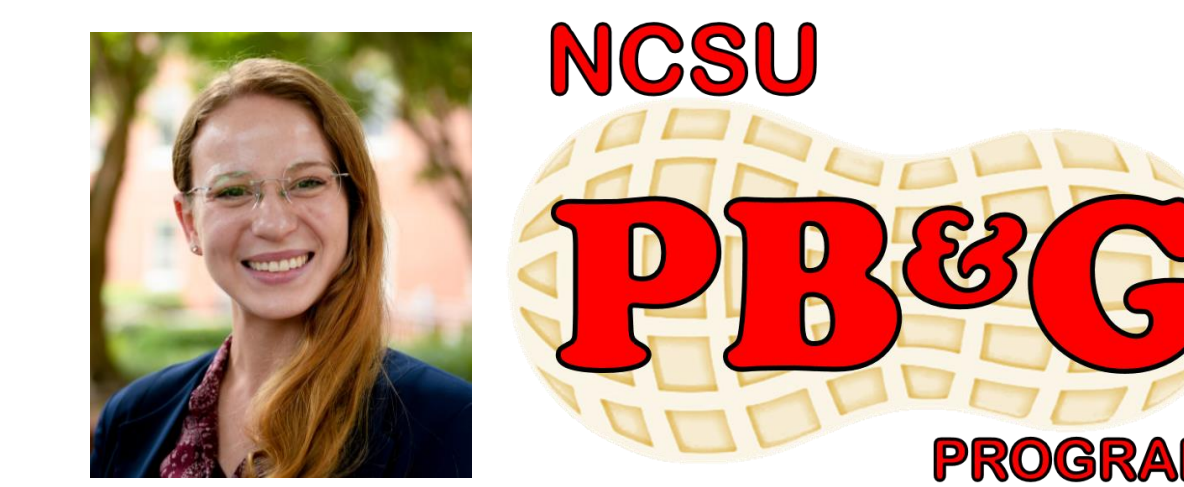


# The Generation and Analysis of Genomics Information Forms the Foundation for Efficient Virginia-Type Peanut Cultivar Development

C.S. NEWMAN\*, A.T. OAKLEY, R.J. ANDRES, K.R. FRITZ, J.C. DUNNE, Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC, 27695; B.E. SCHEFFLER, S.A. SIMPSON, USDA-ARS GBRU, Stoneville, MS, 38776; J.C. CAMPBELL, S.B. CANNON, USDA-ARS, Corn Insects and Crop Genetics Research Unit, Ames, IA, 50011-4014; R.C. YOUNGBLOOD, Institute for Genomics, Biocomputing, and Biotechnology, Mississippi State University, Mississippi State, MS, 39762; A.M. HULSE-KEMP, USDA-ARS Genomics and Bioinformatics Research Unit, Raleigh, NC, 27695

NC STATE UNIVERSITY



csnewma2@ncsu.edu

## Introduction

- Peanut (*Arachis hypogaea* L.,  $2n=4x=40$ ) is a staple crop that is planted and consumed globally as a source of protein and oil
- In the United States, 2,782,290 MT of peanuts were produced in 2020<sup>1</sup>
- Runner-Types, Virginia-Types, Spanish-Types, and Valencia-Types, are grown on 85%, 10%, 2%, and 1% of United States peanut acreage respectively<sup>2</sup>



Figure 1. Peanut Market-Types: Runner, Valencia, Spanish, Virginia (Listed from left to right<sup>3</sup>)

- NCSU has been involved in Virginia-Type peanut breeding and improvement since the 1930s<sup>4</sup>
- Breeding activities within the NCSU Peanut Breeding and Genetics (PB&G) program were informed by visual ratings and observations.

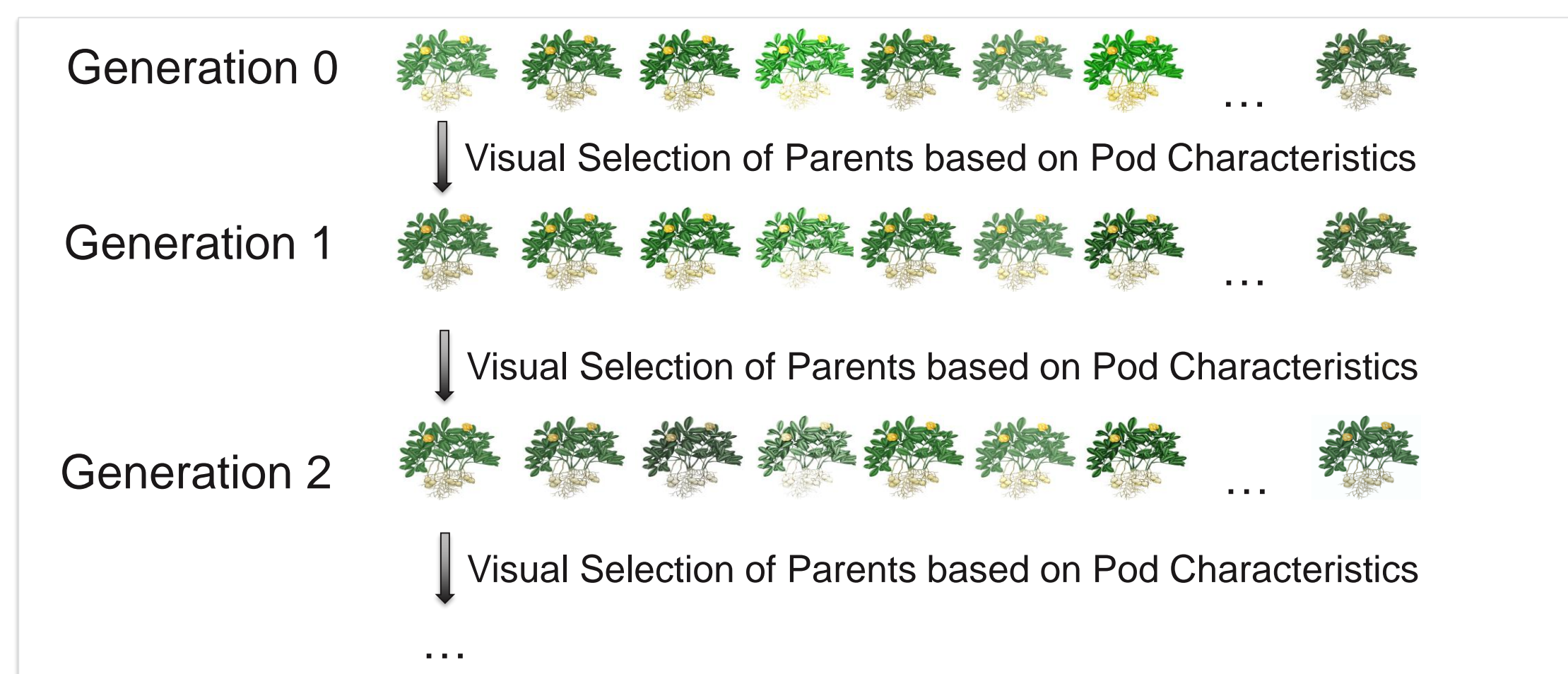


Figure 2. Example of recurrent visual selection for breeding peanut<sup>5</sup> cultivars used in the NCSU PB&G program

- Novel mating schemes, which added genes from wild species into the peanut breeding population is a vital part of the PB&G program history. The genes from wild species allowed the cultivars produced from NCSU to be resistant to economically damaging pathogens.

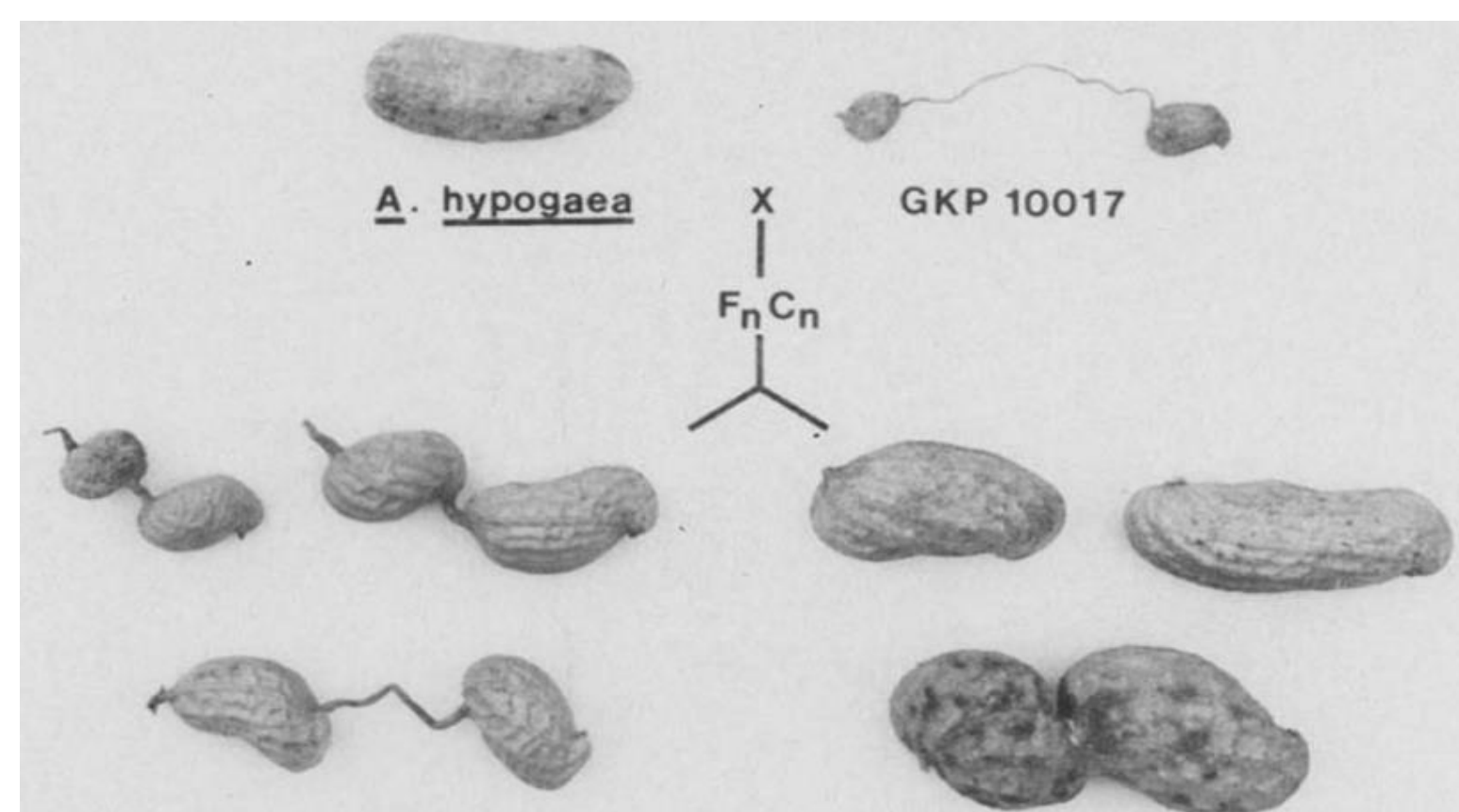


Figure 3. Photograph<sup>6</sup> showing the result of a wild species (*A. cardenasii*) mated with cultivated peanut (*A. hypogaea*) at NCSU in the 60s

- Contemporary genomic data are cheap, high volume and can be generated quickly
- Using genomic data in breeding is known as genomics-assisted breeding. Integrating genomic technology leads to more efficient and accurate cultivar development.

## Objectives

- Create Genomic Resources for the NCSU PB&G Program
- Use Genomic Data to Characterize Peanut Lines within the Breeding Population
- Characterize Historical Wild Species Introgression Events
- Generate the Information Needed for Genomics-Assisted Breeding in the NCSU PB&G Program

## Materials and Methods

- Cultivar Bailey II was chosen to represent Virginia-Type peanut as the reference genomic sequence. This cultivar has excellent product shelf life, outstanding agronomic performance, and good disease resistance to fungal pathogens.
- Genome assembly was done with a series of bioinformatics software packages which took raw genomic sequence data and assembled the data into large pieces. Those pieces were further corrected and connected into chromosome-scale linear sequences. The genome was annotated using a modified BIND<sup>7</sup> pipeline.
- Variant discovery, i.e. discovering genetic differences, was done for 66 peanut lines, of historical and agronomic significance to the breeding program. Variant discovery was largely done following GATK<sup>8</sup> best practices.



Figure 4. The popular Bailey II peanut cultivar, chosen for whole genome sequencing and assembly

## Results

### Bailey II Genome Sequence:

Table 1. Comparison of the Bailey II genome assembly with other cultivated peanut assemblies as a way to contextualize the quality of the genome

Genome Assembly	Number of Contigs	Sequence Length (bp)	Complete BUSCOs
Bailey II	591	2,532,481,151	5181
Tifrunner (v2) <sup>9</sup>	3496	2,534,699,547	5180
Shitoqi <sup>10</sup>	6289	2,506,086,556	5174
Fuhuasheng <sup>11</sup>	31747	2,483,343,588	5078

A 83% decrease in the number of contigs when compared with Tifrunner v2

### Genomic Variant Discovery:

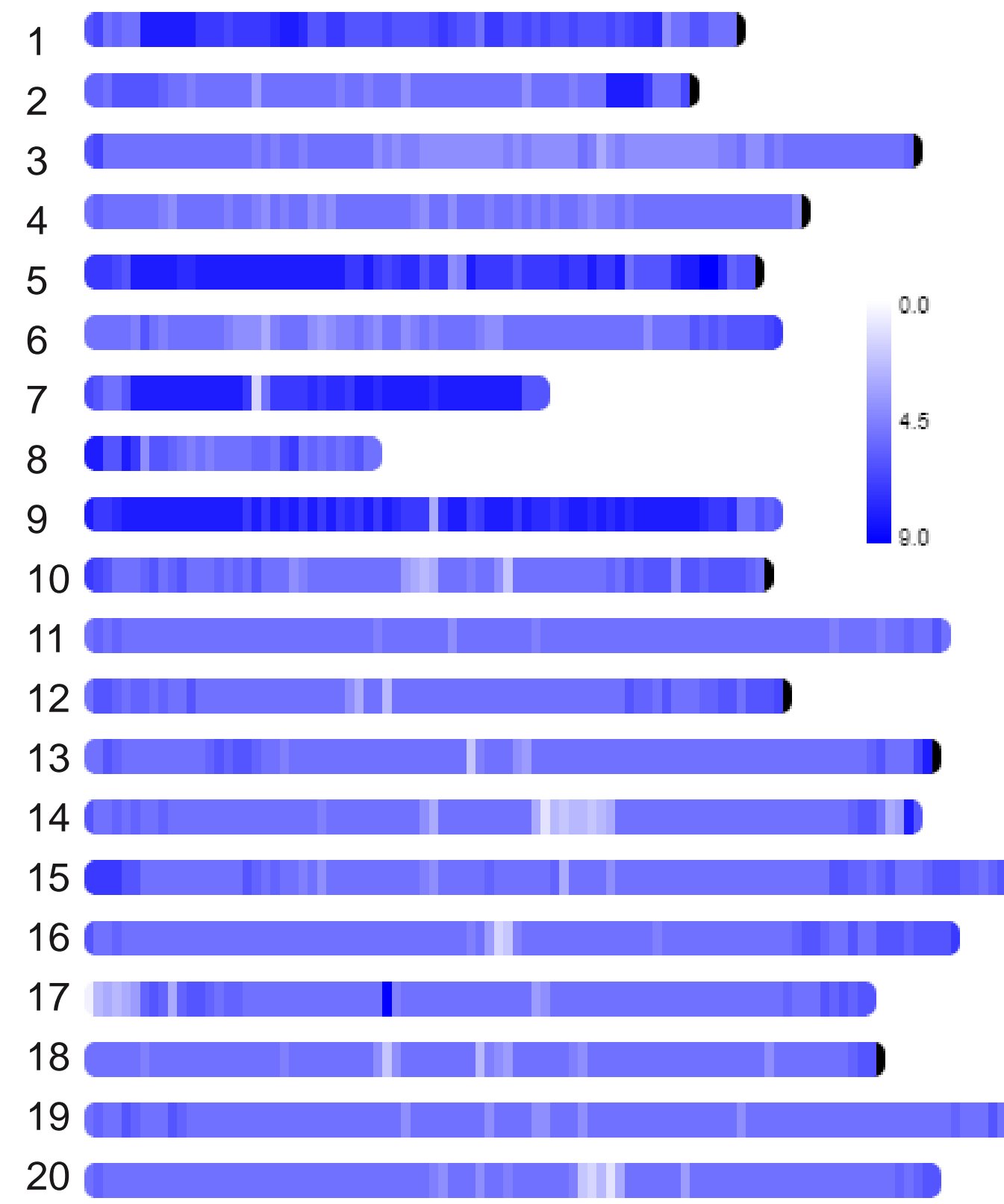


Figure 5. Twenty *hypogaea* chromosomes colored by marker density within windows of 500,000 base pairs. Marker density has been scaled by natural log

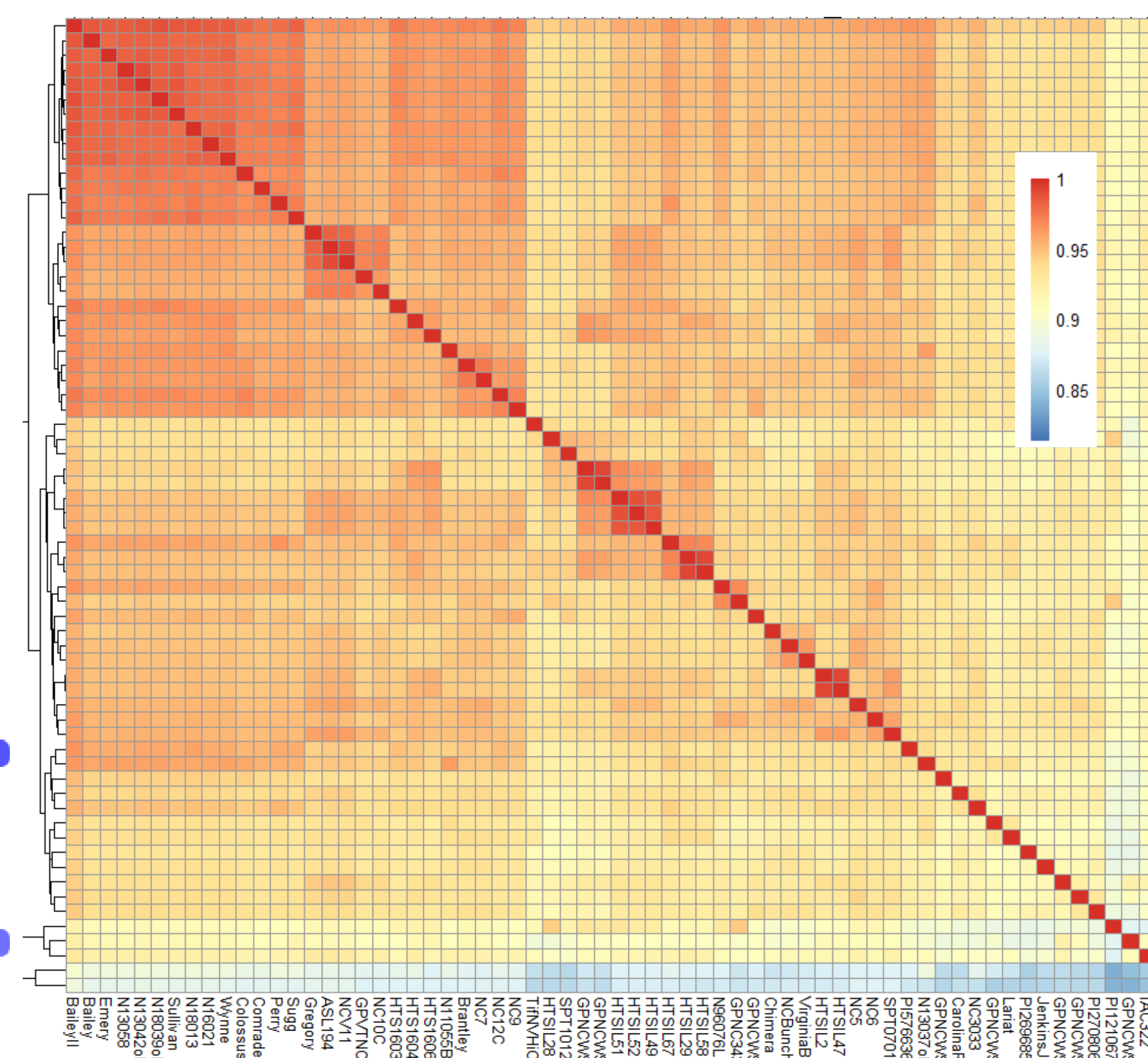


Figure 6. Genomic state identity matrix of individuals important to the NCSU PB&G program. Individuals are organized by Euclidean distance

### Evidence of Wild Species Genome in Cultivated Peanut Lines

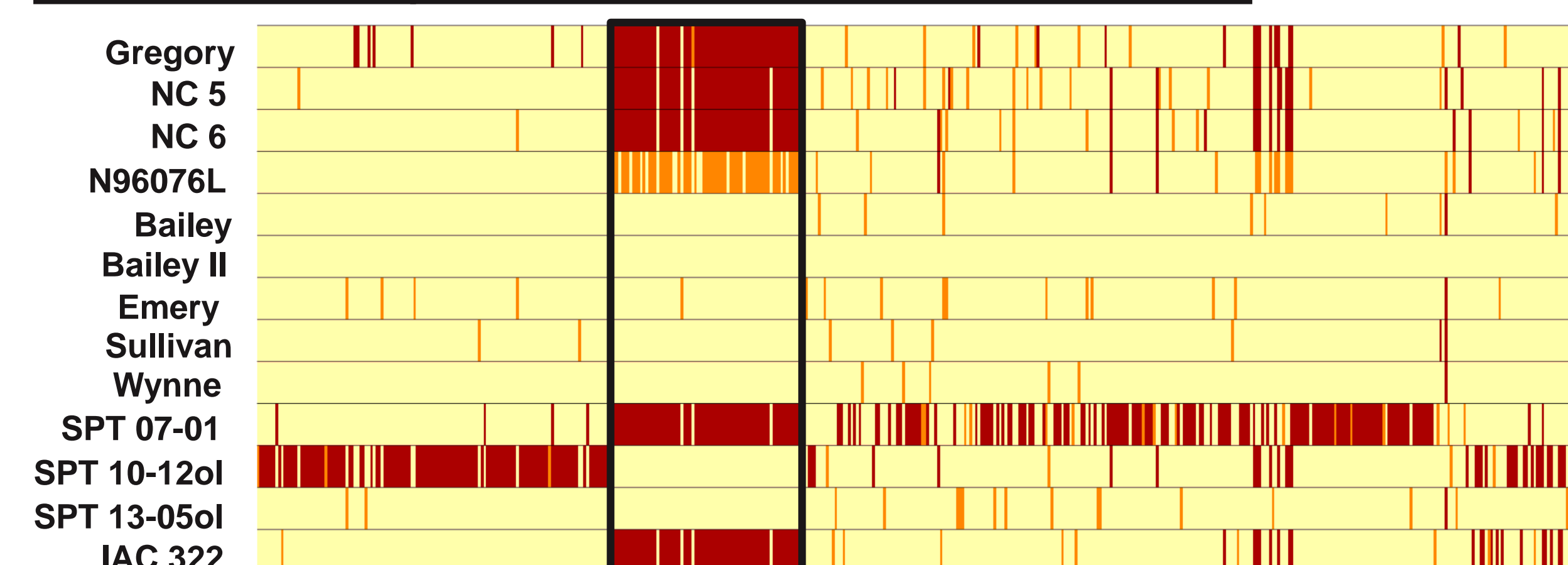


Figure 7. The genomic identity of several peanut lines (represented as rows) across chromosome eight. The dark red and orange within the black box represents foreign genetic material from the wild species mating event, shown in figure 3. Light yellow represents the baseline peanut genome. These blocks of foreign DNA allow for enhanced disease resistance in contemporary cultivars

## Discussion and Conclusion

- Create Genomic Resources for the NCSU PB&G Program
  - Genome reference sequence was generated and represents a large enhancement in terms of contiguity
  - Greater than one million high quality variant sites were identified throughout the genome for the breeding population
- Use Genomic Data to Characterize Peanut Lines within the Breeding Population
  - Markers discovered serve as 'fingerprints' for different individuals
  - The genomic relationship between individuals was quantified
- Characterize Historical Wild Species Introgression Events
  - Genomic blocks were identified and delimited in better resolution when what was previously possible
- Generate the Information Needed to Launch Genomics-Assisted Breeding in the NCSU PB&G program

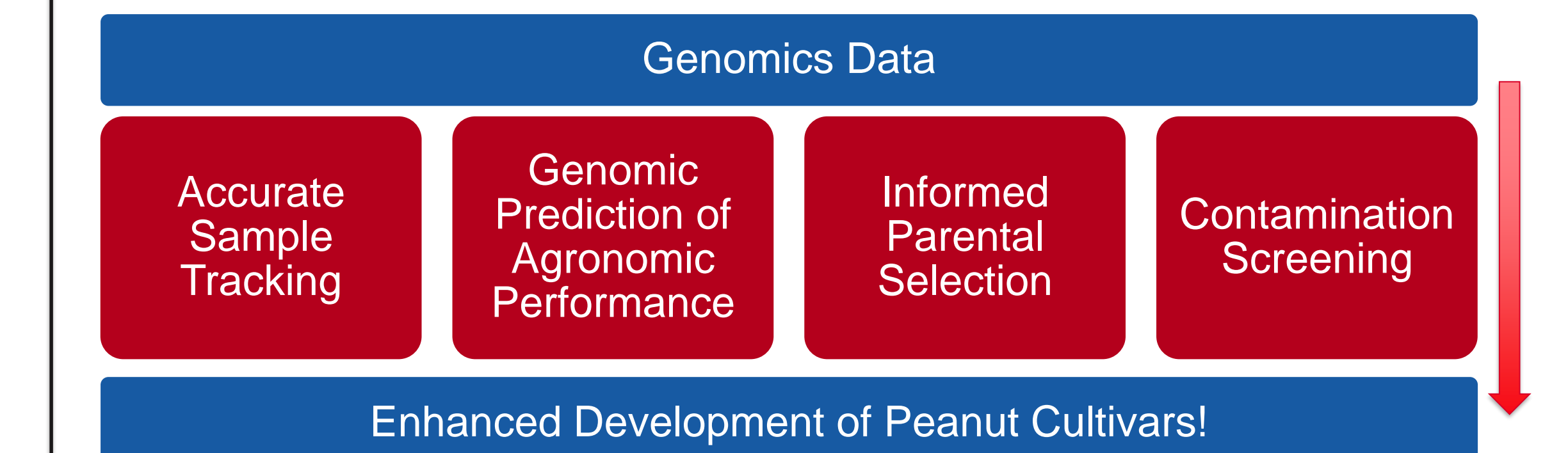


Figure 7. Flowchart demonstrating how the genomic information in this study will lead to timely, accurate peanut cultivar development in the NCSU PB&G program

## References

- Crop Production 2020 Summary 01/12/2021. Crop Prod. 125 (2020).
- Anthony. Peanut Types and Production. American Peanut Council. <https://www.peanutsusa.com/about-peanuts/the-peanut-industry/3/9-peanut-types-and-production.html>
- Market Type Photograph. <https://www.aboutpeanuts.com/peanut-facts/95-types-of-peanuts>. 2019 Virginia-Carolinas Peanut Promotions
- North Carolina State University, College of Agriculture and Life Sciences, Peanut Breeding Program Records, 1938-1992.
- Keith Ward. Image. <https://www.motherearthnews.com/organic-gardening/growing-peanuts-zw0z1312zsto/>.
- Stalker, H. T., Wynne, J. C. & Company, M. VARIATION IN PROGENIES OF AN ARACHIS HYPOGAEA x DIPLOID WILD SPECIES HYBRID. Euphytica 28, 675-684 (1979).
- Jing Li, Urinder Singh, Priyanka Bhandary, Jacqueline Campbell, Zebulun Arendsee, Arun S Seetharam, Eve Syrkin Wurtele, Foster thy young: enhanced prediction of orphan genes in assembled genomes, Nucleic Acids Research, 2021, gkab1238, <https://doi.org/10.1093/nar/gkab1238>
- Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, Shakir K, Thibault J, Chandran S, Whelan C, Lek M, Gabriel S, Daly MJ, Neale B, MacArthur DG, Banks E. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples bioRxiv. 201178. DOI: 10.1101/201178
- Bertioli, D.J., Jenkins, J., Cleveenger, J. et al. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. Nat Genet 51, 877-884 (2019). <https://doi.org/10.1038/s41588-019-0405-z>
- Zhuang, W., Chen, H., Yang, M. et al. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nat Genet 51, 865-876 (2019). <https://doi.org/10.1038/s41588-019-0402-2>
- Xiaoping, Chen Qing Lu, et al. Sequencing of Cultivated Peanut, *Arachis hypogaea*, Yields Insights into Genome Evolution and Oil Improvement, Molecular Plant, Volume 12, Issue 7, 2019, Pages 920-934, ISSN 1674-2052, <https://doi.org/10.1016/j.molp.2019.03.005>.

## Funding

