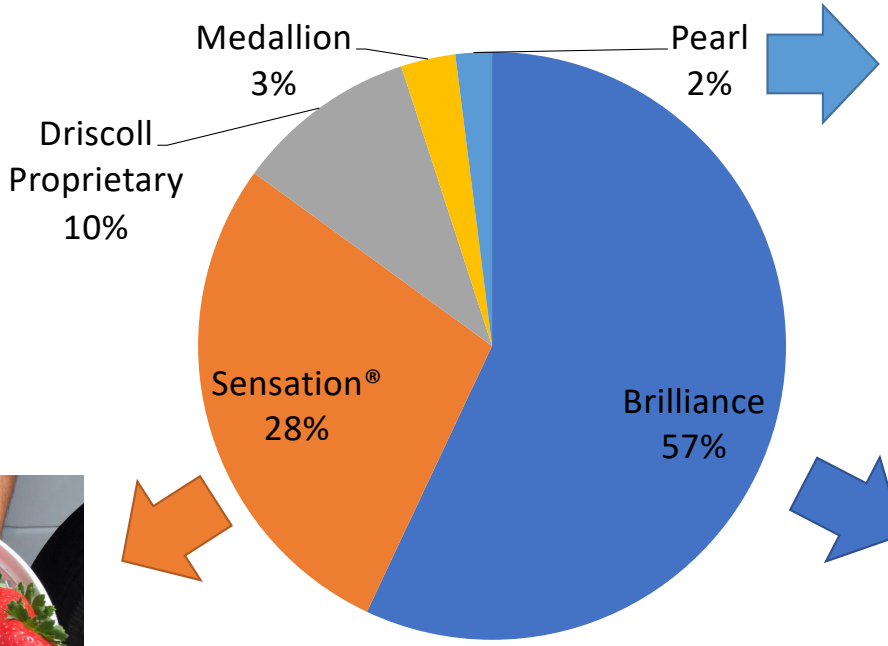
Big data applications in strawberry breeding

Zhen Fan, PhD, Postdoctral Associate
Co-authors: Vance Whitaker, Sujeet Verma, Luis Osorio, Ronald Tapia, Cheryl Dalid
Horticultural Sciences Department
Gulf Coast Research and Education Center
University of Florida

The UF strawberry breeding program



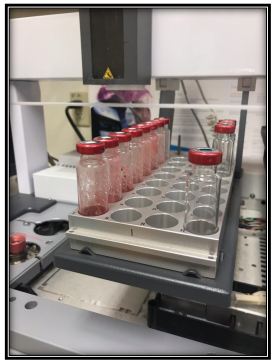
Categories of big data in strawberry breeding

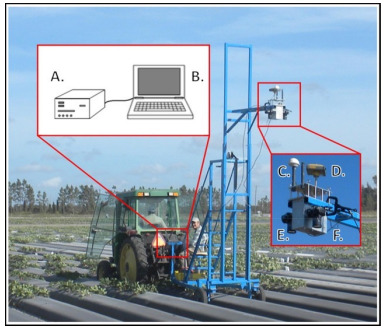
Survey/Field data
Sensory evaluation
Field phenotyping

Genomics data
Short reads (Illumina®)
Long reads (Pacbio® Hifi)
SNP array

Other omics data
Transcriptome
Metabolites (Volatile)

Remote sensing data
Canopy reflectance
High resolution image





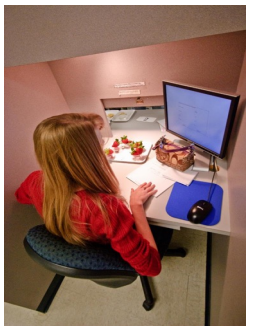
Volatiles modulating eating experience

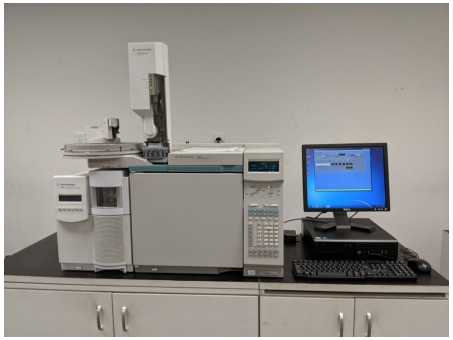
Objectives:

- Identification of volatiles **important for consumer liking**
- **Prediction of consumer liking** with metabolites data

Materials and methods:

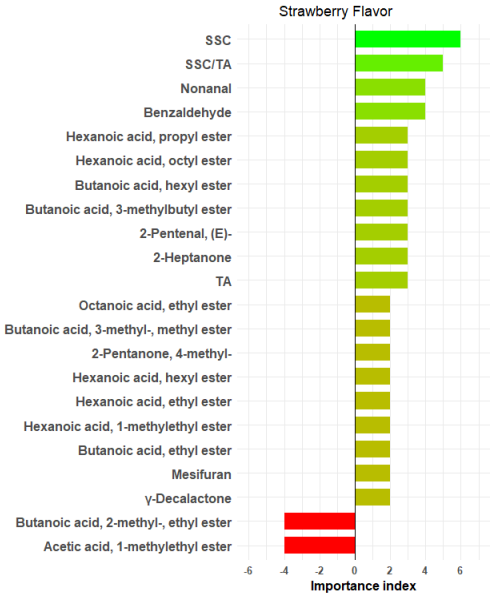
- **154** genotype/harvest date combinations, over **15,400** sensory/liking evaluations (consumer panel)
- **213** genotype/harvest date combinations, over **2,130** sensory evaluations (descriptive panel)





Volatile + Sensory

Volatiles modulating eating experience

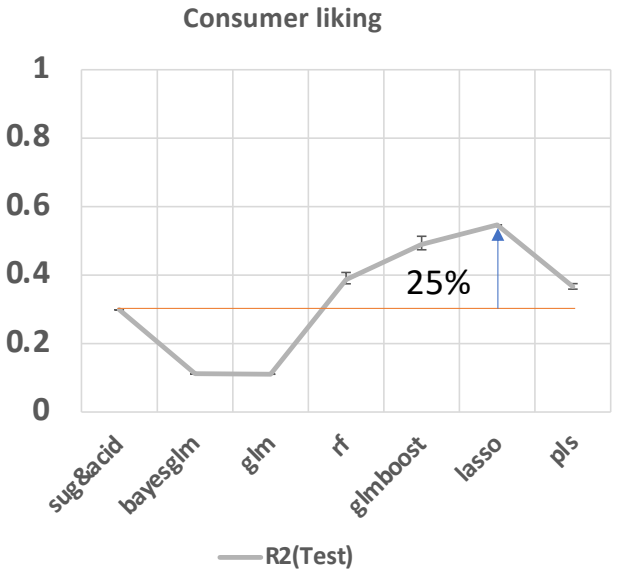


Article | Open Access | Published: 01 April 2021

Strawberry sweetness and consumer preference are enhanced by specific volatile compounds

Zhen Fan, Tomas Hasing, Timothy S. Johnson, Drake M. Garner, Michael L. Schwieterman, Christopher R. Barbey, Thomas A. Colquhoun, Charles A. Sims, Marcio F. R. Resende & Vance M. Whitaker

Horticulture Research **8**, Article number: 66 (2021) | Cite this article



ORIGINAL RESEARCH article
Front. Plant Sci., 17 March 2021 | https://doi.org/10.3389/fpls.2021.640704

Volatiles Influencing Sensory Attributes and Bayesian Modeling of the Soluble Solids–Sweetness Relationship in Strawberry

Zhen Fan[1], Anne Plotto[2], Jinhe Bai[2] and Vance M. Whitaker[1*]

RNAseq + Hifi + Volatile + SNP array

Discovering flavor genes and their regulatory elements

Objectives:

- Identification of **biosynthetic genes** for volatile production
- Exploration of **natural variation** in their **regulatory regions**

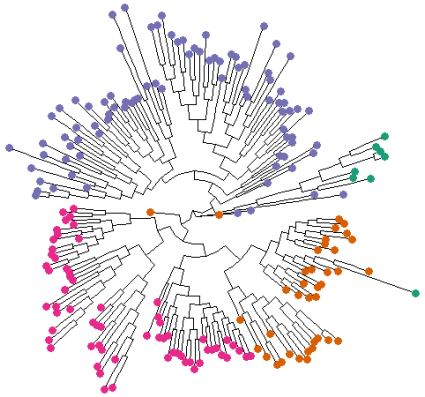
Materials and methods:

- RNAseq data for **196** breeding accessions
- Pacbio Hifi reads for a UF variety; short reads for parents
- Volatile quantification for a GWAS panel with over **300** individuals

RNAseq + Hifi + Volatile + SNP array

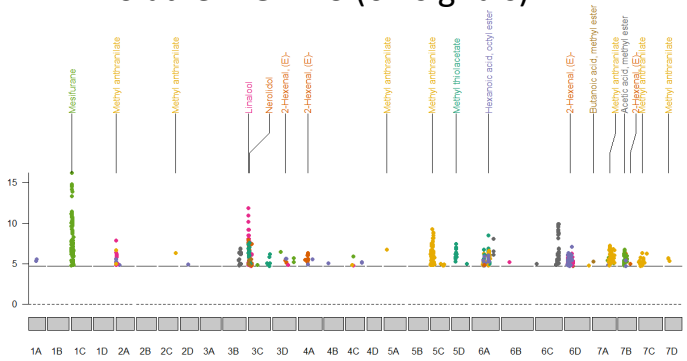
Discovering flavor genes and their regulatory elements

eQTL

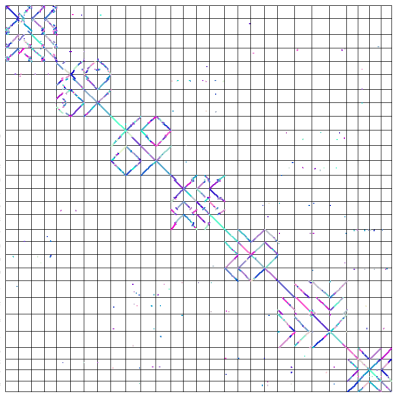


Expressed gene	62181
# eQTL/cis-eQTL/trans-eQTL	68535/23452 /45083
eGenes with cis-eQTL	22731
eGenes with dis-eQTL	20650

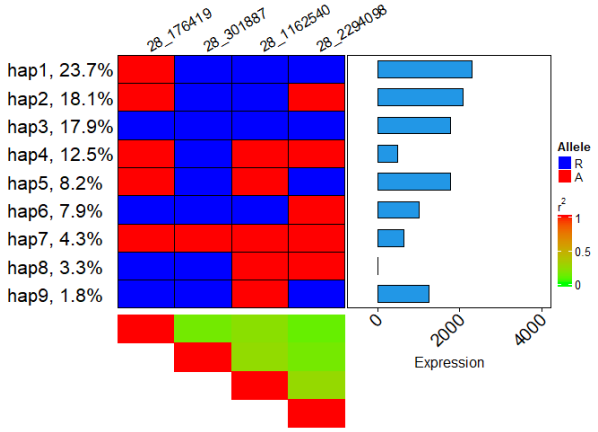
Volatile – GWAS (62 signals)



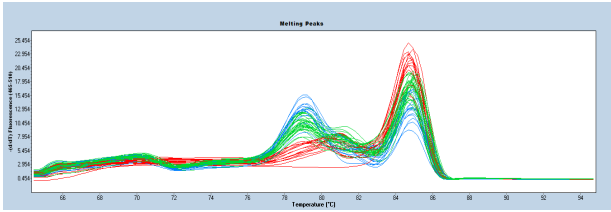
Phased genome (88 & 79 gaps)



Favorable alleles



Functional markers



Paper in review

SNP array + GS

Genomic Selection

Objectives

1. Predict parents from the seedling population
2. Predict parents from current Advanced selection trials

Dr. Luis Osorio

Dr. Sujeet Verma

GS Methodology

- GP Methods: GBLUP, Bayes B, RKHS, Deep learning models
- Software: ASReml-R, BGLR, Tensorflow
- GP Models: Additive
- GBVs: Phenotype Ranking & Selection Index

SNP array + GS

Predictive Ability

Predictive Ability: $corr(y, \hat{y})$

Trait	PBLUP	GS Methods		
		GBLUP	Bayes B	RKHS
AWT	0.44	0.49	0.49	0.52
SSC	0.37	0.43	0.44	0.45
EMY	0.14	0.29	0.3	0.3
TMY	0.24	0.31	0.35	0.33
TC	0.14	0.32	0.35	0.32

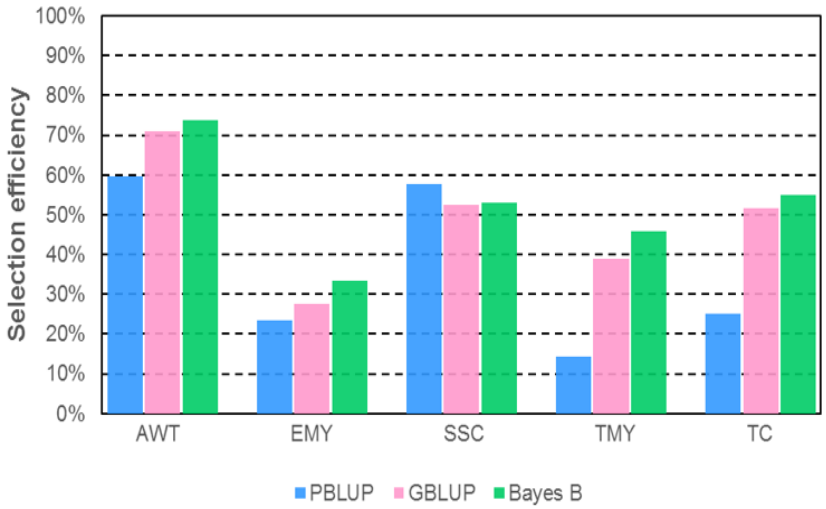
An experimental validation of genomic selection in octoploid strawberry

Salvador A Gezan, Luis F Osorio, Sujeet Verma, Vance M Whitaker

Horticulture Research, Volume 4, 2017, 16070, https://doi.org/10.1038/hortres.2016.70

Published: 11 January 2017 **Article history**

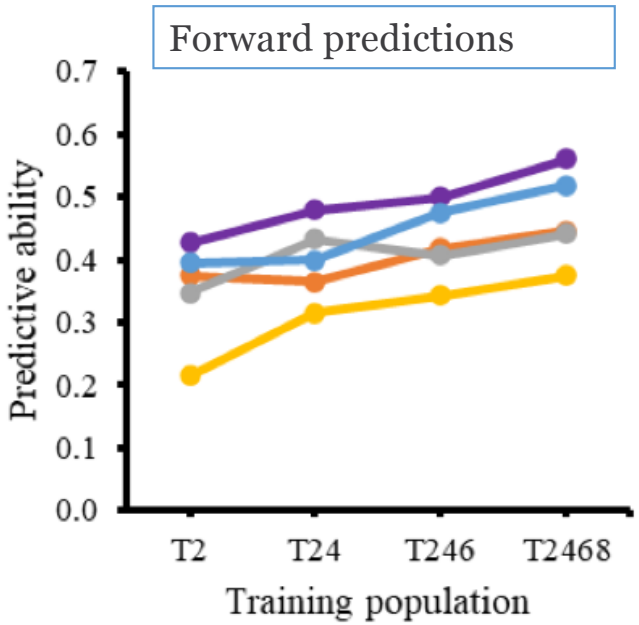
Parent Selection Efficiency (PSE)

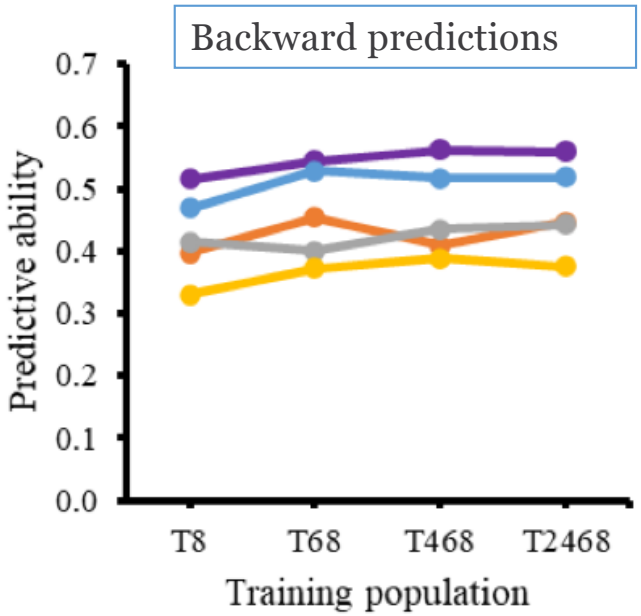


$$PSE = GG_{PRED}/GG_{TOTAL}$$

SNP array + GS

Validation of GS over multiple cycles





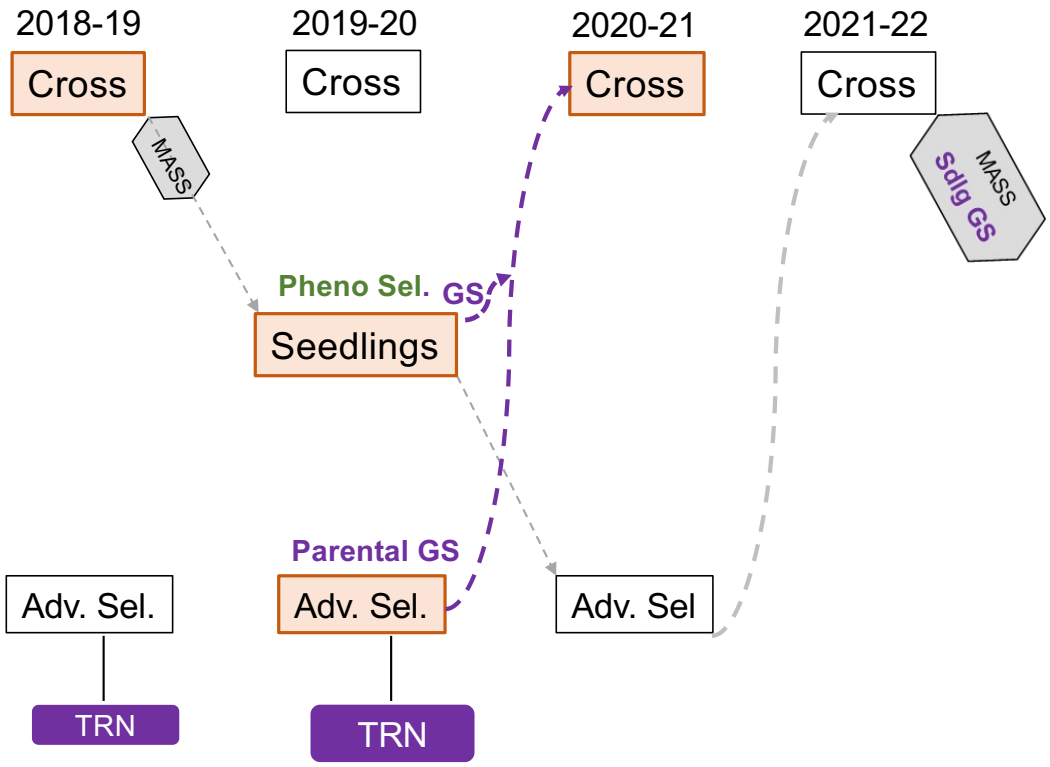
Independent Validation of Genomic Prediction in Strawberry Over Multiple Cycles

Luis F. Osorio[1], Salvador A. Gezan[2†], Sujeet Verma[1] and Vance M. Whitaker[1*]

[1]Gulf Coast Research and Education Center, University of Florida, Wimauma, FL, United States
[2]School of Forest Resources and Conservation, University of Florida, Gainesville, FL, United States

SNP array + GS + MASS

Strawberry breeding cycles



Modeling strawberry biomass and leaf area

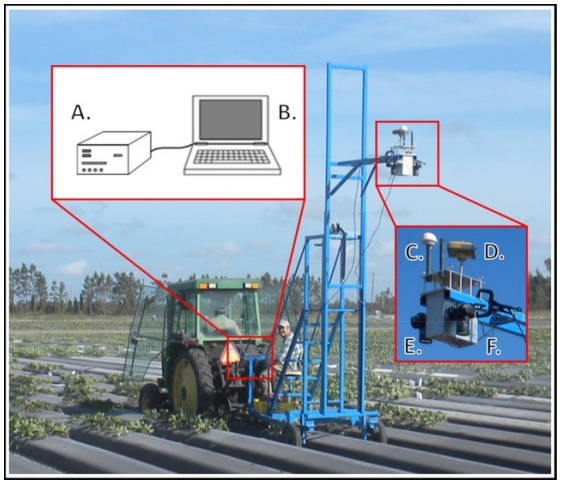
DR. AMR ABD-ELRAHMAN

Objectives:

- Modeling strawberry **biomass** and **leaf area** using object-based analysis of high-resolution images

Materials and methods:

- Vegetation Mobile Mapping System (VMMS) consisted of two digital cameras, a GPS and a GNSS receiver
- Total leaf area and dry biomass of the plants were measured using destructive methods.



High-resolution images

Modeling strawberry biomass and leaf area

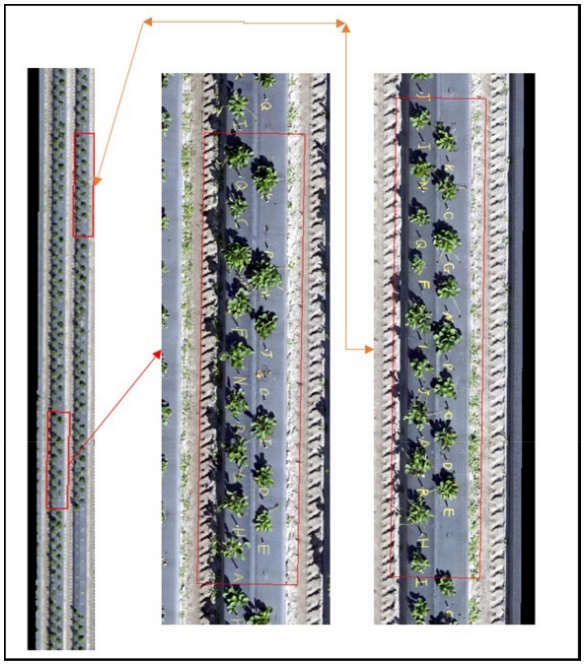


Table 8. LOOCV and ten-fold CV results for the dry biomass and leaf area models.

	Dry Biomass Models			Leaf Area Models		
	R^2	Cross-validation RSME (g)	Cross-validation RSME (%)	R^2	Cross-validation RSME (m^2)	Cross-validation RSME (%)
LOOCV	0.82	7.96	8.75	0.79	5.80E-02	9.28
Ten-fold CV	0.84	7.72	8.48	0.80	5.60E-02	8.96

Modeling strawberry biomass and leaf area using object-based analysis of high-resolution images

Zhen Guan [a, b], Amr Abd-Elrahman [a, b], Zhen Fan [c], Vance M. Whitaker [c], Benjamin Wilkinson [a]

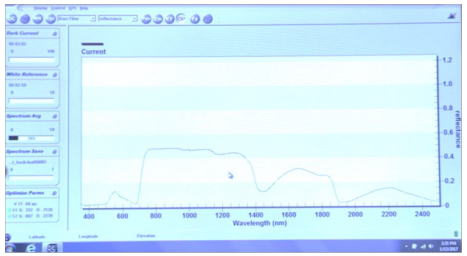
Canopy reflectance spectrometry + GS

Prediction of powdery mildew resistance in strawberry

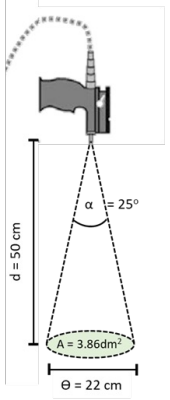
Canopy reflectance measurement

ASD Field Spec 4

Real-time spectral data recording

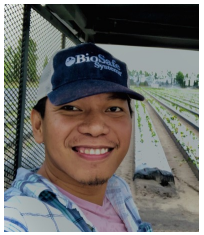


Field of view



Canopy reflectance acquisition:

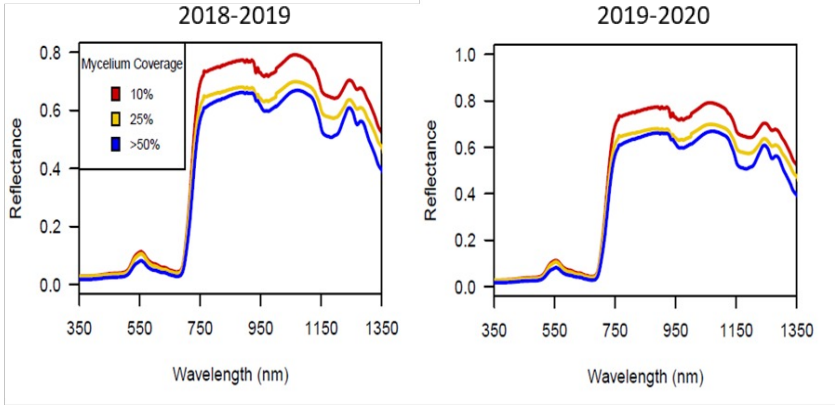
- 350-2500 nm with 1 nm resolution
- 10 AM – 2 PM with full sunlight
- Calibration every 10 mins with white board panel
- Capture 3 times/plant where each is an average of 10 recordings
- >8,000,000 data points for 450 indivduals



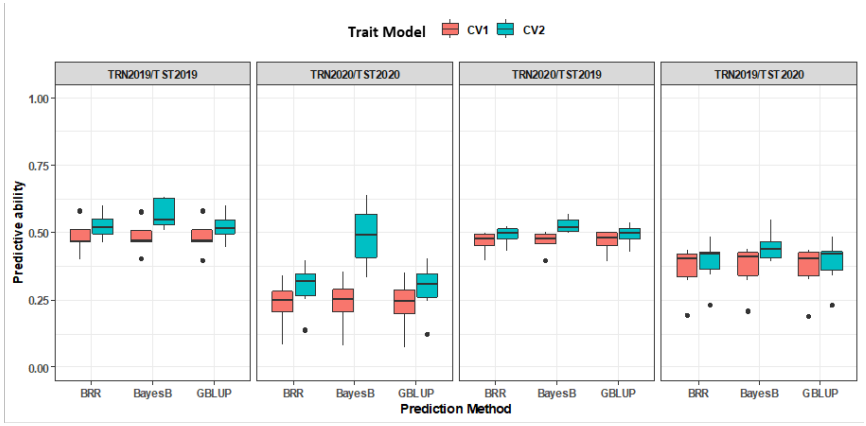
Ronald Tapia
PhD student

Canopy reflectance spectrometry + SNP array

Prediction powdery mildew resistance in strawberry



Canopy-based spectral signatures of powdery mildew disease of strawberry.



PA using simple univariate method with only genomic information, and a combination of genomic and spectral data as predictors.

Combining canopy reflectance spectrometry and genome-wide prediction to increase response to selection for powdery mildew resistance in cultivated strawberry

Ronald Tapia[1,2], Amr Abd-Elrahman[1,3], Luis Osorio[1,2], Vance M. Whitaker[1,2,*] and Seonghee Lee[1,2,*]

[1] Gulf Coast Research and Education Center, Institute of Food and Agricultural Science, University of Florida, 14625 County Road 672, Wimauma, FL 33598, USA
[2] Department of Horticultural Sciences, University of Florida, Gainesville, FL 32611, USA
[3] School of Forest, Fisheries, and Geomatics Sciences, University of Florida, Gainesville, FL 32603, USA

Acknowledgements

UF strawberry breeding lab
UF strawberry molecular genetics and genomics lab
Dr. Abd-Elrahman's research program
Horticultural Research Laboratory, USDA-ARS, Fort Pierce
UC Davis strawberry breeding lab
Dr. Thomas Colouhoun's research program
Dr. Charlie Sims sensory lab

Dr. Vance Whitaker