A QTL-based Phenology Model of the Common Bean With Non-linear Genetic x Environment Effects

Chris Hwang

Crops *in silico* Symposium

University of Illinois at Urban Champaign

May, 3rd, 2019

Population growth & food security

”Global population of 9.7 billion by 2050.” – UN, 2015

”Global arable land stagnated at 10.9%.” – World Bank, 2016

Plant Breeding

Crop Growth Models

”Agriculture accounts for 60% of fresh water withdrawal.” – World Bank, 2016

”Recent extreme weather events resulted in global cereal production losses of 9~10%.” – Lesk et al. 2016, Nature

Population growth & food security

”Global population of 9.7 billion by 2050.” – UN, 2015

”Global arable land stagnated at 10.9%.” – World Bank, 2016

Plant Breeding — Crop Growth Models

”Agriculture accounts for 60% of fresh water withdrawal.” – World Bank, 2016

”Recent extreme weather events resulted in global cereal production losses of 9~10%.” – Lesk et al. 2016, Nature

Gene-based modeling: Linear regression models

Plant Breeding

Crop Growth Models

Linear mixed-effect Model with G x E

$$\hat{y}_{gl} = \mu + \sum_{e=1}^{E} \beta_{E_e} \Delta E_{gle} + \sum_{q=1}^{Q} \beta_{QTL_q} QTL_{gq} + \sum_{e=1}^{E} \sum_{q=1}^{Q} \beta_{QTL_q x E_e} QTL_{gq} \Delta E_{gle}$$

$$\Delta E_{gle} = \left(E_{gle} - \overline{E}_{gle}\right)$$

- Predicts static traits
- Predicts complex traits
- Limited application in training locations
  - Linear response to environmental factors

Boer et al. 2007

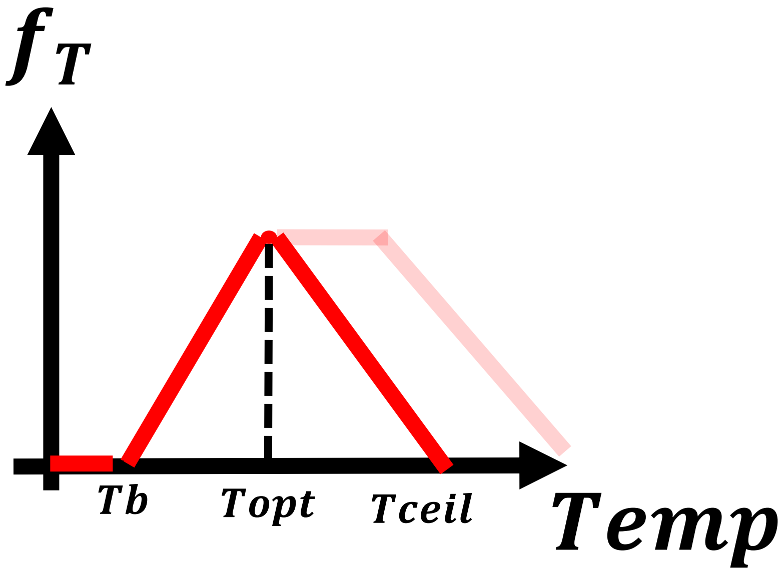
Gene-based modeling: Crop growth models

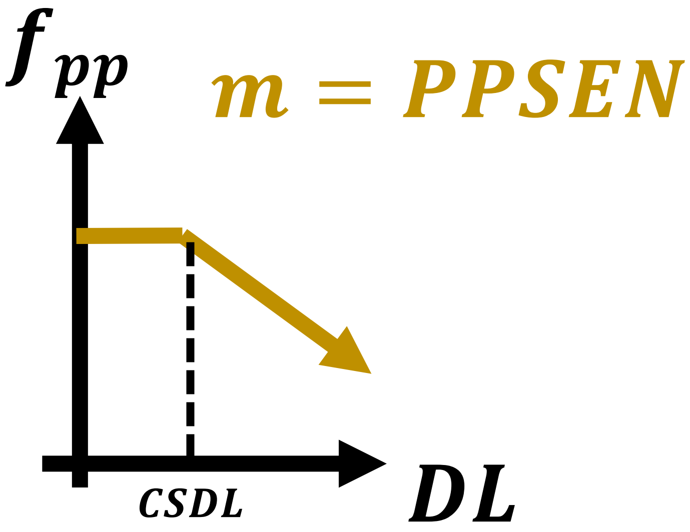
Plant Breeding

Crop Growth Models

Photothermal accumulation functions

$$\text{phys\_days}(t) = \sum_{t=1}^{\text{phys\_days}(t) \leq \textit{Threshold}} \left[f_T(\text{Temp}, Tb, Topt, \ldots) * f_{pp}(CSDL, PPSEN)\Delta t \right]$$





UF UNIVERSITY of FLORIDA

USDA USAID FEED THE FUTURE NSF

Gene-based modeling: Modifying breeding models

Plant Breeding

- Predicts static traits
- Predicts complex traits
- Limited application in training locations
  - Linear response to environmental factors

Crop Growth Models

- Dynamic simulations
- Simulate component traits
- Non-linear environmental response

Non-linear, dynamic breeding models

Data

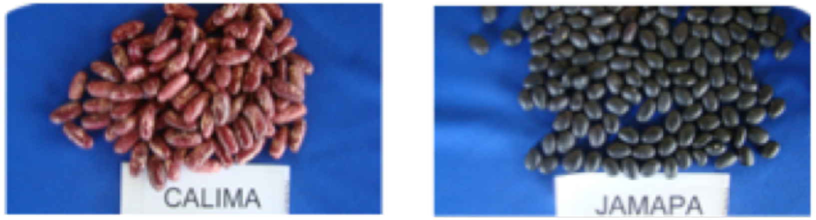
Data: Crop information

I. Common bean (*Phaseolus vulgaris L.*)

C a. *Calima* (Parent A)
   - I. Determinate, Andean
   - II. Sensitive to photoperiod

J b. *Jamapa* (Parent B)
   - I. Indeterminate, Mesoamerican
   - II. Insensitive to photoperiod

c. Recombinant Inbred Lines (RILs)
   - I. 184 individuals (F11: 14)

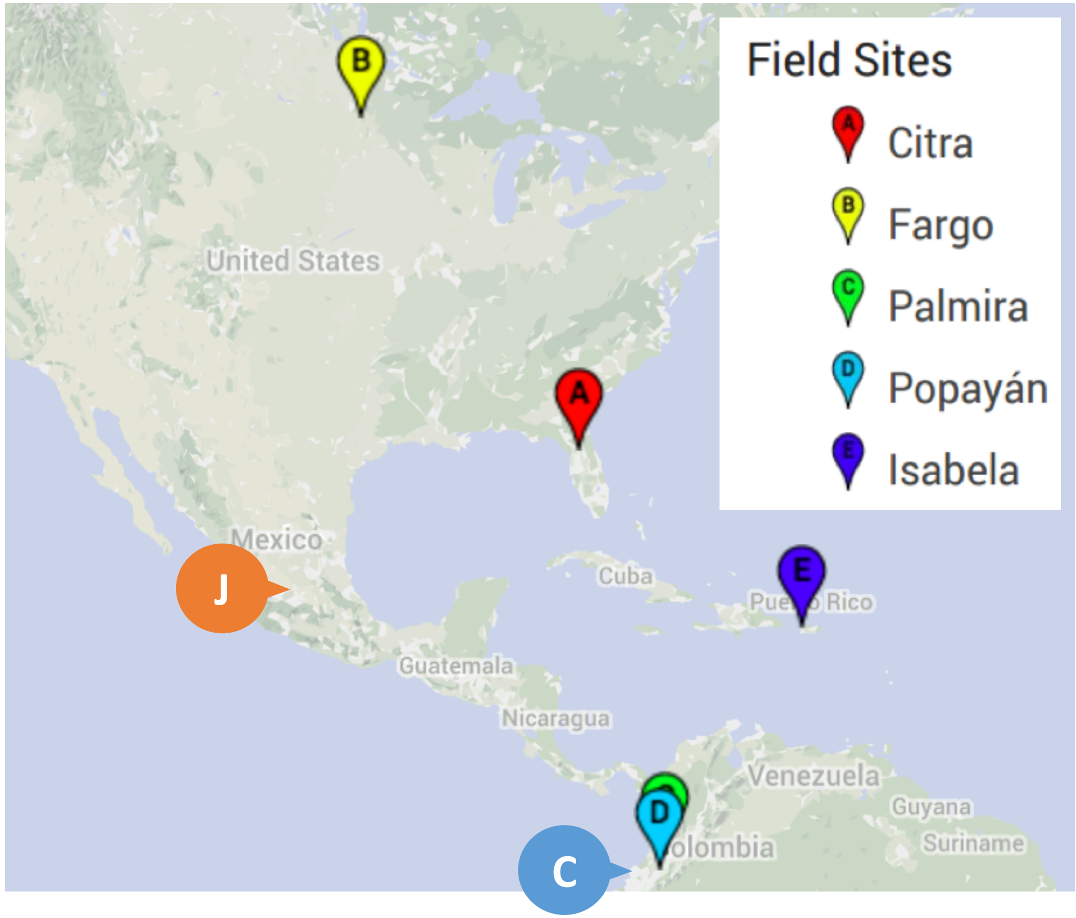


Seed Diversity within RILs

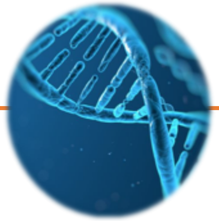
Source: Clavijo Jose, NSF Presentation, 2014

Data: Site information

Location	Day length (h)	Year	Temp [°C]
Citra (CT)	12.5~13.5	10’~11’	30/15
Fargo (ND)	13.5~15.6	12’	23/14
Palmira (PA)	12.0~12.1	11’~12’	29/19
Popayan (PO)	12.1~12.2	12’	22/16
Isabella (PR)	11.5~12.4	12’	29/19



I. Environmental factors
   a. Temperature, °C
      I. Mean, max, min, day, night, day-night-difference
   b. Day length, hr

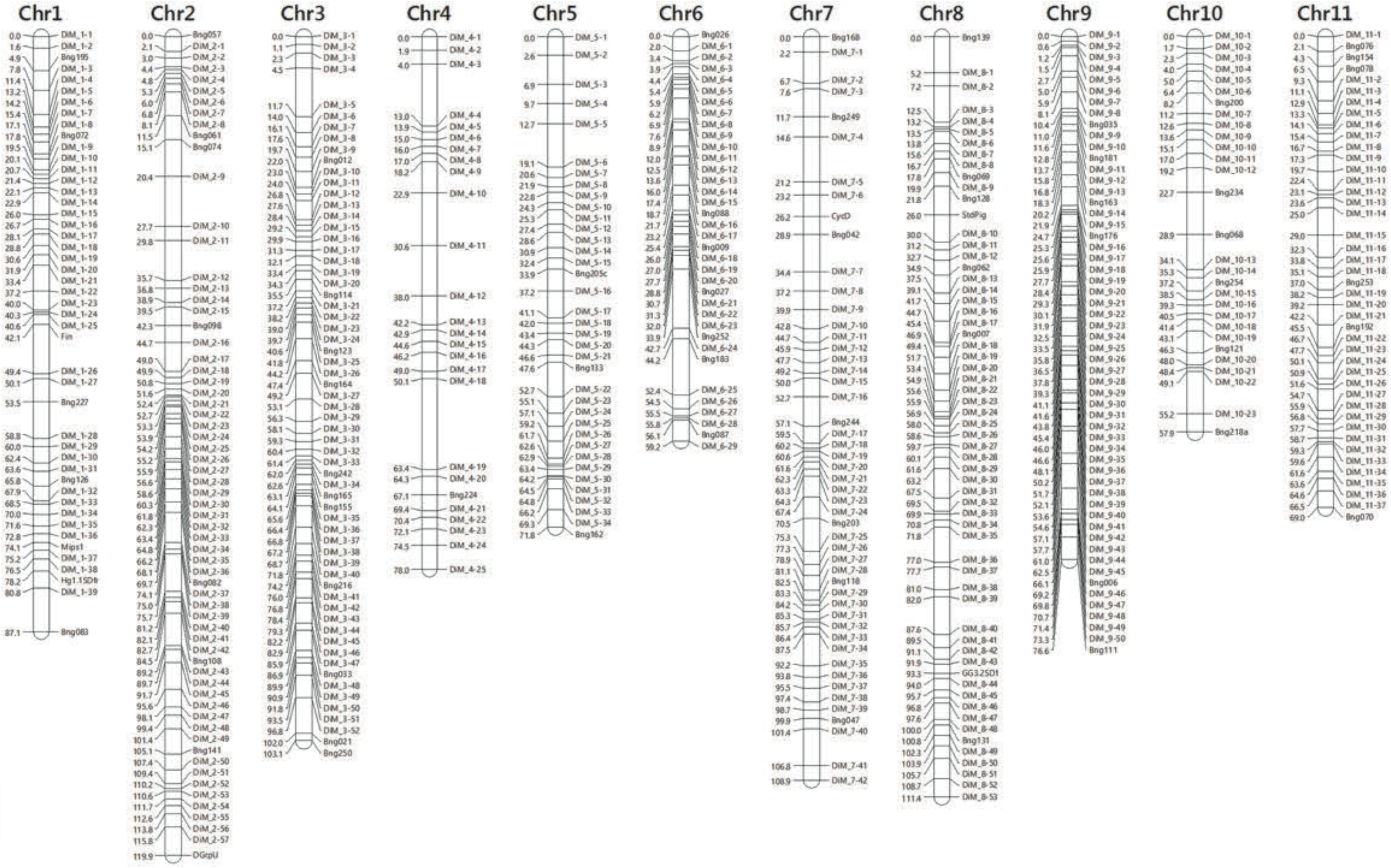


Data: Genetics

I. Linkage map (Bhakta et al. 2015)

a. **513** QTL markers

b. Average interlocus distance of 1.9 cM



QTL Marker$_i$

4	CAL	1
5	JAM	-1
6	RIJC001	-1
7	RIJC002	1
8	RIJC003	1
9	RIJC004	-1
10	RIJC005	1
11	RIJC006	1
12	RIJC007	1

Data: Model fitting & evaluation

I. Fitting
   a. 171 RILs grown across 5 sites
II. Evaluation
   a. 2 parents left out 16 RILs across 5 sites
   b. 2 parents 7 RILs planted in 2016 at Palmira, Colombia (PA16)
   c. 2 parents 106 RILs planted in 2016 at Citra, FL
      I. March sowing (CT16-1)
      II. May sowing (CT16-2)

Multi-stage phenology model

Phenology: Stages of bean development

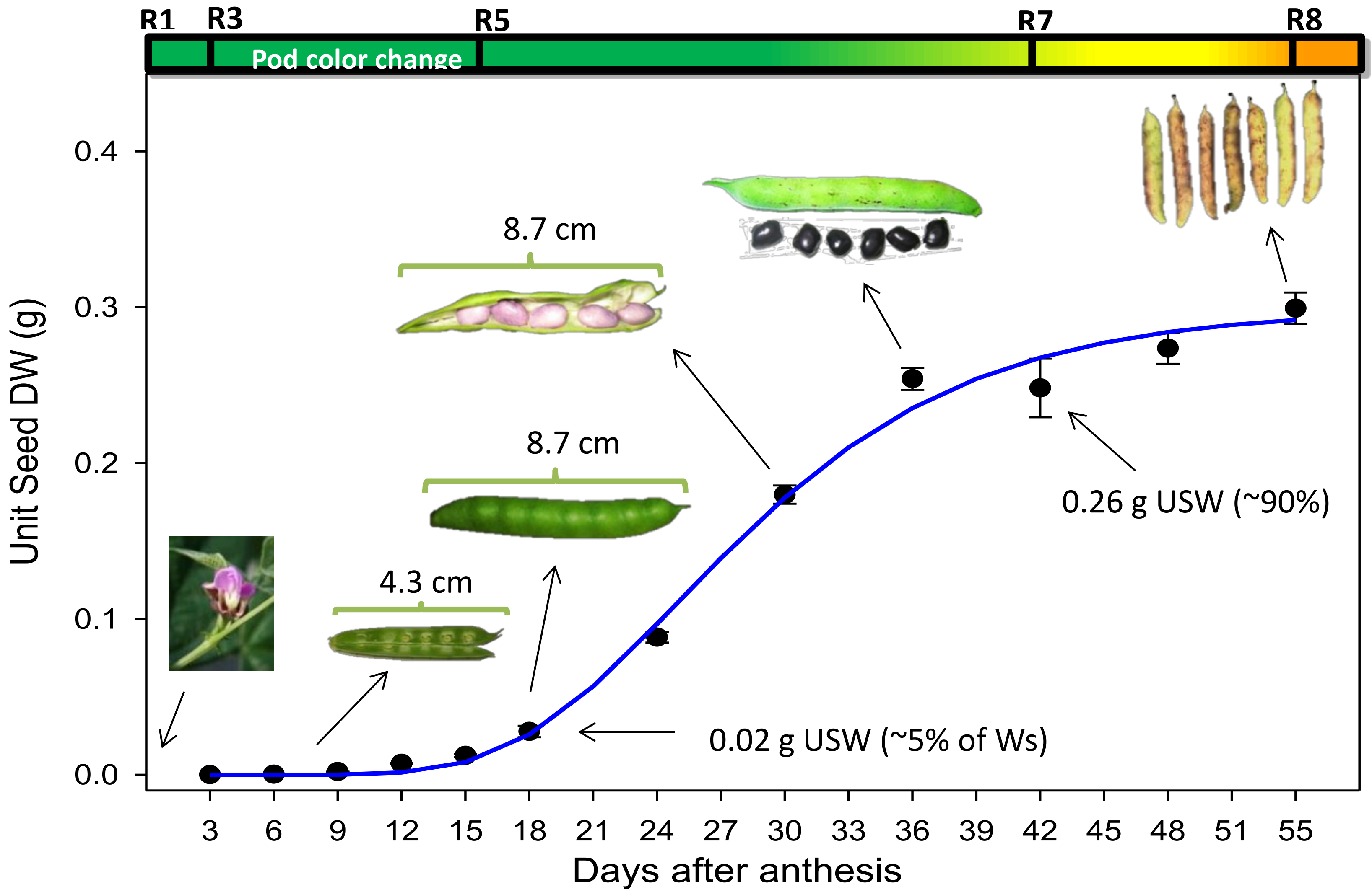
I. Planting

II. Emergence (VE)

III. Flowering (R1)

IV. Seed filling (R5)

V. Seed maturity (R7)



Clavijo Michelangeli et al. 2013

UF UNIVERSITY of FLORIDA

USDA USAID FEED THE FUTURE NSF

Phenology: Stages of bean development

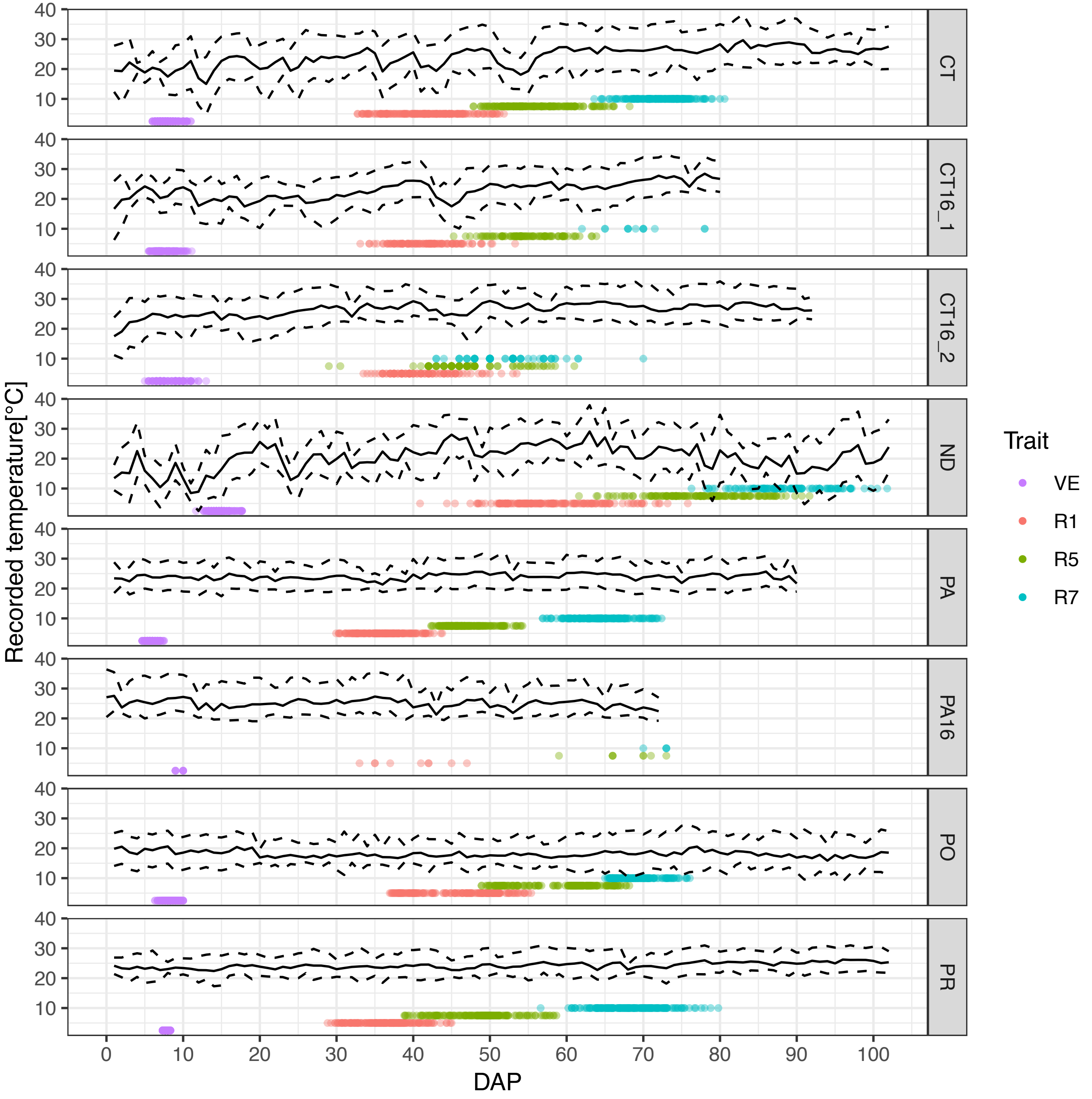
I. Planting

II. Emergence (VE)

III. Flowering (R1)

IV. Seed filling (R5)

V. Seed maturity (R7)

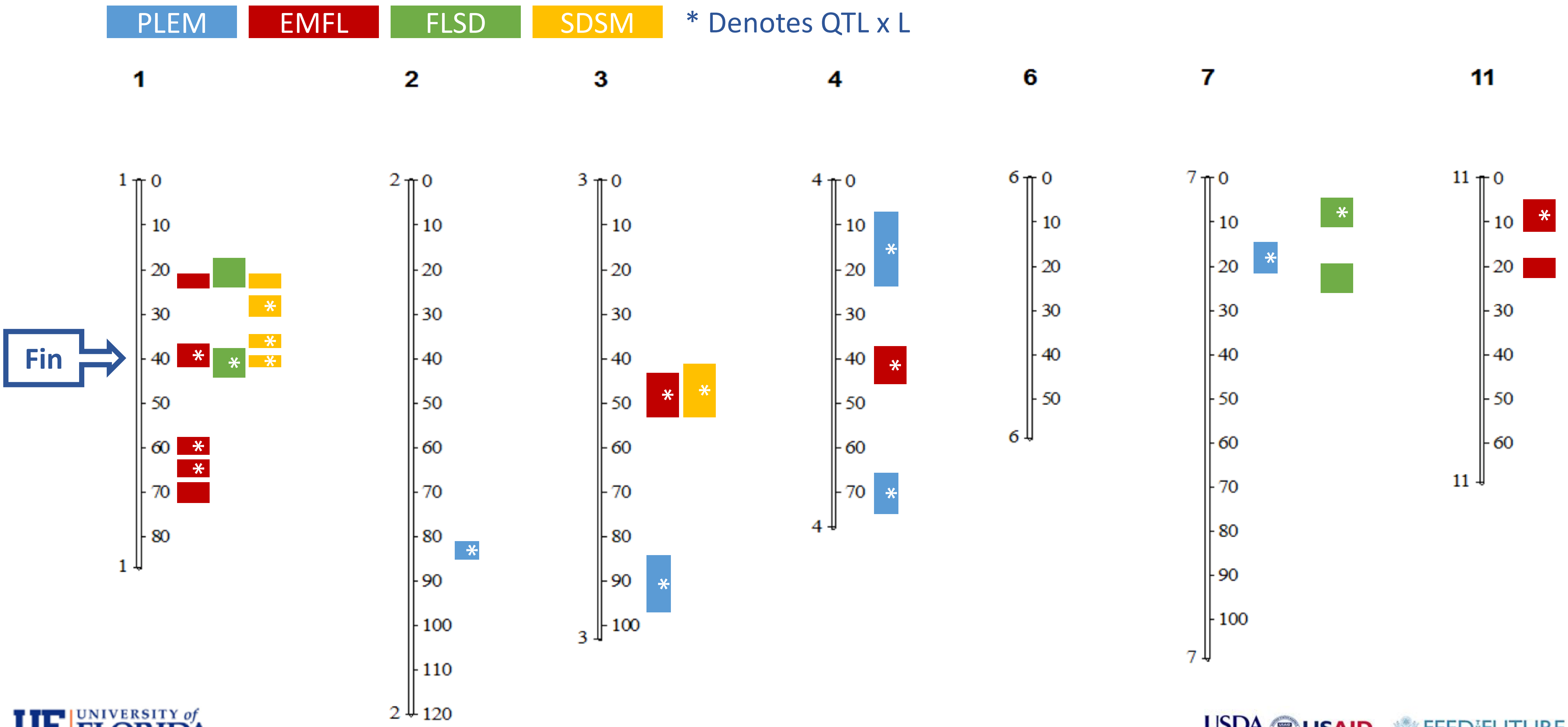


UF UNIVERSITY of FLORIDA

FEED THE FUTURE

NSF

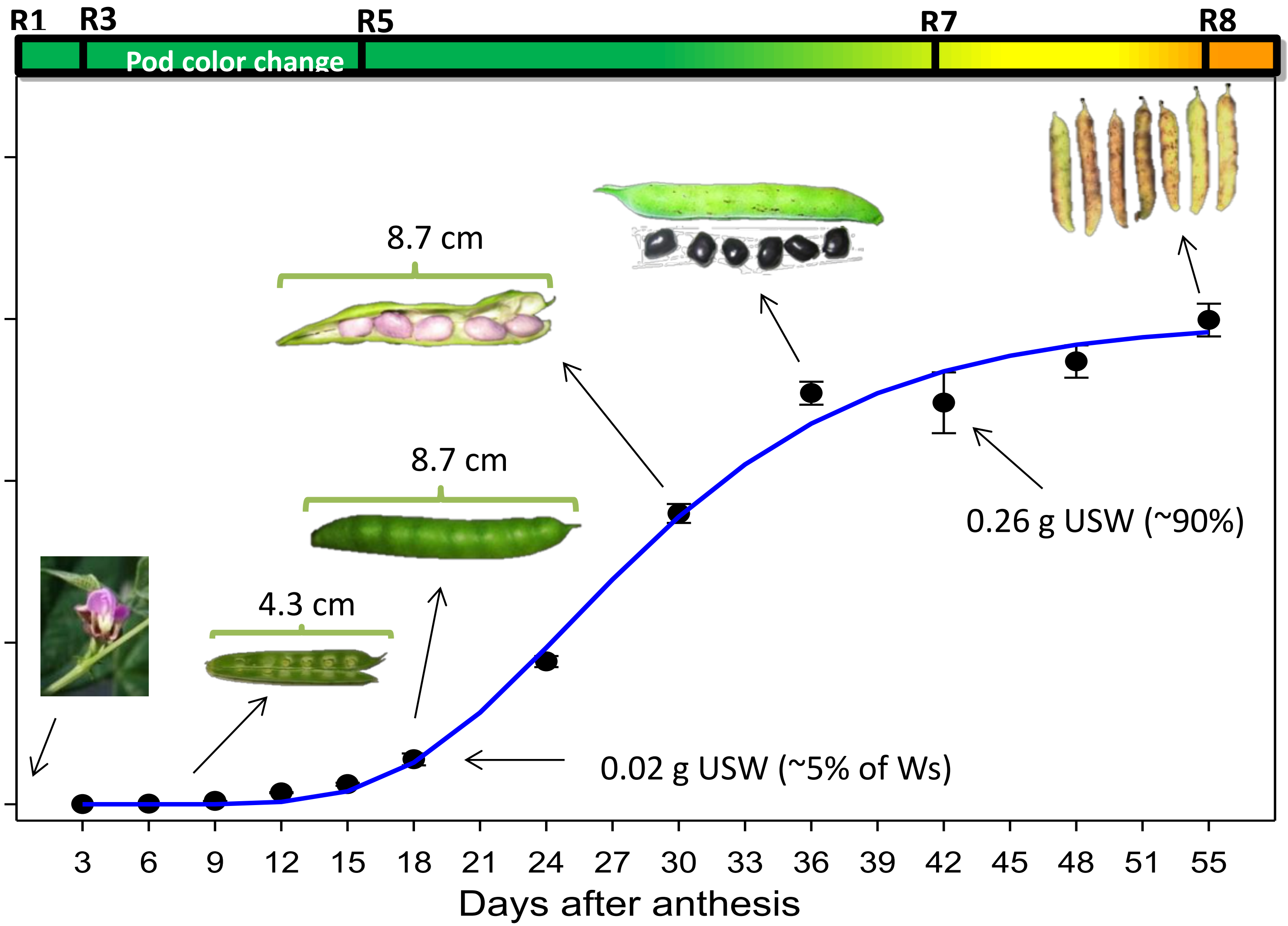
QTLs associated with phenological stages



Additive Nonlinear Rate Modules

Rate modules

I. Planting
   a. RPLEM = $(VEDAP)^{-1}$
II. Emergence (VE)
   a. REMFL = $(R1DAP - VEDAP)^{-1}$
III. Flowering (R1)
   a. RFLSD = $(R5DAP – R1DAP)^{-1}$
IV. Seed filling (R5)
   a. RSDSM = $(R7DAP – R5DAP)^{-1}$
V. Seed maturity (R7)



Clavijo Michelangeli et al. 2013

Rate modules

Accumulating daily rates of progress

$$\hat{Y}_{gls} = \sum_{d\,=\,DAP_{s-1}}^{\hat{Y}_{gls}\,\geq\,1} \max\left[0, rate_{gls}(d)\right]\Delta d$$

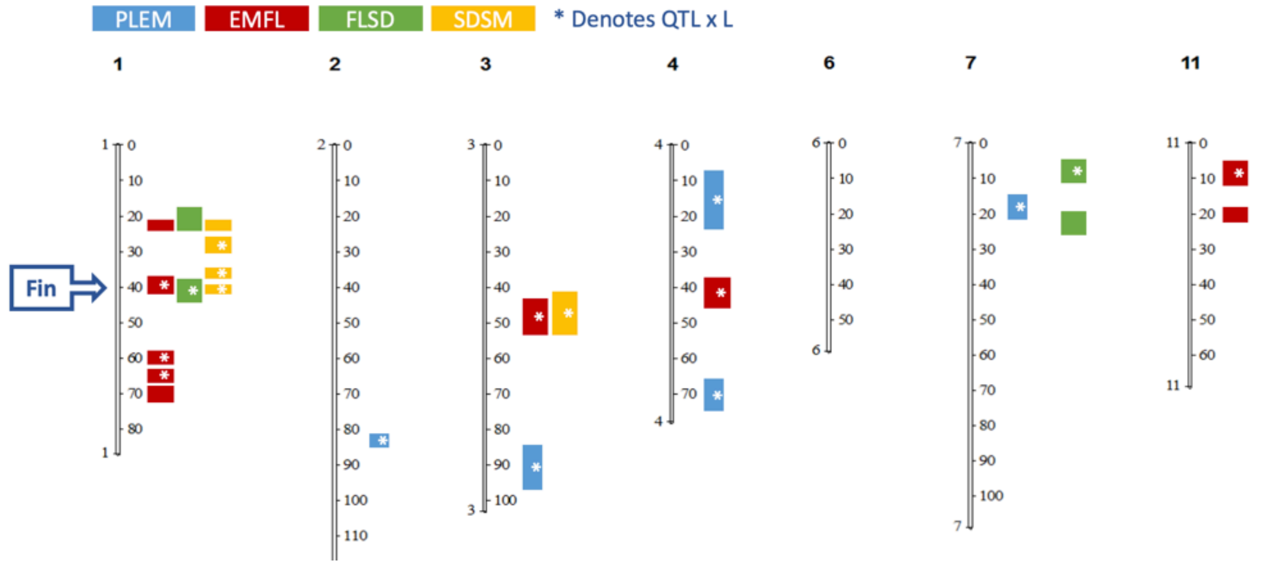
$$\begin{aligned} rate_{gls}(d) = \alpha_s + \textstyle\sum_h \left[TEmax_{s0} f_T\left(Temp_{lh},\ T_{peak_{s0}}\right)\Delta h\right] + PPSEN_{s0} f_{PP}(DL_{ld}, CSDL_s) + \\ \textstyle\sum_q \beta_{QTL_{sq}} QTL_{gq} + \\ \textstyle\sum_q \sum_h \left[TEmax_{sq} QTL_{gq} f_T\left(Temp_{lh},\ T_{peak_{sq}}\right)\Delta h\right] + \\ \textstyle\sum_q PPSEN_{sq} QTL_{gq} f_{PP}(DL_{ld}, CSDL_s) \end{aligned}$$

Rate modules

Computing daily rates of progress

$$\hat{Y}_{gls} = \sum_{d = DAP_{s-1}}^{\hat{Y}_{gls} \geq 1} \max[0, rate_{gls}(d)]\Delta d$$

$$rate_{gls}(d) = \alpha_s + \sum_h [TEmax_{s0} f_T(Temp_{lh}, T_{peak_{s0}})\Delta h] + PPSEN_{s0} f_{PP}(DL_{ld}, CSDL_s) + \sum_q \beta_{QTL_{sq}} QTL_{gq} + \sum_q \sum_h [TEmax_{sq} QTL_{gq} f_T(Temp_{lh}, T_{peak_{sq}})\Delta h] + \sum_q PPSEN_{sq} QTL_{gq} f_{PP}(DL_{ld}, CSDL_s)$$

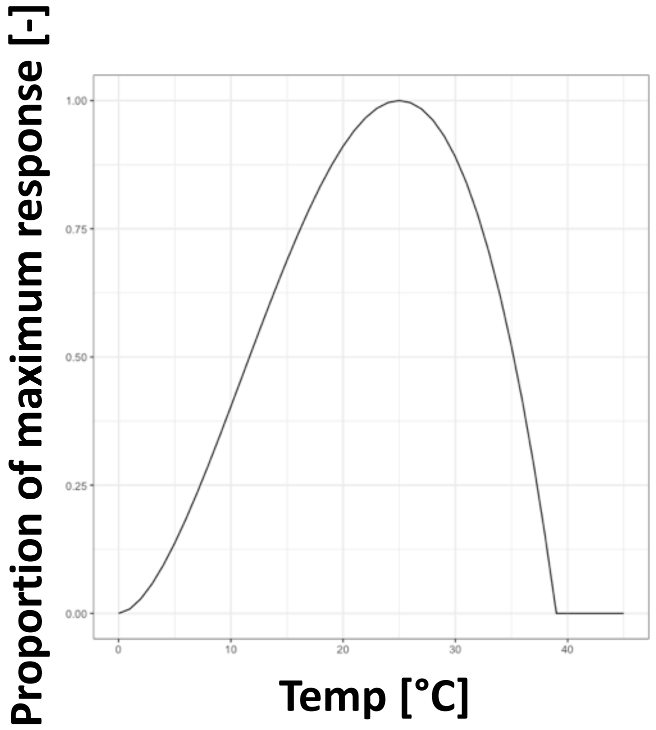


Rate modules

Computing daily rates of progress

$$\hat{Y}_{gls} = \sum_{d = DAP_{s-1}}^{\hat{Y}_{gls} \geq 1} \max[0, rate_{gls}(d)]\Delta d$$

$$rate_{gls}(d) = \alpha_s + \sum_h [TEmax_{s0} f_T(Temp_{lh}, T_{peak_{s0}})\Delta h] + PPSEN_{s0} f_{PP}(DL_{ld}, CSDL_s) + \sum_q \beta_{QTL_{sq}} QTL_{gq} + \sum_q \sum_h [TEmax_{sq} QTL_{gq} f_T(Temp_{lh}, T_{peak_{sq}})\Delta h] + \sum_q PPSEN_{sq} QTL_{gq} f_{PP}(DL_{ld}, CSDL_s)$$

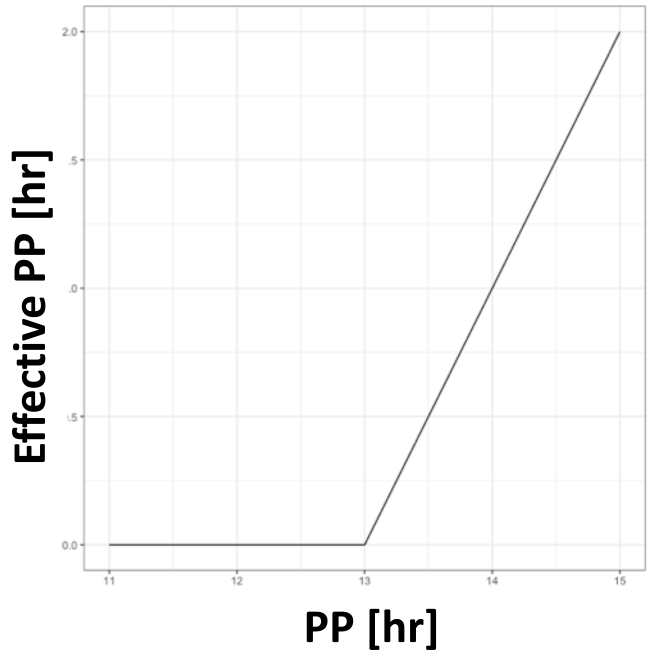


Rate modules

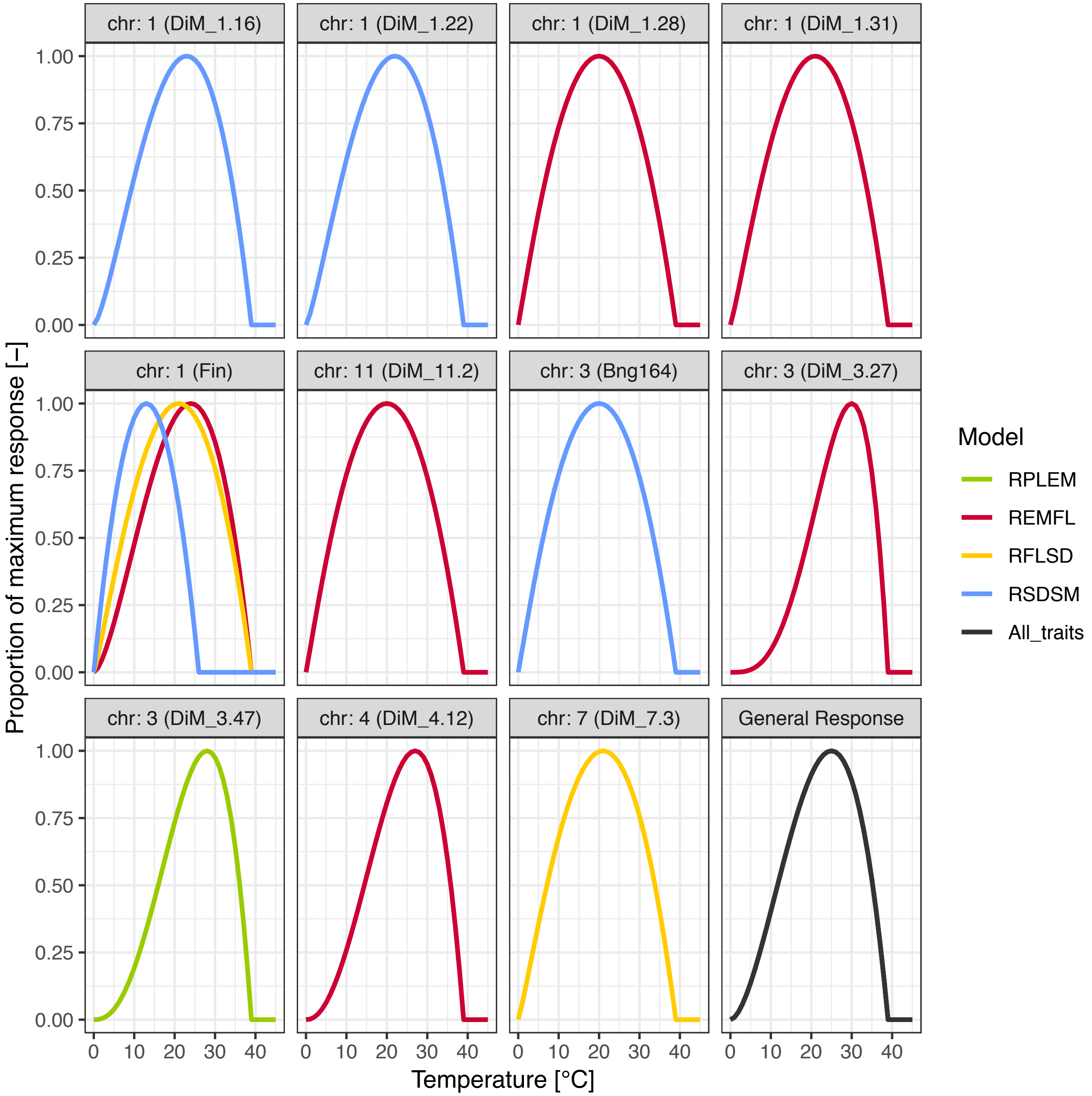
Computing daily rates of progress

$$\hat{Y}_{gls} = \sum_{d = DAP_{s-1}}^{\hat{Y}_{gls} \geq 1} \max[0, rate_{gls}(d)]\Delta d$$

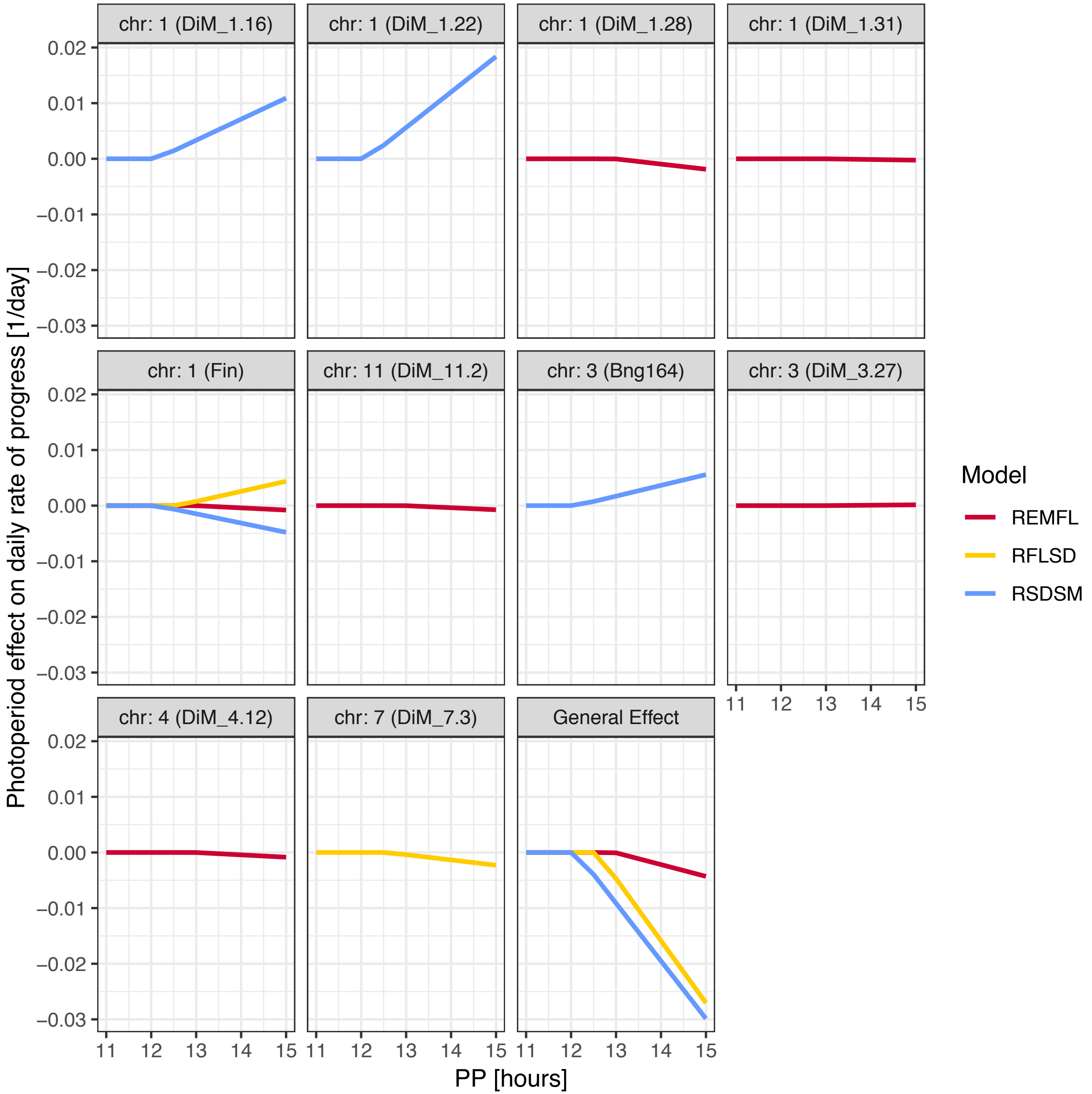
$$\begin{aligned} rate_{gls}(d) = \alpha_s + \textstyle\sum_h \left[TEmax_{s0} f_T(Temp_{lh}, T_{peak_{s0}})\Delta h\right] + PPSEN_{s0} f_{PP}(DL_{ld}, CSDL_s) + \\ \textstyle\sum_q \beta_{QTL_{sq}} QTL_{gq} + \\ \textstyle\sum_q \sum_h \left[TEmax_{sq} QTL_{gq} f_T\left(Temp_{lh}, T_{peak_{sq}}\right)\Delta h\right] + \\ \textstyle\sum_q PPSEN_{sq} QTL_{gq} f_{PP}(DL_{ld}, CSDL_s) \end{aligned}$$



Temperature response profiles



PP effects for *Calima*



UF UNIVERSITY of FLORIDA

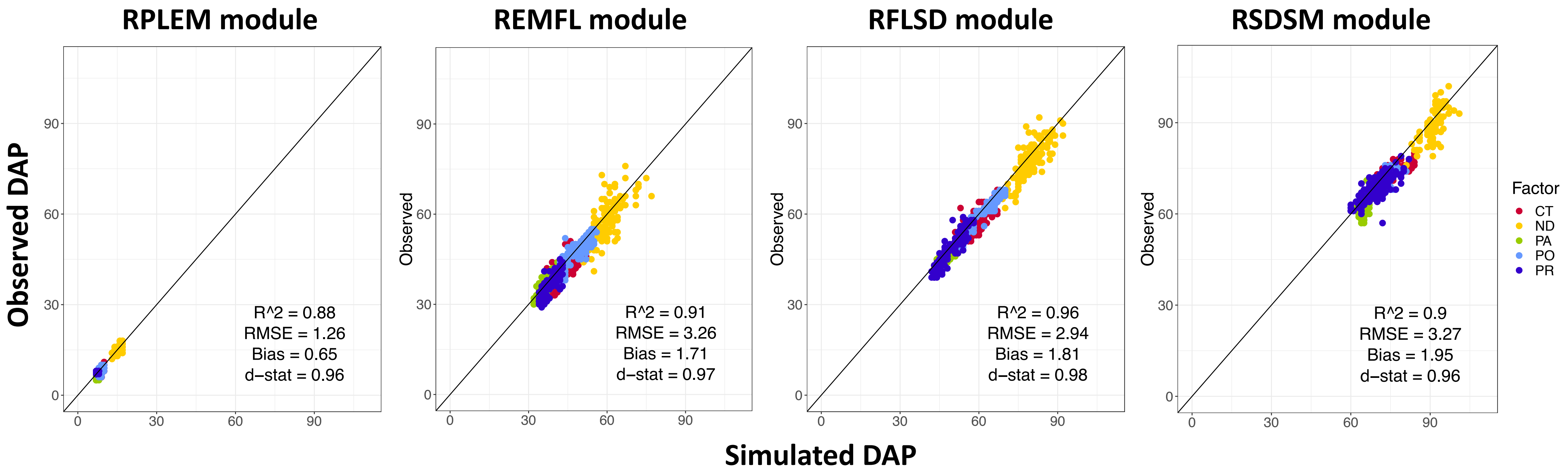
USAID FROM THE AMERICAN PEOPLE

FEED THE FUTURE The U.S. Government's Global Hunger & Food Security Initiative

NSF

Fitted trait module performance

* Each module is initialize at observed DAP of previous stage

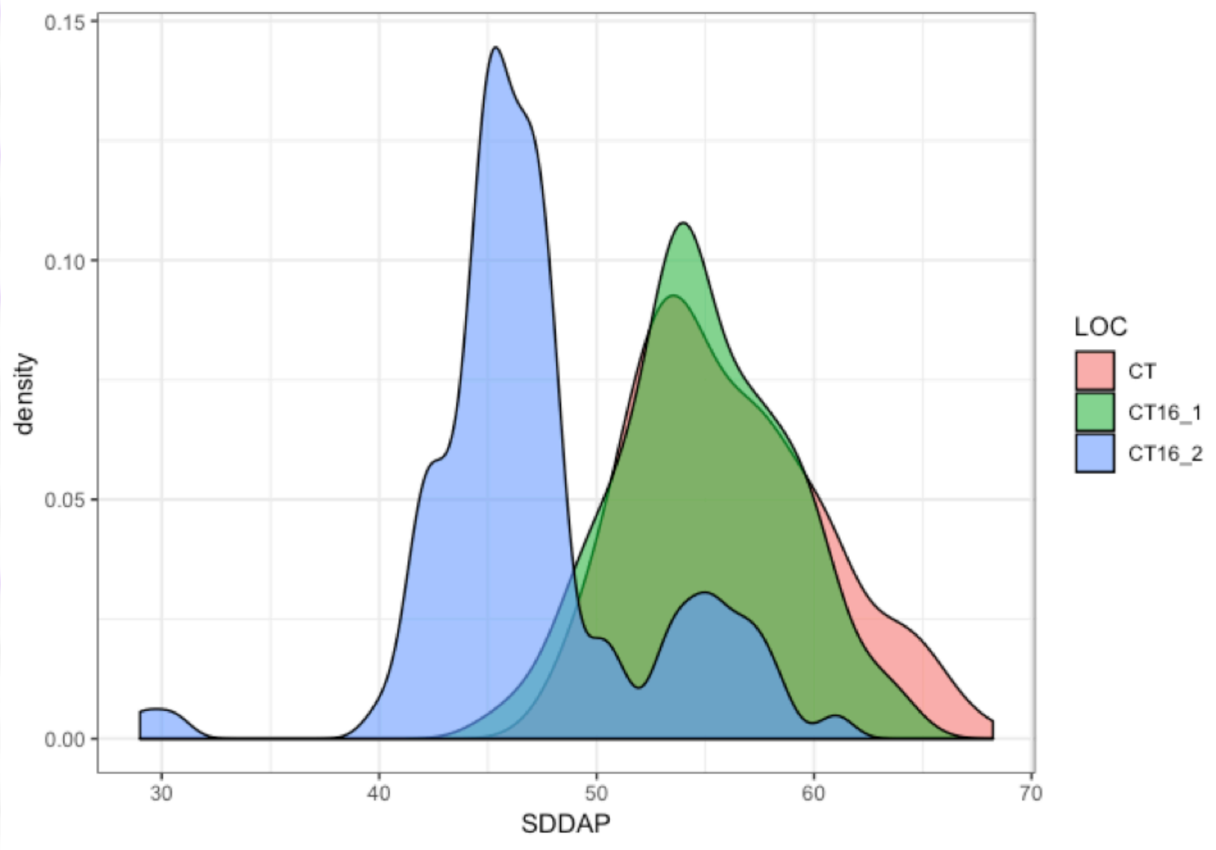


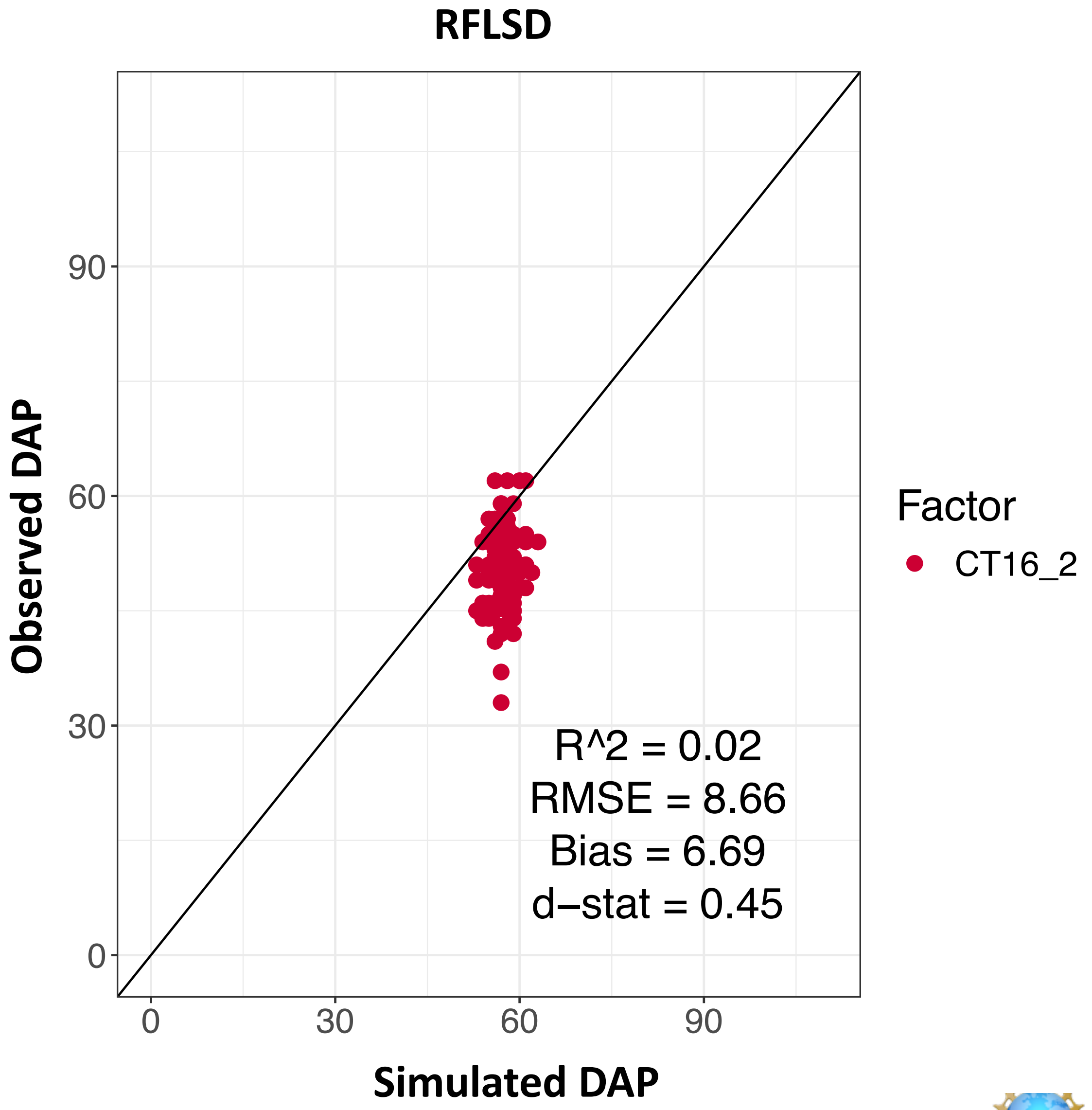
Phenology model evaluation

* Each module is initialize at simulated DAP of previous stage

Site	Stage	R^2	RMSE	Site	Stage	R^2	RMSE
NSF	VE	0.89	1.06	CT16-1	VE	0.01	1.46
	R1	0.89	3.32		R1	0.19	3.91
	R5	0.90	3.96		R5	0.02	5.02
	R7	0.85	4.19		R7	0.95	9.06
PA16	VE	0.10	1.8	CT16-2	VE	0.05	1.64
	R1	0.60	3.51		R1	0.33	3.46
	R5	--			R5	0.02	8.66
	R7	--			R7	0.01	15.45

Phenology model evaluation





Conclusions & What's next

I. Limitations
   a. Assumptions in temperature function (Tref & Toff)
      I. Limited G and G x E from training population and sites
   b. No soil-water-nutrient dynamics
   c. Not a mixed effect model (inflated SSE)

II. Improved breeding tool
   a. Dynamical simulation
   b. Non-linear G x E responses
   c. Simulate component traits

III. Next steps
   a. Implementation in DSSAT CROPGRO
   b. Extension to other reproductive development traits

Acknowledgments

UF | UNIVERSITY of FLORIDA

C. Eduardo Vallejos
James W. Jones
Kenneth J. Boote
Melanie J. Correll
Salvador Gezan
Gerrit Hoogenboom
Alina Zare
Daniel Wallach (visiting scholar)
Jose A. Clavijo Michelangeli
Mehul Bhakta
Li Zhang
Jugpreet Singh
Melissa Carvalho
Subodh Acharya
Vy Nguyen
Ashley Francois
Curtis Waddell
Tara Bongiovani

CIAT

Steve Beebe
Idupulapti Rao
Juamer Ricaurte

NDSU North Dakota State University

Juan Osorno
Raphael Colbert

UNIVERSITY OF PUERTO RICO UPR

Jim Beaver
Elvin Roman
Abiezer Gonzalez

PENNSTATE 1855

Rongling Wu
Yanquan Wang
Ningtao Wang

Stony Brook University

Wei Hou

NSF

#IOS-0923975

USDA

National Needs Fellowship
#2015-38420-23701
NIFA Hatch
#FLAABE005454

NORMAN BORLAUG FOUNDATION

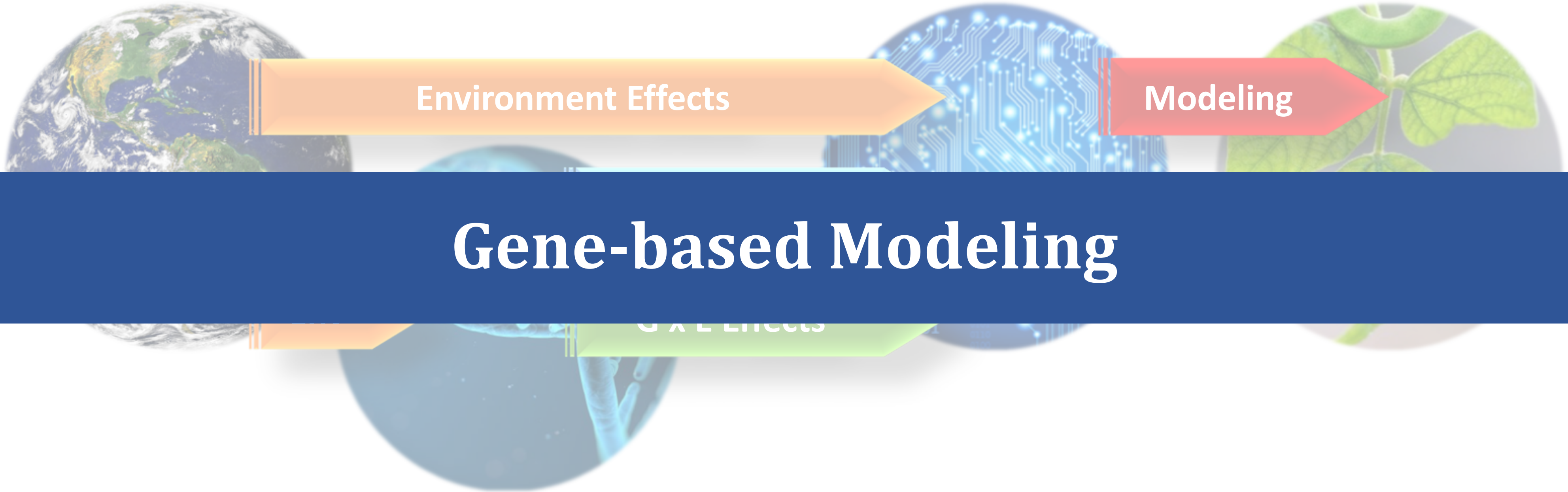
Sponsored
by
Purdue University

UF | IFAS UNIVERSITY of FLORIDA

Climate Change Seed Grant 2016

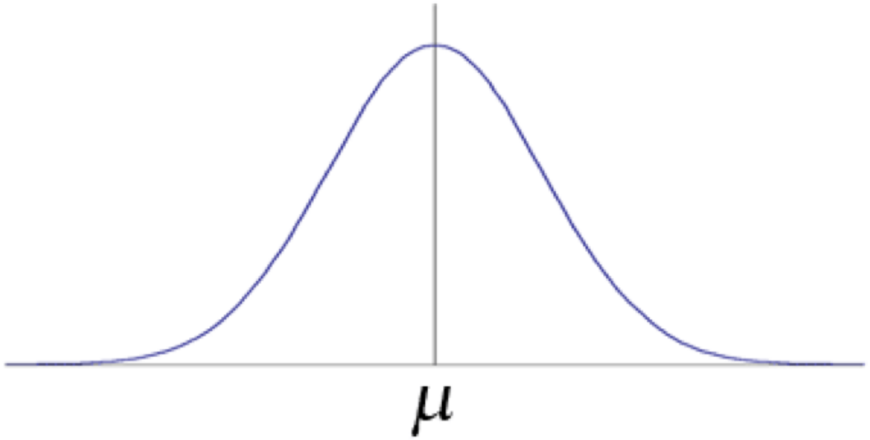
Questions & Suggestions

Gene-based Modeling



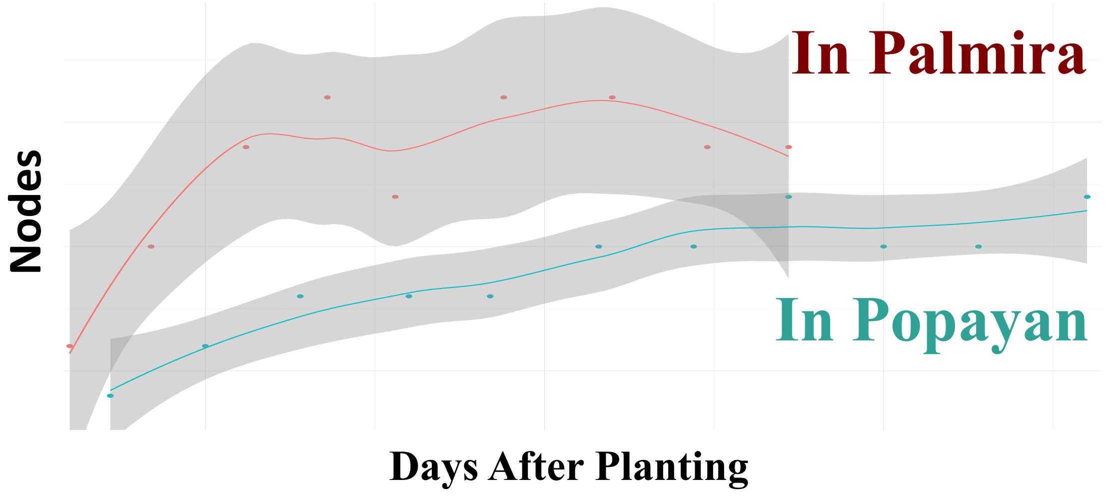
Gene-based modeling

Phenotype Distribution



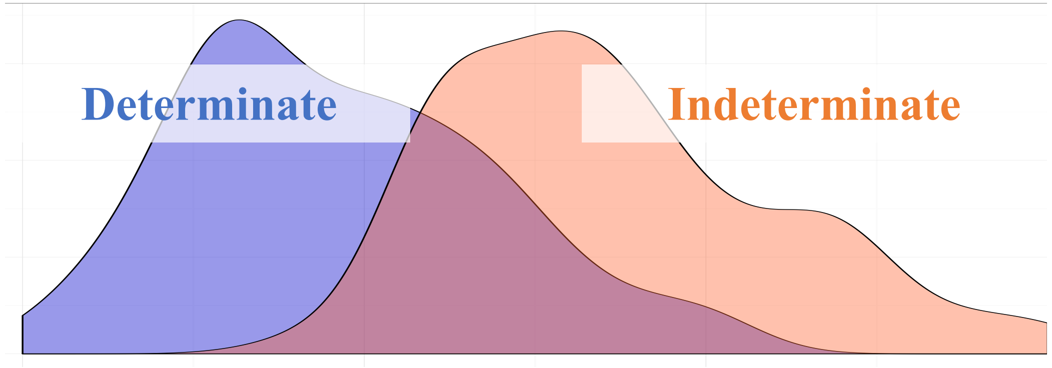
Gene-based modeling

Node number for a single variety (*Calima*)



Gene-based modeling

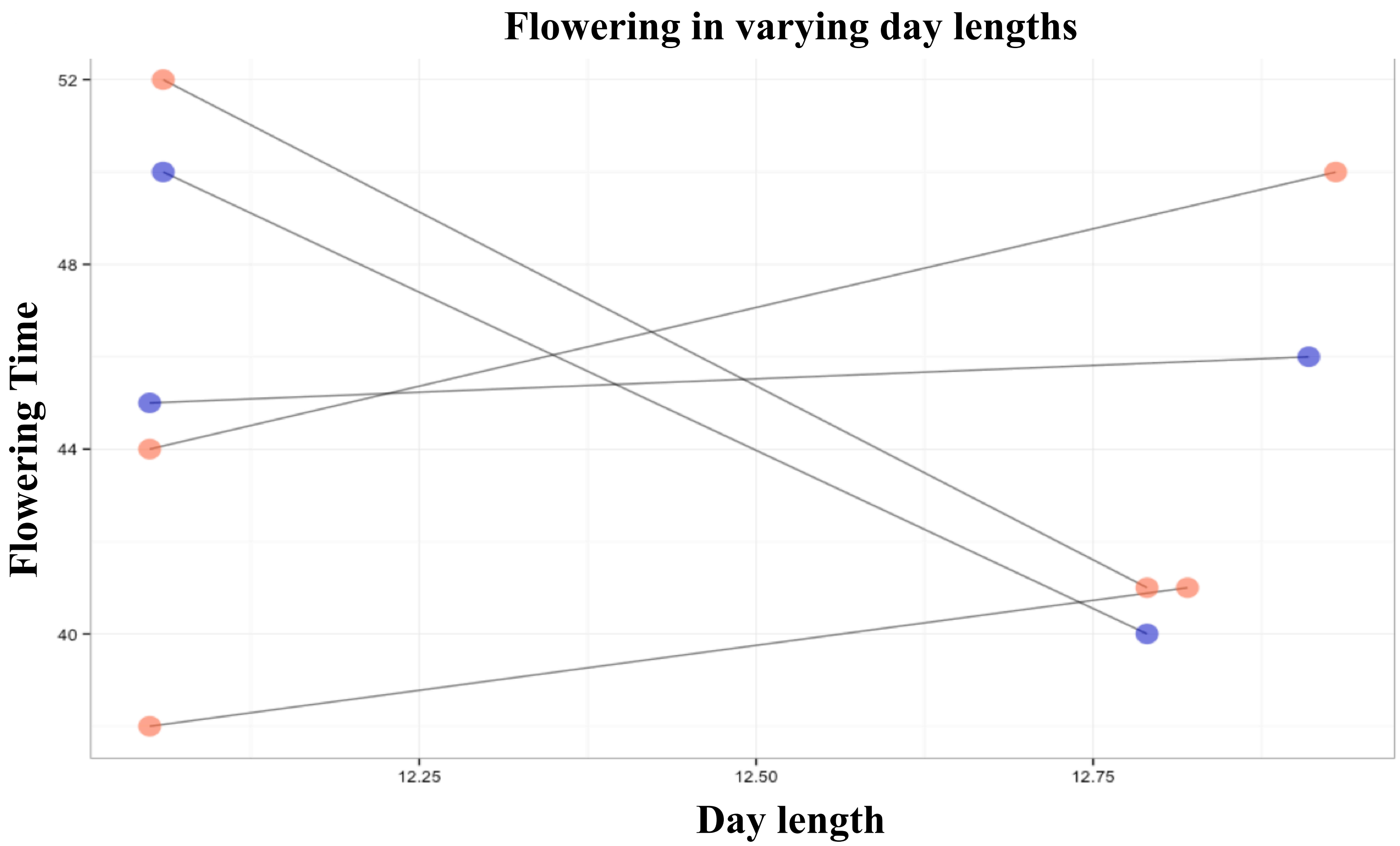
Flowering time in Palmira



Days After Planting

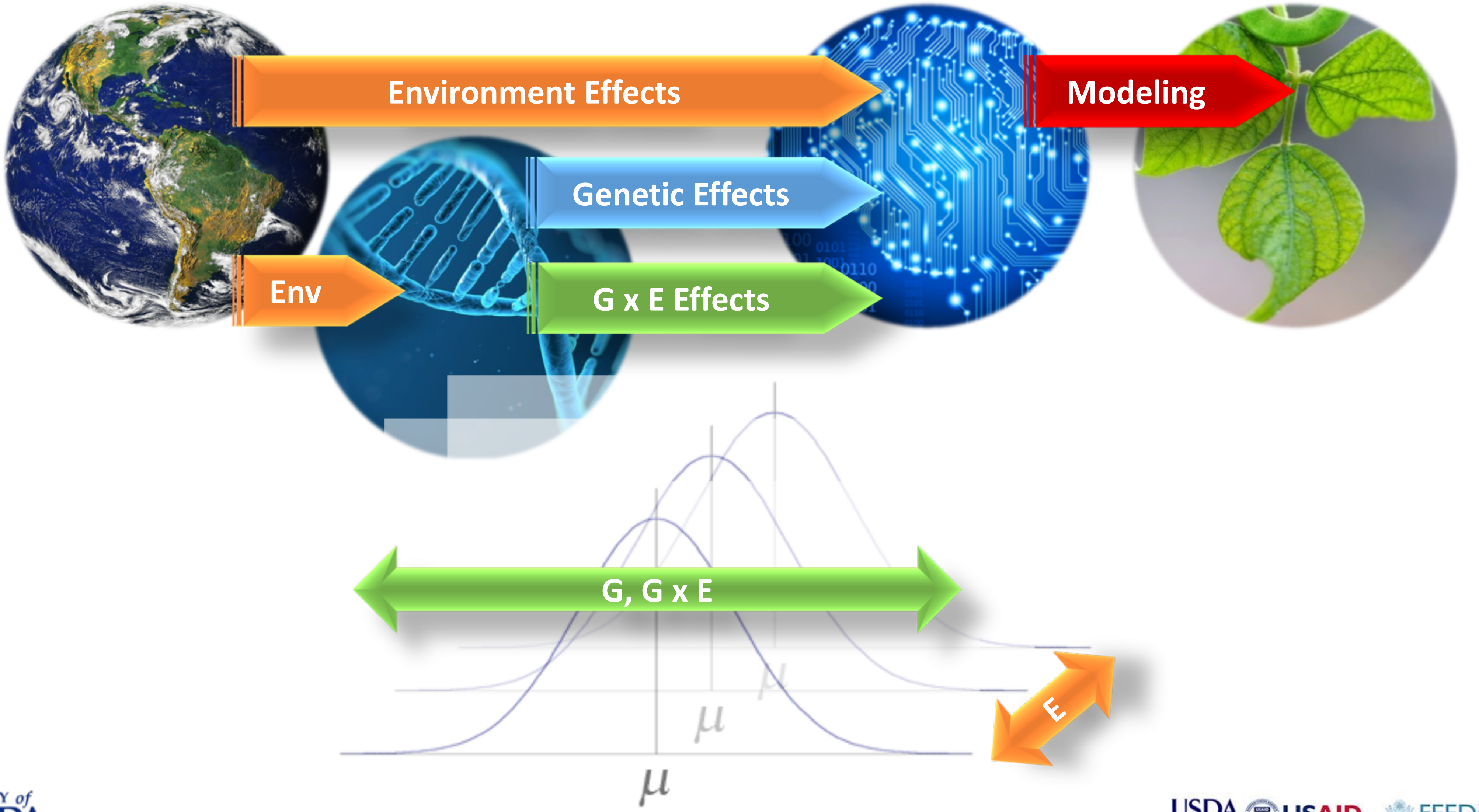
Gene-based modeling

Gene-based modeling



- Determinat
- Indeterminate

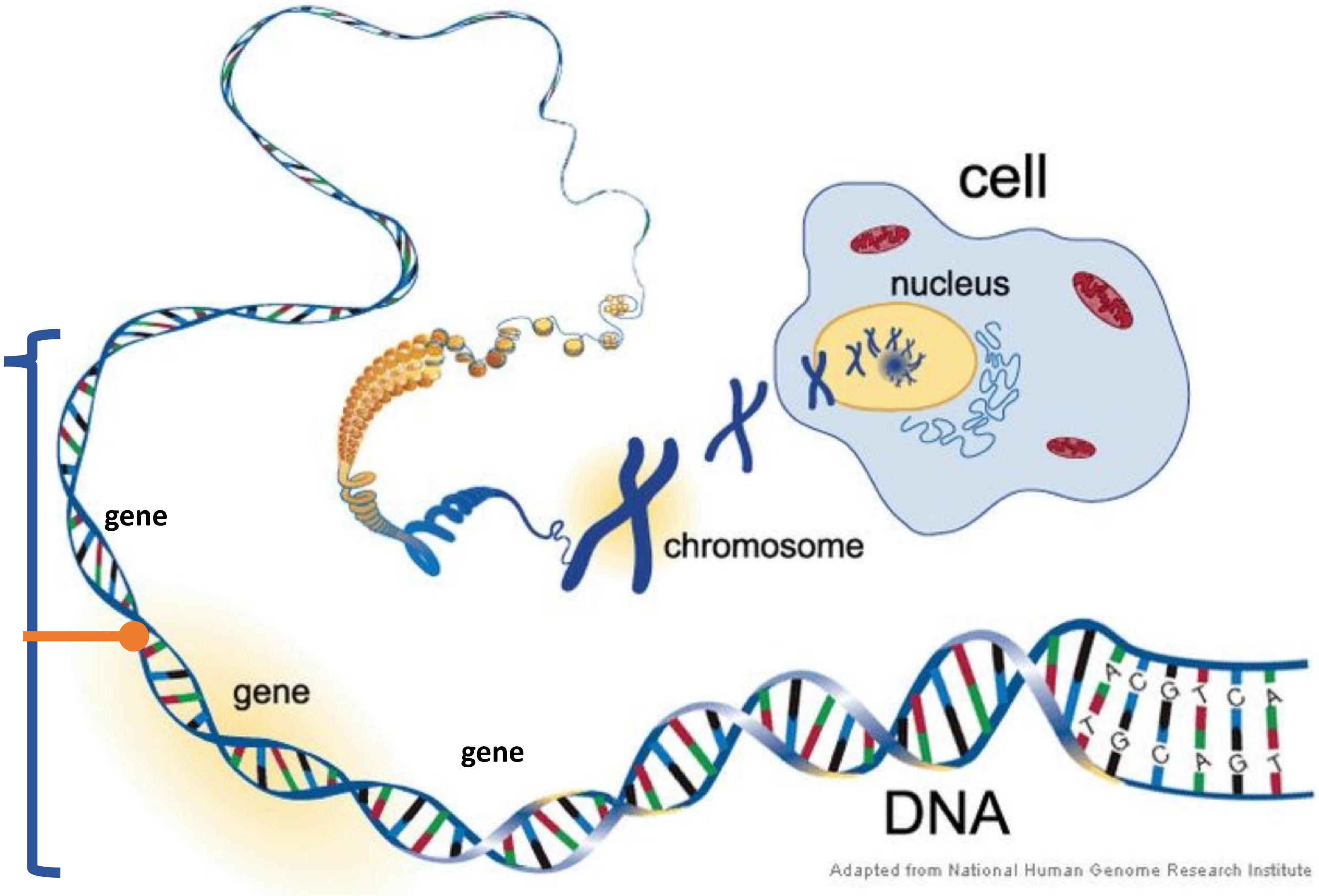
Gene-based modeling



Gene-based modeling: Genetic markers

Quantitative Trait Locus (QTL)

Single Nucleotide Polymorphism (SNP)



Gene-based modeling: Linear regression models

Plant Breeding

Crop Growth Models

Linear mixed-effect Model with G x S

$$\hat{y}_{ij} = \mu + \beta_S S_j + \sum_{q=1}^{Q} \beta_\alpha X_{iq} + \sum_{q=1}^{Q} \beta_{QTLxS} X_{iq} S_j$$

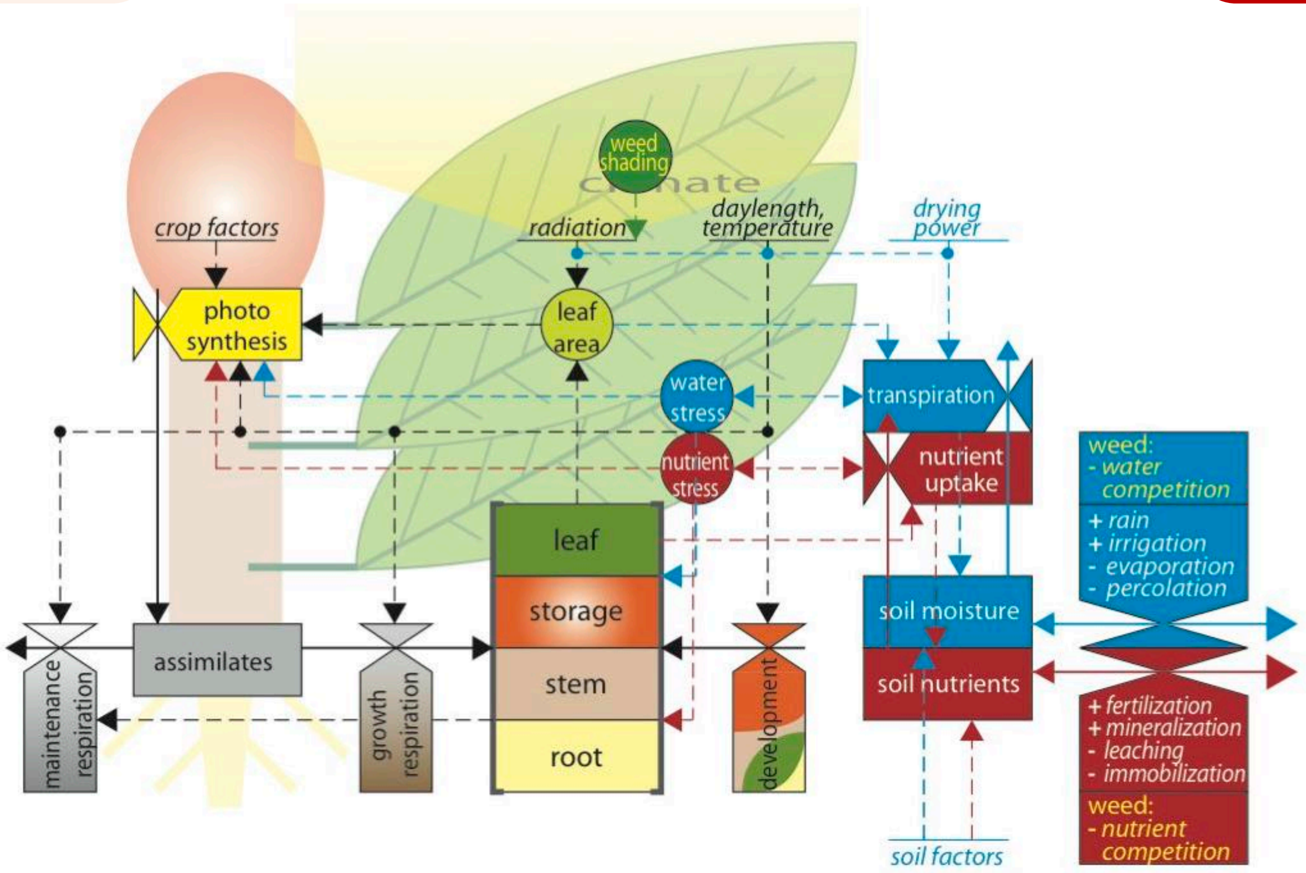
- Predicts static traits
- Predicts complex traits
- Limited application in training locations

Boer et al. 2007

Gene-based modeling: Crop growth models

Plant Breeding

Crop Growth Models



Source: M. van Ittersum

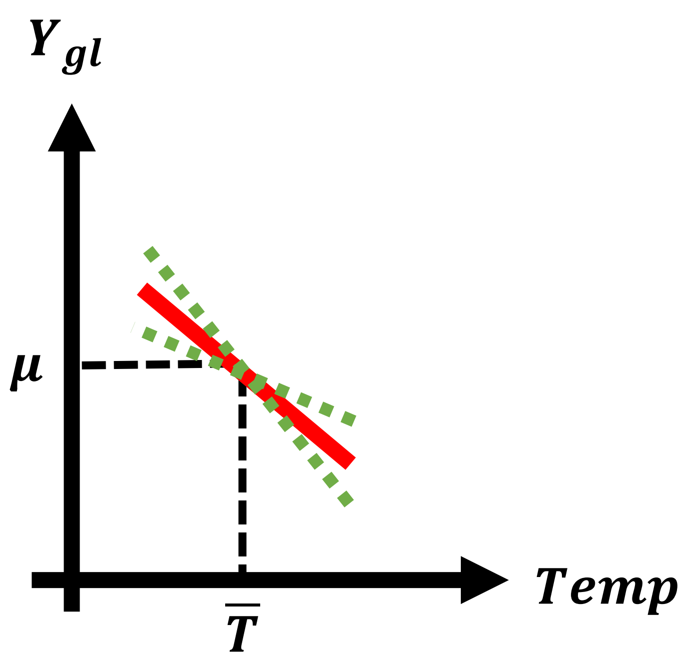
WAGENINGEN UNIVERSITY
WAGENINGEN UR

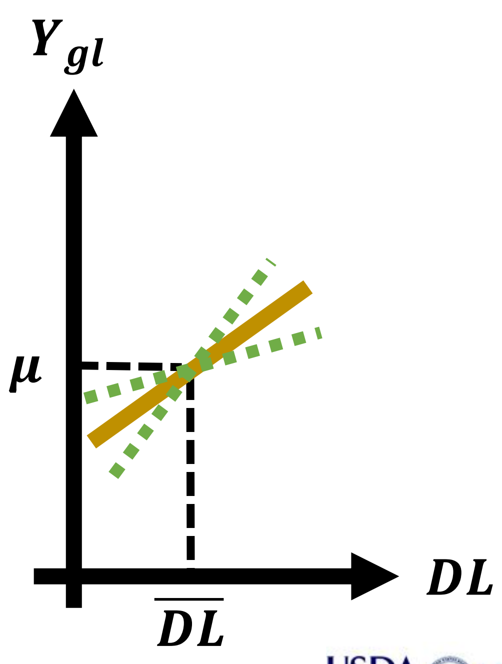
Model conversion

Base linear mixed effect model

$$\hat{Y}_{gl} = \mu + \beta_T \Delta T_{gl} + \beta_{DL} \Delta DL_{gl} + \sum_{q=1}^{Q} \beta_{QTL_q} QTL_{gq} + \sum_{q=1}^{Q} \left(\beta_{QTL_q xT} QTL_{gq} \Delta T_{gl} + \beta_{QTL_q xDL} QTL_{gq} \Delta DL_{gl} \right)$$

$$\Delta T_{gl} = \left(T_{gl} - \overline{T} \right) \quad \Delta DL_{gl} = \left(DL_{gl} - \overline{DL} \right)$$



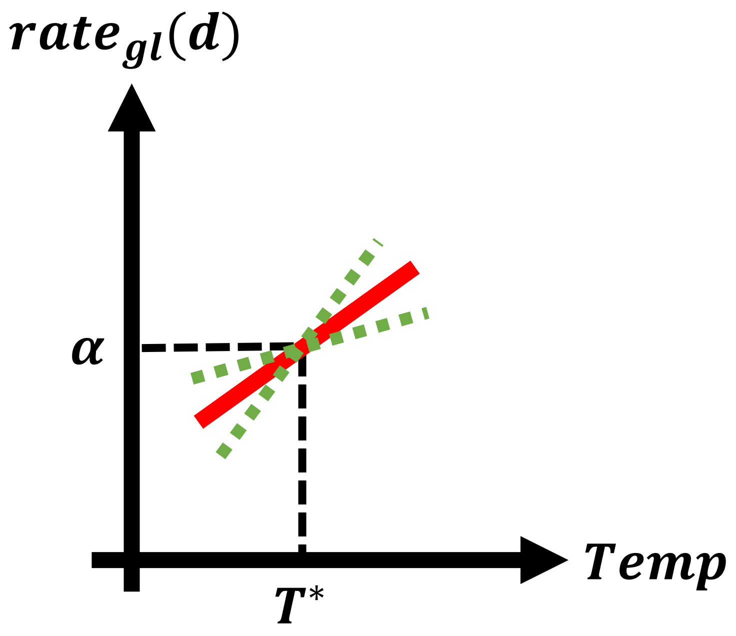


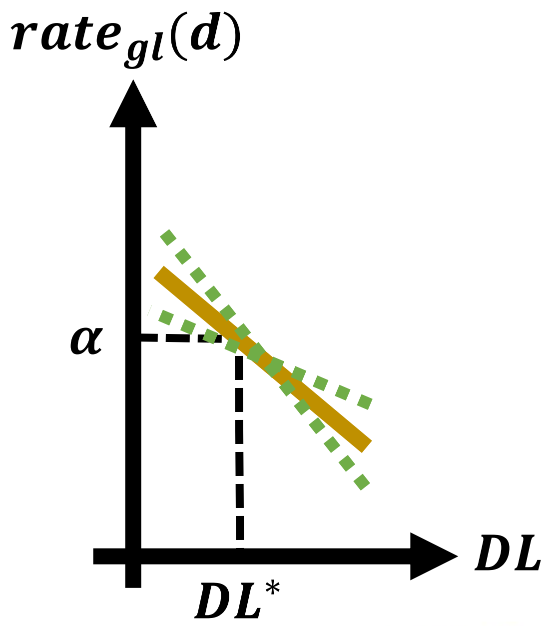
Model conversion

Convert to dynamic model

$$rate_{gl}(d) = \alpha + \beta_T \Delta T_{gl}(d) + \beta_{DL} \Delta DL_{gl}(d) + \sum_{q=1}^{Q} \beta_{QTL_q} QTL_{gq} + \sum_{q=1}^{Q} \left(\beta_{QTL_q xT} QTL_{gq} \Delta T_{gl}(d) + \beta_{QTL_q xDL} QTL_{gq} \Delta DL_{gl}(d) \right)$$

$$\Delta T_{gl}(d) = \left(T_{gl} - T^*\right) \quad \Delta DL_{gl}(d) = \left(DL_{gl} - DL^*\right)$$





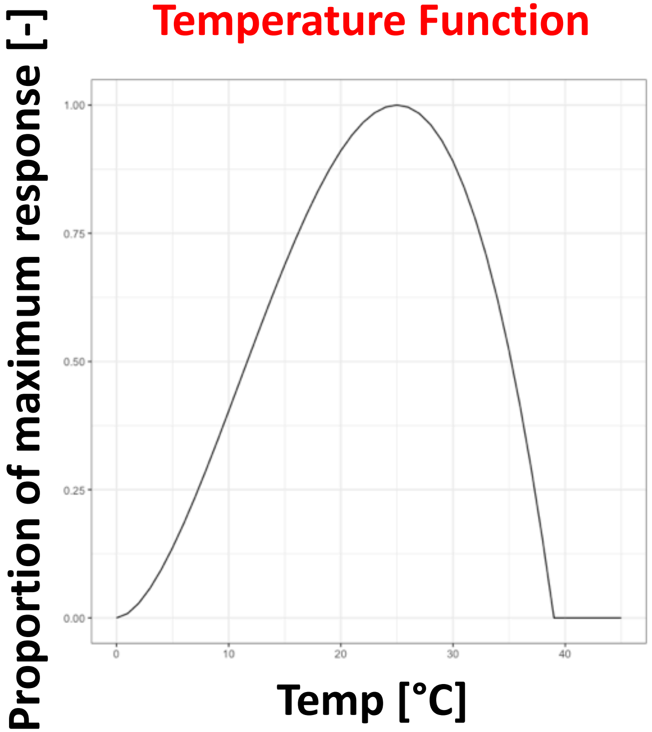
Model conversion

Non-linear E response functions

$$rate_{gl}(d) = \alpha + \beta_T \Delta T_{gl0}(d) + \beta_{DL} \Delta DL_{gl}(d) + \sum_{q=1}^{Q} \beta_{QTL_q} QTL_{gq} + \sum_{q=1}^{Q} \left(\beta_{QTL_q xT} QTL_{gq} \Delta T_{glq}(d) + \beta_{QTL_q xDL} QTL_{gq} \Delta DL_{gl}(d) \right)$$

$$\Delta T_{glq}(d) = \sum_{h=1}^{24} f_T\left(Temp_{d,h}, Tpeak_q\right) \quad \Delta DL_{gl}(d) = f_{pp}(d, CSDL)$$

Temperature Function



Photoperiod Function