MINING QTL FOR FRUIT WEIGHT QUALITY TRAITS IN UNCHARACTERIZED TOMATO GERMPLASM

by

ALEXIS RAMOS

(Under the Direction of Esther van der Knaap)

ABSTRACT

Tomato (*Solanum lycopersicum*) is a vegetable of great economic importance and its use is largely determined by the shape and size of the produce. While five major fruit weight genes controlling fruit size FW2.2 (CNR), FW3.2 (KLUH), FW11.3 (CSR), locule number LC (WUS), and FAS (CLV3) have been cloned, breeders would greatly benefit from new genes controlling fruit size. Two related experiments were conducted to characterize an underused germplasm and to identify new QTL controlling fruit size. Genotyping and phenotyping of 167 accessions collected from the center of origin of tomato revealed a large variation in fruit size that was not explained by the known genes. This knowledge led to the creation of F_2 mapping populations for Bulk Segregant Analysis with QTLseq, and the identification and confirmation of a novel fruit weight QTL at the bottom of chromosome 2. Future efforts will be aimed to clone the underlying gene.

INDEX WORDS: *Solanum lycopersicum*, phenotyping, QTLseq, germplasm evaluation, fruit weight

MINING QTL FOR FRUIT WEIGHT QUALITY TRAITS IN UNCHARACTERIZED TOMATO GERMPLASM

by

ALEXIS RAMOS

B.S., Florida International University, 2015

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2018

MINING QTL FOR FRUIT WEIGHT QUALITY TRAITS IN UNCHARACTERIZED TOMATO GERMPLASM

by

ALEXIS RAMOS

Major Professor:	Esther van der Knaap
Committee:	Wayne Parrott
	Zenglu Li

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2018

DEDICATION

This is for my family. Lo logré familia !

ACKNOWLEDGEMENTS

Esther van der Knaap; I want to thank Esther van der Knaap for giving me the opportunity to pursue my degree in her lab at UGA, providing all the tools needed to achieve my goal, and sharing her expertise with me. I want to thank Wayne Parrott and Zenglu Li for all their feedback and guidance through my studies. Furthermore, I want to thank Jason Wallace for helping me with the statistical analysis and teaching me about experimental design and GWAS. Finally, I want to thank Josh Clevenger for his help with the QTLseq and career advice.

My family, for always being there for me and supporting all my decisions.

I want to thank everyone in my lab for the tremendous help during field season and being so friendly and helpful. Specially Neda Keyhaninejad for being so awesome and a great lab manager and Eudald Illa Berenguer for teaching me a lot of the lab skills and statistical tools. I also want to thank Nathan Taitano for being a great friend through my almost 3 years at UGA, for being a computer wizard, and for helping me with bioinformatics.

Also, I want to thank all PBGG faculty, staff, and students for being so helpful and providing an amazing learning environment.

Finally, I want to thank Kristen Savannah Humphries for being such a wonderful person and providing so much support through my thesis defense week, for holding cherry tomatoes to take a cool image for my presentation, and for buying me some relaxing drinks after my defense.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Tomato (*Solanum lycopersicum* L.) is a member of the large family Solanaceae or Nightshades (Peralta, Knaap, & Spooner, 2007). Tomato is an annual, biennial, or sometimes perennial herb with branches that can grow up to 4 m from centers, which is primarily cultivated for human consumption (Peralta, Knaap, & Spooner, 2007). However, tomato is usually grown as an annual crop for agricultural purposes (Peralta, Knaap, & Spooner, 2007). Tomato is mostly autogamous with inserted stigmas, but there are facultatively allogamous populations that present exerted stigmas. This predominantly selfing reproduction system makes tomato a highly inbred species (Ranc et al., 2012). Normally, tomato fruits are bright red when ripe and have various shapes and sizes. A wide range of tomato varieties presenting different morphologies is currently commercialized (Diez, 1995; Monforte et al., 2014; Rodríguez et al., 2011). Tomato plants have a generation time of about 4-5 months, depending on the variety.

Tomato taxonomy and naming

Tomato was first named as *pomo d'oro* in 1544 by Italian Pietro Andrea (Peralta, Knaap, & Spooner, 2007). Over time, tomato has been renamed and reclassified several times. Between late 1600s and early 1800s there was controversy with the binomial nomenclature of tomato. Turnefort (1694) classified tomato in its own genus *Lycopersicon* and classified cultivated tomato as *Lycopersicum*. Linnaeus (1753) placed tomato in the Solanum genus and named cultivated tomato *S. lycopersicum* using the binomial nomenclature system. Some years later cultivated tomato was reclassified by Turnefort as *Lycopersicon* (genus) *esculentum* (species). Name

changing in tomato lasted hundreds of years until more recently genomic evidence and phylogenetic studies arose. *Solanum* (genus) *lycopersicum* (species) is now the accepted and most used name for cultivated tomato among breeders and scholars alike. Some synonym names of cultivated tomato are *Solanum lycopersicum* L. var. *lycopersicum*, *Lycopersicon esculentum* Mill, and *Solanum lycopersicum* var. *esculentum*.

S. lycopersicum has been further classified into two botanical varieties, *Solanum lycopersicum* L. var. *cerasiforme* (SLC) and *Solanum lycopersicum* L. var. *lycopersicum* (SLL) (Blanca et al., 2012). The controversy about these botanical varieties and the evolutionary relationships of SLC and SLL to each other and to the closest wild relative *Solanum pimpinellifolium* (SP) has largely been resolved. SLC has been proposed to be the immediate ancestor of SLL based on its wide presence in Central America, and the short style length in the flowers (Hancock, 1992). There are two main competing hypotheses about the evolutionary position of SLC. One hypothesis suggests that SLC is an admixture between SP and SLL (Ranc, Muños, Santoni, & Causse, 2008). It has been suggested that SLC is the result of cultivated tomato escapes into the wild, or hybrids between cultivated and weedy species (Peralta, Knaap, & Spooner, 2007). The second, more recent hypothesis, suggests the variety SLC is the evolutionary intermediate between SP and SLL, and regards it as the immediate ancestor of SLL (Blanca et al., 2015) as well as a direct descendant of certain SP from Ecuador.

Tomato origin and domestication

South America is the center of origin of tomato since all its wild relatives including SP originated there. There are 13 wild tomato species, and four of them readily intercross with tomato (Peralta, Knaap, & Spooner, 2007). *Solanum pimpinellifolium* (SP) is the closed wild relative of cultivated tomato (Peralta, Knaap, & Spooner, 2007; Rick & Forbes, 1975; Zuriaga et

al., 2009). Historically, the tomato center domestication was proposed to be located either in Peru and Mexico (Peralta & Spooner, 2006). However, the recently proposed tomato evolution suggests a two-step domestication process. Genetic analysis of populations of SP, SLC, and SLL suggests early tomato domestication in Ecuador, and later domestication in Mexico (see Figure 1.1) (Blanca et al., 2015). Tomato was then taken from Mesoamerica to Europe by Spanish Conquistadors in the 16th century, and most breeding has used this germplasm (Jenkins, 1948).

Genetics and genetic resources

S. lycopersicum is a diploid plant with 2n = 2x = 24 chromosomes and a genome size of ~960 MB (Pavan, van Heusden, & Bai, 2001). The genome of the inbred tomato line *Solanum lycopersicum* var "Heinz 1706" has been sequenced and serves as a reference genome (The Tomato Genome Consortium, 2012). Genetically, wild tomato shows the greatest diversity, followed by SLC, and the least diversity is observed in SLL (Blanca et al., 2015; Ranc et al., 2008). Cultivated tomato has low genetic diversity because of its autogamous nature and its evolutionary history (Yuling & Lindhout 2007).

Uses, economic importance, and breeding

Tomato is mainly grown for its fleshy fruit and its production worldwide has been steadily increasing since 1993; reaching 177 million tonnes in 2016 (FAOSTAT, 2016). Most of the production is localized in Asia, Europe, and Americas; with China and India being the largest producers with 56 and 18 million tonnes respectively in 2016 (FAOSTAT, 2016). Tomato is widely consumed in various forms, and it is classified into two broad categories, fresh or processing, based on its use and harvesting method. Fresh tomatoes include tomatoes consumed without any industrial modifications and fortifications. Processing tomatoes include the tomatoes that are industrially processed for pastes, sauces, soups and canning (Tomatoland Information

Services, 2011). From 1999-2009, about 75% of the total world tomato production was classified as fresh, while 25% was for processing (Tomatoland Information Services, 2011). In contrast, most of the tomato produced in United States is for processing. In 2008, only 11% of tomato produced was for fresh market (USDA, 2016).

Tomato breeding can be separated into two main areas, each concerned with slightly different goals that target either the processing or the fresh tomato market industries. The distinctive breeding goal for processing tomato is concerned with harvesting the tomatoes mechanically. Processing tomatoes are bred to be firm, have jointless pedicel and compact fruit set, uniform ripening, prolonged shelf life, and determinate growth habit (Berry et al., 1991; (Stevens & Rick, 1986). On the other hand, breeding for fresh market tomatoes is more concerned with the fruit appearance (size, shape, flavor, and color), and nutritional attributes (Acquaah, 2009; Bai & Lindhout, 2007, Stevens & Rick, 1986). This difference in machine-harvested processing tomatoes and hand-picked fresh market tomatoes greatly influences their prices; fresh tomato is more expensive on a per pound basis, because the larger production costs (USDA ERS, 2016). In general, tomato breeding also aims to increase yield, biotic and abiotic resistance, and other fruit quality traits as higher percentage of soluble solids to make tomato products more efficiently (USDA, 2016; Berry et al., 1999; Bai & Lindhout, 2007).

Components of tomato fruit weight

Fruit size is determined by cell division and expansion, which cause changes in cell number and cell size (Roth, 1977; Bourdon et al., 2010). The period of fruit development in which these processes influence fruit weight, and the tissues involved, varies among species (Coombe, 1976; Roth, 1977). Fruit growth is generally dictated by an intense period of cell divisions early in fruit development, followed by rapid cell enlargement which culminates with

fruit ripening (Bourdon et al., 2010). In tomato fruits, increased fruit weight is caused by an increase in cell number and cell size, where cell division begins at anthesis and continues for ~1-2 weeks after fertilization (Tanksley, 2004, Xiao et al., 2009). Cell expansion follows the cell division process and lasts until ~1 week before ripening (Tanksley, 2004). When the tomato fruit is in the mature green stage, ~5 weeks post anthesis, both cell division and cell expansion have ceased (Giovannoni, 2004; Xiao et al., 2009). The ripening phase involves chemical changes that determine aroma, color, texture, etc., but no further fruit size or shape changes take place then (Tanksley, 2004). The increase in cell size in tomato has been largely attributed to endoreduplication (Bergervoet et al., 1996; Cheniclet et al. 2005; Mu, 2015). The increases in cell number and size may manifest themselves in the different parts of the fruit. In tomato, these increases result in larger pericarp, septum, columella and placental tissues, causing increased fruit weight (van der Knaap et al., 2014). Lastly, fruit weight is also greatly influenced by locule number, where a higher number of locules causes an increase in fruit weight (Lippman & Tanksley, 2001; van der Knaap et al., 2014).

Major fruit shape and weight genes

Tomato fruit shape and size are quantitative traits, and many candidate QTL have been mapped (Grandillo et al., 1999). However, only few of those fruit shape and size QTL have been confirmed and cloned. Among them are weight genes FW2.2(CNR) in chromosome 2, FW3.2(SlKLUH) in chromosome 3, and FW11.3(CSR) in chromosome 11, and shape genes OVATE in chromosome 2, FASCIATED (FAS, ortholog of CLV3) in chromosome 11, LOCULE NUMBER (LC, ortholog of WUSCHEL) in chromosome 2, and SUN in chromosome 7 (Mu et al., 2017; van der Knaap et al., 2014). These genes are responsible for most of the fruit shape

variation in tomato; and the three fruit weight genes together with FAS and LC, greatly dictate fruit weight.

Major fruit shape genes

OVATE and SUN are responsible for elongation of tomato fruits, and have additive effects (Liu et al., 2002; Wu et al., 2015). OVATE is a negative growth regulator that modifies shape and architecture of ovaries. It controls cell division in the proximal distal axis, at the proximal end of the fruit, causing oblate or oval shaped fruit (Tanksley, 2004). It is expressed at, and up to two weeks after anthesis; creating elongation of the ovaries at very early stages of fruit development (Liu et al., 2002). The derived ovate allele, which causes fruit elongation, has a premature stop codon in an Ovate Family Protein (OFP) (Liu et al., 2002). The ovate mutation does not cause the same phenotype in different genetic backgrounds, which led to the mapping of two suppressors, sov1 and sov2 loci (Rodríguez et al., 2013; van der Knaap et al., 2014).

SUN is more effective than OVATE in controlling fruit elongation in the proximal distal axis (van der Knaap, 2014). It has the greatest impact in fruit development post anthesis, but its effect can be noted at anthesis (van der Knaap & Tanksley 2001; Wu et al., 2011). SUN causes an increase in cell number in proximal distal axis, while decreasing cell number in the medio lateral axis (Wu et al., 2011). However, the mechanism by which it changes patterns in cell division is poorly understood (van der Knaap et al., 2014). SUN encodes a member of the IQD family of calmodulin-binding proteins (Xiao et al., 2008).

LC and FAS are regulators of locule number in tomato and they are both expressed pre-anthesis (van der Knaap et al., 2014). LC is involved in regulation of stem cell fate in plants and the higher locule fruit phenotype is caused by two SNPs located downstream of putative tomato ortholog of WUSCHEL (Muñoz et al., 2011). This derived LC allele causes a change in number

of carpel primordia, increasing locule number. The mechanism is not known, but it is suggested that the mutation causes loss of regulation of WUS expression, increasing stem cell population and consequently locule number (van der Knaap et al., 2014). FAS has a larger effect than LC, and these two genes have epistatic effects (Lippman & Tanksley, 2001). It has been suggested that FAS impacts meristem organization, size, and boundary information (van der Knaap et al., 2014). The FAS allele causing an increase in locule number in tomato is driven by a partial loss in expression of SlCLV3, the tomato putative ortholog of CLV3 (Xu et al., 2015). Even though LC and FAS these are shape genes, an increase in locule number causes an increase in fruit size.

Major fruit weight genes

CNR, SlKLUH, and CSR are responsible for variation in tomato fruit weight and have additive effects. The derived or mutated alleles for these genes cause higher weight by either increasing cell size or cell number. CNR and SlKLUH increase cell number in tomato (Frary et al., 2000; Chakrabarti et al., 2013). CNR effects in fruit size are observed at anthesis, and in latter stages of fruit development (Nesbitt & Tanksley, 2001; Cong & Tanksley, 2002). This gene is associated with changes in cell number in tomato carpel ovary, and responsible for changes in fruit weight by up to 30% (Frary et al., 2000). SNP changes in the promoter of the gene are predicted to be causing the allelic variation at this locus (Frary et al., 2000). CNR encodes negative regulator membrane proteins member of the Cell Number Regulator (CNR) family, which regulate cell number causing enlargement of placenta and columella (Gonzalo et al.,2009; Guo et al., 2010). Little is known how the proteins lead to changes in cell division, however. SlKLUH encodes cytochrome P450, and it increases cell number in pericarp and septum areas (Chakrabarti et al., 2013). It is proposed that a SNP in the gene promoter causes the difference in

fruit mass (Chakrabarti et al., 2013). SlKLUH effect on fruit weight becomes apparent post-anthesis (Chakrabarti et al., 2013).

CSR has a smaller effect on fruit weight than CNR and SlKLUH (van der Knaap & Tanksley, 2003). The CSR locus has been fine mapped to a 13-kb region, and the candidate gene responsible for the increase in fruit weight was named Cell Size Regulator (CSR) (Mu et al., 2017). The derived allele of CSR causes an increase in tomato pericarp cell size, and some increase in cell ploidy levels in pericarp and columella tissues (Mu et al., 2017). CSR has been suggested to be involved in regulating endoreduplication (Mu et al., 2017).

Distribution of weight gene alleles in cultivated tomato germplasm

Even though there is great variation in tomato fruit size and shape, the major genes for these traits are mostly fixed in the cultivated germplasm. The distribution of the major genes LC, FAS, CNR, SlKLUH, and CSR has been determined for a large tomato population containing members of the three tomato groups (530 SLL, 316 SLC, and 145 SP) (Blanca et al., 2015). Overall, the wild type allele is predominantly fixed in SP for all loci; and SLC shows the greatest allelic variation across all loci. The SLL accessions, which represent the cultivated germplasm, however, have a high frequency for the derived allele for the fruit weight loci, while low for FAS, and intermediate for LC. Even though the SLL accessions in the study were derived from active crop improvement programs (Blanca et al., 2015), the data show how most loci have very high frequency for the weight genes, meaning the weight potential for size has been almost capped in the cultivated germplasm. There is some room for increasing weight by introgressing the derived alleles for LC and FAS in this modern germplasm. However, that would be detrimental for the processing market as they prefer ellipsoid or rectangular shaped fruit over flat or round types (van der Knaap, 2013).

Future research and breeding

Even though there have been many advances in understanding how tomato fruit shape and size are controlled, there is much work to do. Future works should focus on determining regulatory elements in the genome that affect expression of these genes. Also, the mechanisms and pathways by which they impact fruit morphology should be further investigated with fluorescence tagging of candidate proteins and gene knockout studies. New information about these genes, as well as the discovery of new genes, can be directly implemented in breeding programs. Exploitation of tomato morphology genetics can be used to develop tomatoes with desired shapes and sizes, and potentially target or create new fresh market classes and consumer niches.

Literature cited

Abràmoff, M. D., Magalhães, P. J., & Ram, S. J. (2004). Image processing with ImageJ. *Biophotonics international*, *11*(7), 36-42.

Acquaah, George. (2009) *Principles of plant genetics and breeding*. John Wiley & Sons. 667-676.

Bai, Y., and Lindhout, P. (2007). Domestication and breeding of tomatoes: what have we gained and what can we gain in the future? *Annals of botany* 100.5, 1085–1094. doi: 10.1093/aob/mcm150

Bergervoet, J. H., Berhoeven, H. A., Gilissen, L. J., & Bino, R. J. (1996). High amounts of nuclear DNA in tomato (Lycopersicon esculentum Mill.) pericarp. *Plant Science*, *116*(2), 141-145.

Berry, S. Z., Gould, W. A., & Wiese, K. L. (1991). OHIO 8245 PROCESSING TOMATO. *Hortscience*, *26*(8), 1093-1093.

Blanca, J., Canizares, J., Cordero, L., Pascual, L., Diez, M. J., & Nuez, F. (2012). Variation Revealed by SNP Genotyping and Morphology Provides Insight into the Origin of the Tomato. *Plos One, 7*(10). doi: 10.1371/journal.pone.0048198

Blanca, J., Montero-Pau, J., Sauvage, C., Bauchet, G., Illa, E., Díez, M. J., . . . Cañizares, J. (2015). Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC Genomics*, *16*(1), 257. doi: 10.1186/s12864-015-1444-1

Chakrabarti, M., Zhang, N., Sauvage, C., Munos, S., Blanca, J., Canizares, J., . . . van der Knaap, E. (2013). A cytochrome P450 regulates a domestication trait in cultivated tomato.

Proceedings of the National Academy of Sciences of the United States of America, 110(42), 17125-17130. doi: 10.1073/pnas.1307313110

Chernet, S., Belew, D., & Abay, F. (2014). Genetic diversity studies for quantitative traits of tomato (Solanum lycopersicon L.) genotypes in Western Tigray, Northern Ethiopia. Journal of Plant Breeding and Crop Science, 6(9), 105-113

Cheniclet, C., Rong, W. Y., Causse, M., Frangne, N., Bolling, L., Carde, J. P., & Renaudin, J. P. (2005). Cell expansion and endoreduplication show a large genetic variability in pericarp and contribute strongly to tomato fruit growth. *Plant Physiology*, *139*(4), 1984-1994.

Coombe, B. G. (1976). The development of fleshy fruits. *Annual Review of Plant Physiology*, *27*(1), 207-228.

Cong, B., Liu, J. P., & Tanksley, S. D. (2002). Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(21), 13606-13611. doi: 10.1073/pnas.172520999

Diez M. (1995) Tipos varietales. In: Nuez F, editor. El cultivo del tomate. Madrid: Mundi-Prensa. pp. 93–129.

Food and Agriculture Organization of the United Nations. 2016. *FAOSTAT*. Rome, Italy: FAO.

Frary, A., Nesbitt, T. C., Frary, A., Grandillo, S., van der Knaap, E., Cong, B., . . . Tanksley, S. D. (2000). fw2.2: A quantitative trait locus key to the evolution of tomato fruit size. *Science, 289*(5476), 85-88. doi: 10.1126/science.289.5476.85

Giovannoni, J. J. (2004). Genetic regulation of fruit development and ripening. *The plant cell*, *16*(suppl 1), S170-S180.

Grandillo, S., Ku, H. M., & Tanksley, S. D. (1999). Identifying the loci responsible for natural variation in fruit size and shape in tomato. *Theoretical and Applied Genetics, 99*(6), 978-987. doi: 10.1007/s001220051405

Grandillo, S., Zamir, D., & Tanksley, S. D. (1999). Genetic improvement of processing tomatoes: A 20 years perspective. *Euphytica, 110*(2), 85-97. doi: 10.1023/a:1003760015485

Hancock JF. Plant Evolution and the Origin of Crop Species. 1992. Prentice Hall, Englewood Cliffs, NJ. pp 275-276

Huang, Z. J., & van der Knaap, E. (2011). Tomato fruit weight 11.3 maps close to fasciated on the bottom of chromosome 11. *Theoretical and Applied Genetics, 123*(3), 465-474. doi: 10.1007/s00122-011-1599-3

Jenkins, J. A. (1948). The Origin of the Cultivated Tomato. *Economic Botany, 2*(4), 379-392.

Ku, H. M., Grandillo, S., & Tanksley, S. D. (2000). fs8.1, a major QTL, sets the pattern of tomato carpel shape well before anthesis. *Theoretical and Applied Genetics, 101*(5-6), 873-878. doi: 10.1007/s001220051555

Lippman, Z., & Tanksley, S. D. (2001). Dissecting the genetic pathway to extreme fruit size in tomato using a cross between the small-fruited wild species Lycopersicon pimpinellifolium and L-esculentum var. giant heirloom. *Genetics, 158*(1), 413-422.

Liu, J. P., Van Eck, J., Cong, B., & Tanksley, S. D. (2002). A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. *Proceedings of the National Academy of Sciences of the United States of America, 99*(20), 13302-13306. doi: 10.1073/pnas.162485999

Monforte, A. J., Diaz, A., Cano-Delgado, A., & van der Knaap, E. (2014). The genetic basis of fruit morphology in horticultural crops: lessons from tomato and melon. *Journal of Experimental Botany, 65*(16), 4625-4637. doi: 10.1093/jxb/eru017

Mu, Q., Huang, Z., Chakrabarti, M., Illa-Berenguer, E., Liu, X., Wang, Y., … Knaap, E. van der. (2017). Fruit weight is controlled by Cell Size Regulator encoding a novel protein that is expressed in maturing tomato fruits. *PLOS Genetics*, *13*(8), e1006930. https://doi.org/10.1371/journal.pgen.1006930

Nesbitt, T. C., & Tanksley, S. D. (2001). fw2.2 directly affects the size of developing tomato fruit, with secondary effects on fruit number and photosynthate distribution. *Plant Physiology, 127*(2), 575-583. doi: 10.1104/pp.010087

Nesbitt, T. C., & Tanksley, S. D. (2002). Comparative sequencing in the genus Lycopersicon: Implications for the evolution of fruit size in the domestication of cultivated tomatoes. *Genetics, 162*(1), 365-379.

Pavan, S., van Heusden, A. W., & Bai, Y. (2001). Solanum lycopersicum (Tomato) *eLS*: John Wiley & Sons, Ltd.

Peralta, I.E., Knapp, S., Spooner, D.M. (2007). The taxonomy of tomatoes: a revision of wild tomatoes (solanum l. section lycopersicon (mill.) wettst.) and their outgroup relatives (solanum sections juglandifolium (rydb.) child and lycopersicoides (child) peralta). *Systematic Botany Monographs*. 84:1-10.

Peralta, I., & Spooner, D. (2006). History, Origin and Early Cultivation of Tomato (Solanaceae) *Genetic Improvement of Solanaceous Crops Volume 2* (pp. 1-24): Science Publishers.

Ranc, N., Munos, S., Xu, J. X., Le Paslier, M. C., Chauveau, A., Bounon, R., . . . Causse, M. (2012). Genome-Wide Association Mapping in Tomato (Solanum lycopersicum) Is

Possible Using Genome Admixture of Solanum lycopersicum var. cerasiforme. *G3-Genes Genomes Genetics, 2*(8), 853-864. doi: 10.1534/g3.112.002667

Rick, C. M., & Fobes, J. F. (1975). Allozyme Variation in the Cultivated Tomato and Closely Related Species. *Bulletin of the Torrey Botanical Club, 102*(6), 376-384. doi: 10.2307/2484764

Rodriguez, G. R., Kim, H. J., & van der Knaap, E. (2013). Mapping of two suppressors of OVATE (sov) loci in tomato. *Heredity, 111*(3), 256-264. doi: 10.1038/hdy.2013.45

Rodríguez, G. R., Moyseenko, J. B., Robbins, M. D., Huarachi Morejón, N., Francis, D. M., van der Knaap, E. Tomato Analyzer: A Useful Software Application to Collect Accurate and Detailed Morphological and Colorimetric Data from Two-dimensional Objects. *J. Vis. Exp.* (37), e1856, doi:10.3791/1856 (2010).

Roth, I. (1977). *Fruits of Angiosperms*. Stuttgart, Germany: Schweizerbart Science Publishers. pp. 14-16.

Sato, S., Tabata, S., Hirakawa, H., Asamizu, E., Shirasawa, K., Isobe, S., . . . Tomato Genome, C. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature, 485*(7400), 635-641. doi: 10.1038/nature11119

Sims, W. L. (1980). *HISTORY OF TOMATO PRODUCTION FOR INDUSTRY AROUND THE WORLD*.

Stevens, M. A., & Rick, C. M. (1986). Genetics and breeding. In *The Tomato Crop* (pp. 35–109). Springer, Dordrecht. https://doi.org/10.1007/978-94-009-3137-4_2

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., ... & Innan, H. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *The Plant Journal*, *74*(1), 174-183.

Tanksley, S. D. (2004). The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. *Plant Cell, 16,* S181-S189. doi: 10.1105/tpc.018119

Tomato Genome Consortium. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, *485*(7400), 635-641.

Tomatoland Information Services. (2011). "*World Consumption: Table Tomato and Processed Tomato*." Retrieved from http://www.tomatoland.com/

USDA Economic Research Service. (2016). "*Tomatoes*". Retrieved from http://www.ers.usda.gov

van der Knaap, E. (2013) Morphology Matters. *International Innovation* 116, 81-83.

van der Knaap, E., Chakrabarti, M., Chu, Y. H., Clevenger, J. P., Illa-Berenguer, E., Huang, Z. J., . . . Wu, S. (2014). What lies beyond the eye: the molecular mechanisms regulating tomato fruit weight and shape. *Frontiers in Plant Science, 5*. doi: 10.3389/fpls.2014.00227

van der Knaap, E., & Tanksley, S. D. (2001). Identification and characterization of a novel locus controlling early fruit development in tomato. *Theoretical and Applied Genetics, 103*(2-3), 353-358. doi: 10.1007/s001220100623

Wu, S., Clevenger, J. P., Sun, L., Visa, S., Kamiya, Y., Jikumaru, Y., . . . van der Knaap, E. (2015). The control of tomato fruit elongation orchestrated by sun, ovate and fs8.1 in a wild relative of tomato. *Plant Science, 238*, 95-104. doi: 10.1016/j.plantsci.2015.05.019

Wu, S., Xiao, H., Cabrera, A., Meulia, T., & van der Knaap, E. (2011). SUN Regulates Vegetative and Reproductive Organ Shape by Changing Cell Division Patterns. *Plant Physiology, 157*(3), 1175-1186. doi: 10.1104/pp.111.181065

Xiao, H., Jiang, N., Schaffner, E., Stockinger, E. J., & van der Knaap, E. (2008). A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science, 319*(5869), 1527-1530. doi: 10.1126/science.1153040

Xiao, H., Radovich, C., Welty, N., Hsu, J., Li, D., Meulia, T., & van der Knaap, E. (2009). Integration of tomato reproductive developmental landmarks and expression profiles, and the effect of SUN on fruit shape. *BMC Plant Biology*, *9*(1), 1.

Zuriaga, E., Blanca, J., & Nuez, F. (2009). Classification and phylogenetic relationships in Solanum section Lycopersicon based on AFLP and two nuclear gene sequences. *Genetic Resources and Crop Evolution, 56*(5), 663-678. doi: 10.1007/s10722-008-9392-0

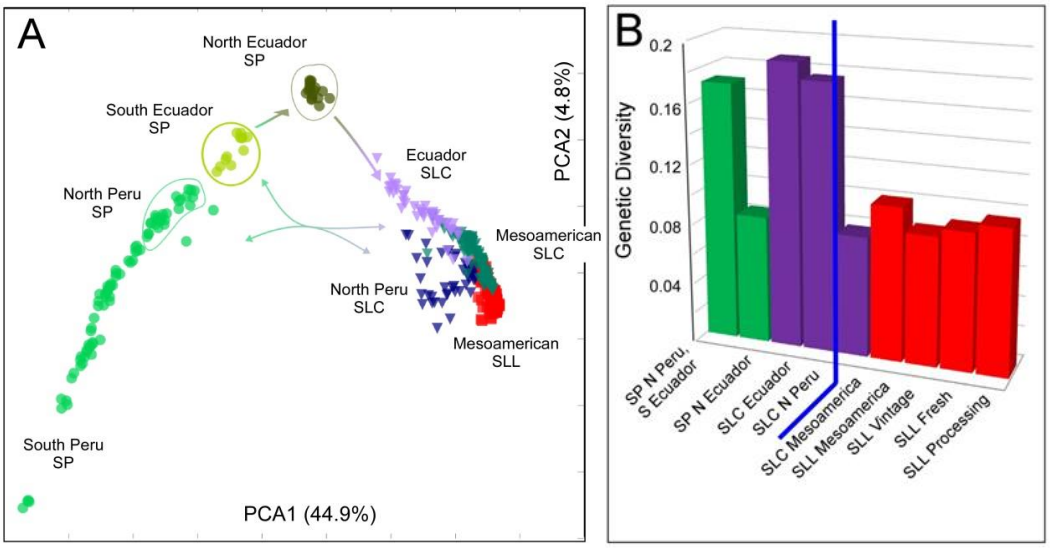


Figure 1.1. Genetic composition of tomato in different geographical regions. (A) SP, SLC, and SLL genetic clustering based on PCA analysis, and (B) genetic diversity (expected heterozygosity).

CHAPTER 2

CHARACTERIZATION OF AN UNDERUSED TOMATO GERMPLASM AND ANALYSIS OF GENETIC AND MORPHOLOGICAL PREDICTORS OF FRUIT WEIGHT [1]

[1] Ramos, A., Illa-Berenguer, E., Razifard, H., Caicedo, A.L., and van der Knaap, E. To be submitted to *Euphytica*.

Abstract

Tomato originated in South America and it was domesticated in a two-step process, first in South America and then in Mexico. The germplasm that made it to Mexico has been extensively used in breeding, and identification of new sources of variation would be highly beneficial. In this study we characterized in depth a tomato population comprised of wild, semi wild, and ancestral cultivated tomato accessions for fruit size traits. We found new sources of variation in fruit size quality traits that were not explained by the major fruit weight genes FW2.2 (CNR), FW3.2 (KLUH), FW11.3 (CSR), LC (WUSCHEL), and FAS (CLV3), providing evidence supporting that genes were either left behind during domestication, or made it to modern germplasm but have not be cloned yet. Our study also supports a two-step domestication process in tomato, but phenotypically we found some discrepancies not reported previously. Also, we show that most fruit area measurements are the best predictors of fruit weight. Furthermore, we found that the increase in fruit weight from *Solanum pimpinellifolium* (SP) to *Solanum lycopersicum* var. *cerasiforme* (SLL) was accompanied by an increase in percentage water content, and a disproportional increase in columella to fruit area ratio, suggesting that the central part of the fruits enlarged more than the remainder of the fruit during domestication. Also, our combined approach lays a foundation that should facilitate GWAS and biparental mapping efforts.

Introduction

Cultivated tomato *Solanum lycopersicum* is a fruit of tremendous economic importance worldwide and its consumption has been increasing yearly reaching 164 million tonnes in 2013 ("Food and Agriculture Organization of the United Nations," 2012). Tomato originated in South America and all wild relatives are native to the Andean region (Peralta, Knaap, & Spooner, 2007) . The history of tomato domestication was debated for some time. However recent studies using

genome-wide analyses support a two-step domestication process, where the first domestication took place in Ecuador and northern Peru and a second domestication that took place in Mexico (Blanca et al., 2015). Blanca et al. (2015) also clarified the biological status of tomato phylogenetic groups *Solanum lycopersicum var. lycopersicum* (SLL), *S. lycopersicum var. cerasiforme* (SLC), and S. *pimpinellifolium* (SP), and provided evidence that SLC arose in the Andean region of Ecuador and Peru from SP, and that SLL arose from SLC in Mexico. Tomato domestication, as in many other crops (Frary & Doğanlar, 2003), was accompanied by an explosion of morphological and physiological traits compared to the wild ancestors. One of the most relevant and important features in tomato domestication was the tremendous increase in fruit size (Bai & Lindhout, 2007).

Tomato fruit morphology is very important, as it determines the culinary purpose of the fruit. Most commercially grown tomatoes belong to SLL and are divided into cultivars that target the fresh (field and greenhouse) or the processing market. Tomato breeding goals vary depending on location, need, and resources (Bai & Lindhout, 2007). Yet, there are desired traits that are common for both market classes such as increased yield, and biotic and abiotic resistance. One of the determining factors for either the processing or fresh market class is the fruit shape and size. To facilitate mechanical harvesting and processing, processing tomatoes are ideally of uniform shape and size (Berry, Gould, & Wiese, 1991). On the other hand, fresh market tomatoes are bred to have compelling or niche-specific shape and size based on consumer preferences (Acquaah, 2009; Bai & Lindhout, 2007, Stevens & Rick, 1986). An example of common fresh market tomatoes are cherry and globe tomatoes. Cherry tomato morphology makes it suitable for snacking and for use in salads, while globe tomato is preferred for slicing.

Fruit size is determined by cell division and expansion, which cause changes in cell number and cell size (Bourdon et al., 2010; Roth, 1977). In tomato, fruit size is determined by a period of rapid cell division at anthesis continuing for ~1-2 weeks after fertilization (Tanksley, 2004; Xiao et al., 2009), followed by cell expansion that lasts ~1 week before ripening (Tanksley, 2004). These increases in cell number and size manifest themselves in different parts of the produce and together give rise to the final fruit size and shape. Also, an increase in the number of locules results in an increase in fruit size while also altering its shape.

Many candidate fruit size and shape QTL have been mapped (Grandillo, Ku, & Tanksley, 1999) and seven major genes have been cloned (Mu et al., 2017; van der Knaap et al., 2014). Two major genes controlling locule number in the fruit, LOCULE NUMBER (LC) (ortholog of WUSCHEL) and FASCIATED (FAS) (ortholog of CLV3), have been cloned (Muños et al., 2011; Xu et al., 2015). FAS has a larger effect than LC, and these two genes have epistatic effects (Barrero & Tanksley, 2004; Lippman & Tanksley, 2001). Also, various genes controlling fruit weight by changing cell size or cell number in different parts of the fruit have been cloned. FW2.2 (Cell Number Regulator, CNR) increases the cell number in the tomato columella and placenta (Frary et al., 2000), while FW3.2 (SlKLUH) causes an increase in cell number in the septum and pericarp (Chakrabarti et al., 2013). FW11.3 (Cell Size Regulator, CSR) causes an increase in tomato pericarp cell size, and is associated with increases in cell ploidy levels (Mu et al., 2017). Combinations of these major genes, affecting different parts of te fruit, greatly dictate the final fruit shape and size in modern tomato.

Despite the knowledge of how fruit size is controlled in tomato, much remains unknown. Even though recent introgressions from distant wild relatives for disease resistance, most modern tomato breeding is currently performed with a narrow germplasm that made it to Europe in the

16th century after the second domestication bottleneck (Jenkins, 1948). Thus, many fruit quality traits are fixed in elite germplasm. Finding new sources of variation would greatly benefit the breeding efforts for new varieties with novel characteristics or improved features. The main objective of this study was to characterize in depth underutilized tomato germplasm comprised of SP, SLC, and SLL from South and Central America, and Mexico with focus on fruit weight components. By doing so, we attempted to simplify fruit weight by separating it into its components, much like it is done with yield (Chernet & Zibelo, 2014). This resource should create the basis for mapping new genes that affect fruit quality traits. We also investigated the association of known genes with fruit weight and fruit weight components, the relationship of weight components to weight, and conclude with the proposed evolution history of tomato.

Materials and methods

Plant materials

A panel of 167 accessions comprised of 119 semi-domesticated *S. lycopersicum* var. *cerasiforme* (SLC), 28 wild S. *pimpinellifolium* (SP), and 20 ancestral landraces *Solanum lycopersicum* var. *lycopersicum* (SLL) was assembled by Joaquin Canizares, Maria Jose Diez, and Jose Blanca (collaborators from the Institute for Conservation & Improvement of Valentian Agrodiversity (COMAV), Valencia, Spain). Most accessions are SLC since they are more likely to have alleles for desired traits that were left behind during the domestication bottleneck. These accessions are distinct from accessions used in other sequencing or breeding projects at present time (a unique panel of accessions selected based on the results from Blanca et al., 2015). Most accessions are from the COMAV germplasm bank, and the rest are from TGRC (UC Davis, CA), USDA (Geneva, NY), AVRDC (Taiwan), and CATIE (Costa Rica).

The 119 SLC entries represented the most diverse accessions from the available collection of cerasiforme (unique SLC panel of accessions selected based on the results from Blanca et al., 2015). From these, 34 are from Peru, 41 from Ecuador, 23 from Mexico, and the remaining are from Colombia and countries in Central America (Costa Rica, Honduras, Nicaragua, and El Salvador). There were 12 SP from Peru and 14 from Ecuador. The SLL group was mostly from Mexico (16 accessions), with one accession from Nicaragua. These accessions represent areas close to the center the species origin, as well as Mexico, where SLL arose. They also show large diversity in morphological components, with tremendous differences in fruit weight. The accessions went through 2 generations of single seed descendant in Valencia, Spain, and Wooster, Ohio before the seeds were bulked for use in this study.

Experimental design

In the summer of 2016, the accessions were grown in Athens, GA for phenotyping. The plants were grown in a randomized complete block design (RCBD) with 3 plants each for a total of 2 replicates. The software JMP (JMP®, Version *13.2.0*.) was used to randomize the accessions. The 167 accessions were planted approximately 2 feet apart from each other in the same row. The row spacing between different experimental units was about 6 feet. The same accessions were also grown the same summer by collaborators in Life Oak, FL in a completely randomized design (CRD) with 3 plants for 1 replicate. The 167 accessions grown in Georgia and Florida in 2016 showed high correlation for fruit weight ($r=0.92$, $p<0.001$) between experiments despite growth in different environments (data not shown). A total of 36 accessions that displayed phenotypes of interest was regrown during the summer of 2017 at Live Oak, FL. They were planted in a CRD with no replicates and using the same protocols used in Athens in 2016. This subset also showed high reproducibility ($r \geq 0.96$) compared to the other years (Supplementary figures S2.1-S2.2).

One-way ANOVA confirmed that there were no statistical differences across years and reps (p-value=0.8533). All plants were grown following standard tomato farming practices and were watered using drip irrigation.

Phenotype collection and image analysis

Non-derived phenotypes

The "non-derived" phenotypes were traits directly evaluated on the fruits. In total 18 weight components were measured (Supplementary table S2.1). Approximately 40 fruits were harvested from each replicate. These 40 fruits came from one to three plants in each replicate. From these 40, 20 were selected to represent the average to slightly larger weight range per accession. These 20 fruits were bulk-weighted using a VWR-3001E top loading balance. Accessions carrying very small fruit (SP mainly) were weighed using a VWR-64B analytical scale. Extreme large and small fruits were excluded from the analysis to avoid skewing the average phenotypic values. Locule number was counted in 40 fruits per accession. For the evaluation of dry matter, 2-10 fruits (depending on fruit size) were weighed per accession before and after drying in an oven at 70-80 ° C for one week and percent dry matter was calculated (Supplementary table S2.1).

For measuring morphological components, images were generated by scanning sliced fruits. Approximately eight fruits per accession were sliced along the medio-lateral axis at the equatorial plane, and both halves were scanned at 300 dpi using an HP Scanjet G4050 and imported into a computer using HP Scanjet G4050 Photo Scanner Full Feature Software and Driver version 14.5.1. For bigger fruited accessions, only a single half per tomato was scanned. A cardboard box was used to create a dark background when scanning.

All the images generated from the scanning were analyzed using the software Tomato Analyzer 4.0 (version 4.0 unreleased, earlier version described in (Rodríguez et al., 2010). Tomato area, pericarp area, pericarp+septum area, and columella+placenta area, were measured (Supplementary table S2.1). The ratios of these phenotypes to the total fruit area were also calculated. For most accessions eight unique tomato halves were selected, adjusted, and measured. Each subsequent phenotype was measured using the same set of fruits. For columella+placenta measurements only the half tomato with the greater area (if they were not the same area to start with) was measured. A smaller area in one half or the other is likely the result of a non-centered cut, and therefore, the larger half would be used for the analysis. Highly loculed accessions (9 accessions total) were not analyzed for these traits since the distinction between columella, septum, and pericarp was not clearly discernable in each fruit.

Four tomatoes at the breaker or mature green stages were collected per accession for the phenotyping of the cellular components such as cell size and number. Ripe fruits were avoided because they were too soft and sample collection damaged the cells. For each fruit, three high quality 0.5-1 mm thick pericarp slices (12 total per accession) were collected by hand using double-edge razor blades that were changed frequently. All slices were collected from the equatorial region of the tomatoes where for two tomatoes, the slices were taken in the medio-lateral axis, and for the other two the slices were taken in the proximal-distal axis. For the accessions grown on 2017, the slices were collected only in the proximal-distal axis. This is because we did not find any statistical difference in cell size taken from the proximal-distal or the medio-lateral axis. The sliced sections were stained with a solution containing one part 0.5% Toluidine Blue to two parts distilled water for a few seconds. The sections were rinsed with ddH2O. Images of the stained sections were taken using an Olympus DP70 camera that was

mounted on an OLYMPUS MVX10 optical microscope using an Olympus MVX-TVO.63XC adapter. Each picture that was taken included a ruler to scale cell size and pericarp thickness correctly. Pictures were taken at 10X to 30X magnification, depending on the size of the tomato pericarp.

All the images generated from the microscopy were analyzed using the software ImageJ bundled with 64-bit Java 1.8.0_112 (Schneider, Rasband, & Eliceiri, 2012) and the MorphoLibJ plugin (Legland, Arganda-Carreras, & Andrey, 2016). Since not all images were taken at the same magnification, the images were calibrated using the scale prior to obtaining the measurements. "Maximum pericarp cell size" was obtained by measuring the five largest visible cells in each slice (12 reps leading to 60 cells per accession). Number of cell layers in the pericarp was obtained by tracing three lines that were drawn perpendicular to the exoderm, and cells that intersected these lines were counted (36 lines per accession). The requirements for drawing the lines were as follows: vascular bundles were avoided when tracing the lanes; and the 2-3 cell layers below the exoderm and the endoderm layer were not counted since in many cases they are not clearly visible which could skew the results between the accessions. If the pericarp thickness was irregular, the first line was placed in the thickest region, the second in the thinnest region, and the last in an intermediate region. The length of these lines was recorded and used as a measure of pericarp thickness.

Derived phenotypes

Certain phenotypes were calculated from the directly measured traits and were designated as "derived phenotypes". Pericarp cell layers and pericarp thickness were used to calculate the number of cells per 1 mm length (Supplementary table S2.1). These cell number per mm measurements were employed to calculate "Circumference cell number 1" (Supplementary table

S2.1). “Circumference cell number 2” was calculated by calculating the diameter of the “Maximum pericarp cell size” using the circle area formula (Supplementary table S2.1). “Pericarp area ratio”, “Pericarp+septum area”, and “Columella+placenta area” were calculated by dividing the phenotype in question by the total fruit area (Supplementary table S2.1).

Ovary collection, staining, and phenotyping

Approximately 10 flowers were collected at the anthesis stage from the 36 accessions regrown during the summer of 2017 in Live Oak, FL. The flowers were brought back to the laboratory for Propidium Iodide (PI) staining and ethanol series. The samples and reagents were kept at 4°C or on ice. On Day 1, petals and sepals were removed, and the ovaries were sliced in half along the proximal distal axis, and fixed in FAA (50% Ethanol, 10% 37% Formaldehyde, 5% glacial acetic acid). On Day 2, the FAA solution was decanted, 50% ethanol was added to submerge all ovaries (~2mL), and they were gently mixed for 30-60 min using a MaxQ 3000 shaker. This step was repeated increasing ethanol concentration to 70%, 85%, 95%, and twice in 100%. The samples were left in 100% ethanol at 4°C over night. On day 3, rehydration series took place. A similar process as above was repeated, but this time the ethanol concentration was gradually decreased (95%, 85%, 70%, 50%, 30%, and 15%) and lastly the samples were rinsed twice with ddH_2O for 20 min each. The ovaries were then stained with 2 mL of 1 mg/mL PI stock in 100 mL water for 1 h and rinsed with water twice for 20 min each. On Day 4, the samples were once again dehydrated as described above. Lastly, ovaries were cleared in a 1:1 solution of ethanol: methyl salicylate for 2 h and then kept on methyl salicylate at 4°C until imaging. For each accession at least five good quality ovaries were imaged. Images were taken from the equatorial plane of the ovaries using a Zeiss LSM 880 Confocal Microscope. “Maximum ovary pericarp cell size” (~20 cells per accession) and “number of cell layers in the pericarp ovary” (~12

lines per accession) were measured using a similar protocol as described for mature fruits, but no cell layers were excluded during measuring.

Statistical analysis and graphs

JMP (JMP®, Version *13.2.0*) and R open source software (version 3.3.1; R Core Team 2014) were used to conduct the statistical analysis. Most phenotypes measured were log10 transformed for the statistical analyses to meet the normality requirements for ANOVA and linear regression tests. Other statistical requirements as outliers and population size were checked and dealt with as necessary to conduct the best possible analysis. The correlation matrix was created using the Corrplot Package in R (N=147, heterozygous or missing/dropped data were excluded) (Wei & Simko, 2016) and the correlation network was plotted using the qgraph package in R (Epskamp, Cramer, Waldorp, Schmittmann, & Borsboom, 2012). Partial correlations which control for the effect of other variables were calculated using the ppcor package in R (Kim, 2015) (data not shown). Histograms, mosaic plots, and box plots were created in JMP and R.

A total of 10 accessions heterozygous for any of the five known fruit weight genes were removed from the statistical analyses (Supplementary figure S2.3). For testing which of the five known genes were statistically significant in predicting the different phenotypes, ANOVA with Type II SS were used. This way the model variance was partitioned in what was uniquely explained by each gene. Initially, ANOVA with Type I Sums of Squares were used to determine how much each gene uniquely contributed to the phenotypes. Also, interactions were tested using ANOVA with Type I SS for each phenotype using a model including the five genes and all the second order interactions were plotted to detect significant interactions. If significant interactions were present, then two models (one with and one without interaction) were compared and the best one was kept.

DNA extraction, whole genome sequencing and genotyping

DNA was extracted from young leaves from the plants that were used to bulk the seeds. Extraction was performed with the Qiagen 96-well DNA extraction kit. DNA libraries were created using the NEBNext Ultra II DNA library prep kit (E7645L) (www.neb.com) and 12 barcoded primers (E7335) (performed by Dr. van der Knaap). Twelve libraries were pooled, and size selected (550-650 bp) using the Pippin at the Molecular and Cellular Imaging Center at OSU. The size selected libraries were sequenced on the Illumina NextSeq 150 PE high throughput at the Georgia Genomics Facility at UGA. All accessions were genotyped with standard molecular markers for the known fruit weight genes CNR , CSR , LC, SlKLUH , and FAS using Kompetitive Allele Specific PCR (KASP) (http://www.lgcgroup.com) or Cleaved Amplified Polymorphic Sequences (CAPS) technologies (Konieczny & Ausubel, 1993) (Supplementary Table S2.2).

GWAS and Phylogenetic analyses

Genome wide Association Studies (GWAS) were conducted on this population by Hamid Razifard under the guidance of Ana Caicedo (collaborators from University of Massachusetts, Amherst, USA). The SNP dataset used for GWAS was derived from the main SNP dataset, then excluding SNPs with Minor Allele Frequency (MAF) < 0.05 and missing rate > 10%. GWAS was conducted using a mixed linear model provided in GEMMA (Zhou & Stephens, 2012). The associations were adjusted for population structure using a genetic relatedness matrix created also using GEMMA. P-values from Likelihood Ratio Test (LRT) were used for assessing significance. The significance cutoff was determined based on the effective number of independent SNPs calculated using GEC (M.-X. Li, Yeung, Cherny, & Sham, 2012). The trait values were quantile-

normalized in R. Due to the strong population structure in the germplasm, the 28 SP accessions were excluded from the final GWAS.

Hamid Razifard also conducted the phylogenetic and population structure analysis where he utilized Whole Genome Sequence Data from our 167 accessions and publicly available sequence data from other 128 accessions (SP, SLC, and SLL) from the Central and South America region (data not shown). The phylogenetic tree was built using the coalescent-based SVDquartets method (Chou et al., 2015) based on 54,726 4D SNPs (four-fold degenerate sites, i.e. the third positions in codons where a substitution to either of the four nucleotides does not change the amino acid coded by that codon) with missing data < 10% in the accessions passing the filtering criteria. The population structure was estimated in fastSTRUCTURE (v.1.0)(Raj, Stephens, & Pritchard, 2014) using 8,871,314 SNPs with < 10% missing data finding an optimal number of ancestral populations ($K = 4$). Populations of each species were delimited based on the results of ancestral structure and phylogenetic analyses. Certain accessions that could not be resolved in specific phylogenetic clades because of admixture were classified as "admixed". The remaining accession clades were named after the geographical provenance of most of accessions (> 50%) composing that population. A separate Peruvian clade was defined (SLC San Martin) because of its distinct phylogenetic placement. Also, the clade with tomatoes representing the mostly modern cultivated varieties was defined as SLL Major. The new phylogenetic groups information was incorporated to the 167 accessions of focus (Figure 2.1A, and Supplementary table S2.3). The groups SLL fresh and SLL processing were incorporated from publicly available data from Blanca et al. 2015. The simplified phylogenetic tree (Figure 2.1B) was built using the R package SNPRelate (Zheng et al., 2012) using recommended settings and vcf files for 10 accessions (one from each of the groups created by Hamid).

Results

Phenotype of tomato accessions that are clustered based on SNP analysis

The population structure analysis (data not shown) led to 10 distinct groups, which were largely assembled based on collection sites (Figure 2.1A). The phylogenetic relationship between these groups showed that the SP is the most ancestral clade while SLC and SLL clades arose more recently (Figure 2.1B). Surprisingly, among the SP species, the SP_S_ECU accessions were as ancestral as SP_PER even though geographically they were not as far south as SP_N_ECU accessions (Figure 2.1). Also, SP separated into three phylogenetic groups even though all SP accessions were geographically close to each other, implying high genetic diversity among tomato's closest wild relatives.

To identify accessions that might carry alleles for crop improvement, we obtained detailed fruit weight and weight-related traits from each accession (Supplementary table S2.1 and Supplementary table S2.3). The traits were evaluated across the SP, SLC, and SLL accessions and the results showed that in general, fruit weight increased as did many of the fruit weight-related components as SP evolved into SLL (Figure 2.2, Supplementary figures S2.4-S2.16). Most accessions in the entire population produced tomatoes with two locules that weighed 1 to 10 grams, which is typical for ancestral and semi domesticated tomato accessions. Because of this, most of the SLL accessions behaved as outliers relative to SP and SLC.

For several traits such as fresh weight, locule number and area measurements, the distribution clearly showed that SP is the smallest, followed by SLC and then SLL. Also, SP showed the lowest variation, followed by SLC, and last SLL (Supplementary figures S2.4-S2.16). Phenotypic variation increased with the further domestication of tomato.

Even though weight varied extensively among the accessions, pericarp thickness, number of cell layers, and maximum cell size showed less variation, and the means were closer to one another compared to fruit weight (Supplementary figures S2.8-S2.10). Overall, the SP, SLC, and SLL differed the least in cell layers and maximum cell size, while showing the greatest differences in fruit weight, locule number, and columella area.

The cellular evaluations of the ovary at anthesis might lead to insights about the developmental and temporal effects of the QTLs on fruit size. Anthesis takes place at the mid-point of development of the organ from the floral meristem to the ripe fruit. Final weight such as controlled by LC and FAS is determined before anthesis, whereas SlKLUH and CSR is determined after anthesis (van der Knaap et al., 2014). At anthesis, maximum cell size and number of cell layers already differed in the tomato pericarps, which in turn resulted in pericarp thickness differences (Figure 2.3A). At breaker stage, however, the cellular differences were more pronounced as the larger fruits showed larger values for most of the phenotypes. Also, the large-fruited accessions did not necessarily have both the largest maximum cell size and highest number of cell layers in the pericarp at ovary or breaker stages. Rather, a combination of these two components determined the final pericarp thickness where heavier fruits clearly featured thicker pericarp (Figure 2.3B). Lastly, since the largest fruits did not have the largest pericarp thickness at anthesis, most of the final weight appeared to be determined post-anthesis in these accessions. Cell size increases between anthesis and mature fruit ranged from 60 to 450 fold, whereas cell layer increases ranged from 1 (no increase) to 1.8 fold, nearly a doubling of cells immediately after anthesis. CNR and SlKLUK were segregating where 15 and 6 accessions respectively carried the derived allele (six accessions had the derived allele at both loci, data not shown); however, they did not explain the trends observed in the cellular components.

Furthermore, CSR was fixed in this subset of accessions and therefore, other cell size genes might be segregating in the population to account for the huge range in cell size increases post-anthesis. Moreover, the relative increase from ovary to mature fruit clearly showed that the cell size increase was the greatest driver of pericarp thickness after anthesis, as in certain accessions cell size increased almost 500 fold (Figure 2.3C).

Known predictors of fruit weight correlate to weight and weight-related traits

The five known fruit weight genes (CNR, SlKLUH, CSR, LC, and FAS) were genotyped in the population to determine how well they were associated with fruit weight (Table 2.1). CSR and LC were highly significantly associated with all the traits. CNR was significant for all traits except pericarp cell layers. SlKLUH was barely significant and only for fruit weight (p-value < 0.03). FAS was significant for fruit weight (p-value < 0.002) and highly significant for locule number (p-value < 2.20E-16).

In this population, CSR and LC were the most significant predictors of fruit weight. None of the major genes were associated with the ovary cellular phenotypes (data not shown) which may be expected since the known fruit weight genes that impact cell division (CNR and SlKLUH) or cell size (CNR) do so after anthesis (van der Knaap et al., 2014). Surprisingly, even though SlKLUH has been shown to regulate weight by increasing cell layers (Chakrabarti et al, 2013) this locus was not associated with the trait in this population. Likewise, CNR is thought to regulate cell number (Frary et al., 2000) but again in this population, it was not associated with cell layers and instead with cell size. On the other hand, CSR is known to regulate cell size (Mu et al., 2017) and the gene was associated with cell size.

When accessions were classified based on the allele combinations of CNR, SlKLUH, CSR, LC, and FAS, the observed phenotypic variation suggested the presence of additional loci

controlling weight in tomato (Supplementary Table S2.3, Supplementary figures S2.17-S2.24). Within the groups that carried the same allele for all five genes, large weight and weight related trait differences were observed. For example, 58 accessions that carried the wild allele at all five loci carried fruit of up to 14.95 g.

Morphological traits that predict final fruit weight

The correlation coefficients among the traits showed that fruit area measurements and locule number were highly correlated with fruit weight ($r \geq 0.86$) (Figure 2.4). On the other hand, only the columella+placenta area ratio was highly correlated to weight, suggesting that only columella and placenta were disproportionally enlarged to account for fruit weight increases. Moreover, the area ratio measurements did not explain actual fruit size and hence were not good indicators of fruit weight. For example, a 100 g and a 1 g fruit could both have identical value for the pericarp area over the total area respectively. However, the area ratios were rather useful in identifying fruit mass allocation. When partial correlations were conducted (data not shown) the correlation coefficients among the phenotypes decreased to nearly zero for most correlations indicating that many of the phenotypes are correlated to each other simply because they are correlated to the other variables (e.g. weight).

Percent dry matter was negatively correlated to all the fruit weight and weight-related traits where fruits of smaller weight showed higher dry weight percentage (Figure 2.4). The pericarp components such as number of cell layers ($r=0.81$, p-value<0.001) and cell size ($r=0.90$, p-value<0.001) were highly correlated to pericarp thickness which was expected because as these components increase, so should the pericarp thickness. Circumference cell number 1 and Circumference cell number 2 were positively correlated to fruit weight ($r=0.54$, p-value<0.001

and r=0.77, p-value<0.001 respectively), and highly correlated to each other (r=0.81, p-value<0.001).

When comparing the SP accessions alone, certain weight components showed weak correlations to overall fruit weight (Supplementary figure S2.25). One exception was columella+placenta area and columella+placenta area ratio, suggesting that the changes in weight in this SP germplasm were nearly entirely driven by enlargement of these specific tissues. In SLL in addition to the area measurements, circumference cell number 1 and 2 were highly associated with weight (Supplementary figure S2.27). Interestingly, pericarp thickness and cell size were negatively correlated to fruit weight. Also, pericarp area ratio and pericarp+septum area ratio showed a strong negative correlation to weight and most other weight components. In SLL the enlargement in columella and placenta area was accompanied by a decrease in pericarp and septum area (and the cellular components of the pericarp). Also, the sample size for SP (n=27) and SLL (n=16) was low compared to SLC (n=104), and this could explain some of the low or negative correlations to fruit weight.

While there was no correlation between pericarp and placenta+columella area ratios in entire population (Figure 2.4), the columella and placenta area ratio increased as the pericarp area ratio decreased in SLL (Supplementary figure S2.27). In other words, the fruit size increase was accompanied by a disproportional columella+placenta area increase, relative to the increase in pericarp and septum area (Supplementary figures S2.14- S2.16).

Allele and phenotypic distributions among the subpopulations including modern tomatoes

We sought to determine the distribution of the alleles of the known genes in the ancestral and semi-domesticated germplasm as well as in modern processing and fresh market tomatoes. Since we know the causal mutations for each gene and only two alleles have been reported for

each gene, this information should shed light on when during domestication the alleles arose as well as their impact on the phenotypes. The genotyping results showed that a higher number of ancestral and semi-domesticated accessions carried the derived alleles for CNR, SlKLUH, and LC compared to CSR and FAS, suggesting that the latter two alleles arose later during domestication (Figure 2.5). SLL accessions carried the most derived alleles and SP the least, indicating strong selection in SLL. All SP accessions carried the wild-type allele for the known fruit weight genes except for LC where two SP N ECU accessions carried the derived allele for this gene. The SLC from Peru carried more LC and CNR derived alleles than SLC from Ecuador and Mexico. On the other hand, Ecuador showed the most derived alleles for SlKLUH among the countries. For SLL MEX and SLL Major, CSR, CNR and LC were nearly fixed for the derived allele, while only a few accessions carried the derived allele for FAS. This trend in SLL became more evident in the modern SLL accessions. In modern tomatoes CNR, SlKLUH and CSR were nearly fixed for the derived allele, and FAS was nearly fixed for the wild allele. The major difference between fresh and processing tomatoes was for LC, where the derived allele is common in fresh market tomatoes while the wild allele is common in processing tomatoes.

In SLC, weight and all weight related components were higher in the accessions from South America compared to accessions from Central America and Mexico (Supplementary figures S2.28-S2.35). This pattern suggested a reversal in domestication as tomato migrated north from its center of origin. The groups SLC ECU, SLC PER, and SLC San Martin showed higher phenotypic values for weight and weight related components relative to the SLC MEX CA NSA and SLC MEX groups. This trend was least evident for locule number, where the differences were minimal and the SLC ECU accessions showed large variation.

QTL detected by GWAS

GWAS identified QTL and the respective significantly associated SNPs for most traits except for cell size, pericarp area, and pericarp+placenta area (Supplementary table S2.4, Manhattan plots not shown). For most chromosomes, certain significant SNPs were shared between fruit weight and the other weight components QTL. However, there were no significant SNP for fruit weight on chromosome 11, while this chromosome harbors both FAS and CSR. The five known fruit weight genes CNR, SlKLUH, CSR, LC, and FAS were also detected in the GWAS. Significant SNPs near the causal SNPs or polymorphisms were found in case the causal SNP was not highly significant.

Discussion

Fruit weight and many of the weight components greatly increased from SP to SLL during tomato domestication, agreeing with previously published work. In this population, the area components and locule number were the best predictors of fruit weight and these traits were highly correlated to each other. The ovary analysis indicated that even though both cell size and cell layers contribute to pericarp thickness, cell size increase plays a more prominent role especially after fruit set. Also, while there were already cellular components differences at anthesis, it is not until the fruit development stage that the larger fruited accessions show larger differences in the weight related phenotypes. Only four of the known weight genes (CNR, CSR, LC, and FAS) were highly associated with weight. Many of the weight-related traits suggested the presence of unknown genes that modify or mask the effects of SlKLUH since the effect of this gene is negligible in this population. Also, the major genes did not explain much of the phenotypic variation in fruit weight, further suggesting the presence of unknown genes (Supplementary Table S2.3). Therefore, this study has given insights about tomato domestication,

how weight components contribute to overall fruit weight, and this will serve as a tool for breeders that will allow the discovery of new fruit weight genes that were likely left behind during tomato domestication or have simply not been discovered yet. Future work will be aimed to create mapping populations to detect and fine map these potential genes.

Emergence of derived alleles and selection of fruit weight during domestication

Altogether, the allele frequency of CNR, SlKLUH, CSR, LC, and FAS and the phenotypes in this population gave insights about the history of tomato domestication and the origin of the mutations responsible for this process. The fruit weight increases from SP to SLL was largely explained by the allele distribution for the major fruit weight genes. The frequency for the derived alleles of CNR, SlKLUH, CSR, and LC was the highest in SLL, demonstrating strong selection for these genes in Mexico. There was also an increase in phenotypic variation as the result of domestication; a trend that has been observed in many crops (Boster, 1985; Frary & Doğanlar, 2003).

The high frequency of the derived allele for CNR, LC, and SlKLUH suggests that these alleles arose earlier in tomato domestication or were highly selected for compared to CSR and FAS. Based on the allelic distributions of CSR and FAS, the derived allele likely arose in Peruvian and Ecuadorian SLC and eventually made it to Mexico. CSR was then highly selected, meaning it was a crucial mutation in the origin of SLL. On the other hand, FAS appeared to be deselected over time. The FAS mutation produces fruits with high locule number featuring uneven shapes. Consumers prefer uniform shaped fruits rather than highly loculed and uneven shaped fruits, which is why selections were made against this trait (van der Knaap, 2013). This is supported by the FAS allele distribution in modern tomato varieties, where the derived allele was in extreme low frequency in fresh tomatoes and nonexistent in processing tomatoes. Even though

our population is smaller than those used in previous studies, our findings further support previous research (Blanca et al. 2015). Also, our study incorporates novel information about the recently cloned gene CSR (Mu et al., 2017), adding a missing piece of information to the tomato domestication history. While the overall tomato domestication trend depicted in our study largely agrees with the work of Blanca et al. 2015, there is a discrepancy. We observed a decrease in CNR, SlKLUH and LC derived-allele frequency and phenotypic components in Mexican SLC relative to Peruvian and Ecuadorian SLC, which is unusual when assuming that domestication is always going in one direction and an increase in fruit weight and weight components was expected. I propose three plausible explanations to our findings. First, fruit weight might have been deselected in Central America and Mexico. From the human selection perspective, smaller fruits generally ripen faster than bigger fruits, which can speed up generation time. Also, humans might have selected smaller fruits for niche specific needs as cooking, transportation, or weather conditions. From the animal selection perspective, rodents and birds may be better at handling smaller fruits than larger fruits, leading to better dispersal of small-fruited plants. Second, because the selection bottlenecks, only the smaller fruited SLC accessions made it to Mexico by chance. It could also be that a more extensive sampling or a bigger population size is needed. While there are 34 Peruvian and 41 Ecuadorian SLC accessions, only a total of 43 accessions represent the SLC from Colombia, Costa Rica, Salvador, Honduras, Nicaragua, and Mexico, where 23 are from Mexico (Figure 2.1). Including more accessions from these countries in future studies would test the validity of our findings and provide a better understanding of tomato domestication. Third, and most likely, we missed some key event in the tomato domestication history. This could be undiscovered major trading events among these countries, pre-Columbian human selection for different religious or practical reasons, or some other natural selection pressure. Anthropological

accounts about the tomato history are lacking, which makes it hard to further investigate the origin of the tomato and to settle these discrepancies.

Fruit weight components dictating fruit weight

In tomato as in other crops, the fruit size increase from wild to domesticated types was accompanied by an increase in area in different parts of the fruit, a locule number increase, and an increase in water content resulting in larger fruits. However, the increase in fruit area was not uniform in all parts of the fruits. Our data showed a disproportional increase in columella+placenta area relative to pericarp+septum area, which was underscored by the distribution of CNR in SLL (Figure 2.5). Furthermore, our correlations indicated high selection for cell circumference number later during domestication in SLL; providing a good opportunity to map new genes controlling this trait. Also, the partial correlations showed that most of our strong correlations were the result of fruit weight, where a larger fruit inherently has larger fruit weight attributes than a smaller fruit.

The increase in columella and placenta area (and the strong correlation to fruit weight) could also be the result of an increase in locule number driven by LC, where accessions with increased locules have both a larger columella+placenta area and a higher weight. However, we excluded accessions with extreme locule number from the analysis, and the increase in columella+placenta may be due to novel genes that regulate the trait. Furthermore, these novel genes could have been the drivers of SLL evolution from SLC.

Our findings showed that most of the weight components explain the increase in fruit weight since they were highly correlated. Most fruits were round and uniform and therefore, cell layers and size (1 dimension) increased without any effect on fruit elongation. If the accessions differed in shape (round to elongated), weight could have varied even though cell layer, cell size,

and fruit areas were the same. From all the weight components, the relationships between maximum cell size, cell layers number, and pericarp thickness were the most complex. In some cases, fruits with similar pericarp thickness showed different values for pericarp cell layers and cell size, yet these two components led to yield similar thickness. This was observed both in the ovary and in mature fruits for the accessions that were evaluated. The relationship between the cellular components at anthesis and mature fruit indicated that the expansion of the pericarp area by increased cell size drove fruit weight differences post anthesis, even when CSR is fixed. This information provides an opportunity to map other genes controlling cellular processes in the pericarp, in particular those that affect cell size.

Relationships of the major known genes to weight and fruit weight components and QTL mapping possibilities

The effect of CNR, SlKLUH, CSR, LC, and FAS genes on fruit weight agrees largely with previous findings. While CNR, CSR, and LC have been associated with weight in previous studies (Lin et al., 2014), many of the fruit weight components had not been previously reported (Table 2.1). Also, it was surprising that both LC and FAS are associated with weight and locule number, but only LC is associated with all the other components. An upregulation of LC (SlWUSHEL) causes meristem enlargement likely by an increase in cell number (Brand, Fletcher, Hobe, Meyerowitz, & Simon, 2000), and maybe this was the case for most accessions and the effect was detectable on mature fruits. Since FAS (SlCLV3) and LC are in a feedback loop in the same pathway (Somssich, Je, Simon, & Jackson, 2016; Xu et al., 2015) this would imply that FAS (SlCLV3) was downregulated, and hence it was not associated with other fruit weight components.

Our associations were also likely obscured by the great fruit weight range in this population. Certain fruits weighed a few grams while others were close to 100 g, and a fruit that is 50 times larger has larger fruit weight components (whether proportionally larger or not). Even though this weight difference is greatly driven by the known fruit weight genes, it is hard to tease their effects apart since they are segregating in diverse genetic backgrounds. Also, the low association of SlKLUH with weight indicates that genes identified from biparental mapping projects are not always highly associated with the same traits in GWAS populations. This gene is not even associated with the pericarp+septum area trait even though it affects cell number in the pericarp and septum of the fruit (Chakrabarti et al., 2013). Thus, it is likely that unknown genes or epigenetic modifiers in this germplasm interact or mask the effect of the major genes such as SlKLUH. This is further supported by the large amount of phenotypic variation not explained by the major genes.

While we detected the major fruit weight genes with GWAS and there was overlap between these loci and the other fruit weight components, this tool was not powerful enough to use fruit weight components as proxy for overall fruit weight. The idea was that the components of weight would be more heritable and qualitative than overall weight. The inherent population structure in our germplasm, the quantitative nature of weight, and the large fruit weight differences and segregating known fruit weight genes made the QTL detection in the GWAS population challenging. Furthermore, we detected a large number of polymorphisms in the FW2.2 locus and many of the small effect QTL reported in this region (Lecomte et al., 2004) are likely segregating, further obscuring our analysis. However, even though GWAS were unable to detect QTL for some traits, the results can be used to guide future mapping efforts for the other traits with identified QTL. With less quantitative loci segregating, however, GWAS should be powerful

enough to detect QTL and facilitate the development of mapping populations since there would be fewer but larger effect QTL.

This germplasm provides a unique opportunity to map new genes that control weight and weight related traits. The combination of in depth phenotyping, knowledge of the alleles for the major fruit weight genes, phenotype consistency, and the GWAS results create a unique and powerful tool for mapping new genes that were left behind during domestication or have not yet been cloned. It is important to note, however, that there are chances that some of these "new" genes could be already fixed in modern cultivated tomato; which would explain why they have not been identified in other mapping studies.

Literature cited

Acquaah, George. (2009) Principles of plant genetics and breeding. John Wiley & Sons. 667-676.

Bai, Y., & Lindhout, P. (2007). Domestication and Breeding of Tomatoes: What have We Gained and What Can We Gain in the Future? *Annals of Botany*, *100*(5), 1085–1094. https://doi.org/10.1093/aob/mcm150

Barrero, L. S., & Tanksley, S. D. (2004). Evaluating the genetic basis of multiple-locule fruit in a broad cross section of tomato cultivars. *Theoretical and Applied Genetics*, *109*(3), 669–679. https://doi.org/10.1007/s00122-004-1676-y

Bernacchi, D., Beck-Bunn, T., Eshed, Y., Lopez, J., Petiard, V., Uhlig, J., … Tanksley, S. (1998). Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from Lycopersicon hirsutum. *Theoretical and Applied Genetics*, *97*(3), 381–397. https://doi.org/10.1007/s001220050908

Berry, S. Z., Gould, W. A., & Wiese, K. L. (1991). "Ohio 8245" processing tomato. *HortScience : A Publication of the American Society for Horticultural Science*. Retrieved from http://agris.fao.org/agris-search/search.do?recordID=US19950152001

Blanca, J., Montero-Pau, J., Sauvage, C., Bauchet, G., Illa, E., Díez, M. J., … Cañizares, J. (2015). Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC Genomics*, *16*, 257. https://doi.org/10.1186/s12864-015-1444-1

Boster, J. S. (1985). Selection for perceptual distinctiveness: Evidence from aguaruna cultivars of<Emphasis Type="Italic">Manihot esculenta</Emphasis>. *Economic Botany*, *39*(3), 310–325. https://doi.org/10.1007/BF02858802

Bourdon, M., Frangne, N., Mathieu-Rivet, E., Nafati, M., Cheniclet, C., Renaudin, J.-P., & Chevalier, C. (2010). Endoreduplication and Growth of Fleshy Fruits. In *Progress in Botany 71* (pp. 101–132). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02167-1_4

Brand, U., Fletcher, J. C., Hobe, M., Meyerowitz, E. M., & Simon, R. (2000). Dependence of Stem Cell Fate in Arabidopsis on a Feedback Loop Regulated by CLV3 Activity. *Science*, *289*(5479), 617–619. https://doi.org/10.1126/science.289.5479.617

Broman, K. W., Wu, H., Sen, Ś., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, *19*(7), 889–890. https://doi.org/10.1093/bioinformatics/btg112

Capel, C., Carmen, A. F. del, Alba, J. M., Lima-Silva, V., Hernández-Gras, F., Salinas, M., … Lozano, R. (2015). Wide-genome QTL mapping of fruit quality traits in a tomato RIL population derived from the wild-relative species <Emphasis Type="Italic">Solanum pimpinellifolium</Emphasis> L. *Theoretical and Applied Genetics*, *128*(10), 2019–2035. https://doi.org/10.1007/s00122-015-2563-4

Causse, M., Saliba-Colombani, V., Lecomte, L., Duffé, P., Rousselle, P., & Buret, M. (2002). QTL analysis of fruit quality in fresh market tomato: a few chromosome regions control the variation of sensory and instrumental traits. *Journal of Experimental Botany*, *53*(377), 2089–2098.

Chakrabarti, M., Zhang, N., Sauvage, C., Muños, S., Blanca, J., Cañizares, J., … van der Knaap, E. (2013). A cytochrome P450 regulates a domestication trait in cultivated tomato. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(42), 17125–17130. https://doi.org/10.1073/pnas.1307313110

Chernet, S., & Zibelo, H. (2014). Evaluation of Tomato Varieties for Fruit Yield and Yield Components in Western Lowland of Tigray, Northern Ethiopia - SciAlert Responsive Version. https://doi.org/10.3923/ijar.2014.259.264

Chou, J., Gupta, A., Yaduvanshi, S., Davidson, R., Nute, M., Mirarab, S., & Warnow, T. (2015). A comparative study of SVDquartets and other coalescent-based species tree estimation methods. *BMC Genomics*, *16*(10), S2. https://doi.org/10.1186/1471-2164-16-S10-S2

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). Qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*. Retrieved from https://dare.uva.nl/search?identifier=cb3d90cf-240a-4838-ab68-603c9ee0013a

Food and Agriculture Organization of the United Nations. (2016). FAOSTAT. Rome, Italy:FAO.

Foolad, M. R. (2007). Genome Mapping and Molecular Breeding of Tomato. *International Journal of Plant Genomics*, *2007*. https://doi.org/10.1155/2007/64358

Fox, J., & Weisberg, S. (2011). *An {R} Companion to Applied Regression* (Second). Thousand Oaks CA: Sage. Retrieved from http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Frary, A., & Doğanlar, S. (2003). Comparative Genetics of Crop Plant Domestication and Evolution. *TURKISH JOURNAL OF AGRICULTURE AND FORESTRY*, *27*(2), 59–69.

Frary, A., Nesbitt, T. C., Frary, A., Grandillo, S., Knaap, E. van der, Cong, B., … Tanksley, S. D. (2000). fw2.2: A Quantitative Trait Locus Key to the Evolution of Tomato Fruit Size. *Science*, *289*(5476), 85–88. https://doi.org/10.1126/science.289.5476.85

Grandillo, S., Ku, H. M., & Tanksley, S. D. (1999). Identifying the loci responsible for natural variation in fruit size and shape in tomato. *Theoretical and Applied Genetics*, *99*(6), 978–987. https://doi.org/10.1007/s001220051405

Grandillo, S., & Tanksley, S. D. (1996). QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species <Emphasis Type="Italic">Lycopersicon pimpinellifolium</Emphasis>. *Theoretical and Applied Genetics*, *92*(8), 935–951. https://doi.org/10.1007/BF00224033

Illa-Berenguer, E., Van Houten, J., Huang, Z., & van der Knaap, E. (2015). Rapid and reliable identification of tomato fruit weight and locule number loci by QTL-seq. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, *128*(7), 1329–1342. https://doi.org/10.1007/s00122-015-2509-x

Jenkins, J. A. (1948). The Origin of the Cultivated Tomato. *Economic Botany*, *2*(4), 379–392.

Kim, S. (2015). ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *Communications for Statistical Applications and Methods*, *22*(6), 665–674. https://doi.org/10.5351/CSAM.2015.22.6.665

Konieczny, A., & Ausubel, F. M. (1993). A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. *The Plant Journal: For Cell and Molecular Biology*, *4*(2), 403–410.

Lecomte, L., Saliba-Colombani, V., Gautier, A., Gomez-Jimenez, M. C., Duffé, P., Buret, M., & Causse, M. (2004). Fine mapping of QTLs of chromosome 2 affecting the fruit architecture and composition of tomato. *Molecular Breeding*, *13*(1), 1–14. https://doi.org/10.1023/B:MOLB.0000012325.77844.0c

Legland, D., Arganda-Carreras, I., & Andrey, P. (2016). MorphoLibJ: integrated library and plugins for mathematical morphology with ImageJ. *Bioinformatics (Oxford, England)*, *32*(22), 3532–3534. https://doi.org/10.1093/bioinformatics/btw413

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv:1303.3997 [q-Bio]*. Retrieved from http://arxiv.org/abs/1303.3997

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, M.-X., Yeung, J. M. Y., Cherny, S. S., & Sham, P. C. (2012). Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Human Genetics*, *131*(5), 747–756. https://doi.org/10.1007/s00439-011-1118-2

Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., … Huang, S. (2014). Genomic analyses provide insights into the history of tomato breeding. *Nature Genetics*, *46*(11), 1220–1226. https://doi.org/10.1038/ng.3117

Lippman, Z., & Tanksley, S. D. (2001). Dissecting the genetic pathway to extreme fruit size in tomato using a cross between the small-fruited wild species Lycopersicon pimpinellifolium and L. esculentum var. Giant Heirloom. *Genetics*, *158*(1), 413–422.

Mansfeld, B., & Grumet, R. (2017). QTLseqr: An R package for bulk segregant analysis with next-generation sequencing. *BioRxiv*, 208140. https://doi.org/10.1101/208140

Mazzucato, A., Wang, A., Li, C., Zhang, C., Zamir, D., Zhu, G., … Zhang, Z. (2014). Genomic analyses provide insights into the history of tomato breeding. *Nature Genetics*, *46*(11), 1220. https://doi.org/10.1038/ng.3117

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing

next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Mu, Q., Huang, Z., Chakrabarti, M., Illa-Berenguer, E., Liu, X., Wang, Y., … Knaap, E. van der. (2017). Fruit weight is controlled by Cell Size Regulator encoding a novel protein that is expressed in maturing tomato fruits. *PLOS Genetics*, *13*(8), e1006930. https://doi.org/10.1371/journal.pgen.1006930

Muños, S., Ranc, N., Botton, E., Bérard, A., Rolland, S., Duffé, P., … Causse, M. (2011). Increase in tomato locule number is controlled by two single-nucleotide polymorphisms located near WUSCHEL. *Plant Physiology*, *156*(4), 2244–2254. https://doi.org/10.1104/pp.111.173997

Musseau, C., Just, D., Jorly, J., Gévaudant, F., Moing, A., Chevalier, C., … Fernandez, L. (2017). Identification of Two New Mechanisms That Regulate Fruit Growth by Cell Expansion in Tomato. *Frontiers in Plant Science*, *8*, 988. https://doi.org/10.3389/fpls.2017.00988

Peralta, I., Knaap, S., & Spooner, D. (2007). THE TAXONOMY OF TOMATOES: A REVISION OF WILD TOMATOES (SOLANUM L. SECTION LYCOPERSICON (MILL.) WETTST.) AND THEIR OUTGROUP RELATIVES (SOLANUM SECTIONS JUGLANDIFOLIUM (RYDB.) CHILD AND LYCOPERSICOIDES (CHILD) PERALTA), (84), 1–10.

Prudent, M., Causse, M., Génard, M., Tripodi, P., Grandillo, S., & Bertin, N. (2009). Genetic and physiological analysis of tomato fruit weight and composition: influence of carbon availability on QTL detection. *Journal of Experimental Botany*, *60*(3), 923–937. https://doi.org/10.1093/jxb/ern338

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics*, *197*(2), 573–589. https://doi.org/10.1534/genetics.114.164350

Ranc, N., Muños, S., Santoni, S., & Causse, M. (2008). A clarified position for solanum lycopersicum var. cerasiformein the evolutionary history of tomatoes (solanaceae). *BMC Plant Biology*, *8*, 130. https://doi.org/10.1186/1471-2229-8-130

Ranc, N., Muños, S., Xu, J., Paslier, M.-C. L., Chauveau, A., Bounon, R., … Causse, M. (2012). Genome-Wide Association Mapping in Tomato (Solanum lycopersicum) Is Possible Using Genome Admixture of Solanum lycopersicum var. cerasiforme. *G3: Genes|Genomes|Genetics*, *2*(8), 853–864. https://doi.org/10.1534/g3.112.002667

Rodríguez, G. R., Moyseenko, J. B., Robbins, M. D., Morejón, N. H., Francis, D. M., & van der Knaap, E. (2010). Tomato Analyzer: a useful software application to collect accurate and detailed morphological and colorimetric data from two-dimensional objects. *Journal of Visualized Experiments: JoVE*, (37). https://doi.org/10.3791/1856

Rodríguez-Leal, D., Lemmon, Z. H., Man, J., Bartlett, M. E., & Lippman, Z. B. (2017). Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing. *Cell*, *171*(2), 470-480.e8. https://doi.org/10.1016/j.cell.2017.08.030

Roth, I. (1977). *Fruits of Angiosperms*. Stuttgart, Germany: Schweizerbart Science Publishers. Retrieved from http://www.schweizerbart.de//publications/detail/isbn/9783443140106/Handbuch_d_Pflanzenanatomie_10_1_Roth

Ruangrak, E., Su, X., Huang, Z., Wang, X., Guo, Y., Du, Y., & Gao, J. (2018). Fine mapping of a major QTL controlling early flowering in tomato using QTL-seq. *Canadian Journal of Plant Science*, 1–11. https://doi.org/10.1139/cjps-2016-0398

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of Image Analysis. *Nature Methods*, *9*(7), 671–675.

Singh, V. K., Khan, A. W., Jaganathan, D., Thudi, M., Roorkiwal, M., Takagi, H., … Varshney, R. K. (2016). QTL-seq for rapid identification of candidate genes for 100-seed weight and root/total plant dry weight ratio under rainfed conditions in chickpea. *Plant Biotechnology Journal*, *14*(11), 2110–2119. https://doi.org/10.1111/pbi.12567

Somssich, M., Je, B. I., Simon, R., & Jackson, D. (2016). CLAVATA-WUSCHEL signaling in the shoot meristem. *Development*, *143*(18), 3238–3248. https://doi.org/10.1242/dev.133645

Stevens, M. A., & Rick, C. M. (1986). Genetics and breeding. In *The Tomato Crop* (pp. 35–109). Springer, Dordrecht. https://doi.org/10.1007/978-94-009-3137-4_2

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., … Terauchi, R. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *The Plant Journal: For Cell and Molecular Biology*, *74*(1), 174–183. https://doi.org/10.1111/tpj.12105

Tanksley, S. D. (2004). The Genetic, Developmental, and Molecular Bases of Fruit Size and Shape Variation in Tomato. *The Plant Cell*, *16*(suppl 1), S181–S189. https://doi.org/10.1105/tpc.018119

van der Knaap, E., Chakrabarti, M., Chu, Y. H., Clevenger, J. P., Illa-Berenguer, E., Huang, Z., … Wu, S. (2014). What lies beyond the eye: the molecular mechanisms regulating tomato

fruit weight and shape. *Frontiers in Plant Science*, *5*. https://doi.org/10.3389/fpls.2014.00227

Wei, Q., Fu, W., Wang, Y., Qin, X., Wang, J., Li, J., … Chen, J. (2016). Rapid identification of fruit length loci in cucumber (*Cucumis sativus* L.) using next-generation sequencing (NGS)-based QTL analysis. *Scientific Reports*, *6*, 27496. https://doi.org/10.1038/srep27496

Xiao, H., Radovich, C., Welty, N., Hsu, J., Li, D., Meulia, T., & van der Knaap, E. (2009). Integration of tomato reproductive developmental landmarks and expression profiles, and the effect of SUN on fruit shape. *BMC Plant Biology*, *9*, 49. https://doi.org/10.1186/1471-2229-9-49

Xu, C., Liberatore, K. L., MacAlister, C. A., Huang, Z., Chu, Y.-H., Jiang, K., … Lippman, Z. B. (2015). A cascade of arabinosyltransferases controls shoot meristem size in tomato. *Nature Genetics*, *47*(7), 784–792. https://doi.org/10.1038/ng.3309

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Zhou, X., & Stephens, M. (2012). Genome-wide Efficient Mixed Model Analysis for Association Studies. *Nature Genetics*, *44*(7), 821–824. https://doi.org/10.1038/ng.2310

Table 2.1. Major fruit weight genes associated with non-derived fruit weight and fruit weight parameters. Significance codes: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' '

Weight parameter	Gene	Sum Sq	Df	F value	Pr(>F)	
Weight	CNR	1.0081	1	19.116	2.29E-05	***
	SlKLUH	0.2564	1	4.862	0.029	*
	CSR	3.7248	1	70.630	3.07E-14	***
	LC	3.3382	1	63.299	4.01E-13	***
	FAS	0.7308	1	13.858	2.78E-04	***
Locule number	CNR	0.07001	1	10.340	1.60E-03	**
	SlKLUH	0.00243	1	0.359	0.550	
	CSR	0.32454	1	47.932	1.21E-10	***
	LC	1.00324	1	148.169	< 2.2e-16	***
	FAS	0.97917	1	144.613	< 2.2e-16	***
Pericarp area	CNR	0.7696	1	16.669	7.40E-05	***
	SlKLUH	0.1347	1	2.917	0.090	.
	CSR	2.5568	1	55.376	8.71E-12	***
	LC	2.3294	1	50.452	5.41E-11	***
	FAS	0.0745	1	1.613	0.206	
Pericarp + septum area	CNR	0.8608	1	16.883	6.69E-05	***
	SlKLUH	0.1138	1	2.232	0.137	
	CSR	2.8998	1	56.876	5.05E-12	***
	LC	2.9572	1	58.002	3.36E-12	***
	FAS	0.143	1	2.805	0.096	.
Columella + placenta area	CNR	1.8996	1	27.339	5.99E-07	***
	SlKLUH	0.044	1	0.633	0.428	
	CSR	4.8832	1	70.279	4.59E-14	***
	LC	4.381	1	63.050	5.56E-13	***
	FAS	0.2437	1	3.507	0.063	.
Pericarp thickness	CNR	0.1475	1	5.592	0.019	*
	SlKLUH	0.1019	1	3.864	0.051	.
	CSR	0.761	1	28.854	2.97E-07	***
	LC	0.7756	1	29.408	2.33E-07	***
	FAS	0.0017	1	0.065	0.800	
Pericarp cell layers	CNR	0.00297	1	0.589	0.444	
	SlKLUH	0.00165	1	0.327	0.568	
	CSR	0.09881	1	19.570	1.87E-05	***
	LC	0.09823	1	19.455	1.97E-05	***
	FAS	0.0002	1	0.039	0.844	
Maximum cell size	CNR	0.4236	1	10.063	1.84E-03	**
	SlKLUH	0.037	1	0.879	0.350	
	CSR	0.4787	1	11.371	9.52E-04	***
	LC	0.8105	1	19.254	2.16E-05	***
	FAS	0.0149	1	0.355	0.552	

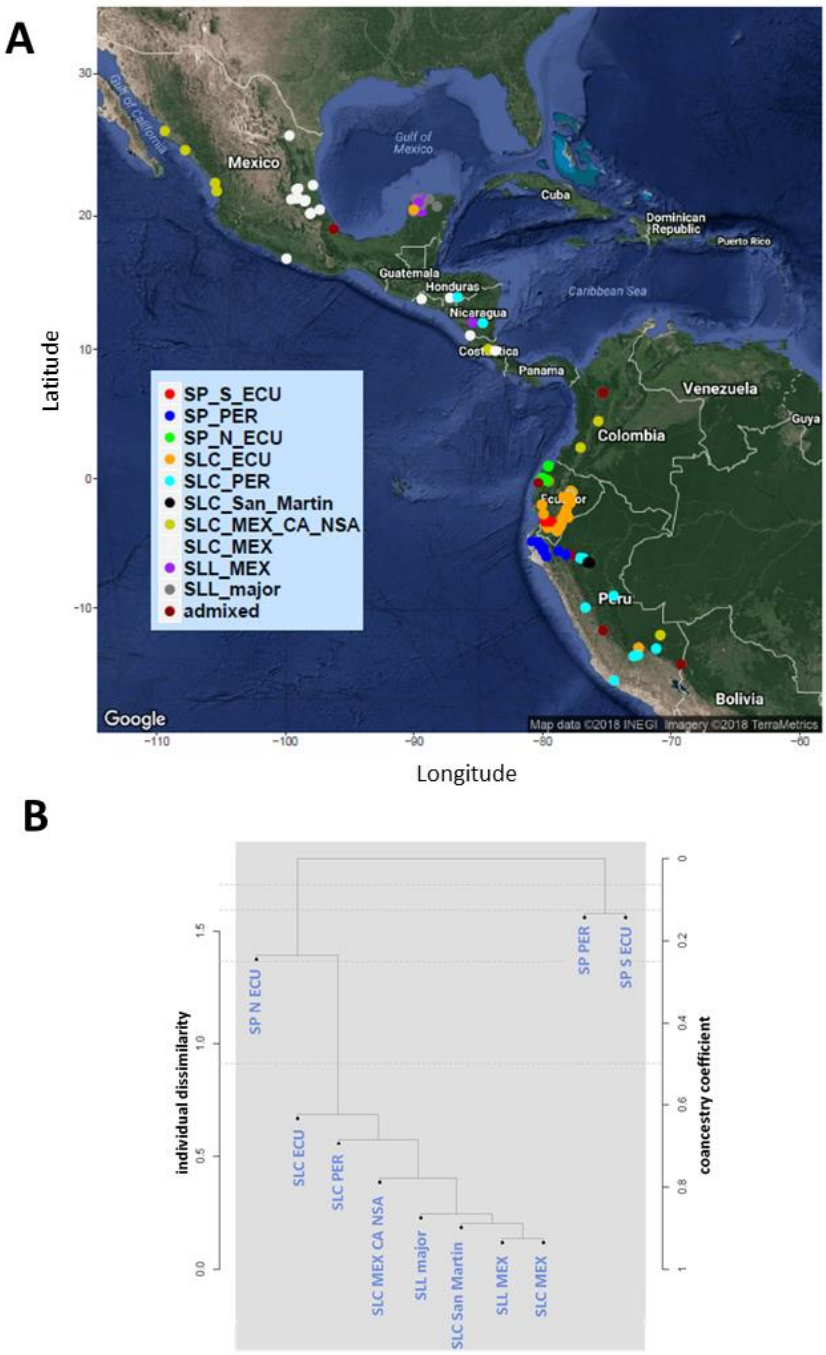


Figure 2.1. Population phylogeny and collecting site. A. Geographical location for the 166 accessions with available passport data. B. Simplified phylogenetic tree (N=10) showing the relationships among the different groups; admixed group was excluded. SP S ECU: Southern Ecuadorian SP, SP N ECU: Northern Ecuadorian SP, SP PER: Peruvian SP, SLC ECU: Ecuadorian SLC, SLC PER: Peruvian SLC, SLC San Martin: SLC from San Martin, Peru, SLC MEX CA NSA: SLC from Mexico, Central America and northern South America, SLC MEX: Mexican SLC, SLL MEX: Mexican SLL, SLL major: the major group of SLL.

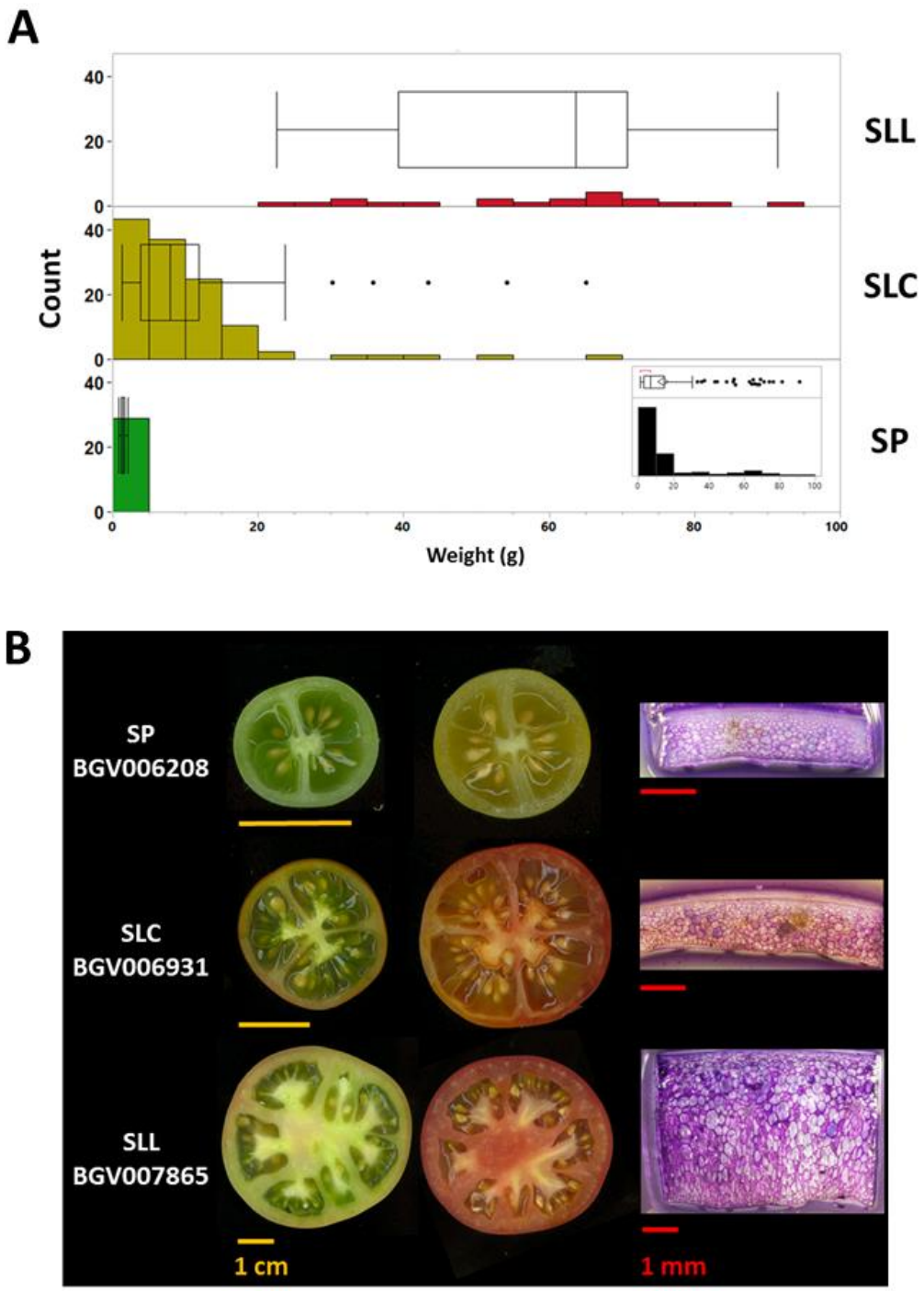
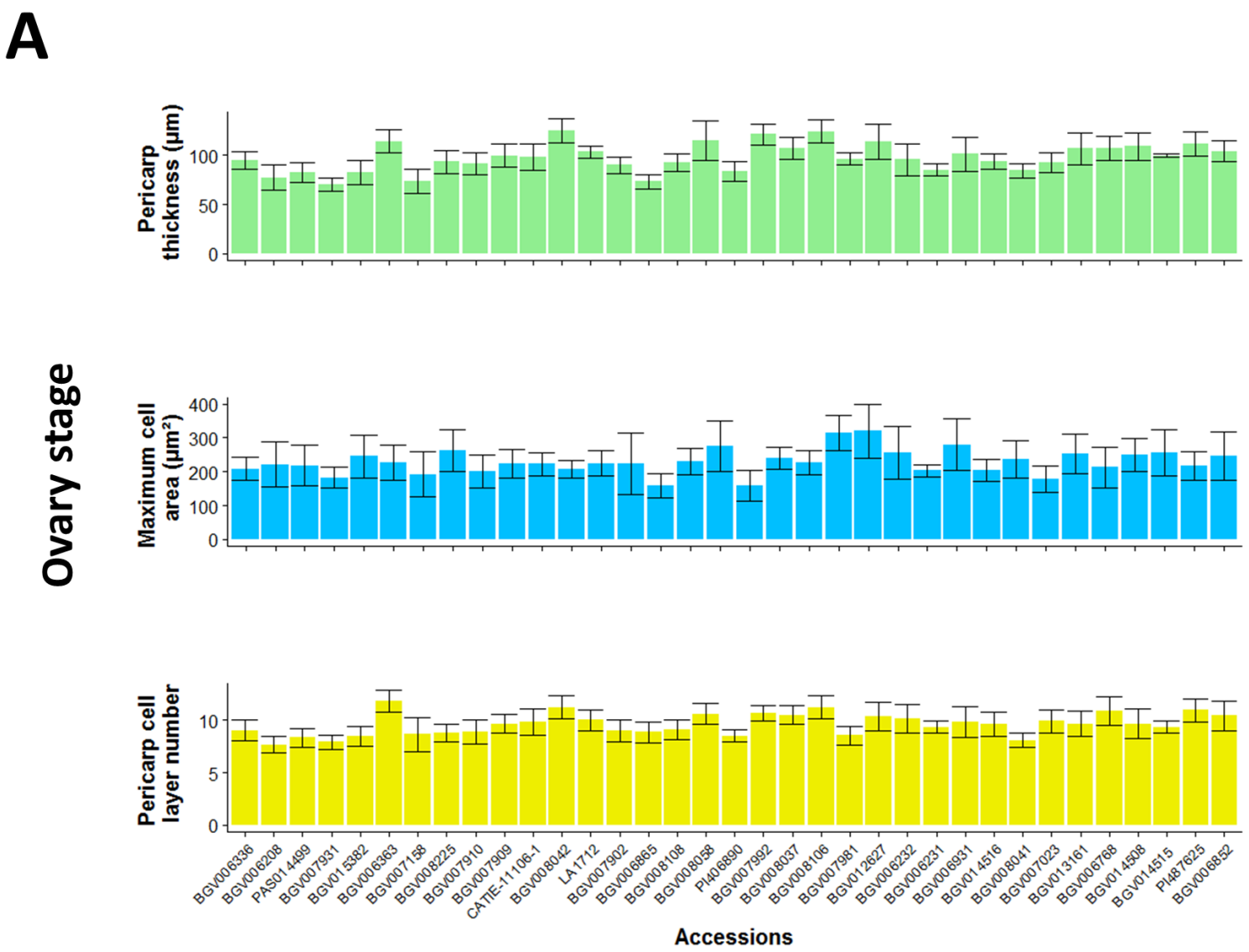
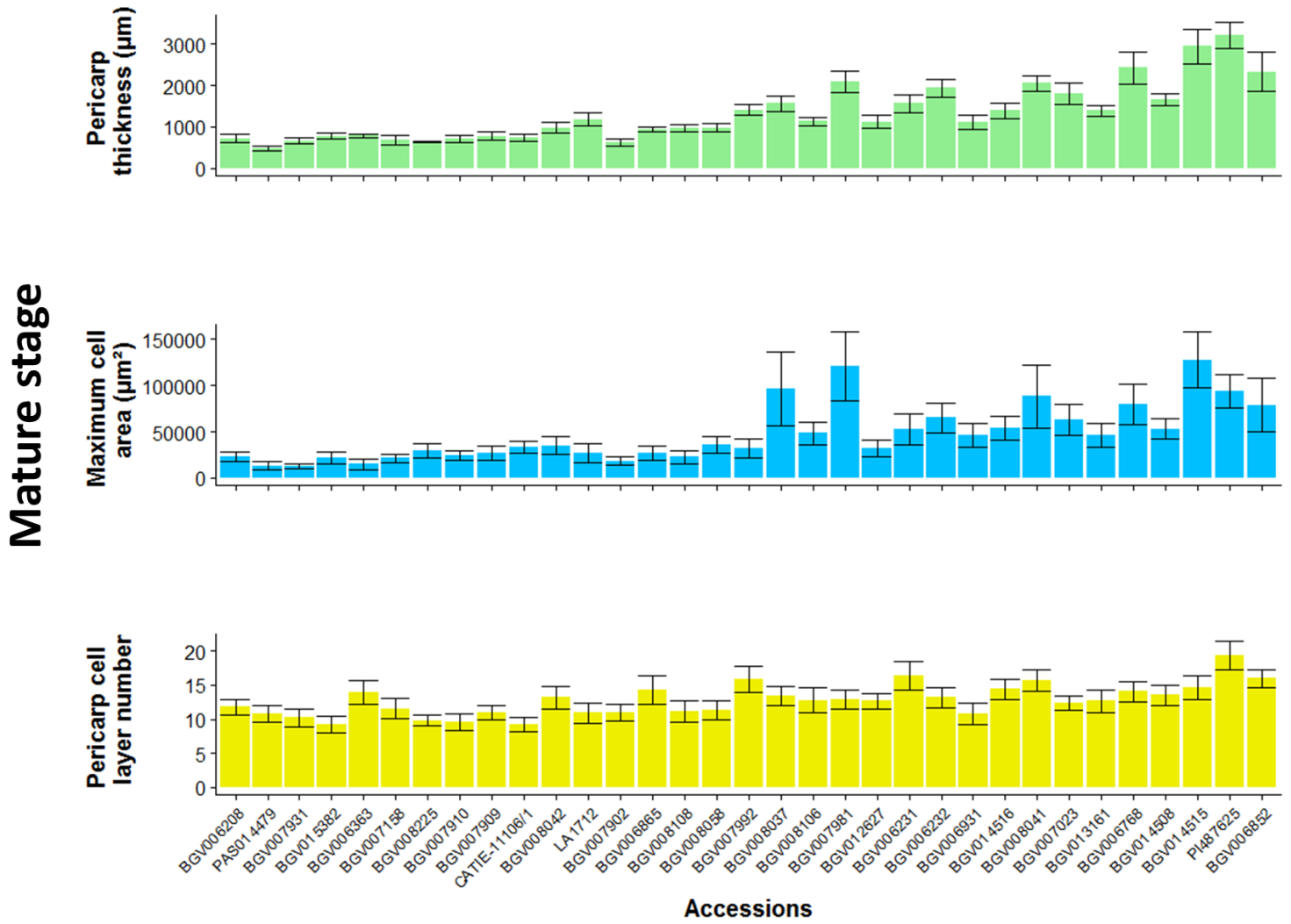


Figure 2.2. Phenotypic distribution for SP, SLC, and SLL fruit weight. A. Fruit weight distributions by tomato group. Small histogram in black is the distribution for the entire population. B. Fruit and pericarp size differences among representative accessions from each tomato group


A
Ovary stage
Pericarp thickness (µm)
0
50
100
Maximum cell area (µm²)
0
100
200
300
400
Pericarp cell layer number
0
5
10
BGV006336
BGV006208
PAS014499
BGV007931
BGV015382
BGV006363
BGV007158
BGV008225
BGV007910
BGV007909
CATIE-11106-1
BGV008042
LA1712
BGV007902
BGV006865
BGV008108
BGV008058
PI406890
BGV007992
BGV008037
BGV008106
BGV007981
BGV012627
BGV006232
BGV006231
BGV006931
BGV014516
BGV008041
BGV007023
BGV013161
BGV006768
BGV014508
BGV014515
PI487625
BGV006852
Accessions

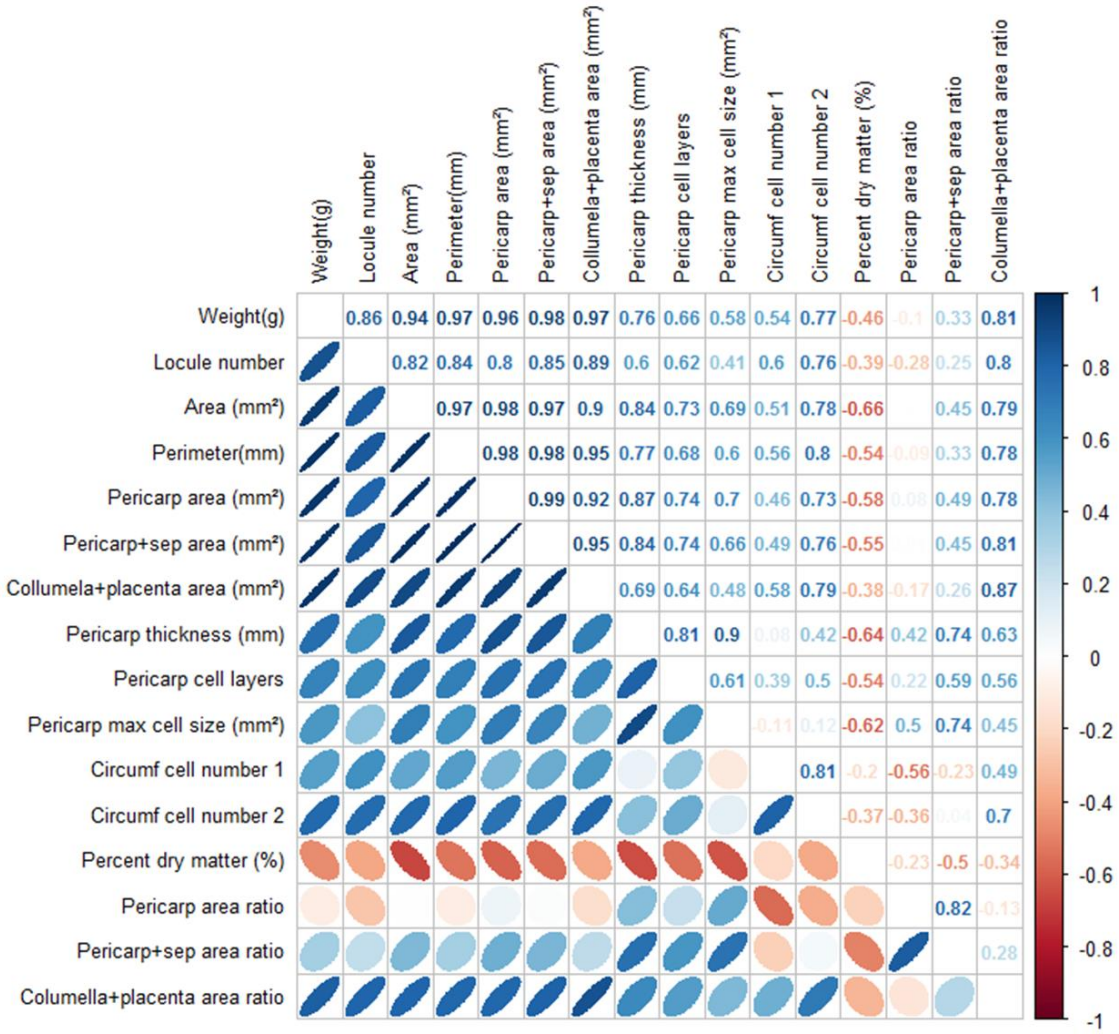
B


Mature stage
Pericarp thickness (µm)
3000
2000
1000
0
Maximum cell area (µm²)
150000
100000
50000
0
Pericarp cell layer number
20
15
10
5
0
BGV006208
PAS014479
BGV007931
BGV015382
BGV006363
BGV007158
BGV008225
BGV007910
BGV007909
CATIE-11106/1
BGV008042
LA1712
BGV007902
BGV006865
BGV008108
BGV008058
BGV007992
BGV008037
BGV008106
BGV007981
BGV012627
BGV006231
BGV006232
BGV006931
BGV014516
BGV008041
BGV007023
BGV013161
BGV006768
BGV014508
BGV014515
PI487625
BGV006852
Accessions

C
Relative increase from ovary to mature stage
Pericarp thickness ratio
30
20
10
0
Cell area ratio
500
400
300
200
100
0
Pericarp cell layer ratio
1.5
1.0
0.5
0.0
BGV006208
PAS014479
BGV007931
BGV015382
BGV006363
BGV007158
BGV008225
BGV007910
BGV007909
CATIE-11106/1
BGV008042
LA1712
BGV007902
BGV006865
BGV008108
BGV008058
BGV007992
BGV008037
BGV008106
BGV007981
BGV012627
BGV006232
BGV006231
BGV006931
BGV014516
BGV008041
BGV007023
BGV013161
BGV006768
BGV014508
BGV014515
PI487625
BGV006852
Accessions

Figure 2.3. Pericarp thickness, maximum pericarp cell size, and pericarp cell layer number for 35 selected accessions. (A) Ovary stage, (B) breaker stage, and (C) breaker to ovary ratio for the three phenotypes. There was no fruit data for BGV006336 and PI406890 at breaker stage. The accessions are arranged in ascending fruit size order in the x axis, where BGV006336 has the smallest fruits (0.82g) and BGV006852 has the largest (23.65g).

A


Weight(g)
Locule number
Area (mm²)
Perimeter(mm)
Pericarp area (mm²)
Pericarp+sep area (mm²)
Collumela+placenta area (mm²)
Pericarp thickness (mm)
Pericarp cell layers
Pericarp max cell size (mm²)
Circumf cell number 1
Circumf cell number 2
Percent dry matter (%)
Pericarp area ratio
Pericarp+sep area ratio
Columella+placenta area ratio
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

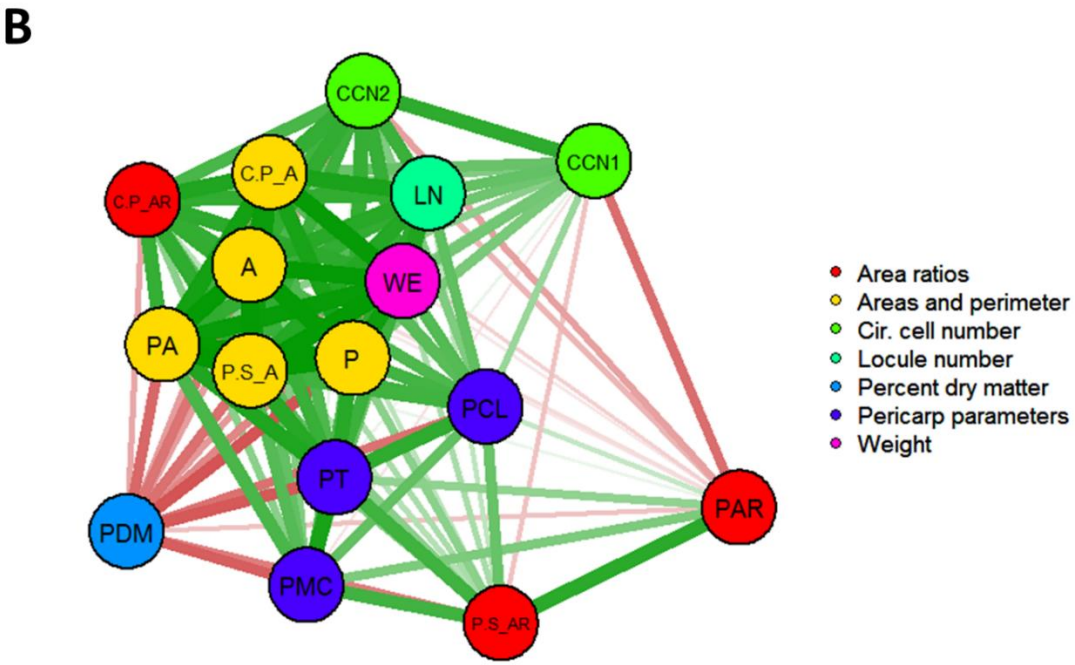


Figure 2.4. Correlations between weight and weight parameters. A. Color indicates the strength of the correlation where dark blue is a positive correlation of 1, and dark red is a negative correlation of 1. Coefficients of correlation (r) are also shown for each combination. B. Correlation network where green is a positive correlation and red a negative correlation. The thickness of the line indicates the strength of the correlation. WE = weight, LN=locule number, P=perimeter, A=area, PA=pericarp area, P.S_A=pericarp plus septum area, C.P_A=columella plus placenta area, PT=pericarp thickness, PCL=pericarp cell number, PMCS=pericarp maximum cell size, CCN1=circumference cell number 1, CNN2= circumference cell number 2, PDM= percent dry matter, PAR=pericarp area ratio, P.S_AR=pericarp plus septum area ratio, C.P_AR=columella plus placenta area ratio.

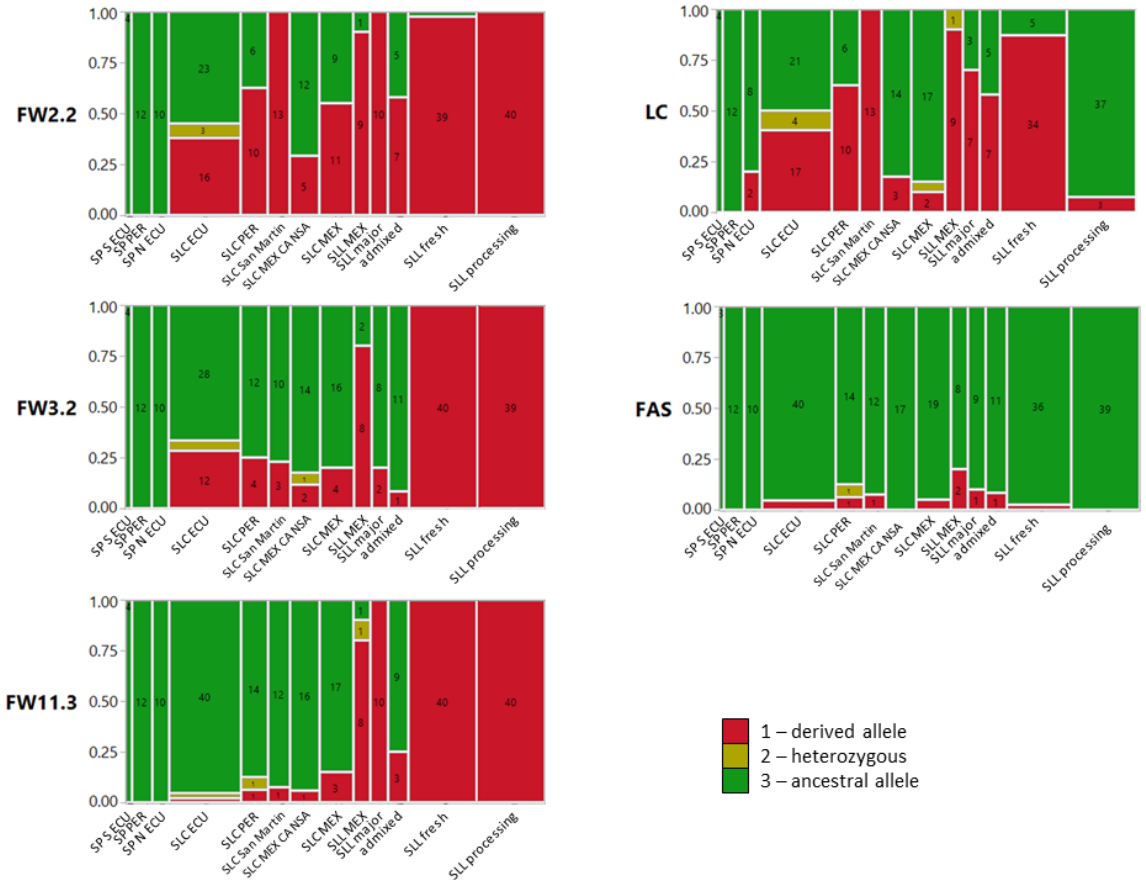


Figure 2.5. Fruit weight and shape allele frequencies for the major genes across the phylogenetic groups. Numbers inside the columns represent how many accessions belong to that category.

Supplementary table S2.1. Weight and weight components measured.

Traits	Units	Description	Measurement/formula	Approximate number of samples per accession	Software
Weight	g	Weight of the fruit	Fruits bulk weighed	20	
Area	mm^2	Area of the fruit	Fruit area measured at the equatorial plane in the medio-lateral axis	8	Tomato Analyzer 4.0
Perimeter	mm	Perimeter of the fruit	Fruit perimeter measured at the equatorial plane in the medio-lateral axis	8	Tomato Analyzer 4.0
Locule number		Locule number of the fruit	Fruits were cut at the equatorial plane in the medio-lateral axis and locule number was counted	40	
Percent dry matter	%	Percent dry matter of the fruit	Calculated using the formula: $\frac{dry\ weight}{fresh\ weight} * 100$ Fruits were bulk weighed before and after drying for one week at 70-80 ° C	2 to 10	
Pericarp area	mm^2	Area of the fruit pericarp	Fruit pericarp area measured at the equatorial plane in the medio-lateral axis	8	Tomato Analyzer 4.0
Pericarp area ratio		Ratio of pericarp area to total fruit area	Calculated using the formula: $\frac{pericarp\ area}{fruit\ area}$	8	
Pericarp + septum area	mm^2	Area of the fruit pericarp and septum	Fruit pericarp and septum area measured at the equatorial plane in the medio-lateral axis	8	Tomato Analyzer 4.0
Pericarp + septum area ratio		Ratio of pericarp and septum area to total fruit area	Calculated using the formula: $\frac{pericarp\ plus\ septum\ area}{fruit\ area}$	8	
Columella + placenta area	mm^2	Area of the fruit columella and placenta	Fruit columella and placenta area measured at the equatorial plane in the medio-lateral axis	8	Tomato Analyzer 4.0
Columella + placenta area ratio		Ratio of columella and placenta area to total fruit area	$\frac{columella\ plus\ placenta\ area}{fruit\ area}$ Calculated using the formula:	8	

Pericarp thickness	mm	Thickness of the pericarp	Length of the lines traced perpendicular to the exocarp	36	ImageJ
Pericarp cell layers		Number of cell layers in the pericarp in the axial-abaxial axis	Lines were traced as perpendicular to the exocarp as possible, avoiding vascular bundles, the endocarp layer, 2-4 small cell layers right below the exocarp, and 0-1 layers right above the endocarp. Number of cells intersected by the line were counted	36	ImageJ
Pericarp max cell size	mm^2	Size of the biggest cells in the pericarp	The area of the biggest cells in the pericarp was measured	60	ImageJ
Pericarp cell number per mm		Number of cells in the pericarp per 1mm in the axial-abaxial axis	Calculated using the formula: $\frac{pericarp\ cell\ layers}{pericarp\ thickness}$	derived trait	
Circumference cell number 1 (cell layers)		Number of cells in the circumference of the fruit in the medio-lateral axis (1)	Calculated using the formula: pericarp cell number per mm*perimeter	derived trait	
Cell diameter	mm	Diameter of the biggest cells in the pericarp	Calculated using the formula: $2\sqrt{\frac{\text{pericarp max cell size}}{\pi}}$	derived trait	
Circumference cell number 2 (cell size)		Number of cells in the circumference of the fruit in the medio-lateral axis (2)	Calculated using the formula: $\frac{fruit\ perimeter}{cell\ diameter}$	derived trait	

Supplementary table S2.2. Markers used for genotyping the known fruit weight genes.

Marker Type	Locus	Gene	Primer ID	Primer sequence (5' -> 3')	Wild-type allele	Derived allele	References
dCAPS	fw3.2	KLUH	12EP239	AAAGTCGAATAAATTAGATGAACTTGA	326 bp	304 bp	Chackrabarti et al. 2013
			12EP240	ATTGGGTCTCTCCTCGCTCT			
dCAPS	fas	CLV3	EP1069	CCAATGATAATTAAGATATTGTGACG	335 bp	466 bp	Rodríguez et al. 2011
			EP1070	ATGGTGGGGTTTTCTGTTCA			
			EP1071	CAGAAATCAGAGTCCAATTCCA			
KASP	*fw2.2*	CNR	16EP13	GAAGGTGACCAAGTTCATGCTCACAACATATAAAGTGTACTGACCCTCAc	92 bp		
			16EP14	GAAGGTCGGAGTCAACGGATTCACAACATATAAAGTGTACTGACCCTCAt			
			16EP15	ATGGATCAAATTAGTCTGAATTAATGTTTC			
KASP	*fw11.3*	CSR	16EP297	GAAGGTGACCAAGTTCATGCTATCGAACACTTTCTCAAACTCTTCTTC	177/192		
			16EP298	GAAGGTCGGAGTCAACGGATTCTTGTCCTCGCTCTCGTTCTCT			
			16EP299	CACCTTCTTCTCACCGTCATCA			
KASP	*lc1*	WUSCHEL	16EP22	GAAGGTGACCAAGTTCATGCTTGAAATGTATAAAGTAGTACGAATTGTCCAATc	173 bp		
			16EP23	GAAGGTCGGAGTCAACGGATTTGAAATGTATAAAGTAGTACGAATTGTCCAATt			
			16EP24	GACATGAATTAGGATTGTGTTTGAGATG			

Supplementary table S2.3. Phenotypes, phylogenetic groups, genotypes for the major fruit weight genes, and country of origin for all accessions. Blanks indicate missing data and NA indicates the phenotype could not be derived by the formula because of missing data. For genotypes 3 is the wild allele, 1 derived allele, and 2 heterozygous.

Name	Class abbreviated	Phylogenetic group	CNR	SlKLUH	CSR	LC	FAS	Weight (g)	Area (mm^2)	Perimeter (mm)	Locule number	Percent dry matter
BGV006208	SP	SP S ECU	3	3	3	3	3	1.1	136.7	43.4	2.0	10.4
BGV006225	SLC	SLC ECU	3	3	3	3	3	8.0	473.3	80.2	2.0	7.1
BGV006230	SLC	admixed	3	3	3	3	3	3.9	281.7	62.2	2.1	9.2
BGV006232	SLC	SLC ECU	3	3	3	3	3	8.2	566.7	88.3	2.0	7.0
BGV006327	SP	SP PER	3	3	3	3	3	1.4	124.0	43.0	2.0	12.3
BGV006336	SP	SP PER	3	3	3	3	3	0.8	120.2	40.8	2.0	11.6
BGV006347	SP	SP PER	3	3	3	3	3	0.9	117.9	40.2	2.0	13.5
BGV006353	SP	SP PER	3	3	3	3	3	1.0	124.1	41.8	2.0	10.3
BGV006363	SP	SP PER	3	3	3	3	3	1.6	173.3	48.4	2.4	12.9
BGV006370	SP	SP PER	3	3	3	3	3	1.3	149.4	45.2	2.1	15.0
BGV006454	SP	SP PER	3	3	3	3	3	1.4	141.1	43.4	2.0	16.0
BGV006457	SP	SP PER	3	3	3	3	3	1.3	159.7	47.0	2.0	13.4
BGV006478	SP	SP PER	3	3	3	3	3	1.1	127.3	42.9	2.0	13.4
BGV006753	SLC	SLC ECU	3	3	3	3	3	7.5	551.5	87.5	2.0	10.6
BGV006768	SLC	SLC ECU	3	3	3	3	3	14.7	818.2	105.7	2.1	6.6
BGV006779	SLC	SLC ECU	3	3	3	3	3	13.3	727.3	100.9	2.1	7.5
BGV006792	SLC	SLC ECU	3	3	3	3	3	15.0	853.2	107.6	2.4	6.1
BGV006859	SLC	SLC ECU	3	3	3	3	3	4.2	269.5	60.9	2.3	8.9
BGV006865	SLC	SLC ECU	3	3	3	3	3	4.1	279.1	62.1	2.0	9.3
BGV006867	SLC	SLC ECU	3	3	3	3	3	5.0	336.7	68.3	2.0	9.2
BGV006901	SLC	SLC ECU	3	3	3	3	3	6.7	487.7	82.4	2.1	8.1

BGV006904	SLC	SLC ECU	3	3	3	3	3	3.6	296.7	64.1	2.0	9.0
BGV006910	SLC	SLC ECU	3	3	3	3	3	4.5	359.3	70.8	2.1	11.5
BGV007109	SP	SP N ECU	3	3	3	3	3	1.1	147.5	45.2	2.0	9.9
BGV007111	SP	SP N ECU	3	3	3	3	3	1.2	143.9	44.5	2.0	10.2
BGV007149	SP	SP N ECU	3	3	3	3	3	1.9	155.5	46.9	2.0	10.7
BGV007151	SP	SP S ECU	3	3	3	3	3	1.5	148.5	45.6	2.0	11.0
BGV007152	SP	SP N ECU	3	3	3	3	3	2.1	168.6	48.7	2.0	10.8
BGV007155	SP		3	3	3	3	3	1.6	160.1	46.6	2.0	11.6
BGV007158	SP	SP N ECU	3	3	3	3	3	1.6	145.4	45.1	2.0	11.9
BGV007161	SP	SP N ECU	3	3	3	3	3	1.7	166.6	48.0	2.0	12.1
BGV007181	SP	SP N ECU	3	3	3	3	3	1.8	170.8	48.8	2.0	10.9
BGV007194	SP	SP N ECU	3	3	3	3	3	1.3	148.6	45.5	2.0	11.2
BGV007198	SP	admixed	3	3	3	3	3	1.8	178.9	49.0	2.0	11.6
BGV007366	SP	SP S ECU	3	3	3	3	3	1.5	143.1	44.2	2.0	12.5
BGV007901	SLC	SLC MEX	3	3	3	3	3	3.1	208.6	56.0	2.1	10.6
BGV007902	SLC	SLC MEX	3	3	3	3	3	3.6	224.3	57.3	2.1	9.7
BGV007909	SLC	SLC MEX	3	3	3	3	3	2.7	205.3	53.6	2.0	9.3
BGV007911	SLC	SLC MEX	3	3	3	3	3	2.9	245.5	58.1	2.0	10.2
BGV007918	SLC	SLC MEX	3	3	3	3	3	2.8	206.3	54.3	2.1	11.0
BGV007927	SLC	SLC MEX CA NSA	3	3	3	3	3	1.5	134.2	42.9	2.0	13.1
BGV007931	SLC	SLC MEX CA NSA	3	3	3	3	3	1.3	131.0	42.9	2.1	15.2
BGV007934	SLC	SLC MEX CA NSA	3	3	3	3	3	1.9	196.4	51.9	2.2	13.3
BGV007935	SLC	SLC MEX CA NSA	3	3	3	3	3	1.9	182.5	50.1	2.3	13.0
BGV007992	SLC	SLC ECU	3	3	3	3	3	5.1	380.2	74.5	2.0	10.2

BGV008218	SLC	SLC MEX CA NSA	3	3	3	3	3	4.9	377.2	72.4	2.1	10.1
BGV008225	SLC	SLC PER	3	3	3	3	3	2.0	186.4	51.1	2.0	11.0
BGV008345	SLC	SLC MEX CA NSA	3	3	3	3	3	4.0	282.1	62.8	2.1	8.7
BGV008348	SLC	SLC MEX CA NSA	3	3	3	3	3	5.7	429.5	74.9	2.3	8.7
BGV012615	SLC	admixed	3	3	3	3	3	2.0	187.0	52.0	2.0	9.7
BGV012626	SLC	admixed	3	3	3	3	3	3.0	286.6	63.7	2.0	9.8
BGV013134	SLC	SLC MEX	3	3	3	3	3	3.0	246.1	59.7	2.0	9.9
BGV013175	SLC	SLC MEX CA NSA	3	3	3	3	3	3.8	333.1	67.5	2.1	10.8
BGV015380	SP	SP PER	3	3	3	3	3	1.1	130.4	43.2	2.2	14.2
BGV015382	SP	SP PER	3	3	3	3	3	1.5	149.2	45.3	2.1	13.2
LA1712	SLC	SLC MEX CA NSA	3	3	3	3	3	3.6	283.1	61.9	2.0	9.7
LA2697	SLC	SLC MEX CA NSA	3	3	3	3	3	3.5	287.9	62.9	2.0	12.0
PAS014479	SP	SP PER	3	3	3	3	3	1.2	132.8	43.0	2.4	12.7
PI129026	SLC	SLC ECU	1	1	3	1	1	65.1		205.0	11.3	7.6
PI129033	SLC	SLC ECU	1	1	3	1	1	43.4		182.3	10.9	7.4
BGV008221	SLC	SLC MEX	3	1	1	3	3	5.1	354.7	70.1	2.1	9.8
BGV012639	SLC	SLC ECU	3	1	1	3	3	14.1	724.3	100.2	2.0	7.0
BGV008041	SLC	SLC PER	1	1	3	3	3	10.2	412.2	76.0	2.1	8.5
PI406890	SLC	SLC PER	1	1	3	3	3	4.9	413.1	75.7	2.1	10.0
BGV007920	SLC	SLC MEX	3	3	1	3	3	4.6	310.3	66.5	2.0	8.7
BGV012613	SLC	admixed	3	3	1	3	3	4.7	474.5	81.5	2.5	8.2
BGV007862	SLL	SLL major	1	1	1	3	3	33.5	1254.5	132.3	2.4	8.0

Tegucigalpa	SLL	SLL major	1	1	1	3	3	22.5	669.3	97.5	2.1	7.9
BGV006148	SLC	SLC ECU	3	1	3	3	3	6.1	491.8	82.8	2.0	7.4
BGV007933	SLC	SLC MEX CA NSA	3	1	3	3	3	2.8	252.9	59.4	2.1	12.9
BGV005895	SLC	SLC MEX	1	1	3	3	3	6.7	452.3	79.3	2.1	9.0
BGV006234	SLC	SLC ECU	1	1	3	3	3	9.7	672.2	97.0	2.3	6.7
BGV007860	SLL	SLL MEX	1	1	1	1	3	67.5	2772.1	202.6	5.3	7.1
BGV007864	SLL	SLL MEX	1	1	1	1	3	69.3	3072.4	212.6	5.4	7.2
BGV007867	SLL	SLL MEX	1	1	1	1	3	66.0	2925.9	206.4	4.9	6.8
BGV007872	SLL	SLL MEX	1	1	1	1	3	68.7	3414.4	219.4	5.9	6.7
BGV007936	SLL	SLL MEX	1	1	1	1	3	91.4		228.6	6.6	6.5
BGV008224	SLL	SLL MEX	1	1	1	1	3	30.4	1306.3	136.1	3.1	6.8
BGV008354	SLC	SLC MEX CA NSA	1	1	1	1	3	17.7	875.4	114.5	3.3	8.4
BGV007854	SLL	SLL major	1	3	1	1	3	44.7	1655.7	152.3	5.6	7.3
BGV007857	SLL	SLL major	1	3	1	1	3	50.1	1602.3	159.3	6.9	8.1
BGV007865	SLL	SLL major	1	3	1	1	3	64.3	2674.4	191.3	7.7	8.5
BGV007871	SLL	SLL MEX	1	3	1	1	3	53.8	2162.1	177.2	5.1	7.7
BGV007875	SLL	SLL major	1	3	1	1	3	74.3	2171.3	179.2	6.6	7.0
BGV007876	SLL	SLL major	1	3	1	1	3	63.1	2602.3	190.4	7.8	8.0
BGV007895	SLL	SLL major	1	3	1	1	3	81.5	2433.6	189.7	9.6	10.0
BGV008077	SLC	SLC PER	1	3	1	1	3	14.5	832.5	108.0	3.2	8.8
PI378994	SLC	admixed	1	3	1	1	3	3.9	387.5	74.1	3.1	9.2
BGV006235	SLC	SLC ECU	1	3	3	1	3	10.6	673.0	96.9	3.4	7.3
BGV006828	SLC	SLC ECU	1	3	3	1	3	15.8	910.8	113.2	4.7	7.9
BGV008036	SLC	admixed	1	3	3	1	3	9.3	555.7	87.2	3.3	9.2
BGV008061	SLC	admixed	1	3	3	1	3	7.2	486.4	82.5	3.6	9.5

BGV008065	SLC	SLC San Martin	1	3	3	1	3	15.1	836.4	109.0	4.4	7.0
BGV008095	SLC	SLC PER	1	3	3	1	3	8.4	530.1	86.1	2.6	7.5
BGV008098	SLC	SLC San Martin	1	3	3	1	3	9.2	570.2	88.4	3.3	8.7
BGV008100	SLC	admixed	1	3	3	1	3	13.6	609.2	93.5	2.9	8.2
BGV008106	SLC	admixed	1	3	3	1	3	6.4	409.8	75.4	2.7	8.8
BGV012640	SLC	SLC PER	1	3	3	1	3	10.1	618.4	92.3	2.5	8.7
BGV014515	SLC	SLC San Martin	1	3	3	1	3	17.0	965.2	116.4	3.7	7.1
BGV014516	SLC	SLC San Martin	1	3	3	1	3	10.0	537.7	87.7	3.1	8.6
BGV014518	SLC	SLC San Martin	1	3	3	1	3	13.5	794.5	104.2	2.9	7.1
BGV014519	SLC	SLC San Martin	1	3	3	1	3	12.9	727.3	101.1	3.0	8.5
BGV014522	SLC	SLC San Martin	1	3	3	1	3	13.6	699.8	98.1	3.2	8.2
BGV015726	SLC	SLC San Martin	1	3	3	1	3	14.9	880.6	111.5	3.4	8.7
LA2309	SLC	SLC San Martin	1	3	3	1	3	15.0	881.6	110.9	3.5	6.2
BGV004584	SLC	SLC MEX CA NSA	1	3	3	3	3	9.6	593.3	91.0	2.4	9.4
BGV006229	SLC	SLC ECU	1	3	3	3	3	8.3	536.0	86.5	2.2	8.8
BGV006806	SLC	SLC ECU	1	3	3	3	3	5.7	344.4	70.1	2.1	8.7
BGV006906	SLC	SLC ECU	1	3	3	3	3	8.9	534.0	85.9	2.3	9.7
BGV007908	SLC	SLC MEX	1	3	3	3	3	2.5	192.6	52.7	2.0	12.3
BGV007910	SLC	SLC MEX	1	3	3	3	3	2.4	207.3	53.7	2.3	11.2
BGV007921	SLC	SLC MEX	1	3	3	3	3	2.5	204.8	53.3	2.0	10.9

BGV007989	SLC	SLC PER	1	3	3	3	3	6.1	411.4	75.3	2.1	9.3
BGV008037	SLC	SLC PER	1	3	3	3	3	6.1	374.2	71.7	2.0	9.9
BGV008051	SLC	SLC MEX	1	3	3	3	3	2.7	214.0	55.3	2.1	10.2
BGV008058	SLC	SLC ECU	1	3	3	3	3	4.8	345.7	67.8	2.1	9.2
BGV008067	SLC	SLC MEX	1	3	3	3	3	2.4	220.0	55.1	2.3	12.1
BGV008070	SLC	SLC MEX	1	3	3	3	3	2.6	240.3	57.3	2.0	12.9
BGV008219	SLC	SLC MEX	1	3	3	3	3	3.9	294.6	64.7	2.0	11.3
BGV008223	SLC	SLC MEX	1	3	3	3	3	3.4	262.0	60.4	2.0	9.2
BGV012627	SLC	SLC MEX CA NSA	1	3	3	3	3	7.9	559.2	88.0	2.4	10.4
CATIE-11106/1	SLC	SLC MEX CA NSA	1	3	3	3	3	2.9	263.7	60.4	2.1	9.5
BGV006231	SLC	SLC ECU	3	3	3	1	3	8.2	534.1	86.4	3.0	7.0
BGV006775	SP	SP N ECU	3	3	3	1	3	1.9	202.0	52.5	2.1	10.6
BGV006777	SLC	SLC ECU	3	3	3	1	3	9.1	501.6	83.8	2.8	9.1
BGV006899	SLC	SLC ECU	3	3	3	1	3	9.0	629.9	93.6	2.9	8.6
BGV006907	SLC	SLC ECU	3	3	3	1	3	11.3	707.2	100.2	5.9	7.8
BGV007015	SLC	SLC ECU	3	3	3	1	3	11.9	756.2	102.5	3.2	9.0
BGV007017	SLC	SLC ECU	3	3	3	1	3	9.6	673.2	96.6	2.9	9.2
BGV007023	SLC	SLC ECU	3	3	3	1	3	10.5	574.2	88.8	2.5	8.8
BGV007169	SP	SP N ECU	3	3	3	1	3	1.7	175.5	48.6	2.3	11.9
BGV007981	SLC	SLC PER	3	3	3	1	3	6.8	533.1	85.3	3.3	8.6
BGV008096	SLC	SLC PER	3	3	3	1	3	9.7	529.8	86.3	2.2	7.2
BGV008108	SLC	SLC MEX CA NSA	3	3	3	1	3	4.2	321.3	66.7	2.4	10.7
BGV013161	SLC	SLC PER	3	3	3	1	3	10.6	560.0	88.6	2.1	6.8
BGV015730	SLC	SLC PER	3	3	3	1	3	10.3	665.4	96.2	2.5	8.0

BGV005912	SLC	SLC ECU	1	1	3	1	3	11.8	804.5	106.2	3.8	5.9
BGV006927	SLC	SLC ECU	1	1	3	1	3	10.8	776.5	104.6	4.5	7.9
BGV006931	SLC	SLC ECU	1	1	3	1	3	9.7	624.7	93.2	3.5	7.7
BGV006934	SLC	SLC ECU	1	1	3	1	3	11.9	843.1	110.1	3.9	8.3
BGV007990	SLC	SLC PER	1	1	3	1	3	9.5	593.9	90.5	3.4	9.8
BGV008042	SLC	admixed	1	1	3	1	3	3.5	260.4	59.9	3.5	10.1
BGV014508	SLC	SLC San Martin	1	1	3	1	3	16.1	786.2	104.8	2.8	4.1
BGV014514	SLC	SLC San Martin	1	1	3	1	3	18.9	1022.2	118.4	4.1	7.3
BGV015727	SLC	SLC San Martin	1	1	3	1	3	12.1	683.6	97.8	2.2	7.5
PI129088	SLC	SLC MEX CA NSA	1	2	3	1	3	19.6	859.0	107.9	2.8	8.2
PI487625	SLC	SLC ECU	1	1	3	1	3	18.7	1111.0	123.5	3.2	6.4
BGV007878	SLL	SLL major	1	3	1	1	1	71.2	2627.6	193.6	10.2	7.2
BGV007900	SLC	admixed	1	3	1	1	1	35.8		168.7	9.5	6.7
BGV015734	SLC	SLC San Martin	1	3	1	1	1	54.2		196.5	9.8	5.7
Voyage	SLL	SLL MEX	1	3	1	1	1	76.5		256.8	8.7	7.8
BGV006175	SLC	SLC ECU	1	1	3	2	3	15.2	943.5	111.8	3.2	6.7
BGV006767	SLC	SLC ECU	2	1	3	2	3	5.4	367.0	71.5	2.2	9.6
BGV006825	SLC	SLC PER	1	3	2	1	2	23.7	1165.9	125.9	4.3	7.6
BGV006852	SLC	SLC ECU	1	3	3	2	3	23.6	1336.4	135.6	4.7	6.8
BGV006881	SLC	SLC ECU	2	2	3	2	3	7.8	547.5	87.7	2.7	7.8
BGV006896	SLC	SLC ECU	2	3	3	3	3	5.3	396.0	74.3	2.4	9.5
BGV007339	SP	SP S ECU	3	3	3	3		1.6	158.3	47.2	2.0	11.5
BGV007863	SLL	SLL major	1	3	1	3	3	25.2	1108.4	123.5	2.4	8.1

BGV007870	SLL	SLL MEX	1	1	2	2	3	37.5				6.9
BGV007894	SLC	SLC MEX	3	1	3	1	3	10.3	587.7	89.9	2.5	8.4
BGV007899	SLC	SLC MEX	1	1	1	1	1	30.2		174.0	8.0	7.7
BGV008189	SLC	SLC PER	3	1	3	1	3	4.0	340.6	69.1	3.2	9.0
BGV012614	SLC	SLC MEX	1	3	3	2	3	3.0	279.5	62.7	2.1	10.3
BGV012625	SLC	SLC ECU	3	2	2	1	3	18.7	996.8	119.5	4.6	7.4
BGV013945	SLC	SLC PER	1	3	3	3	1	8.8	663.2	94.5	4.4	8.4
LA0767	SLL	SLL MEX	3	1	3	1	1	55.1		219.6	7.1	7.7

Supplementary table S2.3 continued.

Name	Pericarp area (mm^2)	Pericarp area ratio	Pericarp+septum area (mm^2)	Pericarp+septum area ratio	Columella+placenta area (mm^2)	Columella+placenta area ratio
BGV006208	45.7	0.3336	51.6	0.3770	5.3	0.0387
BGV006225	167.6	0.3626	200.0	0.4326	33.3	0.0704
BGV006230	85.5	0.3047	100.2	0.3581	17.4	0.0614
BGV006232	202.4	0.3631	247.1	0.4431	43.3	0.0765
BGV006327	39.6	0.2973	43.6	0.3275	6.2	0.0497
BGV006336	35.3	0.2853	40.0	0.3214	4.5	0.0367
BGV006347	39.6	0.3367	44.3	0.3783	5.4	0.0458
BGV006353	44.2	0.3475	50.1	0.3942	6.6	0.0532
BGV006363	55.9	0.3268	64.1	0.3751	11.2	0.0649
BGV006370	48.8	0.3297	56.5	0.3818	10.6	0.0711
BGV006454	41.1	0.2977	47.8	0.3454	8.8	0.0613
BGV006457	56.0	0.3323	62.8	0.3727	11.2	0.0628
BGV006478	48.2	0.3582	55.4	0.4061	9.9	0.0773
BGV006753	217.4	0.3943	274.3	0.4975	40.8	0.0740
BGV006768	339.9	0.4240	394.0	0.4916	63.2	0.0775
BGV006779	295.6	0.4051	340.9	0.4672	59.9	0.0823
BGV006792	330.6	0.4019	406.1	0.4924	52.9	0.0615
BGV006859	78.6	0.2910	93.6	0.3466	24.1	0.0898
BGV006865	83.4	0.2997	99.4	0.3568	19.6	0.0710
BGV006867	110.4	0.3291	129.0	0.3843	30.5	0.0908
BGV006901	168.9	0.3385	198.8	0.3976	26.8	0.0553
BGV006904	101.6	0.3413	121.4	0.4076	22.8	0.0768
BGV006910	118.5	0.3287	140.0	0.3876	29.6	0.0813

BGV007109	50.3	0.3405	55.2	0.3733	7.7	0.0525
BGV007111	47.9	0.3338	52.7	0.3667	7.7	0.0540
BGV007149	44.1	0.2775	49.5	0.3113	10.5	0.0675
BGV007151	44.8	0.2989	50.8	0.3383	10.9	0.0723
BGV007152	50.2	0.2922	55.5	0.3214	10.3	0.0613
BGV007155	48.5	0.3071	53.1	0.3362	10.5	0.0654
BGV007158	43.8	0.2965	48.6	0.3289	11.7	0.0808
BGV007161	52.3	0.3093	59.1	0.3526	13.1	0.0792
BGV007181	44.3	0.2567	48.9	0.2828	12.8	0.0753
BGV007194	45.8	0.3047	49.9	0.3316	8.6	0.0581
BGV007198	50.2	0.2892	55.6	0.3200	13.0	0.0718
BGV007366	42.2	0.2980	46.7	0.3293	11.5	0.0805
BGV007901	66.4	0.2927	70.5	0.3194	15.7	0.0749
BGV007902	69.0	0.2885	73.7	0.3168	15.7	0.0704
BGV007909	65.1	0.3122	72.0	0.3447	14.6	0.0712
BGV007911	73.9	0.2993	83.3	0.3364	16.7	0.0684
BGV007918	60.1	0.2819	66.1	0.3100	14.2	0.0689
BGV007927	45.5	0.3395	52.8	0.3941	7.8	0.0585
BGV007931	42.8	0.3197	48.1	0.3597	10.1	0.0775
BGV007934	57.9	0.2964	67.0	0.3428	15.8	0.0804
BGV007935	53.3	0.2943	63.7	0.3519	13.9	0.0764
BGV007992	156.3	0.3919	183.9	0.4498	33.2	0.0861
BGV008218	124.2	0.3292	136.8	0.3642	35.8	0.0948
BGV008225	59.0	0.3086	66.2	0.3420	12.9	0.0686
BGV008345	90.2	0.3159	100.5	0.3522	18.1	0.0632
BGV008348	152.8	0.3785	172.6	0.4271	41.7	0.0969

BGV012615	55.7	0.2979	63.0	0.3327	16.0	0.0856
BGV012626	98.2	0.3352	117.7	0.4007	26.3	0.0913
BGV013134	84.9	0.3324	91.3	0.3677	15.6	0.0638
BGV013175	114.3	0.3457	127.8	0.3880	29.9	0.0897
BGV015380	44.3	0.3253	47.7	0.3585	8.5	0.0653
BGV015382	47.0	0.3155	52.3	0.3513	9.9	0.0657
LA1712	94.8	0.3389	104.3	0.3734	23.2	0.0804
LA2697	90.5	0.3179	100.7	0.3535	25.3	0.0878
PAS014479	46.2	0.3471	49.6	0.3777	8.3	0.0620
PI129026	NA	NA	NA	NA	NA	NA
PI129033	NA	NA	NA	NA	NA	NA
BGV008221	109.1	0.3094	120.8	0.3426	30.5	0.0856
BGV012639	275.7	0.3823	317.5	0.4413	54.4	0.0755
BGV008041	169.2	0.4084	197.0	0.4781	38.5	0.0933
PI406890	140.1	0.3389	160.9	0.3892	50.8	0.1229
BGV007920	107.2	0.3358	121.5	0.3807	22.7	0.0731
BGV012613	158.8	0.3325	189.4	0.3962	43.1	0.0903
BGV007862	563.8	0.4489	674.0	0.5356	140.8	0.1121
Tegucigalpa	347.6	0.5109	379.9	0.5582	53.5	0.0798
BGV006148	160.6	0.3276	186.9	0.3813	35.3	0.0716
BGV007933	77.2	0.3036	88.4	0.3478	26.2	0.1037
BGV005895	145.5	0.3214	171.4	0.3787	51.0	0.1129
BGV006234	245.4	0.3651	298.4	0.4439	40.6	0.0603
BGV007860	754.1	0.2685	966.6	0.3468	363.6	0.1332
BGV007864	812.2	0.2587	1046.3	0.3271	406.3	0.1327
BGV007867	726.0	0.2480	975.2	0.3337	537.8	0.1897

BGV007872	914.2	0.2790	1257.4	0.3759	561.7	0.1654
BGV007936	NA	NA	NA	NA	NA	NA
BGV008224	521.3	0.3986	625.6	0.4787	180.8	0.1389
BGV008354	306.2	0.3260	383.0	0.3983	163.1	0.1863
BGV007854	570.7	0.3452	785.9	0.4770	302.3	0.1825
BGV007857	536.9	0.2976	827.3	0.4701	272.9	0.1732
BGV007865	780.6	0.3011	1144.5	0.4390	662.3	0.2481
BGV007871	816.0	0.3599	1051.7	0.4643	399.1	0.1866
BGV007875	707.2	0.3090	1059.6	0.4619	417.4	0.1930
BGV007876	760.6	0.2895	1099.8	0.4220	563.6	0.2230
BGV007895	697.3	0.2701	1034.6	0.4122	671.3	0.2797
BGV008077	290.8	0.3478	360.8	0.4315	90.8	0.1090
PI378994	156.8	0.3893	190.2	0.4833	32.3	0.0820
BGV006235	228.4	0.3403	292.6	0.4354	58.4	0.0867
BGV006828	257.3	0.2792	380.1	0.4132	93.8	0.1040
BGV008036	182.9	0.3327	223.6	0.4073	63.3	0.1135
BGV008061	155.8	0.3141	193.9	0.3893	39.6	0.0830
BGV008065	192.1	0.2330	268.5	0.3193	105.7	0.1239
BGV008095	177.8	0.3257	212.5	0.3904	40.2	0.0741
BGV008098	184.5	0.3264	225.2	0.3979	56.7	0.1001
BGV008100	207.1	0.3310	244.7	0.3944	81.0	0.1348
BGV008106	131.2	0.3218	150.5	0.3689	32.1	0.0784
BGV012640	202.1	0.3295	232.5	0.3844	64.9	0.1049
BGV014515	354.7	0.3652	435.0	0.4478	106.4	0.1103
BGV014516	140.3	0.2541	172.8	0.3129	56.5	0.1054
BGV014518	244.6	0.3182	298.8	0.3875	69.7	0.0866

BGV014519	262.0	0.3575	313.4	0.4240	64.5	0.0878
BGV014522	205.7	0.2986	252.6	0.3665	81.7	0.1171
BGV015726	305.7	0.3429	378.4	0.4255	135.1	0.1564
LA2309	253.3	0.2875	317.6	0.3610	72.4	0.0821
BGV004584	195.6	0.3287	237.7	0.3996	93.5	0.1572
BGV006229	182.2	0.3401	222.5	0.4156	47.0	0.0875
BGV006806	131.8	0.3724	156.2	0.4414	21.6	0.0630
BGV006906	147.7	0.2790	179.5	0.3390	54.1	0.1013
BGV007908	62.8	0.3105	70.1	0.3463	14.1	0.0730
BGV007910	59.9	0.2868	69.9	0.3247	16.0	0.0774
BGV007921	67.9	0.3273	75.2	0.3625	16.0	0.0780
BGV007989	153.2	0.3723	181.6	0.4413	50.6	0.1229
BGV008037	146.1	0.3978	166.4	0.4530	32.0	0.0850
BGV008051	70.5	0.3192	75.6	0.3478	16.3	0.0761
GV008058	111.0	0.3329	124.6	0.3729	21.3	0.0610
BGV008067	67.4	0.3064	76.9	0.3494	32.8	0.1494
BGV008070	78.7	0.3293	86.4	0.3590	20.3	0.0842
BGV008219	101.4	0.3334	112.2	0.3689	25.5	0.0867
BGV008223	79.5	0.3027	87.2	0.3321	21.2	0.0807
BGV012627	166.9	0.2991	198.8	0.3578	90.5	0.1576
CATIE-11106/1	67.1	0.2560	77.7	0.2963	22.9	0.0868
BGV006231	180.7	0.3391	225.0	0.4222	29.9	0.0559
BGV006775	60.9	0.3052	69.5	0.3485	18.8	0.0927
BGV006777	191.7	0.3792	236.4	0.4675	29.3	0.0585
BGV006899	227.7	0.3630	281.1	0.4422	65.4	0.1036

BGV006907	199.8	0.2801	272.1	0.3795	89.6	0.1255
BGV007015	267.2	0.3548	339.2	0.4505	86.4	0.1146
BGV007017	235.8	0.3538	291.1	0.4368	70.1	0.1047
BGV007023	193.6	0.3400	245.0	0.4308	79.6	0.1389
BGV007169	49.0	0.2858	54.6	0.3184	12.8	0.0730
BGV007981	180.8	0.3479	243.8	0.4703	36.1	0.0677
BGV008096	180.7	0.3365	209.3	0.3897	47.3	0.0893
BGV008108	92.2	0.2843	102.7	0.3174	29.3	0.0906
BGV013161	200.0	0.3526	235.9	0.4160	42.8	0.0769
BGV015730	241.3	0.3631	286.3	0.4314	57.7	0.0880
BGV005912	229.2	0.2880	302.9	0.3799	57.3	0.0715
BGV006927	264.3	0.3399	320.7	0.4121	68.5	0.0887
BGV006931	173.4	0.2733	223.6	0.3534	53.2	0.0857
BGV006934	259.7	0.3079	339.3	0.4017	73.0	0.0866
BGV007990	169.6	0.2877	210.0	0.3563	79.0	0.1340
BGV008042	78.7	0.3070	98.5	0.3911	17.3	0.0661
BGV014508	286.1	0.3606	349.0	0.4433	77.6	0.0981
BGV014514	296.1	0.2959	397.8	0.3967	114.7	0.1121
BGV015727	272.5	0.3978	307.9	0.4497	60.7	0.0892
PI129088	316.5	0.3792	402.5	0.4806	83.3	0.0943
PI487625	410.0	0.3744	514.5	0.4654	123.9	0.1115
BGV007878	718.0	0.2782	1184.1	0.4567	604.5	0.2260
BGV007900	NA	NA	NA	NA	NA	NA
BGV015734	NA	NA	NA	NA	NA	NA
Voyage	NA	NA	NA	NA	NA	NA
BGV006175	350.2	0.3905	425.8	0.4849	121.2	0.1285

BGV006767	114.5	0.3119	145.0	0.3953	36.7	0.1000
BGV006825	387.4	0.3404	492.4	0.4336	133.3	0.1174
BGV006852	428.0	0.3268	596.3	0.4547	141.2	0.1069
BGV006881	188.5	0.3427	240.1	0.4354	38.3	0.0700
BGV006896	117.6	0.2960	146.5	0.3682	26.8	0.0680
BGV007339	46.3	0.2865	50.8	0.3146	10.0	0.0629
BGV007863	457.7	0.4140	549.5	0.5116	187.7	0.1692
BGV007870	NA	NA	NA	NA	NA	NA
BGV007894	201.0	0.3466	235.2	0.4053	65.9	0.1122
BGV007899	NA	NA	NA	NA	NA	NA
BGV008189	100.7	0.2924	124.0	0.3595	31.6	0.0928
BGV012614	91.6	0.3251	104.8	0.3746	30.1	0.1078
BGV012625	353.8	0.3454	512.6	0.5115	87.6	0.0878
BGV013945	196.4	0.3081	261.4	0.4098	63.1	0.0938
LA0767	NA	NA	NA	NA	NA	NA

Supplementary table S2.3 continued.

Name	Pericarp thickness (mm)	Pericarp cell layers	Pericarp max cell size (mm^2)	Pericarp cell number/mm	Circumference cell number 1 (cell layers)	Cell diameter	circumference cell number 2 (cell size)	Country
BGV006208	0.719	11.778	0.023	16.4	711.5	0.170	255.9	ECU
BGV006225	1.532	14.444	0.046	9.4	755.6	0.242	331.1	ECU
BGV006230	0.938	11.444	0.031	12.2	758.4	0.200	310.5	ECU
BGV006232	1.929	13.182	0.065	6.8	603.4	0.288	306.9	ECU
BGV006327	0.793	15.861	0.013	20.0	859.1	0.129	332.7	PER
BGV006336	0.466	11.111	0.012	20.0	816.8	0.125	327.7	PER
BGV006347	0.751	11.259	0.020	15.0	602.8	0.159	253.5	PER
BGV006353	0.913	12.667	0.024	13.9	579.8	0.175	238.8	PER
BGV006363	0.791	13.944	0.015	17.6	853.4	0.138	351.9	PER
BGV006370	0.873	12.278	0.032	14.1	635.1	0.202	223.1	PER
BGV006454	1.009	11.167	0.037	11.1	480.1	0.217	200.0	PER
BGV006457	0.829	11.333	0.025	13.7	642.6	0.178	263.5	PER
BGV006478				NA	NA		NA	PER
BGV006753	1.738	14.167	0.066	8.2	713.3	0.291	301.1	ECU
BGV006768	2.417	14.111	0.079	5.8	616.9	0.318	332.6	ECU
BGV006779	3.075	18.583	0.129	6.0	609.6	0.405	248.9	ECU
BGV006792	2.663	16.556	0.115	6.2	668.6	0.383	280.5	ECU
BGV006859	1.071	13.611	0.031	12.7	774.6	0.198	308.0	ECU
BGV006865	0.950	14.375	0.027	15.1	939.3	0.186	334.0	ECU
BGV006867	1.396	12.970	0.040	9.3	634.1	0.226	301.5	ECU
BGV006901	1.641	14.389	0.067	8.8	722.3	0.292	282.3	ECU
BGV006904	1.096	14.030	0.051	12.8	821.0	0.254	252.2	ECU

BGV006910	0.959	10.993	0.022	11.5	811.5	0.167	424.6	ECU
BGV007109	0.826	10.000	0.038	12.1	547.4	0.219	206.7	ECU
BGV007111	0.782	10.472	0.024	13.4	596.3	0.175	254.7	ECU
BGV007149	0.718	10.515	0.023	14.7	687.6	0.171	274.1	ECU
BGV007151	0.679	11.750	0.025	17.3	788.8	0.177	257.2	ECU
BGV007152	0.800	10.639	0.034	13.3	647.7	0.207	235.7	ECU
BGV007155	0.679	11.750	0.023	17.3	807.4	0.171	272.7	ECU
BGV007158	0.691	11.583	0.021	16.8	756.6	0.165	274.0	ECU
BGV007161	0.690	11.833	0.023	17.1	822.9	0.171	280.8	ECU
BGV007181	0.553	10.848	0.016	19.6	956.3	0.141	344.8	ECU
BGV007194	0.620	12.681	0.016	20.4	929.6	0.144	316.5	ECU
BGV007198	0.828	12.273	0.023	14.8	726.6	0.171	286.2	ECU
BGV007366	0.568	11.167	0.015	19.6	868.9	0.138	321.4	
BGV007901	0.669	10.583	0.026	15.8	885.8	0.182	307.3	MEX
BGV007902	0.623	11.000	0.018	17.7	1011.0	0.153	375.1	MEX
BGV007909	0.784	10.972	0.027	14.0	750.3	0.185	289.7	MEX
BGV007911	0.815	10.333	0.029	12.7	737.0	0.194	300.4	MEX
BGV007918	0.764	10.611	0.023	13.9	754.2	0.172	316.5	MEX
BGV007927	0.753	10.861	0.017	14.4	618.9	0.146	294.5	MEX
BGV007931	0.668	10.231	0.013	15.3	657.2	0.126	339.4	MEX
BGV007934	0.717	10.056	0.016	14.0	728.6	0.141	368.7	MEX
BGV007935	0.738	8.444	0.031	11.4	573.4	0.198	252.9	MEX
BGV007992	1.414	15.852	0.032	11.2	835.2	0.202	369.0	PER
BGV008218	1.152	11.889	0.049	10.3	747.3	0.249	291.2	CRI
BGV008225	0.643	9.833	0.029	15.3	781.6	0.193	265.1	NIC
BGV008345	1.193	9.833	0.042	8.2	518.2	0.232	270.9	SLV

BGV008348	1.772	11.300	0.069	6.4	477.8	0.297	252.3	CRI
BGV012615	0.730	11.361	0.029	15.6	808.9	0.192	270.4	COL
BGV012626	0.918	13.000	0.026	14.2	902.9	0.182	349.8	COL
BGV013134	0.968	11.458	0.036	11.8	706.7	0.215	277.7	CRI
BGV013175	1.006	10.485	0.044	10.4	703.5	0.238	284.0	COL
BGV015380	0.601	11.333	0.013	18.9	814.8	0.128	337.9	PER
BGV015382	0.779	9.194	0.022	11.8	534.7	0.167	271.9	PER
LA1712	1.180	10.905	0.027	9.2	572.2	0.185	335.2	CRI
LA2697	0.876	9.722	0.037	11.1	698.2	0.217	290.3	COL
PAS014479	0.489	10.750	0.013	22.0	945.0	0.130	331.1	PER
PI129026	2.548	14.917	0.083	5.9	1199.9	0.324	632.1	ECU
PI129033	1.794	13.139	0.053	7.3	1334.7	0.259	702.7	ECU
BGV008221	1.056	11.472	0.038	10.9	761.7	0.221	317.2	SLV
BGV012639	1.898	12.972	0.066	6.8	685.3	0.291	344.7	ECU
BGV008041	2.028	14.694	0.088	7.2	550.5	0.335	226.8	PER
PI406890				NA	NA		NA	HND
BGV007920	0.984	11.472	0.028	11.7	775.3	0.190	350.3	MEX
BGV012613	1.095	12.639	0.042	11.5	941.4	0.232	351.0	PER
BGV007862	4.105	16.750	0.171	4.1	539.7	0.466	283.7	MEX
Tegucigalpa	4.515	15.139	0.175	3.4	327.0	0.473	206.4	
BGV006148	1.338	14.833	0.039	11.1	917.8	0.223	370.5	ECU
BGV007933	0.665	10.722	0.023	16.1	958.1	0.173	343.5	MEX
BGV005895	1.311	12.250	0.047	9.3	741.3	0.244	325.5	ECU
BGV006234	1.889	14.361	0.065	7.6	737.8	0.288	336.6	ECU
BGV007860	2.766	16.611	0.089	6.0	1216.4	0.336	602.4	MEX
BGV007864	2.469	14.056	0.087	5.7	1210.3	0.333	638.9	MEX

BGV007867	2.395	15.083	0.073	6.3	1299.9	0.306	675.1	MEX
BGV007872	3.676	19.750	0.122	5.4	1178.8	0.395	556.1	MEX
BGV007936	3.228	19.778	0.054	6.1	1400.7	0.262	872.4	MEX
BGV008224	2.626	13.667	0.102	5.2	708.2	0.360	378.5	NIC
BGV008354	1.831	15.455	0.037	8.4	966.6	0.217	528.4	CRI
BGV007854	3.082	18.222	0.115	5.9	900.4	0.383	397.9	MEX
BGV007857	3.279	19.194	0.121	5.9	932.8	0.392	406.3	MEX
BGV007865	3.619	22.485	0.107	6.2	1188.3	0.369	517.7	MEX
BGV007871	3.987	22.407	0.069	5.6	995.9	0.297	596.4	MEX
BGV007875	3.152	16.278	0.105	5.2	925.3	0.366	489.1	MEX
BGV007876	3.700	19.182	0.091	5.2	987.4	0.340	559.9	MEX
BGV007895	2.525	19.278	0.067	7.6	1448.6	0.291	651.1	MEX
BGV008077	1.875	13.569	0.072	7.2	782.0	0.303	356.2	PER
PI378994	1.367	13.000	0.047	9.5	705.1	0.244	304.1	
BGV006235	2.043	18.222	0.068	8.9	864.1	0.295	328.2	ECU
BGV006828	1.567	18.111	0.040	11.6	1308.7	0.225	502.4	ECU
BGV008036	1.958	14.000	0.085	7.1	623.2	0.328	265.5	PER
BGV008061	1.319	12.667	0.055	9.6	792.7	0.264	312.5	MEX
BGV008065	1.304	11.917	0.045	9.1	996.3	0.238	457.6	PER
BGV008095	1.215	11.879	0.057	9.8	841.8	0.270	318.5	PER
BGV008098	1.137	11.194	0.045	9.8	870.2	0.240	369.0	PER
BGV008100	1.682	13.700	0.060	8.1	761.6	0.276	338.7	PER
BGV008106	1.134	12.722	0.048	11.2	845.5	0.248	303.9	PER
BGV012640	1.257	10.639	0.067	8.5	781.4	0.292	316.4	PER
BGV014515	2.903	14.444	0.127	5.0	579.3	0.402	289.6	PER
BGV014516	1.386	14.389	0.055	10.4	909.7	0.263	332.7	PER

BGV014518	1.935	14.583	0.103	7.5	785.1	0.363	287.1	PER
BGV014519	1.987	13.278	0.103	6.7	675.6	0.362	279.1	PER
BGV014522	1.467	12.833	0.028	8.7	858.2	0.188	523.1	PER
BGV015726	2.491	14.500	0.100	5.8	648.7	0.357	312.5	PER
LA2309	1.724	12.667	0.071	7.3	814.5	0.301	367.8	PER
BGV004584	2.219	15.194	0.090	6.8	622.9	0.339	268.1	COL
BGV006229	1.746	14.028	0.053	8.0	695.2	0.260	333.1	ECU
BGV006806	1.714	16.778	0.064	9.8	686.1	0.286	244.9	ECU
BGV006906	1.323	15.333	0.039	11.6	995.6	0.223	385.5	ECU
BGV007908	0.730	12.045	0.022	16.5	868.7	0.169	311.9	MEX
BGV007910	0.710	9.639	0.024	13.6	728.7	0.176	305.4	MEX
BGV007921	0.702	10.889	0.020	15.5	826.7	0.158	336.8	MEX
BGV007989	1.421	11.778	0.065	8.3	624.5	0.288	261.6	PER
BGV008037	1.553	12.389	0.096	8.0	571.9	0.350	204.7	PER
BGV008051	0.859	11.778	0.034	13.7	758.1	0.209	264.5	MEX
BGV008058	0.976	11.306	0.036	11.6	785.7	0.214	317.4	MEX
BGV008067	0.682	9.403	0.022	13.8	759.5	0.168	327.9	MEX
BGV008070	0.623	10.417	0.021	16.7	958.6	0.164	349.6	MEX
BGV008219	0.912	11.278	0.031	12.4	800.9	0.200	323.5	CRI
BGV008223	0.874	9.972	0.042	11.4	689.9	0.231	261.4	HND
BGV012627	1.121	12.694	0.032	11.3	996.6	0.202	435.3	COL
CATIE-11106/1	0.699	8.944	0.033	12.8	773.4	0.206	293.9	HND
BGV006231	1.564	16.389	0.052	10.5	905.5	0.256	337.2	ECU
BGV006775	1.043	13.389	0.037	12.8	674.1	0.218	240.5	ECU
BGV006777	2.025	13.417	0.090	6.6	555.3	0.339	247.0	ECU

BGV006899	2.198	15.400	0.074	7.0	655.5	0.307	304.4	ECU
BGV006907	1.373	14.333	0.042	10.4	1046.0	0.232	432.2	ECU
BGV007015	1.788	16.722	0.045	9.4	959.1	0.240	426.8	ECU
BGV007017	2.444	18.288	0.060	7.5	723.0	0.276	350.0	ECU
BGV007023	1.797	12.444	0.063	6.9	615.4	0.284	313.1	ECU
BGV007169	0.693	11.056	0.027	15.9	775.4	0.187	259.9	ECU
BGV007981	2.094	12.933	0.121	6.2	526.9	0.392	217.7	PER
BGV008096	1.629	12.194	0.061	7.5	646.2	0.279	309.2	PER
BGV008108	0.963	11.182	0.023	11.6	774.8	0.170	393.5	PER
BGV013161	1.388	12.667	0.046	9.1	808.6	0.242	365.7	PER
BGV015730	1.494	15.300	0.059	10.2	985.3	0.275	350.3	PER
BGV005912	1.897	18.056	0.057	9.5	1010.4	0.269	394.7	ECU
BGV006927	2.438	17.056	0.071	7.0	732.1	0.301	347.2	ECU
BGV006931	1.121	10.815	0.046	9.7	899.8	0.241	386.1	ECU
BGV006934	1.580	13.472	0.038	8.5	939.0	0.220	501.0	ECU
BGV007990	1.227	9.133	0.058	7.4	673.6	0.272	332.7	PER
BGV008042	0.984	13.167	0.035	13.4	800.9	0.211	283.3	PER
BGV014508	1.662	13.556	0.053	8.2	854.7	0.260	402.9	PER
BGV014514	1.821	13.500	0.056	7.4	877.5	0.267	443.4	PER
BGV015727	2.256	13.903	0.053	6.2	603.0	0.260	376.2	PER
PI129088	2.465	16.389	0.070	6.6	717.1	0.299	361.2	COL
PI487625	3.208	19.400	0.094	6.0	747.1	0.346	356.6	CRI
BGV007878	3.137	17.571	0.065	5.6	1084.4	0.289	670.9	MEX
BGV007900	2.078	14.939	0.067	7.2	1213.0	0.292	578.1	MEX
BGV015734	1.909	11.861	0.088	6.2	1221.0	0.335	585.9	PER
Voyage	2.222	15.333	0.080	6.9	1772.0	0.319	805.0	

BGV006175				NA	NA			NA	ECU
BGV006767	1.338	14.485	0.032	10.8	773.4	0.202		353.9	ECU
BGV006825	2.654	14.639	0.109	5.5	694.5	0.373		337.3	ECU
BGV006852	2.336	15.983	0.079	6.8	927.8	0.317		427.4	ECU
BGV006881	1.237	14.111	0.031	11.4	1000.6	0.197		444.4	ECU
BGV006896	1.671	15.083	0.034	9.0	670.7	0.209		355.6	ECU
BGV007339	0.553	10.333	0.018	18.7	881.4	0.152		310.8	
BGV007863	3.378	16.167	0.114	4.8	590.9	0.381		323.7	MEX
BGV007870				NA	NA			NA	MEX
BGV007894	1.884	14.194	0.062	7.5	677.7	0.282		318.9	MEX
BGV007899	2.388	18.333	0.051	7.7	1336.2	0.255		681.8	MEX
BGV008189	1.177	11.455	0.048	9.7	673.0	0.247		280.1	PER
BGV012614				NA	NA			NA	MEX
BGV012625	2.472	17.521	0.079	7.1	846.6	0.317		377.3	PER
BGV013945	1.339	13.028	0.038	9.7	919.1	0.219		431.8	PER
LA0767	2.872	16.167	0.066	5.6	1235.8	0.289		758.9	

Supplementary table S2.4. Physical position for the significant SNP in the QTL detected with GWAS for the non-derived phenotypes. SNP that overlapped across phenotypes are shown in bold.

Chr*	Known fruit weight genes	Causal mutation/ locus	Weight	Locule number	Cell layers	Thickness	Max cell size	Columella + placenta area	Pericarp area	Pericarp + septum area
0				SL25ch00p 17916153						
2				SL25ch02p 47088126						
2				SL25ch02p 47104838						
2				SL25ch02p 47108457						
2				SL25ch02p 47109224						
2				SL25ch02p 47120157						
2			**SL25ch02p 47123272**	**SL25ch02p 47123272**						
2				SL25ch02p 47124369						
2			**SL25ch02p 47138907**	**SL25ch02p 47138907**						
2				SL25ch02p 47148187						
2				SL25ch02p 47166327						
2			**SL25ch02p 47167029**	**SL25ch02p 47167029**						
2				SL25ch02p						

				47168841						
			SL25ch02p	**SL25ch02p**				**SL25ch02p**		
2			**47171876**	**47171876**				**47171876**		
				SL25ch02p						
2				47173189						
				SL25ch02p						
2				47174126						
				SL25ch02p						
2				47176795						
				SL25ch02p						
2				47177366						
				SL25ch02p						
2				47178313						
				SL25ch02p						
2				47179658						
				SL25ch02p						
2				47182871						
			SL25ch02p	**SL25ch02p**				**SL25ch02p**		
2			**47183456**	**47183456**				**47183456**		
		SL25ch02p	**SL25ch02p**	**SL25ch02p**				**SL25ch02p**		
2	**LC**	**47188498**	**47188498**	**47188498**				**47188498**		
		SL25ch02p	**SL25ch02p**	**SL25ch02p**						
2		**47188504**	**47188504**	**47188504**						
				SL25ch02p						
2				47193596						
				SL25ch02p						
2				47204573						
				SL25ch02p				**SL25ch02p**		
2				**47204967**				**47204967**		
				SL25ch02p				**SL25ch02p**		
2				**47218361**				**47218361**		
2				SL25ch02p						

		47233028						
		SL25ch02p				**SL25ch02p**		
2		**47282426**				**47282426**		
		SL25ch02p				**SL25ch02p**		
2		**47293413**				**47293413**		
		SL25ch02p						
2		47327709						
		SL25ch02p						
2		47335097						
		SL25ch02p						
2		47355580						
		SL25ch02p						
2		47360051						
		SL25ch02p				**SL25ch02p**		
2		**47361497**				**47361497**		
		SL25ch02p				**SL25ch02p**		
2		**47366168**				**47366168**		
		SL25ch02p						
2		47377392						
		SL25ch02p				**SL25ch02p**		
2		**47396239**				**47396239**		
		SL25ch02p						
2		47406245						
		SL25ch02p						
2		47409176						
		SL25ch02p						
2		47414088						
		SL25ch02p						
2		47431348						
		SL25ch02p						
2		47451557						
2		SL25ch02p						

		47456686						
		SL25ch02p						
2		47478528						
		SL25ch02p						
2		47486548						
		SL25ch02p						
2		47513948						
		SL25ch02p						
2		47522462						
		SL25ch02p						
2		47649373						
		SL25ch02p						
2		47652183						
		SL25ch02p						
2		47674763						
		SL25ch02p						
2		47677559						
	SL25ch02p	**SL25ch02p**						
2	**47732780**	**47732780**						
	SL25ch02p	**SL25ch02p**						
2	**47742357**	**47742357**						
		SL25ch02p						
2		47747259						
		SL25ch02p						
2		47767560						
	SL25ch02p	**SL25ch02p**						
2	**47773270**	**47773270**						
		SL25ch02p						
2		47792270						
		SL25ch02p						
2		47999730						
2		SL25ch02p						

		48015576						
		SL25ch02p						
2		48040715						
		SL25ch02p						
2		49587330						
	SL25ch02p							
2	49613592							
		SL25ch02p						
2		49617538						
	SL25ch02p							
2	49617778							
	SL25ch02p							
2	49643914							
	SL25ch02p							
2	49940988							
	SL25ch02p		**SL25ch02p**					
2	**49956080**		**49956080**					
	SL25ch02p							
2	50027634							
	SL25ch02p							
2	50050280							
	SL25ch02p							
2	50055695							
	SL25ch02p							
2	50057236							
	SL25ch02p							
2	50069770							
	SL25ch02p		**SL25ch02p**					
2	**50078984**		**50078984**					
	SL25ch02p		**SL25ch02p**					
2	**50084865**		**50084865**					
2	**SL25ch02p**		**SL25ch02p**					

			50087336		**50087336**					
			SL25ch02p							
2			50101053							
					SL25ch02p					
2					50134793					
			SL25ch02p			**SL25ch02p**				
2			**50155828**			**50155828**				
			SL25ch02p							
2			50219947							
			SL25ch02p							
2			50262525							
				SL25ch02p						
2				50313381						
			SL25ch02p	**SL25ch02p**						
2			**50708874**	**50708874**						
			SL25ch02p							
2			51039990							
			SL25ch02p							
2			51185104							
			SL25ch02p							
2			51303773							
			SL25ch02p							
2			51326634							
		SL25ch02p								
2	**CNR**	52251645								
		SL25ch02p								
2		52251886								
			SL25ch02p			**SL25ch02p**				
2			**52708494**			**52708494**				
						SL25ch02p				
2						53794883				
2						SL25ch02p				

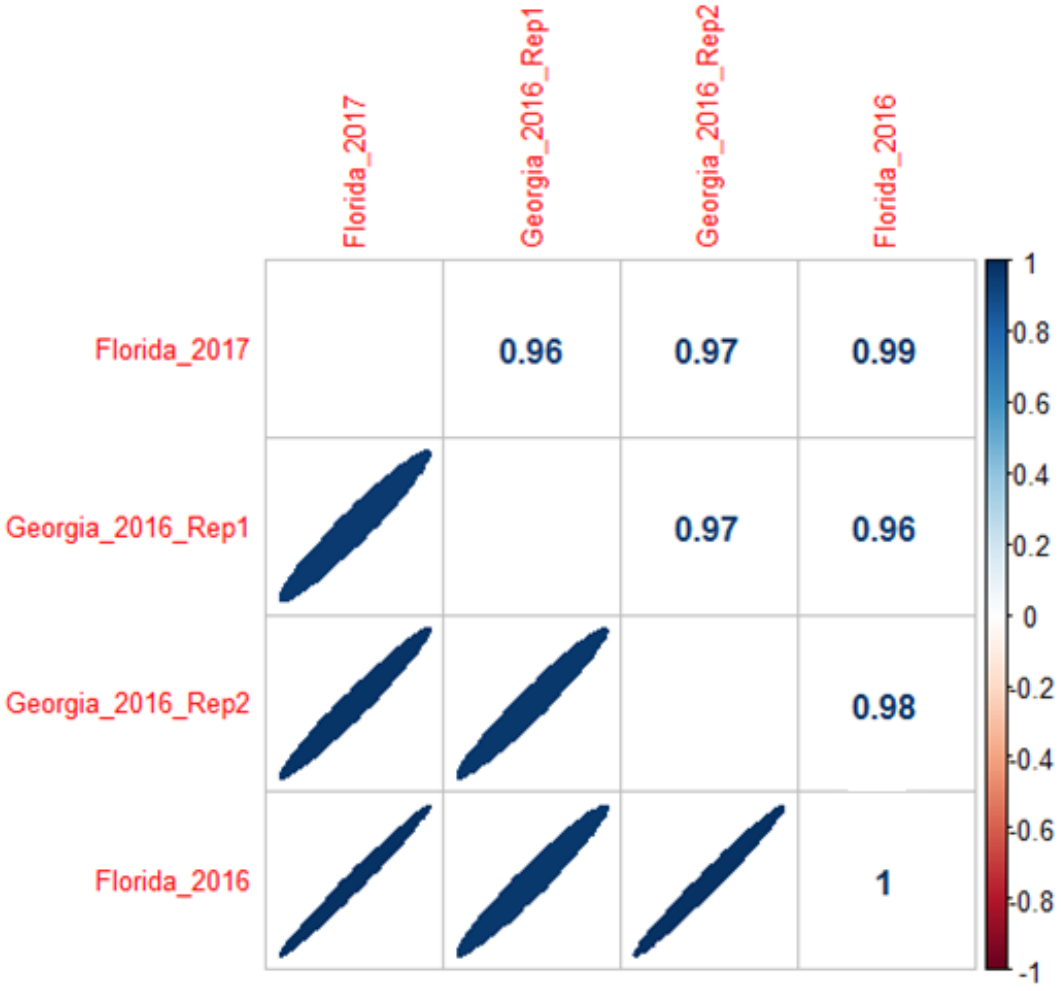
						53795138				
2						SL25ch02p 53800901				
3	**SIKLUH**	SL25ch02p 64799226								
3			SL25ch03p 69254047							
3			SL25ch03p 69258095							
3			SL25ch03p 69271979							
3			SL25ch03p 69582581							
3			SL25ch03p 69628839							
6			SL25ch06p 47398304							
6			SL25ch06p 47408277							
6			SL25ch06p 47414990							
8					SL25ch08p 55303149					
8					SL25ch08p 55327903					
8					SL25ch08p 55334646					
8			SL25ch08p 60568126							
11				SL25ch11p 54838887						

11				SL25ch11p 54843277						
11				SL25ch11p 54872584						
11				SL25ch11p 54873506						
11	**FAS**	SL25ch02p 54877107-55,171,481		SL25ch11p 54924884						
11				SL25ch11p 54934056						
11				SL25ch11p 54934637						
11				SL25ch11p 54938333						
11				SL25ch11p 54946315						
11				SL25ch11p 54954681						
11				SL25ch11p 54984105						
11				SL25ch11p 55003415						
11				SL25ch11p 55020323						
11				SL25ch11p 55021988						
11				SL25ch11p 55023960						

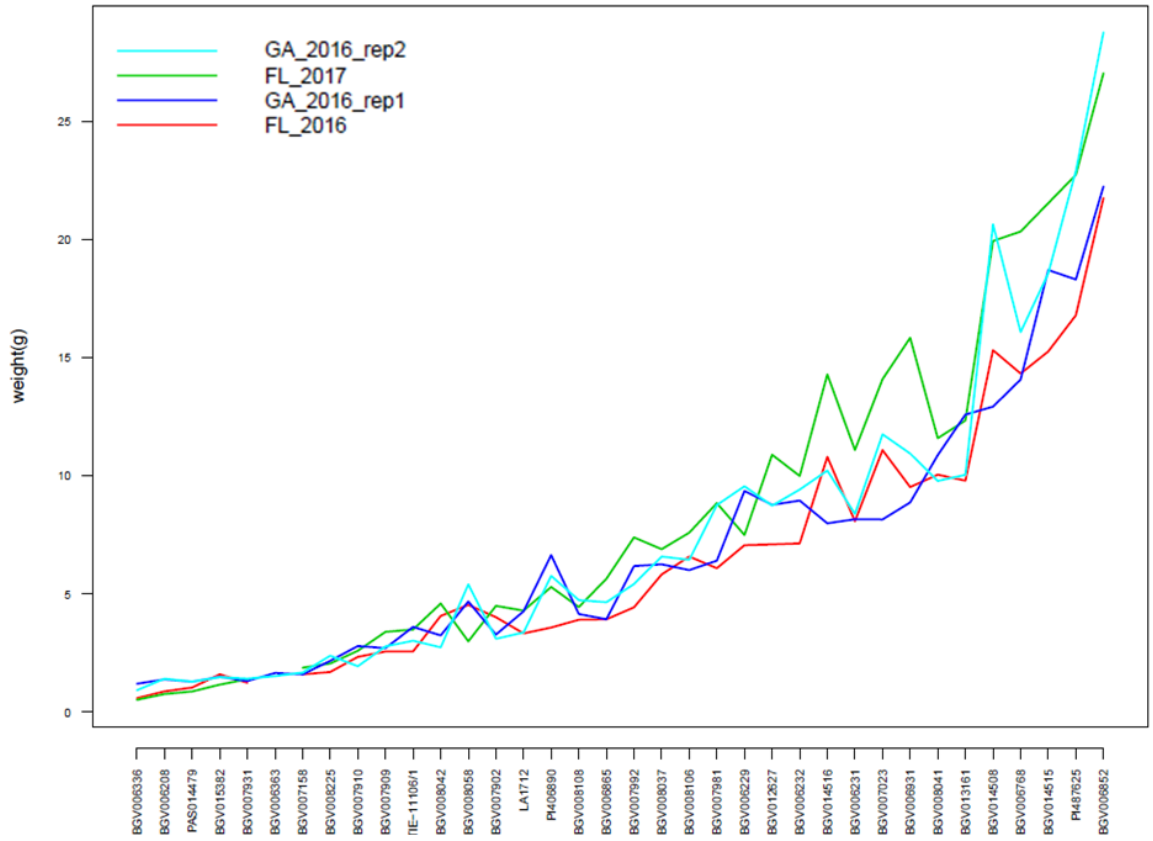
		SL25ch11p						
11		55029083						
		SL25ch11p						
11		55036151						
		SL25ch11p						
11		55055868						
		SL25ch11p						
11		55059726						
		SL25ch11p						
11		55061105						
		SL25ch11p						
11		55073707						
		SL25ch11p						
11		55099447						
		SL25ch11p						
11		55118012						
		SL25ch11p						
11		55151475						
		SL25ch11p						
11		55161262						
		SL25ch11p						
11		55171115						
		SL25ch11p						
11		55174587						
		SL25ch11p						
11		55176476						
		SL25ch11p						
11		55179668						
		SL25ch11p						
11		55182143						
		SL25ch11p						
11		55183870						

				SL25ch11p						
11				55190028						
				SL25ch11p						
11				55199321						
				SL25ch11p						
11				55201893						
				SL25ch11p						
11				55203155						
				SL25ch11p						
11				55203793						
				SL25ch11p						
11				55218629						
		SL25ch02p								
11	**CSR**	55265540								

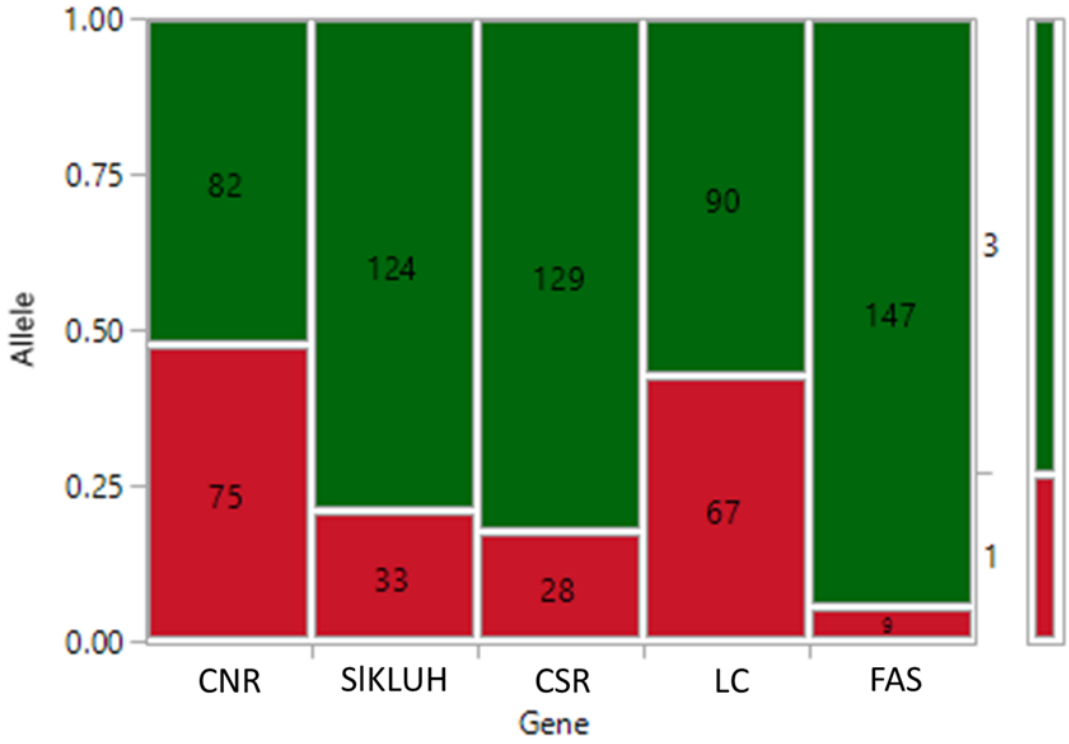
*Chromosome



Supplementary figure S2. 1. Weight correlations across years, locations, and replicates in the format Location_year_replicate. Color indicates the strength of the correlation where dark blue is a positive correlation of 1, and dark red is a negative correlation of 1. Coefficients of correlation (r) are also shown for each combination.

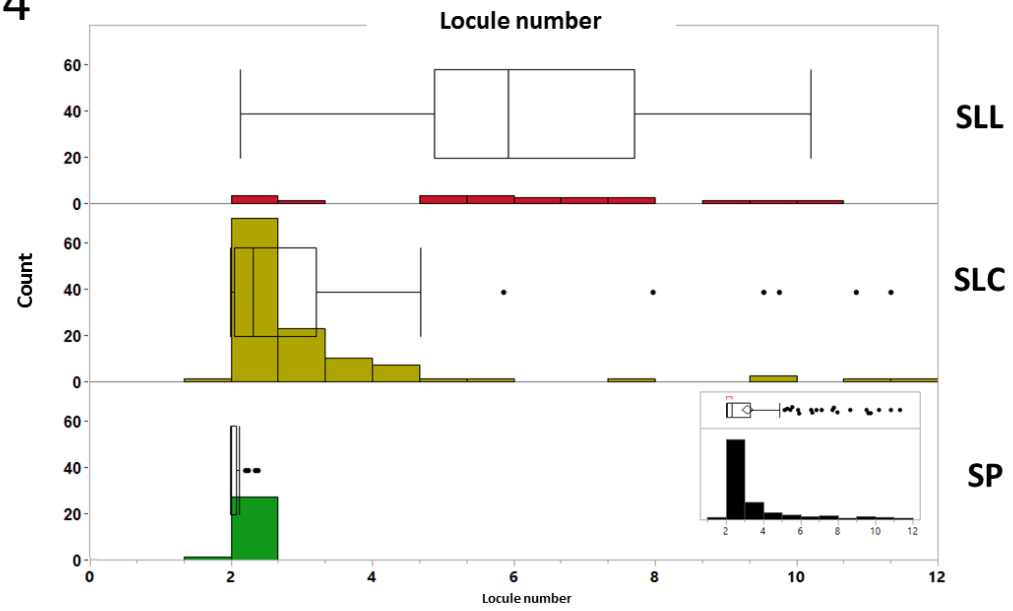


Supplementary figure S2. 2. Fruit weight interaction plots for 37 accessions grown in different years, locations, and replicates. Line labels in the format Location_year_replicate.

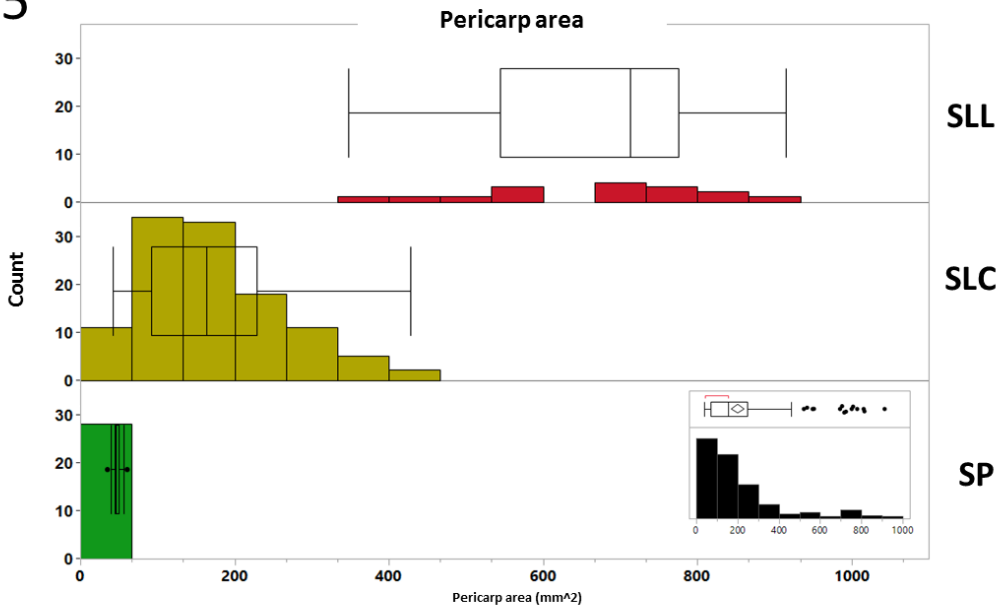


Supplementary figure S2. 3. Allele distribution for the five major fruit weight genes excluding heterozygotes (N =157). Wild allele (3) in green, derived allele (1) in red.

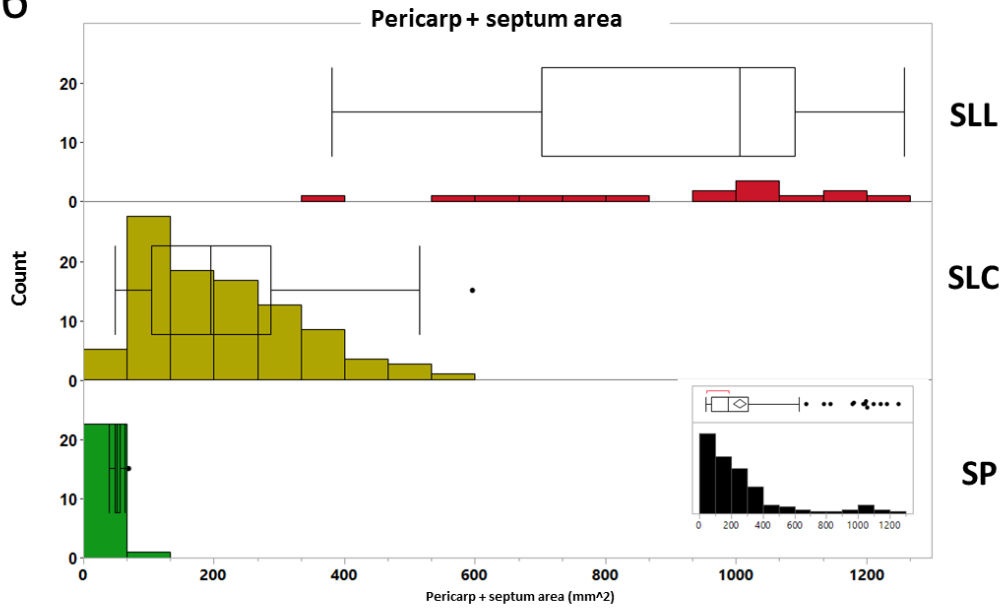
S2.4


Locule number
SLL
SLC
SP
Count
Locule number

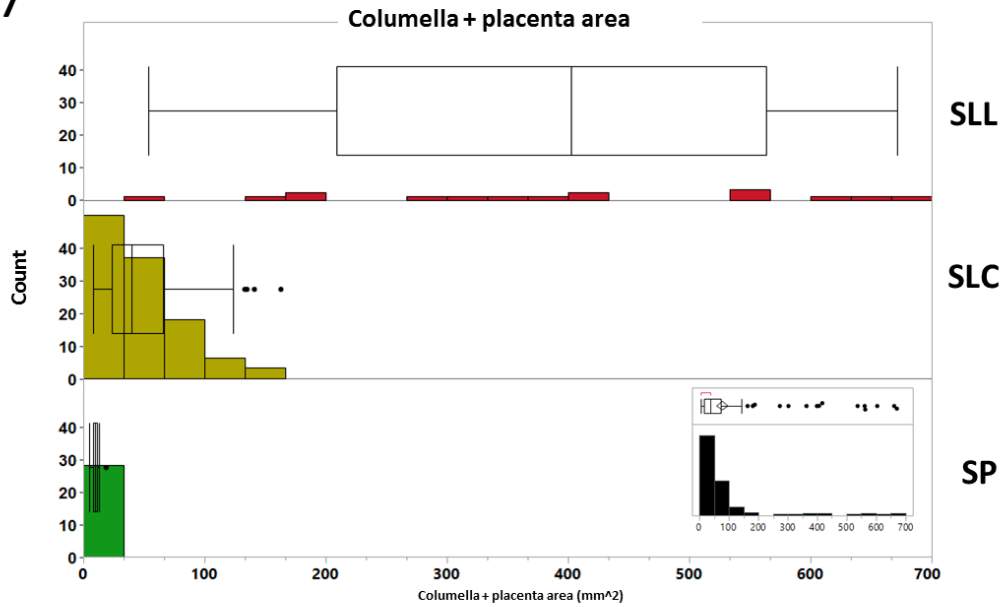
S2.5


Pericarp area
SLL
SLC
SP
Count
Pericarp area (mm^2)

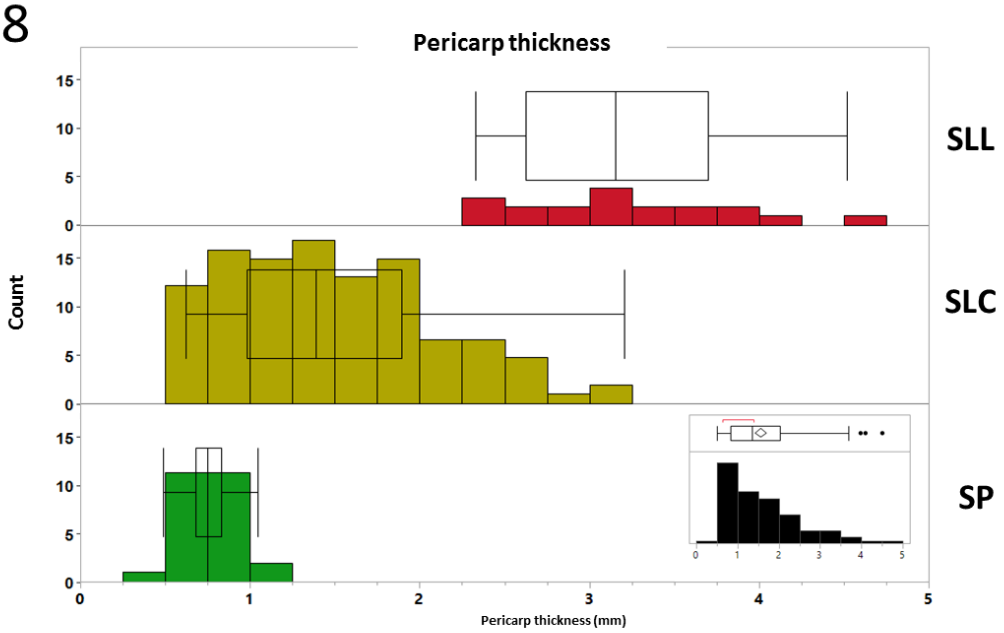
S2.6


Pericarp + septum area
SLL
SLC
SP
Count
Pericarp + septum area (mm^2)

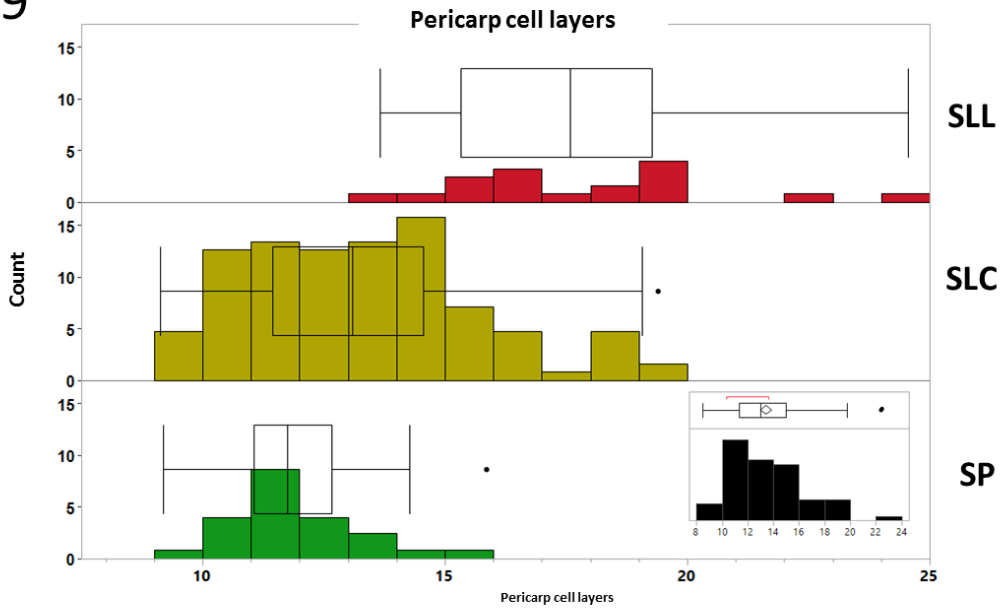
S2.7


Columella + placenta area
SLL
SLC
SP
Count
Columella + placenta area (mm^2)

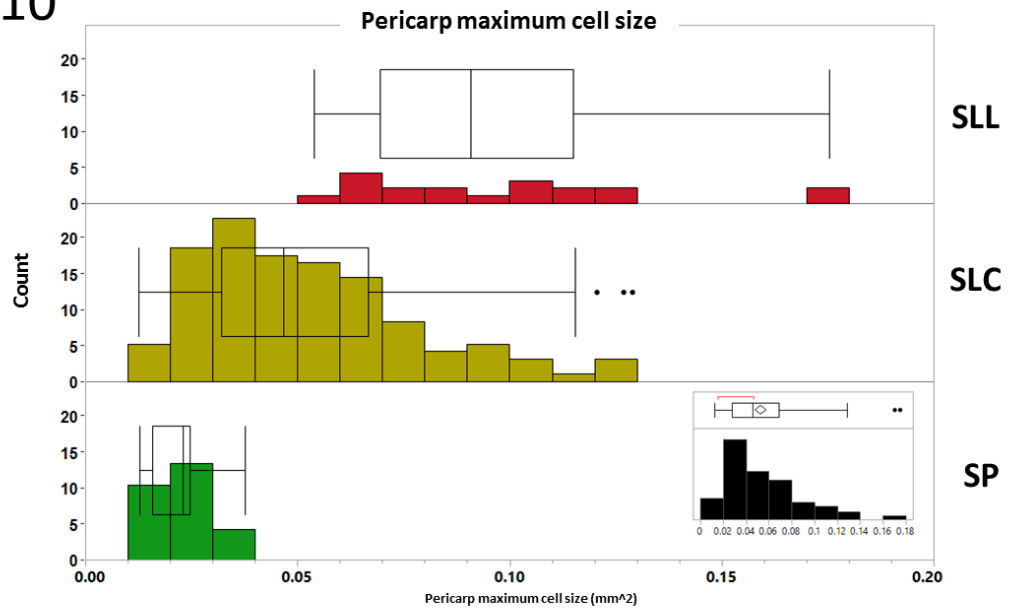
S2.8


Pericarp thickness
SLL
SLC
SP
Count
Pericarp thickness (mm)

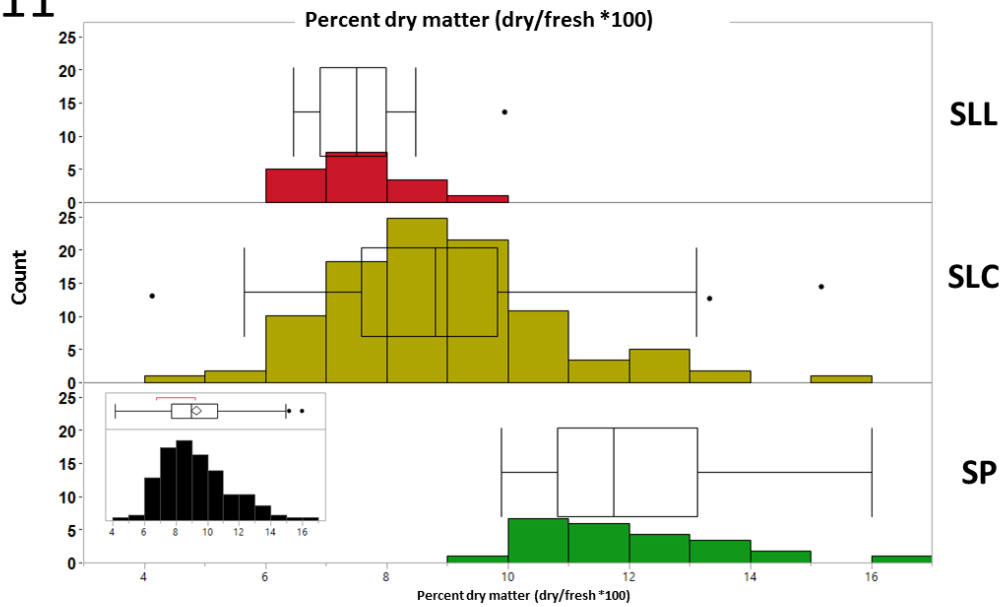
S2.9


Pericarp cell layers
SLL
SLC
SP
Count
Pericarp cell layers

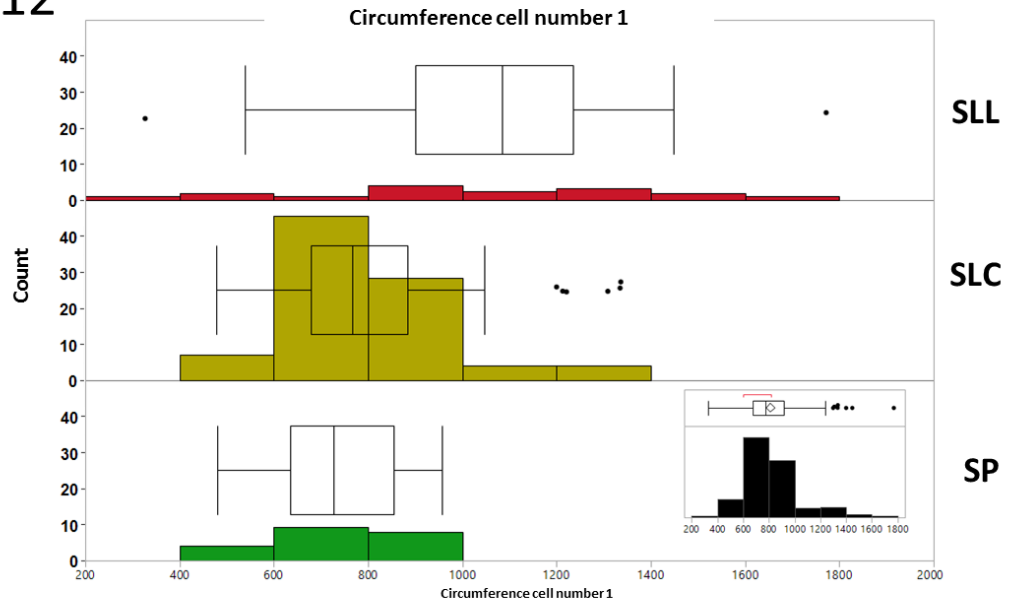
S2.10


Pericarp maximum cell size
SLL
SLC
SP
Count
Pericarp maximum cell size (mm^2)

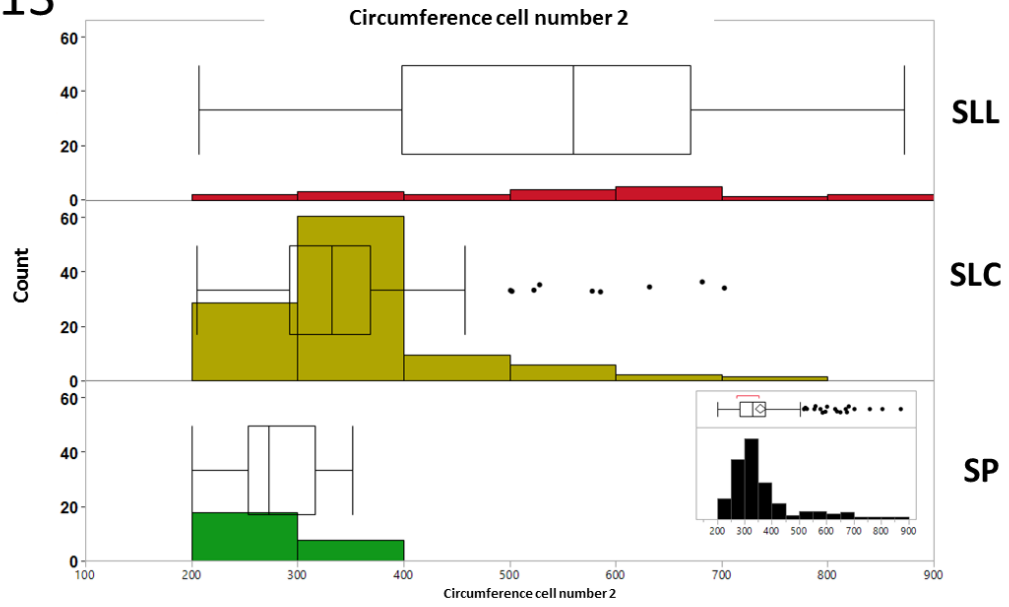
S2.11


Percent dry matter (dry/fresh *100)
SLL
SLC
SP
Count
Percent dry matter (dry/fresh *100)

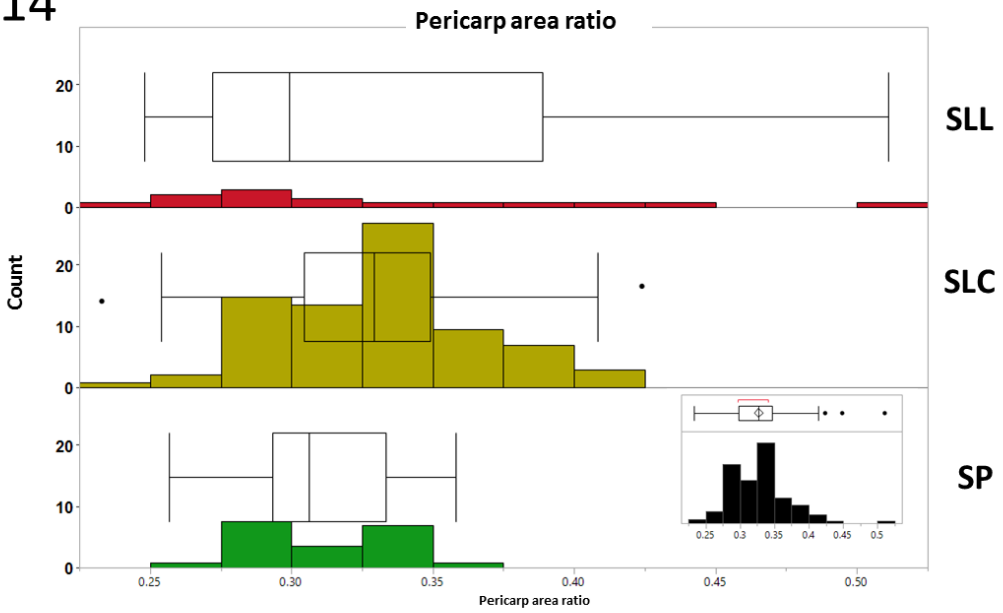
S2.12


Circumference cell number 1
SLL
SLC
SP
Count
Circumference cell number 1

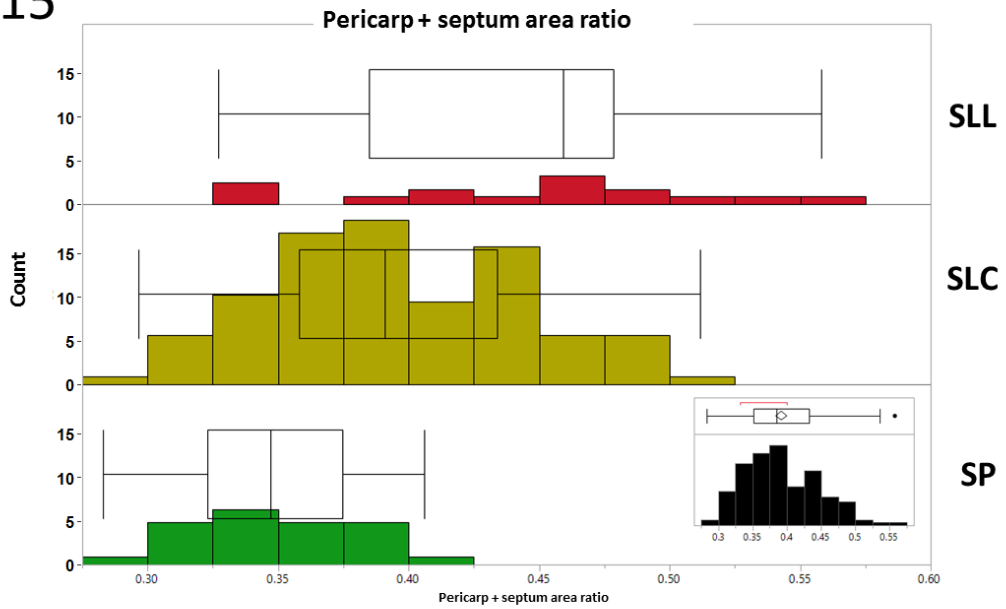
S2.13


Circumference cell number 2
SLL
SLC
SP
Count
Circumference cell number 2

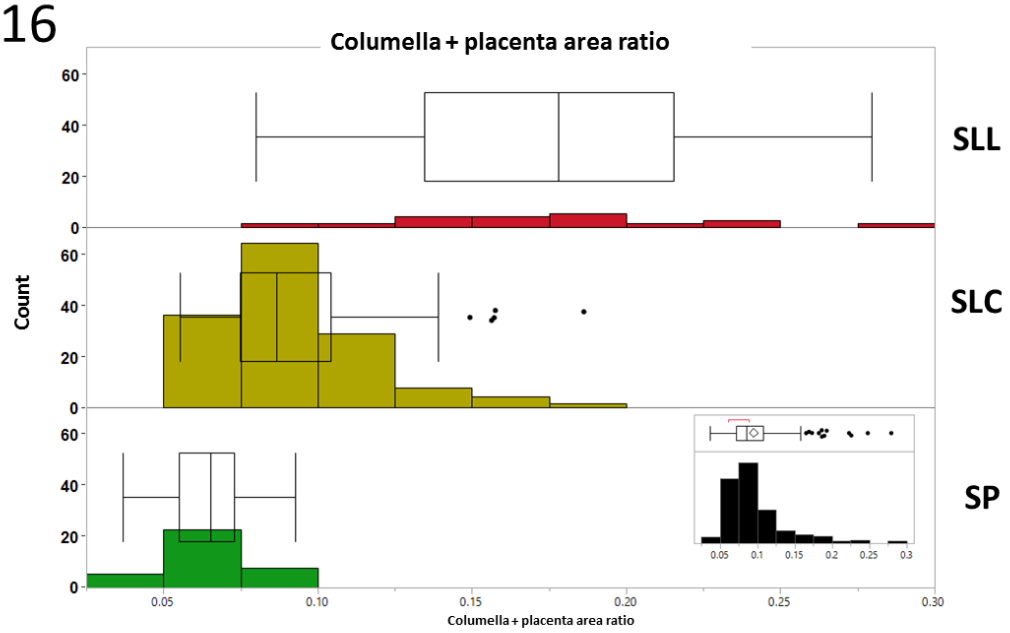
S2.14


Pericarp area ratio
SLL
SLC
SP
Count
Pericarp area ratio

S2.15

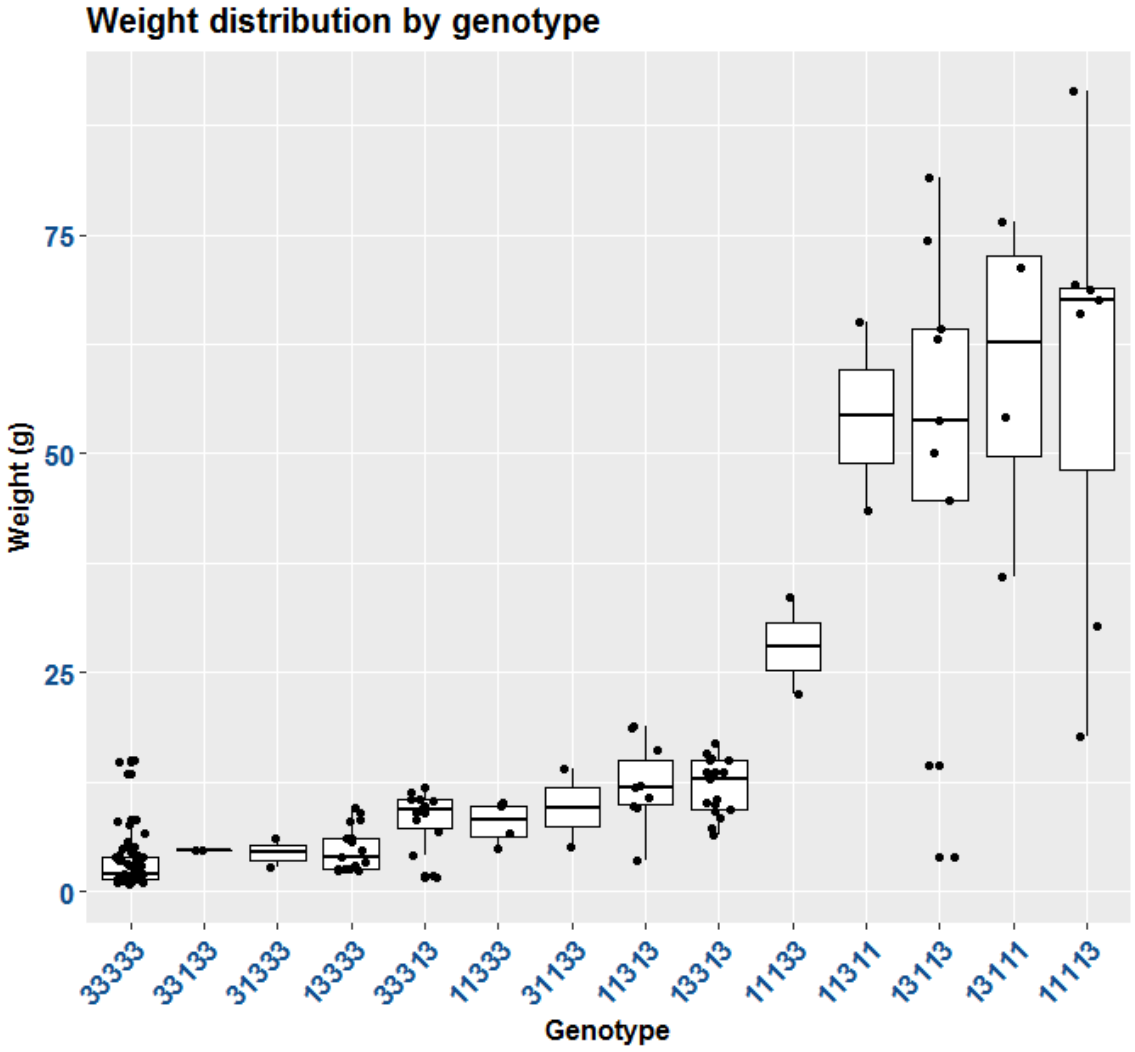

Pericarp + septum area ratio
SLL
SLC
SP
Count
Pericarp + septum area ratio

S2.16

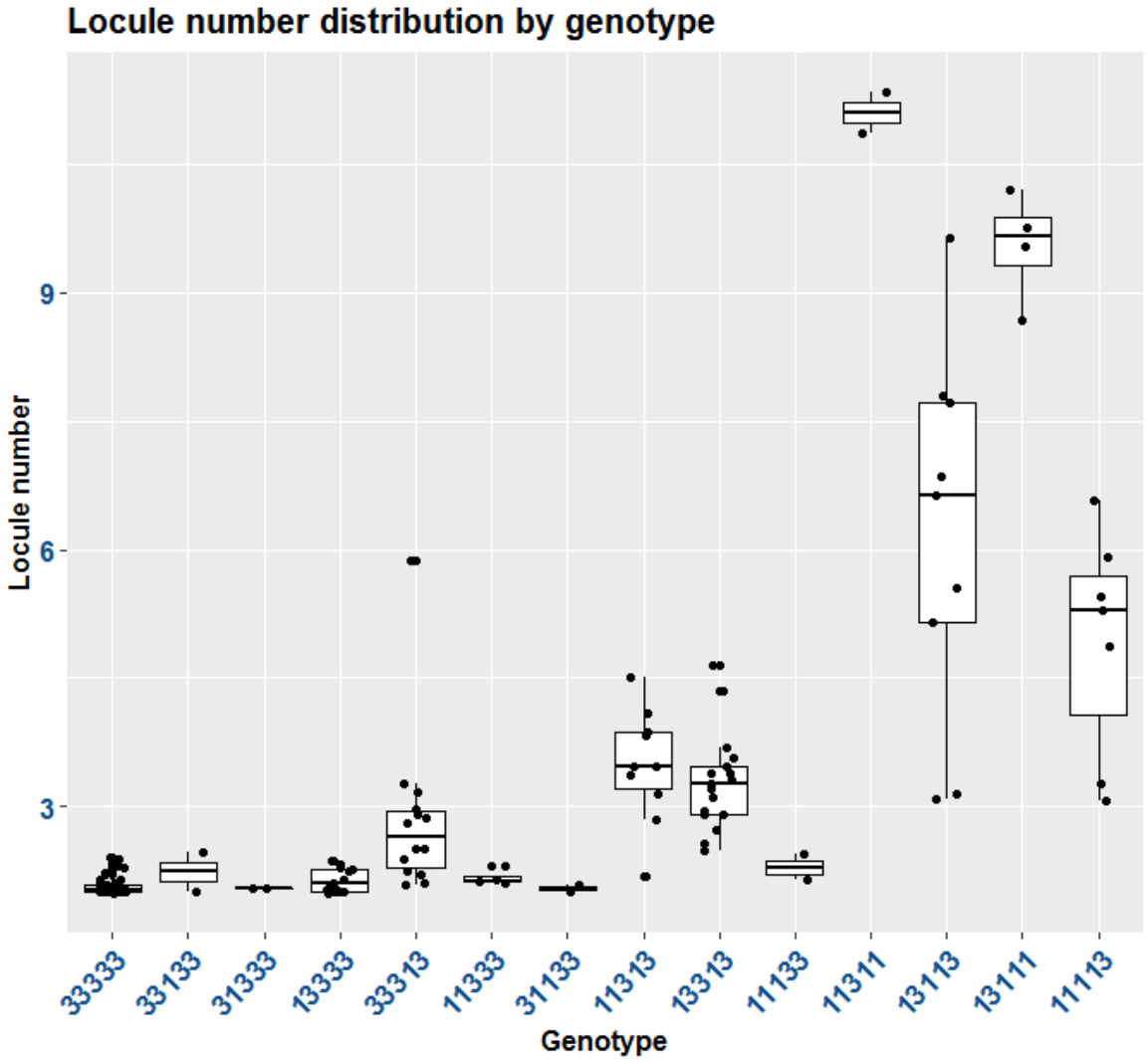


Supplementary figure S2. 4 – S2.16 Weight and weight parameters distributions for SP, SLC, and SLL. Small histogram in black is the distribution for the entire population.

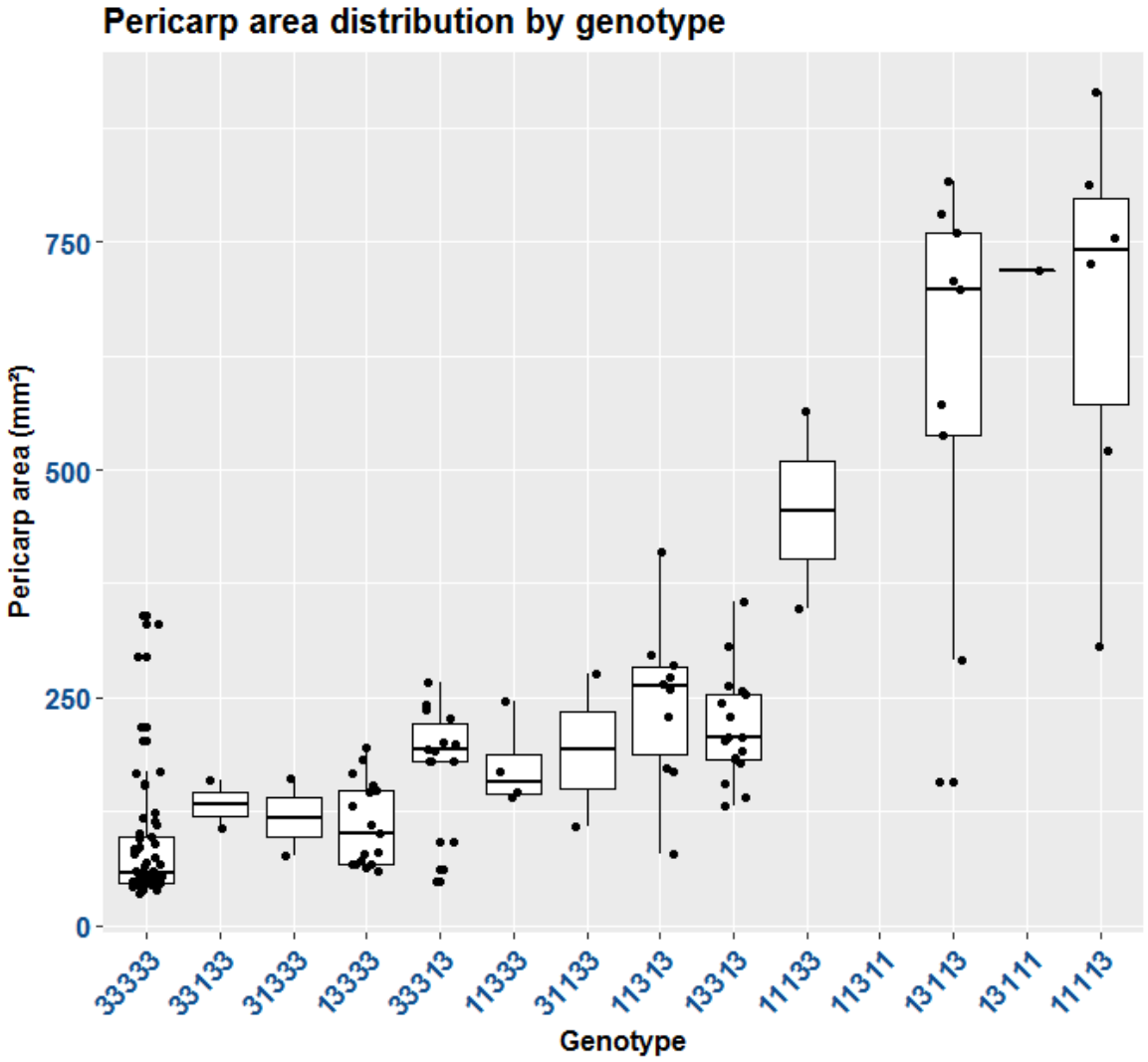
S2.17


Weight distribution by genotype
Weight (g)
75
50
25
0
33333
33133
31333
13333
33313
11333
31133
11313
13313
11133
11311
13113
13111
11113
Genotype

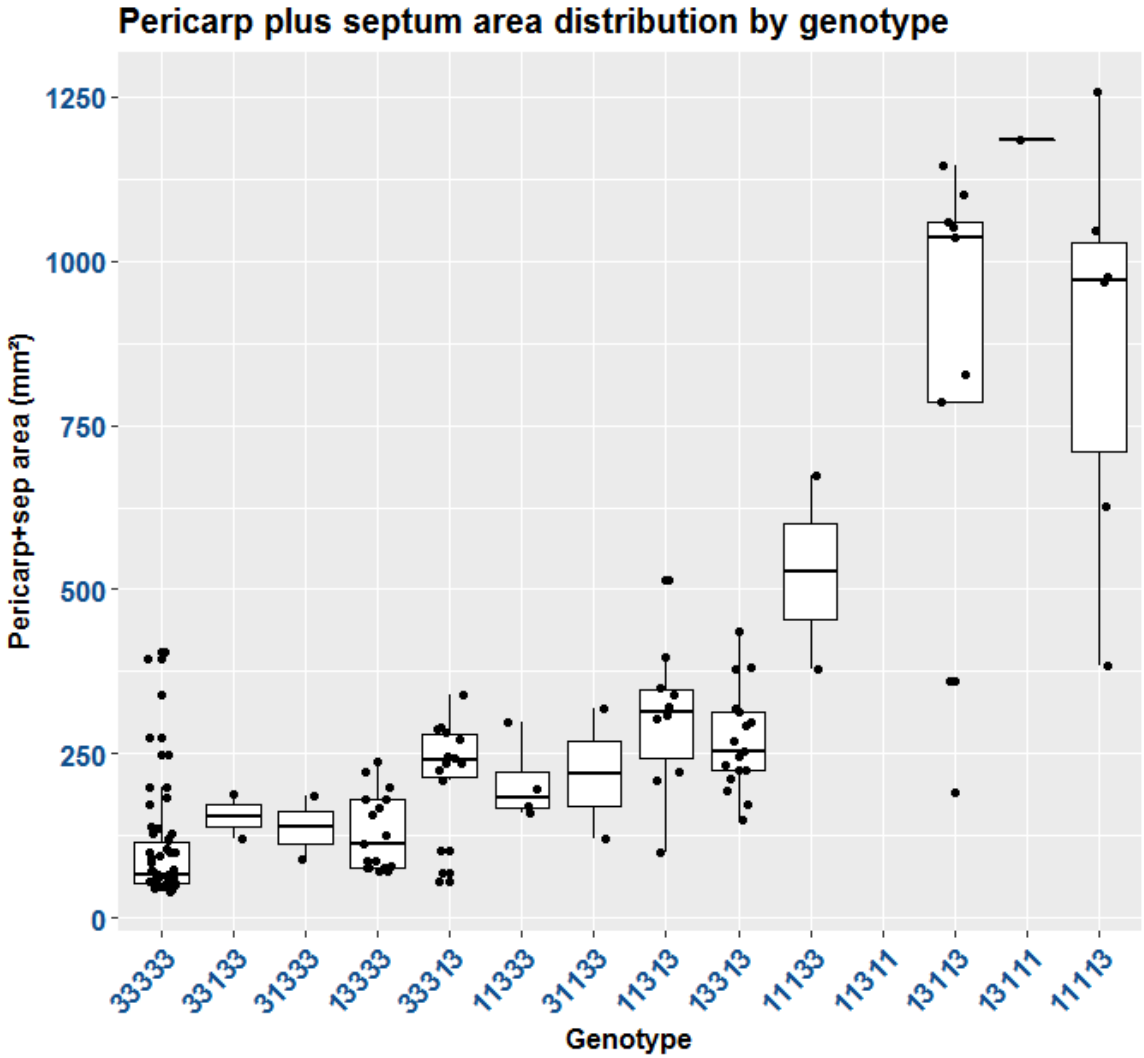
S2.18


Locule number distribution by genotype
Locule number
9
6
3
33333
33133
31333
13333
33313
11333
31133
11313
13313
11133
11311
13113
13111
11113
Genotype

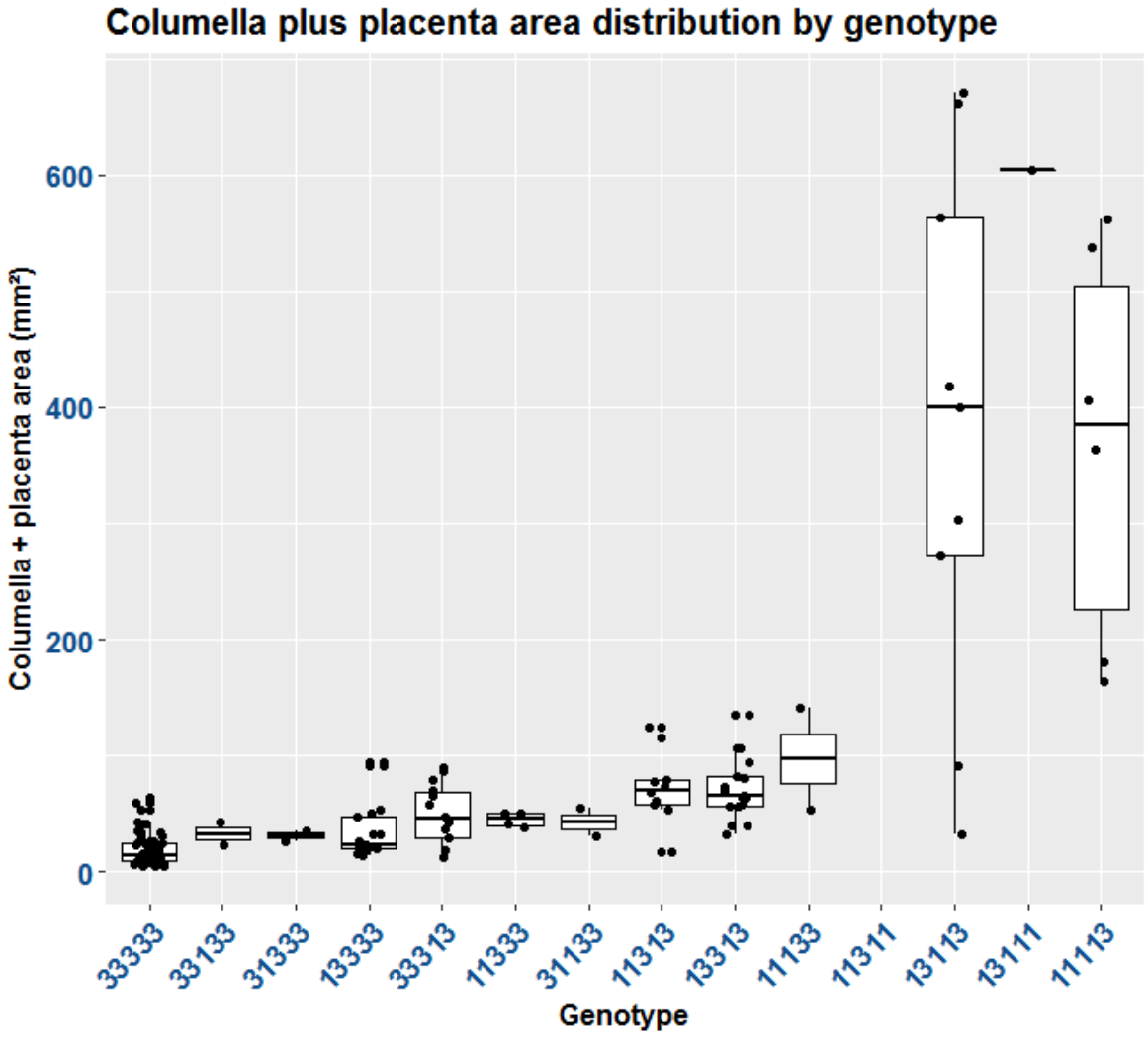
S2.19


Pericarp area distribution by genotype
Pericarp area (mm²)
0
250
500
750
33333
33133
31333
13333
33313
11333
31133
11313
13313
11133
11311
13113
13111
11113
Genotype

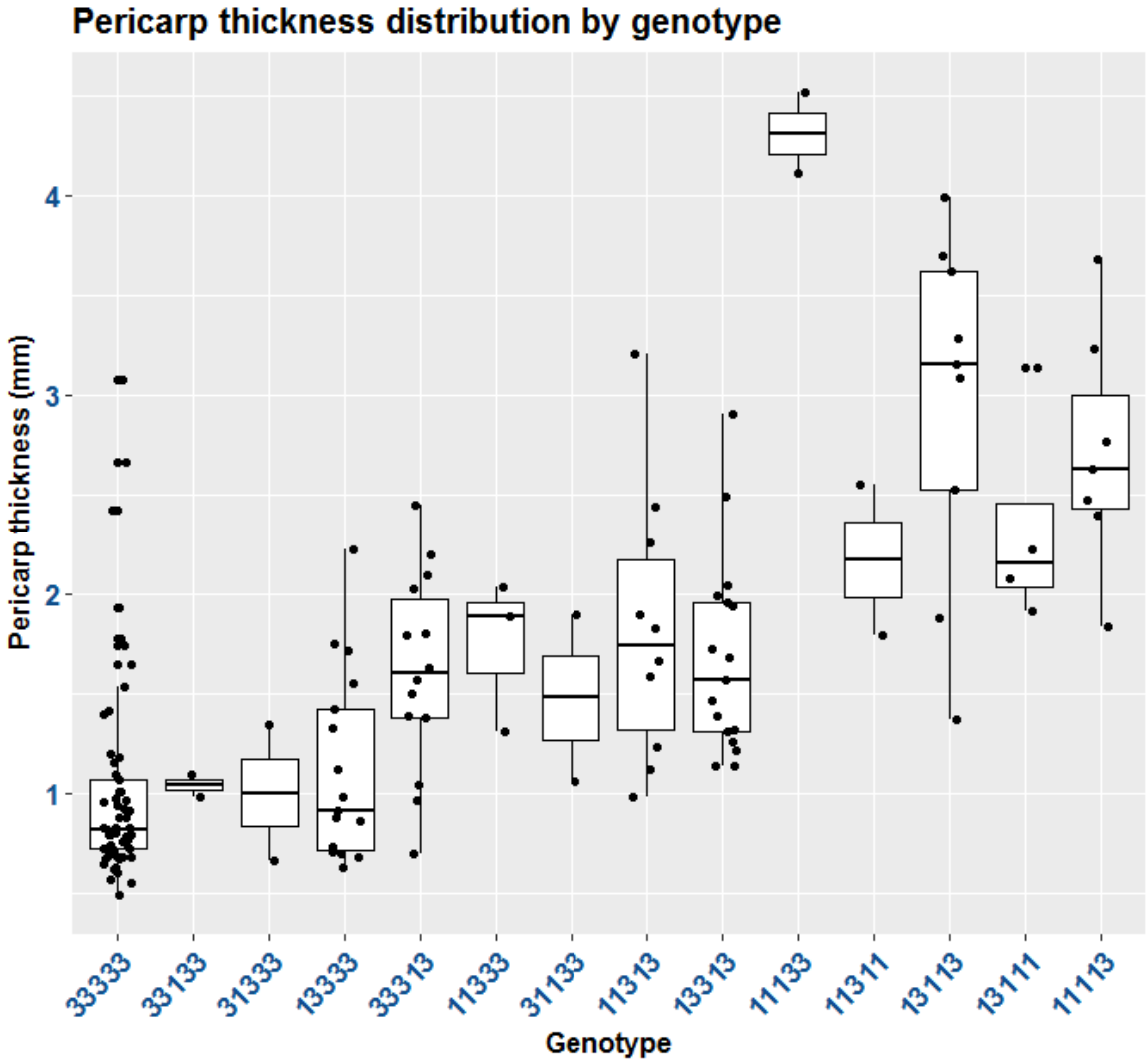
S2.20


Pericarp plus septum area distribution by genotype
Pericarp+sep area (mm²)
1250
1000
750
500
250
0
33333
33133
31333
13333
33313
11333
31133
11313
13313
11133
11311
13113
13111
11113
Genotype

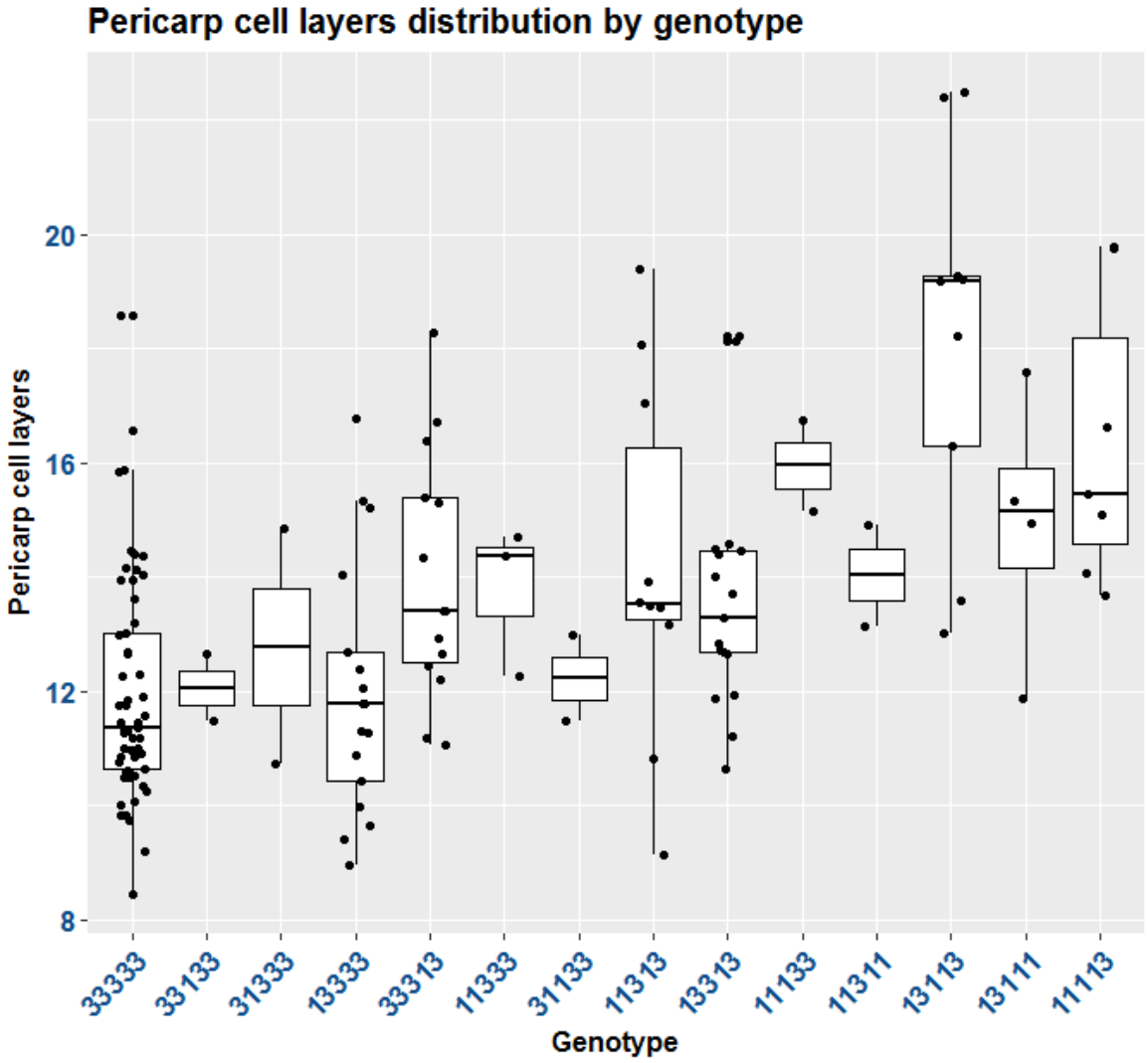
S2.21


Columella plus placenta area distribution by genotype
Columella + placenta area (mm²)
600
400
200
0
33333
33133
31333
13333
33313
11333
31133
11313
13313
11133
11311
13113
13111
11113
Genotype

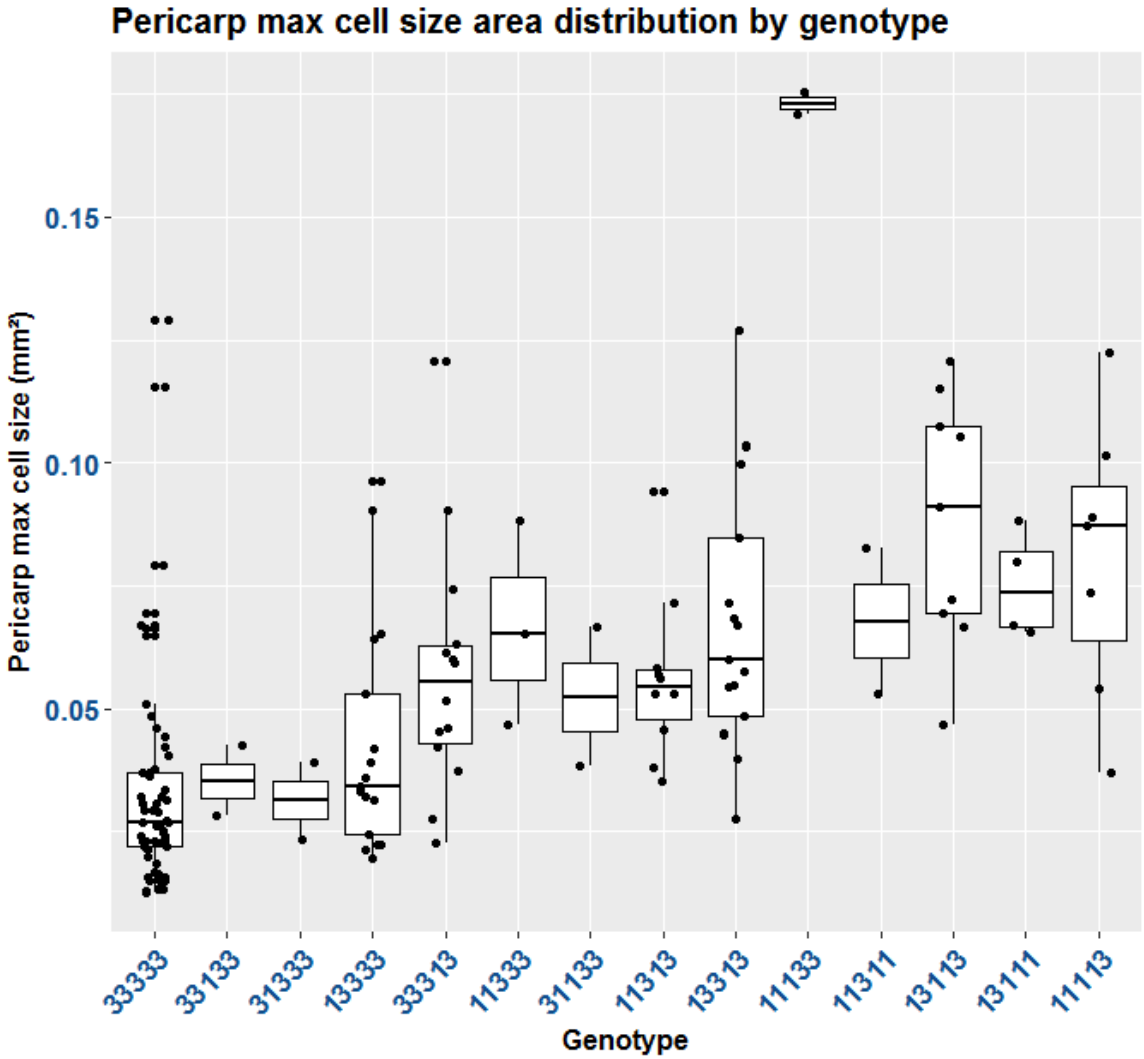
S2.22


Pericarp thickness distribution by genotype
Pericarp thickness (mm)
4
3
2
1
33333
33133
31333
13333
33313
11333
31133
11313
13313
11133
11311
13113
13111
11113
Genotype

S2.23

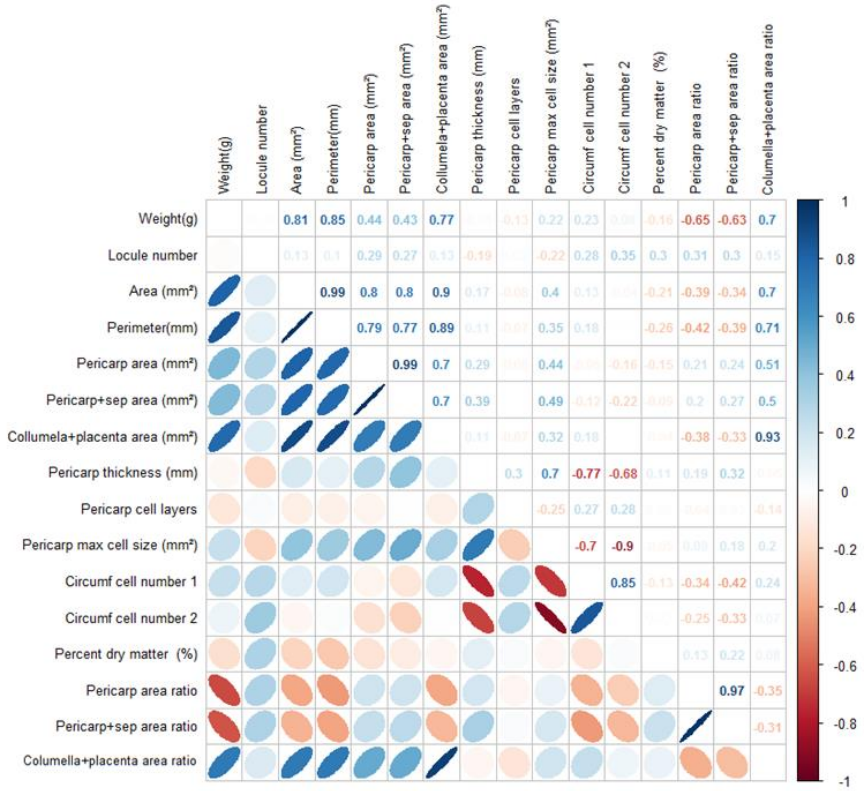

Pericarp cell layers distribution by genotype
Pericarp cell layers
20
16
12
8
33333
33133
31333
13333
33313
11333
31133
11313
13313
11133
11311
13113
13111
11113
Genotype

S2.24

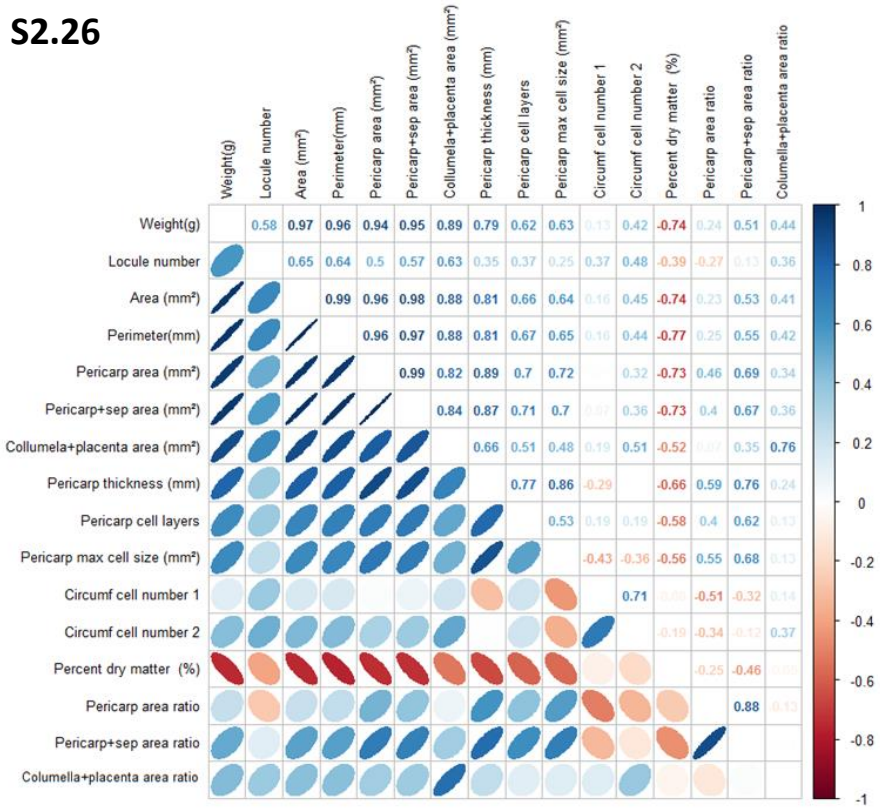


Supplementary figure S2. 17 – 2.24. Weight and weight parameters distributions grouped based on CNR, SlKLUH, CSR, LC, and FAS genotype (in that order). Groups are arranged in order of increasing derived alleles (3 is the wild allele and 1 the derived allele). Groups with only 1 accession or heterozygous for any of the five major genes were not included.

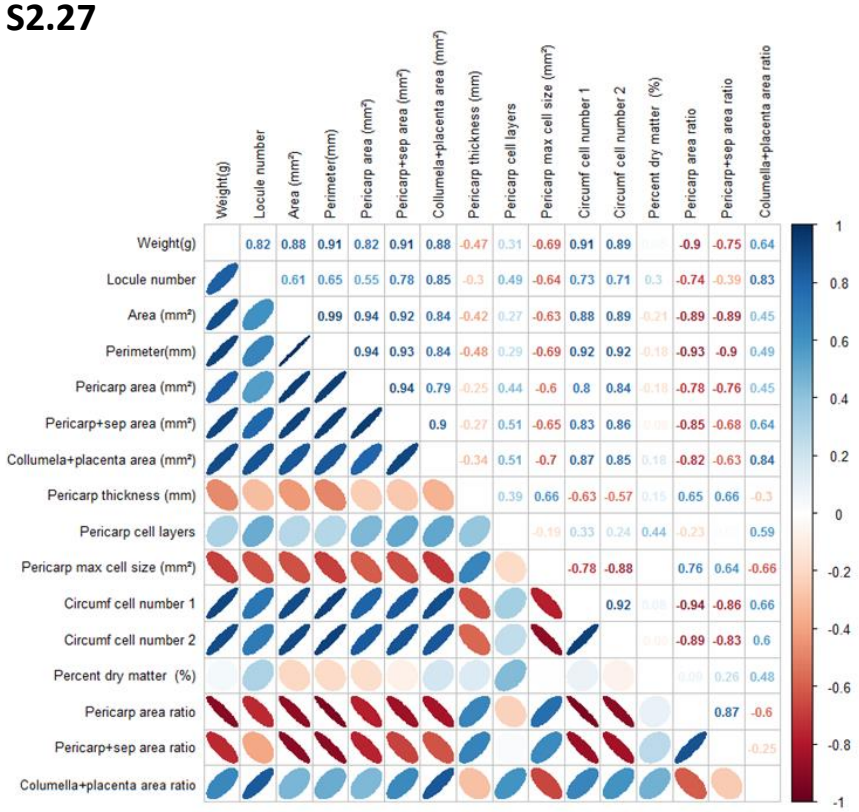
S2.25


Weight(g)
Locule number
Area (mm²)
Perimeter(mm)
Pericarp area (mm²)
Pericarp+sep area (mm²)
Collumela+placenta area (mm²)
Pericarp thickness (mm)
Pericarp cell layers
Pericarp max cell size (mm²)
Circumf cell number 1
Circumf cell number 2
Percent dry matter (%)
Pericarp area ratio
Pericarp+sep area ratio
Columella+placenta area ratio
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

S2.26

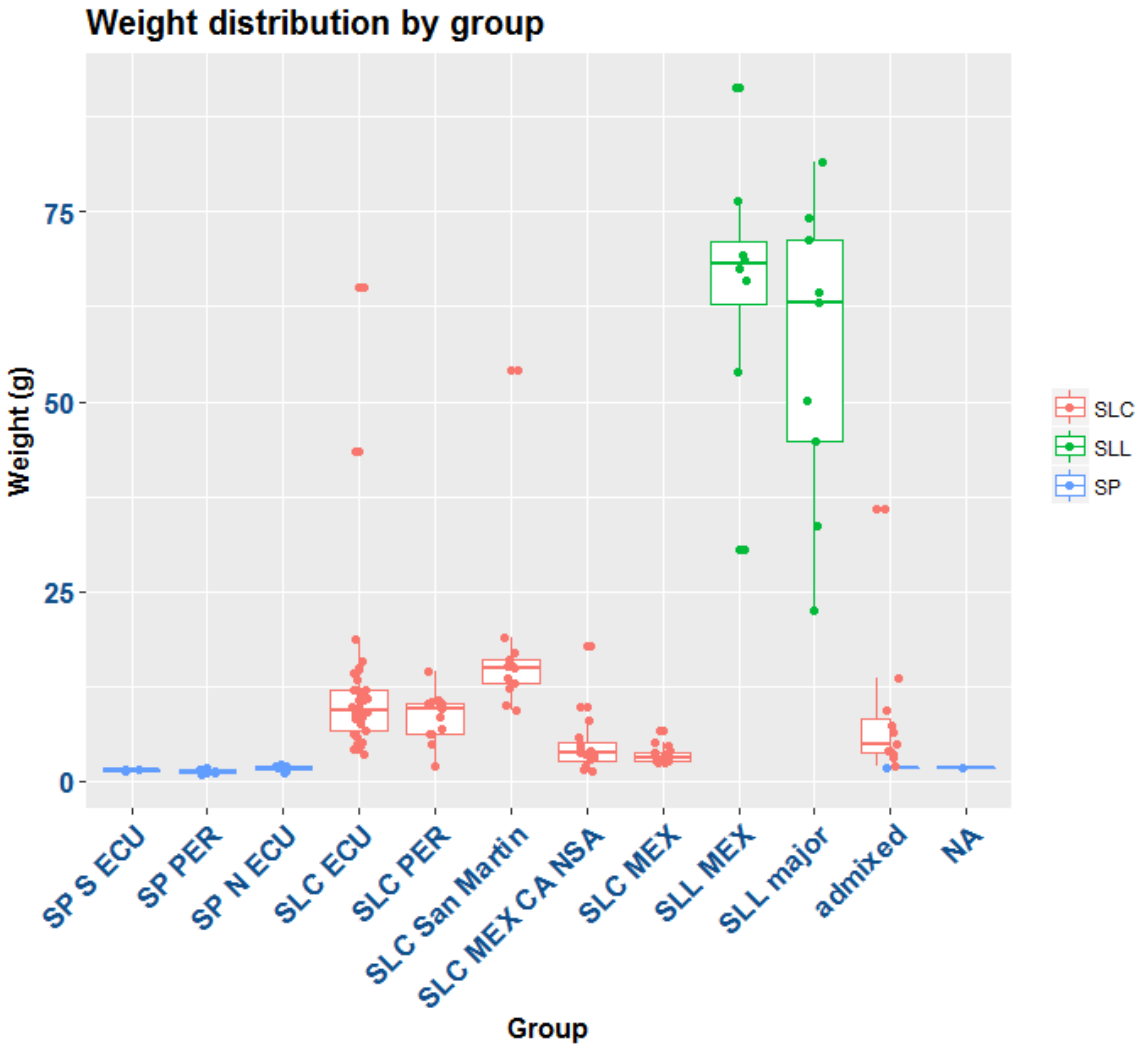

Weight(g)
Locule number
Area (mm²)
Perimeter(mm)
Pericarp area (mm²)
Pericarp+sep area (mm²)
Collumela+placenta area (mm²)
Pericarp thickness (mm)
Pericarp cell layers
Pericarp max cell size (mm²)
Circumf cell number 1
Circumf cell number 2
Percent dry matter (%)
Pericarp area ratio
Pericarp+sep area ratio
Columella+placenta area ratio
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

S2.27

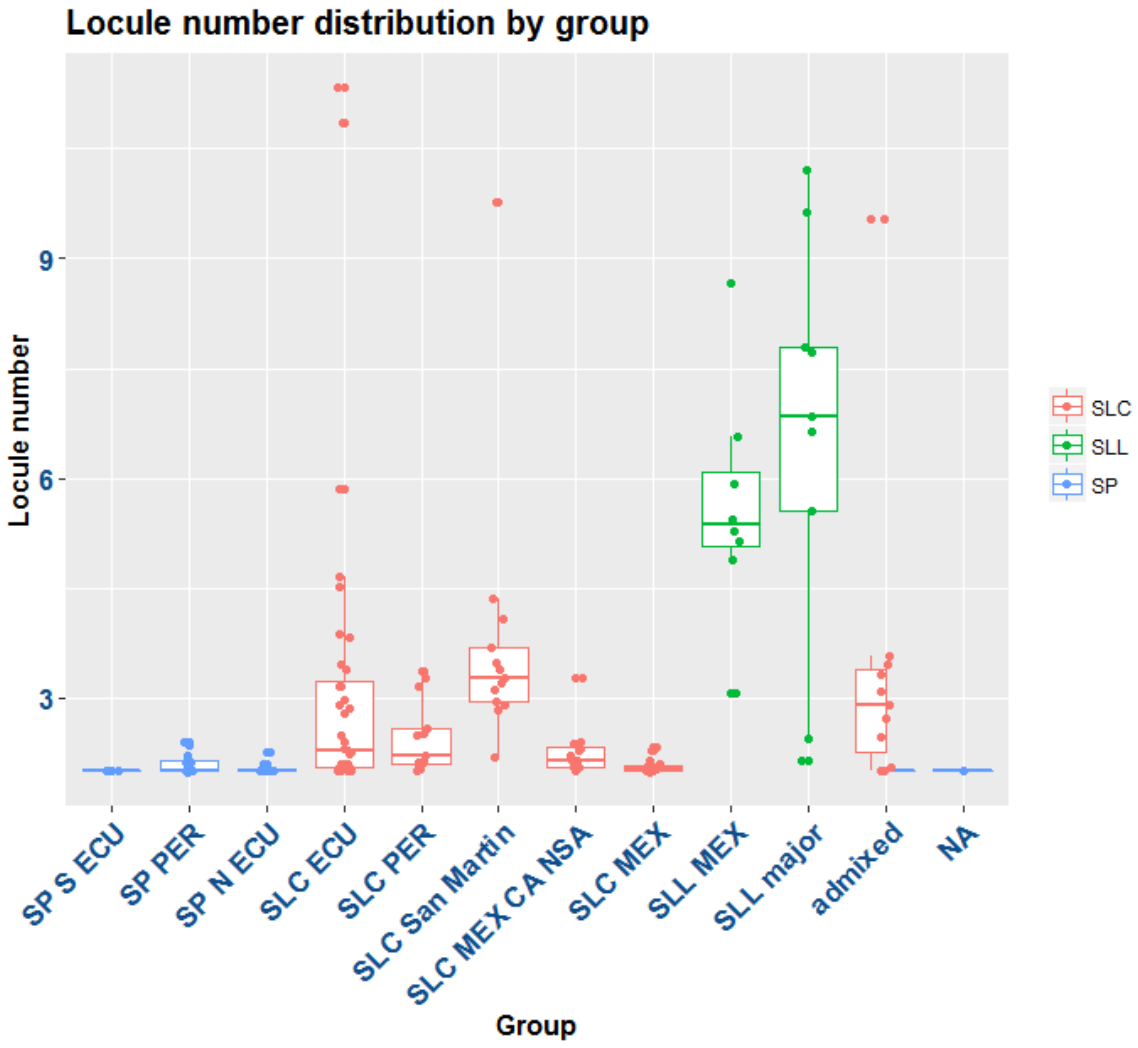


Supplementary figure S2. 25 – 2.27. Correlations between weight and weight parameters for SP (S2.25), SLC (S2.26), and SLL (S2.27). Color indicates the strength of the correlation where dark blue is a positive correlation of 1, and dark red is a negative correlation of 1. Coefficients of correlation (r) values are also shown for each combination.

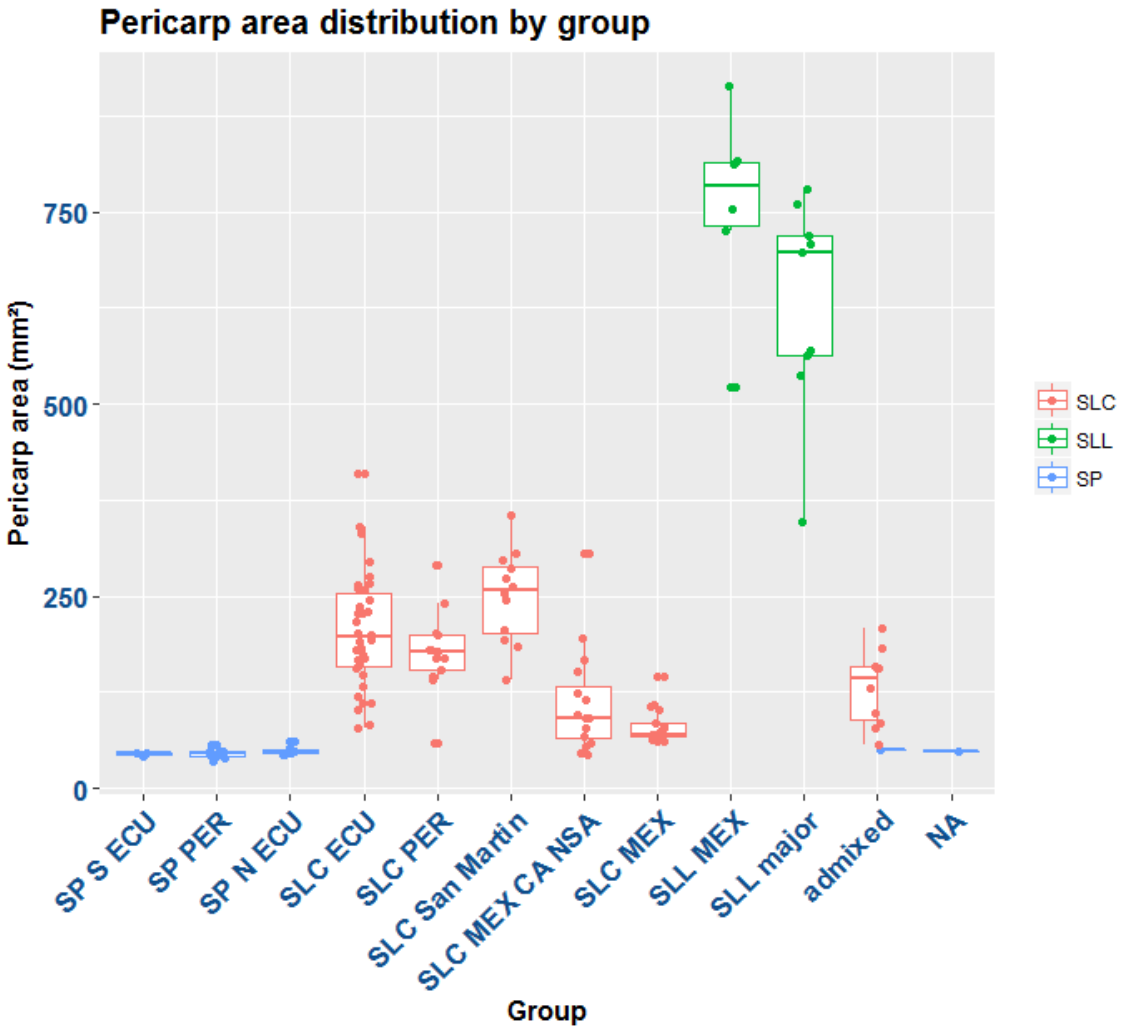
S2.28


Weight distribution by group
Weight (g)
75
50
25
0
SP S ECU
SP PER
SP N ECU
SLC ECU
SLC PER
SLC San Martin
SLC MEX C A NSA
SLC MEX
SLL MEX
SLL major
admixed
NA
Group
SLC
SLL
SP

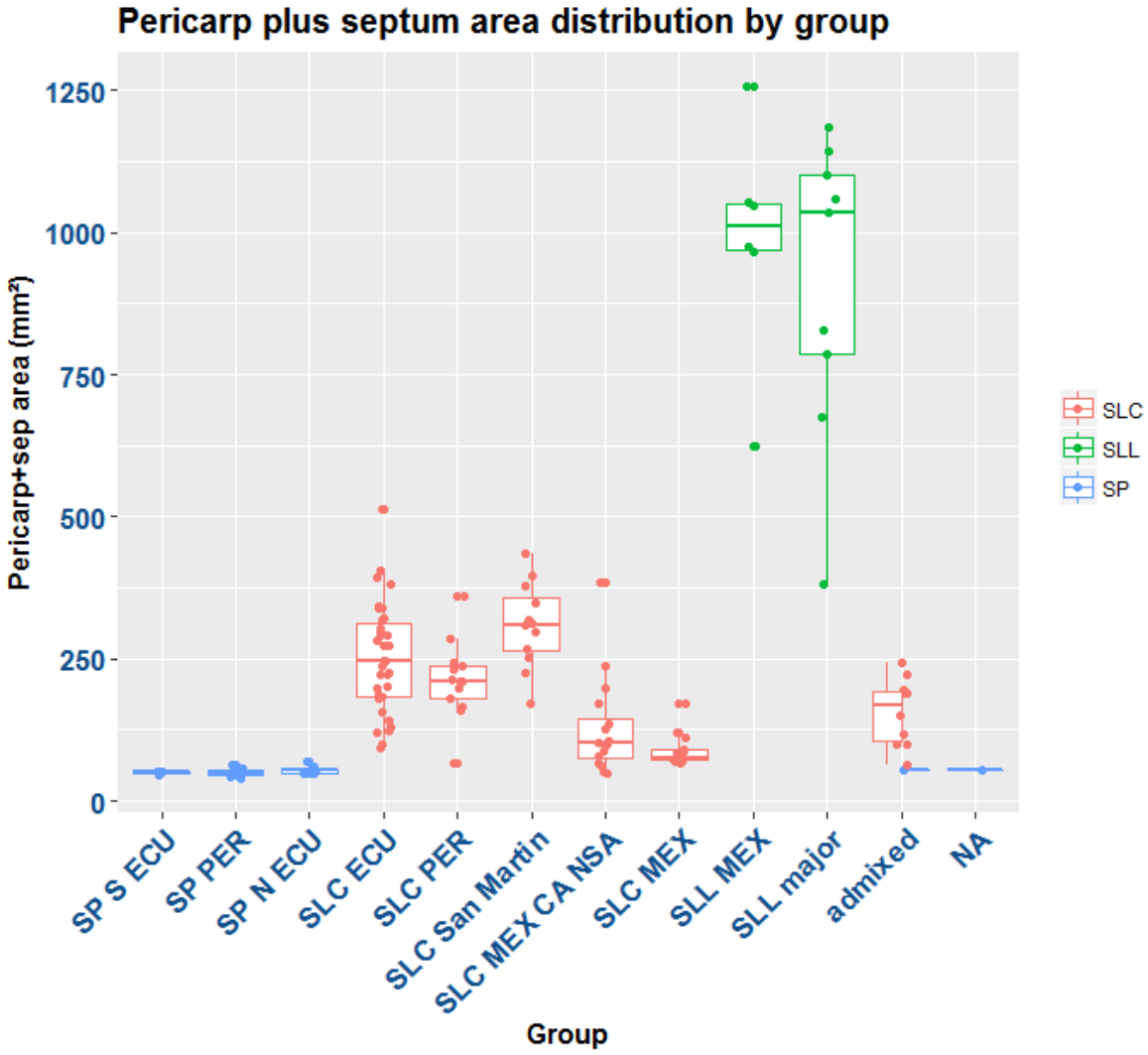
S2.29


Locule number distribution by group
Locule number
9
6
3
SP S ECU
SP PER
SP N ECU
SLC ECU
SLC PER
SLC San Martin
SLC MEX CA NSA
SLC MEX
SLL MEX
SLL major
admixed
NA
Group
SLC
SLL
SP

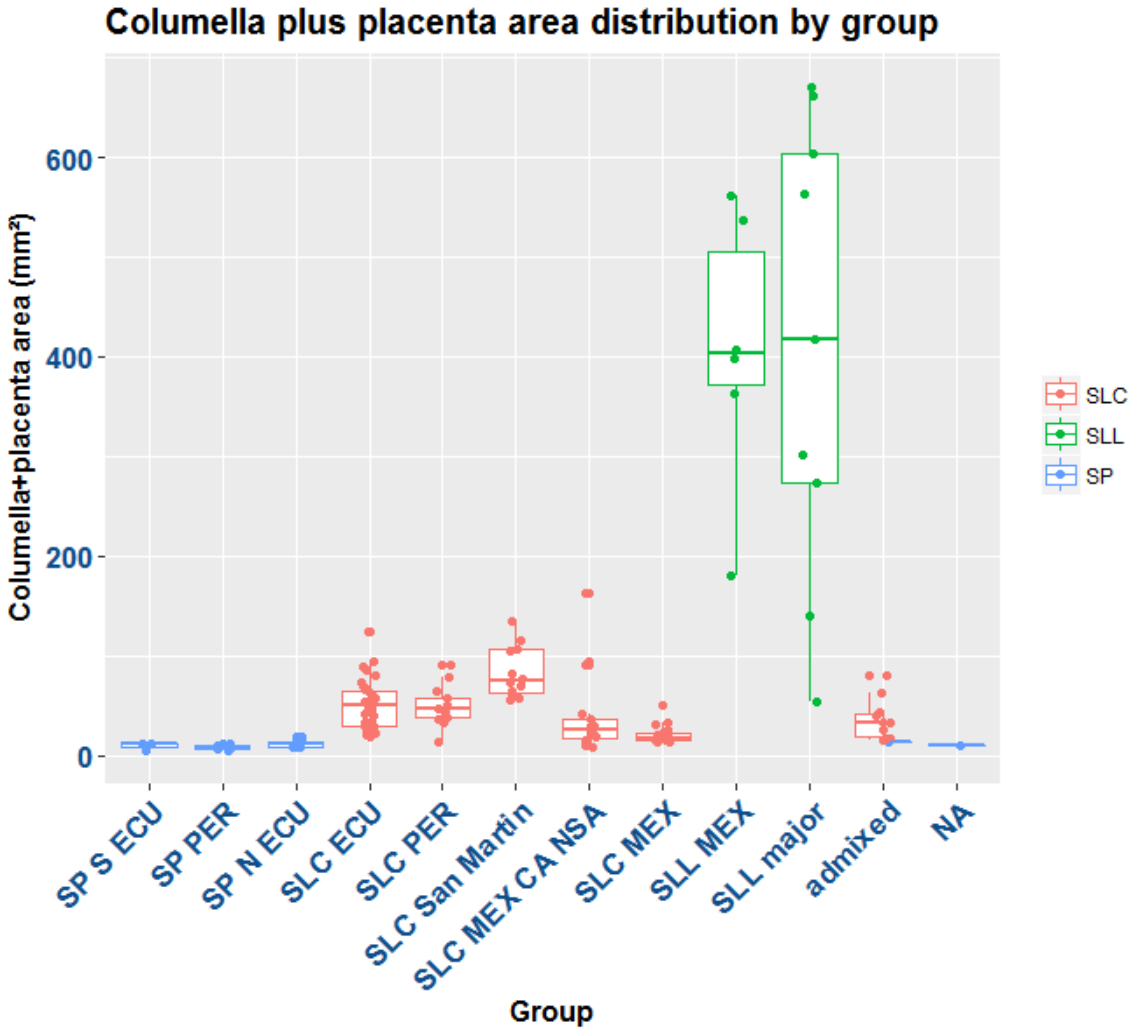
S2.30


Pericarp area distribution by group
Pericarp area (mm²)
750
500
250
0
SP S ECU
SP PER
SP N ECU
SLC ECU
SLC PER
SLC San Martin
SLC MEX CA NSA
SLC MEX
SLL MEX
SLL major
admixed
NA
Group
SLC
SLL
SP

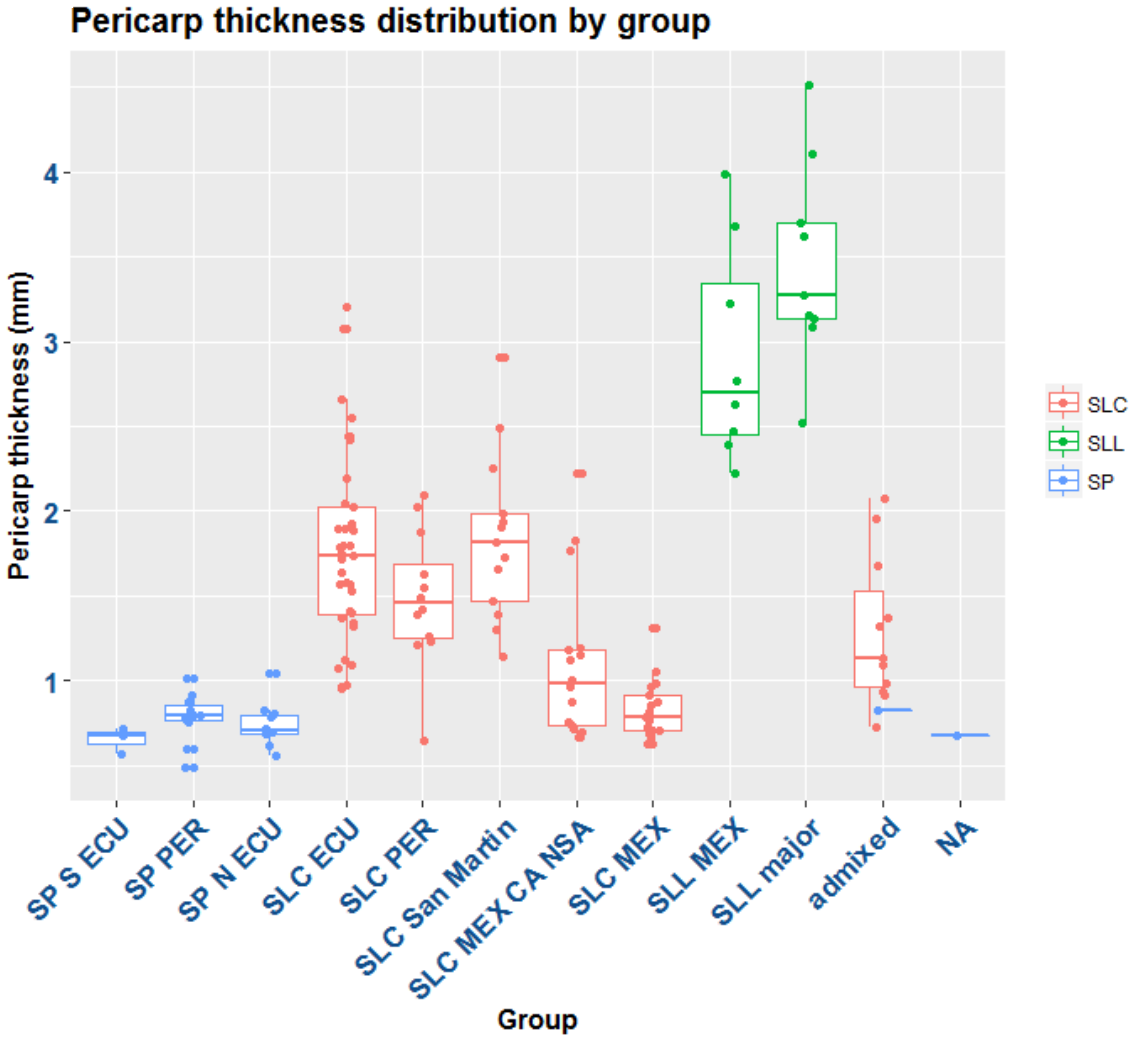
S2.31


Pericarp plus septum area distribution by group
Pericarp+sep area (mm²)
1250
1000
750
500
250
0
SP S ECU
SP PER
SP N ECU
SLC ECU
SLC PER
SLC San Martin
SLC MEX CA NSA
SLC MEX
SLL MEX
SLL major
admixed
NA
Group
SLC
SLL
SP

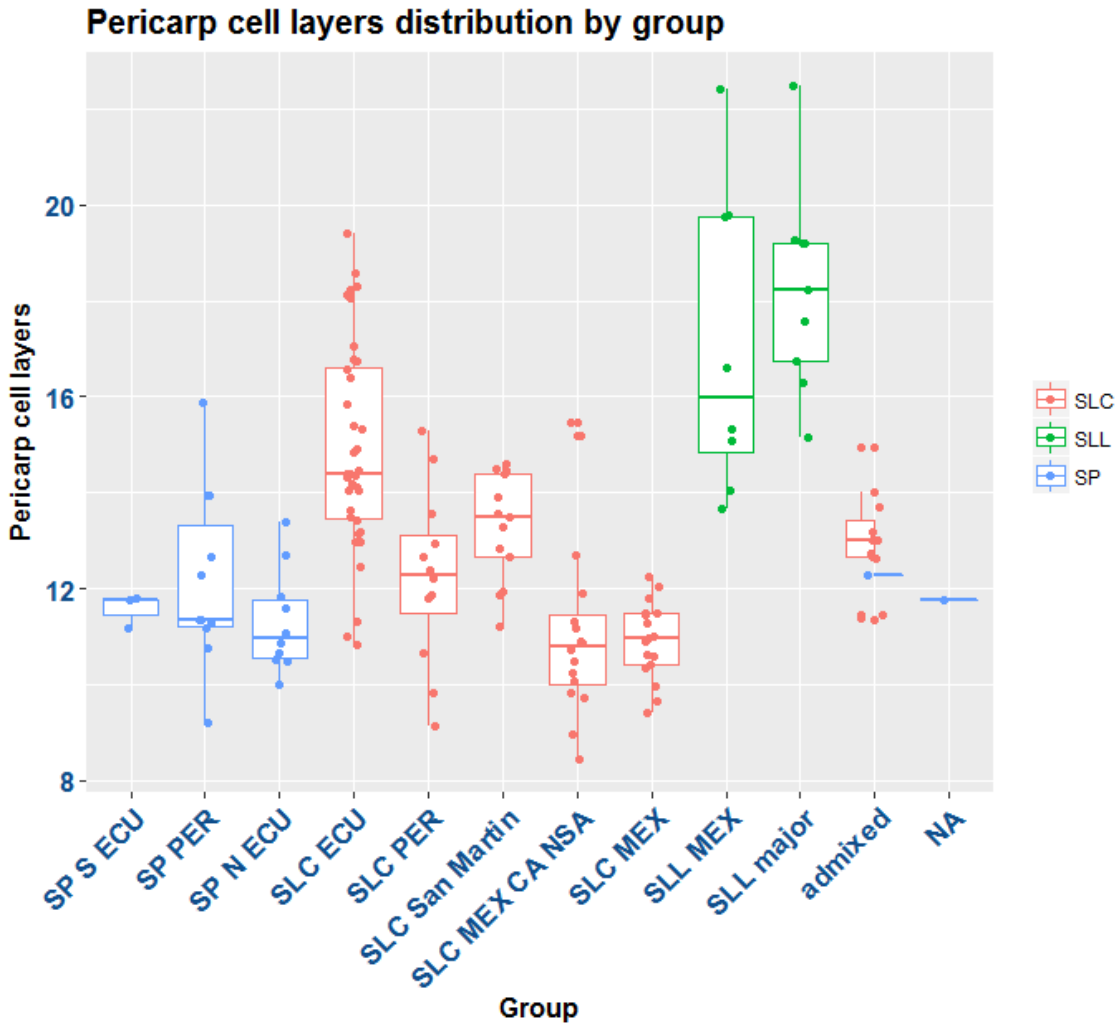
S2.32


Columella plus placenta area distribution by group
Columella+placenta area (mm²)
600
400
200
0
SP S ECU
SP PER
SP N ECU
SLC ECU
SLC PER
SLC San Martin
SLC MEX CA NSA
SLC MEX
SLL MEX
SLL major
admixed
NA
Group
SLC
SLL
SP

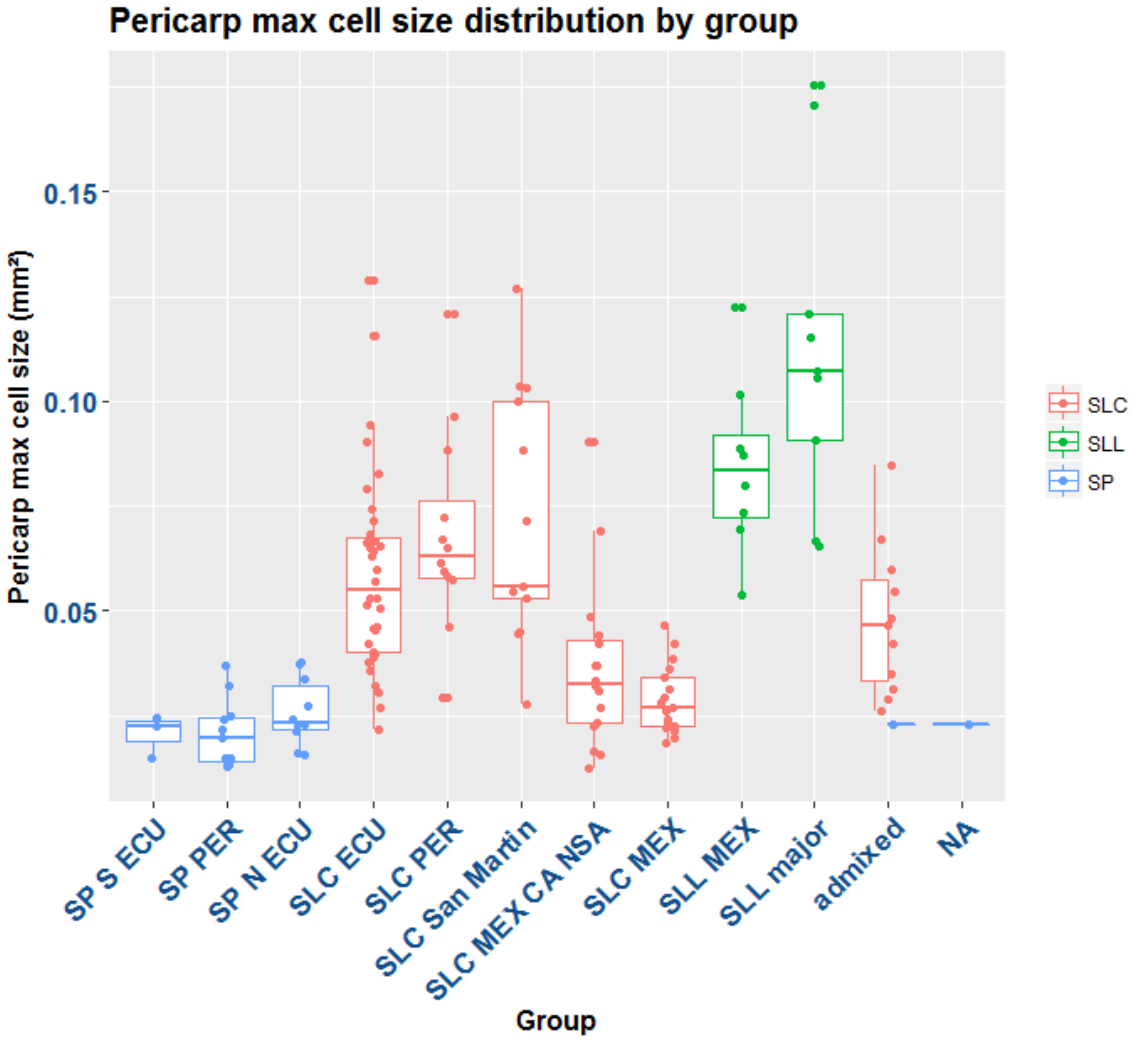
S2.33


Pericarp thickness distribution by group
Pericarp thickness (mm)
4
3
2
1
SP S ECU
SP PER
SP N ECU
SLC ECU
SLC PER
SLC San Martin
SLC MEX CA NSA
SLC MEX
SLL MEX
SLL major
admixed
NA
Group
SLC
SLL
SP

S2.34


Pericarp cell layers distribution by group
Pericarp cell layers
20
16
12
8
SP S ECU
SP PER
SP N ECU
SLC ECU
SLC PER
SLC San Martin
SLC MEX CA NSA
SLC MEX
SLL MEX
SLL major
admixed
NA
Group
SLC
SLL
SP

S2.35



Supplementary figure S2. 28 – 2.35. Weight and weight parameters distributions for SP, SLC, and SLL by phylogenetic group. NA represents accessions with no data available.

CHAPTER 3

IDENTIFICATION OF TOMATO FRUIT WEIGHT QTL BY BULK SEGREGANT ANALYSIS-QTLseq [1]

[1] Ramos, A.,Clevenger, J.P., and van der Knaap, E. To be submitted to *Euphytica*.

Abstract

Tomato fruit weight is one of the most important traits in tomato breeding, and it greatly determines the market niche. Several major genes that control fruit weight have been identified. However, a need for additional genes is always warranted as profit gains tend to be associated with larger fruit sizes. With the new cost-effective sequencing technologies, quantitative trait locus (QTL)-seq has become a widely used tool for QTL mapping. In this study, the QTL-seq approach was performed on extreme bulks for two reciprocal F_2 panels (N=96 and N=144) that were fixed for the five major fruit weight genes CNR (FW2.2), SlKLUH (FW3.2), CSR (FW11.3), FAS (CLV3), and LC (WUSCHEL). We identified a candidate QTL for fruit weight at the bottom of chromosome 2, and a pericarp circumference cell number QTL at the bottom of chromosome 6. The QTL on chromosome 2 was close to the known fruit weight gene CNR. This means that the newly identified QTL could be an allele of the known gene or represent a tightly linked novel weight locus. Association of the QTL with the traits was confirmed in both F_2 panels using molecular markers. Our results detected the fruit weight QTL in both reciprocal panels, suggesting reproducibility. Our results also showed that QTL-seq is dependable even with a relatively small panel size (N=96). This study marks the initial steps towards fine mapping of a potentially novel weight related gene in tomato that can be used in future breeding efforts.

Introduction

Tomato (*Solanum lycopersicum*) is a fruit of great economic importance with around 177 tonnes produced worldwide in 2016 (FAOSTAT, 2016). Tomato is consumed both raw (fresh market) and processed (processing market) where fruit shape

and size are important in the market-specific breeding programs. While cultivated tomato is phenotypically of variable shapes and sizes, one of the striving goals for many breeders is to create cultivars that produce even larger fruits. This is what consumers desire, yet it cannot be achieved with the currently available genetic tools. Tomato fruit size is a quantitative trait, and many candidate QTL have been mapped (Grandillo et al., 1999). Only a few major genes controlling tomato fruit shape and size have been cloned, and they account for a large amount of the morphological variation (van der Knaap et al., 2014). FASCIATED (FAS, ortholog of CLV3) and LOCULE NUMBER (LC, ortholog of WUSCHEL) control the number of locules in the fruit (Muños et al., 2011; Xu et al., 2015) whereas Cell Number Regulator (CNR or FW2.2), FW3.2 (KLUH), and FW11.3 (Cell Size Regulator, CSR) control fruit weight by controlling either cell size or number in the pericarp and septum or columella of the fruit (Chakrabarti et al., 2013; Frary et al., 2000; Mu et al., 2017) . Novel sources of variation are needed to find new fruit shape and size genes that can then be incorporated into the modern varieties.

Historically, linkage mapping has been employed to identify QTL for various traits in different crops. Also, most mapping efforts have been performed using natural diversity, and much of it was done using populations developed from crosses between cultivated varieties and distant wild relatives. Introgression lines (Prudent et al., 2009), recombinant inbred lines (Capel et al., 2015; Causse et al., 2002), and backcross populations (Bernacchi et al., 1998) have all led to the identification of tomato QTL for weight related traits. Various QTL for a wide range of agronomically important traits such as disease resistance and flower characteristics have been mapped (Foolad, 2007). Also, mutagenesis with ethyl methanesulfonate (EMS) (Musseau et al., 2017) and other

technologies as CRISPR/Cas (Rodríguez-Leal, Lemmon, Man, Bartlett, & Lippman, 2017) have been employed to generate additional genetic and phenotypic diversity in tomato, which can be exploited to identify new QTL. Traditional mapping projects need many resources and often take several years to produce results. New emerging sequencing technologies, however, provide a great opportunity to use different techniques that can expedite the mapping efforts. The Bulk Segregant Analysis-QTLseq (BSA-QTLseq) approach has proven to be a reliable and cost-effective tool for QTL mapping for quantitative traits as flowering time and fruit weight in tomato (Illa-Berenguer, Van Houten, Huang, & van der Knaap, 2015; Ruangrak et al., 2018). This technique is also time efficient, as the mapping is performed on a segregating F_2 population.

Tomato originated in South America and was domesticated in a two-step process where the first domestication took place in Ecuador and northern Peru and a second domestication took place in Mexico (Blanca et al., 2015). The bottlenecks in tomato domestication provide a great opportunity to map new genes that control fruit size using natural tomato diversity since there are high chances that beneficial genes were left behind in South America and did not make it to the cultivated germplasm. In this study, we created two F_2 mapping panels that would partially capture the domestication of tomato from Ecuador to Mexico. Using BSA-QTLseq approach combined with the newly developed R package QTLseqr for mapping QTL (Mansfeld et al., 2017), we mapped a new putative locus controlling tomato fruit weight and pericarp circumference cell number on chromosome 2 and 6, respectively.

Materials and methods

Plant materials

In depth phenotyping and genotyping of a panel of 167 accessions (assembled by Joaquin Canizares, Maria Jose Diez, and Jose Blanca, Institute for Conservation & Improvement of Valentian Agrodiversity (COMAV), Valencia, Spain) showed phenotypic variance not explained by the known fruit weight genes (data not shown). In fall 2016, *S. lycopersicum* var. *cerasiforme* (SLC) accessions BGV006768 and BGV007931 were reciprocally crossed to generate F_1s. Both accessions were fixed for the wild allele for the five major fruit weight genes CNR, SlKLUH, CSR, LC, and FAS; yet the fruits average weight was 14 g and 1.6 g, respectively. This difference in fruit weight was consistent across three environments (Life Oak, FL, Watkinsville, GA, and Blairsville, GA) and two summer field seasons (2016 and 2017) (data not shown). One F_1 seedling per cross was used to generate an F_2 panel. The F_1 plants were grown and genotyped for the five major genes for confirmation purposes (data not shown). Also, a polymorphic marker available from other projects was used to confirm that the F_1 were true crosses and no selfing had occurred (Supplementary table S3.1). Panel 17S62 was created from selfing F_1 plants from the parental cross BGV006768 x BGV007931, whereas panel 17S64 was created from the reciprocal cross.

Experimental design

Both F_2 mapping panels 17S62 (N=144) and 17S64 (N=96) where grown in Blairsville, GA in summer 2017. 17S62 was grown in an open field, while 17S64 was grown in raised beds. The seeds were planted in the numerical order that was created

when they were sown. F_1 (N=5) and parental (N=10) checks were included in the field for panel 17S62. No checks were included with panel 17S64 because limited field space.

Phenotyping

Panel 17S62 was phenotyped for total fruit weight, while panel 17S64 was phenotyped for total fruit weight and many additional fruit weight attributes: area, perimeter, pericarp area, pericarp plus septum area, and columella plus placenta area in the medio-lateral axis at the equatorial plane. In addition, ratios of the tissue-specific areas to total area, pericarp thickness, pericarp cell components (maximum cell size and cell layers), and circumference cell number 1 and 2 were measured (Supplementary tables S3.2 and S3.3). Phenotyping methodology is summarized in Supplementary table S3.4. The correlation matrix between the different phenotypes in panel 17S64 were computed using the Corrplot package in R (Wei & Simko, 2016) and the correlation network was plotted using the qgraph package in R (Epskamp et al., 2012). Partial correlations which control for the effect of the other variables were calculated using the ppcor package in R (Kim, 2015) (data not shown). The fruit weight data for both panels were tested with Levene's test (from the car package in R) for equality of variances (p-value= 0.02) which suggested unequal variances (Fox & Weisberg, 2011). Therefore, the panels were kept separate for the downstream analysis.

DNA extraction, library preparation, and sequencing

DNA was extracted from young leaves at the seedling stage using the Qiagen 96-well DNA extraction kit. The DNA for the 10 plants with highest and lowest phenotypes for each trait of interest (17S62 weight, 17S64 weight, 17S64 Circumference cell number 1) was quantified using a Qubit 2.0 Fluorometer and bulked, respectively so each sample

contributed equally to the final bulk (6 bulks total). DNA libraries were created using the NEBNext Ultra II DNA library prep kit (E7645L) (www.neb.com) (performed by Dr. van der Knaap). The bulks were sequenced on the Illumina NextSeq PE 150 High Output platform at the Georgia Genomics Facility at UGA. The full genome sequencing produced approximately 20X coverage per bulk.

Sequence processing and QTLseq

The raw data obtained from the sequencing facility was first explored with FastQC to detect any problems (Andrews, 2010). No filtering was performed at this point and no major issues where detected. Coverage was approximated by using the formula C=LN/G where C is coverage, G is the haploid genome length of tomato (~950 Mb), L is the read length (~150 bp average), and N is the number of reads. The bulk sequences were processed and mapped to the tomato reference genome (SL3.0) using Burrows-Wheeler Aligner-MEM (Li, 2013), SAMtools (Li et al., 2009), Picard Tools (http://broadinstitute.github.io/picard/) , and Genome Analysis Toolkit (GATK) (McKenna et al., 2010) with the recommended default settings. SNP variants were called using GATK Haplotype caller and filtered with the recommended default settings. No missing data were allowed in the final vcf file, and only SNP with QUAL > 30 were kept for downstream analysis. The final vcf file was formatted as a table using the VariantsToTable tool from GATK for the downstream analysis. The BSA-QTLseq analysis was conducted using the QTLseqr R package (Mansfeld & Grumet, 2017). The vcf files were further filtered with this package so the read depth at each SNP in each bulk was at least 10, but no more than 80 when considering both bulks together. Window size was set to 1Mb and 10,000 simulations were run creating simulated deltaSNP

indexes based on the data parameters. The simulations extreme quantiles served as confidence intervals for the ΔSNP. The absolute ΔSNP values were used for plotting the BSA-QTLseq output graphs.

QTL analysis

Kompetitive Allele Specific PCR (KASP) markers (http://www.lgcgroup.com) were created to genotype the entire panels and to validate the QTL detected with QTLseqr (Supplementary table S3.5) (Mansfeld & Grumet, 2017). ANOVA tests were used to associate the QTL with the traits, and the p values were corrected for multiple comparisons using a Bonferroni correction implemented with R (R Core Team 2016). The software R/qtl was employed to estimate the genetic distances, and to calculate the LOD scores for the genotyped markers (Broman, Wu, Sen, & Churchill, 2003). Conditional genetic probabilities were calculated using an error probability of 0.001, step = 0, and map.function = "morgan". Since the data were not normally distributed, a nonparametric interval mapping was employed for calculating LOD scores. Also, 10,000 permutation tests were performed to determine the LOD significance thresholds. A multiple QTL mapping strategy (1,000 imputations) that uses multiple interval mapping (MIM) was implemented using R/qtl. Interaction between QTL was examined using the R/qtl function "scantwo". Based on both analysis of variance, interval mapping, and MIM, statistical models including the significant QTL and interactions where applicable were created for each trait to estimate the amount of phenotypic variance explained by each QTL.

Results

Phenotype analysis and selection for QTLseq

The different weight related phenotypes were carefully investigated to identify good candidate traits for the BSA-QTLseq approach and to understand the major drivers of fruit weight. Only fruit weight was measured for panel 17S62, so this trait was selected for BSA-QTLseq in that panel. The fruits of the parental accessions BGV007931 and BGV006768 represent the extremes of the weight distribution and the fruit weight of the F_1 plants are much closer to the smaller fruited parent (Figure 3.1). The trend of weight segregating towards the smaller parent has been observed previously (Grandillo & Tanksley, 1996; Illa-Berenguer et al., 2015; Lippman & Tanksley, 2001). Even though data from only one BGV006768 parental check was included, it is reliable as similar fruit weights were reported for this accession in previous years and locations (data not shown).

The 16 fruit weight-related attributes that were evaluated for panel 17S64 were examined before performing BSA-QTLseq (Figure 3.1 and Supplementary figure S3.1). Most attributes were slightly skewed to the right, and while no checks were grown with these F_2 plants, their phenotypic distribution was between that of the parents when compared to the values from previous years (data not shown). Fruit weight was selected for BSA-QTLseq since that was the major focus of the study. To determine which other traits to use for Bulk Segregant Analysis-QTLseq, a combined approach was employed. First the F_2 plants with extreme phenotypes for each trait were analyzed revealing a large overlap between plants with extreme weight and plants with extreme values for the most weight-related traits (Supplementary figure S3.2). In other words, large fruits had high values for the fruit weight-related attributes in most cases. Only traits pertaining to

pericarp thickness and the cellular traits did not overlap extensively with fruit weight. The second criterion used was the strength of the relationship between weight and the other traits using correlation coefficients and network analyses (Supplementary figure S3.3). When partial correlations were conducted (data not shown) the correlation coefficients among the phenotypes decreased to nearly zero for most correlations suggesting that most of our strong correlations were driven by fruit weight, where a larger fruit inherently has larger fruit weight attributes than a smaller fruit. Most non-derived traits were strongly or partially positively correlated to fruit weight ($r \geq 0.79$), which was expected. Contrary to most fruit weight-related traits, the two circumference cell number traits were not correlated to fruit weight ($r \leq 0.05$, "circumference cell number 1" p-value=0.83 and "circumference cell number 2" p-value=0.65). The final criterion was the aspect of weight that was captured by the trait. Only the circumference cell number estimates provided a cellular approach to fruit size in the medio-lateral dimension. Circumference cell number 1 was selected for BSA-QTLseq because plants with extreme values did not overlap with plants with extreme fruit weight, the correlation between the trait and fruit weight was negligible, and it offered a cellular approach to fruit size in the medio lateral dimension. Therefore, we sought to map this trait in addition to total fruit weight using the BSA-QTLseq approach.

QTLseq results

The BSA-QTLseq approach revealed a total fruit weight QTL in the same chromosomal position in both panels, and a circumference cell number QTL in panel 17S64. There were other genomic regions that appeared to harbor a QTL, however when genotyped in the entire panel, they were not associated with the respective traits

(Supplementary table S3.6). We also found that markers associated with fruit weight were associated with most of the fruit weight-related traits in 17S64, except cell circumference 1 and 2. Conversely the markers linked to cell circumference 1 and 2 were not linked with most other traits. However, these two traits were linked to pericarp thickness and cell size since these traits were used to estimate the number of cell in the circumference of the fruit.

For fruit weight, the most significant QTL were detected at the bottom of chromosome 2 in both panels (Figure 3.2 A and B). They were highly associated to markers 18EP95 and 18EP107 at positions 53,385,669 bp and 55,261,865 bp at p-value<1.84E-9 and p-value<2.45E-10, respectively (Table 3.1). Association was high in both panels even though panel 17S62 performed poorly in the field and had lower than expected fruit yield per plant and overall plant size. No QTL interaction was detected for the weight QTLseq on 17S62 or the cell circumference number QTLseq on 17S64. However, a minor interaction was detected between the QTL in chromosomes 2 and 9 for the fruit weight QTLseq in 17S64. In panel "17S62 weight" the QTL in chromosome 2 explained 27.8% of the phenotypic variance with LOD=7.36. On the other hand, in panel "17S64 weight" a model including both QTL in chromosome 2 and 9, as well as their interaction explained 56.2% of the phenotypic variance with the highest hit in chromosome 2 with LOD=7.07, supporting linkage to the trait. Even though the LOD scores were not extremely high, in these panels scores above 2.67 and 3.03 respectively indicated a significant linkage between the marker and the trait.

The relatively high QTL peaks detected on chromosomes 3,8,9, and 11 for the weight QTLseq in 17S62 and 17S64 were not statistically associated to the traits (Table

3.1 and Supplementary figures S3.4-S3.6). The peaks observed for fruit weight at the bottom of chromosome 3 for both panels (Supplementary figures S3.4-S3.5) may overlap with *fw3.3*, a previously mapped fruit weight QTL (Illa-Berenguer et al., 2015).

The only QTL detected for cell circumference number in the17S64 panel was at the bottom of chromosome 6 (Figure 3.2 C). Even though there were additional loci that showed linkage close to the significance threshold on other chromosomes (Table 3.1 and Supplementary figure S3.6), they were not linked to the trait when mapped in the entire panel. The cell number circumference QTL on chromosome 6 showed an LOD score of 5.0 indicating significant linkage. However, it appears to be a small-effect QTL compared to the total fruit weight QTL on chromosome 2.

Discussion

Our BSA-QTLseq approach proved to be a reliable alternative to whole genome linkage mapping as it led to the discovery of two new QTL: a fruit weight QTL on chromosome 2 and a cell number in the circumference QTL on chromosome 6. It also showed that BSA-QTLseq can detect the same QTL in a relatively small panel of 96 F_2 plants. In future studies, factors as panel size, bulk size, and read depth should be further considered since they directly affect the quality of genotypic data generated and hence the BSA-QTLseq results. Determining these factors, however, is not an easy task and a wide range of parameters have been successfully employed in different BSA-QTLseq studies in tomato and other crops. Coverages as low as ~6X and up to ~80X on either F_2 or RIL populations with 262 to 531 plants, and bulk sizes ranging from 10 to 50, have all led to QTL discovery (Illa-Berenguer et al., 2015; Ruangrak et al., 2018; Singh et al., 2016; Takagi et al., 2013; Wei et al., 2016).

The QTL in chromosome 2 overlapped with the CNR locus, however both reported causal mutations for this gene were fixed in the parents of these panels (Blanca et al., 2015; Mazzucato et al., 2014). This indicated that we either found a new CNR allele, or a novel tightly linked weight locus. Fine mapping of the gene will allow to determine which was the case. The presence of small effect fruit weight QTL in this region, however, has been previously reported (Grandillo et al., 1999; Lecomte et al., 2004; Mazzucato et al., 2014). Nevertheless, our parents were both *S. lycopersicum* var. *cerasiforme* (SLC) and more closely related to each other than those in previous mapping studies, and likely capture the crucial mutations that led to *Solanum lycopersicum* var. *lycopersicum* (SLL) evolution from SLC. In future works, we will confirm the locus in the next generation and try to identify the underlying gene. Once cloned, we can determine whether it was a beneficial allele that was left behind during domestication. Nevertheless, it could also be that the allele of the underlying gene is fully fixed in modern tomato.

The identification of the fruit weight QTL on chromosome 2 in both reciprocal panels is reassuring of the validity of the locus. It also indicates that field effects as well as parental crossing direction in these panels did not greatly interact with the effect of the QTL. Furthermore, the minor weight QTL that were below significance in 17S62 were also observed in 17S64. Circumference cell number QTL was unlinked to fruit weight in 17S64 ($R^2 = 0.0005$) validating our approach for selecting these traits for BSA-QTLseq. Also, the markers associated with fruit weight were not associated with circumference cell number and vice versa, further supporting that these traits are controlled by different mechanisms. Follow up on this QTL will help to better understand the mechanisms

controlling cell number in the medio lateral axis of the fruit. A novel gene controlling this trait could have a multiplicative effect in overall fruit size, providing a new powerful tool to breeders.

Literature cited

Acquaah, George. (2009) Principles of plant genetics and breeding. John Wiley & Sons. 667-676.

Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at:http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Bai, Y., & Lindhout, P. (2007). Domestication and Breeding of Tomatoes: What have We Gained and What Can We Gain in the Future? *Annals of Botany*, *100*(5), 1085–1094. https://doi.org/10.1093/aob/mcm150

Barrero, L. S., & Tanksley, S. D. (2004). Evaluating the genetic basis of multiple-locule fruit in a broad cross section of tomato cultivars. *Theoretical and Applied Genetics*, *109*(3), 669–679. https://doi.org/10.1007/s00122-004-1676-y

Bernacchi, D., Beck-Bunn, T., Eshed, Y., Lopez, J., Petiard, V., Uhlig, J., … Tanksley, S. (1998). Advanced backcross QTL analysis in tomato. I. Identification of QTLs for traits of agronomic importance from Lycopersicon hirsutum. *Theoretical and Applied Genetics*, *97*(3), 381–397. https://doi.org/10.1007/s001220050908

Berry, S. Z., Gould, W. A., & Wiese, K. L. (1991). "Ohio 8245" processing tomato. *HortScience : A Publication of the American Society for Horticultural Science*. Retrieved from http://agris.fao.org/agris-search/search.do?recordID=US19950152001

Blanca, J., Montero-Pau, J., Sauvage, C., Bauchet, G., Illa, E., Díez, M. J., … Cañizares, J. (2015). Genomic variation in tomato, from wild ancestors to contemporary

breeding accessions. *BMC Genomics*, *16*, 257. https://doi.org/10.1186/s12864-015-1444-1

Boster, J. S. (1985). Selection for perceptual distinctiveness: Evidence from aguaruna cultivars of<Emphasis Type="Italic">Manihot esculenta</Emphasis>. *Economic Botany*, *39*(3), 310–325. https://doi.org/10.1007/BF02858802

Bourdon, M., Frangne, N., Mathieu-Rivet, E., Nafati, M., Cheniclet, C., Renaudin, J.-P., & Chevalier, C. (2010). Endoreduplication and Growth of Fleshy Fruits. In *Progress in Botany 71* (pp. 101–132). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02167-1_4

Brand, U., Fletcher, J. C., Hobe, M., Meyerowitz, E. M., & Simon, R. (2000). Dependence of Stem Cell Fate in Arabidopsis on a Feedback Loop Regulated by CLV3 Activity. *Science*, *289*(5479), 617–619. https://doi.org/10.1126/science.289.5479.617

Broman, K. W., Wu, H., Sen, Ś., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, *19*(7), 889–890. https://doi.org/10.1093/bioinformatics/btg112

Capel, C., Carmen, A. F. del, Alba, J. M., Lima-Silva, V., Hernández-Gras, F., Salinas, M., … Lozano, R. (2015). Wide-genome QTL mapping of fruit quality traits in a tomato RIL population derived from the wild-relative species <Emphasis Type="Italic">Solanum pimpinellifolium</Emphasis> L. *Theoretical and Applied Genetics*, *128*(10), 2019–2035. https://doi.org/10.1007/s00122-015-2563-4

Causse, M., Saliba-Colombani, V., Lecomte, L., Duffé, P., Rousselle, P., & Buret, M. (2002). QTL analysis of fruit quality in fresh market tomato: a few chromosome regions control the variation of sensory and instrumental traits. *Journal of Experimental Botany*, *53*(377), 2089–2098.

Chakrabarti, M., Zhang, N., Sauvage, C., Muños, S., Blanca, J., Cañizares, J., … van der Knaap, E. (2013). A cytochrome P450 regulates a domestication trait in cultivated tomato. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(42), 17125–17130. https://doi.org/10.1073/pnas.1307313110

Chernet, S., & Zibelo, H. (2014). Evaluation of Tomato Varieties for Fruit Yield and Yield Components in Western Lowland of Tigray, Northern Ethiopia - SciAlert Responsive Version. https://doi.org/10.3923/ijar.2014.259.264

Chou, J., Gupta, A., Yaduvanshi, S., Davidson, R., Nute, M., Mirarab, S., & Warnow, T. (2015). A comparative study of SVDquartets and other coalescent-based species tree estimation methods. *BMC Genomics*, *16*(10), S2. https://doi.org/10.1186/1471-2164-16-S10-S2

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). Qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*. Retrieved from https://dare.uva.nl/search?identifier=cb3d90cf-240a-4838-ab68-603c9ee0013a

Food and Agriculture Organization of the United Nations. (2016). FAOSTAT. Rome, Italy:FAO.

Foolad, M. R. (2007). Genome Mapping and Molecular Breeding of Tomato. *International Journal of Plant Genomics*, *2007*. https://doi.org/10.1155/2007/64358

Fox, J., & Weisberg, S. (2011). *An {R} Companion to Applied Regression* (Second). Thousand Oaks CA: Sage. Retrieved from http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Frary, A., & Doğanlar, S. (2003). Comparative Genetics of Crop Plant Domestication and Evolution. *TURKISH JOURNAL OF AGRICULTURE AND FORESTRY*, *27*(2), 59–69.

Frary, A., Nesbitt, T. C., Frary, A., Grandillo, S., Knaap, E. van der, Cong, B., … Tanksley, S. D. (2000). fw2.2: A Quantitative Trait Locus Key to the Evolution of Tomato Fruit Size. *Science*, *289*(5476), 85–88. https://doi.org/10.1126/science.289.5476.85

Grandillo, S., Ku, H. M., & Tanksley, S. D. (1999). Identifying the loci responsible for natural variation in fruit size and shape in tomato. *Theoretical and Applied Genetics*, *99*(6), 978–987. https://doi.org/10.1007/s001220051405

Grandillo, S., & Tanksley, S. D. (1996). QTL analysis of horticultural traits differentiating the cultivated tomato from the closely related species <Emphasis Type="Italic">Lycopersicon pimpinellifolium</Emphasis>. *Theoretical and Applied Genetics*, *92*(8), 935–951. https://doi.org/10.1007/BF00224033

Illa-Berenguer, E., Van Houten, J., Huang, Z., & van der Knaap, E. (2015). Rapid and reliable identification of tomato fruit weight and locule number loci by QTL-seq.

TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik, *128*(7), 1329–1342. https://doi.org/10.1007/s00122-015-2509-x

Jenkins, J. A. (1948). The Origin of the Cultivated Tomato. *Economic Botany*, *2*(4), 379–392.

Kim, S. (2015). ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *Communications for Statistical Applications and Methods*, *22*(6), 665–674. https://doi.org/10.5351/CSAM.2015.22.6.665

Konieczny, A., & Ausubel, F. M. (1993). A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. *The Plant Journal: For Cell and Molecular Biology*, *4*(2), 403–410.

Lecomte, L., Saliba-Colombani, V., Gautier, A., Gomez-Jimenez, M. C., Duffé, P., Buret, M., & Causse, M. (2004). Fine mapping of QTLs of chromosome 2 affecting the fruit architecture and composition of tomato. *Molecular Breeding*, *13*(1), 1–14. https://doi.org/10.1023/B:MOLB.0000012325.77844.0c

Legland, D., Arganda-Carreras, I., & Andrey, P. (2016). MorphoLibJ: integrated library and plugins for mathematical morphology with ImageJ. *Bioinformatics (Oxford, England)*, *32*(22), 3532–3534. https://doi.org/10.1093/bioinformatics/btw413

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv:1303.3997 [q-Bio]*. Retrieved from http://arxiv.org/abs/1303.3997

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format

and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, M.-X., Yeung, J. M. Y., Cherny, S. S., & Sham, P. C. (2012). Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Human Genetics*, *131*(5), 747–756. https://doi.org/10.1007/s00439-011-1118-2

Lin, T., Zhu, G., Zhang, J., Xu, X., Yu, Q., Zheng, Z., … Huang, S. (2014). Genomic analyses provide insights into the history of tomato breeding. *Nature Genetics*, *46*(11), 1220–1226. https://doi.org/10.1038/ng.3117

Lippman, Z., & Tanksley, S. D. (2001). Dissecting the genetic pathway to extreme fruit size in tomato using a cross between the small-fruited wild species Lycopersicon pimpinellifolium and L. esculentum var. Giant Heirloom. *Genetics*, *158*(1), 413–422.

Mansfeld, B., & Grumet, R. (2017). QTLseqr: An R package for bulk segregant analysis with next-generation sequencing. *BioRxiv*, 208140. https://doi.org/10.1101/208140

Mazzucato, A., Wang, A., Li, C., Zhang, C., Zamir, D., Zhu, G., … Zhang, Z. (2014). Genomic analyses provide insights into the history of tomato breeding. *Nature Genetics*, *46*(11), 1220. https://doi.org/10.1038/ng.3117

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Mu, Q., Huang, Z., Chakrabarti, M., Illa-Berenguer, E., Liu, X., Wang, Y., … Knaap, E. van der. (2017). Fruit weight is controlled by Cell Size Regulator encoding a novel protein that is expressed in maturing tomato fruits. *PLOS Genetics*, *13*(8), e1006930. https://doi.org/10.1371/journal.pgen.1006930

Muños, S., Ranc, N., Botton, E., Bérard, A., Rolland, S., Duffé, P., … Causse, M. (2011). Increase in tomato locule number is controlled by two single-nucleotide polymorphisms located near WUSCHEL. *Plant Physiology*, *156*(4), 2244–2254. https://doi.org/10.1104/pp.111.173997

Musseau, C., Just, D., Jorly, J., Gévaudant, F., Moing, A., Chevalier, C., … Fernandez, L. (2017). Identification of Two New Mechanisms That Regulate Fruit Growth by Cell Expansion in Tomato. *Frontiers in Plant Science*, *8*, 988. https://doi.org/10.3389/fpls.2017.00988

Peralta, I., Knaap, S., & Spooner, D. (2007). THE TAXONOMY OF TOMATOES: A REVISION OF WILD TOMATOES (SOLANUM L. SECTION LYCOPERSICON (MILL.) WETTST.) AND THEIR OUTGROUP RELATIVES (SOLANUM SECTIONS JUGLANDIFOLIUM (RYDB.) CHILD AND LYCOPERSICOIDES (CHILD) PERALTA), (84), 1–10.

Prudent, M., Causse, M., Génard, M., Tripodi, P., Grandillo, S., & Bertin, N. (2009). Genetic and physiological analysis of tomato fruit weight and composition: influence of carbon availability on QTL detection. *Journal of Experimental Botany*, *60*(3), 923–937. https://doi.org/10.1093/jxb/ern338

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics*, *197*(2), 573–589. https://doi.org/10.1534/genetics.114.164350

Ranc, N., Muños, S., Santoni, S., & Causse, M. (2008). A clarified position for solanum lycopersicum var. cerasiformein the evolutionary history of tomatoes (solanaceae). *BMC Plant Biology*, *8*, 130. https://doi.org/10.1186/1471-2229-8-130

Ranc, N., Muños, S., Xu, J., Paslier, M.-C. L., Chauveau, A., Bounon, R., … Causse, M. (2012). Genome-Wide Association Mapping in Tomato (Solanum lycopersicum) Is Possible Using Genome Admixture of Solanum lycopersicum var. cerasiforme. *G3: Genes|Genomes|Genetics*, *2*(8), 853–864. https://doi.org/10.1534/g3.112.002667

Rodríguez, G. R., Moyseenko, J. B., Robbins, M. D., Morejón, N. H., Francis, D. M., & van der Knaap, E. (2010). Tomato Analyzer: a useful software application to collect accurate and detailed morphological and colorimetric data from two-dimensional objects. *Journal of Visualized Experiments: JoVE*, (37). https://doi.org/10.3791/1856

Rodríguez-Leal, D., Lemmon, Z. H., Man, J., Bartlett, M. E., & Lippman, Z. B. (2017). Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing. *Cell*, *171*(2), 470-480.e8. https://doi.org/10.1016/j.cell.2017.08.030

Roth, I. (1977). *Fruits of Angiosperms*. Stuttgart, Germany: Schweizerbart Science Publishers. Retrieved from

http://www.schweizerbart.de//publications/detail/isbn/9783443140106/Handbuch_d_Pflanzenanatomie_10_1_Roth

Ruangrak, E., Su, X., Huang, Z., Wang, X., Guo, Y., Du, Y., & Gao, J. (2018). Fine mapping of a major QTL controlling early flowering in tomato using QTL-seq. *Canadian Journal of Plant Science*, 1–11. https://doi.org/10.1139/cjps-2016-0398

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of Image Analysis. *Nature Methods*, *9*(7), 671–675.

Singh, V. K., Khan, A. W., Jaganathan, D., Thudi, M., Roorkiwal, M., Takagi, H., … Varshney, R. K. (2016). QTL-seq for rapid identification of candidate genes for 100-seed weight and root/total plant dry weight ratio under rainfed conditions in chickpea. *Plant Biotechnology Journal*, *14*(11), 2110–2119. https://doi.org/10.1111/pbi.12567

Somssich, M., Je, B. I., Simon, R., & Jackson, D. (2016). CLAVATA-WUSCHEL signaling in the shoot meristem. *Development*, *143*(18), 3238–3248. https://doi.org/10.1242/dev.133645

Stevens, M. A., & Rick, C. M. (1986). Genetics and breeding. In *The Tomato Crop* (pp. 35–109). Springer, Dordrecht. https://doi.org/10.1007/978-94-009-3137-4_2

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., … Terauchi, R. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *The Plant Journal: For Cell and Molecular Biology*, *74*(1), 174–183. https://doi.org/10.1111/tpj.12105

Tanksley, S. D. (2004). The Genetic, Developmental, and Molecular Bases of Fruit Size and Shape Variation in Tomato. *The Plant Cell*, *16*(suppl 1), S181–S189. https://doi.org/10.1105/tpc.018119

van der Knaap, E., Chakrabarti, M., Chu, Y. H., Clevenger, J. P., Illa-Berenguer, E., Huang, Z., … Wu, S. (2014). What lies beyond the eye: the molecular mechanisms regulating tomato fruit weight and shape. *Frontiers in Plant Science*, *5*. https://doi.org/10.3389/fpls.2014.00227

Wei, Q., Fu, W., Wang, Y., Qin, X., Wang, J., Li, J., … Chen, J. (2016). Rapid identification of fruit length loci in cucumber (*Cucumis sativus* L.) using next-generation sequencing (NGS)-based QTL analysis. *Scientific Reports*, *6*, 27496. https://doi.org/10.1038/srep27496

Xiao, H., Radovich, C., Welty, N., Hsu, J., Li, D., Meulia, T., & van der Knaap, E. (2009). Integration of tomato reproductive developmental landmarks and expression profiles, and the effect of SUN on fruit shape. *BMC Plant Biology*, *9*, 49. https://doi.org/10.1186/1471-2229-9-49

Xu, C., Liberatore, K. L., MacAlister, C. A., Huang, Z., Chu, Y.-H., Jiang, K., … Lippman, Z. B. (2015). A cascade of arabinosyltransferases controls shoot meristem size in tomato. *Nature Genetics*, *47*(7), 784–792. https://doi.org/10.1038/ng.3309

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Zhou, X., & Stephens, M. (2012). Genome-wide Efficient Mixed Model Analysis for Association Studies. *Nature Genetics*, *44*(7), 821–824. https://doi.org/10.1038/ng.2310

Table 3. 1. Significance and association of all the tested markers to weight and circumference cell number. Cells were color coded based on significance: light brown (P<0.05), medium brown (P<0.01), and dark brown (P<0.001). LOD scores in blue indicate association.

Population/trait	Chr.*	Position (bp)**	Marker	Genetic distance (cM)	P value	LOD score	Variance explained
17S62 weight	2	44,246,317	18EP16	0	1.00E+00	0.78	
	2	47,861,431	18EP25	15.32	2.13E-03	**3.99**	
	2	49,403,864	18EP28	20.43	**1.71E-04**	**3.87**	
	2	50,265,831	18EP83	23.47	**1.53E-04**	**5.52**	
	2	51,283,924	18EP89	27.37	**9.63E-07**	**5.81**	
	2	52,385,669	18EP95	31.72	**1.84E-09**	**7.36**	27.80%
	2	53,298,093	18EP101	36.22	**2.75E-07**	**6.67**	
	2	55,261,865	18EP107	45.96	**3.20E-07**	**6.28**	
	3	59,088,373	18EP306	0	1.62E-01	1.45	
	7	60,010,835	18EP183	0	8.41E-01	1.83	
	8	59,101,186	18EP312	0	1.26E-01	1.35	
	8	61,135,475	18EP315	8.11	1.00E+00	1.23	
	11	513,622	18EP119	0	1.00E+00	0.08	
	12	6,390,137	17EP240	0	1.18E-01	1.79	
	12	6,404,332	17EP243	0	1.29E-01	1.76	
	12	6,605,941	17EP246	0	1.86E-01	1.72	
17S64 weight	2	44,246,317	18EP16	0	1.00E+00	0.61	
	2	47,861,431	18EP25	16.2	**1.94E-04**	2.98	
	2	49,403,864	18EP28	21.9	**4.77E-07**	**4.39**	
	2	50,265,831	18EP83	26.12	**1.37E-07**	**5.05**	
	2	51,283,924	18EP89	27.9	**1.30E-07**	**5.38**	
	2	52,385,669	18EP95	30.6	**3.00E-09**	**6.03**	
	2	53,298,093	18EP101	33.4	**1.18E-09**	**5.97**	
	2	55,261,865	18EP107	41.9	**2.45E-10**	**7.07**	39.84%
	3	59,088,373	18EP306	0	1.00E+00	0.77	
	6	46,145,655	18EP327	0	1.00E+00	0.58	
	6	49,091,364	18EP330	14.6	7.81E-01	0.84	
	7	60,010,835	18EP183	0	2.19E-01	2.12	
	8	59,101,186	18EP312	0	1.00E+00	0.48	
	8	61,135,475	18EP315	9.67	7.92E-01	1.60	
	9	65,352,211	18EP336	0	6.67E-03	2.14	9.62%
	11	513,622	18EP119	0	4.52E-01	1.74	
	12	6,390,137	17EP240	0	1.00E+00	0.26	

	12	6,404,332	17EP243	0	1.00E+00	0.28	
	12	6,605,941	17EP246	0	1.00E+00	0.25	
17S64 circumference cell number 1	2	44,246,317	18EP16	0	7.99E-02	2.03	
	2	47,861,431	18EP25	16.2	1.00E+00	0.49	
	2	49,403,864	18EP28	21.9	7.80E-01	1.13	
	2	50,265,831	18EP83	26.12	1.00E+00	0.65	
	2	51,283,924	18EP89	27.9	1.00E+00	0.30	
	2	52,385,669	18EP95	30.6	1.00E+00	0.09	
	2	53,298,093	18EP101	33.4	1.00E+00	0.07	
	2	55,261,865	18EP107	41.9	1.00E+00	0.34	
	3	59,088,373	18EP306	0	4.11E-01	1.85	
	6	46,145,655	18EP327	0	**2.24E-05**	**5.00**	26.10%
	6	49,091,364	18EP330	14.6	1.39E-02	**3.28**	
	7	60,010,835	18EP183	0	1.00E+00	0.24	
	8	59,101,186	18EP312	0	1.00E+00	0.78	
	8	61,135,475	18EP315	9.67	1.00E+00	0.21	
	9	65,352,211	18EP336	0	1.00E+00	0.51	
	11	513,622	18EP119	0	1.00E+00	0.09	
	12	6,390,137	17EP240	0	1.00E+00	0.58	
	12	6,404,332	17EP243	0	1.00E+00	0.55	
	12	6,605,941	17EP246	0	1.00E+00	0.57	

*Chromosome
**SL3.0 genome version

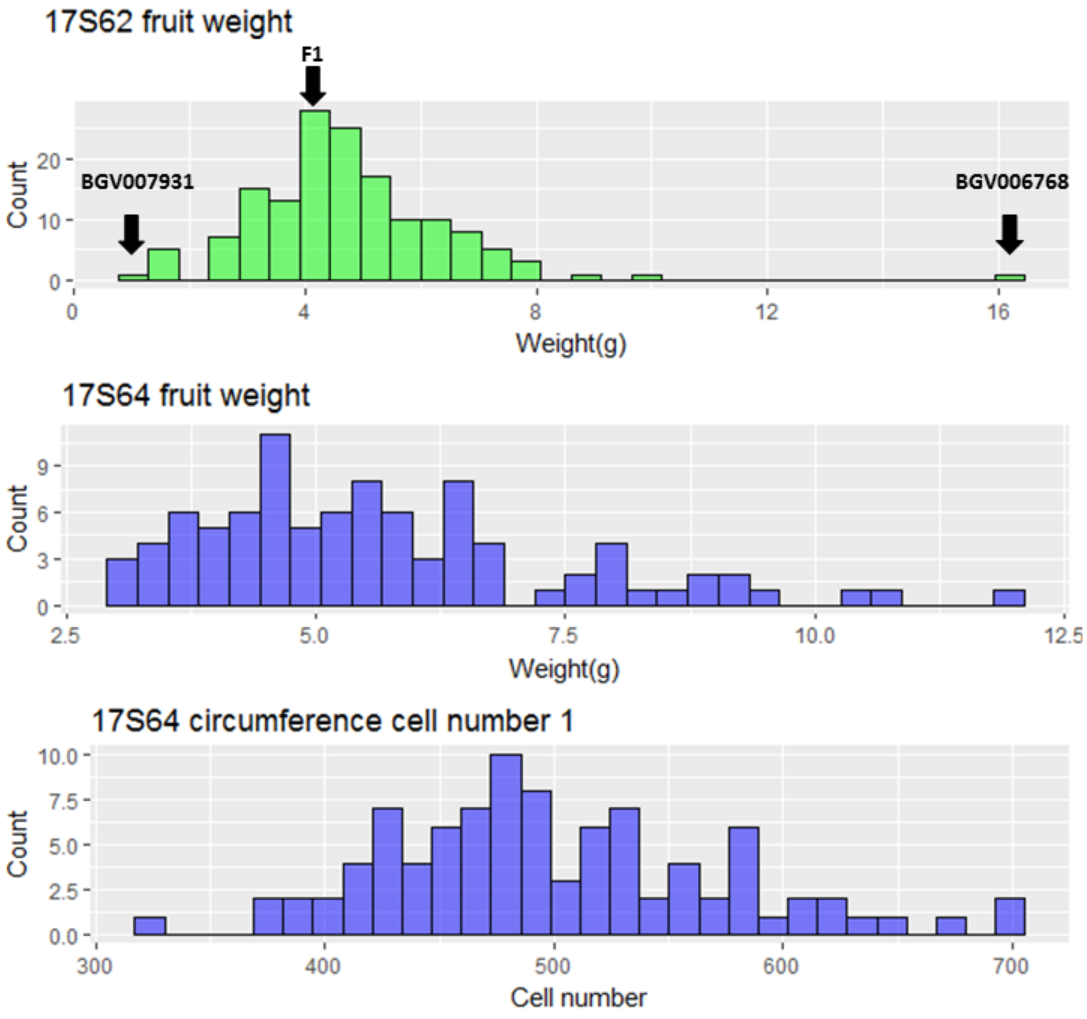


Figure 3. 1. Fruit weight and circumference cell number distribution for the traits selected for QTLseq analysis. Parental lines and F_1 phenotypes are indicated for 17S62 weight.

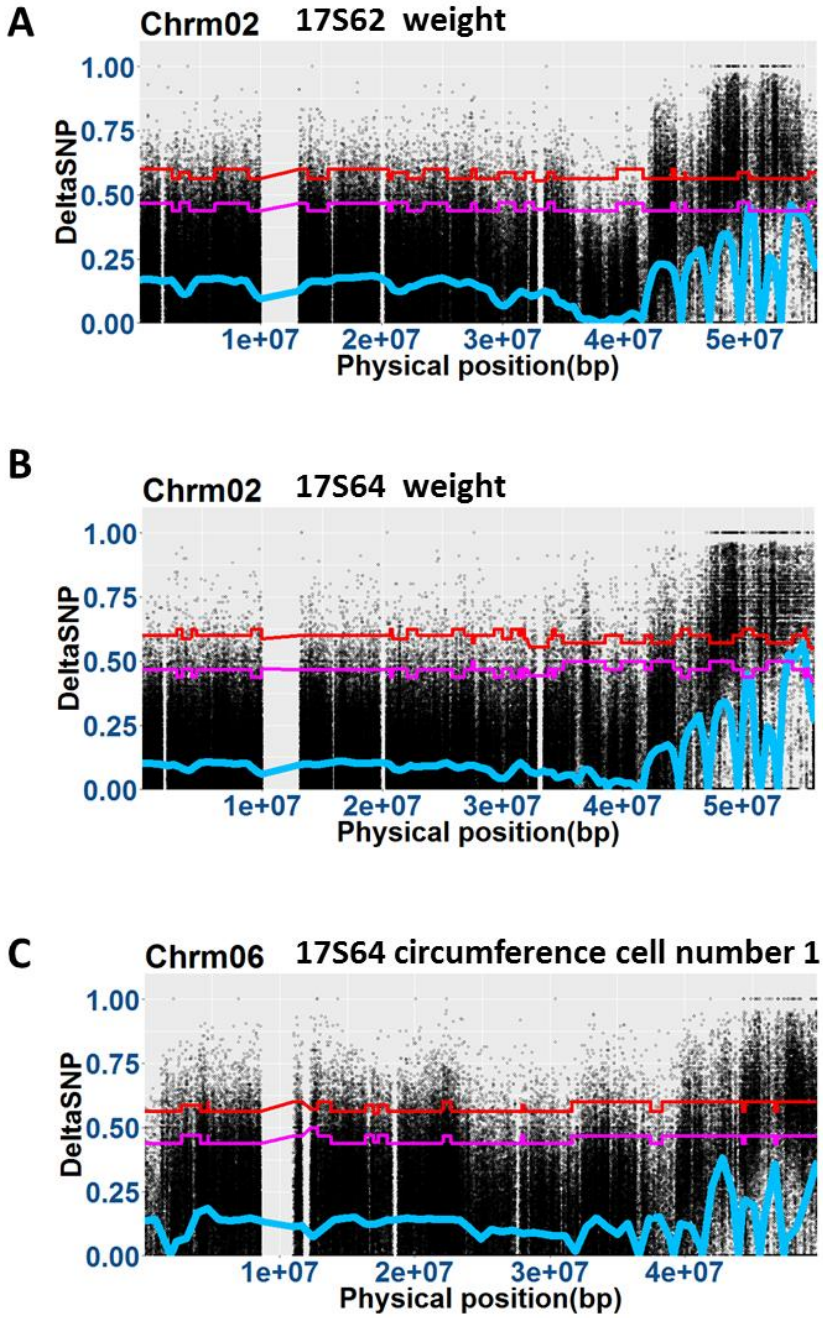


Figure 3. 2. QTLseq outputs for the chromosomes with the highest effect QTL for each trait. The purple and red lines are the 95 and 99 CI for the regions respectively. The blue line represents the tricube smoothed value. Tricube smoothed values higher or close to the CI were considered for further analysis.

Supplementary table S3.1. Markers used for genotyping the F_1 plants to confirm they were not the result of selfing.

Plant	Allele	Parents	Generation	Marker type	Primer	Primer sequence	Restriction enzyme
BGV006768	Orange Strawberry allele	BGV006768	parent	dCAPS	16EP436	ACACCGGTCTCAAAAGTGTG	HinfI
					16EP437	CTAAAGTTAATTAGTTATACAACTAGCGAT	
BGV007931	LA1589 allele	BGV007931	parent	dCAPS	16EP436	ACACCGGTCTCAAAAGTGTG	HinfI
					16EP437	CTAAAGTTAATTAGTTATACAACTAGCGAT	
16S122	heterozygous	BGV007931 x BGV006768	F_1	dCAPS	16EP436	ACACCGGTCTCAAAAGTGTG	HinfI
					16EP437	CTAAAGTTAATTAGTTATACAACTAGCGAT	
16S129	heterozygous	BGV006768 x BGV007931	F_1	dCAPS	16EP436	ACACCGGTCTCAAAAGTGTG	HinfI
					16EP437	CTAAAGTTAATTAGTTATACAACTAGCGAT	

Supplementary table S3.2. Measured phenotypes for population 17S62.

Plant ID	Weight (g)	Plant ID	Weight (g)	Plant ID	Weight (g)
17S62-1	2.77	17S62-55	7.47	17S62-109	5.18
17S62-2	3.56	17S62-56	4.12	17S62-110	2.98
17S62-3	7.55	17S62-57	5.69	17S62-111	3.57
17S62-4	3.09	17S62-58	4.13	17S62-112	4.67
17S62-5	7.78	17S62-59	4.99	17S62-113	
17S62-6	4.30	17S62-60	3.18	17S62-114	2.73
17S62-7	4.87	17S62-61	3.01	17S62-115	4.12
17S62-8	5.92	17S62-62	6.33	17S62-116	4.40
17S62-9	4.24	17S62-63	5.20	17S62-117	3.91
17S62-10	3.37	17S62-64	6.11	17S62-118	6.52
17S62-11	6.06	17S62-65	5.00	17S62-119	4.85
17S62-12	3.81	17S62-66	5.74	17S62-120	5.23
17S62-13	6.75	17S62-67	4.01	17S62-121	4.21
17S62-14	10.04	17S62-68	4.71	17S62-122	4.00
17S62-15		17S62-69	4.54	17S62-123	2.70
17S62-16	3.21	17S62-70	4.65	17S62-124	2.97
17S62-17	5.42	17S62-71	5.03	17S62-125	6.78
17S62-18	6.50	17S62-72	6.29	17S62-126	2.84
17S62-19	3.82	17S62-73	6.46	17S62-127	3.41
17S62-20	5.44	17S62-74	4.85	17S62-128	4.12
17S62-21	2.94	17S62-75	6.20	17S62-129	4.09
17S62-22	6.44	17S62-76	3.11	17S62-130	4.55
17S62-23	5.97	17S62-77	5.61	17S62-131	4.67
17S62-24	5.11	17S62-78	4.74	17S62-132	3.76
17S62-25	3.29	17S62-79	4.13	17S62-133	5.20
17S62-26	4.35	17S62-80	4.73	17S62-134	3.00
17S62-27	3.63	17S62-81	4.69	17S62-135	3.01
17S62-28	5.15	17S62-82	2.83	17S62-136	5.18
17S62-29		17S62-83	6.76	17S62-137	
17S62-30	5.33	17S62-84	4.42	17S62-138	3.68
17S62-31	4.10	17S62-85	5.13	17S62-139	2.92
17S62-32	4.38	17S62-86	7.01	17S62-140	5.21
17S62-33	4.26	17S62-87	7.02	17S62-141	5.76
17S62-34	3.74	17S62-88	4.38	17S62-142	6.61
17S62-35	4.79	17S62-89	5.99	17S62-143	3.31
17S62-36	4.85	17S62-90	4.91	17S62-144	4.15
17S62-37	4.28	17S62-91	6.14	17S63-1	4.49
17S62-38		17S62-92	2.75	17S63-2	4.77
17S62-39	4.62	17S62-93	4.68	17S63-3	4.85

17S62-40	4.82	17S62-94	4.36	17S63-4	3.63
17S62-41	4.03	17S62-95	7.58	17S63-5	3.79
17S62-42	5.28	17S62-96	8.66	17S63-6	4.04
17S62-43	7.58	17S62-97	5.41	BGV006768-1	27.47
17S62-44	3.76	17S62-98		BGV006768-2	26.63
17S62-45	4.47	17S62-99	6.62	BGV006768-3	16.10
17S62-46	5.71	17S62-100	3.01	BGV006768-4	31.73
17S62-47	4.28	17S62-101	6.48	BGV006768-5	27.10
17S62-48	3.93	17S62-102	4.85	BGV006768-6	
17S62-49	7.40	17S62-103	2.51	BGV007931-1	1.46
17S62-50	7.10	17S62-104	5.55	BGV007931-2	1.75
17S62-51	4.00	17S62-105	4.71	BGV007931-3	1.60
17S62-52	4.30	17S62-106	5.72	BGV007931-4	1.58
17S62-53	4.21	17S62-107		BGV007931-5	0.97
17S62-54	7.07	17S62-108	4.69	BGV007931-6	1.60

Supplementary table S3.3. Measured phenotypes for population 17S64.

Plant ID	Weigh t (g)	Area (mm^2)	Perimeter (mm)	Pericarp area (mm^2)	Pericarp area ratio	Pericarp plus septum area (mm^2)	Pericarp plus septum area ratio	Columella plus placenta area (mm^2)	Columella plus placenta area ratio
17S64-1	5.55	384.69	73.96	131.80	0.343	155.33	0.404	37.10	0.096
17S64-2	7.93	508.96	84.62	201.42	0.396	224.60	0.441	58.72	0.115
17S64-3	6.28	399.40	74.72	149.70	0.375	169.63	0.425	34.94	0.087
17S64-4	7.71	498.65	83.45	186.96	0.375	212.86	0.427	52.60	0.105
17S64-5	4.59	323.65	67.24	108.64	0.336	124.09	0.383	31.91	0.099
17S64-6	6.06	406.11	75.21	143.77	0.354	166.64	0.410	41.82	0.103
17S64-7	8.02	482.89	81.93	180.43	0.374	207.29	0.429	54.78	0.113
17S64-8	10.57	620.14	93.10	280.37	0.452	323.85	0.522	66.73	0.108
17S64-9	4.79	365.08	71.38	130.90	0.359	150.93	0.413	40.09	0.110
17S64-10	8.38	518.11	85.12	224.26	0.433	257.78	0.498	44.15	0.085
17S64-11	8.10	536.63	86.72	189.07	0.352	219.55	0.409	52.83	0.098
17S64-12	9.03	528.12	85.88	192.54	0.365	211.93	0.401	53.88	0.102
17S64-13	4.86	345.52	69.64	134.23	0.388	155.90	0.451	31.17	0.090
17S64-14	4.11	332.60	68.14	112.71	0.339	126.48	0.380	37.62	0.113
17S64-15	4.49	354.83	70.34	113.29	0.319	129.90	0.366	34.48	0.097
17S64-16	4.73	342.32	68.82	109.96	0.321	127.60	0.373	31.57	0.092
17S64-17	5.57	375.50	72.35	135.89	0.362	148.56	0.396	40.67	0.108
17S64-18	6.57	437.96	78.13	183.76	0.420	206.36	0.471	46.62	0.106
17S64-19	4.36	341.93	69.07	111.86	0.327	126.63	0.370	50.40	0.147
17S64-20	8.57	488.51	82.58	197.05	0.403	221.52	0.453	54.65	0.112
17S64-21	4.71	346.72	69.42	127.64	0.368	142.05	0.410	46.02	0.133
17S64-22	6.32	391.17	74.09	145.44	0.372	169.31	0.433	34.34	0.088
17S64-23	3.78	290.31	63.67	102.82	0.354	116.57	0.402	26.35	0.091

17S64-24	4.54	342.81	69.06	133.14	0.388	149.44	0.436	35.77	0.104
17S64-25	6.56	439.93	78.49	163.14	0.371	179.75	0.409	61.11	0.139
17S64-26	5.70	390.86	74.01	134.48	0.344	151.21	0.387	36.45	0.093
17S64-28	5.97	406.98	75.20	148.12	0.364	172.78	0.425	45.94	0.113
17S64-29	4.43	341.03	68.90	121.90	0.357	138.42	0.406	25.66	0.075
17S64-30	5.74	418.11	76.45	154.32	0.369	177.62	0.425	46.12	0.110
17S64-31	6.72	406.36	75.11	166.64	0.410	191.65	0.472	34.64	0.085
17S64-32	5.71	420.22	76.43	140.60	0.335	160.05	0.381	44.43	0.106
17S64-33	6.51	420.73	76.56	161.97	0.385	186.04	0.442	44.19	0.105
17S64-34	5.54	370.10	71.91	152.11	0.411	172.06	0.465	30.96	0.084
17S64-35	6.02	416.84	76.71	156.20	0.375	178.47	0.428	37.94	0.091
17S64-36	3.32	262.47	60.54	86.98	0.331	97.82	0.373	22.02	0.084
17S64-37	7.72	460.87	80.22	183.52	0.398	210.81	0.457	42.63	0.092
17S64-38	5.23	371.58	72.12	136.56	0.368	157.75	0.425	34.92	0.094
17S64-39	4.23	353.54	70.22	117.69	0.333	138.89	0.393	34.71	0.098
17S64-40	5.61	398.92	74.82	147.53	0.370	166.91	0.418	34.01	0.085
17S64-41	3.21	253.30	60.06	83.01	0.328	93.53	0.369	20.99	0.083
17S64-42	4.29	322.05	67.32	118.59	0.368	134.79	0.419	29.11	0.090
17S64-43	5.47	371.04	72.13	148.87	0.401	167.70	0.452	39.77	0.107
17S64-44	5.08	356.26	70.50	139.73	0.392	160.11	0.449	34.81	0.098
17S64-45	5.12	357.44	70.59	138.78	0.388	158.28	0.443	29.03	0.081
17S64-46	5.69	386.79	73.59	139.22	0.360	157.74	0.408	47.63	0.123
17S64-47	3.00	276.28	62.21	99.06	0.359	111.86	0.405	20.53	0.074
17S64-48	7.30	472.21	81.26	167.68	0.355	198.91	0.421	49.80	0.105
17S64-49	5.17	346.33	69.71	127.32	0.368	143.83	0.415	40.79	0.118
17S64-50	4.35	340.22	68.92	122.82	0.361	141.09	0.415	27.03	0.079
17S64-51	3.90	317.82	66.82	114.58	0.361	130.44	0.410	32.10	0.101
17S64-52	7.89	513.09	84.69	213.86	0.417	242.89	0.473	52.04	0.101

17S64-53	11.88	616.48	92.87	266.97	0.433	299.92	0.487	60.15	0.098
17S64-54	3.60	289.61	63.96	107.84	0.372	119.07	0.411	25.97	0.090
17S64-55	5.62	391.07	74.27	162.04	0.414	180.24	0.461	32.70	0.084
17S64-57	4.49	330.57	67.94	117.49	0.355	132.47	0.401	31.01	0.094
17S64-58	10.50	623.38	93.31	232.47	0.373	271.73	0.436	76.71	0.123
17S64-59	3.64	280.52	62.81	80.78	0.288	92.61	0.330	31.26	0.111
17S64-60	6.08	417.23	76.16	143.61	0.344	165.52	0.397	45.46	0.109
17S64-62	9.31	550.22	87.84	234.30	0.426	262.45	0.477	44.74	0.081
17S64-63	6.78	428.36	77.42	176.58	0.412	197.82	0.462	37.26	0.087
17S64-64	6.53	427.45	77.42	172.59	0.404	197.71	0.463	42.09	0.098
17S64-65	4.08	322.39	66.97	115.90	0.359	132.93	0.412	28.69	0.089
17S64-66	4.46	346.76	69.54	114.81	0.331	130.72	0.377	32.68	0.094
17S64-67	4.68	367.03	71.55	130.88	0.357	149.73	0.408	31.07	0.085
17S64-68	4.93	340.60	69.54	134.20	0.394	153.41	0.450	29.30	0.086
17S64-69	6.47	442.23	78.62	163.39	0.369	182.43	0.413	48.23	0.109
17S64-70	4.66	365.75	71.56	133.00	0.364	149.29	0.408	35.74	0.098
17S64-71	3.59	301.02	65.02	108.92	0.362	122.65	0.407	31.55	0.105
17S64-72	3.44	241.65	57.87	83.98	0.348	95.45	0.395	18.04	0.075
17S64-73	4.19	324.23	67.11	115.93	0.358	132.83	0.410	31.17	0.096
17S64-74	5.58	403.37	74.92	136.50	0.338	157.14	0.390	40.44	0.100
17S64-75	3.82	314.93	66.31	101.97	0.324	117.75	0.374	27.36	0.087
17S64-76	4.09	321.91	67.24	106.79	0.332	121.84	0.378	32.25	0.100
17S64-77	5.18	367.44	71.49	116.25	0.316	129.81	0.353	39.69	0.108
17S64-78	6.46	450.46	79.46	177.70	0.394	201.21	0.447	37.40	0.083
17S64-79	5.04	350.02	69.77	131.47	0.376	149.11	0.426	28.46	0.081
17S64-80	3.16	282.98	62.89	110.61	0.391	129.05	0.456	21.36	0.075
17S64-81	3.24	273.16	61.75	94.68	0.347	108.60	0.398	25.08	0.092
17S64-82	5.59	391.81	73.79	155.69	0.397	175.38	0.448	40.72	0.104

17S64-83	3.86	280.61	62.67	99.38	0.354	111.27	0.397	27.46	0.098
17S64-84	4.66	353.88	70.14	137.33	0.388	156.15	0.441	30.21	0.085
17S64-85	9.12	533.66	86.46	225.35	0.422	255.71	0.479	43.99	0.082
17S64-86		341.45	69.17	119.08	0.349	133.89	0.392	30.92	0.091
17S64-87	4.97	347.81	69.97	137.64	0.396	155.44	0.447	26.48	0.076
17S64-88	6.68	450.23	79.14	149.28	0.332	171.68	0.381	55.09	0.122
17S64-89	4.69	351.66	69.90	131.56	0.374	149.20	0.424	32.41	0.092
17S64-90	8.94	535.10	86.38	187.56	0.351	212.44	0.397	62.90	0.118
17S64-91	9.57	576.49	89.69	212.33	0.368	235.14	0.408	76.65	0.133
17S64-92	3.76	304.33	65.45	106.86	0.351	120.04	0.394	28.41	0.093
17S64-93	3.24	260.72	61.11	92.13	0.353	102.82	0.394	25.57	0.098
17S64-94	5.77	370.18	71.69	134.80	0.364	157.53	0.426	42.12	0.114
17S64-95	6.81	451.40	79.31	177.97	0.394	199.22	0.441	40.85	0.090
17S64-96	5.23	376.25	72.45	132.67	0.353	148.19	0.394	35.53	0.094

Supplementary table S3.3. continued

Plant ID	Pericarp thickness (mm)	Pericarp cell layers	Pericarp max cell size (mm^2)	Pericarp cell number/mm	Circumference cell number 1 (cell layers)	Cell diameter (mm)	Circumference cell number 2 (cell size)
17S64-1	1.546	12.06	0.065	7.80	576.73	0.287	257.68
17S64-2	2.019	13.58	0.092	6.73	569.24	0.342	247.17
17S64-3	1.622	12.58	0.065	7.76	579.60	0.287	260.16
17S64-4	2.005	12.31	0.112	6.14	512.07	0.378	220.66
17S64-5	1.601	12.78	0.067	7.98	536.79	0.291	230.89
17S64-6	1.530	12.06	0.080	7.88	592.67	0.319	235.57
17S64-7	1.955	12.78	0.084	6.54	535.47	0.326	251.19
17S64-8	2.888	13.26	0.150	4.59	427.33	0.437	213.26
17S64-9	1.791	11.31	0.088	6.31	450.60	0.335	212.87
17S64-10	2.770	13.94	0.128	5.03	428.53	0.403	211.05
17S64-11	1.901	11.92	0.087	6.27	543.47	0.333	260.72
17S64-12	2.017	12.97	0.080	6.43	552.32	0.319	268.93
17S64-13	1.677	11.53	0.073	6.87	478.78	0.306	227.82
17S64-14	1.146	11.81	0.038	10.30	702.04	0.220	309.32
17S64-15	1.277	11.53	0.048	9.03	635.02	0.246	285.92
17S64-16	1.363	10.39	0.059	7.62	524.43	0.275	250.32
17S64-17	2.042	11.79	0.097	5.77	417.69	0.351	206.29
17S64-18	2.284	14.08	0.109	6.17	481.66	0.373	209.61
17S64-19	1.334	11.78	0.049	8.83	609.87	0.250	275.96
17S64-20	2.267	13.46	0.102	5.94	490.16	0.361	228.77
17S64-21	1.676	12.94	0.071	7.72	536.12	0.302	230.24
17S64-22	1.880	11.47	0.100	6.10	452.19	0.356	207.89
17S64-23	1.428	10.56	0.077	7.39	470.84	0.312	203.91

17S64-24	1.788	11.89	0.089	6.65	459.28	0.336	205.31
17S64-25	1.931	11.58	0.100	6.00	470.89	0.357	219.96
17S64-26	1.306	12.39	0.045	9.49	702.22	0.240	308.21
17S64-28	1.794	13.17	0.071	7.34	551.90	0.301	250.14
17S64-29	1.317	11.72	0.052	8.90	613.42	0.258	267.49
17S64-30	1.825	11.78	0.077	6.46	493.54	0.313	244.04
17S64-31	2.252	12.39	0.122	5.50	413.18	0.394	190.78
17S64-32	1.497	12.22	0.056	8.17	624.17	0.267	285.77
17S64-33	1.674	11.19	0.062	6.69	511.95	0.282	271.68
17S64-34	1.682	10.92	0.083	6.49	466.78	0.326	220.67
17S64-35	1.873	12.58	0.076	6.72	515.22	0.312	246.08
17S64-36	1.439	9.89	0.069	6.87	416.14	0.296	204.61
17S64-37	1.903	11.42	0.087	6.00	481.35	0.333	240.58
17S64-38	1.572	11.06	0.067	7.03	507.13	0.292	246.81
17S64-39	1.450	10.78	0.066	7.43	521.97	0.289	242.83
17S64-40	1.525	11.97	0.063	7.85	587.46	0.283	264.07
17S64-41	1.527	10.19	0.074	6.68	400.96	0.307	195.42
17S64-42	1.666	9.72	0.103	5.84	392.95	0.362	185.82
17S64-43	1.947	11.92	0.091	6.12	441.47	0.341	211.34
17S64-44	1.672	11.28	0.070	6.74	475.40	0.299	236.00
17S64-45	1.711	11.46	0.074	6.70	472.61	0.308	229.23
17S64-46	1.510	11.08	0.055	7.34	540.16	0.264	278.37
17S64-47	1.105	10.08	0.050	9.12	567.40	0.253	246.30
17S64-48	2.011	12.06	0.102	5.99	487.08	0.360	225.84
17S64-49	1.412	10.28	0.071	7.28	507.52	0.301	231.95
17S64-50	1.322	11.25	0.045	8.51	586.46	0.240	286.95
17S64-51	1.235	10.26	0.052	8.31	554.98	0.257	260.06
17S64-52	2.218	12.75	0.101	5.75	486.77	0.359	235.64

17S64-53	2.868	13.83	0.129	4.82	447.92	0.405	229.06
17S64-54	1.482	11.34	0.065	7.65	489.48	0.288	222.03
17S64-55	1.959	11.67	0.093	5.95	442.30	0.345	215.57
17S64-57	1.697	11.11	0.091	6.55	444.91	0.340	199.89
17S64-58	1.906	11.44	0.099	6.00	560.15	0.355	262.71
17S64-59	1.392	11.42	0.054	8.20	515.18	0.261	240.36
17S64-60	1.232	10.89	0.046	8.84	673.29	0.243	313.28
17S64-62	2.502	10.86	0.147	4.34	381.30	0.433	202.81
17S64-63	2.014	12.17	0.090	6.04	467.66	0.338	229.05
17S64-64	1.895	10.44	0.100	5.51	426.64	0.357	216.81
17S64-65	1.105	9.69	0.045	8.78	587.75	0.239	280.72
17S64-66	1.280	9.72	0.048	7.59	528.06	0.246	282.17
17S64-67	1.570	10.17	0.079	6.48	463.31	0.316	226.22
17S64-68	1.706	10.50	0.075	6.15	427.99	0.309	225.19
17S64-69	1.900	11.44	0.088	6.02	473.65	0.335	234.36
17S64-70	1.461	10.44	0.069	7.15	511.66	0.297	241.17
17S64-71	1.373	10.11	0.068	7.37	478.97	0.293	221.77
17S64-72	1.418	10.50	0.063	7.41	428.63	0.283	204.15
17S64-73	1.421	10.39	0.058	7.31	490.67	0.272	246.96
17S64-74	1.617	10.69	0.073	6.61	495.56	0.305	245.28
17S64-75	1.208	9.72	0.053	8.05	533.49	0.259	256.02
17S64-76	1.095	10.19	0.047	9.31	625.76	0.244	275.14
17S64-77	1.296	10.45	0.052	8.07	576.70	0.257	278.48
17S64-78	1.799	10.76	0.072	5.98	475.20	0.304	261.77
17S64-79	1.664	10.50	0.076	6.31	440.41	0.311	224.30
17S64-80	1.589	9.73	0.079	6.12	385.08	0.317	198.24
17S64-81	1.723	10.55	0.090	6.12	377.88	0.339	182.15
17S64-82	1.782	11.44	0.079	6.42	473.90	0.318	232.38

17S64-83	1.600	10.94	0.067	6.84	428.79	0.291	215.34
17S64-84	1.600	10.64	0.057	6.65	466.28	0.270	259.56
17S64-85	2.535	12.50	0.138	4.93	426.39	0.420	205.93
17S64-86	1.391	10.06	0.065	7.23	499.98	0.288	240.53
17S64-87	2.123	9.92	0.127	4.67	326.78	0.403	173.81
17S64-88	1.543	12.69	0.049	8.23	651.26	0.250	315.94
17S64-89	1.526	9.97	0.071	6.53	456.68	0.300	233.26
17S64-90	2.034	10.81	0.092	5.31	458.80	0.342	252.79
17S64-91	2.014	11.78	0.099	5.85	524.41	0.355	252.36
17S64-92	1.389	9.81	0.061	7.06	462.13	0.278	235.07
17S64-93	1.555	10.22	0.072	6.57	401.63	0.302	202.41
17S64-94	1.624	10.92	0.065	6.72	481.97	0.288	248.98
17S64-95	2.052	10.67	0.107	5.20	412.19	0.369	214.71
17S64-96	1.719	11.72	0.062	6.82	494.12	0.280	258.85

Supplementary table S3.4. Weight and weight components measured.

Traits	Units	Description	Measurement/formula	Approximate number of samples per accession	Software
Weight	g	Weight of the fruit	Fruits bulk weighed	20	
Area	mm^2	Area of the fruit	Fruit area measured at the equatorial plane in the medio-lateral axis	8	Tomato Analyzer 4.0
Perimeter	mm	Perimeter of the fruit	Fruit perimeter measured at the equatorial plane in the medio-lateral axis	8	Tomato Analyzer 4.0
Locule number		Locule number of the fruit	Fruits were cut at the equatorial plane in the medio-lateral axis and locule number was counted	40	
Percent dry matter	%	Percent dry matter of the fruit	Calculated using the formula: $\frac{dry\ weight}{fresh\ weight} * 100$ Fruits were bulk weighed before and after drying for one week at 70-80 ° C	2 to 10	
Pericarp area	mm^2	Area of the fruit pericarp	Fruit pericarp area measured at the equatorial plane in the medio-lateral axis	8	Tomato Analyzer 4.0
Pericarp area ratio		Ratio of pericarp area to total fruit area	Calculated using the formula: $\frac{pericarp\ area}{fruit\ area}$	8	
Pericarp + septum area	mm^2	Area of the fruit pericarp and septum	Fruit pericarp and septum area measured at the equatorial plane in the medio-lateral axis	8	Tomato Analyzer 4.0
Pericarp + septum area ratio		Ratio of pericarp and septum area to total fruit area	Calculated using the formula: $\frac{pericarp\ plus\ septum\ area}{fruit\ area}$	8	
Columella + placenta area	mm^2	Area of the fruit columella and placenta	Fruit columella and placenta area measured at the equatorial plane in the medio-lateral axis	8	Tomato Analyzer 4.0
Columella + placenta area ratio		Ratio of columella and placenta area to total fruit area	$\frac{columella\ plus\ placenta\ area}{fruit\ area}$ Calculated using the formula:	8	
Pericarp thickness	mm	Thickness of the pericarp	Length of the lines traced perpendicular to the exocarp	36	ImageJ

Pericarp cell layers		Number of cell layers in the pericarp in the axial-abaxial axis	Lines were traced as perpendicular to the exocarp as possible, avoiding vascular bundles, the endocarp layer, 2-4 small cell layers right below the exocarp, and 0-1 layers right above the endocarp. Number of cells intersected by the line were counted	36	ImageJ
Pericarp max cell size	mm^2	Size of the biggest cells in the pericarp	The area of the biggest cells in the pericarp was measured	60	ImageJ
Pericarp cell number per mm		Number of cells in the pericarp per 1mm in the axial-abaxial axis	Calculated using the formula: $\frac{pericarp\ cell\ layers}{pericarp\ thickness}$	derived trait	
Circumference cell number 1 (cell layers)		Number of cells in the circumference of the fruit in the medio-lateral axis (1)	Calculated using the formula: pericarp cell number per mm*perimeter	derived trait	
Cell diameter	mm	Diameter of the biggest cells in the pericarp	Calculated using the formula: $2\sqrt{\frac{\text{pericarp max cell size}}{\pi}}$	derived trait	
Circumference cell number 2 (cell size)		Number of cells in the circumference of the fruit in the medio-lateral axis (2)	Calculated using the formula: $\frac{fruit\ perimeter}{cell\ diameter}$	derived trait	

Supplementary table S3.5. KASP markers genotyped in the population potential QTL.

Chr.*	Physical position (SL3.0 genome)	Primer ID	Primer sequence (5' -> 3')	Primer
2	47,861,431	18EP16	GAAGGTGACCAAGTTCATGCTGGGCTTGTTAATTATCTTATCCGG	forward
		18EP17	GAAGGTCGGAGTCAACGGATTTGGGCTTGTTAATTATCTTATCCGA	forward
		18EP18	TCTTCTCCTACACCTAGAAGAAATACCA	reverse
2	49,403,864	18EP25	GAAGGTGACCAAGTTCATGCTATGGTGTTTTGTTAGTTAGTATGAAGTTGTCTAC	forward
		18EP26	GAAGGTCGGAGTCAACGGATTTGGTGTTTTGTTAGTTAGTATGAAGTTGTCTAT	forward
		18EP27	AAAAGAGAACCACCAAACAAAACTACTT	reverse
2	50,265,831	18EP28	GAAGGTGACCAAGTTCATGCTCCCTTAATTTTGTCATTTAGAGGTAAGC	forward
		18EP29	GAAGGTCGGAGTCAACGGATTGCCCTTAATTTTGTCATTTAGAGGTAAGT	forward
		18EP30	AATGAGCCCAAATAACACAATCATT	reverse
2	44,246,317	18EP83	GAAGGTGACCAAGTTCATGCTACCAAAATATCAGCAACAGAAACCTT	forward
		18EP84	GAAGGTCGGAGTCAACGGATTACCAAAATATCAGCAACAGAAACCTC	forward
		18EP85	CTTGGAGTTCTATTTGCATAATTGAATG	reverse
2	51,283,924	18EP89	GAAGGTGACCAAGTTCATGCTACCGCTCAAGTTGTTCCTTTCTAT	forward
		18EP90	GAAGGTCGGAGTCAACGGATTACCGCTCAAGTTGTTCCTTTCTAC	forward
		18EP91	GCATTTGTTTGGCCTAACCTTAGT	reverse
2	52,385,669	18EP95	GAAGGTGACCAAGTTCATGCTGGTTTTATGCAGTTTAAAGTGCGC	forward
		18EP96	GAAGGTCGGAGTCAACGGATTAGGTTTTATGCAGTTTAAAGTGCGT	forward
		18EP97	GACATAACTGGCATGGTTTTGGT	reverse
2	53,298,093	18EP101	GAAGGTGACCAAGTTCATGCTCGATTTATTAGCTAACATGTTATGGCTAG	forward
		18EP102	GAAGGTCGGAGTCAACGGATTCGATTTATTAGCTAACATGTTATGGCTAA	forward
		18EP103	TTGGTGATTGGTTATGTGTGCAG	reverse
2	55,261,865	18EP107	GAAGGTGACCAAGTTCATGCTTTACACCTTTATCAGTCATAATGGCAAG	forward
		18EP108	GAAGGTCGGAGTCAACGGATTTTACACCTTTATCAGTCATAATGGCAAA	forward
		18EP109	AGTGGGAGTTCTTTTTGACCATG	reverse
3	59,088,373	18EP306	GAAGGTGACCAAGTTCATGCT TTGGTATTAATCAACAAAAGGTCACAG	forward
		18EP307	GAAGGTCGGAGTCAACGGATT TTGGTATTAATCAACAAAAGGTCACAT	forward
		18EP308	TGTATAACGACCGTAATCCTTTTAAGAAA	reverse
6	46,145,655	18EP327	GAAGGTGACCAAGTTCATGCT GACTAATTCCACAGATTAAGTGTGGTTAAT	forward
		18EP328	GAAGGTCGGAGTCAACGGATT GACTAATTCCACAGATTAAGTGTGGTTAAC	forward

		18EP329	CTAGCCATACTTGTCACACCTGTTC	reverse
6	49,091,364	18EP330	GAAGGTGACCAAGTTCATGCT ATCTCCACTTCCTAATATGCACGTAAA	forward
		18EP331	GAAGGTCGGAGTCAACGGATT ATCTCCACTTCCTAATATGCACGTAAC	forward
		18EP332	CAACTTTCCCCTTTAAAGTAGTAACCA	reverse
7	60,010,835	18EP183	GAAGGTGACCAAGTTCATGCTCGCTTCTACAAGGCTTGTTCAAC	forward
		18EP184	GAAGGTCGGAGTCAACGGATTCGCTTCTACAAGGCTTGTTCAAT	forward
		18EP185	ATCGCCTCAGAGTGATTCATCC	reverse
8	59,101,186	18EP312	GAAGGTGACCAAGTTCATGCT CTATAGCGATCCCAAATCATTTAAAGTC	forward
		18EP313	GAAGGTCGGAGTCAACGGATT ATAGCGATCCCAAATCATTTAAAGTG	forward
		18EP314	ATTTGGTCATTACAGAACTGATGGAA	reverse
8	61,135,475	18EP315	GAAGGTGACCAAGTTCATGCT TTATTCCGCCAATTCAACTCATC	forward
		18EP316	GAAGGTCGGAGTCAACGGATT TTTATTCCGCCAATTCAACTCATT	forward
		18EP317	TCCGACTTCAAAACAAATATTTACCA	reverse
9	65,352,211	18EP336	GAAGGTGACCAAGTTCATGCT ACCTCATACATATTTTGCGTGTTACTG	forward
		18EP337	GAAGGTCGGAGTCAACGGATT AACCTCATACATATTTTGCGTGTTACTC	forward
		18EP338	CTATGATGATTGAACAGGAAGATTGG	reverse
11	513,622	18EP119	GAAGGTGACCAAGTTCATGCTTTGCATTACAGGATTCAAAAGTATGAT	forward
		18EP120	GAAGGTCGGAGTCAACGGATTTTGCATTACAGGATTCAAAAGTATGAC	forward
		18EP121	AGCAGGCACTTCCTCTCACC	reverse
12	6,390,137	17EP240	GAAGGTGACCAAGTTCATGCTCTTTATTGGACCCAATGAAACTTTGT	forward
		17EP241	GAAGGTCGGAGTCAACGGATTTTTATTGGACCCAATGAAACTTTGA	forward
		17EP242	ACACCTTGTCAAATGTCCTTTTTCA	reverse
12	6,404,332	17EP243	GAAGGTGACCAAGTTCATGCTAGTCATTCAAGTTACAGGCTGATTTTTT	forward
		17EP244	GAAGGTCGGAGTCAACGGATTGTCATTCAAGTTACAGGCTGATTTTTC	forward
		17EP245	CCTCAGGCTTAAGACATTGAGTTCA	reverse
12	6,605,941	17EP246	GAAGGTGACCAAGTTCATGCTTGTGTGACACTCTCTGATTAAAACAATC	forward
		17EP247	GAAGGTCGGAGTCAACGGATTTGTGTGACACTCTCTGATTAAAACAATT	forward
		17EP248	AAAATCGCCTCATCCTTAATGAAAT	reverse

*Chromosome

Supplementary table S3.6. Bonferroni corrected ANOVA associations of the markers with the phenotypes measured for each panel. Cells were color coded based on significance: light brown (P<0.05), medium brown (P<0.01), and dark brown (P<0.001).

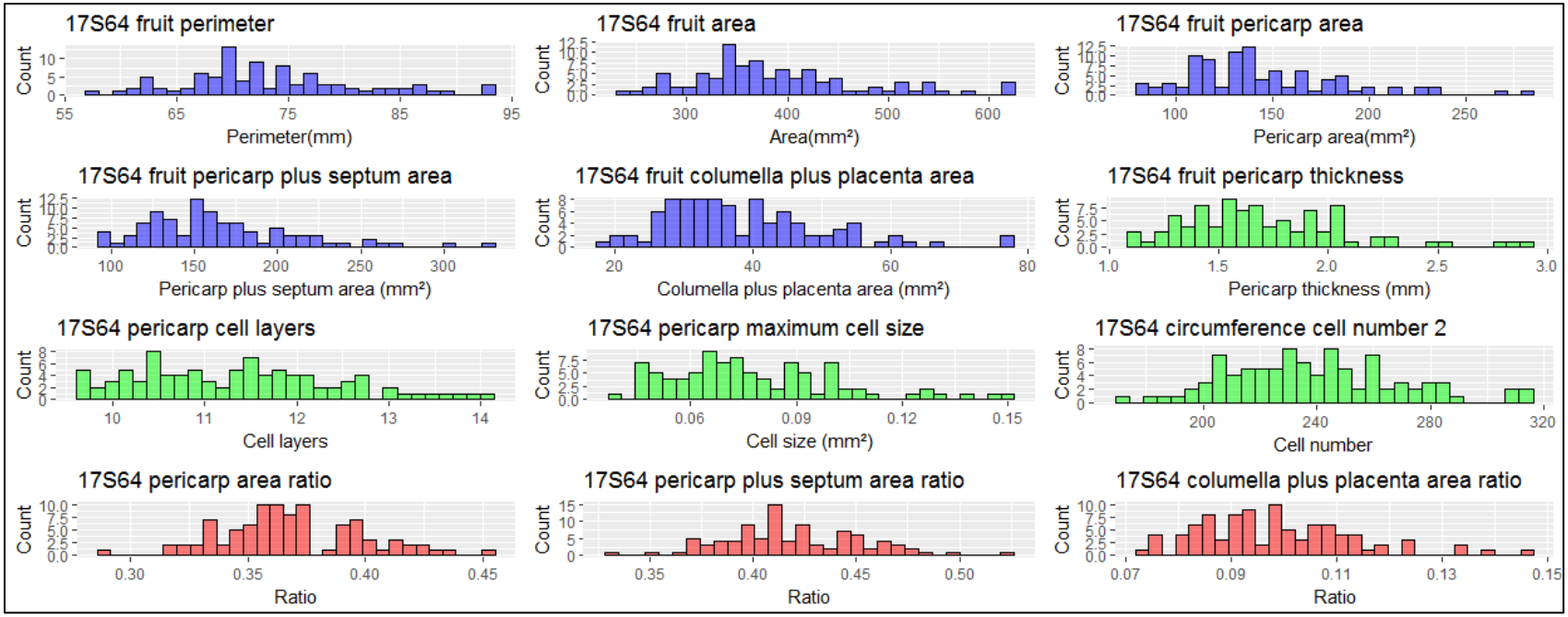
Population	Marker	Chr.*	Weight	Perimeter	Area	Pericarp area	Pericarp thickness	Cell layers	Circumference cell number 1	Circumference cell number 2	Max cell size	Per+sep area**	Col+plac Area***
17S64	18EP16	2	1.00E+00	1.00E+00	9.90E-01	5.52E-01	4.03E-01	1.00E+00	1.33E-01	7.99E-02	3.12E-02	5.76E-01	1.00E+00
17S64	18EP25	2	1.94E-04	3.38E-04	1.54E-04	2.91E-04	1.03E-03	2.00E-02	1.00E+00	1.00E+00	4.83E-03	4.43E-04	1.21E-04
17S64	18EP28	2	4.77E-07	2.22E-07	8.40E-08	1.42E-07	2.58E-06	9.85E-03	1.00E+00	7.80E-01	1.38E-05	1.40E-07	2.70E-06
17S64	18EP83	2	1.37E-07	1.07E-07	5.70E-08	2.22E-07	4.83E-06	3.22E-03	1.00E+00	1.00E+00	5.59E-05	3.08E-07	1.07E-07
17S64	18EP89	2	1.30E-07	9.69E-08	5.09E-08	9.27E-08	1.68E-06	2.36E-03	1.00E+00	1.00E+00	2.74E-05	1.52E-07	5.19E-07
17S64	18EP95	2	3.00E-09	6.06E-09	2.11E-09	1.65E-08	4.18E-06	3.76E-03	1.00E+00	1.00E+00	5.97E-05	3.93E-08	4.58E-10
17S64	18EP101	2	1.18E-09	3.71E-09	9.54E-10	1.10E-08	1.68E-05	2.47E-03	1.00E+00	1.00E+00	3.91E-04	3.95E-08	1.04E-10
17S64	18EP107	2	2.45E-10	1.48E-09	3.46E-10	8.28E-10	1.75E-07	1.76E-04	1.00E+00	1.00E+00	5.97E-05	1.98E-09	5.42E-08
17S64	18EP306	3	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	4.11E-01	1.00E+00	1.00E+00	1.00E+00
17S64	18EP327	6	1.00E+00	1.00E+00	1.00E+00	1.00E+00	3.88E-02	1.00E+00	4.55E-03	2.24E-05	3.06E-04	1.00E+00	1.00E+00
17S64	18EP330	6	7.81E-01	1.00E+00	1.00E+00	3.09E-01	2.52E-03	6.96E-01	6.48E-02	1.39E-02	5.43E-04	3.30E-01	1.00E+00
17S64	18EP183	7	2.19E-01	2.38E-01	2.73E-01	1.45E-01	4.14E-01	8.85E-01	1.00E+00	1.00E+00	1.00E+00	1.68E-01	1.00E+00
17S64	18EP312	8	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	2.90E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
17S64	18EP315	8	7.92E-01	5.27E-01	5.95E-01	8.26E-01	1.00E+00	1.09E-01	1.00E+00	1.00E+00	1.00E+00	8.27E-01	2.25E-01
17S64	18EP336	9	6.67E-03	4.07E-02	2.61E-02	5.40E-03	1.57E-02	1.00E+00	1.00E+00	1.00E+00	2.34E-02	4.21E-03	1.76E-02
17S64	18EP119	11	4.52E-01	6.83E-01	7.70E-01	8.94E-01	9.09E-01	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00

17S64	17EP240	12	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
17S64	17EP243	12	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
17S64	17EP246	12	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00	1.00E+00
17S62	18EP16	2	1.00E+00	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
17S62	18EP25	2	2.13E-03	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
17S62	18EP28	2	1.71E-04	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
17S62	18EP83	2	1.53E-04	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
17S62	18EP89	2	9.63E-07	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
17S62	18EP95	2	1.84E-09	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
17S62	18EP101	2	2.75E-07	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
17S62	18EP107	2	3.20E-07	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
17S62	18EP306	3	1.62E-01	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
17S62	18EP183	7	8.41E-01	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
17S62	18EP312	8	1.26E-01	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
17S62	18EP315	8	1.00E+00	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
17S62	18EP119	11	1.00E+00	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
17S62	17EP240	12	1.18E-01	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
17S62	17EP243	12	1.29E-01	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
17S62	17EP246	12	1.86E-01	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

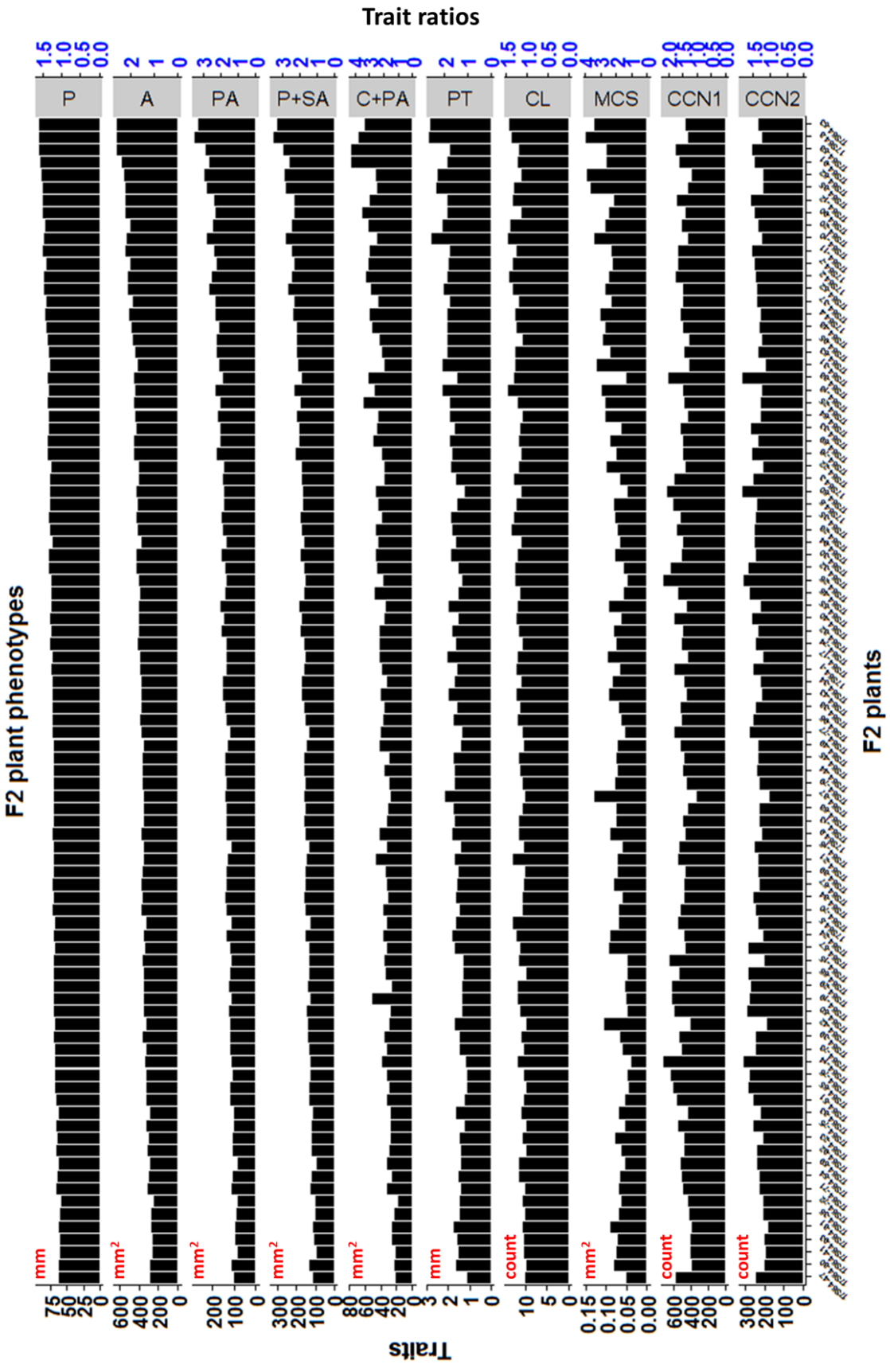
*Chromosome

**Pericarp+septum area

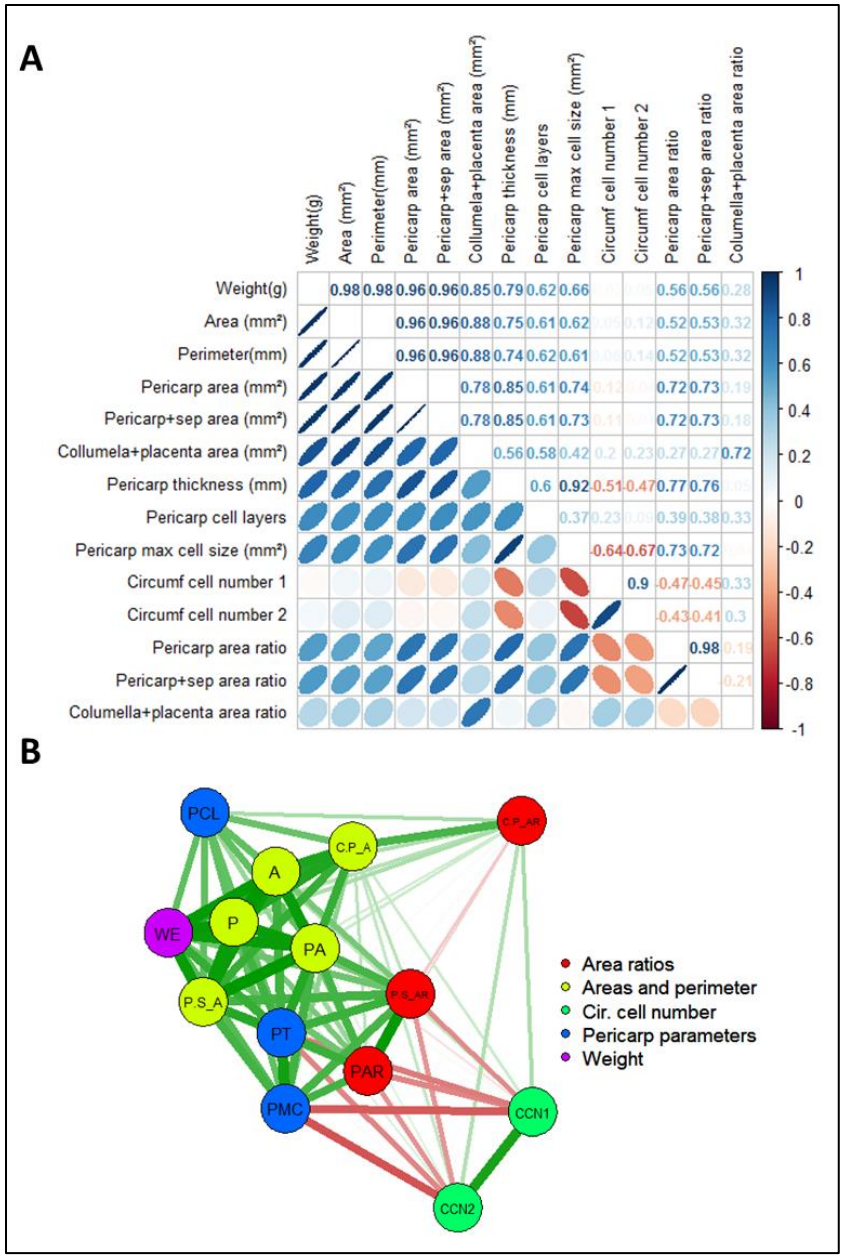
***Columella+placenta area



Supplementary figure S3. 1. Distribution for the different phenotypes recorded for 17S64.


Trait ratios
P
A
PA
P+SA
C+PA
PT
CL
MCS
CCN1
CCN2
mm
mm²
count
Traits
F2 plant phenotypes
F2 plants

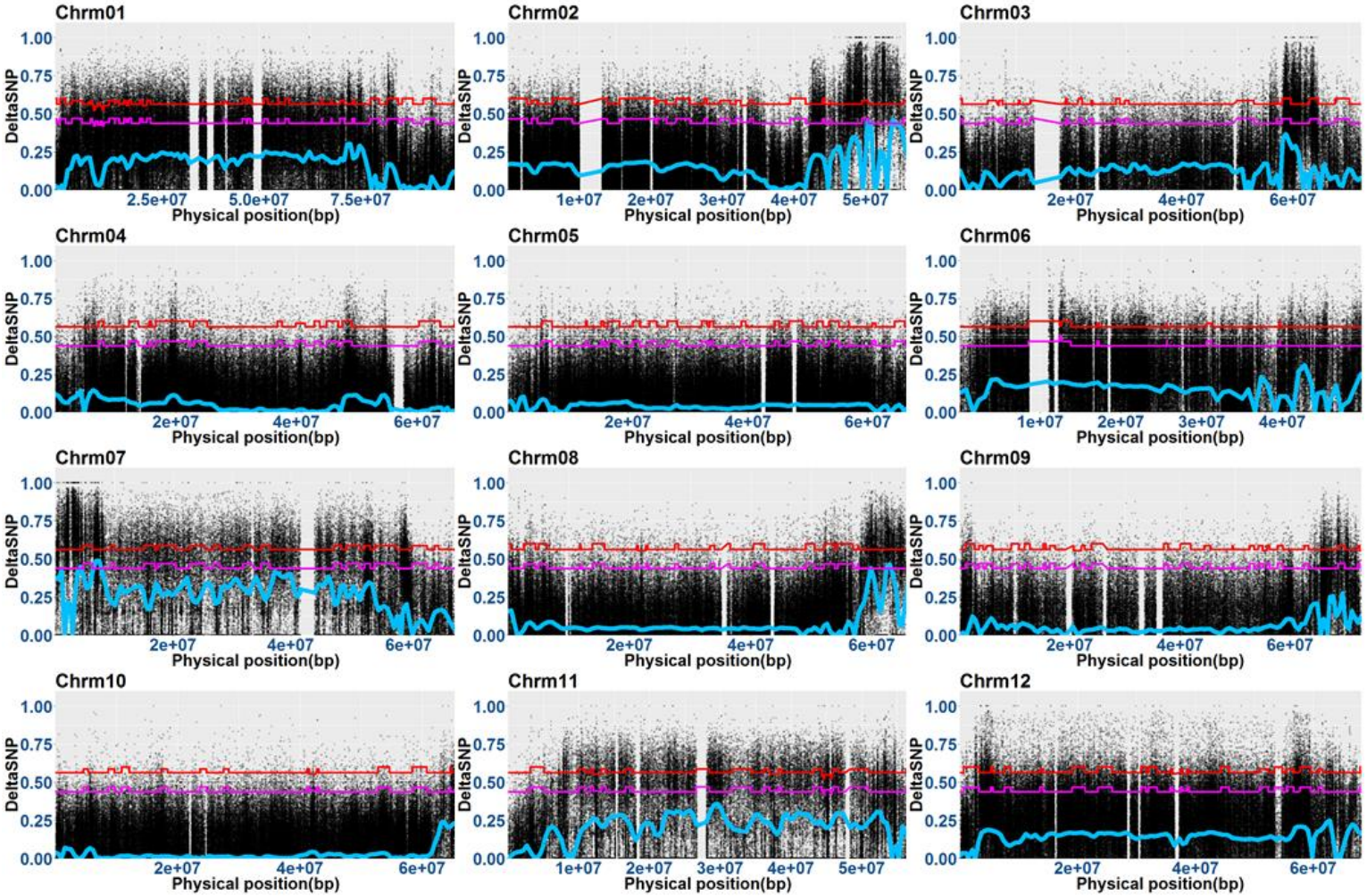
Supplementary figure S3. 2. Phenotypes recorded for 17S64 F_2 plants arranged in order of increasing fruit size in the x-axis. The y-axis "Traits" shows the phenotypic values while y-axis "Trait ratios" shows the ratio of the trait relative to the F_2 plant with the smallest phenotypic value for the trait in question. P= perimeter, A=area, PA=pericarp area, P+SA =pericarp plus septum area, C+PA = columella plus placenta area, PT= pericarp thickness, CL = cell layers, MCS = maximum cell size, CCN1 = cell circumference number 1, and CCN2 = cell circumference number 2.



Supplementary figure S3.3. Correlations between weight and weight parameters. A. Color indicates the strength of the correlation where dark blue is a positive correlation of 1, and dark red is a negative correlation of 1. Coefficients of correlation (r) are also shown for each combination. B. Correlation network where green is a positive correlation and red a negative correlation. The thickness of the line indicates the strength of the correlation. WE = weight, P=perimeter, A=area, PA=pericarp area, P.S_A=pericarp plus septum area, C.P_A=columella plus placenta area, PT=pericarp thickness, PCL=pericarp cell number, PMCS=pericarp maximum cell size, CCN1=circumference cell number 1, CNN2= circumference cell number 2, PDM= percent dry matter, PAR=pericarp area ratio, P.S_AR=pericarp plus septum area ratio, C.P_AR=columella plus placenta area ratio.

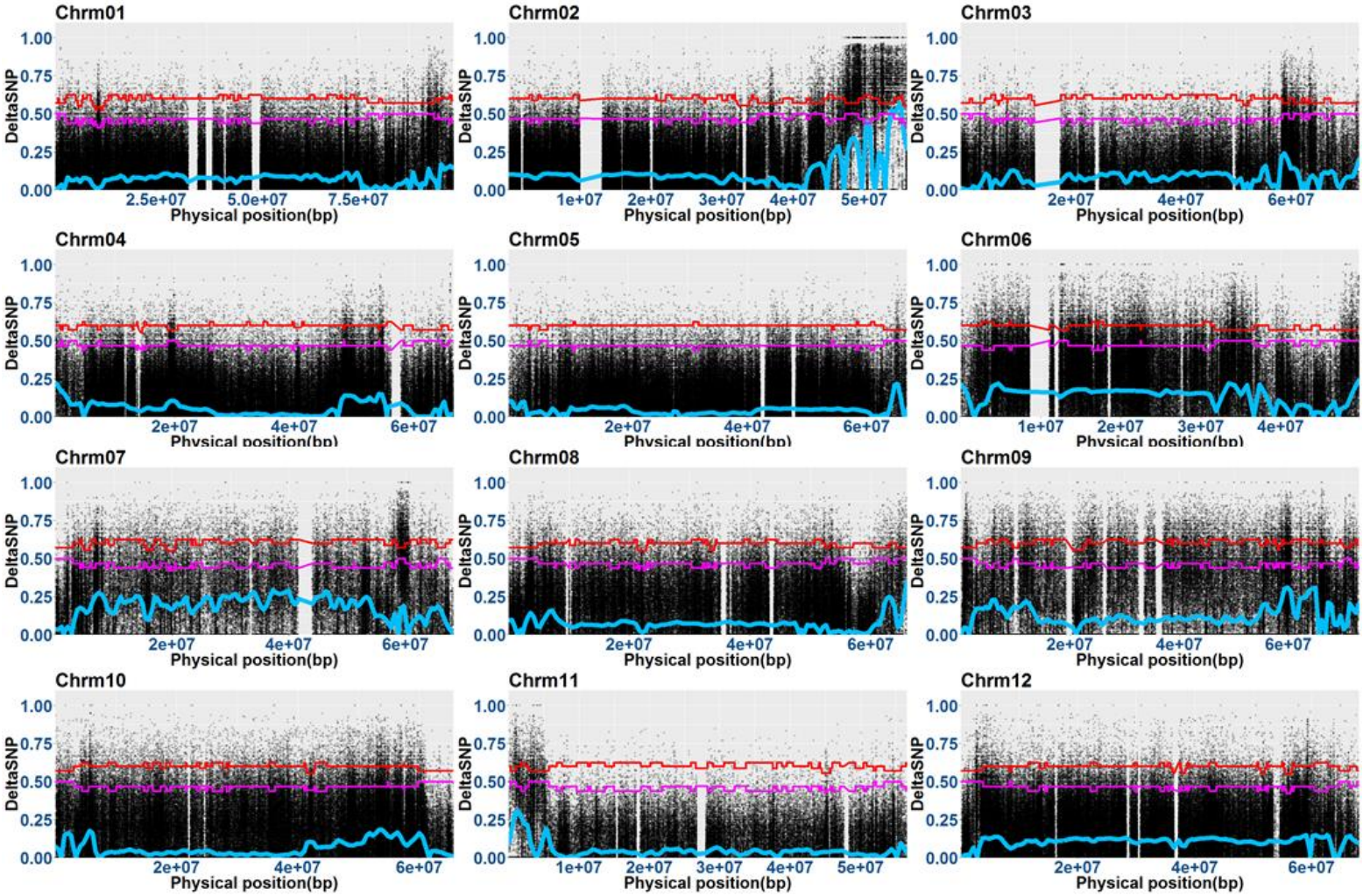
S3.4

17S62 weight



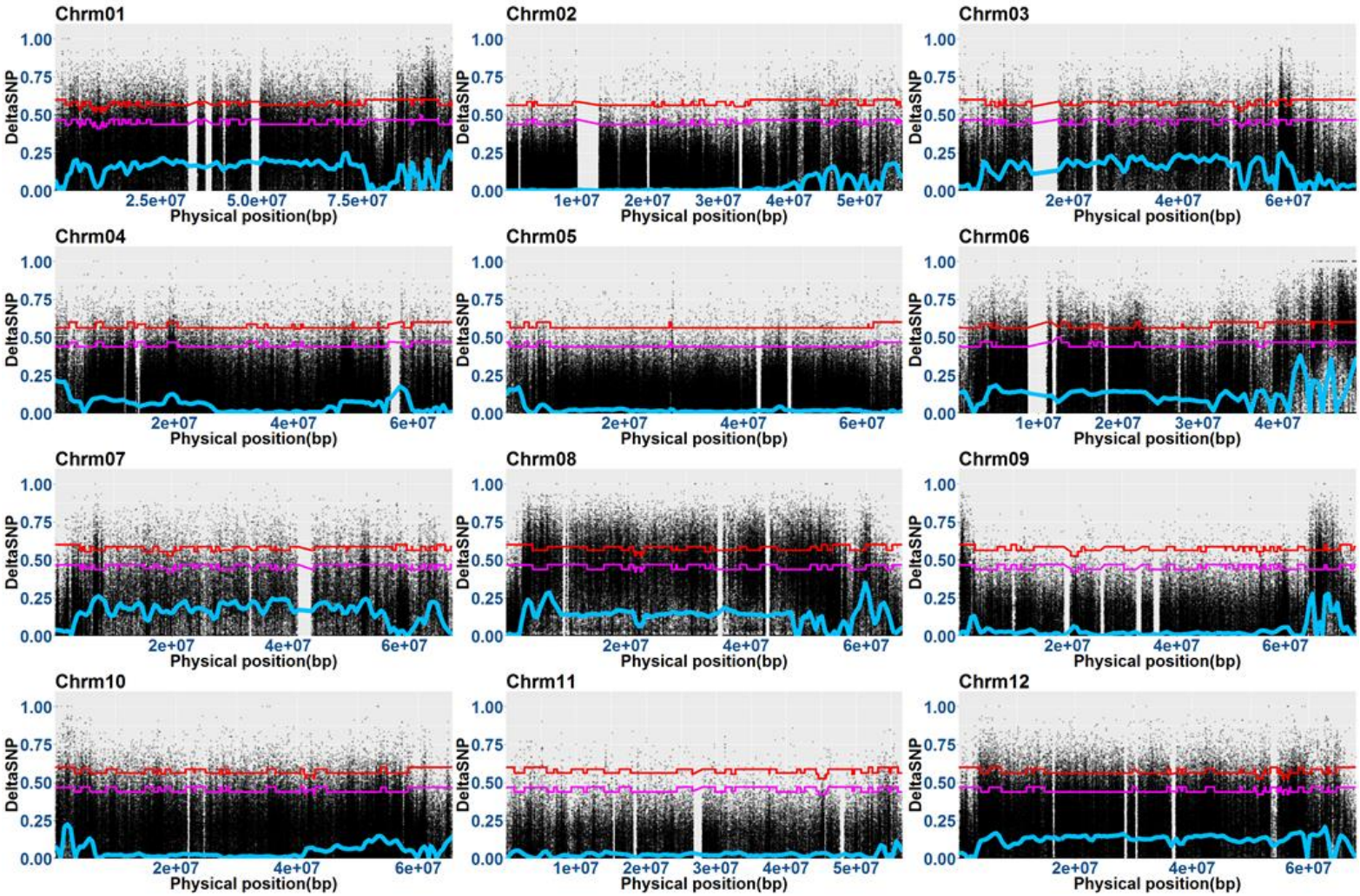
S3.5

17S64 weight



S3.6

17S64 circumference cell number 1



Supplementary figure S3.4 - S3.6 QTLseq outputs for all chromosomes for the 3 QTLseq performed. The purple and red lines are the 95 and 99 CI for the regions respectively. The blue line represents the tricube smoothed value. Tricube smoothed values higher or close to the CI were considered for further analysis.

CHAPTER 4

CONCLUSIONS

Tomato breeders need new sources of variation to create cultivars that yield larger fruits, and with this study we have taken the initial steps to achieve such goal. The two-step domestication history of tomato provided a great opportunity to mine ancestral germplasm for beneficial fruit size genes, which were likely left behind during domestication. Genotyping and extensive phenotyping of this germplasm allowed to find fruit size variation not explained by the known fruit weight genes FW2.2 (CNR), FW3.2 (KLUH), FW 11.3 (CSR), LC (WUSCHEL), and FAS (CLV3). This knowledge was utilized to create F_2 mapping populations which led to the discovery of a fruit weight QTL in chromosome 2, and a circumference cell number QTL in chromosome 6. Further work will be aimed to identify and fine map the candidate genes responsible for the differences in fruit weight. Also, our findings further support the proposed domestication of tomato and indicate that the increase in fruit size was not proportional in all parts of the fruit, where columella and placenta of the fruit enlarged disproportionally.