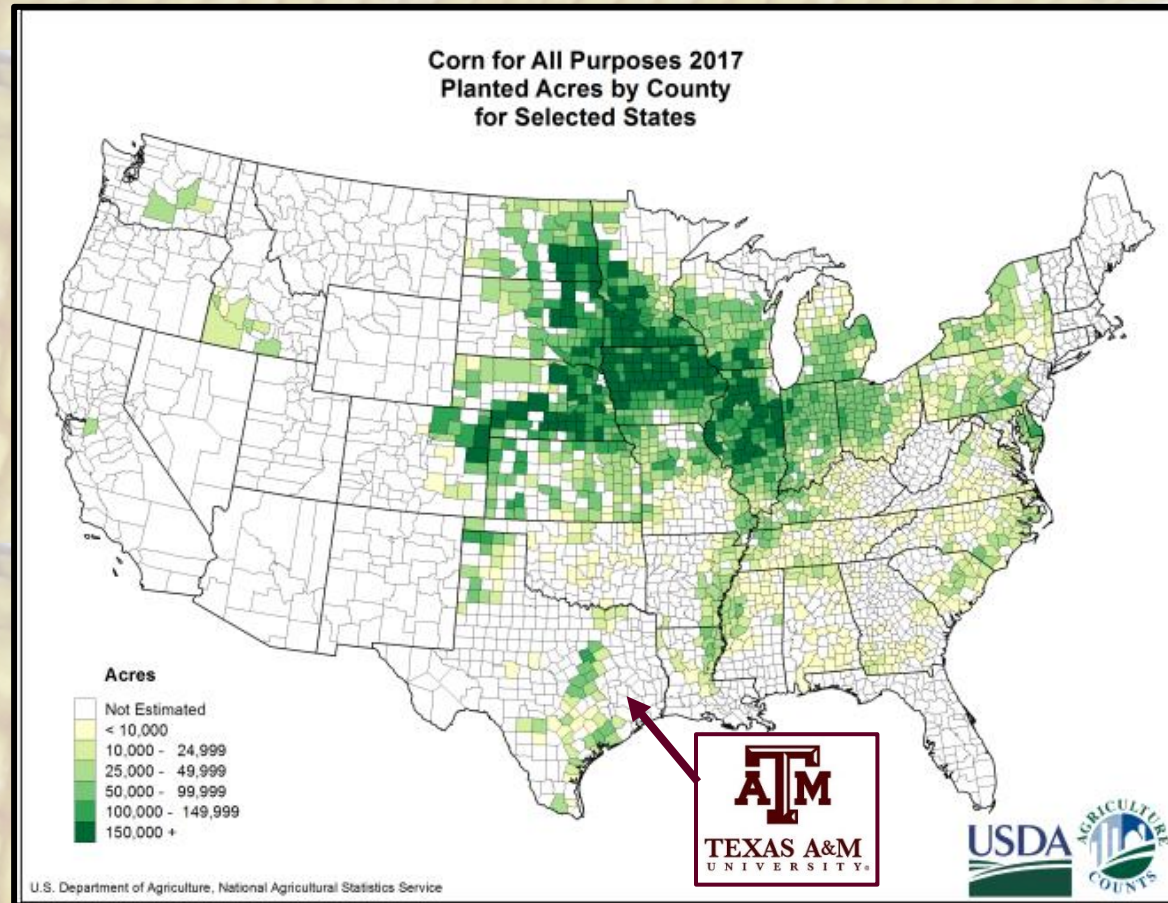# Capturing nature AND nurture with temporal field phenomics to breed better crops

Seth C. Murray, **et al.**

Professor and Eugene Butler Endowed Chair
Department of Soil and Crop Sciences
Texas A&M University

AGBIODATA
AUGUST 7, 2024

TEXAS A&M
AGRILIFE
RESEARCH

@DrSethMurray

Most maize (corn, *Zea mays* L.) is not sweet corn
Corn - $51.5 billion, 89 million acres, 404 million tons
# Sweet corn - $860 million, 495,000 acres, 4 million tons

Corn for All Purposes 2017
Planted Acres by County
for Selected States

Acres

	Not Estimated
	< 10,000
	10,000 - 24,999
	25,000 - 49,999
	50,000 - 99,999
	100,000 - 149,999
	150,000 +

U.S. Department of Agriculture, National Agricultural Statistics Service

TEXAS A&M UNIVERSITY

USDA
AGRICULTURE COUNTS

Zea mays
(Corn)
Commercial
Hybrid

Zea mays
(Corn)
Commercial
Hybrid

Why is a public corn breeding program in Texas needed?

Barrero et al. 2013, Field Crops Research; Data from USDA- NASS
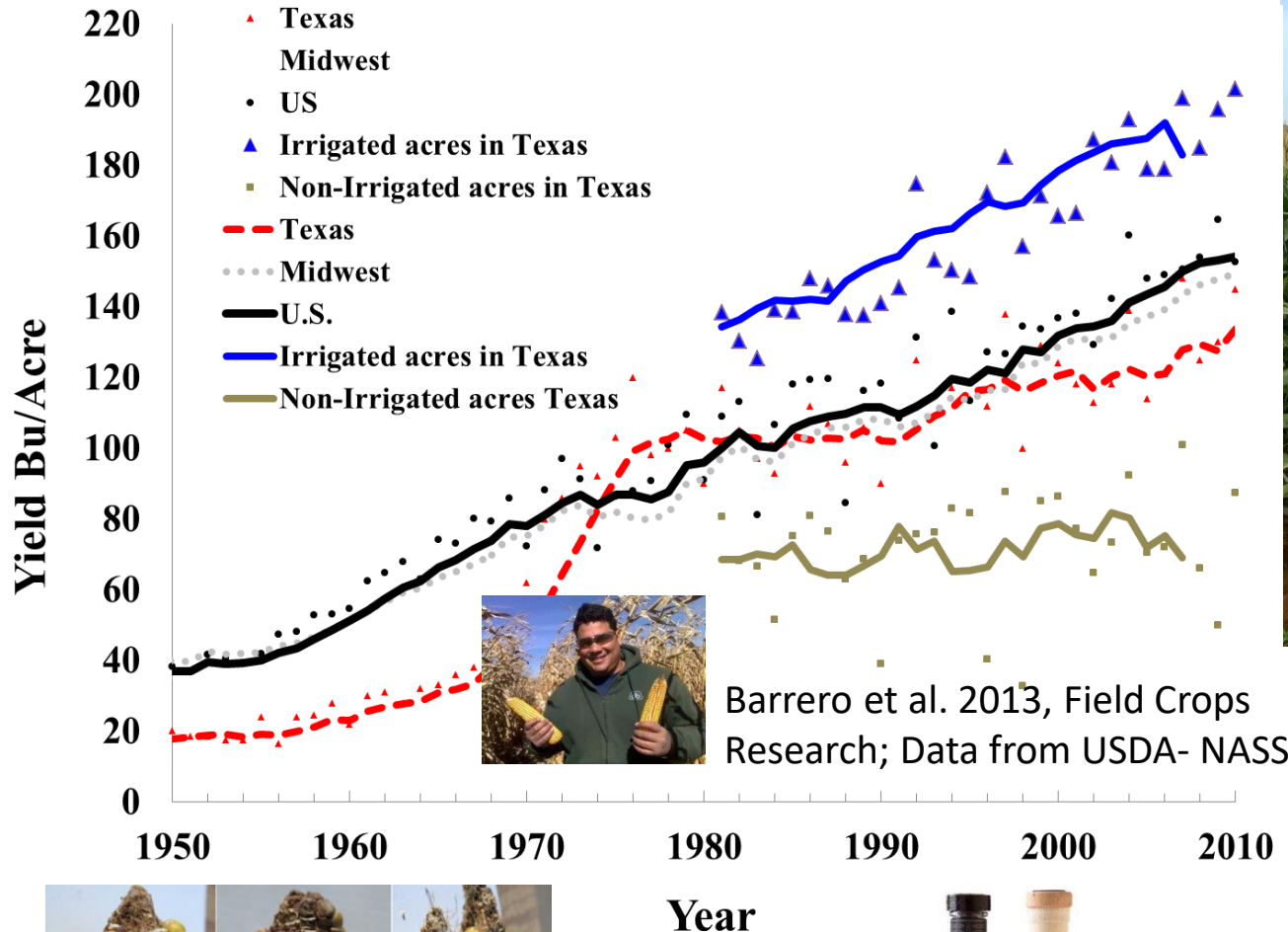
Murray et al. 2019, Journal of Plant Registrations

U.S. corn yield from 1866 to 2016
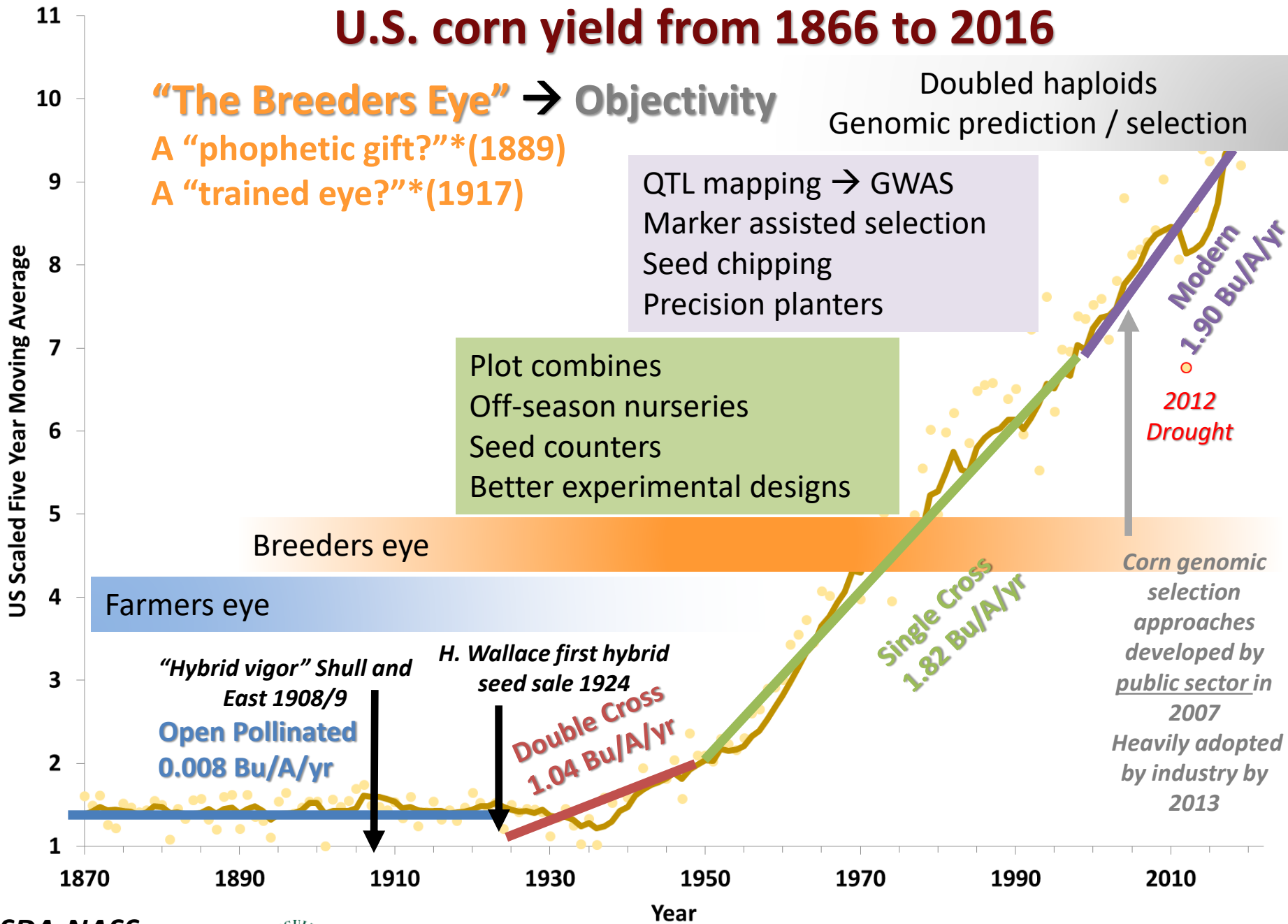
Improvements other than yield are needed

Reduced input use   **Ecosystem service provision**

Human health & nutrition   **Climate change resilience**

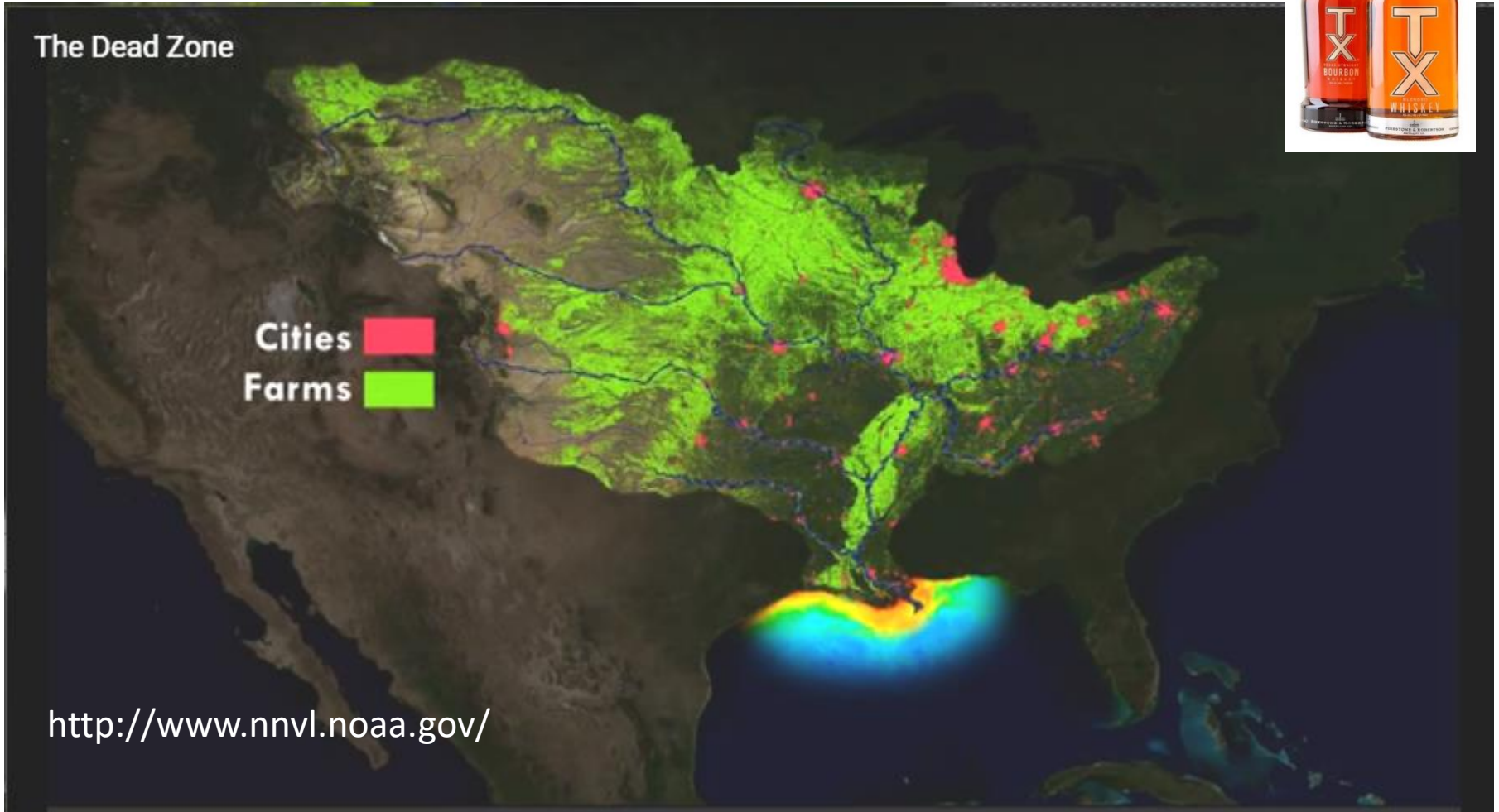**Farmer profitability**   **Flavor**

The Dead Zone

Cities

Farms

http://www.nnvl.noaa.gov/

Importance of language

EXPERT VIEW

Functional phenomics: an emerging field integrating high-throughput phenotyping, physiology, and bioinformatics

Larry M. York

Noble Research Institute, Ardmore, OK 73401, USA

Correspondence: lmyork@noble.org

Phenotype – "…the totality of a plant's observable characteristics"
Measured because believed to be of interest (the dependent trait)

Phene – "…a more elemental component of phenotype"

Phenome – Totality of a plant's measurable and estimable physical characteristics (measured because it can be used for prediction)

Phenomic selection is <u>NOT the same</u> as phenotypic selection

* **Unmanned Aerial Systems (UAS)**
* **Unoccupied Aerial Systems (UAS)**
* **Uncrewed Aerial System (UAS)**
* **Unmanned Aerial Vehicles (UAV)**
* **Drones**

Multi-discipline language issues are not new
* **Multidimensional scaling (MDS)**
* **Principal Component Analyses (PCA)**

Three genomic methodological waves and generalizations

1986	2005	2007
QTL Linkage Mapping	Association Mapping (GWAS)	Genomic Selection / Genomewide Selection (GS)
Find loci Map based cloning → find genes	Find loci → find genes	Select the best individual
Bi-parental families	Diverse populations	Varied, mostly narrow breeding populations
Inference to specific populations – epistasis issues	Alleles work across populations structure	Predictive across training germplasm
Knowledge, publications, ?crop improvement?	Knowledge, publications, ?crop improvement?	Crop improvement, knowledge ?Publication?

Discovery / Estimation   **Prediction**

Simple trait measurements (e.g. heights)   **Value indicies (yield, sustainability, etc.)** $

High throughput field phenotyping

1) **Automate routine measurements**
   - **Plant height**
   - Estimate grain yield (sorghum, wheat)
   - **Estimate disease**

2) **Find new signatures of "eliteness"**
   - **Temporal growth patterns / biomass**
   - **Senescence and grain filling period**

3) **Phenomic selection**

4) **Deep learning and AI tools to directly use images**

5) **Identify stress signatures for farmers management**

6) **Identify new phenotypes and mechanisms of biological importance**

Dr. Dale Cope   Dr. Sorin Popescu   Dr. John Valasek

i = Selection intensity = Grow more plants

Crop	Nursery	Yield trials
Corn (Murray)	10,000 plots	17,000 plots
Sorghum (Rooney)	15,000 plots total	20,000 plots +
Wheat (Ibrahim)	5,000 plots	5,000 plots x reps

Private sector is at a much larger scale for a few major crops but working with less diversity to develop "mega-varieties"

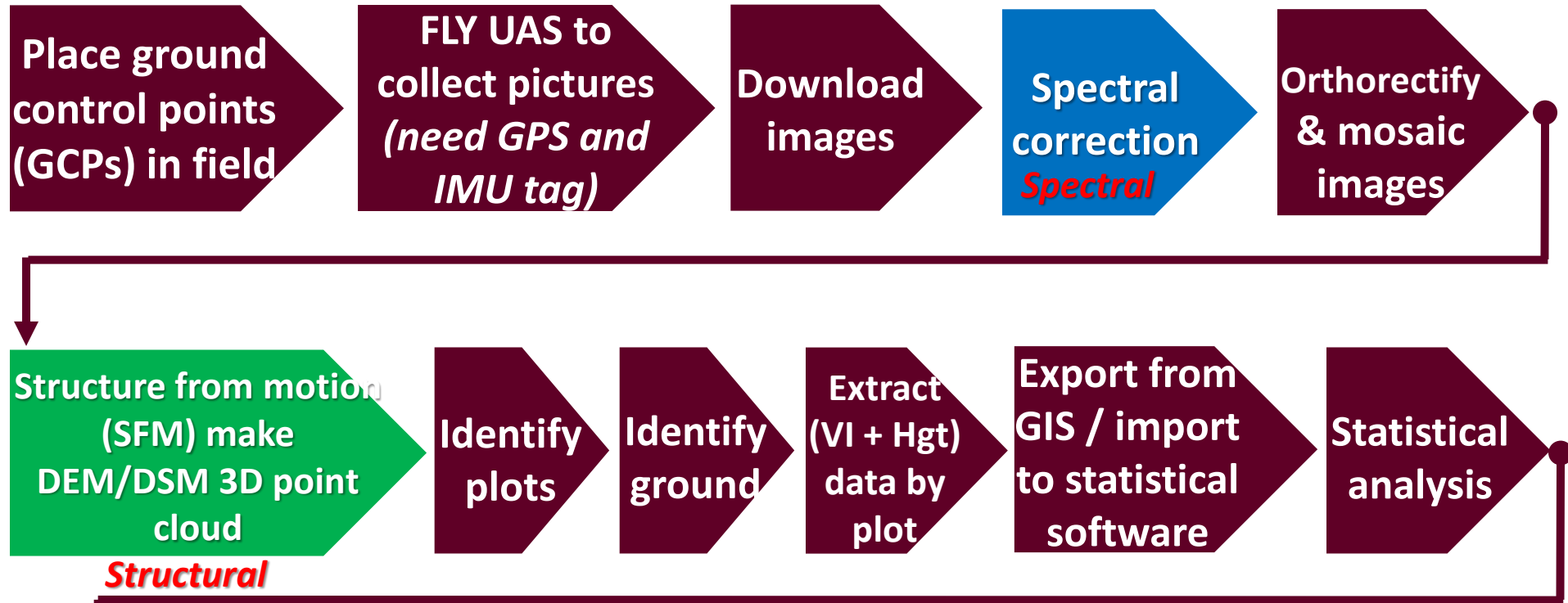# High Throughput can produce better decisions than high accuracy when phenotyping plant populations

LANE AND MURRAY

Crop Science | 3305
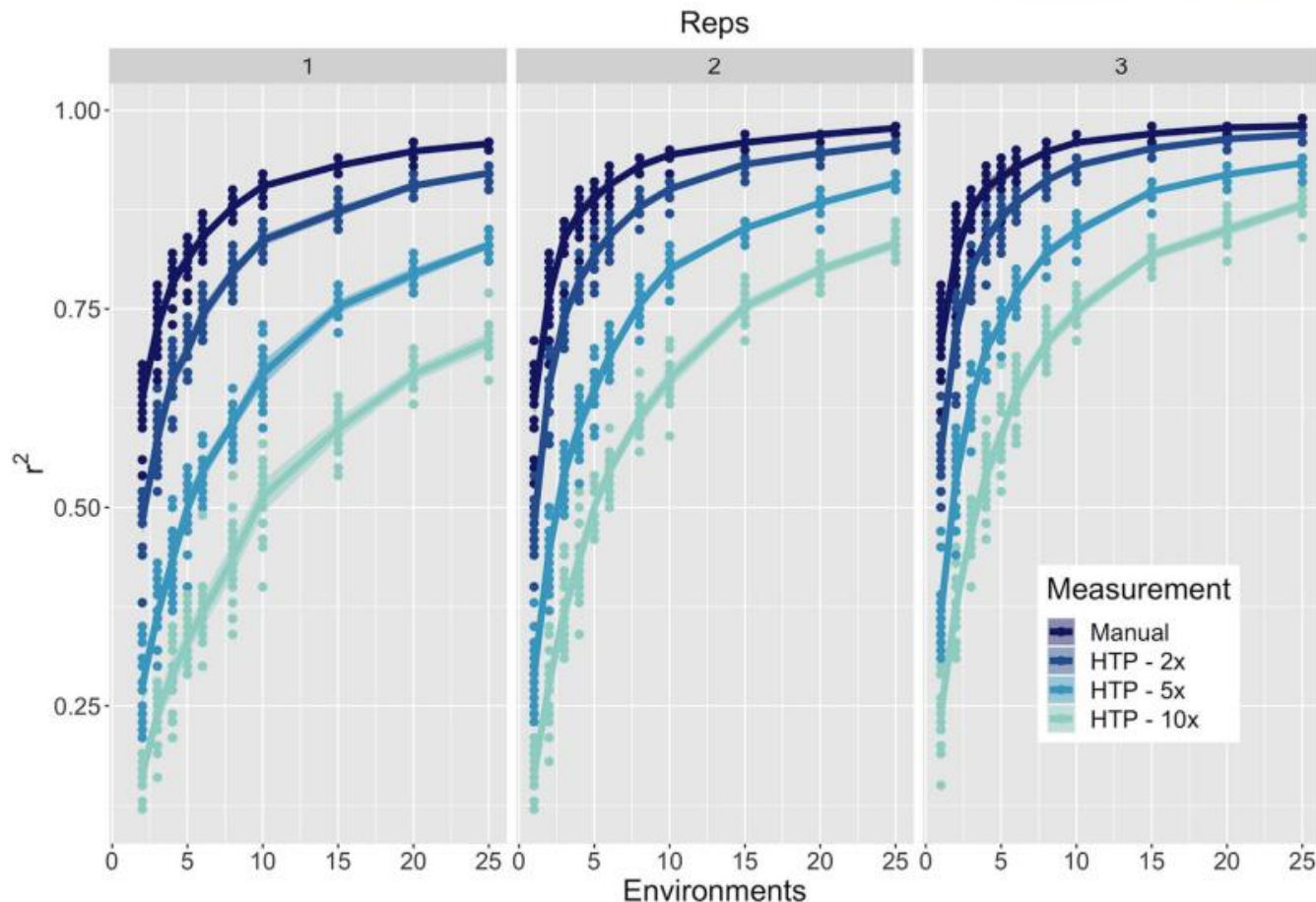
FIGURE 1 Effect of increasing environments and replicates on best linear unbiased estimation (BLUE) accuracy in a simulated F_3 population: Correlations (r^2) between true simulated and statistically estimated genotypic entry values (BLUEs) under four measurement error rates and differing number of environments, divided along the x-axis by number of replicates. Four different error rates of measurement [manual, high throughput (HTP) with 2× error, HTP with 5× error, and HTP with 10× error] were used on populations of 300 genotypes from an F_3 population. Each simulated scenario was repeated 20 times. Shaded areas show a 95% confidence interval for the mean; the lines show the mean r^2 for each scenario for the predicted and simulated true genetic heights for each line
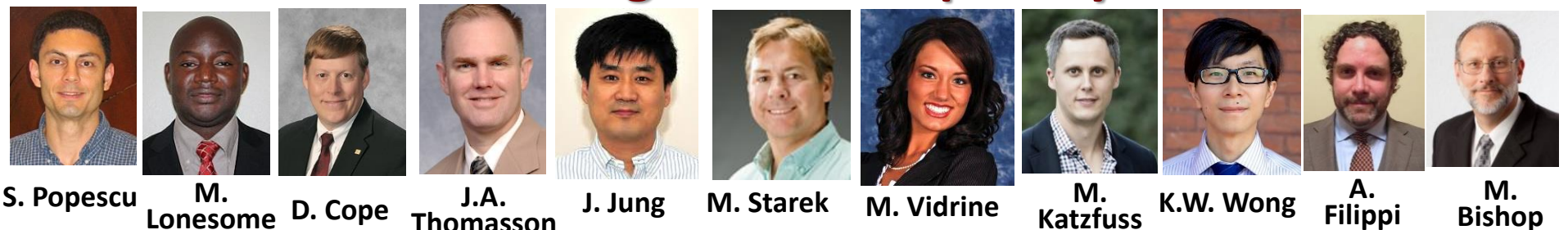
Holly Lane, M.S.

https://acsess.onlinelibrary.wiley.com/doi/10.1002/csc2.20514

Steps to obtain and analyze UAS phenotyping data

Place ground control points (GCPs) in field → **FLY UAS to collect pictures** *(need GPS and IMU tag)* → **Download images** → **Spectral correction** *Spectral* → **Orthorectify & mosaic images**

Structure from motion (SFM) make DEM/DSM 3D point cloud *Structural* → **Identify plots** → **Identify ground** → **Extract (VI + Hgt) data by plot** → **Export from GIS / import to statistical software** → **Statistical analysis**

Estimate genotypic values → Determine measurements repeatability & value

Took a huge transdisciplinary team

S. Popescu | M. Lonesome | D. Cope | J.A. Thomasson | J. Jung | M. Starek | M. Vidrine | M. Katzfuss | K.W. Wong | A. Filippi | M. Bishop

Interdisciplinary challenges posed by agricultural photogrammetry

A)
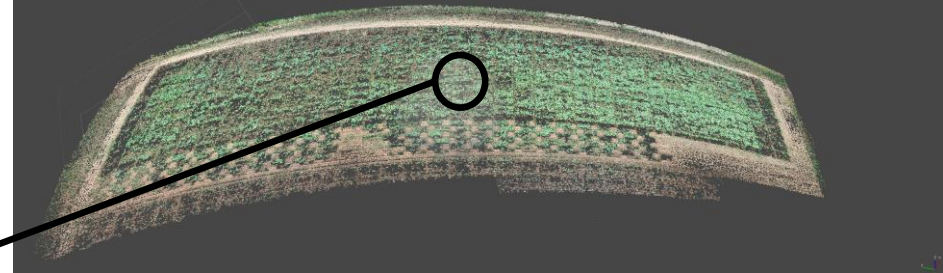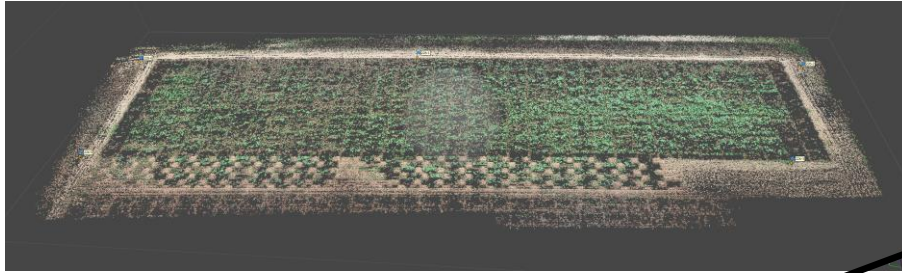
C)
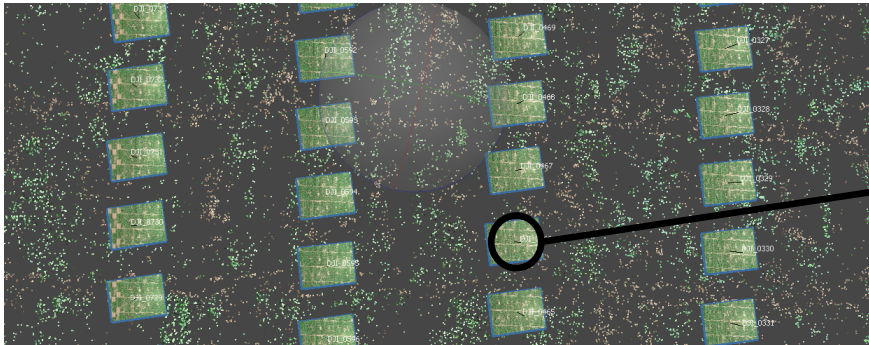
D)

"Hot spots" result in distortions and curvature in orthomosaics
Require corrections using ground control points or RTK UAVs, or obtaining oblique images

Evolution of our UAS phenotyping data pipeline

Plant and place ground control points

↓

FLY UAS to collect pictures (need GPS and IMU tag, plus RTK)
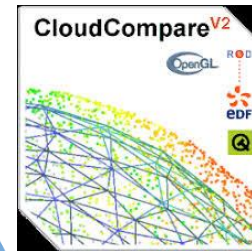
↓

Download images

↓

Orthomosaic

↓

Shape File

↓

Fact 4 Process

↓

Mask Soil

↓

Extract Vegetative Indices

↓

Analyze Data

Metashape

R/UASTools
https://github.com/andersst91/UAStools

GitHub

R

QGIS 3.14 Pi

CloudCompare V2
OpenGL R∙D
eDF
Q

FIELDimageR

jmp
Statistical Discovery.™ From SAS.

USDA
United States Department of Agriculture
National Institute of Food and Agriculture

AG2Pi
Agricultural Genome to Phenome Initiative

Dr. Alper Adak
Postdoc
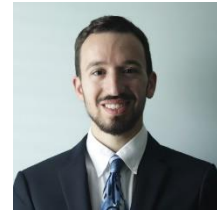
Aaron DeSalvio
PhD student

Mustafa Arik
Research Assistant

R/UASTools

GitHub

AB-line UTM Coordinates

+

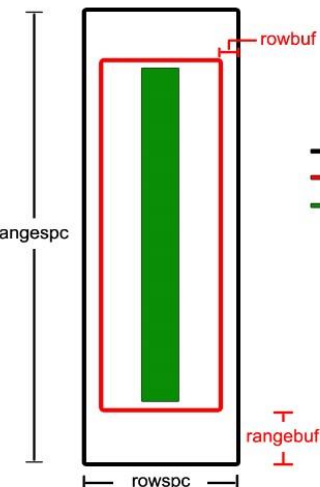**Experimental Field Design**

+

Plot Dimensions

R/UAStools:: plotshpcreate

Anderson and Murray. 2020. Frontiers in Plant Science

frontiers in Plant Science

Dr. Steven Anderson

Prediction of Maize Grain Yield before Maturity Using Improved Temporal Height Estimates of Unmanned Aerial Systems

Steven L. Anderson II, Seth C. Murray,* Lonesome Malambo, Colby Ratcliff, Sorin Popescu, Dale Cope, Anjin Chang, Jinha Jung, and J. Alex Thomasson
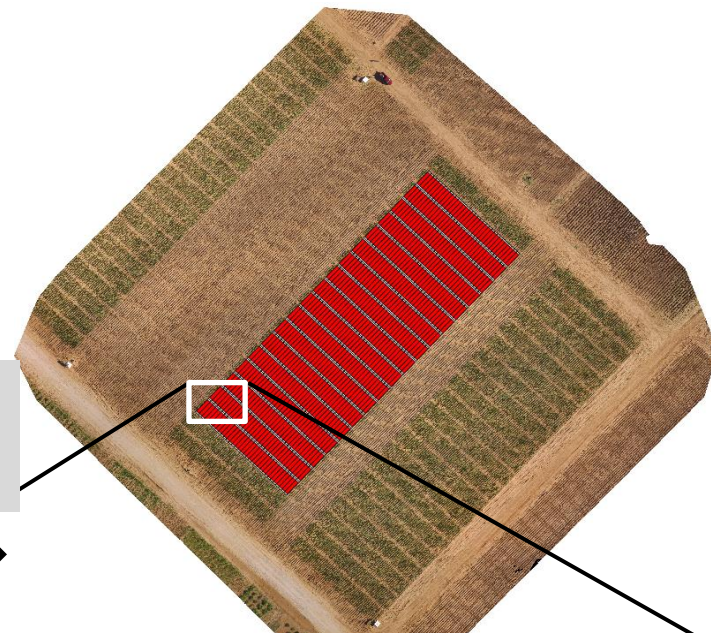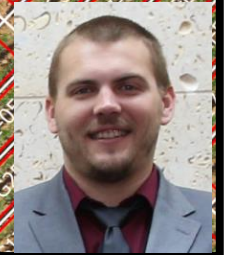
Temporal plant growth curves segregate in a population

S. Anderson PhD student

D. Cope

S. Popescu

M. Lonesome
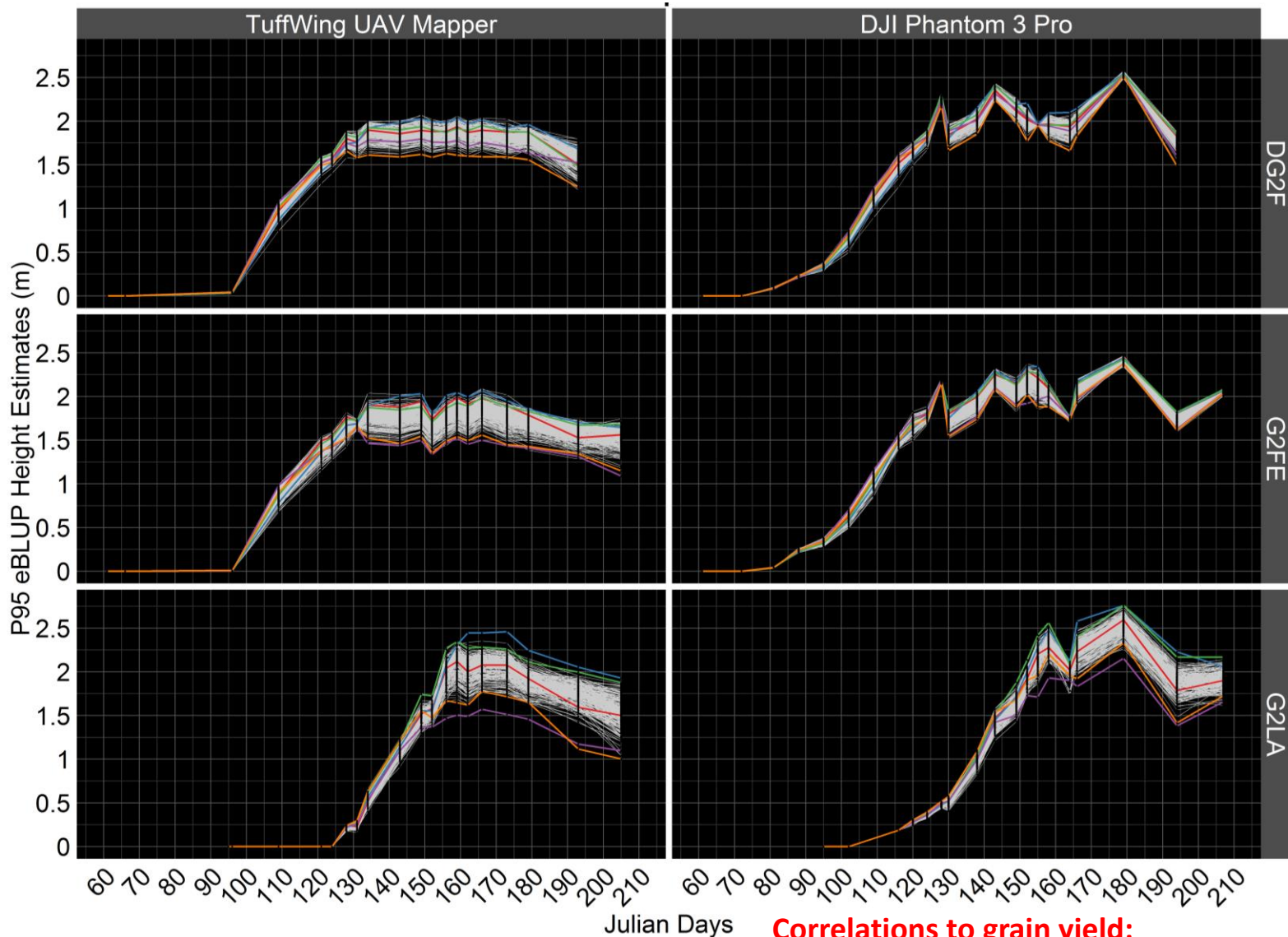
Correlations to grain yield:
UAS (r = 0.36–0.48), manual (r = 0.23–0.28)

Genetic effects from individual loci change over plant growth

Tx740 x NC356 irrigated trials

QTL
5_119
6_35
6_75
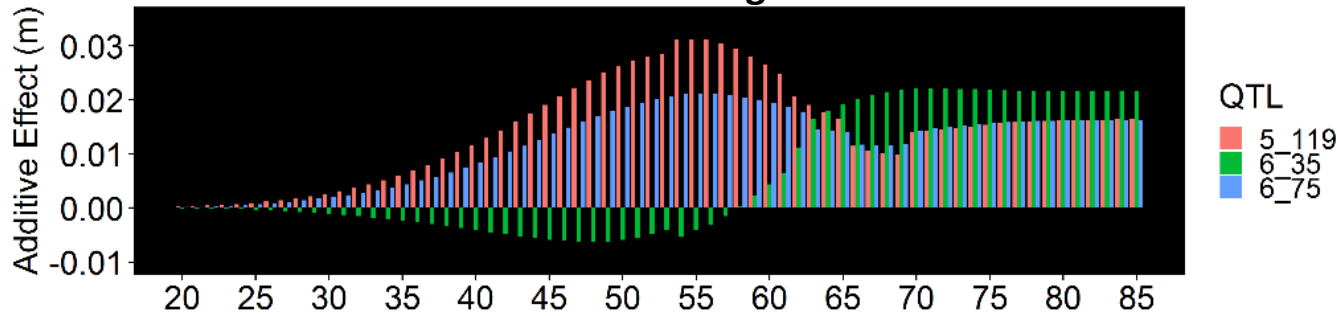
Tx740 x NC356 non- irrigated trials

QTL
1_73
2_55
6_0
6_67

Ki3 x NC356 non- irrigated trials

QTL
4_61
4_62
4_120
5_38

Days After Sowing (DAS)

Dr. Steven Anderson

UAS validated loci for plant height that terminal manual measures could not

1) Early study on discovering candidate gene for PHT and GY
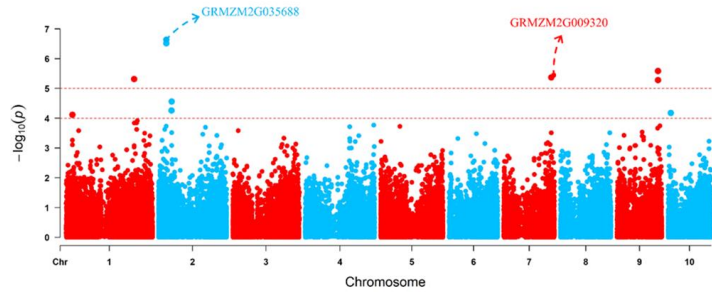
2) KASP marker development to advance the NILs
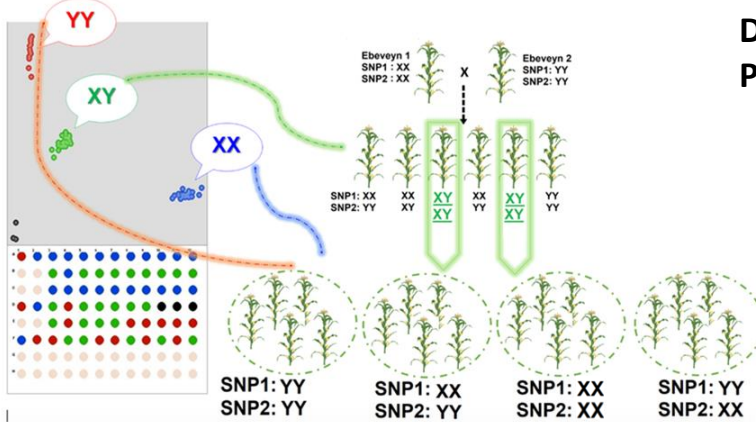
3) Conduct the FHTP to get temporal PHT
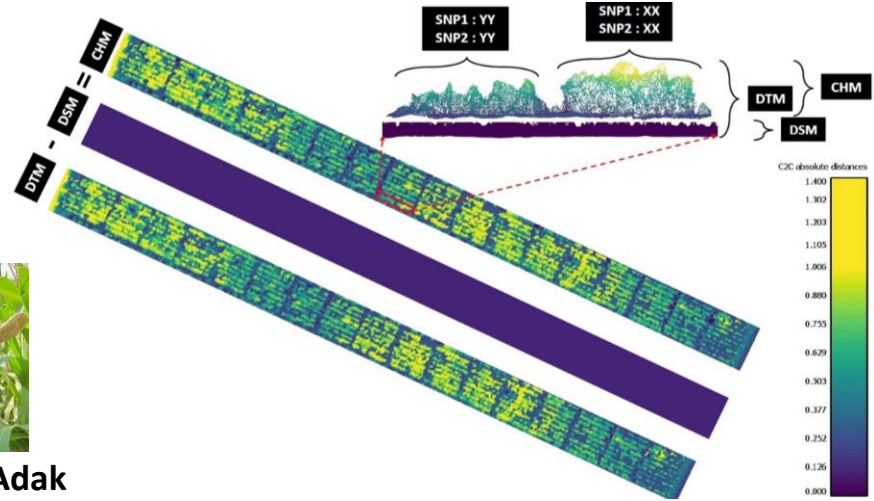
Dr. Alper Adak
Postdoc

4) Compare temporal PHT between NILs temporally
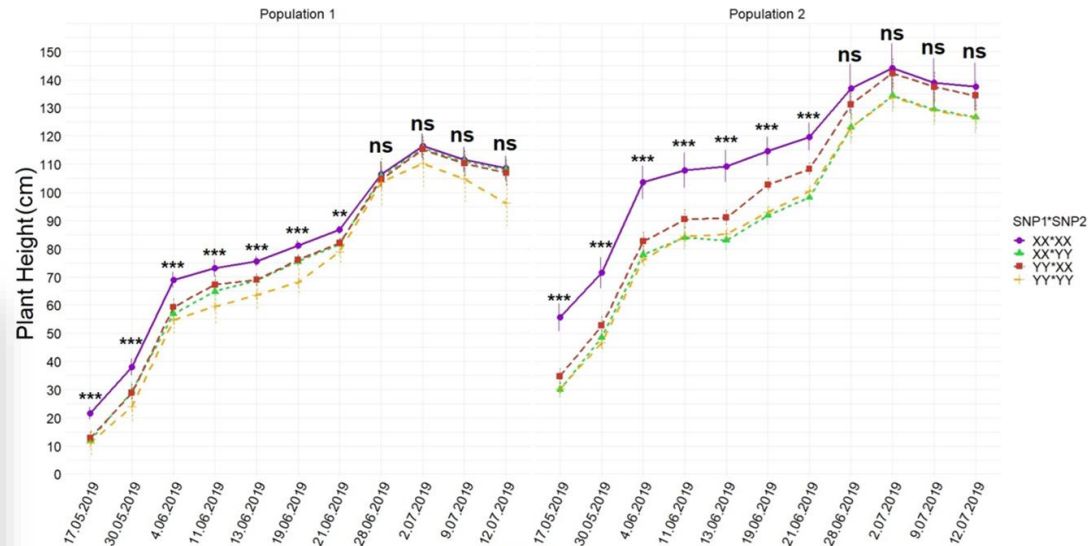
Validation of functional polymorphisms affecting maize plant height by unoccupied aerial systems discovers novel temporal phenotypes

Alper Adak [ID],[1] Clarissa Conrad,[1] Yuanyuan Chen,[1,2] Scott C. Wilde,[1] Seth C. Murray,[1]* Steven L. Anderson II [ID],[1,3] and Nithya K. Subramanian[1]

A large phenomic dataset showing excessive temporal variations in plant height

Weibull fit

$$a \cdot \left(1 - \mathrm{Exp}\left(-\left(\frac{DAP}{b}\right)^{c}\right)\right)$$

a = Asymptote
b = Inflection Point
c = Growth Rate
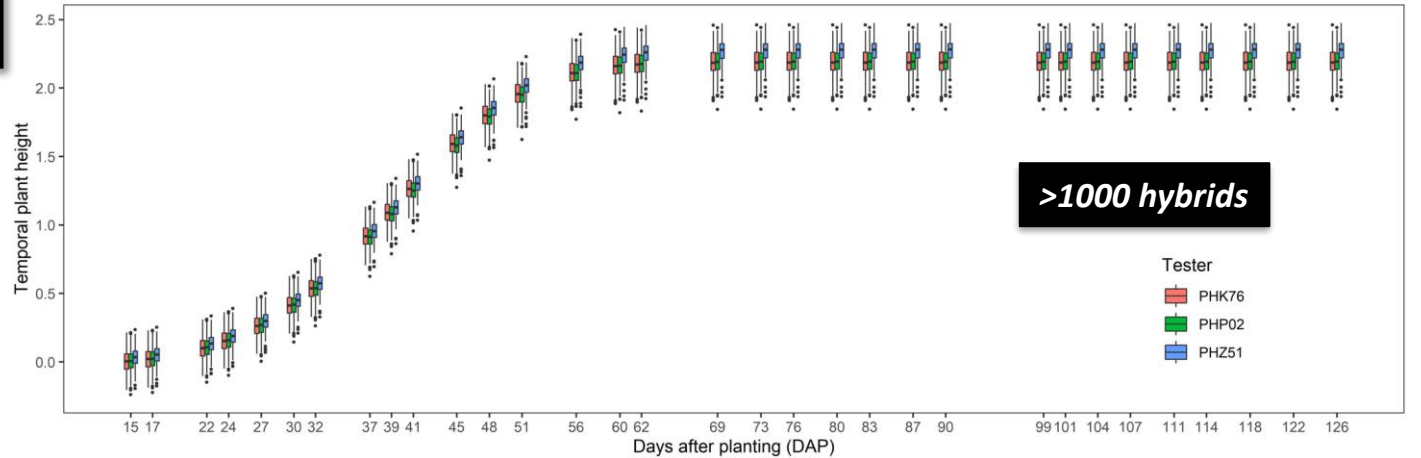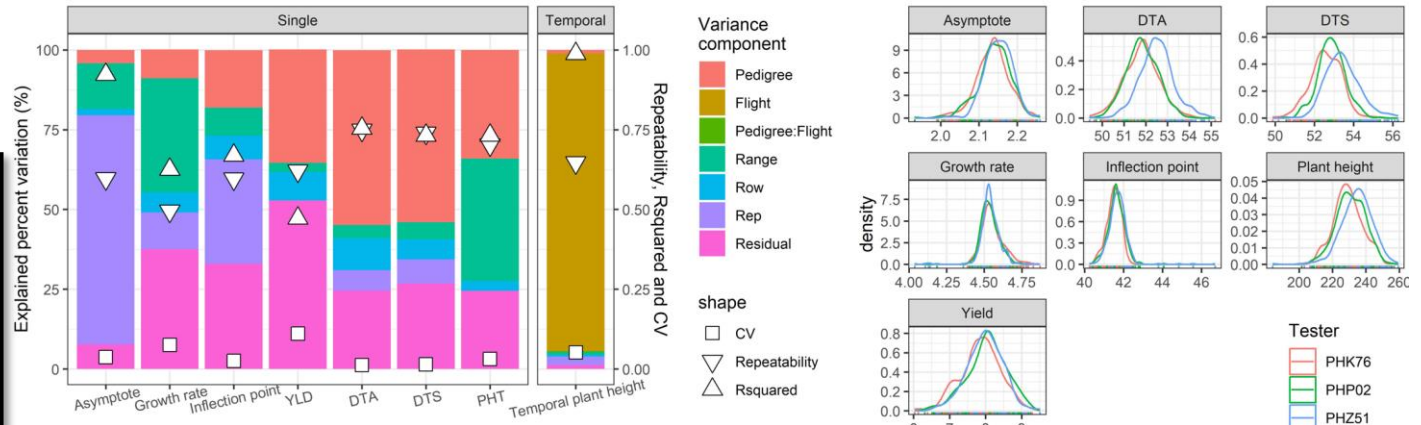
University of Missouri

Dr. Jacob Washburn
USDA-ARS / U of Missouri

Dr. Alper Adak
Postdoc

>1000 hybrids

The Plant Phenome Journal · OPEN ACCESS · ASA | CSSA

ORIGINAL ARTICLE | Open Access | (cc) (i)

High temporal resolution unoccupied aerial systems phenotyping provides unique information between flight dates

Jacob D. Washburn ✉ Alper Adak, Aaron J. DeSalvio, Mustafa A. Arik, Seth C. Murray

Data processing

Table. Available indices in *FIELDimageR*. Any other index can be implemented using the option *myIndex* and the new formula (*FIELDimageR::fieldIndex*).

Description	Index	Formula	Related traits	References
Brightness Index	BI	sqrt((R^2+G^2+B^2)/3)	Vegetation coverage, water content	Richardson and Wiegand (1977)
Soil Color Index	SCI	(R-G)/(R+G)	Soil color	Mathieu et al. (1998)
Green Leaf Index	GLI	(2*G-R-B)/(2*G+R+B)	Chlorophyll	Louhaichi et al. (2001)
Primary Colors Hue Index	HI	(2*R-G-B)/(G-B)	Soil color	Escadafal et al. (1994)
Normalized Green Red Difference Index	NGRDI	(G-R)/(G+R)	Chlorophyll, biomass, water content	Tucker (1979)
Spectral Slope Saturation Index	SI	(R-B)/(R+B)	Soil color	Escadafal et al. (1994)
Visible Atmospherically Resistant Index	VARI	(G-R)/(G+R-B)	Canopy, biomass, chlorophyll	Gitelson et al. (2002)
Overall Hue Index#	HUE	atan(2*(B-G-R)/30.5*(G-R))	Soil color	Escadafal et al. (1994)
Blue Green Pigment Index	BGI	B/G	Chlorophyll, LAI	Zarco-Tejada et al. (2005)
Plant Senescence Reflectance Index	PSRI	(R-G)/(RE)	Chlorophyll, nitrogen, maturity	Merzlyak et al. (1999)
Normalized Difference Vegetation Index	NDVI	(NIR-R)/(NIR+R)	Chlorophyll, LAI, biomass, yield	Rouse et al. (1974)
Green Normalized Difference Vegetation Index	GNDVI	(NIR-G)/(NIR+G)	Chlorophyll, LAI, nitrogen, protein content, water content	Gitelson et al. (1996)

|	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
	Flight_dat	Plot_ID	ID	Red	Green	Blue	BI	GLI	NGRDI	VARI	BGI	BCC	CIVE	COM1	COM2	ExG	ExG2	ExGR	ExR	GmB	GmR	GdB	GdR	GCC	MExG
	20200529	3016763	1	0.30257	0.337805	0.267822	0.304442	0.085687	0.054633	0.092774	0.789688	0.293178	18.75004	20.04575	9.049477	0.105218	0.118398	0.019425	0.085793	0.069983	0.035235	1.277261	1.120272	0.372799	0.075545
	20200529	3016764	2	0.314704	0.340968	0.273255	0.311232	0.076491	0.040289	0.06774	0.795389	0.29185	18.75491	19.9883	9.043298	0.093978	0.105284	-0.00564	0.099617	0.067713	0.026265	1.268985	1.08671	0.368428	0.067122
	20200529	3016765	3	0.276143	0.317359	0.249631	0.28279	0.094419	0.067862	0.116042	0.783413	0.294134	18.74796	20.09213	9.053999	0.108945	0.130815	0.039704	0.06924	0.067728	0.041217	1.286186	1.151421	0.376938	0.078762
	20200529	3016766	4	0.27901	0.312365	0.251604	0.282412	0.083406	0.056013	0.096537	0.799465	0.296116	18.75403	20.0319	9.046817	0.094116	0.115079	0.015867	0.078249	0.060761	0.033355	1.259988	1.123283	0.371693	0.069311
	20200529	3016767	5	0.298039	0.334019	0.247147	0.295612	0.102138	0.055501	0.089443	0.737011	0.279069	18.74315	20.10582	9.057541	0.122853	0.141826	0.039618	0.083235	0.086872	0.03598	1.372513	1.121224	0.380609	0.081203
	20200529	3016768	6	0.268459	0.305387	0.239563	0.272837	0.091359	0.061936	0.105612	0.783194	0.292905	18.7504	20.07178	9.051356	0.102751	0.126468	0.032295	0.070456	0.065823	0.036928	1.287334	1.137827	0.375489	0.073576
	20200529	3016769	7	0.28294	0.315348	0.248568	0.283937	0.086013	0.052473	0.089065	0.784144	0.29141	18.75218	20.03959	9.048098	0.09919	0.118748	0.018423	0.080767	0.066781	0.032409	1.284671	1.115227	0.372916	0.070547
	20200529	3016770	8	0.280534	0.306217	0.242964	0.278135	0.08168	0.04374	0.073061	0.783843	0.289479	18.75637	20.00442	9.044149	0.088935	0.112671	0.002404	0.086531	0.063252	0.025683	1.287967	1.095007	0.37089	0.062891
	20200529	3016771	9	0.277682	0.31262	0.235076	0.277416	0.10056	0.057419	0.093479	0.745828	0.281695	18.74688	20.09556	9.055502	0.112481	0.139725	0.036345	0.076136	0.077543	0.034937	1.358502	1.126756	0.379908	0.075946
	20200529	3016772	10	0.264718	0.301495	0.226427	0.266454	0.1036	0.062603	0.102369	0.744925	0.282415	18.74685	20.1107	9.056861	0.111844	0.144146	0.042733	0.069111	0.075067	0.036777	1.360076	1.139384	0.381382	0.076057
	20200529	3016773	11	0.265673	0.298013	0.226555	0.265407	0.096076	0.055153	0.090697	0.756086	0.284212	18.75015	20.07369	9.052215	0.103798	0.133217	0.02987	0.073929	0.071458	0.03234	1.336736	1.121431	0.377739	0.070779
	20200529	3016774	12	0.264956	0.297998	0.228458	0.265689	0.095265	0.056509	0.094077	0.760577	0.285819	18.75058	20.07169	9.051812	0.102582	0.131942	0.029641	0.07294	0.06954	0.033042	1.326839	1.124414	0.377314	0.070802
	20200529	3016775	13	0.310076	0.329267	0.276202	0.306344	0.064067	0.032107	0.054703	0.823951	0.297434	18.7635	19.92316	9.035253	0.072255	0.087988	-0.03258	0.10484	0.053065	0.019191	1.226599	1.06934	0.362663	0.055528

Temporal phenomic data reveals patterns for vegetative indices (VIs) across the growing season
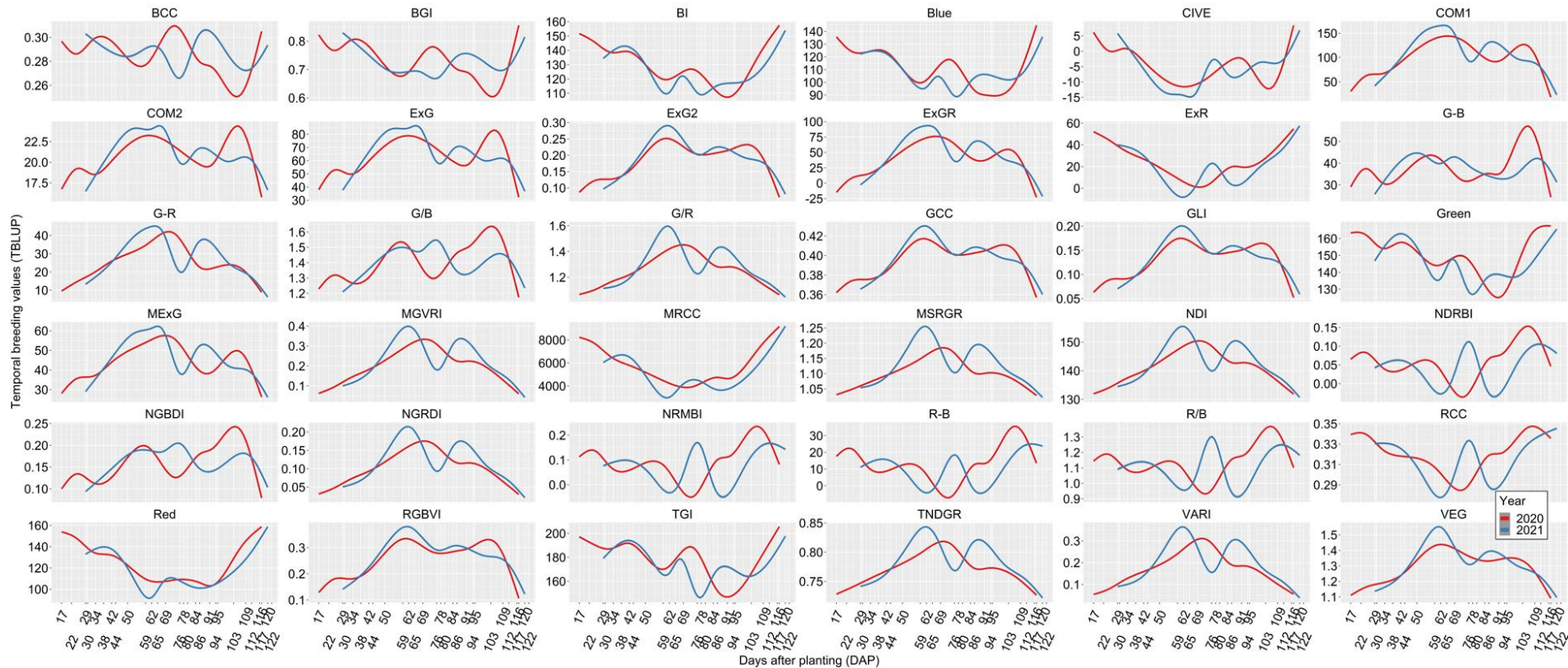
- Different temporal patterns for different VIs across flight times revealed extreme VI values at plant emergence, flowering times, and conclusion of reproductive stage
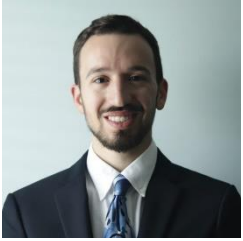
Novel approach to Southern Rust and senescence scoring

Aaron DeSalvio
PhD student

Alper Adak
Postdoc
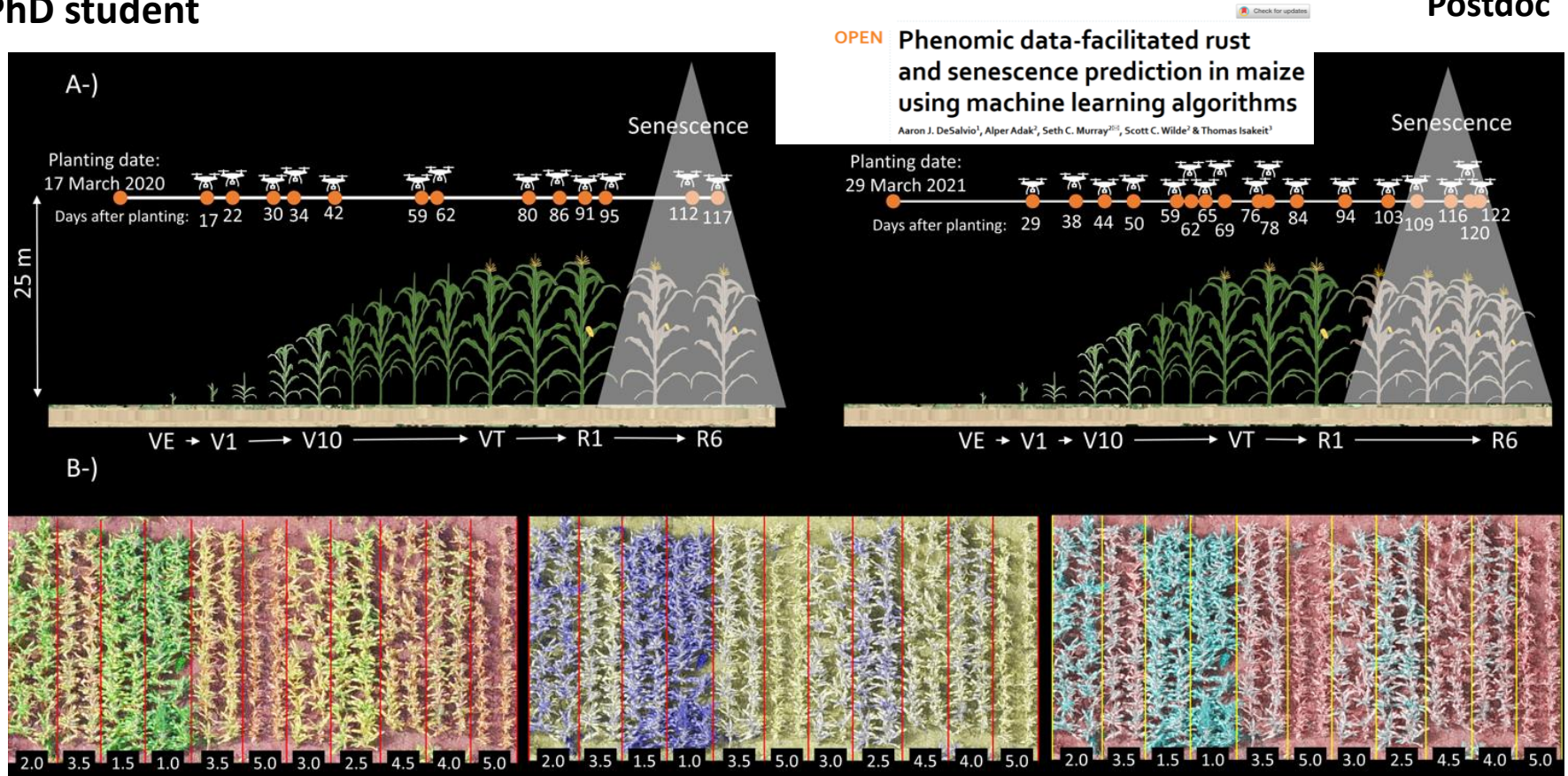
- Senescence annotated using drone images (orthomosaics) vs. hand-scoring in the field

A catalyst for future studies: grain filling and yield

F-) Correlation
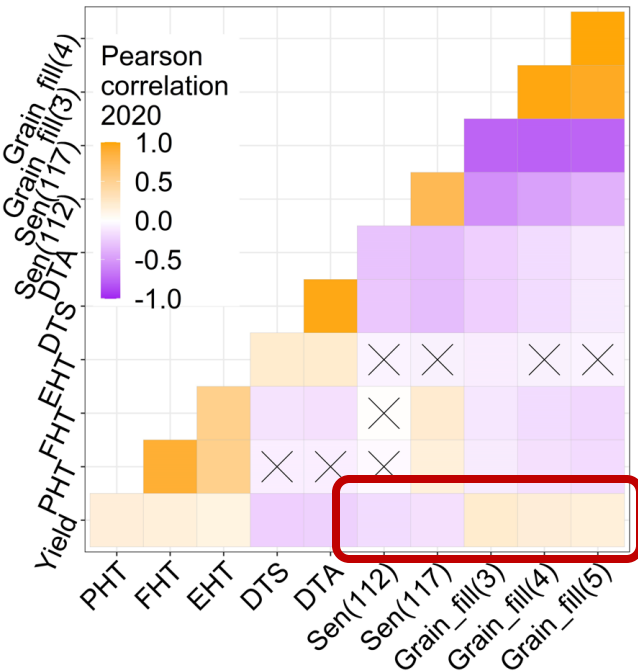
2020
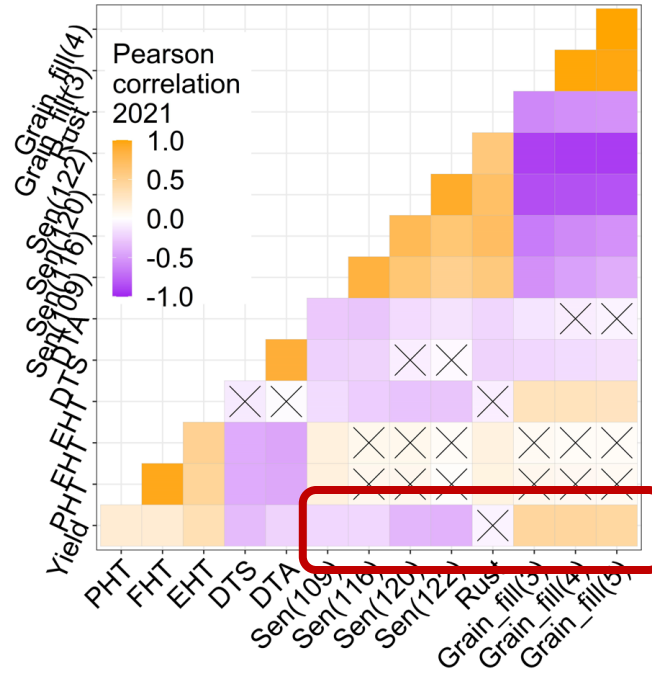
2021

Grain filling period had highest correlations with yield versus other traits (0.22 in 2020; 0.44 in 2021)

Grain filling period calculated as days between DTA and days to senescence as estimated by linear model

Variance explained by nested design
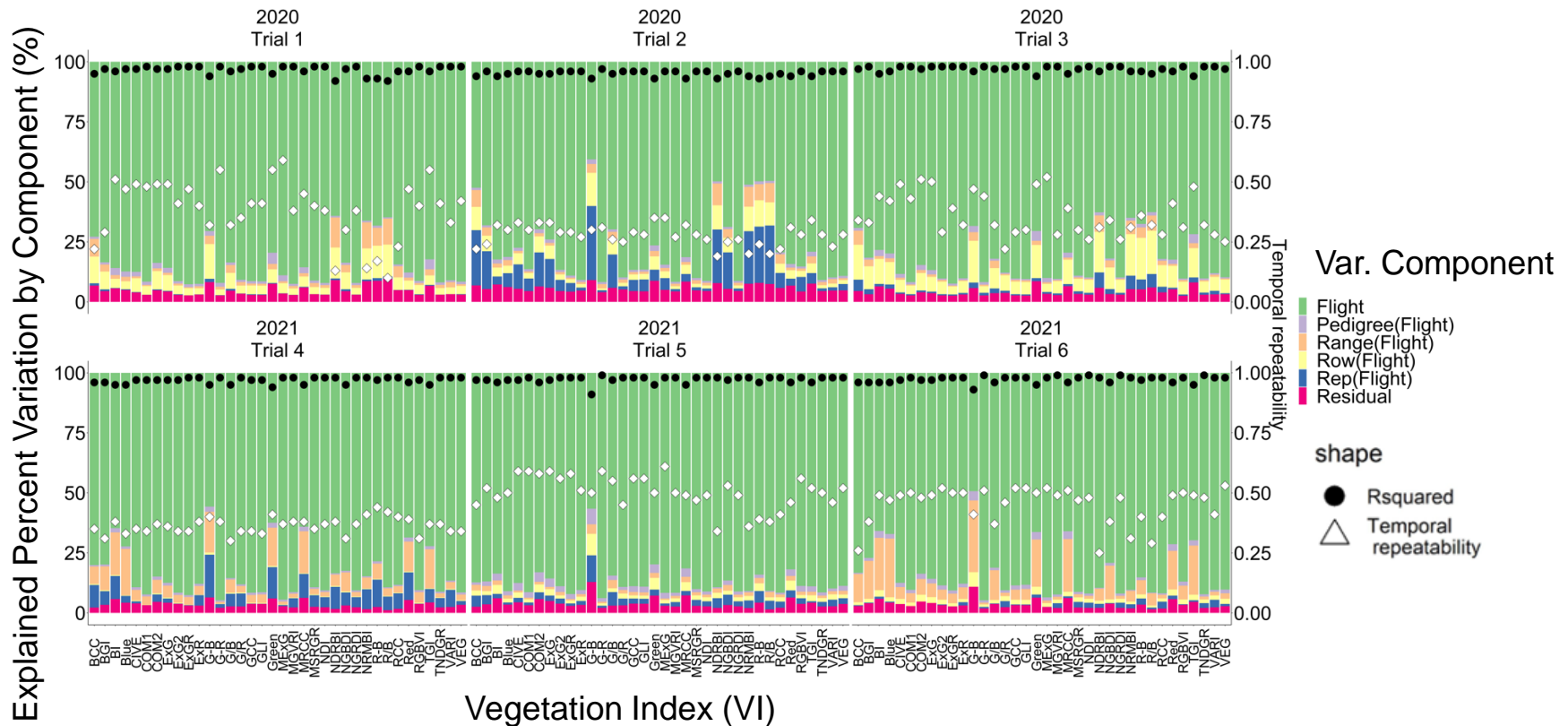
- Flight component in nested design explained highest percent of experimental variation for all VIs in both years

Phenomic prediction model assessment

Phenomic Prediction in Rust
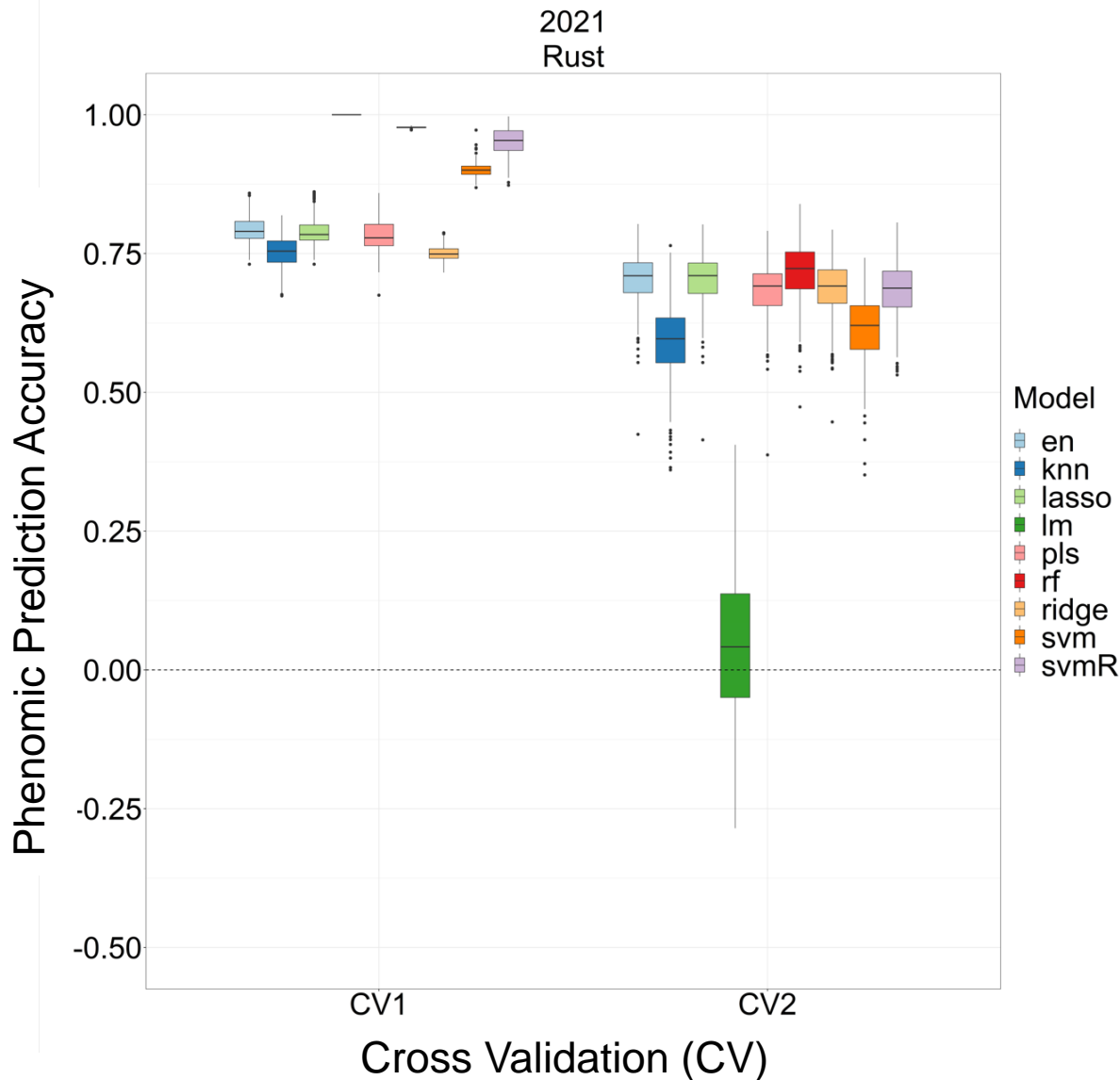
- Prediction accuracies: machine learning models outperformed linear model when predicting performance of untested genotypes

Used cross validation 500 prediction accuracies for each model
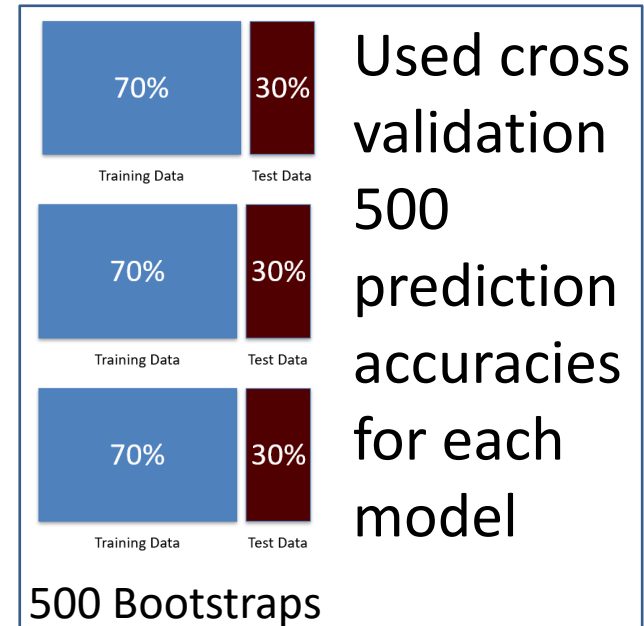
Variable Importance Scores

A-) Variable importance score of Sen(112) in 2020
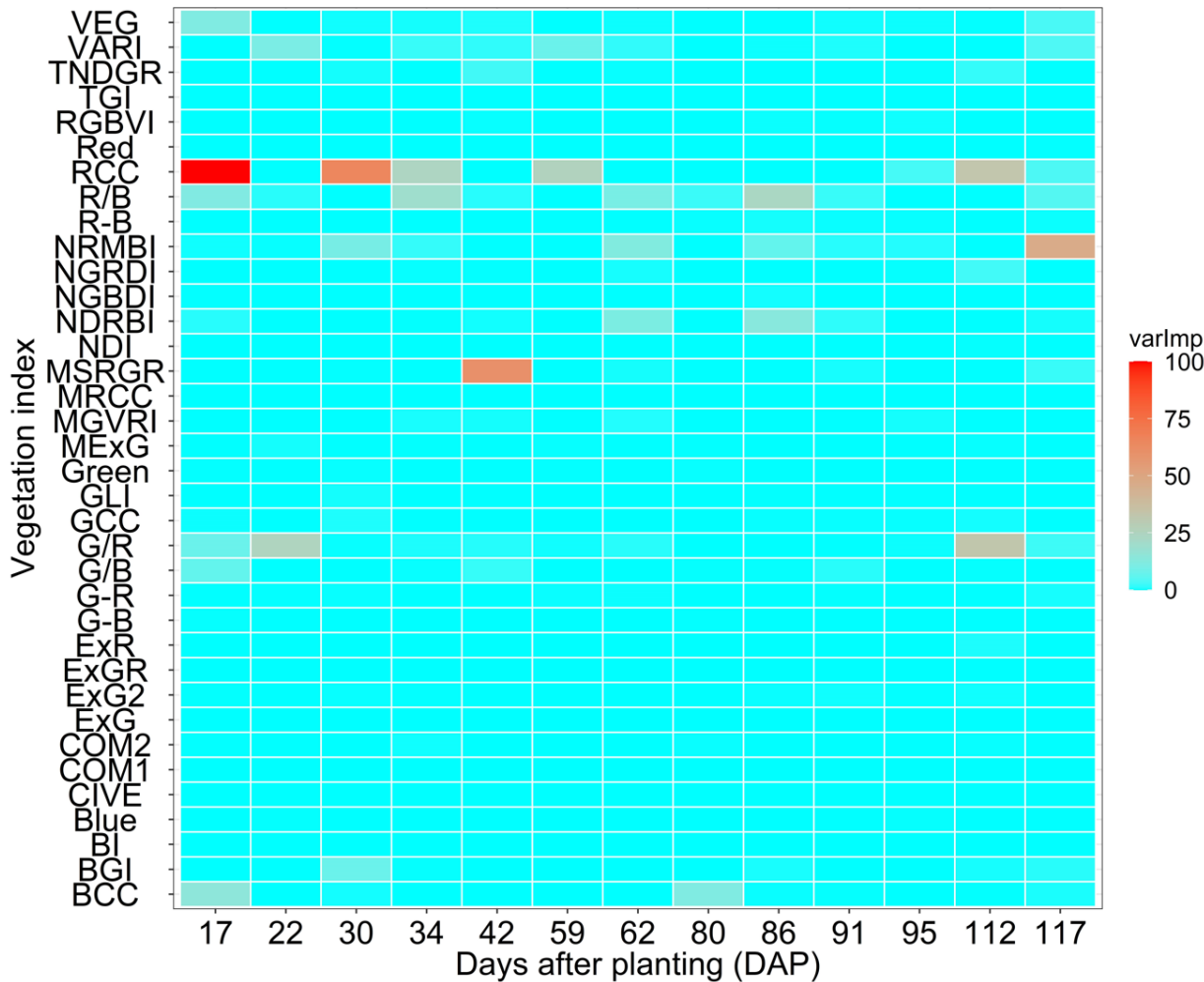
- Early, mid, and late-season phenotypic indicators revealed by variable importance scores

- Brighter red indicates VI was more important for model accuracy (had more predictive power)

May enable early predictions of senescence times for different genetic backgrounds; early warning for rust

Calculation of the days to senescence and grain filling period using the *temporal senescence progression data*

A-)

B-)

$$y = a * (1 - b * exp(-c * Flight))$$

C-)

$$Days\ to\ senescence\ (DTSE) = -\frac{ln\left(\frac{a-y}{a*b}\right)}{c}$$
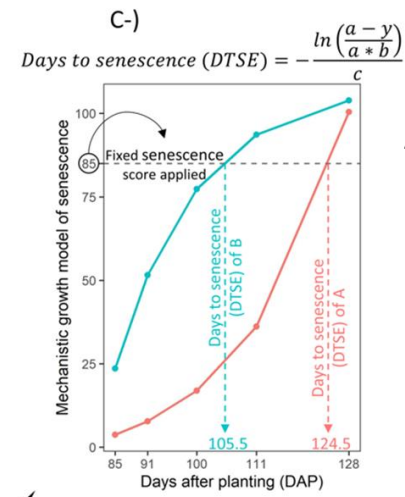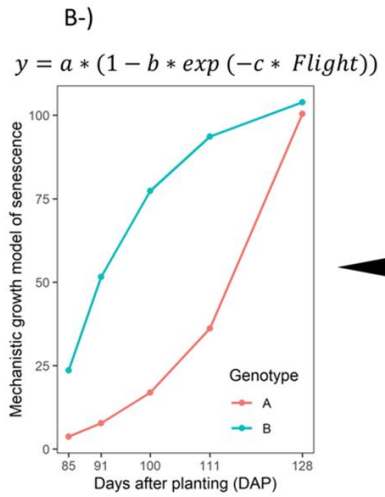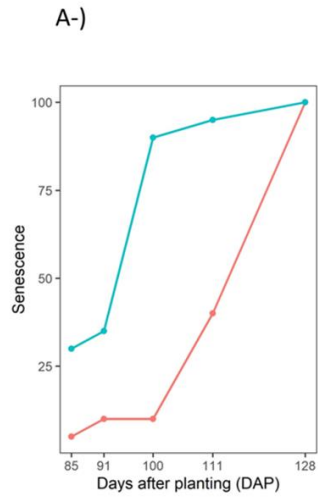
A) Plot based senescence scores

B) Mechanistic growth model

C) Days to senescence (DTSE)

D) Grain filling period

D-)
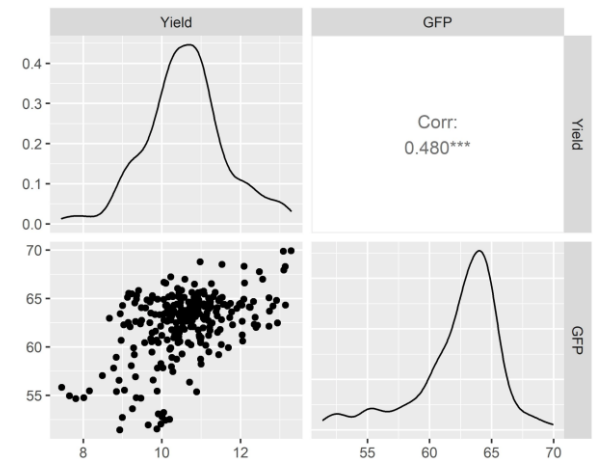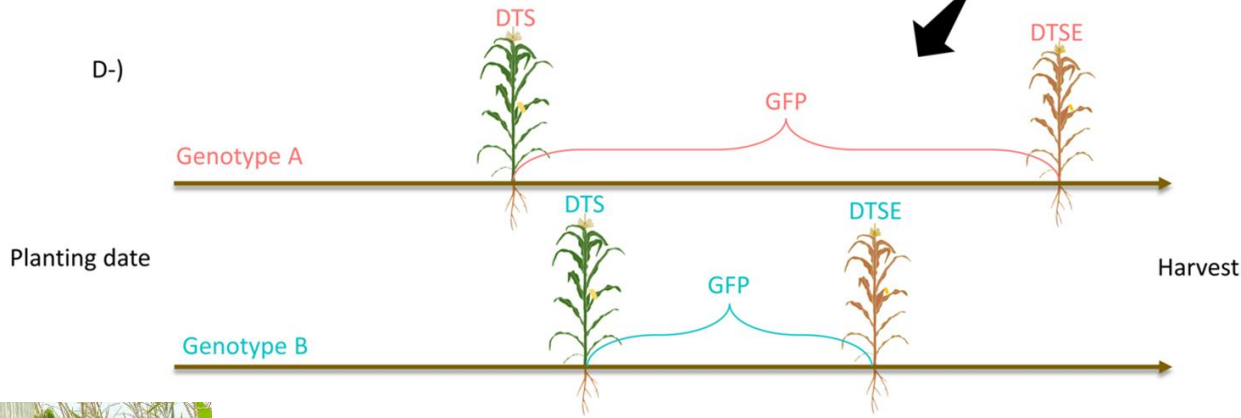
Dr. Alper Adak
Postdoc

Deep learning for senescence scoring 22,000 plots!
TX1, TX2, TX3, WI1, WI2, WI3, MO
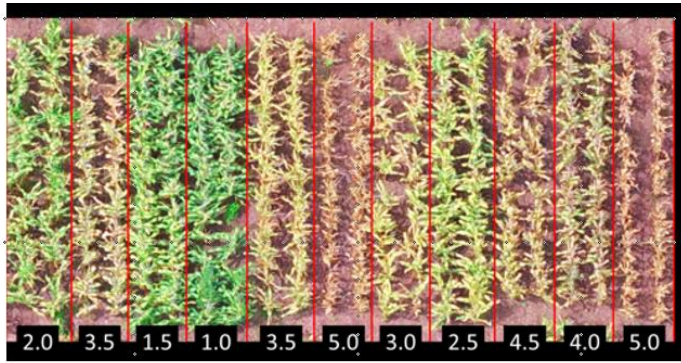
50 train - 50 test
3 classes

C) Confusion Matrix
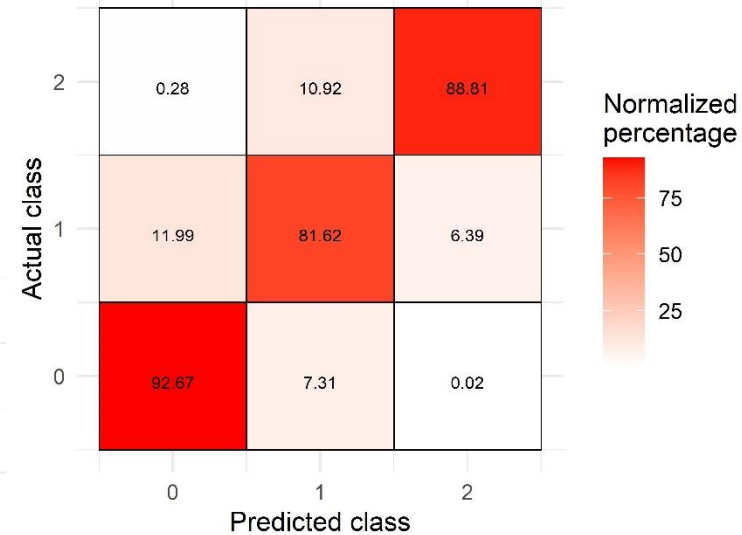
A) 80/20 Accuracy
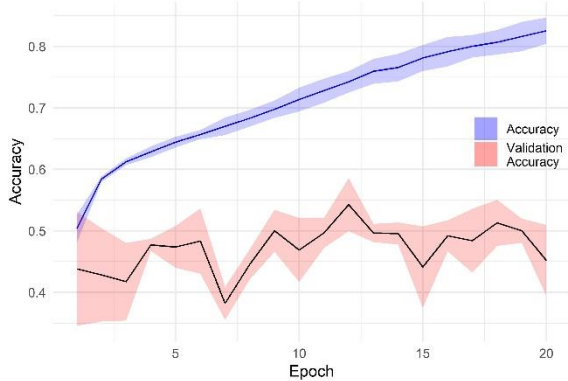
B) 80/20 Loss
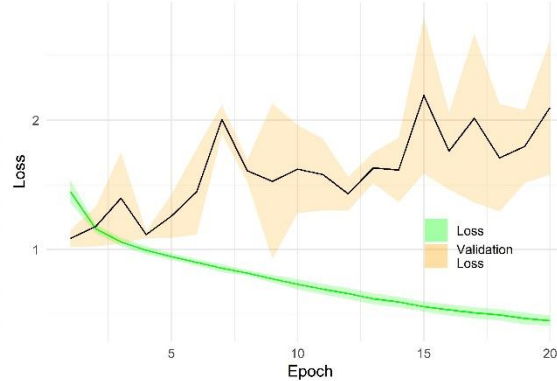
C) Confusion Matrix
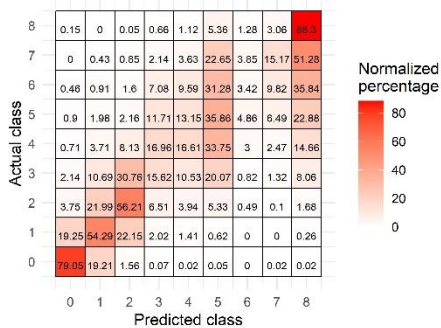
D) Evaluation Metrics
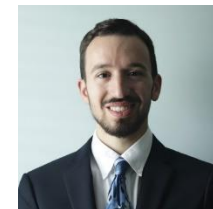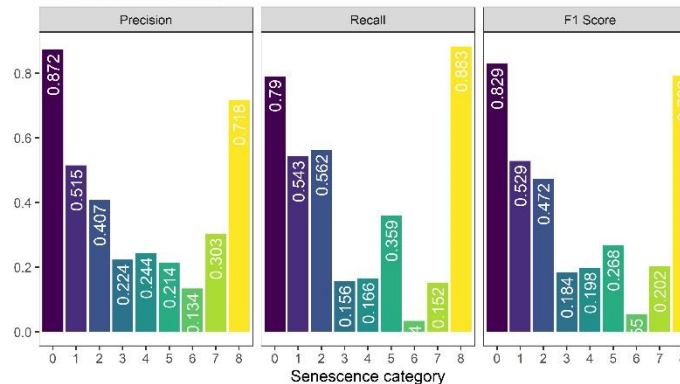
Aaron DeSalvio
PhD student

Cody Kettler
PhD student

Luke Gray
Undergraduate

Nick Shephard
PhD student

Correlation between VI's and Agronomics

Alper Adak
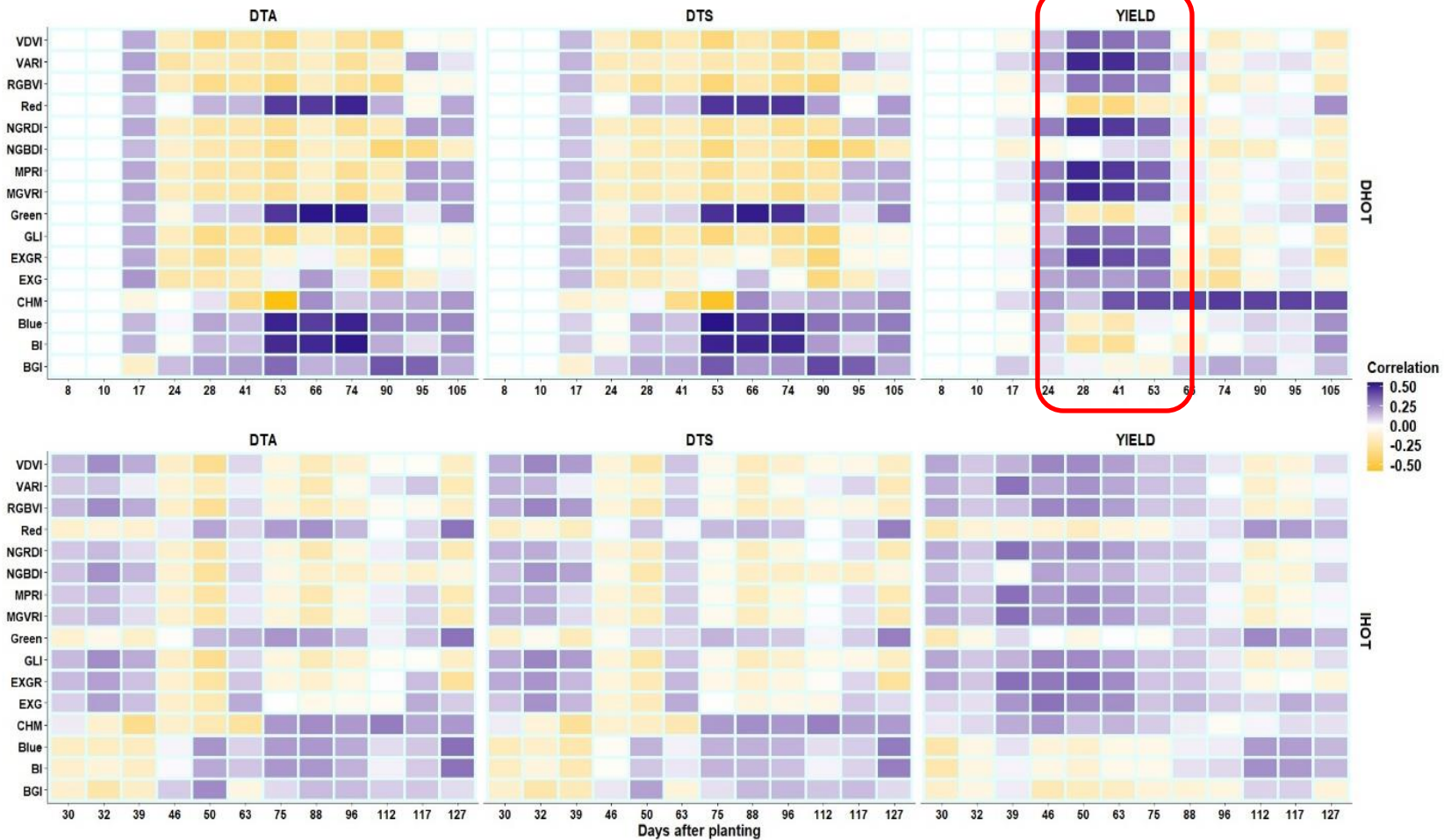
In late (stress) planting, the best predictors of yield were BEFORE flowering, possibly because this is where unselected variation is?

Alper Adak [1], Seth C Murray [1,*], Sofija Božinović [2], Regan Lindsey [1], Shakirah Nakasagga [1], Sumantra Chatterjee [1], Steven L. Anderson II [3] and Scott Wilde [1]

alperadak@tamu.edu for further questions

Basics of Near Infrared Reflectance Spectroscopy (NIRS)

Scan whole kernel corn

Grind and scan ground corn

www.udyone.com

Send for chemical analysis

Partial least squares (PLS) to calibrate between reflectance and wet-chemistry
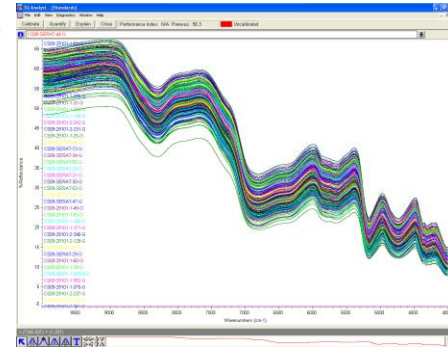
Smpl.	1	... n
Strch.		
Oil		
Prtn.		
Phos.		
Oleic.		

Find predictive equation using independent validation samples

Apply to breeding or genetics research

Crude Protein % Dry Basis
RMSEC: 0.437 Corr. Coeff.: 0.9678
RMSEP: 0.355
8 factors used

Crude Protein

P % Dry Basis
RMSEC: 0.0400 Corr. Coeff.: 0.7043
RMSEP: 0.0278
5 factors used
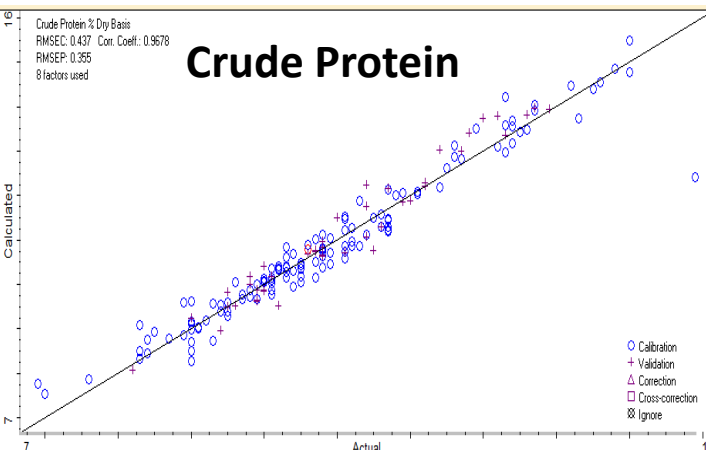
Phosphorus

	Crude Protein	Phosphphorus
σ^2_G	36 %	35 %
σ^2_E	17 %	21 %
σ^2_{Row}	4 %	5 %
σ^2_{erro}	43 %	38 %
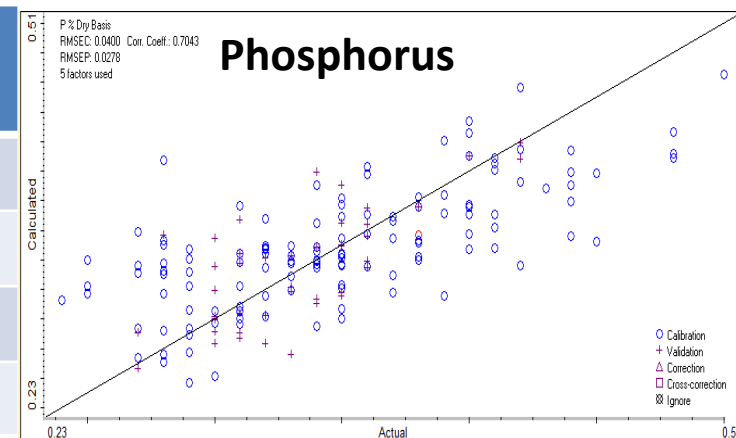
Phenomic selection
(Rincent et al. 2018 G3)

+

GWAS study data
(Barerro et al. 2014 PLOS One)

- **Used NIRS to predict complex traits in wheat and poplar**
- **PS was as accurate as genomic selection but cheaper**

- **346 Hybrids (x Tx714)**
- **4000 samples total**
  - **1700 Ground and 2300 Whole**
- **Dryland and irrigate**

G3
Genes | Genomes | Genetics

GENOMIC PREDICTION

Phenomic Selection Is a Low-Cost and High-Throughput Method Based on Indirect Predictions: Proof of Concept on Wheat and Poplar

Renaud Rincent,* Jean-Paul Charpentier,†‡ Patricia Fakre-Rampant,‡ Etienne Paux,* Jacques Le Gouis,* Catherine Bastien,† and Vincent Segura*†
*GDEC, INRA, UCA, 63000 Clermont-Ferrand, France, †BioForA, INRA, ONF, 45075 Orléans, France, ‡GenoBois analytical platform, INRA, 45075 Orléans, France, and §EPGV, INRA, CEA-IG/CNG, 91057 Évry, France
ORCID IDs: 0000-0003-0885-0969 (R.R.); 0000-0002-6029-0498 (J.-P.C.); 0000-0002-3094-7129 (E.P.); 0000-0001-5726-4902 (J.L.G.); 000-0002-9391-6637 (C..B.); 0000-0003-1860-2256 (V.S.)

ABSTRACT Genomic selection - the prediction of breeding values using DNA polymorphisms - is a disruptive method that has widely been adopted by animal and plant breeders to increase productivity. It was recently shown that other sources of molecular variations such as those resulting from transcripts or metabolites could be used to accurately predict complex traits. These endophenotypes have the advantage of capturing the expressed genotypes and consequently the complex regulatory networks that occur in the different layers between the genome and the phenotype. However, obtaining such omics data at very large scales, such as those typically experienced in breeding, remains challenging. As an alternative, we proposed using near-infrared spectroscopy (NIRS) as a high-throughput, low cost and non-destructive tool to indirectly capture endophenotypic variants and compute relationship matrices for predicting complex traits, and coined this new approach "phenomic selection" (PS). We tested PS on two species of economic interest (Triticum aestivum L. and Populus nigra L.) using NIRS on various tissues (grains, leaves, wood).

KEYWORDS
Poplar
Wheat
breeding
endophenotypes
Near InfraRed
Spectroscopy (NIRS)
Genomic
Prediction
GenPred
Shared Data

PLOS ONE

RESEARCH ARTICLE

Genome Wide Association Study for Drought, Aflatoxin Resistance, and Important Agronomic Traits of Maize Hybrids in the Sub-Tropics

Ivan D. Barrero Farfan[1], Gerald N. De La Fuente[1¤], Seth C. Murray*[1], Thomas Isakeit[2], Pei-Cheng Huang[2], Marilyn Warburton[3], Paul Williams[3], Gary L. Windham[3], Mike Kolomiets[2]

1 Department of Soil and Crop Sciences, Texas A&M University, College Station, Texas, United States of America, 2 Department of Plant Pathology, Texas A&M University, College Station, Texas, United States of America, 3 USDA ARS Corn Host Plant Resistance Research Unit, Mississippi State, Mississippi, United States of America

¤ Current address: Department of Agronomy, Iowa State University, Ames, Iowa, United States of America
* sethmurray@tamu.edu

Abstract

The primary maize (Zea mays L.) production areas are in temperate regions throughout the world and this is where most maize breeding is focused. Important but lower yielding maize growing regions such as the sub-tropics experience unique challenges, the greatest of which are drought stress and aflatoxin contamination. Here we used a diversity panel consisting of 346 maize inbred lines originating in temperate, sub-tropical and tropical areas testcrossed to stiff-stalk line Tx714 to investigate these traits. Testcross hybrids were evaluated under irrigated and non-irrigated trials for yield, plant height, ear height, days to anthe-

NIRS results: whole kernel spectral yield train & predict with 2011 & 2012

Partial Least Squares Results:
Yield (NIRS vs. actual)
R^2 = 0.19 to 0.84
RMSEP < 26 bu/ac
Protein (vs. actual yield)
R^2 = 0.58
Protein + starch + oil
R^2 = 0.64

Functional Regression Results:
Yield (NIRS vs. actual)
R^2 = 0.13 to 0.75

Holly Lane, M.S.   **Dr. José Crossa**

Global PLSR
Model trained on 1573 samples
Tested on remaining 848 samples

r = 0.84

RMSEP = 163 g m^2
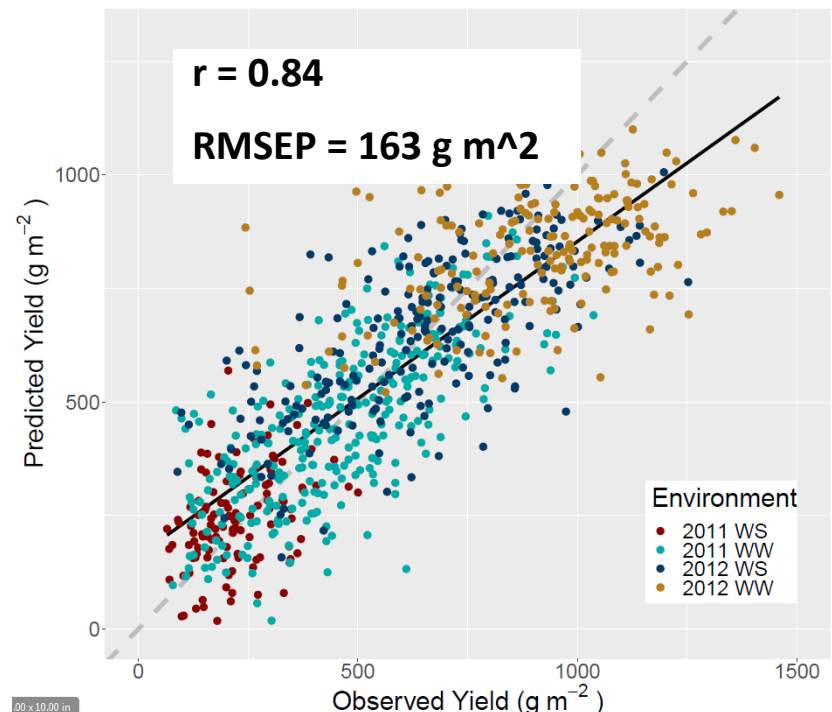
The Plant Phenome Journal   OPEN ACCESS

ORIGINAL RESEARCH | Open Access

Phenomic selection and prediction of maize grain yield from near-infrared reflectance spectroscopy of kernels

Holly M. Lane, Seth C. Murray, Osval A. Montesinos-López, Abelardo Montesinos-López, José Crossa, David K. Rooney, Ivan D. Barrero-Farfan, Gerald N. De La Fuente, Cristine L. S. Morgan

UAS temporal phenomic prediction ≈ genomic prediction

Fig. 3 shows the prediction accuracy results of yield belonging to the three models. GP represents the prediction accuracy of genomic prediction, TPP_Multi represents the prediction accuracy of phenomic prediction using the VIs derived from the multispectral images with low resolution, TPP_RGB represents the prediction accuracy of phenomic prediction using the VIs derived from the RGB images with high resolution. Four cross validation schemes were used: predicting tested genotypes in tested environments (CV1), predicting untested genotypes in tested environments (CV2), tested genotypes in untested environments (CV3), and untested genotypes in untested environments (CV4).
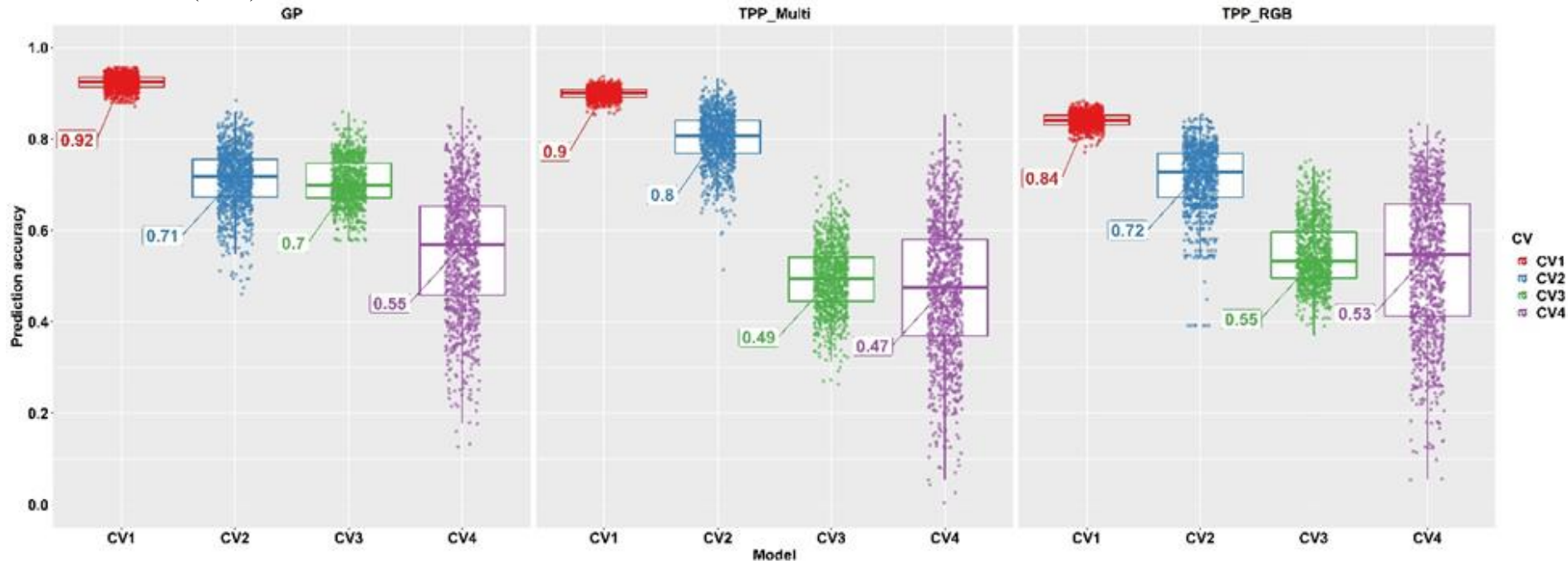
Alper Adak
PhD Student

Genomes to Fields population: 1188 plots, 280 hybrids, 2 TX environments

Only 158 hybrids with genomic data: 70% training, 30% testing

GP = 153,252 markers for genomic prediction

TPP_Multi = 1068 phenomic features (89 VI's x 12 time points)

TPP_RGB = 525 phenomic features (35 VI's x 15 time points)

Adak, A., Murray, S. C., & Anderson, S. L. (2021). Temporal phenomic predictions from unoccupied aerial systems can outperform genomic predictions. *bioRxiv*.

Approaches based on complexity for genetics

Genetic Penetrance

Additive X Genetic Background

Additive X Epistasis

Pure Additive

Here reality is messy

Phenomic Prediction / Selection

Genomic Prediction / Selection + Environment Modeling

Genomic Prediction / Selection

Mapping and Marker Assisted Selection

Biotech / Gene Editing
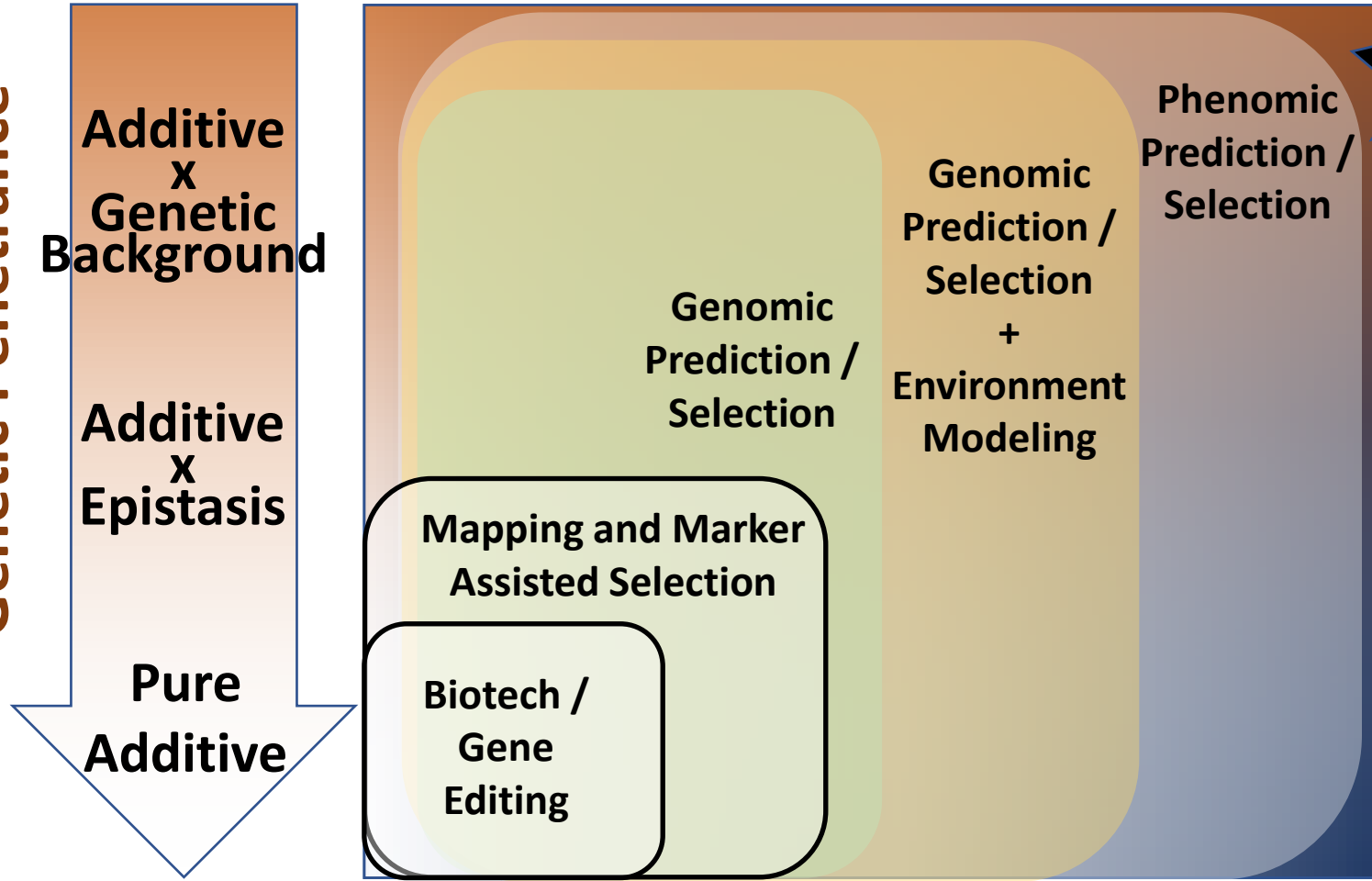
Here reality equals theory

G   G+E+error   G+E+GxE +error

Genetic Heritability
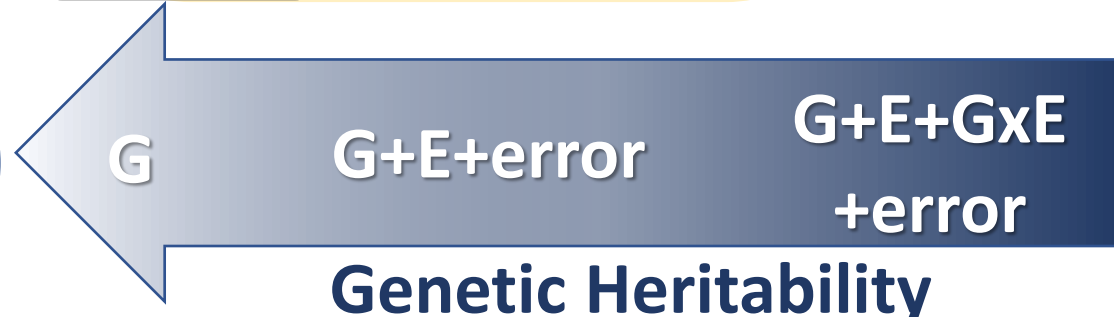
Spectral (NGRDI) time series shows genetic patterns

NGRDI temporal-series: Mean plant NGRDI index versus days after sowing (DAS) by genotype.

Dale Cope
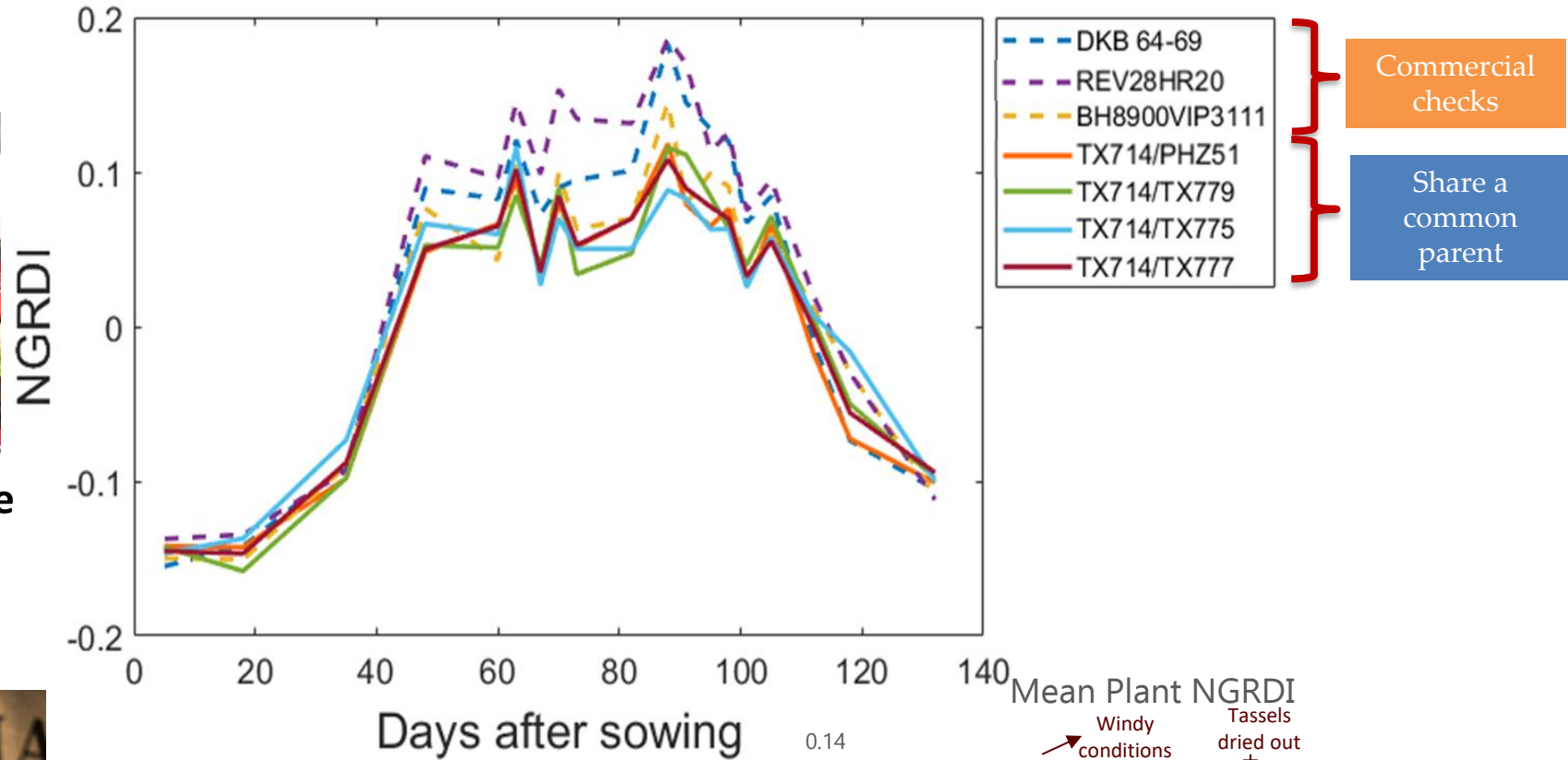
Nathália Cruzato PhD student

Other sources of NGRDI variation
Tassels
Tassels + wind
Tassels drying out
Weeds
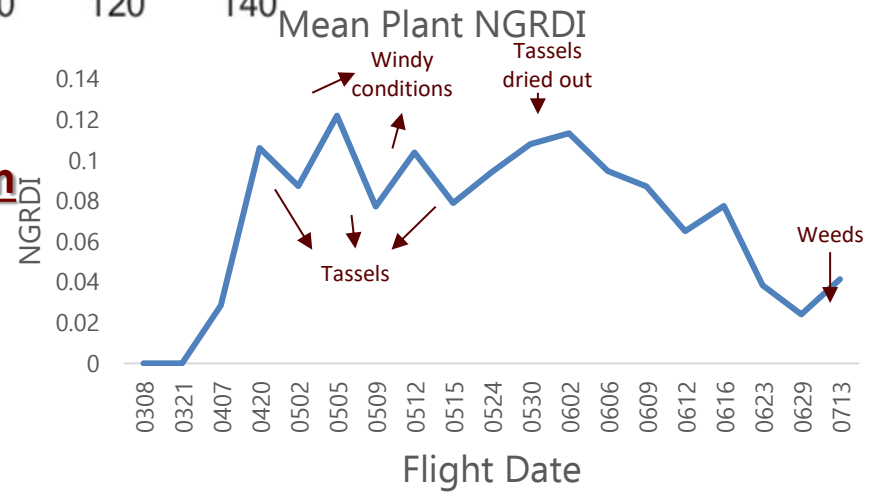
Humans recognize resemblances in parents and siblings appearances and expect similarities in other traits

Humans recognize resemblances but hard to describe the quantitative similarities

Specific resemblance features change over time but maintain resemblance

Appearance Phenotype

Genetics
Environment
G x E
Error
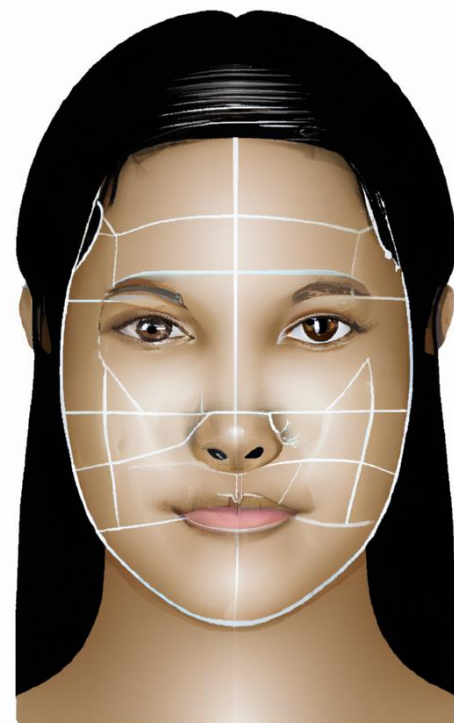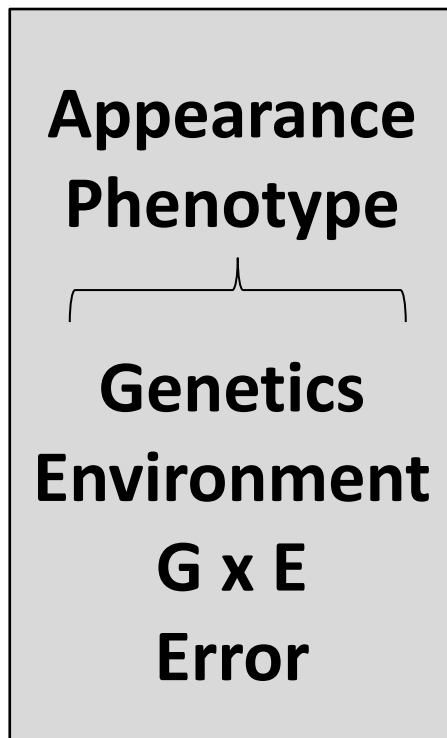
Cartoon of mechanism for why and how phenomics approaches could be more predictive than genomics

Genomics

DNA

Env.

GxE

Management

microbes

stochasticity

Cumulative and Temporal

Phenomics

Metabolism of complex carbohydrates

Biodegradation of xenobiotics

Metabolism of complex lipids

Nucleotide metabolism

Carbohydrate metabolism

Metabolism of other amino acids

Lipid metabolism

Amino acid metabolism

Energy metabolism

Metabolism of cofactors and vitamins

Biosynthesis of secondary metabolites

Lengauer and Hartmann 2007 https://doi.org/10.1016/B0-08-045044-X/00088-2

Müller-Linow, Hilgetag, and Hütt. 2008 https://doi.org/10.1371/journal.pcbi.1000190

Easy to measure phene - 1

Easy to measure phene - 2

Easy to measure phene - 3

Easy to measure phene - 4

Easy to measure phene - 5

Easy to measure phene - 6

Easy to measure phene - 7

Easy to measure phene - 8

Easy to measure phene - 9

...

Easy to measure phene – N=xxxx

ML Models

Difficult phenotype we care about

The more and diverse phenomic features we use as predictors the better our predictions can be

Science has saturated the genome...

"...681,257 single-nucleotide polymorphism (SNP) markers..." – Romay et al. 2013

Research | Open Access | Published: 11 June 2013

Comprehensive genotyping of the USA national maize inbred seed bank

Maria C Romay, Mark J Millard, Jeffrey C Glaubitz, Jason A Peiffer, Kelly L Swarts, Terry M Casstevens, Robert J Elshire, Charlotte B Acharya, Sharon E Mitchell, Sherry A Flint-Garcia, Michael D McMullen, James B Holland, Edward S Buckler ✉ & Candice A Gardner ✉

Genome Biology **14**, Article number: R55 (2013) | Cite this article

40k Accesses | **341** Citations | **19** Altmetric | Metrics

"...Ames panel of ~1,500 inbred lines scored with 12.2 million single-nucleotide polymorphisms..."

Combining GWAS and TWAS to identify candidate causal genes for tocochromanol levels in maize grain ⊘

Di Wu, Xiaowei Li, Ryokei Tanaka, Joshua C Wood, Laura E Tibbs-Cortes, Maria Magallanes-Lundback, Nolan Bornowski, John P Hamilton, Brieanne Vaillancourt, Christine H Diepenbrock ... Show more
Author Notes

Genetics, Volume 221, Issue 4, August 2022, iyac091,
https://doi.org/10.1093/genetics/iyac091
Published: 06 June 2022   Article history ▾

Cost per Raw Megabase of DNA Sequence

https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data

http://www.gramene.org/newsletters/gramnews/julyaugust07.html
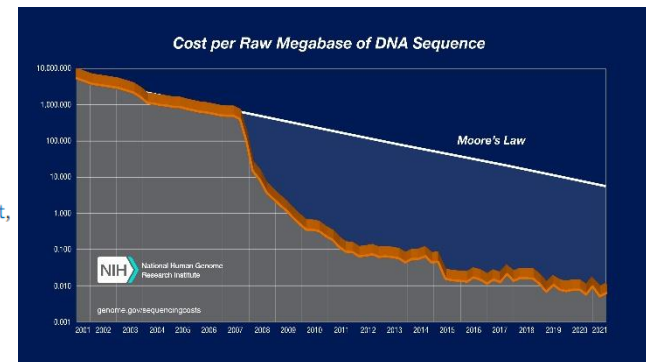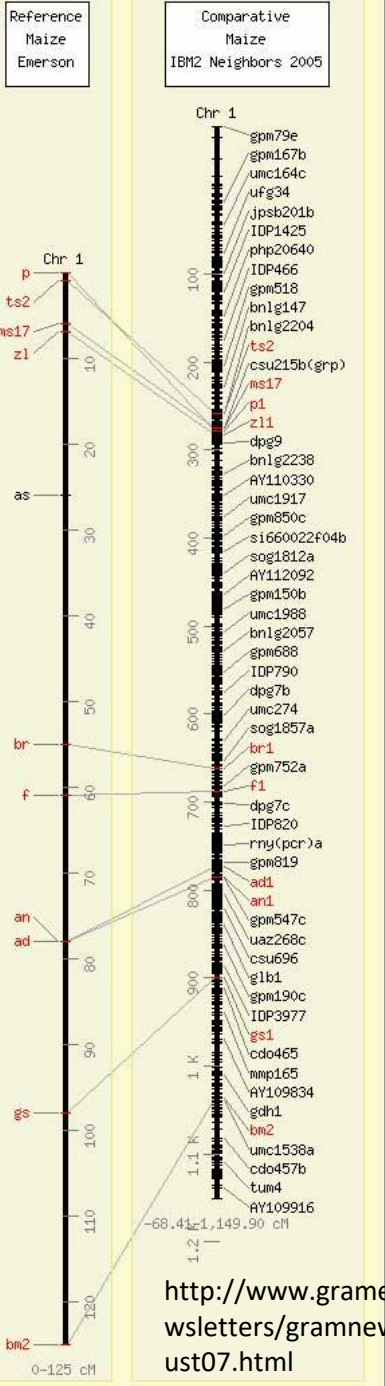
Can we saturate the phenome?

Maize genomes to fields (G2F) project

https://www.genomes2fields.org/   **2023 = 9th year!**

Plant morphology	Agronomic
Plant height Ear height	Stand count Lodging Days to silk Days to anthesis
Productivity	**Weather station**
Grain yield Test weight Grain moisture	Humidity Temperature Solar radiation Soil moisture Wind

180,000 Field Plots

2500+ corn varieties

162 unique environments

Some flew drones

USDA — United States Department of Agriculture — National Institute of Food and Agriculture

IOWA CORN™

TEXAS A&M AGRILIFE RESEARCH

TCB TEXAS CORN PRODUCERS BOARD

USDA ars

G + E + M+ GxE + GxExM

2022 Academic & Federal Institutions

Clemson University
Colorado State University
Cornell University
Iowa State University
Michigan State University
North Carolina State University

Ohio State University
Purdue University
Texas A&M University
University of Delaware
University of Georgia
University of Illinois

University of Minnesota
University of Missouri
University of Nebraska-Lincoln
University of Wisconsin
USDA-ARS

Georg-August-Universität Göttingen
(Not shown on map)

the Genomes to Fields initiative

★ 59

Genomes to Fields (G2F) Genotype by Environment Prediction Competition

Genomes to Field (G2F) : Maize Yield Prediction Competition 2022

Organized by: G2F_2022_GxE_challenge_host
Starts on: Nov 15, 2022 12:00:00 AM
Ends on: Jan 15, 2023 11:59:59 PM

Rank ⇕	Participant team ⇕	Mean_RMSE (↓) ⇕
1	CLAC	2.328863
2	igorkf	2.345147
3	phenomaize	2.374471
4	UCD_MegaLMM	2.387404

Year	Environment	Number of flights for plant height	RGB Vegetation Indices	Multispectral Vegetation Indices	Plant Height Elevation Models	File size: Orthomosaics and plant height point cloud files	File size: All files (including raw image and Agisoft Files)
2020	College Station, TX	12	Yes	na	Yes	566 GB	1.08 TB
2020	Arlington, WI	na	Yes	na	na	56.7 GB	481 GB
2020	Hancock, WI	na	Yes	na	na	38.2 GB	231 GB
2020	Madison, WI	na	Yes	na	na	63 GB	433 GB
2020	Missouri	87	Yes	na	Yes	707 GB	1.74 GB
2020	Michigan	12	Yes	Yes	Yes	73.9 GB	404 GB
2020	New York	na	Yes	Yes	na	39.4 GB	39.4 GB
2020	Minnesota	11	Yes	na	Yes	94 GB	451 GB
2020	Delaware	13	Yes	Yes	Yes	88.2 GB	465 GB
2020	Iowa	na	Yes	na	na	583 GB	722 GB
2020	Colorado	na	TBD	na	na	93.3 GB	564 GB
2021	College Station, TX	na	Yes	Yes	na	47.9 GB	na
2021	Hancock, WI	na	Yes	na	na	110 GB	~1 TB +
2021	Madison, WI	na	Yes	na	na	139 GB	~1 TB +
2021	Arlington, WI	na	Yes	na	na	80.6 GB	~1 TB +
2021	Minnesota	10	Yes	na	Yes	39.5	307 GB
2021	Nebraska	8	Yes	na	Yes	18.1 GB	41.2
2021	Iowa	na	Yes	na	na	92 GB	686 GB

Temporal plant height is heritable across locations but requires advanced statistical methods to integrate

Tested environments used in Genome to Field project (G2F)

Population structure of Hybrid population

Precision of temporal plant height data

CERCA

Circular Economy that Reimagines Corn Agriculture

CERCA Crop Science Team

PI Dr. Edward Buckler
USDA-ARS / Cornell

Springer

de Leon, Kaeppler

Basso

Buckler, Pineros
Costich, Gore, Hau,
Raboy, Romay

Lippman

★ Emmett, Scott

Roston

Castellano,
Hufford

★ Ainsworth

Guan

★ Flint-Garcia,
Washburn

Holland
Rellan-Alvarez

Busch,
Michael,
Noel

★ Chen

Murray

Messina

★ **USDA-ARS**

**University
Partners**

Drone analysis
Grow outs
Phosphorus phenotyping
Perennial germplasm
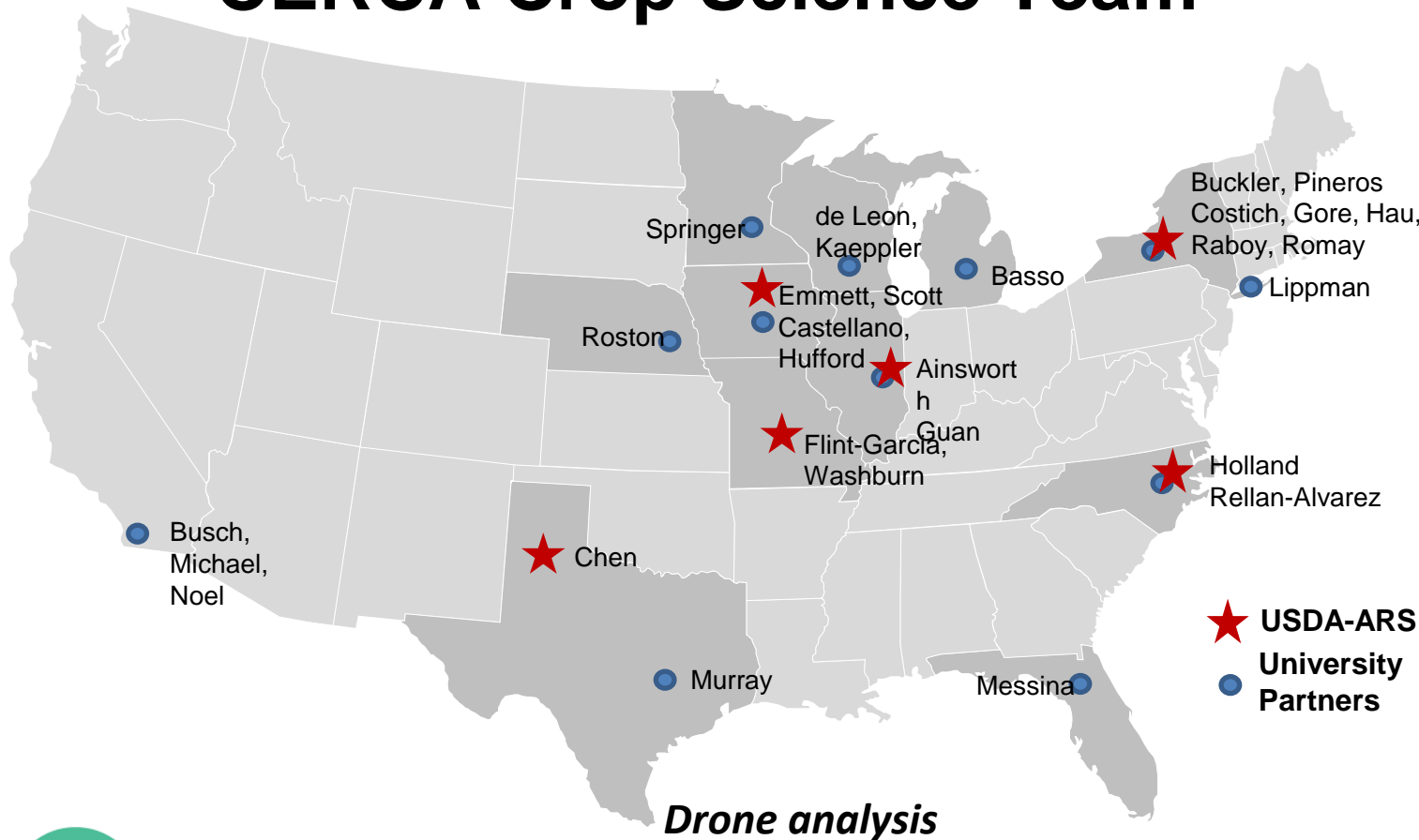
**FOUNDATION FOR
FOOD & AGRICULTURE
RESEARCH**

CERCA

Circular Economy that Reimagines Corn Agriculture
FFAR (50%), Industry (50%), USDA-ARS

Candidate Traits

Nutrient Recycling
- Grain N diver.
- Grain P diver.
- N remobilization
- Decomposition

Extended Season
- Germination
- Cold Tolerance
- Cold Establish.

Modeling

Modeling
(Plant, Soil, Farm, Landscape)

Testing at Agro-ecosystem Scale

Trait Discovery

Germplasm Screens

Evolutionary Mining

Profiling

Traits

Prioritize

Trait Development

Transgenics/Editing

Breeding

Physiology

FS2: Reactome knowledge graph to phenomes of major staple crops for food system threat detection
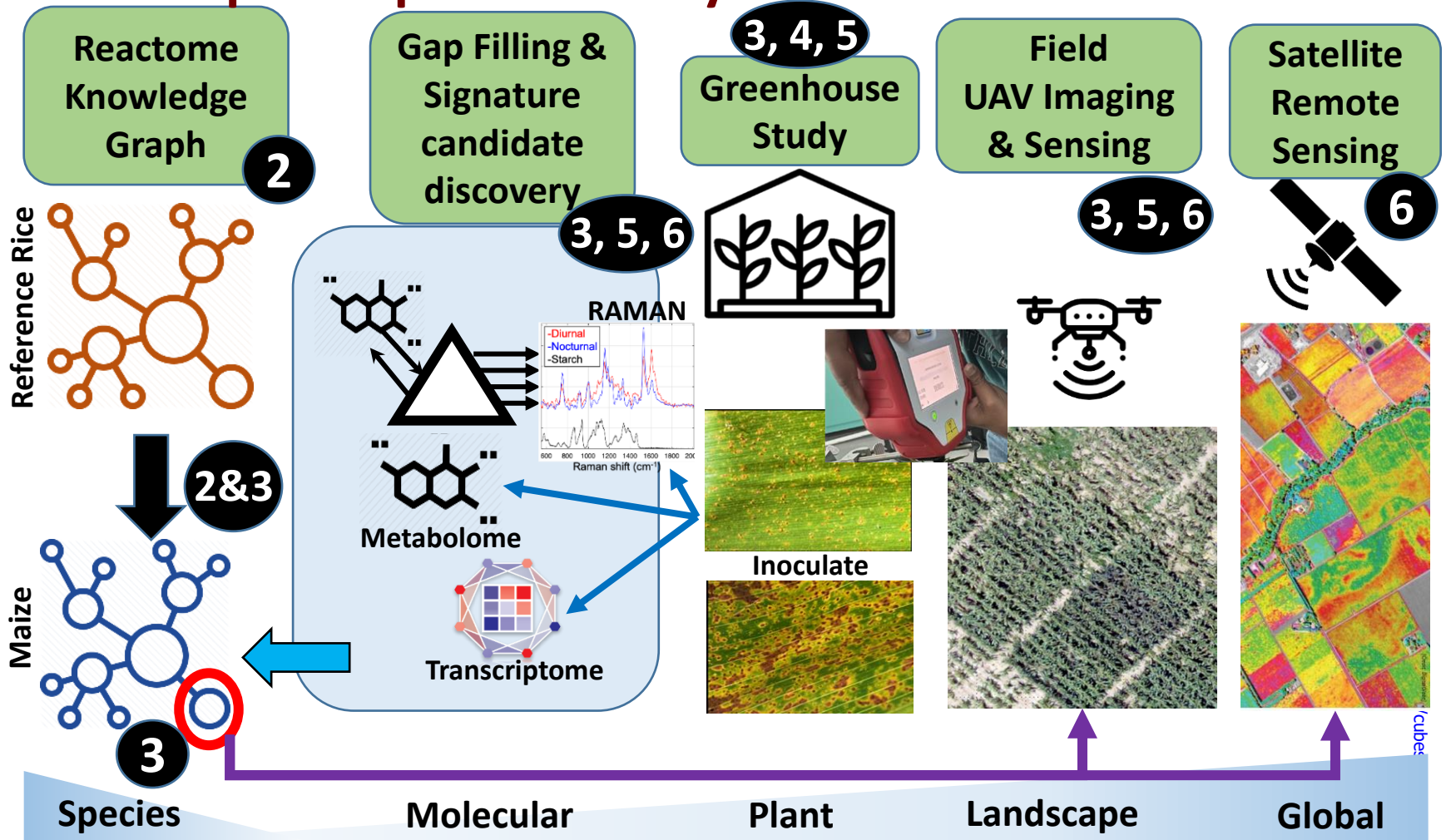
2023 greenhouse plantings

- Planted April 25 in College Station, TX
- ~27 hybrids, 10 inbreds,
- 2 treatments (inoculated vs. uninoculated) x 4 reps,
  (rep = 3 seeds / 1 plant per pot)

https://precisiongreenhouse.tamu.edu/
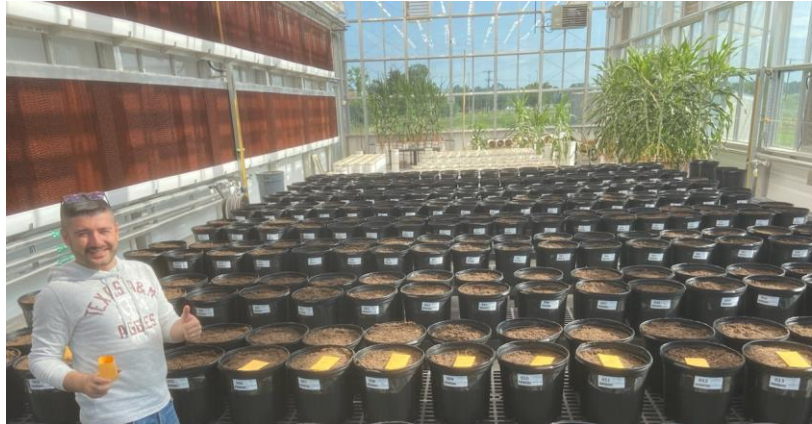
Raman spectra of compounds nominated from reactome

Pure compounds possible to measure *in planta* if sufficient quantities present

Naringenin chalcone Aapigenin Dihydrokaemferol Serotonin

Triptamine (powder) Isopropyl alcohol triptamine (in iso Raman spectrum of C_3H_7OH subtracted)

Limonene Naringenin

Dicamba Methanol Kaemferol (in methanol, Raman spectrum of CH_3OH subtracted)

Methyl jasmonate Ethanol Jasmonate (in ethanol, Raman spectrum of C_2H_5OH subtracted)

Salicylate Methyl salicylate

Corn leaf

Raman frequency (cm^{-1})

Aart Verhoef

PhotosynQ traits
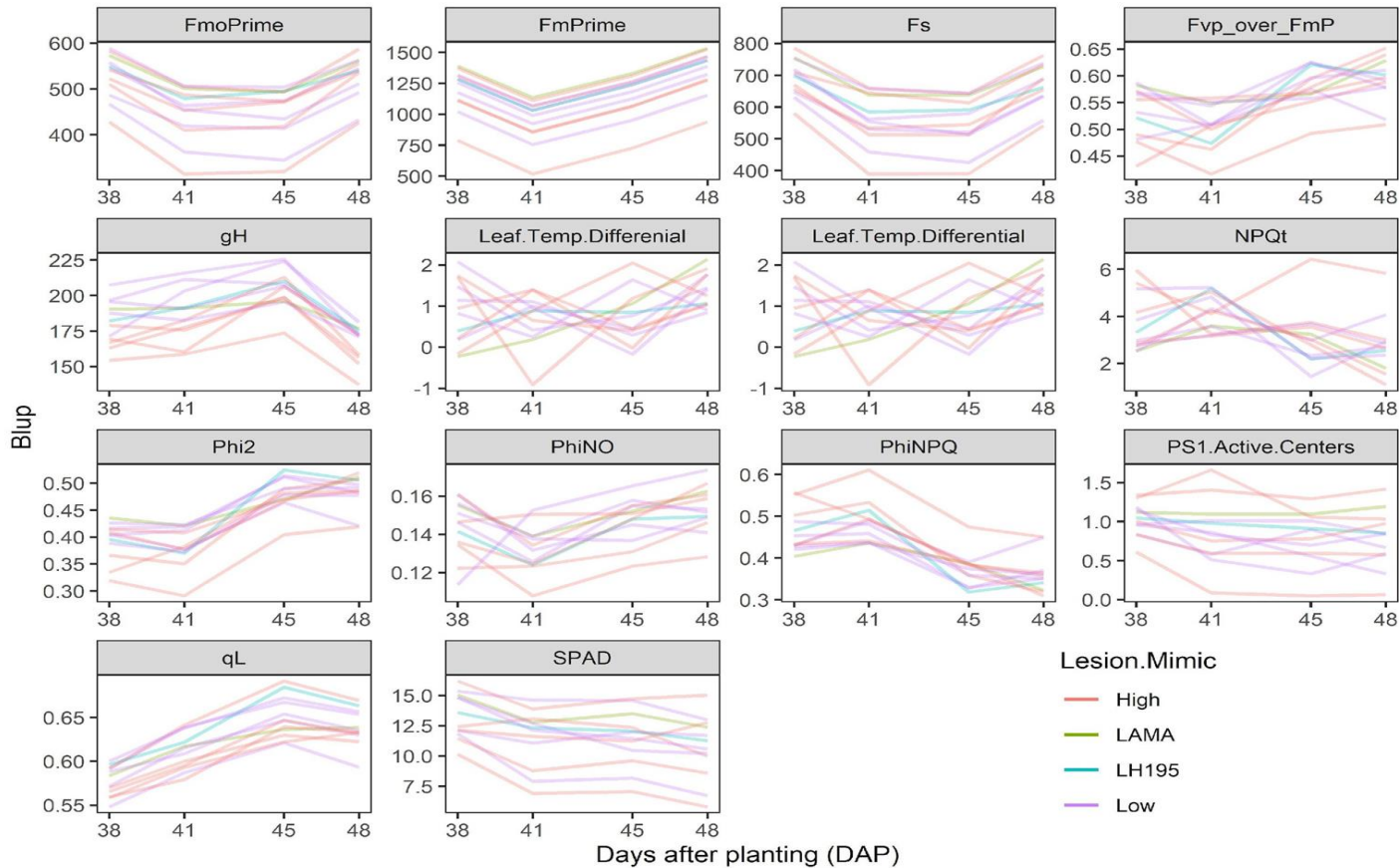
FMOPrime: Minimum Fluorescence
FmPrime: Maximum Fluorescence
Fs: Chlorophyll fluorescence
Fvp_over_FmP: Maximum quantum efficiency
gH+: Proton conductivity
Leaf temperature differential
PhiNPQ: Non photochemical quenching
Phi2: Quantum yield
PhiNO: Non-regulatory Energy Dissipation
PSIactive centers
qL: PSII redox state (Lake model)
SPAD: Relative chlorophyll content

Dimensionality to increase phenome data
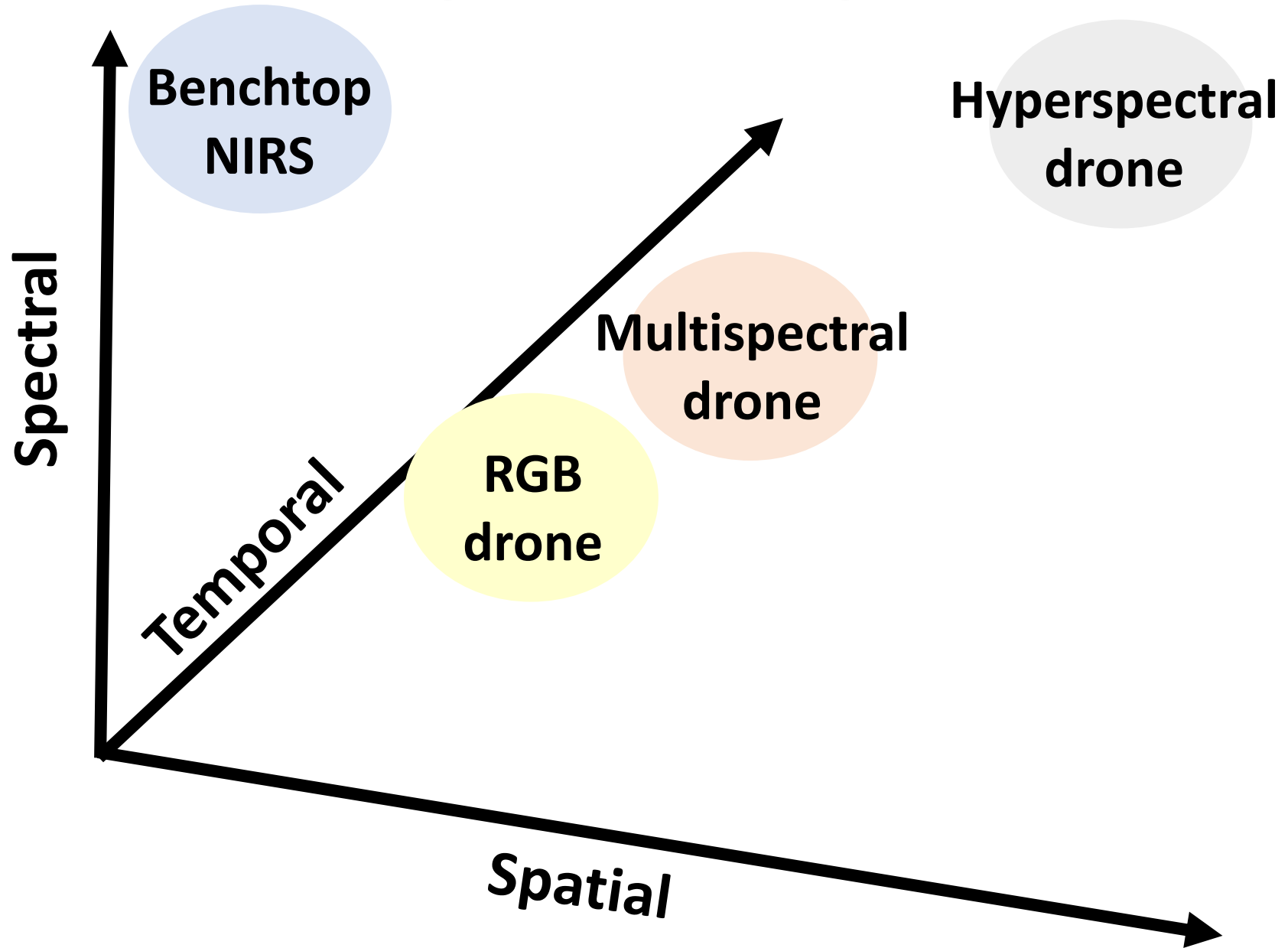
A) Repeated UAS flights and orthomosaic processing

Flight 1
Flight 2
Flight 3
Flight 4
Flight 5
Flight 6
Flight 7

B) Single-plant shapefile generation

C) Single-plant temporal image extraction

Genotype A – Plant 1

Flight1.tif   Flight2.tif   Flight3.tif   Flight4.tif   Flight5.tif

Genotype A – Plant 2

Flight1.tif   Flight2.tif   Flight3.tif   Flight4.tif   Flight5.tif
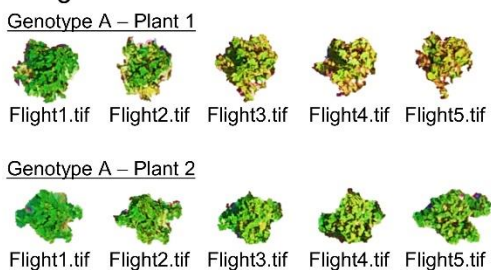
D) Temporal senescence scoring

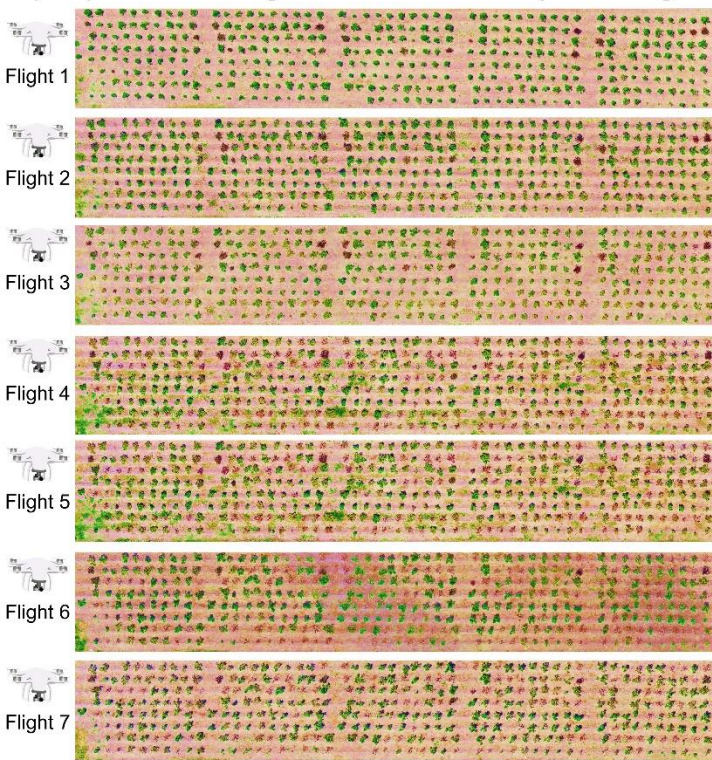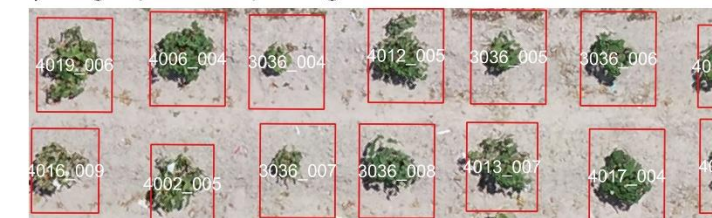| ID	Flight 1	Flight 2	Flight 3	Flight 4
GenoA-Plant1	0	1	2	3
GenoA-Plant2	0	2	2	2
GenoB-Plant1	0	1	1	2
GenoB-Plant2	0	2	3	4

E) Partition data via 80/20 train/test split

train_x

train_y   [0 4 2 0 2 1 5 3]

test_x

test_y   [0 1]
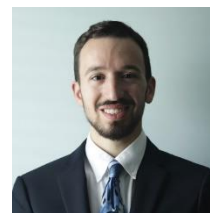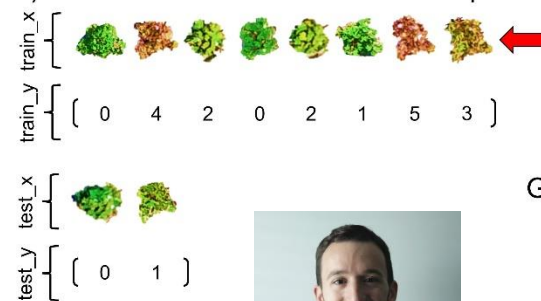
F) Convolutional neural network (CNN)

Input
(163 x 163 x 3)

Conv. layer (3 x 3 kernel, 32 filters)
Max pooling layer (2 x 2 pool)
Conv. layer (3 x 3 kernel, 64 filters)
Max pooling layer (2 x 2 pool)
Conv. layer (3 x 3 kernel, 128 filters)
Max pooling layer (2 x 2 pool)

Flattening layer

Dense layer 1 – 128 units, ReLU activation

Dense layer 2 – 6 units, softmax activation

Senescence score categories

0   1   2   3   4   5

0%   2%   4%   82%   10%   2%

Percent confidence of estimate

CNN senescence score estimate: 3

G) Apply CNN to unseen 20% of data

$$Classification\ accuracy = \frac{True\ positives}{Total\ images\ in\ test\ data\ set}$$

**Aaron DeSalvio
PhD student**

What does the future look like in plant breeding?

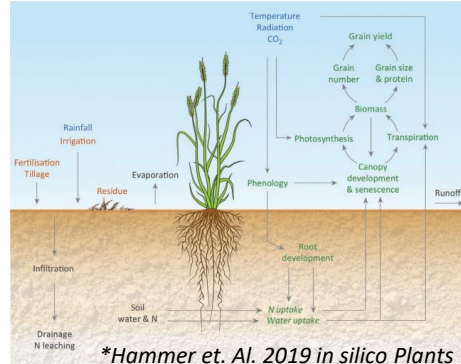|	Dependent "Traits" of interest			Physiological Measures			Temporal Phenomic "Trait" Measures			Segregating Remote Sensing Measures			Genomic Measures			
	Yield	Nutrition	Ecosystem	Flower	Leaf angle	Leaf #	Hgt 30 day	Hgt 60 day	Hgt 90 day	Phene X	Phene Y	Phene Z	SNP 1	SNP 2	SNP 3	SNP 4
Hybrid 1	100	9	3	70	35	16	0.02	0.30	0.61	0.542	0.179	0.449	AA	CC	CT	AG
Hybrid 2	150	10	9	65	32	15	0.04	0.30	0.62	0.068	0.505	0.744	AA	GG	TT	AA
Hybrid 3	125	2	9	68	30	17	0.05	0.33	0.63	0.523	0.949	0.271	AC	CC	CC	AA
Hybrid 4	175	3	2	66	37	16	0.07	0.40	0.70	0.823	0.814	0.537	CC	GG	TT	AA
Hybrid 5	180	7	3	73	40	16	0.10	0.42	0.72	0.637	0.815	0.703	CC	CG	TT	GG
Hybrid 6	75	6	10	67	32	13	0.11	0.54	0.75	0.874	0.368	0.785	CC	GG	TT	AA
Hybrid 7	90	3	10	66	33	19	0.12	0.55	0.81	0.984	0.406	0.867	CC	CC	TT	GG
Hybrid 8	170	5	6	65	32	19	0.16	0.56	0.84	0.491	0.418	0.176	AA	GG	TT	GG
Hybrid 9	160	10	5	70	31	17	0.24	0.58	0.85	0.693	0.287	0.013	CC	CC	TT	GG
Hybrid 10	155	9	7	72	39	15	0.26	0.59	0.99	0.130	0.624	0.566	AA	GG	CC	GG
	Multiple Replicates			*Multiple Replicates*			*Multiple Replicates*			*Multiple Replicates*			na	na	na	na
	Multiple Environments			*Multiple Environments*			*Multiple Environments*			*Multiple Environments*			na	na	na	na
Heritability	0.4	0.2	0.3	0.6	0.3	0.8	0.2	0.6	0.8	0.4	0.01	0.7	1	1	0.98	1

Statistical analysis (e.g. Functional Regression)

Growth models (w/ weather data)

Relationship matrices

Predict best untested hybrids for untested environments

Hammer et. Al. 2019 in silico Plants

Figure 3. Crop model schematic showing the key processes involved in crop growth and development and their interactions with the crop system (Chenu et al. 2017: with permission from Elsevier).

	Hybrid 1	Hybrid 2	Hybrid 3	Hybrid 4	Hybrid 5	Hybrid 6	Hybrid 7	Hybrid 8	Hybrid 9	Hybrid 10
Hybrid 1	~1	0.906	0.331	0.02	0.864	0.373	0.715	0.416	0.248	0.836
Hybrid 2	0.906	~1	0.276	0.323	0.008	0.608	0.364	0.941	0.106	0.009
Hybrid 3	0.331	0.276	~1	0.24	0.163	0.339	0.47	0.172	0.315	0.832
Hybrid 4	0.02	0.323	0.24	~1	0.532	0.647	0.423	0.406	0.73	0.348
Hybrid 5	0.864	0.008	0.163	0.532	~1	0.173	0.363	0.715	0.334	0.397
Hybrid 6	0.373	0.608	0.339	0.647	0.173	~1	0.991	0.127	0.409	0.774
Hybrid 7	0.715	0.364	0.47	0.423	0.363	0.991	~1	0.231	0.219	0.008
Hybrid 8	0.416	0.941	0.172	0.406	0.715	0.127	0.231	~1	0.995	0.13
Hybrid 9	0.248	0.106	0.315	0.73	0.334	0.409	0.219	0.995	~1	0.613
Hybrid 10	0.836	0.009	0.832	0.348	0.397	0.774	0.008	0.13	0.613	~1

Multiple groups cover UAS phenotyping

AG2PI — Agricultural Genome to Phenome Initiative

S1069: Research and Extension for Unmanned Aircraft Systems (UAS) Applications in U.S. Agriculture and Natural Resources
(Multistate Research Project)   Status: Active

NC1212: Exploring the Plant Phenome in Controlled and Field Environments
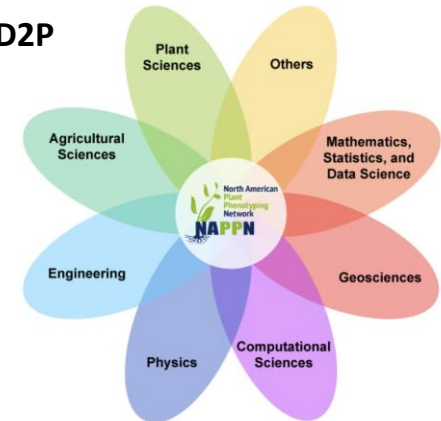(Multistate Research Project)   Status: Active

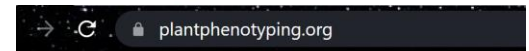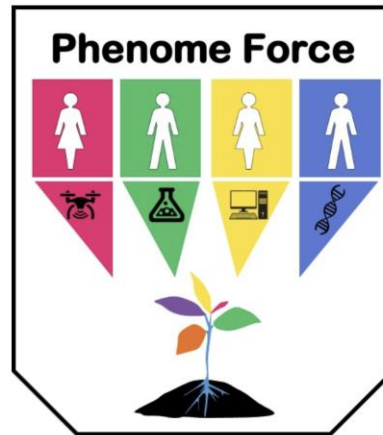## AI IN AGRICULTURE: INNOVATION AND DISCOVERY TO EQUITABLY MEET PRODUCER NEEDS AND PERCEPTIONS

WHEAT CAP — Coordinated Agricultural Project

the Genomes to Fields initiative

ODD-PIGG   Livestock?

D2P

Conference 12539
SPIE. DEFENSE+COMMERCIAL SENSING — Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping VIII

The Plant Phenome Journal — 2021 | VOLUME 4

FIELDimageR

Phenome Force

BREEDBASE

NAPPN — North American Plant Phenotyping Network

plantphenotyping.org

CSSA — Crop Science Society of America

ASA — American Society of Agronomy

ASABE — Engineering a Sustainable Future

Acknowledgements to what was presented

Murray group
Aaron DeSalvio
Dr. Alper Adak
Dr. Steven Anderson
Nick Shepard
Cody Kettler
Nathalia Cruzato*
Holly Lane
Dr. Scott Wilde
Shakirah Nakasagga
Colby Bass
Dr. Sumantra Chaterjee

Many past students and student workers!
*Supported in-part by a Monsanto future leaders fellowship

TAMUS UASAG Team
Sorin Popescu
Lonesome Malambo
Dale Cope
J. Alex Thomasson
Yeyin Shi
William Rooney
N. Ace Pugh

DARPA project team
Aart Verhoef
Alper Adak
Dipankar Sen
Riva Salzman
Pankaj Jaiswal
Sushma Naithani

Others
G2F collaborators
Natalia deLeon
Sofija Božinovic
Matthias Katfuss
Raymond Wong
Patricia Klein
Tom Isakeit
Mike Kolomiets

And many collaborators across other projects not presented