



**HAL**  
open science

## Genes involved in floral meristem in tomato exhibit drastically reduced genetic diversity and signature of selection

Guillaume Bauchet, Stephane Munos, Christopher Sauvage, Julien Bonnet, Laurent Grivet, Mathilde M. Causse

### ► To cite this version:

Guillaume Bauchet, Stephane Munos, Christopher Sauvage, Julien Bonnet, Laurent Grivet, et al.. Genes involved in floral meristem in tomato exhibit drastically reduced genetic diversity and signature of selection. *BMC Plant Biology*, 2014, 14, n.p. 10.1186/s12870-014-0279-2 . hal-02641364

**HAL Id: hal-02641364**

**<https://hal.inrae.fr/hal-02641364v1>**

Submitted on 28 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH ARTICLE

Open Access

# Genes involved in floral meristem in tomato exhibit drastically reduced genetic diversity and signature of selection

Guillaume Bauchet<sup>1,2</sup>, Stéphane Munos<sup>1,3</sup>, Christopher Sauvage<sup>1</sup>, Julien Bonnet<sup>2</sup>, Laurent Grivet<sup>2</sup> and Mathilde Causse<sup>1\*</sup>

## Abstract

**Background:** Domestication and selection of crops have notably reshaped fruit morphology. With its large phenotypic diversity, tomato (*Solanum lycopersicum*) illustrates this evolutive trend. Genes involved in flower meristem development are known to regulate also fruit morphology. To decipher the genetic variation underlying tomato fruit morphology, we assessed the nucleotide diversity and selection footprints of candidate genes involved in flower and fruit development and performed genome-wide association studies.

**Results:** Thirty candidate genes were selected according to their similarity with genes involved in meristem development or their known causal function in *Arabidopsis thaliana*. In tomato, these genes and flanking regions were sequenced in a core collection of 96 accessions (including cultivated, cherry-type and wild relative accessions) maximizing the molecular diversity, using the Roche 454 technology. A total amount of 17 Mb was sequenced allowing the discovery of 6,106 single nucleotide polymorphisms (SNPs). The annotation of the 30 gene regions identified 231 exons carrying 517 SNPs. Subsequently, the nucleotide diversity ( $\pi$ ) and the neutral evolution of each region were compared against genome-wide values within the collection, using a SNP array carrying 7,667 SNPs mainly distributed in coding sequences.

About half of the genes revealed footprints of selection and polymorphisms putatively involved in fruit size variation by showing negative Tajima's D and nucleotide diversity reduction in cultivated tomato compared to its wild relative. Among the candidates, *FW2.2* and *BAM1* sequences revealed selection footprints within their promoter regions suggesting their potential involvement in their regulation. Two associations co-localized with previously identified loci: *LC* (locule number) and *Ovate* (fruit shape).

**Conclusion:** Compared to whole genome genotypic data, a drastic reduction of nucleotide diversity was shown for several candidate genes. Strong selection patterns were identified in 15 candidates highlighting the critical role of meristem maintenance genes as well as the impact of domestication on candidates. The study highlighted a set of polymorphisms putatively important in the evolution of these genes.

**Keywords:** Tomato, *S. lycopersicum*, Meristem, Signature of selection

\* Correspondence: mathilde.causse@avignon.inra.fr

<sup>1</sup>INRA, UR1052, Génétique et Amélioration des Fruits et Légumes (GAFL), 67 Allée des Chênes Domaine Saint Maurice – CS60094, 84143 Montfavet Cedex, France

Full list of author information is available at the end of the article

## Background

Understanding the evolutionary basis of plant variation can be reached through the identification of the molecular mechanisms responsible for the large diversity in plant architecture [1,2]. Evolutionary changes in fruit shape and size has played a key role in the morphological diversification of plant species [3]. Meristem regulation growth is hypothesized to play a major role in sculpting the plant and fruit morphology [4,5]. Its developmental regulation occurs at several levels, including (i) meristem maintenance, (ii) floral organ identity and (iii) floral meristem identity [6-8]. Ovary size partly explained fruit weight, which is first regulated in the meristem [9]. Floral meristem size may impact cell number that will form carpel primordium and subsequent number [10,11]. We hypothesized that variation in genes controlling meristem development and expressed very early in flower/fruit development could be good candidates for fruit size variation.

*Arabidopsis thaliana* is the standard reference for plant biology [12] and a premier model system for molecular and genetic analyses of meristem development [13]. However, the tomato fruit model system proposed by Gillaspay has shown its importance to decipher early developmental determinants, cell cycle steps and organ number determination [14-16]. Together with its ease to cultivate, short life cycle, rich genetic resources, relatively small genome size, available reference genome sequence [17], tomato (*Solanum lycopersicum*) has become a reference in fruit development studies and opens perspectives for a wider understanding of domestication process in fleshy fruit species [18]. Major QTLs involved in the evolution of fruit size and shape have been identified and a few underlying genes cloned [19,20]. For example, the fruit weight QTL *FW2.2* encodes a negative regulator of cell proliferation [21]. Regarding fruit shape, *OVATE* encodes a hydrophilic protein where a single mutation induces a stop codon causing a transition from round to pear shaped tomato fruit [22]. Moreover, two loci, *FASCIATED* and *LC*, that have pleiotropic effects on fruit shape and size [23], determine the locule number: *FASCIATED* encodes a YABBY like transcription factor [24] and *lc* mutation is close to the *WUSCHEL* gene involved in meristem maintenance [25]. Other genes are known [24] or hypothesized to be linked to meristem development [26], but a large genetic potential remains to be revealed [27].

Population genetic studies offer a powerful way to evaluate the molecular evolution of biological mechanisms and to assess the contribution of selection in shaping crop genetic variation and identify related constraints [28,29]. Recently, the Genome Wide Association (GWA) strategy that takes the advantage of natural populations and their increased recombination events [30] has been proposed to decipher the genetic architecture of traits linked to domestication [31]. GWA relies on linkage disequilibrium

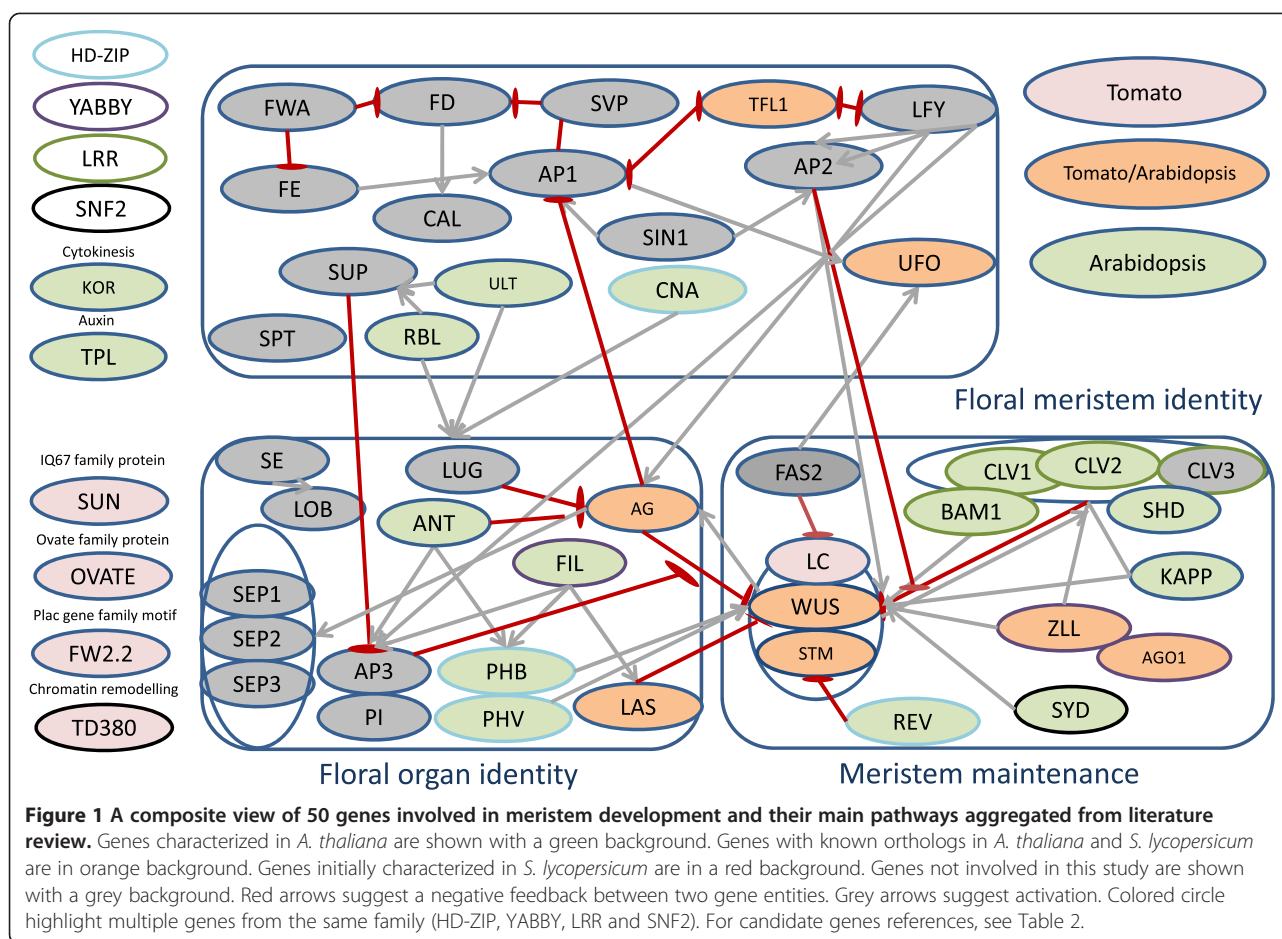
(LD) - non-random association of alleles [32] - and thus on recombination which occurred during meiosis events. In tomato, most recent GWA studies related to fruit size and shape were limited to a single chromosome [33], used a low density marker set [34] or a limited number of agronomical traits collected from public databases [35]. A complementary approach is to compare diversity patterns across species and look for signature of selection over the genome [36].

Here, we describe the patterns of sequence variation of 30 candidate genes in a tomato core collection composed of 96 accessions. The accessions were selected to represent the maximum diversity of a large tomato panel previously described [37]. The set was composed of 17 *S. lycopersicum* (SL) (including Heinz1706, the reference sequenced genome), 63 *S.l. cerasiforme* (SLC), 12 *S. pimpinellifolium* (SP), and four other wild species (WT). Candidate genes were selected for their known function related to tomato fruit size and shape and/or for their involvement in meristem development and maintenance. Using the sequence dataset obtained for the 30 large amplicons covering the genes, nucleotide diversity and signatures of selection were explored. We estimated a set of population genetic parameters (i.e.  $d_N/d_S$ , Tajima's D) to evidence non-neutral processes operating on meristem regulation. We compared these values with those assessed at the whole genome level using a SNP array. Several genes under a strong reduction of diversity in cultivated tomato were identified. Associations with locule number and fruit shape were detected.

## Results

### Candidate gene selection

We first selected 50 candidate genes from the literature. Figure 1 illustrates their classification according to their function and the known interactions between candidates. Among them, 30 genes were retained according to their specificity and success of PCR amplification on the 96 accessions. The 30 candidates included 12 genes involved in meristem maintenance, 6 in floral organ identity and 5 in floral meristem identity. Six other candidates were previously characterized as involved in tomato fruit morphology and two non-coding sequences (one covering the polymorphisms responsible for the *LC* QTL and a non coding region randomly selected, further named as "Non Coding") were also included. Interestingly, we could not identify any ortholog of *CLV3* in tomato. Six genes adjacent to the candidates were also partially covered by long range PCR. They were included in the study as they are closely linked to the candidates. Table 1 lists the candidate genes studied as well as their genomic positions in tomato genome and ortholog ontology in *Arabidopsis thaliana*. Their genomic positions on the reference genome (v2.40) are provided in Additional file 1.



### Sequencing results

About a million of  $\approx 350$  bp reads were generated, while 852,500 reads were aligned onto 174,612 kb of the reference Heinz 1706 sequence (92.5% covered). Roche 454 sequencing process is known to induce a large amount of false INDELs, particularly in homopolymeric regions [108]. For the subsequent analysis, we thus only focused on SNP. Average read depth was 17X while the mapping percentage varied according to taxa from 93.7% in SL to 61.6% for the wild relative *S. pennellii*. Mapping on the reference genome success rate reached 92.7% of the reads. This proportion fell to 61.6%, 76.2%, 83.5%, 86.3% for the four wild accessions of *S. pennellii*, *S. habrochaites*, *S. chmielewskii*, and *S. pimpinellifolium* accessions, respectively. Interestingly *S. cheesmaniae* showed a high mapping rate (94.2%). This result indicates the limit of the alignment procedure for distant accessions. We increased mapping accuracy by *de novo* assembly and aligned 93.5%, 92.7%, 92.2% and 91.9% of WT, respectively, confirming the need to modify the procedure for the wild accessions.

### Sequence annotation

All the amplicons were annotated using ITAG 2.3 and classified as coding regions except the fragment "NON CODING". After Open Reading Frame checking, the exon proportion per fragment ranged from 1% (*LC*) to 44% (*CLV2*). Sequence fragments covered 36 predicted gene entities (30 selected and 6 adjacent genes). Exon number (231) and their average size (170 bp) per candidate gene varied also significantly from 17 bp (*AG*) to 2,600 bp (*BAM1*) and exon number from 1 (*CLV2*, *LAS*) to 25 (*TPL*). Altogether, candidate genes and their flanking unigenes represented 40 kb of coding sequence or 27.8% of the targeted genomic sequence.

### Polymorphism discovery

We detected 3,747 unique SNPs in the three main groups (SL, SLC and SP) and 2,359 SNP by *de novo* assembly in the wild taxa, for a total of 6,106 SNPs. The average SNP density by taxa and accession reached 1 SNP every 2,889 bp for SL, 1 SNP every 1,401 bp for SLC and 1 SNP every 406 bp for SP. Within the wild accessions, *S. cheesmaniae* showed the lowest diversity

**Table 1 Annotation of the candidate genes**

**Targeted tomato candidate genes and related annotation**

| Gene                                  | Gene abbreviation | Arabidopsis gene id | Tomato gene id     | Tomato chromosome | Pathway compartment      | Arabidopsis annotation                | Gene_references |
|---------------------------------------|-------------------|---------------------|--------------------|-------------------|--------------------------|---------------------------------------|-----------------|
| TOPELESS                              | TPL               | AT1G15750.1         | Solyc03g117360.2.1 | SL2.40ch03        | Auxin pathway            | WD-40 repeat protein-like             | [38]            |
| KORRIGAN 1                            | KOR1              | AT5G49720.1         | Solyc01g102580.2.1 | SL2.40ch01        | Cytokinesis              | Endo 1,4 b glucanase                  | [39-41]         |
| CORONA                                | CNA               | AT1G52150.1         | Solyc03g120910.2.1 | SL2.40ch03        | Floral meristem identity | HD ZIP III protein                    | [42-44]         |
| ULTRAPETALA1                          | ULT1              | AT4G28190.1         | Solyc07g054450.2.1 | SL2.40ch07        | Floral meristem identity | SAND-domain transcription factor      | [45]            |
| ARGONAUTE 1                           | AGO1              | AT1G48410.1         | Solyc03g098280.2.1 | SL2.40ch03        | Floral meristem identity | YABBY/AGO                             | [46-48]         |
| REBELOTE                              | RBL               | AT3G55510.1         | Solyc02g081680.2.1 | SL2.40ch02        | Floral meristem identity | unknown function                      | [49]            |
| UNUSUAL FLORAL ORGAN                  | UFO               | AT1G30950.1         | Solyc02g081670.1.1 | SL2.40ch02        | Floral meristem identity | F box protein                         | [50-53]         |
| SELF PRUNING                          | SP/TFL1           | AT5G03840.1         | Solyc06g074350.2.1 | SL2.40ch06        | Floral meristem identity | PEBP protein                          | [54-56]         |
| FILAMENTOUS FLOWER                    | FIL-YAB1          | AT2G45190.1         | Solyc01g091010.2.1 | SL2.40ch01        | Floral organ identity    | YABBY                                 | [57-59]         |
| LATERAL SUPPRESSOR                    | LAS               | AT1G55580.1         | Solyc07g066250.1.1 | SL2.40ch07        | Floral organ identity    | GRAS family transcription factor      | [60-62]         |
| PHABULOSA                             | PHB               | AT2G34710.1         | Solyc02g024070.2.1 | SL2.40ch02        | Floral organ identity    | HD ZIP III                            | [42,63,64]      |
| PHAVOLUTA                             | PHV               | AT1G30490.1         | Solyc08g066500.2.1 | SL2.40ch08        | Floral organ identity    | HD ZIP III                            | [42,63,64]      |
| AINTEGUMENTA                          | ANT               | AT4G37750.1         | Solyc04g077490.2.1 | SL2.40ch04        | Floral organ identity    | AP2/ERF transcription factor family   | [65-67]         |
| AGAMOUS                               | AG                | AT4G18960           | Solyc02g071730.2.1 | SL2.40ch02        | Floral organ identity    | MADS Box transcription factor         | [68-71]         |
| CLAVATA 1                             | CLV1              | AT1G75820.1         | Solyc04g081590.2.1 | SL2.40ch04        | Meristem maintenance     | LRR Receptor kinase                   | [72,73]         |
| KINASE ASSOCIATED PROTEIN PHOSPHATASE | KAPP              | AT5G19280.1         | Solyc01g079720.2.1 | SL2.40ch01        | Meristem maintenance     | kinase-associated protein phosphatase | [74,75]         |
| SHEPERD                               | SHD               | AT4G24190.1         | Solyc04g081570.2.1 | SL2.40ch04        | Meristem maintenance     | ER-resident HSP90-like protein        | [76]            |
| SHOOTMERISTEMLESS                     | STM               | AT1G62360.1         | Solyc02g081120.2.1 | SL2.40ch02        | Meristem maintenance     | KNOX1 homeobox protein                | [77-79]         |
| SPLAYED                               | SYD               | AT2G28290.1         | Solyc02g068560.2.1 | SL2.40ch02        | Meristem maintenance     | SNF2chromatin remodelling protein     | [80]            |
| WUSCHEL                               | WUS               | AT2G17950.1         | Solyc02g083950.2.1 | SL2.40ch02        | Meristem maintenance     | WOX family protein                    | [81-84]         |
| ZWILLE/ ARGONAUTE10                   | ZLL-PNH-AGO10     | AT5G43810.1         | Solyc09g082830.2.1 | SL2.40ch09        | Meristem maintenance     | YABBY/AGO                             | [46,85]         |
| CLAVATA 2                             | CLV2              | AT1G65380.1         | Solyc04g056640.1.1 | SL2.40ch04        | Meristem maintenance     | LRR Receptor like protein             | [86,87]         |
| REVOLUTA                              | REV               | AT5G60690.1         | Solyc11g069470.1.1 | SL2.40ch11        | Meristem maintenance     | HD ZIP III                            | [42,64]         |

**Table 1 Annotation of the candidate genes (Continued)**

| BARELY ANY MERISTEM 1       | BAM1              | AT5G65700.1         | Solyc02g091840.2.1 | SL2.40ch02        | Meristem maintenance | LRR-RLKs kinase                             | [88,89]         |
|-----------------------------|-------------------|---------------------|--------------------|-------------------|----------------------|---|-----------------|
| LOCULE NUMBER               | LC                | AT5G66240.2         | Solyc02g083940.2.1 | SL2.40ch02        | NA                   | non coding region                           | [23,25,26,90]   |
| FRUIT WEIGHT 2.2            | FW2.2/ATPCR2      | AT1G14870.1         | Solyc02g090730.2.1 | SL2.40ch02        | NA                   | Plac gene family motif                      | [21,91-95]      |
| OVATE*                      | OVATE/ATOFP7      | AT2G18500.1         | Solyc02g085500.2.1 | SL2.40ch02        | NA                   | Ovate family Protein                        | [22,58,96-98]   |
| SUN*                        | SUN               | AT5G03960.1         | Solyc10g079240.1.1 | SL2.40ch10        | NA                   | IQ67 family protein                         | [99,100]        |
| TD380*                      | TD380/DDM1        | AT5G66750.1         | Solyc02g085390.2.1 | SL2.40ch02        | NA                   | SNF2 chromatin remodelling protein          | [33,101]        |
| NON CODING                  | NCD               | non coding          | non coding         | SL2.40ch02        | NA                   | non coding region                           | NA              |
| Gene                        | Gene abbreviation | Arabidopsis gene id | Tomato solyc id    | Tomato chromosome |                      | Arabidopsis annotation                      | Gene_references |
| AHRD V1 ***-Q9LIL2_ARATH    | EIF3C             | AT3G56150.2         | Solyc01g102570.2.1 | SL2.40ch01        | NA                   | eukaryotic translation initiation factor 3C | [102]           |
| AHRD V1 *-*-MDTK            | ZF14              | AT1G58340.1         | Solyc02g090740.2.1 | SL2.40ch02        | NA                   | MATE efflux family protein                  | [103]           |
| Ovate protein               | ATOFP6, OFP6      | AT3G52525.1         | Solyc02g085510.1.1 | SL2.40ch02        | NA                   | ovate family protein ... 55 7e-09           | [98]            |
| Adenylosuccinate synthetase | ADSS              | AT3G57610.1         | Solyc02g085520.2.1 | SL2.40ch02        | NA                   | adenylosuccinate synthase                   | [104]           |
| Glucosyltransferase         | UGT71B8           | AT3G21800.1         | Solyc02g081690.1.1 | SL2.40ch02        | NA                   | UDP-glucosyl transferase 71B8               | [105]           |
| ATP dependent RNA helicase  | DEA(D/H) Box      | AT4G16630.1         | Solyc04g081580.2.1 | SL2.40ch04        | NA                   | DEA(D/H)-box RNA helicase family protein    | [106]           |
| Os01g0786800 protein        | Exporter          | AT2G25737.1         | Solyc02g085400.2.1 | SL2.40ch02        | NA                   | Sulfite exporter TauE/Safe family protein   | [107]           |

Gene function, *Arabidopsis thaliana* and tomato gene ID, Tomato chromosome location (v2.40) and bibliographic references. Supplementary genes partially sequenced.

\*Cloned tomato genes.

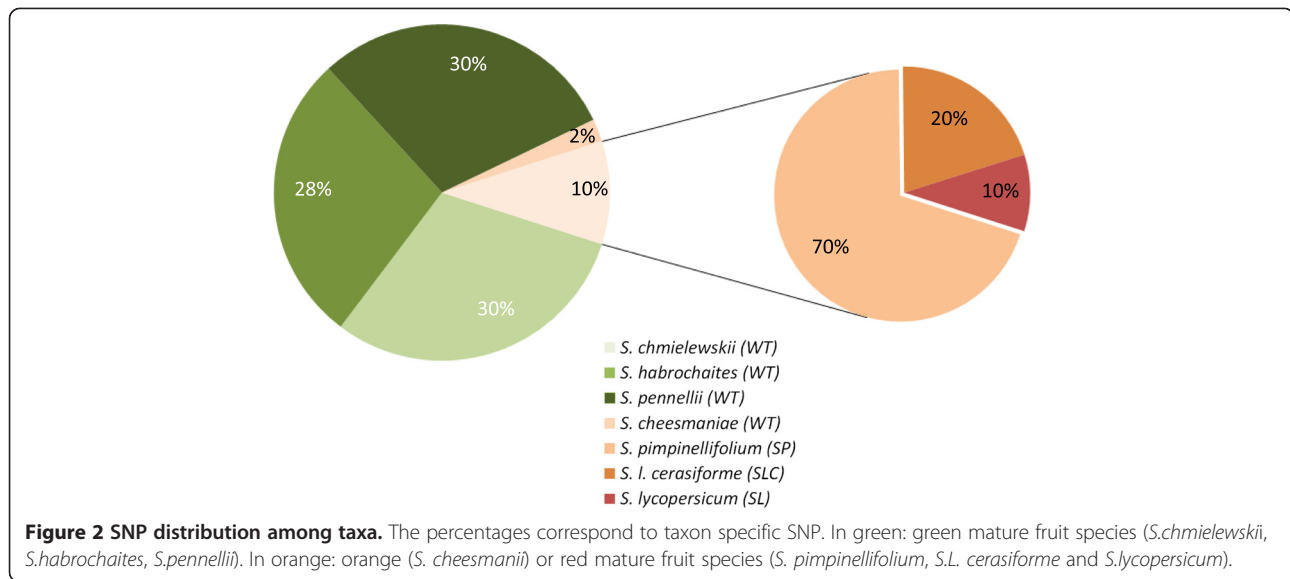
(1 SNP every 1,297 bp). Other wild accessions reached 1 SNP every 96 bp on average (Figure 2). SNP distribution in terms of coding/noncoding region is detailed in Table 2. Regarding SNP identified within the coding sequences (CDS), 423 of these were identified when mapped on the reference (1 SNP every 1,126 bp) and 134 by de novo assembly (1 SNP every 2,169 bp). The most polymorphic locus was *UFO* (1 SNP every 32 bp). The least polymorphic loci were *LC* (1 SNP every 8,807 bp) followed by *AG* (1 SNP every 5,074 bp). We also genotyped the *SL*, *SLC*, and *SP* in the collection (referred further as the 92 accessions) with the SolCAP SNP array, 7,667 SNPs [109]. To perform GWAS, we filtered for rare alleles and missing data and obtained a 5,795 SNPs set. As a cross validation between sequencing and genotyping data, 22 SNP markers of the SolCAP array overlapped the re-sequenced regions. All of them were also identified using the 454 sequencing results. Over the 6,106 SNPs, SnpEff tool identified in the target genes, 432 intragenic (=within CDS) polymorphisms (7%), 568 intergenic (9%), 284

synonymous (4.6%) and 120 corresponded to non-synonymous mutations (2%). More specifically, one synonymous stop (*CNA*), two splice donors (*ZLL*; *REV*), one stop lost (*OVATE*) and one stop gained (*SUN*) were identified. Nine candidate genes (*AG*, *CLV1*, *PHV*, *WUS*, *LC*, *KORI*, *RBL*, *ANT* and *TD380*) did not show any non-synonymous mutation.

#### Population differentiation and structure

Pairwise  $F_{ST}$  on the whole genome dataset was low between *SL* and *SLC* (0.05%) while between *SL* and *SP* and *SLC* and *SP* a stronger differentiation was observed (1.6 and 2% respectively). *SP* and *WT* differentiation rose to 2% and average of *SL* vs *WT* and *SLC* vs *WT* to 5%. These results are supported by the STRUCTURE analysis on red fruit accessions output. Following Evanno's deltaK correction, a two group's population structure was identified as already obtained with a smaller set of SNP [37] (Additional file 2).





### Selection patterns across genetic groups

Nucleotide diversity ( $\pi$  estimates) and neutrality (Tajima's D) were estimated first for each chromosome using the 7,667 SNPs of the SolCAP array (Figure 3) on 92 collection genotypes and related subgroups. Whole chromosome total nucleotide diversity ranged from 0.17 (chromosome 6) to 0.33 (chromosome 4) with an average genome-wide value of 0.27. Intraspecific values were estimated to 0.22, 0.23 and 0.18 for SP, SLC and SL, respectively. The ratio of total nucleotide diversity  $\frac{\pi_{s. lycopersicum}}{\pi_{s. pimpinellifolium}}$  was lower than 1 for all chromosomes but 1 and 2. Tajima's D statistic was positive for all chromosomes but chromosome 9, with significant values for chromosomes 1, 2, 3, 5, 11 ( $P < 0.05$ ) and a whole genome D value of 1.8. Intraspecific D was negative in SL (except on chromosomes 1, 8 and 11) -with lowest value on chromosome 5- and in SP (except on chromosomes 5 and 6). In SLC, D values were all positive except on chromosome 6.

Nucleotide diversity estimation, Tajima's D and codon analysis ( $d_N/d_S$ ) were then performed for the 30 re-sequenced fragments (Additional file 3). Nucleotide diversity between red-orange and green fruited species ( $\pi_{red\_fruit\_species} / \pi_{green\_fruit\_species}$  see Figure 2 for nomenclature) on re-sequenced data was low (0.10). Intraspecific nucleotide diversity estimates were the lowest for SL ( $\pi = 0.0007$ ), followed by SLC ( $\pi = 0.001$ ), SP ( $\pi = 0.002$ ) and the wild types ( $\pi = 0.0120$ ). Re-sequenced fragments showed low and heterogeneous nucleotide diversity, ranging from  $1.65 \times 10^{-4}$  (*FIL*) to  $2 \times 10^{-6}$  (*TPL*). Overall, D-statistic and  $\pi$  values followed a similar trend. Thirteen re-sequenced genes showed a significant D-statistic over the whole collection (Figure 4). The Tajima's D analysis indicated significant evidence for selection in

15 fragments in at least one genetic group (11 fragments for SL, 9 for SLC and 4 for SP) as shown on Figure 5.

According to the codon based analysis, nine fragments showed a  $d_N/d_S$  ratio significantly different from 1 (Table 3). The ratio was lower than one in all gene fragments, six genes showing  $d_N/d_S$  ratio significantly different from 1 displayed also a significant D-statistic on the whole collection. To investigate further the candidates showing multiple signals, we performed a sliding-window approach for the aforementioned tests, allowing an exact positioning of the diversity/neutrality patterns along gene annotation. This allowed the identification of strong negative signals in upstream region of *FW2.2*, *BAM1*, *RBL*, *REV*, *CLV1*, as well as a positive signal in intragenic *OVATE* fragment. Overall, a contrasting pattern within intragenic sequence between SL and SP groups could be observed for *OVATE* and *FW2.2* (Figure 5 and Additional file 4).

### Genome-wide and candidate gene association

Phenotypic data (FW, LC) were previously described for the core collection [33]. Fruit shape index (FSI) was also assessed. Associations were detected by mixed linear model on the dataset including SNP from re-sequenced fragments and SolCAP array. Seven associations were identified after FDR corrections at a whole genome level ( $P < 0.05$ ) involving SNP in 3 fragments and one SolCAP marker, all on chromosome 2 (Table 4). Locule number was associated with six closely linked markers and fruit shape index with only one. Two of these markers were previously identified as causal mutations in *OVATE* and *LC* on chromosome 2 [25,96].

**Table 2 Effects of the SNP detected in coding sequences in the collection of 92 red fruited accessions**

| Gene name | Gene ID            | exon number | CDS lenght | exon length sequenced | SNP in coding sequences | Non synonymous SNPs | Synonymous SNPs | SNP in 3' UTR | SNP in 5' UTR | SNP in splicing site | SNP in STOP codon |
|-----------|--------------------|-------------|------------|-----------------------|-------------------------|---------------------|-----------------|---------------|---------------|----------------------|-------------------|
| AG        | Solyc02g071730.2.1 | 5           | 5074       | 426                   | 3                       | 0                   | 3               | 0             | 0             | 0                    | 0                 |
| ANT       | Solyc04g077490.2.1 | 7           | 3097       | 1893                  | 0                       | 0                   | 0               | 0             | 0             | 0                    | 0                 |
| AGO1      | Solyc03g098280.2.1 | 19          | 6274       | 2650                  | 25                      | 6                   | 19              | 3             | 0             | 0                    | 0                 |
| BAM1      | Solyc02g091840.2.1 | 2           | 3625       | 3514                  | 32                      | 6                   | 26              | 4             | 7             | 0                    | 0                 |
| CLV1      | Solyc04g081590.2.1 | 2           | 3926       | 2436                  | 0                       | 0                   | 0               | 0             | 0             | 0                    | 0                 |
| CLV2      | Solyc04g056640.1.1 | 1           | 2240       | 2240                  | 4                       | 3                   | 1               | 0             | 0             | 0                    | 0                 |
| CNA       | Solyc03g120910.2.1 | 14          | 6463       | 2109                  | 15                      | 4                   | 11              | 2             | 0             | 0                    | 0                 |
| FIL       | Solyc01g091010.2.1 | 7           | 3194       | 1070                  | 7                       | 3                   | 4               | 3             | 3             | 0                    | 0                 |
| FW2.2     | Solyc02g090730.2.1 | 3           | 669        | 610                   | 5                       | 1                   | 4               | 2             | 0             | 0                    | 0                 |
| KAPP      | Solyc01g079720.2.1 | 8           | 14613      | 1118                  | 9                       | 4                   | 5               | 2             | 0             | 0                    | 0                 |
| KOR1      | Solyc01g102580.2.1 | 6           | 3289       | 2509                  | 0                       | 0                   | 0               | 0             | 0             | 0                    | 0                 |
| LAS       | Solyc07g066250.1.1 | 1           | 1286       | 1286                  | 29                      | 6                   | 23              | 0             | 0             | 0                    | 0                 |
| LC        | Solyc02g083940.2.1 | 1           | 5663       | 164                   | 1                       | 0                   | 1               | 0             | 0             | 0                    | 0                 |
| OVATE     | Solyc02g085500.2.1 | 2           | 1383       | 1186                  | 2                       | 2                   | 0               | 3             | 0             | 0                    | 1 (lost)          |
| PHB       | Solyc02g024070.2.1 | 15          | 5882       | 2362                  | 9                       | 4                   | 5               | 0             | 2             | 0                    | 0                 |
| PHV       | Solyc08g066500.2.1 | 15          | 5274       | 2225                  | 0                       | 0                   | 0               | 0             | 0             | 0                    | 0                 |
| RBL       | Solyc02g081680.2.1 | 14          | 4985       | 2650                  | 0                       | 0                   | 0               | 0             | 0             | 0                    | 0                 |
| REV       | Solyc11g069470.1.1 | 18          | 5512       | 2508                  | 43                      | 5                   | 38              | 0             | 0             | 1                    | 0                 |
| SP/TFL1   | Solyc06g074350.2.1 | 2           | 1929       | 510                   | 5                       | 2                   | 3               | 2             | 0             | 0                    | 0                 |
| SHD       | Solyc04g081570.2.1 | 15          | 5102       | 2769                  | 28                      | 6                   | 22              | 4             | 3             | 0                    | 0                 |
| STM       | Solyc02g081120.2.1 | 4           | 3426       | 1592                  | 6                       | 3                   | 3               | 4             | 1             | 0                    | 0                 |
| SYD       | Solyc02g068560.2.1 | 4           | 14883      | 3334                  | 44                      | 22                  | 22              | 1             | 0             | 0                    | 0                 |
| SUN       | Solyc10g079240.1.1 | 5           | 2229       | 1261                  | 15                      | 10                  | 5               | 0             | 0             | 0                    | 1 (gained)        |
| TD380     | Solyc02g085390.2.1 | 5           | 7707       | 577                   | 0                       | 0                   | 0               | 0             | 0             | 0                    | 0                 |
| TPL       | Solyc03g117360.2.1 | 25          | 9952       | 3043                  | 29                      | 4                   | 25              | 9             | 0             | 0                    | 0                 |
| ULT1      | Solyc07g054450.2.1 | 3           | 3273       | 699                   | 4                       | 1                   | 3               | 3             | 2             | 0                    | 0                 |
| UFO       | Solyc02g081670.1.1 | 2           | 1367       | 1367                  | 20                      | 7                   | 13              | 0             | 0             | 0                    | 0                 |
| WUS       | Solyc02g083950.2.1 | 3           | 1238       | 1003                  | 0                       | 0                   | 0               | 0             | 0             | 0                    | 0                 |
| ZLL       | Solyc09g082830.2.1 | 18          | 6153       | 2652                  | 37                      | 8                   | 29              | 0             | 0             | 1                    | 0                 |

**Supplementary genes within resequenced fragments**

|                             |                    |   |      |      |    |   |   |    |   |   |   |
|-----------------------------|--------------------|---|------|------|----|---|---|----|---|---|---|
| Adenylosuccinate synthetase | Solyc02g085520.2.1 | 1 | 8807 | 231  | 1  | 0 | 1 | 0  | 0 | 0 | 0 |
| AHRD V1 - MDTK              | Solyc02g090740.2.1 | 1 | 1608 | 1582 | 13 | 8 | 5 | 5  | 0 | 0 | 0 |
| AHRD V1 - Q9LIL2_ARATH      | Solyc01g102570.2.1 | 1 | 1314 | 62   | 0  | 0 | 0 | 0  | 0 | 0 | 0 |
| ATP dependent RNA helicase  | Solyc04g081580.2.1 | 2 | 6006 | 655  | 4  | 1 | 3 | 15 | 0 | 0 | 0 |
| Glucosyltransferase         | Solyc02g081690.1.1 | 1 | 1445 | 1445 | 11 | 4 | 7 | 0  | 0 | 0 | 0 |
| Os01g0786800 protein        | Solyc02g085400.2.1 | 3 | 2736 | 496  | 3  | 0 | 3 | 3  | 0 | 0 | 0 |
| Ovate protein               | Solyc02g085510.1.1 | 1 | 179  | 179  | 0  | 0 | 0 | 0  | 0 | 0 | 0 |

Effects are detected with SnpEff with tomato annotation ITAG 2.3 version.



## Discussion

We successfully re-sequenced 30 large regions covering candidate loci involved in meristem development and maintenance or corresponding to fruit weight and shape QTL in 96 tomato accessions (92 red-fruited accessions and 4 distant species). We detected a total of 6,106 SNPs within these 30 candidate loci. We also genotyped 5,795 SNPs spread on the whole genome in 92 accessions. Within the wild (SP), admixed (SLC) and cultivated (SL) accessions, the analysis of the nucleotide diversity pattern resulted in two primary conclusions. First, admixed tomato maintained the largest amount of diversity within the collection. Second, the targeted genes showed in average a reduced diversity compared to whole genome values and several strong selection signatures were detected. Moreover, the investigation of selective footprints linked to domestication, in this set of 30 candidate genes related to meristem development, evidenced that a strong purifying selection is at play on this pathway. However the small sample size did not allow us to identify any new association for fruit traits within candidate genes nor in the whole genome data set.

## Polymorphism discovery

Among the studied taxa, polymorphism discovery showed considerable interspecific and intraspecific variations. Red fruited species and *S. cheesmaniae*, showed a drastic reduction of polymorphism compared to green fruited species, as already shown [110]. Overall, 3,747 SNPs were identified by mapping reads on the reference genome and 2,359 SNPs when using *de novo* assembly. Van Deynze and colleagues [111] estimated the nucleotide variation in conserved genes to 1 SNP per 1,627 bp in SLC, 1 per 5,675 bp in fresh market tomatoes and 1 per 851 bp for SP. Our results support these results for SLC (1/1,401 bp) but are sensibly different in SL (1/2,889 bp) and SP (1/406 bp). A possible explanation of this outcome is the difference in the plant material used, as SL and SP in Van Deynze and colleagues [111] are only represented by two and one accessions, respectively. Among the wild type accessions, *S. cheesmaniae* showed the lowest diversity (1 SNP every 1,297 bp). Together with pairwise  $F_{ST}$ , these results support the previously established phylogeny of the *Lycopersicon* complex as well as the domestication scenario and its related bottlenecks [112-114]. Regarding coding mutations, important differences in the number of non-synonymous mutations were observed among candidates. The *OVATE* stop codon was identified as in [22] and could be related to fruit shape variation. SNP modifying splicing sites (*REV* and *ZLL*) may also alter the protein. Lack of polymorphism for some candidates suggested a strong selection pressure especially in meristem maintenance genes (*WUS*, *CLVI*).

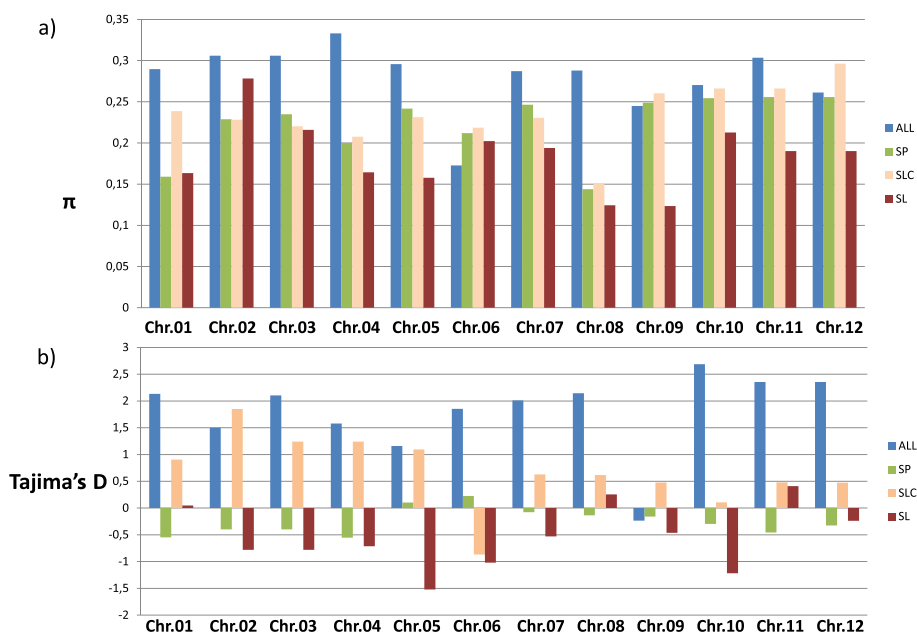
## Nucleotide diversity and selection patterns across genetic groups

Nucleotide diversity ratio showed that rates of alterations varied between genes of the different meristem development pathway compartments with interesting features in the meristem maintenance genes. Intraspecific nucleotide diversity in the SL group is similar to values previously obtained by Labate and colleagues in European germplasm [115]. Over the panel, re-sequenced genes and flanking regions showed a similar profile, with a gradient loss from wild to cultivated species. Nevertheless a large range of variation remains between fragments (Figure 4).

Several significant deviations from the neutral expectation were detected by either analysis, the negative values of Tajima's D and  $d_N/d_S$  ratios smaller than 1 suggested purifying selection, especially on genes from the meristem maintenance compartment where six candidates showed significant D value (Table 3 and Additional file 3). Small sample size, low divergence among lineages and strength of positive selection affect the power of this kind of analysis. However, previous studies in plants suggested that strong purifying selection is a major player in plant genomes. Gossman and colleagues used a genome-wide approach to demonstrate that there is little evidence of adaptive evolution (through the accumulation of adaptive mutations) in many plant species [116]. One of the interpretations suggested by the authors is the small effective size of plant population ( $N_e$ ), which implies that selection may have more impact on the fixation/loss of mutations. In tomato, Städler and colleagues [117] investigated the historical demography of wild tomatoes and demonstrated that the closest wild relative species exhibit concordant signatures of population-size reduction during the evolutionary history. In this context, our results are congruent with these previous observations.

In seven genes (*CLVI*, *FIL*, *LAS*, *TPL*, *REV*, *BAMI*, *SP*), the large and negative Tajima's D test indicated an excess of rare nucleotide polymorphisms with low frequency compared with expectation under neutral theory. This could be explained by the effect of background selection [118], genetic hitchhiking [119] or by an extension of the effective population size ( $N_e$ ) following a bottleneck. For example, *SP* and *LAS* have been previously characterized as key determinants for plant architecture, mutations in these genes inducing strong phenotypic modifications [54,60].

Nucleotide diversity analyses of genes associated with fruit morphology in plants have reported different evolutionary constraints related to gene function and gene fragments. In tomato, the fruit weight QTL *fw3.2* revealed reduced nucleotide diversity in SL and an overall reduced diversity compared with the entire chromosome. The corresponding gene, *SIKLUH*, showed significant local D values (positive and negative), supporting a selective pressure around the gene [120]. In the present



**Figure 3 Diversity revealed by genotyping 5,795 SNP of the SolCAP array. (a)** Chromosomal nucleotide diversity ( $\pi$ ) and **(b)** Tajima's D over the whole collection (ALL) and for 12 *S. pimpinellifolium* (SP), 63 *S.l. cerasiforme* (SLC) and 17 *S. lycopersicum* (SL).

study, candidate genes and their respective chromosomal Tajima's D values were calculated and can be observed taking into account possible genotyping-sequencing platform ascertainment bias as previously observed in other species [121]. In a whole genome comparative transcriptome study of five tomato relatives, Koenig and colleagues identified only 51 genes showing  $d_N/d_S > 1$  [114]. Regarding our gene set, the positive selection evidences underline the rareness of such events. Evolutionary variations of genes involved in traits such as seed/fruit morphology have been reported in other plant species. In *Arabidopsis*, the genetic robustness of cell cycle-related processes was found to be due to functional redundancy more than high selective constraint [29]. In potato, no significant deviation from neutrality was found for genes related to alkaloid pathway,  $d_N/d_S$  ratios close to 1 and negative values of Tajima's D test suggested purifying selection in the gene fragments [122].

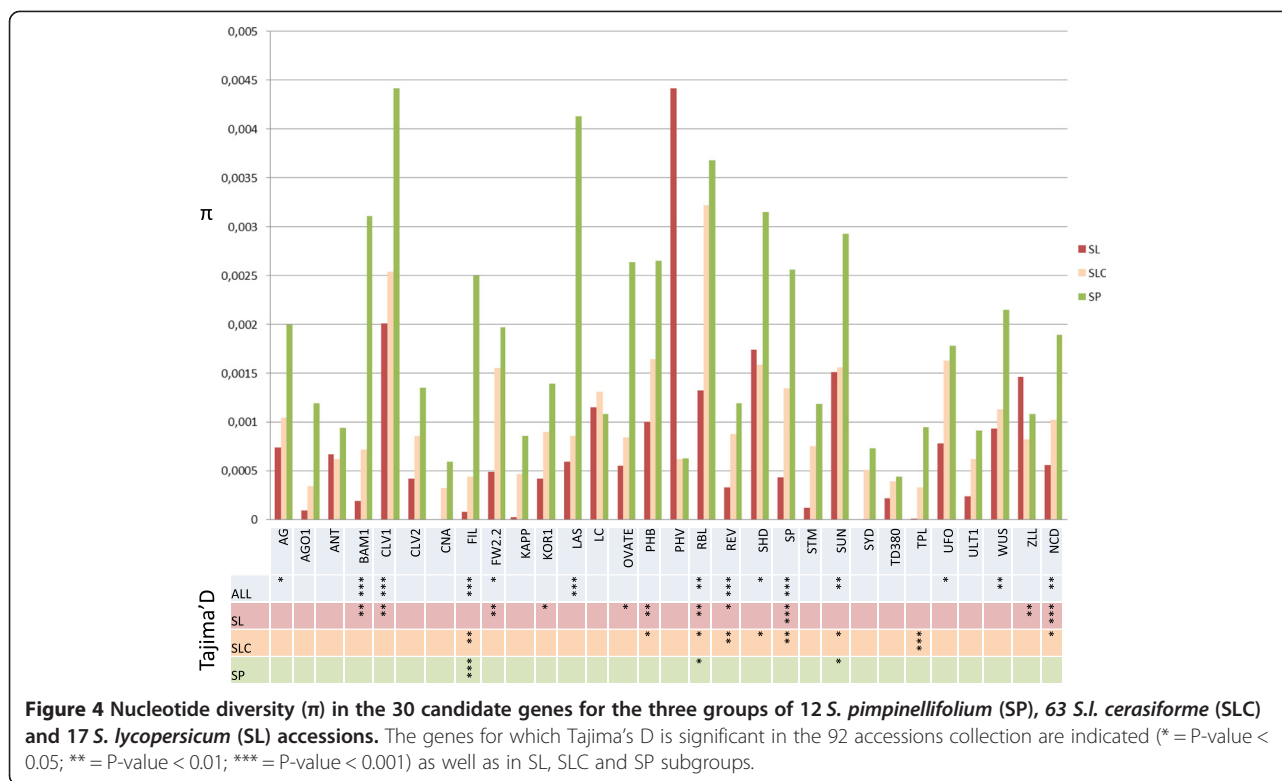
#### Differentiation and population structure

Pairwise  $F_{ST}$  analysis revealed variable trends of differentiation between sub-populations. If differentiation was low between SL and SLC, it was stronger for SL-SP and SLC-SP. These results are consistent with those described by Sim and colleagues [123] between cherry (SLC) and fresh market (SL) tomatoes. However, differentiation between SL and SP had a higher estimate in the aforementioned work. Lower differentiation may be explained by the low sample size of the SP group within our collection. The structure analysis detected two

ancestral groups (SL and SP) and an admixture group composed mainly of SLC accessions. High correlation of the Q estimates (0.94, data not shown) with initial findings on the same panel using a smaller set of SNP markers is comforting results of Ranc and colleagues [33].

#### A few mutations with an important role in fruit size variation

Genome wide association tests for three fruit traits revealed associations with SNP in two intervals surrounding previously described QTL for fruit shape and locule number on chromosome 2. Results from association highlighted two previously identified major loci accounting for fruit shape and size variation, namely *LC* and *OVATE*. We pinpointed the exact mutation of previously identified *LC* and *OVATE* genes (Table 4). We could not detect any other association, particularly with fruit weight, unless decreasing the statistic threshold. Together with the small sample size, a strong relationship between the population structure and fruit weight was shown, hampering the identification of consistent associations new for this trait. Nevertheless non neutral signals of evolution at loci underlying quantitative traits are expected to be different from those due to directional selection [124,125]. Ten genes showed multiple selection signals (Table 4). They include four genes involved in meristem maintenance (*BAMI*, *CLVI*, *REV* and *SYD*), three in floral meristem identity (*RBL*, *CAN* and *UFO*), and three genes previously detected in tomato for their



role in phenotype (*TD380* as a main association with *FW*, *FW2.2* and *OVATE*).

For *FW2.2*, the major tomato fruit weight QTL, the analysis showed signals of selection, including important diversity loss between SL and SP taxa. Tajima's D was strongly significant over the panel (-2.11) and remained significant in the SL group (-2.20). Similarly,  $d_N/d_S$  ratio was close to 1 for the whole fragment and higher than 1 for exon 2, a cysteine-rich motif (24aa) part of a highly conserved core domain [91]. Tajima's D sliding window analysis identified a strong negative peak within the promoter sequence of the *FW2.2* gene (Figure 5a). This finding is supported by the identification of an association signal by Knaap and colleagues in the same region [9]. Taken together, these clues will help to understand the mechanism underlying *FW2.2* regulation which is not yet unravelled. *BAMI* is a *CLAVATA1*-related Leucine rich repeat receptor-like kinases [88]. It is part of the *CLAVATA* regulation complex. It has been demonstrated that *BAM* genes play role in cell division by interacting with *CLAVATA* ligands in the meristem flanking regions [89,126]. *BAMI* has showed the most significant Tajima's D (-2.55) among candidates and low  $d_N/d_S$  (0.0774). Like in *FW2.2* region, a peak was observed in the gene upstream region (Figure 5b). This gene, located in a fruit weight QTL region should be further studied.

## Conclusions

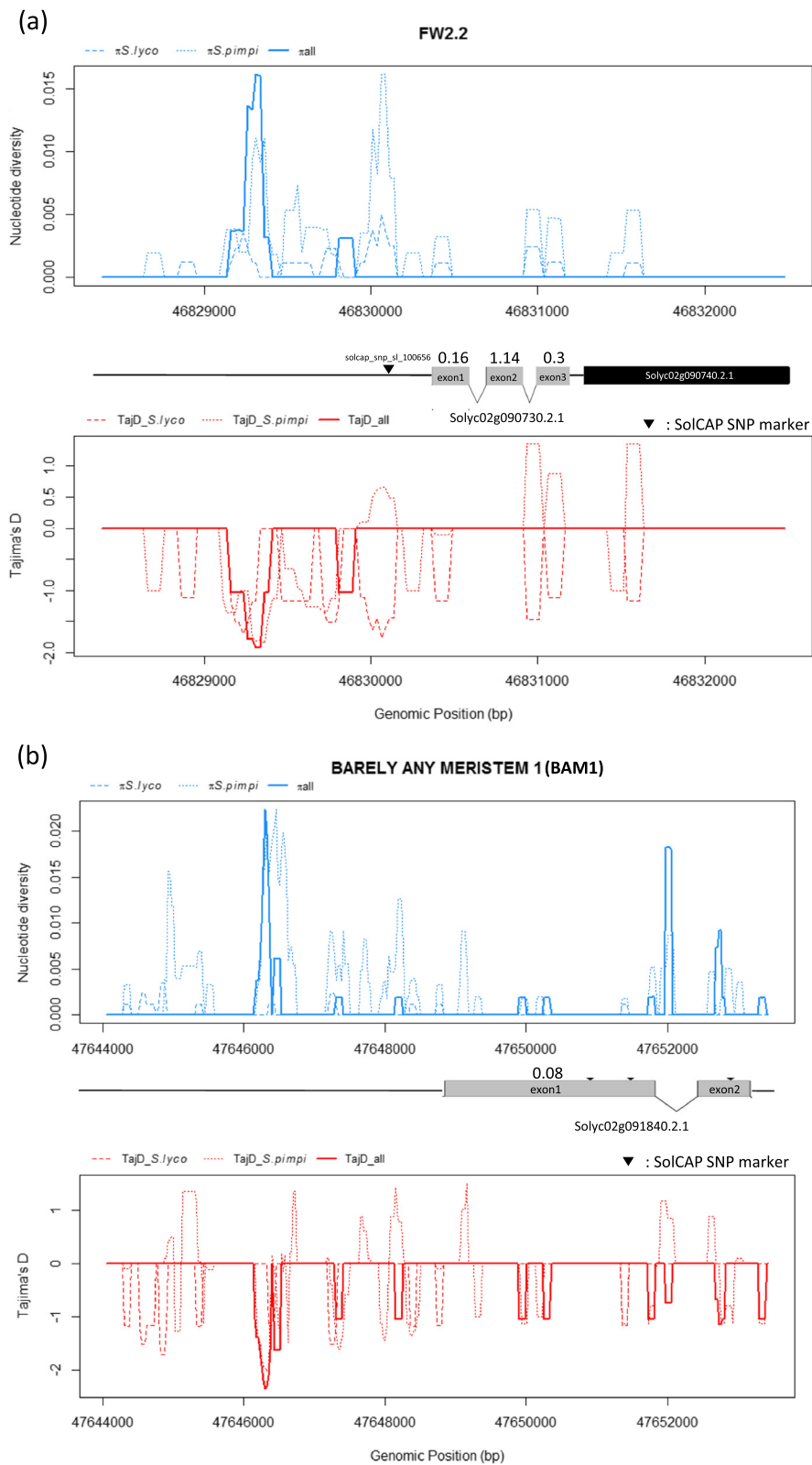
Combining evolutionary metrics and quantitative genetic approach allowed us to decipher the genetic architecture of domestication traits and document their evolutionary history. We identified strong evidence of purifying selection within a few candidate genes with an emphasis on genes related to meristem maintenance. This loss of nucleotide diversity fits previously established domestication scenario [113,114]. Further experiments are required in two ways. The decreasing cost of sequencing will allow large scale GWAS experiments and selective sweep detection at the genome level in a very close future. This will help identifying new candidate loci. For the genes showing patterns of selection, expression profiling and fine scale studies such as methylation studies may uncover their regulation during fruit development as recently shown for the maturation process [127,128].

## Methods

### Selection of candidate genes

Candidate genes were selected following a three steps approach: literature review (1), sequence homology (2) and amplification success rate (3):

First, an extensive literature review identified 50 genes involved in meristem development in *Arabidopsis thaliana*. Related candidate gene protein sequences were extracted to identify their orthologs in tomato. Orthologs



**Figure 5** Sliding-window analysis of nucleotide diversity ( $\pi$ ) -and Tajima's D according to genetic groups for FW2.2 (a) and BAM1 (b) regions. Gene annotation (ITAG 2.3) is displayed. Numbers above exons indicate  $d_N/d_S$  values per exon.

**Table 3 Patterns of variation detected in the genes showing multiple selection signatures, based on 6106 SNP detected in 96 accessions (including wild accessions)**

|  | CLV1     | CNA      | SYD      | REV      | FW2.2    | RBL      | BAM1     | TD380    | UFO      |      |        |      |        |      |        |      |        |      |
|--|----------|----------|----------|----------|----------|----------|----------|----------|----------|------|--------|------|--------|------|--------|------|--------|------|
| <b>Nucleotide diversity (<math>\pi</math>)</b> |          |          |          |          |          |          |          |          |          |      |        |      |        |      |        |      |        |      |
| SL   | 0.001559 | 0        | 0        | 0.000214 | 0.000283 | 0.000745 | 0.000175 | 0.000080 | 0.000518 |      |        |      |        |      |        |      |        |      |
| SLC  | 0.000022 | 0.000007 | 0.000050 | 0.000063 | 0.000015 | 0.000074 | 0.000036 | 0.000080 | 0.000217 |      |        |      |        |      |        |      |        |      |
| SP   | 0.001376 | 0.000079 | 0.000284 | 0.000548 | 0.000579 | 0.000638 | 0.001045 | 0.000171 | 0.000653 |      |        |      |        |      |        |      |        |      |
| WT   | 0.017814 | 0.004996 | 0.010979 | 0.010961 | 0.028111 | 0.011036 | 0.015765 | 0.013036 | 0.012120 |      |        |      |        |      |        |      |        |      |
| <b>Tajima's D</b>                              |          |          |          |          |          |          |          |          |          |      |        |      |        |      |        |      |        |      |
| ALL  | -2.566   | ***      | -1.620   | #        | -1.554   | n.s.     | -2.515   | ***      | -2.112   | *    | -2.214 | **   | -2.555 | ***  | -1.528 | n.s. | -1.897 | *    |
| SL   | -2.177   | **       | n.a.     | n.a.     | n.a.     | n.a.     | -2.042   | *        | -2.205   | **   | -2.131 | **   | -2.302 | **   | -1.069 | n.s. | -1.498 | n.s. |
| SLC  | -1.668   | #        | -1.077   | n.s.     | -1.207   | n.s.     | -2.455   | **       | -1.435   | n.s. | -1.838 | *    | -1.784 | #    | -1.290 | n.s. | -1.007 | n.s. |
| SP   | -0.856   | n.s.     | -1.429   | n.s.     | -1.288   | n.s.     | -0.844   | n.s.     | -1.553   | n.s. | -1.903 | *    | -1.438 | n.s. | -1.712 | #    | -1.490 | n.s. |
| WT   | -0.361   | n.s.     | -0.866   | n.s.     | 0.041    | n.s.     | -0.465   | n.s.     | -0.201   | n.s. | -0.459 | n.s. | -0.708 | n.s. | -0.119 | n.s. | -0.538 | n.s. |
| <b>Syn.-NonSyn. polymorphism</b>               |          |          |          |          |          |          |          |          |          |      |        |      |        |      |        |      |        |      |
| dN   | 0.0076   | 0.0096   | 0.0235   | 0.0046   | 0.0135   | 0.0063   | 0.0013   | 0.0211   | 0.0152   |      |        |      |        |      |        |      |        |      |
| dS   | 0.0511   | 0.0286   | 0.0314   | 0.0228   | 0.0238   | 0.014    | 0.0178   | 0.0351   | 0.0317   |      |        |      |        |      |        |      |        |      |
| dN/dS  | 0.1550   | 0.3356   | 0.7529   | 0.2017   | 0.8074   | 0.4548   | 0.0774   | 0.6015   | 0.4794   |      |        |      |        |      |        |      |        |      |

Nucleotide diversity ( $\pi$ ), Tajima's D and non synonymous (dN), synonymous (dS) and  $dN/dS$  are shown. (n.s., not significant; #,  $P < 0.10$ ; \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ ).

were obtained from NCBI database ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) using TBLASTN procedure (Additional file 5). The output data was sorted according to e-values and bit score. Candidate genes without a match were screened using TBLASTN on the tomato scaffolds genome assembly (v2.40) (see step (2) in the Additional file 1). Reciprocal BLASTN between query and subjects was performed to support the similarity (Additional file 6). For all orthologous sequences, a BLASTN was used to identify their corresponding candidate genes in the tomato genome including the flanking promoter and 3' UTR sequences. Final selection was based on amplification success rate (>90 individuals amplified) and specificity (single PCR product).

#### Plant materials

A total of 96 accessions (Additional file 7) were selected to represent the maximum diversity of a larger collection drawn from 360 accessions previously described in [37]. The set was composed of 63 *S.l. cerasiforme* (SLC); 12 *S. pimpinellifolium* (SP); 17 *S. lycopersicum* (SL) (including Heinz1706, the reference sequenced genome) and four wild relatives (WT) *S. pennellii* (LA716), *S. habrochaites* (PI247087), *S. chmielewskii* (LA1840) and *S. cheesmaniae* (LA1401). Accessions were derived from French researchers' prospecting, breeders' collections, the Tomato Genetics Resource Center (Davis, USA), the Centre for Genetic Resources (Wageningen, The Netherlands), the North Central Regional Plant Introduction Station (Ames,

**Table 4 List of associations detected on 90 accessions with locule number and fruit shape index**

| Phenotype         | SNP_ID              | Gene  | Chromosome | SNP Position (bp) | p-value               |
|-------------------|---------------------|-------|------------|-------------------|-----------------------|
| Locule number     | solcap_snp_sl_23925 |       | 2          | 41345621          | $4.95 \times 10^{-4}$ |
|                   | WUS_SNP_T_A_2117    | WUS   | 2          | 41765967          | $1.28 \times 10^{-3}$ |
|                   | LC_SNP_T_A_3774     | LC    | 2          | 41765967          | $1.28 \times 10^{-3}$ |
|                   | WUS_SNP_G_A_2168    | WUS   | 2          | 41766018          | $1.07 \times 10^{-2}$ |
|                   | WUS_SNP_T_C_3869    | WUS   | 2          | 41767719          | $1.07 \times 10^{-2}$ |
|                   | WUS_SNP_T_A_3979    | WUS   | 2          | 41767829          | $4.95 \times 10^{-4}$ |
| Fruit shape index | OVATE_SNP_A_G_269   | OVATE | 2          | 42944775          | $1.64 \times 10^{-2}$ |

P-values according to the false discovery rate procedure (FDR).



IA) and the N.I. Vavilov Research Institute of Plant Industry (St. Petersburg, Russia). Accessions are characterized and maintained at INRA, Avignon, France. Phenotyped traits (FW and locule number) data were collected from [33]. Tomato Analyzer V2.1.0.0 software [129] was implemented to scan fruit morphology within the 96 accessions. Then, fruit shape index (FSI, ratio of maximum diameter/height) was analyzed. For each the three phenotypic traits, year and accession effect were statistically corrected using Anova using the [R] software ([www.R-project.org](http://www.R-project.org)). Adjusted mean was calculated by “all.effects” procedure package implemented in [R].

#### DNA isolation and sequencing

Genomic DNA was isolated from 100 mg of frozen leaves using the DNeasy Plant Mini Kit (QIAGEN, Valencia, CA) according the manufacturer's recommendations. DNA titration was performed using fluorescence. We used long range PCR (LR-PCR) to amplify DNA sequences (5-10 kb) and cover candidate genes and their potential regulatory regions. Amplification primers were designed in Primer3 (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>), see Additional file 8 for a list of oligonucleotides. Amplification reactions were performed in a final volume of 50  $\mu$ L in a reaction mix composed of 10 ng of template DNA, 10 pmol of each primer, 100 mM concentration of each deoxynucleotide, 5X Taq polymerase buffer P, and 1 unit of Taq polymerase Herculase II (Agilent, CA, USA). After 5 min of denaturation at 95°C, 35 cycles were performed with initial denaturation (20 s at 95°C), annealing during 20 s at 58°C, extension during 6 min at 68°C, followed by a final extension step of 8 min at 68°C. All PCR amplifications were checked on agarose gel (1%, 120 mV, 40 min). All successful and specific PCR products were selected and quantified using Quant'it picogreen dsDNA Assay kit (Invitrogen, Eugene, Oregon, USA) on a fluorescent plate reader (Perkin Elmer 2103 Multilabel reader). Pairs of primers revealing single-band polymerase chain reaction (PCR) product were chosen for sequencing. The thirty PCR fragments were pooled by accession in equimolar quantity. The DNA concentration of each pool was then adjusted to a final concentration of 0.5  $\mu$ M (in a 100  $\mu$ l final volume). These 96 pools were used to obtain the corresponding 454 libraries.

Each DNA library was fragmented by high pressure nitrogen flow to a 300-500 bp size range [130]. Fragmentation quality assessment was performed on an Agilent Bioanalyzer (Agilent technologies, USA). Each library was tagged using a specific sequence tag (GS Rapid Library Prep Kit, Roche diagnostics, Basel, Switzerland). Sequencing experiment was defined as followed: on the 454 sequencing picotiter plate, 8 regions (gaskets), each one containing 12 pools, each pool identified with a

specific sequence tag [131]. Serial dilution and fine quantification was performed with Biomark Slingshot method (Fluidigm, San Francisco, California, USA). Emulsion-based amplification, GS-FLX library sequencing performed as described by Margulies [132]. Library preparation and 454 GS-FLX pyrosequencing (Roche diagnostics, 454 life science corp., Brandford, Connecticut, USA) were performed at Genotoul Genomic (<http://www.genotoul.fr>, INRA Toulouse, France).

#### Read mapping, *de novo* assembly and polymorphism discovery

Checking for contaminants and quality trimming was performed using PyroCleaner software suite [133]. Assembly and polymorphism discovery were performed using NGen<sup>®</sup> version 3 (DNASTAR, Madison, WI, USA) [134]. Reads were mapped on the reference genome V2.4 from the Solanaceae Genomics Network ([http://solgenomics.net/organism/Solanum\\_lycopersicum/genome](http://solgenomics.net/organism/Solanum_lycopersicum/genome)). To improve mapping accuracy of wild accessions, a *de novo* assembly was performed using a BLAST-like Alignment Tool (BLAT) procedure [135]. Genome annotation 2.3 version produced by the International Tomato Annotation Group ([http://solgenomics.net/organism/Solanum\\_lycopersicum/genome](http://solgenomics.net/organism/Solanum_lycopersicum/genome)) was used to predict gene sequence architecture. We used SnpEff [136] to classify polymorphisms into non coding or coding polymorphisms (either synonymous or non-synonymous). Genes were also checked for open reading frame using ORF finder (<http://www.ncbi.nlm.nih.gov/projects/gorf/>). The longest ORF were kept for subsequent analysis. Polymorphisms were selected with a minimal coverage of 10x and polymorphism occurrence higher than 90%.

#### Whole genome genotyping using the SolCAP array

Whole genome SNP genotyping was performed using Infinium assay (Illumina) developed by the Solanaceae Coordinated Agricultural Project -SolCAP- [109,137] as described in [138]. Probe sequences and related information are available from SolCAP (<http://solcap.msu.edu>). The SNP calling rate threshold per locus was set to 90%. Among the 8,784 SNPs from the SolCAP array, 7,667 SNPs passed the quality control. This SNP dataset –without Minor Allele Frequency filtering (MAF) was considered as a neutral dataset, a comparative basis for the candidate genes. To perform GWAS, filtering for low MAF (10% threshold) and missing data (10%), 5,795 SNPs were performed.

#### Estimation of population differentiation, structure and relatedness

Sequencing and genotyping data on the collection of 96 accessions were subjected to genetic diversity indices calculation. Total nucleotide diversity ( $\pi$ ) and Tajima's D test [139] were computed on the collection and genetic



subgroups (SL, SLC and SP) using DNAsp [140] and Variscan [141] software, using global calculation per chromosome for SolCAP array and sliding window on re-sequenced genes. On re-sequencing data, the  $d_N/d_S$  neutrality test was performed on the synonymous ( $d_N$ ) to non-synonymous ( $d_S$ ) substitution rates [142,143] using PAML [144] with the YN00 module and a neighbor joining phylogenetic tree calculated on genotyping dataset to calibrate the  $d_S$  (Additional file 9). This ratio provides insights of selective pressures acting on protein-coding regions and allows identifying positive selection ( $d_N/d_S > 1$ ) or purifying selection ( $d_N/d_S < 1$ ). The pairwise-population fixation differentiation index,  $F_{st}$  [145], was assessed on the core collection between the three groups of accessions SL, SLC and SP and an outgroup constituted of the four wild types.

Population stratification (Q matrix) was defined with STRUCTURE [146] and Evanno's correction [147] using the whole genome genotyping data. Simulations were ran with group number ranging from  $K = 1$  to  $K = 10$ . Ten replicates (burnin period: 100,000 and MCMC step: 500,000) using the Bioportal computing cluster (<http://app3.titan.uio.no/>) with parameters as described in [148] Pairwise kinship coefficient calculation matrix (K matrix) was performed using Spagedi [149] following the Ritland calculation method [150].

### GWA mapping

Associations between polymorphisms and adjusted means of fruit traits were screened with TASSEL v3.0 [151]. For each trait, a mixed linear model (MLM) accounting for kinship (K matrix) and population structure (Q matrix) was used. Significance levels for multiple tests was corrected using FDR [152] correction method with 5% threshold.

### Data availability

The SNP genotype and phenotype datasets as well as *ad hoc* Q and K matrixes are deposited on the GNPis repository hosted <https://urgi.versailles.inra.fr/association/association/viewer.do#panelCard/id=3> [153].

### Additional files

**Additional file 1: Physical location of the candidate genes onto the tomato genome.** All sequences have been aligned on the tomato genome (v2.40) using BLAST. The distances are indicated in Mb.

**Additional file 2: Structure graphical outputs on 90 accessions based on different genotypic datasets.** (a) SolCAP genotyping data; b) re-sequenced genotypic data; (c) SNP data from Xu *et al.* 2012). Fruit weight variation is displayed in black line.

**Additional file 3: Nucleotide diversity ( $\pi$ ), Tajima' D and  $d_N/d_S$  for each candidate, according to the genetic groups: whole collection, SP, SLC and SL groups.**

**Additional file 4: Sliding window analysis of nucleotide diversity ( $\pi$ ) and Tajima's D according to genetic groups for CLV1 (a), RBL (b), REV (c) and OVATE (d) regions.** Gene annotation (ITAG 2.3) is displayed. Numbers above exons indicate  $d_N/d_S$  values per exon. Red '\*' indicates the presence of a STOP codon.

**Additional file 5: Pipeline used for orthologous gene identification from *A. thaliana* to *S. lycopersicum*.**

**Additional file 6: List of Arabidopsis candidate genes proteins (TAIR 10).** Output of the TBLASTn on tomato genome sequence (v2.40).

**Additional file 7: Information on the 96 accessions used in the association study and diversity analysis: List of accessions, subgroups and phenotypic data.** Species are indicated as well as values for fruits weigh (FW), locule number (LCN) and fruit shape index (FSI).

**Additional file 8: Primer list for re-sequenced candidates (v.2.40).**

**Additional file 9: NJ tree (1000 bootstrap) on 96 accessions.** In red, *S. lycopersicum* (SL) accessions; in orange, *S. lycopersicum* var. *cerasiforme* (SLC) accessions; in green: *S. pimpinellifolium* (SP) accessions. In black, outgroup formed of four wild species accessions (WT): *S. chmielewskii*, *S. habrochaites*, *S. peruvianum* and *S. pennellii*.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

SM, LG, JB and GB conceived and designed the experiments. GB and SM performed the experiments. GB, SM and CS analysed the data. GB wrote the paper. MC and CS assisted in editing the article. All authors read and approved the final manuscript.

### Acknowledgements

We acknowledge H el ene Burck and Yolande Carretero from INRA GAFL for their invaluable help in characterizing and maintaining the INRA tomato Genetic Resources collection. Authors are also grateful to Jerome Lluch (INRA, Genotoul), Olivier Bouchez (INRA Genotoul) Jean-Paul Bouchet (INRA GAFL), and Celine Noirot (INRA Genotoul) for sequencing support and pre-analysis. Syngenta Seeds funded this work. We thank the two anonymous reviewers for their useful comments.

### Author details

<sup>1</sup>INRA, UR1052, G en etique et Am elioration des Fruits et L egumes (GAFL), 67 All ee des Ch enes Domaine Saint Maurice – CS60094, 84143 Montfavet Cedex, France. <sup>2</sup>Syngenta Seeds, 12 chemin de l'Hobit, 31790 Saint Sauveur, France. <sup>3</sup>Present address: INRA, UMR CNRS-INRA 441-2594, 24 Chemin de Borde Rouge – Auzeville - CS 52627, 31326 Castanet Tolosan Cedex, France.

Received: 19 June 2014 Accepted: 6 October 2014

Published online: 19 October 2014

### References

1. Doebley JF, Gaut BS, Smith BD: The molecular genetics of crop domestication. *Cell* 2006, **127**(7):1309–1321.
2. Burger JC, Chapman MA, Burke JM: Molecular insights into the evolution of crop plants. *Am J Bot* 2008, **95**(2):113–122.
3. Tanksley SD: The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. *Plant Cell* 2004, **16**(Suppl 1):S181–S189.
4. Girin T, Sorefan K, Østergaard L: Meristematic sculpting in fruit development. *J Exp Bot* 2009, **60**(5):1493–1502.
5. Park SJ, Eshed Y, Lippman ZB: Meristem maturation and inflorescence architecture: lessons from the Solanaceae. *Curr Opin Plant Biol* 2014, **17**:70–77.
6. Becker A, Alix K, Damerval C: The evolution of flower development: current understanding and future challenges. *Ann Bot* 2011, **107**(9):1427–1431.
7. Kaufmann K, Pajoro A, Angenent GC: Regulation of transcription in plants: mechanisms controlling developmental switches. *Nat Rev Genet* 2010, **11**(12):830–842.
8. Casurier B, Schwarz-Sommer Z, Davies B: Floral organ identity: 20 years of ABCs. *Semin Cell Dev Biol* 2010, **21**(1):73–79.

9. van der Knaap E, Chakrabarti M, Chu YH, Clevenger JP, Illa Berenguer E, Huang Z, Keyhaninejad N, Mu Q, Sun L, Wang Y, Wu S: **What lies beyond the eye: the molecular mechanisms regulating tomato fruit weight and shape.** *Front Plant Sci* 2014, **5**(227):1–13.
10. Heidstra R, Sabatini S: **Plant and animal stem cells: similar yet different.** *Nat Rev Mol Cell Biol* 2014, **15**(5):301–312.
11. Perales M, Reddy GV: **Stem cell maintenance in shoot apical meristems.** *Curr Opin Plant Biol* 2012, **15**(1):10–16.
12. Koornneef M, Meinke D: **The development of Arabidopsis as a model plant.** *Plant J* 2010, **61**(6):909–921.
13. Irish VF: **The flowering of Arabidopsis flower development.** *Plant J* 2010, **61**(6):1014–1028.
14. Gillaspay G, Ben-David H, Gruissem W: **Fruits: a developmental perspective.** *Plant Cell Online* 1993, **5**(10):1439–1451.
15. Xiao H, Radovich C, Welty N, Hsu J, Li D, Meulia T, van der Knaap E: **Integration of tomato reproductive developmental landmarks and expression profiles, and the effect of SUN on fruit shape.** *BMC Plant Biol* 2009, **9**(1):49.
16. Bertin N, Causse M, Brunel B, Tricon D, Genard M: **Identification of growth processes involved in QTLs for tomato fruit size and composition.** *J Exp Bot* 2008, **60**(1):237–248.
17. The Tomato Genome Consortium: **The tomato genome sequence provides insights into fleshy fruit evolution.** *Nature* 2012, **485**(7400):635–641.
18. Ranjan A, Ichihashi Y, Sinha N: **The tomato genome: implications for plant breeding, genomics and evolution.** *Genome Biol* 2012, **13**(8):167.
19. Rodriguez GR, Munos S, Anderson C, Sim SC, Michel A, Causse M, Gardener BBM, Francis D, van der Knaap E: **Distribution of SUN, OVATE, LC, and FAS in the Tomato Germplasm and the Relationship to Fruit Shape Diversity.** *Plant Physiol* 2011, **156**(1):275–285.
20. Monforte AJ, Diaz AI, Caño-Delgado A, van der Knaap E: **The genetic basis of fruit morphology in horticultural crops: lessons from tomato and melon.** *J Exp Bot* 2014, doi:10.1093/jxb/eru017.
21. Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD: **fw2.2: a quantitative trait locus key to the evolution of tomato fruit size.** *Science* 2000, **289**(5476):85–88.
22. Liu J, Van Eck J, Cong B, Tanksley SD: **A new class of regulatory genes underlying the cause of pear-shaped tomato fruit.** *Proc Natl Acad Sci U S A* 2002, **99**(20):13302–13306.
23. Lippman Z, Tanksley SD: **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* 2001, **158**(1):413–422.
24. Cong B, Barrero LS, Tanksley SD: **Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication.** *Nat Genet* 2008, **40**(6):800–804.
25. Muñoz S, Ranc N, Botton E, Bérard A, Rolland S, Duffé P, Carretero Y, Le Paslier M-C, Delalande C, Bouzayen M, Brunel D, Causse M: **Increase in tomato locule number is controlled by two single-nucleotide polymorphisms located near WUSCHEL.** *Plant Physiol* 2011, **156**(4):2244–2254.
26. Barrero LS, Cong B, Wu F, Tanksley SD: **Developmental characterization of the fasciated locus and mapping of Arabidopsis candidate genes involved in the control of floral meristem size and carpel number in tomato.** *Genome* 2006, **49**(8):991–1006.
27. Zamir D: **Plant breeders go back to nature.** *Nat Genet* 2008, **40**(3):269–270.
28. Wilson LM, Whitt SR, Ibáñez AM, Rocheford TR, Goodman MM, Buckler ES: **Dissection of maize kernel composition and starch production by candidate gene association.** *Plant Cell* 2004, **16**(10):2719–2733.
29. Sterken R, Kiekens R, Coppens E, Vercauteren I, Zabeau M, Inzé D, Flowers J, Vuylsteke M: **A Population genomics study of the arabidopsis core cell cycle genes shows the signature of natural selection.** *Plant Cell Online* 2009, **21**(10):2987–2998.
30. Korte A, Farlow A: **The advantages and limitations of trait analysis with GWAS: a review.** *Plant Methods* 2013, **9**(1):29.
31. Mandel JR, Nambesani S, Bowers JE, Marek LF, Ebert D, Rieseberg LH, Knapp SJ, Burke JM: **Association mapping and the genomic consequences of selection in sunflower.** *PLoS Genet* 2013, **9**(3):e1003378.
32. Kimura M: **A model of a genetic system which leads to closer linkage by natural selection.** *Evolution* 1956, **10**(278):87.
33. Ranc N, Muñoz S, Xu J, Le Paslier M-C, Chauveau A, Bounon R, Rolland S, Bouchet J-P, Brunel D, Causse M: **Genome-wide association mapping in tomato (*Solanum lycopersicum*) is possible using genome admixture of *Solanum lycopersicum* var. *Cerasiforme*.** *G3: Genes|Genomes|Genetics* 2012, **2**(8):853–864.
34. Xu J, Ranc N, Muñoz S, Rolland S, Bouchet J-P, Desplat N, Paslier M-C, Liang Y, Brunel D, Causse M: **Phenotypic diversity and association mapping for fruit quality traits in cultivated tomato and related species.** *Theor Appl Genet* 2012, **126**(3):567–581.
35. Shirasawa K, Fukuoka H, Matsunaga H, Kobayashi Y, Kobayashi I, Hirakawa H, Isobe S, Tabata S: **Genome-Wide Association studies using single nucleotide polymorphism markers developed by re-sequencing of the genomes of cultivated tomato.** *DNA Research* 2013, **20**(6):593–603.
36. Sim S-C, Van Deynze A, Stoffel K, Douches DS, Zarka D, Ganai MW, Chetelat RT, Hutton SF, Scott JW, Gardner RG, Panthee DR, Mutschler M, Myers JR, Francis DM: **High-Density SNP genotyping of tomato reveals patterns of genetic variation due to breeding.** *PLoS One* 2012, **7**(9):e45520.
37. Ranc N, Muñoz S, Santoni S, Causse M: **A clarified position for *Solanum lycopersicum* var. *cerasiforme* in the evolutionary history of tomatoes (*Solanaceae*).** *BMC Plant Biol* 2008, **8**(1):130.
38. Long JA, Ohno C, Smith ZR, Meyerowitz EM: **TOPLESS regulates apical embryonic fate in Arabidopsis.** *Science* 2006, **312**(5779):1520–1523.
39. Zuo J, Niu Q-W, Nishizawa N, Wu Y, Kost B, Chua N-H: **KORRIGAN, an Arabidopsis endo-1,4-β-glucanase, localizes to the cell plate by polarized targeting and is essential for cytokinesis.** *Plant Cell Online* 2000, **12**(7):1137–1152.
40. Krupková E, Schmölling T: **Developmental consequences of the tumorous shoot development1 mutation, a novel allele of the cellulose-synthesizing KORRIGAN1 gene.** *Plant Mol Biol* 2009, **71**(6):641–655.
41. Brummell DA, Catala C, Lashbrook CC, Bennett AB: **A membrane-anchored E-type endo-1,4-β-glucanase is localized on Golgi and plasma membranes of higher plants.** *Proc Natl Acad Sci* 1997, **94**(9):4794–4799.
42. Prigge MJ, Otsuga D, Alonso JM, Ecker JR, Drews GN, Clark SE: **Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in Arabidopsis development.** *Plant Cell Online* 2005, **17**(1):61–76.
43. Green KA: **CORONA, a Member of the Class III Homeodomain Leucine Zipper gene family in Arabidopsis, regulates stem cell specification and organogenesis.** *Plant Cell Online* 2005, **17**(3):691–704.
44. Barton MK: **Twenty years on: the inner workings of the shoot apical meristem, a developmental dynamo.** *Dev Biol* 2010, **341**(1):95–113.
45. Fletcher JC: **The ULTRAPETALA gene controls shoot and floral meristem size in Arabidopsis.** *Development* 2001, **128**(8):1323–1333.
46. Bai M, Yang G-S, Chen W-T, Mao Z-C, Kang H-X, Chen G-H, Yang Y-H, Xie B-Y: **Genome-wide identification of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families and their expression analyses in response to viral infection and abiotic stresses in *Solanum lycopersicum*.** *Gene* 2012, **501**(1):52–62.
47. Kidner CA, Martienssen RA: **The role of ARGONAUTE1 (AGO1) in meristem formation and identity.** *Dev Biol* 2005, **280**(2):504–517.
48. Bohmert K, Camus I, Bellini C, Bouchez D, Caboche M, Benning C: **AGO1 defines a novel locus of Arabidopsis controlling leaf development.** *EMBO J* 1998, **17**(1):170–180.
49. Prunet N, Morel P, Thierry AM, Eshed Y, Bowman JL, Negrutiu I, Trehin C: **REBELOTE, SQUINT, and ULTRAPETALA1 function redundantly in the temporal regulation of floral meristem termination in Arabidopsis thaliana.** *Plant Cell Online* 2008, **20**(4):901–919.
50. Wilkinson MD, Haughn GW: **UNUSUAL FLORAL ORGANS controls meristem identity and organ primordia fate in Arabidopsis.** *Plant Cell Online* 1995, **7**(9):1485–1499.
51. Hepworth S, Klenz J, Haughn G: **UFO in the Arabidopsis inflorescence apex is required for floral-meristem identity and bract suppression.** *Planta* 2006, **223**(4):769–778.
52. Allen K, Sussex I: **Falsiflora and anantha control early stages of floral meristem development in tomato (*Lycopersicon esculentum* Mill.).** *Planta* 1996, **200**(2):254–264.
53. Lippman ZB, Weigel D, Cohen O, Alvarez JP, Abu-Abied M, Pekker I, Paran I, Eshed Y, Zamir D: **The making of a compound inflorescence in tomato and related Nightshades.** *PLoS Biol* 2008, **6**(11):e288.
54. Pnueli L, Carmel-Goren L, Hareven D, Gutfinger T, Alvarez J, Ganai M, Zamir D, Lifschitz E: **The SELF-PRUNING gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of CEN and TFL1.** *Development* 1998, **125**(11):1979–1989.

55. Yeager A: **Studies on the inheritance and development of fruit size and shape in the tomato.** *J Agric Res* 1937, **55**:141–152.
56. Shannon S, Meeks-Wagner DR: **A Mutation in the Arabidopsis TFL1 gene affects inflorescence meristem development.** *Plant Cell Online* 1991, **3**(9):877–892.
57. Goldshmidt A, Alvarez JP, Bowman JL, Eshed Y: **Signals derived from YABBY gene activities in organ primordia regulate growth and partitioning of arabidopsis shoot apical meristems.** *Plant Cell Online* 2008, **20**(5):1217–1230.
58. Huang Z, Houten J, Gonzalez G, Xiao H, Knaap E: **Genome-wide identification, phylogeny and expression analysis of SUN, OFP and YABBY gene family in tomato.** *Mol Genet Genomics* 2013, **288**(3-4):111–129.
59. Bartholmes C, Hidalgo O, Gleissberg S: **Evolution of the YABBY gene family with emphasis on the basal eudicot *Eschscholzia californica* (Papaveraceae).** *Plant Biol* 2012, **14**(1):11–23.
60. Schumacher K, Schmitt T, Rossberg M, Schmitz G, Theres K: **The Lateral suppressor (Ls) gene of tomato encodes a new member of the VHHD protein family.** *Proc Natl Acad Sci* 1999, **96**(1):290–295.
61. Greb T: **Molecular analysis of the LATERAL SUPPRESSOR gene in Arabidopsis reveals a conserved control mechanism for axillary meristem formation.** *Genes Dev* 2003, **17**(9):1175–1187.
62. Pysh LD, Wysocka-Diller JW, Camilleri C, Bouchez D, Benfey PN: **The GRAS gene family in Arabidopsis: sequence characterization and basic expression analysis of the SCARECROW-LIKE genes.** *Plant J* 1999, **18**(1):111–119.
63. McConnell JR, Emery J, Eshed Y, Bao N, Bowman J, Barton MK: **Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots.** *Nature* 2001, **411**(6838):709–713.
64. Emery JF, Floyd SK, Alvarez J, Eshed Y, Hawker NP, Izhaki A, Baum SF, Bowman JL: **Radial patterning of Arabidopsis shoots by class III HD-ZIP and KANADI genes.** *Curr Biol* 2003, **13**(20):1768–1774.
65. Klucher KM, Chow H, Reiser L, Fischer RL: **The AINTEGUMENTA gene of Arabidopsis required for ovule and female gametophyte development is related to the floral homeotic gene APETALA2.** *Plant Cell Online* 1996, **8**(2):137–153.
66. Mizukami Y, Fischer RL: **Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis.** *Proc Natl Acad Sci* 2000, **97**(2):942–947.
67. Nole-Wilson S, Krizek BA: **AINTEGUMENTA contributes to organ polarity and regulates growth of lateral organs in combination with YABBY Genes.** *Plant Physiol* 2006, **141**(3):977–987.
68. Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM: **The protein encoded by the Arabidopsis homeotic gene *agamous* resembles transcription factors.** *Nature* 1990, **346**(6279):35–39.
69. Pnueli L, Hareven D, Rounsley SD, Yanofsky MF, Lifschitz E: **Isolation of the tomato AGAMOUS gene TAG1 and analysis of its homeotic role in transgenic plants.** *Plant Cell Online* 1994, **6**(2):163–173.
70. Hong RL, Hamaguchi L, Busch MA, Weigel D: **Regulatory elements of the floral homeotic gene AGAMOUS identified by phylogenetic footprinting and shadowing.** *Plant Cell Online* 2003, **15**(6):1296–1309.
71. Pan IL, McQuinn R, Giovannoni JJ, Irish VF: **Functional diversification of AGAMOUS lineage genes in regulating tomato flower and fruit development.** *J Exp Bot* 2010, **61**(6):1795–1806.
72. Clark SE, Williams RW, Meyerowitz EM: **The CLAVATA1 gene encodes a putative receptor kinase that controls shoot and floral meristem size in Arabidopsis.** *Cell* 1997, **89**(4):575–585.
73. Clark SE, Running MP, Meyerowitz EM: **CLAVATA1, a regulator of meristem and flower development in Arabidopsis.** *Development* 1993, **119**(2):397–418.
74. Williams RW, Wilson JM, Meyerowitz EM: **A possible role for kinase-associated protein phosphatase in the Arabidopsis CLAVATA1 signaling pathway.** *Proc Natl Acad Sci* 1997, **94**(19):10467–10472.
75. Stahl Y, Simon R: **Receptor Kinases in Plant Meristem Development.** In *Receptor-Like Kinases in Plants*, Volume 13. Edited by Tax F, Kemmerling B. Berlin Heidelberg: Springer; 2012:23–39.
76. Ishiguro S, Watanabe Y, Ito N, Nonaka H, Takeda N, Sakai T, Kanaya H, Okada K: **SHEPHERD is the Arabidopsis GRP94 responsible for the formation of functional CLAVATA proteins.** *EMBO J* 2002, **21**(5):898–908.
77. Scofield S, Dewitte W, Murray JAH: **The KNOX gene SHOOT MERISTEMLESS is required for the development of reproductive meristematic tissues in Arabidopsis.** *Plant J* 2007, **50**(5):767–781.
78. Chen JJ, Janssen BJ, Williams A, Sinha N: **A gene fusion at a homeobox locus: alterations in leaf shape and implications for morphological evolution.** *Plant Cell Online* 1997, **9**(8):1289–1304.
79. Uchida N, Townsley B, Chung K-H, Sinha N: **Regulation of SHOOT MERISTEMLESS genes via an upstream-conserved noncoding sequence coordinates leaf development.** *Proc Natl Acad Sci* 2007, **104**(40):15953–15958.
80. Kwon CS: **WUSCHEL is a primary target for transcriptional regulation by SPLAYED in dynamic control of stem cell fate in Arabidopsis.** *Genes Dev* 2005, **19**(8):992–1003.
81. Laux T, Mayer KF, Berger J, Jurgens G: **The WUSCHEL gene is required for shoot and floral meristem integrity in Arabidopsis.** *Development* 1996, **122**(1):87–96.
82. Mayer KF, Schoof H, Haecker A, Lenhard M, Jurgens G, Laux T: **Role of WUSCHEL in regulating stem cell fate in the Arabidopsis shoot meristem.** *Cell* 1998, **95**:805–815.
83. Reinhardt D, Frenz M, Mandel T, Kuhlemeier C: **Microsurgical and laser ablation analysis of interactions between the zones and layers of the tomato shoot apical meristem.** *Development* 2003, **130**(17):4073–4083.
84. Wang X, Wang X-G, Ren J-P, Ma Y, Yin J: **Characterization of Tomato Transcription Factor WUSCHEL and Functional Study in Arabidopsis.** *J Integrative Agri* 2012, **11**(8):1257–1265.
85. Moussian B, Schoof H, Haecker A, Jurgens G, Laux T: **Role of the ZWILLE gene in the regulation of central shoot meristem cell fate during Arabidopsis embryogenesis.** *EMBO J* 1998, **17**(6):1799–809.
86. Kayes JM, Clark SE: **CLAVATA2, a regulator of meristem and organ development in Arabidopsis.** *Development* 1998, **125**(19):3843–3851.
87. Wang G, Zhang Z, Angenent GC, Fiers M: **New aspects of CLAVATA2, a versatile gene in the regulation of Arabidopsis development.** *J Plant Physiol* 2011, **168**(4):403–407.
88. Hord CLH: **The BAM1/BAM2 Receptor-Like Kinases are important regulators of arabidopsis early anther development.** *Plant Cell Online* 2006, **18**(7):1667–1680.
89. DeYoung BJ, Clark SE: **BAM receptors regulate stem cell specification and organ development through complex interactions with CLAVATA signaling.** *Genetics* 2008, **180**(2):895–904.
90. Lecomte L, Duffe P, Buret M, Servin B, Hospital F, Causse M: **Marker-assisted introgression of five QTLs controlling fruit quality traits into three tomato lines revealed interactions between QTLs and genetic backgrounds.** *TAG Theor Appl Genet* 2004, **109**(3):658–668.
91. Libault M, Stacey G: **Evolution of FW2.2-like (FWL) and PLAC8 genes in eukaryotes.** *Plant Signal Behav* 2010, **5**(10):1226–1228.
92. Song W-Y, Hörtensteiner S, Tomioka R, Lee Y, Martinoia E: **Common functions or only phylogenetically related? The large family of PLAC8 motif-containing/PCR genes.** *Mol Cells* 2011, **31**(1):1–7.
93. Alpert KB, Grandillo S, Tanksley SD: **fw2.2: a major QTL controlling fruit weight is common to both red- and green-fruited tomato species.** *TAG Theor Appl Genet* 1995, **91**(6):994–1000.
94. Liu J, Cong B, Tanksley SD: **Generation and Analysis of an Artificial Gene Dosage Series in Tomato to Study the Mechanisms by Which the Cloned Quantitative Trait Locus fw2.2 Controls Fruit Size.** *Plant physiology* 2003, **132**(1):292–299.
95. Cong B, Tanksley SD: **FW2.2 and cell cycle control in developing tomato fruit: a possible example of gene co-option in the evolution of a novel organ.** *Plant Mol Biol* 2006, **62**(6):867–880.
96. Li E, Wang S, Liu Y, Chen JG, Douglas CJ: **OVATE FAMILY PROTEIN4 (OFP4) interaction with KIN7 regulates secondary cell wall formation in Arabidopsis thaliana.** *The Plant Journal* 2011, **67**(2):328–341.
97. Hackbusch J, Richter K, Müller J, Salamini F, Uhrig JF: **A central role of Arabidopsis thaliana ovate family proteins in networking and subcellular localization of 3-aa loop extension homeodomain proteins.** *Proc Natl Acad Sci U S A* 2005, **102**(13):4908–4912.
98. Wang S, Chang Y, Guo J, Zeng Q, Ellis BE, Chen J-G: **Arabidopsis Ovate family proteins, a novel transcriptional repressor family, control multiple aspects of plant growth and development.** *PLoS One* 2011, **6**(8):e23896.
99. Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E: **A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit.** *Science* 2008, **319**(5869):1527–1530.
100. Wu S, Xiao H, Cabrera A, Meulia T, van der Knaap E: **SUN regulates vegetative and reproductive organ shape by changing cell division patterns.** *Plant Physiol* 2011, **157**(3):1175–1186.
101. Kaya H, Shibahara K-I, Taoka K-I, Iwabuchi M, Stillman B, Araki T: **FASCIATA genes for chromatin assembly factor-1 in Arabidopsis maintain the cellular organization of apical meristems.** *Cell* 2001, **104**(1):131–142.
102. Dennis MD, Browning KS: **Differential phosphorylation of plant translation initiation factors by Arabidopsis thaliana CK2 holoenzymes.** *J Biol Chem* 2009, **284**(31):20602–20614.



103. Burko Y, Geva Y, Refael-Cohen A, Shleizer-Burko S, Shani E, Berger Y, Halon E, Chuck G, Moshelion M, Ori N: **From organelle to organ: ZRIZI MATE-type transporter is an organelle transporter that enhances organ initiation.** *Plant Cell Physiol* 2011, **52**(3):518–527.
104. Fonné-Pfister R, Chemla P, Ward E, Girardet M, Kreuz KE, Honzatzko RB, Fromm HJ, Schär HP, Grütter MG, Cowan-Jacob SW: **The mode of action and the structure of a herbicide in complex with its target: binding of activated hydantocidin to the feedback regulation site of adenylosuccinate synthetase.** *Proc Natl Acad Sci* 1996, **93**(18):9431–9436.
105. Lim E, Bowles D: **A class of plant glycosyltransferases involved in cellular homeostasis.** *EMBO J* 2004, **23**(15):2915–2922.
106. Linder P: **Dead-box proteins: a family affair—active and passive players in RNP-remodeling.** *Nucleic Acids Res* 2006, **34**(15):4168–4180.
107. Panjabi P, Jagannath A, Bisht N, Padmaja KL, Sharma S, Gupta V, Pradhan A, Pental D: **Comparative mapping of *Brassica juncea* and *Arabidopsis thaliana* using Intron Polymorphism (IP) markers: homoeologous relationships, diversification and evolution of the A, B and C Brassica genomes.** *BMC Genomics* 2008, **9**(1):113.
108. Wicker T, Schlagenhauf E, Graner A, Close TJ, Keller B, Stein N: **454 sequencing put to the test using the complex genome of barley.** *BMC Genomics* 2006, **7**:275.
109. Sim S-C, Durstewitz G, Plieske J, Wieseke R, Ganai MW, Van Deynze A, Hamilton JP, Buell CR, Causse M, Wijeratne S, Francis DM: **Development of a large SNP genotyping array and generation of high-density genetic maps in tomato.** *PLoS One* 2012, **7**(7):e40563.
110. Lucatti A, van Heusden A, de Vos R, Visser R, Vosman B: **Differences in insect resistance between tomato species endemic to the Galapagos Islands.** *BMC Evol Biol* 2013, **13**(1):175.
111. Van Deynze A, Stoffel K, Buell CR, Kozik A, Liu J, van der Knaap E, Francis D: **Diversity in conserved genes in tomato.** *BMC Genomics* 2007, **8**(1):465.
112. Miller JC, Tanksley SD: **RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*.** *TAG Theor Appl Genet* 1990, **80**(4):437–448.
113. Blanca J, Cañizares J, Cordero L, Pascual L, Diez MJ, Nuez F: **Variation revealed by snp genotyping and morphology provides insight into the origin of the tomato.** *PLoS One* 2012, **7**(10):e48198.
114. Koenig D, Jiménez-Gómez JM, Kimura S, Fulop D, Chitwood DH, Headland LR, Kumar R, Covington MF, Devisetty UK, Tat AV, Tohge T, Bolger A, Schneeberger K, Ossowski S, Lanz C, Xiong G, Taylor-Teeple M, Brady SM, Pauly M, Weigel D, Usadel B, Fernie AR, Peng J, Sinha NR, Maloof JN: **Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato.** *Proc Natl Acad Sci U S A* 2013, **110**(28):E2655–E2662.
115. Labate JA, Sheffer SM, Balch T, Robertson LD: **Diversity and population structure in a geographic sample of tomato accessions.** *Crop Sci* 2011, **51**(3):1068–1079.
116. Gossmann TJ, Song B-H, Windsor AJ, Mitchell-Olds T, Dixon CJ, Kapralov MV, Filatov DA, Eyre-Walker A: **Genome wide analyses reveal little evidence for adaptive evolution in many plant species.** *Mol Biol Evol* 2010, **27**(8):1822–1832.
117. Städler T, Roselius K, Stephan W: **Genealogical footprints of speciation processes in wild tomatoes: demography and evidence for historical gene flow.** *Evolution* 2005, **59**(6):1268–1279.
118. Charlesworth D, Charlesworth B, Morgan MT: **The pattern of neutral molecular variation under the background selection model.** *Genetics* 1995, **141**(4):1619–1632.
119. Smith JM, Haigh J: **The hitch-hiking effect of a favourable gene.** *Genet Res* 1974, **23**(01):23–35.
120. Chakrabarti M, Zhang N, Sauvage C, Muñoz S, Blanca J, Cañizares J, Diez MJ, Schneider R, Mazourek M, McClelland J, Causse M, van der Knaap E: **A cytochrome P450 regulates a domestication trait in cultivated tomato.** *Proc Natl Acad Sci* 2013, **110**(42):17125–17130.
121. Albrechtsen A, Nielsen FC, Nielsen R: **Ascertainment biases in SNP chips affect measures of population divergence.** *Mol Biol Evol* 2010, **27**(11):2534–2547.
122. Manrique-Carpintero NC, Tokuhisa JG, Ginzberg I, Holliday JA, Veilleux RE: **Sequence diversity in coding regions of candidate genes in the glycoalkaloid biosynthetic pathway of wild potato species.** *G3: Genes Genomes Genomics* 2013, **3**(9):1467–1479.
123. Sim SC, Robbins MD, Deynze AV, Michel AP, Francis DM: **Population structure and genetic differentiation associated with breeding history and selection in tomato (*Solanum lycopersicum* L.).** *Heredity* 2010, **106**(6):927–935.
124. Pavlidis P, Metzler D, Stephan W: **Selective sweeps in multilocus models of quantitative traits.** *Genetics* 2012, **192**(1):225–239.
125. Chevin L-M, Hospital F: **Selective sweep at a quantitative trait locus in the presence of background genetic variation.** *Genetics* 2008, **180**(3):1645–1660.
126. Durbak AR, Tax FE: **CLAVATA signaling pathway receptors of Arabidopsis regulate cell proliferation in fruit organ formation as well as in meristems.** *Genetics* 2011, **189**(1):177–194.
127. Zhong S, Fei Z, Chen Y-R, Zheng Y, Huang M, Vrebalov J, McQuinn R, Gapper N, Liu B, Xiang J, Shao Y, Giovannoni JJ: **Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening.** *Nat Biotechnol* 2013, **31**:154–159.
128. Manning K, Tör M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, Seymour GB: **A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening.** *Nat Genet* 2006, **38**(8):948–952.
129. Brewer MT, Lang L, Fujimura K, Dujmovic N, Gray S, van der Knaap E: **Development of a controlled vocabulary and software application to analyze fruit shape variation in tomato and other plant species.** *Plant Physiol* 2006, **141**(1):15–25.
130. Roe BA: **Shotgun library construction for DNA sequencing.** *Methods Mol Biol* 2004, **255**:171–187.
131. Wiley G, Macmill S, Qu C, Wang P, Xing Y, White D, Li J, White JD, Domingo A, Roe BA: **Methods for Generating Shotgun and Mixed Shotgun/paired-end Libraries for the 454 dna Sequencer.** In *Current Protocols in Human Genetics*. John Wiley & Sons, Inc; 2001.
132. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell Scott B, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM: **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature* 2005, **437**(7057):376–380.
133. Mariette J, Noiroit C, Klopp C: **Assessment of replicate bias in 454 pyrosequencing and a multi-purpose read-filtering tool.** *BMC Res Notes* 2011, **4**(1):149.
134. Burland TG: **DNASTAR's Lasergene sequence analysis software.** *Molecular Methods* 1999, **132**:71–91.
135. Kent WJ: **BLAT—The BLAST-Like alignment tool.** *Genome Res* 2002, **12**(4):656–664.
136. Cingolani P, Platts A, Wang L, Coon M, Nguyen T, Wang L, Land S, Ruden D, Lu X: **A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3.** *Fly* 2012, **6**(2):1–13.
137. Hamilton JP, Sim S-C, Stoffel K, Van Deynze A, Buell CR, Francis DM: **Single Nucleotide polymorphism discovery in cultivated tomato via sequencing by synthesis.** *Plant Gen* 2012, **5**(1):17–29.
138. Sauvage C, Segura V, Bauchet G, Stevens R, Thi Do P, Nikoloski Z, Fernie AR, Causse M: **Genome Wide Association in tomato reveals 44 candidate loci for fruit metabolic traits.** *Plant Physiology* 2014, pp.114.241521. doi:http://dx.doi.org/10.1104/pp.114.241521.
139. Tajima F: **Statistical method for testing the neutral mutation hypothesis by dna polymorphism.** *Genetics* 1989, **123**(3):585–595.
140. Librado P, Rozas J: **DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.** *Bioinformatics* 2009, **25**:1451–1452.
141. Hutter S, Vilella A, Rozas J: **Genome-wide DNA polymorphism analyses using VarScan.** *BMC Bioinform* 2006, **7**(1):409.
142. Yang Z, Bielawski JP: **Statistical methods for detecting molecular adaptation.** *Trends Ecol Evol* 2000, **15**(12):496–503.
143. Goldman N, Yang Z: **A codon-based model of nucleotide substitution for protein-coding DNA sequences.** *Mol Biol Evol* 1994, **11**(5):725–736.
144. Yang Z: **PAML 4: phylogenetic analysis by maximum likelihood.** *Mol Biol Evol* 2007, **24**(8):1586–1591.
145. Weir BS, Cockerham CC: **Estimating F-statistics for the analysis of population structure.** *Evolution* 1984, **38**(6):1358–1370.
146. Falush D, Stephens M, Pritchard JK: **Inference of population structure using multilocus genotype data: dominant markers and null alleles.** *Mol Ecol Notes* 2007, **7**(4):574–578.
147. Evanno G, Regnaut S, Goudet J: **Detecting the number of clusters of individuals using the software structure: a simulation study.** *Mol Ecol* 2005, **14**(8):2611–2620.

148. Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, Carrecedo Á, Lareu M: An overview of STRUCTURE: applications, parameter settings and supporting software. *Frontiers in Genetics* 2013, **4**(98):1–13.
149. Hardy OJ, Vekemans X: SPAGeDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* 2002, **2**:618–620.
150. Loiselle BA, Sork VL, Nason J, Graham C: Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *Am J Bot* 1995, **82**:1420–1425.
151. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES: TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 2007, **23**(19):2633–2635.
152. Benjamini Y, Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc* 1995, **57**(1):289–300.
153. Steinbach D, Alaux M, Amselem J, Choisne N, Durand S, Flores R, Keliet A-O, Kimmel E, Lapalu N, Luyten I, Michotey C, Mohellibi N, Pommier C, Reboux S, Valdenaire D, Verdelet D, Quesneville H: GnpIS: an information system to integrate genetic and genomic data from plants and fungi. *Database* 2013, doi:10.1093/database/bat058.

doi:10.1186/s12870-014-0279-2

**Cite this article as:** Bauchet et al.: Genes involved in floral meristem in tomato exhibit drastically reduced genetic diversity and signature of selection. *BMC Plant Biology* 2014 **14**:279.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

