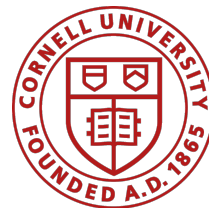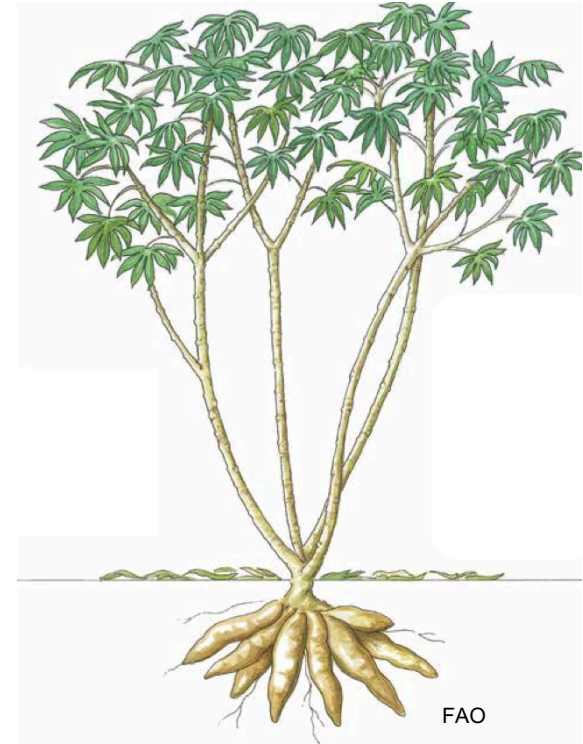# Development and integration of handheld NIRS-based phenotyping for cassava breeding

Jenna Hershberger

Cassava: plant of many uses
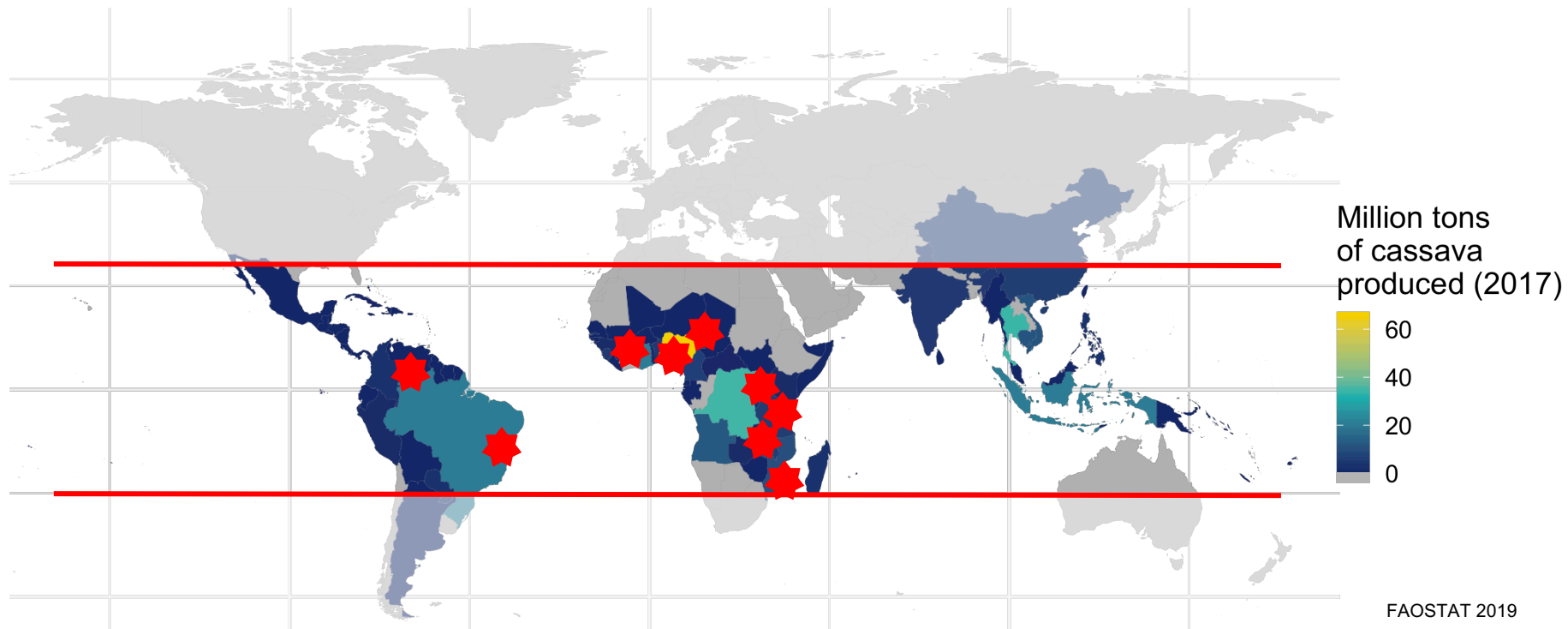

FAO

Cassava: a tropical staple

- Roots consumed by over 800 million people
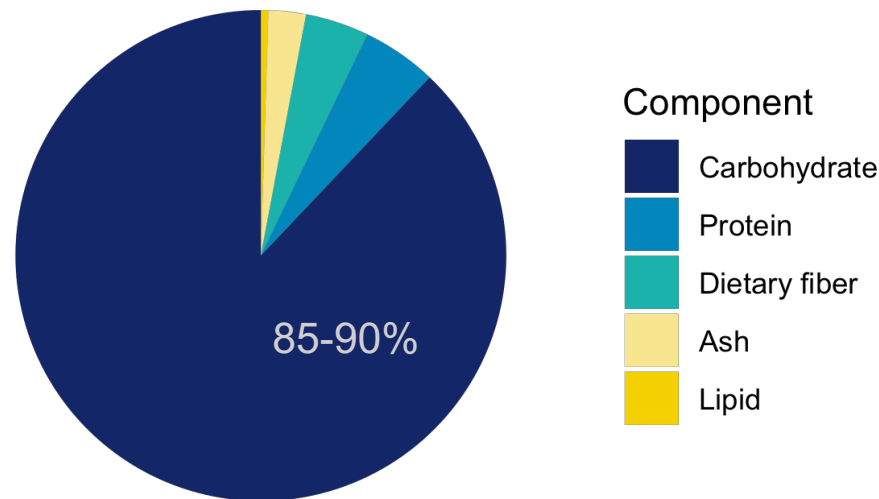- Drought-tolerant subsistence crop

FAO

Cassava: a tropical staple

Million tons
of cassava
produced (2017)

60
40
20
0

FAOSTAT 2019

Cassava breeding targets

- Pest and disease resistance
- Yield and performance
- Quality traits

Photo from Paula Iragaba

Root dry matter content (DMC)

- $\dfrac{dry\,weight}{fresh\,weight} \times 100\%$

- Major component of dry yield and food quality

- Typically ~20-45%

- Not uniform within the root

DMC composition

85-90%

Component

- Carbohydrate
- Protein
- Dietary fiber
- Ash
- Lipid

CornellCALS College of Agriculture and Life Sciences

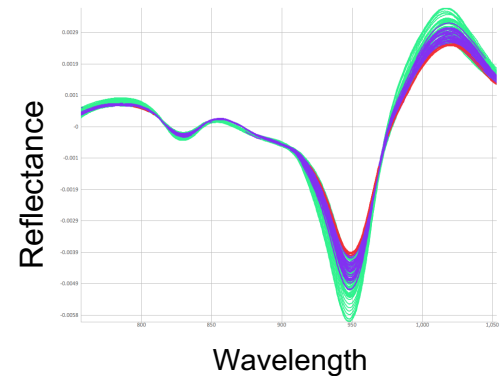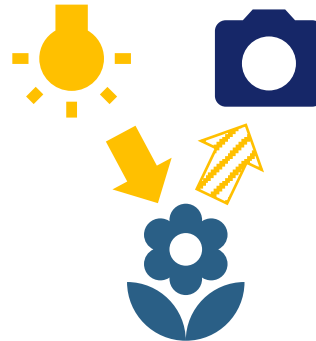# DMC is time-consuming and laborious to measure

7000+ plots per season • Limited electricity • Off-site trials

Near-infrared spectroscopy (NIRS)

- Reflectance from near-infrared region of the electromagnetic spectrum
- Complex signal from biophysical and biochemical properties of sample
- Multivariate statistical techniques required for calibration
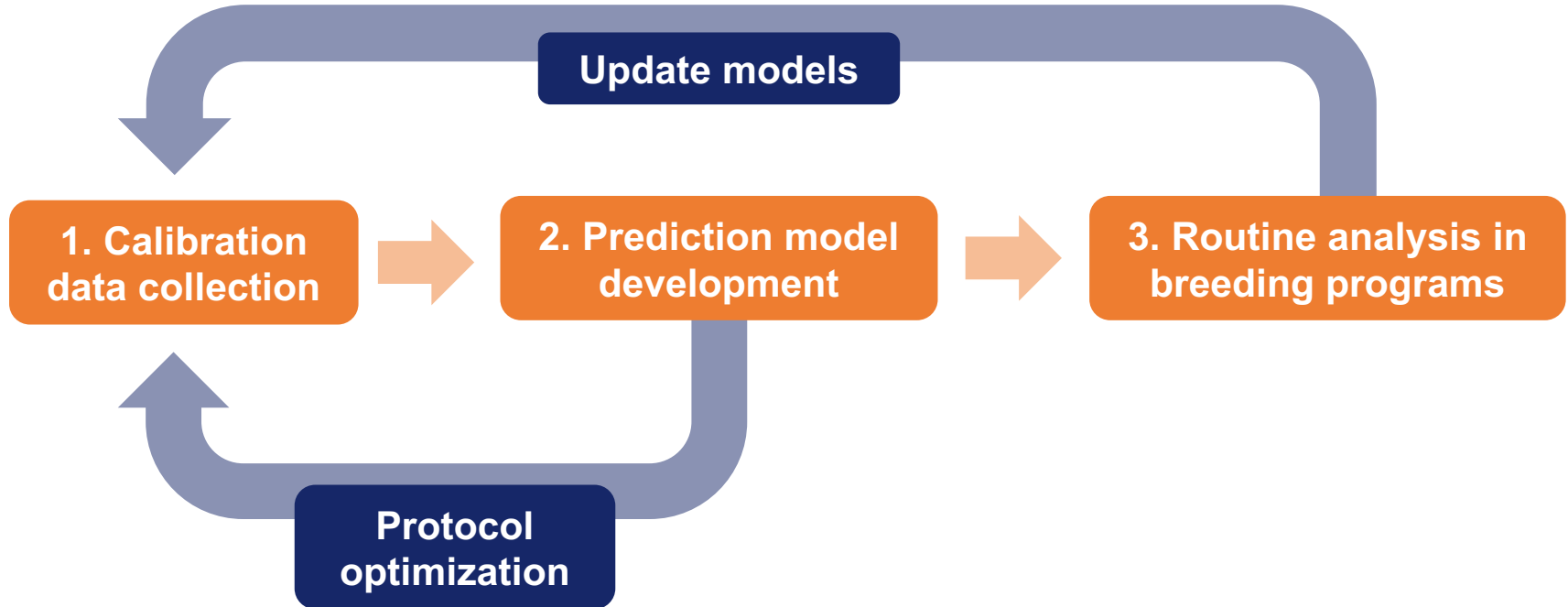- Allows for rapid, non-destructive analyses

NIR – 780:2500nm

| X-ray | UV | | Infrared |

Reflectance

Wavelength

Cornell **CALS** College of Agriculture and Life Sciences

NIRS for cassava phenotyping

- Highly predictive of dry matter in cassava roots
- Validated spectrometers are prohibitively expensive

Foss 6500
Spectral range	400-2500nm
Predictive ability	> 0.94

NIRS validation workflow

Update models

1. Calibration data collection → 2. Prediction model development → 3. Routine analysis in breeding programs

Protocol optimization

1. Calibration data collection

IITA (Nigeria)	NaCRRI (Uganda)	Embrapa (Brazil)

1. Calibration data collection
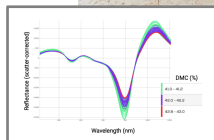
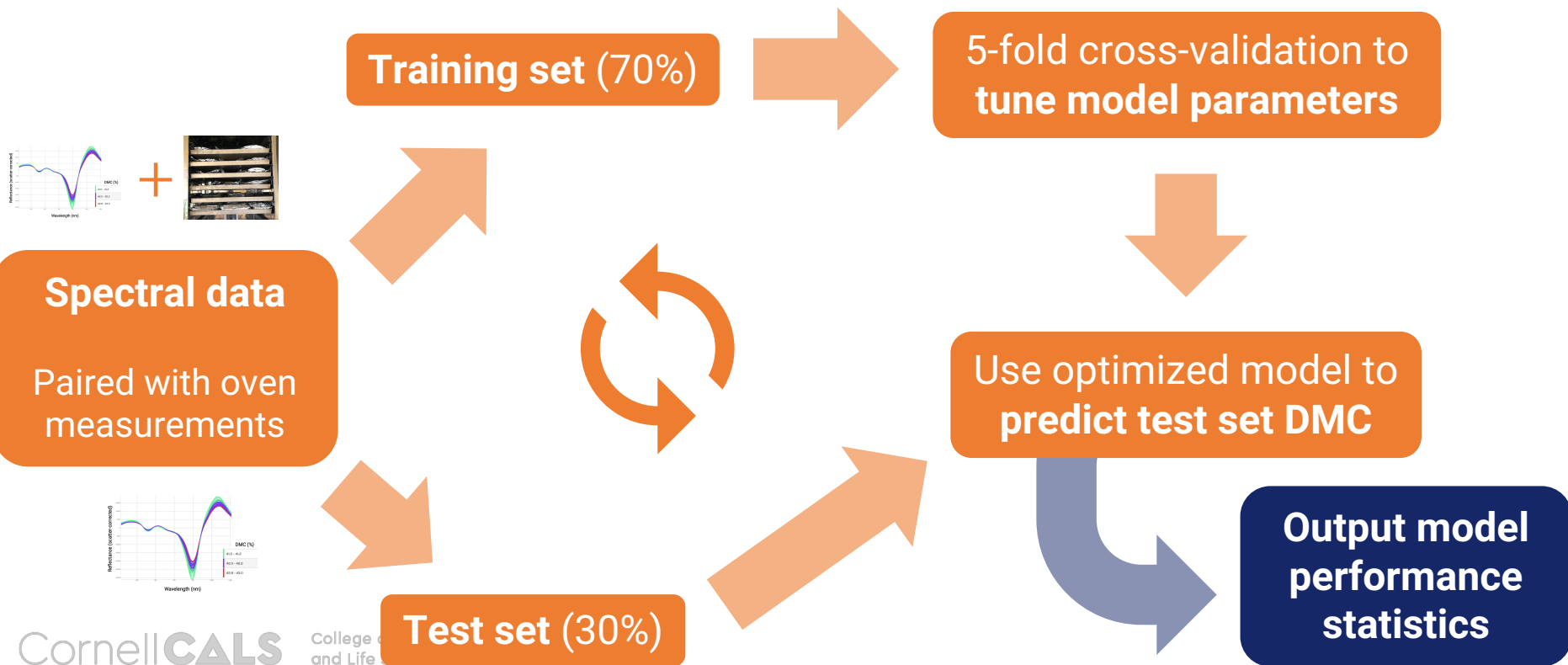| Sample three roots/plot | → | Slice root in three regions | → | Scan the cut surface from each region | → | Measure oven DMC with a representative sample from the entire root |

$$\frac{dry\ weight}{fresh\ weight} \times 100\%$$

College of Agriculture and Life Sciences
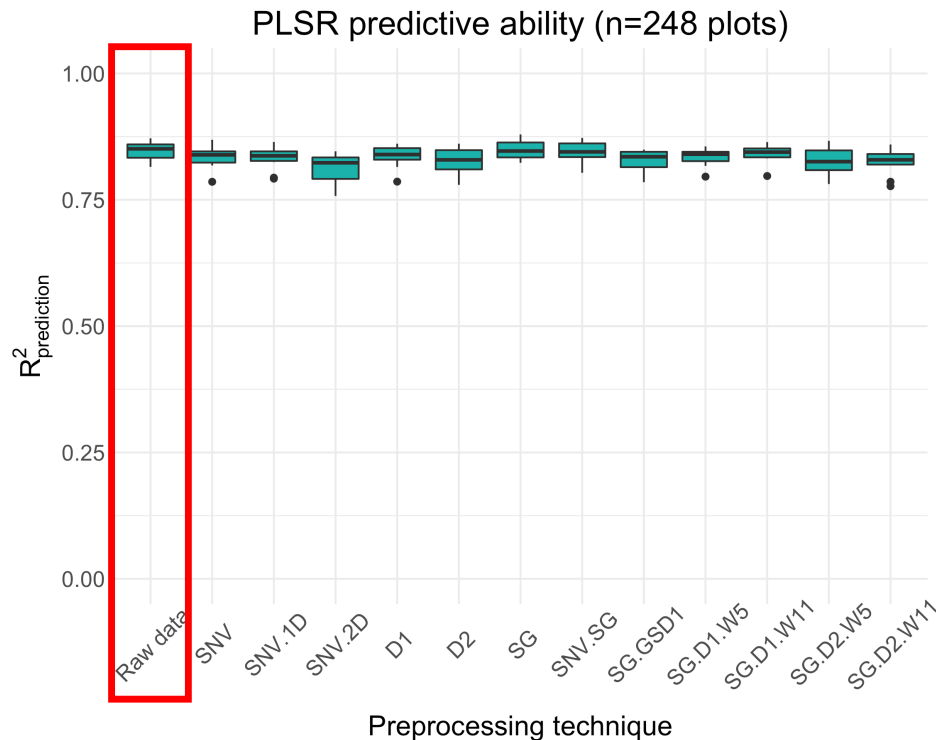
2. Prediction model development

Optimization:
PLSR outperforms other model types

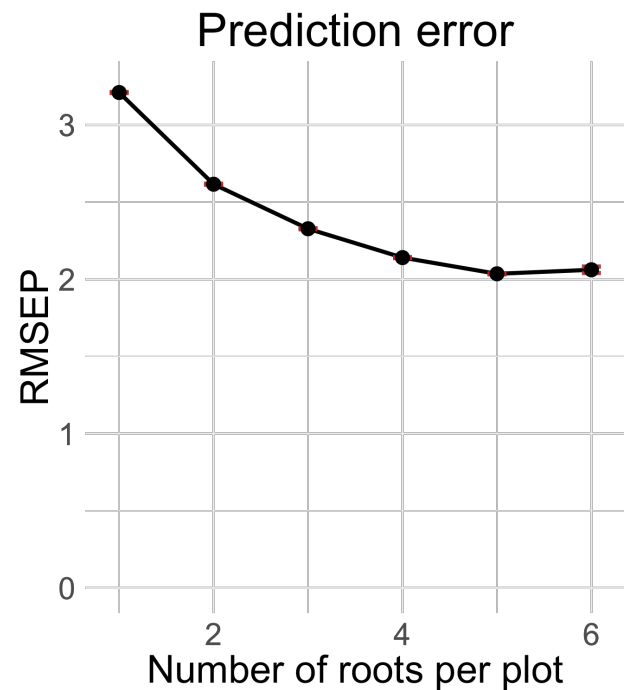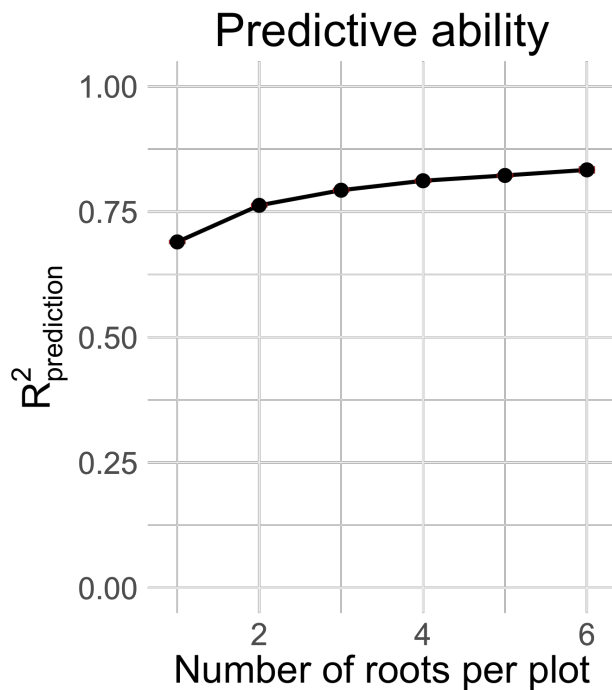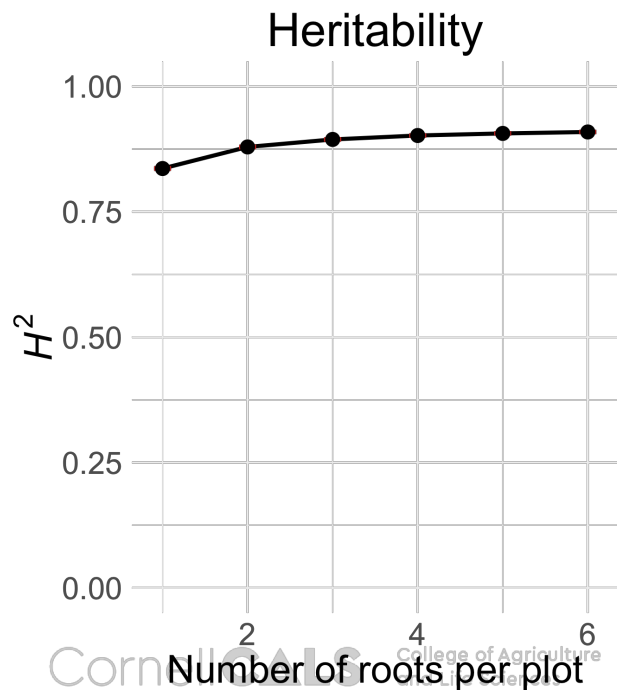| Model type |
|---|
| Partial Least Squares Regression (PLSR) |
| Support Vector Machine (SVM) |
| Random Forest (RF) |

- Models developed and evaluated using IITA dataset
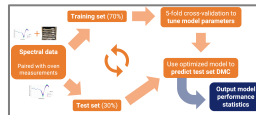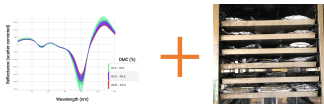- Ten iterations of model development pipeline

Optimization:
No significant change with preprocessing

- Scatter correction
  - Standard normal variate (SNV)

- Derivatives
  - First and second derivatives (1D, 2D)
  - Window size 11 unless noted by "W"
  - Gap segment derivative (GSD)
  - Savitzky-Golay (SG)

PLSR predictive ability (n=248 plots)

Optimization:
Diminishing returns after three roots

1. Calibration data collection → 2. Prediction model development → 3. Routine analysis in breeding programs

Protocol optimization

- PLSR
- No preprocessing
- 3 roots per plot

SCiO

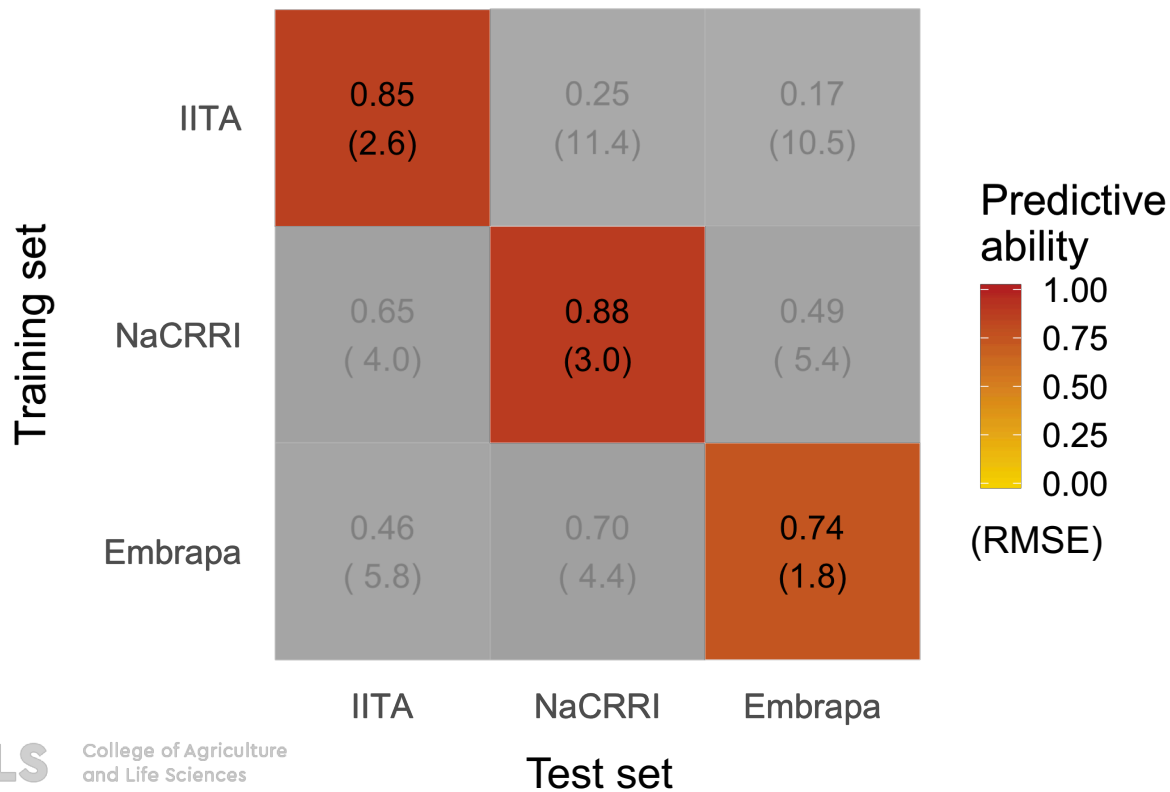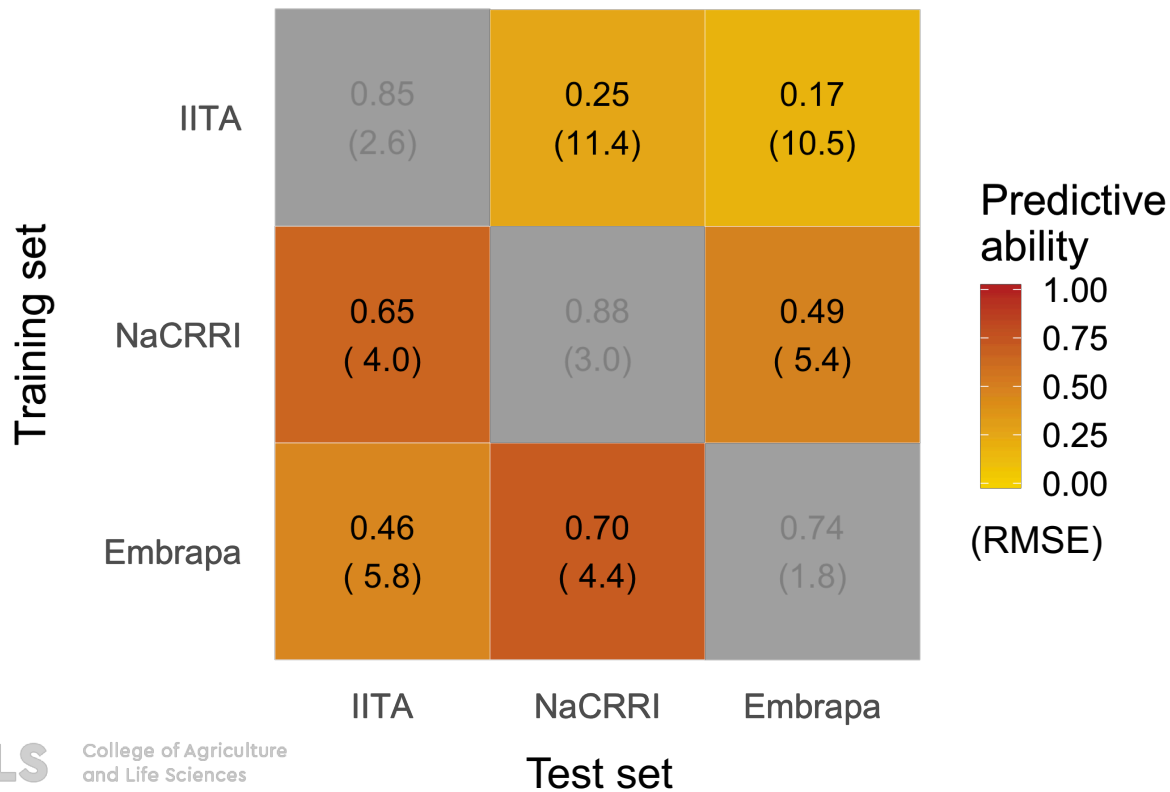|	
Spectral range	740-1070nm
Predictive ability	??
Price per unit	$299

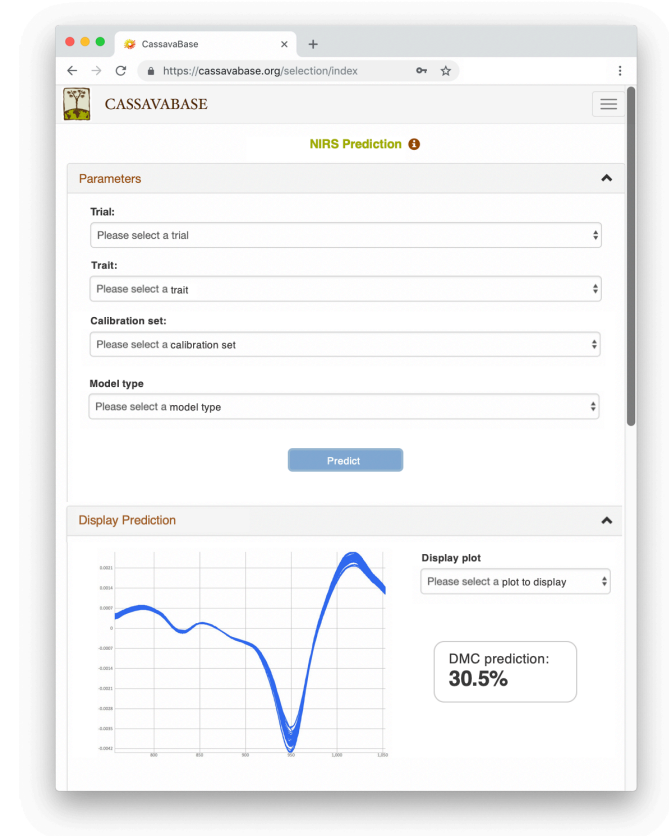# SCiO is highly predictive of DMC within breeding programs

SCiO is moderately predictive of DMC across breeding programs

3. Routine use in breeding programs

- Integrate NIRS data and models into Cassavabase

- Implement model update strategy

Acknowledgements

Gore Lab at Cornell University

NextGen Cassava
- **IITA** – Ismail Rabbi and team
- **NaCRRI** – Robert Kawuki and team
- **Embrapa** – Eder de Oliviera and team
- **Cassavabase** – Lukas Mueller and team

PhenoApps team
- Kansas State University
- Makerere University

Committee Members
- Michael Gore
- Rebecca Nelson
- Tim Setter