



NCSU PEANUT BREEDING PROGRAM PROPOSAL

**Marker-Assisted Selection in Virginia-Type
Peanut for Multiple Disease Resistance**

2020-0236

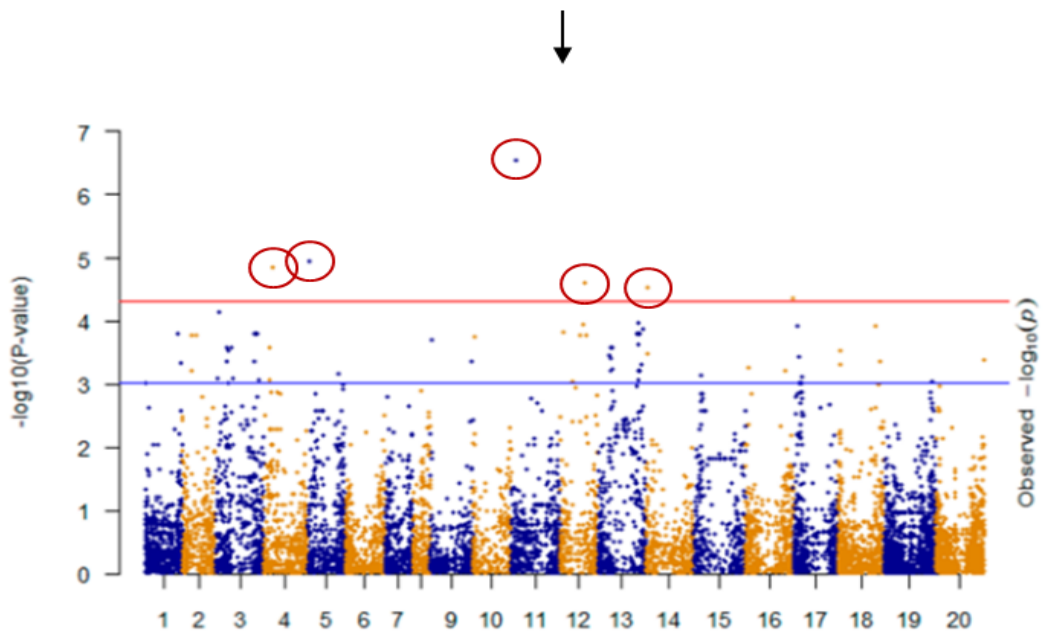
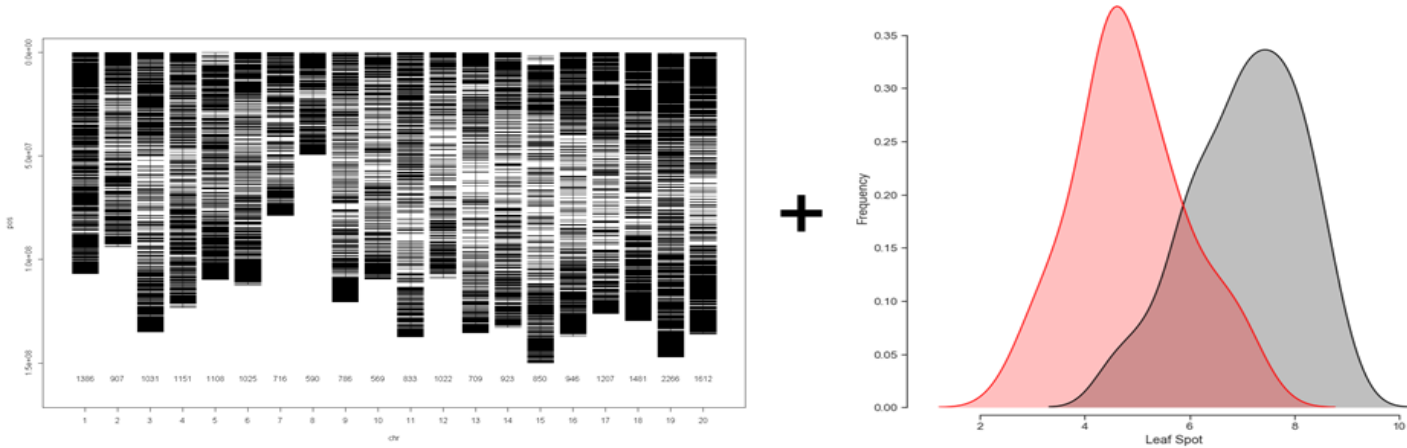
Marker-Assisted Selection Report

- COVID-19 Update on Field Program
- Marker-Assisted Selection Methodology
 1. Seed Chipper/Corer
 2. Crude DNA Isolation
 3. Multiplexing Marker Technology
- Impacts of Methodology
- Greenhouse Speed Breeding

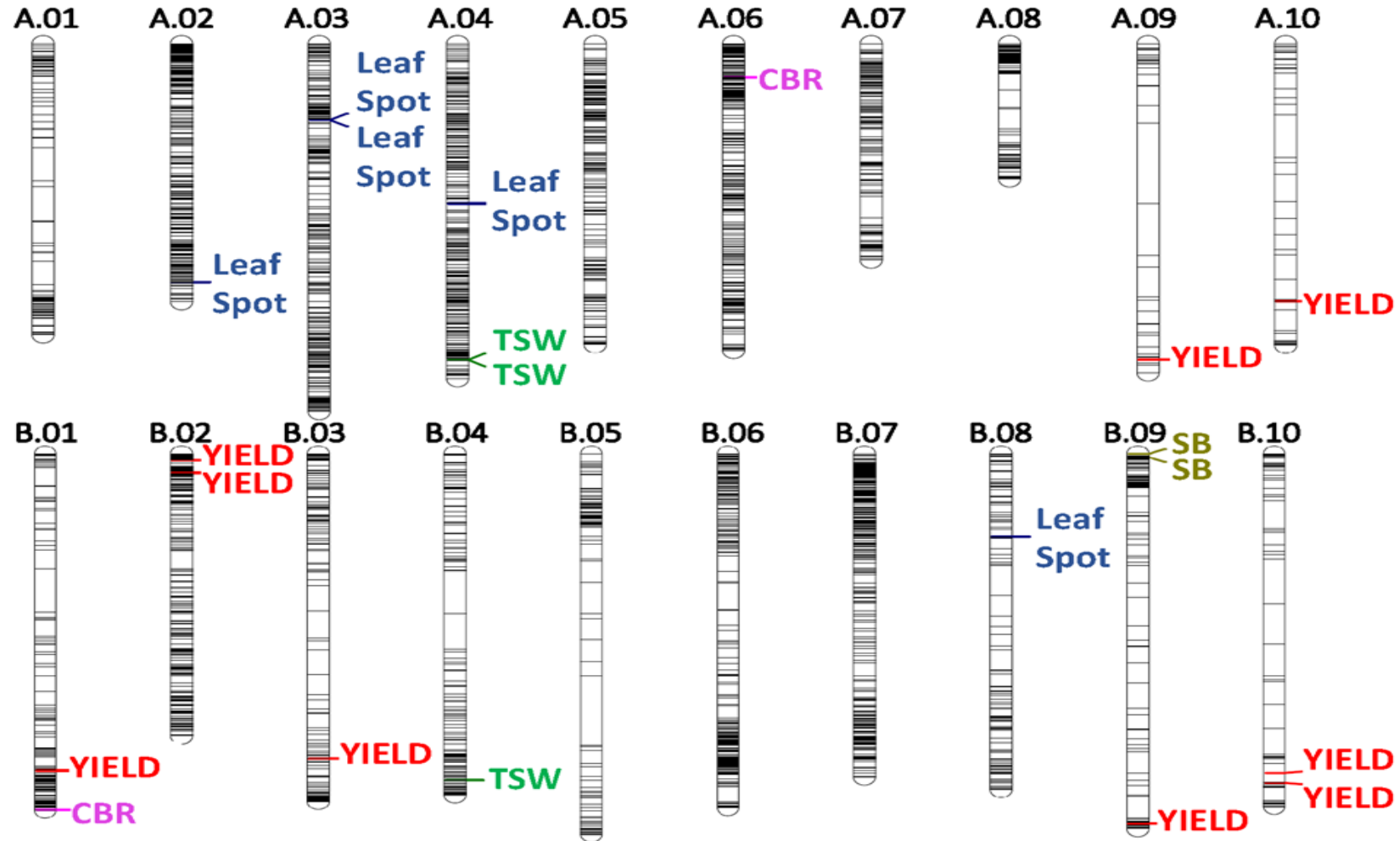
Peanut Breeding Program – Marker Use

1. Verify all parental materials used in crossing block
2. Confirm all crosses (F_1 Plants)
3. Genotype all early generation materials (F_2 through F_5) at large effect markers for marker-assisted selection

Genome-Wide Association Analysis



Marker-Assisted Selection in Peanuts



Identify Marker Trait Associations

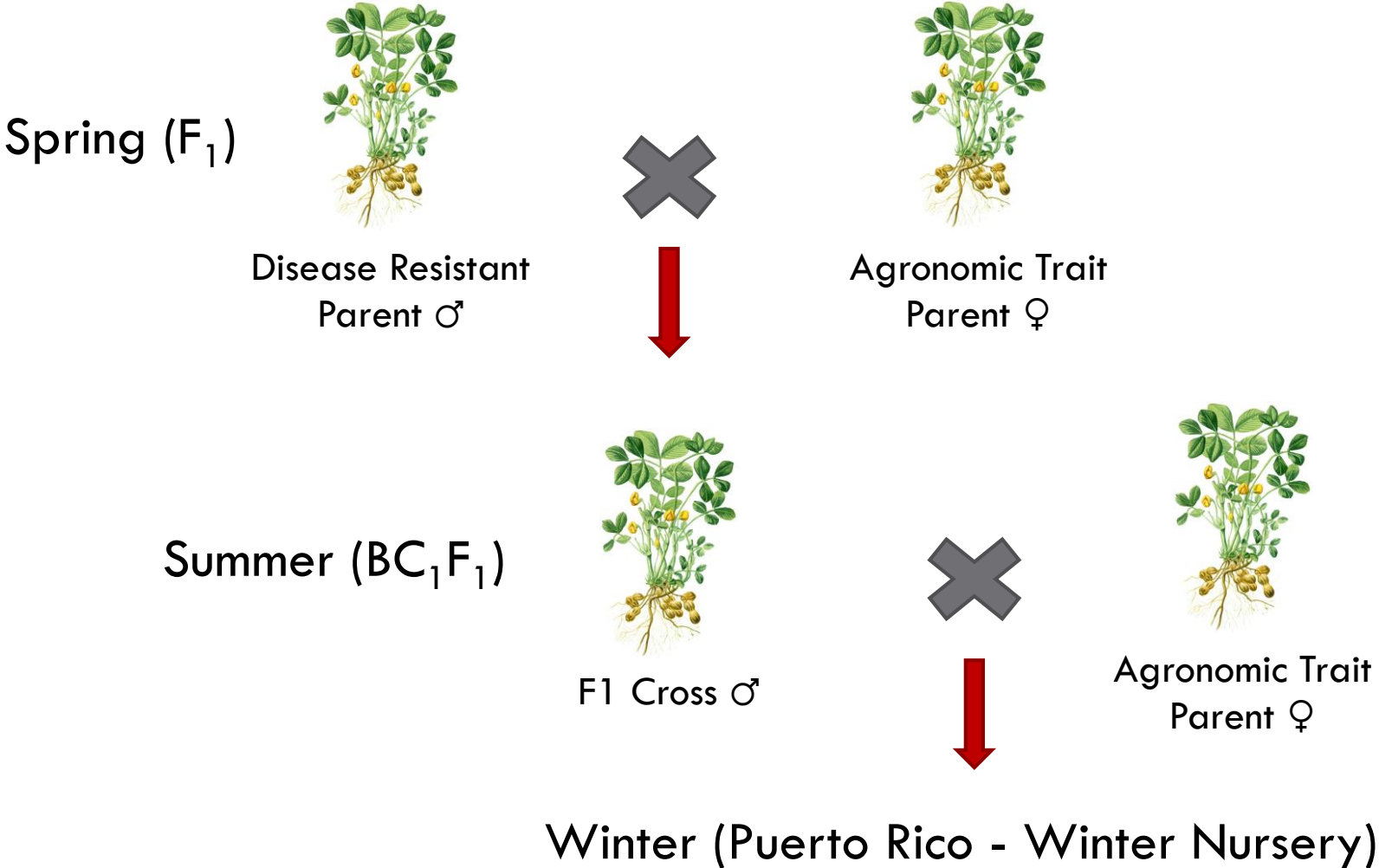
GWAS or QTL Analysis Results

Trait*	A. Hypogaea Chromosome	A. Hypogaea Physical Position (bp)	GWAS Adjusted r2	Frequency of Desirable Allele	Frequency of Undesirable Allele
PY	10	101,147,237	0.07	0.9	0.1
PY	19	153,897,034	0.07	0.61	0.39
PY	8	34,876,837	0.08	0.79	0.21
LS	2	94,810,239	0.1	0.93	0.07
LS	3	30,614,262	0.02	0.97	0.03
LS	18	32,865,575	0.13	0.06	0.94
CBR	13	123,937	0.07	0.5	0.5
CBR	11	148,970,476	0.07	0.2	0.8
SB	19	985,995	0.12	0.5	0.5
TSWV	14	135,149,720	0.06	0.09	0.91
FAD2B	19	154,049,242	N/A	0.89	0.11

* PY - Pod Yield, LS - Leaf Spot, CBR - *Cylindrocladium* black rot, SB - *Sclerotinia* blight, TSWV - Tomato spotted wilt virus, and FAD2B - Gene providing the high-oleic trait

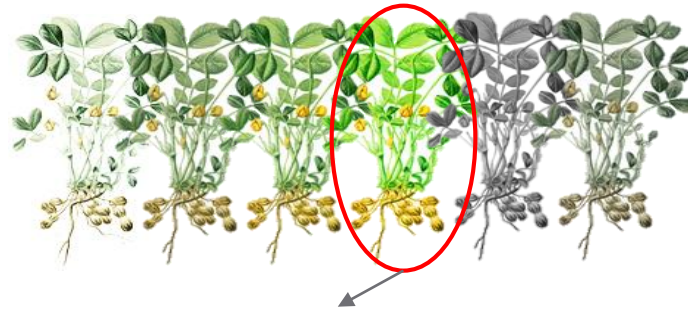
Other Traits? Flavor, Folate, etc.

Greenhouse Crossing Program

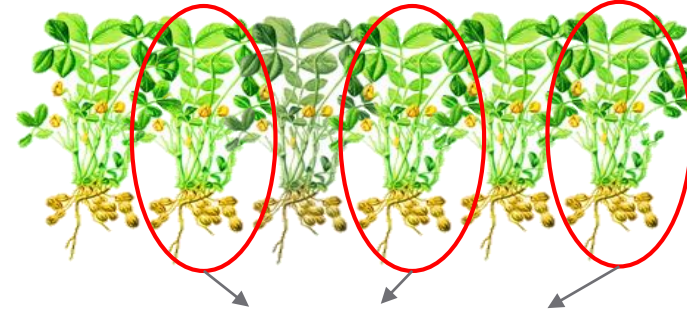


Marker-Assisted Selection in Peanuts

F2 Among and Within Family Selection



Generational Pipeline



Advanced Disease Pipeline

~ 200 Family Lines

F3 Single Plant Selections

~ 350 Family Lines

F4 Single Plant Selections

~ 500 Family Lines

F5 Single Plant Selections

~ 400 Family Lines

F6 Single Plant Selections

~ 250 Family Lines

F7 Bulk Line Selections

~ 150 Family Lines

Preliminary Yield & Disease Screening

Winter Nursery

Progeny Disease Testing + MAS

F_{2:4}

Winter Nursery

Single Plant Selection

F_{4:6}

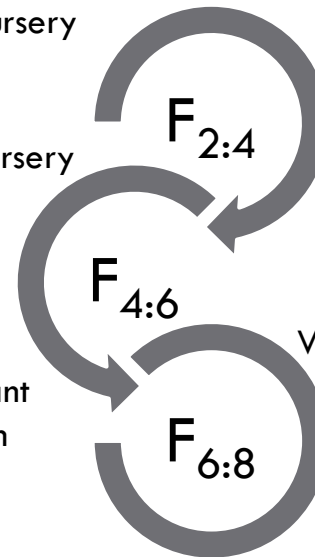
Progeny Disease Testing + MAS

Single Plant Selection

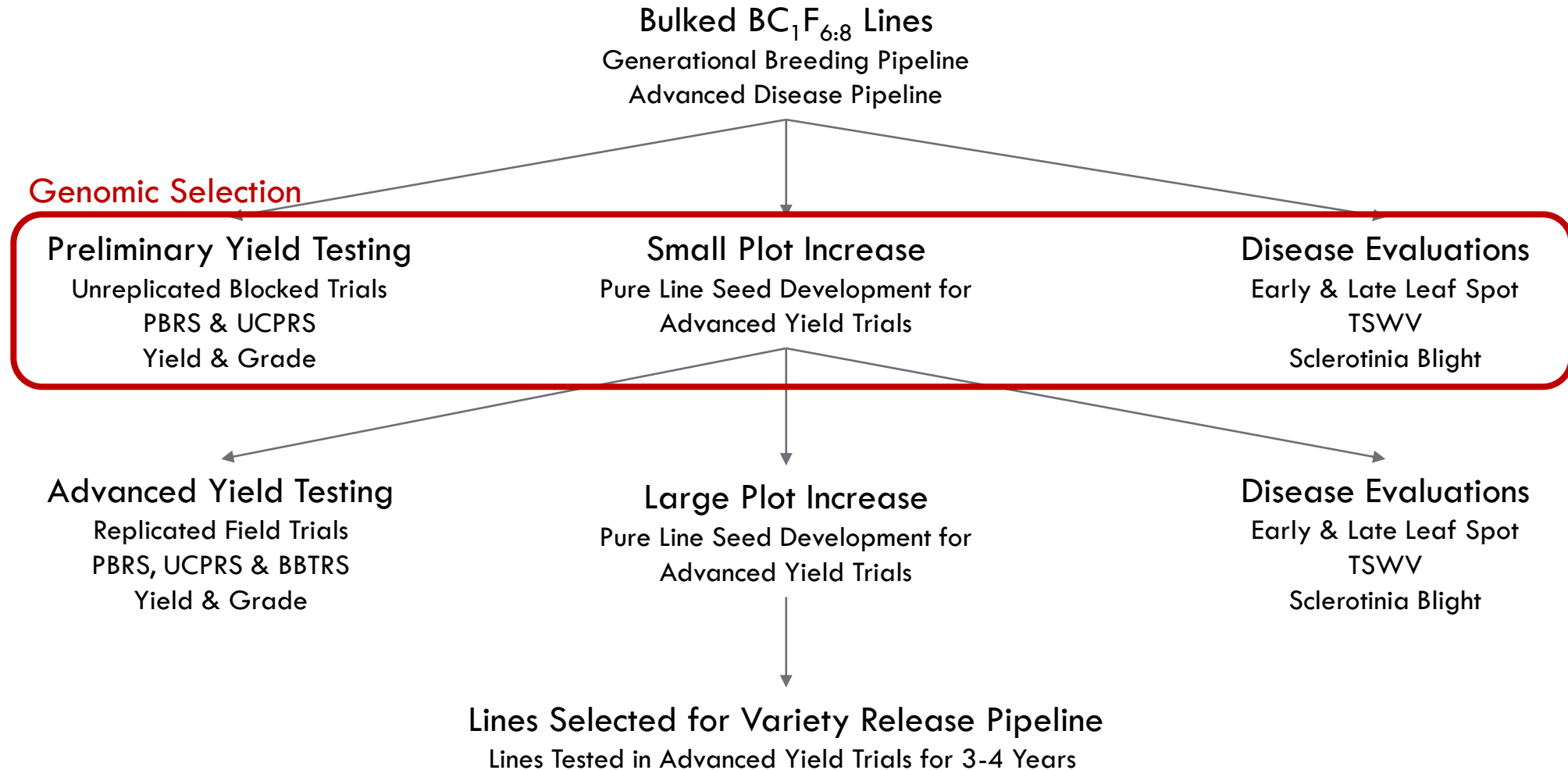
Winter Nursery

F_{6:8}

Bulk Line Harvest



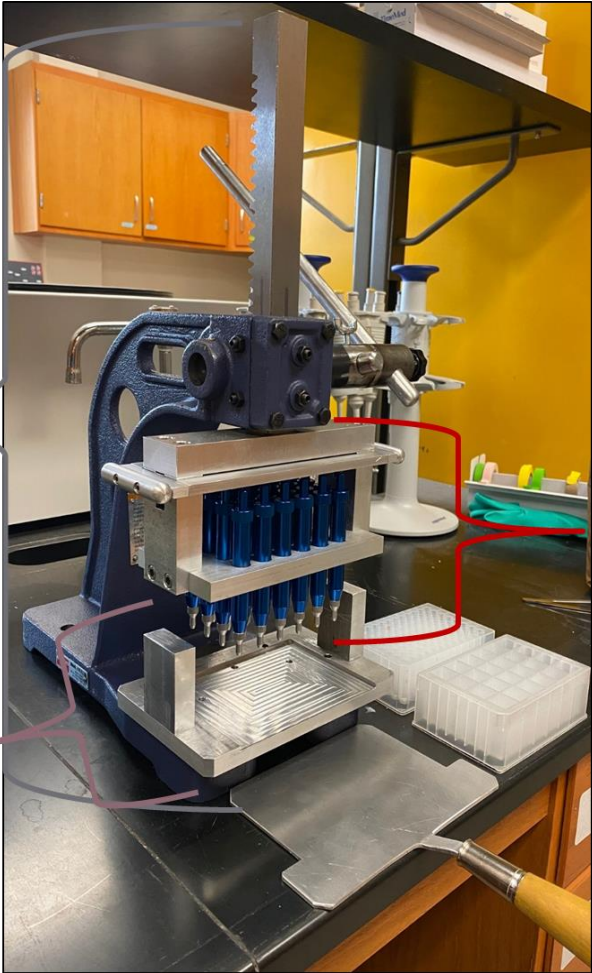
Program Testing Network



Marker Assisted Selection Methodology

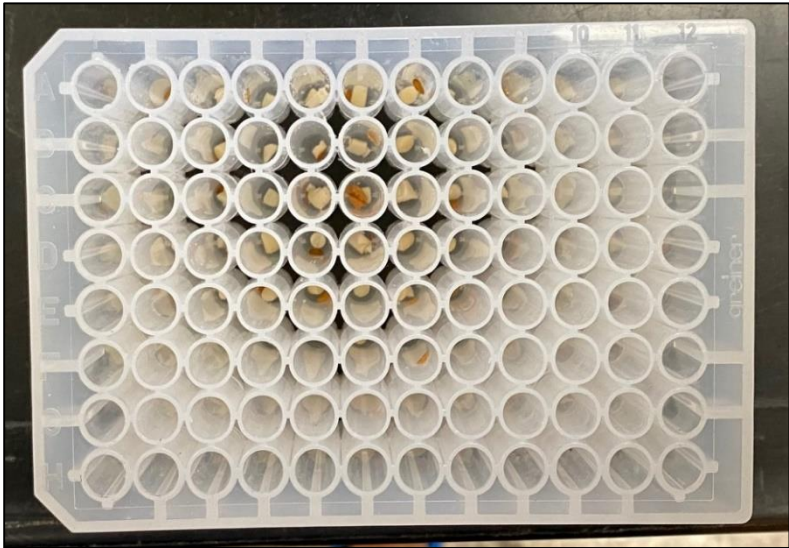
1. Manual Seed Chipper > Sample 384 seeds rapidly
2. Standard “crude” DNA extraction from 384 samples in one hour at minimal costs
3. Custom multiplex SNP genotyping

Seed Chipper – Monsanto Patent Avoided



3. Custom-built base

2. 24 modified biopsy punches in fixed positions (4 x 6 layout) with a custom-built single ejector



Crude DNA Isolation

1. Spin briefly in plate centrifuge.
2. Add 50 μ L Buffer A to each well and mix by inversion.
3. Heat at 65 $^{\circ}$ C for 10 minutes.
4. Add 200 μ L Buffer B and mix by inversion
5. Spin briefly and transfer 60 μ L of supernatant to a new plate
6. Use 2 μ L in PCR reaction, no need to quantify DNA



Buffer A:

Stock	Volume	Final Concentration
10 M NaOH	200 μ L	100mM
Tween 20	400 μ L	2%
Ultrapure H ₂ O	19.4mL	N/A

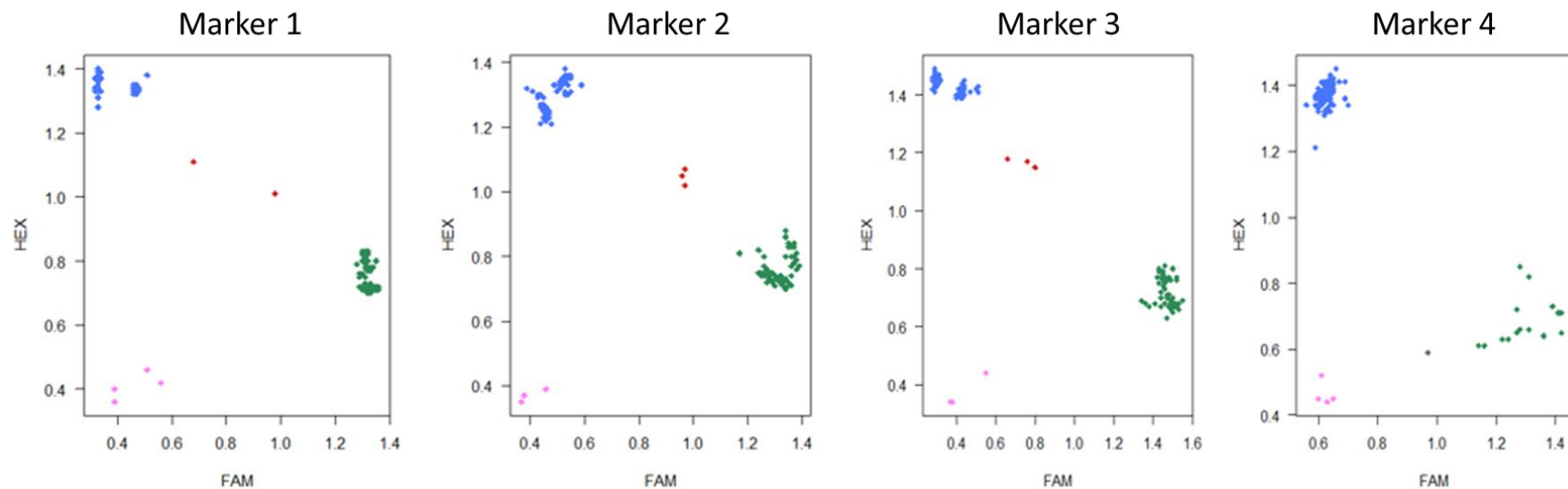
Buffer B:

Stock	Volume	Final Concentration
1 M Tris-HCl#	8mL	100mM
0.5 M EDTA	320 μ L	2mM
Ultrapure H ₂ O	72mL	N/A

KASP (or PACE) Assay

Most Marker Behave Similarly...

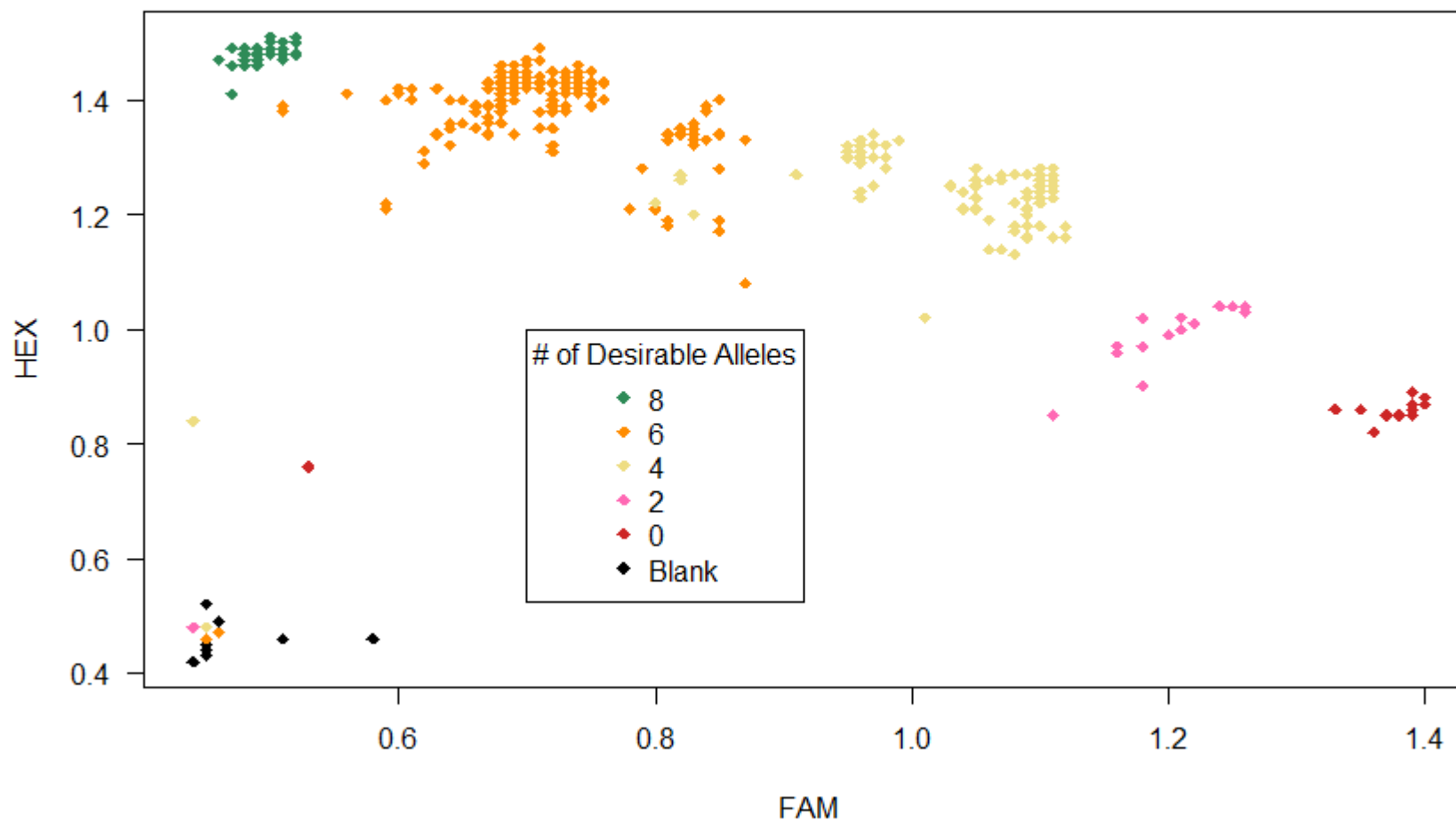
- Markers can often be grouped based on their fluorescence pattern
- What would happen if you arranged the beneficial allele on the same axis across all assays and then pooled the beneficial markers together?



Marker-Assisted Selection

Inbreds w/ high-quality DNA isolated with Qiagen kit, quantified and normalized

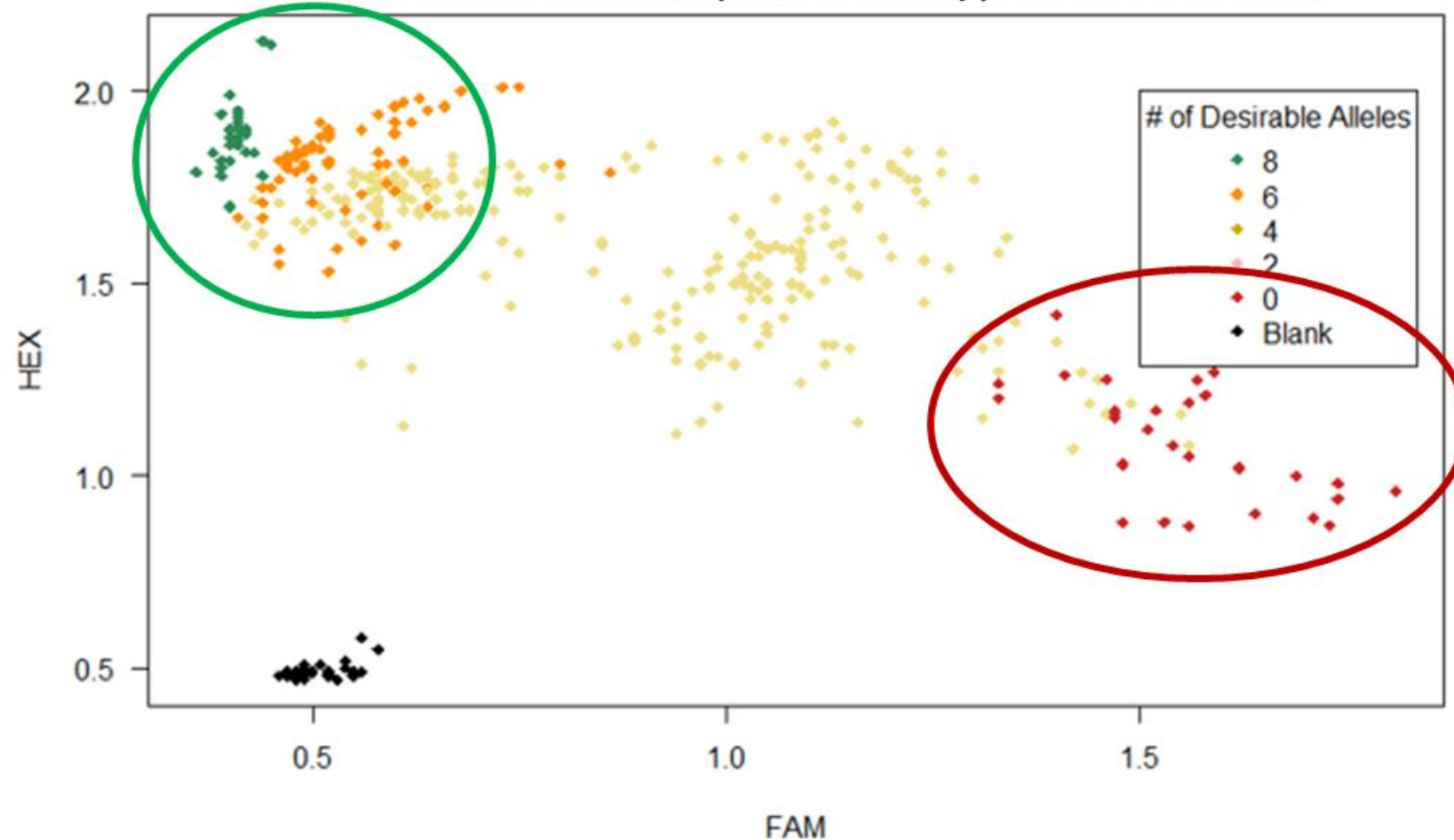
Signal vs. # of Desirable Alleles in a 4-Marker Pool on a Panel of Inbreds



Marker-Assisted Selection

Inbreds w/ crude DNA isolation – not quantified or normalized

Panel of Inbreds Multiplex Genotyped at Four Loci



Impact of Methodology

Pros:

- Reduced genotyping costs
- Increased speed
- Simplified data analysis and selection – selection index just HEX/FAM ratio

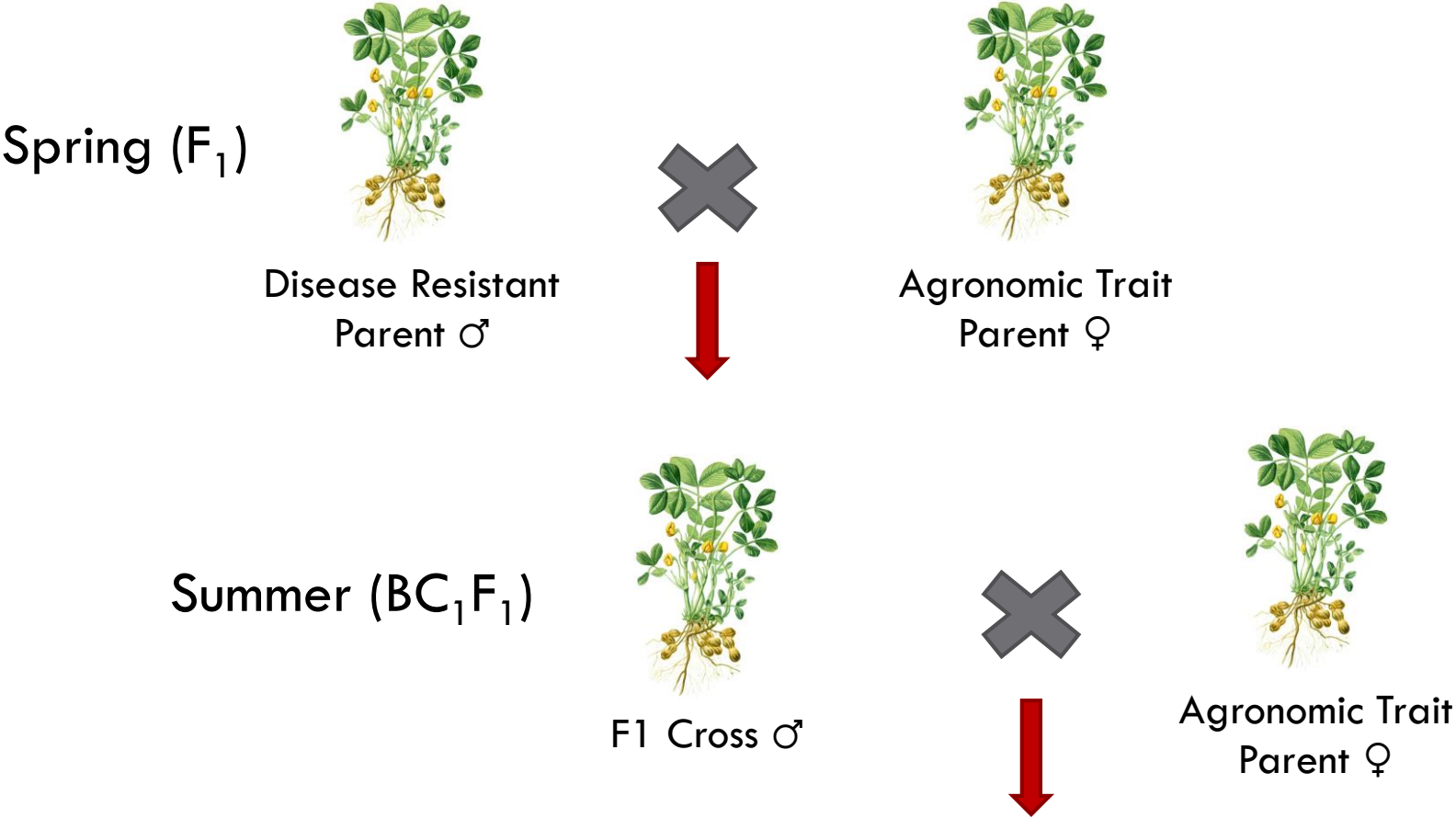
Verify all parental materials used in crossing block

Confirm all crosses (F_1 Plants)

Genotype all early generation materials (F_2 through F_5) at large effect markers for marker-assisted selection

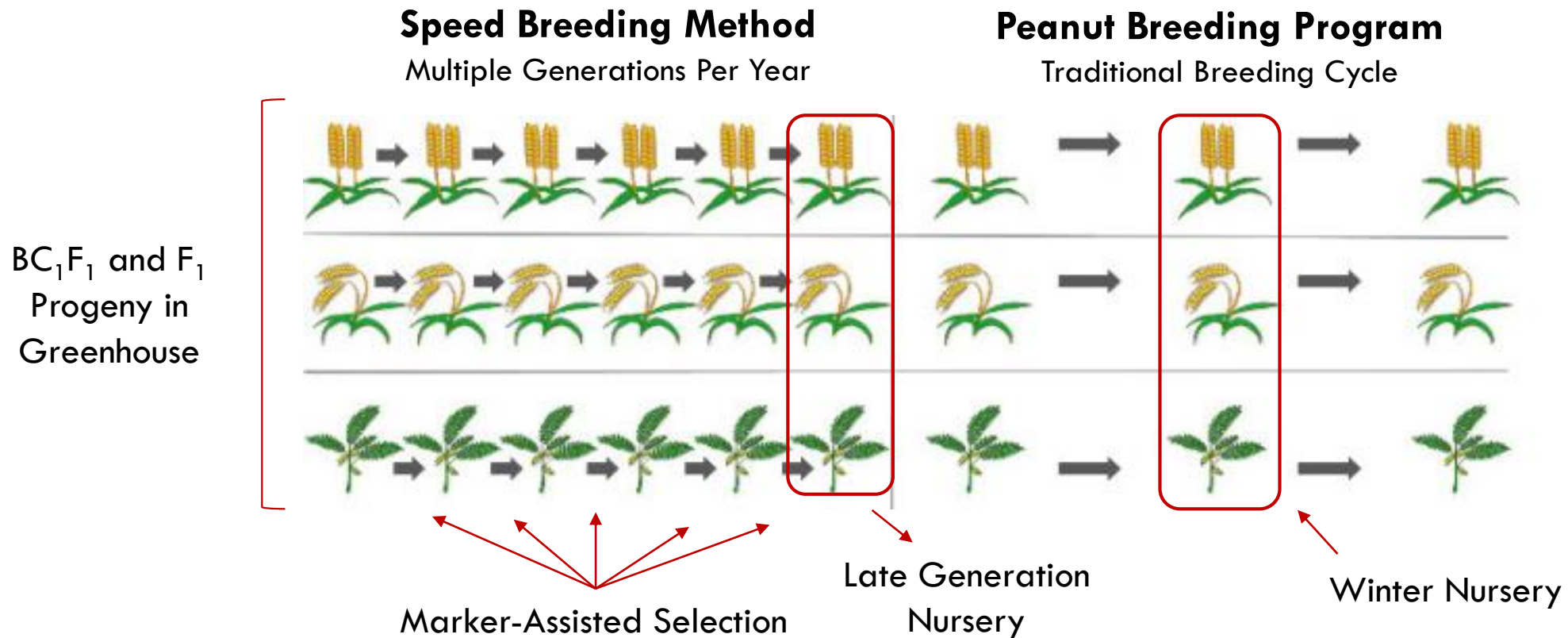
- ❖ Ideally adopt a speed breeding approach to turn 3+ generations per year in the greenhouse
- ❖ Concentrate field resources on genomic selection and testing later generation materials

Greenhouse Speed Breeding w/ MAS

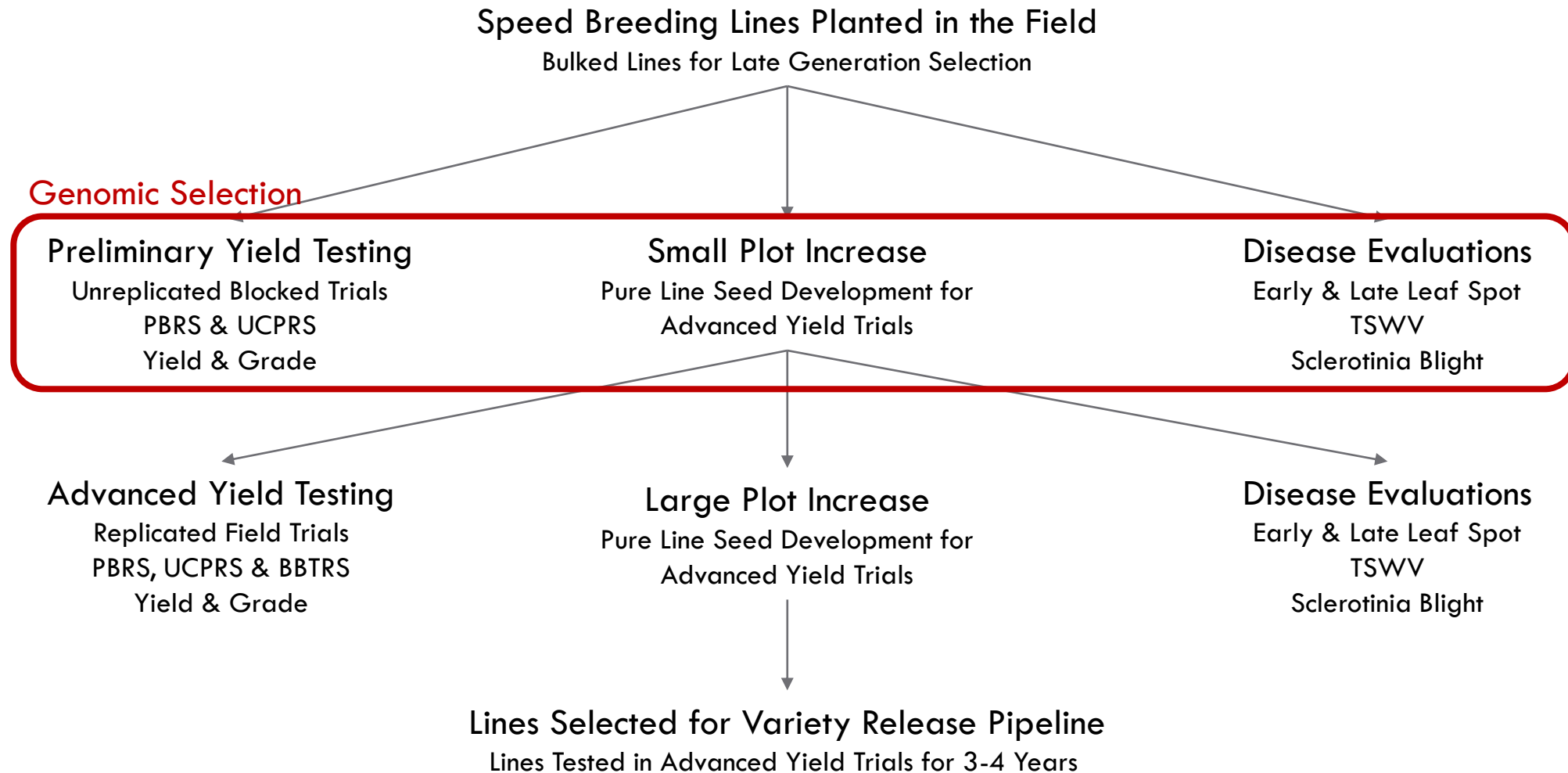


~~Winter (Puerto Rico - Winter Nursery)~~

Greenhouse Speed Breeding w/ MAS



Greenhouse Speed Breeding w/ MAS



2021 Budget

FUNDS REQUESTED:	2021
EPA Salaries and fringe benefits¹	
Graduate Student Stipends and Support²	\$21,000
SPA Salaries and fringe benefits¹	
Part-time Labor (including: 8.60% fringe benefits, and add \$142/year health insurance for hourly / part-time workers who will work ≥ 30hrs/wk. for longer than 3 months)	\$5,000
Supplies (such as: lab and field consumables, fuel, etc.)	\$2,500
Equipment (greater than \$500)	
Travel (for airfare, overnight lodging and meals associated with project tasks)	
Other (such as: motor pool rental, lab analysis, contracted services, greenhouse space, page/publication charges)	
TOTAL in Dollars	\$28,500



NCSU PEANUT BREEDING PROGRAM PROPOSAL

Marker Development through Next-generation
Sequencing (NGS) for Late Generation Selection

NC-42

Next-Generation Sequencing Report

Protocol Development Steps

1. Development of internal reference *A. hypogaea* cv. 'Bailey II'
2. Whole-genome sequencing of diverse breeding lines in NCSU peanut breeding program
3. Align whole-genome sequences to internal reference for single nucleotide polymorphism (SNP) discovery
4. *In silico* digest of *Arachis hypogaea* cv. 'Bailey II' reference genome
5. Use *in silico* site information (alignments) to maximize SNP recovery

Bailey II – Internal Reference Genome

Table 1. Comparison of genome assemblies based on summary statistics compiled for Bailey II, Tifrunner (both versions), Fuhuasheng and Shitoqi for use as internal reference genome for the North Carolina State University peanut breeding and genetics program.

Assembly	Scaffold Total	Contig Total	Max Scaffold Length	Max Contig Length	% Main Genome in Scaffolds	BUSCO*
<i>Bailey II (CANU)</i>	89	1,334	90,278,000	61,544,000	100.00%	5,192
<i>Tifrunner (v.1)</i>	384	4,039	160,880,000	9,488,000	99.79%	5,182
<i>Tifrunner (v.2)</i>	442	4,139	160,028,000	9,488,000	99.78%	5,183
<i>Fuhuasheng (v.1)</i>	20	31,747	168,161,000	1,735,000	100.00%	5,078
<i>Shitoqi (v.1)</i>	21	7,747	159,155,000	8,551,000	100.00%	5,178

* Measure of completeness of genome

Whole-Genome Sequencing

Table 2. List of 48 lines subjected to whole-genome sequencing (10x coverage) and SNP discovery using the Bailey II internal reference genome for alignment.

(EV) Bailey*	(EV) Perry*	(GL) GP-NC 343*	(MC) GP-NC WS 3	(MC) GP-NC WS 14	(MD) HTS IL-028*
(EV) Bailey II*	(EV) Sugg*	(GL) PI 121067	(MC) GP-NC WS 4	(MC) GP-NC WS 15	(MD) HTS IL-029*
(EV) Emery*	(EV) Sullivan*	(GL) PI 269685	(MC) GP-NC WS 5	(MC) GP-NC WS 16	(MD) HTS IL-047*
(EV) N05006	(EV) Wynne*	(GL) PI 270806	(MC) GP-NC WS 6	(MC) HTS 16-03*	(MD) HTS IL-049*
(EV) N11055B*	(EV) Gregory*	(GL) PI 576636*	(MC) GP-NC WS 7	(MC) HTS 16-04*	(MD) HTS IL-051*
(EV) N13042ol*	(FV) NC 5	(GL) PI 665000	(MC) GP-NC WS 11	(MC) HTS 16-06*	(MD) HTS IL-052*
(EV) N16021*	(FV) NC 6	(MC) GP-NC WS 1	(MC) GP-NC WS 12	(MC) SPT 07-01*	(MD) HTS IL-058*
(EV) N96076L*	(GL) Carolina Runner	(MC) GP-NC WS 2	(MC) GP-NC WS 13	(MD) HTS IL-002*	(MD) HTS IL-067*

* Denotes Whole-Genome Sequencing in Leaf Spot Trial

EV - Elite Virginia-type Cultivar or Breeding Line

FV - Virginia-type Founder Cultivar or Breeding line

GL - Germplasm Line

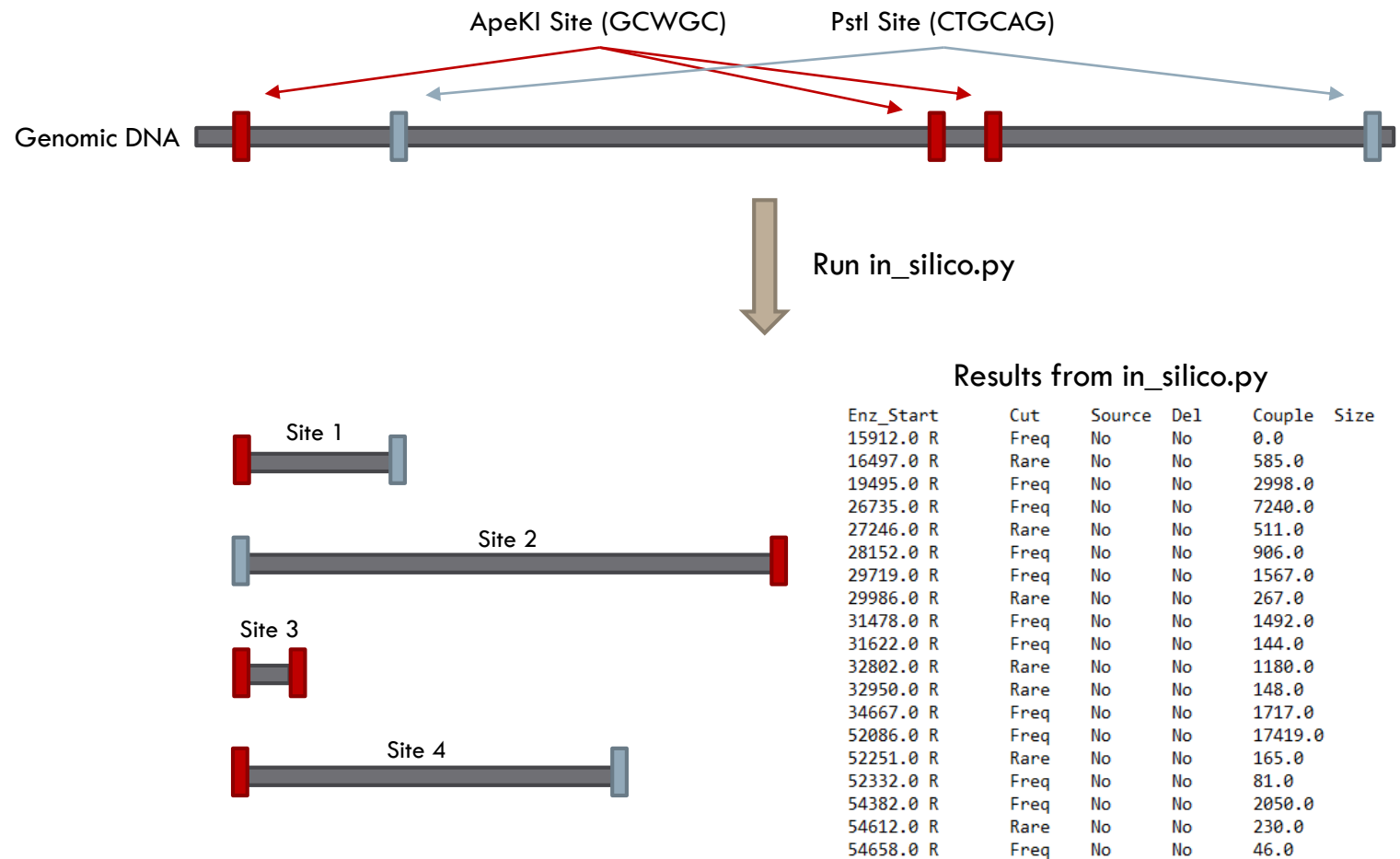
MC - Multiple Disease Resistant Lines with Introgressions from *A. cardenasii*

MD - Multiple Disease Resistant Lines with Introgressions from *A. diogeni*

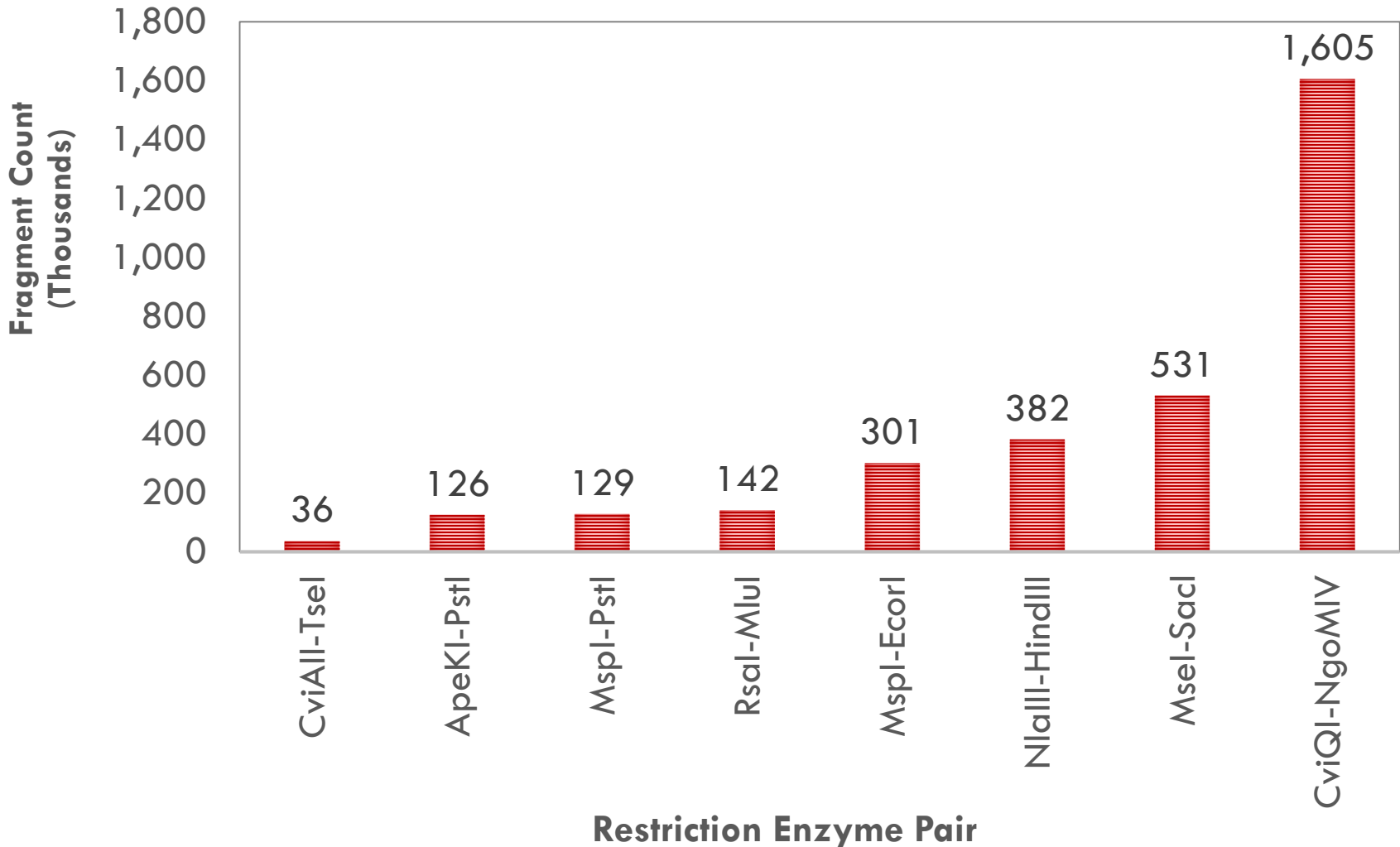
Result: > 4 million SNP markers identified – Maximize SNP recovery with Genotype-By-Sequencing

Genotype-By-Sequencing (GBS)

In silico, construct reduced representation libraries (RRLs) by digesting each DNA sample with a restriction enzyme pair (example: ApeKI – PstI)

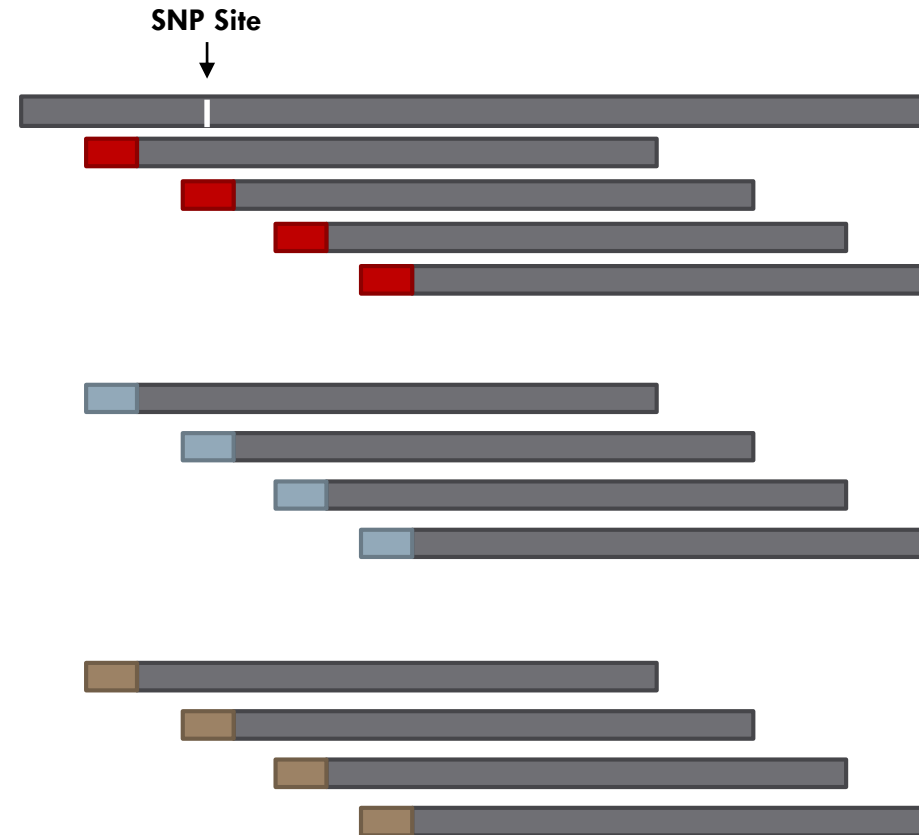


Genotype-By-Sequencing (GBS)



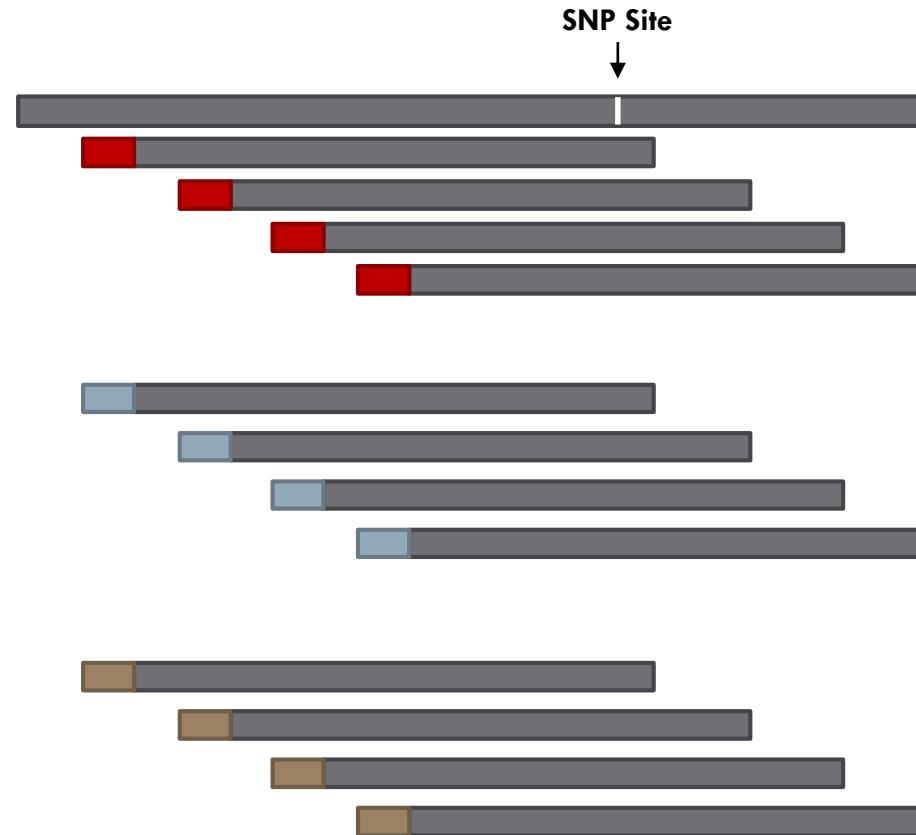
Genotype-By-Sequencing (GBS)

Align whole-genome
sequences to internal
reference 'Bailey II' for
SNP discovery



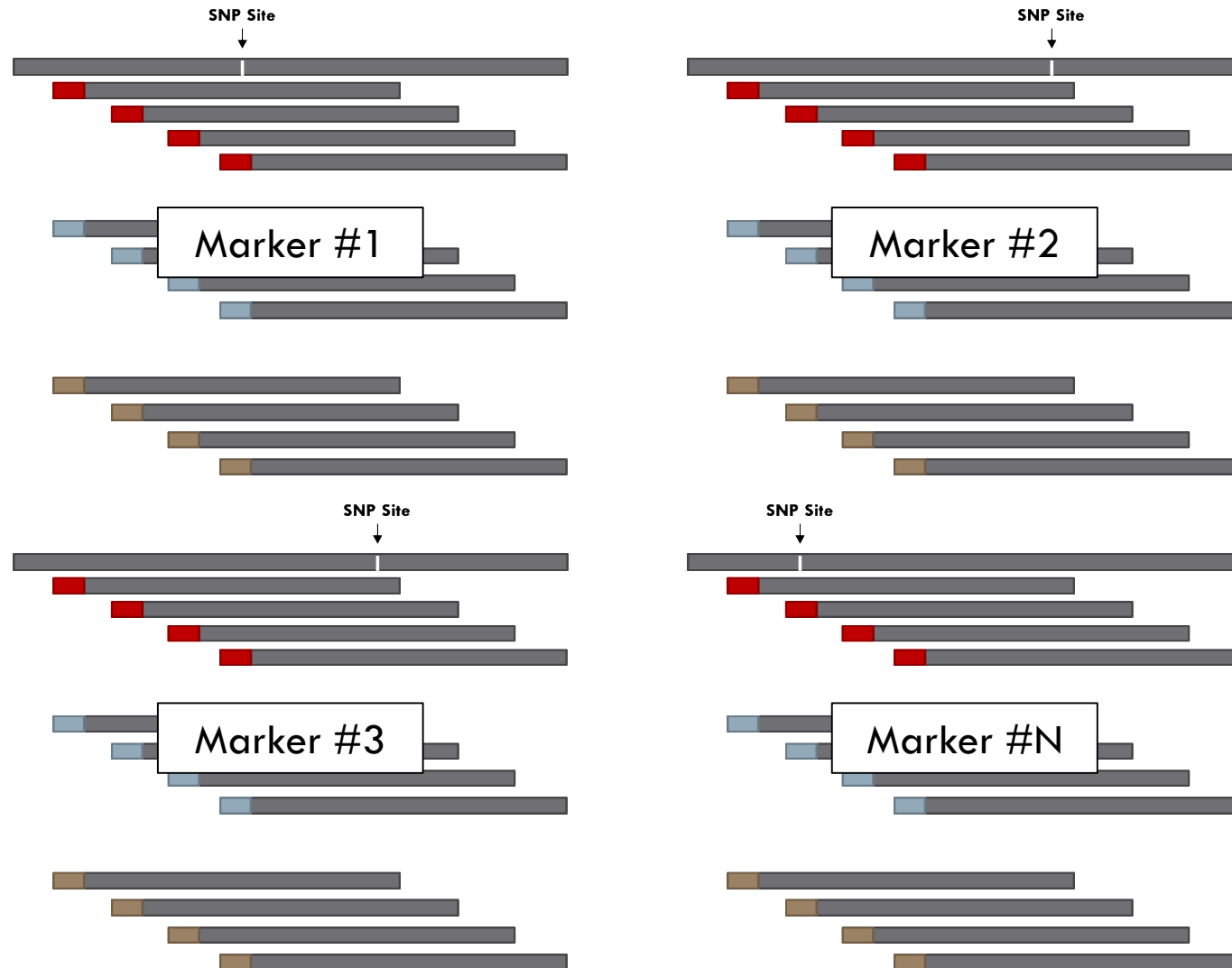
Genotype-By-Sequencing (GBS)

Align whole-genome
sequences to internal
reference 'Bailey II' for
SNP discovery



Genotype-By-Sequencing (GBS)

Align whole-genome
sequences to internal
reference 'Bailey II' for
SNP discovery



Genotype-By-Sequencing (GBS)

Use results from *in silico* digest to count potential recovery sites of SNPs discovered from whole-genome alignment

Results from in_silico.py

Enz_Start	Cut	Source	Del	Couple	Size
15912.0 R	Freq	No	No	0.0	
16497.0 R	Rare	No	No	585.0	
19495.0 R	Freq	No	No	2998.0	
26735.0 R	Freq	No	No	7240.0	
27246.0 R	Rare	No	No	511.0	
28152.0 R	Freq	No	No	906.0	
29719.0 R	Freq	No	No	1567.0	
29986.0 R	Rare	No	No	267.0	
31478.0 R	Freq	No	No	1492.0	
31622.0 R	Freq	No	No	144.0	
32802.0 R	Rare	No	No	1180.0	
32950.0 R	Rare	No	No	148.0	
34667.0 R	Freq	No	No	1717.0	
52086.0 R	Freq	No	No	17419.0	
52251.0 R	Rare	No	No	165.0	
52332.0 R	Freq	No	No	81.0	
54382.0 R	Freq	No	No	2050.0	
54612.0 R	Rare	No	No	230.0	
54658.0 R	Freq	No	No	46.0	

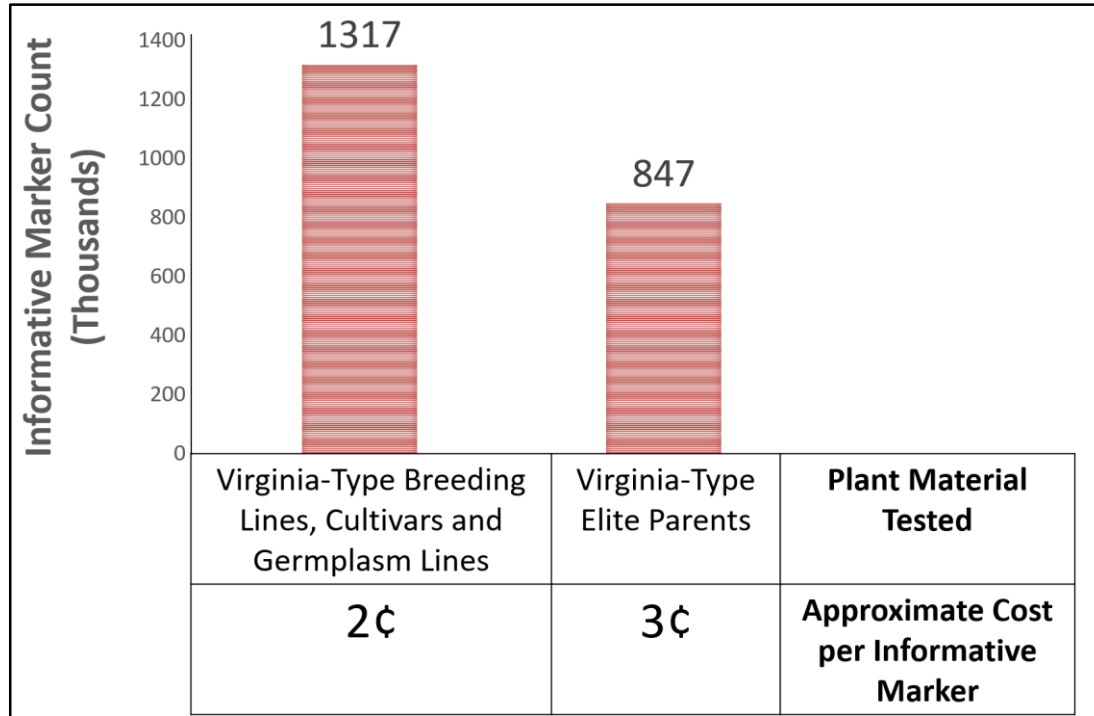


Filtering Results (in_silico.py)

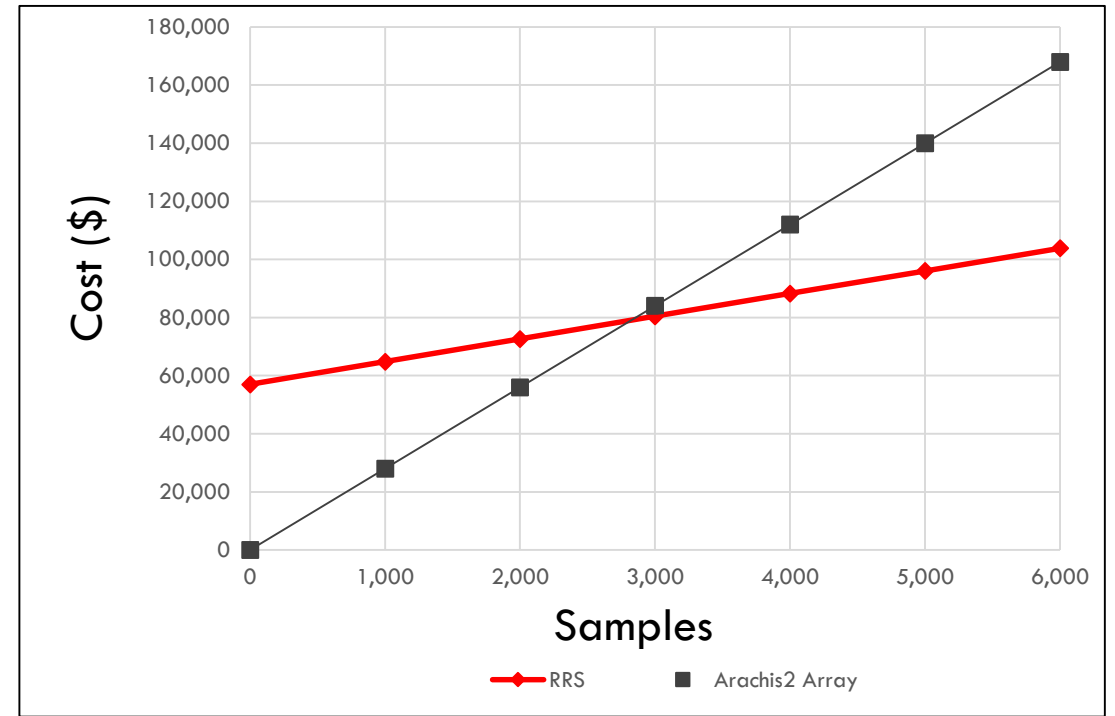
- ❖ Sequence fragment size selection (150-500 bp)
- ❖ **Frequent – Rare** sticky ends
- ❖ Sites identified from whole-genomes sequencing falling between starting positions

Validation: Run GBS protocol with selected enzyme pairs for read depth evaluation and SNP count recovery post alignment

Genotype-By-Sequencing (GBS)



The Genotype-By-Sequencing (GBS) or Reduced Representation Sequencing (RRS) pipeline will reduce the cost per informative marker from \$0.03 to \$0.001



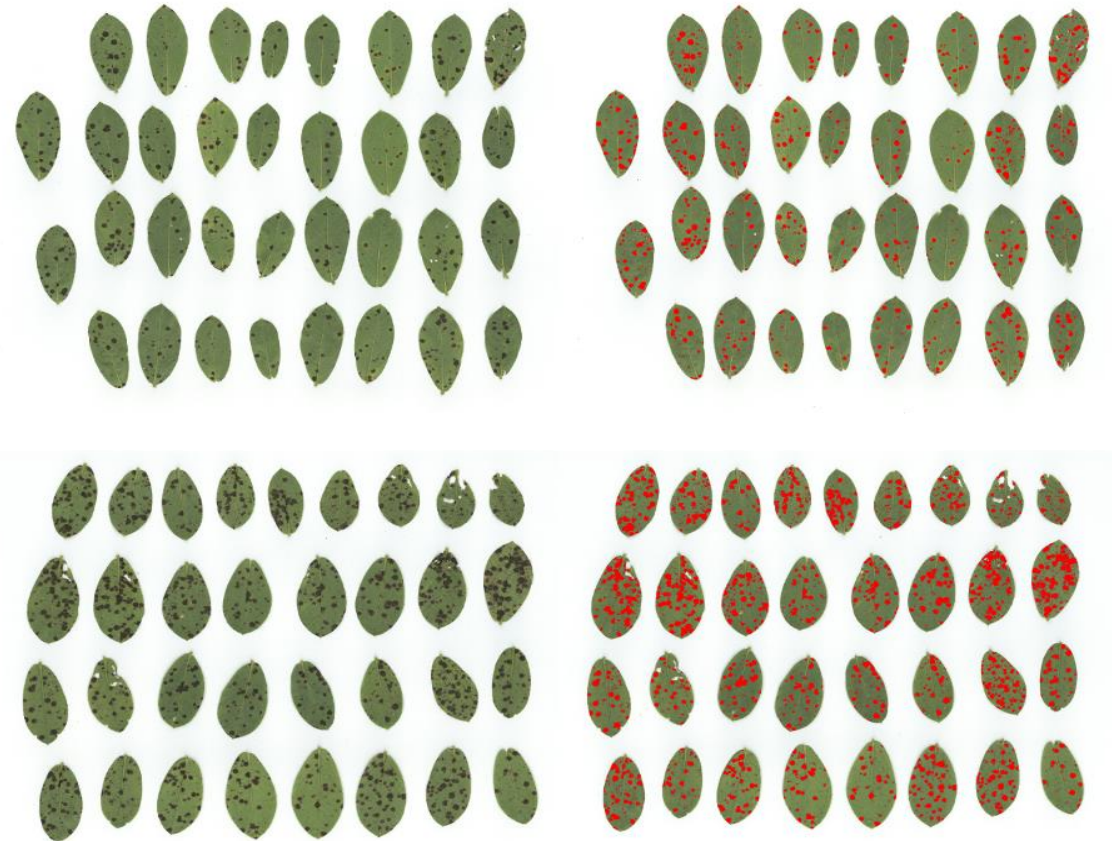
The Genotype-By-Sequencing (GBS) or Reduced Representation Sequencing (RRS) pipeline will break even in 4 years or after ~2,800 samples

Leaf Spot Characterization

Aerial Image of Peanut Defoliation

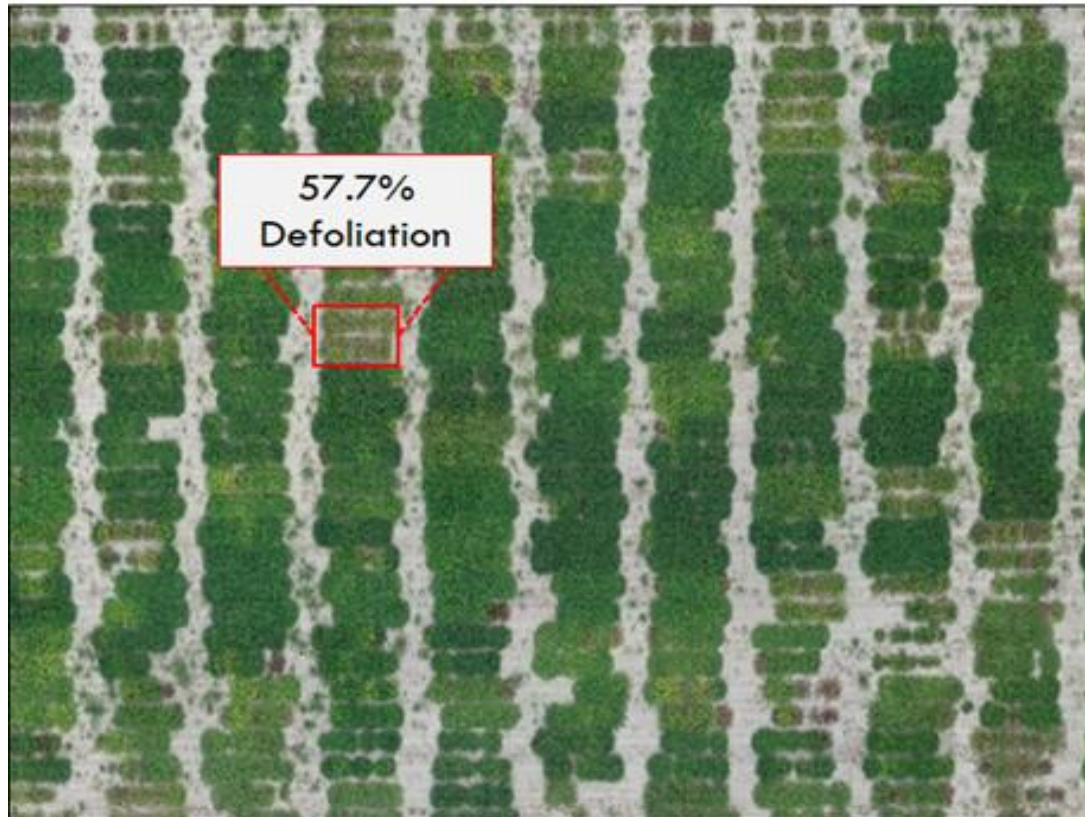


Leaf Spot Tissue Scans for Coverage

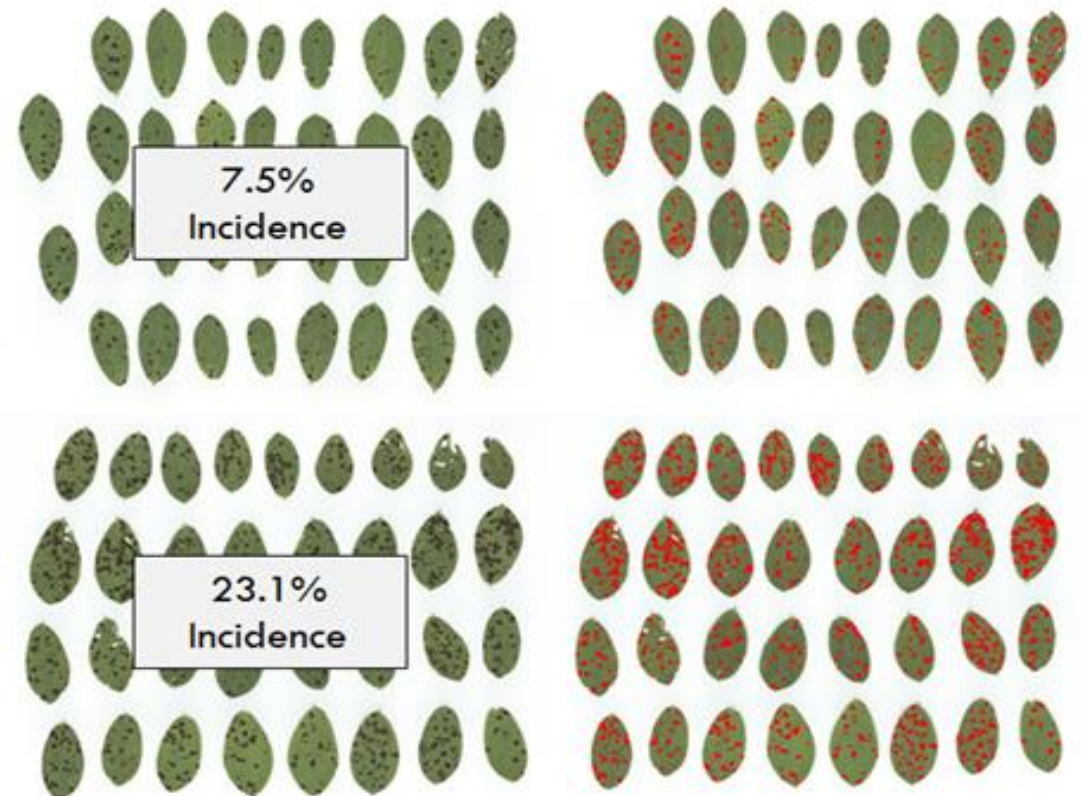


Leaf Spot Characterization

Aerial Image of Peanut Defoliation



Leaf Spot Tissue Scans for Coverage



Comparison of Imaging to Visual Ratings

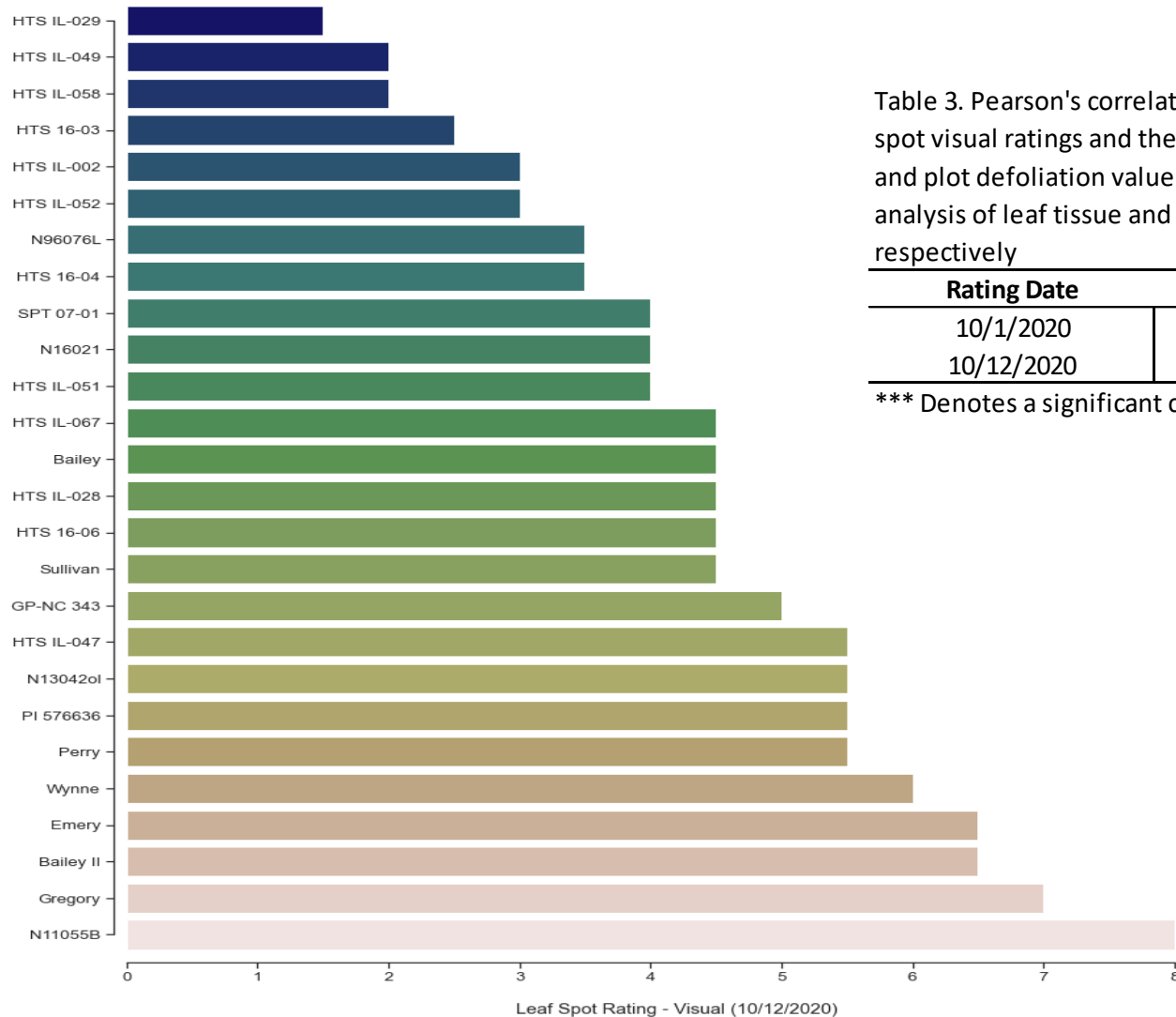


Table 3. Pearson's correlation coefficient between leaf spot visual ratings and the combined percent leaf spot and plot defoliation values determined through image analysis of leaf tissue and aerial drone imaging, respectively

Rating Date	Pearson's Correlation (r)
10/1/2020	0.59***
10/12/2020	0.64***

*** Denotes a significant correlation at the <0.0001 level

2021 Budget

FUNDS REQUESTED:	2021
EPA Salaries and fringe benefits¹	
Graduate Student Stipends and Support²	\$21,000
SPA Salaries and fringe benefits¹	
Part-time Labor (including: 8.60% fringe benefits, and add \$142/year health insurance for hourly / part-time workers who will work ≥ 30hrs/wk. for longer than 3 months)	\$5,000
Supplies (such as: lab and field consumables, fuel, etc.)	\$2,500
Equipment (greater than \$500)	
Travel (for airfare, overnight lodging and meals associated with project tasks)	
Other (such as: motor pool rental, lab analysis, contracted services, greenhouse space, page/publication charges)	
TOTAL in Dollars	\$28,500



SUPPORT FOR PEANUT WILD SPECIES
BREEDING AND GERMPLASM
MAINTENANCE AT NCSU

NORTH CAROLINA PEANUT GROWERS RESEARCH MEETING
NOVEMBER 30, 2020

Objective 1: Maintain Wild Species Greenhouse Collection

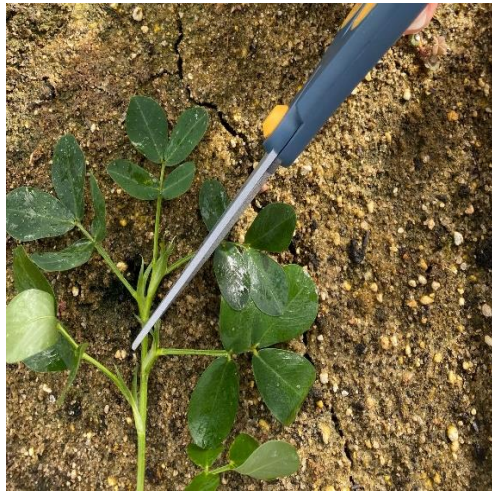
BEFORE



- Started with 325 very weedy and overgrown pots.
- Discarded 171 either because they were duplicates or we already had sufficient seed.
- Did not harvest any seed off these pots.

Objective 1: Maintain Wild Species Greenhouse Collection

- Performed vegetative propagation on the remaining 154 pots.



1) Cut diagonally along the last node of an actively growing lateral branch.



2) Dip cut end in rooting powder.



3) Plant cutting and keep soil wet.



4) In a few weeks, cutting will produce roots and leaves.

Objective 1: Maintain Wild Species Greenhouse Collection



One control box for two benches.

Control box feeds a lateral tube that runs the length of the bench. Every other box has a vertical riser with a sprinkler on top.



Objective 1: Maintain Wild Species Greenhouse Collection

- Harvest seed when its ready.
- Maintain via vegetative propagation if necessary.



AFTER

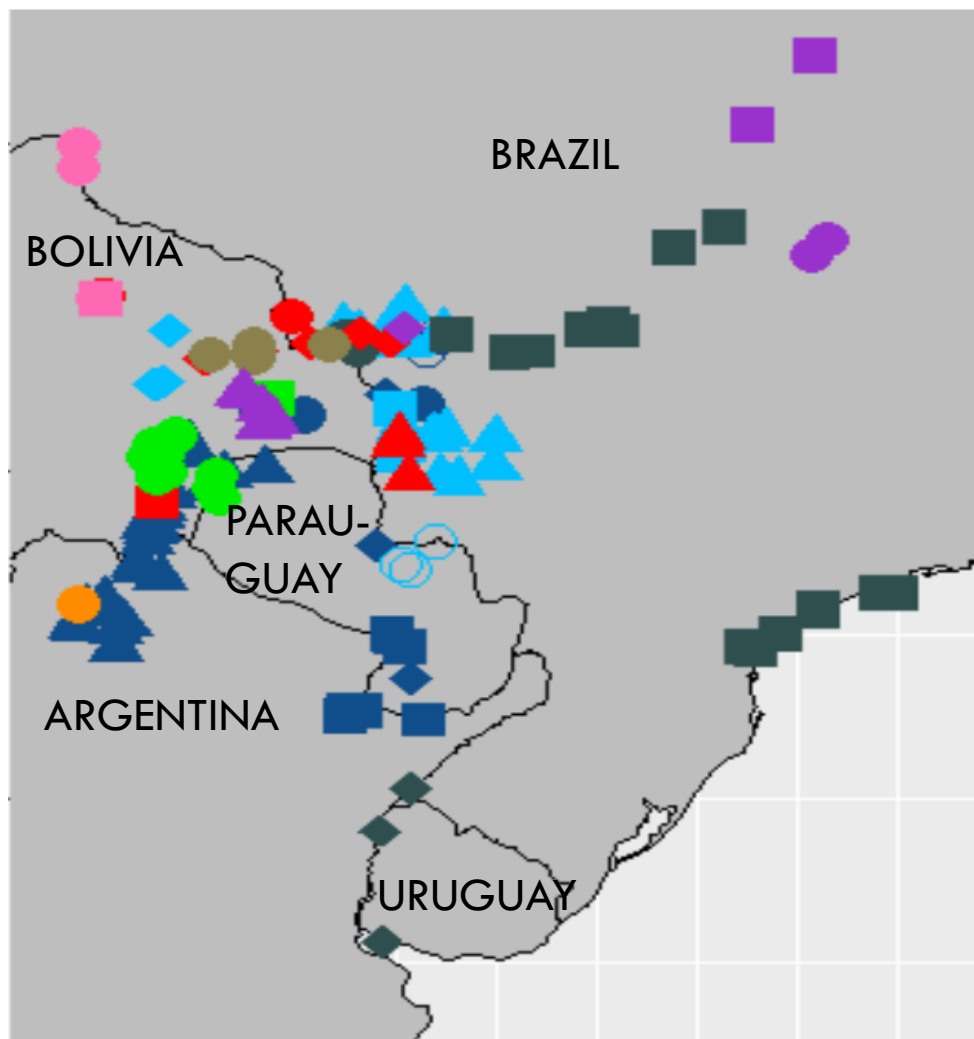


Objective 2: Inventory, Organize, and Refresh the Wild Species Seed Collection

- Refresh: No Wild Species Nursery at Sandhills this year due to Covid-19.
 - No long-term effects.
 - Larger in 2021.
- Inventory: 475 total accessions
 - 22 known duplicates
 - 38 have questionable or incomplete information
 - 415 “high-quality” accessions

Objective 2: Inventory, Organize, and Refresh the Wild Species Seed Collection

Section Arachis – 217 Accessions



A Genome (155)

- *A. cardenasii* (15)
- *A. correntina* (10)
- ◆ *A. diogeni* (6)
- ▲ *A. duranensis* (46)
- *A. helodes* (7)
- *A. herzogii* (1)
- *A. hoehnei* (4)
- ◆ *A. kempff-mercadoi* (6)
- ▲ *A. kuhlmannii* (19)
- *A. microsperma* (4)
- *A. simpsonii* (5)
- *A. stenosperma* (23)
- ◆ *A. villosa* (9)

AB Genome (7)

- *A. monticola* (7)

A. hypogaea

B Genome (14)

- *A. gregoryi* (1)
- *A. ipaensis* (1)
- ◆ *A. magna* (7)
- ▲ *A. valida* (4)
- *A. williamsii* (1)

D Genome (5)

- *A. glandulifera* (5)

F Genome (5)

- *A. benensis* (4)
- *A. trinitensis* (1)

K Genome (17)

- *A. batizocoi* (15)
- *A. cruziana* (2)

Unknown Genome (6)

- *A. decora* (3)
- *A. palustris* (2)
- ◆ *A. praecox* (1)

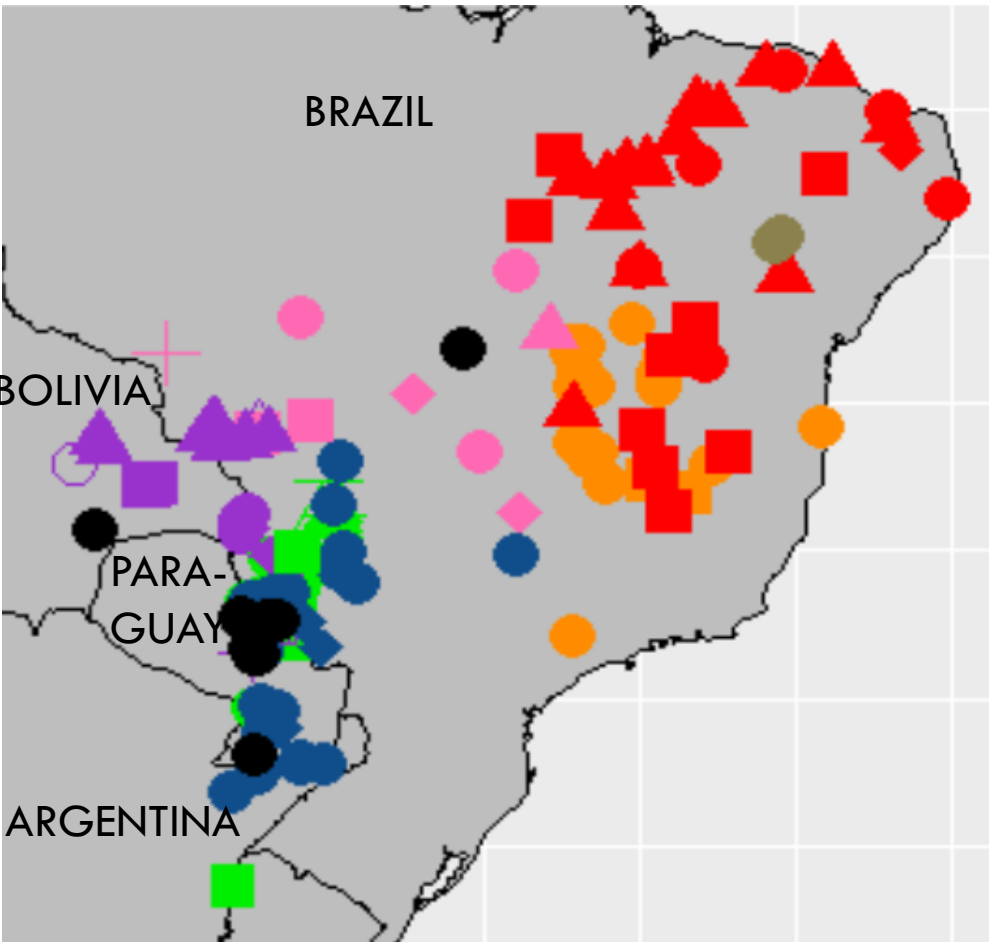
Species Unknown (8)

- ▲ Unknown

Accessions from these 3 species have been successfully used by our program to broaden the genetic diversity of the cultivated breeding program.

Objective 2: Inventory, Organize, and Refresh the Wild Species Seed Collection

Other Sections – 196 Accessions



Section Caulorrhizae (27)

- *A. pintoi* (25)
- *A. repens* (2)

Section Erectoides (42)

- *A. archeri* (1)
- △ *A. benthamii* (2)
- + *A. cryptopotamica* (2)
- ▲ *A. hermannii* (3)
- *A. major* (10)
- *A. paraguariensis* (21)
- *A. porphyrocalyx* (1)
- ◆ *A. stenophylla* (2)

Section Extranervosae (10)

- *A. burchelli* (3)
- *A. lutescens* (2)
- ◆ *A. macedoi* (2)
- ▲ *A. retusa* (1)
- + *A. villosulicarpa* (2)

Section Heteranthae (40)

- *A. dardanoi* (7)
- *A. pusilla* (13)
- ◆ *A. seridoensis* (1)
- ▲ *A. sylvestris* (19)

Section Procumbentes (28)

- *A. appressipila* (6)
- *A. chiquitana* (3)
- ◆ *A. kretschmeri* (4)
- + *A. lignosa* (1)
- ▲ *A. matiensis* (10)
- *A. pflugeae* (2)
- *A. rigonii* (1)
- △ *A. subcoriacea* (1)

Section Rhizomatosae (38)

- ▲ *A. burkartii* (1)
- *A. glabrata* (33)
- ◆ *A. pseudovillosa* (4)

Section Triseminata (2)

- *A. triseminata* (2)

Section Unknown (11)

- Unknown

Used as forage cultivars.

Exploratory research on use of wild peanut as a cover crop because:

- 1) Low-Growing
- 2) Form dense mats
- 3) Nitrogen-fixing

Objective 2: Inventory, Organize, and Refresh the Wild Species Seed Collection

- Organize: Currently have 5 freezers that look like this.
- Discard anything older than 15 years.
 - Includes 2 additional freezers at Reedy Creek
- Keep only 3 most recent bottles of each accession.
- Eventually genotype to eliminate duplicates and misclassifications in Section Arachis.

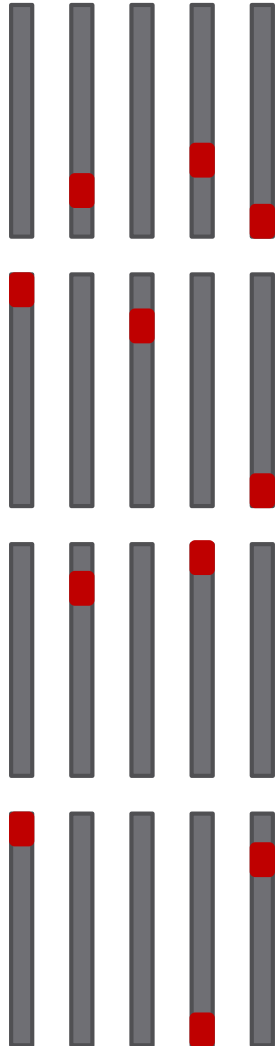


Objective 3: Develop a genotyping strategy to characterize the wild species collection.

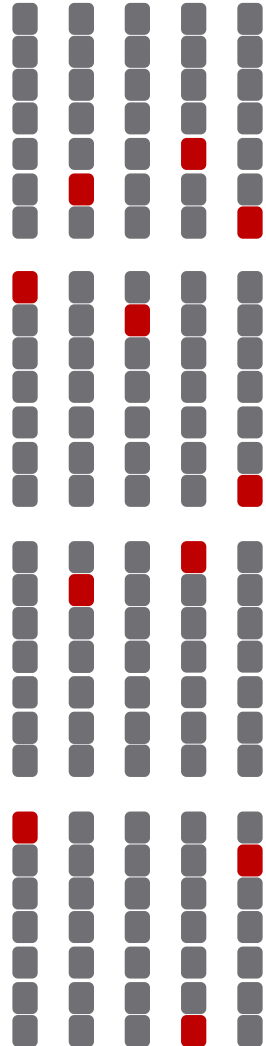
- With 218 wild species in Section Arachis, which ones should we cross with?
- Traditional Route: 1st step would involve full sequencing of 10 wild species.
 - \$50,000 & ~2 years to analyze the data
- Alternate Route: Wild species primary utility is their collection of novel R (Resistance) genes.
 - Not found in cultivated material.
- R genes are very similar to one another and that makes them easy to find in the peanut genome.

Objective 3: Develop a genotyping strategy to characterize the wild species collection.

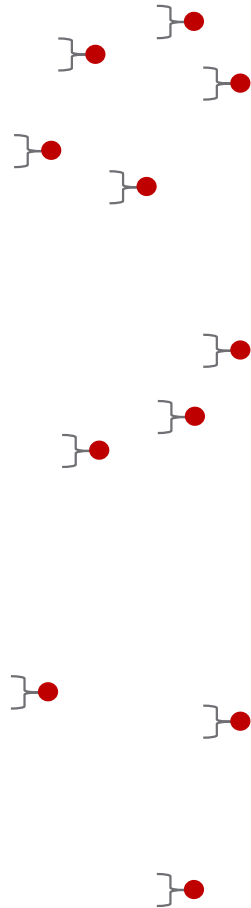
Genome with R Genes



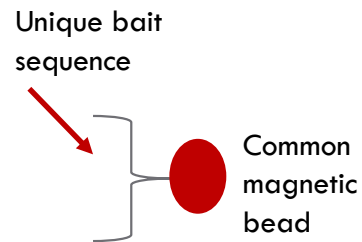
Fragment Genome



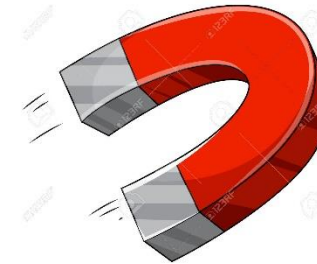
Mix with Bait Library



Baits Bind to Targets



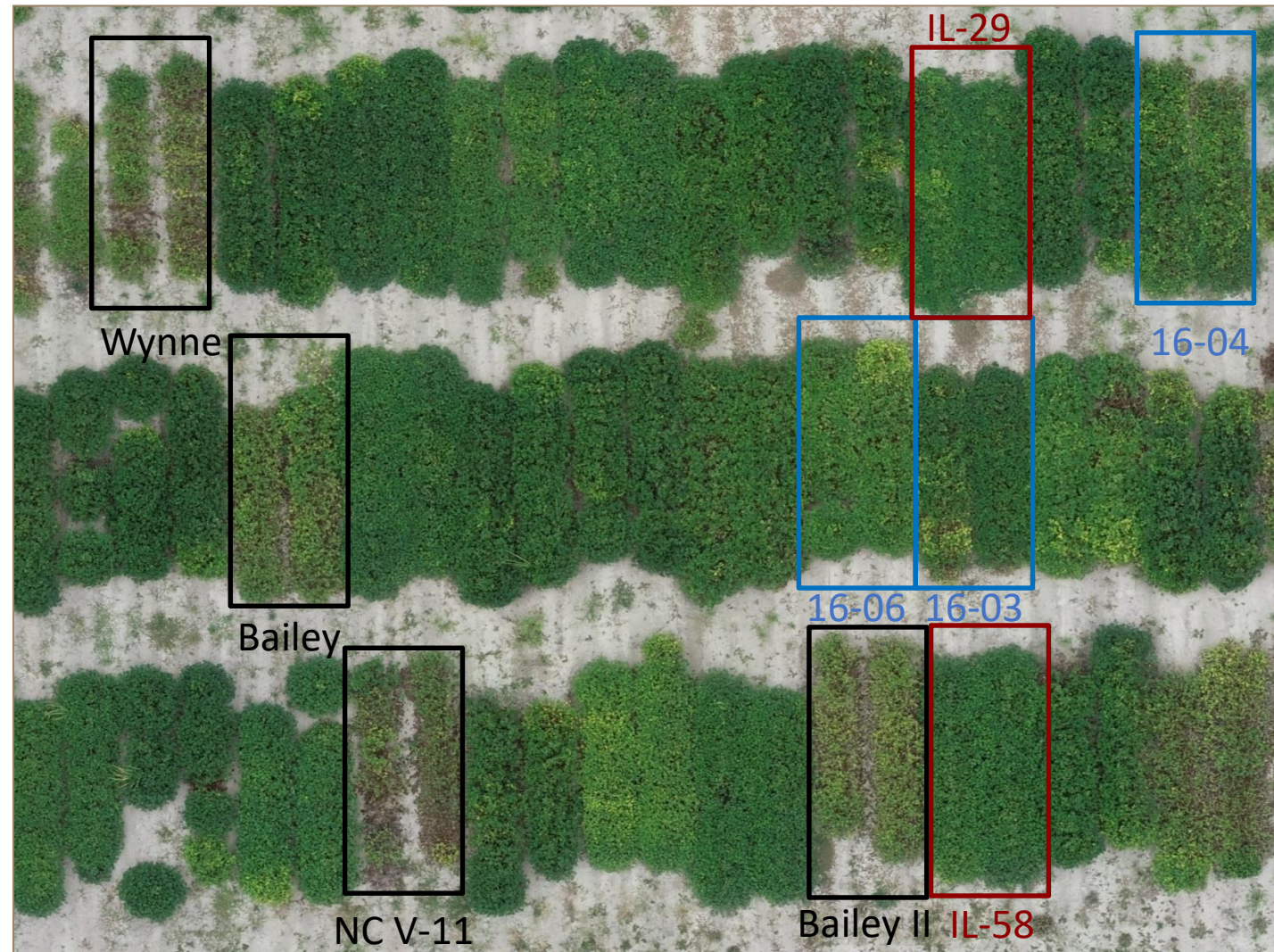
Use Magnet to Capture Baits + Targets



- Sequence just the R genes.
- Get 95% of the relevant data.
- \$8,800
- Data analysis ~2 months.

Objective 4: Continue the wild species crossing program.

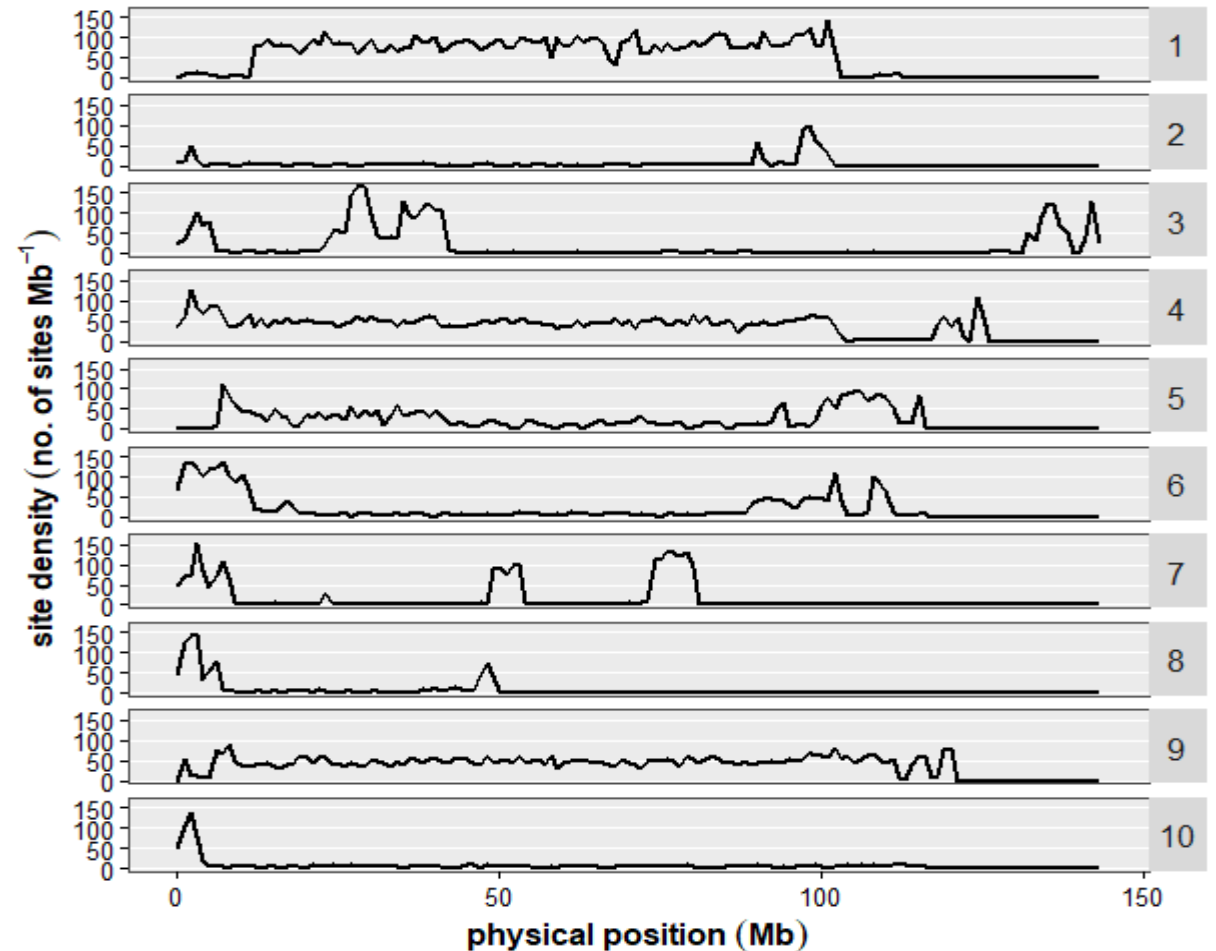
- We have breeding lines with elite leaf spot resistance from both *A. cardenasii* and *A. diogeni* backgrounds.
- The objective is to combine them in a single breeding line.
- This makes it much easier to transfer both sources of resistance to future cultivars.



Objective 4: Continue the wild species crossing program.

- 1st step: Sequenced 18 relevant lines.
- 2nd step: Aligned those sequences to the reference genome.
- Peaks in the graphs on the right represent regions of wild species DNA.
- Design molecular markers to determine which of these blocks lead to resistance.

A. Diogeni Introgression Blocks in IL-29



Objective 4: Continue the wild species crossing program.

- In summer 2000, VA-98R was crossed with the *A. correntina* accession GKP 9530.
- After 20 years we may finally have 3 plants out of this cross with the right number of chromosomes.
- **Send seed from these plants to USDA to check the chromosome number.**



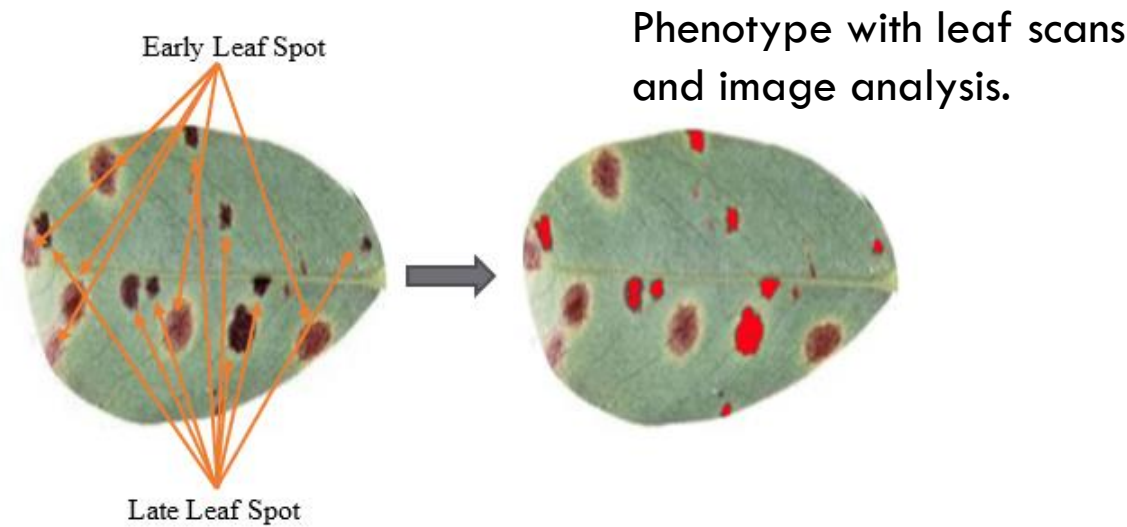
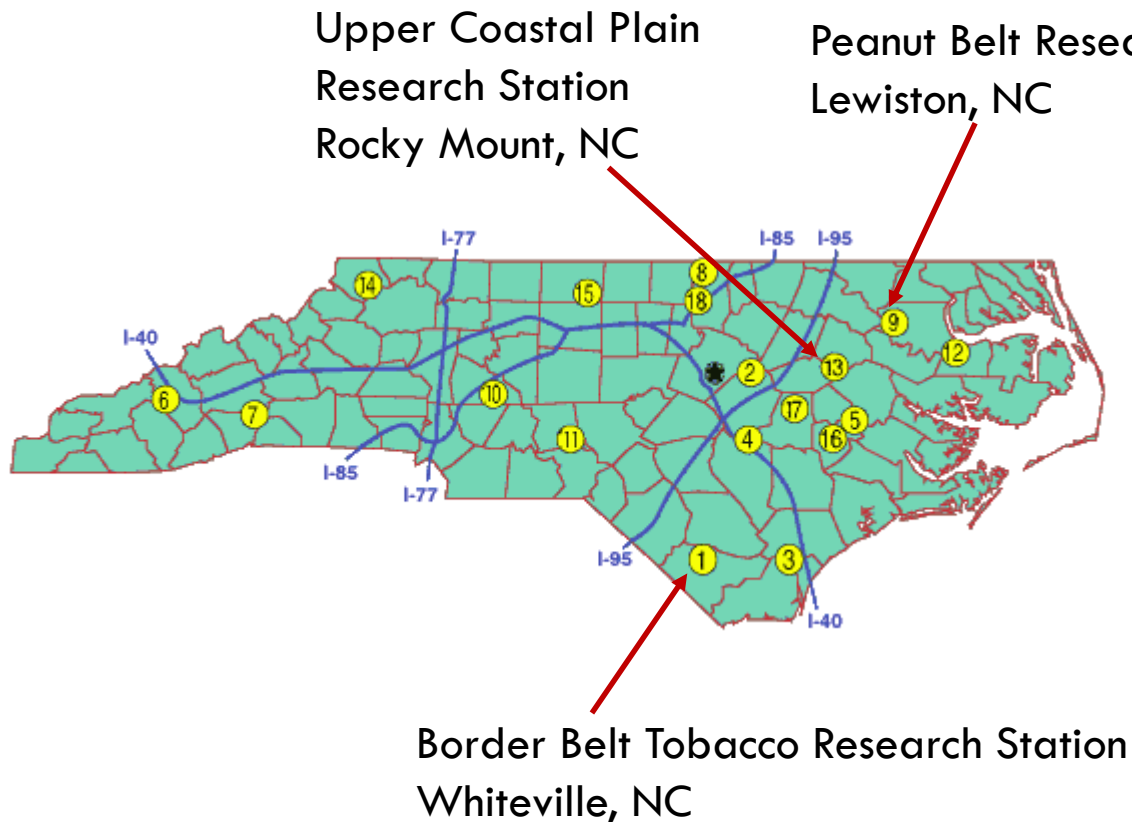
Objective 4: Continue the wild species crossing program.

- 46 synthetic allotetraploids in the greenhouse from 11 different wild species crosses.
- **Keep looking for normal levels of fertility in these plants.**

♀ Parent	Genome	♂ Parent	Genome	# Plants
<i>A. batizocoi</i>	KK	<i>A. cardenasii</i>	AA	2
<i>A. batizocoi</i>	KK	<i>A. correntina</i>	AA	2
<i>A. batizocoi</i>	KK	<i>A. stenosperma</i>	AA	6
<i>A. gregoryi</i>	BB	<i>A. stenosperma</i>	AA	3
<i>A. ipaensis</i>	BB	<i>A. cardenasii</i>	AA	3
<i>A. ipaensis</i>	BB	<i>A. correntina</i>	AA	15
<i>A. ipaensis</i>	BB	<i>A. diogoi</i>	AA	5
<i>A. ipaensis</i>	BB	<i>A. duranensis</i>	AA	1
<i>A. ipaensis</i>	BB	<i>A. stenosperma</i>	AA	4
<i>A. magna</i>	BB	<i>A. cardenasii</i>	AA	4
<i>A. magna</i>	BB	<i>A. stenosperma</i>	AA	1



Objective 5: Evaluate wild species lines for disease resistance and yield potential.



Budget

BUDGET

FUNDS REQUESTED:

EPA Salaries and fringe benefits¹

Graduate Student Stipends and Support²

SPA Salaries and fringe benefits¹

Part-time Labor (including: 9.05% fringe benefits, and add \$151/year health insurance for hourly / part-time workers who will work \geq 30hrs/wk. for longer than 3 months)

Supplies (such as: lab and field consumables, fuel, etc.)

Equipment **(greater than \$500)**

Travel (for airfare, overnight lodging and meals associated with project tasks)

Other (such as: motor pool rental, lab analysis, contracted services, greenhouse space, page/publication charges)

TOTAL in Dollars

	2021	2022	2023
EPA Salaries and fringe benefits ¹	\$21,218		
Graduate Student Stipends and Support ²			
SPA Salaries and fringe benefits ¹			
Part-time Labor (including: 9.05% fringe benefits, and add \$151/year health insurance for hourly / part-time workers who will work \geq 30hrs/wk. for longer than 3 months)			
Supplies (such as: lab and field consumables, fuel, etc.)	\$7,500		
Equipment (greater than \$500)			
Travel (for airfare, overnight lodging and meals associated with project tasks)			
Other (such as: motor pool rental, lab analysis, contracted services, greenhouse space, page/publication charges)			
TOTAL in Dollars	\$28,718	\$	\$