Research Paper

Selected genome regions for fruit weight and shelf life in tomato RILs discernible by markers based on genomic sequence information

Vladimir Cambiaso[†1], Magalí Diana Gimenez[†1], Javier Hernán Pereira da Costa[1,2], Dana Valeria Vazquez[1], Liliana Amelia Picardi[1,2,3], Guillermo Raúl Pratta[1,2] and Gustavo Rubén Rodríguez*[1,2]

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

[2] *Cátedra de Genética, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Campo Experimental Villarino*, S2125ZAA Zavalla, Santa Fe, Argentina

[3] *Consejo de Investigaciones de la Universidad Nacional de Rosario. Universidad Nacional de Rosario*, Maipú 1065, S2000CGK Rosario, Santa Fe, Argentina

Fruit weight (FW) and shelf life (SL) are important traits in commercial fresh market tomatoes. A tomato RIL population was developed by antagonistic and divergent selection for both traits from an interspecific cross between the *Solanum lycopersicum* L. cv. "Caimanta" and the *S. pimpinellifolium* L. accession "LA0722". The objective of this work was to evaluate phenotypic and genetic components for FW and SL. Phenotypic data from RILs were collected during 3-year trials. Sixteen SSR, 62 InDels developed based on the genome sequences of "Caimanta" and "LA0722", and four functional markers for fruit size genes were used. FW and SL had a significant genetic variability, and both traits showed a genotype by year interaction. Genome-wide molecular characterization of the population demonstrated that is genetically structured according to FW. Marker data was used to study changes on allelic frequencies at *loci* between the phenotypic extreme group of RILs for FW and SL. Twenty four markers were associated to FW, the *LC* gene in chromosome 2 and other six markers in chromosomes 1, 2, 6, and 11 presented the most significant associations. Finally, we reported three new genomic regions located on chromosomes 9, 10 and 12 that underlie SL in tomato.

Key Words: allelic frequencies, fruit quality, InDels, *Solanum* section *lycopersicon*.

Introduction

Tomato wild relatives represent a still unexploited source of alleles to develop new cultivars in breeding programs and increase the genetic variability in the crop for several agronomic traits comprising fruit weight and fruit shelf life (Gur and Zamir 2004, McCouch *et al.* 2013, Rodríguez *et al.* 2010). The genetic bases of the fruit weight (FW) in tomato has been well studied through interspecific as well as intraspecific crosses. Although hundreds of QTLs have been identified just seven located on chromosomes 1, 2, 3, 4, 9 and 11 showed large effect (Grandillo and Cammareri 2016, Lippman and Tanksley 2001). Five of them were identified by positional cloning: *FW2.2* (Frary *et al.* 2000) in chromosome 2, *FAS* or fasciated in chromosome 11 (Xu *et al.*

2015), *LC* in chromosome 2 (Muños *et al.* 2011), *FW3.2* in chromosome 3 (Chakrabarti *et al.* 2013), and *FW11.3* in chromosome 11 (Mu *et al.* 2017). An important trait in commercial fresh market tomatoes (*Solanum lycopersicum* L.) is the fruit shelf life. Some tomato mutant genes, such as *rin* (*ripening inhibitor*), *nor* (*non-ripening*), *Nr* (*Never ripe*), and *alc* (*alcobaca*), delay the normal process of ripening but have undesirable pleiotropic effects on other components of fruit quality, even in their heterozygous state (Kovács *et al.* 2009, Mutschler *et al.* 1992, Rodríguez *et al.* 2010, Vrebalov *et al.* 2002). Biotechnology has made considerable progress in modifying metabolic pathways that delay the tomato ripening process and discovering complex networks of transcription factors that regulate ripening (Centeno *et al.* 2011, Meli *et al.* 2010, Smith *et al.* 1988, Uluisik *et al.* 2016). Even though these features could be added to commercial tomatoes through genetic engineering, some consumers do not support genetic engineering of crops due to unanswered questions about food and environmental safety (Lucht 2015, Valente and Chaves 2018). Besides, Zamir (2001) says that "considering the problems of

consumer acceptance of genetically modified (GM) products we should apply GM technologies in plant breeding only in cases in which we have no other classical genetic alternative". Therefore, to overcome limitations caused by either the undesirable pleiotropic effects detected in spontaneous mutations that increase the shelf life in the tomato germplasm and the commercial issues of transgenic cultivars, alleles present in wild species constitute an important resource to improve fruit shelf life in tomato breeding programs.

The cross between the *S. lycopersicum* L. cv. "Caimanta" and the *S. pimpinellifolium* L. accession "LA0722" resulted promising to improve both fruit quality and shelf life in a tomato breeding program. Populations with different genetic structures derived from this interspecific cross are available, such as Recombinant Inbred Lines (RIL, Rodríguez *et al.* 2006a), Second Cycle Hybrids (SCH, Cabodevila *et al.* 2017, Pereira da Costa *et al.* 2016, Pratta *et al.* 2011a), and backcross populations (Pereira da Costa *et al.* 2013, 2014). In previous work, we have demonstrated that RILs obtained from the cross between cv. "Caimanta" and the accession "LA0722" have a longer shelf life than the wild parent and better fruit quality than the cultivated type (Rodríguez *et al.* 2006a, 2006b). These genotypes were characterized by Amplified Fragment Length Polymorphism (AFLP) markers and Quantitative Trait *Loci* (QTL) for fruit weight, solids soluble content, acidity, firmness, and fruit shelf life (Pratta *et al.* 2011b, 2011c). A higher number of QTLs was detected for fruit weight and shelf life, both traits that were used as select target throughout the developmental process of the RILs.

Since their discovery in the 1980s, microsatellites or Single Sequence Repeats (SSR) are extensively employed in plant genetics studies, using both low- and high-throughput genotyping approaches. Today, because these were developed based on known sequences, we can predict their distribution throughout the genome; the primers sequences; sizes of the cultivated and wild alleles; and chromosomal location and physical and genetic distance at reference maps (http://www.solgenomics.net). Also, a tomato genome reference was published based on the sequence assembly of the *S. lycopersicum* L. cv. "Heinz 1706" (The Tomato Genome Consortium 2012). Insertion/Deletion (InDel) polymorphisms, which include SSR, are the second most abundant form of sequence variation in the genome. Recently, Cambiaso *et al.* (2019) sequenced the whole genome of "Caimanta" and "LA0722" and developed a set of InDel DNA markers that were multiplexed and scored using easily accessed genotyping platforms. These markers were used to construct a genetic linkage map as a reference for QTL detection and resource for marker-assisted selection.

The objective of this work was to evaluate the phenotypic and genetic components for fruit weight and fruit shelf life in a 3-year tomato RIL population trial. We also used molecular markers based on genomic sequence information to identify and localize underlying genomic regions.

Materials and Methods

Plant material

Eighteen tomato RILs were developed via five cycles of antagonistic and divergent selection for fruit shelf life and fruit weight from an interspecific cross hybrid between the *S. lycopersicum* cv. "Caimanta" and the *S. pimpinellifolium* accession "LA0722" (Rodríguez *et al.* 2006a). The antagonistic-divergent selection began at the F_2 segregating generation and continued until F_6 (Pratta *et al.* 2011b, 2011c, Rodríguez *et al.* 2006a). Though the number of RILs could be considered small to span all existing recombination, they correctly represent the selected phenotypic extremes for both traits. Therefore, this experiment could be considered a selective genotyping approach (Liu 2004). The parental genotypes, cv. "Caimanta", and accession "LA0722" were used as a tester. These were provided by INTA Cerrillos (Salta, Argentina) and Tomato Genetic Resources Center (TGRC, CA, USA) respectively.

Trials and phenotypic data collection

Trials were conducted at Campo Experimental Villarino, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Argentina (33°S lat. and 61°W long.). During the three years (y1, y2, and y3), trials were conducted to collect phenotype data for RILs with F_6 to F_8 cycles respectively. The genotypes were sown in seedling trays and were transplanted to the field (y1) or greenhouse (y2 and y3) after 45 days in a completely random design. The number of plants per genotype was 14, 8, and 9 for y1, y2, and y3 respectively. The RIL 17 was not evaluated in y1. The crop was grown according to the standard cultural recommendations for the area, and the plants were trained. Ten tomato fruits at the breaker stage defined by Giovannoni (2004) were harvested from each plant to evaluate fruit shelf life (SL, in days) and fruit weight (FW, in g) attributes. SL was measured as the number of days elapsed from the harvesting until the first symptoms of deterioration and excessive softening appeared. To determine this trait, the fruits were stored at $25 \pm 3°C$. Fruits were examined three times per week, and those commercially unacceptable due to exhibiting wrinkled areas or excessive softening were discarded (Pereira da Costa *et al.* 2013, Rodríguez *et al.* 2010).

Molecular characterization

Genomic DNA was extracted from young leaves of each plant in y3 using a commercial Kit (Wizard® Genomic DNA Purification Kit from Promega®, Madison, WI, USA). Multiple markers types were used. Sixteen SSR tested by Pereira da Costa *et al.* (2013), 62 InDel developed based on the genome sequence of "Caimanta" and "LA0722" (Cambiaso *et al.* 2019), and four functional markers for fruit size genes: *FAS* (Rodríguez *et al.* 2011); *FW2.2* (Blanca *et al.* 2015), *LC* (Muños *et al.* 2011), and *FW3.2* (Chakrabarti *et al.* 2013). Markers were selected to cover the entire tomato genome. Details on the molecular markers used are summarized in

Supplemental Table 1. The same standard PCR protocol was used to obtain amplification products from the different kinds of molecular markers with two biological replicates per genotype. Electrophoresis of InDel and functional markers was conducted on 3% w/v agarose gels stained with SYBR® Safe DNA Gel Stain (Thermo Fisher Scientific®, Waltham, MA, USA) for visualization, while SSR markers were run on 6% w/v polyacrylamide gels visualized by a silver staining procedure.

Phenotypic data analysis

The normality of both trait distributions was verified using the Shapiro-Wilk test (Shapiro and Wilk 1965). The genotype by year interaction was estimated by two-way ANOVA. Best Linear Unbiased Predictors (BLUPs) based on random models were estimated for both FW and SL. For each trait, the following linear model was used:

$$Y_{ij} = \mu + g_i + y_j + gy_{ij} + \varepsilon_{ij}$$

where Y_{ij} is the trait measured, μ is the overall mean, g_i is the effect resulting from the ith genotype, y_j is the effect resulting from the jth year, gy_{ij} is the effect resulting from genotype by year interaction, and ε_{ij} is the residual error (effect resulting from experimental error). The statistical model was fitted with the lmer function from the lme4 package (Bates et al. 2015) in the R software (R Core Team 2017). Estimates of variance were retrieved from the summary table generated by the lmer function and were used to calculate the percentage of variance for genetic and environmental factors in the model. Based on the formula proposed by Kearsey and Pooni (1996), who established that non-additive genetic variance is negligible after six cycles of selfing, the narrow sense heritability (h^2) for each trait was calculated as: $h^2 = \dfrac{\sigma_g^2}{\left(\sigma_g^2 + \sigma_{gy}^2 + \sigma_\varepsilon^2\right)}$, with σ_g^2, the variance of the genotypes; σ_{gy}^2 the variance of the interaction between genotype and year, and σ_ε^2, the variance of the error.

Molecular data analysis

Molecular markers were defined as polymorphic when different size bands were visualized among RILs in the electrophoresis gel. The marker's score for each RIL was defined as homozygous for "Caimanta" alleles (LL), homozygous for "LA0722" alleles (PP) and heterozygous (LP). If the biological replicates have a different score, they are defined as a heterozygous or segregating marker. To evaluate the genetic variation in the RIL population, hierarchical cluster analysis with 1,000 bootstrapping was performed using Manhattan distance and Ward method (Murtagh and Legendre 2014). The molecular data from the 18 RILs was also used to analyze genetic relationships employing the software STRUCTURE version 2.3 (Pritchard et al. 2000). To estimate the degree of admixture, a burn-in period length of 10,000 was used and the number of Markov Chain Monte Carlo iterations set to 100,000. Optical K was inferred according to Evanno et al. (2005).

Marker data was used to study changes on allelic frequencies at loci between the phenotypic extreme group of RIL for FW and SL. The phenotypic extreme groups were composed by the RILs with BLUPs values for FW or SL above or under the average values of the RIL population. We applied a G-statistic to characterize the data at each marker that follows a chi-square (χ^2) distribution with one degree of freedom (Magwene et al. 2011).

Results

Phenotypic characterization

The RIL population showed phenotypic variation for FW (**Fig. 1**) and SL (**Fig. 2**). The source of variations genotype and year were significant and the genotype by year interaction was highly significant for both traits ($p < 0.0001$). The narrow sense heritability was 0.69 for FW, so a major proportion of the phenotypic variance is explained by the additive genetic component (**Fig. 1B**). Independently of the year effect, the genotypes with the highest values for FW were RIL 1, RIL 3, and RIL 14, whereas the lowest values were for RIL 6, RIL 7, RIL 8; RIL 9, RIL 10, RIL 11, RIL 12, and RIL 13 (**Fig. 1**). The RIL 2, RIL 16, RIL 17, and RIL 18 mean values were variable throughout the years and only showed higher fruit weight values in the second year of evaluation. Regarding SL (**Fig. 2**), the narrow sense heritability had a value of 0.48. RIL 17 had the highest values for SL independent of the year of evaluation. In a second group, the higher values for the trait were showed by RIL 3, RIL 4, RIL 7, RIL 12, and RIL 13 in y3 and RIL 8 in y1 and y3 (**Fig. 2**). On the other hand, it is noticeable that RIL 8, RIL 17, and RIL 18 had a variable performance for SL throughout the years. In spite of this they always showed higher mean values than the wild parental genotype "LA0722".

Molecular characterization

Each RIL had a unique molecular profile. All markers were polymorphic among the 18 RILs, except IND4-0954 at chromosome 4 and the functional marker for *FAS* at chromosome 11. For *FAS* all RILs were homozygous as the wild accession "LA0722". RILs showed, on average, 14.8% of loci in the heterozygous state, which is higher than expected based on the number of selfing generations. RIL 1, RIL 2, RIL 3, and RIL 18 presented fewer loci in the heterozygous state (3.5, 1.2, 2.3, and 2.3% respectively). On the contrary, RIL 9, RIL 11, and RIL 16 showed the highest values of heterozygosity with values of 34.9, 34.9, and 50% respectively.

Fig. 3A shows a scheme of the RIL population structure along the 12 tomato chromosomes. Most of the genome was 50% of each homozygous genotype, as expected in a RIL population generated without selection. However, in some regions the "LA0722" alleles were overrepresented, for example those at chromosomes 4, 8, 9, 10, and 11, as well as at the bottom of chromosome 5. And also for other regions
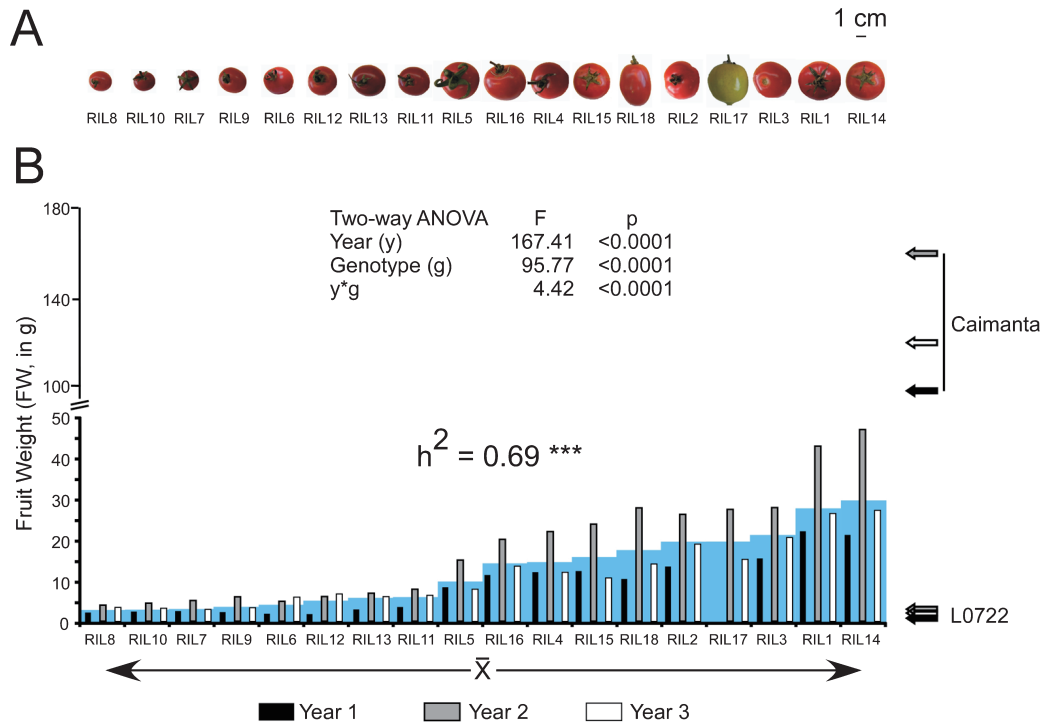
Fig. 1. A) Images of representative fruits for the 18 Recombinant Inbred Lines (RIL); B) Phenotypic data for fruit weight in the RIL population for the three years of trials. Mean values for the first, second, and third years of the trial are represented by black, grey or white bars respectively. Mean values for *Solanum lycopersicum* L. cv. "Caimanta" and *S. pimpinellifoilium* L. accession "LA0722" are indicated by arrows on the right. The Best Linear Unbiased Predictors (BLUPs) for fruit weight are represented by a blue box. h^2: narrow sense heritability; p: probability; F: statistic F. The average value that defines the extreme phenotypic for fruit weight is indicated by $\bar{X}$. *** p < 0.001.
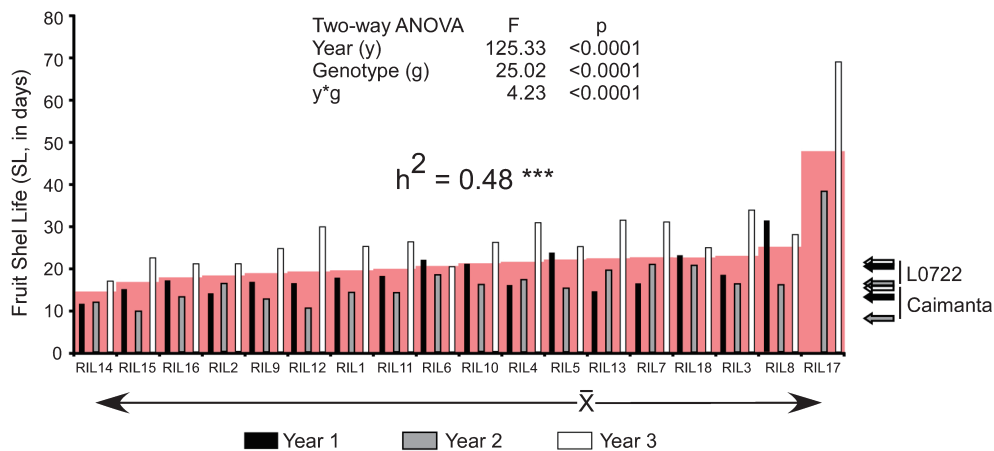

Fig. 2. Phenotypic data for fruit shelf life in the RIL population for the three years of trials. Mean values for the first, second, and third years of the trial are represented by black, grey or white bars respectively. Mean values for *Solanum lycopersicum* L. cv. "Caimanta" and *S. pimpinellifoilium* L. accession "LA0722" are indicated by arrows on the right. The Best Linear Unbiased Predictors (BLUPs) for shelf life are represented by a red box. h^2: narrow sense heritability; p: probability; F: statistic F. The average value that defines the extreme phenotypic for fruit shelf is indicated by $\bar{X}$. *** p < 0.001.

the "Caimanta" alleles were overrepresented, like all along chromosomes 1, 2, 12 and the top of chromosome 3.

The cluster analysis with all markers shows two groups (**Fig. 3B**). One of them was formed by the parental "Caimanta" and seven RILs, whereas a second group consisted of "LA0722" and the others 11 RILs. Within the first group, the parental "Caimanta" was molecularly different

from RIL 1, RIL 2, RIL 3, RIL 14, RIL 15, RIL 17, and RIL 18. The second group can be divided into three minor groups. One of them was integrated by the parental "LA0722" and RIL 6, RIL 7, and RIL 8. Another group was consolidated just for RIL 4 and RIL 5, and the last group by RIL 9, RIL 10, RIL 11, RIL 12, and RIL 16. The three RILs (1, 2, 3, 14 and 17) with the highest mean values for FW throughout the

	Caimanta	RIL15	RIL2	RIL3	RIL17	RIL1	RIL14	RIL18	RIL7	RIL8	LA0722	RIL6	RIL4	RIL5	RIL16	RIL10	RIL9	RIL11	RIL12	RIL13
FAS [a]	LL	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	NA	PP	PP	PP
LC [b]	LL	LL	LL	LL	LL	LL	LL	LL	PP	PP	PP	PP	LL	LL	LP	PP	PP	PP	PP	PP
FW2.2 [c]	LL	LL	LL	LL	LL	LL	LL	LL	LP	LP	PP	LP	LL	LL	PP	PP	LP	LP	PP	LP
FW3.2 [d]	LL	LL	LL	LL	LL	LL	LL	LL	LL	PP	PP	PP	LL	LL	LL	PP	LL	LL	LL	LL

Fig. 3. A) "Caimanta" and "LA0722" allele frequencies distribution scheme in the 18 Recombinant Inbred Lines (RIL) along the 12 tomato chromosomes. B) Cluster analysis based on 16 SSR, 62 InDel, and 4 functional markers genotypic data with alleles distribution at functional markers and inferred ancestry for cluster *Q1* and *Q2* from structure analysis. Mb: megabases; au: *p*-value approximately unbiased; LL homozygous for *S. lycopersicum* allele; PP: homozygous for *S. pimpinellifolium* alleles; LP: heterozygous. [a]: Rodríguez *et al.* 2011; [b]: Muños *et al.* 2011; [c]: Blanca *et al.* 2015; [d]: Chakrabarti *et al.* 2013.

Fig. 4. G-statistic for allelic frequencies throughout tomato genome between phenotypic extreme values for FW (black dots) and SL (white dots). χ^2 (df = 1) with **: $p < 0.01$; ***: $p < 0.001$.

three years of evaluation clustered at the first group, whereas those RILs with the lowest mean values (RIL 6 to 13) belong to the second group. The distribution of genotypes for FW functional markers agrees with the defined clusters. All RILs in the first group carry the cultivated alleles (LL) at the *LC*, *FW2.2*, and *FW3.2* genes. Most of the RILs in the second group have at least one gene homozygous for *S. pimpinellifolium* allele (PP), except RIL 4 and RIL 5.

A K value of 2 was chosen as the most informative of the RIL population structure. The estimated ancestry coefficients were 0.788 in Q1 and 0.212 in Q2 for "Caimanta"

(**Fig. 3B**). "LA0722" had a coefficient of estimated ancestry of 0.990 in Q2. The RIL population presented coefficients of estimated ancestry in Q1, which ranged from 0.002 to 0.999. The estimated coefficients of ancestry had correspondence with the hierarchical cluster analysis. Most of the RILs with high values in Q1 clustered in the first group with "Caimanta" and the RILs with high values in Q2 grouped with "LA0722". However, some RILs (4, 5, 12, 13 and 15) represent more admixture genotypes. Therefore, it is noticeable that this RIL population is mostly structured for FW, one of the selection traits targeted during the development.

Changes in allele frequencies between phenotypic extreme for FW and SL

With the G-statistic analysis, 24 markers resulted significantly associated ($p < 0.001$) with FW throughout almost all tomato chromosomes (Chr), except of Chr 10 and 12 (**Fig. 4**). The largest number of associations, with a total of 4, was found in Chr 2 and 11. The SSR095 (Chr 1), SSR032 (Chr 2), IND6.0468 and IND6.3717 (Chr 6), IND11.0017 and SG036 (Chr 11) markers presented the highest G values. The cloned gene *LC* was detected as segregating and associated with FW ($p < 0.001$) in this RIL population. On the other hand, *FAS* was not segregating and both *FW2.2* and *FW3.2* were segregating and associated with FW at $p < 0.01$. Additionally, other 17 markers were also associated with FW at $p < 0.01$.

Only the SSR318 marker in chromosome 10 was significant associated with SL at $p < 0.001$. Besides, the marker IND9.7010 (Chr 9) and IND12.0379 (Chr 12) were associated with this attribute only at $p < 0.01$ (**Fig. 4**). These three genomic regions are new putative regions that could be involved on the genetic control of this complex trait.

Discussion

FW and SL mean values were affected by genotype, year of evaluation, and the genotype by year interaction, as Pratta *et al.* (2011b) had previously demonstrated for this RIL population. However, some genotypes always perform with the higher mean values for each trait independently of the year of evaluation, for example: RIL 1, RIL 2, RIL 3, and RIL 14 for FW, and RIL 8, RIL 17, and RIL 18 for SL (**Fig. 1**). This categorization agrees with results previously reported by Rodríguez *et al.* (2006a, 2006b). FW and SL have high values of narrow sense inheritability, which indicates a high proportion of phenotypic variance explained by the genetic–mainly the additive–component. Nonetheless, this parameter was overestimated since a large proportion of the genome remains in a heterozygous state. In this study, RILs showed, on average, 14.8% of *loci* in the heterozygous state, which was higher than expected (~1%) based on the number of selfing generations. These values agree with those reported by Truong *et al.* (2014), where the observed heterozygous frequency is much higher than the expected heterozygous frequency in an F_7 RIL population and with some individuals achieving values near 40%. Furthermore, the RIL population used in this study could show a higher percentage of heterozygous *loci* than expected if the target trait under selection presents some non-additive mechanisms in its genetic bases.

The scheme of the RIL population genomic structure along the 12 tomato chromosomes (**Fig. 3A**) shows that some genomic regions are overrepresented by alleles from "Caimanta" and others, by "LA0722" parental genotypes. These regions could be the result of artificial selection for FW and SL. For example, regions on chromosomes 1, 2, 12, and the top of chromosome 3, where "Caimanta" alleles are overrepresented, were under selection pressure for FW and regions on chromosomes 4, 8, 9, 10, 11, and the bottom of chromosomes 5, where "LA0722" alleles are overrepresented, were under selection pressure for SL.

The conformation of the clusters agrees with discrepancies among genotypes for FW for both mean phenotypic values and score for FW functional genes (**Fig. 3B**). This cluster analysis revealed that the grouping is related to FW, one of the target selection traits, and for the known genes controlling this attribute. However, some other results could be due to other selected genomic regions during RILs development. On the other hand, RILs with better performance for SL (RIL 3, RIL 7, RIL 8, RIL 13, RIL 17 and RIL 18) did not follow an associated pattern among the defined clusters. The result of the model-based clustering analysis is in agreement with previous studies where both cultivated and wild tomato genotypes were included (Lin *et al.* 2014, Ranc *et al.* 2008). A K value of 2 is as expected for a tomato population where wild and cultivated genotypes are included in the analysis as well as in population derived from a biparental crossing. The *S. pimpinellifolium* L. accession "LA0722" has a Q2 value higher than 0.99, whereas cv. "Caimanta" has a Q1 value of 0.788 and Q2 of 0.212 indicating that "Caimanta" has introgressed genes from this wild species, as well as some other. The accession "LA0722" used in this study is a genotype that belongs almost exclusively to the wild species group, whereas "Caimanta" has a cultivated genotype with some introgressed genome regions from other wild relatives.

In this experiment, we detected segregation for some known genes or QTL, for FW, such as *fw1.1* (marked by SSR095 in Chr 1), *FW2.2*, *LC* (both in Chr 2), and *FW11.3* (marked by SSRG036 in Chr 11). Some of them were the markers with the highest G values, together with the SSR032 (in Chr 2), IND6.0468 and IND6.3717 (in Chr 6) and IND11.0017 (in Chr 11), which were not previously reported. Lippman and Tanksley (2001) found major QTL located on chromosomes 1, 2, 3, and 11 that explained 67% of FW phenotypic variation and the wild type alleles always decreased the mean value for the trait. Five of them were identified by positional cloning: *FW2.2* (Frary *et al.* 2000) in chromosome 2, *FAS* or fasciated in chromosome 11 (Xu *et al.* 2015), *LC* in chromosome 2 (Muños *et al.* 2011), *FW3.2* in chromosome 3 (Chakrabarti *et al.* 2013), and *FW11.3* in chromosome 11 (Mu *et al.* 2017). *FW3.2* was segregating in this RIL population but associated with FW at $p < 0.01$, confirming that the genetic background plays an important role in the expression of the genes controlling quantitative traits. Besides, *FAS* was not segregating in this RIL population since all genotypes carried the wild allele at homozygous state. This gene is almost complete recessive for the cultivated allele but is also highly dependent on the genetic background (Lippman and Tanksley 2001); therefore, we would expect loss of the cultivated allele when just the 16 highest phenotypes for FW were selected in an F_2 generation composed of 82 individuals (Rodríguez *et al.* 2006a).

The marker with the highest G value for SL was the SSR318 in chromosome 10. Pereira da Costa *et al.* 2013 found a QTL for SL (SSR596) in the same genomic region marked by the IND10.0429. The genomic regions marked by IND9.7010 (Chr 9) and IND12.0379 (Chr 12) were not previously reported linked to SL.

Overall, the results presented in this study demonstrate that FW and SL have significant genetic variability in the RIL population derived from an interspecific cross between *S. lycopersicum* L. cv. "Caimanta" and the *S. pimpinellifolium* accession "LA0722". Both traits show genotype by year interaction, but the performance for FW can be easier tracked among genotypes throughout the years of evaluation. Genome-wide characterization of the RIL population by molecular markers based on the genomic sequence information of the parental genotypes allows us to determine the population structure and the genetic relationships among the RIL that belongs to the population. Phenotypic and molecular data are shown which genes and genomic regions were selected during the development of the population, especially for FW. Finally, we also reported new genomic regions that underlie SL, a complex trait in tomato fruits.

Author Contribution Statement

Acknowledgments

Literature Cited

Bates, D., M. Mächler, B.M. Bolker and S.C. Wlaker (2015) Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67: 1–48.

Blanca, J., J. Montero-Pau, C. Sauvage, G. Bauchet, E. Illa, M.J. Diez, D.M. Francis, M. Causse, E. van der Knaap and J. Cañizares (2015) Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. BMC Genomics 16: 257.

Cabodevila, V.G., L.A. Picardi and G.R. Pratta (2017) A multivariate approach to explore the genetic variability in the F_2 segregating population of a tomato second cycle hybrid. BAG J. Basic Appl. Genet. 28: 7–17.

Cambiaso, V., G.R. Pratta, J.H. Pereira da Costa, R. Zorzoli, D.M. Francis and G.R. Rodríguez (2019) Whole genome re-sequencing analysis of two tomato genotypes for polymorphism insight in cloned genes and a genetic map construction. Sci. Hortic. 247: 58–66.

Centeno, D.C., S. Osorio, A. Nunes-Nesi, A.L.F. Bertolo, R.T. Carneiro, W.L. Araújo, M.C. Steinhauser, J. Michalska, J. Rohrmann, P. Geigenberg *et al.* (2011) Malate plays a crucial role in starch metabolism, ripening, and soluble solid content of tomato fruit and affects postharvest softening. Plant Cell 23: 162–184.

Chakrabarti, M., N. Zhang, C. Sauvage, S. Muños, J. Blanca, J. Cañizares, M.J. Diez, R. Schneider, M. Mazourek, J. McClead *et al.* (2013) A cytochrome P450 regulates a domestication trait in cultivated tomato. Proc. Natl. Acad. Sci. USA 110: 17125–17130.

Evanno, G., S. Regnaut and J. Goudet (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. 14: 2611–2620.

Frary, A., T.C. Nesbitt, A. Frary, S. Grandillo, E. van der Knaap, B. Cong, J. Liu, J. Meller, R. Elber, K.B. Alpert *et al.* (2000) fw2.2: A quantitative trait locus key to the evolution of tomato fruit size. Science 289: 85–88.

Giovannoni, J. (2004) Genetic regulation of fruit development and ripening. Plant Cell Online 16: 170–181.

Grandillo, S. and M. Cammareri (2016) Molecular mapping of quantitative trait loci in tomato. *In*: Causse, M., J. Giovannoni, M. Bouzayen and M. Zouine (eds.) The tomato genome, Springer-Verlag, Berlin.

Gur, A. and D. Zamir (2004) Unused natural variation can lift yield barriers in plant breeding. PLoS Biol. 2: e245.

Kearsey, M.J. and H.S. Pooni (1996) The genetical analysis of quantitative Traits, 1st edn. Chapman & Hall, London, p. 381.

Kovács, K., R.G. Fray, Y. Tikunov, N. Graham, G. Bradley, G.B. Seymour, A.G. Bovy and D. Grierson (2009) Effect of tomato pleiotropic ripening mutations on flavour volatile biosynthesis. Phytochemistry 70: 1003–1008.

Lin, T., G. Zhu, J. Zhang, X. Xu, Q. Yu, Z. Zheng, Z. Zhang, Y. Lun, S. Li, X. Wang *et al.* (2014) Genomic analyses provide insights into the history of tomato breeding. Nat. Genet. 46: 1220–1226.

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

Liu, B.H. (2004) Statistical genomics: linkage, mapping, and QTL analysis. CRC Press, New York, p. 610.

Lucht, J.M. (2015) Public acceptance of plant biotechnology and GM crops. Viruses 7: 4254–4281.

Magwene, P.M., J.H. Willis and J.K. Kelly (2011) The statistics of bulk segregant analysis using next generation sequencing. PLoS Comput. Biol. 7: e1002255.

McCouch, S., G.J. Baute, J. Bradeen, P. Bramel, P.K. Bretting, E. Buckler, J.M. Burke, D. Charest, S. Cloutier, G. Cole *et al.* (2013) Feeding the future. Nature 499: 23–24.

Meli, V.S., S. Ghosh, T.N. Prabha, N. Chakraborty, S. Chakraborty and

A. Datta (2010) Enhancement of fruit shelf life by suppressing N-glycan processing enzymes. Proc. Natl. Acad. Sci. USA 107: 2413–2418.

Mu, Q., Z. Huang, M. Chakrabarti, E. Illa-Berenguer, X. Liu, Y. Wang, A. Ramos and E. van der Knaap (2017) Fruit weight is controlled by *Cell Size Regulator* encoding a novel protein that is expressed in maturing tomato fruits. PLoS Genet. 13: e1006930.

Muños, S., N. Ranc, E. Botton, A. Berard, S. Rolland, P. Duffé, Y. Carretero, M.C. Le Paslier, C. Delalande, M. Bouzayen *et al.* (2011) Increase in tomato locule number is controlled by two single-nucleotide polymorphisms located near *WUSCHEL*. Plant Physiol. 156: 2244–2254.

Murtagh, F. and P. Legendre (2014) Ward's hierarchical agglomerative clustering method : Which algorithms implement Ward's criterion? J. Classif. 31: 274–295.

Mutschler, M.A., D.W. Wolfe, E.D. Cobb and K.S. Yourstone (1992) Tomato fruit quality and shelf life in hybrids heterozygous for the *alc* ripening mutant. HortScience 27: 352–355.

Pereira da Costa, J.H., G.R. Rodríguez, G.R. Pratta, L.A. Picardi and R. Zorzoli (2013) QTL detection for fruit shelf life and quality traits across segregating populations of tomato. Sci. Hortic. 156: 47–53.

Pereira da Costa, J.H., G.R. Rodríguez, G.R. Pratta, L.A. Picardi and R. Zorzoli (2014) Pericarp polypeptides and SRAP markers associated with fruit quality traits in an interspecific tomato backcross. Genet. Mol. Res. 13: 2539–2547.

Pereira da Costa, J.H., G.R. Rodríguez, D.R. Liberatti, S.L. Mahuad, E. Marchionni Basté, L.A. Picardi, R. Zorzoli and G.R. Pratta (2016) Tomato second cycle hybrids as a source of genetic variability for fruit quality traits. Crop Breed. Appl. Biotechnol. 16: 289–297.

Pratta, G.R., S.L. Mahuad, A. Díaz, E. Marchionni Basté, D.R. Liberatti, G.R. Rodríguez and R. Zorzoli (2011a) Molecular genetics of quantitative traits in tomato diallel crosses. Pesqui. Agropecu. Bras. 46: 508–515.

Pratta, G.R., G.R. Rodriguez, R. Zorzoli, E.M. Valle and L.A. Picardi (2011b) Molecular markers detect stable genomic regions underlying tomato fruit shelf life and weight. Crop Breed. Appl. Biotechnol. 11: 157–164.

Pratta, G.R., G.R. Rodriguez, R. Zorzoli, E.M. Valle and L.A. Picardi (2011c) Phenotypic and molecular characterization of selected tomato recombinant inbred lines derived from the cross *Solanum lycopersicum × S. pimpinellifolium*. J. Genet. 90: 229–237.

Pritchard, J.K., M. Stephens and P. Donnelly (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

R Core Team (2017) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ranc, N., S. Muños, S. Santoni and M. Causse (2008) A clarified posi-

tion for *Solanum lycopersicum* var. *cerasiforme* in the evolutionary history of tomatoes (solanaceae). BMC Plant Biol. 8: 130.

Rodríguez, G.R., G.R. Pratta, R. Zorzoli and L.A. Picardi (2006a) Evaluation of plant and fruit traits in recombinant inbred lines of tomato obtained from a cross between *Lycopersicon esculentum* and *L. pimpinellifolium*. Cienc. Investig. Agrar. 33: 111–118.

Rodríguez, G.R., G.R. Pratta, R. Zorzoli and L.A. Picardi (2006b) Recombinant lines obtained from an interspecific cross between *Lycopersicon* species selected by fruit weight and fruit shelf life. J. Am. Soc. Hortic. Sci. 131: 651–656.

Rodríguez, G.R., G.R. Pratta, D.R. Liberatti, R. Zorzoli and L.A. Picardi (2010) Inheritance of shelf life and other quality traits of tomato fruit estimated from F_1's, F_2's and backcross generations derived from standard cultivar, *nor* homozygote and wild cherry tomato. Euphytica 176: 137–147.

Rodríguez, G.R., S. Muños, C. Anderson, S.C. Sim, A. Michel, M. Causse, B.B. McSpadden Gardener, D.M. Francis and E. van der Knaap (2011) Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm and the relationship to fruit shape diversity. Plant Physiol. 156: 275–285.

Shapiro, S.S. and M.B. Wilk (1965) An analysis of variance test for normality (complete samples). Biometrika 52: 591–611.

Smith, C.J.S., C.F. Watson, J. Ray, C.R. Bird, P.C. Morris, W. Schuch and D. Grierson (1988) Antisense RNA inhibition of polygalacturonase gene expression in transgenic tomatoes. Nature 334: 724–726.

The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485: 635–641.

Truong, S.K., R.F. McCormick, D.T. Morishige and J.E. Mullet (2014) Resolution of genetic map expansion caused by excess heterozygosity in plant recombinant inbred populations. G3 (Bethesda) 4: 1963–1969.

Uluisik, S., N.H. Chapman, R. Smith, M. Poole, G. Adams, R.B. Gillis, T.M.D. Besong, J. Sheldon, S. Stiegelmeyer, L. Perez *et al.* (2016) Genetic improvement of tomato by targeted control of fruit softening. Nat. Biotechnol. 34: 950–952.

Valente, M. and C. Chaves (2018) Perceptions and valuation of GM food: A study on the impact and importance of information provision. J. Clean. Prod. 172: 4110–4118.

Vrebalov, J., D. Ruezinsky, V. Padmanabhan, R. White, D. Medrano, R. Drake, W. Schuch and J. Giovannoni (2002) A MADS-box gene necessary for fruit ripening at the tomato *ripening-inhibitor* (*rin*) locus. Science 296: 343–346.

Xu, C., K.L. Liberatore, C.A. MacAlister, Z. Huang, Y.-H. Chu, K. Jiang, C. Brooks, M. Ogawa-Ohnishi, G. Xiong, M. Pauly *et al.* (2015) A cascade of arabinosyltransferases controls shoot meristem size in tomato. Nat. Genet. 47: 784–792.

Zamir, D. (2001) Improving plant breeding with exotic genetic libraries. Nat. Rev. Genet. 2: 983–989.