



Peanut Genome Initiative

2012-2017 Research Accomplishment Report to the U.S. Peanut Industry

December 5, 2017

**Peanut Genome Initiative
Research Final Report to the U.S. Peanut Industry
October, 2017**

Table of Contents

Executive Summary	3
Highlights of Research Accomplishments	6
Introduction	7
Research Technical Accomplishment Reports	9
Component 1: Sequencing and Assembly of Peanut Genomes	9
Component 2: Developing Maps and Markers for finding QTL	8
Component 3: Developing Markers for Specific Genes in QTL	15
Component 4: Evaluating New Sequencing and Assembly Technologies	17
Component 5: Identifying Breeding Lines with QTL for Key Traits	18
Component 6: Creating On-line Tools for Genome Assisted Breeding	23
Where Do We go From Here?	29
Appendices:	34
Appendix 1: Examples of Funding Leading to Additional Grant and Project Support	34
Appendix 2: Meetings Sponsored by PGI	35
Appendix 3: Examples of Key Traits That Reside in a Given Wild Arachis Species	36
Appendix 4: Research Publications by PGI members & Associates Relevant to PGP	37
Appendix 5: Sponsors Who Have Provided Financial Support for the Peanut Genome Initiative	50
Appendix 6: Members of the Peanut Genome Consortium	52
Appendix 7: Terms and Definitions	53

EXECUTIVE SUMMARY

Unlocking the Door to a Limitless Future: The Peanut Genome Initiative (PGI)

In 2012, the U.S. peanut industry charged The Peanut Foundation with initiating a research program to map the genetic code of the peanut plant. The Peanut Genome Initiative (PGI) was — and remains — the largest research project ever funded by our industry, with the \$6M cost shared equally among growers, shellers and manufacturers. This is the final report of that five-year program.

Everyone who played a role in this project should be tremendously proud of his or her contribution. The PGI has given us a map with which we can unlock some of the genetic potential of the peanut plant. We now have the capability to find beneficial genes in cultivated and wild peanuts that can lead to even greater yields, lower production costs, lower losses to disease, improved processing traits, improved nutrition, improved safety, better flavor and virtually anything that is genetically controlled by the peanut plant.

For decades to come, the PGI will lead to improved sustainability and profitability of every segment of the industry and maintain peanut's competitiveness among other crop choices that farmers may have. These accomplishments have opened doors for breeders to control peanut traits like never before, and without using controversial and expensive GMO techniques.

Objectives and Early Findings

The PGI was born out of an industry-wide desire to tap advances in genomic technologies to enhance production and quality to assure long term sustainability of the industry. The peanut genome mapped by the PGI is of extraordinary quality, but in order to be useful in peanut breeding programs, large volumes of genomic data must be available in sophisticated on-line databases. The PGI-funded PeanutBase system provides an excellent platform for making peanut genome data publically available. The focus of the PGI has been to develop and apply new genomic technologies to peanut science, and specifically, to develop marker-assisted-selection (MAS) methodologies that lead to improved varieties for the marketplace.

Peanut proved to be a particularly difficult genome to study for several reasons, including the fact that the total genetic information in peanut is very large compared to other legumes such as soybeans, and approximately as large as the human genome. Also, like many domesticated crop plants, peanut's genetic information is contained in twice the number of chromosomes compared to its wild relatives. Because of PGI research, we now know that modern day peanut is the result of a rare but natural combination of genomes from two wild peanut species. That event happened about 10,000 years ago in the foothills of the Andes Mountains in western Argentina.

The first step in understanding the genome of modern day tetraploid (meaning it has four sets of chromosomes) peanut was to understand the genomes of its two wild diploid (with two

sets of chromosomes) parents. Researchers completed analysis of the two wild species genomes in a remarkably short time, which greatly facilitated the completion of the modern day peanut genome.

Speeding Breeding Programs

While mapping the peanut genome is a huge academic accomplishment in itself, support for breeding programs has always been the ultimate goal of the PGI. With the genome map now available, highly specific regions on that map (markers) can be discovered that confirm genes that confer specific traits. For example, PGI research has already discovered markers for high oleic oil chemistry and resistance to leafspot (early & late), root-knot nematodes and Tomato Spotted Wilt Virus (TSWV).

These high quality markers will allow breeders to quickly determine if their breeding lines contain certain traits and develop cultivars that better meet the specific (and diverse) needs of the industry. This breeding technique — known as marker-assisted selection (MAS) — allows breeders to make selections based on quick genetic tests rather than having to grow out hundreds of breeding lines to determine which ones have the desired trait. MAS does not result in genetically modified organisms (GMOs). The information generated in the PGI has already resulted in the registration of new cultivars (e.g., Tamrun, OL12, Schubert, NuMex-01, Olé, FloRun™, Georgia 14N, and TifNV-HiO/L) as well as several varieties derived from wild species that can now be utilized in cultivated peanut breeding programs. Many more are sure to come.

Advancing Science and the Body of Knowledge

As a result of PGI research, 150 scientific papers have been published to date, with more in various stages of development. That number includes the landmark paper on genome sequences of the wild diploid parents of current cultivated peanut, published in the prestigious journal “Nature Genetics.” PGI funding helped to develop an outstanding website and database known as PeanutBase, which serves as the publically available site to store and utilize peanut genomic data. All PGI data is in the public domain and cannot be owned or patented by anyone. New sources of disease resistant germplasm have been found, and additional accessions of peanuts have been added to the USDA Peanut Germplasm Collection for breeders’ use. Techniques have been developed to move genes (and the resulting traits) from wild peanut species into cultivated peanut with more desirable agronomic characteristics.

Acknowledgments

Many researchers contributed to this project, but The HudsonAlpha Institute coordinated the assembly of the final peanut genome. Dr. Jeremy Schmutz with HudsonAlpha said, “The quality and completeness of the peanut genome sequence exceeds anything to date that has been produced for a tetraploid crop plant....It’s much more complete than our cotton assemblies. It’s really, really good!”

Scientists working on the peanut genome project bring a great deal of experience from other crop genome sequencing projects, and are among the best in the world. Research partnerships were formed by scientists from the U.S., China, Japan, Brazil, Argentina, Australia, India, Israel, and several countries in Africa. The U.S. is represented by scientists at University of California-Davis; the University of Georgia at Athens and Tifton; Texas A&M University; USDA-ARS at Tifton, GA, Griffin, GA, Stillwater, OK, Ames, IA and Stoneville, MS; NC State University; Auburn University; University of Florida at Marianna; and NCGR at Santa Fe, NM. Each U.S. scientist and international collaborator had a very specific role. The research contributions of each PGI member were vital to the overall mission of developing useful genetic tools that will accelerate successful breeding programs.

Ultimately, the success of the PGI was due to the willingness of a talented group of researchers to collaborate rather than compete and was facilitated by timely advances in genomic technology and subsequent reductions in research costs.

Choosing Where to Go from Here

By any reasonable criteria, the peanut industry has already realized benefits that will return much more than their original \$6M investment. Table 5 in the body of this report gives some indication of the potential financial benefits of solving some of the key problems plaguing the industry. Some problems are inherently more difficult to solve than others, but by generating this genetic map, the PGI has laid the foundation for solving many of the genetically controlled issues facing our industry.

The PGI has also opened many doors for future research. Table 6 in body of this report summarizes some of the input already gathered from peanut industry representatives. We now need to decide which of those doors offer the greatest return on future investment and start to use our new genomic tools to make major positive impacts on our industry.

This entire report is available on the Peanut Foundation website at www.peanutfoundation.org

Highlights of Research Accomplishments

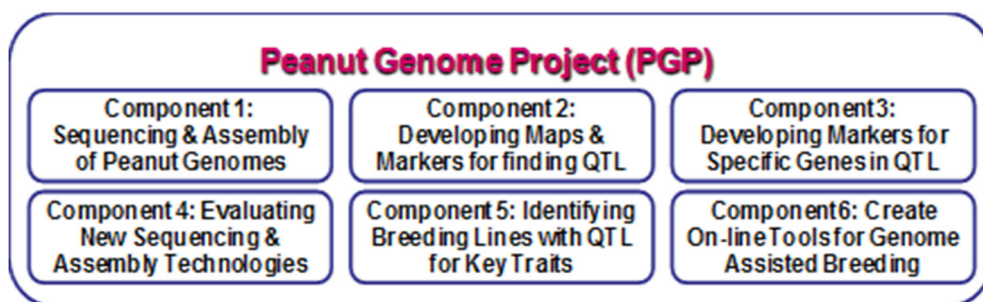
As noted earlier, the peanut industry is already reaping benefits from the PGI. Here are research highlights to date:

- The diploid wild parent species of today's cultivated peanut were sequenced and were utilized to construct the cultivated peanut genome in the proper molecular sequences and positions.
- The cultivated peanut has also been sequenced, and the assembly of the genome is 99.996% complete. HudsonAlpha says the genome is the best tetraploid genome ever assembled.
- Molecular markers have been developed from the sequencing data, which has led to the 2nd generation of a single nucleotide polymorphism (SNP) chip that is currently being utilized by breeders to evaluate breeding populations.
- Molecular markers for genes conveying resistance have been associated with late leaf spot, early leaf spot, white mold, TSWV, root knot nematode, and rust. Markers for high oleic oil chemistry have also been identified. Some of these are already being used in active breeding programs.
- Populations, for breeders use, have been developed with high levels of leaf spot resistance from wild species.
- Many genes have been identified that express traits at different developmental stages of growing peanut.
- Hybrid populations have been generated that contain an array of highly desirable characters for use by breeders to associate molecular markers with specific traits.
- Interspecific hybrid populations have been generated that are being used to introgress desirable genes from diploid peanut species into the cultivated genome.
- Peanut collections from all over the world are being genotyped to document the genetic diversity. Progress has been made in accessing the ICRISAT peanut germplasm collection in India.
- Thousands of different lines of peanut are being phenotyped for dozens of different traits. These phenotypes are being matched with genotypes which will help identify markers for even more desirable traits.
- PeanutBase, the on-line Breeders Toolbox, was developed and is widely used as a resource for genomic information and tools — as well as information about germplasm and the peanut community in general.
- The Peanut Foundation has sponsored Advances in Arachis Genomics and Biotechnology conferences, which have fostered international collaboration on peanut genomics.

Introduction

The Peanut Genome Initiative (PGI) grew out of the International Advances in *Arachis* Genomics and Biotechnology meetings in 2012 as an attempt to add new strategies to work out industry wide problems that have been difficult to solve. Geneticists and peanut breeders investigate many complex problems related to production and processing peanut products, and although progress has been achieved over the years, there have also been many frustrations associated with combining multiple traits and selecting improved traits for complexly inherited characters. When the PGI began there were relatively few molecular markers available to associate with useful genes and knowledge about the peanut genome was rudimentary. The Peanut Genome Initiative was thus organized to establish a path to create the genetic resources needed to enhance cultivar development for more productive, disease resistant, and higher quality peanuts.

Scott Jackson from the University of Georgia and Chair of the Peanut Genome Consortium (PGC) says, “Study of peanut genome structure and order makes a great detective story, where many clues are found and linked together to unlock great mysteries of genetics and gene regulation. This is exciting work”. The strategy of The Peanut Genome Project was organized into 6 components that would merge together to create a highly useful resource. By developing a strategy to concurrently investigate basic genomic research problems, develop phenotypic data for agronomically important traits, assure genetic resources are preserved, and develop data bases and tool sets, a path has been followed to solve critical problems in a timely manner.



Because peanut is a polyploid species with a highly complex genome with many gene duplications between chromosome pairs that is equal in size to the human genome, the first objective in the project was to create a high-quality assembly of peanut chromosome structure in the diploid wild species that gave rise to cultivated peanut (Figure 1). This serves as a backbone assembly that can be used to sequence and annotate the cultivated genome. In addition, the cultivated peanut genome has large numbers of nucleotide sequences that repeat many times and that pose problems for fitting the correct order of the genome in the proper configuration. Completion of the two wild species genome assemblies was accomplished in a remarkably short time. When the project started, the science associated with assembling a polyploid genome was not advanced to the level that it could be accomplished for peanut; however, by evaluating new sequencing and assembly technologies and applying them to peanut has resulted in the genomic sequence of *A. hypogaea* now being more than 99.996% completed.

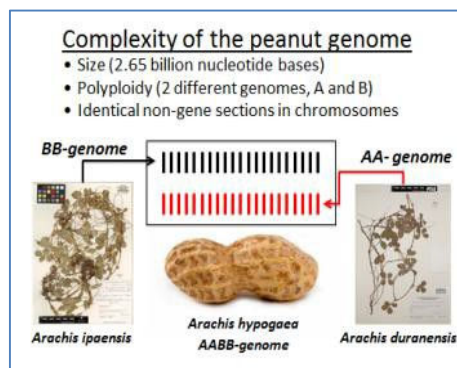


Figure 1. Diploid progenitors and genomes of the wild and cultivated species.

While a group of scientists were working of genomic sequencing, other groups were development molecular maps which are necessary to establish linkage associations with agronomically and commercially useful traits. Populations were developed to associate traits with markers and new marker technologies were developed. Importantly, the information generated in this program is being preserved and organized to be a breeder-friendly tool for crop improvement. Peanut is the only crop species where the scientific community had the foresight to simultaneously combine basic scientific advancement with application of new knowledge to solve real-life problems.

Scientists working on the peanut genome project bring a great deal of experience from other crop genome sequencing projects, and are among the best in the world. Research partnerships were formed by scientists from the U.S., China, Japan, Brazil, Argentina, Australia, India, Israel, and several countries in Africa. The U.S. is represented by scientists at University of California-Davis; the University of Georgia at Athens and Tifton; Texas A&M University; USDA-ARS at Tifton, GA, Griffin, GA, Stillwater, OK, Ames IA and Stoneville MS; NC State University; Auburn University; University of Florida at Marianna; and NCGR at Santa Fe, NM. Each U.S. scientist and their international collaborators have a very specific role within the PGI Action Plan. The research contributions of each PGI member are vital to the overall mission of developing useful genetic tools that will accelerate the breeding programs for traits such as disease resistance and drought tolerance and ones that are difficult to achieve with conventional breeding strategies. PGI members are pioneers, clearing new ground with each deliberate step. This report chronicles individual responsibilities, the current state of the genome, and the strategies to move toward utilizing the cultivated peanut genome sequence.

Peanut Genome Project
Research Technical Accomplishment Report
to the U.S. Peanut Industry
October, 2017

Research Component 1: Sequencing and assembly of the peanut genomes

The primary goal of this research component was to sequence the genome of cultivated peanut, *Arachis hypogaea*, a complicated endeavor due to the duplicated (tetraploid) nature of the genome that consists of two sub-genomes inherited from two wild ancestors. Fortunately, these ancestors were known and available to study (Figure 1). Therefore, as an initial step, the team chose to sequence the two ancestors *A. ipaënsis* and *A. duranensis* that do not have a duplicated genome structure (diploid) and each have 20 chromosomes. This would be more tractable than sequencing the tetraploid cultivated peanut with 40 chromosomes and would immediately provide information that could be used for peanut improvement. Furthermore, it would provide a framework for sequencing cultivated peanut. This was undertaken with the Beijing Genome Institute (BGI) in 2013. Both genome sequences were available to the peanut community in 2014. A summary of traits and markers derived from these two genome sequences is in Research Component 2.

In 2016 in a landmark publication entitled *The genome sequences of Arachis duranensis and Arachis ipaënsis, the diploid ancestors of cultivated peanut* was published in Nature Genetics by David Bertoli and a large group of cooperators. This publication described the assembly of each of the 10 chromosomes in the two ancestor species that contribute the A- and B-genomes to cultivated peanut. What makes this work extraordinary is the fact that the size of each ancestor genome is about the same as soybean. It took 5 years to assemble the soybean genome; the PGP has achieved twice as much in only two years. Among the many key findings was the reason for unexpected patterns of inheritance that have impeded breeding progress for late leaf spot resistance and other traits. It was shown that segments of some chromosomes in the A-genome had unexpectedly replaced its counterpart in the B-genome chromosome as a result of hybridization, leaving a void space in the donating A-genome chromosome. This means that inheritance of a gene affected by such an event would be quite low. This places greater importance on more robust DNA-markers that tag exact gene-fingerprints to find the few hybrid progeny that inherit the trait.

Once the two diploids were completed, the team prepared a roadmap for sequencing the ~2.7 billion base pairs of cultivated peanut. Initial data was provided by BGI, but the assembly of that sequence data resulted in a sequence map that was incomplete. We were unable to adequately partition out the two subgenomes. Advances in sequencing technology led to several other approaches that we were able to combine with the data from BGI to produce a high-quality genome sequence. These technologies included long-read (10s of thousands of base pairs long) PacBio that would allow very similar regions between the two sub genomes to be partitioned and not collapse, a problem with the BGI sequence. This was done in collaboration with Drs. Jeremy Schmutz (HudsonAlpha Institute, Huntsville, AL) and Brian Scheffler (USDA-ARS, Starkville, MS). Begun in 2016, the data was produced and assembled in the summer of 2017 and presented to the research team at the 2017 APRES meeting in New Mexico.

Two other complementary technologies emerged in 2015-2017, Dovetail and HiC. These allowed the generation of even longer-range information that was used to i) confirm the PacBio sequence assembly and ii) build even longer sequences that extend to full chromosome lengths. Part of the overall project was to evaluate new sequencing technologies to ensure that we produced a high quality genome sequence. These activities are summarized in Research Component 4. What has been accomplished is

amazing given that when we began were very uncertain as to whether we would be able to generate a high quality assembly of tetraploid peanut. Current assembly statistics for the cultivated peanut are shown in Table 1, in comparison to the initial BGI assembly. Note the Contig N50 of 1,600 kilobases, this is 1.6 million base pairs! This means that 50% of the contigs (longest, unbroken stretches of assembled base pairs) are longer than 1.6 million base pairs, or most of the genome is in long contigs. Dr. Schmutz, who has worked on many genome sequences including human, maize and soybean, said this was one of the best that he knows of in terms of completeness and quality.

Table 1. Comparison of initial BGI, HudsonAlpha V0 (2015) and HudsonAlpha V1 PacBio (2017) genome assemblies.

	BGI Short Read Based	V0 HudsonAlpha Short Read Based	V1 HudsonAlpha Long Read Based
Total Bases (Mb)	2,628	2,042	2,524.9
Contig N50 (Kb)	14.4	46.6	1,600
Bases in Chrs (Gb)	N/A	N/A	2,410.5 (94.6%)
Completeness	N/A	99.377%	99.996%

Research Component 2: Developing maps and markers for finding QTL

Genes control a plant's characteristics and are positioned on chromosomes. By sequencing and annotating the peanut genome DNA, the locations of almost all peanut genes are specified with coordinates on these genome sequences. Thus, a common framework to compare results from different experiments now exist for peanut from all laboratories at different locations.

Cross referencing this knowledge with genetic mapping allows us to identify chromosome regions that control important traits, like disease resistance or nutritional quality. For instance, we now know that a gene that confers resistance to fungal rust disease is located on chromosome B08 at position 126.6 million bases. The two genes that control the production of oleic acid are positioned on chromosome A09 at 114.7 million bases and chromosome B09 at 142.7 million bases.

Knowledge about gene locations can be used in a breeding program to determine which seeds from a cross are likely to produce better plants. This can be done with DNA markers, the development of which is now much easier with the genome sequences and can be a valuable complement to breeder selections which are based on how a plant performs in the field. It can shorten the time for cultivar development, and can help in breeding using new sources of desirable traits (disease resistance, drought tolerance, etc.) using older land races or wild species.

Prior to the peanut genome project, marker position was inferred from genetic maps consisting of a few hundred Simple Sequence Repeat (SSR) markers. Although still a powerful tool, the sparsity of the genetic maps led to low resolution in marker placement which can lead to markers being in "incomplete linkage" with the trait of interest. This means that selecting plants for crop improvement

with a marker not in a tight linkage with the trait of interest is not perfect, and unwanted genotypes remain in the population. The genome sequences fostered the development of new technologies to identify and develop markers, increased the efficiency of identifying and selecting for wild relative genomic introgressions conferring strong disease resistances, and allowed the precise physical positioning of molecular markers, ensuring complete linkage with traits of interest and vastly improving marker-assisted breeding efforts.

New genomic approaches proved useful in processing over 55 million DNA markers or SNPs found in the DNA from 6 peanut lines representing all market types. The number was culled to 60,000 validated SNPs that provide high definition coverage of each peanut chromosome. The genome origin of 96% of those SNPs was determined. These DNA markers or SNP's were placed on a "Gene Chip" or "SNP Chip" that enables breeders to select several different traits simultaneous, instead of the current limitation of one-at-a-time. Each of the 60,000 DNA fragments is placed in an individual well on the "chips". Useful gene markers are found when a DNA fragment for a specific trait in a breeding line sticks to a chip-fragment in a 'well'. Special software associates the 'matches' with a position on a chromosome, and phenotypic data associates the position with a trait. This tells the breeder if the breeding line contains desired genes, and helps eliminate lines without having to grow them out in multi-year evaluation trials.

A. Trait-linked markers for segments of wild species for introgression into cultivated peanut

Pest and disease control accounts for about ¼ of peanut production costs in the U.S. Development and adoption of resistant cultivars is the most cost efficient and effective way to control the spread of diseases, reduce yield losses and maximize economic gains. Some cultivated peanut accessions have moderate levels of resistance, but the secondary gene pool consisting of wild peanut species is a source of extremely high levels of resistance for multiple pathogens. Wild species, however, have many undesirable agronomic traits that are a disincentive to their use in breeding, including small seeds, long fragile pegs, long pod constriction, and small biomass. The identification of genomic regions that harbor disease resistance in wild species is the first step in the implementation of marker assisted selection that can speed the introgression of wild disease resistances and the elimination of linkage drag with undesirable traits. Using the sequenced genomes of *A. duranensis* and *A. ipaënsis* (that are used as proxies of the A and B subgenomes of peanut, respectively), regions have now been identified that control several disease resistances, and as predicted, they co-localize with major clusters of Resistance Genes (Figure 2).

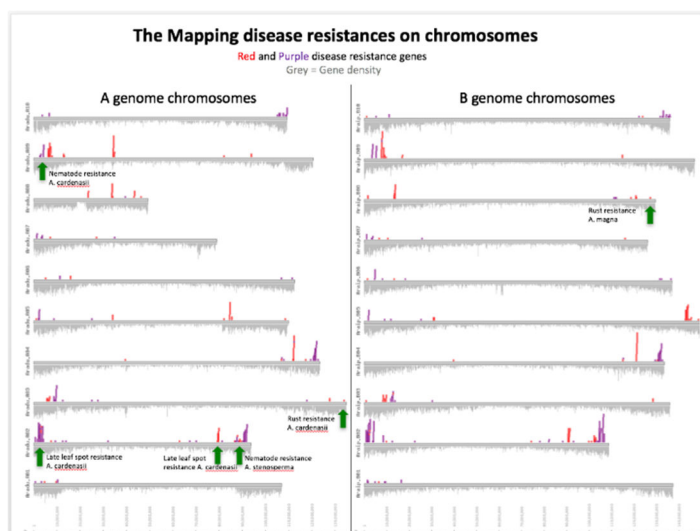


Figure 2. Genomic regions of alleles for resistance derived from wild species mapped onto peanut A and B chromosomes (green arrows), and locations of Resistance Genes (colored blocks). Most disease resistances collocate with clusters of Resistance Genes.

Genomic regions and markers linked to different components of rust, root-knot nematode (RKN) and late leaf spot (LLS) resistance, and domestication traits from different wild species have been identified. These species are diploid, therefore, hybrids with *A. hypogaea* are sterile so introgression of resistance alleles from into allotetraploid cultivated peanut involves special crossing schemes to overcome the ploidy and sterility problems. Several wild species breeding programs have been able to incorporate genes for multiple resistances into cultivated peanut and backcrossing programs are being implemented to increase yields and quality. Marker assisted breeding will greatly facilitate this process in the future. The following is a summary of several examples where the PGI has aided utilization of wild peanut species.

1. Markers for leaf spot resistance from *Arachis cardenasii*

Recently a peanut breeding line was recovered for use in the U.S. that is highly resistant to late leaf spot (CS-16). This line derives from a cross made during the 1960s between cultivated peanut and a highly resistant wild species *A. cardenasii* at NC State University. With the information available from the genome sequences, the genetic contribution of the wild species to this line was identified and then DNA markers specific for the *A. cardenasii* introgressed regions were developed. These lines have been crossed and backcrossed with elite peanut lines from the southeast U.S. (such as TifNV-High O/L and Georgia-12Y), and the markers are being used to select resistant progenies.

Other *A. cardenasii* derived lines have been selected for early leaf spot resistance and hybridized with cultivars and breeding lines at NC State University. Many of these lines now have extremely high levels of resistance (Figure 3) as well as large seeds and moderate to good yields. One line was used as parent to the cultivar Bailey, a predominant Virginia-type peanut in the NC-VC growing region.



Figure 3. *Arachis cardenasii* derived lines resistant to early leaf spot as compared to check cultivar (left, 1st row). Photo recorded on October 10, 2016.

Field resistance to early and late leaf spot segregates in the CAP population called 1801, which is derived from a cross between Florida-07 and GP-NC WS 16. GP-NC WS 16 is a leaf spot resistant line released from NC State University with *A. cardenasii* in the pedigree. This population has been phenotyped for early leaf spot in 2013 and 2014 and late leaf spot in 2012 – 2016.

A genetic analysis named QTL-seq (Quantitative Trait Locus SEquencing) was carried out at the University of Georgia that used bulked DNA from the top 5% resistant lines and the bottom 5%

susceptible lines over the years 2012 – 2014 data. Three genomic regions were identified that explained 15% of the total variance for the traits combined. Markers were developed to select for these regions and were validated in a non-spray test in Tifton, GA (2015). The markers were further validated to select for resistance in the population data from 2015 and 2016 in Tifton and 2015 in Headlands. It was found that one of these markers could be used to select for resistance to late leaf spot and the other two markers could be used to select for resistance to early leaf spot (Figure 4.). These markers have been deployed in Corley Holbrook's breeding program beginning in the spring of 2016 to select for late leaf spot resistance. These data have been submitted for public dissemination.

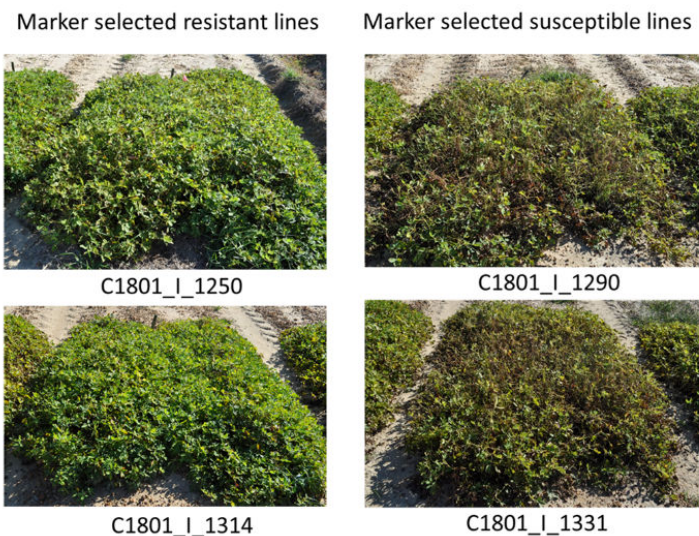


Figure 4. Plots from blind validation test of LLS markers. Left shows genotypes selected as resistant with markers and right shows genotypes selected as susceptible on the last day of rating in a non-sprayed test.

2. Markers for *Meloidogyne arenaria* (RKN) resistance – from *A. cardenasii* and *A. stenoperma*

At present, U.S. farmers rely on a single source of nematode resistance that is originally from *A. cardenasii*. Resistance to nematodes was introduced into cultivated peanut from the wild species *A. cardenasii* and deployed in the cultivars such as NemaTam, Tifguard, Georgia-14N, and TifNV-High O/L. This near immunity resistance was an excellent target for marker-assisted selection due to the rigors of phenotyping for nematode resistance. Molecular markers for this resistance have been used for some time to assist selections in breeding. However, the new information from the PGI has been fundamental in dissecting the genetic basis of this resistance, and discovering that it resides in two genes (one with a strong effect, one with a weak effect), that occasionally have been found to recombine. In the past, this has led to very occasional misleading results with the old markers. The marker that had been successfully used to select for resistance showed broken linkage with the resistance (B. Branch, UGA). This led to a re-evaluation of the physical position of the trait. Using the newly released genome sequences of *A. duranensis* and *A. ipaënsis*, new methods were developed, that allowed a genetics study to be carried out (P. Ozias-Akins & J. Clevenger). With these methods, it was possible to, for the first time, physically map with much higher precision, the wild introgression that gives cultivated peanut resistance. The exact location of the underlying gene conferring resistance was then mapped to a small region including 3.6% (4 million base pairs) of chromosome A09. The marker previously used to select for resistance was physically far away from this small region, which confirms that recombination between that marker and the resistance was possible. A gene expression analysis led to the identification of a candidate gene for resistance. A new, high throughput SNP marker was developed for deployment in breeding programs that is within the newly identified resistance region. This gene is being studied to confirm its function as a nematode resistance gene (Ozias-Akins lab). If it is confirmed as the resistance gene, the newly developed marker will be in complete linkage with resistance and will always select true. To date the new marker has been used in marker-assisted breeding efforts since 2016 in Corley Holbrook and Bill Branch's breeding programs. The results of

the study, including marker information are now in public domain. Now, new markers have been developed for both of these genes making marker assisted selection for nematode resistance faster and more reliable.

The reliance on a single source of nematode resistance makes the industry vulnerable to resistance being broken by the pest. As a safeguard against this, the wild species *A. stenosperma* has been identified as a new source of resistance to several pests and diseases, including RKN. Three genome regions that strongly reduce nematode root galling and egg production have been identified. They reside on chromosomes A02, A04 and A09. In parallel, the allotetraploid hybrids (*A. batizocoi* × *A. stenosperma*)^{4x}, (*A. gregoryi* × *A. stenosperma*)^{4x}, (*A. valida* × *A. stenosperma*)^{4x} and (*A. magna* × *A. stenosperma*)^{4x}, that are at the same chromosome number and compatible with cultivated peanut have been made. The induced allotetraploid (*A. batizocoi* × *A. stenosperma*)^{4x} has been incorporated in a breeding program in Georgia and NC. Crosses with elite cultivars and advanced peanut lines have been done and early selection for resistance using molecular markers. We envisage that we will have a new source of nematode resistance for peanut producers.

3. Markers for rust resistance – from *Arachis magna*

Rust resistance was identified in the wild species *Arachis magna*. The genome region that confers resistance resides at the end of chromosome B08. Markers were developed for this region so that resistance can be tracked and selected in crosses. Since *A. magna* is diploid and, therefore hybrids with the cultivated peanut are sterile, the allotetraploid (*A. magna* × *A. stenosperma*)^{4x} was produced so that fertility will be obtained in crosses. Programs are not using this material for introgression of the rust-resistance region into elite cultivars and the markers are being used for early selection of progenies.

4. Seed size – from *A. cardenasii* and *A. ipaënsis*

Pod and seed size is segregating in the CAP population derived from the cross between Florida-07 and GP-NC WS 16. A major genomic region (variation explained 36 – 63%) was identified on chromosome A05 controlling pod size, seed size, 100 pod weight, 100 seed weight, and pod area over two years of data (2012 and 2014). This region was identified using genotyping data from the Axiom *Arachis* v1 SNP array. Although the wild species, *A. ipaënsis* has smaller seeds than cultivated peanut, a genome region from *A. ipaënsis* at the end of chromosome 06 has been identified that increases seed size in peanut (Figure 5).

Using a separate portion of the population that was not used for QTL analysis, genotypes were selected to be large or small seeded using genotyping information from the Axiom *Arachis* SNP array. The results showed that marker-assisted selection for seed size had an average effect of almost a quarter of a gram difference in single seed weight and a half a gram difference in pod weight.

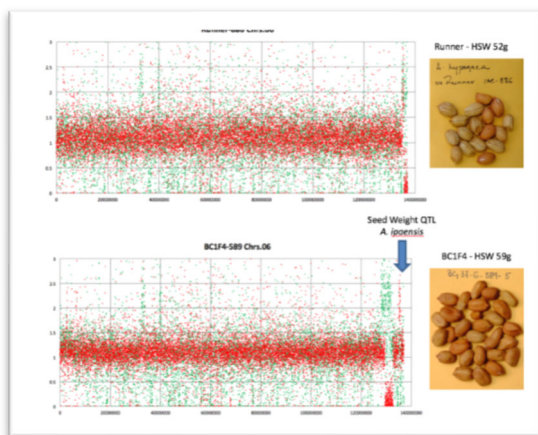


Figure 5. Genomic region coding for seed size at the end of chromosome 6. A BC₁F₄ line, with 25% wild genome has higher HSW (weight of 100 seeds) than its recurrent cultivated peanut parent.

5. Markers for other agronomic characteristics – from *A. ipaënsis*.

Although wild plants have very poor agronomic characteristics, experience with other crops has shown that genes from wild species can, counterintuitively, improve the agronomic characteristics of crop plants. In peanut, research is now beginning to show the same. Using the genome sequences, markers have been developed and are being used to assist selections in breeding. Markers have also been developed for increased haulm. Three cultivars, with larger seed size and increased haulm, derived from wild species have been released in Senegal, and a cultivar release is expected next year in Brazil.

B. Trait-linked markers for segments of cultivated peanut

1) White Mold Resistance

Another CAP population, labeled 1799, which is derived from a cross between Tifrunner and NC 3033 has been investigated for field resistance/tolerance to white mold by Tim Brenneman in Tifton, GA over the years 2013-2016. Based on these data, resistant and susceptible bulked genotypes were selected and subjected to QTL-seq analysis. The analysis resulted in the identification of two QTL that when selected for had a significant effect on white mold rating and percentage of unaffected plants per plot over the three years of data (Figure 6).

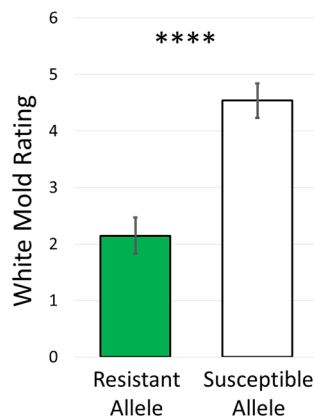


Figure 6. Significant increase in field resistance to white mold over three years of data in Tifton, GA by selecting resistant alleles with only 2 markers.

Markers were developed for selecting for these QTL and a blind validation test is being conducted in Tifton, GA to test the utility of these QTL for inclusion into marker-assisted selection breeding programs. Markers for white mold resistance would be a huge advance for breeders as testing for white mold in the field is confounded heavily by large plot variance. The analysis to identify these markers relies completely on the availability of reference genome sequences and could not have been performed without the PGP.

2) High density SNP-based genetic maps for studying resistance to peanut foliar diseases

Molecular mapping is a prerequisite for genetic dissection of genomic regions linked to the traits of interest and for molecular marker development that will be used in marker-assisted breeding. Host resistance to diseases, such as early leaf spot and late leaf spot (ELS and LLS) and Tomato spotted wilt virus (TSWV), is critical for increasing the yield and reducing the cost for peanut growers. We developed two recombinant inbred line (RIL) mapping populations, which were derived from Tifrunner × GT-C20 (the “T”) and SunOleic 97R × NC 94022 (the “S”). As part of the International PGI, these two RILs were genotyped by whole genome re-sequencing (WGRS) technology using two diploid genome sequences as reference genomes. Using new technologies developed as part of the genome project, high-density SNP-based genetic maps were constructed as tools to identify genomic regions controlling disease resistance. For all three diseases, ELS, LLS and TSWV, there were major genomic

regions identified with over 40% phenotypic variation explained (PVE). The markers linked to TSWV resistance will be a major boon to marker-assisted breeding efforts (Figure 7).



Figure 7. Segregation of disease resistance in the “T” population. High density genetic mapping has led to the discovery of markers tightly linked to resistance that can be deployed in breeding programs for increased efficiency of cultivar improvement.

3) Axiom genotyping array

High throughput genotyping arrays have been deployed in other major crops and have been inestimably beneficial. These arrays allow, for a small cost, the assessment of tens of thousands (in some cases hundreds of thousands to millions) of genomic regions with one analysis. Further, entire populations can be assessed at a time, providing optimal speed for genetic discovery. Once analyzed, these results can be deposited in public databases for community-wide use. For example, Soybean has public data from a 50,000 marker SNP array on its entire germplasm collection. This has allowed *in silico* genetic analyses, where marker information is downloaded along with public phenotyping information and markers linked to beneficial traits are identified and deployed within an afternoon.

The peanut genome sequences laid the framework for this technology to be developed for the peanut community and the array technologies developed rely heavily on the genome project. To date, two SNP arrays have been developed, *Axiom Arachis v1* (58,000 markers), and *Axiom Arachis v2* (48,000 markers). Although version 1 is very successful, knowledge gained from its development was used to vastly improve the marker sets on version 2 which is also half the cost of version 1. To date, the entire US minicore collection has been genotyped using the version 1 and 2 arrays, and multiple populations have also been genotyped. The array technology, which could not have been developed without the genome sequences, has and will continue to markedly increase the efficiency of molecular breeding efforts.

4. Additional mapping populations

In addition, two next-generation mapping populations, MAGIC (Multi-parents Advanced Generation Inter-Crossing) and MARS (marker assisted recurrent selection), are in development for future improvement of mapping resolution and combination of the favorable alleles to pyramid different disease resistance traits into breeding lines by marker-assisted selection. One MAGIC population has eight founders (SunOleic 97R, NC94022, Tifrunner, GT-C20, plus Florida 07, SPT0606, Georgia 13M, and TifNV-High O/L) for future study of gene/trait associations with greater resolution and one MARS population to cumulate the favorable alleles in the “S” population with eight lines. Individuals of the MAGIC from these eight-way F_2 families will be selfed through single seed descent (SSD) with a target population of 1500 recombinant inbred lines (RILs). The number of RILs contributed per family will be kept as equal as possible within the constraints of variable seed set.

Research Component 3: Developing markers for specific genes in QTL

Creating an expressed gene atlas was a major accomplishment under this component. Initial work focused on a developmental series, early vegetative to late reproductive (Figure 8), for the reference tetraploid genotype ‘Tifrunner’. Next generation sequencing technologies enabled deep sequencing of expressed genes by a method known as RNA-seq. The depth of sequencing was important for identifying genes that were expressed at low levels and to support robust statistical analysis of gene expression differences between peanut organs and tissues. Another important use of these data was annotation of the diploid genome sequences (Component 1). Tetraploid expressed gene sequence was more extensive than what had been obtained from the diploid progenitors, and since genes are the more slowly evolving fraction of genomes, the high level of similarity between tetraploid and diploid sequence facilitated annotation of diploid genomes. The annotation information and homeology between genomes was published in Bertoli et al. in 2016 and in *PeanutBase.org*.

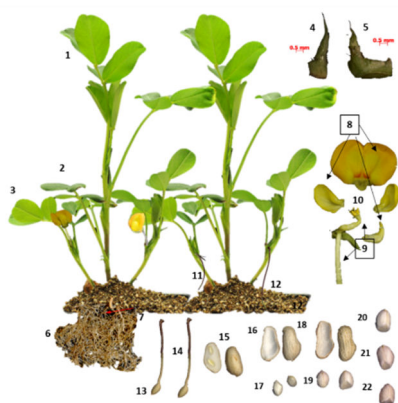


Figure 8. Organs/tissues sampled for developmental series gene expression atlas.

PeanutBase presents tracks of expression data for each annotated gene (Figure 9) including the more than 8800 homeologs. In total these expression data provided evidence to support the identification of 36,734 and 41,840 genes in the two respective progenitor genomes. Finally, data for gene expression differences throughout the developmental series are presented in a second visual format through the University of Toronto hosted eFP Browser (http://bar.utoronto.ca/efp_arachis/cgi-bin/efpWeb.cgi). This browser allows visualization of gene expression levels overlaid on a cartoon of the plant, a format more intuitive for breeders and developmental biologists (Figure 10). All data analyses and information on access were published by Clevenger et al. The expression data provide an extensive resource for breeders and biologists to identify candidate genes that are associated with important growth and developmental processes affecting the industry goals of improved yield and quality.



Figure 9. Visualization of gene expression data at PeanutBase.org (bottom tracks). A gene involved in oil biosynthesis is expressed mainly in developing seeds.

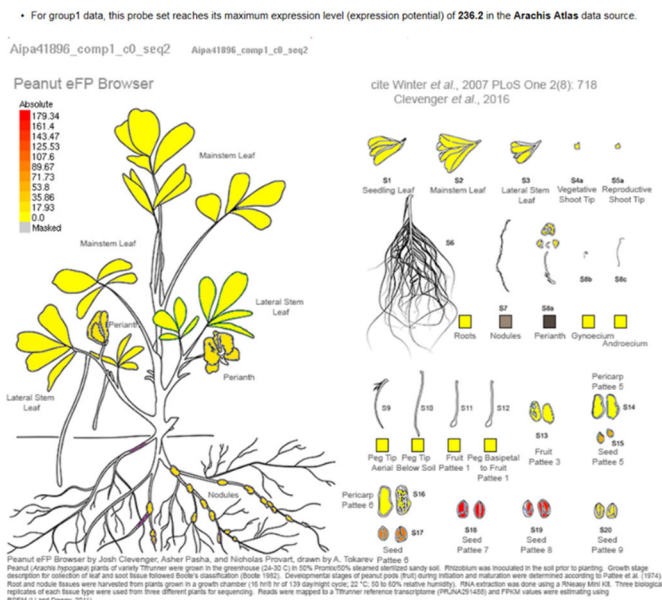


Figure 10. Visualization of gene expression in eFP Browser. A gene involved in oil biosynthesis is expressed mainly in developing seeds.

Expanding upon generation of these initial expression data are other data sets that explore biotic and abiotic stress responses in relevant genotypes (Table 2). Over 1,4 trillion base pairs of gene expression sequence data to date has been generated in the Ozias-Akins lab alone. Other labs such as that of Mark Burow, Texas A&M, Lubbock, also have produced considerable expressed gene data from leaf roots, flowers, and pods of multiple peanut market classes and diploid relatives. These data were primarily used for SNP discovery and diversity analysis (Component 2) rather than gene expression analysis.

Table 2. Summary of RNA-seq experiments.

Genotype/Treatment Series	Libraries Sequenced (lines x tissues/time course/controls x reps)
Tifrunner – normal development	1 x 22 x 2-3=58
Tifguard, Gregory, RILs 46,48 – nematode infected	4 x 5 x 3=60
Florida-07, SPT-06-06 – LLS infected	2 x 6 x 3=36
Tifrunner, Florida-07, C76-16, A72, NC3033, Tifguard – Aspergillus/aflatoxin contaminated	6 x 3 x 2-3=33
Tifrunner – Rhizobium inoculated	1 x 6 x 3=18
Tifrunner, NC3033 – pods/seeds	2 x 7 x 6-7=80

The Ozias-Akins group examined gene expression in inoculated and uninoculated roots of nematode resistant cultivar, Tifguard, compared with nematode susceptible Gregory. Analysis of gene expression data not only led to the characterization of gene expression networks, but also the identification of sequence polymorphisms useful for the development of molecular markers associated with resistance and identification of a candidate gene for resistance (Component 2). The markers linked with nematode resistance have been used extensively in the breeding programs of Corley Holbrook and Bill Branch in Georgia.

Research Component 4: Evaluating new sequencing & assembly technologies

As part of the overall project, the team evaluated new and emerging sequencing technologies to ensure that the end result of the project was a high quality genome sequence that would advance peanut improvement and production. When sequencing began in 2013, Illumina-based sequencing was the standard. While relatively inexpensive, it was limited in that the length of each sequence read was short, usually less than 200 base pairs. We anticipated that difficulties of assembling the large and highly duplicated peanut genome would result from such short reads and the inability to differentiate duplicated regions. In fact, the first Illumina-based peanut genome sequence was composed of short contigs (Table 1) and many collapsed duplicated regions.

Moleculo was one of the first technologies we explored (2014-2015). This was a technology developed, still based on Illumina sequencing, that would produce longer assembled regions. We incorporated this data into the two diploid ancestor genomes, but found that it did not contribute significantly to the genome assembly as the regions that Molecu was able to produce were already represented in the Illumina-based genome sequence. Thus, this was not pursued further.

PacBio sequencing had been around for some time, but the cost, through-put and error rate precluded its use early in the project. However, in 2015-2016, significant advances in cost reduction and read length (number of base pairs) made it feasible for peanut. We undertook with Drs. Jeremy Schmutz and Brian Scheffler to produce 80-90x coverage of the peanut genome. This was finished early summer 2017 and an initial assembly presented at APRES (Table 1). A finalized assembly with genetic information will be public during the winter of 2017/2018.

Two other technologies are being incorporated into the PacBio assembly, **Dovetail** and **HiC**. Both technologies are Illumina-based but the library preparation allows one to make long-range connections between sequences. This has allowed the team to validate the PacBio genome assembly by analyzing longer-range sequence information that the assembly team could use to patch together pieces that belonged together, or to break apart pieces that were incorrectly joined. For example, in the initial PacBio assembly, Dovetail data was used to make ~800 breaks of potentially incorrectly joined pieces of sequence. Just one example of how these data are helping to produce an accurate genome sequence for the peanut research community.

Collaborators at Hudson-Alpha (Huntsville, AL) evaluated three new software programs for ability to assemble wild and cultivated peanut genomes. ‘Meraculous™’ assembler technology (used by JGI in the U.S. Department of Energy) performed best. ‘Meraculous’ assemblies correctly placed 75% of all wild peanut genes in the correct cultivated peanut genome.

Component 5: Identifying breeding lines with QTLs for key traits

Phenotyping and genotyping of genetic resources to identify markers that can be used for Marker Assisted Selection is an important first step in identifying QTLs for agronomically useful traits. This research should result in developing molecular markers for economically significant traits that will greatly improve the speed and efficiency of all peanut breeding programs. Ultimately, this research will supply the peanut industry in the U.S. with new and improved cultivars. The use of these new cultivars will increase production efficiency because of improved yield potential, less loss due to plant diseases and nematodes, and decrease use of expensive pesticides.

Peanut has lagged other crops in use of molecular genetic technology for cultivar development due in part to lack of investment, but also because of low levels of molecular polymorphism among cultivated varieties. The early genomic technologies (isozyme, RFLP [Restriction Fragment Length Polymorphism], AFLP [Amplified Fragment Length Polymorphism], RAPD [Random Amplified Polymorphic DNA], and SCAR [Sequence Characterized Amplified Region]) showed extremely low levels of polymorphism in *A. hypogaea*. In addition to low levels of genetic variation, the tetraploid nature of the genome of cultivated peanut has been responsible for slow progress in the area of developing genomic resources such as molecular markers and genetic maps.

Recent advances in molecular genetic technology have allowed researchers to more precisely measure genetic polymorphism and enabled the identification of molecular markers of QTL's for several economically significant traits. Efforts to shepherd initiatives for increased research in peanut genomics began at the 2001 U.S. Legume Crops Genomics Workshop and then at subsequent meetings of the International Peanut Genome Consortium. These efforts have resulted in quantum leaps of knowledge about the peanut genome, and have facilitated ongoing marker assisted breeding programs. These efforts have also stimulated the development of molecular genetic tools and RIL (recombinant inbred line) populations that should result in additional quantum leaps of knowledge. The peanut genome sequencing project was launched in 2012 using 'Tifrunner' as the reference genotype. The two progenitor diploids (*A. duranensis* and *A. ipaënsis*) were also sequenced. This should result in the development of additional molecular tools that will greatly advance peanut cultivar development.

Translation of genomics to breeding requires association of molecular markers with phenotypes and the implementation of cost-effective marker-assisted selection (MAS). Because of sparse genomic information for peanut, we have only been able to take advantage of MAS for a limited number of traits, primarily nematode resistance and high oleic acid. Yet in such a narrowly focused project, at least 3-fold gain in the speed of selection was achieved leading to the release of TifNV-Hi O/L as a cultivar. These two traits also are relatively simply inherited in peanut which is in contrast to many other traits for which breeders must select, e.g. disease resistance, drought tolerance, and yield. With the initiation of the PGI, the genome sequences of *A. duranensis* (A-genome) and *Arachis ipaënsis* (B-genome) became available through peanutbase.org which greatly expedited marker discovery. Development of the Axiom Single Nucleotide Polymorphism (SNP) array containing 58,233 SNPs provides a genotyping platform for structured and natural populations harboring genetic diversity. The bottleneck has become phenotyping because of the need to collect multiple years of data in multiple locations for complex traits that have significant genotype by environment effects. Reliable phenotyping of populations is essential for association of markers with traits, and phenotyping requires replicated testing over multiple years and ultimately in multiple environments. Anticipating this need, the peanut breeding community initiated the development of 16 recombinant inbred line (RIL) populations in 2008 using a nested mating design (Table 3).

Table 3. Attributes for parents of 16 RIL populations. Levels of disease resistance are indicated as unknown (U) or susceptible (S) or resistant (R) qualified as highly (H) or moderately (M).

Parent	Common Or Unique Parent	Market Class	Oleic Acid	TSWV	Early Leaf Spot	Late Leaf Spot	White Mold	Sclerotinia	CBR
Tifrunner	Common	Runner	L	R	MR	MR	S	U	MR
Florida-07	Common	Runner	H	MR	S	MR	MR	U	HS
N08082olJCT	Unique	Virginia	H	MR	MS	U	U	MR	HS
C76-16	Unique	Runner	L	MR	U	U	U	U	U
NC 3033	Unique	Exotic	L	HS	MS	HS	R	U	HR
NM Valencia A	Unique	Valencia	L	S	S	S	HS	HS	U
Olin	Unique	Spanish	H	MS	S	S	U	R	U
SSD6	Unique	Exotic	L	HR	U	U	U	U	U
SPT 06-06	Unique	Exotic	L	MR	HR	MR	U	U	U
Florunner	Unique	Runner	L	HS	S	S	S	S	S

The development, maintenance, and characterization of these RIL populations has been a monumental team effort. This work was led by Corley Holbrook (USDA-ARS, Tifton, GA) with partners at the University of Georgia-Tifton and Athens, NC State University, University of Florida, Auburn University, and Texas A&M University. Sixteen RIL mapping populations were created with parents that maximized genetic diversity of marker identification and practical breeding objectives. Two modern runner cultivars (Tifrunner and Florida-07) were selected as common parents because runner cultivars account for about 80% of the production in the U.S. In addition, eight unique male parents were selected to supply diversity across market classes and are donors of favorable genes for enhancing drought tolerance and resistance to most important diseases of peanut in the U.S. Seed increase for these large populations was necessary to provide the community with material for extensive phenotyping.

From a cultivar development standpoint, these advances in technology have enabled the identification of molecular markers associated with quantitative trait loci (QTL) for several economically significant traits. The first successful example of marker assisted selection (MAS) was the introgression of nematode resistance through an amphidiploid pathway into cultivated peanut, and the subsequent development of nematode resistant cultivars, 'COAN' and 'NemaTAM'. Recent research has resulted in developing improved markers for nematode resistance and the discovery of molecular markers associated with high oleic acid content, resistance to TSWV, resistance to foliar diseases, rust and late leaf spot, , drought tolerance, and seed biochemical traits.

For traits such as high oleic fatty acid content and nematode resistance, identification of major QTLs have enabled efficient marker-assisted backcrossing. Recently, several peanut cultivars have been developed using marker assisted breeding, and more are anticipated in the near future.

1. Identification of Genetic Markers for Disease Resistance

During the early 1990s tomato spotted wilt, caused by tomato spotted wilt virus (TSWV) became a major limiting factor for peanut production in the southeastern U.S. with loss estimates of over \$40 million in 1997 in Georgia alone. Fortunately, natural resistance to TSWV was available, and development and deployment of TSWV-resistant cultivars has been extremely successful for reduction of disease incidence and subsequent damage. Resistance to TSWV is essential for any cultivar developed for use in the southeastern U.S., and higher levels of resistance are needed. Field scoring for disease incidence and severity is commonly used for resistant line selection. Field TSWV ratings were taken for RIL populations Tifrunner × NC3033 and Florida-07 × SPT06-06 for multiple years, and several QTLs were identified, each explaining over 15% of the phenotypic variation.

Early leaf spot and late leaf spot diseases caused by *Cercospora arachidicola* and *Cercosporidium personatum*, respectively, are among the most destructive diseases of peanut worldwide. One or both of these diseases occur(s) in almost every peanut growing area of the world. Either can cause total defoliation of susceptible plants and yield losses of greater than 50%. In the U.S., these diseases are currently controlled primarily by multiple applications of fungicides, with cost of control estimated at \$26.8 million in 2009 for Georgia alone. Use of resistant cultivars is the only economically feasible method to control these diseases, and resistant cultivars could greatly reduce the cost of production of peanut in the U.S. Multiple years of field ratings for late leaf spot resistance was performed for the RIL population Florida-07 x SPT06-06. DNA from the best five and the worst five RILs in response to late leaf spot challenge were pooled for sequencing.

As for the late leaf spot trait, QTL-seq was performed with the most resistant and most susceptible lines from the Holbrook half of the population. Three KASPar markers linked to late leaf spot resistant QTLs distributed on chromosome B3, A5 and B5 were developed. Twelve resistant or susceptible RILs from Isleib's half of this population were selected according to their late leaf spot marker profile. These genotypically selected RILs and Holbrook's RILs used for QTL-seq analysis were planted at the Gibbs farm without fungicide spray in 2016. Late leaf spot field rating using the Florida rating scale exhibited clear phenotypic separation matching very well with genotypic prediction (Figure 11). These markers are confirmed to be associated with LLS and have been implemented in the ARS/UGA breeding program at Tifton. The development of new elite cultivars packaged with strong late leaf spot resistance will be accelerated with these leaf spot resistant markers.

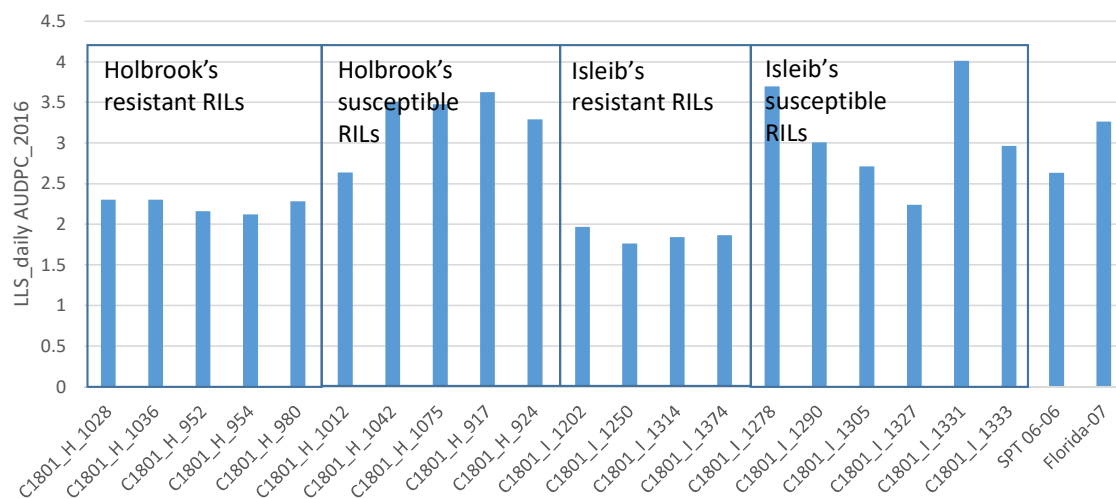


Figure 11. Late leaf spot field trial with C1801=Florida07 x SPT06-06 RILs selected by genetic markers associated with QTL conditioning LLS resistance.

In the U.S., all cultivars with resistance to the peanut root-knot nematode derive their resistance from an introgressed segment on chromosome A09 from the wild species *A. cardenasii*. This introgression was produced at Texas A&M University and effectively saved peanut production in badly infested areas. Other lines from this same wild species were produced at North Carolina State University. These lines were transferred to ICRISAT India in the 1980's, and one, CS 16, was renamed ICGV 86687. From there it was transferred to Campinas Breeding Institute, Brazil and renamed IAC 69007. In Brazil it was crossed with an elite cultivar and selected progeny lines, called the "300 series", show exceptional resistance to LLS and rust. In the U.S., CS 16 was lost years ago. Recently through collaboration, a 300 Series line and the original CS 16 have been reintroduced to the U.S. and are now being used in U.S. breeding programs. These lines contain three well-defined, genetically characterized *A. cardenasii* segments, different from the ones that confer nematode resistance. They have been extensively field tested under very high disease pressure and shown to be resistant (Figure 12). Genetic markers have been developed to detect the presence of these three *A. cardenasii* chromosome segments in progeny. These lines and markers provide the tools necessary to achieve the production of elite USA-adapted

peanut advanced lines that harbor these chromosome segments from *A. cardenasii*. These lines are almost certain to show strong and durable resistance to LLS and rust (and possibly early leaf spot) and should reduce farmer's costs associated with fungicide sprays.



Figure 12. Field of F₃ families descending from crosses *A. hypogaea* x F₄ segregants of 300 Series. No fungicide sprays have been used, and disease pressure is intense. The elite cultivated background, used in this breeding are the mostly dead plants in the two left-hand borders.

2. Identification of Genetic Markers for Yield, and Seed and Pod Traits

In addition to yield, seed and pod traits can significantly affect the economic return for growers and other segments of the US peanut industry. Multiple years of data were generated for Florida-07 x SPT06-06 and Tifrunner x NC3033 populations for yield, pod weight, seed weight, and shelling percentage. Mapping these post-harvest traits has revealed several QTLs contributing to these traits of economic importance.

3. Identification of Genetic Markers for Improving Drought Resistance

The RIL populations from Florida-07 x C76-16 was planted in the field as replicated trials with three replications, and grown under reduced irrigation in West Texas. In this region, mean rainfall in June-August never exceeds 50% of plant evapotranspiration. Drought was imposed by skipping irrigations, and water deficit irrigation began approx. 40 DAP (beginning of flowering) and continued through 105 DAP (mid pod fill, near cessation of flowering) when genetic variability for drought response is low, and rainfall begins to increase (September and October). Target of water deficit was between 25% PET and 50% PET, depending upon rainfall. In a second experiment, approximately 20 contrasting selections with the highest yield and grades, as well as parental checks, were tested in a rainout shelter to estimate T and TE. Seed were sown in pots, irrigated to field capacity daily, and grown to approx. 30 days, after which half the seedlings were harvested to determine initial pre-stress biomass. Plants were grown to approx. 90 DAP or until the permanent wilting point of 50% of the plants was observed, at which time final weights were taken and water usage, dry matter accumulation measured, and T and TE were calculated. Markers were then associated with stress tolerance.

4. Mining the U.S. Germplasm Collection

The U.S. Mini-core collection (112 accessions selected to represent the genetic diversity of the >9,000 accessions in the entire U.S. Peanut Germplasm Collection) were phenotyped for variation in yield, seed size, grade, resistance to early leaf spot, late leaf spot, white mold, TSWV, and seed composition (fatty acids, flavonoids-including resveratrol, and minerals). Four accessions (PI 158854, PI 196622, PI 268868, and PI 371521) were identified that had a major QTL (Ah3) for resistance to late leaf spot. Three accessions (PI 356004, PI 493880, and PI 496401) were identified as resistant sources to TSWV, and genetic markers were identified that are associated with resistance to TSWV. Markers associated with flavor components and chemical content were also identified.

The exomes of the U.S. Mini-core collection have recently been sequenced. Herewith, the vast majority of the biologically relevant gene variants present in the collection becomes accessible for gene hunting and analysis. The data will facilitate establishing the genetic basis of the phenotypic differences studied in the collection and will empower the Mini-core collection as a breeding tool with a cataloged

gene variance. The exome data set can be further mined for millions of gene-associated SNP and other markers.

5. What was learned from phenotypic analysis of the U.S. Peanut Germplasm Collection?

Previous results of genotyping and phenotyping the U. S. minicore collection identified markers for tolerance to drought stress in peanut. These were tested in two breeding populations developed to combine drought tolerance with other important traits, such as high oleic oil chemistry, early maturity, and resistance to root-knot nematodes. Results of testing demonstrated that it was possible to select for markers for multiple traits at the same time. In one population, drought tolerance, pest resistance, and improved oil chemistry were all selected using DNA markers. Different markers for responses under drought were confirmed. Using these markers, it was possible to select for an increased yield under drought early in the breeding program. These markers explained from 5% to 7% of phenotypic variability per marker, and selecting several simultaneously identified plants that out-yielded unselected materials by up to 48% in the F₂ generation (Figure 13). Without markers, such a selection could only have been performed later in the breeding program. In a second population, there were relatively few differences between parents, and selection with markers increased yield by 10%. As

more sequencing data is incorporated, we expect that it will be possible to make more gains in the future.

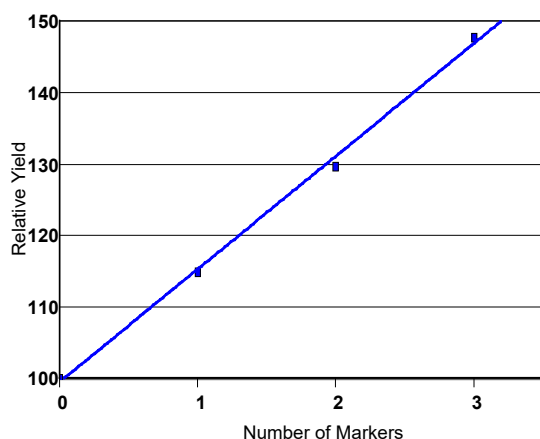


Figure 13. Yield (%) relative to the overall mean, based on the number of markers used for selection.

Summary

The development of molecular markers for economically significant traits should greatly improve the speed and efficiency of all peanut breeding programs. As a result of the PGI, genetic markers for resistance to late leaf spot were validated by field trials and are being integrated into breeding programs. Phenotyping effort have also been carried out for resistance to early and late leaf spot, TSWV, white mold, and preharvest aflatoxin contamination. Phenotyping efforts have also been carried out for yield, grade, pod morphology, root traits, and drought tolerance. These efforts have resulted in the identification of numerous QTL that should be useful in MAS. MAS has been used in the development of several recent peanut cultivars, and more releases from MAS are anticipated.

Component 6: Creating on-line tools for genomic assisted breeding

PeanutBase, which was launched in early summer, 2013 and is now in its 4th year, provides a full-featured genetic and genomic database for peanut researchers and breeders (Figure 14).

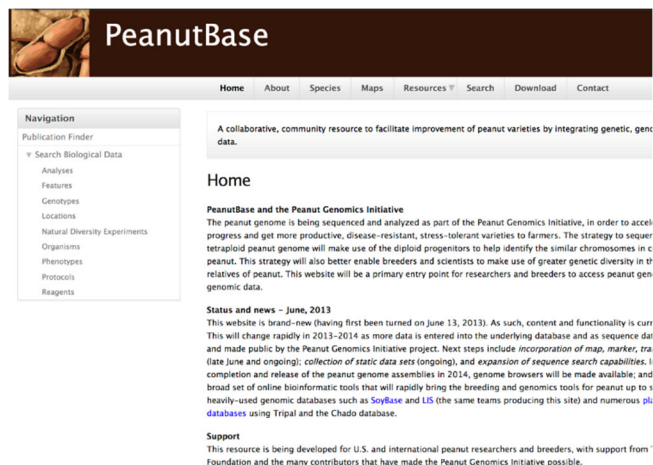
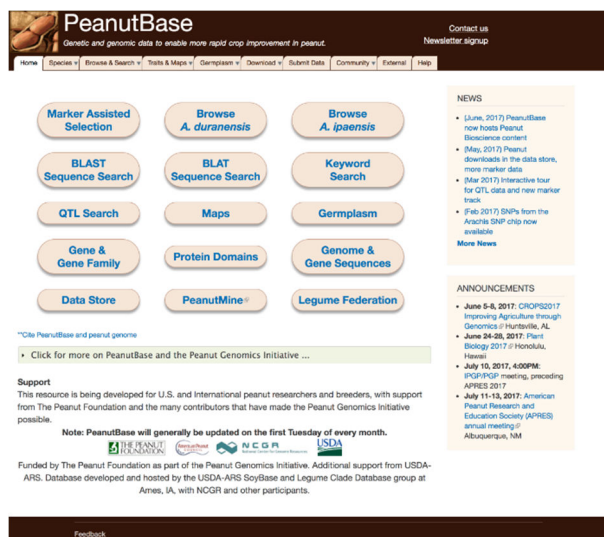


Figure 14. Peanut Base home pages.



The website has several major focal points and services, including:

1. Genome sequences:

- Genome sequences for diploid progenitors available for downloading, for local data mining by researchers.
- In full-featured genome browsers for both wild progenitor genomes, including extensive information about features on the genomes: gene locations, names, expression patterns; genetic markers and positions; alignments with other legume species; link-outs to other resources.
- Accessible for powerful sequence search tools or through keyword searches.
- The cultivated tetraploid peanut genome assembly is nearing completion in Q3 2017, and PeanutBase is primed to provide downloads, browsers, search tools, and aligned genome features as soon as it is released.

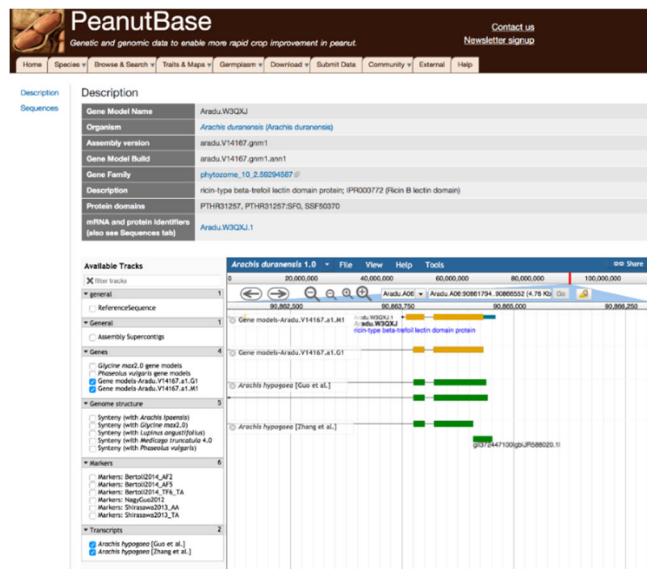


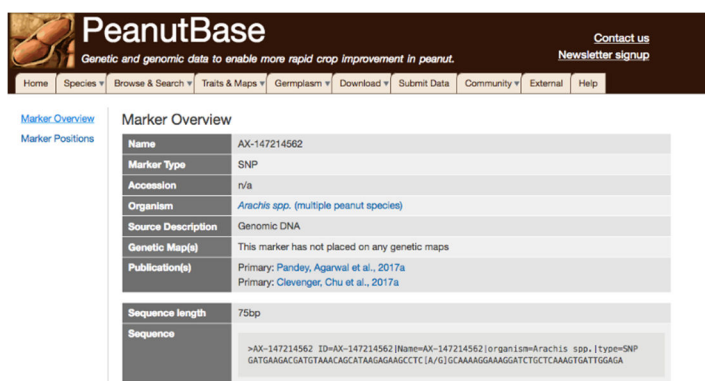
Figure 15. Gene record page for *A. duranensis* gene, showing syntenic soybean gene and transcript data for *A. hypogaea*.

3. Gene information:

- Predicted genes for the diploid progenitors of cultivated peanut, including sequences, and functional descriptions.
- Expression data for each gene, in three collections: 1) developmental, for 22 tissues and developmental stages; 2) under drought conditions; and 3) under nematode response.
- Tools for gene expression analysis: 1) heat-map displays of expression intensity in the genome browsers; 2) "eFP browser" showing expression intensity on an image of the peanut plant and developmental stages; 3) sortable, searchable lists and tables of differential gene expression for all genes in the three large expression collections

4. Genetic markers and maps:

- 25 genetic maps, interactively viewable in the CMap viewer and also all projected onto the diploid genome assemblies.
- 43,000 genetic markers stored in a database and searchable by name (with wildcards) and by publication, and displaying full information about the marker: position, sequence primers, publication, map(s), etc. (Figures 16, 17 and 18).



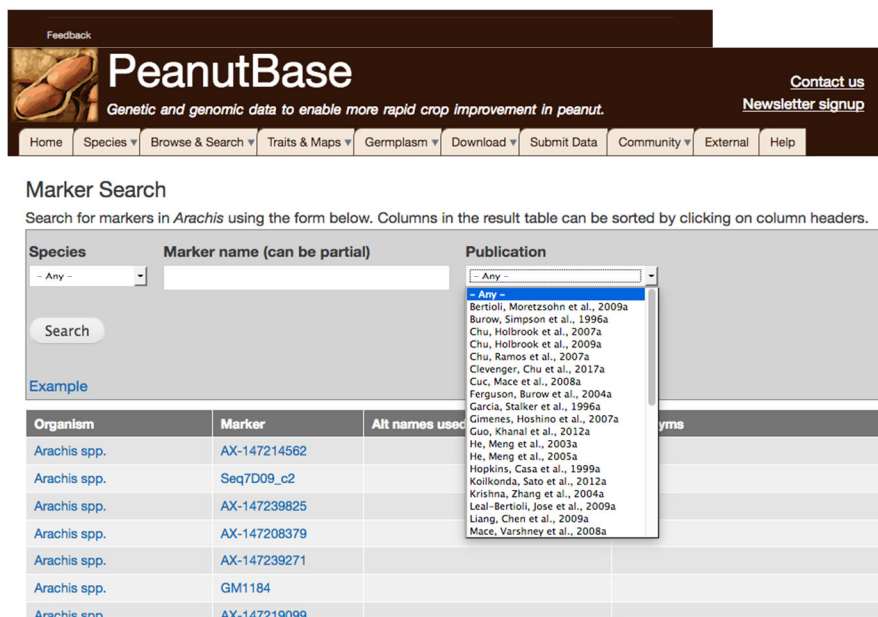
PeanutBase
Genetic and genomic data to enable more rapid crop improvement in peanut.

Home Species Browse & Search Traits & Maps Germplasm Download Submit Data Community External Help

Marker Overview

Name	AX-147214562
Marker Type	SNP
Accession	n/a
Organism	<i>Arachis</i> spp. (multiple peanut species)
Source Description	Genomic DNA
Genetic Map(s)	This marker has not placed on any genetic maps
Publication(s)	Primary: Pandey, Agarwal et al., 2017a Primary: Clevenger, Chu et al., 2017a
Sequence length	75bp
Sequence	>AX-147214562 ID=AX-147214562 (Name=AX-147214562) (Organism=Arachis spp.) (Type=SNP) GATGAAGCAGTGAACAGCATAAGAGAGCCTC(A/G)GCAAAAGGAAGGATCTGCTCAAGTGATTGGAGA

Figure 16. Genetic marker page of PeanutBase.



PeanutBase
Genetic and genomic data to enable more rapid crop improvement in peanut.

Home Species Browse & Search Traits & Maps Germplasm Download Submit Data Community External Help

Marker Search

Search for markers in *Arachis* using the form below. Columns in the result table can be sorted by clicking on column headers.

Species:

Marker name (can be partial):

Publication:

Search

Example

Organism	Marker	Alt names used	Publication
Arachis spp.	AX-147214562		
Arachis spp.	Seq7D09_c2		
Arachis spp.	AX-147239825		
Arachis spp.	AX-147208379		
Arachis spp.	AX-147239271		
Arachis spp.	GM1184		
Arachis spp.	AX-147919099		

Figure 17. Example of a marker search page in PeanutBase.

Marker Overview					
Marker Positions					
Name	AX-147214562				
Physical position(s)	Assembly version	Chromosome	Start	End	View Position
	aradu.V14167.gnm1	Aradu.A02	81898199	81898199	[GBrowse]
Map Position(s)	This marker is not placed on any map.				

Figure 18. Example of a marker position page.

 PeanutBase Genetic and genomic data to enable more rapid crop improvement in peanut.					
Home Species Browse & Search Traits & Maps Germplasm Download Submit Data Community External Help					Contact us Newsletter signup
Overview Map Features					
Map Name	AA_A.duranensis_x_A.stenosperma_d				
Publication Map Name	A. duranensis K7988_x_A. stenosperra V10309 F5 map				
Map Units	cM				
Map view	CMap				
Species	Arachis duranensis, Arachis stenosperra				
Mapping population	AA_A.duranensis_x_A.stenosperma_d				
Parent 1	A. duranensis K7988				
Parent 2	A. stenosperra V10309				
Population Size					
Population Type	F5				
Methods					
Description	The mapping population comprised F5 RILs produced by a cross of A. duranensis K7988 with A. stenosperra V10309. This map, on which the authors identified QTLs for nematode resistance, was constructed using marker genotyping data originating in Shirasawa, Bertoli et al. (2013a) and Bertoli, Ozias-Akins et al. (2014a) as well as with unpublished information. The total length of the genetic map was estimated to be 1036 cM and it contained a total of 872 SNP and SSR mapped loci.				
Publication	Leal-Bertoli, Moretzsohn et al., 2015: Genetic Mapping of Resistance to Meloidogyne arenaria in Arachis stenosperra: A New Source of Nematode Resistance for Peanut.				
Comment					

Figure 19. Example of map features of an interspecific cross.

5. Germplasm (peanut variety) information:

- Searchable mirror of GRIN-Global peanut germplasm and trait data, which includes trait data that is not available at GRIN-Global. Images of accessions are also attached to these records, which will include images that are not yet available at GRIN-Global.
- Images of the peanut accessions in the USDA-ARS collection, with thumbnail images of various collections (mini-core, other wild *Arachis* species, etc.) set up for browsing (Figure 20).
- A powerful interactive geographic map viewer showing the locations of all *Arachis* species and landrace collections in GRIN-Global (Figure 21).

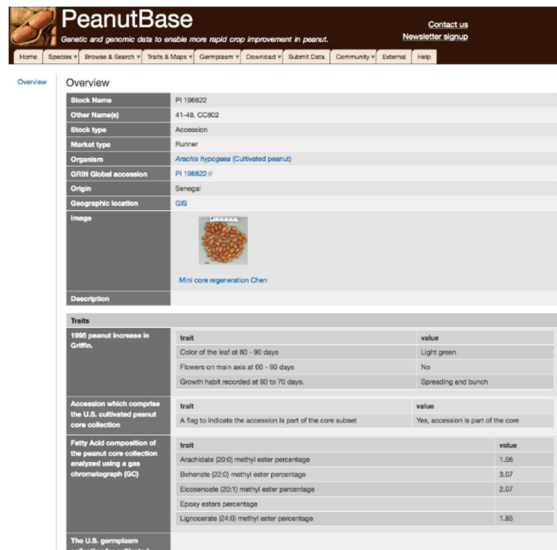


Figure 20. Example of species data for genetic resources.

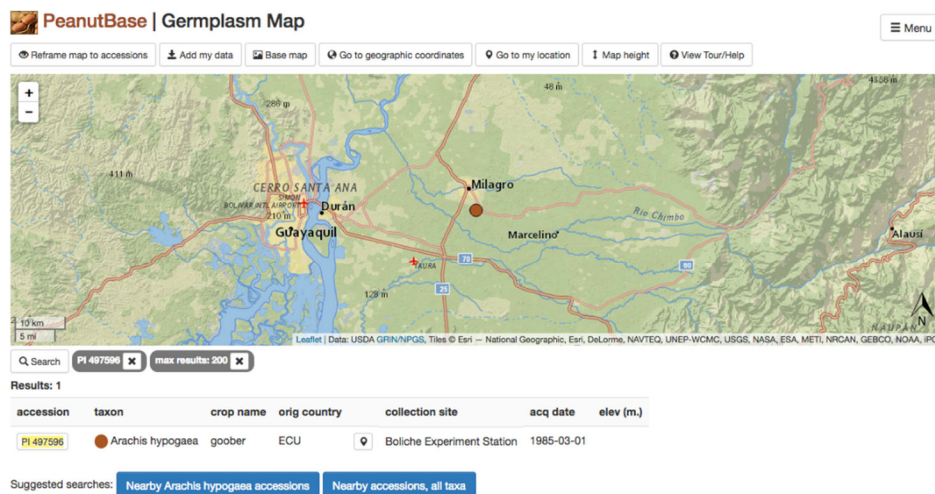


Figure 21. Interactive map of accession location.

6. Detailed information about high-value mapped traits:

- Mapped trait locations (quantitative trait location; QTL) for 232 QTLs for 34 traits, curated from peanut genetic literature, and available for searching by keyword or publication, or for browsing through interactive genetic maps.
- Marker Assisted Selection (MAS) pages: specialized summaries of particular high-value traits, associated markers, and assay methods (for late leaf spot, root knot nematode, rust, high-oleic) (Figure 22).

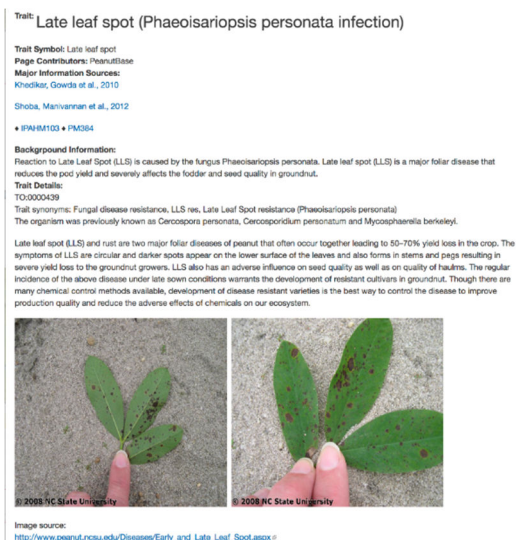


Figure 22. Detailed information about leaf spot in peanut.

7. Gene families, providing gene function information for peanut and other legume species:

- Each peanut gene has been placed in a gene family with genes from 10 other legume species (soybean, common bean, chickpea, etc.), showing evolutionary relationships, functional descriptions, physical locations in the respective genomes, and links from each gene to other information resources (Figure 23).

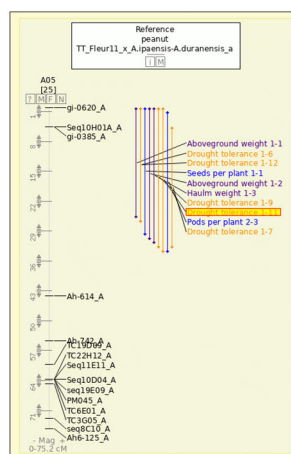


Figure 23. Detailed information about gene locations on peanut molecular map.

8. Guides and help for the features above:

- A Help page, describing each tool and search feature.
- Website tours, guiding users through various analyses on the site. The tours, available both through the Help page and from selected pages (germplasm, gene families, QTL), guides users step by step through the website's features and through typical analysis tasks.

9. Community

- Meeting announcements, minutes, and reports for the IPGI.

- Information and archives for the AAGB conference.
- News of upcoming events of interest to the peanut community.

10. A sophisticated query interface, PeanutMine, for custom analyses and queries:

- PeanutMine is a website interface that uses the InterMine tool to allow users to construct complex queries across various datasets - for example, to find the genes within a QTL region and with increased expression under a particular condition (e.g. drought); or to find genes for which the upstream regions have particular regulatory motifs.
- PeanutMine provides an alternative interface to data held by PeanutBase and is updated whenever a significant dataset has been added.

11. The Data Store

A location for major data sets, with standardized methods for describing the data:
<https://peanutbase.org/data/public/>

Collaborations

PeanutBase benefits considerably through collaborations and in-kind contributions. Major collaborators include:

- The TriPal development team at Washington State University (Dorrie Main and Stephen Ficklin); the PeanutBase website is constructed from TriPal software components.
- The Legume Information System (LIS) developed by the National Center for Genomic Resources (NCGR) at Sante Fe; LIS developers contribute software components.
- SoyBase, at USDA database developed on the Iowa State Campus, providing mainly guidance for collection and display of Quantitative Trait Loci (QTL) data.
- AgBioData, a group of US agricultural databases and resources that meets regularly to identify and solve common problems.

Major in-kind contributions come from the USDA (literature curation, genomic data analyses, genome assembly); NCGR (genome annotation); and the Legume Federation, an NSF-funded project (outreach, software components, and PeanutMine)

PeanutBase is also integrated with LegumeInfo.org (LIS), linking peanut genes to related genes in other species through gene families and genomic synteny.

Website Usage

Overall usage has been fairly stable the past two years. In the past 12 months (7/31/16-7/31/17), PeanutBase had 7657 unique visitors, of which 58% were returning users. On average, visitors looked at 10 pages and stayed on the site for 9 minutes. These last three items mean that people are using PeanutBase, not just glancing at it and moving on.

In the past year PeanutBase has had visitors from 142 countries. Chrome remains by far the most common web browser (70%), and Windows is the most common operating system. There are some visitors using mobile devices, ~1600 sessions vs ~17,000 sessions via desktop browsers.

Google analytics overview for 7/31/16 – 7/31/17

Where do we go from here?

The Peanut genomics project has sequenced the genomes of the cultivated peanut and two of its progenitor species. While providing tremendous advances to the community, we have found that peanut is even more complicated than expected due to a process called tetrasomic inheritance (i.e., the A and B genomes do not differentiate because the 4 alleles are either AAAA or BBBB). This affects only a small portion of the genome and new methodologies must be developed to sort these areas out as we move forward to marker-assisted breeding and transferring new traits from wild relatives. The cultivar sequenced is a powerful model, but all *A. hypogaea* lines and cultivars will be slightly different, this is the genetic variation underlying breeding. In order to more fully leverage genetic variation to accelerate peanut improvement, a breeder will need to know what is variable in their populations. DNA sequencing costs have tremendously decreased, but high levels of resolution may not be needed for breeding, and may be cost-prohibitive. The question is what basic knowledge and data are worth pursuing that will be of benefit for both scientific advances and for solving problems facing the peanut industry? The following paragraphs outline several ideas and directions that need to be explored.

Some objectives can be more easily solved using genomic tools than others. For example, when one or a few genetic markers are associated with traits, selection is rather straight forward. However, research in other crops have indicated that traits such as yield are highly complex and many genetic markers, sometimes hundreds, are associated with productivity each of which have very small effects. Breeding for highly complex traits for which phenotyping data is not highly precise is a long-range endeavor which may or may not bring economic value to the industry. Table 5 illustrates the industry-wide costs of several diseases

Table 5. Estimated savings and revenue per year if all available genomic tools are used based on estimates compiled during 2011. (Information supplied by Marshall Lamb, Dr. Tim Brennenman, Dr. Bob Kemerait, Dr. David Jordan, Dr. John Damicone, Dr. Barbara Shew, Dr. Jason Woodward).

Trait	Savings	Increased Revenue	Total ¹
Drought Tolerance ²	\$8.87 million	\$60.32 million	\$69.19 million
Early Leafspot	\$31.75 million		\$31.75 million
Late Leafspot	\$21.41 million		\$21.31 million
TSWV ³		\$24.99 million	\$24.99 million
White mold (Rhizoctonia)	\$45.15 million		\$45.15 million
Sclerotinia	\$2.18 million	\$6.75 million	\$8.93 million
Nematode	\$12.24 million	\$43.62 million	\$55.86 million
Subtotal	\$121.50 million	\$135.68 million	\$257.18 million
<i>A. flavus/A. paraciticus</i>	\$25.00 million		\$25.00 million
Increase folate & high Oleic		\$100.00 million ⁴	\$100.00 million
Subtotal	\$25.00 million	\$100.00 million	\$125.00 million
TOTAL	\$146.50 million	\$235.68 million	\$257.18 million

¹ Savings and revenues will vary year to year based on acreage and weather conditions. This chart is based on 1.34 million acres.

² Savings estimate based on 36,086,044,968 gal/year in water use efficiency; revenue estimate based on improved drought tolerance.

³ Reflects losses only to TSWV. No estimates were included for expenses related to condensed planting window and potential losses due to later harvest.

⁴ Revenue for each new health claim per JM Smuckers marketing groups.

Table 6 summarizes the thoughts of industry needs as perceived by a group of producers, shellers, manufacturers, and researchers who attended the 2017 APRES meeting. The survey included all the most economically important items in Table 5 as well as other items.

Table 6. Research Needs Brainstorming Session conducted at the Seed Summit, American Peanut Research and Education Society, Albuquerque, NM on July 11, 2017.

Needs of the Peanut Industry (grouped by theme)

Germplasm Support

- Free Germplasm Exchange; Open Availability of Genetic Resources
- Support for Development of Inter-Specific Populations; Introgression of Useful Traits
- More Structured Populations
- Genotype and Phenotype Core Collection
- Re-sequence Entire Germplasm Collection
- Re-sequence drought susceptible and tolerant groups of genotypes

Genomic and Genetic Information Management

- Establish Committee to Review Potential Genome Assembly Releases
- Improved Genomic Techniques (QTL Sequence)
- MAS for Gene Stacking
- Informatics/Information Systems; Maintain/Update PeanutBase

Yield/Grade Improvement

- Yield Limiting Stresses; Yield per Gallon of Water/Drought Tolerance
- Yield and Grade/Economic Return
- Sustainability

Biotic and Abiotic Stress Mitigation

- Leafspot Research
- Nematode Research
- Peanut Smut
- Develop Markers for Key Diseases - TSWV, Sclerotinia, White Mold, CBR

Traits Important to Manufacturers/Consumers

- Flavor
- High O/L; Continue Some Research on Normal Oleic Peanuts
- Uniform Kernel Size Distribution
- Aflatoxin Research
- Maintain/Enhance Health Benefits

Genetic Techniques Development

- Conduct Seminar at APRES on DNA Extraction, genotyping, MAS

General

- Sustain Long-Term Research
-

Where should the genomics project go from here during the next five years? The vision of the genomics project has been to develop genomic resources that will aid in solving industry problems that have been difficult to solve using conventional technologies. If any of the disease or production problems in Table 5 could be eliminated, there would be significant value added to the peanut industry. While drought resistance is at the top of the list, the trait is very tedious to evaluate and mechanisms of drought resistance are not well understood. The following is a brief outline of priorities that should add significant value to the producer and industry if resources are available to

develop multi-disciplinary teams to intensively investigate the areas.

Many of the traits or items listed in Table 6 cannot be solved by genomic research, such as the establishment of committees, while some others are very complex and/or difficult to phenotype traits such as aflatoxin, resistances, drought tolerance and improved yield. As well, sequencing the 9,000+ cultivated lines and all 500+ wild species may not be currently economically feasible given the real economic problems in the industry. However, a select group of lines should be sequenced and characterized as we know that both the cultivated and wild species contain genes that would add value to the peanut crop.

Germplasm Support

1. ***Germplasm base:*** There are numerous cultivated and wild species lines identified by investigators in the U.S. and internationally. These lines need to be accumulated into one collection and tested for the traits of interest (for diseases and insects, testing needs to be done in all three production regions because there is good evidence that there are different strains of virus and diseases in different regions). At least 3 collections need to be maintained:
 - a. Disease collection
 - b. Insect collection
 - c. Quality and drought-tolerant collection

These collections then need to be fully characterized by sequencing and identifying the genes that confirm the traits of interest, including confirming the genetics of different genetic sources of resistances. The materials should be maintained as separate collections at the USDA germplasm facilities in Griffin, GA.

2. ***Utilization of wild species:*** Develop a set of introgression lines with genes from wild species, known to have economic importance, introgressed into a cultivated background. These lines should be advanced to a generation where they are fertile enough to be of use in breeding programs without encountering sterility problems in crosses with advanced breeding lines and, at the same time, the traits of interest should be characterized to such an extent that breeders can utilize the materials. Selected lines need to be characterized on a genomics level and molecular markers identified for traits of agronomic importance, especially disease resistances.
3. ***Pyramiding traits:*** Markers that are associated with individual traits can be used to select for single breeding lines containing multiple traits. Trait phenotypes must be confirmed after marker selection. These breeding lines can be used to make subsequent crosses to select desired genotypes.

Genomic and Genetic Information Management

1. ***Marker development:*** A robust, genome-wide SNP array, an output of the peanut genomics project, has been developed and can be used by breeders to identify markers linked to traits of interest. Additional sequencing data needs to be applied to develop additional markers for rare alleles from both cultivated and wild species. Molecular associations with traits of agronomic importance, white mold, early leaf spot, late leaf spot, etc., should be confirmed and specific markers for individual genes designed.
2. ***PeanutBase*** must continue to be supported as a resource of information to the peanut

community for genomic, germplasm, and trait information.

3. Establish community guidelines for long-term storage of data, spearheaded by PeanutBase.

Favorable Traits

1. **Leaf spot resistance:** Early and Late Leaf Spot are caused by two different fungi and resistance may be conditioned by common genetic systems, but there is good evidence that genes are present in different populations that suppress one or the other pathogen. In regions where “resistant” cultivars are produced, there has been an increase in the other pathogen, so resistance to both diseases need to be combined to solve the leaf spot problem. Several genes have been identified in lines that have *A. cardenasii* in the pedigree, but there is also good evidence that different genes controlling resistance are also found in the *A. hypogaea* collection. Resistance is multigenic and complex, but techniques are available to give good phenotyping data to separate the components of resistance (infection, disease spread, pathogen reproduction, defoliation). Because of the great economic impact for solving the leaf spot problem in peanut, the following research is needed:
 - a. Cultivated lines with resistance to early leaf spot or to late leaf spot need to be hybridized with susceptible genotypes and the genetics of resistance investigated. This will involve field and genomic studies to identify genes conferring resistance common and unique among the lines.
 - b. Components of resistance need to be understood on a genomic level for each pathogen.
 - c. Populations of introgression lines from *Arachis* species need to be characterized genetically to identify genes giving resistance to the leaf spots and molecular markers associated with resistance genes for each pathogen.
 - d. Lines need to be developed to combine resistance to early and to late leaf spot from different genetic sources for plant breeding.
2. **Nematodes and high oleic acid:** The marker systems already available for high oleic acid and nematode resistance have a proven track record and need to be applied in more (all) cultivar development programs. New sources of nematode resistance need to be identified and characterized to ensure durability of resistance.
3. **TSWV:** Because TSWV cannot be controlled by chemicals and only moderately by production practices, genetic solutions are necessary to suppress the disease. The resistance found in cultivated peanut and in introgression populations with highly resistant *Arachis* species need to be fully characterized on a genomic level. Multiple sources of resistance need to be combined to assure long-term and complete suppression of the disease.

Genetic Techniques Development

1. **Establish a central molecular lab to work with peanut breeders:** Plant breeders are highly successful at developing populations and phenotyping plants and lines. However, utilizing molecular markers and genomics is on the periphery of their skill set and establishing a laboratory with a faculty-level scientist who can work with and help guide cultivar development programs would greatly facilitate utilization of genomic data. Other crops (e.g., wheat) have established central laboratories to integrate plant breeding and molecular technologies and their efforts have proven highly successful. This is something the peanut community may want to consider.
2. **Marker confirmation:** Progress has been made in discovery of molecular markers associated with disease resistance, quality and other agronomic traits. These associations must be confirmed through genetic studies and more precise markers developed that can be used routinely used by the plant breeder.

3. **Genome-wide-selection:** By utilizing genomic and marker data it should be possible to make predictions of the “best” parents to use in hybridization programs. Genome-wide-selection (GWS) is used routinely in other polyploid crops such as wheat, but in peanut, populations need to be analyzed to develop prediction models and then to validate the models. If applicable to peanut, many generations and significant cost savings should be possible for peanut breeders.
4. **Training for breeders:** These tools and resources are not useful if breeders don’t know how to use them. Training would also provide a means of improving communication between breeders’ needs and molecular resources that are being developed.

Appendices

Appendix 1. Examples of PGI funding leading to additional grant and project support

The PGI has produced the foundation for leveraging funding from federal programs such as USDA-NIFA, NSF, BARD, and USAID. For example, Peggy Ozias-Akins has led the following projects that co-funded the generation of expressed sequence data contributing to Component 3 and genome sequence contributing to Component 2. A few examples include the following:

- USDA-NIFA, Co-PIs Scott Jackson, George Vellidis, Brian Scheffler, Corley Holbrook. Improving Peanut Seed Quality by Molecular Definition of Stress Thresholds Permissive for Aflatoxin Contamination.
- NSF, CoPIs Soraya Bertoli, Scott Jackson, Tom Stalker, Daniel Fonceka (CERAAS, Senegal). BREAD ABRDC: Genomic approaches to capture novel alleles in cultivated peanut to increase smallholder production.
- BARD, Israeli PI Ran Hovav, US co-PI Scott Jackson. The genetics of pod-filling in peanut under water limiting conditions.
- USAID, Peanut and Mycotoxin Innovation Lab, co-PIs Scott Jackson, David Bertoli, Corley Holbrook, Rajeev Varshney (ICRISAT), Issa Faye (ISRA Senegal), Daniel Fonceka (CERAAS, Senegal). Translational Genomics to Reduce Pre-harvest Aflatoxin Contamination of Peanut.
- The USDA-NIFA project helped to support two graduate students, Josh Clevenger and Walid Korani, who have made significant contributions to both Components 2&3, specifically in the generation and analysis of genome sequence for SNP marker discovery and expressed sequence from aflatoxin contaminated seeds of multiple genotypes. One aflatoxin gene expression study published by Clevenger et al. in 2016 suggested that a repressor of ABA (a plant growth regulator) signaling is a potential susceptibility factor. The USAID project facilitated analysis in the US of genotypes reported by ICRISAT to have reduced aflatoxin contamination. Positive results led to the generation of gene expression data from these genotypes with funding from other projects including the PGP.
- The BARD project has supported one graduate student, Carolina Chavarro, who has made significant contributions to both Components 2&3, specifically in the generation and analysis of genome sequence for SNP marker discovery and expressed sequence from seeds of genotypes that differ in their pod filling characteristics. These data will help to identify genes underlying pod fill traits of importance to production and quality. The collaboration has resulted in one publication (Gupta et al. 2016) with another in preparation.

Appendix 2. Meetings sponsored by the PGI

Advances in Peanut Genomics and Biotechnology conferences

October 2-8, 2012	Hyderabad, India
June 17-21, 2013	Zhengzhou, China
November 11-14, 2014	Savannah, GA
November 5-7, 2015	Brisbane, Australia
March 14-17, 2017	Cordoba, Argentina

CROPS Conference – HudsonAlpha, Huntsville, AL

June, 2016

June, 2017

InterDrought Conference – Hyderabad, India

November, 2016

Appendix 3. Examples of key traits that reside in a given wild *Arachis* species.

A- genome species

<i>A. cardenasii</i>	Aflatoxin, <i>Cylindrocladium parasiticum</i> , Early leaf spot Groundnut rosette virus, Late leaf spot, Peanut Bud Necrosis Virus, Peanut Mottle Virus (PMV), Peanut rust (<i>Puccinia arachidis</i>), Peanut Stripe Virus (PStV), Tomato-Spotted Wilt Virus (TSVW), Peanut Root-Knot Nematode (<i>Meloidogyne arenaria</i>), Northern Root-Knot Nematode (<i>Meloidogyne hapla</i>) Armyworm (<i>Spodoptera</i> spp.), Corn Earworm (<i>Helicoverpa armigera</i>), Leafminer (<i>Aproaerema modicella</i>), Leafhoppers (<i>Empoasca fabae</i>), Southern Corn Rootworm (<i>Diabrotica undecimpunctata howardi</i>)
<i>A. duranensis</i>	Aflatoxin, Late leaf spot, Peanut Stunt Virus Armyworm (<i>Spodoptera</i> spp.), Corn Earworm (<i>Helothis zea</i>), Corn Earworm (<i>Helicoverpa armigera</i>), Groundnut aphid (<i>Aphis craccivora</i>), Leafminer (<i>Aproaerema modicella</i>), Leafhoppers (<i>Empoasca fabae</i>), Nematodes (<i>Meloidogyne arenaria</i>), Thrips (<i>Frankliniella fusca</i>), Thrips (<i>Frankliniella schultzei</i>), Chilli Thrips (<i>Scirtothrips dorsalis</i>)
<i>A. correntina</i>	<i>Cylindrocladium parasiticum</i> , Late leaf spot, Peanut Mottle Virus (PMV), Peanut rust (<i>Puccinia arachidis</i>), Tomato-Spotted Wilt Virus (TSVW) Armyworm (<i>Spodoptera</i> spp.), Corn Earworm (<i>Helothis zea</i>), Groundnut aphid (<i>Aphis craccivora</i>), Leafhoppers (<i>Empoasca fabae</i>), Lesser Cornstalk Borer (<i>Elasmopalpus lignosellus</i>), Spider Mites (<i>Tetranychus urticae</i>), Thrips (<i>Frankliniella fusca</i>)
<i>A. diogeni</i>	Early leaf spot Groundnut Rosette Disease, Peanut Bud Necrosis Virus, Peanut Mottle Virus (PMV), Peanut Ringspot Virus, Peanut rust (<i>Puccinia arachidis</i>), Peanut Stripe Virus (PStV), Tomato-Spotted Wilt Virus (TSVW), Peanut Root-Knot Nematode (<i>Meloidogyne arenaria</i>), Northern Root-Knot Nematode (<i>Meloidogyne hapla</i>) Armyworm (<i>Spodoptera</i> spp.), Corn Earworm (<i>Helothis zea</i>), Groundnut aphid (<i>Aphis craccivora</i>), Leafhoppers (<i>Empoasca fabae</i>), Thrips (<i>Frankliniella fusca</i>), Thrips (<i>Frankliniella schultzei</i>), Chilli Thrips (<i>Scirtothrips dorsalis</i>)
<i>A. stenosperma</i>	Early leaf spot Groundnut Rosette Disease, Late leaf spot, Peanut rust (<i>Puccinia arachidis</i>), Tomato-Spotted Wilt Virus (TSVW), Nematodes (<i>Meloidogyne arenaria</i>), Nematodes (<i>Meloidogyne hapla</i>), Nematodes (<i>Meloidogyne javanica</i>)

Corn Earworm (*Helothis zea*), Corn Earworm (*Helicoverpa armigera*), Groundnut aphid (*Aphis craccivora*), Leafminer (*Aproaerema modicella*), Leafhoppers (*Empoasca fabae*), Thrips (*Enneothrips flavens*), Thrips (*Frankliniella fusca*)

B genome species

<i>A. magna</i>	Early leaf spot, Late leaf spot, Sclerotinia Blight (<i>Sclerotinia minor</i>)
<i>A. ipaënsis</i>	Late leaf spot, Peanut rust (<i>Puccinia arachidis</i>) Armyworm (<i>Spodoptera</i> spp.), Corn Earworm (<i>Helicoverpa armigera</i>), Leafhoppers (<i>Empoasca fabae</i>)

K genome species (previously named a B genome)

<i>A. batizocoi</i>	Early leaf spot, Late leaf spot, Peanut rust (<i>Puccinia arachidis</i>), Tomato-Spotted Wilt Virus (TSVW) Corn Earworm (<i>Helothis zea</i>), Groundnut aphid (<i>Aphis craccivora</i>), Leafhoppers (<i>Empoasca fabae</i>), Nematodes (<i>Meloidogyne arenaria</i>), Thrips (<i>Frankliniella fusca</i>)
---------------------	--

Appendix 4. Research publications by PGI members and associates since 2012 that are relevant to the PGP

Component 1: Sequencing and assembly of the peanut genome

Component 4: Evaluating new DNA sequencing methods

1. He, G., N.A. Barkley, Y. Zhao, M. Yuan, and C.S. Prakash. 2014. Phylogenetic relationships of species of genus *Arachis* based on genic sequences. *Genome* 57(6): 327-334.
2. Chopra, R., G.B. Burow, A. Farmer, J. Mudge, C.E. Simpson, T.A. Wilkins, M.R. Baring, N. Puppala, K.D. Chamberlin, and M.D. Burow. 2015. Next-generation transcriptome sequencing, SNP discovery, and SNP validation in four market classes of peanut (*Arachis hypogaea* L.). *Mol. Genet. and Gen.* 290(3):1169- 1180.
3. Guo, Y., B. Abernathy, Y. Zeng, and P. Ozias-Akins. 2015. TILLING by sequencing to identify induced mutations in stress resistance genes of peanut (*Arachis hypogaea*). *BMC Genomics* 16:1348.
4. Hickey, J.M., G. Gorjanc, R.K. Varshney, and C. Nettelblad. 2015. Imputation of single nucleotide polymorphism genotypes in biparental, backcross, and topcross populations with a hidden markov model. *Crop Sci.* 55 (5):1934-1946.
5. Bertoli D., S.B. Cannon, L. Froenicke, G. Huang, A.D. Farmer, E.K.S. Cannon, X. Liu, D. Gao, J. Clevenger, S. Dash, L. Ren, M.C. Moretzsohn, K. Shirasawa, W. Huang, B. Vidigal, B. Abernathy, Y. Chu, C.E. Niederhuth, P. Umale, A.C.G. Araujo, A. Kozik, K.D. Kim, M.D. Burrow, R.K. Varshney, X. Wang, W. Zhang, N. Barkley, P.M. Guimaraes, S. Isobe, B. Guo, B. Liao, H.T. Stalker, R.J. Schmitz, B.E. Scheffler, S.C.M. Leal-Bertoli, X. Xun, S.A. Jackson, R. Michelmore, and P. Ozias-Akins. 2016. The genome sequences of *Arachis duranensis* and *Arachis ipaënsis*, the diploid ancestors of cultivated peanut. *Nature Genet.* 48: 438-446.
6. Chen, X., Q. Yang, H. Li, H. Li, Y. Hong, L. Pan, N. Chen, F. Zhu, X. Chi, W. Zhu, M. Chen, H. Liu, Z.

- Yang, E. Zhang, T. Wang, N. Zhong, M. Wang, H. Liu, S. Wen, X. Li, G. Zhou, S. Li, H. Wu, R.K. Varshney, X. Liang, and S. Yu. 2016. Transcriptome-wide sequencing provides insights into geocarpy in peanut (*Arachis hypogaea* L.). *Plant Biotech. J.* 14:1215-1224.
7. Clevenger, J., K. Marasigan, V. Liakos, V. Sobolev, G. Vellidis, C. Holbrook, and P. Ozias-Akins. 2016. RNA sequencing of contaminated seeds reveals the state of the seed permissive for pre-harvest aflatoxin contamination and points to a potential susceptibility factor. *Toxins*. 8:317.
 8. Wendel, J.F., S.A. Jackson, B.C. Meyers, and R.A. Wing. 2016. Evolution of plant genome architecture. *Genome Bio.* 17:37.
 9. Padney, M.K., A.W. Khan, V.K. Singh, M.K. Vishwakarma, Y. Shasidhar, V. Kumar, V. Garg, R.S. Bhat, A. Chikitikineni, P. Janila, B. Guo, and R.K. Varshney. 2017. QTL- seq approach identified genomic regions and diagnostic markers for rust and late leaf-spot resistance in groundnut (*Arachis hypogaea* L.). *Plant Biotech. J.* 15:927-941.
 10. Peng, Z., W. Fan, L. Wang, D. Paudel, D. Leventini, B.L. Tillman, J. Wang. 2017. Target enrichment sequencing in cultivated peanut (*Arachis hypogaea* L.) using probes designed from transcript sequences. *Mol. Genet. and Gen.* pp. 1-11.
 11. Varshney, R.K., R. Terauchi, S.R. McCouch. 2014. Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. *PLoS Biol.* 12(6): e1001883.
 12. Wang, H., R.V. Penmetsa, M. Yuan, L. Gong, Y. Zhao, B. Guo, A.D. Farmer, B.D. Rosen, J. Gao, S. Isobe, D.J. Bertioli, R.K. Varshney, D.R. Cook, and G. He. 2012. Development and characterization of BAC-end sequence derived SSRs, and their incorporation into a new higher density genetic map for cultivated peanut (*Arachis hypogaea* L.). *BMC Plant Biol.* 12:10.

Component 2: Discovery of useful DNA markers

13. Grabile, M., L. Chalup, G. Robledo, and G. Seijo. 2012. Genetic and geographic origin of domesticated peanut as evidenced by 5S rDNA and chloroplast DNA sequences. *Plant Systematic and Evol.* 298:1151-1165.
14. Brasileiro, A.C.M., A.C.G. Araujo, S.C.M. Leal-Bertioli, and P.M. Guimarães. 2014. Genomics and genetic transformation in *Arachis*. *International J. Plant Bio. & Res.* 2):1017-1028.
13. Brasileiro, A.C.M., N. Carrasquilla-Garcia, R.V. Penmetsa, D. Cook, M.C. Morezsohn, and D.J. Bertioli. 2015. *Arachis batizocoi*: a study of its relationship to cultivated peanut (*A. hypogaea* L.) and its potential for introgression of wild genes into the peanut crop using induced allotetraploids. *Annals of Botany* 115:237-249.
14. Conner, J.A., M. Muruganantham, H. Huo, K. Chae, and P. Ozias-Akins. 2015. A parthenogenesis gene of apomict origin elicits embryo formation from unfertilized eggs in a sexual plant. *Proc. Natl. Acad. Sci. USA* 112:11205-11210.
15. Gao, D., N. Jiang, R. Wing, J. Jiang, and S.A. Jackson. 2015. Transposons play an important role in evolution and diversification of centromeres among closely related species. *Front Plant Sci.* 6:216.
16. Gao, D., Y. Li, B. Abernathy, and S.A. Jackson. 2015. Landscape and evolutionary dynamics of terminal-repeat retrotransposons in miniature (TRIMs) in 48 whole plant genomes. *Genome Bio.* 17:7.
17. Bertioli, D.S., C.M. Leal-Bertioli, and H.T. Stalker. 2016. pp. 147-161 In: H.T. Stalker and R.F.

- Wilson (eds.). The peanut genome: the history of the consortium and the structure of the genome of cultivated peanut and its diploid ancestors. Peanuts: Genetics, Processing and Utilization. Academic Press and AOCS Press, Waltham, MA.
15. Janila, P., M.K. Pandey, Y. Shasidhar, T.M. Variath, M. Sriswathi, P. Khera, S.S. Manohar, P. Nagesh, M.K. Vishwakarma, G.P. Mishra, T. Radhakrishnan, N. Manivannan, K.L. Dobariya, R.P. Vasanthi, and R.K. Varshney. 2016. Molecular breeding for introgression of fatty acid desaturase mutant alleles (ahFAD2A and ahFAD2B) enhances oil quality in high and low oil containing peanut genotypes. *Plant Sci.* 242:203-213.
 16. Janila, P., M.T. Variath, M.K. Pandey, H. Desmae, B.N. Motagi, P. Okori, S.S. Manohar, A.L. Rathnakumar, T. Radhakrishnan, B. Liao, and R.K. Varshney. 2016. Genomic tools in groundnut breeding program: status and perspectives. *Plant Sci.* 7:289.
 17. Katam, R., V. Gottschalk, P. Survajahala, G. Brewster, P. Paxton, K. Sakata, C. Lee, C. Williams, B. Guo, and L.M. Latinwo. 2014. Advances in proteomics research for peanut genetics and breeding. In: Mallikarjuna, N. and Varshney, R.K., (eds). *Genetics, Genomics and Breeding of Peanuts*. Boca Raton, FL: CRC Press. pp. 161- 177.
 18. Kole, C.D., M. Muthamilarasan, R. Henry, D. Edwards, R. Sharma, M. Abberton, J. Batley, A. Bentley, M. Blakeney, J. Bryant, H. Cai, M. Cakir, L.J. Cseke, J. Cockram, A.C. de Oliveira, C.D. Pace, H. Dempewolf, S. Ellison, P. Gepts, A. Greenland, A. Hall, K. Hori, S. Hughes, M.W. Humphreys, M. Iorizzo, A.M. Ismail, A. Marshall, S. Mayes, H.T. Nguyen. F.C. Ogonnaya, R. Ortiz, A.H. Paterson, P.W. Simon, J. Tohme, R. Tuberosa, B. Valliyodan, R.K. Varshney, S.D. Wulschleger, M.Yano, and M. Prasad. 2015. Application of genomics-assisted breeding for generation of climate resilient crops: Progress and prospects. *Frontiers in Plant Sci.* 6:563.
 19. Pandey, M.K., G. Agarwal, S.M. Kale, J. Clevenger, S.N. Nayak, M. Sriswathi, A. Chititkineni, C. Chavarro, X. Chen, H.D. Upadhyaya, M.K. Vishwakarma, S. Leal-Bertioli, X. Liang, D.J. Bertioli, B. Guo, S.A. Jackson, P. Ozias-Akins, and R.K. Varshney. 2017. Development and evaluation of a high density genotyping 'Axiom *Arachis*' array with 58 K SNPs for accelerating genetics and breeding in groundnut. *Sci. Rep.* 7-40577.
 18. Ratnaparkhe, M.B., T.-H. Lee, X. Tan, X. Wang, J. Li, C. Kim, L.K. Rainville, C. Lemke, R.O. Compton, J. Robertson, M. Gallo, D.J. Bertioli, and A.H. Paterson. 2014. Comparative and evolutionary analysis of major peanut allergen gene families. *Genome Bio. and Evol.* 6 (9):2468-2488.
 19. Saxena, R.K., D. Edwards, and R.K. Varshney. 2014. Structural variation in plant genomes. *Briefings in Functional Genomics* 13(4):296-307.
 20. Singh, A.K. and S.N. Nigam. 2016. *Arachis* gene pools and genetic improvement in groundnut. *Sustainability Development and Biodiversity* 1:17-75.
 21. Suassuna, T.M.F., N.D. Suassuna, M.C. Moretzsohn, S.C.M. Leal-Bertioli, D.J. Bertioli, and E. Paulo de Medeiros. 2015. Yield, market quality, and leaf spots partial resistance of interspecific peanut progenies. *Crop Breed. and Applied Biotech.* 15:175-180.
 20. Sujay, V., M.V. Gowda, M.K. Pandey, R.S. Bhat, Y.P. Khedikar, H.L. Nadaf, B. Gautami, C. Sarvamangala, S. Lingaraju, T. Radhakrishnan, S.J. Knapp, and R.K. Varshney. 2012. Quantitative trait locus analysis and construction of consensus genetic map for foliar disease resistance based on two recombinant inbred line populations in cultivated groundnut (*Arachis hypogaea* L.). *Mol. Breed.* 30:773-788.
 22. Wang, H., P. Khera, B. Huang, M. Yuan, R. Katam, W. Zhuang, K. Harris-Shultz, K.M. Moore, A.K.

- Culbreath, X. Zhang, R.K. Varshney, L. Xie, and B. Guo. 2015. Analysis of genetic diversity and population structure of peanut cultivars and breeding lines from China, India and the US using SSR markers. *J. Intergr. Plant Biol.* 58 (5):452-465.
23. Yeri, S.B., K. Shirasawa, M.K. Pandey, M.V.C. Gowda, V. Sujay, M. Sriswathi, H.L. Nadaf, B.N. Motagi, S. Lingaraju, A.R.S. Bhat, R.K. Varshney, P.U. Krishnaraj, R.S. Bhat. 2014. Development of NILs from heterogeneous inbred families for validating the rust resistance QTLs in peanut (*Arachis hypogaea* L.). *Plant Breeding* 133: 80-85.

Component 3: Discovery of all genes that control targeted traits

24. Koilkonda, P., S. Sato, S. Tabata, K. Shirasawa, H. Hirakawa, H. Sakai, S. Sasamoto, A. Watanabe, T. Wada, Y. Kishida, H. Tsuruoka, T. Fujishiro, M. Yamada, M. Kohara, S. Suzuki, M. Hasegawa, H. Kiyoshima, and S. Isobe. 2012. Large-scale development of expressed sequence tag-derived simple sequence repeat markers and diversity analysis in *Arachis* spp. *Mol. Breed.* 30:125-138.
25. Moretzsohn, M.C., E.G. Gouvea, P.W. Inglis, S.C.M. Leal-Bertioli, J.F.M. Valls, D.J. Bertioli. 2012. A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. *Annals of Botany* 11:113-126.
26. Pandey, M.K., B. Gautami, T. Jayakumar, M. Sriswathi, H.D. Upadhyaya, M.V.C. Gowda, T. Radhakrishnan, D.J. Bertioli, S.J. Knapp, D.R. Cook, and R.K. Varshney. 2012. Highly informative genic and genomic SSR markers to facilitate molecular breeding in cultivated groundnut (*Arachis hypogaea*). *Plant Breed.* 131:139-147.
27. Shirasawa, K., P. Koilkonda, K. Aoki, H. Hirakawa, S. Tabata, M. Watanabe, M. Hasegawa, H. Kiyoshima, S. Suzuki, C. Kuwata, Y. Naito, T. Kuboyama, A. Nakaya, S. Sasamoto, A. Watanabe, M. Kato, K. Kawashima, Y. Kishida, M. Kohara, A. Kurabayashi, C. Takahashi, H. Tsuruoka, T. Wada, and S. Isobe. 2012. In silico polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. *BMC Plant Biol.* 12:80.
28. Shoba D., N. Manivannan, P. Vindhiyavarman, and S.N. Nigam. 2012. SSR markers associated for late leaf spot disease resistance by bulked segregant analysis in groundnut (*Arachis hypogaea* L.). *Euphytica* 188:265- 272.
29. Wang H., M.K. Pandey, L. Qiao, H. Qin, A.K. Culbreath, G. He, R.K. Varshney, and B. Guo. 2013. Genetic mapping and QTL analysis for disease resistance using F₂ and F₅ generation-based genetic maps derived from Tifrunner × GT-C20 in peanut (*Arachis hypogaea* L.). *Plant Genome* 6: doi: 10.3835/plantgenome2013.05.0018.
30. Bertioli, D.J., P. Ozias-Akins, Y. Chu, K.M. Dantas, S.P. Santos, E. Gouvea, P.M. Guimarães, S.C.M. Leal- Bertioli, S.J. Knapp, and M.C. Moretzsohn. 2014. The use of SNP markers for linkage mapping in diploid and tetraploid peanut. *G3: Genes, Genomes, Genetics* 4:89-96.
31. Chamberlin, K.D. 2014. Characterization of ICRISAT peanut mini-core accessions with regards to a molecular marker associated with resistance to Sclerotinia blight. *Peanut Sci.* 41:42-49.
32. Gajjar N.K., G.P. Mishra, T. Radhakrishnan, S.M. Dodia, A.L. Rathnakumar, N. Kumar, S. Kumar, J.R. Dobaria, and A. Kumar. 2014. Validation of SSR markers linked to the rust and late leaf spot diseases resistance in diverse peanut genotypes. *AJCS.* 8(6):927-936.

33. Jakkeral, S., H.L. Nadaf, M.V.C. Gowda, R.S. Bhat, B.N. Motagi, G. Mukri, P. Gangshetty, A. Talwar, and B. Archana. 2014. Genetic variability and marker detection for rust resistance in recombinant inbred lines and backcross inbred lines of groundnut (*Arachis hypogaea* L.). In AAGB 2014 - 7th International Conference on Advances in *Arachis* through Genomics & Biotechnology. Georgia, USA, 29.
34. Pandey, M.K., B. Guo, C.C. Holbrook, P. Janila, X. Zhang, D.J. Bertioli, S. Isobe, X. Liang, and R.K. Varshney. 2014. Molecular markers, genetic maps and QTLs for molecular breeding in peanut. N. Mallikarjuna and R. K. Varshney (eds.) Genetics, Genomics and Breeding of Peanut. CRC Press. pp. 79-113.
35. Pandey M.K., M. Wang, L. Qiao, S. Feng, P. Khera, H. Wang, B. Tonniss, N. Barkley, J. Wang, C. Holbrook, A. Culbreath, R.K. Varshney, and B. Guo. 2014. Identification of QTLs associated with peanut oil contents in RIL populations and mapping FAD2 genes and their relative contribution towards oil quality. BMC Genetics. 15:133.
36. Wilson, J.N., R. Chopra, M.R. Baring, M. Gomez, C.E. Simpson, J.C. Chagoya, and M.D. Burow. 2014. Genetic mapping and QTL analysis for oil concentration in peanut. Amer. Peanut Res. Educ Soc. Annual Meeting.
37. Varshney, R.K., M.K. Pandey, P. Janila, S.N. Nigam, H.K. Sudini, M.V.C. Gowda, S. Manda, T. Radhakrishnan, S.S. Manohar, and P. Nagesh. 2014. Marker-assisted introgression of a QTL region to improve rust resistance in three elite and popular varieties of peanut (*Arachis hypogaea* L.). Theor. Appl. Genet. 127: 1771-1778.
38. Zhou, X., Y. Xia, X. Ren, Y. Chen, L. Huang, S. Huang, B. Liao, Y. Lei, L. Yan, and H. Jiang. 2014. Construction of a SNP-based genetic linkage map in cultivated peanut based on large scale marker development using next- generation double-digest restriction-site-associated DNA sequencing (ddRADseq). BMC Genomics 15:351.
39. Guimarães, P.M., L.G. Arrais, C.V. Morgante, O.B. Silva Jr, A.C.G. Araujo, A.C.Q. Martins, M.A.P. Saraiva, T.N. Oliveira, R.C. Togawa, S.C.M. Leal-Bertioli, D.J. Bertioli, and A.C.M. Brasileiro. 2015. Root transcriptome analysis of wild peanut reveals candidate genes for nematode resistance. PLoS One. <https://doi.org/10.1371/journal.pone.0140937>.
40. Holbrook, A.K. Culbreath, R.K. Varshney, and B. Guo. 2015. Genetic mapping of QTLs controlling fatty acids provided insights into the genetic control of fatty acid synthesis pathway in peanut (*Arachis hypogaea* L.). PlosOne 10(4):e0119454.doi:10.1371/journal.pone.0119454.
41. Jakkeral, S.A., H.L. Nadaf, M.V.C. Gowda, R.S. Bhat, R.K. Patil, B.N. Motagi, P.V. Kenchanagowda, G. Mukri, B. Archana, P. Gangshetty, K. Gangadhar, and L. Jaggal. 2014. Marker detection and genetic analysis for rust resistance of recombinant and backcross inbred Hake, A. A., M. V. C. Gowda, and R. S. Bhat. 2015. Analysis of TMV 2 × NLM derived RIL population towards mapping genomic regions governing taxonomic and morphological traits in peanut. In 5th International Conference on Next Generation Genomics and Integrated Breeding for Crop Improvement. ICRISAT, Hyderabad, India.
42. Peng, Z., M. Gallo, B.L. Tillman, D. Rowland, and J. Wang. 2015. Molecular marker development from transcript sequences and germplasm evaluation for cultivated peanut (*Arachis hypogaea* L.). Mol. & General Genet. 291(1):363-381.
43. Kanyika, B.T.N., D. Lungu, A.M. Mweetwa, E. Kaimoyo, V. M. Njung'e, E.S. Monyo, M. Siambi, G.

- He, Y. Zhao, and S.D.Villiers. 2015. Identification of groundnut (*Arachis hypogaea*) SSR markers suitable for multiple resistance traits QTL mapping in African germplasm. *Electronic Journal of Biotech.* 18(2): 61-67.
44. Smýkal, P., C.J. Coyne, M.J. Ambrose, N. Maxted, H. Schaefer, M.W. Blair, J. Berger, S. Greene, M.N. Nelson, N. Besharat, T. Vymyslický, C. Toker, R.K. Saxena, M. Roorkiwal, M.K. Pandey, J. Hu, Y.H. Li, L.X. Wang, Y. Guo, L.J. Qiu, R.J. Redden, and R.K. Varshney. 2015. Legume crops phylogeny and genetic diversity for science and breeding. *Critical Reviews in Plant Sci.* 34:43-104.
 45. Sukruth, M., S.A. Paratwagh, V. Sujay, V. Kumari, M.V.C. Gowda, H.L. Nadaf, B.N. Motagi, S. Lingaraju, M.K. Pandey, R.K. Varshney, and R.S. Bhat. 2015. Validation of markers linked to late leaf spot and rust resistance, and selection of superior genotypes among diverse recombinant inbred lines and backcross lines in peanut (*Arachis hypogaea* L.). *Euphytica* 204:343-351.
 46. Wang, M.L., P. Khera, M.K. Pandey, H. Wang, L. Qiao, S. Feng, B. Tonniss, N.A. Barkley, D. Pinnow, C.C. Holbrook, A.K. Culbreath, R.K. Varshney, and B. Guo. 2015. Genetic mapping of QTLs controlling fatty acids provided insights into the genetic control of fatty acid synthesis pathway in peanut (*Arachis hypogaea* L.). *PLoS ONE* DOI:10.1371/journal.pone.0119454.
 47. Guo, B.Z., P. Khera, H. Wang, Z. Peng, H. Sudini, X. Wang, M. Osiru, J. Chen, V. Vadez, M. Yuan, C.T. Wang, X. Zhang, F. Waliyar, J. Wang, and R.K. Varshney. 2016. pp. 163-208 In: H.T. Stalker and R.F. Wilson (eds.). *Annotation of trait loci on integrated genetic maps of Arachis species. Peanuts: Genetics, Processing, and Utilization.* Academic Press and AOCS Press, Waltham, MA.
 48. Chu, Y., R. Gill, J. Clevenger, P. Timper, C.C. Holbrook, and P. Ozias-Akins. 2016. Identification of rare recombinants leads to tightly linked markers for nematode resistance in peanut. *Peanut Sci.* 43: 88-93.
 49. Clevenger, J., Y. Chu, B. Scheffler, and P. Ozias-Akins. 2016. A developmental transcriptome map for allotetraploid *Arachis hypogaea*. *Plant Sci.* 7:1446.
 50. Khera, P., M.K. Pandey, H. Wang, S. Feng, L. Qiao, A.K. Culbreath, S. Kale, J. Wang, C.C. Holbrook, W. Zhuang, R.K. Varshney, and B. Guo. 2016. Mapping quantitative trait loci of resistance to tomato spotted wilt virus and leaf spots in a recombinant inbred line population of peanut. (*Arachis hypogaea* L.) from SunOleic 97R and NC94022. *PLoS ONE* 11(7): e0158452.
 51. Kolekar, R.M., V. Sujay, K. Shirasawa, S.B. Yeri, D.B. Chougale, B. Asha, M.V.C. Gowda, R.K. Varshney, and R.S. Bhat. 2015. Mapping late leaf spot and rust resistance using an improved map from the RILs of TAG 24 × GPBD 4 in peanut (*Arachis hypogaea*). In 5th International Conference on Next Generation Genomics and Integrated Breeding for Crop Improvement. ICRISAT, Hyderabad, India.
 52. Kolekar, R. M., V. Sujay, K. Shirasawa, M. Sukruth, M.V.C. Gowda, M.K. Pandey, R.K. Varshney, and R.S. Bhat. 2016. QTL mapping for late leaf spot and rust resistance using an improved genetic map and extensive phenotypic data on a recombinant inbred line population in peanut (*Arachis hypogaea* L.). *Euphytica* 209:147-156.
 53. Kolekar, R.M., V. Sujay, K. Shirasawa, M. Sukruth, Y.P. Khedikar, M.V.C. Gowda, M.K. Pandey, R.K. Varshney, and R.S. Bhat. 2016. QTL mapping for late leaf spot and rust resistance using an improved genetic map and extensive phenotypic data on a recombinant inbred line population

- in peanut (*Arachis hypogaea* L.). *Euphytica* 209:147-156.
54. Leal-Bertioli, S.C.M., M.C. Moretzsohn, P.A. Roberts, C. Ballén-Taborda, T.C.O. Borba, P.A. Valdisser, R.P. Vianello, A.C.G. Araújo, P.M. Guimarães, and D.J. Bertioli. 2016. Genetic mapping of resistance to *Meloidogyne arenaria* in *Arachis stenosperma*: a new source of nematode resistance for peanut. *G3: Genes, Genomes, Genetics* 6:377-390.
 55. Li, Y., L. Li, X. Zhang, K. Zhang, D. Ma, J. Liu, X. Wang, F. Liu, and Y. Wan. 2017. OTL mapping analysis of main stem height and the first lateral branch length in peanut (*Arachis hypogaea* L.). *Euphytica* 213:57.
 56. Liu L., P. Dang, and C. Chen. 2015. Development and utilization of InDel markers to identify peanut (*Arachis hypogaea*) disease resistance. *Frontiers in Plant Sci.* 6:988.
 57. Luo, H., Z. Xu, Z. Li, X. Li, J. Lv, X. Ren, L. Huang, X. Zhou, Y. Chen, J. Yu, W. Chen, Y. Lei, B. Liao, and H. Jiang. 2017. Development of SSR markers and identification of major quantitative trait loci controlling shelling percentage in cultivated peanut (*Arachis hypogaea* L.). *Theoretical and Applied Genetics* 130:1635-1648.
 58. Nagy, E.D., Y. Guo, S. Tang, J.E. Bowers., R.A. Okashah, C.A. Taylor, D. Zang, S. Khanal, A.F. Heesacker, N. Khalilian, A.D. Farmer, N. Carrasquilla-Garcia, R. Varma Penmetsa, D. Cook, H.T. Stalker, N. Nielsen, P. Ozias-Akins, and S.J. Knapp. 2012. A high-density genetic map of *Arachis duranensis*, a diploid ancestor of cultivated peanut. *BMC Genomics* 13:469.
 59. Clevenger, J., Y. Chu, C. Chavarro, G. Agarwal, D.J. Bertioli, S. Leal-Bertioli, M.K. Pandey, J. Vaughn, B. Abernathy, N.A. Barkley, R. Hovav, M. Burow, S.N. Nayak, A. Chitikineni, T.G. Isleib, C.C. Holbrook, S.A. Jackson, R.K. Varshney, and P. Ozias-Akins. 2017. Genome-wide SNP genotyping resolves signatures of selection and tetrasomic recombination of peanut. *Mol. Plant* 10:309-322.
 60. Clevenger, J., Y. Chu, L.A. Guimaraes, T. Maia, D. Bertioli, S. Leal-Bertioli, P. Timper, C. Holbrook, and P. Ozias-Akins. 2017. Gene expression profiling describes the genetic regulation of *Meloidogyne arenaria* resistance in *Arachis hypogaea* and reveals a candidate gene for resistance. *Sci. Rep.* 7:1317.
 61. Shasidhar, Y., M.K. Vishwakarma, M.K. Pandney, P. Janila, M.T. Variath, S.S. Manohar, S.N. Nigam, B. Guo, and R.K. Varshney. 2017. Molecular mapping of oil content and fatty acids using dense genetic maps in groundnut (*Arachis hypogaea* L.). *Plant Sci.* 8:794.
 62. Zhao, C., J. Qiu, G. Agarwal, J. Wang, X. Ren, H. Xia, B. Guo, C. Ma, S. Wan, D.J. Bertioli, R.K. Varshney, M.K. Padney, and X. Wang. 2017. Genome-wide discovery of microsatellite markers from diploid progenitor species, *Arachis duranensis* and *A. ipaënsis*, and their application in cultivated peanut (*A. hypogaea*). *Plant Sci.* 8:1209.
 63. Wang, H., S. Khera, S. Li, Y. Ren, M. Yuan, W. Zhuang, R.K. Varshney, B.Z. Guo, and L. Xie. 2017. Application of highly polymorphic SSR markers in peanut breeding and true hybrid identification. *Journal of Fujian Agriculture and Forestry University (Natural Science)* (in press).

Component 5: Matching up Genes and Traits

64. Gill, R., A. Culbreath, C. Holbrook, and P. Ozias-Akins. 2012. Characterization of a recombinant inbred line (RIL) population for resistance to late leaf spot in cultivated peanut (*Arachis hypogaea*). *Phytopathology* 102:S2.4.
65. Zhao, Y., C.S. Prakash, and G. He. 2012. Characterization and compilation of polymorphic simple

- sequence repeat (SSR) markers of peanut from public database. BMC Research Notes 5: 362.
66. Bertioli, D.J., B. Vidigal, S. Nielen, M.B. Ratnaparkhe, T.H. Lee, S.C. Leal-Bertioli, C. Kim, P.M. Guimaraes, G. Seijo, T. Schwarzacher, A.H. Paterson, P. Heslop-Harrison, and A.C. Araujo. 2013. The repetitive component of the A genome of peanut (*Arachis hypogaea*) and its role in remodeling intergenic sequence space since its evolutionary divergence from the B genome. Ann. Bot. 112:545-559.
 67. Guo, B., M.K. Pandey, G. He, R.K. Varshney, X. Zang, B. Liao, A. Culbreath, R.K. Varshney, V. Nwosu, R.F. Wilson, and H.T. Stalker. 2013. Recent advances in molecular genetic linkage maps of cultivated peanut (*Arachis hypogaea* L.). Peanut Sci. 40:89-94.
 68. Holbrook, C.C., T.G. Isleib, P. Ozias-Akins, Y. Chu, S.J. Knapp, B. Tillman, B. Guo, R. Gill, and M.D. Burow. 2013. Development and phenotyping of recombinant inbred line (RIL) populations for peanut (*Arachis hypogaea*). Peanut Sci. 40:89-94.
 69. Shekoofa, A., J.M. Devi, T.R. Sinclair, C.C. Holbrook, and T.G. Isleib. 2013. Divergence in drought- resistance traits among parents of recombinant peanut inbred lines. Crop Sci. 53:2569-2576.
 70. Stalker, H.T., S. Tallury, P. Ozias-Akins, S. Lea-Bertioli, and D. Bertioli. 2013. The value of diploid peanut relatives for breeding and genomics. Peanut Sci. 40:70-88.
 71. Wang, M.L., C.Y. Chen, B. Tonniss, N.A. Barkley, D.L. Pinnow, R.N. Pittman, J. Davis, C.C. Holbrook, H.T. Stalker, and G.A. Pederson. 2013. Oil, fatty acid, flavonoid, and resveratrol content variability and FAD2A functional SNP genotypes in the U.S. peanut mini-core collection. J. Agric. Food Chem. 61:2875-2882.
 72. Baidouri, M.E.L., M.C. Carpentier, R. Cooke, D. Gao, E. Lasserre, C. Llauro, M. Mirouze, N. Picault, S.A. Jackson, and O. Panaud. Incorporation of disease resistances from wild species. 2014. The flying genes: widespread and frequent horizontal transfers in plants. Gen. Res. 24:831-838.
 73. Burow, M.D., J.L. Starr, C.-H. Park, C.E. Simpson, and A.H. Paterson. 2014. Introgression of homeologous quantitative trait loci (QTLs) for resistance to the root-knot nematode [*Meloidogyne arenaria* (Neal) Chitwood] in an advanced backcross-QTL population of peanut (*Arachis hypogaea* L.) Mol. Breed. 34: 393-406.
 74. Burow, M.D., J.L. Starr, C.-H. Park, C.E. Simpson, and A.H. Paterson. 2014. Identification of QTLs for resistance to the root-knot nematode (*Meloidogyne arenaria* (Neal) Chitwood) in an advanced backcross-QTL population demonstrates introgression of orthologous resistance genes in peanut (*Arachis hypogaea* L.) Mol. Breed. 34:393-406.
 75. Chamberlin, K.D., N.L. Barkley, B.L. Tillman, J.W. Dillwith, R. Madden, M.E. Payton, and R.S. Bennett. 2014. A comparison of methods used to determine the oleic/linoleic acid ratio in cultivated peanut (*Arachis hypogaea* L.). Agricultural Sci. 5(3):227-237.
 76. Chopra, R., G. Burow, A. Farmer, J. Mudge, C.E. Simpson, and M.D. Burow. 2014. Comparisons of de novo transcriptome assemblers in diploid and polyploid species using peanut (*Arachis* spp.) RNA-seq data. PloS ONE. 9(12): e115055. doi: 10.1371/journal.pone.0115055.
 77. Chu, Y., L.A. Guimaraes, P. Timper, C. Holbrook, and P. Ozias-Akins. 2014. A technique to study *Meloidogyne arenaria* resistance in agrobacterium rhizogenes-transformed peanut. Plant Dis. 98:1292-1299.
 78. Holbrook, C.C., T.B. Brenneman, H.T. Stalker, W.C. Johnson, III, P. Ozias-Akins, Y. Chu, G. Vellidis, and D. McClusky. 2014. Peanut. In: S. Smith, B. Diers, and B Carver Editors, pp 101-

122. Yield gains in major U.S. field crops. CSSA, Madison, WI.
79. Kumari, V., M.V.C. Gowda, V. Tasiwal, M.K. Pandey, R.S. Bhat, N. Mallikarjuna, H.D. Upadhyaya, and R.K. Varshney. 2014. Diversification of primary gene pool through introgression of resistance to foliar diseases from synthetic amphidiploids to cultivated groundnut (*Arachis hypogaea* L.) The Crop J. 2:110-119.
80. Lee, S-H., R. Tuberosa, S.A. Jackson, and R.K. Varshney. 2014. Genomics of plant genetic resources: a gateway to a new era of global food security. Plant Genet. Resources: Characterization and Utilization. 12: S2-S5.
81. Rami, J.F., S.M. Leal-Bertioli, D. Foncéka, M. Moretzsohn, and D. Bertioli. 2014. Alien gene transfer in crop plants. pp.:253-279. In Groundnut, A. Pratap and J. Kumar eds.). Springer New York.
82. Tallury, S.P., J. Hollowell, T.G. Isleib, and H.T. Stalker. 2014. Greenhouse evaluation of section *Arachis* wild species for Sclerotinia blight and CBR resistance. Peanut Sci. 41:17-24.
83. Kumari, V., M.V.C. Gowda, V. Tasiwal, M.K. Pandey, R.S. Bhat, N. Mallikarjuna, H.D. Upadhyaya, and R.K. Varshney. 2014. Diversification of primary gene pool through introgression of resistance for foliar diseases from synthetic amphidiploids to cultivated groundnut (*Arachis hypogaea* L.). Crop J. 2:110-119.
84. Wan, L., Y. Wu, J. Huang, X. Dai, Y. Lei, L. Yan, H. Jiang, J. Zhang, R.K. Varshney, and B. Liao. 2014. Identification of ERF genes in peanuts and functional analysis of AhERF008 and AhERF019 in abiotic stress response. Functional & Integrative Genomics 14: 467-477.
85. Zhang, H., L. Scharfenstein, D. Zhang, P.-K. Chang, B.G. Montalbano, B.Z. Guo, X. Meng, and J. Yu. Peanut resistance gene expression in response to *Aspergillus flavus* infection during seed germination. Journal of Phytopathology 163: 212-221.
86. Zhu, F., J. Han, S. Liu, X. Chen, R.K. Varshney, X. Liang. 2014. Cloning, expression pattern analysis and subcellular localization of resveratrol synthase gene in peanut (*Arachis hypogaea* L.). American J. Plant Sci. 5: 3619-3631. Chavarro, C., B. Abernathy, D. Bertioli, S. Jackson, C.C. Holbrook, J. Clevenger, and P. Ozias-Akins. 2014. Identification of SNPs for *Arachis hypogaea* L. genotypes using WGS based on the two diploid reference genomes. 47th Annual Meeting of the American Peanut Res. And Educ. Society, Charleston, SC.
87. Chopra, R., C.E. Simpson, A. Hillhouse, P. Payton, J. Sharm, and M.D. Burow. 2014. Transcript-based SNP map and QTL analysis on plant architecture traits of F2 lines developed from intra-specific cross of *Arachis duranensis* x *Arachis cardenasii*. 7th International Conference on Advances in *Arachis* through Genomics & Biotechnology, Savannah, GA.
88. Varshney, R.K., L. Froencike, M.K. Pandey, S.N. Nayak, L.H.D. Upadhyaya, A. Rathore, B. Liao, N. Barkley, B. Guo, M.V.C. Gowda, C.C. Holbrook, T.B. Brenneman, R.S. Bhat, C. Chen, J. Damicone, M. D. Burrow, T. Isleib, L. Dean, M.L. Wang, V. Vadez, S. Jackson, and R. Micheltmore. 2014. An international initiative to conduct comprehensive genome-wide association studies (GWAS) for an array of agronomic traits in peanut In AAGB 2014 - 7th International Conference on Advances in *Arachis* through Genomics & Biotechnology. Georgia, USA, pp. 41-42.
89. Pandey, M.K., H.D. Upadhyaya, A. Rathore, V. Vadez, M.S. Sheshshaye, M. Sriswathi, M. Govil, A. Kumar, M.V.C. Gowda, S. Sharma, F. Hamidou, V.A. Kumar, P. Khera, R.S. Bhat, A.W. Khan, S. Singh, H. Li, E. Monyo, H.L. Nadaf, G. Mukri, S.A. Jackson, B. Guo, X. Liang, and R.K. Varshney. 2014. Genome-wide association studies for 50 agronomic traits in peanut using the 'reference

- set' comprising 300 genotypes from 48 countries of the semi- arid tropics of the world. PLoS ONE 9(8): e105228.
90. Abberton, M., J. Batley, A. Bentley, J. Bryant, H. Cai, J. Cockram, A. Costa de Oliveira, L.J. Cseke, H. Dempewolf, C. De Pace, D. Edwards, P. Gepts, A. Greenland, A.E. Hall, R. Henry, K. Hori, G.T. Howe, S. Hughes, M. Humphreys, D. Lightfoot, A. Marshall, S. Mayes, H.T. Nguyen, F.C. Ogonnaya, R. Ortiz, A.H. Paterson, R. Tuberosa, B. Valliyodan, R.K. Varshney, and M. Yano. 2016. Global agricultural intensification during climate change: a role for genomics. *Plant Biotech. J.* 14(4):1095-1098.
 91. Bennett, R.S., M.E. Payton, and K.D. Chamberlin. 2015. Response to oxalic acid as a resistance assay for *Sclerotinia minor* in peanut. *Peanut Sci.* 42(1):56-64.
 92. Brasileiro, A.C.M., C. Morgante, A.C.G. Araujo, S.C.M. Leal-Bertioli, A.K. Silva, A.C.Q. Martins, C.C. Vinson, C.M.R. Santos, O. Bonfim, R.C. Togawa, M.A.P. Saraiva, D.K. Bertioli, and P.M. Guimaraes. 2015. Transcriptome profiling of wild *Arachis* from water-limited environments uncovers drought tolerance candidate genes. *Plant Mol. Bio. Reporter* 1-17.
 93. Chandran, M., Y. Chu, S. J. Maleki, and P. Ozias-Akins. 2015. Stability of transgene expression in reduced allergen peanut (*Arachis hypogaea* L.) across multiple generations and at different soil sulfur levels. *J. Agric. Food Chem.* 63:1788-1797.
 94. Chen, H., C. Zhang, Y. Deng, S. Zhou, Y. Zheng, S. Ma, R. Tang, R.K. Varshney, and W. Zhuang. 2015. Identification of low Ca²⁺ stress-induced embryo apoptosis response genes in *Arachis hypogaea* by SSH-associated library lift (SSHALL). *Plant Biotech. J.* 14(2):682-698.
 95. Clevenger, J. and P. Ozias-Akins. 2015. SWEEP: A tool for filtering high-quality SNPs in polyploid crops. *G3* 5: 1797-1803.
 96. Clevenger, J., C. Chavarro, S.A. Pearl, P. Ozias-Akins, and S.A. Jackson. 2015. SNP identification in polyploids: A review, example and recommendations. *Mol. Plant* 8:831-846.
 97. Faye, I., M.K. Pandey, F. Hamidou, A. Rathore, O. Ndoye, V. Vadez, and R.K. Varshney. 2015. Identification of quantitative trait loci for yield and yield related traits of groundnut (*Arachis hypogaea* L.) under different water regimes in Niger and Senegal. *Euphytica*. 206(3):631-647.
 98. Gupta, K., I. Hedvat, A. Faigenboim-Doron, P. Ozias-Akins, J. Clevenger, and R. Hovav. 2015. Transcriptome profiling of peanut developing seed with a focus on duplicate oil related pathways. *BMC Genomics*.
 99. Huang, E.B., K.L. Verbyla, A.P. Verbyla, C. Raghavan, V.K. Singh, P.M. Gaur, H. Leung, R.K. Varshney, and C.R. Cavanagh. 2015. MAGIC populations in crops: current status and future prospects. *Theoretical and Applied Genet.* 128:999-1017.
 100. Isleib, T.G., H.E. Pattee, R.S. Tubbs, T.H. Sanders, L.O. Dean, and K.W. Hendrix. 2015. Intensities of sensory attributes in high- and normal-oleic cultivars in the uniform peanut performance test. *Peanut Sci.* 42(2):83-91.
 101. Johanningsmeier, S.D., G.K. Harris, and C.M. Klevorn. 2016. Metabolomic technologies for improving the quality of food: Practice and promise. *Annual Review of Food Science and Technology*. 7:413-438.
 102. Kanade, S.G., A.A. Shaikh, and J.D. Jadhav. 2015. Sowing environments effect on rust (*P. arachidis*) disease in groundnut (*Arachis hypogaea* L.). *Int. J. Plant Protection*. 8:174-180.
 103. Klevorn, C.M., K.W. Hendrix, T.H. Sanders, and L.L. Dean. 2015. Differences in development

- of oleic and linoleic acid in high- and normal-oleic virginia and runner-type peanuts. *Peanut Sci.* 43:12-23.
104. Leal-Bertioli, S., K. Shirasawa, B. Abernathy, M. Moretzsohn, C. Chavarro, J. Clevenger, P. Ozias-Akins, S. Jackson, and D. Bertioli. 2015. Tetrasomic recombination is surprisingly frequent in allotetraploid *Arachis*. *Genetics* 199:1093-1097.
 105. Leal-Bertioli, S.C.M., S.P. Santos, K.M. Dantas, P.W. Inglis, S. Nielen, A.C.G. Araujo, J.P. Silva, U. Cavalcante, P.M. Guimaraes, A.C.M. Brasileiro, N. Carrasquilla-Garcia, R.V. Penmetza, D. Cook, M.C. Moretzsohn, and D.J. Bertioli. 2015. *Arachis batizocoi*: A study of its relationship to cultivated peanut (*A. hypogaea*) and its potential for introgression of wild genes into the peanut crop using induced allotetraploids. *Annals of Botany* 115 (2):237-249.
 106. Leal-Bertioli, S.C.M., U. Cavalcante, E.G. Gouvea, C. Ballén-Taborda, K. Shirasawa, P.M. Guimarães, S.A. Jackson, D.J. Bertioli, and M.C. Moretzsohn. 2015. Identification of QTL for rust resistance in the peanut wild species *Arachis magna* and the development of KASPar markers for marker assisted selection. *G3: Genes, Genet. and Gen.* 5:1403-1413.
 107. Luis, J.M., P. Ozias-Akins, C.C. Holbrook, R.C. Kemeraït, Jr., J.L. Snider, and V. Liakos. 2016. Phenotyping peanut genotypes for drought tolerance. *Peanut Sci.* 43(1):36-48.
 108. Oellrich A., R.L. Walls, E.K. Cannon, S.B. Cannon, L. Cooper, J. Gardiner, G.V. Gkoutos, L. Harper, M. He, R. Hoehndorf, P. Jaiswal, S.R. Kalberer, J.P. Lloyd, D. Meinke, N. Menda, L. Moore, R.T. Nelson, A. Pujar, C.J. Lawrence, and E. Huala. An ontology approach to comparative phenomics in plants. 2015. *Plant Methods*. 11:10.
 109. Paratwagh, S.A., and R.S. Bhat. 2015. Development of superior introgression lines for resistance to foliar diseases and productivity in groundnut (*Arachis hypogaea* L.). *Electron. J. Plant Breed.* 6(4):1034-1040.
 110. Varshney, R.K. 2015. Exciting journey of 10 years from genomes to fields and markets: Some success stories of genomics-assisted breeding in chickpea, pigeonpea and groundnut. *Plant Sci.* 242:98-107.
 111. Varshney, R.K., H. Kudapa, L. Pazhamala, A. Chitkineni, M. Thudi, A. Bohra, P.M. Gaur, P. Janila, A. Fikre, P. Kimurto, and N. Ellis. 2015. Translational genomics in agriculture: some examples in grain legumes. *Critical Reviews in Plant Sci.* 34:169-194.
 112. Wu, C., R. Gill, Y. Chu, C.C. Holbrook, and P. Ozias-Akins. 2015. Fine phenotyping of pod and seed traits in *Arachis* germplasm accessions using digital image analysis. *Peanut Sci.* 42:65-73.
 113. Bertioli, D., D. Gao, B.S. Vidigal, L. Froenicke, S. Cannon, B. Abernathy, L. Ren, A.C. Guerra Araujo, P.M. Guimarães, S. Bertioli, R. Michelmore, and S. Jackson. 2015. Transposable elements in the A and B genomes of peanut. XXIV Plant and Animal Genome Conference, San Diego, CA. W866
 114. Gao, D., D. Bertioli, A. Iwata, Y. Chu, J.P. Clevenger, P. Ozias-Akins, L. Froenicke, X. Liu, and S. Cannon. 2015. Annotation of transposable elements for peanut improvement and genomics. 47th American Peanut Research and Education Society (APRES) Annual Meeting, Charleston, SC. 2105:28
 115. Chamberlin, K.D., J.P. Damicone, M.R. Baring, M.D. Burow, C.B. Godseye, R.S. Bennett, H.A. Melouk, and C.E. Simpson. 2015. Registration of high-oleic peanut germplasm line ARSOK-S1 (TX996784) with enhanced resistance to sclerotinia blight and pod rot. *J. Plant Regist.* 9:103-107.

116. Barkley, N.A., H.D. Upadhyaya, B. Liao, and C.C. Holbrook. 2016. pp. 67-110. In: H.T. Stalker and R.F. Wilson (eds.). Global resources of genetic diversity in peanut. Peanuts: Genetics, Processing & Utilization. Academic Press and AOCS Press, Waltham, MA.
117. Chu, Y., J. Clevenger, R. Hovav, J. Wang, B. Scheffler, S.A. Jackson, and P. Ozias-Akins. 2016. pp. 209-240. In: R.F. Wilson and H.T. Stalker (eds.). Application of Genomic, Transcriptomic, and Metabolomic Technologies in *Arachis* Species. Peanuts: Genetics, Processing, and Utilization. Academic Press and AOCS Press, Waltham, MA.
118. Holbrook, C.C., M.D. Burow, C. Chen, M.K. Pandey, L. Liu, J.C. Chagoya, Y. Chu, P. Ozias-Akins. 2016. pp. 111-146. In: RF Wilson and HT Stalker (eds.). Recent advances in peanut breeding and genetics. Peanuts: Genetics, Production, Processing and Utilization. Academic Press and AOCS Press, Waltham, MA.
119. Khera, P., K.P. Pandey, N. Mallikarjuna, M. Sriswathi, M. Roorkiwal, P. Janila, K. Shilpa, H. Sudini, B.Z. Guo, R.K. Varshney. Development of introgression lines and advanced backcross QTL analysis for disease resistance, oil quality and yield component traits in peanut (*Arachis hypogaea* L.). Mol. Breed.
120. Stalker, H.T., S. Tallury, G. Seijo, and S. Leal-Bertioli. 2016. pp. 27-66. In: H.T. Stalker and R.F. Wilson (eds.). Biology, speciation and utilization of peanut species. Peanuts: Genetics, Processing and Utilization. Academic Press and AOCS Press, Waltham, MA.
121. Janila, P., M.K. Pandey, S.S. Manohar, M.T. Variath, P. Nallathambi, H.L. Nadaf, H. Sudini, and R.K. Varshney. 2016. Foliar fungal disease-resistant introgression lines of groundnut (*Arachis hypogaea* L.) record higher pod and haulm yield in multilocation testing. Plant Breed. 135:355-366.
122. Huang, L., X. Ren, B. Wu, X. Li, W. Chen, X. Zhou, Y. Chen, M.K. Pandey, Y. Jiao, H. Luo, Y. Lei, R.K. Varshney, B. Liao, and H. Jiang. 2016. Development and deployment of a high-density linkage map identified quantitative trait loci for plant height in peanut (*Arachis hypogaea* L.). Sci. Rep. 6:39478.
123. Chen, X., H. Li, M.K. Pandey, Q. Yang, X. Wang, V. Garg, H. Li, X. Chi, D. Doddamani, Y. Hong, H. Upadhyaya, H. Guo, A.W. Khan, F. Zhu, K.AVS Krishnamohan, M. Chen, N. Zhong, G. Agarwal, S. Li, A. Chitikineni, G.Q. Zhang, S. Sharma, N. Chen, H. Liu, P. Janila, Z. Yang, J. Zhao, C. Zhang, Y. Yu, J. Bi, X. Zhang, Z.J. Liu, A.H. Paterson, S. Wang, X. Liang, R.K. Varshney, and S. Yu. 2016. Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens. Proceedings of the National Academy of Sci. 113(24):6785-6790.
124. Wang, P., C. Gao, X. Bian, S. Zhao, C. Zhao, H. Xia, H. Song, L. Hou, S. Wan, and X. Wang. 2016. Genome-wide identification and comparative analysis of cytosine-5 DNA methyltransferase and demethylase families in wild and cultivated peanut. Plant Sci. 7:7.
125. Katham, R., K. Sakata, P. Suravajhala, T. Pechan, D.M. Kambiranda, K.S. Naik, B. Guo, and S.M. Basha. 2016. Comparative leaf proteomics of drought-tolerant and -susceptible peanut in response to water stress. J. of Proteomics 143:209-226.
126. Willmon, D., A.R. Devireddy, M. Inupakutika, N. Puppala, and Y. Cho. 2017. Stress responses of peanut (*Arachis hypogaea* L.) genotypes as measured by trigonelline content after exposure to UV-B radiation. Plant Sci. 8:998-1010.
127. Bertioli, D.J., S.C. Leal-Bertioli, and H.T. Stalker. 2016. pp. 147-161. The peanut genome: the

- history of the consortium and the structure of the genome of cultivated peanut and its diploid ancestors. *Peanuts: Genetics, Processing, and Utilization*. Academic Press and AOCS Press, Waltham, MA.
128. McKinney, J.L. and B.L. Tillman. 2016. Spotted wilt in peanut as impacted by genotype resistance, planting date, and plant population. *Crop Sci.* 57:130-136.
 129. Leon, R.G., M.J. Mulvaney, and B.L. Tillman. 2016. Peanut cultivars differing in growth habit and canopy architecture respond similarly to weed interference. *Peanut Sci.* 43(2):133-140.
 130. Tseng, Y.C., B.L. Tillman, Z. Peng, and J. Wang. 2016. Identification of major QTLs underlying tomato spotted wilt virus resistance in peanut cultivar Florida-EP '113'. *BMC Genetics.* 17:128.
 131. Klevron, C.M., K.W. Hendrix, T.H. Sanders, and L.L. Dean. 2016. Differences in development of oleic and linoleic acid in high-and normal-oleic virginia and runner-type peanuts. *Peanut Sci.* 43:12-23.
 132. Aninbon, C., S. Jogloy, N. Vorasoot, S. Nuchadomrong, T. Senawong, C.C. Holbrook, and A. Patanothai. 2016. Effect of mid season drought on phenolic compounds in peanut genotypes with different levels of resistance to drought. *Field Crops Res.* 187:127-134.
 133. Pandey, M.K., H. Wang, P. Khera, M.K. Vishwakama, S.M. Kale, A.K. Culbreath, C.C. Holbrook, X. Wang, R.K. Varshney, and B. Guo. 2017. Genetic dissection of novel QTLs for resistance to leaf spots and tomato spotted wilt virus in peanut (*Arachis hypogaea* L.). *Plant Sci.* 8:25.
 134. Luis, J.M., P. Ozias-Akins, C.C. Holbrook, R.C. Kemerait, Jr., J.L. Snider, and Vasilis. 2016. Phenotyping peanut genotypes for drought tolerance. *Peanut Sci.* 43(1): 36-48.
 135. Barkley, N.A., H.D. Upadhyaya, B. Liao, and C.C. Holbrook. 2016. pp. 67-110. *Global Resources of genetic diversity in peanut. Peanuts: Genetics, Proceeding, and Utilization*. Academic Press and AOCS Press, Waltham, MA.
 136. Pandey, M.K., G. Agarwal, S.M. Kale, J. Clevenger, S.N. Nayak, M. Sriswathi, A. Chitikineni, C. Chavarro, X. Chen, H.D. Upadhyaya, M.K. Vishwakarma, S. Leal-Bertioli, X. Liang, D.J. Bertiolo, B. Guo, S.A. Jackson, P. Ozias-Akins, and R.K. Varshney. 2017. Development and evaluation of a high density genotyping 'Axiom_Arachis' array with 58K SNPs for accelerating genetics and breeding in groundnut. *Sci. Rep.* 7:40577.
 137. Korani, W.A., Y. Chu, C. Holbrook, J. Clevenger, and P. Ozias-Akins. 2017. Genotypic regulation of aflatoxin accumulation but not *Aspergillus* fungal growth upon post-harvest infection of peanut (*Arachis hypogaea* L.) seeds. *Toxins.* 9(7): 218.
 138. Luis, J.M., P. Ozias-Akins, C.C. Holbrook, R.C. Kemerait, Jr., J.L. Snider, and V. Liakos. 2016. Phenotyping peanut genotypes for drought tolerance. *Peanut Sci.* 43:36-48.

Component 6: Web-based genome libraries & the Genomic Toolbox; Cultivars Released

141. Shirasawa, K., S. Isobe, S. Tabata, and H. Hirakawa. 2014. Kazusa marker DataBase: a database for genomics, genetics, and molecular breeding in plants. *Breeding Sci.* 64:264.
142. Dash, S., E.K.S. Cannon, S.R. Kalberer, A.D. Farmer, S.B. Cannon. 2016. pp.241-252 In: H.T. Stalker and R.F. Wilson (eds.). *PeanutBase and other bioinformatic resources for peanut. Peanuts: Genetics, Processing, and Utilization*. Academic Press and AOCS Press, Waltham, MA.
143. Burow, M.D., M.R. Baring, J.L. Ayers, A.M. Schubert, Y. López, and C.E. Simpson. 2014. Registration of Tamrun OL12' peanut. *J. Plant Regist.* 8:117-121.

144. Burow, M.D., M.R. Baring, N. Puppala, C.E. Simpson, J.L. Ayers, J. Cason, A.M. Schubert, A. Muiña, and Y. López. 2014. Registration of 'Schubert' peanut. J. Plant Regist. 8:122-126.
145. Chen, C.Y., N.A. Barkley, M.L. Wang, C.C. Holbrook, and P.M. Dang. 2014. Registration of purified accessions for the U.S. peanut mini-core germplasm collection. J. Plant Regist. 8(1):77-85.
146. Tallury, S.P., T.G. Isleib, S.C. Copeland, P. Rosas-Anderson, M. Balota, D. Singh, and H.T. Stalker. 2014. Registration of two multiple disease-resistant peanut germplasm lines derived from *Arachis cardenasii* Krapov. & W.C. Gregory, GKP 10017. J. Plant Regist. 8:86-89.
147. Puppala, N. and S.P. Tallury. 2014. Registration of high oleic valencia peanut cultivar 'NuMex-01'. J. Plant Regist. 8:127-130.
148. Chamberlin, K.D., R. Bennett, J.P. Damicone, C.B. Godsey, H.A. Melouk, K. Keim. 2015. Registration of 'OLé' peanut. J. Plant Regist. 9(2):154-158.
149. Isleib, T.G., S.R. Milla-Lewis, H.E. Pattee, S.C. Copeland, M.C. Zuleta, B.B. Shew, J.E. Hollowell, T.H. Sanders, L.O. Dean, K.W. Hendrix, M. Balota, J.W. Chapin, and W.S. Monfort. 2015. Registration of 'Sugg' peanut. J. Plant Regist. 9: 44-52.
150. Tillman, B.L. and D.W. Gorbet. 2015. Registration of FloRun™ '107'. J. Plant Regist. 9(2):162-167.

Appendix 5. Sponsors who provide financial support for the Peanut Genome Project

U.S. Peanut Sheller Companies:

- American Peanut Sheller's Assoc.
- Birdsong Peanuts
 - Damascus Peanut Company
 - Golden Peanut Company
 - McCleskey Mills
 - Snyder's/Lance
 - Tifton Peanut Company
 - Williston Peanuts
- Southwestern Peanut Sheller's – Birdsong Peanuts
 - Clint Williams Company
 - Golden Peanut Company
 - Wilco Peanut Company
- Virginia Carolina Sheller's Assoc. Birdsong Peanuts
 - Golden Peanut Company
 - Peanut Processors
 - Severn Peanut Company
- American Peanut Growers Group
- Brooks Peanut Company
- Sessions Company
- Tifton Quality Growers

- Diamond Foods, Inc.
- E.J. Cox
- Hampton Farms
- The Hershey Company
- J.B. Sanfilippo
- Jimbo's Jumbo's
- J.M. Smucker
- Kraft – Planters
- Mars Chocolate
- Old Home Foods
- Pardoe's Perky Peanuts
- Peanut Butter & Company
- The Peanut Shop of Williamsburg
- Producers Peanut Company

Food Manufacturing Companies:

- Algood Food Company
- American Blanching
- Arway Confections, Inc.

US Peanut Producer Organizations:

- National Peanut Board
- Florida Peanut Producers Association
- Texas Peanut Producers Association
- Georgia Peanut Commission

Allied Sector Companies:

- B.A.G.
- Bayer CropScience
- Chips Group
- Concordia, LLC
- Dothan Warehouse
- Early Trucking
- Georgia Federal-State Inspection Service
- Hofler Brokerage
- International Service Group
- JLA USA
- Jack Wynn & Company
- J.R. James Brokerage
- Lewis M. Carter

- Kelly Manufacturing Company
- Lovatt & Rushing
- Mazur & Hockman
- M.C. McNeill & Co. LLC
- National Peanut Brokers Assn.
- National Peanut Buying Points Assn.
- Nolin Steel
- O'Connor & Company
- Olam International Limited
- RCB Nuts
- Reed Marketing, LLC
- Satake USA, Inc.
- SGL International, LLC
- Southern Ag Carriers

International Collaborators

BGI-Americas
 Henan Academy of Agricultural
 Sciences Chinese Academy of
 Agricultural Sciences Shandong
 Academy of Agricultural Sciences

Scientific and Technical Contributions to the Peanut Genome Project are provided by:

Auburn University
 BGI-Americas
 Catholic University-Brasilia
 Chinese Academy of
 Agricultural Sciences
 EMBRAPA
 Generation Challenge-Gates
 Foundation
 Henan Academy of
 Agricultural Sciences
 ICRISAT (India, West &
 Central Africa)
 Indian Council of
 Agricultural Research
 (ICAR) Kazusa DNA Research
 Institute (Japan)

National Center Genome Resources
 Peanut Company of Australia
 Shandong Academy of Agricultural Sciences
 North Carolina State University
 Texas A & M University
 University of California-Davis
 University of Florida
 University of Georgia
 USDA-Agricultural Research Service
 Volcani Center (Israel)

Appendix 6. Terms and Definitions

Abridged from <http://www.panzea.org/infor/faq.html>, and <http://www.netsci.org/Science/Bioinform/terms.html>

Allele: Different forms of a gene which occupy the same position on the chromosome.

Allotetraploid: A cell containing two pairs of different chromosomes (i.e. Peanut)

Autotetraploid: A cell containing two pairs of the same chromosomes (i.e. Soybean)

Amplification: The process of repeatedly making copies of the same piece of DNA.

Annotation: Text fields of information about a biosequence which are added to a sequence databases. Annotation (the elucidation and description of biologically relevant features in the sequence) consists of the description of the following items:

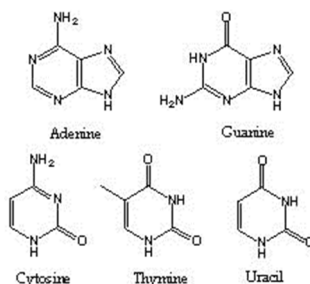
- Function(s) of the protein.
- Post-translational modification(s). For example carbohydrates, phosphorylation, acetylation, GPI- anchor, etc.
- Domains and sites. For example calcium binding regions, ATP-binding sites, zinc fingers, homeobox, kringle, etc.
- Secondary structure.
- Quaternary structure. For example homodimer, heterotrimer, etc.
- Similarities to other proteins.
- Disease(s) associated with deficiency(s) in the protein.
- Sequence conflicts, variants, etc.

Assembly: The process of placing fragments of DNA that have been sequenced into their correct position within the chromosome.

Association Mapping: As in QTL mapping, the goal of association mapping is to find a statistical association between genetic markers and a quantitative trait. However, in association mapping, the genetic markers usually must lie relatively close to a candidate gene. The goal is to identify the actual genes affecting that trait, rather than just (relatively large) chromosomal segments. QTL mapping is performed in a genetically defined population. Association mapping is performed at the population level within a set of unrelated or distantly-related individuals sampled from a population. Association mapping relies on linkage disequilibrium (LD) between the candidate gene markers and the polymorphism in that gene causes the differences in the phenotypic trait.

Bacterial artificial chromosome (BAC): A long sequencing vector which is created from a bacterial chromosome by splicing a DNA fragment from another species. Once the foreign DNA has been cloned into the host bacteria, many copies of the new chromosome can be made.

Base: One of five molecules which are assembled, along with a ribose and a phosphate, to form nucleotides (Figure 1). Adenine (A), guanine (G), cytosine (C), and thymine (T) are found in DNA while RNA is made from adenine (A), guanine (G), cytosine (C), and uracil (U).



Base pair (BP): The complementary bases on opposite strands of DNA which are held together by hydrogen bonding. The atomic structure of these bases preselect the pairing of adenine with thymine and the pairing of guanine with cytosine (or uracil in RNA).

Bioinformatics: An absolute definition of bioinformatics has not been agreed upon. The first level, however, can be defined as the design and application of methods for the collection, organization, indexing, storage, and analysis of biological sequences (both nucleic acids [DNA and RNA] and proteins). The next stage of bioinformatics is the derivation of knowledge concerning the pathways, functions, and interactions of these genes (functional genomics) and proteins (proteomics). Bioinformatics is also referred to as computational biology.

Candidate Genes: The distinction between "random" and "candidate" genes is of great importance. By random genes we refer to genes without any known function of the proteins (or RNAs) that they encode. They may be selected from a random set of expressed DNA sequences (DNA sequences that are copied, or transcribed, into RNA) at a time in cell development. Candidate genes refer to genes of known or suspected function or traits of interest.

Cell: The smallest functional structural unit of living matter. Cells are classed as either procaryotic and eucaryotic.

CentiMorgan (cM): The unit of measurement for distance and recombine frequency on a genetic map. Formally, the length (number of bases) that have a 1% probability of participating in mixing of genes. For humans, the average length of a cM is one million base pairs (or 1 megabase, Mb).

cDNA (complementary DNA): An artificial piece of DNA that is synthesized from an mRNA (messenger RNA) template and is created using reverse transcriptase. The single stranded form of cDNA is frequently used as a probe in the preparation of a physical map of a genome. cDNA is preferred for sequence analysis because the introns found in DNA are removed in translation from DNA ----> mRNA ----> cDNA.

Chromosome: A collection of DNA and protein which organizes the human genome. Each human cell contains 23 sets of chromosomes; 22 pairs of autosomes (non sex determining chromosomes) and one pair of sex determining chromosomes. The human genome within the 23 sets of chromosomes is made of approximately 30,000 genes which are built from over 3 billion base pairs. While eukaryotic chromosomes are complex sets of proteins and DNA, prokaryotic chromosomal DNA is circular with the entire genome on a single chromosome.

Cloning: The technique used to produce copies of a piece of DNA. A DNA fragment that contains a gene of interest is inserted into the genome of a virus or plasmid which is then allowed to replicate.

Cloning vector: A piece of DNA from any foreign body which is grafted into a host DNA strand that can then self replicate. Vectors are used to introduce foreign DNA into host cells for the

purpose of manufacturing large quantities of the new DNA or the protein that the DNA expresses.

Coding region: The portion of a genome that is translated to RNA which in turn codes protein (also see exon).

Codon: The set of three nucleotides along a strand of mRNA that determine (or code) the amino acid placement during protein synthesis. The number of possible arrangements of these three nucleotides (or triplet codes) available for protein synthesis is $(4 \text{ bases})^3 = 64$. Thus, each amino acid can be coded by up to 6 different triplet codes. Three triplet codes (UAA, UAG, UGA) specify the end of the protein. In the example below, three codons are shown.

--- UCA CGU CAU ---
 Ser Arg His

Complementarity: The sequence-specific or shape-specific recognition that occurs when two or more molecules bind together. DNA forms double stranded helixes because the complementary orientation of the bases in each strand facilitate the formation of the hydrogen bonds which hold the strands together.

Computational biology: See bioinformatics

Consensus sequence: The most commonly occurring amino acid or nucleotide at each position of an aligned series of proteins or polynucleotides.

Consensus map: The location of all consensus sequences in a series of multiply aligned proteins or polynucleotides.

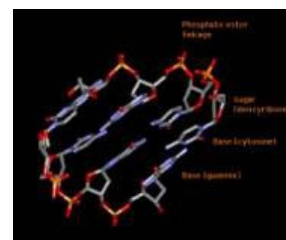
Conserved sequence: A sequence within DNA or protein that is consistent across species or has remained unchanged within the species over its evolutionary period.

Contig maps: The representation of the structure of contiguous regions of the genome (contigs) by specifying overlap relationships among a set of clones.

Contigs: A series of cloning vectors which are ordered in such a way as to have each sequence overlap that of its neighbors. The result is that the assembly of the series provides a contiguous part of a genome.

Diploid: A cell containing two sets of chromosomes.

DNA (deoxyribonucleic acid): A double stranded molecule made of a linear assembly of nucleotides. DNA holds the genetic code for an organism in the arrangement of the bases. The double strand of DNA results from the hydrogen bonds formed between bases when two polynucleotide chains, identical, but running in opposite directions, associate.



DNA polymerase: The enzyme which assembles DNA into a double helix by adding complementary bases to a single strand of DNA. Linkages are formed by adding nucleotides at the 5' hydroxyl group to the phosphate group located on the 3' hydroxyl.

EMBL: The European Molecular Biology Laboratory (<http://www.embl-heidelberg.de>) which is located in Heidelberg Germany.

EMBL Nucleotide Sequence Database: Europe's primary nucleotide sequence resource. Main sources for DNA and RNA sequences are direct submissions from individual researchers, genome sequencing projects and patent applications. The database is produced in collaboration with GenBank and the DNA Database of Japan (DDBJ). Each of the three groups

collects a portion of the total sequence data reported worldwide, and all new and updated database entries are exchanged between the groups on a daily basis.

Endonuclease: An enzyme that cleaves at internal locations within a nucleotide sequence. The enzyme's site of action is generally a sequence of 8 bases. For *E. coli*, treatment with a restriction endonuclease will lead to around 70 fragments. Cleavage of human DNA leads to around 50,000 fragments.

Enzyme: A protein which catalyzes (or speeds the rate of reaction for) biochemical processes, but which does not alter the nature or direction of the reaction.

EST (Expressed Sequence Tag): A partial sequence of a cDNA clone that can be used to identify sites in a gene.

Eukaryote: An organism whose genomic DNA is organized as multiple chromosomes within a separate organelle -- the cell nucleus.

Exon: The region of DNA which encodes proteins. These regions are usually found scattered throughout a given strand of DNA. During transcription of DNA to RNA, the separate exons are joined to form a continuous coding region.

Exonuclease: An enzyme which cleaves nucleotides sequentially starting at the free end of the linear chain of DNA.

FASTA: An alignment program for protein sequences created by Pearson and Lipman in 1988. The program is one of the many heuristic algorithms proposed to speed up sequence comparison. The basic idea is to add a fast prescreen step to locate the highly matching segments between two sequences, and then extend these matching segments to local alignments using more rigorous algorithms such as Smith-Waterman.

Fingerprinting: The process of identifying overlapping regions at the ends of DNA fragments.

FISH: Fluorescence in situ hybridization. A method used to pinpoint the location of a DNA sequence on a chromosome.

Frameshift: Genetic mutation which shifts the reading frame used to translate mRNA (see reading frame).

Functional genomics: The development and application of experimental approaches to assess gene function by making use of the information and reagents provided by structural genomics.

Gene: A section of DNA at a specific position on a particular chromosome that specifies the amino acid sequence for a protein.

Gene expression profiling: Determining specifically which genes are "switched on," with precise definition of the phenotypic trait.

Gene mapping: Determining the relative physical locations of genes on a chromosome. Useful for plant and animal breeding.

GenBank: The NIH genetic sequence database. An annotated collection of all publicly available DNA sequences which is located at <http://www.ncbi.nlm.nih.gov>. There are approximately 2,162,000,000 bases in 3,044,000 sequence records as of December 1998. GenBank is part of the International Nucleotide Sequence Database Collaboration, which is comprised of the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis.

Gene expression: The conversion of the information encoded in a gene to messenger RNA which is in turn converted to protein.

Genetic map (Linkage Map): The linear order of genes on a chromosome of a species. Genetic maps are created by observing the recombination of tagged genetic segments (STSs) during meiosis. The map shows the position of known genes and markers relative to each other, but does not show the specific physical points on the chromosomes.

Genetic mutation: An inheritable alteration in DNA or RNA which results in a change in the structure, sequence, or function of a gene.

Genetic polymorphism: The occurrence of one or more different alleles at the same locus in a one percent or greater of a specific population.

Genome: The total genetic material of a given organism.

Genomics: The mapping, sequencing, and analysis of an organism's genome.

Genomic library: A collection of biomolecules made from DNA fragments of a genome that represent the genetic information of an organism that can be propagated and then systematically screened for particular properties. The DNA may be derived from the genomic DNA of an organism or from DNA copies made from messenger RNA molecules. A computer-based collection of genetic information from these biomolecules can be a virtual genomic library.

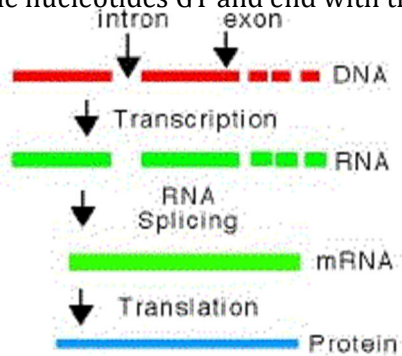
Genotyping: The use of markers to organize the genetic information found in individual DNA samples and to measure the variation between such samples.

Haploid: A cell containing only one set of chromosomes.

Hexaploid: A cell containing three sets of the same chromosomes (i.e. Wheat)

Hybridization: The formation of a double stranded DNA, RNA, or DNA/RNA from two complementary oligonucleotide strands.

Intron: The portion of a DNA sequence which interrupts the protein coding sequences of the gene. Most introns begin with the nucleotides GT and end with the nucleotides AG.



In vitro: Outside a living organism, usually in a test tube.

In vivo: Inside a living organism.

Kilobase (kb): A length of DNA equal to 1,000 nucleotides.

Linkage analysis: The process used to study genotype variations between affected and healthy individuals wherein specific regions of the genome that may be inherited with, or "linked" to, disease are determined.

Linkage Disequilibrium (LD): In population genetics, LD is the association of alleles at two or more loci on same or different chromosome that is greater than random association. Populations where combinations of alleles or genotypes can be found in the expected proportions are said to be in linkage equilibrium.

Linkage map: A map which displays the relative positions of genetic loci on a chromosome.

Loci: The location of a gene or other marker on the surface of a chromosome. The use of locus is sometimes restricted to mean regions of DNA that are expressed.

Mapping: The process of determining the positions of genes and the distances between them on a chromosome. This is accomplished by identifying unique genome markers (ESTs, STSs, etc.) and localizing these to specific sites on the chromosome. There are three types of DNA maps: physical maps, genetic maps, and cytogenetic maps. The types of markers identified differentiate the map produced.

Marker: A physical location on a chromosome which can be reliably monitored during replication and inheritance. Markers on the Human Transcript Map are all STSs.

Microarray: DNA which has been anchored to a chip as an array of microscopic dots, each one of which represents a gene. Messenger RNA which encodes for known proteins is added and will hybridize with its complementary DNA on the chip. The result will be a fluorescent signal indicating that the specific gene has been activated.

Microsatellite: a specific sequence of DNA bases or nucleotides which contains mono, di, tri, or tetra tandem repeats. For example

GGGGGGGG is a (G)8

ACACACAC is referred to as a (AC)4

ATCATCACTACTACT would be referred to as (ATC)5

ATCTATCT would be referred to as (ATCT)2

Microsatellites also are called simple sequence repeats (SSR), short tandem repeats (STR), or variable number tandem repeats (VNTR).

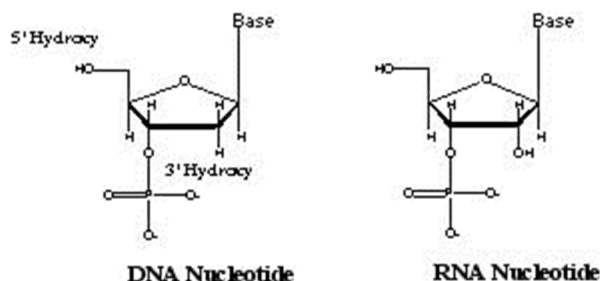
Motifs: A pattern of DNA sequence that is similar for genes of similar function. Also a pattern for protein primary structure (sequence motifs) and tertiary structure that is the same across proteins of similar families.

mRNA (messenger RNA): RNA that is used as the template for protein synthesis. The first codon in a messenger RNA sequence is almost always AUG

NCBI: The National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>), a division of the NIH, is the home of the BLAST and Entrez servers.

NCGR: The National Center for Genome Resources (<http://www.ncgr.org>).

Nucleotide (nt): A molecule which contains three components: a sugar (deoxyribose in DNA, ribose in RNA), a phosphate group, and a heterocyclic base.



Oligos (Oligonucleotides): A chain of nucleotides.

Pairwise alignment: In the first step, two sequences are padded by gaps so that they are the same length and so that they display the maximum similarity on a residue to residue basis. An optimal Pairwise Alignment is an alignment which has the maximum amount of similarity with the minimum number of residue 'substitutions'.

PCR (polymerase chain reaction; in vitro DNA amplification): The laboratory technique for duplicating (or replicating) DNA using the bacterium *Thermus aquaticus*, a heat stable bacterium from the hot springs of Yellowstone. As with the polymerase reaction that occurs in cells, there are three stages of a PCR process: separation of the DNA double helix, addition of the primer to the section of the DNA strand which is to be copied, and synthesis of the new DNA. Since PCR is run in a single reaction vessel, the reactor contains all of the components necessary for replication: the target DNA, nucleotides, the primer, and the bacterial DNA polymerase. PCR is initiated by heating the reaction vessel to 90° which causes the DNA chains to separate. The temperature is lowered to 55° to allow the primers to bind to the section of the DNA that they were designed to recognize. Replication is then initiated by heating the vessel to 75°. The process is repeated until the quantity of new DNA desired is obtained. Thirty cycles of PCR can produce over 1 million copies of a target DNA.

Physical map: The physical locations (and order) on chromosomes of identifiable areas of DNA sequences such as restriction sites, genes, coding regions, etc. Physical maps are used when searching for disease genes by positional cloning strategies and for DNA sequencing.

Polymerase: The process of copying DNA in each chromosome during cell division. In the first step the two DNA chains of the double helix unwind and separate into separate strands. Each strand then serves as a template for the DNA polymerase to make a copy of each strand starting at the 3' end of the chain.

Polymorphic marker: A length of DNA that displays population-based variability so that its inheritance can be followed.

Polymorphism: Individual differences in DNA. Single nucleotide polymorphism (the difference of one nucleotide in a DNA strand) is currently of interest to a number of companies.

Quantitative trait locus (QTL): A locus, or location, on a chromosome for genes that govern a measurable trait with continuous variation, such as height, weight, or color intensity. The presence of a QTL is inferred from genetic mapping, where the total variation is partitioned into components linked to a number of discrete chromosome regions.

QTL mapping: QTLs are detected through QTL mapping populations produced from crossing two inbred lines. The first offspring generation (the F₁) is uniformly heterozygous. However, in the second generation (the F₂) the parental alleles segregate and most chromosomes recombine. Genes and genetic markers that are close together on a chromosome will tend to co-segregate in the F₂ (the

same allele combinations that occurred in one of the parents will tend to occur together in the offspring). The closer together are two markers or genes on a chromosome, the less likely the parental alleles at the two loci will be split up in the F_2 as a result of recombination. This will lead to a statistical association between a gene segregating for alleles that have a measurable difference in their effect on a quantitative trait and segregating alleles at closely linked marker loci. QTLs can thus be localized to specific chromosomal segments if the trait is measured in all the F_2 offspring and if all of these offspring are genotyped at hundreds of genetic markers covering the whole genome.

Reading frame (also open reading frame): The stretch of triplet sequence of DNA that encodes a protein. The reading frame is designated by the initiation or start codon and is terminated by a stop codon. As an example, the sequence CAGAUGAGGUCAGGCAUA potentially can be translated as follows:

Position 1 CAGAUGAGGUCAGGCAUA
gln met arg ser Gly ile

Position 2 C AGAUGAGGUCAGGCAUA
arg trp gly Gln ala

Position 3 CA GAUGAGGUCAGGCAUA
asp glu val Arg his

DNA (through RNA) uses a triplet code to specify the amino acid for a given protein. As can be seen above, a given strand of DNA has three possible starting points (position [or reading frame] one, two, or three). Since both strands of DNA can be translated into RNA and then into protein, a sequence of double helical DNA can specify six different reading frames.

Recombinant Inbred Lines (RIL): RILs are the highly inbred progeny of a segregating population or QTL mapping resource. Two parental inbred lines are crossed, the F_1 are intermated (or selfed) to form an F_2 generation. Numerous individuals from the segregating F_2 generation then serve as the founders of RILs. Each subsequent generation of a given RIL is formed by selfing in the previous generation and with single seed descent. In this manner each RIL, after several generations, will contain two identical copies of each chromosome, with most of them being recombinant.

Scaffold: A series of contigs that are in the correct order, but are not connected in one continuous length.

Sequencing: Determining the order of nucleotides in a gene or the order of amino acids in a protein.

Shotgun method: A method that uses enzymes to cut DNA into hundreds (or thousands) of random bits which are then reassembled by computer so it looks like the original genome. The Human Genome Project shotgun approach is applied to cloned DNA fragments that already have been mapped so that it is known exactly where they are located on the genome, making assembly easier and much less prone to error.

Single nucleotide polymorphism (SNP): The most common type of DNA sequence variation. An SNP is a change in a single base pair at a particular position along the DNA strand. When an SNP occurs, the gene's function may change, as seen in the development of bacterial resistance to antibiotics or of cancer in humans.

Transcriptome: The complete collection of RNA molecules transcribed (or processed) from the DNA of a cell.

Transcription: The process of copying a strand of DNA to yield a complementary strand of RNA

Translation: The process of sequentially converting the codons on mRNA into amino acids which are then linked to form a protein.

Acknowledgements

This document was prepared with the support of The Peanut Foundation, Alexandria VA

International PGI, Executive Committee (July, 2017)

Members

Scott Jackson, UGA (Chair)
 Corley Holbrook, USDA-ARS (Co-chair)
 Rajeev Varshney, ICRISAT (India)
 Raymond Schnell, MARS, Inc.
 Baozhu Guo, USDA-ARS
 Steven Cannon, USDA-ARS
 Tom Stalker, NCSU
 Lutz Froenicke, UC-Davis
 Barry Tillman, University Florida
 Boshou Liao, CAAS, (China)
 Ethy Cannon, Iowa State Univ.
 Xingyou Zhang, HAAS (China)
 Graeme Wright (PCA (Australia)
 Ran Hovav, ARS TVC (Israel)
 Xingjun Wang, SAAS (China)

Peggy Ozias-Akins, UGA (Co-chair)
 Richard Michelmore, UC-Davis
 Howard Valentine, TPF
 Victor Nwosu, MARS, Inc.
 Brian Scheffler, USDA-ARS
 Andrew Farmer, NCGR
 Xin Liu, BGI
 Kelly Chamberlin, USDA-ARS
 Haile Desmae, ICRISAT-WCA
 David Bertioli, U Brasila (Brazil)
 Soraya Bertioli, EMBRAPA
 Mark Burow, TAMU
 Richard Wilson, OBC
 Guillermo Seijo, IBN (Argentina)
 Sachiko Isobe, KDRI (Japan)

Ex Officio

Roy Scott, USDA-ARS-ONP
 Jean-Marcel Ribuat, Director, GCP
 Howard Shapiro, MARS, Inc.
 Steve Brown, Executive Director, TPF
 Fuhe Luo, Vice Chairman CPPCC & CAPD CC

Maricio Lopez, President, EMBRAPA
 David Hoisington, UGA
 Xun Xu, Deputy Director, BGI-Shenzhen
 T. Radhakrishnan, Director, DGR, ICAR
 Jeff Ehlers, Program Officer, Gates Foundation

Annual Report Writing Team

George Birdsong
 Dan Ward
 Victor Nwosu
 Darlene Cowart

•

Jim Elder
 John Takash
 Steve Brown
 Howard Valentine

Editors

H. Thomas Stalker and
 Corley C. Holbrook

Scientific Authors

Josh Clevenger, Peggy Ozias Akins,
 Soraya Leal-Bertioli, David Bertioli
 Scott Jackson, Steven Cannon,
 Ethy Cannon, Mark Burow
 Corley Holbrook

