A shiny app for exploring soybean milestone cultivars

Di Cook & Hadley Wickham
@visnut & @hadleywickham

July 2015

Data

Three sources of information:

- Next-generation sequencing DNA-seq on 79 lines: DNA sequencing libraries were prepared using TruSeq DNA sample prep and NuGENs unamplified prep kits (Illumina Inc., San Diego, CA and NuGEN Technologies Inc., San Carlos, CA).

- Field yield trials: 30/79 + 138 ancestral lines

- Breeding literature, what lines were bred to produce what line

Goals

- Ultimate task is to understand how the genome changed with the breeding of lines, and how this affected other traits.

- Three apps

1. Copy number variation (CNV): 2Gb of analysis files, annotations

2. SNPs: 12Gb of data, 20milll SNPs, 1mil locations, 79 lines

3. Genealogy: Shows the parent to child lineage

CNV app

- Seven tabs containing different functionality
- Four of the tabs, CNV Location, Copy Number, "Search CNVs by Location", and CNV List allow primarily concerned with exploring the identified copy number variants
- The other three tabs, Phenotype Data, Genealogy, and Methodology provide additional information about the soybean cultivars and the experimental methodology
- App written by Susan Vanderplas

Graphics

- Mix of ggplot2 and animint

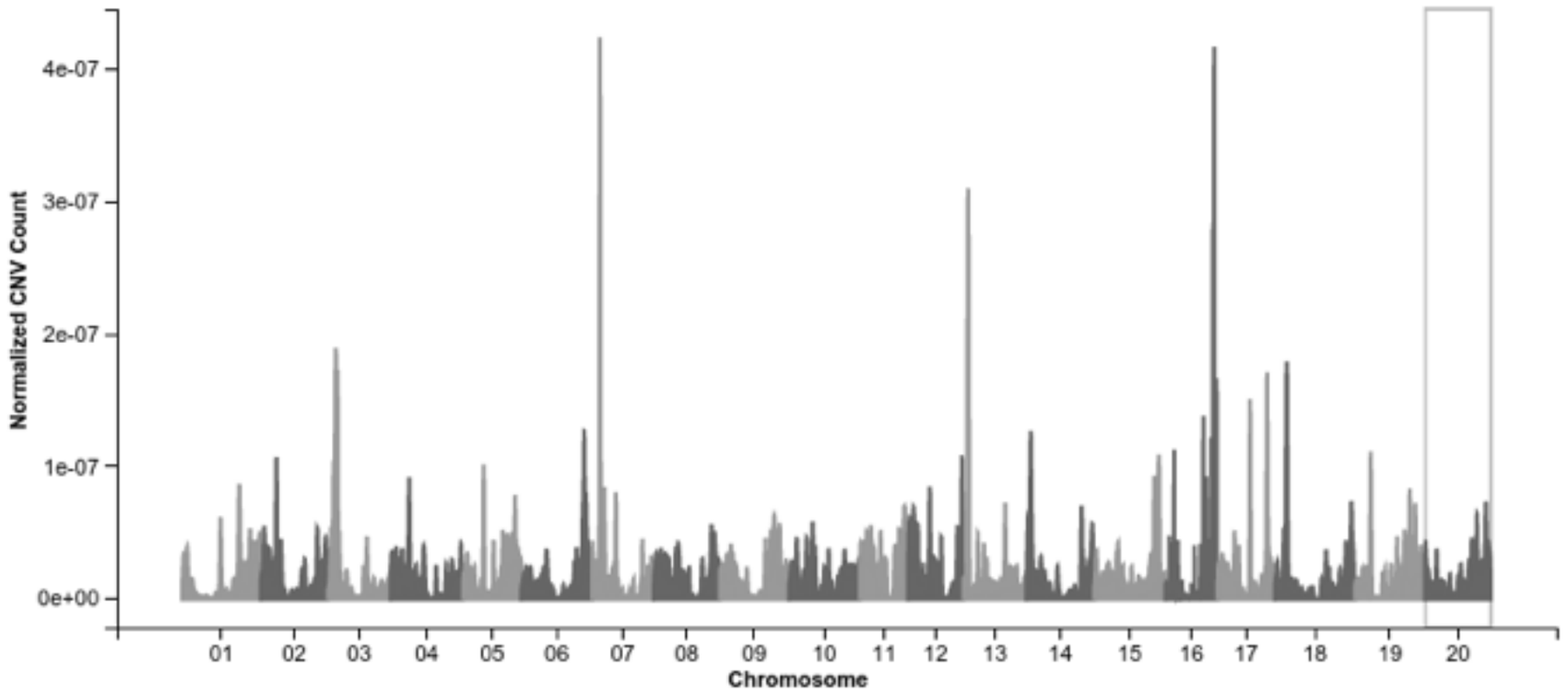
- Animint for interactive graphics, ggvis could be used instead

Use the first two plots to select a chromosome and one or more varieties.

The first plot ('Whole-Genome CNV Density') shows normalized CNV counts for all varieties over all chromosomes. Click on a chromosome region to view more detailed information for that chromosome.

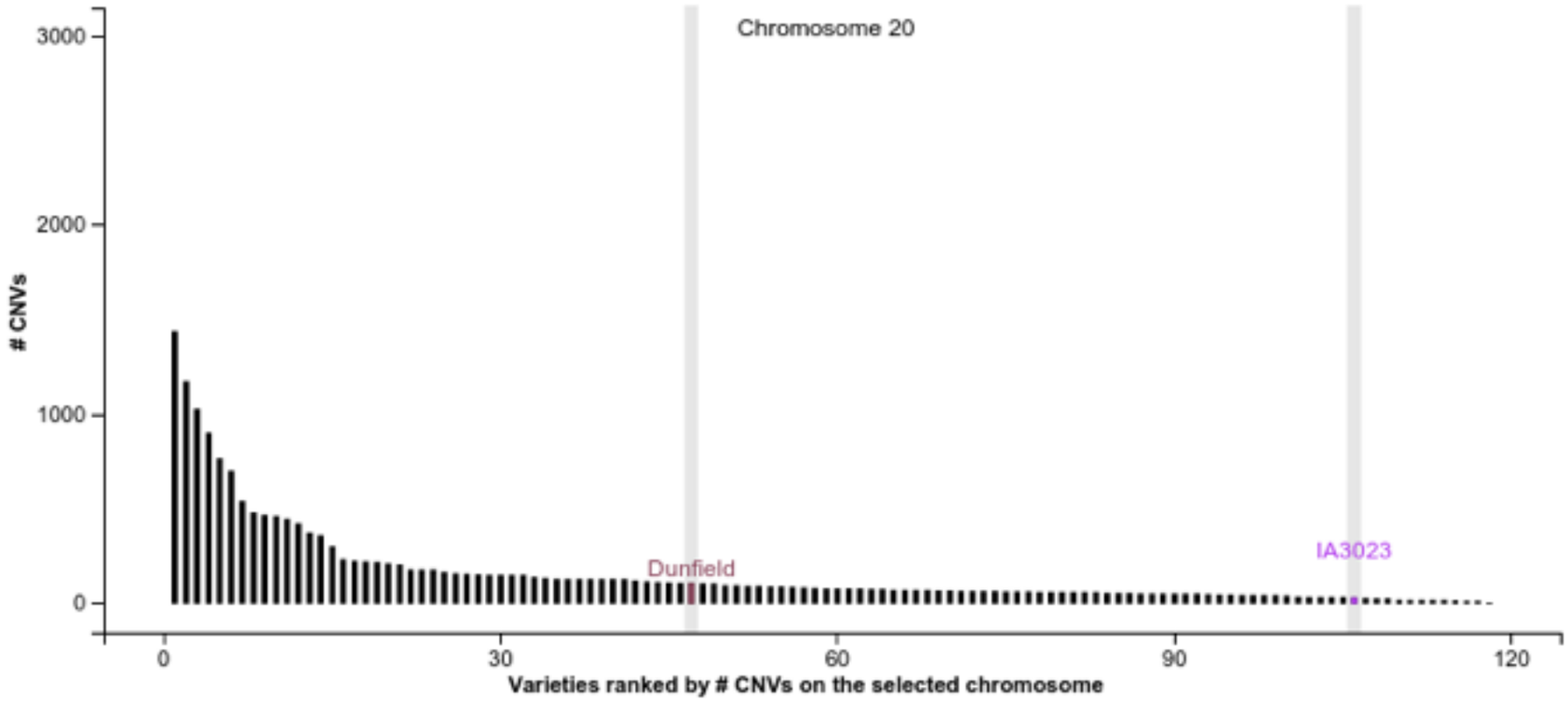
The second plot ('Number of CNVs by Variety') shows CNV counts for each variety, sorted from most CNVs to fewest CNVs. Click on one or more varieties to see more detailed information about which regions of the chromosome contain the most CNVs.

Whole-Genome CNV Density (all varieties)

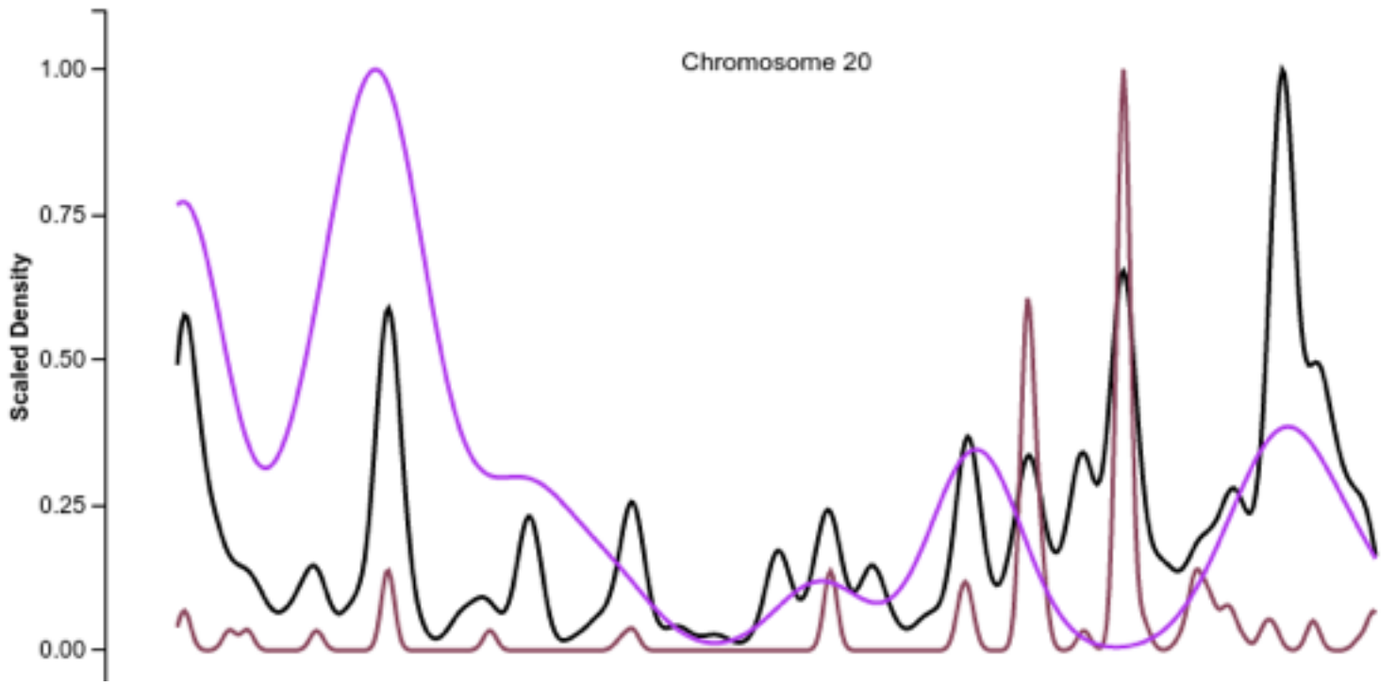


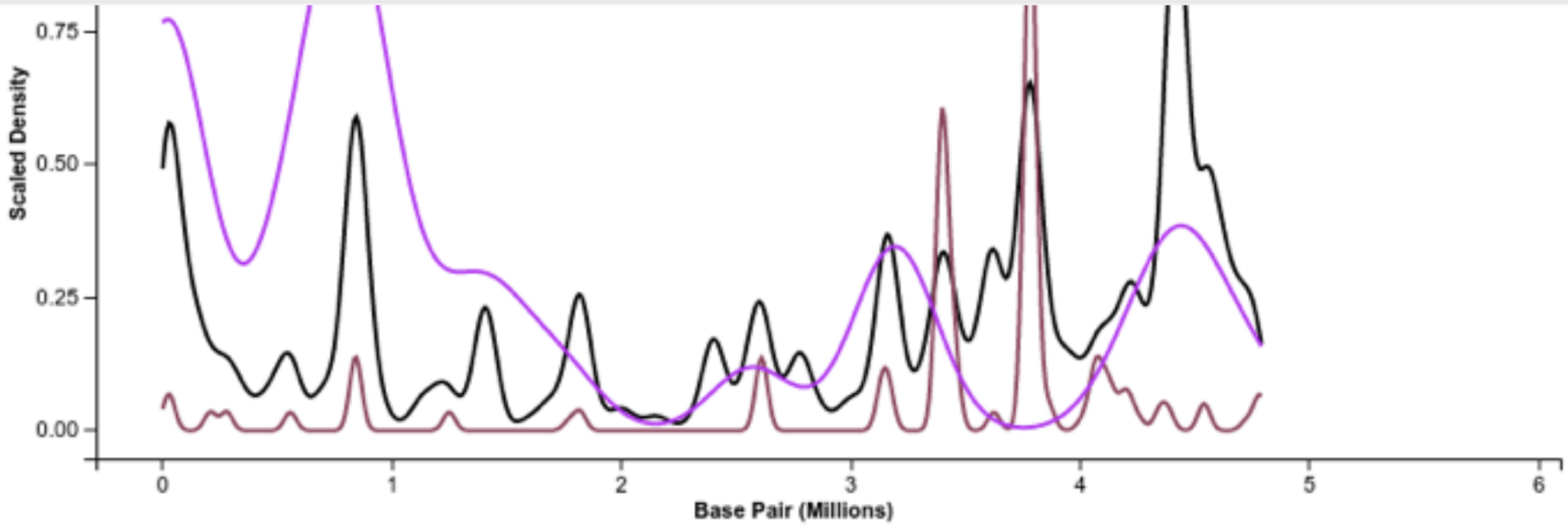
Number of CNVs by Variety

Number of CNVs by Variety

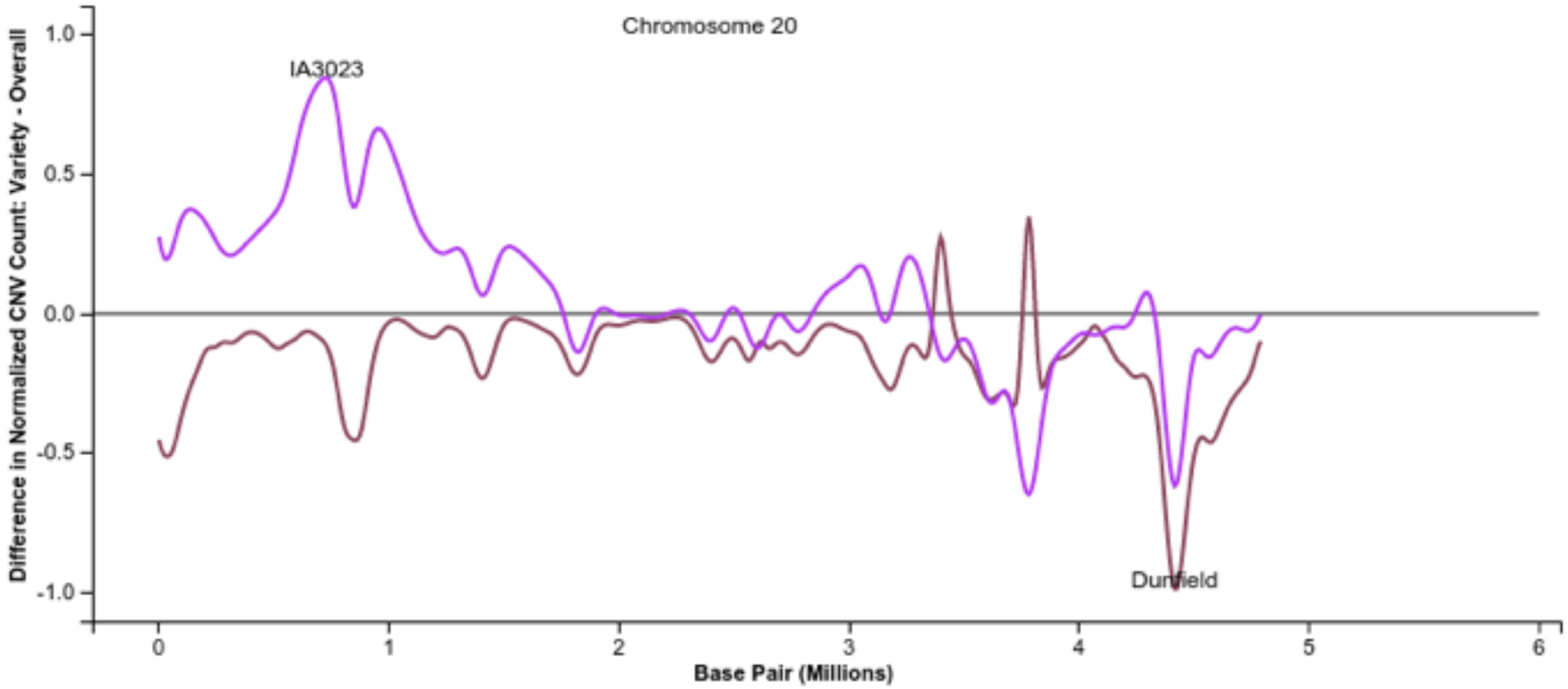


Density of CNVs (overall + selected variety)





Difference in Normalized CNV Count: Variety - Overall



The third plot ('Density of CNVs') shows the overall CNV distribution across all varieties in black, with selected varieties (from the top-right plot) shown in color.

The last plot ('Difference in Normalized CNV Count') shows, for each selected variety, which regions of the chromosome contain relatively more CNVs than the average variety, and which regions of the chromosome contain relatively fewer CNVs.

Clear Selections

Type the chromosome number or click on the text box for options

Choose Chromosomes

Chr20

Type the name of the variety or click on the text box for options

Choose Varieties

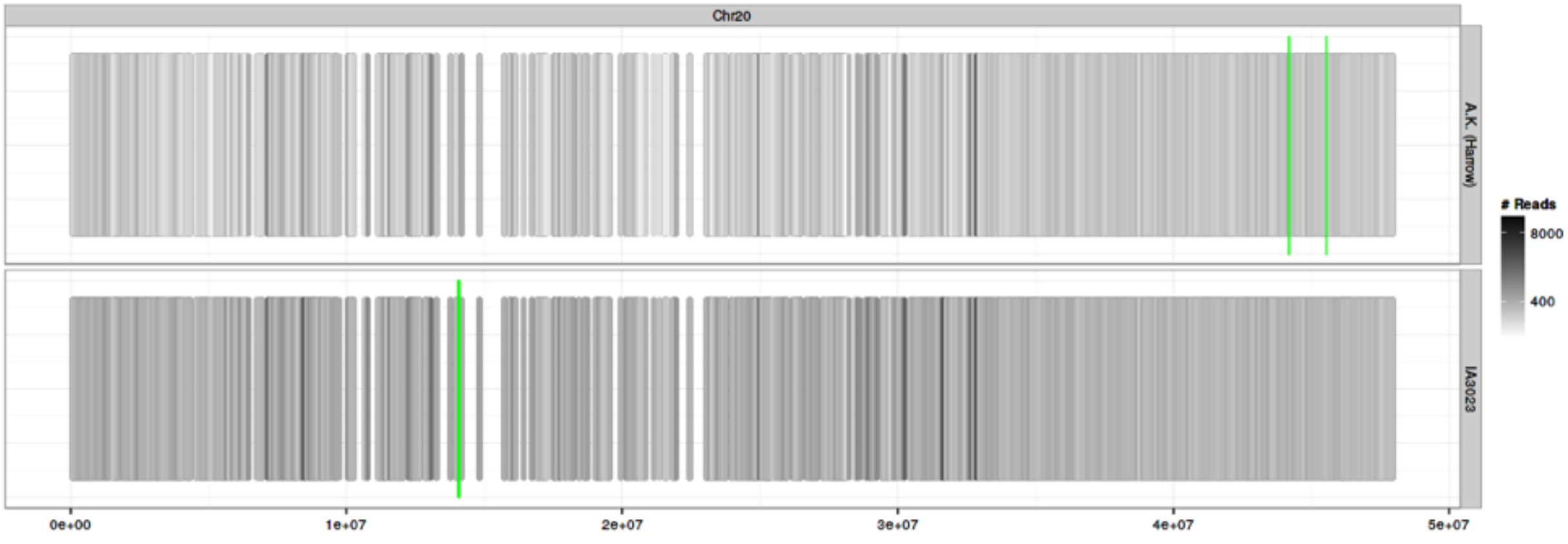
A.K. (Harrow) IA3023

Variety List

Select the types of CNV regions you would like to display

Choose Features (optional)

gene CDS mRNA exon



Green lines indicate a significant CNV at that location

Clear Selections | Download

Use the Search field above the table to filter by Glyma ID.

Type the chromosome number or click on the text box for options

Choose Chromosomes: Chr20

Type the name of the variety or click on the text box for options

Choose Varieties: Perry

Variety List

Select the types of CNV regions you would like to display

Choose Features (optional): gene CDS mRNA exon

Show 10 entries

Search:

Variety	Chromosome	Start	End	Feature	CN	ID
Perry	Chr20	1411768	1413076	gene	0	Glyma.20G015700
Perry	Chr20	14079025	14079718	CDS	1	Glyma.20G056400
Perry	Chr20	1412047	1412239	CDS	0	Glyma.20G015700
Perry	Chr20	43676609	43676727	CDS	2	Glyma.20G199000
Perry	Chr20	44144925	44145370	CDS	0	Glyma.20G204500
Perry	Chr20	44193804	44194628	CDS	4	Glyma.20G205000
Perry	Chr20	44202863	44203072	CDS	4	Glyma.20G205100
Perry	Chr20	1411768	1413076	mRNA	0	Glyma.20G015700.1
Perry	Chr20	1412047	1412239	exon	0	Glyma.20G015700
Perry	Chr20	44144925	44145370	exon	0	Glyma.20G204500
Variety	Chromosome			Feature	CN	ID

Showing 1 to 10 of 13 entries

Previous 1 2 Next

Clear Selections

Number of Generations to Show

1 2 10

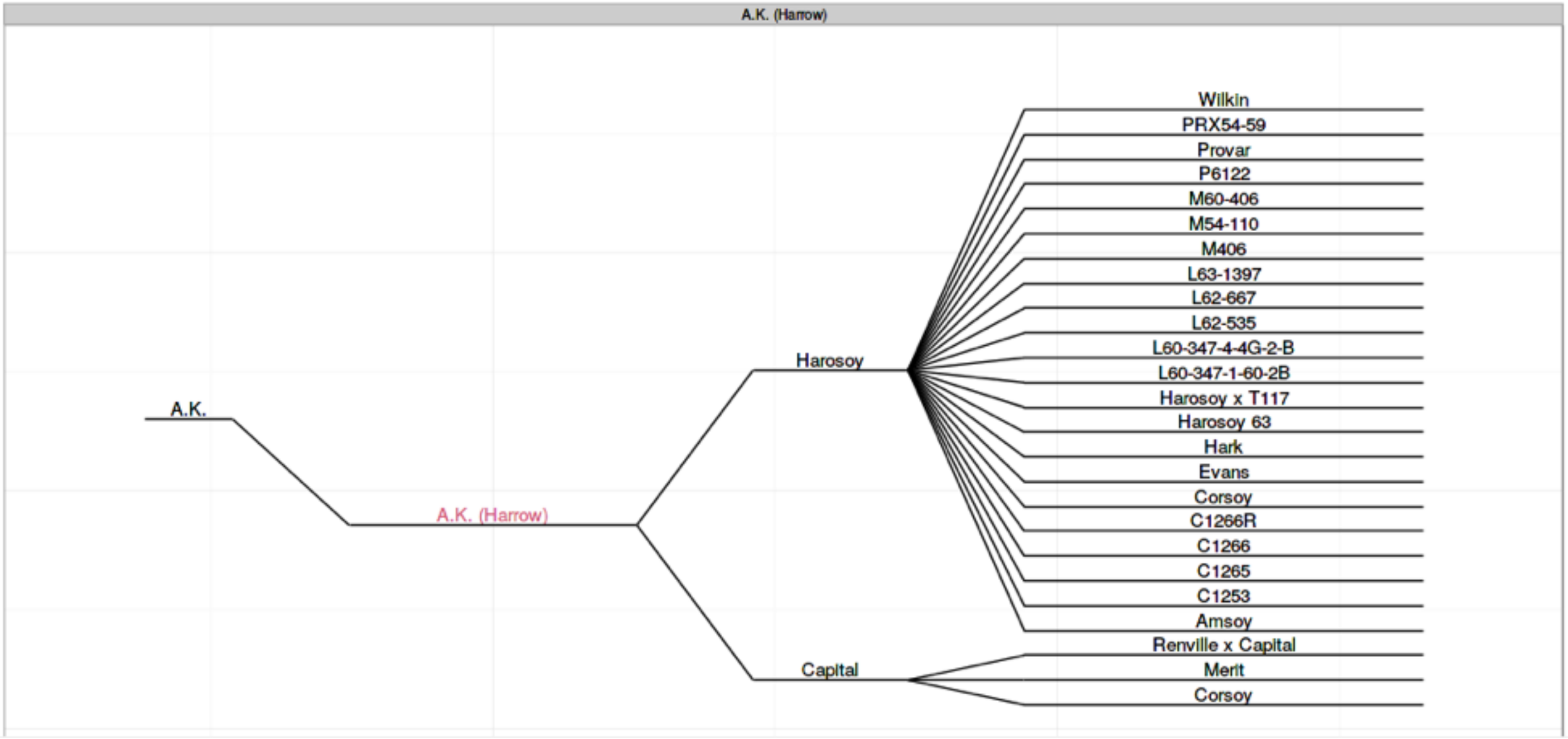
1 2 3 4 5 6 7 8 9 10

Type the name of the variety or click on the text box for options

Choose Varieties

A.K. (Harrow)

Variety List



Click on data points in the plot to see field trial data.

Yield, Protein, and Oil by Year

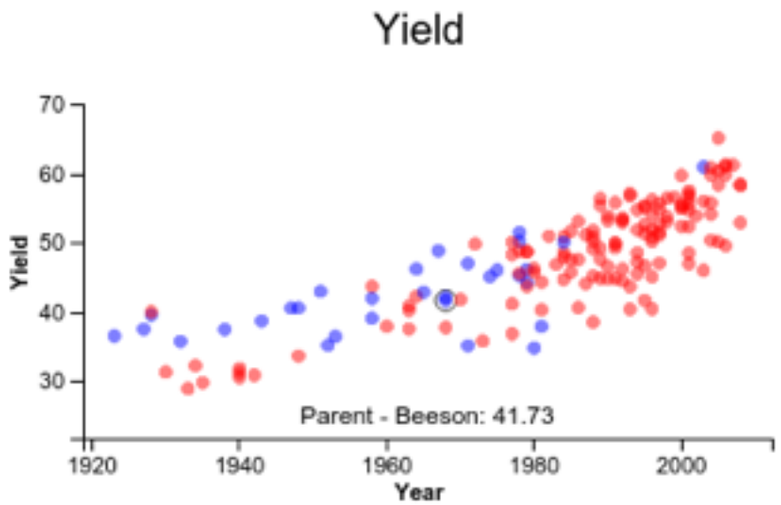
Show/Hide

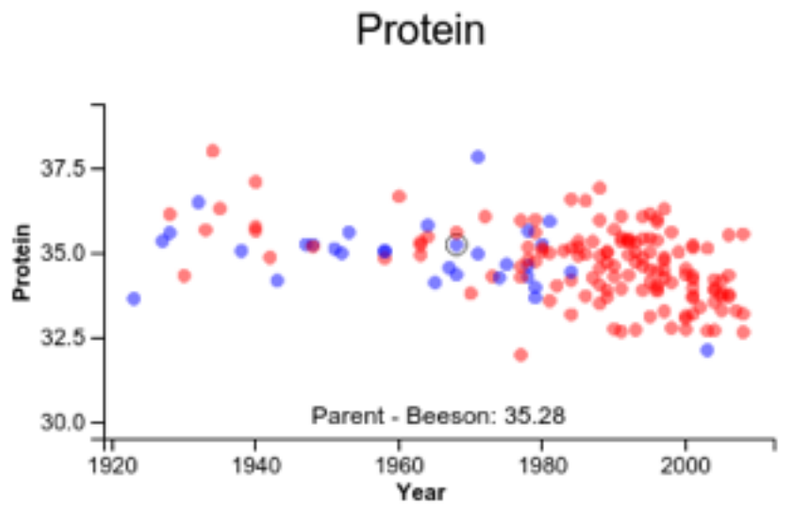
More Field Trial Data by Year

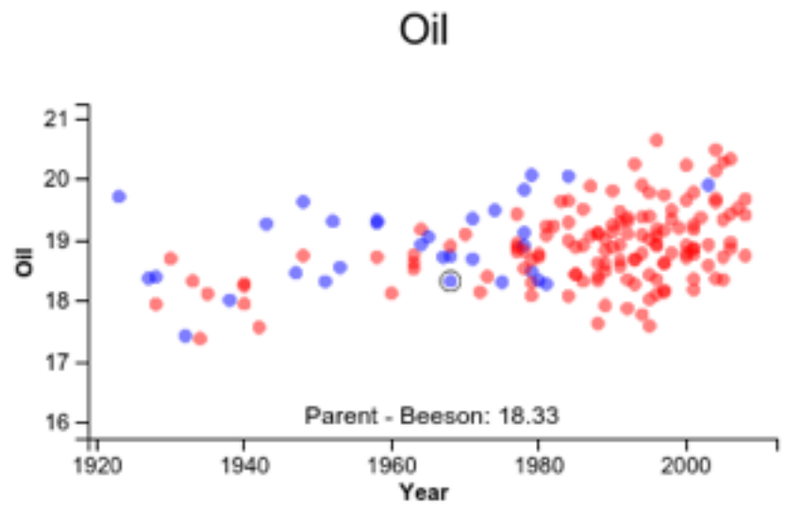
Maturity, Lodging, Seeds Show/Hide

Yield, Protein, Oil Pairwise Plots

Show/Hide







Click on data points in the plot to see field trial data.

Yield, Protein, and Oil by Year

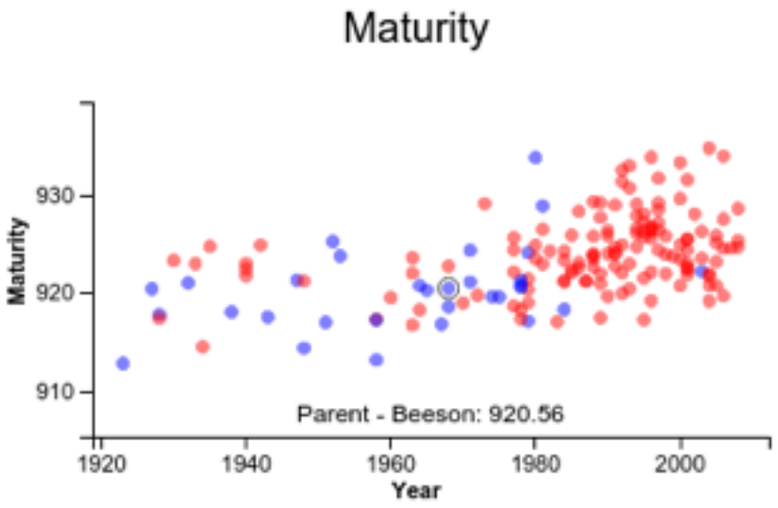
Show/Hide

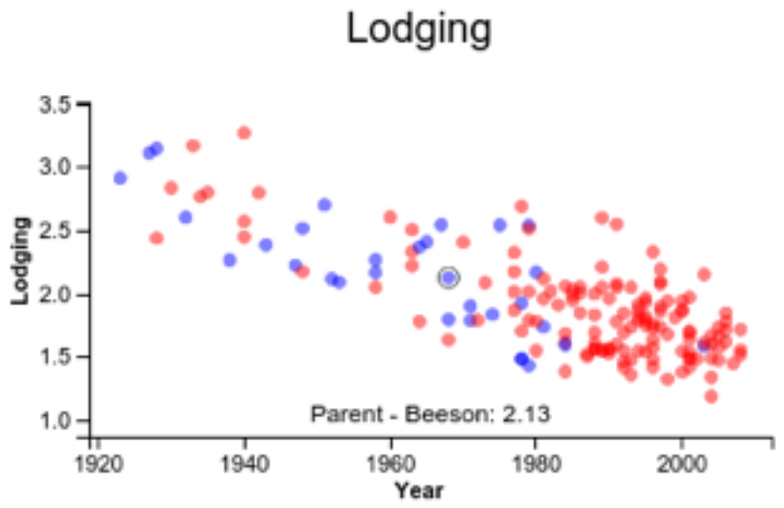
More Field Trial Data by Year

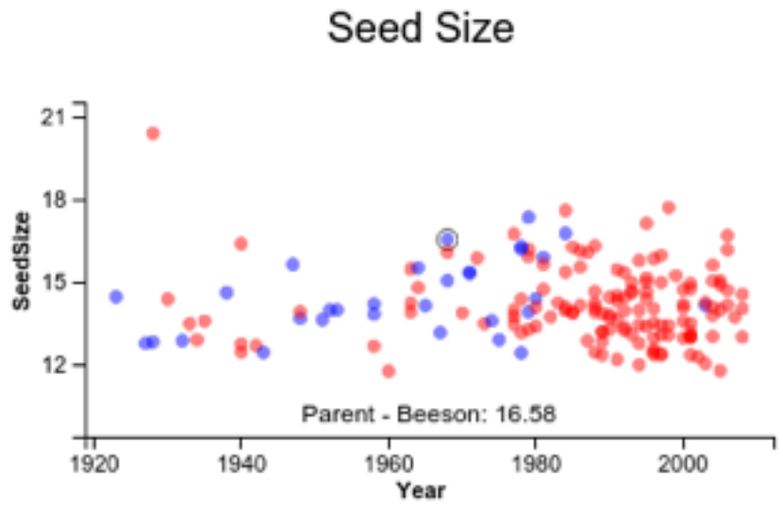
Maturity, Lodging, Seeds Show/Hide

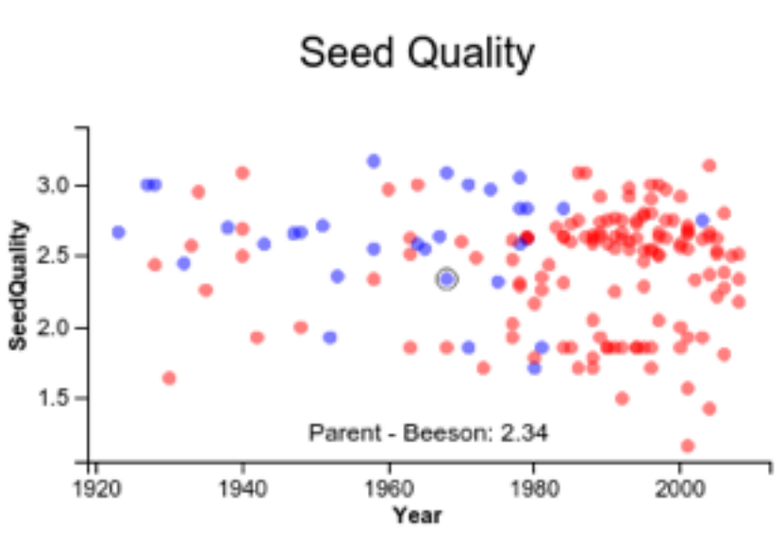
Yield, Protein, Oil Pairwise Plots

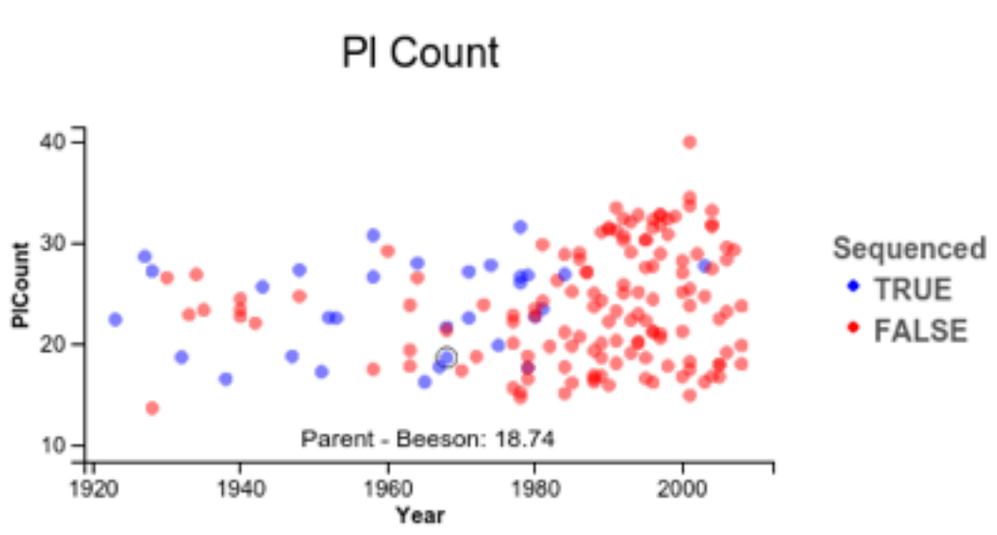
Show/Hide











Click on data points in the plot to see field trial data.

Yield, Protein, and Oil by Year

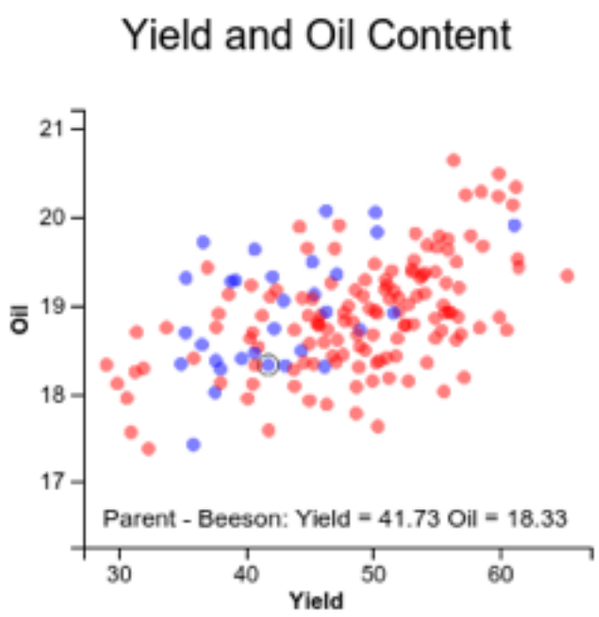
Show/Hide

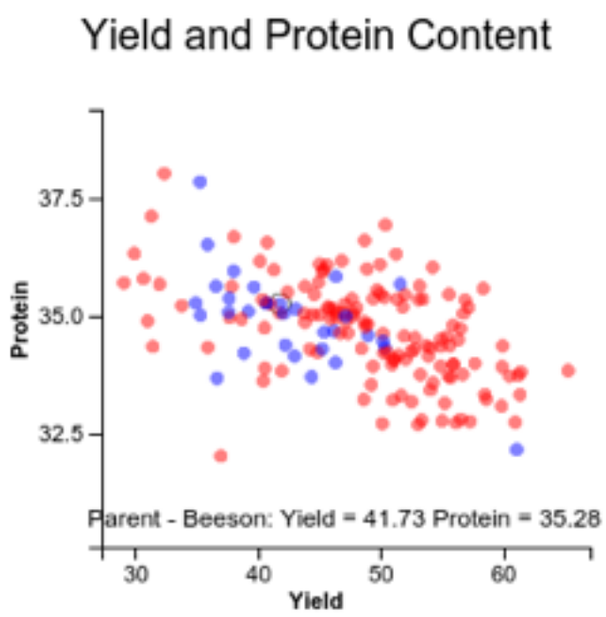
More Field Trial Data by Year

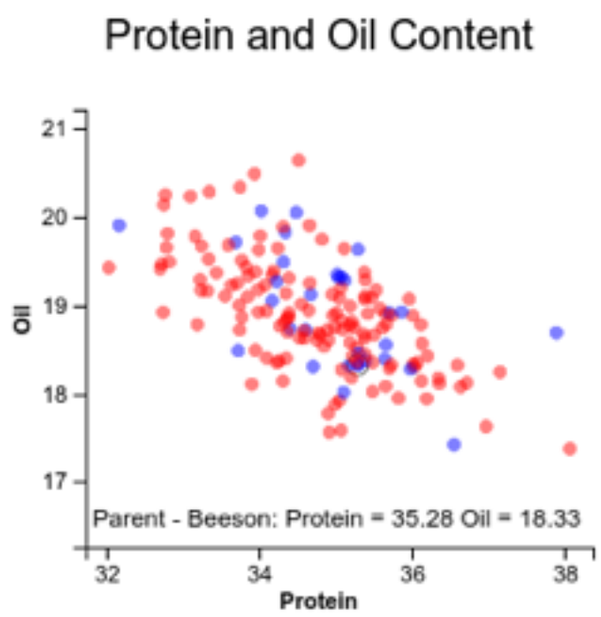
Maturity, Lodging, Seeds Show/Hide

Yield, Protein, Oil Pairwise Plots

Show/Hide







SNP app

- Five tabs containing different functionality
- Overview, SNP locations, search by id, SNP browser, SNP distance and the Kevin Bacon distance

- App written by Susan Vanderplas

GlymaID navigation

Ex: 01g004700 searches Chr 01 for ID 004700

Navigate by GlymaID: 01g004700 ?

 # SNPs shown: 20

Manual navigation

Choose chromosome and position

Choose Chromosome: Chr01

Start point ?: 441820

Matching Glyma IDs

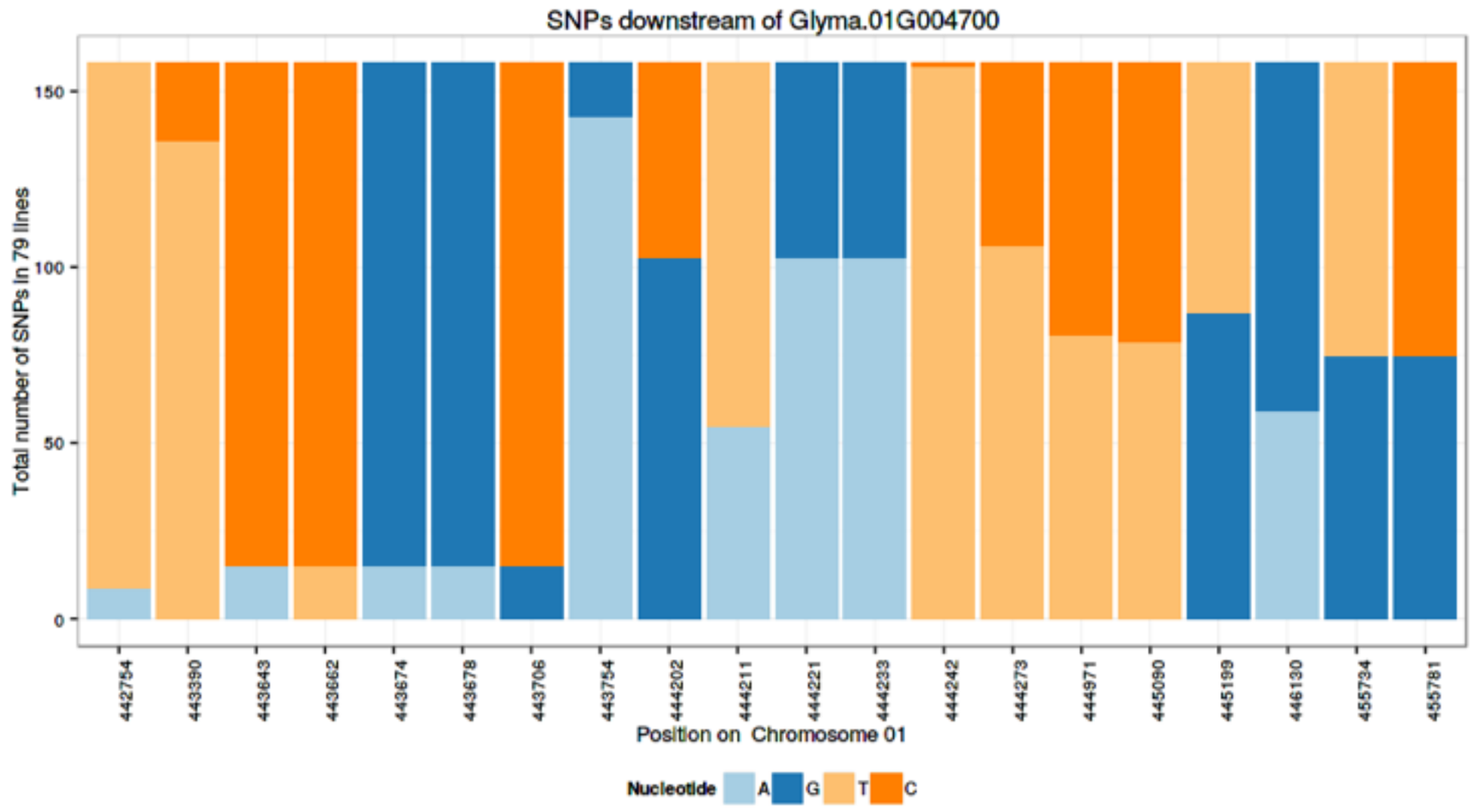
Show 10 entries

GlymaID	Start	End
01G004700	441820	445055

GlymaID | Start | End

Showing 1 to 1 of 1 entries

Previous 1 Next



Move upstream

Move downstream

GlymaID navigation

Ex: 01g004700 searches
Chr 01 for ID 004700

Navigate by GlymaID

01g004700 ?

Show Cultivar

A3127

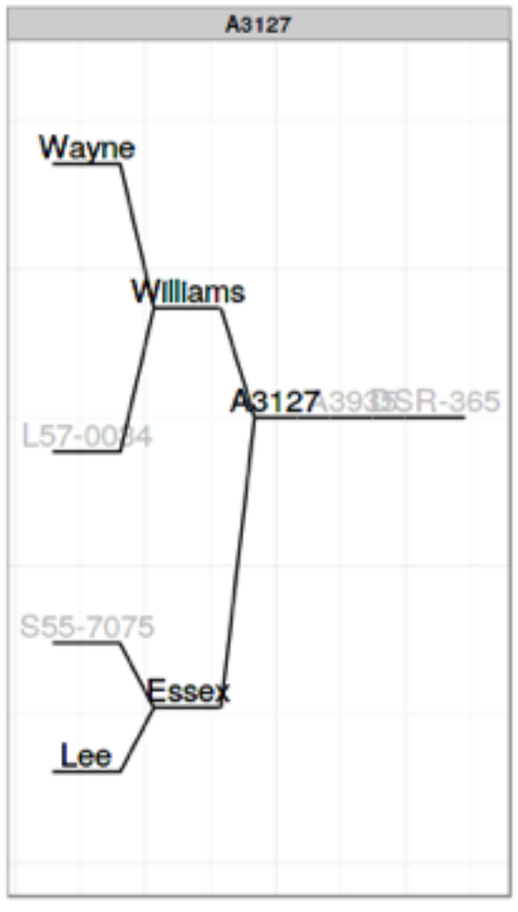
List of Cultivars

Show Ancestors
Show Descendants

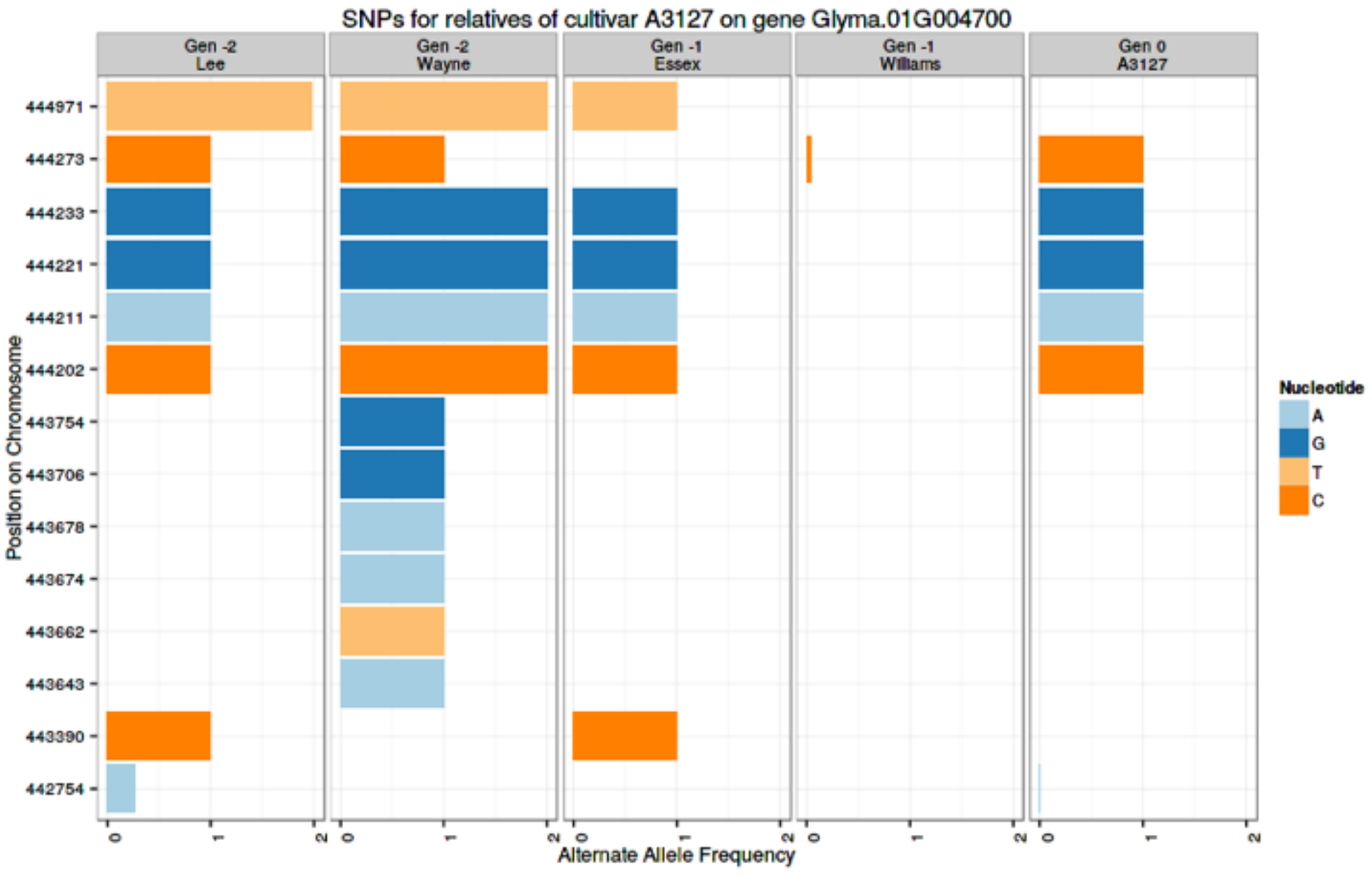
Generations

2

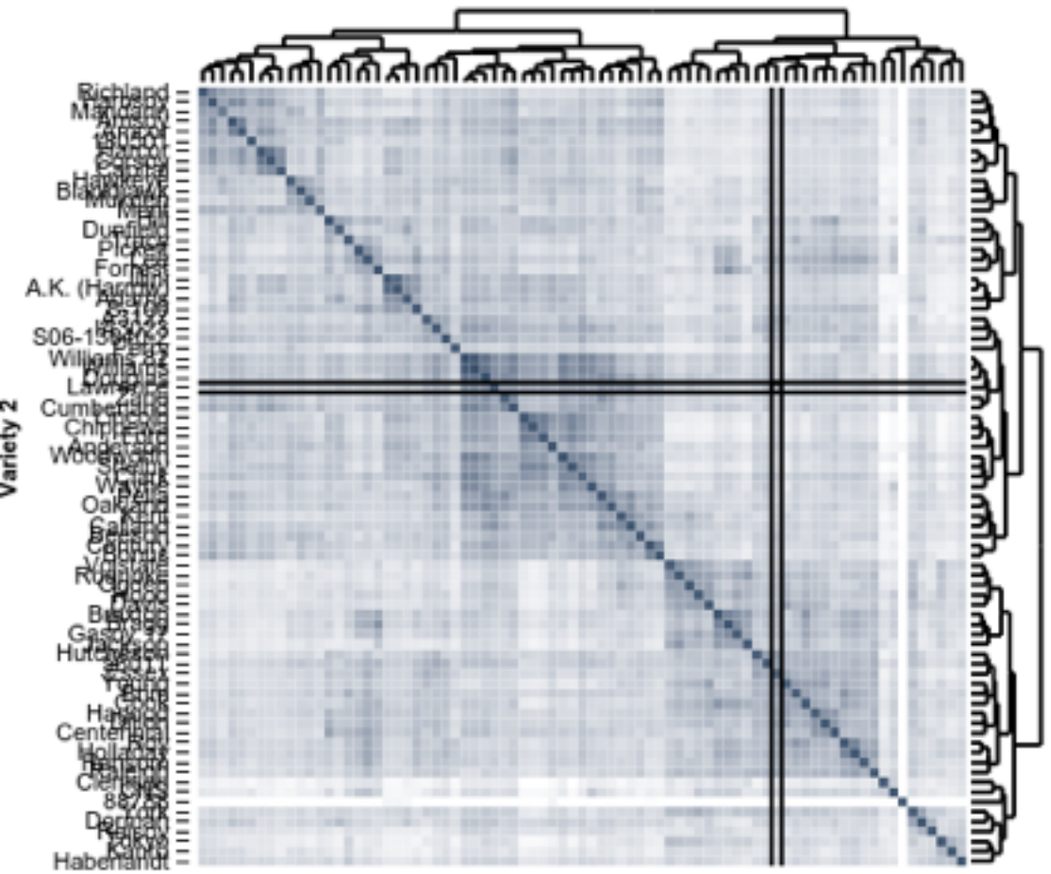
Genealogy



Data is available for cultivars shown in black.

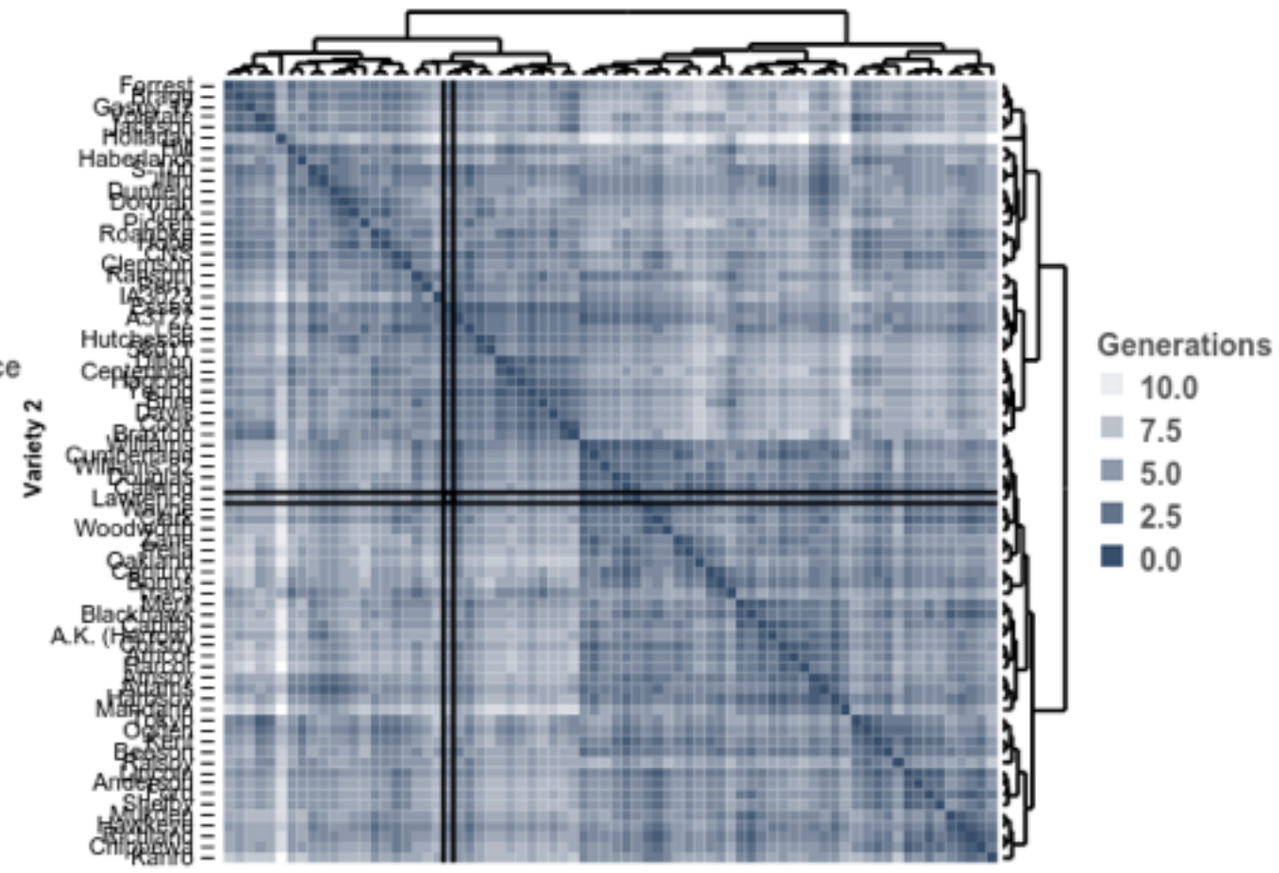


SNP Distance



Essex, Lawrence = 1.556

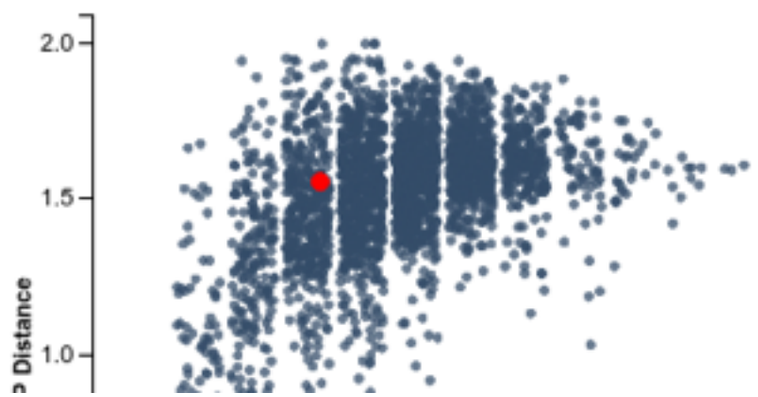
Generational (Kevin Bacon) Distance



Essex, Lawrence = 3 generations apart

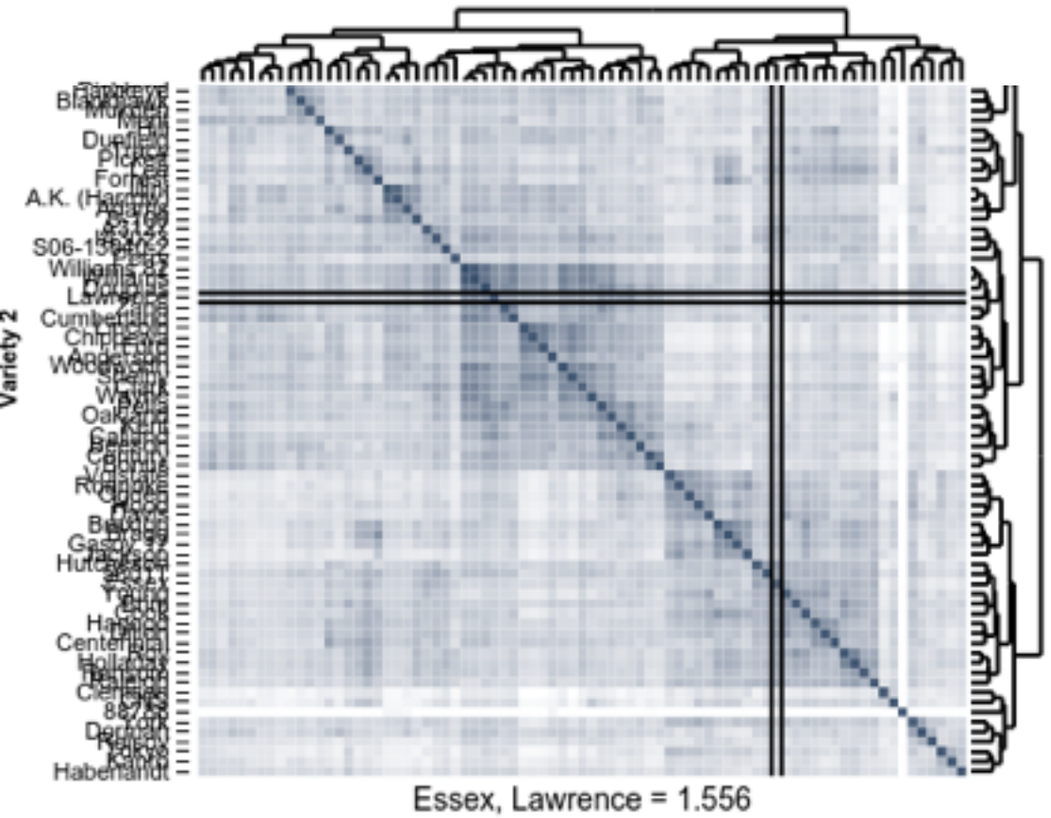
Shortest Path between Varieties

Generations and SNP Distance



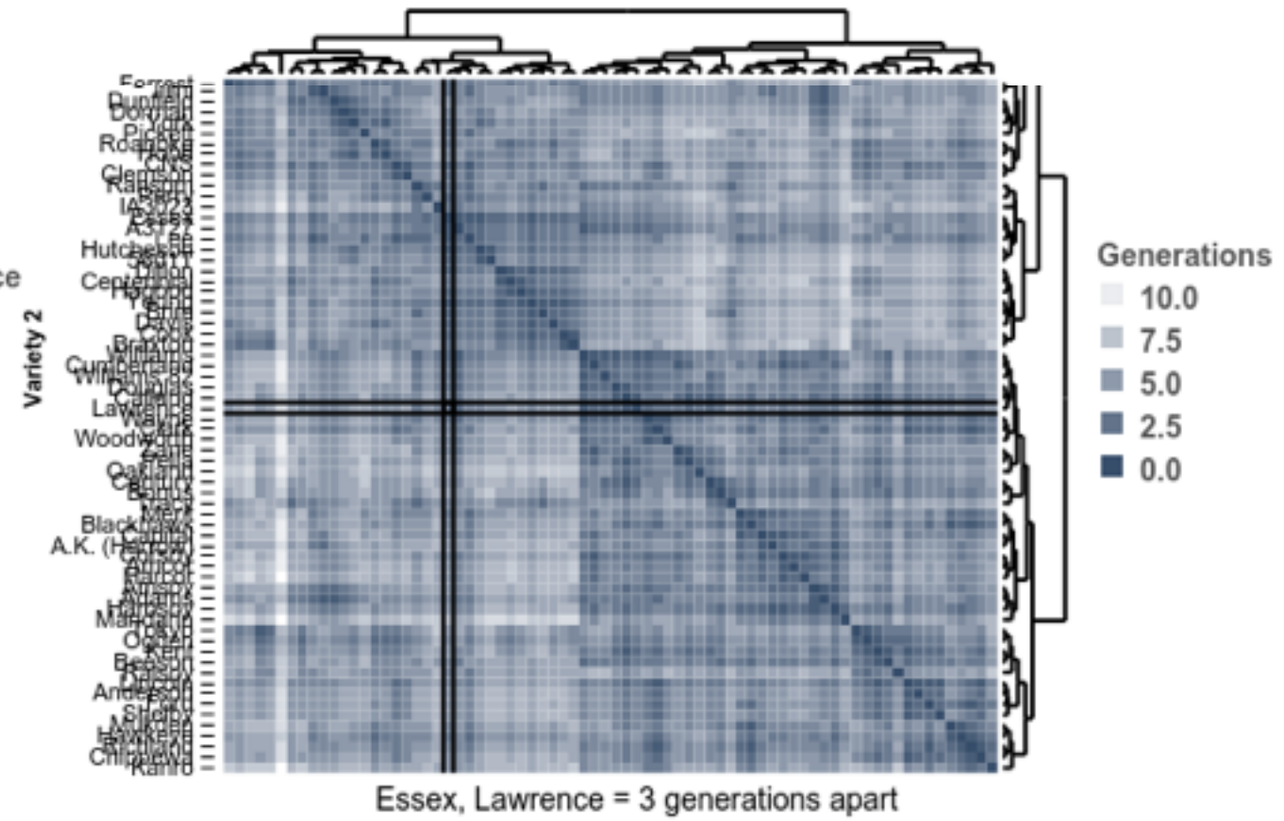
Yield Difference and SNP Distance

SNP Distance



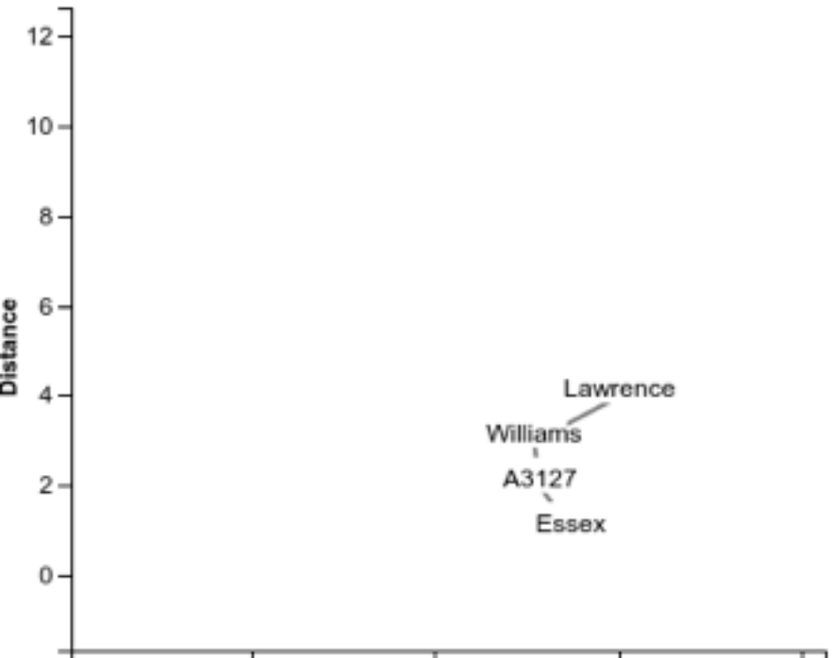
Essex, Lawrence = 1.556

Generational (Kevin Bacon) Distance

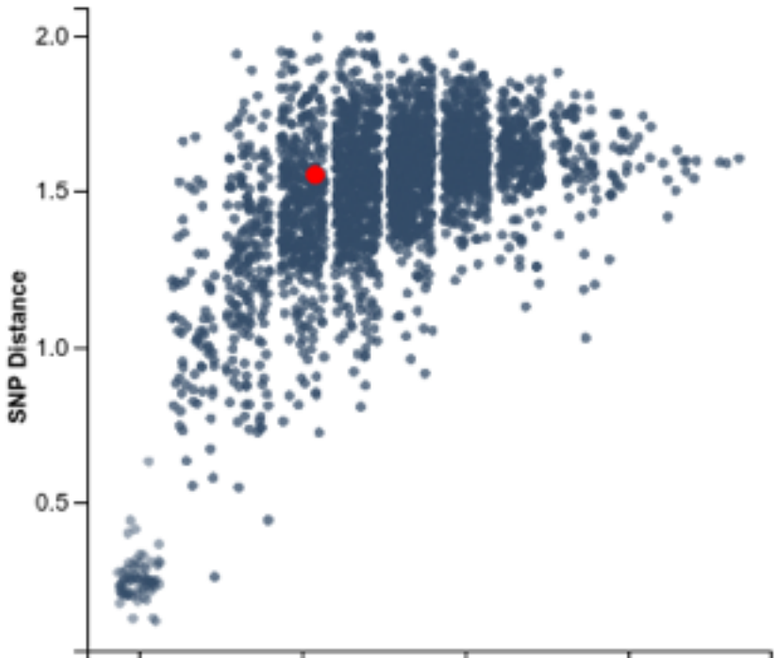


Essex, Lawrence = 3 generations apart

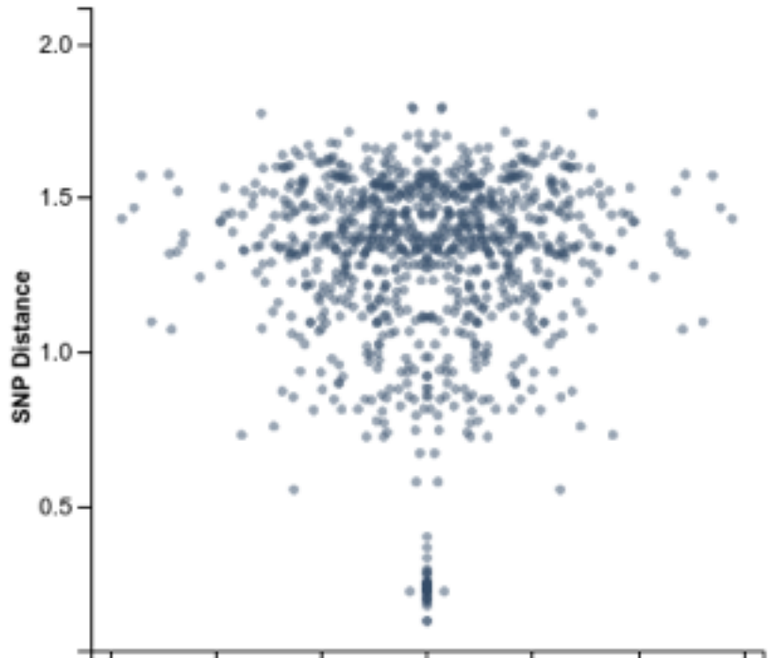
Shortest Path between Varieties



Generations and SNP Distance



Yield Difference and SNP Distance



This work is licensed under the Creative Commons Attribution-Noncommercial 3.0 United States License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/3.0/us/ or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.