Project Sponsor

USDA

United States Department of Agriculture
National Institute of Food and Agriculture

This project is funded by the USDA National Institute of Food and Agriculture. Project number 2009-01929.

Project Participants

Colorado State University

NDSU

University of Nebraska Lincoln

USDA ars Agricultural Research Service
www.ars.usda.gov

Michigan State University

Oregon State University OSU

UC DAVIS

BeanCAP Project Objectives

1. Marker development and application

a. SNP platform
b. Association mapping of agronomic traits

2. Nutritional analysis

a. Phenotypic characterization
b. Association mapping of nutritional traits

3. Phaseolus Genes database development

a. Breeder-based
b. Central access point for bean improvement genetics

4. Early plant breeder training

a. High school
b. First two years of college

5. Educational multimedia development

a. Plant breeding mini-documentaries
b. Animations of physiological processes

2010
57 Active Participants

Phillip McClean, North Dakota State University (NDSU), PD; **Juan Osorno**, NDSU, Co-PD, education lead, plant breeding; **Julie Garden-Robinson**, NDSU, Co-PD, extension; **Michelle Grant**, NDSU, Administrative assistant; **Christina Johnson**, NDSU, Artistic lead; **Shane Reetz**, NDSU, Documentary lead; **Lindsey Duppong**, NDSU, Infrographics artist; **Samira Mafi Moghaddam**, NDSU, Marker development; **Rian Lee**, NDSU, Marker development; **Sujan Mamidi**, NDSU, Statistical analysis; **Alexander Johnson**, NDSU, Undergraduate plant breeding intern; **Mariah Smith**, NDSU, High school plant breeding intern; **Austen Lund**, NDSU, High school plant breeding intern; **Angela Linares**, NDSU, Graduate student, intern training/seed dispersal; **Albert J. Vander Wal**, NDSU, Research technician, seed dispersal; **Carlos Urrea**, University of Nebraska, Lincoln (UNL), Co-PD; **Doug Valade**, UNL, High school summer intern student; **Misty Griffitts**, UNL, Undergraduate summer intern student; **Nicole Schnittger**, UNL, Undergraduate summer intern student; **Tania Torres**, UNL, Undergraduate school term intern student; **Emily Hoehn**, UNL, Undergraduate school term intern student; **Michael Grusak**, USDA/Houston, Co-PD; **David Dworak**, USDA/Houston, Research technician; **Stephanie Mercado**, USDA/Houston, Research technician; **Jim Myers**, Oregon State University; Co-PD, Nutritional analysis, plant breeding; **Deborah Kean**, OSU, Faculty research assistant, field (retired); **Annie Chozinski**, OSU, Faculty research assistant, field; **Joel Davis**, OSU, Faculty research assistant, lab nutrition analysis; **Michelle Bullock**, OSU, Student field worker; **Paul Gepts**, University of California, Davis (UCD), Co-PD, data base development; **Shelby Repinski**, UCD,Graduate student, QTL entry to database; **Adriana Gomez**, UCD, Graduate student, QTL entry to database; **Dawei Lin**, UCD, Bioinformatics and database lead; **Jose Boveda**, UCE, Database/web programmer; **Joe Fass**, UCD, Lead programmer; **Nikhil Joshi**, UCD, Bioinformatics programmer; **Monica Britton**, UCD, Bioinformatics analyst; **Jim Kelly**, Michigan State University (MSU), Plant breeding, education; **Jacob Emmendorfer**, MSU, High school plant breeding intern student; **Damien Johnson**, MSU, High school plant breeding intern student; **Philip Munoz**, MSU, Undergraduate plant breeding intern student; **Mary Harris**, MSU, Undergraduate plant breeding intern student; **Phil Miklas**, USDA/Prosser, Co-PD; Field increase of core population; **Perry Cregan**, USDA/Beltsville,Marker development and screening; **David Hyten**, USDA/Beltsville, Marker development and screening; **Edward Fickus**, USDA/Beltsville, Marker development technician; **Ken Kmecik, Seminis**, Greenhouse/field increase of core population; **Mark Brick**, Colorado State University (CSU), Plant breeding, education, nutrition analysis; **Henry Thompson**, CSU, Nutrition analysis; **Sarah Dominick**, CSU, Undergraduate plant breeding intern; **Hannah Walters**, CSU, Undergraduate plant breeding intern; **Bryan Fisher**, CSU ,High school plant breeding intern; **Colton Heeney**, CSU, High school plant breeding intern; **Jordon Leone**, CSU, High school plant breeding intern; **Leslie Brick**, CSU, Research associate, nutrition analysis; **Dimas Echeveria Moreno**, CSU, Research associate, nutrition analysis.

USDA
United States Department of Agriculture
National Institute of Food and Agriculture

2011
85 Active Participants

North Dakota State Univ: **Phillip McClean**, PD; **Juan Osorno**, Co-PD, education lead, breeding; **Julie Garden-Robinson**, Co-PD, extension; **Michelle Grant**, Administrative assistant**; Bradley Bisek**, intern; **Nicole Dallman**, intern; **Kataryna Cookman**, intern; **Mitchell Bauske**, intern; **Lyndsie Park**, intern; **Peter Totten**, intern; **Christina Johnson**, Artistic lead; **Shane Reetz**, Documentary lead; **Bree Malingnen**, Infrographics artist; **Samira Mafi Moghaddam**, Marker development; **Rian Lee**, Marker development; **Sujan Mamidi**, Statistical analysis; **Stacy Halvorson**, extension associate; **Leah Whigham**, nutrition researcher; **Deb Habedank**, childcare director; **Todd Weinmann**, extension agent; **Steve Sagaser**, extension agent; **Chelsea Langus**, intern; **Alexandra Idso**, intern; **Aimee Henning**, intern, **Kendra Otto**, intern; **Emily Westrom**, intern; **Amy Hutchinson**, intern; **Kayla Bahtiraj**, intern. ***Univ Nebraska, Lincoln***: **Carlos Urrea**, , Co-PD; **Nicole Schnitger**, intern; **Misty Griffitts**, intern; **Scout Wilson**, intern; **Charity Berkey**, intern; **Danielle Becker**, intern; ***USDA/Houston***: **Michael Grusak**, Co-PD, nutritional analysis; **Paz Etcheverry**, cooperator; **David Dworak**, technician; **Lori Center**, technician; **William Carter**, intern; ***Oregon State Univ***: **Jim Myers**,; Co-PD, Nutritional analysis, breeding; **Annie Chozinski**, faculty research assistant; **Kara Young**, intern; **Katrina Maguelli**, intern. ***Univ California, Davis***: **Paul Gepts**, Co-PD, data base development; **Shelby Repinski**, Graduate student, QTL entry to database; **Adriana Navarro Gomez**, Graduate student, QTL entry to database; **Sun Lei**, Graduate student; **Tania Gioia**, Graduate Student; **Dawei Lin**, Bioinformatics and database lead; **Jose Boveda**, Database/web programmer; **Joe Fass**, Lead programmer; **Nikhil Joshi**, Bioinformatics programmer; **Monica Britton**, Bioinformatics analyst; **Zhi-Wei Lu**, Bioinformatics analyst. ***Michigan State Univ***: **Jim Kelly**, Co-PD, Breeding, education; **Evan Wright**, technician; **Amy Lasley**, MSU, graduate student; **Valerio Hoyos Villegas**, graduate student; **Rosa Castanon**, intern; **Brittany Lane**, intern. ***USDA/Prosser***: **Phil Miklas**, Co-PD, Field increase of core population; **Susan Swanson**, technician; **Jennifer Trapp**, technician; **Jeff Coulson**, technician. ***USDA/Beltsville***: **Perry Cregan**, Co-PD, Marker development and screening; **David Hyten**, Marker development and screening; **Edward Fickus**, Marker development technician; **Qijiain Song**, Bioinformatics analysis; **Gaofeng Jia**, marker and bioinformatics analysis; **Josaine Rodriques**, Federal University of Vicosa, Brazil, SSR analysis, graduate; **Charles Quigley**, research DNA sequencing. ***Seminis***: **Ken Kmecik**, Greenhouse/field increase of core population. ***Colorado State Univ***: **Mark Brick**, Breeding, education, nutrition analysis; **Henry Thompson**, Nutrition analysis; **Soni Hueftle**, intern; **Griffin Carpenter**, intern; **Keera Brown**, intern; **Alyssa Bollig**, intern; **Dimas Echeveria Moreno**, Research associate, nutrition analysis. ***USDA/East Lansing***: **Karen Cichy**, Co-PD, Nutritional analysis; **Nicole Butler**, graduate student; ***USDA/Mayquez, PR***: **Tim Porch**, Co-PD, Field stress analysis; **Abraham Montes**, technician; **Franquie Colon**, research assistant; **Gregory Howard**, research; **Edlin Gonzalez**, research.

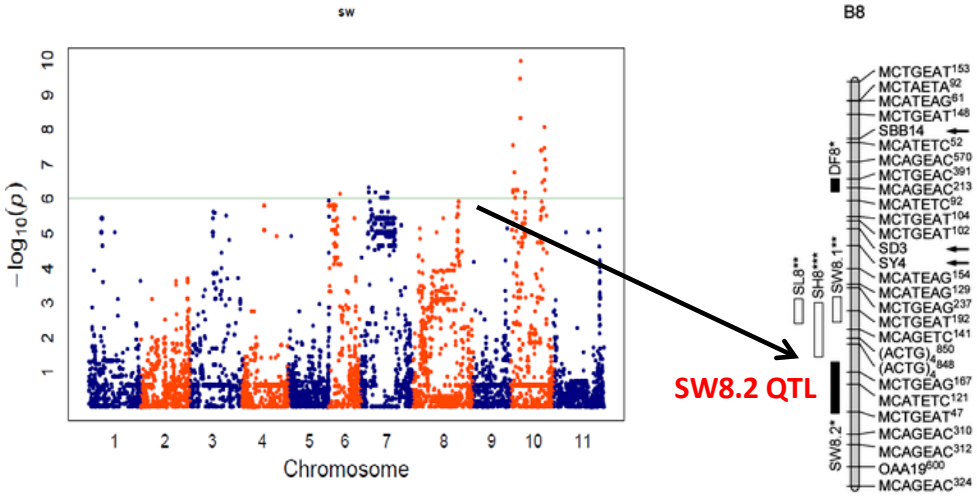
2012
54 Active Participants

North Dakota State University*: Phil McClean**, PD; **Juan Osorno**, Co-PD; **Julie Garden-Robinson**, Co-PD; **Rian Lee**, Marker development organizer; **Shane Reetz**, Documentary lead; **Bree Reetz**, Education video development; **Samira Mafi Moghaddam**, Marker development, Association mapping; **Sujan Mamidi**, Statistical analysis, Association mapping; **Christina Johnson**, Animation and other multimedia development; **Emily Driessen**, Undergraduate, Marker analysis; **Erin Sullivan**, Undergraduate, Marker analysis; **Ryan Lenz**, Undergraduate, Marker analysis; **Lucia Smith**, undergraduate plant breeding intern; **Alex Kallmeyer**, undergraduate plant breeding intern; **Casey Kjera**, Undergraduate nutrition intern; **Abby Plucker**, Undergraduate nutrition intern; **Stephanie Anderson**, Undergraduate nutrition intern; **Johanna Christenson**, Undergraduate nutrition intern; **Brooke Nell**, Undergraduate nutrition intern; **Nicole Seaburg**, Undergraduate nutrition intern; **Kimberly Beauchamp**, Graduate student nutrition education. ***USDA/Houston*: Michael Grusak**, Co-PD; **William Carter**, Micronutrient analysis; **Jenna Emerick**, Micronutrient analysis; **Rida Khan**, Micronutrient analysis. ***University of California/Davis*: Paul Gepts**, Co-PD; **Sarah Kuzay**, Database development; **Paige Hamilton-Conaty**, Database development; **Dawei Lin**, Bioinformatics and database lead; **Jose Boveda**, Database/web programmer; **Joe Fass**, Lead programmer; **Nikhil Joshi**, Bioinformatics programmer;. ***University of Nebraska, Lincoln: **Carlos Urrea**, Co-PD, ***Eastern Wyoming College*: Skye Martin**, undergraduate breeding intern. ***Western Nebraska Community College*: Nathan Marquez**, undergraduate breeding intern, **Scout Wilson**, undergraduate breeding intern. ***Colorado State University*: Mark Brick,** Co-PD; **Henry Thompson**, Co-PD; **Donny Hodgkinson**, Undergraduate intern; **Emily Troxell**, Undergraduate intern; **Nathan Pohl**, High school intern. ***Michigan State University*: Jim Kelly**, Co-PD; **Cynthia Amstutz**, Undergraduate intern; **Lucas Costanza**, Undergraduate intern; **Mary Harris**, Undergraduate intern; Yusong Mu. ***USDA/Beltsville: Perry Cregan***, Co-PD, SNP development; **Qijian Song,** Research Geneticist; **Gaofeng Jia,** Visiting Scientist; **Charles Quigley,** Support Scientist. ***Oregon State University:* Jim Myers, Co-PD,** nutrition and snap bean analysis; **Joel Davis**, faculty research assistant; **Christina Hagerty,** graduate research assistant; **Ceely Will,** under graduate research. ***USDA/East Lansing, MI***: **Karen Cichy**, nutritional analysis.

USDA
United States Department of Agriculture
National Institute of Food and Agriculture

GWAS – Seed Weight

Association mapping result



n=252

markers = 10,287 SNPs

Model = EMMA

Perez-Vega et al. (2010)

Large trial

n=252; 10,827 SNPs;

4 locations

SNP	chr	Position_	P_value	maf
m4922	**8**	**43,827,895**	**3.51E-06**	**0.42**
m6145	8	44,237,412	2.97E-05	0.42
m6144	8	44,246,478	2.97E-05	0.42
m6143	8	44,270,482	2.97E-05	0.42
m6142	8	44,272,503	2.97E-05	0.42
m6141	8	44,273,042	2.97E-05	0.42
m6139	8	44,287,718	2.97E-05	0.42
m6174	8	44,338,667	4.64E-05	0.41
m6173	8	44,399,862	4.64E-05	0.41
m4856	8	44,425,316	4.64E-05	0.41
m4864	**8**	**44,470,617**	**8.35E-06**	**0.40**
m4863	**8**	**44,479,849**	**4.64E-05**	**0.41**
m4862	**8**	**44,480,227**	**8.35E-06**	**0.40**
m4861	**8**	**44,494,182**	**8.35E-06**	**0.40**
m4860	**8**	**44,526,077**	**8.35E-06**	**0.40**
m4859	**8**	**44,532,123**	**8.35E-06**	**0.40**
m4858	8	44,535,382	4.64E-05	0.41
m4029	8	44,605,263	1.73E-05	0.41
m4051	**8**	**44,712,020**	**2.81E-06**	**0.40**
m4050	**8**	**44,724,115**	**2.81E-06**	**0.40**
m4048	**8**	**44,754,521**	**2.81E-06**	**0.40**
m4044	**8**	**44,792,096**	**2.81E-06**	**0.40**
m4046	**8**	**44,804,652**	**2.81E-06**	**0.40**
m7104	**8**	**44,833,382**	**8.35E-06**	**0.40**
m7103	8	44,839,102	4.64E-05	0.41
m7102	8	44,842,481	4.64E-05	0.41
m7101	8	44,853,371	4.64E-05	0.41
m7100	**8**	**44,858,735**	**8.35E-06**	**0.40**
m7099	8	44,862,000	4.64E-05	0.41
m9586	8	44,891,515	1.45E-05	0.40
m4752	**8**	**44,921,533**	**8.35E-06**	**0.40**

USDA
United States Department of Agriculture
National Institute of Food and Agriculture

Leveraging the BeanCAP and Genome Sequencing Projects

Sequencing Project Leads

- *Scott Jackson, University of Georgia, PD*
- *Phil McClean, North Dakota State University, Co-PD*
- *Jeremy Schmutz, Hudson-Alpha, Co-PD*
- *Dan Rokshar, DOE-JGI, Co-PD*
- *Perry Cregan, USDA/Beltsville*

Funding

- USDA, AFRI
- DOE, JGI

USDA
United States Department of Agriculture
National Institute of Food and Agriculture

The Sequencing Project
Goals

Sequencing project goals

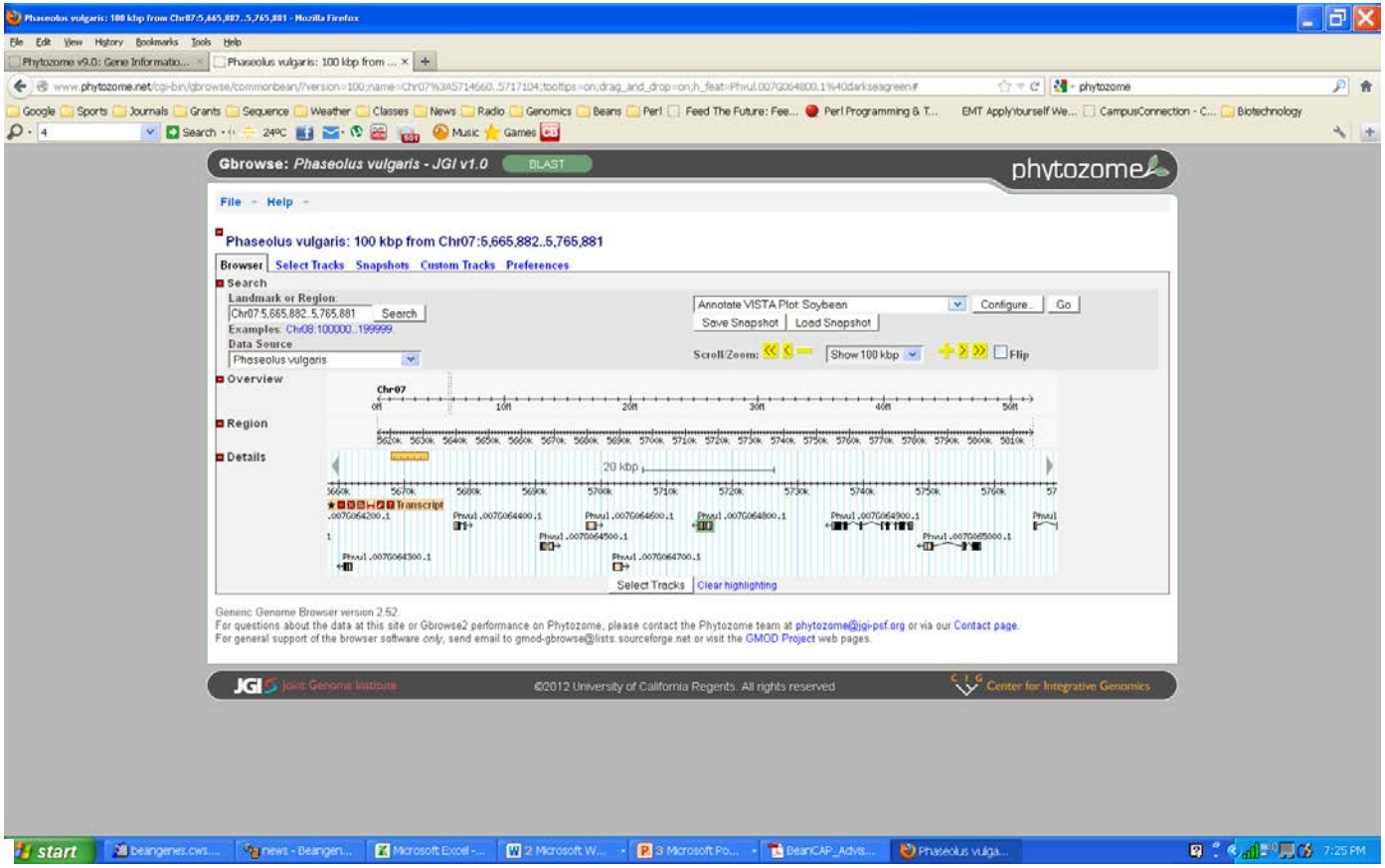
- *Collect next-generation sequence data*
- *Develop a high-density SNP map*
- *Create a sequence level assembly using NG sequence and SNP map*
- *Define gene models*
- *Annotate gene models*

The Sequencing Project

Accomplishments

Genome released publicly

August, 2012



United States Department of Agriculture
National Institute of Food and Agriculture

Sequencing of Pooled DNA Samples

Looking for domestication regions

Pool	Pool size (n)	Depth
Wild		
Middle American	30	3.9x
Andean	30	4.0x
Landraces		
Durango	48	4.1x
Mesoamerican	26	3.4x
Nueva Granada	9	2.6x
Peru	17	2.4x

What loci are selected during development of:

- *Wild gene pools* from ancestral wild population
- *Landraces* from Middle American and Andean wild gene pools
  - Important genes for crop improvement
- *Races* from landrace gene pools
  - Market-class specific genes???

USDA
United States Department of Agriculture
National Institute of Food and Agriculture

Sequencing of Pooled DNA Samples

Looking for domestication regions

Analysis

- SNP variation at all gene (n=27,082) within each pool
- Diversity for each gene within each pool
  - Pairwise diversity: π
- Measured loss of diversity at each stage of improvement
  - Ratio of ancestral to derived pool

$$\frac{\pi_{MesoWild}}{\pi_{MesoLandrace}}$$

- Meaured differentiation
  - F_{ST} between populations

Sequencing of Pooled DNA Samples

Low diversity genes

Population	# Low diversity genes Criteria SNP=0	SNP < 4
All *P. vulgaris*	**142**	677
All Middle American	212	677
Wild Middle America	198	677
Middle American landraces	276	689
Race Durango	293	689
Race Mesoamerican	249	689
All Andean	445	689
Wild Andean	224	689
Andean landraces	528	689
Race Nueva Granada	246	689
Race Peru	247	689

Essential *P. vulgaris* genes

- Not detected as selected genes because
  - Low diversity in ancestral populations
  - Selection criteria do not apply

Sequencing of Pooled DNA Samples

Looking for selected genes

Required features of a domesticated gene

1. Diversity

- Ratio of ancestral to derived pool

$$\frac{\pi_{MesoWild}}{\pi_{MesoLandrace}}$$

  - Ex: ***Gene selected during Mesoamerican domestication***
- Large values represent great differentiation
  - **Upper 10% of empirical distribution**

Sequencing of Pooled DNA Samples

Looking for selected genes

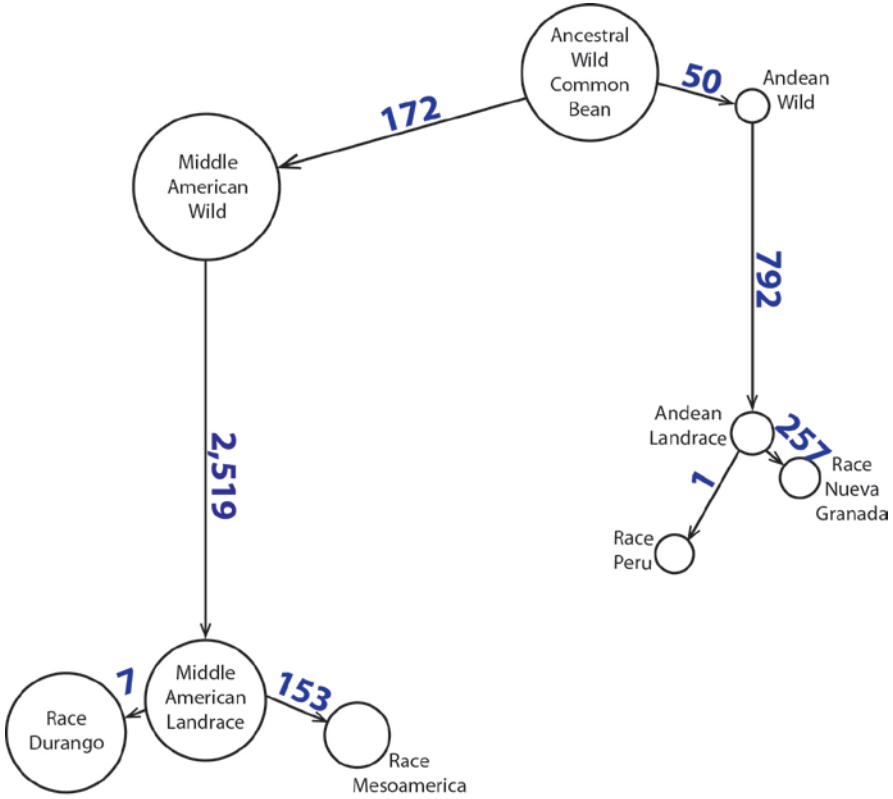
Required features of a domesticated gene

2. Differentiation

- Change in allele frequencies leads to differentiation
- F_{ST} between populations
  - Large values represent differentiation
    - Cutoff: FST > 0.15
      - **Wright: F_{ST} > 0.15 = great population differentiation**

Sequencing of Pooled DNA Samples

Looking for selected genes



Genes selected in at least one population	**3,791**
Genes selected in two populations	**160**
Same genes selected in Middle American and Andean landraces	**94**

United States Department of Agriculture
National Institute of Food and Agriculture

Sequencing of Pooled DNA Samples

Selected Mesoamerican landrace genes

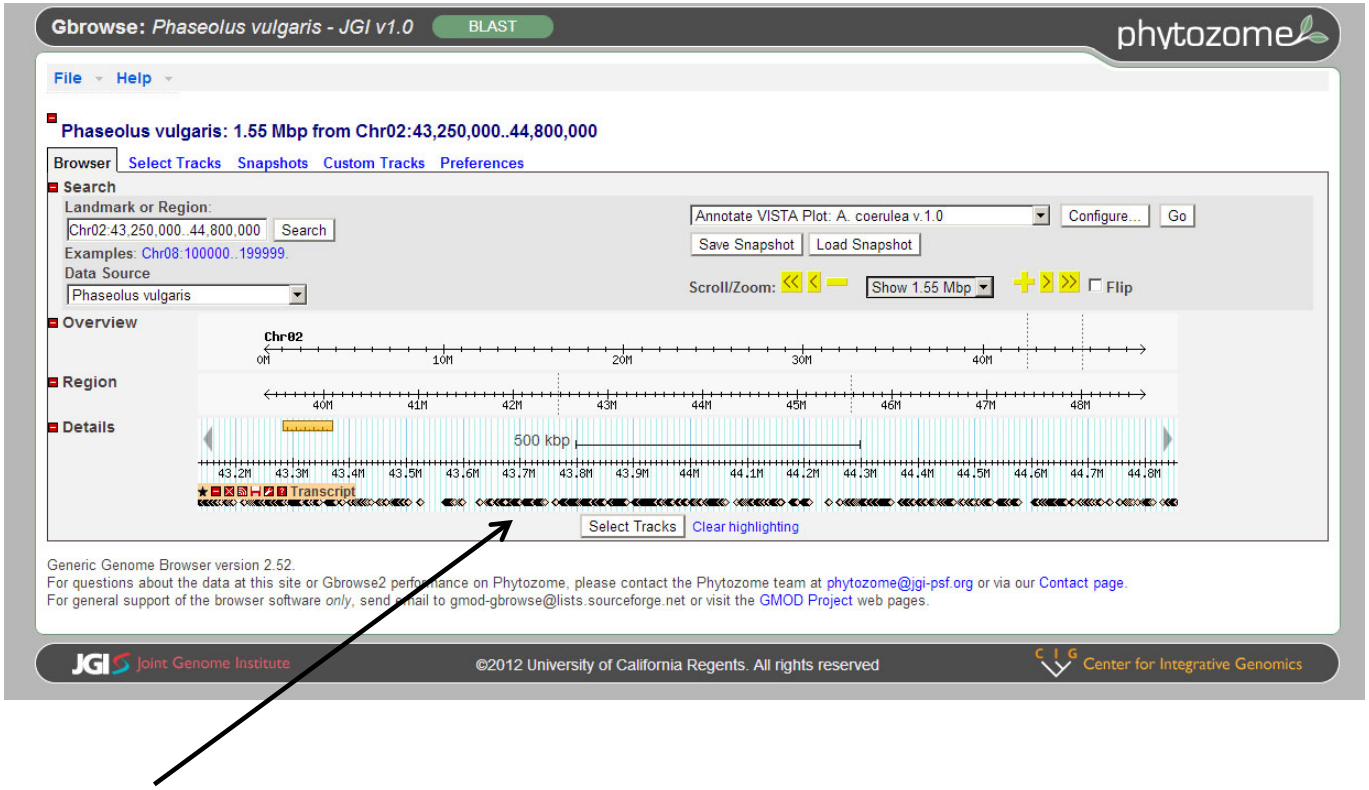
Pv gene model	Chrom	Start position	π ratio	F_{ST}	At symbol	Function
Phvul.001G062100	1	7.66	6.19	0.58	DAG1	Sugar transport
Phvul.007G064800	7	5.71	5.02	0.50	ATGA20OX2	GA 20 oxidase 2 (Mendel's height gene)
Phvul.007G065600	7	5.85	4.99	0.56	AGL42	AGAMOUS-like 42 (early flowering)
Phvul.008G168000	8	43.53	6.01	0.58	NR1	Nitrate reductase 1 (**SEED WEIGHT**)
Phvul.009G162700	9	23.67	7.80	0.23	AtSWEET10	Sucrose transport in phloem

BeanCAP National Association Mapping Trial

- Seed weight an agronomic trait studied
- **Are any SNP loci mapping at/near nitrate reductase on Pv08?**

GWAS – Seed Weight

Finding the locus



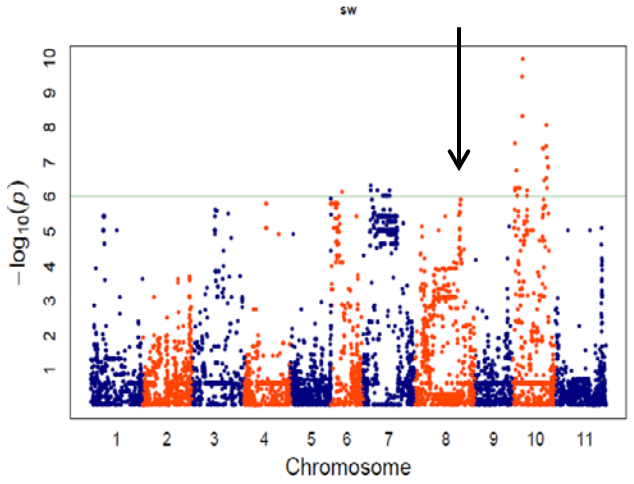
Many genes in the region

- Can we narrow it down??
- Searched for data in pooled sequence data

United States Department of Agriculture
National Institute of Food and Agriculture

GWAS – Seed Weight

Association mapping result



n=252

markers = 10,287 SNPs

Model = EMMA

Large trial

n=252; 10,827 SNPs

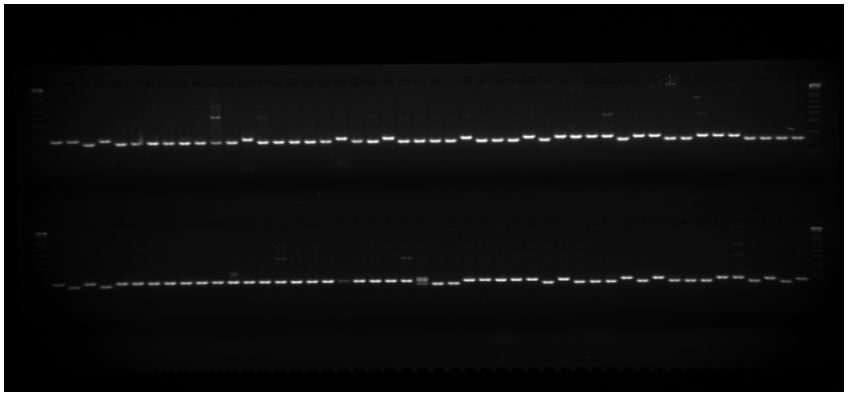
SNP	chr	Position_	P_value	maf
m4922	**8**	**43,827,895**	**3.51E-06**	**0.42**
m6145	8	44,237,412	2.97E-05	0.42
m6144	8	44,246,478	2.97E-05	0.42
m6143	8	44,270,482	2.97E-05	0.42
m6142	8	44,272,503	2.97E-05	0.42
m6141	8	44,273,042	2.97E-05	0.42
m6139	8	44,287,718	2.97E-05	0.42
m6174	8	44,338,667	4.64E-05	0.41
m6173	8	44,399,862	4.64E-05	0.41
m4856	8	44,425,316	4.64E-05	0.41
m4864	**8**	**44,470,617**	**8.35E-06**	**0.40**
m4863	**8**	**44,479,849**	**4.64E-05**	**0.41**
m4862	**8**	**44,480,227**	**8.35E-06**	**0.40**
m4861	**8**	**44,494,182**	**8.35E-06**	**0.40**
m4860	**8**	**44,526,077**	**8.35E-06**	**0.40**
m4859	**8**	**44,532,123**	**8.35E-06**	**0.40**
m4858	8	44,535,382	4.64E-05	0.41
m4029	8	44,605,263	1.73E-05	0.41
m4051	**8**	**44,712,020**	**2.81E-06**	**0.40**
m4050	**8**	**44,724,115**	**2.81E-06**	**0.40**
m4048	**8**	**44,754,521**	**2.81E-06**	**0.40**
m4044	**8**	**44,792,096**	**2.81E-06**	**0.40**
m4046	**8**	**44,804,652**	**2.81E-06**	**0.40**
m7104	**8**	**44,833,382**	**8.35E-06**	**0.40**
m7103	8	44,839,102	4.64E-05	0.41
m7102	8	44,842,481	4.64E-05	0.41
m7101	8	44,853,371	4.64E-05	0.41
m7100	**8**	**44,858,735**	**8.35E-06**	**0.40**
m7099	8	44,862,000	4.64E-05	0.41
m9586	8	44,891,515	1.45E-05	0.40
m4752	**8**	**44,921,533**	**8.35E-06**	**0.40**

USDA
United States Department of Agriculture
National Institute of Food and Agriculture

Application of Indel Markers

Targeting Pv08 SW locus

- Nitrate reductase gene region targeted for marker saturation (43.53 Mb)
  - Saturate with medium-throughput indel markers
    - Why? Gel-based screening
  - 22 indel markers in region selected
    - 42Mb – 46 Mb
  - 12 markers polymorphic
    - GWAS analysis with these 12 markers

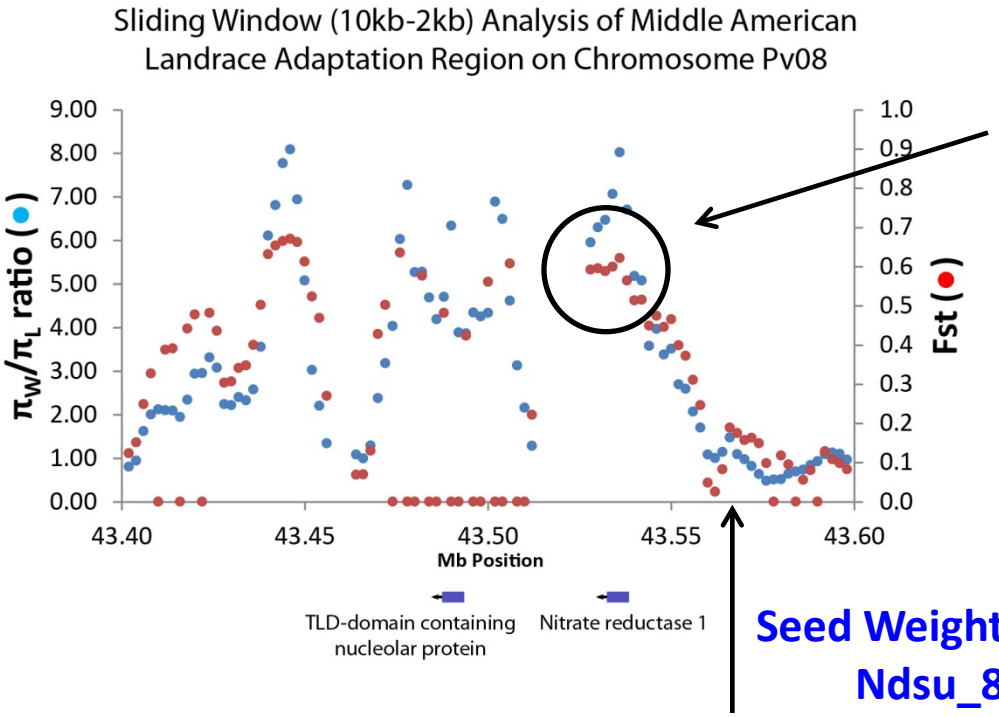


USDA
United States Department of Agriculture
National Institute of Food and Agriculture

Application of Indel Markers

Nitrate Reductase Indel Marker

Indel	chr	Position	P_value	maf
Ndsu_8_43.0997	8	43,099,724	4.32E-02	0.01
Ndsu_8_43.5710	8	43,570,352	3.11E-03	0.40
Ndsu_8_44.5633	8	44,563,293	4.50E-01	0.01
Ndsu_8_44.7084	8	44,708,437	9.01E-04	0.40
Ndsu_8_44.8451	8	44,845,148	9.01E-04	0.40



Very high:

- π ratio
- F_{ST} value

Seed Weight Indel Marker: Ndsu_8_43.5710