**Final Scientific Report
Cover Page**

BARD Project Number: IS-4341-10

Date of Submission of the report: 18.12.2013

Project Title: Whole genome approach for genetic analysis in cucumber:

Fruit size as a test case.

Investigators

Institutions

Principal Investigator (PI):	Sherman, Amir	ARO, Min. Ag.
Co-Principal Investigator (Co-PI):	Grumet, Rebecca	Michigan St. U
Collaborating Investigators:	Ophir, Ron	ARO, Min. Ag.
	Katzir, Nurit	ARO, Min. Ag.
	Weng, Yiqun	USDA, ARS

Keywords *not* appearing in the title and in order of importance. Avoid abbreviations.

Genetic mapping, Fruit Development, Genome assembly, SNP array

Abbreviations commonly used in the report, in alphabetical order:

QTL- Quantative trait loci RIL – recombinant inbreed line SNP- Single nucleotide polymorphism

Budget:	IS: \$	US: \$	Total: \$
	<u>180,000</u>	<u>140,000</u>	<u>320,000</u>

Signature
Principal Investigator

Signature
Authorizing Official, Principal Institution

Publication Summary (numbers)

	Joint IS/US authorship	US Authors only	Israeli Authors only	Total
Refereed (published, in press, accepted) BARD support acknowledged		3	1	4
Submitted, in review, in preparation	2			2
Invited review papers				
Book chapters				
Books				
Master theses				
Ph.D. theses				
Abstracts	1	2		3
Not refereed (proceedings, reports, etc.)				

Postdoctoral Training: List the names and social security/identity numbers of all postdocs who received more than 50% of their funding by the grant.

Cooperation Summary (numbers)

	From US to Israel	From Israel to US	Together, elsewhere	Total
Short Visits & Meetings			1	1
Longer Visits (Sabbaticals)				

Description Cooperation:

The research achievements are results of close cooperation between the U.S labs that study physiology and genetics of cucumber that develop the genetic resources of the RIL's lines(Weng lab) and carefully studied fruit development (Grumet and Weng lab). The ability to grow this population in two different sites enhances the ability to create confidence in the subject. In addition the Weng lab supplied substantial amount of high quality seq of cucumber that helped in putting to together the genomic resources. The Israeli side (Katzir Ophir and Sherman) concentrated in creating genomic resources including a reestablishment of the physical genomic maps of the two cucumber lines that are parents of the RILs population, identifying huge amount of SNPs and creating self-design SNP array for genotyping. The SNP array was used to genotype the RILs population mapping the fruit size QTL (based on the physiological data from U.S labs) in very high accuracy. This close collaboration ensured that all research objectives have been well done or exceeded, thus ensuring the success of the project resulted from the close collaboration among labs of both countries.

Patent Summary (numbers)

	Israeli inventor only	US inventor only	Joint IS/US inventors	Total
Submitted	0	0	0	0
Issued	0	0	0	0

Final Scientific Report

(allowed)				
Licensed	0	0	0	0

Abstract

The Cucurbitaceae family includes a broad array of economically and nutritionally important crop species that are consumed as vegetables, staple starches and desserts. Fruit of these species, and types within species, exhibit extensive diversity as evidenced by variation in size, shape, color, flavor, and others. Fruit size and shape are critical quality determinants that delineate uses and market classes and are key traits under selection in breeding programs. However, the underlying genetic bases for variation in fruit size remain to be determined. A few species the Cucurbitaceae family were sequenced during the time of this project (cucumber was already sequenced when the project started watermelon and melon sequence became available during the project) but functional genomic tools are still missing.

This research program had three major goals: 1. Develop whole genome cucumber and melon SNP arrays. 2. Develop and characterize cucumber populations segregating for fruit size. 3. Combine genomic tools, segregating populations, and phenotypic characterization to identify loci associated with fruit size. As suggested by the reviewers the work concentrated mostly in cucumber and not both in cucumber and melon. In order to develop a SNP (single nucleotide polymorphism) array for cucumber, available and newly generated sequence from two cucumber cultivars with extreme differences in shape and size, pickling GY14 and Chinese long 9930, were analyzed for variation (SNPs). A large set of high quality SNPs was discovered between the two parents of the RILs population (GY14 and 9930) and used to design a custom SNP array with 35000 SNPs using Agilent technology. The array was validated using 9930, Gy14 and F1 progeny of the two parents.

Several mapping populations were developed for linkage mapping of quantitative trait loci (QTL) for fruit size. These includes 145 F₃ families and 150 recombinant inbred line (RILs F₇ or F₈ (Gy14 X 9930) and third population contained 450 F₂ plants from a cross between Gy14 and a wild plant from India. The main population that was used in this study is the RILs population of Gy14 X 9930.

Phenotypic and morphological analyses of 9930, Gy14, and their segregating F₂ and RIL progeny indicated that several, likely independent, factors influence cucumber fruit size and shape, including factors that act both pre-anthesis and post-pollination. These include: amount, rate, duration, and plane of cell division pre- and post-anthesis and orientation of cell expansion. Analysis of F₂ and RIL progeny indicated that factors influencing fruit length were largely determined pre-anthesis, while fruit diameter was more strongly influenced by environment and growth factors post-anthesis. These results suggest involvement of multiple genetically segregating factors expected to map independently onto the cucumber genome.

Using the SNP array and the phenotypic data two major QTLs for fruit size of cucumber were mapped in very high accuracy (around 300 Kb) with large set of markers that should facilitate identification and cloning of major genes that contribute to fruit size in cucumber. In addition, a highly accurate haplotype map of all RILS was created to allow fine mapping of other

traits segregating in this population. A detailed cucumber genetic map with 6000 markers was also established (currently the most detailed genetic map of cucumber).

The integration of genetics physiology and genomic approaches in this project yielded new major infrastructure tools that can be used for understanding fruit size and many other traits of importance in cucumber. The SNP array and genetic population with an ultra-fine map can be used for future breeding efforts, high resolution mapping and cloning of traits of interest that segregate in this population. The genetic map that was developed can be used for other breeding efforts in other populations. The study of fruit development that was done during this project will be important in dissecting function of genes that contribute to the fruit size QTLs. The SNP array can be used as tool for mapping different traits in cucumber. The development of the tools and knowledge will thus promote genetic improvement of cucumber and related cucurbits.

Achievements:

A. Significance of main scientific achievements

The following results have been achieved in research project:

Establishing genetic material:

Several mapping populations were developed for linkage mapping of quantitative trait loci (QTL) for fruit sizes (fruit length and diameter) by the Weng group. These include 145 F₃ families and 150 F₇ or F₈ recombinant inbred line (RILs) of Gy14 X 9930. The third population contained 450 F₂ plants from a cross between Gy14 and a wild plant introduction line from India (WI 7221 x PI 183967, *Cucumis sativus* var. *hardwickii*). Another RIL population with contrasting fruit sizes of both parents is also available for use, which contains 60 F₉ RILs from Gy14 × PI 183967 (*C. sativus* var *hardwickii*).

Detailed phenotypic analysis of fruit size in cucumber:

The Grumet lab detailed fruit growth from 7 days pre-anthesis to 20 days post-anthesis. Phenotypic analyses of ovule number, cell number, plane of cell division, period of cell division, cell size and shape, and period of cell expansion (were coupled with studies of gene expression . The RIL mapping populations from the cross between Gy14 and 9930 were also analyzed in the field for fruit size by the Grumet and Weng labs in two different locations in the U.S for two consecutive years. This type of analysis is very important in order to differentiate between environmental and genetic basis of the traits of fruit size.

Creating genomic infrastructure for cucumber genetics:

The Weng lab donated to the project a large set of genomic sequence of GY14 that was created by 454 technologies. New sequencing was done in this project as collaboration between the Weng lab and the Israeli group using Illumina technology. As there was a published reference of

Final Scientific Report

9930 the Israeli group used the data to create a large set of genetic variation (SNPs) between the two parents that was evenly distributed along the cucumber chromosomes. Using this database a self-designed SNP array based on Agilent microarray technology for cucumber was developed. The SNP arrays were tested on the parents of the mapping populations: 9930, Gy14 and F1 DNA samples and few other cucumber accessions were supplied by the Weng lab. The positive results established that the SNP array can be utilized as a tool for cucumber genomics.

Mapping the fruit size QTL:

Fine mapping of the fruit size QTLs was done by integrating the phenotypic data from the American labs on cucumber fruit size of the GY14 X 9930 RILs with data obtained from genotyping of 140 F7 GY14 X 9930 RILs. Additional important products of this process were an advanced detailed genetic map of cucumber and a very dense haplotype of the RILs that could lead to mapping of other traits in this population. This genetic map also helped create a better draft and physical assembly of the cucumber genome by indicating that some areas had been misassembled and by allowing for addition of DNA fragments that had been previously unmapped. In addition, the genetic map can be used in future cucumber breeding efforts. Further work will be done by the American teams to identify genes that contribute to cucumber fruit size.

B. Agricultural impacts of the research findings:

This research program has several major achievements that will impact cucumber breeding. Establishing RILs populations with a very dense genetic map will allow the cloning of many traits based on this information including some resistance genes for cucumber diseases. Developing a SNP array as a tool for cucumber breeding and the ability to use this knowledge to develop similar tools will allow mapping of many traits of interest. Detailed analysis of fruit development indicates involvement of multiple genetically segregating factors and provides physiological context to evaluate QTL identified from the SNP array analysis. Fine mapping of the major QTLs for cucumber fruit size can be used to facilitate effective cucumber breeding. All these sets of data will be available through future publications.

Details of cooperation:

This project is a synergistic collaborative project among two American (Grumet and

Weng) and three Israeli (Sherman Ophir and Katzir) laboratories. The cooperation among the American and Israeli labs significantly promoted project objectives, and allowed us to develop state of the art genetic and genomic infrastructure for cucumber that were used for mapping QTLs for fruit size, but have many more applications. The BARD project brought together labs with different expertise. The American groups brought expertise in cucumber plant biology, physiology and genetics. The Israeli side provided knowledge and expertise in functional genomics and development of new genomic tools for mapping. This collaboration led to a set of major achievements in the course of this project and establishment of SNP array technology, a valuable tool for many crops which do not have extensive functional genomic resources. The Weng group brought into the project an advanced RILs population of GY14 X 9930 cucumbers, a large DNA sequence data set for the Gy14 RIL parent. During the project more sequence data was generated for the second RIL parent, 9930. This data enabled identification of tens thousands of SNPs and led to development a SNP array tool by the Israeli group. In parallel the Grumet lab characterized fruit development and gene expression of the parents, F1 and F2 progeny. The RILs populations were also grown in the field by the Weng and Grumet groups in two different locations, two years in a row. The fruit data was analyzed and combined with the genetic data ~~that was~~ generated by the Israeli group based on SNP array analysis of RILs DNA was supplied by the Weng lab in order to map QTLs for cucumber fruit size. In addition, using the genomic data the Israeli group improved the physical map of cucumber and established the SNP array as tool for cucumber breeding. At the outset of the project the USA and Israeli investigators all met at the Annual International Plant & Animal Genome Conference held in California, USA in 2011. The researchers will meet again in 2014 in the same place to draw future plans and discuss papers that we are in the process of writing based on the results of this project. These activities in this grant enhance the connection between the Israeli and American scientist and allow combining two different fields of research in order to create new basis for research in cucumber. This research grant has also promoted deeper collaboration and linkage between American and Israeli scientists in agricultural science and technology research, especially in cucurbit field. We expect two new published papers that are the outcome of the shared work to be published during the next year.

Final Scientific Report

Appendix:

List of publications resulting from the BARD project:

- *Colle M, Ando K, Grumet R. 2013. Expression analysis of gene markers for fruit growth in cucumbers differing in shape and size. Amer Soc Plant Biol. P10049. (abstract)
- *Colle M, Hodges-King C, Grumet R. 2012. Factors influencing fruit size and shape differences in cucumber. In: Cucurbitaceae 2012. Sari N, Solmaz I, Aras V (eds). Cukurova Univ. Press. P. 116-125.
- *Colle M, Shaban M, Grumet R. 2011. Characterization of component factors associated with differences in cucumber fruit size and shape. HortScience 46:S281 (abstract).
- *Yang LM, Koo DH, Li YH, Zhang XJ, Luan FS, Havey MJ, Jiang JM, Weng Y (2012) Chromosome rearrangements during domestication of cucumber as revealed by high-density genetic mapping and draft genome assembly. *Plant Journal*, 71:895–906
- *Yang LM, Li DW, Li YH, Gu XF, Huang SW, Garcia-Mas J, Weng Y (2013) A 1,681-locus consensus genetic map of cultivated cucumber including 67 NB-LRR resistance gene homolog and ten gene loci. *BMC Plant Biology*, 2013,13:53.
- *Sherman A, Eshed R, Harel-Beja R, Tzuri G, Portnoy V, Cohen S, Rubinstein M, Schaffer AA, Burger J, Katzir N and Ophir R. 2013. Combining bulk segregation analysis and microarrays for mapping of the pH trait in melon. *Theor Appl Genet.* 126:349-358.
- Ophir R, Rubinstein M, Eshed R, Katzir N, Colle M, Grumet R Yang L, Weng Y, Sherman A. A method of genotyping by pedigree-based training-set for identification of QTLs associated with cucumber fruit size. PAG meeting Jan 2014 (Abstract).

3. Summary of unpublished data:

A. Phenotyping RILS populations. The two groups (Grumet and Weng) measure similar phenotype fruit FL/LD (Fruit length (FL) / Fruit L/D) as major parameter for fruit length. RILs population was phenotyped in two different locations and some time in three locations for fruit size using different parameters, (Figure 1, 2 different locations different years). In addition the Grumet lab has done careful analysis of different stages of fruit development as describe in the last section. One can see the FL/LD had a strong correlation between the different locations and years supporting the basis of the genetic basis of fruit length QTL (Table 1).

Table 1. Correlations among years and locations for RIL ovary and fruit size data. Experiments were performed in randomized complete block design in locations in Michigan and Wisconsin, 2011 and 2012.

Trait	Summer 2011 vs Summer 2012	Michigan 2011 vs Wisconsin 2011
Ovule number (ON)	0.64****	
Ovary length (OL)	0.63****	
Ovary diameter (OD)	0.38****	

Ovary L/D (OL/D)	0.76****	
Fruit length (harvest) (FLH)	0.37****	
Fruit diameter (harvest) (FDH)	-0.21*	
Fruit L/D (harvest) (L/DH)	0.45****	
Fruit length (FL)	0.8****	0.74****
Fruit diameter (FD)	0.55****	0.59****
Fruit L/D (FL/D)	0.78****	0.78****

Final Scientific Report

Figure 1. Frequency distribution of fruit length (L, top), diameter (D, middle) and L/D (bottom) in 124 RILs from Gy14 × 9930 (2012 Hancock filed data-Weng group)

Figure 2. Distribution of ovary and fruit length, diameter and L/D ratio for (Gy14X9930) F6:7 RIL families at anthesis (left) and fruit maturity (right) from Michigan field trial, 2011

B. SNP discovery based on sequence from 9930 and GY14. The sequence that was available to us are scaffolds of GY14 (from “phytozome”) and reads of GY14 (454 technology; from Weng lab), 9930 (Illumina and Sanger technology; seven chromosomes and scaffolds from ICuGI/BGI version 2). To detect the variation within 9930 was resequenced genomic DNA from this accession using Illumina HiSeq 2000 platform by the Weng lab and analyzed by the Ophir group in order to discover SNPs that are homozygous in the two accessions (9930, GY14). Accession 9930 chromosomes that are the first genome of cucumber that was sequenced were used as a reference. All the different data sets mention above were mapped to 9930 chromosomes using *Mosaik* as alignment tool. The total number of SNPs where we discovered

Final Scientific Report

using the criteria of single SNP per window of 35 bps is 97,015 where 80,708 were successfully mapped to the seven chromosomes, and the rest were found on unresolved scaffolds. The definition of fairly good *homozygous* SNP was based on a minimum sequence repeatability (i.e., coverage ≥ 5 for each allele) and reads quality (higher than 57 in Illumina scale) and result in 73,885.

Figure 3. SNPs position on the cucumber genome reference. The positions on 9930 genome reference of the subset of SNPs that were printed on the SNP microarray. Each bar represents a window of 100Kb.

Figure 4. Whole genome lod-score curve of fruit-shape QTL. A single qtl analysis of cucumber fruit length to diameter ratio illustrated over all seven chromosomes. The two main QTLs of fruit size are marked in chromosomes 1, 4.

C. Combining biological data and SNP array for mapping the fruit size QTL. Based on the biological data that was collected around 140 RILs liners of the GY14X9930 were genotyped using the SNP array and around 27000 SNPs were validated as segregating in Mendelian pattern (data not shown). RILs with higher level than expected of heterozygosity were also filter out. A detail genetic map with 6000 SNPs and 120 RILs were created utilizing “Stamp” program creating 1470 cM genetic map. This map is the most detailed cucumber genetic map that was created until now (data not shown). Validation of the map was done by correlating it on the physical map. The two maps display strong colinearity. Finally a QTL analysis was done with around 6000 SNPs using RQTL.- Example of the genome scanning is presented in figure 4 for the fruit size QTL(FL/LD). Table 2 summarized the list of two major QTLs for every trait that were discovered. Major peaks were defined around few of the QTLs in 300 Kb areas. One can dissect these QTLs further finding the gene of responsible for the phenotype using the detailed SNP map and testing it on different populations and germplasm collection.

Table 2: Location of major QTLs for the different traits. The location of the 2 major QTLs for the traits that were tested represented in their physical map and genetic map. Locations represent the peak signal based on bioinformatics analysis.

Trait	QTL 1				QTL 2			
	lod score	physical position		genetic map position	lod score	physical position		genetic map position
		chr	position (bp)	Position (cM)		chr	position (bp)	Position (cM)
Fruit Length	6.72	Chr4	19790790	75	6.16	Chr6	22393678	184
Fruit Diameter	6.85	Chr6	18652459	161				
Fruit L/D	6.58	Chr4	19790790	75	5.44	Chr1	12398321	87
Ovary Length	5.71	Chr4	15630863	92				
Ovary Diameter	4.46	Chr6	19738497	172	4.19	Chr2	3294727	168
Ovary L/D	5.87	Chr6	29481110	211	4.91	Chr4	19790790	75
Fruit Flesh thickness	9.24	Chr6	21982781	194	5.17	Chr4	10927900	119

Final Scientific Report

D. Detailed analysis fruit growth development (done by the grumet lab).

Our studies of cucumber fruit growth indicate that several, likely independent, factors influence cucumber fruit size and shape, including factors that act both pre-anthesis and post-pollination. These include: amount, rate, duration, and plane of cell division pre- and post-anthesis and orientation of cell expansion. Mature CL fruit had a greater number of cells longitudinally, while GY14 had a greater number in cross section, and cells in CL fruit were more elongated than those in GY14. Analysis of F_2 and RIL progeny indicated that factors influencing fruit length were largely determined pre-anthesis, while fruit diameter was more strongly influenced by environment and growth factors post-anthesis. These results indicate involvement of multiple genetically segregating factors expected to map independently onto the cucumber genome and provide physiological context to evaluate QTL identified from the SNP array analysis. The data supporting these conclusions are presented below.

Ovary and fruit growth of CL, Gy14, and their F_1 and F_2 progeny

Ovary and fruit growth analysis:

Greater ovary length for CL was observed as early as 7 days pre-anthesis and became more evident at anthesis (Figure 1a). Cell number and rate of cell division pre-anthesis was higher for CL than Gy14 in the longitudinal direction with an accelerated increase in CL beginning 2 days pre-anthesis (Figure 2a,c). In contrast, the rate of cell division in the transverse direction was higher in Gy14.

Figure 1. Ovary length, diameter, L/D and ovule number of CL, Gy14, and their F_1 progeny. Each value is the mean of 10 ovaries $\pm$ S.E. The experiment was performed twice with equivalent results.

Figure 2. Ovary cell number and cell size in longitudinal (a,c) and cross-section (b,d) for CL, GY14 and their F_1 progeny from 7 days pre-anthesis until anthesis. Each value is the mean of five ovaries $\pm$ S.E.

Figure 3. Fruit growth post-pollination for CL, GY14 and their F_1 hybrid progeny. Each value is the mean of ten fruits.

Final Scientific Report

Fruit growth post-pollination showed a typical sigmoidal pattern of growth for all genotypes (Figure 3). CL continued to increase both in length and diameter until about 20dpp, but Gy14 was largely done by 10-12 dpp. The increase in cell number also occurred for a longer period of time leading to a greater number of cells in the longitudinal direction for CL (Figure 4a). CL also had significantly longer cells in the longitudinal direction (Figure 4b, $P < 0.05$). In contrast, GY14 fruit showed a greater rate of increase in diameter, and a longer period of rapid cell division and greater number of cells in cross section (Figure 3b, 4c). The final fruit volume and cell number per fruit was greater for GY14 than CL (Figure 4d). Both CL and Gy14 initially showed a greater rate of increase in length than diameter but this reversed as the fruit develops. This transition, which is evident in the L/D ratio, occurred sooner in Gy 14 (Figure 3c). F₁ had intermediate values for length and diameter.

Figure 4. Cucumber fruit cell number (a) and length in the longitudinal direction (b), cell number in cross section (c) and total fruit cell number (d) as a function of days post pollination (dpp). Each value is the mean of five fruits + S.E.

Gene expression analysis in CL and GY14

Developing fruit were harvested at 0, 4, 8, 12, 16, 20 dpp and compared for level and timing of gene expression for genes that we have previously identified to show developmental specific expression in cv. Vlasplik (Ando et al., 2012). Genes predominantly expressed at anthesis and the first few days post-pollination included cell-division related genes such as cyclins, cyclin-dependent kinases and histone genes. Expression patterns of *CYCD3* and *His4* were similar in all three genotypes (CL, GY14, Vlasplik) (Figure 5) and coincided with the period of rapid increase in cell number and preceded the rapid increase in cell size (Figure 4).

Figure 5. qRT-PCR analysis of expression of cell division-related (pre-exponential growth-associated) genes in Gy14, Vlasplik and Chinese Long (CL) as a function of fruit age. Each value is the mean of 3 biological replicates/genotype and 3 technical replicates/biological replicate.

Among the cells most highly expressed at peak exponential growth are cuticle associated genes, including the transcription factor, SHINE (Ando et al., 2012). While peak expression occurred at 8-12 dpp for all three genotypes, cuticle-associated gene expression was 3-10 fold lower in CL than in the two pickling cucumber lines (Figure 6a-c). This observation is consistent with the greater cuticular lipid staining in the pickling cucumber epidermal cell layer. Note also that at 16dpp GY14 fruit have largely lost chlorophyll while CL remain dark green (Figures 3, 7).

Final Scientific Report

Figure 6. qRT-PCR analysis of expression of mid-exponential growth, cuticle associated genes in Gy14, VlaspiK and Chinese Long as a function of fruit age. Each value is the mean of 3 biological replicates/genotype and 3 technical replicates/biological replicate. *significantly different, LSD $P < 0.05$.

Figure 7. (a) Changes in cucumber fruit surface of pickling cucumber 'VlaspiK' from 0 to 4, 8, and 16 (d) days post pollination (dpp). Magnification 370X. At 16 dpp there is extensive cuticle penetration into the epidermal cell layer. (b) Cross section of fruit cuticle at 16dpp of Gy14, VlaspiK and CL stained with Sudan IV at 200x magnification. Brackets indicate epidermal cell layer.

Genes expressed predominantly during late and post-exponentially phase growth in the pickling type Vlasplik were markedly enriched for transcription factors (Ando et al., 2012). CL showed a different pattern for two of the three transcription factor genes tested, with increased earlier expression of *ERF3* and *ATHB-2* (Figure 8). *ATHB-2* in Arabidopsis is associated with orientation of cell expansion (Steindler et al., 1999). CL and GY14 exhibit different cell shape; CL mesocarp cells were more elongated in the longitudinal direction (Figure 4b).

Figure 8. qRT-PCR analysis of expression of gene markers observed during late-post exponential growth in pickling cucumber. Each value is the mean of 3 biological replicates/genotype and 3 technical replicates/biological replicate. *significantly different, LSD $P < 0.05$.

Final Scientific Report

Analysis of segregating F_2 and RIL progeny

F_2 progeny from the GY14 X 9930 cross showed a continuous distribution for fruit length and diameter (data not shown), consistent with quantitative traits. Among the F_2 progeny were individuals with significantly longer fruit than CL (range: 16-33cm; CL= 29 ± 1.05 cm) indicating transgressive segregation for length. Length and diameter segregated independently among the F_2 progeny (Table 1), indicating separable controlling factors.

There was a strong correlation among ovule number, ovary length at anthesis, and fruit length at harvest stage (10 dpp) and 20 dpp (Table 1), indicating that ovule number and ovary length are good predictors of fruit length. Ovary and fruit length and L/D were also highly correlated with each other. In contrast, ovule number was not correlated with ovary diameter or fruit diameter, and ovary diameter was minimally or not correlated with fruit diameter, indicating relative importance of factors influencing fruit diameter post-pollination, and a possibly greater role of environment on fruit diameter than fruit length.

RIL analysis:

Distributions of RILs based on length, diameter and L/D from 2011 are shown in Figure 9. Equivalent distributions were observed in 2012 (data not shown). Ovule number, ovary length, ovary L/D, fruit length and fruit L/D were very highly correlated between growing seasons (2011, 2012; Table 2) ($R=0.6-0.8$, $P<0.0001$). Diameter related traits were less highly correlated ($R=0.3-0.6$, $R<0.0001$), suggesting that environmental factors are larger contributors to fruit diameter. The Y. Weng lab also observed the RILs in the field in Wisconsin for fruit size at maturity. The relationships observed among length, diameter and the L/D ratio at maturity were highly correlated between the two locations for all three parameters (L, D, L/D; $R=0.59-0.78$, $P<0.001$; 2011 data), indicating consistency across environments. The diameter values were less highly correlated, again suggesting stronger environmental effect on diameter than length.

Table 1. Correlation coefficients between ovary traits at anthesis and fruit traits at harvest stage (10 days post pollination) and 20 dpp for greenhouse-grown F₂ progeny of GY14 x CL.

Trait	ON	OL	OD	OL/D	FLH	FDH	L/DH	FL	FD
Ovule number (ON)	-								
Ovary length (OL)	0.5***	-							
Ovary diameter (OD)	-0.23ns	0.24ns	-						
Ovary L/D (OL/D)	0.65****	0.76****	-0.43*	-					
Fruit length (harvest) (FLH)	0.61****	0.74****	-0.10ns	0.77****	-				
Fruit diameter (harvest) (FDH)	0.09ns	0.33*	0.31*	0.10ns	0.52***	-			
Fruit L/D (harvest) (L/DH)	0.52***	0.50***	-0.36*	0.71****	0.54***	-0.43**	-		
Fruit length at 20dpp (FL)	0.60***	0.80****	-0.14ns	0.83****	0.97****	0.38*	0.71****	-	
Fruit diameter at 20dpp (FD)	-0.03ns	0.44*	0.35ns	0.18ns	0.36ns	0.58**	-0.05ns	0.39*	-
Fruit L/D (FL/D)	0.63***	0.59***	-0.36ns	0.79****	0.83****	0.07ns	0.79****	0.84****	-0.16ns

ns, *, **, ***, **** R value not significant, or significant at p<0.05, p<0.01, p<0.001, p<0.0001, respectively

Table 2. Correlations among years and locations for RIL ovary and fruit size data. Experiments were performed in randomized complete block design in locations in Michigan and Wisconsin, 2011 and 2012.

Trait	Summer 2011 (MI) vs Summer 2012 (MI)	Michigan 2011 vs Wisconsin 2011
Ovule number (ON)	0.64****	
Ovary length (OL)	0.63****	
Ovary diameter (OD)	0.38****	
Ovary L/D (OL/D)	0.76****	
Fruit length (harvest) (FLH)	0.37****	
Fruit diameter (harvest) (FDH)	-0.21*	
Fruit L/D (harvest) (L/DH)	0.45****	
Fruit length (FL)	0.8****	0.74****
Fruit diameter (FD)	0.55****	0.59****
Fruit L/D (FL/D)	0.78****	0.78****

Final Scientific Report

Figure 9. Distribution of ovary and fruit length, diameter and L/D ratio for (Gy14XCL) F6:7 RIL families at anthesis (left) and fruit maturity (right) from Michigan field trial, 2011.

As for F_2 progeny, length of ovaries and fruit segregated independently from diameter, indicating separate controlling factors (Tables 3, 4). Once again there were strong correlations ($R=0.5 - 0.8$, $P<0.0001$) among ovule number, ovary length, ovary L/D, fruit length and fruit L/D, indicating that ovule number and ovary length are good predictors of fruit length. In contrast, ovule number was not correlated with ovary or fruit diameter, and ovary diameter was not significantly correlated with fruit diameter. Flesh thickness was more highly correlated with fruit diameter than was seed cavity size. Although somewhat negatively correlated, flesh thickness and seed cavity segregated largely independently ($R= -0.19^*$).

Collectively the RIL data support the observation that factors controlling length are largely determined pre-anthesis while factors regulating diameter primarily occur post-anthesis. These

results also indicate that the length and diameter growth rates determining shape proceed at different rates as the fruit develops from harvest stage to maturity, and are consistent with the differential growth curves observed for length and diameter in the parental genotypes.

Table 3. Correlation coefficients between ovary traits at anthesis and fruit traits at harvest stage and maturity for (Gy14xCL)F_{6,7} RILs, Michigan field trial 2011.

Trait	ON	OL	OD	OL/D	FLH	FDH	L/DH	FL	FD
Ovule number (ON)	-								
Ovary length (OL)	0.47****	-							
Ovary diameter (OD)	0.21ns	0.26**	-						
Ovary L/D (OL/D)	0.59****	0.78****	-0.38****	-					
Fruit length (harvest) (FLH)	0.47****	0.48****	-0.29*	0.61****	-				
Fruit diameter (harvest) (FDH)	0.37****	0.41****	-0.22*	0.50****	0.87****	-			
Fruit L/D (harvest) (L/DH)	0.32***	0.43****	-0.12ns	0.46****	0.48****	0.42****	-		
Fruit length (FL)	0.63****	0.61****	-0.15ns	0.68****	0.59****	0.44****	0.49****	-	
Fruit diameter (FD)	-0.025ns	-0.16ns	0.09ns	-0.19*	-0.17ns	-0.15ns	-0.32***	0.21*	-
Fruit L/D (FL/D)	0.61****	0.68****	-0.18*	0.77****	0.64****	0.49****	0.61****	0.84****	-0.34****

ns, *, **, ***, **** R value not significant, or significant at p<0.05, p<0.01, p<0.001, p<0.0001, respectively

Table 4. Correlation coefficient between ovary traits at anthesis and fruit traits at harvest and maturity of (Gy14xCL) F7:8 RILs. Summer 2012 MSU.

Trait	ON	OL	OD	OL/D	FL	FD	FL/D	Carp	SC	FT
Ovule number (ON)	-									
Ovary length (OL)	0.58****	-								
Ovary diameter (OD)	0.14ns	0.54****	-							
Ovary L/D (OL/D)	0.57****	0.62****	-0.31***	-						
Fruit length (FL)	0.54****	0.60****	-0.04ns	0.70****	-					
Fruit diameter (FD)	-0.13ns	-0.15ns	0.09ns	-0.24**	0.13ns	-				
Fruit L/D (FL/D)	0.52****	0.62****	-0.07ns	0.76****	0.81****	-0.46****	-			
Carpel number (Carp)	0.15ns	0.38****	0.44****	0.03ns	0.19*	-0.15ns	0.26**	-		

Seed cavity diameter (SC)	-0.25*	-0.24**	0.34****	-0.56****	-0.27**	0.47****	-0.51****	0.24**	-	
Flesh thickness (FT)	0.03ns	0.01ns	-0.15ns	0.14ns	0.34****	0.78****	-0.14ns	-0.33****	-0.19*	-