



Phenotype database at sol genomics network

- What is a phenotype?
- How do we store a phenotype?
- Different users, different notions
- Linking multiple experiments
- Linking phenotypes and genotypes

What is a phenotype?

Observable trait or characteristic of an organism

- Morphology
- Development
- Behavior
- Biochemical property
- Molecular characteristics??
- Dependent on genotype
- Environment



$$G + E + G \times E = P$$

What is a phenotype?



Observable trait or characteristic of an organism

- Quantitative Vs qualitative
- How is the phenotype recorded?
- What metadata is attached ?
- The phenome or a collection of measurements?



What is a phenotype?

Table: phenotype

A phenotypic statement, or a single atomic phenotypic observation, is a controlled sentence describing observable effects of non-wild type function. E.g. Obs=eye, attribute=color, cvalue=red.

phenotype Structure

F-Key	Name	Type	Description
	phenotype_id	serial	<i>PRIMARY KEY</i>
	uniquename	text	<i>UNIQUE NOT NULL</i>
cvterm	observable_id	integer	The entity: e.g. anatomy_part, biological_process.
cvterm	attr_id	integer	Phenotypic attribute (quality, property, attribute, character) - drawn from PATO.
	value	text	Value of attribute - unconstrained free text. Used only if cvalue_id is not appropriate.
cvterm	cvalue_id	integer	Phenotype attribute value (state).
cvterm	assay_id	integer	Evidence type

What is a phenotype?



phenotype columns

FK	Name	Type	Description
	phenotype_id	serial	<i>PRIMARY KEY</i>
	uniquename	text	<i>UNIQUE</i> <i>NOT NULL</i>
cvterm	observable_id	integer	The entity: e.g. <u>anatomy_part</u> , <u>biological_process</u> .
cvterm	<u>attr_id</u>	integer	Phenotypic attribute (<u>quality</u> , property, attribute, character) - drawn from PATO.
	value	text	<u>Value of attribute</u> - unconstrained free text. Used only if cvalue_id is not appropriate.
cvterm	cvalue_id	integer	Phenotype attribute value (state).
cvterm	assay_id	integer	<u>Evidence type</u> .

A phenotype is what you observe.

This means that stage or time since conception is not part of the phenotype. It is part of the experiment. (Unless you are observing development rate, in which case, stage would be the phenotype.)

That table description in the schema sounds like the old PATO, not the new [EQ](#) model.

How do we store a phenotype?



phenotype columns

FK	Name	Type	Description
	phenotype_id	serial	<i>PRIMARY KEY</i>
	uniquename	text	<i>UNIQUE</i> <i>NOT NULL</i>
<i>cvterm</i>	observable_id	integer	The entity: e.g. anatomy_part, biological_process.
<i>cvterm</i>	attr_id	integer	Phenotypic attribute (quality, property, attribute, character) - drawn from PATO.
	value	text	Value of attribute - unconstrained free text. Used only if cvalue_id is not appropriate.
<i>cvterm</i>	cvalue_id	integer	Phenotype attribute value (state).
<i>cvterm</i>	assay_id	integer	Evidence type.

Entity-Quality-Value

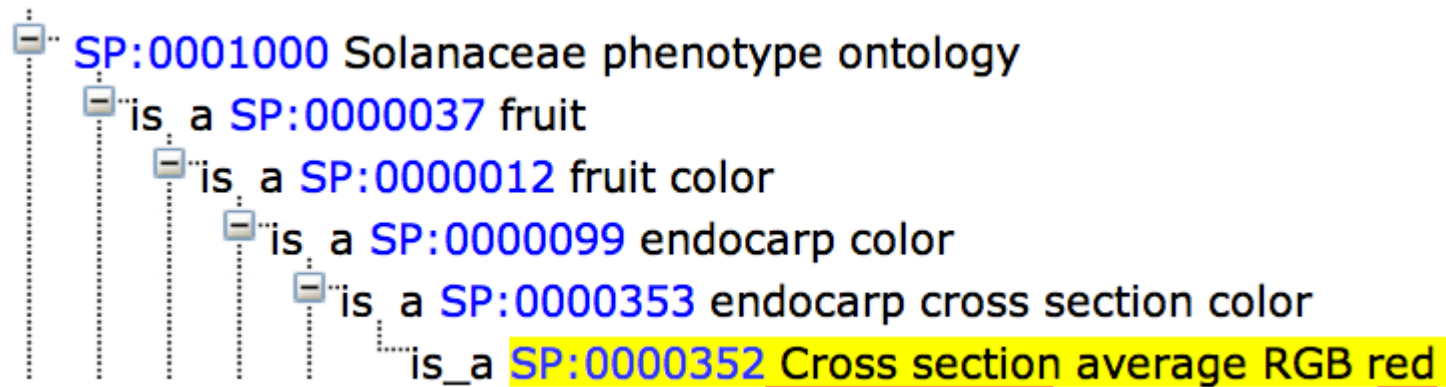
- Entity is an observable ('fruit')
- Quality is an attribute ('color')
- Value is qualitative ('orange') OR quantitative (140.02; unit: RGB red)

How do we store a phenotype?

Entity-Quality-Value

- Entity is an observable ('fruit')
- Quality is an attribute ('color')
- Value is qualitative ('orange') OR quantitative (140.02; unit: RGB red)

Pre-composed Vs Post-composed terms



PO-> fruit

PATO-> red

How do we store a phenotype?

Metadata

- **Unit – may or may not be part of the cvterm**
- **Property of the phenotype or the experiment ?**
Treatments, design, weather, location, person, date
- **Should phenotypes be reused? (one post-composed phenotype will have multiple values)**

How do we store a phenotype?

Metadata

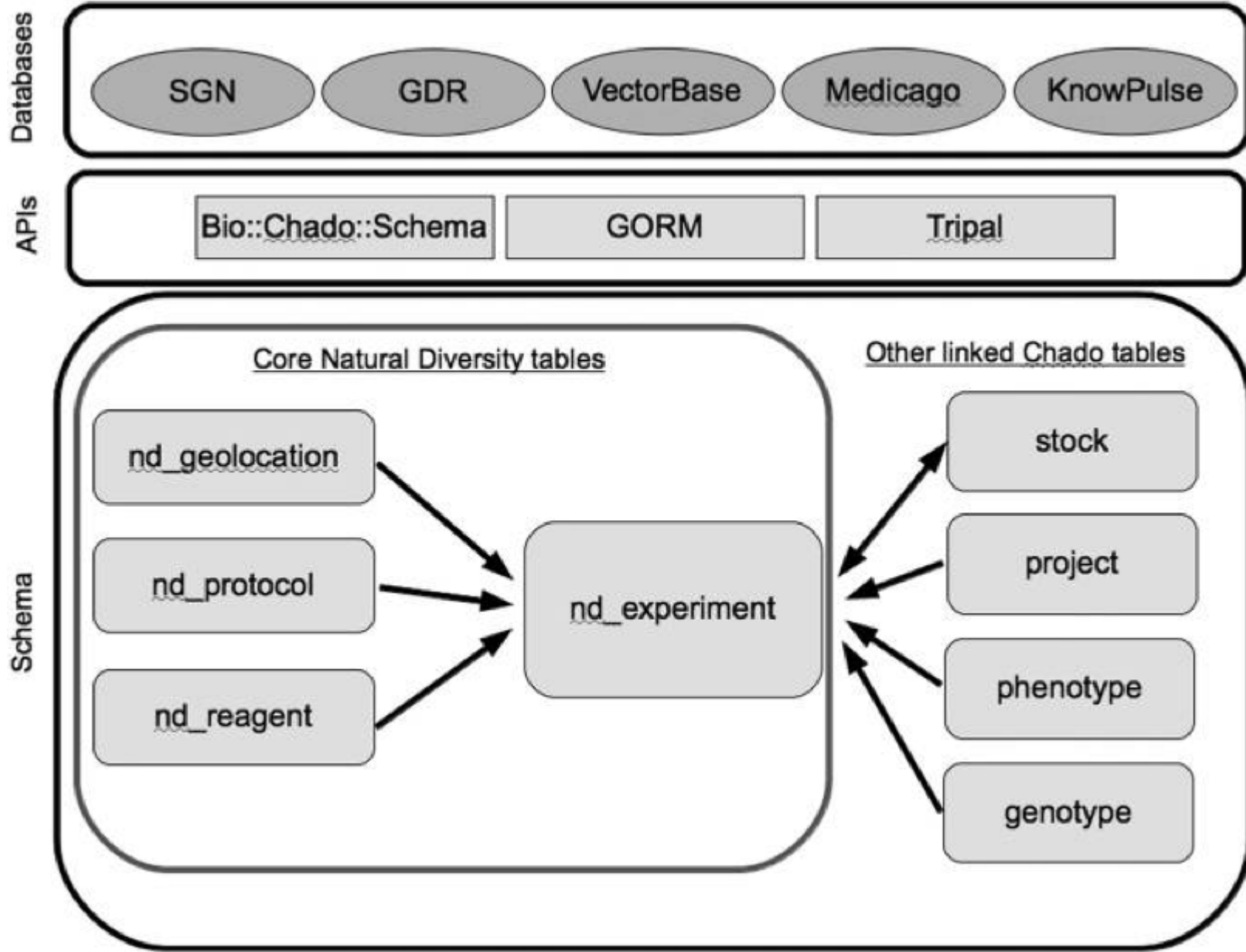
- **Unit** – may or may not be part of the cvterm
- **Property of the phenotype or the experiment ?**

Treatments, design, weather, location, person, date

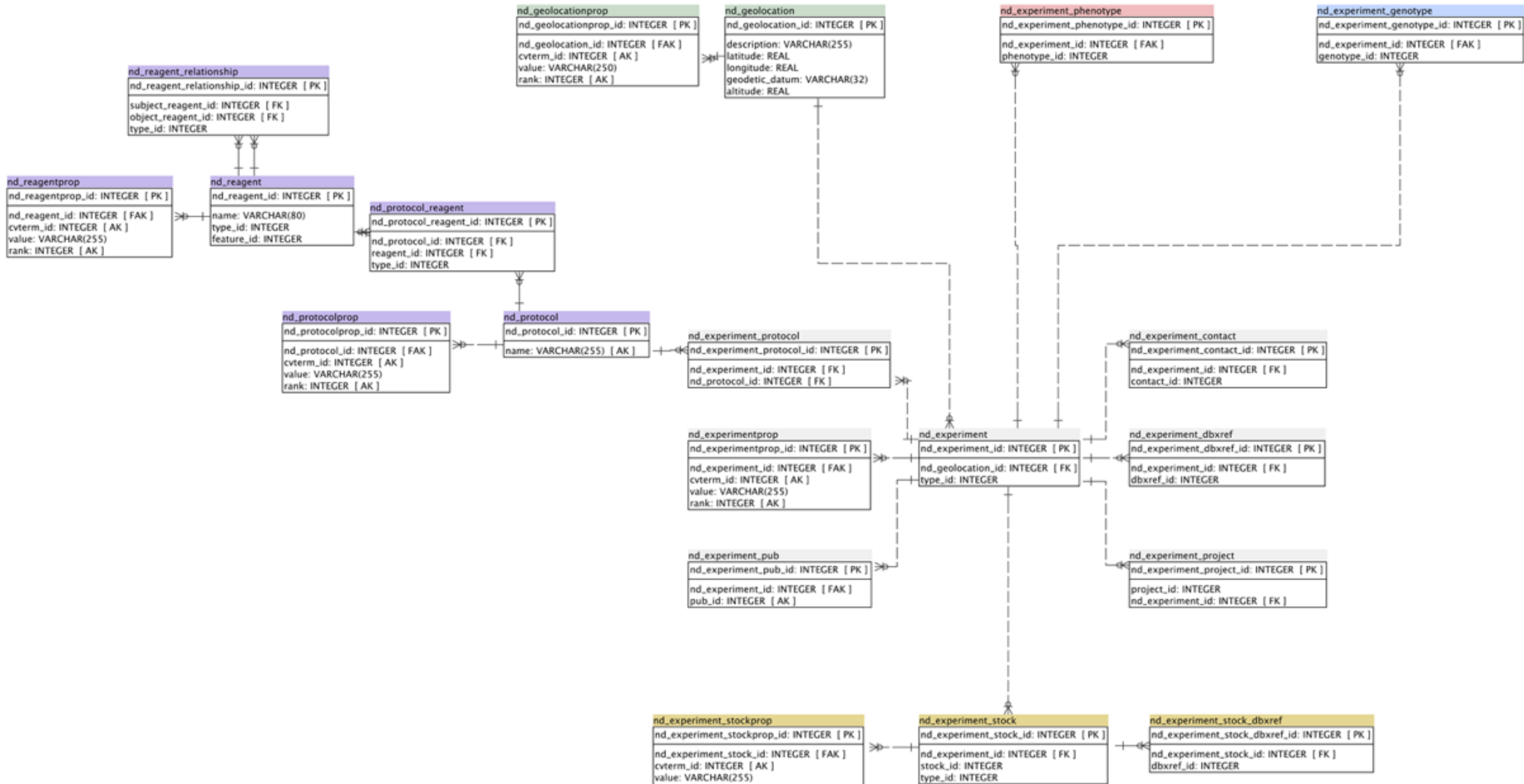
Phenotype can be linked with

- **Accessions**
- **Experiments**
- **Genotypes**

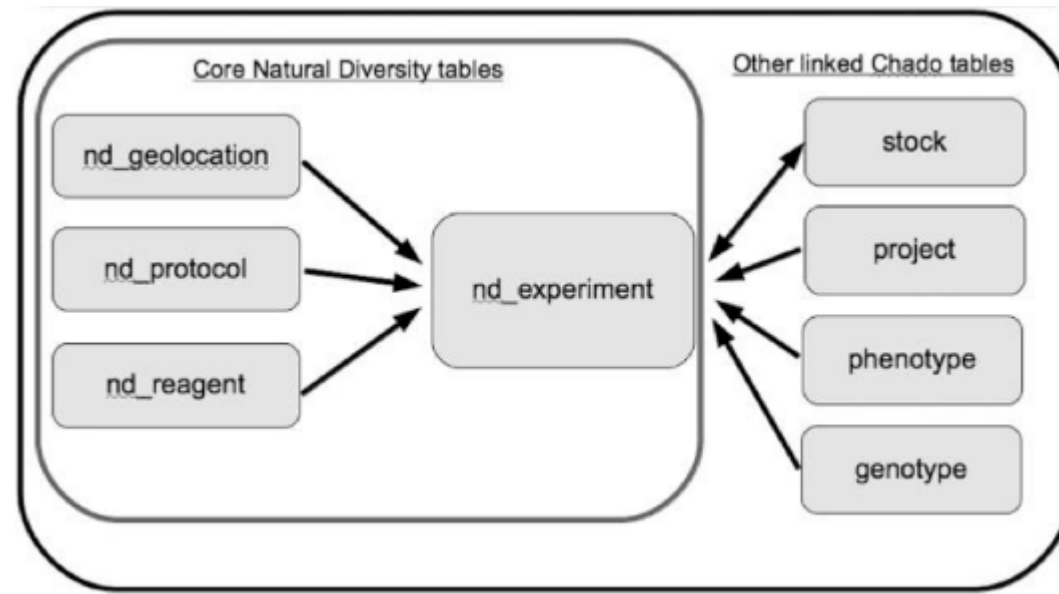
How do we store a phenotype?



How do we store 'natural diversity'?



How do we store 'natural diversity'?



- Multiple experiments of the same accessions/lines/strains
- Collected, treated and evaluated in various locations, environments, and times.
- Each accessions/lines/strains can be scored for a large number of phenotypic traits.
- Genotyped with an array of genetic markers.
- Storing data from experiments that generate new accessions/lines/strains and experimental samples, such as field collection, cross, and treatments.

Different users, different notions



Tomato Data Entry				Quality Measures		
Line #	Plot Number	Replicate Number	Inflorescence	Date of Maturity (third set)	Brix	pH
#	#	#	<i>2nd and 3rd truss - simple/intermediate/compound</i>	<i>MM/DD/YY</i>		
			<i>SP:0000071</i>	<i>SP:0000366</i>	<i>SP:0000165</i>	<i>SP:0000170</i>
SCT_0325	8901	1	simple	9/24/2009	5.8	4.17
SCT_0287	8902	1	simple	9/3/2009	6.6	4.52
SCT_0289	8903	1	simple	9/24/2009	7.3	4.23
SCT_0295	8904	1	simple	8/27/2009	6.7	4.42
SCT_0296	8905	1	simple	9/24/2009	5.7	4.19
SCT_0308	8906	1	simple	9/24/2009	5.3	4.26
SCT_0298	8907	1	simple	9/24/2009	4.6	4.34
SCT_0299	8908	1	simple	9/24/2009	5.8	4.53
SCT_0303	8909	1	simple	8/27/2009	4.9	4.24
SCT_0326	8910	1	simple	8/27/2009	4.9	3.99
SCT_0304	8911	1	simple	9/3/2009	5.7	4.35
SCT_0308	8912	1	simple	9/3/2009	5.5	4.39
SCT_0327	8913.2	1	simple	8/27/2009	5.8	4.24
SCT_0327	8913.1	1	simple	8/27/2009	5	4.16
SCT_0310	8914	1	simple	8/27/2009	5.2	4.25
SCT_0312	8915	1	simple	8/27/2009	5.8	4.31
SCT_0314	8916	1	simple	9/3/2009	5.7	4.35
SCT_0321	8917	1	simple	9/3/2009	6.1	4.32

Different users, different notions



Line #	Plot Number	Tuber size	Tuber appearance	Date data gathered	Sugar Analysis-Glucose	Sugar Analysis-Sucrose
#	#	NE1014 scale 1-9, see appendix	NE1014 scale 1-9, see appendix	mm/yyyy	%FW	%FW
A05141-1	64-17	9	5	10/7/2009	0.006	0.169
A05141-10	62-16	3	7	10/7/2009	0.003	0.081
A05141-100	63-19	7	5	10/7/2009	0.063	0.082
A05141-101	63-18	7	5	10/7/2009	0.005	0.103
A05141-102	63-16	5	5	10/7/2009	0.016	0.127
A05141-105	64-20	7	7	10/7/2009	0.037	0.096
A05141-106	62-02	9	7	10/7/2009	0.017	0.153
A05141-107	63-20	7	7	10/7/2009	0.485	0.917
A05141-109	64-02	7	5	10/7/2009	0.025	0.118
A05141-11	65-22	7	5	10/7/2009	0.052	0.096
A05141-110	62-06	7	5	10/7/2009	0.070	0.088
A05141-111	65-01	3	1	10/7/2009	0.041	0.210
A05141-112	62-09	7	5	10/7/2009	0.039	0.127
A05141-114	64-03	1	3	10/7/2009	0.059	0.122
A05141-115	63-05	5	3	10/7/2009	0.022	0.129
A05141-116	64-13	5	5	10/7/2009	0.037	0.173
A05141-118	63-07	1	1	10/7/2009	0.062	0.195
A05141-119	62-17	7	7	10/7/2009	0.113	0.108
A05141-12	64-23	3	5	10/7/2009	0.006	0.080
A05141-120	65-11	5	7	10/7/2009	0.030	0.139

Different users, different notions



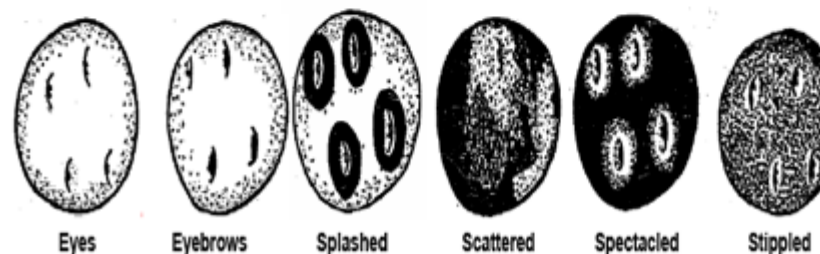
Line #	Plot Number	Tuber size	Tuber appearance	Date data gathered	Sugar Analysis-Glucose	Sugar Analysis-Sucrose
#	#	NE1014 scale 1-9, see appendix	NE1014 scale 1-9, see appendix	mm/yyyy	%FW	%FW
A05141-1	64-17	9	5			
A05141-10	62-16	3	7			
A05141-100	63-19	7	5			
A05141-101	63-18	7	5			
A05141-102	63-16	5	5			
A05141-105	64-20	7	7			
A05141-106	62-02	9	7			
A05141-107	63-20	7	7			
A05141-109	64-02	7	5			
A05141-11	65-22	7	5			
A05141-110	62-06	7	5			
A05141-111	65-01	3	1			
A05141-112	62-09	7	5			
A05141-114	64-03	1	3			
A05141-115	63-05	5	3			
A05141-116	64-13	5	5			
A05141-118	63-07	1	1			
A05141-119	62-17	7	7			
A05141-12	64-23	3	5			
A05141-120	65-11	5	7			

Other Descriptors

Scale	Vine Maturity Ratings	Heat Sprout Rating	Color (Figure 13 below)	Shape (Figure 14)
1	100% dead)	none	eyes	compressed
2	--	swollen eyes	eyebrows	round
3	--	sprouts ca. 1/4"	splashed	oval
4	senesce (yellowing	sprouts ca. 1/2"	scattered	oblong
5	--	sprouts ca. 1"	spectacled	long
6	flowering	sprouts ca. 2"	stippled	--
7	flowers yet in	sprouts ca. 3"	--	--
8	--	new top growth	--	--
9	bloom with buds	chain tubers	--	--

Figure 13. Used for Tuber Skin Pigment Pattern on the Potato Data Entry sheet

Figure 13: Distribution of Secondary Skin Tuber Color



Different users, different notions



Flower Color Scale

Flower Color Code	Flower Color	Flower Feature Code	Flower Feature
R	Red	#	White Dots
P	Purple	@	White Speckles
W	White	*	White Star pattern
B	Blue	^	B
NF	No Flowers (didn't flower or bud abortion prior to flowering)	v^	Very pronounced White tip
		v	also used in extreme cases of flower color shading—i.e., a very deep purple flower would be: vDP
		~	Ruffled margins
Examples:			
MRP^ = Medium Red-Purple Color with White Flower Tips			
W~ = White Flower Color and ruffled margins			

Group	Color (Figure 13 below)	Shape (Figure 14)
eyes	eyes	compressed
eyebrows	eyebrows	round
1/4"	splashed	oval
1/2"	scattered	oblong
a. 1"	spectacled	long
a. 2"	stippled	--
a. 3"	--	--
growth	--	--
leaves	--	--

on the Potato Data Entry sheet

Spectacled Stippled

Multiple phenotyping experiments

Stock details

[New] [Edit] [Delete]

Organism **Solanum lycopersicum**
Stock type **accession**
Stock name **NC2C**
Uniquename **NC2C**
Description

Stock editors: [SolCAP project](#)

Related stocks

Accessions this accession is a member of

Type	Name
population	Cultivars and heirloom lines

Accessions that are plots of this accession

Type	Name
plot	1085_1_2009,Davis, California
plot	2091_2_2009,Davis, California
plot	CAP-039_1_2009,Mills River, North Carolina
plot	CAP-181_2_2009,Mills River, North Carolina
plot	150_1_2010,Davis, California
plot	304_2_2010,Davis, California
plot	5266_1_2010,Wimauma, Florida
plot	6266_2_2010,Wimauma, Florida
plot	391_1_2010,Mills River, NC
plot	392_2_2010,Mills River, NC

Multiple phenotyping experiments

Stock details

[New] [Edit] [Delete]

Organism **Solanum lycopersicum**
Stock type **accession**
Stock name **NC2C**
Uniquename **NC2C**
Description

Stock editors: [SolCAP project](#)

☐ Phenotype data

☐ Member phenotypes solcap fresh tomatoes 2010, NC State University

Quality	Average	Min	Max	Lines/repeats
pH	3.95	3.94	3.96	2
titratable acids (gram per deciliter)	0.53	0.52	0.54	2
total soluble solids	4.75	4.30	5.20	2
Trait	Average	Min	Max	Lines/repeats
days to fruit ripening	131.00	131.00	131.00	2

☐ Member phenotypes solcap fresh tomatoes 2009, NC State University

Quality	Average	Min	Max	Lines/repeats
pH	4.41	4.32	4.50	2
titratable acids (gram per deciliter)	0.41	0.40	0.42	2
total soluble solids	4.50	4.50	4.50	2

Multiple phenotyping experiments

☐ **Member phenotypes** solcap fresh tomatoes 2010. Hutchinson Drive, Davis CA

Fruit color	Average	Min	Max	Lines/repeats
Cross section average 'a' value	23.71	11.90	31.85	18
Cross section average 'b' value	25.76	21.52	29.57	18
Cross section average L value	37.40	30.81	43.04	18
Cross section average RGB blue	43.53	37.85	52.50	18

Fruit cross section	Average	Min	Max	Lines/repeats
fruit lobedness degree	1.15	0.58	1.72	13
pericarp area (square centimeter)	0.45	0.45	0.45	13

Quality	Average	Min	Max	Lines/repeats
pH	4.26	4.24	4.28	2
titratable acids (gram per deciliter)	0.33	0.33	0.34	2
total soluble solids	5.76	5.53	5.98	2

Trait	Average	Min	Max	Lines/repeats
days to first flowering	69.50	67.00	72.00	2
days to fruit ripening	118.50	115.00	122.00	2
distal fruit end scar size	0.80	0.60	1.00	2

Qualitative Trait	Value
flower fasciation	absent, absent
hypocotyl anthocyanin content	present, present
inflorescence structure	intermediate inflorescence, simple inflorescence
peduncle	absent, absent
plant habit index	indeterminate, indeterminate

Multiple phenotyping experiments



Plot: 304_2_2010,Davis, California

Stock details

New QTL population |

[New] [Edit] [Delete]

Organism **Solanum lycopersicum**
Stock type **plot**
Stock name **304**
Uniquename **304_2_2010,Davis, California**
Description

Stock editors: [SolCAP project](#)

[-] Phenotype data

[-] Experiment: solcap fresh tomatoes 2010. Hutchinson Drive, Davis CA

Fruit color	Average	Min	Max	Lines/repeats
Cross section average 'a' value	23.77	20.20	25.63	5
Cross section average 'b' value	26.05	24.63	28.39	5
Cross section average L value	33.50	30.81	36.29	5
Cross section average RGB blue	41.41	37.85	45.53	5
Cross section average RGB green	60.08	52.19	69.76	5

Multiple phenotyping experiments



[-] **Phenotype data**

[-] **Experiment: solcap fresh tomatoes 2010. Hutchinson Drive, Davis CA**

Fruit color	Average	Min	Max	Lines/repeats
Cross section average 'a' value	23.77	20.20	25.63	5
Cross section average 'b' value	26.05	24.63	28.39	5
Cross section average L value	33.50	30.81	36.29	5
Cross section average RGB blue	41.41	37.85	45.53	5
Cross section average RGB green	60.08	52.19	69.76	5

Quality	Average	Min	Max	Lines/repeats
pH	4.28	4.28	4.28	1
titratable acids (gram per deciliter)	0.34	0.34	0.34	1
total soluble solids	5.53	5.53	5.53	1

Trait	Average	Min	Max	Lines/repeats
days to first flowering	72.00	72.00	72.00	1
days to fruit ripening	115.00	115.00	115.00	1
distal fruit end scar size	0.60	0.60	0.60	1

Qualitative Trait	Value
flower fasciation	absent
hypocotyl anthocyanin content	present
inflorescence structure	simple inflorescence
peduncle	absent

Multiple phenotyping experiments



Stock details

[New QTL population](#) | [Back to stock search](#)

[\[New\]](#) [\[Edit\]](#) [\[Delete\]](#)

Organism **Solanum lycopersicum**
Stock type **f2 population**
Stock name **QTL Tomato Howard German x LA1589 F2**
Uniquename **QTL Tomato Howard German x LA1589 F2**
Description **QTL Tomato Howard German**

Stock editors: [Esther van der Knaap](#)

Phenotype data

Member phenotypes recorded for population QTL Tomato Howard German x LA1589 F2 by [Esther van der Knaap](#)

Trait	Average	Min	Max	Lines/repeats	QTL(s)
distal angle macro 10%	142.52	69.77	172.26	113	
distal angle macro 15%	119.24	61.80	145.58	113	
distal angle macro 20%	96.92	52.58	120.90	113	
distal angle macro 25%	76.59	36.86	100.26	113	
distal angle micro 2%	174.31	94.44	295.54	113	
distal angle micro 5%	161.87	82.43	241.80	113	
distal fruit end blockiness 10%	0.61	0.43	0.78	113	
distal fruit end blockiness 20%	0.80	0.70	0.91	113	
distal fruit end blockiness 30%	0.92	0.87	0.97	113	
distal fruit end blockiness 5%	0.45	0.22	0.58	113	
distal fruit end indentation	0.00	0.00	0.00	113	
distal fruit end protrusion	0.00	0.00	0.00	113	
fruit area	418.39	224.50	860.93	113	

Multiple phenotyping experiments

Population details

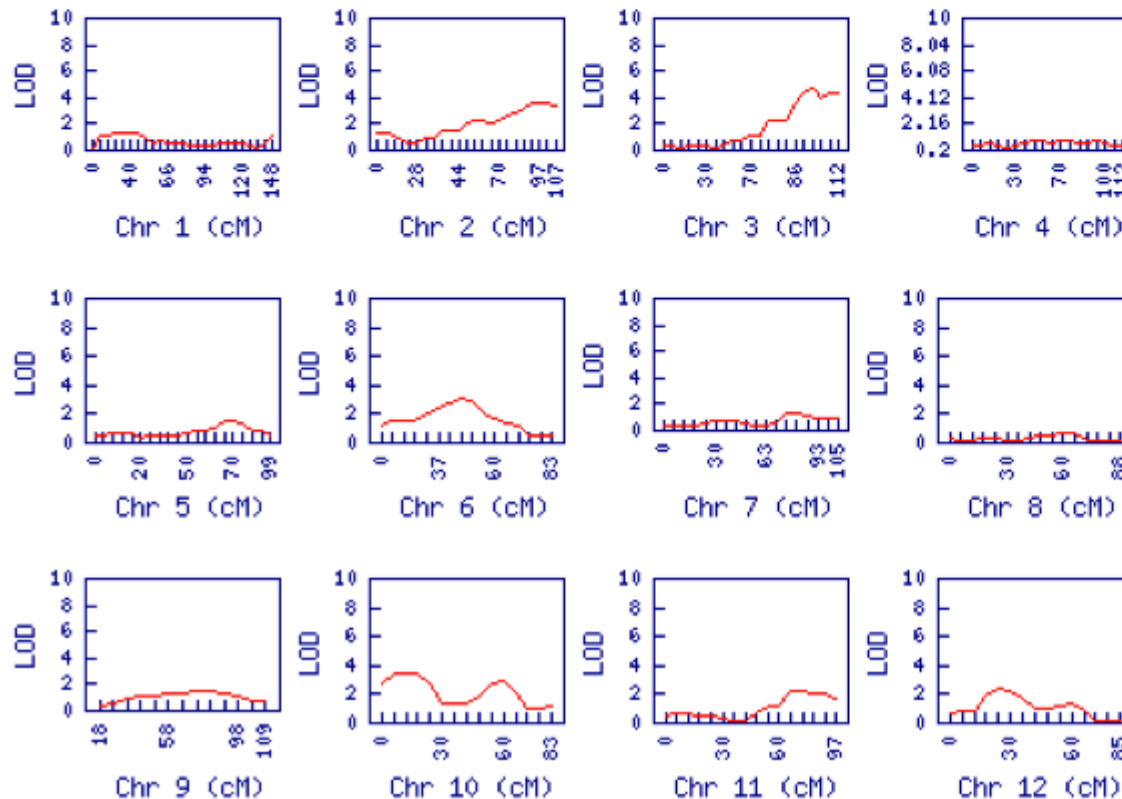
[Edit] [New QTL Population]

Name: **QTL Tomato Howard German x LA1589 F2**

Description: **QTL Tomato Howard German**

Uploaded by: **Esther van der Knaap**

QTL(s)



No. of permutations	1000
QTL genotype significance level	0.05
Genome scan size (cM)	10
QTL genotype probability method	Calculate
Mapping method	Maximum Likelihood
Permutation significance level	0.05

☐ Genotype data

☐ Experiment: Solcap processing tomatoes scored for >8,000 SNP markers 2011

Solcap Infinium array genotypes for stock SCT_0001 (name = 2K1-1439, id = 647)

- SNP genotypes stored in JSON format
 - ```
{"solcap_snp_sl_3060": "AA", "solcap_snp_sl_28593": "GG"}
```
- Data can be downloaded
- Tool for polymorphic SNP detection
- Association mapping of phenotypes and SNP markers
- Genomic selection in tomato/potato ?

# Back to phenotypes..

1. decision on post-composing terms . How?
2. revisit the phenotype module for breaking the value-phenotype ties
3. better definition of how we use the prop tables

phenotype\_cvterm Structure

| F-Key     | Name                | Type    | Description              |
|-----------|---------------------|---------|--------------------------|
|           | phenotype_cvterm_id | serial  | <i>PRIMARY KEY</i>       |
| phenotype | phenotype_id        | integer | <i>UNIQUE#1 NOT NULL</i> |
| cvterm    | cvterm_id           | integer | <i>UNIQUE#1 NOT NULL</i> |

# Back to phenotypes..

1. decision on post-composing terms . How?
2. revisit the phenotype module for breaking the value-phenotype ties
3. better definition of how we use the prop tables

phenotype Structure

| F-Key  | Name          | Type    | Description                                                                              |
|--------|---------------|---------|------------------------------------------------------------------------------------------|
|        | phenotype_id  | serial  | <i>PRIMARY KEY</i>                                                                       |
|        | uniquename    | text    | <i>UNIQUE NOT NULL</i>                                                                   |
| cvterm | observable_id | integer | The entity: e.g. anatomy_part, biological_process.                                       |
| cvterm | attr_id       | integer | Phenotypic attribute (quality, property, attribute, character) - drawn from PATO.        |
|        | value         | text    | Value of attribute - unconstrained free text. Used only if cvalue_id is not appropriate. |
| cvterm | cvalue_id     | integer | Phenotype attribute value (state).                                                       |
| cvterm | assay_id      | integer | Evidence type                                                                            |