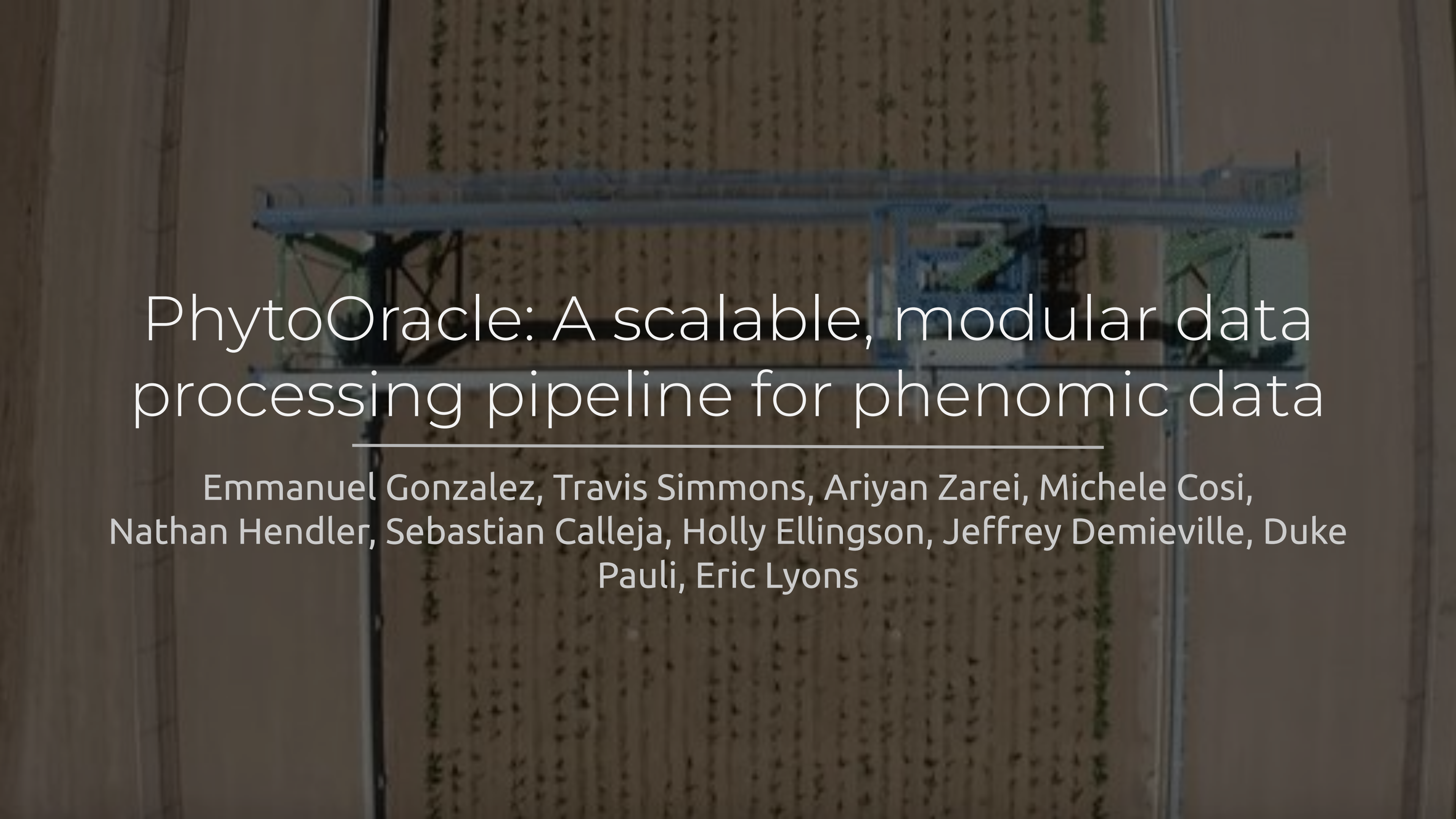
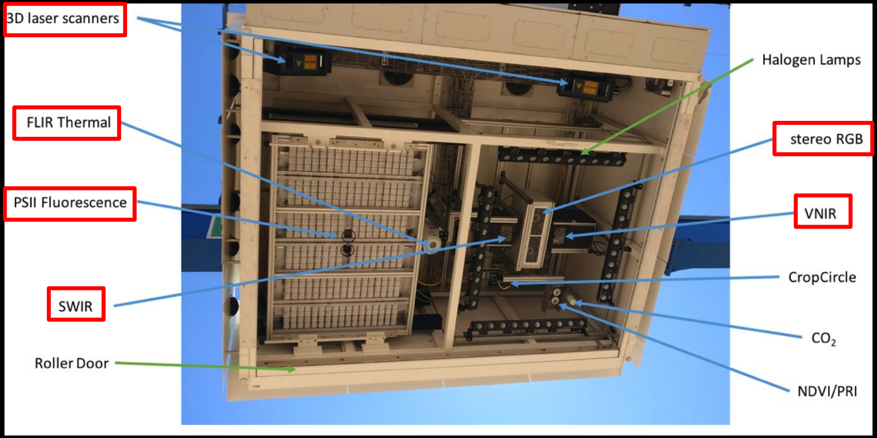


PhytoOracle: A scalable, modular data processing pipeline for phenomic data

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

PhytoOracle

Obtaining and extracting high-quality phenomic data

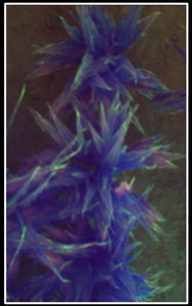


Data volume:

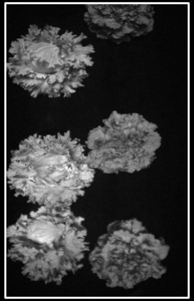
- Max: 10 TB/day
- Typical: 1.5 TB/day

3D

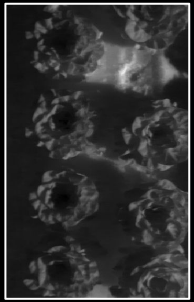
RGB



Hyperspectral



Fluorescence



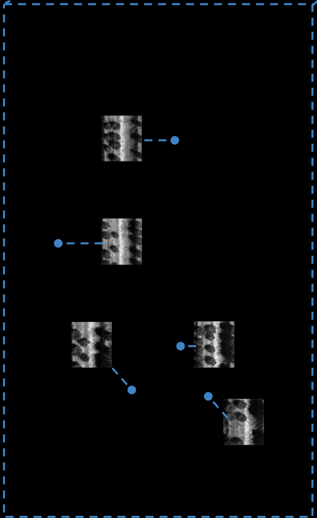
Thermal

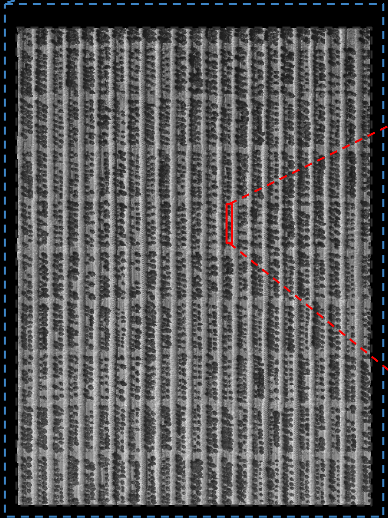
PhytoOracle pipeline workflow

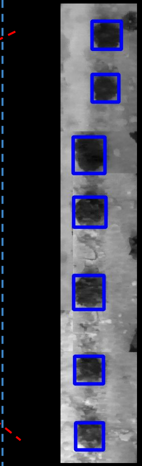
Raw Images

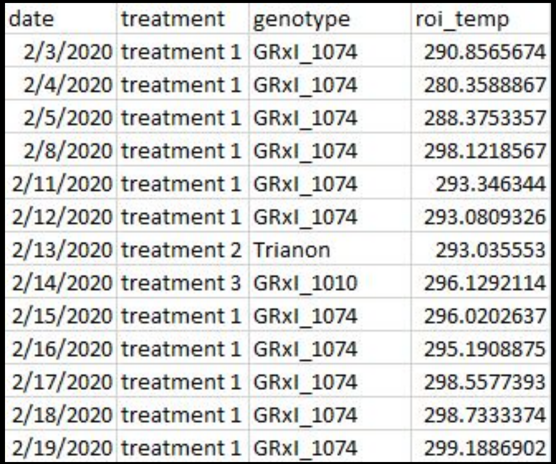
GPS correction, Orthomosaic, Plot clipping

Phenotype Quantification





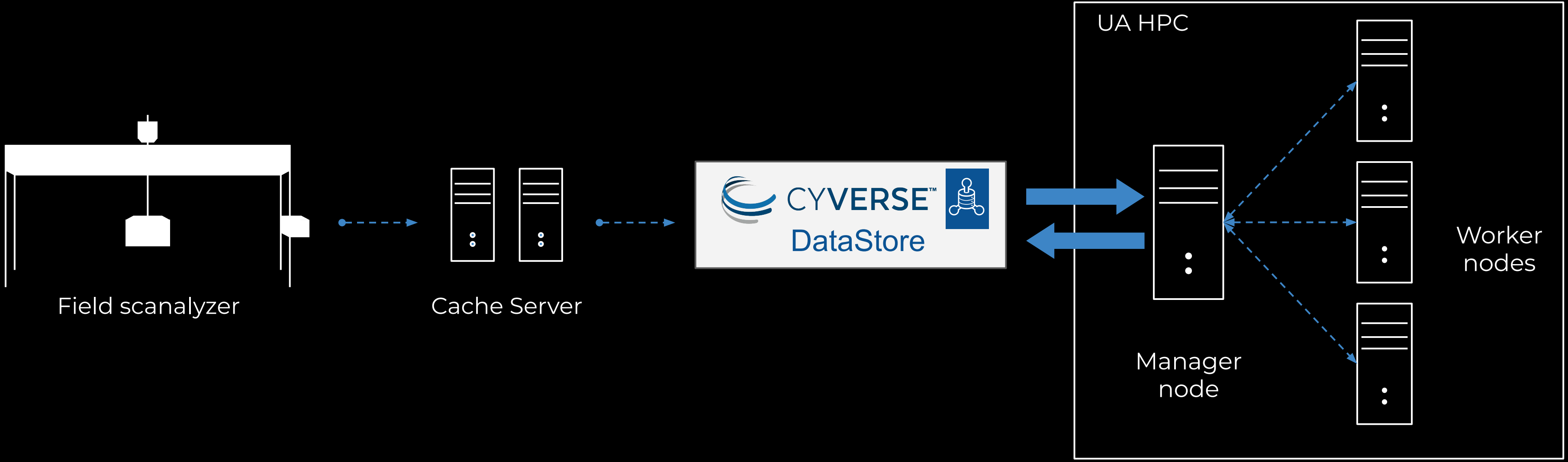




date	treatment	genotype	roi_temp
2/3/2020	treatment 1	GRxI_1074	290.8565674
2/4/2020	treatment 1	GRxI_1074	280.3588867
2/5/2020	treatment 1	GRxI_1074	288.3753357
2/8/2020	treatment 1	GRxI_1074	298.1218567
2/11/2020	treatment 1	GRxI_1074	293.346344
2/12/2020	treatment 1	GRxI_1074	293.0809326
2/13/2020	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

Work Queue



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)?

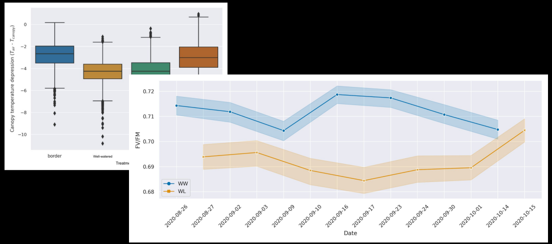
Only 6 days!

---

* From raw data to a quantifiable phenotype

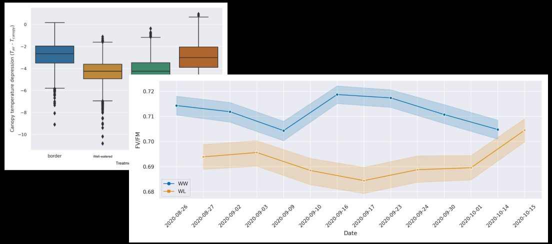
Object detection made easy

- Object detection
  - https://tinyurl.com/objectnotebook



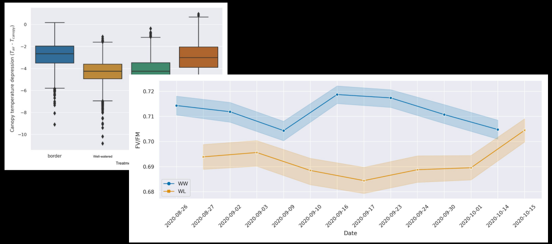
Visualizing output phenotype data

- Visualizing phenotypic trait outputs on streamlit
  - https://tinyurl.com/streamlitapp



Today’s tutorial

- Run 3D workflow using PhytoOracle



Workshop Preparation

Log in to atmosphere

- https://cyverse.org/atmosphere
- Click ‘Projects’
- Click ‘PhytoOracle’
- Click ‘PhytoOracle’
- Click ‘Open Web Shell’

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

Closing Remarks

We thank the NAPPN 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