Full title: Whole genome re-sequencing analysis of two tomato genotypes for polymorphism insight in cloned genes and a genetic map construction

Vladimir Cambiaso[1], Guillermo Raúl Pratta[1], Javier Hernán Pereira da Costa[1], Roxana Zorzoli[2], David Merril Francis[3], Gustavo Rubén Rodríguez[1*]

[1]Instituto de Investigaciones en Ciencias Agrarias de Rosario (IICAR-CONICET-UNR). Campo Experimental Villarino, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, S2125ZAA Zavalla, Argentina.

[2]Consejo de Investigaciones de la Universidad Nacional de Rosario. Cátedra de Genética, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, S2125ZAA Zavalla, Argentina.

[3]Department of Horticulture and Crop Sciences, The Ohio State University, OARDC, 1680 Madison Ave. Wooster, OH 44691, USA

*Corresponding author:

Gustavo Rubén Rodríguez.

Instituto de Investigaciones en Ciencias Agrarias de Rosario (IICAR-CONICET-UNR). Campo Experimental Villarino, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, S2125ZAA Zavalla, Argentina.Tel.: +54 - 341 – 4970080 ext. 1162

Fax: +54 - 341 - 4970199

E-mail: grodrig@unr.edu.ar

Vladimir Cambiaso ORCID: 0000-0002-9326-1570

Guillermo Raúl Pratta ORCID: 0000-0002-3682-0946

Javier Hernán Pereira da Costa ORCID: 0000-0002-9406-6988

David Merril Francis ORCID: 0000-0003-2016-1357

Gustavo Rubén Rodríguez ORCID: 0000-0002-4171-4099

27 # Abstract

28 Next generation sequencing technologies have become affordable for most plant breeding programs.
29 In this study we sequenced the entire genome of the *Solanum lycopersicum* L. cultivar Caimanta and
30 *S. pimpinellifolium* L. accession LA0722 with assembly relative to the Heinz 1706 reference version
31 SL2.50. We present 1) analysis of the amount and distribution of polymorphism in "Caimanta" and
32 "LA0722", 2) examination of alleles in candidate genes affecting disease resistance, fruit shape, fruit
33 weight and fruit quality and 3) development of molecular markers to construct a genetic linkage map
34 based on a F_2 population. A total of 1,397,518 polymorphisms were detected in the comparison
35 between "Caimanta" and "LA0722". A resistant allele for *Rx4/*Xv3 was detected by sequence, and
36 confirmed through inoculation. We developed a set of insertion/deletion (InDel) DNA markers that
37 can be multiplexed and scored using easily accessed genotyping platforms. These markers were used
38 to construct a genetic linkage map. We demonstrate that the whole genome sequencing of parental
39 lines can be successfully used to reveal phenotypes and characterize a reference population through
40 easily accessed genotyping strategies.

41

42 **Keywords**: InDel markers; linkage map; next generation sequencing; *Solanum* spp.; variant calling.

43

44 # 1. Introduction

45 Tomato (*Solanum lycopersicum* L.) is one of the most effective model crop systems due to the short
46 generation time, small genome size and available genetic and genomic resources (Giovannoni, 2004).
47 Wild tomato species have been extensively used in breeding programs as sources of disease resistance
48 and to adapt cultivars to diverse production areas (Blanca et al., 2015; Sim et al., 2011). The feasibility
49 of improving tomato fruit quality through interspecific crosses has been demonstrated (Fridman et al.,
50 2004; Pratta et al., 1996; Rick, 1973; Zorzoli et al., 1998). In populations derived from biparental
51 crosses, the construction of a genetic linkage map provides a reference, facilitates the discovery of

52 quantitative trait *loci* (QTL), and delivers tools for marker-assisted selection (Collard et al., 2005). The
53 first high-density linkage map in tomato was published in 1992 by Tanksley et al. and was based on 67
54 F_2 plants from an interspecific cross between the *S. lycopersicum* L. cultivar VF36-Tm2[a] and the *S.*
55 *pennellii* Correll accession LA716. The map had over 1,000 Restriction Fragment Length
56 Polymorphism (RFLP) markers and reached a total length of 1,276 cM. Since then, intra (Saliba-
57 Colombani et al., 2000) and interspecific (Gonzalo and van der Knaap, 2008) maps have been
58 constructed in order to study fruit quality and fruit shape as well as agronomically relevant traits like
59 fruit weight and yield. In 2012 Sim et al. generated high-density maps for interspecific F_2 populations
60 based on a genotyping array of 7,720 Single Nucleotide Polymorphisms (SNP): EXPEN2012 and
61 EXPIM2012. To date, more than 15 maps are available through the SOL Genomics Network (SGN)
62 database (http://www.solgenomic.net). The construction of high-density linkage maps and the ease of
63 detecting sequence polymorphisms has facilitated the fine-mapping localization of many genes in the
64 genome and the identification of alleles by positional cloning (Causse et al., 2016).
65 A reference tomato genome was published (The Tomato Genome Consortium, 2012) based on the
66 sequence assembly of the inbred *S. lycopersicum* L. cultivar Heinz 1706 consisting of 760 megabases
67 (Mb) from a predicted 900 Mb genome. In the same study a draft genome and *de novo* assembly of
68 739 Mb for the *S. pimpinellifolium* L. accession LA1589 was presented along with a comparison of the
69 two accessions including a list of detected SNP and InDel (Insertion/Deletions) polymorphisms. InDel
70 polymorphisms are the second most abundant form of sequence variation in the genome. The
71 relatively simple and inexpensive technical and equipment resources demanded for InDel DNA marker
72 development and genotyping represent an accessible strategy for breeding programs that do not have
73 access to high-throughput genotyping based on SNP markers (Yang et al., 2014). In order to develop
74 InDel markers two different strategies could be pursued. As already was described (Yang et al., 2014)
75 makers could be developed based on the list of polymorphic InDel detected between "Heinz 1706" and
76 "LA1589", and these polymorphisms can be tested in new biparental populations under study. A
77 second strategy is to sequence the entire genomes of the parental genotypes and to develop InDel
78 markers based on the variant discovery within the new sequence resources.

79 Populations with different genetic structures derived from the cross between the *S. lycopersicum* L.
80 cultivar Caimanta and the *S. pimpinellifolium* L. accession LA0722 were promised to improve both
81 fruit quality and fruit shelf life (Pratta et al., 2003; Rodriguez et al., 2006; Pereira da Costa et al 2013).
82 The aim of this study was to apply next generation sequencing technologies to characterize a reference
83 population derived from a biparental cross of “Caimanta” and “LA0722” as a framework for breeding
84 purposes. We describe a comparison of whole genome sequence between both parental lines with a
85 focus on the amount and distribution of polymorphism. We conducted a further examination of
86 polymorphism in known genes affecting disease resistance, fruit shape, fruit weight and fruit quality.
87 Finally we developed a set of molecular markers based on two different strategies and constructed a
88 genetic linkage map.
89

90 2. Material and Methods

91 2.1. Plant Material

92 The *S. lycopersicum* L. cultivar Caimanta was developed in the late seventies at the Instituto Nacional
93 de Tecnología Agropecuria (INTA) Experimental Station at Cerrillos, Salta, Argentina. The complete
94 breeding scheme is presented in Fig S1. . The *S. pimpinellifolium* L. accession LA0722 was collected
95 in 1959 at Trujillo, La Libertad, Perú and was provided by Tomato Genetic Resources Center (Davis,
96 California). Both materials were maintained by several selfing generations at the experimental field
97 J.F. Villarino, FCA-UNR, located at Zavalla, Santa Fe, Argentina (lat. 33°S, long. 61°W).

98 2.2. Genome sequencing, variant calling and polymorphism
99 distribution

100 Genomic DNA of “Caimanta” and “LA0722” were extracted from young leaves stored at -80°C using
101 a commercial Kit (Wizard® Genomic DNA Purification Kit from Promega®, Madison, WI, USA).
102 Both DNA samples were sequenced at the Genome Technology Access Center (GTAC) facility at

103 Washington University (St Louis, MO, USA) and were pooled into the same lane on the Illumina
104 Hiseq 2500 to obtain 101 base pair (bp) paired-end reads.
105 The quality of FASTQ files were evaluated using the FASTQC program version 0.11.4 (Andrews,
106 2010). The sequence reads were trimmed and filtered for quality. Bowtie 2 with the option "—very-
107 sensitive-local" (Langmead and Salzberg, 2012) was used to align both sequences to the *S.*
108 *lycopersicum* L. cultivar Heinz 1706 reference assembly version SL2.50
109 (https://solgenomics.net/organism/Solanum_lycopersicum/genome). The resulting files were sorted,
110 labeled and converted to BAM files using Picard software version 1.119
111 (http://picard.sourceforge.net). The duplicate records were located using Picard. Around insertion or
112 deletions a local re-alignment was done using GATK version 3.2-2 (DePristo et al., 2011; McKenna et
113 al., 2010). The resulting BAM files were analyzed with Qualimap version 2.0.2 (García-Alcalde et al.,
114 2012). SNP and InDel calling were done using the HaplotypeCaller tool from GATK version 3.2-2
115 (DePristo et al., 2011; McKenna et al., 2010). The sequence data generated in this study have been
116 deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive
117 (SRA) under the accession number SRP128767.
118 Variant calling files were used to calculate the SNP and InDel density of "Caimanta" and "LA0722"
119 relative to the Heinz 1706 reference across the entire genome. Then, further variant calling files were
120 used to compare SNP and InDel variation between "Caimanta" and "LA0722". The visualization of
121 the genetic distance and the relatedness among "Caimanta", "LA0722" and "Heinz 1706" was
122 achieved by a Principal Component Analysis (PCA) performed with the R statistical software
123 environment version 3.1.1 (R Core Team, 2014). A similarity matrix based on the proportion of the
124 total base pairs number shared by each genotype relative to the Heinz 1706 reference was done. The
125 whole-genomic variations stored in variant calling files were plotted using the web based visualization
126 tool CircosVCF (Drori et al. 2017) .
127

128 ## 2.3. Sequence variation polymorphism in cloned genes

129 Examination and visualization of polymorphism in cloned genes was also conducted. Specific genes
130 affecting disease resistance, fruit shape, fruit weight and fruit quality were examined and compared in
131 the sequences of “Caimanta” and “LA0722”. The genomic sequences of cloned genes were extracted
132 for “Caimanta”, “Heinz 1706” and “LA0722”, and were compared to allele sequences available in the
133 NCBI database by a multiple sequence comparison methodology.
134 Details on the sequences used to evaluate all genes were summarized in Supplementary Table S1. The
135 length of the sequences and the target region under analysis for each cloned genes was based on
136 previous studies and available sequence data. Disease resistance genes *TM2*, *Rx4/Xv3*, *VE-1*, and *VE-2*
137 were evaluated and the presence of specific resistance and or discover novel alleles genes were done
138 based on sequences reported in previous studies (Fradin et al., 2009; Kawchuk et al., 2001;
139 Lanfermeijer et al., 2005; Pei et al., 2012). For genes affecting fruit shape, the *OVATE* and *LC*
140 sequences were evaluated. A visual inspection of the alignment surrounding a single sustitution that
141 results in an early stop codon in alleles associated with elongated fruits (Liu et al., 2002; Rodríguez et
142 al., 2011) and two single-nucleotide polymorphisms responsible for increasing locule number (Muños
143 et al., 2011) were specifically inspected in the alignment. For genes affecting fruit size, we examined
144 sequences from *FW2.2*, *ORF44*, and *FW3.2* obtained from previous studies (Chakrabarti et al., 2013;
145 Nesbitt and Tanksley, 2002). To analyzed genes affecting fruit quality, we focused on *LIN5* and the
146 chromoplast-specific lycopene beta-cyclase. Sequences from *LIN5* were extracted from previous
147 studies (Bolger et al., 2014) and a single substitution associated with an amino acid change responsible
148 for enhancing the activity of *LIN5* (Fridman et al., 2004) was analyzed. To evaluate the chromoplast-
149 specific lycopene beta-cyclase, two distinct frame-shift mutations detected in the coding sequence of
150 old-gold and old-gold-crimson alleles responsible for deep red fruit and high lycopene (Ronen et al.,
151 2000) were examined.
152 The multiple sequence comparisons for all cloned genes were performed using log-expectation as
153 implemented using Multiple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar, 2004).

154 Cluster analysis of the distance matrix generated from MUSCLE was performed using the Ward
155 method for hierarchical clustering as implemented by the hclust function in the R statistical software
156 environment version 3.1.1 (R Core Team, 2014).
157

158 2.4. Bacterial spot race T3 allelism determined by inoculation and
159 molecular markers

160 Bacterial spot race T3 inoculations were conducted and a hypersensitive response (HR) was evaluated.
161 The *Xanthomonas perforans* race T3 strain Xcv761 was cultured on yeast, dextrose, and calcium
162 carbonate (YDC) agar medium (Lelliot and Stead, 1987) at 28°C for 48 h. Bacterial cells were
163 removed from the agar plates and suspended in sterile, double-distilled water. The suspensions were
164 standardized to A600 = 0.15 which corresponds to a concentration of approximately 3×10^8 cfu ml-1
165 based on dilution plating. Each six-week old plant was inoculated on four different leaflets by
166 infiltration of approximately 0.1 ml of a $3x10^8$ cfu ml-1 solution into the leaf surface using a syringe
167 without a needle. The line OH087663 was used as a positive control for the *Rx4/Xv3* resistance gene.
168 At least three plants were tested for "Caimanta", "LA0722", and a randomly selected sub-set of four
169 RILs developed from the cross between "Caimanta" and "LA0722" (Rodríguez et al., 2006). The
170 presence of the resistant allele in derived progeny from those parents was evaluated. DNA for each
171 parent and RILs was extracted as described above. Segregation of the putative *Rx4/Xv3* resistance
172 locus was verified using primers PCC12 as described previously (Pei et al., 2012). Symptom
173 evaluation was conducted at 24 and 48 hours post inoculation and expressed as the percentage of
174 inoculations showing a clear hypersensitive response (HR) associated with resistance.

175 2.5. Development of molecular markers and genetic linkage map
176 construction

177 Development of InDel markers was based on two different strategies. As already was described by
178 Yang et al., 2014, the first strategy was based on the published list of polymorphism between the

179 cultivar Heinz 1706 of *S. lycopersicum* L. and the accession LA1589 of *S. pimpinellifolium* L. (The
180 Tomato Genome Consortium, 2012) and the second strategy was based on the InDel calling from the
181 sequence comparison between "Caimanta" and "LA0722".
182 The InDel markers selected from the first strategy, comparing "Heinz 1706" and "LA1589", were
183 corroborated through nucleotide comparisons with BLAST® (Altschul et al., 1990) while for the
184 second strategy the candidate InDel regions were visually evaluated with IGV software version 2.3
185 (Robinson et al., 2011). Polymerase chain reaction (PCR) markers based on insertions and deletions
186 ranging from 15 bp to more than 50 bp were developed. Multiplex PCR assays for InDels were
187 developed creating sub groups including size ranges from 15 to 22 bp, 23 to 40 bp and larger than 41
188 bp. Primer design emphasized markers with an amplified fragment size of 100-200 bp (small size),
189 220-350 bp (medium size) and 400-500 bp (large size) within these groups, respectively. A maximum
190 of three pairs of primers were included in the same reaction mix. For some regions of the genomes,
191 SNPs were detected based on Cleaved Amplified Polymorphic Sequences (CAPS). The online tool
192 "CAPS Designer" from Sol Genomics Network (Available at:
193 solgenomics.net/tools/caps_designer/caps_input.pl) was used to find restriction sites around
194 polymorphic SNP between "Caimanta" and "LA0722" obtained from the SNP calling.
195 The online interface of Primer3 version 0.4.0 (Untergasser et al., 2012) was used for InDel and CAPS
196 primers design. The same standard PCR protocol was followed for the three different kinds of
197 molecular markers (Powell et al., 1996).
198 A genetic linkage map was constructed based on a population of 94 F_2 plants derived from selfed F_1
199 ("Caimanta" x "LA0722") hybrid. Genomic DNA was extracted from young leaves stored at -80°C
200 from all the F_2 plants following the same protocol noted above. Different kinds of markers were used
201 in the molecular characterization of the F_2 population: Single Sequence Repeats (SSR) tested by
202 (Pereira da Costa et al., 2013); InDel developed in this study based on the first strategy; InDel and
203 CAPS developed on the basis of parental sequence polymorphism (second strategy), and 4 functional
204 markers for fruit traits: *fas* (Rodríguez et al., 2011), *fw2.2* (Blanca et al., 2015), *lc* (Muños et al., 2011)

205 and *fw3.2* (Chakrabarti et al., 2013). Details on the molecular markers used are summarized in
206 Supplementary Table S2.
207 Electrophoresis of InDel and CAPS markers was conducted on 3% w/v agarose gels stained with
208 SYBR® Safe DNA Gel Stain (Thermo Fisher Scientific®, Waltham, MA, USA) for visualization,
209 while SSR makers were run on 6 % w/v polyacrylamide gels visualized by a silver staining procedure.
210 The R/Qtl (Broman et al., 2003) package was used to construct the linkage map in the R statistical
211 software environment version 3.1.1 (R Core Team, 2014). Markers were placed in the same linkage
212 group if they had a LOD score greater than 3.8 and an estimated recombination fraction lower than
213 0.35. The distance between markers was calculated using the Kosambi function (Kosambi, 1943). The
214 markers were set in the correct order in each linkage group with the functions "orderMarkers" and
215 "ripple". The change in chromosome length and in log likelihood dropping one marker at the time was
216 investigated with the function "droponemarker". When no recombination between markers, we used
217 the physical position to decide the order. The genetic and physical position of markers was compared.

218

219 3. Results

220 3.1. Genome sequencing, variant calling, and polymorphism
221 distribution for Caimanta and LA0722

222 From the whole genome sequencing, a total of 128,692,024 and 134,466,322 paired reads of 101 bp
223 length were obtained for "Caimanta" and the accession "LA0722", respectively. After quality control
224 and alignment against the tomato genome reference Heinz 1706 version SL2.50 an average depth of
225 coverage of 15.35 fold for Caimanta and 15.79 fold for LA0722 were obtained. The depth coverage
226 across the entire genome and the standard deviation for the 12 chromosomes for both accessions is
227 presented in Supplementary Fig. S2. The graphical depth of coverage comparison from both genotypes

228 reveals some regions in common with extremely high or low coverage. The presence of these regions
229 generated the high standard deviations detected.
230 Polymorphisms were analyzed between "Caimanta" and "LA0722" and the reference genome as well
231 as between "Caimanta" and "LA0722". A total of 65,950 polymorphisms were detected across the
232 entire genome between "Caimanta" and the reference Heinz 1706, while 1,153,384 polymorphisms
233 were detected between "LA0722" and the reference. Fig. 1a and 1b show the total number of SNP and
234 InDel detected for "Caimanta" and "LA0722" relative to the reference Heinz 1706 and also shared
235 between them. Fig. 1c shows the relatedness among "Caimanta", "LA0722" and "Heinz 1706". The
236 first and the second principal component (PC1 and PC2) explained the 99.89% and the 0.11% of the
237 total variation, respectively. The PC1 differentiated the cultivated genotypes, "Caimanta" and "Heinz
238 1706" from the wild accession "LA0722", while the PC2 differentiated between the cultivated
239 genotypes. From the comparison between "Caimanta" and "LA0722" 1,397,518 polymorphisms were
240 detected. Table 1 a summarizes number of SNP and InDel variants detected by chromosome relative to
241 the reference Heinz 1706 whereas Table 1 b details the number of polymorphisms between
242 "Caimanta" and "LA0722" by chromosome. The maximum number of SNP and InDel between
243 "Caimanta" and the reference were detected for Chromosome 4, while the minimum number of SNP
244 and InDel were detected for chromosome 6 and 8, respectively. Surprisingly, for chromosomes 1, 3
245 and 6 InDel polymorphisms were more frequent than SNPs. From the comparison between "LA0722"
246 and the reference the maximum and minimum number of SNP and InDel were detected in
247 chromosome 8 and 3, respectively. From the comparison between "Caimanta" and "LA0722" the
248 maximum number of SNP and InDel polymorphism was found for chromosome 7 and 1, respectively.
249 The minimum number of both, SNP and InDel polymorphism was found on chromosome 3.
250 Polymorphisms are visualized as density plots for SNP and InDel. SNP and InDel density plots
251 obtained through the comparison of "Caimanta" and "LA0722" to the reference are displayed in Fig.
252 2. This Figure also provides an integrative view of the polymorphism across the entire genome
253 detected between "Caimanta" and "LA0722" in relation to the reference genome. Unsurprisingly,
254 "LA0722" has a higher level of polymorphism than "Caimanta" when compared to the reference

255 genome. Some regions with an extremely low rate of polymorphism were found for both genomes at
256 the top of chromosome 2 and in the middle of chromosome 3.

257 3.2. Sequence variation polymorphism in cloned genes

258 Ten known genes affecting disease resistance, fruit shape, fruit weight and fruit quality were analyzed
259 to confirm expected phenotypes and explore new alleles.
260 Fig. 3 presents alignment-based clusters and the results of the visual inspection of specific mutations
261 for *Rx4/Xv3* (Fig. 3a) and *LC* (Fig. 3b) genes. For disease resistance genes, the analysis correctly
262 aggregated the susceptible and resistant genotypes for all cases except for *VE-2*. *VE-2* lacked
263 informative polymorphism, and therefore phenotypic expectations are only based on *VE-1*. For genes
264 affecting fruit shape, fruit weight and fruit quality it was necessary to perform a visual inspection of
265 specific mutations associated to the gene function. The multiple sequence comparison of *TM2* alleles
266 demonstrated a 100% of base pairs shared between "Caimanta", "Heinz 1706" and the susceptible *tm-*
267 *2* allele at the sequence region studied (Supplementary Fig. S3). The sequence for "LA0722" showed
268 99.96% in common with *tm-2*, and polymorphisms indicated a novel allele clustering close to
269 susceptible alleles. For the *Rx4/Xv3* candidate gene, "Caimanta" and "Heinz 1706" shared a 100% of
270 the analyzed sequences with the susceptible elite processing tomato line OH88119. In contrast
271 "LA0722" shared a 99.90% with the resistant allele sequences found in "PI128216" and
272 "Hawaii7981" (Fig. 3a). The visual inspection of the detected polymorphisms determined the presence
273 of the 6-bp InDel associated with resistance (Fig. 3a) (Pei et al., 2012). With respect to the *VE-1*
274 sequences "Caimanta" shared a 99.97% with the susceptible cultivars evaluated. In contrast
275 "LA0722" showed a higher percentage in common with alleles from the resistant cultivars (99.78; S3
276 Fig.). The visual inspection of the reported deletion at the position 1,220 that creates a premature stop
277 codon resulting in truncated protein of 407 amino acids was found in all susceptible cultivars and
278 "Caimanta". "LA0722" carries the key deletion distinguishing resistant varieties from susceptible.
279 The cluster analysis with the sequences of *OVATE* grouped both *S. lycopersicum* cultivars together and
280 the wild accession LA0722 remained separate (Supplementary Fig.S3). The reported functional

281 mutation for *OVATE* is present in "Heinz 1706" and absent in "Caimanta" and "LA0722". The
282 multiple sequence comparison of *LC* shows that the three *S. lycopersicum* cultivars clustered together
283 while the wild accession separate (Fig. 3b). Visual inspection of the two single-nucleotide
284 polymorphisms reported to be responsible for increasing locule number determined that "Caimanta"
285 and "Levovil" carry the mutant allele that produces fruit with many locules, while "Heinz 1706",
286 "LA0722" and "Cervil" have the wild-type allele that produces fruits with mostly two locules (Fig.
287 3b). Comparison of *FW2.2* sequence demonstrate that the three large-fruited *S. lycopersicum* cultivars
288 grouped together (>99.98% sequences in common). Similarly, the two small-fruited accessions of *S.*
289 *pimpinellifollium* shared a 99.70% of the region under study (Supplementary Fig.S3). The small-
290 fruited *S. pennelli* accession LA0716 remained separated in the cluster analysis. The analysis of the
291 polymorphism underlying the functional mutation for *fw2.2* reveals that the three small fruit
292 accessions share the wild-type allele, while the three large fruits cultivars share the large-fruited
293 cultivated allele. At *FW3.2 S. lycopersicum* cultivars were identical and the *S. pimpinellifolium*
294 accessions presented more than 99.68% in common for both sequence fragments. Visual inspection of
295 the most significantly associated SNP (substitution of G by A), shows that the two small fruit
296 accessions share the wild-type small-fruited allele (G), while the cultivars share the large-fruited allele
297 (A).
298 For genes that potentially affect fruit quality, clusters reflected species origin of alleles rather than
299 functional mutations (Supplementary Fig.S3). The multiple sequence comparison showed that a
300 mutation characterized as responsible for enhancing the activity of *LIN5* was only present in the *S.*
301 *pennellii* accesion LA0716 (Supplementary Fig.S3). "Caimanta", "Heinz 1706" and "LA0722"
302 possess different alleles likely associated with reduced BRIX relative to the "LA0716" allele. For the
303 og^c sequence analysis, "LA0722", "Heinz 1706" and "Caimanta" clustered together while the
304 Genebank accession no. AF254793 remained separate (Supplementary Fig.S3). The visual inspection
305 of two distinct frame-shift mutations (Ronen et al., 2000) were used to determine that the three
306 genotypes have functional chromoplast-specific beta-cyclase associated with low lycopene relative to
307 the "crimson" mutations.

3.3. Bacterial spot race T3 resistance confirmation by inoculation and molecular markers

The evaluation of the resistance gene *Rx4/Xv3* in the sequence of “Caimanta” and “LA0722” suggest that “Caimanta” is susceptible while “LA0722” may provide resistance to bacterial spot race T3. Table 2 presents the results of the inoculations and the segregation of the putative *Rx4/Xv3* resistance locus verified using primers PCC12 (Pei et al., 2012) for “Caimanta”, “LA0722”, and a sub-set of four RILs. Both evaluations confirmed the susceptibility of “Caimanta” and the resistance of “LA0722” previously predicted by the sequence information. The presence of the resistant allele was also revealed in derived progeny. RILs L8 and L9 predicted to be resistant showed symptoms HR in 100% of the inoculated leaflets after 24 hours. The other two RILs, L1 and L14 appeared to be susceptible in the genotypic analysis and showed water soaking symptoms in at least 75-100% of the inoculated leaf after 48 hours.

3.4. Development of molecular markers and genetic linkage map construction

For marker development, the discovered InDel were clustered into four groups according to their size in bp (Table 1). Those with polymorphism ≤ 14 bp (difficult to be distinguished in 3% w/v agarose gel); 15–22 bp (used to develop small size markers); 23–40 bp (used to develop medium size markers) and finally those with polymorphism ≥ 41 bp (used to develop large size markers). For all the chromosomes the vast majority of the InDel were shorter than 15 bp. Nevertheless, hundreds of InDels larger than 15 bp were detected for all chromosomes, providing many opportunities to develop fragment-size variation markers.

Predicted polymorphisms between the cultivar Heinz 1706 and the accession LA1589 (The Tomato Genome Consortium, 2012) were used to identify 52 InDels for molecular marker development. On the other hand, 126 InDel were developed based on the detected polymorphism between “Caimanta” and “LA0722”. Only five predicted markers were monomorphic and five failed to amplify by standard

333 PCR protocols. This second strategy had a 92% success rate and the distribution of the detected
334 polymorphism was even across the entire genome. A total of 45 multiplex PCR were designed, 36
335 with three markers each and nine with two markers. In all cases, at least two of three markers included
336 in the same multiplex amplify correctly. In 16 cases (44.4%) all three markers were amplified and
337 scored. Supplementary Fig. S4 shows a 3% w/v agarose gel electrophoresis following multiplex PCR
338 for three InDel markers.
339 The entire F_2 population was characterized with 185 molecular markers: 24 SSR, 156 InDel, 1 CAPS
340 and 4 functional markers for fruit size and shape (Supplementary Table S2). A total of 157 markers
341 were mapped onto 12 linkage groups corresponding to the 12 chromosomes. A total of 28 markers
342 (15%) were excluded from the analysis because: 1) distorted segregation (15 markers); 2) dominance
343 of markers (five markers); 3) more than 15% missing data (five markers); and 4) extreme changes in
344 chromosome length and in log likelihood caused when testing quality by dropping one marker at the
345 time (three markers). Four markers at the top of chromosome 11 present a distorted segregation and a
346 biased transmission in direction of the wild progenitor. These markers were not excluded from the
347 map because the whole region presents a distorted segregation. The 15 markers with a distorted
348 segregation that were excluded from the analysis, were distributed in chromosomes 1, 2, 3, 4, 5, 9 and
349 12, and were not grouped together in any specific region of those chromosomes. The genetic map is
350 shown in Fig. 4. The total length of the linkage map was 1,495 cM with an average distance between
351 markers of 10.3 cM and a maximum spacing of 43.8 cM. Table 3 summarizes the number of markers,
352 the length, the average spacing and the maximum spacing in cM by chromosome. Due to the lack of
353 polymorphism detected on the top of chromosome 2 and in the middle of chromosome 3 (graphically
354 presented on Fig. 2), only markers below the physical position 20,190,400 bp for chromosome 2 and
355 between 6,017,080 bp and 54,701,833 bp for chromosome 3 were developed. The genetic and physical
356 positions of all markers were consistent for all chromosomes (Fig. 4; Supplementary Table S2).

357 # 4. Discussion

358 Sequencing technologies to key parents for polymorphism discovery, insight into specific alleles and
359 creation of a reference genetic map was applied. The resources are based on a biparental cross between
360 an Argentinian fresh market *S. lycopersicum* L. cultivar, Caimanta, and the *S. pimpinellifolium* L.
361 accession LA0722 which has been used as a donor of fruit quality traits (Pratta et al., 2003). Both
362 parental genotypes have been sequenced and aligned against the tomato genome reference. Previous
363 comparison of the *S. pimpinellifolium* accession LA1589 draft genome and the cultivar Heinz 1706
364 reference genome found a total of 5.4 million SNPs (The Tomato Genome Consortium, 2012). In this
365 study we found only 18% of this number SNPs when comparing "LA0722" with "Heinz 1706"
366 (906,360 SNPs) and "LA0722" with "Caimanta" (1,081,626 SNPs). The fewer SNPs identified in this
367 study may reflect methodological differences in the approach. We performed alignment for "LA0722"
368 against the *S. lycopersicum* reference and not a *de novo* assembly as was performed to obtain the *S.*
369 *pimpinellifolium* LA1589 draft genome. In our approach there are unmapped reads against the genome
370 references where SNPs cannot be called. The fewer number of SNPs detected in this study could be
371 also due to difference in the stringency of SNP and INDEL calling and difference between both *S.*
372 *pimpinellifolium* accessions. After assigning the genomic DNA sequence contigs of "LA1589" to
373 "Heinz 1706" only 146,695 InDels were identified (Yang et al., 2014). In this study we detected
374 247,024 InDels between "LA0722" and "Heinz 1706" and 315,892 InDels between "LA0722" and
375 "Caimanta". The lower number of InDels detected when comparing "Heinz 1706" with *S.*
376 *pimpinellifollium* may reflect the introgressions of this wild species on chromosome 4, 9, 11 and 12
377 used to create the compact habit, fruit shape and small fruit core that distinguish processing tomatoes
378 from fresh market tomatoes (The Tomato Genome Consortium, 2012).
379 The whole genome sequence comparison provided information about the amount and distribution of
380 genetic variation. From the SNP and InDel density plots two large regions with an extremely low
381 polymorphism have been revealed, one in the top of the chromosome 2 (from the physical position
382 20,190,400 bp) and the other one in the middle of chromosome 3 (between 6,017,080 bp and

383 54,701,833 bp). These regions may represent genomic introgressions from wild species conserved in
384 cultivated genomes due to the contribution of these regions to desirable characteristics that have been
385 selected during the domestication or the breeding process. Alternatively, they could represent regions
386 with high levels of repetitive sequence affecting alignment and mapping, such as the nucleolar
387 organizing region on chromosome 2.

388 The sequence data also allowed us to inspect alleles at specific genes affecting disease resistance, fruit
389 shape, fruit size and fruit quality. Multiple sequence comparisons between our sequence and control
390 sequences revealed several features relevant as breeding goals. The lack of polymorphism detected
391 between “Caimanta” and “Heinz 1706” for the region on chromosome 9 where *TM2* is located,
392 suggested a lack of introgression for resistance. With respect to *TM2*, we expected “Caimanta” to be
393 resistant based on pedigree and the release notice
394 (https://www.inase.gov.ar/consultaGestion/gestiones, no 1237). However, “Caimanta” clearly
395 possesses the sequence of the susceptible allele, suggesting introgression of *Tm2a* as a breeding target.
396 Similarly, “Caimanta” carries the sequences of the susceptible *Rx4/Xv3* and *Ve* allele. We verified the
397 function of the *Rx4/Xv3* allele from “LA0722”.

398 Examination of genes affecting fruit characteristics was consistent with expectations based on
399 pedigree. Fruit shape and fruit weight alleles in “Caimanta” are all consistent with expectations, based
400 on “Caimanta’s” large multi-loculed fruit. Examination of the fruit quality genes suggests sugar
401 content could be improved by introgression of the LA0716 *LIN5* allele. Improved BRIX with the
402 LA0716 *LIN5* allele is thought to be through increased translocation of sucrose driving unloading into
403 the sink fruit (Fridman et al., 2004). The Michaelis constant (Km[sucrose]) for the LA0716 allele
404 suggests improved hydrolysis of sucrose into fructose and glucose. Thus improved BRIX are imparted
405 by both loading and osmotic changes associated with hydrolysis (Fridman et al., 2004). The crimson
406 frame-shift mutations, *old gold* and *old gold crimson*, are causal for high lycopene content and deep
407 red color desired in some markets. At the same time, improvement in lycopene comes at a cost to the
408 nutritionally desirable carotenoid beta carotene. “Caimanta” contains a functional locus associated

409 with wild-type cultivated varieties, and modification toward high lycopene or high beta carotene
410 would depend on market demand.
411 The sequence comparison between “Caimanta” and “LA0722” increased the success in DNA marker
412 development to 92%. The ability to use sequence data to optimize multiplexing strategies decreased
413 the time, effort and supplies spent on genotyping. Consistent with previous findings, the InDel
414 genotyping was an effective strategy for a breeding program that lacks access to high-throughput SNP
415 platforms (Yang et al., 2014). InDel abundance and distribution across the entire genome provided
416 sufficient markers. The molecular characterization of the F_2 population with 157 molecular markers
417 allowed us to construct a genetic linkage map with a total length of 1,495 cM, an average distance
418 between markers of 10.3 cM and a maximum spacing of 43.8 cM. The further potential to use the
419 markers and map for marker-assisted selection seems likely, especially given the potential for
420 “LA0722” to provide resistance missing from “Caimanta” (this study) and fruit characteristics
421 identified previously (Pratta et al., 1996; Zorzoli et al., 1998). In order to saturate specific regions of
422 interest more molecular markers can be developed based on polymorphisms detected during the whole
423 genome comparison. The potential to use our data to develop makers for intraspecific crosses is also
424 high since InDel between “Caimanta” and “Heinz 1706” were abundant across the entire genome
425 (24,220 InDel; Table 1).
426 We detected some segregation distortion which appears to be consistent with other F_2 populations
427 derived from interspecific crosses between *S. lycopersicum* L and *S. pimpinellifollium* L. (Gonzalo and
428 van der Knaap, 2008; Lippman and Tanksley, 2001; Robbins et al., 2011). In our population,
429 segregation distortion and a biased transmission in direction of the wild progenitor were detected on
430 chromosome 11. In chromosomes 1, 2, 3, 4, 5, 9 and 12 some markers displayed distorted segregation
431 but these were not grouped together in any specific region of those chromosomes.

432

433 # 5. Conclusions

434 In this study we demonstrated the utility of whole genome sequencing from parental lines as a
435 resource to verify alleles in genes controlling parental phenotypes, measuring variation across the
436 genome, and characterizing reference populations through easily accessed genotyping strategies. We
437 detected a high level of polymorphism between the parental lines distributed across the entire genome.
438 We found and confirmed a resistant allele for *Rx4/Xv3* that is already present in derived populations,
439 and have evidence for the presence of a second resistance, the *VE-1* allele from "LA0722". Finally, we
440 developed a set of molecular makers and constructed a linkage map as a genetic reference for QTL
441 detection and validation and also to perform marker-assisted selection. The resources developed will
442 be useful for both interspecific and intraspecific tomato populations.

443

444 # Acknowledgments

445 We thank Tomato Genetic Resources Center, University of California, Davis, CA for kindly providing
446 seeds of *S. pimpinellifolium* accession LA0722 and Estación Experimental Agropecuaria – Instituto
447 Nacional de Tecnología Aagropecuaria EEA Cerrillos, Salta, Argentina for kindly proving seeds of
448 Caimanta cultivar. We thank the BEC.AR Programme for granting a scholarship to VC. We thank to
449 Saranga Wijeratne at the Molecular and Cellular Imaging Center, Ohio Agricultural Research and
450 Development Center, The Ohio State University (http://oardc.osu.edu/mcic/) for the bioinformatics
451 support.

452

453 # References

454 Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search

455 tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
456 Andrews, S., 2010. FastQC a quality control tool for high throughput sequence data. [WWW
457 Document]. URL http://www.bioinformatics.babraham.ac.uk/ projects/fastqc/
458 Blanca, J., Montero-Pau, J., Sauvage, C., Bauchet, G., Illa, E., Díez, M.J., Francis, D., Causse, M., van
459 der Knaap, E., Cañizares, J., 2015. Genomic variation in tomato, from wild ancestors to
460 contemporary breeding accessions. BMC Genomics 16, 257. https://doi.org/10.1186/s12864-015-
461 1444-1
462 Bolger, A., Scossa, F., Bolger, M.E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh, S.,
463 Sørensen, I., Lichtenstein, G., Fich, E. a, Conte, M., Keller, H., Schneeberger, K., Schwacke, R.,
464 Ofner, I., Vrebalov, J., Xu, Y., Osorio, S., Aflitos, S.A., Schijlen, E., Jiménez-Goméz, J.M.,
465 Ryngajllo, M., Kimura, S., Kumar, R., Koenig, D., Headland, L.R., Maloof, J.N., Sinha, N., van
466 Ham, R.C.H.J., Lankhorst, R.K., Mao, L., Vogel, A., Arsova, B., Panstruga, R., Fei, Z., Rose,
467 J.K.C., Zamir, D., Carrari, F., Giovannoni, J.J., Weigel, D., Usadel, B., Fernie, A.R., 2014. The
468 genome of the stress-tolerant wild tomato species Solanum pennellii. Nat. Genet. 46, 1034–1038.
469 https://doi.org/10.1038/ng.3046
470 Broman, K.W., Wu, H., Sen, S., Churchill, G.A., 2003. R/qtl: QTL mapping in experimental crosses.
471 Bioinformatics 19, 889–890. https://doi.org/10.1093/bioinformatics/btg112
472 Causse, M., Giovannoni, J., Bouzayen, M., Zouine, M., 2016. The Tomato Genome. Springer-Verlag
473 GmgH Germany, Berlin. https://doi.org/10.1007/978-3-662-53389-5
474 Chakrabarti, M., Zhang, N., Sauvage, C., Muños, S., Blanca, J., Cañizares, J., Diez, M.J., Schneider,
475 R., Mazourek, M., McClead, J., Causse, M., van der Knaap, E., Knaap, E., 2013. A cytochrome
476 P450 regulates a domestication trait in cultivated tomato. Proc. Natl. Acad. Sci. U. S. A. 110,
477 17125–17130. https://doi.org/10.1073/pnas.1307313110
478 Collard, B.C.Y., Jahufer, M.Z.Z., Brouwer, J.B., Pang, E.C.K., 2005. An introduction to markers,
479 quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The
480 basic concepts. Euphytica 142, 169–196. https://doi.org/10.1007/s10681-005-1681-5
481 DePristo, M.A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J.R., Hartl, C., Philippakis, A. a, del

482 Angel, G., Rivas, M. a, Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko,
483 A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D., Daly, M.J., 2011. A framework for variation
484 discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–8.
485 https://doi.org/10.1038/ng.806
486 Drori, E., Levy, D., Rahimi, O., 2017. Genome analysis CircosVCF : circos visualization of whole-
487 genome sequence variations stored in VCF files 33, 1392–1393.
488 https://doi.org/10.1093/bioinformatics/btw834
489 Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput.
490 Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340
491 Fradin, E.F., Zhang, Z., Juarez Ayala, J.C., Castroverde, C.D.M., Nazar, R.N., Robb, J., Liu, C.-M.,
492 Thomma, B.P.H.J., 2009. Genetic dissection of verticillium wilt resistance mediated by tomato
493 Ve1. Plant Physiol. 150, 320–332. https://doi.org/10.1104/pp.109.136762
494 Fridman, E., Carrari, F., Liu, Y., Fernie, A.R., Zamir, D., 2004. Zooming in on a quantitative
495 interspecific introgressions. Science (80-.). 305, 1786–1790.
496 https://doi.org/10.1126/science.1101666
497 García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Götz, S., Tarazona, S., Dopazo, J.,
498 Meyer, T.F., Conesa, A., 2012. Qualimap: Evaluating next-generation sequencing alignment
499 data. Bioinformatics 28, 2678–2679. https://doi.org/10.1093/bioinformatics/bts503
500 Giovannoni, J.J., 2004. Genetic regulation of fruit development and ripening. Plant Cell 16, 170–180.
501 https://doi.org/10.1105/tpc.019158
502 Gonzalo, M.J., van der Knaap, E., 2008. A comparative analysis into the genetic bases of morphology
503 in tomato varieties exhibiting elongated fruit shape. Theor. Appl. Genet. 116, 647–56.
504 https://doi.org/10.1007/s00122-007-0698-7
505 Kawchuk, L.M., Hachey, J., Lynch, D.R., Kulcsar, F., van Rooijen, G., Waterer, D.R., Robertson, A.,
506 Kokko, E., Byers, R., Howard, R.J., Fischer, R., Prufer, D., 2001. Tomato Ve disease resistance
507 genes encode cell surface-like receptors. Proc. Natl. Acad. Sci. U. S. A. 98, 6511–6515.
508 https://doi.org/10.1073/pnas.091114198

509 Kosambi, D.D., 1943. The estimation of map distances from recombination values. Ann. Eugen. 12,
510 172–175. https://doi.org/10.1111/j.1469-1809.1943.tb02321.x
511 Lanfermeijer, F.C., Warmink, J., Hille, J., 2005. The products of the broken Tm-2 and the durable Tm-
512 22 resistance genes from tomato differ in four amino acids. J. Exp. Bot. 56, 2925–2933.
513 https://doi.org/10.1093/jxb/eri288
514 Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–
515 359. https://doi.org/10.1038/nmeth.1923
516 Lelliot, R.A., Stead, D.E., 1987. Methods for the Diagnosis of Bacterial Diseases of Plants. Blackwell
517 Scientific Publication Ltd., Oxford, UK.
518 Lippman, Z., Tanksley, S.D., 2001. Dissecting the genetic pathway to extreme fruit size in tomato
519 using a cross between the small-fruited wild species Lycopersicon pimpinellifolium and L.
520 esculentum var. Giant Heirloom. Genetics 158, 413–22.
521 Liu, J., Van Eck, J., Cong, B., Tanksley, S.D., 2002. A new class of regulatory genes underlying the
522 cause of pear-shaped tomato fruit. Proc. Natl. Acad. Sci. U. S. A. 99, 13302–6.
523 https://doi.org/10.1073/pnas.162485999
524 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K.,
525 Altshuler, D., Gabriel, S., Daly, M., DePristo, M. a, 2010. The genome analysis toolkit: a
526 mapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20,
527 1297–303. https://doi.org/10.1101/gr.107524.110
528 Muños, S., Ranc, N., Botton, E., Bérard, A., Rolland, S., Duffé, P., Carretero, Y., Paslier, M.-C. Le,
529 Delalande, C., Bouzayen, M., Brunel, D., Causse, M., 2011. Increase in tomato locule number is
530 controlled by two single-nucleotide polymorphisms located near WUSCHEL. Plant Physiol. 156,
531 2244–2254. https://doi.org/10.1104/pp.111.173997
532 Nesbitt, C.T., Tanksley, S.D., 2002. Comparative sequencing in the genus lycopersicon: Implications
533 for the evolution of fruit size in the domestication of cultivated tomatoes. Genetics 162, 365–379.
534 Pei, C., Wang, H., Zhang, J., Wang, Y., Francis, D.M., Yang, W., 2012. Fine mapping and analysis of
535 a candidate gene in tomato accession PI128216 conferring hypersensitive resistance to bacterial

536 spot race T3. Theor. Appl. Genet. 124, 533–542. https://doi.org/10.1007/s00122-011-1726-1
537 Pereira da Costa, J.H., Rodríguez, G.R., Pratta, G.R., Picardi, L.A., Zorzoli, R., 2013. QTL detection
538 for fruit shelf life and quality traits across segregating populations of tomato. Sci. Hortic.
539 (Amsterdam). 156, 47–53.
540 Powell, W., Machray, G.C., Proven, J., 1996. Polymorphism revealed by simple sequence repeats.
541 Trends Plant Sci. 1, 215–222. https://doi.org/10.1016/1360-1385(96)86898-1
542 Pratta, G.R., Zorzoli, R., Picardi, L.A., 2003. Diallel analysis of production traits among domestic,
543 exotic and mutant germplasms of Lycopersicon. Genet. Mol. Res. 2, 206–13.
544 Pratta, G.R., Zorzoli, R., Picardi, L.A., 1996. Evaluación de caracteres de interés agronómico en
545 especies del género Lycopersicon. Hortic. Argentina 15, 25–32.
546 R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for
547 Statistical Computing, Vienna, Austria [WWW Document]. URL http://www.r-project.org/
548 Rick, C.M., 1973. Potential genetic resources in tomato species: clues from observations in native
549 habitats. Genes, Enzym. Popul. 255–269. https://doi.org/10.1007/978-1-4684-2880-3_17
550 Robbins, M.D., Sim, S.-C., Yang, W., Van Deynze, A., Van Der Knaap, E., Joobeur, T., Francis,
551 D.M., 2011. Mapping and linkage disequilibrium analysis with a genome-wide collection of
552 SNPs that detect polymorphism in cultivated tomato. J. Exp. Bot. 62, 1831–1845.
553 https://doi.org/10.1093/jxb/erq367
554 Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P.,
555 2011. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26. https://doi.org/10.1038/nbt.1754
556 Rodríguez, G.R., Muños, S., Anderson, C., Sim, S.-C., Michel, A., Causse, M., Gardener, B.B.M.,
557 Francis, D., van der Knaap, E., 2011. Distribution of SUN, OVATE, LC, and FAS in the tomato
558 germplasm and the relationship to fruit shape diversity. Plant Physiol. 156, 275–85.
559 https://doi.org/10.1104/pp.110.167577
560 Rodríguez, G.R., Pratta, G.R., Zorzoli, R., Picardi, L.A., 2006. Recombinant lines obtained from an
561 interspecific cross between lycopersicon species selected by fruit weight and fruit shelf life. J.
562 Am. Soc. Hortic. Sci. 131, 651–656.

563 Ronen, G., Carmel-Goren, L., Zamir, D., Hirschberg, J., 2000. An alternative pathway to beta -
564 carotene formation in plant chromoplasts discovered by map-based cloning of Beta and old-gold
565 color mutations in tomato. Proc. Natl. Acad. Sci. 97, 11102–11107.
566 https://doi.org/10.1073/pnas.190177497
567 Saliba-Colombani, V., Causse, M., Gervais, L., Philouze, J., 2000. Efficiency of RFLP, RAPD, and
568 AFLP markers for the construction of an intraspecific map of the tomato genome. Genome 43,
569 29–40. https://doi.org/10.1139/gen-43-1-29
570 Sim, S.-C., Deynze, A. Van, Michel, A.P., Francis, D.M., 2011. Population structure and genetic
571 differentiation associated with breeding history and selection in tomato (Solanum lycopersicum
572 L .). Heredity (Edinb). 106, 927–935. https://doi.org/10.1038/hdy.2010.139
573 Sim, S.-C., Durstewitz, G., Plieske, J., Wieseke, R., Ganal, M.W., Van Deynze, A., Hamilton, J.P.,
574 Buell, C.R., Causse, M., Wijeratne, S., Francis, D.M., 2012. Development of a large SNP
575 genotyping array and generation of high-density genetic maps in tomato. PLoS One 7, e40563.
576 https://doi.org/10.1371/journal.pone.0040563
577 Tanksley, S.D., Ganal, M.W., Prince, J.P., de Vicente, M.C., Bonierbale, M.W., Broun, P., Fulton,
578 T.M., Giovannoni, J.J., Grandillo, S., Martin, G.B., 1992. High density molecular linkage maps
579 of the tomato and potato genomes. Genetics 132, 1141–60.
580 The Tomato Genome Consortium, 2012. The tomato genome sequence provides insights into fleshy
581 fruit evolution. Nature 485, 635–41. https://doi.org/10.1038/nature11119
582 Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G., 2012.
583 Primer3-new capabilities and interfaces. Nucleic Acids Res. 40, 1–12.
584 https://doi.org/10.1093/nar/gks596
585 Yang, J., Wang, Y., Shen, H., Yang, W., 2014. In silico identification and experimental validation of
586 insertion-deletion polymorphisms in tomato genome. DNA Res. 21, 429–38.
587 https://doi.org/10.1093/dnares/dsu008
588 Zorzoli, R., Pratta, G.R., Picardi, L.A., 1998. Efecto de los mutantes nor y rin y de genes silvestres
589 sobre características del fruto en Lycopersicon. Mendeliana 13, 12–19.

590

591 # Tables

592 **Table 1.** Number of SNP and InDel by chromosome (Ch) detected for the *Solanum lycopersicum* L.
593 cv. Caimanta and the *S. pimpinellifolium* L. accession LA0722 related to reference Heinz 1706 version
594 SL2.50 (a) and the comparison between them (b)

595

a) **b)**

	Caimanta		**LA0722**		**Caimanta vs LA0722**				
Ch	**SNP**	**InDel**	**SNP**	**InDel**	**SNP**	**INDEL[1]**			
						≤14bp	**15-22bp**	**23-40bp**	**≥41bp**
1	1,908	2,397	107,189	31,706	128,278	38,741	1,002	579	173
2	5,911	2,609	34,984	10,793	50,688	15,161	441	238	76
3	1,863	2,066	31,780	9,804	44,480	14,151	351	188	75
4	6,850	2,761	61,588	17,566	101,058	27,069	632	331	107
5	3,339	1,746	115,552	29,646	126,888	33,544	749	492	205
6	667	1,354	59,357	16,591	64,121	18,507	559	350	110
7	2,216	1,567	119,471	30,857	131,624	34,444	852	520	197
8	2,253	1,248	120,626	31,921	130,590	35,715	877	587	218
9	3,475	2,045	70,907	18,672	79,521	21,845	569	346	123
10	2,638	1,709	79,128	20,477	87,771	24,198	625	382	138
11	4,431	2,196	61,145	16,323	69,948	19,694	490	305	110
12	6,179	2,522	44,633	12,668	66,659	19,112	410	217	87
Total	**41,730**	**24,220**	**906,360**	**247,024**	**1,081,626**	**302,181**	**7,557**	**4,535**	**1,619**

596 [1] The detected number of InDel from the comparison between "Caimanta" and "LA0722" was
597 clustered by size in base pairs (bp).

598

599 **Table 2.** Confirmation of the hypersensitive response (HR) to bacterial spot race T3 (*Xanthomonas*
600 *perforans*) controlled by the *Rx4/Xv3* gene predicted by sequence comparison analysis

Genotypes[a]	Sequence[b]	PCC12[c]	First evaluation (%)[d]	Second evaluation (%)[e]
Caimanta	Susceptible	Susceptible	0	0
LA0722	Resistant	Resistant	100	100
OH087663	Resistant	Resistant	100	100
L1	-	Susceptible	0	0
L8	-	Resistant	100	75
L9	-	Resistant	100	83
L14	-	Susceptible	0	0

601

602 [a] L1, L8, L9, L14 and L18 are derived RILs from the cross between the *Solanum lycopersicum* L. cv.
603 Caimanta and the *S. pimpinellifolium* L. accession LA0722 developed by Rodriguez et al. (2006)
604 [b] Response predicted by sequence comparison analysis
605 [c] Segregation of the putative *Rx4/Xv3* resistance locus verified using primers PCC12 (Pei et al. 2012)
606 [d] Symptom evaluation conducted at 24 hours post inoculation expressed as percentage of inoculated
607 leaf showing a clear HR response associated with resistance.
608 [e] Symptom evaluation conducted at 48 hours post inoculation expressed as percentage of inoculated
609 leaf showing a clear HR response associated with resistance.
610

611 **Table 3.** Summary statistics for F_2 "Caimanta" x "LA0722" map including number of markers, length,

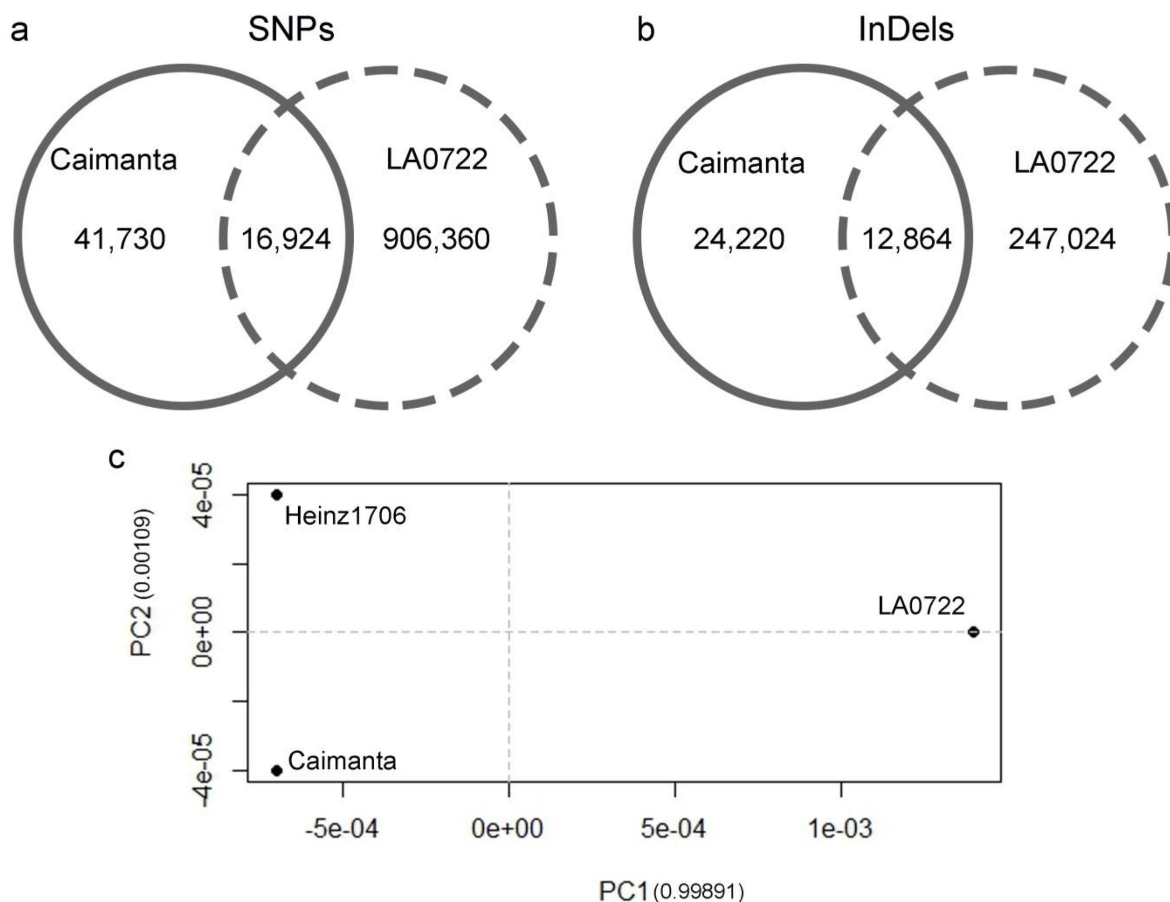
612 average spacing and maximum spacing in centiMorgan (cM) per chromosome

Chromosome	**N° Markers**	**Length (cM)**	**Average Spacing (cM)**	**Maximum Spacing (cM)**
1	22	214.8	10.2	43.8
2	10	146.1	16.2	30.3
3	15	157.8	11.3	33.0
4	8	117.1	16.7	26.7
5	13	97.7	8.1	26.5
6	14	90.7	7.0	15.2
7	15	119.4	8.5	22.2
8	16	99.2	6.6	21.3
9	11	137.5	13.8	33.8
10	14	77.8	6.0	33.6
11	9	154.8	19.4	43.6
12	10	82.7	9.2	34.1
Total	**157**	**1,495.6**	**10.3**	**43.8**

613

614 # Figures

615 **Fig. 1a** Number of SNPs relative to the tomato genome reference cultivar Heinz 1706 (version
616 SL2.50) detected for the *Solanum lycopersicum* L. cv. Caimanta and the *S. pimpinellifolium* L.
617 accession LA0722. **b** Number of InDels relative to the tomato genome reference cultivar Heinz 1706
618 (version SL2.50) detected for the *Solanum lycopersicum* L. cv. Caimanta and the *S. pimpinellifolium*
619 L. accession LA0722. **c** Principal component plot obtained from a similarity matrix among the
620 *Solanum lycopersicum* L. cv. Caimanta and the *S. pimpinellifolium* L. accession LA0722 relative to
621 the tomato genome reference cultivar Heinz 1706 (version SL2.50)

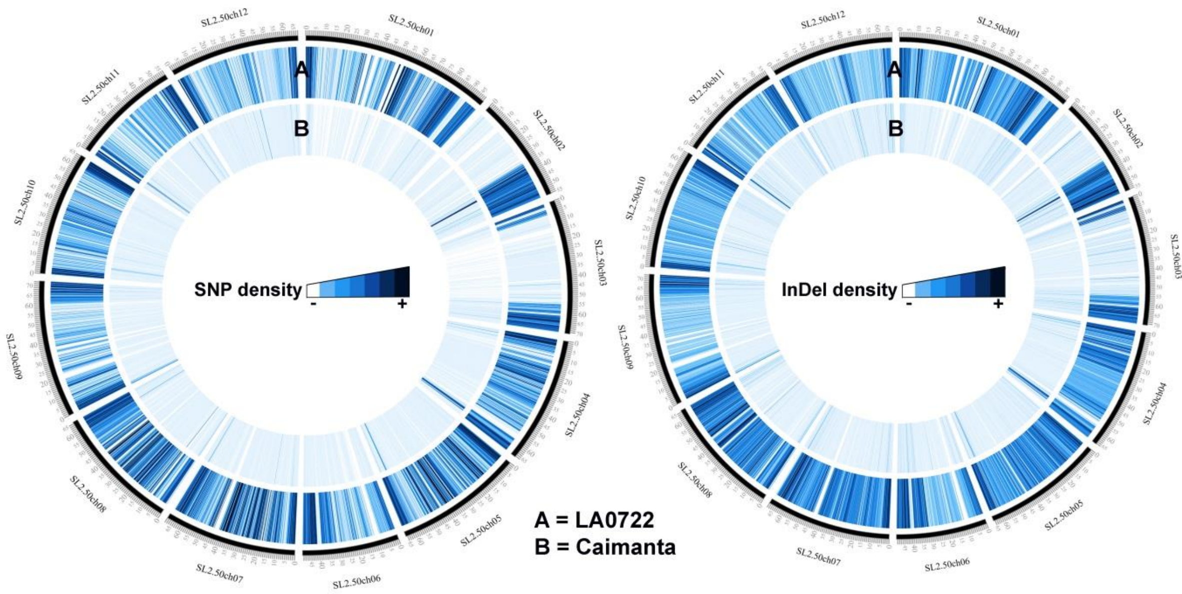


622
623

624 **Fig. 2** SNP and InDel density plots by chromosome for the *Solanum lycopersicum* L. cv. Caimanta and

625 the *S. pimpinellifolium* L. accession LA0722 against the tomato genome reference cultivar Heinz 1706

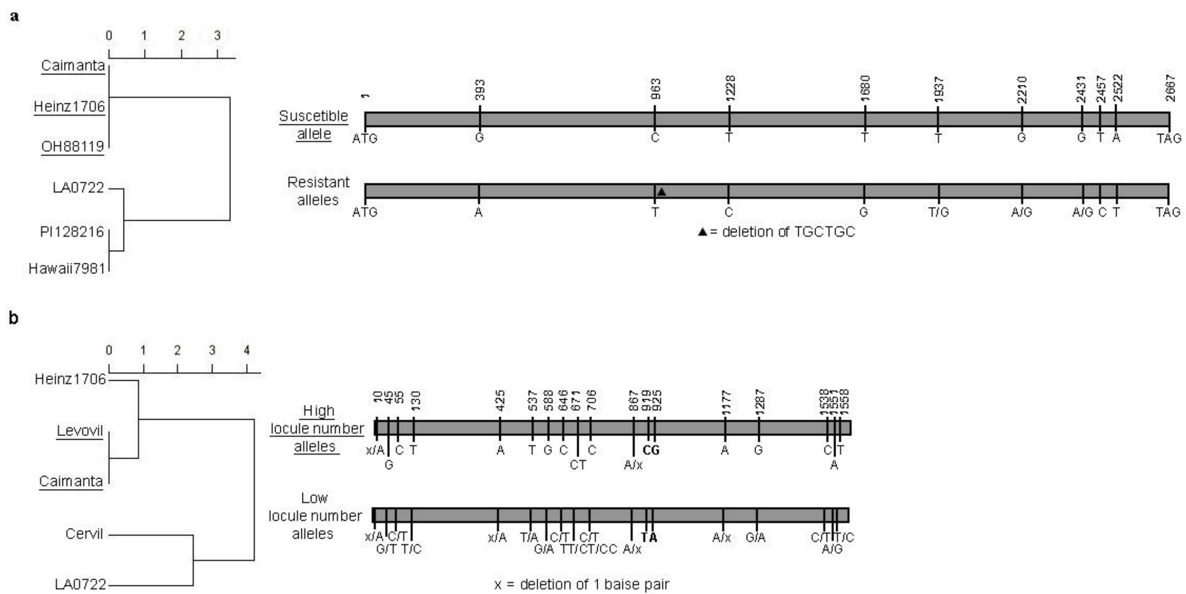
626 (version SL2.50)



627

628

629 **Fig. 3** Sequence analysis of specific candidate genes including alleles in the *Solanum lycopersicum* L.
630 cvs Caimanta and Heinz1706 and in the *S. pimpinellifolium* L accession LA0722. Cluster analysis is
631 based on the distance matrix generated from a multiple sequence comparison using log-expectation as
632 implemented using Multiple Sequence Comparison by Log-Expectation (MUSCLE) using the Ward
633 method for hierarchical clustering. **a** *Rx4/Xv3 bacterial spot resistance*. Underlined genotypes present
634 the susceptible allele. **b** *LC locule number*. Underlined genotypes present the high locule number
635 allele. Bold letters indicate the two single-nucleotide polymorphisms responsible for increasing locule
636 number. The grey bars are graphical representations of the allele sequences. Polymorphic sites are
637 indicated by numbers above the gray bars, specific polymorphic nucleotides are specified under the
638 gray bars

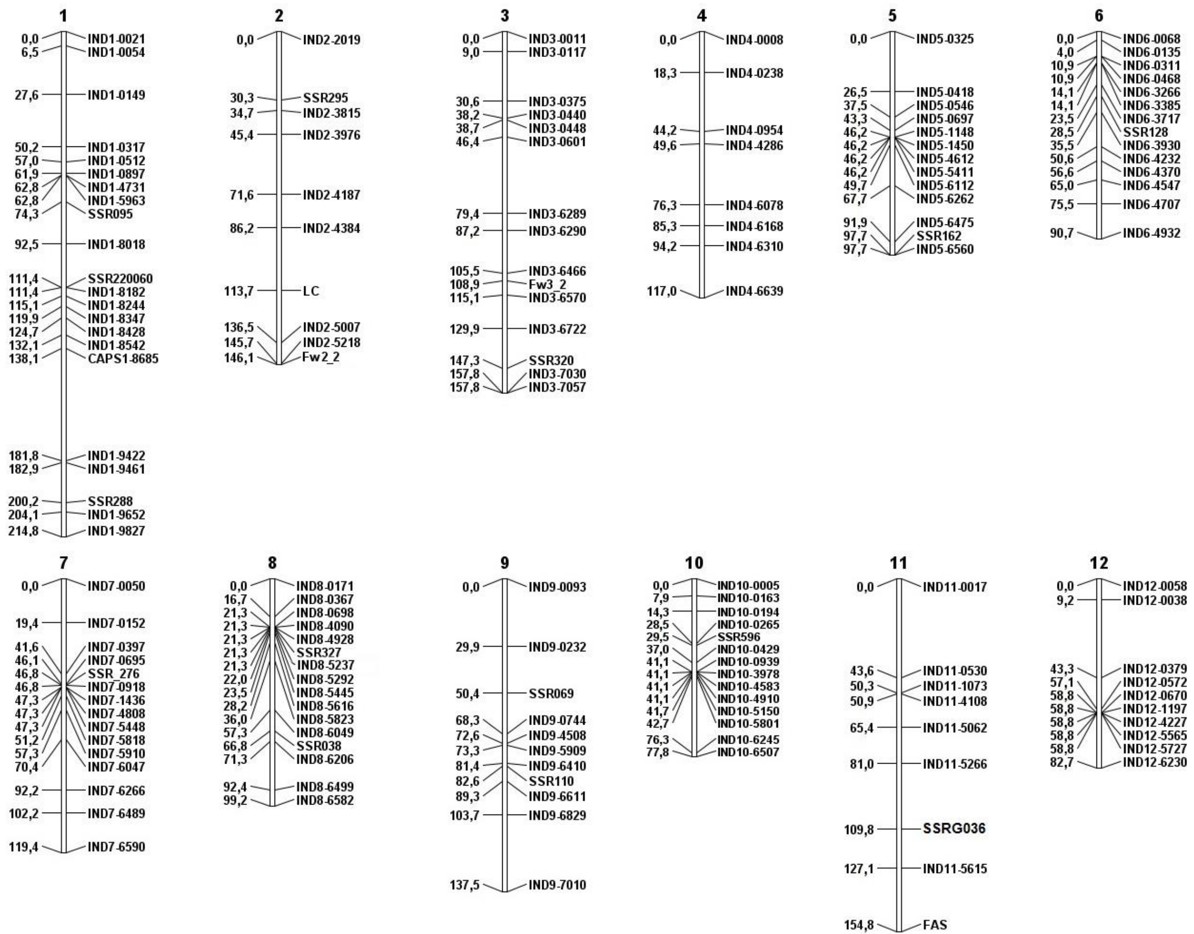


639

640

641 **Fig. 4** Genetic linkage map of the F_2 population derived from the interspecific cross between the

642 *Solanum lycopersicum* L. cv. Caimanta and the *S. pimpinellifolium* L. accession LA0722

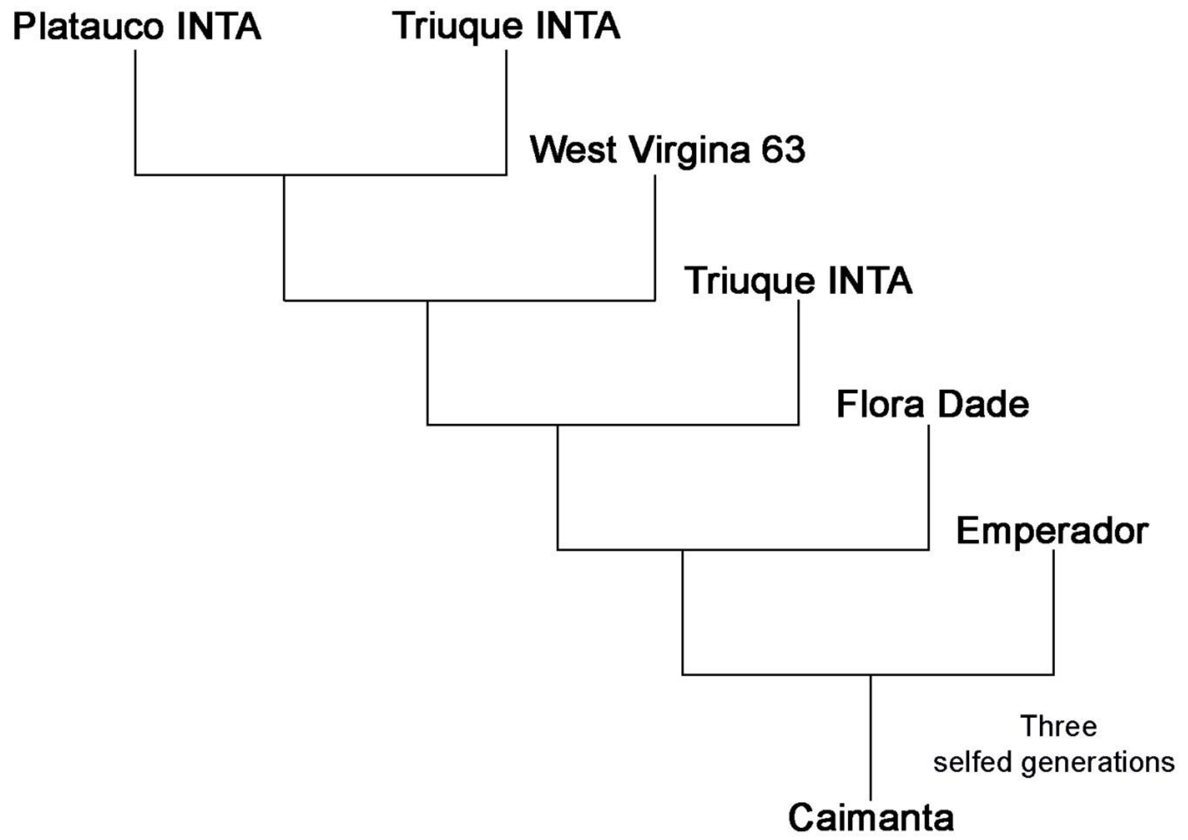


643

644

645 # Supporting information

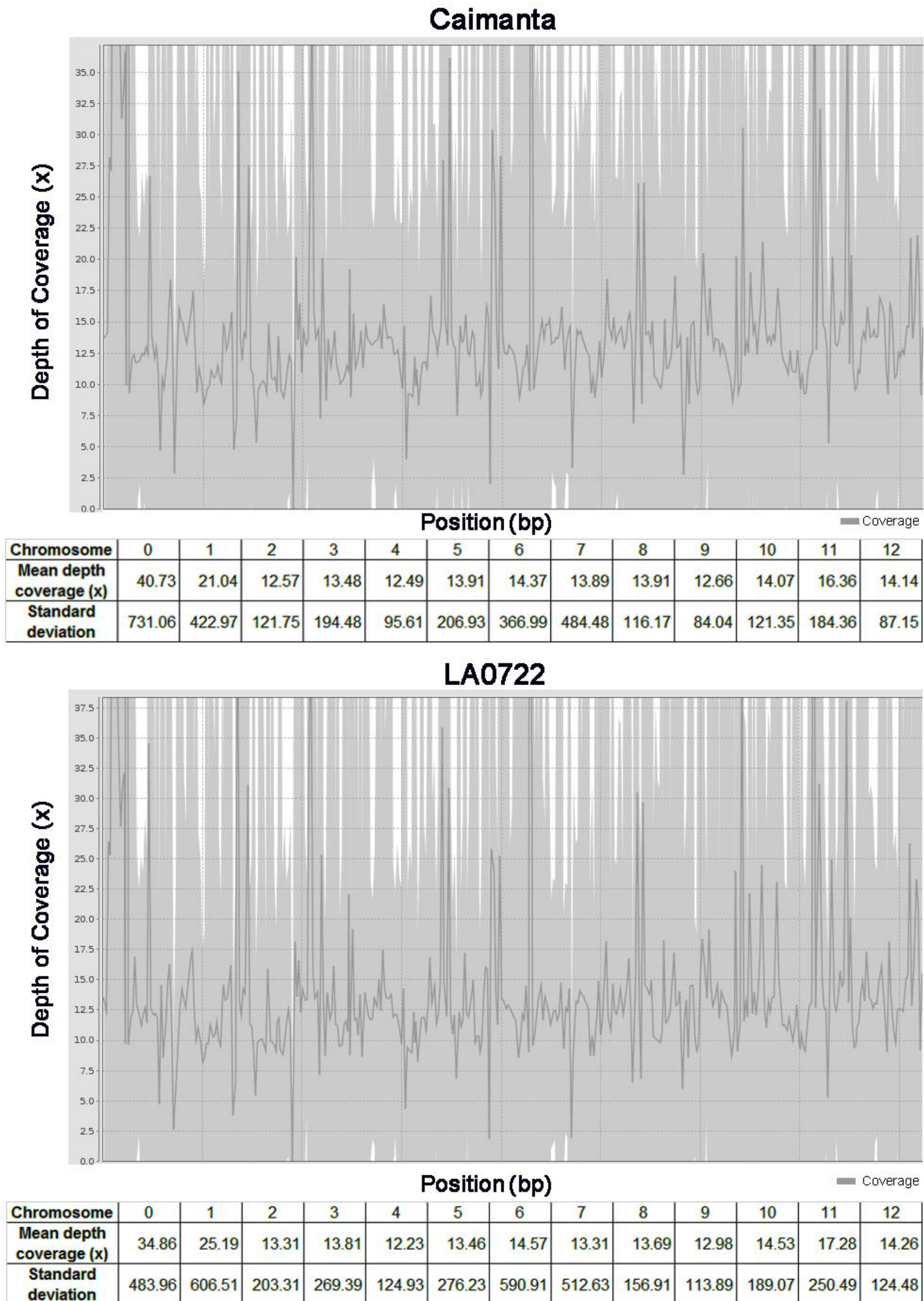
646 **Fig. S1** Pedigree for the *Solanum lycopersicum* L. cv. Caimanta



647

648

649 **Fig. S2** Depth of coverage across the entire genome reference for the *Solanum lycopersicum* L. cv.
650 Caimanta and the *S. pimpinellifolium* L. accession LA0722. The detected mean depth of coverage (x)
651 and standard deviation is shown by chromosome

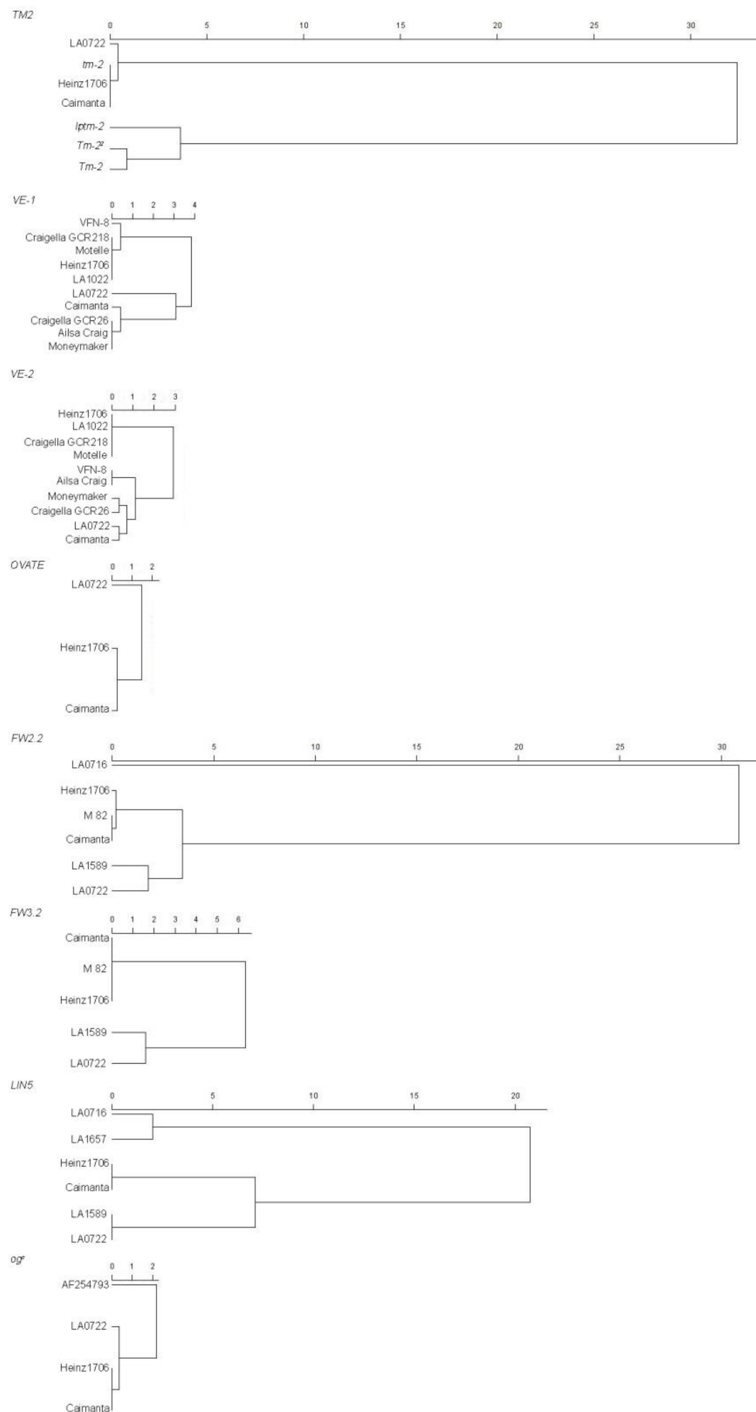


Chromosome	0	1	2	3	4	5	6	7	8	9	10	11	12
Mean depth coverage (x)	40.73	21.04	12.57	13.48	12.49	13.91	14.37	13.89	13.91	12.66	14.07	16.36	14.14
Standard deviation	731.06	422.97	121.75	194.48	95.61	206.93	366.99	484.48	116.17	84.04	121.35	184.36	87.15

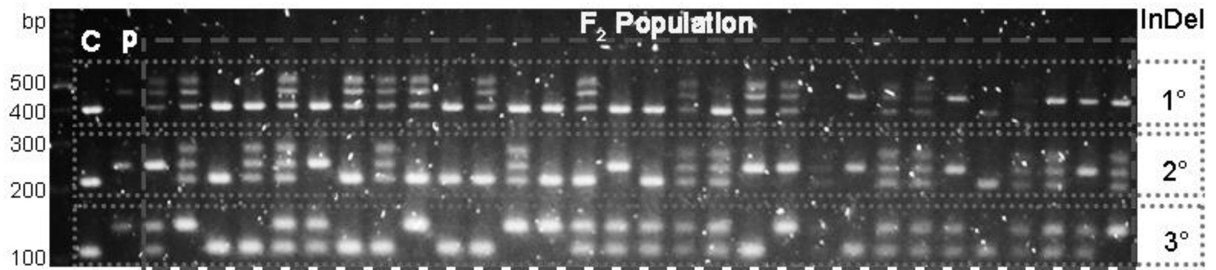
Chromosome	0	1	2	3	4	5	6	7	8	9	10	11	12
Mean depth coverage (x)	34.86	25.19	13.31	13.81	12.23	13.46	14.57	13.31	13.69	12.98	14.53	17.28	14.26
Standard deviation	483.96	606.51	203.31	269.39	124.93	276.23	590.91	512.63	156.91	113.89	189.07	250.49	124.48

652

Fig. S3 Sequence analysis of specific candidate genes including alleles in the *Solanum lycopersicum* L. cvs Caimanta and Heinz1706 and in the *S. pimpinellifolium* L accession LA0722. Cluster analysis is based on the distance matrix generated from a multiple sequence comparison using log-expectation as implemented using Multiple Sequence Comparison by Log-Expectation (MUSCLE) using the Ward method for hierarchical clustering



658 **Fig. S4** 3% w/v agarose gel electrophoresis following multiplex PCR for three InDel markers of
659 different sizes. C: *Solanum lycopersicum* L. cv. Caimanta; P: *S. pimpinellifolium* L. accession
660 LA0722; bp: base pairs

661 

662

663 **Table S1.** Details on the sequences used to evaluate disease resistance, fruit shape, fruit weight and
664 fruit quality genes.

665 **Table S2.** Molecular markers.