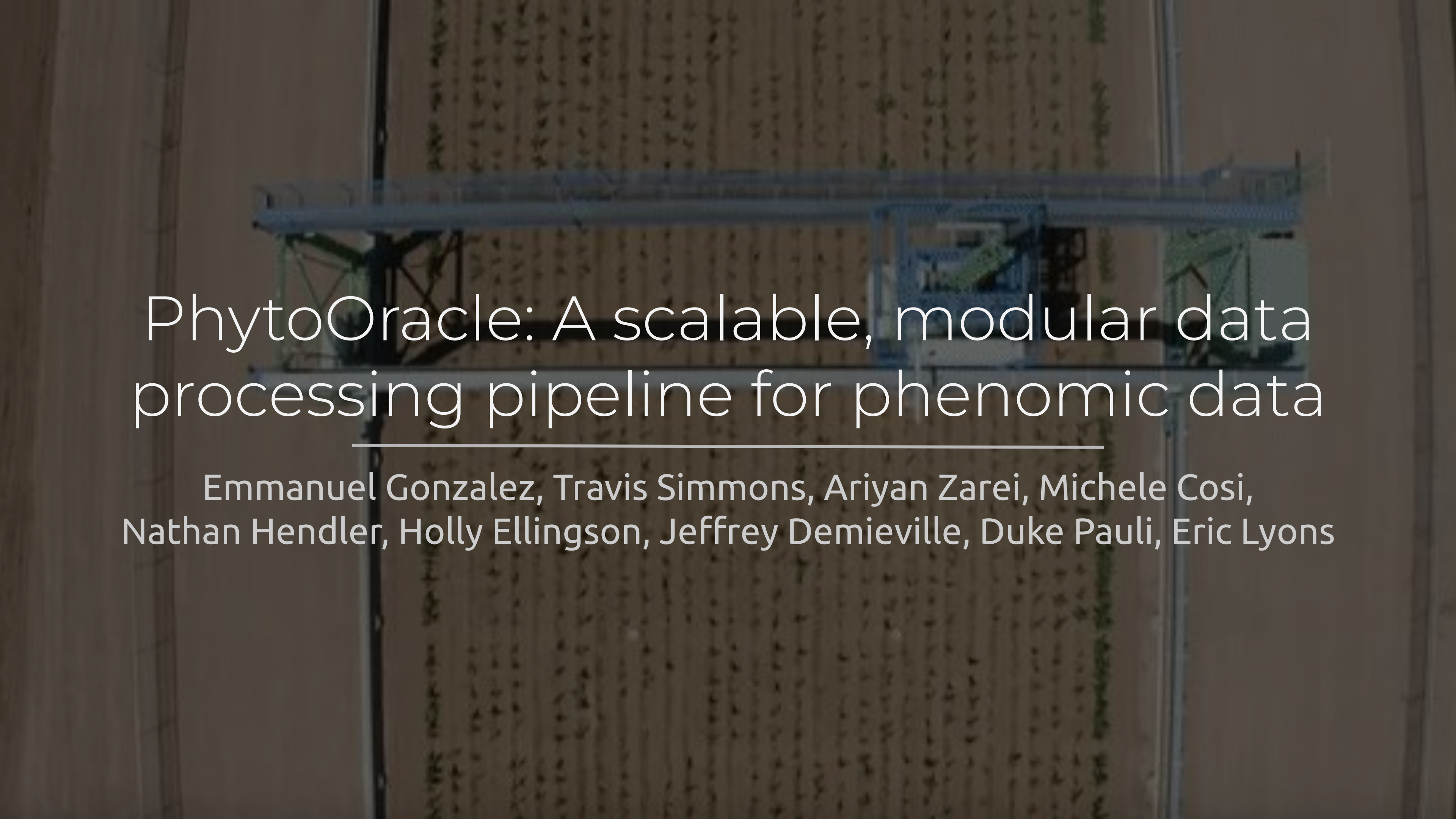


PhytoOracle: A scalable, modular data processing pipeline for phenomic data

Emmanuel Gonzalez, Travis Simmons, Ariyan Zarei, Michele Cosi,
Nathan Hendler, Holly Ellingson, Jeffrey Demieville, Duke Pauli, Eric Lyons

Domestication of teosinte and trait gathering

Teosinte

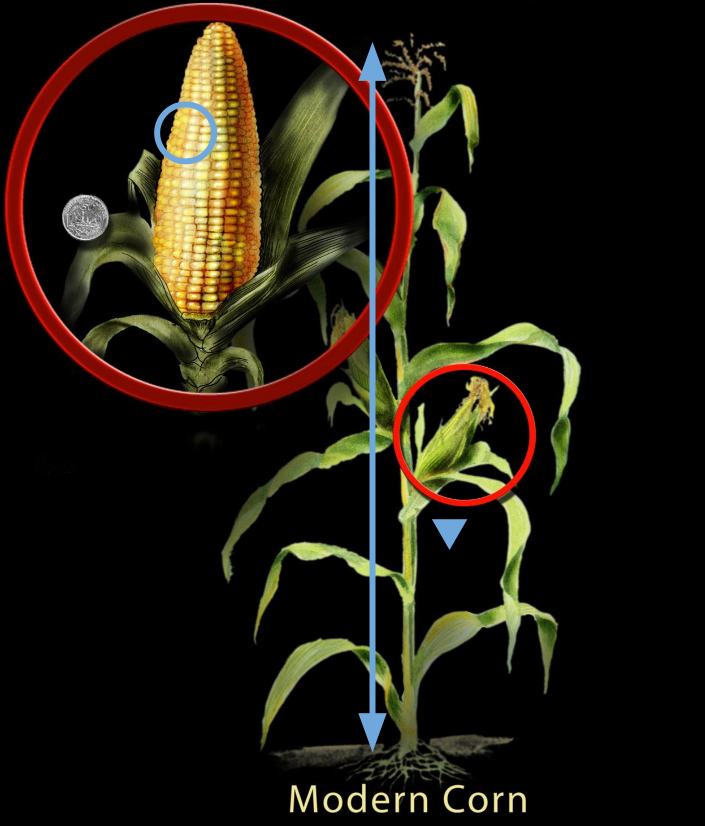
Nicolle Rager Fuller, National Science Foundation

Domestication of teosinte and trait gathering

Nicolle Rager Fuller, National Science Foundation

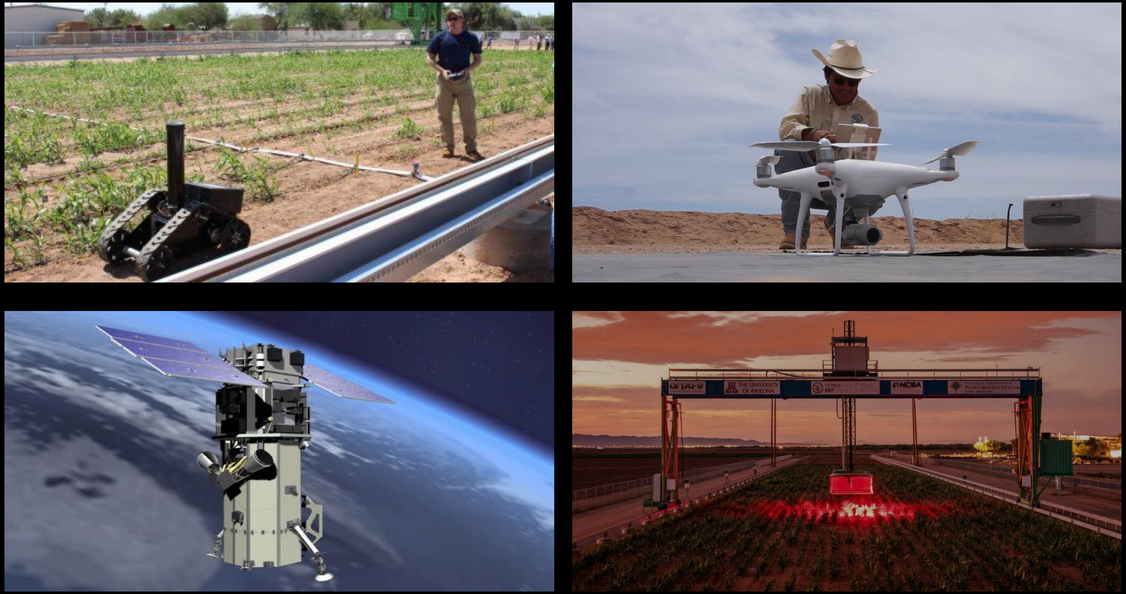
Traits of interest

- Morphological
  - Plant height
- Physiological
  - Photosynthetic efficiency, reflectance, canopy temperature
- Yield
  - Ear per plant, angle of ear, kernel number, kernel firmness
- Quality of yield
  - Kernel colour, sugar content



Nicolle Rager Fuller, National Science Foundation

Trait gathering has modernized

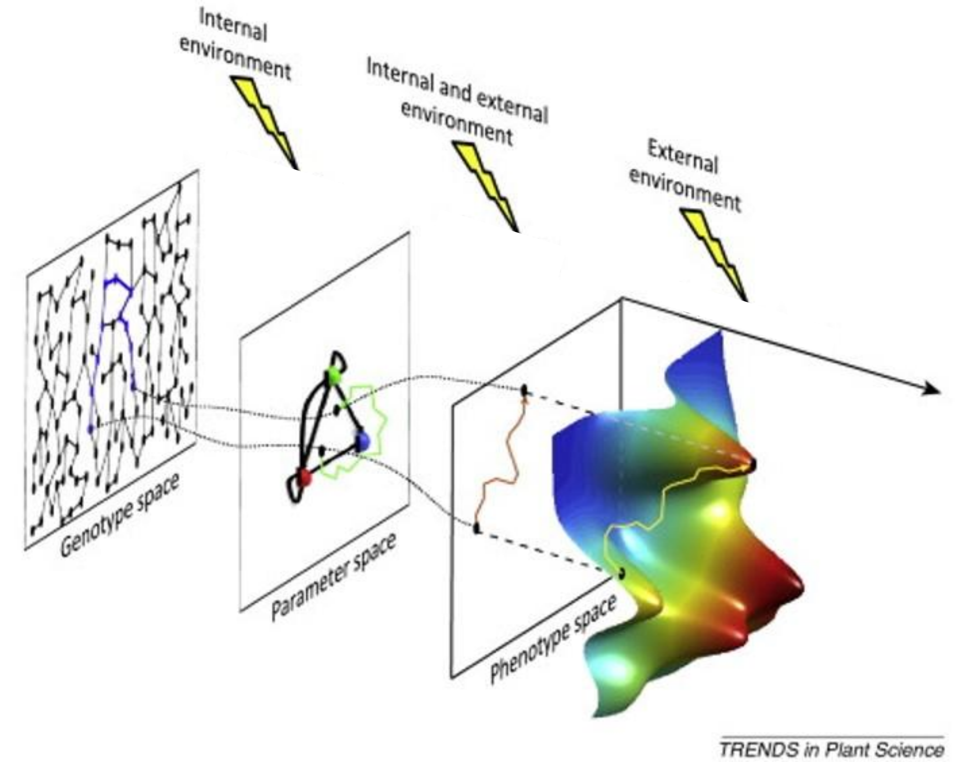


Nicolle Rager Fuller, National Science Foundation

Connecting genotype to phenotype

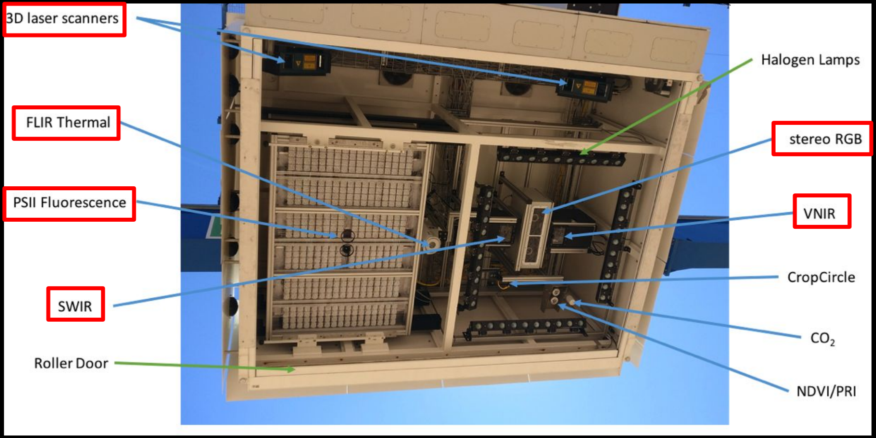
Gene combinations interact with the internal and environment to produce phenotypes

- Optimal to study phenotypes in the environment in which they are expressed



Adapted from Gutierrez and Maere, (2014) Trends in Plant Science 19:5

Obtaining and extracting high-quality phenomic data

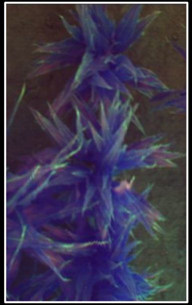


Data volume:

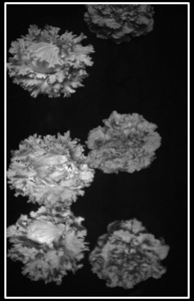
- Max: 10 TB/day
- Typical: 1.5 TB/day

3D

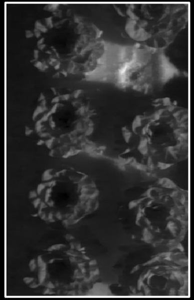
RGB



Hyperspectral



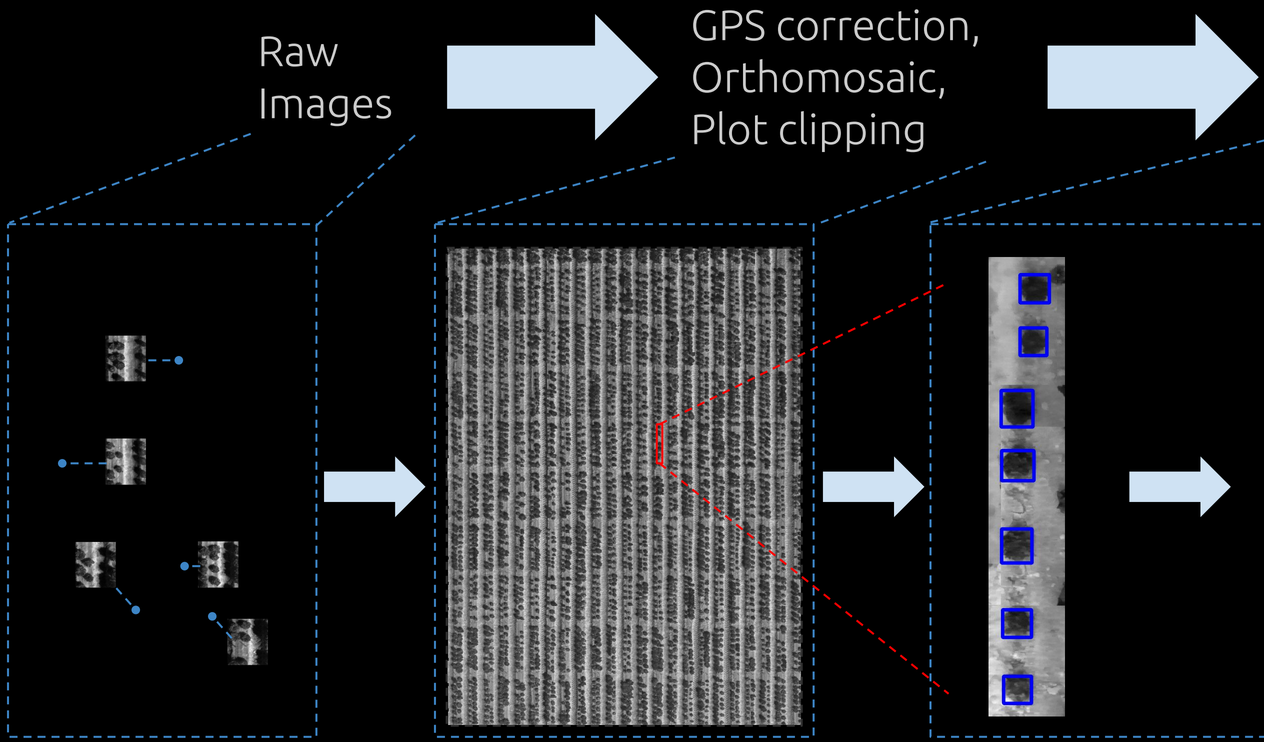
Fluorescence



Thermal

PhytoOracle

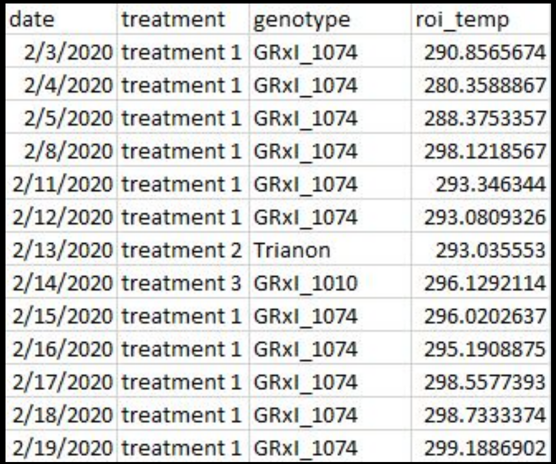
PhytoOracle pipeline workflow



Raw Images

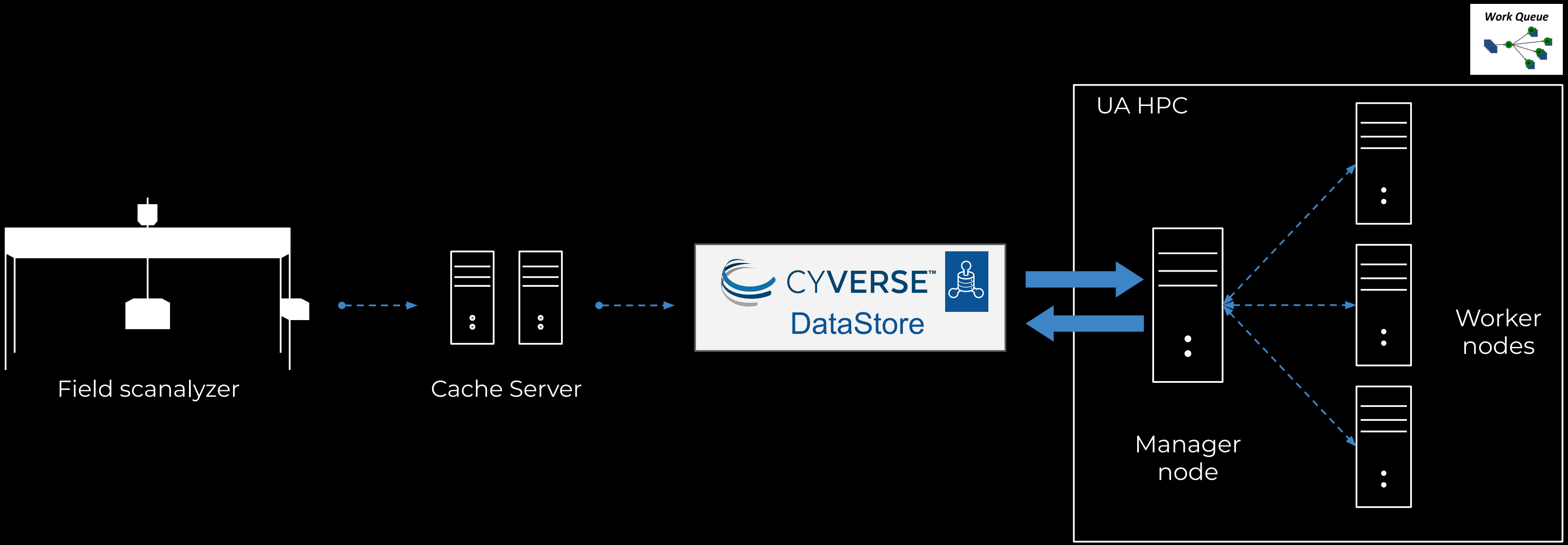
GPS correction, Orthomosaic, Plot clipping

Phenotype Quantification



date	treatment	genotype	roi_temp
2/3/2020	treatment 1	GRxI_1074	290.8565674
2/4/2020	treatment 1	GRxI_1074	280.3588867
2/5/2020	treatment 1	GRxI_1074	288.3753357
2/8/2020	treatment 1	GRxI_1074	298.1218567
2/11/2020	treatment 1	GRxI_1074	293.346344
2/12/2020	treatment 1	GRxI_1074	293.0809326
2/13/2020	treatment 2	Trianon	293.035553
2/14/2020	treatment 3	GRxI_1010	296.1292114
2/15/2020	treatment 1	GRxI_1074	296.0202637
2/16/2020	treatment 1	GRxI_1074	295.1908875
2/17/2020	treatment 1	GRxI_1074	298.5577393
2/18/2020	treatment 1	GRxI_1074	298.7333374
2/19/2020	treatment 1	GRxI_1074	299.1886902

Data flow and distributive framework



Scalability exponentially reduces data processing times

How much time would it take to process* a single season worth of RGB data (50TB) on a 4-core, regular lab computer?

55 years!

* From raw data to a quantifiable phenotype

How much time would it take PhytoOracle to process* a single season worth of RGB data (50TB)?

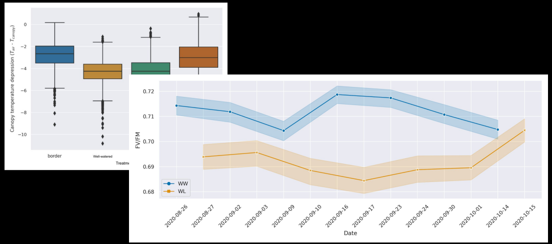
Only 6 days!

---

* From raw data to a quantifiable phenotype

Today's tutorial

- Running sections of PhytoOracle pipelines to extract thermal and reflectance data
- Visualizing intermediate and phenotypic trait outputs on a Jupyter Notebook

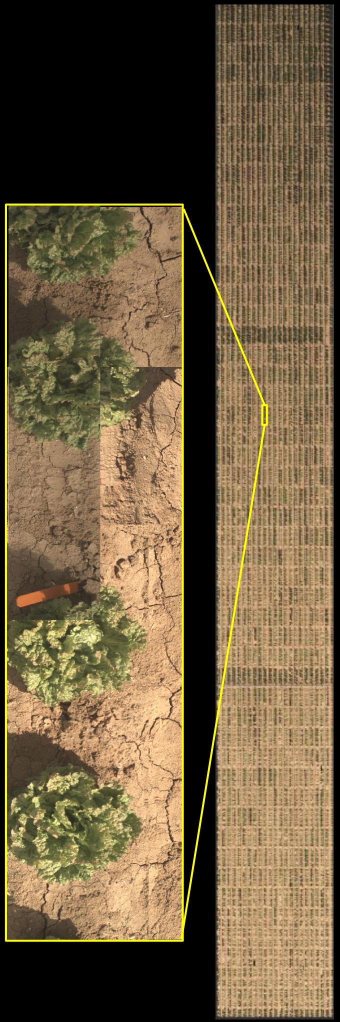


MegaStitch


arpa·e
THE UNIVERSITY OF ARIZONA
TERRA PHENOTYPING REFERENCE PLATFORM
NCSA
PLANT SCIENCE CENTER

GPS errors and image stitching

- Noisy geo-references
  - Relative coordinate system
  - Delays in triggering the cameras
  - Missing images
- Naive stitching not possible
  - GPS error not uniform



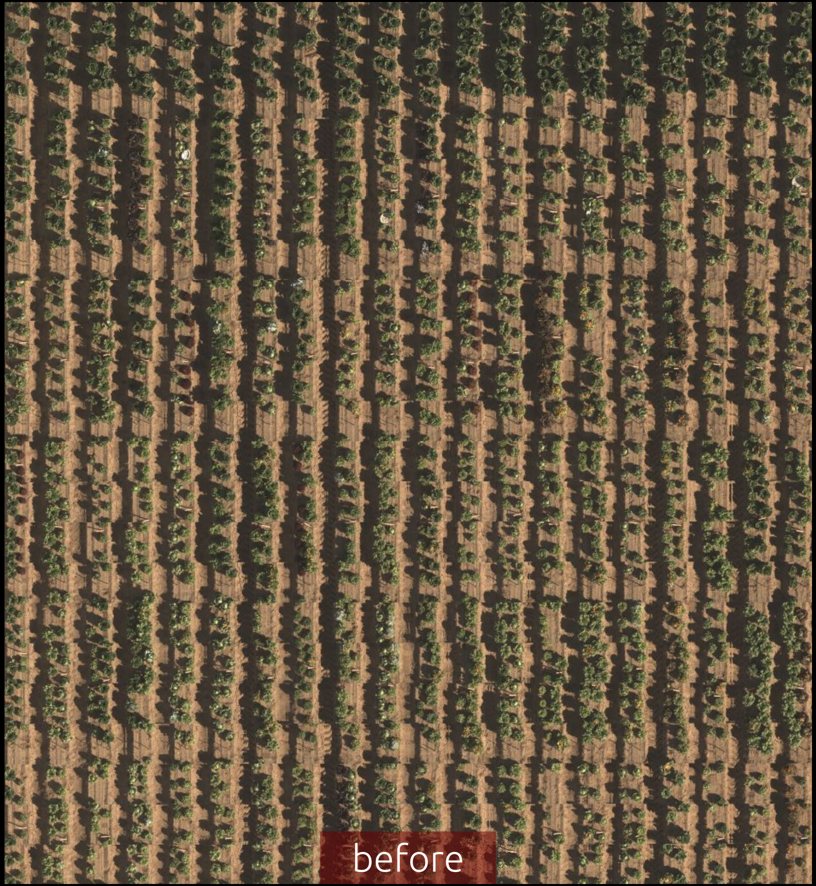
Geo-correction and using GPS metadata

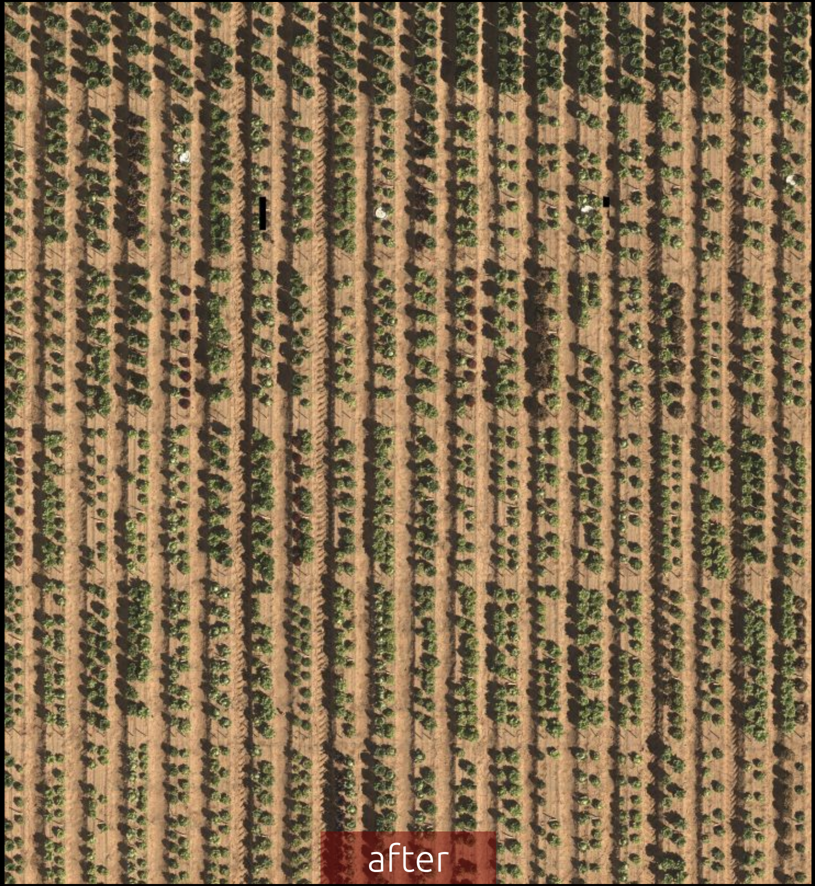
- Large scale image stitching using Megastitch
    - Drift and inconsistency
    - White bucket lids come to rescue!
- Solution?
    - Use noisy GPS as priors
    - Anchor Ground Control Point (GCP) lids to their known location
    - Minimize error globally

Impacts on phenotyping

- Addressed problems
  - Commercial software cannot stitch scanalyzer images
  - Low overlap and few distinct visual features
- Impacts
  - Accurate plot level orthos
  - Identify and match plants from different scans
  - Measure phenotypes accurately

Results





Labelbox Model Training

Image Labeling Example

No annotations yet

Use the tools above to start annotating.

Resources

Documentation:
https://tinyurl.com/phytooracle-rtd

Containers:
https://github.com/phytooracle

Workflows:
https://github.com/LyonsLab/PhytoOracle

Data:
https://tinyurl.com/cyverse-datacommons

Orthomosaics (10% resolution):
https://tinyurl.com/bisque-orthomosaics

ArcGIS map:
https://tinyurl.com/arcgis-phytooracle

Closing Remarks

We thank the AG2PI team.

We also thank the Cyverse staff for the iRODS and virtual machine troubleshooting.

We finally thank Drs. Duke Pauli, Kobus Barnard and Eric Lyons for their support and leadership.

THE UNIVERSITY OF ARIZONA

Work Queue

CYVERSE™

iRODS

Makeflow