Regions Underlying Population Structure and the Genomics of Organ Size Determination in *Capsicum annuum*

Theresa A. Hill, Jareerat Chunthawodtiporn, Hamid Ashrafi, Kevin Stoffel, Allyson Weir, and Allen Van Deynze*

Abstract

Fruits, as an important part of the human diet, have been under strong selection during domestication. In general, continued directed selection has led to varieties having larger fruit with greater shape variation and tremendous increases in fruit mass. Common cultivated peppers (*Capsicum annuum* L.) are found in a wide range of sizes and shapes. Analysis of genetic relatedness and population structure has shown that the large-fruited, nonpungent types have reduced diversity and comprise a highly structured group. To explore this population structure, a statistical method for detecting fixation within subpopulations was applied to a set of 21 pungent and 19 nonpungent lines that represent the pepper breeding germplasm. We have identified 17 blocks within the pepper genome that are conserved among nonpungent large-fruited varieties. To determine if these regions were fixed by selection on fruit size or pungency, quantitative trait loci (QTLs) from seven studies along with capsaicin biosynthesis genes and homologs of organ size regulatory genes were mapped onto the current pepper genome assembly. Of the 17 fixed regions, 14 overlapped with fruit size or shape QTLs. There were seven putative organ size regulators and seven capsaicin biosynthetic genes within these regions. This work defines genomic regions that underly structure within the nonpungent pepper germplasm and QTLs or genes that may have been selected for during the development of large-fruited nonpungent pepper varieties.

Core Ideas

- The genomes of hot peppers versus sweet peppers were compared.
- Genomic regions that were polymorphic in hot peppers but not polymorphic in sweet peppers were detected.
- These regions overlap with quantitative trait loci (QTLs) and genes determining pepper fruit size and pungency.
- There are hotspots for fruit size QTLs in the genome.

Fruits, as an important part of the human diet, have been under strong selection during domestication. In general, continued directed selection has led to varieties having larger fruit with greater shape variation and tremendous increases in fruit mass as well as favorable flavors (Paran and van der Knaap, 2007). Pepper, *Capsicum* spp., which is important as both a vegetable and a spice crop worldwide, has perhaps more variability in fruit size and shape in cultivated types than any other crop. Sizes differ by as much as 20 cm, with great variations in length, diameter, and shape represented.


Abbreviations: CaCNR, CaCELL NUMBER REGULATOR; CaWUS, CaWUSCHEL; DPA, d post anthesis; LC, LOCULE-NUMBER; NP, nonpungent (used in the nomenclature of regions); QTL, quantitative trait locus; RIL, recombinant inbred line.
throughout the breeding germplasm (Bosland et al., 2012). As with most vegetable crops, wild progenitors of the most widely cultivated species, C. annuum, have very small fruit (about 1 cm in length). Valuable traits, such as disease resistance and drought tolerance, have been identified in semidomesticated and wild germplasm. Difficulties in introgressing complex traits such as disease resistance or drought tolerance from wild germplasm is exacerbated by the complex genetics determining commercially important fruit traits: size, yield, and quality (Chunthawodtiporn et al., 2017). Recovering the desired agronomic and fruit quality traits from wide crosses has proven to be challenging. A combinatorial approach integrating genetic, genomic, transcriptomic, and morphometric data will help unravel the molecular and genetic basis of multiple complex domestication traits that affect commercial pepper productivity, enabling the development of marker-assisted selection strategies. Understanding the loci underlying pepper domestication and crop improvement traits, and comparisons of our knowledge with other crop plants such as tomato (Solanum lycopersicum L.) will broaden our general understanding of domestication and breeding of vegetable crops.

Tomato has been a model for understanding the genetic determination of fruit size, finding both size and shape variation to be highly complex traits controlled by a large number of loci (Grandillo et al., 1999). The regulation of fruit growth is inherently complex, as it requires the integration of developmental and environmental information to ensure reproductive success. Studies of organogenesis in Arabidopsis thaliana (L.) Heynh. have identified a large number of genes acting in multiple pathways impinging on cellular proliferation, differentiation, and expansion to produce the final organ size (Fig. 1; reviewed in (Krizek, 2009; Breuninger and Lenhard, 2010; Powell et al., 2012)). Generally, cell divisions occur throughout the young organ primordia. As the organ becomes larger, a reduction in the region of dividing cells occurs until, ultimately, proliferative cell division ceases and the cells progress to differentiation and elongation. Both positive and negative regulators act on cell proliferation, the exit from cell division, and cell elongation. Mutations that increase the amount or duration of cell proliferation result in larger organ size. Morphological and genetic studies in tomato have identified a large number of QTLs for fruit size and shape traits. The six genes underlying QTLs that have been identified demonstrate the developmental and molecular mechanisms controlling tomato fruit growth are consistent with the general models for organ development derived from other species.

Fig. 1. Summary of some organ size regulators and pathways in plants. A schematic diagram showing some of the proteins known to influence cell proliferation, proliferation arrest, and cell expansion during organ development. Arrows and T-bars indicate activating and inhibiting functions, respectively. Green and red lines indicate overall promoters and inhibitors of growth, respectively. Hormones are shown in purple.
Fruit size and shape in pepper have been shown to be quantitatively inherited with complex genetics (Ben Chaim et al., 2001, 2003; Rao et al., 2003; Zygier et al., 2005). A large QTL analysis on field-grown recombinant inbred lines (RILs) derived from an interspecific C. annuum × Capsicum frutescens L. cross demonstrate the complexity of fruit size and shape variation in Capsicum (Yarnes et al., 2013). The large-fruited parent was 29% longer than the longest of the 92 F$_2$ RILs measured. There were 52 fruit size QTLs identified, found on all chromosomes with the exception of chromosomes 8 and 10. Syntenic relationships between pepper and tomato, along with genetic data, have implicated two genes, OVATE and KLUH, as fruit size and shape determinants in both pepper and tomato (Tsaballa et al., 2011; Monforte et al., 2014). Comparison between QTL studies and the identification of candidate genes in pepper has been greatly enabled by the recent development of genetic and molecular resources for pepper. The diploid Capsicum genome consists of $n = 12$ chromosomes with an estimated haploid genome size of 3.6 Gb (Arunuganathan and Earle, 1991; Moscone et al., 2003). Two C. annuum genome assemblies have become available. The two assemblies have a similar number of genes annotated, ~34–35 thousand, with overlapping and assembly specific genes placed on pseudomolecules (Kim et al., 2014; Qin et al., 2014). These genome assemblies and annotations will allow us to determine the physical positions of genes and QTLs and thus identify overlapping QTLs and candidate genes within QTL intervals (Rehrig et al., 2014).

Characterization of the genetic relatedness and diversity among cultivated C. annuum, has been performed using a diversity panel of C. annuum consisting of 21 pungent and 19 nonpungent lines (Hill et al., 2013). The lines were selected with input from breeders to specifically represent a broad range of germplasm used in commercial breeding programs. Among these lines, 6426 polymorphic markers covering 3818 unigenes were identified. An estimated threefold reduction in diversity was detected in nonpungent compared with pungent lines. An analysis of genetic relatedness and diversity using the software Structure (http://web.stanford.edu/group/pritchardlab/structure.html, accessed 15 June 2017) revealed clustering of the germplasm, which was confirmed via statistical support by principal components analysis and phylogenetic analysis. These analyses indicate a grade of decreasing genetic diversity leading to the less pungent, larger fruited types and ultimately to the large nonpungent bell (blocky) types (Hill et al., 2013). In addition, the nonpungent types comprise a highly structured population. A region corresponding to 8.7 cM, centered on the pungency locus, was fixed (36 monomorphic markers) among all nonpungent lines. Here, we present a statistical analysis of relative polymorphism between the pungent and nonpungent diversity panel lines across the entire genome, finding multiple fixed (nonpolymorphic) regions $>5$ cM among the nonpungent lines. These regions contribute to the structure of the nonpungent population and may be caused by selection or genetic drift. To determine if the fixed regions may be caused by linkage disequilibrium at loci selected for large fruit size or lack of pungency, the coincidence of fixed regions with fruit QTLs and genes regulating fruit shape and pungency was determined.

Materials and Methods

Detection of Fixed Regions in Nonpungent Lines

The C. annuum diversity panel of 40 lines consisting of 21 pungent and 19 nonpungent lines used in Hill et al. (2013) was assayed for genomic regions of reduced polymorphism among the nonpungent group using the $S$ statistic (Vaysse et al., 2011). Among the 6426 polymorphic diversity panel markers, there were 3129 marker haplotypes within 2689 unigenes present in the ‘FA’ genetic map (Supplemental Table S1) (Hill et al., 2015). The $S$ statistic was calculated from these markers for all predefined 5-cM sliding windows with a 1-cM step across the pepper genome (Vaysse et al., 2011). Windows with less than five polymorphic markers were not retained in the analysis. For nonpungent versus pungent types, relative heterozygosity as was defined as:

$$S_{np} = \frac{h_n}{h_n + h_p},$$

where $h_n$ is the number of polymorphic markers in the nonpungent group and $h_p$ is the number of polymorphic markers in the pungent group in a given 5-cM window. For each window, $S_{np}$ was then calculated as:

$$S_n = \frac{\theta_{np} - E[\theta_{np}]}{sd[\theta_{np}]},$$

where $E[\theta_{np}]$ is the expected value of $\theta_{np}$, calculated by comparing all of the markers between nonpungent and pungent groups, and $sd[\theta_{np}]$ is the SD of all sliding windows. The positions of the markers in the CM334 version 1.55 genome assembly were determined with GMAP (Wu and Watanabe, 2005).

Fruit Morphology Measurements

Fruits were collected at the breaker stage of development from at least three field-grown plants and latitudinal sections were imaged with a scanner (K13018, Canon USA, Inc., Melville, NY) (Yarnes et al., 2013). The images were analyzed with the Tomato Analyzer version 3.0 freeware (Brewer et al., 2006, Gonzalo et al., 2009). The Tomato Analyzer’s basic measurements (perimeter, area, width mid-height, maximum width, height mid-width, and maximum height) were recorded for each fruit. Statistical tests for differences in line means between pungent and nonpungent types were performed using JMP version 12.0 (SAS, Cary, NC).
Identifying Organ Size and Shape Gene Homologs in the Pepper Genome

Protein sequences for genes known to be involved in organ size and shape determination were extracted from National Center for Biotechnology Information and The Arabidopsis Information Resource databases. The protein sequences were used to query the pepper CM334 version 1.5 annotated coding DNA sequence sequences at the SolGenomics Network (https://solgenomics.net/help/tools/blast, accessed 7 June 2017) using BLASTn (Altschul et al., 1990). The longest and highest scoring hit was then used for a reciprocal search to verify that the original query was its best match. Genes identified in the pepper genome assembly and the corresponding homologs from A. thaliana or tomato are listed in Supplemental Table S2. The corresponding gene names for the published Arabidopsis logs from or tomato are listed in Supplemental Table S2. The CM334 version 1.5 gene annotations were identified using BLASTn with extracted cDNA sequences.

Mapping QTLs onto the Genome

Published fruit size and shape QTL information was gathered from studies that were amenable to mapping the QTL positions onto the pepper genome, (Table 1) (Ben Chaim et al., 2001; Rao et al., 2003; Lu et al., 2012; Mimura et al., 2012; Yarnes et al., 2013; Han et al., 2016; and unpublished data by Chunthawodtiporn et al.). Where necessary, sequences corresponding to QTL flanking markers were extracted from the literature or from the SolGenomics Network (Ben Chaim et al., 2003; Chaim et al., 2003; Lee et al., 2004; Zygier et al., 2005; Minamiyama et al., 2006; Yi et al., 2006; Nagy et al., 2007; Wu et al., 2009; Truong et al., 2010; Borovsky and Paran, 2011). The physical position of the markers was determined by querying the CM334 version 1.55 genome with marker sequences using the BLAST server at the SolGenomics Network. Positions with the top hit value were extracted. Where there were no sequences available for the QTL markers, the nearest outside flanking markers were used. For those QTLs with only a peak, one flanking marker, or both, the high density FA map was used to identify positions corresponding to FA 10-cM regions spanning the QTL peak. FA map markers were placed on the CM334 version 1.55 genome as previously described (Hill et al., 2015).

The fruit traits were standardized across studies into 10 categories: blossom end shape, fruit diameter, fruit length, fruit shape, fruit weight, longitudinal area, longitudinal perimeter, pericarp thickness, transverse area, and transverse perimeter. Within the fruit diameter category are latitudinal width and height traits derived from the Tomato Analyzer (Rodriguez et al., 2010; Yarnes et al., 2013; and unpublished data by Chunthawodtiporn et al.) and fruit width. Fruit length includes the longitudinal height traits from the Tomato Analyzer data (Yarnes et al., 2013; and unpublished data by Chunthawodtiporn et al., 2017). Genomic regions that are important for fruit size and shape were summarized by combining overlapping genomic regions within each study. The consolidation of QTLs by study is elaborated in Supplemental Table S3.

Table 1. Summary of quantitative trait locus studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rao et al. (2003)</td>
<td>C. annuum ‘Maor’ × C. frutescens ‘BG2816’</td>
</tr>
<tr>
<td>Yarnes et al. (2013)</td>
<td>C. frutescens ‘2814–6’ × C. annuum ‘NuMexRNuKy’</td>
</tr>
<tr>
<td>Mimura et al. (2012)</td>
<td>C. annuum ‘LS2341’ × C. annuum ‘California Wonder’</td>
</tr>
<tr>
<td>Ben Chaim et al. (2001)</td>
<td>C. annuum ‘Maor’ × C. annuum ‘Perennial’</td>
</tr>
<tr>
<td>Chunthawodtiporn et al. (unpublished data)</td>
<td>C. annuum ‘Maor’ × C. annuum ‘CM334’</td>
</tr>
<tr>
<td>Han et al. (2016)</td>
<td>C. annuum ‘Perennial’ × C. annuum ‘Dempsey’</td>
</tr>
<tr>
<td>Lu et al. (2012)</td>
<td>C. annuum ‘YCM334’ × C. annuum ‘Taear’</td>
</tr>
</tbody>
</table>

The regions under selection among nonpungent lines, the fruit trait QTLs and genes of interest were visualized on the pepper genome using MapChart (Voorrips 2002). In addition, regions of low recombination frequency to no recombination according to both the FA and ‘NM’ genetic maps (Hill et al., 2015) were painted on the chromosomes.

Capsaicin Biosynthetic Genes

Genes within the capsaicin biosynthesis pathway were extracted from the published CM334 version 1.5 genome (Kim et al., 2014) and the corresponding CM334 version 1.55 gene names were identified using BLASTn.

Results

Identification of Fixed Genomic Regions among Nonpungent C. annuum

In the pepper diversity study presented in Hill et al. (2013), a region was identified corresponding to 8.7 cM, centered on the pungency locus (PUN1) that was fixed among the 21 nonpungent lines assayed. To determine if there has been significant selection around PUN1, we used the S_1 statistic (Vaysse et al., 2011) applied to Dataset S2 from Hill et al. (2013), and identified two fixed regions on chromosome 2, one spanning PUN1 and the second spanning the tomato fruit shape gene ortholog CaO-VATE (Fig. 2). The S_1 test identifies blocks of the genome where one population has little or no variation relative to another, consistent with fixation of a long haplotype resulting from strong positive selection or genetic drift during domestication or subsequent improvement by breeding for specific types. This statistic will identify regions underlying the structure of the non-pungent germplasm. Using the C. annuum diversity panel, representing broad C. annuum breeding germplasm (Hill et al., 2013), the presence of additional fixed regions across the genome within nonpungent (indicated by NP) C. annuum was assayed. There were 17 regions greater than 5 cM on 9 of the 12 chromosomes that were fixed in nonpungent types (Table 2, Supplemental Fig S1, Supplemental Table S1). These regions ranged from 6 to 19 cM, spanning 6.3 to 41.0 Mb, containing 63 to 410 genes. Regions with elevated S_1 values among the non-pungent types are caused by a relatively high number of
Fig. 2. Detection of significantly fixed regions among nonpungent C. annuum on pepper chromosome 2. The profile of the fixation statistic, $S_n$, is plotted by cM position along the length of chromosome 2. A value of $-2$ represents a difference in allele frequency that is greater than 2 SDs below the expected allele frequency. The genes of interest found within fixed regions are shown below.

Table 2. Chromosomal nonpungent (NP) regions and gene content with identified organ size regulators and capsaicin biosynthetic genes.

<table>
<thead>
<tr>
<th>Region</th>
<th>Chromosome</th>
<th>Position (cM)</th>
<th>Pos (bp)</th>
<th># of genes</th>
<th>Candidate gene</th>
<th>Gene name†</th>
<th>Trait category</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP1.1</td>
<td>01</td>
<td>60.2–79.4</td>
<td>35,525,837–41,894,832</td>
<td>149</td>
<td>CaBB</td>
<td>CA01g19860</td>
<td>Organ growth and size</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CaBF9</td>
<td>CA01g18890</td>
<td>Organ growth and size</td>
</tr>
<tr>
<td>NP1.2</td>
<td>01</td>
<td>96.1–110.1</td>
<td>156,557,626–168,853,068</td>
<td>313</td>
<td>KasIIIb</td>
<td>CA01g28560</td>
<td>Capsaicin</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PUN1</td>
<td>CA02g19260</td>
<td>Capsaicin</td>
</tr>
<tr>
<td>NP1.3</td>
<td>01</td>
<td>10.2–21.9</td>
<td>226,143,117–264,420,677</td>
<td>410</td>
<td>CaGRF9</td>
<td>CA02g19160</td>
<td>Organ growth and size</td>
</tr>
<tr>
<td>NP2.1</td>
<td>02</td>
<td>49.7–57.8</td>
<td>147,649,687–153,821,541</td>
<td>330</td>
<td>CaOVATE</td>
<td>CA02g22830</td>
<td>Organ growth and size</td>
</tr>
<tr>
<td>NP2.2</td>
<td>02</td>
<td>61.6–69</td>
<td>157,150,459–159,833,718</td>
<td>159</td>
<td>4CL</td>
<td>CA03g30500</td>
<td>Capsaicin</td>
</tr>
<tr>
<td>NP3.1</td>
<td>03</td>
<td>28.5–36.9</td>
<td>15,589,344–19,212,153</td>
<td>80</td>
<td>HCT</td>
<td>CA03g31230</td>
<td>Organ growth and size</td>
</tr>
<tr>
<td>NP3.2</td>
<td>03</td>
<td>119.6–131.8</td>
<td>244,772,375–250,606,405</td>
<td>308</td>
<td>FatA</td>
<td>CA03g30250</td>
<td>Capsaicin</td>
</tr>
<tr>
<td>NP3.3</td>
<td>03</td>
<td>137.3–143.4</td>
<td>252,107,320–253,197,332</td>
<td>64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP4.1</td>
<td>04</td>
<td>35.8–45.2</td>
<td>16,968,386–58,022,200</td>
<td>272</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP4.2</td>
<td>04</td>
<td>47.6–57.2</td>
<td>177,652,216–189,238,013</td>
<td>112</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP6.1</td>
<td>06</td>
<td>2.3–8.4</td>
<td>81,769–4535,834</td>
<td>199</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP6.2</td>
<td>06</td>
<td>97.9–104.1</td>
<td>219,031,550–225,502,847</td>
<td>247</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP6.3</td>
<td>06</td>
<td>124.3–130.6</td>
<td>233,198,698–235,289,162</td>
<td>146</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP7.1</td>
<td>07</td>
<td>27.1–44.9</td>
<td>9059,522–50,078,775</td>
<td>300</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP9.1</td>
<td>09</td>
<td>9.6–20.0</td>
<td>4192,541–6684,873</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP11.1</td>
<td>11</td>
<td>25.4–42.1</td>
<td>15,134,869–39,159,124</td>
<td>234</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NP12.1</td>
<td>12</td>
<td>16.5–22.8</td>
<td>11,856,333–17,964,619</td>
<td>111</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† Gene name indicates the CM334 version 1.55 gene designation.
polymorphic markers in the nonpungent group. There were no fixed regions among the pungent types.

**Distribution of Organ Size Regulators in the Pepper Genome**

The nonpungent pepper germplasm typically consists of larger fruited types than the pungent germplasm (Paran and van der Knaap, 2007). An analysis of the diversity lines included in this study showed that, on average, pungent types had significantly smaller fruits in both length and width than nonpungent types (Fig. 3). To identify pepper genes that may play a role in organ size determination, known organ size regulators from *A. thaliana*, tomato, or both were used in reciprocal BLAST searches against the annotated pepper genes. A total of 80 proteins were identified, 73 had been placed on chromosome pseudomolecules (Supplemental Table S3). The remaining seven were annotated on unanchored scaffolds. Multiple copies for homologs of *AGL6*, *ANT*, *AP2*, *ARF2*, *ARF8*, *BB*, *BIN2*, *CYCD3*, *EBP*, *GRF1*, *MED25*, *TCP4*, and *TCP13* were identified. Seven organ size regulators fell within six of the NP fixed regions (Table 2). Among these, *CaOVATE* was the only gene that was homologous to a known fruit size or shape determination gene in tomato.

Accompanying the published genome is a survey of gene expression levels representing RNA-seq data for 24 tissues from the small fruited ‘CM334’ variety (Table 22 from Kim et al. (2014)). We extracted the expression profiles for potential organ size regulators (Supplemental Table S2). Values for CM334 pericarp tissue at 6 and 16 d post-anthesis (DPA) were reported. Expression values for *CaCELL NUMBER REGULATOR (CaCNR)*, *CaWUSCHEL (CaWUS)*, *CaSUN* and *CaKLUH* were very low during these stages of fruit development (reads per kb million < 1). Expression values for several key genes (Fig. 1) show detectable differences occurring between 6 and 16 DPA, similar to that reported for developing tomato fruit transitioning from cell division to cell expansion (Fig. 4, Supplemental Table S2). For example, expression values for the promoter of cell division, *CaCYCD3-I,1* were relatively high and decreased from 168 to 61 RPKM from 6 to 16 DPA. Expression values for organ size regulators within NP regions showed *CaYAB5* and *CaGRF9* levels decreasing and *CaUFO* and *CaARF2* levels increasing from 6 to 16 DPA, whereas *CaOVATE*, *CaROT4*, and *CaBB* levels were unchanged (Fig. 4C, D).

**Comparison of Nonpungent Fixed Regions with Pepper Fruit QTLs**

To compare the concordance of NP fixed regions with fruit size and shape QTLs, the genomic positions of QTLs from previous studies were determined. Markers from seven studies were amenable to placement of QTLs on the genome (Table 1) (Ben Chaim et al., 2001; Rao et al., 2003; Lu et al., 2012; Mimura et al., 2012; Yarnes et al., 2013; Han et al., 2016; and unpublished data by Chunthawodtiporn et al.). The QTL regions from all seven studies were grouped into overlapping regions and summarized by phenotypic categories; length, diameter, weight, shape, and pericarp thickness, with length including longitudinal perimeter and area, and width including transverse perimeter and area (Fig. 5, Fig. 6, Supplemental Fig. S2). There were 33 such regions. Diameter or a combination of diameter and weight traits were the most represented (15). Together, these genetic studies indicate fruit size determinants on all 12 pepper chromosomes (Fig. 5, Fig. 6, Supplemental Fig. S2, Supplemental Table S3). The lowest density of coverage was on chromosome 5. There was a notable concentration of QTLs and organ size regulators at the bottom of chromosomes 2 and 3 (Fig. 5, Fig. 6). The bottom of chromosome 2 had 14 organ size gene homologs and QTLs for weight, length, diameter, and shape; the bottom of chromosome 3 contained nine organ size gene homologs and QTLs for weight, length, diameter, and pericarp thickness. These regions contained homologs of the tomato fruit size

![Fig. 3.](image-url) Fruit size among pungent and nonpungent peppers. Nonpungent varieties were significantly larger in perimeter, height, and width than the nonpungent varieties (*, p < 0.05; **, p < 0.01; ***, p < 0.001).
LOCULE-NUMBER (LC), OVATE, CNR, and KLUH, as well as two NP fixed regions on each chromosome. There were 14 of the 17 NP fixed regions that overlapped with fruit size or shape QTLs, 10 of which overlapped with multiple QTLs.

**Capsaicin Biosynthetic Genes within NP Fixed Regions**

A search for genes involved in capsaicin production within the NP fixed regions was also performed. In addition to the capsaicin synthase locus, *PUN1* on chromosome 2, six genes encoding enzymes in the capsaicin biosynthetic pathway were identified within four of the NP fixed regions on chromosomes 1, 3, and 6 (Table 2, Supplemental Table S2) (Kim et al., 2014). These included enzymes in the two major pathways producing substrates for capsaicin synthase (Fig. 7) (Stewart et al., 2005; Mazourek et al., 2009; Aza-Gonzalez et al., 2011; Han et al., 2013). Three enzymes are members of the phenylalanine pathway (C4H, 4CL, and HCT) and three are in the valine pathway (KasIIIb, FatA, and KR). A comparison of the reported expression values (Kim et al., 2014) for these genes in the placenta during fruit development shows significant differences in expression between the pungent CM334 and nonpungent blocky type *C. annuum* 'Early Cal Wonder' (Supplemental Fig S3). Genes in the phenylalanine pathway all displayed reduced expression at 25 DPA in the nonpungent variety, when capsaicin levels are increasing in pungent peppers (Iwai et al., 1979). Of the QTL studies for the fruit traits presented here, Yarnes et al. (2013) also mapped the loci responsible for capsaicin levels, finding a total of 12 capsaicin QTLs. One of the Yarnes et al. capsaicin QTLs, 6.8 (233,387,238–236,837,905 bp) on chromosome 6, overlapped with a nonpungent fixed region, NP6.3. This QTL interval spanned *FatA*, at 234,247,861 bp.
Discussion

Size and shape variation in fruit is determined by complex genetic factors, making them difficult to breed for. In this study, we leverage the genetic structure in commercially grown pepper populations (*C. annuum*) that include a wide variety of fruit types with diverse sizes, shapes, and pungency to identify the genetic loci that may control these traits.

Structure within *Capsicum annuum* Breeding Germplasm

The small diversity analysis performed here showed 17 regions of the genome that are nonpolymorphic (fixed) among the nonpungent pepper breeding germplasm in *C. annuum*. It is important to note that the lines were specifically chosen by breeders from multiple commercial breeding programs worldwide to represent the base germplasm of cultivated *C. annuum*, and thus the panel does not represent a random sample. Structure analysis of the same set of lines shows that the nonpungent lines appear to be drawn from a small common genetic pool (Hill et al., 2013). Therefore, these fixed regions may be caused by linkage disequilibrium from selection for traits common to the nonpungent types or genetic drift. Supporting this, selection for the nonpungency locus has resulted in a fixed region of 8 cM around the *pun1* allele. The nonpungent types share additional traits such as short internode length, vase-shaped architecture, large leaves, and large fruit. Fruit measurements of our diversity lines indicates diversifying selection on fruit size between pungent and nonpungent breeding lines, with pungent types having smaller fruits.

Fig. 5. A representation of pepper chromosome 2 with fruit-related features. The pepper chromosome 2 (Chr02) marked at 5-Mb intervals (gray ticks) with the Mb ruler is shown at the left. Organ development genes are indicated with positions on the chromosome shown as black lines. Quantitative trait loci (QTLs) and nonpungent (NP) fixed regions are indicated by colored bars with the associated traits shown at the right. Note the density of the organ size gene homologs, QTLs, and NP fixed regions at the bottom of the chromosome.
relative to nonpungent types. To identify fixed regions that may correlate with fruit size, shape determination, or both among nonpungent breeding varieties, overlap with known organ size regulators and fruit size and shape QTLs was determined. Of 17 NP fixed regions, 15 showed an overlap with homologs of organ size regulators, and pepper fruit size QTLs, and shape QTLs.

Physical Overlap of Pepper Fruit Size and Shape QTLs from Multiple Studies

Fruit size and shape QTLs from seven published studies were mapped onto the CM334 pseudomolecules (Ben Chaim et al., 2001; Rao et al., 2003; Lu et al., 2012; Mimura et al., 2012; Yarnes et al., 2013; Han et al., 2016; and unpublished data by Chunthawodtiporn et al.). Fruit diameter was the most complex trait, having the largest number of loci by phenotypic classification. A total of 33 QTL regions, clustered by overlap and phenotypic category, were identified in this study. Of these, 15 cosegregated with diameter, 13 with length and diameter, three with length, one with pericarp thickness, and one with shape. Of the 35 summarized QTLs from the Yarnes et al. (2013) study, which used a high density genetic map, there were 23 QTLs for diameter, seven for length, and three for length and diameter with the phenotypic variation explained by a single QTL ranging from 6 to 27%. These results indicate that width determination may be more complex, involving more loci than length. Additionally, the dense map used in Yarnes et al. (2013) appears to have enabled both the detection of additional QTLs with shorter intervals and separation of length and diameter QTLs. High-density genetic maps for detecting fruit size QTLs will be critical for dissecting QTL-dense regions.

Fig. 6. A representation of pepper chromosome 3 with fruit-related features. The pepper chromosome 3 (Chr03) marked at 5-Mb intervals (gray ticks) with the Mb ruler is shown at the left. Organ development genes are indicated with positions on the chromosome shown as black lines. Quantitative trait loci (QTL) and nonpungent (NP) fixed regions are indicated by colored bars with associated traits shown at the right. Note the clusters of organ size gene homologs and a high density of QTLs at the bottom of the chromosome. The legend is the same as that of Fig. 5.
genes was identified in the pepper genome. Although CaOVATE and CaKLH have been associated with fruit shape and weight, respectively, in pepper (Tsaballa et al., 2011; Chakrabarti et al., 2013), no fruit size or shape determinants have been verified or characterized.

**CaOVATE** was found on chromosome 2, where 13 additional organ size regulators and multiple fruit QTLs have been mapped. These additional genes included the homologs of the tomato fruit regulators *LC* and *CNR*. Three QTLs, *Chuntha2.1, Mimura2.2*, and *Yarnes2.2*, span all three of the tomato homologs with a maximum of 27% of the variation in length explained. A fourth

**Known Fruit Size and Shape Genes Overlapping Fruit QTLs**

Our understanding of the genetics of fruit size and/or shape determination in tomato has yielded the isolation of six genes (Monforte et al., 2014). These genes include *LC* and *FASCIATED*, which encode WUSCHEL and CLAVATA3, respectively; and *FW2.2* and *FW3.2*, which encode CNR and SIKLUH, respectively; and *SUN* and *OVATE* (Cong et al., 2002; van der Knaap et al., 2002; Xiao et al., 2009; Muños et al., 2011; Rodriguez et al., 2011; Chakrabarti et al., 2013; Xu et al., 2015). A single copy of each of the six known tomato fruit size and shape genes was identified in the pepper genome.

---

**Fig. 7.** Proposed capsaicin biosynthetic pathway. The two pathways downstream of the amino acids phenylalanine and valine lead to the production of capsaicin. Enzymes found within nonpungent fixed regions are highlighted with yellow. The Capsaicin Synthase (CS) enzyme is at the PUNGENCY locus, encoded by the PUN1 gene [adapted from Aza-Gonzalez et al. (2011)].
QTL, Chaim2.2, including both CaOVATE and CaWUS, explains up to 17% of the variation in diameter. The syntenic region in tomato is also spanned by multiple QTLs (Grandillo et al., 1999; Gonzalo et al., 2009). Pepper QTLs in this region cosegregate with weight, length, diameter, and shape. The location of Han2.2, Chaim2.1, NP2.1, Rao 2.3, Yarnes2.3, and Yarnes2.4 intervals do not include OVATE, LC, or CNR, pointing to additional genes in this region as fruit size and shape determinants. Taken together, this evidence indicates multiple linked genes, including as yet unknown genes, regulating fruit size and shape in this region.

A second region with a high density of QTLs and organ size regulators was observed on the bottom of chromosome 3. One of the genes within this region was CaKLH. A previous study showed that CaKLH cosegregated with fruit size in a C. frutescens × C. annuum interspecific RIL population (Chakrabarti et al., 2013). This was consistent with CaKLH lying within the Yarnes3.5 interval. In addition, CaKLH was within Chuntha3.1, Han3.3, and Rao3.3. It seems unlikely that this is the only gene regulating pepper fruit size in this region. There are at least two discrete QTL regions and Ben Chaim et al. (2001) reports four nonoverlapping QTLs in this region for multiple fruit traits. The QTLs reported for one of the regions, spanning Chaim3.4, were specific to weight or length, with up to 19% of the variation in length accounted for here. The bottom of chromosome 3 is syntetic with tomato. However, the tomato chromosome 3 region corresponding to the pepper chromosome 3 fruit length region lacks any major tomato QTLs. This suggests there may be differences in major fruit length determinants between pepper and tomato.

The remaining known organ size and shape regulators in tomato are CLAVATA3 and SUN. CaCLAVATA3 resides at the bottom of chromosome 6, spanned only by the Chuntha6.1 QTL region. There were no QTLs identified on the bottom of chromosome 10 that overlapped with the CaSUN gene. This region is orthologous to the bottom of tomato chromosome 10, where the ancestral copy of tomato SUN is located (Wu et al., 2009). However, the SUN locus is located on tomato chromosome 7, resulting from a retrotransposon-mediated duplication of chromosome 10 SUN in the tomato lineage. This duplication probably resulted in elevated SUN levels and elongated fruit. There is no indication that SUN has been duplicated in Capsicum, considering the current genome assemblies. The lack of QTLs at chromosome 10 SUN suggest this is not a major gene determining shape in the biparental pepper populations assayed thus far.

**Genes Regulating Organ Size Overlapping NP Fixed Regions**

Of the six tomato genes, only the tomato fruit shape gene OVATE was found under a NP fixed region. Expression levels of CaOVATE, were very low (<2 RPKM) at both 6 and 16 DPA. The expression values were determined from the pericarp of the small, elongate-fruited semidomesticated CM334 lines. Low expression values have been observed for CaOVATE small elongate peppers versus higher expression in small round peppers after anthesis supporting the hypothesis that natural variation at the CaOVATE locus may be a determinant of shape variation in C. annuum (Tsaballa et al., 2011).

In addition to CaOVATE, the organ size regulators BB, GRF9, UFO, ARF2-i2, YABS, and ROT2 were found within NP fixed regions. These genes have not yet been identified as determinants of fruit size variation in crop plants. The dynamic profiles for several of the identified organ size regulatory genes support a role in pepper fruit development, with expression changes indicative of the transition from cell proliferation to cell elongation occurring between 6 and 16 DPA. In general, promoters of cell proliferation were expressed at higher levels at 6 DPA during the proliferation phase, whereas negative regulators of proliferation or promoters of cell expansion were expressed at higher levels at 16 DPA.

The putative promoters of cell proliferation, CaYAB5 and CaGRF9, had higher levels of expression at 6 DPA. Arabidopsis thaliana yab and grf mutants both display narrower leaves (Kim et al., 2003; Sarojam et al., 2010; Kim and Tsukaya, 2015). CaGRF9 within NP1.2 was also within QTLs for weight, length, and diameter from multiple studies. CaYAB5, within the NP7.1 interval, was within the Lu7.1 QTL for pericarp thickness. The putative negative regulator of cell division, CaARF2-i2, has increasing levels of expression from 6 to 16 DPA. Arabidopsis thaliana ARF2 directly interacts with the ANT promoter, inhibiting the activity of ANT, a promoter of cell proliferation (Czesnick and Lenhard, 2015). CaARF2-i2 was within NP3.2 and the QTLs Chuntha3.1 and Yarnes3.6 for both fruit length and width.

Both CaBB and CaROT4 had very low levels of expression (< 2 RPKM) at both 6 and 16 DPA. Their A. thaliana homologs are both negative regulators of proliferation that act at the boundary between cell proliferation and cell expansion (Czesnick and Lenhard, 2015). This specific phase and location of activity may account for the low levels of expression observed at discrete time points in whole pericarp tissues. CaBB was close to CaGRF9, where there are QTLs for weight, length, and diameter, whereas CaROT, in the NP9.1 region is not within any QTL intervals from these studies. The A. thaliana UFO has been implicated as a negative regulator of organ expansion (Liu et al., 2002; Cheng et al., 2013). CaUFO showed a slight increase in expression from 6 to 16 DPA in the small-fruited CM334. CaUFO was found within NP2.1 where there were QTLs for fruit diameter, length, and weight. The general expression patterns of these primary gene candidates under NP fixed regions are consistent with function during fruit development. Integrating emerging sequencing data and higher-resolution genetic studies will help us to identify fruit size determinants in the Capsicum germplasm.
Capsaicin Biosynthesis Genes Overlapping NP Fixed Regions

The NP2.1 region contained the pungency locus, PUN1, encoding capsaicin synthase, which is required for the production of capsaicin (Stewart et al., 2005, Han et al., 2013). Lines most closely related to the nonpungent group in the phylogeny of breeding lines tend to be less pungent, suggesting there may be fixation around capsaicin biosynthesis loci in the nonpungent types as a result of common ancestry (Hill et al., 2013). There are six additional genes within four NP regions that are involved in capsaicin biosynthesis. These include genes in both the phenylalanine (CAH, 4CL, and HCT) and the valine (KasIIIa, FatA, and KR) pathways (Mazourek et al., 2009). A link with capsaicin production for three genes was made by Lee et al. (2016), who reported that the closely linked genes (KR, CAH, and FatA) were within the capsaicin QTL qcap6.1, overlapping NP6.2 and NP6.3. This QTL was identified in a population derived from the highly pungent Capsicum chinense Jacq. ‘Bhut Jolokia’ and a moderately pungent chili pepper, C. annuum ‘NB1’. The QTL analysis by Yarnes et al. (2013), using a population derived from a small-fruited highly pungent type and a large-fruited mildly pungent type along with a high density genetic map found a narrow QTL, within NP6.3, spanning only FatA. Molecular evidence suggests one or more genes in the valine pathway may have causal mutations affecting capsaicin biosynthesis. These genes showed differential expression profiles in the placenta of developing fruit between pungent and nonpungent varieties, with significantly elevated levels of expression for KasIIIa, FatA, and KR in pungent CM334 during the period of capsaicin accumulation, 20 to 40 DPA (Iwai et al., 1979). These data support the phylogenetic and population structure analyses showing that the nonpungent types share common genetics with larger, less pungent C. annuum pepper varieties and were the result of selection for reduced pungency and larger fruit size (Paran and van der Knaap, 2007; Hill et al., 2013). The concordance between the QTL mapping studies, expression data, and the diversity study presented here provide additional support that the NP fixed regions have a functional significance and perhaps that changes in the FatA gene, within NP6.3, directly affect pungency levels.

Conclusion

Our understanding of the development of fruit has been largely dependent on the analysis of segregating populations in tomato, yielding the isolation of six genes associated with fruit size and shape (Monforte et al., 2014). Our study complements these efforts by systematically studying the genetic regulation and molecular candidates of fruit size and shape regulation in pepper. Similarly, the regulation of capsaicin in pepper is just beginning to be understood. Unlike an association study designed and analyzed to detect QTLs after correcting for population structure, this analysis detected loci responsible for the population structure (Hill et al., 2013). As shown, the population size needed to detect these relatively large effects can be much smaller. Understanding the genetic and physiological mechanisms controlling fruit size and capsaicin synthesis will increase our understanding of organ development in Solanaceae and guide hypothesis-driven breeding. The 17 regions identified in this study that are conserved in the nonpungent, large-fruited types would explain the difficulties in recovering fruit size and important horticultural traits in wide crosses with bell pepper types. Markers in these regions may help select for genomic regions that are important for recovering commercially viable cultivated bell pepper types.

Supplemental Information

Table S1. Genotype calls among nonpungent and pungent lines for markers with genetic map positions and physical positions.

Table S2. Physical positions of identified organ growth and capsaicin biosynthetic gene homologs mapped to CM334 version 1.55 chromosome pseudomolecules.

Table S3. Summary of the QTL mapped to the CM334 version 1.55 genome.

Figure S1. Detection of regions under significant selection among nonpungent C. annuum.

Figure S2. A representation of pepper chromosomes with fruit related features.

Figure S3. Expression profiles of capsaicin biosynthesis genes within nonpungent fixed regions.

Conflict of Interest Disclosure

The authors declare that there is no conflict of interest.

References


