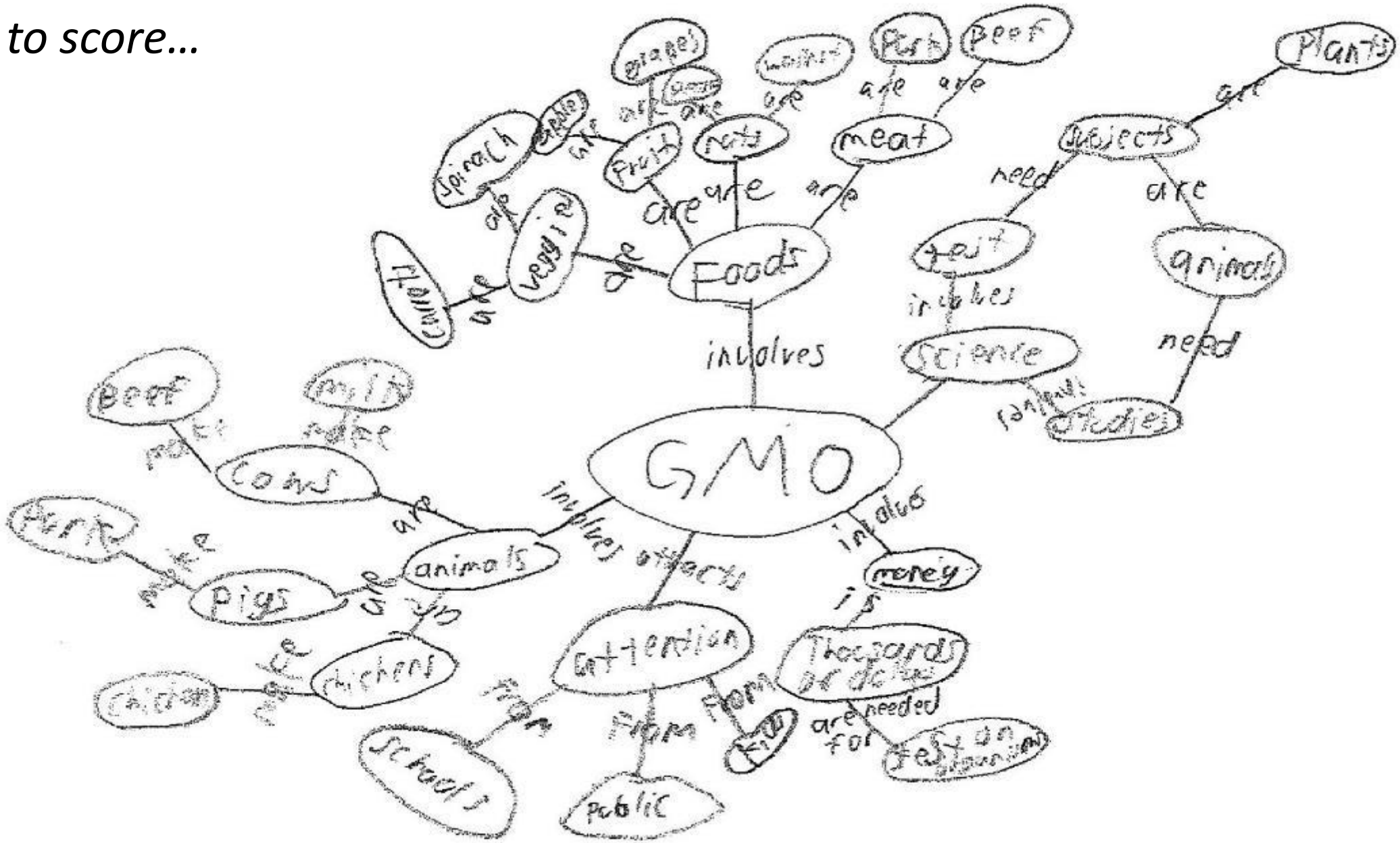
Maps like this are a little more complex to score…



Maps like this are not as easy to score…



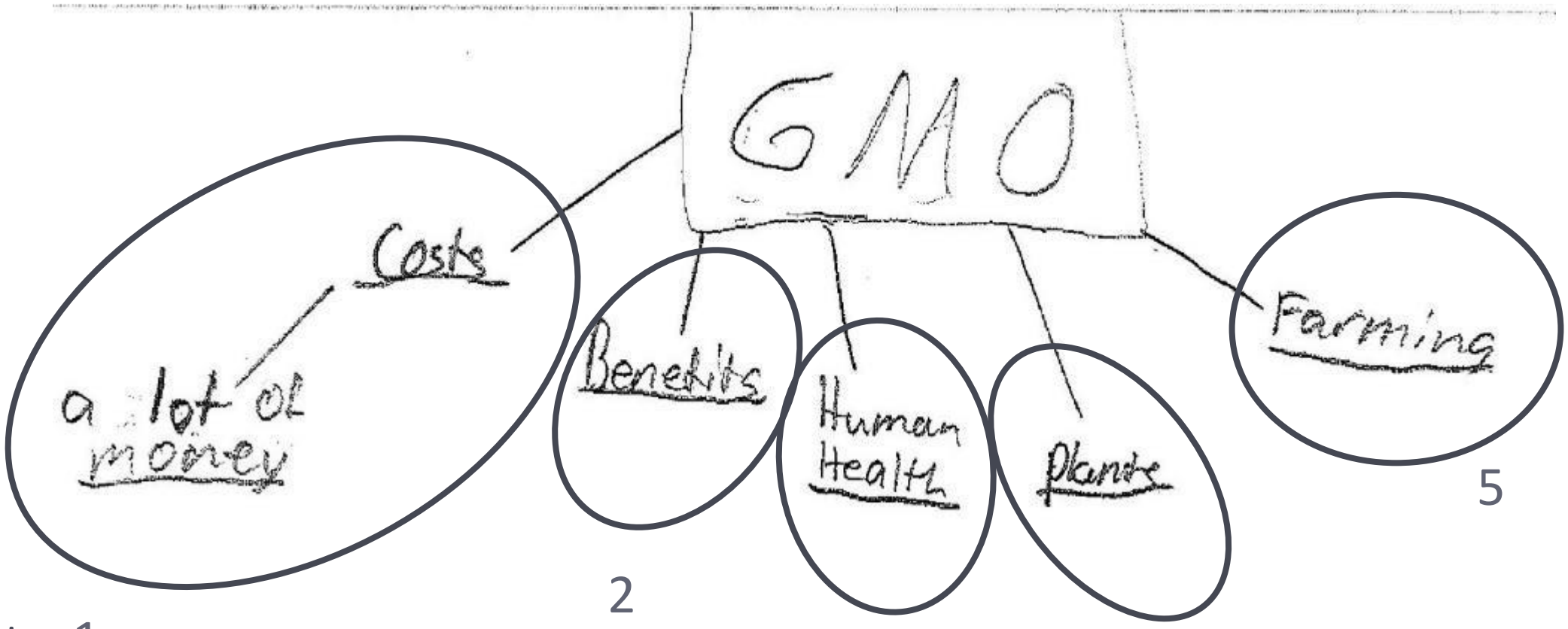
Maps like this are not as easy to score…



Concept maps: What are students' beliefs about GMOs?

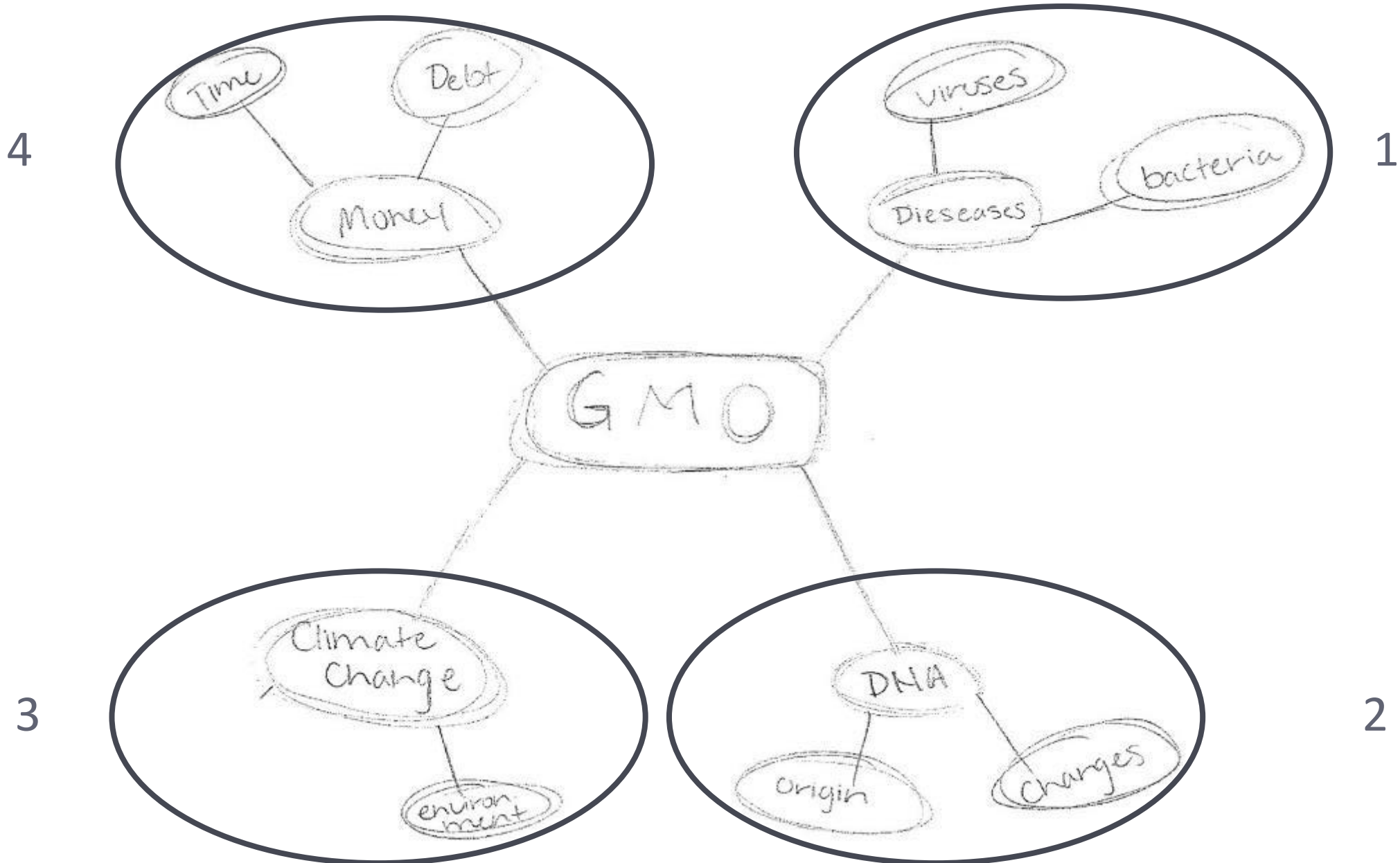
- Many students created a “base map” → not a topic of interest or familiarity?
- Several concepts were common (% of maps; red = word bank):
  - Food (~35% positive; ~10% negative)
  - Trait modification (~25% positive; ~5% negative)
  - Human health (~25% positive; ~40% negative)
  - “Chemicals” (0% positive; ~15% negative)
  - Environmental impacts (~7% positive; ~15% negative)
  - Cost (~10% positive; ~20% negative + ~25% “costs a lot”)
  - Feelings (~25% of maps; more neg than pos)

Concept maps: How much do students elaborate their thoughts on GMOs?

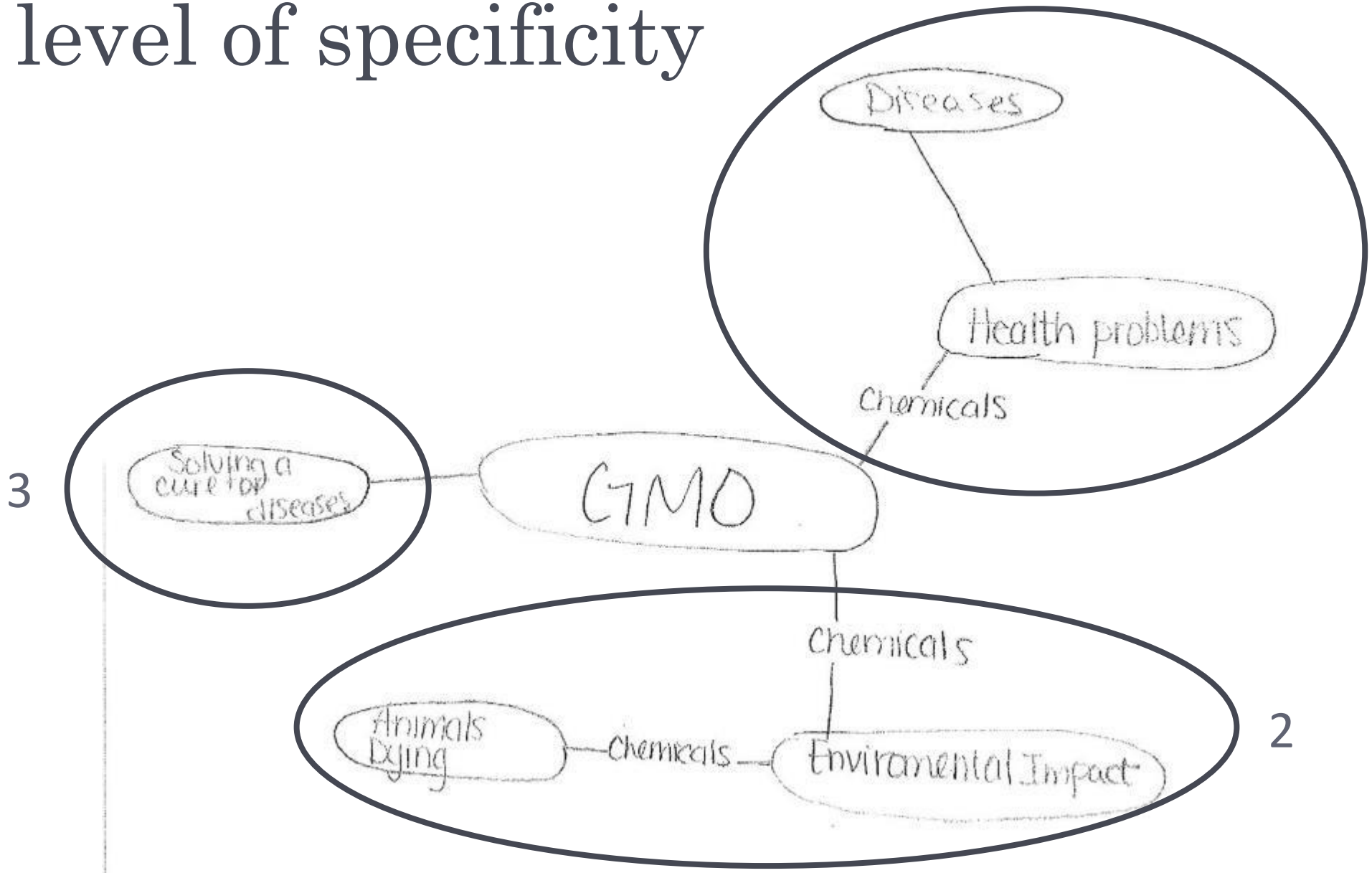


NOTE:
Elaboration is coded at the cluster level

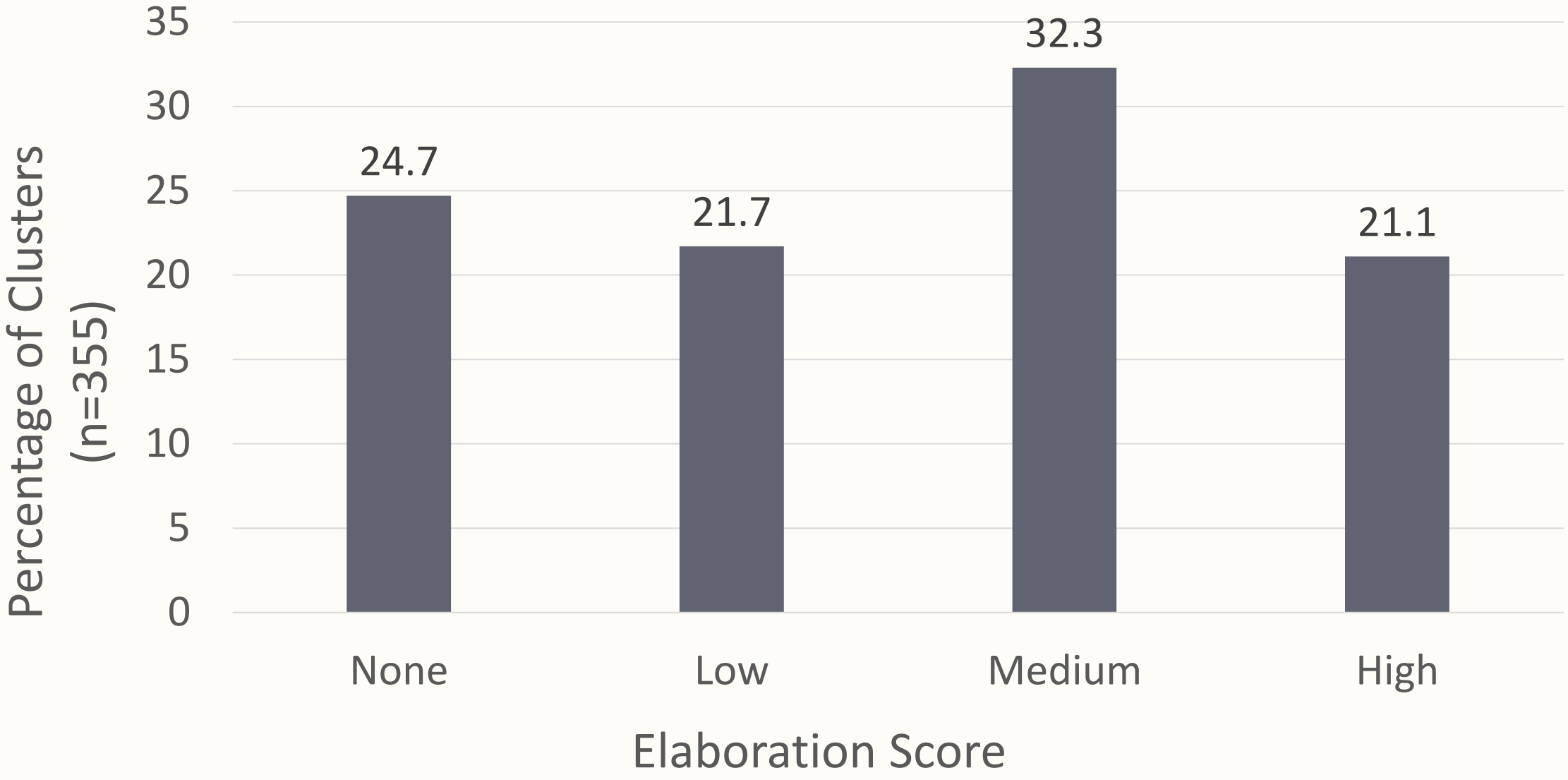
Differences between low and medium elaboration scores



Medium and high scores really distinguished by level of specificity

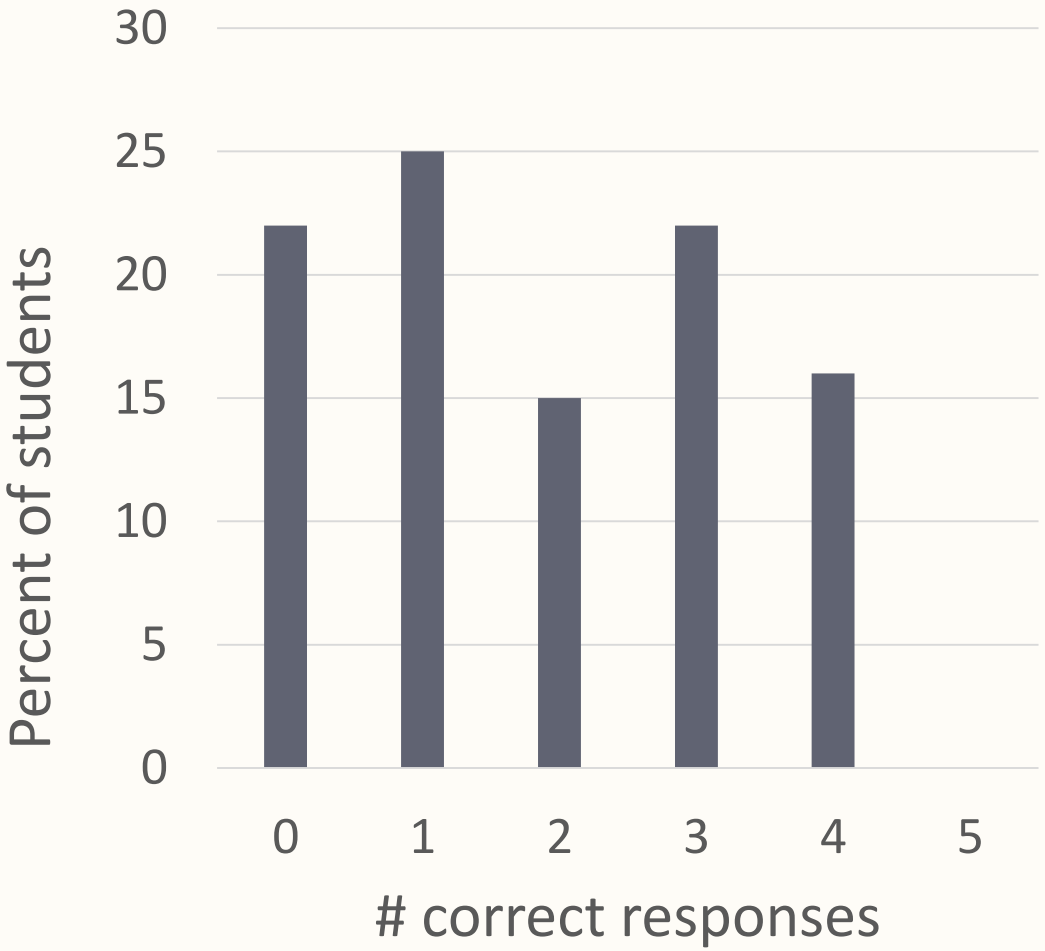


Concept maps: elaboration scores



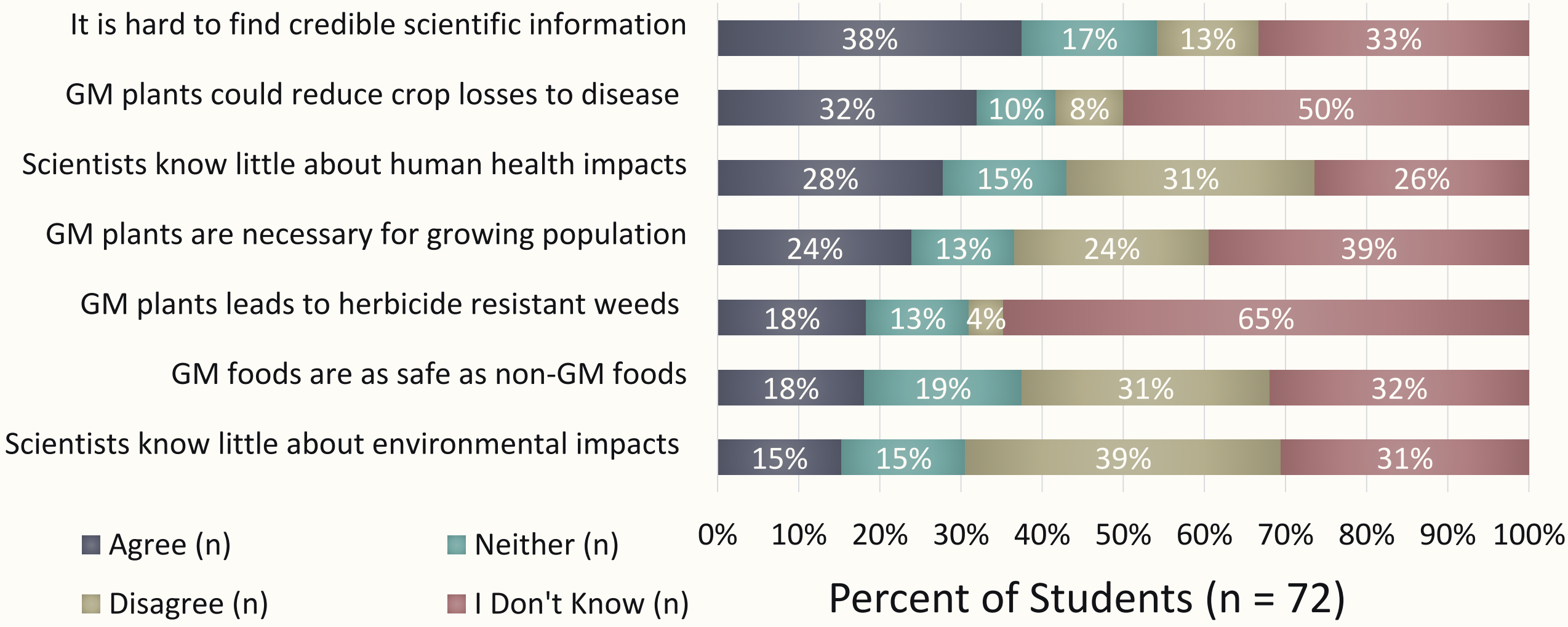
Survey: Knowledge Index

Question	% correct
Some genetically modified plants grow faster than non-genetically modified plants	48
Some plants have been genetically modified to make foods with more minerals and vitamins than traditional crops	44
All food products made from genetically modified plants contain DNA	26
Some plants have been genetically modified to make foods that last longer	16
Traditional crops can become contaminated by pollen from genetically modified plants	16



Low level of knowledge corroborated by data on self-reported knowledge

Survey: Beliefs



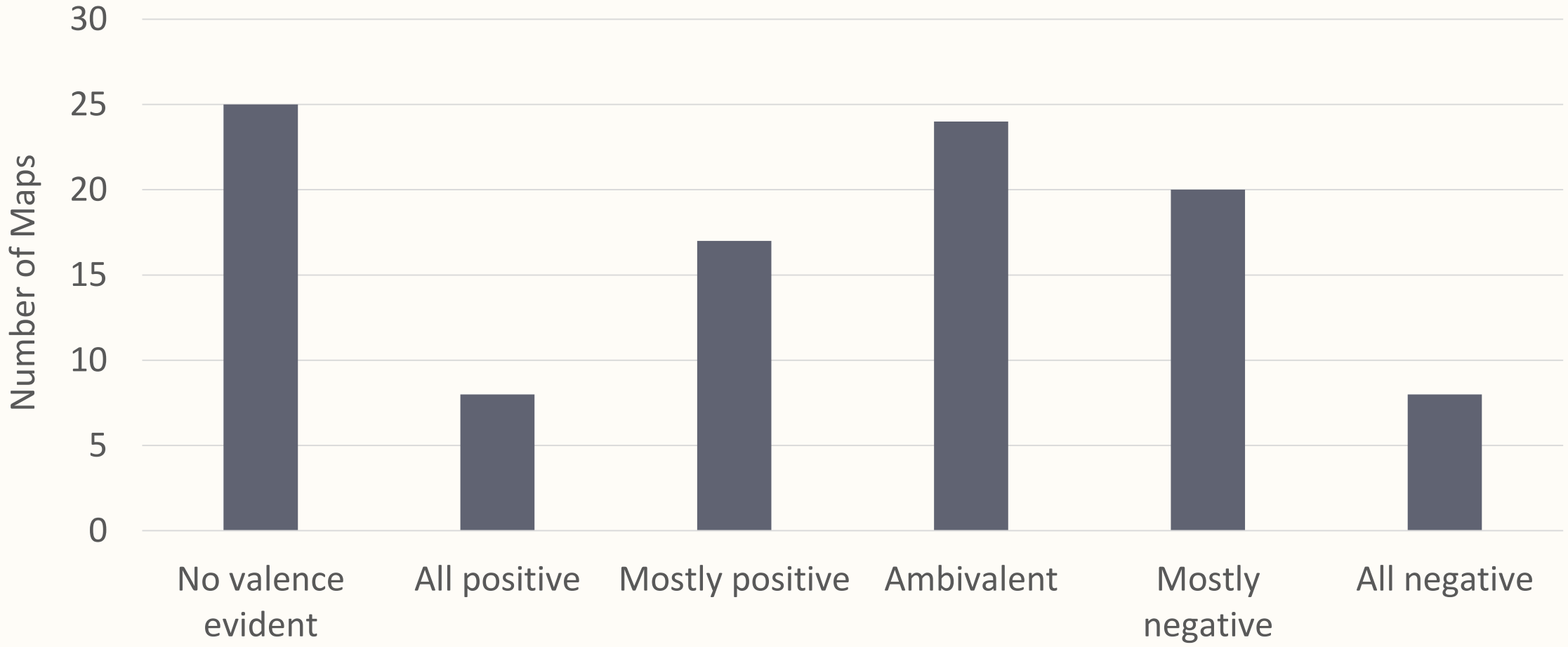
Students do not have strong beliefs about GM

Survey: Attitudes toward GM applications

	Mean*	SD	% support
Genetically modifying plants to produce medicines for humans	2.50	1.09	50
Genetically modifying plants to produce food for humans	2.68	0.95	47
Genetically modifying plants to produce food for farm animals	2.56	1.02	50
Index (Mean)	2.58	0.77	

*1=strongly support; 5 = strongly oppose

Concept maps: Overall attitudes about GM foods



Data did not support expectations of primarily negative attitudes

Overall Conclusions from Initial Assessment:

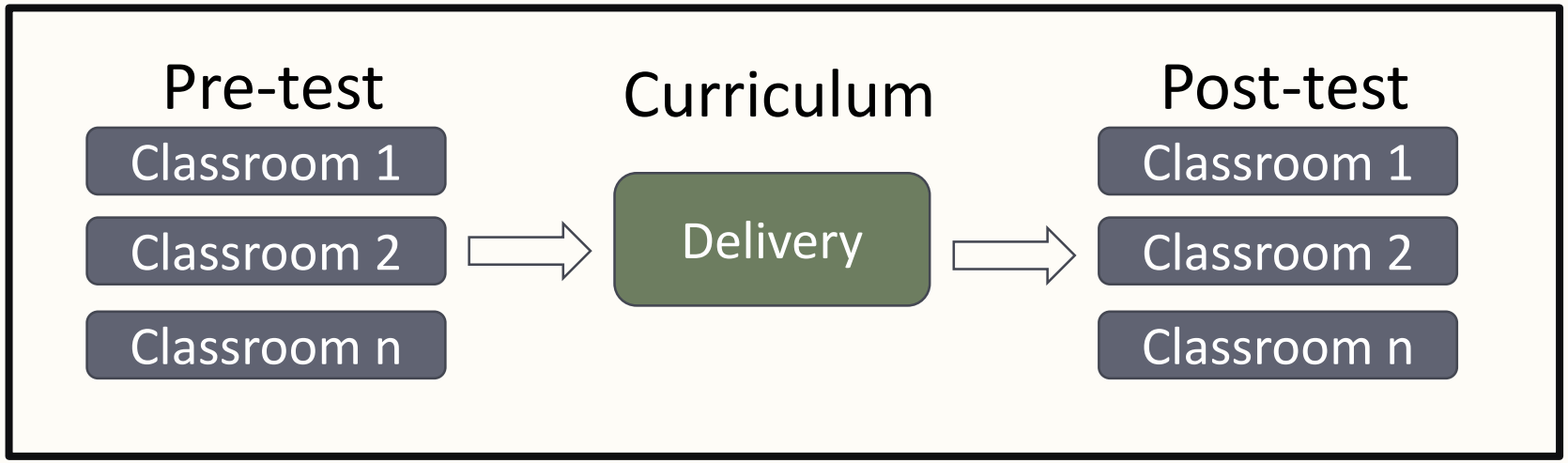
- Students are generally **not very knowledgeable** about GM/GMF
  - Low self-assessed knowledge, low knowledge index scores, and evidence of uncertainty
  - Reliance on seeded words in concept map
  - Other international studies (Taiwan, UK, Netherlands, Australia) also show that students are not very knowledgeable
- Students tend to have **ambivalent attitudes:** they associate both positive and negative outcomes with GM/GMF

Initial Assessment -- Next Steps:

Manuscript for submission to *Journal of Agricultural Education* for review (Fall 2019)

Summative Evaluation (In Progress)

- Use curriculum described previously
- Quasi-experimental design
  - Pre-tests/post-tests (survey & CM) with high school students
  - Interviews with student teachers re: experience using lessons in class



Summative Evaluation (In Progress)

- Data collection: Fall 2019 – Winter 2020
- Advantages to study design
  - Trained instructors – reduces variation in delivery
  - Wide range of classes – large sample
  - Simulates reality of material delivery in classes (time constraints)

Summative Evaluation (In Progress)

- Status
  - Refining questionnaire items to reflect curriculum content
  - Refining concept mapping word bank to address challenges in coding cognitions
  - Determining how we will assess change in concept maps
- Anticipated publication
  - Refereed journal article
  - Conference presentation at International Symposium on Society and Resource Management in 2020 (Cairns, Australia)

Broader Impacts: Where we are

Curriculum

- Curriculum development nearly complete
- Refining curriculum into single block for summative evaluation

Social science

Formative assessment

- Data collection tools refined
- Writing manuscript

Summative assessment

- Submitting IRB application soon
- Recruiting student teacher participants
- Collect data: Fall 2019/Winter 2020
- Analyze data: Spring/Summer 2020
- Develop manuscript/conference presentation for review

Questions?

- Troy Hall, Professor and Department Head, Forest Ecosystems & Society, troy.hall@oregonstate.edu
- Jay Well, Assistant Director, SMILE Program, jay.well@oregonstate.edu
- Betsy Emery, Graduate Research Assistant, Forest Ecosystems & Society, Elizabeth.emery@oregonstate.edu

Appendix:

Detailed Deliverables and Milestones

oal #1: increase HS teachers' content area knowledge, confidence, and materials with mphasis on crop safety and benefits

Proposed Activity	Status (including obstacles & plan revisions)	Output
rmative sessment th teachers	• Conducted focus groups and surveys with SMILE teachers in April 2018 • Used results from those efforts to shape curriculum (found that GMO is not as controversial as expected; teachers have limited confidence to teach this material; students have very brief attention spans and low scientific literacy → significantly down-scaled scope and technical nature of modules). • students' lack of digital literacy and rudimentary knowledge of science led us to develop basic modules on these topics apart from GE lessons.	• Completed ~40 surveys and 4 focus groups with SMILE teachers during Spring 2018 teacher workshop
ILE staff ach material teachers	• Successful teacher workshops in August 2018, January 2019, and August 2019.	• Facilitated 3 separate three-hour workshops with SMILE teachers, totaling 9 hours of professional development for ~40 MS and HS teachers
nual sessment by achers of aterial	• Need to develop plan to obtain their input about teaching the content. • Jay has had informal conversations with teachers that have used the curriculum to gather feedback and efficacy of lessons • Teachers have also provided substantial feedback about lessons during the teacher workshops.	

Goal #2: Increase learners' ability to think critically and introspectively about GT.

Proposed Activity	Status (including obstacles & plan revisions)	Output
Incorporate best practices for curriculum development	• Used teacher input and existing pedagogical techniques for teaching socio-scientific issues to develop 3 separate units, each with 2-3 individual one hour lessons • developed and facilitated an interactive, overnight, educational challenge with 4 separate break out lessons focused on food labeling, how to make a GMO, building a business plan, and merchandising food products for HS students.	• 8 individual one-hour lessons about GE, available at http://people.forestry.oregonstate.edu/steve-strauss/genes-affecting-plant-regeneration-and-transformation-poplar • 4 one-hour interactive breakout sessions that complement each other
Promote open-minded thinking in curriculum	• Incorporated activities into the modules the promote discussion, small group sharing, and team work • Challenge focused on team work and problem solving as a group	
Student assessment	• Had significant delays and challenges with obtaining IRB approval for pilot project in Fall 2018. • Piloted data collection procedure in 9 SMILE clubs (5 MS, 4 HS)	• Collected a total of 63 surveys and 101 concept maps from 125 middle school and high school

Goal #2: Increase learners' ability to think critically and introspectively about GT.

Proposed Activity	Status (including obstacles & plan revisions)	Output
Develop case studies with materials and activities from collaborators. Select cases for curriculum based on audience analysis.	• Reduction in funding led to scaling back to two case studies (originally proposed 5). • Teacher interactions → inability to use herbicide in classrooms, complicating original case study idea to use roundup ready soybeans in activities. • Working with other collaborators to bring in new GE materials into activities (innate potato, arctic apple) • 2019 challenge was focused on students deciding how and why to GE a food crop to populate a new planet. Students developed food packaging and a business plan for that new food product.	• First Unit: Digital Literacy, Scientific Literacy, and GMO Primer (3 individual lessons) • Second Unit: GE Perspectives (1st "Case Study" with 3 separate lessons) • Third Unit: How to Make a GMO (2nd "Case Study" with 2 individual lessons) • Challenge Break Out Sessions (4 separate lessons that are lighter in scope)

Goal #3: increase students' ability to apply science to address complex social problems, esp agriculture

Proposed Activity	Status (including obstacles & plan revisions)	Output
Teachers present curriculum to clubs	• Curriculum used in MS and HS clubs; only have data for first 6 lessons (of 8)	• Each of the lessons were implemented in an average of 10 clubs (out of 30 active MS and HS clubs)
Pre-test/post-test design to assess attitudes and beliefs	• Discovered problems with student attrition and changing attendance at clubs, as well as differences in teachers' implementation of modules. → decided we could not do pre/post design for pilot project. • Still plan to do pre/post study of controlled classroom delivery (direct delivery) in Fall/Winter 2019.	• Collected pre-test surveys and concept maps from 125 MS and HS students (63 surveys; 101 concept maps)
Conduct interviews to assess how cognitive mechanisms of change	• No longer plan to interview students. (Student interest in GE is not very high and the challenges of project team members accessing remote clubs throughout OR are substantial and costly)	

Goal #3: increase students' ability to apply science to address complex social problems, esp agriculture

Proposed Activity	Status (including obstacles & plan revisions)	Output
Club members complete annual post-retrospective evaluation	• No longer plan to do this	• NA
Observations of SMILE meetings and activities	• Jay visited 9 SMILE clubs to facilitate concept mapping exercise and introduce the project. Travel funds were used for this, so no funds for observing SMILE clubs. • Project staff observed high school challenge activities • Project staff will be able to observe student responses during the direct delivery portion of the project.	• NA
Formal summative evaluation with controlled delivery (5 classes)	• Still planned for Fall 2019/Winter 2020 • Partnering with graduate student teachers in Agricultural Education and Math and Science Learning programs at OSU to implement project in their classroom appointments • Potential to implement project in 10 student teachers, each are responsible for 4-6 high school science classes, each with 30 students. Target student sample size is 1,200 students.	• Fall/winter 2019-20

Goal #4: Disseminate case studies

Proposed Activity	Status (including obstacles & plan revisions)	Output
Share curriculum at regional and national science education conferences	• We presented the first lesson from the first module: Fact Checking in an Era of Fake News at the Oregon Science Teachers Association on October 12, 2018 (31 k-12 teachers participated)	• Conference presentation TBD
Disseminate case studies through internet	• All lessons currently developed have been posted on the SMILE website • Will do a bigger push once we have completed direct delivery and finalized all lessons	• See Strauss lab website
Additional broad outputs will include publications in social science journals assessing the success of the curricula based on surveys	• Preparing manuscript using pilot data about students' beliefs and attitudes about GE crops to submit to Journal of Agricultural Education. We are finalizing data analysis for this project. • Plan to prepare a manuscript about efficacy of curriculum using pre-test/post-test data from direct delivery (Fall 2020/Winter 2021)	• Refereed journal article -- TBD

Thank you for listening

287

Posters to date

Oregon State University

Phenomics pipeline for high-throughput image analysis of *in vitro* plant development

Anna Magnuson[1], Chris Dowen[2], Howland Jones[2], Natalie Cochran[2], Dalton Kennedy[2], David Swanson[3], Michael Nagel[1], Cathleen Ma[1], Brett Flenet[1], Rajesh Mewalal[1], Jialin Yuan[3], Nihar A. Dash[3], Fushi Li[3], Yuan Jiang[3] and Steven H. Strauss[1]

[1]Department of Forest Ecosystems & Society, Oregon State University, Corvallis OR; [2]Middleton Spectral Vision, Middleton WI; [3]Department of Computer Science, Oregon State University, Corvallis OR; [4]Statistics Department, Oregon State University, Corvallis OR

Abstract

Collecting quantitative data on in vitro plant material has long been difficult as it relies on scoring systems that are subjective and time consuming. Oregon State University has partnered with Middleton Spectral Vision (MSV, a WI start-up based company) to develop a phenomics pipeline that can efficiently generate high-throughput precision data. This work is part of an NSF Plant Genome Research Program funded study to conduct genome-wide association studies (GWAS) to identify genes that determine the rate of regeneration and transformation in poplar (Populus). The pipeline begins with an RGB and hyperspectral imaging system that is coupled with two modules of data reduction: KernelQuant™, a software developed by MSV that can analyze spectral data, and a machine vision system to identify and quantify tissues from RGB and, ultimately hyperspectral images.

Pipeline overview

Custom imaging system — The MacroPhor™ Array

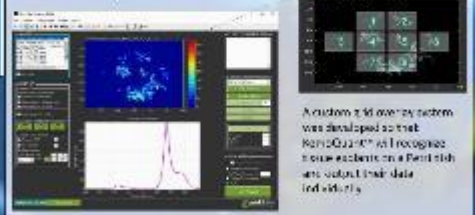
The MacroPhor™ Array is designed to capture RGB and hyperspectral images across a 2x2 Petri dish configuration. Hyperspectral images contain a full 400-800nm emission spectra for each pixel, which allows for rapid and precise detection of fluorescence signals including GFP, and quantitative analysis of spectral components within a sample.

An RGB flat image was developed to allow for RGB images without producing glare on the Petri dish lid. Light was visible at edges of Petri dishes, and ring is visible in overview image to the white debris.

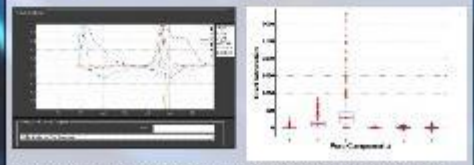
A 488nm wavelength laser scans over the area of a Petri dish to detect GFP, chlorophyll, and other fluorescence signatures in the sample.

Hyperspectral analysis

KernelQuant™ uses multivariate curve resolution to detect pure component fluorescent signals within spectra and automatically exports data on signal intensities and component ratios.

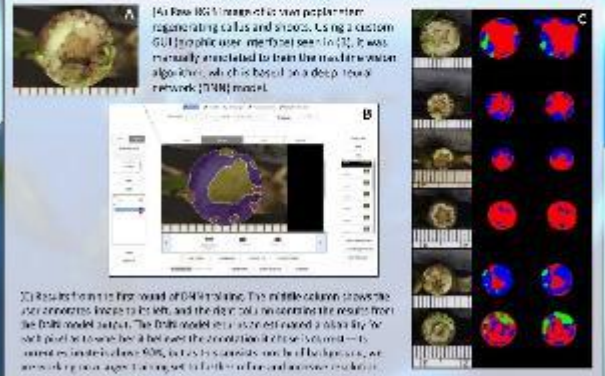


A custom grid overlay system was developed so that KernelQuant™ will recognize tissue explants on a Petri dish and output their data individually.



Pure components for an OSU sample include GFP, chlorophyll a, chlorophyll b, and a unique signature shift for each signal.

Machine vision



A) Raw RGB image of in vitro poplar stem regenerating callus and shoots, using a custom GUI (graphic user interface) seen in (B). It was manually annotated to train the machine vision algorithm, which is based on a deep neural network (DNN) model.

B

C) Results from the first round of DNN training. The middle column shows the user annotations (ground truth) on the left, and the right column shows the results from the DNN model output. The DNN model was trained to classify for each pixel to which part does the annotation belong (callus, shoot, background). The model was trained using a set of 100 images, and a separate image was used to test the performance of the model.

C

Current applications

- The Middleton Spectral Vision system extracts unique spectral signatures and quantifies the amount of fluorescent proteins and chlorophyll signals in OSU tissue samples
- Hyperspectral imaging will aid in a pilot study to determine the effectiveness of the developmental genes LEC1 and ESR1 to increase rates of GWAS for shoot regeneration in poplar
- Hyperspectral imaging will allow us to replace the standard GFP reporter with an auxin-responsive GFP reporter and cytokinin-responsive YFP reporter by early detection of signals
- KernelQuant™ will help screen for the expression of the GFP and YFP reporters, and thus the lab will try to yield biologically informative data about causes of variation in regeneration and transformation
- High-throughput RGB imaging with the MacroPhor™ Array will help increase the accuracy of phenotype-genotype correlation
- Training on the DNN model is ongoing and will ultimately allow us to work in high-throughput mode for an in vitro GWAS study of variation in nutrient by the lab currently in progress

Acknowledgements

We thank the National Science Foundation Plant Genome Research Program for support [IOS #1546900: Analysis of genes affecting plant regeneration and transformation in poplar], and members of the Tree Biosafety and Genomics Research Cooperative at OSU for its long-term support of the Strauss laboratory.

Genome-wide association studies of regeneration in *Populus* with machine vision and hyperspectral phenomics

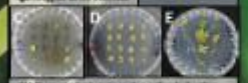
Introduction

- The recalcitrance of many plant species to in vitro regeneration methods prevents researchers from obtaining non-chimaeric transformants following genetic transformation.
- *Populus trichocarpa* is an ideal model organism for studying regeneration due to amenability to in vitro and in vivo regeneration protocols.
- Genome-wide association studies (GWAS) can benefit from advances in phenotyping, particularly high-throughput phenotyping, hyperspectral imaging and machine vision, to capture a breadth and depth of data that would be practically unobtainable by humans unassisted by machines.
- To obtain both RGB and hyperspectral data on a high-throughput scale for in vitro plant tissue cultures, we are using the macroPlor™ Array, a custom instrument from Middleton Spectral Vision, and the accompanying software suite KernoQuest™ to deconvolute spectra into individual components, including chlorophyll, fluorescent reporters and spectral shifts.
- Deep learning models using convolutional neural networks are trained to segment images by tissue type (i.e. callus, shoot) after learning from images annotated by a graphical user interface, thus enabling high-throughput and precise analysis of images.

Regeneration

In vitro regeneration

Optimized callus and shoot induction hormone treatments for in vitro regeneration from stem and petiole were selected following herbicide testing to determine which cause the widest range in response across genotypes (data not shown). Tissue culturing and plant phenotyping has begun and will run through 2019, with 1200 genotypes to be studied.



(A) Stem explant undergoing callus induction and (B) shoot regeneration; petri dishes of stem explants undergoing (C) necrosis, (D) various stages of callus induction, and (E) later stages of callus growth and shoot regeneration

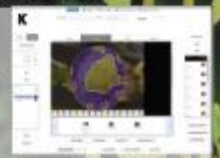
Stem regeneration

To enhance the natural wounding and regeneration response, 0.5mg/mL thidiazuron (a synthetic cytokinin) was applied to the tips of cut stems. Phenotyping has been completed for 593 of 902 genotypes to be included in the stem regeneration study, and this partial dataset is being used to test and refine GWAS methods.

(F-G) Side view of shoots regenerating from tips of cut stems placed in water; View of stem tip showing several phenotypes recognized by machine vision: (H) green callus; (I) green callus and shoot; (J) red callus and shoot

Machine vision

(K) Through a graphical user interface, specific tissues of interest are annotated by the user with assistance by an edge detection algorithm. These annotated images are then used to train convolutional neural networks for image segmentation. Project aims includes deployment of this interface on a web server to assist researchers using machine vision.



After learning from user annotations, deep learning models classify sections of images by the type of tissue recognized (L). The accuracy is measurable by comparison to user-annotated images not included in the training set and depends on the type of deep learning architectures used (VGG-16 and PSPNet). Images below are segmented into callus (blue), shoot (green) stem (blue) and background).

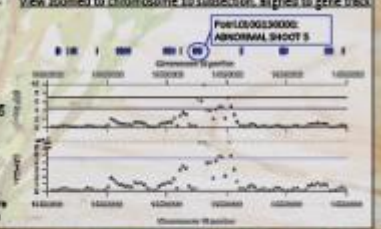
Association testing

The GWAS method Sequence Kernel Association Test (SKAT) was used to collapse SNPs into 5kb windows based on physical location and test for associations with traits of interest. To control for non-normality of data, resampling was applied. This test reveals a possible association of shoot area with a homolog of a known shoot regulator in Arabidopsis.

Genetic markers associated with shoot area:

Genome-wide view

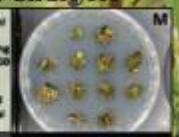
View zoomed to chromosome 10 subsection, aligned to gene track

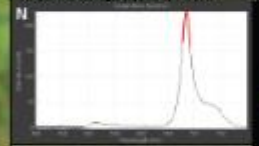


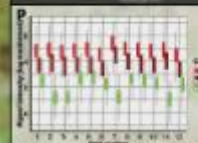
*The lowest p-values (green) will require more resampling to validate

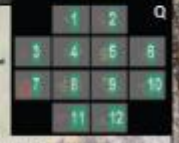
Hyperspectral analysis

The macroPlor™ Array (middletonspectralvision.com) is used to capture RGB (M) and hyperspectral images (N-Q) for trays of petri dishes. Deep learning models, while depending on user annotation of RGB images, will also incorporate channels for spectral components provided by hyperspectral analysis. Fake color applied to GFP and chlorophyll spectral peaks (N) enables qualitative inspection of spectral components of images (O) prior to analysis.









With the software KernoQuest™, images are segmented by explant (Q) and spectral components are identified by multivariate curve resolution and quantified by least squares linear regression (P). For each component, signal intensities are computed per pixel (providing additional image channels to be used for machine vision training and prediction of tissue type) and per explant (to be used directly in GWAS).

Outlook

- Convolutional neural networks and hyperspectral imaging provide new opportunities for genetic discovery by enabling precise, high-throughput phenotyping of complex traits.
- While currently, most genes in forest species and other non-model plants remain uncharacterized, the advent of high-throughput hyperspectral phenotyping with machine vision may enable rapid elucidation of these genes via improved power and accuracy for experiments to study gene-function relationships.
- Characterization of the genetic basis of regeneration offers opportunities for converting poor responders to regeneration into efficient responders via overexpression or knockdown/out of developmental regulators. This may enable robust transformation of genotypes and species which genetic engineering methods cannot be efficiently applied to yet.

Acknowledgements

We thank the National Science Foundation Plant Genome Research Program for support (IOS # 1546900), Analysis of genes affecting plant regeneration and transformation in poplar), and members of the Tree Biosafety and Genomics Research Cooperative at OSU for its long-term support of the Strauss laboratory.

Analysis of Genes Affecting Plant Regeneration and Transformation

Amanda Goddard[1], Elizabeth Emery[1], Michael Nagle[1], Damanpreet Kaur[2], Jialin Yuan[2], Ekaterina Peremyslova[1], Cathleen Ma[1], Troy Hall[1], Yuan Jiang[3], Fuxin Li[2], Jay Wells[4], and Steven H. Strauss[1]

[1]Department of Forest Ecosystems & Society, Oregon State University, Corvallis, OR; [2]Department of Computer Science, Oregon State University, Corvallis, OR; [3]Statistics Department, Oregon State University, Corvallis, OR; [4]Science and Math Investigative Learning Experiences (SMILE), Office of Precollege Programs, Oregon State University, Corvallis, OR, 97331

Summary

- The goal of this project is to develop high throughput phenomics assays for studying genetic control of regeneration and transformation in poplar using GWAS.
- A secondary goal is to conduct social science research, and create new education materials informed by this research, relevant to understanding GMOs, agriculture, and scientific literacy in the online/social media world. Key developments to date include: 1) A customized imaging system for high throughput visible and hyperspectral images of in vitro plant materials (prototype is M dilation Spectral Vision). 2) In depth analysis of in vitro regeneration conditions suitable for producing highly heritable regeneration and transformation responses in nearly co-isogenic. 3) A machine vision/phenomics system for rapid phenotyping based on RGB images is nearly complete. 4) Preparation of in vivo and in vitro regeneration responses is nearly complete for three GWAS experiments (~1,000 genotypes). 5) GWAS of preliminary data on in vivo shoot regeneration (from machine vision-predictions) suggests regeneration-control-genes can be identified.

Extensive natural variation in regeneration responses in wild cottonwoods

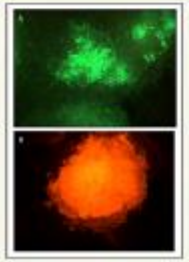
Our GWAS study relies on naturally high levels of SNP and trait polymorphism accompanied by low linkage disequilibrium in wild outcrossed trees. Shown to the left are six wild genotypes (columns) as they regenerate shoots in vivo from cut, dormant stem surfaces over a 5 week experiment (rows).

Shown below are four wild genotypes growing in vitro as they regenerate shoots after callus induction. Note the variation in callus size, color, and subsequent shoot size and number.

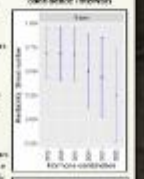
In Vitro Treatment and Heritability Optimization

To identify conditions that will maximize regeneration, and increase our power to detect causative genes via GWAS, we have explored 1,757 unique combinations of in vitro medium-hormone-explant treatments x genotypes over 12 distinct experiments. The figure above summarizes the treatments and explant sources studied or under study.

Variation in Phenotypic Responses



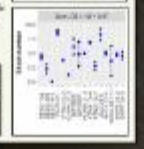
In vitro treatment optimization (above box) has produced a wide range of regeneration phenotypes useful for GWAS.



Heritability of shoot regeneration and 95% confidence intervals

To the left are fluorescent reporters are being used to monitor transformation and regeneration responses, including tissue and embryogenesis [3]. The reporters can be readily quantified by hyperspectral image analysis, using the macroPhor Array (see next box).

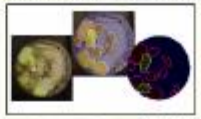
The use of two light genotypes allows us to calculate heritability and identify the best treatment for maximizing genetic expression for GWAS. GWAS combines high HF and marker genotype variation.



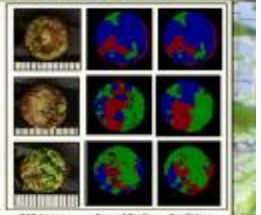
GFP (A) and Ds Red (B) fluorescent markers indicate the presence of a transgenic shoot meristems or somatic embryos in response to hormone or GENE gene treatments such as ARUSONEL.

Treatment	CK1	CK2	BA4	BA5	BA6	CK6
ZmD (mg/L)	0	0.05	0.1	0.25	0.05	0.1
BAP (mg/L)	0.5	0.5	0.5	1	1	1

Machine Vision, Genotype Prediction & Hyperspectral Phenotyping



macroPhor Array™ Custom instrument for high-throughput hyperspectral & RGB imaging



RGB Image | Ground Truth | Prediction

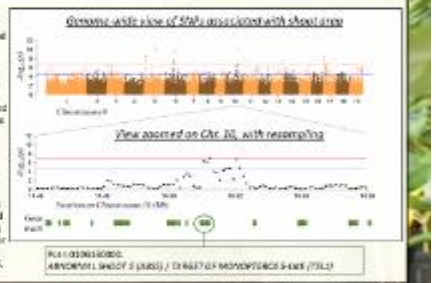
	Background	Stem	Callus	Shoot	Mean IoU
Training	99.71%	91.12%	89.71%	94.48%	93.95%
Validation	99.65%	79.03%	72.68%	77.27%	81.38%

Table 1. Results of the DeepLab 3+ model after it has been trained via transfer learning on our hand-annotated ground truth dataset. The validation dataset percentages represent the accuracy of the current best model in annotating the differences between background, stem, callus, and shoot.

"Ground Truth" images above are the result of the hand annotation of 117 RGB images of explants denoting stem, callus, and shoot growth on an in vivo explant, as well as the surrounding background. These images were used to train the model. The images labeled 'Prediction' are a result our model utilizing semantic segmentation in DeepLab 3+. Prediction accuracy (mean IoU) is currently 81.4%.

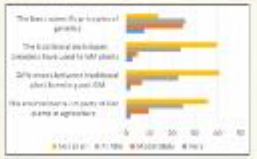
Preliminary GWAS mapping of in vivo shoot regeneration

SNP-set Sequence Kernel Association Test (SKAT) was used to identify 3kb windows of SNPs (indicated by dots) associated with the proportion of each plant's area classified by machine vision as shoot. Accounting (up to 10PC corrected and) corrected for non-normality of the trait. A top association appears in the promoter and 5'UTR of a homolog of a known regulator of shoot development, ARR5.



Rco1 (ORR5/ARR9): ARABIDOPSIS RESPONSE REGULATOR 5 (ARR5) / TOMATO RR MONOPHYLOGENY 5 (RR5)

Broader Impacts through SMILE



To understand how students think about genetic modification and GM foods, 183 students were asked to create their own concept map to display their knowledge about these topics. One of the most complex maps is shown above.

We are working with the SMILE's state wide network of middle and high school programs to study and educate underrepresented students and teachers around the topic of agricultural genetic modification (GM). Survey results (left) of 73 students show that they rated their knowledge about topics related to agricultural GM as mostly low. Other than understanding the basic principles of genetics, they expressed little confidence in their literacy about GM.

As part of this project, we have developed 6 lessons about the science behind genetic modification and the various perspectives on the technology. These have been used in classrooms across Oregon and can be accessed at smile.oregonstate.edu/lesson-resources.

Acknowledgements

We thank the National Science Foundation Plant Genome Research Program (IOS # 1546900) for support of this project. We also thank the members of the Tree Biosafety and Genomics Research Cooperative at OSU for its long term investment in our transformation and regeneration studies. We thank Middleton Spectral Vision for high quality imaging system support.

Web- Based Deep Segmentation Tools for Phenotyping

Jialin Yuan[1], Zheng Zhou[2], Michael Nagle[1], Ekaterina Peremyslova[1], Ali Behnoudfar[1], Nihar A. Doshi[3], Ritesh Mewalal[1], Cathleen Ma[1], Anna Carlina Magnuson[1], Yuan Jiang[1], Steven H. Strauss[1] and Fuxin Li[1]

1. Oregon State University, Corvallis, OR, 2. Tencent Inc., ShenZhen, China, 3. University of Southern California, Los Angeles, CA

Motivation

- Quantifying phenotypes of complex biological tissues, such as during in vitro regeneration, is slow and imprecise. Machine vision methods can give a major improvement, but need a user friendly interface (GUI) for annotation of tissues of interest
- Non-web-based IA (image annotation) tools rely on installation and configuration, often difficult to use
- Existing web-based IA tools, i.e. LabelMe[1], Labelbox[2], are expensive and time-consuming to annotate objects at pixel-level accuracy
- Deep learning method is the state-of-art method for segmentation, it is expected to be more efficient and robust to obtain pixel-level annotations

Image Annotator

class/object configuration

- **Class Panel:** User can specify the names of the classes
- **Hierarchy Panel:** User can specify objects and the classes (parts) they belong to
  - Add multiple classes into an object.
  - Use "add to" to directly add classes.

Toolkit

- Two pens for drawing: *posPen* and *negPen*
- Two modes to annotate target objects:
  - **DL-ObjectSelect mode**[3], draw positive strokes inside one object and negative strokes outside, the UI then automatically generate precise edges along labeling result.
  - **Manual mode**, mark objects by directly painting on them.

Image Annotation GUI

IO system

- Load multiple images and view them in the file gallery
- Import and export the configuration of classes and objects
- Save the generated annotation result

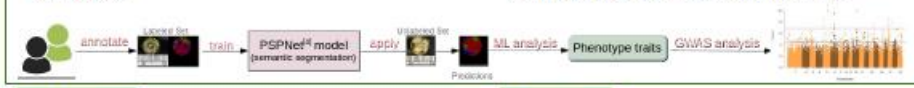
History management

- *Undo* and *redo* previous drawings to retract wrong operations
- Clear all unwanted *posPen* or *negPen* drawing in one click

Example of Annotating one Image

A GWAS Application

1. 103 annotated Images from the Image Annotator
2. Trained the PSPNet[4] for semantic segmentation of background, callus, shoot and stem
3. Compute phenotype traits: portion of callus, shoots over stem
4. GWAS analysis: normality and association testing (see poster by Michael Nagle for detailed results on GWAS associations)



Conclusion

- We developed a web-based interactive image annotation tool
  - There is no need to install any software or configure anything.
  - The user can work from anywhere, anytime with any device or platform.
- Our IA tool is integrated with the DL-ObjectSelect[3] algorithm, which is easy to use for object segmentation
- We adopted the IA tool in a GWAS study, which proves to be effective in practice, though improvements in annotation databases and prediction efficiency are still under study

References

[1]. "LabelMe". Available: http://labelme.csail.mit.edu/Release3.0/
[2]. "LabelBox". Available: https://www.labelbox.com
[3]. Xu, Ning, et al. "Deep interactive object selection."CVPR. 2016.
[4]. Zhao, Hengshuang, et al. "Pyramid scene parsing network." CVPR. 2017.

Acknowlegement

We thank the National Science Foundation Plant Genome Research Program for fundin, "Analysis of Genes Affecting Plant Regeneration and Transformation in Poplar." IOS # 1546900.

GWAS Identification of Loci Associated with Rooting in *Populus*

Anna C. Mamidipudi, Sabrina Mai, Christine Jacobson, Alexander Myburg, Luan Kuo, Yuan Liang, Kelly Li, Heather S. Dunning, Stephen P. DiFazio, and Steven H. Strauss

Abstract

Two independent experiments were performed at Oregon State University and Washington State University to identify genetic loci associated with root production and architecture. Cuttings were taken from a field site in Corvallis, Oregon and a 335 replicate clonal common garden in Clatskanie, Oregon. The experiments followed similar methodologies. Cuttings were rooted in soilless media under controlled environmental conditions. Root phenotypes were collected using destructive harvesting, machine vision, and manual measurements. Heritability estimates were calculated for each trait. Genome wide association studies (GWAS) were performed using a mixed linear model approach. GWAS analyses of the Clatskanie clonal common garden identified a number of SNPs associated with rooting traits at the 0.05 FDR threshold. Candidate genes were identified within 20 kb of significant SNPs. Several of these genes have been previously implicated in root development, while others represent novel associations. Genomic prediction models were also developed using genome wide SNP data. Model accuracy was assessed using cross validation. These results provide insight into the genetic architecture of root traits in *Populus* and may facilitate marker assisted selection for improved rooting ability in breeding populations.

Project Overview

Oregon State University:

- Two-node cuttings were collected from 335 unique genotypes in a clonal common garden in Clatskanie, OR
- Cuttings were potted in soilless media in individual pots, and placed in a greenhouse with supplemental lighting
- Cuttings were rooted in a growth chamber (22/15 °C, 16 h photoperiod, 250 µmol m⁻² s⁻¹ PAR) for 3 weeks
- Data collected on survival, foliar health, dormancy, and rooting quality score
- Before rooting, cuttings were gravity-watered for 1 min, given 1 of 3 soil moisture treatments (Control, Drought, Flood), and returned to the growth chamber
- Destructive harvest after 3 weeks
- Roots washed over a 2 mm mesh sieve, scanned for root architecture and average stem diameter using WinRHIZO Pro 2019a
- Data collected for principal component analysis to generate new variables for GWAS

Washington State University:

- 400 rooted cuttings from seven genotypes were collected
- Cuttings were rooted in soilless media in the Fall (3 °C) for 5 weeks (2400 µmol m⁻² s⁻¹ PAR, 16 h photoperiod, supplemental lighting)
- Cuttings were rooted in a growth chamber (22/15 °C, 16 h photoperiod, 250 µmol m⁻² s⁻¹ PAR) for 3 weeks
- Once roots were 2.5 cm, they were photographed and images were analyzed using ImageJ
- Data collected manually included:
  - Dry Root Biomass (DRB)
  - Longest Root Length (LRL)
  - Root Density Type 1 (RD1)
  - Root Density Type 2 (RD2)
  - Root Branch Count (Root-14) (RBC)
- GWAS analysis is in progress

Phenotyping

Figure 1. Experimental setup at OSU. A) Water treatments cutting, after one of drought, or flood treatments cutting. C) Rooting after three weeks.

Figure 2. Root sequence at Washington State University. A) Cuttings in soilless media after three weeks of growth. Treatments of Root Imaging. Rooting took 1-2 weeks and cuttings were imaged when roots measured 2.5 cm. C) Image analysis.

Machine Vision Analysis

Color-based clustering was used on each image to segment out the background and leaf tissue. Then the color, shape, and root topology, as well as combined traits from multiple images, were used. Root area (mm²) and average stem diameter measurements were compared for 423 genotypes to assess the overall variation in rooting ability.

Figure 3. *Left image*: Original image of harvested root cuttings. *Right image*: Segment mode produced by color based morphology segmentation. Root area and average stem diameter were calculated from the segmented image.

Heritability

Correlations between trait and the heritability and measures of central tendency between OSU and WSU across different rooting platforms. A high genetic correlation explains about 10-25 % of the variation.

Variable	Y Root Diameter	Y Root Density
Combined	0.16	0.13
Water treatment	0.04	1.00
Soil		
Treatment	0.00	0.15

Variable	
Days to Root Initiation (DRI)	0.35
Longest Root Length (LRL)	0.27
Total Root Length (TRL)	0.16
Density Parameter (RD1/RD2)	0.13
Root Count Number (RBC)	0.12

Table 1. Generalized linear mixed model results for each trait. Random effects were included for the OSU block, WSU tray, and WSU position. Fixed effects included the OSU treatment and WSU root count. Conditional and marginal R² values are reported.

Table 2. Heritability estimates for five traits collected in the WSU experiment. Values were calculated using a mixed linear model with 399 genotypes. Root diameter and root density were measured at two different times (early and late).

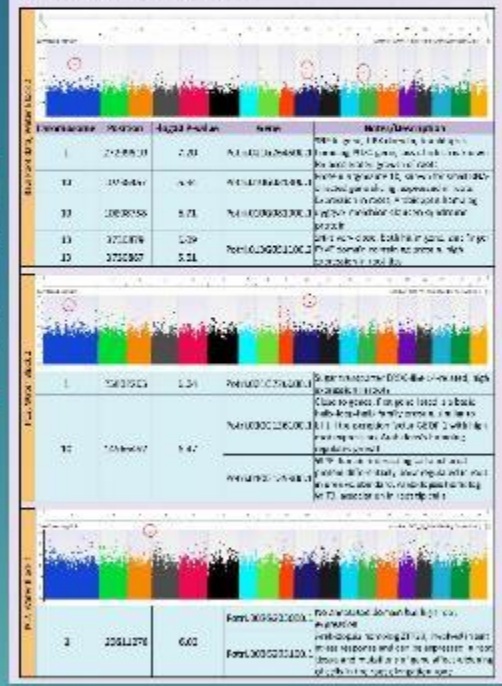
Principal Components Analysis

	PC1	PC2	PC3	PC4
Height	0.44	0.43	0.77	0.15
Diameter	0.55	0.3	-0.75	-0.15
Ave. Stem Diameter	0.6	0.25	0.56	0.66
Root Area	0.33	0.83	0.33	0.08

Genome Wide Association Study

GWAS (genome-wide association study) was performed using the mixed linear model approach implemented in the GEMMA software package. A total of 1,248,000 SNPs were used in the analysis after filtering for minor allele frequency (MAF > 0.05) and missing data. Population structure was controlled using a centered relatedness matrix. Significance thresholds were determined using a false discovery rate (FDR) of 0.05. Manhattan plots were generated for each trait, and candidate genes were identified within 20 kb of significant SNPs using the *Populus trichocarpa* v4.1 genome annotation.

Below, some of the SNPs of interest are shown with their chromosome, position, significance values, and gene annotations, for the clonal garden.



Chromosome	Position	-log10 P-value	Gene	Nearby Gene Description
1	21,458,210	7.83	Potri.001G264800.1	DNA-binding bromodomain-containing protein; may regulate chromatin remodeling and transcription
10	17,904,567	6.91	Potri.010G178700.1	Auxin-responsive protein; involved in lateral root development and auxin signaling pathway
10	18,032,455	6.71	Potri.010G183300.1	Cytochrome P450 family protein; may participate in hormone biosynthesis
13	3,711,876	5.89	Potri.013G031200.2	Ethylene-responsive transcription factor; may regulate root development
13	3,722,907	5.81		
1	7,581,953	5.54	Potri.001G078100.1	Serine/threonine-protein kinase; may regulate root growth
10	7,654,567	5.47	Potri.010G126000.1	Transcription factor; may regulate root architecture
			Potri.010G126100.1	
			Potri.010G126200.1	
3	20,811,278	6.02	Potri.003G203300.1	Pentatricopeptide repeat-containing protein
			Potri.003G203400.1	Protein kinase superfamily protein

Summary

- Root architecture is a highly polygenic trait that can be measured using multiple methods
- Heritability estimates were moderate to high for most traits, indicating that genetic factors play a significant role in root development
- Both the additive GWAS results and the principal component data yielded many statistically significant SNPs. Upon investigation, some SNPs were found to be associated genes with functions in auxin signaling, transcription regulation, and root development, which may have implications for improving root traits in breeding programs
- Genomic prediction models showed moderate accuracy, with the highest accuracy observed for traits related to root architecture and lowest accuracy for traits related to root biomass

Acknowledgements

We thank the Oregon State University Forest Biotechnology Group for their support. This work was supported by the National Science Foundation Plant Genome Research Program (Award #IOS-1238202). Analysis of genotyping and genomic data was performed using resources provided through the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science. We also thank the members of the Tree Biotechnology and Genomics Cooperative at OSU for their support and feedback.

Development of an imaging-based phenomics system for *in vitro* GWAS studies of plant regeneration and transformation

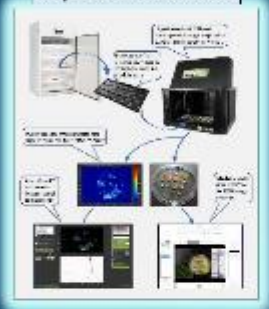
Anna Magnuson[1], Chris Graves[2], Haviland Jones[2], Natalie Crothers[2], Gabor Kemeny[2], David Braumann[2], Michael Nagle[1], Cathleen Ma[1], Brett Pierce[1], Ritesh Mewalal[3], Jiahn Huan[2], Nihar A. Deshi[2], Fuxin Li[4], Yuan Jiang[4] and Steven H. Strauss[1]

[1]Department of Forest Ecosystems & Society, Oregon State University, Corvallis OR; [2]Middleton Spectral Vision, Middleton WI; [3]Department of Computer Science, Oregon State University, Corvallis OR; [4]Statistics Department, Oregon State University, Corvallis OR

Abstract

Collecting quantitative data on *in vitro* plant material has long been difficult as it relies on scoring systems that are subjective and time consuming. Oregon State University has partnered with Middleton Spectral Vision (MSV), a Wisconsin-based company, to develop a phenomics pipeline that can efficiently generate high-throughput, precision data. This work is part of an NSF Plant Genome Research Program funded study to conduct genome-wide association studies (GWAS) to identify genes that determine the rate of regeneration and transformation in poplar (*Populus*). The pipeline begins with an RGB and hyperspectral imaging system that is coupled with two routes of data extraction: KernoQuant™, a software developed by MSV that can analyze spectral data, and a machine vision system to identify and classify tissues from RGB images. Ultimately hyperspectral images

Pipeline overview



Custom imaging system — The MacroPhor™ Array

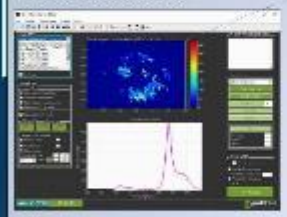
The MacroPhor™ Array is designed to capture RGB and hyperspectral images across a 2x3 Petri dish configuration. Hyperspectral images contain a full 400-800nm emission spectra for each pixel, which allows for rapid and precise detection of fluorescence signals including GFP, and quantitative analysis of spectral components within a sample.

An LED lighting system developed to allow for RGB images without producing glare on the Petri dish lid. Light is visible on the edges of Petri dishes, and ring is visible in chamber image to the right of dishes.

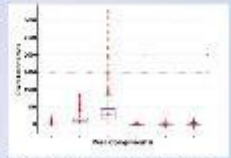
A 488nm wavelength laser scans over the area of a Petri dish to detect GFP, chlorophyll, and other fluorescence signatures in the sample.

Hyperspectral analysis

KernoQuant™ uses multivariate curve resolution to detect pure component fluorescent signals within spectra and automatically exports data on signal intensities and composition ratios.



A custom grid overlay system was developed so that KernoQuant™ will recognize tissue explants on a Petri dish and output tissue data individually.



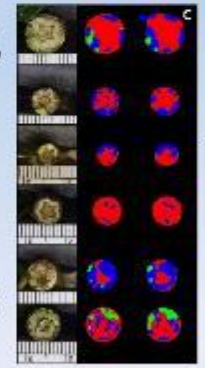
Pure component factors of GUS sample include GFP, chlorophyll a, chlorophyll b, and a unique signature shift for each signal.

Machine vision

A

[A] Raw RGB image of a cross poplar stem regenerating callus and shoots. Using a custom built graphic user interface seen in [B], it was manually annotated to train the machine vision algorithm, which is based on a deep neural network (DNN) model.

B



C

[C] Results from the first round of DNN training. The middle column shows the user annotated image to the left, and the right column contains the results from the DNN model output. The DNN model returns an estimated probability for each pixel as to whether it belongs to the annotation it should resemble—in current version it considers above 90%, but as it is dominated by all background, we are adding more augmented training sets to further refine and increase resolution.

Summary

- The Middleton Spectral Vision system enables consistent, high throughput, and high resolution RGB and hyperspectral imaging of tissues in Petri dishes.
- Its hyperspectral image files are software enabled unique spectral signatures, including of GFP and chlorophyll, to be developed, extracted, and quantified for user-fit explants.
- Hyperspectral imaging of the senescence process to quantify transformation rate and genetic variation in response reactivity using with auxin- and cytokinin-responsive GFP/YFP.
- A gradient boosting machine learning model on machine vision has been developed that allows us to specify target classes of tissue types within cross-sections (e.g. callus, leaves, shoots, roots).
- Training on the deep neural network model is ongoing and will ultimately allow us to obtain high-throughput data for our Jr. Level GWAS study, which will be in vitro regeneration and mortality that is currently in progress.
- Phenotyping of in vitro regeneration response for GWAS is expected to begin this fall through winter.

Acknowledgements

We thank the National Science Foundation Plant Genome Research Program for support (IOS # 1546900). And lots of gratitude to all of the Tree Biosafety and Genomics Research Cooperative at OSU for its long-term support of the Strauss lab (2000's).

Analysis of Genes Affecting Plant Regeneration and Transformation

Brett B. Pierce[1], Cathleen Ma[1], Anna Magnuson[1], Amy Klocko[1], Michael Nagle[1], Wellington Muchero[2], Jialin Yuan[3], Yuan Jiang[3], Fuxin Li[3], and Steven H. Strauss[1]

[1]Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR; [2]Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN; [3]School of Electrical Engineering and Computer Science, Oregon State University; and [4]Department of Statistics, Oregon State University

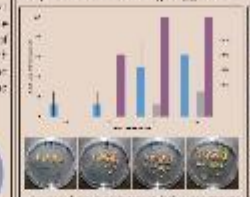
Project Overview

Regeneration of differentiated organisms from single cells is a critical need for functional genomics and for the construction of genetically engineered organisms. The ability to manipulate genes responsible for regeneration of tree species will be a key to understanding the genetic control of [illegible] and transformation in [illegible] Populus, which is one of the few [illegible] studied species with the potential to [illegible] these traits. The project will identify genetic elements that control [illegible] develop novel phenomic methods based on image analysis, and develop new tools, science and education methods for teaching about genetic engineering, to diverse high school students and teachers.

Our specific objectives are to: (1) evaluate a variety of [illegible] methods to assess regeneration in tree seedlings, (2) develop new phenomic tools, including an image capture and automated data management system, to massively enhance the ability to phenotype cloned GWAS mapping populations with regard to [illegible] shoots and new root tips, develop new shoots and new root tip materials, deliver them to molecular biology communities for [illegible] and [illegible] through publications, social media, and conferences, share the projects, insights and learning models internationally.

Genetic Dissection of Regeneration

The goal of the phase of work is to identify a series of treatments that maximize genetic variance and minimize environmental variance in the GWAS population. [illegible] study of treatments that give the most heritability and thus will allow GWAS efficiency. In a factorial experiment, we tested 23 basal media types and a number of other [illegible] factors on a small number of genotypes. Here is one [illegible] treatments of one to 16 genotypes. Factors tested included nitrogen and carbohydrate, auxins [illegible] 2,4-D, IAA, IBA, light, genotype, and some types of combinations. Methods included two examples below. High [illegible] concentrations of CaCl promote shoot organogenesis on basal [illegible] callus, and [illegible] are [illegible] light also [illegible] shoot organogenesis.



In Vitro Culture and Shoot Induction

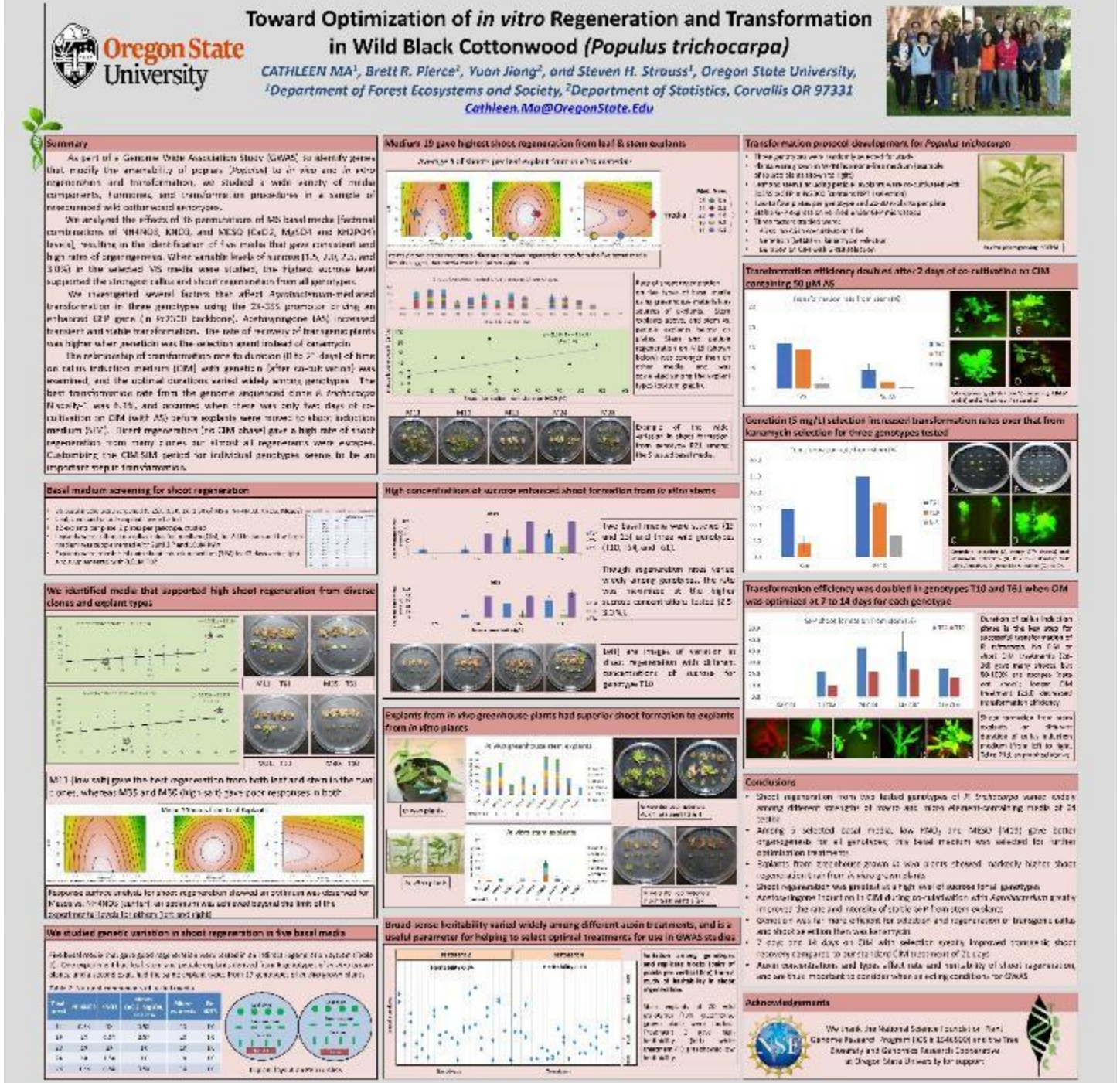
Morphological and Biochemical Identification

The goal is to accurately recognize and quantify individual tissue types and sort them into predefined categories (e.g., roots, shoots, and calli), as well as to recognize more complex and unique or physiological states (e.g., nutrient stressed, pigmented, necrotic, and modified organs). The [illegible] is to [illegible] using the Gabor algorithm to [illegible] segmentation which [illegible] can [illegible] but [illegible] specific [illegible] object of the analyzed based on color distribution of the target object. A deep learning algorithm [illegible] automatic [illegible] be assigned to further the [illegible] capabilities. The integration of the deep learning algorithm and the [illegible] will also increase production speed as well as drastically reduce processing time, therefore enhancing high-throughput analysis.

Image Acquisition System

The call to obtain high-resolution quantitative, precision data that reflect the complex biochemical biology of regenerating tissues is challenging, but critical for GWAS of regeneration traits. We therefore set out to design a new [illegible] system that would [illegible] and physiologically [illegible] measurements.

We contracted with Vermillion Specialty Products (Vermillion, WI) to design and building a [illegible] capable of imaging [illegible] of this project. The [illegible] will be equipped with one high-resolution [illegible] camera [illegible] and [illegible] [illegible] [illegible]


Toward Optimization of in vitro Regeneration and Transformation in Wild Black Cottonwood (Populus trichocarpa)
Oregon State University
CATHLEEN MA[1], Brett R. Pierce[2], Yuan Jiang[2], and Steven H. Strauss[1], Oregon State University,
[1]Department of Forest Ecosystems and Society, [2]Department of Statistics, Corvallis OR 97331
Cathleen.Ma@OregonState.Edu
Summary
Medium 19 gave highest shoot regeneration from leaf & stem explants
Transformation protocol development for Populus trichocarpa
Transformation efficiency doubled after 2 days of co-cultivation on CIM containing 50 µM AS
Genetic (5 mg/L) selection increased transformation rates over that from kanamycin selection for three genotypes tested
Basal medium screening for shoot regeneration
High concentrations of sucrose enhanced shoot formation from in vitro stems
We identified media that supported high shoot regeneration from diverse clones and explant types
Transformation efficiency was doubled in genotypes T10 and T61 when CIM was optimized at 7 to 14 days for each genotype
Explants from in vivo greenhouse plants had superior shoot formation to explants from in vitro plants
Conclusions
We studied genetic variation in shoot regeneration in five basal media
Broad-sense heritability varied widely among different auxin treatments, and is a useful parameter for helping to select optimal treatments for use in GWAS studies
Acknowledgements

Advanced phenotypic analysis of *in vitro* development and transformation for GWAS in *Populus*: *Machine vision analysis of RGB and hyperspectral images*

Michael Hoge[1], Cathleen Ma[2], Anna Magnuson[2], Julie Haun[2], Nitin A. Doshi[2], Fadhli Ali[2], Yuan Jiang[2], Nora Ireland[2], Daniel Tran[2], Steven H. Strauss[2]
[1]Department of Computer Science, Oregon State University Corvallis OR, [2]Forestry, Department, Oregon State University Corvallis OR

Overview

- Recalcitrance to *in vitro* regeneration: an obstacle to transformation of many species, particularly woody plants
- *Populus trichocarpa* is an ideal model organism for GWAS of regeneration due to variation in transformability, regenerative capacity across genotypes, and diversity across wild populations
- Power and accuracy of GWAS can benefit from new phenotyping methods (high-throughput, hyperspectral, machine vision)
- Through GWAS, we propose that polymorphism in regulation of a known shoot developmental regulator affects shoot regeneration

Plant treatments

In vivo

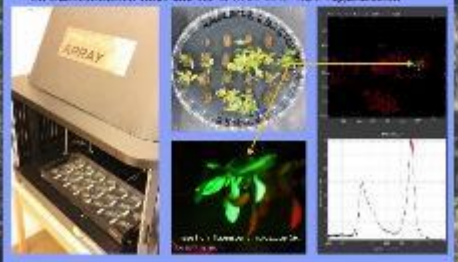
- Cut stem tips placed in water and treated with cytokinin to promote regeneration
- Imaging of stem tips (below) showing various stages of callus and shoot development

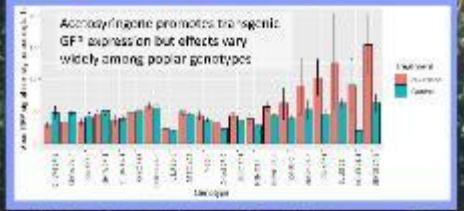
In vitro

- *In vitro* treatments used in GWAS were selected for high heritability of regeneration
- Imaging of whole plates with each explant at various stages of regeneration

Spectral analysis

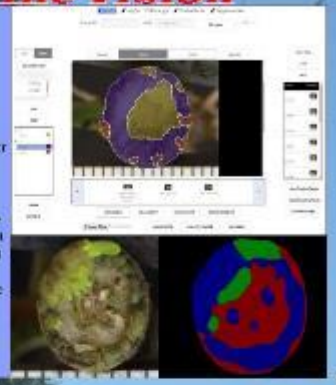
- Laser excites fluorophores (including endogenous GFP)
- Hyperspectral camera collects visible-IR spectra of fluorescence for each pixel; each fluorophore is identified, quantified
- Statistical analysis determines effects of treatments on GFP signal, an indicator of transformation efficiency
- System currently being used to test effects of particular treatments (e.g., acetosyringone, lipoic acid, *Agrobacterium* strain) on transformation rates and for GWAS of *in vitro* regeneration



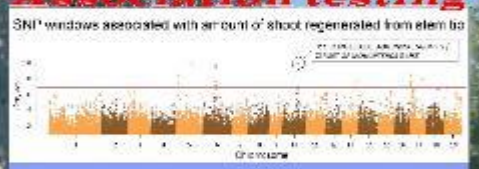


Machine vision

- Through a GUI, different types of tissue (callus, shoot, root) are annotated
- Annotated images are used to train neural networks for segmentation and classification
- Trained networks divide images into callus (blue), shoot (green), stem (red) and background
- Calculations of the relative size of different types of tissues are used as traits for GWAS



Association testing



- The GWAS method Sequence Kernel Association Test (SKAT) was used to collapse SNPs into 5kb windows (shown by dots above) and test for their combined effect on shoot regenerated from cut stem tips
- Shoot area was significantly associated with a homolog of a known shoot regulator that mediates auxin-cytokinin crosstalk in Arabidopsis
- Epistasis analysis (BOOST method in PLINK) indicates that the effect of this gene depends on a homolog of *LONESOME HIGHWAY* (*LHW*; $p = 5.19 \times 10^{-8}$, But fewer significant with 545,450 pairwise tests for epistasis between SNPs of interest). This suggests conservation of function from Arabidopsis, in which *ABNORMAL SHOOT 2 (ABS2)*, and *LHW* must interact with one another to regulate transcription in pluripotent cells of the procambium (Ohashi-Ito et al., *Dev Cell* 2013)

Outlook

- Convolutional neural networks and hyperspectral imaging provide new opportunities for *in vitro* optimization and genetic discovery by enabling precise, high-throughput phenotyping of complex traits
- Characterization of the genetic basis of regeneration offers opportunities for converting poor responders to regenerate into efficient responders via overexpression or knockdown/out of developmental regulators
- Efficient transformation of recalcitrant genotypes may involve a combination of *in vitro* treatments including with acetosyringone, as well as genetic treatments such as overexpression of *ABS2* and other developmental regulators

Acknowledgements

We thank the National Science Foundation Plant Genome Research Program for support (IOS #1546868, Analysis of genes affecting plant regeneration and transformation in poplar), and members of the Tree Biosafety and Genomics Research Cooperative at OSU for its long-term support of the Strauss laboratory.