Genetic Mapping, Germplasm Evaluation and Development of Genomic Tools for Mango to Accelerate Breeding of Improved Cultivars

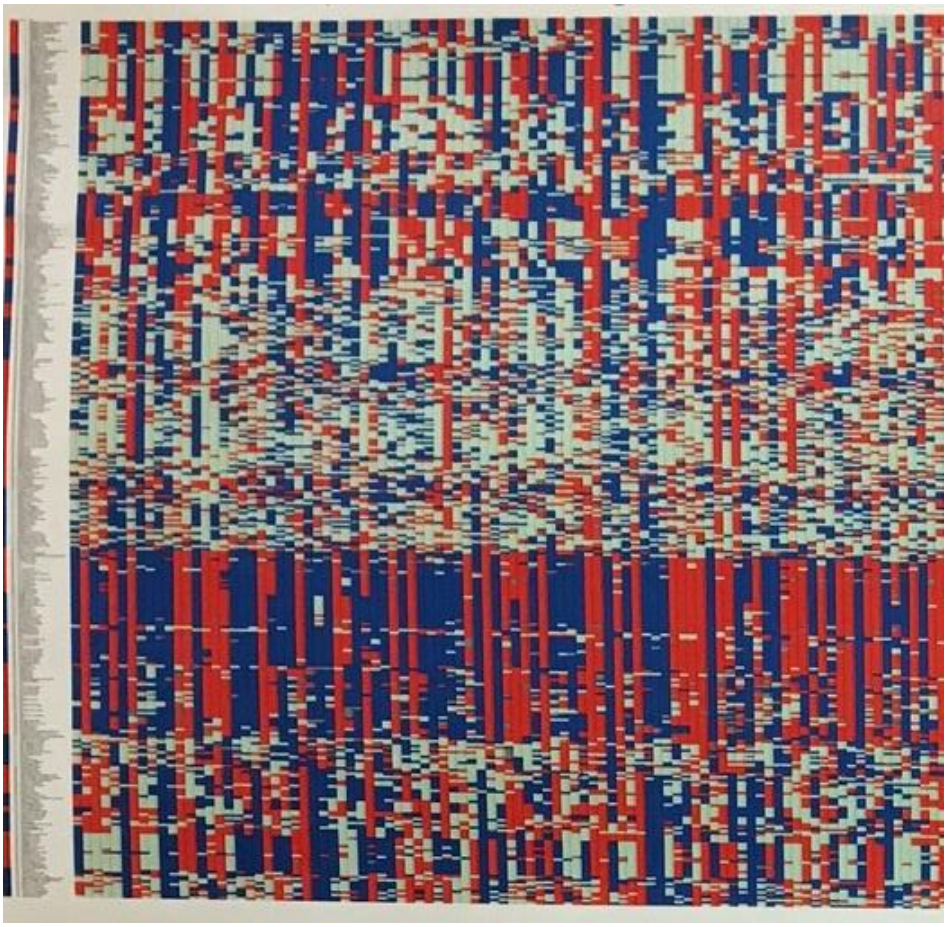
Principal Investigator: David N. Kuhn

Research Molecular Biologist

USDA-ARS Subtropical Horticulture Research Station

Miami, FL

July 2017



Abstract

Genomics is the study of the complete genome of an organism rather than individual genes, traits or processes. The advantage of a genomics approach to mango is that it allows the use of all of the extensive genomic information for other plants such as Arabidopsis, rice, maize, grape, etc. The reason that the genomic information from other plants is so useful is that all plants share a basic set of genes, metabolic pathways, transcriptional regulators, hormone-mediated responses and stress responses. Knowing something about these shared attributes in one plant means that it is likely that similar genes, pathways, etc. will also be found in mango. To be able to leverage this ever increasing amount of plant genomic knowledge, we must understand more about mango through: building a molecular genetic map from single nucleotide polymorphism (SNP) markers; associating horticultural traits with regions of the map and individual SNP markers; and gernotyping mango germplasm with mapped SNP markers to estimate genetic diversity and identify new parents for breeding and selection programs.

Mango (*Mangifera indica*) is an economically and nutritionally important tropical/subtropical tree fruit crop. Most of the current commercial cultivars are selections rather than the products of breeding programs. To improve the efficiency of mango breeding, molecular markers have been used to create a consensus genetic map that identifies all 20 linkage groups in seven mapping populations. Polyembryony is an important mango trait, used for clonal propagation of cultivars and rootstocks. In polyembryonic mango cultivars, in addition to a zygotic embryo, several apomictic embryos develop from maternal tissue surrounding the fertilized egg cell. This trait has been associated with linkage group 8 in our consensus genetic map and has been validated in two of the seven mapping populations. In addition, we have observed a significant association between trait and single nucleotide polymorphism (SNP) markers for the vegetative trait of branch habit and the fruit traits of bloom, ground skin color, blush intensity, beak shape, and pulp color.

Assessing the genetic diversity and relatedness of available mango germplasm accessions is essential to identification of genetically distant parents with favorable horticultural traits to produce hybrid populations for selection of improved cultivars. From germplasm collections from Australia, Senegal, Thailand and the United States, 1911 individuals of *M. indica* and other species have been genotyped with 384 SNP markers. Analysis of the more than 730,000 genotypic data points indicates that essentially all the genetic diversity available for mango has been captured in the current germplasm collections and that genetic diversity in the current commercial cultivars is very limited. It also identifies significant mislabeling and misidentification in these germplasm collections and among the parents used in breeding and selection programs. Horticulturalists should use this data to select more diverse parents for breeding and selection programs and to make the identification of improved cultivars more efficient.

Introduction

Mango (*Mangifera indica*) is one of the most important fruit crops of the world due to its large fruit with a soft, sweet pulp. World mango production is fifth among all fruits, and second only to banana among tropical fruits (Galán Saúco 2015). A subtropical group in the Indian sub-

continent is characterized by monoembryonic seed and a tropical group in the south-east-Asia region is characterized by polyembryonic seed (Mukherjee and Litz 2009)

Mango has been widely cultivated in India and Southeast Asia for thousands of years. In the 15th and 16th centuries, Portuguese and Spanish traders spread mango to other tropical and subtropical regions of the world. (Litz 2009). Early in the 20th century, cultivars from the Indian and Asian regions were combined in a new center of mango development in Florida, where many cultivars were selected and disseminated. These cultivars, selected for milder taste and aroma, colorful skin and larger fruit size, are still the major cultivars used today in international trade.

Mango is now grown throughout the sub-tropical and tropical world in over 100 countries with a total fruit production of 43.3 million tons in 2013 (Galán Saúco 2015). The majority (76%) of world production comes from Asia, with the Americas (12%) and Africa (11.8%) the second and third largest producers. India is the largest producer, growing over 18 million tons (MT) primarily for domestic consumption, followed by China (4.5 MT) Thailand (3.1 MT), Indonesia (2.6 MT) and Mexico (1.9 MT) (Galán Saúco 2015). Although Mexico is fifth in production it is first in export to the USA, which is 43% of the global import market.

Around the world there are hundreds and possibly thousands of different mango cultivars and selections, most of which are only grown and marketed locally. Relatively few cultivars are traded internationally due to the highly specific requirements for cultivars with favorable color, storage and shipping traits.

To date the development of genetic and genomic resources in mango have been limited and have not greatly contributed to mango breeding around the world. An early, very limited genetic map of mango produced by Kashkush et al. (2001) was not sufficiently resolved to be useful for marker assisted selection (MAS) or trait association to markers. Recently, a high resolution map of mango has been produced by Luo, Shu et al. (2016) that may prove more useful. Several transcriptomes from different mango tissues have been produced (Pandit, Kulkarni et al. 2010, Azim, Khan et al. 2014, Luria, Sela et al. 2014, Wu, Jia et al. 2014, Dautt-Castro, Ochoa-Leyva et al. 2015, Sherman, Rubinstein et al. 2015). In 2016, Kuhn, Dillon et al. (2016) identified ~400,000 single nucleotide polymorphism (SNP) markers using a reference transcriptome from 'Tommy Atkins' and expressed RNA from 17 genetically diverse cultivars. The genetic diversity of mango has been explored by different groups with a variety of markers, who all found a narrow genetic basis among the commercial cultivars grown and traded internationally (Schnell, Brown et al. 2006, Dillon, Bally et al. 2013, Sherman, Rubinstein et al. 2015). An increase in the number of unbiased markers and a highly resolved genetic map are essential molecular tools for mango breeders if the power of genomics is to drive future progress of breeding for improved mango cultivars.

The current improved commercial cultivars have typically been selected from open pollinated seedling progeny and then vegetatively propagated to maintain genetic uniformity (Bally, Lu et al. 2009). The continual demand for new and improved cultivars with superior production and quality traits is a challenge for breeders relying on traditional breeding techniques. Factors that limit progress in traditional fruit tree breeding are the long juvenile phase, long generation time, and large resource requirements in field area and personnel for maintaining and evaluating hybrid populations. In addition to these restraints, mango breeders are faced with high

heterozygosity, polyembryony, low crossing rates (0.1%) from high numbers of flowers per panicle, a very high level of fruitlet drop, and only a single seed per flower resulting in a low number of fruit (0.1% of flowers), all of which makes the task of active manual crosses challenging (Bally, Lu et al. 2009). There is also little knowledge of the heritability of most of the important horticultural traits in mango (Schnell, Brown et al. 2006). Finally, the lack of genotypic and phenotypic diversity among the current commercial cultivars may reduce breeding efficiency if they are continued to be used as parents in breeding programs. Adoption of molecular genomic tools has the potential to estimate genetic diversity of potential parents, identify markers associated with important horticultural traits and, in general, improve the efficiency of mango breeding programs.

In this project, we generated a mango consensus genetic map, a valuable tool that can be used to improve the efficiency and overcome the challenges facing mango breeding programs. We used the genetic map to identify markers and regions of the genome that are associated with important horticultural traits such as embryo type, branch habit, bloom, ground skin color, blush intensity, beak shape, and pulp color. We also used 384 SNP markers to genotype all accessions from 10 domestic and international germplasm collections to get an accurate estimate of the available mango germplasm, to identify offtypes and mislabeling in the collections, and to provide genetic evidence to assist in distinguishing the numerous species of mango.

OBJECTIVES

1. The production of a high resolution genetic map for mango.

- Genotype 775 individuals from seven mapping populations with 1054 SNP genetic markers.
- Produce a high resolution consensus genetic map with 20 linkage groups.
- Associate qualitative horticultural traits with map regions and SNP markers.
- Mapping Populations (female parent first):

2. Screening with genetic markers of all mango germplasm to identify trees with favorable traits by genotype to use in future breeding crosses.

- Select a subset of 384 SNP markers from mapped markers evenly distributed across the mango genetic map including SNP markers associated with horticultural traits.
- Genotype 1911 individuals from worldwide germplasm collections with 384 SNP genetic markers (>730,000 genotypic data points).
- Estimate genetic diversity in germplasm collections from genotype data.

MATERIALS AND METHODS

Mapping populations:

Seven mapping populations were used to make the consensus map (Table 1). The four mapping populations from Australia share a common paternal parent, Kensington Pride (KP). In addition, the cultivar NMBP1243, the maternal parent of one of the mapping populations, is a progeny of the Irwin (I) x KP population. The Brazilian population (Haden (H) x Tommy Atkins (TA) share

both parents with the self pollinated populations of H and TA from the Subtropical Horticulture Research Station (SHRS). The TA self pollinated population was generated by germinating and genotyping fruit from a commercial grove planted with only TA. The H self pollinated population was generated by germinating and genotyping fruit from an isolated tree at SHRS.

Table 1. Number of progeny and the sources of seven hybrid mapping populations used to create the consensus genetic map. Populations were named maternal parent x paternal parent.

Population Name	Number of individuals	Source of Population
Tommy Atkins x Tommy Atkins (TA x TA) (Self-pollinated))	60	USDA-ARS, SHRS, USA[1]
Tommy Atkins x Kensington Pride (TA x KP)	100	DAFQ, Australia[2]
Haden x Tommy Atkins (H x TA)	225	Embrapa, Brazil[3]
Haden x Haden (H x H) (Self-pollinated)	40	USDA-ARS, SHRS,USA[1]
Irwin x Kensington Pride (I x KP)	180	DAFQ, Australia[2]
NMBP1243 x Kensington Pride (NMBP1243 x KP)	100	DAFQ, Australia[2]
Creeper x Kensington Pride (Cr x KP)	70	DAFQ, Australia[2]

[1] United States Department of Agriculture-Agricultural Research Service, Subtropical Horticulture Research Station, United States of America
[2] Department of Agriculture and Fisheries, Queensland, Australia
[3] Brazilian Agricultural Research Corporation (Embrapa), Pernambuco, Brazil

Germplasm Collections:

Leaves from each individual tree were collected into labeled paper bags. International samples were sent by express delivery with paper bags or envelopes in a Styrofoam cooler with ice packs in the bottom.

Table 2. Mango germplasm collections genotyped.

Population	Station	Location	Number of Individuals
Germplasm	SHRS ARS	Miami, FL	210
Germplasm	Fairchild Tropical Botanical Garden	Miami, FL	109
Polycross seedlings	SHRS ARS	Miami, FL	386
Germplasm	Zill private collection	Boynton Beach, FL	48
Open pollinated seedlings	Zill private collection	Boynton Beach, FL	56
Germplasm	Fruit and Spice Park	Homestead, FL	171
Germplasm	SRS and WRS	Mareeba, Australia	685
Mangifera laurina hybrids	SRS and WRS	Mareeba, Australia	84
Germplasm		Senegal	63

Germplasm		Thailand	40
Germplasm and other species	Florida International University (E. Warschefsky)	Miami, FL	59
		Total	**1911**

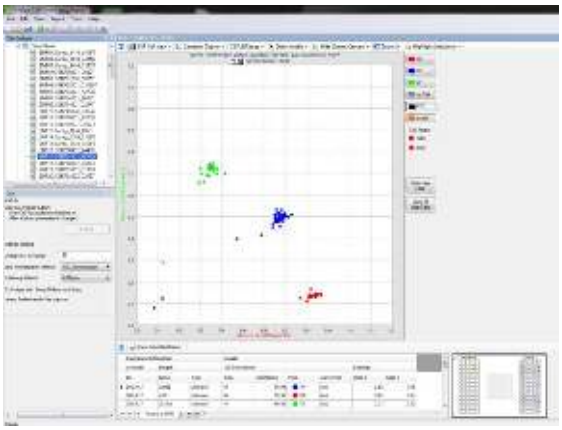
SNP containing sequences:

SNP containing sequences came from three different sources: Department of Agriculture and Fisheries, Queensland (DAFQ, Australia), SHRS, USA and the Agriculture Research Organization (ARO), Israel (Table 2). The SHRS SNP markers were identified as described in Kuhn et al. (2016). The ARO SNP markers were identified as described in Sherman et al. (2015). The DAFQ SNP markers were identified from sequence data described in Hoang et al. (2015).

DNA Isolation:

DNA for genotyping was isolated from the leaves of individual progeny in the mapping populations as in Kuhn et al. (2016). Once isolated the DNA was quantified by fluorescence on a fluorescence plate reader (BioMark, Inc.) and normalized to 10ng/uL on a liquid handling robot (Hamilton, Inc., Reno, NV, USA).

SNP Assays:



All 1054 SNP assays were produced from SNP containing sequences by Fluidigm (South San Francisco, CA, USA) and assayed on a Fluidigm EP-1 platform. Genotyping is done on the Fluidigm EP-1, a high throughput microfluidics SNP assay platform. All individuals are genotyped 96 markers at a time.

Typical Fluidigm EP-1 output for genotype of a population at one SNP marker. Green is homozygous for Hex labeled allele, Red is homozygous for Fam labeled allele, and Blue is heterozygous.

Genetic mapping

Two mapping programs, JoinMap4 (Kyazma B.V.®, Wageningen, Netherlands) and OneMap (Margarido, Souza et al. 2007) were used to create genetic maps for each of the seven mapping populations (Table1).

Germplasm genotype analysis

SNP genotypes from germplasm accessions were produced as described above. Genotypes were coded as 1 (homozygous allele 1, blue), 2 (homozygous allele 2, orange), 3 (heterozygous, green) and 0 (missing data). Pairwise distance matrices were calculated with DistClust using a Hamming distance algorithm. Groupings were defined using k_medoid analysis. Grouped

accessions were exported into Excel to reapply complete accession names and color code the genotypes for each accession. In the Excel spreadsheet, rows are accessions, columns are SNP markers.

Trait Association

Phenotype data for 14 qualitative traits were available for TA x KP, Cr x KP and I x KP populations. In all cases KP was the pollen donor as it is polyembryonic. The qualitative traits measured were: stage of fruit ripeness, fruit shape, ground skin color, blush color, blush intensity, bloom, stem end shape, cleavage, beak shape, pulp color, embryo type, flavor, branch habit, tree vigor, beak shape and cleavage (Table 3). Embryo type was measured by visual inspection of the seed without seed coat from the F_1 mapping population parent (Aron, Czosnek et al. 1998).

Of the 14 traits, the twelve fruit traits were assessed on a sample of ten randomly picked at fruit maturity from each individual genotype within the three mapping populations. Fruit were ripened at 26 C° and assessed at the eating ripe stage (Holmes, Hofman et al. 2010) using the criteria detailed in Table 3.

Associating traits with the mapped SNP markers was done using MapQTL6 (Kyazma B.V.®, Wageningen, Netherlands) using Cross Pollinated (CP) for population type and Interval Mapping (IM) for association statistic. All calculation parameters were set to MapQTL6 defaults. Global thresholds were calculated as described in MapQTL6 (permutation tests of 10,000 rounds) and only traits that showed higher association probabilities than the global threshold were considered to be significant.

Results

The manuscript of the mango genetic map work and trait association, “Genetic map of mango: a tool for mango breeding”, was accepted for publication in the journal Frontiers in Plant Science (Front. Plant Sci., 20 April 2017 | https://doi.org/10.3389/fpls.2017.00577) (Kuhn, Bally et al. 2017). NMB was acknowledged as follows:

Funding: DK, AG, JR were funded by USDA-ARS CRIS #6631-21000-022-00D and the **National Mango Board NACA#58-6038-5-001.**

Acknowledgments**: Special thanks to Leo Ortega and the National Mango Board (USA) for their exceptional support in funding and encouraging this research.**

Consensus genetic map

To include all markers in the consensus map, we employed the strategy detailed in Materials and Methods, using the strengths of both JoinMap4 and OneMap. We produced a consensus map with 726 SNP markers distributed across 20 LGs shown in Figure 1.

Figure 1. The consensus genetic map of mango. Vertical lines represent linkage groups. Horizontal lines crossing the vertical lines depict the name and position in cM of SNP markers on the linkage group.

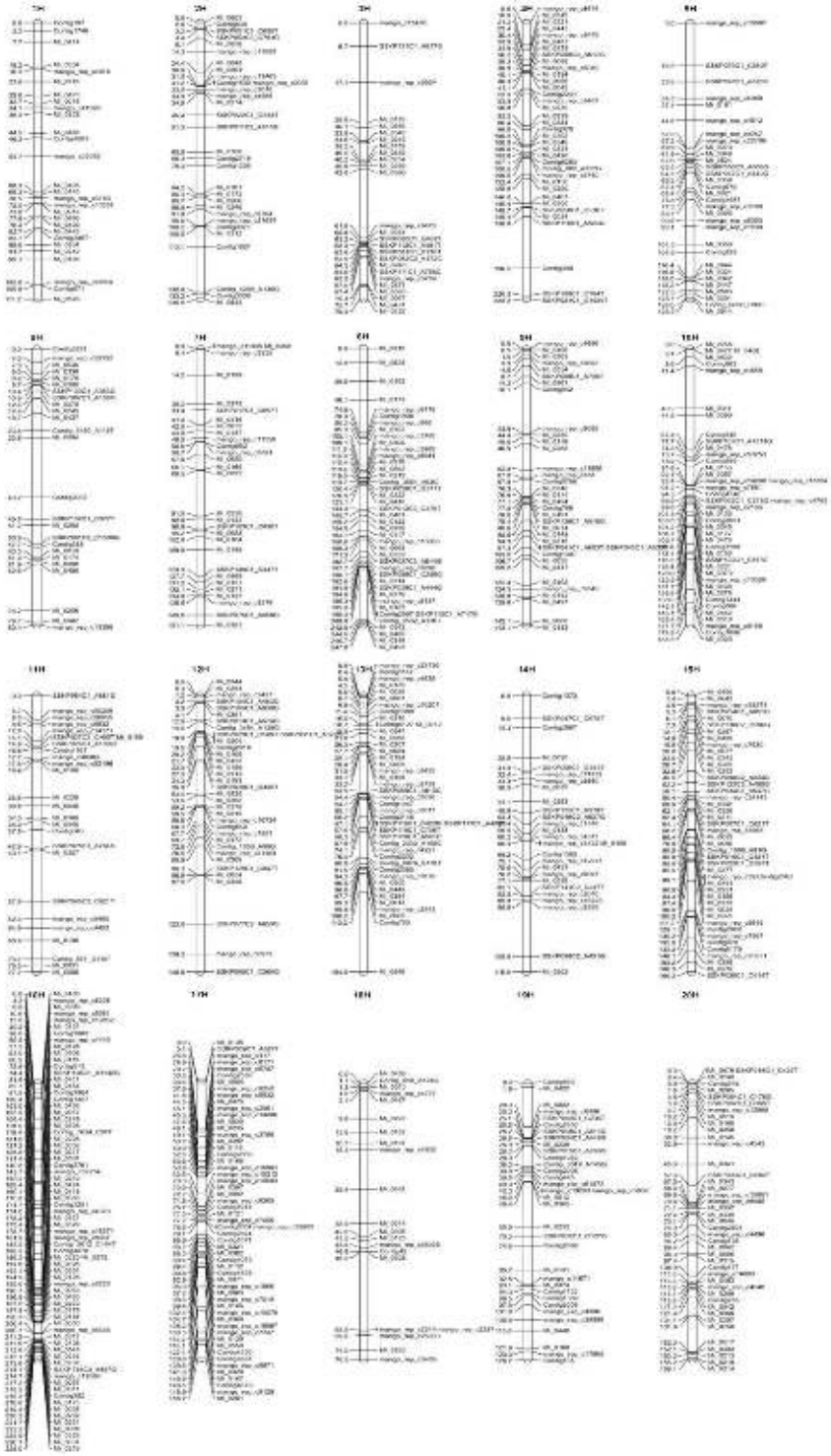


Table 3 shows the calculated length in centimorgans (cM) and the number of markers for each of the 20 LGs. Linkage group 8 was the longest at 247.8 cM and LG 16 had the greatest number of markers at 71. Average distance between markers for each LG is also shown in Table 6 and the overall average distance between markers was 4.095 cM. Greatest distance between markers as 44.775 cM on LG 13 and shortest distance was 0.001 cM on LG 8 and 13 not including identical markers (0.000 cM distance).

Table 3. Consensus Map Statistics. Summary of the final consensus linkage map containing 726 markers across 20 linkage groups.

LG	Number of markers per linkage group	Length of each linkage group (cM)	Ave Distance between markers (cM)	Max Distance between markers (cM)	Min Distance between markers (cM)
1	28	111.2	4.120	14.629	0.058
2	31	135.6	4.520	22.779	0.053
3	26	79.4	3.177	19.760	0.081
4	36	223.2	6.377	41.634	0.073
5	31	126.3	4.209	19.450	0.185
6	25	80.4	3.351	17.440	0.166
7	29	151.1	5.397	24.997	0.002
8	42	247.8	6.045	32.912	0.001
9	35	143.1	4.210	25.709	0.012
10	42	186.5	4.549	28.816	0.001
11	26	77.2	3.090	14.414	0.002
12	35	148.8	4.378	26.070	0.000
13	43	154.9	3.689	44.775	0.001
14	27	114.9	4.422	22.593	0.020
15	45	166.2	3.778	17.991	0.000
16	71	228.0	3.257	17.854	0.000
17	56	156.7	2.849	26.688	0.000
18	21	76.5	3.826	21.565	0.002
19	34	126.7	3.839	20.472	0.000
20	43	156.1	3.716	20.124	0.023
Total	**726**	**2890.6**			
Min	**21**	**76.5**			
Max	**71**	**247.8**			
Ave	**36.3**	**144.5**			

Associating Qualitative Traits with the Map

Qualitative phenotypic data were available for three of the mapping populations (TA x KP, I x KP, and Cr x KP). Interval mapping testing using MapQTL found seven of the 14 qualitative traits used in the association study had significant LOD scores in at least one of the populations. Table 7 shows the seven qualitative traits with significant LOD scores and their position on the map associated with the trait. Reported LOD scores are all above the thresholds determined by permutation tests for the trait in the respective population.

Embryo type was the only trait to have significant LOD scores at the same marker (Mi_0173) across two different populations (Figure 2). Marker Mi_0173 was unable to be mapped in the I x KP population, which prevented testing for a significant signal for embryo type in that population. For trait association, only genotype data from mapped markers in the population were used to ensure that the phasing specific to the population was correct.

Bloom, pulp color and branch habit traits showed significant association to markers in two different populations. The marker association was on different LGs in each population (Table 4). For example, the bloom trait showed a significant association to a marker on LG 9 in I x KP and on LG 13 in TA x KP. The ground skin color, blush intensity and beak shape traits showed a significant association to markers on a single LG in only one population (Table 4).

Table 4. Trait Association in Three Mapping Populations. Abbreviations: linkage group (LG), Tommy Atkins x Kensington Pride (TA x KP), Creeper x Kensington Pride (Cr x KP), Irwin x Kensington Pride (I x KP), likelihood of the odds (LOD).

Trait	LG	Marker	Position (cM)	TA x KP LOD	Cr x KP LOD	I x KP LOD
Embryo Type	8	Mi_0173	46.1	4.96	8.82	
	8	mango_rep_c6716	74.8		7.70	
	8	Contig1936	78.3		7.40	
	8	mango_rep_c886	80.2		7.23	
	8	Mi_0102	85.3		6.65	
Ground Skin Colour	17	Mi_0135	0.0	5.61		
	17	SSKP009C1_A627T	0.1	5.61		
Blush Intensity	20	Mi_0450	19.2			4.62
	20	Mi_0145	30.8			5.83
	20	mango_rep_c4542	33.9			6.17
	20	Mi_0341	45.6			6.65
	20	SSKP003C1_C682T	57.6			5.99
	20	Mi_0343	67.5			5.75
	20	Mi_0277	68.6			5.69
	20	mango_rep_c15051	69.6			5.62
	20	mango_rep_c8905	70.4			5.60
	20	Mi_0357	71.1			5.57
	20	Mi_0330	72.4			5.49
	20	Mi_0046	73.1			5.43

	20	Contig2601	74.0		5.33
Bloom	13	Contig1142	0.4	5.80	
	9	Mi_0417	109.2		4.86
	9	Mi_0402	122.4		8.05
	9	mango_rep_c9549	124.5		7.91
	9	Mi_0142	128.8		7.14
	9	Mi_0497	129.6		7.03
Beak Shape	11	mango_c48384	17.7	6.16	
	11	mango_rep_c52196	17.8	6.16	
Pulp Colour	16	Mi_0217	125.8	5.18	
	13	Mi_0029	5.6		4.36
Branch Habit	8	Mi_0192	29.6	4.90	
	16	Contig3904	97.5		4.48
	16	Contig1327	100.4		4.42

Genotyping Mango Germplasm

Figure 2. Grouped mango germplasm genotypes.

Leaves were collected from 1911 individual trees from 10 germplasm collections (Table 3). Samples were from 31 *Mangifera* species, three different groups of interspecific hybrids (*M. laurina* x *M. indica*), a polycross experiment involving open pollinated progeny from six mango cultivars, commercial cultivars, and named accessions from germplasm collections. The dataset consisting of 733,824 genotype data points was edited and individuals and markers with greater than 5% missing data were removed. The entire color coded dataset is shown in Figure 2.

A more detailed analysis of the data would be difficult to present in this final report format. Instead a visual representation of the entire dataset has been provided to allow an intuitive understanding of the genotyping results. Banding and patterns in Figure 2 represent clustering of germplasm accessions by genotype. Bands that are predominantly blue and orange represent highly homozygous individuals. The second band from the top in Figure 2 contains 142 Southeast Asian cultivars such as Nam Doc Mai, Carabao, Nam Tam Teen, and Okrong. The band below the middle of Figure 2 with clearly delineated columns contains 86 supposed hybrids where the maternal parent was polyembryonic. They share

identical genotypes for 263 of 267 SNP markers making them clones of the maternal parent rather than hybrids. The lower band that is predominantly blue and orange contains 119 accessions and encompasses 27 of the 31 *Mangifera* species in the study. The 267 SNP markers in the edited dataset produced genotypes with less than 5% missing data for all 31 *Mangifera* species as well as closely related individuals from other genera such as *Bouea macrophylla*. This suggests that all of the potential genetic diversity in the genus *Mangifera* has been captured by the SNP marker set. The dataset was further divided into subsets: species only, hybrids only, Miami germplasm plus polycross progeny and worldwide germplasm only. In the species only subset, group numbers in the k_medoid analysis were altered to allow resolution of all possible species (Figure 3). Numerous accessions with different species names appeared to have identical genotypes, calling into question the accuracy of the classification of these species. Interestingly, there were often larger differences in genotype between *M. indica* accessions than between different species. This also calls into question the accuracy of classification of individuals as *M. indica* based solely on phenotypic characters.

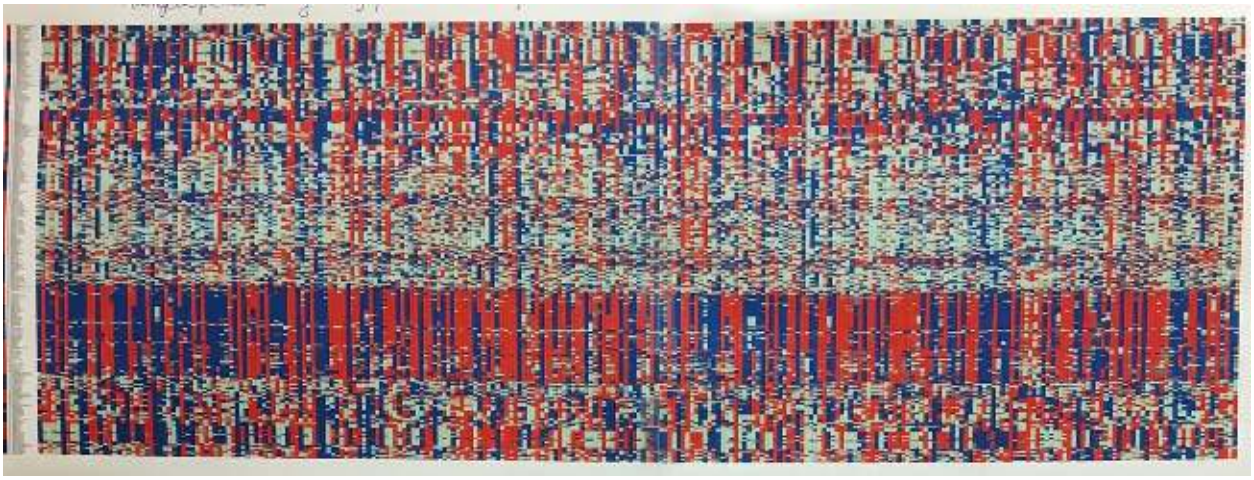


Figure 3. Species only subset of germplasm genotype dataset.

The 267 SNP markers were able to easily resolve the identity of the most common commercial cultivars. In addition, the paternal parent of the Florida cultivars could be inferred. The conclusion of the genotyping project is that the current commercial cultivars are very closely related, which may explain why so little progress has been made in identifying improved varieties. The varieties released from the Australian breeding program over the last 20 years are almost identical genotypically, due to the similarity of the parents used in the crosses. This similarity or even identity of genotypes is also seen for favorite cultivars of particular cultures, such as Carabao for people from the Phillipines. All accessions of Carabao from germplasm collections from around the world were essentially genetically identical, but easily distinguishable from other *M. indica* varieties. This suggests that human selection has played a significant role in defining the what we think of as a mango.

DISCUSSION AND CONCLUSIONS

A genetic map of mango from SNP markers

Genetic maps are built from genotyping the progeny of crosses (mapping populations) between trees chosen due to their interesting and differing phenotypes (traits). For example, a female parent that produces green mangos and is anthracnose resistant is crossed with a male parent that produces red mangos and is anthracnose susceptible. The simple breeding goal is to identify a tree among the progeny that produces red mangos and is anthracnose resistant. The more advanced molecular breeding goal is to identify the regions of the mango genome that regulate fruit color and anthracnose resistance. To do this, a genetic map is needed that will give us more information about the distance between genes on each linkage group (chromosome), which genes are associated with the traits and which alleles of the genes are inherited from the parental genomes.

Genetic maps can only be made from a cross between two known parents. Genotyping open pollinated seed from a single female parent tree will not produce a map. Doing the crosses, identifying the true hybrids, growing the trees until mature and collecting the phenotypic data takes at least seven years. Thus, we have sought pre-existing mapping populations, described in Table 1, from our international collaborators to be able to produce a consensus map and associate horticultural traits to it.

The 1054 SNP markers used provided sufficient resolution to identify 20 linkage groups for the consensus map, representing the 20 chromosomes in the haploid mango genome. The SNP markers were highly reproducible and reliable with genotyping of all mapping populations individuals with less than 5% missing data. This allowed us to use seven mapping populations to produce a consensus map, which increased the number of mapped markers overall. Having more parents increases the chances of at least one parent in one mapping population being heterozygous for the SNP marker, which is necessary for the marker to be mapped.

We have completed the mango genetic map with 726 SNP markers and identified all 20 linkage groups (chromosomes) of the mango genome. Seven horticultural traits were associated to mapped genetic markers: polyembryony, branch habit, bloom, ground skin color, blush intensity, beak shape and pulp color.

Impact of accomplishment: Making a genetic map is a necessary step in identifying genetic markers associated with horticultural traits to be used in marker assisted selection (MAS) to improve the efficiency of mango breeding. For example, associating polyembryony with a genetic marker allows the breeder to identify the polyembryony trait in seedlings without having to wait years for trees to grow, flower and fruit. Identifying polyembryonic trees means the breeder can rapidly increase the favorable selection by planting the clonal embryos. Similarly, important commercial traits such as fruit color and pulp color can be identified at the seedling level so that breeders can focus their efforts on seedlings that have a greater chance to be improved cultivars.

Genotyping mango germplasm

We have genotyped 1911 mango germplasm accessions from national and international collections with 384 SNP genetic markers. The results support the conclusion that current germplasm collections capture essentially all the genetic diversity available for mango. Mango breeding programs are not taking advantage of the genetic diversity available to them.

Impact of accomplishment: The genetic diversity available in mango germplasm collections was compared to the genetic diversity of parents used in current mango breeding programs. Parents used for breeding show little genetic diversity, although genetically diverse cultivars are easily available. Commercially viable cultivars that are genetically different from current breeding program parents and have favorable disease resistance, color and flavor should be introduced into breeding programs to increase the success of selecting improved mango cultivars. In essence, breeders are not taking advantage of "hybrid vigor" in their breeding programs.

The germplasm genotyping data analysis also suggests that culture and ethnicity play an important role in the mango industry. Mango cultivars from different areas such as Thailand or the Philippines show very little genetic diversity within the area but are genetically distinct when different areas are compared. This suggests that human selection over many centuries has played a large role in the current commercial varieties available worldwide. The US has large immigrant populations from traditional mango consuming areas. Providing these populations with their favorite mango cultivars may aid in promoting mango consumption in the broader US market that does not have a history or cultural traditions involving mango.

LITERATURE CITED

Aron, Y., H. Czosnek, S. Gazit and C. Degani (1998). "Polyembryony in mango (Mangifera indica L.) is controlled by a single dominant gene." Hortscience **33**(7): 1241-1242.

Azim, M. K., I. A. Khan and Y. Zhang (2014). "Characterization of mango (Mangifera indica L.) transcriptome and chloroplast genome." Plant Molecular Biology **85**(1-2): 193-208.

Bally, I. S. E., P. Lu and P. Johnson (2009). Mango Breeding. Breeding plantation tree crops: tropical species. S. M. Jain and P. M. Priyadarshan. New York, Springer. **1:** 51-82.

Dautt-Castro, M., A. Ochoa-Leyva, C. A. Contreras-Vergara, M. A. Pacheco-Sanchez, S. Casas-Flores, A. Sanchez-Flores, D. N. Kuhn and M. A. Islas-Osuna (2015). "Mango (Mangifera indica L.) cv. Kent fruit mesocarp de novo transcriptome assembly identifies gene families important for ripening." Frontiers in plant science **6**: 62.

Dillon, N. L., I. S. E. Bally, C. L. Wright, L. Hucks, D. J. Innes and R. G. Dietzgen (2013). "Genetic diversity of the Australian National Mango Genebank." Scientia Horticulturae **150**: 213-226.

Galán Saúco, V. (2015). "Trends in world mango production and marketing." Acta Horticulturae **in press**.

Hoang, V. L. T., D. J. Innes, P. N. Shaw, G. R. Monteith, M. J. Gidley and R. G. Dietzgen (2015). "Sequence diversity and differential expression of major phenylpropanoid-flavonoid biosynthetic genes among three mango varieties." BMC Genomics **16**(1): 1-12.

Holmes, R., P. Hofman and L. Barker (2010). Mango quality assessmsnt manual, A guide to assessing the post-harvest quality of Australian mangoes. Brisbane, The state of Queensland, Department of Employment, Economic Development and Innovation**:** 59.

Kuhn, D. N., I. S. Bally, N. L. Dillon, D. Innes, A. M. Groh, J. Rahaman, R. Ophir, Y. Cohen and A. Sherman (2017). "Genetic map of mango: a tool for mango breeding." Frontiers in plant science **8**.

Kuhn, D. N., N. L. Dillon, D. J. Innes, L.-S. Wu and K. Mockaitis (2016). "Development of single nucleotide polymorphism (SNP) markers from the mango (Mangifera indica) transcriptome for mapping and estimation of genetic diversity." Acta Horticulturae **1111**: 315-322.

Litz, R. E. (2009). The mango: botany, production and uses, CABI.

Luo, C., B. Shu, Q. Yao, H. Wu, W. Xu and S. Wang (2016). "Construction of a High-Density Genetic Map Based on Large-Scale Marker Development in Mango Using Specific-Locus Amplified Fragment Sequencing (SLAF-seq)." Frontiers in plant science **7**: 1310.

Luria, N., N. Sela, M. Yaari, O. Feygenberg, I. Kobiler, A. Lers and D. Prusky (2014). "De-novo assembly of mango fruit peel transcriptome reveals mechanisms of mango response to hot water treatment." BMC genomics **15**(1): 1.

Margarido, G. R. A., A. P. Souza and A. A. F. Garcia (2007). "OneMap: software for genetic mapping in outcrossing species." Hereditas **144**(3): 78-79.

Mukherjee, S. K. and R. E. Litz (2009). Introduction: botany and importance. The Mango; Botany, Production and Uses 2nd Edition. R. E. Litz. Wallingford, Oxen, CAB International. **1:** 1-18.

Pandit, S. S., R. S. Kulkarni, A. P. Giri, T. G. Kollner, J. Degenhardt, J. Gershenzon and V. S. Gupta (2010). "Expression profiling of various genes during the fruit development and ripening of mango." Plant Physiology and Biochemistry **48**(6): 426-433.

Schnell, R. J., J. S. Brown, C. T. Olano, A. W. Meerow, R. J. Campbell and D. N. Kuhn (2006). "Mango genetic diversity analysis and pedigree inferences for Florida cultivars using microsatellite markers." Journal of the American Society for Horticultural Science **131**(2): 214-224.

Sherman, A., M. Rubinstein, R. Eshed, M. Benita, M. Ish-Shalom, M. Sharabi-Schwager, A. Rozen, D. Saada, Y. Cohen and R. Ophir (2015). "Mango (Mangifera indica L.) germplasm diversity based on single nucleotide polymorphisms derived from the transcriptome." BMC Plant Biology **15**(1): 1-11.

Wu, H.-x., H.-m. Jia, X.-w. Ma, S.-b. Wang, Q.-s. Yao, W.-t. Xu, Y.-g. Zhou, Z.-s. Gao and R.-l. Zhan (2014). "Transcriptome and proteomic analysis of mango (Mangifera indica Linn) fruits." Journal of proteomics **105**: 19-30.